Jigsaw percolation

Joint work with David Sivakoff

Univerza na Primorskem

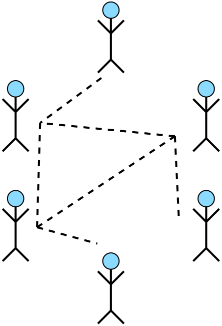
May 29, 2017

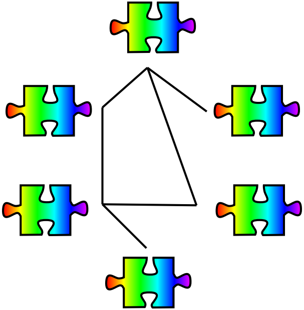
Motivation

- 2011 DARPA shredder challenge: UCSD team used crowdsourcing to piece together shredded paper.
- Polymath Project: Timothy Gowers' experiment with "massively collaborative mathematics."
- How might people *cooperatively* combine their individual ideas to solve a problem?
- Coagulation model to form coalitions: discrete Smoluchowski dynamics with additional restrictions, say kin relations or spatial proximity.

People and ideas (puzzle pieces)

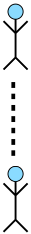
A dynamic on two graphs with the same vertex set but different edges, introduced in a 2015 paper by Brummitt, Chatterjee, Dey, and Sivakoff, henceforth referred to as BCDS.





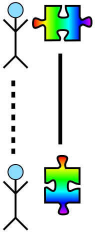
People and ideas (puzzle pieces)

- If two people know each other...



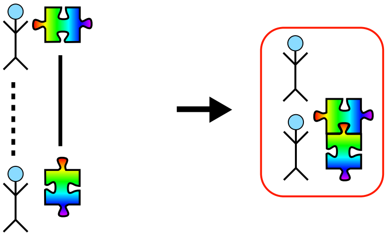
People and ideas (puzzle pieces)

- If two people know each other...
- and have compatible ideas...



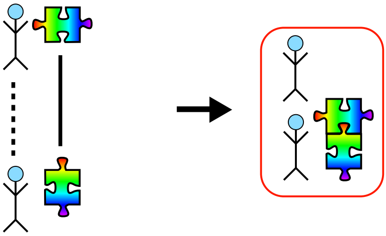
People and ideas (puzzle pieces)

- If two people know each other...
- and have compatible ideas...
- then they merge their ideas.

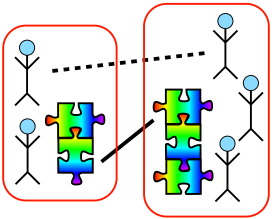


People and ideas (puzzle pieces)

- If two people know each other...
- and have compatible ideas...
- then they merge their ideas.

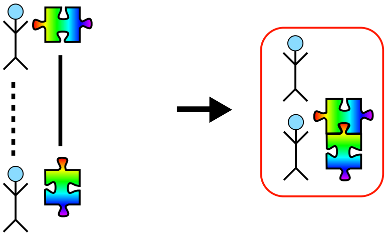


- Generally, if two groups with merged ideas know each other and have compatible ideas...

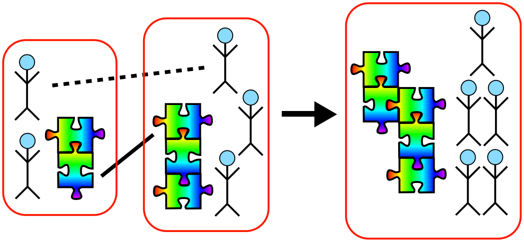


People and ideas (puzzle pieces)

- If two people know each other...
- and have compatible ideas...
- then they merge their ideas.

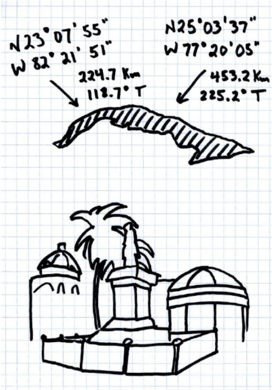


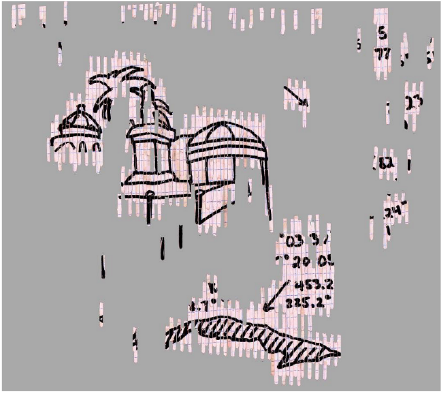
- Generally, if two groups with merged ideas know each other and have compatible ideas...
- then they merge their ideas.



