Duality in inhomogeneous random graphs, and the cut metric

Svante Janson* Oliver Riordan†

April 26, 2009; slightly revised May 1, 2009

Abstract

The classical random graph model $G(n, \lambda/n)$ satisfies a 'duality principle', in that removing the giant component from a supercritical instance of the model leaves (essentially) a subcritical instance. Such principles have been proved for various models; they are useful since it is often much easier to study the subcritical model than to directly study small components in the supercritical model. Here we prove a duality principle of this type for a very general class of random graphs with independence between the edges, defined by convergence of the matrices of edge probabilities in the cut metric.

1 Introduction and results

Throughout, a matrix denoted A_n is assumed to be symmetric, n-by-n, and to have non-negative entries. Given such a matrix $A_n = (a_{ij})$, let $G(A_n)$ denote the random graph on $[n] = \{1, 2, \dots, n\}$ in which edges are present independently and the probability that ij is an edge is $\min\{a_{ij}/n, 1\}$. If A_n is itself random, then $G(A_n)$ denotes the random graph whose conditional distribution, given A_n, is as above. As shown by Bollobás, Janson and Riordan [7], if the matrices A_n converge (in probability) in a certain sense defined below, then the random graph 'model' $G(A_n)$ may be seen as a generalization of many earlier inhomogeneous models, such as that introduced in [5]. Furthermore, results for $G(A_n)$ generalize corresponding results for percolation on sequences of dense finite graphs of the type proved by Bollobás, Borgs, Chayes and Riordan [4].

It is well known that in the classical random graph $G(n, p)$, $p = \lambda/n$, the small components of the supercritical graph behave like a subcritical instance of the same model; this fact was first exploited by Bollobás [3]. It was also used by Łuczak [13], who stated it explicitly as the 'symmetry rule'; see also [11]. It is also sometimes known as a (discrete) 'duality principle'; see, for example, Alon and Spencer [1]. Corresponding results have been proved for several other

*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

models, for example by Molloy and Reed [14] for the configuration model of Bollobás [2], and by Bollobás, Janson and Riordan [5] for their inhomogeneous model. Our aim in this note is to prove such a result for the very general model $G(A_n)$ described above.

First we need a few definitions, mainly from [7], although many of the important concepts are from earlier papers. Let $(\mathcal{S}, \mu)$ be a measure space with $0 < \mu(\mathcal{S}) < \infty$. Almost all the time, μ will be a probability measure; in fact, 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 $\kappa : \mathcal{S}^2 \to [0, \infty)$. Adapting a definition of Frieze and Kannan [10], for $W \in L^1(\mathcal{S}^2)$ we define the *cut norm* $\|W\|_\square$ of W by

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

(This is equivalent within a factor 4 to the variant where f and g are 0/1-valued functions.) A *rearrangement* of the kernel κ is any kernel $\kappa^{(\tau)}$ defined by

$$\kappa^{(\tau)}(x, y) = \kappa(\tau(x), \tau(y)), \tag{2}$$

where $\tau : \mathcal{S} \to \mathcal{S}$ is a measure-preserving bijection. Given two kernels κ, κ' on $[0, 1]$, the *cut metric* of Borgs, Chayes, Lovász, Sós and Vesztergombi [9] may be defined by

$$\delta_\square(\kappa_1, \kappa_2) = \inf_\tau \|\kappa_1 - \kappa_2^{(\tau)}\|_\square, \tag{3}$$

where the infimum is over all rearrangements of κ_2. (Of course, it makes no difference if we rearrange κ_1 instead, or both κ_1 and κ_2.)

Probabilistically, it is more natural to define $\delta_\square$ via couplings, as discussed in [9]; see also [8]. Given two measure spaces $(\mathcal{S}_1, \mu_1)$, $(\mathcal{S}_2, \mu_2)$ with $0 < \mu_1(\mathcal{S}_1) = \mu_2(\mathcal{S}_2) < \infty$, a *coupling* of these spaces is simply a measure space $(\mathcal{S}, \mu)$ together with measure preserving maps $\sigma_i : \mathcal{S} \to \mathcal{S}_i$, $i = 1, 2$. Given kernels κ_i on $\mathcal{S}_i$, the corresponding *pull-backs* $\kappa_i^{(\sigma_i)}$ are the kernels on $(\mathcal{S}, \mu)$ defined by

$$\kappa_i^{(\sigma_i)}(x, y) = \kappa_i(\sigma_i(x), \sigma_i(y)),$$

and the cut metric may be defined by

$$\delta_\square(\kappa_1, \kappa_2) = \inf \|\kappa_1^{(\sigma_1)} - \kappa_2^{(\sigma_2)}\|_\square,$$

where the infimum is taken over all couplings. It is not obvious that this definition agrees with (3) for kernels on $[0, 1]$, but this turns out to be the case, as shown in [9].

Although the coupling definition is perhaps more natural (and is forced on us if we consider probability spaces with atoms), the rearrangement definition seems intuitively simpler, and is often notationally simpler. Where possible, we shall work with rearrangements rather than couplings. However, we shall still need to consider kernels on different spaces. In this setting a *rearrangement* of a kernel κ on $(\mathcal{S}_1, \mu_1)$ is any kernel $\kappa^{(\tau)}$ on $(\mathcal{S}_2, \mu_2)$, where $\tau : \mathcal{S}_2 \to \mathcal{S}_1$ is a measure preserving bijection and $\kappa^{(\tau)}$ is defined by (2) as before. In fact, for technical reasons it is convenient to allow τ to be a measure-preserving bijection between $\mathcal{S}_2 \setminus N_2$ and $\mathcal{S}_1 \setminus N_1$, where the N_i are null sets: $\mu_i(N_i) = 0$.

