

The component structure of dense random subgraphs of the hypercube.

Colin McDiarmid
Department of Statistics,
University of Oxford,
24 - 29 St Giles',
Oxford, OX1 3LB, UK.
cmcd@stats.ox.ac.uk

Alex Scott*
Mathematical Institute,
University of Oxford,
Radcliffe Observatory Quarter,
Woodstock Road,
Oxford, OX2 6GG, UK.
scott@maths.ox.ac.uk

Paul Withers
Mathematical Institute,
University of Oxford,
Radcliffe Observatory Quarter,
Woodstock Road,
Oxford, OX2 6GG, UK.
paul.n.withers@gmail.com

June 17, 2018

Abstract

Given $p \in (0, 1)$, we let $Q_p = Q_p^d$ be the random subgraph of the d -dimensional hypercube Q^d where edges are present independently with probability p . It is well known that, as $d \rightarrow \infty$, if $p > \frac{1}{2}$ then with high probability Q_p is connected; and if $p < \frac{1}{2}$ then with high probability Q_p consists of one giant component together with many smaller components which form the 'fragment'.

Here we fix $p \in (0, \frac{1}{2})$, and investigate the fragment, and how it sits inside the hypercube. In particular we give asymptotic estimates

*Supported by a Leverhulme Research Fellowship

for the mean numbers of components in the fragment of each size, and describe their asymptotic distributions and indeed their joint distribution, much extending earlier work of Weber.

1 Introduction

The hypercube $Q = Q^d$ is the graph with vertex set $\{0, 1\}^d$ and with two vertices adjacent when they differ in exactly one co-ordinate. Alternatively it can be considered as the graph on the power set of $[d] = \{1, 2, \dots, d\}$ in which two sets are adjacent when their symmetric difference is a singleton. We consider the random subgraph $Q_p = Q_p^d$ where the edges appear independently with fixed probability p , and examine the component structure as $d \rightarrow \infty$. We say that Q_p has a property *with high probability (or whp)* if the property holds with probability tending to 1 as $d \rightarrow \infty$, and Q_p has a property *with very high probability (or whp)* if it holds with probability $1 - e^{-\Omega(d)}$.

Burtin [10] considered dense subgraphs, and showed that for a fixed $p < 1/2$, whp Q_p is disconnected, and for a fixed $p > 1/2$, whp Q_p is connected. Erdős and Spencer [11] showed that for $p = 1/2$, Q_p is connected with probability tending to e^{-1} (see also Theorem 14.3 of Bollobás [4]). Also Weber [15] considered the dense case – we will discuss his work shortly. Ajtai, Komlós and Szemerédi [1] looked at sparse subgraphs and demonstrated that a phase transition occurs at $p = 1/d$; for $p = \lambda/d$ with $\lambda > 1$, whp the largest component of Q_p has size $\Omega(2^d)$ and the second largest has size $o(2^d)$, while for $\lambda < 1$ whp the largest component has size $o(2^d)$. Bollobás, Kohayakawa and Łuczak [5, 6, 7, 8] gave more detailed results around the phase transition at $p = 1/d$, and investigated the minimum degree, connectedness and the existence of a complete matching in the series of subgraphs of Q^d formed by adding edges randomly, one at a time. They showed that, almost surely, this graph process becomes connected at the same time as the disappearance of the last isolated vertex, and at this time a complete matching emerges. See [9] for more recent work concerning behaviour around the phase transition and for further references.

This paper looks at the sizes of the largest and other components of Q_p for a fixed p with $0 < p < 1/2$. These graphs Q_p will be disconnected with a single large component whp. Our methods are different from the methods used in the papers mentioned above. Note that we cannot expect some sort of elegant ‘symmetry rule’ as for Erdős-Rényi random graphs $G(n, p)$, where (roughly speaking), given the size of the largest component in a supercritical random graph $G(n, p)$, the rest of the graph looks like a subcritical $G(n', p')$,

see for example [13] section 5.6: the geometry of the cube makes life more interesting and complicated.

We denote the number of vertices in a graph G by $v(G)$, and call this the *size* of G ; and denote the number of edges by $e(G)$. In Q_p , we order the components by size (where components having the same size are ordered say by the position of the ‘smallest’ vertex of each component in some canonical ordering of the vertices). Denote the j -th component by \mathcal{L}_j and let $L_j = v(\mathcal{L}_j)$, the size of \mathcal{L}_j . The *giant* component is \mathcal{L}_1 . The *fragment* \mathcal{Z} is the graph formed by all the components other than \mathcal{L}_1 , and we let $Z = v(\mathcal{Z}) = 2^d - L_1$. Let X_t denote the number of components of Q_p of size t , and let $\mu_t = \mathbb{E}[X_t]$. Let $X = \sum_{t \geq 1} X_t$, the total number of components of Q_p . Finally let $q = 1 - p$.

Observe that $\mu_1 = (2q)^d$; and that $\mu_1 \rightarrow \infty$ as $d \rightarrow \infty$, since $2q > 1$. The quantity m_p defined by

$$m_p = \lfloor 1/\log_2(1/q) \rfloor \tag{1}$$

is central to our results. For an integer t , we have $2q^t \geq 1 \Leftrightarrow t \leq m_p$. In particular, we always have $m_p \geq 1$ since $2q > 1$; and $m_p \geq 2$ if and only if $2q^2 \geq 1$, that is $p \leq 1 - 1/\sqrt{2} \approx 0.29$.

Weber [15] showed that whp the fragment size Z satisfies $Z \sim \mu_1$, the second largest component size L_2 satisfies $L_2 = m_p$, and the number X_t of components of size t satisfies $X_t \sim \mu_t = \Theta(d^{t-1}(2q^t)^d)$ for each $t = 1, \dots, m_p$; and it follows that the total number X of components satisfies $X \sim \mu_1$ whp. We extend and amplify these results, presenting our results in five theorems. Weber’s results are contained within Theorems 1 and 4 below. (Weber later introduced also a probability for vertices to appear in the random subgraph of Q^d [16], but we do not pursue that extension here.)

Our first three theorems, Theorems 1, 2 and 3, concern the global behaviour of components in Q_p ; and the last two theorems which we present in this section, Theorems 4 and 5, concern their local behaviour (and are needed to prove the earlier ones). In Section 6, we will present our sixth and last main theorem, Theorem 17, concerning joint distributions of random variables like the X_t .

Throughout, we fix $0 < p < 1/2$ and let $q = 1 - p$. The first theorem describes the number X of components in Q_p , the size $Z = 2^d - L_1$ of the fragment, and the size L_2 of the second largest component. Note that, as $d \rightarrow \infty$, we have $d \ll \mu_1$ and so $\sqrt{d\mu_1} \ll \mu_1$.

Theorem 1. *For fixed $0 < p < 1/2$, the random graph $Q_p = Q_p^d$ satisfies the following.*

- (a) *Let Y be either the number X of components of Q_p , or the fragment size Z : then for each $\varepsilon > 0$, we have $|Y - \mu_1| < \varepsilon\sqrt{d\mu_1}$ whp.*

- (b) The second largest component size L_2 in Q_p satisfies $L_2 = m_p$ wvhp, where m_p is as in (1). Also, the mean and variance satisfy $|\mathbb{E}[L_2] - m_p| = e^{-\Omega(d)}$ and $\text{Var}(L_2) = e^{-\Theta(d)}$.

Our second theorem concerns how the fragment sits in Q^d . How much do the components of the fragment cluster together? How far is it typically from a fixed vertex to the fragment \mathcal{Z} of Q_p ? Given a vertex u in Q^d and $r > 0$, the r -ball $B_r(u)$ around u is the set of vertices v at graph distance at most r from u (in Q^d). Recall that, for $0 < \eta < 1$, the entropy $h(\eta)$ is defined to be $-\eta \log_2 \eta - (1 - \eta) \log_2 (1 - \eta)$, and it is strictly increasing on $(0, \frac{1}{2})$ with image $(0, 1)$. Let $\eta^* = \eta^*(p)$ be the unique solution to $h(\eta) = \log_2 \frac{1}{1-p}$ with $0 < \eta < \frac{1}{2}$. For example, if $p = \frac{1}{4}$ then $\eta^* \approx 0.08$.

Theorem 2. For fixed $0 < p < 1/2$, the random graph $Q_p = Q_p^d$ satisfies the following.

- (a) There exists $\delta = \delta(p) > 0$ such that wvhp each δd -ball in Q^d contains at most m_p vertices of the fragment.
- (b) For each $\varepsilon > 0$ there is $\gamma = \gamma(\varepsilon, p) > 0$ such that wvhp a proportion at most $e^{-\gamma d}$ of the vertices in Q^d are within distance $(\eta^* - \varepsilon)d$ of the fragment \mathcal{Z} , but all vertices are within distance $(\eta^* + \varepsilon)d$. (All distances are in Q^d .)

In part (a) above, clearly wvhp there are δd -balls containing at least m_p vertices of the fragment – consider for example any ball with centre in a component of size m_p . Thus the statement that wvhp no δd -ball in Q^d contains strictly more than m_p vertices of the fragment is saying strongly that the components of the fragment \mathcal{Z} do not cluster together in Q^d . For example, wvhp no two components of \mathcal{Z} of size $> m_p/2$ are within distance δn of each other; and more generally, for each $k = 2, \dots, m_p+1$, no k components of \mathcal{Z} of size $> m_p/k$ are all within distance δn of each other.

In part (b), many vertices in Q^d are at a short distance from the fragment \mathcal{Z} , including of course the vertices in \mathcal{Z} , but only a very small proportion of the total are at distance at most $(\eta^* - \varepsilon)d$. However, when $r = (\eta^* + \varepsilon)d$, wvhp every r -ball contains a vertex in \mathcal{Z} (and indeed contains $2^{\Omega(d)}$ vertices in \mathcal{Z}). Overall, the giant gets everywhere, and indeed the fragment is heavily outnumbered everywhere.

