Entropy and Graphs

by

Seyed Saeed Changiz Rezaei

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Mathematics
in
Combinatorics and Optimization

Waterloo, Ontario, Canada, 2014

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

The entropy of a graph is a functional depending both on the graph itself and on a probability distribution on its vertex set. This graph functional originated from the problem of source coding in information theory and was introduced by J. Körner in 1973. Although the notion of graph entropy has its roots in information theory, it was proved to be closely related to some classical and frequently studied graph theoretic concepts. For example, it provides an equivalent definition for a graph to be perfect and it can also be applied to obtain lower bounds in graph covering problems.

In this thesis, we review and investigate three equivalent definitions of graph entropy and its basic properties. Minimum entropy colouring of a graph was proposed by N. Alon in 1996. We study minimum entropy colouring and its relation to graph entropy. We also discuss the relationship between the entropy and the fractional chromatic number of a graph which was already established in the literature.

A graph G is called *symmetric with respect to a functional* $F_G(P)$ defined on the set of all the probability distributions on its vertex set if the distribution P^* maximizing $F_G(P)$ is uniform on $V(G)$. Using the combinatorial definition of the entropy of a graph in terms of its vertex packing polytope and the relationship between the graph entropy and fractional chromatic number, we prove that vertex transitive graphs are symmetric with respect to graph entropy. Furthermore, we show that a bipartite graph is symmetric with respect to graph entropy if and only if it has a perfect matching. As a generalization of this result, we characterize some classes of symmetric perfect graphs with respect to graph entropy. Finally, we prove that the line graph of every bridgeless cubic graph is symmetric with respect to graph entropy.

Acknowledgements

I would like to thank my advisor Chris Godsil for his guidance and support throughout my graduate studies in Combinatorics and Optimization Department. I would like to thank Joseph Cheriyan and Kevin Purbhoo for their valuable comments about the thesis. I am also grateful to the Department of Combinatorics and Optimization for providing me with a motivating academic environment.

Table of Contents

List of Figures

Chapter 1

Introduction

The entropy of a graph is an information theoretic functional which is defined on a graph with a probability density on its vertex set. This functional was originally proposed by J. Körner in 1973 to study the minimum number of codewords required for representing an information source (see J. Körner [19]).

J. Körner investigated the basic properties of the graph entropy in several papers from 1973 till 1992 (see J. Körner [19]-[25]).

Let F and G be two graphs on the same vertex set V. Then the union of graphs F and G is the graph $F \cup G$ with vertex set V and its edge set is the union of the edge set of graph F and the edge set of graph G. That is

$$\begin{aligned} V(F \cup G) &= V, \\ E(F \cup G) &= E(F) \cup E(G). \end{aligned}$$

The most important property of the entropy of a graph is that it is sub-additive with respect to the union of graphs. This leads to the application of graph entropy for graph covering problem as well as the problem of perfect hashing.

The graph covering problem can be described as follows. Given a graph G and a family of graphs $\mathcal{G}$ where each graph $G_i \in \mathcal{G}$ has the same vertex set as G, we want to cover the edge set of G with the minimum number of graphs from $\mathcal{G}$. Using the sub-additivity of graph entropy one can obtain lower bounds on this number.

Graph entropy was used in a paper by Fredman and Komlós for the minimum number of perfect hash functions of a given range that hash all k-element subsets of a set of a given size (see Fredman and Komlós [14]).

As another application of graph entropy, Kahn and Kim in [18] proposed a sorting algorithm based on the entropy of an appropriate comparability graph.

In 1990, I. Csiszár, J. Körner, L. Lovász, K. Marton, and G. Simony, characterized minimal pairs of convex corners which generate the probability density $P = (p_1, \cdots, p_k)$ in a k-dimensional space. Their study led to another definition of the graph entropy in terms of the vertex packing polytope of the graph. They also gave another characterization of a perfect graph using the sub-additivity property of graph entropy.

The sub-additivity property of the graph entropy was further studied in J. Körner [20], J. Körner and G. Longo [22], J. Körner et. al. [23], and J. Körner and K. Marton [24]. Their studies led to the notion of a class of graphs which is called *normal graphs*.

Noga Alon, and Alon Orlitsky studied the problem of source coding in information theory using the minimum entropy colouring of the characteristic graph associated with a given information source. They investigated the relationship between the minimum entropy colouring of a graph and the graph entropy (see N. Alon and A. Orlitsky [1]).

This thesis is organized as follows. In Chapter 2, we define the entropy of a random variable. We also briefly investigate the application of entropy in counting problems. In chapter 3, we define the entropy of a graph. Let $VP(G)$ be the *vertex packing polytope* of a given graph G which is the convex hull of the characteristic vectors of its independent sets. Let $|V(G)| = n$ and P be a probability density on $V(G)$. Then the *entropy of* G *with respect to the probability density* P is defined as

$$H_k(G, P) = \min_{\mathbf{a} \in VP(G)} \sum_{i=1}^{n} p_i \log(1/a_i).$$

This is the definition of graph entropy which we work with throughout this thesis and was given by I. Csiszár and et. al. in [9]. It is shown by I. Csiszár et. al. in [9] that the above two definitions are equal. We also investigate the basic properties of graph entropy and explain the relationship between the the graph entropy and perfect graphs and fractional chromatic number of a graph. Chapter 4 is devoted to minimum entropy colouring of a given graph and its connection to the graph entropy. G. Simonyi in [36] showed that the maximum of the graph entropy of a given graph over the probability density of its vertex set is equal to its fractional chromatic number. We call a graph is symmetric with respect to graph entropy if the uniform density maximizes its entropy. We show that vertex transitive graphs are symmetric. In Chapter 5, we study some other classes of graphs which are symmetric with respect to graph entropy. Our main results are the following theorems.

5.1.1 Theorem. Let G be a bipartite graph with parts A and B, and no isolated vertices.

Then, uniform probability distribution U over the vertices of G maximizes $H_k(G, P)$ if and only if G has a perfect matching.

As a generalization of this result we show that

5.2.3 Theorem. Let $G = (V, E)$ be a perfect graph and P be a probability distribution on $V(G)$. Then G is symmetric with respect to graph entropy $H_k(G, P)$ if and only if G can be covered by its cliques of maximum size.

A. Schrijver [34] calls a graph G a *k-graph* if it is k-regular and its fractional edge coloring number $\chi'_f(G)$ is equal to k. We show that

5.3.3 Theorem. Let G be a k-graph with $k \geq 3$. Then the line graph of G is symmetric with respect to graph entropy.

As a corollary to this result we show that the line graph of every bridgeless cubic graph is symmetric with respect to graph entropy.

Chapter 2

Entropy and Counting

In this chapter, we explain some probabilistic preliminaries such as the notions of probability spaces, random variables and the entropy of a random variable. Furthermore, we give some applications of entropy methods in counting problems. Specifically, we elaborate the proof of the Brégman's Theorem for counting the number of perfect matchings of a bipartite graph using entropy.

2.1 Probability Spaces, Random Variables, and Density functions

Let Ω be a set of *outcomes*, let $\mathcal{F}$ be a family of subsets of Ω which is called the set of *events*, and let $P : \mathcal{F} \to [0, 1]$ be a function that assigns probabilities to events. The triple $(\Omega, \mathcal{F}, P)$ is a *probability space* and $(\Omega, \mathcal{F})$ is a *measurable space*. A *measure* is a nonnegative countably additive set function, that is a function $\mu : \mathcal{F} \to \mathbb{R}$ such that

(i). $\mu(A) \geq \mu(\emptyset) = 0$ for all $A \in \mathcal{F}$, and

(ii). if $A_i \in \mathcal{F}$ is a countable sequence of disjoint sets, then

$$\mu\left(\bigcup_i A_i\right) = \sum_i \mu\left(A_i\right).$$

If $\mu(\Omega) = 1$, we call μ a *probability measure*. Throughout this thesis, probability measures are denoted by $P(.)$. A probability space is discrete if Ω is countable. In this thesis, we only consider discrete probability spaces. Then having $p(\omega) \geq 0$ for all $\omega \in \Omega$ and $\sum_{\omega \in \Omega} p(\omega) = 1$, for all event $A \in \mathcal{F}$, the probability of the event A is denoted by $P(A)$, which is

$$P(A) = \sum_{\omega \in A} p(w)$$

Note that members of $\mathcal{F}$ are called *measurable sets* in measure theory; they are also called *events* in a probability space.

On a finite set Ω, there is a natural probability measure P, called the (discrete) *uniform* measure on 2^{Ω}, which assigns probability $\frac{1}{|\Omega|}$ to singleton $\{\omega\}$ for each ω in Ω. Coin tossing gives us examples with $|\Omega| = 2^n$, $n = 1, 2, \cdots$. Another classical example is a fair *die*, a perfect cube which is thrown at random so that each of the six faces, marked with the integers 1 to 6, has equal probability $\frac{1}{6}$ of coming up.

Probability spaces become more interesting when random variables are defined on them. Let $(S, \mathcal{S})$ be a measurable space. A function

$$X : \Omega \to S,$$

is called a *measurable map* from $(\Omega, \mathcal{F})$ to $(S, \mathcal{S})$ if

$$X^{-1}(B) = \{\omega : X(\omega) \in B\} \in \mathcal{F}.$$

If $(S, \mathcal{S}) = (\mathbb{R}, \mathcal{R})$, the real valued function X defined on Ω is a *random variable*.

For a discrete probability space Ω any function $X : \Omega \to \mathbb{R}$ is a random variable. The *indicator function* $1_A(\omega)$ of a set $A \in \mathcal{F}$ which is defined as

$$1_A(\omega) = \begin{cases} 1, & \omega \in A, \\ 0, & \omega \notin A. \end{cases} \tag{2.1}$$

is an example of a random variable. If X is a random variable, then X induces a probability measure on $\mathbb{R}$ called its *probability density function* by setting

$$\mu(A) = P(X \in A)$$

for sets A. For a comprehensive study of probability spaces see R. M. Duddley [10] and Rick Durrett [11].

In this thesis, we consider discrete random variables. Let X be a discrete random variable with range $\mathcal{X}$ and probability density function $p_X(x) = \Pr(X = x)$, for $x \in \mathcal{X}$. For the sake of convenience, we use $p(x)$ instead of $p_X(x)$. Thus, $p(x)$ and $p(y)$ refer to two different random variables and are in fact different probability density functions, $p_X(x)$ and $p_Y(y)$, respectively.

2.2 Entropy of a Random Variable

Let X be a random variable X with probability density $p(x)$. We denote the *expectation* by E. Then *expected value of the random variable* X is written

$$E\left(X\right) = \sum_{x \in \mathcal{X}} xp(x),$$

and for a function $g(.)$, the *expected value of the random variable* $g(X)$ is written

$$E_p\left(g(X)\right) = \sum_{x \in \mathcal{X}} g(x)p(x),$$

or more simply as $E\left(g(X)\right)$ when the probability density function is understood from the context.

Let X be a random variable which drawn according to probability density function $p(x)$. The *entropy* of X, $H(X)$ is defined as the expected value of the random variable $\log \frac{1}{p(x)}$, therefore, we have

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log p(x).$$

The log is to the base 2 and entropy is expressed in bits. Furthermore, by convention $0 \log 0 = 0$. Since $0 \leq p(x) \leq 1$, we have $\log \frac{1}{p(x)} \geq 0$ which implies that $H(X) \geq 0$. Let us recall our coin toss example in previous section with $n = 1$ where the coin is not necessarily fair. That is denoting the event *head* by H and the event *tail* by T, let $P(H) = p$ and $P(T) = 1 - p$. Then the corresponding random variable X is defined as $X(H) = 1$ and $X(T) = 0$. That is we have

$$X = \begin{cases} 1, & \Pr\{X = 1\} = p; \\ 0, & \Pr\{X = 0\} = 1 - p. \end{cases}$$

Then,

$$H(X) = -p \log p - (1-p)\log(1-p).$$

Note that the maximum of $H(X)$ is equal to 1 which is attained when $p = \frac{1}{2}$. Thus, the entropy of a fair coin toss, i.e., $P(H) = P(T) = \frac{1}{2}$ is 1 bit. More generally for any random variable X,

$$H(X) \leq \log |\mathcal{X}|, \tag{2.2}$$

with equality if and only if X is uniformly distributed.

The *joint entropy* $H(X,Y)$ of a pair of discrete random variables (X,Y) with a joint probability density function $p(x,y)$ is defined as

$$H\left(X,Y\right) = -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log p(x,y).$$

Note that we can also express $H(X,Y)$ as

$$H(X,Y) = -E\left(\log p(X,Y)\right).$$

We can also define the conditional entropy of a random variable given another. Let

$$H(Y|X=x) = -\sum_{y\in\mathcal{Y}} p(y|x)\log p(y|x),$$

Then *Conditional Entropy* $H(Y|X)$ is defined as

$$H(Y|X) = \sum_{x\in\mathcal{X}} p(x)H(Y|X=x). \tag{2.3}$$

Now we can again obtain another description of the conditional entropy in terms of the conditional expectation of random variable as follows.

$$\begin{aligned} H\left(Y|X\right) &= -\sum_{x\in\mathcal{X}} p(x)\sum_{y\in\mathcal{Y}} p(y|x)\log p(y|x) \\ &= -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log p(y|x) \\ &= -E\log p(Y|X). \end{aligned}$$

The following theorem is proved by T. Cover and J. Thomas in [8] pages 17 and 18.

2.2.1 Theorem. *Let X, Y, and Z be random variables with joint probability distribution $p(x, y, z)$. Then we have*

$$H(X, Y) = H(X) + H(Y|X),$$
$$H(X, Y|Z) = H(X|Z) + H(Y|X, Z).$$

□

Furthermore, letting $f(.)$ be any function (see T. Cover and J. Thomas [8] pages 34 and 35), we have

$$0 \le H(X|Y) \le H(X|f(Y)) \le H(X). \tag{2.4}$$

2.3 Relative Entropy and Mutual Information

Let X be a random variable and consider two different probability density functions $p(x)$ and $q(x)$ for X. The *relative entropy* $D(p||q)$ is a measure of the distance between two distributions $p(x)$ and $q(x)$. The *relative entropy* or *Kullback-Leibler distance* between two probability densities $p(x)$ and $q(x)$ is defined as

$$D(p||q) = \sum_{x \in \mathcal{X}} p(x) \log \frac{p(x)}{q(x)}, \tag{2.5}$$

We can see that $D(p||q) = E_p \log \frac{p(X)}{q(X)}$.

Now consider two random variables X and Y with a joint probability densities $p(x, y)$ and marginal densities $p(x)$ and $p(y)$. The *mutual information* $I(X; Y)$ is the relative entropy between the joint distribution and the product distribution $p(x)p(y)$. More precisely, we have

$$\begin{aligned} I(X; Y) \quad &= \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} \\ &= D(p(x, y)||p(x)p(y)). \end{aligned}$$

It is proved in T. Cover and J. Thomas [8], on pages 28 and 29, that we have

$$\begin{aligned} I(X;Y) &= H(X) - H(X|Y), \\ I(X;Y) &= H(Y) - H(Y|X), \\ I(X;Y) &= H(X) + H(Y) - H(X,Y), \\ I(X;Y) &= I(Y;X), \\ I(X;X) &= H(X). \end{aligned} \tag{2.6}$$

2.4 Entropy and Counting

In this section we consider the application of entropy method in counting problems. The following lemmas are two examples of using entropy in solving well-known combinatorial problems (see J. Radhakrishnan [33]).

2.4.1 Lemma. (Shearer's Lemma). *Suppose n distinct points in $\mathbb{R}^3$ have n_1 distinct projections on the XY-plane, n_2 distinct projections on the XZ-plane and n_3 distinct projections on the YZ-plane. Then, $n^2 \leq n_1 n_2 n_3$.*

Proof. Let $P = (A, B, C)$ be one of the n points picked at random with uniform distribution, and $P_1 = (A, B)$, $P_2 = (A, C)$, and $P_3 = (B, C)$ are its three projections. Then we have

$$H(P) = H(A) + H(B|A) + H(C|A, B), \tag{2.7}$$

Furthermore,

$$\begin{aligned} H(P_1) &= H(A) + H(B|A), \\ H(P_2) &= H(A) + H(C|A), \\ H(P_3) &= H(B) + H(C|B). \end{aligned}$$

Adding both sides of these equations and considering (2.4) and (2.7), we have $2H(P) \leq H(P_1) + H(P_2) + H(P_3)$. Now, noting that $H(P) = \log n$, and $H(P_i) \leq \log n_i$, the lemma is proved. □

As another application of the entropy method, we can give an upper bound on the number of the perfect matchings of a bipartite graph (see J. Radhakrishnan [33]).

2.4.2 Theorem. (Brégman's Theorem).*Let G be a bipartite graph with parts V_1 and V_2 such that $|V_1| = |V_2| = n$. Let $d(v)$ denote the degree of a vertex v in G. Then, the number of perfect matchings in G is at most*

$$\prod_{v \in V_1} (d(v)!)^{\frac{1}{d(v)}}.$$

Proof. Let $\mathcal{X}$ be the set of perfect matchings of G. Let X be a random variable corresponding to the elements of $\mathcal{X}$ with uniform density. Then

$$H(X) = \log |\mathcal{X}|.$$

The following remark is useful in our discussion. Let Y be any random variable with the set of possible values $\mathcal{Y}$. First note that the conditional entropy $H(Y|X)$ is obtained using (2.3). Let $\mathcal{Y}_x$ denote the set of possible values for the random variable Y given $x \in \mathcal{X}$, that is

$$\mathcal{Y}_x = \{y \in \mathcal{Y} : P(Y = y|X = x) > 0\}.$$

We partition the set $\mathcal{X}$ into sets $\mathcal{X}_1, \mathcal{X}_2, \cdots, \mathcal{X}_r$ such that for $i = 1, 2, \cdots, r$ and all $x \in \mathcal{X}_i$, we have

$$|\mathcal{Y}_x| = i. \tag{2.8}$$

Letting Y_x be a random variable taking its value on the set $\mathcal{Y}_x$ with uniform density, and noting equations (2.2) and (2.8) for all $x \in \mathcal{X}_i$ we have

$$H(Y_x) = \log i. \tag{2.9}$$

But note that

$$H(Y|X) \quad = \quad E_X(H(Y_x)). \tag{2.10}$$

Then using (2.9) and (2.10), we get

$$H(Y|X) \leq \sum_{i}^{r} P(X \in \mathcal{X}_i) \log i. \tag{2.11}$$

We define the random variable $X(v)$ for all $v \in V_1$ as

$$X(v) := u \text{ such that } u \in V_2 \text{ and } u \text{ is matched to } v \text{ in } X, \ \forall v \in V_1.$$

For a fixed ordering vertices $v_1, \cdots, v_n$ of V_1

$$\begin{aligned} \log|\mathcal{X}| &= H(X) \qquad (2.12) \\ &= H\left(X(v_1)\right) + H\left(X(v_2)|X(v_1)\right) + \cdots + H\left(X(v_n)|X(v_1), \cdots, X(v_{n-1})\right), \end{aligned}$$

Now, pick a random permutation

$$\tau : [n] \to V_1,$$

and consider X in the order determined by τ. Then for every permutation τ, we have

$$H(X) = H\left(X(\tau(1))\right) + H\left(X(\tau(2))|X(\tau(1))\right) + \cdots + H\left(X(\tau(n))|X(\tau(1)), \cdots, X(\tau(n-1))\right).$$

By averaging over all τ, we get

$$H(X) = E_\tau\left(H\left(X(\tau(1))\right) + H\left(X(\tau(2))|X(\tau(1))\right) + \cdots + H\left(X(\tau(n))|X(\tau(1)), \cdots, X(\tau(n-1))\right)\right).$$

For a fixed τ, fix $v \in V_1$ and let $k = \tau^{-1}(v)$. Then we let $\mathcal{Y}_{v,\tau}$ to be the set of vertices u in V_2 which are adjacent to vertex $v \in V_1$ and

$$u \notin \{x\left(\tau(1)\right), x\left(\tau(2)\right), \cdots, x\left(\tau(k-1)\right)\}$$

Letting $\mathcal{N}(v)$ be the set of neighbours of $v \in V_1$ in V_2, we have

$$\mathcal{Y}_{v,\tau} = \mathcal{N}(v) \setminus \{x\left(\tau(1)\right), x\left(\tau(2)\right), \cdots, x\left(\tau(k-1)\right)\}.$$

Letting $d(v)$ be the degree of vertex v and $Y_{v,\tau} = |\mathcal{Y}_{v,\tau}|$ be a random variable taking its value in $\{1, \cdots, d(v)\}$, that is

$$Y_{v,\tau} = j, \text{ for } j \in \{1, \cdots, d(v)\}.$$

Using (2.10) and noting that $P_{X(v),\tau}(Y_{v,\tau} = j) = \frac{1}{d(v)}$, we have

$$\begin{aligned}
H\left(X\right) &= \sum_{v \in V_1} E_\tau \left(X(v) | X(\tau(1)), X(\tau(2)), \cdots, X(\tau(k-1))\right) \\
&\leq \sum_{v \in V_1} E_\tau \left(\sum_{j=1}^{d(v)} P_{X(v)}\left(Y_{v,\tau} = j\right) . \log j \right) \\
&= \sum_{v \in V_1} \sum_{j=1}^{d(v)} E_\tau \left(P_{X(v)}\left(Y_{v,\tau} = j\right)\right) . \log j \\
&= \sum_{v \in V_1} \sum_{j=1}^{d(v)} P_{X(v),\tau}\left(Y_{v,\tau} = j\right) . \log j \\
&= \sum_{v \in V_1} \sum_{j=1}^{d(v)} \frac{1}{d(v)} \log j \\
&= \sum_{v \in V_1} \log \left(d(v)!\right)^{\frac{1}{d(v)}} .
\end{aligned}$$

Then using (2.13), we get

$$|\mathcal{X}| \leq \left(d(v)!\right)^{\frac{1}{d(v)}} .$$

$\square$

Chapter 3

Graph Entropy

In this chapter, we introduce and study the entropy of a graph which was defined in [19] by J. Körner in 1973. We present several equivalent definitions of this parameter. However, we will focus mostly on the combinatorial definition which is going to be the main theme of this thesis. We explain the proof of the theorem by G. Simonyi relating the entropy of a graph with its fractional chromatic number in more detail. Furthermore, we obtain the fractional chromatic number of a vertex transitive graph using its properties along with the Lagrange Multiplier technique in convex optimization.

