Discrete Mathematics 309 (2009) 3113–3124

Contents lists available at ScienceDirect

Discrete Mathematics

journal homepage: www.elsevier.com/locate/disc

Large components in random induced subgraphs of n-cubes

Christian M. Reidys *

Center for Combinatorics, LPMC-TJKLC, Nankai University, Tianjin 300071, PR China

ARTICLE INFO

Article history:
Received 31 July 2007
Received in revised form 6 August 2008
Accepted 18 August 2008
Available online 20 November 2008

Keywords:
Random graph
n-cube
Giant component
Vertex boundary

ABSTRACT

In this paper we study random induced subgraphs of the binary n-cube, Q_2^n. This random graph is obtained by selecting each Q_2^n-vertex with independent probability λ_n. Using a novel construction of subcomponents we study the largest component for $\lambda_n = \frac{1+\chi_n}{n}$, where $\epsilon \geq \chi_n \geq n^{-\frac{1}{3}+\delta}$, $\delta > 0$. We prove that there exists a.s. a unique largest component $C_n^{(1)}$. We furthermore show that for $\chi_n = \epsilon$, we have $|C_n^{(1)}| \sim \alpha(\epsilon)\frac{1+\chi_n}{n}2^n$ and for $o(1) = \chi_n \geq n^{-\frac{1}{3}+\delta}$, $|C_n^{(1)}| \sim 2\,\chi_n \frac{1+\chi_n}{n} 2^n$ holds. This improves the result of [B. Bollobás, Y. Kohayakawa, T. Luczak, On the evolution of random boolean functions, Extremal Problems Finite Sets (1991) 137–156] where constant $\chi_n = \chi$ is considered. In particular, in the case of $\lambda_n = \frac{1+\epsilon}{n}$, our analysis implies that a.s. a unique giant component exists.

1. Introduction and statement of results

1.1. Background

Burtin was the first [9] to study the connectedness of random subgraphs of n-cubes, Q_2^n, obtained by selecting all Q_2^n-edges independently (with probability p_n). He proved that a.s. all such subgraphs are connected for $p > 1/2$ and are disconnected for $p < 1/2$. Erdős and Spencer [11] refined Burtin's result and, more importantly in our context, they conjectured that there exists a.s. a giant component for $p_n = \frac{1+\epsilon}{n}$ and $\epsilon > 0$. Their conjecture was proved by Ajtai, Komlós and Szemerédi [2] who established the existence of a unique giant component for $p_n = \frac{1+\epsilon}{n}$. Key ingredients in their proof are Harper's isoperimetric inequality [13] and a two round randomization, used for showing the nonexistence of certain splits. Several variations including the analysis of the giant component in random graphs with given average degree sequence have been studied [1,16,17]. Bollobás, Kohayakawa and Luczak [7] analyzed the behavior for ϵ tending to 0 and showed in particular that the constant for the giant component for fixed $\epsilon > 0$ coincides with the probability of infinite survival of the associated Poisson branching process. Spencer et al. [8] refined their results, using specific properties of the n-cube as for instance the isoperimetric inequality [13] and Ajtai et al.'s two round randomization idea. Considerably less is known for random induced subgraphs of the n-cube obtained by independently selecting each Q_2^n-vertex with probability λ_n. The main result here is the paper of Bollobás et al. who have shown in [6] for constant χ that $C_n^{(1)} = (1+o(1))\kappa\,\chi\,\frac{1+\chi}{n}2^n$. In this paper we improve this result. We show that for $\chi_n \geq n^{-\frac{1}{3}+\delta}$, where $\delta > 0$, a unique largest component exists and determine its size. The key idea is a novel construction for small subcomponents given in Lemma 2.

Random induced subgraphs arise in the context of molecular folding maps [21], where the neutral networks of molecular structures are modeled as random induced subgraphs of n-cubes [18]. They also occur in the context of neutral evolution of populations (i.e. families of Q_2^n-vertices) consisting of erroneously replicating bit strings. Here, we work of course in Q_4^n,

* Tel.: +86 22 2350 5133x6800; fax: +86 22 2350 9272.
E-mail address: reidys@nankai.edu.cn.

0012-365X/$ – see front matter © 2009 Published by Elsevier B.V.
doi:10.1016/j.disc.2008.08.015

since we have the alphabet $\{\mathbf{A}, \mathbf{U}, \mathbf{G}, \mathbf{C}\}$. Random induced subgraphs of n-cubes have had impact on conceptual level [20] and led to experimental work identifying sequences that realize two distinct ribozymes [19]. A systematic computational analysis of neutral networks of molecular folding maps can be found in [12].

The main result of this paper is the following:

Theorem. *Let Q^n_{2,λ_n} be the random graph consisting of Q^n_2-subgraphs, Γ_n, induced by selecting Q^n_2-vertices with independent probability $\lambda_n = \frac{1+\chi_n}{n}$, where $\epsilon \geq \chi_n \geq n^{-\frac{1}{3}+\delta}$ and $\epsilon, \delta > 0$. Then we have for $\chi_n = \epsilon$*

$$\lim_{n\to\infty} \mathbb{P}\left(|C_n^{(1)}| \sim \alpha(\epsilon)\frac{1+\epsilon}{n}2^n \text{ and } C_n^{(1)} \text{ is unique}\right) = 1 \tag{1.1}$$

and for $o(1) = \chi_n \geq n^{-\frac{1}{3}+\delta}$

$$\lim_{n\to\infty} \mathbb{P}\left(|C_n^{(1)}| \sim 2\,\chi_n\frac{1+\chi_n}{n}2^n \text{ and } C_n^{(1)} \text{ is unique}\right) = 1. \tag{1.2}$$

For $\chi_n = \epsilon$ the above theorem (combined with a straightforward argument for $\lambda_n \leq \frac{1-\epsilon}{n}$) implies

$$\lim_{n\to\infty} \mathbb{P}(\Gamma_n \text{ has a unique giant component}) = \begin{cases} 1 & \text{for } \lambda_n \geq \dfrac{1+\epsilon}{n} \\ 0 & \text{for } \lambda_n \leq \dfrac{1-\epsilon}{n}. \end{cases} \tag{1.3}$$

This is the random induced subgraph analogue of Ajtai et al.'s [2] result. We present in Lemma 2 a novel construction of subcomponents using branching processes inductively. We prove the main result using a generic vertex-boundary result due to Aldous [3,4]. All results proved in this paper can straightforwardly be generalized to n-cubes over arbitrary, finite alphabets.

1.2. Notation and terminology

The binary n-cube, Q^n_2, is a combinatorial graph with vertex set $\mathbb{F}^n_2$ in which two vertices are adjacent if they differ in exactly one coordinate. Let $d(v, v')$ be the number of coordinates by which v and v' differ. We set

$$\forall A \subset \mathbb{F}^n_2,\ j \leq n; \quad \mathsf{B}(A, j) = \{v \in \mathbb{F}^n_2 \mid \exists\, \alpha \in A;\ d(v, \alpha) \leq j\} \tag{1.4}$$

$$\forall A \subset \mathbb{F}^n_2,\ j \leq n; \quad \mathsf{S}(A, j) = \{v \in \mathbb{F}^n_2 \mid \exists\, \alpha \in A;\ d(v, \alpha) = j\} \tag{1.5}$$

$$\forall A \subset \mathbb{F}^n_2; \quad \mathsf{d}(A) = \{v \in \mathbb{F}^n_2 \mid \exists\, \alpha \in A;\ d(v, \alpha) = 1\} \tag{1.6}$$

and call $\mathsf{B}(A, j)$ and $\mathsf{d}(A)$ the ball of radius j around A and the vertex boundary of A in Q^n_2, respectively. If $A = \{\alpha\}$ we simply write $\mathsf{B}(\alpha, j)$. Let $D, E \subset \mathbb{F}^n_2$, we call $D\ell$-dense in E if $\mathsf{B}(v, \ell) \cap D \neq \varnothing$ for any $v \in E$. Q^n_2 can also be viewed as the Cayley graph $\mathsf{Cay}(\mathbb{F}^n_2, \{e_i \mid i = 1, \ldots, n\})$ where e_i is the canonical base vector. We will view $\mathbb{F}^n_2$ as an $\mathbb{F}_2$-vectorspace and denote the linear hull over $\{v_1, \ldots, v_h\}$, $v_j \in \mathbb{F}^n_2$ by $\langle v_1, v_2, \ldots, v_h\rangle$. Furthermore there exists a natural linear order "$\leq$" over Q^n_2 given by

$$v \leq v' \iff (d(v, 0) < d(v', 0)) \vee (d(v, 0) = d(v', 0) \wedge v <_{\text{lex}} v'), \tag{1.7}$$

where $<_{\text{lex}}$ denotes the lexicographical order. Any notion of minimal or smallest element in a subset $A \subset Q^n_2$ is refers to the linear order $\leq$ of Eq. (1.7).

