

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

BY CHARLES D. BRUMMITT^{§,*}, SHIRSHENDU CHATTERJEE[¶] PARTHA S. DEY^{¶,†} AND DAVID SIVAKOFF^{||,‡}

University of California, Davis[§], Courant Institute of Mathematical Sciences[¶] and Duke 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 high probability, 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 begins to show what social networks stifle creativity and what networks collectively innovate.

1. Introduction. Creativity and innovation can seem like magic [32]. But what unifies the way people innovate and solve complex problems is that they collectively combine ideas and solutions. In short, innovation and problem solving are often social endeavors. The idea of lone geniuses undergoing eureka moments is a myth; instead, most breakthroughs in science, art and business emerge from connected groups of people exchanging and merging ideas [29, 52, 15].

Consider first that successful creative works—from inventions [29] to Broadway musicals [52, 51] to software [23]—combine ideas from different disciplines. With new digital tools, creativity is even more collaborative [39]. One can collect and prioritize ideas from thousands of people [47]; iteratively improve ideas using crowds [43]; and collaboratively invent and design home products, clothing and cars with thousands of people [37, 19].

Next consider problem solving. Finding creative solutions to urban planning [2], software design [23] and open questions in science [53, 30] increasingly requires teams of people who combine diverse expertise and ideas. Again, digital tools empower large networks of people to collaboratively solve problems by combining ideas and partial solutions. Examples include addressing climate change [28], designing software [33], and collaborating using a wiki to prove mathematical conjectures [21]. With mild incentives and simple Internet tools, one can leverage social networks to collectively assemble solutions, such as finding objects in a large country [45], piecing together shredded paper [44], folding proteins [25], or collaboratively solving complex tasks [31].

*Funded by the Statistical and Applied Mathematical Sciences Institute (SAMSI), the Department of Defense (DoD) through the National Defense Science & Engineering Graduate Fellowship (NDSEG) Program, the Defense Threat Reduction Agency Basic Research Award HDTRA1-10-1-0088, and the Army Research Laboratory Cooperative Agreement W911NF-09-2-0053.

†Funded by Simons Postdoctoral Fellowship.

‡Funded in part by NSF grant DMS-10-57675 and by SAMSI.

AMS 2000 subject classifications: Primary 60K35 (percolation), 91D30 (social networks); secondary 05C80 (random graphs)

Keywords and phrases: percolation, social networks, random graph, phase transition

Companies increasingly recognize the importance of collaboration for innovation. Many companies connect their employees using internal social networks (such as IBM’s Beehive [40]) and expertise location systems (such as Tacit [12]) to match compatible ideas and expertise. When they cannot find the right expertise internally, companies outsource their most difficult R&D problems to leverage knowledge worldwide, using services like [Innocentive](#) and [Kaggle](#).

Technology is not the only reason for the increase in collaboration. The consistent rise in scientific collaboration among universities, for instance, predates the Internet and has grown stably over time [53]. As knowledge and specialization grow, so does the need to collaborate (often across disciplines) in order to advance science [53, 30] and other creative endeavors [52, 51].

Researchers have begun to study how the structure of social networks correlates empirically with their creative and team output, such as in Broadway musicals [52, 51], scientific breakthroughs [10, 20, 34] and sports teams [17]. For instance, producers of the most successful Broadway musicals have an intermediate number of triangles in their collaboration network, to balance the creative input of new collaborators with the creative stagnation (yet greater familiarity) of past collaborators [52, 51]. In general, teams solve problems better than individuals do [26, 13, 38].

However, the dynamics occurring in creative social networks remains elusive. Researchers have noted the importance yet difficulty in modeling social creativity: “...little is known about how teamwork leads to greater creativity” [17]; “...discovery involves complex social and cognitive processes and may not be predictable in detail” [7]; “...traditional epidemiological approaches need to be refined to take knowledge heterogeneity into account and preserve the system’s ability to promote creative processes of novel recombinations of ideas” [34]. Epidemic models of single ideas that spread like a slow, hard-to-catch disease in a social network [6, 8] do not model the space of ideas and how ideas recombine to form new knowledge [34]. Other models of scientific discovery and innovation include a branching process of new ideas mating with old ones [50]; ant colony models of scientists seeking papers to cite like ants seeking food; information foraging theory, to model the costs and benefits of hunting for references in bibliographic habitats; the A–B–C model of finding triadic closure among ideas; bridging structural holes (gaps between dense communities of networks) in networks of people and ideas [11].

Although some of these frameworks model the space of ideas, none also model how people collaborate as they combine ideas with one another. A better understanding of idea combination may elucidate phase transitions in the creativity of social networks. Such phase transitions may explain why some communities in history (such as ancient Greece, Renaissance Florence and Elizabethan England) [3], companies (Apple, 3M, Google) [35] and cities (Silicon Valley, Tel Aviv) [9, 35] are so innovative and why other social networks lie in sub-optimal creative states. Our main results, Theorems 1 and 2, characterize such a phase transition. We find, roughly speaking, the required number of interactions among a group of people for them to collectively solve a large puzzle.

1.1. *Motivation for the jigsaw dynamic.* Here we introduce a new kind of percolation on finite graphs that models a creative network of people merging compatible ideas into bigger and better ideas. The model is reminiscent of other models of percolation on graphs, such as bond percolation [24] and bootstrap percolation [27], but jigsaw percolation has more complex dynamics.

Consider a social network of n people with vertex set $V = \{1, 2, \dots, n\}$, 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 5).

Next we propose a natural dynamic for people to merge their compatible ideas (puzzle pieces). If two people u, w know one another 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 one another 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.