People and ideas (puzzle pieces)

This generates larger and larger partial solutions, as in this intermediate step from the DARPA challenge.



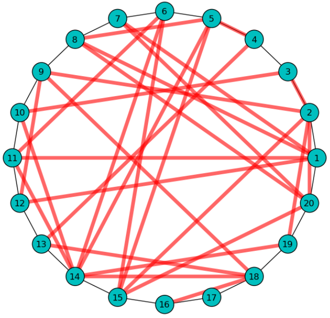


Jigsaw percolation model

The vertex set V has N vertices.

- People Graph: Erdős-Rényi random graph $(V, E_{\text{ppl}}) \sim G(N, p)$.
- Puzzle Graphs: Connected deterministic graphs (V, E_{puz}).

Example: $N = 20$, $p = 0.15$, and (V, E_{puz}) is the ring graph on 20 vertices (i.e., $\mathbb{Z}_{20}$).



How connected must the people graph be to solve the puzzle?

Basic jigsaw percolation

Start by the partition of V into singleton clusters. Iteratively create coarser partitions by merging existing clusters.

Basic jigsaw percolation

Start by the partition of V into singleton clusters. Iteratively create coarser partitions by merging existing clusters.

Basic jigsaw merging rule [BCDS]

Merge two clusters, W_1 and W_2, if there is a puzzle edge between a pair $v_1 \in W_1$, $v_2 \in W_2$; and a people edge between a pair $v_1' \in W_1$, $v_2' \in W_2$.

Basic jigsaw percolation

Start by the partition of V into singleton clusters. Iteratively create coarser partitions by merging existing clusters.

Basic jigsaw merging rule [BCDS]

Merge two clusters, W_1 and W_2, if there is a puzzle edge between a pair $v_1 \in W_1$, $v_2 \in W_2$; and a people edge between a pair $v_1' \in W_1$, $v_2' \in W_2$.

The event `Solve` happens if eventually all vertices are in the same cluster.

Adjacent-Edge (AE) jigsaw percolation

Adjacent-Edge Merging Rule [BCDS]

Merge any two clusters, W_1 and W_2, if there is a vertex $v_1 \in W_1$ and vertices $v_2, v_2' \in W_2$ such that there is a puzzle edge between v_1 and v_2, and a people edge between v_1 and v_2'.

Adjacent-Edge (AE) jigsaw percolation

Adjacent-Edge Merging Rule [BCDS]

Merge any two clusters, W_1 and W_2, if there is a vertex $v_1 \in W_1$ and vertices $v_2, v_2' \in W_2$ such that there is a puzzle edge between v_1 and v_2, and a people edge between v_1 and v_2'.

Open question: Is Adjacent-Edge jigsaw percolation distinguishable from basic jigsaw percolation?
That is, are in some substantial way group connections more important than individual ones?
Our results (and their proofs) apply to both versions.

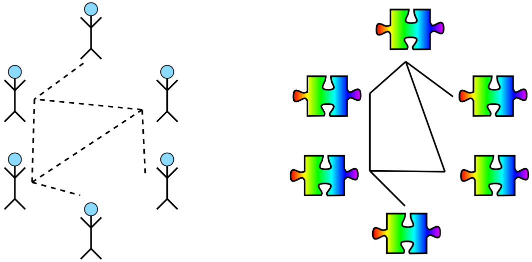
AE jigsaw dynamics: a simple example

Each person has one unique piece of the puzzle.

People Graph of who knows whom: (V, E_{ppl}).

Puzzle Graph of compatible ideas: (V, E_{puz}).

