Journal of Statistical Physics (2020) 181:305–328
https://doi.org/10.1007/s10955-020-02582-4

Phase Transitions in Finite Random Networks

Joe Neeman[1] · **Charles Radin**[1] · **Lorenzo Sadun**[1]

Received: 2 March 2020 / Accepted: 29 May 2020 / Published online: 19 June 2020

Abstract
We analyze ensembles of random networks with fixed numbers of edges, triangles, and nodes. In the limit as the number of nodes goes to infinity, this model is known to exhibit sharp phase transitions as the density of edges and triangles is varied. In this paper we study finite size effects in ensembles where the number of nodes is between 30 and 66. Despite the small number of nodes, the phases and phase transitions are clearly visible. Concentrating on 54 nodes and one particular phase transition, we use spectral data to examine the structure of a typical graph in each phase, and see that is it very similar to the graphon that describes the system as n diverges. We further investigate how graphs change their structure from that of one phase to the other under a natural edge flip dynamics in which the number of triangles is either forced downwards or allowed to drift upwards. In each direction, spectral data shows that the change occurs in three stages: a very fast stage of about 100 edge flips, in which the triangle count reaches the targeted value, a slower second stage of several thousand edge flips in which the graph adopts the qualitative features of the new phase, and a very slow third stage, requiring hundreds of thousands of edge flips, in which the fine details of the structure equilibrate.

Keywords Random graph · Phase transition · Finite size effect · Markov Chain Monte Carlo · Graphon

1 Introduction

We analyze the 'edge/triangle' random network model popularized by Strauss in 1986 [21], focusing on the emergence of phases and phase transitions as system size increases. Such transitions have been found in a variety of parametric random network models (see [8,14]

Communicated by Federico Ricci-Tersenghi.

✉ Joe Neeman
jneeman@math.utexas.edu

Charles Radin
radin@math.utexas.edu

Lorenzo Sadun
sadun@math.utexas.edu

[1] Department of Mathematics, The University of Texas at Austin, Austin, TX 78712, USA

and references therein). Specifically, consider a graph with n vertices, where n is large. Fix the fraction e of the $\binom{n}{2}$ possible edges that are actually present, and fix the fraction t of the $\binom{n}{3}$ triples of vertices that form triangles. How many such graphs exist? What is the structure of a typical graph? Especially, how does this structure change as the parameters e and t are varied?

In the $n \to \infty$ limit, large deviations theory reformulates these questions into questions about optimal *graphons*. Fixing the values of e and t is equivalent to applying integral constraints to the graphon, and finding the structure of a typical graph is equivalent to maximizing an entropy functional subject to those constraints. The entropy and the optimizing graphons are piecewise analytic functions of (e, t). We call each region of analyticity in the (e, t) plane a *phase*, and the boundaries between them *phase transitions*. At some phase transitions there is a unique optimizing graphon, and we say that the transition is *continuous*. At others there are two optimizers, one for each phase, and the graphon jumps discontinuously from one to the other as the constraints are varied.

Unfortunately, analyzing graphons can only tell us what happens in the $n \to \infty$ limit. In this paper we analyze what happens for finite values of n, specifically for n between 30 and 66.

We study a particular discontinuous phase transition, between the so-called $B(1, 1)$ and $A(3, 0)$ phases (defined precisely below). The optimum graphons for the two phases are different, with the $B(1, 1)$ graphon having only one negative eigenvalue and the $A(3, 0)$ graphon having two negative eigenvalues. The corresponding finite graphs then have adjacency matrices with one or two large ($O(n)$) negative eigenvalues. By studying the distribution of the second-largest negative eigenvalue of the adjacency matrix in the ensemble of graphs with fixed (n, e, t), we can identify the phase structure for different values of n. The structure gets sharper as n increases, but is already clear for n as low as 30; this allows for computer simulation. The fact that ensembles of random graphs with such small values of n already exhibit the qualitative structures described by the optimizing graphons is the main result of the first half of this paper.

The second half of the paper concerns transitions from one phase to the other. Restricting our attention to edge density $e = 0.67$ and $n = 54$, we set up an edge flip dynamics that samples the ensemble for fixed triangle density t. We then consider what happens when we introduce a bias that either allows t to increase from values associated with the $A(3, 0)$ phase to those associated with $B(1, 1)$, or that forces t to decrease from $B(1, 1)$ to $A(3, 0)$. Along the way, we track several related quantities:

(1) The triangle density t.
(2) The second most negative eigenvalue λ_2.
(3) The density of another subgraph, namely a tetrahedron with one edge removed, which we call "two-ears".
(4) A plot of the eigenvectors of the adjacency matrix corresponding to the two most negative eigenvalues. This is a scatter plot in $\mathbb{R}^2$, with each point being one of the n vertices. In the $A(3, 0)$ phase this scatter plot consists of three clusters, each of size around $n/3$, forming an equilateral triangle. In the $B(1, 1)$ phase it consists of a single cluster consisting of about $0.4n$ vertices and a vertical bar containing the other $0.6n$. By watching this plot evolve as the graph undergoes its edge flip dynamics, we can literally see a graph changing from one structure to the other.
(5) Graphons, of the sort associated with the $A(3, 0)$ and $B(1, 1)$ phases, that come closest to fitting the actual adjacency matrix.

In each direction, the transition involves three distinct stages. In Stage 1, requiring on the order of 100 edge flips, t reaches the desired value. However, the underlying structure of the graphs continues to resemble the starting phase more than the target phase. In Stage 2, requiring a few thousand edge flips, the graphs acquire the qualitative features of the new phase. This process typically involves a latency period of random length, followed by a period of more rapid change; the duration of the latency period is interesting in its own right. The final Stage 3 requires hundreds of thousands of edge flips to equilibrate the number of vertices in each cluster of the scatter plot.

Our analysis is one realization of the attempt to understand how a system consisting of many similar components in one homogeneous configuration manages to rearrange its components as a parameter is varied a small amount, to become a qualitatively different homogeneous configuration. The classic (harder!) case is where a 'liquid' configuration, say of H_2O molecules, rearranges into a (crystalline) 'solid' configuration as the temperature crosses a precise freezing point.

2 Background

We begin with some notation. Given n abstract labelled nodes, we denote by $E(G)$ and $T(G)$ the number of edges and triangles in the simple network G. (Simple means that two nodes can have at most one edge connecting them and there are no loops connecting a node with itself.) Strauss put on the set of all such networks the 2-parameter family of unnormalized probability mass functions:

$$\exp[s_1 e(G) + s_2 t(G)], \tag{1}$$

where

$$e(G) = \frac{E(G)}{\binom{n}{2}} \text{ and } t(G) = \frac{T(G)}{\binom{n}{3}} \tag{2}$$

are the edge and triangle densities in G, with value 1 in the complete network. The model parameters s_1, s_2 are analogues of $\beta\mu$ and β in a grand canonical distribution of statistical mechanics, where β is inverse temperature and μ is chemical potential.

A network G is described by its adjacency matrix $M(G)$. This is a symmetric $n \times n$ matrix whose (i, j) entry is 1 if nodes i and j are connected by an edge, and zero otherwise. The statistical properties of the matrix are independent of the (arbitrary) ordering of the nodes. Permutation-invariant data such as the eigenvalues of $M(G)$ are particularly informative.

As in statistical mechanics, phases in parametric models of random networks require an infinite-system-size limit formalism. Instead of looking at a symmetric 0–1 matrix $M(G)$, we consider a symmetric function g on the unit square $[0, 1]^2$ taking values in the interval $[0, 1]$. (The points in $[0, 1]$ are thought of as representing infinitely many nodes.) Such a *graphon* g is the kernel of an integral operator $M(g)$ acting on $L^2([0, 1])$. In an important series of papers [2,3,11–13], Borgs et al. used a certain metric on the space of networks to construct a completion, called the space of (reduced) graphons, with many useful properties. Used together with an important large deviation result [5] of Chatterjee/Varadhan, we may study convergence of distributions to infinite size through optimal-free-energy graphons g_{s_1,s_2}. See for instance [4].

These network models are mean-field in the sense that there is no space cutoff limiting the direct influence of variables on one another. For traditional statistical mechanics models with short range forces there is an equivalence of ensembles, connected by Legendre transforms.

However, as is common in mean-field models, Legendre transforms are not invertible for these graphon models and lose information contained in the microcanonical ensemble. Therefore, we will use the microcanonical version of the model [6,7]. That is, we specify the parameters $e(G)$ and $t(G)$ and consider the uniform distribution on networks with the fixed values of e and t.

For fixed values of (e, t), the graphon formalism allows us to take a limit in the number of nodes, $n \to \infty$, with 'most' constrained networks G converging to an entropy-optimal graphon $g_{(e,t)}$, in the sense that their (suitably normalized) adjacency matrices $M(G)$, viewed as operators on $\mathbb{R}^n$, converge in a weak sense to the compact linear integral operator $M(g_{(e,t)})$ on $L^2([0, 1]^2)$, with the (suitably normalized) ordered sets of eigenvalues of the $M(G)$'s converging to those of $M(g_{(e,t)})$; see section 11.6 in [10]. In [6–9,15–18] this and similar models have been studied in depth; we describe some results next.

