The cut metric, random graphs, and branching processes

Béla Bollobás[*†‡] Svante Janson[§] Oliver Riordan[¶]

January 14, 2009; revised February 06, 2010

Abstract

In this paper we study the component structure of random graphs with independence between the edges. Under mild assumptions, we determine whether there is a giant component, and find its asymptotic size when it exists. We assume that the sequence of matrices of edge probabilities converges to an appropriate limit object (a kernel), but only in a very weak sense, namely in the cut metric. Our results thus generalize previous results on the phase transition in the already very general inhomogeneous random graph model introduced by the present authors in [4], as well as related results of Bollobás, Borgs, Chayes and Riordan [3], all of which involve considerably stronger assumptions. We also prove corresponding results for random hypergraphs; these generalize our results on the phase transition in inhomogeneous random graphs with clustering [5].

1 Introduction and results

Throughout this paper we consider random graphs with independence between the edges. The distribution of a random n-vertex graph with this property is of course specified by the matrix of edge probabilities; here we are interested in the asymptotic behaviour of the component structure as $n \to \infty$, so we shall consider a sequence of such matrices. Our main focus is to determine when there is whp a giant component, i.e., a component containing $\Theta(n)$ vertices. Here, as usual, an event holds *with high probability*, or *whp*, if it holds with probability $1 - o(1)$ as $n \to \infty$. When there is a giant component, we shall also find its asymptotic size.

For these questions it is natural to focus on (extremely) *sparse* graphs, with $\Theta(n)$ edges, so we shall normalize by considering matrices A_n whose entries are

[*]Department of Mathematical Sciences, University of Memphis, Memphis TN 38152, USA
[†]Trinity College, Cambridge CB2 1TQ, UK
[‡]Research supported in part by NSF grants DMS-0906634, CNS-0721983 and CCF-0728928, and ARO grant W911NF-06-1-0076
[§]Department of Mathematics, Uppsala University, PO Box 480, SE-751 06 Uppsala, Sweden
[¶]Mathematical Institute, University of Oxford, 24–29 St Giles', Oxford OX1 3LB, UK

n times the corresponding edge probabilities. Thus the case in which each A_n has all (off-diagonal) entries equal to some $c > 0$ corresponds to the classical sparse model $G(n, c/n)$. Without some further assumptions, it seems difficult to prove asymptotic results, although Alon [1] did so for some questions concerning connectedness. As in previous work, the natural additional assumption turns out to be convergence to a suitable limiting object, namely a *kernel*, i.e., a symmetric non-negative function on $[0, 1]^2$. Our aim is to relate the asymptotic size of the giant component to a suitable function of this kernel.

The aim described above was also one of the aims of [4], and of Bollobás, Borgs, Chayes and Riordan [3]. We shall prove a common generalization of the corresponding results from these papers by weakening the assumptions: we shall work with convergence in the cut metric (defined below) as in [3], while allowing unbounded matrices and kernels, as in [4]. It turns out that these very weak, natural assumptions suffice to allow us to relate the giant component of the random graph to the kernel.

To state our results we shall need a few definitions. By a *kernel* on $[0, 1]$ we simply mean an integrable, symmetric function $\kappa : [0, 1]^2 \to [0, \infty)$. We regard kernels as elements of L^1, so two kernels that are equal almost everywhere are considered to be the same.

Throughout, A_n will denote a symmetric n-by-n matrix with non-negative entries. If $A_n = (a_{ij})$ is such a matrix, then there is a piecewise constant kernel κ_{A_n} naturally associated to A_n: this takes the value a_{ij} on the square $((i-1)/n, i/n] \times ((j-1)/n, j/n]$. We call κ an *n-by-n kernel* if it is of the form κ_{A_n} for some A_n.

There is a (sparse) random graph naturally associated to A_n, namely the graph $G(A_n) = G_{1/n}(n, A_n)$. This graph has vertex set $[n] = \{1, 2, \dots, n\}$, the events that different edges are present are independent, and the probability that ij is present is $\min\{a_{ij}/n, 1\}$. If some of the a_{ii} are non-zero then $G(A_n)$ may contain loops; this will be irrelevant for us here, since we study only the component structure of $G(A_n)$. Often, it is convenient to consider minor variants of these definitions: in the *Poisson multi-graph* variant, $G^{\mathrm{m}}_{\mathrm{Po}}(A_n)$, the number of copies of each possible edge ij is Poisson with mean a_{ij}/n. In the *Poisson simple graph* variant, $G_{\mathrm{Po}}(A_n)$, the probability that ij is present is $1 - \exp(-a_{ij}/n)$; in both cases the numbers of copies of different edges are independent. Thus $G_{\mathrm{Po}}(A_n)$ is the simple graph underlying $G^{\mathrm{m}}_{\mathrm{Po}}(A_n)$. Most of the time it makes no difference which variant we consider. Indeed, whenever $a_{ij} < n/2$, say, for all i and j, then

$$G(A_n) =_{\mathrm{d}} G_{\mathrm{Po}}(A'_n) \tag{1}$$

where $=_{\mathrm{d}}$ denotes equality in distribution, and A'_n is the matrix with entries

$$a'_{ij} = -n \log(1 - a_{ij}/n) = a_{ij} + O(a_{ij}^2/n). \tag{2}$$

In the typical case considered here, the entries a_{ij} are small compared to n, so switching between $G(\cdot)$ and $G_{\mathrm{Po}}(\cdot)$ thus corresponds to a minor change in the edge probability parameters. Moreover, under the rather weak assumptions $\max_{ij} a_{ij} < n/2$ and $\sum_{i,j=1}^n a_{ij}^3 = o(n^3)$, the random graphs $G(A_n)$ and

$G_{\mathrm{Po}}(A_n)$ are asymptotically equivalent in the strong sense that they can be coupled so that they are equal whp; see [17, Corollary 2.13].

Having described the limit object (a kernel), and the random graph, it remains to describe the notion of convergence. In doing so it is convenient to consider somewhat more general kernels.

Let $(\mathcal{S}, \mu)$ be a probability space; most of the time we shall take $\mathcal{S}$ to be $[0, 1]$ (or $(0, 1]$) with μ Lebesgue measure. A *kernel* on $\mathcal{S}$ is an integrable, symmetric function $\kappa : \mathcal{S}^2 \to [0, \infty)$. Following Frieze and Kannan [15], for $W \in L^1(\mathcal{S}^2)$ we define the *cut norm* $\|W\|_\square$ of W by

$$\|W\|_{\square,1} := \sup_{S,T} \Bigl| \int_{S\times T} W(x, y)\, d\mu(x)\, d\mu(y) \Bigr|, \tag{3}$$

where the supremum is taken over all pairs of measurable subsets of $\mathcal{S}$. Alternatively, one can take

$$\|W\|_{\square,2} := \sup_{\|f\|_\infty, \|g\|_\infty \le 1} \Bigl| \int_{\mathcal{S}^2} f(x)W(x, y)g(y)\, d\mu(x)\, d\mu(y) \Bigr|. \tag{4}$$

In taking the supremum in (4) one can restrict to functions f and g taking only the values ± 1; it follows that

$$\|W\|_{\square,1} \le \|W\|_{\square,2} \le 4\|W\|_{\square,1}.$$

Thus the two norms $\|\cdot\|_{\square,1}$ and $\|\cdot\|_{\square,2}$ are equivalent, and it will almost never matter which one we use. We shall write $\|\cdot\|_\square$ for either norm, commenting in the few cases where the choice matters. (There are further, equivalent versions of the cut-norm; see Borgs, Chayes, Lovász, Sós and Vesztergombi [9].)

Note that for either definition of the cut norm we have

$$\Bigl| \int W \Bigr| \le \|W\|_\square \le \|W\|_{L^1}.$$

The definition (4) is natural for a functional analyst: this norm is the dual of the projective tensor product norm in $L^\infty \hat{\otimes} L^\infty$, and is thus the injective tensor product norm in $L^1 \check{\otimes} L^1$; equivalently, it is equal to the operator norm of the corresponding integral operator $L^\infty \to L^1$. One advantage of this version is the simple "Banach module" property we shall note later in (23). On the other hand, (3) is probably more familiar in combinatorics, and (surprisingly) occasionally has a tiny advantage; see Section 3.

Given a kernel κ and a measure-preserving bijection $\tau : \mathcal{S} \to \mathcal{S}$, let $\kappa^{(\tau)}$ be the kernel defined by

$$\kappa^{(\tau)}(x, y) = \kappa(\tau(x), \tau(y));$$

we call $\kappa^{(\tau)}$ a *rearrangement* of κ. We write $\kappa \sim \kappa'$ if κ' is a rearrangement of κ. Given two kernels κ, κ' on $[0, 1]$, the *cut metric* of Borgs, Chayes, Lovász, Sós and Vesztergombi [9] is defined by

$$\delta_\square(\kappa, \kappa') = \inf_{\kappa'' \sim \kappa'} \|\kappa - \kappa''\|_\square. \tag{5}$$

If we wish to specify which version of the cut norm is involved, we write $\delta_{\square,1}$ or $\delta_{\square,2}$. Usually, this is irrelevant.

As in [9], one can also define $\delta_\square$ using couplings between different kernels, rather than rearrangements. In this case it is irrelevant that the kernels are on the same probability space. In particular, we may regard a matrix A_n as a kernel on the discrete space with n equiprobable elements. Then (by an obvious coupling) $\delta_\square(A_n, \kappa_{A_n}) = 0$, where κ_{A_n} is the n-by-n kernel on $[0,1]$ corresponding to A_n. Thus $\delta_\square(A_n, \kappa) = \delta_\square(\kappa_{A_n}, \kappa)$ for any kernel κ on any probability space $(\mathcal{S}, \mu)$. In the light of this we shall often identify a matrix with the corresponding kernel on $[0,1]$.

Throughout this paper, we shall consider sequences (A_n) of matrices such that for some kernel κ we have $\delta_\square(A_n, \kappa) \to 0$. It follows from the results of [16] that for any kernel κ on a probability space $(\mathcal{S}, \mu)$, there exists a kernel κ' on $[0,1]$ with $\delta_\square(\kappa, \kappa') = 0$. Hence we lose no generality by taking $(\mathcal{S}, \mu)$ to be the *standard ground space* in which $\mathcal{S} = [0,1]$ (or $(0,1]$) and μ is Lebesgue measure. In this case it is natural to identify A_n with κ_{A_n} as above, and we may use the more down-to-earth formula (5) as the definition of $\delta_\square$.

To state our results we need two further definitions, from [4]. Given a kernel κ on a probability space $(\mathcal{S}, \mu)$, let $\mathfrak{X}_\kappa$ be the multi-type Galton–Watson branching process defined as follows. We start with a single particle in generation 0, whose type has the distribution μ. A particle in generation t of type x gives rise to children in generation $t+1$ whose types form a Poisson process on $\mathcal{S}$ with intensity $\kappa(x,y)\,d\mu(y)$. The children of different particles are independent, and independent of the history.

We shall also consider the branching processes $\mathfrak{X}_\kappa(x)$, $x \in \mathcal{S}$, defined as above except that $\mathfrak{X}_\kappa(x)$ starts with a single particle of the given type x.

Let $\rho(\kappa)$ denote the *survival probability* of $\mathfrak{X}_\kappa$, i.e., the probability that all generations are non-empty. It is easily seen that this is the same as the probability that the total number $|\mathfrak{X}_\kappa|$ of particles in $\mathfrak{X}_\kappa$ is infinite. For basic results about $\rho(\kappa)$, we refer the reader to [4].

Finally, as in [4], a kernel κ is *reducible* if there exists $A \subset \mathcal{S}$ with $0 < \mu(A) < 1$ such that κ is zero almost everywhere on $A \times (\mathcal{S} \setminus A)$. Otherwise, κ is *irreducible*.

Throughout, we use standard notation for probabilistic asymptotics as in [18]. For example, $\xrightarrow{\mathrm{p}}$ denotes convergence in probability, and $X_n = o_{\mathrm{p}}(f(n))$ means $X_n/f(n) \xrightarrow{\mathrm{p}} 0$.

1.1 Main results

In this subsection we state our main results; we shall give corresponding results for hypergraphs in Section 3. Recall that any matrix denoted by A_n is assumed to be a symmetric n-by-n matrix with non-negative entries. Given a graph G and an $i \geq 1$, we write $C_i(G)$ for the number of vertices in the ith largest component of G, with $C_i(G) = 0$ if G has fewer then i components. We shall see later that our results imply corresponding results for the Poisson variants of

$G(A_n)$; for simplicity we state them only in the original formulation, where the edge probabilities are $\min\{a_{ij}/n, 1\}$. The theorems are valid for a kernel κ on any probability space $(\mathcal{S}, \mu)$, but as noted above we may assume without loss of generality that $\mathcal{S} = [0, 1]$, and we shall do so in the proofs for convenience.

Theorem 1.1. *Let κ be a kernel and (A_n) a sequence of symmetric non-negative n-by-n matrices such that $\delta_\square(A_n, \kappa) \to 0$. Then $C_1(G(A_n))/n \le \rho(\kappa) + o_\mathrm{p}(1)$. If κ is irreducible, then $C_1(G(A_n))/n \xrightarrow{\mathrm{p}} \rho(\kappa)$ and $C_2(G(A_n)) = o_\mathrm{p}(n)$.*

Of course, as usual we do not require A_n to be defined for every n, only for a subsequence.

Let $\rho_\kappa(x)$ denote the survival probability of the process $\mathfrak{X}_\kappa(x)$ started with a particle of type x. Let T_κ be the integral operator on $\mathcal{S}$ with kernel κ, defined by

$$(T_\kappa f)(x) = \int_{\mathcal{S}} \kappa(x, y) f(y)\, d\mu(y), \tag{6}$$

for any (measurable) function f such that this integral is defined (finite or $+\infty$) for a.e. x. Note that this class of functions includes every (measurable) function $f \ge 0$. Also, let

$$\|T_\kappa\| = \sup\bigl\{\|T_\kappa f\|_2 : \|f\|_2 \le 1,\ f \ge 0\bigr\} \le \infty;$$

clearly if $\|T_\kappa\| < \infty$, then $\|T_\kappa\|$ is simply the norm of T_κ as an operator on $L^2(\mathcal{S}, \mu)$.

Recall from [4, Theorem 6.2] that $\rho(\kappa) > 0$ if and only if $\|T_\kappa\| > 1$, and that if $\|T_\kappa\| > 1$, then ρ_κ is the unique non-zero solution $f \ge 0$ to the functional equation

$$f = 1 - \exp(-T_\kappa f).$$

Using Theorem 1.1, we shall deduce the following slight extension, describing the 'critical' value of c above which a giant component appears in $G(cA_n)$.

Theorem 1.2. *Let κ be a kernel, (A_n) a sequence of symmetric non-negative n-by-n matrices such that $\delta_\square(A_n, \kappa) \to 0$, and $c > 0$ a constant, and set $G_n = G(cA_n)$.*

(a) *If $c \le \|T_\kappa\|^{-1}$, then $C_1(G_n) = o_\mathrm{p}(n)$.*

(b) *If $c > \|T_\kappa\|^{-1}$, then $C_1(G_n) = \Theta(n)$ whp. Furthermore, if κ is bounded, then for any constant $\alpha < (c\|T_\kappa\| - 1)/(c \sup \kappa)$ we have $C_1(G_n) \ge \alpha n$ whp.*

(c) *If κ is irreducible, then $C_1(G_n)/n \xrightarrow{\mathrm{p}} \rho(c\kappa)$ and $C_2(G_n) = o_\mathrm{p}(n)$.*

This clearly generalizes the main result, Theorem 1, of Bollobás, Borgs, Chayes and Riordan [3], which is simply the special case in which κ and the entries of the matrices A_n are uniformly bounded. As we shall see in the next subsection, Theorem 1.2 also generalizes Theorem 3.1 of [4]. Note, however, that to prove this requires various results from [4].

Returning to the irreducible case, we shall also prove a 'stability' result analogous to Theorem 3.9 of [4].

Theorem 1.3. *Let κ be an irreducible kernel and (A_n) a sequence of non-negative symmetric n-by-n matrices such that $\delta_\square(A_n, \kappa) \to 0$. For every $\varepsilon > 0$ there is a $\delta = \delta(\kappa, \varepsilon) > 0$ such that, whp,*

$$\rho(\kappa) - \varepsilon \leq C_1(G'_n)/n \leq \rho(\kappa) + \varepsilon$$

for every graph G'_n that may be obtained from $G_n = G(A_n)$ by deleting at most δn vertices and their incident edges, and then adding or deleting at most δn edges.

As we shall show in Subsection 2.6, using this result it is not hard to deduce exponential tail bounds on the size of the giant component.

Theorem 1.4. *Let κ be an irreducible kernel and $\varepsilon > 0$ a real number. There is a $\gamma = \gamma(\kappa, \varepsilon) > 0$ such that whenever (A_n) is sequence of non-negative symmetric n-by-n matrices with $\delta_\square(A_n, \kappa) \to 0$, then setting $G_n = G(A_n)$ we have*

$$\mathbb{P}\big(|C_1(G_n) - \rho(\kappa)n| \geq \varepsilon n\big) \leq e^{-\gamma n}$$

and

$$\mathbb{P}\big(C_2(G_n) \geq \varepsilon n\big) \leq e^{-\gamma n}$$

for all large enough n.

For the very special case of $G(n,p)$, $p = c/n$, much stronger results are known, establishing the correct dependence of γ on ε in the upper and lower bounds. Indeed, such a 'large deviation principle' for $C_1(G(n, c/n))$ was obtained by O'Connell [23], and Biskup, Chayes and Smith [2] proved a corresponding result for the number of vertices in 'large' components. One might ask whether these results can be generalized to $G(A_n)$; this is likely to be rather hard. Indeed, it is not even clear whether they extend to $G(A_n)$ with A_n converging to a constant kernel κ.

Remark 1.5. We have stated all our results for a deterministic sequence A_n with $\delta_\square(A_n, \kappa) \to 0$. In applications, however, the matrices A_n are often random, and G_n is defined by first conditioning on A_n, and then taking the entries as giving the conditional probabilities of the edges, which are conditionally independent. The conclusions of Theorems 1.1–1.3 are all of the form that $G(A_n)$ has certain properties whp. Having proved such a result assuming $\delta_\square(A_n, \kappa) \to 0$, the corresponding result with A_n random and $\delta_\square(A_n, \kappa) \xrightarrow{\text{p}} 0$ follows immediately. One way of seeing this is to note that a sequence E_n of events holds whp if and only if every subsequence has a subsubsequence holding whp. If $\delta_\square(A_n, \kappa) \xrightarrow{\text{p}} 0$, then given a subsequence (with deterministic indices) of the random sequence (A_n), one can find a subsubsequence such that $\delta_\square(A_n, \kappa) \to 0$ holds a.s., condition on the matrices in this subsubsequence, and apply the result for the deterministic case.

The rest of the paper is organized as follows. In the next few subsections we discuss various applications and consequences of the results above. In Section 2 we prove Theorems 1.1–1.4: as the proofs are somewhat lengthy we shall break this section into subsections. Finally, in Section 3 we present extensions of our main results to the *hyperkernels* and corresponding random (hyper)graphs considered in [5].

1.2 Relationship to the sparse inhomogeneous model

In this subsection we shall prove a simple lemma which, together with Theorem 1.2, implies Theorem 3.1 of [4]. This latter result states that (essentially) the conclusions of Theorems 1.1 and 1.2 (with $c = 1$) hold when the random graph G_n is an instance of the general sparse inhomogeneous model $G^{\mathcal{V}}(n, \kappa_n)$ of [4]. Since the full definitions of [4] are rather cumbersome, for this subsection only we assume a certain familiarity with the terminology of [4].

We say that a kernel κ on $(\mathcal{S}, \mu)$ is *of finite type* if there is a finite partition $(S_1, \ldots, S_r)$ of $\mathcal{S}$ into measurable sets such that κ is constant on each of the sets $S_i \times S_j$. A key strategy we used in [4] was to reduce results about the general case to the finite-type case; we shall use the same approach in this subsection. In the rest of this paper we follow a different strategy, using cut convergence to directly prove results about the general case.

The sparse inhomogeneous model $G^{\mathcal{V}}(n, \kappa_n)$ is defined in terms of a *ground space* $\mathcal{V} = (\mathcal{S}, \mu, (\mathbf{x}_n))$, and a sequence (κ_n) of kernels on $(\mathcal{S}, \mu)$. Here $(\mathcal{S}, \mu)$ is a probability space (satisfying some additional assumptions) and each $\mathbf{x}_n$ is a (deterministic or) random sequence of n points of $\mathcal{S}$, satisfying certain technical assumptions. The sequence (κ_n) is assumed to converge to a kernel κ in a certain sense, and must also satisfy a certain 'graphicality' assumption that involves the sequences $\mathbf{x}_n$. For the full technical details, which will not be relevant here, see [4].

As noted in [4, Remark 8.8], in proving results about this model one may always assume that the vertex types are deterministic. In this case $G^{\mathcal{V}}(n, \kappa_n)$ has the distribution of $G(A_n)$, where A_n is the matrix obtained by sampling the kernel according to the vertex types: A_n has entries $a_{ij} = a_{ij}^{(n)}$ given by $a_{ij} = \kappa_n(x_i^{(n)}, x_j^{(n)}) \wedge n$ for $i \neq j$ and $a_{ii} = 0$, where $x \wedge y = \min\{x, y\}$. We refer the reader to [4] for the formal definition of $G^{\mathcal{V}}(n, \kappa_n)$, and in particular for the precise definitions of a (generalized) vertex space and a graphical (sequence of) kernel(s).

The next lemma shows that the matrices A_n associated to $G^{\mathcal{V}}(n, \kappa_n)$ do converge in probability to the limit kernel κ in the cut metric. Although our main interest is in the cut distance, we in fact obtain a result for the L^1 norm, modulo rearrangements. Given two kernels κ, κ' on the standard ground space, let

$$\delta_1(\kappa, \kappa') = \inf_{\kappa'' \sim \kappa'} \|\kappa - \kappa''\|_{L^1}, \tag{7}$$

in analogy with (5). More generally, for two kernels on arbitrary (not necessarily equal) probability spaces, we may define $\delta_1(\kappa, \kappa')$ as a certain infimum over couplings of these probability spaces; we omit the details.

Lemma 1.6. *Let $\mathcal{V} = (\mathcal{S}, \mu, (\mathbf{x}_n))$ be a vertex space, and let (κ_n) be a sequence of kernels that is graphical on $\mathcal{V}$ with limit κ. Let A_n be the matrix with entries $a_{ij} = \kappa_n(x_i^{(n)}, x_j^{(n)}) \wedge n$ for $i \neq j$ and $a_{ii} = 0$. Then $\delta_1(\kappa_{A_n}, \kappa) \xrightarrow{\mathrm{p}} 0$ and $\delta_\square(A_n, \kappa) = \delta_\square(\kappa_{A_n}, \kappa) \xrightarrow{\mathrm{p}} 0$.*

Proof. Since $\|\kappa'\|_\square \leq \|\kappa'\|_{L^1}$ for any κ', we have $\delta_\square(\kappa_1, \kappa_2) \leq \delta_1(\kappa_1, \kappa_2)$ for any two kernels, so it suffices to prove the first statement.

Conditioning on the vertex types, we may and shall assume that the vertex types are deterministic. For convenience we assume that $\mathcal{S}$ is the standard ground space $[0, 1]$. (The general case requires couplings of κ and A_n, but is otherwise the same.)

