Information Processing Letters 110 (2010) 356–359

Contents lists available at ScienceDirect

Information Processing Letters

www.elsevier.com/locate/ipl

On the connectivity threshold for general uniform metric spaces

Gady Kozma [a], Zvi Lotker [b,*], Gideon Stupp [c]

[a] Department of Mathematics, Weizmann Institute of Science, PO Box 26, Rehovot 76100, Israel
[b] Department of Communication Systems Engineering, Ben Gurion University, Beer Sheva 84105, Israel
[c] Cisco systems, 32 Hamelacha st, Netanya, Israel

ARTICLE INFO

Article history:
Received 8 December 2008
Received in revised form 20 January 2010
Accepted 24 February 2010
Available online 26 February 2010
Communicated by A. Tarlecki

Keywords:
Random networks
Sensor networks
Connectivity threshold
Fractal

ABSTRACT

Let μ be a measure supported on a compact connected subset of an Euclidean space, which satisfies a uniform d-dimensional decay of the volume of balls of the type

$$\alpha\delta^d \leqslant \mu\big(B(x,\delta)\big) \leqslant \beta\delta^d \tag{1}$$

where d is a fixed constant. We show that the maximal edge in the minimum spanning tree of n independent samples from μ is, with high probability $\approx (\frac{\log n}{n})^{1/d}$. While previous studies on the maximal edge of the minimum spanning tree attempted to obtain the exact asymptotic, we on the other hand are interested only on the asymptotic up to multiplication by a constant. This allows us to obtain a more general and simpler proof than previous ones.

1. Introduction

The laws governing the behavior of the minimum spanning tree (MST) on points in the unit disk and ball have been thoroughly researched and applied in the field of mobile computing; see [12] for a survey of the worst case. For the average case (or points taken randomly), extremely fine results have been obtained: see [4] for the central limit theorem for the total length, [2,1] for the dimension of a typical path, [3, Chapter 6] for an "objective" approach, [13] for intriguing simulation results, and [6] for efforts to prove them. See also the book [9] for the strongly related continuum percolation. The exact asymptotics of the length of the longest edge of the MST was studied in [10,11,7].

Although adequate for modeling mobile systems in man-made environments such as inside a room or a building, Euclidean geometry is perhaps too restrictive for modeling systems in natural settings, such as woods or rugged terrains. However, there has been much less research on MST behavior in other metric spaces and in particular on fractals. In [8], we studied the worst case problem for the total weighted MST length. Here, we switch to the average case and are interested in the length of the longest edge, which, by the greedy algorithm, is the same as the connectivity threshold, i.e., the minimal number r such that the graph in which two points are connected if and only if their metric distance is $\leqslant r$, is connected. In the setting of a ball in $\mathbb{R}^d$ this is known to be, with high probability

$$\approx \left(\frac{\log n}{n}\right)^{1/d}$$

where $\approx$ means that the ratio of the two quantities is bounded between two absolute constants. We wish to extend this result to fractal sets.

Clearly, to get any kind of estimate, one has to assume that the fractal is connected. Further, it is clear that some kind of regularity is needed. To see why regularity is needed, it might be instructive to consider the following example: in $\mathbb{R}^2$ take the set $F = \bigcup_{i=1}^{\infty}(A_i \cup B_i)$, consisting of a set of "thick" vertical slabs $A_i = [\frac{1}{2^i}, \frac{1}{(2^{i-1})}] \times [0,1]$ connected by "thin" horizontal bridges $B_i = [\frac{1}{(2^{i+1})}, \frac{1}{2^i}] \times [0, 3^{-i}]$; see Fig. 1. Take the normalized Lebesgue mea-

* Corresponding author.
E-mail addresses: gady.kozma@weizmann.ac.il (G. Kozma), zvilo@cse.bgu.ac.il (Z. Lotker), gstupp@cisco.com (G. Stupp).

0020-0190/$ – see front matter
doi:10.1016/j.ipl.2010.02.015

Fig. 1. The set F.

sure on this set. Now take n random points and connect them by their MST. Since the bridges become thin very fast, there will be no points in the very thin bridges that connect the slabs starting from $[1/2\log n, 1/(2\log n - 1)] \times [0,1]$. Thus, the MST will contain edges that are $\geqslant c/\log n$ long, despite the fact that this set has Hausdorff dimension 2 and is, in fact, a monofractal (a degenerate multifractal spectrum), which indicates strong regularity in some sense.

Definition 1.1. Let d be a fixed constant. The metric probability space (F, ρ, μ) is *semi-uniform of dimension d* if there exist some numbers δ_0 and $0 < \alpha \leqslant \beta < \infty$ such that for every point $x \in F$ and for every $0 < \delta < \delta_0$,

$$\alpha\delta^d \leqslant \mu\big(B(x,\delta)\big) \leqslant \beta\delta^d$$

where $B(x, r)$ is a ball of radius r centered at x in the metric ρ. We call δ_0, α and β the *parameters* of the metric probability space.