The optimal-entropy graphons have been shown to have a very simple form. If we break the interval [0, 1] into finitely many sub-intervals, called "podes", and thereby break the square $[0, 1]^2$ into finitely many rectangular blocks, then the optimal graphon is constant on each block. As noted above, the interval [0, 1] is thought of as representing infinitely many nodes, so we are envisioning a decomposition of the set of nodes into a finite number of collections, also called podes, of 'equivalent' nodes, equivalent in the sense that two equivalent nodes have the same probability of connection with any other node. It only takes a few parameters to describe such a "multipodal" graphon, namely the sizes of the sub-intervals and the value of g on each block; this vast reduction in degrees of freedom is chiefly responsible for the proliferation of results in these models. Remarkably, this sort of multipodal graphon has turned up independently in other areas, most notably in community detection [1].

By a *phase* in our network model we mean an open connected region in the space of pairs $\{(e, t)\}$, at each point of which there is a unique optimal-entropy graphon $g_{(e,t)}$, whose parameters are *smooth* functions of (e, t). At a phase boundary there may be two or more optimal-entropy graphons, typically with qualitatively different structures, in which case we say the phase transition is *discontinuous*. Alternatively there may be a single optimal-entropy graphon, in which case we say the phase transition is *continuous*. These are in some sense analogous to first-order and second-order phase transitions in statistical mechanics, but the analogy can only be taken so far, especially since this is a mean-field model. (In particular, the expansion of the entropy as a power series in e and t sometimes involves fractional powers.)

In Fig. 1 we sketch the phase diagram of the microcanonical edge/triangle model [6]. This model exhibits three infinite families of phases labelled $A(m, 0)$, $B(m, 1)$ and $C(m, 2)$ as well as a singleton $F(1, 1)$. An $A(m, 0)$ graphon (that is, the optimizing graphon in the $A(m, 0)$ phase) has m podes and is invariant under permutation of these indistinguishable podes; see Fig. 2. A $B(m, 1)$ graphon has m indistinguishable podes and one pode that is different. A $C(m, 2)$ graphon has m indistinguishable podes and two additional podes that are different from the m podes but are indistinguishable from each other. The differences between these phases are reflected in the different eigenvectors of the adjacency matrices of generic networks in each phase.

We will be concentrating on the transition between the $B(1, 1)$ and $A(3, 0)$ phases as we go through the point $e = 0.67$, $t \approx 0.2596$ by varying t. This transition is particularly interesting because it has been proven to be discontinuous; the optimizing graphons for t slightly above 0.2596 are very different from the optimizing graphons for t slightly below 0.2596. (See Fig. 2.)

In switching from an arrangement described by one graphon to an arrangement described by the other, nodes need to rearrange their connections in a process analogous to nucleation

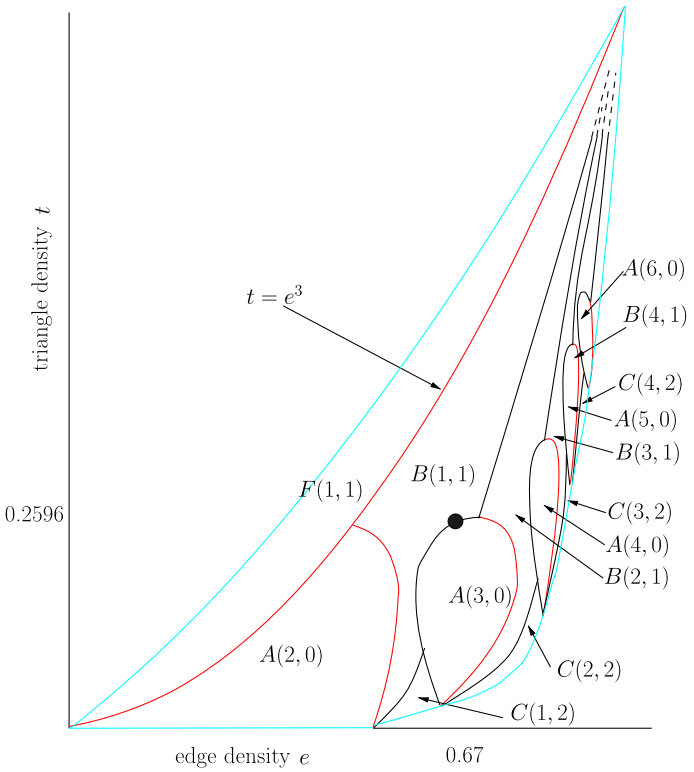

Fig. 1 Schematic sketch of 14 of the phases in the edge/triangle model. Continuous phase transitions are shown in red, and discontinuous phase transitions in black. The boundary of the phase space is in blue. The transition point studied in this paper is indicated by a black dot (Color figure online)

in a liquid/crystal transition. Patterns are broken, patterns are formed, and once sufficiently many nodes follow the new pattern, the rest follow suit.

Our study of this transition involved two major challenges. First, we needed to work with finite networks, yet the very notion of phases requires taking infinite-size limits. Although the graphon formalism ensures that systems of large enough size are modelled well by graphons, there is no mechanism in the formalism to determine how large is 'large enough'. Fortunately, we were able to see unambiguous evidence of the emergence of the $B(1,1)/A(3,0)$ transition in systems with as few as 30 nodes, well below the 100-or-so nodes that our computers can comfortably handle; we describe this in Sect. 3. The fact that such transitions can be seen in systems of such low size is significant both for the computer analysis of network transitions and for the modelling of scientific data of such size, and is the **first major result** of this paper.

Second, we needed to introduce a random dynamics to play the role of cooling or heating a material from an equilibrium state at one set of thermodynamic parameters to an equilibrium state at another set. Our dynamics respects the constraint on the edge density while introducing enough randomness to drive the triangle density across a transition.

The solutions we found are described in Sect. 4, both for going from $B(1,1)$ to $A(3,0)$ and for going from $A(3,0)$ to $B(1,1)$. Along the way we produced videos, available at https://www.youtube.com/playlist?list=PLZcI2rZdDGQqx3WEY6BXoJXcxqQXF5ZzQ

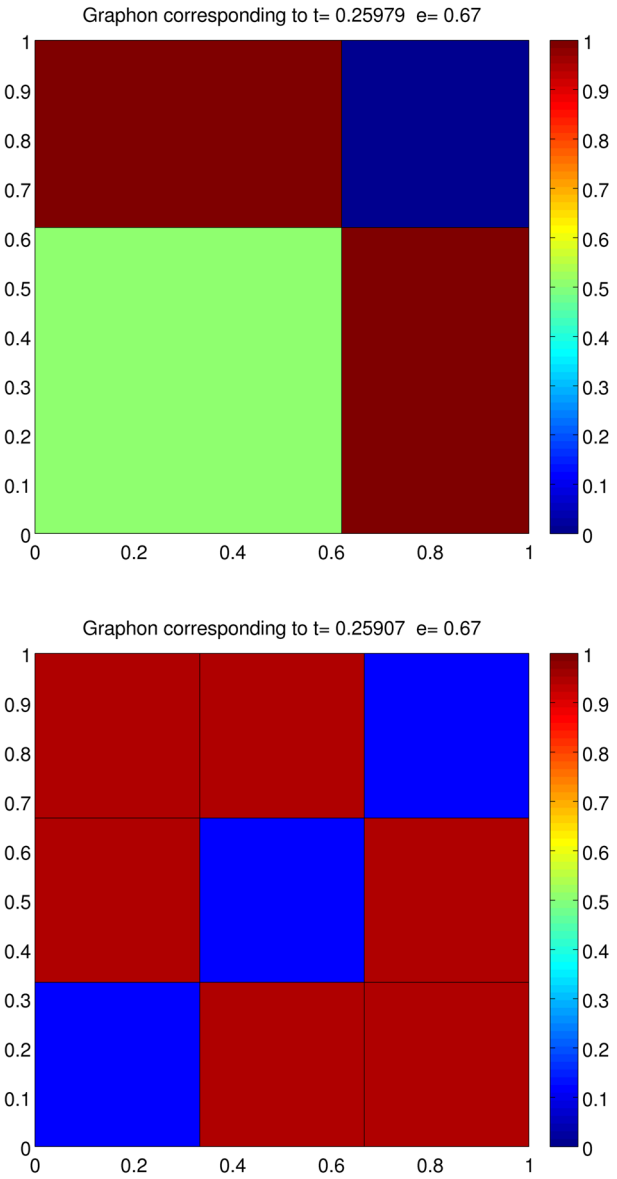

Fig. 2 Transition between phases $B(1,1)$ and $A(3,0)$ at edge density 0.670. Triangle densities of the two graphons: 0.25979, 0.25907. Triangle density at transition: 0.25963868

that literally show nodes rearranging into the pode structure appropriate to prescribed constraints. (The rearrangement is not in physical space, which plays no role in this mean-field model. Rather, the podal structure and its changes are seen in the eigenvectors of the adjacency matrix, as noted above.)

Understanding this process is the **second major result** of this paper. In each direction, the evolution from one phase structure to the other occurs in three distinct stages. The first

stage is very quick and brings the density t of triangles to its desired value. The second stage is a factor of 10–100 slower and results in a structure that can be viewed as "new phase plus defects", e.g. with podes that are slightly too big or too small. The third phase is a factor of 1000–10,000 slower than the first and achieves true equilibrium.

Finally, the time required for crystal nucleation in physical systems is random, as are other features of the process. We have begun an statistical analysis of the time required for each of the three stages of network nucleation (both for $A(3,0) \to B(1,1)$ and for $B(1,1) \to A(3,0)$), with details in Sect. 4.

3 Equilibrium in Finite Systems

Our goal is to understand how a network changes, under an appropriate dynamics, from the structure of one phase to that of another. Before we can do this, we must understand the structure of the two phases. We begin with the entropy-maximizing graphons that describe infinite (or sufficiently large) random networks, and then consider finite-size effects. For more about these graphons and the precise transition between them, see [6].

