Combinatorics, Probability and Computing (2014) 23, 331–366. © Cambridge University Press 2014
doi:10.1017/S0963548314000017

Local Limit Theorems for the Giant Component of Random Hypergraphs[†]

MICHAEL BEHRISCH[1§], AMIN COJA-OGHLAN[2¶] and MIHYUN KANG[3‖]

[1]Institute of Transportation Systems, German Aerospace Center, Rutherfordstrasse 2, 12489 Berlin, Germany
(e-mail: michael.behrisch@dlr.de)

[2]Goethe University, Mathematics Institute, 60054 Frankfurt am Main, Germany
(e-mail: acoghlan@math.uni-frankfurt.de)

[3]TU Graz, Institut für Optimierung und Diskrete Mathematik (Math B), Steyrergasse 30, 8010 Graz, Austria
(e-mail: kang@math.tugraz.at)

Received 9 January 2009; revised 16 December 2013; first published online 13 February 2014

Let $H_d(n,p)$ signify a random d-uniform hypergraph with n vertices in which each of the $\binom{n}{d}$ possible edges is present with probability $p = p(n)$ independently, and let $H_d(n,m)$ denote a uniformly distributed d-uniform hypergraph with n vertices and m edges. We derive local limit theorems for the joint distribution of the number of vertices and the number of edges in the largest component of $H_d(n,p)$ and $H_d(n,m)$ in the regime $(d-1)\binom{n-1}{d-1}p > 1+\varepsilon$, resp. $d(d-1)m/n > 1+\varepsilon$, where $\varepsilon > 0$ is arbitrarily small but fixed as $n \to \infty$. The proofs are based on a purely probabilistic approach.

2010 *Mathematics subject classification*: Primary 05C80
Secondary 05C65

1. Introduction and results

1.1. The phase transition and the giant component

This paper deals with the connected components of random graphs and hypergraphs. Recall that a *d-uniform hypergraph* H is a set $V(H)$ of *vertices* together with a set $E(H)$ of edges $e \subset V(H)$ of size $|e| = d$. The *order* of H is the number of vertices of H, and the *size* of H is the number of edges. A 2-uniform hypergraph is just a graph.

† An extended abstract version of this work appeared in the proceedings of *RANDOM 2007*, Vol. 4627 of *Lecture Notes in Computer Science*, Springer, pp. 341–352.
§ Supported by the DFG Research Center MATHEON in Berlin.
¶ Supported by DFG CO 646.
‖ Supported by the Deutsche Forschungsgemeinschaft (DFG Pr 296/7-3, KA 2748/3-1)

We say that a vertex $v \in V(H)$ is *reachable* from $w \in V(H)$ if there exist edges $e_1, \ldots, e_k \in E(H)$ such that $v \in e_1$, $w \in e_k$ and $e_i \cap e_{i+1} \neq \emptyset$ for all $1 \leqslant i < k$. Reachability is an equivalence relation, and the equivalence classes are called the *components* of H. If H has only a single component, then H is *connected*. We let $\mathcal{N}(H)$ signify the maximum order (*i.e.*, the number of vertices) of a component of H. For all hypergraphs H that we deal with, the vertex set $V(H)$ will consist of integers. Therefore, the subsets of $V(H)$ can be ordered lexicographically, and we call the lexicographically first component of H that has order $\mathcal{N}(H)$ the *largest component* of H. In addition, we denote by $\mathcal{M}(H)$ the size (*i.e.*, the number of edges) of the largest component.

We will consider two models of random d-uniform hypergraphs. The random hypergraph $H_d(n,p)$ has the vertex set $V = \{1,\ldots,n\}$, and each of the $\binom{n}{d}$ possible edges is present with probability p independently. Moreover, $H_d(n,m)$ is a uniformly distributed d-uniform hypergraph with vertex set $V = \{1,\ldots,n\}$ and with exactly m edges. In the case $d = 2$, the notation $G(n,p) = H_2(n,p)$, $G(n,m) = H_2(n,m)$ is common. Finally, we say that the random hypergraph $H_d(n,p)$ enjoys a certain property $\mathcal{P}$ *with high probability* ('w.h.p.') if the probability that $\mathcal{P}$ holds in $H_d(n,p)$ tends to 1 as $n \to \infty$; a similar terminology is used for $H_d(n,m)$.

Since the pioneering work of Erdős and Rényi [15, 16] on the evolution of $G(n,m)$, the component structure of random discrete objects (*e.g.*, graphs, hypergraphs, digraphs) has been among the main subjects of probabilistic combinatorics. One reason for the relevance of this subject is the connection to statistical physics and percolation ('mean field models'). Another reason is the impact on computer science (*e.g.*, in the study of complex networks or computational problems such as Max Cut or Max 2-Sat [14]).

More precisely, Erdős and Rényi [16] studied (among other things) the component structure of *sparse* random graphs with $O(n)$ edges. The main result is that the order $\mathcal{N}(G(n,m))$ of the largest component undergoes a *phase transition* as $2m/n \sim 1$. Let us state a more general version from Schmidt-Pruzan and Shamir [28], which covers d-uniform hypergraphs. Let either $H = H_d(n,m)$ and $c = dm/n$, or $H = H_d(n,p)$ and $c = \binom{n-1}{d-1}p$; we refer to c as the *average degree* of H. Then the result is the following.

(i) If $c < (d-1)^{-1} - \varepsilon$ for an arbitrarily small but fixed $\varepsilon > 0$, then $\mathcal{N}(H) = O(\ln n)$ w.h.p.

(ii) By contrast, if $c > (d-1)^{-1} + \varepsilon$, then H features a unique component of order $\Omega(n)$ w.h.p., which is called the *giant component*. More precisely, $\mathcal{N}(H) = (1-\rho)n + o(n)$ w.h.p., where ρ is the unique solution to the transcendental equation

$$\rho = \exp(c(\rho^{d-1} - 1)) \tag{1.1}$$

that lies strictly between 0 and 1. Furthermore, the second largest component has order $O(\ln n)$ w.h.p.

In this paper we present a new, purely probabilistic approach for investigating the precise limiting behaviour of the order and size of the largest component of sparse random graphs and, more generally, hypergraphs. We obtain *local limit theorems* for the joint distribution of the order and size of the largest component of $H = H_d(n,p)$ or $H = H_d(n,m)$ (Theorems 1.1 and 1.3). Whereas in the case of *graphs* (*i.e.*, $d = 2$) these results are either known or can be derived from prior work (in particular, Bender, Canfield and McKay [8]), all our results are new for d-uniform hypergraphs with $d > 2$. Besides, we believe that our probabilistic approach is interesting in the case of graphs as well, because we completely avoid the use of involved enumerative methods.

1.2. Main results

Our first result provides the *local limit theorem* for the joint distribution of $\mathcal{N}(H_d(n,p))$ and $\mathcal{M}(H_d(n,p))$.

Theorem 1.1. *Let $d \geqslant 2$ be a fixed integer. For any two compact sets $\mathcal{I} \subset \mathbb{R}^2$, $\mathcal{J} \subset ((d-1)^{-1}, \infty)$, and for any $\delta > 0$, there exists $n_0 > 0$ such that the following holds. Let $p = p(n)$ be a sequence such that $c = c(n) = \binom{n-1}{d-1}p \in \mathcal{J}$ for all n, and let $0 < \rho = \rho(n) < 1$ be the unique solution to* (1.1). *Further, let*

$$\sigma_{\mathcal{N}}^2 = \frac{\rho(1-\rho+c(d-1)(\rho-\rho^{d-1}))}{(1-c(d-1)\rho^{d-1})^2} \cdot n, \tag{1.2}$$

$$\sigma_{\mathcal{M}}^2 = c^2\rho^d \cdot \frac{2+c(d-1)(\rho^{2d-2}-2\rho^{d-1}+\rho^d)-\rho^{d-1}-\rho^d}{(1-c(d-1)\rho^{d-1})^2} \cdot n + (1-\rho^d)\frac{c}{d} \cdot n,$$

$$\sigma_{\mathcal{N}\mathcal{M}} = c\rho \cdot \frac{1-\rho^d-c(d-1)\rho^{d-1}(1-\rho)}{(1-c(d-1)\rho^{d-1})^2} \cdot n.$$

Suppose that $n \geqslant n_0$ and that ν, μ are integers such that

$$x = \nu - (1-\rho)n \quad \textit{and} \quad y = \mu - (1-\rho^d)\binom{n}{d}p$$

satisfy $n^{-1/2}(x,y) \in \mathcal{I}$. Then, letting

$$P(x,y) = \frac{1}{2\pi\sqrt{\sigma_{\mathcal{N}}^2\sigma_{\mathcal{M}}^2 - \sigma_{\mathcal{N}\mathcal{M}}^2}} \cdot \exp\left[-\frac{\sigma_{\mathcal{N}}^2\sigma_{\mathcal{M}}^2}{2(\sigma_{\mathcal{N}}^2\sigma_{\mathcal{M}}^2 - \sigma_{\mathcal{N}\mathcal{M}}^2)}\left(\frac{x^2}{\sigma_{\mathcal{N}}^2} - \frac{2\sigma_{\mathcal{N}\mathcal{M}}xy}{\sigma_{\mathcal{N}}^2\sigma_{\mathcal{M}}^2} + \frac{y^2}{\sigma_{\mathcal{M}}^2}\right)\right], \tag{1.3}$$

we have

$$(1-\delta)P(x,y) \leqslant \mathbb{P}\left[\mathcal{N}(H_d(n,p)) = \nu,\ \mathcal{M}(H_d(n,p)) = \mu\right] \leqslant (1+\delta)P(x,y).$$

Theorem 1.1 characterizes the joint limiting distribution of $\mathcal{N}(H_d(n,p))$ and $\mathcal{M}(H_d(n,p))$ *precisely*, because it actually yields the asymptotic probability that $\mathcal{N}$ and $\mathcal{M}$ attain any two values $\nu = (1-\rho)n + x$, $\mu = (1-\rho^d)\binom{n}{d}p + y$. Namely, the theorem shows that

$$\mathbb{P}\left[\mathcal{N}(H_d(n,p)) = \nu,\ \mathcal{M}(H_d(n,p)) = \mu\right] \sim P(x,y) \tag{1.4}$$

uniformly for average degrees $c = \binom{n-1}{d-1}p \in \mathcal{J}$ and deviations (x,y) such that $n^{-1/2}(x,y) \in \mathcal{I}$. Hence, the average degree c is assumed to be bounded and also bounded away from $(d-1)^{-1}$. We emphasize that $P(x,y)$ is of order n^{-1} as $n \to \infty$, as $\sigma_{\mathcal{N}}^2, \sigma_{\mathcal{M}}^2, \sigma_{\mathcal{N}\mathcal{M}}$ are of order n. Since $P(x,y)$ is the density function of a bivariate normal distribution, Theorem 1.1 readily yields the following 'macroscopic' *central limit theorem.*

Corollary 1.2. *With the notation and the assumptions of Theorem 1.1, suppose that the limit*

$$\Xi = \lim_{n\to\infty} \frac{\sigma_{\mathcal{NM}}}{\sigma_{\mathcal{N}}\sigma_{\mathcal{M}}}$$

exists. Then the joint distribution of

$$\frac{\mathcal{N}(H_d(n,p)) - (1-\rho)n}{\sigma_{\mathcal{N}}} \quad \text{and} \quad \frac{\mathcal{M}(H_d(n,p)) - (1-\rho^d)\binom{n}{d}p}{\sigma_{\mathcal{M}}}$$

converges in distribution to the bivariate normal distribution with mean 0 *and covariance matrix*

$$\begin{pmatrix} 1 & \Xi \\ \Xi & 1 \end{pmatrix}.$$

Let us stress that Theorem 1.1 is significantly stronger than Corollary 1.2, since the latter result yields the asymptotic probability that

$$x\sigma_{\mathcal{N}} \leqslant \mathcal{N}(H_d(n,p)) - (1-\rho)n \leqslant x'\sigma_{\mathcal{N}} \quad \text{and} \tag{1.5}$$

$$y\sigma_{\mathcal{M}} \leqslant \mathcal{M}(H_d(n,p)) - (1-\rho^d)\binom{n}{d}p \leqslant y'\sigma_{\mathcal{M}} \tag{1.6}$$

for any *fixed* $x, x', y, y' \in \mathbb{R}$ with $x < x', y < y'$. Note that $\sigma_{\mathcal{M}}, \sigma_{\mathcal{N}} = \Theta(\sqrt{n})$. Therefore, for any fixed $\delta > 0$, setting $x' = x + \delta\sigma_{\mathcal{N}}$ and $y' = y + \delta\sigma_{\mathcal{M}}$, we could use (1.5)–(1.6) to determine the asymptotic probability that

$$|\mathcal{N}(H_d(n,p)) - (1-\rho)n - x\sigma_{\mathcal{N}}| < \delta\sqrt{n} \quad \text{and}$$

$$\left|\mathcal{M}(H_d(n,p)) - (1-\rho^d)\binom{n}{d}p - y\sigma_{\mathcal{M}}\right| < \delta\sqrt{n}.$$

However, (1.5)–(1.6) do not yield the probability that $\mathcal{N}(H_d(n,p))$ and $\mathcal{M}(H_d(n,p))$ hit certain values ν, μ *exactly*, in contrast to (1.4).

The second main result is a local limit theorem for $\mathcal{N}(H_d(n,m))$ and $\mathcal{M}(H_d(n,m))$.

Theorem 1.3. *Let $d \geqslant 2$ be a fixed integer. For any two compact sets $\mathcal{I} \subset \mathbb{R}^2$, $\mathcal{J} \subset ((d-1)^{-1}, \infty)$, and for any $\delta > 0$, there exists $n_0 > 0$ such that the following holds. Let $m = m(n)$ be a sequence of integers such that $c = c(n) = dm/n \in \mathcal{J}$ for all n, and let $0 < \rho = \rho(n) < 1$ be the unique solution to* (1.1)*. Further, let*

$$\tau_{\mathcal{N}}^2 = \rho n \cdot \frac{1 - (c+1)\rho - c(d-1)\rho^{d-1} + 2cd\rho^d - cd\rho^{2d-1}}{(1 - c(d-1)\rho^{d-1})^2},$$

$$\tau_{\mathcal{M}}^2 = \frac{c\rho^d n}{d(1 - c(d-1)\rho^{d-1})^2} \cdot \big[1 - c(d-2)\rho^{d-1} - (c^2 d - cd + 1)\rho^d - c^2(d-1)\rho^{2d-2} + 2c(cd-1)\rho^{2d-1} - c^2\rho^{3d-2}\big],$$

$$\tau_{\mathcal{NM}} = c\rho^d n \cdot \frac{1 - c\rho - c(d-1)\rho^{d-1} + (c + cd - 1)\rho^d - c\rho^{2d-1}}{(1 - c(d-1)\rho^{d-1})^2}.$$

https://doi.org/10.1017/S0963548314000017 Published online by Cambridge University Press

Suppose that $n \geqslant n_0$ and that ν, μ are integers such that $x = \nu - (1-\rho)n$ and $y = \mu - (1 - \rho^d)m$ satisfy $n^{-1/2}(x, y) \in \mathcal{I}$. Then, letting

$$Q(x,y) = \frac{1}{2\pi\sqrt{\tau_{\mathcal{N}}^2\tau_{\mathcal{M}}^2 - \tau_{\mathcal{N}\mathcal{M}}^2}} \exp\left[-\frac{\tau_{\mathcal{N}}^2\tau_{\mathcal{M}}^2}{2(\tau_{\mathcal{N}}^2\tau_{\mathcal{M}}^2 - \tau_{\mathcal{N}\mathcal{M}}^2)}\left(\frac{x^2}{\tau_{\mathcal{N}}^2} - \frac{2\tau_{\mathcal{N}\mathcal{M}}xy}{\tau_{\mathcal{N}}^2\tau_{\mathcal{M}}^2} + \frac{y^2}{\tau_{\mathcal{M}}^2}\right)\right],$$

we have

$$(1-\delta)Q(x,y) \leqslant \mathbb{P}\big[\mathcal{N}(H_d(n,m)) = \nu,\ \mathcal{M}(H_d(n,m)) = \mu\big] \leqslant (1+\delta)Q(x,y).$$

Again, $Q(x, y)$ is the density function of a bivariate normal distribution and hence Theorem 1.3 yields the following central limit theorem.

