Robust Entropy Estimation Strategies Based on Edge Weighted Random Graphs (with corrections)

Alfred O. Hero[a] and Olivier Michel [b]

[a]Dept. EECS, University of Michigan, Ann Arbor, MI, 48109-2122 U.S.A.

[b]Ecole Normale Supérieure, 46 allee d'Italie 69394, Lyon, FRANCE

ABSTRACT

In this paper we treat the problem of robust entropy estimation given a multidimensional random sample from an unknown distribution. In particular, we consider estimation of the Rényi entropy of fractional order which is insensitive to outliers, e.g. high variance contaminating distributions, using the k-point minimal spanning tree (k-MST). A greedy algorithm for approximating the NP-hard problem of computing the k-minimal spanning tree is given which is a generalization of the potential function partitioning method of Ravi etal.[1] The basis for our approach is an asymptotic theorem establishing that the log of the overall length or weight of the greedy approximation is a strongly consistent estimator of the Rényi entropy. Quantitative robustness of the estimator to outliers is established using Hampel's method of influence functions.[2] The structure of the influence function indicates that the k-MST is a natural extension of the one dimensional α-trimmed mean for multi-dimensional data.

Keywords: pattern recognition, entropy estimation, random graph theory, spatial processes, mutual information

1. INTRODUCTION

In Hero and Michel[3] the asymptotic behavior of a general class of minimal k-point graphs over a set of random points in $\mathbb{R}^d$ was shown to lead to consistent estimates of the Rényi entropy of the underlying distribution. This paper provides an overview of these results with emphasis on motivating examples, applications, and simulations. All of the results presented in Hero and Michel[3] are applicable to entropy estimation using the weight function of any minimal graph which satisfies the so-called *quasi-additive property* of Redmond and Yukich.[4] This includes graphs such as those arising from the Traveling Salesman Problem, the Steiner Tree, the Minimal Spanning Tree, or the Minimal Pairwise Matching Problem. Here we concentrate on the the minimal spanning tree (MST) due to its computational advantages for entropy estimation applications.

Assume that we are given a set $\mathcal{X}_n = \{x_1, \ldots, x_n\}$ of n points in $\mathbb{R}^d$. Fix k and denote by $\mathcal{X}_{n,k}$ a k-point subset of $\mathcal{X}_n$, $0 < k \leq n$. The elements of the subset $\mathcal{X}_{n,k}$ are distinct and there are $\binom{n}{k}$ possible k-point subsets of $\mathcal{X}_n$. A spanning tree is connected acyclic graph over $\mathcal{X}_n$ defined as a set of edges $\{e_i\}$ that connects all n points such that there are no paths in the graph that lead back to any given point. For a given edge weight exponent γ the minimal spanning tree (MST) is the spanning tree which minimizes the total edge weight $L(\mathcal{X}_n) = \sum_e |e|^\gamma$ of the graph. The MST arises for $d = 2$ in VLSI circuit layout and network provisioning,[5,6] two sample matching,[7] pattern recognition,[8] clustering,[9] nonparametric regression[10] and testing for randomness.[11] The MST can be constructed in time polynomial in n.

The minimal k-point spanning tree (k-MST) problem is to find the subset of points $\mathcal{X}_{n,k}$ and the set of edges connecting these points such that the resultant tree has minimum total weight $L(\mathcal{X}_{n,k})$. This problem arises in competitive bidding for network routing contracts and other combinatorial optimization problems.[1] The planar k-MST problem was shown to be NP-complete by Zelikovsky and Lozevanu[12] and Ravi, Sundaram, Marathe, Rosenkrantz and Ravi.[1] Ravi etal proposed a polynomial time greedy approximation algorithm for the planar k-MST with approximation ratio $O(k^{1/4})$. In Hero and Michel[3] we gave a polynomial time approximation algorithm for the general d dimensional problem with approximation ratio $O\left(k^{[(d-1)/d]^2}\right)$.

Other author information: (Send correspondence to A. Hero)
A. Hero: E-mail: hero@eecs.umich.edu
O. Michel: E-mail: omichel@ens-lyon.fr

Furthermore, as shown in Hero and Michel,[3] when the set of points $\mathcal{X}_n$ is a random sample from a continuous density f, the log-length of the k-MST is a robust strongly consistent estimator of the Rényi entropy R_ν of of the density

$$R_\nu(f) = \frac{1}{1-\nu} \log \int_{\mathbb{R}^d} f^\nu(x)dx$$

where the order $\nu \in (0,1)$ of the Rényi entropy is determined by the edge weight function and the dimension d.

Entropy estimation has been of interest for pattern analysis, process complexity assessment, model identification, tests of distributions, and other applications where invariance to scale, translation and other invertible transformations is desired in the discriminant.[13–15] Another application is in vector quantization where Rényi entropy is related to asymptotic quantizer distortion via the Panter-Dite factor and Bennett's integral.[16,17] Among the many other applications of entropy estimation are: estimation of Lyapounov exponents in non-linear models,[18,19] multi-modality image registration using mutual information matching criteria,[20] stopping criteria for regression and classification trees,[21] and other general entropy estimation problems[22,13]

This paper provides an overview of the k-MST approach to entropy estimation drawing heavily on results established in Hero and Michel.[3] We start out by illustrating the MST and its relation to entropy estimation and non-linear cluster analysis in the context of some numerical examples. These examples are used to point out the nonrobustness of the MST to outliers, and to introduce the k-MST as a robust alternative and the k-dependent length k-MST curve as a discriminant for selecting the number of outliers to reject. Then we describe the greedy algorithm for approximating the k-MST in detail and discuss two asymptotic theorems results proven in Hero and Michel.[3] Finally we use influence functions to establish quantitative robustness of the greedy k-MST to outliers and draw an analogy to α-trimmed mean estimators for one dimensional problems.