In the A(3,0) phase, an optimizing graphon $g_{e,t}$ is shown in the bottom of Fig. 2. If we denote the value of the graphon along the diagonal by a and the off-diagonal value by b, then this describes an infinite network with the following properties.

(1) The nodes group into three podes of equal size, which we might imagine as red, blue, and green nodes.
(2) The edges within each pode, e.g. connecting two red nodes, appear with probability a.
(3) The edges between different podes appear with probability $b > a$.
(4) The corresponding integral operator $M(g_{e,t})$ has rank three. Its nonzero eigenvalues are $(a + 2b)/3$ with multiplicity 1 and $(a - b)/3$ with multiplicity 2.
(5) The positive eigenvalue $(a + 2b)/3$ is the overall edge density, and the corresponding eigenfunction is constant.
(6) The triangle density is

$$t = \frac{1}{9}a^3 + \frac{2}{3}ab^2 + \frac{2}{9}b^3 = e^3 + 2\left(\frac{a-b}{3}\right)^2,$$

which is the trace of the cube of the integral operator. Given e and t, it is easy to compute a and b, the eigenvalues and eigenfunctions of $M(g_{e,t})$, and the densities of all subnetworks.
(7) The eigenspace corresponding to the negative eigenvalue $(a - b)/3$ consists of functions that are constant on each pode and integrate to zero. If we pick an orthonormal basis $\{\xi_1(x), \xi_2(x)\}$ for this eigenspace and pick an arbitrary point $x \in [0, 1]$, the ordered pair $(\xi_1(x), \xi_2(x))$ can lie at three possible points in $\mathbb{R}^2$, and these points form the vertices of an equilateral triangle centered at the origin. A change of basis only serves to rotate the triangle.

If we pick a network G with n nodes whose adjacency matrix $M(G)$ is close to this graphon (which necessarily requires n large), then $M(G)$ will have two large negative eigenvalues of approximate size $n(a - b)/3$ and one large positive eigenvalue of approximate size $n(a + 2b)/3$. The eigenvectors for the negative eigenvalues will be approximately constant on each pode, so that if we plot the values $\{(\xi_1(v_i), \xi_2(v_i))\}$ for the n nodes $\{v_i\}$, we will get (with high probability) three clusters of roughly equal size, located at the vertices of an equilateral

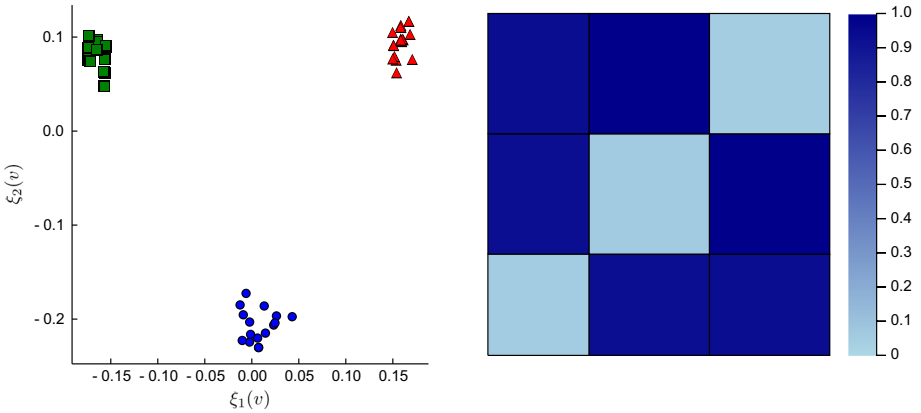

Fig. 3 $(\xi_1(v), \xi_2(v))$ plot and best-fit tripodal graphon for a typical graph with 54 nodes and triangle density $t = 0.24$. Each mark on the left corresponds to a node in the graph. Using spatial clustering on the left-hand plot, these nodes were divided into three podes (denoted by different colors and shapes in the left-hand plot). The plot on the right shows the empirical graphon corresponding to this division into three: the division of the axes represents the three podes, and the color of a region represents the density of edges between the corresponding podes

triangle. Moreover, the second largest negative eigenvalue λ_2 will be nearly as large as the largest negative eigenvalue λ_1, while all other negative eigenvalues will be of order $\sqrt{n}$.

We now turn to the $B(1, 1)$ phase, with an optimal graphon shown at the top of Fig. 2. The corresponding integral operator has rank two, and in particular only has one negative eigenvalue, whose eigenvector is constant on each pode. If we pick a large random network that is close to this structure, then the most negative eigenvalue λ_1 of the adjacency matrix will be $O(n)$, and the values of $\xi_1(v)$ will distinguish the two podes, but the second most negative eigenvalue λ_2 will only be $O(\sqrt{n})$. ξ_2 will be essentially zero on the smaller pode and random on the larger pode. A scatter plot of $\{(\xi_1(v_i), \xi_2(v_i))\}$ will then give a tight cluster of points corresponding to the small pode and a vertical bar corresponding to the large pode.

The upshot is that both the size of λ_2 and the distribution of $(\xi_1(v), \xi_2(v))$ are strong indicators of the phase we are in. Another indicator is the size of the podes themselves, as indicated by the size of the clusters in the (ξ_1, ξ_2) plot. In the $A(3, 0)$ phase, there are three podes of equal size. In $B(1, 1)$, there are two podes with a roughly 60-40 split between the larger and the smaller pode.

The left side of Fig. 3 shows the distribution of $(\xi_1(v), \xi_2(v))$ for a typical graph with $n = 54$, $e = 0.67$ and $t = 0.24$. Grouping nodes, we then generate a tripodal graphon by simply counting the number of edges that are blue-blue, blue-green, etc. Figure 4 is similar, only for a sample graph at $t = 0.26$, which is in the B(1,1) phase. In the first plot of Fig. 4, we attempt to segregate the nodes into three podes, as before, and to approximate this with a tripodal graphon, but there is no clear dividing line between the reddish triangles and squares. Based on this lack of clear separation, it may be more natural to treat such graphs as in the second plot, where the entire vertical bar is considered a single pode, yielding a bipodal graphon.

Of course, the uniform ensemble of networks of a fixed node number n and fixed numbers of edges and triangles are only proven to look like the optimal graphons in the $n \to \infty$ limit. We need a procedure to generate the ensemble for finite n. To do this, we implement a dynamics on the space of networks with given e, described in Subsect. 3.1, which has as

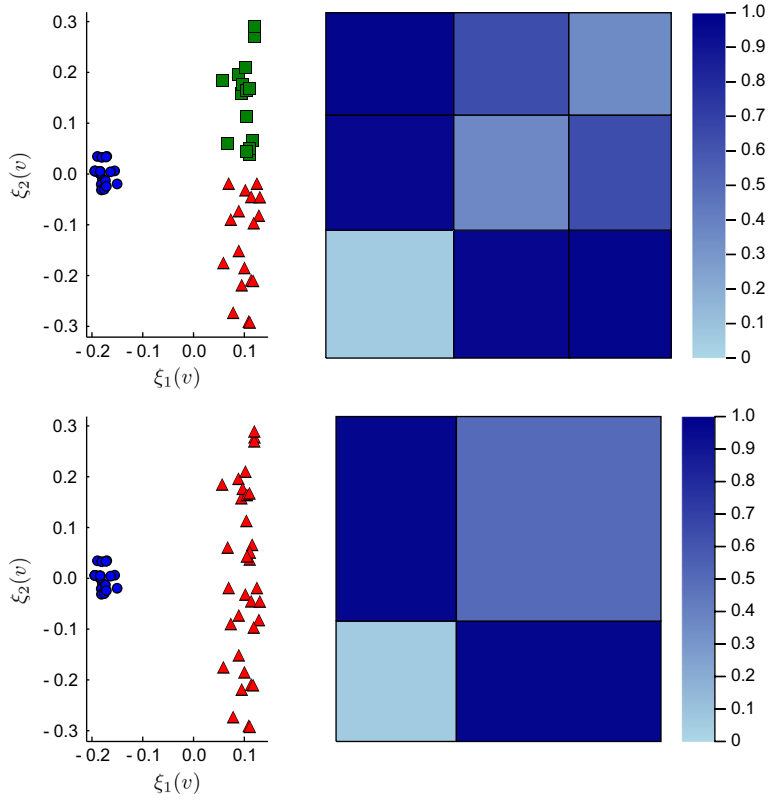

Fig. 4 $(\xi_1(v), \xi_2(v))$ plot and best fit tri- and bipodal graphons for a typical graph with 54 nodes and triangle density $t = 0.26$. While it is possible to group nodes into three podes, as in the first plot, it may be more natural to use two podes, as in the second

limit set the subset of those with given t. (The dynamics then plays the role of an adjustable heat bath.) We then run the dynamics 'sufficiently long' to obtain samples of this limit set. We next examine the histogram of values of λ_2 we obtained this way.

The histograms for $n = 54$, $e = 0.67$ and equal jumps of t from 0.24 to 0.26 are shown in Fig. 5. When $t = 0.248$ or $t = 0.250$ the variance in λ_2 is large, but when $t > 0.250$ or $t < 0.248$ the variance is much smaller. We interpret this as meaning that when $t > 0.250$ we are in one phase, when $t < 0.248$ we are in a different phase, and where the variance is large we have some sort of transition. Note that λ_2 is much more negative for values of t below the transition than above, which is what we would expect if the two phases corresponded to $B(1, 1)$ and $A(3, 0)$. The behavior of the eigenvectors is harder to quantify but easy to spot. As noted earlier, Figs. 3 and 4 show the distributions of $(\xi_1(v), \xi_2(v))$ for two representative sample networks, one substantially below and one substantially above the transition. The identification of the upper phase with $B(1, 1)$ and the lower phase with $A(3, 0)$ is clear.

