Cover time of random subgraphs of the hypercube

Colin Cooper* Alan Frieze† Wesley Pegden‡

June 3, 2025

Abstract

$Q_{n,p}$, the random subgraph of the n-vertex hypercube Q_n, is obtained by independently retaining each edge of Q_n with probability p. We give precise values for the cover time of $Q_{n,p}$ above the connectivity threshold.

1 Introduction

Let Q_n be the hypercube with $n = 2^d$ vertices and $m = dn/2$ edges where $d = \log_2 n$ is the degree of any vertex. Let $Q_{n,p}$ denote the random subgraph of the hypercube Q_n with n vertices where we retain each edge independently with probability p. The threshold probability p_c for connectivity in $Q_{n,p}$ has been the object of extensive study. The original question as to whether connectivity enjoys a threshold property was answered by Burtin in [5], who proved that $p = 1/2$ is the threshold for connectedness. This study culminated in a proof by Bollobás [4], in the random hypercube process, that w.h.p. the hitting time for connectivity equals the hitting time for minimum degree one. For more on this topic see e.g., [14].

The cover time of a connected graph is the maximum over the start vertex of the expected time for a simple random walk to visit every vertex of the graph. There is a large literature on this subject see for example [2], [15], including results [6]–[9] on various models of random graphs by the authors of this note.

*Department of Informatics, King's College, London WC2B 4BG, England. Research supported at the University of Hamburg, by a Mercator fellowship from DFG Project 491453517
†Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh PA 15213, USA. Research supported in part by NSF grant DMS1952285
‡Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh PA 15213, USA. Research supported in part by NSF grant DMS1700365

Because of its relationship with the Ehrenfest model of diffusion, the random walk on the hypercube has long been an object of study. Diaconis and Shahshahani [10] proved the existence of a cutoff phenomenon for the lazy walk at $T = \frac{1}{4} d \log d$, and Diaconis, Graham and Morrison [11] established the rate of convergence (of the total variation distance) to uniformity in the cutoff window. Matthews [19] proved that the cover time of the hypercube Q_n is $t_{\text{cov}} = (1 + o(1))n \log n$. The proof uses results on the Matthews bound from the paper [18] by the same author. This note gives the w.h.p. cover time of $Q_{n,p}$, the random subgraph of the hypercube, above the connectivity threshold.

Denote $p = \frac{1}{2}(1 + \varepsilon)$, where ε is a parameter used subsequently with this unique meaning. The condition for $Q_{n,p}$ to have minimum degree one, occurs w.h.p. when $d\varepsilon = \omega$ where $\omega \to \infty$ slowly. As this is rather imprecise, and as our proofs are parameterized in terms of $d\varepsilon$, we will consider values of p where $d\varepsilon \geq \theta \log d$, for some small positive constant θ. We assume henceforth that this holds, and thus $Q_{n,p}$ is connected w.h.p.

Theorem 1. *Let $p_c = \frac{1}{2}(1 + \theta \log d/d)$ for some small positive constant θ. Let $t_{\text{cov}}(Q_{n,p})$ denote the cover time of $Q_{n,p}$. For $p \geq p_c$, w.h.p.*

$$t_{\text{cov}}(Q_{n,p}) = (1 + o(1)) \left(\frac{p}{\log 2} \log \frac{2p}{2p - 1} \right) n \log n. \tag{1}$$

Remarks. If $p = (1/2)(1 + \varepsilon)$ where $\varepsilon \to 0$ then

$$t_{\text{cov}} \sim \left(\frac{1}{2 \log 2} \log \frac{1}{\varepsilon} \right) n \log n, \tag{2}$$

so if $d\varepsilon = \ell \log d$, ℓ constant, then $t_{\text{cov}} \sim (1/2 \log 2)\, n \, \log n \, \log \log n$. On the other hand, if ε is constant then $t_{\text{cov}} = \Theta(n \log n)$, and as $p \to 1$ then t_{cov} tends to $n \log n$.

Notation. $G = (V, E)$ is the graph with vertex set V and edge set $E = E(G)$, where we take $V = [n]$ throughout. The degree of a vertex $v \in V$ is denoted by d_v. For $S \subseteq V$, $\deg(S)$ is the degree of set S, where $\deg(S) = \sum_{v \in S} d_v$ and $N(S) = \{w \notin S : \exists v \in S \ s.t. \ \{v, w\} \in E(G)\}$ is the disjoint neighbour set of S. We use $\log x$ for the natural logarithm of x, and $\log_2 x$ for the logarithm base 2. The degree of a vertex in the n-vertex hypercube Q_n is $d = \log_2 n$.

We use $t = 0, 1, ...$ to index time steps, reserve T for a mixing time, and t_{cov} for cover time. We assume $d\varepsilon$ is integer, and if not use the term 'vertices of degree $d\varepsilon$' to denote the vertices of degree $\lfloor d\varepsilon \rfloor$ and $\lceil d\varepsilon \rceil$. We use $dist(u, v)$ as the minimum distance between vertices u, v of a graph.

A sequence of events $\mathcal{E}_n$ occurs *with high probability*, (w.h.p.), if $\lim_{n \to \infty} \mathbb{P}(\mathcal{E}_n) = 1$. We use the standard notation $O(\cdot), o(\cdot)$ etc, this denoting $o_n(\cdot)$ and so on. We use $A_n \sim B_n$ to denote $A_n = (1 + o(1))B_n$ and thus $\lim_{n \to \infty} A_n/B_n = 1$. We use ω to denote a quantity which tends to infinity with n more slowly than any other functions in the given expression. The expression $f(n) \ll g(n)$ indicates $f(n) = o(g(n))$.

2 Background to cover time proof

2.1 The first visit time lemma

Let $G = (V, E)$ be a connected n-vertex graph with $m = |E|$ edges. Let $u \in V$ be arbitrary. Let $\mathcal{W}_u$ denote the random walk $(X(t), t \geq 0)$ starting from $X(0) = u$. The walk defines a reversible Markov chain with state space V. Let P be the matrix of transition probabilities, and $\pi_v = d_v/2m$ the stationary distribution of P. Considering a walk $\mathcal{W}_v$, starting at v, let $r_t = \mathbb{P}(X(t) = v)$ be the probability the walk returns to v at step $t \geq 0$, and thus $r_0 = 1$. Let $R(z)$ generate the sequence $(r_t, t \geq 0)$, and $R(t, z)$ generate the first t entries, $(r_0, \ldots, r_{t-1})$. Thus

$$R(z) = \sum_{t=0}^{\infty} r_t z^t, \qquad\qquad R(t, z) = \sum_{j=0}^{t-1} r_j z^j.$$

Finally, for a fixed value of T to be specified, let $R_v = R(T, 1)$, and note that $R_v \geq r_0 = 1$.