AE jigsaw dynamics: a simple example

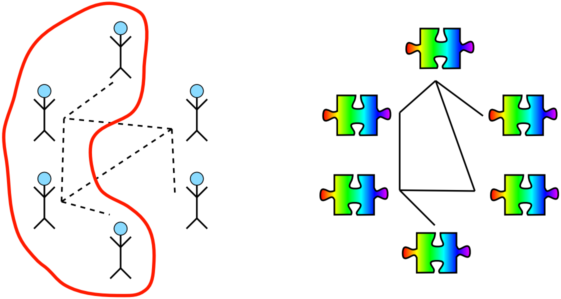


Each person has one unique piece of the puzzle.

People Graph of who knows whom: (V, E_{ppl}).

Puzzle Graph of compatible ideas: (V, E_{puz}).

AE jigsaw dynamics: a simple example



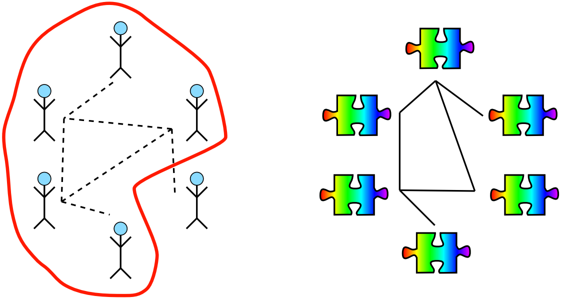
Each person has one unique piece of the puzzle.

People Graph of who knows whom: (V, E_{ppl}).

Puzzle Graph of compatible ideas: (V, E_{puz}).

Successively merge groups that know one another and have compatible puzzle pieces.

AE jigsaw dynamics: a simple example



Each person has one unique piece of the puzzle.

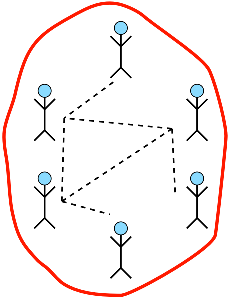
People Graph of who knows whom: (V, E_{ppl}).

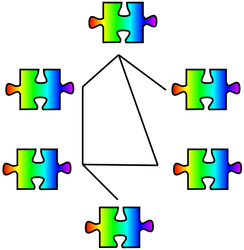
Puzzle Graph of compatible ideas: (V, E_{puz}).

Successively merge groups that know one another and have compatible puzzle pieces.

AE jigsaw dynamics: a simple example

Solved the puzzle!





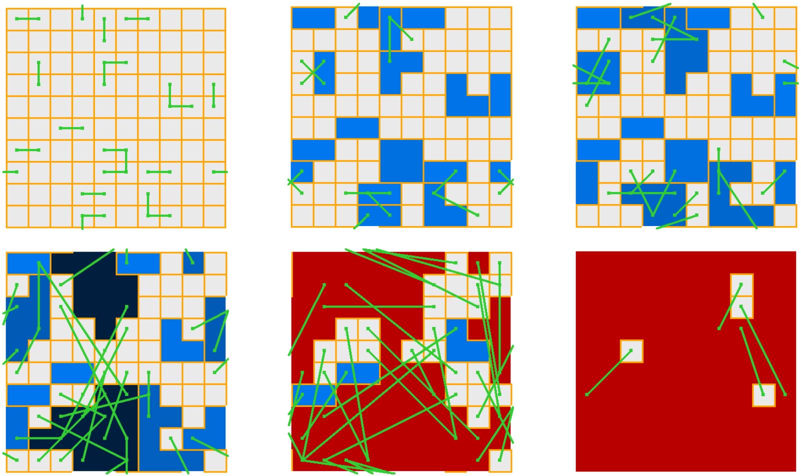
Each person has one unique piece of the puzzle.

People Graph of who knows whom: (V, E_{ppl}).

Puzzle Graph of compatible ideas: (V, E_{puz}).

Successively merge groups that know one another and have compatible puzzle pieces.

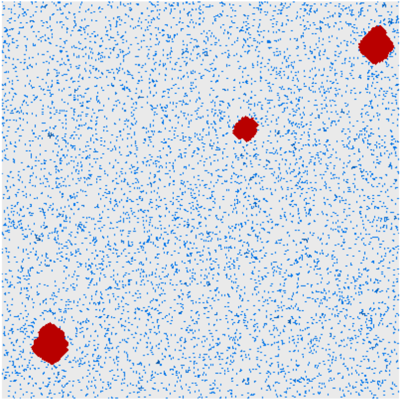
AE jigsaw dynamics on a torus



Adjacent-Edge JP on 10 $\times$ 10 torus (puzzle graph), with $p = 0.11$ (people graph is $G(10^2, 0.11)$), at times $t = 0, \ldots, 5$.

