

Article

## Properties of Branch Length Similarity Entropy on the Network in $\mathbb{R}^k$

Oh Sung Kwon \* and Sang-Hee Lee

National Institute for Mathematical Sciences, 70, Yuseong-daero 1689 beon-gil, Yuseon-gu, Daejeon, Korea; E-Mail: sunchaos@nims.re.kr

\* Author to whom correspondence should be addressed; E-Mail: jamjjari@gmail.com; Tel.: +82-42-717-5741; Fax: +82-42-717-5758.

Received: 7 November 2013; in revised form: 20 December 2013 / Accepted: 3 January 2014 / Published: 16 January 2014

---

**Abstract:** Branching network is one of the most universal phenomena in living or non-living systems, such as river systems and the bronchial trees of mammals. To topologically characterize the branching networks, the Branch Length Similarity (BLS) entropy was suggested and the statistical methods based on the entropy have been applied to the shape identification and pattern recognition. However, the mathematical properties of the BLS entropy have not still been explored in depth because of the lack of application and utilization requiring advanced mathematical understanding. Regarding the mathematical study, it was reported, as a theorem, that all BLS entropy values obtained for simple networks created by connecting pixels along the boundary of a shape are exactly unity when the shape has infinite resolution. In the present study, we extended the theorem to the network created by linking infinitely many nodes distributed on the bounded or unbounded domain in  $\mathbb{R}^k$  for  $k \geq 1$ . We proved that all BLS entropies of the nodes in the network go to one as the number of nodes,  $n$ , goes to infinite and its convergence rate is  $1 - O(1/\ln n)$ , which was confirmed by the numerical tests.

**Keywords:** branch length similarity entropy; BLS entropy; convergence rate

**MSC Classification:** 62B10, 94A17

---

## 1. Introduction

Branching networks can be frequently observed in nature, such as river systems [1,2], the arterial and bronchial trees of mammals [3] and phylogenetic trees [4]. They consist of nodes and branches. Nodes are connection points between branches. Many researchers have extensively studied these networks to characterize them, by using concepts that represent the length of an edge or the strength (or linkage) of a node-node connection or self-similarity. Geological scientists and hydrologists [5–8] were interested in analyzing the complex ordering of these networks. They performed topological and morphometric analyses [9], which can be applied to all branching networks that are organized into a hierarchy. Recently, these approaches have been extended to economic and social systems [10], which are composed of abstractly defined nodes representing the elements of the system and branches representing the interaction between them.

Unlike the aforementioned approaches, Lee *et al.* [11] suggested a new concept and approach, the Branch Length Similarity (BLS) entropy and its profile, to topologically characterize branching systems. The BLS entropy was defined on a simple branching network consisting of a single node and branches. The simple network was referred to as a Unit Branching Network (UBN). The outline of an object's shape in a digitized image is composed of a series of pixels, and a UBN can be built by joining each pixel with every other pixel on the object's outline. Therefore, a BLS entropy profile can be obtained from a series of pixels. In the study of [11], the authors, as an application example of a BLS entropy profile, used the shapes of 20 battle tanks with a pixel resolution of  $460 \times 350$  and showed that the BLS profiles were successful in the identification of the tank shapes. As another example, Kang *et al.* [12] calculated the BLS entropy profile for the wings of butterflies to identify the species, which is often emphasized as the primary step for an understanding of ecology. The identification process has some important practical applications, such as agriculture and border control, in which pests and invaders must be identified and eradicated before they become established as unwanted visitors in agricultural areas (see [13]). The authors showed that a back-propagation neural network system based on the BLS entropy profile has good performance in both accuracy and computational efficiency.

In contrast to the engineering examples mentioned above, the BLS entropy and its profile could be used to characterize and analyze the spatial distribution of elements of a system. In the area of ecology, ecologists have explored statistical methods to characterize the spatial distribution of the ecological elements, such as population density, to infer the existence of underlying processes, such as movement or responses to environmental heterogeneity. This is because the spatial distribution is likely to indicate intraspecific and interspecific interactions, such as competition, predation and reproduction [14,15]. In addition, the importance of characterizing the spatial distribution comes from its central role in ecological theories and its practical role in population sampling theory [16]. In fact, some ecological theories and models use the assumption that the spatial structure of the ecological elements that are close to one another in space or in time are more likely to be affected by the same generating process.

Recently, along with the increase of accessibility and accuracy in remote sensing technology, large-scale analysis and space-time data collection, it has been known that the spatial distribution is strongly scale-dependent in many systems (see [17]). In other words, the random spatial distribution of the physical or abstract elements observed at a small spatial scale in many systems can be identified as the

aggregated distribution at a large spatial scale. For this reason, the novel statistical approaches analyzing the spatial distribution obtained from the multi-scale levels have been required. One of the promising approaches is to form the networks among the elements by cutting and linking the elements [18,19]. Through the investigation of the network properties, such as the connectivity and concentration, one can understand various aspects of the systems at the multi-scale level.

