JIGSAW PERCOLATION: WHAT SOCIAL NETWORKS CAN COLLABORATIVELY SOLVE A PUZZLE?

By Charles D. Brummitt[1], Shirshendu Chatterjee,
Partha S. Dey[2] and David Sivakoff[3]

*University of California, Davis, New York University,
University of Warwick and Ohio State University*

We introduce a new kind of percolation on finite graphs called *jigsaw percolation*. This model attempts to capture networks of people who innovate by merging ideas and who solve problems by piecing together solutions. Each person in a social network has a unique piece of a jigsaw puzzle. Acquainted people with compatible puzzle pieces merge their puzzle pieces. More generally, groups of people with merged puzzle pieces merge if the groups know one another and have a pair of compatible puzzle pieces. The social network solves the puzzle if it eventually merges all the puzzle pieces. For an Erdős–Rényi social network with n vertices and edge probability p_n, we define the critical value $p_c(n)$ for a connected puzzle graph to be the p_n for which the chance of solving the puzzle equals $1/2$. We prove that for the n-cycle (ring) puzzle, $p_c(n) = \Theta(1/\log n)$, and for an arbitrary connected puzzle graph with bounded maximum degree, $p_c(n) = O(1/\log n)$ and $\omega(1/n^b)$ for any $b > 0$. Surprisingly, with probability tending to 1 as the network size increases to infinity, social networks with a power-law degree distribution cannot solve any bounded-degree puzzle. This model suggests a mechanism for recent empirical claims that innovation increases with social density, and it might begin to show what social networks stifle creativity and what networks collectively innovate.

1. Introduction. Solving difficult problems and creating new ideas are sometimes compared to merging the pieces of a puzzle [2, 25]. Often these breakthroughs are achieved not by one person working in isolation but rather by a collection of people who exchange and merge partial solutions and ideas [25]. As a result, the structure of collaboration networks (who collaborates with whom) can

affect the success of the network's creative output, as found empirically for scientific breakthroughs [9, 18, 27] and for hit Broadway musicals [38, 39]. In business, some companies connect their employees using internal social networks [30] and expertise location systems [12] to match compatible ideas and expertise. Some companies outsource their most difficult R&D problems to leverage knowledge worldwide using services such as Innocentive and Kaggle. Digital tools for massive collaboration are also being used to solve problems in mathematics [19], climate change [24] and software design [26].

Here we formalize this metaphor of a large group of people collaboratively solving a puzzle by introducing a new kind of percolation on finite graphs that aims to model a network of people who merge compatible ideas into bigger and better ideas. The model is reminiscent of other models of percolation on graphs, such as bond percolation [22] and bootstrap percolation [23], but jigsaw percolation has more complex dynamics.

Consider a social network of n people with vertex set $V = \{1, 2, \ldots, n\}$, each of whom has a unique "partial idea" that could merge with one or more other partial ideas belonging to other people. These "partial ideas" can be thought of as pieces of a jigsaw puzzle: an idea is compatible with certain other ideas, just as a piece of a jigsaw puzzle can join with certain other puzzle pieces (in the correct solution of the puzzle). Thus we use "ideas" and "puzzle pieces" interchangeably. The two networks are:

- the *people graph* (V, E_{people}), denoting who knows and communicates with whom;
- the *puzzle graph* (V, E_{puzzle}), denoting which ideas are compatible and thus can merge to form a bigger, better idea.

In this paper, we assume each person has a unique idea, so there are n ideas (puzzle pieces), and the system of people and their compatible ideas is a graph with two sets of edges, E_{people} and E_{puzzle}. Allowing a person to have multiple ideas or multiple people to have the same idea requires two vertex sets, which we leave for future work; see Section 6.

Next we propose a natural dynamic for people to merge their compatible ideas (puzzle pieces). If two people u, w know each other and have compatible puzzle pieces (i.e., $uw \in E_{\text{people}} \cap E_{\text{puzzle}}$), then they merge their puzzle pieces. After u, w merge their puzzle pieces, we say that u, w belong to the same *jigsaw cluster* $U \subseteq V$. The general rule is that two jigsaw clusters U, W merge if at least two people (one from each cluster) know each other, and at least two people (one from each cluster) have compatible puzzle pieces. More precisely, we say that jigsaw clusters U, W are *people-adjacent* if $uw \in E_{\text{people}}$ for some $u \in U, w \in W$. Similarly, U, W are *puzzle-adjacent* if $u'w' \in E_{\text{puzzle}}$ for some $u' \in U, w' \in W$. Jigsaw clusters U, W merge if they are both people-adjacent and puzzle-adjacent.
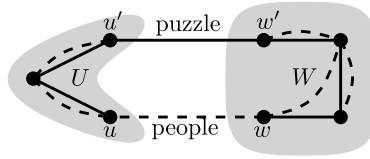
FIG. 1. *Illustration of the jigsaw dynamic. Dashed and solid edges denote the people graph and puzzle graph, respectively. Jigsaw clusters U and W contain three and four nodes each. Nodes u, w know each other but do not have compatible puzzle pieces. However, they have merged their puzzle pieces with nodes u′, w′, who do have compatible puzzle pieces. Thus U and W merge.*

The motivation for this dynamic is the notion that after merging their ideas, a group of people can use any of those ideas to merge with the ideas of other people whom they know. We illustrate this in Figure 1. Here two nodes u, w in different jigsaw clusters U, W know each other ($uw \in E_{\text{people}}$), but their puzzle pieces are incompatible ($uw \notin E_{\text{puzzle}}$). However, u and w have merged their puzzle pieces with those of u' and w', respectively, and u' and w' do have compatible puzzle pieces ($u'w' \in E_{\text{puzzle}}$). Thus u can tell w about her friend u', and w can tell u about his friend w'. Then u' and w' merge their compatible puzzle pieces, and the jigsaw clusters U and W merge.

Our main results, Theorems 1 and 2, characterize a phase transition in the probability that a random graph solves a jigsaw puzzle in the manner described above. We find, roughly speaking, the required number of interactions among a group of people for them to collectively solve a large puzzle. This phase transition might begin to inform what properties of social networks facilitate their ability to collaboratively solve problems and to innovate.

1.1. *Related literature.* Previous models of scientific discovery and innovation can be roughly partitioned into three sets. Models in the first set focus on the structure of the social network but not on the space of ideas; an example is an epidemic model of a single idea that spreads like a slow, hard-to-catch disease in a social network [5, 7]. Models in the second set focus on the space of ideas but not on the social network; an example is a branching process of new ideas mating with old ones [37]. Models in the third set attempt to capture both the social network and how it interacts with some space of ideas. One example is a model of people trading and gifting ideas with neighbors in a social network to obtain certain ideas needed to produce an output [14]. Four other models in this set are reviewed in [10]: an ant colony model of scientists seeking papers to cite like ants seeking food; the costs and benefits of hunting for references in bibliographic habitats ("information foraging theory"); the A–B–C model of finding triadic closure among ideas; and bridging structural holes (gaps between dense communities of graphs) in networks of people and ideas. However, researchers have noted the difficulty in modeling how teamwork and collaboration lead to greater collective creativity

and discovery [6, 16]. Our contribution to this literature is a model that focuses on the way people might collaboratively merge their partial solutions to a difficult problem (or their partial ideas that combine to form a better idea).

1.2. *Road map for the paper.* In Section 2, we define the jigsaw percolation process formally. We present the main results in Section 3 and prove them in Sections 4–5. In Section 6, we discuss simulations and open questions.

2. **Formal definition of jigsaw percolation.** Formally, jigsaw percolation on $(V, E_{\text{people}}, E_{\text{puzzle}})$ proceeds in steps as follows. At every step $i \geq 0$, we have a partition $\mathcal{C}_i$ of the vertex set V. The elements of $\mathcal{C}_i$, called "jigsaw clusters," are labels on vertices that denote which puzzle pieces have merged by step i:

(1) Initially, $\mathcal{C}_0$ is the set of singletons $\{\{v\} : v \in V\}$.
(2) At step $(i + 1) \geq 1$, we merge every pair of jigsaw clusters in $\mathcal{C}_i$ that are both puzzle- and people-adjacent; see Figure 2.

For example, after the first step, $\mathcal{C}_1$ is the set of connected components in the graph $(V, E_{\text{people}} \cap E_{\text{puzzle}})$. Note that three or more jigsaw clusters can merge simultaneously, as illustrated in Figure 2.

It is useful to write jigsaw percolation as a dynamical system as follows. At step i, let $\mathcal{E}_i$ be the unordered pairs of clusters in $\mathcal{C}_i$ that are people-adjacent and puzzle-adjacent. Then the jigsaw clusters in $\mathcal{C}_{i+1}$ are the connected components of the graph $(\mathcal{C}_i, \mathcal{E}_i)$:

$$(2.1) \qquad \mathcal{C}_{i+1} = \left\{ \bigcup_{U \in A} U : A \text{ is a connected component of } (\mathcal{C}_i, \mathcal{E}_i) \right\}.$$

Given $(V, E_{\text{people}}, E_{\text{puzzle}})$, we merge jigsaw clusters until no more merges can be made, that is, iterate equation (2.1) to a fixed point $\mathcal{C}_\infty$. After finitely many steps, no more merges can be made. We say that *the people graph solves the puzzle* if all nodes belong to the same jigsaw cluster at the end of the process (i.e., $\mathcal{C}_\infty = \{V\}$). Figure 3 illustrates a people graph that fails to solve a 2×2 puzzle.
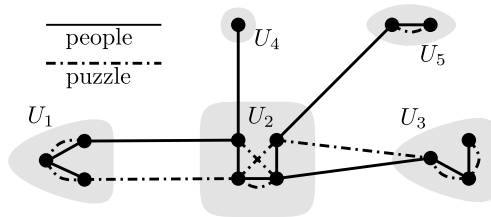

FIG. 2.   *Jigsaw clusters $U_1, U_2, U_3, U_4, U_5 \in \mathcal{C}_i$ at stage i. At stage $i + 1$, jigsaw clusters U_1, U_2, U_3 merge.*
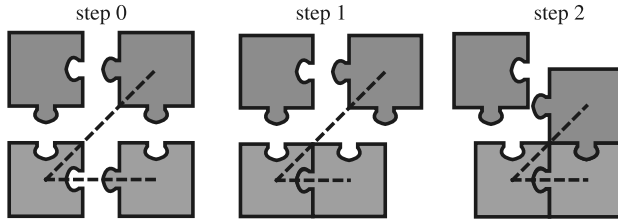
FIG. 3. *A complete trajectory of the jigsaw dynamics. The people graph (dashed edges) does not solve this* 2×2 *puzzle.*

An equivalent definition of the process that is elegant and simple to code on the computer is to iteratively contract nodes that are adjacent in $E_{\text{people}} \cap E_{\text{puzzle}}$ until no more contractions are possible. The people graph solves the puzzle if this procedure ends with a single node.