2. ENTROPY ESTIMATION VIA MINIMAL GRAPHS

2.1. A Motivating MST Example

In Figures **??** and 2 we show an example which motivates the use of minimal spanning trees as entropy discriminants between two different distributions. The two columns of Figure **??** correspond to two different distributions on the unit square $[0,1]^2$ – the left column corresponds to a uniform density while the right column corresponds to a triangular density with a maximum at the point $(0.5, 0.5)$. The top row of Figure **??** presents the results of a single experiment generating 100 random samples from the uniform and triangular distributions, respectively. The middle row presents the corresponding MST's for each of these realizations constructed by minimizing the sum of the edge lengths $\sum_e |e|$ in the tree. Note that for this experiment the overall length of the MST for the uniform sample is larger than that of the more concentrated triangular sample. The mean length of the MST for each of the distributions is shown in the bottom row of Figure **??**, computed on the basis of a large number of repeated independent experiments of the type illustrated in the first row. (5% confidence intervals are shown). Note that for large n the mean length curves appear to increase with sublinear rates with rate constants that depend on the underlying distribution of the random sample. The left panel of Figure 2 shows a more direct comparison of these two mean length curves plotted simultaneously as a function of n. The right panel shows the length curves normalized by $\sqrt{n}$ and transformed by $2\log(\cdot)$. It is evident that for both the unifom and the triangular distributions the normalized and transformed length of the MST converges to two different constant levels. Furthermore, the asymptote for the uniform distribution is significantly larger than that for the triangular distribution. In fact, as was remarked in Hero and Michel,[23] the difference between the asymptotes is equal to the difference between the Rényi entropies of order $\nu = (d-\gamma)/d = (2-1)/2 = 1/2$ of the respective distributions.

2.2. Robustification of the MST via the k-MST

In the previous section we illustrated that the MST provides a consistent estimator of Rényi entropy. Here we illustrate the sensitivity of the MST to outlier contamination and introduce a robustification via the k-MST.