Given a symmetric n-by-n matrix A_n, there is a piecewise constant kernel κ_{A_n} on $[0, 1]$ naturally associated to A_n, taking the value a_{ij} on the square $((i-1)/n, i/n] \times ((j-1)/n, j/n]$. (When working with couplings, one can simply view A_n itself as a kernel on a finite space with n points.) We often identify A_n and κ_{A_n}, writing, for example, $\delta_\square(A_n, \kappa)$ for $\delta_\square(\kappa_{A_n}, \kappa)$. Throughout we consider the following random graph 'model': we have a kernel κ on $[0, 1]$ and a sequence A_n of (deterministic or random) matrices with $\delta_\square(A_n, \kappa) \xrightarrow{\mathrm{p}} 0$, and study $G_n = G(A_n)$. We shall show that deleting the giant component from such a graph G_n, when it exists, leaves another instance of the same model. To make sense of this requires some further definitions.

Let T_κ denote the integral operator associated to κ, defined by $(T_\kappa f)(x) = \int_{\mathcal{S}} \kappa(x, y) f(y)\, d\mu(y)$.

Given a kernel κ on a type space $(\mathcal{S}, \mu)$, where μ is a probability measure, let $\mathfrak{X}_\kappa$ be the Poisson Galton–Watson branching process naturally associated to κ: we start with a single particle whose type is distributed according to μ, particles have children independently of each other and of the history, and the types of the children of a particle of type x form a Poisson process on $\mathcal{S}$ with intensity $\kappa(x, y)\, d\mu(y)$. We write $\mathfrak{X}_\kappa(x)$ for the same process started with a single particle of type x.

As in [5], let $\rho(\kappa)$ denote the survival probability of $\mathfrak{X}_\kappa$ and $\rho(\kappa; x)$ that of $\mathfrak{X}_\kappa(x)$. Also, let $\rho_k(\kappa; x)$ and $\rho_k(\kappa)$ denote respectively the probabilities that $\mathfrak{X}_\kappa(x)$ or $\mathfrak{X}_\kappa$ consists of exactly k particles in total.

We now turn to the 'dual' of a kernel κ on a probability space $(\mathcal{S}, \mu)$, giving two versions with slightly different normalization. First, let $\widehat{\kappa}$ be the kernel that is equal to κ as a function, but defined on the space $(\mathcal{S}, \widehat{\mu})$, where $\widehat{\mu}$ is the measure defined by

$$d\widehat{\mu}(x) = (1 - \rho(\kappa; x))\, d\mu(x). \tag{4}$$

Note that $\widehat{\mu}(\mathcal{S}) = 1 - \rho(\kappa)$. Second, to return to a probability space, let $\widehat{\mu}'$ be the normalized measure $\widehat{\mu}/(1 - \rho(\kappa))$, and let $\widehat{\kappa}'$ be the kernel on $(\mathcal{S}, \widehat{\mu}')$ equal to κ as a function. Finally, let $\widetilde{\kappa} = (1 - \rho(\kappa))\widehat{\kappa}'$ be the kernel on $(\mathcal{S}, \widehat{\mu}')$ given by $\widetilde{\kappa}(x, y) = (1 - \rho(\kappa))\kappa(x, y)$. The kernels $\widehat{\kappa}$ and $\widetilde{\kappa}$ are equivalent in a certain natural sense; for example, the operators $T_{\widehat{\kappa}}$ and $T_{\widetilde{\kappa}}$ coincide.

Finally, the kernel κ is *reducible* if there is some $A \subset \mathcal{S}$ with $0 < \mu(A) < 1$ such that κ is zero a.e. on $A \times A^{\mathrm{c}}$, and *irreducible* otherwise.

We write $\mathcal{C}_i(G)$ for the ith largest component of a graph; for definiteness, if there is a tie, we order components of equal sizes according to any fixed ordering on the subsets of $[n]$. Let $\widetilde{G}$ denote the graph formed from G by deleting $\mathcal{C}_1(G)$. Recall from [7] that if $\delta_\square(A_n, \kappa) \xrightarrow{\mathrm{p}} 0$ and κ is irreducible, then

$$|\mathcal{C}_1(G_n)|/n \xrightarrow{\mathrm{p}} \rho(\kappa) \tag{5}$$

and

$$|\mathcal{C}_2(G_n)|/n \xrightarrow{\mathrm{p}} 0, \tag{6}$$

where $G_n = G(A_n)$. Recall also from [5] that $\rho(\kappa) > 0$ if and only if $\|T_\kappa\| > 1$.

Given a (symmetric, n-by-n, non-negative, as always) matrix A_n, let $\widetilde{A}_n$ denote the random $|\widetilde{G}|$-by-$|\widetilde{G}|$ sub-matrix of A_n corresponding to $\widetilde{G}$, where $G = G(A_n)$. More precisely, $\widetilde{A}_n$ may be defined ordering the vertices of $\widetilde{G}$ arbitrarily, and setting $\widetilde{A}_{ij} = a_{vw}$ where v and w are the ith and jth vertices of $\widetilde{G}$.

Our aim in this paper is to prove the following 'duality' result.

Theorem 1.1. *Let (A_n) be a (random or deterministic) sequence of symmetric, non-negative matrices with $\delta_\square(A_n, \kappa) \xrightarrow{\mathrm{p}} 0$ for some irreducible kernel κ on $[0, 1]$. Then $\delta_\square(\widetilde{A}_n, \widehat{\kappa}') \xrightarrow{\mathrm{p}} 0$.*

The main significance is the following consequence.