3. Statement of results.

3.1. *Erdős–Rényi random graphs solving ring and bounded-degree puzzles.* In most of this paper, we consider people graphs that are Erdős–Rényi random graphs $\mathcal{G}(n, p_n)$, in which each possible edge appears independently with probability p_n, with associated probability distribution $\mathbb{P}_{p_n}$. (The exception is Section 5, in which we consider power-law random graphs rather than Erdős–Rényi random graphs.) For a fixed, connected puzzle graph of size n, we are interested in the probability of the event

$$\texttt{Solve} := \{\text{the people graph solves the puzzle}\} = \{\mathcal{C}_\infty = \{V\}\}.$$

We denote this probability by $\mathbb{P}(\texttt{Solve})$ or by $\mathbb{P}_{p_n}(\texttt{Solve})$ to make explicit the value of p_n. Note that the jigsaw dynamic is *monotonic*, in that adding more edges to the people graph or to the puzzle graph cannot decrease the chance of solving the puzzle. Thus, for fixed n, $\mathbb{P}_p(\texttt{Solve})$ is nondecreasing with p. Trivially, $\mathbb{P}_0(\texttt{Solve}) = 0$ and $\mathbb{P}_1(\texttt{Solve}) = 1$. Furthermore, $\mathbb{P}_p(\texttt{Solve})$ is a polynomial in p of degree at most $\binom{n}{2}$. Thus for each n there exists a unique $p \in (0, 1)$ such that $\mathbb{P}_p(\texttt{Solve}) = 1/2$, and we make the following definition.

DEFINITION 1. The *critical value* $p_c(n)$ for solving a connected puzzle is the unique value of $p_n \in (0, 1)$ such that $\mathbb{P}_{p_n}(\texttt{Solve}) = 1/2$.

REMARK 1. There is nothing special about the number $1/2$. For our results, we could have taken any fixed positive real number strictly smaller than 1. However, the critical value $p_c(n)$ depends on the choice of the puzzle graph, which we suppress in the notation $p_c(n)$.

REMARK 2. If the people graph is not connected, then the puzzle cannot be solved. Thus $p_c(n) \geq t_n$, where t_n is the unique real number such that $\mathbb{P}(\mathcal{G}(n, t_n) \text{ is connected}) = 1/2$. Asymptotically we have $t_n \approx (\log n - \log \log 2)/n$; see [17]. Note that the equality $p_c(n) = t_n$ holds when the puzzle graph is the star graph ($\{1, 2, \ldots, n\}, \{(i, n) : 1 \leq i < n\}$), because in this case the puzzle can be solved if and only if the people graph is connected.

We use the following standard notation for describing sequences of nonnegative real numbers a_n and b_n: $a_n = O(b_n)$ means there exists $C > 0$ so that $a_n \leq Cb_n$ for all sufficiently large n; $a_n = \Theta(b_n)$ means $a_n = O(b_n)$ and $b_n = O(a_n)$; $a_n = o(b_n)$ means $a_n/b_n \to 0$ as $n \to \infty$; and $a_n = \omega(b_n)$ means $b_n = o(a_n)$.

Our main results are the following two theorems.

THEOREM 1 (Ring puzzle). *If the people graph is the Erdős–Rényi random graph and the puzzle graph is the n-cycle, then*

$$\frac{1}{27 \log n} \leq p_c(n) \leq \frac{\pi^2}{6 \log n}(1 + o(1)).$$

Moreover, for $p_n = \lambda/\log n$, $\mathbb{P}_{p_n}(\texttt{Solve}) \to 0$ or 1 according as $\lambda < 1/27$ or $\lambda > \pi^2/6$.

REMARK 3. We believe that our upper bound is tight; see Section 6. We did not attempt to optimize the constant $1/27$ in the lower bound; this value was chosen to make the proof easier to read. We do not think that our proof method will yield an optimal lower bound.

THEOREM 2 (Connected puzzle of bounded degree). *For an Erdős–Rényi people graph solving a connected puzzle with bounded maximum degree, $p_c(n) = O(1/\log n)$ and $p_c(n) = \omega(1/n^b)$ for any $b > 0$. In particular, we have $\mathbb{P}_{p_n}(\texttt{Solve}) \to 0$ for $p_n = O(1/n^b)$ for any $b > 0$, and $\mathbb{P}_{p_n}(\texttt{Solve}) \to 1$ for $p_n = \lambda/\log n$ with $\lambda > \pi^2/6$.*

REMARK 4. The upper bound for $p_c(n)$ in Theorem 2 holds for any connected puzzle graph, even with maximum degree growing with n as $n \to \infty$; see Proposition 2. The star graph example in Remark 2 provides a counterexample to the lower bound when the maximum degree is unbounded.

REMARK 5. The jigsaw dynamic is symmetric under swapping the people and puzzle graphs. Thus Theorems 1 and 2 also apply to a ring and bounded-degree people graph (resp.) solving an Erdős–Rényi puzzle.

Some of the techniques in our proofs resemble those used for long-range percolation and for bootstrap percolation, but our arguments differ in key ways. In our

proof of the lower bound on $p_c(n)$ for the ring puzzle graph, we show that a set of cut points, which must separate jigsaw clusters in the final configuration $\mathcal{C}_\infty$, exists with high probability for sufficiently small p. This is similar in spirit to finding a positive density of points over which no edge crosses in the context of one-dimensional long range percolation [13, 35] to show that no infinite component exists.

In our proof of the upper bound on $p_c(n)$, we use the fact that once a sufficiently large, solved cluster emerges, that cluster will inevitably continue to merge and ultimately solve the puzzle. As in bootstrap percolation on the lattice graph [1, 23], our upper bound arises from a sufficient condition for the formation of a large cluster.

3.2. *Power-law random graphs solving bounded-degree puzzles.* As a model of social networks, the Erdős–Rényi random graph assumes no structure other than the average number of connections (neighbors) per person. However, in many social networks—from scientific citations [33] to scientific collaborations [3, 31, 32] to sexual partners [28]—some people have orders of magnitude more connections than others. The broad-scale degree distributions of such networks are well described by a power-law (or by a power-law with a cutoff), in which the fraction of vertices having degree k is proportional to $k^{-\alpha}$ for some power $\alpha > 2$. In light of these findings, we consider jigsaw percolation on people graphs that are given by the configuration model [29] with limiting power-law degree distribution $\mathbf{p} = \{p_k\}$ satisfying

$$\text{(3.1)} \qquad \begin{aligned} &p_k = 0 \qquad \text{for } k < d_{\min} \text{ for some } d_{\min} \geq 3 \quad \text{and} \\ &p_k \asymp k^{-\alpha + o(1)} \qquad \text{as } k \to \infty \text{ for some power } \alpha > 2. \end{aligned}$$

The condition $d_{\min} \geq 3$ is imposed to ensure that the resulting people graph is connected with high probability. Here and later the phrase "with high probability" refers to "with probability tending to 1 as the size of the graph (network) grows to infinity."

In the configuration model, the people graph (V, E_{people}) is constructed in two stages. Assuming $|V| = n$, first the degrees $d_1, d_2, \ldots, d_n$ are chosen to be i.i.d. from the aimed degree distribution $\mathbf{p}$, and d_i many half-edges are assigned to vertex i, $1 \leq i \leq n$. We make the sum of the degrees even by possibly adding one to d_n. This has no effect on the analysis that follows. Then, conditioned on $\{d_i\}_{i=1}^n$, (V, E_{people}) is chosen uniformly from the collection of (multi-)graphs having degree sequence $(d_1, d_2, \ldots, d_n)$ by randomly matching the half-edges at each vertex.

Surprisingly, such heterogeneous social networks cannot solve a large class of puzzles.

PROPOSITION 1. *For any $\alpha > 2$, if (V, E_{people}) is given by the configuration model on n vertices with power-law degree distribution $\mathbf{p}$ satisfying* (3.1), *and if (V, E_{puzzle}) has bounded maximum degree, then* $\mathbb{P}(\texttt{Solve}) \to 0$ *as $n \to \infty$.*

REMARK 6.   Because collaboration networks in science [3, 31, 32] manage to collectively solve puzzles despite their degree distributions being well modeled by power-laws with exponential decay, more realistic assumptions, such as unbounded-degree puzzles and randomly grown collaboration networks, merit future work; see Section 6 for more details.

For degree exponent $\alpha > 2$ of the social network, we expect Proposition 1 to hold for models of power-law random graphs other than the configuration model as well. It is easy to check that the maximum of n i.i.d. random variables from the distribution given in (3.1) is tight under the scaling $n^{-1/(\alpha-1)}$. Thus one expects to couple the power-law random graph as a subgraph of an Erdős–Rènyi random graph with edge probability $1/n^b$ with $b < 1/(\alpha - 1)$ and deduce Proposition 1 from Theorem 2 and a monotonicity argument. This conclusion is indeed true for the Chung–Lu power-law random graph model (cf. [11]) with $\alpha > 3$. However, for $\alpha < 3$ the power-law random graphs contain large cliques having size polynomial in n. This excludes the possibility of the above coupling, as the maximum size of a clique in the Erdős–Rényi random graph $\mathcal{G}(n, n^{-b})$ is at most poly-logarithmic in n.