Jigsaw percolation on $\mathbb{Z}_n^2$: nucleation

AE JP on $\mathbb{Z}_n^2$ with $n = 400$,
$p = 0.021$ at $t = 31$.



Apparently, solving the puzzle is caused by a local concentration of highly connected individuals that create a gradually growing partial solution by adding boundary pieces. This phenomenon is called nucleation.

Results for general puzzle graphs

Setting: a sequence of connected puzzle graphs on N vertices and maximum degree D, with $N \to \infty$; and Erdős-Rényi people graph with edge density p on the same vertex set.

Theorem [BCDS]

If $p = \lambda / \log N$ with $\lambda > \pi^2/6$, then

$$\mathbb{P}_p(\texttt{Solve}) \to 1.$$

Theorem

If $p = \mu/(D \log N)$ with $\mu < 1/30$, then

$$\mathbb{P}_p(\texttt{Solve}) \to 0.$$

Corollary: For puzzles of bounded degree, the transition between low and high probability of `Solve` occurs when $p = \Theta(1/\log N)$.

Rule of thumb

On vertex-transitive graphs with N vertices of degree D, the transition between low and high $\mathbb{P}(\texttt{Solve})$ typically occurs as $pD \log N$ changes between a small constant and a large constant.

Rule of thumb

On vertex-transitive graphs with N vertices of degree D, the transition between low and high $\mathbb{P}(\texttt{Solve})$ typically occurs as $pD \log N$ changes between a small constant and a large constant.

This rule does not hold universally: if G_{puz} is the complete graph on N vertices, $p \approx 1/(N \log N)$ is too small for connectivity of G_{ppl}!

Rule of thumb

On vertex-transitive graphs with N vertices of degree D, the transition between low and high $\mathbb{P}(\texttt{Solve})$ typically occurs as $pD \log N$ changes between a small constant and a large constant.

This rule does not hold universally: if G_{puz} is the complete graph on N vertices, $p \approx 1/(N \log N)$ is too small for connectivity of G_{ppl}!

It does hold for many famous graphs.

Theorem: Hypercube puzzle

Assume the puzzle graph is the hypercube $\{0, 1\}^n$. There exist constants $c_1, c_2 > 0$ so that

$$\mathbb{P}(\texttt{Solve}) \to \begin{cases} 0 & \text{when } p \leq c_1/n^2 \\ 1 & \text{when } p \geq c_2/n^2 \end{cases}$$

Rule of thumb

On vertex-transitive graphs with N vertices of degree D, the transition between low and high $\mathbb{P}(\texttt{Solve})$ typically occurs as $pD \log N$ changes between a small constant and a large constant.

This rule does not hold universally: if G_{puz} is the complete graph on N vertices, $p \approx 1/(N \log N)$ is too small for connectivity of G_{ppl}!

It does hold for many famous graphs.

Theorem: Hypercube puzzle

Assume the puzzle graph is the hypercube $\{0,1\}^n$. There exist constants $c_1, c_2 > 0$ so that

$$\mathbb{P}(\texttt{Solve}) \to \begin{cases} 0 & \text{when } p \leq c_1/n^2 \\ 1 & \text{when } p \geq c_2/n^2 \end{cases}$$

Open question: Sharp constant?

Proof of the upper bound: local growth

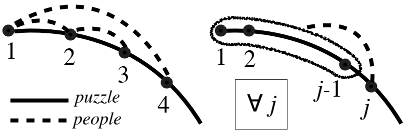
$$p > (1+\epsilon)\frac{\pi^2}{6\log N} \implies \mathbb{P}_p(\texttt{Solve}) \to 1$$

Proof of the upper bound: local growth

$$p > (1+\epsilon)\frac{\pi^2}{6\log N} \implies \mathbb{P}_p(\texttt{Solve}) \to 1$$

`Grow`: j is people-connected to $\{1, 2, \ldots, j-1\}$, for all $j \leq K$.