Corollary 1.4. *With the notation and the assumptions of Theorem 1.3, suppose that the limit*

$$\Xi = \lim_{n\to\infty} \frac{\tau_{\mathcal{N}\mathcal{M}}}{\tau_{\mathcal{N}}\tau_{\mathcal{M}}}$$

exists. Then the joint distribution of

$$\frac{\mathcal{N}(H_d(n,m)) - (1-\rho)n}{\tau_{\mathcal{N}}} \quad \textit{and} \quad \frac{\mathcal{M}(H_d(n,m)) - (1-\rho^d)m}{\tau_{\mathcal{M}}}$$

converges in distribution to the bivariate normal distribution with mean 0 *and covariance matrix*

$$\begin{pmatrix} 1 & \Xi \\ \Xi & 1 \end{pmatrix}.$$

1.3. Related work

1.3.1. Graphs. Bender, Canfield and McKay [8] were the first to compute the asymptotic probability that a random graph $G(n,m)$ is connected for *any* ratio m/n. Although they employ a probabilistic result from Łuczak [21] to simplify their arguments, their proof is based on enumerative methods. In addition, using their formula for the probability of $G(n,m)$ being connected, Bender, Canfield and McKay [9] inferred the probability that $G(n,p)$ is connected as well as a central limit theorem for the number of edges of $G(n,p)$ given connectivity. Moreover, it is possible (though somewhat technical) to derive local limit theorems for $G(n,m)$ and $G(n,p)$ from the main result of [8]. In fact, Pittel and Wormald [24, 25] recently used enumerative arguments to derive an improved version of the main result of [8] and to obtain a local limit theorem that, in addition to $\mathcal{N}$ and $\mathcal{M}$, also includes the order and size of the 2-core. In summary, in [8, 9, 24, 25] enumerative results on the number of connected graphs of given order and size were used to infer the distributions of the order and size of the largest component of $G(n,m)$ and $G(n,p)$. By contrast, in the present paper we use the converse approach: employing probabilistic methods, we first determine the joint distribution of the order and size of the largest component. From this it is possible to derive the number of connected graphs with a given order and size [7]. Recently, Bollobás and Riordan proved, using random walk and martingale arguments, that the (properly rescaled and centred) number of vertices in the giant component in $G(n,p)$ converges in distribution to the normal distribution in the supercritical regime [11].

The asymptotic probability that $G(n,p)$ is connected was first computed by Stepanov [29] (this problem is significantly simpler than computing the probability that $G(n,m)$ is connected). He also obtained a local limit theorem for $\mathcal{N}(G(n,p))$, but his methods are insufficient to obtain the joint distribution of $\mathcal{N}(G(n,p))$ and $\mathcal{M}(G(n,p))$. Moreover, Pittel [23] derived central limit theorems for $\mathcal{N}(G(n,p))$ and $\mathcal{N}(G(n,m))$ from his result on the joint distribution of the numbers of trees of given sizes outside the giant component. The arguments in [23, 29] are of an enumerative and analytic nature.

A few authors have applied probabilistic arguments to problems related to the present work. For instance, O'Connell [22] employed the theory of large deviations in order to estimate the probability that $G(n,p)$ is connected up to a factor $\exp(o(n))$. Whereas this result is significantly less precise than Stepanov's, O'Connell's proof is simpler. In addition, Barraez, Boucheron and Fernandez de la Vega [4] exploited the analogy between the component structure of $G(n,p)$ and branching processes to derive a central limit theorem for the joint distribution of $\mathcal{N}(G(n,p))$ and the *total* number of edges in $G(n,p)$. However, their techniques do not yield a *local* limit theorem. Finally, van der Hofstad and Spencer [17] used a novel perspective on the branching process argument to rederive the formula of Bender, Canfield and McKay [8] for the number of connected graphs. Hence, it is possible to derive bivariate local limit theorems for the order and size of the largest component of $G(n,p)$ and $G(n,m)$ from the results of [17].

1.3.2. Hypergraphs. By comparison with the case of graphs ($d = 2$), little is known about the phase transition and the connectivity probability of random d-uniform hypergraphs with $d > 2$. In fact, to our knowledge, the arguments used in most of the aforementioned papers do not extend to the case $d > 2$.

Karoński and Łuczak [19] derived an asymptotic formula for the number of connected d-uniform hypergraphs of order n and size $m = \frac{n}{d-1} + o(\ln n/\ln\ln n)$ via combinatorial techniques. Since the minimum number of edges necessary for connectivity is $\frac{n-1}{d-1}$, this formula addresses *sparsely* connected hypergraphs. Using this result, Karoński and Łuczak [20] investigated the phase transition in $H_d(n,m)$ and $H_d(n,p)$. They obtained local limit theorems for the joint distribution of the order and size of the largest component in both $H_d(n,m)$ and $H_d(n,p)$ in the *early supercritical phase*, That is, their results apply to the case

$$m = \binom{n}{d}p = \frac{n}{d(d-1)} + o(n^{2/3}(\ln n/\ln\ln n)^{1/3}).$$

Furthermore, Andriamampianina and Ravelomanana [2] extended the result from [19] to the regime $m = \frac{n}{d-1} + o(n^{1/3})$ via enumerative techniques. In addition, relying on [2], Ravelomanana and Rijamamy [26] extended [20] to

$$m = \binom{n}{d}p = \frac{n}{d(d-1)} + o(n^{7/9}).$$

Note that all of these results either deal with *very sparsely* connected hypergraphs, that is,

$$m = \frac{n}{d-1} + o(n),$$

or with the *early* supercritical phase, that is,

$$m = \binom{n}{d} p = \frac{n}{d(d-1)} + o(n).$$

By contrast, the results of this paper concern the component structure of random hypergraphs $H_d(n,m)$ or $H_d(n,p)$ with

$$m = \binom{n}{d} p = \frac{n}{d(d-1)} + \Omega(n).$$

Thus, our results and those from [2, 19, 20, 26] are complementary. Indeed, it would be interesting to see if the techniques of the present work can be extended into the 'scaling window' to close the gap left by [26]. Recently, Bollobás and Riordan established that the (properly rescaled and centred) number of vertices in the giant component in $H_d(n,p)$ converges in distribution to the normal distribution throughout the supercritical regime, using the same arguments as in [12].

The regime of m and p that we deal with in the present work was previously studied by Coja-Oghlan, Moore and Sanwalani [13] via probabilistic arguments. Setting up an analogy between a certain branching process and the component structure of $H_d(n,p)$, they computed the expected order and size of the largest component of $H_d(n,p)$ along with the variance of $\mathcal{N}(H_d(n,p))$. Furthermore, they computed the probability that $H_d(n,m)$ or $H_d(n,p)$ is connected up to a constant factor. Whereas the arguments of [13] by themselves are not strong enough to yield local limit theorems, combining the branching process arguments with further probabilistic techniques, in [6] we inferred a local limit theorem for $\mathcal{N}(H_d(n,p))$. Theorems 1.1 and 1.3 extend this result significantly by giving local limit theorems for the *joint* distribution of $\mathcal{N}$ and $\mathcal{M}$.

1.4. Techniques and outline

To prove Theorems 1.1 and 1.3, we build upon a qualitative result on the connected components of $H_d(n,p)$ from [13], and a local limit theorem for $\mathcal{N}(H_d(n,p))$ from our previous paper [6] (Theorems 2.2 and 2.3: see Section 2). The proofs of both of these ingredients rely solely on probabilistic reasoning (mostly branching process arguments).

In Section 3 we show that (somewhat surprisingly) the *univariate* local limit theorem for $\mathcal{N}(H_d(n,p))$ from [6] can be converted into a *bivariate* local limit theorem for $\mathcal{N}(H_d(n,m))$ and $\mathcal{M}(H_d(n,m))$. To this end, we observe that the local limit theorem for $\mathcal{N}(H_d(n,p))$ implies a bivariate local limit theorem for the joint distribution of $\mathcal{N}(H_d(n,p))$ and the number $\bar{\mathcal{M}}(H_d(n,p))$ of edges *outside* the largest component. Then, we will set up a relationship between the joint distribution of $\mathcal{N}(H_d(n,p))$ and $\bar{\mathcal{M}}(H_d(n,p))$ and the joint distribution of $\mathcal{N}(H_d(n,m))$ and $\bar{\mathcal{M}}(H_d(n,m))$. This will put us in a position to infer the joint distribution of $\mathcal{N}(H_d(n,m))$ and $\bar{\mathcal{M}}(H_d(n,m))$ via Fourier analysis. As in $H_d(n,m)$ the *total* number of edges is fixed (namely, m), we have $\bar{\mathcal{M}}(H_d(n,m)) = m - \mathcal{M}(H_d(n,m))$.

https://doi.org/10.1017/S0963548314000017 Published online by Cambridge University Press

Hence, we obtain a local limit theorem for the joint distribution of $\mathcal{N}(H_d(n,m))$ and $\mathcal{M}(H_d(n,m))$, that is, Theorem 1.3. Further, Theorem 1.3 easily implies Theorem 1.1. We believe that this Fourier-analytic approach may have further applications to related problems.

2. Preliminaries

Throughout the paper we let

$$\phi_{\mu,\sigma^2}(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp(-(x-\mu)^2/2\sigma^2)$$

denote the density of the normal distribution. We let $\phi = \phi_{0,1}$ denote the density of the standard normal distribution.

We will make use of the following *Chernoff bound* on the tails of a binomially distributed variable $X = \mathrm{Bin}(v,q)$ (see [18, p. 26] for a proof): for any $t > 0$ we have

$$\mathbb{P}\big[|X - \mathbb{E}(X)| \geqslant t\big] \leqslant 2\exp\left(-\frac{t^2}{2(\mathbb{E}(X)+t/3)}\right). \tag{2.1}$$

Moreover, we employ the following *local limit theorem* for the binomial distribution (see [10, Chapter 1]).

Proposition 2.1. *Suppose that $0 \leqslant p = p(n) \leqslant 1$ is a sequence such that $np(1-p) \to \infty$ as $n \to \infty$. Let $X = \mathrm{Bin}(n,p)$. Then for any sequence $x = x(n)$ of integers such that $|x - np| = o(np(1-p))^{2/3}$,*

$$\mathbb{P}\big[X = x\big] \sim \frac{1}{\sqrt{2\pi np(1-p)}}\exp\left(-\frac{(x-np)^2}{2p(1-p)n}\right) \quad \text{as } n \to \infty.$$

The following theorem summarizes results from [13, Section 6] on the component structure of $H_d(n,p)$. (The theorem is a slight variation of [13, Theorem 5], which is proved in [13, Section 6].)

Theorem 2.2. *Let $c = c(n)$ be a sequence of non-negative reals and let $p = c\binom{n-1}{d-1}^{-1}$ and $m = \binom{n}{d}p = cn/d$. Then for both $H = H_d(n,p)$ and $H = H_d(n,m)$, the following holds.*

(i) *For any $c_0 < (d-1)^{-1}$ there is a number n_0 such that for all $n > n_0$ for which $c = c(n) \leqslant c_0$ we have*

$$\mathbb{P}\big[\mathcal{N}(H) \leqslant 300(d-1)^2(1-(d-1)c_0)^{-2}\ln n\big] \geqslant 1 - n^{-100}.$$

(ii) *For any $c_0 > (d-1)^{-1}$ there are numbers $n_0 > 0$, $0 < c_0' < (d-1)^{-1}$ such that, for all $n > n_0$ for which $c_0 \leqslant c = c(n) < \ln n/\ln\ln n$, the following holds. The transcendental equation (1.1) has a unique solution $0 < \rho = \rho(n) < 1$, which satisfies*

$$\rho^{d-1}c < c_0'.$$

Furthermore, with probability $\geqslant 1 - n^{-100}$ *there exists precisely one component of order* $(1-\rho)n + o(n)$ *in* H*, while all other components have order* $\leqslant \ln^2 n$*. In addition,*

$$\mathbb{E}\big[\mathcal{N}(H)\big] = (1-\rho)n + o(\sqrt{n}).$$

We also need the following local limit theorem for $\mathcal{N}(H_d(n,p))$ from [6].

Theorem 2.3. *Let* $d \geqslant 2$ *be a fixed integer. For any two compact intervals* $\mathcal{I} \subset \mathbb{R}$, $\mathcal{J} \subset ((d-1)^{-1}, \infty)$*, and for any* $\delta > 0$*, there exist* $n_0 > 0$ *and* $C_0 > 0$ *such that the following holds. Let* $p = p(n)$ *be a sequence such that* $c = c(n) = \binom{n-1}{d-1} p \in \mathcal{J}$ *for all* n*. Then, for all* $n \geqslant n_0$ *the following two statements are true.*

(i) *We have* $\mathbb{P}\big[\mathcal{N}(H_d(n,p)) = v\big] \leqslant C_0/\sqrt{n}$ *for all* v*.*

(ii) *Let* $0 < \rho = \rho(n) < 1$ *be the unique solution to* (1.1)*, and let* $\sigma_{\mathcal{N}}$ *be as in* (1.2)*. If* v *is an integer such that* $\sigma_{\mathcal{N}}^{-1}(v - (1-\rho)n) \in \mathcal{I}$*, then*

$$\frac{1-\delta}{\sqrt{2\pi}\sigma_{\mathcal{N}}} \exp\left[-\frac{(v-(1-\rho)n)^2}{2\sigma_{\mathcal{N}}^2}\right] \leqslant \mathbb{P}\big[\mathcal{N}(H_d(n,p)) = v\big]$$
$$\leqslant \frac{1+\delta}{\sqrt{2\pi}\sigma_{\mathcal{N}}} \exp\left[-\frac{(v-(1-\rho)n)^2}{2\sigma_{\mathcal{N}}^2}\right].$$

We will use some properties of the Fourier transform (see [27]). Given a measurable function $f : \mathbb{R} \to \mathbb{C}$, we define

$$\|f\|_p = \left(\int_{-\infty}^{\infty} |f(z)|^p \, dz\right)^{1/p}, \quad \text{for } 1 \leqslant p < \infty.$$

Here and throughout the paper dz denotes Lebesgue measure. As usual, we let

$$\|f\|_\infty = \inf\{C \geqslant 0 : |f(z)| \leqslant C \text{ for almost every } z \in \mathbb{R}\}.$$

We let $L_p(\mathbb{R})$ consist of all measurable $f : \mathbb{R} \to \mathbb{C}$ for which $\|f\|_p < \infty$ for $1 \leqslant p \leqslant \infty$. For $f \in L_1(\mathbb{R})$, its Fourier transform is defined by

$$\hat{f}(\xi) = \int_{-\infty}^{\infty} f(z)\, e^{-i\xi z} \frac{dz}{\sqrt{2\pi}}, \quad \text{for } \xi \in \mathbb{R}.$$

The Fourier transform translates convolution into pointwise product: given $f, g \in L_1(\mathbb{R})$, the convolution of f and g defined by

$$(f * g)(\zeta) = \int_{-\infty}^{\infty} f(\zeta - z) g(z)\, dz = \int_{-\infty}^{\infty} f(z) g(\zeta - z)\, dz, \quad \text{for all } \zeta \in \mathbb{R}$$

satisfies

$$\widehat{f * g}(\xi) = \sqrt{2\pi} \hat{f}(\xi) \hat{g}(\xi), \quad \text{for all } \xi \in \mathbb{R}. \tag{2.2}$$

We shall use the Plancherel theorem stating that, for $f \in L_1(\mathbb{R}) \cap L_2(\mathbb{R})$,

$$\|f\|_2 = \|\hat{f}\|_2, \tag{2.3}$$

and the inversion theorem asserting that, for $f, \hat{f} \in L_1(\mathbb{R})$,

$$f(z) = \int_{-\infty}^{\infty} \hat{f}(\xi)\, e^{i\xi z}\, \frac{d\xi}{\sqrt{2\pi}}, \quad \text{for almost every } z \in \mathbb{R}. \tag{2.4}$$

We also need (the following special case of) Young's inequality:

$$\|f * g\|_2 \leqslant \|f\|_1 \|g\|_2, \tag{2.5}$$

for any $f \in L_1(\mathbb{R})$ and $g \in L_2(\mathbb{R})$.

We use the 'O-notation' to express asymptotic estimates as $n \to \infty$. Typically we will apply this notation to expressions that do not only depend on n, but also on various other parameters. Suppose that $f(x_1, \ldots, x_k, n)$, $g(x_1, \ldots, x_k, n)$ are functions of n and further parameters x_i are from domains $D_i \subset \mathbb{R}$ $(1 \leqslant i \leqslant k)$, and that $g \geqslant 0$. Then we say that the estimate $f(x_1, \ldots, x_k, n) = O(g(x_1, \ldots, x_k, n))$ holds *uniformly in* $x_1, \ldots, x_k$ if the following is true: there exist numbers $C > 0$ and $n_0 > 0$ such that

$$|f(x_1, \ldots, x_k, n)| \leqslant C \cdot g(x_1, \ldots, x_k, n) \quad \text{for all } n \geqslant n_0 \text{ and } (x_1, \ldots, x_k) \in \prod_{j=1}^{k} D_j.$$

Similarly, we say that $f(x_1, \ldots, x_k, n) \sim g(x_1, \ldots, x_k, n))$ holds *uniformly in* $x_1, \ldots, x_k$ if for any $\varepsilon > 0$ there exists $n_0 > 0$ such that, for all $n > n_0$,

$$\sup_{(x_1, \ldots, x_k) \in D_1 \times \cdots \times D_k} \left| \frac{f(x_1, \ldots, x_k, n)}{g(x_1, \ldots, x_k, n)} - 1 \right| < \varepsilon.$$

We define uniformity analogously for the other Landau symbols Ω, Θ, *etc.*

3. The local limit theorem for $H_d(n, m)$: proof of Theorem 1.3

Throughout this section, we fix two compact sets $\mathcal{J} \subset ((d-1)^{-1}, \infty)$ and $\mathcal{I} \subset \mathbb{R}^2$. Let $\delta > 0$ be arbitrarily small but fixed (*i.e.*, independent of n). In addition, $0 < p = p(n) < 1$ is a sequence of edge probabilities such that $c = \binom{n-1}{d-1} p \in \mathcal{J}$ for all n. Then, by Theorem 2.2 there exists a unique $0 < \rho = \rho(n) < 1$ such that

$$\rho = \exp\left[c(\rho^{d-1} - 1)\right].$$

Let

$$\sigma = \sqrt{\binom{n}{d} p(1-p)}.$$

Let $v = v(n)$ and $\bar{\mu} = \bar{\mu}(n)$ be two sequences of integers. We set

$$x = x(n) = v - (1-\rho)n \quad \textit{and} \quad y = y(n) = \rho^d \binom{n}{d} p - \bar{\mu}. \tag{3.1}$$

We assume that $n^{-1/2}(x, y) \in \mathcal{I}$.

