Article

Informational Entropy of B-ary Trees after a Vertex Cut

Lorentz Jäntschi [1,2,]* and Sorana D. Bolboacă [1,2]

[1] Technical University of Cluj-Napoca, 103-105 Muncii Bvd, 400641 Cluj-Napoca, Cluj, Romania. Phone: +4-0264-401775. Fax: +4-0264-415054. E-mail: lori@academicdirect.org

[2] "Iuliu Haţieganu" University of Medicine and Pharmacy Cluj-Napoca, Department of Medical Informatics and Biostatistics, 6 Louis Pasteur, 400349 Cluj-Napoca, Cluj, Romania. Phone: +4-0264-431697. Fax: +4-0264-593847. E-mail: sbolboaca@umfcluj.ro

* Author to whom correspondence should be addressed.

Abstract: Together with stars and paths, b-ary trees are one of the most studied acyclic graph structures. As any other structure, a b-ary tree can be seen as containing information. The aim of the present research was to assess through informational entropy the structural information changes in b-ary trees after removal of an arbitrary vertex.

Keywords: B-ary tree, informational entropy, vertex cut.

AMS Mathematics Subject Classification: 05C05 (Trees), 37B40 (Topological entropy), 94A17 (Measures of information, entropy), 11T06 (Polynomials).

1. Introduction

B-ary trees are, together with stars and paths, one of the most studied acyclic graph structures. A survey on the available literature showed that almost 20% of papers dealing with graphs deal also with trees. The statistics calculated on trees are used in machine learning [1] and solving of nondeterministic polynomial-time hard (NP-hard) problems [2]. The concept of entropy in conjunction with regular trees was used as measure of similarity between two grammars [3]. Moreover, the information flows in b-ary trees was subject of some theoretical results, including a converse in [4]. General formulas characterizing the random walk in a b-ary tree were obtained in [5]. A general
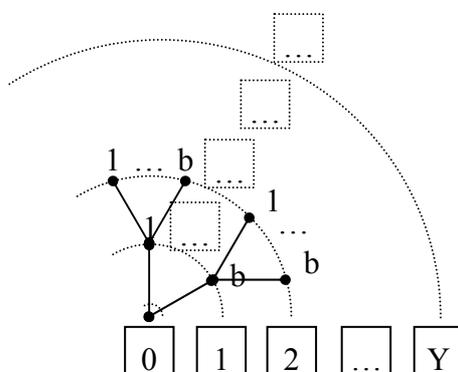
formula giving the number of substructures by their sizes by systematically removal of a vertex in a b-ary tree was also obtained [6].

As any other structure, a tree can be seen as a system that contains information. The aim of the present research was to assess through informational entropy the structural information changes in b-ary trees after removal of an arbitrary vertex.

2. Material

Let T(b,Y) be a regular b-ary tree with Y levels (Figure 1). The removal of an arbitrary vertex applied on a b-ary tree will generate one (when the removal is applied to a leaf), two (when the removal is applied to a root), or three (when the removal is applied to an inside node) sub-graphs.

Figure 1. A b-ary tree with Y levels, T(b,Y)

It is possible to evaluate the number of sub-graphs by removing every vertex from the tree one at a time. A statistical motivation is associated to this procedure. If we are able to obtain the entire population of sub-structures by their sizes after this procedure, then the probability of appearances of a substructure of a given size can be obtained (by dividing its occurrence to the total number of substructures).

It could be said that systematically removal of every vertex is not a usual procedure for a given practical problem. Indeed, but usually in a real application, an issue may be the removal of one vertex when the specification of which one vertex is to be remove is unknown. One example may be a tree network topology (let us say a Regional Internet Registry providing Internet resource allocations). In this case, the assessment of the impact of removing a node is an important issue for projecting the network.

The assumptions applied in the present research were as follows: all vertices and edges in the b-ary tree are present (or available) at the initial moment (before removing the first vertex), the tree is balanced, and the vertices and edges have no precedence/weight one to another. Then we will move from complete characterization of a set (the set of all sub-graphs obtained by removing of an arbitrary vertex of a tree) to an issue of probability (obtaining the appearance probability of a set of a given size). In terms of the entire structure (the tree), the appearance probability of a set of a given size may not be a relevant information but a global parameter which to characterize the tree under the given circumstances (removal of one vertex) could be of interest and was investigated.

3. Method

Entropy is the parameter that characterizes the amount of disorder. Rényi [15] gave a generalization of the entropy through a family of functional for quantifying the diversity, uncertainty or randomness of a system.

Let X be a discrete variable that take the values {x1, …, xn} and p(·) the probability mass function (gives the probability that a discrete random variable is exactly equal to a value). The Rényi entropy (abbreviated as $H_\alpha(X)$ for any non-negative order α, where α ≥ 0) is given by:

$$H_\alpha(X) = \frac{1}{1-\alpha} \log_2 \sum_{i=1}^{n} p^\alpha(x_i)$$

(1)

Two particular cases of Eq(1) are taken into consideration: α → 1 (or Shannon's entropy [8]), and α = 2 (negative logarithm of Simpson's diversity index D [10] or Onicescu's energy [9]):

$$H_1(X) = -\sum_{i=1}^{n} p(x_i) \log_2(p(x_i))$$

$$D(X) = \sum_{i=1}^{n} p^2(x_i); \quad H_2(X) = -\log_2 D(X)$$

(2)

4. Results

Let us recall the structure from Figure 1. The following formula (abbreviated as NSSP) describes the number of substructures (coefficient of the X variable) by sizes (power of the X variable) for systematically removal of a vertex from a b-ary tree [6]:

$$NSSP(b, Y) = \sum_{k=1}^{Y} b^k \left(X^{\frac{b^{Y+1-k}-1}{b-1}} + X^{\frac{b^{Y+1}-b^{Y+1-k}}{b-1}} \right) + \frac{b^{Y+1}-1}{b-1} X^0$$

(3)

The Eq(3) give the full description of the counts and sizes of substructures for any Y,b>0. The Table 1 presents the substructures and sizes for some particular cases of trees. Note that the coefficient of X^0 counts the total number of vertices in the original structure.