The proof of Proposition 1, presented in Section 5, circumvents this issue with a direct argument without the need for any coupling. Furthermore, for $\alpha \in (1, 2)$, we expect the power-law random graph given by the configuration model to solve any bounded-degree puzzle with high probability, because then the people graph has very small diameter; cf. [40]. However, we do not have a rigorous proof for that conjecture.

3.3. *Subsequent work.*   After this work appeared as a preprint, Slivken [36] proved a related result for random puzzle graph. In this model, both the people and the puzzle graphs are Erdős–Rényi with edge probabilities p_{ppl} and p_{puz}, respectively, which satisfy $p_{ppl} \wedge p_{puz} \geq (1 + \varepsilon) \log n/n$ for some $\varepsilon > 0$ to ensure that both graphs are connected with high probability. It is shown in [36] that the probability of solving the puzzle is close to zero if $p_{ppl} \cdot p_{puz} \leq c/(n \log n)$ and is close to one if $p_{ppl} \cdot p_{puz} \geq \log \log n/(cn \log n)$, for some constant $c > 0$. In another subsequent paper [21], Gravner and one of the present authors proved that for an Erdős–Rényi people graph solving a general puzzle graph with bounded maximum degree D, the critical value p_c is $\Theta(1/\log n)$, where the constants depend only on D.

4. Erdős–Rényi random graphs solving ring and bounded-degree puzzles.
In this section, we prove Theorems 1 and 2, in which the people graph is the Erdős–Rényi random graph. In Section 4.1, we prove the upper bound on the critical value $p_c(n)$ for both Theorems 1 and 2. In Section 4.2, we prove the lower bound for the ring puzzle in Theorem 1, and in Section 4.3 we prove the lower bound for arbitrary puzzles with bounded maximum degree.

4.1. *Upper bound on the critical value.* In this section, we prove that the critical value has upper bound $\pi^2/(6 \log n)$ for any connected puzzle graph.

PROPOSITION 2 (Upper bound for the critical value). *For an Erdős–Rényi people graph and any connected puzzle graph on n vertices, if $\lambda > \pi^2/6$ and $p_n = \lambda/\log n$, then*

$$\lim_{n \to \infty} \mathbb{P}_{p_n}(\texttt{Solve}) = 1.$$

REMARK 7. A close look at the proof of Proposition 2 reveals that the same conclusion is true as long as $p_n \geq \pi^2/(6 \log n) \cdot (1 + c \log \log n/\log n)$ for some constant $c \in (0, \infty)$.

For simplicity, one can look at the ring puzzle graph (the n-cycle), with

$$E_{\text{puzzle}} = \big\{(1,2), (2,3), \ldots, (n-1,n), (n,1)\big\}.$$

The idea of the proof is the following sufficient condition to solve the ring puzzle, illustrated in Figure 4. Suppose that in the people graph, node 2 is adjacent to node 1; node 3 is adjacent to 1 or 2; node 4 is adjacent to 1, 2 or 3; and so on, so that node j is people-adjacent to at least one of $\{1, 2, \ldots, j-1\}$ for all $2 \leq j \leq n$ (as illustrated in Figure 4). Then the people graph solves the puzzle.

However, to obtain a good bound, we do not consider solving the whole puzzle in the manner depicted in Figure 4. Instead, we partition the puzzle graph into disjoint blocks and use the sufficient condition depicted in Figure 4 within each block. If the blocks are sufficiently large, then solving just one block suffices to solve the whole puzzle. We call a set $U \subseteq V$ *internally solved* if the people graph induced on U can solve the puzzle graph induced on U and prove the existence of a "large" internally solved set. We use the following lemma to partition the puzzle graph into disjoint blocks. The motivation comes from analyzing the ring puzzle graph.
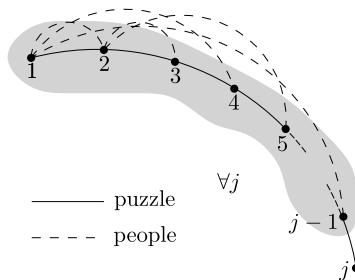

FIG. 4. *Illustration of the sufficient condition to solve the ring puzzle*: j *is people-adjacent to* $\{1, 2, \ldots, j-1\}$ *for all* $j = 2, 3, \ldots, n$. *This event is contained in the event* Solve.

LEMMA 1.   *Let $m \geq 1$ be a fixed integer. For any connected graph G with vertex set V, there exists an integer $k \geq |V|/(2m)$ and subsets $\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_k$ of V such that*:

   (i) $V = \bigcup_{i=1}^{k} \mathcal{B}_i$;
   (ii) $|\mathcal{B}_i| \in [m, 2m]$ for $i = 1, 2, \ldots, k-1$ and $|\mathcal{B}_k| < 2m$;
   (iii) *the induced subgraph on $\mathcal{B}_i$ is connected for all $i = 1, 2, \ldots, k$;*
   (iv) $\mathcal{B}_i$ *and* $\mathcal{B}_j$ *share at most one vertex in common for all* $1 \leq i < j \leq k$.

PROOF.   The proof proceeds by induction on $n := |V|$. The lemma is obviously true for $n \leq 2m$, so let us assume that $n \geq 2m + 1$.

For any connected graph G of size n, fix a spanning tree T of G. Removing a single vertex v_0 from the tree T results in finitely many disjoint components $C_1, C_2, \ldots, C_k$, each of which has a unique marked vertex adjacent to v_0 in T. We consider three disjoint cases.

CASE 1.   If one of the components has size between $[m, 2m]$, we define this component as $\mathcal{B}_1$ and use induction on the graph G with the vertex set $\mathcal{B}_1$ removed, which is still connected.

CASE 2.   If all of the components have size $< m$, define l as the smallest integer such that $|C_1| + |C_2| + \cdots + |C_{l-1}| < m$ and $|C_1| + |C_2| + \cdots + |C_l| \geq m$. Such an l exists, because $|C_1| + |C_2| + \cdots + |C_k| = n - 1 > m$. Necessarily we have $|C_1| + |C_2| + \cdots + |C_l| < 2m$, because $|C_i| < m$ for all i. We take $\mathcal{B}_1 := \bigcup_{i=1}^{l} C_i \cup \{v_0\}$ and use induction on the graph G with vertex set $\bigcup_{i=1}^{l} C_i$ removed (note that v_0 will appear in more than one subset because it has not yet been removed from G).

CASE 3.   If none of the components has size between $[m, 2m]$ and at least one component has size $> 2m$, we choose one such component (and ignore the other components), call it V_1, and remove the marked vertex v_1 from it. Removing v_1 creates several new components, each containing a marked vertex adjacent to v_1 in T. We repeat this procedure until reaching the following situation: the size of V_k is $> 2m$, but if we remove the marked vertex v_k from it, then all the resulting components have size $\leq 2m$. If one of them has size more than m, then we take that component as $\mathcal{B}_1$, and we continue by induction with the rest of the tree, which is connected by construction. If all of the components have size $< m$, we follow the steps in Case 2 to define $\mathcal{B}_1$ and continue by induction.

To complete the proof we need to check properties (iii) and (iv) for each block $\mathcal{B}_i$, which follow easily from the spanning tree and marked vertex construction.   □

PROOF OF PROPOSITION 2.   Using Lemma 1, we partition the puzzle graph into blocks $\mathcal{B}_1$, $\mathcal{B}_2, \ldots, \mathcal{B}_k$ of size $\leq 2m$ (where m is determined later) with $|\mathcal{B}_i| \geq m$ for all $i < k$. Note that $k \geq n/(2m)$. Let B_i be the event that block $\mathcal{B}_i$ is solved using only people edges in block $\mathcal{B}_i$. Let $S := \sum_{i=1}^{k-1} \mathbb{1}_{B_i}$ be the number of blocks (excluding the last block $\mathcal{B}_k$) that are solved using people edges only within each block (i.e., internally solved). The events B_i are independent because the blocks use disjoint sets of edges, and they are Bernoulli random variables with mean $\mathbb{P}(B_i)$.

Next we show that if $p_n = \lambda / \log n$ with $\lambda > \pi^2/6$, then

$$\mathbb{P}(S \geq 1) \to 1 \qquad \text{as } n \to \infty.$$

Consider the subgraph of the puzzle graph induced by $\mathcal{B}_i$. We can fix a rooted spanning tree and label the vertices with integers $1, 2, \ldots, |\mathcal{B}_i|$ in such a way that the vertex with label j is puzzle-adjacent to the set of vertices with labels $\{1, 2, \ldots, j - 1\}$ in the spanning tree for all $j \geq 1$. As illustrated in Figure 4, a sufficient condition for the event B_i to occur is the event

$$\overline{B_i} := \big\{ \text{for all } 1 \leq j \leq |\mathcal{B}_i|, \text{ the vertex labeled } j \text{ is people-adjacent}$$

$$\text{to the set of vertices labeled } \{1, 2, \ldots, j - 1\} \big\} \subset B_i.$$

[Note that there could be other ways to solve the puzzle. For example, in the case of a ring puzzle, j is people-adjacent to $j + 1$, and $j + 1$ (but not j) is people-adjacent to $\{1, \ldots, j - 1\}$. Thus $\overline{B_1}$ is not a necessary condition for B_1 to occur, that is, $\overline{B_1} \subsetneq B_1$.] The events that $j + 1$ is people-adjacent to $\{1, 2, \ldots, j\}$ occur independently with probability $\geq 1 - (1 - p_n)^j$, so

$$\mathbb{P}(\overline{B_i}) \geq \prod_{j=1}^{|\mathcal{B}_i|-1} \big(1 - (1 - p_n)^j\big) \geq \prod_{j=1}^{2m} \big(1 - (1 - p_n)^j\big).$$

Thus the random variable S stochastically dominates

$$S' \sim \text{Binomial}\left(k - 1, \prod_{j=1}^{2m} (1 - (1 - p_n)^j)\right).$$

For $n \in \mathbb{N}$, let $\varepsilon_n := -\log(1 - p_n)$, so that $1 - p_n = e^{-\varepsilon_n}$. We use the next lemma to obtain a lower bound on

$$\log \mathbb{E} S' = \log(k - 1) + \sum_{j=1}^{2m} \log\big(1 - e^{-j\varepsilon_n}\big).$$

The proof of Lemma 2 follows the present proof.