The next theorem amplifies part (a) of Theorem 1, concerning the number X of components and the fragment size Z . Recall first that, for two integer valued random variables Y and Y' , the *total variation distance* between their

distributions is given by

$$d_{TV}(Y, Y') = \frac{1}{2} \sum_k |\mathbb{P}(Y = k) - \mathbb{P}(Y' = k)|.$$

We use $d_{TV}(Y, \text{Po}(\lambda))$ to denote $d_{TV}(Y, Y')$ where Y' has the Poisson distribution $\text{Po}(\lambda)$ with mean λ . Our main interest and effort will be involved in bounding $d_{TV}(Y, \text{Po}(\lambda))$ for relevant random variables Y , where $\lambda = \mathbb{E}[Y]$. Also, given a (non-trivial) random variable $Y = Y_d$ we let Y^* denote the natural centred and rescaled version $(Y - \mathbb{E}[Y])/\sqrt{\text{Var}(Y)}$. It is well known that if Y_n is a sequence of random variables with mean λ_n , such that $d_{TV}(Y_n, \text{Po}(\lambda_n)) \rightarrow 0$ and $\lambda_n \rightarrow \infty$ as $n \rightarrow \infty$, then Y_n^* is asymptotically standard normal (that is, Y_n^* converges in distribution to a standard normal random variable): see inequality (1.39) in [2] for a stronger version of this result.

Theorem 3. *Fix $0 < p < 1/2$, and let $q = 1 - p$. In $Q_p = Q_p^d$, let Y either be the number X of components or be the fragment size Z . Then the following properties hold as $d \rightarrow \infty$.*

- (a) *Both $\lambda := \mathbb{E}[Y]$ and $\text{Var}(Y)$ are $(1 + O(dq^d)) \mu_1$.*
- (b) *Both $d_{TV}(Y, \text{Po}(\lambda))$ and $d_{TV}(Y, \text{Po}(\mu_1))$ are $O(dq^d)$, and Y^* is asymptotically standard normal.*

Note that since $\text{Var}(Y) \sim \mu_1$ by part (a), Chebyshev's inequality shows that $|Y - \lambda| \leq \omega(d) \sqrt{\mu_1}$ whp if $\omega(d) \rightarrow \infty$ as $d \rightarrow \infty$. This result may be compared with Theorem 1 part (a).

The final two theorems concern local behaviour. The first counts small components by size. It is needed in order to prove the earlier theorems. Recall that X_t is the number of components of size t in Q_p , and $\mu_t = \mathbb{E}[X_t]$. We noted earlier that $\mu_1 = (2q)^d$: it is not hard to give exact formulae also for μ_2 and μ_3 (assuming $d \geq 2$); namely

$$\mu_2 = (p/2q^2) d(2q^2)^d \quad \text{and} \quad \mu_3 = 2(p^2/q^4) d(d-1)(2q^3)^d \quad (2)$$

(see also the discussion following Theorem 5).

Theorem 4. *Fix $0 < p < \frac{1}{2}$, let $q = 1 - p$, and let $1 \leq t \leq m_p$. Then the following results concerning the number X_t of components of size t in $Q_p = Q_p^d$ hold, as $d \rightarrow \infty$.*

- (a) *$\mu_t = (1 + O(\frac{1}{d})) \frac{t^{t-2}}{t!} (\frac{p}{q^2})^{t-1} d^{t-1} (2q^t)^d$, and $\text{Var}(X_t) = (1 + O(d^t q^{td})) \mu_t$.*

(b) For each $\varepsilon > 0$, we have $|X_t - \mu_t| < \varepsilon\sqrt{d\mu_t}$ wvhp, and so also $|X_t - \mu_t| < \varepsilon\mu_t$ wvhp.

(c) $d_{TV}(X_t, \text{Po}(\mu_t)) = O(d^t q^{td})$, and X_t^* is asymptotically standard normal.

Observe from part (a) that $\mu_t = \Omega(d)$ (and indeed $\mu_t \gg d$ unless $t = 2$ and $p = 1 - 1/\sqrt{2}$, so $2q^2 = 1$), so the first half of part (b) above implies the second half. For a partial local limit result corresponding to part (c), see Proposition 13 at the end of Section 3.

These results help us to visualise the asymptotic disappearance of small components in Q_p as p increases from 0 to $1/2$. For each fixed p , there are wvhp a giant component and many small components of every size up to a maximum size m_p . In particular $\mu_t \rightarrow \infty$ as $d \rightarrow \infty$ for each $t \leq m_p$. Observe that m_p is large for small p and decreases to 1 as p increases to $1/2$. The typical number of components decreases exponentially as p increases and the size L_2 of a largest component of the fragment drops as m_p falls below each integer value. In particular, the last components of size 2 disappear as p increases past $1 - 1/\sqrt{2} \approx 0.29$ and the last isolated vertices disappear as p increases past $1/2$. We recall that $Q_{1/2}$ is connected with probability tending to e^{-1} as $d \rightarrow \infty$. Indeed, whp $Q_{1/2}$ consists of X isolated vertices and a connected component of $2^d - X$ vertices, where X has mean value 1 and asymptotic distribution $\text{Po}(1)$ (see [11]).

Ambient isomorphisms

We shall in fact prove a much finer and more detailed version of Theorem 4, namely Theorem 5, which uses a natural restricted version of isomorphism for subgraphs of the cube, so that we can consider also how components ‘sit’ in the host hypercube. We then deduce Theorem 4 from Theorem 5.

We call a graph a *cube subgraph* if it is a subgraph of the cube Q^d for some d . Let H be a connected cube subgraph. The *support* $S(H)$ is the set of indices i such that there is an edge xy in H with $x_i = 0$ and $y_i = 1$ (that is, H meets both top and bottom faces in the i -th coordinate direction). Call $|S(H)|$ the *span* of H , $\text{span}(H)$. Note that if H consists of a single vertex then $\text{span}(H) = 0$, and otherwise $\text{span}(H) \geq 1$. Indeed, if $v(H)$ is 1, 2 or 3 then $\text{span}(H) = v(H) - 1$, whereas for example if H is a 4-vertex path then $\text{span}(H)$ could be 2 or 3.

The *canonical copy* H^* of H is defined as follows. If H is a single vertex then its canonical copy is the graph Q^0 (consisting of a single vertex). Suppose that H has at least one edge, so $\text{span}(H) = s \geq 1$. Let ϕ be the increasing injection from $[s]$ to $[d]$ with image $S(H)$. Given $x = (x_1, x_2, \dots, x_d) \in Q^d$ let $\phi(x) = (x_{\phi(1)}, x_{\phi(2)}, \dots, x_{\phi(s)}) \in Q^s$. Then the vertices of the canonical copy H^* are the points $\phi(x)$ where x is a vertex of H ; and the edges of H^*

are the pairs $\phi(x)\phi(y)$ such that xy is an edge of H . (Note that the canonical copy is a subgraph of Q^s .) See figure 1 for an illustration.

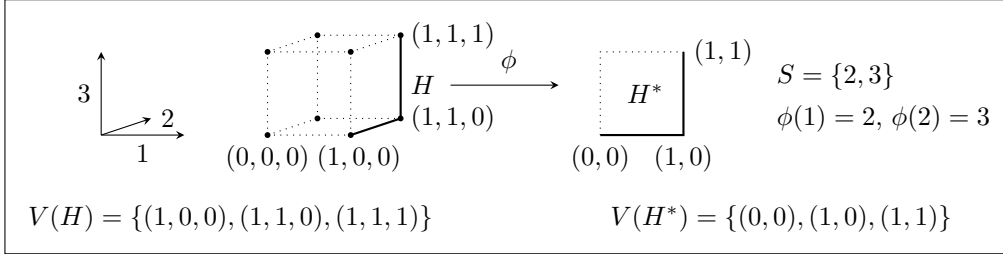


Figure 1: Example of a canonical copy in Q^3

We say that connected subgraphs H_1 of Q^{d_1} and H_2 of Q^{d_2} are *ambient-isomorphic* if they have the same canonical copy. Of course, if H_1 and H_2 are ambient-isomorphic then they are isomorphic, but this definition is stronger in that it requires the copies to ‘sit in the cube’ in the same way. For example, let O denote the zero d -vector and let e_k denote the k th unit d -vector: if $i < j$ then the three vertex path $O, e_i, e_i + e_j$ in Q^d has canonical copy the path $(0, 0), (1, 0), (1, 1)$ in Q^2 as in Figure 1, and so the original path in Q^d is not ambient isomorphic to the path $O, e_j, e_i + e_j$ which has canonical copy the path $(0, 0), (0, 1), (1, 1)$. Observe that if $s = \text{span}(H)$ then there is a unique subgraph of Q^s ambient-isomorphic to H (namely the canonical copy of H).

Our fifth theorem concerns numbers of components ambient isomorphic to given connected cube subgraphs H_i . Recall that $(d)_k$ means $d(d-1)\cdots(d-k+1)$.

Theorem 5. *Fix $0 < p < 1/2$, and let $q = 1 - p$. Let $r \geq 1$ and let H_1, H_2, \dots, H_r be pairwise non-ambient-isomorphic connected cube subgraphs each of size at most m_p . Let t be the minimum size $v(H_i)$ of a graph H_i , and let s be the maximum span of a graph H_i of size t . For each i , let $Y_i = Y_i(d)$ be the number of components of Q_p^d ambient-isomorphic to H_i . Finally, let $Y = Y(d) = \sum_i Y_i$ and let $\lambda = \lambda(d) = \mathbb{E}[Y]$. Then the following hold.*

- (a) *For the constant $c > 0$ given in equations (3) and (4) below, we have $\lambda = (1 + O(1/d))c(d)_s(2q^t)^d$; and if t is 1, 2 or 3 then $s = t - 1$, and we may replace the error bound $O(1/d)$ by $O(dq^d)$. Also $\text{Var}(Y) = (1 + O(d^t q^{td}))\lambda$.*
- (b) *For each $\varepsilon > 0$, we have $|Y - \lambda| < \varepsilon\sqrt{d\lambda}$ wvhp, and so also $|Y - \lambda| < \varepsilon\lambda$ wvhp.*
- (c) *$d_{TV}(Y, \text{Po}(\lambda)) = O(d^t q^{td})$, and Y^* is asymptotically standard normal.*

By part (a), λ is $\Omega(d)$ (and indeed λ is $\Omega(d^2)$ except if $t = 2$ and $p = 1 - 1/\sqrt{2}$), so the first half of part (b) implies the second half (as with Theorem 4). See Lemma 11 for a fuller version of Theorem 5, which considers also the numbers of vertices in the components counted. That lemma, together with the estimates of μ_t from Lemma 12 (a), will yield Theorem 4, by letting H_1, \dots, H_r list all the t -vertex connected canonical cube subgraphs, so that the random variable Y in Theorem 5 is X_t .