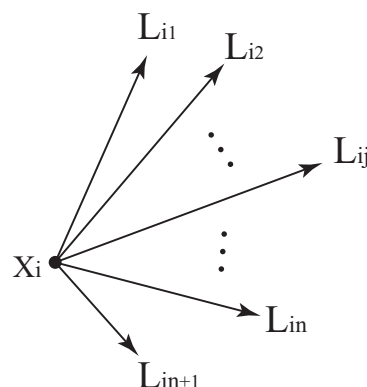
In this viewpoint, the statistical method based on the BLS entropy and its profile, providing a way to make the network and a measure to characterize the network, could be an effective alternative approach. However, although the statistical method could be reliably used in the issues mentioned above, the mathematical properties of the BLS entropy should be extensively explored on a preferential basis to provide a solid ground for applications to a wide range of spatial systems. One of the basic mathematical properties is: what is the value of the BLS entropy for networks consisting of an infinite number of nodes? This question is directly related to the performance and efficiency of the statistical methods based on the BLS entropy profile in the application problems. Jeon and Lee [20] provided a mathematical theorem that shows how the BLS entropy profile changes under the condition that the number of branches goes to infinity. In this study, as the extension study of [20], we explored another theorem for the BLS entropy on the network, which is created by linking infinitely many nodes distributed in the domain in  $\mathbb{R}^k$ .

## 2. Main Result

We define the Branch Length Similarity (BLS) entropy as the property of simple branching networks composed of  $n + 1$  nodes. Such a network is referred to as a Unit Branch Network (UBN) in this paper. Let  $\mathbf{x}_i$  be the position vector of the  $i$ -th node and  $L_{ij}$  be the distance between  $\mathbf{x}_i$  and  $\mathbf{x}_j$ , such that  $L_{ij} = |\mathbf{x}_i - \mathbf{x}_j|$ . For any  $i$ -th node, we consider the UBN as in Figure 1. The probability of the  $j$ -th branch of the  $i$ -th UBN is defined as  $P_{ij}$  in Equation (1). By the mathematical form of the Boltzmann entropy, the BLS entropy,  $S_i$ , of the  $i$ -th UBN is defined by:

$$S_i = \frac{-1}{\ln n} \sum_{i \neq j=1}^{n+1} P_{ij} \ln P_{ij} \text{ with } P_{ij} = L_{ij} / \sum_{i \neq j=1}^{n+1} L_{ij} \quad (1)$$

**Figure 1.** The  $i$ -th Unit Branch Network (UBN).



## 2.1. Theoretical Results

Applying this notion to the nodes placed on an arbitrary bounded domain in  $\mathbb{R}^k$  for  $k \geq 1$ , we obtain the following result.

**Theorem 1.** *Let  $\Omega$  be the bounded domain in  $\mathbb{R}^k$  for  $k \geq 1$ . Suppose there are distinct  $(n + 1)$  nodes in  $\Omega$ . Then, the BLS entropy,  $S_n$ , of any node in  $\Omega$  satisfies:*

$$\lim_{n \rightarrow \infty} S_n = 1 \quad (2)$$

*Proof.* Let  $O$  be any node in  $\Omega$ , which is arbitrarily chosen. Let  $n$  be the natural number greater than one and  $R$  be the longest distance from  $O$  to the boundary of  $\Omega$ . Then,  $\Omega$  belongs to the ball  $B_R := \{\mathbf{x} \in \mathbb{R}^k : |\mathbf{x} - \mathbf{x}_0| \leq R\}$ , where  $\mathbf{x}_0$  is the position vector of  $O$ . Let  $\epsilon_n = \alpha / \ln n$  with a sufficiently small constant  $\alpha \ll R \ln 2$  and  $\Omega_n := \Omega \cap \{r > \epsilon_n\}$ , where  $r = |\mathbf{x} - \mathbf{x}_0|$ . Then,  $\epsilon_n \rightarrow 0$  and  $\Omega_n \rightarrow \Omega$  as  $n \rightarrow \infty$ .