The following first visit time lemma bounds the probability a vertex has not been visited at steps $T, T+1, \ldots, t$.

Lemma 2. The first visit time lemma *[7]*
Let G be a graph satisfying the following conditions

(i) For all $t \geq T$, $\max_{u,x \in V} |P_u^{(t)}(x) - \pi_x| \leq n^{-3}$.

(ii) For some (small) constant $\theta > 0$ and some (large) constant $K > 0$,

$$\min_{|z| \leq 1 + \frac{1}{KT}} |R(T, z)| \geq \theta.$$

(iii) $T\pi_v = o(1)$ and $T\pi_v = \Omega(n^{-2})$.

Let $\mathcal{A}_v(t)$ be the event that the random walk $\mathcal{W}_u$ on graph G does not visit vertex v at steps $T, T+1, \ldots, t$. Then, uniformly in v,

$$\mathbb{P}(\mathcal{A}_v(t)) = \frac{(1 + O(T\pi_v))}{(1 + p_v)^t} + O(T^2 \pi_v e^{-t/KT})$$

where p_v is given by the following formula, with $R_v = R_v(T, 1)$:

$$p_v = \frac{\pi_v}{R_v(1 + O(T\pi_v))}.$$

For the cover time of $Q_{n,p}$ we use the following w.h.p. values of the parameters in Lemma 2. The total degree $2m = (1 + O(1/\sqrt{n}))dnp$, and $\pi_v = d_v/2m$, where $1 \leq d_v \leq d$. The value of $T = O(\log^k n)$ for some constant $k \leq 7$, and thus $T\pi_v = O(\log^k n/n)$. The value of $R_v = 1 + O(1/\log d)$, so $p_v = (d_v/2ndp)(1 + O(1/\log d))$.

As we consider values of $t \geq n \log n$, there are values $\nu_1, \nu_2 = O(1/\log d)$, and $\nu_1 \leq \nu_2$ such that $1 - \nu_1 \leq dnp/2mR_v \leq 1 - \nu_2$, and

$$e^{-(1-\nu_1)\delta_v t/dnp} \leq \mathbb{P}(\mathcal{A}_v(t)) = (1 + O(T\pi_v))e^{-t\pi_v/R_v} \leq e^{-(1-\nu_2)d_v t/dnp}.$$

To tidy things up, write

$$\frac{dnp}{2mR_v} = 1 - \nu \text{ where } \nu = O(1/\log d), \tag{3}$$

is to be understood as a variable which abbreviates the inequality $\nu_1 \leq \nu \leq \nu_2$, and

$$\mathbb{P}(\mathcal{A}_v(t)) = e^{-(1-\nu)d_v t/dnp}. \tag{4}$$

The value of T and Condition (i) of the first visit time lemma will be established in Section 2.3. The claim that $R_v = 1 + O(1/\log d)$ is proved in Section 2.4. Condition (iii) holds as $1 \leq d_v \leq d$ and $m \sim dnp$ where $p \geq 1/2$. We rely on the following lemma (Lemma 18 of [9]) to establish Condition (ii).

Lemma 3. *Let v be a vertex of a connected n-vertex graph G. Let T be a mixing time satisfying Condition* (i) *of Lemma 2. If $T = o(n^3)$, $T\pi_v = o(1)$ and R_v is bounded above by a constant, then Condition* (ii) *of Lemma 2 holds for $\theta = 1/4$ and any constant $K \geq 3R_v$.*

2.2 Properties of $Q_{n,p}$ used in the proofs

Vertices of degree $d\varepsilon$ have a particular significance in the proofs, as values around $d\varepsilon$ determine the cover time of the random walk. For convenience we assume $d\varepsilon$ is integer, and if not take this to mean the union of vertices of degree $\lfloor d\varepsilon \rfloor$ and $\lceil d\varepsilon \rceil$.

For $p \geq p_c$ the following properties of $Q_{n,p}$ hold w.h.p.

P1. Conductance. The conductance of $Q_{n,p}$ is $\Phi = \Omega\left(\frac{1}{d^3 \log d}\right)$.

P2. Minimum degree. For $d\varepsilon > \theta \log d$, the minimum degree at least one.

P3. Distance between low degree vertices. Let $S_L = \{v \in V : d_v \leq L\}$. A vertex v is of low degree if $d_v \leq L$, given in (5) below. Fix the values of h, L to

$$h = \frac{d}{2 \log d}, \qquad L = \frac{100d}{\log d}. \tag{5}$$

No two vertices of degree at most L are within distance h of each other.

P4. Degree of last to be visited vertices. Vertices of degree $\sim d\varepsilon$ are last to be visited.

P5. The number of vertices degree $d\varepsilon$. The number $X(d\varepsilon)$ of vertices of degree $d\varepsilon$ satisfies $X(d\varepsilon) = \mathbb{E}X(d\varepsilon)(1 + o(1))$, where $\mathbb{E}X(d\varepsilon)$ is given by (33).

P6. Distance between vertices of degree $d\varepsilon$. If $\varepsilon \leq 1/100$ no two vertices of degree $d\varepsilon$ are within distance h of each other.

The proofs of these properties are given in the Appendix; P1 in Section 4.1, P2 in Lemma 6.1, P3 in Lemma 7.1, P4 in Section 4.3, P5 in Section 4.4, and P6 in Lemma 7.2.

2.3 Mixing time of the random walk

The *conductance* $\Phi(G)$ of a graph G is

$$\Phi = \min_{\pi(S) \leq 1/2} \frac{|E(S : \overline{S})|}{\deg(S)}.$$

Here $\deg(S) = \sum_{v \in S} d(v)$, $\pi(S) = \frac{\deg(S)}{\deg(G)}$, and $E(S : \overline{S})$ is the set of edges between S and $V \setminus S$ in the G. It follows from [17] that

$$|P_u^{(t)}(x) - \pi_x| \leq (\pi_x/\pi_u)^{1/2}(1 - \Phi^2/2)^t. \tag{6}$$

As we assume $Q_{n,p}$ is connected and the maximum degree is d we have $\pi_x/\pi_u = O(\log n)$. It follows from P1 (see Section 4.1 for the proof) that $\Phi = \Omega(1/\log^3 n \log\log n)$. To satisfy Condition (i) of Lemma 2, we take

$$T = \log^7 n. \tag{7}$$

A walk is *lazy*, if it only moves to a neighbour with probability 1/2 at any step. There are several technical points in our cover time proof which require us to consider lazy walks. Firstly the hypercube is bipartite, and hence periodic. To remove the periodicity we can make the walk lazy. Secondly the bound (6) assumes the walk is lazy.