The constant c in part (a) may be specified as follows. Let $I^* = \{i \in [r] : v(H_i) = t, \text{span}(H_i) = s\}$. For each $i \in I^*$, recall that $e(H_i)$ is the number of edges of H_i , let $e'(H_i)$ be the number of edges of Q^d not in H_i but with both end vertices in H_i , and let

$$\beta_i = \frac{1}{2^s s!} \left(\frac{p}{q^2}\right)^{e(H_i)} \left(\frac{1}{q}\right)^{e'(H_i)} : \quad (3)$$

now we let

$$c = \sum_{i \in I^*} \beta_i. \quad (4)$$

If $t = 1$ then $c = 1$. If $t = 2$ then $c = p/2q^2$, so $\lambda \sim (p/2q^2) d(2q^2)^d$. If $t = 3$ then $1 \leq |I^*| \leq 4$ and each $\beta_i = \frac{1}{2}(p/q^2)^2$, so if $|I^*| = 4$ we have $\lambda \sim 2(p^2/q^4) d^2(2q^3)^d$. These results are in accord with (2).

By Theorem 4 we saw that wvhp in Q_p , there are components of each size up to m_p : now by Theorem 5 we see in much more detail that each connected cube subgraph of size at most m_p , with its way of sitting within the host hypercube, appears wvhp as a component of Q_p .

What we call ambient isomorphism could be called ‘ordered ambient isomorphism’, since we insist that the injection ϕ in the definition is increasing. If we drop this requirement then essentially the same results hold (*mutatis mutandis*), since new isomorphism classes are unions of old ones. When we deduce Theorem 4 from Theorem 5/Lemma 11, we may think of this as relaxing *ambient isomorphism* all the way to *isomorphism*.

Given a connected cube subgraph H , let $p_H = p_H(d)$ be the probability that Q_p has a component ambient isomorphic to H . When p is fixed with $0 < p < \frac{1}{2}$, by Theorem 5,

$$\text{either } p_H \text{ or } 1 - p_H \text{ is } e^{-\Omega(d)}. \quad (5)$$

To see this, let $t = v(H)$, let Y be the number of components ambient isomorphic to H and $\lambda = \mathbb{E}[Y]$. If $t > m_p$ then $2q^t < 1$, so $\mathbb{P}(Y \geq 1) \leq \lambda = e^{-\Omega(d)}$; and if $t \leq m_p$ then $\lambda \rightarrow \infty$ (as we saw above), and by part (b) of Theorem 5 wvhp $Y \geq \lambda/2 > 0$. The statement (5) contrasts with the

situation at $p = \frac{1}{2}$, when (as we noted earlier) the number of isolated vertices has asymptotic distribution $\text{Po}(1)$.

Notation

We use standard notation throughout. For non-negative functions f and g , we say that $f(d) = \Omega(g(d))$ if $\liminf_{d \rightarrow \infty} f(d)/g(d) > 0$, and $f(d) = \Theta(g(d))$ if both $f(d) = \Omega(g(d))$ and $g(d) = \Omega(f(d))$. Also, we write $f \ll g$ if $f(d) = o(g(d))$.

Plan of the paper

Section 2 gives preliminary results, first concerning subgraphs in the hypercube Q^d , and then concerning the variance of counting random variables and their closeness to a Poisson distribution. In Section 3, Lemma 11 gives several results concerning numbers of components ambient-isomorphic to a given list of connected cube subgraphs H_i . Lemma 12 gives quite precise results on the expected value of X_t for $1 \leq t \leq m_p$. These lemmas allow us to prove Theorem 5, and then Theorem 4, at the end of the section.

In order to prove Theorems 1, 2 and 3 we must show that with tiny failure probability there is just one component of size strictly greater than m_p . To do this, in Section 4 we call a vertex ‘good’ if its degree in Q_p is at least half the expected value dp : we show that, with tiny failure probability, all good vertices are in the same component; and then deduce that, for a suitable constant N , with tiny failure probability each component of the fragment has size at most N . From this result, we see in particular that wvhp m_p is an upper bound for the size L_2 of a second largest component. In Section 5 we complete the proofs of Theorems 1, 2 and 3.

In Section 6 we consider joint distributions of random variables like the X_t in Theorem 4 or the Y_i in Theorem 5, and we present our sixth and last main theorem, Theorem 17. Finally, Section 7 contains some very brief concluding remarks.

These investigations arose from work on multicommodity flows in the cube Q^d when edges have independent random capacities, see [14].

2 Preliminary results

2.1 Preliminary results on the hypercube Q^d

Let us first consider $\text{span}(H)$ for a connected cube subgraph H . We have already noted that $\text{span}(H) = v(H) - 1$ if $v(H)$ is 1, 2 or 3. It is easy to see that always $\text{span}(H) \leq v(H) - 1$, and the inequality is strict if H is not a

tree (since any cycle contains at least two edges in some dimension). If we have equality we call H a *spreading* tree. (We use the term ‘spreading’ since each edge sits in a distinct dimension.)

What are the subcubes in Q^d ? If we are given $S \subseteq [d]$ and $z \in \{0, 1\}^{[d] \setminus S}$, then clearly the vertices x such that $x_j = z_j$ for each $j \in [d] \setminus S$ form a subcube isomorphic to $Q^{|S|}$. We shall need to consider such ‘cylinder’ subcubes, for example in the proof of Lemma 7. As an aside, let us note that each cube subgraph H isomorphic to a hypercube Q^s is obtained in this way. Since $v(H) = 2^s$, this is easily seen to be equivalent to showing that H has span s ; and it is a straightforward exercise to show the latter.

Proposition 6. *Let H be a subgraph of Q^d isomorphic to a hypercube Q^s . Then $\text{span}(H) = s$. \square*

Next we investigate the number $n_H = n_H(d)$ of subgraphs of Q^d ambient-isomorphic to a given subgraph H , the number of subgraphs which are spreading trees of a given size t , and the total number of connected subgraphs of size t .

Lemma 7. (a) *For each connected subgraph H of Q^d , $n_H = 2^{d-s} \binom{d}{s}$, where $\text{span}(H) = s$.*

(b) *For each $d \geq t - 1 \geq 0$, the number of ambient-isomorphism classes of spreading trees of size t in Q^d is $2^{t-1} t^{t-3}$.*

(c) *For each $d \geq t - 1 \geq 0$, the number of subgraphs of Q^d which are spreading trees of size t is $2^d t^{t-3} \binom{d}{t-1}$.*

(d) *For each fixed $t \geq 1$, the number of connected subgraphs of Q^d of size t is $2^d t^{t-3} \binom{d}{t-1} (1 + O(d^{-1}))$.*

We see from parts (c) and (d) above that the population of connected subgraphs of a given size t in Q^d is asymptotically dominated by spreading trees.

Proof. We first recall that any cube subgraph of size t can be embedded in Q^{t-1} and so, for $d \geq t - 1$, the number of pairwise non-ambient-isomorphic connected cube subgraphs of size t depends only on t .

(a) There is a single ambient-isomorphic copy of H in each (cylinder) subcube Q^s of Q^d ; and so there are $2^{d-s} \binom{d}{s}$ copies of Q^s in Q^d .

(b) By Cayley’s formula there are t^{t-2} trees on the set $\{0, 1, 2, \dots, t-1\}$ of t vertices. Given one of these trees, call vertex 0 the root and move the other

vertex labels onto the edge leading towards the root. This constructs a vertex-rooted, edge-labeled tree, with edge-labels $1, 2, \dots, t-1$. The construction is reversible, so there are exactly t^{t-2} such trees.

Given such a vertex-rooted, edge-labeled tree T , we choose a vertex in Q^{t-1} for the root, then use the labels of the edges to specify the ‘dimension’ in which that edge exists. This defines a t -vertex rooted spreading tree, and all the rooted trees constructed are distinct; and furthermore every t -vertex rooted tree in Q^{t-1} can be constructed in this way. Thus there are $2^{t-1}t^{t-2}$ t -vertex rooted spreading trees in Q^{t-1} , and so $2^{t-1}t^{t-3}$ t -vertex unrooted spreading trees; and of these unrooted trees, no two distinct ones are ambient-isomorphic since they have span $t-1$ and so are their own canonical copies.

(c) By parts (a) and (b), the number of t -vertex spreading trees in Q^d is

$$2^{t-1}t^{t-3} \cdot 2^{d-(t-1)} \binom{d}{t-1} = 2^d t^{t-3} \binom{d}{t-1}.$$

(d) If T is a spreading tree of size t , and H is a connected cube subgraph of size t with $\text{span}(H) < t-1 = \text{span}(T)$, then $n_H/n_T = O(d^{-1})$ by part (1). The number of ambient-isomorphism classes of connected subgraphs of Q^d of size t does not depend on d for $d \geq t-1$; and thus the contribution to the total number of connected subgraphs of Q^d of size t by those with span less than $t-1$ is $O(d^{-1})$ of the total. \square

We will need one more lemma which we will apply to the hypercube Q^d . It may be ‘folk knowledge’, but we give a proof here for completeness.

Lemma 8. *Let the graph G be rooted at vertex r and have maximum degree at most d . Then for each positive integer t , the number of subtrees containing r and exactly t other vertices is at most $(ed)^t$.*

Proof. We first show (a) that the number of $(t+1)$ -vertex subtrees in G containing r is at most the number of $(t+1)$ -vertex subtrees containing the root in an infinite d -ary tree T^∞ ; and then show (b) that the latter number is at most the number of points $x \in \{0, 1\}^{td}$ with t 1’s. Clearly the number of such points is $\binom{td}{t} \leq (ed)^t$.

The *path tree* $T(G, r)$ [12] has a node for each path P from r , adjacent to each node corresponding to a path extending P by one edge; and has root the node for the path with the single vertex r . It is easy to see that, for each tree in G containing r , there is a corresponding tree in $T(G, r)$ containing the root. Thus the the number of $(t+1)$ -vertex subtrees in G containing r is

at most the number of $(t+1)$ -vertex subtrees containing the root in $T(G, r)$; and since $T(G, r)$ embeds in T^∞ , part (a) of the proof follows.

