Combinatorica **27** (6) (2007) 645–658
DOI: 10.1007/s00493-007-2241-5

COMBINATORICA
Bolyai Society – Springer-Verlag

A PHASE TRANSITION FOR THE METRIC DISTORTION OF PERCOLATION ON THE HYPERCUBE

OMER ANGEL, ITAI BENJAMINI

Received August 11, 2005

Let H_n be the hypercube $\{0,1\}^n$, and denote by $H_{n,p}$ Bernoulli bond percolation on H_n, with parameter $p = n^{-\alpha}$. It is shown that at $\alpha = 1/2$ there is a phase transition for the metric distortion between H_n and $H_{n,p}$. For $\alpha < 1/2$, the giant component of $H_{n,p}$ is likely to be quasi-isometric to H_n with constant distortion (depending only on α). For $1/2 < \alpha < 1$ the minimal distortion tends to infinity as a power of n. We argue that the phase $1/2 < \alpha < 1$ is an analogue of the non-uniqueness phase appearing in percolation on non-amenable graphs.

1. Introduction

The hypercube H_n is the graph with vertex set $\{0,1\}^n$ and edges between any two vertices that differ in a single coordinate. Bernoulli bond percolation on a graph G with parameter p is a distribution on subgraphs $H \subset G$ where each edge is in H with probability p independently of all other edges. Let $H_{n,p}$ be the random graph resulting from Bernoulli bond percolation on H_n. Any graph G may be viewed as a metric space. The vertex set $V(G)$ is endowed with the induced graph metric, i.e., $d(x,y)$ is the length of the shortest path from x to y. The main focus of this paper is the effect of Bernoulli percolation on the geometry of the hypercube.

In [7] Hastad, Leighton and Newman show that if p is taken to be constant, then for large n the distortion (defined below) between H_n and $H_{n,p}$ is at most a constant with probability tending to 1. Their motivation was consideration of a graph as a representation of a computer network, with

Mathematics Subject Classification (2000): 05D40; 68R10, 60D05 processors at the vertices and communication channels along edges. They were interested in the latency of embedding H_n in $H_{n,p}$, which measures the loss of computational power resulting from eliminating a subset of the processors (vertices) or communication channels (edges). The latency of an embedding between graphs is closely related to metric distortion. It is slightly stronger (in the sense that small latency implies small distortion), since it is also takes into account possible communication bottlenecks.

Bond percolation on the hypercube is first analyzed in the classic paper [1] of Ajtai, Komlós and Szemerédi. It is shown that if $p = an^{-1}$ with $a < 1$, then all connected components of $H_{n,p}$ have size polynomial in n (hence logarithmic in the size of H_n). On the other hand, if $a > 1$ then $H_{n,p}$ contains a single giant component of size linear in the size of H_n, and of diameter of order n.

In the regime $p < n^{-1}$, when all connected components are small it is evident that H_n is completely different geometrically from any of the components of $H_{n,p}$. However, when a giant component exists it is a priori possible that the distortion between the cube and the giant component is constant. This leaves a gap between $p = n^{-1}$ where a giant component appears, and p constant where the distortion was previously known to be small.

Metric distortion as defined below is just one of the ways to quantify similarity between metric spaces. We use metric distortion to describe the transition at $\alpha = 1/2$, though the transition is fundamental and would be reflected in any other measure of similarity. Thus, for example, the latency exhibits a transition at $\alpha = 1/2$ similar to that of the distortion. Small metric distortion between two graphs (or metric spaces) implies similarity in several geometric respects. For more information on metric distortion see [8].

Theorem 1. *Fix $\alpha \in [0,1]$, and let $p = n^{-\alpha}$.*

A. *If $\alpha < 1/2$ then there is some constant $c = c(\alpha)$ such that*

$$\mathbb{P}\big(D(H_n, H_{n,p}) < c\big) \xrightarrow[n\to\infty]{} 1.$$

B. *If $\alpha > 1/2$ then there is some constant $\beta = \beta(\alpha) > 0$ such that*

$$\mathbb{P}\big(D(H_n, H_{n,p}) > n^{\beta}\big) \xrightarrow[n\to\infty]{} 1.$$

A result similar to Theorem 1 holds for site percolation on the cube, as well as for mixed percolation (with threshold at $p_b p_s = n^{-1/2}$). The proof is essentially identical.