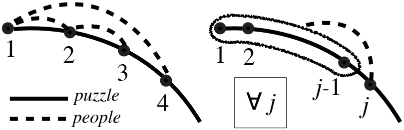
Then, with $g(x) = -\log(1-e^{-x})$,



Proof of the upper bound: local growth

$$p > (1+\epsilon)\frac{\pi^2}{6\log N} \implies \mathbb{P}_p(\texttt{Solve}) \to 1$$

$\texttt{Grow}$: j is people-connected to $\{1, 2, \ldots, j-1\}$, for all $j \leq K$.
Then, with $g(x) = -\log(1 - e^{-x})$,



$$\mathbb{P}_p(\texttt{Grow}) \geq \prod_{j=1}^{\infty}\left(1-(1-p)^j\right) \geq \exp\left(-\sum_j g(pj)\right)$$

$$\geq \exp\left(-p^{-1}\int_0^{\infty} g(x)\,dx\right) = \exp\left(-p^{-1}\cdot\frac{\pi^2}{6}\right)$$

$$\geq N^{-1/(1+\epsilon)}$$

Proof of the upper bound: unstoppable clusters

Assume $K \geq C(\log N)^2$ for a large enough C.

Assume you can create a successful instance of `Grow` on some set S of K connected vertices, and thus generate a cluster of size K. This cluster is unstoppable! Why?

Proof of the upper bound: unstoppable clusters

Assume $K \geq C(\log N)^2$ for a large enough C.

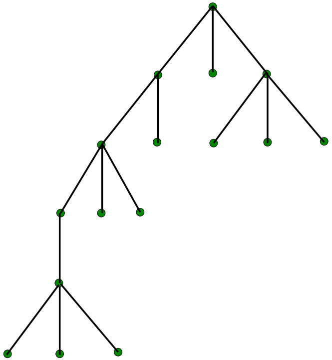
Assume you can create a successful instance of `Grow` on some set S of K connected vertices, and thus generate a cluster of size K. This cluster is unstoppable! Why?

With high probability, all other vertices are people-connected to it:

$$\mathbb{P}_p(\text{some vertex is not people-connected to } S)$$
$$\leq N(1-p)^K \leq Ne^{-pK} \leq Ne^{-C\log N} = N^{1-C}.$$

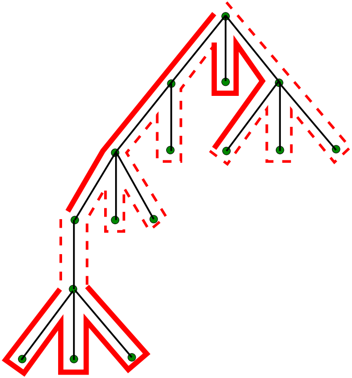
Upper bound: independent opportunities for `Grow`

Take a spanning tree for (V, E_{puz}), draw it on the plane,



Upper bound: independent opportunities for `Grow`

Take a spanning tree for (V, E_{puz}), draw it on the plane, then doubly-traverse the edges to find $\Omega(N/K^2) \gg N^{1/(1+\epsilon)}$ edge-disjoint subtrees with K vertices. (Example: $K = 4$.)



Lower bound: a necessary condition for `Solve`

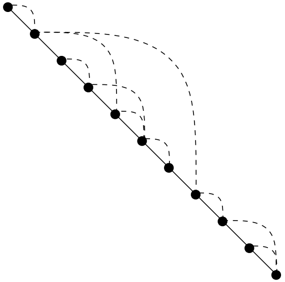
If c is small enough, and D is the maximum degree of puzzle graph,

$$p < \frac{c}{D \log N} \implies \mathbb{P}_p(\texttt{Solve}) \to 0$$

Lower bound: a necessary condition for `Solve`

If c is small enough, and D is the maximum degree of puzzle graph,

$$p < \frac{c}{D \log N} \implies \mathbb{P}_p(\texttt{Solve}) \to 0$$



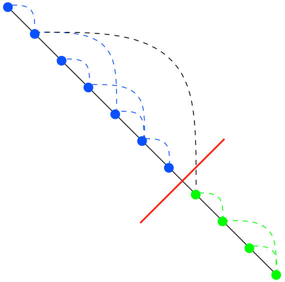
Necessary condition:

For any k there is a subset of vertex set with size $\in [k, 2k]$ that is internally solved. This set must be connected in both graphs.

