J. Appl. Prob. **54,** 1261–1277 (2017)
*doi:*10.1017/jpr.2017.62
© *Applied Probability Trust* 2017

JIGSAW PERCOLATION ON RANDOM HYPERGRAPHS

BÉLA BOLLOBÁS,* *University of Cambridge, University of Memphis, and London Institute for Mathematical Sciences*
OLIVER COOLEY,** ***
MIHYUN KANG ** **** AND
CHRISTOPH KOCH,***** *Graz University of Technology*

Abstract

The jigsaw percolation process on graphs was introduced by Brummitt *et al.* (2015) as a model of collaborative solutions of puzzles in social networks. Percolation in this process may be viewed as the joint connectedness of two graphs on a common vertex set. Our aim is to extend a result of Bollobás *et al.* (2017) concerning this process to hypergraphs for a variety of possible definitions of connectedness. In particular, we determine the asymptotic order of the critical threshold probability for percolation when both hypergraphs are chosen binomially at random.

Keywords: Jigsaw percolation; random graph; hypergraph; high-order connectedness; breadth-first search

2010 Mathematics Subject Classification: Primary 05C80
Secondary 60K35; 05C65; 60C05

1. Introduction and main result

Jigsaw percolation on graphs was introduced by Brummitt *et al.* [4] as a model for interactions within a social network. It was inspired by collaboration networks and the idea that advances are often the result of a group of people combining their ideas, skills, or knowledge to achieve better results than would have been possible individually, as seen in scientific research (see [1], [2], and [8]–[11]), among other fields.

Jigsaw percolation models such collaboration as a collective puzzle-solving process. The premise is that each of n people has a piece of a puzzle which must be combined in a certain way to solve the puzzle. Using the model, Brummit *et al.* [4] were able to show that some networks are better than others at solving certain puzzles, and that this is fundamentally due to the contrasting properties of the network. This motivates both a solid understanding of the jigsaw process and detailed knowledge of the social network in order to optimise performance and problem solving.

In the model there are two possibly overlapping sets of edges coloured red and blue defined on a common set of vertices. (In particular, any pair of vertices may form both a red and a blue edge at the same time.) Jigsaw percolation is a deterministic process on *clusters* of vertices

Received 29 March 2016; revision received 5 July 2017.
* Postal address: Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, Wilberforce Road, Cambridge CB3 0WA, UK. Email address: b.bollobas@dpmms.cam.ac.uk
** Postal address: Institute of Discrete Mathematics, Graz University of Technology, Steyrergasse 30, 8010 Graz, Austria.
*** Email address: cooley@math.tugraz.at
**** Email address: kang@math.tugraz.at
***** Current address: Mathematics Institute, University of Warwick, Zeeman Building, CV4 7AL Coventry, UK. Email address: c.koch@warwick.ac.uk

https://doi.org/10.1017/jpr.2017.62 Published online by Cambridge University Press

that evolves in discrete time. Initially, each vertex forms its own cluster and in each subsequent time-step two clusters merge if they are joined by at least one edge of each colour. The process stops when no two clusters can be merged. The jigsaw process *percolates* if we end in a single cluster. In particular, the process cannot percolate if either of the graphs given by blue or red edges is not connected.

More generally, given integers $1 \leq r \leq s$, define (r, s)-*jigsaw percolation* as follows. Let $G_1, \ldots, G_s$ be graphs on the same vertex set V. At each discrete time $t = 0, 1, \ldots$, we have a partition of V into *clusters*. At time $t = 0$ this is the finest partition: every vertex forms its own cluster. At time t, let $\widetilde{G}_t$ be the graph whose vertices are the clusters, with two vertices joined by an edge if the corresponding clusters are joined by an edge in at least r of the graphs G_i, $1 \leq i \leq s$. The clusters of the jigsaw process at time $t + 1$ are the unions of the clusters that belong to the same component of $\widetilde{G}_t$. The process *percolates* if eventually we arrive at a single cluster. Note that a $(1, s)$-process percolates if and only if the union of the graphs G_i is connected (and, in particular, the $(1, 1)$ process on a graph G percolates if and only if G is connected). On the other hand, if we have (s, s)-percolation then each G_i must be connected. However, the connectedness of each G_i is far from sufficient for (s, s)-percolation. Thus, we may view percolation of the (s, s) process as a notion of the 'joint connectedness' of the graphs $G_1, \ldots, G_s$, which requires some interaction between them rather than merely connectedness of each G_i or information about their union. So far, only $(2, 2)$-jigsaw percolation has been considered: in most of this paper, we shall do the same (and drop $(2, 2)$ from the notation), but we shall consider (s, s)-jigsaw percolation for larger s in Section 4.

Returning to the motivation, the blue graph may represent a *puzzle graph* of how the pieces of the puzzle may be combined to reach a solution, while the red graph may represent the *people graph*, modelling friendships between the people who hold the puzzle pieces.

Brummitt *et al.* [4] studied the model when the red graph is the binomial random graph and with various deterministic possibilities for the blue graph, including a Hamilton cycle, or other connected graphs of bounded maximum degree, and provided upper and lower bounds for the percolation threshold probabilities.

Gravner and Sivakoff [6] improved on these results for many different puzzle graphs of bounded degree, and also introduced a generalised process with redundancy parameters in the number of neighbours required for clusters to merge.

The setting in which both graphs are binomial random graphs was studied by Bollobás *et al.* [3], who determined the asymptotic order of the threshold for percolation in terms of the product of the two associated probabilities.

Before we state their result, we set the scene. Let $G(n, p_1)$ and $G(n, p_2)$ denote the pair of random graphs on the (common) vertex set $[n]$ (for $a \in \mathbb{N}$, we define $[a] := \{1, \ldots, a\}$), where each edge is present independently with probability p_1 or p_2, respectively. Throughout the paper any *unspecified* asymptotic is with respect to $n \to \infty$ and, in particular, we use the phrase *with high probability* (w.h.p.) to mean with probability tending to 1 as $n \to \infty$. For a probability distribution $\mathcal{P}$, we use the notation $X \sim \mathcal{P}$ to mean that X is a random variable with distribution $\mathcal{P}$. The main result of Bollobás *et al.* [3] is the following.

Theorem 1. (See [3].) *Let $0 \leq p_1 = p_1(n)$, $p_2 = p_2(n) \leq 1$ and let $G_1 \sim G(n, p_1)$ and $G_2 \sim G(n, p_2)$ independently. Then there exists a constant $c > 0$ such that:*

(i) *if $p_1 p_2 \leq \dfrac{1}{cn \log n}$ then w.h.p. (G_1, G_2) does not percolate;*

(ii) *if $p_1 p_2 \geq \dfrac{c}{n \log n}$ and $\min\{p_1, p_2\} \geq \dfrac{c \log n}{n}$ then w.h.p. (G_1, G_2) percolates.*

https://doi.org/10.1017/jpr.2017.62 Published online by Cambridge University Press

Our main aim in this paper is to extend this result to a hypergraph setting.

1.1. Notation and model

Following Erdős, we denote by $V^{(i)}$ the set of i-element subsets of a set V and call its elements *i-sets*.

Hypergraphs and high-order connectedness. Given an integer $k \geq 2$, a *k-uniform hypergraph* (abbreviated to *k-graph*) H consists of a set $V = V(H)$ of *vertices* and a set $E = E(H)$ of *edges*, where $E \subset V^{(k)}$. (Thus, for $k = 2$ this defines a graph.) There are several natural possibilities for the concept of connectedness of a hypergraph: here we define some of them. Given an integer $1 \leq j \leq k-1$, we say that two distinct *j-sets* $J, J' \in V^{(j)}$ are *j-tuple-connected* in H, abbreviated to *j-connected*, if there is a sequence of edges $e_1, \ldots, e_m \in E$ such that

- $J \subset e_1$ and $J' \subset e_m$;
- for all $1 \leq i \leq m-1$, some j-set $J_i \in V^{(j)}$ is contained in $e_i \cap e_{i+1}$.

In other words, we may walk from J to J' using edges which consecutively intersect in at least j vertices. Furthermore, any j-set is always j-connected to itself. (For a j-set which is not contained in any edge, this would not be implied by the previous definition.) A *j-component* of H is a maximal subset of $V^{(j)}$ such that any two of its elements are j-connected in H. We say that H is *j-connected* if all the j-sets of $V^{(j)}$ lie in a single j-component. Note that in the case $k = 2$ and $j = 1$ this is simply the usual definition of connectedness (and the vertex set of a component) of a graph.

