

SQUARE PERCOLATION AND THE THRESHOLD FOR QUADRATIC DIVERGENCE IN RANDOM RIGHT-ANGLED COXETER GROUPS

JASON BEHRSTOCK, VICTOR FALGAS-RAVRY, AND TIM SUSSE

ABSTRACT. Given a graph Γ , its auxiliary *square-graph* $\square(\Gamma)$ is the graph whose vertices are the non-edges of Γ and whose edges are the pairs of non-edges which induce a square (i.e., a 4-cycle) in Γ . We determine the threshold edge-probability $p = p_c(n)$ at which the Erdős–Rényi random graph $\Gamma = \Gamma_{n,p}$ begins to asymptotically almost surely have a square-graph with a connected component whose squares together cover all the vertices of $\Gamma_{n,p}$. We show $p_c(n) = \sqrt{\sqrt{6} - 2}/\sqrt{n}$, a polylogarithmic improvement on earlier bounds on $p_c(n)$ due to Hagen and the authors. As a corollary, we determine the threshold $p = p_c(n)$ at which the random right-angled Coxeter group $W_{\Gamma_{n,p}}$ asymptotically almost surely becomes strongly algebraically thick of order 1 and has quadratic divergence.

1. INTRODUCTION

In this paper we investigate the phase transition for a variant of “square percolation”, with motivation coming from both previous work on clique percolation and from questions in geometric group theory.

Clique percolation was introduced by Derényi, Palla and Vicsek [14] as a simple model for community detection, and quickly became well-studied in network science, from computational, empirical, and theoretical perspectives, see e.g. [9, 14, 24, 26, 27, 28]. In (k, ℓ) -clique percolation, to investigate the “community structure” of a graph or network Γ one studies the auxiliary (k, ℓ) -clique graph whose vertices are the k -cliques of Γ and whose edges are those pairs of k -cliques having at least ℓ vertices in common.

One of the main research questions in the area was determining the threshold $p = p(n)$ for the emergence of a “giant component” in the auxiliary (k, ℓ) -clique graph when the original graph $\Gamma \in \mathcal{G}(n, p)$ is an Erdős–Rényi random graph on n vertices with edge-probability p . This was completely resolved in 2009 by Bollobás and Riordan [9], in a highly impressive paper making sophisticated use of branching processes. In the concluding remarks of their paper, Bollobás and Riordan suggested a study of “square percolation” as a natural extension of their work. More precisely, given a graph Γ they suggested studying the component structure of the auxiliary graph whose vertices are the *not necessarily induced* 4-cycles in Γ , and whose edges are pairs of 4-cycles with a diagonal¹ in common. For $\Gamma \in \mathcal{G}(n, p)$, they stated that they believed the threshold for the associated auxiliary graph to contain a giant component containing a positive proportion of all squares of Γ should be λ_c/\sqrt{n} , where $\lambda_c = \sqrt{\sqrt{6} - 2}$ (see the discussion around equation (19) in Section 2.3 of [9]).

A related (but slightly different) notion of “square percolation” arose independently in joint work of the authors with Hagen [6] on the divergence of the random right-angled Coxeter group, providing motivation from geometric group theory for understanding the phase transition in an

2010 *Mathematics Subject Classification.* 05C80, 20F65, 57M15, 60B99, 20F55, 20F69.

To appear in *Random Structures & Algorithms*.

¹Note that given a 4-cycle in a graph, we use the term “diagonal” to refer to the pair of vertices of a diagonal, even though they may not span an edge in the graph; indeed most of this paper concerns induced 4-cycles so that the edge spanned by the diagonal is *not* in the graph.

auxiliary graph formed from the *induced* 4-cycles of an Erdős–Rényi random graph $\Gamma \in \mathcal{G}(n, p)$. To make this more precise, we make the following definition.

Definition 1.1. To any graph Γ , we associate an auxiliary *square-graph*, $\square(\Gamma)$, whose vertices are the non-edges of Γ , and whose edges are the pairs of non-edges of Γ that together induce a 4-cycle (a.k.a. *square*) in Γ .

Thus for vertices a, b, c, d in a graph Γ , the pair $\{ac, bd\}$ is an edge of $\square(\Gamma)$ if and only if (i) ac and bd are non-edges of Γ (and thus vertices of $\square(\Gamma)$), and (ii) ab, bc, cd and da are all edges of Γ .

Remark 1.2. This definition of the auxiliary square-graph $\square(\Gamma)$ differs slightly from the one used in the related papers [6, 12]. In those papers, the auxiliary graph had the induced 4-cycles as its vertices, and its edges were those pairs of induced 4-cycles having a diagonal in common. These two variants of auxiliary square-graphs encode essentially the same information, but the formulation above is more natural from a combinatorial perspective and more convenient for the exploration processes we shall consider in this paper.

We investigate the component structure of $\square(\Gamma)$, albeit with an unusual twist. With a view to applications in geometric group theory, we will be interested in the question of whether or not $\square(\Gamma)$ has a component that “covers” all of the vertex-set of the original graph Γ .

Definition 1.3. We refer to connected components of $\square(\Gamma)$ as *square-components* of Γ . Given a square-component C we define its *support* to be the collection of vertices of Γ given by:

$$\text{supp}(C) = \bigcup_{vw \in C} \{v, w\},$$

and say that the component C of $\square(\Gamma)$ *covers* the vertex set $\text{supp}(C) \subseteq V(\Gamma)$. If C covers all of $V(\Gamma)$, we say it is a square-component with *full support*

Write $\Gamma \in \mathcal{G}(n, p)$ to denote that Γ is an instance of the Erdős–Rényi random graph model with parameters n and p , i.e., that Γ is a graph on n vertices obtained by including each edge at random with probability p , independently of all the others.

Our main combinatorial result in this paper is pinpointing the precise threshold $p_c(n)$ at which $\Gamma \in \mathcal{G}(n, p)$ asymptotically almost surely² experiences a phase transition from having only square-components with support of logarithmic order to having a square-component with full support. Throughout this paper we set $\lambda_c = \sqrt{\sqrt{6} - 2}$. The following two results establish that the critical threshold probability is $p(n) = \lambda_c n^{-1/2}$ by showing highly disparate behavior on either side of this threshold as given by the following two contrasting results.

Theorem 1.4 (Subcritical Behavior). *Let $\lambda < \lambda_c$ be fixed. Suppose that $p(n) \leq \lambda n^{-1/2}$. Then for $\Gamma \in \mathcal{G}(n, p)$, a.a.s. every square-component of Γ covers at most $O((\log n)^{2^{32}})$ vertices.*

Theorem 1.5 (Supercritical Behavior). *Let $\lambda > \lambda_c$ be fixed, and let $f: \mathbb{N} \rightarrow \mathbb{R}^+$ be a function with $f(n) \rightarrow 0$ and $f(n)n^2 \rightarrow \infty$ as $n \rightarrow \infty$. Let $p = p(n)$ be an edge-probability with*

$$\lambda n^{-1/2} \leq p(n) \leq 1 - f(n).$$

Then for $\Gamma \in \mathcal{G}(n, p)$, a.a.s. there is a square-component of Γ covering all vertices of Γ .

Our proofs of Theorems 1.4 and 1.5 confirm the conjecture of Bollobás and Riordan regarding the location of the phase transition for their version of (non-induced) square percolation — see Corollary 6.8 in Section 6.5. Further, Theorems 1.4 and 1.5 have a direct application to the study of the geometric properties of random right-angled Coxeter group, which we now describe.

²As usual, *asymptotically almost surely* or *a.a.s.* is shorthand for “with probability tending to 1 as $n \rightarrow \infty$.”

Given a graph Γ , we define the associated *right-angled Coxeter group* (RACG) W_Γ by taking the free group on $V(\Gamma)$ and adding the relations $a^2 = 1$ and $ab = ba$ for all $a \in V(\Gamma)$, $ab \in E(\Gamma)$. In this way, the graph Γ encodes a *finite presentation* for the right-angled Coxeter group W_Γ . Given graphs Γ and Λ , it is well-known that the associated groups W_Γ and W_Λ are isomorphic if and only if the graphs Γ and Λ are isomorphic, see [25]. Thus algebraic and geometric properties of W_Γ can be studied via purely graph-theoretic means, as we do in this paper. Indeed, a number of geometric properties of a right-angled Coxeter group W_Γ admit encodings as graph-theoretic properties of the presentation graph Γ . Such properties include thickness and having quadratic divergence, which are both important in geometric group theory (see Section 3 below for a formal definition of these notions). An investigation of right-angled Coxeter groups with quadratic divergence was the main motivation for the work undertaken in this paper.

The correspondence between right-angled Coxeter groups and graphs allows one to define models of random groups based on random graph models. In particular, in this paper we consider the *random right-angled Coxeter group*, W_Γ where the presentation graph $\Gamma \in \mathcal{G}_{n,p}$ is an Erdős–Rényi random graph. Using Theorems 1.4 and 1.5 above on square-components in Erdős–Rényi random graphs, we prove the following.

Theorem 1.6 (Criticality for quadratic divergence of RACGs). *Let $\epsilon > 0$. If*

$$\frac{\lambda_c + \epsilon}{\sqrt{n}} \leq p(n) \leq 1 - \frac{(1 + \epsilon) \log n}{n}$$

and $\Gamma \in \mathcal{G}(n, p)$, then, a.a.s. the right-angled Coxeter group W_Γ has quadratic divergence and is strongly algebraically thick of order exactly 1.

On the other hand, if $p(n)$ satisfies

$$0 \leq p(n) \leq \frac{\lambda_c - \epsilon}{\sqrt{n}}$$

then the right-angled Coxeter group W_Γ a.a.s has at least cubic divergence and is not strongly algebraically thick of order 0 or 1.

The geometric properties of W_Γ when $\Gamma \in \mathcal{G}_{n,p}$ and $p = 1 - \theta(n^{-2})$ was described in detail by Behrstock, Hagen and Sisto in [8, Theorem V]. Together with their work, our results give an essentially complete picture of quadratic and linear divergence in random right-angled Coxeter groups.

Organization of the paper. In Section 3 we provide additional background material on the geometry of random groups and derive Theorem 1.6 from Theorems 1.4–1.5. In Section 4, we recall some basic facts about branching processes and give an outline of the proof strategy we follow for our main results, and of the ways in which it departs from the framework used by Bollobás and Riordan in their study of clique percolation in random graphs. Theorem 1.4 is proved in Section 5, while Theorem 1.5 is derived in Section 6. We end the paper in Section 7 with some discussion of the results and of further work and related problems.

Acknowledgments. The authors are extremely grateful to two anonymous referees, whose careful work helped improve the correctness and clarity of the paper. The authors thank Mark Hagen for discussions during the early stages of this work. Behrstock was supported by NSF grant DMS-1710890. Falgas-Ravry was supported by Swedish Research Council grant VR 2016-03488. The authors thank Ela Behrstock for her skillful rendering of Figures 1 and 2. Aspects of this work were motivated by output from software written by the authors (available upon request from the authors, some available online at <http://comet.lehman.cuny.edu/behstock/random.html>). Accordingly, this research was supported, in part, by a grant of computer time from the City University of New York High Performance Computing Center under NSF Grants CNS-0855217, CNS-0958379, and ACI-1126113.

2. GRAPH-THEORETIC NOTATION AND STANDARD NOTIONS

Given a set A and $r \in \mathbb{N}$, let $A^{(r)}$ denote the collection of all subsets of A of cardinality r . So for example $A^{(2)}$ is the collection of unordered distinct pairs of elements of A . As a notational convenience, we set $[n] := \{1, 2, \dots, n\}$, and we often denote the unordered set $\{u, v\}$ by uv .

A graph is a pair $\Gamma = (V, E)$, where $V = V(\Gamma)$ is a set of *vertices* and $E = E(\Gamma)$ is a collection of pairs of vertices referred to as the *edges* of Γ . A subgraph of Γ is a graph G with $V(G) \subseteq V(\Gamma)$ and $E(G) \subseteq E(\Gamma)$. If $V(G) = X$ and $E(G) = E(\Gamma) \cap X^{(2)}$, then we say G is the subgraph of Γ induced by X and denote this fact by $G = \Gamma[X]$. When there is no risk of confusion, we may abuse notation and use X to refer to both the subset of $V(\Gamma)$ and the associated induced subgraph $\Gamma[X]$. The complement of a graph $\Gamma = (V, E)$ is the graph $\Gamma^c = (V, V^{(2)} \setminus E)$.

A path of length ℓ in a graph Γ is an ordered sequence of $\ell + 1$ distinct vertices v_0, v_1, \dots, v_ℓ together with a set of ℓ edges $\{v_{i-1}v_i : i \in [\ell]\} \subseteq E(\Gamma)$. Such a path is said to join v_0 to v_ℓ . Two vertices are said to be connected in Γ if there is a path joining them. Being connected is an equivalence relation on the vertices of Γ . A (*connected*) *component* of Γ is then a nonempty set of vertices from $V(\Gamma)$ that forms an equivalence class under this relation.

In this paper we study *squares* in graphs. A square, or 4-cycle, in Γ is a copy of the graph $C_4 = (\{a, b, c, d\}, \{ab, bc, cd, da\})$ as a subgraph of Γ . In an abuse of notation, we will denote such a C_4 by $abcd$. In other words, if we say “ $abcd$ is a copy of C_4 /a square in Γ ”, we mean “ $ab, bc, cd, da \in E(\Gamma)$ ”. Further if we say “ $abcd$ is an induced C_4 /square in Γ ”, we mean that $abcd$ is a square in Γ and that in addition $ac, bd \notin E(\Gamma)$. A useful notion for studying squares in graphs is that of a link graph: given a vertex $x \in V(\Gamma)$, the *link graph* Γ_x of x is the collection of *neighbors* of x in Γ , i.e., $\Gamma_x = \{y \in V(\Gamma) : xy \in E(\Gamma)\}$.

By $\Gamma \in \mathcal{G}(n, p)$ we mean that Γ is a random graph on the vertex set $[n]$ obtained by including each edge uv in $E(\Gamma)$ with probability p , independently of all the others. This is known as the Erdős–Rényi random graph model. Given a sequence of edge probabilities $p = p(n)$ and a graph property \mathcal{P} , we say that a typical instance of $\Gamma \in \mathcal{G}(n, p)$ has property \mathcal{P} , or, equivalently, that $\Gamma \in \mathcal{P}$ holds *asymptotically almost surely (a.a.s.)* if

$$\lim_{n \rightarrow \infty} \mathbb{P}(\Gamma \in \mathcal{P}) = 1.$$

Throughout the paper, we use standard Landau notation: given functions $f, g: \mathbb{N} \rightarrow \mathbb{R}^+$, we write $f = o(g)$ for $\lim_{n \rightarrow \infty} f(n)/g(n) = 0$ and $f = O(g)$ if there exists a constant $C > 0$ such that $\limsup_{n \rightarrow \infty} f(n)/g(n) \leq C$. Further we write $f = \omega(g)$ for $g = o(f)$, $f = \Omega(g)$ for $g = O(f)$. Finally if $f = O(g)$ and $f = \Omega(g)$ both hold, we denote this fact by $f = \Theta(g)$.

 3. GEOMETRIC GROUP THEORY AND THE \mathcal{CFS} PROPERTY

Our main result in this paper establishes that $p(n) = \lambda_c/\sqrt{n}$ is the threshold for a typical instance Γ of the Erdős–Rényi random graph model $\mathcal{G}(n, p)$ to have a square-graph with a component covering all of $V(\Gamma)$. This property is a.a.s. equivalent to possessing the \mathcal{CFS} -property, defined below.

3.1. Background. Recall that the graph join $\Gamma_1 * \Gamma_2$ of two graphs Γ_1 and Γ_2 is the graph obtained by taking disjoint unions of Γ_1 and Γ_2 , and adding in all edges from Γ_1 to Γ_2 .

Definition 3.1. A finite graph Γ is defined to be \mathcal{CFS} (*constructed from squares*) if Γ has induced subgraphs K and Γ' with K a (possibly empty) clique so that:

- $\Gamma = \Gamma' * K$, and
- $\square(\Gamma')$ has a component C with $\text{supp}(C) = V(\Gamma')$.

Dani–Thomas were the first to introduce a special case of the \mathcal{CFS} property for triangle-free graphs in [12]. The \mathcal{CFS} property for arbitrary graphs was then studied by Hagen and

the authors in [6], with an eye towards establishing when this property holds a.a.s. in random graphs, while in [21] Levcovitz studied the geometric properties of right-angled Coxeter groups whose presentation graphs do not possess the \mathcal{CFS} property.

With Hagen, the authors determined in [6] the threshold for the \mathcal{CFS} property to hold a.a.s. in Erdős–Rényi random graphs up to a polylogarithmic factor.

Theorem 3.2 (Theorems 5.1 and 5.7 in [6]). *If $p(n) \leq (\log n)^{-1}/\sqrt{n}$, then a.a.s. a graph $\Gamma \in \mathcal{G}(n, p)$ does not have the \mathcal{CFS} property. On the other hand if $p(n) \geq 5\sqrt{\log n}/\sqrt{n}$ and $(1-p)n^2 \rightarrow \infty$, then a.a.s. $\Gamma \in \mathcal{G}(n, p)$ does have the \mathcal{CFS} property.*

Our contribution in this paper is to eliminate the polylogarithmic gap in Theorem 3.2 and thus to determine the precise threshold for the \mathcal{CFS} property in random graphs.

The \mathcal{CFS} property is closely linked to the large scale geometry of right-angled Coxeter groups, connected to *divergence* and (*strong algebraic*) *thickness*. Divergence is a quasi-isometry invariant of groups introduced by Gersten [17] and further developed by Druţu, Mozes and Sapir [15], while thickness was introduced by Behrstock–Druţu–Moshier in [5] and then further refined by Behrstock–Druţu in [4]. We define these notions and explain how they are related below.

Definition 3.3. Let (X, d) be a geodesic metric space, let $o \in X$ and let $\rho \in (0, 1]$. Given $x, y \in X$ with $d(x, o) = d(y, o) = r$, we define $d_{\rho r}(x, y)$ to be the infimum of the lengths of paths in $X \setminus B(o, \rho r)$ between x and y , if such a path exists, and ∞ otherwise; here $B(o, \rho r)$ denotes the ball of radius ρr about o . We then set

$$\delta_{\rho}(r) = \sup_{o \in X} \sup_{x, y \in \partial B(o, r)} d_{\rho r}(x, y).$$

The *divergence* of X is defined to be the collection of functions $\delta_{\rho} : r \mapsto \delta_{\rho}(r)$,

$$\text{Div}(X) := \{\delta_{\rho} : \rho \in (0, 1]\}.$$

Given two non-decreasing functions $f, g : \mathbb{N} \rightarrow \mathbb{R}^+$, we say that $f \lesssim g$ if there exists $C \geq 1$ so that:

$$f(r) \leq C \cdot g(Cr + C) + Cr + C,$$

and we say $f \sim g$ if $f \lesssim g$ and $g \lesssim f$. Importantly, two polynomials that are non decreasing $\mathbb{N} \rightarrow \mathbb{R}^+$ and have the same degree are equivalent under this relation, and further for $a, b \in \mathbb{N}$ we have $x^a \sim x^b$ if and only if $a = b$.

