entropy

Article

A Conjecture Regarding the Extremal Values of Graph Entropy Based on Degree Powers

Kinkar Chandra Das [1,*] and Matthias Dehmer [2]

1 Department of Mathematics, Sungkyunkwan University, Suwon 440-746, Korea

2 Department of Biomedical Computer Science and Mechatronics, University for Health Sciences, Medical Informatics and Technology (UMIT), Eduard Wallnoefer Zentrum, A-6060 Hall in Tyrol, Austria; matthias.dehmer@unibw.de

* Correspondence: kinkardas2003@gmail.com; Tel.: +82-31-299-4528

Academic Editor: Andreas Holzinger

Received: 14 March 2016; Accepted: 10 May 2016; Published: 13 May 2016

Abstract: Many graph invariants have been used for the construction of entropy-based measures to characterize the structure of complex networks. The starting point has been always based on assigning a probability distribution to a network when using Shannon's entropy. In particular, Cao *et al.* (2014 and 2015) defined special graph entropy measures which are based on degrees powers. In this paper, we obtain some lower and upper bounds for these measures and characterize extremal graphs. Moreover we resolve one part of a conjecture stated by Cao *et al.*

Keywords: graphs; information theory; entropy; graph entropy; degree sequences; degree powers

MSC: 05C07

1. Introduction

Graph entropy measures have played an important role in a variety of fields, including information theory, biology, chemistry, and sociology. The entropy of a graph was first introduced by Mowshowitz [1] and Trucco [2]. Afterwards, Dehmer and Mowshowitz [3] interpreted the entropy of a graph based on vertex orbits as its structural information content. Indeed, this measure has been used as a graph complexity measure and is a measure for symmetry. Note that several graph entropies have been used extensively to characterize the topology of networks [3].

Dehmer [4] presents some novel information functionals that capture, in some sense, the structural information of the underlying graph G. Several graph invariants, such as the number of vertices, edges, distances, the vertex degree sequences, extended degree sequences (*i.e.*, the second neighbor, third neighbor, *etc.*), degree powers and connections, have been used for developing entropy-based measures [3–6]. In fact, the degree power is one of the most important graph invariants in graph theory. In [5,7], Cao *et al.* studied properties of graph entropies which are based on an information functional by using degree powers of graphs. To study results dealing with investigating properties of degree powers, we refer to [8–12]. In view of the vast of amount of existing graph entropy measures [4,13], there has been very little work to find their extremal values [14]. A reason for this might be the fact that Shannon's entropy represents a multivariate function and all probability values are not equal to zero when considering graph entropies. Inspired by Dehmer and Kraus [14], it turned out that determining minimal values of graph entropies is intricate because there is a lack of analytical methods to tackle this particular problem. In this paper we study novel properties of graph entropies which are based on an information functional by using degree powers of graphs. In particular, we proved that the path P_n gives the maximal graph entropy for any tree T (one part of the conjecture given in [5]). Moreover, we obtain some bounds on graph entropy in terms of the maximum degree and minimum degree of graphs.

2. Preliminaries

Let $G = (V, E)$ be a graph with n vertices. For $v_i \in V(G)$, d_i is the degree of the vertex v_i in G. The maximum vertex degree is denoted by Δ and the minimum vertex degree δ.

The vertex degree is an important graph invariant, which is related to many properties of graphs. Let G be a graph of order n with degree sequence $d_1, d_2, \ldots, d_n$. The sum of degree powers of a graph G is defined by $\sum_{i=1}^{n} d_i^k$, where k is an arbitrary real number. Sharp bounds for the sum of the k-th powers of the degrees of the vertices of graph G were obtained by Cioabă in [15].