Figure 6 shows what happens when $n = 30, 48$ or 66. Qualitatively, the picture is the same as for $n = 54$. Quantitatively, there are some small differences. The transition is sharper when n is larger, but is still visible when n is as small as 30. The specific location of the transition varies somewhat in n. The bigger n is, the closer the transition point is to $t = 0.2596$, which is the known location of the $B(1, 1)/A(3, 0)$ phase transition for infinite n.

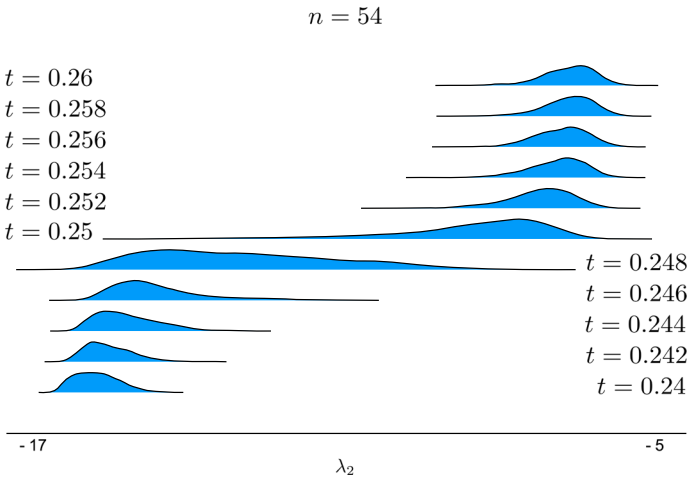

Fig. 5 Density plots showing the distribution of the second most negative eigenvalue λ_2, for graphs with 54 nodes, edge density 0.67, and various triangle densities t

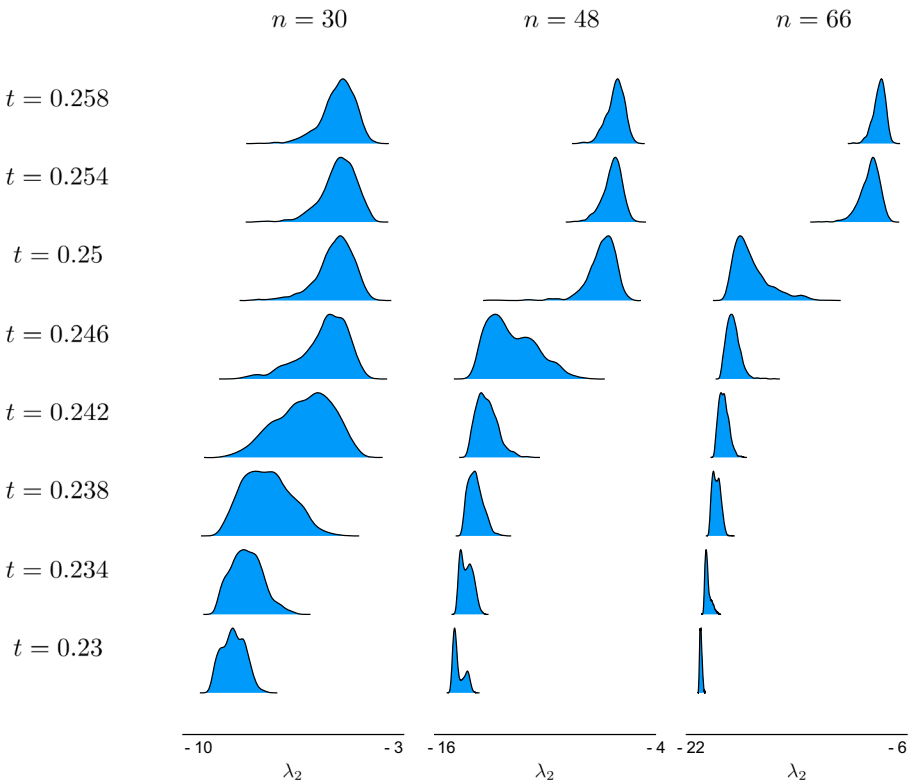

Fig. 6 Density plots showing the distribution of the second most negative eigenvalue λ_2 for $n = 30$, 48 and 66 nodes, edge density 0.67, and various triangle densities t

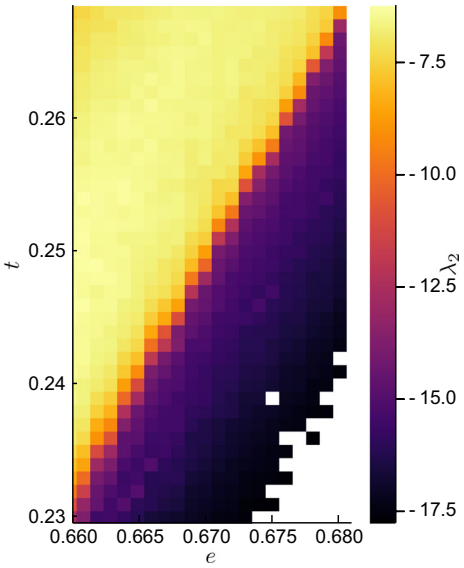

Fig. 7 The second-most-negative eigenvalue λ_2 of a typical 54-node graph, as a function of edge density e and triangle density t. The phase transition is clearly visible as a curve along which λ_2 changes rapidly as a function of e and t

The location of the transition for small n is particularly intriguing. There is a lower limit to the value of t achievable by a bipodal network in which each pode has an Erdős-Rényi structure. For the range of values of n considered in this paper (30 to 66), the transition actually occurs *below* this point! In this range of t-values, the larger pode develops some sort of n-dependent internal structure that we do not yet understand. This is a subject of continuing research.

Finally, we examine whether our results are sensitive to changes in the edge density $e = 0.67$. They are not. Figure 7 shows the mean value of λ_2 as a function of (e, t). There is a sharp transition for all values of e. All that changes with e is the specific value of t at which it occurs.

Having established that the phenomena we are investigating are insensitive to changes in e or n, we will henceforth restrict attention to $n = 54$ and $e = 0.67$.

3.1 Sampling from Equilibrium

We use Markov Chain Monte Carlo (MCMC) to (approximately) uniformly sample graphs with specific numbers of edges and triangles. In order to ensure that our results are not sensitive to the sampling procedure, we carry this out in several different ways, all of which follow the same general procedure: given fixed number n of vertices, and given fixed edge and triangle counts E and T, we define the following three objects:

(1) a constraint set Ω which contains all graphs on n vertices with E edges and T triangles,
(2) a function d_Ω measuring the "distance" from a graph to Ω, and
(3) for each graph G, a probability measure Q_G on graphs with n nodes (the "proposal distribution").

We use this data to define a Metropolis-Hastings Markov chain with the following update rule: given the current state G (a graph on n nodes), we draw the graph G' from Q_G. If $G' \in \Omega$, it becomes the next state of the Markov chain. If $G' \notin \Omega$ but $d_\Omega(G') \leq d_\Omega(G)$, G' becomes the next state of the Markov chain. Otherwise, the Markov chain remains at G.

The Markov chain defined above is reversible, and the uniform measure on Ω is a stationary measure. Moreover, if Ω is connected under the proposal distributions then the uniform measure on Ω is the unique stationary measure; if in addition the starting point is connected to Ω under the proposal distribution then the Markov chain will converge to the uniform measure on Ω. Finally, we generate a random graph by running the Markov chain for a long time (for most of the empirical results presented here, 10^9 accepted steps were sufficient) and accepting the final graph if it had exactly E edges and T triangles.

All of the results presented here were checked with the following three choices of Ω, d_Ω, and Q:

(1) Ω is the set of graphs with exactly E edges and at most T triangles; $d_\Omega(G) = \max\{0, T(G) - T\}$; and Q_G is the random graph obtained from G by deleting an edge at random and adding an edge at random;
(2) Ω and d_Ω are as above; Q_G is the random graph obtained from G by choosing a vertex at random, then deleting an incident edge at random and adding an incident edge at random;
(3) Ω is the set of graphs with $|E(G) - E| \leq C$ and $|T(G) - T| \leq C$ for some constant C; $d_\Omega(G) = \max\{0, |E(G) - E| + |T(G) - T| - 2C\}$; and Q_G is the random graph obtained from G by either deleting a random edge, adding a random edge, or deleting a random edge and adding a random edge.

Note that in all three cases, Ω is strictly larger than the set of graphs with E edges and T triangles. Therefore, the final step of our sampling algorithm is to accept the final graph if it meets the edge/triangle constraint, and to try again if it does not. This final step turns out to succeed often, particularly for the first two choices above, because when $T < E^3$ then a substantial fraction of the graphs with at most T triangles turn out to have exactly T triangles.

The figures and numbers that we present in the main body of this work are all for the first choice of dynamics. The other two choices of dynamics showed the same qualitative behavior but mixed more slowly; the third choice of dynamics in particular has a tendency to "get stuck" for long periods, since after deleting an edge it sometimes struggles to find another move that respects the triangle constraints.