Since Theorems 1.1 and 1.3 are statements that hold for large n, throughout this section we will assume implicitly that $n > n_0$ for some sufficiently large number $n_0 = n_0(d, \delta, \mathcal{I}, \mathcal{J})$. We will use asymptotic notation with respect to $n \to \infty$, and all asymptotics are understood to hold uniformly for $c = c(n) \in \mathcal{J}$ and $x = x(n)$, $y = y(n)$ such that $n^{-1/2}(x, y) \in \mathcal{I}$.

3.1. Outline

In this section we outline the proof of Theorem 1.3. Our starting point is Theorem 2.3, that is, the local limit theorem for the order $\mathcal{N}(H_d(n,p))$ of the largest component. We shall convert this *univariate* local limit theorem into a *bivariate* one that covers both $\mathcal{N}$ and $\mathcal{M}$. At first glance, this may seem implausible as the univariate local limit theorem seems to contain 'too little information' to also infer the precise distribution of the number of edges in the largest component. However, perhaps surprisingly, two simple observations will allow us to show that the univariate local limit theorem does indeed 'encode' the distribution of the size of the largest component implicitly.

The first observation is that Theorem 2.3 implies a local limit theorem for the joint distribution of $\mathcal{N}(H_d(n,p))$ and the number $\bar{\mathcal{M}}(H_d(n,p))$ of edges *outside* the largest component. As we will elaborate below, the reason for this is the well-known *duality principle* for the largest component (*e.g.*, Alon and Spencer [1]). This principle states that the hypergraph obtained from $H_d(n,p)$ by *removing* the largest component is close in distribution to a random hypergraph $H_d(n-\mathcal{N}(H_d(n,p)),p)$ on the remaining vertices. In particular, given that $\mathcal{N}(H_d(n,p))=v$, the number $\bar{\mathcal{M}}(H_d(n,p))$ of edges outside the largest component has (approximately) a binomial distribution $\mathrm{Bin}(\binom{n-v}{d},p)$. Indeed, we will show that for integers v 'close to' $\mathbb{E}[\mathcal{N}(H_d(n,p))]$ and $\bar{\mu}$ 'close to' $\mathbb{E}[\bar{\mathcal{M}}(H_d(n,p))]$ we have

$$\mathbb{P}\big[\mathcal{N}(H_d(n,p))=v,\ \bar{\mathcal{M}}(H_d(n,p))=\bar{\mu}\big] \sim \mathbb{P}\big[\mathcal{N}(H_d(n,p))=v\big]\cdot\mathbb{P}\left[\mathrm{Bin}\left(\binom{n-v}{d},p\right)=\bar{\mu}\right] \tag{3.2}$$

(see Lemma 4.2 below for a precise statement).

The two factors on the right-hand side are known: the local limit theorem for $\mathcal{N}(H_d(n,p))$ (Theorem 2.3) gives an asymptotic expression for $\mathbb{P}[\mathcal{N}(H_d(n,p))=v]$. Moreover, the well-known local limit theorem for the binomial distribution (Proposition 2.1) yields an explicit expression for $\mathbb{P}\big[\mathrm{Bin}\big(\binom{n-v}{d},p\big)=\bar{\mu}\big]$. Thus, we can easily obtain a bivariate local limit theorem for $\mathcal{N}(H_d(n,p))$, $\bar{\mathcal{M}}(H_d(n,p))$ from (3.2).

However, (3.2) does not (yet) yield the joint distribution of $\mathcal{N}(H_d(n,p))$ and the number $\mathcal{M}(H_d(n,p))$ of edges *inside* the largest component. This is because in $H_d(n,p)$ the *total* number of edges

$$|E(H_d(n,p))| = \mathcal{M}(H_d(n,p)) + \bar{\mathcal{M}}(H_d(n,p))$$

is a random variable. In fact, $|E(H_d(n,p))|$ is quite non-trivially correlated to $\mathcal{N}(H_d(n,p))$.

The second key observation of our proof is that we can get around this problem by working with the $H_d(n,m)$ model. In $H_d(n,m)$ the step from $\mathcal{M}$ to $\bar{\mathcal{M}}$ is easy because the total number of edges is fixed to be m. Therefore,

$$\bar{\mathcal{M}}(H_d(n,m)) = m - \mathcal{M}(H_d(n,m)). \tag{3.3}$$

Furthermore, for *any* edge probability p' the two random hypergraph models $H_d(n,p')$ and $H_d(n,m)$ are closely related: given that the total number of edges in $H_d(n,p')$ equals

https://doi.org/10.1017/S0963548314000017 Published online by Cambridge University Press

m, $H_d(n,p')$ is distributed exactly as $H_d(n,m)$. Consequently,

$$\mathbb{P}\big[\mathcal{N}(H_d(n,p')) = v,\ \bar{\mathcal{M}}(H_d(n,p')) = \bar{\mu}\big]$$
$$= \sum_{m=0}^{\binom{n}{d}} \mathbb{P}\left[\mathrm{Bin}\left(\binom{n}{d}, p'\right) = m\right] \cdot \mathbb{P}\big[\mathcal{N}(H_d(n,m)) = v,\ \bar{\mathcal{M}}(H_d(n,m)) = \bar{\mu}\big]. \tag{3.4}$$

Recall that our goal is to compute $\mathbb{P}[\mathcal{N}(H_d(n,m)) = v,\ \bar{\mathcal{M}}(H_d(n,m)) = \bar{\mu}]$, that is, the last term in (3.4). Equation (3.4) puts us in a position to do so. The reason for this is that (3.4) holds for *any* $p' \in (0,1)$, while $\mathbb{P}[\mathcal{N}(H_d(n,m)) = v,\ \bar{\mathcal{M}}(H_d(n,m)) = \bar{\mu}]$ is independent of p'. Thus, we could view (3.4) as an infinite system of linear equations (one for each value of p') that we aim to 'solve' for $\mathbb{P}[\mathcal{N}(H_d(n,m)) = v,\ \bar{\mathcal{M}}(H_d(n,m)) = \bar{\mu}]$.

The way to solve this 'system of equations' is via Fourier inversion. To apply this technique, we need to parametrize (3.4) suitably. This means that we are going to work with certain values of the edge probability p' that will turn out to be particularly convenient. More precisely, given the sequence $p = p(n)$ of edge probabilities such that $c = c(n) = \binom{n-1}{d-1} p \in \mathcal{J}$, we let

$$p_z = p + z\sigma \binom{n}{d}^{-1} \quad \text{and} \quad m_z = \left\lceil \binom{n}{d} p_z \right\rceil = \left\lceil \binom{n}{d} p + z\sigma \right\rceil, \quad \text{for } z \in \mathbb{R}, \tag{3.5}$$

and we set $z^* = \ln^2 n$. Then (3.4) implies that for any $z \in [-z^*, z^*]$ and for n sufficiently large to ensure that $p_z \in [0,1]$ for all such z, we have

$$\mathbb{P}\big[\mathcal{N}(H_d(n,p_z)) = v,\ \bar{\mathcal{M}}(H_d(n,p_z)) = \bar{\mu}\big]$$
$$= \sum_{m=0}^{\binom{n}{d}} \mathbb{P}\left[\mathrm{Bin}\left(\binom{n}{d}, p_z\right) = m\right] \cdot \mathbb{P}\big[\mathcal{N}(H_d(n,m)) = v,\ \bar{\mathcal{M}}(H_d(n,m)) = \bar{\mu}\big]. \tag{3.6}$$

We choose this parametrization (and this value of z^*) because it will allow us to approximate the terms $\mathbb{P}\big[\mathrm{Bin}\big(\binom{n}{d}, p_z\big) = m\big]$ by a Gaussian distribution. More precisely, we are going to rephrase the right-hand side of (3.4) as a convolution of a Gaussian distribution, corresponding to $\mathbb{P}\big[\mathrm{Bin}\big(\binom{n}{d}, p_z\big) = m\big]$, with a certain function g that encodes the terms

$$\mathbb{P}\big[\mathcal{N}(H_d(n,m)) = v,\ \bar{\mathcal{M}}(H_d(n,m)) = \bar{\mu}\big].$$

Since (3.2) gives us an explicit expression for the left-hand side, this will allow us to determine the function g and thus the 'unknowns' $\mathbb{P}[\mathcal{N}(H_d(n,m)) = v,\ \bar{\mathcal{M}}(H_d(n,m)) = \bar{\mu}]$. Let us point out that, since taking the Fourier transform corresponds to a basis transformation, this approach can be seen quite directly as solving the 'system of linear equations' (3.4) via diagonalization.

To carry out this approach, we define two functions f and g, with f corresponding to the left-hand side of (3.6) and with g encoding $\mathbb{P}[\mathcal{N}(H_d(n,m)) = v,\ \bar{\mathcal{M}}(H_d(n,m)) = \bar{\mu}]$: we let

$$f(z) = \begin{cases} n \cdot \mathbb{P}\big[\mathcal{N}(H_d(n,p_z)) = v,\ \bar{\mathcal{M}}(H_d(n,p_z)) = \bar{\mu}\big] & \text{if } z \in [-z^*, z^*], \\ 0 & \text{if } z \in \mathbb{R} \setminus [-z^*, z^*], \end{cases}$$

https://doi.org/10.1017/S0963548314000017 Published online by Cambridge University Press

$$g(z) = \begin{cases} n \cdot \mathbb{P}\left[\mathcal{N}(H_d(n,m_z)) = \nu,\ \bar{\mathcal{M}}(H_d(n,m_z)) = \bar{\mu}\right] & \text{if } z \in [-z^*, z^*], \\ 0 & \text{if } z \in \mathbb{R} \setminus [-z^*, z^*]. \end{cases}$$

The scaling factor n will turn out to be appropriate to ensure that $f(0), g(0) = \Theta(1)$. In this notation, our objective is to compute $g(0)$ explicitly.

To this end, we are going to proceed in four steps. First, we are going to exhibit a function F such that $\|f - F\|_2 = o(1)$ explicitly. Second, we are going to show that (3.6) can be restated as $\|f - g * \phi\|_2 = o(1)$. Third, we are going to determine a function h such that $F = h * \phi$. Finally, we are going to infer that $|g(0) - h(0)| = o(1)$, thereby obtaining the desired explicit formula for $g(0)$. From this the proof of Theorem 1.3 will be immediate.

Let us now carry out the details of this plan. To get the function F as above, we use Theorem 2.3 and Proposition 2.1. More precisely, in Section 4 we will prove the following; recall the definition of x, y from (3.1).

Proposition 3.1. *The function f has the following properties.*

(i) *There exists $\gamma_0 = \gamma_0(d, \mathcal{I}, \mathcal{J}) > 0$ such that $f(z) \leqslant \gamma_0$ for all $z \in \mathbb{R}$ and $\|f\|_1, \|f\|_2 \leqslant \gamma_0$. Moreover, $|f(z)| \leqslant \exp(-z^2/\gamma_0) + O(n^{-90})$ for $|z| > \gamma_0$.*

(ii) *Let $\gamma > 0$ be arbitrarily large but fixed as n grows. Let*

$$\lambda = \frac{d\sigma(\rho - \rho^d)}{\sigma_{\mathcal{N}}(1 - c(d-1)\rho^{d-1})}. \tag{3.7}$$

Then $\lambda > 0$ and the function

$$F(z) = \frac{n}{2\pi\rho^{d/2}\sigma\sigma_{\mathcal{N}}} \cdot \exp\left[-\frac{1}{2}\left(\lambda^2(z - x\lambda^{-1}\sigma_{\mathcal{N}}^{-1})^2 + \rho^d(z + y\rho^{-d}\sigma^{-1} - c\rho^{-1}\sigma^{-1}x)^2\right)\right]$$

is such that $|f(z) - F(z)| = o(1)$ for all $z \in [-\gamma, \gamma]$.

Thus, part (i) of Proposition 3.1 shows that $\|f\|_1, \|f\|_2, \|f\|_\infty$ are bounded and that $f(z) \to 0$ rapidly as $z \to \infty$. In addition, part (ii) provides an explicit expression $F(z)$ that approximates $f(z)$ well on compact sets. (In Lemma 3.1 and throughout, the exponents in the error terms such as $O(n^{-90})$ are not best possible: they are detailed merely for the sake of concreteness.)

In Section 5 we will prove the following.

Proposition 3.2. *The function g enjoys the following properties.*

(i) *There is a number $\gamma_0 = \gamma_0(d, \mathcal{I}, \mathcal{J}) > 0$ such that $g(z) \leqslant \gamma_0 \cdot \exp(-z^2/\gamma_0)$ for all $z \in \mathbb{R}$.*

(ii) *Consequently, $\|g\|_1, \|g\|_2 = O(1)$.*

(iii) *For any $\alpha > 0$ there are $\beta > 0$ and $n_1 > 0$ so that for all $n \geqslant n_1$ and any $z, z' \in [-\beta, \beta]$ we have $|g(z') - g(z)| < \alpha$.*

In Section 3.2 we will establish the following relationship between f and g.

Lemma 3.3. *For almost every $z \in \mathbb{R}$ we have*

$$f(z) = (1 + o(1))(g * \phi(z)) + O(n^{-18}).$$

Furthermore, $\|f - g * \phi\|_2 = o(1)$.

Using Proposition 3.1 and Lemma 3.3, we find a function h such that $\|f - h * \phi\|_2 = o(1)$. More precisely, in Section 3.3 we will prove the following.

Lemma 3.4. *Let λ be as in* (3.7), *and define*

$$\varsigma = \lambda^2 + \rho^d, \quad \kappa = -\left[\frac{\lambda}{\sigma_{\mathcal{N}}} + \frac{c\rho^{d-1}}{\sigma}\right]x + \frac{y}{\sigma}, \quad \theta = \frac{x^2}{\sigma_{\mathcal{N}}^2} + \frac{(c\rho^{d-1}x - y)^2}{\rho^d \sigma^2}.$$

Then $0 < \varsigma \leqslant 1 - \Omega(1)$ *and the function*

$$h(z) = \frac{n}{2\pi\rho^{d/2}\sqrt{1-\varsigma}\sigma_{\mathcal{N}}\sigma} \exp\left[-\frac{(z + \kappa\varsigma^{-1})^2}{2(\varsigma^{-1} - 1)} - \frac{\varsigma\theta - \kappa^2}{2\varsigma}\right] \tag{3.8}$$

satisfies $\|f - h * \phi\|_2 = o(1)$.

From Lemma 3.3 and Lemma 3.4, we have the two relations

$$\|f - g * \phi\|_2 = o(1) \quad \text{and} \quad \|f - h * \phi\|_2 = o(1).$$

In Section 3.4 we shall see that these bounds imply the following.

Lemma 3.5. *We have* $|g(0) - h(0)| = o(1)$.

Proof of Theorem 1.3 (assuming Propositions 3.1 and 3.2 and Lemmas 3.3–3.5). Let $m = m(n)$ be a sequence of integers such that $c = dm/n \in \mathcal{J}$. Moreover, recall that $\nu = \nu(n)$, $\mu = \mu(n)$ are sequences of integers such that $x = \nu - (1-\rho)n$ and $y = \mu - (1-\rho^d)m$ satisfy $n^{-1/2}(x, y) \in \mathcal{I}$. Let $p = p(n)$ be such that $\binom{n-1}{d-1}p = c$. This choice of p ensures that $m_0 = m$ (see (3.5)). In addition, let $\bar{\mu} = m_0 - \mu$. Since $\mathcal{M}(H_d(n,m)) = m - \bar{\mathcal{M}}(H_d(n,m))$, the definition of g ensures that

$$\mathbb{P}\big[\mathcal{N}(H_d(n,m)) = \nu,\ \mathcal{M}(H_d(n,m)) = \mu\big] = g(0)/n. \tag{3.9}$$

By Lemma 3.5, for any $\delta > 0$ there exists $n_0 > 0$ such that for all $n > n_0$ we have

$$|h(0) - g(0)| < \delta. \tag{3.10}$$

Finally, one verifies

$$h(0) = n \cdot Q(\nu - (1-\rho)n, \mu - (1-\rho^d)cn/d) = n \cdot Q(x, y), \tag{3.11}$$

where Q is the function defined in Theorem 1.3.[1] Combining (3.9)–(3.11), we see that for any $\delta > 0$ there exists n_0 such that for all $n > n_0$ we have

$$|\mathbb{P}\big[\mathcal{N}(H_d(n,m)) = \nu,\ \mathcal{M}(H_d(n,m)) = \mu\big] - Q(x, y)| < \delta/n.$$

[1] Full details can be found in [5, Chapter 4].

https://doi.org/10.1017/S0963548314000017 Published online by Cambridge University Press

Note that $Q(x,y) = \Theta(1/n)$, because $\tau^2_{\mathcal{N}}, \tau^2_{\mathcal{M}}, \tau_{\mathcal{N}\mathcal{M}}$ are of order n. This implies Theorem 1.3. □

Thus, our remaining task is to prove Propositions 3.1 and 3.2 and Lemmas 3.3–3.5. The proofs of Lemmas 3.3 and 3.5, which largely rely on arguments from Fourier analysis, can be found in Sections 3.2 and 3.4. Moreover, in Section 3.3 we prove Lemma 3.4. Sections 4 and 5 contain the proofs of Propositions 3.1 and 3.2, which rely on techniques from probabilistic combinatorics.