Theorem 1.2. *Let (A_n) be a sequence of symmetric, non-negative matrices with $\delta_\square(A_n, \kappa) \xrightarrow{\mathrm{p}} 0$ for some irreducible kernel κ with $\rho(\kappa) > 0$, and let $G_n = G(A_n)$. Then there is a random sequence (B_n) of matrices such that $\widetilde{G}_n$ and $G(B_n)$ may be coupled to agree whp, with B_n $m(n)$-by-$m(n)$, $m(n)/n \xrightarrow{\mathrm{p}} 1 - \rho(\kappa)$, and $\delta_\square(B_n, \widetilde{\kappa}) \xrightarrow{\mathrm{p}} 0$,*

Proof. Conditioning on the A_n, we may assume without loss of generality that the A_n are deterministic, with $\delta_\square(A_n, \kappa) \to 0$.

The result is essentially immediate from Theorem 1.1 and the uniqueness of the giant component in G_n. Indeed, we simply take $B_n = \frac{m(n)}{n} \widetilde{A}_n$. Note that $m(n) = n - |\mathcal{C}_1(G_n)|$ satisfies $m(n)/n \xrightarrow{\mathrm{p}} 1 - \rho(\kappa)$ by (5). Since $\delta_\square(\widetilde{A}_n, \widehat{\kappa}') \xrightarrow{\mathrm{p}} 0$ by Theorem 1.1, this implies $\delta_\square(B_n, \widetilde{\kappa}) \xrightarrow{\mathrm{p}} 0$. Note that B_n depends on G_n, but only via the vertex set of $\mathcal{C}_1(G_n)$. Conditioning on this vertex set, we see that the distribution of $\widetilde{G}_n$ is exactly that of $G(B_n)$ conditioned on containing no component larger than $\mathcal{C}_1(G_n)$ (or of the same size but earlier in our fixed order). However, the unconditional probability of $G(B_n)$ containing such a component tends to 0, as otherwise G_n would have positive probability of containing two components of order $\Theta(n)$, contradicting (6). For full details of a related argument see [5, page 79]. $\square$

In turn, Theorem 1.2 implies, for example, that the number of edges in the giant component of $G(A_n)$ is 'what one would expect', i.e., that Theorem 3.5 of [5] extends to this more general setting.

Corollary 1.3. *Let κ be an irreducible kernel, and let $G_n = G(A_n)$, where $\delta_\square(A_n, \kappa) \xrightarrow{\mathrm{p}} 0$. Then*

$$\frac{1}{n} e(\mathcal{C}_1(G_n)) \xrightarrow{\mathrm{p}} \zeta(\kappa),$$

where

$$\zeta(\kappa) := \frac{1}{2} \int_{\mathcal{S}^2} \kappa(x, y) \big(\rho(\kappa; x) + \rho(\kappa; y) - \rho(\kappa; x)\rho(\kappa; y)\big)\, d\mu(x)\, d\mu(y).$$

Proof. As usual, we condition on the A_n assuming that $\delta_\square(A_n, \kappa) \to 0$. Next we eliminate 'large' entries (in particular those exceeding n), as well as any diagonal entries.

If $\delta_\square(A_n, \kappa) \to 0$, then, as shown in [7, Lemma 2.1], there is some $M(n)$ with $M(n)/n \to 0$ such that the sum of the entries of A_n exceeding $M(n)$ is $o(n^2)$. Define A'_n by setting all such entries, and all diagonal entries, to 0. Noting that the sum of the diagonal entries of A_n not exceeding $M(n)$ is at most $nM(n) = o(n^2)$, we have $\delta_\square(A_n, A'_n) \le \|\kappa_{A_n} - \kappa_{A'_n}\|_{L^1} = o(1)$, and in the natural coupling $G(A_n)$ and $G(A'_n)$ agree in all but $o_{\mathrm{p}}(n)$ edges. The expected number of edges in $G(A'_n)$ is simply $n/2$ times $\int \kappa_{A'_n}$. Since the actual number is a sum of independent indicator variables, its variance is at most its mean, and hence $O(n)$. Thus

$$n^{-1}e(G(A_n)) = n^{-1}e(G(A'_n)) + o_{\mathrm{p}}(1) = \frac{1}{2}\int_{\mathcal{S}^2} \kappa_{A'_n} + o_{\mathrm{p}}(1) = \frac{1}{2}\int_{\mathcal{S}^2} \kappa + o_{\mathrm{p}}(1).$$

Applying this result to $\widetilde{G}_n$, which agrees whp with $G(B_n)$, we see that

$$\begin{aligned}
\frac{1}{n}e(\widetilde{G}_n) &= \frac{|\widetilde{G}_n|}{2n}\int_{\mathcal{S}^2} \widetilde{\kappa}(x,y)\, d\widehat{\mu}'(x)\, d\widehat{\mu}'(y) + o_{\mathrm{p}}(1) \\
&= (1-\rho(\kappa))\frac{1}{2}\int_{\mathcal{S}^2} (1-\rho(\kappa))\widehat{\kappa}'(x,y)\, d\widehat{\mu}'(x)\, d\widehat{\mu}'(y) + o_{\mathrm{p}}(1) \\
&= \frac{1}{2}\int_{\mathcal{S}^2} \widehat{\kappa}(x,y)\, d\widehat{\mu}(x)\, d\widehat{\mu}(y) + o_{\mathrm{p}}(1) \\
&= \frac{1}{2}\int_{\mathcal{S}^2} \kappa(x,y)(1-\rho(\kappa;x))(1-\rho(\kappa;y))\, d\mu(x)\, d\mu(y) + o_{\mathrm{p}}(1).
\end{aligned}$$