When X is the Cayley graph of a right-angled Coxeter group, it is straightforward to see that $\delta_{\rho}(r) \sim \delta_1(r)$. Therefore when we are referring to the *divergence function* of W_{Γ} , we will mean $\delta_1(r)$. We say that a RACG W_{Γ} has *quadratic divergence* if $\delta_1(r) \sim r^2$ and *linear divergence* if $\delta_1(r) \sim r$.

Definition 3.4. Let G be a finitely generated group.

- We say that G is strongly algebraically thick of order 0 if it has linear divergence.
- We say that G is strongly algebraically thick of order at most n if G has a collection of subgroups $\mathcal{H} = \{H_{\alpha}\}$ so that:
 - $\langle \bigcup_{\alpha} H_{\alpha} \rangle$ has finite index in G
 - for $H_{\alpha}, H_{\beta} \in \mathcal{H}$ there exists a sequence $H_0 = H_{\alpha}, H_1, \dots, H_{k-1}, H_k = H_{\beta}$ of elements of \mathcal{H} so that $H_{i-1} \cap H_i$ is infinite for each $1 \leq i \leq k$
 - there exists a constant $M > 0$ so that each $H_{\alpha} \in \mathcal{H}$ is M -*quasiconvex*, that is to say, every pair of points in H_{α} can be connected by an (M, M) -quasigeodesic contained in H_{α} .
 - each $H_{\alpha} \in \mathcal{H}$ is strongly algebraically thick of order at most $n - 1$.

Further, we say that G is strongly algebraically thick of order *exactly* n if it is strongly algebraically thick of order at most n and *not* strongly algebraically thick of order at most $n - 1$. We also usually write “thick” as a shorthand for “strongly algebraically thick”.

Behrstock and Druţu discovered that the order of thickness provides upper bounds on the divergence of a metric space. In particular they proved:

Proposition 3.5 ([4, Corollary 4.17]). *Let G be a finitely generated group which is strongly algebraically thick of order at most n . Then for every $\rho \in (0, 1]$, $\delta_\rho(r) \lesssim r^{n+1}$.*

The group theoretic motivation for studying the \mathcal{CFS} property is that it provides a graph theoretical proxy for certain geometric properties of right-angled Coxeter groups, such as their divergence. To see that W_Γ has quadratic divergence when Γ has the \mathcal{CFS} property is straightforward, since interpreting the definition of \mathcal{CFS} in the Cayley graph yields a chain of linearly many spaces with linear divergence with each intersecting the next in an infinite diameter set. Indeed, it is an immediate consequence of the definitions that if G is the direct product of two infinite groups then G has linear divergence, just as a path avoiding a linear-sized ball in the plane has linear length. Hence every finitely generated abelian group of rank at least 2 has linear divergence (and is thick of order 0).

Now, if $\Gamma = \Gamma' * K$ where K is a clique, then $W_\Gamma \cong W_{\Gamma'} \times \mathbb{Z}_2^{|K|}$. In such a case $W_{\Gamma'}$ is a finite-index subgroup of W_Γ and thus, up to finite index, we can assume that Γ does not contain a vertex sending an edge to all other vertices of Γ .

Now, W_Γ contains a network of convex subgroups generated by the induced squares in the full-support component of $\square(\Gamma)$. Each of these groups is *virtually* \mathbb{Z}^2 , that is to say has a finite index subgroup which is a copy of \mathbb{Z}^2 . Further, two induced squares in Γ correspond to incident edges in $\square(\Gamma)$ if and only if the intersection of the associated virtual \mathbb{Z}^2 subgroups is *virtually* \mathbb{Z} , that is to say has a finite index subgroup which is a copy of \mathbb{Z} .

Thus, paths in the full-support component of $\square(\Gamma)$ give the connecting sequences needed in Definition 3.4. Hence if Γ has the \mathcal{CFS} property, W_Γ is thick of order at most 1 and has at most quadratic divergence.

As shown by Dani–Thomas in the triangle-free case [12, Theorem 1.1 and Remark 4,8], and by the present authors with Hagen in the general case (as above) [6, Proposition 3.1], if Γ has the \mathcal{CFS} property then the associated right-angled Coxeter group W_Γ has thickness of order at most 1, and hence has at most quadratic divergence. Further, in [8], Behrstock, Hagen and Sisto show that a right-angled Coxeter group W_Γ has linear divergence (and is thick of order 0) if and only if Γ is the join of two non-complete graphs. Right-angled Coxeter groups feature a type of non-positive curvature captured by the notion of being $\text{CAT}(0)$, see, e.g., [13]. In most cases, $\text{CAT}(0)$ spaces satisfy the dichotomy that their divergence must either be linear or at least quadratic (the seminal result in this direction was proved by Kapovich and Leeb [20, Proposition 3.3] under the assumption that there exists a periodic geodesic with superlinear divergence; when there exists a Morse geodesic the statement follows from [4, Theorem 6.6] or [11, Theorem 2.14]; for right-angled Coxeter groups it follows from rank-rigidity [10] or the fact that they are hierarchically hyperbolic groups and thus have linear or at least quadratic divergence [7, 1]). Finally, Levcovitz proved that any graph without \mathcal{CFS} has at least cubic divergence [21], and so we see that W_Γ has exactly quadratic divergence if and only if Γ is not the join of two non-complete graphs and has the \mathcal{CFS} property.

3.2. Proof of threshold for quadratic divergence in random RACGs. Assuming our main theorems about square percolation, Theorems 1.4 and 1.5, we are now in a position to provide a proof of Theorem 1.6 on the threshold for quadratic divergence in RACGs:

Proof of Theorem 1.6 from Theorems 1.4 and 1.5. Let $\epsilon > 0$, and suppose that

$$\frac{\lambda_c + \epsilon}{\sqrt{n}} \leq p(n) \leq 1 - \frac{(1 + \epsilon) \log n}{n}.$$

By Theorem 1.5, the graph Γ a.a.s. has the \mathcal{CFS} property. Thus, by [6, Proposition 3.1], W_Γ a.a.s. has at most quadratic divergence and is thick of order at most 1. Further, since $1 - p(n) \geq \frac{(1+\epsilon)\log n}{n}$, standard results on the connectivity of Erdős–Rényi random graphs tell us that a.a.s. the complement of Γ is connected, and thus that Γ itself is a.a.s. not the join of two non-trivial graphs. Thus, [8] implies that W_Γ is not thick of order 0 and hence is thick of order exactly one and has precisely quadratic divergence.

On the other hand, if

$$p(n) \leq \frac{\lambda_c - \epsilon}{\sqrt{n}},$$

then by Theorem 1.4 no component of the square graph can have full support, and thus the graph Γ is not \mathcal{CFS} . It then follows from [21], that W_Γ has at least cubic divergence, and thus by Proposition 3.5 that it is not thick of order 1. \square

4. BRANCHING PROCESSES AND PROOF STRATEGY

4.1. Branching processes. We recall here some basic facts and definitions from the theory of branching processes that we will use in our argument; for a more general treatment of such processes, see e.g. [3].

Definition 4.1. A Galton–Watson branching process $\mathbf{W} = (W_t)_{t \in \mathbb{Z}_{\geq 0}}$ with offspring distribution X is a sequence of non-negative integer-valued random variables with $W_0 = 1$ and for all $t \geq 1$, $W_t = \sum_{i=1}^{W_{t-1}} X_{i,t}$, where the $X_{i,t}$: $i, t \in \mathbb{N}$ are independent, identically-distributed random variables with $X_{i,t} \sim X$ for all i, t .

A Galton–Watson branching process can be viewed as a random rooted tree: in the zero-th generation there is a root or ancestor, who begets a random number $X_{1,1} \sim X$ of children that form the first generation. In every subsequent generation, each child independently begets a random number of children, with the i -th member of generation t begetting $X_{i,t} \sim X$ children.

Galton–Watson branching processes are a widely studied family of random processes and are the subject of much probabilistic research; see e.g. [3] and the references therein. Here we introduce only some fairly standard elements of the theory that are needed for our argument. A Galton–Watson process \mathbf{W} is said to become *extinct* if $W_t = 0$ for some $t \in \mathbb{N}$. The *total progeny* of \mathbf{W} is the total number of vertices in the associated tree, which we denote by $W = \sum_{t=0}^{\infty} W_t$; this quantity is finite if and only if \mathbf{W} becomes extinct.

A key tool in the study of \mathbf{W} is the generating function of its offspring distribution, $f_X(t) = \mathbb{E}t^X$. The following standard results from the theory of branching processes relate the probability of extinction for \mathbf{W} to the mean and generating function of its offspring distribution X .

Proposition 4.2 (See e.g. [3]). *Let \mathbf{W} be a Galton–Watson branching process with offspring distribution X . Let $\mu = \mathbb{E}X$ and $f(t) = f_X(t)$. Then the following hold:*

(i) (**subcritical regime**) *if $\mu < 1$, then almost surely \mathbf{W} becomes extinct, and what is more,*

$$\mathbb{P}(\mathbf{W} \text{ has not become extinct by generation } k) = \mathbb{P}(W_k \neq 0) \leq \mu^k;$$

(ii) (**supercritical regime**) *if $\mu > 1$, then the probability θ_e that \mathbf{W} becomes extinct is the smallest solution $\theta \in [0, 1]$ to the equation*

$$f(\theta) = \theta,$$

and satisfies $\theta_e < 1$.

We shall also need the following result on the distribution of the total progeny W of \mathbf{W} .

Proposition 4.3 (Dwass’s formula [16]). *Let \mathbf{W} be a Galton–Watson branching process with offspring distribution X . Then the total progeny W satisfies*

$$\mathbb{P}(W = k) = \frac{1}{k} \mathbb{P}(X_1 + X_2 + \cdots + X_k = k - 1),$$

where X_1, X_2, \dots, X_k are independent, identically distributed random variables with $X_i \sim X$ for all $i \in [k]$.

4.2. Departures from the Bollobás–Riordan framework. Bollobás and Riordan in [9] developed a powerful branching process framework for the study of clique percolation. Much of that framework can be adapted to the study of the non-induced square percolation we are concerned with in this paper. However there remains a number of significant hurdles which need to be overcome in order to extend their techniques to the present setting.

In the subcritical regime, the structure of squares makes the analysis of exceptional edges and offspring distributions (which are the crux of the argument) differ significantly from the Bollobás–Riordan paper; care is needed to handle the resulting complications correctly. Indeed, Bollobás and Riordan are able to model clique percolation using a Galton–Watson branching process whose offspring distribution is roughly Poisson; however, for square percolation, the offspring distribution is more heavy-tailed, forcing us to resort to somewhat delicate technical arguments.

Further, in the supercritical regime, because of our motivation from geometric group theory, we are interested in the study of *induced* square percolation. In particular, adding new edges to a graph could destroy some induced squares and hence split apart square-components even as we are trying to build a giant square-component. This situation is quite unlike that in clique percolation, and we have to use a completely different sprinkling argument to obtain our results (*inter alia* sprinkling vertices rather than edges). Thus here again there are significant complications and major departures from Bollobás and Riordan’s framework in [9].

4.3. Proof strategy. Our results rely on the analysis of a branching process exploration of the square-components of a graph $\Gamma \in \mathcal{G}(n, p)$ for some fixed $p = \lambda n^{-1/2}$ where $\lambda > 0$.

We begin with an arbitrary induced square $S_1 = abcd$ in Γ . Its diagonals ac and bd give us two pairs of non-edges which can be used to discover further non-edges of Γ belonging to the same square-component. The size of the set $(\Gamma_a \cap \Gamma_c) \setminus \{b, d\}$ of common neighbors of a and c in $V(\Gamma) \setminus \{b, d\}$ is a binomially distributed random variable $Z \sim \text{Binom}(n - 4, p^2)$. Assuming that $\Gamma_a \cap \Gamma_c$ is an *independent set* (i.e., contains no edge of Γ) these common neighbors together with b, d give rise to $\binom{Z+2}{2}$ non-edges that lie in the same square-component as ac ; however, since we already knew about the pair bd , only $X = \binom{Z+2}{2} - 1$ of these are new. We then pursue our exploration of the square-component of ac by iterating this procedure: for each as-yet untested non-edge xy in our square-component, we can first find the common neighbors of xy , and add as “children” of xy all the new non-edges discovered in this way.

This can be viewed as a Galton–Watson branching process \mathbf{W} with offspring distribution X in a natural way. Assuming the past exploration does not greatly interfere with the distribution of the number of children in our process, the expected number of children at each step is roughly equal to

$$\mathbb{E}X = \mathbb{E} \left(\binom{Z+2}{2} - 1 \right) = \mathbb{E} \left(\frac{Z^2 + 3Z}{2} \right).$$

The expected value of X is readily computed from the first and second moments of the binomial distribution of Z , yielding

$$\mathbb{E}X = \frac{1}{2} \lambda^4 + 2 \lambda^2 + o(1).$$

The Galton–Watson process \mathbf{W} becomes critical when the expectation of its offspring distribution is 1. Solving

$$\frac{1}{2}\lambda^4 + 2\lambda^2 = 1$$

and selecting the non-negative root $\lambda_c = \sqrt{\sqrt{6} - 2} = 0.6704\dots$, we thus see that for any fixed $\lambda < \lambda_c$, our branching process \mathbf{W} is subcritical. We thus expect it to terminate a.a.s. after a fairly small number of steps, from which one can hope to, in turn, deduce that a.a.s. all square-components are small. On the other hand, for any fixed $\lambda > \lambda_c$, \mathbf{W} is supercritical, and with probability strictly bounded away from zero it does not terminate before we have discovered a reasonably large number of non-edges. A second-moment argument can then be used to show that a strictly positive proportion of non-edges must lie in reasonably large square-components. With a little glueing work, we can then hope to show that in fact a strictly positive proportion of non-edges lie in a giant square-component that covers all the vertices of Γ .

The above is however a simplification of what is actually required to make the arguments go through, and the situation turns out to be considerably more nuanced than what we described above. A first issue is our assumption that the vertices in Z form an independent set: in the subcritical regime, we need to consider what happens if the set Z of common neighbors of some non-edge xy which we are testing interacts with some other previously discovered vertices, or with vertices in Z . In particular, any “exceptional” edge from Z to previously discovered vertices other than x, y could potentially create many additional squares, and hence add many new pairs to our square-component which are not accounted for by our branching process. Bollobás and Riordan faced a similar problem in their work on clique percolation. However, as stated in the previous subsection, the way they dealt with “exceptional edges” does not quite work for us in the square percolation setting. One issue is that in a copy of C_4 , vertices on opposite sides of a diagonal are not adjacent, so that the number of 4-cycles created by an exceptional edge cannot always be bounded by the degree of a newly discovered vertex. In addition, we note that if it is not dealt with properly, the presence of exceptional edges could significantly affect the future distribution of the number of children in our branching process: if in the example above a, c had *three* common neighbors among the already discovered vertices rather than two, then the correct number of children for ac in the exploration process would be $\binom{Z+3}{2} - 3 = \binom{Z+2}{2} - 1 + Z$, which has expectation equal to $1 + \lambda_c^2 > 1$ when $\lambda = \lambda_c$. Finally, for the argument to work, we need not only for a Galton–Watson branching process with offspring distribution X to become a.a.s. extinct within a few generations (which is an easy first moment argument): we also need its total progeny to be a.a.s. small. Here the fact that X is a quadratic function of the binomial random variable Z (and thus rather heavy-tailed) causes complicated issues, which were not faced in [9] (where the offspring distribution was essentially Poisson with mean < 1). Overcoming these problems is the main work done in Section 5.

Secondly, in the supercritical argument, after establishing the a.a.s. existence of many non-edges in reasonably large square-components, we must prove the a.a.s. existence of a giant square-component covering all vertices and a strictly positive proportion of non-edges of Γ . Here the crucial point is that, because of the applications in geometric group theory motivating our work, we are considering *induced* square percolation. The size of a largest square-component in Γ is not monotone with respect to the addition of edges to the graph — adding an edge could very well destroy an induced square, thus potentially breaking a large square-component into several smaller pieces. So we have to use a completely new sprinkling argument to be able to agglomerate all “reasonably large” square-components into a single giant square-component. To do this we reserve some vertices for sprinkling, rather than edges. We use these vertices to build bridges between reasonably large square-components in a sequence of rounds until all such components are joined into one. Finally once we have established the a.a.s. existence of a giant square-component, some care is needed to ensure this square-component covers every vertex of

Γ . Assembling a giant square-component and ensuring it has full support in this way involves overcoming a number of interesting obstacles, and is the main work done in Section 6.

5. THE SUBCRITICAL REGIME: PROOF OF THEOREM 1.4

Theorem 1.4 will be established as an immediate consequence of a stronger result, Theorem 5.1, which we state and prove below after providing a few preliminary definitions.

Given a graph Γ , in addition to the square-graph $\square(\Gamma)$ from Definition 1.1 we shall consider a different but closely related auxiliary graph $\boxtimes(\Gamma)$ that includes information about all squares in Γ (rather than just the induced squares). Explicitly we let $\boxtimes(\Gamma) := (V(\Gamma)^{(2)}, \{\{ac, bd\} : ab, bc, cd, ad \in E(\Gamma)\})$ be the graph whose vertices are pairs of vertices from $V(\Gamma)$ and whose edges correspond to (not necessarily induced) copies of C_4 . The *support* $\text{supp}(C)$ of a component C of $\boxtimes(\Gamma)$ is defined as in Definition 1.3, *mutatis mutandis*.

Note that the square graph $\square(\Gamma)$ is exactly the subgraph of $\boxtimes(\Gamma)$ induced by the set $\{ab \in V(\Gamma)^{(2)} : ab \notin E(\Gamma)\}$ of non-edges of Γ . In particular, for every square-component C in $\square(\Gamma)$, there is a component C' in $\boxtimes(\Gamma)$ with $C \subseteq C'$ and thus $\text{supp}(C) \subseteq \text{supp}(C')$. To establish Theorem 1.4, it is thus enough to prove the following stronger theorem that bounds the size of the support in Γ of components of $\boxtimes(\Gamma)$.