High-order jigsaw percolation. A *double k-graph* (V, E_1, E_2) consists of a set $E_1 \subset V^{(k)}$ of *red* edges and a set $E_2 \subset V^{(k)}$ of *blue* edges on a common set V of vertices. We call the k-graphs (V, E_1) and (V, E_2) the *red k-graph* and the *blue k-graph*, respectively.

Let (V, E_1, E_2) be a double k-graph, let $\mathcal{J} \subset V^{(j)}$, and let $\mathcal{C}_0$ be a partition of $\mathcal{J}$ into $\kappa_0 := |\mathcal{C}_0|$ partition classes. (Unless stated otherwise, $\mathcal{C}_0$ will be the partition in which every j-set forms its own partition class and so $\kappa_0 = |\mathcal{J}|$.) Then *j-jigsaw percolation* is a deterministic, discrete-time process which is characterised by a sequence of partitions $\{\mathcal{C}_t\}_{t=0,1,\ldots}$ of $\mathcal{J}$. We write $\kappa_t := |\mathcal{C}_t|$ for the number of partition classes at time t and denote the partition classes by $\mathcal{C}_{t,1}, \ldots, \mathcal{C}_{t,\kappa_t}$. The partition $\mathcal{C}_{t+1}$ is obtained from $\mathcal{C}_t$ as follows.

(I) We define an auxiliary graph $\widetilde{G}_t$ on the vertex set $[\kappa_t]$, where the edge $\{i, i'\}$ is present if and only if there are j-sets $J_{i,1}, J_{i,2} \in \mathcal{C}_{t,i}$ and $J_{i',1}, J_{i',2} \in \mathcal{C}_{t,i'}$ and edges $e_1 \in E_1$ and $e_2 \in E_2$ such that $J_{i,1} \cup J_{i',1} \subset e_1$ and $J_{i,2} \cup J_{i',2} \subset e_2$.

(II) If $\widetilde{G}_t$ is an empty graph then STOP.

(III) Otherwise, we set κ_{t+1} to be the number of components of $\widetilde{G}_t$ and denote these components by $\widetilde{C}_1, \ldots, \widetilde{C}_{\kappa_{t+1}}$. The partition $\mathcal{C}_{t+1}$ is then given by

$$\mathcal{C}_{t+1,i} = \bigcup_{w \in \widetilde{C}_i} \mathcal{C}_{t,w} \quad \text{for } 1 \leq i \leq \kappa_{t+1}.$$

(IV) If $\kappa_{t+1} = 1$ then STOP.

(V) Otherwise, we proceed to time $t+1$.

https://doi.org/10.1017/jpr.2017.62 Published online by Cambridge University Press

If the process stopped in step (IV), we say that $\mathcal{J}$ *percolates on* (V, E_1, E_2); otherwise, i.e. if the process stopped in step (II), we say that $\mathcal{J}$ *does not percolate on* (V, E_1, E_2). We refer to the partition classes $\mathcal{C}_{t,i}$ as *clusters*.

The main results in this paper deal with the setting when $\mathcal{J} = V^{(j)}$. In this case we simply say that (V, E_1, E_2) *j-percolates* or (V, E_1, E_2) *does not j-percolate*. The generalised version is required in the proofs of our results.

Throughout the paper we will ignore floors and ceilings whenever this does not significantly affect the argument.

1.2. Main result

In this paper we consider the *random binomial double k-graph* $H_k(n, p_1, p_2)$ on vertex set $[n]$, where every k-set is present as a red edge with probability p_1 and present as a blue edge with probability p_2, independently of each other and of all other k-sets.

Similarly to the graph case, a necessary condition for j-jigsaw percolation on a double k-graph is that both the red and the blue k-graphs are j-connected. In this setting the red/blue k-graph is a copy of the *binomial random k-graph* $H_k(n, p_1)$ or $H_k(n, p_2)$, respectively, i.e. the vertex set is $[n]$ and each k-set is present independently with probability p_1 or p_2, respectively. It was shown in [5] that the (sharp) threshold for j-connectedness in $H_k(n, p)$ is

$$p_{\text{conn}} = p_{\text{conn}}(n) := \frac{j \log n}{\binom{n}{k-j}}. \tag{1}$$

As in Theorem 1, we impose a slightly stronger condition on $\min\{p_1, p_2\}$ to ensure the j-connectedness in the supercritical regime. We prove the following extension of Theorem 1.

Theorem 2. *For integers* $1 \le j \le k-1$, *let* $0 \le p_1 = p_1(n), p_2 = p_2(n) \le 1$ *and let* $H \sim H_k(n, p_1, p_2)$. *Then there exists a constant* $c = c(k, j) > 0$ *such that*

(i) *if* $p_1 p_2 \le \dfrac{1}{cn^{2k-2j-1} \log n}$ *then w.h.p. H does not j-percolate;*

(ii) *if* $p_1 p_2 \ge \dfrac{c}{n^{2k-2j-1} \log n}$ *and* $\min\{p_1, p_2\} \ge \dfrac{c \log n}{n^{k-j}}$ *then w.h.p. H j-percolates.*

In other words, the threshold $\hat{p} = \hat{p}(n)$ for j-jigsaw percolation on $H_k(n, p_1, p_2)$ in terms of the product $p = p_1 p_2$ is of order

$$\hat{p} = \Theta\left(\frac{1}{n^{2k-2j-1} \log n}\right).$$

Somewhat surprisingly, the subcritical case (Theorem 2(i)) does not seem to be easy to prove. The corresponding subcritical case for graphs (i.e. $k = 2$ and $j = 1$) was almost trivial, but the general case requires significantly more involved analysis.

On the other hand, the supercritical case (Theorem 2(ii)) becomes much easier; we prove a neat reduction to the graph case.

The methods that we apply in this paper can also be used, with minimal additional work, to prove some generalisations of Theorem 2 and related results. We present these results and outlines of the proofs in Section 4.

We will prove the subcritical case of Theorem 2 in Section 2, and the supercritical case in Section 3

https://doi.org/10.1017/jpr.2017.62 Published online by Cambridge University Press

2. Subcritical regime: proof of Theorem 2(i)

Let $1 \leq j \leq k-1$ be integers. As a first step, we develop necessary conditions for any double k-graph H on vertex set V to j-percolate. Then, in the second step, we show that if $H \sim H_k(n, p_1, p_2)$, where $p_1 p_2 \leq 1/(cn^{2k-2j-1}\log n)$ then w.h.p. it fails to satisfy the weakest of these necessary conditions.

2.1. Necessary conditions for j-percolation

The initial idea is to modify the algorithm (given by (I)–(V) in Section 1.1) slightly, so that for each cluster it additionally keeps track of two sets of edges, one red and one blue, which allow us to obtain this cluster by a sequence of merges.

In this spirit, we consider triples of the form $(\mathcal{J}_0, \mathcal{E}_1, \mathcal{E}_2)$, where $\mathcal{J}_0 \subset V^{(j)}$, $\mathcal{E}_1 \subset V^{(k)}$, and $\mathcal{E}_2 \subset V^{(k)}$, i.e. $\mathcal{E}_1$ and $\mathcal{E}_2$ are embedded into the complete double k-graph $(V, V^{(k)}, V^{(k)})$. We call the edges in $\mathcal{E}_1$ *red* and those in $\mathcal{E}_2$ *blue*; note that edges may be red and blue at the same time. The *size* of a triple is given by $|\mathcal{J}_0|$. Furthermore, we call $(\mathcal{J}_0, \mathcal{E}_1, \mathcal{E}_2)$ *internally spanned* if $\mathcal{J}_0$ percolates on the double k-graph $(V, \mathcal{E}_1, \mathcal{E}_2)$. Note that these triples play a crucial role for j-jigsaw percolation on double k-graphs (comparable to *internally spanned sets* in [3]).

Claim 1. *For every positive integer $N \leq \binom{|V|}{j}/\binom{k}{j}$, if a double k-graph H j-percolates then it must contain an internally spanned triple of size ℓ for some $N \leq \ell \leq \binom{k}{j}N$.*

Proof. We shall think of the edges of H arriving one at a time: whenever an edge arrives, we check whether it can be used to merge some clusters. Note that each edge e that arrives can be used to merge a cluster C with some others if e contains some j-set of C. Thus, with the arrival of an edge, at most $\binom{k}{j}$ clusters will be merged. Therefore, when the first cluster of size at least N forms, it was created by merging at most $\binom{k}{j}$ clusters, each of size at most $N-1$, and, therefore, it has size at most $\binom{k}{j}N$.