Let the underlying density be the mixture

$$f(x) = (1-\epsilon)f_1(x) + \epsilon f_0(x), \qquad 0 < \epsilon < 1 \tag{1}$$

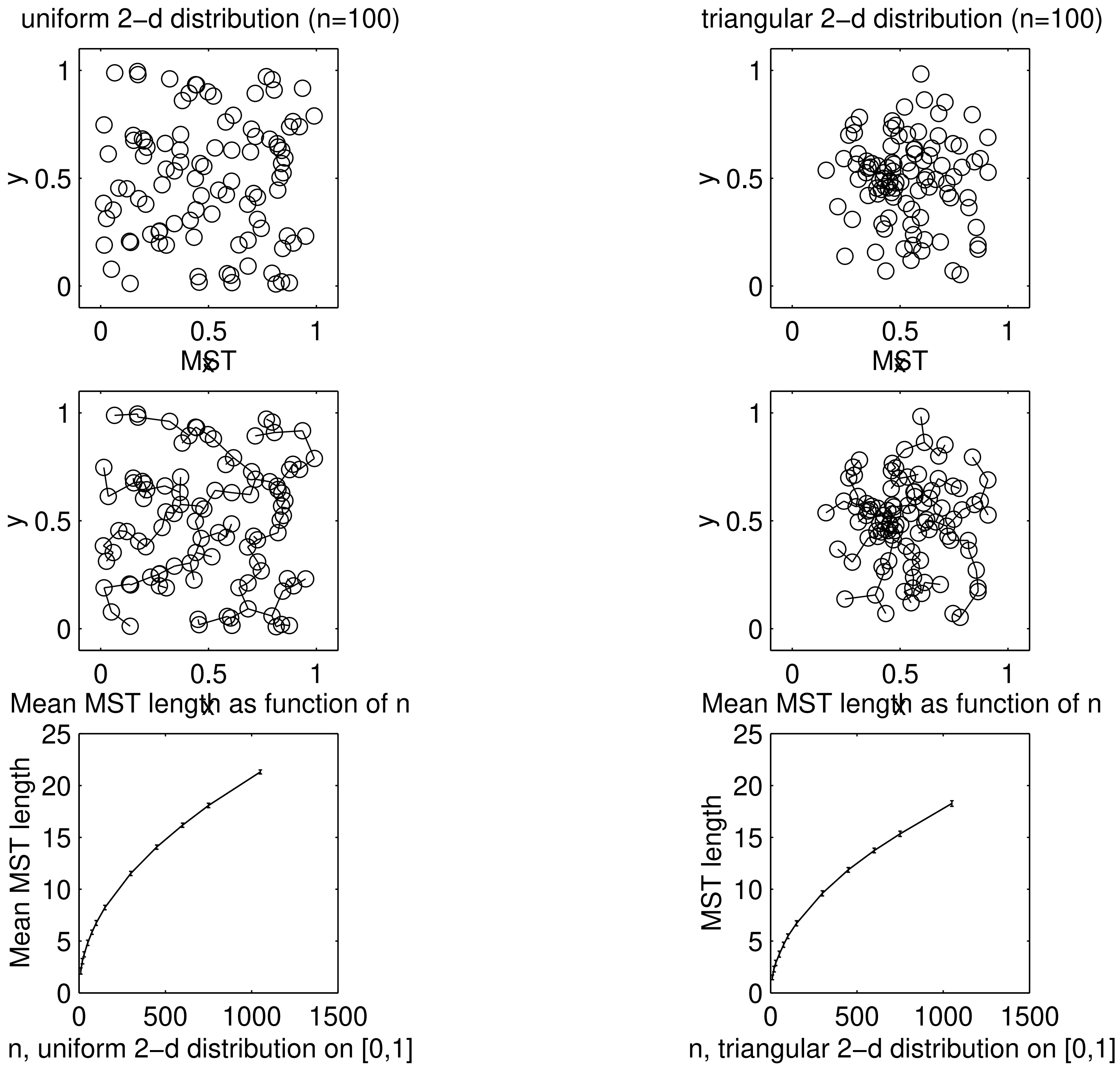

Figure 1. *2D Triangular vs. Uniform sample study.*

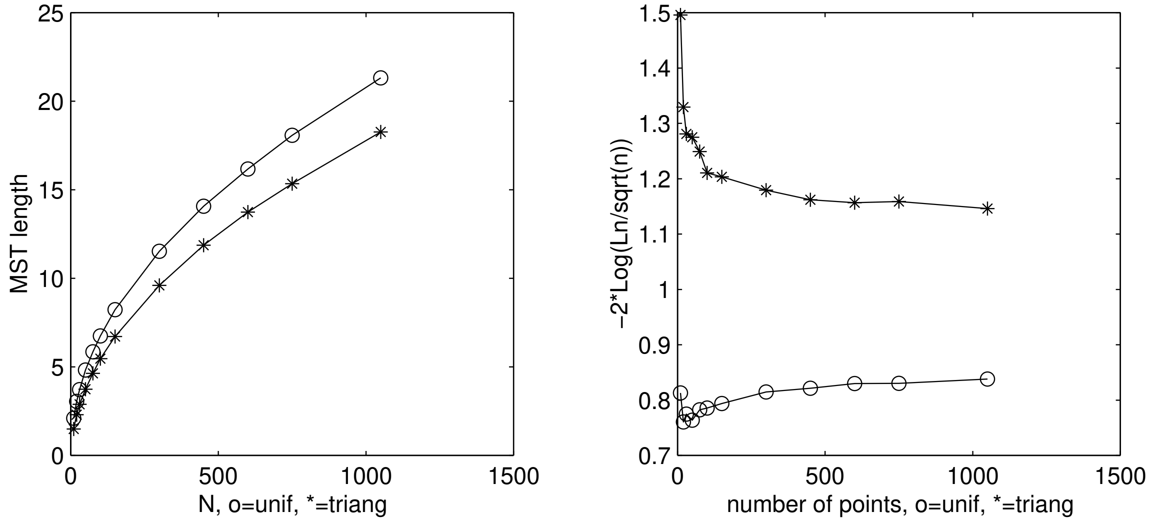

Figure 2. *MST and log MST weights as function of number of samples for 2D uniform vs. triangular study.*

where f_1 can be interpreted as a density of interest (generating the signal) and f_0 is a contaminating distribution (generating noise or outliers). In Figure 3 the left panel shows 50 realizations from a density f_1 of the form

$$f_1(x) = ce^{-\frac{1}{2}225(\|x-[0.4,0.4]\|-0.25)^2}$$

where c is a normalizing constant, $\|x\|^2 = x_1^2 + x_2^2$ is the magnitude squared of $x = (x_1, x_2)$. The constant contours of this density are circles for which the maximum contour is a circle of radius 0.25 and center $[0.4, 0.4]$ and the other contours specify an annulus. Hence we call this an annulus density. In the right panel of Figure 3 the same 50 realizations from f_1 are shown contaminated with 50 samples from the uniform density f_0. This corresponds to the rather severe case of $\epsilon = 0.5$ in equation (1). The bottom row of Figure 3 shows the MST's for each of these cases. Notice that while the MST on the left panel captures the shape of the uncontaminated density, and its length could be used as a reliable entropy estimator, the MST on the right panel is severly influenced by the addition of the uniform noise. Thus the MST length function is not robust to outliers.

A solution to this lack of robustness was presented in Banks etal[10] in the context of non-parametric non-linear regression where the authors proposed a robustification obtained by pruning the $n-k$ largest edges from the MST and reconnecting any remaining isolated subtrees. The resulting tree is not typically an optimal tree passing through the remaining points. Here we propose a different robustification based on the optimal k point MST which, as will be shown in the next section, has provable convergence properties and for which quantitative robustness can be established.

In Figure 4 we illustrate the application of the k-MST to the same experiment as shown in Figure 3 for the MST. It is evident form the figure that as the number of points eliminated by the k-MST increases from 1 to 2 to 38

