



# CHORUS

This is the accepted manuscript made available via CHORUS. The article has been published as:

## Spectra of networks containing short loops

M. E. J. Newman

Phys. Rev. E **100**, 012314 — Published 26 July 2019

DOI: [10.1103/PhysRevE.100.012314](https://doi.org/10.1103/PhysRevE.100.012314)

# Spectra of networks containing short loops

M. E. J. Newman

*Department of Physics and Center for the Study of Complex Systems,  
University of Michigan, Ann Arbor, Michigan, USA*

The spectrum of the adjacency matrix plays several important roles in the mathematical theory of networks and network data analysis, for example in percolation theory, community detection, centrality measures, and the theory of dynamical systems on networks. A number of methods have been developed for the analytic computation of network spectra, but they typically assume that networks are locally tree-like, meaning that the local neighborhood of any node takes the form of a tree, free of short loops. Empirically observed networks, by contrast, often have many short loops. Here we develop an approach for calculating the spectra of networks with short loops using a message passing method. We give example applications to some previously studied classes of networks and find that the presence of loops induces substantial qualitative changes in the shape of network spectra, generating asymmetries, multiple spectral bands, and other features.

## I. INTRODUCTION

The adjacency matrix of an undirected network or graph with  $n$  nodes is the  $n \times n$  symmetric matrix  $\mathbf{A}$  having elements  $A_{uv} = 1$  if node  $u$  is connected to node  $v$  by an edge and  $A_{uv} = 0$  otherwise. The spectral properties of such matrices play a central role both in the mathematical theory of networks and in practical methods for the analysis of network data. The largest (most positive) eigenvalue of the matrix, for instance, is related to the percolation threshold on a network [1, 2], the corresponding eigenvector is widely used as a centrality measure [3], and spectral properties play a role in graph partitioning and community detection [4–6], in localization [7, 8], in the behavior of dynamical systems on networks [9], and in detectability transitions [10, 11].

Spectra can be calculated using well established, though computationally demanding, numerical methods, such as the QR algorithm, but there are also a number of analytic approaches, all closely related, that make use of message passing or cavity techniques [12–15]. In these approaches, the spectrum is expressed in terms of a set of complex-valued “messages” that are passed between adjacent nodes of the network, such that the values of the messages a node sends can be calculated, via fairly simple closed-form equations, from those it receives. In practical applications the equations are usually solved numerically, so that these methods are not fully analytic. Nonetheless, they can provide an estimate of the spectrum of the network that is accurate to any desired degree of precision, and they also form the foundation for a variety of additional analytic calculations, for instance of the spectra of random graphs [13–18].

One disadvantage of these methods, however, is that in most cases they are restricted to networks that are *locally tree-like*, meaning that the neighborhood of any node in the network takes the form of a tree, free of short loops, out to arbitrarily large distances in the limit of large network size. (By “short loops,” we mean loops whose size remains constant as the network becomes large—a triangle would be an example of a short loop in a net-

work.) Unfortunately, while model networks such as random graphs are typically tree-like in this sense, real-world networks usually are not. We would like to understand what effect the presence of short loops in networks might have on network spectra.

In this paper, we present a message passing method that allows us to compute adjacency matrix spectra for networks that contain short loops. Specifically, we show how to compute the spectra of networks that are made up of a collection of finite subgraphs or *motifs* which may contain loops, joined together via shared nodes. Such networks have been studied previously, for instance in the context of random graph models in which motifs are placed at random [19–22], and message passing methods (or their equivalent) have been applied to them to calculate things such as percolation properties [21, 22] or the behavior of the Ising model [23]. The spectrum of the special case in which the only loop is a triangle has been calculated in the limit of large average degree using random matrix methods [24], but as far as we are aware no calculation of the spectrum has been performed in the general case studied here. There are, of course, calculations going back many decades of the spectra of more specialized networks containing loops, such as regular lattices. Husimi graphs, a generalization of Bethe lattices built out of short loops of the same length, have also been studied using message passing methods [15, 25, 26]. Our work, however, focuses on more general classes of networks, closer to those observed in empirical studies.

## II. NETWORKS CONTAINING SHORT LOOPS

There are a number of ways of generating networks with short loops. One common approach involves “triadic closure,” meaning any of several processes in which loops are added to a network that initially has none by looking for pairs of nodes with a common neighbor and connecting them to form a triangle [27–30]. This approach has the advantage of mirroring directly a mechanism by which loops are believed to form in some real networks. On the other hand, it turns out to be quite dif-

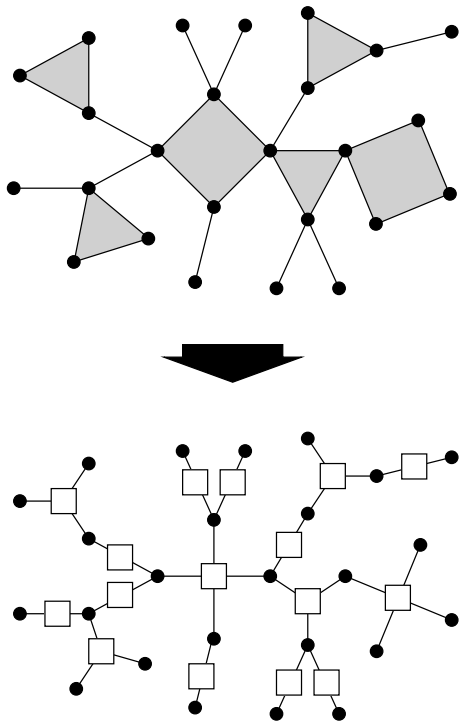


FIG. 1: Top: a network built from a collection of motifs. In this case there are three motifs: single edges, triangles, and loops of length four. The shading is intended only to highlight the motifs; the network itself consists of edges between nodes alone. Bottom: the same network represented as a factor graph, a bipartite network with two sets of nodes. One set, the filled circles, represents the nodes of the original network. The other, the open squares, represents the motifs.

difficult to treat analytically: even describing the ensemble of networks generated by such processes is non-trivial.