Suppose first that κ is regular finitary; roughly speaking, this means that κ is of finite type. (More precisely, κ must be of finite type and must satisfy an additional technical condition; see [4].) Suppose also that $\kappa_n = \kappa$ for every n. In this case the result is essentially trivial: we may assume that there is a partition of $\mathcal{S}$ into sets $S_1, \ldots, S_k$ such that κ is constant on each set $S_r \times S_s$. The definition of a vertex space ensures that for each r there are $\mu(S_r)n + o(n)$ vertices i such that $x_i \in S_r$. Rearranging (or coupling) appropriately, we may assume that each S_r is an interval $I_r \subseteq \mathcal{S} = [0, 1]$. We may then order the vertices so that for all but $o(n)$ vertices i the interval $(i - 1/n, i/n]$ lies entirely inside the interval I_r containing x_i. After doing so, κ and κ_{A_n} differ on a set of measure $o(1)$. Since both are bounded by $\sup \kappa < \infty$, it follows that $\kappa_{A_n} \to \kappa$ in L^1 and hence in $\delta_\square$.

To treat the general case, we approximate by finite-type kernels, as so often in [4]. Indeed, by Lemma 7.3 of [4] there is a sequence of regular finitary kernels κ_m^- such that $\kappa_m^- \leq \kappa_n$ for all $n \geq m$ and $\kappa_m^-(x, y) \nearrow \kappa(x, y)$ for a.e. $(x, y) \in \mathcal{S}^2$. By monotone convergence, we have $\int \kappa_m^- \to \int \kappa$ as $m \to \infty$. Fix $\varepsilon > 0$. Then there is some m such that $\kappa^- = \kappa_m^-$ satisfies $\kappa^- \leq \kappa$ and $\int(\kappa - \kappa^-) \leq \varepsilon$.

Let A_n^- be the matrix with entries $a_{ij}^- = \kappa^-(x_i^{(n)}, x_j^{(n)}) \wedge n$, $i \neq j$, and $a_{ii}^- = 0$. Considering from now on only $n \geq m$, we then have $a_{ij}^- \leq a_{ij}$ and thus $\kappa_{A_n^-} \leq \kappa_{A_n}$ pointwise. After conditioning on the vertex types, the expected number of edges in $G^{\mathcal{V}}(n, \kappa_n)$ is exactly

$$\frac{1}{2} \sum_i \sum_{j \neq i} \frac{a_{ij}}{n} = \frac{1}{2} \sum_i \sum_j \frac{a_{ij}}{n} = \frac{n}{2} \int \kappa_{A_n},$$

using $a_{ii} = 0$ for the first equality. Thus, by Lemma 8.7 of [4], $\int \kappa_{A_n} \to \int \kappa$. Similarly (since a finite-type kernel is always graphical), $\int \kappa_{A_n^-} \to \int \kappa^-$. Hence,

$$\|\kappa_{A_n} - \kappa_{A_n^-}\|_{L^1} = \int (\kappa_{A_n} - \kappa_{A_n^-}) \to \int (\kappa - \kappa^-) \leq \varepsilon.$$

By the finite-type case above, we have $\delta_1(\kappa_{A_n^-}, \kappa^-) \to 0$. Since $\|\kappa - \kappa^-\|_{L^1} \leq \varepsilon$ it follows that $\limsup \delta_1(\kappa_{A_n}, \kappa) \leq 2\varepsilon$. Recalling that $\varepsilon > 0$ was arbitrary, the result follows. $\square$

Recall that Theorem 3.1 of [4] states (essentially) that the random graphs $G_n = G^{\mathcal{V}}(n, \kappa_n)$ satisfy the conclusions of Theorems 1.1 and 1.2. Using Lemma 1.6, by Remark 1.5 the vertex space case of this result follows immediately from Theorems 1.1 and 1.2. As noted in [4, Section 8.1], the apparent extra generality of generalized vertex spaces makes no essential difference, so Theorem 3.1 of [4] then follows. In other words, we have shown that Theorem 3.1 of [4] may be deduced from our present Theorems 1.1 and 1.2, using various results from [4] mentioned above. Let us remark that in practice, the conditions of Theorem 3.1 of [4] will often be easier to verify than those of Theorems 1.1 and 1.2.

1.3 Further applications

As noted in [5], the definitions in [4] exclude one simple case to which the results clearly extend, namely the case of an arbitrary integrable kernel κ, and i.i.d. vertex types: given a kernel κ, one may define the random graph $G(n,\kappa) = G_{1/n}(n,\kappa)$ on $[n]$ by taking $x_1,\ldots,x_n$ to be independent and uniformly distributed on $[0,1]$, and given these 'vertex types', joining each pair $\{i,j\}$ of vertices with probability $\min\{\kappa(x_i,x_j)/n, 1\}$, independently of all other pairs. With κ bounded, a corresponding dense random graph was studied by Lovász and Szegedy [19].

Our next lemma shows that Theorems 1.1–1.3 apply (unsurprisingly) to the graphs $G(n,\kappa)$, since the (random) matrices of edge probabilities associated to $G(n,\kappa)$ converge to κ in probability in $\delta_\square$.

Lemma 1.7. *Let κ be a kernel. For $n \ge 1$ let $x_1,\ldots,x_n$ be i.i.d. uniform points from $\mathcal{S}$, and let A_n be the n-by-n matrix with entries $a_{ij} = \kappa(x_i,x_j)$ for $i \ne j$, and $a_{ii} = 0$. Then $\delta_1(A_n,\kappa) \xrightarrow{\mathrm{p}} 0$ and $\delta_\square(A_n,\kappa) \xrightarrow{\mathrm{p}} 0$.*

Proof. As before, we have $\delta_\square \le \delta_1$, so it suffices to prove the first statement. Fix $\varepsilon > 0$. By standard results there is a finite-type kernel κ' such that $\|\kappa - \kappa'\|_{L^1} \le \varepsilon^2$. Indeed, this follows by the construction of the product measure, since the rectangular sets $A \times B$ generate an algebra $\mathcal{F}_0$ that generates the product σ-field, and it is easily seen that finite linear combinations of indicator functions of sets in $\mathcal{F}_0$ are dense in $L^1(\mathcal{S}^2)$.

Let A_n' be the matrix with entries $a_{ij}' = \kappa'(x_i,x_j)$, $i \ne j$, and $a_{ii}' = 0$. Then

$$\mathbb{E}\|\kappa_{A_n} - \kappa_{A_n'}\|_{L^1} = \frac{n(n-1)}{n^2}\|\kappa - \kappa'\|_{L^1} \le \varepsilon^2,$$

so with probability at least $1-\varepsilon$ we have

$$\delta_1(A_n, A_n') = \delta_1(\kappa_{A_n}, \kappa_{A_n'}) \le \|\kappa_{A_n} - \kappa_{A_n'}\|_{L^1} \le \varepsilon. \tag{8}$$

Since κ' is of finite type, it is essentially trivial that $\delta_1(A_n', \kappa') \xrightarrow{\mathrm{p}} 0$ as $n \to \infty$; the argument is similar to one in the previous subsection, so we omit the details. Since $\delta_1(\kappa,\kappa') \le \|\kappa - \kappa'\|_{L^1} \le \varepsilon^2$,

$$\delta_1(A_n,\kappa) \le \delta_1(A_n,A_n') + \delta_1(A_n',\kappa') + \delta_1(\kappa',\kappa),$$

and $\varepsilon > 0$ was arbitrary, it follows that $\delta_1(A_n,\kappa) \xrightarrow{\mathrm{p}} 0$, as claimed. $\square$

So far we have shown that the results in Subsection 1.1 imply many existing results about the giant component in various sparse random graphs. We now turn to a new application, giving an example that we believe is not covered by known results.

Let $p = p(n)$ be some normalizing function, with $0 < p \leq 1$ and $p(n) \to 0$. Let G_n be a sequence of graphs in which G_n has n vertices and $\Theta(pn^2)$ edges, and let κ be a kernel. Following the terminology of [6, 7], we say that $\delta_\Box(G_n, \kappa) \to 0$ if $\delta_\Box(A_n, \kappa) \to 0$, where A_n is $1/p$ times the adjacency matrix of G_n. A sequence (G_n) satisfying this condition may be thought of as a sequence of inhomogeneous sparse quasi-random graphs. For graphs which are dense and homogeneous, there are many equivalent definitions of quasi-randomness, or pseudo-randomness; see Thomason [25, 26] or Chung, Graham and Wilson [12], for example. In the sparse case these notions are no longer equivalent, as discussed by Chung and Graham [11] in the homogeneous case, and Bollobás and Riordan [6] in general; when κ is constant, normalizing so that $\kappa = 1$, we have $\delta_\Box(G_n, \kappa) \to 0$ if and only if

$$\sup_{V \subset V(G_n)} \left| e(G_n[V]) - p|V|^2/2 \right| = o(pn^2); \tag{9}$$

this condition is called DISC in [11]. Other, stronger conditions have also been considered, in particular by Thomason [25, 26]. Our next result establishes the threshold for percolation on an arbitrary sequence of inhomogeneous sparse quasi-random graphs.

Theorem 1.8. *Let $c > 0$ be a constant, let $p = p(n)$ be any function with $c/n \leq p(n) \leq 1$, let κ be an irreducible kernel on $[0,1]^2$, and let (G_n) be a sequence of graphs with $|G_n| = n$ and $\delta_\Box(G_n, \kappa) \to 0$. Writing G'_n for the random subgraph of G_n obtained by selecting each edge independently with probability $c/(pn)$, we have $C_1(G'_n)/n \xrightarrow{\mathrm{p}} \rho(c\kappa)$. In particular, the threshold value of c above which a giant component appears in G'_n is given by $1/\|T_\kappa\|$.*

Proof. As above, let A_n be $1/p$ times the adjacency matrix of G_n. Then, by assumption, $\delta_\Box(A_n, \kappa) \to 0$, so $\delta_\Box(cA_n, c\kappa) \to 0$. The random subgraph G'_n is exactly $G(cA_n)$, so the result follows from Theorem 1.1. $\Box$

As noted in [6], one way to construct inhomogeneous sparse quasi-random graphs is to consider appropriate *random* graphs, but this is not so interesting in the present context: the random subgraphs of such graphs end up being the graphs $G(n, \kappa)$ considered at the start of the subsection. A more interesting application of Theorem 1.8 is to deterministic quasi-random graphs. In the homogeneous case, where $\kappa = 1$ is constant, many such sequences are known. One example is given by the 'polarity graphs' of Erdős and Rényi [14], defined (for suitable n) by taking as vertices the points of the projective plane over $GF(q)$, q a prime power, and joining $x = (x_0, x_1, x_2)$ and $y = (y_0, y_1, y_2)$ if and only if $x_0y_0 + x_1y_1 + x_2y_2 = 0$ in $GF(q)$. Here $n = q^2 + q + 1$ and $p = (q+1)/n = \Theta(n^{-1/2})$. Other examples are the coset graphs of Chung [10] and the Ramanujan graphs of Lubotzky, Phillips and Sarnak [20]. In all these examples the limiting kernel is constant, so Theorem 1.8 says that on any of these graphs, the threshold for percolation is when the average degree of the random subgraph is equal to 1.

Note that in the examples above, the matrices (A_n) to which Theorem 1.1 or Theorem 1.2 is applied are very far from satisfying the uniform boundedness condition assumed in Bollobás, Borgs, Chayes and Riordan [3]. Indeed, each A_n has all entries either 0 or $1/p$, where $p = p(n) \to 0$. This also implies that the corresponding kernels κ_{A_n}, which do converge to $\kappa = 1$ in the cut norm, do not converge in various natural stronger senses, such as pointwise or in L^1.

In general, it is very hard to compute the cut distance between two kernels. Indeed, if A_1 and A_2 are the adjacency matrices of two graphs, then the general problem of computing $\delta_\square(\kappa_{A_1}, \kappa_{A_2})$ includes as a special case deciding whether G_1 and G_2 are isomorphic. Thus applications of Theorems 1.1 and 1.2 are likely to involve special cases where cut convergence is guaranteed for some simple reason, such as the example in the previous subsection.

1.4 Consequences for branching processes

Theorem 1.1 has an interesting consequence purely concerning branching processes. Recall that if κ is a kernel, then $\rho(\kappa)$ denotes the survival probability of the multi-type Poisson Galton–Watson process $\mathfrak{X}_\kappa$.

Theorem 1.9. *Let κ_m, $m \geq 1$, and κ be kernels with $\delta_\square(\kappa_m, \kappa) \to 0$ as $m \to \infty$. Then $\rho(\kappa_m) \to \rho(\kappa)$.*

Proof. Let us first note that the result is not really a statement about the cut metric $\delta_\square$, but rather about the cut norm $\|\cdot\|_\square$. Indeed, by definition of $\delta_\square$ there are rearrangements κ'_m of κ_m with $\|\kappa'_m - \kappa\|_\square \leq \delta_\square(\kappa_m, \kappa) + 1/m$, say, and hence $\|\kappa'_m - \kappa\|_\square \to 0$. Since $\rho(\kappa'_m) = \rho(\kappa_m)$, in proving the result we may assume if we like that $\|\kappa_m - \kappa\|_\square \to 0$.

We shall prove the result in three steps.

Step 1: suppose that all κ_m are irreducible; this case is the heart of the proof. For each m we may find a sequence $A_n^{(m)}$ of symmetric n-by-n matrices with $\delta_\square(A_n^{(m)}, \kappa_m) \to 0$ as $n \to \infty$. Indeed, this is an immediate consequence of Lemma 1.7. By Theorem 1.1, if n is large enough, then

$$\mathbb{P}\Big(\big|C_1(G(A_n^{(m)}))/n - \rho(\kappa_m)\big| \geq 1/m\Big) \leq 1/m^2, \tag{10}$$

say. Pick $n(m)$ such that (10) holds and $\delta_\square(A_{n(m)}^{(m)}, \kappa_m) \leq 1/m$, and let $A_m = A_{n(m)}^{(m)}$. By (10), with probability 1 we have

$$\left|\frac{C_1(G(A_m))}{|G(A_m)|} - \rho(\kappa_m)\right| \to 0. \tag{11}$$

Now $\delta_\square(A_m, \kappa_m) \leq 1/m$ by our choice of $n(m)$, while $\delta_\square(\kappa_m, \kappa) \to 0$, so $\delta_\square(A_m, \kappa) \to 0$. Applying Theorem 1.1 again, we have $C_1(G(A_m))/|G(A_m)| \leq$

$\rho(\kappa) + o_{\mathrm{p}}(1)$. Together with (11) this implies that

$$\limsup \rho(\kappa_m) \le \rho(\kappa). \tag{12}$$

If κ is irreducible, then we have $C_1(G(A_m))/|G(A_m)| \xrightarrow{\mathrm{p}} \rho(\kappa)$, so $\rho(\kappa_m) \to \rho(\kappa)$, as required. We shall return to the lower bound in the case that κ is reducible later.

Step 2: we now consider the general case, where some of κ and the κ_m may be reducible. By Theorem 6.4(i) of [4], given a kernel κ' and a sequence κ'_n tending pointwise down to κ', we have $\rho(\kappa'_n) \to \rho(\kappa')$. Applying this with $\kappa' = \kappa_m$ and $\kappa'_n = \kappa_m + 1/n$, say, we see that for each m there is an $\varepsilon_m < 1/m$ such that $|\rho(\kappa'_m) - \rho(\kappa_m)| \le 1/m$, where $\kappa'_m = \kappa_m + \varepsilon_m$. Now κ'_m is irreducible, and $\|\kappa'_m - \kappa_m\|_\square \le 1/m \to 0$, so $\delta_\square(\kappa'_m, \kappa) \to 0$, and the results of Step 1 apply. In particular, the upper bound (12) holds, and if κ is irreducible, then $\rho(\kappa_m) \to \rho(\kappa)$, as required.

Step 3: in the case where κ is reducible, it remains to prove the lower bound corresponding to (12). For this we decompose κ into irreducible kernels as in [4]. As shown there (in Lemma 5.17), given any κ there is a finite or countable partition $(S_i)_{i=0}^N$, $N \le \infty$, of $\mathcal{S}$ into measurable sets such that $\kappa = \sum_{i\ge1} \kappa^{(i)}$ holds a.e., where each $\kappa^{(i)}$ is zero off $\mathcal{S}_i \times \mathcal{S}_i$ and irreducible when restricted to $\mathcal{S}_i \times \mathcal{S}_i$. Fix $\varepsilon > 0$. Since $\rho(\kappa) = \sum \rho(\kappa^{(i)})$, there is some $k < \infty$ such that $\sum_{i=1}^k \rho(\kappa^{(i)}) \ge \rho(\kappa) - \varepsilon$. Define $\kappa_m^{(i)}$ to be the kernel that is equal to κ_m on $\mathcal{S}_i \times \mathcal{S}_i$ and zero off this set, and let $\kappa'_m = \sum_{i=1}^k \kappa_m^{(i)}$. Then $\kappa_m \ge \kappa'_m$, so $\rho(\kappa_m) \ge \rho(\kappa'_m) = \sum_{i=1}^k \rho(\kappa_m^{(i)})$. Since $\|\kappa_m - \kappa\|_\square \ge \|\kappa_m^{(i)} - \kappa^{(i)}\|_\square$ for each i, we have $\|\kappa_m^{(i)} - \kappa^{(i)}\|_\square \to 0$ for each i. Since $\kappa^{(i)}$ is irreducible, by the result of Step 2 we have $\rho(\kappa_m^{(i)}) \to \rho(\kappa^{(i)})$. Summing over i from 1 to k it follows that

$$\liminf_{m\to\infty} \rho(\kappa_m) \ge \sum_{i=1}^k \rho(\kappa^{(i)}) \ge \rho(\kappa) - \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary we thus have $\liminf_{m\to\infty} \rho(\kappa_m) \ge \rho(\kappa)$. Together with (12), this completes the proof. $\square$

Note that Theorem 1.9 is a purely analytic statement about branching processes and the cut metric (or cut norm – rearrangements change nothing here). However, the only proof we know is that above, which goes via graphs! Corresponding results with much stronger assumptions (monotone convergence, either upwards or downwards) were proved in [4]; these weaker results were all that was needed there.

We close this section by giving a direct proof of a weaker form of Theorem 1.9, assuming L^1 convergence. As above, rearrangement is irrelevant, so it makes no difference whether we suppose that $\delta_1(\kappa_n, \kappa) \to 0$ or $\|\kappa_n - \kappa\|_{L^1} \to 0$.

Theorem 1.10. *Let κ_n, $n \ge 1$, and κ be kernels on a probability space $(\mathcal{S}, \mu)$, with $\|\kappa_n - \kappa\|_{L^1} \to 0$ as $n \to \infty$. Then $\rho(\kappa_n) \to \rho(\kappa)$.*

The proof will be based on weak-$*$ convergence. Let f_n, $n \geq 1$, and f be functions in $L^\infty(\mathcal{S}, \mu)$. The definition of the weak-$*$ topology on $L^\infty(\mathcal{S}, \mu)$ is that $f_n \xrightarrow{\text{w}*} f$ if and only if

$$\int g(x) f_n(x)\, d\mu(x) \to \int g(x) f(x)\, d\mu(x) \text{ for every } g \in L^1(\mathcal{S}, \mu). \tag{13}$$

Lemma 1.11. *Suppose that $\kappa \in L^1(\mathcal{S} \times \mathcal{S})$ and $f_n \in L^\infty(\mathcal{S}, \mu)$ with $f_n \xrightarrow{\text{w}*} 0$. Let $h_n = T_\kappa f_n$, so $h_n(x) = \int \kappa(x, y) f_n(y)\, d\mu(y)$. Then $h_n \to 0$ in $L^1(\mathcal{S}, \mu)$.*

Proof. Note first that by the uniform boundedness principle we have $C = \sup \|f_n\|_\infty < \infty$. (In fact, in the application, each f_n is bounded by 1.)

Let $\varepsilon > 0$. As in the proof of Lemma 1.7, there is a finite-type kernel κ' such that $\|\kappa - \kappa'\|_{L^1} < \varepsilon$. We may express κ' as $\kappa'(x, y) = \sum_{i=1}^N \varphi_i(x) \psi_i(y)$ for φ_i, $\psi_i \in L^1$. (In fact, we may take each φ_i or ψ_i to be a constant times a characteristic function.) Now

$$\begin{aligned}
&\|h_n\|_{L^1} = \Big\| \int \kappa(x, y) f_n(y)\, d\mu(y) \Big\|_{L^1} \\
&\leq \int |(\kappa(x, y) - \kappa'(x, y)) f_n(y)|\, d\mu(x)\, d\mu(y) + \sum_{i=1}^N \Big\| \int \varphi_i(x) \psi_i(y) f_n(y)\, d\mu(y) \Big\|_{L^1}.
\end{aligned}$$

The first term above is at most $\|\kappa - \kappa'\|_{L^1} \|f_n\|_\infty \leq \varepsilon C$. The second term is exactly

$$\sum_{i=1}^N \|\varphi_i\|_{L^1} \left| \int \psi_i(y) f_n(y)\, d\mu(y) \right|.$$

Each integral tends to zero by the definition (13) of the weak-$*$ topology, so it follows that $\limsup \|h_n\|_{L^1} \leq \varepsilon C$. Since $\varepsilon > 0$ was arbitrary, the result follows. $\square$

With this preparation behind us, we turn to the proof of Theorem 1.10.

Proof of Theorem 1.10. We may assume without loss of generality that the σ-field $\mathcal{F}$ on $\mathcal{S}$ where μ is defined is countably generated, and thus $L^1(\mathcal{S}, \mu)$ is separable. One way to see this is to note that otherwise we can replace $\mathcal{F}$ by a countably generated sub-σ-field $\mathcal{F}_0$ such that each κ_n is $\mathcal{F}_0 \times \mathcal{F}_0$-measurable; alternatively, by the results of [16] we may assume without loss of generality that $\mathcal{S} = [0, 1]$, with μ Lebesgue measure.

Suppose for simplicity that κ is irreducible; arguing as in the proof of Theorem 1.9, it is not hard to reduce the general case to this case.

Suppose for a contradiction that $\|\kappa_n - \kappa\|_{L^1} \to 0$ but $\rho(\kappa_n) \not\to \rho(\kappa)$. Passing to a subsequence, we may assume that $|\rho(\kappa_n) - \rho(\kappa)|$ is bounded away from zero. To obtain a contradiction it then suffices to show that for some subsequence (κ_{n_i}) of (κ_n) we have $\rho(\kappa_{n_i}) \to \rho(\kappa)$.

Let $\rho_n(x) = \rho_{\kappa_n}(x)$ be the survival probability of the branching process $\mathfrak{X}_{\kappa_n}(x)$, started with a single particle of type x. As shown in [4], the function ρ_n satisfies

$$\rho_n = 1 - \exp(-T_{\kappa_n}\rho_n). \tag{14}$$

It is well known that the unit ball of $L^\infty(\mathcal{S}, \mu)$ is sequentially compact in the weak-$*$ topology when $L^1(\mathcal{S}, \mu)$ is separable. (The unit ball of L^∞ is always compact, but not necessarily sequentially compact otherwise.) For the special case $\mathcal{S} = [0, 1]$, let (f_n) be a sequence in the unit ball of $L^\infty([0, 1])$. This sequence has a subsequence (f_{n_k}) such that $\int_I f_{n_k}$ converges for each of the countably many intervals I with rational endpoints. Since the f_{n_k} are uniformly bounded, this is enough to ensure weak-$*$ convergence.

Since $\|\rho_n\|_\infty \le 1$ for every n, by sequential compactness there is some $\rho^* \in L^\infty(\mathcal{S}, \mu)$ and some subsequence of (κ_n) along which $\rho_n \xrightarrow{\text{w}*} \rho^*$. From now on we restrict our attention to such a subsequence.