Table 1. Substructures obtained by remove of an arbitrary vertex: regular b-ary trees

b	Y	Formula	Comments
1	1	$2X^0+2X^1$	A graph with two vertices and one edge
1	>1	$(Y+1)X^0+2X(X^Y-1)/(X-1)$	For b=1 tree degenerates in a path
>1	1	$(b+1)X^0+b(X+X^b)$	For Y=1 tree degenerates in a star
b	2	$(b^2+b+1)X^0+b(X^{b+1}+X^{b2})+b^2(X+X^{b2+b})$	Powers of X's may be repeated in the series

Y = number of levels; b = ramification degree of the b-ary tree

By using the Eq(3), we want to find the roots of the following equation:

$$b^{Y+1-k_1} - 1 = b^{Y+1} - b^{Y+1-k_2} \quad \text{i.e.} \quad 1 + b^{k_2-k_1} = b^{k_2} + b^{k_2-(Y+1)} \tag{4}$$

Excluding b=1 (b, Y, k_1, and k_2 are natural not null numbers; and 1 is also a natural number) the equality Eq(4) can be satisfied just when the only solutions of Eq(5) is for k_1=Y+1 (that implying as consequence to have the b=1). The conclusion that can be drawn from here is that for b>1 all terms from Eq(3) are distinct.

Since the problem of occurrences of sub-graphs sizes (n_o) was solved in the general case for b>1 (n_o repeats in terms of Eq(3)) we are able to perform the calculations for a given Y and b (note that the problem can be extended to any b and Y less than or equal to a given value):

$$n_o\left(\frac{b^{Y+1-k}-1}{b-1}\right) = \text{coef}(X^{\frac{b^{Y+1-k}-1}{b-1}}) = n_o\left(\frac{b^{Y+1}-b^{Y+1-k}}{b-1}\right) = \text{coef}(X^{\frac{b^{Y+1}-b^{Y+1-k}}{b-1}}) \tag{5}$$

in addition, the probability (of observing a structure of a given size in a random cut) is given by:

$$p\left(\frac{b^{Y+1-k}-1}{b-1}\right) = p\left(\frac{b^{Y+1}-b^{Y+1-k}}{b-1}\right) = \frac{b^k}{2\sum\limits_{k=1}^{Y} b^k} = \frac{b^{k-1}(b-1)}{2(b^Y-1)} \tag{6}$$

The replacement of Eq(6) into Eq(1) and Eq(2) leads to the following formulas:

$$H_1(b,Y) = -2\sum\limits_{k=1}^{Y}\frac{b^{k-1}(b-1)}{2(b^Y-1)}\log_2\left(\frac{b^{k-1}(b-1)}{2(b^Y-1)}\right) = \log_2\frac{2(b^Y-1)}{b-1} - \frac{b^{Y+1}(Y-1)-Yb^Y+b}{(b^Y-1)(b-1)}\log_2 b \tag{7}$$

$$H_2(b,Y) = -\log_2 2\sum\limits_{k=1}^{Y}\frac{b^{2k-2}(b-1)^2}{4(b^Y-1)^2} = -\log_2\frac{1}{2}\frac{b^Y+1}{b^Y-1}\frac{b-1}{b+1} = 1-\log_2\left(\frac{b^Y+1}{b^Y-1}\frac{b-1}{b+1}\right) \tag{8}$$

where $H_1(b,Y)$ is the Rényi entropy of order 1; $H_2(b,Y)$ is the Rényi entropy of order 2 of a b-ary tree T(b,Y) after removal of a vertex, when b>1. The value of 2 is given by the fact that we have equal probabilities for fragments of size (b^{Y+1-k} - 1)/(b - 1), and ($b^{Y+1} - b^{Y+1-k}$)/(b - 1) respectively (see Eq(6)).

For b=1 (path) an entry from Table 1 can help us to obtain the values for $H_1(1,Y)=H_1(Y)$ and $H_2(1,Y)=H_2(Y)$, respectively:

$$H_1(Y) = -\sum\limits_{k=1}^{Y}\frac{2}{2Y}\log_2\left(\frac{2}{2Y}\right) = -\sum\limits_{k=1}^{Y}\frac{1}{Y}\log_2\left(\frac{1}{Y}\right) = \log_2(Y) \tag{9}$$

$$H_2(Y) = -\log_2\sum\limits_{k=1}^{Y}\left(\frac{2}{2Y}\right)^2 = -\log_2\frac{1}{Y} = \log_2(Y) = H_1(Y) \tag{10}$$

5. Discussion

In terms of statistics, the random event subject of the investigation was to remove a vertex from a b-ary tree, and the observation was to count the size of a connected structure (the size of a connected structure being the random variable). This statistical experiment allowed obtaining an important informational statistics: entropy and diversity index.

Equations (7)-(10) allow one to compute the values and based on the obtained results, to plot the dependence of the $H_x(b,Y)$ (where x = 1, 2; $H_1(b,Y)$ = Rényi entropy of order 1; $H_2(b,Y)$ = Rényi entropy of order 2) and $D(b,Y)$ by the number of levels (Y) and ramification degree (b) of the tree. These dependencies ranging from 1 to 20 for Y and b are presented in Figure 2.