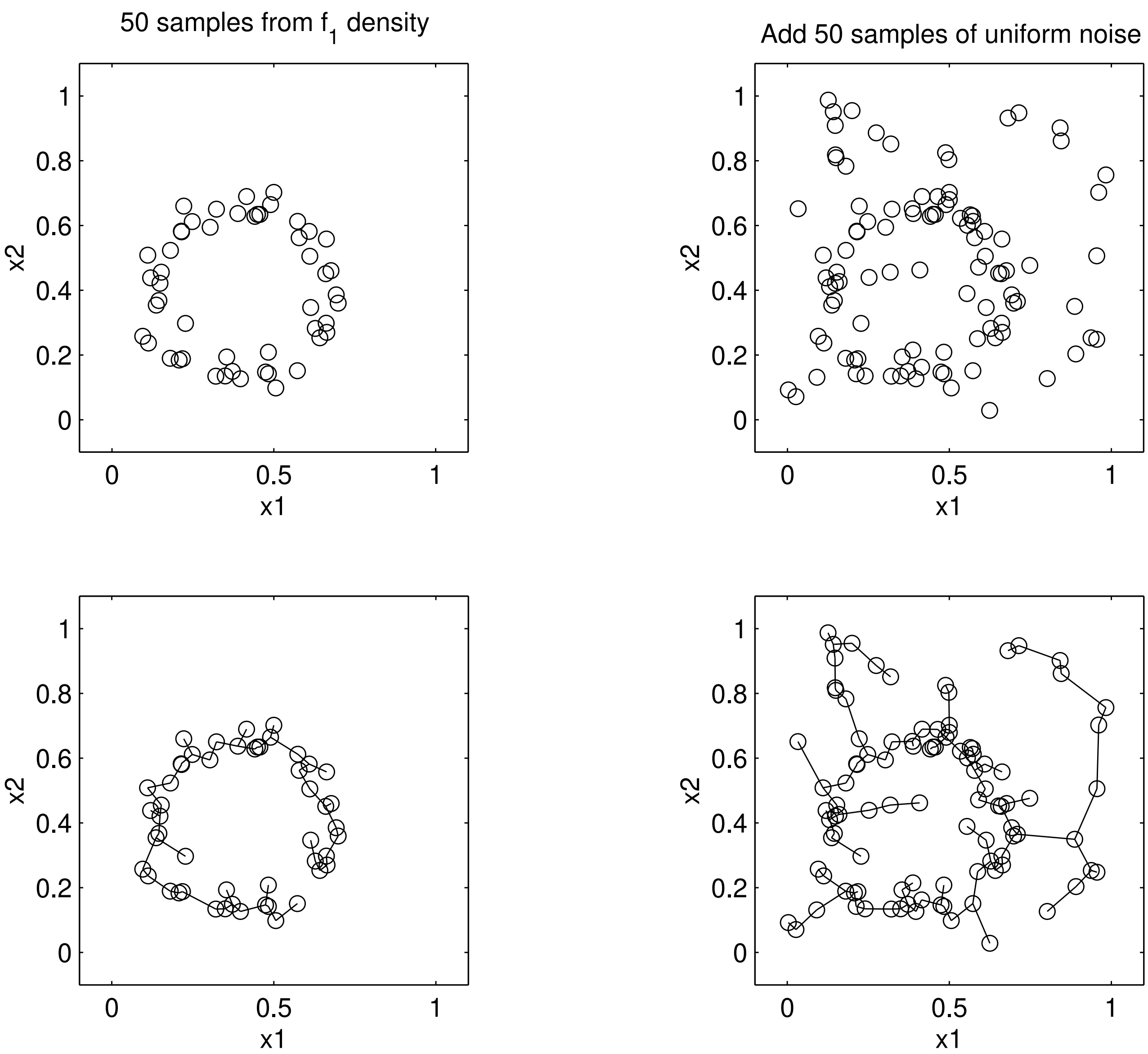

Figure 3. *1st row: 2D torus density with and without the addition of uniform "outliers." 2nd row: corresponding MST's.*

the k-MST rejects an increasing number of outliers from the contaminating density. Indeed for the case of $k = 62$ (38 outliers rejected) the k-MST appears to have almost completely recovered the MST for the signal alone annular distribution. However, as the number of rejected points increases beyond 38 to 25 the k-MST begins eliminating points which come from the desired annular distribution. The key to a practical k-MST robustification algorithm will be accurate detection of the correct number of points to reject.

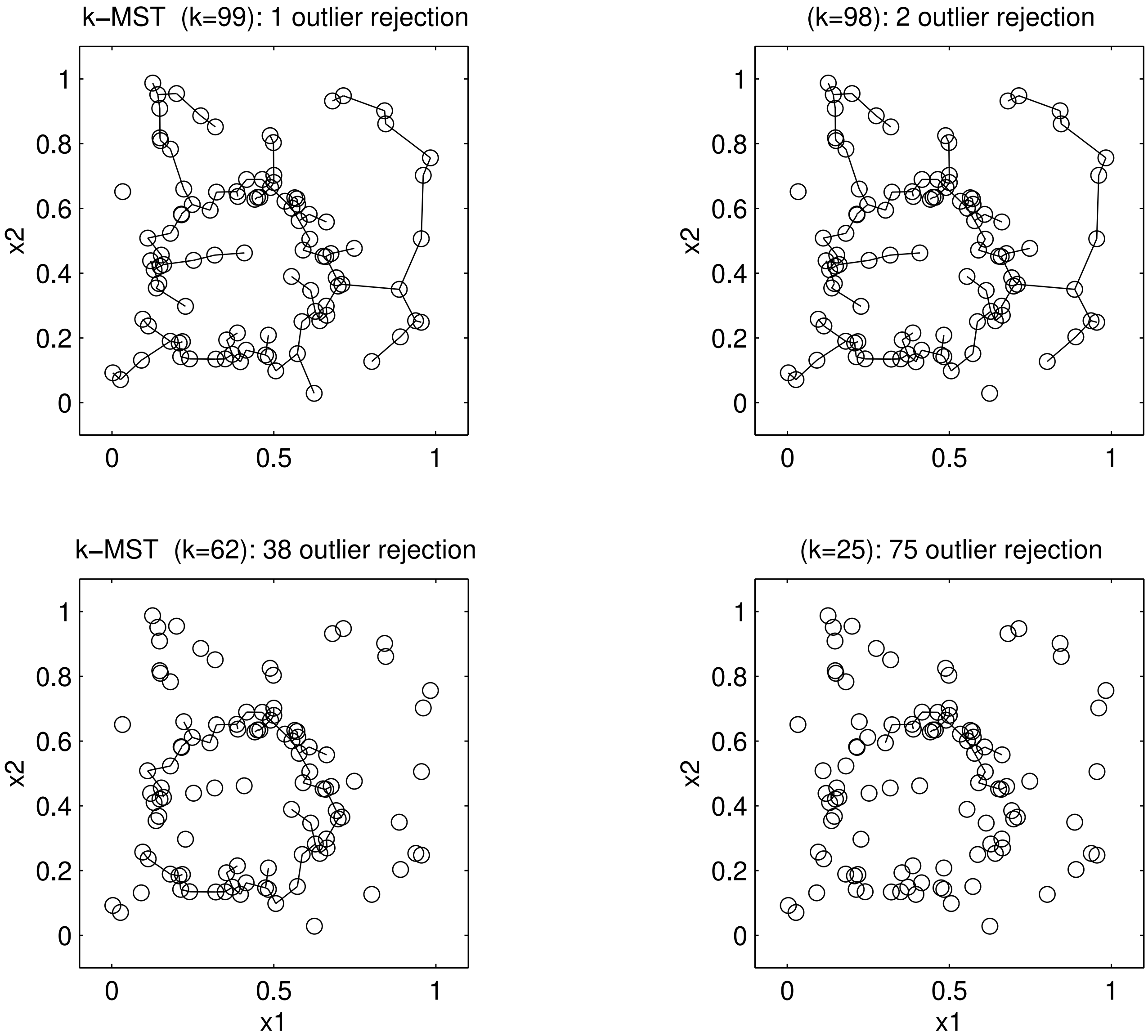

Figure 4. *MST for 2D torus density with and without the addition of uniform "outliers".*

A natural detection criterion can be constructed on the basis of the k-MST length curve $L(\mathcal{X}_{n,k})$ plotted as a function of the number of points rejected $n-k$. In Figure 5 this curve is plotted for a realization, shown in left panel, from a uniform density on $[0, 15] \times [0, 15]$. Observe that the curve appears to decrease more or less linearly as $n-k$ increases. We give theory in the next section that establishes that this is always the case for uniform densities. On the other hand, in Figure 6 the same curve is plotted based on the noise contaminated sample from an annulus density. Note that this latter curve appears to separate into two piecewise linear segments with a break at approximately $n - k = 40$. If we can reliably detect the breakpoint, or the knee, of the k-MST length curve then we can reliably implement to k-MST as a robust entropy estimator.