The definition of Shannon's entropy [16]: Let $\mathbf{p} = (p_1, p_2, \ldots, p_n)$ be a probability vector, namely, $0 \le p_i \le 1$ and $\sum_{i=1}^{n} p_i = 1$. The Shannon's entropy of p is defined as

$$I(p) = -\sum_{i=1}^{n} p_i \, log \, p_i.$$

Recently, Cao *et al.* [5] introduced the following special graph entropy:

$$I_f(G) = -\sum_{i=1}^{n} \frac{d_i^k}{\sum_{j=1}^{n} d_j^k} \log \left(\frac{d_i^k}{\sum_{j=1}^{n} d_j^k} \right), \tag{1}$$

where d_i is the degree of the vertex v_i in G. According to [3], we see that $f(v_i) := d^{k_i}$. Throughout the paper all logarithms have base 2.

A graph G is said to be r-regular graph if all of its vertices have same degree r. For r-regular graph,

$$I_f(G) = \log n.$$

Throughout this paper we use P_n and S_n to denote the path graph and the cycle graph on n vertices, respectively. We obtain

$$I_f(P_n) = \frac{(n-2)\,2^k}{(n-2)2^k + 2} \log \left(\frac{2^k}{(n-2)\,2^k + 2} \right) + \frac{2}{(n-2)\,2^k + 2} \log \left((n-2)\,2^k + 2 \right).$$

The following conjecture has been published in [5]:

Conjecture 1. Let T be a tree with n vertices and $k > 0$. Then $I_f(T) \le I_f(P_n)$ with equality holding if and only if $T \cong P_n$; $I_f(T) \ge I_f(S_n)$ with equality holding if and only if $T \cong S_n$.

3. Results and Discussion

In this section we prove one part of the Conjecture 1. Moreover, we give some lower and upper bounds on $I_f(G)$ in terms of n, Δ and δ.

3.1. Proof of Conjecture on Entropy

Method of Lagrange Multipliers

To find the maximum and minimum values of $f(x, y, z)$ subject to the constraints $g(x, y, z) = k$ (assuming that these extreme values exist and $\nabla g \ne 0$ on the surface $g(x, y, z) = k$):

(a) Find all values of x, y, z, and λ such that

$$\begin{aligned} &\nabla f(x, y, z) = \lambda \nabla g(x, y, z) \\ \text{and} \qquad &g(x, y, z) = k. \end{aligned}$$

(b) Evaluate f at all the points (x, y, z) that result from step (a). The largest of these values is the maximum value of f; the smallest is the minimum value of f.

Lemma 1. *For $k \geq 2$,*

$$\frac{2^{k+1} - k}{2^{k+1} + 2} > \frac{k+3}{3^k}.$$

Proof. Let us consider a function

$$h(x) = \frac{x+2}{2^{x+1}+2} + \frac{1}{3^{x-1}} + \frac{x}{3^x}, \quad x \geq 2.$$

Then we have $h(x)$ is a strictly decreasing function on $x \geq 2$ and hence

$$h(x) < h(2) = 0.4 + \frac{1}{3} + \frac{2}{9} < 1.$$

This gives the required result. □

Lemma 2. *For $n \geq 4$, $k \geq 2$, we have*

$$\frac{(n-1)^k}{(n-1)^k + n - 1} \log \left(\frac{(n-1)^k}{(n-1)^k + n - 1} \right) - \frac{n-3}{(n-1)^k + n - 1} \log \left((n-1)^k + n - 1 \right)$$
$$> \frac{(n-2)\, 2^k}{(n-2)2^k + 2} \log \left(\frac{2^k}{(n-2)\, 2^k + 2} \right). \tag{2}$$

Proof. Since

$$\log \left(n - 2 + \frac{1}{2^{k-1}} \right) > \log (n-2),$$

we have to prove that

$$\frac{2^k (n-2)}{(n-2)\, 2^k + 2} \log (n-2) > \frac{(n-1)^k}{(n-1)^k + n - 1} \log \left(1 + \frac{1}{(n-1)^{k-1}} \right) + \frac{n-3}{(n-1)^k + n - 1} \times \log \left((n-1)^k + n - 1 \right),$$