Subtracting from $e(G_n)$ gives the result. ∎

Theorem 1.2 has more substantial applications, allowing other quantities associated to the small components of a suitable random graph G_n to be studied in a simple way. For one example, concerning susceptibility, see [12]. For another, consider Theorem 3 in [4]. Translated to the present notation, this result concerns the graphs $G_n = G(A_n)$, where the matrices A_n have uniformly bounded entries and $\delta_\square(A_n, \kappa) \to 0$. It makes two statements: (a) when $\|T_\kappa\| < 1$ then $|\mathcal{C}_1(G_n)| \le B \log n$ holds whp for some constant B (depending on κ and the bound on the entries of the A_n) and (b) when $\|T_\kappa\| > 1$ and κ is irreducible, then $|\mathcal{C}_2(G_n)| \le B' \log n$ whp for some B'. The proof of part (a) in [12] is very simple, that of part (b) rather lengthy. Using Theorem 1.2 it is easy to deduce part (b) from part (a); one only needs the simple fact that in this setting, since κ is bounded and hence T_κ is Hilbert–Schmidt, the dual kernel is strictly subcritical; see [5, Theorem 6.7].

Remark 1.4. Theorems 1.1 and 1.2 extend *mutatis mutandis* to the graphs $G(H_n)$ studied in [7, Section 3], which may be seen as the simple graphs underlying random (non-uniform) hypergraphs whose 'hypermatrices' of edge probabilities converge in a suitable sense to a 'hyperkernel', i.e., a sequence of symmetric functions κ_r on $\mathcal{S}^r$, $r = 2, 3, \ldots$. Since the changes needed are very simple, but complicate the notation, we do not give the details. Note that for the analogue of Corollary 1.3, one needs an additional condition, called 'edge integrability' in [7, Remark 3.5], as well as convergence in the corresponding version of the cut metric.

2 Proofs

The main idea is to prove an analogue of [5, Theorem 9.10]. The statement, Theorem 2.5 below, is a little awkward, as we are trying to formulate a result about the 'type' of a vertex in a setting where individual vertices don't really have types.

We start with a much simpler statement concerning branching processes. As in [7] we write $\mathcal{W}$ for the set of all integrable non-negative functions $W : \mathcal{S} \times \mathcal{S} \to [0, \infty)$, and $\mathcal{W}_{\mathrm{sym}}$ for the subset of symmetric functions, i.e., kernels. For $W \in \mathcal{W}$, we write λ_W and λ'_W for the marginals of W with respect to the first and second variables:

$$\lambda_W(x) := \int W(x, y)\, d\mu(y), \qquad \lambda'_W(y) := \int W(x, y)\, d\mu(x).$$

Of course, for $W \in \mathcal{W}_{\mathrm{sym}}$ we have $\lambda_W = \lambda'_W$.

Given a finite graph F with vertex set $\{1, \ldots, r\}$, integrable functions $f_i : \mathcal{S} \to \mathbb{R}$, and $W \in \mathcal{W}_{\mathrm{sym}}$, let

$$t^{\times}_{\mathrm{isol}}(F, (f_i), W) := \int_{\mathcal{S}^r} \prod_{ij \in E(F)} W(x_i, x_j) \prod_{k=1}^{r} f_k(x_k) e^{-\lambda_W(x_k)}\, d\mu(x_1) \cdots d\mu(x_r). \tag{7}$$

Note that this differs from the quantity $t_{\mathrm{isol}}(F, W)$ considered in [7] by the inclusion of the factors $f_k(x_k)$, $k = 1, \ldots, r$.

Lemma 2.1. *Let F be a tree and $f_1, \ldots, f_{|F|}$ bounded functions on $\mathcal{S}$. Then $W \mapsto t^{\times}_{\mathrm{isol}}(F, (f_k), W)$ is a bounded map on $\mathcal{W}_{\mathrm{sym}}$ that is Lipschitz continuous in the cut norm. More specifically, there exists a constant C (depending on F only) such that $|t^{\times}_{\mathrm{isol}}(F, (f_k), W)| \le C \prod_k \|f_k\|_\infty$ for all $W \in \mathcal{W}_{\mathrm{sym}}$, and $|t^{\times}_{\mathrm{isol}}(F, (f_k), W) - t^{\times}_{\mathrm{isol}}(F, (f_k), W')| \le C\|W - W'\|_\square \prod_k \|f_k\|_\infty$ for all $W, W' \in \mathcal{W}_{\mathrm{sym}}$.*

Proof. The proof is a simple extension of [7, Theorem 2.3], so we only outline the differences.

Firstly, writing each f_k as the sum of its positive and negative parts, we may assume without loss of generality that $f_k \ge 0$ for each k. Also, we may rescale so that $\|f_k\|_\infty = 1$ for all k.

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

$$t_0\big(F, (W_{ij})_{ij \in E(F)}\big) := \int_{\mathcal{S}^r} \prod_{ij \in E(F)} W_{ij}(x_i, x_j)\, d\mu(x_1) \cdots d\mu(x_r). \tag{8}$$

Note that we have omitted both the exponential factors and the factors $f_k(x_k)$ from (7). As in [7], given $W \in \mathcal{W}$ let

$$W^{(a,b)}(x,y) := e^{-a\lambda_W(x)} W(x,y) e^{-b\lambda'_W(y)}. \tag{9}$$

Also, let

$$W_{(ij)} := f_i(x)^{1/d_i} W(x,y) f_j(y)^{1/d_j},$$

where d_i is the degree of vertex i in F. It is shown in [7, Lemma 2.4] that the map $W \mapsto W^{(a,b)}$ is Lipschitz continuous with respect to the cut norm, with the constant independent of a and b. Since $\|f_i\|_\infty, \|f_j\|_\infty \le 1$, the linear map $W \mapsto W_{(ij)}$ cannot increase the cut norm, so it and the composition $W \mapsto W_{(ij)}^{(a,b)}$ are Lipschitz continuous. Noting that

$$t^\times_{\mathrm{isol}}(F, (f_k), W) = t_0\big(F, (W_{(ij)}^{(1/d_i, 1/d_j)})_{ij}\big),$$

and that the marginals of $W_{(ij)}^{(a,b)}$ are at most those of $W^{(a,b)}$ and are hence bounded by constants depending only on a and b, the rest of the proof of [7, Theorem 2.3] goes through unchanged. $\square$

