

Multi-coloured jigsaw percolation on random graphs

OLIVER COOLEY* AND ABRAHAM GUTIERREZ†

The jigsaw percolation process, introduced by Brummitt, Chatterjee, Dey and Sivakoff, was inspired by a group of people collectively solving a puzzle. It can also be seen as a measure of whether two graphs on a common vertex set are “jointly connected”. In this paper we consider the natural generalisation of this process to an arbitrary number of graphs on the same vertex set. We prove that if these graphs are random, then the jigsaw percolation process exhibits a phase transition in terms of the product of the edge probabilities. This generalises a result of Bollobás, Riordan, Slivken and Smith.

AMS 2000 SUBJECT CLASSIFICATIONS: 05C80.

1. Introduction

1.1. Jigsaw percolation

In recent years there has been significant research inspired by the observation that certain advances are only possible as a result of the collaboration of a group of people, rather than the work of one individual e.g. [2, 12, 13, 14].

In order to model this mathematically, Brummitt, Chatterjee, Dey and Sivakoff [6] introduced the jigsaw percolation process. The premise is that a group of people each have one piece of a puzzle which must be combined in a certain way to solve the puzzle. The individuals (and their associated puzzle pieces) are represented by a set of vertices, and there are two graphs on these vertices: a *people graph*, with an edge if the two people know each other; and a *puzzle graph* with an edge if the two puzzle pieces are compatible. In the jigsaw percolation process, we begin with each vertex forming its own *cluster* and we merge two clusters if there is an edge between them in both the people and the puzzle graph – this represents these two people sharing all their information. The new merged cluster inherits all the incident edges

arXiv: [1712.00992](https://arxiv.org/abs/1712.00992)

*Supported by Austrian Science Fund (FWF): I3747.

†Supported by Austrian Science Fund (FWF): P29355-N35.

of the original clusters. (The process will be described more formally later.) This process is iterated until either there is only one cluster remaining, in which case we say that the process *percolates* indicating that the puzzle has been solved, or no more clusters can be merged, in which case we say that the process *does not percolate*. More generally, if the two graphs are G_1 and G_2 , we say that the *double-graph* (G_1, G_2) percolates or does not percolate respectively.

This process was introduced by Brummitt, Chatterjee, Dey and Sivakoff in [6] and was also studied by Gravner and Sivakoff in [11]. Bollobás, Rioridan, Slivken and Smith [5] considered the case when the people graph and the puzzle graph are independent binomial random graphs, and proved that the property of the two graphs percolating undergoes a phase transition in terms of the product of the two associated edge probabilities. More precisely, their result can be stated as follows. Let $G(n, p)$ denote the Erdős-Rényi binomial random graph on vertex set $[n] := \{1, 2, \dots, n\}$ in which each pair of vertices forms an edge with probability p independently of each other. We say that a property or event holds *with high probability* (abbreviated to *whp*), if it holds with probability tending to 1 as n tends to infinity.

Theorem 1 ([5]). *There exists a constant c such that the following holds: let $G_1 = G(n, p_1), G_2 = G(n, p_2)$ be independent binomial random graphs on the same vertex set, where $0 \leq p_1 = p_1(n), p_2 = p_2(n) \leq 1$. Then*

- (i) *if $p_1 p_2 \leq \frac{1}{cn \ln n}$ or $\min\{p_1, p_2\} \leq \frac{\ln n}{cn}$ then whp (G_1, G_2) does not percolate;*
- (ii) *if $p_1 p_2 \geq \frac{c}{n \ln n}$ and $\min\{p_1, p_2\} \geq \frac{c \ln n}{n}$, then whp (G_1, G_2) percolates.*

Note that this theorem is not quite stated as it appeared in [5], but it is easy to derive this form from the original. We also observe that connectedness of each of the two graphs is a necessary (but not sufficient) condition for percolation of the double-graph. The conditions on $\min\{p_1, p_2\}$ determine whether this necessary condition is satisfied whp, since the threshold for connectivity is at $p = \frac{\ln n}{n}$ as famously proved by Erdős and Rényi in [9].

Theorem 1 was extended to hypergraphs by Bollobás, Cooley, Kang and Koch [4], with a whole family of generalisations of the percolation process to k -uniform hypergraphs in which the clusters consist of j -sets of vertices for $1 \leq j \leq k - 1$. Theorem 1 was also strengthened by Cooley, Kapetanopoulos and Makai [8], who determined that the sharp threshold for jigsaw percolation occurs when the product of the two probabilities is $p_1 p_2 = \frac{1}{4n \ln n}$.

In this paper, we extend in a different direction, namely to an arbitrary number of graphs on the same vertex set.

Definition. An r -fold graph is an $(r + 1)$ -tuple $\mathbf{G} := (V, E_1, \dots, E_r)$, where $V := [n]$ is the set of vertices and $E_i \subseteq \binom{V}{2}$ for each $i \in [r]$. We will call $1, 2, \dots, r$ the *colours* of \mathbf{G} and the graph $G_i := (V, E_i)$ will be said to be of colour i for every $i \in [r]$.

The multi-coloured jigsaw algorithm on an r -fold graph is the natural generalisation of the 2-coloured version in which clusters must be joined by an edge of each colour in order to merge. A formal description of this algorithm is given later in Algorithm 3.

It is easy to see that percolation for $r = 1$ is equivalent to connectedness of the graph. Thus, percolation of the jigsaw process is a generalised notion of connectedness of multiple graphs on the same vertex set. Therefore Theorem 1 and the main result of this paper (Theorem 2) may be seen as generalisations of the connectedness threshold result of Erdős and Rényi [9].

1.2. Main theorem

To state the main result of the paper, we introduce the following generalisation of the binomial model for random graphs.

Definition. An r -fold binomial random graph $\mathbf{G}(n, p_1, \dots, p_r)$ is an r -fold graph $([n], E_1, \dots, E_r)$ where $([n], E_i) \sim G(n, p_i)$ are independent binomial random graphs for every $i \in [r]$.

The main result of this paper is a generalisation of Theorem 1 to an arbitrary number of colours r . In fact we will even allow r to tend to infinity sufficiently slowly as a function of n .

Theorem 2. *Let $2 \leq r = o(\sqrt{\ln \ln n})$ and $C_r := 2^{8r^2}$. Then the following holds: suppose that p_1, \dots, p_r are functions of n such that $0 \leq p_1 \leq p_2 \leq \dots \leq p_r \leq 1$ and $\mathbf{G} = \mathbf{G}(n, p_1, p_2, \dots, p_r)$. For $i \in [r]$ let $P_i := p_1 p_2 \dots p_i$. Then*

- (i) *if $P_i \leq \frac{1}{C_r n (\ln n)^{i-1}}$ for some $2 \leq i \leq r$ or $P_1 \leq \frac{\ln n}{C_r n}$ then whp \mathbf{G} does not percolate;*
- (ii) *if $P_i \geq \frac{C_r}{n (\ln n)^{i-1}}$ for every $2 \leq i \leq r$ and $P_1 \geq \frac{C_r \ln n}{n}$, then whp \mathbf{G} percolates.*

Let us note that if r is dependent on n , then also C_r is explicitly dependent on n . However, if r is constant, then so is C_r . Note also that both in the proof of Theorem 1 in [5] and in the proof of Theorem 2 in this paper, no attempt is made to optimise the constants c and C_r , and the value given in Theorem 2 is probably far from best possible. In [8], Theorem 1 was strengthened by optimising the constants and thus proving the exact location of the threshold, which is at $p_1 p_2 = \frac{1}{4n \ln n}$.

