Spectral density of random graphs with topological constraints

2010 J. Phys. A: Math. Theor. 43 195002

(http://iopscience.iop.org/1751-8121/43/19/195002)

IOP PUBLISHING JOURNAL OF PHYSICS A: MATHEMATICAL AND THEORETICAL

J. Phys. A: Math. Theor. **43** (2010) 195002 (20pp) doi:10.1088/1751-8113/43/19/195002

Spectral density of random graphs with topological constraints

Tim Rogers[1], **Conrad Pérez Vicente**[2], **Koujin Takeda**[3] **and Isaac Pérez Castillo**[1]

[1] Department of Mathematics, King's College London, Strand, London, WC2R 2LS, UK
[2] Department de Física Fondamental, Facultat de Física, Universitat de Barcelona, 08028 Barcelona, Spain
[3] Department of Computational Intelligence and Systems Science, Tokyo Institute of Technology, Yokohama 226-8502, Japan

Received 18 November 2009, in final form 18 March 2010
Published 23 April 2010
Online at stacks.iop.org/JPhysA/43/195002

Abstract
The spectral density of random graphs with topological constraints is analysed using the replica method. We consider graph ensembles featuring generalized degree–degree correlations, as well as those with a community structure. In each case, a formal exact solution is found for the spectral density in the form of consistency equations depending on the statistical properties of the graph ensemble in question. We highlight the effect of these topological constraints on the resulting spectral density.

PACS numbers: 89.20.−a, 89.75.Hc, 46.65.+g, 87.18.Sn

(Some figures in this article are in colour only in the electronic version)

1. Introduction

Since Wigner's seminal work during the 1950s, random matrix theory (RMT) established itself as a cornerstone of modern theoretical physics, with innumerable applications (see, for example, [1] and references therein). One central problem of RMT is the determination of the mean spectral density of an ensemble of random matrices. In some cases, well-understood universal laws governing the mean spectral density have been known for some time (see, for instance, [2–4]). However, for ensembles of sparse matrices, i.e. matrices with many entries being zero, the picture is rather different. Studied first by Rodgers and Bray [5], the spectral density of real symmetric sparse random matrices has been extensively researched [6–12], although exact results have only been obtained relatively recently [13–17].

One area of research, in which sparse random matrices feature heavily, is in the modelling of real-world complex networks, with applications in fields as diverse as bioinformatics and finance. In the investigation of spectral density, the most commonly studied random graph

ensembles are those in which the degrees of neighbouring vertices become independent in the limit $N \to \infty$. This class includes the classical Erdős–Rényi (or Poissonian) random graphs [5–7, 10–12, 15] and graphs with a specified degree distribution [14, 17]. Unfortunately, these simple ensembles may not provide realistic models for real-world complex networks which can include features such as correlations between non-neighbouring vertices, or groups of vertices organized into highly connected communities [18]. Very few results have been obtained for random graph ensembles with more complex topologies, though there exists an approximation scheme for graphs with degree–degree correlations [8] and numerical investigations of some other ensembles [19].

In this paper, we extend the analysis of the spectral density of random graphs to complex networks with the features described above. In order to achieve this we consider graphs with hierarchically constrained topologies. Introduced in [20, 21], such ensembles can be tuned to more closely reflect the statistics of a real graph, whilst remaining in a form for which large N computations are tractable. We also propose a simple generalization of this ensemble to one featuring a community structure.

Applying the replica method, expressions are found for spectral densities of these ensembles in terms of the statistical properties of the graphs. For the purpose of clarity, we only show here the calculations for the spectral density of connectivity matrices; however, the techniques can just as well be applied to more general sparse random matrices. This is possible even in the case of non-Hermitian matrices following a scheme similar to that set out in [22].

The ensemble definitions are given in section 2. In the third section, we prepare for the calculation of the spectral density by first computing relevant statistical properties of the graph ensembles, gaining some insight to the problem on the way. Section 4 contains the replica calculation itself, both for the hierarchically constrained ensemble and for graphs with a community structure. Several applications are discussed in section 5, where analytical findings of the resulting spectral densities are compared with numerical diagonalizations. The final section contains a summary and discussion.

2. Ensemble definitions

Let $G = (V, E)$ be a graph on a set $V = \{1, \ldots, N\}$ of vertices and $E \subseteq V \times V$ of edges. The latter set is usually represented by the connectivity matrix C, whose entries are $c_{ij} = 1$ if $(i, j) \in E$, and $c_{ij} = 0$ otherwise $\forall\, i, j \in V$. We define the ℓth generalized degree of a vertex i, denoted as $k_i^{(\ell)}(C)$, to be the total number of walks of length ℓ starting at vertex i. In terms of the connectivity matrix, this is given by the recursive definition

$$k_i^{(0)}(C) = 1, \qquad k_i^{(\ell)}(C) = \sum_{j=1}^{N} c_{ij} k_j^{(\ell-1)}(C), \qquad \ell = 1, \ldots, L, \tag{1}$$

for some integer L. We denote $\boldsymbol{k}_i(C) = \left(k_i^{(1)}(C), \ldots, k_i^{(L)}(C)\right)$. Note that the first component of $\boldsymbol{k}_i(C)$ is simply the degree of vertex i, so we will usually refer to it as $k_i(C)$, rather than $k_i^{(1)}(C)$. The generalized degree of a particular vertex contains information about the generalized degrees of its neighbours. For instance, if vertex i has degree k and neighbours $\{j_1, \ldots, j_k\}$, then

$$k_i^{(\ell)}(C) = \sum_{t=1}^{k} k_{j_t}^{(\ell-1)}(C) \qquad \text{for} \quad \ell = 1, \ldots, L. \tag{2}$$

In what follows we study graph ensembles in which the generalized degrees are constrained. To achieve this, we follow the scheme set out in [20, 21] and define the weight of the graph ensemble as

$$W_N(C) = \prod_{i<j} \left[\frac{c}{N} Q(\boldsymbol{k}_i, \boldsymbol{k}_j) \delta_{c_{ij},1} + \left(1 - \frac{c}{N} Q(\boldsymbol{k}_i, \boldsymbol{k}_j)\right) \delta_{c_{ij},0} \right] \prod_{i=1}^{N} \delta_{\boldsymbol{k}_i(C),\boldsymbol{k}_i}, \tag{3}$$

where the $\{\boldsymbol{k}_i\}_{i=1}^N$ are taken to be arbitrary, c is the average $c = (1/N)\sum_{i=1}^N k_i$, and $Q(\boldsymbol{k}, \boldsymbol{k}')$ is a symmetric, non-negative function. Note from (3) that for each vertex i, its generalized degree $\boldsymbol{k}_i(C)$ is constrained to be precisely $\boldsymbol{k}_i$. For this reason, we also refer to the $\{\boldsymbol{k}_i\}_{i=1}^N$ as generalized degrees.

As mentioned in the introduction, this ensemble is designed to be amenable to analysis in the large N limit, whilst allowing the topology of the graphs to be tuned by choosing the generalized degrees $\{\boldsymbol{k}_i\}$ and the function $Q(\boldsymbol{k}, \boldsymbol{k}')$. This was demonstrated in [21] with the computation of the entropy of the ensemble, and in [23] the Ising model is analysed on such graphs in the simpler case of $L = 1$.

We will also consider an ensemble whose graphs feature community structures. In such graphs, vertices are organized into densely intra-connected clusters, also called modules or communities, with a sparse distribution of inter-community edges[4]. To incorporate this structure without sacrificing the solvability of the model, we propose a generalization of the previously introduced ensemble[5].

Consider a graph composed of N communities, each of size M (giving a total of NM vertices). We decompose the connectivity matrix $\mathcal{C}$ of this graph into three sets of smaller matrices: a single $N \times N$ connectivity matrix C with entries $c_{ij} = 1$ if communities i and j are connected, and zero otherwise; a collection of $M \times M$ matrices B_{ij} encoding the connections between vertices in communities i and j; and a collection of $M \times M$ matrices A_i specifying the internal connections of community i. For our ensemble, we take the B_{ij} and A_i to be drawn randomly and independently according to weights $\mu(B)$ and $\nu(A)$, respectively, whilst C is taken from the constrained generalized degree ensemble. All together, the weight for this graph ensemble can be written as follows:

$$W_{MN}^{\mathrm{com}}(\mathcal{C}) = \prod_{i<j} \left[\frac{c}{N} Q(\boldsymbol{k}_i, \boldsymbol{k}_j) \delta_{c_{ij},1} \mu(B_{ij}) + \left(1 - \frac{c}{N} Q(\boldsymbol{k}_i, \boldsymbol{k}_j)\right) \delta_{c_{ij},0} \right] \prod_{i=1}^{N} \nu(A_i) \delta_{\boldsymbol{k}_i(C),\boldsymbol{k}_i}. \tag{4}$$

For ease of calculation, we take μ to satisfy $\mu(B) = \mu(B^T)$.

Exploiting the bridge to statistical mechanics introduced by Edwards and Jones [27], we will compute the mean spectral density of the connectivity matrices of graphs from both ensembles in the limit $N \to \infty$ by the replica method. It will be instructive to first analyse the local statistics of the graphs in the same limit, as this will help us to simplify the subsequent analysis.

3. Asymptotic graph statistics

For the ensembles of graphs under study, the relevant statistical properties are captured by the following joint distribution:

$$P\left(\{\boldsymbol{q}_t\}_{t=1}^k, \boldsymbol{k}\right) = \lim_{N\to\infty} \left\langle \frac{k}{cN} \sum_{i=1}^{N} \delta_{\boldsymbol{k}_i(C),\boldsymbol{k}} \sum_{j_1<\cdots<j_k} \delta_{\boldsymbol{k}_{j_1}(C),\boldsymbol{q}_1} \cdots \delta_{\boldsymbol{k}_{j_k}(C),\boldsymbol{q}_k} c_{ij_1} \cdots c_{ij_k} \right\rangle_C, \tag{5}$$

[4] For a nice and complete review on community structures and its importance in complex networks see [18].

[5] Very recently, the spectral density of matrices with a particular type of modular form has been studied in [24–26].

where here, and hereafter, we use $\langle\cdots\rangle_C$ to denote the ensemble average. Note that (5) is the probability of finding a vertex whose neighbour has generalized degree $\boldsymbol{k}$ and is connected to k vertices with generalized degrees $\{\boldsymbol{q}_t\}_{t=1}^k$. Other relevant probability distributions can be obtained by marginalizing expression (5). In particular, by summing with respect to $\{\boldsymbol{q}_t\}_{t=1}^k$ and denoting the resulting distribution as $P(\boldsymbol{k})$, we obtain

$$P(\boldsymbol{k}) = \lim_{N\to\infty}\left\langle \frac{k}{cN}\sum_{i=1}^{N}\delta_{\boldsymbol{k},\boldsymbol{k}_i(C)}\right\rangle_C, \tag{6}$$

which is the probability of finding a vertex connected to a vertex with generalized degree $\boldsymbol{k}$ (see, for instance, [28]). Note that since in our ensemble the generalized degrees are constrained to the arbitrary values $\{\boldsymbol{k}_i\}_{i=1}^N$, we have that $P(\boldsymbol{k}) = p(\boldsymbol{k})k/c$ with $p(\boldsymbol{k}) = \lim_{N\to\infty}(1/N)\sum_{i=1}^N \delta_{\boldsymbol{k},\boldsymbol{k}_i}$ being the generalized degree distribution. Other probability distributions, however, may not have such straightforward forms.

