MATCH
Communications in Mathematical
and in Computer Chemistry

MATCH Commun. Math. Comput. Chem. **85** (2021) 275-284

ISSN 0340 - 6253

Topological Structure of Extremal Graphs on the First Degree–Based Graph Entropies

Jingzhi Yan

School of Information Science and Engineering
Lanzhou University, Lanzhou 730000, PR China
yanjingzhi@lzu.edu.cn

(Received July 21, 2020)

Abstract

Recently different entropy-based measures are employed to determine the complexities of relational structures. In this paper, we investigate the extremal properties of the first degree-based graph entropy for general graphs and characterize the structure of extremal graphs, which lead to a method of obtaining extremal graphs for any given (n, m) graph. With this method, the minimum value of the entropy for graphs with fewer cycles and the maximum value for general graphs are given. Finally, we make a conjecture on the minimum value for any (n, m) graph.

1 Introduction

In many fields, including biology, computer science, and chemistry, graph entropy is a powerful tool for the analysis of complexities of relational structures [14–16]. Since graph entropy plays an increasingly important role in a variety of problem areas, different graph measures based on Shannon entropy have been developed to characterize the structure of graphs, involving basic invariants such as vertices, edges, degrees, and distances. More information on the topic can be found in [8].

One of the highlights is Dehmer's graph entropies [6, 7] based on information functionals, which has inspired a lot of important research findings in information science, graph theory, and network science. Cao, Dehmer, and Kang [2] introduced graph entropy measures based on independent sets and matchings of graphs, and calculated the values of entropies of the complete graphs, star graphs, and complete bipartite graphs. Later Wan et al. [17] established some upper and lower bounds for these information-theoretic quantities. Distance is one of the most noticeable graph invariants. In [4], the number of vertices with a fixed distance to a given vertex served as an information functional, and in [10] eccentricity of a vertex, the maximum distance between the vertex and other vertices of a graph, is used for developing a new version of graph entropy. Ilić and Dehmer proved a sharp upper bound [13] for the distance-based graph entropies introduced in [4]. Similar information functionals also include the number of spanning forests of c-cyclic graphs [18], in which the authors proved formulas of the entropy for certain graph families and showed the maximal value of the entropy for some unicyclic graphs. The graph entropy we are focusing on in this paper is based on degree powers, which is due to Dehmer [6]. It is directly connected to the sum of powers of vertex degree known as the oldest degree-based structure descriptor in mathematical chemistry [9, 11]. Cao, Dehmer, and Shi proved some extremal values for trees, unicyclic graphs, bicyclic graphs, and other special graphs for the first degree-based graph entropy and further proposed conjectures for higher orders to determine extremal values in trees. In [12], Ilić proved one part of the conjecture about upper and lower bounds of the graph entropy. Das and Shi gave some extremal properties for trees [5]. These papers imply that it is intricate to determine minimal values of graph entropies.

Inspired by the work of [3], we investigate the extremal properties of the first degree-based graph entropy for general graphs and characterize the graphs attaining the extrema in Section 3. Based on the results, a method for seeking the extrema for any given (n, m) graph is proposed. As an application of the method, the minimum value of the entropy for graphs with fewer cycles and the maximum value for general graphs are given in Section 4. In the next section, we will start by introducing the basic concepts and notions concerning graph theory that will be used in the paper. And then definitions of degree-based graph entropies are reproduced.

2 Preliminaries

For all other terminologies, notations and concepts not mentioned here can be referred to [3] and any graph theory textbook.

2.1 Terminology and notation

A *graph* G is an ordered pair $(V(G), E(G))$ consisting of a non-empty set $V(G)$ of vertices and a set $E(G)$ of edges, where $E(G)$ is made up of unordered pair of vertices in G. If vertices u and v form an edge e, then we say u and v are *adjacent* and e can be denoted by uv; the two vertices are called the *ends* of the edge. For the sake of simplicity, we assume that $u \neq v$. Obviously, no two different edges have the same ends in a graph. $G - uv$ is the graph obtained from G by deleting the edge uv. Similarly, the graph obtained from G by adding uv is denoted by $G + uv$.

A graph H is a *subgraph* of G if $V(H) \subseteq V(G)$ and $E(H) \subseteq E(G)$. Let S be a non-empty set of $V(G)$. $G[S]$ denotes the subgraph of G whose vertex set is S and whose edge set is the set of those edges of G that have both ends in S, which is called the subgraph of G induced by S.