For part (b), let T be a $(t+1)$ -vertex subtree in T^∞ containing the root. We may suppose that T^∞ is embedded in the plane, with the root at the top and children listed in order from left to right. We construct $x^T \in \{0, 1\}^{td}$ with t 1's as follows. Initially the vector x is null and the list L contains just the root. We repeat the following t times. Remove the first vertex v in L , and let $y \in \{0, 1\}^d$ indicate its children (with a 1 for each child): append y to x and append the children to L (listed in order). The output x^T is the final value of x . Clearly we can reconstruct T from x^T , so the number of possible trees T is at most the number of possible vectors x^T , which completes the proof. \square

2.2 Preliminary results on variance and approximation to Poisson distribution

Let $(A_i : i \in I)$ be a family of events with a dependency graph L (so that A_i and A_j are independent if i and j are not adjacent in L and $i \neq j$). Write $i \sim j$ if i and j are adjacent in L . For each i , let $\pi_i = \mathbb{P}(A_i)$ and let \mathbb{I}_i be the indicator of A_i ; and let $X = \sum_i \mathbb{I}_i$. (In this subsection we do not use X as the number of components in Q_p .) Then

$$\begin{aligned} \text{Var}(X) &= \sum_i \sum_j (\mathbb{P}(A_i \wedge A_j) - \pi_i \pi_j) \\ &= \sum_i (\pi_i - \pi_i^2) + \sum_i \sum_{j \sim i} (\mathbb{P}(A_i \wedge A_j) - \pi_i \pi_j) \\ &= \mathbb{E}[X] + \Delta^+ - \Delta^-, \end{aligned} \tag{6}$$

where

$$\Delta^+ = \sum_i \sum_{j \sim i} \mathbb{P}(A_i \wedge A_j) \tag{7}$$

and

$$\Delta^- = \sum_i \pi_i^2 + \sum_i \sum_{j \sim i} \pi_i \pi_j. \tag{8}$$

The following lemma is essentially Theorem 6.23 of [13], proved by the Stein-Chen method, which shows that a sum X as above has close to a Poisson distribution.

Lemma 9. *With notation as above, and letting $\lambda = \mathbb{E}[X]$, we have*

$$d_{TV}(X, \text{Po}(\lambda)) \leq \min\{\lambda^{-1}, 1\} (\Delta^+ + \Delta^-).$$

We shall also need a minor extension of the above. Suppose that we are given also a family $(t_i : i \in I)$ of positive integers, and let $\tilde{X} = \sum_i t_i \mathbb{I}_i$. Then much as above, we have

$$\begin{aligned} \text{Var}(\tilde{X}) &= \sum_i \sum_j t_i t_j (\mathbb{P}(A_i \wedge A_j) - \pi_i \pi_j) \\ &= \sum_i t_i^2 (\pi_i - \pi_i^2) + \sum_i \sum_{j \sim i} t_i t_j (\mathbb{P}(A_i \wedge A_j) - \pi_i \pi_j) \\ &= \mathbb{E}[\tilde{X}] + \tilde{\Delta}^+ - \tilde{\Delta}^- \end{aligned} \tag{9}$$

where

$$\tilde{\Delta}^+ = \sum_i t_i (t_i - 1) \pi_i + \sum_i \sum_{j \sim i} t_i t_j \mathbb{P}(A_i \wedge A_j) \tag{10}$$

and

$$\tilde{\Delta}^- = \sum_i t_i^2 \pi_i^2 + \sum_i \sum_{j \sim i} t_i t_j \pi_i \pi_j. \tag{11}$$

Lemma 10. *With notation as above, and now letting $\lambda = \mathbb{E}[\tilde{X}]$, we have*

$$d_{TV}(\tilde{X}, \text{Po}(\lambda)) \leq \min\{\lambda^{-1}, 1\} (\tilde{\Delta}^+ + \tilde{\Delta}^-).$$

Proof. Replace each event A_i by t_i identical (not independent) copies. Note that, for each i , the t_i copies of A_i are dependent, and so they are adjacent to each other in the natural extended dependency graph. Now apply Lemma 9. \square

3 The numbers of small components

The first lemma in this section, Lemma 11, gives expected values and variances for the numbers of small components in certain ambient-isomorphism classes, and for the number of vertices in such components; and gives some results on approximation by a Poisson distribution. The second lemma uses Lemma 11, together with counting results from Subsection 2.1, to deduce results corresponding to those in Lemma 11 when we consider *all* components of a given size. Using these lemmas we prove Theorem 5 and then Theorem 4.

In Lemma 11, we consider both the numbers of components in Q_p ambient isomorphic to given graphs, and the total numbers of vertices in such components.

Lemma 11. *Let $0 < p < \frac{1}{2}$ and let $q = 1 - p$. Let r be a positive integer and let H_1, H_2, \dots, H_r be pairwise non-ambient-isomorphic connected cube subgraphs. For each $i \in [r]$, let $s_i = \text{span}(H_i)$, and let $e'(H_i)$ be the number of cube edges not in H_i but with both end vertices in H_i .*

For each $i \in [r]$, let Y_i be the number of components of Q_p ambient-isomorphic to H_i . Then the following hold.

(a) *For each $i \in [r]$, once $d \geq s_i$ we have*

$$\mathbb{E}[Y_i] = (p/q^2)^{e(H_i)} q^{-e'(H_i)} 2^{d-s_i} \binom{d}{s_i} q^{v(H_i)d}.$$

Let $t = \min_i v(H_i)$, and let $s = \max\{s_i : v(H_i) = t\}$. Let $I^ = \{i \in [r] : v(H_i) = t, s_i = s\}$, and let*

$$c = \frac{1}{2^s s!} \sum_{i \in I^*} (p/q^2)^{e(H_i)} q^{-e'(H_i)}.$$

(b) (i) *The sum $Y = \sum_{i=1}^r Y_i$ satisfies $\mathbb{E}[Y] = (1 + O(1/d)) c(d)_s (2q^t)^d$; and if t is 1, 2 or 3 then $s = t - 1$ and we can replace $O(1/d)$ by $O(dq^d)$. Also*
(ii) *$\text{Var}(Y) = (1 + O(d^t q^{td})) \mathbb{E}[Y]$, and (iii) $d_{TV}(Y, \text{Po}(\mathbb{E}[Y])) = O(d^t q^{td})$.*

(c) (i) *The weighted sum $\tilde{Y} = \sum_{i=1}^r v(H_i) Y_i$ satisfies $\mathbb{E}[\tilde{Y}] = (1 + O(dq^d)) t \mathbb{E}[Y]$; and if $t = 1$ then (ii) $\text{Var}(\tilde{Y}) = (1 + O(dq^d)) \mathbb{E}[\tilde{Y}]$ and (iii) $d_{TV}(\tilde{Y}, \text{Po}(\mathbb{E}[\tilde{Y}])) = O(dq^d)$.*

Proof. (a) Consider a fixed graph H_j . Let G be a subgraph of Q^d which is ambient-isomorphic to H_j , and let A be the event that the subgraph of Q_p induced by the vertices of G is exactly G , and it is also a component of Q_p . Then

$$\mathbb{P}(A) = p^{e(H_j)} q^{e'(H_j)} q^{v(H_j)d - 2e(H_j) - 2e'(H_j)} = (p/q^2)^{e(H_j)} q^{-e'(H_j)} q^{v(H_j)d}. \quad (12)$$

Hence, by Lemma 7 part (a)

$$\mathbb{E}[Y_j] = 2^{d-s_j} \binom{d}{s_j} (p/q^2)^{e(H_j)} q^{-e'(H_j)} q^{v(H_j)d},$$

completing the proof of part (a).

(b) Observe from part (a) that $\mathbb{E}[Y_j] = \Theta(d^{s_j} (2q^{v(H_j)})^d)$. Thus the dominant contribution to $\mathbb{E}[Y]$ is from graphs H_j with $j \in I^*$; for if $j \in I^*$ and $i \in I \setminus I^*$, then $\mathbb{E}[Y_i] = O(1/d) \mathbb{E}[Y_j]$. Using part (a) we now see that

$$\mathbb{E}[Y] = (1 + O(1/d)) c(d)_s (2q^t)^d = (1 + O(1/d)) cd^s (2q^t)^d.$$

Now suppose that t is 1, 2 or 3. If $j \in I^*$ and $i \in I \setminus I^*$, then $v(H_i) > t$ so $\mathbb{E}[Y_i] = O(dq^d) \mathbb{E}[Y_j]$ (note that if $v(H_i) = t+1$ then $s_i \leq s+1$). Hence $\mathbb{E}[Y] = (1 + O(dq^d)) c(d)_s (2q^t)^d$. (If $t \geq 4$ then there could be t -vertex graphs H_i with different spans, and if one has span $s-1$ then $\mathbb{E}[Y] = (1 + \Theta(1/d)) c(d)_s (2q^t)^d$.)

Now we prove parts (b)(ii) and (b)(iii). Given d , let $\mathcal{S} = \mathcal{S}(d)$ be the set of subgraphs of Q^d ambient isomorphic to one of the graphs H_1, \dots, H_r . List the members of \mathcal{S} as G_1, \dots, G_N (where $N = N(d)$); and let A_i be the event that G_i is a component of Q_p . For distinct $i, j \in [N]$ let $i \sim j$ if either the vertex sets $V(G_i)$ and $V(G_j)$ meet or there is an edge of Q^d between them. Observe that if $i \neq j$ and $i \not\sim j$ then the events A_i and A_j are independent, so we have a dependency graph. Now by (6) $\text{Var}(Y) = \mathbb{E}[Y] + \Delta^+ - \Delta^-$, where Δ^+ and Δ^- are defined in (7) and (8) respectively. We next bound Δ^+ then Δ^- .