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 one another ($uw \in E_{\text{people}}$), 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.

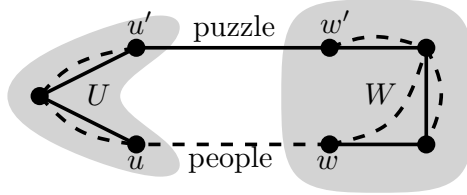


Fig 1: Illustration of the jigsaw dynamic. Dashed and solid edges denote the people graph and puzzle graph, respectively. Jigsaw clusters U and W contain three and four nodes each. Nodes u, w know one another but do not have compatible puzzle pieces. However, they have merged their puzzle pieces with nodes u', w' , who do have compatible puzzle pieces. Thus U and W merge.

1.2. *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 collection \mathcal{C}_i of disjoint subsets of V , where the elements of \mathcal{C}_i , the 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) After the first step, \mathcal{C}_1 is the set of connected components in the graph $(V, E_{\text{people}} \cap E_{\text{puzzle}})$.
- 3) After step $i \geq 1$, we have a collection of jigsaw clusters \mathcal{C}_i that partition the set of vertices V . At step $(i+1)$, we merge every pair of jigsaw clusters in \mathcal{C}_i that are both puzzle- and people-adjacent (see Figure 2).

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 $(\mathcal{C}_i, \mathcal{E}_i)$:

$$\mathcal{C}_{i+1} = \{\cup_{U \in A} U : A \text{ is a connected component of } (\mathcal{C}_i, \mathcal{E}_i)\}. \quad (1.1)$$

Given $(V, E_{\text{people}}, E_{\text{puzzle}})$, we merge jigsaw clusters until no more merges can be made, i.e., iterate Eq. (1.1) to a fixed point \mathcal{C}_∞ . 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.

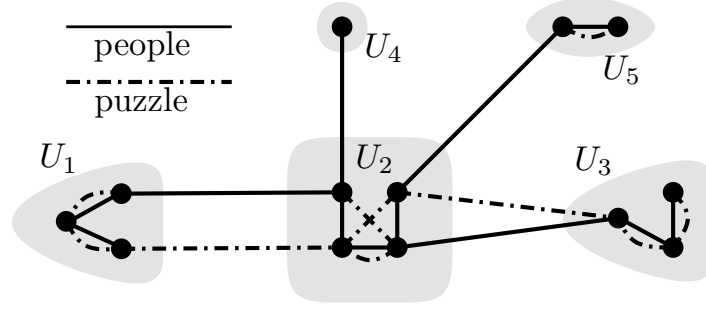


Fig 2: Jigsaw clusters $U_1, U_2, U_3, U_4, U_5 \in \mathcal{C}_i$ at stage i . At stage $i+1$, jigsaw clusters U_1, U_2, U_3 merge.

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.

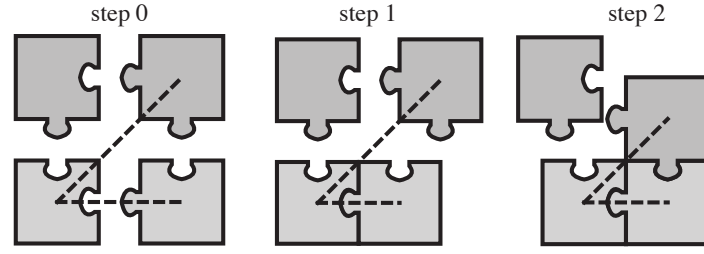


Fig 3: A complete trajectory of the jigsaw dynamics. The people graph (dashed edges) does not solve this 2×2 puzzle.

1.3. *Statement of results.* In 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} . For a fixed, connected puzzle graph of size n , we are interested in the probability of the event

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

We denote this probability by $\mathbb{P}(\text{Solve})$ or by $\mathbb{P}_{p_n}(\text{Solve})$ 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 , $\mathbb{P}_p(\text{Solve})$ is nondecreasing with p . Trivially, $\mathbb{P}_0(\text{Solve}) = 0$ and $\mathbb{P}_1(\text{Solve}) = 1$. Furthermore, $\mathbb{P}_p(\text{Solve})$ 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 $\mathbb{P}_p(\text{Solve}) = 1/2$, 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 $\mathbb{P}_{p_n}(\text{Solve}) = 1/2$.*

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 [18]). Note that the equality $p_c(n) = t_n$ holds when the puzzle graph is the star graph $(\{1, 2, \dots, n\}, \{(i, n) : 1 \leq i < n\})$, because in this case the puzzle can be solved iff the people graph is connected.*

We use the following standard notation for describing sequences of non-negative 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 $a_n/b_n \rightarrow 0$ as $n \rightarrow \infty$; 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 $p_n = \lambda / \log n$, $\mathbb{P}_{p_n}(\text{Solve}) \rightarrow 0$ or 1 according as $\lambda < 1/27$ or $\lambda > \pi^2/6$.

REMARK 3. *We believe that our upper bound is tight (see Section 5). 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 $\mathbb{P}_{p_n}(\text{Solve}) \rightarrow 0$ for $p_n = O(1/n^b)$ for any $b > 0$, and $\mathbb{P}_{p_n}(\text{Solve}) \rightarrow 1$ 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 $n \rightarrow \infty$ (see Proposition 1). 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 (respectively) solving an Erdős-Rényi puzzle.*

As a model of social networks, the Erdős-Rényi random graph assumes no structure other than the average number of connections per person. However, in many social networks—from scientific citations [46] to scientific collaborations [4, 41, 42] to sexual partners [36]—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 a random vertex has degree k with probability proportional to $k^{-\gamma}$ for some $\gamma > 2$.