An expression for the distribution (5) can be found through a saddle-point computation, the details of which will prove to be of great use in the later replica analysis of the spectral density, as our final equations for the spectral density will be written in terms of certain marginals of $P\big(\{\boldsymbol{q}_t\}_{t=1}^k, \boldsymbol{k}\big)$.

3.1. General calculation

To find an expression for $P\big(\{\boldsymbol{q}_t\}_{t=1}^k, \boldsymbol{k}\big)$, we need to calculate $\langle c_{ij_1}\cdots c_{ij_k}\rangle_C$. To do so, we introduce the generating function

$$Z_N(\boldsymbol{h}) = \sum_C W_N(C)\prod_{i<j}\mathrm{e}^{h_{ij}c_{ij}}, \tag{7}$$

with generating fields $\boldsymbol{h} = \{h_{ij}\}$. This allows us to write

$$\big\langle c_{ab_1}\cdots c_{ab_k}\big\rangle_C = \frac{1}{Z_N}\left.\frac{\partial^k}{\partial h_{ab_1}\cdots\partial h_{ab_k}}Z_N(\boldsymbol{h})\right|_{\boldsymbol{h}=\boldsymbol{0}}, \tag{8}$$

with $Z_N = Z_N(\boldsymbol{0})$. To carry out the calculation of $Z_N(\boldsymbol{h})$, we introduce a Fourier representation for the Kronecker delta constraints appearing in the definition of the weight (3):

$$\delta_{\boldsymbol{k}_i(C),\boldsymbol{k}_i} = \int_{-\pi}^{\pi}\frac{\mathrm{d}\boldsymbol{w}_i}{(2\pi)^L}\exp\left(\mathrm{i}\boldsymbol{w}_i\cdot\boldsymbol{k}_i - \mathrm{i}\sum_{\ell=1}^{L}w_i^{(\ell)}\sum_{j=1}^{N}c_{ij}k_j^{(\ell-1)}\right), \tag{9}$$

with $\boldsymbol{w}_i = \big(w_i^{(1)},\ldots,w_i^{(L)}\big)$. The sum over C may now be performed explicitly, obtaining for large N

$$\begin{aligned}\big\langle c_{ab_1}\cdots c_{ab_k}\big\rangle_C = {}& \frac{1}{Z_N}\left(\prod_{t=1}^{k}\frac{c}{N}Q\big(\boldsymbol{k}_a,\boldsymbol{k}_{b_t}\big)\right)\int_{-\pi}^{\pi}\left[\prod_{i=1}^{N}\frac{\mathrm{d}\boldsymbol{w}_i}{(2\pi)^L}\right]\mathrm{e}^{\mathrm{i}\sum_i\boldsymbol{w}_i\cdot\boldsymbol{k}_i}\\ &\times\exp\left[\frac{c}{2N}\sum_{i,j=1}^{N}Q(\boldsymbol{k}_i,\boldsymbol{k}_j)\big(\mathrm{e}^{-\mathrm{i}\sum_{\ell=1}^{L}(w_i^{(\ell)}k_j^{(\ell-1)}+w_j^{(\ell)}k_i^{(\ell-1)})}-1\big)\right]\\ &\times\exp\left[-\mathrm{i}\sum_{\ell=1}^{L}\left(w_a^{(\ell)}\sum_{t=1}^{k}k_{b_t}^{(\ell-1)}+k_a^{(\ell-1)}\sum_{t=1}^{k}w_{b_t}^{(\ell)}\right)\right].\end{aligned} \tag{10}$$

To integrate out the w-variables and apply saddle-point integration we first need to decouple terms comprising vertex indices. To achieve this, we introduce the order parameter

$$\phi(\boldsymbol{k},\boldsymbol{q}) = \frac{1}{N}\sum_{i=1}^{N}\delta_{\boldsymbol{k}_i,\boldsymbol{k}}\exp\left(-\mathrm{i}\sum_{\ell=1}^{L}w_i^{(\ell)}q^{(\ell-1)}\right), \tag{11}$$

allowing us to write the average $\langle c_{ab_1}\cdots c_{ab_k}\rangle_C$ as follows:

$$\langle c_{ab_1}\cdots c_{ab_k}\rangle_C = \left\langle \frac{\frac{k_a!}{c^{k_a}}\big[\prod_{j=1}^{k_a}\frac{k_{b_j}}{N}Q(\boldsymbol{k}_a,\boldsymbol{k}_{b_j})\big]\mathbb{I}_{\boldsymbol{k}_a}\big(\{\boldsymbol{k}_{b_j}\}_{j=1}^{k_a}\big)}{\sum_{\boldsymbol{q}_1,\ldots,\boldsymbol{q}_{k_a}}\psi(\boldsymbol{k}_a,\boldsymbol{q}_1)\cdots\psi(\boldsymbol{k}_a,\boldsymbol{q}_{k_a})\mathbb{I}_{\boldsymbol{k}_a}\big(\{\boldsymbol{q}_j\}_{j=1}^{k_a}\big)} \times\prod_{j=1}^{k_a}\frac{\sum_{\boldsymbol{q}_1,\ldots,\boldsymbol{q}_{k_{b_j}-1}}\psi\big(\boldsymbol{k}_{b_j},\boldsymbol{q}_1\big)\cdots\psi\big(\boldsymbol{k}_{b_j},\boldsymbol{q}_{k_{b_j}-1}\big)\mathbb{I}_{\boldsymbol{k}_{b_j}}\big(\boldsymbol{k}_a,\{\boldsymbol{q}_j\}_{j=1}^{k_{b_j}-1}\big)}{\sum_{\boldsymbol{q}_1,\ldots,\boldsymbol{q}_{k_{b_j}}}\psi\big(\boldsymbol{k}_{b_j},\boldsymbol{q}_1\big)\cdots\psi\big(\boldsymbol{k}_{b_j},\boldsymbol{q}_{k_{b_j}}\big)\mathbb{I}_{\boldsymbol{k}_{b_j}}\big(\{\boldsymbol{q}_j\}_{j=1}^{k_{b_j}}\big)}\right\rangle_{\Phi(\phi,\psi)}, \tag{12}$$

where we have taken $k = k_a$ and have introduced the indicator function

$$\mathbb{I}_{\boldsymbol{k}}\big(\{\boldsymbol{q}_t\}_{t=1}^{k}\big) = \prod_{\ell=1}^{L}\delta_{k^{(\ell)},\sum_{r=1}^{k}q_r^{(\ell-1)}}, \tag{13}$$

which enforces the relationship between the generalized degrees of the neighbours of a given vertex, as noted in (2). The measure in (12) is defined by

$$\langle\cdots\rangle_{\Phi(\phi,\psi)} = \frac{\int\{\mathrm{d}\phi\,\mathrm{d}\psi\}\,\mathrm{e}^{N\Phi(\phi,\psi)}(\cdots)}{\int\{\mathrm{d}\phi\,\mathrm{d}\psi\}\,\mathrm{e}^{N\Phi(\phi,\psi)}}, \tag{14}$$

with

$$\Phi(\phi,\psi) = -c\sum_{\boldsymbol{k},\boldsymbol{q}}\psi(\boldsymbol{k},\boldsymbol{q})\phi(\boldsymbol{k},\boldsymbol{q}) + \frac{c}{2}\sum_{\boldsymbol{k},\boldsymbol{q}}Q(\boldsymbol{k},\boldsymbol{q})\phi(\boldsymbol{k},\boldsymbol{q})\phi(\boldsymbol{q},\boldsymbol{k}) + \sum_{\boldsymbol{k}}p(\boldsymbol{k})\ln\sum_{\boldsymbol{q}_1,\ldots,\boldsymbol{q}_k}\psi(\boldsymbol{k},\boldsymbol{q}_1)\cdots\psi(\boldsymbol{k},\boldsymbol{q}_k)\mathbb{I}_{\boldsymbol{k}}\big(\{\boldsymbol{q}_t\}_{t=1}^{k}\big). \tag{15}$$

In the limit $N\to\infty$ this measure converges to a functional Dirac delta centred at the saddle point of $\Phi(\phi,\psi)$. Extremizing $\Phi(\phi,\psi)$, we find saddle-point equations

$$\psi(\boldsymbol{k},\boldsymbol{q}) = Q(\boldsymbol{q},\boldsymbol{k})\phi(\boldsymbol{q},\boldsymbol{k}), \tag{16a}$$

$$\phi(\boldsymbol{k},\boldsymbol{q}) = P(\boldsymbol{k})\frac{\sum_{\boldsymbol{q}_1,\ldots,\boldsymbol{q}_{k-1}}\psi(\boldsymbol{k},\boldsymbol{q}_1)\cdots\psi(\boldsymbol{k},\boldsymbol{q}_{k-1})\mathbb{I}_{\boldsymbol{k}}\big(\boldsymbol{q},\{\boldsymbol{q}_t\}_{t=1}^{k-1}\big)}{\sum_{\boldsymbol{q}_1,\ldots,\boldsymbol{q}_k}\psi(\boldsymbol{k},\boldsymbol{q}_1)\cdots\psi(\boldsymbol{k},\boldsymbol{q}_k)\mathbb{I}_{\boldsymbol{k}}\big(\{\boldsymbol{q}_t\}_{t=1}^{k}\big)}. \tag{16b}$$

Finally, returning to the definition of $P\big(\{\boldsymbol{q}_t\}_{t=1}^{k},\boldsymbol{k}\big)$, using the preceding result together with the saddle-point equations (16*a*) and (16*b*), we reach

$$\begin{aligned}P\big(\{\boldsymbol{q}_t\}_{t=1}^{k},\boldsymbol{k}\big) &= \lim_{N\to\infty}\frac{k}{cN}\sum_{i=1}^{N}\delta_{\boldsymbol{k}_i,\boldsymbol{k}}\sum_{j_1<\cdots<j_k}\delta_{\boldsymbol{k}_{j_1},\boldsymbol{q}_1}\ldots\delta_{\boldsymbol{k}_{j_k},\boldsymbol{q}_k}\langle c_{ij_1}\cdots c_{ij_k}\rangle_C\\ &= P(\boldsymbol{k})\frac{\psi(\boldsymbol{k},\boldsymbol{q}_1)\cdots\psi(\boldsymbol{k},\boldsymbol{q}_k)\mathbb{I}_{\boldsymbol{k}}\big(\{\boldsymbol{q}_t\}_{t=1}^{k}\big)}{\sum_{\boldsymbol{q}_1,\ldots,\boldsymbol{q}_k}\psi(\boldsymbol{k},\boldsymbol{q}_1)\cdots\psi(\boldsymbol{k},\boldsymbol{q}_k)\mathbb{I}_{\boldsymbol{k}}\big(\{\boldsymbol{q}_t\}_{t=1}^{k}\big)}.\end{aligned} \tag{17}$$

It is not our intention to quantify (17) as this requires the solution of the saddle-point equations, which is a cumbersome task for general values L. It is nonetheless interesting to keep in mind this result as it will help us to understand the later results for the spectral density. The following marginals will also be relevant for our subsequent discussion:

