ENTROPY AND ENERGY OF SUBSTRUCTURES OBTAINED BY VERTEX CUTTING IN REGULAR TREES

LORENTZ JÄNTSCHI AND SORANA-DANIELA BOLBOACĂ

The entropy (a quantitative measure of disorder in a system) and informational energy (informational "disorder") of substructures obtained by cutting the vertex in regular trees was investigated and is presented. In a regular tree every vertex has the same number of children and leafs had no children at all. The information energy was defined as Energy $= \sum p_i^2$, where $p_i =$ the probability of apparition of a substructure of i size. The entropy was defined as Entropy $= -p_i \cdot log_2 \cdot p_i$, where p_i has the signification described above. Regarding the entropy the following remarks can be done: (a) the entropy decrease with ramification; (b) the entropy increase with increasing of the number of levels; and (c) the decreasing with ramification is more accentuate compared with the increasing of the number of levels. Regarding the information energy a decrease with the decrease of ramification and with the increase of number of levels was observed.

References

[1] Jäntschi L, Diudea M V. Subgraphs of pair vertices. J Math Chem DOI:10.1007/s10910-008-9411-6.
[2] Bolboacă S D, Jäntschi L. How good can the characteristic polynomial be for correlations? Int J Mol Sci 2007;8(4):335-345.
[3] Jäntschi L. Characteristic and Counting Polynomials of Nonane Isomers. AcademicDirect, Internet. 2007, p.101.

Technical University of Cluj-Napoca, Romania, 15 Constantin Daicoviciu Street, 400020 Cluj-Napoca, Romania.
E-mail address: lori@j.academicdirect.org
URL: http://lori.academicdirect.org

"Iuliu Hatieganu" University of Medicine and Pharmacy Cluj-Napoca, 13 Emil Isac Street, 400023 Cluj-Napoca, Romania.
E-mail address: sbolboaca@umfcluj.ro
URL: http://sorana.academicdirect.ro

1991 *Mathematics Subject Classification.* 05C05 (Trees); 94A17 (Measures of Information, Entropy); 60B99 (Entropy, Energy).
Key words and phrases. Entropy; Informational Energy; Regular Trees.

ENTROPY AND ENERGY OF SUBSTRUCTURES OBTAINED BY VERTEX CUTTING IN REGULAR TREES

Lorentz JÄNTSCHI
and
Sorana-Daniela BOLBOACĂ

Technical University of Cluj-Napoca
and
"Iuliu Hațieganu" University of Medicine and Pharmacy Cluj-Napoca,
Romania

Having a regular b-tree with Y levels (Figure 1) the removal of a arbitrary vertex from the tree will generate from one (when removal are applied to a leaf) to two (when removal are applied to the root), and to three (when removal are applied to a inside node) sub-graphs.

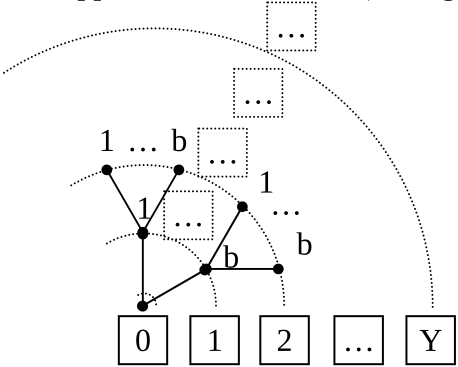


Figure 1. A b-Tree with Y levels, T(b,Y)

By removing once at the time every vertex from the tree, it is possible to evaluate the number of sub-graphs by size which may be obtained.

Somebody may say that systematically removing 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 removed is unknown. One example may be then having a network topology like a tree (let us say a Regional Internet Registry providing Internet resource allocations). In these cases assessing of the impact when a node is removed it is an important issue for projecting the network. 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 of the appearance probability of a set of a given size). In terms of the entire structure (the tree), even the appearance probability of a set of a given size may not be a relevant information and we may want to obtain a global parameter which to characterize the tree under the given circumstances (removal of one vertex).

Two parameters are often used for characterization of the mess: the information entropy (or Shannon's entropy [1], a measure of the uncertainty associated with a random variable) and information energy (Onicescu's energy [2], a supplement and a complement of the Shannon's entropy, a formula obtained complementing to 1 the Simpson's measure of diversity [3]). Both measures proved to be very useful for characterizing physical and chemical processes [4].