Surprisingly, such heterogeneous social networks cannot solve a large class of puzzles. We write the following corollary to Theorem 2 for a random power-law degree people graph, but it applies to any random people graph with a degree distribution that has finite $(1 + \epsilon)^{\text{th}}$ moment for some $\epsilon > 0$ and that, given a degree sequence, is uniform over all simple graphs with that sequence.

COROLLARY 1. *For a random power-law degree people graph with exponent $\gamma > 2$ and any connected, bounded-degree puzzle graph, the probability of solving the puzzle tends to 0 as $n \rightarrow \infty$.*

REMARK 6. *Because collaboration networks in science [4, 41, 42] 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, merit future work (see Section 5 for details).*

PROOF OF COROLLARY 1. Let d_i be the degree of vertex i in the people graph. If $f(n) \rightarrow \infty$ as $n \rightarrow \infty$, then

$$\mathbb{P} \left(\max_{1 \leq i \leq n} d_i \geq f(n)n^{1/(\gamma-1)} \right) \leq \frac{\mathbb{E} d_1^{\gamma-1}}{f(n)^{\gamma-1}} \rightarrow 0,$$

because $\mathbb{E} d_1^{\gamma-1}$ is finite. Choose $0 < b < (\gamma - 2)/(\gamma - 1)$. By Theorem 2, if the people graph is Erdős-Rényi with $p = n^{-b}$, then the probability of solving a bounded-degree puzzle tends to 0 as $n \rightarrow \infty$. Since the minimum degree of this Erdős-Rényi people graph is bounded below by $n^{1-b}(1-o(1)) = \omega(n^{1/(\gamma-1)})$ with high probability, the Erdős-Rényi people graph stochastically dominates the power-law people graph in terms of edge density. By monotonicity of the event **Solve**, this implies that the power-law people graph cannot solve the puzzle with high probability. \square

Some of the techniques in our proofs resemble those used for long range percolation and 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}_∞ , 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 to show that no infinite component exists in one-dimensional long range percolation [48, 14].

In our proof of the upper bound on $p_c(n)$, we use the fact that once a sufficiently large, solved cluster emerges, then that cluster will inevitably continue to merge and ultimately solve the puzzle. As in bootstrap percolation on the lattice graph [1, 27], our upper bound arises from a sufficient condition for the formation of a large cluster.

Before proceeding, we note similarities between this model and several real world phenomena. First, it vaguely captures how people solve jigsaw puzzles: Dump the puzzle pieces on the table so that they land randomly, look for local connections, and then iteratively merge nearby clumps of merged puzzle pieces. Jigsaw percolation also resembles more consequential problems in several scientific fields. In protein assembly, peptides merge into meta-proteins, which further merge to form proteins [54]. Whether certain molecules merge together or tile the plane underlies the computations of many DNA self-assembly systems [49]. Finally, in creative social networks ranging from companies to communities of scientists, people combine their compatible ideas into bigger, better ideas, in a manner vaguely similar to jigsaw percolation. The phase transition in the probability that a random graph solves a jigsaw puzzle begins to inform what properties of these social networks facilitate their ability to collaboratively solve problems and to innovate.

1.4. *Road map for the paper.* In Section 2, we prove the upper bound on the critical value $p_c(n)$ for both Theorems 1 and 2. In Section 3, we prove the lower bound for the ring puzzle in Theorem 1, and in Section 4 we prove the lower bound for arbitrary puzzles with bounded maximum degree. In Section 5, we discuss simulations and open questions.

2. 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 1 (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*

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

REMARK 7. *A close look at the proof of Proposition 1 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

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

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 $\{1, 2, \dots, j-1\}$ for all $2 \leq j \leq n$ (as illustrated in Figure 4). Then the people graph solves the puzzle.

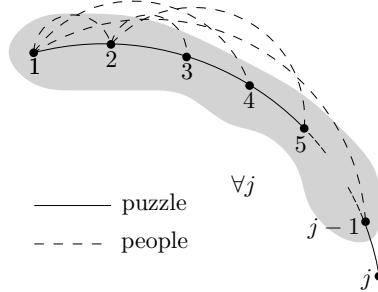


Fig 4: Illustration of the sufficient condition to solve the ring puzzle: j is people-adjacent to $\{1, 2, \dots, j-1\}$ for all $j = 2, 3, \dots, n$. This event is contained in the event **Solve**.

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 .

LEMMA 1. Consider an Erdős-Rényi people graph and any connected puzzle graph on n vertices. Then for any fixed $\delta > 0$,

$$\begin{aligned} \mathbb{P}_{p_n}(\text{Solve} \mid \exists \text{ an internally solved subset of size } \geq (1 + \delta)\log n/\epsilon_n) \\ \geq 1 - n^{-\delta}, \end{aligned}$$

where $\epsilon_n := -\log(1 - p_n)$.

PROOF. Suppose that there exists a *internally solved* subset U of size $m \geq (1 + \delta)\log n/\epsilon_n$. The probability that all the remaining $n - m$ vertices in $V \setminus U$ are connected to U by a people edge is

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

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 implies **Solve**. The proof is now done, because the event that a particular set of vertices forms an *internally solved* subset depends only on the edges among those vertices. \square

We use the following lemma to partition the puzzle graph into disjoint blocks. The motivation comes from analyzing the ring puzzle graph.