LEMMA 2. *Let* $\theta(x) := -\int_0^x \log(1 - e^{-t})\,dt$ *for* $x \in [0, \infty]$. *If* $\lim_{\varepsilon \to 0} m_\varepsilon \varepsilon = x \in [0, \infty]$, *then*

$$\lim_{\varepsilon \to 0} \varepsilon \sum_{i=1}^{m_\varepsilon} \log(1 - e^{-i\varepsilon}) = -\theta(x).$$

Moreover, for all $m \geq 1$ *and* $\varepsilon > 0$,

$$(4.1) \qquad \left| \sum_{i=1}^{m} \log(1 - e^{-i\varepsilon}) + \frac{\pi^2}{6\varepsilon} \right| \leq \frac{1}{2} \log \frac{2e^2}{\varepsilon} + \frac{\pi^2}{6\varepsilon e^{m\varepsilon}}.$$

Fix $\delta > 0$, and let $m := \lceil (1 + \delta)(\log n)/\varepsilon_n \rceil$. Here we tacitly assume that n is large, so that $2m < n$. Using Lemma 2, we estimate

$$\log \mathbb{E}(S') \geq \log\left(\frac{n}{2m} - 1 \right) - \frac{\pi^2}{6\varepsilon_n} + \left(\sum_{j=1}^{2m} \log(1 - e^{-j\varepsilon_n}) + \frac{\pi^2}{6\varepsilon_n} \right)$$

$$\geq \left(1 - \frac{\pi^2}{6\lambda} \right) \log n - \log \frac{2m}{1 - 2m/n} - \frac{1}{2} \log \frac{2e^2}{\varepsilon_n} - \frac{\pi^2}{6\varepsilon_n e^{2m\varepsilon_n}}$$

$$\geq \left(1 - \frac{\pi^2}{6\lambda} \right) \log n - \log \frac{m}{\sqrt{\varepsilon_n}} - O(1)$$

$$\geq \left(1 - \frac{\pi^2}{6\lambda} \right) \log n - \frac{5}{2} \log \log n - O(1)$$

$$\to \infty \qquad \text{as } n \to \infty.$$

In the last inequality we used the fact that $m = O(\log n / \varepsilon_n)$ and $\varepsilon_n \geq p_n = \lambda / \log n$. Since S' is binomial, $\mathbb{E}(S') \to \infty$ implies that $\mathbb{P}(S' \geq 1) \to 1$.

Let $I := \inf\{i \geq 1 : \mathcal{B}_i \text{ is internally solved}\}$ be the random index such that $\mathcal{B}_I$ is the first block among $\mathcal{B}_1, \mathcal{B}_2, \ldots$ that is internally solved. We define $I = \infty$ when no internally solved block exists. Thus we have $\mathbb{P}(I < \infty) = \mathbb{P}(S \geq 1) \geq \mathbb{P}(S' \geq 1) \to 1$ as $n \to \infty$.

Let U be a deterministic set of size m. The probability that all the remaining $n - m$ vertices in $V \setminus U$ are connected to U by a people edge is

$$(1 - (1 - p_n)^m)^{n-m} \geq (1 - e^{-\varepsilon_n m})^n \geq 1 - n e^{-\varepsilon_n m} \geq 1 - n^{-\delta}.$$

Note that by connectivity of the puzzle graph and people graph, the event that all vertices in $V \setminus U$ are connected to U by people edges and U is internally solved implies Solve. Moreover the event that a particular set of vertices forms an *internally solved* subset or not depends only on the edges among those vertices. Thus we have

$$\mathbb{P}(\text{Solve}) \geq \mathbb{P}(\text{Solve}, I < \infty)$$

$$\geq \sum_{i=1}^{k} \mathbb{P}(\text{Solve}|I = i)\mathbb{P}(I = i) \geq (1 - n^{-\delta})\mathbb{P}(I < \infty) \to 1$$

as $n \to \infty$. The proof is complete. $\square$

PROOF OF LEMMA 2. Note that

$$-\varepsilon \sum_{i=1}^{k} \log(1 - e^{-i\varepsilon}) = \varepsilon \sum_{i=1}^{k} \sum_{j=1}^{\infty} \frac{e^{-ij\varepsilon}}{j} = \varepsilon \sum_{j=1}^{\infty} \frac{1 - e^{-jk\varepsilon}}{j(e^{j\varepsilon} - 1)}$$

$$= \sum_{j=1}^{\infty} \frac{1 - e^{-jk\varepsilon}}{j^2} - \sum_{j=1}^{\infty} \frac{(1 - e^{-jk\varepsilon})(e^{j\varepsilon} - 1 - j\varepsilon)}{j^2(e^{j\varepsilon} - 1)}.$$

Using the power series expression of e^x, it is easy to see that $(e^x - 1 - x)/(e^x - 1) \le \min\{x/2, 1\}$. Applying the last inequality, we have

$$\sum_{j=1}^{\infty} \frac{(1 - e^{-jk\varepsilon})(e^{j\varepsilon} - 1 - j\varepsilon)}{j^2(e^{j\varepsilon} - 1)} \le \sum_{j=1}^{\infty} \frac{\min\{j\varepsilon/2, 1\}}{j^2} \le \sum_{j \le m} \frac{\varepsilon}{2j} + \sum_{j > m} \frac{1}{j^2}$$

$$\le \frac{\varepsilon}{2}(\log m + 1) + \frac{1}{m} = \frac{\varepsilon}{2} \log \frac{2e^2}{\varepsilon}$$

using $m = 2/\varepsilon$. Thus, combining the last two displays,

$$(4.2) \qquad \left| \sum_{i=1}^{k} \varepsilon \log(1 - e^{-i\varepsilon}) + \sum_{j=1}^{\infty} \frac{1 - e^{-jk\varepsilon}}{j^2} \right| \le \frac{\varepsilon}{2} \log \frac{2e^2}{\varepsilon}.$$

In particular, if $\lim_{\varepsilon \to 0} k_\varepsilon \varepsilon = x \in [0, \infty]$, then interchanging the sum and the integral

$$\lim_{\varepsilon \to 0} \varepsilon \sum_{i=1}^{k_\varepsilon} \log(1 - e^{-i\varepsilon}) = -\sum_{j=1}^{\infty} \frac{1 - e^{-jx}}{j^2}$$

$$= -\sum_{j=1}^{\infty} \frac{1}{j} \int_0^x e^{-jt} \, dt = \int_0^x \log(1 - e^{-t}) \, dt,$$

which completes the proof. The bound (4.1) follows from (4.2) and the fact that $e^{-jk\varepsilon} \le e^{-k\varepsilon}$ for all $j \ge 1$. $\square$

4.2. *Lower bound for the ring puzzle.* In this section, we prove a matching-order lower bound for an Erdős–Rényi people graph solving the ring puzzle. The idea of the proof is to show the existence of a cut set that divides the ring into pieces that never merge.

PROPOSITION 3. *For the ring puzzle graph, if $\lambda \le 1/27$ and $p_n = \lambda/\log n$, then $\mathbb{P}_{p_n}(\text{Solve}) \to 0$. Therefore $p_c(n) \ge 1/(27 \log n)$.*

PROOF.   Let x be a positive integer to be chosen later [it will be $\Theta(\log n)$]. We will identify the vertices in the ring puzzle graph (V, E_{puzzle}) with elements from $\mathbb{Z}_n$, so that two vertices $u, v \in \mathbb{Z}_n$ are neighbors if and only if $u - v = \pm 1$, where all additions and subtractions in $\mathbb{Z}_n$ are modulo n. We denote the interval $\{a, a + 1, \ldots, b\} \subseteq \mathbb{Z}_n$ by $[a, b]$ and its length by $|[a, b]| = b - a + 1$.

Given an interval $I = [a, b] \subset \mathbb{Z}_n$, we call it x-good if there is a vertex $u \in I$ such that u is not people-adjacent to any vertex in the interval $[a - x, b + x]$. We call the vertex $u \in I$ an x-good vertex in I. The proof hinges on the following observation. Loosely speaking, if throughout the puzzle there are people unacquainted with anyone in a sufficiently large neighborhood of the puzzle, then these people obstruct the growing solution, and the social network cannot solve the puzzle.

LEMMA 3.   *Suppose that there exist integers $0 = a_0 < a_1 < \cdots < a_k = n$ such that, for all $j = 0, 1, \ldots, k - 1$, the interval $I_j := [a_j + 1, a_{j+1}]$ is x-good and has length $|I_j| \leq x$. Then the puzzle cannot be solved.*

PROOF.   Let $v_j \in I_j$ be an x-good vertex in I_j for $j = 0, 1, \ldots, k - 1$. Clearly $1 \leq v_0 < v_1 < \cdots < v_{k-1} \leq n$. Furthermore, each v_j has no people edges with $[v_{j-1}, v_{j+1}]$ (where $j + \ell$ is taken modulo k) because $|I_j| \leq x$ for all $j = 0, 1, \ldots, k - 1$.

Suppose for contradiction that the puzzle can be solved. Then there must exist a first stage, i, after which there exists an index j such that two distinct vertices, $u \in [v_j, v_{j+1}]$ and $v \in [v_{j+1}, v_{j+2}]$, belong to the same cluster in $\mathcal{C}_i$. One of these vertices must be v_{j+1} (without loss of generality, $u = v_{j+1}$), because otherwise v_{j+1} would have to belong to a larger cluster in $\mathcal{C}_{i-1}$, and therefore v_{j+1} would have merged at an earlier stage of the process, which is a contradiction. Since v_{j+1} is not people-adjacent to any other vertices in $[v_{j+1}, v_{j+2}]$, v must be in a component in $\mathcal{C}_{i-1}$ that contains vertices outside of $[v_{j+1}, v_{j+2}]$, but this is also a contradiction. Thus the puzzle cannot be solved.   □

In light of Lemma 3, to complete the proof we need to show the existence of such intervals with probability tending to 1. Suppose $n \geq x^2$. Define $k := \lfloor n/(x - 1) \rfloor \leq n$. Define

$$l_i := x \qquad \text{for } 1 \leq i \leq n - k(x - 1),$$

$$l_i := x - 1 \qquad \text{for } n - k(x - 1) < i \leq k,$$

and $a_i := l_1 + l_2 + \cdots + l_i$ for $i = 0, 1, \ldots, k$. Note that $a_k = n$. Clearly all the intervals $I_i := [a_i + 1, a_{i+1}], 0 \leq i \leq k - 1$ are of length $x - 1$ or x. Let Z be the number of intervals that are not x-good,

$$Z := \sum_{i=0}^{k-1} \mathbb{1}_{\{\text{the interval } I_i \text{ is NOT x-good}\}}.$$

It suffices to show that $\mathbb{P}(Z > 0) \to 0$ as $n \to \infty$ for appropriate choice of x. We will use Lemma 4 to estimate the probability that an interval is not x-good.