that is,

$$\frac{2^k (n-2)}{(n-2)\, 2^k + 2} \log (n-2) > \frac{(n-1)^k + n - 3}{(n-1)^k + n - 1} \log \left(1 + \frac{1}{(n-1)^{k-1}} \right) + \frac{(n-3)\, k}{(n-1)^k + n - 1} \times \log (n-1),$$

that is,

$$\frac{2^k (n-2)}{(n-2)\, 2^k + 2} \log (n-2) > \frac{1}{(n-1)^{k-1}} + \frac{(n-3)\, k}{(n-1)^k + n - 1} \log (n-1) \quad \text{as} \quad \log (1+x) < x,$$

that is,

$$\frac{2^k\,(n-2)-(n-3)\,k}{(n-2)\,2^k+2}\,\log\,(n-2) > \frac{1}{(n-1)^{k-1}} + \frac{(n-3)\,k}{(n-1)^k+n-1}\,\log\left(1+\frac{1}{n-2}\right),$$

that is,

$$\frac{2^k\,(n-2)-(n-3)\,k}{(n-2)\,2^k+2}\,\log\,(n-2) > \frac{1}{(n-1)^{k-1}} + \frac{k}{(n-1)^k+n-1},$$

that is,

$$\frac{2^k\,(n-2)-(n-3)\,k}{(n-2)\,2^k+2}\,\log\,(n-2) > \frac{n+k-1}{(n-1)^k}\,. \tag{3}$$

For $n=4$, by Lemma 1, the Inequality (3) is satisfied. Otherwise, $n \geq 5$. Since $k \geq 2$, one can easily check that

$$\frac{2^k\,(n-2)-(n-3)\,k}{(n-2)\,2^k+2} > \frac{1}{2} \quad \text{and} \quad (n-1)^k > 2\,(n+k-1).$$

The Inequality (3) is satisfied. This completes the proof of the Lemma. □

We can assume that $d_1 \geq d_2 \geq \cdots \geq d_n$.

Theorem 1. *Let T be a tree of order n (>1) and $k \geq 2$. Then $I_f(T) \leq I_f(P_n)$.*

Proof. For $n = 2, 3$, we have $T \cong P_n$ and hence the equality holds. Otherwise, $n \geq 4$. If $T \cong P_n$, then the equality holds. Otherwise, $\Delta \geq 3$. It is well-known: for any tree T, $d_{n-1} = d_n = 1$. From Equation (1), we have

$$I_f(T) = -\sum_{i=1}^{n-2} \frac{d_i^k}{\sum\limits_{j=1}^{n-2} d_j^k + 2}\,\log\left(\frac{d_i^k}{\sum\limits_{j=1}^{n-2} d_j^k + 2}\right) + \frac{2}{\sum\limits_{j=1}^{n-2} d_j^k + 2}\,\log\left(\sum_{j=1}^{n-2} d_j^k + 2\right). \tag{4}$$

Claim 1.

$$\frac{1}{\sum\limits_{j=1}^{n-2} d_j^k + 2}\,\log\left(\sum_{j=1}^{n-2} d_j^k + 2\right) \leq \frac{1}{(n-2)\,2^k+2}\,\log\left((n-2)\,2^k+2\right). \tag{5}$$

Proof of Claim 1. For tree T, we have

$$(n-2)\,2^k \leq \sum_{i=1}^{n-2} d_i^k \leq (n-1)^k + n - 3.$$

Let us consider a function

$$f(x) = \frac{1}{x}\,\log\,x \quad \text{for} \quad (n-2)\,2^k+2 \leq x \leq (n-1)^k+n-1.$$

Then we have

$$f'(x) = \frac{1}{x^2}\,(\log\,e - \log\,x) < 0.$$

Therefore $f(x)$ is a decreasing function on $(n-2)\,2^k+2 \leq x \leq (n-1)^k+n-1$ and hence we get the required result in Inequality (5). □

Claim 2.

$$\sum_{i=1}^{n-2} \frac{d_i^k}{\sum\limits_{j=1}^{n-2} d_j^k + 2} \log \left(\frac{d_i^k}{\sum\limits_{j=1}^{n-2} d_j^k + 2} \right) > \frac{(n-2)\,2^k}{(n-2)2^k + 2} \log \left(\frac{2^k}{(n-2)\,2^k + 2} \right). \tag{6}$$

