

网络熵

新网络熵：图的支配熵

[1]Şahin Bünyamin. New network entropy: The domination entropy of graphs[J]. Information Processing Letters,2022,174:

1.起源

熵entropy：香农在1948年提出的概念。

图熵graph entropy：Rashevsky在1955年提出的概念。用于衡量顶点相对于顶点度的等效类的划分。

介绍一下用熵来衡量某些拓扑结构的优势

以前：化学、物理、生物分子拓扑结构，通过一个拓扑指数（topological indices）来展示不同分子的差异。以下是一些拓扑指数：

1. Wiener index：维纳指数等于图中每对顶点之间总距离的二分之一。
2. Hosoya index：Hosoya 指数等于匹配总数。
3. Merrifield-Simmons index：Merrifield-Simmons 指数等于独立集的总数。

存在的问题：它们通常与分子的相对分子特性相关，但相同的指标不能对不同的分子有很高的区分能力。

后来：运用信息论中的熵，提出了信息指数（information indices）。结果表明，信息指数比各自的拓扑指数对分子有更大的区分能力。

再后来：Dehmer 介绍了一些新的图顶点信息泛函（information functionals，更准确地说，信息泛函基于导出的概率分布量化图的结构信息）。

然后就出现了很多图熵度量（graph entropy measure），基于图的一些不变的属性，如顶点数、边数、顶点的度序列、顶点的度幂（degree powers），还有基于匹配和独立集的。

2.关于本文

支配（domination）也是图的一个重要的属性。

Haynes等人证明，图的这些不变的属性，在树中，对图的微小变化是十分敏感的。

基于以上种种原因，作者定义了一个基于支配集的图熵的度量。

要得到熵，作者使用了图的支配多项式（domination polynomials）。

除了一些众所周知的支配多项式之外，我们还定义了细分星图的支配多项式。此外，我们将支配熵与基于匹配、独立集、顶点度和给出的 21 个图的自同构组的图熵度量进行了比较。这 21 个五阶图根据其拓扑复杂度进行排序。

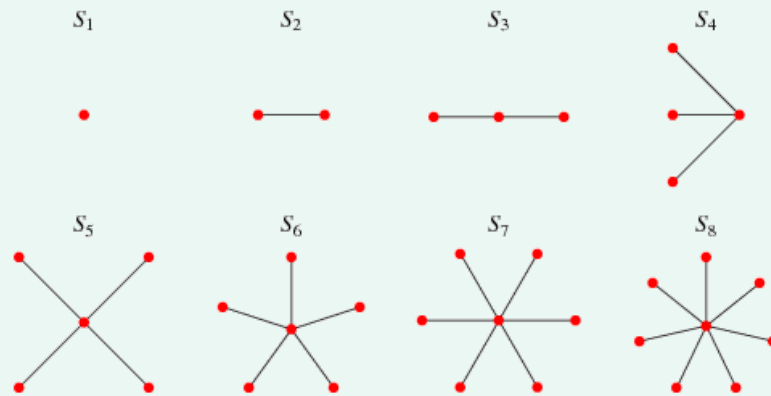
3.名词、概念

cardinality: [n] 基数; 集合的势

subdivide: [vi]细分; [vt]把...细分

star graph

n 阶星图 S_n , 或者称为“ n -star”, 是一个 n 个节点的树, 其中一个节点的度为 $n-1$, 剩下的 $n-1$ 个节点的度为1。因此星形图 S_n 与完全二部图 $K_{1,n-1}$ 同构。



double star graph

双星图, 是由两个星图联合以及连接他们中心的线组成的图形。

n 阶双星图 $S_{p,q}$, 由 $S_{1,p}$ and $S_{1,q}$ 组成, $n = p + q + 2$ 。

subdivided star graph

细分的星图 S_k^* , 由星图 S_{k+1} 得来, 通过在 k 个顶点的星图中度为1的顶点上再添加一个点, 因此 S_k^* 的阶数等于 $2k+1$ 。