3.1.1 Evaluating Convergence

Since we are not able to prove bounds on the mixing time of any of the Markov chains above, we validated the convergence of our Markov chains by verifying that the distribution of certain statistics of interest (for example, the second-smallest eigenvalue of the resulting graphs) were independent of the Markov chains' initial states. Specifically, we initialized our various Markov chains from three different initial distributions:

(1) a uniformly random graph with exactly E edges,
(2) a graph drawn from a $B(1, 1)$ graphon, and
(3) a graph drawn from an $A(3, 0)$ graphon.

Note that these three input distributions have very different spectral properties. After running the Markov chains above, however, we verified that the spectral statistics of the outputs did not depend on which of the three initializations we chose.

4 Structural Rearrangement Under Change of Phase

We now turn to the central question of this paper. We have seen that for $t = t_{low} = 0.24$, the ensemble is dominated by $A(3, 0)$ networks, as evidenced by λ_2 being very negative, and by

(ξ_1, ξ_2) being clustered around three points, with equal numbers of nodes near each point. When $t = t_{high} = 0.26$, the ensemble is dominated by networks with λ_2 much smaller, with the (ξ_1, ξ_2) plot being a vertical bar (representing 60% of the nodes) and a separate cluster (40%). But how does the network change from one kind of network to the other under our dynamics, as described in Sect. 3?

The supplementary material contains several animations that show the evolution of (ξ_1, ξ_2) and various other statistics across the three stages. In this section, we do our best to summarize the main features using figures and text. However, we highly recommend viewing the actual animations. A picture is worth a thousand words, and a movie is worth a thousand pictures.

4.1 Increasing *t*

We display 100 short runs of 10,000 steps in Fig. 8 and two long runs of 1,000,000 steps in Fig. 9. Both figures show λ_2 as a function of time. Figure 9 also shows the number of embedded "two ear" graphs, where a "two ear" is two triangles sharing a common edge, or equivalently a tetrahedron with one edge missing. The data in Fig. 9 has been smoothed, with both λ_2 and the 2-ear count averaged over 1000 successive times. The data in Fig. 8 has been smoothed by averaging over 50 successive times.

A typical transition occurs in three dynamical stages. The first two are visible in Fig. 8 and the third is visible in Fig. 9. In the first stage, which only takes about 50 steps, the network retains its $A(3, 0)$ structure as t increases from t_{low} to t_{high}. In this stage λ_2 increases somewhat as the parameters a and b change, but the distribution of (ξ_1, ξ_2) does not significantly change. This stage is visible in Fig. 8 as a sudden, but modest, increase in λ_2 at the very beginning of each plot.

In the second stage, which usually takes between 1000-6000 steps, the network loses its $A(3, 0)$ structure as two of the clusters merge. This stage is marked by an increase in λ_2 and a dramatic change in the distribution of (ξ_1, ξ_2). At the end of this stage, we have a structure that is very similar to the optimal $B(1, 1)$ graphon, except that the podes do not have the optimal sizes, being closer to $2 : 1$ than to 60:40. This stage is visible in Fig. 8, and accounts for most of the visible change in λ_2. In the last stage, which takes hundreds of thousands of steps, the relative sizes of the podes adjust back and forth among a number of possibilities with nearly equal entropies, resulting in fluctuations in λ_2 and the 2-ear count over very long time scales. The beginning of this stage can be seen in Fig. 8 as the point at which λ_2 becomes approximately constant over time. In order to actually see the pode sizes changing, we have to view the trajectories at the much larger time-scale shown in Fig. 9.

The first stage is driven entirely by entropy. Since all networks are well below the target number of triangles, all swaps of one edge for another are allowed. This results in edges simply moving from where they are concentrated to where they are not. Since $a < b$, a gets bigger and b gets smaller, but (aside from random noise) the graphon remains piecewise constant, with the three podes remaining indistinguishable. This is reflected in the fact that plots of (ξ_1, ξ_2) basically do not change. Since λ_2 is proportional to $a - b$, λ_2 increases somewhat. For $n = 54$, it goes from a little below -16 to around -14.5.

In the second stage, the network explores the available phase space for $t = t_{high}$ starting from an $A(3, 0)$ structure. Within each cluster, the variances in ξ_1 and ξ_2 increase, and an occasional vertex may move from one cluster to another. In a typical run, there is a latency period during which λ_2 gradually increases from around -14.5 to around -11, after which two of the clusters join and λ_2 rises quickly to around -7. During the merger, all three clusters

Fig. 8 100 repetitions showing the trajectory of λ_2 as a function of time, as the triangle density is increased from 0.24 to 0.26 over 10,000 steps. The red and green lines show the least-squares best fit of the trajectory by a linear function (the red line) followed by a constant function (the green line)

spread out, and sometimes the third cluster gains or loses a vertex or two. Figure 10 shows the evolution of ξ_1 and ξ_2 during this second stage.

The third stage involves punctuated equilibrium. Most of the time, both λ_2 and the 2-ear density fluctuate in a range, and the distribution of (ξ_1, ξ_2) does not change qualitatively. Once in a while, a node moves from one pode to another, triggering sudden changes in both λ_2 and the two-ear count. When the new ratio of pode sizes is close to the optimal ratio, the system can stabilize in the new state for almost 100,000 steps. If the new ratio of pode sizes is far from optimal, however, entropy drives them to change back more quickly, usually within 10,000 steps. Such an extreme excursion is visible at time 150,000 in the first run, where the smaller pode grows to 24 nodes and remains there for about 10,000 steps before eventually making its way back to 21 nodes at time 170,000. On its way back, the size of the smaller pode briefly (for about 10,000 steps) stabilizes at 23, corresponding to the two-ear count stabilizing at around 111,150.

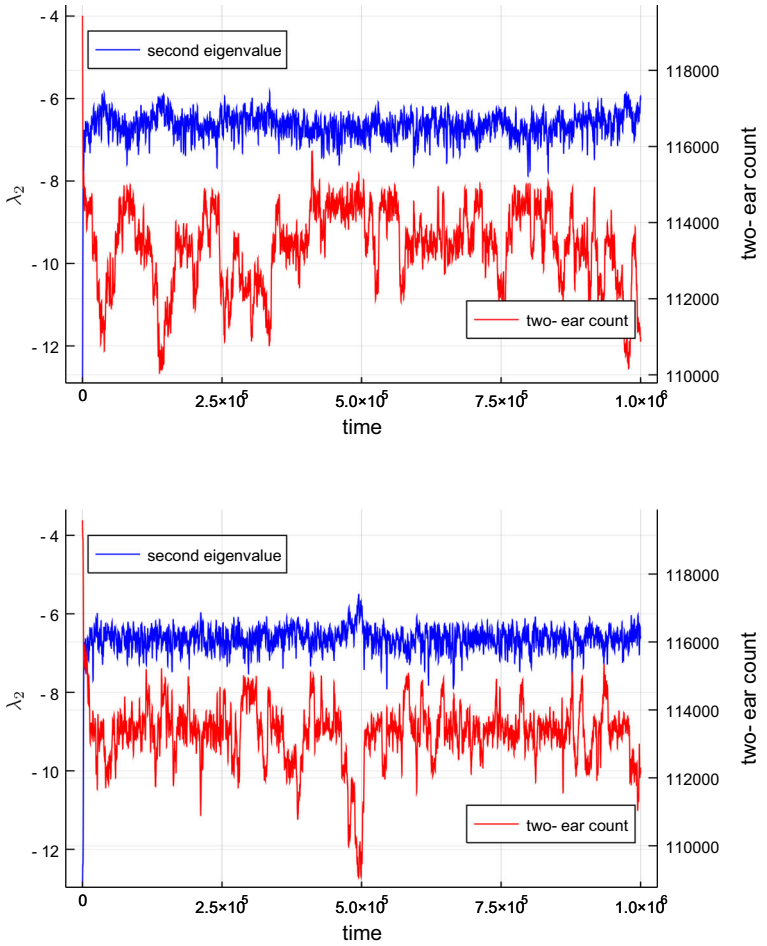

Fig. 9 Two different random trajectories of λ_2 (in blue), and the number of embedded "two ear" graphs (in red) as a function of time, as the triangle count is increased from 0.24 to 0.26 over 10^6 steps. The data is smoothed by averaging over windows of 1000 steps

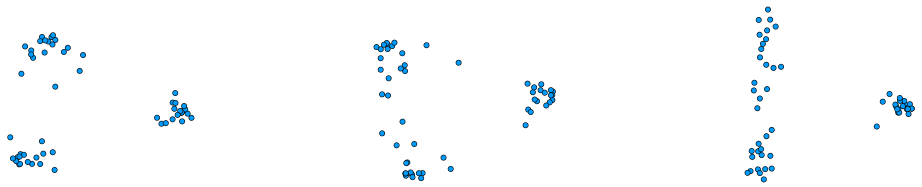

Fig. 10 Snapshots of (ξ_1, ξ_2) after 500, 800, and 1500 steps during an upward trajectory. Each dot corresponds to a vertex v; the coordinates of the dot are given by $\xi_1(v)$ and $\xi_2(v)$. The transition can also be viewed in the animations