Making the walk lazy halves the conductance but (7) still holds, and the value of π_v is unchanged. Using a lazy walk asymptotically doubles the cover time, as half the steps are wasted. It also doubles the value of R_v; as the expected number of steps before an exit from v is two. Thus the ratio of these values cancels in (4). Other then this it has a negligible effect on the analysis, and we will ignore it for the rest of the paper and continue as though there are no lazy steps.

2.4 The number of returns in the mixing time

For a random walk X_t starting from a vertex v of a graph G, let $R_v(T)$ denote the expected number of visits to v in T steps. As $X_0 = v$, we have

$$R_v(T) = 1 + \sum_{t=1}^{T} \mathbb{P}(X_t = v).$$

Lemma 4. *Let $p \geq p_c$, and let X_t be a random walk on $Q_{n,p}$. Then w.h.p. for all $v \in V$ and all $T = O(\log^k n)$, k constant, $R_v(T) = 1 + O(1/\log d)$.*

Proof. As in (5) of P3, fix the values of h, L to $h = \frac{d}{2\log d}$ and $L = \frac{100d}{\log d}$. Let $t_0 = d/\log^2 d$. The proof is in three steps, from $t \leq t_0$, from $t_0 < t \leq h$, and from $h \leq t \leq T = \log^k n$.

We first consider the case where, with the possible exception of v itself, no vertex within distance h of v has degree at most L.

$$\sum_{\tau=1}^{t_0} \mathbb{P}(X_\tau = v) \leq \sum_{\tau=1}^{t_0} \frac{1}{L}\mathbb{P}(X_{\tau-1} \in N(v)) \leq \frac{t_0}{L} = O\left(\frac{1}{\log d}\right).$$

Let v_i be a vertex at distance $1 \leq i \leq h$ from v. The probability the distance to v decreases to $i-1$ at the next step is at most i/L, and the probability it increases to $i+1$ is at least $(L-i)/L$. Thus the drift away from v per step is at least

$$\mu \geq \frac{L-h}{L} - \frac{h}{L} = \frac{L-2h}{L} = \frac{99}{100}.$$

In $t \leq h$ steps the expected displacement of the walk from v is at least $t\mu$. Let $dist(X_t, v)$ be the actual displacement. For $\delta > 0$ constant

$$\mathbb{P}(dist(X_t, v) \leq (1-\delta)t\mu) \leq e^{-\Omega(\delta^2 t\mu)}.$$

Thus

$$\sum_{\tau=t_0}^{h} \mathbb{P}(X_\tau = v) \leq he^{-\Omega(\delta^2 t_0 \mu)} = o\left(\frac{1}{\log d}\right).$$

It follows that, w.h.p., in h steps the walk is at least distance $H = \mu h(1-2\delta)$ from v, where $H \geq 2d/(5\log d)$, say. Let

$$\widehat{q} = \frac{L-h}{L} = \frac{199}{200}, \qquad \widehat{p} = \frac{h}{L} = \frac{1}{200}.$$

Consider a biassed random walk with transition probabilities $\widehat{p}$ of one step left and $\widehat{q}$ of one step right, setting out from $H-1$ on the integer line $\{0, 1, ..., H\}$. The probability the walk reaches the origin v before returning to H is

$$\frac{\left(\frac{\widehat{q}}{\widehat{p}}\right) - 1}{\left(\frac{\widehat{q}}{\widehat{p}}\right)^H - 1} = O(\mu^{-h}) = O\left(\left(\frac{199}{200}\right)^{2d/5\log d}\right).$$

Thus, with $T = \log^k n$, for any constant k,

$$\sum_{\tau=h}^{T} \mathbb{P}(X_\tau = v) \leq O\left(\left(\frac{199}{200}\right)^{2d/5\log d}\right) = o\left(\frac{1}{\log d}\right).$$

Next consider the case where vertex w is one of the at most 2 vertices of degree at most L is within distance h of v. If $w \in N(v)$ this can increase the expected returns by $O(1/L)$. Suppose w is a distance $i \geq 2$ from v. In the worst case assume the walk always returns to level $i-1$ (a wasted move). Deleting all edges between w and its neighbours leaves all vertices within distance h of v with degree at least $L-1$. This has a negligible effect on the analysis given above. $\square$

3 The cover time of $Q_{n,p}$. Proof of Theorem 1

3.1 Proof outline

Before proceeding we give a quick sketch of the upper and lower bound proofs, as this will motivate the subsequent calculations.

Let $X_p(i)$ be the number of vertices of degree i in $Q_{n,p}$. Let $q = 1-p$, then

$$\mathbb{E}X_p(i) = n\binom{d}{i}p^i q^{d-i}. \tag{8}$$

Recall that $\mathcal{A}_v(t)$ given in (4) is an upper bound on the probability vertex v is unvisited at step t. Let $S(t)$ be the vertices 'still surviving' at step t.

$$S(t) = \sum_{v\in V}\mathbb{P}(\mathcal{A}_v(t)) \sim \sum_{v\in V} e^{-d_v t/dnp}.$$

Thus

$$\mathbb{E}S(t) \sim \sum_{i\geq 1}\mathbb{E}X_p(i)e^{-it/dnp} \quad \sim \quad n(1-p+pe^{-t/ndp})^d.$$

Put $t = \alpha ndp$ and equate $\mathbb{E}S(t) = 1$. Using $d = \log_2 n = \log_e n/\log_e 2$,

$$\log \mathbb{E}S \sim \log n + d\log(1-p+pe^{-\alpha}) = \frac{\log n}{\log 2}(\log 2 + \log(1-p+pe^{-\alpha})).$$

$$\log \mathbb{E}S = 0 \iff \log 2 + \log(1-p+pe^{-\alpha}) = 0 \iff (1-p+pe^{-\alpha}) = 1/2.$$

Solving this gives $\alpha = \log 2p/(2p-1)$, which suggests the following result.

$$t_{\text{cov}} \sim ndp \cdot \log \frac{2p}{2p-1}.$$

For the corresponding lower bound we prove that vertices of degree $d\varepsilon$ maximize the above calculations, where $d\varepsilon$ is somewhat larger than the minimum degree. We prove by direct construction that w.h.p. at some step t slightly below t_{cov} there are many vertices of degree $d\varepsilon$ which are unvisited by the walk. We now proceed to the details of the above proof idea.