As remarked above, when $p < n^{-1}$, there are no large components in $H_{n,p}$, so the distortion is trivially large. Thus Theorem 1 establishes three phases.

For $\alpha > 1$ there is no giant component. For $\alpha < 1/2$ there is a giant component that is geometrically similar to H_n. When $1/2 < \alpha < 1$ we find an intermediate phase, where a giant component exists, but it is geometrically very different from the complete cube.

Roughly speaking, a mapping between two metric spaces is said to have small distortion if it changes all distances by approximately a constant factor. The distortion between two spaces is small if there exists a mapping between them with small distortion. To make or statements precise, we use the following measure of the distortion between metric spaces.

Definition 2. The *distortion* of a map f between metric spaces (X, d_X), (Y, d_Y) is given by $D(f) = \frac{D_+(f)}{D_-(f)}$ where

$$D_+(f) = 1 \vee \left(\sup_{a,b \in X} \frac{d_Y\big(f(a), f(b)\big)}{d_X(a,b)} \right),$$

$$D_-(f) = \inf_{a,b \in X} \frac{1 \vee d_Y\big(f(a), f(b)\big)}{d_X(a,b)}.$$

The metric distortion between metric spaces X and Y, is given by

$$D(X,Y) = \inf_{f: X \to Y} D(f).$$

Note that since we deal with finite metric spaces of different sizes, we necessarily deal with functions that are not 1-to-1. The above definition accommodates the possibility that two points a, b are mapped to the same point in Y. This does not constitute a significant change from more common definitions. It should also be noted that this definition is not symmetric, as the functions are not required to have a full image. Finally, note that the distance between disjoint connected components of a graph is infinite, so that only functions that map connected components into connected components are of interest.

When estimating the distortion between a graph G and the bond percolation sub-graph G_p there is a special map to consider since the two have the same vertex sets. Thus one can ask whether the identity map from G_p to G is close to optimal. However, there is no reason to expect this map to give the minimal distortion or even close. Indeed, generally the identity embedding does not give a good estimate for the distortion. However, it will turn out that in the super-critical case we study – when the metric distortion is small – there is a map that approximates the identity map which gives a small distortion.

At this point it is interesting to compare the behavior of the metric distortion of percolation in the hypercube to the behavior in some other graphs. For a cube of side length n in the lattice $\mathbb{Z}^d$, the distortion between the box and the giant component is $O(\log n)$. This can be seen by mapping any x to the point of the giant component nearest to x. While this does not give constant distortion, the distortion is small in relation to the diameter of the graph. More importantly, there is no additional phase transition beyond the percolation threshold: the once the giant component exists it is similar to the entire cube.

For percolation on the complete graph, the metric distortion is given by the diameter of the giant component. This diameter only varies between constant and $O(\log n)$.

In the hypercube the giant component has diameter $\theta(n)$ a.s. for any $p > an^{-1}$ with $a > 1$. While our lower bound for the distortion is n^β, we believe that the distortion in this case is in fact linear in the diameter. The difficulty in proving this stems from the need to eliminate maps that themselves do not preserve the geometry of the cube.

The ideas in the proofs for the two phases are quite different. In the super-critical regime, with $\alpha < 1/2$, the main idea is that two vertices that are nearby in H_n are likely to stay nearby in $H_{n,p}$. Thus the identity map is almost good enough, as it gives a constant distortion for most pairs of vertices. The few misbehaving vertices require some tweaking of the map, though the distance from x to $f(x)$ is still bounded.

In the sub-critical case, Theorem 1 implies that there is some fundamental geometric difference between H_n and $H_{n,p}$. Our proof is based on demonstrating such a difference. Specifically, we show that while the cube contains many geodesic cycles (i.e., isometrically embedded cycles), the giant component for $\alpha > 1/2$ is locally very treelike and contains much fewer geodesic cycles.