Figure 2. Rényi entropy of order 1 - $H_1 = H_1(b,Y)$, Rényi entropy of order 2 - $H_2 = H_2(b,Y)$, and Simpson's diversity index - $D = D(b,Y)$ of a b-ary tree with Y levels after a vertex removal
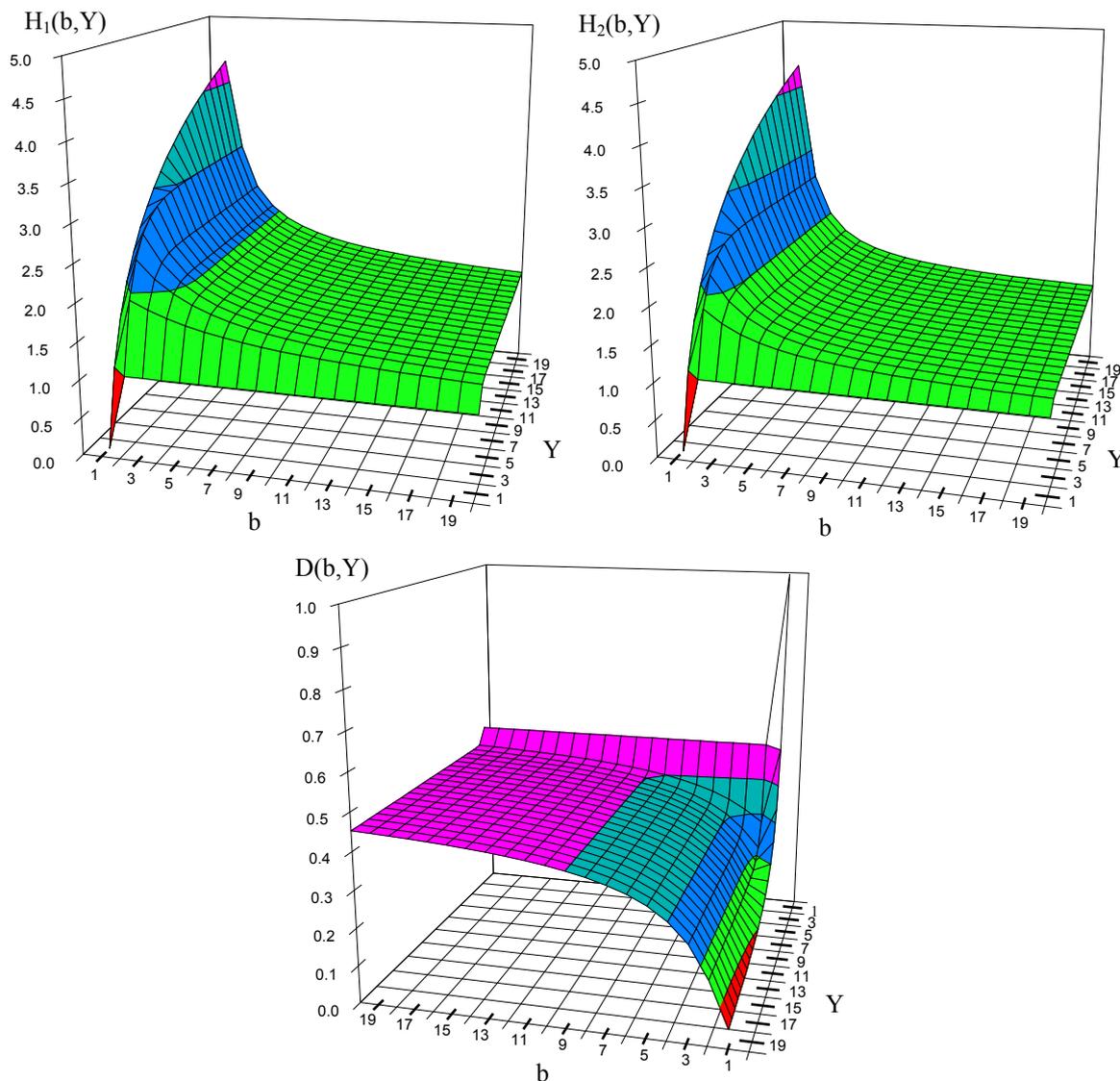

The analysis of the calculated values obtained by applying the formulas for Rényi entropy of order 1 and 2, and of Simpson's diversity index revealed the followings:

÷ For a given value of b (ramification degree of the b-ary tree):

  o The values of Rényi entropy of order 1 are equal with the values of and Rényi entropy of order 2 for b=1 and 1≤Y≤20.

 o The distribution of D(b,Y), $H_1(b,Y)$ and $H_2(b,Y)$ follow a similar pattern for b > 2 and 1≤Y≤20.

- o Starting with b=4 the values of D(b,Y), H_1(b,Y) and H_2(b,Y) become a monotone function with the increases of the number of levels. Very small differences were observed for 7≤b≤20.
- o If the b=1 is excluded from the analysis, the values of D(b,Y) decrease with the increasing of Y while the values of H_1(b,Y) and H_2(b,Y) increase with increasing of Y (see Figure 3).