3.2 Upper bound on the cover time

Let $T(u)$ be the time taken by the random walk $\mathcal{W}_u$ to visit every vertex of a connected graph G, and $t_{\text{cov}}(u) = \mathbb{E}T(u)$. Let U_t be the number of vertices of G which have not been visited by $\mathcal{W}_u$ at step t. We note the following:

$$\begin{aligned} t_{\text{cov}}(u) = \mathbb{E}(T(u)) &= \sum_{t>0} \mathbb{P}(T(u) \geq t), && (9) \\ \mathbb{P}(T(u) \geq t) = \mathbb{P}(T(u) > t-1) &= \mathbb{P}(U_{t-1} > 0) \leq \min\{1, \mathbb{E}(U_{t-1})\}. && (10) \end{aligned}$$

As in (4), let $\mathcal{A}_v(t)$, $t \geq T$ be the event that $\mathcal{W}_u(t)$ has not visited v in the interval $[T, t]$. It follows from (9), (10) that for all $t \geq T$,

$$t_{\text{cov}}(u) \leq t + 1 + \sum_{s \geq t} \mathbb{E}(U_s) \leq t + 1 + \sum_{v \in V} \sum_{s \geq t} \mathbb{P}(\mathcal{A}_s(v)) \tag{11}$$

and

$$\sum_{s \geq t} \mathbb{P}(\mathcal{A}_v(s)) \leq \sum_{s \geq t} e^{-(1-\nu)d_v s/dnp} \leq \frac{dnp}{(1-\nu)d_v} e^{-(1-\nu)d_v t/dnp}, \tag{12}$$

where $\nu = O(1/\log d)$. Let $X(i)$ be the number of vertices of degree i. The above argument implies that

$$Q(t) = \sum_{v \in V} \sum_{s \geq t} \mathbb{P}(\mathcal{A}_s(v)) = O\left(dnp\right) \sum_{i=1}^{d} X(i) e^{-(1-\nu)it/dnp}. \tag{13}$$

The argument given in Section 3.1 can now be adapted to give an upper bound on the cover time. Let $X(i)$ be the number of vertices of degree i in $Q_{n,p}$. Let $b = d\omega$, then

$$\mathbb{P}(X(i) \geq b\mathbb{E}X(i)) \leq \frac{1}{b}, \tag{14}$$

so with probability $1 - O(1/\omega)$ this upper bound of $d\omega\mathbb{E}X(i)$ on the number of vertices of degree i holds simultaneously for all $i \in \{0, 1, ..., d\}$.

Let $\delta = (\log db)/\log n = o(1)$, and $t_U = \alpha\ (ndp)/(1-\nu)$ where

$$\alpha = \log \frac{p}{p - 1 + \left(\frac{1}{2}\right)^{1+\delta}} = \log \frac{2p}{2p - 1 - O(\delta)}. \tag{15}$$

Let $S(t)$ be the number of unvisited vertices at t. By the above estimate (8), (13), (14), and with $t_U = \alpha dnp/(1-\nu))$, then w.h.p.,

$$\begin{aligned} S(t_U) &= \sum_{v\in V} \mathbb{P}(\mathcal{A}_v(t_U)) && (16)\\ &\leq b \sum_{i\geq 1} \mathbb{E}X(i) e^{-it_U(1-\nu)/dnp} = bn(1-p+pe^{-\alpha})^d, && (17)\\ Q(t_U) &= O(dnp)\ S(t_U) \ = \ O(dnp)bn(1-p+pe^{-\alpha})^d. && (18) \end{aligned}$$

However, by (15)

$$(1-p+pe^{-\alpha})^d = 2^{-(1+\delta)d} = n^{-(1+\delta)},$$

so

$$Q(t_U) = O(1)n^2 dbn^{-(1+\delta)} = O(n).$$

Thus by (11), for any $u \in V$,

$$t_{\rm cov}(u) \leq t_U + 1 + Q(t_u) = t_U + O(n)$$

Finally

$$t_{\rm cov} \leq (1+o(1))\ ndp\ \log \frac{2p}{2p-1}.$$

3.3 Lower bound on the cover time

Let $S(0)$ be the set of vertices of degree $d\varepsilon$ in $Q_{n,p}$. We construct a subset of $S(0)$ which is still unvisited w.h.p. at $t_L = t_U(1-o(1))$. By (33) below

$$\mathbb{E}|S(0)| = \mathbb{E}X(d\varepsilon) \sim \frac{1}{\sqrt{2\pi d\varepsilon(1-\varepsilon)}} \left(\frac{1+\varepsilon}{\varepsilon}\right)^{d\varepsilon}, \tag{19}$$

and by property P5, $|S(0)| = (1+o(1))\mathbb{E}S(0)$.

The function $f(x) = ((1+x)/x)^x$ is monotone increasing from one for $x \in (0,1]$. Thus for any $d\varepsilon \geq \theta \log d$, (i.e., $p \geq p_c$, see paragraph preceding Theorem 1), the value of $|S(0)|$ is much greater than $T = \log^7 n$, as given in (7). We remove any vertices visited during T from $S(0)$ to apply the results of Lemma 2.

Let $t_L = (1-\delta)ndp \log 2p/(2p-1)$ where $\delta = o(1)$ is given by (22) below. At step t_L, given the value of $|S(0)|$,

$$\mathbb{E}|S(t_L)| \sim |S(0)| e^{-(1-\nu)d\varepsilon t_L/dnp} \sim \frac{1}{\sqrt{2\pi d\varepsilon(1-\varepsilon)}} \left(\frac{1+\varepsilon}{\varepsilon}\right)^{\delta d\varepsilon(1-\nu)}, \tag{20}$$

where the expectation is with respect to the random walk.

Case where $\varepsilon \leq 1/100$. It follows from Lemma 7 that w.h.p. for $\varepsilon \leq 1/100$ all vertices of degree $d\varepsilon$ are at least a distance $h = d/2\log d$ apart. Choose two vertices $v, w \in S$, let the graph distance between them be $\ell \geq h$. Coalesce these into a single vertex $\gamma = \gamma(v,w)$, to form a graph $\Gamma(v,w)$. We claim $R_\gamma = 1 + O(1/\log d)$. The proof is similar to that of Lemma 4. Project the walk starting from γ onto an integer line length $h/2$, with γ identified with zero, and a loop at $h/2$.

Let $Y_x = Y_x(t)$ be the indicator that $\mathcal{W}_u$ has not visited vertex x at t. As $R_v, R_w, R_\gamma = 1 + O(1/\log d)$, and $d_\gamma = 2d\varepsilon$ it follows that

$$\mathbb{E}Y_vY_w = e^{-(1-o(1))2d\varepsilon t/ndp} = \left(e^{-(1-o(1))d\varepsilon t/ndp}\right)^2 = (1+o(1))\mathbb{E}Y_v\mathbb{E}Y_w,$$

and so

$$\mathbb{E}|S(t)^2| = |S(0)|(|S(0)|-1)e^{-(1+o(1))2d\varepsilon t/ndp} + |S(0)|e^{-(1+o(1))d\varepsilon t/ndp}. \tag{21}$$

Choose δ so that $\sqrt{d}\varepsilon^{\delta d\varepsilon} = o(1)$. This is satisfied by

$$\delta = \frac{\log d}{d\varepsilon \log 1/\varepsilon} = O\left(\frac{1}{\log\log d}\right) = o(1). \tag{22}$$

By (20) and (21),

$$\mathbb{P}(|S(t_L)| \neq 0) \geq (1-o(1))\frac{(\mathbb{E}|S(t_L)|)^2}{\mathbb{E}|S(t_L)|^2} = 1 - \frac{O(1)}{\mathbb{E}|S(t_L)|}.$$

Using (22) in (20), we see that $\mathbb{E}|S(t_L)| \to \infty$ and thus $\mathbb{P}(|S(t_L)| \neq 0) = 1 - o(1)$ as required.