An alternative approach is to create networks by directly placing loops in them. For instance, one can create a random graph model in which not only single edges but also complete triangles are strewn among a set of nodes, each triangle connecting three randomly chosen nodes [19]. This approach creates ensembles of networks that are relatively straightforward to describe and analyze. One can, for instance, calculate the expected sizes of components, or percolation properties [19, 20]. One can take this approach further and create random graphs in which not merely triangles but also other, larger motifs are added to the network—sets of four, five, or more nodes, connected in any of a variety of different ways [21, 22].

Here we consider networks of this latter type, but we will not focus on any particular model for generating them. We will not, for instance, assume that they are generated randomly (although they could be). Instead, we demand only a weaker condition on our networks, as follows. We consider networks built of motifs, as we have described, where each motif is a connected subgraph joining together a specific set of nodes in some specific

pattern of edges. A single edge connecting two nodes is itself considered to be a motif, the simplest possible example. A triangle is another example, and there are many larger possibilities. Networks built in this fashion can be represented by a *factor graph*, a bipartite graph having two distinct sets of nodes. One set represents the nodes of the original network, the other represents the motifs, and there are edges connecting each node to the motifs to which it belongs—see Fig. 1. In our calculations we will assume that this factor graph is locally tree-like. The original network itself is not in general tree-like—it may contain many loops of various lengths that fall within the individual motifs—but the factor graph is tree-like. In effect, we are saying that all the short loops in the network are accounted for within the motifs. There are no additional short loops other than these. Our approach works by writing the spectrum of the network in terms of a message passing algorithm that acts not upon the original network but upon the factor graph, whose locally tree-like structure then makes the calculation exact, at least in the limit of large network size.

### III. MESSAGE PASSING

In this section we develop our message passing approach for the general case of networks formed of motifs of any size and structure, arranged on a locally tree-like factor graph as in Fig. 1. In Section IV we apply the resulting equations to the specific example of a network formed of just two motifs, single edges and triangles, for which the equations take a particularly simple form.

#### A. Spectral density

Our goal will be the calculation of the *spectral density* or *density of states*  $\rho(x)$  of a given network of  $n$  nodes, which is the function

$$\rho(x) = \frac{1}{n} \sum_{i=1}^n \delta(x - \lambda_i), \quad (1)$$

where  $\lambda_i$  is the  $i$ th eigenvalue of the adjacency matrix and  $\delta(x)$  is the Dirac delta function. Following a standard line of argument, we write the delta function as the limit of a suitably normalized Lorentzian (or Cauchy) distribution as its width  $\eta$  tends to zero:

$$\delta(x) = \lim_{\eta \rightarrow 0^+} \frac{\eta/\pi}{x^2 + \eta^2} = -\frac{1}{\pi} \lim_{\eta \rightarrow 0^+} \text{Im} \frac{1}{x + i\eta}, \quad (2)$$

where the notation  $\eta \rightarrow 0^+$  indicates that  $\eta$  tends to zero from above. Substituting this form into Eq. (1) we get

$$\rho(x) = -\frac{1}{n\pi} \lim_{\eta \rightarrow 0^+} \text{Im} \sum_{i=1}^n \frac{1}{x - \lambda_i + i\eta}. \quad (3)$$

It will be convenient to define a generalization of the spectral density to a complex argument  $z$  thus:

$$\rho(z) = -\frac{1}{n\pi} \sum_{i=1}^n \frac{1}{z - \lambda_i} = -\frac{1}{n\pi} \text{Tr}(z\mathbf{I} - \mathbf{A})^{-1}. \quad (4)$$

The standard real spectral density  $\rho(x)$  of Eq. (3) is then given by the imaginary part of this quantity where  $z = x + i\eta$  and we take the limit as  $\eta$  goes to zero from above. The quantity  $\eta$  acts as a broadening parameter that broadens the delta-function peaks in the spectrum by an amount roughly equal to  $\eta$ , and in practical calculations on finite networks it is sometimes convenient to retain a small non-zero value of  $\eta$  in order to make  $\rho(x)$  a smooth function of its argument.

Expanding the matrix inverse in Eq. (4) as a geometric series  $(z\mathbf{I} - \mathbf{A})^{-1} = z^{-1} \sum_{r=0}^{\infty} (\mathbf{A}/z)^r$  and taking the trace term by term, we have

$$\rho(z) = -\frac{1}{n\pi z} \sum_{r=0}^{\infty} \frac{\text{Tr} \mathbf{A}^r}{z^r}. \quad (5)$$

The quantity  $\text{Tr} \mathbf{A}^r$  is equal to the number of closed walks of length  $r$  on the network, a closed walk being any (possibly self-intersecting) path across the network that starts and ends at the same node. If we can count the number of such walks for all lengths  $r$  then we can calculate the spectral density from Eq. (5).