Figure 3. Variation of Simpson's diversity index (D) and Rényi's entropy of order 1 (H_1) for a given value of ramification degree
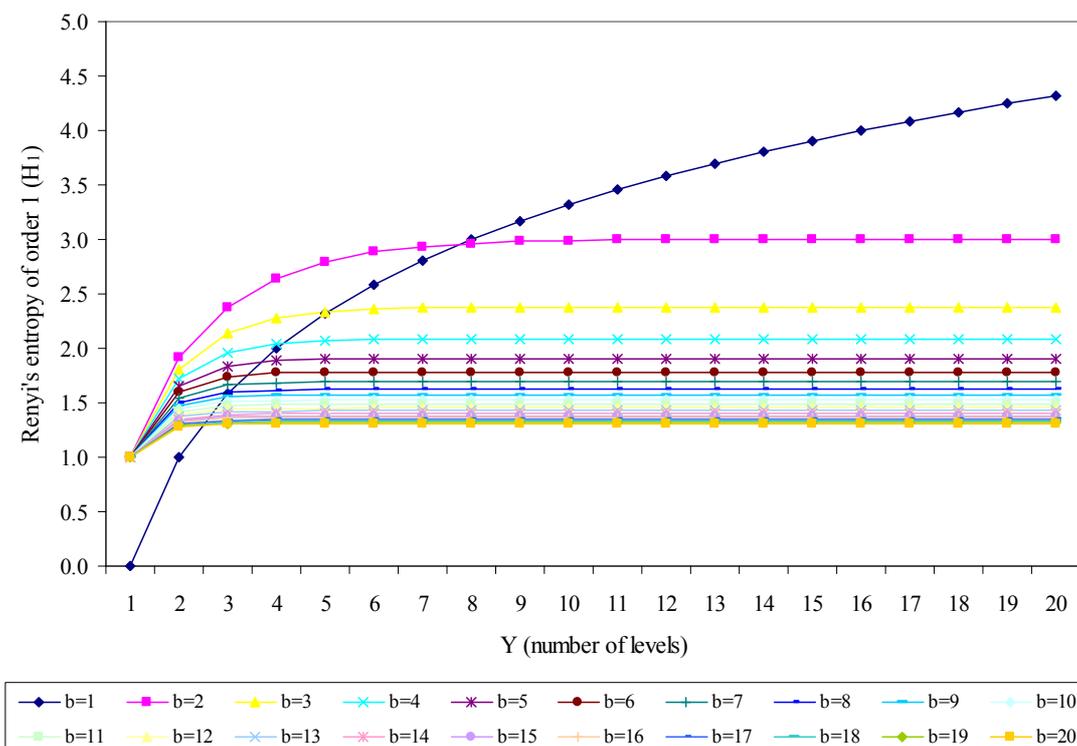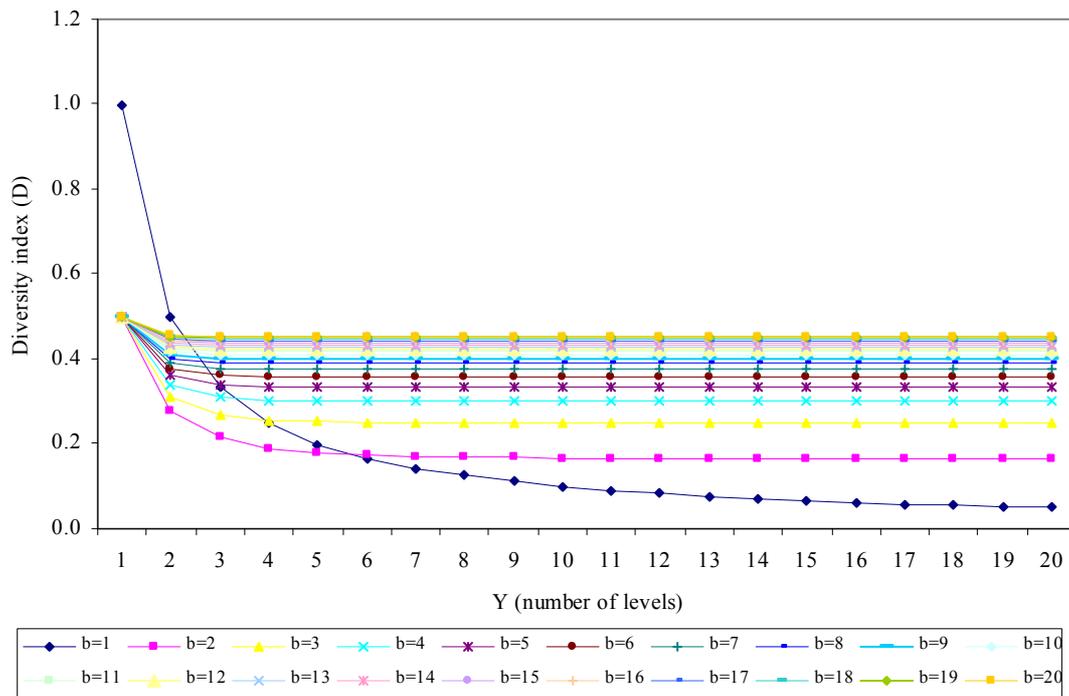
o   The maximum value is systematically obtained by b=2 for D(b,Y) and b=20 for H_1(b,Y) and H_2(b,Y). The D(b,Y) values tend to varied with very small quantities (the differences are not observed at the level of first four decimals) starting with a certain value of Y (D(b,Y): D(2,12), D(3,9), D(4,7), D(5,7), D(6,7), D(7,5), D(8,5), D(9,5), D(10,5), D($\geq$11,4)). The variation with very small quantities are also observed for H_1(b,Y) (H_1(2,19); H_1(3,12); H_1(4,10); H_1(5,8); …) and H_2(b,Y) (H_2(2,18); H_2(3,10); H_2(4,9); H_2(5,8); …).

÷   For a given value of Y (number of levels of the b-ary tree):

o   For Y=1: the values of D(b,Y) were equal with 0.5 for 2≤b≤20; the values of H_1(b,Y) and H_2(b,Y) were equal to 1 for 2≤b≤20.

o   The distribution of D(b,Y) follows a certain pattern for b>2 and 1≤Y≤20. A similar pattern is observed for distribution of H_1(b,Y) and H_2(b,Y).

o   The values of D(b,Y) increase with increases of b while the values of H_1(b,Y) and H_2(b,Y) decrease with increasing of b after an initial increase (see b>2 and Y>8, Figure 4).

Figure 4. Variation of Rényi's entropy of order 1 (H_1) and 2 (H_2) for a given value of Y
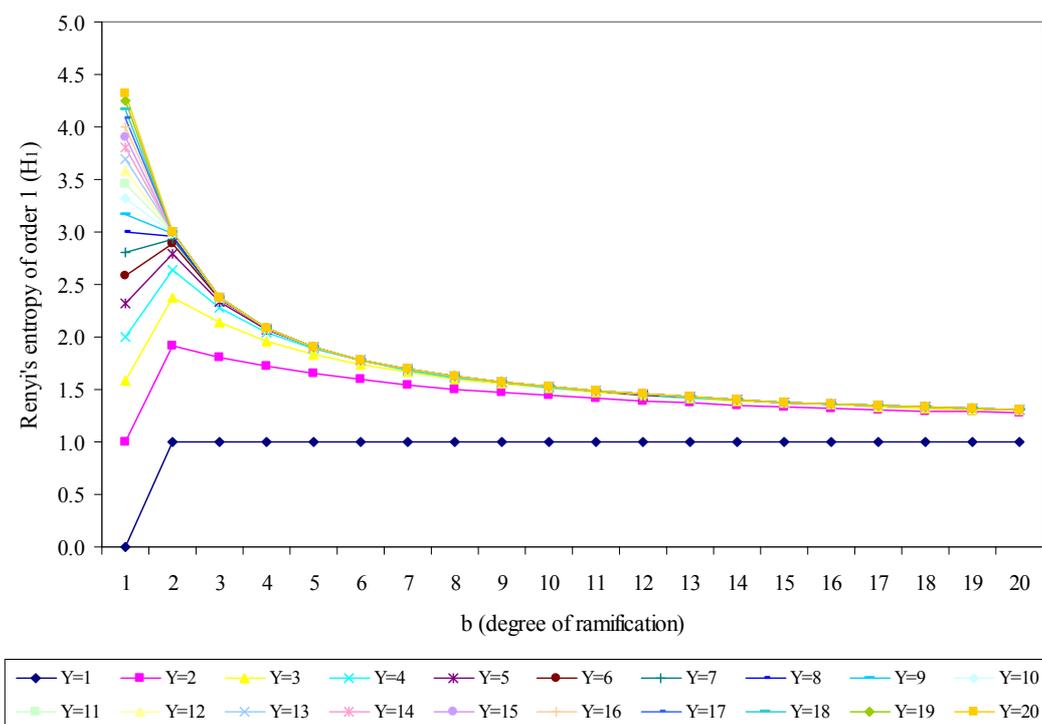
The maximum value of Rényi entropy of order 1 was equal with the maximum value of Rényi entropy of order 2 (4.3219) for T(1,20). With a few exceptions (for T(1,<5)) the values of Rényi entropy were higher than 2 which leads to the conclusion that the most disorder structure were the ones from the path (unary-tree) class (T(1,5), T(1,6), T(1,7), T(1,8), T(1,9), T(1,10), T(1,11), T(1,12), T(1,13), T(1,14), T(1,15), T(1,16), T(1,17), T(1,18), T(1,19), and T(1,20)). The structures become more stable as the degree of the ramification of the b-ary tree increase (structures become more complex). The values of Rényi entropy of order 1 range from 1.2762 to 1.3015 for T(20, 1<Y≤20), and the values of Rényi entropy of order 2 range from 1.1372 to 1.1444 for T(20, 1<Y≤20).