friendship graph

友谊图 F_k , 由循环图 C_3 的 k 个副本组成, 因此它们都共享一个公共顶点。友谊图 F_k 的阶数等于 $2k+1$ 。

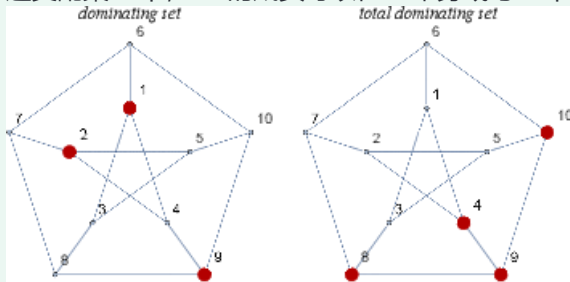
度序列degree sequence

假设 $V(G) = \{v_1, \dots, v_n\}$ 和 $\deg(v_k) = d_k, k = 1, \dots, n$ 。序列 $D(G) = (d_1, d_2, \dots, d_n)$ 称为 G 的度序列。

总支配集total dominating set

对于图 G 和顶点集 $V(G)$ 的子集 S^t ，用 $N_G^t[S^t]$ 表示 G 中与 S^t 中的顶点相邻的顶点集（ G 中的所有点都至少与 S^t 中的一个点相邻）。如果 $N_G^t[S^t] = V(G)$ ，则称 S^t 是一个总支配集（ G 中的顶点）。因为总支配集的成员必须与另一个顶点相邻，所以没有为具有孤立顶点的图定义总支配集。

总支配集与普通支配集的区别在于，在全支配集 S^t 中，要求 S^t 的成员自身与 S^t 中的一个顶点相邻，而在普通支配集 S 中， S 的成员可以在 S 本身或与 S 中的顶点相邻。



总支配数total domination number, $\gamma_t(G)$

用于表示图 G 的总支配数（the total domination number of a graph G ），这是最小总支配集的基数（which is the cardinality of a total dominating set with minimum order）。

例如，在上图所示的彼得森图中， $\gamma(P) = 3$ ，因为集合 $S = \{1, 2, 9\}$ 是最小支配集（左图），而 $\gamma_t(P) = 4$ 因为 $S^t = \{4, 8, 9, 10\}$ 是最小总支配集（右图）。

支配多项式domination polynomial, $D(G, x)$

$D(G, i)$ 表示基数为 i 的 G 的支配集族（也是一个集合）， d_i 表示 $D(G, i)$ 的基数，使得 $d_i(G) = |D(G, i)|$ 。

$$D(G, x) = \sum_{i=\gamma(G)}^{V(G)} d_i(G) x^i$$

例子：

考虑路径图 $P_4: v_1 v_2 v_3 v_4$ 。为了支配图 P_4 ，取能够支配其他两个顶点的两个顶点就足够了。因此 $\gamma(P_4) = 2$ 。 P_4 的总支配集是 $\{v_2, v_3\}$ 并且 $\gamma_t(P_4) = 2$ 。基数为 2 的 P_4 的支配集是 $D(G, 2) = \{\{v_1, v_3\}, \{v_1, v_4\}, \{v_2, v_3\}, \{v_2, v_4\}\}$ 且 $d_2(G) = 4$ 。此外，基数为 3 的 P_4 的支配集为 $D(G, 3) = \{\{v_1, v_2, v_3\}, \{v_1, v_2, v_4\}, \{v_1, v_3, v_4\}, \{v_2, v_3, v_4\}\}$ ， $d_3(G) = 4$ 。最后，基数为 4 的 P_4 的支配集是 $D(G, 4) = \{v_1, v_2, v_3, v_4\}$ 和 $d_4(G) = 1$ 。