Remark. Given an r -fold graph $\mathbf{G} = ([n], E_1, \dots, E_r)$ it is easy to see that percolation of every i -fold graph $([n], E_{j_1}, \dots, E_{j_i})$ obtained by considering a subset of i colours is a necessary condition for percolation of \mathbf{G} (but not sufficient). For $i = 1$, we guarantee connectedness by taking $p_1 = P_1 \geq \frac{C_r \ln n}{n}$. For $2 \leq i \leq r$ the inequalities $P_i \geq \frac{C_r}{n(\ln n)^{i-1}}$ together with $p_1 \leq p_2 \leq \dots \leq p_r$ ensure that every such i -fold graph percolates whp.

While much of the proof of Theorem 2 follows the proof of Theorem 1 in [5], there are important differences for the multi-coloured case which present additional difficulty. We will point out these differences in the course of the proof.

1.3. The multi-coloured jigsaw algorithm

The multi-coloured jigsaw process is formally described as follows.

Algorithm 3 (The multi-coloured jigsaw algorithm).

Input: r -fold graph $\mathbf{G} := ([n], E_1, \dots, E_r)$.

At time $t \geq 0$ there is a partition $\mathcal{C}_t = \{C_t^1, C_t^2, \dots, C_t^{k_t}\}$ of the vertex set $[n]$, which we construct inductively as follows:

1. We take $k_0 = n$, set $C_0^j := \{j\}$ and $\mathcal{C}_0 = \{\{1\}, \dots, \{n\}\}$ for all $j \in [n]$ i.e. we begin at time 0 with the discrete partition into single vertices.
2. At time $t \geq 0$, construct a graph H_t on vertex set \mathcal{C}_t by joining C_t^j to $C_t^{j'}$ if there exist edges $e_i := \{v_{j,i}, v_{j',i}\} \in E_i$ for all $i \in [r]$ such that $v_{j,i} \in C_t^j$ and $v_{j',i} \in C_t^{j'}$.
3. If $E(H_t) = \emptyset$, then STOP. Otherwise, construct the partition

$$\mathcal{C}_{t+1} = \{C_{t+1}^1, \dots, C_{t+1}^{k_{t+1}}\},$$

where $C_{t+1}^1, \dots, C_{t+1}^{k_{t+1}}$ are obtained by merging the connected components of H_t i.e. if $D_t^j \subseteq \mathcal{C}_t$ induces a connected component in H_t then $C_{t+1}^j = \bigcup_{C \in D_t^j} C$.

4. If $|\mathcal{C}_{t+1}| = 1$ STOP. Otherwise, go to step 2.

Definition. • We say that the r -fold graph $\mathbf{G} = (V, E_1, E_2, \dots, E_r)$ percolates if Algorithm 3 applied to \mathbf{G} ends with one single cluster. Otherwise we say that \mathbf{G} does not percolate.

- We say that a subset $W \subseteq V$ is a percolating subset (or that it percolates) in $\mathbf{G} = (V, E_1, E_2, \dots, E_r)$ if the induced r -fold subgraph $\mathbf{G}[W] := (W, E_1[W], \dots, E_r[W])$ percolates.

The definition of a percolating subset corresponds to the definition in [5] of an internally spanned set.

1.4. Intuition

Let us consider heuristically how the jigsaw process might be expected to evolve. For simplicity we discuss the case $r = 2$, although the intuition is transferrable to a larger number of colours.

We begin with n clusters each containing a single vertex. Initially clusters can only merge if there is a double-edge (i.e. both a red and a blue edge) between the corresponding vertices. Although such double-edges are rare, the fact that there are many vertices will mean that some clusters will indeed merge.

Subsequently clusters may continue to merge and grow larger. Indeed, the larger a cluster becomes, the more likely it is to merge with other clusters and continue growing. Thus we might expect that after a certain size we encounter a snowball effect, and the growth of the largest cluster accelerates until it contains all vertices.

Indeed, this intuition turns out to be correct: there is a *bottleneck* in the percolation process, which occurs at size $\Theta(\ln n)$ (this was observed by Bollobás, Riordan, Slivken and Smith in [5]). More precisely, in the subcritical case we show that the largest cluster in the percolation process will not exceed size $\ln n$ whp. On the other hand, in the proof of the supercritical process, the hardest part is proving that there is a cluster of size slightly larger than $\ln n$ — then it is fairly easy to prove that this cluster will eventually merge with all other clusters whp, and therefore we have percolation.

We will ignore floors and ceilings throughout the paper whenever they do not significantly affect the arguments (this is usually the case since we consider graphs on n vertices, where $n \rightarrow \infty$). We also assume that n is sufficiently large in calculations.

2. Proof of the subcritical case

In this section we prove part (i) of Theorem 2. This proof is an obvious generalisation of the corresponding proof for 2 colours in [5].

We first handle the case when $p_1 \leq \frac{\ln n}{C_r n}$. Since the threshold for connectedness in $G(n, p)$ is at $p = \frac{\ln n}{n}$ as proved by Erdős and Rényi in [9], the graph (V, E_1) is disconnected whp (since $C_r = 2^{8r^2} > 1$) and since connectedness of each (V, E_i) is a necessary condition for percolation, whp the r -fold graph will not percolate, as claimed.

In the case when $P_i \leq \frac{1}{C_r n \ln^{i-1} n}$ for some $2 \leq i \leq r$, we just need to prove that whp the i -fold graph (V, E_1, \dots, E_i) does not percolate. We first observe the following necessary condition for percolation.

Claim 4. *For every natural number $k \leq n/2$ and every percolating i -fold graph $\mathbf{G} = (V, E_1, E_2, \dots, E_i)$ there is a subset W of V of size at least k but not larger than $2k$ such that $(W, E_j[W])$ is connected for each $1 \leq j \leq i$.*

Proof. We can modify step (3) in Algorithm 3 so that instead of merging entire connected components of sets C_t^j (connected in \mathcal{G}_t), we only merge an arbitrary pair of sets C_t^j which are joined by an edge of each of the r colours.

We begin to merge pairs of clusters and consider the first time that a cluster W of size at least k appears. This cluster cannot be of size larger than $2k$ since it is the union of two clusters of size less than k . Since $\mathbf{G}[W]$ percolates, every $(W, E_j[W])$ is connected. \square

We define the random variable \mathcal{T}_k to be the number of i -fold subgraphs $(W, E'_1, E'_2, \dots, E'_i)$ where $W \subseteq V$ and $E'_j \subseteq E_j$ and the graphs (W, E'_j) are trees with exactly k vertices. Thus, defining \mathcal{A} to be the event that the i -fold graph $\mathbf{G} = (V, E_1, E_2, \dots, E_i)$ percolates, Claim 4 and a union bound tells us that

$$\mathbb{P}[\mathcal{A}] \leq \sum_{k=\ln n}^{2 \ln n} \mathbb{P}[\mathcal{T}_k \geq 1] \leq \sum_{k=\ln n}^{2 \ln n} \mathbb{E}[\mathcal{T}_k],$$

where we have also applied Markov's inequality to obtain the final expression.