LEMMA 2. 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 $\mathcal{B}_1, \mathcal{B}_2, \dots, \mathcal{B}_k$ of V such that

- i) $V = \cup_{i=1}^k \mathcal{B}_i$;
- ii) $|\mathcal{B}_i| \in [m, 2m]$ for $i = 1, 2, \dots, k-1$ and $|\mathcal{B}_k| < 2m$;
- iii) the induced subgraph on \mathcal{B}_i is connected for all $i = 1, 2, \dots, k$;
- 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, \dots, 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 $|C_1| + |C_2| + \dots + |C_{l-1}| < m$ and $|C_1| + |C_2| + \dots + |C_l| \geq m$. Such an l exists, because $|C_1| + |C_2| + \dots + |C_k| = n - 1 > m$. Necessarily we have $|C_1| + |C_2| + \dots + |C_l| < 2m$, because $|C_i| < m$ for all i . We take $\mathcal{B}_1 := \cup_{i=1}^l C_i \cup \{v_0\}$ and use induction on the graph G with vertex set $\cup_{i=1}^l C_i$ 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. \square

PROOF OF PROPOSITION 1. Using Lemma 2, we partition the puzzle graph into blocks $\mathcal{B}_1, \mathcal{B}_2, \dots, \mathcal{B}_k$ 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 $p = \lambda/\log n$ with $\lambda > \pi^2/6$, then

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

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 $1, 2, \dots, |\mathcal{B}_i|$ in such a way that the vertex with label j is puzzle-adjacent to the set of vertices with labels $\{1, 2, \dots, j-1\}$ 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

$$\begin{aligned} \overline{B}_i := & \{\text{for all } 1 \leq j \leq |\mathcal{B}_i|, \text{ the vertex labeled } j \text{ is people-adjacent} \\ & \text{to the set of vertices labeled } \{1, 2, \dots, j-1\} \subset \mathcal{B}_i. \end{aligned}$$

(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 $\{1, \dots, j-1\}$. Thus \overline{B}_1 is not a necessary condition for B_1 to occur, i.e., $\overline{B}_1 \subsetneq B_1$.) The events that $j+1$ is people-adjacent to $\{1, 2, \dots, j\}$ occur independently with probability $\geq 1 - (1 - p_n)^j$, so

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

Thus the random variable S stochastically dominates

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

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

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

The proof of Lemma 3 follows the present proof.

LEMMA 3. *Let $\theta(x) := -\int_0^x \log(1 - e^{-t}) dt$ for $x \in [0, \infty]$. If $\lim_{\epsilon \rightarrow 0} m_\epsilon \epsilon = x \in [0, \infty]$, then*

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

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

$$\left| \sum_{i=1}^m \log(1 - e^{-i\epsilon}) + \frac{\pi^2}{6\epsilon} \right| \leq \frac{1}{2} \log \frac{2e^2}{\epsilon} + \frac{\pi^2}{6\epsilon e^{m\epsilon}}. \quad (2.1)$$

Fix $\delta \in (0, n/\lceil \log n/\epsilon_n \rceil - 1)$, and let $m = \lceil (1+\delta)(\log n)/\epsilon_n \rceil$. Using Lemma 3, we estimate

$$\begin{aligned} & \log \mathbb{E} S' \\ &= \log \left(\frac{n}{2m} - 1 \right) - \frac{\pi^2}{6\epsilon} + \left(\sum_{j=1}^{2m} \log(1 - e^{-j\epsilon}) + \frac{\pi^2}{6\epsilon} \right) \\ &\geq \left(1 - \frac{\pi^2}{6\lambda} \right) \log n - \log \frac{2m}{1 - 2m/n} - \frac{1}{2} \log \frac{2e^2}{\epsilon} - \frac{\pi^2}{6\epsilon e^{2m\epsilon}} \\ &\geq \left(1 - \frac{\pi^2}{6\lambda} \right) \log n - \frac{m}{n-2m} - \frac{5}{2} \log \log n - \frac{1}{2} \log \frac{8e^2(1+\delta)^2}{\lambda^3} - \frac{\pi^2 \log n}{6\lambda n^{2+2\delta}} \\ &\rightarrow \infty \text{ as } n \rightarrow \infty. \end{aligned}$$

Since S' is binomial, $\mathbb{E} S' \rightarrow \infty$ implies that $\mathbb{P}(S' \geq 1) \rightarrow 1$. Thus we have $\mathbb{P}(S \geq 1) \geq \mathbb{P}(S' \geq 1) \rightarrow 1$. On the event $\{S \geq 1\}$, the whole ring is solved with probability tending to one by Lemma 1 and the fact that $m \geq (1+\delta)(\log n)/\epsilon_n$. The proof is complete. \square

PROOF OF LEMMA 3. Note that

$$\begin{aligned} -\epsilon \sum_{i=1}^k \log(1 - e^{-i\epsilon}) &= \epsilon \sum_{i=1}^k \sum_{j=1}^{\infty} \frac{e^{-ij\epsilon}}{j} = \epsilon \sum_{j=1}^{\infty} \frac{1 - e^{-jk\epsilon}}{j(e^{j\epsilon} - 1)} \\ &= \sum_{j=1}^{\infty} \frac{1 - e^{-jk\epsilon}}{j^2} - \sum_{j=1}^{\infty} \frac{(1 - e^{-jk\epsilon})(e^{j\epsilon} - 1 - j\epsilon)}{j^2(e^{j\epsilon} - 1)}. \end{aligned}$$

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

$$\begin{aligned} & \sum_{j=1}^{\infty} \frac{(1 - e^{-jk\epsilon})(e^{j\epsilon} - 1 - j\epsilon)}{j^2(e^{j\epsilon} - 1)} \leq \sum_{j=1}^{\infty} \frac{\min\{j\epsilon/2, 1\}}{j^2} \\ & \leq \sum_{j \leq m} \frac{\epsilon}{2j} + \sum_{j > m} \frac{1}{j^2} \leq \frac{\epsilon}{2} (\log m + 1) + \frac{1}{m} = \frac{\epsilon}{2} \log \frac{2e^2}{\epsilon} \end{aligned}$$