Two graphs G and H are said to be *isomorphic*, written as $G \cong H$, if there is a bijection $\phi : V(G) \to V(H)$ such that $uv \in E(G)$ if and only if $\phi(u)\phi(v) \in E(H)$ for all $u, v \in V(G)$.

We define a *neighbor* set $N(v)$ of a vertex v in G to be the set of all vertices adjacent to v. The *degree* $d_G(v)$ of v in G is just the size of $N(v)$. $\Delta(G)$ and $\delta(G)$ denote the maximum and minimum degrees of vertices in G, respectively. When vertices are distinguished by subscripts, to simplify the notation, we often write $d_1, d_2, ..., d_n$ instead of $d_G(v_1), d_G(v_2), ..., d_G(v_n)$ for $V(G) = \{v_1, v_2, ..., v_n\}$. $(d_{i_1}^{a_{i_1}}, d_{i_1}^{a_{i_2}}, \ldots, d_{i_k}^{a_{i_k}})$ is called a *degree sequence* of G, where a_{i_j} denotes the number of vertices with degree d_{i_j} for $j = 1, 2, \ldots, k$ and $\sum\limits_{j=1}^{k} a_{i_j} = n$.

A *path* P in G is a finite non-empty vertex sequence $v_1, v_2, ..., v_k$ such that v_i and v_{i+1} are adjacent, for $1 \leq i \leq k-1$, where v_i are all distinct. We say that P is a path from v_1 to v_k and the vertices v_1 and v_k are called the origin and terminus of P, respectively. A *cycle* is defined as a path with the same origin and terminus. Sometimes the term 'cycle' refers to a graph corresponding to a cycle. If there exists a path between any vertices in graph G, then G is *connected*. If G is a connected graph with n vertices and $n + c - 1$ edges, then G is called *c-cycle graph*. Particularly, when $c = 0$, G is called *acyclic*, also known as a *tree*; when $c = 1, 2$, G is called *unicyclic* and *bicyclic*, respectively. A *star* is a tree with $n - 1$ vertices of degree 1, denoted by S_n. A *complete graph* K_n is a graph with n vertices such that all the vertices are pairwise adjacent.

To simplify the notation, we often use an ordered pair of integers (n, m) to denote a graph with n vertices and m edges. All graphs considered in this paper are connected and finite.

2.2 Degree–based graph entropy

As we have seen, there are competing notions of graph entropy. In fact, graph entropy in one form or another plays an important role in a variety of problem areas. The vertex degree sequence is one of the most important graph invariants, which seems relatively easy to calculate in large-scale networks. It has also been proven useful in information theory, social networks, and mathematical chemistry [1, 3]. Cao, Dehmer, and Shi in [3] introduced the following degree-based graph entropy, in which degree powers are used as the information functionals:

$$\begin{aligned} I_k(G) &= \log\left(\sum_{i=1}^{n} {d_i}^k\right) - \sum_{i=1}^{n} \frac{{d_i}^k}{\sum_{j=1}^{n} {d_j}^k} \log {d_i}^k \\ &= \log\left(\sum_{i=1}^{n} {d_i}^k\right) - \frac{1}{\sum_{j=1}^{n} {d_j}^k} \sum_{i=1}^{n} {d_i}^k \log {d_i}^k \end{aligned}$$

where G is an (n, m) graph, $V(G) = \{v_1, v_2, ..., v_n\}$ and k is an arbitrary real number. Note that logarithms are always taken to the base two in this paper.

Since $\sum\limits_{i=1}^{n} d_i$ is the constant $2m$, $I_1(G) = \log(2m) - \frac{1}{2m}\sum\limits_{i=1}^{n} d_i \log d_i$, which is called the *first degree-based graph entropy.* Let $h(G) = \sum\limits_{i=1}^{n} d_i \log d_i$. The extremal values of $I_1(G)$ can be easily obtained from extremal values of $h(G)$.

3 Topological structures of extremal graphs for $h(G)$

We are first going to present the following important lemma, which is necessary to get the main results of this paper.