(i) The probability $P(\boldsymbol{q},\boldsymbol{k})$ of finding a pair of connected vertices with generalized degrees $\boldsymbol{k}$ and $\boldsymbol{q}$:

$$\begin{aligned}P(\boldsymbol{q},\boldsymbol{k}) &= \lim_{N\to\infty}\left\langle\frac{1}{cN}\sum_{i,j=1}^{N}\delta_{\boldsymbol{k}_i(C),\boldsymbol{k}}\,\delta_{\boldsymbol{k}_j(C),\boldsymbol{q}}\,c_{ij}\right\rangle_C\\ &= \sum_{\boldsymbol{q}_1,\ldots,\boldsymbol{q}_{k-1}}P\big(\{\boldsymbol{q}_t\}_{t=1}^{k-1},\boldsymbol{q},\boldsymbol{k}\big) = Q(\boldsymbol{k},\boldsymbol{q})\phi(\boldsymbol{k},\boldsymbol{q})\phi(\boldsymbol{q},\boldsymbol{k}),\end{aligned} \tag{18}$$

and its conditional distribution

$$P(\boldsymbol{q}|\boldsymbol{k}) = P(\boldsymbol{q},\boldsymbol{k})/P(\boldsymbol{k}) = Q(\boldsymbol{k},\boldsymbol{q})\phi(\boldsymbol{k},\boldsymbol{q})\phi(\boldsymbol{q},\boldsymbol{k})\left(\frac{k}{c}p(\boldsymbol{k})\right)^{-1}. \tag{19}$$

(ii) The conditional distribution $P\big(\{\boldsymbol{q}_t\}_{t=1}^{k}|\boldsymbol{k}\big)$:

$$\begin{aligned} P\big(\{\boldsymbol{q}_t\}_{t=1}^{k}\big|\boldsymbol{k}\big) &= P\big(\{\boldsymbol{q}_t\}_{t=1}^{k},\boldsymbol{k}\big)\big/P(\boldsymbol{k}) \\ &= \frac{\psi(\boldsymbol{k},\boldsymbol{q}_1)\cdots\psi(\boldsymbol{k},\boldsymbol{q}_k)\mathbb{I}_{\boldsymbol{k}}\big(\{\boldsymbol{q}_t\}_{t=1}^{k}\big)}{\sum_{\boldsymbol{q}_1,\ldots,\boldsymbol{q}_k}\psi(\boldsymbol{k},\boldsymbol{q}_1)\cdots\psi(\boldsymbol{k},\boldsymbol{q}_k)\mathbb{I}_{\boldsymbol{k}}\big(\{\boldsymbol{q}_t\}_{t=1}^{k}\big)}. \end{aligned} \tag{20}$$

(iii) The conditional distribution $P\big(\{\boldsymbol{q}_t\}_{t=1}^{k-1}\big|\boldsymbol{q},\boldsymbol{k}\big)$:

$$\begin{aligned} P\big(\{\boldsymbol{q}_t\}_{t=1}^{k-1}\big|\boldsymbol{q},\boldsymbol{k}\big) &= P\big(\{\boldsymbol{q}_t\}_{t=1}^{k},\boldsymbol{k}\big)\big/P(\boldsymbol{q},\boldsymbol{k}) \\ &= \frac{\psi(\boldsymbol{k},\boldsymbol{q}_1)\cdots\psi(\boldsymbol{k},\boldsymbol{q}_{k-1})\mathbb{I}_{\boldsymbol{k}}\big(\boldsymbol{q},\{\boldsymbol{q}_t\}_{t=1}^{k-1}\big)}{\sum_{\boldsymbol{q}_1,\ldots,\boldsymbol{q}_{k-1}}\psi(\boldsymbol{k},\boldsymbol{q}_1)\cdots\psi(\boldsymbol{k},\boldsymbol{q}_{k-1})\mathbb{I}_{\boldsymbol{k}}\big(\boldsymbol{q},\{\boldsymbol{q}_t\}_{t=1}^{k-1}\big)}. \end{aligned} \tag{21}$$

3.2. Case $L = 1$

It is instructive to consider the particular case of $L = 1$ as it has also been discussed in previous works (see, for instance, [21, 23]). In this case, the generalized degree $\boldsymbol{k}$ is simply the degree k. As we only constrain the degrees, the indicator functions in (16*b*) are identically one, giving

$$\phi(k,q) = \frac{P(k)}{\sum_q \psi(k,q)}, \qquad \psi(k,q) = Q(k,q)\phi(q), \tag{22}$$

with $P(k) = p(k)k/c$ being the degree distribution of the nearest neighbour of a vertex. Since the right-hand side of the first preceding equation is independent of q, so is $\phi(k,q)$, and therefore we write $\phi(k,q) = \phi(k)$. Besides, introducing $\psi(k) = \sum_q \psi(k,q)$, we reach

$$\phi(k) = \frac{P(k)}{\psi(k)} \qquad \text{and} \qquad \psi(k) = \sum_q Q(k,q)\phi(q). \tag{23}$$

From this, we obtain simplified expressions for (18) and (19):

$$P(q,k) = P(k)P(q)\frac{Q(k,q)}{\psi(k)\psi(q)} = P(k)\frac{\psi(k,q)}{\psi(k)}, \tag{24a}$$

$$P(q|k) = P(q)\frac{Q(k,q)}{\psi(k)\psi(q)} = \frac{\psi(k,q)}{\psi(k)}. \tag{24b}$$

Moreover, this also induces factorization in (20) and (21), giving

$$P\big(\{q_t\}_{t=1}^{k}\big|k\big) = \prod_{t=1}^{k} P(q_t|k) \qquad \text{and} \qquad P\big(\{q_t\}_{t=1}^{k-1}\big|q,k\big) = P\big(\{q_t\}_{t=1}^{k-1}\big|k\big). \tag{25}$$

Suppose further that we have a separable function $Q(k,q) = v(k)v(q)$. Then, from (23) it is easy to see that $\sum_q v(q)\phi(q) = \pm 1$, giving the solution $\psi(k) = \pm v(k)$. This implies in turn that $P(q|k) = P(q)$ as expected.

Having gained some understanding of the typical order parameters involved in the problem, we move on to the explicit calculation of the spectral density.

4. The spectral density of constrained graphs

Suppose C is the connectivity matrix of a random graph belonging to the ensemble under consideration. It is real and symmetric and hence has N real eigenvalues, which we denote by $\{\lambda_i^C\}_{i=1}^N$. The natural object of study is the mean spectral density in the limit $N \to \infty$:

$$\rho(\lambda) = \lim_{N\to\infty} \left\langle \frac{1}{N} \sum_{i=1}^{N} \delta\left(\lambda - \lambda_i^C\right) \right\rangle_C . \tag{26}$$

To compute (26), we first recast the problem in terms of a Gaussian integral. Following Edwards and Jones [27], one may write

$$\rho(\lambda) = -\lim_{\varepsilon\to 0^+} \lim_{N\to\infty} \frac{2}{\pi N} \operatorname{Im} \frac{\partial}{\partial\lambda} \langle \ln \mathcal{Z}_C(\lambda_\varepsilon)\rangle_C , \tag{27}$$

where $\lambda_\varepsilon = \lambda - \mathrm{i}\varepsilon$ and $\mathcal{Z}_C$ is given by

$$\mathcal{Z}_C(\lambda_\varepsilon) = \int \left[\prod_{i=1}^{N} \mathrm{d}x_i\right] \exp\left(-\mathrm{i}\frac{\lambda_\varepsilon}{2} \sum_{i=1}^{N} x_i^2 + \mathrm{i} \sum_{i<j} c_{ij} x_i x_j\right) . \tag{28}$$

Ignoring the imaginary units, this object is reminiscent of the partition function of a system of dynamical variables interacting on a graph, much like the Gaussian ferromagnetic model introduced in [29].

Reasoning along these lines, the tools of statistical mechanics can be brought to bear on the calculation of spectral density. In particular, the replica method has been frequently applied, leading either to approximative schemes [7, 10–12], or more recent exact solutions [14, 15, 26]. We proceed with the analysis for our ensemble, paying close attention to the impact of constraining the generalized degrees.

4.1. General calculation

To evaluate the average of the logarithm in (27), we apply the replica method, writing

$$\langle \ln \mathcal{Z}_C(\lambda_\varepsilon)\rangle_C = \lim_{n\to 0} \frac{1}{n} \ln \left\langle \mathcal{Z}_C^n(\lambda_\varepsilon)\right\rangle_C, \tag{29}$$

where the replicated partition function reads

$$\left\langle \mathcal{Z}_C^n(\lambda_\varepsilon)\right\rangle_C = \int \left[\prod_{i=1}^{N} \mathrm{d}\boldsymbol{x}_i\right] \exp\left(-\mathrm{i}\frac{\lambda_\varepsilon}{2} \sum_{i=1}^{N} \boldsymbol{x}_i^2\right) \left\langle \exp\left(\mathrm{i} \sum_{i<j} c_{ij} \boldsymbol{x}_i \cdot \boldsymbol{x}_j\right)\right\rangle_C , \tag{30}$$

with $\boldsymbol{x}_i = \left(x_i^{(1)}, \ldots, x_i^{(n)}\right)$ and $\mathrm{d}\boldsymbol{x}_i = \mathrm{d}x_i^{(1)} \cdots \mathrm{d}x_i^{(n)}$. As in the previous calculation, we use a Fourier representation for the Kronecker deltas in the weight (3) to enable us to perform the ensemble average explicitly, writing, for large N,

$$\begin{aligned}\left\langle \mathcal{Z}_C^n(\lambda_\varepsilon)\right\rangle_C = \frac{1}{Z_N} \int \left[\prod_{i=1}^{N} \mathrm{d}\boldsymbol{x}_i\right] \int_{-\pi}^{\pi} \left[\prod_{i=1}^{N} \frac{\mathrm{d}\boldsymbol{w}_i}{(2\pi)^L}\right] \mathrm{e}^{\mathrm{i}\sum_{i=1}^{N} \boldsymbol{w}_i\cdot\boldsymbol{k}_i - \frac{\mathrm{i}\lambda_\varepsilon}{2}\sum_{i=1}^{N} \boldsymbol{x}_i^2} \\ \times \exp\left[\frac{c}{2N} \sum_{i,j=1}^{N} Q(\boldsymbol{k}_i, \boldsymbol{k}_j)\left(\mathrm{e}^{-\mathrm{i}\sum_{\ell=1}^{L}(w_i^{(\ell)} k_j^{(\ell-1)} + w_j^{(\ell)} k_i^{(\ell-1)}) + \mathrm{i}\boldsymbol{x}_i\cdot\boldsymbol{x}_j} - 1\right)\right] .\end{aligned} \tag{31}$$