Lower bound: a necessary condition for `Solve`

If c is small enough, and D is the maximum degree of puzzle graph,

$$p < \frac{c}{D \log N} \implies \mathbb{P}_p(\texttt{Solve}) \to 0$$



Necessary condition:

For any k there is a subset of vertex set with size $\in [k, 2k]$ that is internally solved. This set must be connected in both graphs.

Lower bound: connected sets of logarithmic size

1. The number of puzzle-connected sets of k vertices containing a fixed vertex is at most $(3D)^k$.

Lower bound: connected sets of logarithmic size

1. The number of puzzle-connected sets of k vertices containing a fixed vertex is at most $(3D)^k$.
2. The probability that a fixed set of $k = \alpha \log N$ people contains at least k edges is at most $\binom{k^2}{k} p^k \leq \frac{(3k^2)^k}{k^k} p^k = (3kp)^k \leq (3c\alpha/D)^{\alpha \log N}$.

Lower bound: connected sets of logarithmic size

1. The number of puzzle-connected sets of k vertices containing a fixed vertex is at most $(3D)^k$.
2. The probability that a fixed set of $k = \alpha \log N$ people contains at least k edges is at most $\binom{k^2}{k} p^k \le \frac{(3k^2)^k}{k^k} p^k = (3kp)^k \le (3c\alpha/D)^{\alpha \log N}$.
3. If `Solve` happens, there exists a set of size $\alpha \log N$, for some $\alpha \in [1, 2]$ that is connected in the puzzle graph and contains at least $\alpha \log N$ people edges.

Lower bound: connected sets of logarithmic size

1. The number of puzzle-connected sets of k vertices containing a fixed vertex is at most $(3D)^k$.
2. The probability that a fixed set of $k = \alpha \log N$ people contains at least k edges is at most $\binom{k^2}{k} p^k \leq \frac{(3k^2)^k}{k^k} p^k = (3kp)^k \leq (3c\alpha/D)^{\alpha \log N}$.
3. If `Solve` happens, there exists a set of size $\alpha \log N$, for some $\alpha \in [1, 2]$ that is connected in the puzzle graph and contains at least $\alpha \log N$ people edges.
4. Thus $P(\texttt{Solve}) \leq N \log N \sup_\alpha (3D)^{\alpha \log N} (3c\alpha/D)^{\alpha \log N} \leq N^2 (18c)^{\log N}$.

Critical probability

The *critical probability* p_c is the solution p to $\mathbb{P}_p(\texttt{Solve}) = 1/2$.

Critical probability

The *critical probability* p_c is the solution p to $\mathbb{P}_p(\texttt{Solve}) = 1/2$.

Sharp transition occurs if, for every $\epsilon > 0$,

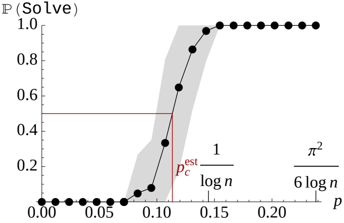
$$\mathbb{P}_{(1-\epsilon)p_c}(\texttt{Solve}) \to 0 \text{ and } \mathbb{P}_{(1+\epsilon)p_c}(\texttt{Solve}) \to 1.$$

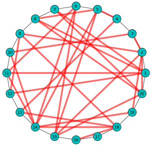
Sharp transition is thought to be a nearly universal phenomenon, due to results by E. Friedgut, G. Kalai, and others. Their theorems do not cover Jigsaw percolation, as they depend on transitivity of random bits.

We can prove sharp transition only when we can establish precise asymptotics for p_c.

Sharp transition for the ring puzzle

AE JP on $\mathbb{Z}_n$ with $n = 1000$, averaged over 200 trials [BCSD].





Theorem: ring puzzle.

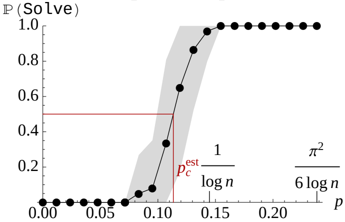
If $G_{\text{puz}} = \mathbb{Z}_n$ then

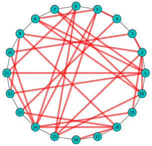
$$p_c \log n \to \frac{\pi^2}{6},$$

with sharp transition.