Assume that there are distinct  $n$  nodes in  $\Omega_n$ , except  $O$ . Letting  $L_j (j \leq n)$  be the distance from  $O$  to each  $j$ -th node, it satisfies that  $\epsilon_n \leq L_j \leq R$ . Hence:

$$\sigma := \sum_{j=1}^n L_j \in [\epsilon_n n, Rn] \quad (3)$$

Let  $P_j$  be the probability of a  $j$ -th branch, such that  $P_j = L_j / \sum_{i=1}^n L_i$ . Inserting  $P_j$  into Equation (1),  $S_n$  is:

$$S_n = \frac{\ln \sigma}{\ln n} - \frac{\delta}{\sigma \ln n} \quad (4)$$

where  $\delta = \sum_{j=1}^n L_j \ln L_j$ . By Equation (3), the first term of the right-hand side of Equation (4) is bounded by:

$$\frac{\ln \epsilon_n + \ln n}{\ln n} = \frac{\ln(\epsilon_n n)}{\ln n} \leq \frac{\ln \sigma}{\ln n} \leq \frac{\ln(Rn)}{\ln n} \leq \frac{\ln R + \ln n}{\ln n} \quad (5)$$

By the L'Hôpital's rule, one has:

$$\lim_{n \rightarrow \infty} \frac{\ln \ln n}{\ln n} = \lim_{n \rightarrow \infty} \frac{1}{\ln n} = 0 \quad (6)$$

Thus:

$$\lim_{n \rightarrow \infty} \frac{\ln \epsilon_n}{\ln n} = \lim_{n \rightarrow \infty} \frac{\ln \alpha - \ln(\ln n)}{\ln n} = 0 \quad (7)$$

which shows  $\lim_{n \rightarrow \infty} (\ln \sigma / \ln n) = 1$  by the squeeze theorem.

Now, consider the second term of Equation (4). Since  $L_j \in [\epsilon_n, R]$ , one has:

$$\sigma \ln \epsilon_n \leq \delta \leq \sigma \ln R,$$

which yields  $\ln \epsilon_n / \ln n \leq \delta / (\sigma \ln n) \leq \ln R / \ln n$ . As  $n \rightarrow \infty$  the upper and lower bounds go to zero by Equation (7). Hence, the second term of Equation (4) goes to zero, and Equation (2) follows.  $\square$

Theorem 1. can be extended to the unbounded domain in  $\mathbb{R}^k$ .

**Corollary 1.** *Suppose that there are distinct  $(n + 1)$  nodes on the unbounded domain,  $\Omega$ , in  $\mathbb{R}^k$ . Then, the BLS entropy,  $S_n$ , of the node in  $\Omega$  satisfies Equation (2).*

*Proof.* All procedures of the proof are similar to Theorem 1.. Let  $\epsilon$  and  $R$  be the shortest and longest distances from  $O$  to the other nodes, except  $O$  in  $\Omega$ . Let  $\Omega_n = \Omega \cap \{\epsilon_n < r < R_n\}$  with  $\epsilon_n = \epsilon \frac{\ln 2}{\ln n}$  and  $R_n = R \frac{\ln n}{\ln 2}$ . Then,  $\Omega_n \rightarrow \Omega$  as  $n \rightarrow \infty$ . Let  $L_j (j \leq n)$  be the distance from  $O$  to each  $j$ -th node. Using  $\epsilon_n \leq L_j \leq R_n$  and Equation (6), we can show that  $\ln \sigma / \ln n \rightarrow 1$  as  $n \rightarrow \infty$  by a similar way as Equation (5). Furthermore,  $\lim_{n \rightarrow \infty} \delta / (\sigma \ln n) = 0$  by  $(\ln \epsilon + \ln \ln 2 - \ln \ln n) \sigma \leq \delta \leq (\ln R + \ln \ln n - \ln \ln 2) \sigma$ . Thus, the required result follows by Equation (4).  $\square$

From Equations (4) and (5), the convergence rate behaves like:

$$1 - S_n = O\left(\frac{1}{\ln n}\right) \quad (8)$$

This result is confirmed by the following numerical tests.

