Surveys in Mathematics and its Applications

ISSN 1842-6298 (electronic), 1843-7265 (print)
Volume **4** (2009), 169 – 177

ENTROPY DUE TO FRAGMENTATION OF DENDRIMERS

Lorentz Jäntschi and Sorana D. Bolboacă

Abstract. Subgraphs can results through application of criteria based on matrix which characterize the entire graph. The most important categories of criteria are the ones able to produce connected subgraphs (fragments). Based on theoretical frame on graph theory, the fragmentation algorithm on pair of vertices containing the largest fragments (called MaxF) are exemplified. The counting polynomials are used to enumerate number of all connected substructures and their sizes. For a general class of graphs called dendrimers general formulas giving counting polynomials are obtained and characterized using informational measures.

1 Introduction

Molecular Topology creates one of the most important junction between Graph Theory and Organic Chemistry. Some books were published in this interdisciplinary field, just one of them being [3]. Two main operational tools on this field are square matrices (the cell of matrix contains a graph theoretical property associated to the pairs of vertices) and graph polynomials. Counting polynomials were recently introduced as a tool for substructures count and sizes for a given fragmentation criteria [5] applied on a graph-type structure. The paper presents the counting polynomial formulas obtained on a regular type structure - dendrimers by applying of a maximum fragment size algorithm. The counting polynomials were further investigated from the informational perspective when another series of entropy and energy formulas of polynomials were obtained. Finally, a structure - activity relationships study was conducted using the entropy and the energy of a polynomial formula as structure descriptors.

2000 Mathematics Subject Classification: 05C10; 11T06.
Keywords: Graph theory; Subgraphs; Graph polynomials; Entropy.
This work was supported by UEFISCSU Romania through grants (ID0458/206/2007 and ID1051/202/2007).

2 Graph Theory Fundamentals

Let V be a set and $E \subseteq V \times V$ a subset of the Cartesian product $V \times V$. Then $G = (V, E)$ is an un-oriented graph. A graph is connected if there is a path from one to any other vertex. Let G = (V, E) be an un-oriented *connected* graph. We will note with $V(G)$ the set of vertices from G, with $E(G)$ the set of edges from G, and with $d(G)$ the distance matrix of G. The element $d(G)_{i,j}$ of the distance matrix $d(G)$ is the length of a shortest path joining veritices i and j, if any: $min(l(p_{i,j}))$ otherwise ∞. In terms of distance matrix $d(G)$ the connectivity of G is translated in: $d(G)_{i,j} < \infty$ for any $i, j \in V(G)$. The largest subgraph obtained from G containing the i vertex and not containing the j vertex is considered [7] to be a maximal connected subgraph (relative to i and j vertices). Let us call this subgraph $MaxF(G)_{i,j}$. Maximal connected subgraphs can be defined through construction (2.1).

$$\begin{aligned} MaxF(G)_{i,j} &= (V(MaxF(G)_{i,j}), E(MaxF(G)_{i,j})) \\ VTmp(G)_{i,j} &= \{s \in V(G) \mid s \neq j\} \\ ETmp(G)_{i,j} &= \{(u,v) \in E(G) \mid u, v \neq j\} \\ V(MaxF(G)_{i,j}) &= \{s \in VTp(G)_{i,j} \mid d(VTmp(G)_{i,j})_{s,i} < \infty\} \\ E(MaxF(G)_{i,j}) &= \{(s,t) \in E(G) \mid s, t \in V(MaxF(G)_{i,j})\} \end{aligned} \tag{2.1}$$

The counting matrices containing the number of subgraphs vertices are formally known as pair-based matrices in graphs. Note that, in general these are unsymmetrical matrices. We will note the associated matrix using brackets $[\cdot]$. The definition for $[MaxF]$ is given by (2.2):

$$[\mathrm{MaxF}]_{i,j} = |MaxF(G)_{i,j}| \tag{2.2}$$

3 Counting Polynomials

A counting polynomial for a graph structure is a polynomial formula depending on the structure, and on the applied criteria; it is frequently expressed using an undetermined variable X. Counting polynomials can be defined and investigated based on square matrices as well as the fragmentation criteria defined above. For matrices (as above defined, (2.2), we have:

$$CP(G, [\mathrm{C}], X) = \sum_{i \geqslant 0} coef(G, i, [\mathrm{C}])X^i \tag{3.1}$$