To decouple vertices we introduce the following order parameter:

$$\Phi(\boldsymbol{x}, \boldsymbol{k}, \boldsymbol{q}) = \frac{1}{N} \sum_{i=1}^{N} \delta_{\boldsymbol{k}_i,\boldsymbol{k}} \delta(\boldsymbol{x}_i - \boldsymbol{x}) \exp\left(-\mathrm{i} \sum_{\ell=1}^{L} w_i^{(\ell)} q^{(\ell-1)}\right) , \tag{32}$$

with $\boldsymbol{x} = (x^{(1)}, \ldots, x^{(n)})$ and $\delta(\boldsymbol{x}_i - \boldsymbol{x}) = \delta\big(x_i^{(1)} - x^{(1)}\big) \cdots \delta\big(x_i^{(n)} - x^{(n)}\big)$. After enforcing the order parameter using a functional Dirac delta and rearranging terms, we can write a compact expression for the replicated partition function

$$\big\langle \mathcal{Z}_C^n(\lambda_\varepsilon) \big\rangle_C = \int \{\mathrm{d}\Phi\, \mathrm{d}\Psi\}\, \mathrm{e}^{N\mathcal{F}(\Phi,\Psi)}, \tag{33}$$

where

$$\begin{aligned}\mathcal{F}(\Phi,\Psi) = -c\sum_{\boldsymbol{k},\boldsymbol{q}} \int \mathrm{d}\boldsymbol{x}\, \Psi(\boldsymbol{x},\boldsymbol{k},\boldsymbol{q})\Phi(\boldsymbol{x},\boldsymbol{k},\boldsymbol{q}) - \frac{c}{2}\sum_{\boldsymbol{k},\boldsymbol{q}} p(\boldsymbol{k})p(\boldsymbol{q})Q(\boldsymbol{k},\boldsymbol{q}) \\ + \frac{c}{2}\sum_{\boldsymbol{k},\boldsymbol{q}} Q(\boldsymbol{k},\boldsymbol{q}) \int \mathrm{d}\boldsymbol{x}\, \mathrm{d}\boldsymbol{y} \Phi(\boldsymbol{x},\boldsymbol{k},\boldsymbol{q})\Phi(\boldsymbol{y},\boldsymbol{q},\boldsymbol{k})\, \mathrm{e}^{\mathrm{i}\boldsymbol{x}\cdot\boldsymbol{y}} + \sum_{\boldsymbol{k}} p(\boldsymbol{k}) \ln \frac{c^k}{k!} \\ \times \int \mathrm{d}\boldsymbol{x}\, \mathrm{e}^{-\mathrm{i}\frac{\lambda_\varepsilon}{2}\boldsymbol{x}^2} \sum_{\boldsymbol{q}_1,\ldots,\boldsymbol{q}_k} \Psi(\boldsymbol{x},\boldsymbol{k},\boldsymbol{q}_1)\cdots\Psi(\boldsymbol{x},\boldsymbol{k},\boldsymbol{q}_k)\mathbb{I}_{\boldsymbol{k}}\big(\{\boldsymbol{q}_t\}_{t=1}^k\big).\end{aligned} \tag{34}$$

The integral (33) can now be evaluated by steepest descent. Extremizing $\mathcal{F}$ with respect to Φ and Ψ, we obtain the saddle-point equations

$$\Psi(\boldsymbol{x},\boldsymbol{k},\boldsymbol{q}) = Q(\boldsymbol{k},\boldsymbol{q}) \int \mathrm{d}\boldsymbol{y}\, \Phi(\boldsymbol{y},\boldsymbol{q},\boldsymbol{k})\, \mathrm{e}^{\mathrm{i}\boldsymbol{x}\cdot\boldsymbol{y}}, \tag{35a}$$

$$\Phi(\boldsymbol{x},\boldsymbol{k},\boldsymbol{q}) = P(\boldsymbol{k}) \frac{\mathrm{e}^{-\mathrm{i}\frac{\lambda_\varepsilon}{2}\boldsymbol{x}^2} \sum_{\boldsymbol{q}_1,\ldots,\boldsymbol{q}_{k-1}} \Psi(\boldsymbol{x},\boldsymbol{k},\boldsymbol{q}_1)\cdots\Psi(\boldsymbol{x},\boldsymbol{k},\boldsymbol{q}_{k-1})\mathbb{I}_{\boldsymbol{k}}\big(\boldsymbol{q},\{\boldsymbol{q}_t\}_{t=1}^{k-1}\big)}{\int \mathrm{d}\boldsymbol{y}\, \mathrm{e}^{-\mathrm{i}\frac{\lambda_\varepsilon}{2}\boldsymbol{y}^2} \sum_{\boldsymbol{q}_1,\ldots,\boldsymbol{q}_k} \Psi(\boldsymbol{y},\boldsymbol{k},\boldsymbol{q}_1)\cdots\Psi(\boldsymbol{y},\boldsymbol{k},\boldsymbol{q}_k)\mathbb{I}_{\boldsymbol{k}}\big(\{\boldsymbol{q}_t\}_{t=1}^k\big)}. \tag{35b}$$

The natural next step in the calculation is to make a replica symmetric ansatz. In the case of graphs with unconstrained topologies, the correct form for the order parameters has been recently established as a superposition of Gaussians [13–15]. With a modest amount of foresight we extend this to the correlated case by writing

$$\Phi(\boldsymbol{x},\boldsymbol{k},\boldsymbol{q}) = \phi(\boldsymbol{k},\boldsymbol{q}) \int \mathrm{d}\Delta\, \phi(\Delta|\boldsymbol{k},\boldsymbol{q}) \prod_{\alpha=1}^{n} \frac{\mathrm{e}^{-\frac{1}{2\Delta}(x^{(\alpha)})^2}}{\sqrt{2\pi\,\Delta}}, \tag{36a}$$

$$\Psi(\boldsymbol{x},\boldsymbol{k},\boldsymbol{q}) = \psi(\boldsymbol{k},\boldsymbol{q}) \int \mathrm{d}\Delta\, \psi(\Delta|\boldsymbol{k},\boldsymbol{q}) \prod_{\alpha=1}^{n} \frac{\mathrm{e}^{-\frac{\Delta}{2}(x^{(\alpha)})^2}}{\sqrt{2\pi/\Delta}}, \tag{36b}$$

where we assume that the densities $\psi(\Delta|\boldsymbol{k},\boldsymbol{q})$ and $\phi(\Delta|\boldsymbol{k},\boldsymbol{q})$ are normalized, i.e. $\int \mathrm{d}\Delta\, \psi(\Delta|\boldsymbol{k},\boldsymbol{q}) = 1$ and similarly for $\phi(\Delta|\boldsymbol{k},\boldsymbol{q})$. Note that the parameter Δ is generally a complex variable and $\mathrm{d}\Delta = \mathrm{d}\,\mathrm{Re}\Delta\, \mathrm{d}\,\mathrm{Im}\Delta$. Plugging the ansatz into the saddle-point equations and taking the replica limit $n \to 0$, we obtain

$$\psi(\Delta|\boldsymbol{k},\boldsymbol{q})\psi(\boldsymbol{k},\boldsymbol{q}) = Q(\boldsymbol{k},\boldsymbol{q})\phi(\Delta|\boldsymbol{q},\boldsymbol{k})\phi(\boldsymbol{q},\boldsymbol{k}), \tag{37a}$$

$$\begin{aligned}\phi(\Delta|\boldsymbol{k},\boldsymbol{q}) = \sum_{\boldsymbol{q}_1,\ldots,\boldsymbol{q}_{k-1}} \left\{ \frac{P(\boldsymbol{k})}{\phi(\boldsymbol{k},\boldsymbol{q})} \frac{\psi(\boldsymbol{k},\boldsymbol{q}_1)\cdots\psi(\boldsymbol{k},\boldsymbol{q}_{k-1})\mathbb{I}_{\boldsymbol{k}}\big(\boldsymbol{q},\{\boldsymbol{q}_t\}_{t=1}^{k-1}\big)}{\sum_{\boldsymbol{m}_1,\ldots,\boldsymbol{m}_k} \psi(\boldsymbol{k},\boldsymbol{m}_1)\cdots\psi(\boldsymbol{k},\boldsymbol{m}_k)\mathbb{I}_{\boldsymbol{k}}\big(\{\boldsymbol{m}_t\}_{t=1}^k\big)} \right\} \\ \times \int \left[\prod_{t=1}^{k-1} \mathrm{d}\Delta_t\, \psi(\Delta_t|\boldsymbol{k},\boldsymbol{q}_t)\right] \delta\left(\Delta - \frac{1}{\mathrm{i}\lambda_\varepsilon + \sum_{t=1}^{k-1}\Delta_t}\right).\end{aligned} \tag{37b}$$

Integrating with respect to Δ in (37*a*) and (37*b*) reveals

$$\psi(\boldsymbol{k},\boldsymbol{q}) = Q(\boldsymbol{k},\boldsymbol{q})\phi(\boldsymbol{q},\boldsymbol{k}), \tag{38a}$$

$$\phi(\boldsymbol{k},\boldsymbol{q}) = P(\boldsymbol{k})\frac{\sum_{\boldsymbol{q}_1,\ldots,\boldsymbol{q}_{k-1}} \psi(\boldsymbol{k},\boldsymbol{q}_1)\cdots\psi(\boldsymbol{k},\boldsymbol{q}_{k-1})\mathbb{I}_{\boldsymbol{k}}\left(\boldsymbol{q},\{\boldsymbol{q}_t\}_{t=1}^{k-1}\right)}{\sum_{\boldsymbol{m}_1,\ldots,\boldsymbol{m}_k} \psi(\boldsymbol{k},\boldsymbol{m}_1)\cdots\psi(\boldsymbol{k},\boldsymbol{m}_k)\mathbb{I}_{\boldsymbol{k}}\left(\{\boldsymbol{m}_t\}_{t=1}^{k}\right)}, \tag{38b}$$

and we recognize these equations as precisely the saddle-point equations (16*a*) and (16*b*) in the earlier calculation of the asymptotic graph statistics. This fact is not surprising since these objects depend only upon the generalized degrees $(\boldsymbol{k},\boldsymbol{q})$, and in this regard the order parameters in both calculations are the same. In turn, this also reveals that the expression between braces in (37*b*) is precisely the conditional distribution $P\left(\{\boldsymbol{q}_t\}_{t=1}^{k-1}|\boldsymbol{q},\boldsymbol{k}\right)$. All in all, we can write a self-consistency equation for the density $\psi(\Delta|\boldsymbol{q},\boldsymbol{k})$:

$$\psi(\Delta|\boldsymbol{q},\boldsymbol{k}) = \sum_{\boldsymbol{q}_1,\ldots,\boldsymbol{q}_{k-1}} P\left(\{\boldsymbol{q}_t\}_{t=1}^{k-1}|\boldsymbol{q},\boldsymbol{k}\right) \times \int\left[\prod_{t=1}^{k-1} \mathrm{d}\Delta_t\, \psi(\Delta_t|\boldsymbol{k},\boldsymbol{q}_t)\right]\delta\left(\Delta - \frac{1}{\mathrm{i}\lambda_\varepsilon + \sum_{t=1}^{k-1}\Delta_t}\right). \tag{39}$$

The simple form of this equation is possible thanks to the link between the order parameters in this calculation and the asymptotic form of the conditional distribution (21). Moreover, written in this way, the physical interpretation is clear: $\psi(\Delta|\boldsymbol{q},\boldsymbol{k})$ is the conditional density of the parameter Δ for a vertex of generalized degree $\boldsymbol{k}$, given that it has a neighbour of generalized degree $\boldsymbol{q}$.

All that remains now is to compute the spectral density. From (27) we can write

$$\begin{aligned}\rho(\lambda) &= -\lim_{\varepsilon\to 0^+}\lim_{n\to 0}\frac{2}{\pi n}\,\mathrm{Im}\,\frac{\partial}{\partial\lambda}\mathcal{F}(\Phi,\Psi)\\ &= \lim_{\varepsilon\to 0^+}\frac{1}{\pi}\mathrm{Re}\sum_{\boldsymbol{k}} p(\boldsymbol{k})\int \mathrm{d}\Delta\,\psi_{\mathrm{phys}}(\Delta|\boldsymbol{k})\Delta,\end{aligned} \tag{40}$$

with

$$\psi_{\mathrm{phys}}(\Delta|\boldsymbol{k}) = \sum_{\boldsymbol{q}_1,\ldots,\boldsymbol{q}_k} P\left(\{\boldsymbol{q}_t\}_{t=1}^{k}|\boldsymbol{k}\right)\int\left[\prod_{t=1}^{k}\mathrm{d}\Delta_t\,\psi(\Delta_t|\boldsymbol{k},\boldsymbol{q}_t)\right]\delta\left(\Delta - \frac{1}{\mathrm{i}\lambda_\varepsilon + \sum_{t=1}^{k}\Delta_t}\right). \tag{41}$$

The calculation is now complete. We have expressed the limiting mean spectral density for the connectivity matrices of graphs from the ensemble defined by (3) in terms of densities $\psi(\Delta|\boldsymbol{q},\boldsymbol{k})$; moreover, we have found that the consistency equation (39) for these densities is phrased simply in terms of the asymptotic generalized degree statistics of the graphs.