Then the elements of our internally spanned triple consist of the set $\mathcal{J}$ of j-sets in the cluster and the sets $\mathcal{E}_1, \mathcal{E}_2$ of red and blue edges, respectively. Note that this may contain many more edges than necessary, but certainly $\mathcal{J}$ percolates on $\mathcal{E}_1$ and $\mathcal{E}_2$, and, therefore, $(\mathcal{J}, \mathcal{E}_1, \mathcal{E}_2)$ is an internally spanned triple of the required size. □

Counting all possible internally spanned triples of a given size does not seem easy. Instead, we consider a relaxation of internally spanned triples. To do so, we first define a strengthened notion of j-connectedness on k-graphs. Let (V, E) be a k-graph and let $\mathcal{J} \subset V^{(j)}$ be a collection of j-sets. We say that a subset $\mathcal{J}^* \subset \mathcal{J}$ is *$\mathcal{J}$-traversable* (in (V, E)) if, for every two distinct j-sets $J, J' \in \mathcal{J}^*$, there is a sequence of edges $e_1, \ldots, e_m \in E$ such that

- $J \subset e_1$ and $J' \subset e_m$;
- for all $1 \leq i \leq m-1$, some j-set $J_i \in \mathcal{J}$ is contained in $e_i \cap e_{i+1}$.

In other words, we may walk from J to J' using edges such that the intersection of two consecutive edges contains at least one j-set from $\mathcal{J}$. Furthermore, any singleton $\{J\} \subset \mathcal{J}$ is $\mathcal{J}$-traversable. To shorten notation, a collection $\mathcal{J}$ is called *traversable* if it is $\mathcal{J}$-traversable. Note that $V^{(j)}$-traversable collections of j-sets are precisely j-connected collections. Now we say that a pair $(\mathcal{J}, \mathcal{E})$ is *traversable* if $\mathcal{J}$ is traversable in the k-graph $(V, \mathcal{E})$. Furthermore, a triple $(\mathcal{J}_0, \mathcal{E}_1, \mathcal{E}_2)$ is *traversable* if both $(\mathcal{J}_0, \mathcal{E}_1)$ and $(\mathcal{J}_0, \mathcal{E}_2)$ are traversable, i.e. if $\mathcal{J}_0$ is traversable in both the red and blue k-graphs, $(V, \mathcal{E}_1)$ and $(V, \mathcal{E}_2)$, respectively.

https://doi.org/10.1017/jpr.2017.62 Published online by Cambridge University Press

Fact 1. *Every internally spanned triple is traversable and, thus, contains an edge-minimal traversable triple.*

Similarly to internally spanned triples, the *size* of a traversable triple $(\mathcal{J}, \mathcal{E}_1, \mathcal{E}_2)$ is defined to be $|\mathcal{J}|$. Let us denote the set of all edge-minimal traversable triples of size ℓ within the complete double k-graph on vertex set V by $\mathcal{T}_\ell$. Furthermore, we partition $\mathcal{T}_\ell$ according to the number of red and blue edges: for all integers ℓ, r, b, we define

$$\mathcal{T}_{\ell,r,b} := \{(\mathcal{J}_0, \mathcal{E}_1, \mathcal{E}_2) \in \mathcal{T}_\ell \mid |\mathcal{E}_1| = r \text{ and } |\mathcal{E}_2| = b\}.$$

To summarise, we have established a necessary condition for percolation of an arbitrary double k-graph H: by Claim 1 and Fact 1, the existence of a triple $T \in \mathcal{T}_{\ell,r,b}$ of size $\log n \leq \ell \leq \binom{k}{j} \log n$ is necessary for H to percolate. Aiming for a first moment argument, we will therefore provide an upper bound on $|\mathcal{T}_{\ell,r,b}|$.

2.2. Edge-minimal traversable triples

Our next goal is to establish a general upper bound on $|\mathcal{T}_{\ell,r,b}|$ in an arbitrary double hypergraph. This will be carried out in three steps: firstly, we observe that, for every $T \in \mathcal{T}_{\ell,r,b}$, its number of edges is bounded in terms of its size (Section 2.2.1); secondly, we characterise every $T \in \mathcal{T}_{\ell,r,b}$ by a pair of matrices, which we call its *blueprint* (Section 2.2.2); and, thirdly, we use these matrices to algorithmically construct a superset of $\mathcal{T}_{\ell,r,b}$ and bound its size from above (Section 2.2.3).

2.2.1. *Number of edges.* Crucially, the number of edges in each edge-minimal traversable triple is bounded as a function of its size.

Claim 2. *Every edge-minimal traversable triple of size* $1 \leq \ell \leq \binom{|V|}{j}$ *contains at most* $\ell - 1$ *edges of each colour. In particular,* $|\mathcal{T}_{\ell,r,b}| = 0$ *unless* $0 \leq r, b \leq \ell - 1$.

Proof. By the definition of traversable triples, it suffices to prove that every edge-minimal traversable pair $(\mathcal{J}, \mathcal{E})$ of size $1 \leq |\mathcal{J}| \leq \binom{|V|}{j}$ satisfies $|\mathcal{E}| \leq |\mathcal{J}| - 1$.

We proceed by induction over $|\mathcal{J}|$; for $|\mathcal{J}| = 1$, the statement is trivial. So let $|\mathcal{J}| \geq 2$ and assume that the statement holds for all $1 \leq i \leq |\mathcal{J}|-1$. Let $e \in \mathcal{E}$ be an arbitrary edge. By edge-minimality, the collection $\mathcal{J}$ is not traversable in the k-graph $(V, \mathcal{E} \setminus \{e\})$, i.e. there exists a partition $\mathcal{J} = X_1 \dot\cup \cdots \dot\cup X_m$ with $m \geq 2$ and $1 \leq |X_1|, \ldots, |X_m| \leq |\mathcal{J}| - 1$ such that no edge $e' \in \mathcal{E} \setminus \{e\}$ contains a j-set from more than one partition class. Moreover, by edge-minimality, each edge in $\mathcal{E}$ contains at least one j-set from $\mathcal{J}$, and, thus, we obtain a partition of the edge set $\mathcal{E} \setminus \{e\} = Y_1 \dot\cup \cdots \dot\cup Y_m$ by setting $Y_i := \{e' \in \mathcal{E} \mid \text{ there exists } J \in X_i \colon J \subset e'\}$. Since (X_i, Y_i) is an edge-minimal traversable pair and $|X_i| \leq |\mathcal{J}| - 1$, we have, by the induction hypothesis, $|Y_i| \leq |X_i| - 1$ and by summing over all $1 \leq i \leq m$, we obtain $|\mathcal{E}| - 1 \leq |\mathcal{J}| - m \leq |\mathcal{J}| - 2$, as required. If, for some $e^* \in Y_i$, the pair $(X_i, Y_i \setminus \{e^*\})$ was traversable then $(\mathcal{J}, \mathcal{E} \setminus \{e^*\})$ would be traversable as well, contradicting edge-minimality. □

2.2.2. *Blueprints.* Next, in order to extract further structural information from edge-minimal traversable triples, we fix an arbitrary pair $\sigma = (\sigma_j, \sigma_k)$ of total orders. Here σ_j is an order on $V^{(j)}$ and σ_k on $V^{(k)}$. Clearly, σ induces a total order on each of the sets $\mathcal{J}_0$, $\mathcal{E}_1$, and $\mathcal{E}_2$.

Given a triple $T = (\mathcal{J}_0, \mathcal{E}_1, \mathcal{E}_2) \in \mathcal{T}_{\ell,r,b}$, we explore $\mathcal{J}_0$ via a breadth-first search (BFS) using only red edges, and colouring j-sets white once they have been discovered. More precisely, we study the following exploration process starting from the minimal element $J_{(1)}$ of $\mathcal{J}_0$ with respect to σ. Initially, we colour $J_{(1)}$ white and make it active. In each subsequent step,

https://doi.org/10.1017/jpr.2017.62 Published online by Cambridge University Press

there is an active j-set, say J, and we consider all edges of $\mathcal{E}_1$ containing J which have not been considered previously. For each of these edges in turn (according to σ), we consider all the j-sets of $\mathcal{J}_0$ that it contains and colour them white, if they are not yet coloured white, in the order given by σ. We have then finished exploring J and move on to the next j-set which was coloured white in this process. Since $\mathcal{J}_0$ is traversable in $(V, \mathcal{E}_1)$, this induces a new total order on $\mathcal{J}_0$ which we call the *BFS-order* of T (with respect to σ) and denote by $\tau_\sigma = \tau_\sigma(T) := (J_{(1)}, \dots, J_{(|\mathcal{J}_0|)})$. (Note that τ_σ will, in general, be different from the order induced by σ on $\mathcal{J}_0$.)