3.1 Entropy of a Convex Corner

A subset $\mathcal{A}$ of $\mathbb{R}_+^n$ is called a *convex corner* if it is compact, convex, has non-empty interior, and for every $\mathbf{a} \in \mathcal{A}$, $\mathbf{a}' \in \mathbb{R}_+^n$ with $\mathbf{a}' \leq \mathbf{a}$, we have $\mathbf{a}' \in \mathcal{A}$. For example, the *vertex packing polytope* $VP(G)$ of a graph G, which is the convex hull of the characteristic vectors of its independent sets, is a convex corner.

Now, let $\mathcal{A} \subseteq \mathbb{R}_+^n$ be a convex corner, and $P = (p_1, \cdots, p_n) \in \mathbb{R}_+^n$ a probability density, i.e., its coordinates add up to 1. The entropy of P with respect to $\mathcal{A}$ is

$$H_{\mathcal{A}}(P) = \min_{a \in \mathcal{A}} \sum_{i=1}^{n} p_i \log \frac{1}{a_i}.$$

Consider the convex corner $\mathcal{S} := \{x \geq 0, \sum_i x_i \leq 1\}$, which is called a *unit corner*. The following lemma relates the entropy of a random variable defined in the previous chapter to the entropy of the unit corner.

3.1.1 Lemma. *The entropy $H_{\mathcal{S}}(P)$ of a probability density P with respect to the unit corner $\mathcal{S}$ is just the regular (Shannon) entropy $H(P) = -\sum_i p_i \log p_i$.*

Proof. From Remark **??**, we have

$$H_{\mathcal{S}}(\mathbf{p}) = \min_{\mathbf{s}\in\mathcal{S}} -\sum_i p_i \log s_i = \min_{\mathbf{s}\in\{x\geq 0,\ \sum_i x_i = 1\}} -\sum_i p_i \log s_i$$

Thus the above minimum is attained by a probability density vector $\mathbf{s}$. More precisely, we have

$$H_{\mathcal{S}}(\mathbf{p}) = D(\mathbf{p}||\mathbf{s}) + H(\mathbf{p}).$$

Noting that $D(\mathbf{p}||\mathbf{s}) \geq 0$ and $D(\mathbf{p}||\mathbf{s}) = 0$ if and only if $\mathbf{s} = \mathbf{p}$, we get

$$H_{\mathcal{S}}(\mathbf{p}) = H(\mathbf{p}).$$

□

There is another way to obtain the entropy of a convex corner. Consider the mapping $\Lambda : \text{int } \mathbb{R}^n_+ \to \mathbb{R}^n$ defined by

$$\Lambda(x) := (-\log x_1, \cdots, -\log x_n).$$

It is easy to see using the concavity of the log function that if $\mathcal{A}$ is a convex corner, then $\Lambda(\mathcal{A})$ is a closed, convex, full-dimensional set, which is *up-monotone*, i.e., $a \in \Lambda(\mathcal{A})$ and $a' \geq a$ imply $a' \in \Lambda(\mathcal{A})$. Now, $H_{\mathcal{A}}(P)$ is the minimum of the linear objective function $\sum_i p_i x_i$ over $\Lambda(\mathcal{A})$. Now we have the following lemma (See [9]).

3.1.2 Lemma. (I. Csiszár, J. Körner, L. Lovás , K. Marton, and G. Simonyi). *For two convex corners $\mathcal{A}, \mathcal{C} \subseteq \mathbb{R}^k_+$, we have $H_{\mathcal{A}}(P) \geq H_{\mathcal{C}}(P)$ for all P if and only if $\mathcal{A} \subseteq \mathcal{C}$.*

Proof. The "if" part is obvious. Assume that $H_{\mathcal{C}}(P) \leq H_{\mathcal{A}}(P)$ for all P. As remarked above, we have

$$H_{\mathcal{A}}(P) = \min\{P^T\mathbf{x} : \mathbf{x} \in \Lambda(\mathcal{A})\},$$

and hence it follows that we must have $\Lambda(\mathcal{A}) \subseteq \Lambda(\mathcal{C})$. This clearly implies $\mathcal{A} \subseteq \mathcal{C}$. □

Then we have the following corollary.

3.1.3 Corollary. *We have $0 \leq H_{\mathcal{A}}(P) \leq H(P)$ for every probability distribution P if and only if $\mathcal{A}$ contains the unit corner and is contained in the unit cube.*

3.2 Entropy of a Graph

Let G be a graph on vertex set $V(G) = \{1, \cdots, n\}$, let $P = (p_1, \cdots, p_n)$ be a probability density on $V(G)$, and let $VP(G)$ denote the vertex packing polytope of G. The *entropy of G with respect to P* is then defined as

$$H_k(G, P) = \min_{\mathbf{a} \in VP(G)} \sum_{i=1}^{n} p_i \log(1/a_i).$$

Let $G = (V, E)$ be a graph with vertex set V and edge set E. Let V^n be the set of sequences of length n from V. Then the graph $G^{(n)}$ is the *n-th conormal power* graph with vertex set V^n, and two distinct vertices x and y of $G^{(n)}$ are adjacent in $G^{(n)}$ if there is some $i \in n$ such that x_i and y_i are adjacent in G, that is

$$E^{(n)} = \{(x, y) \in V^n \times V^n : \exists i : (x_i, y_i) \in E\}.$$

For a graph F and $Z \subseteq V(F)$ we denote by $F[Z]$ the induced subgraph of F on Z. The chromatic number of F is denoted by $\chi(F)$.

Let

$$T_\epsilon^{(n)} = \{U \subseteq V^n : P^n(U) \geq 1 - \epsilon\}.$$

We define the functional $H(G, P)$ with respect to the probability distribution P on the vertex set $V(G)$ as follows.

$$H(G, P) = \lim_{n \to \infty} \min_{U \in T_\epsilon^{(n)}} \frac{1}{n} \log \chi(G^{(n)}[U]). \tag{3.1}$$

Let X and Y be two discrete random variables taking their values on some (possibly different) finite sets and consider the vector-valued random variable formed by the pair (X, Y) (see Thomas M. Cover, and Joy A. Thomas [8] page 16).

Now let X denote a random variable taking its values on the vertex set of G and Y be a random variable taking its values on the independent sets of G. Having a fixed distribution P over the vertices, the set of feasible joint distributions $\mathcal{Q}$ consists of the joint distributions Q of X and Y such that

$$\sum_{y \in \mathcal{Y}} Q(X, Y = y) = P(X).$$

As an example let the graph G be a 5-cycle C_5 with the vertex set

$$V(C_5) = \{x_1, x_2, x_3, x_4, x_5\},$$

and let $\mathcal{Y}$ denote the set of independent sets of G. Let P be the uniform distribution over the vertices of G, i.e.,

$$P(X = x_i) = \frac{1}{5},\ \forall i \in \{1, \cdots, 5\},$$

Noting that each vertex of C_5 lies in two maximal independent sets, we define the joint distribution Q as

$$Q(X = x, Y = y) = \begin{cases} \frac{1}{10}, & y \text{ maximal and } y \ni x, \\ 0, & \text{Otherwise.} \end{cases} \tag{3.2}$$

is a feasible joint distribution.

Now given a graph G, we define the functional $H'(G, P)$ with respect to the probability distribution P on the vertex set $V(G)$, as

$$H'(G, P) = \min_{Q} I(X; Y). \tag{3.3}$$

The following lemmas relate the functionals defined above.

3.2.1 Lemma. (I. Csiszár, et. al.). *For every graph G we have $H_k(G, P) = H'(G, P)$.*

Proof. First, we show that $H_k(G, P) = H'(G, P)$. Let X be a random variable taking its values on the vertices of G with probability density $P = (p_1, \cdots, p_n)$. Furthermore, let Y be the random variable associated with the independent sets of G and $\mathcal{F}(G)$ be the family of independent sets of G. Let q be the conditional distribution of Y which achieves the minimum in (3.3) and r be the corresponding distribution of Y. Then we have

$$H'(G, P) = I(X; Y) = -\sum_i p_i \sum_{i \in F \in \mathcal{F}(G)} q(F|i) \log \frac{r(F)}{q(F|i)}.$$

From the concavity of the log function we have

$$\sum_{i \in F \in \mathcal{F}(G)} q(F|i) \log \frac{r(F)}{q(F|i)} \leq \log \sum_{i \in F \in \mathcal{F}(G)} r(F).$$

Now we define the vector $\mathbf{a}$ by setting

$$a_i = \sum_{i \in F \in \mathcal{F}(G)} r(F).$$

Note that $\mathbf{a} \in VP(G)$. Hence,

$$H'(G,P) \geq -\sum_i p_i \log a_i.$$

and consequently,

$$H'(G,P) \geq H_k(G,P).$$

Now we prove the reverse inequality. Let $\mathbf{a} \in VP(G)$. Then letting s be a probability density on $\mathcal{F}(G)$, we have

$$a_i = \sum_{i \in F \in \mathcal{F}(G)} s(F).$$

We define transition probabilities as

$$q(F|i) = \begin{cases} \frac{s(F)}{a_i} & i \in F, \\ 0 & i \notin F. \end{cases} \tag{3.4}$$

Then, setting $r(F) = \sum_i p_i q(F|i)$, we get

$$H'(G,P) \leq \sum_{i,F} p_i q(F|i) \log \frac{q(F|i)}{r(F)}$$

By the concavity of the log function, we get

$$-\sum_F r(F) \log r(F) \leq -\sum_F r(F) \log s(F),$$

Thus,

$$-\sum_{i,F} p_i q(F|i) \log r(F) \leq -\sum_{i,F} p_i q(F|i) \log s(F).$$

And therefore,

$$H'(G,P) \leq \sum_{i,F} p_i q(F|i) \log \frac{q(F|i)}{s(F)} = -\sum_i p_i \log a_i.$$

□

3.2.2 Lemma. (J. Körner). *For every graph G we have $H'(G,P) = H(G,P)$.*

Proof. See Appendix A. □

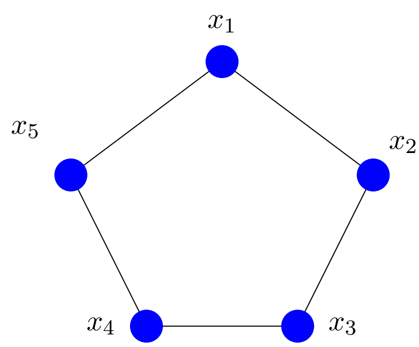


Figure 3.1: A characteristic graph of an information source with 5 alphabets

.

3.3 Graph Entropy and Information Theory

A *discrete memoryless* and *stationary information source* X is a sequence $\{X_i\}_{i=1}^{\infty}$ of independent, identically distributed discrete random variables with values in a finite set $\mathcal{X}$. Let $\mathcal{X}$ denote the set of the alphabet of a discrete memoryless and stationary information source with five elements. That is

$$\mathcal{X} = \{x_1, x_2, x_3, x_4, x_5\}.$$

We define a characteristic graph G corresponding to $\mathcal{X}$ as follows. The vertex set of G is

$$V(G) = \mathcal{X}.$$

Furthermore, two vertices of G are adjacent if and only if the corresponding elements of $\mathcal{X}$ are distinguishable. As an example one can think of the 5-cycle of Figure 3.1 as a characteristic graph of an information source $\mathcal{X}$. In the source coding problem, our goal is to label the vertices of the characteristic graph with minimum number of labels so that we can recover the elemnets of a given alphabet in a unique way. This means that we should colour the vertices of the graph properly with minimum number of colours. More precisely, one way of encoding the elements of the source alphabet $\mathcal{X}$ in Figure 3.1 is

$$\begin{aligned} \{x_1, x_3\} &\rightarrow \text{red}, \\ \{x_2, x_4\} &\rightarrow \text{blue}, \\ \{x_5\} &\rightarrow \text{green}. \end{aligned} \tag{3.5}$$

Now, let X be a random variable takes its values from $\mathcal{X}$ with the following probability density

$$P(X = x_i) = p_i,\ \forall i \in \{1, \cdots, 5\}.$$

Now consider the graph $G^{(n)}$, and let $\epsilon > 0$. Then neglecting vertices of $G^{(n)}$ having a total probability less than ϵ, the encoding of vertices of $G^{(n)}$ essentially becomes the colouring of a sufficiently large subgraph of $G^{(n)}$. And therefore, the minimum number of codewords is

$$\min_{U \in T_\epsilon^{(n)}} \chi(G^{(n)}(U)).$$

Taking logarithm of the above quantity, normalizing it by n, and making n very large, we get the minimum number of required information bits which is the same as the graph entropy of G. The characteristic graph of a regular source where distinct elements of the source alphabet are distinguishable is a complete graph. We will see in section 3.5 that the entropy of a complete graph is the same as the entropy of a random variable.

3.4 Basic Properties of Graph Entropy

The main properties of graph entropy are *monotonicity*, *sub-additivity*, and *additivity under vertex substitution*. Monotonicity is formulated in the following lemma.

3.4.1 Lemma. (J. Körner). *Let F be a spanning subgraph of a graph G. Then for any probability density P we have $H_k(F, P) \leq H_k(G, P)$.*

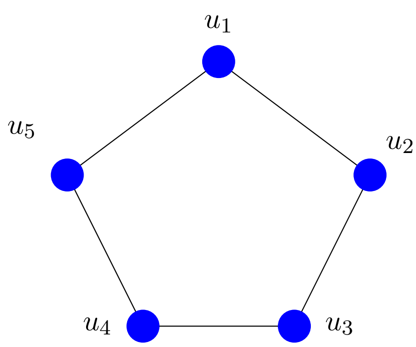
Proof. For graphs F and G mentioned above, we have $VP(G) \subseteq VP(F)$. This immediately implies the statement by the definition of graph entropy. □

The sub-additivity was first recognized by Körner in [21] and he proved the following lemma.

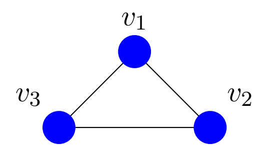
3.4.2 Lemma. (J. Körner). *Let F and G be two graphs on the same vertex set V and $F \cup G$ denote the graph on V with edge set $E(F) \cup E(G)$. For any fixed probability density P we have*

$$H_k\left(F \cup G, P\right) \leq H_k\left(F, P\right) + H_k\left(G, P\right).$$

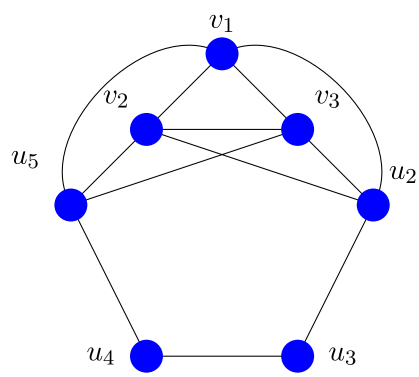
Proof. Let $\mathbf{a} \in VP(F)$ and $\mathbf{b} \in VP(G)$ be the vectors achieving the minima in the definition of graph entropy for $H_k(F, P)$ and $H_k(G, P)$, respectively. Notice the vector $\mathbf{a} \circ \mathbf{b} = (a_1 b_1, a_2 b_2, \cdots, a_n b_n)$ is in $VP(F \cup G)$, simply because the intersection of an



(a) A 5-cycle G.



(b) A triangle F.



(c) The graph $G_{u_1 \longleftarrow F}$

Figure 3.2

independent set of F with an independent set of G is always an independent set in $F \cup G$. Hence, we have

$$\begin{aligned} H_k(F,P) + H_k(G,P) &= \sum_{i=1}^{n} p_i \log \frac{1}{a_i} + \sum_{i=1}^{n} p_i \log \frac{1}{b_i} \\ &= \sum_{i=1}^{n} p_i \log \frac{1}{a_i b_i} \\ &\geq H_k(F \cup G, P). \end{aligned}$$

□

The notion of substitution is defined as follows. Let F and G be two vertex disjoint graphs and v be a vertex of G. By substituting F for v we mean deleting v and joining every vertex of F to those vertices of G which have been adjacent with v. We will denote the resulting graph $G_{v \leftarrow F}$. We extend this concept also to distributions. If we are given a probability distribution P on $V(G)$ and a probability distribution Q on $V(F)$ then by $P_{v \leftarrow Q}$ we denote the distribution on $V\left(G_{v \leftarrow F}\right)$ given by $P_{v \leftarrow Q}(x) = P(x)$ if $x \in V(G) \setminus v$ and $P_{v \leftarrow Q}(x) = P(x)Q(x)$ if $x \in V(F)$. This operation is illustrated in Figure 3.2.

Now we state the following lemma whose proof can be found in J. Körner, et. al. [23].

3.4.3 Lemma. (J. Körner, G. Simonyi, and Zs. Tuza). *Let F and G be two vertex disjoint graphs, v a vertex of G, while P and Q are probability distributions on $V(G)$ and $V(F)$, respectively. Then we have*

$$H_k\left(G_{v \leftarrow F}, P_{v \leftarrow Q}\right) = H_k\left(G, P\right) + P(v) H_k\left(F, Q\right).$$

□

Notice that the entropy of an empty graph (a graph with no edges) is always zero (regardless of the distribution on its vertices). Noting this fact, we have the following corollary as a consequence of Lemma 3.4.3.