In the sub-critical regime there is a constant probability that two adjacent vertices that are both in the giant component are not joined by any short path in $H_{n,p}$, but instead have distance proportional to the diameter of $H_{n,p}$. Of course, since for the cube (and in much greater generality, [1,3]) the giant component is known to be unique, the distance between the two vertices is finite. However, there is no longer a direct correlation between the metric of the cube and the graph metric of the giant component.

Since giant components are the analogue in finite graphs of infinite percolation clusters, this behavior can be thought of as the finite graph analogue of non-uniqueness of infinite clusters. For graphs where p_u (the threshold for uniqueness of the infinite cluster) is strictly greater than p_c there is an intermediate regime where two nearby vertices have a positive probability of being in disjoint infinite components. Thus being connected "through infinity" in the non-uniqueness regime is replaced here by being connected through distant parts of the giant component. The inequality $p_c < p_u$ is conjectured to hold for any non-amenable graph, and has been established in many cases, see e.g. [9,5].

Thus we propose that having the metric in the giant component very different from the original graph metric is the following as an analogue in finite graphs of having non-unique infinite components. In particular, a possible criterion is that neighbours in the base graphs are at distance of the order of the diameter in the giant component with probability bounded away from 0. For further information regarding percolation on finite graphs see e.g. [3].

We now proceed to present the proof of Theorem 1. The proof is separated into the super-critical and the sub-critical cases, as each requires a different approach. Section 4 includes some open questions raised by our results.

2. The super-critical case: $\alpha < 1/2$

To illustrate the proof we first show a weaker result, namely that the typical distortion of the identity map is constant. This has some inherent interest in light of the relation to the non-uniqueness phase discussed above.

Proposition 3. *Let $p = n^{-\alpha}$ with $\alpha < 1/2$. For some $\ell = \ell(\alpha)$, with probability tending to 1 the percolation distance between two given neighbours in H_n is at most $2\ell+1$.*

Sketch of Proof. Let x and y be two neighbours in H_n. We consider only paths of length $2\ell+1$ between them of the following form: The path makes ℓ steps in distinct coordinates, then a step in the coordinate in which x and y differ, and then the first ℓ steps are retraced in reversed order to reach y. The number of such paths is $\frac{(n-1)!}{(n-1-\ell)!} \sim n^\ell$. Each such path is open with probability $p^{2\ell+1}$. Let X be the number of open paths of this form between x and y. The expectation of X is $n^{(1-2\alpha)\ell-\alpha}(1+o(1))$. For some ℓ depending only on α the exponent is positive.

A pair of paths is unlikely to have a large intersection. The second moment of X can be approximated by the square of the expectation. The largest error term comes from pairs of paths that intersect only in their first step, so that $\mathbb{E}X^2 = (\mathbb{E}X)^2(1+n^{2\alpha-1}+o(n^{2\alpha-1}))$. The second moment method (see [2]) yields that

$$\mathbb{P}(X=0) \le 1 - \frac{(\mathbb{E}X)^2}{\mathbb{E}X^2} = \big(1+o(1)\big)n^{2\alpha-1} \to 0. \qquad \blacksquare$$

The idea in the proof of Theorem 1(A) is similar. We argue that there are many paths of some fixed length between nearby vertices, and that one of them is very likely to be open. However, to get a path between any pair of vertices we need an exponentially small probability of failure. To achieve this we must choose the set of paths we consider carefully.

Fix ℓ so that $(1-2\alpha)\ell > 9\alpha$, which is possible as $\alpha < 1/2$. Set $m = \lfloor \frac{n-1}{\ell+2} \rfloor$ and fix throughout this section some partial partition of the coordinates into $\ell+2$ disjoint sets of size m denoted $A, B, C_1, \dots, C_\ell$ (with at least one coordinate left unused). This allows us to talk about B-neighbours (vertices that differ only in a single coordinate in B) and A-edges with obvious meaning.