Additionally, we introduce *marks* on the j-sets of $\mathcal{J}_0$. As we shall see, marking a j-set in the blue process is similar to colouring it white in the red process. Initially, $J_{(1)}$ is marked and all other j-sets are unmarked. We then go through $\mathcal{J}_0$ according to τ_σ. In the ith step, i.e. when $J_{(i)}$ is active, we reveal all blue edges that contain $J_{(i)}$ and none of $\{J_{(i+1)}, \dots, J_{(|\mathcal{J}_0|)}\}$, one by one according to σ. Whenever we reveal a blue edge in this way, we mark all the still unmarked j-sets in $\mathcal{J}_0$ that it contains.

The reason for colouring and marking the j-sets in $\mathcal{J}_0$ is that it allows us to additionally keep track of two sets of parameters, $r_{i,z}$ and $b_{i,z}$, for $1 \le i \le |\mathcal{J}_0|$ and $0 \le z \le \binom{k}{j}$. We say that a red (or blue) edge performs *z-duty* if we colour (respectively, mark) precisely z many j-sets when it is revealed. Then $r_{i,z}$ denotes the number of red z-duty edges that were revealed while $J_{(i)}$ was active. Similarly, $b_{i,z}$ denotes the number of blue z-duty edges that were revealed when $J_{(i)}$ was coloured white. We encode this information in matrices $R := (r_{i,z})_{i,z}$ and $B := (b_{i,z})_{i,z}$ and call the pair $\pi_\sigma(T) := (R, B)$ the *blueprint* of T with respect to σ.

For our upcoming arguments we introduce some notation for such matrices. Given a positive integer a and an $a \times (\binom{k}{j} + 1)$ matrix $M = (m_{i,z})_{i,z}$ with nonnegative integer entries, we define

$$f(M) := \sum_{i=1}^{a} \sum_{z=0}^{\binom{k}{j}} z m_{i,z} \quad \text{and} \quad g(M) := \sum_{i=1}^{a} \sum_{z=0}^{\binom{k}{j}} m_{i,z}.$$

Furthermore, for a nonnegative integer m, we define

$$\mathcal{M}_{a,m} := \left\{ M \in \mathbb{Z}_{\ge 0}^{a \times (\binom{k}{j}+1)} \;\middle|\; f(M) = a - 1 \text{ and } g(M) = m \right\}.$$

(We shall only ever use these definitions with $a = \ell$ and m either b or r.) The intuition behind these definitions is the following: f allows us to reconstruct the total number of white (or marked) j-sets (apart from $J_{(1)}$) from the blueprint of a triple, while g provides the number of red (respectively, blue) edges. More formally, this is stated in the following lemma.

Lemma 1. *Let $\ell > r, b \ge 0$ be integers and σ be a pair of total orders on $V^{(j)}$ and $V^{(k)}$. For any $T = (\mathcal{J}_0, \mathcal{E}_1, \mathcal{E}_2) \in \mathcal{T}_{\ell,r,b}$, the blueprint $\pi_\sigma(T) = (R, B)$ satisfies*

- $f(R) = f(B) = |\mathcal{J}_0| - 1 = \ell - 1$;
- $g(R) = |\mathcal{E}_1| = r$;
- $g(B) = |\mathcal{E}_2| = b$.

In other words, $\pi_\sigma\left(\mathcal{T}_{\ell,r,b}\right) \subset \mathcal{M}_{\ell,r} \times \mathcal{M}_{\ell,b}$.

https://doi.org/10.1017/jpr.2017.62 Published online by Cambridge University Press

Proof. Let $T = (\mathcal{J}_0, \mathcal{E}_1, \mathcal{E}_2)$ and recall that $\mathcal{J}_0$ contains precisely $\ell - 1$ many j-sets excluding $J_{(1)}$ (which is already initially both white and marked) and is traversable in both the red and blue k-graphs $(V, \mathcal{E}_1)$ and $(V, \mathcal{E}_2)$, respectively. Therefore, each j-set in $\mathcal{J}_0 \setminus \{J_{(1)}\}$ is coloured white by at least one red edge and likewise receives a mark from at least one blue edge. On the other hand, no j-set is coloured or marked more than once. Since $f(R)$ counts the number of j-sets coloured white by red edges and $f(B)$ counts the number of marks given by blue edges, the first statement follows.

On the other hand, recall that T contains r red edges and b blue edges. Since T is edge-minimal, any edge, either red or blue, was revealed in the process and, thus, was counted in precisely one of the $r_{i,z}$ and one of the $b_{i,z}$. Thus, the second and third statements follow. □

Next we give a (crude) upper bound on the number of such matrices.

Claim 3. *Let $a > m \geq 0$ be integers. If $\mathcal{M}_{a,m} \neq \varnothing$ then*

$$m \geq \binom{k}{j}^{-1} (a-1).$$

Furthermore, there is a constant $C' = C'(k, j) > 0$ (independent of a, m) such that

$$|\mathcal{M}_{a,m}| \leq (C')^{a-1}.$$

Note that for edge-minimal traversable triples, certainly the number of j-sets is larger than the number of edges of each colour by Claim 2. Hence, the condition $a > m$ is fulfilled whenever we want to apply this result in this paper.

Proof. For the first statement, let M be any matrix in $\mathcal{M}_{a,m}$. Then

$$a - 1 = f(M) = \sum_{i=1}^{a} \sum_{z=0}^{\binom{k}{j}} z m_{i,z} \leq \binom{k}{j} \sum_{i=1}^{a} \sum_{z=0}^{\binom{k}{j}} m_{i,z} = \binom{k}{j} m. \tag{2}$$

For the second statement, first note that all entries of a matrix in $\mathcal{M}_{a,0}$ must be 0, hence $|\mathcal{M}_{a,0}| \leq 1$. Therefore, let us assume that $m \geq 1$. Choosing an arbitrary $a \times (\binom{k}{j} + 1)$ matrix M with nonnegative integer entries satisfying $g(M) = m$ can be seen as having a set of m indistinguishable elements and partitioning it into

$$a\left(\binom{k}{j} + 1\right) \overset{(2)}{\leq} 2m\binom{k}{j}^2$$

potentially empty distinguishable partition classes. Thus, (using the fact that there are

$$\binom{t+s-1}{t-1} = \binom{t+s-1}{s}$$

ways of partitioning s indistinguishable elements into t distinguishable classes) we certainly obtain the upper bound

$$|\mathcal{M}_{a,m}| \leq \binom{2\binom{k}{j}^2 m + m - 1}{m} \leq \left(\frac{3e\binom{k}{j}^2 m}{m}\right)^m \leq \left(9\binom{k}{j}^2\right)^{a-1},$$

as claimed. □

https://doi.org/10.1017/jpr.2017.62 Published online by Cambridge University Press

2.2.3. *Triples obtained from blueprints.* Given ℓ, r, and b, we provide an upper bound on $|\mathcal{T}_{\ell,r,b}|$. To this end, we construct a superset of $\mathcal{T}_{\ell,r,b}$: we first choose a blueprint $(R, B) \in \mathcal{M}_{\ell,r} \times \mathcal{M}_{\ell,b}$ and then construct all possible triples $(\mathcal{J}_0, \mathcal{E}_1, \mathcal{E}_2)$ with $|\mathcal{J}_0| = \ell$, $|\mathcal{E}_1| = r$, and $|\mathcal{E}_2| = b$ (within the complete double k-graph) such that its collection of white j-sets $\mathcal{J}_0$ is traversable in the red k-graph $(V, \mathcal{E}_1)$ (we do not demand that $\mathcal{J}_0$ is traversable in the blue k-graph $(V, \mathcal{E}_2)$, which means we will be overcounting). More precisely, we construct $(\mathcal{J}_0, \mathcal{E}_1, \mathcal{E}_2)$ using Algorithm 1 consisting of two phases: a white-and-red phase (WR) followed by a blue phase (B). Both WR and B terminate after ℓ steps, or if at any point we do not have a valid choice; in the latter case, the current instance of the procedure is discarded.

Algorithm 1. The algorithm proceeds as follows.
Input: $(R, B) \in \mathcal{M}_{\ell,r} \times \mathcal{M}_{\ell,b}$.
Output: $(\mathcal{J}_0, \mathcal{E}_1, \mathcal{E}_2) \subset V^{(j)} \times V^{(k)} \times V^{(k)}$.
Phase WR.
Input: $R \in \mathcal{M}_{\ell,r}$.
Output: $(\mathcal{J}_0, \mathcal{E}_1) \subset V^{(j)} \times V^{(k)}$ and a total order ϑ on $\mathcal{J}_0$.
Phase B.
Input: $(\mathcal{J}_0, \mathcal{E}_1)$ and ϑ from phase WR, $B \in \mathcal{M}_{\ell,b}$.
Output: $\mathcal{E}_2 \subset V^{(k)}$.