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

$$\left| \sum_{i=1}^k \epsilon \log(1 - e^{-i\epsilon}) + \sum_{j=1}^{\infty} \frac{1 - e^{-jk\epsilon}}{j^2} \right| \leq \frac{\epsilon}{2} \log \frac{2e^2}{\epsilon}. \quad (2.2)$$

In particular, if $\lim_{\epsilon \rightarrow 0} k\epsilon = x \in [0, \infty]$, then interchanging the sum and the integral

$$\begin{aligned} \lim_{\epsilon \rightarrow 0} \epsilon \sum_{i=1}^{k\epsilon} \log(1 - e^{-i\epsilon}) &= - \sum_{j=1}^{\infty} \frac{1 - e^{-jx}}{j^2} \\ &= - \sum_{j=1}^{\infty} \frac{1}{j} \int_0^x e^{-jt} dt = \int_0^x \log(1 - e^{-t}) dt, \end{aligned}$$

which completes the proof. The bound (2.1) follows from (2.2) and the fact that $e^{-jk\epsilon} \leq e^{-k\epsilon}$ for all $j \geq 1$. \square

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

PROOF OF PROPOSITION 2. Let x be a fixed 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 iff $u - v = \pm 1$, where all additions and subtractions in \mathbb{Z}_n are modulo n . We denote the interval $\{a, a+1, \dots, b\} \subseteq \mathbb{Z}_n$ by $[a, b]$ and its length by $|[a, b]|$.

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 4. *Suppose that there exist integers $0 = a_0 < a_1 < \dots < a_k = n$ such that, for all $j = 0, 1, \dots, k-1$, 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 $j = 0, 1, \dots, k-1$. Clearly $1 \leq v_0 < v_1 < \dots < v_{k-1} \leq n$. Furthermore, each v_j has no people edges with $[v_{j-1}, v_{j+1}]$ (where $j + \ell$ is taken modulo k), because $|I_j| \leq x$.

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. \square

In light of Lemma 4, 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

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

and $a_i := l_1 + l_2 + \dots + l_i$ for $i = 0, 1, \dots, k$. 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,

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

It suffices to show that $\mathbb{P}(Z > 0) \rightarrow 0$ as $n \rightarrow \infty$ for appropriate choice of x . We will use Lemma 5 to estimate the probability that an interval is not x -good.

LEMMA 5. *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*

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

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

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

In particular, if $p = p_n = \lambda / \log n$ and $x = t \log n / \lambda$ for some $t < 1/3$, we have

$$\begin{aligned} \mathbb{P}(Z > 0) & \leq \exp \left[\log n - \frac{t \log n}{2\lambda} (2 \log(\sqrt{1 + 1/t} - 1) + 7t - 2t\sqrt{1 + 1/t} - 1) \right] \\ & \rightarrow 0 \text{ as } n \rightarrow \infty \end{aligned}$$

when

$$\lambda < \frac{t}{2} \left[2 \log(\sqrt{1 + 1/t} - 1) + 7t - 2t\sqrt{1 + 1/t} - 1 \right]. \quad (3.1)$$

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

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

Thus given $\lambda \leq 1/27$, we can choose $t \in (0, 1/3)$ such that (3.1) holds, and taking $x = t \log n / \lambda$ we have

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

This completes the proof. \square

PROOF OF LEMMA 5. 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

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

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

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

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

$$\begin{aligned} \mathbb{P}(X + 2Y' \geq lx) &\leq e^{-\theta lx} (1 - p + pe^\theta)^{2lx^2} (1 - p + pe^{2\theta})^{l^2 x^2 / 2} \\ &\leq \exp[-lx(\theta - 2t(e^\theta - 1) - lt(e^{2\theta} - 1)/2)], \end{aligned} \tag{3.2}$$

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 (3.2), we finally have

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

This completes the proof. \square

Propositions 1 and 2 give Theorem 1.

4. 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 $n \rightarrow \infty$.

PROPOSITION 3. *For any sequence of connected puzzle graphs with bounded maximum degree as $|V| = n \rightarrow \infty$, $p_c(n) = \omega(1/n^b)$ 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 $U = U_0, U_1, \dots, U_\ell = U' \in \mathcal{C}_i$ 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 $\mathcal{A}_i \subset \mathcal{C}_i$ 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 $i = 0, \dots, k$ as

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

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 $\mathbb{P}(E_k) \rightarrow 0$ and $\mathbb{P}(F_k) \rightarrow 0$ as $n \rightarrow \infty$, 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 $\mathbb{P}(E_i) \rightarrow 0$ and $\mathbb{P}(F_i) \rightarrow 0$ as $n \rightarrow \infty$ for some $i \in \{0, 1, \dots, k-1\}$, which implies that $\mathbb{P}(E_i^c \cap F_i^c) \rightarrow 1$. 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

$$1 - (1 - n^{-b})^{(DL_i)^2} \leq 1 - e^{-2(DL_i)^2 n^{-b}} \leq 2(DL_i)^2 n^{-b}, \quad (4.1)$$