## B. Counting closed walks

Let  $u$  denote an arbitrary node in the network and  $\sigma$  denote one of the motifs to which it belongs. Because the factor graph is locally tree-like, any closed walk that starts at node  $u$  and takes its first step across one of the edges in motif  $\sigma$  must, on its final step, return to node  $u$  also across one of the edges in  $\sigma$  (although not necessarily the same edge). Were this not the case, were the walk to return via a different motif, then it would in the process complete a loop on the factor graph, of which, by hypothesis, there are none, and hence such walks cannot exist.

Let us define  $N_r^{u\sigma}$  to be the number of closed walks that start at node  $u$ , take their first step along one of the edges in motif  $\sigma$ , and return to node  $u$  for the first time, also via  $\sigma$ , exactly  $r$  steps later. Any node other than  $u$  may be visited as many times as we wish during the walk, but node  $u$  is visited only twice, once at the start of the walk and once at the end. We will call such walks *excursions*.

The total number of distinct possible excursions of length  $r$  from node  $u$  is given by the sum of  $N_r^{u\sigma}$  over the set  $\mathcal{S}_u$  of all motifs  $\sigma$  to which  $u$  belongs:  $\sum_{\sigma \in \mathcal{S}_u} N_r^{u\sigma}$ . Walks of length  $r$  that visit their starting node more than twice—say  $m$  times other than at the start of the walk—are made up of  $m$  distinct excursions with

lengths  $r_1 \dots r_m$  such that  $\sum_{i=1}^m r_i = r$ . Thus the number  $N_{rm}^u$  of such walks is

$$N_{rm}^u = \sum_{r_1=1}^{\infty} \dots \sum_{r_m=1}^{\infty} \delta(r, \sum_i r_i) \prod_{i=1}^m \sum_{\sigma \in \mathcal{S}_u} N_{r_i}^{u\sigma}, \quad (6)$$

where  $\delta(i, j)$  is the Kronecker delta. (Terms with  $r_i = 1$  can only appear in networks that have self-loops—edges that connect a node to itself—which is rare in real-world situations. In all other networks the shortest possible excursion has length 2. We leave these terms in the expression for the sake of completeness, however.)

Summing Eq. (6) over all possible values of  $m$  and all nodes  $u$  now gives us the total number of closed walks of length  $r$  in the whole network with any number of excursions, which is precisely equal to the quantity  $\text{Tr} \mathbf{A}^r$  that we are trying to calculate. Substituting into Eq. (5) we then get

$$\begin{aligned} \rho(z) &= -\frac{1}{n\pi z} \sum_{r=0}^{\infty} \frac{1}{z^r} \sum_{u=1}^n \sum_{m=0}^{\infty} N_{rm}^u \\ &= -\frac{1}{n\pi z} \sum_{r=0}^{\infty} \frac{1}{z^r} \sum_{u=1}^n \sum_{m=0}^{\infty} \sum_{r_1=1}^{\infty} \dots \sum_{r_m=1}^{\infty} \delta(r, \sum_i r_i) \\ &\quad \prod_{i=1}^m \sum_{\sigma \in \mathcal{S}_u} N_{r_i}^{u\sigma} \\ &= -\frac{1}{n\pi z} \sum_{u=1}^n \sum_{m=0}^{\infty} \prod_{i=1}^m \sum_{r_i=1}^{\infty} \sum_{\sigma \in \mathcal{S}_u} \frac{N_{r_i}^{u\sigma}}{z^{r_i}}, \end{aligned} \quad (7)$$

where we adopt the convention that the empty product  $\prod_{i=1}^0$  is equal to 1. Defining the useful quantity

$$\mu_{u\sigma}(z) = \sum_{r=1}^{\infty} \frac{N_r^{u\sigma}}{z^{r-1}}, \quad (8)$$

we can write Eq. (7) as

$$\begin{aligned} \rho(z) &= -\frac{1}{n\pi z} \sum_{u=1}^n \sum_{m=0}^{\infty} \left[ \sum_{\sigma \in \mathcal{S}_u} \frac{\mu_{u\sigma}(z)}{z} \right]^m \\ &= -\frac{1}{n\pi} \sum_{u=1}^n \frac{1}{z - \sum_{\sigma \in \mathcal{S}_u} \mu_{u\sigma}(z)}. \end{aligned} \quad (9)$$

We regard  $\mu_{u\sigma}(z)$  as a “message,” sent by motif  $\sigma$  to node  $u$ , whose value can be calculated as we now demonstrate.