A regression analysis for a give value of levels (Y) was applied in order to identify the relationship between Simpson's diversity index ($D(b,Y)$) and Rényi entropy of order 1 ($H_1(b,Y)$), and between Rényi entropy of order 1 ($H_1(b,Y)$) and 2 ($H_2(b,Y)$). The obtained results expressed as equations and associated squared correlation coefficients are presented in Table 2. The relative variation (min=0; max=1) of regression equation coefficients from Table 2 for H1 = f(D) and H2 = f(H1) are graphically presented in Figure 5.

Table 2. Substructures by removal of an arbitrary vertex for some regular b-ary trees

Y	$H_1 = f(D)$ Equation	R^2 (for 1<b<21)	$H_2 = f(H_1)$ Equation	R^2 (for 1<b<21)
2	$-4.68 \cdot D^2 - 0.17 \cdot D + 2.32$	0.999904	$0.77 \cdot H_1^2 - 1.37 \cdot H_1 + 1.63$	0.999860
3	$-1.50 \cdot D^2 - 3.50 \cdot D + 3.20$	0.999948	$0.39 \cdot H_1^2 - 0.45 \cdot H_1 + 1.07$	0.999863
4	$0.82 \cdot D^2 - 5.58 \cdot D + 3.66$	0.999930	$0.26 \cdot H_1^2 + 0.08 \cdot H_1 + 0.81$	0.999934
5	$2.49 \cdot D^2 - 6.94 \cdot D + 3.94$	0.999855	$0.19 \cdot H_1^2 + 0.13 \cdot H_1 + 0.66$	0.999947
6	$3.61 \cdot D^2 - 7.83 \cdot D + 4.11$	0.999738	$0.15 \cdot H_1^2 + 0.25 \cdot H_1 + 0.56$	0.999989
7	$4.32 \cdot D^2 - 8.37 \cdot D + 4.22$	0.999614	$0.13 \cdot H_1^2 + 0.32 \cdot H_1 + 0.51$	0.999966
8	$4.76 \cdot D^2 - 8.71 \cdot D + 4.28$	0.999511	$0.12 \cdot H_1^2 + 0.36 \cdot H_1 + 0.47$	0.999936
9	$5.02 \cdot D^2 - 8.90 \cdot D + 4.31$	0.999436	$0.11 \cdot H_1^2 + 0.38 \cdot H_1 + 0.45$	0.999911
10	$5.17 \cdot D^2 - 9.02 \cdot D + 4.34$	0.999387	$0.11 \cdot H_1^2 + 0.40 \cdot H_1 + 0.44$	0.999893
11	$5.25 \cdot D^2 - 9.08 \cdot D + 4.35$	0.999357	$0.10 \cdot H_1^2 + 0.41 \cdot H_1 + 0.43$	0.999882
12	$5.30 \cdot D^2 - 9.12 \cdot D + 4.35$	0.999340	$0.10 \cdot H_1^2 + 0.41 \cdot H_1 + 0.43$	0.999875
13	$5.33 \cdot D^2 - 9.14 \cdot D + 4.36$	0.999330	$0.10 \cdot H_1^2 + 0.41 \cdot H_1 + 0.43$	0.999871
14	$5.34 \cdot D^2 - 9.15 \cdot D + 4.36$	0.999324	$0.10 \cdot H_1^2 + 0.41 \cdot H_1 + 0.43$	0.999868
15	$5.35 \cdot D^2 - 9.16 \cdot D + 4.36$	0.999321	$0.10 \cdot H_1^2 + 0.42 \cdot H_1 + 0.43$	0.999867
16	$5.36 \cdot D^2 - 9.16 \cdot D + 4.36$	0.999319	$0.10 \cdot H_1^2 + 0.42 \cdot H_1 + 0.42$	0.999866
17	$5.36 \cdot D^2 - 9.16 \cdot D + 4.36$	0.999318	$0.10 \cdot H_1^2 + 0.42 \cdot H_1 + 0.42$	0.999866
18	$5.36 \cdot D^2 - 9.16 \cdot D + 4.36$	0.999318	$0.10 \cdot H_1^2 + 0.42 \cdot H_1 + 0.42$	0.999866
19	$5.36 \cdot D^2 - 9.16 \cdot D + 4.36$	0.999317	$0.10 \cdot H_1^2 + 0.42 \cdot H_1 + 0.42$	0.999866
20	$5.36 \cdot D^2 - 9.16 \cdot D + 4.36$	0.999317	$0.10 \cdot H_1^2 + 0.42 \cdot H_1 + 0.42$	0.999866
	$coef(D^2) = 5.441 - 23.71 \cdot e^{-Y/2.39}$	0.999	$coef(H_1^2) = 0.1 + 2.9 \cdot e^{-Y/1.35}$	0.997
	$coef(D) = -9.195 + 23.08 \cdot e^{-Y/2.14}$	0.999	$coef(H_1) = 0.4 - 7.47 \cdot e^{-Y/1.39}$	0.994
	$Intercept = 4.361 - 5.92 \cdot e^{-Y/1.87}$	0.999	$Intercept = 0.43 + 3.7 \cdot e^{-Y/1.77}$	0.999

The analysis of the results presented in Table 2 revealed the followings:

÷ Very good relationships in terms of determination coefficient are obtained between Simpson's diversity index and Rényi entropy of order 1, and Rényi entropy of order 1 and 2, respectively.