where we assume that n is large so that $n^{-b} < \frac{1}{2}$, and we use the fact that $e^{-2x} \leq 1 - x$ for $x \in (0, 1/2)$. For convenience, we now order the clusters in \mathcal{C}_i so that $A_1, A_2, \dots, A_{|\mathcal{A}_i|} \in \mathcal{A}_i$ and $A_{|\mathcal{A}_i|+1}, A_{|\mathcal{A}_i|+2}, \dots, A_{|\mathcal{C}_i|} \in \mathcal{C}_i \setminus \mathcal{A}_i$. 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 $X_i \sim \text{Binomial}(DL_i C_i n^{1-ib}, 2(DL_i)^2 n^{-b})$. 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 puzzle graph. If we let $C_{i+1} = 4(DL_i)^3 C_i$ (this is $2\mathbb{E} X_i / n^{1-(i+1)b}$), then by Chebychev's inequality

$$\begin{aligned} \mathbb{P}(E_{i+1} | E_i^c \cap F_i^c) &= \mathbb{P}\left(\sum_{j=1}^{|\mathcal{A}_i|} \sum_{\ell=j+1}^{|\mathcal{C}_i|} I_{\{A_j, A_\ell\}}^{i+1} \geq C_{i+1} n^{1-(i+1)b} \middle| E_i^c \cap F_i^c\right) \\ &\leq \mathbb{P}\left(X_i \geq C_{i+1} n^{1-(i+1)b}\right) \\ &= \mathbb{P}(X_i - \mathbb{E} X_i \geq \mathbb{E} X_i) \\ &\leq (\mathbb{E} X_i)^{-1} = O(n^{-1+(i+1)b}) \rightarrow 0. \end{aligned}$$

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

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 $U = U_0, U_1, \dots, U_\ell = U' \in \mathcal{C}_i$ such that U_j and U_{j+1} are puzzle-adjacent for all $j \in \{0, \dots, \ell-1\}$. For a fixed cluster $A \in \mathcal{C}_i$, let Y_A^i denote the number of paths of length k that start at A (so $U_0 = A$) and such that U_j will merge with

U_{j+1} at stage $i+1$ for each $j \in \{0, \dots, k-1\}$. For any cluster path U_0, \dots, U_k , the probability that U_j and U_{j+1} merge at stage $i+1$ is bounded above by $2(DL_i)^2 n^{-b}$ on the event F_i^c , by inequality (4.1). The number of cluster paths of length k in \mathcal{C}_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,

$$\begin{aligned} \mathbb{P} \left(\sum_{A \in \mathcal{C}_i} Y_A^i \geq 1 \mid F_i^c \right) &\leq n \mathbb{P} (Y_A^i \geq 1 \mid F_i^c) \\ &\leq n \left[(DL_i)^k (2(DL_i)^2 n^{-b})^k \right] = O(n^{1-kb}) \rightarrow 0. \end{aligned}$$

This implies that there are no cluster paths of length k or longer that merge at stage $i+1$. In turn, this implies that the largest cluster after stage $i+1$ is smaller than $L_{i+1} := L_i^k$ with high probability on the event F_i^c , so $\mathbb{P}(F_{i+1}) \rightarrow 0$, which completes the proof. \square

Propositions 1 and 3 give Theorem 2.

5. 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 5a). However, as with bootstrap percolation [22], 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*

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

If true, this means that estimating c_1 to within 1% via simulation would require taking n to be at least $\exp[(100c_2/c_1)^{1/b}]$, 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.

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 1 (i.e., the one-sided requirement that j is connected to $\{1, 2, \dots, j-1\}$ 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 5b). 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 Lemma 1 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 2 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 5b is an interesting open problem.

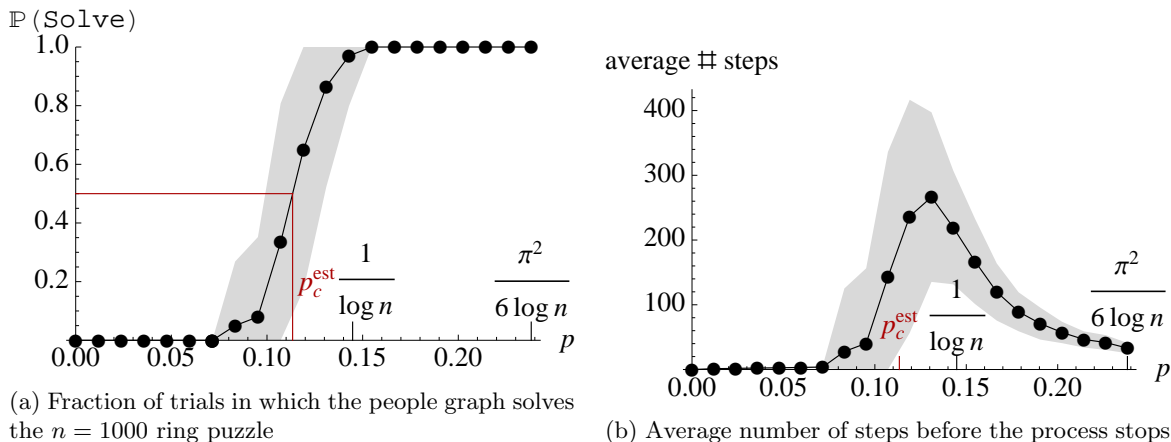


Fig 5: Simulations of jigsaw percolation on a ring of size $n = 1000$, with 200 trials for 21 equally spaced values of $p \in [0, 1.05 \times \pi^2 / (6 \log n)]$ (which took 57 days on a department server). Dots are averages of 200 trials, while shaded gray areas denote ± 1 standard deviation. The estimated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

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

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

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 $p_\epsilon(n)$ as the unique p for which $\mathbb{P}_p(\text{Solve}) = \epsilon$. Then

$$p_\epsilon(n) / p_{1-\epsilon}(n) \rightarrow 1$$

as $n \rightarrow \infty$ for any $\epsilon \in (0, 1)$ 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 [16], 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 -cycles 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 and hence innovate more [5, 9]. Furthermore, most social networks wish to minimize communication overhead; the critical value $p_c(n)$ indicates the minimal communication needed to coallaboratively solve large puzzles.