A vertex in $H_{n,p}$ will be called *good* if it has at least $2m$ vertices at distance 2 (in $H_{n,p}$) differing only in A-coordinates. We first show that good vertices are dense:

Lemma 4. *Let $p = n^{-\alpha}$ for some $\alpha < 1/2$. With probability tending to 1, every vertex v has some good B-neighbour u.*

Proof. First we show that the probability that any fixed vertex v is good tends to 1 as $n \to \infty$, with α fixed. To this end we count open paths of length 2 using only A-coordinates starting at v. If there are more than $4m$ such paths, then they connect v to at least $2m$ vertices and v is good (since each vertex can be reached in at most two ways). Since $|A| = m$, the probability that v is incident on at least $\frac{1}{2}pm = c_1 n^{1-\alpha}$ open A-edges tends to 1 as $n \to \infty$. Each such edge can be extended in $m-1$ ways to a path of length 2. Therefore, on this event there are at least $(m-1)c_1 n^{1-\alpha}$ edges that extend the open edges to paths of length 2. With probability tending to 1 at least fraction $\frac{1}{2}p$ of these are open giving a total of at least $c_2 n^{2-2\alpha}$ open paths. Since for n large enough, $c_2 n^{2-2\alpha} > 4n > 4m$, the probability that v is good tends to 1.

Fix some v and let $u_1, \dots, u_m$ be u's B-neighbours. Since the u_i's all differ in their B-coordinates, and since goodness of a vertex depends only on A-edges, the events of the u_i's being good are all independent. For large enough n each u_i is good with probability at least $1 - 3^{-\ell}$ and then the probability that none of the u_i's is good is at most $3^{-ml} \ll 2^{-n}$. Thus, a union bound over all vertices v shows that with probability tending to 1 every v has a good B-neighbour u. ∎

Proof of Theorem 1(A). We construct a map f that takes each vertex x to some good B-neighbour of x. By the above lemma, with probability tending to 1 such a map exists. Since $d(x, f(x)) = 1$, it follows that $d(f(x), f(y)) \geq d(x, y) - 2$ and so $D_-(f) \geq 1/3$. It remains to show that (with high probability) distances do not increase by more then some constant factor. For this, it suffices to show that with probability tending to 1, for any two neighbours x, y, the percolation distance between $f(x)$ and $f(y)$ is bounded by $2\ell+9$.

The H_n-distance between $f(x)$ and $f(y)$ is at most 3. First, fix some coordinate e where $f(x)$ and $f(y)$ differ. In the remainder there are some slight variations according as to which set of the partition contains e (or rather the coordinate corresponding to e). Since $f(x)$ is good it has $2m$ vertices at percolation distance 2, differing only in the A-coordinates. Let $x_1, \dots, x_m$ be m of those that do not differ from $f(x)$ in the e coordinate. This is possible even if $e \in A$ since at most $m-1$ of the $2m$ vertices are ruled out by this constraint. Otherwise all $2m$ are suitable. We similarly define $y_1, \dots, y_m$ at distance 2 from $f(y)$. Note that the H_n distance $d(x_i, y_i)$ is at most 7.

We now look for an open path from x_i to y_i of length at most $2\ell+9$ for some i, and bound the probability that there is no such path. The second moment method is used to bound the probability of having no path for some i. The sets of paths we consider are chosen so that distinct i's will be independent.

To each i we associate a unique coordinate $b_i \in B$. If $e \in B$ we use instead of e the one of the remaining coordinates not in any of the sets. For a sequence $\mathbf{c} = (c_1, \dots, c_\ell)$ of coordinates with $c_k \in C_k$ we define the path $P_i(\mathbf{c})$ from x_i to y_i as the path making ℓ steps in the coordinates given by $\mathbf{c}$, making a step in the coordinate b_i, making steps in the coordinates where x_i and y_i differ, and then retracing the $\ell+1$ steps given by i and $\mathbf{c}$ in reversed order, thereby arriving at y_i. If $e \in C_k$ for some k then we again replace it by the unused coordinate.

We now claim that for $i \neq i'$ and any $\mathbf{c}, \mathbf{c}'$, the paths $P_i(\mathbf{c})$ and $P_{i'}(\mathbf{c}')$ are disjoint. Indeed, suppose that some v is visited by $P_i(\mathbf{c})$. According to the e coordinate of v we can say whether v is in the first or second half of the path. Assume w.l.o.g. it is the first. By the distance from $f(x)$ we can determine the exact position of v in the path. Finally, either v differs from $f(x)$ in a unique B-coordinate, and then i is uniquely determined by v, or else v differs from $f(x)$ in exactly 2 A-coordinates and nothing else, and then $v = x_i$ for some i. In any case, we can determine i from v, so paths for different i's cannot intersect.