÷ The relationships are polynomial of second order for both investigated situations (($D(b,Y)$)-($H_1(b,Y)$) and ($H_1(b,Y)$)-($H_2(b,Y)$).

÷ The $H_2(b,Y) = f(H_1(b,Y))$ obtained slightly better performances in terms of determination coefficients compared with $H_1(b,Y) = f(D(b,Y))$ (the maximum value of $R^2 = 0.999989$ – Y=4 compared with $R^2 = 0.999948$ – Y=3). Moreover, the distribution of the determination coefficient

is more narrow for $H_2(b,Y) = f(H_1(b,Y))$ (a difference between maximum and minimum of 0.000129) compared to $H_1(b,Y) = f(D(b,Y))$ (a difference between maximum and minimum of 0.000631).

Figure 5. Relative variation (min=0; max=1) of regression equation coefficients from Table 2 for $H_1 = f(D)$ and $H_2 = f(H_1)$
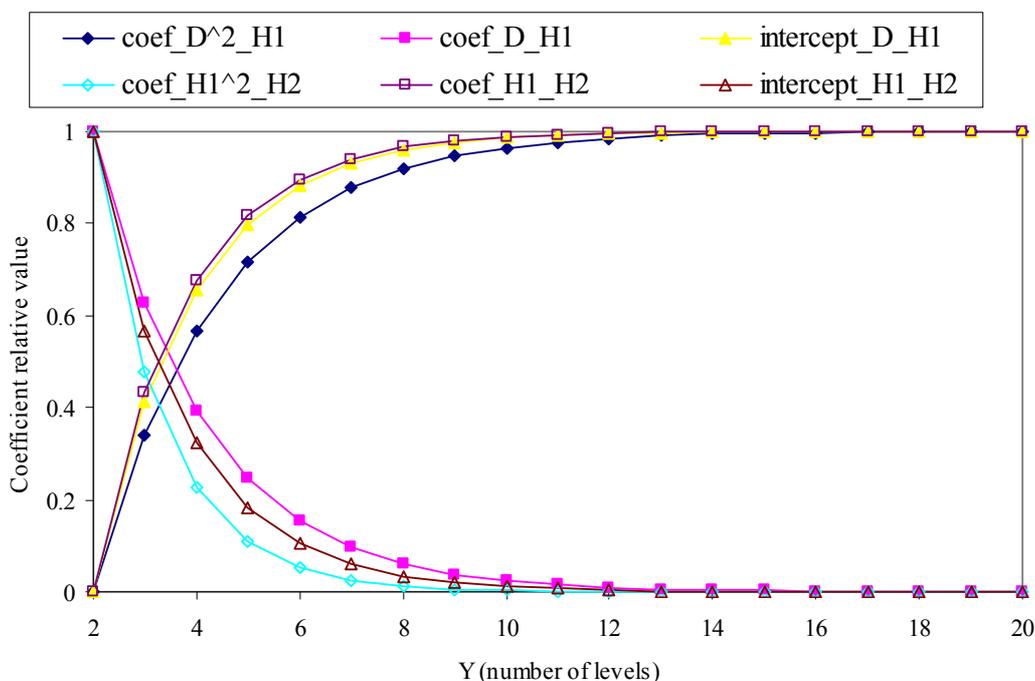

÷ The polynomial of second order dependency of $H_1(b,Y)$ by the $D(b,Y)$ revealed to be done by the same equation if the number of decimals are reduced to two for $Y \geq 13$.

÷ The polynomial of second order dependency of $H_2(b,Y)$ by the $H_1(b,Y)$ revealed to be done by the same equation if the number of decimals are reduced to two for $Y \geq 11$. Moreover, the determination coefficient of 0.999866 was systematically obtained for $Y \geq 16$.

The calculated values of information entropy proved to be negatively correlated for the arbitrary vertex removal from a b-ary tree. The variation of the correlations between these measures of information is graphically presented in Figure 6.

The linear relationship between Rényi's entropy of order 1 (H_1) and 2 (H_2) of b-ary trees is perfect (equal to 1) for b=1 due to the equality of the values of these two parameters. The analysis of the obtained values of correlation coefficients revealed the followings:
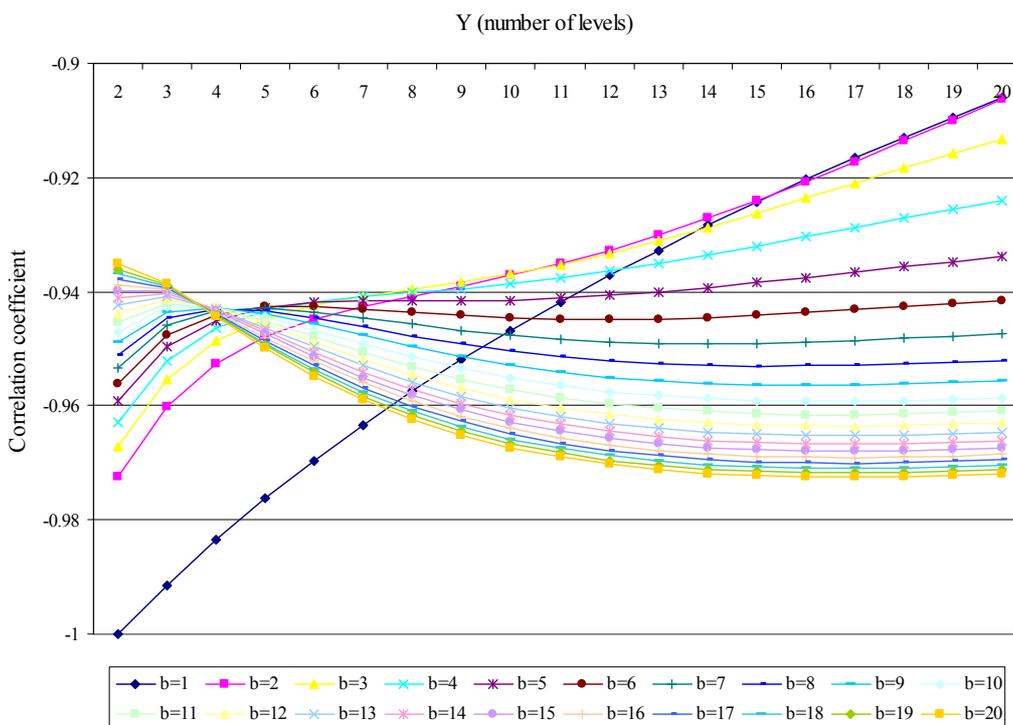
÷ Degree of ramification (b):

  o $r(H_1,H_2)$: the minimum as well as the maximum values of correlation coefficient (expressed as absolute values) increase with increasing of ramification degree. Moreover, a certain pattern of variation could be observed.