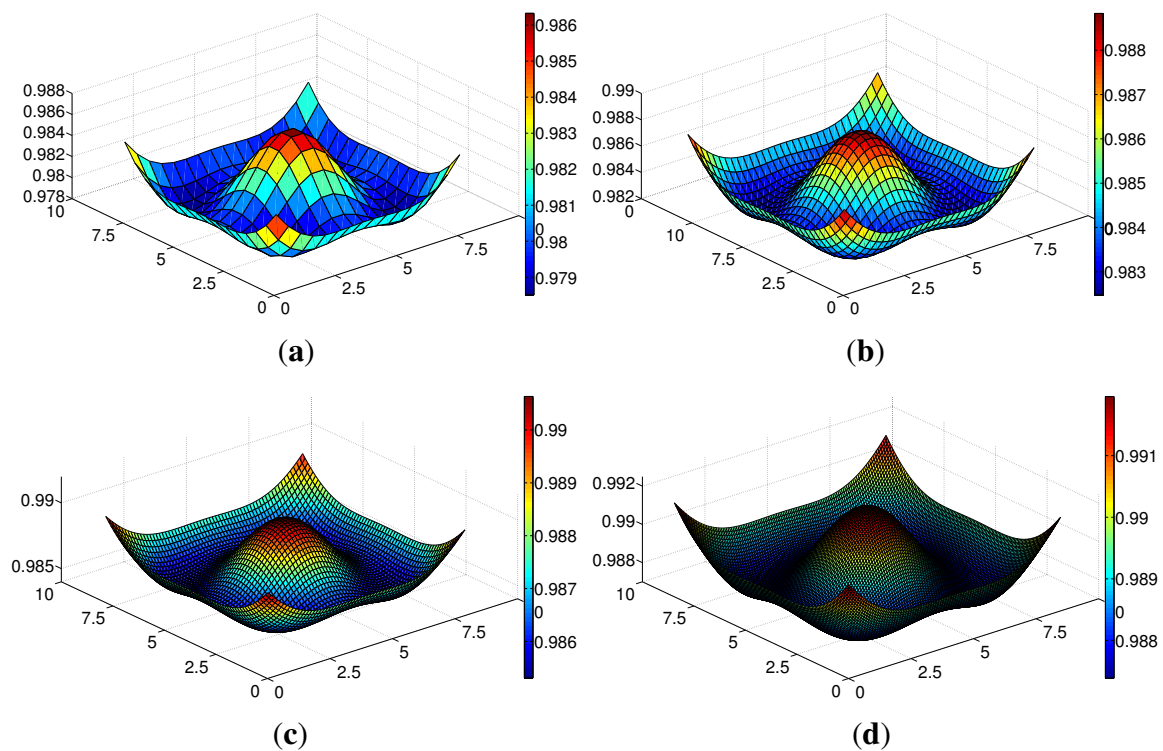
## 2.2. Numerical Tests

We calculate the BLS entropies of the nodes on the bounded regions by increasing the number of nodes. To see the effect of the domain shape and the distribution of the nodes, we consider two regions (rectangle and triangle) and uniform and random distributions.

**Test 1.** First, consider the uniform network on the rectangle region  $R = [0, 10] \times [0, 10]$ . Let  $M$  be a given natural number and  $h = 10/M$ . Dividing the rectangle region,  $R$ , by  $M^2$  squares, there are  $N := (M + 1)^2$  nodes on  $R$ , and the positions of the nodes are given by  $\mathbf{x}_{ij} = (hi, hj)$  for  $0 \leq i, j \leq M$ . For any  $(i, j)$ -th node, first, calculate the distances from the  $(i, j)$ -th node to others. Applying Equation (1), we obtain the BLS entropies for all nodes, and the BLS entropy profiles are drawn as Figure 2 for a given  $N$ . The local maximums are detected at the center and the four vertices. Since the BLS entropy grows up at the sharp corner, the local maximums occur at the vertices. The outlines of the profiles in Figure 2 are similar, but the interval of the BLS entropy grows up from  $N = 256$  to  $N = 16,384$ . To observe this behavior, we calculate the convergence rate in Table 1. Since the convergence rate is a log type, it converges slowly to one as  $N \rightarrow \infty$ . Hence, we calculate the fractional order of the convergence rate by using the  $2^k$  nodes on the uniform network. In particular, we use  $N = 2^{2n}$  for an integer,  $n$ , since  $N$  is a square number. Here,  $|\cdot|_\infty$  is a  $L^\infty$  norm, such that  $|v|_\infty = \sup_{\mathbf{x} \in R} |v(\mathbf{x})|$ , and  $Rate$  denotes  $|1 - S_{N_j}|_\infty / |1 - S_{N_{j-1}}|_\infty$ . *Optimal Rate*, that is, the theoretical convergence rate by Equation (8), is defined by  $\ln 2^{2k} / \ln 2^{2(k+1)} = k / (k + 1)$ . In Table 1, we can observe that  $Rate$  goes to *Optimal Rate* as  $N$  grows greater (see Figure 3). Thus, the convergence rate Equation (8) is confirmed.

**Test 2.** Next is a result on the non-uniform network on the rectangle region,  $R$ . In this case, we spread the nodes randomly on  $R$  by the uniform distribution, that is, a symmetric probability distribution, whereby a finite number of values are equally likely to be observed; every one of the  $n$  values has equal probability  $1/n$ . In the same way, we calculate the BLS entropies for all nodes. In this case, we take the mean value of the 10 times results in obtaining the norm,  $|1 - S_N|_\infty$ , to reduce the effect of the randomness. For the ease of calculation, the convergence rate we set is the number of nodes  $N = 3^k$  for an integer  $k$ . The figures in Figure 4 are depicted by using the interpolating method on the uniform grid (we used “griddata”, which is a built-in function of Matlab). For a small  $N$ , the profile of the BLS entropy has a complex shape, but the shape of the profile resembles the uniform case in Figure 2, as  $N$  is greater. By Figure 4 and Table 2, we can see that the interval of the profile grows up, and the convergence rate goes to *Optimal Rate*, though it is not as clear as Test 1 (see Figure 3).

