Demystification of Graph and Information Entropy

Bryce Frederickson
Utah State University

DEMYSTIFICATION OF GRAPH AND INFORMATION ENTROPY

by

Bryce Frederickson

Capstone submitted in partial fulfillment of the requirements for graduation with

University Honors

with a major in

Mathematics

in the Department of Mathematics and Statistics

Approved:

Capstone Mentor
Dr. David Brown

Departmental Honors Advisor
Dr. David Brown

Committee Member [optional: type name or N/A]
Dr. Brent Thomas

University Honors Program Director
Dr. Kristine Miller

UTAH STATE UNIVERSITY
Logan, UT

Spring 2020

Acknowledgements

This work was done as a capstone project for the Utah State University Honors Program under the mentorship of Dr. David Brown and with the help of Dr. Brent Thomas of the Department of Mathematics and Statistics in the College of Science.

Demystification of Graph and Information Entropy

Bryce Frederickson

May 5, 2020

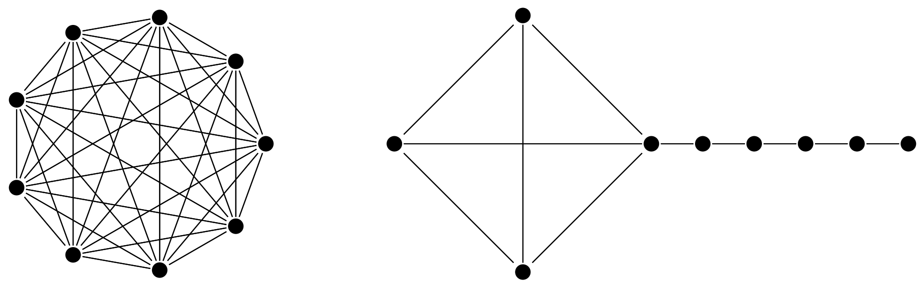
Abstract

Shannon entropy is an information-theoretic measure of unpredictability in probabilistic models. Recently, it has been used to form a tool, called the von Neumann entropy, to study quantum mechanics and network flows by appealing to algebraic properties of graph matrices. But still, little is known about what the von Neumann entropy says about the combinatorial structure of the graphs themselves. This paper gives a new formulation of the von Neumann entropy that describes it as a rate at which random movement settles down in a graph. At the same time, this new perspective gives rise to a generalization of von Neumann entropy to directed graphs, thus opening a new branch of research. Finally, it is conjectured that a directed cycle maximizes von Neumann entropy for directed graphs on a fixed number of vertices.

Introduction

A thief is loose in a building. From what we can tell, her movement is random, with the caveat that the more accessible places there are from where she is, the more likely she is to quickly move on to a new place, but she may return at some point. The thief was last seen at position x a short time ago. What is the probability that she's still there now? Are there structural properties of the building layout that would make this probability greater? In other words, do some building structures lend themselves to predictability of the thief's movement than others?

Consider the following two floor plan structures for buildings with nine rooms. The first consists of all nine rooms stacked on top of each other with an elevator making each room accessible from any other room. The second has four rooms stacked in the same way with an elevator, and on the bottom floor there's a door leading to another room, which has a door leading to another, and so on until the last of the nine rooms is reached. Figure 1 shows these connections graphically. Later on in this paper, we will develop a tool for quantifying the unpredictability of these graphs in the sense described, called the von Neumann entropy. The first building has a von Neumann entropy of 3, while the second has a von Neumann entropy of about 2.68, indicating that the structure of the first floor plan is less predictable than the second. The higher the entropy, the more unpredictable.



$H(G_1) = 3.00$ $\quad$ $H(G_2) = 2.68$

Figure 1: A graphical depiction of two building floor plans. Rooms are represented by dots, and a line is drawn between two dots if there is a direct connection between the corresponding rooms by a door or elevator. By our analysis, the first floor plan has higher entropy and is less predictable.

We will build our discussion from the Shannon entropy of a probability distribution as a measure of unpredictability. To illustrate what we mean, let us compare the weather patterns in two imaginary locations. In these locations, there are only three types of weather: sunny, cloudy, and rainy. In the first location, any of the three weather types is equally likely to occur on any given day, independently of the days surrounding it. In the second location, 98% of the days are sunny, 1% are cloudy, and 1% are rainy, again, with weather between days completely independent of other days. You might think that weather in the second location is more predictable, and if so, then Shannon entropy would agree. We will come back to this example later in more detail.

The rest of this paper will be a rigorous mathematical development of Shannon and von Neumann entropy in the context of probability distributions, graphs, and directed graphs.

What is entropy anyway?

To begin, let $\mathbf{p} = (p_1, p_2, \ldots, p_n)$ be a *discrete probability distribution*; that is, each $p_i \geq 0$, and

$$\sum_{i=1}^{n} p_n = 1.$$

The *Shannon entropy* of $\mathbf{p}$ is given by

$$H(\mathbf{p}) = \sum_{i=1}^{n} -p_i \log_2 p_i,$$

where $0 \log_2 0$ is defined to be 0 since

$$\lim_{x \to 0} x \log_2 x = 0.$$

A probability distribution, such as $\mathbf{p}$, should be thought of in the context of disjoint events $X_1, \ldots X_n$ (events that have no effect on one another) that occur with probability $p_1, \ldots, p_n$, respectively. The *surprisal* of X_i is given by

$$-\log_2 p_i, \quad \text{or} \quad \log_2 \frac{1}{p_i}.$$

The surprisal is an attempt to quantify the amount of surprise felt upon experiencing event X_i So, for example, the surprisal of rolling a 1 on a standard die (probability 1/6) is about 2.6, while the probability of rolling two 1's from two dice (probability 1/36) is twice that: about 5.2. Similarly, if you win a 1,000,000:1 lottery, the surprisal of your win is about 20, while if you win a 2,000,000:1 lottery, the surprisal is about 21. (Really, is it *that* much more surprising?)

The expected, or average, surprisal is precisely the Shannon entropy. Therefore, the Shannon entropy measures how unpredictable a system or collection of events, such as $X_1, \ldots, X_n$, is.

Let's see the computations from our weather example. The weather in the first location can be modeled by the discrete probability distribution $\mathbf{p_1} = (1/3, 1/3, 1/3)$, where sunny, cloudy, and rainy weather each have probability 1/3. Then no matter what the weather, the surprisal is $\log_2 3 \approx 1.585$, so the Shannon entropy is

$$H(\mathbf{p}_1) \approx 1.585$$

as well. The weather in the second location can be modeled by $\mathbf{p_2} = (.98, .01, .01)$. Then with 98% probability, the weather is sunny and the surprisal is only $-\log_2 .98 \approx 0.0291$. On the other hand, 2% of the days are much more "surprising", and the surprisal is $\log_2 100 \approx 6.644$. But in the end, these more surprising days are so few and far between, the expected surprisal is merely

$$H(\mathbf{p}_2) \approx .98(0.0291) + .02(6.644) \approx 0.161.$$

Where the weather is described by $\mathbf{p_2} = (0.98, 0.01, 0.01)$, a cloudy or rainy day is quite surprising, but in general, one is not surprised by the weather.

We now establish some characterizing properties of the Shannon entropy function $H \colon \mathcal{P}^n \to \mathbb{R}$, where $\mathcal{P}^n \subset \mathbb{R}^n$ is the set of all discrete probability distributions of the form $(p_1, \ldots, p_n)$, see [13].