If $i \neq j$ and the vertex sets $V(G_i)$ and $V(G_j)$ meet, then $\mathbb{P}(A_i \wedge A_j) = 0$, so in the sum for Δ^+ in (7) we need consider only the case where the two vertex sets $V(G_i)$ and $V(G_j)$ are disjoint but have connecting edges in Q^d (of which there can be at most $v(G_i)v(G_j)$). By (12), there is a constant α such that

$$\mathbb{P}(A_k) \leq \alpha q^{v(G_k)d} \quad \text{for each } k. \quad (13)$$

Thus, if $i \neq j$ then

$$\mathbb{P}(A_i \wedge A_j) \leq \mathbb{P}(A_i) \mathbb{P}(A_j) q^{-v(G_i)v(G_j)} \leq \mathbb{P}(A_i) \alpha q^{v(G_j)d} q^{-v(G_i)v(G_j)}. \quad (14)$$

For a given graph G_i of size t_1 , the number of vertices v in Q^d adjacent to vertices in G_i is at most $t_1 d$, and by Lemma 8 each of these vertices v could be in at most $(ed)^{t_2-1}$ possible components of size t_2 . In the sums below, t_1 and t_2 run over the possible sizes of the graphs G_k . From the definition (7), and using (14), we have

$$\begin{aligned} \Delta^+ &= \sum_{t_1} \sum_{t_2} \sum_{i:v(G_i)=t_1} \sum_{j:j \sim i, v(G_j)=t_2} \mathbb{P}(A_i \wedge A_j) \\ &\leq \sum_{t_1} \sum_{t_2} \sum_{i:v(G_i)=t_1} \mathbb{P}(A_i) (t_1 d) (ed)^{t_2-1} \alpha q^{t_2 d} q^{-t_1 t_2} \\ &\leq (1 + o(1)) \sum_{t_1} \sum_{i:v(G_i)=t_1} \mathbb{P}(A_i) (t_1 d) (ed)^{t_1-1} \alpha q^{td} q^{-t_1 t} \\ &= \mathbb{E}[Y] O(d^t q^{td}), \end{aligned}$$

that is

$$\Delta^+ = \mathbb{E}[Y] O(d^t q^{td}). \quad (15)$$

Now consider Δ^- . By (13)

$$\sum_i \mathbb{P}(A_i)^2 \leq \sum_i \mathbb{P}(A_i) \cdot \alpha q^{td} = \mathbb{E}[Y] \cdot \alpha q^{td},$$

and, as for Δ^+ except without the factor $q^{-t_1 t_2}$ (also including pairs i, j with $V(G_i) \cap V(G_j) \neq \emptyset$), we have

$$\sum_i \sum_{j \sim i} \mathbb{P}(A_i) \mathbb{P}(A_j) = \mathbb{E}[Y] O(d^t q^{td});$$

thus

$$\Delta^- = \mathbb{E}[Y] O(d^t q^{td}). \quad (16)$$

Now that we have (15), (16), from (6), we have $\text{Var}(Y) = \mathbb{E}[Y](1 + O(q^{td} d^t))$, and by Lemma 9 we have $d_{TV}(Y, \text{Po}(\mathbb{E}[Y])) = O(d^t q^{td})$, as required.

(c) The contribution to $\mathbb{E}[Y]$ from graphs H_j with $v(H_j) > t$ is $O(dq^d) \cdot \mathbb{E}[Y]$, and similarly for $\mathbb{E}[\tilde{Y}]$. This gives equation (c)(i).

For parts (c) (ii) and (iii), we may argue as for parts (b) (ii) and (iii), but using Lemma 10 instead of Lemma 9. Since the t_i are sizes of the graphs H_j , they are uniformly bounded, so $\tilde{\Delta}^-$ (as in (11)) is at most a constant times the unweighted version Δ^- , and similarly for the second term in $\tilde{\Delta}^+$ (as in (10)). For the first term in $\tilde{\Delta}^+$, there is no contribution from the isolated vertices (graphs H_i with $v(H_i) = 1$), so this term is $O(d(2q^2)^d)$: but $\mathbb{E}[Y] \geq \mu_1 = (2q)^d$, so the term is $O(\mathbb{E}[Y] dq^d)$. Hence by (15) and (16), these terms together are $O(\mathbb{E}[Y] dq^d)$. Equation (9) and Lemma 10 now complete the proof. \square

Recall that X_t denotes the number of components of size t in Q_p , and that $\mu_t = \mathbb{E}[X_t]$. We noted earlier (more than once) that $\mu_1 = (2q)^d$, and the precise values of μ_2 and μ_3 are given in equation (2).

Lemma 12. *Let $0 < p < \frac{1}{2}$ and let $q = 1 - p$. Let $1 \leq t \leq m_p$. Then*

$$\mu_t = (1 + O(\frac{1}{d})) \frac{t^{t-2}}{t!} (\frac{p}{q^2})^{t-1} d^{t-1} (2q^t)^d = \Theta(d^{t-1} (2q^t)^d).$$

Proof. If H_j is a spreading tree of size t , then $s_j = t - 1$ and $e'(H_j) = 0$, and so by Lemma 11 (a),

$$\mathbb{E}[Y_j] = (p/q^2)^{t-1} 2^{d-t+1} \binom{d}{t-1} q^{td}. \quad (17)$$

To calculate μ_t we need to sum $\mathbb{E}[Y_j]$ over all the ambient-isomorphism classes of t -vertex connected cube subgraphs H_j . We see from Lemma 11 (a) (and

equation (17)) that if H_j is a spreading tree and $H_{j'}$ is not (so $s_{j'} \leq t - 2$) then $\mathbb{E}[Y_{j'}] = O(d^{-1}) \mathbb{E}[Y_j]$. Thus the only significant terms are those corresponding to ambient-isomorphism classes of spreading trees, and by Lemma 7 (b) there are $2^{t-1}t^{t-3}$ such classes. Hence

$$\begin{aligned}\mu_t &= (1 + O(\frac{1}{d})) 2^{t-1}t^{t-3} 2^{d-t+1} \binom{d}{t-1} q^{td} (p/q^2)^{t-1} \\ &= (1 + O(\frac{1}{d})) d^{t-1} (2q^t)^d (t^{t-3}/(t-1)!) (p/q^2)^{t-1},\end{aligned}$$

as required. \square

Let us now complete the proof of Theorem 5 and then of Theorem 4.

Proof of Theorem 5. In part (a), the expected value is from Lemma 11 part (b)(i), and the variance is from Lemma 11 part (b)(ii); and the first half of part (c) (on Poisson approximation) is from Lemma 11 part (b)(iii).

Consider part (b). By a Chernoff inequality (see for example inequality (2.9) and Remark 2.6 of [13]),

$$\begin{aligned}\mathbb{P}(|Y - \lambda| \geq \varepsilon(d\lambda)^{\frac{1}{2}}) &\leq \mathbb{P}(|\text{Po}(\lambda) - \lambda| \geq \varepsilon(d\lambda)^{\frac{1}{2}}) + d_{TV}(Y, \text{Po}(\lambda)) \\ &\leq 2e^{-\varepsilon^2 d/3} + O(d^t q^{td}),\end{aligned}$$

by the Poisson approximation bound. Thus $\mathbb{P}(|Y - \lambda| \geq \varepsilon(d\lambda)^{\frac{1}{2}}) = e^{-\Omega(d)}$, as required.

Finally, consider the second half of part (c). Since $d_{TV}(Y, \text{Po}(\lambda)) \rightarrow 0$ and $\lambda \rightarrow \infty$ as $d \rightarrow \infty$, it follows that $(Y - \lambda)/\sqrt{\lambda}$ is asymptotically standard normal. But $\text{Var}(Y) \sim \lambda$, so also Y^* is asymptotically standard normal. This concludes the proof of Theorem 5. \square

Proof of Theorem 4. The expression for the mean μ_t in part (a) is from Lemma 12 part (a). The rest follows directly from Theorem 5, with H_1, \dots, H_r listing a representative of each ambient-isomorphism class of t -vertex connected cube subgraphs. \square

We have now proved Theorem 4, which says in particular that the number X_t of components in Q_p of size t has close to the Poisson distribution $\text{Po}(\mu_t)$. From what we have already proved, we can quickly give a first corresponding local limit result, showing that for suitable t we have $\mathbb{P}(X_t = \nu) \sim \mathbb{P}(\text{Po}(\mu_t) = \nu)$ uniformly over the ‘central range’ of integers ν . Recall from Theorem 4 that $\mu_t = \Theta(d^{t-1}(2q^t)^d)$.

Proposition 13. *Let $0 < p < 1/2$, and let t be an integer with $m_p/3 < t \leq m_p$. Then for any fixed $c > 0$*

$$\sup_{\nu} |\mathbb{P}(X_t = \nu)/\mathbb{P}(\text{Po}(\mu_t) = \nu) - 1| = e^{-\Omega(d)}$$

where the sup is over integers ν with $|\nu - \mu_t| \leq c\sqrt{\mu_t}$.

Proof. Note first that $\mathbb{P}(\text{Po}(\mu_t) = \nu) = \Theta(\mu_t^{-\frac{1}{2}})$, uniformly over integers ν with $|\nu - \mu_t| \leq c\sqrt{\mu_t}$. By Theorem 4 part (c), $d_{TV}(X_t, \text{Po}(\mu_t)) = O(d^t q^{td})$, so $|\mathbb{P}(X_t = \nu) - \mathbb{P}(\text{Po}(\mu_t) = \nu)| = O(d^t q^{td})$ uniformly over integers ν ; and hence

$$|\mathbb{P}(X_t = \nu)/\mathbb{P}(\text{Po}(\mu_t) = \nu) - 1| = O(d^t q^{td} \mu_t^{-\frac{1}{2}}),$$

uniformly over integers ν with $|\nu - \mu_t| \leq c\sqrt{\mu_t}$. But $d^t q^{td} \mu_t^{1/2} = O(d^{3t/2} (2q^{3t})^{d/2}) = o(1)$ provided $2q^{3t} < 1$. Finally, we have $2q^{3t} < 1$ if $t > m_p/3$ (and indeed if $t = m_p/3$ unless $(2q)^{m_p} = 1$). \square

4 The fragment \mathcal{Z} has no large components

It will be straightforward to handle components of any fixed size $t > m_p$: we need to show that wvhp there are no larger components in \mathcal{Z} , see Lemma 16 below. We use two preliminary lemmas. Given a spanning subgraph Q' of Q , call a vertex Q' -good if its degree in Q' is at least $dp/2$ and bad otherwise.