Theorem 5.1. *Let $\lambda < \lambda_c$ be fixed. Suppose that $p(n) \leq \lambda n^{-1/2}$. Then for $\Gamma \in \mathcal{G}(n, p)$, a.a.s. every component of $\boxtimes(\Gamma)$ has a support of size $O((\log n)^{2^{31}})$.*

Since the order of the support of the largest component in $\boxtimes(\Gamma)$ is monotone non-decreasing with respect to the addition of edges to Γ , we may assume in the remainder of this section that $p(n) = \lambda n^{-1/2}$. Further, since $\lambda < \lambda_c$ is fixed, there exists a constant $\varepsilon > 0$ such that for a binomially distributed random variable $Z \sim \text{Binom}(n, p^2)$ we have

$$(5.1) \quad \mathbb{E} \left(\binom{Z+2}{2} - 1 \right) = 1 - \varepsilon.$$

With this last equality in hand, we are now ready to present and analyse the exploration process that lies at the heart of our proof of Theorem 5.1.

We shall discover a superset of the component of $\boxtimes(\Gamma)$ which contains some fixed pair $v_1 v_2 \in V_\Gamma^{(2)}$. We begin our exploration by finding common neighbors of v_1 and v_2 , then adding all pairs of such newly discovered vertices to a set of active pairs. These new pairs obviously lie in the same component of $\boxtimes(\Gamma)$ as $v_1 v_2$.

After this initial step in the exploration, we proceed as follows. First, we choose a new pair a_t from our set of active pairs. By assumption, there exists a previously explored pair b_t such that all 4 edges from a_t to b_t are present. We continue our exploration by finding the set Z_t of common neighbors of the vertices in a_t among the *previously undiscovered* vertices of Γ . We then add all pairs $(Z_t \cup b_t)^{(2)} \setminus \{b_t\}$ to our set of active pairs — these obviously lie in the same component of $\boxtimes(\Gamma)$ as a_t — and delete a_t from that set. We then repeat the procedure, choosing a new active pair a_{t+1} , finding its common neighbors among undiscovered vertices, etc.

This, however, is not enough to discover the totality of the component of $v_1 v_2$ in $\boxtimes(\Gamma)$. Indeed, it is possible that the pair a_t has additional common neighbors among already discovered vertices (in addition to the two neighbors in b_t), which could give rise to additional active pairs that lie in the same component of $\boxtimes(\Gamma)$ as a_t . To deal with this possibility, we have to add in an exceptional phase in our exploration, which takes care of potential additional edges among the vertices we have discovered. In this exceptional phase, we generously overestimate how many new active pairs could be discovered, and for each of these new pairs we start new and essentially independent versions of our exploration process, which we think of as children processes of our original exploration process. (Each of these children processes could potentially trigger an

exceptional phase; however we keep track of how many such exceptional phases have been triggered — we call the non-exceptional phases of exploration between consecutive exceptional phases *epochs*, and show that with high probability our overall exploration process terminates within five epochs.)

The key is that, with the exceptional phase factored in, we do discover a superset of the collection of all pairs in the same component of $\boxtimes(\Gamma)$ as our starting pair v_1v_2 . We compare the non-exceptional phase of our exploration to a subcritical branching process and give upper bound on its total progeny, which is small. Obtaining this bound is somewhat tricky (due to the nature of our offspring distribution) and relies on a rather technical application of Dwass's formula (Proposition 4.3). The remainder of the proof is provided in Lemma 5.10 which shows that, with high probability, we do not run through more than five exceptional phases. In particular, we do not start too many child processes, so that with high probability our overall exploration process stops before we have discovered a large number of vertices.

5.1. An exploration process. Our exploration process will proceed by considering the following for each time $t \geq 0$:

- **(Discovered vertices.)** An ordered set of vertices: $D_t = \{v_1, v_2, \dots, v_{d_t}\}$.
- **(Active pairs.)** An ordered set of pairs of vertices from D_t : $A_t = \{x_1y_1, x_2y_2, \dots, x_{a_t}y_{a_t}\}$.
- **(Discovered pairs.)** A set of pairs of vertices: $S_t \subseteq D_t^{(2)}$.
- **(Explored edge set.)** A set of edges: $E_t \subseteq D_t^{(2)} \cap E(\Gamma)$.
- **(Epoch.)** An integer: $e_t \in \{0, 1, 2, 3, 4, 5\}$.

These sets will satisfy:

- (\star) for all $t \geq 0$ and for every active pair $x_iy_i \in A_t$, the vertices x_i and y_i have at most 2 common neighbors in the graph (D_t, E_t) .

The initial state of the exploration consists of the following data, which is seeded by a choice of v_1v_2 , an arbitrary pair of vertices from $V(\Gamma)$ (note that this pair can, alternatively, be thought of as a vertex of $\boxtimes(\Gamma)$). We set $D_0 = \{v_1, v_2\}$, $A_0 = S_0 = \{v_1v_2\}$, $E_0 = \emptyset$ and $e_0 = 0$.

At each time step t our exploration proceeds as follows, with ε as given in equation (5.1):

1. If $|D_t| < 2^{2^{10}} \varepsilon^{-2^{10}} (\log n)^{2^{31}}$ and $A_t \neq \emptyset$, let $a = xy$ be the first pair in A_t . For each $z \in V(\Gamma) \setminus D_t$ we test whether or not z sends an edge in Γ to both vertices of a . Set $Z_t := \{z \in V(\Gamma) \setminus D_t : zx, zy \in E(\Gamma)\}$ and F_t to be the collection of joint neighbors of x and y in (D_t, E_t) . (Note, by property (\star), the set F_t consists of a set of at most two discovered vertices.)

We arbitrarily label the vertices in Z_t as $\{v_{d_{t+1}}, v_{d_{t+2}}, \dots, v_{d_{t+1}}\}$, and add them to D_t to form D_{t+1} . We then set

$$A_{t+1} = \left(A_t \cup (F_t \cup Z_t)^{(2)} \right) \setminus \left(a \cup F_t^{(2)} \right)$$

to be the new collection of active pairs, and extend the ordering on A_t to an ordering on A_{t+1} by letting the new pairs from $(F_t \cup Z_t)^{(2)}$ be prior to the old pairs from A_t . We further set $E_{t+1} = E_t \cup \{zx, zy : z \in Z_t\}$, set $e_{t+1} = e_t$, $S_{t+1} = S_t \cup A_{t+1}$ and then proceed to the next time step $t + 1$ of the process. Note that since the only new edges being added are ones connecting a new vertex to x and y and since $xy \notin A_{t+1}$ each pair in A_{t+1} has at most 2 common neighbors in (D_{t+1}, E_{t+1}) , i.e., property (\star) is still satisfied in the next time-step.

2. If $|D_t| \geq 2^{2^{10}} \varepsilon^{-2^{10}} (\log n)^{2^{31}}$, then we terminate the process and declare *large stop*.
3. If $|D_t| < 2^{2^{10}} \varepsilon^{-2^{10}} (\log n)^{2^{31}}$ and $A_t = \emptyset$, then we consider $i = |E(\Gamma[D_t]) \setminus E_t|$.

If $i = 0$ or $e_t + i \geq 5$, then we terminate our exploration process and declare *extinction stop* or *exceptional stop*, respectively.

Otherwise we set $e_{t+1} = e_t + i$, update E_t by setting $E_t = E(\Gamma[D_t])$, and set $i_1 = i$ (which by assumption is > 0). We then run the following subroutines:

3A. Set

$$Z_t^1 := \{z \in V(\Gamma) \setminus D_t : z \text{ sends at least three edges into } D_t\}.$$

and let $i_1 := |Z_t^1|$.

- If $i_1 > 0$ and $e_{t+1} + i_1 > 5$, then we terminate the whole exploration process and declare *exceptional stop*.
- Else if $i_1 > 0$ and $e_{t+1} + i_1 \leq 5$, we add Z_t^1 to D_t , update E_t by setting $E_t = E(\Gamma[D_t])$, update S_t by setting $S_t = D_t^{(2)}$. We then update e_{t+1} to $e_{t+1} + i_1$ and run through subroutine **3A.** again.
- Otherwise $i_1 = 0$ and we proceed to subroutine **3B.**

3B. Let

$$Z_t^2 := \{z \in V(\Gamma) \setminus D_t : z \text{ sends at least two edges into } D_t\},$$

Since subroutine **3A.** terminated with $i_1 = 0$ each vertex in Z_t^2 sends exactly two edges into D_t . We set $D_{t+1} = D_t \cup Z_t^2$, add all edges lying between Z_t^2 and D_t to E_t to form E_{t+1} , and let A_{t+1} consist of all pairs of vertices in D_{t+1} containing at least one vertex of Z_t^2 . Further, we set $S_{t+1} = S_t \cup A_{t+1}$.

Once this is done, we proceed to the next time-step $t + 1$ in the overall exploration process, observing that property (\star) has been preserved (since by construction every vertex in Z_t^2 has degree exactly two in (D_{t+1}, E_{t+1})).

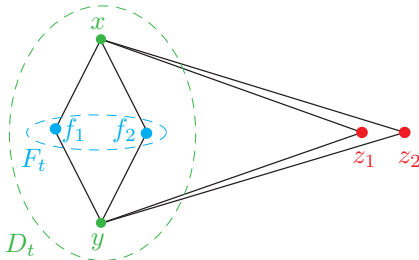


FIGURE 1. An illustration of Stage 1 of the Exploration Process, with exploration from the active pair $a = xy$ at time t . In this example, $F_t = \{f_1, f_2\}$ and the set of newly discovered vertices is $Z_t = \{z_1, z_2\}$. We thus have $A_{t+1} \setminus A_t = \{f_1z_1, f_2z_1, f_1z_2, f_2z_2, z_1z_2\}$ and $A_t \setminus A_{t+1} = \{xy\}$.

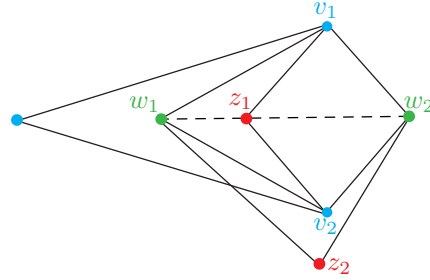


FIGURE 2. Illustration of Stage 3 of the Exploration Process. Suppose the pairs $w_1w_2, w_1z_1, w_2z_1 \in S_t$ were discovered while v_1v_2 was active. When w_1w_2 is active in Stage 1 the pairs $\{v_1z_2, v_2z_2\}$ will be discovered. However the pair z_1z_2 will only be discovered in an instance of Stage 3, if the two dashed edges are revealed to be present (thus making $w_1z_1w_2z_2$ a square).

5.2. Analysing the process. The exploration process defined in the previous subsection can terminate for one of three reasons:

- (1) $|D_t| \geq 2^{2^{10}} \varepsilon^{-2^{10}} (\log n)^{2^{31}}$ (*large stop*);
- (2) $e_t \geq 5$ (*exceptional stop*);

(3) $A_t = \emptyset$ and $E_t = \Gamma[D_t]$ (*extinction stop*).

It follows from the above that the process must in fact terminate within $O\left((\log n)^{2^{32}}\right)$ time-steps. We begin our analysis by noting that, given our aim of proving Theorem 1.4, extinction stops are good for us:

Lemma 5.2. *Suppose that the exploration from v_1v_2 terminates at time T with an extinction stop. Let C be the component of $\boxtimes(\Gamma)$ containing v_1v_2 . Then $C \subseteq S_T$. Furthermore, the number of vertices in the support of C is at most $2^{2^{10}} \epsilon^{-2^{10}} (\log n)^{2^{31}}$ and $|C| \leq 2^{2^{11}} \epsilon^{-2^{11}} (\log n)^{2^{32}}$.*

Proof. We perform our exploration process from the pair v_1v_2 , and assume it terminates with an extinction stop at time T . It is enough to show that given $a = u_1u_2 \in C \cap S_t$, for every neighbor $b = w_1w_2$ of a in $\boxtimes(\Gamma)$, there is some $t' \leq T$ such that $b \in S_{t'}$.

If a was discovered at a time-step where **1.** applied or if $a \in A_0$, then $a \in A_t$ and was an active pair at some time $t \geq 0$. Thus, at some later time-step t' where **1.** applies, our exploration process selects a as its “exploration pair” and discover all neighbors z_1z_2 of a in $\boxtimes(\Gamma)$ with $z_1z_2 \in (Z_{t'} \cup F_{t'})^{(2)}$ — where $F_{t'}$ is the pair we used to discover a , and $Z_{t'}$ is the collection of joint neighbors of u_1 and u_2 that lie in $V(\Gamma) \setminus D_{t'}$. If b is in this set, then $b \in S_{t'}$.

Otherwise, if we failed to find $b = w_1w_2$ at this time t' , b must contain at least one vertex from $D_{t'} \setminus F_{t'}$. By property (\star) of our exploration, at least one of the edges u_iw_j , $i, j \in \{1, 2\}$ lies outside $E_{t'}$.

In particular, since we do not end with a large or exceptional stop, this edge will be uncovered at later time step t'' where **3A.** applies. But by the end of **3B.**, all vertices sending at least two edges into $D_{t''}$ have been added to $D_{t''}$. Thus all common neighbors of a will be found (since $a \subseteq D_{t''}$, and a common neighbor of a has at least two neighbors in $D_{t''}$). Hence after the updates $b \subseteq D_{t''}$. There are two options: if both vertices of b are present after **3A.**, then $b \in S_{t''}$, since after **3A.** all possible pairs of discovered vertices (not already tested) are added to $S_{t''}$. Otherwise, both vertices of b are present after **3B.**, and since at least one of them was discovered in **3B.**, b is added to $A_{t''+1} \subseteq S_{t''+1}$, and we are done again.

If on the other hand a was discovered at a step t where **3B.** applies, then a is added to A_{t+1} , and the above applies. Finally, if a was discovered at a time step t where **3A.** applies, then in **3A.** and **3B.**, all common neighbors of a are added to D_t , and all such pairs are added to S_t or S_{t+1} .

Either way, since $S_t \subseteq S_T$ for all $t \leq T$, we see that $b \in S_T$. Thus every neighbor of a in $\boxtimes(\Gamma)$ is eventually discovered by our exploration process, and $C \subseteq S_T$ as claimed. Further, since $S_T \subseteq D_T^{(2)}$ by construction, and since our exploration ends with an extinction stop by the hypothesis of the lemma, we have $|C| \leq \frac{1}{2}|D_T|^2 < 2^{2^{11}} \epsilon^{-2^{11}} (\log n)^{2^{32}}$ as claimed. \square

We now turn to the technical crux of the analysis. We will need Harris’s lemma on correlations between monotone events, which we state below after recalling some definitions. Given a set F , a *decreasing event* in the space $\{0, 1\}^F$ is a subset \mathcal{D} of $\{0, 1\}^F$ such that for all $\mathbf{x} \in \mathcal{D}$ and all $\mathbf{y} \in \{0, 1\}^F$ with $y_f \leq x_f$ for all $f \in F$, we have $\mathbf{y} \in \mathcal{D}$. Similarly, an *increasing event* is a subset \mathcal{U} of $\{0, 1\}^F$ such that for all $\mathbf{x} \in \mathcal{D}$ and all $\mathbf{y} \in \{0, 1\}^F$ with $x_f \leq y_f$ for all $f \in F$, we have $\mathbf{y} \in \mathcal{D}$. (So decreasing events are closed under decreasing coordinates, while increasing events are closed under increasing coordinates.) A *principal increasing event* is an increasing event of the form $\{\mathbf{x} : x_f = 1 \text{ for all } f \in F'\}$, where F' is some fixed subset of F .

Lemma 5.3 (Harris’s lemma [18]). *Let F be a finite set and $p \in [0, 1]$. Let \mathcal{D} be a decreasing event in $\{0, 1\}^F$ and let \mathcal{U} be an increasing event in $\{0, 1\}^F$. Let \mathbf{x} be a random element of $\{0, 1\}^F$ chosen by setting $x_f = 1$ with probability p and $x_f = 0$ otherwise, independently at random for each $f \in F$. Then*

$$\mathbb{P}(\mathbf{x} \in \mathcal{D} \cap \mathcal{U}) \leq \mathbb{P}(\mathbf{x} \in \mathcal{D}) \mathbb{P}(\mathbf{x} \in \mathcal{U}).$$

Corollary 5.4. *Let F be a finite set and $p \in [0, 1]$. Let F' be a subset of F and let $\mathcal{U}_{F'} = \{\mathbf{x} : x_f = 1 \text{ for all } f \in F'\} \subseteq \{0, 1\}^F$ denote the corresponding principal increasing event. Let $\tilde{\mathcal{U}}$ be an increasing event in $\{0, 1\}^{F \setminus F'}$, and let \mathcal{U} be the collection of elements of $\{0, 1\}^F$ whose projection onto $\{0, 1\}^{F \setminus F'}$ is an element of $\tilde{\mathcal{U}}$. Finally, let \mathcal{D} be a decreasing event in $\{0, 1\}^F$, and let \mathbf{x} be a random element of $\{0, 1\}^F$ chosen by setting $x_f = 1$ with probability p and $x_f = 0$ otherwise, independently at random for each $f \in F$. Denote by $\tilde{\mathbf{x}}$ the projection of \mathbf{x} onto $\{0, 1\}^{F \setminus F'}$. Then, provided $\mathbb{P}(\mathbf{x} \in \mathcal{D} \cap \mathcal{U}_{F'}) > 0$, we have:*

$$\mathbb{P}(\mathbf{x} \in \mathcal{U} | \mathbf{x} \in \mathcal{D} \cap \mathcal{U}_{F'}) \leq \mathbb{P}(\tilde{\mathbf{x}} \in \tilde{\mathcal{U}}) = \mathbb{P}(\mathbf{x} \in \mathcal{U}).$$

Proof. Let $\tilde{\mathcal{D}}$ denote the projection of $\mathcal{D} \cap \mathcal{U}_{F'}$ onto $\{0, 1\}^{F \setminus F'}$; observe that, since \mathcal{D} is a decreasing event in $\{0, 1\}^F$, $\tilde{\mathcal{D}}$ is a decreasing event in $\{0, 1\}^{F \setminus F'}$. Let $\mathbf{x}_{F'} = \mathbf{1}$ denote the event that $x_f = 1$ for all $f \in F'$, and note this event is independent of $\tilde{\mathbf{x}}$. Then

$$\begin{aligned} \mathbb{P}(\mathbf{x} \in \mathcal{U} | \mathbf{x} \in \mathcal{D} \cap \mathcal{U}_{F'}) &= \frac{\mathbb{P}(\mathbf{x} \in \mathcal{U} \cap \mathcal{D} | \mathbf{x} \in \mathcal{U}_{F'})}{\mathbb{P}(\mathbf{x} \in \mathcal{D} | \mathbf{x} \in \mathcal{U}_{F'})} = \frac{\mathbb{P}(\tilde{\mathbf{x}} \in \tilde{\mathcal{U}} \cap \tilde{\mathcal{D}} | \mathbf{x}_{F'} = 1)}{\mathbb{P}(\tilde{\mathbf{x}} \in \tilde{\mathcal{D}} | \mathbf{x}_{F'} = 1)} \\ &= \mathbb{P}(\tilde{\mathbf{x}} \in \tilde{\mathcal{U}} | \tilde{\mathbf{x}} \in \tilde{\mathcal{D}}) \leq \mathbb{P}(\tilde{\mathbf{x}} \in \tilde{\mathcal{U}}), \end{aligned}$$

with the inequality coming from Harris's lemma. \square

With this result in hand we can show that our exploration process is dominated by a subcritical branching process.