## C. Message passing equations

To evaluate the messages (8) we need to compute the number of excursions  $N_r^{u\sigma}$ —the number of walks from  $u$  that take their first (and last) step via motif  $\sigma$  and visit  $u$  only at the start and end of the walk. The structure of such walks is illustrated in Fig. 2. Each consists of

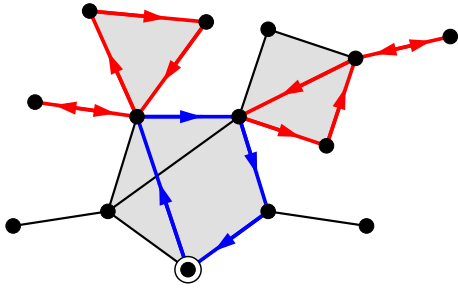


FIG. 2: A complete excursion from a starting node (bottom, circled) consists of a closed walk within a single motif, here having four steps, plus, optionally, any number of sub-excursions along the way that leave the motif and return to it some time later via the same node. Each sub-excursion is itself of the same form, which allows us to write a self-consistent expression (10) for the total number of excursions.

a closed walk around motif  $\sigma$  itself, which visits  $u$  only at its start and end, plus any number of excursions from nodes in  $\sigma$  (other than  $u$ ) to the rest of the network.

Let  $w$  be a walk of length  $k + 1$  around motif  $\sigma$ , which we will represent by the list  $v_1 \dots v_k$  of  $k$  nodes (not necessarily distinct) that the walk passes through other than node  $u$ , and let  $r_1^v, r_2^v, \dots$  be the lengths of the excursions from node  $v$ . Then a complete walk of length  $r$ , excursions included, must have  $r = k + 1 + \sum_{v \in w} \sum_i r_i^v$ , and the total number of complete walks of length  $r$  that have  $w$  as their foundation is given by

$$\left[ \sum_{m_{v_1}=0}^{\infty} \sum_{r_1^{v_1}=1}^{\infty} \dots \sum_{r_{m_{v_1}}^{v_1}=1}^{\infty} \right] \dots \left[ \sum_{m_{v_k}=0}^{\infty} \sum_{r_1^{v_k}=1}^{\infty} \dots \sum_{r_{m_{v_k}}^{v_k}=1}^{\infty} \right] \delta(r, k + 1 + \sum_{v \in w} \sum_{i=1}^{m_v} r_i^v) \prod_{v \in w} \prod_{i=1}^{m_v} \sum_{\substack{\tau \in \mathcal{S}_v \\ \tau \neq \sigma}} N_{r_i^v}^{v\tau}, \quad (10)$$

where  $m_{v_1} \dots m_{v_k}$  represent the numbers of excursions from each of the nodes  $v_1 \dots v_k$ .

The total number  $N_r^{u\sigma}$  of paths is now equal to the sum of Eq. (10) first over the complete set  $\mathcal{W}_{u\sigma}^k$  of walks of length  $k + 1$  in motif  $\sigma$  that start and end at node  $u$ , and second over all  $k = 0 \dots \infty$ . Taking the resulting expression and substituting it into Eq. (8), we find that

$$\begin{aligned} \mu_{u\sigma}(z) &= \sum_{k=0}^{\infty} \frac{1}{z^k} \sum_{w \in \mathcal{W}_{u\sigma}^k} \prod_{v \in w} \sum_{m=0}^{\infty} \prod_{i=1}^m \sum_{\substack{r_i^v=1 \\ \tau \in \mathcal{S}_v \\ \tau \neq \sigma}} \frac{N_{r_i^v}^{v\tau}}{z^{r_i^v}} \\ &= \sum_{k=0}^{\infty} \frac{1}{z^k} \sum_{w \in \mathcal{W}_{u\sigma}^k} \prod_{v \in w} \sum_{m=0}^{\infty} \left[ \sum_{\substack{\tau \in \mathcal{S}_v \\ \tau \neq \sigma}} \sum_{r=1}^{\infty} \frac{N_r^{v\tau}}{z^r} \right]^m \\ &= \sum_{k=0}^{\infty} \sum_{w \in \mathcal{W}_{u\sigma}^k} \prod_{v \in w} \frac{1}{z - \sum_{\substack{\tau \in \mathcal{S}_v \\ \tau \neq \sigma}} \mu_{v\tau}(z)}. \end{aligned} \quad (11)$$

Defining a new message  $g_{\sigma u}(z)$ , passed from node  $u$  to motif  $\sigma$ , by

$$g_{\sigma u}(z) = \frac{1}{z - \sum_{\substack{\tau \in \mathcal{S}_u \\ \tau \neq \sigma}} \mu_{u\tau}(z)}, \quad (12)$$

we can write Eq. (11) as

$$\mu_{u\sigma}(z) = \sum_{k=0}^{\infty} \sum_{w \in \mathcal{W}_{u\sigma}^k} \prod_{v \in w} g_{\sigma v}(z). \quad (13)$$

Equations (12) and (13) give us a complete set of self-consistent equations whose solutions give the values of the messages  $\mu_{u\sigma}(z)$ . For any given network we can solve these equations, for example by simple iteration, to find the values of the  $\mu_{u\sigma}(z)$  for any  $z$  and then substitute the results into Eq. (9) to get the spectral density.

#### IV. EXAMPLES

In practice the message passing equations (12) and (13) can be difficult to solve because they require us to enumerate all closed walks of a given length for every motif in the network, which in many cases is not an easy task. One important case that is relatively straightforward, however, is the case of a network composed solely of single edges and triangles, which has been studied in other contexts in the past [19, 20, 22].

For networks of this kind, which we will call *edge-triangle networks*, there are only two distinct motifs: single edges, which connect two nodes to one another, and triangles, which connect three.