---

[1] Claude E. SHANNON, A Mathematical Theory of Communication, Bell System Technical Journal, 1948, 27:379-423, 623-656.

[2] Octav ONICESCU, Energie informationelle, Comptes Rendus de la Academie des Sciences Paris Ser. A, 1966, 263:841-841.

[3] E. H. SIMPSON, Measurement of Diversity, Nature 1949, 163(4148):688-688.

[4] Sorana D. BOLBOACĂ, Elena M. PICĂ, Claudia V. CIMPOIU, LORENTZ JÄNTSCHI, Statistical Assessment of Solvent Mixture Models Used for Separation of Biological Active Compounds, Molecules, 2008, 13(8):1617-1639.

Let us define these two measures of information. Let be X a discrete variable taking the values $\{x_1, \ldots, x_n\}$ and $p(\cdot)$ the probability mass function (gives the probability that a discrete random variable is exactly equal to a value).

The Shannon's entropy H(X) is given by (where logarithm is taken to base 2 to give a value in bits):

$$H(X) = - \Sigma_{1\leq i\leq n} p(x_i)\log_2(p(x_i)) \qquad \text{(eq1)}$$

The Onicescu's energy E(X) is given by:

$$E(X) = \Sigma_{1\leq i\leq n} p^2(x_i) \qquad \text{(eq2)}$$

Let us recall the structure from figure 1. The following formula (NSSP) describes completely the number of substructures (coefficients of the X variable) by sizes (powers of the X variable):

$$\mathrm{NSSP}(T(b,Y)) = \sum_{k=1}^{Y} b^k \left(X^{\frac{b^{Y+1-k}-1}{b-1}} + X^{\frac{b^{Y+1}-b^{Y+1-k}}{b-1}} \right) + \frac{b^{Y+1}-1}{b-1} X^0 \qquad \text{(eq3)}$$

The formula (eq3) give the complete description of the substructures counts and sizes for any Y,b>0. The following table (Table 1) gives 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 by removal of an arbitrary vertex for some regular 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

Recalling

$$\mathrm{NSSP}(T(b,Y)) = \sum_{k=1}^{Y} b^k \left(X^{\frac{b^{Y+1-k}-1}{b-1}} + X^{\frac{b^{Y+1}-b^{Y+1-k}}{b-1}} \right) + \frac{b^{Y+1}-1}{b-1} X^0 \qquad \text{(eq3)}$$

We may want to find roots of:

$$b^{Y+1-k_1} - 1 = b^{Y+1} - b^{Y+1-k_2} \text{ i.e. } 1 + b^{k_2-k_1} = b^{k_2} + b^{k_2-(Y+1)} \qquad \text{(eq4)}$$

Excluding b=1, since b, Y, k_1, and k_2 are natural not null numbers, and natural too are 1 and b^{k_2} the equality (eq4) can be satisfied only for $b^{k_2-k_1} = b^{k_2-(Y+1)}$ which implies that the only solutions of (eq5) is for k_1=Y+1 (implying as consequence b=1). The conclusion that can be drawn from here is that for b>1 all terms inside the sum from (eq3) are distinct.

Since the problem of sub-graphs sizes occurrences (n_o) was solved in the general case for b>1 (no repeats in terms of eq3) we are able to calculate it for a given Y and a given b (note that the problem can be extended to any b less or equal to a given value and for a Y less or equal to a given value):

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

(eq5)

It follows that probabilities are given by:

$$p\left(\frac{b^{Y+1-k}-1}{b-1}\right)=p\left(\frac{b^{Y+1}-b^{Y+1-k}}{b-1}\right)=\frac{b^k}{2\sum_{k=1}^{Y}b^k}=\frac{b^k(b-1)}{2(b^Y-1)} \quad \text{(eq6)}$$

Replacing of (eq6) into (eq1) and (eq2) is now only a matter of calculation:

$$H(T(b,Y))=-2\sum_{k=1}^{Y}\frac{b^k(b-1)}{2(b^Y-1)}\log_2\left(\frac{b^k(b-1)}{2(b^Y-1)}\right)=\sum_{k=1}^{Y}\frac{b^k(b-1)}{(b^Y-1)}\log_2\left(\frac{2(b^Y-1)}{b^k(b-1)}\right)$$

(eq7)