Therefore we now want to estimate $\mathbb{E}[\mathcal{T}_k]$. There are $\binom{n}{k}$ ways of choosing the vertices, and for each colour j , there are k^{k-2} possible spanning trees and p_j^{k-1} is the probability that the edges of such spanning tree are present. Thus $\mathbb{E}[\mathcal{T}_k] = \binom{n}{k} \prod_{j=1}^i k^{k-2} p_j^{k-1}$. Thus we obtain

$$\begin{aligned} \mathbb{P}[\mathcal{A}] &\leq \sum_{k=\ln n}^{2 \ln n} \mathbb{E}[\mathcal{T}_k] = \sum_{k=\ln n}^{2 \ln n} \binom{n}{k} \prod_{j=1}^i k^{k-2} p_j^{k-1} \leq \sum_{k=\ln n}^{2 \ln n} \left(\frac{en}{k}\right)^k \prod_{j=1}^i \frac{(kp_j)^{k-1}}{k} \\ &= \sum_{k=\ln n}^{2 \ln n} \frac{en}{k^{i+1}} (enk^{i-1} P_i)^{k-1} \\ &\leq en \sum_{k=\ln n}^{2 \ln n} \left(\frac{2^{i-1} e}{C_r}\right)^{k-1} \\ &\leq en \sum_{k=\ln n}^{\infty} \left(\frac{1}{e^2}\right)^{k-1} = o(1), \end{aligned}$$

where the last line follows since $C_r = 2^{8r^2} \geq 2^{r-1}e^3 \geq 2^{i-1}e^3$. This completes the proof of the subcritical case.

3. Proof of the supercritical case

In this section we will prove part (ii) of Theorem 2. The main idea for the proof is to construct an increasing sequence of percolating subsets $V_1 \subseteq V_2 \subseteq V_3 = V$. Therefore we will divide the proof into three parts, and we aim to prove the following:

Part I: whp there is a percolating subset $V_1 \subseteq V$ of size at least $t_1 := (\ln n)^{1+\frac{1}{r}}$;

Part II: conditioned on the existence of a percolating subset $V_1 \subseteq V$ of size at least t_1 , whp there exists a percolating subset $V_2 \supset V_1$ of size at least $\frac{n}{2^{r+2}}$;

Part III: conditioned on the existence of a percolating subset V_2 of size at least $\frac{n}{2^{r+2}}$, whp the whole set V percolates.

The independence between the three parts of the proof is guaranteed by independent rounds of exposure. More precisely, let

$$\mathbf{G}^{(j)} := \left([n], E_1^{(j)}, E_2^{(j)}, \dots, E_r^{(j)} \right) \sim G \left(n, \frac{p_1}{3}, \frac{p_2}{3}, \dots, \frac{p_r}{3} \right)$$

independently for $j = 1, 2, 3$. Then we will view \mathbf{G} as the union $\mathbf{G}^{(1)} \cup \mathbf{G}^{(2)} \cup \mathbf{G}^{(3)}$.¹

In Part j of the proof we will work only with $\mathbf{G}^{(j)}$, effectively *exposing* an r -fold probability of $(p_1/3, p_2/3, \dots, p_r/3)$ in each round.

3.1. Preliminaries

We begin with some basic observations.

Proposition 5. *Suppose $r, C_r, p_1, p_2, \dots, p_r$ satisfy the conditions of Theorem 2 (ii). Then for n large enough there exist real numbers $0 \leq p'_1 \leq p'_2 \leq \dots \leq p'_r \leq 1$ that also satisfy conditions of Theorem 2 (ii) and such that*

- $p'_i \leq p_i$ for every i ,
- $p'_1 p'_2 \dots p'_r = \frac{C_r}{n(\ln n)^{r-1}}$.

¹Note that this is not quite true, since the union of three independent copies of $G(n, p/3)$ is distributed as $G(n, p^*)$, where $p^* = p - p^2/3 + p^3/27$. However, since $p^* < p$ we can couple $G(n, p^*)$ with $G(n, p)$ such that $G(n, p^*) \subseteq G(n, p)$, and since percolation is a monotone increasing property, this will be sufficient.

We omit the proof of this intuitively obvious result – for details see [10].

Since percolation is a monotone property, by Proposition 5 we may assume that

$$(1) \quad P_r = p_1 \dots p_r = \frac{C_r}{n(\ln n)^{r-1}}.$$

From this, and recalling that $p_1 \geq \frac{C_r \ln n}{n}$, we can deduce that

$$(2) \quad p_2 \leq \left(\frac{p_1 p_2 p_3 \dots p_r}{p_1} \right)^{1/(r-1)} \leq \left(\frac{1}{\ln n} \right)^{\frac{r}{r-1}}.$$

Remark. In the two-coloured case, i.e. $r = 2$, we obtain the bound $p_1 \leq p_2 \leq (\ln n)^{-2}$. In the general case, the analogous calculation only yields the bound $p_i \leq (\ln n)^{-1}$ (for $i \geq 3$). This seemingly minor difference leads to significant extra difficulties, as some approximations are no longer valid. We will therefore have to distinguish between “small” and “large” p_i (see Lemmas 8 and 9 in Section 3.2).

3.2. Part I

We will construct a large percolating subset V_1 by “trial and error”. Algorithm 6 will start from a single vertex and add one vertex at a time in an attempt to construct V_1 . We will make several attempts to construct V_1 – each such attempt is called a *round*; each round consists of a number of *steps*. We divide the proof into two stages:

- I.a First, we will bound from below the probability that the algorithm constructs a percolating subset of size at least $t_0 := \frac{\ln n}{c_r}$ (in one round, see Lemma 10) where $c_r := C_r^{\frac{1}{r-1}}$.
- I.b Second, conditioned on the algorithm constructing a percolating subset of size at least t_0 , we will bound from below the probability that the algorithm constructs a percolating subset of size at least $t_1 = (\ln n)^{1+\frac{1}{r}}$ (in one round, see Lemma 11).

The probability that Algorithm 6 reaches t_1 in one round is bounded from below by the product of the probabilities of the two stages. This product turns out to be small, but crucially Algorithm 6 makes many attempts to reach t_1 . The probability that at least one of these rounds succeeds will be large (see Lemma 13).

In step t of round k of Algorithm 6, we have a trial set X_k^t which is a percolating set. If the algorithm finds a suitable vertex to add to the trial set X_k^t , we create the new trial set X_k^{t+1} and proceed to step $t + 1$ of round k . If not, we discard the vertices of the trial set X_k^t and begin the new round $k + 1$. We stop if a round has reached step t_1 or if we have had $\frac{n}{2t_1}$ rounds.

The formal description of the algorithm is as follows:

Algorithm 6 (The 1-by-1 algorithm). *The algorithm is divided into rounds, indexed by k , and each round is divided into steps, indexed by t . At the start of the k -th round there is a set $A_k^0 \subseteq [n]$ of active vertices and a set $D_k \subseteq [n]$ of discarded vertices. We begin with $A_1^0 = [n]$ and $D_1 = \emptyset$. The procedure of the k -th round is as follows:*

At the start of the t -th step of the k -th round there are sets of trial and dormant vertices:

- $X_k^t = \{x_k^1, x_k^2, \dots, x_k^t\} \subseteq A_k^0$ (trial vertices);
- $U_k^t \subseteq A_k^0$ (dormant vertices),

where $A_k^0 = X_k^t \cup A_k^t \cup U_k^t$.

(1) For $t = 0$, we move an arbitrary active vertex $x_k^1 \in A_k^0$ to the trial set:

- $X_k^1 := \{x_k^1\};$ • $U_k^1 := \emptyset;$ • $A_k^1 := A_k^0 \setminus \{x_k^1\};$ • $R_k^0 := \emptyset,$

and set $t := 1$.