Let us treat these in turn. The case of single edges is straightforward. There is only one closed walk in such a motif, having two steps, along the edge and back again. In this case Eq. (13) simplifies to

$$\mu_{u\sigma}(z) = g_{\sigma v}(z), \quad (14)$$

where  $v$  is the node at the other end of edge  $\sigma$  from  $u$ .

The case of triangles is only a little more complicated. Every closed walk from  $u$  in a triangle has the same form: we walk from  $u$  to one of the two other nodes in the triangle—let us call them  $v$  and  $w$ —then we alternate back and forth between  $v$  and  $w$  some number of times before returning to  $u$  on the final step. For even values of  $k$  in Eq. (13) nodes  $v$  and  $w$  are visited the same number of times, namely  $\frac{1}{2}k$  each. For odd values, one node is visited  $\frac{1}{2}(k + 1)$  times and the other  $\frac{1}{2}(k - 1)$  times. Setting  $k = 2l$  and summing over integer values of  $l$ , Eq. (13) can then be written

$$\begin{aligned} \mu_{u\sigma}(z) &= \sum_{l=1}^{\infty} [2g_{\sigma v}^l g_{\sigma w}^l + g_{\sigma v}^l g_{\sigma w}^{l-1} + g_{\sigma v}^{l-1} g_{\sigma w}^l] \\ &= \frac{2g_{\sigma v}(z)g_{\sigma w}(z) + g_{\sigma v}(z) + g_{\sigma w}(z)}{1 - g_{\sigma v}(z)g_{\sigma w}(z)}. \end{aligned} \quad (15)$$

Combining Eq. (14) (for single edges) and (15) (for triangles) with Eq. (12), we now have our complete message passing equations for edge–triangle networks.

### A. Regular graphs

As an example, consider a random network in which every node belongs to exactly one single edge and one triangle. This is the equivalent for an edge–triangle network of the regular graphs of traditional graph theory—networks in which every node has the same degree. The spectrum of a random regular graph obeys the well-known Kesten–McKay distribution [31]. Here we calculate the equivalent result for the edge–triangle case.

A random edge–triangle network necessarily has a locally tree-like factor graph for the same reason that traditional random graphs are locally tree-like: in a random graph the probability of creating a loop vanishes in the limit of large network size. This means our message passing equations are applicable to the random case and will be exact in the limit of large network size.

Moreover, in a *regular* edge–triangle network every node is equivalent, having the same local neighborhood, out to arbitrary distances in the limit of large network size. Similarly, every single edge has the same neighborhood as every other, as does every triangle. This means in practice that  $\mu_{u\sigma}(z)$  takes only two values, one for single edges and one for triangles, which we will denote by  $\mu(z)$  and  $\nu(z)$  respectively. Similarly,  $g_{\sigma u}(z)$  takes only two values, which we will denote  $g(z)$  and  $h(z)$ . In terms of these quantities, the message passing equations read

$$\mu(z) = g(z), \quad \nu(z) = \frac{2h(z)}{1 - h(z)}, \quad (16)$$

and

$$g(z) = \frac{1}{z - \nu(z)}, \quad h(z) = \frac{1}{z - \mu(z)}. \quad (17)$$

Eliminating  $g$  and substituting into Eq. (9), we find that the complex spectral density  $\rho(z)$  is given by

$$\rho(z) = -\frac{1/\pi}{z - \mu(z) - \nu(z)} = \frac{1/\pi}{\mu - 1/\mu}. \quad (18)$$

At the same time, eliminating  $g$ ,  $h$ , and  $\nu$  from Eqs. (16) and (17) yields a quadratic equation for  $\mu$  with solution

$$\mu(z) = \frac{z^2 - z - 1 \pm \sqrt{z^4 - 2z^3 - 5z^2 + 6z + 1}}{2z}. \quad (19)$$

We substitute this result into (18) and, after some manipulation, find that

$$\rho(z) = \frac{z^2 - z - 1 \pm (2z - 1)\sqrt{z^4 - 2z^3 - 5z^2 + 6z + 1}}{2\pi(z^4 - 2z^3 - 5z^2 + 6z)}. \quad (20)$$

Letting  $z$  go to the real line at  $x$ , taking the imaginary part, and noting that the discriminant within the square root can be written in the form  $[(x - \frac{1}{2})^2 - \frac{13}{4}]^2 - 8$ , we then recover the real spectral density

$$\rho(x) = \frac{1}{\pi} \left| x - \frac{1}{2} \right| \frac{\sqrt{8 - [(x - \frac{1}{2})^2 - \frac{13}{4}]^2}}{9 - [(x - \frac{1}{2})^2 - \frac{13}{4}]^2}. \quad (21)$$

This is the analog of the Kesten–McKay distribution for this random regular edge–triangle network.

In order for (21) to be real, we require that the quantity within the square root be positive. The zeros of this quantity satisfy  $(x - \frac{1}{2})^2 - \frac{13}{4} = \pm\sqrt{8}$ , or

$$x = \frac{1}{2} \left( 1 \pm \sqrt{13 \pm 8\sqrt{2}} \right), \quad (22)$$

and hence the spectral density is non-vanishing in two different bands, the first between the  $x$  values  $\frac{1}{2}(1 - \sqrt{13 \pm 8\sqrt{2}})$ , which is approximately  $x \in [-1.965, -0.149]$ , and the second between the values  $\frac{1}{2}(1 + \sqrt{13 \pm 8\sqrt{2}})$ , which is approximately  $x \in [1.149, 2.965]$ . Moreover, since Eq. (21) is a function of  $|x - \frac{1}{2}|$  only, the spectrum must be symmetric about the point  $x = \frac{1}{2}$ , meaning that the two bands are mirror images of one another. Figure 3 shows the shape of the spectrum computed from Eq. (21).