Proof of Claim 2. Let us consider a function

$$g(y_1, y_2, \ldots, y_{n-2}) = \sum_{i=1}^{n-2} \frac{y_i^k}{\sum\limits_{j=1}^{n-2} y_j^k + 2} \log \left(\frac{y_i^k}{\sum\limits_{j=1}^{n-2} y_j^k + 2} \right) \quad \text{for} \quad 1 \le y_i \le n-1,\, i = 1, 2, \ldots, n-2$$

with integer y_i such that

$$h(y_1, y_2, \ldots, y_{n-2}) = \sum_{i=1}^{n-2} y_i = 2\,(n-2).$$

Now,

$$\begin{aligned} \frac{\partial g}{\partial y_i} &= \frac{k\, y_i^{k-1}}{\sum\limits_{j=1}^{n-2} y_j^k + 2} \log \left(\frac{y_i^k}{\sum\limits_{j=1}^{n-2} y_j^k + 2} \right) - \frac{k\, y_i^{2\,k-1}}{\left(\sum\limits_{j=1}^{n-2} y_j^k + 2 \right)^2} \log \left(\frac{y_i^k}{\sum\limits_{j=1}^{n-2} y_j^k + 2} \right) \\ &\quad + \left[\frac{k\, y_i^{k-1}}{\sum\limits_{j=1}^{n-2} y_j^k + 2} - \frac{k\, y_i^{2\,k-1}}{\left(\sum\limits_{j=1}^{n-2} y_j^k + 2 \right)^2} \right] \log e \\ &= \left(1 - \frac{y_i^k}{\sum\limits_{j=1}^{n-2} y_j^k + 2} \right) \frac{k\, y_i^{k-1}}{\sum\limits_{j=1}^{n-2} y_j^k + 2} \left[\log \left(\frac{y_i^k}{\sum\limits_{j=1}^{n-2} y_j^k + 2} \right) + \log e \right], i = 1, 2, \ldots, n-2. \end{aligned}$$

By using the method of Lagrange multiplier, we have

$$\frac{\partial g}{\partial y_1} = \frac{\partial g}{\partial y_2} = \cdots = \frac{\partial g}{\partial y_{n-2}} = \lambda.$$

Therefore we yield $y_1 = y_2 = \cdots = y_{n-2} = 2$. Again by using the method of Lagrange multiplier, we conclude that $g(2, 2, \ldots, 2)$ gives either minimum or maximum value. By Lemma 2, we have $g(n-1, 1, \ldots, 1) > g(2, 2, \ldots, 2)$. Therefore we obtain the required result in Inequality (6). □

Using Inequalities (5) and (6) in Equation (4), we get $I_f(T) < I_f(P_n)$. This completes the proof. □

3.2. Bounds on I_f

In this subsection we obtain lower and upper bounds for I_f in terms of n, Δ and δ.

Theorem 2. *Let G be a graph of order n with maximum degree Δ and minimum degree δ. Then*

$$\log\left(1+(n-1)\left(\frac{\delta}{\Delta}\right)^k\right) \le I_f(G) \le \log\left(1+(n-1)\left(\frac{\Delta}{\delta}\right)^k\right). \tag{7}$$

Both inequalities hold if and only if G is a regular graph.