Although as this calculation has shown that the spectral density of the ensemble studied here is governed entirely by its generalized degree statistics, we should point out that this is by no means the general rule.

4.2. Case $L = 1$

Let us again consider the particular case $L = 1$. As before the generalized degree $\boldsymbol{k}$ is simply the degree k, and we have $P\left(\{q_t\}_{t=1}^{k-1}|q,k\right) = P\left(\{q_t\}_{t=1}^{k-1}|k\right)$. This implies that the left-hand side of (39) does not depend on q, that is $\psi(\Delta|q,k) = \psi(\Delta|k)$. This implies in turn on the right-hand side of (39) that $\psi(\Delta_t|k,q_t) = \psi(\Delta_t|q_t)$. Moreover, since $P\left(\{q_t\}_{t=1}^{k-1}|k\right) = \prod_{t=1}^{k-1}P(q_t|k)$ and upon defining $\varphi(\Delta|k) = \sum_q P(q|k)\psi(\Delta|q)$, we can write the following self-consistency equation for $\varphi(\Delta|k)$:

$$\varphi(\Delta|k) = \sum_q P(q|k)\int\left[\prod_{t=1}^{q-1}\mathrm{d}\Delta_t\,\varphi(\Delta_t|q)\right]\delta\left(\Delta - \frac{1}{\mathrm{i}\lambda_\varepsilon + \sum_{t=1}^{q-1}\Delta_t}\right). \tag{42}$$

The spectral density is given by

$$\rho(\lambda) = \lim_{\varepsilon\to 0^+} \frac{1}{\pi}\mathrm{Re}\sum_k p(k)\int \mathrm{d}\Delta\, \psi_{\mathrm{phys}}(\Delta|k)\Delta, \tag{43}$$

with

$$\psi_{\mathrm{phys}}(\Delta|k) = \int\left[\prod_{t=1}^{k}\mathrm{d}\Delta_t\,\varphi(\Delta_t|k)\right]\delta\left(\Delta - \frac{1}{\mathrm{i}\lambda_\varepsilon + \sum_{t=1}^{k}\Delta_t}\right). \tag{44}$$

The situation is further simplified if Q is taken to be separable, in which case $P(q|k) = P(q)$ and we recover the results of [14] for the uncorrelated case.

4.3. Graphs with community structure

We turn our attention to graphs with communities whose ensemble weight is given by (4). In this situation it is again possible to compute expressions for the mean spectral density of the ensemble using the replica method. Let us briefly outline the main features of the calculation.

Introducing N vectors $\{X_i\}_{i=1}^{N}$, of M components each, $X_i = (x_{i,1},\ldots,x_{i,M})$, one may write

$$\rho(\lambda) = -\lim_{\varepsilon\to 0^+}\lim_{N\to\infty}\frac{2}{\pi NM}\,\mathrm{Im}\,\frac{\partial}{\partial\lambda}\langle\ln\mathcal{Z}_{\mathcal{C}}(\lambda_\varepsilon)\rangle_{\mathcal{C}}, \tag{45}$$

where $\langle\cdots\rangle_{\mathcal{C}}$ denotes the ensemble average and

$$\mathcal{Z}_{\mathcal{C}}(\lambda_\varepsilon) = \int\left[\prod_{i=1}^{N}\mathrm{d}X_i\right]\exp\left(-\frac{\mathrm{i}}{2}\sum_{i=1}^{N}X_i\,(\lambda_\varepsilon I_M - A_i)\,X_i^T + \mathrm{i}\sum_{i<j}c_{ij}X_iB_{ij}X_j^T\right), \tag{46}$$

and $\mathrm{d}X_i = \prod_{m=1}^{M}\mathrm{d}x_{i,m}$. We consider the thermodynamic limit to be given by $N\to\infty$, whilst M remains fixed and finite. Treating X_i as individual vector-valued dynamical variables the calculation proceeds as usual via the replica method. The order parameters take the same form as in the previous calculation; however, with the replica symmetric ansatz parameterized by $M\times M$ matrices $\boldsymbol{\Delta}$. In the end, one obtains the following self-consistency equation:

$$\begin{aligned}\psi(\boldsymbol{\Delta}|\boldsymbol{q},\boldsymbol{k}) = \sum_{\boldsymbol{q}_1\ldots\boldsymbol{q}_{k-1}}P\big(\{\boldsymbol{q}_t\}_{t=1}^{k-1}\big|\boldsymbol{q},\boldsymbol{k}\big)\int\left[\prod_{t=1}^{k-1}\mathrm{d}\boldsymbol{\Delta}_t\psi(\boldsymbol{\Delta}_t|\boldsymbol{k},\boldsymbol{q}_t,)\,\mathrm{d}B_t\mu(B_t)\right]\\ \times\int\mathrm{d}A\nu(A)\delta\left(\boldsymbol{\Delta} - \left(\mathrm{i}(\lambda_\varepsilon I_M - A) + \sum_{t=1}^{k-1}B_t\boldsymbol{\Delta}_tB_t^T\right)^{-1}\right).\end{aligned} \tag{47}$$

The spectral density is recovered via

$$\rho(\lambda) = \lim_{\varepsilon\to 0^+}\frac{1}{\pi M}\,\mathrm{Re}\sum_{\boldsymbol{k}}p(\boldsymbol{k})\int\mathrm{d}\boldsymbol{\Delta}\,\psi_{\mathrm{phys}}(\boldsymbol{\Delta}|\boldsymbol{k})\,\mathrm{Tr}\,\boldsymbol{\Delta}, \tag{48}$$

where

$$\begin{aligned}\psi_{\mathrm{phys}}(\boldsymbol{\Delta}|\boldsymbol{k}) = \sum_{\boldsymbol{q}_1\ldots\boldsymbol{q}_k}P\big(\{\boldsymbol{q}_t\}_{t=1}^{k}\big|\boldsymbol{k}\big)\int\left[\prod_{t=1}^{k}\mathrm{d}\boldsymbol{\Delta}_t\psi(\boldsymbol{\Delta}_t|\boldsymbol{k},\boldsymbol{q}_t)\,\mathrm{d}B_t\mu(B_t)\right]\\ \times\int\mathrm{d}A\nu(A)\delta\left(\boldsymbol{\Delta} - \left(\mathrm{i}(\lambda_\varepsilon I_M - A) + \sum_{t=1}^{k}B_t\boldsymbol{\Delta}_tB_t^T\right)^{-1}\right).\end{aligned} \tag{49}$$

Note that in the case $M = 1$ (that is, when each community is a single vertex), we have $\nu(A) = \delta_{A,0}$ and $\mu(B) = \delta_{B,1}$, and we recover the result of the previous calculation. Also, for general M, taking $L = 1$ induces the same simplifications as observed previously.

5. Numerics

In general, the self-consistency equation (39) is not exactly solvable. However, a numerical solution may be efficiently obtained using population dynamics [14, 30]. Although we will consider mainly the case $L = 1$, we describe this numerical procedure for the general case. First, for all possible pairs of generalized degrees $(\boldsymbol{q}, \boldsymbol{k})$, each density $\psi(\Delta|\boldsymbol{q}, \boldsymbol{k})$ in (39) is represented by a population of $\mathcal{N}$ variables $\{\Delta_i(\boldsymbol{q}, \boldsymbol{k})\}_{i=1}^{\mathcal{N}}$. The following procedure is then repeated a predefined number of iteration steps:

(1) Choose a pair of generalized degrees $(\boldsymbol{q}, \boldsymbol{k})$ and a variable $\Delta_a(\boldsymbol{q}, \boldsymbol{k})$ uniformly at random from its population.
(2) Randomly select a set of generalized degrees $\{\boldsymbol{q}\}_{t=1}^{k-1}$ according to the distribution $P\left(\{\boldsymbol{q}_t\}_{t=1}^{k-1}\middle|\boldsymbol{q}, \boldsymbol{k}\right)$.
(3) Choose $k-1$ variables $\left\{\Delta_{\ell_1}(\boldsymbol{k}, \boldsymbol{q}_1), \ldots, \Delta_{\ell_{k-1}}(\boldsymbol{k}, \boldsymbol{q}_{k-1})\right\}$ uniformly at random from their populations.
(4) Assign $\left(\mathrm{i}\lambda_\varepsilon + \sum_{t=1}^{k-1} \Delta_{\ell_t}(\boldsymbol{k}, \boldsymbol{q}_t)\right)^{-1} \to \Delta_a(\boldsymbol{q}, \boldsymbol{k})$.

This procedure is adapted straightforwardly to find a numerical solution of $\psi_{\text{phys}}(\Delta|\boldsymbol{k})$ using (41), which is then used to calculate the spectral density from (40). To assess our results we compare our analytical findings with results from numerical diagonalization of graphs. Although the starting point of our calculations was the ensemble definition (3), the final equations are phrased only in terms of the resulting generalized degree statistics. It is therefore appropriate to compare our results directly to data coming from random graphs with prescribed generalized degree statistics, without concerning ourselves with intermediate step of determining a suitable choice of weight $W_N(C)$. For $L = 1$, we heuristically adapt the Steger and Wormald algorithm [31][6] to generate graphs with a given connected degree–degree distribution $P(k, k')$ in the following way: given a degree sequence $\overline{k} = (k_1, \ldots, k_N)$ with a number of edges $m = \frac{1}{2}\sum_{i=1}^{N} k_i$ iterate the following procedure:

(1) Let E be a set of assigned edges, $\widehat{k} = (\widehat{k}_1, \ldots, \widehat{k}_N)$ an N-tuple of integers.
(2) Initialize $E = \emptyset, \widehat{k} = \overline{k}$.
(3) Choose two vertices $v_i, v_j \in V$ with probability $p_{ij} \propto P(k_i, k_j)\widehat{k}_i\widehat{k}_j$ and $(v_i, v_j) \notin E$. Reduce $\widehat{k}_i, \widehat{k}_j$ by 1.
(4) Repeat step 3 until no more edges can be added to E.
(5) If $|E| < m$ report failure otherwise output graph.

The input for this algorithm is the degree sequence of the graphs to be generated; however, our results are expressed in terms of the degree distribution $p(k)$. We therefore need to generate degree sequences which are compatible with $p(k)$. We discuss two possibilities:

- *Random degree sequence*. For each instance, the degrees are randomly drawn from $p(k)$. There is a chance that no graph can be generated exactly fitting the resulting degree sequence $\overline{k}$. In this case, the degree sequence is said to be non-graphical. To deal with this, one may choose either to check the graphicality of $\overline{k}$ before generating graphs, or simply accept all graphs generated regardless of whether they fail step 5. There are various ways to check graphicality, for instance, a theorem of Erdős and Gallai states that a degree sequence $\overline{k}$ with $k_1 \geqslant \cdots \geqslant k_N$ and $\sum_{i=1}^{N} k_i$ even is graphical if and only if, for all $n = 1, \ldots, N-1$,

$$\sum_{i=1}^{n} k_i \leqslant n(n-1) + \sum_{i=n+1}^{N} \min\{k_i, n\}. \tag{50}$$

[6] We have also tried an adapted version of the algorithm suggested in [32] with similar results.

For a discussion on graphicality and the generation of random graphs, see [33].

- *Fixed degree sequence*. Select a set of positive integers $\{N, N_1, N_2, \ldots\}$ such that $N = \sum_k N_k$ and $p(k) \simeq N_k/N$. We then generate random degree sequences with N_1 vertices of degree one, N_2 vertices of degree 2, and so on.