3.2. Proof of Lemma 3.3

Set $m_- = m_0 - z^*\sigma$, $m_+ = m_0 + z^*\sigma$, and let

$$P(m) = n \cdot \mathbb{P}\big[\mathcal{N}(H_d(n,m)) = \nu,\ \bar{\mathcal{M}}(H_d(n,m)) = \bar{\mu}\big],$$
$$B_z(m) = \mathbb{P}\left[\mathrm{Bin}\left(\binom{n}{d}, p_z\right) = m\right].$$

Then, by the definition of f and (3.6) for all $z \in [-z^*/2, z^*/2]$, we have

$$\begin{aligned} f(z) = \sum_{m=0}^{\binom{n}{d}} P(m)B_z(m) &= \sum_{m_- \leqslant m \leqslant m_+} P(m)B_z(m) + \sum_{m \notin [m_-,m_+]} P(m)B_z(m) \\ &\leqslant \sum_{m_- \leqslant m \leqslant m_+} P(m)B_z(m) + \sum_{m \notin [m_-,m_+]} B_z(m) \qquad (3.12)\\ &\leqslant \sum_{m_- \leqslant m \leqslant m_+} P(m)B_z(m) + \mathbb{P}\left[\mathrm{Bin}\left(\binom{n}{d}, p_z\right) \notin [m_-, m_+]\right]. \qquad (3.13) \end{aligned}$$

For $z \in [-z^*/2, z^*/2]$ the mean of the binomial distribution in the last summand satisfies

$$\binom{n}{d} p_z = m_0 + z\sigma \in \left[m_0 - \frac{z^*\sigma}{2}, m_0 + \frac{z^*\sigma}{2}\right].$$

Hence, applying the Chernoff bound (2.1) and recalling that $z^* = \ln^2 n$, for $n > n_0$ large enough we have

$$\mathbb{P}\left[\mathrm{Bin}\left(\binom{n}{d}, p_z\right) \notin [m_-, m_+]\right] \leqslant 2\exp\left(-\frac{z^{*2}\sigma^2}{20m_0}\right) \leqslant n^{-100}.$$

Plugging this bound back into (3.13), we see that for $z \in [-z^*/2, z^*/2]$,

$$f(z) = O(n^{-100}) + \sum_{m_- \leqslant m \leqslant m_+} P(m)B_z(m). \qquad (3.14)$$

Furthermore, for $z^*/2 \leqslant |z| \leqslant z^*$, we can bound $f(z)$ via part (i) of Proposition 3.1, which shows that there exists $\gamma_0 = O(1)$ such that $f(z) \leqslant \exp(-z^2/\gamma_0) + O(n^{-90})$ for $|z| > \gamma_0$. As $z^* = \ln^2 n$, we therefore obtain for $n > n_0$ large

$$f(z) \leqslant \exp(-\Omega(z^{*2})) + O(n^{-90}) \leqslant O(n^{-90}), \quad \text{for } z^*/2 \leqslant |z| \leqslant z^*. \qquad (3.15)$$

https://doi.org/10.1017/S0963548314000017 Published online by Cambridge University Press

Combining (3.14) and (3.15), we obtain for $|z| \leqslant |z^*|$

$$f(z) = O(n^{-90}) + \sum_{m_- \leqslant m \leqslant m_+} P(m)B_z(m). \tag{3.16}$$

For each integer $0 \leqslant m \leqslant \binom{n}{2}$, let J_m be the set of all reals z such that $m_z = \lceil m_0 + z\sigma \rceil = m$. Then J_m has length $\int_{J_m} 1 d\zeta = 1/\sigma$. Since $g(z) = P(m)$ for all $z \in J_m$ by the definition of g, we get

$$P(m) = \frac{\int_{J_m} g(\zeta)d\zeta}{\int_{J_m} 1 d\zeta} = \sigma \int_{J_m} g(\zeta)d\zeta, \quad \text{for any } m_- \leqslant m \leqslant m_+.$$

Furthermore, by the local limit theorem for the binomial distribution, for each $m_- \leqslant m \leqslant m_+$ we have

$$\begin{aligned} B_z(m) &\sim \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(m - (m_0 + z\sigma))^2}{2\sigma^2}\right] \\ &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(m_\zeta - (m_0 + z\sigma))^2}{2\sigma^2}\right], \quad \text{for all } \zeta \in J_m. \end{aligned}$$

Consequently, for any $m_- \leqslant m \leqslant m_+$ and any $\zeta \in J_m$ we have

$$\begin{aligned} B_z(m) &\sim \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(\lceil m_0 + \zeta\sigma \rceil - (m_0 + z\sigma))^2}{2\sigma^2}\right] \\ &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(\zeta - z)^2}{2}\right] \\ &\quad \cdot \exp\left[\frac{(\lceil m_0 + \zeta\sigma \rceil - (m_0 + z\sigma))^2 - (m_0 + \zeta\sigma - (m_0 + z\sigma))^2}{2\sigma^2}\right]. \end{aligned}$$

Since $m_0 + \zeta\sigma \leqslant \lceil m_0 + \zeta\sigma \rceil \leqslant m_0 + \zeta\sigma + 1$, we have

$$|(\lceil m_0 + \zeta\sigma \rceil - (m_0 + z\sigma))^2 - (m_0 + \zeta\sigma - (m_0 + z\sigma))^2| \leqslant 2|z - \zeta|\sigma + 1.$$

Therefore,

$$\exp\left[\frac{(\lceil m_0 + \zeta\sigma \rceil - (m_0 + z\sigma))^2 - (m_0 + \zeta\sigma - (m_0 + z\sigma))^2}{2\sigma^2}\right] \sim 1$$

and thus

$$B_z(m) \sim \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(\zeta - z)^2}{2}\right] = \sigma^{-1}\phi(z - \zeta), \quad \text{for all } m_- \leqslant m \leqslant m_+,\ \zeta \in J_m.$$

Hence,

$$P(m)B_z(m) \sim \int_{J_m} g(\zeta)\phi(z - \zeta)\, d\zeta.$$

Summing up, we obtain

$$f(z) = (1 + o(1)) \int_{\mathbb{R}} g(\zeta)\phi(z - \zeta)d\zeta + O(z^* n^{-90}) = (1 + o(1))g * \phi(z) + O(n^{-89})$$

for all $z \in \mathbb{R}$. This proves the first part of Lemma 3.3. The second part follows from the dominated convergence theorem because $\|f\|_2 = O(1)$.

https://doi.org/10.1017/S0963548314000017 Published online by Cambridge University Press

3.3. Proof of Lemma 3.4

We start by manipulating the function $F(z)$ a little: we have

$$\begin{aligned}
F(z) &= \frac{n}{2\pi\rho^{d/2}\sigma\sigma_{\mathcal{N}}}\exp\left[-\frac{1}{2}\left(\lambda^2(z - x\lambda^{-1}\sigma_{\mathcal{N}}^{-1})^2 + \rho^d(z + y\rho^{-d}\sigma^{-1} - c\rho^{-1}\sigma^{-1}x)^2\right)\right] \\
&= \frac{n}{2\pi\rho^{d/2}\sigma\sigma_{\mathcal{N}}} \\
&\quad \cdot \exp\left[-\frac{1}{2}\left(\left(\lambda^2 + \rho^d\right)z^2 - 2\left(\frac{x\lambda}{\sigma_{\mathcal{N}}} + \frac{xc\rho^{d-1}}{\sigma} - \frac{y}{\sigma}\right)z + \frac{x^2}{\sigma_{\mathcal{N}}^2} + \frac{\left(xc\rho^{d-1} - y\right)^2}{\rho^d\sigma^2}\right)\right] \\
&= \frac{n}{2\pi\rho^{d/2}\sigma\sigma_{\mathcal{N}}}\exp\left[-\frac{1}{2}\left(\varsigma z^2 + 2\kappa z + \theta\right)\right] \\
&= \frac{n}{2\pi\rho^{d/2}\sigma\sigma_{\mathcal{N}}}\exp\left[-\frac{(z + \kappa\varsigma^{-1})^2}{2\varsigma^{-1}} - \frac{\theta\varsigma - \kappa^2}{2\varsigma}\right].
\end{aligned} \tag{3.17}$$

As a next step, we are going to infer from (3.17) that $\varsigma < 1 - \Omega(1)$. To prove this claim it is convenient to use Lemma 3.3. Let $\mathbb{P}_F = \|F\|_1^{-1}F dz$ be the probability distribution on $\mathbb{R}$ defined by the density function $\|F\|_1^{-1}F$. Then (3.17) shows that $\mathrm{Var}(\mathbb{P}_F) = \varsigma^{-1}$. Thus, we need to show that $\mathrm{Var}(\mathbb{P}_F) > 1 + \Omega(1)$. By the triangle inequality,

$$\|F - g * \phi\|_1 \leqslant \|F - f\|_1 + \|f - g * \phi\|_1. \tag{3.18}$$

As Proposition 3.1 and Lemma 3.3 show that f converges to F as well as to $g * \phi$ pointwise almost everywhere, and as $f, F, g \in L_1(\mathbb{R})$, the dominated convergence theorem implies that f converges to both F and $g * \phi$ in L_1. Hence, (3.18) entails that $\|F - g * \phi\|_1 = o(1)$. As the convolution of two probability measures is a probability measure, we thus obtain $\|g\|_1 \sim \|F\|_1$. Therefore, letting $\mathbb{P}_{\|g\|_1^{-1}g*\phi} = \|g\|_1^{-1}g * \phi\, dz$ and recalling from Propositions 3.1 and 3.2 that both $F(z)$ and $g(z)$ decay exponentially as $z \to \infty$, we see that

$$\varsigma^{-1} = \mathrm{Var}(\mathbb{P}_{\|F\|_1^{-1}F}) \sim \mathrm{Var}(\mathbb{P}_{\|g\|_1^{-1}g*\phi}). \tag{3.19}$$

The probability distribution $\mathbb{P}_{\|g\|_1^{-1}g*\phi}$ equals the distribution of the sum of two independent random variables, one with distribution $\|g\|_1^{-1}g\, dz$ and the second with distribution $\phi\, dz$. Since the variance of the sum of two independent random variables equals the sum of their separate variances, we get

$$\mathrm{Var}(\mathbb{P}_{\|g\|_1^{-1}g*\phi}) = \mathrm{Var}(\mathbb{P}_{\|g\|_1^{-1}g}) + \mathrm{Var}(\mathbb{P}_\phi) = \mathrm{Var}(\mathbb{P}_{\|g\|_1^{-1}g}) + 1 > 1, \tag{3.20}$$

where the last (strict) inequality just follows from the fact that $g \in L_1(\mathbb{R})$ (see Lemma 3.3), as this rules out the possibility of $\mathbb{P}_{\|g\|_1^{-1}g}$ being a point measure. Combining (3.20) and (3.19), we see that $\varsigma < 1$. Furthermore, since ρ and therefore also ς is a continuous function of $c = c(n)$ by the implicit function theorem, and since c ranges over the compact set $\mathcal{J}$ by assumption, the strict inequality $\varsigma < 1$ implies that indeed $\varsigma < 1 - \Omega(1)$, that is, ς remains bounded away from 1 as $n \to \infty$.

Knowing that $\varsigma < 1 - \Omega(1)$, we can define

$$\eta_1 = \frac{n}{2\pi\rho^{d/2}\sqrt{1-\varsigma\sigma_{\mathcal{N}}\sigma}}\exp\left(-\frac{\varsigma\theta-\kappa^2}{2\varsigma}\right),$$
$$\eta_2 = -\kappa\varsigma^{-1},$$
$$\eta_3 = \varsigma^{-1} - 1,$$
$$\eta_4 = \eta_1\sqrt{2\pi\eta_3},$$

so that $h(z) = \eta_4\phi_{\eta_2,\eta_3}$. Hence, it is clear that $h * \phi = \eta_4\phi_{\eta_2,\eta_3+1}$. Finally, a straight computation shows that $F = \eta_4\phi_{\eta_2,\eta_3+1}$. As $f, F \in L_2(\mathbb{R})$ and $f(z) \to F(z)$ pointwise almost everywhere by Proposition 3.1, the dominated convergence theorem yields $\|f - h * \phi\|_2 = o(1)$, as desired.

3.4. Proof of Lemma 3.5

If it *were* true that $f = g * \phi$ and $f = h * \phi$, then we could immediately infer that $g = h$. Indeed, if $f = g * \phi = h * \phi$, then taking the Fourier transform yields $\hat{g}\hat{\phi} = \hat{h}\hat{\phi}$ (using (2.2)). Dividing by $\hat{\phi}$ gives $\hat{g} = \hat{h}$, whence Fourier transforming once more shows $g = h$.

However, knowing only $\|f - g * \phi\|_2, \|f - h * \phi\|_2 = o(1)$ (by Lemmas 3.3 and 3.4), we have to work a little harder. Since $\|f - g * \phi\|_2, \|f - h * \phi\|_2 = o(1)$, there is a function $\omega = \omega(n)$ such that $\lim_{n\to\infty}\omega(n) = \infty$ and

$$\|f - g * \phi\|_2, \|f - h * \phi\|_2 < \frac{1}{2}\exp(-2\omega^2).$$

By the triangle inequality,

$$\|(g - h) * \phi\|_2 < \exp(-2\omega^2) = o(1). \tag{3.21}$$

To compare g and h, the crucial step is to establish that $\|(g-h) * \phi_{0,\tau^2}\|_2 = o(1)$ for 'small' $\tau \leqslant 1$.