3.4.4 Corollary. *Let the connected components of the graph G be the subgraphs G_i's and P be a probability distribution on $V(G)$. Set*

$$P_i(x) = P(x)\left(P(V(G_i))\right)^{-1}, x \in V(G_i).$$

Then

$$H_k\left(G, P\right) = \sum_i P\left(V(G_i)\right) H_k\left(G_i, P_i\right).$$

Proof. Consider the empty graph on as many vertices as the number of connected components of G. Let a distribution be given on its vertices by $P\left(V(G_i)\right)$ being the probability of the vertex corresponding to the ith component of G. Now substituting each vertex by the component it belongs to and applying Lemma 3.4.3 the statement follows. □

3.5 Entropy of Some Special Graphs

Now we look at entropy of some graphs which are also mentioned in G. Simonyi [35] and [36] . The first one is the complete graph.

3.5.1 Lemma. *For K_n, the complete graph on n vertices, one has*

$$H_k(K_n, P) = H(P).$$

Proof. By definition of entropy of a graph, $H_k(K_n, P)$ has the form $\sum_{i=1}^{n} p_i \log \frac{1}{q_i}$ where $q_i \geq 0$ for all i and $\sum_{i=1}^{n} q_i = 1$. This expression is well known to take its minimum at $q_i = p_i$. Indeed, by the concavity of the log function $\sum_{i=1}^{n} p_i \log \frac{p_i}{q_i} \leq \log \sum_{i=1}^{n} q_i = 0$. □

And the next one is the complete multipartite graph.

3.5.2 Lemma. *Let $G = K_{m_1, m_2, \cdots, m_k}$, i.e., a complete k-partite graph with maximal independent sets of size $m_1, m_2, \cdots, m_k$. Given a distribution P on $V(G)$ let Q be the distribution on $S(G)$, the set of maximal independent sets of G, given by $Q(J) = \sum_{x \in J} P(x)$ for each $J \in S(G)$. Then $H_k(G, P) = H_k(K_k, Q)$.*

Proof. The statement follows from Lemma 3.4.3 and substituting independent sets of size $m_1, m_2, \cdots, m_k$ for the vertices of K_k. □

A special case of the above Lemma is the entropy of a complete bipartite graph with uniform probability distribution over its vertex set which is equal to 1. Now, let G be a bipartite graph with color classes A and B. For a set $D \subseteq A$, let $\mathcal{N}(D)$ denotes the the set of neighbours of D in B, that is a subset of the vertices in B which are adjacent to a vertex in A.

Given a distribution P on $V(G)$ we have

$$P(D) = \sum_{i \in D} p_i \quad \forall D \subseteq V(G),$$

Furthermore, defining the binary entropy as

$$h(x) := -x \log x - (1-x) \log(1-x), \quad 0 \leq x \leq 1,$$

J. Körner and K. Marton proved the following theorem in [24].

3.5.3 Theorem. (J. Körner and K. Marton). *Let G be a bipartite graph with no isolated vertices and P be a probability distribution on its vertex set. If*

$$\frac{P(D)}{P(A)} \leq \frac{P(\mathcal{N}(D))}{P(B)},$$

for all subsets D of A, then

$$H_k(G, P) = h(P(A)).$$

And if

$$\frac{P(D)}{P(A)} > \frac{P(\mathcal{N}(D))}{P(B)},$$

then there exists a partition of $A = D_1 \cup \cdots \cup D_k$ *and a partition of* $B = U_1 \cup \cdots \cup U_k$ *such that*

$$H_k(G, P) = \sum_{i=1}^{k} P(D_i \cup U_i)\, h\left(\frac{P(D_i)}{P(D_i \cup U_i)}\right).$$

Proof. Let us assume the condition in the theorem statement holds. Then, using max-flow min-cut theorem (see A. Schrijver [34] page 150), we show that there exists a probability density Q on the edges of G such that for all vertices $v \in A$, we have

$$\sum_{v \in e \in E(G)} Q(e) = \frac{p(v)}{P(A)}, \tag{3.6}$$

We define a digraph D' by

$$V(D') = V(G) \cup \{s, t\},$$

and joining vertices s and t to all vertices in parts A and B, respectively. The edges between A and B are the exactly the same edges in G. Furthermore, we orient edges from s toward A and from A toward B and from B to t. We define a capacity function $c : E(D') \to \mathbb{R}_+$ as

$$c(e) = \begin{cases} \frac{p(v)}{P(A)}, & e = (s, v),\ v \in A, \\ 1, & e = (v, u),\ v \in A \text{ and } u \in B, \\ \frac{p(u)}{P(B)}, & e = (u, t),\ u \in B. \end{cases} \tag{3.7}$$

By the definition of c, we note that the maximum st-flow is at most 1. Now, by showing that the minimum capacity of an st-cut is at least 1, we are done.

Let $\delta(U)$ be a st-cut for some subset $U = \{s\} \cup A' \cup B'$ of $V(D')$ with $A' \subseteq A$ and $B' \subseteq B$. If

$$\mathcal{N}(A') \nsubseteq B',$$

then

$$c(\delta(U)) \geq 1.$$

So suppose that

$$\mathcal{N}(A') \subseteq B'.$$

Then using the assumption

$$\frac{P(A')}{P(A)} \leq \frac{P\mathcal{N}(A')}{P(A)},$$

we get

$$\begin{aligned} c\left(\delta(U)\right) &\geq \frac{P(B')}{P(B)} + \frac{P(A \setminus A')}{P(A)} \\ &\geq \frac{P(A')}{P(A)} + \frac{P(A \setminus A')}{P(A)} = 1. \end{aligned} \tag{3.8}$$

Now, we define the vector $\mathbf{b} \in \mathbb{R}_+^{|V(G)|}$, as follows,

$$(\mathbf{b})_v := \frac{p(v)}{P(A)}.$$

Then using (3.6), we have

$$\mathbf{b} \in VP\left(\overline{G}\right),$$

Thus,

$$H_k(\overline{G}, P) \leq \sum_{v \in V(G)} p(v) \log \frac{1}{b_v} = H(P) - h\left(P(A)\right).$$

Then, using Lemma 3.4.1 and Lemma 3.5.2, we have

$$H_k(G, P) \leq h(P(A)),$$

Now, adding the last two inequalities we get

$$H_k\left(G, P\right) + H_k\left(\overline{G}, P\right) \leq H(P). \tag{3.9}$$

On the other hand, by Lemma 3.4.2, we also have

$$H(P) \leq H_k\left(G, P\right) + H_k\left(\overline{G}, P\right), \tag{3.10}$$

Comparing (3.9) and (3.10), we get

$$H(P) = H_k\left(G, P\right) + H_k\left(\overline{G}, P\right),$$

which implies that

$$H_k(G, P) = h(P(A)).$$

This proves the first part of the theorem.

Now, suppose that the condition does not hold. Let D_1 be a subset of A such that

$$\frac{P(D_1)}{P(A)} . \frac{P(B)}{P(\mathcal{N}(D_1))}$$

is maximal. Now consider the subgraph $(A \setminus D_1) \cup (B \setminus \mathcal{N}(D_1))$ and for $i = 2, \cdots, k$ let

$$D_i \subseteq A \setminus \bigcup_{j=1}^{i-1} D_j,$$

such that

$$\frac{P(D_i)}{P(A \setminus \bigcup_{j=1}^{i-1} D_j)} . \frac{P(B \setminus \bigcup_{j=1}^{i-1} \mathcal{N}(D_j))}{P(\mathcal{N}(D_i))},$$

is maximal. Let us

$$U_i = \mathcal{N}(D_i) \setminus \mathcal{N}(D_i \cup \cdots \cup D_{i-1}), \quad \text{for } i = 1, \cdots, k.$$

Consider the independent sets $J_0, \cdots, J_k$ of the following form

$$J_0 = B,\ J_1 = D_1 \cup B \setminus U_1, \cdots,\ J_i = D_1 \cup \cdots \cup D_i \cup B \setminus U_1 \setminus \cdots \setminus U_i, \cdots,\ J_k = A.$$

Set

$$\begin{aligned}
\alpha(J_0) &= \frac{P(U_1)}{P(U_1 \cup D_1)}, \\
\alpha(J_i) &= \frac{P(U_{i+1})}{P(U_{i+1} \cup D_{i+1})} - \frac{P(U_i)}{P(U_i \cup D_i)}, \quad \text{for } i = 1, \cdots, k-1, \\
\alpha(J_k) &= 1 - \frac{P(U_k)}{P(U_k \cup D_k)}.
\end{aligned}$$

Note that by the choice of D_i's, all $\alpha(J_i)$'s are non-negative and add up to one. This implies that the vector $\mathbf{a} \in \mathbb{R}_+^{|V(G)|}$ defined as

$$a_j = \sum_{j \in J_r} \alpha(J_r), \quad \forall j \in V(G),$$

is in $VP(G)$. Furthermore,

$$a_j = \begin{cases} \frac{P(D_i)}{P(D_i \cup U_i)}, & j \in D_i, \\ \frac{P(U_i)}{P(D_i \cup U_i)}, & j \in U_i. \end{cases}$$

By the choice of the D_j's and using the same max-flow min-cut argument we had, there exists a probability density Q_i on edges of $G[D_i \cup U_i]$ such that

$$\begin{aligned}
b'_j &= \sum_{j \in e \in E(G[D_i \cup U_i])} Q_i(e) = \frac{p_j}{P(D_i)}, \quad \forall j \in D_i, \\
b'_j &= \sum_{j \in e \in E(G[D_i \cup U_i])} Q_i(e) = \frac{p_j}{P(U_i)}, \quad \forall j \in U_i.
\end{aligned}$$

Now we define the probability density Q on the edges of G as follows

$$Q(e) = \begin{cases} P(D_i \cup U_i)Q_i(e), & e \in E\left(G[D_i \cup U_i]\right), \\ 0, & e \notin E\left(G[D_i \cup U_i]\right). \end{cases}$$

The corresponding vector $\mathbf{b} \in VP\left(\overline{G}\right)$ is given by

$$b_j = P\left(D_i \cup U_i\right) b'_j, \quad \text{for } j \in D_i \cup U_i.$$

The vectors $\mathbf{a} \in VP\left(G\right)$ and $\mathbf{b} \in VP\left(\overline{G}\right)$ are the minimizer vectors in the definition of $H_k\left(G, P\right)$ and $H_k\left(\overline{G}, P\right)$, respectively. Suppose that is not true. Then noting that the fact that by the definition of $\mathbf{a}$ and $\mathbf{b}$, we have

$$\sum_{j \in V(G)} p_j \log \frac{1}{a_j} + \sum_{j \in V(G)} p_j \log \frac{1}{b_j} = \sum_{j \in V(G)} p_j \log \frac{1}{p_j} = H(P).$$

the sub-additivity of graph entropy is violated. Now, it can be verified that $H_k\left(G, P\right)$ is equal to what stated in the theorem statement. $\square$

3.6 Graph Entropy and Fractional Chromatic Number

In this section we investigate the relation between the entropy of a graph and its fractional chromatic number which was already established by G. Simonyi [36]. First we recall that the *fractional chromatic number* of a graph G is denoted by $\chi_f\left(G\right)$ is the minimum sum of nonnegative weights on the independent sets of G such that for any vertex the sum of the weights on the independent sets of G containing that vertex is at least one (see C. Godsil and G. Royle [17]). I.Csiszár and et. al. [9] showed that for every probability density P, the entropy of a graph G is attained by a point $\mathbf{a} \in VP(G)$ such that there is not any other point $\mathbf{a}' \in VP(G)$ majorizing the point $\mathbf{a}$ coordinate-wise. Furthermore, for any such point $\mathbf{a} \in VP(G)$ there is some probability density P on $VP(G)$ such that the value of $H_k\left(G, P\right)$ is attained by $\mathbf{a}$. Using this fact G. Simonyi [36] proved the following lemma.

3.6.1 Lemma. (G. Simonyi). *For a graph G and probability density P on its vertices with fractional chromatic number $\chi_f(G)$, we have*

$$\max_P H_k(G, P) = \log \chi_f(G).$$

Proof. Note that for every graph G we have $\left(\frac{1}{\chi_f(G)}, \cdots, \frac{1}{\chi_f(G)}\right) \in VP(G)$. Thus for every probability density P, we have

$$H(G, P) \leq \log \chi_f(G).$$

Now, we show that graph G has an induced subgraph G' with $\chi_f(G') = \chi_f(G) = \chi_f$ such that if $\mathbf{y} \in VP(G')$ and $\mathbf{y} \geq \frac{\mathbf{1}}{\chi_f}$, then $\mathbf{y} = \frac{\mathbf{1}}{\chi_f}$.

Suppose the above statement does not hold for graph G. Consider all $\mathbf{y} \in VP(G)$ such that

$$\mathbf{y} \geq \frac{\mathbf{1}}{\chi_f(G)}.$$

Note that there is not any $\mathbf{y} \in VP(G)$ such that

$$\mathbf{y} > \frac{\mathbf{1}}{\chi_f(G)},$$

because then we have a fractional colouring with value strictly less than $\chi_f(G)$. Thus for every $\mathbf{y} \geq \frac{\mathbf{1}}{\chi_f(G)}$ there is some $v \in V(G)$ such that $y_v = \frac{1}{\chi_f(G)}$. For such a fixed $\mathbf{y}$, let

$$\Omega_{\mathbf{y}} = \left\{ v \in V(G) : y_v > \frac{1}{\chi_f(G)} \right\}.$$

Let $\mathbf{y}^*$ be one of those $\mathbf{y}$'s with $|\Omega_{\mathbf{y}}|$ of maximum size. Let

$$G' = G\left[V(G) \setminus \Omega_{\mathbf{y}^*}\right].$$

From our the definition of G' and fractional chromatic number, we have either

$$\chi_f(G') < \chi_f(G).$$

or

$$\exists \mathbf{y} \in VP(G'), \text{such that } \mathbf{y} \geq \frac{\mathbf{1}}{\chi_f} \text{ and } \mathbf{y} \neq \frac{\mathbf{1}}{\chi_f}.$$

Suppose

$$\chi_f(G') < \chi_f(G).$$

Therefore

$$\mathbf{z} = \frac{\mathbf{1}}{\chi_f(G')} \in VP(G')$$

and consequently

$$\mathbf{z} > \frac{\mathbf{1}}{\chi_f(G)}.$$

Without loss of generality assume that

$$V(G) \setminus V(G') = \{1, \cdots, |V(G) \setminus V(G')|\}.$$

Set

$$\epsilon := \frac{1}{2}\left(\min_{v \in \Omega_{\mathbf{y}^*}} y_v - \frac{1}{\chi_f(G)}\right) > 0,$$
$$\mathbf{z}^* = \left(\mathbf{0}^T_{|V(G)\setminus V(G')|}, \mathbf{z}^T\right)^T \in VP(G).$$

Then

$$(1-\epsilon)\mathbf{y}^* + \epsilon \mathbf{z}^* \in VP(G),$$

which contradicts the maximality assumption of $\Omega_{\mathbf{y}^*}$. Thus, we have

$$\chi_f(G') = \chi_f(G).$$

Now we prove that if $y \in VP(G')$ and $\mathbf{y} \geq \frac{\mathbf{1}}{\chi_f}$, then $\mathbf{y} = \frac{\mathbf{1}}{\chi_f}$.

Suppose $\mathbf{z}'$ be a point in $VP(G')$ such that $\mathbf{z}' \geq \frac{1}{\chi_f}$ but $\mathbf{z}' \neq \frac{1}{\chi_f}$. Set

$$\mathbf{y}' = \left(\mathbf{0}^T_{|V(G)\setminus V(G')|}, \mathbf{z}'^T\right)^T \in VP(G).$$

Then using the $\epsilon > 0$ defined above, we have

$$(1-\epsilon)\mathbf{y}^* + \epsilon \mathbf{y}' \in VP(G),$$

which contradicts the maximality assumption of $\Omega_{\mathbf{y}^*}$.

Now, by I.Csiszár et. al. [9], there exists a probability density P' on $VP(G')$ such that $H(G', P') = \log \chi_f$. Extending P' to a probability distribution P as

$$p_i = \begin{cases} p'_i, & i \in V(G), \\ 0, & i \in V(G) \setminus V(G'). \end{cases} \tag{3.11}$$

the lemma is proved. Indeed, suppose that $H(G,P) < H(G', P')$ and let $\bar{\mathbf{y}} \in VP(G)$ be a point in $VP(G)$ which gives $H(G,P)$. Let $\bar{\mathbf{y}}_{VP(G')}$ be the restriction of $\bar{\mathbf{y}}$ in $VP(G')$. Then there exists $\mathbf{z} \in VP(G')$ such that

$$\mathbf{z} \geq \bar{\mathbf{y}}_{VP(G')}.$$

This contradicts the fact that

$$H(G', P') = \log \chi_f.$$

$\square$

3.6.1 Remark. Note that the maximizer probability distribution of the graph entropy is not unique. Consider C_4 with vertex set $V(C_4) = \{v_1, v_2, v_3, v_4\}$ with parts $A = \{v_1, v_3\}$ and $B = \{v_2, v_4\}$. Using Theorem 3.5.3, probability distributions $P_1 = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$ and $P_2 = (\frac{1}{8}, \frac{1}{4}, \frac{3}{8}, \frac{1}{4})$ give the maximum graph entropy which is 1.

Using the above lemma we compute the fractional chromatic number of a vertex transitive graph in the following corollary.

3.6.2 Corollary. *Let G be a vertex transitive graph with $|V(G)| = n$, and let $\alpha(G)$ denote the size of a coclique of G with maximum size. Then*

$$\chi_f(G) = \frac{n}{\alpha(G)}.$$

Proof. First note that since G is a vertex transitive graph, there exists a family of cocliques $S_1, \cdots, S_b$ of size $\alpha(G)$ that cover the vertex set of G, i.e., $V(G)$ uniformly. That is each vertex of G lies in exactly r of these cocliques, for some constant r. Thus we have

$$b\alpha(G) = nr, \tag{3.12}$$

Now, we define a fractional coloring $\mathbf{f}$ (see C. Godsil and G. Royle [17] pages 137 and 138) as follows

$$f_i = \begin{cases} \frac{1}{r}, & i \in \{1, \cdots, b\}, \\ 0, & \text{Otherwise.} \end{cases} \tag{3.13}$$

Thus, from the definition of the fractional chromatic number of a graph, (3.12), and (3.13), we have

$$\log \chi_f(G) \leq \log \sum_i f_i = \log \frac{b}{r} = \log \frac{n}{\alpha(G)}. \tag{3.14}$$

Now suppose that the probability density $\mathbf{u}$ of the vertex set $V(G)$ is uniform and let $\mathbf{B}$ be the 01-matrix whose columns are the characteristic vectors of the independent sets in G. Then

$$VP(G) = \{\mathbf{x} \in \mathbb{R}_+^n \ : \ \mathbf{B}\lambda = \mathbf{x}, \ \sum_i \lambda_i = 1, \ \lambda_i \geq 0, \forall i\}$$

Consider the function

$$g(\mathbf{x}) = -\frac{1}{n} \sum_{i=1}^{n} \log x_i.$$

We want to minimize $g(\mathbf{x})$ over $VP(G)$. So we use the vector λ in the definition of $VP(G)$ above. Furthermore, from our discussion above, note that each vertex of a vertex transitive graph lies in a certain number of independent sets m. Thus, we rewrite the function $g(.)$ in terms of λ as

$$g(\lambda) = -\frac{1}{n}\log(\lambda_{i_1} + \cdots + \lambda_{i_m}) - \cdots - \frac{1}{n}\log(\lambda_{j_1} + \cdots + \lambda_{j_m}).$$

Now let $\mathcal{S}$ be the set of independent sets of G, and $\nu, \gamma_i \geq 0$ for all $i \in \{1, \cdots, |\mathcal{S}|\}$ be the Lagrange multipliers. Then the Lagrangian function $L_g(\nu, \gamma_1, \cdots, \gamma_{|\mathcal{S}|})$ is

$$L_g(\nu, \gamma_1, \cdots, \gamma_{|\mathcal{S}|}) = g(\lambda) + \nu \left(\sum_{i=1}^{|\mathcal{S}|} \lambda_i - 1 \right) - \sum_{i}^{|\mathcal{S}|} \gamma_i \lambda_i,$$

Now using Karush-Kuhn-Tucker conditions for our convex optimization problem (see S. Boyd and L. Vanderberghe[4]) we get

$$\begin{aligned} &\nabla L_g(\nu, \gamma_1, \cdots, \gamma_{|\mathcal{S}|}) = 0, \\ &\gamma_i \geq 0, \ i \in \{1, \cdots, |\mathcal{S}|\}, \\ &\gamma_i \lambda_i = 0, \ i \in \{1, \cdots, |\mathcal{S}|\}. \end{aligned} \tag{3.15}$$

Then considering the co-clique cover $\{S_1, \cdots, S_b\}$ above with $|S_i| = \alpha(G)$ for all i, one can verify that λ^* defined as

$$\lambda_i^* = \begin{cases} \frac{\alpha(G)}{nr}, & i \in \{1, \cdots, b\}, \\ 0, & \text{Otherwise.} \end{cases} \tag{3.16}$$

is an optimum solution to our minimization problem. Since setting $\gamma_i = 0$ for $i \in \mathcal{S} \setminus \{1, \cdots, b\}$ along with λ^* gives a solution to (3.15). Substituting λ^* into $g(\lambda)$

$$H_k(G, U) = \log \frac{n}{\alpha(G)}.$$

Using (3.14) and Lemma 3.6.1, the corollary is proved. □

The above corollary implies that the uniform probability density is a maximizer for $H_k(G, P)$ for a vertex transitive graph. We will give another proof of this fact at the end of the next chapter using *chromatic entropy*.