Each $A \subset \mathbb{F}^n_2$ induces a unique induced subgraph in Q^n_2, denoted by $Q^n_2[A]$, in which $a_1, a_2 \in A$ are adjacent iff a_1, a_2 are adjacent in Q^n_2. Let Q^n_{2,λ_n} be the random graph consisting of Q^n_2-subgraphs, Γ_n, induced by selecting each Q^n_2-vertex with independent probability λ_n. That is, Q^n_{2,λ_n} is the finite probability space $(\{Q^n_2[A] \mid A \subset \mathbb{F}^n_2\}, \mathbb{P})$, with the probability measure

$$\mathbb{P}(A) = \lambda_n^{|A|}(1-\lambda_n)^{2^n-|A|}. \tag{1.8}$$

A property M is a subset of induced subgraphs of Q^n_2 closed under graph isomorphisms. The terminology "M holds a.s." is equivalent to $\lim_{n\to\infty} \mathsf{Prob}(\mathsf{M}) = 1$. A component of Γ_n is a maximal, connected, induced Γ_n-subgraph, C_n. The largest Γ_n-component is denoted by $C_n^{(1)}$. It is called a giant component if and only if

$$\exists\, \kappa > 0, \quad |C_n^{(1)}| \geq \kappa|\Gamma_n|, \tag{1.9}$$

and $x_n \sim y_n$ is equivalent to (a) $\lim_{n\to\infty} x_n/y_n$ exists and (b) $\lim_{n\to\infty} x_n/y_n = 1$. Let $Z_n = \sum_{i=1}^n \xi_i$ be a sum of mutually independent indicator random variables (r.v.), ξ_i having values in $\{0, 1\}$. Then we have, [10], for $\eta > 0$ and $c_\eta = \min\{-\ln(e^\eta[1+\eta]^{-[1+\eta]}), \frac{\eta^2}{2}\}$

$$\mathsf{Prob}(|Z_n - \mathbb{E}[Z_n]| > \eta\,\mathbb{E}[Z_n]) \leq 2e^{-c_\eta \mathbb{E}[Z_n]}. \tag{1.10}$$

n is always assumed to be sufficiently large and ϵ is a positive constant satisfying $0 < \epsilon < \frac{1}{3}$. We use the notation $B_m(\ell,\lambda_n) = \binom{m}{\ell}\lambda_n^{\ell}(1-\lambda_n)^{m-\ell}$ and write $g(n) = O(f(n))$ and $g(n) = o(f(n))$ for $g(n)/f(n) \to \kappa$ as $n \to \infty$ and $g(n)/f(n) \to 0$ as $n \to \infty$, respectively.

2. Preliminaries

Let us briefly recall some basic facts about branching processes [14,15]. Suppose ξ is a r.v. and $(\xi_i^{(t)})$, $i, t \in \mathbb{N}$ counts the number of offspring of the ith-individual of generation $t-1$. We consider the family of r.v. $(Z_i)_{i\in\mathbb{N}_0}$: $Z_0 = 1$ and $Z_t = \sum_{i=1}^{Z_{t-1}} \xi_i^{(t)}$ for $t \geq 1$ and interpret Z_t as the number of individuals "alive" in generation t. We shall be interested in the limit probability $\lim_{t\to\infty} \mathrm{Prob}(Z_t > 0)$, i.e. the probability of infinite survival.

Theorem 1. *Let $u_n = n^{-\frac{1}{3}}$, $\lambda_n = \frac{1+\chi_n}{n}$, $m = n - \lfloor \frac{3}{4}u_n n\rfloor$ and $\mathrm{Prob}(\xi = \ell) = B_m(\ell,\lambda_n)$. Then for $\chi_n = \epsilon$ the r.v. ξ becomes asymptotically Poisson, i.e. $\mathrm{Prob}(\xi = \ell) \sim \frac{(1+\epsilon)^{\ell}}{\ell!}\,\mathrm{e}^{-(1+\epsilon)}$ and*

$$0 < \lim_{t\to\infty} \mathrm{Prob}(Z_t > 0) = \alpha(\epsilon) < 1. \tag{2.1}$$

For $o(1) = \chi_n \geq n^{-\frac{1}{3}+\delta}$, $\delta > 0$ we have

$$\lim_{t\to\infty} \mathrm{Prob}(Z_t > 0) = (2+o(1))\chi_n. \tag{2.2}$$

In view of Theorem 1, we introduce the notation

$$\pi(\chi_n) = \begin{cases} \alpha(\epsilon) & \text{for } \chi_n = \epsilon \\ 2(1+o(1))\chi_n & \text{for } o(1) = \chi_n \geq n^{-\frac{1}{3}+\delta}. \end{cases} \tag{2.3}$$

We proceed by labeling the indices of a Q_2^n-vertex $v = (x_1,\ldots,x_n)$. For this purpose set

$$\nu_n = \left\lfloor \frac{1}{2k(k+1)} u_n n \right\rfloor, \qquad \iota_n = \left\lfloor \frac{k}{2k+1} u_n n \right\rfloor, \quad \text{and} \quad z_n = k\nu_n + \iota_n. \tag{2.4}$$

We write a Q_2^n-vertex $v = (x_1,\ldots,x_n)$ as

$$(\underbrace{x_1^{(1)},\ldots,x_{\nu_n}^{(1)}}_{\nu_n \text{ coordinates}}, \underbrace{x_1^{(2)},\ldots,x_{\nu_n}^{(2)}}_{\nu_n \text{ coordinates}},\ldots, \underbrace{x_1^{(k+1)},\ldots,x_{\iota_n}^{(k+1)}}_{\iota_n \text{coordinates}}, \underbrace{x_{z_n+1},\ldots,x_n}_{\substack{n-z_n \geq \\ n-\lfloor \frac{1}{2}u_n n\rfloor \text{ coordinates}}}). \tag{2.5}$$

For any $1 \leq s \leq \nu_n$, $r = 1,\ldots,k$, we set $e_s^{(r)}$ to be the $(s+(r-1)\nu_n)$th-unit vector, i.e. $e_s^{(r)}$ has exactly one 1 at its $(s+(r-1)\nu_n)$th coordinate. Similarly, let $e_s^{(k+1)}$, $1 \leq s \leq \iota_n$ denote the $(s+k\nu_n)$th-unit vector. We use the standard notation for the $z_n+1 \leq t \leq n$ unit vectors, i.e. e_t denotes the vector where $x_t = 1$ and $x_j = 0$, otherwise.

In our first lemma, we use Theorem 1 in order to obtain small tree-components in Γ_n. The main observation here is that, although easily larger subcomponents could be constructed, one is content with those of size $\lfloor \frac{1}{4}u_n n\rfloor$. The latter, however, can be constructed with probability at least $\pi(\chi_n)$. This fact will be crucial in the proof of Lemma 4, which eventually allows us to compute the size of the largest component.

Lemma 1. *Suppose $\lambda_n = \frac{1+\chi_n}{n}$ and $\epsilon \geq \chi_n \geq n^{-\frac{1}{3}+\delta}$, where $\delta > 0$. Then each Γ_n-vertex is contained in a Γ_n-subcomponent of size $\lfloor \frac{1}{4}u_n n\rfloor$ with probability at least $\pi(\chi_n)$.*