Lemma 1 *Let v_i, v_j, and v_k be three vertices in a graph G. Then $d_G(v_i) \geq d_G(v_j)$ if and only if $h(G + v_iv_k - v_kv_j) > h(G)$.*

Proof. Let $G' = G + v_iv_k - v_kv_j$. Clearly, $V(G) = V(G')$, $d_{G'}(v_i) = d_G(v_i) + 1$, $d_{G'}(v_j) = d_G(v_j) - 1$ and other vertices in G' have the same degrees as in G. Therefore, by $d_G(v_i) \geq$

$d_G(v_j)$, we have

$$
\begin{aligned}
h(G') - h(G) &= [(d_i+1)\log(d_i+1)) - d_i \log d_i] + [(d_j-1)\log(d_j-1) - d_j \log d_j] \\
&= \log\left(\frac{(d_i+1)^{(d_i+1)}}{d_i^{\,d_i}} \cdot \frac{(d_j-1)^{(d_j-1)}}{d_j^{\,d_j}}\right) \qquad (1) \\
&\geq \log \frac{(d_j+1)^{(d_j+1)}(d_j-1)^{(d_j-1)}}{d_j^{\,2d_j}}
\end{aligned}
$$

We have to show that the function $f(x) = \frac{(x+1)^{(x+1)}(x-1)^{(x-1)}}{x^{2x}} > 1$ over the set of natural numbers. By a simple calculation, we have

$$\frac{df}{dx} = (x-1)^{(x-1)}x^{(-2x)}(x+1)^{(x+1)}(\log(x-1) - 2\log x + \log(x+1)).$$

It is easy to see that $\frac{df}{dx} < 0$, so $f(x)$ is a monotonously decreasing function. Since $\lim\limits_{x\to\infty} f(x) = 1$, $f(x) > 1$. Thus $h(G') > h(G)$.

Conversely, suppose $h(G') > h(G)$. Then, by equality (1), we have

$$\frac{(d_i+1)^{(d_i+1)}}{d_i^{\,d_i}} > \frac{d_j^{\,d_j}}{(d_j-1)^{(d_j-1)}}$$

It follows that $d_G(v_i) \geq d_G(v_j)$. ■

For the sake of brevity, let G^* and $\overline{G}$ always denote graphs attaining the maximum and minimum values of h in (n,m) graphs, respectively. Next, the properties of G^* and $\overline{G}$ will be discussed to characterize extremal graphs.

Property 2 *Let v_i and v_j be two vertices in G^*. Then $d_i \geq d_j$ if and only if $N(v_j) \subseteq N(v_i)$.*

Proof. The sufficiency is obvious. We shall prove the necessity by contradiction. Suppose that v_i and v_j satisfy the hypotheses, but $N(v_j) \nsubseteq N(v_i)$. Let $v_k \in N(v_j)$ and $v_k \notin N(v_i)$. Then by Lemma 1, $h(G^* + v_iv_k - v_kv_j) > h(G^*)$, contradicting the definition of G^*. ■

From Property 2 we can see that two vertices with the same degree have the same neighbor set in G^*, and $\Delta(G^*) = n-1$ as well.

Property 3 *Let v_i and v_j be adjacent vertices in G^* with $d_i \geq d_j$, and S the set of all vertices of degree not less than d_i. Then $G^*[S]$ is a complete graph.*

Proof. Suppose, to the contrary, that there are two disadjacent vertices v_k and v_l in $G^*[S]$. Clearly $d_l \geq d_i$, and therefore $N(v_i) \subseteq N(v_l)$ by Property 2. Because v_i and v_j are adjacent, $v_j \in N(v_i)$. Thus, v_j and v_l are adjacent. Since $d_k \geq d_j$, it follows from Lemma 1 that $h(G^* + v_kv_l - v_lv_j) > h(G^*)$, a contradiction. ■

Before presenting the structure of G^*, we shall find it convenient to adopt the following notion and notation in our description.

Definition 1 *A K_aT graph is one whose vertex set can be partitioned into two disjoint sets S and T, where $|S| = a$, so that each pair of distinct vertices in S is joint by an edge, no edge has both ends in T and if $d(u) \geq d(v)$ for $u, v \in T$, then $N(v) \subseteq N(u)$.*

By the definition of K_aT graph, it is easy to see that S induces the complete graph K_a.

Theorem 4 *There must exist a K_aT graph such that $G^* \cong K_aT$.*

Proof. Let $u \in V(G^*)$ with $d(u) = \min\limits_{e \in E(G^*)} \{\max\limits_{e=uv}\{d(u), d(v)\}\}$, and let S be the set of all vertices of degree not less than $d(u)$ and $T = V - S$. Then, by Property 3, S induces the complete graph K_a, where a is the size of S. We say that there is no edge between any two vertices in T. Otherwise, it contradicts the choice of u. Furthermore, by Property 2, $N(v) \subseteq N(u)$ for any vertices u, v with $d(u) \geq d(v)$ in T. We conclude that $G^* \cong K_aT$. ■

The next result suggests the structure of graphs attaining the minimum value of h.