Proposition 1. *The Shannon entropy H is the unique function that satisfies the following properties:*

1. *For each n, H is continuous.*
2. *For each n, $H(1/n, 1/n, \ldots, 1/n) = \log_2 n$.*

3. *If* $\mathbf{q} = (q_1, \ldots q_\ell)$*,* $\mathbf{p}_1 = (p_{1,1}, \ldots, p_{1,k_1}), \ldots, \mathbf{p}_\ell = (p_{\ell,1}, \ldots, p_{\ell,k_\ell})$ *are discrete probability distributions, then so is* $(q_1\mathbf{p}_1, \ldots, q_\ell\mathbf{p}_\ell)$*, and*

$$H(q_1\mathbf{p}_1, \ldots, q_\ell\mathbf{p}_\ell) = H(\mathbf{q}) + \sum_{i=1}^{\ell} q_i H(\mathbf{p}_i).$$

Proof. We first show that H indeed satisfies those properties. It is clear that H is continuous, and we can simply evaluate, for any n,

$$\begin{aligned} H\left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right) &= \sum_{i=1}^{n} -\frac{1}{n}\log_2\frac{1}{n} \\ &= n\left(-\frac{1}{n}\log_2\frac{1}{n}\right) \\ &= -\log_2\frac{1}{n} \\ &= \log_2 n. \end{aligned}$$

Now we verify the third property. Note first that $q_i p_{i,j} \geq 0$ for all i, j, and

$$\begin{aligned} \sum_{i=1}^{\ell}\sum_{j=1}^{k_i} q_i p_{i,j} &= \sum_{i=1}^{n} q_i \sum_{j=1}^{k_i} p_{i,j} \\ &= \sum_{i=1}^{n} q_i \\ &= 1, \end{aligned}$$

so $(q_1\mathbf{p}_1, \ldots, q_\ell\mathbf{p}_\ell)$ is a discrete probability distribution, as claimed. Furthermore,

$$\begin{aligned} H(q_1\mathbf{p}_1, \ldots, q_\ell\mathbf{p}_\ell) &= \sum_{i=1}^{\ell}\sum_{j=1}^{k_i} -q_i p_{i,j}\log_2 q_i p_{i,j} \\ &= \sum_{i=1}^{\ell}\sum_{j=1}^{k_i} (-q_i p_{i,j}\log_2 q_i - q_i p_{i,j}\log_2 p_{i,j}) \\ &= \sum_{i=1}^{\ell} -q_i\log_2 q_i \sum_{j=1}^{k_i} p_{i,j} + \sum_{i=1}^{\ell} q_i \sum_{j=1}^{k_i} -p_{i,j}\log_2 p_{i,j} \\ &= \sum_{i=1}^{\ell} -q_i\log_2 q_i \cdot 1 + \sum_{i=1}^{\ell} q_i \sum_{j=1}^{k_i} -p_{i,j}\log_2 p_{i,j} \\ &= H(\mathbf{q}) + \sum_{i=1}^{\ell} q_i H(\mathbf{p}_i). \end{aligned}$$

Now we show that no other functional has these properties. Suppose that F is a functional satisfying all three properties. Then for every $\mathbf{p} \in \mathcal{P}^n \cap \mathbb{Q}^n$, we can write $\mathbf{p} = \left(\frac{a_1}{b}, \ldots, \frac{a_n}{b}\right)$, where $a_1, \ldots, a_n$ are nonnegative integers adding to $b > 0$. Then by the second and third properties,

$$\begin{aligned} \log_2 b &= F\left(\frac{1}{b}, \ldots, \frac{1}{b}\right) \\ &= F\left(\frac{a_1}{b}, \ldots, \frac{a_n}{b}\right) + \sum_{i=1}^{n} \frac{a_i}{b} F\left(\frac{1}{a_i}, \ldots, \frac{1}{a_i}\right) \\ &= F\left(\frac{a_1}{b}, \ldots, \frac{a_n}{b}\right) + \sum_{i=1}^{n} \frac{a_i}{b} \log_2 a_i, \end{aligned}$$

so

$$\begin{aligned} F\left(\frac{a_1}{b}, \ldots, \frac{a_n}{b}\right) &= \log_2 b - \sum_{i=1}^{n} \frac{a_i}{b} \log_2 a_i \\ &= \sum_{i=1}^{n} \frac{a_i}{b} \log_2 b - \sum_{i=1}^{n} \frac{a_i}{b} \log_2 a_i \\ &= \sum_{i=1}^{n} -\frac{a_i}{b} \left(\log_2 a_i - \log_2 b\right) \\ &= \sum_{i=1}^{n} -\frac{a_i}{b} \log_2 \frac{a_i}{b} \\ &= H\left(\frac{a_1}{b}, \ldots, \frac{a_n}{b}\right). \end{aligned}$$

Finally, since for each n, F and H are continuous functions that agree on $\mathcal{P}^n \cap \mathbb{Q}^n$, and since $\mathcal{P}^n \cap \mathbb{Q}^n$ is dense in $\mathcal{P}^n$, it follows that $F = H$. $\square$

In [13], Alfréd Rényi generalized the Shannon entropy by weakening condition 3 of Proposition 1. The *Rényi α-entropy* is given by

$$H_\alpha(p_1, \ldots, p_n) = \frac{1}{1-\alpha} \log_2 \sum_{i=1}^{n} p_i^\alpha$$

for any $\alpha > 1$.

Proposition 2. *The functional H_α satisfies the following properties:*

1. *For each n, H_α is continuous.*
2. *For each n, $H_\alpha(1/n, \ldots, 1/n) = \log_2 n$.*
3. *If $\mathbf{q} = (q_1, \ldots, q_\ell)$ and $\mathbf{p} = (p_1, \ldots, p_k)$ are discrete probability distributions, then so is $\mathbf{qp} := (q_1\mathbf{p}, \ldots, q_\ell\mathbf{p})$, and*

$$H(\mathbf{qp}) = H(\mathbf{q}) + H(\mathbf{p}).$$

Proof. It is clear that H_α is continuous, and we can compute

$$\begin{aligned} H_\alpha(1/n,\ldots,1/n) &= \frac{1}{1-\alpha}\log_2\sum_{i=1}^{n}\frac{1}{n^\alpha} \\ &= \frac{1}{1-\alpha}\log_2\frac{n}{n^\alpha} \\ &= \frac{1}{1-\alpha}\log_2 n^{1-\alpha} \\ &= \log_2 n. \end{aligned}$$

That $\mathbf{qp}$ is a discrete probability distribution follows from property 3 of Proposition 1, so we just need to show $H(\mathbf{qp}) = H(\mathbf{q}) + H(\mathbf{p})$. To this end, we compute:

$$\begin{aligned} H(\mathbf{qp}) &= \frac{1}{1-\alpha}\log_2\sum_{i=1}^{\ell}\sum_{j=1}^{k}(q_i p_j)^\alpha \\ &= \frac{1}{1-\alpha}\log_2\left(\sum_{i=1}^{\ell} q_i^\alpha\right)\left(\sum_{j=1}^{k} p_j^\alpha\right) \\ &= \frac{1}{1-\alpha}\left(\log_2\sum_{i=1}^{\ell} q_i^\alpha + \log_2\sum_{j=1}^{k} p_j^\alpha\right) \\ &= \frac{1}{1-\alpha}\log_2\sum_{i=1}^{\ell} q_i^\alpha + \frac{1}{1-\alpha}\log_2\sum_{j=1}^{k} p_j^\alpha \\ &= H(\mathbf{q}) + H(\mathbf{p}). \end{aligned}$$

□

The next relationship connects the Shannon and Rényi entropy functionals, and shows that the Shannon entropy $H(\mathbf{p})$ is a special limiting case of the Rényi entropy $H_\alpha(\mathbf{p})$.