We first run phase WR as follows. Initially, we choose an arbitrary j-set $J_{(1)} \in V^{(j)}$ and colour it white, meaning that $\mathcal{J}_0 \leftarrow \{J_{(1)}\}$ and set $J_{(1)} \prec_\vartheta J_{(1)}$. No red edges are chosen yet, so $\mathcal{E}_1 \leftarrow \varnothing$. Then in step $i = 1, \dots, \ell$, we consider the j-set $J_{(i)} \in \mathcal{J}_0$ to be active. For each $z = 1, \dots, \binom{k}{j}$ and each $y = 1, \dots, r_{i,z}$, we do the following:

- choose an edge $e_{i,z,y} \in V^{(k)}$ containing $J_{(i)}$;
- if $e_{i,z,y} \notin \mathcal{E}_1$ then $\mathcal{E}_1 \leftarrow \mathcal{E}_1 \cup e_{i,z,y}$; otherwise, discard the instance;
- for $x = 1, \dots, z$,
  - we choose a j-set $J_{i,z,y,x} \subset e_{i,z,y}$;
  - if $J_{i,z,y,x} \notin \mathcal{J}_0$ then $\mathcal{J}_0 \leftarrow \mathcal{J}_0 \cup \{J_{i,z,y,x}\}$ and for all $J \in \mathcal{J}_0$ set $J \prec_\vartheta J_{i,z,y,x}$; otherwise, discard the instance.

If $i = \ell$, terminate WR with output $(\mathcal{J}_0, \mathcal{E}_1)$ and ϑ. Otherwise, set $J_{(i+1)}$ to be the minimal element of $\mathcal{J}_0$ with respect to ϑ such that $J_{(i)} \prec_\vartheta J$ and $J \not\prec_\vartheta J_{(i)}$ (if this minimal element does not exist, discard the instance), and proceed with step $i + 1$.

We observe that, for any instance that is not discarded, the output satisfies $|\mathcal{J}_0| = \ell$ and $|\mathcal{E}_1| = r$, since $R \in \mathcal{M}_{\ell,r}$, and in step i precisely $\sum_z z r_{i,z}$ many distinct j-sets are added to $\mathcal{J}_0$, and precisely $\sum_z r_{i,z}$ many distinct k-sets are added to $\mathcal{E}_1$. Moreover, we note that the ordering ϑ obtained in this way is simply the lexicographical order on $\mathcal{J}_0$ with respect to i, z, y, x, i.e. $J_{i,z,y,x} \prec_\vartheta J_{i',z',y',x'}$ if and only if $\{i < i'\}$ or $\{i = i'$ and $z < z'\}$ or $\{i = i',\ z = z',$ and $y < y'\}$ or $\{i = i',\ z = z',\ y = y'$ and $x \le x'\}$ (where $J_{(1)}$ is seen as $J_{0,1,1,1}$). This completes phase WR.

It remains to embed the blue edges in a fashion which is consistent with ϑ and B. Observe that it will actually be crucial for the argument, that the order ϑ is already fixed before we start embedding the blue edges.

We next run phase B as follows. We consider the ordered set $(\mathcal{J}_0, \prec_\vartheta)$ as the sequence $(J_{(1)}, J_{(2)}, \dots, J_{(\ell)})$, and initialise $\mathcal{E}_2 \leftarrow \varnothing$. In step $i = 1, \dots, \ell$, we consider the j-set $J_{(i)}$ to

https://doi.org/10.1017/jpr.2017.62 Published online by Cambridge University Press

be active. For every $z = 1, \dots, \binom{k}{j}$ and $y = 1, \dots, b_{i,z}$,

- choose an edge $e'_{i,z,y} \in V^{(k)}$ which contains $J_{(i)}$ and another $J_{(i^*)}$ with $1 \le i^* < i$;
- if $e'_{i,z,y} \notin \mathcal{E}_2$ then $\mathcal{E}_2 \leftarrow \mathcal{E}_2 \cup \{e'_{i,z,y}\}$; otherwise, discard the instance.

If $i = \ell$, terminate B with output $(\mathcal{J}_0, \mathcal{E}_1, \mathcal{E}_2)$. Otherwise, proceed with step $i + 1$.

If phase B terminates without being discarded then the output of Algorithm 1 consists of $(\mathcal{J}_0, \mathcal{E}_1, \mathcal{E}_2)$, where $\mathcal{J}_0$ and $\mathcal{E}_1$ are given in the output of phase WR and $\mathcal{E}_2$ is the output of phase B.

We observe that for any instance that does not get discarded, we have $|\mathcal{E}_2| = b$, since $B \in \mathcal{M}_{\ell,b}$ and in every step precisely $\sum_z b_{i,z}$ distinct k-sets are added to $\mathcal{E}_2$.

Note also that in terms of marking the j-sets, phase B is less restrictive than the construction of a traversable triple. More precisely, it does not actively mark the j-sets, meaning that the same j-sets might be re-used multiple times. Furthermore, we only insist on one previously seen j-set $J_{(i^*)}$ being present in the edge $e'_{i,z,y}$, where in fact there should be z (previously seen but unmarked) such j-sets.

Now let $\mathcal{Q}_{\ell,r,b}$ denote the set of all outputs of instances of Algorithm 1 that did not get discarded, i.e. every one of our choices in both phases was valid.

Lemma 2. *For all integers $\ell > r, b \ge 0$, we have $\mathcal{T}_{\ell,r,b} \subset \mathcal{Q}_{\ell,r,b}$; furthermore, there is a constant $C > 0$ independent of ℓ, r, b such that*

$$|\mathcal{Q}_{\ell,r,b}| \le |V|^j C^{\ell-1} (|V|^{k-j})^r (\ell |V|^{k-j-1})^b.$$

Proof. The first assertion is simple. Fix a pair σ of total orders on $V^{(j)}$ and $V^{(k)}$. Now consider a triple $T \in \mathcal{T}_{\ell,r,b}$ and note that, by Lemma 1, its blueprint satisfies $\pi_\sigma(T) = (R, B)$ for some $(R, B) \in \mathcal{M}_{\ell,r} \times \mathcal{M}_{\ell,b}$. Knowing T, σ, and, therefore, also $\tau_\sigma(T)$, it is straightforward to provide an instance of Algorithm 1 with input (R, B) and output T that does not get discarded.

For the upper bound on the total number of outputs, let us first fix any particular blueprint $(R, B) \in \mathcal{M}_{\ell,r} \times \mathcal{M}_{\ell,b}$. For the red edges, as far as an upper bound is concerned, in each step and for each z, we may choose $r_{i,z}$ elements from a set of at most $|V|^{k-j}$ with replacement and then there are at most $\binom{k}{j}^z$ ways to choose the white j-sets within any red edge. (Of course these bounds are quite crude in general, but sufficient for our result.) Since we have at most $|V|^j$ choices for the initial white j-set, we have at most

$$\begin{aligned} |V|^j \prod_{i,z} \left(\binom{k}{j}^z |V|^{k-j} \right)^{r_{i,z}} &= |V|^j \binom{k}{j}^{f(R)} (|V|^{k-j})^{g(R)} \\ &= |V|^j \binom{k}{j}^{\ell-1} (|V|^{k-j})^r \quad \text{(by Lemma 1)} \end{aligned}$$

instances which can be distinguished by their combined white and red structure.