Examples of such a semi-uniform space are the Cantor's dust and the Sierpinski's carpet, endowed with the metric of their embedding in $\mathbb{R}^n$ and natural probability measures [5]. We show that for such probability spaces the classical asymptotic laws of maximal MST edge still hold with fractional powers, if the space is connected. For example, in the case of Sierpinski's carpet, with dimension $d = \log 8/\log 3$, the longest edge of the MST is $\approx (\frac{\log n}{n})^{1/d}$ with probability $1-\epsilon$ for all $\epsilon > 0$. Formally, the statement is

Theorem 1.1. *Let F be a compact connected subset of $\mathbb{R}^k$ and let μ be a semi-uniform measure of dimension d on F. Then, there exist two constants $C > c > 0$, such that for all $\epsilon > 0$ and for any sufficiently large m (depending on ϵ and d), if $X^m = X_1, \ldots, X_m$ are independent samples from μ, then:*

1. *If $r > C(\frac{\log m}{m})^{1/d}$ then the set $\bigcup B(X^m, r)$ is connected with probability at least $1-\epsilon$.*
2. *If $r < c(\frac{\log m}{m})^{1/d}$ then the set $\bigcup B(X^m, r)$ is connected with probability at most ϵ.*

Here and below we will use the convention that c and C denote constants whose value might change from formula to formula and even inside the same formula. c and C might depend on d, α and β. c will usually denote constants that are "small enough" and C, constants which are "large enough".

2. Proof

To prove the first part of the theorem let $\delta = (\frac{2\log m}{\alpha m})^{1/d}$ and let $\{B(p_i, \delta)\}_{i=1}^{N(\delta)}$ be a maximal set of disjoint balls with centers $p_i \in F$. Then for each X_i and j,

$$\mathbf{Pr}\big[X_i \in B(p_j, \delta)\big] \geqslant 2\frac{\log m}{m}.$$

Since the X^m are i.i.d. it follows that $\mathbf{Pr}$ is a product measure. Denote by $\mathbf{A}_\delta$ the event that there is at least one of X_i in each of $B(p_j, \delta)$.

Lemma 2.1. *For large enough m, event $\mathbf{A}_\delta$ occurs with probability at least $1-\epsilon$.*

Proof. Had all the measures of all the balls been equal, we could immediately have used the coupon collector. Where the bins are the $B(p_j, \delta)$, and the random "coupons" are the X_i throw. However, the measures of the balls are only equal up to a constant factor. Still, for all p_j,

$$2\frac{\log m}{m} \leqslant \mathbf{Pr}\big[B(p_j, \delta)\big].$$

Take $m/(2\log m)$ equiprobable balls. Denote those balls by $B(p_i, \delta_i')$. Note that $\mu(B(p_i, \delta_i')) = 2\frac{\log m}{m}$. Define the event $\mathbf{A}'$ to have at least one random point X_i in each $B(p_j, \delta_j')$ after m trials. Note that this is possible since semi-uniform spaces do not have atoms. Clearly since we only reduced the volume, it follows that:

$$\mathbf{Pr}[\mathbf{A}_\delta] \geqslant \mathbf{Pr}\big[\mathbf{A}'\big].$$

Now we can bound $\mathbf{Pr}[A']$ using coupon collector arguments, where the number of balls is m and the number of bins is $\lceil \frac{m}{2\log m} \rceil$. Fix i and consider the probability of not having points in $B(p_i, \delta_i')$ after we had thrown m balls

$$\mathbf{Pr}\big[B\big(p_i, \delta_i'\big) \cap X^m = \emptyset\big] = \left(1 - 2\frac{\log m}{m}\right)^m \leqslant \frac{1}{m^2}$$

for sufficiently large m. Now we use union bound over i,

$$\mathbf{Pr}\big[\mathbf{A}'\big] = 1 - \mathbf{Pr}\big[\overline{\mathbf{A}'}\big] \geqslant 1 - \frac{m}{2\log m}\left(1 - \frac{2\log m}{m}\right)^m$$
$$\geqslant 1 - \frac{m}{\log m} \cdot \frac{1}{m^2} \geqslant 1 - \epsilon$$

and the lemma follows. □

Lemma 2.2. *Conditioned on $\mathbf{A}_\delta$, $\bigcup B(X^m, 3\delta)$ is connected.*

Proof. As is well known (and easy to see), the maximality of the family $B(p_i, \delta)$ implies that the family $\bigcup B(p_i, 2\delta)$ is a cover of F. Since it is given that the event $\mathbf{A}_\delta$, happens it follows that the set $\bigcup B(X_i, 3\delta)$ covers the set $\bigcup B(p_i, 2\delta)$ and therefore it also covers the set F. Now the lemma follows directly from the definition of connectivity. Suppose by contradiction that $\bigcup B(X_i, 3\delta)$ is separated, i.e., there exist two open sets U, V s.t. $U \cup V = B(X_i, 3\delta)$ and $U \cap V = \emptyset$. Using U, V, it follows that F is also separated, leading to a contradiction. □