Proposition 3. *For any discrete probability distribution* $\mathbf{p}$*,*

$$\lim_{\alpha\to 1} H_\alpha(\mathbf{p}) = H(\mathbf{p}).$$

Proof. Let $\mathbf{p} \in \mathcal{P}^n$. Then because

$$\left.\log_2\sum_{i=1}^{n} p_i^\alpha\right|_{\alpha=1} = 0$$

and

$$\frac{\partial}{\partial \alpha} \log_2 \sum_{i=1}^{n} p_i^{\alpha} \bigg|_{\alpha=1} = \frac{1}{\log 2} \frac{1}{\sum_{i=1}^{n} p_i^{\alpha}} \sum_{i=1}^{n} p_i^{\alpha} \log p_i \bigg|_{\alpha=1}$$
$$= \sum_{i=1}^{n} p_i \log_2 p_i$$
$$= -H_{\alpha}(\mathbf{p}),$$

we have that

$$\log_2 \sum_{i=1}^{n} p_i^{\alpha} = -H(\mathbf{p})(\alpha - 1) + g(\alpha),$$

where

$$\lim_{\alpha \to 1} \frac{g(\alpha)}{\alpha - 1} = 0.$$

Therefore,

$$H_{\alpha}(\mathbf{p}) = \frac{-H(\mathbf{p})(\alpha - 1) + g(\alpha)}{1 - \alpha} \to H(\mathbf{p}) \quad \text{as } \alpha \to 1.$$

$\square$

In this paper, we will explore objects, like the graphs that represented the buildings in the introduction, and determine what properties of those objects maximize and minimize entropies derived from the Shannon and Rényi entropies we have discussed. To these ends the following proposition will be helpful.

Proposition 4. *For any* $\mathbf{p} \in \mathcal{P}^n$*,*

$$0 \leq H(\mathbf{p}) \leq \log_2 n,$$

and for any $\alpha > 1$*,*

$$0 \leq H_{\alpha}(\mathbf{p}) \leq \log_2 n.$$

Furthermore, $H(\mathbf{p}) = 0$ *iff* $H_{\alpha}(\mathbf{p}) = 0$ *iff* $\mathbf{p}$ *is a permutation of* $(1, 0, \ldots, 0)$*, and* $H(\mathbf{p}) = \log_2 n$ *iff* $H_{\alpha}(\mathbf{p}) = \log_2 n$ *iff* $\mathbf{p} = (1/n, \ldots, 1/n)$*.*

Before we prove this proposition, recall that a function $f : [a, b] \to \mathbb{R}$ is *strictly convex* if for all $x, y \in [a, b]$ with $x \neq y$ and for all $t \in (0, 1)$,

$$tf(x) + (1 - t)f(y) > f(tx + (1 - t)y).$$

Informally speaking, f is strictly convex if every secant line stays above the curve. It follows from the Mean Value Theorem that if $f''(x) > 0$ for all $x \in (a, b)$, then f is strictly convex.

Lemma 1. *If $g\colon \mathcal{P}^n \to \mathbb{R}$ is a function of the form*

$$g(p_1, \dots, p_n) = \sum_{i=1}^{n} f(p_i),$$

where $f\colon [0,1] \to \mathbb{R}$ is a strictly convex function, then g is minimized iff $p_1 = \cdots = p_n = 1/n$.

Proof. Note that because f is strictly convex, for all $\mathbf{p} = (p_1, \dots, p_n) \in \mathcal{P}^n$ with $p_i \neq p_j$ for some $1 \leq i < j \leq n$, we have for any $t \in (0,1)$ that

$$\begin{aligned}
& g(p_1, \dots, p_i, \dots, p_j, \dots, p_n) \\
& = (1-t+t)f(p_i) + (1-t+t)f(p_j) + \sum_{\substack{k=1 \\ k\neq i,j}}^{n} f(p_k) \\
& = (tf(p_i) + (1-t)f(p_j)) + ((1-t)f(p_i) + tf(p_j)) + \sum_{\substack{k=1 \\ k\neq i,j}}^{n} f(p_k) \\
& > f(tp_i + (1-t)p_j) + f((1-t)p_i + tp_j) + \sum_{\substack{k=1 \\ k\neq i,j}}^{n} f(p_k) \\
& = g(p_1, \dots, tp_i + (1-t)p_j, \dots, (1-t)p_i + tp_j, \dots, p_n).
\end{aligned}$$

Now consider the following algorithm. Let $\mathbf{p} = (p_1, \dots, p_n) \in \mathcal{P}^n$.

1. If $p_1 = \cdots p_n = 1/n$, terminate the algorithm.
2. Otherwise, let $i = \min\{k \mid p_k \neq 1/n\}$.
3. If $p_i < 1/n$, there must exist $j > i$ such that $p_j > 1/n$. Similarly, if $p_i > 1/n$, there must exist $j > i$ such that $p_j < 1/n$. In either case, there exists $t \in (0,1)$ such that $tp_i + (1-t)p_j = 1/n$.
4. Replace p_i with $1/n$ and p_j with $(1-t)p_i + tp_j$ so that the value of $g(\mathbf{p})$ decreases and $\mathbf{p}$ is still in $\mathcal{P}^n$.
5. Go back to step 1.

It is clear that this algorithm terminates in at most $n-1$ iterations with $\mathbf{p} = (1/n, \dots, 1/n)$. Thus $\mathbf{p}$ minimizes g on $\mathcal{P}^n$ iff $\mathbf{p} = (1/n, \dots, 1/n)$. □

Now we are in a position to give a simple proof of Proposition 4.

Proof. Let $\mathbf{p} \in \mathcal{P}^n$. First note that $-x \log_2 x \geq 0$ for all $x \in [0,1]$ with equality iff $x \in \{0,1\}$. Thus

$$H(\mathbf{p}) \geq 0,$$

with equality iff $\mathbf{p} \in \{0,1\}^n$, which in $\mathcal{P}^n$ means $\mathbf{p}$ is a permutation of $(1,0,\dots,0)$.

Similarly, we know that for $\alpha > 1$

$$0 < \sum_{i=1}^{n} p_i^{\alpha} \le \sum_{i=1}^{n} p_i = 1,$$

where the inequality is equality iff $\mathbf{p} \in \{0,1\}^n$. Then since $\log_2$ is a strictly increasing function and $1-\alpha < 0$, we have

$$H_\alpha(\mathbf{p}) \ge \frac{1}{1-\alpha} \log_2 1 = 0,$$

with equality iff $\mathbf{p}$ is a permutation of $(1, 0, \dots, 0)$.

Next, we note that the functions $f_1, f_2 \colon [0,1] \to \mathbb{R}$ given by

$$f_1(x) = \begin{cases} x \log_2 x & \text{if } x > 0, \\ 0 & \text{if } x = 0 \end{cases} \qquad \text{and} \qquad f_2(x) = x^\alpha, \quad \alpha > 0$$

are strictly convex since

$$f_1''(x) = \frac{1}{x \log 2} > 0$$

and

$$f_2''(x) = \alpha(\alpha - 1)x^{\alpha-2} > 0.$$

Therefore, by Lemma 1

$$\sum_{i=1}^{n} p_i \log_2 p_i \qquad \text{and} \qquad \sum_{i=1}^{n} p_i^{\alpha}$$

are minimized iff $\mathbf{p} = (1/n, \dots, 1/n)$. Taking this one step further,

$$H(\mathbf{p}) = -\sum_{i=1}^{n} p_i \log_2 p_i \qquad \text{and} \qquad H_\alpha(\mathbf{p}) = \frac{1}{1-\alpha} \log_2 \sum_{i=1}^{n} p_i^{\alpha}$$

are maximized iff $\mathbf{p} = (1/n, \dots, 1/n)$, since $\log_2$ is a strictly increasing function and $1 - \alpha < 0$. Finally, we note that indeed

$$H(1/n, \dots, 1/n) = H_\alpha(1/n, \dots, 1/n) = \log_2 n,$$

as we've seen in Propositions 1 and 2. □

Graph Preliminaries

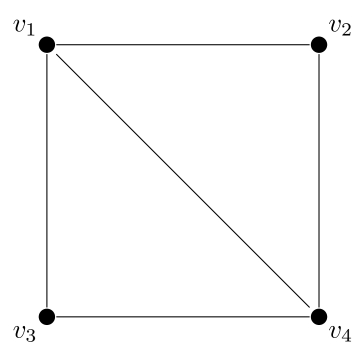
The focus of our study is to use entropy somehow to quantify unpredictability in other types of systems. We start with combinatorial objects called graphs.