Proof. First part: Let

$$A = \max_{1\le i\le n}\left\{\frac{d_i^k}{\sum\limits_{j=1}^{n} d_j^k}\right\}.$$

Then we obtain

$$\log\left(\frac{d_i^k}{\sum\limits_{j=1}^{n} d_j^k}\right) \le \log(A), \quad \text{for any } i,\ 1\le i\le n.$$

Therefore

$$\begin{aligned}\sum_{i=1}^{n}\frac{d_i^k}{\sum\limits_{j=1}^{n} d_j^k}\log\left(\frac{d_i^k}{\sum\limits_{j=1}^{n} d_j^k}\right) &\le \sum_{i=1}^{n}\frac{d_i^k}{\sum\limits_{j=1}^{n} d_j^k}\log(A) \\ &= \log(A).\end{aligned} \tag{8}$$

Since

$$A = \max_{1\le i\le n}\left\{\frac{d_i^k}{\sum\limits_{j=1}^{n} d_j^k}\right\} \le \frac{\Delta^k}{\Delta^k+(n-1)\,\delta^k},$$

from the above, we get

$$\begin{aligned} I_f(G) &= -\sum_{i=1}^{n}\frac{d_i^k}{\sum\limits_{j=1}^{n} d_j^k}\log\left(\frac{d_i^k}{\sum\limits_{j=1}^{n} d_j^k}\right) \\ &\ge \log\left(\frac{1}{A}\right) \\ &\ge \log\left(1+(n-1)\left(\frac{\delta}{\Delta}\right)^k\right). \end{aligned} \tag{9}$$

Thus we get the left of the Inequality (7).

Suppose that the left inequality holds in (7). Then all the inequalities must be equalities. From the equality in (8), we have $d_1 = d_2 = \cdots = d_n$. From the equality in (9), we have $d_2 = d_3 = \cdots = d_n$. Hence G is a regular graph.

Conversely, one can see that the left equality holds in (7) for regular graph.

Second part: for the right inequality in (7), we assume that

$$B = \min_{1 \le i \le n} \left\{ \frac{d_i^k}{\sum\limits_{j=1}^{n} d_j^k} \right\} \ge \frac{\delta^k}{\delta^k + (n-1)\,\Delta^k}\,.$$

Using the above result as before, we get the right inequality in (7). Moreover, the right equality holds in (7) if and only if G is a regular graph. $\square$

4. Conclusions

In this paper, we studied a special graph entropy measure which is based on vertex degrees. We proved one part of the Conjecture 1 that the path P_n gives the maximal graph entropy for any tree T. Moreover, we give lower and upper bounds for this measure $I_f(G)$ in terms of n, Δ and δ. The characterization of minimal entropy remains an open problem and constitutes future work. We see that characterizing extremal graphs when using graph entropies is intricate because the problem depends on the underlying entropy measure and graph invariant. In this case, finding the minimal entropy is quite challenging as $I_f(G)$ can be interpreted as a multivariate function in terms of the $p(v_i)$. Studying these problems for special and rather simple graph classes gives us an idea about the complexity of the problem when considering general graphs.

In this paper, we tackled a theoretical problem when dealing with graph entropy. However, as already demonstrated, graph entropies have been applied for solving problems in machine learning and knowledge discovery in several disciplines. Interesting application areas are health and bioinformatics, see [17–20]. So far, graph entropy measures and other classical information-theoretic measures have been also employed for solving special problems of machine learning such as parameter selection and explorative data analysis of publication data [20–22]. A next step could be to demonstrate the potential for graph entropies in nursing and health informatics more extensively and, hence, to add and demonstrate more conceptional rigor and interdisciplinarity when dealing with applied problems in the mentioned fields.

Acknowledgments: The authors are much grateful to three anonymous referees for their valuable comments on our paper, which have considerably improved the presentation of this paper.