We assume for simplicity that $d(x_i, y_i) = 7$. If it is smaller, then $2\ell+9$ below is replaced by $2\ell+2+d(x_i, y_i)$, and a similar change applies to for the second moment calculation. The result is the same bound, but with slightly shorter paths.

Let X_i be the number of open paths from x_i to y_i.

$$\mathbb{E}X_i = p^{2\ell+9} m^\ell = cn^{(1-2\alpha)\ell-9\alpha},$$

where c is a constant depending only on ℓ. For $\ell=\ell(\alpha)$ as specified above the exponent is positive, and so the expected number of paths tends to infinity.

The only way for two paths P, P' to have an intersection is if for some k the initial k edges and final k edges agree. The second moment of X_i is given by

$$\begin{aligned}\mathbb{E}X_i^2 &= \sum_{P,P'} p^{4\ell+18-|P\cap P'|} \\ &\le (\mathbb{E}X_i)^2 \left(\sum_{k=0}^{\ell-1} p^{-2k} m^{-k} + m^{-\ell} p^{-2\ell-9}\right) \\ &\le (\mathbb{E}X_i)^2 \big(1 + cn^{2\alpha-1}\big).\end{aligned}$$

And so the probability of having $X_i > 0$ tends to 1 as $n \to \infty$. In particular, for large enough n we have $\mathbb{P}(X_i = 0) < 3^{-\ell}$. Since the X_i's are independent, the probability of having no open path for any of the i is less then $(3^{-\ell})^m$. In this case there is an open path of length $2\ell+13$ between $f(x)$ and $f(y)$. Since the number of pairs x, y is $n2^n \ll 3^{\ell m}$, a union bound shows that asymptotically almost surely distances increase by a factor of no more than $2\ell+13$. ∎

3. The Sub-Critical Case: $\alpha > 1/2$

For two vertices x, y at some fixed distance from each other, the number of paths of length ℓ between them grows like $n^{\ell/2}$, and so the probability of having an open path tends to 0 as $n \to \infty$. This shows that any map with $d(x, f(x))$ bounded cannot have small distortion. However, the image of a map need not cover all pairs x, y, and there will be some pairs where paths will exist. Thus to show that every map has a large distortion some deeper understanding of the geometric difference between H_n and $H_{n,p}$ is needed. Geodesic cycles provide this understanding.

A *geodesic cycle* in a graph is an isometric embedding of a cycle in the graph. The cube has a great number of geodesic cycles passing through any given vertex. These can be found by making any number of steps in distinct coordinates, and then repeating the same sequence of coordinates to return to the starting point.

Geodesic cycles (as well as a geodesic paths) are structures that are roughly preserved by any map with a small distortion. The key idea of our proof is to show that the percolated cube does not have many cycles, restricting the possible maps.

Lemma 5. *If a map from G to H has distortion D, and G has a geodesic cycle C of length $\ell > 4D$ passing through a vertex v, then there is a simple (self avoiding) cycle $C' \subset H$ of length $\ell' \in [\ell/2D, D\ell]$ so that $f(v)$ is at distance at most D^2 from C'.*

The key idea is to look at a cycle in H that is the "image" of C, and remove any small loops it has to extract a simple cycle.

Proof. Recall that the distortion is $D = D_+/D_-$. Suppose the vertices of C are $v_0, v_1, \ldots, v_\ell$ with $v_0 = v_\ell = v$, and their images in H are $x_i = f(v_i)$. Since $d(x_i, x_{i+1}) \leq D_+$, there are simple paths of length at most D_+ between them, and these paths may be joined to form a cycle C_0 in H of length at most $D_+\ell$, passing through all the x_i's in order. Generally C_0 is not simple. The paths chosen from x_i to x_{i+1} for various i's will be called arcs.