A *graph* consists of a set V of *vertices* and a set E of pairs of vertices, called *edges*. We typically write the edge $\{u, v\}$ simply as uv. If $uv \in E$, we say the vertices u and v are *adjacent*, and that u is *incident* to uv. If vertex v is adjacent to k other vertices, we say the *degree* of vertex v is k, and we write $d(v) = k$. We will typically refer to a specific graph (the vertices and edges that comprise it) by a single capital letter, probably G.

We draw a graph by placing each vertex at a point in the plane, and we draw a curve or line segment with endpoints at u and v whenever uv is an edge. However, a graph has no geometric properties, so the physical distance between vertices and the angles between edges, as well as the straightness of edges, are irrelevant. What matters is which vertices are adjacent to which other vertices. This information can be represented by other means, such as via a matrix. For example, if we label the vertices so that $V = \{v_1, v_2, \ldots, v_n\}$, then we can construct the *adjacency matrix* $A = [a_{ij}]$ of graph G by

$$a_{ij} = \begin{cases} 1 & \text{if } v_i v_j \in E \\ 0 & \text{otherwise} \end{cases} .$$

We can take this a step further to construct the *Laplacian matrix* L of G in the following way. Let $D = [d_{ij}]$ be the diagonal matrix such that $d_{ii} = d(v_i)$ (and $d_{ij} = 0$ if $i \neq j$). Then define $L = D - A$. If the graph has at least one edge, we can then construct the *normalized Laplacian matrix* $\bar{L}$ of G by dividing L by its trace (i.e. the sum of its diagonal entries).



$$A = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix}$$

$$L = \begin{bmatrix} 3 & -1 & -1 & -1 \\ -1 & 2 & 0 & -1 \\ -1 & 0 & 2 & -1 \\ -1 & -1 & -1 & 3 \end{bmatrix}$$

$$\bar{L} = \begin{bmatrix} 3/10 & -1/10 & -1/10 & -1/10 \\ -1/10 & 2/10 & 0 & -1/10 \\ -1/10 & 0 & 2/10 & -1/10 \\ -1/10 & -1/10 & -1/10 & 3/10 \end{bmatrix}$$

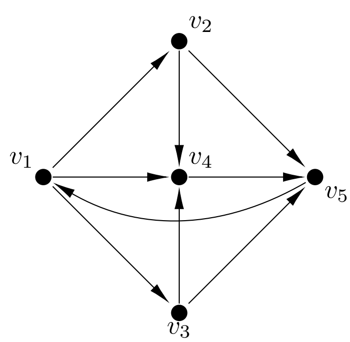
Figure 2: A graph with its adjacency, Laplacian, and normalized Laplacian matrices. The adjacency matrix has a 1 in row i column j when there's an edge between v_i and v_j. In the Laplacian matrix, those 1's become -1's, and the degrees go on the diagonal. The Laplacian matrix is divided by 10 To obtain the normalized Laplacian matrix so that its diagonal entries add to 1.

Now we move onto a slightly more general class of objects called directed graphs. A *directed graph*, or *digraph*, consists of a set V of *vertices* and a set A of *ordered* pairs of vertices, called *arcs*. If $(u, v) \in A$, we write $u \to v$, which is read u *beats* v. In our discussion, we will only consider *loopless* digraphs, where there are no arcs of the form (v, v). The number of vertices beaten by vertex v is called the *out-degree* of v, written $d^+(v)$, and the number of vertices that beat v is called the *in-degree* of v, written $d^-(v)$. We will also refer to a digraph (the vertices and arcs that comprise it) by a single letter, probably Γ.

We draw a digraph similar to how we draw a graph. The vertices are placed at points in a plane, and an arrow is drawn from u to v whenever $u \to v$. After giving the vertices a labeling $v_1, \ldots, v_n$, we can also define the adjacency matrix $A = [a_{ij}]$ of Γ by

$$a_{ij} = \begin{cases} 1 & \text{if } v_i \to v_j \\ 0 & \text{otherwise} \end{cases} .$$

We can even discuss the *Laplacian matrix* L of Γ by defining the diagonal matrix $D = [d_{ij}]$, where $d_{ii} = d^+(v_i)$, and taking $L = D - A$, similar to the graph case. And again, dividing L by its trace gives the *normalized Laplacian matrix* $\bar{L}$ of Γ.



$$A = \begin{bmatrix} 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$L = \begin{bmatrix} 3 & -1 & -1 & -1 & 0 \\ 0 & 2 & 0 & -1 & -1 \\ 0 & 0 & 2 & -1 & -1 \\ 0 & 0 & 0 & 1 & -1 \\ -1 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\bar{L} = \begin{bmatrix} 3/9 & -1/9 & -1/9 & -1/9 & 0 \\ 0 & 2/9 & 0 & -1/9 & -1/9 \\ 0 & 0 & 2/9 & -1/9 & -1/9 \\ 0 & 0 & 0 & 1/9 & -1/9 \\ -1/9 & 0 & 0 & 0 & 1/9 \end{bmatrix}$$

Figure 3: A digraph with its adjacency, Laplacian, and normalized Laplacian matrices. The adjacency matrix has a 1 in row i column j when $v_i \to v_j$. In the Laplacian matrix, those 1's become -1's, and the out-degrees go on the diagonal. The Laplacian matrix is divided by 9 To obtain the normalized Laplacian matrix so that its diagonal entries add to 1.

In fact, these digraph constructions can be viewed as generalizations of the graph constructions. If G is a graph with vertex set V and edge set E, then consider Γ with vertex set V and arc set A such that for each edge uv of G, we have $u \to v$ and $v \to u$ in Γ. Then G and Γ have the same adjacency matrix, Laplacian matrix, and normalized Laplacian matrix, and for our purposes in this paper, G and Γ may be considered the same object.

Graph Entropy

The idea to use entropy to study graphs is not new, and its exact implementation has varied to fit several contexts, see [7]. One use, from which we base our discussion, arises from the field of quantum information theory, as in [11] and [17]. A quantum state has an associated density matrix, which is Hermitian, positive semi-definite, and has unit trace (trace equal to 1). This implies, by a theorem in Linear Algebra and Matrix Theory, that the spectrum (the collection of eigenvalues) of the density matrix can be treated as a discrete probability distribution ('Hermetian' guarantess eigenvalues are real, 'positive semi-definite' means eigenvalues are nonnegative, 'unit trace' means the eigenvalues sum to 1). The von Neumann entropy of the state is obtained by applying the Shannon entropy to that spectrum. The normalized Laplacian matrix of an undirected graph has similar properties, so one may think of it as a density matrix of some physical system, and it makes sense to define the von Neumann entropy of a graph in this way. This concept has been widely studied in recent years, see [2, 4, 5, 6, 18].

This paper will extend those ideas to the context of directed graphs, for which a new approach is needed since the spectrum of the normalized Laplacian is complex-valued and has little hope of being interpreted as a discrete probability distribution. But before we get to that, we will highlight some observations about the well-known undirected case.

If a graph G has at least one edge, we define the *von Neumann entropy* of G by

$$H(G) = \sum_{\lambda \in \bar{\Lambda}} -\lambda \log_2 \lambda,$$

where $\bar{\Lambda}$ is the spectrum of $\bar{L}$; that is, the multiset of its eigenvalues (with multiplicity).