Lemma 5.5. *If we are at a time-step t of the process where **1.** applies, then given the past history of the process, the random variable $|A_{t+1} \setminus A_t|$ counting the number of new active pairs discovered by $a_1 = x_1 y_1$ is stochastically dominated by a random variable $X = \frac{Z^2 + 3Z}{2}$, where Z is a binomial random variable with parameters n and p^2 .*

Proof. Our analysis in this proof is similar to that of Bollobás and Riordan in [9, Inequality (3)]. Suppose we are at a time-step t of the process where **1.** applies. The past history of the process then consists of the following information:

- (1) a certain set E_t of edges are present in Γ ;
- (2) for a certain T of pairs $(z, xy) \in V(\Gamma) \times V(\Gamma)^{(2)}$, at least one of the edges zx , zy is missing from Γ .

The information in (1) can be encoded as a principal increasing event \mathcal{U}_{E_t} in $\{0, 1\}^{V(\Gamma)^{(2)}}$ in the natural way, while (2) can be viewed as a decreasing event \mathcal{D} .

Let $a = xy$ be the active pair we test at time-step t . For every vertex $z \in V(\Gamma) \setminus D_t$, the events $\tilde{\mathcal{U}}_z := \{zx, zy \in E(\Gamma)\}$ are independent increasing events in the space $\{0, 1\}^{V(\Gamma)^{(2)} \setminus E_t}$. Let \mathcal{U}_z denote the corresponding increasing event in $\{0, 1\}^{V(\Gamma)^{(2)}}$. Applying Corollary 5.4, we have that

$$\mathbb{P}(U_z | \mathcal{D} \cap \mathcal{U}_{E_t}) \leq \mathbb{P}(U_z) = p^2.$$

(We note here that the fact $z \notin D_t$ is essential — we have no control over the conditional probabilities of edges inside the set of discovered vertices D_t , which could potentially lie in E_t .) It follows that the number of hitherto undiscovered vertices $z \in V(\Gamma) \setminus D_t$ which the active pair a discovers (i.e. the size of $|Z_t|$) is stochastically dominated by a binomial random variable Z with parameters n and p^2 .

By property (\star) , we have that x and y have $|F_t| \leq 2$ common neighbours in D_t , whence by definition of our exploration process the number X of new active pairs discovered by $a = xy$ is stochastically dominated by $\binom{Z+2}{2} - 1 = \frac{Z^2 + 3Z}{2}$, as claimed. \square

Property (\star) is key to the proof of Lemma 5.5 — without it, we do not have the requisite stochastic domination, which is we need the exceptional phase of our exploration. However, having used property (\star) in this proof, we shall not need it in the remainder of this section. We turn instead to the problem of controlling how much of $V(\Gamma)$ we can discover using our branching processes.

Lemma 5.6. *Let $Z \sim \text{Binom}(n, p^2)$ and $k \in \mathbb{N}$. Then $\mathbb{P}(Z \geq 9 \log n + 9 \log k) \leq n^{-5} k^{-6}$.*

Proof. Recall that in this section, $p = \lambda n^{-1/2}$, for some constant $\lambda < \lambda_c$. Since $Z \sim \text{Binom}(n, p^2)$, we have:

$$\begin{aligned} \mathbb{P}\left(Z \geq 9(\log n + \log k)\right) &= \sum_{r=\lceil 9(\log n + \log k) \rceil}^n \binom{n}{r} p^{2r} (1-p^2)^{n-r} < \sum_{r=\lceil 9(\log n + \log k) \rceil}^n n^r \left(\lambda_c n^{-1/2}\right)^{2r} \\ &= \sum_{r=\lceil 9(\log n + \log k) \rceil}^n (\lambda_c)^{2r} < n \lambda_c^{2\lceil 9(\log n + \log k) \rceil} \leq n \exp(18 \log \lambda_c (\log n + \log k)), \end{aligned}$$

where in the last two inequalities we used the fact $(\lambda_c)^2 = \sqrt{6} - 2 < 1$. Since $18 \log \lambda_c < -6$, this immediately gives us the desired bound

$$\mathbb{P}(Z \geq 9 \log n + 9 \log k) < n e^{-6 \log n - 6 \log k} = n^{-5} k^{-6}.$$

□

Corollary 5.7. $\mathbb{P}(\exists xy \in V(\Gamma)^{(2)} : |\Gamma_x \cap \Gamma_y| \geq 9 \log n) \leq n^{-3}$.

Proof. Fix $xy \in V(\Gamma)^{(2)}$. By Lemma 5.6 with $k = 1$, we have

$$\mathbb{P}(|\Gamma_x \cap \Gamma_y| \geq 9 \log n) = \sum_{r=\lceil 9 \log n \rceil}^{n-2} \binom{n-2}{r} p^{2r} (1-p^2)^{n-2r} < \mathbb{P}(Z \geq 9 \log n) \leq n^{-5}.$$

Taking a union bound over all $\binom{n}{2} < n^2$ possible choices of the pair xy , the lemma follows. □

We now analyse the total progeny of the Galton–Watson branching process with offspring distribution given by the random variable X from the statement of Lemma 5.5. By (5.1), $\mathbb{E}X = 1 - \epsilon$ and thus the branching process is subcritical. Unfortunately, combining the Markovian bound on the extinction time from Proposition 4.2(i), with the bounds on the maximum degree in $\boxtimes(\Gamma)$ from Corollary 5.7 does not give us sufficiently good control on the extinction time and total progeny of our Galton–Watson process. Thus we turn to an application of Dwass’s formula to obtain the tighter bounds needed for the proof of Theorem 1.4.

Lemma 5.8. *Let $\mathbf{W} = (W_t)_{t \in \mathbb{Z}_{\geq 0}}$ be a Galton–Watson branching process with an offspring distribution X as in Lemma 5.5. Set $k_0 = 2^{26} \epsilon^{-2} (\log n)^5$. Then \mathbf{W} is subcritical, and its total progeny $W = \sum_{t=0}^{\infty} W_t$ satisfies*

$$\mathbb{P}(W \geq k_0) = O(n^{-5}).$$

Proof. Let $\{X_{k,j} : k \in \mathbb{N}, j \in [k]\}$ be an infinite family of independent, identically distributed copies of X . For each $k \in \mathbb{N}$ and every $j \in [k]$, write $X_{k,j}$ as $X_{k,j} = X_{k,j}^a + X_{k,j}^b$, where $X_{k,j}^a = \min(X_{k,j}, 2^8 (\log n + \log k)^2)$. Set $\mu_k^a = \mathbb{E}(X_{k,1}^a)$. By construction, $\mu_k^a \leq \mathbb{E}[X] = 1 - \epsilon$. Since $X = (Z^2 + 3Z)/2 \leq 2Z^2$, where $Z \sim \text{Binom}(n, p^2)$, and since $2(9(\log n + \log k))^2 < 2^8 (\log n + \log k)^2$, Lemma 5.6 implies that for every $k \in \mathbb{N}$ and $j \in [k]$,

$$(5.2) \quad \mathbb{P}\left(X_{k,j}^b > 0\right) \leq \mathbb{P}\left(2Z^2 \geq 2^8 (\log n + \log k)^2\right) \leq \mathbb{P}(Z \geq 9(\log n + \log k)) \leq n^{-5} k^{-6}.$$

Applying Dwass's formula, Proposition 4.3, to our branching process \mathbf{W} , we have that for any $k \in \mathbb{N}$,

$$(5.3) \quad \begin{aligned} \mathbb{P}(W = k) &= \frac{1}{k} \mathbb{P} \left(\sum_{j=1}^k X_{k,j} = k - 1 \right) \\ &\leq \frac{1}{k} \left(\mathbb{P} \left(\sum_{j=1}^k \frac{X_{k,j}^a - \mu_k^a}{2^8(\log n + \log k)^2} \geq \frac{k(1 - \mu_k^a) - 1}{2^8(\log n + \log k)^2} \right) + \mathbb{P} \left(\sum_{j=1}^k X_{k,j}^b > 0 \right) \right). \end{aligned}$$

Since $\mu_k^a \leq 1 - \varepsilon$, for $k \geq 2\varepsilon^{-1}$ we have that $2(1 - \mu_k^a) - \frac{2}{k} \geq 2(1 - \mu_k^a) - \varepsilon \geq \varepsilon$ and hence $k(1 - \mu_k^a) - 1 \geq \frac{\varepsilon k}{2}$. Thus we have

$$s := \frac{k(1 - \mu_k^a) - 1}{2^8(\log n + \log k)^2} > \frac{\varepsilon k}{2^9(\log n + \log k)^2} := s'.$$

Since the random variables $(X_{k,j}^a - \mu_k^a)/(2^8(\log n + \log k)^2)$ are by construction independent random variables with mean zero and absolute value at most 1, we can apply a standard Chernoff bound [2, Theorem A.1.16] in (5.3) to obtain:

$$\mathbb{P} \left(\sum_{j=1}^k \frac{X_{k,j}^a - \mu_k^a}{2^8(\log n + \log k)^2} \geq s \right) \leq e^{-\frac{(s')^2}{2k}} = e^{-\frac{\varepsilon^2 k}{2^{19}(\log n + \log k)^4}}.$$

Letting, $k_0 = k_0(n) = \lceil 2^{26}\varepsilon^{-2}(\log n)^5 \rceil$ we have that for n sufficiently large, all $k \geq k_0(n)$ satisfy $k \geq 5\varepsilon^{-2} \cdot 2^{19}(\log n + \log k)^5$. Indeed, for $k \geq \log n$ we have $5\varepsilon^{-2} \cdot 2^{19}(\log n + \log k)^5 \leq 5\varepsilon^{-2} \cdot 2^{24}(\log k)^5$, which for k sufficiently large is less than k . Thus picking n sufficiently large (which in turn makes k_0 sufficiently large) and $k = k_0(n)$, we have:

$$(5.4) \quad \mathbb{P} \left(\sum_{j=1}^k \frac{X_{k,j}^a - \mu_k^a}{2^8(\log n + \log k)^2} \geq s \right) \leq e^{-\frac{\varepsilon^2 k}{2^{19}(\log n + \log k)^4}} \leq e^{-5(\log n + \log k)} \leq n^{-5}k^{-5}.$$

Further, applying inequality (5.2) and a union bound, we get that

$$(5.5) \quad \mathbb{P} \left(\sum_{j=1}^k X_{k,j}^b > 0 \right) \leq k\mathbb{P}(X_{k,1}^b > 0) \leq n^{-5}k^{-5}.$$

Combining (5.3), (5.4) and (5.5), we finally have

$$\mathbb{P}(W \geq k_0) \leq \sum_{k=k_0}^{\infty} \mathbb{P}(W = k) \leq \sum_{k=k_0}^{\infty} \frac{1}{k} 2n^{-5}k^{-5} = O(n^{-5}).$$

□

We now use Lemma 5.8 to estimate the probability that our exploration ends with a large stop. The key is to view our exploration as a branching process of branching processes: we begin with a subcritical branching process, corresponding to step **1.** in our exploration process. Call this the ancestral branching process. When this process becomes extinct, we run through step **3.** of our exploration process, potentially adding new active pairs to our otherwise empty set of active pairs A_t . For each of these new pairs, we start an independent child branching process. For each of these we repeat the same procedure as for the ancestral branching process (so the child processes can generate their own child processes, and so forth). Thus to bound the total number of pairs discovered over the entire course of our exploration, we must control the growth of this branching process of branching processes.

Lemma 5.9. *The probability that the exploration from v_1v_2 terminates with a large stop is $O(n^{-3})$.*

Proof. We view our exploration process as a kind of branching process of branching processes, with parent processes begetting children processes as described at the start of this section. Beginning from the single pair v_1v_2 , the exploration of its component in $\boxtimes(\Gamma)$ undertaken at time-steps t when **1.** applies is dominated by a subcritical branching process \mathbf{W} as in the statement of Lemma 5.8. When that process terminates, we have discovered a certain set D_{t_0} of vertices of Γ . If **3.** applies, then we add a certain number of vertices to D_{t_0} to form D_{t_0+1} and then add a subset of the pairs from $D_{t_0+1}^{(2)}$ to form A_{t_0+1} . We may view each of the pairs a_i added to A_{t_0+1} at this time as the root of a subcritical independent branching process \mathbf{W}_i^3 . There are at most $|D_{t_0+1}|^2$ of these ‘‘child-processes’’, and they are stochastically dominated by independent copies \mathbf{W}_i of the subcritical branching process \mathbf{W} from Lemma 5.8. When all these branching processes have become extinct, we now have a (larger) set D_{t_1} of discovered vertices and we may be back at a time step where **3.** applies. We repeat our procedure — adding vertices to form D_{t_1+1} , adding new pairs to form A_{t_1+1} , etc. The whole procedure can begin again at most 5 times (for otherwise an exceptional stop must have occurred).

We run our exploration ignoring large stops and only applying **1.** and **3.** until the process terminates (with an exceptional stop or an extinction stop), and show that if an exceptional stop does not occur then with probability $1 - O(n^{-3})$ the final size of the discovered set of vertices is at most $2^{2^{10}} \varepsilon^{-2^{10}} (\log n)^{2^{31}}$.

Since we never can consider more than $\binom{n}{2}$ different active pairs, it follows that we never start more than n^2 branching processes \mathbf{W}_i in the course of our exploration. By Lemma 5.8, and a union bound, the probability that one of our at most n^2 branching processes \mathbf{W}_i has a total progeny of more than $k_0 = 2^{26} \varepsilon^{-2} (\log n)^5$ is $O(n^{-3})$. Assume from now on this does not happen. Since the progeny of our processes correspond to pairs $xy \in V(\boxtimes(\Gamma)) = V(\Gamma)^{(2)}$, none of our processes can add more than $2k_0$ vertices to the set of discovered vertices D_t .

Further, by Corollary 5.7, the probability that there is any pair $xy \in V(\boxtimes(\Gamma)) = V(\Gamma)^{(2)}$ such that $|\Gamma_x \cap \Gamma_y| \geq 9 \log n$ is at most n^{-3} . Assume from now on this does not happen. Then in the first time-step t_0 where **3.** applies, we can bound the number of vertices added to D_{t_0} : each pair xy can contribute at most $9 \log n$ vertices to Z_t^1 or Z_t^2 , and as we do not have an exceptional stop we can repeat the addition procedure at most 5 times. In particular we have

$$\begin{aligned} |D_{t_0+1}| &\leq \left(\left(\left(\left(\left((|D_{t_0}|)^2 9 \log n \right)^2 9 \log n \right)^2 9 \log n \right)^2 9 \log n \right)^2 9 \log n \right)^2 9 \log n \\ &\leq (|D_{t_0}| 9 (\log n))^{2^6} \leq (18(\log n)k_0)^{2^6} . \end{aligned}$$

The number of child processes started at that time step is at most $|D_{t_0+1}|^2$; by our assumption, each of these discovers at most $2k_0$ vertices in total, so that by the next time-step t_1 when **3.** applies, we have

$$|D_{t_1}| \leq |D_{t_0+1}|^2 2k_0 \leq (18(\log n)k_0)^{2^7} 2k_0 := 2k_1.$$

Repeating the analysis above, we obtain

$$|D_{t_1+1}| \leq (18(\log n)k_1)^{2^6} ,$$

³To be more precise: as we explore our active pairs in a depth-first manner in time-steps t where **1.** applies, adding newly discovered active pairs prior to previously discovered active pairs in A_t in the ordering on A_{t+1} , it follows that for pairs a, a' added to A_{t_0+1} the branching process associated to a will die off before the branching process associated to a' begins. Thus the two processes do not mix; we thank the referee for suggesting that we clarify this point.

and in the next time-step t_2 where **3.** applies we have

$$|D_{t_2}| \leq |D_{t_1+1}|^2 2k_0 \leq (18(\log n)k_1)^{2^7} 2k_0 =: 2k_2.$$

and we can keep going in this way, defining k_3, k_4 mutatis mutandis. If we avoid an exceptional stop, then we must terminate by the fifth time-step t_4 when **3.** applies. Iterating our analysis, we see that the size of the final set of discovered vertices D_{t_4} is at most

$$\begin{aligned} |D_{t_4}| &\leq (18(\log n)k_3)^{2^7} 2k_0 \\ &= \left(9(\log n) \left(9(\log n) \left(9(\log n) (18(\log n)k_0)^{2^7} 2k_0 \right)^{2^7} 2k_0 \right)^{2^7} 2k_0 \right)^{2^7} 2k_0 \\ &< (18(\log n)k_0)^{2^{29}} < 2^{2^{10}} \varepsilon^{-2^{10}} (\log n)^{2^{31}}. \end{aligned}$$

This shows that the probability our process terminates with a large stop is $O(n^{-3})$. \square

Finally, we compute the probability that an exploration ends with an exceptional stop.

Lemma 5.10. *The probability that the exploration from $v_1 v_2$ terminates with an exceptional stop is $O((\log n)^{30 \cdot 2^{31}} n^{-5/2})$.*

Proof. Suppose that the exploration from $v_1 v_2$ terminates at time T with an exceptional stop. Then we must have discovered at least 5 exceptional edges at time-steps $t \leq T$ when **3.** applied, where we call an edge exceptional if it appeared in $E(\Gamma[D_t]) \setminus E_t$ (type 1) or as the third or above edge from some vertex $z \in Z_t^1$ to D_t (type 2), and where such edges are ordered according to the ordering of the vertices of D_t .