Now consider the number of ways of choosing a blue edge together with the previously seen j-set $J_{(i^*)}$. A blue edge contains at least $j + 1$ already embedded vertices, for instance, those in $J_{(i)} \cup J_{(i^*)}$, and there are at most ℓ choices for $J_{(i^*)}$. Thus, we have at most $\ell |V|^{k-j-1}$ choices for this edge. Therefore, the number of choices for the blue edges is at most

$$\prod_{i,z} (\ell |V|^{k-j-1})^{b_{i,z}} = (\ell |V|^{k-j-1})^{g(B)} = (\ell |V|^{k-j-1})^b.$$

https://doi.org/10.1017/jpr.2017.62 Published online by Cambridge University Press

Finally, we have already counted the number of matrices $R \in \mathcal{M}_{\ell,r}$ and $B \in \mathcal{M}_{\ell,b}$ in Claim 3: there is a constant $C' > 0$ such that we have

$$|\mathcal{M}_{\ell,r}| \le (C')^{\ell-1} \quad \text{and} \quad |\mathcal{M}_{\ell,b}| \le (C')^{\ell-1}.$$

Combining this with the previous calculations provides the desired upper bound with $C = (C')^2\binom{k}{j}$. □

Remark 1. While the definition of (edge-minimal) traversable triples is symmetric in the two colours red and blue, the bound in Lemma 2 is not. However, by exchanging the roles of the two colours, we also obtain

$$|\mathcal{Q}_{\ell,r,b}| \le |V|^j C^{\ell-1}(|V|^{k-j})^b(\ell|V|^{k-j-1})^r.$$

Proof of Theorem 2(i). Suppose that

$$p_1 p_2 = \frac{1}{cn^{2k-2j-1}\log n},$$

where $c > 0$ is a sufficiently large constant. As mentioned earlier, it is a necessary condition for j-percolation of $H \sim H_k(n, p_1, p_2)$ that both its red and blue k-graphs are j-connected. Therefore, by (1), we may, without loss of generality, assume that

$$\frac{j(k-j)!\log n}{2n^{k-j}} \le p_1, p_2 \le \frac{2}{j(k-j)!\, cn^{k-j-1}(\log n)^2},$$

where the upper bound follows directly from the lower bound and the assumption on the product $p := p_1 p_2$. We will prove that $H_k(n, p_1, p_2)$ does not j-percolate by showing that there is a bottleneck in the process.

By Claim 1 (for $N = \log n$), Fact 1, and Claim 2, we obtain the upper bound

$$\mathbb{P}(H \ j\text{-percolates}) \le \sum_{\ell,r,b} \sum_{T \in \mathcal{T}_{\ell,r,b}} \mathbb{P}(T \subset H),$$

where (*a priori*) the first sum ranges over all integers ℓ, r, and b satisfying $\log n \le \ell \le \binom{k}{j}\log n$ and $0 \le r, b \le \ell-1$. Furthermore, Lemma 1 and Claim 3 imply that $(\ell-1)/\binom{k}{j} \le r, b$ whenever $\mathcal{T}_{\ell,r,b} \ne \varnothing$.

It will be convenient to split the summation into two parts, depending on whether the edge-minimal traversable triple $T = (\mathcal{J}_0, \mathcal{E}_1, \mathcal{E}_2)$ contains more red or blue edges. So let us first consider the term

$$q_1 := \sum_{r \le b < \ell} \sum_{T \in \mathcal{T}_{\ell,r,b}} \mathbb{P}(T \subset H).$$

Since any triple $T \in \mathcal{T}_{\ell,r,b}$ contains precisely r red edges and b blue edges, the probability that it is contained in H is $p_1^r p_2^b$ and, in particular, this probability depends only on the parameters r and b. Furthermore, note that, by Lemma 2, we have

$$|\mathcal{T}_{\ell,r,b}| \le n^j C^{\ell-1}(n^{k-j})^r(\ell n^{k-j-1})^b$$

for some positive constant $C > 0$ and, thus,

$$\sum_{T \in \mathcal{T}_{\ell,r,b}} \mathbb{P}(T \subset H) \le n^j C^{\ell-1}(n^{k-j}p_1)^r(\ell n^{k-j-1}p_2)^b \le n^j C^{\ell-1}(\ell n^{2k-2j-1}p)^r, \qquad (3)$$

https://doi.org/10.1017/jpr.2017.62 Published online by Cambridge University Press

since $r \le b$ and $\ell n^{k-j-1} p_2 = o(1)$. Now recall that

$$\ell n^{2k-2j-1} p \le \frac{1}{c}\binom{k}{j} < 1 \quad \text{for sufficiently large } c,$$

and, thus, the bound in (3) is maximal when r is minimal, i.e. $r = \binom{k}{j}^{-1}(\ell - 1)$. This provides the upper bound

$$q_1 \le \sum_{r \le b < \ell} n^j \left(\frac{1}{c}\binom{k}{j} C^{\binom{k}{j}}\right)^{(\ell-1)/\binom{k}{j}} \le \left(\binom{k}{j}\log n\right)^3 n^j \left(\frac{1}{c}\binom{k}{j} C^{\binom{k}{j}}\right)^{(\log n - 1)/\binom{k}{j}} = o(1)$$

for any sufficiently large constant c dependent on k and j (since C is also only dependent on k and j).

To complete the proof consider

$$q_2 := \sum_{b \le r < \ell} \sum_{T \in \mathcal{T}_{\ell,r,b}} \mathbb{P}(T \subset H).$$

Swapping the roles of red and blue (as indicated in Remark 1), we can use precisely the same argument to show that $q_2 = o(1)$ (using $p_1' := p_2$, $p_2' := p_1$, $\ell' := \ell$, $r' := b$, and $b' := r$). Thus,

$$\mathbb{P}(H \ j\text{-percolates}) \le q_1 + q_2 = o(1),$$

i.e. the binomial random double k-graph $H_k(n, p_1, p_2)$ w.h.p. does not j-percolate. □

3. Supercritical regime: proof of Theorem 2(ii)

As we shall see, the supercritical regime is easier to prove, since we can reduce to the graph case $k = 2$ and $j = 1$. In the first step, we provide the reduction for any $k \ge 3$ but $j = 1$. We then show how to obtain the result for arbitrary pairs (k, j) from the result for $(k-1, j-1)$ provided $j \ge 2$.

Given integers $1 \le j \le k-1$, we define a statement $S(k, j)$ as follows.

$S(k, j)$: there exists a constant $c = c(k, j) > 0$ and an integer $n_0 = n_0(k, j)$ such that for any functions $p_1 = p_1(n)$ and $p_2 = p_2(n)$ which satisfy

$$\frac{c \log n}{n^{k-j}} \le p_1, p_2 \le 1, \quad \text{and} \quad p_1 p_2 \ge \frac{c}{n^{2k-2j-1}\log n} \quad \text{for } n \ge n_0,$$

the double k-graph $H \sim H_k(n, p_1, p_2)$ j-percolates w.h.p.

Showing that $S(k, j)$ holds for all pairs of integers $1 \le j < k$ proves Theorem 2(ii).

We proceed inductively, with the base case being the result on graphs, proved in Theorem 1(i).

Remark 2. We note that $S(2, 1)$ holds.

We split the induction step into two parts.

Claim 4. *If $k \ge 3$ then $S(2, 1)$ implies $S(k, 1)$.*

https://doi.org/10.1017/jpr.2017.62 Published online by Cambridge University Press

Proof. Let $k \geq 3$ be an integer. Assume that $S(2,1)$ holds and let $c' = c(2,1)$. Set $c = c(k,1) := 3(2k)^{2k}c'$. By monotonicity, in order to show $S(k,1)$ it suffices to prove that $H \sim H_k(n, p_1, p_2)$ 1-percolates for probabilities $p_1 = p_1(n)$ and $p_2 = p_2(n)$ satisfying

$$p = p_1 p_2 = \frac{c}{n^{2k-3}\log n} \quad \text{and} \quad \frac{c\log n}{n^{k-1}} \leq p_1, p_2 \leq \frac{1}{n^{k-2}(\log n)^2},$$

where the upper bound is an immediate consequence of the lower bound and the first condition.

We use the following construction to reduce to the graph case. Split the vertex set $[n]$ into two disjoint sets, say $Q := [\lfloor n/2 \rfloor]$ and $Q^* := [n] \setminus Q$. Let H' be the double graph on Q whose edges are any pair in Q contained in an edge of H whose remaining $k-2$ vertices are all in Q^*. Note that, if H' 1-percolates then all vertices of Q must lie in a single cluster $\mathcal{C}$ of the final partition $\mathcal{C}_\infty$ of the 1-jigsaw percolation process in H. Note that $H' \sim H_2(n', p_1', p_2')$ with $n' := \lfloor n/2 \rfloor$ and the edge probabilities $p_1 = p_1(n')$ and $p_2 = p_2(n')$ satisfy

$$p_i' := 1 - (1-p_i)^{\binom{n/2}{k-2}} \geq 1 - \exp\left(-p_i\left(\frac{n}{2k-4}\right)^{k-2}\right) \geq \frac{p_i n^{k-2}}{(2k)^k} \tag{4}$$

for $i \in \{1,2\}$ and sufficiently large n, since $p_i = o(n^{-(k-2)})$. In particular, we have

$$p_1'p_2' \geq \frac{p_1p_2n^{2k-4}}{(2k)^{2k}} \geq \frac{c'}{n'\log n'}, \quad \text{and} \quad p_i' \geq \frac{p_i n^{k-2}}{(2k)^k} \geq \frac{c\log n}{(2k)^k n} \geq \frac{c'\log n'}{n'} \quad \text{for } i \in \{1,2\}.$$