Now

$$\|T_{\kappa_n}\rho_n - T_\kappa\rho_n\|_{L^1} \le \|\kappa_n - \kappa\|_{L^1}\|\rho_n\|_\infty \le \|\kappa_n - \kappa\|_{L^1} \to 0.$$

Also, by Lemma 1.11, $\|T_\kappa\rho_n - T_\kappa\rho^*\|_{L^1} \to 0$. Hence $T_{\kappa_n}\rho_n \to T_\kappa\rho^*$ in L^1. Passing to a subsequence, we may assume that $T_{\kappa_n}\rho_n \to T_\kappa\rho^*$ a.e. But then, using (14),

$$\rho_n = 1 - e^{-T_{\kappa_n}\rho_n} \to 1 - e^{-T_\kappa\rho^*} \text{ a.e.}$$

From (13) and dominated convergence, it follows that

$$\rho_n \xrightarrow{\text{w}*} 1 - e^{-T_\kappa\rho^*}.$$

Since $\rho_n \xrightarrow{\text{w}*} \rho^*$, it follows that $\rho^* = 1 - e^{-T_\kappa\rho^*}$ a.e.

Let $\rho(x)$ denote the survival probability of $\mathfrak{X}_\kappa(x)$. Since κ is irreducible, by [4, Theorem 6.2], either $\rho^* = \rho$ a.e. or $\rho^* = 0$ a.e. In the first case,

$$\rho(\kappa_n) = \int \rho_n(x)\,d\mu(x) \to \int \rho^*(x)\,d\mu(x) = \rho(\kappa),$$

as desired. In the second case, we have $\rho(\kappa_n) \to 0$ similarly.

All that remains is to rule out the possibility that $\rho(\kappa_n) \to 0 < \rho(\kappa)$. This is not hard using the results in [4]. For $M > 0$, let κ^M denote the pointwise minimum of κ and M, and define κ_n^M similarly. Suppose that $\rho(\kappa) > 0$. Then $\|T_\kappa\| > 1$. As shown in the proof of [4, Lemma 5.16], we have $\|T_{\kappa^M}\| \nearrow \|T_\kappa\|$ as $M \to \infty$, so there is some M with $c = \|T_{\kappa^M}\| > 1$. Fix such an M. Since

$$\|\kappa_n^M - \kappa^M\|_{L^1} \le \|\kappa_n - \kappa\|_{L^1} \to 0, \tag{15}$$

and the kernels κ_n^M and κ^M are uniformly bounded, we have $\|T_{\kappa_n^M}\| \to \|T_{\kappa^M}\| = c > 1$. In particular, for all large enough n we have $\|T_{\kappa_n^M}\| > (c+1)/2 > 1$. Finally, it follows from [4, Remark 5.14] that we have

$$\rho(\kappa_n^M) \ge \frac{\|T_{\kappa_n^M}\| - 1}{\sup \kappa_n^M} \ge \frac{(c-1)/2}{M} > 0.$$

Since $\rho(\kappa_n) \ge \rho(\kappa_n^M)$ it follows that $\rho(\kappa_n) \not\to 0$, and the proof is complete. $\square$

If we assume cut convergence instead of L^1 convergence, then using the fact that

$$\Big\| \int \kappa(x,y) f(y)\, d\mu(y) \Big\|_{L^1} \le \|\kappa\|_\square \|f\|_\infty$$

in place of the corresponding observation for the L^1 norm, the first part of the proof above goes through unchanged, showing that $\rho^* \to \rho$ a.e. or $\rho^* \to 0$. Unfortunately, we do not know how to exclude the possibility that $\rho(\kappa_n) \to 0 < \rho(\kappa)$, except by appealing to Theorem 1.1, i.e., working with graphs. The problem is that the relation equivalent to (15) for the cut norm rather than the L^1 norm does not hold in general. Of course, given that Theorem 1.9 is true, it is almost guaranteed that it has a direct analytic proof.

As discussed in [6, Section 2], until recently there was another example of an analytic fact about kernels whose only known proof involved graphs (and the cut metric), namely that two bounded kernels may be coupled to agree a.e. if and only if their 'graphical moments' (or subgraph counts) are equal. This follows from the results of Borgs, Chayes, Lovász, Sós and Vesztergombi [9] concerning metrics for graphs (see [6]). However, by now there are analytic proofs: Janson and Diaconis [13] showed that it also follows from results of Hoover and Kallenberg on exchangeable arrays. A direct (and far from simple) proof has recently been given by Borgs, Chayes and Lovász [8].

2 Proofs of Theorems 1.1–1.4

In this section we shall prove our main results; the strategy of the proof of Theorem 1.1 is as follows. First, in Subsection 2.1, we shall show that if each κ_n is an n-by-n kernel and $\delta_\square(\kappa_n, \kappa) \to 0$, then almost all of the weight of κ_n comes from values that are $o(n)$. This will allow us to assume that all edge probabilities in $G(A_n)$ are $o(1)$. It then follows that the expected number of small tree components in $G(A_n)$ is close to what it 'should be', i.e., n times a certain function of the kernel κ_{A_n}. In Subsection 2.2 we show that this function is continuous with respect to the cut metric. This then tells us that we have almost the 'right' number of vertices in small components; the details are given in Subsection 2.3. Finally, in Subsection 2.4 we complete the proof of Theorem 1.1 by showing that in the irreducible case, almost all vertices in large components are in a single component, using a method from Bollobás, Borgs, Chayes and Riordan [3]. In Subsection 2.5 we treat the reducible case, proving Theorem 1.2. Finally, in Subsection 2.6 we prove our stability and concentration results, Theorems 1.3 and 1.4.

For convenience, in this section we assume, as we may, that all kernels are on $[0,1]$, unless explicitly stated otherwise.

2.1 Eliminating large edge weights

In Theorem 2.1 of [7] it was shown that if (G_n) is a sequence of graphs in which G_n has n vertices and $O(n)$ edges, A_n is the adjacency matrix of G_n, κ is a kernel and $\delta_\square(nA_n, \kappa) \to 0$, then $\kappa = 0$ a.e. and $e(G_n) = o(n)$. A simple modification of the proof gives the following lemma. Recall that a matrix denoted A_n is assumed to be n-by-n.

Lemma 2.1. *Suppose that κ is a kernel and (A_n) a sequence of non-negative matrices such that $\delta_\square(A_n, \kappa) \to 0$. Then there is some function $M(n)$ with $M(n) = o(n)$ such that only $o(n)$ entries of A_n exceed $M(n)$, and the sum of these entries is $o(n^2)$.*

A consequence of this is that if A'_n is obtained from A_n by taking the point-wise minimum with $M(n)$, then $\delta_\square(A'_n, \kappa) \to 0$.

Proof. Although the details are almost exactly the same as in [7], we spell them out. We write κ_n for κ_{A_n}.

Since $\delta_\square(\kappa_n, \kappa) \to 0$, we may choose rearrangements $\kappa^{(\tau_n)}$ of κ such that

$$\|\kappa_n - \kappa^{(\tau_n)}\|_\square \to 0. \tag{16}$$

It suffices to show that for any $c > 0$, the sum of the entries of A_n exceeding cn is at most c^2n^2 for n large enough. This implies that there are at most cn such entries, and the result then follows by letting c tend to 0.

Suppose for a contradiction that there is some $c > 0$ such that, for infinitely many n, the sum of the entries of A_n exceeding cn is at least c^2n^2; from now on we fix such a c and restrict our attention to the corresponding values of n. Let G_n be the graph whose edges correspond to those entries of A_n which exceed cn. Let M_n be a largest matching in G_n.

Suppose first that $|V(M_n)|/n \to 0$. Let S_n be the subset of $[0, 1]$ corresponding to the vertex set of M_n, so $\mu(S_n) = |V(M_n)|/n \to 0$. Every edge of weight at least cn meets a vertex of M_n, so

$$\int_{S_n\times[0,1]} \kappa_n = \frac{1}{n^2} \sum_{v\in V(M_n)} \sum_w a_{vw} \geq \frac{1}{2n^2}(cn)^2 = c^2/2,$$

where the factor 2 accounts for the double counting of edges within $V(M_n)$.

From (16), writing S'_n for $\tau_n(S_n)$, we have

$$\int_{S'_n\times[0,1]} \kappa = \int_{S_n\times[0,1]} \kappa^{(\tau_n)} \geq \int_{S_n\times[0,1]} \kappa_n - o(1) \geq c^2/2 - o(1),$$

so $\int_{S'_n\times[0,1]} \kappa \not\to 0$. Since $\mu(S'_n \times [0, 1]) = \mu(S'_n) = \mu(S_n) \to 0$, this contradicts integrability of κ.

Passing to a subsequence, we may thus assume that for some $a > 0$, every maximal matching M_n meets at least an vertices.

Since κ is integrable, we have $\int \kappa 1_{\{\kappa>C\}} \to 0$ as $C \to \infty$, where $1_{\{\kappa>C\}} : [0, 1]^2 \to \{0, 1\}$ is the indicator of the event that $\kappa(x, y) > C$. In particular, there is a $C < \infty$ with $\int \kappa 1_{\{\kappa>C\}} \leq ac/4$. Fix an n with $n > 4C/(ac)$, noting that if $S \subset [0, 1]^2$ satisfies $\mu(S) \leq 1/n$, then

$$\int_S \kappa \leq C\mu(S) + \int \kappa 1_{\{\kappa>C\}} \leq C/n + ac/4 \leq ac/2. \tag{17}$$

Choosing n large enough, we may assume from (16) that there is a $\kappa' = \kappa^{(\tau_n)} \sim \kappa$ with

$$\|\kappa_n - \kappa'\|_\square \le ac/25. \tag{18}$$

Given subsets U and V of $[n]$, let

$$A_n(U,V) = \sum_{u\in U}\sum_{v\in V} a_{uv}.$$

Let $M_n = \{u_1v_1, \ldots, u_rv_r\}$ be a matching in G_n with $r \ge an$, and set $U = \{u_i\}$ and $V = \{v_i\}$. Identifying subsets of $[n]$ with the corresponding unions of intervals of length $1/n$, from (18) we have

$$\left| \int_{U\times V} \kappa' - \frac{A_n(U,V)}{n^2} \right| \le ac/25.$$

Let U' be a random subset of U obtained by selecting each vertex independently with probability $1/2$, and let V' be the complementary subset of V, defined by $V' = \{v_i : u_i \notin U_i\}$. The edges of our matching M_n never appear as edges from U' to V'. On the other hand, any other edge u_iv_j, $i \ne j$, from U to V has probability $1/4$ of appearing. Hence,

$$\mathbb{E}\big(A_n(U',V')\big) = \frac{A_n(U,V)}{4} - \frac{1}{4}\sum_i A_{u_iv_i}.$$

Similarly, writing $S \subset [0,1]^2$ for the union of the r $1/n$-by-$1/n$ squares corresponding to the edges u_iv_i, we have

$$\mathbb{E}\left(\int_{U'\times V'} \kappa'\right) = \frac{1}{4}\int_{U\times V}\kappa' - \frac{1}{4}\int_S \kappa'.$$

Combining the last three displayed equations using the triangle inequality, and noting that $\mu(S) = r/n^2 \le 1/n$, it follows that

$$\begin{aligned}
\left|\mathbb{E}\left(\int_{U'\times V'}\kappa'\right) - \frac{1}{n^2}\mathbb{E}\big(A_n(U',V')\big)\right| &\ge \frac{1}{4n^2}\sum_i A_{u_iv_i} - \frac{1}{4}\int_S \kappa' - ac/100 \\
&\ge \frac{(an)(cn)}{4n^2} - ac/8 - ac/100 > ac/16,
\end{aligned}$$

using (17). On the other hand, from (18),

$$\left|\int_{U'\times V'}\kappa' - \frac{A_n(U',V')}{n^2}\right| \le ac/25$$

always holds, which implies a corresponding upper bound on the difference of the expectations. Since $ac/25 < ac/16$, we obtain a contradiction, completing the proof. ∎

2.2 Tree integrals and the cut metric

In this subsection we shall show that a certain function of a kernel whose role will become clear later is continuous (in fact Lipschitz) with respect to the cut metric. Here there is no particular reason to consider only the standard ground space; instead we consider an arbitrary probability space.

Let $(\mathcal{S}, \mathcal{F}, \mu)$ be a probability space. Let $\mathcal{W}$ be the set of all integrable non-negative functions $W : \mathcal{S} \times \mathcal{S} \to [0, \infty)$, and let $\mathcal{W}_{\mathrm{sym}}$ be the subset of symmetric functions. The integrability assumption is for convenience only; the results extend to arbitrary measurable non-negative functions if one is a little careful with infinities in the proofs. However, we shall only need the integrable case.

For $W \in \mathcal{W}$, let

$$\lambda_W(x) := \int_{\mathcal{S}} W(x, y)\, d\mu(y) \tag{19}$$

and

$$\lambda'_W(y) := \int_{\mathcal{S}} W(x, y)\, d\mu(x) \tag{20}$$

denote the marginals of W; we allow the value $+\infty$, although by our assumption that W is integrable, $\lambda_W(x) < \infty$ a.e. and $\lambda'_W(y) < \infty$ a.e. Note that λ_W and λ'_W are measurable functions from $\mathcal{S}$ to $[0, \infty]$.

Throughout this subsection we work with (4) as the definition of the cut norm: if $W \in L^1(\mathcal{S}^2)$, then

$$\|W\|_\square := \sup_{\|f\|_\infty \le 1,\, \|g\|_\infty \le 1} \left| \int_{\mathcal{S}^2} f(x) g(y) W(x, y)\, d\mu(x)\, d\mu(y) \right|. \tag{21}$$

It is immediate from the definition (21) that

$$\|W\|_\square \le \|W\|_{L^1(\mathcal{S}^2)} \tag{22}$$

and that, for any bounded functions h and k on $\mathcal{S}$,

$$\|h(x)k(y)W(x, y)\|_\square \le \|h\|_\infty \|k\|_\infty \|W\|_\square. \tag{23}$$

Before stating the main result of this subsection, let us note that if two kernels are close in cut norm, then their marginals are close in L^1. (This is doubtless well known, but in any case very easy to see.)

Lemma 2.2. *If $W_1, W_2 \in \mathcal{W}$, then $\|\lambda_{W_1} - \lambda_{W_2}\|_{L^1(\mathcal{S})} \le \|W_1 - W_2\|_\square$.*

Proof. If $f \in L^\infty(\mathcal{S})$, then

$$\int_{\mathcal{S}} \big(\lambda_{W_1}(x) - \lambda_{W_2}(x)\big) f(x)\, d\mu(x) = \int_{\mathcal{S}^2} f(x) \big(W_1(x, y) - W_2(x, y)\big)\, d\mu(x)\, d\mu(y)$$

and the result follows from (21), letting $g(y) = 1$ and taking the supremum over all f with $\|f\|_\infty \le 1$. (Or simply taking $f(x)$ equal to the sign of $\lambda_{W_1}(x) - \lambda_{W_2}(x)$.) $\square$

We now turn to the integrals we shall consider, one for each finite graph F. Given a finite graph F with vertex set $\{1, \ldots, r\}$ and $W \in \mathcal{W}_{\mathrm{sym}}$, let

$$t_{\mathrm{isol}}(F, W) := \int_{\mathcal{S}^r} \prod_{ij \in E(F)} W(x_i, x_j) \prod_{k=1}^{r} e^{-\lambda_W(x_k)} \, d\mu(x_1) \ldots d\mu(x_r). \tag{24}$$

The reason for the notation is that $t_{\mathrm{isol}}(F, W)$ corresponds roughly to $1/n$ times the expected number of isolated copies of F in a certain random graph defined from W.

Our aim in this subsection is to prove the following result.

Theorem 2.3. *Let F be a tree. Then $W \mapsto t_{\mathrm{isol}}(F, W)$ is a bounded map on $\mathcal{W}_{\mathrm{sym}}$ that is Lipschitz continuous in the cut norm. In other words, there exists a constant C (depending on F only) such that $t_{\mathrm{isol}}(F, W) \le C$ for all $W \in \mathcal{W}_{\mathrm{sym}}$, and $|t_{\mathrm{isol}}(F, W) - t_{\mathrm{isol}}(F, W')| \le C \|W - W'\|_\square$ for all $W, W' \in \mathcal{W}_{\mathrm{sym}}$.*

We shall prove Theorem 2.3 via a sequence of lemmas. The first step will be to transform (24) to an integral of a product over edges only, rather than over edges and vertices. This will involve considering asymmetric kernels, as well as different kernels for different edges of F.

Given a tree F with r vertices in which each edge has an arbitrary direction, and for every edge $ij \in F$ a (not necessarily symmetric) kernel $W_{ij} \in \mathcal{W}$, set

$$t_0\big(F, (W_{ij})_{ij \in E(F)}\big) := \int_{\mathcal{S}^r} \prod_{ij \in E(F)} W_{ij}(x_i, x_j) \, d\mu(x_1) \ldots d\mu(x_r). \tag{25}$$

Note that the exponential factors $e^{-\lambda_W(x_k)}$ present in (24) are missing from (25).

We shall reintroduce the exponential factors by attaching them to the kernels W_{ij}. Recalling the definitions of the marginals λ_W and λ'_W in (19) and (20), for real $a, b \ge 0$ let

$$W^{(a,b)}(x, y) := e^{-a\lambda_W(x)} W(x, y) e^{-b\lambda'_W(y)}. \tag{26}$$

Finally, let d_i be the (total) degree of vertex i in F. Then, comparing (24) and (25), for every symmetric $W : \mathcal{S}^2 \to [0, \infty)$ we have

$$t_{\mathrm{isol}}(F, W) = t_0\big(F, (W^{(1/d_i, 1/d_j)})_{ij}\big). \tag{27}$$

To study $t_{\mathrm{isol}}(F, W)$, we shall first study the map $W \mapsto W^{(a,b)}$, and then study the behaviour of t_0 on the restricted set of asymmetric kernels that arise as images of this map.

Lemma 2.4. *For every fixed $a, b \ge 0$, the map $W \mapsto W^{(a,b)}$ is Lipschitz continuous on $\mathcal{W}$ in the cut norm; more precisely,*

$$\|W_1^{(a,b)} - W_2^{(a,b)}\|_\square \le 7 \|W_1 - W_2\|_\square$$

for all $W_1, W_2 \in \mathcal{W}$. Also, for every $W \in \mathcal{W}$, $\sup_x \lambda_{W^{(a,b)}}(x) \le e^{-1}/a$ and $\sup_y \lambda'_{W^{(a,b)}}(y) \le e^{-1}/b$.

Surprisingly, this turns out to be the hardest part of the proof of Theorem 2.3.

Proof. Let us start with the final inequalities, which are immediate consequences of the inequality $te^{-t} \leq e^{-1}$. Indeed,

$$\begin{aligned}\lambda_{W^{(a,b)}}(x) &:= \int_{\mathcal{S}} W^{(a,b)}(x,y)\,d\mu(y) \leq \int_{\mathcal{S}} e^{-a\lambda_W(x)} W(x,y)\,d\mu(y) \\ &= e^{-a\lambda_W(x)}\lambda_W(x) \leq e^{-1}/a,\end{aligned}$$

and similarly $\lambda'_{W^{(a,b)}}(y) \leq e^{-1}/b$.

Turning to the main assertion, let $W_1, W_2 \in \mathcal{W}$. To simplify the notation set $\lambda_j := \lambda_{W_j}$ and $\lambda'_j := \lambda'_{W_j}$ for $j = 1, 2$. It will turn out that we have to argue separately according to which of $\lambda_1(x)$ and $\lambda_2(x)$ is larger, and similarly for $\lambda'_1(y)$ and $\lambda'_2(y)$. Accordingly, define the indicator functions

$$\begin{aligned}I_1(x) &:= \mathbf{1}[\lambda_1(x) \leq \lambda_2(x)], & I_2(x) &:= \mathbf{1}[\lambda_1(x) > \lambda_2(x)], \\ I'_1(y) &:= \mathbf{1}[\lambda'_1(y) \leq \lambda'_2(y)], & I'_2(y) &:= \mathbf{1}[\lambda'_1(y) > \lambda'_2(y)],\end{aligned}$$

so $I_1(x) + I_2(x) = I'_1(y) + I'_2(y) = 1$.

We may write $W_1^{(a,b)} - W_2^{(a,b)}$, a difference of two three-term products, as a telescopic sum of three terms in the usual way. In particular, we have

$$\begin{aligned}W_1^{(a,b)} - W_2^{(a,b)} = &\big(e^{-a\lambda_1(x)} - e^{-a\lambda_2(x)}\big)e^{-b\lambda'_1(y)}W_1(x,y) \\ &+ e^{-a\lambda_2(x)}\big(e^{-b\lambda'_1(y)} - e^{-b\lambda'_2(y)}\big)W_1(x,y) \\ &+ e^{-a\lambda_2(x)}e^{-b\lambda'_2(y)}\big(W_1(x,y) - W_2(x,y)\big).\end{aligned} \tag{28}$$

It will turn out that this decomposition is only useful when $\lambda_1(x) \leq \lambda_2(x)$ and $\lambda'_1(y) \leq \lambda'_2(y)$, so we shall multiply by the indicator function $I_1(x)I'_1(y)$.

To bound the final term in (28), note that $0 \leq I_1(x)e^{-a\lambda_2(x)} \leq 1$ and $0 \leq I'_1(y)e^{-a\lambda'_2(y)} \leq 1$, so from (23) we have

$$\big\| I_1(x)I'_1(y)e^{-a\lambda_2(x)}e^{-b\lambda'_2(y)}\big(W_1(x,y) - W_2(x,y)\big)\big\|_\square \leq \|W_1 - W_2\|_\square. \tag{29}$$

For the remaining terms we estimate the L^1 norm, recalling (22). Turning to the first term, by the mean value theorem, if $\lambda_1(x) \leq \lambda_2(x)$ then for some $y \in [\lambda_1(x), \lambda_2(x)]$ we have

$$e^{-a\lambda_1(x)} - e^{-a\lambda_2(x)} = a|\lambda_1(x) - \lambda_2(x)|e^{-ay} \leq a|\lambda_1(x) - \lambda_2(x)|e^{-a\lambda_1(x)},$$

where $\lambda_1(x) \leq \lambda_2(x)$ is used in the final inequality. It follows that

$$I_1(x)\big|e^{-a\lambda_1(x)} - e^{-a\lambda_2(x)}\big| \leq a|\lambda_1(x) - \lambda_2(x)|e^{-a\lambda_1(x)}.$$

Thus,

$$\begin{aligned}
&\big\|I_1(x)I_1'(y)\big(e^{-a\lambda_1(x)}-e^{-a\lambda_2(x)}\big)e^{-b\lambda_1'(y)}W_1(x,y)\big\|_{L^1(\mathcal{S}^2)}\\
&\qquad\qquad\le \big\|a\,|\lambda_1(x)-\lambda_2(x)|\,e^{-a\lambda_1(x)}W_1(x,y)\big\|_{L^1(\mathcal{S}^2)}\\
&\qquad\qquad= \int_{\mathcal{S}^2} a\,|\lambda_1(x)-\lambda_2(x)|\,e^{-a\lambda_1(x)}W_1(x,y)\,d\mu(y)\,d\mu(x)\\
&\qquad\qquad= \int_{\mathcal{S}} a\,|\lambda_1(x)-\lambda_2(x)|\,e^{-a\lambda_1(x)}\lambda_1(x)\,d\mu(x)\\
&\qquad\qquad\le e^{-1}\int_{\mathcal{S}} |\lambda_1(x)-\lambda_2(x)|\;d\mu(x) = e^{-1}\|\lambda_1-\lambda_2\|_{L^1(\mathcal{S})}\\
&\qquad\qquad\le e^{-1}\|W_1-W_2\|_{\square},
\end{aligned}$$

where we used $te^{-t}\le e^{-1}$ for the second last step and Lemma 2.2 for the final step.