Also visible in the figure are two delta-function peaks in the spectrum at  $x = -2$  and  $x = 0$ . These appear in the solution as divergences in  $\rho(z)$  at the corresponding points, which map to delta functions via Eq. (2). We could, if we wish, introduce a small Lorentzian broadening via the parameter  $\eta$  in (2) to make these peaks visible in the figure, but we have not done that here. Instead we have merely added the peaks to the figure by hand at their calculated positions.

We also show in Fig. 3 a histogram of the spectrum, calculated by direct numerical diagonalization of the adjacency matrix of a single example network randomly generated from the model with  $n = 10\,002$  nodes. (The number of nodes must be a multiple of six to satisfy the requirement that the number of ends of edges be even and the number of corners of triangles be a multiple of three.) As the figure shows, the exact solution and numerical results agree well.

In principle, one could construct a closed-form solution analogous to (21) for the spectrum of any random regular edge–triangle network such that every node has the same number of edges and triangles. A few other cases are relatively simple to solve, such as the case where there are triangles only and no single edges, which is essentially equivalent to the Husimi graphs studied in Refs. [15, 26]. The general case, however, involves solving a quartic rather than a quadratic equation, and, while this can be done, the solution is complicated and we do not give it here.

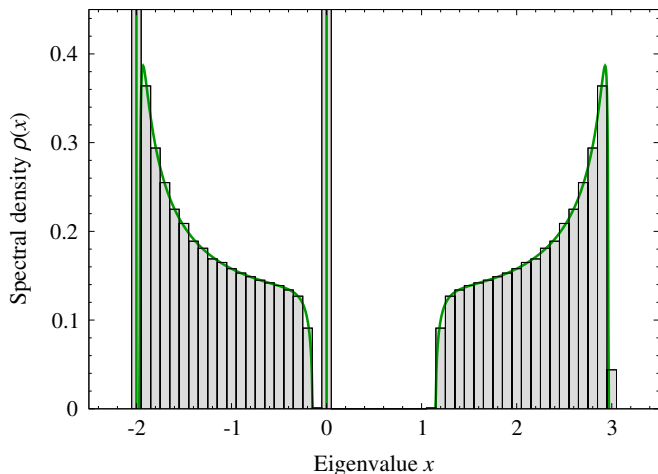


FIG. 3: The spectrum of a random regular network in which every node belongs to exactly one single edge and one triangle. The histogram shows the distribution of eigenvalues calculated by numerical diagonalization of the adjacency matrix of one random realization of the network with 10 002 nodes. The solid curves show the analytic solution, with the two delta-function peaks added by hand at their calculated positions.

### B. Random graphs

A more complex example is a general random edge-triangle network in which we specify separately the number of edges and triangles that each node participates in. In recent work, Peron *et al.* [24] have calculated spectra for such networks using a random matrix method, though their result is valid only in the limit of large average network degree. Our message passing method, by contrast, allows us to study all cases, including cases with arbitrarily small average degree.

To compute the spectral density of these networks we must solve the full message passing equations (12), (14), and (15), which we do here by simple iteration. For a given network we choose any convenient starting values for the messages (setting them all to zero works well), then numerically iterate the equations until they converge to the desired degree of accuracy. This process does not give us an analytic form for the spectrum, but it can in principle give us a solution accurate to any required precision.

Figure 4a shows the results of one such calculation. In this example we have generated a network of  $n = 10\,000$  nodes, each having a number of single edges drawn from a Poisson distribution with mean 2, and a number of triangles drawn from a Poisson distribution also with mean 2. The result is a random network with average degree 6 and  $\frac{2}{3}n$  triangles in total, on average. The histogram in Fig. 4a shows the spectrum of the network calculated by conventional numerical diagonalization of the adjacency matrix. The continuous curve shows the spectral density