Lemma 14. *The probability that there is a pair of Q_p -good vertices at distance at most 3 in Q which are not joined by a path (of length at most 7) in Q_p is $2^{-\Omega(d^2)}$.*

Proof. Let $\Gamma(w)$ denote the set of neighbours in Q_p of a vertex w . Fix vertices $u \neq v$ in Q at distance at most 3. Consider the case when $d_Q(u, v) = 3$ - the other cases are similar. We may suppose wlog that $u = \emptyset$ and $v = \{1, 2, 3\}$. Let A and B be sets of at least $dp/2$ neighbours in Q of u and v respectively.

For each $i \neq j$ in $\{4, \dots, d\}$ with $\{i\} \in A$ and $v \cup \{j\} \in B$, there is a path

$$\{i\}, \{i, j\}, \{i, j, 1\}, \{i, j, 1, 2\}, \{i, j, 1, 2, 3\}, \{j, 1, 2, 3\}$$

in Q , not using any edges incident with u or v . These form at least $(|A| - 3)(|B| - 4) \geq (pd/2 - 3)(pd/2 - 4)$ paths in Q of length 5 between A and B ; and the paths are pairwise edge-disjoint since each edge identifies the pair (i, j) . But the number of paths is at least $p^2 d^2 / 5$ for d sufficiently large, and then

$$\begin{aligned} & \mathbb{P}(\text{no } A-B \text{ path of length 5 in } Q_p \mid \Gamma(u) = A, \Gamma(v) = B) \\ & \leq (1 - p^5)^{p^2 d^2 / 5} \leq e^{-p^7 d^2 / 5}. \end{aligned}$$

But $\mathbb{P}(\text{no } u-v \text{ path of length 5 in } Q_p \mid u, v \text{ are } Q^p\text{-good})$ is a weighted average of such probabilities, so

$$\begin{aligned} & \mathbb{P}(\text{no } u-v \text{ path of length 5 in } Q_p) \wedge (u, v \text{ are } Q^p\text{-good}) \\ & \leq \mathbb{P}(\text{no } u-v \text{ path of length 5 in } Q_p \mid u, v \text{ are } Q^p\text{-good}) \leq e^{-p^7 d^2/5}. \end{aligned}$$

Now, by a union bound, the probability that there is a pair of Q^p -good vertices at distance 3 in Q which are not joined by a path of length 7 in Q_p is at most

$$2^d d^3 e^{-p^7 d^2/5} = 2^{-\Omega(d^2)}.$$

Similarly, if $d_Q(u, v) = 2$ then there is a $u-v$ path of length 6, and if $d_Q(u, v) = 1$ then there is a $u-v$ path of length 1 or 5. \square

The second preliminary lemma is deterministic.

Lemma 15. *Let Q' be a (fixed) spanning subgraph of Q . Suppose that (a) each vertex has a Q' -good neighbour in Q , and (b) for each pair u, v of Q' -good vertices at distance at most 3 in Q there is a $u-v$ path in Q' . Then for each pair u, v of Q' -good vertices there is a $u-v$ path in Q' ; and so all Q' -good vertices are in the same component of Q' .*

Proof. Let u, v be Q' -good vertices at distance $t > 3$ in Q . We must show that there is a $u-v$ path in Q' . Let $u = x_0, x_1, \dots, x_{t-1}, x_t = v$ be a $u-v$ path in Q of length t . For each $i = 1, \dots, t-1$, let y_i be a Q' -good neighbour in Q of x_i , where we choose $y_1 = u$ and $y_{t-1} = v$. Then since $d_Q(y_i, y_{i+1}) \leq 3$ for each $i = 1, \dots, t-2$ there is a $y_i - y_{i+1}$ path in Q' . Hence there is a $u-v$ path in Q' . \square

By a Chernoff bound and a union bound, we have

$$\begin{aligned} \mathbb{P}(\text{some vertex has no } Q_p\text{-good neighbour in } Q) & \leq 2^d \mathbb{P}(\text{Bin}(d, p) < pd/2)^d \\ & \leq 2^d e^{-(pd/8)d} = 2^{-\Omega(d^2)}. \end{aligned}$$

Let A be the event that all Q_p -good vertices in Q_p are in the same component. From the above bound and the last two lemmas

$$\mathbb{P}(\bar{A}) = 2^{-\Omega(d^2)}. \tag{18}$$

We may now deduce an upper bound for L_2 as required. We shall later typically set the parameter γ as 3, so that failure probabilities will be negligibly small.

Lemma 16. *Let $0 < p < 1/2$ and let $\gamma > 0$. Then there is a constant N such that $\mathbb{P}(L_2 > N) = o(2^{-\gamma d})$.*

Proof. Let $N = \lfloor \frac{16(1+\gamma)}{p} \rfloor$. If some component of the fragment has size at least $N + 1$, then also the giant component has size at least $N + 1$: hence, if the event A holds then there is a component with size at least $N + 1$ consisting entirely of bad vertices, and so there is a subtree T . Since Q is bipartite there is a set W of at least $(N + 1)/2$ vertices of T which forms an independent (stable) set in Q ; and the probability that each vertex in such a set W is bad is

$$\mathbb{P}(\text{Bin}(d, p) < pd/2)^{|W|} \leq e^{-\frac{pd}{8} \frac{N+1}{2}} = e^{-(N+1)pd/16}$$

by a Chernoff bound. Hence by Lemma 8 and a union bound, the probability that there is a subtree of Q_p with $N + 1$ vertices each of which is bad is at most

$$2^d (ed)^N e^{-(N+1)pd/16} = o(2^{-\gamma d})$$

since $(N + 1)pd/16 \geq (1 + \gamma)d$. Finally, using also (18), we have

$$\mathbb{P}(L_2 > N) \leq \mathbb{P}((L_2 > N) \wedge A) + \mathbb{P}(\bar{A}) = o(2^{-\gamma d}),$$

which completes the proof. \square

5 Completing proofs of Theorem 1, 2 and 3

5.1 Proof of Theorem 1

It is convenient to prove part (b) first. Let N be as in Lemma 16 for $\gamma = 3$, so that $\mathbb{P}(L_2 > N) = o(2^{-3d})$. Consider an integer t with $m_p < t \leq N$. By Markov's inequality and Lemma 12 part (a),

$$\mathbb{P}(X_t \geq 1) \leq \mathbb{E}[X_t] = O((2q^t)^d d^t) = e^{-\Omega(d)},$$

where the last step follows since $2q^t < 1$. Hence wvhp the fragment \mathcal{Z} has no component containing exactly t vertices. Putting these results together, we see $L_2 \leq m_p$ wvhp; and that

$$\mathbb{E}[L_2] \leq m_p + N \mathbb{P}(m_p < L_2 \leq N) + 2^d \mathbb{P}(L_2 > N) = m_p + e^{-\Omega(d)}.$$

But $L_2 \geq m_p$ wvhp by Theorem 4 part (b) with $t = m_p$; and so $L_2 = m_p$ wvhp. It follows that $\mathbb{E}[L_2] \geq m_p - e^{-\Omega(d)}$, and hence $|\mathbb{E}[L_2] - m_p| = e^{-\Omega(d)}$.

To bound $\text{Var}(L_2)$ we have

$$\mathbb{E}[(L_2 - m_p)^2 I_{L_2 \leq N}] \leq N^2 \mathbb{P}(L_2 \neq m_p) = e^{-\Omega(d)},$$

and

$$\mathbb{E}[(L_2 - m_p)^2 I_{L_2 > N}] \leq 2^{2d} \mathbb{P}(L_2 > N) = e^{-\Omega(d)}.$$

Hence

$$\text{Var}(L_2) \leq \mathbb{E}[(L_2 - m_p)^2] = e^{-\Omega(d)},$$

which completes the proof of part (b).

Now let us prove part (a). We have seen that $L_2 \leq m_p$ wvhp. Let $X^- = \sum_{t \leq m_p} X_t$ and $Z^- = \sum_{t \leq m_p} tX_t$. Then $X^- = X$ wvhp and $Z^- = Z$ wvhp. Let $\varepsilon > 0$. By Theorem 5 part (b), we have $|X^- - \mu_1| \leq \varepsilon \mu_1$ wvhp, and $|X_1 - \mu_1| \leq (\varepsilon/2)\mu_1$ wvhp. Also, $\sum_{t=2}^{m_p} tX_t \leq (\varepsilon/2)\mu_1$ wvhp by a first moment argument, since μ_t is exponentially smaller than μ_1 for each $t = 2, \dots, m_p$; so $|Z^- - \mu_1| \leq \varepsilon \mu_1$ wvhp. Thus $|Y - \mu_1| \leq \varepsilon \mu_1$ wvhp, as required.

5.2 Proof of Theorem 2

We prove the two parts of the theorem separately. We denote the r -ball $B_r(\mathbf{0})$ centred on $\mathbf{0}$ by B_r for short.