Lemma 3.6. *Suppose that* $\omega^{-1/8} \leqslant \tau \leqslant 1$. *Then* $\|(g-h) * \phi_{0,\tau^2}\|_2 \leqslant \exp(-\omega/3)$ *for n sufficiently large.*

Proof of Lemma 3.6. By Proposition 3.1(i) and Lemma 3.3, $f, g \in L_1(\mathbb{R}) \cap L_2(\mathbb{R})$ and thus we can apply the Plancherel theorem (2.3). Let $\psi = \hat{\phi}_{0,\tau^2} = \phi_{0,\tau^{-2}}$. Then

$$\begin{aligned}\|(g-h) * \phi_{0,\tau^2}\|_2^2 &\overset{(2.3)}{=} \|\widehat{(g-h) * \phi_{0,\tau^2}}\|_2^2 \overset{(2.2)}{=} 2\pi\|(\hat{g}-\hat{h})\psi\|_2^2 \\ &= 2\pi\int_{-\omega}^{\omega}|(\hat{g}-\hat{h})\psi|^2 + 2\pi\int_{\mathbb{R}\setminus[-\omega,\omega]}|(\hat{g}-\hat{h})\psi|^2. \end{aligned} \tag{3.22}$$

Since $\hat{\phi} = \phi = \phi_{0,1}$, we obtain

$$\begin{aligned}\int_{-\omega}^{\omega}|(\hat{g}-\hat{h})\psi|^2 &\leqslant \frac{\|\psi\|_\infty^2}{\inf_{-\omega\leqslant t\leqslant\omega}|\hat{\phi}(t)|^2}\int_{-\omega}^{\omega}|(\hat{g}-\hat{h})\hat{\phi}|^2 \\ &\leqslant \exp(\omega^2)\|(\hat{g}-\hat{h})\hat{\phi}\|_2^2\end{aligned}$$

$$= \exp(\omega^2)\|(g-h)*\phi\|_2^2 \quad \text{(by (2.2) and Plancherel (2.3))}$$
$$\overset{(3.21)}{\leqslant} \exp(-\omega^2). \tag{3.23}$$

In addition, by the Cauchy–Schwarz inequality,

$$\int_{\mathbb{R}\setminus[-\omega,\omega]} |(\hat{g}-\hat{h})\psi|^2 \leqslant \left[\int_{\mathbb{R}} |(\hat{g}-\hat{h})^2|^2\right]^{1/2} \cdot \left[\int_{\mathbb{R}\setminus[-\omega,\omega]} |\psi|^4\right]^{1/2}. \tag{3.24}$$

As $\tau^{-2} \leqslant \omega^{1/4}$, we have

$$\int_{\mathbb{R}\setminus[-\omega,\omega]} |\psi|^4 \leqslant \tau^{-4}\int_{\omega}^{\infty} \exp(-2\tau^2\zeta^2)\,d\zeta \leqslant \exp(-\omega). \tag{3.25}$$

Moreover,

$$\begin{aligned} 2\pi\int_{\mathbb{R}} |(\hat{g}-\hat{h})^2|^2 &= 2\pi\|(\hat{g}-\hat{h})^2\|_2^2 \\ &= \|(g-h)*(g-h)\|_2^2 \quad \text{(by (2.2) and (2.3))} \\ &= \|g*g - 2g*h + h*h\|_2^2 \\ &\leqslant \left[\|g*g\|_2 + 2\|g*h\|_2 + \|h*h\|_2\right]^2. \end{aligned} \tag{3.26}$$

Lemma 3.3 shows that $\|g\|_1, \|g\|_2 = O(1)$ and Lemma 3.4 implies $\|h\|_1, \|h\|_2 = O(1)$. Therefore, invoking Young's inequality (2.5), we obtain

$$\begin{aligned} \|g*g\|_2 &\leqslant \|g\|_1\|g\|_2 = O(1), \\ \|g*h\|_2 &\leqslant \|g\|_1\|h\|_2 = O(1), \\ \|h*h\|_2 &\leqslant \|h\|_1\|h\|_2 = O(1). \end{aligned}$$

Plugging these estimates into (3.26), we see that $\int_{\mathbb{R}} |(\hat{g}-\hat{h})^2|^2 = O(1)$. Hence, (3.24) and (3.25) yield

$$\int_{\mathbb{R}\setminus[-\omega,\omega]} |(\hat{g}-\hat{h})\psi|^2 \leqslant O(\exp(-\omega/2)). \tag{3.27}$$

Finally, combining (3.22), (3.23), and (3.27), we obtain the desired bound on $\|(g-h)*\phi_{0,\tau^2}\|_2$. □

Proof of Lemma 3.5. We are going to use Lemma 3.6 to show that $g(0)$ must be close to $h(0)$. The basic idea is as follows. For small τ the function ϕ_{0,τ^2} is just a narrow peak above the origin. Therefore, the continuity property of g established in Proposition 3.2 implies that the convolution $g*\phi_{0,\tau^2}(0)$ is close to $g(0)$. Similarly, $h*\phi_{0,\tau^2}(0)$ is approximately the same as $h(0)$. As $g*\phi_{0,\tau^2}(0)$ and $h*\phi_{0,\tau^2}(0)$ are close by Lemma 3.6, we will be able to conclude that $|h(0)-g(0)| = o(1)$. Let us carry out the details.

Assume for contradiction that there is a positive $\alpha = \Omega(1)$ such that $g(0) > h(0)+\alpha$ for arbitrarily large n (an analogous argument applies in the case $g(0) < h(0)-\alpha$). Let $\tau = \omega^{-1/8}$. Our goal is to show

$$\|(h-g)*\phi_{0,\tau^2}\|_2 > \exp(-\omega/3), \tag{3.28}$$

in contradiction to Lemma 3.6.

https://doi.org/10.1017/S0963548314000017 Published online by Cambridge University Press

To show (3.28), note that Proposition 3.2 and Lemma 3.4 imply $\|g\|_\infty = O(1)$ and $\|h\|_\infty = O(1)$. Hence, there is a number $1 < \Gamma = O(1)$ such that $g(\zeta), h(\zeta) \leqslant \Gamma$, for almost all $\zeta \in \mathbb{R}$. Further, again by Proposition 3.2 and Lemma 3.4 and because h is uniformly continuous on all of $\mathbb{R}$, there is a number $\beta = \beta(\alpha) > 0$ (independent of n) such that

$$|g(0) - g(z)| \leqslant 0.01\alpha,\ |h(0) - h(z)| \leqslant 0.01\alpha, \quad \text{for all } z \text{ such that } |z| < 2\beta. \tag{3.29}$$

Let $\gamma = \int_{\mathbb{R}\setminus[-\beta/2,\beta/2]} \phi_{0,\tau^2}$. Then, for sufficiently large n we have $\gamma < 0.01\alpha\Gamma^{-1}$, because $\tau = \omega^{-1/8} \to 0$ as $n \to \infty$. (Intuitively, the narrow 'spike' that ϕ_{0,τ^2} represents falls into the interval $[-\beta/2, \beta/2]$ around the origin.) Therefore, for any z such that $|z| < \beta/2$, we have

$$\begin{aligned} g * \phi_{0,\tau^2}(z) &= \int_{\mathbb{R}} g(z-\zeta)\phi_{0,\tau^2}(\zeta)d\zeta \geqslant \int_{-\beta}^{\beta} g(z-\zeta)\phi_{0,\tau^2}(\zeta)d\zeta \\ &\overset{(3.29)}{\geqslant} (g(0) - 0.01\alpha)(1-\gamma) \geqslant g(0) - 0.03\alpha, \end{aligned} \tag{3.30}$$

and similarly

$$\begin{aligned} h * \phi_{0,\tau^2}(z) &= \int_{\mathbb{R}} h(z-\zeta)\phi_{0,\tau^2}(\zeta)d\zeta \\ &\leqslant \int_{-\beta}^{\beta} h(z-\zeta)\phi_{0,\tau^2}(\zeta)d\zeta + \Gamma \int_{\mathbb{R}\setminus(-\beta,\beta)} \phi_{0,\tau^2}(\zeta)d\zeta \\ &\leqslant h(0) + 0.01\alpha + \Gamma\gamma \overset{(3.29)}{\leqslant} h(0) + 0.03\alpha. \end{aligned} \tag{3.31}$$

Since (3.30) and (3.31) are true for all z such that $|z| < \beta/2$, our assumption $g(0) > h(0) + \alpha$ yields

$$\|(g-h) * \phi_{0,\tau^2}\|_2^2 \geqslant \int_{-\beta/2}^{\beta/2} |g * \phi_{0,\tau^2}(z) - h * \phi_{0,\tau^2}(z)|^2 dz \geqslant 0.5\alpha^2\beta. \tag{3.32}$$

As α, β remain bounded away from 0 while $\omega(n) \to \infty$ as $n \to \infty$, for sufficiently large n we have $0.5\alpha^2\beta > 2\pi\exp(-\omega/6)$, and thus (3.32) implies (3.28). □

4. Analysis of f: proof of Proposition 3.1

Throughout this section, we keep the notation and the assumptions from Section 3.

Recall the definition of the function f: for $|z| \leqslant z^* = \ln^2 n$ we have

$$f(z) = n \cdot \mathbb{P}\big[\mathcal{N}(H_d(n,p_z)) = \nu,\ \bar{\mathcal{M}}(H_d(n,p_z)) = \bar{\mu}\big],$$

while $f(z) = 0$ for $z \notin [-z^*, z^*]$. Thus, to understand f we need to study the joint distribution of the order $\mathcal{N}(H_d(n,p_z))$ of the largest component and of the number $\bar{\mathcal{M}}(H_d(n,p_z))$ of edges *outside* this component. What we are going to show is that the two events $\{\mathcal{N}(H_d(n,p_z)) = \nu\}$, $\{\bar{\mathcal{M}}(H_d(n,p_z)) = \bar{\mu}\}$ are almost independent. From this Proposition 3.1 follows rather immediately, as we already have a local limit theorem for $\mathcal{N}(H_d(n,p_z))$ (Theorem 2.3), and as $\bar{\mathcal{M}}(H_d(n,p_z))$ will turn out to be a binomial random variable.

The intuitive reason why $\mathcal{N}(H_d(n,p_z)) = v$, $\bar{\mathcal{M}}(H_d(n,p_z)) = \bar{\mu}$ are essentially independent is that for v 'close' to the expected order $(1-\rho)n$ of the largest component, the number $n-v$ of remaining vertices is sufficiently small that the average degree $\binom{n-v-1}{d-1}p_z$ of a random hypergraph $H_d(n-v,p_z)$ on $n-v$ vertices is strictly smaller than $1/(d-1)$. In effect, such a random hypergraph does not typically have a component of order greater than $\ln^2 n$.

As a first step, we study how the expected order of the largest component of $H_d(n,p_z)$ depends on z. Recall that $p_z = p + z\sigma/\binom{n}{d}$ and let $c_z = \binom{n-1}{d-1}p_z$. As $c_z \sim c > (d-1)^{-1}$, there is a unique $\rho_z \in (0,1)$ such that

$$\rho_z = \exp(c_z(\rho_z^{d-1} - 1)).$$

As in (3.7) we let

$$\lambda = \frac{d\sigma(\rho - \rho^d)}{\sigma_{\mathcal{N}}(1 - c(d-1)\rho^{d-1})}.$$

(We know that $1 - c(d-1)\rho^{d-1} \neq 0$ because this term occurs in the denominator of the variance of $\mathcal{N}(H_d(n,p))$: see Theorem 2.3.)

Lemma 4.1. *Let $z \in [-z^*, z^*]$. Then*

$$\mathbb{E}\big[\mathcal{N}(H_d(n,p_z))\big] = (1-\rho_z)n + o(\sqrt{n}) = (1-\rho)n + \lambda\sigma_{\mathcal{N}} z + o(\sqrt{n})$$

and

$$\max\{\mathbb{P}\big[\mathcal{N}(H_d(n-v,p_z)) > \ln^2 n\big], \mathbb{P}\big[\mathcal{N}(H_d(n-v,\bar{\mu})) > \ln^2 n\big]\} = O(n^{-100}).$$

Proof of Lemma 4.1. By Theorem 2.2 we have

$$\mathbb{E}\big[\mathcal{N}(H_d(n,p_z))\big] = (1-\rho_z)n + o(\sqrt{n}). \tag{4.1}$$

Furthermore, the function $z \mapsto \rho_z$ has two continuous derivatives by the implicit function theorem. Consequently, we can Taylor-expand ρ_z at $z = 0$ by differentiating both sides of the transcendental equation $\rho_z = \exp(c_z(\rho_z^{d-1} - 1))$. By the chain rule,

$$\begin{aligned}\frac{d\rho_z}{dz} &= \exp(c_z(\rho_z^{d-1} - 1)) \cdot \frac{d}{dz}\big[c_z(\rho_z^{d-1} - 1)\big] = \rho_z \cdot \frac{d}{dz}\big[c_z(\rho_z^{d-1} - 1)\big] \\ &= d\sigma(\rho_z^d - \rho_z) + c_z(d-1)\rho_z^{d-1} \cdot \frac{d}{dz}\rho_z.\end{aligned}$$

Hence,

$$\rho_z = \rho - \lambda\sigma_{\mathcal{N}} n^{-1} z + o(n^{-1/2}). \tag{4.2}$$

Thus, the first assertion follows from (4.1).

For the second part, we observe that by (4.2) we have $v \sim (1-\rho_z)n \sim (1-\rho)n$ for all $z \in [-z^*, z^*]$. Therefore the average degree c' of $H_d(n-v,p_z)$ satisfies

$$c' = \binom{n-v-1}{d-1}p_z \sim \rho^{d-1}c_0 < (d-1)^{-1},$$

https://doi.org/10.1017/S0963548314000017 Published online by Cambridge University Press

and the average degree c'' in $H_d(n-v,\bar{\mu})$ satisfies

$$c'' = \frac{d\bar{\mu}}{n-v} \sim \rho^{d-1}c_0 < (d-1)^{-1}.$$

By part (i) of Theorem 2.2, the probability that a random hypergraph of average degree smaller than and bounded away from $(d-1)^{-1}$ has a component of order greater than $\ln^2 n$ is bounded by $(n-v)^{-100} = O(n^{-100})$, and hence the second assertion follows. □

Let $G \subset V = \{1,\ldots,n\}$ be a subset of size v. We would like to condition on the event that G is the *largest* component of $H_d(n,p_z)$ and study the conditional distribution of the number of edges in the hypergraph $H_d(n,p_z) - G$ obtained by removing the vertices in G. The problem with this is that once we condition on G being the largest component, the edges of $H_d(n,p_z) - G$ may no longer occur independently, because the conditioning implies that $H_d(n,p_z) - G$ has no component of order greater than v. By contrast, if we just condition on the event that *G is a component* (but not necessarily the largest one), then the conditioning does not affect the edges in $H_d(n,p_z) - G$ at all. Thus, given that G is a component, $H_d(n,p_z) - G$ is identical to a random hypergraph $H_d(n-v,p_z)$, and therefore the number of edges in $H_d(n,p_z) - G$ has a binomial distribution $\mathrm{Bin}(N,p_z)$, where

$$N = \binom{n-v}{d}.$$

The proof of the following lemma, which is reminiscent of arguments used in [22], is based on the observation that conditioning on G being a component is essentially equivalent to conditioning on G being the largest component.

Lemma 4.2. *For all $z \in [-z^*, z^*]$ we have*

$$\frac{f(z)}{n \cdot \mathbb{P}\big[\mathrm{Bin}(N,p_z) = \bar{\mu}\big]\mathbb{P}\big[\mathcal{N}(H_d(n,p_z)) = v\big]} = 1 + O(n^{-100}).$$

Proof of Lemma 4.2. Let $\mathcal{G} = \{G \subset V : |G| = v\}$. For $G \in \mathcal{G}$ we let $\mathcal{C}_G$ denote the event that G is a component in $H_d(n,p_z)$. Then, by the union bound,

$$\begin{aligned}\frac{f(z)}{n} &= \mathbb{P}\big[\mathcal{N}(H_d(n,p_z)) = v,\ \bar{\mathcal{M}}(H_d(n,p_z)) = \bar{\mu}\big] \\ &\leqslant \sum_{G\in\mathcal{G}} \mathbb{P}\big[\mathcal{C}_G,\ |E(H_d(n,p_z) - G)| = \bar{\mu}\big]. \end{aligned} \tag{4.3}$$

The event $\mathcal{C}_G$ merely imposes a condition on the hypergraph spanned by G (it must be connected), and imposes the absence of edges between G and $V \setminus G$. But $\mathcal{C}_G$ imposes no conditioning on the hypergraph $H_d(n,p_z) - G$ spanned by the remaining $n-v$ vertices. In particular, $|E(H_d(n,p_z) - G)|$ is binomially distributed with parameters N and p_z. Hence, (4.3) yields

$$\frac{f(z)}{n} \leqslant \sum_{G\in\mathcal{G}} \mathbb{P}\big[\mathcal{C}_G\big]\mathbb{P}\big[|E(H_d(n,p_z)-G)| = \bar{\mu}\big]$$
$$= \mathbb{P}\big[\mathrm{Bin}(N,p_z) = \bar{\mu}\big]\sum_{G\in\mathcal{G}} \mathbb{P}\big[\mathcal{C}_G\big]. \tag{4.4}$$

Furthermore, by the same token, given that G is a component in $H_d(n, p_z)$, $H_d(n, p_z) - G$ is identical to the random hypergraph $H_d(n - v, p_z)$ on $n - v$ vertices. As a consequence,

$$\mathbb{P}\big[\mathcal{C}_G,\, \mathcal{N}(H_d(n,p_z)-G) < v\big] = \mathbb{P}\big[\mathcal{C}_G\big]\mathbb{P}\big[\mathcal{N}(H_d(n,p_z)-G) < v\big].$$

Therefore, (4.4) yields

$$\frac{f(z)}{n} \leqslant \mathbb{P}\big[\mathrm{Bin}(N,p_z) = \bar{\mu}\big]\sum_{G\in\mathcal{G}} \frac{\mathbb{P}\big[\mathcal{C}_G,\, \mathcal{N}(H_d(n,p_z)-G) < v\big]}{\mathbb{P}\big[\mathcal{N}(H_d(n,p_z)-G) < v\big]}. \tag{4.5}$$

Furthermore, $\mathbb{P}[\mathcal{N}(H_d(n,p_z)-G) < v] \geqslant 1 - O(n^{-100})$ by the second part of Lemma 4.1. Thus, (4.5) entails

$$\frac{f(z)}{n} \leqslant \frac{\mathbb{P}\big[\mathrm{Bin}(N,p_z) = \bar{\mu}\big]}{1 - O(n^{-100})}\sum_{G\in\mathcal{G}} \mathbb{P}\big[\mathcal{C}_G,\, \mathcal{N}(H_d(n,p_z)-G) < v\big]$$
$$= \frac{\mathbb{P}\big[\mathrm{Bin}(N,p_z) = \bar{\mu}\big]}{1 - O(n^{-100})} \cdot \mathbb{P}\big[\exists G \in \mathcal{G} : \mathcal{C}_G,\, \mathcal{N}(H_d(n,p_z)-G) < v\big]$$
$$\leqslant \frac{\mathbb{P}\big[\mathrm{Bin}(N,p_z) = \bar{\mu}\big]\mathbb{P}\big[\mathcal{N}(H_d(n,p_z)) = v\big]}{1 - O(n^{-100})}. \tag{4.6}$$