Sharp transition for the ring puzzle

AE JP on $\mathbb{Z}_n$ with $n = 1000$, averaged over 200 trials [BCSD].





Theorem: ring puzzle.

If $G_{\text{puz}} = \mathbb{Z}_n$ then

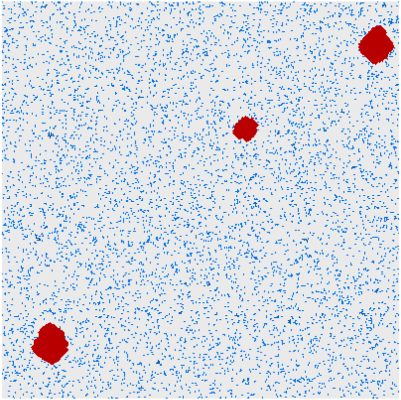
$$p_c \log n \to \frac{\pi^2}{6},$$

with sharp transition.

Open question: Note that the sharp transition in the picture is significantly below $\pi^2/(6 \log n)$. Why?

Jigsaw percolation on $\mathbb{Z}_n^2$

AE JP on $\mathbb{Z}_n^2$ with $n = 400$,
$p = 0.021$ at $t = 31$.



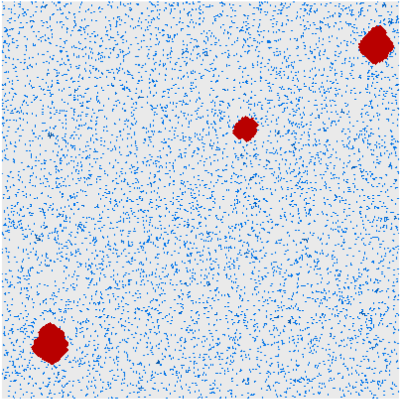
Theorem: 2d-torus AE JP

Assume $G_{\text{puz}} = \mathbb{Z}_n^2$. For large enough n,

$$\frac{0.0388}{\log n} < p_c < \frac{0.303}{\log n}.$$

Jigsaw percolation on $\mathbb{Z}_n^2$

AE JP on $\mathbb{Z}_n^2$ with $n = 400$, $p = 0.021$ at $t = 31$.



Theorem: 2d-torus AE JP

Assume $G_{\text{puz}} = \mathbb{Z}_n^2$. For large enough n,

$$\frac{0.0388}{\log n} < p_c < \frac{0.303}{\log n}.$$

Open question: Can these bounds be improved? Can sharp transition be proved? One obstacle: we known of no useful necessary condition other than connectivity!

Other results: stricter verification

Merging Rules

Merge two clusters, W_1 and W_2, if at least one of the following holds:

1. there are doubly connected vertices $v_1 \in W_1$ and $v_2 \in W_2$; or

Other results: stricter verification

Merging Rules

Merge two clusters, W_1 and W_2, if at least one of the following holds:

1. there are doubly connected vertices $v_1 \in W_1$ and $v_2 \in W_2$; or
2. there is a vertex $v_1 \in W_1$ with a puzzle neighbor and at least σ people neighbors in W_2. (Thus σ is the minimal number of “referees” needed to verify that a piece fits.)

Other results: stricter verification

Merging Rules

Merge two clusters, W_1 and W_2, if at least one of the following holds:

1. there are doubly connected vertices $v_1 \in W_1$ and $v_2 \in W_2$; or
2. there is a vertex $v_1 \in W_1$ with a puzzle neighbor and at least σ people neighbors in W_2. (Thus σ is the minimal number of “referees” needed to verify that a piece fits.)

Theorem: scaling for large σ.

Assume that the sequence of puzzle graphs has bounded degree. For $N \geq N_0(\sigma)$, p_c is between two constants times $\sigma^2 / \log N$.

Requiring a lot of “referees” to verify a fit is very costly!