Proof of Theorem 2 part (a). Let $s = m_p + 1$ and let $V = V(Q)$. Recall that $L_2 \leq m_p$ wvhp. We use $\deg(v)$ for the degree of a vertex v in Q_p . Also, for $v \in V$ and $W \subseteq V$, let $e(v, W)$ be the number of edges in Q_p between v and W . For each subset $S \subseteq V$ with $|S| = s$ we have

$$\begin{aligned} \mathbb{P}((S \subseteq V(\mathcal{Z})) \wedge (L_2 \leq m_p)) &\leq \mathbb{P}(\deg(v) \leq m_p - 1 \quad \forall v \in S) \\ &\leq \mathbb{P}(e(v, V \setminus S) \leq m_p - 1 \quad \forall v \in S) \\ &= (\mathbb{P}(\text{Bin}(d - s, p) \leq m_p - 1))^s \\ &\leq \left(\binom{d - s}{m_p - 1} q^{d - s - (m_p - 1)} \right)^s \\ &\leq (d^{m_p - 1} q^{d - 2m_p})^s \leq (d/q^2)^{m_p s} q^{sd}. \end{aligned}$$

Hence, for any $r > 0$,

$$\begin{aligned} &\mathbb{P}(|V(\mathcal{Z}) \cap B_r(u)| \geq s \text{ for some } u \in V) \\ &= \mathbb{P}(\cup_{u \in V} \cup_{S \subseteq B_r(u), |S|=s} (S \subseteq V(\mathcal{Z}))) \\ &\leq \mathbb{P}(\cup_{u \in V} \cup_{S \subseteq B_r(u), |S|=s} (S \subseteq V(\mathcal{Z})) \wedge (L_2 \leq m_p)) + \mathbb{P}(L_2 > m_p) \\ &\leq 2^d \binom{|B_r|}{s} (d/q^2)^{m_p s} q^{sd} + \mathbb{P}(L_2 > m_p) \\ &\leq (d/q^2)^{m_p s} |B_r|^s (2q^s)^d + \mathbb{P}(L_2 > m_p). \end{aligned} \tag{19}$$

Since $s > m_p$ we have $2q^s < 1$, and so $1 > \log_2 \frac{1}{q} - \frac{1}{s} > 0$. Let η_1 be the unique $x \in (0, \frac{1}{2})$ such that $h(x) = \log_2 \frac{1}{q} - \frac{1}{s}$. Let $0 < \eta < \eta_1$: then $h(\eta) < \log_2 \frac{1}{q} - \frac{1}{s}$, and so

$$2(2^{h(\eta)} q)^s < 1.$$

Set $r = \eta d$. Then $|B_r| = 2^{h(\eta)d+o(d)}$ by standard estimates. Thus, by the last inequality,

$$|B_r|^s (2q^s)^d = (2(2^{h(\eta)}q)^s)^d 2^{o(d)} = 2^{-\Omega(d)}.$$

Hence, by (19) and using $\mathbb{P}(L_2 > m_p) = 2^{-\Omega(d)}$, we have

$$\mathbb{P}(|V(\mathcal{Z}) \cap B_r(u)| \geq s \text{ for some } u \in V) = 2^{-\Omega(d)}$$

as required. \square

Consider η_1 in the above proof: if $\eta > \eta_1$ then the expected number of ηd -balls containing more than m_p vertices in \mathcal{Z} tends to ∞ as $d \rightarrow \infty$.

Proof of Theorem 2 part (b). We may assume that $\varepsilon > 0$ is sufficiently small that $\eta^* - \varepsilon > 0$ and $\eta^* + \varepsilon < \frac{1}{2}$. Given $0 < \eta \leq \frac{1}{2}$, we have $|B_{\eta d}| = 2^{h(\eta)d+o(d)}$, as we noted above. Also, $2^{-h(\eta^*)} = q$. Hence, by Theorem 1 (a), wvhp

$$\begin{aligned} |B_{(\eta^*-\varepsilon)d}| \cdot Z &\leq 2^{h(\eta^*-\varepsilon)d+o(d)} \cdot 2\mu_1 \\ &= 2^{(h(\eta^*-\varepsilon)-h(\eta^*)+o(1))d} \cdot 2^d \\ &= 2^{-\Omega(d)} \cdot 2^d, \end{aligned}$$

But the number of vertices within distance at most $(\eta^* - \varepsilon)d$ of \mathcal{Z} is at most $|B_{(\eta^*-\varepsilon)d}| \cdot Z$, so this proves the first half of part (b).

For the second half, let B' denote $B_{(\eta^*+\varepsilon)d}$. By the definition of η^* , and recalling that $h(\eta)$ is strictly increasing on $(0, \frac{1}{2})$, we have $q^d |B'| = e^{\Omega(d)}$. Since Q^d is bipartite, there is a stable subset B'' of B' with $|B''| \geq \frac{1}{2}|B'|$; and the probability that no vertex of \mathcal{Z} is in B' is at most the probability that no vertex in B'' is isolated, which equals

$$(1 - q^d)^{|B''|} \leq \exp(-\frac{1}{2}q^d |B'|) = \exp(-e^{\Omega(d)}).$$

This bound refers to the ball B' centred at $\mathbf{0}$, and indeed to any fixed centre vertex. Taking a union bound over all 2^d possible centre vertices shows that the probability that some vertex is not within distance $(\eta^* + \varepsilon)d$ of \mathcal{Z} is $\exp(-e^{\Omega(d)})$, and thus completes the proof. \square

In the last part of the proof above, the number of isolated vertices in B'' has distribution $\text{Bin}(|B''|, q^d)$, with mean at least $\frac{1}{2}|B'|q^d = e^{\Omega(d)}$. Hence, by a Chernoff bound, the probability that there are at most $\frac{1}{4}|B'|q^d$ isolated vertices in the ball B' is at most $e^{-e^{\Omega(d)}}$; and so, by a union bound, wvhp each $(\eta^* + \varepsilon)d$ -ball contains exponentially many isolated vertices.

5.3 Proof of Theorem 3

By Lemma 16 we may choose a fixed positive integer N such that $\mathbb{P}(L_2 > N) \leq 2^{-3d}$.

Proof of Theorem 3 part (a). Note that $Z \leq 2^d$ and so

$$\mathbb{E}[ZI_{L_2 > N}] \leq 2^d \mathbb{P}(L_2 > N) \leq 2^{-2d}.$$

For each $2 \leq t \leq N$, $\mathbb{E}[X_t]$ is $O(d(2q^2)^d)$. Hence,

$$\begin{aligned} \mathbb{E}[Z] &= \mathbb{E}[ZI_{L_2 \leq N}] + \mathbb{E}[ZI_{L_2 > N}] \\ &\leq \mu_1 + O(d(2q^2)^d) + 2^{-2d} = (1 + O(dq^d))\mu_1. \end{aligned}$$

Also, of course, $\mu_1 \leq \mathbb{E}[X] \leq \mathbb{E}[Z]$, which completes the proof for the expected values.

Now consider variances. Let $X_{\leq N} = \sum_{t=1}^N X_t$, the total number of components in Q_p of size at most N ; and similarly let $Z_{\leq N} = \sum_{t=1}^N tX_t$, the total size of the components of size at most N . Then

$$\text{Var}(Y) - \text{Var}(Y_{\leq N}) \leq \mathbb{E}[Y^2 - Y_{\leq N}^2] \leq 2^{2d} \mathbb{P}(L_2 > N) \leq 2^{-d},$$

and

$$\text{Var}(Y_{\leq N}) - \text{Var}(Y) \leq \mathbb{E}[Y + Y_{\leq N}] \mathbb{E}[Y - Y_{\leq N}] \leq 2\mathbb{E}[Y] 2^d \mathbb{P}(L_2 > N) = o(2^{-d}),$$

and so

$$|\text{Var}(Y) - \text{Var}(Y_{\leq N})| = O(2^{-d}).$$

Hence by Lemma 11, with H_1, \dots, H_r listing a representative of each ambient-isomorphism class of connected cube subgraphs with at most N vertices, we see that $\text{Var}(Y) = (1 + O(dq^d))\mu_1$, as required. \square

Proof of Theorem 3 part (b). Let us show first that

$$d_{TV}(Y, \text{Po}(\lambda)) = O(dq^d). \tag{20}$$

Write $\lambda_{\leq N}$ for $\mathbb{E}[Y_{\leq N}]$. Now $d_{TV}(Y, \text{Po}(\lambda))$ is at most

$$d_{TV}(Y, Y_{\leq N}) + d_{TV}(Y_{\leq N}, \text{Po}(\lambda_{\leq N})) + d_{TV}(\text{Po}(\lambda_{\leq N}), \text{Po}(\lambda)).$$

We consider the three terms in the sum in order. First we have

$$d_{TV}(Y, Y_{\leq N}) \leq \mathbb{P}(Y \neq Y_{\leq N}) = \mathbb{P}(L_2 > N) \leq 2^{-3d} = o(q^d).$$

Next, by Lemma 11 (with H_i as above)

$$d_{TV}(Y_{\leq N}, \text{Po}(\lambda_{\leq N})) = O(dq^d).$$

Finally, for $\mu, \delta > 0$ the sum of independent $\text{Po}(\mu)$ and $\text{Po}(\delta)$ random variables has distribution $\text{Po}(\mu + \delta)$; and so

$$d_{TV}(\text{Po}(\mu), \text{Po}(\mu + \delta)) \leq \mathbb{P}(\text{Po}(\delta) \neq 0) = 1 - e^{-\delta} \leq \delta.$$

Thus

$$d_{TV}(\text{Po}(\lambda_{\leq N}), \text{Po}(\lambda)) \leq \lambda - \lambda_{\leq N} \leq 2^d \mathbb{P}(L_2 > N) \leq 2^{-2d} = o(q^d).$$

Putting this together we obtain (20).

Also, by Lemma 12, it is easy to see that

$$0 \leq \mathbb{E}[X_{\leq N}] - \mu_1 \leq \mathbb{E}[Z_{\leq N}] - \mu_1 = \sum_{t=2}^N t\mu_t \leq 2d(2q^2)^d$$

for d sufficiently large. Thus in particular we have $\lambda_{\leq N} - \mu_1 = o(q^d)$, and so $\lambda - \mu_1 = o(q^d)$ by the above; and we obtain $d_{TV}(Y, \text{Po}(\mu_1)) = O(dq^d)$.

Finally, since also $\text{Var}(Z) \sim \lambda \rightarrow \infty$ as $d \rightarrow \infty$, it follows from (20) that \tilde{Z} is asymptotically standard normal. This completes the proof of part (b), and thus of Theorem 3. \square

6 Joint distribution of components

We saw in Theorem 4 that, for each $t = 1, \dots, m_p$ the number X_t of components of Q_p of size t has close to the Poisson distribution $\text{Po}(\mu_t)$, where $\mu_t = \mathbb{E}[X_t]$. But what about the joint distribution of X_1, \dots, X_{m_p} ?

We can in fact handle this in much the same way as we handled the distribution of a single X_t , but based on Lemma 18 below rather than on the results in Section 2.2. It turns out that the joint distribution of X_1, \dots, X_{m_p} is close to a product of Poisson distributions. Write $\mathcal{L}(X_1, \dots, X_{m_p})$ for the joint law of X_1, \dots, X_{m_p} ; and write $\prod_{j=1}^{m_p} \text{Po}(\mu_j)$ for the distribution of independent random variables $\text{Po}(\mu_j)$. We shall see that

$$d_{TV}(\mathcal{L}(X_1, \dots, X_{m_p}), \prod_{j=1}^{m_p} \text{Po}(\mu_j)) = O(d^2 q^d). \quad (21)$$

Thus, the numbers of components in the fragment of each size t are asymptotically independent, with a Poisson distribution for $t \leq m_p$, and identically 0 for $t > m_p$. Indeed, we have the following much more detailed theorem concerning the small components, in the spirit of Theorem 5. Note that there is a finite set of canonical cube subgraphs with at most m_p vertices.