In (3.1) $coef(G, i, [\mathrm{C}])$ counts the number of matching of value i in [C] matrix, [C] is the unsymmetrical matrix constructed from the criterion C. We used as C the abode-defined $MaxF$ criterion (2.1). For a connected graph G, always $coef(G, 0, [C]) = |V(G)|$ for any discussed criterion C. If someone counts the number

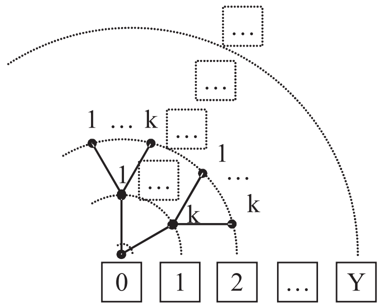


Figure 1: Dendrimers, $D_{k,Y}$

of zero's in the matrix defined by (2.2) for the dendrimer (see Figure 1), it will found the formula giving the X^0 term in counting polynomials (left over for the further analysis), this is equal to the number of vertices (due to the 0's from the diagonal of the matrix), and are given by (3.2).

$$\frac{k^{Y+1}-1}{k-1} \tag{3.2}$$

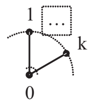
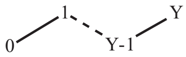
Taking here a probabilistic approach, the polynomial formula characterizes the entire graph though its monomes and every monome describes the results of an independent event on which counting operation is applied. Thus, a rank of a monome express the magnitude of the event - the size of the subgraphs obtained from the graph structure by applying the criteria, while a coefficient of a monome express the frequency of an event - number of subgraphs obtained from the graph structure having same size by applying the criteria. Obtaining of polynomial formulas have important applications in chemical graph theory and structure-activity relationships [1, 6]. Thus, information theory approaches on counting polynomials may have important application in these fields. The most complex task is obtaining of polynomials formulas. The definition formula of counting polynomials found some applications in investigation of structure-activity relationships [6], characterization of nanostructures [4], investigation of indeterminate over a finite field and with bounded degree polynomials [8], and others in [12]. The formula 3.1 for counting polynomial $CP_{MaxF}(k,Y,X)$ due to $MaxF$ fragmentation is as follows (3.3):

$$CP_{MaxF}(k,Y,X) = \sum_{i=1}^{Y}\left(\frac{k^{Y+1-i}-1}{(k-1)/k^{Y+1}}X^{\frac{k^{Y+1-i}-1}{(k-1)/k^{i}}} + \frac{k^{Y+1-i}-1}{(k-1)/k^{i}}X^{\frac{k^{Y+1-i}-1}{k-1}}\right) \tag{3.3}$$

4 Proof of the Counting Polynomial Formula

In order to proof the polynomial formula given above, is enough to see that is true for a large enough particular cases. The $D_{k,Y}$ dendrimer (see figure 1) has two important particular cases: $k = 1$ - when become a path and $Y = 1$ - when become a star. Table 1 contains the polynomial formulas for stars and paths, as is are obtained from (3.3).

Table 1: Degeneration of dendrimers into stars (S_k) and paths (P_Y)

Graph		Polynomial
	S_k	$CP_{MaxF}(k,1,X) = k^2X^k + kX$
	S_1	$CP_{MaxF}(1,1,X) = 2X$
	S_2	$CP_{MaxF}(2,1,X) = 4X^2 + 2X$
	S_3	$CP_{MaxF}(3,1,X) = 9X^3 + 3X$
	P_Y	$CP_{MaxF}(1,Y,X) = \frac{YX^{Y+2}-(Y+1)X^{Y+1}+X}{(X-1)^2/2}$
	P_1	$CP_{MaxF}(1,1,X) = 2X$
	P_2	$CP_{MaxF}(1,2,X) = 4X^2 + 2X$
	P_3	$CP_{MaxF}(1,3,X) = 6X^3 + 4X^2 + 2X$