Similarly, for the second term in (28) we obtain the bound

$$\big\|I_1(x)I_1'(y)e^{-a\lambda_2(x)}\big(e^{-b\lambda_1'(y)}-e^{-b\lambda_2'(y)}\big)W_1(x,y)\big\|_{L^1(\mathcal{S}^2)}\le e^{-1}\|W_1-W_2\|_{\square}.$$

Putting these two bounds together with (29), comparing with (28) we see that

$$\Big\|I_1(x)I_1'(y)\big(W_1^{(a,b)}(x,y)-W_2^{(a,b)}(x,y)\big)\Big\|_{\square}\le(1+2e^{-1})\|W_1-W_2\|_{\square}. \tag{30}$$

So far we treated the case $\lambda_1(x)\le\lambda_2(x)$, $\lambda_1'(y)\le\lambda_2'(y)$. The remaining three cases are treated similarly.

More precisely, for $\lambda_1(x)\le\lambda_2(x)$, $\lambda_1'(y)>\lambda_2'(y)$, we use

$$\begin{aligned}
W_1^{(a,b)}-W_2^{(a,b)} &= \big(e^{-a\lambda_1(x)}-e^{-a\lambda_2(x)}\big)e^{-b\lambda_1'(y)}W_1(x,y)\\
&\quad+e^{-a\lambda_2(x)}e^{-b\lambda_1'(y)}\big(W_1(x,y)-W_2(x,y)\big)\\
&\quad+e^{-a\lambda_2(x)}\big(e^{-b\lambda_1'(y)}-e^{-b\lambda_2'(y)}\big)W_2(x,y)
\end{aligned}$$

in place of (28) to prove the equivalent of (30) with $I_1(x)I_2'(y)$ in place of $I_1(x)I_1'(y)$.

For $\lambda_1(x)>\lambda_2(x)$, $\lambda_1'(y)\le\lambda_2'(y)$ we use

$$\begin{aligned}
W_1^{(a,b)}-W_2^{(a,b)} &= e^{-a\lambda_1(x)}\big(e^{-b\lambda_1'(y)}-e^{-b\lambda_2'(y)}\big)W_1(x,y)\\
&\quad+e^{-a\lambda_1(x)}e^{-b\lambda_2'(y)}\big(W_1(x,y)-W_2(x,y)\big)\\
&\quad+\big(e^{-a\lambda_1(x)}-e^{-a\lambda_2(x)}\big)e^{-b\lambda_2'(y)}W_2(x,y)
\end{aligned}$$

to obtain a bound with $I_2(x)I_1'(y)$ as the indicator function.

Finally, for $\lambda_1(x)>\lambda_2(x)$, $\lambda_1'(y)>\lambda_2'(y)$ we use

$$\begin{aligned}
W_1^{(a,b)}-W_2^{(a,b)} &= e^{-a\lambda_1(x)}e^{-b\lambda_1'(y)}\big(W_1(x,y)-W_2(x,y)\big)\\
&\quad+\big(e^{-a\lambda_1(x)}-e^{-a\lambda_2(x)}\big)e^{-b\lambda_1'(y)}W_2(x,y)\\
&\quad+e^{-a\lambda_2(x)}\big(e^{-b\lambda_1'(y)}-e^{-b\lambda_2'(y)}\big)W_2(x,y)
\end{aligned}$$

for $I_2(x)I_2'(y)$.

The key point is that in all cases, when we come to apply the bound obtained from the mean value theorem, when dealing with a term $e^{-a\lambda_1(x)} - e^{-a\lambda_2(x)}$ we obtain a bound involving $e^{-\lambda_i(x)}$ for $i = 1$ or 2 depending on which of $\lambda_1(x)$ and $\lambda_2(x)$ is larger. For the rest of the argument to work, it is important that the term we consider contains a factor $W_i(x, y)$ rather than $W_{3-i}(x, y)$. Similar comments apply to the $e^{-b\lambda_1'(y)} - e^{-b\lambda_2'(y)}$ terms. Fortunately, we can ensure that this is always the case, as shown by the decompositions above. Informally speaking, we simply choose the right moment to switch from W_1 to W_2.

Combining (30) and its equivalents, noting that $I_1(x)I_1'(y) + I_1(x)I_2'(y) + I_2(x)I_1'(y) + I_2(x)I_2'(y) = 1$, we see that

$$\|W_1^{(a,b)} - W_2^{(a,b)}\|_\square \le (4 + 8e^{-1})\|W_1 - W_2\|_\square \le 7\|W_1 - W_2\|_\square. \qquad \square$$

Remark 2.5. Although we do not care about the constant, let us note that the four estimates (29) above can be combined into a single application of (23), with $h(x) = I_1(x)e^{-\lambda_2(x)} + I_2(x)e^{-\lambda_1(x)}$ and $k(y) = I_1'(y)e^{-\lambda_2'(y)} + I_2'(y)e^{-\lambda_1'(y)}$. This gives $1 + 8e^{-1} < 4$ in place of $4 + 8e^{-1}$.

We next turn to the study of $t_0(F, \cdot)$ as defined by (25), restricting our attention to kernels with bounded marginals. It turns out that we must first study a related function t_1, which may be seen as a rooted version of t_0.

Given a rooted directed graph F with vertex set $\{1, 2, \dots, r\}$ and root 1, and functions $W_{ij} \in \mathcal{W}$, let

$$t_1\big(F, (W_{ij})_{ij\in E(F)}; x_1\big) := \int_{\mathcal{S}^{r-1}} \prod_{ij\in E(F)} W_{ij}(x_i, x_j)\, d\mu(x_2)\dots d\mu(x_r).$$

Note that this is a function of $x_1 \in \mathcal{S}$, and that

$$t_0\big(F, (W_{ij})_{ij\in E(F)}\big) = \int_{\mathcal{S}} t_1\big(F, (W_{ij})_{ij\in E(F)}; x\big)\, d\mu(x). \tag{31}$$

Let $\mathcal{W}_B := \{W \in \mathcal{W} : \sup_x \lambda_W(x), \sup_y \lambda_W'(y) \le B\}$.

Lemma 2.6. *Let F be a rooted directed tree and $(W_{ij})_{ij\in E(F)}$ a family with $W_{ij} \in \mathcal{W}_B$ for all ij. Then for all $x \in \mathcal{S}$,*

$$t_1\big(F, (W_{ij})_{ij\in E(F)}; x\big) \le B^{e(F)}.$$

Proof. A simple induction on the number $e(F)$ of edges of F. If $e(F) = 0$, so F consists of just a single vertex, then both sides are equal to 1. For $e(F) > 0$, pick a leaf v of F that is not the root, with neighbour w. We may assume without loss of generality that the edge wv is oriented from w to v. In the integrand appearing in the left hand side above, there is only one factor that depends on x_v, namely $W_{wv}(x_w, x_v)$. Integrating out over x_v, this integrates to $\lambda_{W_{wv}}(x_w)$. Replacing $\lambda_{W_{wv}}(x_w)$ by B, which is an upper bound by assumption, we see that that $t_1(F, \cdot; x) \le Bt_1(F - v, \cdot; x)$, and the result follows by induction. $\square$

Returning to the unrooted case, we are now ready for the final step in the proof of Theorem 2.3.

Lemma 2.7. *Let F be a directed tree, and $B < \infty$ a constant. For all families $(W_{ij})_{ij \in E(F)}$ and $(W'_{ij})_{ij \in E(F)}$ with $W_{ij}, W'_{ij} \in \mathcal{W}_B$, we have*

$$t_0\big(F, (W_{ij})_{ij \in E(F)}\big) \le B^{e(F)} \tag{32}$$

and

$$\big|t_0\big(F, (W_{ij})_{ij \in E(F)}\big) - t_0\big(F, (W'_{ij})_{ij \in E(F)}\big)\big| \le B^{e(F)-1} \sum_{ij \in E(F)} \|W_{ij} - W'_{ij}\|_{\square}. \tag{33}$$

Proof. The bound (32) is immediate from (31) and Lemma 2.6 by choosing an arbitrary root.

For the Lipschitz estimate (33), it suffices to treat the case where the families W_{ij} and W'_{ij} differ only on a single edge ij, say $ij = 12$. In this case, let F_1 and F_2 be the two components of $F \setminus \{12\}$, and regard these as rooted trees with roots 1 and 2, respectively. Then, simplifying the notation,

$$t_0\big(F, (W_{ij})_{ij}\big) = \int_{\mathcal{S}^2} t_1(F_1; x_1) t_1(F_2; x_2) W_{12}(x_1, x_2)\, d\mu(x_1)\, d\mu(x_2)$$

and similarly for (W'_{ij}). Thus, by (21),

$$\begin{aligned}
&\big|t_0\big(F, (W_{ij})_{ij}\big) - t_0\big(F, (W'_{ij})_{ij}\big)\big| \\
&\quad = \left| \int_{\mathcal{S}^2} t_1(F_1; x_1) t_1(F_2; x_2) \big(W_{12}(x_1, x_2) - W'_{12}(x_1, x_2)\big)\, d\mu(x_1)\, d\mu(x_2) \right| \\
&\qquad \le \|t_1(F_1)\|_\infty \|t_1(F_2)\|_\infty \|W_{12} - W'_{12}\|_{\square}.
\end{aligned}$$

The result follows by Lemma 2.6. ∎

Putting the pieces together, Theorem 2.3 follows.

Proof of Theorem 2.3. In the light of (27), this is immediate from Lemmas 2.4 and 2.7. ∎

2.3 Small components

Let $N_k(G)$ denote the number of vertices of a graph G in components of order k, and let $\rho_k(\kappa)$ denote the probability that $\mathfrak{X}_\kappa$ consists of exactly k particles in total. Our next aim is to prove the following lemma. Recall that A_n is always assumed to be n-by-n.

Lemma 2.8. *Let (A_n) be a sequence of non-negative symmetric matrices converging in $\delta_\square$ to a kernel κ, and let $k \ge 1$ be fixed. Then $\mathbb{E} N_k(G(A_n))/n \to \rho_k(\kappa)$.*

As usual in sparse random graphs, the dominant contribution will be from tree components. We start with a simple lemma showing that cyclic components can be neglected.

Let us call a sequence (A_n) of non-negative symmetric matrices (in which A_n is n-by-n as usual) *well behaved* if all the diagonal entries are zero, and $\max A_n = o(n)$, where $\max A_n$ is the largest entry in A_n. One useful property of such sequences is that for them, the models $G(A_n)$ and $G_{\mathrm{Po}}(A_n)$ are essentially equivalent, as shown by the following simple lemma.

Lemma 2.9. *Let κ be a kernel and let (A_n) be a sequence of well-behaved matrices with $\delta_\square(A_n, \kappa) \to 0$. Let A'_n be the matrix with entries defined by* (2). *Then $\delta_\square(A'_n, \kappa) \to 0$.*

Proof. For n large enough that $\max a_{ij} \le n/2$, say, from (2) we have $|a_{ij} - a'_{ij}| = O(a_{ij}^2/n)$, with the implicit constant C absolute. It follows that

$$\sum_{ij} |a_{ij} - a'_{ij}| \le C \sum_{ij} a_{ij}^2/n \le C \max\{a_{ij}/n\} \sum_{ij} a_{ij} = o(1) \sum_{ij} a_{ij},$$

using the well-behavedness assumption. Since $\delta_\square(A_n, \kappa) \to 0$, we have $\sum a_{ij} \sim n^2 \int \kappa = O(n^2)$. Hence

$$\delta_\square(\kappa_{A_n}, \kappa_{A'_n}) \le \|\kappa_{A_n} - \kappa_{A'_n}\|_{L^1} = n^{-2} \sum_{ij} |a_{ij} - a'_{ij}| = o(1),$$

and the result follows. $\square$

The point of Lemma 2.9 is that if we can prove that $G_{\mathrm{Po}}(A_n)$ has a certain property whenever $\delta_\square(A_n, \kappa) \to 0$, then the same result for $G(A_n)$ follows: we simply express $G(A_n)$ as $G_{\mathrm{Po}}(A'_n)$ as in (1), and apply our result for $G_{\mathrm{Po}}(\cdot)$ to the sequence (A'_n).

Our next lemma shows that the graphs we consider have few vertices in small components containing cycles. Let $N_k^{\mathsf{t}}(G)$ denote the number of vertices of a graph G in tree components of order k, and $N_k^{\mathsf{c}}(G)$ the number in cyclic components of order k, so $N_k(G) = N_k^{\mathsf{t}}(G) + N_k^{\mathsf{c}}(G)$.

Lemma 2.10. *Let (A_n) be a sequence of well-behaved matrices and $k \ge 2$ an integer. Then $\mathbb{E} N_k^{\mathsf{c}}(G_n) = o(n)$, where $G_n = G_{\mathrm{Po}}^{\mathrm{m}}(A_n)$.*

Note that in this lemma there is no convergence assumption. Note also that Lemma 2.10 immediately implies a corresponding result for $G_{\mathrm{Po}}(A_n)$, which is simply the simple graph underlying $G_{\mathrm{Po}}^{\mathrm{m}}(A_n)$, and so satisfies $N_k^{\mathsf{c}}(G_{\mathrm{Po}}(A_n)) \le N_k^{\mathsf{c}}(G_{\mathrm{Po}}^{\mathrm{m}}(A_n))$. It also implies a corresponding result for $G(A_n)$; this may be deduced from the result for $G_{\mathrm{Po}}(A_n)$ by expressing $G(A_n)$ as $G_{\mathrm{Po}}(A'_n)$ as above.

Proof. We shall consider an evolving version $G_n(t)$ of G_n. To define this, for each possible edge ij, construct a Poisson process on $[0, 1]$ with intensity a_{ij}/n; the points of these processes will be the birth times of the ij edges. Let $G_n(t)$

be the graph formed by all edges born by time t, noting that the number of ij edges in $G_n(1)$ is Poisson with mean a_{ij}/n. Taking the processes independent, $G_n(1)$ thus has the distribution of $G_n = G^{\mathrm{m}}_{\mathrm{Po}}(A_n)$.

Let $M_{\le k}(G)$ denote the number of cyclic components of a (multi-)graph G of order at most k; thus $N^{\mathrm{c}}_k(G) \le kM_{\le k}(G)$.

Let $f(t)$ denote the expectation of $M_{\le k}(G_n(t))$; then $f(0) = 0$ and $f(1) = \mathbb{E}M_{\le k}(G_n)$, so $\mathbb{E}N^{\mathrm{c}}_k(G_n) \le kf(1)$, and it suffices to show that the derivative of f is bounded above by $o(n)$. Condition on $G_n(t)$, and consider the edges born in a short time interval $[t, t + \mathrm{d}t]$. Taking $\mathrm{d}t$ small enough, the probability that there is more than one such edge in any interval $[t, t + \mathrm{d}t]$ is negligible. The only way we can have $M_{\le k}(G + e) \ge M_{\le k}(G)$ is if e joins two vertices i, j in some component of G of order at most k. There are at most kn such pairs of vertices. Since the a_{ij} are uniformly bounded by $o(n)$, the probability $a_{ij}\mathrm{d}t/n$ of adding $e = ij$ is $o(\mathrm{d}t)$, and the probability of adding some such edge is $o(kn\mathrm{d}t) = o(n\mathrm{d}t)$. Adding such an edge increases $M_{\le k}$ by at most 1, so the expected increase in time $\mathrm{d}t$ is at most $o(n\mathrm{d}t)$ as required. □

We are now ready to prove Lemma 2.8.

Proof of Lemma 2.8. We claim that it suffices to prove the lemma under the assumption that (A_n) is well behaved, i.e., $\max A_n = o(n)$, and the diagonal entries are 0.

To see this, note that by Lemma 2.1 there is some $\delta = \delta(n) \to 0$ such that at most δn entries of A_n exceed δn, and the sum of these entries is at most δn^2. Define $A'_n = (a'_{ij})$ by setting $a'_{ij} = 0$ if $a_{ij} > \delta n$ or if $i = j$, and setting $a'_{ij} = a_{ij}$ otherwise. Then

$$\delta_\square(A_n, A'_n) \le \frac{1}{n^2}\sum |a_{ij} - a'_{ij}| = \frac{1}{n^2}\sum_{a_{ij}>\delta n} a_{ij} + \frac{1}{n^2}\sum_{i:a_{ii}\le\delta n} a_{ii} \le \delta + \delta = o(1).$$

Hence $\delta_\square(A'_n, \kappa) \to 0$, so the sequence A'_n and kernel κ satisfy the assumptions of the lemma, and (A'_n) is well behaved. In establishing our claim we may thus assume that

$$\mathbb{E}N_k(G(A'_n))/n \to \rho_k(\kappa). \tag{34}$$

But then the same result for $G(A_n)$ follows almost immediately. Indeed, we may assume that $G(A'_n) \subset G(A_n)$, and we have

$$\mathbb{E}\big(E(G(A_n)) \setminus E(G(A'_n))\big) = \mathbb{E}\big(e(G(A_n)) - e(G(A'_n))\big) \le \frac{1}{n}\sum |a_{ij} - a'_{ij}| = o(n).$$

Since adding an edge to a graph G changes $N_k(G)$ by at most $2k$, it follows that

$$\mathbb{E}|N_k(G(A_n)) - N_k(G(A'_n))| = o(n),$$

which with (34) proves the same statement for A_n, establishing the claim.

From now on we suppose as we may that (A_n) is well behaved. In the light of Lemma 2.9 we may work with $G_{\mathrm{Po}}(A_n)$ instead of $G(A_n)$. In fact, we shall work with $G_n = G^{\mathrm{m}}_{\mathrm{Po}}(A_n)$, which has exactly the same component structure as $G_{\mathrm{Po}}(A_n)$.

Given a loopless multi-graph F on $[k]$ and a sequence $\mathbf{v} = (v_1, \ldots, v_k)$ with $1 \le v_i \le n$ for each i, set

$$p_{\mathbf{v}}(F) = p_{\mathbf{v}}(F, A_n) = \prod_{ij \in E(F)} \frac{a_{v_i v_j}}{n} \prod_{uw : \{u,w\} \cap \{v_i\} \neq \emptyset} e^{-a_{uw}/n}, \tag{35}$$

where the second product is over all edges uw of the complete graph on $[n]$ meeting $\{v_1, \ldots, v_k\}$.

Let us call a sequence $\mathbf{v} = (v_1, \ldots, v_k)$ *good* if the v_i are distinct, and *bad* otherwise. If F is a simple graph and $\mathbf{v}$ is good, then $p_{\mathbf{v}}(F)$ is the probability that the vertices $v_1, \ldots, v_k$ of $G_n = G^{\mathrm{m}}_{\mathrm{Po}}(A_n)$ form a component isomorphic to F, with the ith vertex of F mapped to v_i. Hence, writing $n_F(G_n)$ for the number of components of G_n isomorphic to F, for simple F we have

$$\mathbb{E} n_F(G_n) = \frac{1}{\mathrm{aut}(F)} \sum_{\mathbf{v} \text{ good}} p_{\mathbf{v}}(F).$$

Our aim is to relate this sum with F a tree to $t_{\mathrm{isol}}(F, \kappa_{A_n})$, and hence to $t_{\mathrm{isol}}(F, \kappa)$.

Let $\lambda_\kappa(x)$ denote the marginal of κ, defined by (19). For $1 \le i \le n$, set

$$\lambda_n(i) = \frac{1}{n} \sum_j a_{ij},$$

so λ_n is essentially the marginal of κ_{A_n}. (More precisely, $\lambda_n(i)$ gives the value of the marginal of κ_{A_n} at any point of the interval of length $1/n$ corresponding to vertex $i \in [n]$.)

Given a multi-graph F and a (not necessarily good) sequence $\mathbf{v}$, let

$$p^0_{\mathbf{v}}(F) = p^0_{\mathbf{v}}(F, A_n) = \prod_{ij \in E(F)} \frac{a_{v_i v_j}}{n} \prod_{i=1}^{k} e^{-\lambda_n(v_i)}. \tag{36}$$

Expanding each term $\lambda_n(v_i)$ and then comparing (35) and (36), we see that if $\mathbf{v}$ is good then the only difference is that certain factors $\exp(-a_{uw}/n)$ appear twice in (36) and only once in (35), namely such factors with $u, w \in \{v_1, \ldots, v_k\}$. Since there are $\binom{k}{2} = O(1)$ such factors and each is (by our well-behavedness assumption) $1 + o(1)$, we have

$$p_{\mathbf{v}}(F) \sim p^0_{\mathbf{v}}(F) \tag{37}$$

uniformly in good sequences $\mathbf{v}$. Hence, for simple F,

$$\mathbb{E} n_F(G_n) \sim \frac{1}{\mathrm{aut}(F)} \sum_{\mathbf{v} \text{ good}} p^0_{\mathbf{v}}(F). \tag{38}$$

Specializing now to the case of a tree T on $[k]$, recalling (24) we have

$$t_{\mathrm{isol}}(T, \kappa_{A_n}) = n^{-k} \sum_{\mathbf{v}} \prod_{ij \in E(T)} a_{v_i v_j} \prod_{i=1}^{k} e^{-\lambda_n(v_i)},$$

so

$$\sum_{\mathbf{v}} p^0_{\mathbf{v}}(T) = n t_{\mathrm{isol}}(T, \kappa_{A_n}).$$

Our next aim is to show that

$$\sum_{\mathbf{v} \text{ bad}} p^0_{\mathbf{v}}(T) = o(n). \tag{39}$$

Once we have done so, it follows from the formulae above that

$$\mathbb{E} n_T(G_n) = o(n) + (1 + o(1)) n \frac{t_{\mathrm{isol}}(T, \kappa_{A_n})}{\mathrm{aut}(T)}. \tag{40}$$

In any sequence $\mathbf{v}$ contributing to (39), at least one pair v_i, v_j coincides. Since $a_{ii} = 0$ for every i, we may assume that if $ij \in E(T)$, then $v_i \neq v_j$. Let us fix a *pattern* of coincidences, i.e., decide for which pairs $\{i, j\}$ we have $v_i = v_j$. The contribution to (39) from a given pattern may be bounded by

$$X(F) = \sum_{\mathbf{w} \text{ good}} p^0_{\mathbf{w}}(F), \tag{41}$$

where F is the multi-graph formed from T by identifying the appropriate vertices, and $w_1, \ldots, w_s$ runs over the distinct vertices among $v_1, \ldots, v_r$. Indeed, the only difference is that in the contribution to (39) we have factors $e^{-d_i \lambda_n(w_i)}$ rather than $e^{-\lambda_n(w_i)}$ in (41), where $d_i \geq 1$ is the number of the v_j that are mapped to w_i.

Note that F is connected. If F is simple, then using (38) again we have

$$X(F) \sim \mathrm{aut}(F) \mathbb{E} n_F(G_n) = O(n),$$

since $n_F(G_n) \leq n$. Moreover, if F is simple and not a tree, then by Lemma 2.10 we have $X(F) = o(n)$.

If F is not simple, let F' be the underlying simple graph. Then the terms of the sums defining F' and F are in one-to-one correspondence, and each term for F' is the term for F multiplied by $e(F) - e(F') \geq 1$ factors of the form a_{ij}/n. Each such factor is $o(1)$, so we have $X(F) = o(X(F'))$. We have just seen that $X(F') = O(n)$ for any connected simple F', so if F is not simple we have $X(F) = o(n)$.