Lemma 2.1 corresponds roughly to counting tree components of a given size in a certain random graph by a weight which is a product of the weights of their vertices. In fact, we wish to count *vertices* in such trees by a certain weight, i.e., to count trees by a weight that is the sum of the weights of their vertices.

Given a finite graph F with vertex set $\{1, \dots, r\}$, an integrable function $f : \mathcal{S} \to \mathbb{R}$, and $W \in \mathcal{W}_{\mathrm{sym}}$, let

$$t^+_{\mathrm{isol}}(F, f, W) := \int_{\mathcal{S}^r} \sum_{k=1}^r f(x_k) \prod_{ij \in E(F)} W(x_i, x_j) \prod_{k=1}^r e^{-\lambda_W(x_k)}\, d\mu(x_1)\cdots d\mu(x_r). \tag{10}$$

Lemma 2.2. *Let F be a tree and f a bounded function on $\mathcal{S}$. Then $W \mapsto t^+_{\mathrm{isol}}(F, f, W)$ is a bounded map on $\mathcal{W}_{\mathrm{sym}}$ that is Lipschitz continuous in the cut norm. More specifically, there exists a constant C (depending on F only) such that $|t^+_{\mathrm{isol}}(F, f, W)| \le C\|f\|_\infty$ for all $W \in \mathcal{W}_{\mathrm{sym}}$, and $|t^+_{\mathrm{isol}}(F, f, W) - t^+_{\mathrm{isol}}(F, f, W')| \le C\|W - W'\|_\square \|f\|_\infty$ for all $W, W' \in \mathcal{W}_{\mathrm{sym}}$.*

Proof. Write $t^+_{\mathrm{isol}}(F, f, W)$ as a sum of $|F|$ terms $t^\times_{\mathrm{isol}}(F, (f_k), W)$; in each, one of the f_k is equal to f, and the others are the constant function 1. $\square$

Although we shall not use this, let us note a corollary.

Corollary 2.3. *Let κ_n be a sequence of kernels with $\|\kappa_n - \kappa\|_\square \to 0$. Then for each fixed k we have $\|\rho_k(\kappa_n; \cdot) - \rho_k(\kappa; \cdot)\|_{L^1} \to 0$, and $\|\rho(\kappa_n; \cdot) - \rho(\kappa; \cdot)\|_{L^1} \to 0$.*

Proof. It is not hard to check that for any kernel κ' and any bounded f we have

$$\int_{\mathcal{S}} \rho_k(\kappa'; x) f(x)\, d\mu(x) = \sum_T \frac{1}{\mathrm{aut}(T)} t^+_{\mathrm{isol}}(T, f, \kappa'),$$

where the sum is over all isomorphism classes of trees on k vertices. (This generalizes (43) in [7]; it is perhaps most easily seen by considering a finite random graph associated to κ'.) Lemma 2.2 thus gives

$$\left| \int_{\mathcal{S}} (\rho_k(\kappa_n; x) - \rho_k(\kappa; x)) f(x)\, d\mu(x) \right| \leq C' \|f\|_\infty \|\kappa_n - \kappa\|_\square$$

for some constant C'. Taking $f(x)$ to be the sign of $(\rho_k(\kappa_n; x) - \rho_k(\kappa; x))$, the first statement follows.

Turning to the second statement, first note that, summing over $k' \leq k$, we have

$$\|\rho_{\leq k}(\kappa_n; \cdot) - \rho_{\leq k}(\kappa; \cdot)\|_{L^1} \to 0 \tag{11}$$

for any fixed k. Let

$$\Delta_k(\kappa') := \rho(\kappa') - \rho_{\leq k}(\kappa') = \|\rho(\kappa'; x) - \rho_{\leq k}(\kappa'; x)\|_{L^1}.$$

From (11) and the triangle inequality, for any k we have

$$\limsup_{n\to\infty} \|\rho(\kappa_n; \cdot) - \rho(\kappa; \cdot)\|_{L^1} \leq \Delta_k(\kappa) + \limsup_{n\to\infty} \Delta_k(\kappa_n).$$

With k fixed, from (11) we have $\rho_{\leq k}(\kappa_n) \to \rho_{\leq k}(\kappa)$, so $\limsup_{n\to\infty} \Delta_k(\kappa_n) \leq \Delta_k(\kappa) + \limsup_{n\to\infty} |\rho(\kappa_n) - \rho(\kappa)|$. Theorem 1.9 from [7] tells us that $\rho(\kappa_n) \to \rho(\kappa)$, so this gives $\limsup_{n\to\infty} \Delta_k(\kappa_n) \leq \Delta_k(\kappa)$, and hence

$$\limsup_{n\to\infty} \|\rho(\kappa_n; \cdot) - \rho(\kappa; \cdot)\|_{L^1} \leq 2\Delta_k(\kappa)$$

for any k. Letting $k \to \infty$, noting that $\rho_{\leq k}(\kappa) \nearrow \rho(\kappa)$, we have $\Delta_k(\kappa) \to 0$, and the result follows. $\square$