LEMMA 4. *Fix an integer $x \geq 1$. Let I be an interval of length lx for some number $l > 0$. Suppose that $t := px \in (0, 1/(l+2))$. Then we have*

$$\mathbb{P}(I \text{ is NOT } x\text{-good}) \leq \exp\left[-\frac{t}{2p}\left(2l\log(\sqrt{1+l/t}-1) + (l^2 + 4l + 2)t\right.\right.$$
$$\left.\left. - 2t\sqrt{1+l/t} - 2l\log l - l\right)\right].$$

In our case, all intervals are of length $x - 1$ or x, so $l \in [1 - 1/x, 1]$. If we suppose that $t := px < 1/3$, then

$$\mathbb{P}(Z > 0) \leq \mathbb{E}(Z)$$
$$\leq n \exp\left[-\frac{t}{2p}\left(2\log(\sqrt{1+1/t}-1) + 7t - 2t\sqrt{1+1/t} - 1 + \eta(x)\right)\right],$$

where $\eta(x) \to 0$ when $x \to \infty$. In particular, if $p = p_n = \lambda/\log n$ and $x = t\log n/\lambda$ for some $t < 1/3$, we have

$$\mathbb{P}(Z > 0) \leq \exp\left[\log n - \frac{t\log n}{2\lambda}\left(2\log(\sqrt{1+1/t}-1) + 7t\right.\right.$$
$$\left.\left. - 2t\sqrt{1+1/t} - 1 + \eta(t\log n/\lambda)\right)\right]$$
$$\to 0 \qquad \text{as } n \to \infty$$

when

$$(4.3) \qquad \lambda < \frac{t}{2}\left[2\log(\sqrt{1+1/t}-1) + 7t - 2t\sqrt{1+1/t} - 1\right].$$

One can easily check (by taking $t = 0.07$) that

$$\sup_{t \in (0,1/3)} \frac{t}{2}\left[2\log(\sqrt{1+1/t}-1) + 7t - 2t\sqrt{1+1/t} - 1\right] > 1/27.$$

Thus given $\lambda \leq 1/27$, we can choose $t \in (0, 1/3)$ such that (4.3) holds, and taking $x = t\log n/\lambda$ we have

$$\mathbb{P}\left(\sum_{i=0}^{k-1} \mathbb{1}_{\{\text{the interval } I_i \text{ is NOT } x\text{-good}\}} > 0\right) \to 0 \qquad \text{as } n \to \infty.$$

This completes the proof. $\square$

PROOF OF LEMMA 4. Without loss of generality, suppose that the interval I is $[1, lx]$. Recall that I is x-good if there is a vertex $u \in I$ such that

u has no people edges with $I_x := [1 - x, lx + x]$. Thus I is not x-good implies that all vertices in I have at least one people edge with I_x, in other words $\sum_{j \in I_x} \mathbb{1}_{\{i \text{ has a people edge with } j\}} \geq 1$ for all $i \in I$, and thus

$$\sum_{i \in I} \sum_{j \in I_x} \mathbb{1}_{\{i \text{ has a people edge with } j\}} \geq lx.$$

The number of distinct pairs of vertices between I and $I_x \setminus I$ is $2lx^2$, and the number of distinct pairs of vertices within I is $\binom{lx}{2}$. Therefore

$$\sum_{i \in I} \sum_{j \in I_x} \mathbb{1}_{\{i \text{ has a people edge with } j\}} \stackrel{\mathrm{d}}{=} X + 2Y,$$

where $X \sim \mathrm{Bin}(2lx^2, p)$, $Y \sim \mathrm{Bin}(\binom{lx}{2}, p)$ and X, Y are independent. In particular, we have

$$\mathbb{P}(I \text{ is not } x\text{-good}) \leq \mathbb{P}(X + 2Y \geq lx)$$
$$\leq \mathbb{P}(X + 2Y' \geq lx) \leq e^{-\theta lx} \mathbb{E}(e^{\theta X + 2\theta Y'})$$

for any $\theta > 0$, where $Y' \sim \mathrm{Bin}(l^2 x^2/2, p)$ is independent of X. We have

$$(4.4) \quad \begin{aligned} \mathbb{P}(X + 2Y' \geq lx) &\leq e^{-\theta lx}(1 - p + pe^\theta)^{2lx^2}(1 - p + pe^{2\theta})^{l^2 x^2/2} \\ &\leq \exp[-lx(\theta - 2t(e^\theta - 1) - lt(e^{2\theta} - 1)/2)], \end{aligned}$$

where $t := px$. Note that we have

$$\frac{\mathbb{E}(X + 2Y')}{lx} = (l + 2)px = (l + 2)t.$$

Hence, under the assumption $t \in (0, 1/(l + 2))$, we have $lx > \mathbb{E}(X + 2Y')$ and $\sqrt{1 + l/t} - 1 > l$. Taking $\theta = \log[(\sqrt{1 + l/t} - 1)/l]$ in (4.4), we finally have

$$\begin{aligned} \mathbb{P}(I \text{ is not } x\text{-good}) \leq \exp\Big[-\frac{t}{2p}\Big(2l \log(\sqrt{1 + l/t} - 1) + (l^2 + 4l + 2)t \\ - 2t\sqrt{1 + l/t} - 2l \log l - l\Big)\Big]. \end{aligned}$$

This completes the proof. $\quad \square$

Propositions 2 and 3 give Theorem 1.

4.3. *Lower bound for puzzles with bounded degree.* In this section, we prove the lower bound in Theorem 2 for arbitrary puzzle graphs with bounded degree as $n \to \infty$.

PROPOSITION 4. *For any sequence of connected puzzle graphs with bounded maximum degree as $|V| = n \to \infty$, $p_c(n) = \omega(1/n^b)$ for any $b > 0$.*

PROOF. Let $p = n^{-b}$ such that $k \geq 2$ and $b \in (\frac{1}{k}, \frac{1}{k-1})$ are fixed, and suppose that the maximum degree of (V, E_{puzzle}) is at most D for all n. After stage i we have a collection of jigsaw clusters $\mathcal{C}_i$. Initially $\mathcal{C}_0 = \{\{v\} : v \in V\}$, and after the first stage $\mathcal{C}_1$ is the set of connected components in the graph $(V, E_{\text{people}} \cap E_{\text{puzzle}})$. Thereafter, two clusters $U, U' \in \mathcal{C}_i$ merge if there is an edge between the two clusters in E_{people} and an edge between the two clusters in E_{puzzle}. Therefore, if $U, U' \in \mathcal{C}_i$, then $U, U' \subset W \in \mathcal{C}_{i+1}$ if and only if there is some nonnegative integer ℓ and a sequence of clusters $U = U_0, U_1, \ldots, U_\ell = U' \in \mathcal{C}_i$ such that U_j merges with U_{j+1} at stage $i + 1$.

Observe that for $i \geq 1$, every merge event in stage $i + 1$ must involve at least one cluster that was formed by a merge in stage i. Inspired by this observation, we let $\mathcal{A}_i \subseteq \mathcal{C}_i$ be the set of *active* clusters that were the result of at least one merge in stage i when $i \geq 1$, and let $\mathcal{A}_0 = \mathcal{C}_0$. Next we define the events E_i and F_i for $i = 0, \ldots, k$ as

$$E_i = \{|\mathcal{A}_i| \geq C_i n^{1-ib}\},$$
$$F_i = \{\max\{|W| : W \in \mathcal{C}_i\} \geq L_i\},$$

where C_i and L_i are constants that depend on d and k, which we will define later. In words, E_i is the event that there are at least $C_i n^{1-ib}$ active clusters following stage i, which is contained in the event that at least $C_i n^{1-ib}$ merges occur at stage i, because each active cluster must be the result of at least one merge. F_i is the event that the largest cluster following stage i has at least L_i vertices. For sufficiently large n, the event E_k is equivalent to the event that at least one merge occurs at stage k, because $kb > 1$. Therefore, our goal is to show that $\mathbb{P}(E_k) \to 0$ and $\mathbb{P}(F_k) \to 0$ as $n \to \infty$, which implies that no merges occur after stage k and that the largest cluster has size at most L_k, so the puzzle remains unsolved.

Our strategy is to prove this by induction on i. It is trivially true that $\mathbb{P}(E_0) = 0$ and $\mathbb{P}(F_0) = 0$ with $C_0 = 2$ and $L_0 = 2$. Now, let us assume that $\mathbb{P}(E_i) \to 0$ and $\mathbb{P}(F_i) \to 0$ as $n \to \infty$ for some $i \in \{0, 1, \ldots, k-1\}$, which implies that $\mathbb{P}(E_i^c \cap F_i^c) \to 1$. On the event $E_i^c \cap F_i^c$, we know that the number of active clusters is $|\mathcal{A}_i| < C_i n^{1-ib}$, and the largest cluster has at most L_i vertices. The latter implies that every cluster has fewer than DL_i neighboring clusters in (V, E_{puzzle}) because each vertex has at most D total neighboring vertices in the puzzle graph. We will use this fact in two ways. First, we will show that the number of merges at stage $i + 1$ is small because each active cluster after stage i has relatively few opportunities to merge. Second, we will show that no path of neighboring clusters longer than length $k - i$ merge at stage $i + 1$ because few such paths exist.