(2) For $t \geq 1$, we reveal all edges of $E_1^{(1)}$ between A_k^t and x_k^t and edges of $E_i^{(1)}$ ($i = 1, \dots, r$) between any neighbour of x_k^t in $E_1^{(1)}$ and x_k^1, \dots, x_k^t .
Let

- $R_k^t := \{x \in A_k^t : xx_k^t \in E_1^{(1)}\};$
- $B_k^t := \{x \in R_k^t : \forall i \in \{2, 3, \dots, r\} \exists s_i \leq t \text{ such that } xx_k^{s_i} \in E_i^{(1)}\}.$

(3) If $B_k^t \neq \emptyset$, then let x_k^{t+1} be an arbitrary element of B_k^t . Then set:

- $X_k^{t+1} := X_k^t \cup \{x_k^{t+1}\};$
- $A_k^{t+1} := A_k^t \setminus R_k^t;$
- $U_k^{t+1} := U_k^t \cup (R_k^t \setminus \{x_k^{t+1}\}).$

If $t \geq t_1 = (\ln n)^{1+\frac{1}{r}}$ then STOP, otherwise set $t = t + 1$ and go to step (3).

(4) If $B_k^t = \emptyset$, then set

- $A_{k+1}^0 := A_k^0 \setminus X_k;$ • $D_{k+1} := D_k \cup X_k^t.$

(5) If

$$k \geq \frac{n}{2(\ln n)^{1+\frac{1}{r}}}$$

then STOP, otherwise set $k := k + 1$ and $t := 0$, and go to step (1).

We reveal edges and non-edges as they are exposed in the algorithm, e.g. when defining R_t^k we test each pair (x_0^k, a) for $a \in A_k^t$ to reveal whether it lies in $E_1^{(1)}$. Note that since every tested pair has at least one of its endpoints in the trial set, we guarantee independence between rounds by discarding the trial set at the end of each round. We also have independence within each round, because no pair is tested twice within a round.

Since we consider at most $n/(2(\ln n)^{1+\frac{1}{r}})$ rounds, and stop each with a trial set of size at most $(\ln n)^{1+\frac{1}{r}}$ vertices, we start each new round with at least $n/2$ vertices, i.e.

$$|A_k^0| \geq \frac{n}{2}.$$

We will need the following definitions:

Definition.

- Let \mathcal{X}_k^t be the event that X_k^t is defined (i.e. we reach step t in round k).
- Let $\mathcal{S}_k^t := \{|R_k^s| \leq \frac{n}{4t_1} \text{ for } s = 0, 1, 2, \dots, t\}$.
- Let $\mathcal{Y}_k^t := \mathcal{X}_k^t \cap \mathcal{S}_k^t$.

The event \mathcal{X}_k^t means that we found a percolating subset of size t formed with only edges of the first round of exposure. Conditioned on getting to round k the event \mathcal{X}_k^1 always holds. For $t \geq 2$ the event \mathcal{X}_k^t is equivalent to the event that B_k^{t-1} is non-empty. The event \mathcal{S}_k^t guarantees that within a round k , we do not discard too many vertices by step t . More specifically, if the event \mathcal{Y}_k^{t-1} holds, we have

$$|A_k^t| \geq |A_k^0| - (t-1) \frac{n}{4t_1} \geq \frac{n}{2} - \frac{t}{t_1} \frac{n}{4} \geq \frac{n}{4}.$$

Note that if we get to round k , the event \mathcal{S}_k^0 always holds, since $R_k^0 = \emptyset$.

We will use the following easily verified inequalities to approximate some expressions.

Fact 7. For $t \geq 0$, $p \leq 1$ we have:

- a) If $1 - pt \geq 0$ then $1 - (1-p)^t \geq pt(1-pt)$;
- b) If $1 - pt \leq \frac{1}{2}$ then $1 - (1-p)^t \geq \frac{1}{5}$.

Note that a) was used in [5], but that b) is only needed for the multi-coloured case. We will also use the following observation: for $t \leq t_1$,

$$(3) \quad p_1 t \leq p_2 t_1 \stackrel{(2)}{\leq} \frac{(\ln n)^{1+\frac{1}{r}}}{(\ln n)^{\frac{r}{r-1}}} = (\ln n)^{\frac{1}{r}-\frac{1}{r-1}} = o(1).$$

The following parameter will help us distinguish between “small” and “large” p_i ’s, something that is not needed in the 2-coloured case since both p_1 and p_2 are “small”.

Definition. For $t \leq t_1$, let $i_t := \max\{i \in [2, r] : 1 - \frac{p_i t}{3} \geq \frac{1}{2}\}$.

Note that by (3), i_t is well defined. We now calculate a lower bound on the probability of “one-step success” i.e. the probability of being able to add a vertex to the percolating set in Algorithm 6. Recall that $P_i = p_1 p_2 \dots p_i$ for each $1 \leq i \leq r$.

Lemma 8. *For n large enough and $1 \leq t \leq t_1 = (\ln n)^{1+\frac{1}{r}}$ we have that independently for each $x \in \mathcal{A}_k^t$ the following holds:*

$$\mathbb{P}[x \in B_k^t] \geq \left(\frac{1}{5}\right)^{r-1} \frac{P_{i_t}}{3^{i_t}} t^{i_t-1}.$$

Proof. We have

$$\begin{aligned} \mathbb{P}[x \in B_k^t] &= \frac{p_1}{3} \prod_{j=2}^r \left(1 - \left(1 - \frac{p_j}{3}\right)^t\right) \\ &\stackrel{(F. 7)}{\geq} \frac{p_1}{3} \left(\frac{1}{5}\right)^{r-i_t} \prod_{j=2}^{i_t} \frac{p_j t}{3} \left(1 - \frac{p_j t}{3}\right) \\ &\geq \left(\frac{1}{5}\right)^{r-i_t} \frac{P_{i_t}}{3^{i_t}} t^{i_t-1} \left(\frac{1}{2}\right)^{i_t-1} \geq \left(\frac{1}{5}\right)^{r-1} \frac{P_{i_t}}{3^{i_t}} t^{i_t-1}. \quad \square \end{aligned}$$

We now make use of the lower bound of Lemma 8 and the fact that the events $\{x \in B_k^t\}$ are independent for different vertices x . Recall that $c_r := C_r^{\frac{1}{r-1}}$.

Lemma 9. *For $1 \leq t \leq t_1 = (\ln n)^{1+\frac{1}{r}}$ and n large enough the following holds.*

$$(a) \quad \mathbb{P}[\mathcal{Y}_k^t | \mathcal{Y}_k^{t-1}] \geq 1 - \exp\left(-\left(\frac{1}{16}\right)^r \left(\frac{c_r t}{\ln n}\right)^{i_t-1}\right).$$

(b) If $(\frac{1}{16})^r (\frac{c_r t}{\ln n})^{i_t-1} \leq 1$ we have

$$\mathbb{P} [\mathcal{Y}_k^t | \mathcal{Y}_k^{t-1}] \geq \left(\frac{1}{32}\right)^r \left(\frac{c_r t}{\ln n}\right)^{i_t-1}.$$

Proof. Since case (b) follows from case (a) and the inequality $1 - \frac{x}{2} \geq \exp(-x)$, valid for $x \in [0, 1]$, we only need to prove case (a).