Proof. We consider the following branching process in the subcube $Q_2^{n-z_n}$, using the notation of Eq. (2.5). Without loss of generality we initialize the process at $v = (0,\ldots,0)$ (abusing notation we shall denote $(0,\ldots,0)$ by 0) and set $E_0 = \{e_{z_n+1},\ldots,e_n\}$ and $L_*[0] = \{(0,\ldots,0)\}$. We consider the $n - \lfloor \frac{3}{4}u_n n\rfloor$ smallest neighbors of v. Starting with the smallest, we select each of them with independent probability $\lambda_n = \frac{1+\chi_n}{n}$. Suppose $v + e_j$ is the first being selected. Then we set $E_1 = E_0 \setminus \{e_j\}$, $N_1[0] = \{v+e_j\}$ and proceed inductively setting $E_t = E_{t-1}\setminus\{e_w\}$ and $N_t[0] = N_{t-1}[0] \cup \{v+e_w\}$ for each neighbor $v+e_w$ being selected, subject to the condition $|E_t| > n - (\lfloor \frac{3}{4}u_n n\rfloor - 1)$. This procedure generates the set containing all selected 0-neighbors, which we denote by $N_*[0]$. We consider $L_*[1] = N_*[0] \cup L_*[0] \setminus \{0\}$. If $\varnothing \neq L_*[1]$ we proceed by choosing its smallest element, v_1^*. By construction, v_1^* has at least $n - \lfloor \frac{3}{4}u_n n\rfloor$ neighbors of the form $v_1^* + e_r$, where $e_r \in E_t$. We iterate the process selecting from the smallest $n - \lfloor \frac{3}{4}u_n n\rfloor$ neighbors of v_1^* and set $L_*[2] = (N_*[1]\cup L_*[1])\setminus\{v_1^*\}$. We then proceed inductively, setting $L_*[r+1] = (N_*[r]\cup L_*[r])\setminus\{v_r^*\}$. By construction, this process generates an induced subtree of $Q_2^{n-z_n}$. It stops in the case of $L_*[r] = \varnothing$ for some $r \geq 1$ or

$$|E_s| = n - \left(\left\lfloor \frac{3}{4}u_n n\right\rfloor - 1\right),$$

in which case $\lfloor \frac{1}{4} u_n n \rfloor - 1$ vertices have been connected. Theorem 1 guarantees that this $Q_2^{n-z_n}$-tree has size $\lfloor \frac{1}{4} u_n n \rfloor$ with probability at least $\pi(\chi_n)$. $\square$

We refer to the particular branching process used in Lemma 1 as γ-process. The γ-process produces a subcomponent of size $\lfloor \frac{1}{4} u_n n \rfloor$, which we refer to as γ-subcomponent.

3. Small subcomponents

The γ-process of Lemma 1 did by construction not involve the first z_n coordinates. In the following lemma we will use the first $k\, \nu_n$ of them in order to build inductively larger subcomponents.

Lemma 2. *Let $k \in \mathbb{N}$ be arbitrary but fixed, $\lambda_n = \frac{1+\chi_n}{n}$, $\nu_n = \lfloor \frac{u_n n}{2k(k+1)} \rfloor$ and $\varphi_n = \pi(\chi_n)\nu_n(1 - \mathrm{e}^{-(1+\chi_n)u_n/4})$. Then there exists $\rho_k > 0$ such that each Γ_n-vertex is with probability at least*

$$\pi_k(\chi_n) = \pi(\chi_n)\left(1 - \mathrm{e}^{-\rho_k \varphi_n}\right) \tag{3.1}$$

contained in a Γ_n-subcomponent of size at least $c_k(u_n n)\varphi_n^k$, where $c_k > 0$.

Proof. Since all translations are Q_2^n-automorphisms we can, without loss of generality assume that $v = (0, \ldots, 0)$ (abusing notation we shall denote $(0, \ldots, 0)$ by 0). Using the notation of Eq. (2.5) we recruit the $n - z_n$-unit vectors e_t for a γ-process. The γ-process of Lemma 1 yields a γ-subcomponent, $C_0^{(0)}$, of size $\lfloor \frac{1}{4} u_n n \rfloor$ with probability $\geq \pi(\chi_n)$. We consider for $1 \leq i \leq k$ the sets of ν_n elements $B_i = \{e_1^{(i)}, \ldots, e_{\nu_n}^{(i)}\}$ and set $H = \langle e_{z_n+1}, \ldots, e_n \rangle$. By construction we have

$$\left\langle B_i \cup \left\langle \bigcup_{1 \leq j \leq i-1} B_j \right\rangle \oplus H \right\rangle = \langle B_i \rangle \oplus \left\langle \bigcup_{1 \leq j \leq i-1} B_j \right\rangle \oplus H. \tag{3.2}$$

In particular, for any $1 \leq s \neq j \leq \nu_n$: $e_s^{(1)} - e_j^{(1)} \in H$ is equivalent to $e_s^{(1)} = e_j^{(1)}$. Since all vertices are selected independently and $|C_0^{(0)}| = \lfloor \frac{1}{4} u_n n \rfloor$, for fixed $e_s^{(1)} \in B_1$ the probability of not selecting a vertex $v' \in e_s^{(1)} + C_0^{(0)}$ is given by

$$\mathbb{P}\left(\left\{e_s^{(1)} + \xi \mid \xi \in C_0^{(0)}\right\} \cap \Gamma_n = \varnothing\right) = \left(1 - \frac{1 + \chi_n}{n}\right)^{\lfloor \frac{1}{4} u_n n \rfloor} \sim \mathrm{e}^{-(1+\chi_n)\frac{1}{4}u_n}. \tag{3.3}$$

We set $\mu_n = (1 - \mathrm{e}^{-(1+\chi_n)\frac{1}{4}u_n})$, i.e. $\mu_n = \mathbb{P}\left((e_s^{(1)} + C_0^{(0)}) \cap \Gamma_n \neq \varnothing\right)$ and introduce the r.v.

$$X_1 = \left|\left\{e_s^{(1)} \in B_1 \mid \exists\, \xi \in C_0^{(0)};\, e_s^{(1)} + \xi \in \Gamma_n\right\}\right|. \tag{3.4}$$

Obviously, $\mathbb{E}(X_1) = \mu_n \nu_n$ and using the large deviation result of Eq. (1.10) we can conclude that

$$\exists\, \rho > 0; \quad \mathbb{P}\left(X_1 < \frac{1}{2}\mu_n \nu_n\right) \leq \mathrm{e}^{-\rho\, \mu_n \nu_n}. \tag{3.5}$$

Suppose for $e_s^{(1)}$ there exists some $\xi \in C_0^{(0)}$ such that $e_s^{(1)} + \xi \in \Gamma_n$ (that is, $e_s^{(1)}$ is counted by X_1). We then select the smallest element of the set $\{e_s^{(1)} + \xi \mid \xi \in C_0^{(0)}, e_s^{(1)} + \xi \in \Gamma_n\}$, say $e_s^{(1)} + \xi_{0,e_s^{(1)}}$ and initiate a γ-process using the $n - z_n$ elements $\{e_{z_n+1}, \ldots, e_n\}$ at $e_s^{(1)} + \xi_{0,e_s^{(1)}}$. The process yields a γ-subcomponent, $C^{(1)}_{e_s^{(1)}+\xi_{0,e_s^{(1)}}}$ of size $\lfloor \frac{1}{4} u_n n \rfloor$ with probability at least $\pi(\chi_n)$. For any two elements $e_s^{(1)}, e_j^{(1)}$ with $e_s^{(1)} + \xi_{0,e_s^{(1)}}, e_j^{(1)} + \xi_{0,e_j^{(1)}} \in \Gamma_n$ the respective γ-subcomponent, $C^{(1)}_{e_s^{(1)}+\xi_{0,e_s^{(1)}}}$ and $C^{(1)}_{e_j^{(1)}+\xi_{0,e_j^{(1)}}}$ are vertex disjoint since $\langle B_1 \cup H \rangle = \langle B_1 \rangle \oplus H$. Let $\tilde{X}_1$ be the r.v. counting the number of these new, pairwise vertex disjoint sets of γ-subcomponents of size $\lfloor \frac{1}{4} u_n n \rfloor$. By construction each of them is connected to 0. We immediately observe $\mathbb{E}(\tilde{X}_1) \geq \pi(\chi_n)\mu_n \nu_n$ and set $\varphi_n = \pi(\chi_n)\mu_n \nu_n$. Using the large deviation result in Eq. (1.10) we derive

$$\exists\, \rho_1 > 0; \quad \mathbb{P}\left(\tilde{X}_1 < \frac{1}{2}\varphi_n\right) \leq \mathrm{e}^{-\rho_1 \varphi_n}. \tag{3.6}$$

We proceed by proving that for each $1 \leq i \leq k$ there exists a sequence of r.v.s $(\tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_i)$ where $\tilde{X}_j$ counts the number of pairwise disjoint sets of γ-subcomponents added at step j, where $1 \leq j \leq i$, such that:

(a) all sets, $C_\alpha^{(j)}$, $1 \le j \le i$, are pairwise vertex disjoint and have size $\lfloor \frac{1}{4} u_n n \rfloor$
(b) all $C_\alpha^{(j)}$ are connected to 0 and

$$\exists\, \rho_i > 0; \quad \mathbb{P}\left(\tilde{X}_i < \frac{1}{2^i}\varphi_n^i\right) \le \mathrm{e}^{-\rho_i \varphi_n}, \quad \text{where } \varphi_n = \pi(\chi_n)\mu_n \nu_n. \tag{3.7}$$