To meet our first goal, we define a random variable $I_{\{A,B\}}^{i+1}$ for each pair of an active cluster $A \in \mathcal{A}_i$ and a neighboring cluster $B \in \mathcal{C}_i$ such that $B \neq A$, and there is an edge in E_{puzzle} between A and B. The random variable $I_{\{A,B\}}^{i+1}$ is the indicator of the event that A and B merge at stage $i + 1$. On the event F_i^c, the probability

that A merges with B is at most

$$(4.5) \qquad 1 - \left(1 - n^{-b}\right)^{(DL_i)^2} \leq 1 - \left(1 - (DL_i)^2 n^{-b}\right) = (DL_i)^2 n^{-b},$$

where we use the fact that $(1 - x)^n \geq 1 - nx$ for $x \in (0, 1)$. For convenience, we now order the clusters in $\mathcal{C}_i$ so that $A_1, A_2, \ldots, A_{|\mathcal{A}_i|} \in \mathcal{A}_i$ and $A_{|\mathcal{A}_i|+1}, A_{|\mathcal{A}_i|+2}, \ldots, A_{|\mathcal{C}_i|} \in \mathcal{C}_i \setminus \mathcal{A}_i$. Therefore, on $E_i^c \cap F_i^c$, the total number of merges that occur in stage $i + 1$,

$$\sum_{j=1}^{|\mathcal{A}_i|} \sum_{\ell=j+1}^{|\mathcal{C}_i|} I_{\{A_j, A_\ell\}}^{i+1},$$

is stochastically dominated by $X_i \sim \text{Binomial}(DL_i C_i n^{1-ib}, (DL_i)^2 n^{-b})$. This is because there are at most $DL_i C_i n^{1-ib}$ distinct pairs of neighboring clusters, at least one of which is active, and the events that each of these pairs merges at stage $i + 1$ are independent because they depend on disjoint sets of edges in the people graph. If we let $C_{i+1} = 2(DL_i)^3 C_i$ (this is $2\mathbb{E}X_i / n^{1-(i+1)b}$), then by Chebyshev's inequality

$$\mathbb{P}(E_{i+1} | E_i^c \cap F_i^c) = \mathbb{P}\left(\sum_{j=1}^{|\mathcal{A}_i|} \sum_{\ell=j+1}^{|\mathcal{C}_i|} I_{\{A_j, A_\ell\}}^{i+1} \geq C_{i+1} n^{1-(i+1)b} \,\middle|\, E_i^c \cap F_i^c\right)$$

$$\leq \mathbb{P}\left(X_i \geq C_{i+1} n^{1-(i+1)b}\right)$$

$$= \mathbb{P}(X_i - \mathbb{E}X_i \geq \mathbb{E}X_i)$$

$$\leq (\mathbb{E}X_i)^{-1} = O\left(n^{-1+(i+1)b}\right) \to 0.$$

Since $\mathbb{P}(E_i^c \cap F_i^c) \to 1$, we have that $\mathbb{P}(E_{i+1}) \to 0$.

Next we must show that the largest cluster after stage $i + 1$ has size at most L_{i+1}. Define a *cluster path* of length $\ell \geq 0$ between $U, U' \in \mathcal{C}_i$ to be a sequence of distinct clusters $U = U_0, U_1, \ldots, U_\ell = U' \in \mathcal{C}_i$ such that U_j and U_{j+1} are puzzle-adjacent for all $j \in \{0, \ldots, \ell - 1\}$. For a fixed cluster $A \in \mathcal{C}_i$, let Y_A^i denote the number of cluster paths of length k that start at A (meaning that $U_0 = A$) and such that U_j will merge with U_{j+1} at stage $i + 1$ for each $j \in \{0, \ldots, k - 1\}$. For any cluster path $U_0, \ldots, U_k$, the probability that U_j and U_{j+1} merge at stage $i + 1$ is bounded above by $(DL_i)^2 n^{-b}$ on the event F_i^c, by inequality (4.5). The number of cluster paths of length k in after stage i that start at A is bounded by $(DL_i)^k$ on F_i^c because each cluster has at most DL_i neighboring clusters. Therefore, by Markov's inequality,

$$\mathbb{P}\left(\sum_{A \in \mathcal{C}_i} Y_A^i \geq 1 \,\middle|\, F_i^c\right) \leq n \mathbb{P}(Y_A^i \geq 1 | F_i^c)$$

$$\leq n\left[(DL_i)^k \left((DL_i)^2 n^{-b}\right)^k\right] = O\left(n^{1-kb}\right) \to 0.$$

This implies that there are no cluster paths of length k or longer that merge at stage $i + 1$. Note that clustering can occur in any tree-like pattern, and the maximum size of a rooted tree with depth (maximum distance from the root) k and maximum degree DL_i is $L_i(1 + (DL_i)^1 + (DL_i)^2 + \cdots + (DL_i)^k) = L_i((DL_i)^{k+1} - 1)/(DL_i - 1)$.

In turn, this implies that the largest cluster after stage $i + 1$ is smaller than $L_{i+1} := L_i((DL_i)^{k+1} - 1)/(DL_i - 1)$ with high probability on the event F_i^c, so $\mathbb{P}(F_{i+1}) \to 0$, which completes the proof. $\square$

Propositions 2 and 4 give Theorem 2.

5. People graphs with limiting power-law degree distributions. In this section, we prove Proposition 1, which states that a configuration model random people graph with limiting power-law degree distribution having exponent $\alpha > 2$ cannot solve bounded-degree puzzles with high probability. Recall that a set $U \subseteq V$ is *internally solved* if the people graph induced on U can solve the puzzle graph induced on U. We will call this event $\texttt{Solve}_U$. The idea is to show that with high probability no set of vertices of a certain, finite size is internally solved.

LEMMA 5. *Suppose $U \subseteq V$ such that $|U| = m > 1 + \frac{2\alpha}{\alpha-2}$ is constant. Then*

$$\mathbb{P}(\texttt{Solve}_U) = o(n^{-1}).$$

PROOF. Without loss of generality, suppose that $U = [m]$. Fix $\gamma := \alpha/2 \in (1, \alpha - 1)$ and $\varepsilon := 1/2 - 1/\alpha$, so that $(1 - \varepsilon)\gamma > 1$. It is easy to see that $\mathbb{E}d_1^\gamma < \infty$. Define the event

$$\mathcal{D}_{n,m} := \{\text{there exists a pair of indices } 1 \le i < j \le m, \text{ such that } d_i d_j \ge n^{1-\varepsilon}\}.$$

By union bound and Markov's inequality, we have

$$(5.1) \quad \mathbb{P}(\mathcal{D}_{n,m}) \le \binom{m}{2} \mathbb{P}(d_1 d_2 \ge n^{1-\varepsilon}) \le \binom{m}{2} \frac{\mathbb{E}(d_1^\gamma)\mathbb{E}(d_2^\gamma)}{n^{(1-\varepsilon)\gamma}} = o(n^{-1}).$$

Observe that the event $\texttt{Solve}_U$ implies that the people graph induced by U is connected, which in turn implies that it contains at least $m - 1$ (nonloop) edges. Partitioning on $\mathcal{D}_{n,m}$, we have

$$(5.2) \quad \begin{aligned} &\mathbb{P}(\texttt{Solve}_U) \\ &\le \mathbb{P}(\mathcal{D}_{n,m}) + \mathbb{P}\big(E_{\text{people}}|_U \text{ has } \ge m - 1 \text{ nonloop edges}, \mathcal{D}_{n,m}^c\big). \end{aligned}$$

Let $\mathcal{F}_{\mathbf{k}} := \{d_1 = k_1, \ldots, d_m = k_m\}$ be the event that the degrees of the vertices in U are $\mathbf{k} := (k_1, \ldots, k_m)$. On the event $\mathcal{F}_{\mathbf{k}}$, label the half-edges at vertex $u \in U$ as $(u, 1), (u, 2), \ldots, (u, k_u)$. Let

$$(5.3) \quad \begin{aligned} &\mathcal{E} = \mathcal{E}(\mathbf{k}) \text{ denote the set of all pairs of half-edges } \big\{(u, \ell_u), (v, \ell_v)\big\} \\ &\text{such that } 1 \le u < v \le m, 1 \le \ell_u \le k_u \text{ and } 1 \le \ell_v \le k_v. \end{aligned}$$

Note that $\mathcal{E}$ does not contain any pairs of half-edges that would form a self-loop if joined.

Conditional on $\mathcal{F}_\mathbf{k}$, for each $e \in \mathcal{E}$, let Y_e be the indicator that the half-edges in e are matched in the construction of the configuration model graph, so E_people contains an edge between the vertices of e. The number of nonloop people edges between vertices of U is then $X_m = \sum_{e \in \mathcal{E}} Y_e$. By Markov's inequality, the probability of $\{X_m \geq m - 1\}$ given $\mathcal{F}_\mathbf{k}$ is at most the expected number of subsets of $\mathcal{E}$ with size $m - 1$ such that all half-edge pairs in the subset get matched in the construction of the configuration model graph. Therefore,

$$\mathbb{P}(X_m \geq m - 1 | \mathcal{F}_\mathbf{k}) \leq |\mathcal{E}|^{m-1} \max \mathbb{P}(Y_{e_1} = \cdots = Y_{e_{m-1}} = 1 | \mathcal{F}_\mathbf{k}),$$

where the maximum is taken over all subsets of size $m - 1$ of $\mathcal{E}$. If $\mathcal{F}_\mathbf{k} \subseteq \mathcal{D}_{n,m}^c$, then on the event $\mathcal{F}_\mathbf{k}$,

$$|\mathcal{E}| = \sum_{1 \leq u < v \leq m} k_u k_v \leq m^2 n^{1-\varepsilon}.$$