We recall that $\mathcal{Y}_k^t = \mathcal{X}_k^t \cap \mathcal{S}_k^t$, and therefore

$$(4) \quad \mathbb{P} [\mathcal{Y}_k^t | \mathcal{Y}_k^{t-1}] \geq 1 - \mathbb{P} [\overline{\mathcal{X}}_k^t | \mathcal{Y}_k^{t-1}] - \mathbb{P} [\overline{\mathcal{S}}_k^t | \mathcal{Y}_k^{t-1}].$$

We bound the two probability terms on the right-hand side of the inequality separately.

Let \mathcal{Z}_k^t be the random variable that represents the number of sets Z of size $\frac{n}{4t_1}$ such that $Z \subseteq R_k^t$. If $\overline{\mathcal{S}}_k^t \cap \mathcal{Y}_k^{t-1}$ holds then $\mathcal{Z}_k^t \geq 1$, therefore we deduce using Markov's inequality that

$$(5) \quad \mathbb{P} [\overline{\mathcal{S}}_k^t | \mathcal{Y}_k^{t-1}] \leq \mathbb{E} [\mathcal{Z}_k^t | \mathcal{Y}_k^{t-1}] \leq \binom{n}{\frac{n}{4t_1}} \left(\frac{p_1}{3}\right)^{\frac{n}{4t_1}} \leq \left(\frac{4}{3} e t_1 p_1\right)^{\frac{n}{4t_1}} \leq e^{-\sqrt{n}}.$$

For the last inequality we used that $n/(4t_1) \geq \sqrt{n}$ and $p_1 t_1 \stackrel{(3)}{=} o(1) \leq \frac{3}{4e^2}$ for n large enough.

For the second term in (4), we use Lemma 8 and the observation that $|A_k^{t-1}| \geq n/4$ to obtain

$$\begin{aligned} \mathbb{P} [\overline{\mathcal{X}}_k^t | \mathcal{Y}_k^{t-1}] &= \prod_{x \in A_k^{t-1}} \mathbb{P}[x \notin B_k^{t-1}] \leq \left(1 - \left(\frac{1}{5}\right)^{r-1} \frac{P_{i_t}}{3^{i_t}} t^{i_t-1}\right)^{\frac{n}{4}} \\ &\leq \exp\left(-\frac{n}{4} \left(\frac{1}{5}\right)^{r-1} \frac{P_{i_t}}{3^{i_t}} t^{i_t-1}\right). \end{aligned}$$

From the assumptions of Theorem 2, we have $P_{i_t} \geq \frac{C_r}{n(\ln n)^{i_t-1}} \geq \frac{c_r^{i_t-1}}{n(\ln n)^{i_t-1}}$. We deduce that

$$(6) \quad \mathbb{P} [\overline{\mathcal{X}}_k^t | \mathcal{Y}_k^{t-1}] \leq \exp\left(-\left(\frac{1}{15}\right)^r \left(\frac{c_r t}{\ln n}\right)^{i_t-1}\right).$$

Substituting (5) and (6) into (4) gives

$$\mathbb{P}[\mathcal{Y}_k^t | \mathcal{Y}_k^{t-1}] \geq 1 - \exp\left(-\left(\frac{1}{15}\right)^r \left(\frac{c_r t}{\ln n}\right)^{i_t-1}\right) - \exp(-\sqrt{n}).$$

To complete the proof we recall that $t \leq t_1$, $2 \leq i_t \leq r$ and observe that

$$\left(\frac{t}{\ln n}\right)^{i_t-1} \leq \left(\frac{t_1}{\ln n}\right)^{r-1} = (\ln n)^{\frac{r-1}{r}} = o(\sqrt{n}),$$

and conclude that

$$(7) \quad \mathbb{P}[\mathcal{Y}_k^t | \mathcal{Y}_k^{t-1}] \geq 1 - \exp\left(-\left(\frac{1}{16}\right)^r \left(\frac{c_r t}{\ln n}\right)^{i_t-1}\right). \quad \square$$

Recall that $t_0 = \frac{\ln n}{C_r^{1/(r-1)}} = \frac{\ln n}{c_r}$. In order to calculate a lower bound on the probability of “proceeding to step t_1 ” we use Lemma 9 to calculate lower bounds for the events “proceeding to step t_0 ” and “proceeding to step t_1 given that we already proceeded to step t_0 ”. We formally express this in Lemmas 10 and 11.

Lemma 10. $\mathbb{P}[\mathcal{Y}_k^{t_0} | \mathcal{X}_k^0] \geq n^{-\frac{7(r-1)}{c_r}}$.

Proof. Since $\frac{c_r t}{\ln n} \leq 1$ for $1 \leq t \leq t_0$, we can use Lemma 9 (b):

$$\begin{aligned} \mathbb{P}[\mathcal{Y}_k^{t_0} | \mathcal{X}_k^0] &= \mathbb{P}[\mathcal{Y}_k^{t_0} | \mathcal{Y}_k^0] = \prod_{t=1}^{t_0} \mathbb{P}[\mathcal{Y}_k^t | \mathcal{Y}_k^{t-1}] \geq \prod_{t=1}^{t_0} \left(\frac{1}{32}\right)^r \left(\frac{C_r t}{\ln n}\right)^{i_t-1} \\ &\geq \prod_{t=1}^{t_0} \left(\frac{1}{32^2}\right)^{r-1} \left(\frac{c_r t}{\ln n}\right)^{r-1} \\ &\geq \left(\frac{c_r t_0}{1024 \ln n}\right)^{(r-1)t_0} \\ &= \left(\frac{1}{1024}\right)^{(r-1)\frac{\ln n}{c_r}} \geq n^{-\frac{7(r-1)}{c_r}}, \end{aligned}$$

since $\frac{1}{1024} \geq \frac{1}{e^7}$. □

Lemma 11. $\mathbb{P}[\mathcal{X}_k^{t_1} | \mathcal{Y}_k^{t_0}] \geq n^{-2^{8r+2}/c_r}$.

In the proof of Lemma 11 we will use the following claim.

Claim 12. For any real numbers $\alpha \geq 1$ and $0 \leq y \leq 1 - \frac{1}{\alpha}$, we have

$$1 - y \geq e^{-\alpha y}.$$

Proof. From the hypothesis we deduce that

$$\alpha \geq \frac{1}{1 - y} = \sum_{i=0}^{\infty} y^i \stackrel{(y \geq 0)}{\geq} \frac{1}{y} \sum_{i=1}^{\infty} \frac{y^i}{i} = \frac{-\ln(1 - y)}{y},$$

and the desired inequality follows. □

Proof of Lemma 11. We begin by applying Lemma 9 (a):

$$\begin{aligned} \mathbb{P}[\mathcal{X}_k^{t_1} | \mathcal{Y}_k^{t_0}] &\geq \mathbb{P}[\mathcal{Y}_k^{t_1} | \mathcal{Y}_k^{t_0}] = \prod_{t=t_0+1}^{t_1} \mathbb{P}[\mathcal{Y}_k^t | \mathcal{Y}_k^{t-1}] \\ &\geq \prod_{t=t_0}^{t_1} \left(1 - \exp\left(-\left(\frac{1}{16}\right)^r \left(\frac{c_r t}{\ln n}\right)^{i_t-1}\right) \right). \end{aligned}$$