Conversely, if $G \in \mathcal{G}$ is a component of $H_d(n, p_z)$ and $\mathcal{N}(H_d(n, p_z) - G) < v$, then G is the unique largest component of $H_d(n, p_z)$. Therefore,

$$\frac{f(z)}{n} \geqslant \sum_{G\in\mathcal{G}} \mathbb{P}\big[\mathcal{C}_G,\, \mathcal{N}(H_d(n,p_z)-G) < v,\, |E(H_d(n,p_z)-G)| = \bar{\mu}\big]$$
$$= \sum_{G\in\mathcal{G}} \mathbb{P}\big[\mathcal{C}_G\big]\mathbb{P}\big[(\mathcal{N}(H_d(n,p_z)-G) < v),\, |E(H_d(n,p_z)-G)| = \bar{\mu}\big]. \tag{4.7}$$

Further, given that

$$|E(H_d(n,p_z)-G)| = \bar{\mu},$$

$H_d(n, p_z) - G$ is just a random hypergraph $H_d(n - v, \bar{\mu})$. Hence, (4.7) yields

$$\frac{f(z)}{n} \geqslant \mathbb{P}\big[\mathcal{N}(H_d(n-v,\bar{\mu})) < v\big]\mathbb{P}\big[\mathrm{Bin}(N,p_z) = \bar{\mu}\big]\sum_{G\in\mathcal{G}} \mathbb{P}\big[\mathcal{C}_G\big]$$
$$\geqslant \mathbb{P}\big[\mathcal{N}(H_d(n-v,\bar{\mu})) < v\big]\mathbb{P}\big[\mathrm{Bin}(N,p_z) = \bar{\mu}\big]\mathbb{P}\big[\mathcal{N}(H_d(n,p_z)) = v\big], \tag{4.8}$$

where the last estimate follows from the union bound. Now,

$$\mathbb{P}\big[\mathcal{N}(H_d(n-v,\bar{\mu})) \geqslant v\big] \leqslant O(n^{-100})$$

https://doi.org/10.1017/S0963548314000017 Published online by Cambridge University Press

by the second part of Lemma 4.1. Plugging this into (4.8), we get

$$\frac{f(z)}{n} \geqslant (1 - O(n^{-100}))\mathbb{P}\big[\mathrm{Bin}(N, p_z) = \bar{\mu}\big]\mathbb{P}\big[\mathcal{N}(H_d(n, p_z)) = v\big]. \tag{4.9}$$

Combining (4.6) and (4.9) completes the proof. □

Proof of Proposition 3.1. Since $f(z) = 0$ if $|z| > z^*$, we only need to consider $z \in [-z^*, z^*]$. By Theorem 2.3 (the local limit theorem for $\mathcal{N}(H_d(n, p))$) and Proposition 2.1 (the local limit theorem for the binomial distribution) there is a number $\gamma_1 = \gamma_1(d, \mathcal{J}) > 0$ such that, for all $z \in [-z^*, z^*]$,

$$\mathbb{P}\big[\mathcal{N}(H_d(n, p_z)) = v\big] \leqslant \gamma_1 n^{-1/2} \quad \text{and} \quad \mathbb{P}\big[\mathrm{Bin}(N, p_z) = \bar{\mu}\big] \leqslant \gamma_1 n^{-1/2}. \tag{4.10}$$

Therefore, Lemma 4.2 implies

$$f(z)/n = (1 + O(n^{-100}))\mathbb{P}\big[\mathcal{N}(H_d(n, p_z)) = v\big]\mathbb{P}\big[\mathrm{Bin}(N, p_z) = \bar{\mu}\big], \tag{4.11}$$

whence

$$\sup_{z \in \mathbb{R}} f(z) \leqslant 2\gamma_1^2$$

for large enough n. More precisely, (4.10), (4.11) and Proposition 2.1 imply that for n large,

$$\begin{aligned} f(z) &\leqslant \gamma_1 \cdot \sqrt{n}\, \mathbb{P}\big[\mathrm{Bin}(N, p_z) = \bar{\mu}\big] + O(n^{-100}) \\ &\leqslant 2\gamma_1 \sqrt{\frac{n}{2\pi N p_z(1 - p_z)}} \cdot \exp\left[-\frac{(Np_z - \bar{\mu})^2}{2Np_z(1 - p_z)}\right] + O(n^{-100}) \\ &\leqslant 2\gamma_1 \sqrt{\frac{n}{2\pi N p_z(1 - p_z)}} \cdot \exp\left[-\frac{\big(Np_0 - \bar{\mu} + z\sigma \cdot N/\binom{n}{d}\big)^2}{2Np_z(1 - p_z)}\right] + O(n^{-100}). \end{aligned}$$

As $\bar{\mu}, v$ are such that $n^{-1/2}(x, y) \in \mathcal{I}$ for a compact set $\mathcal{I}$, the first summand decays exponentially as z grows. Indeed, there is a number $\gamma_2 = \gamma_2(d, \mathcal{I}, \mathcal{J})$ such that

$$f(z) \leqslant \gamma_2 \exp(-z^2/\gamma_2) + O(n^{-100}), \quad \text{for all } z.$$

Hence, there exists $\gamma_3 = \gamma_3(d, \mathcal{I}, \mathcal{J})$ such that $\|f\|_1, \|f\|_2 \leqslant \gamma_3$. Setting $\gamma_0 = \gamma_1 + \gamma_2 + \gamma_3$ completes the proof of assertion (i).

Let $\gamma > 0$ be arbitrarily large but fixed as n grows, and consider $z \in [-\gamma, \gamma]$. We are left to prove that $|f(z) - F(z)| = o(1)$. To this end, all we need to do is to plug in the explicit expressions for the two factors on the right-hand side of (4.11) and simplify. Let

$$\mu_{\mathcal{N},z} = (1 - \rho_z)n, \quad \sigma_{\mathcal{N},z} = \frac{\sqrt{\rho_z\big(1 - \rho_z + c_z(d-1)(\rho_z - \rho_z^{d-1})\big)n}}{1 - c_z(d-1)\rho_z^{d-1}},$$

so that $\sigma_{\mathcal{N}} = \sigma_{\mathcal{N},0}$. Then Theorem 2.3 implies that

$$\mathbb{P}\big[\mathcal{N}(H_d(n, p_z)) = v\big] \sim \frac{1}{\sqrt{2\pi}\sigma_{\mathcal{N},z}} \exp\left(-\frac{(v - \mu_{\mathcal{N},z})^2}{2\sigma_{\mathcal{N},z}^2}\right). \tag{4.12}$$

Since we know from Lemma 4.1 that

$$1 - \rho_z = 1 - \rho + z \cdot \frac{\lambda \sigma_{\mathcal{N}}}{n} + o(n^{-1/2}),$$

(4.12) yields

$$\mathbb{P}\left[\mathcal{N}(H_d(n, p_z)) = v\right] \sim \frac{1}{\sqrt{2\pi}\sigma_{\mathcal{N}}} \exp\left(-\frac{(v - (1-\rho)n - z\lambda\sigma_{\mathcal{N}})^2}{2\sigma_{\mathcal{N}}^2}\right). \tag{4.13}$$

Further, since $v = (1-\rho)n + x$ by (3.1) and as $x = O(\sqrt{n})$, we have

$$\begin{aligned}
Np_z(1-p_z) = Np_z + O(1) &= \binom{n-v}{d}\left[p + z\sigma\binom{n}{d}^{-1}\right] \\
&= \frac{(n-v)^d}{d!}\left[p + z\sigma\binom{n}{d}^{-1}\right] + O(1) \\
&= (\rho - x/n)^d \binom{n}{d}\left[p + z\sigma\binom{n}{d}^{-1}\right] + O(1) \\
&= (\rho - x/n)^d\left[m_0 + z\sigma\right] + O(1) = \left(\rho^d - \frac{d\rho^{d-1}x}{n}\right)\left[m_0 + z\sigma\right] + O(1) \\
&= \rho^d(m_0 + z\sigma - c\rho^{-1}x) + O(1).
\end{aligned}$$

Hence, Proposition 2.1 entails

$$\begin{aligned}
\mathbb{P}\left[\mathrm{Bin}(N, p_z) = \bar{\mu}\right] &= \frac{1}{\sqrt{2\pi N p_z(1-p_z)}} \exp\left[-\frac{\bar{\mu} - Np_z}{2Np_z(1-p_z)}\right] \\
&\sim \frac{1}{\sqrt{2\pi\rho^d m_0}} \exp\left(-\frac{(\bar{\mu} - \rho^d(m_0 + z\sigma - c\rho^{-1}x))^2}{2\rho^d m_0}\right).
\end{aligned} \tag{4.14}$$

Plugging (4.13) and (4.14) into (4.11), we obtain

$$\begin{aligned}
\frac{f(z)}{n} &\sim \mathbb{P}\left[\mathcal{N}(H_d(n, p_z)) = v\right]\mathbb{P}\left[\mathrm{Bin}(N, p_z) = \bar{\mu}\right] \\
&\sim \frac{1}{2\pi\sqrt{\rho^d m_0}\sigma_{\mathcal{N}}} \exp\left(-\frac{(v - (1-\rho)n - z\lambda\sigma_{\mathcal{N}})^2}{2\sigma_{\mathcal{N}}^2} - \frac{(\bar{\mu} - \rho^d(m_0 + z\sigma - c\rho^{-1}x))^2}{2\rho^d m_0}\right) \\
&\sim \frac{1}{2\pi\rho^{d/2}\sigma\sigma_{\mathcal{N}}} \exp\left(-\frac{(x - z\lambda\sigma_{\mathcal{N}})^2}{2\sigma_{\mathcal{N}}^2} - \frac{(-y - \rho^d\sigma z + c\rho^{d-1}x)^2}{2\rho^d\sigma^2}\right) \\
&= \frac{F(z)}{n},
\end{aligned}$$

thereby completing the proof. □

5. Analysis of g: proof of Proposition 3.2

Throughout this section, we keep the notation and the assumptions from Section 3. Let $0 < \alpha < 0.1$ *be given. We will always assume that* $n > n_1$ *for some large enough* $n_1 = n_1(\alpha)$.

The function g is a bit ‘unwieldy’ because it is defined in terms of the random hypergraph $H_d(n, m_z)$ with a *fixed* number of edges. To prove Proposition 3.2, we are going to represent g in terms of $H_d(n, p_z)$ instead. As a first step, we are going to express

$g(z)$ in terms of the number $C_d(v, m_z - \bar{\mu})$ of connected d-uniform hypergraphs of order v and size $m_z - \bar{\mu}$ for $z \in [-z^*, z^*]$. To this end, we use a similar argument as in [22].

Lemma 5.1. *Uniformly for $z \in [-z^*, z^*]$, we have*

$$g(z) = (1 + O(n^{-100}))n\binom{n}{v}\binom{\binom{n-v}{d}}{\bar{\mu}} C_d(v, m_z - \bar{\mu}) \Big/ \binom{\binom{n}{d}}{m_z}.$$

Proof of Lemma 5.1. We claim that

$$\frac{g(z)}{n} \leqslant \binom{n}{v} C_d(v, m_z - \bar{\mu}) \binom{\binom{n-v}{d}}{\bar{\mu}} \binom{\binom{n}{d}}{m_z}^{-1}. \tag{5.1}$$

The reason is that $g(z)/n$ is the probability that the largest component of $H_d(n, m_z)$ has order v and size $m_z - \bar{\mu}$, while the right-hand side equals the *expected* number of such components. There are $\binom{n}{v}$ ways to choose v vertices on which to place such a component. Then, there are $C_d(v, m_z - \bar{\mu})$ ways to choose the component itself. Moreover, there are $\binom{\binom{n-v}{d}}{\bar{\mu}}$ ways to choose the hypergraph induced on the remaining $n - v$ vertices, while the total number of d-uniform hypergraphs of order n and size m_z is $\binom{\binom{n}{d}}{m_z}$. Conversely,

$$\frac{g(z)}{n} \geqslant \binom{n}{v} C_d(v, m_z - \bar{\mu}) \binom{\binom{n-v}{d}}{\bar{\mu}} \mathbb{P}\big[\mathcal{N}(H_d(n - v, \bar{\mu})) < v\big] \binom{\binom{n}{d}}{m_z}^{-1}, \tag{5.2}$$

since the right-hand side equals the probability that $H_d(n, m_z)$ has one component of order v and size $m_z - \bar{\mu}$, while all other components have order $< v$. Since

$$\mathbb{P}\big[\mathcal{N}(H_d(n - v, \bar{\mu})) < v\big] = 1 - O(n^{-100})$$

by Lemma 4.1, the assertion follows from (5.1) and (5.2). □

To perform the transition from $H_d(n, m_z)$ to $H_d(n, p_z)$, we will now express $C_d(v, m_z - \bar{\mu})$ in terms of the probability of a certain event in $H_d(n, p_z)$.

Lemma 5.2. *Let $z, z' \in [-z^*, z^*]$ and set*

$$\mathcal{P}(z, z') = \mathbb{P}\big[\mathcal{N}(H_d(n, p_{z'})) = v,\ \mathcal{M}(H_d(n, p_{z'})) = m_z - \bar{\mu}\big].$$

Then

$$\mathcal{P}(z, z') = (1 - O(n^{-100}))\binom{n}{v} p_{z'}^{m_z - \bar{\mu}} (1 - p_{z'})^{\binom{n}{d} - \binom{n-v}{d} - (m_z - \bar{\mu})} \cdot C_d(v, m_z - \bar{\mu}).$$

Proof. We observe that

$$\mathcal{P}(z, z') \leqslant \binom{n}{v} C_d(v, m_z - \bar{\mu}) p_{z'}^{m_z - \bar{\mu}} (1 - p_{z'})^{\binom{n}{d} - \binom{n-v}{d} - (m_z - \bar{\mu})}, \tag{5.3}$$

because the right-hand side equals the *expected* number of components of order v and size $m_z - \bar{\mu}$ in $H_d(n, p_{z'})$. (There are $\binom{n}{v}$ ways to choose the v vertices on which to place the component and $C_d(v, m_z - \bar{\mu})$ ways to choose the component itself. Furthermore, edges are present with probability $p_{z'}$ independently, and thus the $p_{z'}^{m_z-\bar{\mu}}$ factor accounts for the presence of the $m_z - \bar{\mu}$ desired edges among the selected v vertices. Moreover, the $(1 - p_{z'})$-factor rules out further edges among the v chosen vertices and between the v chosen and the $n - v$ remaining vertices.) Conversely,

$$\mathcal{P}(z,z') \geqslant \binom{n}{v} C_d(v, m_z - \bar{\mu}) p_{z'}^{m_z-\bar{\mu}} (1-p_{z'})^{\binom{n}{d}-\binom{n-v}{d}-(m_z-\bar{\mu})} \cdot \mathbb{P}\big[\mathcal{N}(H_d(n-v, p_{z'})) < v\big], \quad (5.4)$$

since the right-hand side is the probability that there occurs exactly one component of order v and size $m_z - \bar{\mu}$, while all other components have order $< v$. As Lemma 4.1 entails that

$$\mathbb{P}\big[\mathcal{N}(H_d(n-v, p_{z'}) < v)\big] = 1 - O(n^{-100}),$$

(5.3) and (5.4) yield

$$\mathcal{P}(z,z') = (1 - O(n^{-100})) \binom{n}{v} C_d(v, m_z - \bar{\mu}) p_{z'}^{m_z-\bar{\mu}} (1-p_{z'})^{\binom{n}{d}-\binom{n-v}{d}-(m_z-\bar{\mu})},$$

as claimed. □

Combining Lemmas 5.1 and 5.2, we obtain the following representation of $g(z)$ in terms of $H_d(n, p_z)$.

Corollary 5.3. *Let $z, z' \in [-z^*, z^*]$. Then*

$$g(z) = (1 + O(n^{-100})) \cdot n\mathcal{P}(z,z') \cdot \frac{\mathbb{P}\big[\mathrm{Bin}\big(\binom{n-v}{d}, p_{z'}\big) = \bar{\mu}\big]}{\mathbb{P}\big[\mathrm{Bin}\big(\binom{n}{d}, p_{z'}\big) = m_z\big]}.$$

Lemma 5.4. *We have $\sup_{z\in[-z^*,z^*]} \mathcal{P}(z,z) = O(1/n)$.*

We defer the proof of Lemma 5.4 to Section 5.1. From Corollary 5.3 it is relatively straightforward to obtain the following bound on g.