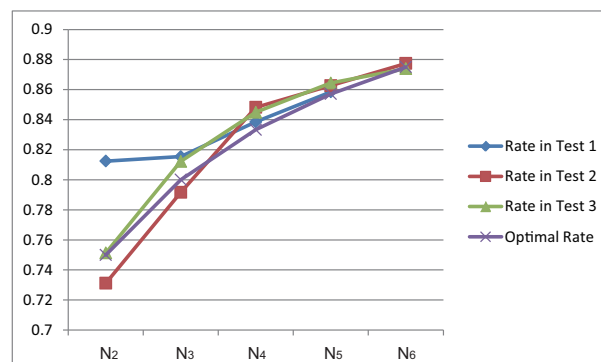
**Figure 2.** The Branch Length Similarity (BLS) entropy profiles of the uniformly distributed network on the rectangle,  $R$ . (a)  $N = 256$ ; (b)  $N = 1,024$ ; (c)  $N = 4,096$ ; (d)  $N = 16,384$ .



**Table 1.** The convergence rate of the uniformly distributed network on the rectangle,  $R$ .

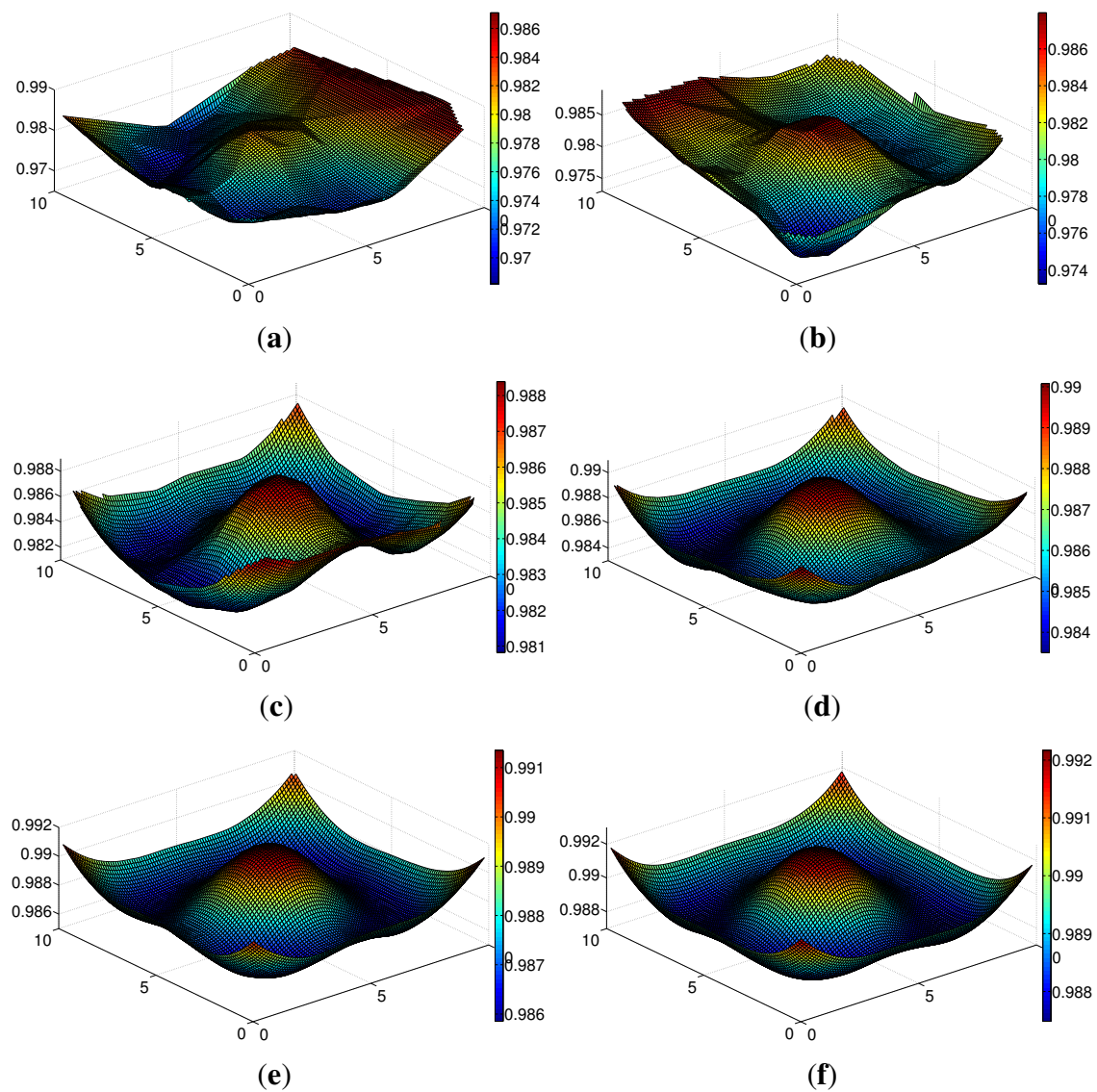
$N$	$ 1 - S_N _\infty$	Rate	Optimal Rate
$N_1 = 8^2 = (2^3)^2$	2.6452E-2	-	-
$N_2 = 16^2 = (2^4)^2$	2.1490E-2	0.81241	0.75000
$N_3 = 32^2 = (2^5)^2$	1.7523E-2	0.81540	0.80000
$N_4 = 64^2 = (2^6)^2$	1.4693E-2	0.83849	0.83333
$N_5 = 128^2 = (2^7)^2$	1.2613E-2	0.85844	0.85714