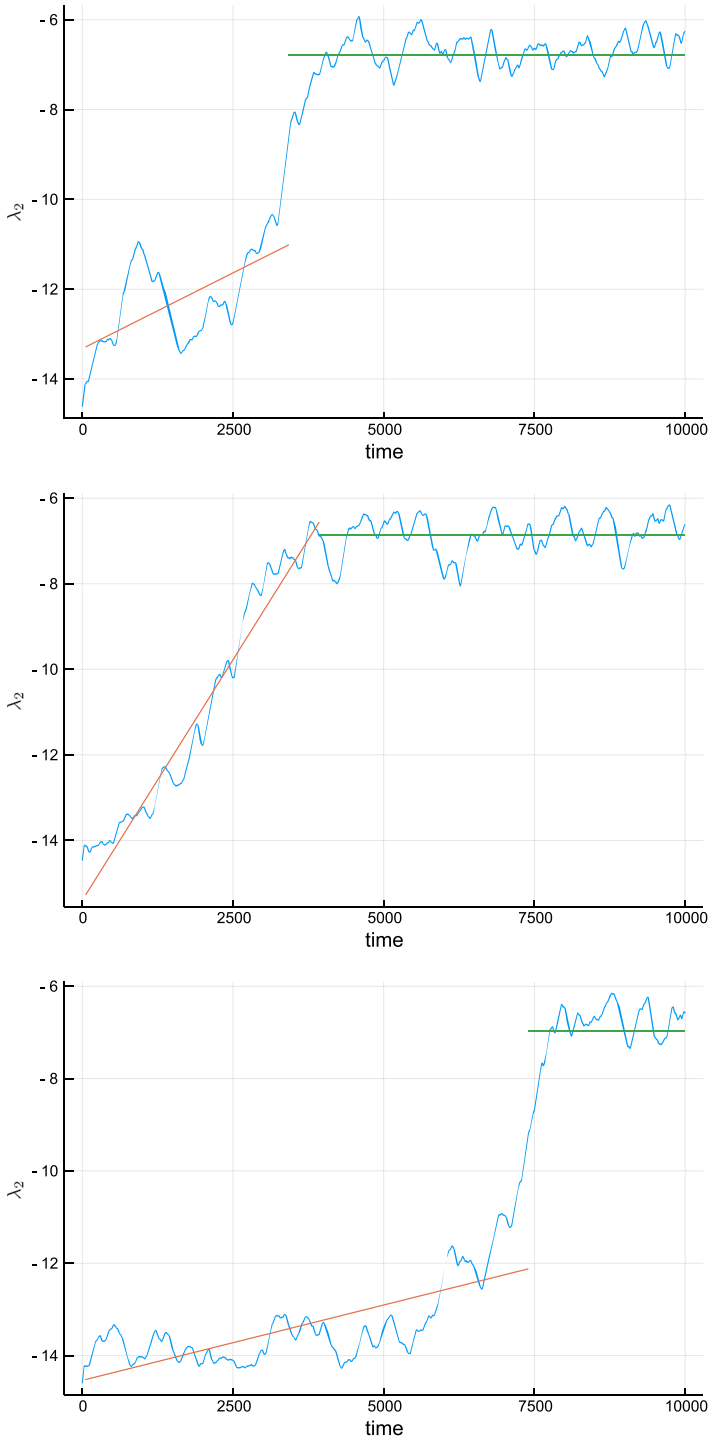

Fig. 11 Close-ups of the 17th, 23rd, and 65th trajectories from Fig. 8, smoothed over 100 steps. These three trajectories can also be seen in the animations

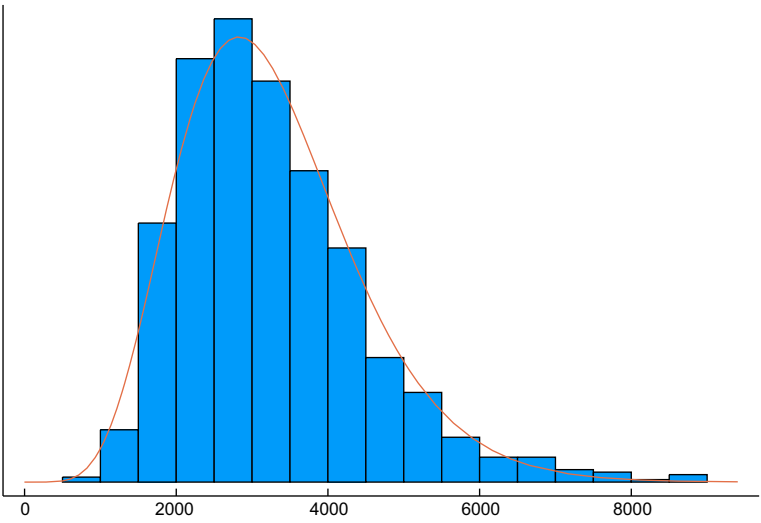

Fig. 12 The histogram of the length of stage 2 for upwards trajectories, and a best-fit Gamma function with ($\alpha = 7.5$ and $\theta = 430$)

The end of Stage 1 and the beginning of Stage 2 is easy to define. This is when t first hits its target value, after which the constraint on t begins to affect which edge swaps are allowed. This consistently takes between 45 and 60 steps, with an average of 50.

The end of Stage 2 / beginning of Stage 3 is harder to define. To do this, we did a least-squares fit of the first 10,000 steps, minus Stage 1, via a line of arbitrary slope followed by a horizontal line. These best fit lines are shown in Fig. 8. We interpret the domain of the sloped line as Stage 2 and the domain of the horizontal line as Stage 3.

The evolution of λ_2 (smoothed over 100 steps) in runs 17, 23 and 65 is shown in more detail in Fig. 11. In a run with a long latency, such as run 65, the sloped line tracks the gradual increase during the latency period, and the subsequent rapid rise shows up as a discontinuity between the lines. In runs with short latencies, such as run 23, the entire increase from -14.5 to -7 is captured by the sloped line, so there is no discontinuity. In a few runs, such as run 17, the latency period involves some false starts, where λ_2 first increases and then decreases, and the fit to a sloped line is not good at all. However, even in these exceptional cases, the beginning of the horizontal line provides a credible measure of the beginning of Stage 3.

We have plotted the distribution of the length of Stage 2 in Fig. 12. The shape of this histogram is modeled well by a Gamma distribution with $\alpha = 7.5$ and $\theta = 430$ (i.e. the function $x^{6.5}e^{-x/430}$), which we have superimposed on the histogram.

4.2 Decreasing *t*

As with increasing t, the trajectories for decreasing t exhibit three stages of increasing length. We present the data, much as before, in Figs. 13, 14, 16, and 17. These show 100 trajectories for 5,000 steps, two trajectories for 1,000,000 steps, close-ups of three of the short downwards trajectories, and a histogram of the lengths of Stage 2, respectively. The long runs show both λ_2 and the 2-ear count, while the short runs only show λ_2. The short runs illustrate what happens in Stages 1 and 2, while the long runs explore Stage 3.

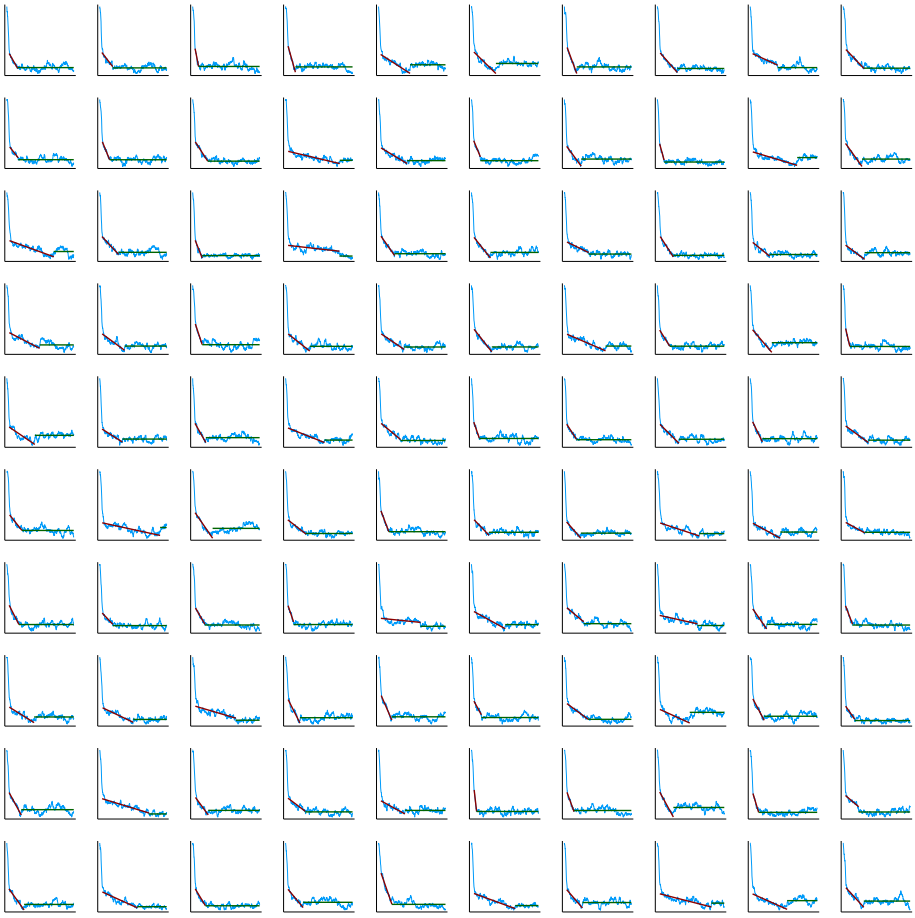

Fig. 13 100 repetitions showing the trajectory of λ_2 as a function of time, as the triangle density is decreased from 0.26 to 0.24 over 5000 steps. The red and green lines show the least-squares best fit of the trajectory by a linear function (the red line) followed by a constant function (the green line). The trajectories are smoothed over 100 timesteps (Color figure online)

However, there are some important differences between the dynamics for decreasing t and the dynamics for increasing t. First, because the entropy at $t = t_{low}$ is substantially lower than the entropy at $t = t_{high}$, the high-to-low transition has to be forced. As long as t is above the target value, the dynamics only allow edge swaps that decrease the number of triangles, or at worst keep the number constant. The initial stage of the transition, rather than being an entropy-driven relaxation, is essentially a greedy algorithm for decreasing t. This stage is a little slower than the first stage of the upward transitions, taking around 180 steps instead of about 50.