The following proposition says that we can talk about $H(G)$ without ambiguity or potential misunderstanding assuming calculations are done correctly. In other terms, it is saying that if one person calculates $H(G)$, their result will be the same as someone else who calculates $H(G)$, even if the names of vertices or edges of G are labeled differently.

Proposition 5. *$H(G)$ is well-defined.*

Proof. There are two concerns here.

First, we note that the construction of $\bar{L}$ is dependent on the labelling of the vertices of G. However, we will show that $\bar{\Lambda}$ is the same regardless of the choice of labelling. Second, we must show that $\bar{L}$ is positive semi-definite; that is, all its eigenvalues are nonnegative.

First, let $(v_1, \ldots, v_n)$ and $(w_1, \ldots, w_n)$ be labellings of V. Clearly, one may be obtained from the other by permutation, so there exists a permutation matrix

P such that

$$\begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} = P \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}.$$

Now if $\bar{L}_v$ is the normalized Laplacian matrix constructed from labelling $(v_1, \dots, v_n)$, and $\bar{L}_w$ is the normalized Laplacian matrix constructed from labelling $(w_1, \dots, w_n)$, then

$$\bar{L}_w = P\bar{L}_v P^{-1}.$$

Thus the matrices are similar and have the same spectrum, and our first concern is resolved.

Now we use the labels $v_1, \dots, v_n$, and we also label the edges of G by $e_1, \dots, e_m$. We can now construct a signed incidence matrix $B = [b_{ik}]$ of G by taking

$$b_{ik} = \begin{cases} 1 & \text{if } e_k = v_i v_j \text{ for some } j > i \\ -1 & \text{if } e_k = v_i v_j \text{ for some } j < i \\ 0 & \text{otherwise} \end{cases}$$

Now consider the n by n matrix $C = BB^\intercal = [c_{ij}]$. We know that $c_{ii} = d(v_i)$ since row i of B has ± 1 for each edge incident to v_i. If $i \neq j$, then c_{ij} is the dot product of rows i and j of B. Vertices v_i and v_j are not adjacent, then they share no edge and $c_{ij} = 0$. Otherwise, they share exactly one edge, namely e_k. Because either $i < j$ or $j < k$, we know that b_{ik} and b_{jk} have opposite signs, so $c_{ij} = b_{ik}b_{jk} = -1$. Now we can see that L agrees with C at every entry, so $L = BB^\intercal$ is positive semi-definite. Finally, we note that since G has at least one edge, L has positive trace, so $\bar{L}$ is also positive semi-definite. □

We note that the von Neumann entropy formula is actually just an application of the Shannon entropy to $\bar{\Lambda}$. As we've seen, $\bar{L}$ is positive semi-definite, and because it has trace 1, the eigenvalues are all nonnegative and add to 1, so $\bar{\Lambda}$ can be treated as a discrete probability distribution.

Furthermore, we note that 0 is always an eigenvalue of $\bar{L}$ since the elements in each row of $\bar{L}$ add to 0. Therefore, if G has n vertices, it has at most $n-1$ nonzero eigenvalues, so

$$H(G) \leq \log_2(n-1)$$

by Proposition 4. Dairyko et al. showed in [5] that this bound is tight and that it's achieved only by K_n: the *complete graph* on n vertices, where every possible edge is present. We give our own proof of that here. We use $\mathcal{G}_n$ to denote the collection of all graphs on n vertices with at least one edge.

Theorem 1. *The von Neumann entropy is maximized on $\mathcal{G}_n$ by K_n.*

Proof. First suppose that $H(G) = \log_2(n-1)$ for some $G \in \mathcal{G}_n$. We will show that $G = K_n$. By Proposition 4, $\bar{L}$ has spectrum

$$\bar{\Lambda} = \left\{ \frac{1}{n-1}^{(n-1)}, 0^{(1)} \right\},$$

where a superscript (m) denotes an algebraic multiplicity of m. Then L has spectrum

$$\Lambda = \left\{ t^{(n-1)}, 0^{(1)} \right\},$$

where $t = \mathrm{tr}L/(n-1) > 0$.

Let x_{ij} denote the vector with 1 in the ith component, -1 in the jth component, and 0 everywhere else. Since L is symmetric and $\vec{1}$, the vector of all ones, is an eigenvector of L corresponding to 0, the eigenspace corresponding to the eigenvalue t contains all vectors orthogonal to $\vec{1}$. In particular, each x_{ij} is an eigenvector of L corresponding to t. Now let v_i, v_j, v_k be three distinct vertices of G. If we write $L = [\ell_{ij}]$, the kth row of the computation $Lx_{ij} = tx_{ij}$ gives

$$\ell_{ki} - \ell_{kj} = 0.$$

Thus $v_i v_k \in E$ iff $v_j v_k \in E$. A similar computation reveals $v_i v_k \in E$ iff $v_i v_j \in E$. Therefore, v_i, v_j, and v_k either have no edges between them or form a triangle. Since this is true of any three vertices in G, and since G has at least one edge by assumption, we conclude that $G = K_n$.

Now we verify that $H(K_n) = \log_2(n-1)$. We note that the normalized Laplacian matrix of K_n is

$$\bar{L} = \frac{1}{n(n-1)} \begin{pmatrix} n-1 & -1 & \cdots & -1 \\ -1 & n-1 & & -1 \\ \vdots & & \ddots & \vdots \\ -1 & -1 & \cdots & n-1 \end{pmatrix} = \frac{nI - J}{n(n-1)},$$

where I is the n by n identity matrix, and J is the n by n all ones matrix.

The spectrum of J is easy to determine. First, $J\vec{1} = n\vec{1}$, so n is an eigenvalue of J. Also, since J has rank 1, 0 takes up the remaining $n-1$ eigenvalues. Thus the spectrum of J is $\{0^{(n-1)}, n^{(1)}\}$. $\bar{L}$ then has spectrum

$$\bar{\Lambda} = \left\{ \frac{n-0}{n(n-1)}^{(n-1)}, \frac{n-n}{n(n-1)}^{(1)} \right\} = \left\{ \frac{1}{n-1}^{(n-1)}, 0^{(1)} \right\}.$$

Thus

$$H(K_n) = H\left(\frac{1}{n-1}, \ldots, \frac{1}{n-1} \right) = \log_2(n-1).$$

□

We note that $H(G)$ is minimized on $\mathcal{G}_n$ by any graph with just one edge. It is then of interest to look at the class $\mathcal{CG}_n$ of *connected* graphs on n vertices, i.e. those for which there is a finite sequence of edges

$$uw_1, w_1w_2, w_2w_3, \ldots, w_{k-1}w_k, w_kv,$$

called a *path*, between any two vertices u and v. In [5], Dairyko et al. conjectured that $H(G)$ is minimized on $\mathcal{CG}_n$ by the *star* graph $K_{1,n-1}$, which consists of one central vertex adjacent to all others, with no other edges present.

An alternative approach to von Neumann entropy

Now that we've identified which graph has the highest von Neumann entropy, a natural question arises: What does that even mean? It would appear that high entropy seems to correlate with qualitative notions such as "connectedness" and "regularity", but can we make that more concrete? So far, we've looked at the normalized eigenvalues as a discrete probability distribution, but what do those probabilities correspond to? Well, at least for now, the answer is unclear.

However, we can give a reformulation of the von Neumann entropy that not only has a meaningful probabilistic interpretation, but also has a natural extension to directed graphs, whose Laplacian matrices may not even have real eigenvalues.

For a directed graph Γ with at least one arc, we define the *von Neumann entropy* of Γ to be

$$H(\Gamma) = \frac{1}{\log 2}\left(\operatorname{tr}(M) - \sum_{j=2}^{\infty} \frac{\operatorname{tr}(M^j)}{j(j-1)}\right),$$

where $M = (I - \bar{L})^\intercal$.