For any fixed set of half-edge pairs, $e_1, \ldots, e_{m-1} \in \mathcal{E}$, we consider the probability of matching each of these pairs sequentially in the configuration model. Since $d_\text{min} \geq 3$, each vertex outside of U has at least 3 half-edges, so each half-edge among the first $2(m - 1)$ that get matched have at least $3(n - m) - 2(m - 1) \geq n$ (for large n) choices for half-edges to get matched with. Therefore,

$$\mathbb{P}(Y_{e_1} = \cdots = Y_{e_{m-1}} = 1 | \mathcal{F}_\mathbf{k}) \leq \left(\frac{1}{n}\right)^{m-1}.$$

The last three displays imply that

$$\mathbb{P}(X_m \geq m - 1 | \mathcal{F}_\mathbf{k}) \leq m^{2m} n^{-\varepsilon(m-1)},$$

provided $\mathcal{F}_\mathbf{k} \subseteq \mathcal{D}_{n,m}^c$. Therefore,

(5.4)
$$\begin{aligned} &\mathbb{P}\big(E_\text{people}|_U \text{ has } \geq m - 1 \text{ nonloop edges}, \mathcal{D}_{n,m}^c\big) \\ &= \sum_{\mathbf{k}\,:\,\mathcal{F}_\mathbf{k} \subseteq \mathcal{D}_{n,m}^c} \mathbb{P}(\mathcal{F}_\mathbf{k})\mathbb{P}(X_m \geq m - 1 | \mathcal{F}_\mathbf{k}) \leq m^{2m} n^{-\varepsilon(m-1)}. \end{aligned}$$

Choosing m such that $\varepsilon(m - 1) > 1$, and combining equations (5.1), (5.2) and (5.4) show that $\mathbb{P}(\texttt{Solve}_U) = o(n^{-1})$. $\square$

Finally we are ready to complete the proof of Proposition 1. First, observe that the jigsaw percolation process can be slowed down, such that at every step only a single pair of clusters is merged. The final set of clusters after all possible merges are made will be the same as in the original formulation, but in the slowed down version, the size of the largest cluster can at most double at each step. This means that for any $k \leq n/2$,

(5.5)
$$\mathbb{P}(\texttt{Solve}) \leq \mathbb{P}\bigg(\bigcup_{m \in [k, 2k]} \bigcup_{U \subset V, |U| = m} \texttt{Solve}_U\bigg).$$

Furthermore, observe that the second union on the right-hand side can be restricted to only those subsets $U \subset V$ that are connected in (V, E_{puzzle}). The number of connected subsets of vertices in (V, E_{puzzle}) of size m is crudely bounded above by $n \cdot (m-1)! D^{m-1}$. This bound is obtained by building a connected set U of size m by first choosing a starting vertex v, in n ways, then adding one vertex at a time to U until U contains m vertices. When U contains ℓ vertices, there are at most ℓD vertices that are adjacent to a vertex in U that can be added in the next step. If we fix $k > 1 + \frac{2\alpha}{\alpha-2}$, then (5.5) and Lemma 5 imply that

$$\mathbb{P}(\texttt{Solve}) \leq (k+1)(2k)! D^{2k} \cdot n \cdot \max_{m \in [k, 2k]} \max_{U \subset V, |U| = m} \mathbb{P}(\texttt{Solve}_U) = o(1).$$

6. Discussion and future directions. In our early attempts to understand jigsaw percolation on the ring graph, we tried to use simulations to inform our conjectures about the critical value $p_c(n)$ [Figure 5(a)]. However, as with bootstrap percolation [20], we expect a slow rate of convergence to the critical value.

CONJECTURE 1. *For jigsaw percolation on the ring puzzle graph with an Erdős–Rényi people graph, there exist constants $b > 0$, $c_1 > 0$ and c_2 such that*

$$p_c(n) = \frac{c_1}{\log n} + \frac{c_2}{(\log n)^{1+b}} + o\big((\log n)^{-1-b}\big).$$

If true, this means that estimating c_1 to within 1% via simulation would require taking n to be at least $\exp[(100 c_2/c_1)^{1/b}]$, which is prohibitively large if $|c_2/c_1|$ is much larger than 0.1, and b is at most 1. However, we expect our upper bound on $p_c(n)$ to be tight for the ring graph.

(a) Fraction of trials in which the people graph solves the $n = 1000$ ring puzzle
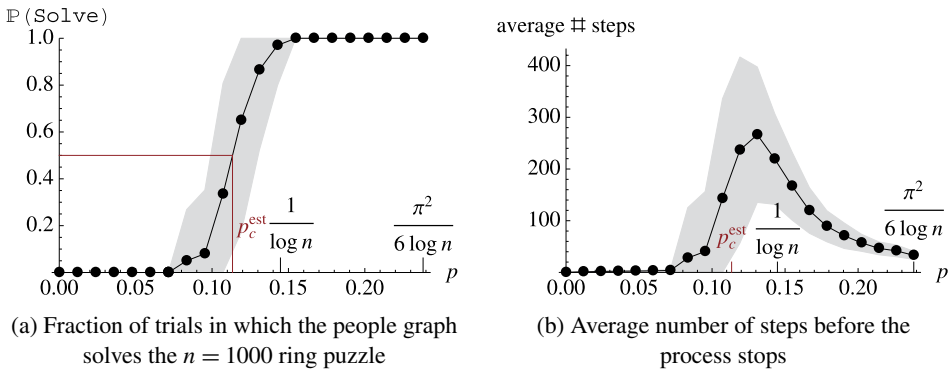
(b) Average number of steps before the process stops

FIG. 5. *Simulations of jigsaw percolation on a ring of size $n = 1000$, with 200 trials for 21 equally spaced values of $p \in [0, 1.05 \times \pi^2/(6 \log n)]$ (which took 57 days on a department server). Dots are averages of 200 trials, while shaded gray areas denote ± 1 standard deviation. The estimated critical value $p_c^{\text{est}} \approx 0.11$, denoted in red, is obtained by fitting a line between the two data points with $\mathbb{P}_p(\texttt{Solve})$ just below and above $1/2$. Characterizing the average number of time steps before the process terminates (b) remains an open question.*

CONJECTURE 2.   *For jigsaw percolation on the ring puzzle graph, $c_1 = \pi^2/6$.*

This conjecture is based on a computation (not shown here) that implies that a two-sided growth version of the sufficient condition used in the proof of Proposition 2 (i.e., the one-sided requirement that j is connected to $\{1, 2, \ldots, j-1\}$ for each j) yields the same upper bound of $\pi^2/(6 \log n)$ but with a correction of order $(\log n)^{-3/2}$. Of course, even when the two-sided growth process fails starting from every vertex, it may still be possible to solve the puzzle by merging the clusters formed. However, if none of these "two-sided growth clusters" intersect, then the puzzle is unlikely to be solved, so we suspect that $c_1 = \pi^2/6$ is the correct lower bound.

Of particular interest for future study, the number of steps until the process stops measures how efficiently the network solves the puzzle or determines that it cannot be solved. We numerically simulated the average number of steps until the process terminates for the ring puzzle [Figure 5(b)]. As expected, the number of steps increases around the phase transition $p_c(n)$. The process terminates quickly when the puzzle is not solved, and the proof of Proposition 2 implies that the number of steps is at most $O(\log n/p_n)$, though this is not the best bound possible. The proof of Proposition 3 shows that for the ring puzzle with $p_n \leq 1/(27 \log n)$, the largest jigsaw cluster (and hence number of steps) is smaller than $\log n$. As p_n increases near $p_c(n)$, the puzzle may be solved, but just barely, so the number of steps required is largest. As p_n increases further, more people-edges leads to larger clusters early in the process. Determining the form of the function in Figure 5(b) is an interesting open problem.

OPEN PROBLEM 1.   *For the ring puzzle, let N_n be the smallest value of i such that $C_i = C_{i+1}$. Determine the asymptotic behaviors of*

$$\mathbb{E}_{p_n}[N_n | \texttt{Solve}^c] \quad and \quad E_{p_n}[N_n | \texttt{Solve}]$$

as functions of p_n.

Finally, we suspect that the phase transition at $p_c(n)$ is sharp, in the following sense.

CONJECTURE 3.   *Define $p_\varepsilon(n)$ as the unique p for which $\mathbb{P}_p(\texttt{Solve}) = \varepsilon$. Then*

$$p_\varepsilon(n)/p_{1-\varepsilon}(n) \to 1$$

as $n \to \infty$ for any $\varepsilon \in (0, 1)$ fixed.

Other avenues of future study include extensions and modifications of jigsaw percolation. Different people and puzzle graphs (especially ones with unbounded degree) are one natural direction, with mathematical and practical interest.

OPEN PROBLEM 2. *Consider other people and puzzle graphs, especially puzzles with unbounded degree.*

Another natural direction is to modify the model to make it more realistic. For example, by analogy with the "adjacent-edge" modification of explosive percolation [15], in the "adjacent-edge" (AE) version of jigsaw percolation, the rule for merging two clusters U and W requires that the people- and puzzle-edges between U and W coincide on at least one vertex. That is, in the AE rule, two jigsaw clusters U and W merge only if there exist $u \in U$ and $w, w' \in W$ such that $(u, w) \in E_{\text{puzzle}}$ and $(u, w') \in E_{\text{people}}$. In this version, a single person must determine whether her friends' jigsaw clusters fit with her piece of the puzzle, but she does not need to be aware of how her entire jigsaw cluster fits with the clusters of her acquaintances. This process is slightly more local, so we suspect that more detailed, rigorous results are possible. Note that all of our results for jigsaw percolation also hold for AE jigsaw percolation.

OPEN PROBLEM 3. *Does the behavior of AE jigsaw percolation differ significantly from that of jigsaw percolation for some class of puzzle graphs? Can more precise statements be made about the behavior of AE jigsaw percolation on the ring graph?*

Another potentially interesting modification is to change the map from people to puzzle pieces so that it is no longer bijective. This would allow many people to have the same idea and a single person to have multiple ideas.