Next we extract from C_0 a simple cycle by erasing loops it includes. Note that the procedure we use is not the standard procedure known as loop erasure. Suppose that C_0 visits some vertex u twice. Since u cannot be used twice in the same arc, it appears in distinct arcs. If u appears in the arc (x_i, x_{i+1}) and in (x_j, x_{j+1}) then there is a path of length at most $2D_+$ between x_i and x_j. However, $d(x_i, x_j) \geq D_- d(v_i, v_j)$, and so $d(v_i, v_j) \leq 2D$. Since the v_i's form a geodesic cycle, breaking C_0 at the two occurrences of u results in one segment containing all but at most $2D$ of the x_i's.

Note that it is possible that a vertex u appears in more than two segments. However, the above considerations apply to any two appearances of u. This allows us to define the first and last appearances of u in C_0, so that from the first to the last the cycle passes through at most $2D$ of the x_i's. Thus simplifying C_0 by skipping from the first appearance directly to the last will cause at most $2D$ of the x_i's to be dropped. We will apply this recursively until a simple cycle is found.

Some care must be taken in the order such loops are erased. As noted, for each u that is visited multiple times there is a segment of C_0 that can be removed. It is harder to control the number of such steps, and so the length of the remaining simple cycle. To achieve this, at each step we pick some vertex u such that the corresponding segment is not contained in the segment of any other repeated vertex. After the loops rooted at u are removed, the vertex u is not contained in any of the short loops corresponding to any other vertex of the remaining cycle. Thus at each stage we pick some vertex with a maximal segment from its first to last appearances, and remove everything in between. Each vertex so selected will not be removed in any subsequent step. Since each such operation removes at least one and at most $2D$ of the x_i's from the cycle, the remaining cycle must have length at least $\frac{\ell}{2D}$.

Finally, $x_0 = f(v)$ might not be within the simple cycle that we found. This happens if x_0 is part of a loop that was removed from the cycle at one of the stages. Since each such loop spans at most $2D$ arcs, its length is at most $2DD_+ \leq 2D^2$. The vertex where such a loop is connected to the cycle is part of the final simple cycle. It follows that $f(v)$ is close to the cycle. ∎

Lemma 6. *Fix parameters $0 < \beta < \gamma < 2\alpha - 1$, and let $p = n^{-\alpha}$. For large enough n, the probability that a vertex v is included in at least one open simple cycle in $H_{n,p}$ of length 2ℓ for at least one $\ell \in [n^\beta, n^\gamma]$ is at most $n^{1-(2\alpha-1-\beta)n^\beta}$. Furthermore, the probability that v is within distance δ of such a cycle is at most $n^{1+\delta-(2\alpha-1-\beta)n^\beta}$.*

Proof. First we estimate the number of simple cycles of length 2ℓ in H_n that visit v. Since a cycle makes an even number of moves in each coordinate, the steps of a cycle of length 2ℓ can be partitioned into ℓ pairs of (non consecutive) moves in the same coordinate. The number of partitions of 2ℓ elements into pairs is $(2\ell-1)!!$. The number of ways to choose a coordinate for each of the pairs is n^ℓ. This is an overestimate since not every pairing generates a simple cycle, and cycles which use a coordinate more than twice are counted multiple times.

The probability that any cycle of length 2ℓ is open is $p^{2\ell} = n^{-2\alpha\ell}$. Thus the probability of v being in any open cycle of length exactly 2ℓ is at most $(2\ell-1)!!\,(n^{1-2\alpha})^\ell$. This bound is decreasing in ℓ as long as $2\ell < n^{2\alpha-1}$. Sum over ℓ in the given range. For large enough n the first term is the largest, and using $(2m-1)!! < m^m$ we find

$$\sum_{\ell=n^\beta}^{n^\gamma} (2\ell-1)!!\,\left(n^{1-2\alpha}\right)^\ell < n\left(2n^\beta - 1\right)!!\,\left(n^{1-2\alpha}\right)^{n^\beta} < n^{1-(2\alpha-1-\beta)n^\beta}.$$

Finally, the number of vertices within distance δ of v is at most n^δ, and the second claim follows from the first. ∎

Proof of Theorem 1(B). Given α, fix positive constants β, γ so that $\beta + \gamma < 2\alpha - 1$ and $\gamma > 3\beta$. Assume that there is a map $f : H_n \to H_{n,p}$ with distortion at most n^β. Every vertex in H_n is contained in a geodesic cycle of length $2n^\gamma$. By Lemma 4 every vertex in the range is at distance at most $n^{2\beta}$ from a simple cycle of length $\ell \in [n^{\gamma-\beta}, 2n^{\gamma+\beta}]$.