We have also the following corollary.

3.6.3 Corollary. *For any graph G and probability density P, we have*

$$H_k(G, P) \leq \log \chi(G).$$

Equality holds if $\chi(G) = \chi_f(G)$ and P maximizes the left hand side above.

Note that (3.1), Lemma 3.2.1, Lemma 3.2.2, and the sub-multiplicative nature of the chromatic number, also results in the above corollary.

3.7 Probability Density Generators

For a pair of vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}_+^k$, $\mathbf{a} \circ \mathbf{b}$ denotes the *Schur product* of $\mathbf{a}$ and $\mathbf{b}$, i.e.,

$$(\mathbf{a} \circ \mathbf{b})_i = a_i.b_i, \;\; i = 1, \cdots, k.$$

Then for two sets $\mathcal{A}$ and $\mathcal{B}$, we have

$$\mathcal{A} \circ \mathcal{B} = \{\mathbf{a} \circ \mathbf{b} : \mathbf{a} \in \mathcal{A}, \;\; \mathbf{b} \in \mathcal{B}\}.$$

We say a pair of sets $\mathcal{A}, \;\; \mathcal{B} \in \mathbb{R}_+^k$ is a *generating pair*, if every probability density vector $\mathbf{p} \in \mathbb{R}_+^k$ can be represented as the schur product of the elements of $\mathcal{A}$ and $\mathcal{B}$, i.e.,

$$\mathbf{p} = \mathbf{a} \circ \mathbf{b}, \;\; \mathbf{a} \in \mathcal{A}, \;\; \mathbf{b} \in \mathcal{B}.$$

In this section we characterize a *pair of generating convex corners*. First, we recall the definition of the *antiblocker* of a convex corner (see D. R. Fulkerson [16]). The *antiblocker* of a convex corner $\mathcal{A}$ is defined as

$$\mathcal{A}^* := \left\{\mathbf{b} \in \mathbb{R}_+^n : \mathbf{b}^T\mathbf{a} \leq 1, \;\; \forall \mathbf{a} \in \mathcal{A}\right\},$$

which is itself a convex corner.

The following lemma relates entropy to antiblocking pairs (see I. Csiszár and et. al. [9]).

3.7.1 Lemma. (I. Csiszár and et. al.). *Let $\mathcal{A}, \mathcal{B} \subseteq \mathbb{R}_+^n$ be convex corners and $\mathbf{p} \in \mathbb{R}_+^n$ a probability density. Then*

(i) *If $\mathbf{p} = \mathbf{a} \circ \mathbf{b}$ for some $\mathbf{a} \in \mathcal{A}$ and $\mathbf{b} \in B$, then*

$$H(\mathbf{p}) \geq H_{\mathcal{A}}(\mathbf{p}) + H_{\mathcal{B}}(\mathbf{p}),$$

with equality if and only if $\mathbf{a}$ and $\mathbf{b}$ achieve $H_{\mathcal{A}}(\mathbf{p})$ and $H_{\mathcal{B}}(\mathbf{p})$.

(ii) *If $\mathcal{B} \subseteq \mathcal{A}^*$ then*

$$H(\mathbf{p}) \leq H_{\mathcal{A}}(\mathbf{p}) + H_{\mathcal{B}}(\mathbf{p}).$$

with equality if and only if $\mathbf{p} = \mathbf{a} \circ \mathbf{b}$ for some $\mathbf{a} \in \mathcal{A}$ and $\mathbf{b} \in \mathcal{B}$.

Proof. (i) We have

$$\begin{aligned} H(\mathbf{p}) &= -\sum_i p_i \log a_i b_i \\ &= -\sum_i p_i \log a_i - \sum_i p_i \log b_i \\ &\geq H_{\mathcal{A}}(\mathbf{p}) + H_{\mathcal{B}}(\mathbf{p}). \end{aligned} \tag{3.17}$$

We have equality if and only if $\mathbf{a}$ and $\mathbf{b}$ achieve $H_{\mathcal{A}}(\mathbf{p})$ and $H_{\mathcal{B}}(\mathbf{p})$.
(ii) Let $\mathbf{a} \in \mathcal{A}$ and $\mathbf{b} \in \mathcal{B}$ achieve $H_{\mathcal{A}}(\mathbf{p})$ and $H_{\mathcal{B}}(\mathbf{p})$, respectively. Then the strict concavity of the log function and the relation $\mathbf{b}^T\mathbf{a} \leq 1$ imply

$$H_{\mathcal{A}}(\mathbf{p}) + H_{\mathcal{B}}(\mathbf{p}) - H(\mathbf{p}) = -\sum_i p_i \log \frac{a_i b_i}{p_i} \geq -\log \sum_i a_i b_i \geq 0.$$

Equality holds if and only if $a_i b_i = p_i$ whenever $p_i > 0$. But then since

$$1 \geq \sum_i a_i b_i \geq \sum_i p_i = 1,$$

equality also holds for those indices with $p_i = 0$. $\square$

The following theorem which was previously proved in I. Csiszár and et. al. [9] characterizes a pair of generating convex corners.

3.7.2 Theorem. (I. Csiszár and et. al.). *For convex corners $\mathcal{A}$, $\mathcal{B} \subseteq \mathbb{R}^k_+$ the following are equivalent:*

(i) $\mathcal{A}^* \subseteq \mathcal{B}$,

(ii) $(\mathcal{A}, \mathcal{B})$ *is a generating pair,*

(iii) $H(\mathbf{p}) \geq H_{\mathcal{A}}(\mathbf{p}) + H_{\mathcal{B}}(\mathbf{p})$ *for every probability density* $\mathbf{p} \in \mathbb{R}^k_+$.

3.8 Additivity and Sub-additivity

If $\mathbf{a} \in \mathbb{R}_+^k$ and $\mathbf{b} \in \mathbb{R}_+^l$ then their *Kronecker product* $\mathbf{a} \otimes \mathbf{b} \in \mathbb{R}_+^{kl}$ is defined by

$$(\mathbf{a} \otimes \mathbf{b})_{ij} = a_i.b_j, \quad i = 1, \cdots, k, \; j = 1, \cdots, l.$$

Note that if $\mathbf{p}$ and $\mathbf{q}$ are probability distributions then $\mathbf{p} \otimes \mathbf{q}$ is the usual product distribution. If $k = l$, then also the Schur product $\mathbf{a} \circ \mathbf{b} \in \mathbb{R}_+^k$ is defined by

$$\mathbf{a} \circ \mathbf{b} = a_i.b_i, \quad i = 1, \cdots, k.$$

Let $\mathcal{A} \subseteq \mathbb{R}_+^k$ and $\mathcal{B} \subseteq \mathbb{R}_+^l$ be convex corners. Their *Kronecker product* $\mathcal{A} \otimes \mathcal{B} \subseteq \mathbb{R}_+^{kl}$ is the convex corner spanned by the Kronecker products $\mathbf{a} \otimes \mathbf{b}$ such that $\mathbf{a} \in \mathcal{A}$ and $\mathbf{b} \in \mathcal{B}$. The *Schur product* $\mathcal{A} \odot \mathcal{B}$ of the convex corners $\mathcal{A}, \; \mathcal{B} \subseteq \mathbb{R}_+^k$ is the convex corner in that same space spanned by the vectors $\mathbf{a} \circ \mathbf{b}$ such that $\mathbf{a} \in \mathcal{A}$ and $\mathbf{b} \in \mathcal{B}$. Thus

$$\mathcal{A} \odot \mathcal{B} = \text{Convex Hull of } \{\mathbf{a} \circ \mathbf{b} : \; \mathbf{a} \in \mathcal{A}, \; \mathbf{b} \in \mathcal{B}\}.$$

I. Csiszár et. al. proved the following lemma and theorem in [9].

3.8.1 Lemma. (I. Csiszár et. al.). *Let $\mathcal{A}, \; \mathcal{B} \subseteq \mathbb{R}_+^k$ be convex corners. The pair $(\mathcal{A}, \mathcal{B})$ is an antiblocking pair if and only if*

$$H(\mathbf{p}) = H_{\mathcal{A}}(\mathbf{p}) + H_{\mathcal{B}}(\mathbf{p})$$

for every probability distribution $\mathbf{p} \in \mathbb{R}_+^k$.

3.8.2 Theorem. (I. Csiszár et. al.). *Let $\mathcal{A} \subseteq \mathbb{R}_+^k$ and $\mathcal{B} \subseteq \mathbb{R}_+^l$ be convex corners, and $\mathbf{p} \in \mathbb{R}_+^k, \; \mathbf{q} \in \mathbb{R}_+^l$ probability distributions. Then, we have*

$$H_{\mathcal{A} \otimes \mathcal{B}}(\mathbf{p} \otimes \mathbf{q}) = H_{\mathcal{A}}(\mathbf{p}) + H_{\mathcal{B}}(\mathbf{q}) = H_{(\mathcal{A}^* \otimes \mathcal{B}^*)^*}(\mathbf{p} \otimes \mathbf{q}),$$

Furthermore, for convex corners $\mathcal{A}, \; \mathcal{B} \subseteq \mathbb{R}_+^k$, and a probability distribution $\mathbf{p} \in \mathbb{R}_+^k$, we have

$$H_{\mathcal{A} \odot \mathcal{B}}(\mathbf{p}) \leq H_{\mathcal{A}}(\mathbf{p}) + H_{\mathcal{B}}(\mathbf{q}).$$

Proof. For $a \in \mathcal{A}$ and $b \in \mathcal{B}$, we have $a \otimes b \in \mathcal{A} \otimes \mathcal{B}$, which implies

$$\begin{aligned} H_{\mathcal{A} \otimes \mathcal{B}}(\mathbf{p} \otimes \mathbf{q}) &\leq -\sum_{i=1}^{k} \sum_{j=1}^{l} p_i q_j \log a_i b_j \\ &= -\sum_{i=1}^{k} p_i \log a_i - \sum_{j=1}^{l} q_j \log b_j. \end{aligned}$$

Hence $H_{\mathcal{A}\otimes\mathcal{B}}(\mathbf{p}\otimes\mathbf{q}) \le H_{\mathcal{A}}(\mathbf{p}) + H_{\mathcal{B}}(\mathbf{q})$. By Lemma 3.8.1,

$$H(\mathbf{p}\otimes\mathbf{q}) = H_{\mathcal{A}\otimes\mathcal{B}}(\mathbf{p}\otimes\mathbf{q}) + H_{(\mathcal{A}\otimes\mathcal{B})^*}(\mathbf{p}\otimes\mathbf{q}).$$

Since $(\mathcal{A})^*\otimes(\mathcal{B})^* \subseteq (\mathcal{A}\otimes\mathcal{B})^*$, we obtain

$$\begin{aligned} H(\mathbf{p}\otimes\mathbf{q}) &\le H_{\mathcal{A}\otimes\mathcal{B}}(\mathbf{p}\otimes\mathbf{q}) + H_{\mathcal{A}^*\otimes\mathcal{B}^*}(\mathbf{p}\otimes\mathbf{q}) \\ &\le H_{\mathcal{A}}(\mathbf{p}) + H_{\mathcal{B}}(\mathbf{q}) + H_{\mathcal{A}^*}(\mathbf{p}) + H_{\mathcal{B}^*}(\mathbf{q}) \\ &\le H(\mathbf{p}) + H(\mathbf{q}) \\ &= H(\mathbf{p}\otimes\mathbf{q}). \end{aligned} \tag{3.18}$$

Thus we get equality everywhere in (3.18), proving

$$H_{\mathcal{A}\otimes\mathcal{B}}(\mathbf{p}\otimes\mathbf{q}) = H_{\mathcal{A}}(\mathbf{p}) + H_{\mathbf{B}}(\mathbf{q}),$$

and consequently,

$$H_{(\mathcal{A}\otimes\mathcal{B})^*}(\mathbf{p}\otimes\mathbf{q}) = H_{\mathcal{A}^*\otimes\mathcal{B}^*}(\mathbf{p}\otimes\mathbf{q}) = H_{\mathcal{A}^*}(\mathbf{p}) + H_{\mathbf{B}^*}(\mathbf{q}).$$

The second claim of the theorem is obviously true. □

As an example let $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ be two graphs. The *OR* product of G_1 and G_2 is the graph $G_1 \bigvee G_2$ with vertex set $V(G_1 \bigvee G_2) = V_1 \times V_2$ and (v_1, v_2) is adjacent to (u_1, u_2) if and only if v_1 is adjacent to u_1 or v_2 is adjacent to u_2. It follows that $VP(G_1 \bigvee G_2) = VP(G_1) \otimes VP(G_2)$. From the above theorem we have

$$H_k\left(G_1 \bigvee G_2, \mathbf{p}\otimes\mathbf{q}\right) = H_k(G_1, \mathbf{p}) + H_k(G_2, \mathbf{q}).$$

Thus if uniform probability densities on the vertices of G_1 and G_2 maximize $H_k(G_1, \mathbf{p})$ and $H_k(G_2, \mathbf{q})$ then the uniform probability density on the vertex of $G_1 \bigvee G_2$ maximizes $H_k(G_1 \bigvee G_2, \mathbf{p}\otimes\mathbf{q})$.

3.9 Perfect Graphs and Graph Entropy

A graph G is *perfect* if for every induced subgraph G' of G, the chromatic number of G' equals the maximum size of a clique in G'. Perfect graphs introduced by Berge in [3] (see C. Berge [3] and L. Lovász [26]).

We defined the vertex packing polytope $VP(G)$ of a graph, in the previous sections. Here, we need another important notion from graph theory, i.e, the *fractional vertex packing polytope* of a graph G. The *fractional vertex packing polytope* of G is defined as

$$FVP(G) = \{\mathbf{b} \in \mathbb{R}^{|V|} : \mathbf{b} \geq 0, \sum_{i \in K} b_i \leq 1 \text{ for all cliques } K \text{ of } G\}$$

It is easy to see that, similar to $VP(G)$, the fractional vertex packing polytope $FVP(G)$ is also a convex corner and $VP(G) \subseteq FVP(G)$ for every graph G. Equality holds here if and only if the graph is perfect (See V. Chvátal [7] and D. R. Fulkerson [16]). Also note that

$$FVP(G) = \left(VP(\overline{G})\right)^*.$$

3.9.1 Lemma. (I. Csiszár and et. al.). *Let $S = \{\mathbf{x} \geq 0, \sum_i x_i \leq 1\}$. Then we have*

$$S = VP(G) \odot FVP(\overline{G}) = FVP(G) \odot VP(\overline{G}).$$

Furthermore,

$$VP(G) \odot VP(\overline{G}) \subseteq FVP(G) \odot FVP(\overline{G})$$

A graph $G = (V, E)$ is *strongly splitting* if for every probability distribution P on V, we have

$$H(P) = H_k(G, P) + H_k\left(\overline{G}, P\right).$$

Körner and Marton in [25] showed that bipartite graphs are strongly splitting while odd cycles are not.

Now, consider the following lemma which was previously proved in I. Csiszár and et. al. [9].

3.9.2 Lemma. *Let G be a graph. For a probability density P on $V(G)$, we have $H(P) = H_k(G, P) + H_k\left(\overline{G}, P\right)$ if and only if $H_{VP(G)}(P) = H_{FVP(G)}(P)$.*

Proof. We have $\left[VP(\overline{G})\right]^* = FVP(G)$. Thus, Lemma 3.2.1 and Lemma 3.8.1 imply

$$\begin{aligned} H_k(G, P) + H_k\left(\overline{G}, P\right) - H(P) &= H_{VP(G)}(P) + H_{VP(\overline{G})}(P) - H(P) \\ &= H_{VP(G)}(P) - H_{FVP(G)}(P). \end{aligned}$$

□

The following theorem conjectured by Körner and Marton in [25] first and proved by I. Csiszár and et. al. in [9].

3.9.3 Theorem. (I. Csiszár and et. al.). *A graph is strongly splitting if and only if it is perfect.*

Proof. By Lemmas 3.1.2 and 3.9.2, G is strongly splitting if and only if $VP(G) = FVP(G)$. This is equivalent to the perfectness of G. $\square$

Let $G = (V, E)$ be a graph with vertex set V and edge set E. The graph

$$G^{[n]} = (V^n, E^{[n]})$$

is the *n-th normal power* where V^n is the set of sequences of length n from V, and two distinct vertices x and y are adjacent in $G^{[n]}$ if all of their entries are adjacent or equal in G, that is

$$E^{[n]} = \{(x, y) \in V^n \times V^n : x \neq y,\ \forall i\ (x_i, y_i) \in E \text{ or } x_i = y_i\}.$$

The *π-entropy of a graph* $G = (V, E)$ with respect to the probability density P on V is defined as

$$H_\pi(G, P) = \lim_{\epsilon \to 0} \lim_{n \to \infty} \min_{U \subseteq V^n, \mathbf{p}^n(U) \geq 1-\epsilon} \frac{1}{n} \log \chi(G^{[n]}(U)).$$

Note that $\overline{G^{[n]}} = \overline{G}^{(n)}$.

The following theorem is proved in G. Simonyi [36].

3.9.4 Theorem. (G. Simonyi). *If $G = (V, E)$ is perfect, then $H_\pi(G, P) = H_k(G, P)$.*

Chapter 4

Chromatic Entropy

A graph G with probability distribution P on its vertex set is called a *Probabilistic Graph* (G, P). In this chapter, we investigate minimum entropy colouring of the vertex set of a probabilistic graph (G, P) which was previously studied by N. Alon and A. Orlitsky [1]. The minimum number of colours $\chi_H(G, P)$ required in a minimum entropy colouring of $V(G)$ was studied by J. Cardinal et. al. [5] and [6]. We state their results and further investigate $\chi_H(G, P)$. For a Kneser graph with uniform probability distribution on its vertex set $(K_{v:r}, U)$ and based on induction on v, we prove that

$$\chi_H(K_{v:r}, U) = \chi(K_{v:r}).$$

Furthermore, using the relationship between the chromatic entropy and the entropy of a vertex transitive graph, we prove that uniform probability distribution is a maximizer of the entropy of a vertex transitive graph.

4.1 Minimum Entropy Coloring

Let X be a random variable distributed over a countable set V and π be a partition of V, i.e., $\pi = \{C_1, \cdots, C_k\}$ and $V = \cup_{i=1}^{k} C_i$. Then π induces a probability distribution on its cells, that is

$$p(C_i) = \sum_{v \in C_i} p(v), \forall i \in \{1, \cdots, k\}.$$

Therefore, the cells of π have a well-defined entropy as follows:

$$H(\pi) = \sum_{i=1}^{k} p(C_i) \log \frac{1}{p(C_i)},$$

If we consider V as the vertex set of a probabilistic graph (G, P) and π as a partitioning of the vertices of G into colour classes, then $H(\pi)$ is the entropy of a proper colouring of $V(G)$.