We prove the assertion by induction on i. Without loss of generality we may assume $i < k$. Indeed, in our construction of $\tilde{X}_1$, we established the induction basis. In order to define $\tilde{X}_{i+1}$ we use the set $B_{i+1} = \{e_1^{(i+1)}, \ldots, e_{\nu_n}^{(i+1)}\}$. For each $C_\alpha^{(i)}$ counted by $\tilde{X}_i$ (i.e. the subcomponents that were connected in step i) we form the set $e_s^{(i+1)} + C_\alpha^{(i)}$. By induction hypothesis two different $C_\alpha^{(i)}, C_{\alpha'}^{(i)}$, counted by $\tilde{X}_i$, are vertex disjoint and connected to 0. Since $\langle B_{i+1}\rangle \bigoplus \langle \bigcup_{1\le j\le i} B_j\rangle \bigoplus H$ are disjoint we can conclude

$$(s \ne s' \vee \alpha \ne \alpha') \quad \Longrightarrow \quad (e_s^{(i+1)} + C_\alpha^{(i)}) \cap (e_{s'}^{(i+1)} + C_{\alpha'}^{(i)}) = \varnothing.$$

Furthermore, the probability that we have for fixed $C_\alpha^{(i)}$: $(e_s^{(i+1)} + C_\alpha^{(i)}) \cap \Gamma_n = \varnothing$, for some $e_s^{(i+1)} \in B_{i+1}$, is exactly as in Eq. (3.3)

$$\mathbb{P}\left((e_s^{(i+1)} + C_\alpha^{(i)}) \cap \Gamma_n = \varnothing\right) = \left(1 - \frac{1+\chi_n}{n}\right)^{\lfloor \frac{1}{4} u_n n \rfloor} \sim \mathrm{e}^{-(1+\chi_n)\frac{1}{4}u_n}.$$

As it is the case for the induction basis, $\mu_n = (1 - \mathrm{e}^{-(1+\chi_n)\frac{1}{4}u_n})$ is the probability that $(e_s^{(i+1)} + C_\alpha^{(i)}) \cap \Gamma_n \ne \varnothing$. We proceed by defining the r.v.

$$X_{i+1} = \sum_{C_\alpha^{(i)}} \left| \left\{ e_s^{(i+1)} \in B_{i+1} \mid \exists\, \xi \in C_\alpha^{(i)};\, e_s^{(i+1)} + \xi \in \Gamma_n \right\} \right|. \tag{3.8}$$

The r.v. X_{i+1} counts the number of events where $(e_s^{(i+1)} + C_\alpha^{(i)}) \cap \Gamma_n \ne \varnothing$ for each $C_\alpha^{(i)}$, respectively. Equivalently, for fixed $C_\alpha^{(i)}$ and $e_s^{(i+1)} \in B_{i+1}$ let

$$e_s^{(i+1)} + \xi_{\alpha, e_s^{(i+1)}} = \min\left\{ e_s^{(i+1)} + \xi_\alpha \mid \xi_\alpha \in C_\alpha^{(i)},\ e_s^{(i+1)} + \xi_\alpha \in \Gamma_n \right\}.$$

Then X_{i+1} counts exactly the minimal elements

$$e_s^{(i+1)} + \xi_{\alpha, e_s^{(i+1)}},\, e_{s'}^{(i+1)} + \xi_{\alpha', e_{s'}^{(i+1)}}, \ldots$$

for all $C_\alpha^{(i)}, C_{\alpha'}^{(i)}, \ldots$ and any two can be used to construct pairwise vertex disjoint γ-subcomponents of size $\lfloor \frac{1}{4} u_n n \rfloor$. We next define $\tilde{X}_{i+1}$ to be the r.v. counting the number of events that the γ-process in H initiated at the $e_s^{(i+1)} + \xi_{\alpha, e_s^{(i+1)}} \in \Gamma_n$ yields a γ-subcomponent of size $\lfloor \frac{1}{4} u_n n \rfloor$. By construction each of these is connected to a unique $C_\alpha^{(i)}$. Since $\langle B_{i+1}\rangle \bigoplus \langle \bigcup_{1\le j\le i} B_j\rangle \bigoplus H$ all newly added sets are pairwise vertex disjoint to all previously added subcomponents. We derive

$$\begin{aligned}
\mathbb{P}\left(\tilde{X}_{i+1} < \frac{1}{2^{i+1}}\varphi_n^{i+1}\right) &\le \underbrace{\mathbb{P}\left(\tilde{X}_i < \frac{1}{2^i}\varphi_n^i\right)}_{\text{failure at step } i} + \underbrace{\mathbb{P}\left(\tilde{X}_{i+1} < \frac{1}{2^{i+1}}\varphi_n^{i+1} \wedge \tilde{X}_i \ge \frac{1}{2^i}\varphi_n^i\right)}_{\text{failure at step } i+1 \text{ conditional to } \tilde{X}_i \ge \frac{1}{2^i}\varphi_n^i} \\
&\le \mathrm{e}^{-\rho_i \varphi_n} + \mathrm{e}^{-\rho\, \varphi_n^{i+1}}(1 - \mathrm{e}^{-\rho_i \varphi_n}), \quad \rho > 0 \\
&\le \mathrm{e}^{-\rho_{i+1}\varphi_n}.
\end{aligned}$$

Therefore each Γ_n-vertex is, with probability at least $\pi(\chi_n)(1 - \mathrm{e}^{-\rho_k \varphi_n})$, contained in a subcomponent of size at least $c_k(u_n n)\varphi_n^k$, for $c_k > 0$ and the proof of the lemma is complete. $\square$

Lemma 2 gives rise to introduce the induced subgraph $\Gamma_{n,k} = Q_2^n[A]$, where

$$A = \{v \mid v \text{ is contained in a } \Gamma_n\text{-subcomponent of size } \ge c_k(u_n n)\varphi_n^k, c_k > 0\}. \tag{3.9}$$

In the case of $\epsilon \ge \chi_n \ge n^{-\frac{1}{3}+\delta}$ we have $1 - \mathrm{e}^{-\frac{1}{4}(1+\chi_n)u_n} \ge u_n/4$ and consequently

$$\varphi_n \ge c'(1 + o(1))\chi_n u_n^2 n \ge c_0 n^\delta$$

for some $c', c_0 > 0$. Furthermore

$$\left\lfloor \frac{1}{4} u_n n \right\rfloor \varphi_n^k \ge c_k n^{\frac{2}{3}} n^{k\delta}, \quad c_k > 0. \tag{3.10}$$

Accordingly, choosing k sufficiently large, each Γ_n-vertex is contained in a subcomponent of arbitrary polynomial size with probability at least

$$\pi(\chi_n)\left(1-\mathrm{e}^{-\rho_k n^{\delta}}\right), \quad 0<\delta, 0<\rho_k. \tag{3.11}$$

We next prove a technical lemma which will be instrumental for the proof of Lemma 4. We show that the number of vertices not contained in $\Gamma_{n,k}$ is sharply concentrated, using a strategy similar to that in Bollobás et al. [7]. Let U_n denote the complement of $\Gamma_{n,k}$ in Γ_n.

Lemma 3. *Let $k\in\mathbb{N}$ and $\lambda_n=\frac{1+\chi_n}{n}$, where $\epsilon\geq\chi_n\geq n^{-\frac{1}{3}+\delta}$. Then we have*

$$\mathbb{P}\left(||U_n|-\mathbb{E}[|U_n|]|\geq\frac{1}{n}\mathbb{E}[|U_n|]\right)=o(1). \tag{3.12}$$

Proof. Let C be a Q_2^n-component of size strictly smaller than $\tau=c_k(u_n n)\varphi_n^k$ and let v be a fixed C-vertex. We shall denote the ordered pair (C,v) by C_v and the indicator variable of the pair C_v by X_{C_v}. Clearly, we have

$$|U_n|=\sum_{C_v}X_{C_v},$$

where the summation is taken over all ordered pairs (C,v) with $|C|<\tau$. Considering isolated points, we immediately obtain $\mathbb{E}[U_n]\geq c|\Gamma_n|$ for some $1\geq c>0$.

Claim. The random variable $|U_n|$ is sharply concentrated.