Since the exploration did not terminate with a large stop, we had $|D_t| \leq 2^{2^{10}} \varepsilon^{-2^{10}} (\log n)^{2^{31}} := \Delta$ at the start of each time-step t . Also, since we did not terminate with a large stop, the time T at which the process terminated must satisfy $T \leq \Delta^2$ (this is an upper bound on the number of pairs we could have tested at time steps where **1.** or **3.** applied).

In any time-step t where **3.** applies and we are testing for membership in one of the (at most five) sets Z_t^1 considered in that turn, the probability that a vertex in $V(\Gamma) \setminus D_t$ sends at least three edges of Γ to the set D_t , conditional on the history of the process up to that point, is by Corollary 5.4 at most

$$\sum_{i=3}^{|D_t|} \binom{|D_t|}{i} p^i \leq \sum_{i=3}^{\Delta} \Delta^i p^i = O(\Delta^4 p^3).$$

Since we have at most n vertices in $V(\Gamma) \setminus D_t$ and at most $T \leq \Delta^2$ time-steps to choose from, the probability of having found at least j type 2 exceptional edges for some $1 \leq j \leq 5$ is

$$(5.6) \quad O\left(T^j (n\Delta^4 p^3)^j\right) = O\left(\Delta^{6j} n^{-j/2}\right).$$

A similar (but simpler) calculation yields that the probability of having found $5 - j$ edges of type 1 is:

$$(5.7) \quad O\left(T^{5-j} (\Delta^2 p)^{5-j}\right) = O\left(\Delta^{4(5-j)} n^{-(5-j)/2}\right).$$

Adding the bounds (5.6) and (5.7) together and substituting in the value of Δ , the Lemma follows. \square

We are now in a position to prove Theorem 5.1, which, as previously noted, immediately implies Theorem 1.4.

Proof of Theorem 5.1. Let v_1v_2 be an arbitrary pair of vertices from $V(\Gamma)$. By Lemmas 5.9 and 5.10, with probability $1 - O((\log n)^{30 \cdot 2^{31}} n^{-5/2}) = 1 - o(n^{-2})$ the exploration from v_1v_2 terminates with an extinction stop. By Lemma 5.2 we obtain a bound on the size of each component of v_1v_2 in $\boxtimes(\Gamma)$ found by this exploration, and a bound on its support as well. By a simple union bound, with probability $1 - o(1)$, all pairs v_1v_2 lie in components of $\boxtimes(\Gamma)$ supported on sets of size at most $2^{2^{10}} \varepsilon^{-2^{10}} (\log n)^{2^{31}}$ in $V(\Gamma)$. \square

6. THE SUPERCRITICAL REGIME: PROOF OF THEOREM 1.5

Fix $\lambda > \lambda_c$. Suppose $\lambda n^{-1/2} \leq p(n) \leq (1 - f(n))$, where $f(n)$ is a function with $f(n) = o(1)$ and $f(n) = \omega(n^{-2})$. Let $\Gamma \in \mathcal{G}(n, p)$. By [6, Theorem 5.1], we know that if $p(n) \geq 5\sqrt{\log n/n}$, then a.s. there is a square-component covering all of $V(\Gamma)$. We may thus restrict our attention in the proof of Theorem 1.5 to the range $\lambda n^{-1/2} \leq p(n) \leq 5\sqrt{\log n/n}$. For such $p(n)$, there exists $\varepsilon > 0$ such that if $Z \sim \text{Binom}(n, p^2)$, then

$$(6.1) \quad \mathbb{E} \left(\frac{Z^2 + 3Z}{2} \right) \geq 1 + \varepsilon.$$

We shall prove the existence of a giant square-component with full support in four stages: first, we define an exploration process in $\square(\Gamma)$. In a second stage, we analyse the process to show that a.s. a large proportion of non-edges of Γ lie in ‘‘somewhat large’’ square-components of $\square(\Gamma)$. Next, in the (more involved) third stage of the argument, we perform *vertex-sprinkling* to show a.s. a large proportion of non-edges of Γ lie in a giant square-component. Finally, we show there exists a.s. a giant square-component covering all of $V(\Gamma)$.

6.1. An exploration process. We consider an exploration process consisting of the following data at each time $t \geq 0$:

- **(Discovered vertices.)** An ordered set of vertices: $D_t = \{v_1, v_2, \dots, v_{d_t}\} \subseteq V(\Gamma)$.
- **(Active pairs.)** A set of non-edges of $\Gamma[D_t]$: $A_t = \{x_1y_1, x_2y_2, \dots, x_{a_t}y_{a_t}\} \subseteq D_t^{(2)} \setminus E(\Gamma)$.
- **(Reached pairs.)** A set of non-edges of $\Gamma[D_t]$: $R_t \subseteq D_t^{(2)} \setminus E(\Gamma)$.

These sets will satisfy:

- (\star) for every $t > 0$ and every active pair $x_iy_i \in A_t$, there is a reached pair $uv \in R_t$ such that x_iy_iuv induces a copy of C_4 in Γ .

The initial state $t = 0$ of the exploration consists of an arbitrary induced C_4 of Γ , denoted $v_1v_2v_3v_4$, and the sets: $D_0 = \{v_1, v_2, v_3, v_4\}$, $A_0 = \{v_1v_3, v_2v_4\}$, and $R_0 = \emptyset$.

Our exploration then proceeds as follows:

1. **If** $|R_t| + |A_t| > (\log n)^4$, then we terminate the process.
2. **If** $|R_t| + |A_t| \leq (\log n)^4$ **and** $A_t \neq \emptyset$, then for each $z \in V(\Gamma) \setminus D_t$ we test whether or not z sends an edge in Γ to both of $\{x_1, y_1\}$ which are the vertices of the first pair $a_1 = x_1y_1$ in the ordered set A_t . We then set $Z_t := \{z \in V(\Gamma) \setminus D_t : zx_1, zy_1 \in E(\Gamma_t)\}$. Denote by F_t the set of common neighbors of x_1 and y_1 in D_t , which by property (\star) (for $t > 0$) and the initial state of our exploration (for $t = 0$) has size at least 2 and contains at least one non-edge of Γ .

We then set $A_{t+1} = (A_t \setminus \{x_1y_1\}) \cup \left((F_t \cup Z_t)^{(2)} \setminus \left(F_t^{(2)} \cup E(\Gamma) \right) \right)$, $R_{t+1} = R_t \cup \{x_1y_1\}$ and $D_{t+1} = D_t \cup Z_t$. We then proceed to the next time-step in the exploration process, noting that property (\star) is maintained. (Note that once we have added v_1v_3 to R_1 at the conclusion of the $t = 0$ step, it is clear that all active pairs in A_1 induce a copy C_4 in Γ when taken together with v_1v_3 , so that (\star) is satisfied at time $t = 1$.)

3. **If** $|R_t| \leq (\log n)^4$ **and** $A_t = \emptyset$, then we terminate the process.

6.2. Many non-edges in somewhat large components. The exploration process defined in the previous subsection can terminate for one of two reasons:

- (1) (**Large stop.**) $|R_t| + |A_t| > (\log n)^4$, or
- (2) (**Extinction stop.**) $A_t = \emptyset$.

The process always terminates after some number $T \leq (\log n)^4$ of time-steps. By construction (and more specifically by property (\star)), at all times $t \geq 0$ the collection of pairs $A_t \cup R_t$ is a subset of the square-component of v_1v_3 and v_2v_4 in $\square(\Gamma)$. Our aim is to show that with somewhat large probability $|A_T| + |R_T| > (\log n)^4$.

Lemma 6.1. *At any time-step $t \geq 0$, the distribution conditional on the past history of the process of the random variable $X_t = |A_{t+1} \setminus A_t|$ counting the number of new active pairs discovered by $a_1 = x_1y_1$ stochastically dominates a random variable X' with mean $\mathbb{E}(X') \geq 1 + \varepsilon + o(1)$.*

Proof. Our arguments in the proof of this Lemma resemble those used in Lemma 5.5 and the analysis of Bollobás and Riordan in [9, Inequality (3)].

Given a vertex z and a vertex-pair a , let $U_{z,a}$ denote the increasing event that z sends edges to both vertices in a in Γ , and $\overline{U_{z,a}}$ the complementary, decreasing event that at least one of the edges from z to a is missing in Γ . Suppose we are at a time-step t of the process where **2.** applies. The past history of the process then consists of the following information:

- (1) a certain set $E_t \subseteq D_t^{(2)}$ of edges are present in Γ ;
- (2) a certain set $A_t \cup R_t \subseteq D_t^{(2)}$ of vertex-pairs are non-edges in Γ
- (3) for every reached pair $a \in R_t$, and we have a subset of vertices $T_a \supseteq V(\Gamma) \setminus D_t$ such that for every $z \in T_a$, the event $\overline{U_{z,a}}$ occurs.

The information in (1) can be encoded as a principal increasing event \mathcal{U}_{E_t} in $\{0, 1\}^{V(\Gamma)^{(2)}}$ in the natural way, while (2) and (3) can together be viewed as a principal decreasing event \mathcal{D}_1 and an intersection of decreasing events (and hence itself a decreasing event) \mathcal{D}_∞ .

Given an as yet-undiscovered vertices $z \in V(\Gamma) \setminus D_t$, we claim that the probability of the event U_{z,x_1y_1} that z sends edges to both vertices in our current active pair x_1y_1 conditional on the past history of the process is not much smaller than the unconditional probability p^2 . To begin with, observe that since $z \notin D_t$, we have $zx_1, zy_1 \notin E_t \cup A_t \cup R_t$. Further, for any $z' \neq z$ and any $a, a' \in A_t \cup R_t$, the events $U_{z,a}$ and $U_{z',a'}$ are independent. Let $\mathcal{D}_z := \bigcap_{a \in R_t} \overline{U_{z,a}}$. It follows from our discussion and the properties of product measures (i.e. that disjoint edge-sets are independent) first of all that

$$(6.2) \quad \mathbb{P}(U_{z,x_1y_1} | \mathcal{U} \cap \mathcal{D}_1 \cap \mathcal{D}_2) = \mathbb{P}(U_{z,x_1y_1} | \mathcal{D}_z),$$

and secondly that for any pair of distinct vertices $z, z' \in V(\Gamma) \setminus D_t$ the events U_{z,x_1y_1} and U_{z',x_1y_1} are independent when we condition on the past history of our exploration process.

Thus what remains to be done is to bound $\mathbb{P}(U_{z,x_1y_1} | \mathcal{D}_z)$. Let $\mathcal{D}'_z := \bigcap_{a \in D_t^{(2)} \setminus \{x_1y_1\}} \overline{U_{z,a}}$. Clearly $\mathcal{D}'_z \subseteq \mathcal{D}_z$, and $\mathcal{D}'_z \cap U_{z,x_1y_1}$ is exactly the event that zx_1, zy_1 are the only edges of Γ that z sends into D_t . Thus we have

$$(6.3) \quad \mathbb{P}(U_{z,x_1y_1} | \mathcal{D}_z) \geq \mathbb{P}(U_{z,x_1y_1} \cap \mathcal{D}_z) \geq \mathbb{P}(U_{z,x_1y_1} \cap \mathcal{D}'_z) = p^2(1-p)^{|D_t|-2},$$

which is equal to $p^2(1 - O(|D_t|p)) = p^2 - o(p^2)$ (since $|D_t| \leq 2(|R_t| + |A_t|) \leq 2(\log n)^4$).

Combining (6.2) with (6.3), we see that conditional on the past history of the exploration process, at time t the indicator function of the event $Y_z = \{z \in Z_t\}$ stochastically dominates a Bernoulli random variable with mean $p^2 - o(p^2)$. Further the $(Y_z)_{z \in V(\Gamma) \setminus D_t}$ are independent events given our conditioning (since, as remarked after (6.2), each such event is only affected by the state of edges from D_t to z). The random variable $|Z_t|$ thus stochastically dominates

a random variable $Z' \sim \text{Binom}(n - |D_t|, p^2 - o(p^2))$. Let X' denote the sum of $\frac{(Z')^2 + 3Z'}{2}$ independent Bernoulli random variables with parameter $1 - p$. For $|D_t| \leq 2(\log n)^4$, we have

$$\begin{aligned} \mathbb{E}|A_{t+1} \setminus A_t| &= \mathbb{E} \left(\frac{|Z_t|^2 + 3|Z_t|}{2} - |E(\Gamma) \cap (F_t \cup Z_t)^{(2)}| \right) \\ &\geq \mathbb{E}X' = (1 - p)\mathbb{E} \left(\frac{|Z'|^2 + 3|Z'|}{2} \right) \geq 1 + \varepsilon + o(1), \end{aligned}$$

where the inequality follows from the stochastic domination of Z' by Z_t , and where in the last line we have used (6.1) and the fact that a binomial distribution with parameters $n - |D_t|$ and $p^2 + o(p^2)$ is close to $\text{Binom}(n, p^2)$. \square

Let $\theta_e = \theta_e(n, p)$ denote the extinction probability of the supercritical branching process \mathbf{W} with the offspring distribution X' given in the proof of Lemma 6.1. Note that by Proposition 4.2(b), θ_e is bounded away from 1.

Up to the time when it terminates, our exploration process on the square-component of $v_1v_2v_3v_4$ stochastically dominates \mathbf{W} . We now use this fact to show many non-edges of Γ lie in “somewhat large” square-components.

Lemma 6.2 (Many squares in large square-components). *Fix $\lambda > \lambda_c$, $p(n)$ satisfying $\lambda n^{-1/2} \leq p(n) \leq 5n^{-1/2}\sqrt{\log n}$, and $\theta_e = \theta_e(n, p)$ as above. Then, with probability $1 - O(n^{-1})$ the number N of induced C_4 's in Γ which are part of square-components of order at least $(\log n)^4$ satisfies*

$$N = (1 + o(1))\mathbb{E}N \geq 3p^4(1 - p)^2 \binom{n}{4} (1 - \theta_e)(1 + o(1)).$$

Proof. Given a collection of 4 vertices $S \in V(\Gamma)^{(4)}$, let E_S be the indicator function of the event that $\Gamma[S] \cong C_4$ and that our exploration process from S terminates with a large stop (which is equivalent to S being part of a square-component of order at least $(\log n)^4$). Conditional on $\Gamma[S] \cong C_4$, Lemma 6.1 implies that $\mathbb{P}(E_S = 1) \geq 1 - \theta_e$ (which is the probability that the branching process \mathbf{W} does not become extinct). Applying Wald's identity, the expectation μ_N of N thus satisfies

$$\begin{aligned} \mu_N = \mathbb{E}N &= \mathbb{E} \sum_{S \in V(\Gamma)^{(4)}} E_S = \sum_{S \in V(\Gamma)^{(4)}} \mathbb{P}(\Gamma[S] \cong C_4) \mathbb{P}(E_S = 1 | \Gamma[S] \cong C_4) \\ (6.4) \quad &\geq 3p^4(1 - p)^2 \binom{n}{4} (1 - \theta_e) = \Omega(p^4 n^4). \end{aligned}$$

(For the last equality, we used the fact that $(1 - p) = 1 - o(1)$ by our upper bound on p and that $1 - \theta_e = \Omega(1)$ by Lemma 6.1 and Proposition 4.2.) We now use Chebyshev's inequality to show N is concentrated around its mean. To do this, we must bound $\mathbb{E}N^2 = \sum_{S, S' \in V(\Gamma)^{(4)}} \mathbb{E}E_S E_{S'}$. Consider two collections of 4 vertices $S, S' \in V(\Gamma)^{(4)}$.

Claim 1. If $S \cap S' = \emptyset$, then E_S and $E_{S'}$ satisfy

$$\mathbb{E}(E_S E_{S'}) = \mathbb{E}(E_S) \mathbb{E}(E_{S'}) + O((\log n)^9 n^{-1} \mathbb{E}(E_S)).$$

Proof. Our claim is that E_S and $E_{S'}$ are essentially independent. Indeed, let us first perform our exploration process from S (stopping immediately if $\Gamma[S]$ does not induce a copy of C_4). For $Z \sim \text{Binom}(n, p^2)$ (and $\lambda n^{-1/2} \leq p \leq 5n^{-1/2}(\log n)^{1/2}$ as everywhere in this section), we

have

$$(6.5) \quad \mathbb{P}(Z \geq 2^7 \log n) = \sum_{r=\lceil 2^7 \log n \rceil}^n \binom{n}{r} p^{2r} (1-p^2)^{n-r} \leq \sum_{r \geq 2^7 \log n} \left(\frac{en}{r} \cdot 25 \frac{\log n}{n} \right)^r \\ \leq \sum_{r \geq 2^7 \log n} \left(\frac{25e}{2^7} \right)^r = o(n^{-5}).$$

Thus with probability $1 - o(n^{-5})$, the number of vertices added to D_t in the last stage of the exploration process from S is at most $2^7 \log n$, implying that the set D_S of vertices discovered by the process from S has size at most $2(\log n)^4 + 2^7 \log n$. Further, the exploration process from S tests at most $(\log n)^4$ pairs in total. This allows us to bound the probability that the exploration process from S' interacts with the exploration process from S .

First of all, by a union bound and Harris's Lemma, the probability that a vertex in S' is discovered by the process from S is at most

$$4p^2 (\log n)^4 = O((\log n)^5 n^{-1}).$$

(We have at most $(\log n)^4$ steps during which we could try to discover a vertex $z \in S'$, and by Lemma 5.3 at each of these steps, conditional on the decreasing event that we have not discovered any vertex from S' yet, the probability that z sends edges to both vertices in the current active pair is at most p^2 .)

Condition now on $D_S \cap S' = \emptyset$, and begin the exploration process from S' . Suppose that we reach a time step in that exploration process at which we have yet to discover any vertex from D_S . Then by Corollary 5.4, conditional on the past history of the exploration processes the probability that some vertex in D_S sends edges to both vertices the current active pair in the exploration process from S' is at most p^2 . Since the exploration process from S' tests at most $(\log n)^4$ pairs xy , it follows that conditional on $D_S \cap S' = \emptyset$, the probability that some $z \in D_S$ sends an edge to both vertices of such an active pair before the end of the exploration process from S' is at most

$$(\log n)^4 |D_S| p^2 = O((\log n)^9 n^{-1}).$$

In particular, the probability of $E_{S'} = 1$ given $E_S = 1$ differs from $\mathbb{E}E_{S'}$ by at most $O((\log n)^9 n^{-1})$, as claimed. \square

Claim 2. For any $S \in V(\Gamma)^{(4)}$, we have

$$\sum_{S' \in V(\Gamma)^{(4)}: S \cap S' \neq \emptyset} \mathbb{E}(E_S E_{S'}) = O(n^3 p^4 \mathbb{E}(E_S)). \quad (\dagger)$$

Proof. Fix S and consider the various ways in which S and S' could intersect non-trivially.