The *chromatic entropy* of a probabilistic graph (G, P) is defined as

$$H_\chi(G, P) := \min\{H(\pi) : \pi \text{ is a colouring of G}\},$$

i.e. the lowest entropy of any colouring of G.

Example. The empty graph is a graph with a vertex set without any edges. We can colour the vertices of an empty graph with one colour. Thus, an empty graph has chromatic entropy 0. On the other hand, in a proper colouring of the vertices of a complete graph, we require distinct colours for distinct vertices. Hence, a complete graph has chromatic entropy $H(X)$.

Now consider a 5-cycle with two different probability distributions over its vertices, i.e., uniform distribution and another one given by $p_1 = 0.3$, $p_2 = p_3 = p_5 = 0.2$, and $p_4 = 0.1$. In both of them we require three colours. In the first one, a colour is assigned to a single vertex and each of the other two colours are assigned to two vertices. Therefore, the first probabilistic 5-cycle has chromatic entropy

$$H(0.4, 0.4, 0.2) \approxeq 1.52.$$

For the second probabilistic 5-cycle, the chromatic entropy is attained by choosing the colour classes as $\{1, 3\}$, $\{2, 5\}$, and $\{4\}$. Then, its chromatic entropy is

$$H(0.5, 0.4, 0.1) \approxeq 1.36.$$

4.2 Entropy Comparisons

A *source code* ϕ for a random variable X is a mapping from the range of X, i.e., $\mathcal{X}$, to the set of finite-length strings, i.e., $\mathcal{D}^*$, of a D-ary alphabet. Let $\phi(x)$ denote the codeword corresponding to x and let $l(x)$ be the length of $\phi(x)$. Then the average length $L(\phi)$ of the source code ϕ is

$$L(\phi) = \sum_{x \in \mathcal{X}} p(x) l(x).$$

The source coding problem is the problem of representing a random variable by a sequence of bits such that the expected length of the representation is minimized.

N. Alon and A. Orlitsky [1] considered a source coding problem in which a sender wants to transmit an information source to a receiver with some correlated information to the intended information source. Motivated by this problem, they considered the *OR/co-normal product* of graphs, as we stated in the previous chapter. We recall this graph product here.

Let $G_1, \cdots, G_n$ be graphs with vertex sets $V_1, \cdots, V_n$. The *OR product* of $G_1, \cdots, G_n$ is the graph $\bigvee_{i=1}^n G_i$ whose vertex set is V^n and where two distinct vertices $(v_1, \cdots, v_n)$ and $(v'_1, \cdots, v'_n)$ are adjacent if for some $i \in \{1, \cdots, n\}$ such that $v_i \neq v'_i$, v_i is adjacent to v'_i in G_i. The n-fold OR product of G with itself is denoted by $G^{\bigvee n}$.

N. Alon and A. Orlitsky [1] proved the following lemma which relates chromatic entropy to graph entropy.

4.2.1 Lemma. (N. Alon and A. Orlitsky). $\lim_{n\to\infty} \frac{1}{n} H_\chi(G^{\bigvee n}, P^{(n)}) = H_k(G, P)$.

Let $\Omega(G)$ be the collection of cliques of a graph G. For every vertex x we choose a conditional probability distribution $p(z'|x)$ ranging over the cliques containing x. This determines a joint probability distribution of X and a random variable Z' ranging over all cliques containing X. Then, the *clique entropy* is the maximal conditional entropy of X given Z'. That is

$$H_\omega(G, P) := \max\{H(X|Z') : X \in Z' \in \Omega(G)\}.$$

Example. The only cliques of an empty graph are singletons. Thus for an empty graph, we have

$$Z' = \{X\},$$

which implies

$$H_\omega(G, P) = 0.$$

On the other hand, for a complete graph, we can take Z' to be the set of all vertices. Thus, for a probabilistic complete graph (G, P), we have

$$H_\omega(G, P) = H(X).$$

For a 5-cycle, every clique is either a singleton or an edge. Thus, for a probabilistic 5-cycle with uniform distribution over the vertices, we have

$$H_\omega(G,P) \le 1.$$

Now, if for every x we let Z' be uniformly distributed over the two edges containing x, then by symmetry we get

$$H(X|Z') = 1,$$

which implies

$$H_\omega(G,P) = 1.$$

N. Alon and J. Cardinal et. al. proved the following lemmas in [1] and [6].

4.2.2 Lemma. (N. Alon and A. Orlitsky). *Let U be the uniform distribution over the vertices $V(G)$ of a probabilistic graph (G,U) and $\alpha(G)$ be the independence number of the graph G. Then,*

$$H_\chi(G,U) \ge \log \frac{|V(G)|}{\alpha(G)}.$$

□

4.2.3 Lemma. (N. Alon and A. Orlitsky). *For every probabilistic graph (G,P)*

$$H_\omega(G,P) = H(P) - H_k(\overline{G},P).$$

□

4.2.4 Lemma. (J. Cardinal et. al.). *For every probabilistic graph (G,P), we have*

$$-\log \alpha(G,P) \le H_k(G,P) \le H_\chi(G,P) \le \log \chi(G).$$

Here $\alpha(G,P)$ denotes the maximum weight $P(S)$ of an independent set S of (G,P).

□

It may seem that non-uniform distribution decreases chromatic entropy $H_\chi(G,P)$, but the following example shows that this is not true. Let us consider 7-star with $deg(v_1) = 7$ and $deg(v_i) = 1$ for $i \in \{2,\cdots,8\}$. If $p(v_1) = 0.5$ and $p(v_i) = \frac{1}{14}$ for $i \in \{2,\cdots,8\}$, then $H_\chi(G,P) = H(0.5,0.5) = 1$, while if $p(v_i) = \frac{1}{8}$ for $i \in \{1,\cdots,8\}$, then $H_\chi(G,P) = H(\frac{1}{8},\frac{7}{8}) \le H(0.5,0.5) = 1$.

4.3 Number of Colours and Brooks' Theorem

Here, we investigate the minimum number of colours $\chi_H(G, P)$ in a minimum entropy colouring of a probabilistic graph (G, P). First, we have the following definition.

A *Grundy colouring* of a graph is a colouring such that for any colour i, if a vertex has colour i then it is adjacent to at least one vertex of colour j for all $j < i$. The Grundy number $\Gamma(G)$ of a graph G is the maximum number of colours in a Grundy colouring of G. Grundy colourings are colourings that can be obtained by iteratively removing maximal independent sets.

The following theorem was proved in J. Cardinal et. al. [6].

4.3.1 Theorem. (J. Cardinal et. al.) *Any minimum entropy colouring of a graph G equipped with a probability distribution on its vertices is a Grundy colouring. Moreover, for any Grundy colouring of G, there exists a probability mass function P over $V(G)$ such that ϕ is the unique minimum entropy colouring of (G, P).*

We now consider upper bounds on $\chi_H(G, P)$ in terms of the maximum valency of G, i.e., $\Delta(G)$. The following theorems were proved in J. Cardinal et. al. [5] and J. Cardinal et. al. [6].

4.3.2 Theorem. (J. Cardinal et. al.). *For any probabilistic graph (G, P), we have $\chi_H(G, P) \leq \Delta(G) + 1$.*

4.3.3 Theorem. (Brooks' Theorem for Probabilistic Graphs). *If G is connected graph different from a complete graph or an odd cycle, and U is a uniform distribution on its vertices, then $\chi_H(G, U) \leq \Delta(G)$.*

4.4 Grundy Colouring and Minimum Entropy Colouring

Let $\phi : v_1, v_2, \cdots, v_n$ be an ordering of the vertices of a graph G. A proper vertex colouring $c : V(G) \to \mathbb{N}$ of G is a $\phi-$colouring of G if the vertices of G are coloured in the order ϕ, beginning with $c(v_1) = 1$, such that each vertex $v_{i+1} (1 \leq i \leq n-1)$ must be assigned a colour that has been used to colour one or more of the vertices $v_1, v_2, \cdots, v_i$ if possible. If v_{i+1} can be assigned more than one colour, then a colour must be chosen which results in using the fewest number of colours needed to colour G. If v_{i+1} is adjacent to vertices of every currently used colour, then $c(v_{i+1})$ is defined as the smallest positive integer not yet used. The parsimonious $\phi-$colouring number $\chi_\phi(G)$ of G is the number of colours in a $\phi-$colouring of G. The maximum value of $\chi_\phi(G)$ over all orderings ϕ of the vertices of G is the ordered chromatic number or, more simply, the ochromatic number of G, which is denoted by $\chi^o(G)$.

Paul Erdös, William Hare, Stephen Hedetniemi, and Renu Lasker proved the following lemma in [12].

4.4.1 Lemma. (Erdös et. al.). *For every graph G, $\Gamma(G) = \chi^o(G)$.*

Now we prove the following lemma.

4.4.2 Lemma. *For every probabilistic graph (G, P), we have*

$$\max_P \chi_H(G, P) = \Gamma(G).$$

Proof. Due to Theorem 4.3.1 any minimum entropy colouring of a graph G equipped with a probability distribution on its vertices is a Grundy colouring, and for any Grundy colouring ϕ of G, there exists a probability distribution P over $V(G)$ such that ϕ is the unique minimum entropy colouring of (G, P). □

4.4.3 Corollary.

$$\max_P \chi_H(G, P) = \chi^o(G, P).$$

Note that every greedy colouring of the vertices of a graph is a Grundy colouring.

It is worth mentioning that the chromatic number of a vertex transitive graph is not achieved by a Grundy colouring. Let G be a 6-cycle. Consider the first Grundy colouring of G with colour classes $\{v_1, v_4\}$, $\{v_3, v_6\}$, and $\{v_2, v_5\}$ and the second Grundy colouring with colour classes $\{v_1, v_3, v_5\}$ and $\{v_2, v_4, v_6\}$.

It may seem that every Grundy colouring of a probabilistic graph is a minimum entropy colouring, but the following example shows that is not true. Consider a probability distribution for the 6-cycle in the above example as $p(v_1) = p(v_3) = 0.4$ and $p(v_2) = p(v_4) = p(v_5) = p(v_6) = 0.05$. Then, denoting the first Grundy colouring in the example above by c_A and the second one by c_B, we have $H(c_A) = 0.44$ and $H(c_B) = 0.25$ which are not equal.

4.4.1 Remark. Let (G, P) and (G', P') be two probabilistic graphs, and $\phi : G \to G'$ a homomorphism from G to G' such that for every $v' \in V(G')$, we have

$$p'(v') = \sum_{v: v \in \phi^{-1}(v')} p(v).$$

Then, can we say that

$$\chi_H(G,P) \le \chi_H(G',P')?$$

The following example shows that is not true. Let (G,P) be a probabilistic 6-cycle and (G',P') be a probabilistic K_2 with the corresponding probability distributions as follows. $p(v_1) = p(v_4) = 0.4$, $p(v_2) = p(v_3) = p(v_5) = p(v_6) = 0.05$. Then, $\chi_H(C_6,P) = 3$ while $\chi_H(K_2,P') = 2$, i.e., $\chi_H(C_6,P) \ge \chi_H(K_2,P')$. It is worth noting that even as a result of simple operations like deleting an edge, we cannot have the above conjecture. To see this just add an edge between v_1 and v_4 in this example.

4.5 Minimum Entropy Colouring and Kneser Graphs

In this section, we study the minimum entropy colouring of a Kneser graph $K_{v:r}$ and prove the following Theorem.

4.5.1 Theorem. *Let $(K_{v:r},U)$ be a probabilistic Kneser graph with uniform distribution U over its vertices and $v \ge 2r$. Then, for the minimum number of colours in a minimum entropy colouring of $(K_{v:r},U)$, i.e. $\chi_H(K_{v:r},U)$, we have*

$$\chi_H(K_{v:r},U) = \chi(K_{v:r}).$$

Before proving the above theorem, we explain some preliminaries and a lemma which were previously given in J. Cardinal et. al. [6].

Consider a probabilistic graph (G,P). Let S be a subset of the vertices of G, i.e.,

$$S \subseteq V(G).$$

Then $P(S)$ denotes

$$P(S) := \sum_{x\in S} p(x).$$

Note that a *colouring* of $V(G)$ is a map ϕ from the vertex set $V(G)$ of G to the set of positive integers $\mathbb{N}$, that is

$$\phi : V(G) \to \mathbb{N}.$$

Then $\phi^{-1}(i)$ denotes the set of vertices coloured with colour i. Let c_i be the probability of the i-th colour class. Hence, letting X be a random vertex with distribution P ranging over the vertices of G, we get

$$c_i = P(\phi^{-1}(i)) = P(\phi(X) = i).$$

The *colour sequence* of ϕ with respect to P is the infinite vector $c = (c_i)$.

Let (G, P) be a probabilistic graph. A sequence c is said to be *colour-feasible* if there exists a colouring ϕ of $V(G)$ having c as colour sequence. We consider non-increasing colour sequences, that is, colour sequences c such that

$$c_i \geq c_{i+1}, \ \forall \ i.$$

Note that colour sequences define discrete probability distributions on $\mathbb{N}$. Then the *entropy of colour sequence* c of a colouring ϕ, i.e., $H(c)$ is

$$H(c) = -\sum_{i \in \mathbb{N}} c_i \log c_i.$$

The following lemma was proved in N. Alon and A. Orlitsky[1].

4.5.2 Lemma. (N. Alon and A. Orlitsky). *Let c be a non-increasing colour sequence, let i, j be two indices such that $i < j$ and let α a real number such that $0 < \alpha \leq c_j$. Then we have $H(c) > H(c_1, \cdots, c_{i-1}, c_i + \alpha, c_{i+1}, \cdots, c_{j-1}, c_j - \alpha, c_{j+1}, \cdots)$.*

We now examine the consequences of this lemma. We say that a colour sequence c *dominates* another colour sequence d if $\sum_{i=1}^{j} c_j \geq \sum_{i=1}^{j} d_i$ holds for all j. We denote this by $c \succeq d$.Note that $\succeq$ is a partial order. We also let $\succ$ denote the strict part of $\succeq$. The next lemma which was proved in J. Cardinal et. al. [6] shows that colour sequences of minimum entropy colourings are always maximal colour feasible.

4.5.3 Lemma. (J. Cardinal et. al.). *Let c and d be two non-increasing rational colour sequences such that $c \succ d$. Then we have $H(c) < H(d)$.*

Now we prove Theorem 4.5.1.

Proof of Theorem 4.5.1. The proof is based on induction on v. For $v = 2r$ the assertion holds. We prove the assertion for $v > 2r$. Due to Erdös-Ko-Rado theorem, the colour sequence corresponding to the grundy colouring achieving the chromatic number of a Kneser graph dominates all colour feasible sequences. Hence using Lemma 4.5.3, we have $\chi_H(K_{v-1:r}, U) = \chi(K_{v-1:r})$. Now, removing the maximum size coclique in $K_{v:r}$, due to induction hypothesis, we get a minimum entropy colouring of $K_{v-1:r}$. Thus we have $\chi_H(K_{v:r}, U) - 1 = \chi_H(K_{v-1:r}, U) = \chi(K_{v-1:r})$. Noting that $\chi(K_{v:r}) = \chi(K_{v-1:r}) + 1$, we have $\chi_H(K_{v:r}, U) = \chi(K_{v:r})$.

4.5.4 Corollary. *Let $G_1 = (K_{v:r}, U)$, and (G_2, U) is homomorphically equivalent, in the sense of Remark 4.4.1, to G_1, then we have $\chi_H(G_1) = \chi_H(G_2) = \chi(G_1) = \chi(G_2)$.*

4.6 Further Results

As we mentioned in the previous chapter, for a probabilistic graph (G, P), we have

$$\max_P H_k(G, P) = \log \chi_f(G). \tag{4.1}$$

In this section, we prove the following theorem for vertex transitive graphs using chromatic entropy. Recall that we gave another proof of the following theorem using the structure of vertex transitive graphs and convex optimization techniques in previous chapter.

4.6.1 Theorem. *Let G be a vertex transitive graph. Then the uniform distribution over vertices of G maximizes $H_k(G, P)$. That is $H_k(G, U) = \log \chi_f(G)$.*

Proof. First note that for a vertex transitive graph G, we have $\chi_f(G) = \frac{|V(G)|}{\alpha(G)}$, and the n-fold OR product $G^{\bigvee n}$ of a vertex transitive graph G is also vertex transitive. Now from Lemma 4.2.2, Lemma 4.2.4 , and equation 4.1, we have

$$H_k\left(G^{\bigvee n}, U\right) \leq \log \chi_f\left(G^{\bigvee n}\right) \leq H_\chi\left(G^{\bigvee n}, U\right), \tag{4.2}$$

From [1] and [38], we have $H_k\left(G^{\bigvee n}, U\right) = nH_k(G, U)$, $\chi_f\left(G^{\bigvee n}\right) = \chi_f(G)^n$, and $\log \chi_f(G) = \lim_{n\to\infty} \frac{1}{n} \log \chi\left(G^{\bigvee n}\right)$. Hence, applying Lemma 4.2.1 to equation 4.2 and using squeezing theorem, we get

$$H_k(G, U) = \log \chi_f(G) = \lim_{n\to\infty} \frac{1}{n} \log \chi\left(G^{\bigvee n}\right) = \lim_{n\to\infty} \frac{1}{n} H_\chi\left(G^{\bigvee n}, U\right). \tag{4.3}$$

□

The following example shows that the converse of the above theorem is not true. Consider $G = C_4 \cup C_6$, with vertex sets $V(C_4) = \{v_1, v_2, v_3, v_4\}$ and $V(C_6) = \{v_5, v_6, v_7, v_8, v_9, v_{10}\}$, and parts $A = \{v_1, v_3, v_5, v_7, v_9\}$, $B = \{v_2, v_4, v_6, v_8, v_{10}\}$. Clearly, G is not a vertex transitive graph, however, using Theorem 3.5.3, one can see that the uniform distribution $U = \left(\frac{1}{10}, \cdots, \frac{1}{10}\right)$ gives the maximum graph entropy which is 1.

Now note that we can describe the chromatic entropy of a graph in terms of the graph entropy of a complete graph as

$$H_\chi(G, P) = \min\{H_k(K_n, P') : (G, P) \to (K_n, P')\}.$$

A graph G is called *symmetric with respect to a functional* $F_G(P)$ defined on the set of all the probability distributions on its vertex set if the distribution P^* maximizing $F_G(P)$ is uniform on $V(G)$. We study this concept in more detail in the next chapter.

Chapter 5

Symmetric Graphs

A graph G with distribution P on its vertices is called *symmetric with respect to graph entropy* $H_k(G, P)$ if the uniform probability distribution on its vertices maximizes $H_k(G, P)$.

In this chapter we characterize different classes of graphs which are symmetric with respect to graph entropy. Using Hall's, König's, and the entropy of bipartite graphs, we characterize bipartite graphs which are symmetric with respect to graph entropy. We also characterize symmetric perfect graphs using weak perfect graph theorem and basic properties of graph entropy. Furthermore, we find a class of line graphs which are symmetric with respect to graph entropy using Lagrange multipliers in convex optimization. As a corollary to this result, we show that the line graph of every bridgeless cubic is also symmetric with respect to graph entropy.