Let $S = S(\beta,\gamma)$ be the set of vertices of $H_{n,p}$ at percolation distance at most n^β from a cycle of length $\ell \in [n^{\gamma-\beta}, 2n^{\gamma+\beta}]$. Thus the range of the postulated function f must be contained in S. Vertices at distance greater than n^β are mapped by f to distinct images. So the pre-image of any vertex has size at most n^{n^β} and the range of f has size at least $2^n n^{-n^\beta}$. Thus if such f exists, then $|S| > 2^n n^{-n^\beta}$.

However, Lemma 6 bounds the probability that any given vertex is in S:

$$\mathbb{P}(x \in S) \leq n^{1+n^{2\beta}-(2\alpha-1-\gamma+\beta)n^{\gamma-\beta}}.$$

Consequently the probability that a function with distortion n^β exists is at most

$$\begin{aligned}\mathbb{P}\big(|S| > 2^n n^{-n^\beta}\big) \leq \frac{\mathbb{E}|S|}{2^n n^{-n^\beta}} &= \frac{2^n \mathbb{P}(x \in S)}{2^n n^{-n^\beta}} \\ &\leq n^{1+n^\beta+n^{2\beta}-(2\alpha-1-\gamma+\beta)n^{\gamma-\beta}}.\end{aligned}$$

When $\gamma > 3\beta$, the dominant term in the exponent is $n^{\gamma-\beta}$ with a negative coefficient. Thus the bound tends to 0 as $n \to \infty$, and so the probability of having a map with distortion n^β tends to 0. ∎

4. Open questions and further directions

The phase transition at $\alpha = 1/2$ involves a fundamental change in the geometry of the percolation cluster. It is reasonable to expect that many measurements of the percolation cluster will exhibit a change at that point. One such measurement is the latency of embedding the cube in the percolated cube. This is motivated by the problem of estimating the loss of computational power in a network when some of the communication channels are faulty.

It follows from the definition of the latency of an embedding [7], that if the distortion is at most c then the latency is at most n^c. Thus we have:

Corollary 7. *If $\alpha < 1/2$, then a.a.s. the latency between H_n and $H_{n,p}$ is at most polynomial in n.*

It seems plausible that having a large distortion between the graphs implies that the latency is also large. In particular, we conjecture:

Conjecture 8. If $\alpha > 1/2$, then a.a.s. the latency between H_n and $H_{n,p}$ is at least e^{n^β} for some $\beta = \beta(\alpha)$.

An additional aspect of the phase transition at $\alpha = 1/2$ involves the routing complexity. The routing problem is to find a path in $H_{n,p}$ between two given vertices, preferably a short one. The routing problem may be defined under one of several models, the strictest of which involves starting at x and only being allowed to query edges incident on vertices that have already been reached from x. We refer to this as the local model.

For $p = n^{-\alpha}$ with $\alpha < 1/2$, it turns out that there is a polynomial time (in n) algorithm for routing in the local model, outputting a path of length $O(n)$ whenever x, y are in the same component. This is a consequence of the fact that distance between neighbouring vertices is typically bounded, and so to get from x to a nearby y only a ball of constant radius (and polynomial volume) needs to be explored. On the other hand when $\alpha > 1/2$ no polynomial algorithm exists. It turns out that in the intermediate regime there is a lower bound on the routing complexity of e^{n^β} for some $\beta > 0$. For more details on complexity of routing in the hypercube and in other scenarios see [4].

The goal of this paper is to point and establish a new qualitative phenomena. There are many open questions remaining. First, one could hope for a better understanding of the new transition, and of the two regimes it separates.

A first question, involves the sub-critical regime. Our proof above shows that the distortion is at least n^β, and it works for any $\beta < \frac{2\alpha-1}{4}$. However, the only clear upper bound on the distortion is of order n – the diameter of $H_{n,p}$. We believe that this is indeed the typical distortion.