Author Contributions: Kinkar Chandra Das and Matthias Dehmer wrote the paper. Both authors have read and approved the final manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Mowshowitz, A. Entropy and the complexity of the graphs: I. An index of the relative complexity of a graph. *Bull. Math. Biophys.* **1968**, *30*, 175–204.
2. Trucco, E. A note on the information content of graphs. *Bull. Math. Biophys.* **1956**, *18*, 129–135.
3. Dehmer, M.; Mowshowitz, A. A history of graph entropy measures. *Inf. Sci.* **2011**, *181*, 57–78.
4. Dehmer, M. Information processing in complex networks: Graph entropy and information functionals. *Appl. Math. Comput.* **2008**, *201*, 82–94.
5. Cao, S.; Dehmer, M.; Shi, Y. Extremality of degree-based graph entropies. *Inf. Sci.* **2014**, *278*, 22–33.
6. Konstantinova, E.V. On Some Applications of Information Indices in Chemical Graph Theory. In *General Theory of Information Transfer and Combinatorics;* Springer: Berlin/Heidelberg, Germany, 2006.
7. Cao, S.; Dehmer, M. Degree-based entropies of networks revisited. *Appl. Math. Comput.* **2015**, *261*, 141–147.
8. Das, K.C. Maximizing the sum of the squares of the degrees of a graph. *Discret. Math.* **2004**, *285*, 57–66.
9. Das, K.C.; Xu, K.; Nam, J. Zagreb indices of graphs. *Front. Math. China* **2015**, *10*, 567–582.
10. Das, K.C. On Comparing Zagreb Indices of Graphs. *MATCH Commun. Math. Comput. Chem.* **2010**, *63*, 433–440.

11. Das, K.C.; Gutman, I.; Zhou, B. New upper bounds on Zagreb indices. *J. Math. Chem.* **2009**, *46*, 514–521.
12. Xu, K.; Das, K.C.; Balachandran, S. Maximizing the Zagreb Indices of (n, m)-Graphs. *MATCH Commun. Math. Comput. Chem.* **2014**, *72*, 641–654.
13. Bonchev, D.G. *Information Theoretic Indices for Characterization of Chemical Structures*; Research Studies Press: Chichester, UK, 1983.
14. Dehmer, M.; Kraus, V. On extremal properties of graph entropies. *MATCH Commun. Math. Comput. Chem.* **2012**, *68*, 889–912.
15. Cioabă, S.M. Sums of powers of the degrees of a graph. *Discret. Math.* **2006**, *306*, 1959–1964.
16. Shannon, C.E.; Weaver, W. *The Mathematical Theory of Communication*; University of Illinois Press: Urbana, IL, USA, 1949.
17. Holzinger, A.; Dehmer, M.; Jurisica, I. Knowledge discovery and interactive data mining in Bioinformatics—State-of-the-Art, future challenges and research directions. *BMC Bioinform.* **2014**, *15* (Suppl. S6), I1, doi:10.1186/1471-2105-15-S6-I1.
18. Holzinger, A. Trends in interactive knowledge discovery for personalized medicine: Cognitive science meets machine learning. *IEEE Intell. Inform. Bull.* **2014**, *15*, 6–14.
19. Mayer, C.; Bachler, M.; Holzinger, A.; Stein, P.K.; Wassertheurer, S. The effect of threshold values and weighting factors on the association between entropy measures and mortality after myocardial infarction in the cardiac arrhythmia suppression trial (CAST). *Entropy* **2016**, *18*, 129, doi:10.3390/e18040129.
20. Mayer, C.; Bachler, M.; Hörtenhuber, M.; Stocker, C.; Holzinger, A.; Wassertheurer, S. Selection of entropy-measure parameters for knowledge discovery in heart rate variability data. *BMC Bioinform.* **2014**, *15* (Suppl. 6), S2, doi:10.1186/1471-2105-15-S6-S2.
21. Holzinger, A.; Hörtenhuber, M.; Mayer, C.; Bachler, M.; Wassertheurer, S.; Pinho, A.; Koslicki, D. On entropy-based data mining. In *Interactive Knowledge Discovery and Data Mining in Biomedical Informatics*; Holzinger, A., Jurisica, I., Eds.; Springer: Berlin/Heidelberg, Germany, 2014; pp. 209–226.
22. Holzinger, A.; Ofner, B.; Stocker, C.; Valdez, A.C.; Schaar, A.K.; Ziefle, M.; Dehmer, M. On graph entropy measures for knowledge discovery from publication network data. In *Availability, Reliability, and Security in Information Systems and HCI*; Cuzzocrea, A., Kittl, C., Simos, D.E., Weippl, E., Xu, L., Eds.; Springer: Berlin/Heidelberg, Germany, 2013; pp. 354–362.