5.1 Symmetric Bipartite Graphs

5.1.1 Theorem. *Let G be a bipartite graph with parts A and B, and no isolated vertices. The uniform probability distribution U over the vertices of G maximizes $H_k(G, P)$ if and only if G has a perfect matching.*

Proof. Suppose G has a perfect matching, then $|A| = |B|$, and due to Hall's theorem we have

$$|D| \leq |\mathcal{N}(D)|, \quad \forall D \subseteq A.$$

Now assuming $P = U$, we have

$$p(D) = \frac{|D|}{|V(G)|}, \qquad p(A) = \frac{|A|}{|V(G)|} = \frac{|B|}{|V(G)|} = p(B),$$

Thus, the condition of Theorem 3.5.3 is satisfied, that is

$$\frac{p(D)}{p(A)} \leq \frac{p(\mathcal{N}(D))}{p(B)}, \quad \forall D \subseteq A,$$

Then, due to Theorem 3.5.3, we have

$$H_k\left(G, U\right) = h\left(p(A)\right) = h\left(\frac{1}{2}\right) = 1.$$

Noting that $H_k\left(G, P\right) \leq \log \mathcal{X}_f(G),\ \forall P$, and $\log \mathcal{X}_f(G) = 1$ for a bipartite graph G, the assertion holds.

Now suppose that G has no perfect matching, then we show that $H_k\left(G, U\right) < 1$. First, note that from König's theorem we can say that a bipartite graph $G = (V, E)$ has a perfect matching if and only if each vertex cover has size at least $\frac{1}{2}|V|$. This implies that if a bipartite graph G does not have a perfect matching, then G has an independent set with size $> \frac{|V|}{2}$.

Furthermore, as mentioned in [34], the vertex packing polytope of a graph G is determined by the following inequalities if and only if G is bipartite.

$$\begin{gathered} 0 \leq x_v \leq 1, \quad \forall v \in V(G), \\ x_u + x_v \leq 1, \quad \forall e = uv \in E(G). \end{gathered}$$

We show $\max_{x\in\ VP(G)} \prod_{v\in V} x_v > 2^{-|V|}$. Let S denote an independent set in G with $|S| > \frac{|V|}{2}$. We define a vector $\overline{\mathbf{x}}$ such that $\overline{x}_v = \frac{|S|}{|V|}$ if $v \in S$ and $\overline{x}_v = 1 - \frac{|S|}{|V|}$ otherwise. Since $|S| > \frac{|V|}{2}$, $\overline{\mathbf{x}}$ is feasible. Letting $t := \frac{|S|}{|V|}$ and noting that $|S| > \frac{|V|}{2}$, we have

$$-\left(t\log t + (1-t)\log(1-t)\right) < 1,$$

This shows that $H_k\left(G, U\right) < 1$ which contradicts the fact G is symmetric with respect to graph entropy. $\square$

5.2 Symmetric Perfect Graphs

Let $G = (V, E)$ be a graph. Recall that the fractional vertex packing polytope of G,i.e, $FVP(G)$ is defined as

$$FVP(G) := \{\mathbf{x} \in \mathbb{R}_+^{|V|} : \sum_{v \in K} x_v \leq 1 \text{ for all cliques K of G}\}.$$

Note that $FVP(G)$ is a convex corner and for every graph G, $VP(G) \subseteq FVP(G)$. The following theorem was previously proved in [7] and [16].

5.2.1 Theorem. (V. Chvátal and D. R. Fulkerson). *A graph G is perfect if and only if $VP(G) = FVP(G)$.*

The following theorem which is called *weak perfect graph theorem* is useful in the following discussion. This theorem was proved by Lovász in [27] and [28] . We state the theorem here.

5.2.2 Theorem. (L. Lovász). *A graph G is perfect if and only if its complement is perfect.*

Now, we prove the following theorem which is a generalization of our bipartite symmetric graphs with respect to graph entropy.

5.2.3 Theorem. *Let $G = (V, E)$ be a perfect graph and P be a probability distribution on $V(G)$. Then G is symmetric with respect to graph entropy $H_k(G, P)$ if and only if G can be covered by its cliques of maximum size.*

Proof. Suppose G is covered by its maximum-sized cliques, say $Q_1, \cdots, Q_m$. That is $V(G) = V(Q_1)\dot{\cup} \cdots \dot{\cup} V(Q_m)$ and $|V(Q_i)| = \omega(G),\ \forall i \in [m]$.

Now, consider graph T which is the disjoint union of the subgraphs induced by $V(Q_i)$ $\forall i \in [m]$. That $T = \dot{\bigcup}_{i=1}^{m} G[V(Q_i)]$. Noting that T is a disconnected graph with m components, using Corollary 3.4.4 we have

$$H_k(T, P) = \sum_i P(Q_i) H_k(Q_i, P_i).$$

Now, having $V(T) = V(G)$ and $E(T) \subseteq E(G)$, we get $H_k(T, P) \leq H_k(G, P)$ for every distribution P. Using Lemma 3.6.1, this implies

$$H_k(T,P) = \sum_i P(Q_i) H_k(Q_i, P_i) \leq \log \chi_f(G), \ \forall P, \tag{5.1}$$

Noting that G is a perfect graph, the fact that complete graphs are symmetric with respect to graph entropy, $\chi_f(Q_i) = \chi_f(G) = \omega(G) = \chi(G)$, $\forall i \in [m]$, and (5.1), we conclude that uniform distribution maximizes $H_k(G,P)$.

Now, suppose that G is symmetric with respect to graph entropy. We prove that G can be covered by its maximum-sized cliques. Suppose this is not true. We show that G is not symmetric with respect to $H_k(G,P)$.

Denoting the minimum clique cover number of G by $\gamma(G)$ and the maximum independent set number of G by $\alpha(G)$, from perfection of G and weak perfect theorem, we get $\gamma(G) = \alpha(G)$. Then, using this fact, our assumption implies that G has an independent set S with $|S| > \frac{|V(G)|}{\omega(G)}$.

We define a vector $\overline{\mathbf{x}}$ such that $\overline{x}_v = \frac{|S|}{|V|}$ if $v \in S$ and $\overline{x}_v = \frac{1-\frac{|S|}{|V|}}{\omega-1}$ if $v \in V(G)\backslash S$. Then, we can see that $\overline{\mathbf{x}} \in FVP(G) = VP(G)$. Let $t := \frac{|S|}{|V|}$. Then, noting that $t > \frac{1}{\omega}$,

$$\begin{aligned}
H_k(G,U) &\leq -\frac{1}{|V|}\sum_{v\in V}\log \overline{x}_v \\
&= -\frac{1}{|V|}\left(\sum_{v\in S}\log \overline{x}_v + \sum_{v\in V\backslash S}\overline{x}_v\right) \\
&= -\frac{1}{|V|}\left(|S|\log\alpha + (|V|-|S|)\log\frac{1-\alpha}{\omega-1}\right) \\
&= -t\log t - (1-t)\log\frac{1-t}{\omega-1} \\
&= -t\log t - (\omega-1)\left(\frac{1-t}{\omega-1}\log\frac{1-t}{\omega-1}\right) < \log\omega(G).
\end{aligned}$$

□

Note that we have

$$\gamma(G) = \alpha(G).$$

Now, considering that finding the clique number of a perfect graph can be done in polynomial time and using weak perfect graph theorem we conclude that one can decide in polynomial time whether a perfect graph is symmetric with respect to graph entropy.

5.2.4 Corollary. *Let G be a connected regular line graph without any isolated vertices with valency $k > 3$. Then if G is covered by its disjoint maximum-size cliques, then G is symmetric with respect to $H_k(G, P)$.*

Proof. Let $G = L(H)$ for some graph H. Then either H is bipartite or regular. If H is bipartite, then G is perfect (See [40]) and because of Theorem 5.2.3 we are done. So suppose that H is not bipartite. Then each clique of size k in G corresponds to a vertex v in $V(H)$ and the edges incident to v in H and vice versa. That is because any such cliques in G contains a triangle and there is only one way extending that triangle to the whole clique which corresponds to edges incident with the corresponding vertex in H. This implies that the covering cliques in G give an independent set in H which is also a vertex cover in H. Hence H is a bipartite graph and hence G is perfect. Then due to Theorem 5.2.3 the theorem is proved. □

5.3 Symmetric Line Graphs

In this section we introduces a class of line graphs which are symmetric with respect to graph entropy. Let G_2 be a line graph of some graph G_1, i.e, $G_2 = L(G_1)$. Let $|V(G_1)| = n$ and $|E(G_1)| = m$. We recall that a vector $\mathbf{x} \in \mathbb{R}_+^m$ is in the matching polytope $MP(G_1)$ of the graph G_1 if and only if it satisfies (see A. Schrijver [34]).

$$\begin{aligned} x_e &\geq 0 && \forall e \in E(G_1), \\ x(\delta(v)) &\leq 1 && \forall v \in V(G_1), \\ x\left(E[U]\right) &\leq \lfloor \tfrac{1}{2}|U| \rfloor, && \forall U \subseteq V(G_1) \text{ with } |U| \text{ odd.} \end{aligned} \tag{5.2}$$

Let $\mathcal{M}$ denote the family of all matchings in G_1, and for every matching $M \in \mathcal{M}$ let the characteristic vector $\mathbf{b}_M \in \mathbb{R}_+^m$ be as

$$(\mathbf{b}_M)_e = \begin{cases} 1, & e \in M, \\ 0, & e \notin M. \end{cases} \tag{5.3}$$

Then the *fractional edge-colouring number* $\chi'_f(G_1)$ of G_1 is defined as

$$\chi'_f(G_1) := \min\{\sum_{M \in \mathcal{M}} \lambda_M | \lambda \in \mathbb{R}_+^{\mathcal{M}}, \sum_{M \in \mathcal{M}} \lambda_M \mathbf{b}_M = \mathbf{1}\}.$$

If we restrict λ_M to be an integer, then the above definition give rise to the edge colouring number of G_1, i.e., $\chi'(G_1)$. Thus

$$\chi'_f(G) \leq \chi'(G).$$

As an example considering G_1 to be the peterson graph, we have

$$\chi'_f(G) = \chi'(G) = 3.$$

5.3.1 Remark. Note that every matching in G_1 corresponds to an independent set in G_2 and every independent set in G_2 corresponds to a matching in G_1. Note that the fractional edge-colouring number of G_1, i.e., $\chi'_f(G_1)$ is equal to the fractional chromatic number of G_2, i.e.,$\chi_((G_2)$. Thus

$$\chi'_f(G_1) = \chi_f(G_2).$$

Furthermore, note that the vertex packing polytope $VP(G_2)$ of G_2 is the matching polytope $MP(G_1)$ of G_1 (see L. Lovász and M. D. Plummer [30]). That is

$$VP(G_2) = MP(G_1).$$

The following theorem which was proved by Edmonds, gives a characterization of the fractional edge-colouring number $\chi'_f(G_1)$ of a graph G_1 (see A. Schrijver [34]).

5.3.1 Theorem. *Let $\Delta(G_1)$ denote the maximum degree of G_1. Then the fractional edge-colouring number of G_1 is obtained as*

$$\chi'_f(G_1) = \max\{\Delta(G_1), \max_{U\subseteq V,\ |U|\geq 3} \frac{|E(U)|}{\lfloor \frac{1}{2}|U| \rfloor}\}.$$

Following A. Schrijver [34] we call a graph G_1 a *k-graph* if it is k-regular and its fractional edge coloring number $\chi'_f(H)$ is equal to k. The following colloray characterizes a k-graph (see A. Schrijver [34]).

5.3.2 Corollary. *Let $G_1 = (V_1, E_1)$ be a k-regular graph. Then $\chi'_f(G_1) = k$ if and only if $|\delta(U)| \geq k$ for each odd subset U of V_1.*

The following theorem introduces a class of symmetric line graphs with respect to graph entropy. The main tool in the proof of the following theorem is Karush-Kuhn-Tucker (KKT) optimality conditions in convex optimization (see S. Boyd and L. Vanderberghe [4]).

5.3.3 Theorem. *Let G_1 be a k-graph with $k \geq 3$. Then the line graph $G_2 = L(G_1)$ is symmetric with respect to graph entropy.*

Proof. From our discussion in Remark 5.3.1 above we have

$$H_k\left(G_2, P\right) = \min_{\mathbf{x}\in MP(G_1)} \sum_{e\in E(G_1)} p_e \log \frac{1}{x_e},$$

Let $\lambda_v,\ \gamma_U \geq 0$ be the Lagrange multipliers corresponding to inequalities $x(\delta(v)) \leq 1$ and $x\left(E[U]\right) \leq \lfloor \frac{1}{2}|U| \rfloor$ in the description of the matching polytope $MP(G_1)$ in (5.2) for all $v \in V(G_1)$ and for all $U \subseteq V(G_1)$ with $|U|$ odd, and $|U| \geq 3$, respectively.
Now, note that the function

$$-\sum_{e\in E(G_1)} p_e \log x_e$$

is a convex function and tends to infinity at the boundary of the non-negative orthant and tends monotonically to $-\infty$ along the rays from the origin. Thus, the Lagrange multipliers corresponding to inequalities $x_e \geq 0$ are all zero.

Set

$$g(\mathbf{x}) = -\sum_{e\in E(G_1)} p_e \log x_e,$$

Then the Lagrangian of $g(\mathbf{x})$ is

$$\begin{aligned} L\left(\mathbf{x}, \lambda, \gamma\right) \quad &= -\sum_{e\in E(G_1)} p_e \log x_e + \sum_{e=\{u,v\}} \left(\lambda_u + \lambda_v\right)\left(x_e - 1\right) \\ &+ \sum_{e\in E(G_1)} \sum_{\substack{U\subseteq V, \\ U\ni e, |U| \text{ odd}, \ |U|\geq 3}} \gamma_U x_e - \sum_{\substack{U\subseteq V, \\ |U| \text{ odd}, \ |U|\geq 3}} \lfloor \frac{1}{2}|U| \rfloor, \end{aligned} \tag{5.4}$$

Using KKT conditions (see S. Boyd, and L. Vanderberghe [4]), the vector $\mathbf{x}^*$ minimizes $g(\mathbf{x})$ if and only if it satisfies

$$\begin{aligned} &\frac{\partial L}{\partial x_e^*} = 0, \\ &\to -\frac{p_e}{x_e^*} + (\lambda_u + \lambda_v) + \sum_{\substack{U\subseteq V, \\ U\ni e, |U| \text{ odd}, \ |U|\geq 3}} \gamma_U = 0 \text{ for } e = \{u, v\}. \end{aligned} \tag{5.5}$$

Fix the probability density to be uniform over the edges of G_1, that is

$$p_e = \frac{1}{m}, \quad \forall e \in E(G_1),$$

Note that the vector $\frac{\mathbf{1}}{k}$ is a feasible point in the matching polytope $MP(G_1)$. Now, one can verify that specializing the variables as

$$\begin{aligned}
&\mathbf{x}^* = \frac{\mathbf{1}}{k}, \\
&\gamma_U = 0 \qquad\qquad \forall U \subseteq V,\ |U| \text{ odd},\ |U| \geq 3 \\
&\lambda_u = \lambda_v = \frac{k}{2m} \qquad\qquad \forall\ e = \{u, v\}.
\end{aligned}$$

satisfies the equations (5.5). Thus

$$H_k\left(G_2, \mathbf{u}\right) = \log k.$$

Then using Lemma 3.6.1 and the assumption $\chi_f(G_2) = k$ the theorem is proved. □

It is well known that cubic graphs have a lot of interesting structures. For example, it can be checked that every edge in a bridgeless cubic graph is in a perfect matching. Furthermore, L. Lovász and M. D. Plummer [30] conjectured that every bridgeless cubic graph has an exponentially many perfect matching. This conjecture was proved by Louis Esperet, et. al. [13] recently. Now we have the following interesting statement for every cubic bridgeless graph.

5.3.4 Corollary. *The line graph of every cubic bridgeless graph $G_1 = (V_1, E_1)$ is symmetric with respect to graph entropy.*

Proof. We may assume that G_1 is connected. Let $U \subseteq V_1$ and let $U_1 \subseteq U$ consist of vertices v such that $\delta(v) \cap \delta(U) = \emptyset$. Then using handshaking lemma for $G_1[U]$, we have

$$3|U_1| + 3|U \setminus U_1| - |\delta(U)| = 2|E(G_1[U])|.$$

And consequently,

$$3|U| = |\delta(U)| \mod 2,$$

Assuming $|U|$ is odd and noting that G_1 is bridgeless, we have

$$\delta(U) \geq 3.$$

Then, considering Corollary 5.3.2, the corollary is proved. □

Figure 5.1 shows a bridgeless cubic graph which is not edge transitive and its edges are not covered by disjoint copies of stars and triangles. Thus the line graph of the shown graph in Figure 5.1 is neither vertex transitive nor covered by disjoint copies of its maximum size cliques. However, it is symmetric with respect to graph entropy by Corollary 5.3.4.

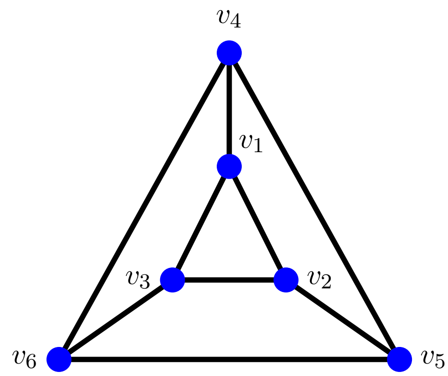


Figure 5.1: A bridgeless cubic graph.

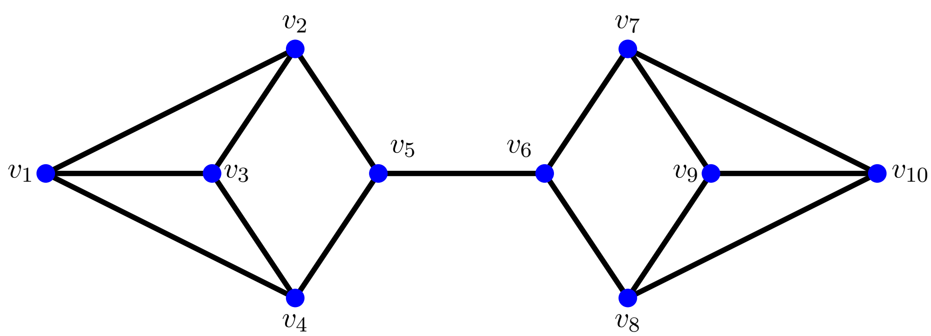


Figure 5.2: A cubic one-edge connected graph.

Figure 5.2 shows a cubic graph with a bridge. The fractional edge chromatic number of this graph is 3.5 while the entropy of its line graph is 1.75712, i.e., $\log_2 3.5 = 1.8074 > 1.75712$. Thus, its line graph is not symmetric with respect to graph entropy, and we conclude that Corollary 5.3.4 is not true for cubic graphs with bridge.

Chapter 6

Future Work

In this chapter we explain two possible research directions related to the entropy of graphs discussed in previous chapters. Since these directions are related to a superclass of perfect graphs which are called *normal graphs* and *Lovász* ϑ, we explain the corresponding terminologies and results in the sequel.

6.1 Normal Graphs

Let G be a graph. A set $\mathcal{A}$ of subsets of $V(G)$ is a *covering*, if every vertex of G is contained in an element of $\mathcal{A}$.

We say that graph G is *Normal* if there exists two coverings $\mathcal{C}$ and $\mathcal{S}$ such that every element C of $\mathcal{C}$ is a clique and every element S of $\mathcal{S}$ is an independent set and the intersection of any element of $\mathcal{C}$ and any element of $\mathcal{S}$ is nonempty, i.e.,

$$C \cap S \neq \emptyset, \ \forall C \in \mathcal{C}, \ S \in \mathcal{S}.$$

Recall from the sub-additivity of Graph Entropy, we have

$$H(P) \leq H_k(G, P) + H_k(\overline{G}, P). \tag{6.1}$$

A probabilistic graph (G, P) is *weakly splitting* if there exists a nowhere zero probability distribution P on its vertex set which makes inequality (6.1) equality. The following lemma was proved in J. Körner et. al. [23].

6.1.1 Lemma. (J. Körner, G. Simonyi, and Zs. Tuza) *A graph G is weakly splitting if and only if it is normal.*

Furthermore, we call (G, P) is strongly splitting if inequality (6.1) becomes equality for every probability distribution P. The following lemma was proved in I. Csiszár et. al. [9].