Corollary 5.5. *There is a number $\gamma_0 = \gamma_0(d, \mathcal{I}, \mathcal{J})$ such that*

$$g(z) \leqslant \gamma_0 \cdot \exp(-z^2/\gamma_0), \quad \textit{for all } z \in \mathbb{R}.$$

Proof. With $z = z'$, Corollary 5.3 and Proposition 2.1 yield

$$\begin{aligned} g(z) &= (1 + O(n^{-100})) \cdot n\mathcal{P}(z,z) \cdot \frac{\mathbb{P}\big[\mathrm{Bin}\big(\binom{n-v}{d}, p_z\big) = \bar{\mu}\big]}{\mathbb{P}\big[\mathrm{Bin}\big(\binom{n}{d}, p_z\big) = m_z\big]} \\ &\leqslant O(n) \cdot \mathcal{P}(z,z) \cdot \exp\left[-\frac{(\bar{\mu} - \binom{n-v}{d} p_z)^2}{2\binom{n-v}{d} p_z(1-p_z)}\right] \\ &\leqslant O(1) \cdot \exp\left[-\frac{(\bar{\mu} - \binom{n-v}{d} p_z)^2}{2\binom{n-v}{d} p_z(1-p_z)}\right] \quad \text{(by Lemma 5.4).} \end{aligned}$$

https://doi.org/10.1017/S0963548314000017 Published online by Cambridge University Press

Since $(\bar{\mu} - \binom{n-v}{d}p)/n^{1/2} = O(1)$ by our assumption on $\bar{\mu}$ and because

$$\binom{n-v}{d}p_z - \binom{n-v}{d}p \sim z\rho^d\sigma,$$

we obtain

$$g(z) \leqslant O(1)\exp\left[-\frac{(1+o(1))z^2\sigma^2\rho^{2d}}{2\binom{n-v}{d}p_z(1-p_z)}\right] \leqslant O(1)\cdot\exp\left[-(1+o(1))z^2\rho^{2d}/2\right],$$

as desired. □

To establish the asymptotic continuity of g at 0 (part (iii) of Proposition 3.2), we prove the following continuity statement for $\mathcal{P}$ in Section 5.1.

Lemma 5.6. *For any $\alpha > 0$, there exists $\beta > 0$ such that for all $z, z' \in [-\beta, \beta]$ we have*

$$n \cdot |\mathcal{P}(z, z') - \mathcal{P}(z', z')| \leqslant \alpha + o(1).$$

Proof of Proposition 3.2. Statements (i) and (ii) follow immediately from Corollary 5.5. With respect to claim (iii), Corollary 5.3 yields

$$\begin{aligned} g(z) - g(z') = \sqrt{n}\mathbb{P}\left[\operatorname{Bin}\left(\binom{n-v}{d}, p_{z'}\right) = \bar{\mu}\right] \\ \cdot \left[\frac{n\mathcal{P}(z,z')}{\sqrt{n}\mathbb{P}\left[\operatorname{Bin}\left(\binom{n}{d}, p_{z'}\right) = m_z\right]} - \frac{n\mathcal{P}(z',z')}{\sqrt{n}\mathbb{P}\left[\operatorname{Bin}\left(\binom{n}{d}, p_{z'}\right) = m_{z'}\right]}\right] + o(1). \end{aligned}$$

Thus, by Proposition 2.1,

$$\begin{aligned} g(z) - g(z') &= \sqrt{2\pi\sigma} \cdot \mathbb{P}\left[\operatorname{Bin}\left(\binom{n-v}{d}, p_{z'}\right) = \bar{\mu}\right] \\ &\quad \cdot \left[n\mathcal{P}(z,z')\exp\left(\frac{(z-z')^2}{2}\right) - n\mathcal{P}(z',z')\right] + o(1) \\ &\leqslant C \cdot \left[n\mathcal{P}(z,z')\exp\left(\frac{(z-z')^2}{2}\right) - n\mathcal{P}(z',z')\right] + o(1). \end{aligned}$$

for a certain number $C = C(d, \mathcal{I}, \mathcal{J}) > 0$. Hence, the assertion follows from Lemma 5.6. □

5.1. Sprinkling: proof of Lemmas 5.4 and 5.6

Let $z, z' \in [-z^*, z^*]$. Let $\varepsilon > 0$ be a small enough number that remains fixed as $n \to \infty$. Moreover, set $q_1 = (1-\varepsilon)p_{z'}$, and let $q_2 \sim \varepsilon p_{z'}$ be such that $q_1 + q_2 - q_1q_2 = p_{z'}$. Choosing $\varepsilon > 0$ sufficiently small, we can ensure that $\binom{n-1}{d-1}q_1 > (d-1)^{-1} + \varepsilon$. Now, we construct $H_d(n, p_{z'})$ in three rounds, as follows.

- (R1) Construct a random hypergraph H_1 with vertex set $V = \{1, \ldots, n\}$ by including each of the $\binom{n}{d}$ possible edges with probability q_1 independently. Let G_1 be the largest component of H_1.
- (R2) Let $H_2 \supset H_1$ be the hypergraph obtained by adding with probability q_2 independently each possible edge $e \notin H_1$ that is not entirely contained in G_1 (*i.e.*, $e \not\subset G_1$) to H_1. Let G_2 signify the largest component of H_2.

https://doi.org/10.1017/S0963548314000017 Published online by Cambridge University Press

(R3) Finally, obtain $H_3 \supset H_2$ by adding each edge $e \notin H_1$ such that $e \subset G_1$ with probability q_2 independently. Let $\mathcal{F}$ denote the set of edges added in this way.

Since for each of the $\binom{n}{d}$ possible edges the overall probability of being contained in H_3 is $q_1 + (1-q_1)q_2 = p_{z'}$, H_3 is just a random hypergraph $H_d(n, p_{z'})$. Moreover, since in (R3) we only add edges that fall completely into the component of H_2 that contains G_1, we have

$$\mathcal{N}(H_d(n, p_{z'})) = \mathcal{N}(H_3) = \mathcal{N}(H_2).$$

Furthermore, $|\mathcal{F}|$ has a binomial distribution

$$|\mathcal{F}| = \mathrm{Bin}\left(\binom{|G_1|}{d} - \mathcal{M}(H_1), q_2\right). \tag{5.5}$$

We are going to apply the local limit theorem for the binomially distributed $|\mathcal{F}|$ (Proposition 2.1). Loosely speaking, we shall observe that most likely G_1 is contained in the largest component of H_3. If this is indeed the case, then $\mathcal{M}(H_3) = |\mathcal{F}| + \mathcal{M}(H_2)$, and therefore

$$\mathcal{M}(H_3) = m_z - \mu \quad \Leftrightarrow \quad |\mathcal{F}| = m_z - \mu - \mathcal{M}(H_2). \tag{5.6}$$

Finally, since

$$\mathbb{P}\big[|\mathcal{F}| = m_{z'} - \mu - \mathcal{M}(H_2)\big] \quad \text{and} \quad \mathbb{P}\big[|\mathcal{F}| = m_z - \mu - \mathcal{M}(H_2)\big]$$

will turn out to be 'close' for $|z - z'|$ small (by the local limit theorem for the binomial distribution), we will find that the same is true of $\mathcal{P}(z, z')$ and $\mathcal{P}(z', z')$.

To implement this sketch, let $\mathcal{Q}(\alpha', z, z')$ be the set of all pairs $(\mathcal{H}_1, \mathcal{H}_2)$ of hypergraphs that satisfy the following three conditions.

(Q1) $\mathcal{N}(\mathcal{H}_2) = v$.
(Q2) $\mathbb{P}\big[\mathcal{M}(H_3) = m_z - \mu | H_1 = \mathcal{H}_1, H_2 = \mathcal{H}_2\big] \geqslant \alpha' n^{-1/2}$.
(Q3) The largest component of $\mathcal{H}_2$ contains the largest component of $\mathcal{H}_1$.

The next lemma shows that the processes (R1)–(R3) such that $(H_1, H_2) \in \mathcal{Q}(\alpha', z, z')$ dominate.

Lemma 5.7. *For any $\alpha > 0$, there exists $\alpha' > 0$ such that for all $z, z' \in [-z^*, z^*]$ the following is true. Let*

$$\Pi(\alpha', z, z') = \mathbb{P}\big[\mathcal{M}(H_3) = m_z - \mu,\ (H_1, H_2) \in \mathcal{Q}(\alpha', z, z')\big].$$

Then

$$\Pi(\alpha', z, z') \leqslant \mathcal{P}(z, z') \leqslant \Pi(\alpha', z, z') + \alpha/n + o(1/n).$$

Proof of Lemma 5.7. The inequality $\Pi(\alpha', z, z') \leqslant \mathcal{P}(z, z')$ is immediate from the definitions. To obtain the second inequality, let $\mathcal{R}$ signify the set of all pairs $(\mathcal{H}_1, \mathcal{H}_2)$ such that (Q1) is satisfied. Since $H_3 = H_d(n, p_{z'})$, we have

$$\mathcal{P}(z, z') = \mathbb{P}\big[\mathcal{M}(H_3) = m_{z'} - \mu,\ (H_1, H_2) \in \mathcal{R}\big].$$

https://doi.org/10.1017/S0963548314000017 Published online by Cambridge University Press

Therefore, letting $\bar{\mathcal{Q}}_2$ (resp. $\bar{\mathcal{Q}}_3$) denote the set of all $(\mathcal{H}_1, \mathcal{H}_2) \in \mathcal{R}$ that violate (Q2) (resp. (Q3)), we have

$$\begin{aligned}
&\mathcal{P}(z,z') - \Pi(\alpha',z,z') \\
&\quad \leqslant \mathbb{P}\big[\mathcal{M}(H_3) = m_z - \mu,\ (H_1,H_2) \in \mathcal{R} \setminus \mathcal{Q}(\alpha',z,z')\big] \\
&\quad \leqslant \mathbb{P}\big[\mathcal{M}(H_3) = m_z - \mu | (H_1,H_2) \in \bar{\mathcal{Q}}_2\big] \cdot \mathbb{P}\big[\mathcal{N}(H_2) = v\big] + \mathbb{P}\big[(H_1,H_2) \in \bar{\mathcal{Q}}_3\big] \\
&\quad \overset{(\mathrm{Q2})}{\leqslant} \alpha' n^{-1/2} \cdot \mathbb{P}\big[\mathcal{N}(H_2) = v\big] + \mathbb{P}\big[(H_1,H_2) \in \bar{\mathcal{Q}}_3\big] \\
&\quad \leqslant \alpha/n + \mathbb{P}\big[(H_1,H_2) \in \bar{\mathcal{Q}}_3\big] \quad \text{(by Theorem 2.3)}, \qquad (5.7)
\end{aligned}$$

provided that $\alpha' \leqslant \alpha/C$ for some large enough number $C = C(\mathcal{I}, \mathcal{J})$. Further, if $(H_1, H_2) \in \bar{\mathcal{Q}}_3$, then either H_1 does not feature a component of order $\Omega(n)$, or H_2 has two such components. Since $\binom{n-1}{d-1} q_1 > (d-1)^{-1} + \varepsilon$ due to our choice of $\varepsilon > 0$, Theorem 2.2 entails that the probability of either event is $O(n^{-100})$. Thus, the assertion follows from (5.7). □

Lemma 5.8. *For any $\alpha, \alpha' > 0$, there exists $\beta = \beta(\alpha) > 0$ such that for any $z, z', z'' \in [-\beta, \beta]$ we have*

$$\Pi(\alpha', z, z') \leqslant (1+\alpha)\mathcal{P}(z'', z').$$

Proof of Lemma 5.8. Consider $(\mathcal{H}_1, \mathcal{H}_2) \in \mathcal{Q}(\alpha', z, z')$ and let us condition on the event $(H_1, H_2) = (\mathcal{H}_1, \mathcal{H}_2)$. Let $N_1 = \binom{v}{d} - \mathcal{M}(H_1)$. Moreover, let $\Delta = m_z - \mu - \mathcal{M}(H_2)$, $\Delta'' = m_{z''} - \mu - \mathcal{M}(H_2)$. We claim that there is a number $\omega = \omega(\alpha')$ such that

$$|N_1 q_2 - \Delta| \leqslant \omega n^{1/2}. \qquad (5.8)$$

Indeed, if $|N_1 q_2 - \Delta| > \omega n^{1/2}$ for a large $\omega = \omega(\alpha')$, then

$$\begin{aligned}
&\mathbb{P}\big[\mathcal{M}(H_3) = m_z - \mu | (H_1,H_2) = (\mathcal{H}_1,\mathcal{H}_2)\big] \\
&\quad \overset{(5.6)}{=} \mathbb{P}\big[|\mathcal{F}| = \Delta | (H_1,H_2) = (\mathcal{H}_1,\mathcal{H}_2)\big] \\
&\quad \leqslant \exp\big[-\omega^2/3\big] n^{-1/2} \quad \text{(by Proposition 2.1 and (5.5))} \\
&\quad < \alpha' n^{-1/2},
\end{aligned}$$

in contradiction to (Q2). Thus, there exists $\beta = \beta(\alpha, \alpha') > 0$ such that, for $|z - z'| < \beta$, Proposition 2.1 yields

$$\begin{aligned}
\frac{\mathbb{P}\big[|\mathcal{F}| = \Delta | (H_1,H_2) = (\mathcal{H}_1,\mathcal{H}_2)\big]}{\mathbb{P}\big[|\mathcal{F}| = \Delta'' | (H_1,H_2) = (\mathcal{H}_1,\mathcal{H}_2)\big]} &\overset{(5.5)}{=} \frac{\mathbb{P}\big[\mathrm{Bin}\big(N_1,q_2\big) = m_z - \mu - \mathcal{M}(H_2)\big]}{\mathbb{P}\big[\mathrm{Bin}\big(N_1,q_2\big) = m_{z''} - \mu - \mathcal{M}(H_2)\big]} \\
&\sim \exp\left[\frac{\big(N_1 q_2 - \Delta''\big)^2 - \big(N_1 q_2 - \Delta\big)^2}{2N_1 q_2(1-q_2)}\right] \\
&\sim \exp\left[\frac{2(\Delta - \Delta'')(N_1 q_2 - \Delta) + (\Delta - \Delta'')^2}{2N_1 q_2(1-q_2)}\right]
\end{aligned}$$

https://doi.org/10.1017/S0963548314000017 Published online by Cambridge University Press

$$\sim \exp\left[\frac{2\sigma(z-z'')(N_1q_2-\Delta)+\sigma^2(z-z'')^2}{2N_1q_2(1-q_2)}\right]$$
$$\leqslant \exp\left[\frac{2\sigma\sqrt{n}\omega\beta+\sigma^2\beta^2}{\Omega(n)}\right] \leqslant 1+\alpha.$$

Consequently,

$$\begin{aligned}
&\Pi(\alpha', z, z') \\
&= \sum_{(\mathcal{H}_1,\mathcal{H}_2)\in\mathcal{Q}(\alpha',z,z')} \mathbb{P}\big[\mathcal{M}(H_3)=m_z-\mu|(H_1,H_2)=(\mathcal{H}_1,\mathcal{H}_2)\big]\cdot\mathbb{P}\big[(H_1,H_2)=(\mathcal{H}_1,\mathcal{H}_2)\big] \\
&= \sum_{(\mathcal{H}_1,\mathcal{H}_2)\in\mathcal{Q}(\alpha',z,z')} \mathbb{P}\big[|\mathcal{F}|=\Delta|(H_1,H_2)=(\mathcal{H}_1,\mathcal{H}_2)\big]\cdot\mathbb{P}\big[(H_1,H_2)=(\mathcal{H}_1,\mathcal{H}_2)\big] \\
&\leqslant (1+\alpha)\sum_{(\mathcal{H}_1,\mathcal{H}_2)\in\mathcal{Q}(\alpha',z,z')} \mathbb{P}\big[|\mathcal{F}|=\Delta''|(H_1,H_2)=(\mathcal{H}_1,\mathcal{H}_2)\big]\cdot\mathbb{P}\big[(H_1,H_2)=(\mathcal{H}_1,\mathcal{H}_2)\big] \\
&\leqslant (1+\alpha)\mathcal{P}(z'',z'),
\end{aligned}$$

as claimed. □

Proof of Lemmas 5.4 and 5.6. Let $z, z' \in [-z^*, z^*]$. Let $(\mathcal{H}_1,\mathcal{H}_2)\in\mathcal{Q}(\alpha',z,z')$ for a small enough $\alpha'>0$. Then (5.5) and (5.6) together with Proposition 2.1 imply that

$$\begin{aligned}
&\mathbb{P}\big[\mathcal{M}(H_3)=m_z-\mu|(H_1,H_2)=(\mathcal{H}_1,\mathcal{H}_2)\big] \\
&\quad= \mathbb{P}\left[\mathrm{Bin}\left(\binom{v}{d}-\mathcal{M}(H_1),q_2\right)=m_z-\mu-\mathcal{M}(H_2)\right] \\
&\quad= O(n^{-1/2}).
\end{aligned}$$

Hence, by Theorem 2.3

$$\begin{aligned}
\Pi(\alpha',z,z') &\leqslant O(n^{-1/2})\cdot\mathbb{P}\big[(H_1,H_2)\in\mathcal{Q}(\alpha',z,z')\big] \\
&\leqslant O(n^{-1/2})\cdot\mathbb{P}\big[\mathcal{N}(H_1)=v\big] = O(1/n).
\end{aligned}$$

Thus, Lemma 5.7 implies that $\mathcal{P}(z,z')=O(1/n)$, whence Lemma 5.4 follows.