- There is one choice of S' with $S = S'$, for which we have $\mathbb{E}(E_S E_{S'}) = \mathbb{E}(E_S)$.
- Next, we have at most $4n$ choices of S' with $|S \cap S'| = 3$. Write $S = \{a, b, c, d\}$ and $S' = \{a, b, c, d'\}$. For $E_S E_{S'}$ to be non-zero, both S and S' must induce copies of C_4 in Γ and moreover E_S must occur. This is only possible if $E_S = 1$ and d' sends edges to the two neighbors of d in $\{a, b, c\}$. Arguing as in Claim 1, these two events are almost independent and occur with probability $(1 + o(1))p^2 \mathbb{E}E_S$. Thus the contribution of S' with $|S \cap S'| = 3$ to the left-hand side of (\dagger) is at most $O(np^2 \mathbb{E}(E_S))$.
- There are at most $4n^3$ choices of S' with $|S \cap S'| = 1$. For $E_S E_{S'}$ to be non-zero, it is necessary for S' to induce a copy of C_4 in Γ and for $E_S = 1$. Arguing as in Claim 1, these two events are almost independent and occur with probability $(1 + o(1))p^4 \mathbb{E}E_S$. Thus the contribution of S' with $|S \cap S'| = 1$ to the left-hand side of (\dagger) is at most $O(n^3 p^4 \mathbb{E}(E_S))$.

- Finally, there are at most $6n^2$ choices of S with $|S \cap S'| = 2$. For $E_S E_{S'}$ to be non-zero, it is necessary for the vertices in $S' \setminus S$ to be incident at least three edges in $\Gamma[S']$ and for $E_S = 1$. Arguing as in Claim 1, these two events are almost independent and occur with probability at most $(1 + o(1))p^3 \mathbb{E}E_S$. Thus the contribution of S' with $|S \cap S'| = 2$ to the left-hand side of (†) is at most $O(n^2 p^3 \mathbb{E}(E_S))$.

Since $np = \omega(1)$, we have $O(1 + np^2 + n^2 p^3 + n^3 p^4) = O(n^3 p^4)$, and the analysis above shows the left-hand side of (†) is at most $O(n^2 p^3 \mathbb{E}(E_S))$, as claimed. \square

Together, Claims 1 and 2 imply $\mathbb{E}N^2 \leq (\mathbb{E}N)^2 + O(n^3 p^4 \mathbb{E}N)$. By (6.4), we know $\mu_N = \mathbb{E}N = \Omega(n^4 p^4)$. Since $p = \Omega(n^{-1/2})$, it follows that

$$\text{Var}(N) = O(n^3 p^4 \mathbb{E}(N)) = O\left(\frac{(\mathbb{E}N)^2}{n}\right).$$

Applying Chebyshev's inequality yields that for any $\eta > 0$ with probability at least $1 - O(\eta^{-2} n^{-1})$,

$$(1 + \eta)\mu_N \geq N \geq (1 - \eta)\mu_N \geq (1 - \eta)3p^4(1 - p)^2 \binom{n}{4} (1 - \theta_e)(1 - o(1)),$$

as desired, with the lower bound on μ_N coming from (6.4). \square

Corollary 6.3. *Let $\lambda > \lambda_c$ be fixed, and let $p = p(n)$ be an edge-probability satisfying $\lambda n^{-1/2} \leq p \leq 5n^{-1/2} \sqrt{\log n}$. Then for all $\varepsilon_1 > 0$ sufficiently small, there exist a constant $\varepsilon_2 > 0$ such that if $\Gamma_1 \in \mathcal{G}((1 - 3\varepsilon_1)n, p(n))$, then with probability $1 - O(n^{-1})$ the number N_v of non-edges of Γ_1 that lie in square-components of $\square(\Gamma_1)$ of order at least $(\log n)^4$, satisfies*

$$N_v = (1 + o(1))\mathbb{E}(N_v) \geq \varepsilon_2 n^2.$$

Proof. Let $\lambda' = (\lambda + \lambda_c)/2$. For ε_1 sufficiently small, we have $p(n) \geq \lambda'(n(1 - 3\varepsilon_1))^{-1/2}$. We now consider $\Gamma_1 \in \mathcal{G}(n(1 - 3\varepsilon_1), p)$.

Ideally, we would now like to directly apply Lemma 6.2 in Γ_1 . However, to ensure the stochastic domination in Lemma 6.1, we started our exploration process from an induced C_4 rather than a non-edge — so we know that $\Omega(n^4 p^4)$ induced C_4 's are part of square-components of order at least $(\log n)^4(1 + o(1))$ whereas we want to show $\Omega(n^2)$ non-edges lie in such components. Since some non-edges could have as many as $\Omega(\log n)$ common neighbors in Γ_1 , it would in principle be possible for p of order $n^{-1/2}$ that, for example, the collection of the diagonals of the induced C_4 's contained in such “large” components consists of a set of only $O(n^2/(\log n)^2)$ non-edges. We must thus rule out this situation.

The simplest way to do this is to run through our proof of Lemma 6.2 again, but this time for the variant of our exploration process from Section 6.1 where we begin with an arbitrary non-edge $v_1 v_2$ of Γ_1 , set $D_0 = \{v_1, v_2\}$, $A_0 = \{v_1 v_2\}$ and $R_0 = \emptyset$. We say such an exploration *survives infancy* if at the first time-step the pair $v_1 v_2$ discovers a set Z_1 of joint neighbors that spans at least one non-edge $v_3 v_4$ of Γ .

For p in the range we are considering the random graph Γ a.a.s. does not contain a complete graph on 6 vertices, and we can use this to give a constant order lower bound on θ_S , the probability the process survives infancy:

$$\begin{aligned} \theta_S &\geq \mathbb{P}(|\Gamma_{v_1} \cap \Gamma_{v_2}| \geq 6 \mid v_1 v_2 \notin E(\Gamma)) - \mathbb{P}(\Gamma \text{ contains a clique on 6 vertices}) \\ &\geq \sum_{i=6}^{n-2} \binom{n-2}{i} p^{2i} (1-p^2)^{n-2-i} - \binom{n}{6} p^{\binom{6}{2}} \\ &\geq \binom{n-2}{6} (\lambda n^{-1/2})^{12} (1 - \lambda^2 n^{-1})^{n-8} - n^6 \left(5n^{-1/2} \sqrt{\log n}\right)^{15} = \frac{(\lambda)^{12}}{6!} e^{-\lambda^2} - o(1), \end{aligned}$$

which is bounded away from zero for n large enough.

Conditional on surviving infancy, by Lemma 6.1 the exploration process from v_3v_4 stochastically dominates a supercritical branching process \mathbf{W} with extinction probability $\theta_e = \theta_e((1 - \varepsilon_1)n, p)$. Applying Wald's identity, this implies that the number N_v of non-edges of Γ_1 that belong to square-components of order at least $(\log n)^4$ satisfies

$$\mathbb{E}(N_v) \geq \binom{(1 - 3\varepsilon_1)n}{2} (1 - p)\theta_S(1 - \theta_e) = \Omega(n^2).$$

We now bound $\mathbb{E}(N_V)^2$ much as we did in Lemma 6.2. Given a pair $xy \in V(\Gamma)^{(2)}$, write E_{xy} for the event that xy is a non-edge and that our exploration process from xy terminates with a large stop. Claim 1 from the proof of Lemma 6.2 shows mutatis mutandis that if $\{x, y\} \cap \{x', y'\} = \emptyset$ then

$$\mathbb{E}(E_{xy}E_{x'y'}) = \mathbb{E}(E_{xy})\mathbb{E}(E_{x'y'}) + O((\log n)^9 n^{-1}).$$

For non-disjoint pairs $\{x, y\}$ and $\{x', y'\}$, the situation is actually easier than it was in Claim 2: such pairs contribute at most $2n\mathbb{E}N_v$ to $\mathbb{E}((N_v)^2)$. Thus

$$\text{Var}(N_v) = O(n\mathbb{E}(N_v)) = O(n^{-1}(\mathbb{E}(N_v))^2),$$

and we conclude that with probability $1 - O(n^{-1})$ there are $(1 + o(1))\mathbb{E}(N_v) = \Omega(n^2)$ non-edges contained in square-components of order at least $(\log n)^4$. The Corollary then follows from a suitable choice of the constant ε_2 . \square

6.3. A connecting lemma. The key to our sprinkling argument is the following, which we use to connect the somewhat large square-components into even larger square-components. We connect square-components by sprinkling in vertices, and looking for complete bipartite graphs with bipartition $\{x_1, x_2, y_1, y_2\} \sqcup \{z_1, z_2\}$, where x_1x_2, y_1y_2 are non-edges in distinct square-components, and z_1z_2 is a non-edge inside the set of newly sprinkled vertices — see Figure 3 below.

Recall that a p -random bipartite graph with partition $V \sqcup W$ is a graph on the vertex set $V \sqcup W$ obtained by including each pair $\{v, w\}$ with $v \in V, w \in W$ as an edge independently at random with probability p .

Lemma 6.4 (Connecting Lemma). *Let $\lambda > \lambda_c$, $\delta \in (0, \frac{1}{2})$ and $\varepsilon_1, \varepsilon_2 > 0$ be fixed. Let V be a set of $(1 - \delta)n$ vertices, and W be a set of $\frac{\varepsilon_1 n}{\log n}$ vertices disjoint from V . Suppose we are given disjoint subsets C_1, C_2, \dots, C_r of $V^{(2)}$ and a subset $S \subseteq W^{(2)}$ with the following properties:*

- (1) $|S| \geq \frac{(\varepsilon_1)^2 n^2}{8(\log n)^2}$;
- (2) $|C_i| \geq M$ for every $i: 1 \leq i \leq r$, and some M satisfying: $(\log n)^4 \leq M \leq \frac{\varepsilon_2}{4} n^2$;
- (3) $\sum_i |C_i| \geq \varepsilon_2 n^2$.

Let $p = p(n)$ be an edge probability with

$$\lambda \frac{1}{\sqrt{n}} < p(n) < 5 \frac{\sqrt{\log n}}{\sqrt{n}}.$$

Consider the p -random bipartite graph $B_p(V, W)$ with bipartition $V \sqcup W$. Let *Boost* be the event that for every C_i with $|C_i| \leq 2M$ there exists $C_j \neq C_i$ and a triple $(x_1x_2, y_1y_2, z_1z_2) \in C_i \times C_j \times S$ such that the restriction $B_p(\{x_1, x_2, y_1, y_2\}, \{z_1, z_2\})$ of $B_p(V, W)$ to $\{x_1, x_2, y_1, y_2\} \sqcup \{z_1, z_2\}$ is complete. Then for all n sufficiently large we have

$$\mathbb{P}(\text{Boost}) \geq 1 - \exp\left(-\frac{\varepsilon_2(\varepsilon_1)^2}{2^{17}}(\log n)^2\right).$$

The proof of the connecting lemma relies on a celebrated inequality of Janson and some careful book-keeping.

Proposition 6.5 (The extended Janson inequality [19]). *Let U be a finite set and U_q a q -random subset of U for some $q \in [0, 1]$. Let \mathcal{F} be a family of subsets of U , and for every $F \in \mathcal{F}$ let I_F be the indicator function of the event $\{F \subseteq U_q\}$. Set $I_{\mathcal{F}} = \sum_{F \in \mathcal{F}} I_F$, and let $\mu = \mathbb{E}I_{\mathcal{F}}$ and $\Delta = \sum_{F, F' \in \mathcal{F}: F \cap F' \neq \emptyset} \mathbb{E}(I_F I_{F'})$. Then*

$$\mathbb{P}(I_{\mathcal{F}} = 0) \leq \exp\left(-\frac{\mu^2}{2\Delta}\right).$$

Proof of Lemma 6.4. Fix C_i with $M \leq |C_i| \leq 2M$. Set $M' = \min(2M, n)$. Let \mathcal{F}_0 denote the collection of *connecting triples* $(x_1x_2, y_1y_2, z_1z_2) \in C_i \times \bigcup_{j \neq i} C_j \times S$. Further let

$$\mathcal{F} = \{\{x_iz_j : i, j \in [2]\} \cup \{y_iz_j : i, j \in [2]\} : (x_1x_2, y_1y_2, z_1z_2) \in \mathcal{F}_0\}.$$

Observe that the elements of \mathcal{F} are subsets of either 6 or 8 edges (depending on whether the pairs x_1x_2 and y_1y_2 overlap or not) of the complete bipartite graph $B(V, W)$ with bipartition $V \sqcup W$. We shall apply Janson's inequality to \mathcal{F} to give an upper bound on the probability that C_i does not connect to $\bigcup_{j \neq i} C_j$ via a pair of squares of $B_p(V, W)$. To this end, we must compute and bound the μ and Δ parameters for \mathcal{F} . The first of these is straightforward:

$$(6.6) \quad \mu := \mathbb{E}I_{\mathcal{F}} = \sum_{F \in \mathcal{F}} \mathbb{E}I_F \geq |C_i| \cdot \left| \bigcup_{j \neq i} C_j \right| \cdot |S| p^8 \geq M \frac{\varepsilon_2(\varepsilon_1)^2}{16} \frac{n^4}{(\log n)^2} p^8$$

To bound the Δ parameter, fix a connecting triple $t = (x_1x_2, y_1y_2, z_1z_2) \in C_i \times \bigcup_{j \neq i} C_j \times S$, and consider the contribution to Δ made by pairs (t, t') of connecting triples that share at least one edge of $B(V, W)$; call such pairs of connecting triples *dependent*.

Write $L(t)$ for the set $\{x_1, x_2, y_1, y_2\}$ (which can have size either 4 or 3 — the latter if one of the x_i is equal to one of the y_j) and $R(t)$ for the pair $\{z_1, z_2\}$. Also let $F_t \in \mathcal{F}$ be the collection of edges of $B(V, W)$ from $L(t)$ to $R(t)$. Clearly if $L(t) \cap L(t') = \emptyset$ or $R(t) \cap R(t') = \emptyset$, then (t, t') do not form a pair of dependent connecting triples.

For $(i, j) \in [4] \times [2]$, let $D_{i,j}(t)$ denote the collection of connecting triples t' with $|L(t) \cap L(t')| = i$ and $|R(t) \cap R(t')| = j$. Further let $D_{i,j}^a(t)$ and $D_{i,j}^b(t)$ denote the collection of t' in $D_{i,j}(t)$ with $|L(t)| = 4$ and $|L(t')| = 3$ respectively. We shall bound the sizes of the sets $D_{i,j}^a(t)$ and $D_{i,j}^b(t)$. Note to begin with that there are at most $\frac{2\varepsilon_1 n}{\log n}$ ways of deleting a vertex in $R(t)$ and replacing it by a different vertex in W . In particular, for all connecting triples t and all $i \in [|L(t)|]$, we have

$$|D_{i,1}^a(t)| \leq |D_{i,2}^a(t)| \cdot \frac{2\varepsilon_1 n}{\log n} \quad \text{and} \quad |D_{i,1}^b(t)| \leq |D_{i,2}^b(t)| \cdot \frac{2\varepsilon_1 n}{\log n}.$$

Thus we may focus on bounding the sizes of $D_{i,j}^a(t)$ and $D_{i,j}^b(t)$ in the case where $j = 2$.

Case 1, $|L(t)| = 4$:

- there are at most 6 ways of splitting $L(t)$ into a pair from C_i and a pair from $\bigcup_{j \neq i} C_j$, and at most 24 ways of deleting a vertex from $L(t)$ and viewing the remaining 3 vertices as the union of a pair from C_i and an (overlapping) pair from $\bigcup_{j \neq i} C_j$, whence $|D_{4,2}^a(t)| \leq 6$ and $|D_{3,2}^b(t)| \leq 24$;
- there are at most $4n$ ways of deleting one vertex from $L(t)$ and replacing it by another vertex from V . As noted above, there are most 6 ways of splitting the resulting 4-set into a pair from C_i and a pair from $\bigcup_{j \neq i} C_j$, whence $|D_{3,2}^a(t)| \leq 24n$;
- there are at most $6(n^2/2) = 3n^2$ ways of deleting two vertices from $L(t)$ and replacing them by two other vertices from V , whence (similarly to the above) we have $|D_{2,2}^a(t)| \leq 18n^2$; further, there are at most $6n$ ways of deleting a pair of vertices from V and replacing them by a single vertex from V , whence (similarly to the above, since there

are most 6 ways of viewing three vertices of a pair from C_i and an (overlapping) pair from $\bigcup_{j \neq i} C_j$ we get $|D_{2,2}^b(t)| \leq 36n$;

- since C_i contains at most $2M$ pairs, there are at most $4(n^2/2)M' = 2n^2M'$ ways of deleting three vertices in $L(t)$ and replacing them by another three vertices from V in such a way that the resulting set can still be viewed as the union of a pair from C_i and a pair from $\bigcup_{j \neq i} C_j$, whence (similarly to the above) we have $|D_{1,2}^a(t)| \leq 12n^2M'$; further and similarly there are at most $4M + 4M'n \leq 8M'n$ ways of deleting three vertices in $L(t)$ and replacing them by a pair of vertices from V , whence (as before) we have $|D_{1,2}^b(t)| \leq 48M'n$;

Case 2, $|L(t)| = 3$:

- there are at most 6 ways of splitting $L(t)$ into a pair from C_i and a pair from $\bigcup_{j \neq i} C_j$, whence $|D_{3,2}^b(t)| \leq 6$; further there are at most $6n$ ways of adding a vertex to $L(t)$ and splitting the resulting 4-set into two disjoint pairs, whence $|D_{3,2}^a(t)| \leq 6n$;
- there are at most $3n$ ways of deleting one vertex from $L(t)$ and replacing it by another vertex from V , whence (as above) $|D_{2,2}^b(t)| \leq 18n$; further, there are at most $3(n^2/2)$ ways of deleting one vertex from $L(t)$ and replacing it by a pair from V , whence (as above) $|D_{2,2}^a(t)| \leq 9n^2$;
- since C_i contains at most $2M$ pairs, there are at most $3M'n$ ways of deleting two vertices in $L(t)$ and replacing them by another two vertices from V in such a way that the resulting set can still be viewed as the union of a pair from C_i and a pair from $\bigcup_{j \neq i} C_j$, whence $|D_{1,2}^b(t)| \leq 18M'n$; similarly, there are at most $3M'(n^2/2)$ ways of deleting two vertices in $L(t)$ and replacing them by a triple of vertices from V in such a way that the resulting set can be viewed as the disjoint union of a pair from C_i and a pair from $\bigcup_{j \neq i} C_j$, whence $|D_{1,2}^a(t)| \leq 9M'n^2$.