6.1.2 Lemma. (I. Csiszár et. al.) *For a probabilistic graph (G, P), we have*

$$H(P) = H_k(G, P) + H_k(\overline{G}, P) \text{ if and only if } H_{VP(G)}(P) = H_{FVP(G)}(P).$$

Furthermore, it is shown in I. Csiszár et. al. [9] that

6.1.3 Lemma. (I. Csiszár et. al.) *A graph G is perfect if and only*

$$H_{VP(G)}(P) = H_{FVP(G)}(P).$$

Using Lemmas 6.1.1, 6.1.2, and 6.1.3, we conclude that every perfect graph is also a normal graph. This fact was previously proved in J. Körner [20]. It is shown in [32] that the line graph of a cubic graph is normal. Furthermore, it is shown in J. Körner [20] that every odd cycle of length at least nine is normal. Smaller odd cycles are either perfect like a triangle or not perfect nor normal like C_5 and C_7. If we require that every induced subgraph of a normal graph to be normal, we obtain the notion of hereditary normality. The following conjecture was proposed in C. De Simone and J. Körner [15].

6.1.4 Conjecture. *Normal Graph Conjecture* A graph is hereditarily normal if and only if the graph nor its complement contains C_5 or C_7 as an induced subgraph.

A *circulant* C_n^k is a graph with vertex set $\{1, \cdots, n\}$, and two vertices $i \neq j$ are adjacent if and only if

$$i - j \equiv k \bmod n.$$

We assume $k \geq 1$ and $n \geq 2(k+1)$ to avoid cases where C_n^k is an independent set or a clique. The following theorem was proved in L. E. Trotter, jr. [37].

6.1.5 Theorem. (L. E. Trotter, jr.) *The circulant $C_{n'}^{k'}$ is an induced subgraph of C_n^k if and only if*

$$\frac{k+1}{k'+1} n' \leq n \leq \frac{k}{k'}.$$

Note that

$$C_{n'}^{k'} \subset C_n^k,$$

implies $k' < k$ and $n' < n$. Particularly, the following lemma was proved in A. K. Wagler [39].

6.1.6 Lemma. (A. K. Wagler)

(i) $C_5 \subseteq C_n^k$ *if and only if* $\frac{5(k+1)}{2} \leq n \leq 5k$.

(ii) $C_7 \subseteq C_n^k$ *if and only if* $\frac{7(k+1)}{2} \leq n \leq 7k$.

(iii) $C_7^2 \subseteq C_n^k$ *if and only if* $\frac{7(k+1)}{3} \leq n \leq \frac{7k}{2}$.

Using the above theorem and lemma, A. K. Wagler [39] proved the Normal Graph Conjecture for circulants C_n^k.

One direction for future research is investigating the Normal Graph Conjecture for general circulants and Cayley graphs.

6.2 Lovász ϑ Function and Graph Entropy

An old problem in information and graph theory is to determine the *zero error Shannon capacity* $C(G)$ of a graph G. Let G be a graph with vertex set $V(G)$ and edge set $E(G)$. The n-th normal power of G is the graph G^n with vertex set $V(G^n) = (V(G))^n$ and two vertices $(x_1, \cdots, x_n) \neq (y_1, \cdots, y_n)$ are adjacent if and only if

$$x_i = y_i \text{ or } \{x_i, y_i\} \in E(G) \; \forall i \in \{1, \cdots, n\}.$$

The *zero error Shannon capacity* $C(G)$ of a graph G is defined as

$$C(G) = \limsup_{n \to \infty} \frac{1}{n} \log \alpha \left(G^n\right).$$

Let P denote the probability distribution over the vertices of G, and $\epsilon > 0$. Let $x^n \in V^n$ be an n-sequence whose entries are from $\mathcal{X}$, and $N\left(a|x^n\right)$ denote the number of occurrences of an element $a \in \mathcal{X}$. We call the set of (P, ϵ)-typical sequences $T^(P, \epsilon)$ to be the set of n-sequences $x^n \in V^n$ such that

$$|N(a|x^n) - P(X = a)| \leq n\epsilon.$$

Then the *capacity of the graph relative to* P is

$$C(G, P) = \lim_{\epsilon \to 0} \limsup_{n \to \infty} \frac{1}{n} \log \alpha \left(G^(P, \epsilon)\right).$$

Given a probabilistic graph (G, P), K. Marton in K. Marton [31] introduced a functional $\lambda(G, P)$ which is analogous to Lovász's bound $\vartheta(G)$ on Shannon capacity of graphs. Similar to $\vartheta(G)$, the probabilistic functional $\lambda(G, P)$ is based on the concept of *orthonormal representation* of a graph which is recalled here.

Let $U = \{\mathbf{u}_i : i \in V(G)\}$ be a set of unit vectors of a common dimension d such that

$$\mathbf{u}_i^T \mathbf{u}_j = 0 \text{ if } i \neq j \text{ and } \{i, j\} \notin E(G).$$

Let $\mathbf{c}$ be a unit vector of dimension d. Then, the system $(U, \mathbf{c})$ is called an orthonormal representation of the graph G with handle $\mathbf{c}$.

Letting $T(G)$ denote the set of all orthonormal representations with a handle for graph G, L. Lovász [29] defined

$$\vartheta(G) = \min_{(U,\mathbf{c}) \in T(G)} \max_{i \in V(G)} \frac{1}{(\mathbf{u}_i, \mathbf{c})^2}.$$

Then it is shown in L. Loász [29] that zero error Shannon capacity $C(G)$ can be bounded above by $\vartheta(G)$ as

$$C(G) \leq \log \vartheta(G).$$

A probabilistic version of $\vartheta(G)$ denoted by $\lambda(G, P)$ is defined in K. Marton [31] as

$$\lambda(G, P) := \min_{(U,\mathbf{c}) \in T(G)} \sum_{i \in V(G)} P_i \log \frac{1}{(\mathbf{u}_i, \mathbf{c})^2}.$$

K. Marton [31] showed that

6.2.1 Theorem. (K. Marton) *The capacity of a probabilistic graph* (G, P) *is bounded above by* $\lambda(G, P)$*, i.e.,*

$$C(G, P) \leq \lambda(G, P).$$

The following theorem was proved in K. Marton [31] which relates $\lambda(G, P)$ to $H_k(G, P)$.

6.2.2 Theorem. (K. Marton) *For any probabilistic graph* (G, P)*,*

$$\lambda\left(\overline{G}, P\right) \leq H_k\left(G, P\right).$$

Furthermore, equality holds if and only if G *is perfect.*

K. Marton [31] also related $\lambda(G,P)$ to $\vartheta(G)$ by showing

$$\max_{P} \lambda(G,P) = \log \vartheta(G). \tag{6.2}$$

It is worth mentioning that $\vartheta(G)$ can be defined in terms of graph homomorphisms as follows.

Let $d \in \mathbb{N}$ and $\alpha < 0$. Then we define $S(d,\alpha)$ to be an infinite graph whose vertices are unit vectors in $\mathbb{R}^d$. Two vertices $\mathbf{u}$ and $\mathbf{v}$ are adjacent if and only if $\mathbf{u}\mathbf{v}^T = \alpha$. Then

$$\vartheta\left(\overline{G}\right) = \min\left\{1 - \frac{1}{\alpha} : G \to S\left(d,\alpha\right),\ \alpha < 0\right\}. \tag{6.3}$$

Thus, noting (6.2) and (6.3) and the above discussion, investigating the relationship between graph homomorphism and graph entropy which may lead to investigating the relationship between graph homomorphism and graph covering problem seems interesting.

Appendix A

Proof of Lemma 3.2.2

Here we explain the proof of Lemma 3.2.2 with more detail compare to the proof already stated in the literature.

First, we state a few lemmas as follow.

A. 1. Lemma. The chromatic number of a graph G, i.e., $\chi(G)$ is equal to the minimum number of maximal independent sets covering G.

Proof. Let $\kappa(G)$ be the minimum number of maximal independent sets covering the vertices of G. Then $\kappa(G) \leq \chi(G)$, since the colour classes of any proper colouring of $V(G)$ can be extended to maximal independent sets. On the other hand, consider a covering system consisting of maximal independent sets $\mathcal{S}$ with a minimum number of maximal independent sets. Let $\mathcal{S} = \{S_1, \cdots, S_{\kappa(G)}\}$. We define a colouring c of the vertices of graph G as

$$c(v) = i, \ \forall v \in S_i \setminus S_{i-1}, \text{ and } \forall i \in \{1, \cdots, \kappa(G)\},$$

The proposed colouring is a proper colouring of the vertices of $V(G)$ in which each colour class corresponds to a maximal independent set in our covering system $\mathcal{S}$. That is

$$\kappa(G) \geq \chi(G).$$

□

Let $\mathcal{X}$ be a finite set and let P be a probability density on its elements. Let K be a constant. Then, a sequence $\mathbf{x} \in \mathcal{X}^n$ is called *P-typical* if for every $y \in \mathcal{X}$ and for the number of occurrences of the element y in $\mathbf{x}$, i.e., $N(y|\mathbf{x})$, we have

$$|N(y|\mathbf{x}) - np(y)| \leq K\sqrt{p(y)}.$$

Then we have the following lemma.

A. 2. Lemma. Let $T^n(P)$ be the set of the P-typical n-sequences. Then,

(i) For all $\epsilon > 0$ there exists $K > 0$ such that

$$P\left(\overline{T^n(P)}\right) < \epsilon, \quad \text{for this } K.$$

(ii) For every typical sequence $\mathbf{x}$ we have

$$2^{-(nH(P)+C\sqrt{n})} \leq P(\mathbf{x}) \leq 2^{-(nH(P)-C\sqrt{n})},$$

for some constant $C > 0$ depending on $|\mathcal{X}|$ and $\epsilon > 0$ and independent of n and P,

(iii) The number of typical sequences $N(n)$ is bounded as

$$2^{nH(P)-C\sqrt{n}} \leq N(n) \leq 2^{nH(P)+C\sqrt{n}}.$$

for some constant $C > 0$ depending on $|\mathcal{X}|$ and $\epsilon > 0$ and independent of n and P.

□

Having $\mathcal{X}$ defined as above, let (G, P) be a probabilistic graph with vertex $V(G) = \mathcal{X}$. We define the relation e as

$$xey \Longleftrightarrow \text{either}\{x, y\} \in E(G) \text{ or } x = y.$$

If e determines an equivalence relation on the vertex set $V(G)$, then graph G is the union of pairwise disjoint cliques. Let $H(P|e)$ denote the conditional entropy given the equivalence class e, i.e.,

$$H(P|e) = \sum_{x\in\mathcal{X}} p(x) \log \frac{\sum_{y:xey} p(y)}{p(x)},$$

Let $\mathcal{A}$ denote the collection of equivalence classes under e. Let P_e be the probability density on the elements A of $\mathcal{A}$ given by

$$p_e(A) = \sum_{x\in A} p(x),$$

Then we have the following lemma (see V. Anantharam [2] and J. Körner [19]).

A. 3. Lemma. (V. Anantharam). The number of P-typical n-sequences in a P_e-typical n-sequence of equivalence classes is bounded below by $2^{nH(P|P_e)-C\sqrt{n}}$ and bounded above by $2^{nH(P|P_e)+C\sqrt{n}}$ for some constant $C > 0$.

Proof. Let $\mathbf{A} = (A_1, \cdots, A_n)$ be a P_e-typical n-sequence. That is for each $A \in \mathcal{A}$

$$|N(A|\mathbf{A}) - np_e(A)| \leq K\sqrt{np_e(A)}, \tag{6.4}$$

Then for all $A \in \mathcal{A}$, we have

$$np(A) \leq \max(4K^2, 2N(A|\mathbf{A})), \tag{6.5}$$

The proof is as follows. Suppose $np(A) \geq 4K^2$. Then $np(A) \geq 2K\sqrt{np_e(A)}$ and therefore,

$$N(A|\mathbf{A}) \geq np_e(A) - K\sqrt{np_e(A)} \geq \frac{np_e(A)}{2},$$

Let $\mathbf{x} = (x_1, \cdots, x_n)$ be a P-typical n-sequence in $\mathbf{A}$, i.e.,

$$x_i \in A_i,\ 1 \leq i \leq n.$$

From P-typicality of $\mathbf{x}$, we have

$$|N(x|\mathbf{x}) - np(x)| \leq K\sqrt{np(x)}.$$

Now, we prove that for each $A \in \mathcal{A}$, the restriction of $\mathbf{x}$ to those co-ordinates having $A_i = A$ is $\left(\frac{p(x)}{p_e(A)} : x \in A\right)$-typical. For $x \in A$, we have

$$\begin{aligned}
\left|N(x|\mathbf{x}) - N(A|\mathbf{A})\frac{p(x)}{p_e(A)}\right| &\leq |N(x|\mathbf{x}) - np(x)| + \left|np(x) - N(A|\mathbf{A})\frac{p(x)}{p_e(A)}\right| \\
&\leq K\sqrt{np(x)} + \frac{p(x)}{p_e(A)}K\sqrt{np_e(A)} \\
&= K\left(\sqrt{\frac{p(x)}{p_e(A)}} + \frac{p(x)}{p_e(A)}\right)\sqrt{np_e(A)}.
\end{aligned}$$

Using (6.5), and noting $N(A|\mathbf{A}) \geq 1$ and $\frac{p(x)}{p_e(A)} \leq \sqrt{\frac{p(x)}{p_e(A)}}$, we get

$$\begin{aligned}
\left|N(x|\mathbf{x}) - N(A|\mathbf{A})\frac{p(x)}{p_e(A)}\right| &\leq K\left(\sqrt{\frac{p(x)}{p_e(A)}} + \frac{p(x)}{p_e(A)}\right)\sqrt{\max(4K^2, 2N(A|\mathbf{A}))} \\
&\leq K\left(\sqrt{\frac{p(x)}{p_e(A)}} + \frac{p(x)}{p_e(A)}\right)\sqrt{\max\left(\frac{4K^2}{N(A|\mathbf{A})}, 2\right)}.\sqrt{N(A|\mathbf{A})} \\
&\leq \max(2K^2, \sqrt{2}K)\left(\sqrt{\frac{p(x)}{p_e(A)}} + \frac{p(x)}{p_e(A)}\right)\sqrt{N(A|\mathbf{A})} \\
&\leq 2\max(2K^2, \sqrt{2}K)\sqrt{\frac{N(A|\mathbf{A})p(x)}{p_e(A)}}.
\end{aligned} \tag{6.6}$$

Now, letting $H(P|e = A)$ denote $\sum_{x\in A} \frac{p(x)}{p_e(A)} \log \frac{p_e(A)}{p(x)}$, we give the following lower and upper bounds on the number of P-typical n-sequences $\mathbf{x}$ in $\mathbf{A}$. Let $C > 0$ be some constant depending on K and $|\mathcal{X}|$ as in Lemma 6.2, then using Lemma 6.2 and (6.4) we get the following upper bound on the P-typical n-sequences $\mathbf{x}$ in $\mathbf{A}$

$$
\begin{aligned}
&\prod_{A\in\mathcal{A}} 2^{N(A|\mathbf{A})H(\frac{P}{P_e(A)})} + C\sqrt{N(A|\mathbf{A})} \\
&= 2^{n\sum_{A\in\mathcal{A}}\left(\frac{N(A|\mathbf{A})}{n}H(P|e=A)+C\sqrt{N(A|\mathbf{A})}\right)} \\
&\leq 2^{n\sum_{A\in\mathcal{A}}\left(p_e(A)+\frac{K}{n}\sqrt{np_e(A)}\right)H(P|e=A)+\sum_{A\in\mathcal{A}}C\sqrt{n}} \\
&\leq 2^{n\sum_{A\in\mathcal{A}}p_e(A)H(P|P_e)+K\sum_{A\in\mathcal{A}}\sqrt{np_e(A)}H(P|e=A)+C|\mathcal{X}|\sqrt{n}} \\
&\leq 2^{nH(P|P_e)+\sqrt{n}\left(C|\mathcal{X}|+K\sum_{A\in\mathcal{A}}\log|A|\right)} \\
&= 2^{nH(P|P_e)+\sqrt{n}(C|\mathcal{X}|+K|\mathcal{X}|)},
\end{aligned}
$$

Now, setting

$$C_1 = C|\mathcal{X}| + K|\mathcal{X}|,$$

Thus, the number of P-typical n-sequences $\mathbf{x}$ in $\mathbf{A}$ is upper bounded by

$$2^{nH(P|e)+C_1\sqrt{n}},$$

Similarly, the number of P-typical n-sequences $\mathbf{x}$ in $\mathbf{A}$ is lower bounded by

$$2^{nH(P|e)-C_1\sqrt{n}}.$$

□

Proof of Lemma 3.2.2.

Let $0 < \epsilon < 1$, and $M(n, \epsilon)$ denote

$$\min_{U\in T_\epsilon^{(n)}} \chi(G^{(n)}[U]),$$

for sufficiently large n. Let $\lambda > 0$ be a positive number. First, we show that

$$M(n, \epsilon) \geq 2^{(H'(G,P)-\lambda)}.$$

Consider $G^{(n)}[U]$ for some $U \in T_\epsilon^{(n)}$. Using Lemma A. 2, for any $\delta > 0$ there is a $K > 0$ such that for any sufficiently large n, we have

$$P\left(T^n(P)\right) \geq 1 - \delta.$$

First, note that

$$1-\delta-\epsilon \leq P\left(U \cap T^n(P)\right). \tag{6.7}$$

Now, we estimate the chromatic number of $G^{(n)}[U \cap T^n(P)]$. Let $\mathcal{S}^n$ denote the family of the maximal independent sets of $G^{(n)}$. Note that every colour class in a minimum colouring of graph can be enlarged to a maximal independent set. Thus,

$$P\left(U \cap T^n(P)\right) \leq \chi\left(G^{(n)}[U \cap T^n(P)]\right). \max_{\mathbf{S} \in \mathcal{S}^n} P\left(\mathbf{S} \cap T^n(P)\right), \tag{6.8}$$

Furthermore, we have

$$\max_{\mathbf{S} \in \mathcal{S}^n} P\left(\mathbf{S} \cap T^n(P)\right) \leq \max_{\mathbf{x} \in T^n(P)} p(\mathbf{x}). \max_{S \in \mathcal{S}^n} |\mathbf{S} \cap T^n(P)|. \tag{6.9}$$

It is worth mentioning that $|\mathbf{S} \cap T^n(P)|$ is the number of typical sequences contained in $\mathbf{S}$. Furthermore, note that $\mathbf{S}$ can be considered as an n-sequence of maximal independent sets taken from $\mathcal{S}$.

Let $N(y, R|\mathbf{x}, \mathbf{S})$ denote the number of occurrences of the pair (y, R) in the following double n-sequence

$$\begin{pmatrix} x_1 & x_2 & \cdots & x_n \\ S_1 & S_2 & \cdots & S_n \end{pmatrix}$$

In other words, $N(y, R|\mathbf{x}, \mathbf{S})$ is the number of occurrences of the letter y selected from the maximal independent set R in the n-sequence $\mathbf{x}$ taken from the maximal independent sequence $\mathbf{S}$. Similarly, $N(y|\mathbf{x})$ denotes the number of occurrences of the source letter y in the n-sequence $\mathbf{x}$.