Theorem 5 $\Delta(\overline{G}) = \delta(\overline{G})$ *or* $\Delta(\overline{G}) = \delta(\overline{G}) + 1$.

Proof. By contradiction. Suppose $\Delta(\overline{G}) \geq \delta(\overline{G}) + 2$. Let v_i and v_j be two vertices in $\overline{G}$ with $d_i = \delta(\overline{G}), d_j = \Delta(\overline{G})$. By the hypotheses, $d_i < d_j$. Clearly, there exists another vertex v_k such that $v_kv_j \in E(\overline{G})$ but $v_iv_k \notin E(\overline{G})$. Then, by Lemma 1, $h(\overline{G} + v_iv_k - v_kv_j) < h(\overline{G})$. This contradiction establishes the theorem. ■

4 Extrema of the first degree–based entropy

By Theorem 4, we know that G^* must be isomorphic to some K_aT graph. As an application for the result, we shall discuss the maximum value of $h(G)$ for a c-cycle graph G with n vertices.

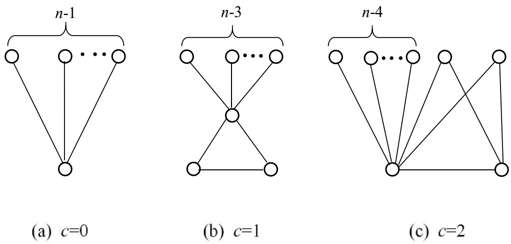


Figure 1. Graphs maximizing $h(G)$ of $c-$cycle graphs G with n vertices

Case: $c = 0$. Since there is no cycle in K_aT, $a = 1$ and $d(v) = 1$ for all $v \in T$. Clearly, $K_aT \cong S_n \cong G^*$, see Figure 1 (a).

Case: $c = 1$. A cycle in K_aT graphs can only lie in K_a or between K_a and T. So if the unique cycle lies in K_a, then $a = 3$ and $d(v) = 1$ for all $v \in T$; otherwise, $a = 2$ and each vertex in T has degree 1 except one of degree 2. Thus, there is a unique K_aT graph with degree sequence: $(n-1, 2^2, 1^{n-3})$. Figure 1 (b) shows the graph isomorphic to G^*.

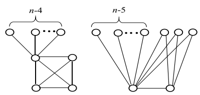
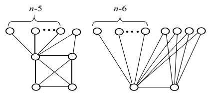
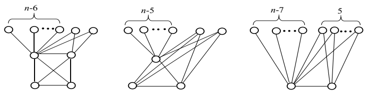
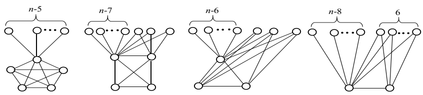
Case: $c = 2$. When the same ideas are applied here, we get a unique degree sequence of K_aT: $(n-1, 3, 2^2, 1^{n-4})$. Then G^* is isomorphic to Figure 1 (c).

As c gets larger, the number of non-isomorphic K_aT graphs is increased. Therefore, we have to find all possible K_aT graphs and compare the values of h among these K_aT graphs to obtain the largest one. Table 1 lists the results for each c from 3 to 6.

In this way, we can get extremal graphs for any given (n, m) graph. However, it is impractical because the enumeration of all possible sequences will get more difficult with the increase of m. So we hope to find a further connection between the maximum value and the structure of K_aT graphs. Lots of examples show that the maximum graphs are isomorphic to the K_aT graphs such that $a = 2$ and $d(v) \leq 2$ for each $v \in T$ as $m \geq n+9$, but we still lack the appropriate methods to prove it, see the following conjecture.

Conjecture 6 *Let G be an (m, n) graph, where $m \geq n+9$. Then $h(G) \leq h(G^*)$ if and only if and the degree sequence of G^* is $(n-1, m-n+2, 2^{m-n+1}, 1^{2n-m-3})$.*

Table 1. The maximum graphs of h values for c-cycle graphs.

c	K_aT graph	degree sequence	maximum
3		$(n-1,3^3,1^{n-4})$ $(n-1,4,2^3,1^{n-5})$	$(n-1,3^3,1^{n-4})$
4		$(n-1,4,3^2,2,1^{n-5})$ $(n-1,5,2^4,1^{n-6})$	$(n-1,5,2^4,1^{n-6})$
5		$(n-1,5,3^2,2^2,1^{n-6})$ $(n-1,4^2,3^2,1^{n-5})$ $(n-1,6,2^5,1^{n-7})$	$(n-1,4^2,3^2,1^{n-5})$ $(n-1,6,2^5,1^{n-7})$
6		$(n-1,4^4,1^{n-5})$ $(n-1,6,3^2,2^3,1^{n-7})$ $(n-1,5,4,3^2,2,1^{n-6})$ $(n-1,7,2^6,1^{n-8})$	$(n-1,4^4,1^{n-5})$

Finally, we close this section by giving graphs that minimize $h(G)$ for (n,m) graph G.