Case where $\varepsilon \geq 1/100$. Let $F_\ell(v) = \{w : \text{dist}(v,w) \leq \ell\}$; where $\text{dist}(v,w)$ is graph distance in Q_n, and ℓ is to be determined. Let V^* be some maximal set of vertices of Q_n such that for all $u, v \in V^*$, $\text{dist}(u,v) > \ell$. Then $|F_\ell(v)| \leq d^\ell$, and so $|V^*| \geq n/2d^\ell$.

Let $B(0) = \{v : v \in V^*,\ d_v = d\varepsilon\}$, where d_v is the degree of v in $Q_{n,p}$. Then $B(0) \subseteq S(0)$ as defined above, and as w.h.p. $|B(0)| = (1+o(1))\mathbb{E}|B(0)|$,

$$|B(0)| \geq |V^*|\frac{1}{n}\frac{1}{3\sqrt{d}}\left(\frac{1+\varepsilon}{\varepsilon}\right)^{d\varepsilon} \geq \frac{1}{6d^{\ell+1/2}}\left(\frac{1+\varepsilon}{\varepsilon}\right)^{d\varepsilon}.$$

Thus at $t = t_L$,

$$\mathbb{E}|B(t)| \geq \frac{1}{6d^{\ell+1/2}}\left(\frac{1+\varepsilon}{\varepsilon}\right)^{\delta d\varepsilon},$$

where we require $\mathbb{E}|B(t)| = \omega \to \infty$, say. This is satisfied for large ℓ by any

$$\delta \geq \frac{10}{\varepsilon\log(1+1/\varepsilon)}\frac{\ell\log d}{d} \geq \frac{C\ell\log d}{d},$$

for some constant C, as $\varepsilon \geq 1/100$. Choose $\ell = d/(\omega\log d)$, then $\delta = O(1/\omega)$ and $\mathbb{E}|B(t_L)| \to \infty$ as required. With this value of ℓ, for any pair $v, w \in B(t)$ an argument similar to Section 2.4, (with the simplifying fact from Lemma 6.2, that if ε is constant, then $\delta = \alpha_0 d\varepsilon$ for some constant $\alpha_0 \in (0,1)$), that will ensure that $R_{\gamma(v,w)} = 1 + o(1)$. The rest of the proof is similar to the previous case.

References

[1] M. Ajtai, J. Komlós, and E. Szemerédi. Largest random component of a k–cube. Combinatorica, 2(1):1–7, (1982).

[2] D. Aldous and J. A. Fill. Reversible Markov Chains and Random Walks on Graphs. Unfinished monograph, 2002 (Recompiled version, 2014).

[3] A. Beveridge, A. Frieze, and C. McDiarmid. Random minimum length spanning trees in regular graphs. Combinatorica, 18(3):311–333, (1998).

[4] B. Bollobás, Complete matchings in random subgraphs of the cube. Random Structures and Algorithms 1 (1990) 95-104.

[5] Y. D. Burtin, On the probability of connectedness of a random subgraph of the n-cube. (in Russian) Problemy Peredaci Informacii 13 (1977) 90-95.

[6] C. Cooper and A. M. Frieze, The cover time of random regular graphs. SIAM Journal on Discrete Mathematics, 18 (2005) 728-740.

[7] C. Cooper and A. M. Frieze, The cover time of the giant component of $G_{n,p}$. Random Structures and Algorithms, 32 (2008) 401-439.

[8] C. Cooper and A. M. Frieze, The cover time of random geometric graphs. Random Structures and Algorithms, 38 (2011) 324-349.

[9] C. Cooper, A. M. Frieze and T. Radzik, The cover times of random walks on hypergraphs. Theoretical Computer Science, 509 (2013) 51-69.

[10] P. Diaconis and M. Shahshahani, Time to reach stationarity in the Bernoulli-Laplace diffusion model, SIAM J. Math. Anal., 18, 208–218 (1987).

[11] P. Diaconis, R. L. Graham and J. A. Morrison, Asymptotic analysis of a random walk on a hypercube with many dimensions, Random Structures and Algorithms, 1(1) (1990) 51-72.

[12] J. Erde, M. Kang and M. Krivelevich. Expansion in supercritical random subgraphs of the cube and its consequences. Random Structures and Algorithms, 65 (2024) 576-600.

[13] W. Feller, *An Introduction to Probability Theory, Volume I.* (Second edition) Wiley (1960).

[14] A. M. Frieze and M. Karoński. *Introduction to Random Graphs.* Cambridge University Press, (2015).

[15] D. A. Levin and Y. Peres. *Markov Chains and Mixing Times.* American Mathematical Society, (2017).

[16] L. H. Harper. Optimal numberings and isoperimetric problems on graphs. J. Combinatorial Theory 1, 385–393, (1966).

[17] M. Jerrum and A. Sinclair. The Markov chain Monte Carlo method: an approach to approximate counting and integration. In *Approximation Algorithms for NP-hard Problems.* (D. Hochbaum ed.) PWS (1996) 482-520

[18] P. Matthews. Covering problems for Brownian motion on spheres. Ann. Prob. 16 (1988) 189-199.

[19] P. Matthews. Some sample path properties of a random walk on the cube. J. Theor. Probab. 2 (1989) 129–146.

4 Appendix: Conductance and other technical details

4.1 Conductance of $Q_{n,p}$

The conductance $\Phi = \Phi_G$ of a graph $G = (V, E)$ is defined as

$$\Phi_G = \min_{\substack{S \subset V(G) \\ 0 < \pi(S) \leq 1/2}} \frac{e(S : \overline{S})}{d_G(S)},$$

where $d(S) = \sum_{v \in S} d_v$ is the degree of a set of vertices S in the graph G, $\overline{S} = V \setminus S$ and $e(S : \overline{S}) = |E(S : \overline{S})|$. The expression $\pi(S) \leq 1/2$ is equivalent to $d(S) \leq |E(G)|$; multiply the former by $2|E(G)|$ to obtain the latter.

The edge isoperimetric inequality for the hypercube, Harper [16], states that

$$\min_{\substack{S \subseteq V \\ |S| \leq n/2}} \left\{ |E(S, \overline{S})| \right\} \geq |S|(d - \log_2 |S|). \tag{23}$$

The bound is tight for sets S which are vertices of a subcube of Q_n. For random sub-hypercubes we have the following lower bound.