Second, the qualitative form of the graph cannot stay the same throughout Stage 1. Below $t = 0.26$, it is mathematically impossible to achieve the desired number of triangles with two podes and with Erdős-Rényi structure within each pode. Instead, as noted above, the system has to develop additional structure within the larger pode. As t drops, vertices in this pode start to segregate into two sub-podes, with edges being more likely between sub-podes than

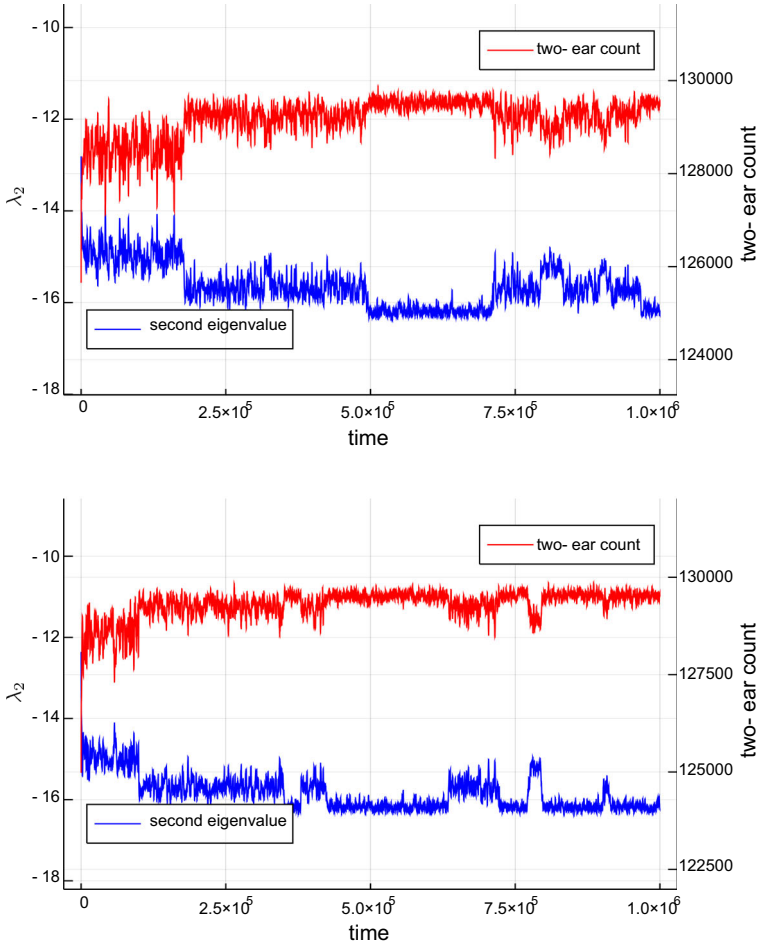

Fig. 14 Two different random trajectories of λ_2 (in blue), and the number of embedded "two ear" graphs (in red) as a function of time, as the triangle count is decreased from 0.26 to 0.24 over 10^6 steps. The data is smoothed by averaging over windows of 1000 steps

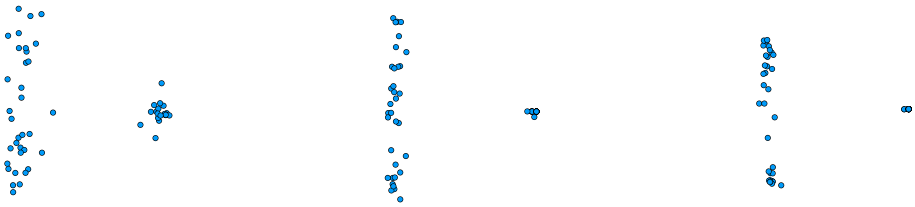

Fig. 15 Snapshots of (ξ_1, ξ_2) after 10, 100, and 200 steps during an downward trajectory. Each dot corresponds to a vertex v; the coordinates of the dot are given by $\xi_1(v)$ and $\xi_2(v)$. The podes are not actually changing size. Rather, values of (ξ_1, ξ_2) are becoming so closely positioned that multiple nodes only show up as a single dot. This transition can also be seen in the animations

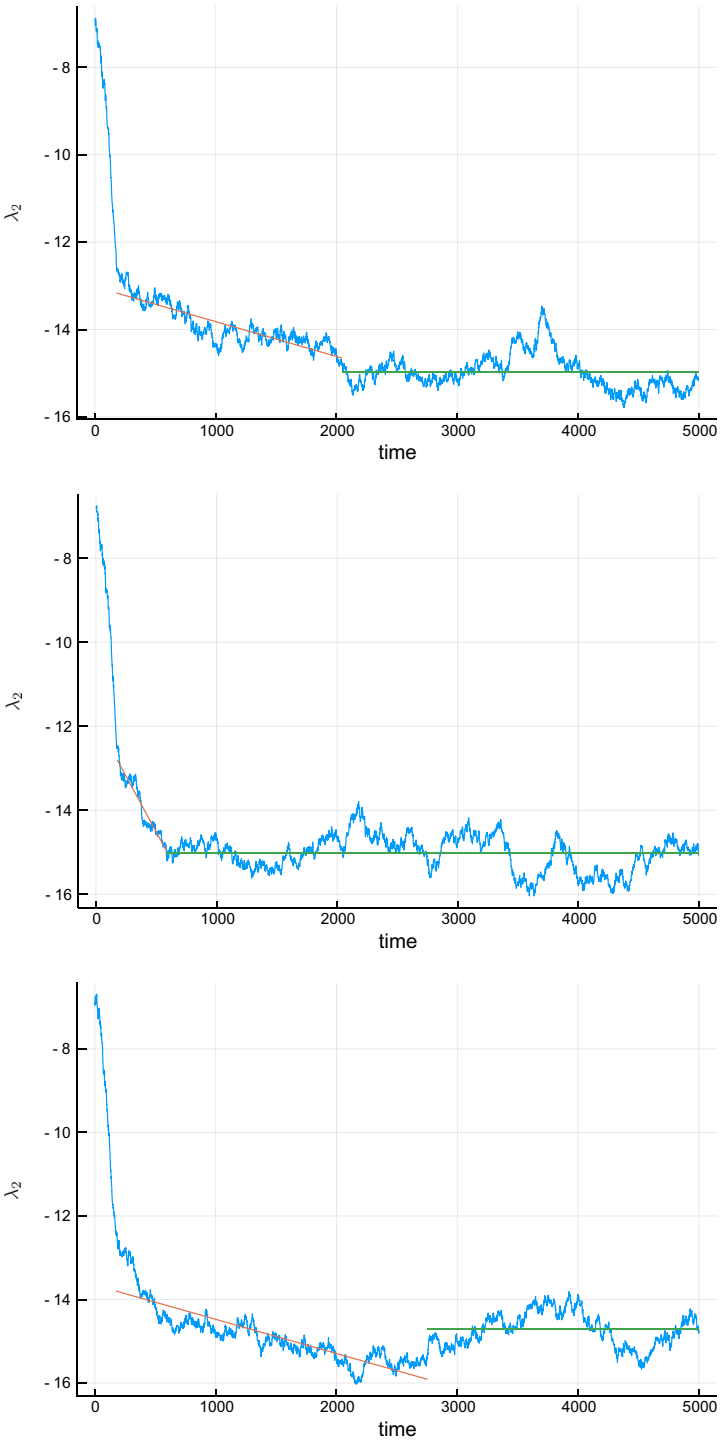

Fig. 16 Close-ups of the 9th, 46th, and 99th trajectories in Fig. 13, not smoothed. These three trajectories can also be viewed in the animations

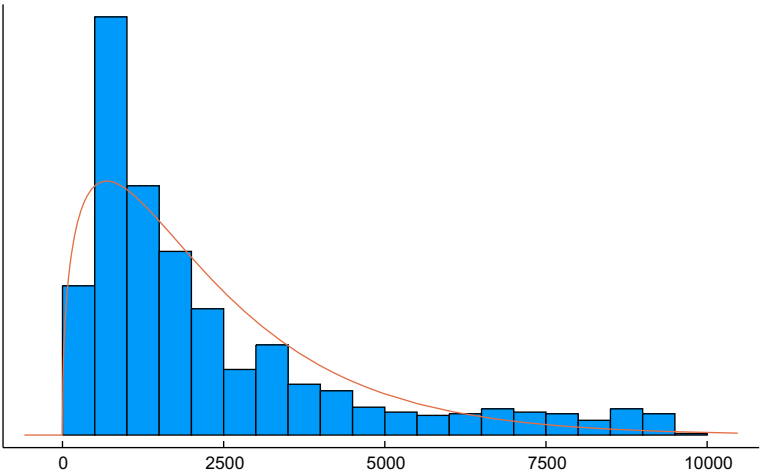

Fig. 17 The histogram of the length of Stage 2 for downwards trajectories, and a best-fit Gamma function with parameters $\alpha = 2.25$ and $\theta = 660$

within a sub-pode. When we reach t_{low}, this segregation is still not complete, but the vertical bar in the (ξ_1, ξ_2) plot has become more like a dumbbell, with most points near one end or the other, as can be seen in Fig. 15. At this point, λ_2 has dropped to around -13.