Surprisingly, social networks with power law degree distributions lack the connectivity needed to solve bounded-degree puzzles (Corollary 1). However, scientific collaboration networks manage to solve puzzles despite their heavy-tailed degree distributions [4, 41, 42]. This highlights the importance of considering more realistic assumptions in the model and of drawing from (still nascent) studies on knowledge spaces [11].

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.

Acknowledgements. We thank Rick Durrett, M. Puck Rombach, Peter J. Mucha, Raissa M. D'Souza, Alex Waagen, Pierre-André Noël and Madeleine Däpp for useful discussions.

References.

- [1] M. Aizenman and J. L. Lebowitz, *Metastability effects in bootstrap percolation*, Journal of Physics A: Mathematical and General **21** (1988), 3801.
- [2] Ernesto Arias, Hal Eden, Gerhard Fischer, Andrew Gorman, and Eric Scharff, *Transcending the individual human mind—creating shared understanding through collaborative design*, ACM Trans. Comput.-Hum. Interact. **7** (2000), no. 1, 84–113.
- [3] David Banks, *The problem of excess genius*, Newsletter of the Classification Society of North America **48** (1997), 1–6, <http://www.monad.com/sdg/Journal/genius.html>.
- [4] A. L. Barabási, H. Jeong, Z. Néda, E. Ravasz, A. Schubert, and T. Vicsek, *Evolution of the social network of scientific collaborations*, Physica A **311** (2002), 590–614.
- [5] L. Bettencourt, J. Lobo, D. Helbing, C. Kühnert, and G. B. West, *Growth, innovation, scaling, and the pace of life in cities*, Proc. Natl. Acad. Sci. U.S.A. **104** (2007), no. 17, 7301.
- [6] Luís M A Bettencourt, Ariel Cintrón-Arias, David I Kaiser, and Carlos Castillo-Chávez, *The power of a good idea: Quantitative modeling of the spread of ideas from epidemiological models*, Physica A **364** (2006), 513–536.
- [7] Luís M A Bettencourt, David I Kaiser, and Jasleen Kaur, *Scientific discovery and topological transitions in collaboration networks*, Journal of Informetrics **3** (2009), no. 3, 210–221.
- [8] Luís M A Bettencourt, David I Kaiser, Jasleen Kaur, Carlos Castillo-Chávez, and David E Wojick, *Population modeling of the emergence and development of scientific fields*, Scientometrics **75** (2008), no. 3, 495–518.
- [9] Luís M A Bettencourt, José Lobo, Deborah Strumsky, and Geoffrey B West, *Urban Scaling and Its Deviations: Revealing the Structure of Wealth, Innovation and Crime across Cities*, PloS one **5** (2010), no. 11, e13541.
- [10] Sen Chai and Lee Fleming, *Emergence of Breakthroughs*, DIME-DRUID ACADEMY Winter Conference (2011), 1–47.