Proposition 5. *With high probability $Q_{n,p}$ has conductance*

$$\Phi_{Q_{n,p}} = \Omega\left(\frac{1}{d^3 \log d}\right).$$

Proof. **Case 0.** $1 \leq |S| \leq \sqrt{d}$. By Lemma 7.1, w.h.p. vertices of degree at most $L = 100d/\log d$ are distance at least $d/2 \log d$ apart. Let $S_1 \subseteq S$ be vertices of degree at most L and $S_2 = S \setminus S_1$. Then

$$e(S_1, \overline{S}) = d(S_1) \quad \text{and} \quad e(S_2, \overline{S}) \geq 100|S_2|d/\log d - 2|S_2| \log_2 |S_2| \geq |S_2|d/\log d,$$

and $d(S) \leq d(S_1) + d|S_2|$. It follows that

$$\Phi_S \geq \frac{1}{\log d}. \tag{24}$$

Case 1: $\sqrt{d} \leq |S| \leq n/3d$. Referring to (23), as $p > 1/2$ the number of retained edges $e(S : \overline{S})$ is at least $X \sim Bin(s(d - \log_2 s), 1/2)$. Thus,

$$\mathbb{P}(\exists S : \sqrt{d} \leq |S| \leq n/3d, e(S : \overline{S}) \leq s(d - \log_2 s)/d)$$

$$\leq \sum_{s=\sqrt{d}}^{n/3d} n(ed)^{s-1}\, \mathbb{P}(X \leq s(d - \log_2 s)/d) \tag{25}$$

$$\leq \sum_{s=\sqrt{d}}^{n/3d} n(ed)^{s} \frac{(ed)^{s(d-\log_2 s)/d}}{2^{s(d-\log_2 s)}} \tag{26}$$

$$= \sum_{s=\sqrt{d}}^{n/3d} \left(\frac{s\, 2^{d/s}}{(ed)^{(\log_2 s)/d}} \frac{(ed)^2}{n} \right)^s \tag{27}$$

$$\leq \sum_{s=\sqrt{d}}^{n/3d} \left(\frac{e^{1+o(1)}}{3} \right)^s = o(1).$$

In (25) we used the estimate $(ed)^{s-1}$ as an upper bound on the number of trees of size s in Q_n, rooted at a fixed vertex, see [3]. If S induces more than one component this can only increase the number of edges to $\overline{S}$. Equation (26) used the following.

$$\left(\frac{k}{N}\right)^k \sum_{i=0}^{k} \binom{N}{i} \leq \left(\frac{k}{N}\right)^k \sum_{i=0}^{k} \frac{N^i}{i!} = \sum_{i=0}^{k} \frac{k^i}{i!} \left(\frac{k}{N}\right)^{k-i} \leq e^k. \tag{28}$$

Here $N = s(d - \log_2 s)$ and $k = N/d$. The bracketed term in (27) has a unique minimum at $s = d \log 2/(1 - (\log_2 ed)/d) > n/3d$. Therefore the maximum value in the bracket in the sum occurs at $s = n/3d$.

If $s \leq n/3d$, then $s(d - \log_2 s)/d \geq s \log_2 d/d$. Thus for $|S| \leq n/3d$,

$$\frac{e(S : \overline{S})}{d(S)} = \Omega\left(\frac{\log_2 d}{d^2}\right). \tag{29}$$

Case 2: $|S| \geq n/3d$. It follows from [1], and Theorem 1.4 of [12] respectively that given $\delta > 0$ there exists constants $c_1, c_2 > 0$ such that if $q = \frac{c}{d}, c \geq c_1$ then $Q_{n,q}$ contains a subgraph H such (i) $|V(H)| \geq (1 - \delta)n$ and (ii) H is a $(c_2/d^2 \log d)$-expander. A graph G is an α-expander if $|N(S)| \geq \alpha |S|$ for all $S \subseteq V(G)$ for which $|S| \leq |V(G)|/2$.

By the above, $Q_{n,p}$ contains the union of $h \sim dp/c$ independent and uniformly chosen vertex subsets $H_1, H_2, \ldots, H_h \subseteq Q_n$, each of which induces an expander. Let $\Gamma = \cup_{i=1}^{h} H_i$, so that $\Gamma \subseteq Q_{n,p}$. The graph Γ and each independent copy $H = H_i$ have the following properties w.h.p.:

P(i). $|E(H_i)| \sim \frac{1}{2}cn$.

P(ii). $\displaystyle\sum_{v:d_H(v)\notin[.99c,1.01c]} d_H(v) \leq ne^{-\Omega(c)}$.

P(iii). $\displaystyle\sum_{v:d_\Gamma(v)\notin[.99dp,1.01dp]} d_\Gamma(v) \leq ne^{-\Omega(d)}$.

P(iv). $|E(\Gamma)| \sim \frac{1}{2}ndp$.

We need to estimate

$$\Phi_\Gamma = \min\left\{\Phi_S : d_\Gamma(S) \leq |E(\Gamma)|\right\}, \text{ where } \Phi_S = \frac{e_\Gamma(S:\overline{S})}{d_\Gamma(S)}.$$

It follows from P(iii) that for $n/3d \leq |S| \leq 2n/3$

$$0.98dp|S| \leq 0.99dp|S| - ne^{-\Omega(d)} \leq d_\Gamma(S) \leq 1.01dp|S| + ne^{-\Omega(d)} \leq 1.02dp|S|.$$

So $d_\Gamma(S) = \kappa d|S|$ for some constant κ, $0.98p \leq \kappa \leq 1.02p$, and $d_\Gamma(S) \leq |E(\Gamma)|$ holds for $|S| \leq 3n/5$.

A similar argument using P(ii) implies that if $S \subseteq V(H_i)$ and $n_i = |V(H_i)|$ then

$$0.98c|S| \leq 0.99c|S| - n_ie^{-\Omega(c)} \leq d_\Gamma(S) \leq 1.01c|S| + n_ie^{-\Omega(c)} \leq 1.02c|S|.$$

Also, with $h = dp/c$ and $\delta = .0001$,

$$\begin{aligned}
&\mathbb{P}(n/3d \leq |S| \leq 3n/5 : |S \cap V(H_i)| < 0.99|S| \text{ for all values } i = 1, ..., h) \\
&\quad\leq \sum_{s=n/3d}^{3n/5} \binom{n}{s} [\mathbb{P}(Bin(s, 1-\delta) < 0.99s)]^h \\
&\quad\leq \sum_{s=n/3d}^{3n/5} \left(\frac{ne}{s}\right)^s e^{-\Omega(ds/c)} = \sum_{s=n/3d}^{3n/5} \left(\frac{ne}{s}e^{-\Omega(d/c)}\right)^s = o(1). \qquad (30)
\end{aligned}$$

It follows that if $n/3d \leq |S| \leq 3n/5$, there exists H_i such if $T = S \cap V(H_i)$ then $|T| \geq 0.99|S|$.