Once we reach t_{low}, the dynamics no longer force us to decrease t. The second stage is shorter than for increasing t, and lasts on the order of 1000 edge swaps. In this stage, λ_2 continues to drop to a value near -15 as the two sub-podes become better defined. (The division of the original large cluster is not always exactly even, and this affects the resulting values of λ_2.) However, the decrease in λ_2 is not as rapid as in Stage 1 and is no longer monotonic. Indeed, the onset of this stage can be viewed not only as the point where t reaches the target value, but also as the point where the plot of λ_2 versus time starts to become noisy. This is clearly seen in Fig. 16, where the data has not been smoothed. At the end of the second stage, our networks now involve two (more-or-less) identical clusters, each a subset of the original large pode, and a third cluster that is different, with no evidence of any internal structure within these three new podes.

There is considerable variation in the length of Stage 2, as can be seen in Fig. 13. As before, we define this length by doing a fit of the data after the end of Stage 1, and until the 10,000th step, by a slanted line followed by a horizontal line. Three such fits (cut off after the 5000th step) are shown in Fig. 16. In most cases, such as run 9 with a long Stage 2 and run 46 with a very short Stage 2, the lines fit together clearly, with little discontinuity. In a few cases, such as run 99, λ_2 overshoots the mark and rebounds, or shows other behavior that does not really fit a linear model, making it difficult to pinpoint the end of Stage 2. The distribution of the length of Stage 2 is shown in Fig. 17, together with a best-fit Gamma functions, with parameters $\alpha = 2.25$ and $\theta = 660$. Compared to the similar histogram for the upwards trajectory, the value of α is much smaller, and the fit is not as good, perhaps because of outliers such as run 99.

The third stage is much more stable than for increasing t. There are still frequent excursions, where a node attempts to leave its pode and usually comes back. These appear as short upwards spikes in λ_2, with simultaneous downward spikes in the 2-ear count. Successful migrations are much less common. For instance, in the first plot of Fig. 14, the first clear

migration only occurs after about 180,000 steps. There is another one after 480,000 steps, and a third at about 700,000 steps, plus less clear events at about 790,000, 830,000, and 960,000 steps. The lowest values of λ_2, around -16.3, are associated with the most symmetric configurations, with 18 nodes in each pode, while higher plateaus around -15.7 and -15.2 represent 19-18-17 or 20-17-17 or 19-19-16 splits.

5 Conclusion and Open Problems

In this paper we have considered phases and phase transitions in ensembles of finite random graphs as the densities of edges and triangles is varied. We have seen that

(1) The theory of large dense random graphs is applicable to remarkably small random graphs. While the terms "phase" and "phase transition" only have a precise mathematical definition in the $n \to \infty$ limit, we can meaningly speak of finite graphs being in one phase or the other in the same way that we speak of a finite sample of H_2O being water or ice. In both settings, this means that the structure of the finite system is well-approximated by the idealized structure of the infinite system. Near the "B(1,1) to A(3,0)" transition, the ensemble of random graphs with n as small as 30 exhibits the same phase structure as in the $n \to \infty$ limit.
(2) Spectral data, in particular the two most negative eigenvalues of the adjacency matrix and their corresponding eigenvectors, provide useful "order parameters" for distinguishing between phases.
(3) For fixed $e = 0.67$, the transition for n between 30 and 100 occurs at a lower value of t than in the $n \to \infty$ limit. As n increases, the transition becomes both sharper and higher. This phenomenon seems to occur for many transitions in the edge-triangle model, not just B(1,1) $\leftrightarrow$ A(3,0), and is the subject of ongoing research.
(4) By setting up a random process involving edge flips, with contraints keyed to driving t from one value to another, we can model the process by which a graph in one phase changes into a graph in the other.
(5) The transition from $A(3, 0)$ to $B(1, 1)$, with t increasing, occurs in three stages.
    (a) Stage 1 is very quick, taking only about 50 flips. In this stage t reaches a value normally associated with the $B(1, 1)$ phase, but the structure of the graph remains that of the $A(3, 0)$ phase. Specifically, the second eigenvalue only gets a little smaller in magnitude, while the eigenvectors do not change at all.
    (b) In Stage 2, which takes 1000–6000 flips, the second eigenvalue becomes much smaller and the scatter plot of the two most negative eigenvectors starts to resemble that of a $B(1, 1)$ graph. However, the number of vertices associated to each pode is not optimal. This stage typically shows a latency period followed by a period of rapid change. The length of this latency period is modeled well by a Gamma distribution.
    (c) It takes hundreds of thousands of edge flips to equilibrate the sizes of the podes. We call this process Stage 3.
(6) The transition from $B(1, 1)$ to $A(3, 0)$, with t decreasing, likewise occurs in three stages. These stages are similar to the three stages when t decreases, but their description isn't quite as simple.
    (a) Stage 1, where t achieves its target value, is longer than for increasing t, taking around 180 steps instead of 50. In this stage the graph remains bipodal, but one of

the podes develops an internal structure that is reflected in the number of triangles, but is not reflected in the two leading eigenvectors.

(b) Stage 2, where the scatter plot of the eigenfunctions aggregates into three distinct clusters, is faster than the other direction, taking on the order of 1000 steps. The description of Stage 2 as latency followed by rapid change is not as clean as in the other direction, and the length of the latency period is no longer well modeled by a Gamma distribution.

(c) Stage 3 is a process of punctuated equilibrium as the sizes of the three podes change among 18–18–18 and 19–18–17, with occasional excursions to 20–17–17 or 19–19–16.

(7) Verbal descriptions cannot match actually seeing the clusters break up and form a bar (in the $A(3, 0)$ to $B(1, 1)$ transition) or seeing a bar break up and form two clusters (in the $B(1, 1)$ to $A(3, 0)$ transition). The best way to appreciate what is happening in these transitions is simply to watch the videos.

Challenges for future work include:

(1) Quantifying the sharpness of the transition.
(2) Seeing how this sharpness, and the deviation of the transition point from the $n \to \infty$ limit, scale with n.
(3) Determining whether this scaling behavior applies to all phase transitions in the edge-triangle model, or whether different kinds of transitions show different behavior.

References

1. Abbe, E.: Community detection and stochastic block models: recent developments. J. Mach. Learn. Res. **18**, 6446–6531 (2017)
2. Borgs, C., Chayes, J., Lovász, L.: Moments of two-variable functions and the uniqueness of graph limits. Geom. Funct. Anal. **19**, 1597–1619 (2010)
3. Borgs, C., Chayes, J., Lovász, L., Sós, V.T., Vesztergombi, K.: Convergent graph sequences I: subgraph frequencies, metric properties, and testing. Adv. Math. **219**, 1801–1851 (2008)
4. Chatterjee, S., Diaconis, P.: Estimating and understanding exponential random graph models. Ann. Stat. **41**, 2428–2461 (2013)
5. Chatterjee, S., Varadhan, S.R.S.: The large deviation principle for the Erdős–Rényi random graph. Eur. J. Comb. **32**, 1000–1017 (2011)
6. Kenyon, R., Radin, C., Ren, K., Sadun, L.: Multipodal structure and phase transitions in large constrained graphs. J. Stat. Phys. **168**, 233–258 (2017)
7. Kenyon, R., Radin, C., Ren, K., Sadun, L.: Bipodal structure in oversaturated random graphs. Int. Math. Res. Notices **2016**, 1009–1044 (2018)
8. Kenyon, R., Radin, C., Ren, K., Sadun, L.: The phases of large networks with edge and triangle constraints. J. Phys. A **50**, 435001 (2017)
9. Koch, H.: Vertex order in some large constrained random graphs. SIAM J. Math. Anal. **48**, 2588–2601 (2016)
10. Lovász, L.: Large Networks and Graph Limits. American Mathematical Society, Providence (2012)
11. Lovász, L., Szegedy, B.: Limits of dense graph sequences. J. Comb. Theory Ser. B **98**, 933–957 (2006)
12. Lovász, L., Szegedy, B.: Szemerédi's lemma for the analyst. GAFA **17**, 252–270 (2007)
13. Lovász, L., Szegedy, B.: Finitely forcible graphons. J. Comb. Theory Ser. B **101**, 269–301 (2011)
14. Newman, M.E.J.: Networks: An Introduction. Oxford University Press, Oxford (2010)
15. Radin, C., Sadun, L.: Phase transitions in a complex network. J. Phys. A **46**, 305002 (2013)
16. Radin, C., Sadun, L.: Singularities in the entropy of asymptotically large simple graphs. J. Stat. Phys. **158**, 853–865 (2015)
17. Radin, C., Ren, K., Sadun, L.: The asymptotics of large constrained graphs. J. Phys. A **47**, 175001 (2014)
18. Radin, C., Ren, K., Sadun, L.: A symmetry breaking transition in the edge/triangle network model. Ann. Inst. H. Poincaré D **5**, 251–286 (2018)

19. Radin, C., Swinney, H.: Phases of granular matter. J. Stat. Phys. **175**, 542–553 (2019)
20. Rietz, F., Radin, C., Swinney, H., Schröter, M.: Nucleation in sheared granular matter. Phys. Rev. Lett. **120**, 055701 (2018)
21. Strauss, D.: On a general class of models for interaction. SIAM Rev. **28**, 513–527 (1986)
22. Uhlenbeck, G.E.: In: Cohen, E.G.D. (ed.) Fundamental Problems in Statistical Mechanics II, pp. 16–17. Wiley, New York (1968)