Now, let $\alpha>0$. Let $\alpha'>0$ be the number from Lemma 5.7. Moreover, let $C=C(d,\mathcal{I},\mathcal{J})>0$ be sufficiently large that $\mathcal{P}(z,z')\leqslant C/n$ for all $z,z'\in[-z^*,z^*]$; such a C exists by our above proof of Lemma 5.4. Furthermore, by Lemma 5.8 there exists $\beta>0$ such that

$$\Pi(\alpha',z,z') \leqslant (1+\alpha/C)\mathcal{P}(z'',z'), \quad \text{for all } z,z',z''\in[-\beta,\beta]. \tag{5.9}$$

Then, for any $z,z',z''\in[-\beta,\beta]$ we have

$$\begin{aligned}
\mathcal{P}(z,z') &\leqslant \Pi(\alpha',z,z')+(\alpha+o(1))/n \quad \text{(Lemma 5.7)} \\
&\leqslant (1+\alpha/C)\mathcal{P}(z'',z')+(\alpha+o(1))/n \quad \text{(Lemma 5.8)} \\
&\leqslant \mathcal{P}(z'',z')+\frac{\alpha}{C}\cdot\mathcal{P}(z'',z')+(\alpha+o(1))/n \\
&\leqslant \mathcal{P}(z'',z')+(2\alpha+o(1))/n,
\end{aligned}$$

where the last step follows from our choice of C. Thus, we have established Lemma 5.6 as well. □

6. Proof of Theorem 1.1

We shall derive Theorem 1.1 from Theorem 1.3 and (3.2).

Suppose that $\nu = (1-\rho)n + x$ and $\mu = (1-\rho^d)\binom{n}{d}p + y$, where $n^{-1/2}(x, y) \in \mathcal{I}$. Let $\alpha > 0$ be arbitrarily small but fixed, and let $\Gamma = \Gamma(\alpha) > 0$ be a sufficiently large number. Moreover, set

$$\mathcal{P} = \mathbb{P}\big[\mathcal{N}(H_d(n,p)) = \nu,\ \mathcal{M}(H_d(n,p)) = \mu\big],$$
$$\mathcal{B}(m) = \mathbb{P}\left[\mathrm{Bin}\left(\binom{n}{d}, p\right) = m\right],$$
$$\mathcal{Q}(m) = \mathbb{P}\big[\mathcal{N}(H_d(n,m)) = \nu,\ \mathcal{M}(H_d(n,m)) = \mu\big].$$

Let $L = \ln^{0.9} n$. Then, letting m range over integers, we define

$$S_1 = \sum_{m:|m-m_0|\leqslant\Gamma\sigma} \mathcal{B}(m)\mathcal{Q}(m),$$
$$S_2 = \sum_{m:\Gamma\sigma<|m-m_0|\leqslant L\sqrt{n}} \mathcal{B}(m)\mathcal{Q}(m),$$
$$S_3 = \sum_{m:|m-m_0|>L\sqrt{n}} \mathcal{B}(m)\mathcal{Q}(m),$$

so that

$$\mathcal{P} = S_1 + S_2 + S_3. \tag{6.1}$$

We shall estimate the three summands S_1, S_2, S_3 separately.

Let us first deal with S_3. As $m_0 = O(n)$, the Chernoff bound (2.1) entails that

$$\sum_{m:|m-m_0|>L\sqrt{n}} \mathcal{B}(m) \leqslant n^{-2}.$$

Since $0 \leqslant \mathcal{Q}(m) \leqslant 1$, this implies

$$S_3 \leqslant n^{-2}. \tag{6.2}$$

To bound S_2, we need the following lemma.

Lemma 6.1. *There is a number $K = K(d, \mathcal{I}, \mathcal{J}) > 0$ such that $\mathcal{Q}(m) \leqslant Kn^{-1}$ for all m such that $|m - m_0| \leqslant L\sqrt{n}$.*

Proof of Lemma 6.1. Let $z = \sigma^{-1}(m - m_0)$, so that $m = m_z$. Then $|z| = O(L) = o(z^*)$, because $\sigma = \Omega(\sqrt{n})$. In addition, let $\bar{\mu}_m = m - \mu$, so that

$$\mathcal{Q}(m) = \mathbb{P}\big[\mathcal{N}(H_d(n,m)) = \nu,\ \bar{\mathcal{M}}(H_d(n,m)) = \bar{\mu}_m\big] = g(z)/n. \tag{6.3}$$

Let $c_z = dm_z/n = dm/n$. Then by Proposition 3.2 there exists $K = K(d, \mathcal{I}, \mathcal{J}) > 0$ such that $g(z) \leqslant K$. Thus, the assertion follows from (6.3). □

Choosing $\Gamma = \Gamma(d, \mathcal{I}, \mathcal{J}) > 0$ large enough, we can achieve that

$$\sum_{m:|m-m_0|>\Gamma\sigma} \mathcal{B}(m) \leqslant \alpha/K.$$

Therefore, Lemma 6.1 entails that

$$S_2 = \sum_{m:\Gamma\sigma<|m-m_0|\leqslant L\sqrt{n}} \mathcal{B}(m)\mathcal{Q}(m) \leqslant \alpha n^{-1}. \tag{6.4}$$

Concerning S_1, we employ Proposition 2.1 to obtain

$$\mathcal{B}(m) \sim \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(m-m_0)^2}{2\sigma^2}\right] \quad \text{if } |m-m_0| \leqslant \Gamma\sigma. \tag{6.5}$$

In addition, let $0 < \varrho_m < 1$ signify the unique number such that

$$\varrho_m = \exp\left(\frac{dm}{n}(\varrho_m^{d-1} - 1)\right).$$

Then Lemma 4.1 yields

$$\varrho_m = \rho + \Delta_m/n + o(n^{-1/2}), \quad \text{with } \Delta_m = -\frac{m-m_0}{\sigma} \cdot \sigma_{\mathcal{N}}\lambda.$$

Hence,

$$1 - \varrho_m^d = 1 - \rho^d - \Xi_m/m + o(n^{-1/2}), \quad \text{with } \Xi_m = \frac{m_0 d}{n}\Delta_m \rho^{d-1}.$$

Thus, Theorem 1.3 entails that $\mathcal{Q}(m) \sim \varphi(m)$, where

$$\begin{aligned}\varphi(m) = {} & \frac{1}{2\pi\sqrt{\tau_{\mathcal{N}}^2\tau_{\mathcal{M}}^2 - \tau_{\mathcal{N}\mathcal{M}}^2}} \cdot \exp\left[-\frac{\tau_{\mathcal{N}}^2\tau_{\mathcal{M}}^2}{2(\tau_{\mathcal{N}}^2\tau_{\mathcal{M}}^2 - \tau_{\mathcal{N}\mathcal{M}}^2)}\right. \\ & \left. \cdot \left(\frac{(x+\Delta_m)^2}{\tau_{\mathcal{N}}^2} - \frac{2\tau_{\mathcal{N}\mathcal{M}}(x+\Delta_m)(y+\Xi_m)}{\tau_{\mathcal{N}}^2\tau_{\mathcal{M}}^2} + \frac{(y+\Xi_m)^2}{\tau_{\mathcal{M}}^2}\right)\right].\end{aligned} \tag{6.6}$$

Now, combining (6.5) and (6.6), we can approximate the sum S_1 by an integral as follows:

$$\begin{aligned}S_1 &\sim \sum_{m:|m-m_0|\leqslant\Gamma\sigma} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(m-m_0)^2}{2\sigma^2}\right]\varphi(m) \\ &= o(1/n) + \frac{1}{\sqrt{2\pi}\sigma}\int_{m_0-\Gamma\sigma}^{m_0+\Gamma\sigma} \exp\left[-\frac{(m-m_0)^2}{2\sigma^2}\right]\varphi(m)dm \\ &\sim \int_{-\Gamma}^{\Gamma} \varphi(m_z)\phi(z)\,dz.\end{aligned} \tag{6.7}$$

Further, note that

$$\Delta_{m_z} = -z\sigma_{\mathcal{N}}\lambda = -z\Theta(\sqrt{n}) \quad \text{and} \tag{6.8}$$

$$\Xi_{m_z} = -z\sigma_{\mathcal{N}}\lambda\rho^{d-1}m_0 d/n = -z\sigma_{\mathcal{N}}\lambda\rho^{d-1}c = -z\Theta(\sqrt{n}). \tag{6.9}$$

https://doi.org/10.1017/S0963548314000017 Published online by Cambridge University Press

Since $\tau_{\mathcal{N}}, \tau_{\mathcal{M}}, \sqrt{\tau_{\mathcal{NM}}} = \Theta(\sqrt{n})$, the function $\varphi(m_z)$ decays exponentially as $z \to \infty$. Therefore, choosing Γ large enough, we can achieve that

$$\int_{\mathbb{R}\setminus[-\Gamma,\Gamma]} \varphi(m_z)\phi(z)\,dz < \alpha/n. \tag{6.10}$$

Combining (6.1), (6.2), (6.4), (6.7), and (6.10), we obtain

$$\left|\mathcal{P} - (1+o(1))\int_{-\infty}^{\infty} \varphi(m_z)\phi(z)\,dz\right| \leqslant 3\alpha/n.$$

Finally, we need to verify that $\int_{-\infty}^{\infty} \varphi(m_z)\phi(z)\,dz$ coincides with the expression $P(x,y)$ from Theorem 1.1. To carry this out, we set

$$\begin{aligned}
\theta(z) &= \frac{1}{2\pi\sqrt{\tau_{\mathcal{N}}^2\tau_{\mathcal{M}}^2 - \tau_{\mathcal{NM}}^2}},\\
\delta(z) &= \sigma_{\mathcal{N}}\lambda, \quad \xi(z) = \sigma_{\mathcal{N}}\lambda\rho^{d-1}c,\\
t_1(z) &= \tau_{\mathcal{N}}^{-2}, \quad t_2(z) = \frac{2\tau_{\mathcal{NM}}}{\tau_{\mathcal{N}}^2\tau_{\mathcal{M}}^2}, \quad t_3(z) = \tau_{\mathcal{M}}^{-2}, \quad v(z) = \frac{\tau_{\mathcal{N}}^2\tau_{\mathcal{M}}^2}{2(\tau_{\mathcal{N}}^2\tau_{\mathcal{M}}^2 - \tau_{\mathcal{NM}}^2)},\\
u_0(z) &= t_1^2(z)x^2 + t_2(z)xy + t_3(z)y^2,\\
u_1(z) &= 2t_1(z)\delta x + t_2\xi(z)x + t_2(z)\delta y + 2t_3(z)\xi y,\\
u_2(z) &= t_1(z)\delta^2 + t_2(z)\delta\xi + t_3(z)\xi^2 - \frac{1}{2v(z)}.
\end{aligned}$$

Whereas the coefficients $u_0(z), u_1(z), u_2(z), v(z)$ vary with z, these variations are negligible in the limit $n \to \infty$: we have $u_j(z) \sim u_j(0)$ and $v(z) \sim v(0)$ uniformly for $|z| \leqslant z^*$. Therefore, plugging (6.8) and (6.9) into (6.6) yields

$$\begin{aligned}
\varphi(m_z)\phi(z) &\sim \theta \exp\left[-v\left(z^2u_2 - zu_1 + u_0\right)\right]\\
&\sim \theta \exp\left[-v\left(u_2\left(z - \frac{u_1}{2u_2}\right)^2 - \frac{u_1^2}{4u_2} + u_0\right)\right].
\end{aligned}$$

Hence, with

$$\theta' = \theta\int_{\mathbb{R}} \exp\left[-u_2v\left(z - \frac{u_1}{2u_2}\right)^2\right]dz \sim \sqrt{\frac{\pi}{u_2v}}\,\theta,$$

we obtain

$$\int_{\mathbb{R}} \varphi(m_z)\phi(z)\,dz \sim \theta' \exp\left[-v\left(u_0 - \frac{u_1^2}{4u_2}\right)\right].$$

To complete the proof, we just need to check that the exponent

$$-v\left(u_0 - \frac{u_1^2}{4u_2}\right),$$

viewed as a quadratic function of x, y, matches the exponent in (1.3). This is an elementary, albeit tedious, matter of algebraic manipulations.[2]

[2] Full details can be found in [5, Chapter 4].

https://doi.org/10.1017/S0963548314000017 Published online by Cambridge University Press

Acknowledgement

We thank Johannes Michaliček for helpful discussions on the use of Fourier analysis for proving Theorem 1.1. We also owe a great dept to an anonymous reviewer, whose suggestions have led to a substantially improved presentation.

References

[1] Alon, N. and Spencer, J. (2000) *The Probabilistic Method*, second edition, Wiley.
[2] Andriamampianina, T. and Ravelomanana, V. (2005) Enumeration of connected uniform hypergraphs. In *Proc. FPSAC 2005*.
[3] Barbour, A. D., Karoński, M. and Ruciński, A. (1989) A central limit theorem for decomposable random variables with applications to random graphs. *J. Combin. Theory Ser. B* **47** 125–145.
[4] Barraez, D., Boucheron, S. and Fernandez de la Vega, W. (2000) On the fluctuations of the giant component. *Combin. Probab. Comput.* **9** 287–304.
[5] Behrisch, M. (2007) Stochastical models for networks in the life sciences. PhD thesis, Humboldt Universität zu Berlin.
[6] Behrisch, M., Coja-Oghlan, A. and Kang, M. (2010) The order of the giant component of random hypergraphs. *Random Struct. Alg.* **36** 149–184.
[7] Behrisch, M., Coja-Oghlan, A. and Kang, M. (2014) The asymptotic probability that a random d-uniform hypergraph is connected. *Combin. Probab. Comput.* doi:10.1017/S0963548314000029.
[8] Bender, E. A., Canfield, E. R. and McKay, B. D. (1990) The asymptotic number of labeled connected graphs with a given number of vertices and edges. *Random Struct. Alg.* **1** 127–169.
[9] Bender, E. A., Canfield, E. R. and McKay, B. D. (1992) Asymptotic properties of labeled connected graphs. *Random Struct. Alg.* **3** 183–202.
[10] Bollobás, B. (2001) *Random Graphs*, second edition, Cambridge University Press.
[11] Bollobás B. and Riordan, O. (2012) Asymptotic normality of the size of the giant component via a random walk. *J. Combin. Theory Ser. B* **102** 53–61.
[12] Bollobás, B. and Riordan, O. (2012) Asymptotic normality of the size of the giant component in a random hypergraph. *Random Struct. Alg.* **41** 441–450.
[13] Coja-Oghlan, A., Moore, C. and Sanwalani, V. (2007) Counting connected graphs and hypergraphs via the probabilistic method. *Random Struct. Alg.* **31** 288–329.
[14] Coppersmith, D., Gamarnik, D., Hajiaghayi, M. and Sorkin, G. B. (2004) Random MAX SAT, random MAX CUT, and their phase transitions. *Random Struct. Alg.* **24** 502–545.
[15] Erdős, P. and Rényi, A. (1959) On random graphs I. *Publicationes Math. Debrecen* **5** 290–297.
[16] Erdős, P. and Rényi, A. (1960) On the evolution of random graphs. *Publ. Math. Inst. Hungar. Acad. Sci.* **5** 17–61.
[17] van der Hofstad, R. and Spencer, J. (2006) Counting connected graphs asymptotically. *European J. Combin.* **27** 1294–1320.
[18] Janson, S., Łuczak, T. and Ruciński, A. (2000) *Random Graphs*, Wiley.
[19] Karoński, M. and Łuczak, T. (1997) The number of connected sparsely edged uniform hypergraphs. *Discrete Math.* **171** 153–168.
[20] Karoński, M. and Łuczak, T. (2002) The phase transition in a random hypergraph. *J. Comput. Appl. Math.* **142** 125–135.
[21] Łuczak, T. (1990) On the number of sparse connected graphs. *Random Struct. Alg.* **1** 171–173.
[22] O'Connell, N. (1998) Some large deviation results for sparse random graphs. *Probab. Theory Rel. Fields* **110** 277–285.
[23] Pittel, B. (1990) On tree census and the giant component in sparse random graphs. *Random Struct. Alg.* **1** 311–342.
[24] Pittel, B. and Wormald, N. C. (2003) Asymptotic enumeration of sparse graphs with a minimum degree constraint. *J. Combin. Theory Ser. A* **101** 249–263.

[25] Pittel, B. and Wormald, N. C. (2005) Counting connected graphs inside out. *J. Combin. Theory Ser. B* **93** 127–172.
[26] Ravelomanana, V. and Rijamamy, A. L. (2006) Creation and growth of components in a random hypergraph process. In *Computing and Combinatorics*, Vol. 4112 of *Lecture Notes in Computer Science*, Springer, pp. 350–359,
[27] Rudin, W. (1987) *Real and Complex Analysis*, third edition, McGraw-Hill.
[28] Schmidt-Pruzan, J. and Shamir, E. (1985) Component structure in the evolution of random hypergraphs. *Combinatorica* **5** 81–94.
[29] Stepanov, V. E. (1970) On the probability of connectedness of a random graph $g_m(t)$. *Theory Probab. Appl.* **15** 55–67.