We prove the claim by estimating $\mathbb{V}[|U|]$ via computing the correlation terms $\mathbb{E}[X_{C_v}X_{D_{v'}}]$ and applying Chebyshev's inequality. Suppose $C_v\neq D_{v'}$. There are two ways by which $X_{C_v},X_{D_{v'}}$ viewed as r.v. over Q^n_{2,λ_n}, can be correlated. First v,v' can belong to the same component, i.e. $C=D$, in which case we write $C_v\sim_1 D_{v'}$. Clearly,

$$\sum_{C_v\sim_1 D_{v'}}\mathbb{E}[X_{C_v}X_{D_{v'}}]\leq\tau\mathbb{E}[|U_n|]. \tag{3.13}$$

Second, correlation arises when v,v' belong to two different components $C_v,D_{v'}$ having minimal distance 2 in Q_2^n. In this case we write $C_v\sim_2 D_{v'}$. Then there exists some Q_2^n-vertex, w, such that $w\in\mathrm{d}(C_v)\cap\mathrm{d}(D_{v'})$ and we derive

$$\begin{aligned}\mathbb{P}(d(C_v,D_{v'})=2)&=\frac{1-\lambda_n}{\lambda_n}\,\mathbb{P}(C_v\cup D_{v'}\cup\{w\}\text{ is a }\Gamma_n\text{-component})\\&\leq n\,\mathbb{P}(C_v\cup D_{v'}\cup\{w\}\text{ is a }\Gamma_n\text{-component}).\end{aligned}$$

We can now immediately give the upper bound

$$\sum_{C_v\sim_2 D_{v'}}\mathbb{E}[X_{C_v}X_{D_{v'}}]\leq n(2\tau+1)^3|\Gamma_n|. \tag{3.14}$$

The uncorrelated pairs $(X_{C_v},X_{D_{v'}})$, writing $C_v\not\sim D_{v'}$, can easily be estimated by

$$\sum_{C_v\not\sim D_{v'}}\mathbb{E}[X_{C_v}X_{D_{v'}}]=\sum_{C_v\not\sim D_{v'}}\mathbb{E}[X_{C_v}]\mathbb{E}[X_{D_{v'}}]\leq\mathbb{E}[|U_n|]^2. \tag{3.15}$$

Consequently we arrive at

$$\begin{aligned}\mathbb{E}[|U_n|(|U_n|-1)]&=\sum_{C_v\sim_1 D_{v'}}\mathbb{E}[X_{C_v}X_{D_{v'}}]+\sum_{C_v\sim_2 D_{v'}}\mathbb{E}[X_{C_v}X_{D_{v'}}]+\sum_{C_v\not\sim D_{v'}}\mathbb{E}[X_{C_v}X_{D_{v'}}]\\&\leq\tau\mathbb{E}[|U_n|]+n(2\tau+1)^3|\Gamma_n|+\mathbb{E}[|U_n|]^2.\end{aligned}$$

Using $\mathbb{V}[|U_n|]=\mathbb{E}[|U_n|(|U_n|-1)]+\mathbb{E}[|U_n|]-\mathbb{E}[|U_n|]^2$ and $\mathbb{E}[U_n]\geq c|\Gamma_n|$ we obtain

$$\frac{\mathbb{V}[|U_n|]}{\mathbb{E}[|U_n|]^2}\leq\frac{c_k(u_n n)\varphi_n^k+\frac{1}{c}n(2c_k(u_n n)\varphi_n^k+1)^3+1}{|\mathbb{E}[U_n]|}=o\left(\frac{1}{n^2}\right).$$

Chebyshev's inequality guarantees $\mathbb{P}(||U_n|-\mathbb{E}[|U_n|]|\geq\frac{1}{n}\mathbb{E}[|U_n|])\leq n^2\frac{\mathbb{V}[|U_n|]}{\mathbb{E}[|U_n|]^2}$, whence the claim and the lemma follows. $\square$

Lemmas 2 and 3 indicate that there are many Γ_n-vertices that are contained in components of at least arbitrary polynomial size. We proceed by studying Γ_n-vertices contained in components of size $<c_k\,u_n n\varphi_n^k$. For this purpose, we

use a strategy introduced by Bollobás et al. [7] and consider the n-regular rooted tree T_n. Let v^* denote the root of T_n. Then v^* has n descendants and all other T_n-vertices have $n-1$. Selecting the T_n-vertices with independent probability λ_n we obtain the probability space T_{n,λ_n}, whose elements, A_n, are random induced subtrees. We shall be interested in the A_n-component which contains the root, denoted by C_{v^*}. Let ξ_{v^*} and ξ_v, for $v \neq v^*$ be two r.v. such that $\mathrm{Prob}(\xi_{v^*} = \ell) = B_n(\ell, \lambda_n)$ and $\mathrm{Prob}(\xi_v = \ell) = B_{n-1}(\ell, \lambda_n)$, respectively. We assume that ξ_{v^*} and ξ_v count the offspring produced at v^* and $v \neq v^*$. Then the induced branching process initialized at v^*, $(Z_i)_{i\in\mathbb{N}_0}$ constructs C_{v^*}. Let $\pi_0(\chi)$ denote its survival probability, then we have in view of Theorem 1 and [7], Corollary 6:

$$\pi_0(\chi_n) = (1 + o(1))\pi(\chi_n). \tag{3.16}$$

Lemma 4. *Let* $\lambda_n = \frac{1+\chi_n}{n}$ *where* $\epsilon \geq \chi_n \geq n^{-\frac{1}{3}+\delta}$. *Then we have for sufficiently large* $k \in \mathbb{N}$

$$(1 - o(1))\pi(\chi_n)|\Gamma_n| \leq |\Gamma_{n,k}| \leq (1 + o(1))\pi(\chi_n)|\Gamma_n| \quad \textit{a.s.} \tag{3.17}$$

Proof. *Claim* 1. $|\Gamma_{n,k}| \geq ((1 - o(1))\pi(\chi_n))\,|\Gamma_n|$ a.s.

According to Lemma 2, we have $\mathbb{E}[|U_n|] < (1 - \pi_k(\chi_n))|\Gamma_n|$ and we can conclude using Lemma 3 and $\mathbb{E}[|U_n|] = O(|\Gamma_n|)$

$$|U_n| < \left(1 + O\left(\frac{1}{n}\right)\right)\mathbb{E}[|U_n|] < \left(1 - \left(\pi_k(\chi_n) - O\left(\frac{1}{n}\right)\right)\right)|\Gamma_n| \quad \text{a.s.} \tag{3.18}$$

In view of Eq. (3.1) and $\chi_n \geq n^{-\frac{1}{3}+\delta}$, we have for arbitrary, fixed k,

$$\pi_k(\chi_n) - O\left(\frac{1}{n}\right) = (1 - o(1))\pi(\chi_n).$$

Therefore we derive

$$|\Gamma_{n,k}| \geq (1 - o(1))\pi(\chi_n)|\Gamma_n| \quad \text{a.s.,} \tag{3.19}$$

and Claim 1 follows.

Claim 2. For sufficiently large k, $|\Gamma_{n,k}| \leq ((1 + o(1))\pi(\chi_n))\,|\Gamma_n|$ a.s. holds.

We first observe that for any fixed Q_2^n-vertex, v, we have the inequality

$$\mathbb{P}\left(|C_{v^*}| \leq \ell\right) \leq \mathbb{P}\left(|C_v| \leq \ell\right). \tag{3.20}$$

Indeed, we can obtain C_v by inductively constructing a spanning tree as follows: suppose the set of all C_v-vertices at distance h is $M_h^{C_v}$. Starting with the smallest $w \in M_h^{C_v}$ $(h \geq 1)$ there are at most $(n - 1)$ w-neighbors contained in $M_{h+1}^{C_v}$ that are not neighbors for some smaller $w' \in M_h^{C_v}$. Hence for any $w \in M_h^{C_v}$ at most $n - 1$ vertices have to be examined. The A_n-component C_{v^*} is generated by the same procedure. Then for each $w \in M_h^{C_{v^*}}$ there are exactly $(n - 1)$ neighbors in $M_{h+1}^{C_{v^*}}$. Since the process adds at each stage less or equally many vertices to C_v, we have by construction $|C_v| \leq |C_{v^*}|$. Standard estimates for binomial coefficients allow us to estimate the numbers of T_n-subtrees containing the root [7], Corollary 3. Since vertex boundaries in T_n are easily obtained we can accordingly compute $\mathbb{P}(|C_{v^*}| = \ell)$. Choosing k sufficiently large, the estimates in [7], Lemma 22, guarantee

$$\mathbb{P}\left(|C_{v^*}| < c_k\, u_n n \varphi_n^k\right) = (1 - \pi_0(\chi_n)) + o(\mathrm{e}^{-n}). \tag{3.21}$$