In our experience, the adapted Steger–Wormald algorithm yields graphs with the desired properties and produces almost no failures, provided one selects the appropriate method of generating degree sequences.

5.1. Case $L = 1$. Correlated degrees

Let us consider a graph ensemble in which the degrees of neighbouring vertices are correlated, that is, $P(k, k')$ does not factorize. For the sake of simplicity, we consider graphs whose vertices can only have degrees 2, 3 and 4, with the following degree distribution $p(k)$ and conditional distribution $P(k'|k)$:

$$\begin{aligned} p(k) &= \tfrac{18}{37}\delta_{k,2} + \tfrac{4}{37}\delta_{k,3} + \tfrac{15}{37}\delta_{k,4}, \\ P(k'|k) &= \tfrac{2}{3}\delta_{k,2}\delta_{k',2} + \delta_{k,3}\delta_{k',3} + \tfrac{4}{5}\delta_{k,4}\delta_{k',4} + \tfrac{1}{5}\delta_{k,4}\delta_{k',2} + \tfrac{1}{3}\delta_{k,2}\delta_{k',4}. \end{aligned} \tag{51}$$

To compute the spectral density of the resulting ensemble, we solve the self-consistency equation (42) using population dynamics as described previously. In this particular case, for each value of $k \in \{2, 3, 4\}$ the corresponding density $\varphi(\Delta|k)$ is represented by a population of $\mathcal{N} = 10^4$ variables $\{\Delta_i(k)\}_{i=1}^{\mathcal{N}}$, which are iterated over 200 MC steps. The spectral density is then computed via (44) and (43). To obtain smooth results, the spectral density is averaged over a further 50 MC steps.

For comparison, we have also calculated the spectral density by numerically diagonalizing 1000 graphs of size $N = 2000$. In this case, each instance is produced by first generating a degree sequence $\overline{k} = (k_1, \ldots, k_N)$ according to $p(k)$ and then applying the adapted Steger–Wormald algorithm as described previously. We have checked that both methods of generating the degree sequences produce equivalent results for large samples and that the number of failures of the algorithm is negligible compared to the sample size.

The results of population dynamics and numerical diagonalization are presented in figure 1 which, apart from peaks at $\lambda = 3$ and $\lambda \simeq 3.7$ due to finite size effects, shows excellent agreement.

Note from the choice of $P(k'|k)$ that there is a bias towards edges between vertices of the same degree, which suggests that the spectral density should share some features with the spectral densities of regular graphs of degrees 2, 3 and 4. Indeed, there are peaks close to ± 2 and $\pm\sqrt{8}$, coming from peaks in the spectral density of regular graphs of degree 2 and 3, and the domain of the spectral density is approximately $[-2\sqrt{3}, 2\sqrt{3}]$, the domain of the spectral density of the regular graph of degree 4.

We consider next an ensemble with a power-law degree distribution and correlated degrees. Such models arise often in the study of real world complex networks. We have chosen $P(k, k') \propto \tau^{kk'}$, where $\tau < 1$, with a maximum value for k being $k_{\max}$. In the limits $k_{\max} \to \infty$ and $\tau \to 1$, this choice results in a power-law degree distribution with exponent 2, though for the purpose of simulations, we will take $\tau = 0.999$, and keep $k_{\max}$ finite.

We use the previously explained algorithm to generate graphs with distribution $P(k, k')$. For graphs of size $N = 2000$, it is necessary to take a rather low maximum degree $k_{\max} = 45(\simeq\sqrt{N})$; if $k_{\max}$ is taken any larger, additional correlations occur between high degree vertices [34], and the failure rate becomes unacceptable. Alternatively, we could have larger values of $k_{\max}$ by increasing the graph size, but that would make the numerical

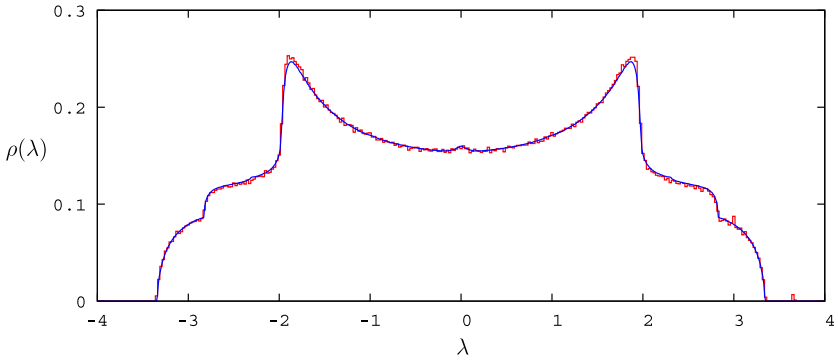


Figure 1. Comparison of the results of population dynamics (blue line) and direct diagonalization (red histogram) for the choice of $P(k'|k)$ given in (51). To construct the histogram, we use the adapted Steger–Wormald algorithm to generate 1000 graphs of size $N = 2000$. The degree sequences were generated randomly according to $p(k)$ given in (51) and only eight failures were reported.

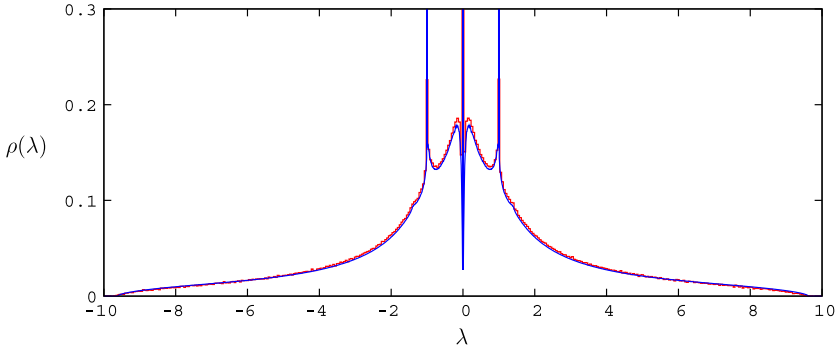


Figure 2. Comparison of the results of population dynamics (blue line) and direct diagonalization (red histogram) for the choice $P(k, k') \propto \tau^{kk'}$, where $\tau = 0.999$ and $k_{\max} = 45$. The adapted Steger–Wormald algorithm was used to generate 500 graphs of size $N = 2000$, whose eigenvalues were used to construct the histogram.

diagonalization computationally expensive. The degree sequences are generated randomly and checked for graphicality before generating the graph.

The spectral density can again be computed via population dynamics. Figure 2 shows a comparison between the results of population dynamics and a histogram of eigenvalues from 500 random graphs of size $N = 2000$. In this case, we take a population of $\mathcal{N} = 10^3$ variables $\{\Delta_i(k)\}_{i=1}^{\mathcal{N}}$, which are iterated over 100 MC steps. The spectral density is then averaged over a further 50 MC steps.

As before some of the salient features of the spectral density can be intuitively explained in terms of the underlying graph structure. First we note that the spectral density for $k_{\max} = 45$ and $\tau = 0.999$ has a bounded support as expected due to the Perron–Frobenius theorem, but as we take $k_{\max} \to \infty$ and $\tau \to 1$, the mean degree diverges and the spectral density presents heavy tails. It is interesting to analyse the contribution of vertices of high degree to the spectral density. As mentioned in [8], for such vertices it may be sufficient to consider only the mean behaviour of the neighbouring vertices. In this effective medium approximation

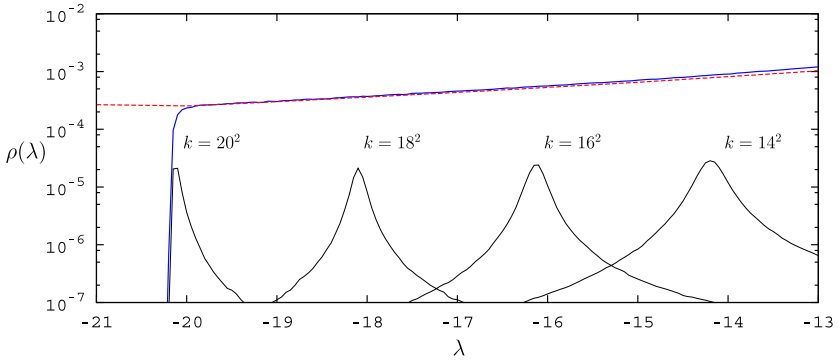


Figure 3. Detail of the tail of the spectral density for the choice $P(k, k') \propto \tau^{kk'}$, where $\tau = 0.999$ and $k_{\max} = 400$. The continuous blue line is the full result from population dynamics, with the labelled black lines being isolated contributions from vertices of high degree. The dashed red line shows an estimate for the tail derived from the effective medium approximation.

(EMA) the approximate behaviour of the tails for very large $|\lambda|$ reads $\rho(\lambda) \simeq 2k_\lambda p(k_\lambda)/|\lambda|$, where $k_\lambda = \lambda^2 + \mathcal{O}(1)$. A more rigorous analysis in [35] states that the largest eigenvalues of graphs with heavy-tailed degree distributions occur close to the square roots of the largest degrees.

In figure 3, we show the results of population dynamics simulations in the tail of the spectral density for a much larger maximum degree of $k_{\max} = 400$. The approximate curve given by the EMA gives a reasonable fit with the result of the simulation and, as expected, the density drops dramatically shortly after $\sqrt{k_{\max}} = 20$. The contributions to the density coming from high degree vertices can be isolated in the output of the population dynamics algorithm; we have included in figure 3 contributions from several high degrees k, each of which exhibits sharp peak close to $\sqrt{k}$.

The other main feature of the density shown in figure 2 is the presence of Dirac delta peaks at -1, 0 and 1 whose weight may be bounded by using the distributions $p(k)$ and $P(k|k')$. For instance, the weight to the peaks at ± 1 has contributions from connected pair of vertices of degree 1, which for the choice $k_{\max} = 45$ and $\tau = 0.999$ have a likelihood of $p(1)P(1|1)/2 = 0.0031$, not far from the exact weight of 0.0037 obtained from numerical diagonalization. A similar intuitive argument can be used to obtain a bound for the weight of the Dirac delta peak at zero, whose appearance is due to dead-end vertices [8, 36].

5.2. *Case $L = 2$. Levels of approximation*

Suppose that the exact knowledge of a graph ensemble is reduced solely to a set of statistical properties captured by, for instance, the degree distribution $p(k)$ and the conditional distribution $P(k|k')$. Whilst in a few cases such quantities suffice to fully characterize the graph ensemble, this is not generally true. We would like to understand in which way the lack of more accurate information affects the spectral density.

With this in mind let us consider a graph ensemble with generalized degrees $\{\boldsymbol{k}_i\}_{i=1}^N$ such that $\boldsymbol{k}_i \in \{\kappa_2, \kappa_3, \kappa_4\}$, where

$$\kappa_2 = \begin{pmatrix} 2 \\ 7 \end{pmatrix}, \qquad \kappa_3 = \begin{pmatrix} 3 \\ 6 \end{pmatrix}, \qquad \kappa_4 = \begin{pmatrix} 4 \\ 8 \end{pmatrix}. \tag{52}$$

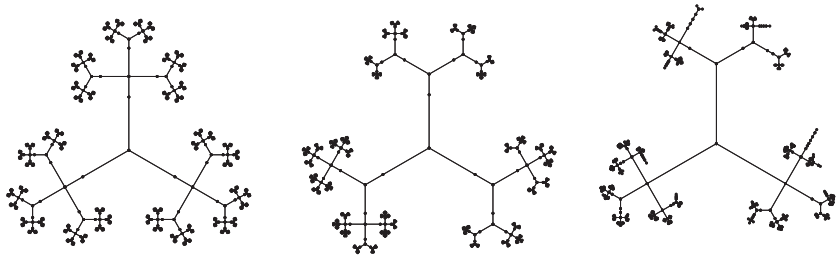


Figure 4. Left: a typical neighbourhood of a vertex of degree 3 of a random graph specified by (52). Middle: a neighbourhood of a vertex of degree 3 in a random graph with degree distribution (53*a*) and $P(k|k')$ given by (53*b*). Right: a neighbourhood of a vertex of degree 3 in an uncorrelated random graph with degree distribution (53*a*).