Next let T' be the smaller of $|T|, |V(H_i) \setminus T|$ and note that $|T'| \geq |S|/2$. By Theorem 1.4 of [12] the set T' has at least $c_2|T'|/d^2 \log d$ neighbours in $|V(H_i) \setminus T|$, and as by definition $S \setminus T$ is disjoint from H_i,

$$e(S:\overline{S}) \geq e(T' : V(H_i) \setminus T') \geq \frac{c_2|T'|}{d^2 \log d} \geq \frac{c_2|S|}{2d^2 \log d}.$$

In summary, for $n/3d \leq$S$\leq 3n/5$,

$$\frac{e(S:\overline{S})}{d_\Gamma(S)} = \Omega\left(\frac{1}{d^3 \log d}\right).$$

The claim of Proposition 5 then follows from (24), (29) and the above. □

4.2 Various supporting lemmas

Minimum degree: General bounds

Lemma 6. *The following hold w.h.p. in $Q_{n,p}$, for $p = (1+\varepsilon)/2$.*

1. *If $d\varepsilon = \omega \to \infty$ there are no vertices of degree zero. Moreover, if $d\varepsilon = (i-1+\theta)\log d$, where i is a fixed integer and $\theta \in (0,1)$ constant, the minimum degree δ is i.*
2. *If ε is constant then $\delta \geq \alpha_0 d\varepsilon$ for some constant $\alpha_0 \in (0,1)$.*

Proof. **Case 1.** The expected number of vertices of degree zero is nq^d which tends to zero for any $d\varepsilon = \omega \to \infty$. Let X_j denote the number of vertices of degree j. Then

$$\mathbb{E}X_j = n\binom{d}{j}p^j q^{d-j} = \binom{d}{j}(1+\varepsilon)^j(1-\varepsilon)^{d-j} \tag{31}$$

So, if $j \leq i-1$ then

$$\mathbb{E}X_j \leq \left(\frac{de(1+\varepsilon)}{(1-\varepsilon)j}\right)^j d^{-(i-1+\theta)} = o(1).$$

Whereas

$$\mathbb{E}X_i \geq \left(\frac{d(1+\varepsilon)}{(1-\varepsilon)i}\right)^i d^{-(i-1+\theta)} \to \infty.$$

An application of the Chebychev inequality will show that $X_i > 0$ w.h.p.

Case 2. Putting $j = \alpha d\varepsilon$, where $\alpha < 1/3$, we obtain from (31) that

$$\mathbb{E}X_j \leq \binom{d}{\alpha d\varepsilon}e^{-(d-2\alpha)\varepsilon^2} \leq \left(\frac{e}{\alpha\varepsilon}\right)^{\alpha d\varepsilon} e^{-d\varepsilon^2/3} \leq e^{-d\varepsilon^2/4},$$

for small α. Taking the union bound over at most d values for j, we see that the minimum degree is at least $\alpha d\varepsilon$ w.h.p. for some small $\alpha > 0$ constant. ☐

Low degree vertices. The following argument concerns the distance h between low degree vertices. Fix the values of h, L to

$$h = \frac{d}{2\log d}, \qquad L = \frac{100d}{\log d}.$$

Say a vertex v is of 'low degree' if $d_v \leq L$ and let $S_L = \{v \in V : d_v \leq L\}$. For large values of p, by Lemma 6.2 above, Lemma 7.1 holds with $S_L = \emptyset$.

Lemma 7. *Let $p = \frac{1}{2}(1+\varepsilon)$, $p \geq p_c$, then the following hold w.h.p.:*

1. *No two vertices of degree at most L are within distance h of each other.*

2. *If $\varepsilon \leq 1/100$, no two vertices of degree at most $(101/100)d\varepsilon$ are within distance h of each other.*

Proof. The probability there exist two vertices of S_L are within distance $\ell \leq h$ is

$$\begin{aligned} P(h) &\leq n\sum_{i=1}^{h} d^i \left(\sum_{\ell \leq L}\binom{d}{\ell}p^\ell q^{d-\ell}\right)^2 \\ &= O(1)nd^h\left(\binom{d}{L}p^L q^{d-L}\right)^2 \\ &\leq \frac{O(1)}{n}\left(\frac{dep}{Lq}\right)^{2L} d^h(1-\varepsilon)^{2d} \\ &\leq \frac{e^{d/2}}{n}\left(\frac{e\log d}{50}\right)^{200d/\log d} \qquad = o(1). \end{aligned}$$

Indeed, let $u, w \in S_L$ and let $uv_1 \cdots v_\ell w$ be a path between them of length $\ell \leq h$. The number of paths length $\ell \leq h$ is at most hd^h. The n on the first line upper bounds the number of choices for u, and the last term upper bounds the probability that the vertices u, w have degree at most L. The third line follows from $p/q \leq 2$ provided $\varepsilon \leq 1/3$, and $d^h = e^{d/2} = n^{1/\log 4}$.

The second case is similar but requires the further information (see (34) of Section 4.3 below) that the probability a vertex has degree at most $101d\varepsilon/100$ is at most $\Theta(1)((1+\varepsilon)/\varepsilon)^{101\varepsilon d/100}/n$ in which case the probability $P(h)$ of the stated event satisfies

$$P(h) \leq \frac{O(1)}{n}\left(\frac{1+\varepsilon}{\varepsilon}\right)^{202\varepsilon d/100} d^h.$$

For $P(h) = o(1)$, we require that

$$\left(\log 2 - \frac{1}{2} - \frac{202}{100}\varepsilon\log\frac{1+\varepsilon}{\varepsilon} - o(1)\right) > 0.$$

The function $x\log(1+x)/x$ is monotone increasing for $x \in (0, 1/(e-1)]$, so the above condition holds for $\varepsilon \leq 1/100$. □

4.3 Degree of the last to be visited vertices.

Let

$$N(i) = \frac{d^d}{i^i(d-i)^{d-i}}\left(\frac{p}{q}\right)^i q^d,$$

so that

$$\mathbb{E}X_p(i) = N(i)\, n\sqrt{\frac{d}{2\pi i(d-i)}}(1+o_d(1)).$$

For $p \geq p_c$, $\varepsilon d \geq \theta \log d$ (for some $\theta > 0$ constant). Thus for $\varepsilon < 1$, $\sqrt{d/(d\varepsilon(d - d\varepsilon))} = o(1)$, whereas if $i \geq 1$ constant then $\sqrt{d/i(d-i)} = \Theta(1)$.