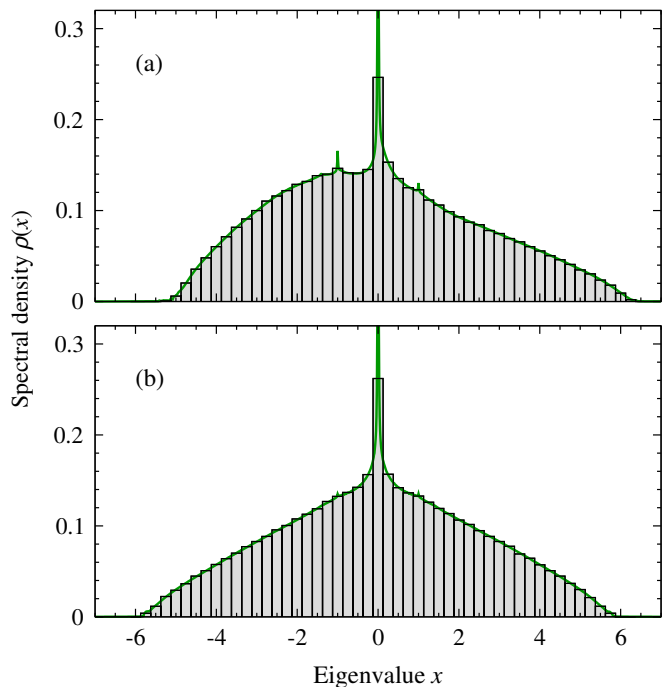


FIG. 4: (a) The spectrum of a random network composed of single edges and triangles, with a Poisson distribution for both: the number of both edges and triangles is Poisson distributed at every node with mean 2. (b) The spectrum of a configuration model with the same node degrees as the network in (a). In each panel the solid curve shows the spectral density for a single randomly generated network with 10 000 nodes, calculated by direct iteration of the message passing equations with Lorentzian broadening parameter  $\eta = 0.01$ . The histogram shows the spectrum of the same network calculated by numerical diagonalization of the adjacency matrix.

calculated using the message passing equations, with a Lorentzian broadening parameter  $\eta = 0.01$  (see Eq. (2) and the following discussion). As the figure shows, the agreement between the two calculations is good. Note in particular how the delta-function peaks in the spectrum at  $x = 0$  and  $\pm 1$  emerge clearly in the message passing calculation, made visible by the broadening. (These peaks arise because of the presence of small components in the network. Components of size one, for instance, contribute a single zero eigenvalue each.) Note also the very different overall shape of the spectrum in this case from that of Fig. 3 for the random regular edge-triangle network.

How does the presence of loops in the network affect the spectrum in practice? To shed light on this question, we show in Fig. 4b the spectrum of a random graph with the same node degrees as in Fig. 4a, but without loops. This network is generated using the standard configuration model in which single edges only are placed between nodes in the appropriate numbers, but no triangles [32, 33]. In the limit of large network size this produces a locally tree-like network. Figures 4a and 4b

are plotted on the same scales and, as we can see, there are some general similarities between the two but also some clear differences. The approximate ranges spanned by the spectra are similar and both show a prominent peak at  $x = 0$ . But without loops the spectrum is symmetric, whereas for the network with loops it has a distinctly asymmetric shape—the presence of loops produces a clear redistribution of values, for instance shifting the upper edge of the band to a higher value. A similar but more pronounced asymmetry is visible in the spectrum of the regular edge–triangle network in Fig. 3.

Mathematically the asymmetry of the spectrum has a straightforward explanation. In a loopless network all closed walks have an even number of steps, and hence the series expansion in Eq. (5) contains only even powers of  $z$ . Once triangles are introduced into the network we can have walks of odd length and hence the function can contain both even and odd powers of  $z$ .

The difference between the spectra for networks with and without loops indicates that calculations in which network spectral properties are approximated using model networks like the configuration model will in general produce not only quantitative but also qualitative errors because of their neglect of loops.

## V. CONCLUSIONS

In this paper we have derived and demonstrated a message passing method for calculating the adjacency matrix spectra of networks that contain short loops. Most previous analytic approaches to calculating network spectra have made the assumption that the networks studied were locally tree-like, an assumption that is strongly violated by most real-world networks. (There have been some prior studies of non-tree-like cases, but they have typically been limited to specialized classes of networks such as regular lattices, dense graphs, or Husimi graphs.) Our approach gets around this assumption by representing networks as a collection of motifs, which are allowed to contain loops, but assuming that the factor graph describing the connections between motifs is tree-like. This

assumption is enough to allow us to write down message passing equations for the spectrum of such a network, even though the network itself contains loops.

We have given two applications of our method to example networks composed of just the two simplest motifs, single edges and triangles. In our first example we consider a random regular graph model in which each node belongs to exactly one edge and one triangle. For this model we are able to solve the message passing equations exactly and derive a closed-form expression for the spectral density. In our second example, we generate random networks with Poisson distributions of their numbers of edges and triangles, and compute their spectra by numerical solution of the message passing equations. In both cases we find good agreement with direct calculations of the eigenvalues by numerical diagonalization.

The calculations described here could be extended in a number of ways. We have considered only one example of the random regular edge–triangle network but, as mentioned in Section IV A, it would be possible in principle to calculate a closed-form expression for the spectra of such networks in the general case where nodes belong to any number of edges and triangles, although such an expression, whose derivation involves the solution of a quartic equation, seems likely to be complicated. One could also consider more elaborate networks built of larger motifs than the simple edges and triangles we have considered here. To do this, one would have to enumerate all closed walks within each allowed motif in order to evaluate Eq. (13), which becomes progressively harder as the motifs become larger. Some cases are relatively tractable, however. Walks on motifs consisting of simple loops of any length, for instance, are quite straightforward to enumerate. These issues, however, we leave for future work.

## Acknowledgments

This work was funded in part by the US National Science Foundation under grant DMS–1710848.