OPEN PROBLEM 4. *What is the effect of changing the map between people and puzzle pieces on a network's ability to solve the puzzle?*

In this paper, each person has one unique puzzle piece (or idea). The critical value $p_c(n)$ marks the phase transition in the connectivity of the Erdős–Rényi people graph at which it begins to solve the puzzle with high probability. For a large class of puzzle graphs (n-cyles in Theorem 1, bounded-degree puzzles in Theorem 2), we show that this phase transition decreases with n. However, the *critical average degree*, $np_c(n)$, increases with the size n of the social network and of the puzzle. Thus, as social networks and the puzzles they try to solve grow commensurately in size, people must interact with more people in order to realize enough compatible, partial solutions. This model therefore suggests a mechanism for the recent statistical claims that as cities become more dense, people interact more [34] and hence innovate more [4, 8]. Furthermore, most social networks wish to minimize communication overhead; the critical value $p_c(n)$ indicates the minimal communication needed to collaboratively solve large puzzles.

Surprisingly, social networks with power-law degree distributions lack the connectivity needed to solve bounded-degree puzzles (Proposition 1). However, scientific collaboration networks manage to solve puzzles despite their heavy-tailed

degree distributions [3, 31, 32]. This highlights the importance of considering more realistic assumptions in the model and of drawing from (still nascent) studies on knowledge spaces [10].

This work, the first step in analyzing a rich, mathematical model, begins to suggest why certain social networks stifle creativity and why others innovate. With a homogeneous degree distribution and sufficiently many interactions, a social network can collectively merge the pieces of a large puzzle—and perhaps merge the ideas that lead to a great idea.

REFERENCES

[1] AIZENMAN, M. and LEBOWITZ, J. L. (1988). Metastability effects in bootstrap percolation. *J. Phys. A* **21** 3801–3813. MR0968311

[2] BALL, P. (2014). Crowd-sourcing: Strength in numbers. *Nature* **506** 422–423.

[3] BARABÁSI, A. L., JEONG, H., NÉDA, Z., RAVASZ, E., SCHUBERT, A. and VICSEK, T. (2002). Evolution of the social network of scientific collaborations. *Phys. A* **311** 590–614. MR1943379

[4] BETTENCOURT, L., LOBO, J., HELBING, D., KÜHNERT, C. and WEST, G. B. (2007). Growth, innovation, scaling, and the pace of life in cities. *Proc. Natl. Acad. Sci. USA* **104** 7301.

[5] BETTENCOURT, L. M. A., CINTRÓN-ARIAS, A., KAISER, D. I. and CASTILLO-CHÁVEZ, C. (2006). The power of a good idea: Quantitative modeling of the spread of ideas from epidemiological models. *Phys. A* **364** 513–536.

[6] BETTENCOURT, L. M. A., KAISER, D. I. and KAUR, J. (2009). Scientific discovery and topological transitions in collaboration networks. *J. Informetr.* **3** 210–221.

[7] BETTENCOURT, L. M. A., KAISER, D. I., KAUR, J., CASTILLO-CHÁVEZ, C. and WOJICK, D. E. (2008). Population modeling of the emergence and development of scientific fields. *Scientometrics* **75** 495–518.

[8] BETTENCOURT, L. M. A., LOBO, J., STRUMSKY, D. and WEST, G. B. (2010). Urban scaling and its deviations: Revealing the structure of wealth, innovation and crime across cities. *PLoS ONE* **5** e13541.

[9] CHAI, S. and FLEMING, L. (2011). Emergence of Breakthroughs. In *DIME-DRUID ACADEMY Winter Conference* 1–47. DRUID-DIME Academy, Aalborg, Denmark.

[10] CHEN, C., CHEN, Y., HOROWITZ, M., HOU, H., LIU, Z. and PELLEGRINO, D. (2009). Towards an explanatory and computational theory of scientific discovery. *J. Informetr.* **3** 191–209.

[11] CHUNG, F. and LU, L. (2002). The average distances in random graphs with given expected degrees. *Proc. Natl. Acad. Sci. USA* **99** 15879–15882 (electronic). MR1944974

[12] CIOINSIGHT (2004). *Web extra: Who's on first*?, CIOinsight (2004), 1–2. Available at http://www.cioinsight.com/c/a/Past-News/Web-Extra-Whos-on-First/.

[13] COPPERSMITH, D., GAMARNIK, D. and SVIRIDENKO, M. (2002). The diameter of a long-range percolation graph. *Random Structures Algorithms* **21** 1–13. MR1913075

[14] COWAN, R. and JONARD, N. (2007). Structural holes, innovation and the distribution of ideas. *J. Econ. Interac. Coord.* **2** 93–110.

[15] D'SOUZA, R. M. and MITZENMACHER, M. (2010). Local cluster aggregation models of explosive percolation. *Phys. Rev. Lett.* **104** 195702.

[16] DUCH, J., WAITZMAN, J. S. and AMARAL, L. A. N. (2010). Quantifying the performance of individual players in a team activity. *PLoS ONE* **5** e10937.

[17] ERDŐS, P. and RÉNYI, A. (1961). On the strength of connectedness of a random graph. *Acta Math. Acad. Sci. Hung.* **12** 261–267. MR0130187

[18] GERSTEIN, M. and DOUGLAS, S. M. (2007). RNAi development. *PLoS Comput. Biol.* **3** e80.

[19] GOWERS, T. and NIELSEN, M. (2009). Massively collaborative mathematics. *Nature* **461** 879–881.

[20] GRAVNER, J. and HOLROYD, A. E. (2008). Slow convergence in bootstrap percolation. *Ann. Appl. Probab.* **18** 909–928. MR2418233

[21] GRAVNER, J. and SIVAKOFF, D. (2013). Nucleation scaling in jigsaw percolation. Preprint. Available at arXiv:1310.2194.

[22] GRIMMETT, G. (1999). *Percolation*, 2nd ed. Springer, Berlin. MR1707339

[23] HOLROYD, A. E. (2003). Sharp metastability threshold for two-dimensional bootstrap percolation. *Probab. Theory Related Fields* **125** 195–224. MR1961342

[24] INTRONE, J., LAUBACHER, R., OLSON, G. and MALONE, T. (2011). The Climate CoLab: Large scale model-based collaborative planning. In *International Conference on Collaboration Technologies and Systems* (*CTS*), *May* 2011 40–47. MIT Center for Collective Intelligence, Cambridge, MA.

[25] JOHNSON, S. (2010). *Where Good Ideas Come from*: *The Natural History of Innovation*. Riverhead Hardcover, New York.

[26] LAKHANI, K. R., GARVIN, D. A. and LONSTEIN, E. (2010). TopCoder (A): Developing software through crowdsourcing. Harvard Business School General Management Unit 610–032 1–18.

[27] LAMBIOTTE, R. and PANZARASA, P. (2009). Communities, knowledge creation, and information diffusion. *J. Informetr.* **3** 180–190.

[28] LILJEROS, F., EDLING, C. R., AMARAL, L. A. N., STANLEY, H. E. and ABERG, Y. (2001). The web of human sexual contacts. *Nature* **411** 907–908.

[29] MOLLOY, M. and REED, B. (1995). A critical point for random graphs with a given degree sequence. *Random Structures Algorithms* **6** 161–180.

[30] MOORE, K. and NEELY, P. (2011). From social networks to collaboration networks: The next evolution of social media for business. *Forbes* September 15 1–3.

[31] NEWMAN, M. (2001). Scientific collaboration networks. I and II. *Phys. Rev. E* **64** 016131, 016132.

[32] NEWMAN, M. E. J. (2001). The structure of scientific collaboration networks. *Proc. Natl. Acad. Sci. USA* **98** 404–409 (electronic). MR1812610

[33] REDNER, S. (1998). How popular is your paper? An empirical study of the citation distribution. *Eur. Phys. J. B* **4** 131–134.

[34] SCHLÄPFER, M., BETTENCOURT, L. M. A., RASCHKE, M., CLAXTON, R., SMOREDA, Z., WEST, G. B. and RATTI, C. (2014). The scaling of human interactions with city size. *Journal of The Royal Society Interface* **11** 98.

[35] SCHULMAN, L. S. (1983). Ong range percolation in one dimension. *J. Phys. A* **16** L639–L641.

[36] SLIVKEN, E. (2013). Jigsaw percolation of Erdös–Rènyi random graphs. Preprint. Available at http://www.math.washington.edu/~slivken/jigsawER.pdf.

[37] SOOD, V., MATHIEU, M., SHREIM, A., GRASSBERGER, P. and PACZUSKI, M. (2010). Interacting branching process as a simple model of innovation. *Phys. Rev. Lett.* **105** 178701.

[38] UZZI, B. (2008). A social network's changing statistical properties and the quality of human innovation. *J. Phys. A* **41** 224023, 12. MR2453835

[39] UZZI, B. and SPIRO, J. (2005). Collaboration and creativity: The small world problem. *Am. J. Sociol.* **111** 447–504.

[40] VAN DEN ESKER, H., VAN DER HOFSTAD, R., HOOGHIEMSTRA, G. and ZNAMENSKI, D. (2005). Distances in random graphs with infinite mean degrees. *Extremes* **8** 111–141. MR2275914

C. D. BRUMMITT
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CALIFORNIA
ONE SHIELDS AVENUE
DAVIS, CALIFORNIA 95616
USA
E-MAIL: brummitt@gmail.com
URL: www.math.ucdavis.edu/~cbrummitt/

S. CHATTERJEE
COURANT INSTITUTE
    OF MATHEMATICAL SCIENCES
NEW YORK UNIVERSITY
251 MERCER STREET
NEW YORK, NEW YORK 10012
USA
E-MAIL: shirshendu@cims.nyu.edu
URL: www.cims.nyu.edu/~chatterj/

P. S. DEY
DEPARTMENT OF STATISTICS
UNIVERSITY OF WARWICK
GIBBET HILL ROAD
COVENTRY
CV4 7AL
UNITED KINGDOM
E-MAIL: psdey1@gmail.com
URL: www2.warwick.ac.uk/fac/sci/statistics/staff/academic-research/dey/

D. SIVAKOFF
DEPARTMENT OF STATISTICS
    AND DEPARTMENT OF MATHEMATICS
OHIO STATE UNIVERSITY
1958 NEIL AVENUE, 404 COCKINS HALL
COLUMBUS, OHIO 43210
USA
E-MAIL: dsivakoff@stat.osu.edu
URL: www.stat.osu.edu/~dsivakoff/