Recall that we could write the sum in (39) as a sum of over $O(1)$ patterns of terms each bounded by $X(F)$ for some graph F arising from identifying some sets of non-adjacent vertices of T. Any such graph contains either a cycle or one or more multiple edges, so $X(F) = o(n)$ in all cases, establishing (39). As noted above, (40) follows.

Recall that $\delta_\square(\kappa_{A_n}, \kappa) \to 0$. By Theorem 2.3 we thus have $t_{\mathrm{isol}}(T, \kappa_{A_n}) \to t_{\mathrm{isol}}(T, \kappa) < \infty$, so

$$\mathbb{E} n_T(G_n) = n t_{\mathrm{isol}}(T, \kappa)/\operatorname{aut}(T) + o(n). \tag{42}$$

Let $\mathfrak{X}_\kappa \cong T$ denote the event that the branching process $\mathfrak{X}_\kappa$ when viewed as a tree is isomorphic to T (which implies that it has total size k). We claim that

$$\mathbb{P}(\mathfrak{X}_\kappa \cong T) = \frac{k}{\operatorname{aut}(T)} t_{\mathrm{isol}}(T, \kappa). \tag{43}$$

In fact, the version of (43) for a *rooted* tree T, which is the same except that the factor k is omitted, is easily proved using induction on k (see [3]), and then (43) follows easily by summing over the different rootings of T.

Hence, summing over all isomorphism types of trees on k vertices,

$$\rho_k(\kappa) = k \sum_T \frac{t_{\mathrm{isol}}(T, \kappa)}{\operatorname{aut}(T)},$$

and from (42),

$$\mathbb{E} N_k^{\mathsf{t}}(G_n) = \mathbb{E}\left(k \sum_T n_T(G_n)\right) = kn \sum_T \frac{t_{\mathrm{isol}}(T, \kappa)}{\operatorname{aut}(T)} + o(n) = \rho_k(\kappa) n + o(n).$$

Since $\mathbb{E} N_k^{\mathsf{c}}(G_n) = o(n)$ by Lemma 2.10, it follows that $\mathbb{E} N_k(G_n) = \rho_k(\kappa)n + o(n)$ as required, where $G_n = G^{\mathrm{m}}_{\mathrm{Po}}(A_n)$. Since $G^{\mathrm{m}}_{\mathrm{Po}}(A_n)$ and $G_{\mathrm{Po}}(A_n)$ have the same components, the corresponding statement for $G_{\mathrm{Po}}(A_n)$ follows immediately, so we have proved a version of Lemma 2.8 for the model $G_{\mathrm{Po}}(\cdot)$. As noted earlier, by Lemma 2.9, Lemma 2.8 follows. □

Lemma 2.11. *Let (A_n) be a sequence of non-negative symmetric matrices converging in $\delta_\square$ to a kernel κ, and let $k \ge 1$ be fixed. Then $N_k(G(A_n))/n \xrightarrow{\mathrm{p}} \rho_k(\kappa)$.*

Proof. As in [4] or [3], this extension of Lemma 2.8 requires almost no extra work: simply repeat the proof of Lemma 2.8 but considering pairs of components of order k to show that with $N = N_k(G(A_n))$ we have $\mathbb{E} N^2/n^2 \to \rho_k(\kappa)^2$. Since $\mathbb{E} N/n \to \rho_k(\kappa)$ by Lemma 2.8, it follows that $\operatorname{Var}(N/n) = o(1)$, so N/n is concentrated about its mean. □

As in [4] or [3] we have the following corollary, where $N_{\ge\omega} = \sum_{k \ge \omega} N_k$.

Corollary 2.12. *Let (A_n) be a sequence of symmetric n-by-n matrices converging in $\delta_\square$ to a kernel κ. Then whenever $\omega = \omega(n)$ tends to ∞ sufficiently slowly we have $N_{\ge\omega}(G(A_n))/n \xrightarrow{\mathrm{p}} \rho(\kappa)$.*

When we have completed the proof of Theorem 1.1, it will follow (arguing as in the proof of Theorem 1.2 in the reducible case) that Corollary 2.12 in fact holds for every $\omega(n) \to \infty$ with $\omega(n) = o(n)$.

2.4 Connecting the large components

To complete the proof of Theorem 1.1 we shall use a modified form of the Erdős–Rényi 'sprinkling' argument to show that almost all vertices in 'large' components are in fact in a single component. We need a strengthened form of a lemma implicit in Bollobás, Borgs, Chayes and Riordan [3]. Before stating this, let us recall another lemma from [3] (again modified, but this time in a trivial way). By an (a, b)-*cut* in a kernel κ we mean a partition (A, A^c) of $[0, 1]$ with $a \le \mu(A) \le 1 - a$ such that $\int_{A\times A^c} \kappa \le b$.

Lemma 2.13. *Let κ be an irreducible kernel, and let $0 < a < \frac{1}{2}$ be given. There is some $b = b(\kappa, a) > 0$ such that κ has no (a, b)-cut.*

Proof. The same statement is proved in [3, Lemma 7], but for *graphons*, i.e., bounded kernels; all kernels considered in [3] were bounded. Although as it happens we shall only use the bounded case, we may as well note that the restriction is entirely irrelevant. Indeed, irreducibility of a kernel κ depends only on whether certain integrals are 0, and hence only on the set where $\kappa > 0$. So if κ is irreducible, so is the pointwise minimum κ' of κ and 1. If κ has an (a, b)-cut, then so does κ', so the result follows from the bounded case. □

Here then is the key lemma that we shall need.

Lemma 2.14. *Let κ be an irreducible kernel and $\delta > 0$ a constant. There are positive constants $\alpha = \alpha(\kappa, \delta)$ and $c = c(\kappa, \delta)$ such that for every sequence (A_n) of non-negative symmetric matrices with $\delta_\square(A_n, \kappa) \to 0$, for all large enough n we have*

$$\mathbb{P}(X \sim_{\alpha n} Y) \ge 1 - \exp(-cn)$$

for all disjoint $X, Y \subset [n]$ with $|X|, |Y| \ge \delta n$, where $X \sim_k Y$ denotes the event that the graph $G(A_n)$ contains at least k vertex disjoint paths starting in X and ending in Y.

A version of this lemma, but with the additional condition that the kernel κ and entries of the matrices A_n are uniformly bounded, is implicit in [3] (see [5, Lemma 4.2]). Although the basic strategy of the proof of Lemma 2.14 is the same as that in [3], dealing with unbounded kernels requires considerable care, so we shall write out the proof in full.

Proof. We write (a_{ij}) for the entries of A_n, suppressing the dependence on n. As before, by Lemma 2.1 we may assume that $\max a_{ij} = o(n)$, and in particular that $a_{ij} \le n/100$, say. We may also assume that $\delta < 1/10$, say.

Throughout this proof we view A_n as a (dense) weighted graph. In particular, given sets V and W of vertices of A_n, i.e., subsets of $[n]$, we write

$$e(V, W) = \sum_{v\in V} \sum_{w\in W} a_{vw}$$

for the total edge weight from V to W. Similarly, for $v \in [n]$ and $W \subset [n]$,

$$e(v, W) = e(\{v\}, W) = \sum_{w \in W} a_{vw}.$$

Let $\kappa^- = \kappa \wedge 1$ be the pointwise minimum of κ and 1. Since $\delta_\square(A_n, \kappa) \to 0$, there are rearrangements κ_n of κ such that

$$\|\kappa_{A_n} - \kappa_n\|_\square \to 0. \tag{44}$$

Let $\kappa_n^- = \kappa_n \wedge 1$, noting that κ_n^- is a rearrangement of κ^-.

Identifying a subset of $[n]$ with the union of the corresponding intervals of length $1/n$ in $[0, 1]$, for subsets V and W of $[n]$ we set

$$e_0(V, W) = n^2 \int_{V \times W} \kappa_n(x, y)\, dx\, dy$$

and

$$e_0^-(V, W) = n^2 \int_{V \times W} \kappa_n^-(x, y)\, dx\, dy.$$

From (44) there is some $\eta(n) \to 0$ such that

$$\big|e(V, W) - e_0(V, W)\big| = n^2 \left| \int_{V \times W} (\kappa_{A_n} - \kappa_n) \right| \le n^2 \eta(n)$$

for all V and W. Since $\kappa \ge \kappa^-$, so $e_0(V, W) \ge e_0^-(V, W)$, it follows that

$$e(V, W) \ge e_0^-(V, W) - n^2 \eta(n). \tag{45}$$

By Lemma 2.13 there is some $b > 0$ such that κ^- has no $(\delta/2, b)$-cut. We may and shall assume that $b < 1/10$, say. Since each κ_n^- is a rearrangement of κ^-, no κ_n^- has a $(\delta/2, b)$-cut.

Fix disjoint sets X and Y of vertices, each of size at least δn. Arguing as in [3], we shall inductively define an increasing sequence S_0, S_1, …, S_ℓ of sets of vertices in a way that depends on A_n, X and Y, but not on the random graph $G(A_n)$. There will be some additional complications due to unbounded matrix entries; it turns out we can sidestep these with appropriate use of the inequality (45).

We start with $S_0 = X$, noting that $|S_0| \ge \delta n$. We shall stop the sequence when $|S_t|$ first exceeds $(1 - \delta/2)n$. Thus, in defining S_{t+1} from S_t, we may assume that $\delta n \le |S_t| \le (1 - \delta/2)n$. Since κ_n^- has no $(\delta/2, b)$-cut, we have

$$\sum_{v \notin S_t} e_0^-(v, S_t) = e_0^-(S_t^{\mathrm{c}}, S_t) = n^2 \int_{S_t^{\mathrm{c}} \times S_t} \kappa_n^- \ge bn^2.$$

Let

$$T_{t+1} = \{v \notin S_t : e_0^-(v, S_t) \ge bn/2\}.$$

Since $\kappa_n^- \leq 1$ holds pointwise, $e_0^-(v, S_t) \leq |S_t| \leq n$ for any v. Thus

$$bn^2 \leq e_0^-(S_t^c, S_t) \leq \frac{bn}{2}\big|[n] \setminus (S_t \cup T_{t+1})\big| + n|T_{t+1}| \leq \frac{bn^2}{2} + n|T_{t+1}|.$$

Hence $|T_{t+1}| \geq \frac{bn}{2}$. Set $S_{t+1} = S_t \cup T_{t+1}$, and continue the construction until we reach an S_ℓ with $|S_\ell| \geq (1 - \delta/2)n$. Note that $\ell \leq 2/b = O(1)$.

We shall now turn to the random graph $G(A_n)$, uncovering the edges between T_t and S_{t-1}, working backwards from T_ℓ. It will be convenient to set $T_0 = S_0$, so $S_t = \bigcup_{j=0}^{t} T_j$. Since $|S_\ell| \geq (1 - \delta/2)n$, while $|Y| \geq \delta n$, the set S_ℓ contains at least $\delta n/2$ vertices from Y. Since $S_0 = T_0 = X$ is disjoint from Y, it follows that there is some t_0, $1 \leq t_0 \leq \ell$, for which T_{t_0} contains a subset Y_0 of Y with

$$|Y_0| \geq \delta n/(2\ell).$$

Next, we aim to construct a set $X_0 \subset S_{t_0-1}$ with $|X_0| \geq b|Y_0|/10$ such that every $x \in X_0$ is joined to some $y \in Y_0$ by an edge of $G(A_n)$. In fact, we shall look for a partial matching from Y_0 to S_{t_0-1} of size exactly

$$N = b|Y_0|/10;$$

we ignore the irrelevant rounding to integers. Let us list the vertices of Y_0 as $\{y_1, \ldots, y_s\}$. We shall test each y_i in turn to see whether it has a neighbour in S_{t_0-1}; the complication is that we must avoid vertices of S_{t_0-1} that are neighbours of earlier y_j. We shall also stop looking for new neighbours if we already have a large enough matching.

Formally, we inductively define subsets $Z_0, Z_1, \ldots, Z_s$ of S_{t_0-1}, starting with $Z_0 = \emptyset$. For $1 \leq i \leq s$, if $|Z_{i-1}| = N$ then we set $Z_i = Z_{i-1}$. If $|Z_{i-1}| < N$ and y_i has a neighbour $z \in S_{t_0-1} \setminus Z_{i-1}$, we set $Z_i = Z_{i-1} \cup \{z\}$ for any such neighbour z. If no such neighbour exists, we set $Z_i = Z_{i-1}$. Note that $Z_0 \subset Z_1 \subset \cdots \subset Z_s$ is a random sequence of sets, and $|Z_s| \leq N$.

We claim that the following statement holds deterministically: if n is large enough, then there are at least $s/2$ values of i for which

$$e(y_i, S_{t_0-1} \setminus Z_{i-1}) \geq bn/4. \tag{46}$$

Suppose that this claim does not hold, and let $Y' \subset Y_0$ be a set of at least $s/2$ vertices y_i for which $e(y_i, S_{t_0-1} \setminus Z_{i-1}) < bn/4$. Since $Z_{i-1} \subset Z_s$, for all $y \in Y'$ we have $e(y, S_{t_0-1} \setminus Z_s) < bn/4$. Summing over y, we have

$$e(Y', S_{t_0-1} \setminus Z_s) < bn|Y'|/4.$$

From (45) it follows that

$$e_0^-(Y', S_{t_0-1} \setminus Z_s) < bn|Y'|/4 + n^2\eta(n).$$

On the other hand, since $Y' \subset T_{t_0}$, we have

$$e_0^-(Y', S_{t_0-1}) \geq bn|Y'|/2.$$

Consequently,

$$e_0^-(Y', Z_s) = e_0^-(Y', S_{t_0-1}) - e_0^-(Y', S_{t_0-1} \setminus Z_s) > bn|Y'|/4 - n^2\eta(n).$$

Since $|Y'| \geq |Y|/2 = \Theta(n)$, we see that if n is large enough, then $e_0^-(Y', Z_s) \geq bn|Y'|/5$. But κ^- is bounded by 1, so

$$e_0^-(Y', Z_s) \leq |Y'||Z_s| \leq |Y'|N = |Y'|(b|Y_0|/10) < bn|Y'|/5.$$

This contradiction establishes the claim.

Suppose that for some i we have $e(y_i, S_{t_0-1} \setminus Z_{i-1}) \geq bn/4$. Then the expected number of edges of $G(A_n)$ from y to $S_{t_0-1} \setminus Z_{i-1}$ is at least $b/4$, so the probability that there is at least one such edge is at least $b/5$.

From the claim above, and independence of edges from different vertices y, it follows that unless we reach $|Z_i| = N$ at some stage, the number of edges in the matching we find stochastically dominates a Binomial distribution D with parameters $|Y_0|/2$ and $b/4$. More precisely, the probability that $|Z_s| < N$ is at most the probability that $D < N$. But D has mean $|Y_0|b/8 \geq N = |Y_0|b/10$. Since $|Y_0| = \Theta(n)$, it follows (by Chernoff's inequality) that with probability $1 - \exp(-\Theta(n))$ we have $|Z_s| \geq N$.

In summary, with probability at least $1 - \exp(-\Theta(n))$ we find a set $X_0 = Z_s$ of at least $b|Y_0|/10$ vertices of S_{t_0-1} such that every $x \in X_0$ is joined to some $y = y(x) \in Y_0$ by an edge of $G(A_n)$, with the $y(x)$ distinct.

Suppose we do find such an X_0. As $|X_0| \geq b|Y_0|/10$, there is some $t_1 < t_0$ such that $Y_1 = X_0 \cap T_{t_1}$ contains at least $b|Y_0|/(10\ell)$ vertices. If $t_1 \geq 1$ then, arguing as above, with probability $1 - \exp(-\Theta(n))$ we find a t_2 and a set Y_2 of at least $b^2|Y_0|/(10\ell)^2$ vertices of T_{t_2} joined in $G(A_n)$ to Y_1, and so on. As the sequence $t_0, t_1, \ldots$ is decreasing, this process terminates after $s \leq \ell$ steps with $t_s = 0$. Hence, with probability $1 - \exp(-\Theta(n))$ we find a set Y_s of at least $(b/(10\ell))^\ell |Y_0| = \Theta(n)$ vertices of $T_0 = S_0 = X$ joined in $G(A_n)$ by vertex disjoint paths to vertices in Y, completing the proof of Lemma 2.14. $\square$

As in [3], Corollary 2.12 and Lemma 2.14 easily combine to give Theorem 1.1.

Proof of Theorem 1.1. Let $G_n = G(A_n)$. By Corollary 2.12 there is some $\omega = \omega(n)$ with $\omega(n) \to \infty$ such that $N_{\geq\omega}(G_n)/n \xrightarrow{\mathrm{p}} \rho(\kappa)$. We may and shall assume that $\omega = o(n)$. Since

$$C_1(G_n) + C_2(G_n) \leq \max\{2\omega, N_{\geq\omega}(G_n) + \omega\} \leq \rho(\kappa)n + o_{\mathrm{p}}(n),$$

it suffices to prove that if κ is irreducible then

$$C_1(G_n) \geq \rho(\kappa)n + o_{\mathrm{p}}(n). \tag{47}$$

If $\rho(\kappa) = 0$, then this statement holds vacuously, so suppose that κ is irreducible and $\rho(\kappa) > 0$.

Fix $0 < \varepsilon < \rho(\kappa)/10$. By [4, Theorem 6.4] we have $\rho((1-\gamma)\kappa) \nearrow \rho(\kappa)$ as $\gamma \to 0$. Fix $0 < \gamma < 1$ such that $\rho((1-\gamma)\kappa) > \rho(\kappa) - \varepsilon$.

Let $G_n' = G((1-\gamma)A_n)$ and $G_n'' = G(\gamma A_n)$ be independent. We may and shall assume that $G_n' \cup G_n'' \subseteq G_n$. Applying Corollary 2.12 to the sequence $(1-\gamma)A_n$, which tends to $(1-\gamma)\kappa$ in $\delta_\square$, we see that there is an $\omega = \omega(n)$ tending to infinity such that

$$N_{\geq \omega}(G_n') \geq (\rho((1-\gamma)\kappa) - \varepsilon)n \geq (\rho(\kappa) - 2\varepsilon)n \tag{48}$$

holds whp. Let us condition on G_n' assuming that (48) does hold. Let B be the set of vertices of G_n' in components of size at least ω (we call these components *large*), so $|B| \geq (\rho(\kappa) - 2\varepsilon)n$.

If $C_1(G_n) \leq (\rho(\kappa) - 3\varepsilon)n$ then there is a partition (X, Y) of B such that $|X|, |Y| \geq \varepsilon n$, with no path in G_n joining X to Y. Let us call such a partition *bad*. Since $G_n' \subset G_n$, each of X and Y must be a union of large components of G_n', so there are at most $2^{n/\omega(n)}$ choices for (X, Y). But the probability that a given pair (X, Y) is bad is at most the probability that there is no path in $G_n'' \subset G_n$ from X to Y; by Lemma 2.14 this probability is $\exp(-\Theta(n))$. Hence the expected number of bad partitions is $o(1)$, and whp there is no such partition. Thus $C_1(G_n) \geq (\rho(\kappa) - 3\varepsilon)n$ whp. Letting $\varepsilon \to 0$, the bound (47) follows, and this is all that is required to complete the proof of Theorem 1.1. $\square$

2.5 The reducible case: proof of Theorem 1.2

In this subsection we shall justify the terminology by showing that one can reduce the reducible case to the irreducible case. Surprisingly, in this setting (unlike that of [5]), this is not quite immediate.

The key step is a lemma allowing us to partition a sequence of matrices converging to a reducible kernel. By the *restriction* $\kappa_{\mathcal{S}}$ of a kernel κ to a set $\mathcal{S} \subset [0, 1]$ we simply mean the function obtained by restricting κ to $\mathcal{S} \times \mathcal{S}$, which we may think of as a kernel on a measure space that is no longer a probability space. It will often be convenient to consider the rescaled restriction $\kappa'_{\mathcal{S}}$: when $\mathcal{S}$ is an interval (which we can always assume) this is the kernel on $[0, 1]^2$ obtained by linearly 'stretching' $\kappa_{\mathcal{S}}$ in the obvious way.

Lemma 2.15. *Let κ be a reducible kernel and $(\mathcal{S}_1, \mathcal{S}_2)$ a partition of $[0, 1]$ with $0 < \mu(\mathcal{S}_1), \mu(\mathcal{S}_2) < 1$ such that $\kappa_{\mathcal{S}_1}$ is irreducible and κ is zero a.e. on $\mathcal{S}_1 \times \mathcal{S}_2$. If (A_n) is a sequence of non-negative symmetric matrices such that $\delta_\square(A_n, \kappa) \to 0$ then we may find for each n complementary subsets $V_{n,1}$ and $V_{n,2}$ of $[n]$ such $|V_{n,i}| \sim \mu(\mathcal{S}_i)n$ and $\delta_\square(A_{n,i}, \kappa_i') \to 0$, where $\kappa_i' = \kappa'_{\mathcal{S}_i}$ is the rescaled restriction of κ to $\mathcal{S}_i$ and $A_{n,i}$ is the principal minor of A_n obtained by selecting the rows and columns indexed by $V_{n,i}$. Moreover, the sum of the entries of A_n corresponding to $(i, j) \in V_{n,1} \times V_{n,2}$ is $o(n^2)$.*

In other words, we may split the vertex set of the random graph $G(A_n)$ into $V_{n,1}$ and $V_{n,2}$ so that the corresponding random graphs have edge probability matrices converging to the restrictions of κ to $\mathcal{S}_1$ and $\mathcal{S}_2$ respectively (after suitable rescaling).

Proof. Suppose that $\delta_\square(A_n,\kappa)\to 0$. Let (τ_n) be a sequence of measure-preserving bijections from $[0,1]$ to itself, corresponding to rearrangements of the kernels κ_{A_n}. Let $I_{n,i}=((i-1)/n,i/n]$ denote the subinterval of $[0,1]$ corresponding to vertex i, i.e., to the ith row/column of A_n. Then, in the rearrangement, $I_{n,i}\cap\tau_n(\mathcal{S}_j)$ is the portion of $I_{n,i}$ that is rearranged to correspond to part of $\mathcal{S}_j$. We write

$$s_{n,i}=\min_{j=1,2}\mu\big(I_{n,i}\cap\tau_n(\mathcal{S}_j)\big)$$

for the extent that $I_{n,i}$ is *split* between $\mathcal{S}_1$ and $\mathcal{S}_2$, noting that $0\le s_{n,i}<\mu(I_{n,i})=1/n$.

We call the sequence (τ_n) *good* if

$$\|\kappa_{A_n}^{(\tau_n)}-\kappa\|_\square\to 0, \tag{49}$$

and

$$s_n=\sum_{i=1}^n s_{n,i}=o(1).$$

Such a good sequence corresponds to rearranging A_n to be close to κ in the cut norm, while mapping almost every vertex either almost entirely into $\mathcal{S}_1$ or almost entirely into $\mathcal{S}_2$. It is not too hard to check that if such a sequence exists, then the first conclusion of the lemma follows; we omit the tedious details, noting only that since κ is integrable, for any subsets X_n of $[0,1]^2$ with measure tending to 0 we have $\int_{X_n}\kappa\to 0$. This shows that changing our rearrangement on a set of measure $o(1)$ will not affect cut norm convergence. To see that the final statement follows, let $U_{n,j}$ be the subset of $[0,1]$ corresponding to $V_{n,j}$. Then

$$\begin{aligned}\int_{U_{n,1}\times U_{n,2}}\kappa_{A_n}&=\int_{\tau_n^{-1}(U_{n,1})\times\tau_n^{-1}(U_{n,2})}\kappa_{A_n}^{(\tau_n)}\\&\le\|\kappa_{A_n}^{(\tau_n)}-\kappa\|_\square+\int_{\tau_n^{-1}(U_{n,1})\times\tau_n^{-1}(U_{n,2})}\kappa=o(1),\end{aligned}$$

since $\tau_n^{-1}(U_{n,j})$ differs from $\mathcal{S}_j$ in a set of measure $o(1)$.