Setting $\alpha := \frac{1}{1 - \exp\{-\left(\frac{1}{16}\right)^r\}} > 1$ and $y := \exp\left(-\left(\frac{1}{16}\right)^r \left(\frac{c_r t}{\ln n}\right)^{i_t-1}\right)$ and noting that $\frac{c_r t}{\ln n} \geq \frac{c_r t_0}{\ln n} = 1$ for $t \geq t_0$, we deduce that $y \leq \exp\left(-\left(\frac{1}{16}\right)^r\right) = 1 - \frac{1}{\alpha}$, therefore we can apply Claim 12. Thus

$$\begin{aligned} \mathbb{P}[\mathcal{X}_k^{t_1} | \mathcal{Y}_k^{t_0}] &\geq \exp\left(-\alpha \sum_{t=t_0}^{t_1} \exp\left(-\left(\frac{1}{16}\right)^r \left(\frac{c_r t}{\ln n}\right)^{i_t-1}\right)\right) \\ &\geq \exp\left(-\alpha \sum_{t=t_0}^{\infty} \exp\left(-\left(\frac{1}{16}\right)^r \left(\frac{c_r t}{\ln n}\right)\right)\right) \\ &= \exp\left(-\frac{\alpha \exp\left(-\left(\frac{1}{16}\right)^r\right)}{1 - \exp\left(-\left(\frac{1}{16}\right)^r \frac{c_r}{\ln n}\right)}\right). \end{aligned}$$

We now simplify the denominator by using the inequality $e^{-x} \leq 1 - x/2$ valid for $x \leq 1$:

$$\begin{aligned} \mathbb{P}[\mathcal{X}_k^{t_1} | \mathcal{Y}_k^{t_0}] &\geq \exp\left(-\frac{\alpha \exp\left(-\left(\frac{1}{16}\right)^r\right)}{1 - \left(1 - \frac{1}{2}\left(\frac{1}{16}\right)^r \frac{c_r}{\ln n}\right)}\right) \\ &= \exp\left(-\frac{2^{4r+1} \alpha \exp\left(-\left(\frac{1}{16}\right)^r\right) \ln n}{c_r}\right). \end{aligned}$$

We now observe that

$$2^{4r+1}\alpha \exp\left(-\left(\frac{1}{16}\right)^r\right) = 2^{4r+1} \frac{\exp\left(-\left(\frac{1}{16}\right)^r\right)}{1 - \exp\left(-\left(\frac{1}{16}\right)^r\right)} \leq 2^{4r+1} \frac{1}{\frac{1}{2}\left(\frac{1}{16}\right)^r} = 2^{8r+2},$$

and the result follows. \square

Using Lemmas 10 and 11, we can complete the proof of Part I

Lemma 13. $G^{(1)}$ contains a percolating subset of size $(\ln n)^{1+\frac{1}{r}}$ with probability at least $1 - e^{-\sqrt{n}}$.

Proof. Let $k \leq \frac{n}{2(\ln n)^{1+\frac{1}{r}}}$. Applying Lemmas 10 and 11, and since $\mathcal{Y}_k^{t_0} \subset \mathcal{Y}_k^0 \subset \mathcal{X}_k^0$, the probability that in round k we find a percolating subset of size $(\ln n)^{1+\frac{1}{r}}$ is at least

$$\mathbb{P}[\mathcal{X}_k^{t_1} | \mathcal{X}_k^0] \geq \mathbb{P}[\mathcal{X}_k^{t_1} | \mathcal{Y}_k^{t_0}] \cdot \mathbb{P}[\mathcal{Y}_k^{t_0} | \mathcal{X}_k^0] \geq n^{-\frac{7(r-1)}{c_r}} \cdot n^{-\frac{2^{8r+2}}{c_r}} \geq n^{-\frac{2^{8r+3}}{c_r}}.$$

We conclude that the probability of not finding a percolating subset of size $t_1 = (\ln n)^{1+\frac{1}{r}}$ in each of the $n / \left(2(\ln n)^{1+\frac{1}{r}}\right)$ rounds is at most

$$\left(1 - n^{-\frac{2^{8r+3}}{c_r}}\right)^{\frac{n}{2(\ln n)^{1+\frac{1}{r}}}} \leq \exp\left(-\frac{n^{1-\frac{2^{8r+3}}{c_r}}}{2(\ln n)^{1+\frac{1}{r}}}\right) \leq \exp(-\sqrt{n}).$$

These inequalities hold since $c_r \geq 2^{8r+5}$, provided n is large enough compared to c_r . \square

Remark. We note that as r becomes larger, Algorithm 6 has a harder time constructing a percolating set larger than $\ln n$. While for two colours we reach size $(\ln n)^{\frac{3}{2}}$ whp, for r colours we must settle for size $(\ln n)^{1+\frac{1}{r}}$.

3.3. Part II

In this subsection we aim to prove that conditioned on the existence of a percolating set of size t_1 in $G^{(1)}$, whp there is a percolating set of size at least $\frac{n}{2^{r+2}}$ in $G^{(1)} \cup G^{(2)}$ (see Lemma 17).

We will attempt to construct a percolating set of linear size with the following algorithm:

Algorithm 14 (The doubling algorithm).

Input: an r -fold graph $G^{(2)}$ and a subset X_0 which is percolating with respect to $G^{(1)}$.

For $t \geq 0$, we construct X_t inductively as follows:

- *Let $A_t := V \setminus X_t$ be the set of active vertices.*

(1) *At step $t \geq 0$ we reveal all edges of $G^{(2)}$ between A_t and $X_t \setminus X_{t-1}$, where $X_{-1} := \emptyset$. We define*

- $B_t := \{v \in A_t : \forall i \in [r] \exists v_i \in X_t \setminus X_{t-1} \text{ such that } vv_i \in E_i^{(2)}\}$.

In other words, B_t is the set of active vertices joined to $X_t \setminus X_{t-1}$ by an edge of each colour from the second round of exposure.

(2) *If $|B_t| < |X_t|$ we STOP. Otherwise, we set*

- $X_{t+1} := X_t \cup B_t$,
- $A_{t+1} := A_t \setminus B_t$.

If $|X_{t+1}| \geq n/2^{r+2}$ then STOP, otherwise go to (1) for step $t + 1$.

We set $b_t := |B_t|$ and $x_t := |X_t|$ for all t .

Remark.

- (i) If we reach step $t + 1$ in Algorithm 14, then $b_s \geq x_s$ for every $s \in [t]$ and therefore

$$x_s = x_{s-1} + b_{s-1} \geq 2x_{s-1} \text{ for every } s \in [t + 1].$$

Thus $b_t \geq x_t \geq 2x_{t-1} \geq 2^2x_{t-2} \geq \dots \geq 2^t x_0 = 2^t t_1$.

- (ii) If we reach step $t + 1$, then

$$(8) \quad x_{t+1} = b_t + x_t \leq 2b_t.$$

Let $t_2 := \max \{t \in \mathbb{N} \cup \{0\} : x_t < \frac{n}{2^{r+2}}\}$. Note that if Algorithm 14 constructs a percolating set X_t of size $\geq \frac{n}{2^{r+2}}$, then it will stop at time $t = t_2 + 1$; otherwise it will stop at time t_2 . Furthermore, by the previous remark we know that $2^{t_2} \leq \frac{b_{t_2}}{t_1} \leq n$, so

$$(9) \quad t_2 \leq \log_2(n) \leq 2 \ln n.$$

Given an r -fold graph \mathbf{G} , we denote the event that V contains a percolating subset of size at least m by $\mathcal{E}(\mathbf{G}, m)$. The general idea to prove the main result of this section (Lemma 17) is as follows: we first prove in

Claim 15 that the expected number of “suitable” vertices B_t is at least twice the size of the percolating set X_t constructed in step $t - 1$ (see Steps 1&2 of Algorithm 14). Subsequently, in Lemma 16 we prove a lower bound on the conditional probability that Algorithm 14 proceeds to step $t + 1$ conditioned on it reaching step t . Finally, we apply this lower bound multiple times to obtain Lemma 17.