因此， P_4 的支配多项式为 $D(G, P_4) = x^4 + 4x^3 + 4x^2$ 。

通过这个多项式，可得知 P_4 的总支配集个数是 9 ($4+4+1$)。

由支配多项式，又可推出一个公式：

$$\gamma_s(G) = \sum_{i=\gamma(G)}^{|V(G)|} d_i(G)$$

γ_s 来表示支配集的总数 (We use the notation γ_s to denote the total number of dominating sets.) 。**注意，不要和总支配数 γ_t 混淆。**很明显， γ_s 等于图 G 的支配多项式的**系数之和**。

图熵

设 G 是一个图， $f: S \rightarrow R_+$ 是定义在 $S = \{s_1, s_2, \dots, s_k\}$ 上的信息泛函，使得 S 是 G 的一组元素。定义如下(log以2为底)：

$$\begin{aligned} I_f(G) &= - \sum_{i=1}^k \frac{f(s_i)}{\sum_{j=1}^k f(s_j)} \log \left(\frac{f(s_i)}{\sum_{j=1}^k f(s_j)} \right) \\ &= \log \left(\sum_{i=1}^k f(s_i) \right) - \frac{\sum_{i=1}^k f(s_i) \log f(s_i)}{\sum_{j=1}^k f(s_j)} \end{aligned}$$

支配熵

对图 G ， $|V| = n$ 且没有孤立的顶点，我们引入信息泛函使得 $f := d_i(G)$ ，其中 $d_i(G) = |D(G, i)|$ 。

Definition 4. For a graph G with $|V| = n$ without an isolated vertex, we introduce the information functional such that

$$f := d_i(G)$$

where $d_i(G) = |D(G, i)|$. Thus, for each $i < \gamma(G)$ $d_i(G) = 0$, $d_{n-1}(G) = n$, and $d_n(G) = 1$. Then by using Definition 3, we obtain the domination entropy

$$\begin{aligned} I_{dom}(G) &= I_f(G) = - \sum_{i=1}^n \frac{d_i(G)}{\gamma_s(G)} \log \left(\frac{d_i(G)}{\gamma_s(G)} \right) \\ &= \log(\gamma_s(G)) - \frac{1}{\gamma_s(G)} \sum_{i=1}^n d_i(G) \log(d_i(G)) \\ &= \log(\gamma_s(G)) - \frac{1}{\gamma_s(G)} \sum_{i=1}^{n-2} d_i(G) \log(d_i(G)) - \frac{n \log n}{\gamma_s(G)}. \end{aligned}$$

基于匹配的熵度量(Hosoya index, or Z-index)

具有 k 条边的匹配数用 $z_k(G)$ 表示。假设空集是匹配的，并且 $z_0(G) = 1$ 。

图 G 的 Hosoya 指数 (或 Z 指数) 计算为 $Z(G) = \sum_{k=0}^m z_k(G)$ 。

Definition 5. For a graph G with $|E(G)| = m$, the entropy measure—which is based on the number of matchings—is denoted by I_{nm} and computed by [29]

$$I_{nm}(G) = - \sum_{k=0}^m \frac{z_k(G)}{Z(G)} \log \left(\frac{z_k(G)}{Z(G)} \right).$$

基于独立集的熵度量 (Merrifield-Simmons index , or σ -index)

$\sigma_k(G)$ 表示基数为 k 的独立集合的数量。空集可以认为是一个独立的集, $\sigma_0(G) = 1$ 。

图 G 的 Merrifield-Simmons 指数 (或 σ -index) 计算为 $\sigma(G) = \sum_{k=0}^n \sigma_k(G)$ 。

Definition 6. For a graph G of order n , the entropy measure – which is based on the number of independent sets – is denoted by I_{nis} and computed by [29]

$$I_{nis}(G) = - \sum_{k=0}^n \frac{\sigma_k(G)}{\sigma(G)} \log \left(\frac{\sigma_k(G)}{\sigma(G)} \right).$$

基于度幂(degree power)的熵度量

相关文章: [Extremality of degree-based graph entropies](#)

这个度幂degree power, 就是字面意思, 度的 n 次方。

$$I_d^k(G) = - \sum_{i=1}^n \frac{\deg(v_i)^k}{\sum_{j=1}^n \deg(v_j)^k} \log \left(\frac{\deg(v_i)^k}{\sum_{j=1}^n \deg(v_j)^k} \right).$$