  o $r(H_1,D)$: the correlation coefficients analysed as absolute value increase with increasing of ramification degree (the range of this variation is wider (0.036699) compared with the narrow variation of $r(H_1, H_2)$ (0.005394)). The distribution of $r(H_1,D)$ follows two patterns according with the value of b (see Figure 7).

Figure 6. Correlation between Rényi's entropy of order 1 (H_1) and 2 (H_2), and between Rényi's entropy of order 1 (H_1) and Simpson's diversity index (D) of a b-ary tree after a vertex removal
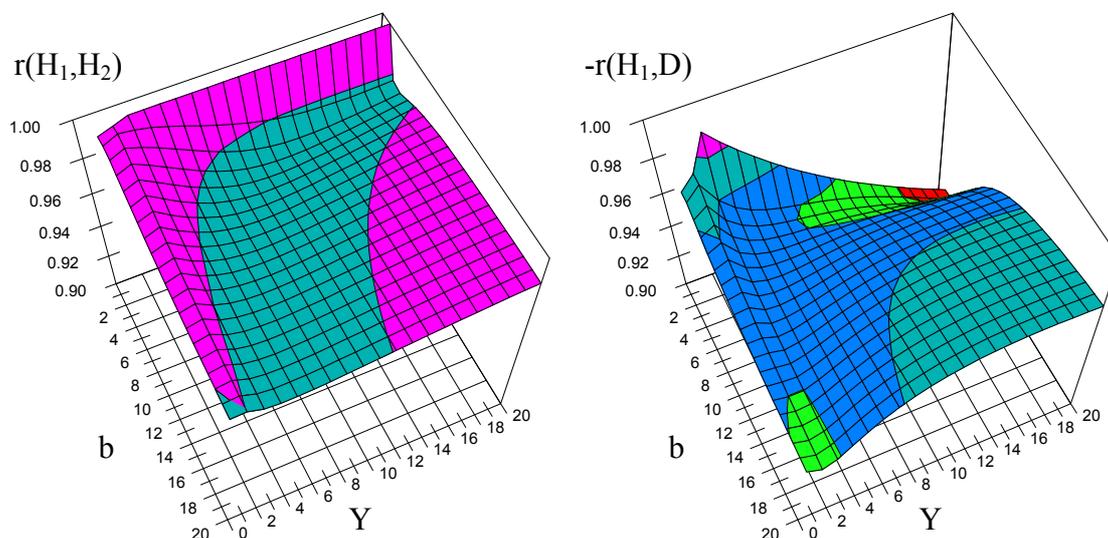
The correlations between H_1 and H_2, and between H_1 and D were calculated by using the following formula:
$r(H_1,H_2) = r(b,Y) = r([H_1(bb,YY)]_{1 \le bb \le b; 1 \le YY \le Y}, [H_2(bb,YY)]_{1 \le bb \le b; 1 \le YY \le Y})$
where [·] denotes a matrix having b columns and Y rows.

Figure 7. Distribution of $r(H_1,D)$ with number of levels (Y)

÷   Number of levels (Y):
- o   The minimum and maximum value (expressed as absolute values) increase with number of levels for $r(H_1,H_2)$ as well as for $r(H_1,D)$.
- o   The maximum value of $r(H_1,H_2)$ is obtained for T(20,20). The highest correlation coefficient

of $r(H_1,D)$ was also obtained by $T(\geq16,20)$.

- o The absolute values of $r(H_1,H_2)$ follows a similar pattern for $Y\leq9$ with a slowly decrease until $b=6$ followed by an increase. For $Y>9$, the absolute value of correlation coefficients increase with increasing of b.
- o The variation of the $r(H_1,D)$ is presented in Figure 8.

Overall, the correlations coefficients revealed to be good, being always higher than 0.9730 ($r(H_1,H_2)$) or smaller than -0.8724 ($r(H_1,D)$).

The linear relationships between H_1 and H_2, as well as between H_1 and D became weaker with the increases of the degree of ramification (b) and of the number of levels (Y). The highest correlation expressed as the absolute value was of $r(H_1,D) = -0.9724$ for $T(2,2)$ and of $r(H_1,H_2) = 0.9996$ for $T(2,2)$).