Claim 15. *Let $t \leq t_2$. Then*

$$\mathbb{E}[b_t] \geq 2x_t.$$

Proof. Let $q_{t,i}$ denote the probability that a vertex $v \in A_t$ is joined to $B_{t-1} = X_t \setminus X_{t-1}$ by at least one edge of $G_i^{(2)}$. From (8) we know that $b_{t-1} \geq x_t/2$ for $0 \leq t \leq t_2$, where $b_{-1} := x_0$, and so we obtain

$$\begin{aligned} q_{t,i} &= 1 - \left(1 - \frac{p_i}{3}\right)^{b_{t-1}} \geq 1 - \left(1 - \frac{p_i}{3}\right)^{x_t/2} \\ (10) \qquad &\geq 1 - \exp\left(-\frac{p_i x_t}{6}\right) \geq \begin{cases} \frac{p_i x_t}{12} & \text{if } p_i x_t \leq 6; \\ \frac{1}{2} & \text{otherwise.} \end{cases} \end{aligned}$$

Let $j_t := \max\{j \in [r] \cup \{0\} : p_j x_t \leq 6\} \geq 0$, where $p_0 := 0$. Recalling that $A_t \geq n/2$ for $t \leq t_2$, we obtain

$$\begin{aligned} \mathbb{E}[b_t] &= |A_t| \left(\prod_{j=1}^r q_{t,j}\right) \stackrel{(10)}{\geq} \frac{n}{2} \left(\prod_{j=1}^{j_t} \frac{p_j x_t}{12}\right) \left(\frac{1}{2}\right)^{r-j_t} \\ &= \begin{cases} 2 \left(\frac{n}{2^{r+2}}\right) \geq 2x_t & \text{for } j_t = 0; \\ n \left(\frac{x_t}{3}\right)^{j_t} P_{j_t} \left(\frac{1}{2}\right)^{r+j_t+1} & \text{otherwise.} \end{cases} \end{aligned}$$

Thus we may assume that $j_t \geq 1$ (otherwise we are done). Making a further case distinction we obtain

Case 1: $j_t = 1$. We recall that $P_1 = p_1 \geq \frac{C_r \ln n}{n}$, thus for n large enough we have:

$$\mathbb{E}[b_t] \geq \left(\frac{C_r \ln n}{2^{r+2}}\right) \frac{x_t}{3} \geq 2x_t,$$

since $\frac{C_r}{2^{r+2}} \geq 1$.

Case 2: $j_t \geq 2$. We recall that $P_i = p_1 \dots p_i \geq C_r / (n(\ln n)^{i-1})$ for all $2 \leq i \in [r]$ and $x_t \geq 2^t t_1$ for all $0 \leq t \leq t_2$. Thus

$$\begin{aligned} \mathbb{E}[b_t] &\geq \frac{n}{2^{r+j_t+1}} \frac{(2^t t_1)^{j_t-1}}{3^{j_t}} x_t \left(\frac{C_r}{n(\ln n)^{j_t-1}} \right) = C_r \frac{2^{t(j_t-1)}}{2^{r+j_t+1} 3^{j_t}} \left(\frac{t_1}{\ln n} \right)^{j_t-1} x_t \\ &\geq \frac{C_r}{2^{2r+1} 3^r} x_t \geq 2x_t, \end{aligned}$$

where the last two inequalities are valid since $\frac{t_1}{\ln n} \geq 1$ and $C_r \geq 2^{8r^2} \geq 2^{2r+2} 3^r$. \square

We apply Claim 15 to bound the probability that we are able to double the size of the percolating set in each step.

Lemma 16. *For each integer $1 \leq t \leq t_2$, we have*

$$\mathbb{P}[b_t \geq x_t | X_t \neq \emptyset] \geq 1 - \exp\left(-\frac{t_1}{4}\right).$$

Proof. For $t \leq t_2$, the trial set X_t is of size at most $n/2^{r+2}$. This means that there are at least $n - n/2^{r+2} \geq n/2$ vertices in the set of active vertices A_t .

We note that the events that $v \in B_t$ are independent for different $v \in A_t$, so b_t is distributed as $\text{Bi}(|A_t|, q_{t,1}q_{t,2}\dots q_{t,r})$. Note that the distribution of b_t is dependent on both $|A_t|$ and $b_{t-1} = |X_t \setminus X_{t-1}|$. In what follows we will suppress the conditioning on these two variables for ease of notation.

Now the Chernoff bound (see e.g. [1]) tells us that

$$(11) \quad \mathbb{P}[\text{Bi}(m, q) \leq (1 - \delta)mq] \leq \exp\left(-\frac{mq\delta^2}{2}\right) \quad \text{for all } 0 < \delta < 1.$$

From Claim 15 we deduce that

$$\begin{aligned} \mathbb{P}[b_t \geq x_t | X_t \neq \emptyset] &\geq \mathbb{P}\left[b_t > \frac{\mathbb{E}[b_t]}{2}\right] \stackrel{(11)}{\geq} 1 - \exp(-\mathbb{E}[b_t]/8) \\ &\geq 1 - \exp(-x_t/4) \geq 1 - \exp(-x_0/4). \end{aligned}$$

Recalling that $x_0 = t_1$, this completes the proof. \square

We apply Lemma 16 multiple times to obtain the main result of this section.

Lemma 17. *For n large enough,*

$$\mathbb{P} \left[\mathcal{E} \left(G^{(1)} \cup G^{(2)}, \frac{n}{2^{r+2}} \right) \mid \mathcal{E} \left(G^{(1)}, (\ln n)^{1+\frac{1}{r}} \right) \right] \geq 1 - \exp\left(-\frac{t_1}{5}\right).$$

Proof. Since $t_2 \leq 2 \ln n$ by (9), we deduce from Lemma 16 that

$$\begin{aligned} \mathbb{P} \left[\mathcal{E} \left(G^{(1)} \cup G^{(2)}, n/2^{r+2} \right) \mid \mathcal{E} \left(G^{(1)}, t_1 \right) \right] &\geq \prod_{t=0}^{t_2} \mathbb{P} [b_t \geq x_t \mid X_t \neq \emptyset] \\ &\geq \left(1 - \exp \left(-\frac{t_1}{4} \right) \right)^{2 \ln n} \\ &\geq 1 - 2(\ln n) \exp \left(-\frac{t_1}{4} \right) \\ &= 1 - \exp \left(\ln(2 \ln n) - \frac{t_1}{4} \right) \\ &\geq 1 - \exp \left(-\frac{t_1}{5} \right), \end{aligned}$$

where the last inequality is valid since $t_1 = \Omega(\ln n)$. □

3.4. Part III

Finally we prove that $\mathbf{G}^* := G^{(1)} \cup G^{(2)} \cup G^{(3)}$ percolates whp.

Lemma 18. *Conditioned on $G^{(1)} \cup G^{(2)}$ containing a percolating subset X of size at least $n/2^{r+2}$, \mathbf{G}^* percolates whp.*

Indeed, we will prove that whp every vertex in $V \setminus X$ is connected to X by edges of every colour by using the final round of exposure $G^{(3)}$.