This ensemble is composed of graphs with vertices of degrees 2, 3 and 4. Moreover, those vertices of degree 2 must be connected to one vertex of degree 3 and one of degree 4, and those of degrees 3 and 4 must be connected to vertices of degree 2 only. A portion of such graph is shown in the leftmost part of figure 4.

For this graph ensemble, a quick counting argument gives the following expressions for $p(k)$ and $P(k|k')$:

$$p(k) = \frac{12}{19}\delta_{k,2} + \frac{4}{19}\delta_{k,3} + \frac{3}{19}\delta_{k,4}, \tag{53a}$$

$$P(k|k') = \delta_{k,2}\delta_{k',4} + \delta_{k,2}\delta_{k',3} + \left(\frac{1}{2}\delta_{k,3} + \frac{1}{2}\delta_{k,4}\right)\delta_{k',2}, \tag{53b}$$

knowledge of which does not fully characterize the graph ensemble. Note that this example is such that the set of self-consistency equations (39) can be solved exactly. To do so, we first need the conditional distribution $P\left(\{\boldsymbol{q}_t\}_{t=1}^{k-1}\middle|\boldsymbol{q}, \boldsymbol{k}\right)$ which in this case reads

$$P(\kappa_3|\kappa_4, \kappa_2) = 1, \qquad P(\kappa_4|\kappa_3, \kappa_2) = 1,$$
$$P(\kappa_2, \kappa_2|\kappa_2, \kappa_3) = 1, \qquad P(\kappa_2, \kappa_2, \kappa_2|\kappa_2, \kappa_4) = 1,$$

or zero otherwise. This results in a set of self-consistency equations for the densities $\{\psi(\Delta|\kappa_3, \kappa_2), \psi(\Delta|\kappa_2, \kappa_3), \psi(\Delta|\kappa_4, \kappa_2), \psi(\Delta|\kappa_2, \kappa_4)\}$ that admits a solution of the type $\psi(\Delta|\kappa_a, \kappa_b) = \delta(\Delta - \Delta_{a,b})$, with $\Delta_{a,b}$ obeying a simple set of algebraic equations

$$\Delta_{3,2} = \frac{1}{\mathrm{i}\lambda_\varepsilon + \Delta_{2,4}}, \quad \Delta_{4,2} = \frac{1}{\mathrm{i}\lambda_\varepsilon + \Delta_{2,3}}, \quad \Delta_{2,3} = \frac{1}{\mathrm{i}\lambda_\varepsilon + 2\Delta_{3,2}}, \quad \Delta_{2,4} = \frac{1}{\mathrm{i}\lambda_\varepsilon + 3\Delta_{4,2}}. \tag{54}$$

Finally, to find an expression for the spectral density, we first need the distribution $P\left(\{\boldsymbol{q}_t\}_{t=1}^{k}\middle|\boldsymbol{k}\right)$ which in this case reads

$$P(\kappa_3, \kappa_4|\kappa_2) = \tfrac{1}{2}, \qquad P(\kappa_4, \kappa_3|\kappa_2) = \tfrac{1}{2},$$
$$P(\kappa_2, \kappa_2, \kappa_2|\kappa_3) = 1, \qquad P(\kappa_2, \kappa_2, \kappa_2, \kappa_2|\kappa_4) = 1,$$

or zero otherwise. This yields

$$\rho(\lambda) = \lim_{\varepsilon\to 0^+} \frac{1}{19\pi}\,\mathrm{Re}\left[\frac{12}{\mathrm{i}\lambda_\varepsilon + \Delta_{2,3} + \Delta_{2,4}} + \frac{4}{\mathrm{i}\lambda_\varepsilon + 3\Delta_{3,2}} + \frac{3}{\mathrm{i}\lambda_\varepsilon + 4\Delta_{4,2}}\right]. \tag{55}$$

Upon solving (54), plugging the solutions into (55) and carefully analysing the poles, we can write

$$\rho(\lambda) = \frac{5}{19}\delta(\lambda) + \frac{1}{19}\delta(\lambda+\sqrt{3}) + \frac{1}{19}\delta(\lambda-\sqrt{3}) + \frac{12|2\lambda^2-7|\sqrt{-25-\lambda^2(-7+\lambda^2)(14-7\lambda^2+\lambda^4)}}{19\pi|\lambda(\lambda^2-4)(\lambda^2-7)(\lambda^2-3)|}\mathbb{I}_{\mathcal{R}}(\lambda), \tag{56}$$

with $\mathcal{R} = \left[\lambda_+^-, \lambda_-^-\right] \cup \left[\lambda_-^+, \lambda_+^+\right]$ and

$$\lambda_\sigma^\mu = \tfrac{1}{2}\sqrt{14 + 2\mu\sqrt{21+8\sigma\sqrt{6}}}, \tag{57}$$

with $\sigma, \mu \in \{-,+\}$ and where $\mathbb{I}_{\mathcal{R}}(\lambda) = 1$ if $|\lambda| \in \mathcal{R}$ or zero otherwise.

This example was chosen specifically to keep the local structure of the graphs deterministic and thus make equations (39) exactly solvable in the case $L = 2$. If we reduce the statistical information we have about the ensemble, either by taking $L = 1$, or assuming the degrees to be uncorrelated, the local graph structure becomes random, and the resulting expressions are no longer exactly solvable. Figure 4 shows the randomizing effect of these simplifications.

We have used population dynamics to compute the spectral density of random graphs with degree distribution (53*a*) and degree–degree correlations (53*b*), as well as uncorrelated random graphs with degree distribution (53*a*). For each degree, the population size is $\mathcal{N} = 10^4$, iterated over 200 MC steps. The spectral density is then averaged over 50 MC steps. The results are shown in figures 5(*b*) and (*c*), alongside histograms of the eigenvalues of 1000 graphs of size $N = 1900$, generated using the adapted Steger–Wormald algorithm[7]. Figure 5(*a*) shows the exact spectral density given by (56), plotted alongside a histogram of eigenvalues obtained from randomly generated graphs of that type.

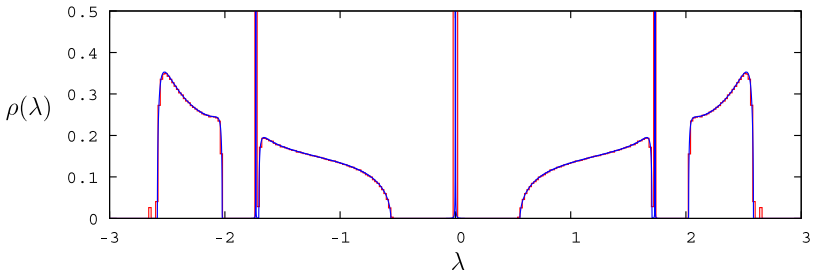
In figure 5, the effects of reducing knowledge of an ensemble (and hence increasing randomness) are clearly visible. In addition to a general smoothing effect, which one might expect, the most striking feature is the appearance of gaps in the spectral density. When the ensemble is fully specified by (52), the continuous part of the spectral density is divided into four disjoint components. When one specifies only the degree distribution and degree–degree correlations, the number of components reduces to two, and when only the degree distribution is known, there is no gap in the density at all.

The appearance of the gap can be traced back to the periodicity in the random graph ensembles (see [37]). For instance, in the original ensemble a walker moving away from a central vertex will repeatedly visit vertices of degree sequence $\{\ldots, 2, 4, 2, 3, 2, 4, 2, 3, \ldots\}$ of periodicity 4 which is also manifestly explicit in equations (54). In the case that the degree–degree correlations are specified, whilst the degree sequence of a walk is now random and therefore not strictly periodic, it is still true that every other vertex visited will have degree 2 (see figure 4), periodicity which is enough to split the spectral density into two components. In the last case, where only the degree distribution is known, the sequence is fully random (see figure 4), and the resulting spectral density has no gap.

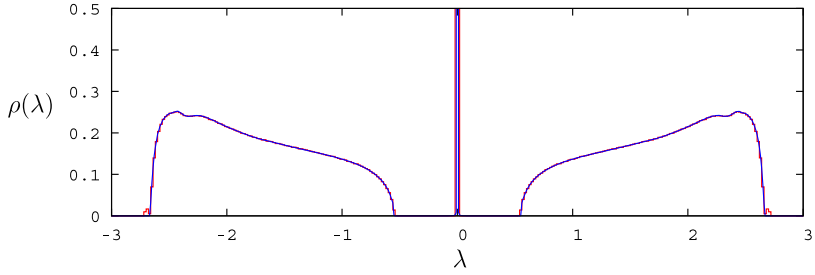
5.3. Community structure

The population dynamics algorithm used to solve (39) is easily adapted to solve the equivalent self-consistency equation (47) for the communities model; one simply initializes populations of $M \times M$ matrices $\boldsymbol{\Delta}$ and updates them according to (47).

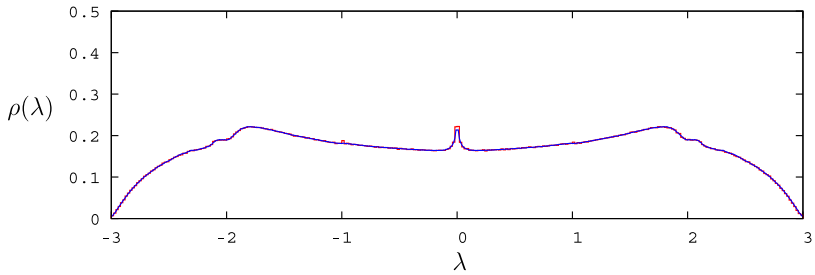
[7] It is necessary to slightly modify the adapted Steger–Wormald algorithm to make sure the constraints specified by (52) are met.



(a) Comparison of the exact spectral density given by equation (56) (blue line) and direct diagonalisation (red histogram). We have checked the isolated peaks at $\lambda \simeq \pm\sqrt{7}$ are due to finite size effects.



(b) Comparison of population dynamics with degree distribution (56a) and degree-degree correlations (53b) (blue line) and direct diagonalisation (red histogram). To visualise the Dirac delta peak we have taken $\varepsilon = 10^{-6}$.



(c) Comparison of population dynamics with degree distribution only (blue line) and direct diagonalisation (red histogram).

Figure 5. Comparison of theoretical results (blue lines) and direct diagonalization (red histograms) for the different levels of approximation to the ensemble specified by (52). In plot (a), the blue line shows the exact result for the spectral density, given by (56); in plots (b) and (c) the result of population dynamics is shown. To generate graphs we use the adapted Steger–Wormald algorithm explained in the text. In each case, 1000 graphs of size $N = 1900$ were generated and diagonalized.