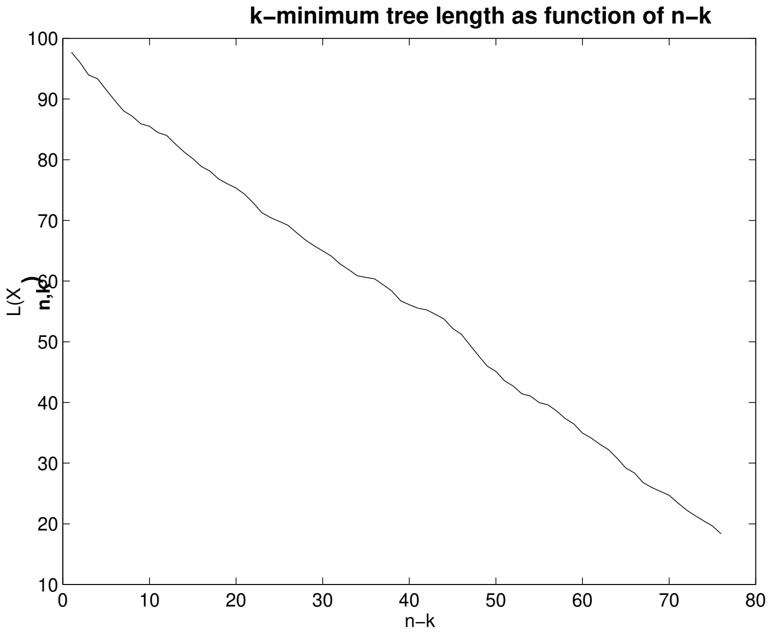

Figure 5. *k-MST curve for 2D uniform density on $[0, 15] \times [0, 15]$ is apparently linear.*

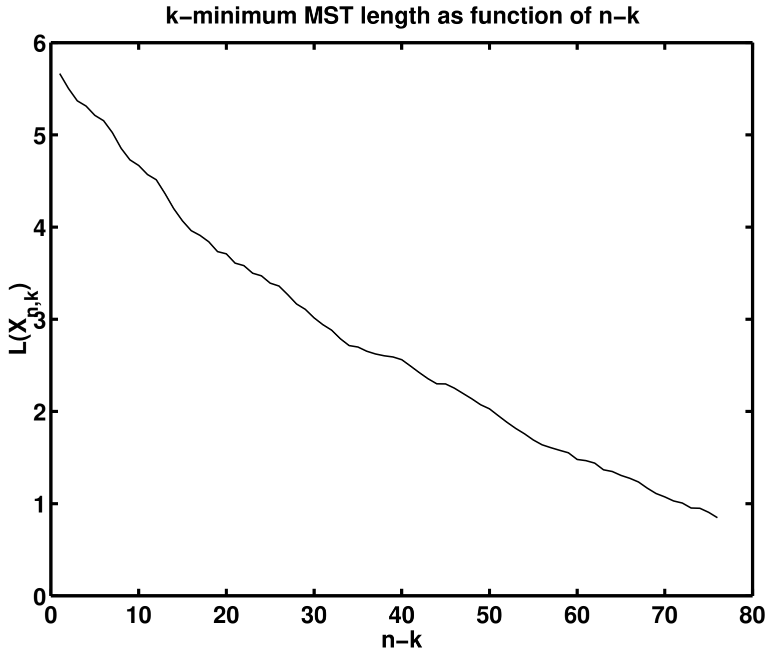

Figure 6. *k-MST curve for 2D annulus density with addition of uniform "outliers" has a knee in the vicinity of* $n - k = 38$

3. GREEDY ALGORITHM FOR K-MST IN $\mathbb{R}^D$

The greedy algorithm for approximating the k-MST is implemented in three steps: 0) a positive integer m is specified; 1) the user specifies a uniform partition $\mathcal{Q}^m$ of $[0,1]^d$ having m^d cells Q_i of resolution $1/m$; 2) the algorithm find the smallest subset $B_k^m = \cup_i Q_i$ of partition elements containing at least k points; 3) on this smallest subset the algorithm selects the k points $\mathcal{X}_{n,k}$ out of this subset which minimize $L(\mathcal{X}_{n,k})$. Stage 3 requires finding a k-point minimal graph on a much reduced set of points, which is typically only slightly larger than k if m is suitably chosen, which can be performed in polynomial time.

The smallest subset mentioned in Stage 2 of the algorithm is not unique. Figures 7 and 8 show an example for which $m = 5$, $k = 17$ for which there are two possible smallest subsets, in this case both contain 18 points.