In view of $\mathbb{P}\left(|C_{v^*}| \leq \ell\right) \leq \mathbb{P}\left(|C_v| \leq \ell\right)$ and Eq. (3.16) we can conclude from Eq. (3.21)

$$(1 - (1 + o(1))\pi(\chi_n))|\Gamma_n| + o(1) \leq \mathbb{E}[|U_n|]. \tag{3.22}$$

According to Lemma 3 we have $(1 - O(\frac{1}{n}))\mathbb{E}[|U_n|] < |U_n|$ a.s. and therefore

$$\left(1 - \left(1 + o(1) + O\left(\frac{1}{n}\right)\right)\pi(\chi_n)\right)|\Gamma_n| \leq |U_n| \quad \text{a.s.} \tag{3.23}$$

Combining Eqs. (3.19) and (3.23) we derive

$$(1 - o(1))\pi(\chi_n)|\Gamma_n| \leq |\Gamma_{n,k}| \leq (1 + o(1))\pi(\chi_n)|\Gamma_n| \quad \text{a.s.,} \tag{3.24}$$

whence the lemma. □

In the last lemma of this section we establish an additional property of $\Gamma_{n,k}$. We show that $\Gamma_{n,k}$ is a.s. 2-dense in Q_2^n with the exception of $2^n \mathrm{e}^{-\tilde{\Delta} n^{\delta}}$ vertices, where $\tilde{\Delta} > 0$. The 2-density of $\Gamma_{n,k}$ plays a key role in the proof of Lemma 7, where we establish the existence of many vertex disjoint, short paths between certain splits of the $\Gamma_{n,k}$. The result shows in addition that $\Gamma_{n,k}$ is uniformly distributed in Γ_n.

Lemma 5. *Let $k \in \mathbb{N}$, $\lambda_n = \frac{1+\chi_n}{n}$ and $\epsilon \geq \chi_n \geq n^{-\frac{1}{3}+\delta}$. Then we have*

$$\exists\, \Delta > 0;\ \forall v \in \mathbb{F}_2^n,\quad \mathbb{P}\left(|\mathsf{S}(v,2) \cap \Gamma_{n,k}| < \frac{1}{2}\left(\frac{k}{2(k+1)}\right)^2 n^{\delta}\right) \leq \mathrm{e}^{-\Delta n^{\delta}}. \tag{3.25}$$

Let $D_\delta = \left\{v \mid |\mathsf{S}(v,2) \cap \Gamma_{n,k}| < \frac{1}{2}\left(\frac{k}{2(k+1)}\right)^2 n^{\delta}\right\}$, then there exists some $\Delta > \tilde{\Delta} > 0$ such that

$$|D_\delta| \leq 2^n \mathrm{e}^{-\tilde{\Delta} n^{\delta}} \quad \textit{a.s.} \tag{3.26}$$

Proof. To prove the lemma, we use the last (see Eq. (2.5)) $\iota_n = \lfloor \frac{k}{2(k+1)} u_n n \rfloor$ elements $e_1^{(k+1)}, \ldots, e_{\iota_n}^{(k+1)}$. We consider for arbitrary $v \in Q_2^n$

$$\mathsf{S}^{(k+1)}(v,2) = \left\{v + e_i^{(k+1)} + e_j^{(k+1)} \mid 1 \leq i < j \leq \iota_n\right\}. \tag{3.27}$$

Clearly, $|\mathsf{S}^{(k+1)}(v,2)| = \binom{\iota_n}{2}$ holds. By construction, for any two $\mathsf{S}^{(k+1)}(v,2) \cap \Gamma_n$-vertices, the Γ_n-subcomponents of size $\geq c_k(u_n n)\varphi_n^k$ constructed via Lemma 2, are vertex disjoint. Furthermore each Γ_n-vertex belongs to $\Gamma_{n,k}$ with probability $\geq \pi_k(\chi_n)$. Let Z be the r.v. counting the number of vertices in $\mathsf{S}^{(k+1)}(v,2) \cap \Gamma_{n,k}$. Then we have

$$\mathbb{E}[Z] \sim \left(\frac{k}{2(k+1)}\right)^2 \frac{u_n^2}{2} n\pi(\chi_n).$$

Eq. (3.25) follows from Eq. (1.10), $u_n^2 n \chi_n \geq n^{\delta}$ and

$$\mathbb{P}(|\mathsf{S}(v,2) \cap \Gamma_{n,k}| < \eta) \leq \mathbb{P}(|\mathsf{S}^{(k+1)}(v,2) \cap \Gamma_{n,k}| < \eta).$$

Let $D_\delta = \left\{v \mid |\mathsf{S}(v,2) \cap \Gamma_{n,k}| < \frac{1}{2}\left(\frac{k}{2(k+1)}\right)^2 n^{\delta}\right\}$. By linearity of expectation $\mathbb{E}(|D_\delta|) \leq 2^n \mathrm{e}^{-\Delta n^{\delta}}$ holds and using Markov's inequality, $\mathbb{P}(X > t\mathbb{E}(X)) \leq 1/t$ for $t > 0$, we derive $|D_\delta| \leq 2^n \mathrm{e}^{-\tilde{\Delta} n^{\delta}}$ a.s. for any $0 < \tilde{\Delta} < \Delta$. □

4. Vertex boundary and paths

The following proposition is due to [5] used for Sidon sets in groups in the context of Cayley graphs. In the following, G denotes a finite group and M a finite set acted upon by G.

Proposition 1. *Suppose G acts transitively on M and let $A \subset M$, then we have*

$$\frac{1}{|G|} \sum_{g \in G} |A \cap gA| = |A|^2/|M|. \tag{4.1}$$

Proof. We prove Eq. (4.1) by induction on $|A|$. For $A = \{x\}$ we derive $\frac{1}{|G|} \sum_{gx=x} 1 = |G_x|/|G|$, since $|M| = |G|/|G_x|$. We next prove the induction step. We write $A = A_0 \cup \{x\}$ and compute

$$\begin{aligned}
\frac{1}{|G|} \sum_g |A \cap gA| &= \frac{1}{|G|} \sum_g (|A_0 \cap gA_0| + |\{gx\} \cap A_0| + |\{x\} \cap gA_0| + |\{gx\} \cap \{x\}|) \\
&= \frac{1}{|G|} (|A_0|^2 |G_x| + 2|A_0||G_x| + |G_x|) \\
&= \frac{1}{|G|} ((|A_0| + 1)^2 |G_x|) = \frac{|A|^2}{|M|}. \quad \square
\end{aligned}$$

Aldous [3,4] observed how to use Proposition 1 for deriving a very general lower bound for vertex boundaries in Cayley graphs:

Lemma 6. *Suppose G acts transitively on M and let $A \subset M$, and let S be a generating set of the Cayley graph $\mathsf{Cay}(G,S)$ where $|S| = n$. Then we have*

$$\exists\, s \in S;\quad |sA \setminus A| \geq \frac{1}{n}|A|\left(1 - \frac{|A|}{|M|}\right). \tag{4.2}$$

Proof. We compute

$$|A| = \frac{1}{|G|}\sum_g (|gA \setminus A| + |A \cap gA|) = \frac{1}{|G|}\sum_g |gA \setminus A| + |A|\frac{|A|}{|M|} \tag{4.3}$$

and hence $|A|(1 - \frac{|A|}{|M|}) = \frac{1}{|G|}\sum_g |gA \setminus A|$. From this we can immediately conclude

$$\exists\, g \in G; \quad |gA \setminus A| \geq |A|\left(1 - \frac{|A|}{|M|}\right).$$

Let $g = \prod_{j=1}^k s_j$. Since each element of $gA \setminus A$ is contained in at least one set $s_jA \setminus A$ we obtain

$$|gA \setminus A| \leq \sum_{j=1}^k |s_jA \setminus A|.$$

Hence there exists some $1 \leq j \leq k$ such that $|s_jA \setminus A| \geq \frac{1}{k}|gA \setminus A|$ and the lemma follows. □

The next lemma proves the existence of many vertex disjoint paths connecting the boundaries of certain splits of $\Gamma_{n,k}$. The lemma is related to a result in [8] but much stronger since the actual length of these paths is ≤ 3. The shortness of these paths results from the 2-density of $\Gamma_{n,k}$ (Lemma 5) and is a consequence of our particular construction of small subcomponents in Lemma 2.