Theorem 17. *Let H_1, \dots, H_r be a list of $r \geq 1$ distinct canonical cube subgraphs with at most m_p vertices (perhaps all such graphs); and for each j let Y_j be the random number of components of $Q_p = Q_p^d$ ambient isomorphic to H_j , with mean λ_j . Let $t^* = \min_j v(H_j)$, so $t^* \geq 1$. Then*

$$d_{TV}(\mathcal{L}(Y_1, \dots, Y_r), \prod_{j=1}^r \text{Po}(\lambda_j)) = O(d^{t^*+1} q^{t^*d}). \quad (22)$$

When the H_j include all the canonical cube subgraphs of size up to m_p (so $t^* = 1$), Theorem 17 directly implies (21). We cannot quite use Theorem 17 to deduce our earlier individual bounds on d_{TV} , for example on $d_{TV}(X_t, \text{Po}(\mu_t))$ in Theorem 4 part (c), since in the bound (22) there is an ‘extra’ factor d .

To prove Theorem 17 we shall use the following lemma. As in Subsection 2.2, let $(A_i : i \in I)$ be a family of events with a dependency graph L , and write $i \sim j$ if i and j are adjacent in L . For each i , let $\pi_i = \mathbb{P}(A_i)$ and let \mathbb{I}_i be the indicator of A_i . Now we let I be partitioned into $I_1 \cup \dots \cup I_r$ for some $r \geq 1$. For each $j \in [r]$, let $X_j = \sum_{i \in I_j} \mathbb{I}_{A_i}$ and let $\lambda_j = \mathbb{E}[X_j]$. The following lemma is essentially the special case of Theorem 10.K of Barbour, Holst and Janson [2] when all means $\lambda_j \rightarrow \infty$. Sums and products over j or j' always mean over j or j' in $[r]$.

Lemma 18. *With notation as above, assume that each $\lambda_j \rightarrow \infty$ as $d \rightarrow \infty$. Then for d sufficiently large*

$$\begin{aligned} & d_{TV}(\mathcal{L}(X_1, \dots, X_r), \prod_j \text{Po}(\lambda_j)) \\ & \leq \sum_j \frac{\ln(\lambda_j)}{\lambda_j} \sum_{i \in I_j} \pi_i^2 + \sum_j \sum_{j'} \frac{\ln(\lambda_j \lambda_{j'})}{\sqrt{\lambda_j \lambda_{j'}}} \sum_{i \in I_j} \sum_{i' \in I_{j'}} \mathbb{I}_{i \sim i'} (\mathbb{P}(A_i \wedge A_{i'}) + \pi_i \pi_{i'}). \end{aligned}$$

Proof of Theorem 17. As earlier, given d let $\mathcal{S} = \mathcal{S}(d)$ be the set of subgraphs of Q^d ambient isomorphic to one of the graphs H_1, \dots, H_r . List the members of \mathcal{S} as G_1, \dots, G_N ; and let A_i be the event that G_i is a component of Q_p . We let i, i' run over $[N]$ and j, j' run over $[r]$. For distinct i, i' let $i \sim i'$ if either the vertex sets $V(G_i)$ and $V(G_{i'})$ meet or there is an edge of Q^d between them; and note that this gives a dependency graph L . For each j , let $I_j = \{i : G_i \text{ is ambient isomorphic to } H_j\}$.

Now we can apply Lemma 18. We must bound the two terms in the bound in the lemma. First, by (13), there is a constant α such that, for each j ,

$$\sum_{i \in I_j} \pi_i^2 \leq \sum_{i \in I_j} \pi_i \cdot \alpha q^{v(G_i)d} = \lambda_j \cdot \alpha q^{v(H_j)d}.$$

Hence

$$\sum_j \frac{\ln(\lambda_j)}{\lambda_j} \sum_{i \in I_j} \pi_i^2 \leq \alpha \sum_j \ln(\lambda_j) q^{v(H_j)d} = O(dq^{t^*d}) \quad (23)$$

since $\ln(\lambda_j) = O(d)$ uniformly over j .

For the second term, let $j, j' \in [r]$ (not necessarily distinct). For $i \in I_j$ and $i' \in I_{j'}$, as in (14) we have

$$\mathbb{P}(A_i \wedge A_{i'}) \leq \pi_i \pi_{i'} q^{-v(H_j)v(H_{j'})} \leq \pi_i \alpha q^{v(H_{j'})d} q^{-v(H_j)v(H_{j'})}.$$

Hence, arguing as in the proof of (15),

$$\begin{aligned} \beta(j, j') &:= \sum_{i \in I_j} \sum_{i' \in I_{j'}} \mathbb{I}_{i \sim i'} (\mathbb{P}(A_i \wedge A_{i'}) + \pi_i \pi_{i'}) \\ &\leq \sum_{i \in I_j} \pi_i \cdot \alpha q^{v(H_{j'})d} (q^{-v(H_j)v(H_{j'})} + 1) v(H_j)d (ed)^{v(H_{j'})-1} \\ &= \lambda_j \cdot O(d^{v(H_{j'})} q^{v(H_{j'})d}) = \lambda_j \cdot O((dq^d)^{v(H_{j'})}). \end{aligned}$$

Similarly, swapping j and j' , we have

$$\beta(j, j') \leq \lambda_{j'} \cdot O((dq^d)^{v(H_j)});$$

and so

$$\beta(j, j') \leq \sqrt{\lambda_j \lambda_{j'}} \cdot O((dq^d)^{t_{jj'}^*}),$$

where $t_{jj'}^* = \min\{v(H_j), v(H_{j'})\}$. Hence,

$$\frac{\ln(\lambda_j \lambda_{j'})}{\sqrt{\lambda_j \lambda_{j'}}} \beta(j, j') = O(d) O((dq^d)^{t_{jj'}^*}) = O(d^{t^*+1} q^{t^*d}).$$

So, summing over the bounded number of choices of j and j' , we obtain

$$\begin{aligned} &\sum_j \sum_{j'} \frac{\ln(\lambda_j \lambda_{j'})}{\sqrt{\lambda_j \lambda_{j'}}} \sum_{i \in I_j} \sum_{i' \in I_{j'}} \mathbb{I}_{i \sim i'} (\mathbb{P}(A_i \wedge A_{i'}) + \pi_i \pi_{i'}) \\ &= \sum_j \sum_{j'} \frac{\ln(\lambda_j \lambda_{j'})}{\sqrt{\lambda_j \lambda_{j'}}} \beta(j, j') = O(d^{t^*+1} q^{t^*d}). \end{aligned}$$

This result, together with (23) lets us use Lemma 18 to complete the proof of Theorem 17. \square

7 Concluding remarks

In Theorems 1 to 5 and Theorem 17, we have seen quite a full picture of the rich component structure of the random graph $Q_p = Q_p^d$, for fixed p with $0 < p < \frac{1}{2}$. In particular, in Theorem 4 we saw that the number X_t of components in Q_p of size t , with mean μ_t , has close to the Poisson distribution $\text{Po}(\mu_t)$, and thus the standardised version X_t^* has close to the standard normal distribution. In Proposition 13 we gave a partial corresponding local limit result for convergence to the Poisson distribution: it would be interesting to learn more on such local behaviour.

References

- [1] M. Ajtai, J. Komlós and E. Szemerédi, Largest random component of a k -cube, *Combinatorica* **2** (1982), 1-7.
- [2] A. D. Barbour, L. Holst and S. Janson, *Poisson Approximation*, Clarendon Press, Oxford 1992.
- [3] A.J. Bernstein, Maximally connected arrays on the n -cube, *SIAM J. Appl. Math.* **15** (1967), 1485-1489.
- [4] B. Bollobás, *Random Graphs*, 2nd ed., Cambridge University Press, 2001.
- [5] B. Bollobás, Complete matchings in random subgraphs of the cube, *Random Structures and Algorithms* **1** (1990), 95-104.
- [6] B. Bollobás, Y. Kohayakawa and T. Łuczak, The evolution of random subgraphs of the cube, *Random Structures and Algorithms* **3** (1992), 55-90.
- [7] B. Bollobás, Y. Kohayakawa and T. Łuczak, On the diameter and radius of random subgraphs of the cube, *Random Structures and Algorithms* **5** (1994), 627-648.
- [8] B. Bollobás, Y. Kohayakawa and T. Łuczak, Connectivity properties of random subgraphs of the cube, *Random Structures and Algorithms* **6** (1995), 221-230.
- [9] C. Borgs, J. Chayes, R. van der Hofstadt, G. Slade and J. Spencer, Random subgraphs of finite graphs: III the phase transition for the n -cube. *Combinatorica* **26** (2006), 395 – 410.

- [10] Yu. D. Burtin, On the probability of connectedness of a random subgraph of the n -cube, *Problemy Pered. Inf.* (Problems of Information Transmission) **13** (1977), 90 – 95 (in Russian).
- [11] P. Erdős and J. Spencer, Evolution of the n -cube, *Comp. and Math with Appl.* **5** (1979), 33-39.
- [12] C. D. Godsil, Matchings and walks in graphs, *J. Graph Th.* **5** (1981) 285 – 297.
- [13] S. Janson, T. Łuczak and A. Ruciński, *Random Graphs*, Wiley, 2000.
- [14] C. McDiarmid, A. Scott and P. Withers, Uniform multicommodity flow through the hypercube with random edge-capacities. *Random Structures and Algorithms* **50** (2017) 437 – 463 (published online 7 November 2016 in Wiley Online Library).
- [15] K. Weber, On components of random graphs in the n -cube. *Elektron. Inf. verarb. Kybern. EIK* **22** (12) (1986), 601 – 613.
- [16] K. Weber, On components of random subgraphs of the n -cube. In *Random Graphs, Volume 2*, A. Frieze and T. Łuczak eds, pages 263 – 278, Wiley, 1992.