We now turn to the random graph equivalent of Lemma 2.2, again using methods from [7]. In the sequel we will for convenience take $(\mathcal{S}, \mu)$ to be $[0,1]$ with Lebesgue measure, but we will continue to write $\mathcal{S}$ and μ to emphasize that the results easily extend to general spaces $\mathcal{S}$. Let (A_n) be a sequence of matrices and κ a kernel on $\mathcal{S} = [0,1]$ with $\delta_\square(A_n, \kappa) \to 0$, and let $\kappa'_n = \kappa_{A_n}^{(\tau_n)}$ be a rearrangement of κ_{A_n} chosen so that $\|\kappa'_n - \kappa\|_\square \to 0$. We write

$$S_v = S_{v,n} = \tau_n^{-1}((v-1)/n, v/n]) \tag{12}$$

for the subset of $\mathcal{S}$ corresponding to the vertex v under this rearrangement. Given a sequence f_n of integrable functions on $\mathcal{S}$, for $v \in V(G_n)$ set

$$f_n(v) = n \int_{S_{v,n}} f_n(x)\, d\mu(x), \tag{13}$$

so $f_n(v)$ is the average of f_n over S_v. Note that if κ is finite type, f_n depends only on the type, and the rearrangement τ_n^{-1} maps each vertex into a single type, then $f_n(v)$ is simply f_n evaluated at the type of v.

Lemma 2.4. *With the definitions above, if the functions f_n are uniformly integrable, then for each fixed k we have*

$$\left| \frac{1}{n} \sum_{v \,:\, |C_v|=k} f_n(v) - \int_{\mathcal{S}} f_n(x)\rho_k(\kappa; x)\, d\mu(x) \right| \xrightarrow{\mathrm{p}} 0,$$

where C_v is the component of $G_n = G(A_n)$ containing the vertex v.

(For reader who prefers to define the cut metric via couplings, the corresponding formulation of this lemma concerns functions f_n defined on the spaces on which the kernels κ and κ_{A_n} are coupled.)

Proof. We claim that it suffices to consider the case where the f_n are uniformly bounded. Indeed, given any $\varepsilon > 0$, we may find uniformly bounded approximations f_n' to f_n with $\|f_n' - f_n\|_{L^1} \le \varepsilon$ for every n. Applying the uniformly bounded case, and then letting $\varepsilon \to 0$, the result follows. Rescaling, we may and shall assume that $\|f_n\|_\infty \le 1$ for all n.

Using Lemma 2.2 in place of [7, Theorem 2.3], the proof is now essentially the same as that of [7, Lemma 2.11], *mutatis mutandis*. We only outline the changes. Let

$$X_n = X_n(G_n) := \frac{1}{n} \sum_{v \,:\, |C_v|=k} f_n(v).$$

Adding or deleting an edge of G_n changes X_n by at most $2k\|f_n\|_\infty/n \le 2k/n$. It follows that, arguing as in the proof of [7, Lemma 2.8], we may assume that the matrices A_n are *well behaved*, meaning that all diagonal entries are zero, and the maximum entry of A_n is $o(n)$ as $n \to \infty$. As in [7], we may then switch to the Poisson multigraph version of $G(A_n)$; we omit the details. Using [7, Lemma 2.10], the contribution to X_n from components C_v that contain cycles is then $o_{\mathrm{p}}(1)$. On the other hand, the contribution from components isomorphic to some particular tree T has expectation

$$o(1) + (1 + o(1)) \frac{t^+_{\mathrm{isol}}(T, f_n, \kappa_n')}{\mathrm{aut}(T)};$$

the argument is as for the corresponding relation (40) in [7]. Continuing as in [7], but using Lemma 2.2, it follows that $|\mathbb{E}X_n - a_n| \to 0$, where $a_n = \int_{\mathcal{S}} f_n(x)\rho_k(\kappa; x)\, d\mu(x)$. Considering sums over pairs of disjoint components, one obtains $|\mathbb{E}X_n^2 - a_n^2| \to 0$, giving $|X_n - a_n| \xrightarrow{\mathrm{p}} 0$ as claimed. □

The corresponding result for the giant component is an immediate consequence; this is the natural analogue of [5, Theorem 9.10] in the present context.

Theorem 2.5. *Let (A_n) be a (deterministic or random) sequence of matrices and κ an irreducible kernel on $[0,1]$ with $\delta_\square(A_n, \kappa) \xrightarrow{\mathrm{p}} 0$, let $\kappa_n' = \kappa_{A_n}^{(\tau_n)}$ be a* *(random) rearrangement of κ_{A_n} chosen so that $\|\kappa'_n - \kappa\|_\square \xrightarrow{\mathrm{p}} 0$, and let f_n be a uniformly integrable sequence of functions $f_n : [0,1] \to \mathbb{R}$. Then*

$$\left| \frac{1}{n} \sum_{v \in \mathcal{C}_1(G_n)} f_n(v) - \int_{\mathcal{S}} f_n(x) \rho(\kappa; x)\, d\mu(x) \right| \xrightarrow{\mathrm{p}} 0,$$

where $G_n = G(A_n)$, and $f_n(v)$ is defined by (12) *and* (13).

Proof. As usual, by conditioning on the sequence (A_n) (and now also on the τ_n), we may assume that the A_n are deterministic and $\|\kappa'_n - \kappa\|_\square \to 0$.