**Figure 3.** Convergence rates for each test:  $N_j$  is the number of nodes in Tables 1–3.





**Figure 4.** The BLS entropy profiles of the randomly distributed network on the rectangle,  $R$ . (a)  $N = 81$ ; (b)  $N = 243$ ; (c)  $N = 729$ ; (d)  $N = 2,187$ ; (e)  $N = 6,561$ ; (f)  $N = 19,683$ .



**Table 2.** The convergence rate of the randomly distributed network on the rectangle,  $R$ .

$N$	$ 1 - S_N _\infty$	<i>Rate</i>	<i>Optimal Rate</i>
$N_1 = 3^4$	3.3683E-2	-	-
$N_2 = 3^5$	2.4627E-2	0.73114	0.75000
$N_3 = 3^6$	1.9495E-2	0.79161	0.80000
$N_4 = 3^7$	1.6536E-2	0.84822	0.83333
$N_5 = 3^8$	1.4265E-2	0.86266	0.85714
$N_6 = 3^9$	1.2516E-2	0.87739	0.87500

**Table 3.** The convergence rate of the randomly distributed network on the triangle,  $T$ .

$N$	$ 1 - S_N _\infty$	Rate	Optimal Rate
$N_1 = 3^4$	3.5863E-2	-	-
$N_2 = 3^5$	2.6943E-2	0.75128	0.75000
$N_3 = 3^6$	2.1883E-2	0.81220	0.80000
$N_4 = 3^7$	1.8489E-2	0.84490	0.83333
$N_5 = 3^8$	1.5982E-2	0.86441	0.85714
$N_6 = 3^9$	1.3968E-2	0.87398	0.87500

**Test 3.** Next is the result of the non-uniform network on the triangular region,  $T$ , by spreading the nodes randomly using the uniform distribution.  $T$  is composed of three points,  $(0, 0)$ ,  $(10, 0)$  and  $(5, 10)$ , and the BLS entropies of the nodes on  $T$  are calculated as in Figure 5. In this case, the local maximums are found at the center and three vertices and the profile are stable as  $N$  becomes greater. Its convergence rate also goes to *Optimal Rate* as  $N$  becomes greater.

**Figure 5.** The BLS entropy profiles of the randomly distributed network on the triangle,  $T$ .

(a)  $N = 81$ ; (b)  $N = 243$ ; (c)  $N = 729$ ; (d)  $N = 2,187$ ; (e)  $N = 6,561$ ; (f)  $N = 19,683$ .