Lemma 7. *Suppose* $\lambda_n = \frac{1+\chi_n}{n}$ *where* $\epsilon \geq \chi_n \geq n^{-\frac{1}{3}+\delta}$. *Let* (A, B) *be a split of* $\Gamma_{n,k}$ *with the properties*

$$\exists\, 0 < \sigma_0 \leq \sigma_1 < 1; \quad \frac{1}{n^2}2^n \leq |A| = \sigma_0|\Gamma_{n,k}| \quad \textit{and} \quad \frac{1}{n^2}2^n \leq |B| = \sigma_1|\Gamma_{n,k}|. \tag{4.4}$$

Then there exists some $t > 0$ *such that a.s.* $\mathsf{d}(A)$ *is connected to* $\mathsf{d}(B)$ *in* Q_2^n *via at least*

$$\frac{t}{n^4}2^n \Big/ \binom{n}{7} \tag{4.5}$$

vertex disjoint (independent) paths of length ≤ 3.

Proof. We consider $\mathsf{B}(A, 2)$ and distinguish the cases

$$|\mathsf{B}(A, 2)| \leq \frac{2}{3}2^n \quad \text{and} \quad |\mathsf{B}(A, 2)| > \frac{2}{3}2^n. \tag{4.6}$$

Suppose first $|\mathsf{B}(A, 2)| \leq \frac{2}{3}2^n$ holds. According to Lemma 6 and Eq. (4.4), we have

$$\exists\, d_1 > 0; \quad |\mathsf{d}(\mathsf{B}(A, 2))| \geq \frac{d_1}{n^3}2^n. \tag{4.7}$$

Lemma 5 guarantees that a.s. all except of at most $2^n e^{-\tilde{\Delta}n^\delta}$ Q_2^n-vertices are within distance 2 to some $\Gamma_{n,k}$-vertex. Hence there exist at least $\frac{d}{n^3}2^n$ vertices of $\mathsf{d}(\mathsf{B}(A, 2))$, that are contained in $\mathsf{B}(B, 2)$, i.e.

$$|\mathsf{d}\mathsf{B}(A, 2) \cap \mathsf{B}(B, 2)| \geq \frac{d}{n^3}2^n \quad \text{a.s.} \tag{4.8}$$

For each $\beta_2 \in \mathsf{d}(\mathsf{B}(A, 2)) \cap \mathsf{B}(B, 2)$ there exists a path $(\alpha_1, \alpha_2, \beta_2)$, starting in $\mathsf{d}(A)$ with terminus β_2. In view of $\mathsf{B}(B, 2) = \mathsf{d}(\mathsf{B}(B, 1))\dot{\cup}\mathsf{B}(B, 1)$, we distinguish the following cases

$$|\mathsf{d}(\mathsf{B}(A, 2)) \cap \mathsf{d}(\mathsf{B}(B, 1))| \geq \frac{1}{n^3}d_{2,1}2^n \quad \text{and} \quad |\mathsf{d}(\mathsf{B}(A, 2)) \cap \mathsf{B}(B, 1)| \geq \frac{1}{n^3}\, d_{2,2}2^n. \tag{4.9}$$

Suppose we have $|\mathsf{d}(\mathsf{B}(A, 2)) \cap \mathsf{d}(\mathsf{B}(B, 1))| \geq \frac{1}{n^3}d_{2,1}2^n$. For each $\beta_2 \in \mathsf{d}(\mathsf{B}(B, 1))$, we select some element $\beta_1(\beta_2) \in \mathsf{d}(B)$ and set $B^* \subset \mathsf{d}(B)$ to be the set of these endpoints. Clearly at most n elements in $\mathsf{B}(B, 2)$ can produce the same endpoint, whence

$$|B^*| \geq \frac{1}{n^4}d_{2,1}2^n.$$

Let $B_1 \subset B^*$ be maximal subject to the condition that for any pair of B_1-vertices (β_1, β_1') we have $d(\beta_1, \beta_1') > 6$. Then we have $|B_1| \geq |B^*|/\binom{n}{7}$ since $|\mathsf{B}(v, 7)| = \binom{n}{7}$. Any two of the paths from $\mathsf{d}(A)$ to $B_1 \subset \mathsf{d}(B)$ are of the form $(\alpha_1, \alpha_2, \beta_2, \beta_1)$ and vertex disjoint since each of them is contained in $\mathsf{B}(\beta_1, 3)$. Therefore there are a.s. at least

$$\frac{1}{n^4} d_{2,1} 2^n \Big/ \binom{n}{7} \tag{4.10}$$

vertex disjoint paths connecting $\mathsf{d}(A)$ and $\mathsf{d}(B)$. Suppose next $|\mathsf{d}(\mathsf{B}(A, 2)) \cap \mathsf{B}(B, 1)| \geq \frac{1}{n^3} d_{2,2} 2^n$. We conclude in complete analogy that there exist a.s. at least

$$\frac{1}{n^3} d_{2,2} 2^n \Big/ \binom{n}{5} \tag{4.11}$$

vertex disjoint paths of the form $(\alpha_1, \alpha_2, \beta_2)$ connecting $\mathsf{d}(A)$ and $\mathsf{d}(B)$. It remains to consider the case $|\mathsf{B}(A, 2)| > \frac{2}{3} 2^n$. By construction both A and B satisfy Eq. (4.4), respectively, whence we can without loss of generality assume that also $|\mathsf{B}(B, 2)| > \frac{2}{3} 2^n$ holds. In this case we have

$$|\mathsf{B}(A, 2) \cap \mathsf{B}(B, 2)| > \frac{1}{3} 2^n$$

and for each $\alpha_2 \in \mathsf{B}(A, 2) \cap \mathsf{B}(B, 2)$ we select $\alpha_1 \in \mathsf{d}(A)$ and $\beta_1 \in \mathsf{d}(B)$. We derive in analogy to the previous arguments that there exist a.s. at least

$$\frac{1}{n^2} d_2 2^n \Big/ \binom{n}{5} \tag{4.12}$$

pairwise vertex disjoint paths of the form $(\alpha_1, \alpha_2, \beta_1)$ and the proof of the lemma is complete. $\square$

5. The largest component

Theorem 2. *Let Q^n_{2,λ_n} be the random graph consisting of Q^n_2-subgraphs, Γ_n, induced by selecting each Q^n_2-vertex with independent probability λ_n. Suppose $\lambda_n = \frac{1+\chi_n}{n}$, where $\epsilon \geq \chi_n \geq n^{-\frac{1}{3}+\delta}$, $\delta > 0$. Then we have*

$$\lim_{n\to\infty} \mathbb{P}\left(|C_n^{(1)}| \sim \pi(\chi_n) \frac{1+\chi_n}{n} 2^n \text{ and } C_n^{(1)} \text{ is unique}\right) = 1. \tag{5.1}$$

Proof. *Claim.* We have $|C_n^{(1)}| \sim |\Gamma_{n,k}|$ a.s.

To prove the Claim we use an idea introduced by Ajtai et al. [2] and select Q^n_2-vertices in two rounds. First we select Q^n_2-vertices with independent probability $\frac{1+\chi_n/2}{n}$ and subsequently with $\frac{\chi_n}{2n}$. The probability for some vertex not to be chosen in both randomizations is

$$\left(1 - \frac{1+\chi_n/2}{n}\right)\left(1 - \frac{\chi_n/2}{n}\right) = 1 - \frac{1+\chi_n}{n} + \frac{(1+\chi_n/2)\chi_n/2}{n^2} \geq 1 - \frac{1+\chi_n}{n}.$$

Hence selecting first with probability $\frac{1+\chi_n/2}{n}$ (first round) and then with $\frac{\chi_n/2}{n}$ (second round) a vertex is selected with probability less than $\frac{1+\chi_n}{n}$ (all preceding lemmas hold for the first randomization $\frac{1+\chi_n/2}{n}$). We now select in our first round each Q^n_2-vertex with probability $\frac{1+\chi_n/2}{n}$. According to Lemma 4, we have

$$|\Gamma_{n,k}| \sim \pi(\chi_n)|\Gamma_n| \quad \text{a.s.} \tag{5.2}$$

Suppose $\Gamma_{n,k}$ contains a component, A, such that

$$\frac{1}{n^2} 2^n \leq |A| \leq (1-b)|\Gamma_{n,k}|, \quad b > 0.$$

Then there exists a split of $\Gamma_{n,k}$, (A, B), satisfying the assumptions of Lemma 7 (and $\mathsf{d}(A) \cap \mathsf{d}(B) = \varnothing$). We observe that Lemma 2 limits the number of ways these splits can be constructed. In view of

$$\left\lfloor \frac{1}{4} u_n n \right\rfloor \varphi_n^k \geq c_k n^{\frac{2}{3}} n^{k\delta}, \quad c_k > 0, \tag{5.3}$$

each A-vertex is contained in a component of size at least $c_k n^{\frac{2}{3}} n^{k\delta}$. Therefore there are at most

$$2^{\left(2^n/(c_k n^{\frac{2}{3}} n^{k\delta})\right)} \tag{5.4}$$

ways to choose A in such a split. According to Lemma 7 there exists $t > 0$ such that a.s. $\mathsf{d}(A)$ is connected to $\mathsf{d}(B)$ in Q^n_2 via at least $\frac{t}{n^4} 2^n / \binom{n}{7}$ vertex disjoint paths of length ≤ 3. We now select Q^n_2-vertices with probability $\frac{\chi_n/2}{n}$. None of the above