Lemma 2.4 extends immediately to a corresponding result summing over all components of size $1 \le k \le K$ for any fixed K, and hence for $K = K(n) \to \infty$ sufficiently slowly. But the results of [7] show that only $o_{\mathrm{p}}(n)$ vertices in components of size more than $K(n)$ are not in $\mathcal{C}_1$, and conversely, trivially, at most $K(n) = o(n)$ vertices of $\mathcal{C}_1$ are not in such components, so we obtain

$$\left| \frac{1}{n} \sum_{v \notin \mathcal{C}_1(G_n)} f_n(v) - \int_{\mathcal{S}} f_n(x) (1 - \rho(\kappa; x))\, d\mu(x) \right| \xrightarrow{\mathrm{p}} 0.$$

It remains only to note that

$$\frac{1}{n} \sum_{v \in V(G_n)} f_n(v) = \int_{\mathcal{S}} f_n(x)\, d\mu(x)$$

by definition of $f_n(v)$. ☐

We shall need the following simple observation concerning the cut norm. In this we write $\|\kappa\|_{\square,\mu}$ for the cut norm of κ defined with respect to a measure μ.

Lemma 2.6. *Let κ be a kernel on a measure space $(\mathcal{S}, \mu)$ with $0 < \mu(\mathcal{S}) < \infty$, and let h be a non-negative measurable function on $(\mathcal{S}, \mu)$. Let ν be the measure defined by $d\nu(x) = h(x)\, d\mu(x)$. Then*

$$\|\kappa\|_{\square,\nu} \le \|h\|_\infty^2 \|\kappa\|_{\square,\mu}.$$

Proof. Essentially immediate from (1). Indeed, for any f, g with $\|f\|_\infty, \|g\|_\infty \le 1$,

$$\begin{aligned} \Big| \int_{\mathcal{S}^2} & f(x) \kappa(x,y) g(y)\, d\nu(x)\, d\nu(y) \Big| \\ &= \Big| \int_{\mathcal{S}^2} f(x) h(x) \kappa(x,y) g(y) h(y)\, d\mu(x)\, d\mu(y) \Big| \\ &\qquad = \|h\|_\infty^2 \Big| \int_{\mathcal{S}^2} \tilde{f}(x) \kappa(x,y) \tilde{g}(y)\, d\mu(x)\, d\mu(y) \Big|, \end{aligned}$$

where $\tilde{f}(x) = f(x)h(x)/\|h\|_\infty$ has $\|\tilde{f}\|_\infty \le 1$, and similarly for $\tilde{g}$. The final integral is bounded by $\|\kappa\|_{\square,\mu}$ by definition. ☐

Using Theorem 2.5, it is not hard to prove Theorem 1.1.

Proof of Theorem 1.1. Given a kernel κ and real number δ, let

$$m_\delta(\kappa) = \sup_{\mu(A)\le\delta} \int_{A\times\mathcal{S}} \kappa(x,y)\,d\mu(x)\,d\mu(y),$$

so $m_\delta(\kappa)$ is the integral of the marginal of κ over the set with measure δ where this marginal is maximal. Note that

$$|m_\delta(\kappa_1) - m_\delta(\kappa_2)| \le \delta_\square(\kappa_1,\kappa_2).$$

Also, if κ is integrable, then $m_\delta(\kappa)\to 0$ as $\delta\to 0$.

Suppose now that κ is irreducible, and, conditioning as usual, that the A_n are deterministic with $\delta_\square(A_n,\kappa)\to 0$. Suppose also that $\rho(\kappa)>0$; otherwise, $\widehat{\kappa}'=\kappa$, while by (5) the matrices $\widetilde{A}_n$ are obtained from A_n by deleting $o_{\mathrm{p}}(n)$ rows and columns, and the result follows easily.

Let $\kappa'_n = \kappa_n^{(\tau_n)}$ be a rearrangement of κ_{A_n} chosen so that $\|\kappa'_n-\kappa\|_\square\to 0$. As before, let $S_v = S_{v,n} = \tau_n^{-1}((v-1)/n, v/n])$ be the subset of $\mathcal{S}$ corresponding to a vertex v of G_n under the rearrangement τ_n.

Let ν_n be the *random* measure that agrees with Lebesgue measure μ on each S_v, $v\notin\mathcal{C}_1$, and is zero otherwise. Noting that $\nu_n(\mathcal{S}) = (1-|\mathcal{C}_1|/n) \xrightarrow{\mathrm{p}} \widehat{\mu}(\mathcal{S}) = 1-\rho(\kappa)$, let $\nu'_n = \nu_n/\nu_n(\mathcal{S})$ be the rescaled version of ν_n.

Although it may appear that we have done our best to disguise this fact, the kernel κ'_n on the measure space $(\mathcal{S},\nu'_n)$ is simply a rearrangement of the kernel $\kappa_{\widetilde{A}_n}$, where $\widetilde{A}_n$ is the submatrix of A_n obtained by deleting rows and columns corresponding to vertices in $\mathcal{C}_1(G_n)$. Since $\delta_\square$ is unchanged by rearrangement, indicating now the measure on the space (always $\mathcal{S}$) on which our kernels are defined, our aim is exactly to show that

$$\delta_\square((\kappa'_n,\nu'_n),(\kappa,\widehat{\mu}')) \xrightarrow{\mathrm{p}} 0. \tag{14}$$

Fix $\varepsilon>0$. From the comments at the start of the proof there is some δ such that $m_\delta(\kappa)<\varepsilon/2$, and then

$$m_\delta(\kappa'_n) < \varepsilon \tag{15}$$

for n large enough.

Let κ_f be a finite-type kernel approximating κ within ε in the L^1 norm, and hence in $\delta_\square$:

$$\|\kappa_f-\kappa\|_\square \le \|\kappa_f-\kappa\|_{L^1} \le \varepsilon, \tag{16}$$

with κ_f constant on the sets $A_i\times A_j$ for some partition $A_1,\dots,A_r$ of $\mathcal{S}$ into measurable sets.