Question 9. What is the typical value of the distortion from H_n to $H_{n,p}$ when $\alpha > 1/2$? Is it linear in n?

The super-critical regime $\alpha < 1/2$ also has some further questions. Specifically, our proof above shows that the distortion is at most $O((1/2-\alpha)^{-1})$. It is possible to argue along lines similar to our proof above that the distortion does diverge when $\alpha \to 1/2$. We do not have a matching lower bound on the distortion, but we believe this does give the correct rate.

Question 10. Is it true that as $\alpha \to 1/2$ from below, the typical distortion is $\theta((1/2-\alpha)^{-1})$?

At the critical point $\alpha = 1/2$, it would be interested to get a better understanding of the critical window and the transition itself, in the spirit of e.g. [6] where the transition at $\alpha = 1$ is analyzed in depth. In particular one may ask how big is the critical window?

Question 11. What happens if $p=an^{-1/2}$? Is there some a_c so that above a_c the distortion is constant while below it the distortion tends to infinity with n?

Finally, there is the fundamental question of characterizing those families of graphs where such an intermediate phase exists. Non-uniqueness of infinite percolation clusters in infinite graphs is a wide spread phenomena, closely related to non-amenability of the graphs. One possible extension would be to finite analogues of infinite graph for which it is known that $p_c < p_u < 1$, e.g. the product of a large cycle and a large girth graph, which will be a finite analogue of $T \times \mathbb{Z}$. And for large girth graphs the distortion between G and G_p will be proportional to the girth for any p bounded away from 1.

One of the first steps in development of such a theory would be to extend the relation between the typical percolation distance between neighbours in G, and the metric distortion between G and G_p. In particular, we ask:

Question 12. Suppose G is a vertex transitive graph, and that for some p, the typical percolation distance between neighbours in G is m. Under what conditions on G, n, p is the metric distortion from G to G_p of order m?

Acknowledgments. Thanks to Eran Ofek and Udi Wieder for useful discussions.

References

[1] M. Ajtai, J. Komlós and E. Szemerédi: Largest random component of a k-cube, *Combinatorica* **2(1)** (1982), 1–7.

[2] N. Alon and J. Spencer: *The Probabilistic Method*, John Wiley and Sons, New York, 1992.

[3] N. Alon, I. Benjamini and A. Stacey: Percolation on finite graphs and isoperimetric inequalities, *Ann. Probab.* **32(3A)** (2004), 1727–1745. arXiv:math.PR/0207112.

[4] O. Angel, I. Benjamini, E. Ofek and U. Wieder: Routing Complexity in Faulty Networks, *Rand. Str. and Algo.*, to appear (2007). http://dx.doi.org/10.1002/rsa.20163

[5] I. Benjamini and O. Schramm: Percolation beyond $\mathbb{Z}^d$, many questions and a few answers; *Elec. Comm. Prob.* **1(8)** (1996), 71–82 (electronic).

[6] C. Borgs, J. Chayes, R. van der Hofstad, G. Slade and J. Spencer: Random subgraphs of finite graphs, I. The scaling window under the triangle condition; *Rand. Str. and Algo.* **27(2)** (2005), 137–184.

[7] J. Hastad, T. Leighton and M. Newman: Reconfiguring a Hypercube in the Presence of Faults (Extended Abstract), in *Proc. of the 19th ACM Symp. on Theory of Computing*, 274–284, 1987.

[8] N. Linial: Finite Metric Spaces – Combinatorics, Geometry and Algorithms; in *Proc. of the Inter. Cong. of Math. III*, 573–586, Beijing, 2002.

[9] R. Lyons: Phase transitions on nonamenable graphs, *J. Math. Phys.* **41(3)** (2000), 1099–1126. (Probabilistic techniques in equilibrium and nonequilibrium statistical physics.)

Omer Angel

Department of Mathematics
University of Toronto
Toronto ON, M5S 2E4
Canada
angel@math.utoronto.ca

Itai Benjamini

Department of Mathematics
Weizmann Institute of Science
Rehovot 76100
Israel
itai.benjamini@weizmann.ac.il