To make sense of this, we first look at this new matrix M. Recall that the diagonal entries of $\bar{L}$ are nonnegative numbers that add to 1, and that the off-diagonal entries of $\bar{L}$ are nonpositive. Therefore, the entries of M are all nonnegative. Furthermore, the entries of each row of $\bar{L}$ add to 0, so the entries of each column of M add to 1. In fact, each column of M is a discrete probability distribution! A matrix with this property is called a *Markov matrix*, and much is known about them.

Before we do more with this definition of von Neumann entropy for digraphs, we need to ensure that it doesn't conflict with the definition of von Neumann entropy for graphs.

Suppose G is a graph with associated digraph Γ. Because $M = (I - \bar{L})^\intercal$, we know that for every eigenvalue λ of $\bar{L}$, $1 - \lambda$ is an eigenvalue of M. Thus by our new definition,

$$\begin{aligned} H(\Gamma) &= \frac{1}{\log 2}\left(\operatorname{tr}(M) - \sum_{j=2}^{\infty} \frac{\operatorname{tr}(M^j)}{j(j-1)}\right) \\ &= \frac{1}{\log 2}\left(\sum_{\lambda \in \bar{\Lambda}} (1-\lambda) - \sum_{j=2}^{\infty} \sum_{\lambda \in \bar{\Lambda}} \frac{(1-\lambda)^j}{j(j-1)}\right) \\ &= \frac{1}{\log 2} \sum_{\lambda \in \bar{\Lambda}} \left((1-\lambda) - \sum_{j=2}^{\infty} \frac{(1-\lambda)^j}{j(j-1)}\right). \end{aligned}$$

Now recall that $-x \log x$ has power series expansion

$$1 - x - \sum_{j=2}^{\infty} \frac{(1-x)^j}{j(j-1)}$$

for $|x-1| \leq 1$. In our case, because M is a Markov matrix, $|1-\lambda| \leq 1$ for all $\lambda \in \bar{\Lambda}$, so we can simplify what we have to

$$\begin{aligned} H(\Gamma) &= \frac{1}{\log 2} \sum_{\lambda \in \Lambda} -\lambda \log \lambda \\ &= \sum_{\lambda \in \Lambda} -\lambda \log_2 \lambda \\ &= H(G) \end{aligned}$$

by our first definition, so we see that there's no conflict.

Now we look deeper into the meaning of this new approach. The matrix M describes a family of what are called *random walks* on Γ. A random walk starts at some vertex v_k, then at discrete time steps, it either moves to another vertex, or it stays put, according to some probability distribution associated with the current vertex. That probability distribution is given by the kth column of M. In particular, if $g = \mathrm{tr}\bar{L}$, then for each vertex v_ℓ such that $v_k \to v_\ell$, the walk has a probability of $1/g$ of moving from v_k to v_ℓ. The walk stays at v_k otherwise, with probability $1 - d^+(v_k)/g$.

More generally, M^j describes the random walk after j time steps. Let $w_j(v_k)$ denote a random walk of length j starting at v_k. The k, ℓ entry of M^j is the probability that $w_j(v_k)$ will end at v_ℓ. As $j \to \infty$, the probability distribution for each $w_j(v_k)$ will converge to a stable state dictated by an eigenvector of M corresponding to an eigenvalue of 1, (which is the same as a left eigenvector of $\bar{L}$ corresponding to an eigenvalue of 0.)

Now relating this to the von Neumann entropy, we have that

$$\mathrm{tr}(M^j) = \sum_{k=1}^{n} \mathrm{P}(w_j(v_k) \text{ ends at } v_k).$$

Therefore, the von Neumann entropy can be expressed as

$$H(\Gamma) = \frac{1}{\log 2} \left(n - 1 - \sum_{k=1}^{n} \sum_{j=2}^{\infty} \frac{P(w_j(v_k) \text{ ends at } v_k)}{j(j-1)} \right).$$

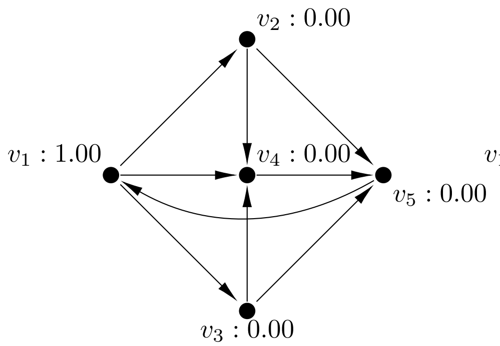
In this sense, the von Neumann entropy is a measure of how quickly a random walk will move away from its initial state and settle in to its limiting state. The more quickly it settles, the higher the entropy. Thus the entropy is, in fact, a measure of how "unpredictable" a graph is. Recall the metaphor of the thief moving randomly in the buildings described in the introduction.

This viewpoint also allows us to place general bounds on the von Neumann entropy. Recall that $g = \mathrm{tr}\bar{L}$.

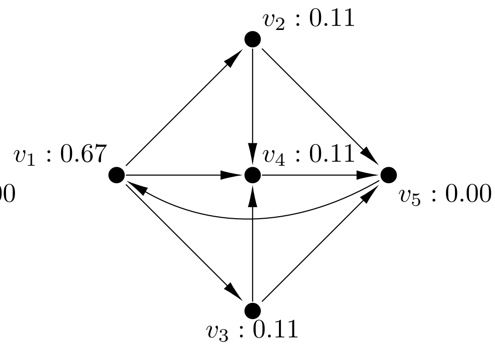
Theorem 2. *For any directed graph* Γ*,* $H(\Gamma) \leq H(\mathbf{d}^+)$*, where*

$$\mathbf{d}^+ = (d^+(v_1)/g, \ldots, d^+(v_n)/g)$$

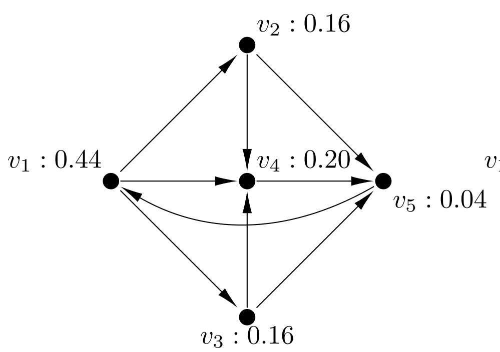
is the distribution of out-degrees in Γ*, and equality holds if and only if* Γ *has no (directed) cycles.*



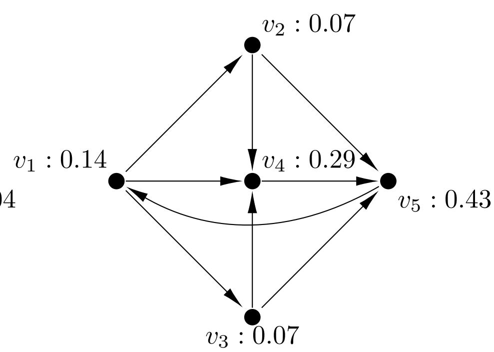
$$M^0 = \left[\begin{array}{c|cccc} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{array}\right]$$



$$M^1 = \left[\begin{array}{c|cccc} 0.67 & 0 & 0 & 0 & 0.11 \\ 0.11 & 0.78 & 0 & 0 & 0 \\ 0.11 & 0 & 0.78 & 0 & 0 \\ 0.11 & 0.11 & 0.11 & 0.89 & 0 \\ 0 & 0.11 & 0.11 & 0.11 & 0.89 \end{array}\right]$$



$$M^2 = \left[\begin{array}{c|cccc} 0.44 & 0.01 & 0.01 & 0.01 & 0.17 \\ 0.16 & 0.60 & 0 & 0 & 0.01 \\ 0.16 & 0 & 0.60 & 0 & 0.01 \\ 0.20 & 0.19 & 0.19 & 0.79 & 0.01 \\ 0.04 & 0.20 & 0.20 & 0.20 & 0.79 \end{array}\right]$$



$$M^{50} = \left[\begin{array}{c|cccc} 0.14 & 0.14 & 0.14 & 0.14 & 0.14 \\ 0.07 & 0.07 & 0.07 & 0.07 & 0.07 \\ 0.07 & 0.07 & 0.07 & 0.07 & 0.07 \\ 0.29 & 0.29 & 0.29 & 0.29 & 0.29 \\ 0.43 & 0.43 & 0.43 & 0.43 & 0.43 \end{array}\right]$$

Figure 4: Progression of $w_j(v_1)$ for $j = 0, 1, 2, 50$ on a digraph. The number next to v_k is the probability that the walk ends at v_k, taken from the $(k, 1)$ entry of M^j.