Thus, by the choice of c', H' does indeed 1-percolate w.h.p. Similarly, reversing the roles of Q and Q^*, i.e. considering only edges with precisely two vertices from Q^*, we deduce that w.h.p. there is a cluster $\mathcal{C}^* \in \mathcal{C}_\infty$ such that $Q^* \subset \mathcal{C}^*$. Even though these two events may not be independent (at least for $k = 4$), applying a union bound we can guarantee that they happen simultaneously w.h.p. Furthermore, there are at least two edges, one red and one blue, present in H'. By definition, such edges can only exist if there are corresponding edges (which may not be uniquely determined) of H which each contain at least one vertex from Q and Q^* (since $k \geq 3$). Thus, the clusters $\mathcal{C}$ and $\mathcal{C}^*$ must coincide and contain all vertices, i.e. $\mathcal{C} = \mathcal{C}^* = [n]$. In other words, H 1-percolates w.h.p. □

Claim 5. *If $k > j \geq 2$ then $S(k-1, j-1)$ implies $S(k,j)$.*

Proof. Fix integers $k > j \geq 2$. Assume that $S(k-1, j-1)$ holds and let $c' = c(k-1, j-1)$. Set $c := \max\{5c', (2k)^k\}$. Again, by monotonicity, in order to show $S(k,j)$ it suffices to prove that $H \sim H_k(n, p_1, p_2)$ j-percolates for probabilities $p_1 = p_1(n)$ and $p_2 = p_2(n)$ satisfying

$$p = p_1 p_2 = \frac{c}{n^{2k-2j-1}\log n} \quad \text{and} \quad p_1, p_2 \geq \frac{c\log n}{n^{k-j}}.$$

We expose the edges of H in two rounds, i.e. for $i = 1, 2$, we set $p_i' := 1 - \sqrt{1-p_i} \geq p_i/2$. Note that $p_i = 2p_i' - (p_i')^2$ and, thus, we obtain H as the union of two independent copies H_α and H_β of $H_k(n, p_1', p_2')$. (In particular, we obtain the final partition of j-jigsaw percolation on H by running j-jigsaw percolation on H_α and using its final partition, denoted by $\mathcal{C}_\infty[H_\alpha]$, as the initial partition for j-jigsaw percolation on H_β.)

In the following, we will consider certain *link double $(k-1)$-graphs* associated to H_α. They are constructed as follows. Given a vertex $v \in [n]$, we first delete from H_α all edges (i.e. k-sets) that do not contain the vertex v. Then we delete v from the vertex set and replace every remaining edge e with the $(k-1)$-set $e \setminus v$. We denote by $H_{\alpha,v}$ the resulting random double $(k-1)$-graph on vertex set $[n] \setminus \{v\}$, and call it the *link double $(k-1)$-graph of v*.

https://doi.org/10.1017/jpr.2017.62 Published online by Cambridge University Press

Now note that there is a natural bijection mapping the set of $(j-1)$-sets (respectively, $(k-1)$-sets) in $H_{\alpha,v}$ to the set of j-sets (respectively, k-sets) containing v in H_α. Therefore, any cluster in the final partition of $(j-1)$-jigsaw percolation on $H_{\alpha,v}$ corresponds to a collection of j-sets (once we have added v to each) which must be contained in a cluster of $\mathcal{C}_\infty[H_\alpha]$. Therefore, whenever $H_{\alpha,v}$ $(j-1)$-percolates, there is a unique cluster in $\mathcal{C}_\infty[H_\alpha]$ which contains all j-sets which include v, and, thus, we call it *the corresponding cluster to* v. We call a vertex v *good* if $H_{\alpha,v}$ $(j-1)$-percolates; vertices that are not good are called *exceptional*. This notion is motivated by the following observation. The corresponding clusters of any two good vertices u and u' overlap in all j-sets containing both u and u', thus, indeed they must coincide (since $j \geq 2$). In other words, the final partition $\mathcal{C}_\infty[H_\alpha]$ contains a cluster $\mathcal{C}$ which includes every j-set with at least one good vertex.

Hence, it remains to study j-sets containing only exceptional vertices. For this, we observe that $H_{\alpha,v}$ is distributed as $H_{k-1}(n', p_1', p_2')$, where $n' := n-1$ and the probabilities $p_1' = p_1'(n')$ and $p_2' = p_2'(n')$ satisfy

$$p_1' p_2' \geq \frac{c}{4n^{2k-2j-1}\log n} \geq \frac{c'}{(n')^{2(k-1)-2(j-1)-1}\log(n')},$$

and

$$p_i' \geq \frac{c\log n}{2n^{k-j}} \geq \frac{c'\log(n')}{(n')^{(k-1)-(j-1)}} \quad \text{for } i \in \{1,2\} \text{ and sufficiently large } n.$$

Consequently, by the choice of c', $H_{\alpha,v}$ $(j-1)$-percolates w.h.p. and, therefore, the expected number of exceptional vertices is $o(n)$, say n/ω, for some function $\omega \to \infty$. Thus, w.h.p. there are at most $n/\sqrt{\omega}$ exceptional vertices, by Markov's inequality.

Now we run j-jigsaw percolation on H_β with initial partition $\mathcal{C}_0[H_\beta] := \mathcal{C}_\infty[H_\alpha]$ and show that w.h.p. $\mathcal{C}_1[H_\beta] = \mathcal{C}_\infty[H_\beta] = [n]$, i.e. there is percolation in a single step. For this, it is sufficient to show that w.h.p., for every j-set $J = \{u_{(1)}, \dots, u_{(j)}\}$ containing *only* exceptional vertices, there are edges e_1 (red) and e_2 (blue) in H_β each containing J and a good vertex v_1 and v_2, respectively, since $J' := \{u_{(1)}, \dots, u_{(j-1)}, v_i\}$ satisfies $J' \in \mathcal{C}_\infty[H_\alpha]$ and $J' \subset e_i$. For any $i \in \{1,2\}$ and any such j-set J, the probability that no edge e_i exists is at most

$$(1-p_i')^{(1-1/\sqrt{\omega})n\binom{n-j-1}{k-j-1}} \leq \exp\left(-\frac{p_i' n^{k-j}}{(2k)^{k-j}}\right) \leq \exp(-(2k)^{j-k} c\log n) \leq n^{-(2k)^j} = o(n^{-j})$$

for sufficiently large n and by the choice of c. Hence, we may take the union bound over $i \in \{1,2\}$ and all (such) j-sets. Therefore, $H_k(n, p_1, p_2)$ j-percolates w.h.p., completing the proof. □

Proposition 2 and Claims 4 and 5 imply that $S(k, j)$ holds for all pairs of integers $1 \leq j < k$ and this proves Theorem 2(ii).

4. Related models

With some small alterations these methods can also be applied for other models, for instance, line graphs or in a setting with any fixed number of colours.

4.1. Line graphs

We consider the following random graph $L(n, p_1, p_2)$ that has a vertex for every pair of elements from the set $[n]$, i.e. $V = [n]^{(2)}$, and any two vertices that intersect form a red/blue edge with probabilities p_1 and p_2, respectively, independently of each other and of all pairs of vertices; disjoint vertices cannot form an edge. Note that (graph)-jigsaw percolation on this

https://doi.org/10.1017/jpr.2017.62 Published online by Cambridge University Press

model is closely related to 2-jigsaw percolation on $H_3(n, p_1, p_2)$. Following the lines of our proof for $k = 3$ and $j = 2$, we obtain the following result.

Theorem 3. *Let $0 \le p_1 = p_1(n), p_2 = p_2(n) \le 1$ and let $H \sim L(n, p_1, p_2)$. Then there is a constant $c > 0$ such that*

(i) *if $p_1 p_2 \le \dfrac{1}{cn \log n}$ then w.h.p. H does not j-percolate;*

(ii) *if $p_1 p_2 \ge \dfrac{c}{n \log n}$ and $\min\{p_1, p_2\} \ge \dfrac{c \log n}{n}$ then w.h.p. H j-percolates.*

In other words, the threshold $\hat{p}_L = \hat{p}_L(n)$ for jigsaw percolation on $L(n, p_1, p_2)$ in terms of the product $p = p_1 p_2$ is of the order

$$\hat{p}_L = \Theta\left(\frac{1}{n \log n}\right).$$

Indeed, the proof for the subcritical regime will be simplified since there cannot be any edges doing multiple duty (neither in red nor in blue). For the supercritical regime, one reduction step is enough, since the link graph of a vertex in the line graph is a binomial random double graph and, thus, Theorem 1 applies.