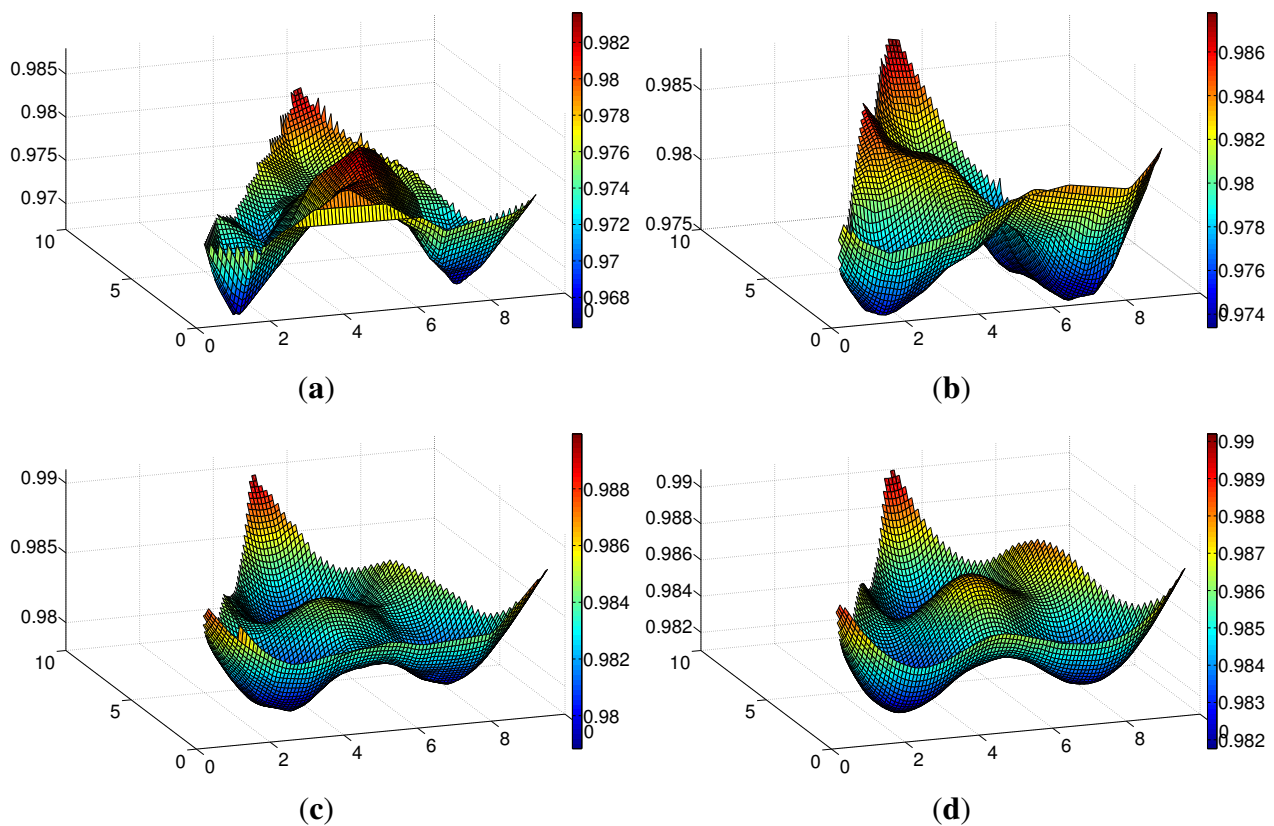
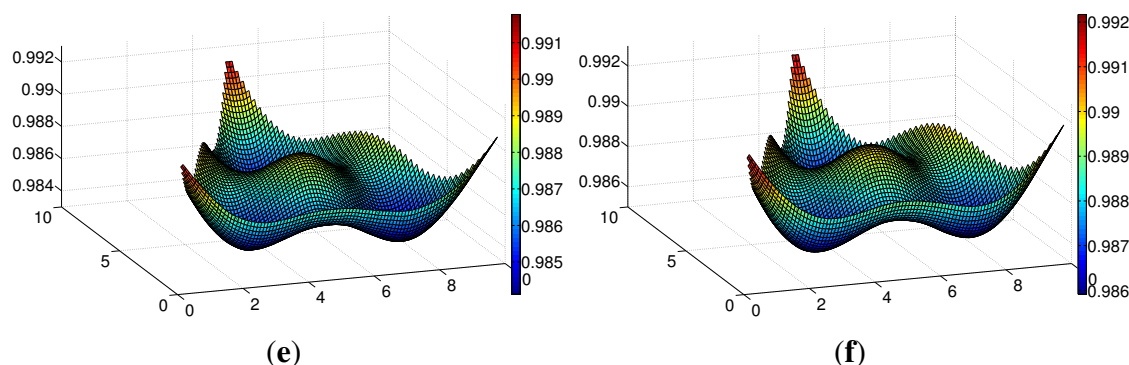




Figure 5. Cont.



### 3. Conclusions

In this paper, we showed that the BLS entropy of any network in  $\mathbb{R}^k$  increases at every node and, finally, converges to one as the number of nodes,  $N$ , increases, and we confirmed it by the numerical tests on the rectangle and triangle. Besides, the following relations are obtained by comparing Tests 1–3. From Tests 1 and 2, the different distributions of the networks on the same region do not effect the convergence rate for a sufficiently large  $N$ . Furthermore, the shape of the region has no effect on the convergence rate by Tests 2 and 3. However, the BLS entropy profile is characterized by the shape of the region. This is confirmed by Tests 1 and 2. Particularly, the BLS entropy profiles resemble each other as  $N$  increases.