Given $t' \in D_{i,j}^a(t)$ and considering the edges between $L(t) \cup L(t')$ and $R(t) \cup R(t')$, we see that

$$(6.7) \quad \mathbb{E}I_{F_t}I_{F_{t'}} = \mathbb{E}I_{F_t}p^{2(4-i)+i(2-j)} = \mathbb{E}I_{F_t}p^{8-ij}.$$

Similarly, for $t' \in D_{i,j}^b(t)$ we have

$$(6.8) \quad \mathbb{E}I_{F_t}I_{F_{t'}} = \mathbb{E}I_{F_t}p^{2(3-i)+i(2-j)} = \mathbb{E}I_{F_t}p^{6-ij}.$$

With the bounds on the size of $D_{i,j}^a(t)$ and $D_{i,j}^b(t)$ derived above and equalities (6.7) and (6.8) in hand, we are now ready to bound the contribution to Δ from a connecting triple t .

Case 1: $|L(t)| = 4$.

$$(6.9) \quad \begin{aligned} & \sum \left\{ \mathbb{E}I_{F_t}I_{F_{t'}} : (t, t') \text{ form a pair of dependent connecting triples} \right\} \\ &= \mathbb{E}I_{F_t} \left(\sum_{i=1}^4 \left(|D_{i,2}^a| p^{8-2i} + |D_{i,2}^b| p^{6-2i} \right) + \sum_{i=1}^4 \left(|D_{i,1}^a| p^{8-i} + |D_{i,1}^b| p^{6-i} \right) \right) \\ &\leq \mathbb{E}I_{F_t} \left(((12n^2M'p^6 + 48M'np^4) + (18n^2p^4 + 36np^2) + (24np^2 + 24) + (6)) \right. \\ &\quad \left. + \frac{2\varepsilon_1 n}{\log n} ((12n^2M'p^7 + 48M'np^5) + (18n^2p^6 + 36np^4) + (24np^5 + 24p^3) + (6p^4)) \right) \\ &\leq (\mathbb{E}I_{F_t}) 2^{10} (np^2)^3 \max \left(1, \frac{\varepsilon_1}{\log n} M'p \right). \end{aligned}$$

(Note in the last line we use the fact that $np^2 \geq (\lambda_c)^2 > 1/4$, whence $(np)^{-1} \leq 4p$.)

Case 2: $|L(t)| = 3$.

$$\begin{aligned}
 & \sum \left\{ \mathbb{E} I_{F_t} I_{F_{t'}} : (t, t') \text{ form a pair of dependent connecting triples} \right\} \\
 &= \mathbb{E} I_{F_t} \left(\sum_{i=1}^3 \left(|D_{i,2}^a| p^{8-2i} + |D_{i,2}^b| p^{6-2i} \right) + \sum_{i=1}^3 \left(|D_{i,1}^a| p^{8-i} + |D_{i,1}^b| p^{6-i} \right) \right) \\
 &\leq \mathbb{E} I_{F_t} \left((9n^2 M' p^6 + 18n M' p^4) + (9n^2 p^4 + 18n p^2) + (6n p^2 + 6) \right) \\
 &\quad + \frac{2\varepsilon_1 n}{\log n} \left((9n^2 M' p^7 + 18n M' p^5) + (9n^2 p^6 + 18n p^4) + (6n p^5 + 6p^3) \right) \\
 (6.10) \quad &\leq (\mathbb{E} I_{F_t}) 2^{10} (np^2)^3 \max \left(1, \frac{\varepsilon_1}{\log n} M' p \right).
 \end{aligned}$$

Together, inequalities (6.9) and (6.10) yield that

$$(6.11) \quad \Delta \leq \frac{1}{2} \mu \cdot 2^{10} (np^2)^3 \max \left(1, \frac{\varepsilon_1}{\log n} M' p \right).$$

Applying the Extended Janson Inequality, Proposition 6.5, together with the bounds (6.6) and (6.11) on μ and Δ , we get:

$$\begin{aligned}
 \mathbb{P}(I_{\mathcal{F}} = 0) &\leq \exp \left(-\frac{\mu^2}{2\Delta} \right) \leq \exp \left(-\frac{\mu}{2^{10} (np^2)^3 \max \left(1, \frac{\varepsilon_1}{\log n} M' p \right)} \right) \\
 &\leq \exp \left(-\frac{\varepsilon_2 (\varepsilon_1)^2}{2^{14}} \frac{M (np^2)}{(\log n)^2 \max \left(1, \frac{\varepsilon_1}{\log n} M' p \right)} \right) \\
 (6.12) \quad &\leq \begin{cases} \exp \left(-\frac{\varepsilon_2 (\varepsilon_1)^2}{2^{16}} \frac{M}{(\log n)^2} \right) & \text{if } 2M \leq \frac{p^{-1} \log n}{\varepsilon_1}, \\ \exp \left(-\frac{\varepsilon_2 \varepsilon_1}{2^{16}} \frac{p^{-1}}{\log n} \right) & \text{if } \frac{p^{-1} \log n}{\varepsilon_1} \leq 2M \leq n, \\ \exp \left(-\frac{\varepsilon_2 \varepsilon_1}{2^{16}} \frac{p^{-1} M}{n \log n} \right) & \text{if } n \leq 2M. \end{cases}
 \end{aligned}$$

Now, the probability that C_i fails to connect to $\cup \{C_j : j \neq i\}$ via a connecting triple is exactly the probability that $I_{\mathcal{F}} = 0$. Applying Markov's inequality together with (6.12) and using our assumptions that $M \geq (\log n)^4$ and that C_1, \dots, C_r are disjoint, we have for n sufficiently large that

$$\mathbb{P}(\text{Boost}) \geq 1 - r \mathbb{P}(I_{\mathcal{F}} = 0) \geq 1 - \frac{n^2}{M} \mathbb{P}(I_{\mathcal{F}} = 0) \geq 1 - \exp \left(-\frac{\varepsilon_2 (\varepsilon_1)^2}{2^{17}} (\log n)^2 \right),$$

provided n is sufficiently large. This concludes the proof of the connecting lemma. \square

6.4. Sprinkling vertices. With Lemma 6.4 in hand, we can return to the proof of Theorem 1.5. To complete the proof, we shall use a multiple-round vertex-sprinkling argument. We partition $\Gamma \in \mathcal{G}(n, p)$ into the union of

- (i) $\Gamma_1 \in \mathcal{G}((1 - 3\varepsilon_1)n, p)$ on $V_1 = [(1 - 3\varepsilon_1)n]$,
- (ii) $\Gamma_2 \in \mathcal{G}(3\varepsilon_1 n, p)$ on $V_2 = [n] \setminus V_1$, and
- (iii) a p -random bipartite graph $B = B_p(V_1, V_2)$ with bipartition $V_1 \sqcup V_2$.

We further partition V_2 into $3 \log n > 2 \log_2 n$ sets of size $\frac{\varepsilon_1 n}{\log n}$, $V_2 = \sqcup_{i=1}^{3 \log n} V_{2,i}$ (and ignore rounding errors). We say that Γ_1 is a *good configuration* if it satisfies the conclusion of Corollary 6.3, i.e., if at least $\varepsilon_2 n^2$ non-edges of Γ_1 lie in square-components in $\square(\Gamma_1)$ of order at least $M_0 := (\log n)^4$ (this is actually slightly weaker than what Corollary 6.3 gives us, but is all we need here).

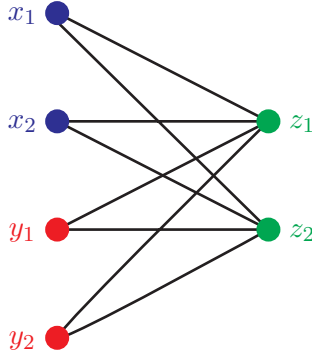


FIGURE 3. In this connecting triple, both $x_1z_1x_2z_2$ and $y_1z_1y_2z_2$ form induced copies of C_4 , joining up the square-components containing the non-edges x_1x_2 and y_1y_2 via the non-edge of sprinkled vertices z_1z_2 .

We shall condition on Γ_1 being a good configuration when we perform our vertex-sprinkling. By Corollary 6.3, this occurs with probability $1 - O(n^{-1})$. A key observation is that the state of the edges in Γ_2 and B are independent of our conditioning. Our strategy is then to reveal the $3 \log n$ sprinkling sets $V_{2,k}$ one by one, and use them to create bridges between “somewhat large” square-components and thereby increase the minimum order of all “somewhat large” square-components.

More precisely, before stage $k \geq 1$ we have revealed all the edges inside

$$V_{1,k-1} := V_1 \cup \left(\bigcup_{i=1}^{k-1} V_{2,i} \right).$$

At this stage, we deem a square-component “large” if it contains at least M_{k-1} non-edges of Γ , and “very large” if it contains at least $2M_{k-1}$ non-edges of Γ (which constitute, as we recall, the vertices of the square-graph). Now in stage k , we reveal the set S_k of non-edges of Γ that lie inside $V_{2,k}$ and the edges between $V_{1,k}$ and $V_{2,k}$. We then merge components as follows: given two square-components C_i and C_j , a *connecting triple* is a triple $(x_1x_2, y_1y_2, z_1z_2) \in C_i \times C_j \times S_k$. Such a connecting triple is *active* if all edges between the sets $\{x_1, x_2, y_1, y_2\}$ and $\{z_1, z_2\}$ are in Γ ; in this case the components C_i and C_j lie inside the same square-component C in $\square(\Gamma[V_{1,k}])$ (see Figure 3). In particular, if both C_i and C_j contained at least M_{k-1} non-edges, then C must contain at least $M_k = 2M_{k-1}$ non-edges.

The connecting lemma we proved in the previous subsection immediately implies that with high probability at each stage k , all components which are “large” but not “very large” must join up with at least one other “large” component. We make this explicit with a lemma below. Recall that throughout this section, $\lambda > \lambda_c$ is fixed and the edge-probability $p = p(n)$ satisfies $\lambda \frac{1}{\sqrt{n}} < p(n) < 5 \frac{\sqrt{\log n}}{\sqrt{n}}$. Let $\varepsilon_1, \varepsilon_2 > 0$ be the constants whose existence is guaranteed by Corollary 6.3.

Lemma 6.6 (Sprinkling lemma). *Suppose that before stage k , at least $\varepsilon_2 n^2$ non-edges of $\Gamma[V_{1,k-1}]$ lie in square-components of order at least $M_{k-1} = 2^{k-1} M_0$ in $\square(\Gamma[V_{1,k-1}])$. Suppose $M_{k-1} \leq \frac{\varepsilon_2}{4} n^2$. Then with probability at least*

$$1 - 2 \exp \left(- \frac{\varepsilon_2 (\varepsilon_1)^2}{2^{17}} (\log n)^2 \right)$$

when we have revealed the edges from $V_{2,k}$ to $V_{1,k-1} \cup V_{2,k} := V_{1,k}$ at least $\varepsilon_2 n^2$ non-edges of $\Gamma[V_{1,k}]$ lie in square-components of order at least $M_k = 2M_{k-1}$ in $\square(\Gamma[V_{1,k}])$.

In particular, with probability at least

$$1 - 6 \log(n) \exp\left(-\frac{\varepsilon_2(\varepsilon_1)^2}{2^{17}}(\log n)^2\right)$$

we have that starting from a good configuration $\Gamma[V_1]$, the sprinkling process described above discovers a giant square-component containing at least $\varepsilon_2 n^2$ non-edges, and all non-edges from $\Gamma[V_1]$ that lie inside components of $\square(\Gamma[V_1])$ of size at least $(\log n)^4$.

Proof. Let S_k denote the set of non-edges of $V_{2,k}$. We have

$$\mathbb{E}|S_k| = (1-p) \binom{\frac{\varepsilon_1}{\log n} n}{2} = (1-o(1)) \frac{(\varepsilon_1)^2}{(2 \log n)^2} n^2.$$

By a standard Chernoff bound,

$$\mathbb{P}\left(|S_k| \leq \frac{(\varepsilon_1)^2 n^2}{8(\log n)^2}\right) \leq \exp\left(-\frac{(\varepsilon_1)^2}{16} \frac{n^2}{(\log n)^2}\right)$$

If $|S_k| \geq \frac{\varepsilon_1 n^2}{8(\log n)^2}$ holds, we can apply Lemma 6.4, concluding that every component of size M_{k-1} is joined with at least one other, resulting in a component of size $M_k = 2M_{k-1}$. Thus the desired conclusion for the first part of the lemma holds with probability at least

$$1 - \exp\left(-\frac{(\varepsilon_1)^2}{16} \frac{n^2}{(\log n)^2}\right) - \exp\left(-\frac{\varepsilon_2(\varepsilon_1)^2}{2^{17}}(\log n)^2\right) \geq 1 - 2 \exp\left(-\frac{\varepsilon_2(\varepsilon_1)^2}{2^{17}}(\log n)^2\right),$$

as desired.

For the ‘‘in particular’’ part, we first apply a simple union bound to the first $2 \log_2 n - 4 \log_2 \log n < 3 \log n$ steps of the process, to show that with probability at least

$$(6.13) \quad 1 - 2(2 \log_2(n) - 4 \log_2(\log n)) \exp\left(-\frac{\varepsilon_2(\varepsilon_1)^2}{2^{17}}(\log n)^2\right)$$

our sprinkling process has uncovered a collection of square-components, each of which contains at least $\frac{\varepsilon_2 n^2}{2}$ non-edges, and whose union contains at least $\varepsilon_2 n^2$ non-edges and includes all non-edges of $\Gamma[V_1]$ coming from components of $\square(\Gamma[V_1])$ of size at least $(\log n)^4$. There can be at most $\frac{1}{2}(\varepsilon_2)^{-1}$ such components. By (6.12), the probability that a fixed pair of such components fails to join up in the next round of sprinkling is at most

$$\exp\left(-\frac{(\varepsilon_2)^2 \varepsilon_1 p^{-1} n}{2^{16} \log n}\right) \leq \exp\left(-\frac{(\varepsilon_2)^2 \varepsilon_1}{2^{16}} \frac{n^{\frac{3}{2}}}{5(\log n)^{\frac{3}{2}}}\right).$$

Taking the union bound over the at most $\frac{1}{8}(\varepsilon_2)^{-2}$ pairs of components, we have that the probability any pair of these components fail to join up is, for large n , a lot less than the last term in equation (6.13):

$$8 \log_2 \log(n) \exp\left(-\frac{\varepsilon_2(\varepsilon_1)^2}{2^{17}}(\log n)^2\right).$$

Combining this with (6.13), and the inequality $2 \log_2(n) < 3 \log(n)$, we get the claimed bound on the probability of having discovered a giant square-component containing at least $\varepsilon_2 n^2$ non-edges and all non-edges of $\Gamma[V_1]$ contained in components of $\square(\Gamma[V_1])$ of size at least $(\log n)^4$. \square

6.5. Covering the whole world. All that now remains to complete the proof of Theorem 1.5 is to show that a.a.s. there is a square-component covering all vertices of $\square(\Gamma)$. A natural way of doing this would be to subdivide $V(\Gamma)$ into $O(\eta^{-1})$ pieces of order ηn , for some suitable small $\eta > 0$, and to show that a.a.s. in each piece U all vertices connect up to the giant in $\Gamma[V(\Gamma) \setminus U]$. This is essentially what we shall do, with a slight technical twist: while in Lemma 6.6 we showed that $\square(\Gamma)$ has a giant component, we have not quite shown it is unique. In principle, one could stitch together a rival giant component at the last stage of sprinkling by building numerous bridges between small components. Then the natural approach described above could fail to ensure that vertices in a piece U connect to the “right” giant, and that we have a square-component with full support.

This is of course highly unlikely, and one could show uniqueness of the giant by exploiting the fact that the number of non-edges of Γ lying in square-components of order at least $(\log n)^4$ is a.a.s. concentrated around its expectation (as shown in Corollary 6.3). However we do not have a nice form for this expectation, so a little care would be needed to show it changes continuously with n to make the argument above fully rigorous. As this paper is already sufficiently long and as the uniqueness of the giant is not our main concern here, we eschew this and focus instead on the problem of ensuring we have a giant component whose support covers all the vertices. We sidestep the issue of the uniqueness of the giant by considering a partition of $[n]$ which allows us to both build a preferred giant and, crucially, to ensure this preferred giant has full support. We begin by establishing a useful corollary of the work in the previous subsections.

Corollary 6.7. *Let $\lambda > \lambda_c$ be fixed, and let $p = p(n)$ be an edge-probability satisfying $\lambda n^{-1/2} \leq p \leq 5n^{-1/2}\sqrt{\log n}$. Let $\Gamma \in \mathcal{G}(n, p)$. Then for every $\varepsilon_3 > 0$ sufficiently small, there exists a constant $\varepsilon_4 > 0$ such that given fixed sets $U \subseteq U' \subseteq [n]$ with $|U| = \lfloor (1-2\varepsilon_3)n \rfloor$, $|U'| = \lfloor (1-\varepsilon_3)n \rfloor$ all of the following hold with probability $1 - O(n^{-1})$:*

- (i) *there are at least $\varepsilon_4 n^2$ non-edges in $\Gamma[U]$ contained in square-components of $\square(\Gamma[U])$ of order at least $(\log n)^4$;*
- (ii) *there is a unique square-component in $\square(\Gamma[U'])$ containing all non-edges in $\Gamma[U]$ contained in square-components of $\square(\Gamma[U])$ of order at least $(\log n)^4$;*
- (iii) *there is a unique square-component in $\square(\Gamma)$ containing all non-edges contained in square-components of $\square(\Gamma[U])$ or $\square(\Gamma[U'])$ of order at least $(\log n)^4$. \square*

Proof. The corollary is immediate for sufficiently small $\varepsilon_3 > 0$ from an application of Corollary 6.3 inside U (for part (i)) and two applications of Lemma 6.6 (for parts (ii) and (iii) respectively), together with a suitable choice of the constant ε_4 . \square

As an immediate corollary of Corollary 6.7, Lemma 6.2 and Theorem 5.1, we have the following result, confirming the conjecture of Bollobás and Riordan on the location of the phase transition for non-induced square percolation.