$\geq \frac{t}{n^4} 2^n / \binom{n}{7}$ paths can be selected during this process. Since any two paths are vertex disjoint the expected number of such splits is less than

$$2^{\left(2^n/(c_k n^{\frac{2}{3}} n^{k\delta})\right)} \left(1 - \left(\frac{\chi_n/2}{n}\right)^4\right)^{\frac{t}{n^4} 2^n/\binom{n}{7}} \sim 2^{\left(2^n/(c_k n^{\frac{2}{3}} n^{k\delta})\right)} e^{-\frac{t\chi_n^4}{2^4 n^8} 2^n/\binom{n}{7}}. \tag{5.5}$$

Hence choosing k sufficiently large, we can conclude that a.s. there cannot exist such a split. Therefore $|C_n^{(1)}| \sim |\Gamma_{n,k}|$, a.s. and the Claim is proved. According to Lemma 4 we consequently have

$$|C_n^{(1)}| \sim \pi(\chi_n)|\Gamma_n|. \tag{5.6}$$

In particular, for $\chi_n = \epsilon$, Theorem 1 ($0 < \alpha(\epsilon) < 1$) implies that there exists a giant Γ_n-component. It remains to prove that $C_n^{(1)}$ is unique. By construction any largest component, C'_n, is necessarily contained in $\Gamma_{n,k}$. In the proof of the Claim we have shown that a.s. there cannot exist a component C'_n in Γ_n with the property $|C'_n| \geq \frac{1}{n}|\Gamma_n|$. Therefore $C_n^{(1)}$ is unique and the proof of the theorem is complete. □

Theorem 3 is the analogue of Ajtai et al.'s result [2] (for random subgraphs of n-cubes obtained by selecting Q_2^n-edges independently).

Theorem 3. *Let Q^n_{2,λ_n} be the random graph consisting of Q_2^n-subgraphs, Γ_n, induced by selecting each Q_2^n-vertex with independent probability λ_n. Then*

$$\lim_{n\to\infty} \mathbb{P}(\Gamma_n \textit{ has a unique giant component}) = \begin{cases} 1 & \textit{for } \lambda_n \geq \dfrac{1+\epsilon}{n} \\ 0 & \textit{for } \lambda_n \leq \dfrac{1-\epsilon}{n}. \end{cases} \tag{5.7}$$

Proof. We proved the first assertion in Theorem 2. It remains to consider the case $\lambda_n = \frac{1-\epsilon}{n}$.

Claim. Suppose $\lambda_n = \frac{1-\epsilon}{n}$, then there exists $\kappa' > 0$ such that $|C_n^{(1)}| \leq \kappa' n$ holds.

The expected number of components of size ℓ is less than

$$\frac{1}{\ell} 2^n n^{\ell-1} \left(\frac{1-\epsilon}{n}\right)^{\ell} = \frac{1}{\ell n} 2^n (1-\epsilon)^{\ell} \tag{5.8}$$

since there are 2^n ways to choose the first element and at most n-vertices to choose from subsequently. This component is counted ℓ times corresponding to all ℓ choices for its "initial" vertex. Let $X_{\kappa' n}$ be the r.v. counting the number of components of size $\geq \kappa' n$. Choosing κ' sufficiently large, we can satisfy $(1-\epsilon)^{\kappa'} < 1/4$ and obtain

$$\mathbb{E}(X_{\kappa' n}) \leq \sum_{\ell \geq \kappa' n} \frac{1}{\ell n} 2^n (1-\epsilon)^{\ell} \leq \frac{1}{\kappa' n^2} 2^n (1-\epsilon)^{\kappa' n} \sum_{\ell \geq 0} (1-\epsilon)^{\ell} < \frac{1}{\kappa' n^2} 2^{-n} \frac{1}{1-(1-\epsilon)}. \tag{5.9}$$

This proves the Claim and the proof of the theorem is complete. □

Acknowledgments

We thank E.Y. Jin, J. Qin, R.R. Wang and L.C. Zuo for discussions. Special thanks to the referees whose feedback and suggestions have led to several improvements. Their help is greatly appreciated. This work was supported by the 973 Project, the PCSIRT Project of the Ministry of Education, the Ministry of Science and Technology, and the National Science Foundation of China.

References

[1] M. Aizenman, D.J Barsky, Sharpness of the phase transition in percolation models, Comm. Math. Phys. 108 (1987) 489–526.
[2] M. Ajtai, J. Komlós, E. Szemerédi, Largest random component of a k-cube, Combinatorica 2 (1982) 1–7.
[3] D. Aldous, P. Diaconis, Strong uniform times and finite random walks, Adv. Appl. Math. 2 (1987) 69–97.
[4] L. Babai, Local expansion of vertex transitive graphs and random generation in finite groups, in: Proc 23 ACM Symposium on Theory of Computing, vol. 1, ACM, New York, 1991, pp. 164–174.
[5] L. Babai, V.T. Sos, Sidon sets in groups and induced subgraphs of cayley graphs, European J. Combin. 1 (1985) 1–11.
[6] B. Bollobás, Y. Kohayakawa, T. Luczak, On the evolution of random boolean functions, Extremal Problems Finite Sets (1991) 137–156.
[7] B. Bollobás, Y. Kohayakawa, T. Luczak, The evolution of random subgraphs of the cube, Random Structures Algorithm 3 (1992) 55–90.
[8] C. Borgs, J.T. Chayes, H. Remco, G. Slade, J. Spencer, Random subgraphs of finite graphs: III. The phase transition for the n-cube.
[9] J.D. Burtin, The probability of connectedness of a random subgraph of an n-dimensional cube, Problems Infom. Transmission 13 (1977) 147–152.
[10] H. Chernoff, A measure of the asymptotic efficiency for tests of a hypothesis based on the sum of observations, Ann. Math. Statist. 23 (1952) 493–509.
[11] P. Erdős, J. Spencer, The evolution of the n-cube, Comput. Math. Appl. 5 (1979) 33–39.

[12] W. Grüner, R. Giegerich, D. Strothmann, C.M. Reidys, J. Weber, I.L. Hofacker, P.F. Stadler, P. Schuster, Analysis of rna sequence structure maps by exhaustive enumeration I. Neutral networks, Chem. Monthly 127 (1996) 355–374.
[13] L.H. Harper, Minimal numberings and isoperimetric problems on cubes, in: Theory of Graphs, International Symposium, Rome, 1966.
[14] T.E. Harris, The Theory of Branching Processes, Springer, 1963.
[15] V.F. Kolchin, Random Mappings, Optimization Software, New York, 1986, xiv +206pp.
[16] M.V. Meshikov, Coincidence of critical points in percolation problems, Soviet Math. Dokl. 33 (1986) 856–859.
[17] M. Molloy, B. Reed, The size of the giant component of a random graph with given degree sequence, Combin. Probab. Comput. 7 (1998) 295–305.
[18] C.M. Reidys, P.F. Stadler, P.K. Schuster, Generic properties of combinatory maps and neutral networks of RNA secondary structures, Bull. Math. Biol. 59 (2) (1997) 339–397.
[19] E.A. Schultes, D.P. Bartel, One sequence, two ribozymes: Implications for the emergence of new ribozyme folds, Science 289 (5478) (2000) 448–452.
[20] P. Schuster, in: Michael Laessig, Angelo Valeriani (Eds.), A Testable Genotype-Phenotype Map: Modeling Evolution of RNA Molecules, Springer, 2002.
[21] P. Schuster, W. Fontana, P.F. Stadler, I.L. Hofacker, From sequences to shapes and back: A case study in RNA secondary structures, Proc. Roy. Soc. B 255 (1994) 279–284.