当 $k=1$ 时, 可得到 first-degree entropy。

Definition 7. For a graph G of order n , the first-degree entropy is defined as

$$I_{fd}(G) = - \sum_{i=1}^n \frac{\deg(v_i)}{\sum_{j=1}^n \deg(v_j)} \log \left(\frac{\deg(v_i)}{\sum_{j=1}^n \deg(v_j)} \right).$$

拓扑信息内容 (topological information content)

Definition 8. Let G be a simple graph of order n and n_i ($1 \leq i \leq k$) be the cardinality of the i -th orbit of G . Then, the topological information content I_α is defined as follows:

$$I_\alpha(G) = - \sum_{i=1}^k \frac{n_i}{n} \log \left(\frac{n_i}{n} \right).$$

It is clear that $I_\alpha(G)$ has the maximum value for graphs, which have no symmetry [32]. This means that the automorphism group of G is trivial.

什么是 orbit of G ??

4.一些图的支配熵

假设 H_1 和 H_2 是两个连通图, $H = H_1 \cup H_2$ 分别是 H_1 和 H_2 的不相交并集。 我们得到:

$$I_{dom} = \log \left(\sum_{i=\gamma(H)}^n d_i(H) \right) - \frac{1}{\sum_{i=\gamma(H)}^n d_i(H)} \sum_{i=1}^n d_i(H) \log (d_i(H))$$

使得 $\gamma(H) = \gamma(H_1) + \gamma(H_2)$ 和 $d_i(H) = \sum_{j=\gamma(H_1)}^i d_j(H_1)d_{i-j}(H_2)$. 怎么得来的?

完全图 K_n , 可得:

$$I_{dom}(K_n) = \log(2^n - 1) - \frac{1}{2^n - 1} \sum_{i=1}^n C_n^i \log(C_n^i)$$

证明: 支配多项式 $D(G, x) = (1 + x)^n - 1$, 通过二项式展开定理, 可得 $\gamma_s(G) = 2^n - 1$, $d_i(G) = C_n^i, 1 \leq i \leq n$. (为什么要减1? 答: 因为二项式展开的第一项为 $C_n^0 1^n \times x^0 = 1$, 而 $d_0(G) = 0$, 因此这项对于该图来说并不存在, 所以要减1)

星图 S_n , 可得:

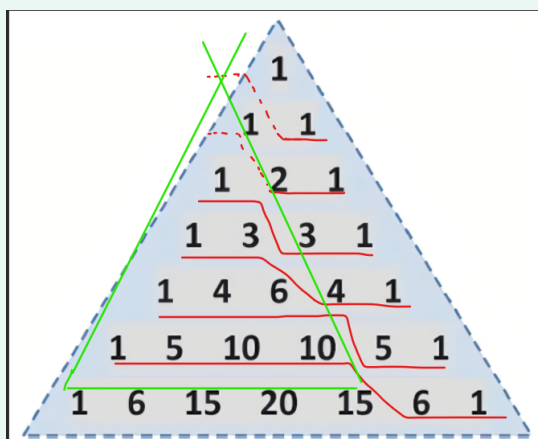
$$I_{dom}(S_n) = \log(2^{n-1} + 1) - \frac{1}{2^{n-1} + 1} \left(\sum_{i=1}^{n-3} C_{n-1}^i \log(C_{n-1}^i) \right) - \frac{n \log n}{2^{n-1} + 1}$$

推导:

先推导 S_n 的支配多项式(找规律):

1. $n=1: d_1(G) = 1$
2. $n=2: d_1(G) = 2, d_2(G) = 1$
3. $n=3: d_1(G) = 1, d_2(G) = 3, d_3(G) = 1$
4. $n=4: d_1(G) = 1, d_2(G) = 3, d_3(G) = 4, d_4(G) = 1$
5. $n=5: d_1(G) = 1, d_2(G) = 4, d_3(G) = 6, d_4(G) = 5, d_5(G) = 1$
6.