- [11] C. Chen, Y. Chen, M. Horowitz, H. Hou, Z. Liu, and D. Pellegrino, *Towards an explanatory and computational theory of scientific discovery*, Journal of Informetrics **3** (2009), no. 3, 191–209.
- [12] CIOinsight, *Web extra: Who's on first?*, CIOinsight (2004), 1–2, <http://www.cioinsight.com/c/a/Past-News/Web-Extra-Whos-on-First/>.
- [13] S. H. Clearwater, B. A. Huberman, and T. Hogg, *Cooperative solution of constraint satisfaction problems*, Science **254** (1991), no. 5035, 1181–1183.
- [14] Don Coppersmith, David Gamarnik, and Maxim Sviridenko, *The diameter of a long-range percolation graph*, Random Structures & Algorithms **21** (2002), no. 1, 1–13.
- [15] Mihaly Csikszentmihalyi, *Flow and the psychology of discovery and invention*, Harper Perennial, New York, NY, May 1997.
- [16] Raissa M D'Souza and Michael Mitzenmacher, *Local Cluster Aggregation Models of Explosive Percolation*, Physical Review Letters **104** (2010), no. 19, 195702.
- [17] J. Duch, J. S. Waitzman, and L. A. N. Amaral, *Quantifying the performance of individual players in a team activity*, PLoS one **5** (2010), no. 6, e10937.
- [18] P. Erdős and A. Rényi, *On the strength of connectedness of a random graph*, Acta Mathematica Hungarica **12** (1961), no. 1, 261–267.
- [19] Christopher M Ford, *Twitter, Facebook, and Ten Red Balloons: Social Network Problem Solving and Homeland Security*, Homeland Security Affairs **7** (2011), 1–7.
- [20] M Gerstein and S. M. Douglas, *RNAi development*, PLoS Computational Biology **3** (2007), no. 4, e80.
- [21] Timothy Gowers and Michael Nielsen, *Massively collaborative mathematics*, Nature **461** (2009), 879–881.
- [22] Janko Gravner and Alexander E Holroyd, *Slow convergence in bootstrap percolation*, The Annals of Applied Probability **18(3)** (2008), 909–928.
- [23] Joan Greenbaum and Morten Kyng (eds.), *Design at Work: Cooperative Design of Computer Systems*, Lawrence Erlbaum Associates, Inc., Hillsdale, NJ, 1991.
- [24] Geoffrey R. Grimmett, *Percolation*, second ed., Springer-Verlag, New York, NY, May 1999.
- [25] Eric Hand, *People power*, Nature **466** (2010), 685–687.
- [26] G. W. Hill, *Group versus individual performance: Are N+1 heads better than one?*, Psychological Bulletin **91** (1982), no. 3, 517–539.
- [27] Alexander E. Holroyd, *Sharp metastability threshold for two-dimensional bootstrap percolation*, Probability Theory and Related Fields **125** (2003), no. 2, 195–224.
- [28] J. Introne, R. Laubacher, G. Olson, and T. Malone, *The Climate CoLab: Large scale model-based collaborative planning*, 2011 International Conference on Collaboration Technologies and Systems (CTS), May 2011, pp. 40–47.
- [29] Steven Johnson, *Where good ideas come from: The natural history of innovation*, Riverhead Hardcover, October 2010.
- [30] B. F. Jones, S. Wuchty, and B. Uzzi, *Multi-university research teams: Shifting impact, geography, and stratification in science*, Science **322** (2008), no. 5905, 1259–1262.
- [31] Aniket Kittur, Boris Smus, Susheel Khamkar, and Robert E. Kraut, *Crowdforge: crowdsourcing complex work*, Proceedings of the 24th annual ACM symposium on User interface software and technology (New York, NY, USA), UIST '11, ACM, 2011, pp. 43–52.
- [32] D. Klahr and H. A. Simon, *Studies of scientific discovery: Complementary approaches and convergent findings.*, Psychological Bulletin **125** (1999), no. 5, 524.
- [33] Karim R. Lakhani, David A. Garvin, and Eric Lonstein, *TopCoder (A): Developing Software through Crowdsourcing*, Harvard Business School General Management Unit **610-032** (2010), 1–18.
- [34] R Lambiotte and P Panzarasa, *Communities, knowledge creation, and information diffusion*, Journal of Informetrics **3** (2009), no. 3, 180–190.
- [35] Jonah Lehrer, *Imagine: How creativity works*, Houghton Mifflin Harcourt, New York, NY, March 2012.
- [36] F. Liljeros, C. R. Edling, L. A. N. Amaral, H. E. Stanley, and Y. Aberg, *The web of human sexual contacts*, Nature **411** (2001), 907–908.
- [37] Mary L. Maher, *Design creativity research: From the individual to the crowd*, Design creativity (2010), 1–6.
- [38] Winter Mason and Duncan J. Watts, *Collaborative learning in networks*, Proc. Natl. Acad. Sci. U.S.A. (2011), 1–6.
- [39] William J. Mitchell, Alan S. Inouye, and Marjory S. Blumenthal (eds.), *Beyond productivity: Information, technology, innovation, and creativity*, The National Academies Press, 2003.
- [40] Karl Moore and Peter Neely, *From social networks to collaboration networks: The next evolution of social media for business*, Forbes (2011), 1–3.
- [41] Mark Newman, *Scientific collaboration networks. I and II.*, Phys. Rev. E **64** (2001), papers 016131, 016132.
- [42] ———, *The structure of scientific collaboration networks*, Proc. Natl. Acad. Sci. USA **98** (2001), 404–409.
- [43] J. V. Nickerson, Y. Sakamoto, and L. Yu, *Structures for creativity: The crowdsourcing of design*, CHI Workshop on Crowdsourcing and Human Computation (2011), 1–4.
- [44] Chris Palmer, *UCSD team's effort in DARPA's shredder challenge derailed by sabotage*, California Institute for Telecommunications and Information Technology (2011), 1–3.
- [45] G Pickard, W Pan, I Rahwan, M Cebrian, R Crane, A Madan, and A Pentland, *Time-Critical Social Mobilization*, Science **334** (2011), no. 509, 509–512.
- [46] Sidney Redner, *How popular is your paper? An empirical study of the citation distribution*, Eur. Phys. J. B **4** (1998), 131–134.
- [47] M. J. Salganik and K. E. C. Levy, *Wiki surveys: Open and quantifiable social data collection*, ArXiv e-prints (2012),

- 1–29.
- [48] L. S. Schulman, *Long range percolation in one dimension*, J. Phys. A. **16** (1983), L639–L641.
 - [49] Nadrian Seeman, *Structural DNA nanotechnology: growing along with Nano Letters*, Nano Letters **10** (2010), no. 6, 1971–1978.
 - [50] Vishal Sood, Mylène Mathieu, Amer Shreim, Peter Grassberger, and Maya Paczuski, *Interacting Branching Process as a Simple Model of Innovation*, Physical Review Letters **105** (2010), no. 17, 178701.
 - [51] B. Uzzi, *A social network’s changing statistical properties and the quality of human innovation*, Journal of Physics A: Mathematical and Theoretical **41** (2008), 224023.
 - [52] B. Uzzi and J. Spiro, *Collaboration and Creativity: The Small World Problem*, American Journal of Sociology **111** (2005), no. 2, 447–504.
 - [53] S. Wuchty, B. F. Jones, and B. Uzzi, *The increasing dominance of teams in production of knowledge*, Science **316** (2007), no. 5827, 1036–1039.
 - [54] Bing Zhang, Matthew C. Chambers, and David L. Tabb, *Proteomic parsimony through bipartite graph analysis improves accuracy and transparency*, Journal of Proteome Research **6** (2007), no. 9, 3549–3557.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CALIFORNIA
ONE SHIELDS AVENUE
DAVIS, CA 95616, USA
E-MAIL: cbrummitt@math.ucdavis.edu
URL: www.math.ucdavis.edu/~cbrummitt/

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

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

DEPARTMENT OF MATHEMATICS
DUKE UNIVERSITY, Box 90320
DURHAM, NC 27708, USA
E-MAIL: djsivy@math.duke.edu
URL: <http://math.duke.edu/~djsivy/>