Other results: free corner fit

Merging Rules

Merge two clusters, W_1 and W_2, if at least one of the following holds:

1. there are doubly connected vertices $v_1 \in W_1$ and $v_2 \in W_2$; or
2. there is a vertex $v_1 \in W_1$ with a puzzle neighbor and at least σ people neighbors in W_2; or

Other results: free corner fit

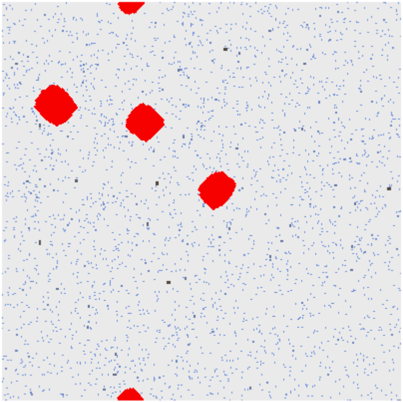
Merging Rules

Merge two clusters, W_1 and W_2, if at least one of the following holds:

1. there are doubly connected vertices $v_1 \in W_1$ and $v_2 \in W_2$; or
2. there is a vertex $v_1 \in W_1$ with a puzzle neighbor and at least σ people neighbors in W_2; or
3. there is a $v_1 \in W_1$ with at least 2 puzzle neighbors in W_2. (On two-dimensional lattice graphs, corner pieces require no verification at all.)

Free corner fit on $\mathbb{Z}_n^2$

JP ($\sigma = 1$, free corner fit) with $n = 400$, $p = 0.009$, at $t = 31$.



Theorem: 2d-torus JP with free corner fit

Let $\tau = 1, \sigma \geq 1$, $\theta = 2$, and $g(x) = -\log(1 - e^{-x})$. Let

$$\lambda_c = \int_0^\infty g\left(\frac{x^{2\sigma+1}}{\sigma!}\right) dx = \frac{(\sigma!)^{\frac{1}{2\sigma+1}}\Gamma(\frac{1}{2\sigma+1})\zeta(\frac{2\sigma+2}{2\sigma+1})}{(2\sigma+1)}.$$

Then as $n \to \infty$,

$$p_c(\log n)^{2+\frac{1}{\sigma}} \to \lambda_c^{2+\frac{1}{\sigma}},$$

with sharp transition.

Origin of the power of log n

Why is $p_c \approx (logn)^{2+1/\sigma}$?

1. Consider an $L \times L$ square with $L \approx p^{-\sigma/(2\sigma+1)} \ll p^{-1/2}$. Then, the probability that a point on the boundary is G_{ppl}-connected to a point inside is on the order $L(L^2p)^\sigma = L^{2\sigma+1}p^\sigma \approx 1$.

Origin of the power of log n

Why is $p_c \approx (logn)^{2+1/\sigma}$?

1. Consider an $L \times L$ square with $L \approx p^{-\sigma/(2\sigma+1)} \ll p^{-1/2}$. Then, the probability that a point on the boundary is G_{ppl}-connected to a point inside is on the order $L(L^2p)^\sigma = L^{2\sigma+1}p^\sigma \approx 1$.
2. This order of L is critical, and the probability of the formation of clusters that traverse such sizes is about $\exp(-Cp^{-\sigma/(2\sigma+1)})$, for some constant C.

Origin of the power of log n

Why is $p_c \approx (logn)^{2+1/\sigma}$?

1. Consider an $L \times L$ square with $L \approx p^{-\sigma/(2\sigma+1)} \ll p^{-1/2}$. Then, the probability that a point on the boundary is G_{ppl}-connected to a point inside is on the order $L(L^2 p)^\sigma = L^{2\sigma+1} p^\sigma \approx 1$.
2. This order of L is critical, and the probability of the formation of clusters that traverse such sizes is about $\exp(-Cp^{-\sigma/(2\sigma+1)})$, for some constant C.
3. This probability must exceed $1/n^2$ for the puzzle to be solved, which gives the claimed power for p_c.

More

1. C. D. Brummitt, S. Chatterjee, P. S. Dey, D. Sivakoff, *Jigsaw percolation: What social networks can collaboratively solve a puzzle?* Annals of Applied Probability 25 (2015), 2013–2038. arXiv:1207.1927.

2. J. Gravner, D. Sivakoff, *Nucleation scaling in jigsaw percolation*. Annals of Applied Probability 27 (2017), 395–438. arXiv:1310.2194.

3. B. Bollobás, O. Riordan, E. Slivken, P. Smith, *The threshold for jigsaw percolation on random graphs*. arXiv:1503.05186