借助杨辉三角, 可得以下规律:



支配多项式的系数为 $n-1$ 行, 所以写成 $(1 + x)^{n-1}$;

但这样所有 x 的幂都少了1, 所以再乘一个 x , 得 $x(1 + x)^{n-1}$;

而第 $n-1$ 个的系数需要加1, 因此加一个 x^{n-1} , 得 $D(S_n, x) = x^{n-1} + x(1 + x)^{n-1}$.

那么也就易得 $\gamma_s(S_n) = \sum_{i=\gamma(S_n)=1}^n d_i(G) = 2^{n-1} + 1$.

带入之前的公式, 即可得出。

双星图 $S_{a,b}$, 可得:

$$I_{dom}(S_{a,b}) = \log(\gamma_s(S_{a,b})) - \frac{1}{\gamma_s(S_{a,b})} \left(\sum_{i=2}^{a+b-3} d_i(G) \log(d_i(G)) \right) \\ - \frac{(C_{a+b-2}^2 + a + b - 1) \log(C_{a+b-2}^2 + a + b - 1) + (a + b) \log(a + b)}{\gamma_s(S_{a,b})}$$

梳子图 E_k , 有 $n=2k$ 个顶点, 可得:

$$I_{dom}(E_k) = \log(3^k) - \frac{1}{3^k} \left(\sum_{i=k}^{2k} C_{i-k}^k 2^{2k-i} \log(C_{i-k}^k 2^{2k-i}) \right)$$

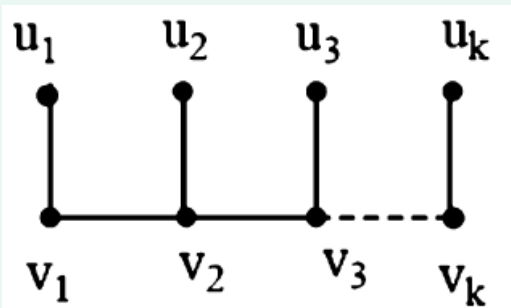


Fig. 1. Comb graph E_k .

友谊图 F_k , 有 $n=2k+1$ 个顶点, 可得:

$$I_{dom}(F_k) = \log(3^k + 2^{2k}) \\ - \frac{1}{3^k + 2^{2k}} \left(\sum_{i=0}^{2k} (C_{2k}^{i-1} + C_k^{i-k} 2^{2k-i}) \log(C_{2k}^{i-1} + C_k^{i-k} 2^{2k-i}) \right)$$

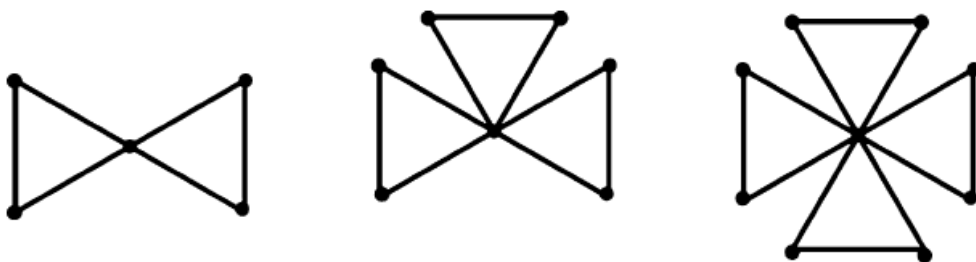


Fig. 2. Friendship graphs F_2 , F_3 and F_4 .