It remains to prove that a good sequence exists. By hypothesis, there is a sequence (τ_n) such that (49) holds; as we shall see, any such sequence must be good! Indeed, suppose s_n does not tend to zero. Then passing to a subsequence, we may assume that $s_n\ge\delta$ for every n, for some $\delta>0$.

For every n in our (sub)sequence, and each $i\in[n]$, pick subsets $E_{i,1},E_{i,2}$ of $I_{n,i}$ of measure $s_{n,i}$ with $E_{i,j}\subset\tau_n(\mathcal{S}_j)$; this is possible by the definition of $s_{n,i}$. Finally, for $j=1,2$, let $E_j=\bigcup_{i=1}^n E_{i,j}$, noting that E_j depends on n, and that $\mu(E_j)=s_n\ge\delta$.

Since $\tau_n^{-1}(E_2)\subset\mathcal{S}_2$, we have $\int_{\tau_n^{-1}(E_2)\times\mathcal{S}_1}\kappa=0$. From (49) and the definition of the cut norm it follows that $\int_{E_2\times\tau_n(\mathcal{S}_1)}\kappa_{A_n}=o(1)$. But

$$\int_{E_1\times\tau_n(\mathcal{S}_1)}\kappa_{A_n}=\int_{E_2\times\tau_n(\mathcal{S}_1)}\kappa_{A_n},$$

since $\kappa_{A_n}(x,y)$ depends on x only through which interval $I_{n,i}$ the point x lies in, and E_1 and E_2 intersect each $I_{n,i}$ in sets of the same measure. Hence, $\int_{E_1\times\tau_n(\mathcal{S}_1)}\kappa_{A_n}=o(1)$, and, using (49) again, $I=\int_{\tau_n^{-1}(E_1)\times\mathcal{S}_1}\kappa=o(1)$.

But $\kappa_{\mathcal{S}_1}$ is irreducible, so for a.e. x in $\mathcal{S}_1$ we have $f(x)=\int_{\mathcal{S}_1}\kappa(x,y)\,dy>0$. It follows that there is some $\gamma>0$ such that the integral of f over any subset of $\mathcal{S}_1$ of measure at least δ is at least γ. But I is exactly such an integral, since $\tau_n^{-1}(E_1)\subset\mathcal{S}_1$, giving a contradiction. This contradiction shows that (τ_n) is indeed good, completing the proof. □

Using Lemma 2.15, it is not hard to deduce Theorem 1.2 from Theorem 1.1.

Proof of Theorem 1.2. Multiplying the kernel κ by c, we may and shall assume that $c=1$.

Part (a) of Theorem 1.2 follows from the first statement of Theorem 1.1; part (c) is a restatement of the second statement of Theorem 1.1, so it remains only to prove part (b).

As shown in [4, Lemma 5.17], we may decompose κ into irreducible kernels. More precisely, there is a partition $(\mathcal{S}_i)_{i=0}^N$ of $[0,1]$ with $0\le N\le\infty$ such that each $\mathcal{S}_i$ has positive measure, the restriction κ_i of κ to $\mathcal{S}_i\times\mathcal{S}_i$ is irreducible for each $i\ge 1$, and κ is zero a.e. off $\bigcup_{i=1}^N\mathcal{S}_i\times\mathcal{S}_i$.

By assumption, $\delta_\square(A_n,\kappa)\to 0$. Applying Lemma 2.15 repeatedly, for any finite $N'\le N$ we may split the vertex set $[n]$ of the graph G_n into $N'+1$ subsets $V_{n,i}$, $i=0,1,\ldots,N'$, such that, for each $i\ne 0$, $|V_{n,i}|\sim\mu(\mathcal{S}_i)n$ and $\delta_\square(A'_{n,i},\kappa'_i)\to 0$, where $A'_{n,i}$ is the submatrix of A_n corresponding to $V_{n,i}$, and $\kappa'_i=\kappa'_{\mathcal{S}_i}$ is the rescaled restriction of κ to $\mathcal{S}_i$. Let $G_{n,i}$ be the subgraph of G_n induced by $V_{n,i}$.

In what follows it is convenient to add zero rows and columns to $A'_{n,i}$ to obtain an n-by-n matrix $A_{n,i}$, and to consider the kernel κ_i on $[0,1]^2$ agreeing with κ on $\mathcal{S}_i^2$ and equal to zero off this set. It is easy to check that $\delta_\square(A'_{n,i},\kappa'_i)\to 0$ implies $\delta_\square(A_{n,i},\kappa_i)\to 0$. Although κ_i is formally reducible, it is so only in a trivial sense (called quasi-irreducible in [4]), and by rescaling suitably it is easy to check that Theorem 1.1 applies to such kernels (with, as it happens, no extra factors from the rescaling), so by Theorem 1.1 we have $C_1(G_{n,i})/n\xrightarrow{\mathrm{p}}\rho(\kappa_i)$ for each $i\ge 1$.

By assumption, $\|T_\kappa\|>1$. But

$$\|T_\kappa\|=\sup_i\|T_{\kappa_i}\|, \tag{50}$$

so there is some i with $\|T_{\kappa_i}\|>1$. We choose $N'\ge i$. Since $C_1(G_n)\ge C_1(G_{n,i})$, it follows that $C_1(G_n)=\Theta(n)$ whp as claimed. Finally, suppose that κ is bounded, by M, say. Since $\|T_{\kappa_i}\|\le M\mu(\mathcal{S}_i)$, only finitely many of the T_{κ_i} can have norm exceeding any constant, and the supremum in (50) is attained, say at $i=j$. As noted in [3], the bound $\rho(\kappa)\ge(\|T_\kappa\|-1)/\sup\kappa$ is implicit in [4]. Applying this to κ_j, the final part of Theorem 1.2(b) follows. □

Note that we cannot say what the limiting size of the giant component is in the reducible case: we know that there are $o_{\mathrm{p}}(n)$ edges joining different $G_{n,i}$, but have no further control on these edges (which may be completely absent), so we do not know whether they link the largest components in the different $G_{n,i}$ or not. Thus $C_1(G_n)/n$ may be as small as $\max_i \rho(\kappa_i) + o_p(1)$, or as large as $\rho(\kappa) + o_p(1) = \sum_i \rho(\kappa_i) + o_p(1)$.

Let us close this subsection with a conjecture. By a *rearrangement* B_n of a matrix A_n we simply mean a matrix obtained from A_n by applying some permutation to the columns, and the same permutation to the rows.

Conjecture 1. *Let κ be a kernel, and (A_n) a sequence of non-negative symmetric matrices in which A_n is n-by-n, such that $\delta_\square(A_n, \kappa) \to 0$. Then there exist rearrangements B_n of each A_n such that $\|\kappa_{B_n} - \kappa\|_\square \to 0$.*

A proof of this conjecture would give a simpler reduction of the irreducible case to the reducible one. We can prove versions of this conjecture with various additional assumptions. Suppose first that κ is of finite type. Then the proof of Lemma 2.15 adapts easily to give the desired rearrangements: first show that in rearrangements (almost) realizing the cut distance, there is no significant splitting of vertices between the parts of κ (unless two parts of κ are 'equivalent', but then they may be united into a single part). This leads eventually to a rearrangement mapping almost every vertex to some subset of some part of κ; since κ is constant on its parts, the subset is irrelevant and may be taken to be an interval, leading to the required B_n.

On the other hand, suppose that both κ and the entries of all A_n are uniformly bounded, without loss of generality by 1. Then approximating κ by some n-by-n kernel, and using a result of Borgs, Chayes, Lovász, Sós and Vesztergombi [9] that if two n-by-n kernels bounded by 1 are within distance δ in the cut metric, then there are rearrangements of the corresponding matrices that are within $32\delta^{1/67}$ in the cut norm, one can find B_n with $\|B_n - \kappa\|_\square \to 0$.

2.6 Stability

In this subsection we shall prove our stability result, Theorem 1.3, and deduce Theorem 1.4. As in [4], we adapt an argument of Luczak and McDiarmid [21] showing that for $c > 1$ constant, whp the giant component of $G(n, c/n)$ has the property that if its vertex set is divided into two pieces that are not too small, then there are many edges from one piece to the other. We shall need the following deterministic lemma from [21].

Lemma 2.16. *For any $\varepsilon > 0$, there exist $\eta_0 = \eta_0(\varepsilon) > 0$ and n_0 such that the following holds. For all $n \geq n_0$, and for all connected graphs G with n vertices, there are at most $(1+\varepsilon)^n$ bipartitions of G with at most $\eta_0 n$ cross edges.* $\square$

Using this and Lemma 2.14, we shall prove the following lemma, which corresponds roughly to the edge deletion case of Theorem 1.3.

Lemma 2.17. *Let κ be an irreducible kernel and (A_n) a sequence of non-negative symmetric matrices such that $\delta_\square(A_n, \kappa) \to 0$. For every $\varepsilon > 0$ there is* *a $\delta = \delta(\kappa, \varepsilon) > 0$ such that, whp,*

$$C_1(G_n') \geq (\rho(\kappa) - \varepsilon)n$$

for every graph G_n' that may be obtained from $G(A_n)$ by deleting at most δn edges.

Proof. We may assume that $\rho(\kappa) > 0$, as otherwise there is nothing to prove. Reducing ε if necessary, we may and shall assume that $\varepsilon < \rho(\kappa)/10$.

Let B_δ be the 'bad' event that it is possible to delete at most δn edges from $G_n = G(A_n)$ so that in what remains no component contains more than $(\rho(\kappa) - \varepsilon)n$ vertices; our aim is to show that for some constant $\delta > 0$ we have $\mathbb{P}(B_\delta) \to 0$.

Suppressing the dependence on n, given $0 < \gamma < 1$, let $G_1 = G((1-\gamma)A_n)$ and $G_2 = G(\gamma A_n)$. As before, taking G_1 and G_2 independent we may assume that $G_1 \cup G_2 \subseteq G_n = G(A_n)$. As noted earlier, by [4, Theorem 6.4] we have $\rho((1-\gamma)\kappa) \nearrow \rho(\kappa)$ as $\gamma \to 0$. Fix $0 < \gamma < 1$ such that $\rho((1-\gamma)\kappa) > \rho(\kappa) - \varepsilon/2$.

As in [21], let U_1 denote the largest component G_1, chosen according to any rule if there is a tie, and consider the event

$$A_1 := \{|U_1| \geq (\rho(\kappa) - \varepsilon/2)n\}.$$

Since $\rho((1-\gamma)\kappa) > \rho(\kappa) - \varepsilon/2$, applying Theorem 1.1 to G_1 we see that A_1 holds whp.

By Lemma 2.14, applied with $\gamma\kappa$ in place of κ, there exist constants $\alpha > 0$ and $c > 0$ such that, given two disjoint sets X, Y of vertices of G_2 with $|X|, |Y| \geq \varepsilon n/2$, we have

$$\mathbb{P}(X \sim_{\alpha n} Y) \geq 1 - e^{-cn} \tag{51}$$

for all large enough n, where $X \sim_k Y$ is the event that there are at least k vertex disjoint paths from X to Y in G_2. Let $\eta = \eta_0(c/2)$, where $\eta_0(\cdot)$ is the function appearing in Lemma 2.16, and set

$$\delta = \min\{(\rho(\kappa) - \varepsilon/2)\eta, \alpha/2\}.$$

Suppose that $B = B_\delta$ and A_1 both hold. Then there is a set E of at most δn edges of G_n such that in $G_n' = G_n - E$ there is no component with more than $(\rho(\kappa) - \varepsilon)n \leq |U_1| - \varepsilon n/2$ vertices. In particular, there is a bipartition (X, Y) of U_1 with $|X|$, $|Y| \geq \varepsilon n/2$ such that there is no path in G_n' from X to Y. But then two conditions must hold: (i) in G_1 there are at most $\delta n \leq \eta |U_1|$ edges from X to Y, and (ii) it is possible to separate X from Y in G_2 by deleting at most $\delta n < \alpha n$ edges.

Let us condition on G_1, assuming that A_1 holds. Then by Lemma 2.16, if n is large enough, there are at most $(1+c/2)^{|U_1|} \leq (1+c/2)^n \leq e^{cn/2}$ bipartitions (X, Y) of U_1 with $|X|, |Y| \geq \varepsilon n/2$ satisfying property (i). By (51), for each of these bipartitions the probability that it has property (ii) is at most e^{-cn}. It follows that $\mathbb{P}(B \cap A_1) \leq e^{cn/2} e^{-cn} = o(1)$. Since A_1 holds whp, we thus have $\mathbb{P}(B) = o(1)$, as required. □

To handle the deletion of vertices rather than edges we simply show that whp all small sets of vertices meet few edges.

Lemma 2.18. *Let κ be a kernel and $\delta > 0$ a real number. Then there is a $\gamma > 0$ such that, if (A_n) a sequence of non-negative symmetric matrices with $\delta_\square(A_n, \kappa) \to 0$, then whp every set of at most γn vertices of $G(A_n)$ meets at most δn edges.*

Proof. For $0 < \gamma < 1$ let $f(\alpha) = \sup \int_{A\times[0,1]} \kappa(x, y)\, d\mu(x)\, d\mu(y)$, where the supremum is over all subsets A of $[0, 1]$ with $\mu(A) \leq \gamma$. Since κ is integrable, we have $f(\gamma) \to 0$ as $\gamma \to 0$, and there is some γ_0 with $f(\gamma_0) < \delta/4$. Let us fix $\gamma \leq \gamma_0$ chosen small enough that $(e/\gamma)^\gamma \leq e^{\delta/20}$, say.

Given a set U of vertices of $G_n = G(A_n)$, let $\nu(U)$ denote the expectation of the sum of the degrees of the vertices in U. If $|U| \leq \gamma n$, then from the definition of the cut metric we have

$$\nu(U)/n \leq f(\gamma) + \delta_\square(A_n, \kappa),$$

so for n large enough we have $\nu(U) \leq \delta n/2$ for all such U. The number of edges incident with U has expectation at most $\nu(U)$, and is a sum of independent indicator variables. It follows from the Chernoff bounds that the probability that a given U meets at least δn edges is at most $e^{-\delta n/10}$, say. Since there are at most $\binom{n}{\gamma n} \leq (e/\gamma)^{\gamma n} \leq e^{\delta n/20}$ choices for U with $|U| = \lfloor \gamma n \rfloor$, the result follows. □

We are now ready to prove Theorem 1.3.

Proof of Theorem 1.3. Recall that G'_n will be obtained from $G_n = G(A_n)$ by deleting at most δn vertices, and then adding and deleting at most δn edges. Considering when $C_1(G'_n)$ is maximized or minimized, it clearly suffices to prove that if δ is chosen small enough, then whp $C_1(G'_n) \geq (\rho(\kappa) - \varepsilon)n$ for all such G'_n obtained by deletion only, and that whp $C_1(G'_n) \leq (\rho(\kappa) + \varepsilon)n$ for such G'_n obtained by adding edges to G_n.

The first statement is immediate from Lemmas 2.17 and 2.18 as in [4]; we omit the simple details.

The second statement follows easily Lemma 2.11; the argument is identical to that in [4]. Simply choose k such that $\sum_{k'\leq k} \rho_{k'}(\kappa) \geq 1 - \rho(\kappa) - \varepsilon/3$; then by Lemma 2.11 there are whp at least $(1 - \rho(\kappa) - \varepsilon/2)n$ vertices of G_n in components of size at most k. Set $\delta = \varepsilon/(4k)$, and note that adding at most δn edges changes the number of vertices in components of size at most k by at most $2k\delta n = \varepsilon n/2$. □

We now turn to the proof of Theorem 1.4, giving exponential tail bounds on the size of $C_1(G_n)$.

Proof of Theorem 1.4. In proving the lower bound on $C_1(G_n)$, we may assume that $\varepsilon < \rho(\kappa)$, and in particular that $\rho(\kappa) > 0$. Given a graph G, let $D = D(G)$ be the minimal d such that it is possible to delete d vertices from G to obtain a graph G' with $C_1(G') \le (\rho(\kappa) - \varepsilon)n$. Note that if G_1 and G_2 differ only in the set of edges incident with some vertex v, then $|D(G_1) - D(G_2)| \le 1$. Theorem 1.3 implies that for some $\delta > 0$ we have $\mathbb{E}D(G_n) \ge \delta n$ for all large enough n. Constructing G_n by making n independent choices, where the ith choice is the set of edges ji, $j < i$, it follows from McDiarmid's inequality [22] that

$$\mathbb{P}\big(C_1(G_n) \le (\rho(\kappa) - \varepsilon)n\big) = \mathbb{P}(D(G_n) = 0) \le e^{-2(\delta n)^2/n} = e^{-2\delta^2 n}. \tag{52}$$

(Of course, one can instead use the Hoeffding–Azuma inequality, in which case the factor two in the exponent is in the denominator.)

Turning to the upper bounds on $C_1(G_n)$ and $C_2(G_n)$, fix $k \ge 1$ with $\rho_{\le k}(\kappa) = \sum_{k' \le k} \rho_{k'}(\kappa) \ge 1 - \rho(\kappa) - \varepsilon/4$, and consider $N_n = N_{\le k}(G_n)$. We have $\mathbb{E}N_n/n \to \rho_{\le k}(\kappa)$ by Lemma 2.8, so for n large enough we have $\mathbb{E}N_n \ge (1 - \rho(\kappa) - \varepsilon/3)n$. We shall show that

$$\mathbb{P}\big(|N_n - \mathbb{E}N_n| \ge \varepsilon n/2\big) \le e^{-\gamma n} \tag{53}$$

for some $\gamma > 0$; then, for n large enough,

$$\begin{aligned}\mathbb{P}\big(C_1(G_n) + C_2(G_n) \ge (\rho(\kappa) + \varepsilon)n\big) &\le \mathbb{P}\big(N_{>k}(G_n) + 2k \ge (\rho(\kappa) + \varepsilon)n\big) \\ &\le \mathbb{P}\big(N_n \le \mathbb{E}N_n - \varepsilon n/2\big) \le e^{-\gamma n}.\end{aligned}$$

Together with (52) this gives the required bounds on $C_1(G_n)$. For the bound on $C_2(G_n)$, we use (52) to bound $C_1(G_n)$ from below, and replace ε by $\varepsilon/2$.

In our proof of (53) the key point is that $N_{\le k}(G)$ is edge-Lipschitz: if G and G' differ in one edge, then $|N_{\le k}(G) - N_{\le k}(G')| \le 2k$. To prove concentration, we apply Talagrand's inequality [24] in the form of [18, Theorem 2.29]. With $N = \binom{n}{2}$, the independent variables $Z_1, \dots, Z_N$ are the indicator functions of the events that the individual edges are present. Let $f(G_n) = f(Z_1, \dots, Z_N) = n - N_n = N_{>k}(G_n)$. Then changing one Z_i changes N_n, and hence f, by at most $c_i = 2k$. Whenever $f(G_n) \ge r$, then taking (the edge set of) one spanning tree for each component of size greater than k, there is a certificate of size at most n for the event that $f(G_n) \ge r$. Hence we may take $\psi(r) = (2k)^2 n$ for all r, and Talagrand's inequality gives

$$\mathbb{P}(|f(G_n) - m| \ge t) \le 4e^{-t^2/(16k^2 n)},$$

where m is the median value of $f(G_n)$. As usual (see, e.g., [18]), it then follows that the mean and median are close (within $O(\sqrt{n})$), and recalling that $N_n = n - f(G_n)$, for n large enough we obtain (53) with $\gamma = \varepsilon^2/(70k^2)$, say. □

3 Extension to hypergraphs

In this section we shall prove an extension of Theorems 1.1 and 1.2 to hypergraphs. Alternatively, this may be thought of as an extension of the random graph model with clustering introduced in [5]. Most of our arguments are simple modifications of those in previous sections, so we shall only outline them. There are one or two places where adapting the proof is not so easy, and there we shall give more detail.

Let $(\mathcal{S}, \mu)$ be a probability space. We write $\mathcal{W}_r$ for the set of all integrable non-negative functions $W : \mathcal{S}^r \to [0, \infty)$, and $\mathcal{W}_{r,\mathrm{sym}}$ for the subset of such functions that are symmetric under permutations of the coordinates. Often we shall call a function $\kappa_r \in \mathcal{W}_{r,\mathrm{sym}}$ an *r-kernel.* A *hyperkernel* $\underset{\sim}{\kappa}$ is simply a sequence $(\kappa_r)_{r\geq 2}$, where κ_r is an r-kernel. The *integral* $i(\underset{\sim}{\kappa})$ of a hyperkernel is defined to be

$$i(\underset{\sim}{\kappa}) = \sum_{r\geq 2} r \int_{\mathcal{S}^r} \kappa_r,$$

and a hyperkernel $\underset{\sim}{\kappa}$ is *integrable* if $i(\underset{\sim}{\kappa}) < \infty$.

The cut norm has a natural extension to r-kernels or indeed to $L^1(\mathcal{S}^r) \supset \mathcal{W}_r$. As before, we consider two slightly different definitions: for $W \in L^1(\mathcal{S}^r)$ set

$$\|W\|_{\square,1} := \sup_{S_1,\dots,S_r} \Big| \int_{S_1\times\cdots\times S_r} W(x_1, \dots, x_r) \Big|, \tag{54}$$

where the supremum is over all r-tuples of measurable subsets of $\mathcal{S}$.

Alternatively, we may consider

$$\|W\|_{\square,2} := \sup_{\|f_1\|_\infty,\cdots,\|f_r\|_\infty\leq 1} \Big| \int_{\mathcal{S}^r} f_1(x_1)\cdots f_r(x_r) W(x_1, \dots, x_r) \Big|. \tag{55}$$

Much of the time it makes no difference which version of $\|\cdot\|_\square$ we consider: as before, in the supremum in (55) we may assume that each f_i is a ± 1 function, and we see that

$$\|W\|_{\square,1} \leq \|W\|_{\square,2} \leq 2^r \|W\|_{\square,1}.$$

While (55) is the more natural definition from the point of view of functional analysis, we shall in fact take (54) as the definition for most of this section, writing $\|W\|_\square$ for $\|W\|_{\square,1}$ – it turns out that we obtain a very slightly stronger result this way.

Given a family $\underset{\sim}{W} = (W_r)_{r\geq 2}$ with $W_r \in \mathcal{W}_r$, set

$$i(\underset{\sim}{W}) = \sum_{r\geq 2} r \int_{\mathcal{S}^r} W_r,$$

$$\|\underset{\sim}{W}\|_{L^1} = \sum_{r\geq 2} r \|W_r\|_{L^1},$$

and

$$\|\underset{\sim}{W}\|_\square = \sum_{r\geq 2} r \|W_r\|_\square, \tag{56}$$

where $\|\cdot\|_\square = \|\cdot\|_{\square,1}$. The reason for the factors of r above will become clear shortly.

Note that while considering a single value of r, it is irrelevant whether we use $\|\cdot\|_{\square,2}$ or $\|\cdot\|_{\square,1}$. However, as soon as we sum cut norms for different r, the potential factor of up to 2^r may make a difference. All our results will apply using $\|\cdot\|_{\square,2}$ instead of $\|\cdot\|_{\square,1}$, but they would then be slightly weaker, as fewer sequences of hyperkernels converge in the resulting norm.