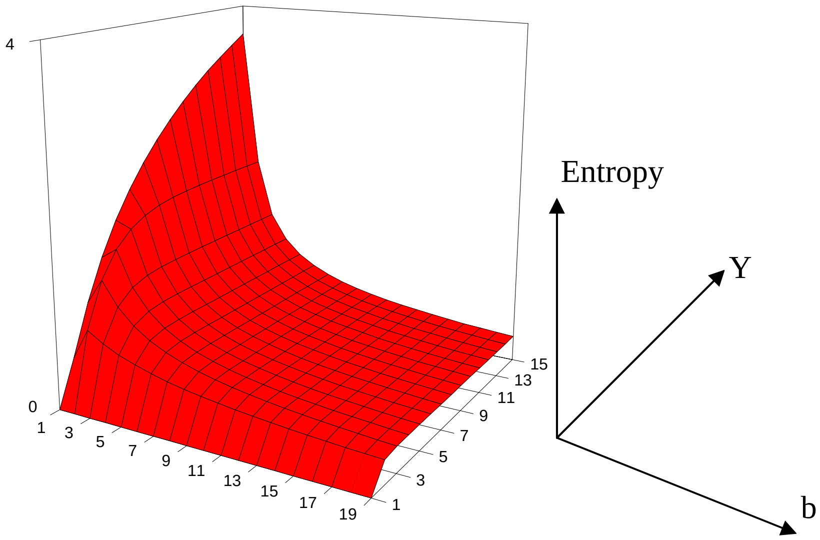
$$E(T(b,Y))=2\sum_{k=1}^{Y}\frac{b^{2k}(b-1)^2}{4(b^Y-1)^2}=\frac{b^2}{2}\frac{b^Y+1}{b^Y-1}\frac{b-1}{b+1} \quad \text{(eq8)}$$

where H(T(b,Y)) is the Shanon's entropy and E(T(b,Y)) is the Onicescu's energy of removal of a vertex from a T(b,Y) tree, when b>1.

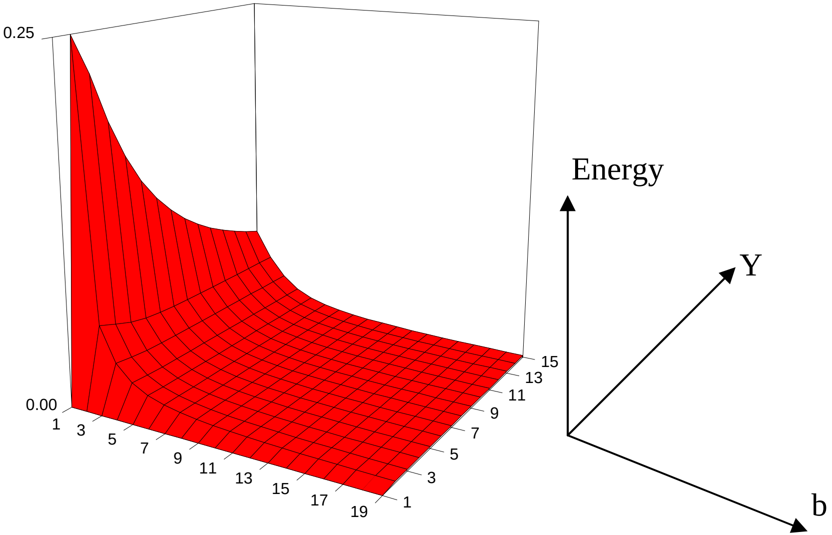
For b=1 (path) a entry from Table 1 can help us to obtain the values for H(T(1,Y))=H(P(Y)) and E(T(1,Y))=H(P(Y)):

$$H(P(Y)) = -\sum_{k=1}^{Y}\frac{2}{2Y}\log_2\left(\frac{2}{2Y}\right) = -\sum_{k=1}^{Y}\frac{1}{Y}\log_2\left(\frac{1}{Y}\right) = \log_2 Y \qquad \text{(eq9)}$$

$$E(P(Y)) = \sum_{k=1}^{Y}\left(\frac{2}{2Y}\right)^2 = \frac{1}{Y} \qquad \text{(eq10)}$$



4
0
1
3
5
7
9
11
13
15
17
19
1
3
5
7
9
11
13
15
Entropy
Y
b


0.25
0.00
1 3 5 7 9 11 13 15 17 19
1 3 5 7 9 11 13 15
Energy
Y
b