Theorem 7 *Let G be an (n,m) graph. Then $h(G) \geq h(\overline{G})$ if and only if the degree sequence of $\overline{G}$ is $(2^{n-2},1^2)$ for $m=n-1$ and $((a+3)^{2m-(a+2)n},(a+2)^{(a+3)n-2m})$ for $\frac{(a+2)n}{2} < m \leq \frac{(a+3)n}{2}$, where a is a natural number from 0 to $n-4$.*

Proof. It follows directly from Theorem 5. ■

5 Conclusion

In this paper, we have explored the extremal properties of the first degree-based graph entropy and further characterized the structure of extremal graphs. As a direct result of the structural characteristic, a method of seeking extremal results for any (n,m) graph was given. In fact, this is only feasible in theory. With the increase of the graph size, this enumeration for all possible K_aT graphs isomorphic to the maximum graph seems impractical. Obviously, we still need to have a deeper insight into the extremal properties of the graph entropy. Also, similar structural features should be investigated for the higher-order graph entropies.

References

[1] B. Bollobás, V. Nikiforov, Degree powers in graphs: the Erdős–Stone theorem, *Comb. Probab. Comput.* **21** (2012) 89–105.

[2] S. Cao, M. Dehmer, Z. Kang, Network entropies based on independent sets and matchings, *Appl. Math. Comput.* **307** (2017) 265–270.

[3] S. Cao, M. Dehmer, Y. Shi, Extremality of degree–based graph entropies, *Inf. Sci.* **278** (2014) 22–33.

[4] Z. Chen, M. Dehmer, Y. Shi, A note on distance–based graph entropies, *Entropy* **16** (2014) 5416–5427.

[5] K. C. Das, Y. Shi, Some properties on entropies of graphs, *MATCH Commun. Math. Comput. Chem.* **78** (2017) 259–272.

[6] M. Dehmer, Information processing in complex networks: graph entropy and information functionals, *Appl. Math. Comput.* **201** (2008) 82–94.

[7] M. Dehmer, F. Emmert-Streib, Structural information content of networks: graph entropy based on local vertex functionals, *Comput. Biol. Chem.* **32** (2008) 131–138.

[8] M. Dehmer, V. Kraus, On extremal properties of graph entropies, *MATCH Commun. Math. Comput. Chem.* **63** (2012) 889–912.

[9] T. Došlić, B. Furtula, A. Graovac, I. Gutman, S. Moradi, Z. Yarahmadi, On vertex degree based molecular structure descriptors, *MATCH Commun. Math. Comput. Chem.* **66** (2011) 613–626.

[10] M. Ghorbani, M. Dehmer, S. Zangi, Graph operations based on using distance–based graph entropies, *Appl. Math. Comput.* **333** (2018) 547–555.

[11] I. Gutman, K. Das, The first zagreb index 30 years after, *MATCH Commun. Math. Comput. Chem.* **50** (2004) 83–92.

[12] A. Ilić, On the extremal values of general degree–based graph entropies, *Inf. Sci.* **370–371** (2016) 424–427.

[13] A. Ilić, M. Dehmer, On the distance based graph entropies, *Appl. Math. Comput.* **269** (2015) 647–650.

[14] A. Mowshowitz, Entropy and the complexity of graphs I: an index of the relative complexity of a graph, *Bull. Math. Biophys.* **30** (1968) 175–204.

[15] N. Rashevsky, Life, information theory, and topology, *Bull. Math. Biophys.* **17** (1955) 229–235.

[16] E. Trucco, A note on the information content of graphs, *Bull. Math. Biol.* **18** (1965) 129–135.

[17] P. Wan, X. Chen, J. Tu, M. Dehmer, S. Zhang, F. Emmert-Streib, On graph entropy measures based on the number of independent sets and matchings, *Inf. Sci.* **516** (2020) 491–504.

[18] P. Wan, J. Tu, M. Dehmer, S. Zhang, F. Emmert-Streib, Graph entropy based on the number of spanning forests of c-cyclic graphs, *Appl. Math. Comput.* **363** (2019) #124616.