Note that for $W \in L^1(\mathcal{S}^r)$ we trivially have

$$\Big|\int_{\mathcal{S}^r} W\Big| \leq \|W\|_\square \leq \|W\|_{L^1},$$

so

$$|i(\underset{\sim}{W})| \leq \|\underset{\sim}{W}\|_\square \leq \|\underset{\sim}{W}\|_{L^1}.$$

As in [5], the quantity $i(\underset{\sim}{W})$ will play a key role in various approximation arguments; the inequality $|i(\underset{\sim}{W})| \leq \|\underset{\sim}{W}\|_\square$ is key to making these arguments work here.

Given a hyperkernel $\underset{\sim}{\kappa}$ and a measure-preserving bijection $\tau : \mathcal{S} \to \mathcal{S}$, let $\underset{\sim}{\kappa}^{(\tau)} = (\kappa_r^{(\tau)})_{r\geq 2}$ be the hyperkernel defined by

$$\kappa_r^{(\tau)}(x_1, \dots, x_r) = \kappa_r(\tau(x_1), \dots, \tau(x_r)).$$

We call a $\underset{\sim}{\kappa}^{(\tau)}$ a *rearrangement* of $\underset{\sim}{\kappa}$, and write $\underset{\sim}{\kappa}' \sim \underset{\sim}{\kappa}$ if $\underset{\sim}{\kappa}'$ is a rearrangement of $\underset{\sim}{\kappa}$. The cut metric extends to hyperkernels on $[0,1]$ as follows:

$$\delta_\square(\underset{\sim}{\kappa}, \underset{\sim}{\kappa}') = \inf_{\underset{\sim}{\kappa}'' \sim \underset{\sim}{\kappa}'} \|\underset{\sim}{\kappa} - \underset{\sim}{\kappa}''\|_\square.$$

For hyperkernels on general probability spaces, which need not be the same, we use couplings to define $\delta_\square$.

Turning to graphs, our next aim is to define an extension of the random graph $G(A_n)$.

By an *n-by-n hypermatrix* H_n we mean a sequence $(H_{n,r})_{r\geq 2}$ where each $H_{n,r}$ is an r-dimensional array with entries $h_{i_1 i_2 \dots i_r} \geq 0$, $1 \leq i_1, \dots, i_r \leq n$, that is symmetric under all permutations of the coordinates. There is a hyperkernel $\underset{\sim}{\kappa} = \underset{\sim}{\kappa}(H_n) = (\kappa_r)_{r\geq 2}$ naturally associated to a hypermatrix H_n: each κ_r is a piecewise constant function on $[0,1]^r$ whose value on a certain hypercube of side $1/n$ is given by the appropriate entry of $H_{n,r}$.

Turning to the random hypergraph, as in [5], the natural normalization in the hypergraph case is unfortunately not the same as in the graph case. Roughly speaking, for each entry $h_{i_1 i_2 \dots i_r}$ of each $H_{n,r}$, we shall add a hyperedge on the corresponding vertices to our hypergraph with probability $h_{i_1 i_2 \dots i_r}/n^{r-1}$. Unfortunately this means that the probability that a particular r-vertex hyperedge is present is then (roughly) $r! h_{i_1 i_2 \dots i_r}/n^{r-1}$, and in particular $2h_{ij}/n$ in the graph case.

Formally, given a hypermatrix H_n, let $\mathcal{H}(H_n)$ be the random hypergraph on $[n]$ in which edges are present independently, and for any $2 \leq r \leq n$ and $i_1 < i_2 < \cdots < i_r$, the probability that the hyperedge $i_1 i_2 \cdots i_r$ is present is

$$\min\{r! h_{i_1 i_2 \dots i_r}/n^{r-1}, 1\}.$$

Alternatively, it is often to convenient to consider the *Poisson multi-hypergraph* version of $\mathcal{H}(H_n)$: here the number of copies of a hyperedge $i_1 i_2 \cdots i_r$ is simply Poisson with mean $r! h_{i_1 i_2 \dots i_r}/n^{r-1}$, and these numbers are independent for different hyperedges.

Turning to the graph, let $G(H_n)$ be the simple graph underlying $\mathcal{H}(H_n)$, obtained by replacing each r-vertex hyperedge by a complete graph on r vertices, and replacing any multiple edges by single edges. In the Poisson multi-hypergraph variant, we keep multiple edges.

Remark 3.1. We call an entry $h_{i_1 i_2 \dots i_r}$ of some $H_{n,r}$ *diagonal* if $i_k = i_\ell$ for some $k \neq \ell$. Note that in the definitions of $\mathcal{H}(H_n)$ and $G(H_n)$, such entries play no role. We shall see later that, as in the graph case, convergence of (H_n) to $\underset{\sim}{\kappa}$ in $\delta_\square$ is unaffected by setting all diagonal entries to 0, so (once we have shown this), we may assume without loss of generality that all diagonal entries are 0. However, we do *not* impose this as a condition of our results, since there is no need to do so.

Given a hyperkernel $\underset{\sim}{\kappa}$, let $\mathfrak{X}_{\underset{\sim}{\kappa}}$ be the compound Poisson Galton–Watson branching process associated to $\underset{\sim}{\kappa}$; for the formal definition see [5]. We write $\rho(\underset{\sim}{\kappa})$ for the survival probability of $\mathfrak{X}_{\underset{\sim}{\kappa}}$.

As in [5], let κ_{e} be the *edge kernel* corresponding to $\underset{\sim}{\kappa} = (\kappa_r)$, defined by

$$\kappa_{\mathrm{e}}(x,y) = \sum_{r\geq 2} r(r-1) \int_{\mathcal{S}^{r-2}} \kappa_r(x,y,x_3,x_4,\dots,x_r)\, d\mu(x_3)\cdots d\mu(x_r). \qquad (57)$$

Note that κ_{e} may be viewed as a (rescaled) 2-dimensional marginal of the hyperkernel $\underset{\sim}{\kappa}$. As in [5], a hyperkernel $\underset{\sim}{\kappa}$ is *irreducible* if the corresponding edge kernel is irreducible. The natural extension of Theorem 1.1 to hyperkernels is as follows.

Theorem 3.2. *Let $\underset{\sim}{\kappa}$ be an irreducible, integrable hyperkernel and (H_n) a sequence of hypermatrices such that $\delta_\square(H_n, \underset{\sim}{\kappa}) \to 0$. Then $C_1(G(H_n))/n \xrightarrow{\mathrm{p}} \rho(\underset{\sim}{\kappa})$, and $C_2(G(H_n)) = o_{\mathrm{p}}(n)$.*

Arguing as in the proof of Lemma 1.7, one can show that Theorem 3.2 extends the corresponding result of [5].

In Theorem 3.2 we define $\delta_\square$ using $\|\cdot\|_{\square,1}$ for the cut norm. Since $\|\cdot\|_{\square,1} \leq \|\cdot\|_{\square,2}$, the corresponding result for the more natural definition using $\|\cdot\|_{\square,2}$ follows immediately.

The heart of the proof of Theorem 3.2 will be Lemma 3.3 below, showing that under an additional assumption, the number of vertices in components of each fixed size is 'what it should be'. Later we shall first remove the additional assumption, and then pass from 'large' components to a single giant component.

We say that a hyperkernel $\underset{\sim}{\kappa} = (\kappa_r)$ is *R-bounded* if κ_r is zero for $r > R$, in which case we shall often speak of the hyperkernel $\underset{\sim}{\kappa} = (\kappa_r)_{r=2}^R$. Correspondingly, a hypermatrix $H_n = (H_{n,r})_{r\geq 2}$ is *R-bounded* if $H_{n,r}$ is the zero matrix for $r > R$.

As in [5], we write $\rho_k(\underset{\sim}{\kappa})$ for the probability that the branching process $\mathfrak{X}_{\underset{\sim}{\kappa}}$ consists of k particles in total. Recall that $N_k(G)$ denotes the number of vertices of a graph G in components of order k.

Lemma 3.3. *Let $R \geq 2$ be fixed. Suppose that $\underset{\sim}{\kappa}$ is an R-bounded hyperkernel and (H_n) is a sequence of R-bounded hypermatrices such that $\delta_\square(H_n, \underset{\sim}{\kappa}) \to 0$. Then for each $k \geq 1$ we have $N_k(G(H_n))/n \xrightarrow{\mathrm{p}} \rho_k(\underset{\sim}{\kappa})$.*

The proof of this lemma will take up the next several subsections. The deduction of Theorem 3.2 will then be relatively easy.

3.1 Eliminating large edge probabilities

Given a hypermatrix H_n, for $r \geq 2$ let $A_{n,r}$ be the matrix with entries

$$a_{ij}^{(r)} = n^{-(r-2)} \sum_{i_3} \sum_{i_4} \cdots \sum_{i_r} h_{iji_3i_4\dots i_r}, \tag{58}$$

and let

$$A_n = \sum_{r\geq 2} r(r-1)A_{n,r} \tag{59}$$

be the *marginal matrix* corresponding to H_n, with entries a_{ij}. Note that the kernel κ_{A_n} defined from A_n is simply the edge kernel κ_{e} corresponding to $\underset{\sim}{\kappa}(H_n)$. Also, in the Poisson multi-graph form of our model, if all diagonal entries are zero, then the expected number of ij edges in $G(H_n)$ is exactly a_{ij}/n. (See Remark 3.1.)

Given $W_r \in L^1(\mathcal{S}^r)$, let $\widehat{W}_r$ be its marginal with respect to the first two coordinates, defined by

$$\widehat{W}_r(x,y) = \int_{\mathcal{S}^{r-2}} W_r(x,y,x_3,\dots,x_r)\, d\mu(x_3)\cdots d\mu(x_r).$$

Note that

$$\|\widehat{W}_r\|_\square \leq \|W_r\|_\square. \tag{60}$$

Indeed, to see this simply take $S_3,\dots,S_r = \mathcal{S}$ in (54), or $f_3,\dots,f_r = 1$ in (55).

An immediate consequence is the following lemma.

Lemma 3.4. *Let $R \geq 2$ be fixed, and suppose that (H_n) is a sequence of R-bounded hypermatrices and $\underset{\sim}{\kappa}$ an R-bounded hyperkernel with $\delta_\square(H_n, \underset{\sim}{\kappa}) \to 0$. Then $\delta_\square(A_n, \kappa_{\mathrm{e}}) \to 0$, where A_n is the marginal matrix of H_n, and κ_{e} is the edge kernel of $\underset{\sim}{\kappa}$.*

Proof. By definition of $\delta_\square$, there are measure-preserving bijections $\tau_n : \mathcal{S} \to \mathcal{S}$ such that $\|\underset{\sim}{\kappa}(H_n) - \underset{\sim}{\kappa}^{(\tau_n)}\|_\square \to 0$. With $\underset{\sim}{\kappa} = (\kappa_r)_{r=2}^R$, writing κ_r' for the r-kernel corresponding to $H_{n,r}$, this says exactly that $\sum_{r=2}^R r\|\kappa_r' - \kappa_r^{(\tau_n)}\|_\square \to 0$. Using

(60), and noting that taking marginals commutes with rearrangement, it follows that $\sum_{r=2}^{R} r\|\kappa_{A_{n,r}} - \widehat{\kappa}_r^{(\tau_n)}\|_\square \to 0$. Since $\|\cdot\|_\square$ is a norm on $L^1(\mathcal{S}^2)$, we have

$$\|\kappa_{A_n} - \kappa_{\mathrm{e}}{}^{(\tau_n)}\|_\square \le \sum_{r=2}^{R} r(r-1)\|\kappa_{A_{n,r}} - \widehat{\kappa}_r^{(\tau_n)}\|_\square \to 0,$$

since changing the factor r to $r(r-1)$ does not affect convergence to zero. Hence $\delta_\square(A_n, \kappa_{\mathrm{e}}) \to 0$. □

Remark 3.5. To obtain a result analogous to (3.4) without the R-boundedness assumption, we would have to redefine $\delta_\square$ for hyperkernels, replacing the factor r in (56) by a factor $r(r-1)$, and only considering 'edge-integrable' limits $\underset{\sim}{\kappa}$, i.e., hyperkernels with $\sum_r r(r-1)\int \kappa_r$ finite.

Let us call a sequence (H_n) of hypermatrices *well behaved* if two conditions hold: every diagonal entry is zero, and $\max A_n/n \to 0$ as $n \to \infty$, where $\max A_n$ is the largest entry of the n-by-n marginal matrix A_n corresponding to H_n. Note that if (H_n) is well behaved, then the probability that some particular edge ij is present in $G(H_n)$ is $o(1)$ as $n \to \infty$, where the bound is uniform over edges.

Lemma 3.6. *Let $R \ge 2$ be fixed, and suppose that (H_n) is a sequence of R-bounded hypermatrices and $\underset{\sim}{\kappa}$ is an R-bounded hyperkernel with $\delta_\square(H_n, \underset{\sim}{\kappa}) \to 0$. Then there is a sequence of well-behaved R-bounded hypermatrices (H'_n) such that $\|\underset{\sim}{\kappa}(H_n) - \underset{\sim}{\kappa}(H'_n)\|_{L^1} \to 0$ and $\delta_\square(H'_n, \underset{\sim}{\kappa}) \to 0$.*

Proof. Let A_n be the marginal matrix corresponding to H_n and let κ_{e} the edge kernel corresponding to $\underset{\sim}{\kappa}$. Then by Lemma 3.4 we have $\delta_\square(A_n, \kappa_{\mathrm{e}}) \to 0$. By Lemma 2.1 there is a function $M(n)$ with $M(n) = o(n)$ such that only $o(n)$ entries of A_n exceed $M(n)$, and the sum of these entries is $o(n^2)$. This immediately implies that the sum of any n entries of A_n is $o(n^2)$.

Call an entry a_{ij} of A_n *bad* if either $a_{ij} > M(n)$ or $i = j$. Let S be the sum of the bad entries, so $S = o(n^2)$. To define H'_n, simply modify H_n by setting to 0 any entry $h_{i_1 i_2 \ldots i_r}$ of $H_{n,r}$ such that $a_{i_k i_\ell}$ is bad for some pair i_k, i_ℓ, $k < \ell$. (In other words, we replace all entries contributing to bad entries a_{ij} in the marginal by zero.) Then H'_n is a hypermatrix, and its marginal $A'_n = (a'_{ij})$ satisfies $a'_{ij} \le a_{ij}$ with $a'_{ij} = 0$ whenever a_{ij} is bad. Thus (H'_n) is well behaved.

Finally, for each r, we may think of modifying $H_{n,r}$ to obtain $H'_{n,r}$ in $\binom{r}{2}$ stages, in each one fixing k and ℓ and setting to zero entries $h_{i_1 i_2 \ldots i_r}$ for which $a_{i_k i_\ell}$ is bad. The sum of the entries set to zero at each stage is at most $n^{r-2}S$. It follows easily that

$$\|\underset{\sim}{\kappa}(H_n) - \underset{\sim}{\kappa}(H'_n)\|_{L^1} \le \sum_{r=2}^{R} \binom{r}{2} S n^{-2} = O(S/n^2) = o(1).$$

The final statement follows immediately, since

$$\delta_\square(H_n, H'_n) = \delta_\square(\underset{\sim}{\kappa}(H_n), \underset{\sim}{\kappa}(H'_n)) \le \|\underset{\sim}{\kappa}(H_n) - \underset{\sim}{\kappa}(H'_n)\|_\square \le \|\underset{\sim}{\kappa}(H_n) - \underset{\sim}{\kappa}(H'_n)\|_{L^1}.$$

□

An immediate consequence of Lemma 3.6 is the following rather informally worded corollary.

Corollary 3.7. *In proving Lemma 3.3, we may assume that* (H_n) *is well behaved.*

Proof. Let (H_n) and $\underset{\sim}{\kappa}$ satisfy the assumption of Lemma 3.3, and define (H'_n) as in Lemma 3.6. Let $G'_n = G(H'_n)$ and $G_n = G(H_n)$. There is a natural coupling of $\mathcal{H}(H'_n)$ and $\mathcal{H}(H_n)$ in which the expected number of r-vertex hyperedges in the symmetric difference is at most $n\|\kappa_{H'_{n,r}} - \kappa_{H_{n,r}}\|_{L^1}$ (with equality if all diagonal entries are zero, at least in the Poisson multi-hypergraph version); by Lemma 3.6 this number is $o(n)$. Since each hyperedge has at most R vertices, and so contributes at most $\binom{R}{2} = O(1)$ edges, summing over $2 \le r \le R$ we have $\mathbb{E}|E(G'_n) \triangle E(G_n)| = o(n)$.

Now $\delta_\square(H'_n, \underset{\sim}{\kappa}) \to 0$, so if Lemma 3.3 holds in the well-behaved case, then $N_k(G'_n)/n \xrightarrow{\mathrm{p}} \rho_k(\underset{\sim}{\kappa})$. Since adding or deleting an edge to a graph G changes the number of vertices in components of order k by at most $2k$, we have $\mathbb{E}|N_k(G_n) - N_k(G'_n)| = o(n)$, so $N_k(G_n)/n \xrightarrow{\mathrm{p}} \rho_k(\underset{\sim}{\kappa})$ follows. ☐

3.2 Hypertree integrals

Throughout this subsection, we fix an integer $R \ge 2$. All hyperkernels will be R-bounded, and all edges of all hypergraphs will have size at most R.

A *hypertree* is simply a connected hypergraph containing no cycles, or, equivalently, a connected hypergraph $\mathcal{H}$ in which $|\mathcal{H}| = 1 + \sum(|E_i| - 1)$, where the sum runs over all edges E_i of $\mathcal{H}$.

Given a hyperkernel $\underset{\sim}{\kappa} = (\kappa_r)_{r\ge 2}$ and a hypertree $\mathcal{H}$, we shall define $t_{\mathrm{isol}}(\mathcal{H}, \underset{\sim}{\kappa})$ in analogy with (24). Unfortunately, there is a difference in the normalization, and the marginals need some further explanation. For the latter, given $W_r \in L^1(\mathcal{S}^r)$, let

$$\lambda_{W_r}(x) = \lambda^{(1)}_{W_r}(x) = \int_{\mathcal{S}^{r-1}} W_r(x, x_2, \dots, x_r)\, d\mu(x_2) \cdots d\mu(x_r).$$

The marginal $\lambda^{(i)}_{W_r}$ of W_r with respect to the ith coordinate is defined similarly.

Given $\underset{\sim}{\kappa} = (\kappa_r)_{r=2}^R$, let

$$\lambda(x) = \lambda_{\underset{\sim}{\kappa}}(x) = \sum_r r\lambda_{\kappa_r}(x). \tag{61}$$

The reason for the extra factor r is that, as noted earlier, we essentially add a hyperedge on each ordered r-tuple $v_1, \dots, v_r$ with a probability κ_r/n^{r-1}, and because a particular vertex could appear in r places in the ordered r-tuple, it is then $\lambda(x)$ that gives the expected number of hyperedges containing a given vertex.

We now define $t_{\mathrm{isol}}(\mathcal{H}, \underset{\sim}{\kappa})$ as an integral over $\mathcal{S}^{|\mathcal{H}|}$ with one variable x_i for each vertex i of $\mathcal{H}$. The integrand has a factor $r!\kappa_r(x_{i_1}, \dots, x_{i_r})$ for each r-element hyperedge $E = i_1 i_2 \dots i_r$ of $\mathcal{H}$, and a factor $e^{-\lambda_{\underset{\sim}{\kappa}}(x_i)}$ for each i

With this definition, Theorem 2.3 extends to the hyperkernel context.

Theorem 3.8. *Let $R \geq 2$ be fixed, and let $\mathcal{H}$ be a hypertree in which each hyperedge has at most R elements. Then $\underset{\sim}{\kappa} \mapsto t_{\mathrm{isol}}(\mathcal{H}, \underset{\sim}{\kappa})$ is a bounded map on the space $\mathcal{W}_{\mathrm{sym}}^{(R)}$ of R-bounded hyperkernels and is Lipschitz continuous in the cut norm. In other words, there exists a constant C (depending on R and $\mathcal{H}$ only) such that $t_{\mathrm{isol}}(\mathcal{H}, \underset{\sim}{\kappa}) \leq C$ for all $\underset{\sim}{\kappa} \in \mathcal{W}_{\mathrm{sym}}^{(R)}$, and $|t_{\mathrm{isol}}(\mathcal{H}, \underset{\sim}{\kappa}) - t_{\mathrm{isol}}(\mathcal{H}, \underset{\sim}{\kappa}')| \leq C\|\underset{\sim}{\kappa} - \underset{\sim}{\kappa}'\|_\square$ for all $\underset{\sim}{\kappa}, \underset{\sim}{\kappa}' \in \mathcal{W}_{\mathrm{sym}}^{(R)}$.*

Rather than give a formal proof, we shall briefly describe the modifications needed to the arguments in Subsection 2.2. Note that we make take $\|\cdot\|_\square = \|\cdot\|_{\square,1}$ or $\|\cdot\|_\square = \|\cdot\|_{\square,2}$ in Theorem 3.8; on R-bounded hyperkernels, these norms are equivalent. As in Subsection 2.2, in this subsection we use the norm $\|\cdot\|_{\square,2}$.

Firstly, note that Lemma 2.2 extends immediately: if $W_r, W_r' \in L^1(\mathcal{S}^r)$, then

$$\|\lambda_{W_r} - \lambda_{W_r'}\|_{L^1} \leq \|W_r - W_r'\|_\square. \tag{62}$$

(Perhaps the nicest way to see this is to note that, generalizing (60) in the natural way, the cut norm of any d-dimensional marginal of some $W \in L^1(\mathcal{S}_r)$ is at most $\|W\|_\square$, and that on $L^1(\mathcal{S})$, the L^1 norm and cut norm coincide.)

Fix $\mathcal{H}$. Extending (25), suppose that for each r-element hyperedge E of $\mathcal{H}$ we have a $W_E \in \mathcal{W}_r$, where $\mathcal{W}_r$ is the set of (not necessarily symmetric) non-negative functions $W_r \in L^1(\mathcal{S}^r)$. Then we may define $t_0(\mathcal{H}, (W_E)_{E \in E(\mathcal{H})})$ in analogy with (25), again without the exponential factors in $t_{\mathrm{isol}}(\mathcal{H}, \underset{\sim}{\kappa})$. To reintroduce these, given any $W_r \in \mathcal{W}_r$ and $\mathbf{a} = (a_1, \ldots, a_r)$ with each $a_i \geq 0$, set

$$W_r^{\mathbf{a}}(x_1, \ldots, x_r) = W_r(x_1, \ldots, x_r) \prod_{i=1}^{r} \exp\bigl(-a_i \lambda_{W_r}^{(i)}(x_i)\bigr),$$

in analogy with (26).

The proof of Lemma 2.4 extends *mutatis mutandis* to give the following result.

Lemma 3.9. *For every fixed $\mathbf{a} \geq 0$, the map $W \mapsto W^{\mathbf{a}}$ is Lipschitz continuous on $\mathcal{W}_r$ in the cut norm; more precisely,*

$$\|W_1^{\mathbf{a}} - W_2^{\mathbf{a}}\|_\square \leq (2^r + r2^r/e)\|W_1 - W_2\|_\square$$

for all $W_1, W_2 \in \mathcal{W}_r$. Also, for every $W \in \mathcal{W}_r$, the ith marginal of $W^{\mathbf{a}}$ is bounded by e^{-1}/a_i. $\square$

As before, the first 2^r can be replaced by 1, but we do not care about the constant.