Corollary 6.8 (Phase transition for non-induced square percolation⁴). *Let λ be fixed.*

- (i) *If $\lambda < \lambda_c$ and $p(n) \leq \lambda n^{-1/2}$, then for $\Gamma \in \mathcal{G}_{n,p}$, a.a.s. every component of $\boxtimes(\Gamma)$ has order $O\left((\log n)^{2^{32}}\right)$.*
- (ii) *On the hand, if $\lambda > \lambda_c$ and $p(n)$ satisfies $\lambda n^{-1/2} \leq p(n) \leq 5\sqrt{\log(n)}n^{-1/2}$, then for $\Gamma \in \mathcal{G}_{n,p}$, a.a.s. there exists a giant component in $\boxtimes(\Gamma)$ of order $\Omega(n^4 p^4)$.*

Proof. Part (i) is immediate from the proof of Theorem 5.1 — indeed, it is what we actually establish. Part (ii) follows from Lemma 6.2 and Corollary 6.7. \square

⁴Note that by a standard second-moment argument, for $p \gg n^{-1}$ the number of (non-induced) squares in Γ is a.a.s. of order $p^4 n^4$, so part (ii) of the corollary is saying that for $\lambda n^{-1/2} \leq p(n) \leq 5\sqrt{\log(n)}n^{-1/2}$ the largest component in $\boxtimes(\Gamma)$ contains a $\Omega(1)$ proportion of the vertices in $\boxtimes(\Gamma)$, while part (i) is saying that for $n^{-1+\eta} \leq \lambda n^{-1/2}$ and $\eta > 0$ fixed, all components in $\boxtimes(\Gamma)$ contain a $o(1)$ proportion of vertices

We now return to the main business of this subsection and apply Corollary 6.7 to prove Theorem 1.5.

Proof of Theorem 1.5. First note that if $f(n)$ is any function with $f(n) = o(1)$ and $f(n) = \Omega(n^{-2})$, and $5n^{-1/2}\sqrt{\log n} \leq p(n) \leq 1 - f(n)$, then $\Gamma \in \mathcal{G}(n, p)$ has the \mathcal{CFS} property, by [6, Theorem 5.1]. Now assume that $\lambda > \lambda_c$ and $\lambda n^{-1/2} \leq p(n) \leq 5n^{-1/2}\sqrt{\log n}$. Pick $\varepsilon_3 > 0$ sufficiently small, and let $\varepsilon_4 > 0$ be the constant whose existence is guaranteed by Corollary 6.7. Partition $[n]$ into $K = \lceil (\varepsilon_3)^{-1} \rceil$ sets

$$U_i = \{x \in [n] : (i-1)\varepsilon_3 n < x \leq i\varepsilon_3 n\}.$$

For each pair (i, j) of distinct elements of $[K]$, we apply Corollary 6.7 to the sets $U = [n] \setminus (U_i \cup U_j)$ and $U' = [n] \setminus U_i$; taking a union bound over all such pairs (i, j) , we see that with probability $1 - O(K^2 n^{-1}) = 1 - O(n^{-1})$, for every pair of distinct elements $i, j \in [K]$ the following hold:

- (1) at least $\varepsilon_4 n^2$ non-edges in $\Gamma[[n] \setminus (U_i \cup U_j)]$ are contained in components of $\square(\Gamma[[n] \setminus (U_i \cup U_j)])$ of order at least $(\log n)^4$;
- (2) there is a unique component C'_{ij} of $\square(\Gamma[[n] \setminus U_i])$ containing all non-edges of $\Gamma[[n] \setminus (U_i \cup U_j)]$ contained in components of $\square(\Gamma[[n] \setminus (U_i \cup U_j)])$ of order at least $(\log n)^4$;
- (3) there is a unique component C_i of $\square(\Gamma)$ containing C'_{ij} as well as all non-edges of $\Gamma[[n] \setminus U_i]$ contained in components of $\square(\Gamma[[n] \setminus U_i])$ of order at least $(\log n)^4$.

We claim that $\forall i, j \in [K]$ we have $C_i = C_j$. Indeed for $i \neq j$, note that $C_i \supseteq C'_{ij}$ and $C_j \supseteq C'_{ji}$. Since both C'_{ij} and C'_{ji} contain all of the at least $\varepsilon_4 n^2$ non-edges contained in components of $\square(\Gamma[[n] \setminus (U_i \cup U_j)])$ of order at least $(\log n)^4$, it follows that $(C_i \cap C_j) \supseteq (C'_{ij} \cap C'_{ji}) \neq \emptyset$. Since their intersection is non-empty, C_i and C_j are the same component of $\square(\Gamma)$, as claimed. We may thus let C_\star denote the a.a.s. unique square-component with $C_\star = C_i$ for all $i \in [K]$.

We now show that a.a.s. the support of this component C_\star is the whole vertex set $[n]$. Pick $i \in [K]$ and condition on the event that there is a square-component C'_i in $\square(\Gamma[[n] \setminus U_i])$ of order at least $\varepsilon_4 n^2$ (an event which occurs with probability $1 - O(n^{-1})$, as we saw in (2) above). If two or more such components exist, pick a largest one. Further, condition on each vertex $x \in U_i$ having at least $\frac{\varepsilon_3}{2}n$ non-neighbors in $\Gamma[U_i]$. By a standard application of Chernoff bounds and a union-bound, this event occurs with probability $1 - O(n^{-1})$.

Having thus conditioned on the state of pairs in $\Gamma[U_i]$ and $\Gamma[[n] \setminus U_i]$, we now show that a.a.s. for every vertex $x \in U_i$, there exist $y \in U_i$ and $uv \in C'_i$ such that $xyuv$ induces a copy of C_4 — so that that xy belongs to the component C' of $\square(\Gamma)$ containing C'_i . Combining this with (3) above (which implies that a.a.s. $C' = C_\star$) and a simple union bound will then yield Theorem 1.5.

Given non-edges $xy \in U_i^{(2)} \setminus E(\Gamma[U_i])$ and $uv \in C'_i$, let $X_{xy,uv}$ be the indicator function of the event that all of ux, uy, vx and vy are edges of Γ . Observe that this event is independent of our conditioning. For $x \in U_i$, set

$$X_x = \sum_{y \in U_i: xy \notin E[U_i]} \sum_{uv \in C'_i} X_{xy,uv}$$

to be the number of induced C_4 's $xyuv$ of Γ with $y \in U_i$ and $uv \in C'_i$. We shall again apply the Extended Janson Inequality (Proposition 6.5) to bound $\mathbb{P}(X_x = 0)$. Given our conditioning, the expectation of X_x is

$$\mu := |\{y \in U_i : xy \notin E(\Gamma)\}| \cdot |C'_i| p^4 \geq \frac{\varepsilon_3 \varepsilon_4 (\lambda_c)^4}{2} n = \Omega(n).$$

We now compute the corresponding Δ -parameter in Janson's inequality. For $y, y' \in U_i \setminus \Gamma_x[U_i]$ and $uv, u'v' \in C'_i$, write $\{xy, uv\} \sim \{xy', u'v'\}$ if $\{ux, uy, vx, vy\} \cap \{u'x, u'y', v'x, v'y\} \neq \emptyset$. Note

that the random variables $X_{xy,uv}$ and $X_{xy',u'v'}$ are independent unless $\{xy, uv\} \sim \{xy', u'v'\}$, and further that $\{xy, uv\} \sim \{xy', u'v'\}$ implies $\{u, v\} \cap \{u', v'\} \neq \emptyset$.

Pick and fix $y \in U_i \setminus \Gamma_x[U_i]$ and $uv \in C'_i$, which can be done in at most $|U_i| \cdot |C'_i|$ ways. We perform a case analysis to determine the contributions to the Δ -parameter from terms of the form $\mathbb{E}X_{xy,uv}X_{xy',u'v'}$ with $\{xy, uv\} \sim \{xy', u'v'\}$:

- there are at most $|U_i|$ choices of $y' \in U_i \setminus \Gamma_x[U_i]$ with $y' \neq y$ such that $\{xy, uv\} \sim \{xy', uv\}$, and for such y' we have $\mathbb{E}[X_{xy,uv}X_{xy',uv}] = p^6$;
- there are at most n choices of $v' \in [n] \setminus U_i$ with $v' \neq v$ such that $\{xy, uv\} \sim \{xy, uv'\}$, and for such v' we have $\mathbb{E}[X_{xy,uv}X_{xy,uv'}] = p^6$ (with the contribution from the symmetric cases $\{xy, uv\} \sim \{xy, u'v\}$ analysed similarly);
- finally, there are at most $n|U_i|$ choices of (v', y') with $v' \in [n] \setminus U_i$, $y' \in U_i \setminus \Gamma_x[U_i]$, $v' \neq v$ and $y \neq y'$ such that $\{xy, uv\} \sim \{xy', uv'\}$, and for such (v', y') we have $\mathbb{E}[X_{xy,uv}X_{xy',uv'}] = p^7$ (with the contribution from the symmetric cases $\{xy, uv\} \sim \{xy', u'v\}$ analysed similarly).

Putting it all together, we see

$$\Delta = \sum_{\{xy, uv\} \sim \{xy', u'v'\}} \mathbb{E}[X_{xy,uv}X_{xy',uv}] \leq |U_i| \cdot |C'_i| \left(|U_i|p^6 + 2np^6 + 2n|U_i|p^7 \right).$$

Since $|U_i| \leq \varepsilon_3 n$, $|C'_i| \leq n^2/2$ and $p \leq 5\sqrt{\frac{\log n}{n}}$, the above implies

$$\Delta \leq 5^7 (\varepsilon_3)^2 n \sqrt{n} (\log n)^{7/2} (1 + o(1)),$$

which for n large enough is at most $\frac{2^{20}\varepsilon_3}{\varepsilon_4(\lambda_c)^4} \mu n^{1/2} (\log n)^{7/2}$. Applying Janson's inequality,

$$\mathbb{P}(X_x = 0) \leq \exp\left(-\frac{\mu^2}{2\Delta}\right) \leq \exp\left(-\frac{\varepsilon_4 \lambda_c^4 \mu}{2^{20}(\log n)^{3/2} \sqrt{n}}\right) = o(n^{-1}).$$

Thus with probability $1 - o(n^{-1})$, $X_x > 0$ and the vertex x is covered by the square-component in $\square(\Gamma)$ that contains C'_i . Taking a union bound over $x \in U_i$, with probability $1 - o(1)$, every vertex $x \in U_i$ is covered by this square-component (which as we showed is the a.a.s uniquely determined giant component C_\star). Taking a union bound over $i \in [K]$ and combining this with **(1)**–**(3)** above, we see that with probability $1 - o(1)$ the square-component C_\star covers all of $[n]$. This concludes the proof of Theorem 1.5. \square

7. CONCLUDING REMARKS

There are several natural questions arising from our work. To begin with, one could ask for more information about the component structure in $\square(\Gamma)$: in the subcritical regime, can one get a better upper bound on the order of square-components? In the supercritical regime, can one give good bounds on the order of the second-largest square-component? In particular, can one give better bounds than just $o(n^2)$, and can one show its support has size $o(n)$? This may be feasible albeit technically challenging.

Another question on the probabilistic side is determining the range of $p = p(n)$ for which the square graph $\square(\Gamma)$ of $\Gamma \in \mathcal{G}(n, p)$ is a.a.s. connected, a very different question from the ones considered in this paper. Investigating other parameters such as the diameter of $\square(\Gamma)$ may also lead in interesting directions.

Further afield, one could consider percolation problems for similar structures. We could for example consider the graph on all triples of independent vertices in Γ obtained by setting two such triples to be adjacent if they induce a copy of the 6-cycle C_6 or of the complete balanced bipartite graph $K_{3,3}$ in Γ . We would guess the techniques in this paper would adapt well to the

latter problem, as Bollobás and Riordan also suggest, but that new ideas would be needed for the former.

One could also consider a similar problem for a p -random r -uniform hypergraph Γ , by setting a set of r vertex-disjoint non-edges $F_1, F_2, \dots, F_r \notin E(\Gamma)$ to be adjacent if all of the r^2 edges meeting each F_i in exactly one vertex are present in Γ (in other words, if and only if the union of the F_i 's induces a copy of the complete balanced r -partite r -uniform hypergraph on r^2 vertices). Note the case $r = 2$ corresponds exactly to square percolation. Levcovitz [22] has provided a quasi-isometry invariant for right-angled Coxeter groups by associating a hypergraph to any such group, so analysis of suitable variants of square percolation for hypergraphs may yield interesting applications in geometric group theory (besides constituting a challenging and rather natural problem in combinatorial probability).

Finally it would be interesting to study other properties of the right-angled Coxeter group W_Γ when $\Gamma \in \mathcal{G}(n, p)$ using tools from random graph theory. In particular, determining the threshold for algebraic thickness of every order, or the exact rate of divergence of W_Γ for all p would be of great interest (see [8, Question 1]). Doing so will require new group theoretic ideas to translate these properties into graph theoretic language, and the identification of suitably tractable graph theoretic proxies for these in $\mathcal{G}(n, p)$. Work of Levcovitz [22] provides promising progress towards finding combinatorial properties to encode higher rates of polynomial divergence in right-angled Coxeter groups; indeed, as we finalized this paper, Levcovitz released a new preprint [23] that provides such a translation, which we expect will be of use in future work on this problem. As Levcovitz's work involves hypergraphs, developing new techniques for generalizations of square percolation to hypergraphs will likely be key to further progress.

Finally, one could study thickness and relative hyperbolicity in random right-angled Coxeter groups with presentation graphs drawn from other distributions than the Erdős–Rényi random graph model, such as random regular graphs. We do not know of any work which has been done in this direction at the present time.

REFERENCES

- [1] ABBOTT, C., BEHRSTOCK, J., AND DURHAM, M. G. Largest acylindrical actions and Stability in hierarchically hyperbolic groups. *Trans. Amer. Math. Soc. Ser. B* 8 (2021), 66–104.
- [2] ALON, N., AND SPENCER, J. H. *The probabilistic method*, fourth ed. Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons, Inc., Hoboken, NJ, 2015.
- [3] BALISTER, P. Branching processes. *Lecture Note Series, IMS, NUS* (2006).
- [4] BEHRSTOCK, J., AND DRUȚU, C. Divergence, thick groups, and short conjugators. *Illinois J. Math.* 58, 4 (2014), 939–980.
- [5] BEHRSTOCK, J., DRUȚU, C., AND MOSHER, L. Thick metric spaces, relative hyperbolicity, and quasi-isometric rigidity. *Mathematische Annalen* 344, 3 (2009), 543–595.
- [6] BEHRSTOCK, J., FALGAS-RAVRY, V., HAGEN, M. F., AND SUSSE, T. Global structural properties of random graphs. *Int. Math. Res. Not.* 2018, 5 (2018), 1411–1441.
- [7] BEHRSTOCK, J., HAGEN, M., AND SISTO, A. Hierarchically hyperbolic spaces I: curve complexes for cubical groups. *Geom. Topol.* 21 (2017), 1731–1804.
- [8] BEHRSTOCK, J., HAGEN, M., AND SISTO, A. Thickness, relative hyperbolicity, and randomness in Coxeter groups. *Algebr. Geom. Topol.* 17, 2 (2017), 705–740.
- [9] BOLLOBÁS, B., AND RIORDAN, O. Clique percolation. *Random Structures Algorithms* 35, 3 (2009), 294–322.
- [10] CAPRACE, P.-E., AND SAGEEV, M. Rank rigidity for CAT(0) cube complexes. *Geom. Funct. Anal.* 21, 4 (2011), 851–891.
- [11] CHARNEY, R., AND SULTAN, H. Contracting boundaries of CAT(0) spaces. *J. Topol.* 8, 1 (2015), 93–117.
- [12] DANI, P., AND THOMAS, A. Divergence in right-angled Coxeter groups. *Trans. Amer. Math. Soc.* 367, 5 (2015), 3549–3577.
- [13] DAVIS, M. W. *The geometry and topology of Coxeter groups*, vol. 32 of London Mathematical Society Monographs Series. Princeton University Press, 2008.
- [14] DERÉNYI, I., PALLA, G., AND VICSEK, T. Clique percolation in random networks. *Phys. Rev. Lett.* 94 (Apr 2005), 160202.

- [15] DRUȚU, C., MOZES, S., AND SAPIR, M. Divergence in lattices in semisimple Lie groups and graphs of groups. *Trans. Amer. Math. Soc.* 362, 5 (2010), 2451–2505.
- [16] DWASS, M. The total progeny in a branching process and a related random walk. *Journal of Applied Probability* 6, 3 (1969), 682–686.
- [17] GERSTEN, S. M. Divergence in 3-manifold groups. *Geom. Funct. Anal.* 4, 6 (1994), 633–647.
- [18] HARRIS, T. E. A lower bound for the critical probability in a certain percolation process. In *Mathematical Proceedings of the Cambridge Philosophical Society* (1960), vol. 56, Cambridge University Press, pp. 13–20.
- [19] JANSON, S., ŁUCZAK, T., AND RUCIŃSKI, A. *Random graphs*, vol. 45. John Wiley & Sons, 2011.
- [20] KAPOVICH, M., AND LEEB, B. 3-manifold groups and nonpositive curvature. *Geom. Funct. Anal.* 8, 5 (1998), 841–852.
- [21] LEVCOVITZ, I. Divergence of CAT(0) cube complexes and Coxeter groups. *Algebr. Geom. Topol.* 18, 3 (2018), 1633–1673.
- [22] LEVCOVITZ, I. A quasi-isometry invariant and thickness bounds for right-angled Coxeter groups. *Groups Geom. Dyn.* 13, 1 (2019), 349–378.
- [23] LEVCOVITZ, I. Characterizing divergence and thickness in right-angled Coxeter groups. arXiv:2007.13796, 2020.
- [24] LI, M., DENG, Y., AND WANG, B.-H. Clique percolation in random graphs. *Phys. Rev. E* (3) 92, 4 (2015), 042116–042122.
- [25] MÜHLHERR, B. Automorphisms of graph-universal coxeter groups. *Journal of Algebra* 200, 2 (1998), 629–649.
- [26] PALLA, G., DERÉNYI, I., FARKAS, I., AND VICSEK, T. Uncovering the overlapping community structure of complex networks in nature and society. *Nature* 435 (June 2005), 814–818.
- [27] TÓTH, B., VICSEK, T., AND PALLA, G. Overlapping modularity at the critical point of k -clique percolation. *J. Stat. Phys.* 151, 3-4 (2013), 689–706.
- [28] WANG, B., CAO, L., SUZUKI, H., AND AIHARA, K. Impacts of clustering on interacting epidemics. *J. Theoret. Biol.* 304 (2012), 121–130.

LEHMAN COLLEGE AND THE GRADUATE CENTER, CUNY, NEW YORK, NEW YORK, USA

E-mail address: jason.behrstock@lehman.cuny.edu

UMEÅ UNIVERSITET, UMEÅ, SWEDEN

E-mail address: victor.falgas-ravry@umu.se

BARD COLLEGE AT SIMON’S ROCK, GREAT BARRINGTON, MASS., USA

E-mail address: tsusse@simons-rock.edu