One important point is: what is the optimal  $N$  in characterizing the shape of the region? Too small an  $N$  is likely to lose information regarding shape, while too large an  $N$  dilutes the characteristics of the region, since all entropies go to one as  $N \rightarrow \infty$ . Thus, finding the optimal number for  $N$  could provide positive support for the shape matching methods used and spreading the engineering applications.

Consequently, our result is meaningful in that it not only shows the convergence rate of the BLS entropy on the networks in  $\mathbb{R}^k$ , but also, it provides solid ground for the development of BLS entropy profile methods that could be practically used as a simple and useful tool for recognizing and characterizing shapes.

### Acknowledgments

This work was supported by the National Institute for Mathematical Sciences.

### Conflicts of Interest

The authors declare no conflicts of interest.

### References

1. Benda, L.; Poff, N.L.; Miller, D.; Dunne, T.; Reeves, G.; Pess, G.; Pollock, M. The network dynamics hypothesis: How channel networks structure riverine habitats. *Bioscience* **2004**, *54*, 413–427.
2. Fisher, S.G. Creativity, idea generation, and the functional morphology of streams. *J. N. Am. Benthol. Soc.* **1997**, *16*, 305–318.

3. West, G.B.; Brown, J.H.; Enquist, B.J. A general model for the origin of allometric scaling laws in biology. *Science* **1997**, *276*, 122–126.
4. Restrepo, J.G.; Ott, E.; Hunt, B.R. Emergence of coherence in complex networks of heterogeneous dynamical systems. *Phys. Rev. Lett.* **2006**, *96*, 254103.
5. Horton, R. Erosional development of streams and their drainage basins; hydrophysical approach to quantitative morphology. *Geol. Soc. Am. Bull.* **1945**, *56*, 275–370.
6. Schumm, S. Evolution of drainage systems and slopes in badlands at perth amboy. *N. J. Geol. Soc. Am. Bull.* **1956**, *67*, 597–646.
7. Shreve, R. Stream lengths and basin areas in topologically random channel networks. *J. Geol.* **1969**, *77*, 397–414.
8. Strahler, A. Quantitative analysis of watershed geomorphology. *Trans. Am. Geophys. Union* **1957**, *38*, 913–920.
9. Kirshen, D.; Bras, R. The linear channel and its effect on the geomorphologic IUH. *J. Hydrol.* **1983**, *65*, 175–208.
10. Arthur, W.B.; Holland, J.H.; LeBaron, B.; Palmer, R.; Tyler, P. *The Economy as an Evolving Complex System II*; Addison-Wesley: Boston, MA, USA, 1997.
11. Lee, S.-H.; Su, N.-Y.; Bardunias, P. Novel approach to shape recognition using shape outline. *J. Kor. Phys. Soc.* **2010**, *56*, 1016–1019.
12. Kang, S.H.; Jeon, W.; Lee, S.-H. Butterfly species identification by branch length similarity entropy. *J. Asia-Pac. Entomol.* **2012**, *15*, 437–441.
13. Gaston, K.J.; May, R.M. Taxonomy of taxonomists. *Nature* **1992**, *356*, 281–282.
14. Dale, M.R.T. *Spatial Pattern Analysis in Plant Ecology*; Cambridge University Press: Cambridge, England, 2000.
15. Dale, M.R.T.; Zbigniewicz, M.W. Spatial pattern in boreal shrub communities: Effects of a peak in herbivore densities. *Can. J. Bot.* **1997**, *75*, 1342–1348.
16. Burrough. P.A. Spatial Aspects of Ecological Data. In *Data Analysis in Community and Landscape Ecology*; Jongman, R.H.G., Ter Braak, C.J.F., van Tongeren, O.F.R., Eds.; Cambridge University Press: Cambridge, England, 1995; 213–251.
17. Epperson, B.K. Spatial and space-time correlation in ecological models. *Ecol. Model.* **2000**, *132*, 63–76.
18. Anselin, L.; Florax, R.; Rey, S.J. *Advances in Spatial Econometrics: Methodology, Tools and Applications*; Springer-Verlag: Berlin, Germany, 2004.
19. Anselin, L.; Syabri, I.; Kho, Y., GeoDa: An introduction to spatial data analysis. *Geograph. Anal.* **2006**, *38*, 5–22.
20. Jeon, W.; Lee, S.-H. Mathematical investigations of branch length similarity entropy profiles of shapes for various resolutions. *J. Kor. Phys. Soc.* **2012**, *61*, 1906–1910.