Proof. Clearly,

$$P(w_j(v_k) \text{ ends at } v_k) \geq P(w_j(v_k) \text{ never leaves } v_k) = (1 - d_k^+/g)^j,$$

with equality if and only if Γ has no directed cycles. Therefore,

$$\begin{aligned}
H(\Gamma) &\leq \frac{1}{\log 2}\left(n-1-\sum_{k=1}^{n}\sum_{j=2}^{\infty}\frac{(1-d^+(v_k)/g)^j}{j(j-1)}\right) \\
&= \frac{1}{\log 2}\left(n-1-\sum_{k=1}^{n}\left(\frac{d^+(v_k)}{g}\log\frac{d^+(v_k)}{g}+1-\frac{d^+(v_k)}{g}\right)\right) \\
&= -\sum_{k=1}^{n}\frac{d^+(v_k)}{g}\log_2\frac{d^+(v_k)}{g} \\
&= H(\mathbf{d}^+).
\end{aligned}$$

□

Note that the condition for equality is equivalent to $\bar{L}$ being permutation equivalent to an upper-triangular matrix. This makes sense, since in that case the eigenvalues of L_Γ are the out-degrees of the vertices of Γ.

Corollary 1. *For any directed graph Γ on n vertices, $H(\Gamma) < \log_2 n$.*

Proof. Since the out-degrees are real-valued, we have $H(\Gamma) \leq H(\mathbf{d}^+) \leq \log_2 n$. If $H(\mathbf{d}^+) = \log_2 n$, then $d^+(v_1) = \ldots = d^+(v_n) > 0$, and Γ must have a directed cycle, so $H(\Gamma) < H(\mathbf{d}^+)$. □

Maximizing von Neumann entropy for digraphs

Now that we've established that $\log_2 n$ is an upper bound for the von Neumann entropy of digraphs on n vertices, it's natural to ask how good this bound is. We've seen that the von Neumann entropy of a simple graph never exceeds $\log_2(n-1)$. So, are there directed graphs with entropy between $\log_2(n-1)$ and $\log_2 n$? The answer appears to be yes for $n \geq 3$.

Let's consider the directed cycle on n vertices, which we will denote DC_n. With the vertex set of DC_n labelled $\{v_1, \ldots, v_n\}$, arcs in DC_n are given by

$$v_i \to v_j \qquad \text{iff} \qquad j - i \equiv 1 \pmod{n}.$$

The Laplacian matrix of DC_n is given by

$$L = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & -1 & & 0 \\ & \vdots & & \ddots & \vdots \\ -1 & 0 & 0 & \cdots & 1 \end{bmatrix}.$$

Since L is a circulant matrix, its eigenvalues are known to be $1 - e^{i2\pi j/n}$, for $j = 0, \ldots, n-1$. Thus we can say

$$\bar{\Lambda} = \left\{\frac{1}{n}\left(1 - e^{i2\pi j/n}\right)\right\}_{j=0}^{n-1}.$$

It's not clear that this suggests high entropy, but we can employ the new perspective to make some sense of the von Neumann entropy of DC_n. For each vertex v_k, we know that

$$P(w_j(v_k) \text{ ends at } v_k) = P(w_j(v_k) \text{ never leaves } v_k)$$

for all $j < n$ because DC_n has no directed cycles of length less than n. Thus we have

$$H(DC_n) = \log_2 n - \frac{1}{\log 2} \sum_{j=n}^{\infty} \frac{\operatorname{tr}(M^j) - n\left(1 - \frac{1}{n}\right)^j}{j(j-1)},$$

and we can see that $H(DC_n)$ isn't much less than $H(\mathbf{d}^+) = \log_2 n$. In fact, here are some values of $H(DC_n)$:

n	$H(DC_n)$
3	$\log_2 2.34$
4	$\log_2 3.52$
5	$\log_2 4.63$
6	$\log_2 5.69$
7	$\log_2 6.74$
8	$\log_2 7.78$
9	$\log_2 8.80$
10	$\log_2 9.82$
20	$\log_2 19.91$
50	$\log_2 49.97$
100	$\log_2 99.98$
200	$\log_2 199.99$

We now provide evidence for the following conjecture.

Conjecture 1. *DC_n has the highest von Neumann entropy of any directed graph on n vertices.*

First, because of Theorem 2, it makes sense that von Neumann entropy is maximized by a graph with $H(\mathbf{d}^+) = \log_2 n$. Such graphs are *regular*, with $d^+(v_k)$ constant for all v_k. I randomly generated 10,000 regular digraphs for each possible value of $d = d^+(v_k)$ and each number of vertices up to 8. To randomly generate the digraph, I selected, for each vertex v_k, a random subset of size d of the remaining vertices for v_k to beat, with all choices independent. Here are the maximum values of $2^{H(\Gamma)}$ for each:

n	DC_n	$d=1$	$d=2$	$d=3$	$d=4$	$d=5$	$d=6$	$d=7$
3	2.343	2.343	2.000					
4	3.522	3.522	3.101	3.000				
5	4.627	4.627	4.398	4.048	4.000			
6	5.695	5.695	5.484	5.222	5.032	5.000		
7	6.742	6.742	6.514	6.353	6.117	6.021	6.000	
8	7.776	7.776	7.608	7.359	7.207	7.078	7.014	7.000

Conclusion

Now that we have a probabilistic and combinatorial interpretation of the von Neumann entropy, we can see that it gives a good measure of how unpredictable a system is in the sense of random movement in the direction of arcs. Furthermore, we can see at least a weak relation between high entropy and regularity in a graph or directed graph since $H(\Gamma) \leq H(\mathbf{d}^+)$.

There are many situations where a high-entropy structure is desirable. For instance, in experimental design, the subjects of an experiment should be placed on as equal footing as possible, though an asymetric relation between subjects may make this difficult.

Consider professional sports scheduling as an example, though I hope it's clear that the structure itself is not dependent on context. The National Basketball Association has 30 teams, and each team plays 82 games per season: 41 at home (in their own arena) and 41 away (at the other team's arena). It is generally accepted that the location of the game gives an advantage to the home team, known as "home court advantage", so each individual game presents a strong bias. To minimize bias over the course of the season, a high-entropy schedule is desired. Ignoring when the games take place, we can represent the schedule as a directed multigraph with 30 vertices (one for each team), where for each game, there is an arc going from the home team to the away team. Then each vertex has in-degree 41 and out-degree 41. Additional restrictions may be imposed as desired. If entropy really measures unpredictability for these graphs in the sense we intend, the graphs with the highest entropy will provide the fairest possible schedules.

There is more work to be done to justify the use of von Neumann entropy for such purposes, but for now, we've at least given some meaning to entropy in directed graphs.