- 
- [1] B. Bollobás, C. Borgs, J. Chayes, and O. Riordan, Percolation on dense graph sequences. *Annals of Probability* **38**, 150–183 (2010).
  - [2] B. Karrer, M. E. J. Newman, and L. Zdeborová, Percolation on sparse networks. *Phys. Rev. Lett.* **113**, 208702 (2014).
  - [3] P. F. Bonacich, Power and centrality: A family of measures. *Am. J. Sociol.* **92**, 1170–1182 (1987).
  - [4] M. Fiedler, Algebraic connectivity of graphs. *Czech. Math. J.* **23**, 298–305 (1973).
  - [5] A. Pothen, H. Simon, and K.-P. Liou, Partitioning sparse matrices with eigenvectors of graphs. *SIAM J. Matrix Anal. Appl.* **11**, 430–452 (1990).
  - [6] M. E. J. Newman, Modularity and community structure in networks. *Proc. Natl. Acad. Sci. USA* **103**, 8577–8582 (2006).
  - [7] M. Cucuringu and M. W. Mahoney, Localization on low-order eigenvectors of data matrices. Preprint arxiv:1109.1355 (2011).
  - [8] T. Martin, X. Zhang, and M. E. J. Newman, Localization and centrality in networks. *Phys. Rev. E* **90**, 052808 (2014).
  - [9] M. A. Porter and J. Gleeson, *Dynamical Systems on Networks: A Tutorial*. Springer, Berlin (2016).
  - [10] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová, Inference and phase transitions in the detection of modules



- in sparse networks. *Phys. Rev. Lett.* **107**, 065701 (2011).
- [11] R. R. Nadakuditi and M. E. J. Newman, Graph spectra and the detectability of community structure in networks. *Phys. Rev. Lett.* **108**, 188701 (2012).
- [12] S. N. Dorogovtsev, A. V. Goltsev, J. F. F. Mendes, and A. N. Samukhin, Spectra of complex networks. *Phys. Rev. E* **68**, 046109 (2003).
- [13] T. Rogers, I. Pérez Castillo, R. Kühn, and K. Takeda, Cavity approach to the spectral density of sparse symmetric random matrices. *Phys. Rev. E* **78**, 031116 (2008).
- [14] T. Rogers, C. Pérez Vicente, K. Takeda, and I. Pérez Castillo, Spectral density of random graphs with topological constraints. *J. Phys. A* **43**, 195002 (2010).
- [15] F. L. Metz, I. Neri, and D. Bollé, Spectra of sparse regular graphs with loops. *Phys. Rev. E* **84**, 055101 (2011).
- [16] G. Semerjian and L. F. Cugliandolo, Sparse random matrices: The eigenvalue spectrum revisited. *J. Phys. A* **35**, 4837–4852 (2002).
- [17] R. Kühn, Spectra of sparse random matrices. *J. Phys. A* **41**, 295002 (2008).
- [18] M. E. J. Newman, X. Zhang, and R. R. Nadakuditi, Spectra of random networks with arbitrary degrees. Preprint arxiv:1901.02029 (2019).
- [19] M. E. J. Newman, Random graphs with clustering. *Phys. Rev. Lett.* **103**, 058701 (2009).
- [20] J. C. Miller, Percolation and epidemics in random clustered networks. *Phys. Rev. E* **80**, 020901 (2009).
- [21] J. P. Gleeson, Bond percolation on a class of clustered random networks. *Phys. Rev. E* **80**, 036107 (2009).
- [22] B. Karrer and M. E. J. Newman, Random graphs containing arbitrary distributions of subgraphs. *Phys. Rev. E* **82**, 066118 (2010).
- [23] S. Yoon, A. V. Goltsev, S. N. Dorogovtsev, and J. F. F. Mendes, Belief-propagation algorithm and the Ising model on networks with arbitrary distributions of motifs. *Phys. Rev. E* **84**, 041144 (2011).
- [24] T. K. D. Peron, P. Ji, J. Kurths, and F. A. Rodrigues, Spectra of random networks in the weak clustering regime. *EPL* **121**, 68001 (2018).
- [25] M. Eckstein, M. Kollar, K. Byczuk, and D. Vollhardt, Hopping on the Bethe lattice: Exact results for densities of states and dynamical mean-field theory. *Phys. Rev. B* **71**, 235119 (2005).
- [26] D. Bollé, F. L. Metz, and I. Neri, On the spectra of large sparse graphs with cycles. In *Spectral Analysis, Differential Equations, and Mathematical Physics*, pp. 35–58, American Mathematical Society, Providence, RI (2013).
- [27] E. M. Jin, M. Girvan, and M. E. J. Newman, The structure of growing social networks. *Phys. Rev. E* **64**, 046132 (2001).
- [28] P. Holme and B. J. Kim, Growing scale-free networks with tunable clustering. *Phys. Rev. E* **65**, 026107 (2002).
- [29] K. Klemm and V. M. Eguiluz, Highly clustered scale-free networks. *Phys. Rev. E* **65**, 036123 (2002).
- [30] M. A. Serrano and M. Boguñá, Tuning clustering in random networks with arbitrary degree distributions. *Phys. Rev. E* **72**, 036133 (2005).
- [31] B. D. McKay, The expected eigenvalue distribution of a large regular graph. *Linear Algebra Appl.* **40**, 203–216 (1981).
- [32] M. Molloy and B. Reed, A critical point for random graphs with a given degree sequence. *Random Structures and Algorithms* **6**, 161–179 (1995).
- [33] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Random graphs with arbitrary degree distributions and their applications. *Phys. Rev. E* **64**, 026118 (2001).