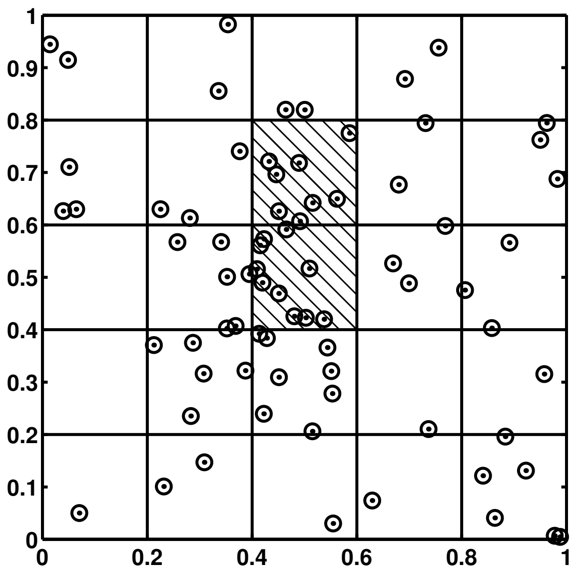

Figure 7. *A sample of* 75 *points from the mixture density* $f(x) = 0.25 f_1(x) + 0.75 f_o(x)$ *where* f_o *is a uniform density over* $[0,1]^2$ *and* f_1 *is a bivariate Gaussian density with mean* $(1/2, 1/2)$ *and diagonal covariance* diag(0.01). *A smallest subset* B_k^m *is the union of the two cross hatched cells shown for the case of* $m = 5$ *and* $k = 17$.

Similarly to Ravi etal[24,1] we specify a small subset by the following greedy algorithm: i) find a reindexing $\{Q_{(i)}\}_{i=1}^{m^d}$ of the cells in $[0,1]^d$ ranked in decreasing order of the number of contained points, $\mathrm{card}(\mathcal{X}_n \cap Q_{(1)}) \geq \ldots \geq \mathrm{card}(\mathcal{X}_n \cap Q_{(m^d)})$ (if there are equalities arrange these in lexicographical order); ii) select the subset specified in Stage 2 by the recursion:

Greedy Subset Selection Algorithm

Intialize: $B = \phi$, $j = 1$

Do until $\mathrm{card}\{\mathcal{X}_n \cap B\} \geq k$

$B = B \cup Q_{(j)}$

End $j = j + 1$

At termination of the algorithm $j = \tilde{q} \leq m^d$ and we have a minimal subset $B_{\lfloor \alpha n \rfloor}^m \stackrel{\mathrm{def}}{=} B = \cup_{i=1}^{\tilde{q}} Q_{(i)}$ containing at least k points.

We prove a variant of the following theorem in Hero and Michel.[3] Here $1 - \alpha$ is the proportion of points that the k-MST rejects, i.e. $k = \alpha n$.

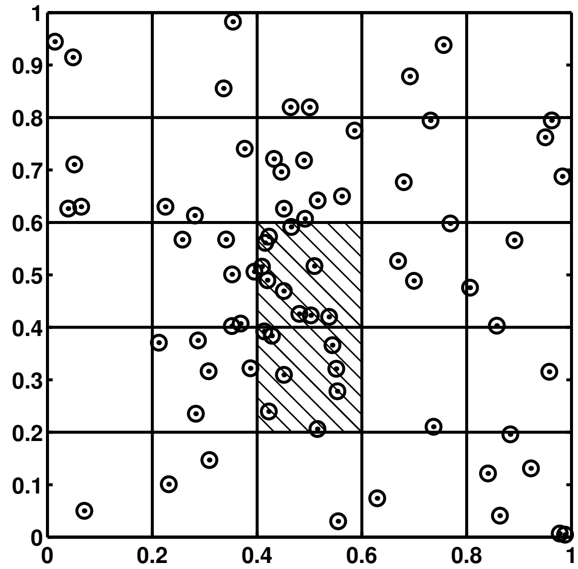

Figure 8. *Another smallest subset B_k^m containing at least $k = 17$ points for the mixture sample shown in Fig 7.*

THEOREM 1. *Fix $\alpha \in [0,1]$, $\gamma \in (0,d)$. Let $f^{(d-\gamma)/d}$ be of bounded variation over $[0,1]^d$ and denote by v its total variation over $[0,1]^d$. Then, the total edge weight $L(\mathcal{X}_{n,k})$ of a k-point graph constructed by the resolution-$1/m$ greedy algorithm satisfies*

$$\limsup_{n\to\infty} \left| L(\mathcal{X}_{n,\lfloor \alpha n \rfloor})/n^{(d-\gamma)/d} - \beta_{L,\gamma} \int_{A_\alpha^m} f^{(d-\gamma)/d}(x)dx \right| < \delta, \qquad (a.s.), \tag{2}$$

where A_α^m is the minimum volume set formed from the resolution $1/m$ partition elements satisfying: $P(A_\alpha^m) = \int_{A_\alpha^m} f(x)dx \geq \alpha$, and

$$\delta \quad = \quad (2\beta_{L,\gamma} m^{-d} + C_3 m^{(\gamma-d)})v = O(m^{\gamma-d}),$$

for constants $\beta_{L,\gamma}$ and C_3 that can be computed without knowledge of $f(x)$.

The significance of the theorem is described in detail in Hero and Michel.[3] For the moment we simply want to draw the reader's attention to the relation between the theorem and entropy estimation.

First we note that as n, m go to infinity Theorem 1 implies that $L(\mathcal{X}_{n,\lfloor \alpha n \rfloor})/n^{(d-\gamma)/d}\ 1/\beta_{L,\gamma}$ convereges (a.s.) to:

$$\begin{aligned} \lim_{m\to\infty} \int_{A_\alpha^m} f^{(d-\gamma)/d}(x)dx &= \inf_{A\in\mathcal{B}:P(A)\geq\alpha} \int_A f^{(d-\gamma)/d}(x)dx \\ &= \inf_{A\in\mathcal{B}:P(A)=\alpha} \int_A f^{(d-\gamma)/d}(x)dx. \end{aligned} \tag{3}$$

where $\mathcal{B}$ is the class of Borel sets.