References

[1] Kartik Anand and Ginestra Bianconi. Entropy measures for networks: Toward an information theory of complex topologies. *Physical Review E*, 80(4):045102, 2009.

[2] Kartik Anand, Ginestra Bianconi, and Simone Severini. Shannon and von Neumann entropy of random networks with heterogeneous expected degree. *Physical Review E*, 83(3):036109, 2011.

[3] Abdelilah Belhaj, Adil Belhaj, Larbi Machkouri, Moulay Brahim Sedra, and Soumia Ziti. Weighted graph theory representation of quantum information inspired by Lie algebras. *arXiv preprint arXiv:1609.03534*, 2016.

[4] S. L. Braunstein, S. Ghosh, and S. Severini. The Laplacian of a graph as a density matrix: a basic combinatorial approach to separability of mixed states. *Annals of Combinatorics*, 10(3):291–317, 2006.

[5] M. Dairyko, L. Hogben, J. C-H Lin, J. Lockhart, D. Roberson, S. Severini, and M. Young. Note on von Neumann and Rényi entropies of a graph. *Linear Algebra and its Applications*, 521:240–253, 2017.

[6] Niel de Beaudrap, Vittorio Giovannetti, Simone Severini, and Richard Wilson. Interpreting the von Neumann entropy of graph Laplacians, and co-entropic graphs. *A Panorama of Mathematics: Pure and Applied*, 658:227, 2016.

[7] Matthias Dehmer and Abbe Mowshowitz. A history of graph entropy measures. *Information Sciences*, 181(1):57–78, 2011.

[8] HG Landau. On dominance relations and the structure of animal societies: Iii the condition for a score structure. *Bulletin of Mathematical Biology*, 15(2):143–148, 1953.

[9] Hyman G Landau. On dominance relations and the structure of animal societies: I. effect of inherent characteristics. *The bulletin of mathematical biophysics*, 13(1):1–19, 1951.

[10] Nati Linial and Avraham Morgenstern. On the number of 4-cycles in a tournament. *Journal of Graph Theory*, 83(3):266–276, 2016.

[11] M. A. Nielsen and I. L. Chuang. *Quantum computation and quantum information*. Cambridge university press, 2010.

[12] F Passerini and S Severini. Quantifying complexity in networks: the von Neumann entropy. *Int. J. Agent Technol. Syst*, 1:58, 2009.

[13] Alfréd Rényi. On measures of entropy and information. Technical report, HUNGARIAN ACADEMY OF SCIENCES Budapest Hungary, 1961.

[14] John Riordan. The number of score sequences in tournaments. *Journal of Combinatorial Theory*, 5(1):87–89, 1968.

[15] N.J.A Sloane. The On-Line Encyclopedia of Integer Sequences, 2018.

[16] NJA Sloane and Simon Plouffe. The encyclopedia of integer sequences academic press. *San Diego, CA*, 1995.

[17] John von Neumann. *Mathematical Foundations of Quantum Mechanics: New Edition*. Princeton university press, 2018.

[18] Cheng Ye, Richard C Wilson, and Edwin R Hancock. A Jensen-Shannon divergence kernel for directed graphs. In *Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)*, pages 196–206. Springer, 2016.

Reflection

I never applied to the Honors Program, but I ended up here anyway. Before my freshman year, I heard about the Undergraduate Research Fellowship program, and I understood that they would help me to work closely with the faculty, so I applied to that. After I was accepted, I was informed that the fellowship automatically inducted me in the Honors program, which I wasn't excited about. I didn't see a way out, so I just rolled with it for as long as I could. Honors wasn't *my* goal, and that made it really hard for me to want to do a capstone project.

Nevertheless, I enrolled in the HONR 3900 capstone preparation course to keep my options open, and I submitted a proposal at the end of the semester related to some open questions I wanted to explore related to previous research I had done. My tentative plan was to keep researching with my mentor Dr. Brown in the same way that I always had and to see which of my questions I could make headway toward answering. However, Dr. Brown was promoted in the math department at the same time, and there was a long period of time where I didn't research with him much at all, as he was busy with administrative business. I tried to keep looking into the entropy project on my own, but the results I came up with didn't seem to relate much to my overall goal. I didn't have a good direction, and progress was minimal.

Eventually, Dr. Brown had some time to do research again. I told him I just wanted to get back on the horse, and I would be okay meeting with a larger group. However, my time was soon running out to finish my project, and our group meetings started to move away from my capstone topic.

At this point, I was becoming increasingly undecided if I wanted to finish my capstone project at all. Honestly, Honors was never the goal I was working toward. I just wanted to learn as much math as I could, and Honors credit had always come along the way without me having to do much more than fill out tedious paperwork for the work I was already doing, but I was never doing anything *for* honors credit, so what reason did I have to finish? I met with the Honors advisor to talk through my options, then I set up an appointment with Dr. Brown via email. I told him in my message that I wanted to talk, and that I was ready to give up on doing Honors. It was too late in the game, and anything I pulled together at this point would just make me a phony.

But before our meeting came around, I had a realization. I didn't get this far by accident. I could have fallen by the wayside years ago, but I didn't. People were working hard to ensure my success. Dr. Brown was one of them. God was another. Miracles had occurred all along the way, and now was my chance to show a bit of appreciation. Finally, I found my why. Really, I don't care if my transcript says "University Honors" or not. I don't care if my capstone is meaningful or not in a broader context. I'm just here to finish what *others* have helped me start.

When I got to Dr. Brown's office, he started "So I guess I need to walk you off the cliff—". "Nah," I interrupted, "I've decided that I'm going to move forward with my capstone after all. It's not going to be the best project I've ever done, but I'm going to finish." That day, we sat down and made an actual plan. We were going to meet one-on-one every Wednesday, and we had deadlines for when everything needed to be done. It still seemed like a lot to do in not much time, but at least it seemed feasible for once.

And so I started to make progress for a little while. But then Spring Break came around, and Dr. Brown was away at a conference the week following. And to top it off, that was the week that the university announced that classes would be moving online due to the COVID-19 pandemic. Face-to-face meetings would stop, and everybody with teaching responsibilities was scrambling to figure out how to run their class moving forward. I was a recitation leader for a linear algebra class, and just figuring out where I could record myself every day was a huge challenge. My classroom was locked, the library study rooms were available, then not, and I didn't have the technology at home to do things on my own. My capstone project suddenly dropped on the priority list, and even worse, I didn't get back with Dr. Brown to set up a new regular meeting time.

More time passed with little progress, and as my capstone project started to creep back into my mind, I opened my work plan to see how my progress was coming along. Again, I needed to readjust my expectations for myself and make a new plan. I started by writing down everything I knew about my topic and fleshing out as many details as I could. Video calls with Dr. Brown helped me to steer the paper and stay motivated to keep working on it. Slowly, the paper started to gain more focus. Ironically, at the same time, my own focus was drawn to exploring the research questions I had never been able to fully answer. This really slowed things down, but in the end, my project was much more complete because of it.

At one point, I got an email from Honors with submission criteria for formatting and what I needed to include. I was too overwhelmed to worry about that right then, so I pushed it off. When I looked at it again the last week of school, it seemed like way too much for me to handle, and I had a little breakdown. Somehow, I finished anyway, and things seemed to turn out alright.

Word count: 1023

Bryce Frederickson is a graduating senior in Utah State University's class of 2020 with a major in mathematics and a minor in computer science. He has been actively involved in research during his undergraduate career as an Undergraduate Research Fellow and honors student. He has presented at regional and national mathematics conferences, including the 2019 Joint Mathematics Meetings. He was a 2019 Goldwater Scholar, and he was named the Peak Prize Undergraduate Researcher of the Year at USU's 2019 Robins awards. In 2020, he was given a Graduate Research Fellowship by the National Science Foundation, which he will use to continue research in the mathematics PhD program at Emory University in the fall.