细分的星图 S_k^* ，有 $n=2k+1$ 个顶点，可得：

$$I_{dom}(S_k^*) = \log(2 \times 3^k - 1) - \frac{1}{2 \times 3^k - 1} ((2^k - 1) \log(2^k - 1) + \sum_{i=0}^k (C_k^i 2^{k-i} + C_k^{i+1} 2^{k-i-1}) \log(C_k^i 2^{k-i} + C_k^{i+1} 2^{k-i-1}))$$

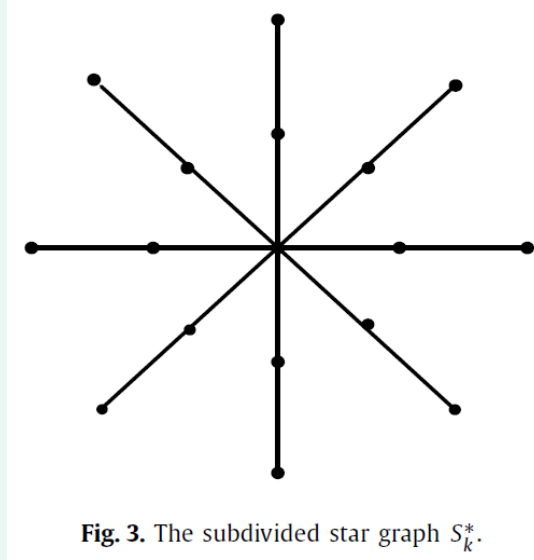


Fig. 3. The subdivided star graph S_k^* .

推导：该图的支配多项式是由作者推导出的，其他图的支配多项式别人已经推导出了

\therefore 要支配 S_k^* ，那么每个臂（arm）都至少要有有一个点在支配集中。

$\therefore \gamma(S_k^*) = k$ ，并有以下两种情形：

1. 在每个臂上，有1个基数为2的集合(取臂上的2个点)，2个基数为1的集合(臂上的2个点2选1)，中心点可选可不选，则可得到 $(x^2 + 2x)^k(x + 1) = x^k(x + 2)^k(x + 1)$ 。
2. 满足1的情况下，只有一种情况不是支配集：就是全部由度为1的点组成的集合。因此要把这种情况去除。因而可得 $x^k(x + 2)^k(x + 1) - x^k$ 。

$$\therefore D(S_k^*, x) = \sum_{i=k}^{n=2k+1} d_i(G) x^i = x^k(x + 2)^k(x + 1) - x^k。$$

5.支配熵与其他熵度量的比较

Table 1
Measures of I_α , I_{fd} , I_{nis} , I_{nm} , I_{dom} and TC of 21 graphs.

Graphs	I_α	I_{fd}	I_{nis}	I_{nm}	I_{dom}	TC
1	1.522	2.25	1.643	1.406	1.712	60
2	1.922	2.156	1.727	1.379	1.688	76
3	0.722	2	2.022	0.722	2.022	100
4	0	2.322	1.349	1.349	1.704	160
5	1.922	2.242	1.349	1.361	1.741	172
6	1.922	2.246	1.650	1.361	1.713	290
7	1.522	2.171	1.650	1.352	1.712	212
8	1.522	2.122	1.760	1.299	1.952	230
9	1.522	2.292	1.361	1.314	1.719	482
10	0.722	2.252	1.361	1.325	1.890	292
11	0.971	2.292	1.677	1.314	1.719	504
12	1.922	2.292	1.361	1.314	1.741	511
13	1.922	2.189	1.677	1.322	1.927	566
14	1.522	2.306	1.352	1.272	1.709	1278
15	1.522	2.271	1.352	1.287	1.890	1316
16	0.971	2.236	1.685	1.296	1.719	1394
17	1.371	2.217	1.352	1.296	1.923	1396
18	0.722	2.311	1.298	1.236	1.863	3216
19	1.522	2.281	1.299	1.252	1.709	3290
20	0.971	2.308	1.149	1.207	2.001	7806
21	0	2.322	0.650	1.169	2.060	18180