There are just four distinct structures between (S_k, P_Y) for $k, Y = 1..3$ as formulas from Table 1 shows too ($S_1 = P_1$; $S_2 = P_2$ after reindexing of vertices).
Starting to check the formulas given in Table 1 no proof is necessary to see that $\{\{0,1\},\{(0,1)\}\}$ graph (S_1 and P_1 in Table 1) has only $MaxF_{0\ vs.\ 1} = \{0\}$ and $MaxF_{1\ vs.\ 0} = \{1\}$ as fragments and their count gives the $2X$ polynomial formula.
For S_2 star ($S_2 = \{\{0,1,2\},\{(0,1),\ (0,2)\}\}$), $MaxF_{0\ vs.\ 1} = \{0,2\}$, $MaxF_{0\ vs.\ 2} = \{0,1\}$, $MaxF_{1\ vs.\ 0} = \{1\}$, $MaxF_{2\ vs.\ 0} = \{2\}$, $MaxF_{1\ vs.\ 2} = \{0,1\}$, and $MaxF_{2\ vs.\ 1} = \{0,2\}$. Counting gives four occurences of size 2 and two occurences of size 1, thus the polynomial is consistent with the formula given in Table 1, $4X^2 + 2X$.
Taking now the general case S_k, fragments of size 1 are $MaxF_{i\ vs.\ 0} = \{i\}$, $i = 1..k$ and all others (in number of ${}_{k+1}P_2 - k = k^2$) gives fragments of size k and $CP_{Max}(k,1,X)$ from (3.3) is thus verified.
The formula for $CP_{MaxF}(1,Y,X)$ (3.3) is the compacted expression for the progression observed in the series $CP_{MaxF}(1,1,X)$, $CP_{MaxF}(1,2,X)$, $CP_{MaxF}(1,3,X)$, ... and it correspond to the $2X\sum_{i=1}^{Y} iX^{i-1}$ series sum, and

$$CP_{MaxF}(1,Y+1,X) = CP_{MaxF}(1,Y,X) + 2(Y+1)X^{Y+1}$$

occures. Listing of $MaxF$ fragments for S_{Y+1} of size $Y+1$ has the following result: $MaxF_{Y+1\ vs.\ Z}$, $Z = 0..Y$ and $MaxF_{Z\ vs.\ Y+1}$, $Z = 0..Y$ - all others being of size less or equal to Y, and the proof of $CP_{MaxF}(1,Y,X)$ (3.3) is thus given by recurrence.
Let's take now a more complex graph (Figure 2) to count it's fragments by size after applying of $MaxF$ criterion. The $MaxF$ fragments of $D_{2,2}$ may be ennumerated as follows (using notations from Figure 2):

- All from and to $v_{1,\cdot}$ and v_0: $k|MaxF_{v_{1,\cdot}}\ vs.\ v_0| = k\sum_{i=0}^{Y-1} k^i = 2X^3$; $k|MaxF_{v_0\ vs.\ v_{1,\cdot}}| = k\sum_{i=0}^{Y} k^i - k\sum_{i=0}^{Y-1} k^i = k^Y = 2X^4$;

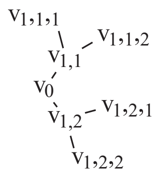


Figure 2: $D_{2,2}$ Dendrimer

- All from and to $v_{1,\cdot,\cdot}$ and v_0: $k^2|MaxF_{v_{1,\cdot,\cdot}}\ vs.\ v_0| = k^2\sum_{i=0}^{Y-1} k^i = 4X^3$; $k^2|MaxF_{v_0\ vs.\ v_{1,\cdot,\cdot}}| = 4X^6$;
- All from $v_{1,i}$ to $v_{1,j}$ $(i \neq j)$: ${}_2P_k|MaxF_{v_{1,i}}\ vs.\ v_{1,j}| = 2X^4$;
- All from and to $v_{1,i}$ to $v_{1,j,\cdot}$ $(i \neq j)$: $2 \cdot 2X^6$; $2 \cdot 2X^4$;
- All from and to $v_{1,i}$ to $v_{1,i,\cdot}$: $2 \cdot 2X^6$; $2 \cdot 2X^1$;
- All from $v_{1,\cdot,i}$ to $v_{1,\cdot,j}$ $(i \neq j)$: $4 \cdot 3X^4$;

Cumulating all above enumerated, $24X^6 + 8X^4 + 6X^3 + 4X$ which is consistent with $CP_{MaxF}(2,2,X)$ - (3.3).

5 Informational Analysis

Let a counting polynomial be given by the general formula:

$$P = \sum_{1\leq i\leq n} a_i X^{b_i},\ A = (a_i)_{1\leq i\leq n},\ B = (b_i)_{1\leq i\leq n} \tag{5.1}$$