Figure 8. Variation of $r(H_1,D)$ with the degree of ramification (b)
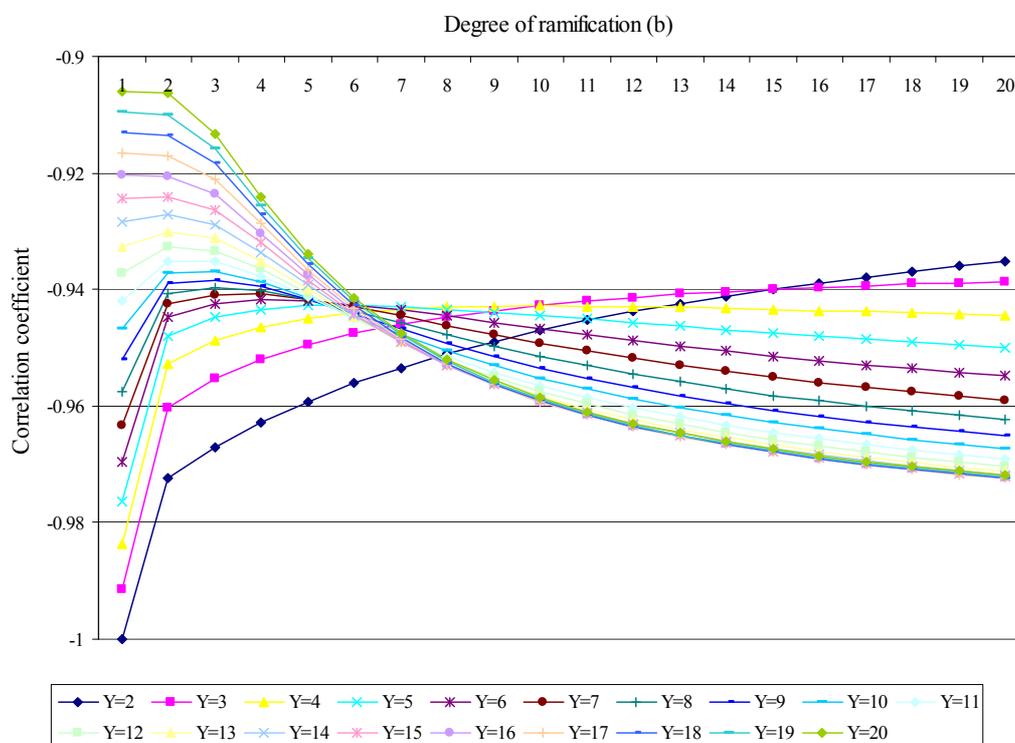
The informational entropy computed on tress found its application on pattern recognition [16], clarification on self-similarity properties of river networks [17], characterization of kinetic equation [18], evaluation of comminution process [19], and so on. Moreover, the informational entropy analysis provide the information about the total loss of information of the investigated structure (the maximum entropy of a system correspond to the total loss of information), asymmetry or symmetry as well as about simplicity [20].

The formulas given by Eq(7) and Eq(8) may be used for larger b and Y as approximation formulas when the b-ary tree is not complete (having few gaps), by using an averaged (approximated) value for its number of levels, since few for larger trees may affect in small amount the entire information contained in the tree. Figures 2 and 6 sustain these assumptions; the entropy of the b-ary tree varies very slowly for larger values of b and Y.

As it could be observed, just the balanced b-ary trees were considered in the present research and this could be a limitation of the study due to the possibility to obtain a generalization. The behaviour obtained in this research on balanced b-ary trees will be assessing in further researches on unbalanced trees (note that the unbalanced trees are particular cases).

Conclusions

The values of Rényi entropy of order 1 are equal with the values of and Rényi entropy of order 2 for a value of the ramification degree of a b-ary tree equal to 1. The values obtained by Rényi entropy of order 1 and 2 decreases with the increases of the ramification degree of the b-ary tree. The Rényi entropy of order 1 and 2 increase with the increases of the number of levels for a give value of the ramification degree of a b-ary tree. The relationship between Rényi entropy of order 1 and 2, and between Rényi entropy of order 1 and Simpson's diversity index revealed to be of order two polynomial types rather than of a linear type.

Acknowledgements

References

1. Carrasco, R. C.; Oncina, J.; Calera-Rubio, J. Stochastic inference of regular tree languages. *Mach. Learn.* **2001**, *44*, 185-197.
2. Hu, Y.; Shi, Z. A subdiffusive behaviour of recurrent random walk in random environment on a regular tree. *Probab. Theory. Relat. Fields.* **2007**, *138*, 521-549.
3. Calera-Rubio, J.; Carrasco, R. C. Computing the relative entropy between regular tree languages. *Inform. Process. Lett.* **1998**, *68*, 283-289.
4. Mossel, E.; Peres, Y. Information flow on trees. *Ann. Appl. Probab.* **2003**, *13*, 817-844.
5. Aldous, D. J. Random walk covering of some special trees. *J. Math. Anal. Appl.* **1991**, *157*, 271-283.
6. Jäntschi, L.; Stoenoiu, C. E.; Bolboacă, S. D. A Formula for Vertex Cuts in b-Trees. *Int. J. Pure App. Math.* **2008**, *47*, 17-22.
7. Estrada, E.; Hatano, N. Statistical-mechanical approach to subgraph centrality in complex networks. *Chem. Phys. Lett.* **2007**, *439*, 247-251.
8. Shannon, C. E. A Mathematical Theory of Communication. *Bell. Syst. Techn. J.* **1948**, *27*, 379-423, 623-656.
9. Onicescu, O. Energie informationelle. *Compt. Rend. Acad. Sci. Paris Ser.* **1966**, *263*, 841-841.
10. Simpson, E. H. Measurement of Diversity. *Nature* **1949**, *163*, 688-688.
11. Bolboacă, S. D.; Pică, E. M.; Cimpoiu, C. V.; Jäntschi, L. Statistical Assessment of Solvent Mixture Models used for Separation of Biological Active Compounds. *Molecules* **2008**, *8*, 1617-1639.

12. Friedman, W. F. *The Index of Coincidence and Its Applications in Cryptography*; Riverbanks Publication: Geneva, Switzerland, Illinois, USA, 1922; 22.

13. Gass, F. Solving a Jules Verne Cryptogram. *Mathem. Mag.* **1986**, *59*, 3-11.

14. Harremoes, P.; Topsoe, F. Inequalities between entropy and index of coincidence derived from information Diagrams. *IEEE Trans. Inform. Theory* **2001**, *47*, 2944-2960.

15. Rényi, A. On measures of information and entropy. *Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability*, **1961**, 547-561.

16. Jumarie, G. Informational similarity of graphs in syntactic pattern recognition. *Pattern Recogn. Lett.* **1994**, *15*, 1177-1181.

17. Claps, P.; Fiorentino, M.; Oliveto, G. Informational entropy of fractal river networks. *J. Hydrol.* **1996**, *187*, 145-156.

18 Masiuk, S.; Rakoczy, R. Kinetic equation of grinding process in mixing of granular material using probability density functions, transient operators and informational entropy. *Chem. Eng. Process* **2008**, *47*, 200-208.

19. Otwinowski, H. Energy and population balances in comminution process modelling based on the informational entropy. *Powder Technol.* **2006**, *167*, 33-44.

20. Lin, S-K. Correlation of Entropy with Similarity and Symmetry. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 367-376.