4.2. More colours

Returning to the original motivation of jigsaw percolation, one might ask whether the social network is able to collectively solve multiple puzzles simultaneously. In this spirit we define a binomial random s-fold k-graph $H_k(n, p_1, \dots, p_s)$, in which the vertex set is $[n]$ and any k-set forms an i-edge with probability p_i independently for all $1 \le i \le s$ and all other k-sets.

Now in the process of j-jigsaw percolation, clusters merge if, for each colour, there is at least one edge connecting them. Hence, based on the same intuition, we analyse internally spanned $(s+1)$-tuples (the s-coloured analogue of internally spanned triples) and observe that the sequence of white j-sets is already determined by the set of edges of the first colour, say red. Now any further colour essentially behaves like blue and, in particular, independently of the other colours, given the sequence of white j-sets. With this intuition we obtain the following generalisation of Theorem 2.

Theorem 4. *For integers $1 \le j < k$ and $s \ge 2$, let $0 \le p_1 = p_1(n) \le \cdots \le p_s = p_s(n) \le 1$ and let $H \sim H_k(n, p_1, \dots, p_s)$. Then there is a constant $c = c(s, k, j) > 0$ such that*

(i) *if $p_1 \le \dfrac{\log n}{cn^{k-j}}$, or, for any $2 \le r \le s$, we have*

$$\prod_{i=1}^{r} p_i \le \frac{1}{cn^{r(k-j-1)+1}(\log n)^{r-1}},$$

then w.h.p. H does not j-percolate;

(ii) *if $p_1 \ge \dfrac{c \log n}{n^{k-j}}$, and, for each $2 \le r \le s$, we have*

$$\prod_{i=1}^{r} p_i \ge \frac{c}{n^{r(k-j-1)+1}(\log n)^{r-1}},$$

then w.h.p. H j-percolates.

https://doi.org/10.1017/jpr.2017.62 Published online by Cambridge University Press

In other words, the threshold $\hat{p}_s = \hat{p}_s(n)$ for j-jigsaw percolation on the s-fold k-graph $H_k(n, p_1, \dots, p_s)$ in terms of $p = \prod_{i=1}^{s} p_i$ is of order

$$\hat{p}_s = \Theta\left(\frac{1}{n^{s(k-j-1)+1}(\log n)^{s-1}}\right).$$

The condition is stated for $2 \le r \le s$ since percolation of any subset of r colours is necessary for the s-fold k-graph to percolate. The condition on p_1 corresponds to the $r = 1$ case; the j-percolation process for one colour simply merges j-components into one cluster, thus, it is important whether any of the p_i are below the j-connectedness threshold given in (1).

Proof of Theorem 4. (Outline.) It turns out that there is a minor technical obstacle when determining the upper bound on the probabilities p_i for all $i \in [s]$ from some necessary conditions for j-percolation of an s-fold k-graph. As before, the k-graph of (any) colour $i \in [s]$ has to be j-connected, i.e. by (1), we may assume that

$$p_i = \Omega\left(\frac{\log n}{n^{k-j}}\right).$$

However, this alone will not yield useful upper bounds. Instead, we observe that additionally, for any proper subset $I \subsetneq \{1, \dots, s\}$ of size at least two, it is necessary that the $|I|$-fold k-graph $H_k(n, (p_i)_{i\in I})$ j-percolates. Theorem 4 is then proved by induction over s with Theorem 2 providing the base case. Hence, assume that $\prod_{j\in[s]:\, j\neq i} p_j = \Omega(\hat{p}_{s-1})$ for all $i \in [s]$. Additionally, by monotonicity, we may also assume that $\prod_{j\in[s]} p_j = \Theta(\hat{p}_s)$ and, thus, obtain the following upper bounds on the probabilities p_i:

$$p_i = \frac{\prod_{j\in[s]} p_j}{\prod_{j\in[s]:\, j\neq i} p_j} = O\left(\frac{\hat{p}_s}{\hat{p}_{s-1}}\right) = O\left(\frac{1}{n^{k-j-1}\log n}\right).$$

Even though this bound is slightly weaker than in the two-colour case (where we had another factor of $1/\log n$), it turns out to be sufficient for adapting our proof. More precisely, it is used in the following two arguments.

- First, in the subcritical regime, in order to derive (3). There we only need

$$\ell n^{k-j-1} p_i \le 1 \quad \text{for all } \ell \le \binom{k}{j} \log n \text{ and all colours } i \in [s].$$

  This holds if the constant $c(s, k, j)$ is chosen sufficiently large compared to $c(s-1, k, j)$.

- Second, in the supercritical regime, for the last estimate in (4). The argument does not change at all, since having a single $(1/\log n)$-factor is already enough here.

Apart from these occurrences, the upper bound is only used to reprove the statement (in the supercritical regime) for graphs [3] in the setting of multiple colours. Some of the asymptotic approximations used in the proof of the two-colour graph case are no longer useful with only the weaker upper bound. However, these technical issues can be dealt with and the details can be found in [7]. □

5. Concluding remarks

Theorem 2 holds for a large enough constant c, which we made no attempt to optimise (nor was any such attempt made for graphs in [3]). It would be interesting to know the exact threshold,

https://doi.org/10.1017/jpr.2017.62 Published online by Cambridge University Press

and, in particular, whether it is sharp, i.e. the upper and lower thresholds are asymptotically the same.

Note also that in the supercritical regime there was an extra condition on $\min\{p_1, p_2\}$ which contained a factor of c. However, there is no intrinsic reason why this c should be the same as the c in the product. Indeed, the reason for this condition is to ensure that each hypergraph is j-connected w.h.p, but, as mentioned in the introduction, the asymptotic threshold for this was determined in [5] to be

$$\frac{j(k-j)!\log n}{n^{k-j}}.$$

It is therefore natural to conjecture that this condition can be replaced by

$$\min\{p_1, p_2\} \geq \frac{c'\log n}{n^{k-j}}$$

for any $c' > j(k-j)!$. Indeed, even if, say

$$p_1 = \frac{j(k-j)!\log n}{n^{k-j}},$$

there is a certain probability, bounded away from 0 and 1, that the red k-graph is j-connected. It is then natural to conjecture that, conditioned on it being j-connected (note that w.h.p. the blue k-graph will also be j-connected), w.h.p. the double k-graph $H_k(n, p_1, p_2)$ j-percolates.

Acknowledgements

We would like to thank the anonymous referees for their comments and suggestions, which improved the quality of the paper. The first author is supported by the National Science Foundation (NSF): DMS 1600742. The second, third, and fourth authors are supported by the Austrian Science Fund under P26826 and W1230, and Doctoral Program 'Discrete Mathematics'. The fourth author is also supported by the European Research Council (grant number 639046).

References

[1] Ball, P. (2014). Crowd-sourcing: strength in numbers. *Nature* **506,** 422–423.
[2] Barabási, A. L. *et al.* (2002). Evolution of the social network of scientific collaborations. *Phys. A* **311,** 590–614.
[3] Bollobás, B., Riordan, O., Slivken, E. and Smith, P. (2017). The threshold for jigsaw percolation on random graphs. *Electron. J. Combin.* **24,** 2.36.
[4] Brummitt, C. D., Chatterjee, S., Dey, P. S. and Sivakoff, D. (2015). Jigsaw percolation: what social networks can collaboratively solve a puzzle? *Ann. Appl. Prob.* **25,** 2013–2038.
[5] Cooley, O., Kang, M. and Koch, C. (2016). Threshold and hitting time for high-order connectedness in random hypergraphs. *Electron. J. Combin.* **23,** 2.48.
[6] Gravner, J. and Sivakoff, D. (2017). Nucleation scaling in jigsaw percolation. *Ann. Appl. Prob.* **27,** 395–438.
[7] Gutiérrez Sanchez, A. (2017). Multi-colored jigsaw percolation on random graphs. *Master's Thesis*. Graz University of Technology.
[8] Newman, M. E. J. (2001). Scientific collaboration networks. I. Network construction and fundamental results. *Phys. Rev. E* **64,** 016131.
[9] Newman, M. E. J. (2001). Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. *Phys. Rev. E* **64,** 016132.
[10] Newman, M. E. J. (2001). The structure of scientific collaboration networks. *Proc. Nat. Acad. Sci. USA* **98,** 404–409.
[11] Tebbe, J. (2011). Book review: Where good ideas come from: the natural history of innovation. *J. Psychological Issues Organizational Culture* **2,** 106–110.

https://doi.org/10.1017/jpr.2017.62 Published online by Cambridge University Press