The meaning of a monome from (5.1) is as follows: the coefficient a_i is the frequency of a fragment (connected subgraph) of size b_i. The polynomial formula defines the frequency space A and events space B. This formula allows obtaining entropies (from A) and energies (from both A and B) of the polynomial formula. The Rényi [9] entropy H_α is defined by the formula (5.2):

$$H_\alpha = \frac{1}{1-\alpha} log_2 \sum_{i=1}^{n} p_i^\alpha,\ p_i = \frac{a_i}{\sum_{j=1}^{n} a_j},\ \alpha > 0 \tag{5.2}$$

$$H_1 = \lim_{\alpha\to 1} H_\alpha = -\sum_{i=1}^{n}(p_i log_2(p_i)) \tag{5.3}$$

$$H_2 = -log_2(\sum_{i=1}^{n}(p_i^2)) \tag{5.4}$$

where the entropies were expressed in bits. Note that Shannon's [10] entropy $H_1(P)$ ($\alpha = 1$, (5.3) is the one related with Boltzmann entropy [2], thus giving a physical meaning of the results, and the term inside of the logarithm in H_2 (5.4) is the Simpson Diversity Index [11], giving thus a biological meaning of the results.
The entropy of a polynomial formula is defined by the frequency of apparition of molecular fragments of specified size and it is comprises in the coefficients of the polynomial formula.
By applying of the (5.3) and (5.4) formulas to the coefficients (see 5.1 and 5.2) of the polynomial formula of the dendrimers (3.3) two variable (k and Y) Shannon (5.3) and -log(Simpson) (5.4) entropies are obtained. Figure 3 plots the Shannon entropy (in bits) where Y parameter is on horizontal axis and k parameter is on frontal axis and the Shannon entropy is on the surface. Both entropies rapidly decreases with

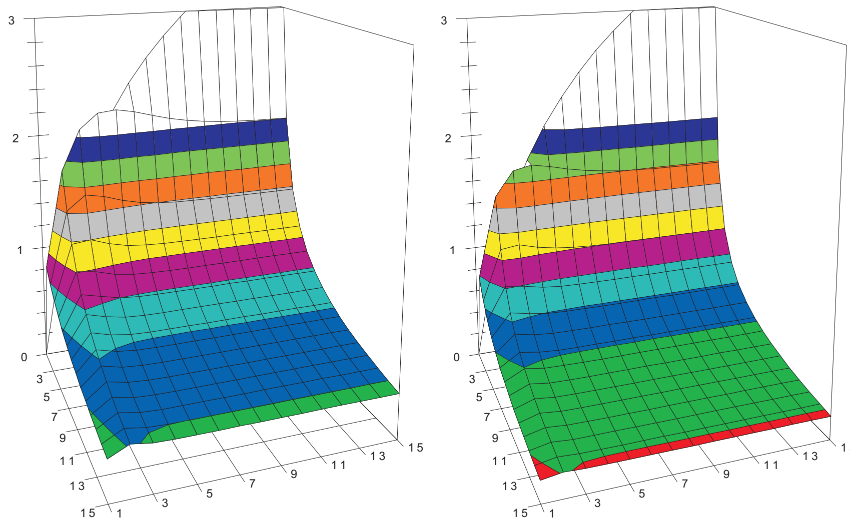


Figure 3: Shannon (5.3) entropy (left) and -log(Simpson) (5.4) entropy (right) of $D_{k,Y}$ due to $MaxF$ (2.1 then (3.3) fragmentation

ramification degree. The coordinates of highest entropies by the number of levels are given by following coordinates (k, Y):

1. For Shannon entropy (5.3): $(1, \infty)$ - infinite path, $(2, 5)$ - 5 levels binary tree, $(3, 3)$ - three levels tertiary tree, $(4, 3)$ - tree lavels ternary tree, $(k \geq 5, 2)$ - two levels any other tree;

2. For -log(Simpson) entropy (5.4): $(1, \infty)$ - infinite path, $(2, 4)$ - 4 levels binary tree, $(3, 3)$ - three levels tertiary tree, $(k \geq 4, 2)$ - two levels any other tree;

Between stars, the path-star $S_2 = P_2$ has the highest entropy on both measures of (5.3) and (5.4).

6 Appendix

In order to test the validity of the formula (3.3) a program implementing equations (2.1) and (2.2) were made. Table 2 contains obtained results. The supplementary file (DkY.chm file) contains plots, distance matrix, maximal fragments and their polynomials for following dendrimers: $D_{2,1}$, $D_{3,1}$, $D_{4,1}$, $D_{5,1}$, $D_{2,2}$, $D_{3,2}$, $D_{4,2}$, $D_{5,2}$, $D_{2,3}$, $D_{3,3}$, and $D_{2,4}$. All obtained results were consistent with the (3.3) formula. Note that in the supplementary file and in Table 2 are counted the number of the verices from initial graph too (3.2) through X^0 terms of the polynolials.