Fix (for the moment) $1\le i\le r$. Applying Theorem 2.5 with every f_n equal to the indicator function of A_i, we see that

$$\sum_{v\in\mathcal{C}_1} \mu(S_v\cap A_i) \xrightarrow{\mathrm{p}} \int_{A_i} \rho(\kappa;x)\,d\mu(x).$$

Let

$$\nu_{n,i} = \sum_{v \notin \mathcal{C}_1} \mu(S_v \cap A_i) = \mu(A_i) - \sum_{v \in \mathcal{C}_1} \mu(S_v \cap A_i).$$

Then, recalling (4), we have

$$\nu_{n,i} \xrightarrow{\mathrm{p}} \int_{A_i} (1 - \rho(\kappa; x))\, d\mu(x) = \widehat{\mu}(A_i)$$

for each i, and hence

$$\Delta := \sum_{i=1}^{r} |\nu_{n,i} - \widehat{\mu}(A_i)| \xrightarrow{\mathrm{p}} 0.$$

Since our aim is to prove an 'in probability' result, coupling appropriately, we may condition on the random graphs G_n, and assume that $\Delta \to 0$. Note that all quantities we consider are now deterministic.

Recall that $\nu'_n = \nu_n/\nu_n(\mathcal{S})$ is the rescaled version of ν_n, and $\widehat{\mu}' = \widehat{\mu}/(1 - \rho(\kappa)) = \widehat{\mu}/\widehat{\mu}(\mathcal{S})$ is the rescaled version of $\widehat{\mu}$. Since $\nu_n(A_i) = \nu_{n,i}$, it follows that

$$\Delta' := \sum_{i=1}^{r} |\nu'_n(A_i) - \widehat{\mu}'(A_i)| \to 0.$$

Let ν^*_n be obtained by 'tweaking' ν'_n so that $\nu^*_n(A_i) = \widehat{\mu}'(A_i)$ for every i. More precisely, recalling that the unnormalized measure ν_n had a 0/1-valued density function $f(x) = d\nu_n/\,d\mu$, we change f on a set of measure Δ to obtain a 0/1-valued f' with $\int_{A_i} f'(x)\, d\mu(x) = \widehat{\mu}'(A_i)$, and use f' to define the (normalized) measure ν^*_n. Since the normalizing factors are bounded (in the limit), it is not hard to check that for some constant C we have

$$\delta_\Box((\kappa'_n, \nu'_n), (\kappa'_n, \nu^*_n)) \le 2m_{C\Delta}(\kappa'_n)$$

for all large enough n. Indeed, we may couple the measures ν'_n and ν^*_n to agree with probability at least $1 - C\Delta$. Alternatively, we may rearrange the kernels to differ only where one or both coordinates fall into some set of measure at most $C\Delta$. (If we had $\int f = \int f'$, we could take this set to be simply the set where f and f' differ.)

Since $\Delta \to 0$, (15) shows that the right hand side above is $O(\varepsilon)$. Using κ_f, it is now easy to complete the proof. Note that $\|\kappa'_n - \kappa_f\|_\Box \le \|\kappa'_n - \kappa\|_\Box + \|\kappa - \kappa_f\|_\Box \le \varepsilon + o(1)$, from (16) and our convergence assumption. By Lemma 2.6, we thus have

$$\delta_\Box((\kappa'_n, \nu^*_n), (\kappa_f, \nu^*_n)) \le O(\varepsilon) + o(1).$$

Since κ_f is constant on each set $A_i \times A_j$, the kernel κ_f 'only cares how much measure falls in each A_i', and we have

$$\delta_\Box((\kappa_f, \nu^*_n), (\kappa_f, \widehat{\mu}')) = 0.$$

But by Lemma 2.6 again,

$$\delta_\Box((\kappa_f, \widehat{\mu}'), (\kappa, \widehat{\mu}')) \le \|\kappa_f - \kappa\|_{\Box, \widehat{\mu}'} \le \|\kappa_f - \kappa\|_\Box \le \varepsilon.$$

Putting the last four displayed inequalities together and using the triangle inequality, (14) follows. □

Acknowledgement. This research was carried out during a visit of both authors to the programme "Discrete Probability" at Institut Mittag-Leffler, Djursholm, Sweden, 2009.

References

[1] N. Alon and J.H. Spencer, *The Probabilistic Method*, 2nd. ed. Wiley, New York, 2000.

[2] B. Bollobás, A probabilistic proof of an asymptotic formula for the number of labelled regular graphs, *European J. Combinatorics* **1** (1980), 311–316.

[3] B. Bollobás, The evolution of random graphs, *Trans. Amer. Math. Soc.* **286** (1984), 257–274.

[4] B. Bollobás, C. Borgs, J. Chayes and O. Riordan, Percolation on dense graph sequences, to appear in *Annals of Probability*. `arXiv:0701346.`

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

[6] B. Bollobás, S. Janson and O. Riordan, Sparse random graphs with clustering, preprint (2008). `arXiv:0807.2040.`

[7] B. Bollobás, S. Janson and O. Riordan, The cut metric, random graphs, and branching processes, preprint (2009). `arXiv:0901.2091.`

[8] B. Bollobás and O. Riordan, Metrics for sparse graphs. Preprint (2007). `arXiv:0708.1919.`

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

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

[12] S. Janson and O. Riordan, Susceptibility in inhomogeneous random graphs. `arXiv:0905.0437.`

[13] T. Łuczak, Component behavior near the critical point of the random graph process, *Random Struct. Alg.* **1** (1990), 287–310.

[14] M. Molloy and B. Reed, The size of the giant component of a random graph with a given degree sequence, *Combin. Probab. Comput.* **7** (1998), 295–305.