Next, for any Borel set A in $[0,1]^d$ having $P(A) > 0$ define the conditional density $f(x|A) = f(x)/P(A)I_A(x)$ where $I_A(x)$ is the indicator function of A. The Rényi entropy of $f(x|A)$ of order $\nu \in (0,1)$ is defined as

$$R_\nu(f|A) = \frac{1}{1-\nu} \log \int f^\nu(x|A)dx. \tag{4}$$

This is also called the conditional Rényi entropy given A. Let A_o be the probability-at-least-α Borel subset of $[0,1]^d$ which minimizes $R_\nu(f|A)$

$$R_\nu(f|A_o) = \inf_{\{A\in\mathcal{B}:P(A)\geq\alpha\}} R_\nu(f|A). \tag{5}$$

For $\nu = (d-\gamma)/d$ define the following function of $L(\mathcal{X}_{n,\lfloor\alpha n\rfloor})$

$$\hat{R}_\nu \stackrel{\text{def}}{=} \frac{1}{1-\nu}\left(\log L(\mathcal{X}_{n,\lfloor\alpha n\rfloor})/(\lfloor\alpha n\rfloor)^\nu - \log\beta_{L,\gamma}\right) \tag{6}$$

Then we have the following main result:

THEOREM 2. *Under the assumptions of Theorem 1 $\hat{R}_\nu$ is a strongly consistent estimator of the minimum conditional Rényi entropy $R_\nu(f|A_o)$ of order $\nu\in(0,1)$ as $m,n\to\infty$.*

Some important implications of these Theorems are:

1. The k-MST function $\hat{R}_\nu$ has very desirable asymptotic properties as an entropy estimator including unbiasedness and vanishing variance.
2. The conditional entropy of a mixture $f = (1-\epsilon)f_1 + \epsilon f_0$ is approximately equal to the uncodnitional entropy of f_1 for small ϵ. Thus the k-MST entropy estimator is robust to outliers.
3. The constant $\beta_{L,\gamma}$ does not need to be computed if only relative entropy is of interest, e.g. in signal classification problems.
4. Given a user-specified maximum tolerated approximation error ϵ, and an upper bound $\overline{v}$ on the total variation of the underlying p.d.f.'s f , Theorem 1 can be manipulated to give a selection rule for choosing the required partition resolution
$$1/m \approx \frac{\epsilon}{(2+C_3)\overline{v}}.$$
5. It can be argued on the basis of Theorems 1 and 2 (see Hero and Michel[3]) that estimates of Rényi entropy of lower orders ($\nu < 1/d$) converge faster than estimates of higher orders.

4. INFLUENCE FUNCTION FOR ENTROPY ESTIMATOR

Influence functions have long been used to study quantitative robustness of estimators to outliers and other contaminating densities.[2] These functions provide a quantitative measure of outlier sensitivity of an estimator. An unbounded influence curve implies that the effect of an outlier on the estimator can be very severe. Robust estimators, such as the trimmed mean estimator which rejects observations which exceed a given sample quantile, have bounded influence curves. Here we outline the results of Hero and Michel[3] where we established that the influence function of the k-MST entropy estimator is bounded.

Let P_n be the empirical distribution function of the n samples $\mathcal{X}_n = \{x_1,\ldots,x_n\}$

$$P_n(A) \stackrel{\text{def}}{=} \frac{1}{n}\int_A I_{x_i}(x)dx$$

for arbitrary Borel set A. For any statistic $T_n = T(P_n)$ converging a.s. to $T = T(P)$ the influence function (called an influence curve for one dimensional samples x_i) is defined as[25]

$$\mathrm{IC}(\mathrm{x_o}) = \lim_{\mathrm{s}\to 0}\frac{\mathrm{T}(\,(1-\mathrm{s})\mathrm{P} + \mathrm{s}\delta_{\mathrm{x_o}}) - \mathrm{T}(\mathrm{P})}{\mathrm{s}}. \tag{7}$$

where δ_{x_o} is a concentrated distribution centered at $x_o\in\mathbb{R}^d$ and $s\in[0,1]$. For small s, $(1-s)P + s\delta_{x_o}$ is interpreted as a perturbed distribution resulting from exchanging sn of the n samples x_i from distribution P with sn samples from the concentrated distribution δ_{x_o}. Thus IC(x_o) can be used to probe the asymptotic sensitivity of the estimator T_n to localized perturbations of P.

The influence function for the k-MST weight function is computed in Hero and Michel[3] by identifying $T_n = L(\mathcal{X}_{n,\lfloor\alpha n\rfloor})/(\lfloor\alpha n\rfloor)^{(d-\gamma)/d}\ 1/\beta_{L,\gamma}$ and invoking Theorem 2 which asserts that T_n converges a.s. to the integral $T(P) = \int f^\nu(x|A_o)dx$. After some manipulations we obtain the following form for the influence function for the normalized k-MST weight function T_n

$$\mathrm{IC} = \begin{cases} \frac{\alpha-1}{\alpha^\nu}\zeta(\alpha) - \nu\int f^\nu(x|A_o)dx + \frac{\nu}{\alpha}f^{\nu-1}(x_o|A_o), & x_o \in A_o \\ \frac{\alpha}{\alpha^\nu}\zeta(\alpha) - \nu\int f^\nu(x|A_o)dx, & x_o \notin A_o \end{cases} \tag{8}$$

where ζ is a non-negative function and $A_o = \{x : f(x) \geq \eta\}$ is the entropy mimimizing set of probability α.

Note that when the rejection proportion $1-\alpha$ is greater than zero and x_o is outside of the set A_o: IC is bounded.

We illustrate this in Figure 9 where IC is plotted as a function of $x_o \in \mathbb{R}^2$ for the case of the bivariate Gaussian distribution. Two cases are shown, the figure on the left is the influence function for $\alpha = 1$, i.e., for the minimal graph spanning all points (labeled MST), and the figure on the right is for $\alpha = 0.8$, i.e. for the minimal k-point graph (labeled k-MST) spanning only 80% of the n points. Note that, as expected, the influence function is bounded for the k-point graph but unbounded for the graph spanning all n points. This suggests that the greedy k-point minimal graph is a natural multi-dimensional extension of rank order statistical methods such as the trimmed mean. This complements the comments of Friedman and Rafsky[7] in which they proposed the MST as a natural generalization of one dimensional rank order statistical tests of Smirnov and Wald Wolfowitz.