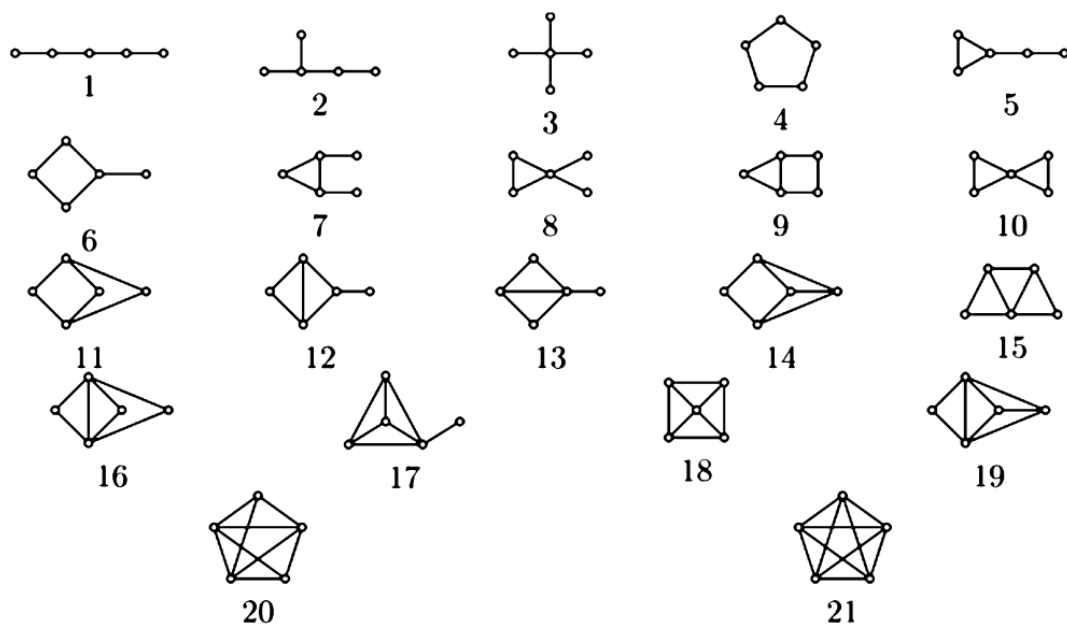


Fig. 4. All connected graphs with 5 vertices which have increased topological complexity measures.

第一个重要结果是图3的 I_{nis} 等于图3的 I_{dom} 。图3是星形图 S_5 。

$$\begin{aligned}
I_{dom}(S_n) &= \log(2^{n-1} + 1) - \frac{1}{2^{n-1} + 1} \left(\sum_{i=1}^{n-3} \binom{n-1}{i} \log \left(\binom{n-1}{i} \right) \right) - \frac{n \log n}{2^{n-1} + 1} \\
&= \log(2^{n-1} + 1) - \frac{n \log n}{2^{n-1} + 1} \\
&\quad - \frac{1}{2^{n-1} + 1} \left[\binom{n-1}{1} \log \binom{n-1}{1} + \cdots + \binom{n-1}{n-3} \log \binom{n-1}{n-3} \right].
\end{aligned}$$

I_{nis} of the stars is defined in [29], such that

$$\begin{aligned}
I_{nis}(S_n) &= \log(2^{n-1} + 1) - \frac{n \log n}{2^{n-1} + 1} - \frac{1}{2^{n-1} + 1} \left(\sum_{i=2}^{n-1} \binom{n-1}{i} \log \left(\binom{n-1}{i} \right) \right) \\
&= \log(2^{n-1} + 1) - \frac{n \log n}{2^{n-1} + 1} \\
&\quad - \frac{1}{2^{n-1} + 1} \left[\binom{n-1}{2} \log \binom{n-1}{2} + \cdots + \binom{n-1}{n-2} \log \binom{n-1}{n-2} \right].
\end{aligned}$$

Because the equation $\binom{n-1}{2} = \binom{n-1}{n-2}$ is attained, the third terms of I_{dom} and I_{nis} are equal. Therefore, we obtain $I_{dom}(S_n) = I_{nis}(S_n)$.