The presence of communities in a graph typically results in a very different spectral density; to illustrate this, we consider a simple choice for the community structure ensemble. Suppose we have communities given by the complete graph on M vertices, connected in a Poissonian random graph of average degree c, in which connected communities are joined by a single randomly drawn edge. In the weight (4), this corresponds to the choices $Q(\boldsymbol{k}_i, \boldsymbol{k}_j) = 1$ for all i, j, $\nu(A) = \big(\prod_a \delta_{A_{aa},0}\big)\big(\prod_{a\neq b} \delta_{A_{ab},1}\big)$, and $\mu(B) = M^{-2} \sum_{a,b} \delta_{B_{ab},1} \prod_{(c,d)\neq(a,b)} \delta_{B_{cd},0}$.

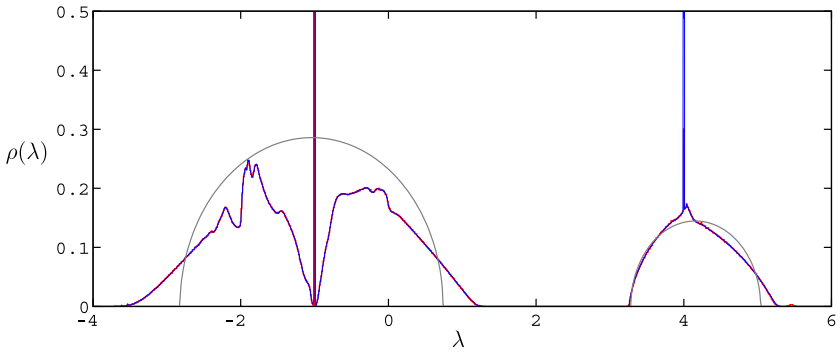


Figure 6. Comparison of the results of population dynamics (blue line) and direct diagonalization (red histogram) for the community structure ensemble described in the text. The grey curve shows the high connectivity limit for this ensemble.

Taking $M = 5$ and $c = 5$, we use population dynamics to solve (47) for this ensemble; the spectral density is then computed using (48). To compare with the results of direct diagonalization, 1000 graphs of size $N = 5000$ were generated. A histogram of their eigenvalues, alongside the result of population dynamics, is shown in figure 6.

It is well known that in the limit $c \to \infty$ the spectral density of a Poissonian random graph with average degree c, and edges of weight $1/\sqrt{c}$, converges to Wigner's semi-circular distribution [5, 13]. We can compute a generalization of this result for the community ensemble considered here through an appropriate treatment of the self-consistency equation (47). It is necessary to re-weight the edges between communities, in order to keep the spectral density bounded as $c \to \infty$, we take $\mu(B) = M^{-2} \sum_{a,b} \delta_{B_{ab},\sqrt{M/c}} \prod_{(c,d)\neq(a,b)} \delta_{B_{cd},0}$. Keeping only the terms relevant in the $c \to \infty$ limit, we obtain an expression for the mean $\mathbf{\Delta}$,

$$\langle \mathbf{\Delta} \rangle = (\mathrm{i}(\lambda_\varepsilon I_M - K_M) + c\langle B \langle \mathbf{\Delta} \rangle B^T \rangle_B)^{-1}, \tag{58}$$

where K_M is the connectivity matrix of the complete graph on M vertices. For the above choice of $\mu(B)$, we have $\langle B\mathbf{\Delta}B^T \rangle = (1/cM)\Delta I_M$, where $\Delta = (1/M)\,\mathrm{Tr}\langle \mathbf{\Delta} \rangle$. Diagonalizing K_M, we obtain a cubic equation for Δ,

$$\Delta = \left(\frac{M-1}{M}\right) \frac{1}{\mathrm{i}\lambda_\varepsilon + \Delta + \mathrm{i}} + \left(\frac{1}{M}\right) \frac{1}{\mathrm{i}\lambda_\varepsilon + \Delta - (M-1)\mathrm{i}}. \tag{59}$$

In the case $M = 5$, we solve (59) to find the following expression:

$$\rho(\lambda) = \frac{\sqrt{3}}{2\pi} \left| u - \frac{\lambda^2 - 3\lambda + 18}{9u} \right| \mathbb{I}_{\mathcal{D}}(\lambda)\,, \tag{60}$$

where

$$u = \left| \tfrac{1}{27}\lambda^3 - \tfrac{1}{6}\lambda^2 - 2\lambda + 4 + \tfrac{1}{18}\sqrt{3d} \right|^{1/3}, \tag{61}$$

and $d = -25\lambda^4 + 156\lambda^3 + 72\lambda^2 - 1296\lambda + 864$. The domain is given by $\mathbb{I}_{\mathcal{D}}(\lambda) = 1$ if $d > 0$, and zero otherwise.

6. Conclusions

Past work on the spectral density of random graphs has typically been confined to simple ensembles in which at most the degree distribution is specified and even then exact results

have only been obtained relatively recently. At the same time, the field of complex networks has gained a great deal of attention from all over the scientific community. Before the analysis of spectral density can be used to provide insights into the behaviour and characterization of complex networks, progress must first be made to expand the class of ensembles for which exact results are obtainable.

In this paper, we have sought to do just this through the calculation of the spectral density of random graphs with constrained topologies. Complex correlations between the degrees of non-neighbouring vertices are incorporated in the constrained generalized degree ensemble, and we also introduce a simple extension of this model to one featuring a community structure. The important statistical properties of the constrained degree ensemble are captured in the distribution (5), which we compute via a saddle point analysis in the large N limit. This calculation foreshadows the replica calculation of the spectral density and provides important insight.

For the problem of determining the mean spectral density, we take standard steps to map the problem onto one of the interacting systems of dynamical variables to which the replica method is applied. Following recent advances, the form of replica symmetric ansatz is identified as a superposition of Gaussians. Exploiting the insights gained in the earlier calculation, we obtain closed expressions for the spectral density in terms of the statistical properties of the graph ensemble (39). Similar equations are found for the community structure ensemble (47).

Though the resulting equations may not often have easily found analytic solutions, they can be efficiently solved numerically using the population dynamics algorithm described earlier. In this way, it is possible to analyse the spectral density of a given graph ensemble without the need to generate and diagonalize large numbers of graphs. An instance of this is provided by the discussion of the tails in the example with a power-law degree distribution; here the results of population dynamics could not feasibly be obtained by diagonalizing random matrices (for $k_{\max} = 400$, we would require graphs of around 1.6×10^5 vertices). We hope that the methods discussed here will prove useful tools in the study of complex networks.

Although as we have demonstrated the statistics of generalized degrees can have a significant impact on the spectral density of the graph ensemble, this is certainly not the only factor at work. One aspect of the topology of complex networks which may play an important role, but which we have not considered so far, is the statistics of the loops in the graph. Unfortunately, knowledge of the generalized degrees of a graph gives no information about loops, and hence the constrained generalized degree ensemble is not likely to be useful to study of their effect on spectral density. Looking to the future, it seems the next major step forward in the analysis of spectral density of random graphs will require techniques capable of handling the effects of loops. Several new techniques to correct for the presence of loops in related problems have a appeared recently [38–40], and we hope that similar ideas may also be applied to the study of spectral density.

Acknowledgments

KT acknowledges hospitality from the Disordered Systems Group, at the department of Mathematics, King's College London. KT is supported by a grant-in-aid Scientific Research on Priority Areas 'Deepening and Expansion of Statistical Mechanical Informatics (DEX-SMI)' from MEXT, Japan (no. 18079006). Conrad Pérez Vicente acknowledges funding from 'Ministerio de Educación y Ciencia' FIS2006-13321-C02-01.

References

[1] Guhr T, Müller-Groeling A and Weidenmüller H A 1998 *Phys. Rep.* **299** 190
[2] Wigner E P 1958 *Ann. Math.* **67** 325
[3] Dyson F J 1962 *J. Math. Phys.* **3** 140
[4] Mehta M L 1991 *Random Matrices* (New York: Academic)
[5] Rodgers G J and Bray A J 1988 *Phys. Rev.* B **37** 3557
[6] Bauer M and Golinelli O 2001 *J. Stat. Phys.* **103** 301
[7] Biroli G and Monasson R 1999 *J. Phys. A: Math. Gen.* **32** L255
[8] Dorogovtsev S N, Goltsev A V, Mendes J F F and Samukhin A N 2003 *Phys. Rev.* E **68** 046109
[9] Mirlin A D and Fyodorov Y V 1991 *J. Phys. A: Math. Gen.* **24** 2273
[10] Nagao T and Tanaka T 2007 *J. Phys. A: Math. Theor.* **40** 4973
[11] Nagao T and Rodgers G J 2008 *J. Phys. A: Math. Theor.* **41** 265002
[12] Semerjian G and Cugliandolo L F 2002 *J. Phys. A: Math. Gen.* **35** 4837
[13] Rogers T, Pérez Castillo I, Kühn R and Takeda K 2008 *Phys. Rev.* E **78** 031116
[14] Kühn R 2008 *J. Phys. A: Math. Theor.* **41** 295002
[15] Dean D S 2002 *J. Phys. A: Math. Gen.* **35** L153
[16] Ciliberti S, Grigera T S, Martín-Mayor V, Parisi G and Verrocchio P 2005 *Phys. Rev.* B **71** 153104
[17] Bordenave C and Lelarge M 2007 arXiv:0801.0155
[18] Fortunato S 2010 *Phys. Rep.* **486** 75–174
[19] Farkas I J, Derenyi I, Barabasi A L and Vicsek T 2001 *Phys. Rev.* E **64** 026704
[20] Skantzos N private communication
[21] Bianconi G, Coolen A C C and Pérez Vicente C J 2008 *Phys. Rev.* E **78** 016114
[22] Rogers T and Pérez Castillo I 2009 *Phys. Rev.* E **79** 012101
[23] Pérez Vicente C J and Coolen A C C 2008 *J. Phys. A: Math. Theor.* **41** 255003
[24] Oraby T 2007 *J. Theor. Probab.* **20** 1572
[25] Jalan S 2009 *Phys. Rev.* E **80** 04610
[26] Ergun G and Kühn R 2009 *J. Phys. A: Math. Theor.* **42** 395001
[27] Edwards S F and Jones R C 1976 *J. Phys. A: Math. Gen.* **9** 1595
[28] Dorogovtsev S N, Goltsev A V and Mendes J F F 2008 *Rev. Mod. Phys.* **80** 1275
[29] Berlin T H and Kac M 1952 *Phys. Rev.* **86** 821
[30] Mézard M and Parisi G 2001 *Eur. Phys. J.* B **20** 217
[31] Steger A and Wormald N C 1999 *Comb. Prob. Comput.* **8** 377
[32] Bayati M, Kim J H and Saberi A 2007 *Proc. 10th Int. Workshop on Approximation and the 11th Int. Workshop on Randomization, and Combinatorial Optimization. Algorithms and Techniques* (Berlin: Springer) p 326
[33] Blitzstein J and Diaconis P 2006 Unpublished manuscript
[34] Catanzaro M, Boguñá M and Pastor-Satorras R 2005 *Phys. Rev.* E **71** 027103
[35] Mihail M and Papadimitriou C 2002 On the Eigenvalue Power Law *Randomization and Approximation Techniques in Computer Science (Lecture Notes in Computer Science (LNCS))* vol 2483 ed J D P Rolim and S Vadhan (Berlin: Springer) pp 254–62
[36] Golinelli O 2003 arXiv:cond-mat/0301437
[37] Viswanath V S and Müller G 1994 *The Recursion Method: Application to Many-Body Dynamics* (Berlin: Springer)
[38] Montanari A and Rizzo T 2005 *J. Stat. Mech.: Theor. Exp.* P10011
[39] Parisi G and Slanina F 2006 *J. Stat. Mech.: Theor. Exp.* L02003
[40] Chertkov M and Chernyak V Y 2006 *J. Stat. Mech.: Theor. Exp.* P06009