Proof of Lemma 18. We begin by defining \mathcal{K} to be the event that there is at least one vertex $v \in V \setminus X$ and one colour $i \in [r]$ such that $E_i^{(3)}$ contains no edge between v and X . Thus

$$\begin{aligned} \mathbb{P}[\mathcal{K}] &\leq \sum_{i \in [r]} \sum_{v \in V \setminus X} \left(1 - \frac{p_i}{3} \right)^{|X|} \leq rn \left(1 - \frac{p_1}{3} \right)^{\frac{n}{2^{r+2}}} \leq rne^{-\frac{C_r}{3 \cdot 2^{r+2}} \ln n} \\ &= rn^{1-\frac{C_r}{3 \cdot 2^{r+2}}} \leq \frac{r}{n} = o(1), \end{aligned}$$

where the last inequality holds since $C_r \geq 2^{8r^2} \geq 3 \cdot 2^{r+3}$. Since $\bar{\mathcal{K}}$ implies that \mathbf{G}^* percolates, this completes the argument. □

4. Concluding remarks

A number of open questions naturally present themselves.

4.1. Optimising C_r

Similar to Bollobás, Riordan, Slivken and Smith [5], we made no attempt to optimise the constant C_r in Theorem 2. As a result, the bounds on P_r for the subcritical and supercritical case are a long way apart. It is natural to expect them to be asymptotically equal, leading to the following strengthening of Theorem 2 (in the case when r is constant):

Conjecture 19. Let $r \in \mathbb{N}$. There exist constants C_2^*, \dots, C_r^* such that the following holds: suppose that p_1, \dots, p_r are functions of n such that $0 \leq p_1 \leq p_2 \leq \dots \leq p_r \leq 1$ and $\mathbf{G} = \mathbf{G}(n, p_1, p_2, \dots, p_r)$. For $i \in [r]$ let $P_i := p_1 p_2 \dots p_i$. Then for any constant $\varepsilon > 0$:

- (i) If $P_i \leq \frac{(1-\varepsilon)C_i^*}{n(\ln n)^{i-1}}$ for some $2 \leq i \leq r$ or $P_1 \leq \frac{(1-\varepsilon)\ln n}{n}$ then whp \mathbf{G} does not percolate.
- (ii) If $P_i \geq \frac{(1+\varepsilon)C_i^*}{n(\ln n)^{i-1}}$ for every $2 \leq i \leq r$ and $P_1 \geq \frac{(1+\varepsilon)\ln n}{n}$, then whp \mathbf{G} percolates.

The case $r = 1$ is simply the graph case, where the classical result of Erdős and Rényi shows that the conjecture is true. In the case when $r = 2$, the conjecture was proved in [8], and the value of C_2^* was determined to be $1/4$. It would be interesting to determine whether this conjecture is true, and if so to determine the exact values of the C_i^* for $i \geq 3$, and indeed to generalise to the case when r is allowed to grow as a function of n .

4.2. Size of the critical window

If the C_i^* can be determined precisely, the next parameter to optimise would be the parameter ε in Conjecture 19. More precisely, does the result still hold if rather than ε being a constant it is allowed to be a function of n which tends to 0 sufficiently slowly. This has already been extensively studied in the case $r = 1$, which corresponds to connectedness of the graph, but is an open problem in general.

4.3. Speed of the jigsaw process

In the supercritical case of Theorem 2, we know that whp the jigsaw percolation algorithm will terminate with just one cluster, but how many steps does this process require?

More precisely, in each step we create an auxiliary graph on the clusters of vertices, with an edge between clusters if there are edges between them of every colour in the r -fold graph, and merge each connected component of this auxiliary graph. How many iterations of this process are required before we have one single remaining cluster?

An analysis of the proof shows that, for the random graphs considered in the supercritical case, whp at most $(1 + o(1))(\ln n)^{1+1/r}$ steps are required. However, this was not optimised and it would be natural to conjecture that actually $\Theta(\ln n)$ steps are sufficient. It would also be interesting to determine the constant in this $\Theta(\ln n)$ term, which would most likely be dependent on how close the probability product P_r is to the jigsaw percolation threshold.

Acknowledgments

We would like to thank Christoph Koch for his helpful comments on an earlier draft of this paper, including pointing out the simple proof of Claim 12.

References

- [1] N. Alon, J. H. Spencer, The probabilistic method (Second Edition), *John Wiley & Sons* (2000). [MR1885388](#)
- [2] P. Ball, Crowd-sourcing: Strength in numbers, *Nature* **506**, 422–423 (2014).
- [3] A. L. Barabási, H. Jeong, Z. Néda, E. Ravasz, A. Schubert, and T. Vicsek, Evolution of the social network of scientific collaborations, *Phys. A* **311**, 590–614 (2002). [MR1943379](#)
- [4] B. Bollobás, O. Cooley, M. Kang and C. Koch, Jigsaw percolation on random hypergraphs, *Journal of Applied Probability* **54**, no. 4, 1261–1277 (2017). [MR3731294](#)
- [5] B. Bollobás, O. Riordan, E. Slivken, P. Smith, The threshold for jigsaw percolation on random graphs, *The Electronic Journal Of Combinatorics* **24**, no. 2 (2017). [MR3665569](#)
- [6] C. D. Brummitt, S. Chatterjee, P. S. Dey, and D. Sivakoff, Jigsaw percolation: What social networks can collaboratively solve a puzzle? *The Annals of Applied Probability* **25**, no. 4, 2013–2038 (2015). [MR3349000](#)
- [7] O. Cooley, M. Kang, and C. Koch, Threshold and hitting time for high-order connectedness in random hypergraphs, *Electron. J. Combin.* **23**, P48 (2016). [MR3512670](#)

- [8] O. Cooley, T. Kapetanopoulos, and T. Makai, The sharp threshold for jigsaw percolation in random graphs, *Advances in Applied Probability* **51**, no. 2, 378–407 (2019). [MR3989519](#)
- [9] P. Erdős, A. Rényi, On random graphs, *Publicationes Mathematicae Debrecen* **6**, 290–297 (1959). [MR0120167](#)
- [10] A. Gutierrez, Multi-coloured Jigsaw Percolation On Random Graphs, *Master's Thesis, TUGraz* (2017).
- [11] J. Gravner, D. Sivakoff, Nucleation scaling in jigsaw percolation, *Annals of Applied Probability* **27**, no. 1, 395–438 (2017). [MR3619791](#)
- [12] M. E. J. Newman, The structure of scientific collaboration networks, *Proc. Natl. Acad. Sci. USA* **98**, 404–409 (2001). [MR1812610](#)
- [13] M. E. J. Newman Scientific collaboration networks. I and II, *Phys. Rev. E* **64**, 016131, 016132 (2001). [MR2676073](#)
- [14] J. Tebbe, Where good ideas come from: The natural history of innovation, *Journal of Psychological Issues in Organizational Culture* **2**, 106–110 (2011).

OLIVER COOLEY
INSTITUTE OF DISCRETE MATHEMATICS
TU GRAZ
STEYRERGASSE 30
8010 GRAZ
AUSTRIA
E-mail address: cooley@math.tugraz.at

ABRAHAM GUTIERREZ
INSTITUTE OF DISCRETE MATHEMATICS
TU GRAZ
STEYRERGASSE 30
8010 GRAZ
AUSTRIA
E-mail address: a.gutierrez@math.tugraz.at

RECEIVED NOVEMBER 30, 2017