Table 2: $\sum_i a_i X^{b_i}$ counting polynomials of maximal fragments ($MaxF$) for firsts $D_{k,Y}$ dendrimers

k,Y	i	12	11	10	9	8	7	6	5	4	3	2	1	0
2,1	B											2	1	0
2,1	A											4	2	3
2,2	B									6	4	3	1	0
2,2	A									24	8	6	4	7
2,3	B							14	12	8	7	3	1	0
2,3	A							112	48	16	14	12	8	15
2,4	B					30	28	24	16	15	7	3	1	0
2,4	A					480	224	96	32	30	28	24	16	31
2,5	B			62	60	56	48	32	31	15	7	3	1	0
2,5	A			1984	960	448	192	64	62	60	56	48	32	63
2,6	B	126	124	120	112	96	64	63	31	15	7	3	1	0
2,6	A	8064	3968	1920	896	384	128	126	124	120	112	96	64	127
3,1	B											3	1	0
3,1	A											9	3	4
3,2	B									12	9	4	1	0
3,2	A									108	27	12	9	13
3,3	B							39	36	27	13	4	1	0
3,3	A							1053	324	81	39	36	27	40
3,4	B					120	117	108	81	40	13	4	1	0
3,4	A					9720	3159	972	243	120	117	108	81	121
4,1	B											4	1	0
4,1	A											16	4	5
4,2	B									20	16	5	1	0
4,2	A									320	64	20	16	21
4,3	B							84	80	64	21	5	1	0
4,3	A							5376	1280	256	84	80	64	85

Acknowledgement. The authors would like to thank Prof. M. V. Diudea for the helpful discussions related to counting polynomials.

References

[1] S. D. Bolboacă and L. Jäntschi, *How Good the Characteristic Polynomial Can Be for Correlations?*, International Journal of Molecular Sciences, **8** (2007), 335-345. DOI 10.3390/i8040335.

[2] L. Boltzmann, *Weitere Studien über das Wärmegleichgewicht unter Gasmolekülen*, Wiener Berichte, **66** (1872), 275-370. JFM 04.0566.01.

[3] M. V. Diudea, I. Gutman and L. Jäntschi, *Molecular Topology*, Nova Science Publishers, 2001, LoC 2001031282.

[4] M. V. Diudea and C. L. Nagy, *Counting polynomials on nanostructures. In: Periodic Nanostructures*. Series: Developments in Fullerene Science, **7** (2007), 69-114. DOI 10.1007/978-1-4020-6020-5_4.

[5] M. V. Diudea, A. E. Vizitiu and D. Janezic, *Cluj and related polynomials applied in correlating studies*, Journal of Chemical Information and Modeling, **47** (2007), 864-874. DOI 10.1021/ci600482j.

[6] L. Jäntschi, S. D. Bolboacă and C. M. Furdui, *Characteristic and Counting Polynomials: Modelling Nonane Isomers Properties*, Molecular Simulation, **35** (2009), 220-227. DOI 10.1080/08927020802398892.

[7] L. Jäntschi and M. V. Diudea, *Subgraphs of pair vertices*, Journal of Mathematical Chemistry, **45** (2009), 364-371. MR2470467. Zbl 1171.05433.

[8] J. F. Ragot, *Counting Polynomials with Zeros of Given Multiplicities in Finite Fields*, Finite Fields and Their Applications, **5** (1999), 219-231. MR1702893. Zbl 1024.11076.

[9] A. Rényi, *On measures of information and entropy*, Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability, **1** (1961), 547-561. MR0132570(24 #A2410). Zbl 0106.33001.

[10] C. E. Shannon, *A Mathematical Theory of Communication*, Bell System Technical Journal, **27**(1948), 379-423 and 623-656. MR0026286(10,133e). Zbl 1154.94303.

[11] E. H. Simpson, *Measurement of Diversity*, Nature **163** (1949), 688-688. Zbl 0032.03902.

[12] R. P. Stanley, *Enumerative Combinatorics (I and II)*, In: Cambridge Studies in Advanced Mathematics, **49** and **62**, 1997 and 1999. MR1442260(98a:05001) and MR1676282(2000k:05026). Zbl 0945.05006 and Zbl 0978.05002.

Lorentz Jäntschi
Faculty of Materials Science and Engineering, Technical University of Cluj-Napoca,
400641 Cluj, Romania.
e-mail: lori@academicdirect.org
http://lori.academicdirect.org

Sorana D. Bolboacă
Department of Medical Informatics and Biostatistics,
"Iuliu Haţieganu" University of Medicine and Pharmacy,
400349 Cluj, Romania.
e-mail: sorana@j.academicdirect.org
http://sorana.academicdirect.ro