There is one minor additional complication not present in the graph case, which we now describe. Given a hyperkernel $\underset{\sim}{\kappa} = (\kappa_r)_{r=2}^R$, for each hyperedge E of $\mathcal{H}$ with r vertices define $W_E \in \mathcal{W}_r$ by

$$W_E(x_1, \ldots, x_r) = \kappa_r(x_1, \ldots, x_r) \prod_{i=1}^{r} \exp\bigl(-\lambda_{\underset{\sim}{\kappa}}(x)/d_i\bigr), \tag{63}$$

where d_i is the degree in $\mathcal{H}$ of the ith vertex of E (in some arbitrary ordering). Then we have

$$t_{\mathrm{isol}}(\mathcal{H}, \underset{\sim}{\kappa}) = t_0(\mathcal{H}, (W_E)_{E \in E(\mathcal{H})}), \tag{64}$$

corresponding to (27). In the graph case we simply had $W_{ij} = \kappa^{(1/d_i, 1/d_j)}$, but this no longer holds, since the marginals appearing in (63) are those of $\underset{\sim}{\kappa}$, not simply those of the kernel κ_r appropriate for r-element hyperedges. The extra complication is dealt with by Lemma 3.10 below.

Given $B > 0$, let $\mathcal{W}_{r,B}$ be the set of $W \in \mathcal{W}_r$ with all marginals bounded by B. If $f \in L^1(\mathcal{S})$ and $W \in \mathcal{W}_r$, define fW by

$$(fW)(x_1, \dots, x_r) = f(x_1) W(x_1, \dots, x_r).$$

Suppose that $W \in \mathcal{W}_{r,B}$ and $f_1, f_2 \in L^1(\mathcal{S})$. Then

$$\|(f_1 - f_2)W\|_\square \le \|(f_1 - f_2)W\|_{L^1} = \|(f_1 - f_2)\lambda\|_{L^1} \le B\|(f_1 - f_2)\|_{L^1}, \tag{65}$$

where λ is the first marginal of W. Now suppose that $f_1, \dots, f_r, f_1', \dots, f_r' \in L^1(\mathcal{S})$ with $\|f_i\|_\infty, \|f_i'\|_\infty \le 1$ for each i, and that $W, W' \in \mathcal{W}_{r,B}$. Defining $f_1 \cdots f_r W$ and $f_1' \cdots f_r' W'$ in the obvious way, we have

$$\|(f_1 \cdots f_r W) - (f_1' \cdots f_r' W')\|_\square \le \|W - W'\|_\square + B \sum_{i=1}^r \|f_i - f_i'\|_{L^1}. \tag{66}$$

Indeed, we may write the difference as $(f_1 \cdots f_r)(W - W')$ plus r terms whose cut norms may be bounded by (65); the cut norm of the first term is at most $\|W - W'\|_\square$ by the analogue of (23).

With $\mathcal{H}$ fixed, let $B = \Delta(\mathcal{H})/e$.

Lemma 3.10. *For each hyperedge E of $\mathcal{H}$, the map $\underset{\sim}{\kappa} \mapsto W_E$ is Lipschitz continuous with respect to the cut norm, and W_E belongs to $\mathcal{W}_{r,B}$.*

Proof. Let r be the number of vertices in E, and let $\underset{\sim}{\kappa} = (\kappa_s)_{s=2}^R$. Let $\widetilde{W}_E = \kappa_r^{\mathbf{a}}$, where $\mathbf{a} = (r/d_1, \dots, r/d_r)$. Since each κ_s is symmetric, all its marginals are equal; we write λ_s for any of these marginals. Then $W_E = f_1 \cdots f_r \widetilde{W}_E$, where

$$f_i(x_i) = \exp\bigl(-\lambda_{\underset{\sim}{\kappa}}(x_i)/d_i + r\lambda_r(x_i)/d_i\bigr) = \exp\Bigl(-\sum_{s \ne r} s\lambda_s(x_i)/d_i\Bigr).$$

Since all marginals λ_s are non-negative, we have $0 < f_i(x) \le 1$. Applying Lemma 3.9 to κ_r tells us that $\widetilde{W}_E \in \mathcal{W}_{r,B}$, and that the map $\underset{\sim}{\kappa} \mapsto \widetilde{W}_E$ is Lipschitz continuous. Summing (62) over $2 \le s \le R$, $s \ne r$, tells us that each f_i varies continuously (in L^1) with $\underset{\sim}{\kappa}$, and Lipschitz continuity of $\underset{\sim}{\kappa} \mapsto W_E$ then follows from (66). Finally, $\widetilde{W}_E \in \mathcal{W}_{r,B}$ and $0 < f_i \le 1$ for each i trivially implies $W_E \in \mathcal{W}_{r,B}$. $\square$

In the light of (64) and Lemma 3.10, it remains only to prove an analogue of Lemma 2.7, showing that $t_0(\mathcal{H}, (W_E)_{E \in \mathcal{H}})$ is Lipschitz continuous with respect to the cut norm when we assume that each $W_E \in \mathcal{W}_{r,B}$. The proofs of

Lemma 2.6 and Lemma 2.7 carry over with trivial modifications, noting that for the latter when we delete a single hyperedge E with r vertices, our hypertree splits into r hypertrees (some of which may be trivial).

3.3 Small components

With the preparation above behind us, the argument of Subsection 2.3 goes through easily. Let us comment very briefly on the changes. Firstly, it is more convenient in this subsection to consider hypergraphs throughout.

Given a hypergraph $\mathcal{H}$, we write $N_k(\mathcal{H})$ for the number of vertices in components of order k, $N_k^{\mathsf{t}}(\mathcal{H})$ for the number in tree components of order k, and $N_k^{\mathsf{c}}(\mathcal{H})$ for the number in non-tree components.

The proof of Lemma 2.10 carries over easily to give the following result.

Lemma 3.11. *Let (H_n) be a well-behaved R-bounded sequence of hypermatrices, and $\mathcal{H}_n = \mathcal{H}(H_n)$ the corresponding random (Poisson multi-)hypergraphs. Then for any fixed k we have $\mathbb{E}N_k^{\mathsf{c}}(\mathcal{H}_n) = o(n)$.*

Proof. As in the graph case, we consider the number $M_{\leq k}(\mathcal{H})$ of components of a hypergraph $\mathcal{H}$ that contain a cycle and have at most k vertices. Since $N_k^{\mathsf{c}}(\mathcal{H}_n) \leq kM_{\leq k}(\mathcal{H})$, it suffices to prove that $\mathbb{E}M_{\leq k}(\mathcal{H}_n) = o(n)$.

When adding a hyperedge E to a hypergraph $\mathcal{H}$, the quantity $M_{\leq k}$ can increase only if E creates a cycle, i.e., contains at least two vertices i and j from some component C of $\mathcal{H}$, and after adding $\mathcal{H}$, the component containing E has order at most k. This certainly implies that E contains a pair $\{i, j\}$ of distinct vertices from some component of order at most k. The rest of the proof follows that of Lemma 2.10, using the fact that (H_n) well behaved guarantees that the expected number of edges of $\mathcal{H}_n$ containing a particular pair $\{i, j\}$ of vertices is $o(1)$, uniformly in i and j. $\square$

The remaining arguments in Subsection 2.3 carry over easily.

Proof of Lemma 3.3. Let (H_n) be a sequence of R-bounded hypermatrices converging in $\delta_\square$ to an R-bounded hyperkernel $\underset{\sim}{\kappa}$. By Corollary 3.7 we may assume that (H_n) is well behaved.

Given a hyperedge $E = i_1 \dots i_r$ with vertices contained in $[n]$, let $h_E = h_{i_1 \dots i_r}$ be the corresponding entry of $H_{n,r}$, and $\mu_E = r! h_E n^{-(r-1)}$ the expected number of copies of E in $\mathcal{H}_n = \mathcal{H}(H_n)$. Given a connected simple hypergraph $\mathcal{F}$ on $[k]$ and a sequence $\mathbf{v} = (v_1, \dots, v_k)$ of vertices of $\mathcal{H}_n$, for each hyperedge $E = i_1 \dots i_r$ of $\mathcal{F}$ let $\mathbf{v}(E) = v_{i_1} \dots v_{i_r}$ be the image of E under the map $i \mapsto v_i$.

As before, for a good sequence $\mathbf{v}$, let $p_{\mathbf{v}}(\mathcal{F}) = p_{\mathbf{v}}(\mathcal{F}, H_n)$ be the probability that the image of $\mathcal{F}$ under $i \mapsto v_i$ is present in $\mathcal{H}_n$, and forms a component of $\mathcal{H}_n$. Thus

$$p_{\mathbf{v}}(\mathcal{F}) = \prod_{E \in E(\mathcal{F})} \mu_{\mathbf{v}(E)} \prod_{E \in E_0} \exp(-\mu_E),$$

where E_0 is the set of all potential edges of $\mathcal{H}_n$ that share at least one vertex with $\{v_1, \ldots, v_k\}$. For any $\mathbf{v}$, set

$$p_{\mathbf{v}}^0(\mathcal{F}) = \prod_{E \in E(\mathcal{F})} \mu_{\mathbf{v}(E)} \prod_{i=1}^{k} \exp(-\lambda_n(v_i)),$$

where $\lambda_n(v)$ is the sum of the probabilities of all hyperedges meeting v. Note that λ_n is exactly the marginal of the hyperkernel corresponding to H_n, but here viewed as a function on $[n]$ rather than on $[0,1]$.

If $\mathbf{v}$ is good, the only difference between $p_{\mathbf{v}}^0(\mathcal{F})$ and $p_{\mathbf{v}}(\mathcal{F})$ is that for each $E \in E_0$ sharing $s \geq 2$ vertices with $\{v_1, \ldots, v_k\}$, the factor $\exp(-\mu_E)$ appears s times in $p_{\mathbf{v}}^0(\mathcal{F})$ but only once in $p_{\mathbf{v}}(\mathcal{F})$. Since (H_n) is well behaved, for any $i \neq j$ the sum of μ_E over hyperedges E containing both i and j is $o(1)$, so it follows as before that $p_{\mathbf{v}}^0(\mathcal{F}) \sim p_{\mathbf{v}}(\mathcal{F})$.

Let $\mathcal{T}$ be a hypertree. Summing $p_{\mathbf{v}}^0(\mathcal{T})$ over *all* sequences $\mathbf{v}$ we obtain exactly $nt_{\mathrm{isol}}(\mathcal{T}, \underset{\sim}{\kappa})$. The rest of the proof of Lemma 2.8 goes through essentially unchanged to show that the contribution from bad sequences $\mathbf{v}$ is negligible, and summing over hypertrees $\mathcal{T}$, and using Lemma 3.11, it follows that $\mathbb{E}N_k(\mathcal{H}_n)/n \to \rho_k(\underset{\sim}{\kappa})$. (Note that (43) holds unchanged for hypergraphs too, with the normalizations used here.) As before, considering disjoint copies of two trees gives convergence in probability, as required. $\square$

Finally, we note that the result we have just proved extends from R-bounded hyperkernels to general hyperkernels.

Corollary 3.12. *Let $\underset{\sim}{\kappa}$ be an integrable hyperkernel and (H_n) a sequence of hypermatrices with $\delta_\square(H_n, \underset{\sim}{\kappa}) \to 0$, and set $G_n = G(H_n)$. Then $N_k(G_n)/n \xrightarrow{\mathrm{p}} \rho_k(\underset{\sim}{\kappa})$.*

Proof. Firstly, it makes no difference whether we work with the hypergraphs $\mathcal{H}_n = \mathcal{H}(H_n)$ or the underlying graphs $G_n = G(H_n)$, as these have exactly the same components.

Fix $k \geq 1$. Let $\underset{\sim}{\kappa} = (\kappa_r)_{r \geq 2}$. For $R \geq 2$, set $\underset{\sim}{\kappa}^R = (\kappa_r)_{r=2}^R$, and similarly define H_n^R by omitting all matrices $H_{n,r}$ with $r > R$. Fix $\varepsilon > 0$. Since $\underset{\sim}{\kappa}$ is integrable, we have $i(\underset{\sim}{\kappa}^R) \nearrow i(\underset{\sim}{\kappa})$ as $R \to \infty$. By Theorem 2.13(i) of [5], we have $\rho_k(\underset{\sim}{\kappa}^R) \to \rho_k(\underset{\sim}{\kappa})$. Hence there is some R such that $i(\underset{\sim}{\kappa} - \underset{\sim}{\kappa}^R) \leq \varepsilon$ and

$$|\rho_k(\underset{\sim}{\kappa}) - \rho_k(\underset{\sim}{\kappa}^R)| \leq \varepsilon. \tag{67}$$

Fix such an R. From the definition of $\delta_\square$, we have

$$\begin{aligned} i\big(\underset{\sim}{\kappa}(H_n) - \underset{\sim}{\kappa}(H_n^R)\big) &\leq i(\underset{\sim}{\kappa} - \underset{\sim}{\kappa}^R) + \delta_\square\big(\underset{\sim}{\kappa}(H_n) - \underset{\sim}{\kappa}(H_n^R), \underset{\sim}{\kappa} - \underset{\sim}{\kappa}^R\big) \\ &\leq \varepsilon + \delta_\square(\underset{\sim}{\kappa}(H_n), \underset{\sim}{\kappa}) = \varepsilon + o(1). \end{aligned}$$

Coupling $\mathcal{H}_n$ and $\mathcal{H}_n^R = \mathcal{H}(H_n^R)$ in the natural way so that the former contains the latter, the expected sum of the sizes of the extra hyperedges in $\mathcal{H}_n$ is at most $ni\big(\underset{\sim}{\kappa}(H_n) - \underset{\sim}{\kappa}(H_n^R)\big) \leq (\varepsilon + o(1))n$. Since adding a clique of size r to a graph G changes the number of vertices in components of size at most k by at most rk, it follows that for k fixed we have $\mathbb{E}|N_k(\mathcal{H}_n) - N_k(\mathcal{H}_n^R)| \le k\varepsilon n + o(n)$, so for n large enough,

$$\mathbb{P}\big(|N_k(\mathcal{H}_n) - N_k(\mathcal{H}_n^R)| \ge k\sqrt{\varepsilon}\big) \le 2\sqrt{\varepsilon}, \tag{68}$$

say. Applying Lemma 3.3 to the sequence (H_n^R), we have $N_k(\mathcal{H}_n^R) = \rho_k(\underset{\sim}{\kappa}^R) + o_{\mathrm{p}}(n)$. Since $\varepsilon > 0$ was arbitrary, the result follows from this, (67) and (68). □

3.4 Proof of Theorem 3.2

We have just seen that for each k we have the 'right' number of vertices of $G(H_n)$ in components of order k; it remains only to show, using the additional assumption of irreducibility, that almost all vertices in large components in fact form a single giant component.

Proof of Theorem 3.2. As usual, Corollary 3.12 implies that there is some $\omega = \omega(n) \to \infty$, which we may take to be $o(n)$, such that

$$N_{\ge\omega}(G(H_n))/n \xrightarrow{\mathrm{p}} \rho(\underset{\sim}{\kappa}). \tag{69}$$

Let $G_n = G(H_n)$. As in the proof of Theorem 1.1, in the light of (69) it suffices to show that $C_1(G_n) \ge \rho(\underset{\sim}{\kappa})n + o_{\mathrm{p}}(n)$. In doing so we may of course assume that $\rho(\underset{\sim}{\kappa}) > 0$.

Fix $\varepsilon > 0$. Theorem 2.12(i) of [5] tells us that as $\gamma \to 0$ we have $\rho((1-\gamma)\underset{\sim}{\kappa}) \nearrow \rho(\underset{\sim}{\kappa})$, so there is some γ with $\rho((1-\gamma)\underset{\sim}{\kappa}) > \rho(\underset{\sim}{\kappa}) - \varepsilon$. In the Poisson multi-hypergraph form, we may write $\mathcal{H}_n = \mathcal{H}(H_n)$ as $\mathcal{H}_n' \cup \mathcal{H}_n''$ where $\mathcal{H}_n' = \mathcal{H}((1-\gamma)H_n)$, $\mathcal{H}_n'' = \mathcal{H}(\gamma H_n)$, and $\mathcal{H}_n'$ and $\mathcal{H}_n''$ are independent.

Writing G_n' for the graph corresponding to $\mathcal{H}_n'$, applying (69) to $(\mathcal{H}_n')$ there is some $\omega = \omega(n) \to \infty$ such that

$$N_{\ge\omega}(G_n') \ge (\rho((1-\gamma)\underset{\sim}{\kappa}) - \varepsilon)n \ge (\rho(\underset{\sim}{\kappa}) - 2\varepsilon)n$$

holds whp. We shall attempt to use the hyperedges of $\mathcal{H}_n''$ to join up the large components of G_n'.

As in [5], the trick is to select one edge from each hyperedge, to obtain a graph. More precisely, let G_n'' be the random multi-graph obtained from $\mathcal{H}_n''$ by replacing each hyperedge E of order r by one of the $\binom{r}{2}$ corresponding edges, chosen uniformly at random. From the Poisson nature of the model, different edges in G_n'' are present independently.

Let $B_n = 2\sum_{r\ge2} A_{n,r}$, where $A_{n,r}$ is the matrix defined by (58). The edge probabilities in G_n'' are given by γ times the entries of B_n. (Note that the coefficient of $A_{n,r}$ is smaller here than in (59), by a factor $1/\binom{r}{2}$, corresponding to choosing one out of $\binom{r}{2}$ edges.)

Let τ be the rescaled edge-kernel defined by

$$\tau(x,y) = 2\sum_{r\ge2}\int_{\mathcal{S}^{r-2}} \kappa_r(x,y,x_3,x_4,\ldots,x_r)\,d\mu(x_3)\cdots d\mu(x_r),$$

i.e., by replacing the factor $r(r-1)$ in (57) by a factor 2. Using (60) and arguing as in the proof of Lemma 3.4, but replacing each appearance of $r(r-1)$ by 2, it is easy to check that $\delta_{\square}(\kappa_{B_n}, \tau) \to 0$; this time, since $2 \le r$, there is no need to truncate the sums over r.

Now $\underset{\sim}{\kappa}$ is irreducible by assumption, which means exactly that κ_{e} is irreducible. Since κ_{e} and τ are non-zero in the same places, it follows that τ is irreducible. Since the graphs G_n'' have the distribution $G(\gamma B_n)$, and $\delta_{\square}(B_n, \tau) \to 0$, Lemma 2.14 tells us that given any two sets X and Y of εn vertices of G_n'', the probability that there is no path in G_n'' from X to Y is exponentially small. As before we may apply this to all partitions of the large components of G_n' into two sets each containing at least εn vertices to deduce that whp we have $C_1(G_n) \ge (\rho(\underset{\sim}{\kappa}) - 3\varepsilon)n$, completing the proof. □

Theorem 3.2 implies a result for branching processes corresponding to Theorem 1.9; we leave the details to the reader.

Finally, let us note that using the trick of selecting one edge from each hyperedge above, it is very easy to extend Theorem 1.3 to the graphs $G(H_n)$ considered in Theorem 3.2.

Acknowledgement. Part of this research was conducted during the programme 'Combinatorics and Statistical Mechanics' at the Isaac Newton Institute, Cambridge; we are grateful to the Institute and to the programme organizers.

References

[1] N. Alon, A note on network reliability, in *Discrete probability and algorithms (Minneapolis, MN, 1993)*, IMA Vol. Math. Appl., 72, Springer, New York, 1995, pp. 11–14.

[2] M. Biskup, L. Chayes and S.A. Smith, Large-deviations/thermodynamic approach to percolation on the complete graph, *Random Structures Algorithms* **31** (2007), 354–370.

[3] B. Bollobás, C. Borgs, J. Chayes and O. Riordan, Percolation on dense graph sequences, *Annals of Probability* **38** (2010), 150–183.

[4] B. Bollobás, S. Janson and O. Riordan, The phase transition in inhomogeneous random graphs, *Random Structures and Algorithms* **31** (2007), 3–122.

[5] B. Bollobás, S. Janson and O. Riordan, Sparse random graphs with clustering, to appear in *Random Structures and Algorithms.* `arXiv:0807.2040.`

[6] B. Bollobás and O. Riordan, Metrics for sparse graphs, in *Surveys in Combinatorics 2009*, London Math. Soc. Lecture Note Series **365**, S. Huczynska, J.D. Mitchell and C.M.Roney-Dougal eds, CUP (2009), pp. 212-287.

[7] B. Bollobás and O. Riordan, Sparse graphs: metrics and random models, to appear in *Random Structures and Algorithms*. `arXiv:0812.2656`.

[8] C. Borgs, J.T. Chayes and L. Lovász, Moments of two-variable functions and the uniqueness of graph limits, to appear in *Geom. Funct. Anal.* `arXiv:0803.1244`.

[9] C. Borgs, J. T. Chayes, L. Lovász, V. T. Sós and K. Vesztergombi, Convergent sequences of dense graphs I: Subgraph frequencies, metric properties and testing, *Advances in Math.* **219** (2008), 1801–1851.

[10] F.R.K. Chung, Diameters and eigenvalues, *J. Amer. Math. Soc.* **2** (1989), 187–196.

[11] F. Chung and R. Graham, Sparse quasi-random graphs, *Combinatorica* **22** (2002), 217–244.

[12] F.R.K. Chung, R.L. Graham and R.M. Wilson, Quasi-random graphs, *Combinatorica* **9** (1989), 345–362.

[13] P. Diaconis and S. Janson, Graph limits and exchangeable random graphs, *Rendiconti di Matematica* **28** (2008), 33–61.

[14] P. Erdős and A. Rényi, On a problem in the theory of graphs, *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **7** (1962), 623–641.

[15] A. Frieze and R. Kannan, Quick approximation to matrices and applications, *Combinatorica* **19** (1999), 175–220.

[16] S. Janson, Standard representation of multivariate functions on a general probability space, *Electron. Commun. Probab.* **14** (2009), 343–346.

[17] S. Janson, Asymptotic equivalence and contiguity of some random graphs, *Random Structures and Algorithms* **36** (2010), 26–45.

[18] S. Janson, T. Łuczak and A. Ruciński, *Random Graphs*. Wiley, New York, 2000.

[19] L. Lovász and B. Szegedy, Limits of dense graph sequences, *J. Combin. Theory B* **96** (2006), 933–957.

[20] A. Lubotzky, R. Phillips and P. Sarnak, Ramanujan graphs, *Combinatorica* **8** (1988), 261–277.

[21] M.J. Luczak and C. McDiarmid, Bisecting sparse random graphs, *Random Structures and Algorithms* **18** (2001), 31–38.

[22] C. McDiarmid, On the method of bounded differences, in *Surveys in combinatorics, 1989*, LMS Lecture Note Series **141**, Cambridge Univ. Press (1989), pp 148–188.

[23] N. O'Connell, Some large deviation results for sparse random graphs, *Probab. Theory Related Fields* **110** (1998), 277–285.

[24] M. Talagrand, Concentration of measure and isoperimetric inequalities in product spaces, *Inst. Hautes Études Sci. Publ. Math.* **81** (1995), 73–205.

[25] A. Thomason, Pseudo-random graphs, in *Proceedings of Random Graphs* (M. Karonski, ed.), Poznań, 1985, *Annals of Discrete Mathematics*, **33** (1987) 307–331.

[26] A. Thomason, Random graphs, strongly regular graphs and pseudorandom graphs, in *Surveys in Combinatorics 1987 (New Cross, 1987)*, London Math. Soc. Lecture Note Ser., **123**, Cambridge Univ. Press, Cambridge (1987), pp 173–195.