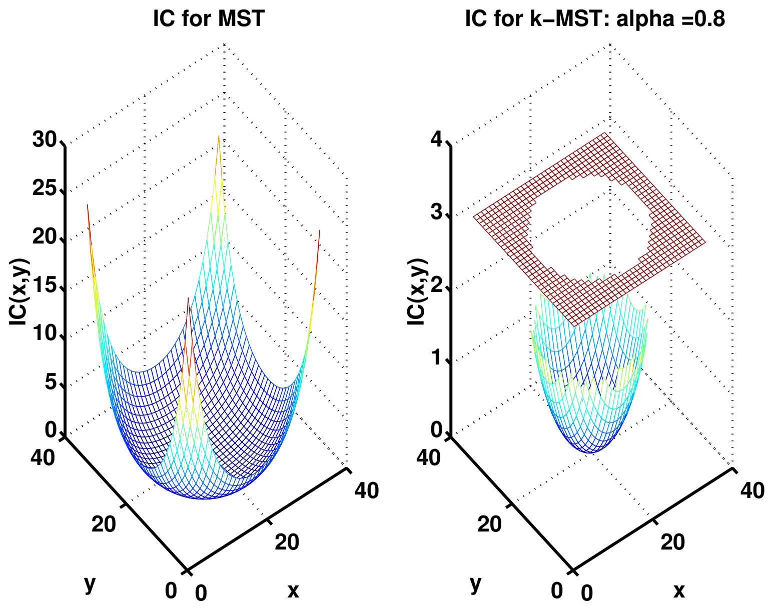

Figure 9. *MST and k-MST influence curves for bivariate Gaussian density on the plane.*

ACKNOWLEDGMENTS

This research was supported in part by AFOSR grant F49620-97-0028.

REFERENCES

1. R. Ravi, M. Marathe, D. Rosenkrantz, and S. Ravi, "Spanning trees short or small," in *Proc. 5th Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 546–555, (Arlington, VA), 1994.
2. F. R. Hampel, *Contributions to the theory of robust estimation.* PhD thesis, Univ. of California - Berkeley, 1968.
3. A. Hero and O. Michel, "Asymptotic theory of greedy approximations to minimal K-point random graphs," Tech. Rep. 315, Communications and Signal Processing Laboratory (CSPL), Dept. EECS, The Univ. of Michigan, Ann Arbor MI 48109-2122, June 1998.
4. C. Redmond and J. E. Yukich, "Limit theorems and rates of convergence for Euclidean functionals," *Ann. Applied Probab.* **4**(4), pp. 1057–1073, 1994.
5. A. Jain and J. Mamer, "Approximations for the random minimal spanning tree with applications to network provisioning," *Oper. Res.* **36**(4), pp. 575–584, 1988.
6. R. C. Prim, "Shortest connection networks and some generalizations," *Bell Syst. Tech. Journ.* , pp. 1389–1401, 1957.
7. J. H. Friedman and L. C. Rafsky, "Multivariate generalizations of the wald-wolfowitz and smirnov two-sample tests," *Annals of Statistics* (4), pp. 697–717, 1979.
8. G. Toussaint, "The relative neighborhood graph of a finite planar set," *Pattern Recognition* , pp. 261–268, 1980.
9. C. Zahn, "Graph-theoretical methods for detecting and describing Gestalt clusters," *IEEE Trans. on Computers* , pp. 68–86, 1971.
10. D. Banks, M. Lavine, and H. J. Newton, "The minimal spanning tree for nonparametric regression and structure discovery," in *Computing Science and Statistics. Proceedings of the 24th Symposium on the Interface*, H. J. Newton, ed., pp. 370–374, 1992.
11. R. Hoffman and A. K. Jain, "A test of randomness based on the minimal spanning tree," *Pattern Recognition Letters* , pp. 175–180, 1983.
12. A. A. Zelikovsky and D. D. Lozevanu, "Minimal and bounded trees," in *Proceedings of Tezele Congres XVIII Acad. Romano-Americaine*, pp. 25–26, (Kishinev), 1993.
13. I. Ahmad and P.-E. Lin, "A nonparametric estimation of the entropy for absolutely continuous distributions," *IEEE Trans. on Inform. Theory* , pp. 664–668, 1976.
14. H. Joe, "On the estimation of entropy and other functionals of a multivariate density," *Ann. Inst. Statist. Math.* , pp. 683–697, 1989.
15. S. C. Hall, Peter Morton, "On the estimation of entropy," *Ann. Inst. Statist. Math.* , pp. 69–88, 1993.
16. A. Gersho, "Asymptotically optimal block quantization," *IEEE Trans. on Inform. Theory* **IT-28**, pp. 373–380, 1979.
17. D. N. Neuhoff, "On the asymptotic distribution of th eerros in vector quantization," *IEEE Trans. on Inform. Theory* **IT-42**, pp. 461–468, March 1996.
18. J. Eckman and D. Ruelle, "Ergodic theory of chaos and strange attractors," *Rev. Mod. Phys* **57**, pp. 617–656, 1985.
19. J. Farmer, "Dimension, fractal measures and chaotic dynamics," in *Evolution of order and chaos*, pp. 228–246, 1982.
20. P. Viola and W. Wells, "Alignment by maximization of mutual information," in *Proc. of 5th Int. Conf. on Computer Vision, MIT*, vol. 1, pp. 16–23, 1995.
21. J. C. Chambers and T. J. Hastie, *Statistical models in S*, Wadsworth, Pacific Grove, CA, 1992.
22. O. Vasicek, "A test for normality based on sample entropy," *J. Royal Statistical Society, Ser. B* , pp. 54–59, 1976.
23. A. Hero and O. Michel, "Robust estimation of point process intensity features using K-minimal spanning trees," in *Proc. of IEEE Symposium on Information Theory*, (Ulm, Germany), July 1997.
24. R. Ravi, M. Marathe, D. Rosenkrantz, and S. Ravi, "Spanning trees – short or small," *SIAM Journal on Discrete Math* **9**, pp. 178–200, 1996.
25. P. J. Huber, *Robust Statistics*, Wiley, New York, 1981.