If $i = dp$ then $d - i = dq$ so

$$N(dp) = \frac{d^d}{(dp)^{dp}(dq)^{dq}}\, p^{dp} q^{dq} = 1.$$

For $0 \leq x < p$, as $d - d(p + x) = d(q + x)$,

$$\begin{aligned} N(d(p-x)) &= \frac{d^d p^{d(p-x)} q^{d(q+x)}}{(d(p-x))^{d(p-x)}(d(q+x))^{d(q+x)}} \\ &= \frac{p^{d(p-x)}}{(p-x)^{d(p-x)}} \frac{q^{d(q+x)}}{(q+x)^{d(q+x)}}. \end{aligned} \tag{32}$$

Now,

$$N(d\varepsilon) = \frac{1}{2^d}\left(\frac{1+\varepsilon}{\varepsilon}\right)^{\varepsilon d} = \frac{1}{n}\left(\frac{2p}{2p-1}\right)^{d(2p-1)}.$$

Thus

$$\mathbb{E}X(d\varepsilon) \sim \frac{1}{\sqrt{2\pi d\varepsilon(1-\varepsilon)}} \left(\frac{1+\varepsilon}{\varepsilon}\right)^{\varepsilon d}. \tag{33}$$

Next, for $\alpha \in (-1, 1)$,

$$\begin{aligned} [N(d\varepsilon(1-\alpha))]^{1/d} &= \frac{1}{2}\left(\frac{(1+\varepsilon)}{\varepsilon(1-\alpha)}\right)^{\varepsilon(1-\alpha)} \left(\frac{(1-\varepsilon)}{1-\varepsilon(1-\alpha)}\right)^{1-\varepsilon(1-\alpha)} \\ &= \frac{1}{2}\left(\frac{1+\varepsilon}{\varepsilon}\right)^{\varepsilon(1-\alpha)} \frac{1}{(1-\alpha)^{\varepsilon(1-\alpha)}} \left(\frac{1-\varepsilon}{1-\varepsilon+\alpha\varepsilon}\right)^{1-\varepsilon+\alpha\varepsilon} \\ &= \frac{1}{2}\left(\frac{1+\varepsilon}{\varepsilon}\right)^{\varepsilon(1-\alpha)} G_\varepsilon(\alpha). \end{aligned} \tag{34}$$

We next prove that the function $G_\varepsilon(\alpha)$ has a maximum $G_\varepsilon(0) = 1$ at $\alpha = 0$, and that

$$G_\varepsilon(\alpha) = e^{-\Theta(\alpha^2\varepsilon)},$$

is monotone decreasing from this for $\alpha \in (-1, 1)$.

Let $F(\alpha) = \log G_\varepsilon(\alpha)$, then $F(0) = 0$,

$$F'(\alpha) = \varepsilon \log \frac{(1-\varepsilon)(1-\alpha)}{1-\varepsilon+\alpha\varepsilon}, \qquad F''(\alpha) = -\frac{\varepsilon}{(1-\alpha)(1-\varepsilon(1-\alpha))}$$

so $F'(0) = 0$ and, provided $\alpha < 1$, for some $\theta \in [0, 1]$,

$$F(\alpha) = (\alpha^2/2) \cdot F''(\theta\alpha) = -\alpha^2 \varepsilon \Theta(1).$$

Let $C_\alpha = 1/\left(2\pi d\varepsilon(1-\alpha)(1-\varepsilon(1-\alpha))\right)^{1/2}$. When $t = ndp \log 2p/(2p-1)$ the value of $e^{-d\varepsilon(1-\alpha)t/ndp}$ is $(\varepsilon/(1+\varepsilon))^{d\varepsilon(1-\alpha)}$. Thus the expected number of vertices of degree $d\varepsilon(1-\alpha)$ still unvisited at time t (see (12)) is asymptotic to

$$n\, C_\alpha\, N(d\varepsilon(1-\alpha))e^{-d\varepsilon(1-\alpha)t/ndp} = C_\alpha\, (G_\varepsilon(\alpha))^d = C_\alpha\, e^{-\alpha^2 \varepsilon d \Theta(1)}.$$

Arguing as in Section 3.3 we see that vertices of degree $\sim d\varepsilon$ should be last to be visited.

4.4 Concentration of the number of vertices of a given degree

Recall that $X(i) = X_p(i)$ is the number of vertices of degree i in $Q_{n,p}$. The expected value of $X(i)$ is given in (8).

Variance of $X(i)$. Let $X = X(i)$ then

$$\begin{aligned}\mathbb{E}X(X-1) = n(n-d) &\left(\binom{d}{i}p^i q^{d-i}\right)^2 \\ &+\ n\binom{d}{i}p^i q^{d-i}\left\{ i\binom{d-1}{i-1}p^{i-1}q^{d-i} + (d-i)\binom{d-1}{i}p^i q^{d-1-i}\right\}.\end{aligned}$$

The first term is for vertices v at distance at least two from vertex u in Q_n. The second term is for those vertices at distance one from vertex u in Q_n which are a neighbour of u in $Q_{n,p}$, and those v which are not, respectively.

Thus with $\eta = \binom{d}{i}p^i q^{d-i}$,

$$\begin{aligned}\mathbb{E}X(X-1) &= n^2\eta^2 - nd\eta^2 + n\eta^2\left(\frac{i^2}{dp} + \frac{(d-i)^2}{dq}\right) \\ &= n^2\eta^2 + n\eta^2 \frac{(pd-i)^2}{dpq},\end{aligned}$$

and

$$\begin{aligned}\mathbb{V}X =& \mathbb{E}X(X-1) + \mathbb{E}X - (\mathbb{E}X)^2 \\ =& n\eta + n\eta^2 \frac{(pd-i)^2}{dpq} \\ =& \mathbb{E}X\left(1 + \mathbb{E}X \frac{1}{n}\frac{(dp-i)^2}{dpq}\right).\end{aligned}$$

Concentration of vertices of degree $d\varepsilon$. The expected value of $X(d\varepsilon)$ is given by (19). From the above, as $dp - d\varepsilon = dq$

$$\mathbb{V}X(d\varepsilon) = \mathbb{E}X(d\varepsilon)\left(1 + \mathbb{E}X(d\varepsilon)\frac{1}{n}\frac{dq}{p}\right) = (1+o(1))\mathbb{E}X(d\varepsilon).$$

The probability that $X(d\varepsilon)$ deviates significantly above $\mathbb{E}X(d\varepsilon)$ is therefore

$$\mathbb{P}(X(d\varepsilon) \geq \mathbb{E}X(d\varepsilon) + \sqrt{\omega\mathbb{E}X(d\varepsilon)}) \leq \frac{1+o(1)}{\omega}.$$