感觉作者是不是写错了？应该是蓝线部分相等，才有 $I_{nis} = I_{dom}$ 。

如表1所示， I_α 的度量顺序为1.522七次、1.922五次、0.971三次、0.722三次、0两次和1.371一次。图4和图21是正则图，它们的自同构群根据它们的对称结构由一个轨道组成(their automorphism groups consist of one orbit based on their symmetry structure, 不太懂什么意思)。图21在21图中具有最大的拓扑复杂度度量。但是，它的 I_α 值为零。

从表1可以看出，由于图4和图21是正则图，对于n个顶点，它们的first-degree熵等于logn。

可以看出这些图的拓扑复杂度存在显著差异，可这些图的first-degree熵和拓扑复杂度值之间没有显著相关性。而且，表中 P_5 的复杂度最小，完全图 K_5 的拓扑复杂度最大。所以研究一下，用first-degree熵度量 P_n 和 K_n ，看看隐藏着什么秘密or问题。

初步观察：

From the last equation, we can compute I_{fd} of P_n , and we add the values of I_{fd} for K_n ($3 \leq n \leq 6$).

$I_{fd}(P_3) = 1.5$, $I_{fd}(K_3) = 1.585$ and $I_{fd}(K_3) - I_{fd}(P_3) = 0.085$

$I_{fd}(P_4) = 1.919$, $I_{fd}(K_4) = 2$ and $I_{fd}(K_4) - I_{fd}(P_4) = 0.081$

$I_{fd}(P_5) = 2.25$, $I_{fd}(K_5) = 2.322$ and $I_{fd}(K_5) - I_{fd}(P_5) = 0.072$

$I_{fd}(P_6) = 2.521$, $I_{fd}(K_6) = 2.585$ and $I_{fd}(K_6) - I_{fd}(P_6) = 0.064$.

It is understood that if the order of paths increases, the weight of the vertices with degree two also increases. Therefore, the difference $I_{fd}(K_n) - I_{fd}(P_n)$ decreases, while n increases.

有点规律，但好像又说明不了什么。

再看看支配熵的表现：

It is time to mention the domination polynomials of the paths. For every $n \geq 4$, the domination polynomial of a path graph is defined by [38]:

$$D(P_n, x) = x[D(P_{n-1}, x) + D(P_{n-2}, x) + D(P_{n-3}, x)]$$

with initial conditions $D(P_1, x) = x$, $D(P_2, x) = x^2 + 2x$, and $D(P_3, x) = x^3 + 3x^2 + x$. Therefore, the total number of dominating sets of paths for $n \geq 4$ is attained as

$$\gamma_s(P_n) = \gamma_s(P_{n-1}) + \gamma_s(P_{n-2}) + \gamma_s(P_{n-3})$$

with initial conditions $\gamma_s(P_1) = 1$, $\gamma_s(P_2) = 3$ and $\gamma_s(P_3) = 5$.

Now, we calculate I_{dom} of P_n and K_n for $3 \leq n \leq 6$.

$I_{dom}(P_3) = 1.371$, $I_{dom}(K_3) = 1.45$ and $I_{dom}(K_3) - I_{dom}(P_3) = 0.079$

$I_{dom}(P_4) = 1.390$, $I_{dom}(K_4) = 1.809$ and $I_{dom}(K_4) - I_{dom}(P_4) = 0.419$

$I_{dom}(P_5) = 1.712$, $I_{dom}(K_5) = 2.060$ and $I_{dom}(K_5) - I_{dom}(P_5) = 0.348$

$I_{dom}(P_6) = 1.829$, $I_{dom}(K_6) = 3.110$ and $I_{dom}(K_6) - I_{dom}(P_6) = 1.281$.

K_n 和 P_n 的 I_{dom} 差异比较明显，对比3~6阶相同图的 I_{fd} 差异。差异是相对明显了，但没有规律。作者也没有继续进行说明。

6.结论

1. 我们得到了完全图、星图、双星图、梳状图、友谊图和细分星图的支配熵；
2. 我们定义了细分星图的支配多项式。
3. 我们对21个图表的五个熵度量进行了一些观察。
4. 我们发现星星的 I_{dom} 和 I_{nis} 是相等的。