Setting

$$q(y, R) = \frac{N(y, R|\mathbf{x}, \mathbf{S})}{N(y|\mathbf{x})}.p(y), \tag{6.10}$$

we have

$$\begin{aligned} |N(y, R|\mathbf{x}, \mathbf{S}) - nq(y, R)| &= \left|\frac{nq(y, R)}{np(y)}\right|.|N(y|\mathbf{x}) - np(y)| \\ \leq \left|\frac{q(y, R)}{p(y)}\right|.K\sqrt{np(y)} &= K\sqrt{n.\frac{q^2(y, R)}{p(y)}} \leq K\sqrt{nq(y, R)}, \end{aligned}$$

since $\mathbf{x}$ is a P-typical sequence. Let

$$a(R) = \sum_{y: y \in R} q(y, R). \tag{6.11}$$

Then

$$N(R|\mathbf{S}) - na(R) = \sum_{y\in R} N(y,R|\mathbf{x},\mathbf{S}) - nq(y,R),$$

And therefore using 6.11,

$$\begin{aligned} |N(R|\mathbf{S}) - na(R)| \quad &\leq \sum_{y\in\mathcal{X}} K\sqrt{nq(y,R)} \leq K_1\sqrt{n\sum_{y\in\mathcal{X}} q(y,R)} \\ &= K_1\sqrt{na(R)}. \end{aligned} \tag{6.12}$$

Now, we define an auxiliary graph Γ of G as follows. Letting S be a maximal independent set of G containing a vertex x of G, the vertex set of Γ consists of pairs (x,S). Furthermore, two vertices (x,S) and (y,R) are adjacent if and only if $S \neq R$. Let $K_2 > 0$ be some constant. Then, applying Lemma A.3 with the equivalence relation a which is

$$((x,S),(y,R)) \notin E(\Gamma),$$

and probability density Q for the graph $\overline{\Gamma}$, the number of Q-typical n-sequences in each a-typical equivalence class A which is a maximal independent set of G lies in the interval

$$\left[2^{nH(Q|a)-K_2\sqrt{n}}, 2^{nH(Q|a)+K_2\sqrt{n}}\right]. \tag{6.13}$$

Noting that every pair (y,R) may occur zero, one,$\cdots$, or n-times in the n-sequence $(\mathbf{x},\mathbf{S})$ and for a given y knowing $N(y,R|\mathbf{x},\mathbf{S})$ for all R uniquely determines $N(y|\mathbf{x})$, there are at most $(n+1)^{|V(\Gamma)|}$ different auxiliary densities of the type given by (6.10). Now we bound $\max_{\mathbf{S}\in\mathcal{S}^n} |\mathbf{S}\cap T^n(P)|$ as follows. Note that $\mathbf{S}\cap T^n(P)$ is the set of P-typical n-sequenences which are contained in a given maximal independent set $\mathbf{S}$ in $G^{(n)}$. Then letting $\mathcal{Q}$ be the feasible joint distribution for (X,S), for all $\mathbf{S}\in\mathcal{S}^n$ and all $Q\in\mathcal{Q}$, set

$$T^n(S,Q) := \{\mathbf{x} : \mathbf{x}\in\mathcal{X}^n, x_i\in S_i, (\mathbf{x},\mathbf{S}) \text{ is } Q\text{-typical.}\}$$

From (6.10), for all $\mathbf{S}\in\mathcal{S}^n$ and for all $\mathbf{x}$ in $|\mathbf{S}\cap T^n(P)|$ there is some $Q\in\mathcal{Q}$ such that $\mathbf{x}\in T^n(S,Q)$. Therefore, for all $\mathbf{S}\in\mathcal{S}^n$, we get

$$\begin{aligned} |\mathbf{S}\cap T^n(P)| \quad &\leq \quad |\bigcup_{Q\in\mathcal{Q}} T^n(S,Q)| \\ &\leq \quad \sum_{Q\in\mathcal{Q}} |T^n(S,Q)| \\ &\leq \quad |\mathcal{Q}| \max_{Q\in\mathcal{Q}} |T^n(S,Q)|, \end{aligned}$$

Then, using (6.13), we obtain

$$\max_{\mathbf{S}\in\mathcal{S}^n} |\mathbf{S}\cap T^n(P)| \leq (n+1)^{|V(\Gamma)|}.2^{n.\max_{Q'\in\mathcal{Q}} H(Q'|a)+K_2\sqrt{n}}. \tag{6.14}$$

Further,

$$\sum_{R:y\in R} q(y,R) = \frac{p(y)}{N(y|\mathbf{x})}.\sum_{R:y\in R} N(y,R|\mathbf{x},\mathbf{S}) = p(y).$$

From the Lemma A.2 part (ii), we get

$$\max_{\mathbf{x}\in T^n(P)} p(\mathbf{x}) \leq 2^{-(nH(P)-C\sqrt{n})}. \tag{6.15}$$

Thus, using the inequalities (6.7)-(6.9), (6.14) and (6.15) we have

$$\begin{aligned}(1-\lambda-\epsilon) \quad &\leq \chi\left(G^{(n)}[U\cap T^n(\mathbf{p})]\right)\\ &.\exp_2\left(n.\left(\max_{Q'\in\mathcal{Q}} H(Q'|a)-H(P)\right)+K_2\sqrt{n}+|V(\Gamma)|.\log_2(n+1)\right),\end{aligned}$$

And consequently,

$$\begin{aligned}\chi\left(G^{(n)}[U\cap T^n(P)]\right) \quad &\geq (1-\lambda-\epsilon) \\ &.\exp_2\left(n(H(P)-\max_{Q'\in\mathcal{Q}} H(Q'|a)-K_2\sqrt{n}-|V(\Gamma)|.\log_2(n+1))\right).\end{aligned} \tag{6.16}$$

Note that

$$H(P)-\max_{Q'\in\mathcal{Q}} H(Q'|a) = \min_{Q'\in\mathcal{Q}}\sum_{x,S} q'(x,S)\log_2\frac{q'(x,S)}{p(x).q'(S)} = \min_{Q'\in\mathcal{Q}} I(Q').$$

Now, considering

$$\chi\left(G^{(n)}[U]\right) \geq \chi\left(G^{(n)}[U\cap T^n(P)]\right),$$

and using (6.16), for every $U\in T_\epsilon^{(n)}$ we get

$$\chi(G^{(n)}[U]) \geq (1-\lambda-\epsilon).\exp_2\left(nH'(G,P)-K_2\sqrt{n}-|V(\Gamma)|.\log_2(n+1)\right).$$

Thus,

$$\frac{1}{n}\log_2\left(\min_{U\in T_\epsilon^n}\chi\left(G^{(n)}[U]\right)\right) \geq \frac{1}{n}\log_2(1-\lambda-\epsilon)+H'(G,P)-\frac{K_2}{\sqrt{n}}-\frac{|V(\Gamma)|}{n}\log_2(n+1),$$

Therefore, we get

$$\liminf_{n\to\infty} \frac{1}{n} \log_2 M(n,\epsilon) \geq H'(G,P). \tag{6.17}$$

Now we show that for every $0 < \epsilon < 1$ and $\delta > 0$ and sufficiently large n, there exists subgraphs $G^{(n)}[U]$ of $G^{(n)}$, for some $U \subseteq V(G^{(n)})$, such that

$$\chi\left(G^{(n)}[U]\right) \leq 2^{n(H'(G,P)+\delta)}.$$

Let Q^* be the joint density on vertices and independent sets of G which minimizes the mutual information $I(Q^*)$. That is

$$I(Q^*) = H'(G,P).$$

Then the probability of every maximal independent set S is

$$Q^*(S) = \sum_{y:y\in S} Q^*(y,S).$$

Letting $\mathbf{S}$ be $\mathbf{S} = (S_1, S_2, \cdots, S_n) \in \mathcal{S}^n$, we have

$$\mathbf{Q}^*(\mathbf{S}) = \prod_{i=1}^{n} Q^*(S_i),$$

Let L be a fixed parameter. For a family of L maximal independent sets, not necessarily distinct and not necessarily covering, we define the corresponding probability density $\mathbf{Q}^*_L$ as follows. We assume that the L maximal independent sets of a given system of maximal independent sets are chosen independently. Thus,

$$Q^*_L(\mathbf{S}_1, \mathbf{S}_2, \cdots, \mathbf{S}_L) = \prod_{j=1}^{L} Q^*(\mathbf{S}_j).$$

Now consider a fixed n. Let $G^{(n)}$ be the n-th conormal power graph of graph G. Consider systems of maximal independent sets consisting of L maximal independent sets each in the form of a n-sequence of maximal independent sets. We call this system of maximal independent sets an L-system.

For each L-system $(\mathbf{S}_1, \mathbf{S}_2, \cdots, \mathbf{S}_L)$ let $U(\mathbf{S}_1, \mathbf{S}_2, \cdots, \mathbf{S}_L)$ be the union of all vertices of $V(G^{(n)})$ which are not covered by the L-system $(\mathbf{S}_1, \mathbf{S}_2, \cdots, \mathbf{S}_L)$. For a given L, we show that the expected value of $P(U(\mathbf{S}_1, \mathbf{S}_2, \cdots, \mathbf{S}_L))$ is less than ϵ. This implies that there exists at least one system $\mathbf{S}_1, \cdots, \mathbf{S}_L$ covering a subgraph of $G^{(n)}$ with probability greater than or equal to $1-\epsilon$. For an L-system chosen with probability Q_L^*, let $Q_{L,\mathbf{x}}^*$ be the probability that a given n-sequence $\mathbf{x}$ is not covered by an L-system, that is

$$\begin{aligned} Q_{L,\mathbf{x}}^* &= Q_L^*\left(\left\{\left(\mathbf{S}_1,\cdots,\mathbf{S}_L\right):\mathbf{x}\in U\left(\mathbf{S}_1,\cdots,\mathbf{S}_L\right)\right\}\right) \\ &= \sum_{\left(\mathbf{S}_1,\cdots,\mathbf{S}_L\right)\ni\mathbf{x}} Q_L^*\left(\mathbf{S}_1,\cdots,\mathbf{S}_L\right). \end{aligned}$$

Then we have

$$\begin{aligned} E\left(P\left(U\left(\mathbf{S}_1,\cdots,\mathbf{S}_L\right)\right)\right) &= \sum_{\mathbf{S}_1,\cdots,\mathbf{S}_L} Q_L^*\left(\mathbf{S}_1,\cdots,\mathbf{S}_L\right).P\left(U(\mathbf{S}_1,\cdots,\mathbf{S}_L)\right) \\ &= \sum_{\left(\mathbf{S}_1,\cdots,\mathbf{S}_L\right)} Q_L^*\left(\mathbf{S}_1,\cdots,\mathbf{S}_L\right)\left(\sum_{\mathbf{x}\in U\left(\mathbf{S}_1,\cdots,\mathbf{S}_L\right)} P(\mathbf{x})\right) \\ &= \sum_{\mathbf{x}\in\mathcal{X}^n} P(\mathbf{x})\left(\sum_{U\left(\mathbf{S}_1,\cdots,\mathbf{S}_L\right)\ni\mathbf{x}} Q_L^*\left(\mathbf{S}_1,\cdots,\mathbf{S}_L\right)\right) \\ &= \sum_{\mathbf{x}\in\mathcal{X}^n} P(\mathbf{x}).Q_{L,\mathbf{x}}^*. \end{aligned} \tag{6.18}$$

For a given ϵ with $0<\epsilon<1$, by Lemma A. 2 there exists a set of typical sequences with total probability greater than or equal to $1-\frac{\epsilon}{2}$. Then we can write the right hand of the above equation as

$$\begin{aligned} &\sum_{\mathbf{x}\in\mathcal{X}^n} P(\mathbf{x}).Q_{L,\mathbf{x}}^* \\ &= \sum_{\mathbf{x}\in T^n(P)} P(\mathbf{x}).Q_{L,\mathbf{x}}^* \\ &+ \sum_{\mathbf{x}\in\overline{T^n(P)}} P(\mathbf{x}).Q_{L,\mathbf{x}}^*. \end{aligned} \tag{6.19}$$

The second term in (6.19) is upper-bounded by $P\left(\overline{T^n(P)}\right)$ which is less than $\frac{\epsilon}{2}$. We give an upper bound for the first term and show that for $L=2^{n(H'(G,P)+\delta)}$ it tends to 0 as $n\to\infty$. Now

$$\begin{aligned} &\sum_{\mathbf{x}\in T^n(P)} P(\mathbf{x}).Q_{L,\mathbf{x}}^* \leq \\ &P\left(T^n(P)\right).\max_{\mathbf{x}\in T^n(P)} Q_{L,\mathbf{x}}^* \leq \\ &\max_{\mathbf{x}\in T^n(P)} Q_{L,\mathbf{x}}^*. \end{aligned}$$

If an n-sequence $\mathbf{x}$ is not covered by an L-system, then $\mathbf{x}$ is not covered by any element of this system. Letting $\mathcal{S}_{\mathbf{x}}$ be the set of maximal independent sets covering the n-sequence $\mathbf{x}$, we have

$$\max_{\mathbf{x}\in T^n(P)} Q^*_{L,\mathbf{x}} = \max_{\mathbf{x}\in T^n(P)} \left(1 - Q^*(\mathcal{S}_{\mathbf{x}})\right)^L . \tag{6.20}$$

We obtain a lower bound for $Q^*(\mathcal{S}_{\mathbf{x}})$ by counting the Q^*-typical n-sequences of maximal independent sets covering $\mathbf{x} \in T^n(P)$. This number is greater than or equal to the Q^*-typical sequences $(\mathbf{y}, \mathbf{B})$ with the first coordinate equal to $\mathbf{x}$. The equality of the first coordinate of the ordered pairs in $V(\Gamma)$ is an equivalence relation p on the set $V(\Gamma)$. Thus, using Lemma A. 3, the number of the Q^*-typical n-sequences of maximal independent sets is bounded from below by

$$2^{nH(Q^*|q)-K_3\sqrt{n}}, \tag{6.21}$$

Let K_4 be a constant independent of n and the density $a(Q^*)$. Then, applying Lemma A.2 to $\mathcal{S}$ and the marginal distribution $a(Q^*)$ of Q^* over the maximal independent sets, we obtain the following lower bound on the probability Q^* of the $a(Q^*)$-typical n-sequences of maximal independent sets,

$$Q^* \geq 2^{-\left(nH(a(Q^*))+K_4\sqrt{n}\right)}. \tag{6.22}$$

Combining (6.20),(6.21), and (6.22), we get

$$\begin{aligned}&\max_{\mathbf{x}\in T^n(P)} Q^*_{L,\mathbf{x}} \leq \\ &\left(1 - \exp_2(-(nH(a(Q^*)) + K_4\sqrt{n}) + nH\left(Q^*|p\right) - K_3\sqrt{n})\right)^L .\end{aligned} \tag{6.23}$$

Note that using (2.6) we have

$$H'\left(G, P\right) = I(Q^*) = H\left(a(Q^*) - H(Q^*|p)\right),$$

Therefore,

$$\max_{\mathbf{x}\in T^n(P)} Q^*_{L,\mathbf{x}} \leq \left(1 - 2^{-(nH'(G,P)+K_5\sqrt{n})}\right)^L, \qquad \text{for some constant } K_5.$$

Then, using the inequality $(1-x)^L \leq \exp_2(-Lx)$, the above inequality becomes

$$\max_{\mathbf{x}\in T^n(P)} Q^*_{L,\mathbf{x}} \leq \exp_2\left(-L.2^{-(nH'(G,P)+K_5\sqrt{n})}\right). \tag{6.24}$$

Setting $L = 2^{(nH'(G,P)+\delta)}$, (6.24) becomes

$$\begin{aligned}\max_{\mathbf{x}\in T^n(P)} Q^*_{L,\mathbf{x}} &\leq \exp_2\left(-(2^{nH'(G,P)+\delta}-1).2^{-(nH'(G,P)+K_5\sqrt{n})}\right)\\ &\leq \exp_2\left(-2^{n\delta-K_6\sqrt{n}}\right). \qquad (6.25)\end{aligned}$$

Substituting (6.19) into (6.25), we get

$$\sum_{\mathbf{x}\in\mathcal{X}^n} P(\mathbf{x}).Q^*_{L,\mathbf{x}} \leq \exp_2\left(-2^{n\delta-K_6\sqrt{n}}\right)+\frac{\epsilon}{2},$$

for $L = 2^{(nH'(G,P)+\delta)}$. For sufficiently large n the term $\exp_2\left(-2^{n\delta-K_6\sqrt{n}}\right)$ tends to zero, and (6.18) implies

$$\sum_{\mathbf{S}_1,\cdots,\mathbf{S}_L} Q^*_L\left(\mathbf{S}_1,\cdots,\mathbf{S}_L\right).P\left(U(\mathbf{S}_1,\cdots,\mathbf{S}_L)\right)\leq\epsilon.$$

Thus, we conclude that for every $0<\epsilon<1$ and $\delta>0$, there exists a $\left(2^{n(H'(G,P)+\delta)}\right)$-system covering a subgraph $G^{(n)}[U]$ of $G^{(n)}$ with probability of U at least $1-\epsilon$. Now, from Lemma A.1, the chromatic number of a graph is equal to the minimum number of maximal independent sets covering the graph. Therefore, for every $\delta>0$ there exists a subgraph $G^{(n)}[U]$ of $G^{(n)}$ with $U\in T^{(n)}_\epsilon$ such that

$$\chi\left(G^{(n)}[U]\right)\leq 2^{n(H'(G,P)+\delta)},$$

Consequently,

$$\min_{U\subset V(G^{(n)}),U\in T^n_\epsilon}\chi\left(G^{(n)}[U]\right)\leq 2^{n(H'(G,P)+\delta)},\quad \text{for every } \delta>0,$$

Then, using the definition of $M(n,\epsilon)$, we get

$$\frac{1}{n}\log_2 M(n,\epsilon)\leq H'(G,P)+\delta,\quad \text{for every } \delta>0.$$

And consequently, we get

$$\limsup_{n\to\infty}\frac{1}{n}\log_2 M(n,\epsilon)\leq H'(G,P). \qquad (6.26)$$

Comparing (6.17) and (6.26), we obtain

$$\lim_{n\to\infty}\frac{1}{n}\log_2 M(n,\epsilon)=H'(G,P),\quad \text{for every } \epsilon \text{ with } 0<\epsilon<1.$$

□

Bibliography

[1] Noga Alon, and Alon Orlitsky, "Source Coding and Graph Entropies," *IEEE Trans. on Information Theory*, Vol. 42, No. 5, pp. 1329-1339, September 1996.

[2] V. Anantharam, "Error Exponent in a Source Coding Problem of Körner", *available at* www.eecs.berkeley.edu/ ananth/1993-1995/Self/Korner.pdf

[3] C. Berge, Graphs and hypergraphs, 2nd edition, North Holland, Amesterdam, 1976.

[4] Stephen Boyd, and Lieven Vanderberghe, Convex Optimization, Published in the United States of America by Cambridge University Press, New York, First published in 2004, Seventh printing with corrections in 2009.

[5] Jean Cardinal, Samuel Fiorini, and G. Van Assche, "On minimum entropy graph colorings," In *Proc. of IEEE Int. Symposium on Information Theory*, p 43, 2004.

[6] Jean Cardinal, Samuel Fiorini, and Gwenael Joret, "Minimum Entropy Coloring," *IEEE Trans. on Information Theory*, Vol. 42, No. 5, pp. 1329-1339, September 2008.

[7] V. Chvátal, "On certain polytopes associated with graphs", *J. Comb. Theory B*, **18**, (1975), 138-154.

[8] Thomas M. Cover, and Joy A. Thomas, Elements of Information Theory, Second Edition 2006, A Wiley-Interscience publication.

[9] I. Csiszár, J. Körner, L. Lovás, K. Marton, and G. Simonyi, "Entropy Splitting for antiblocking corners and perfect graphs," *Combinatorica*, Vol. 10, pp. 27-40, 1990.

[10] R. M. Dudley, Real Analysis and Probability, First published by Wadsworth, Inc. in 1989, Cambridge University Press Edition published in 2002.

[11] Rick Durrett, Probability: Theory and Example, Fourth Edition published 2010 by Cambridge University Press.

[12] P. Erdös, S. T. Hedetniemi, R. C. Laskar, and G. C. E Prins, "On the equality of the partial Grundy and Upper Ochromatic Numbers of Graphs," *Discrete Math* 272(1): 53-64, in honor of Frank Haray, 2003.

[13] Louis Esperet, František Kardoš, and Andrew D. King, Daniel Král, and Serguei Norine, "Exponentially many perfect matchings in cubic graphs", Advances in Mathematics 227 (4): 1646-1664, 2011.

[14] M. Fredman, and J. Komlós, "On the size of separating systems and perfect hash functions", SIAM J. Algebraic Discrete Methods, 5 (1984), pp. 61-68.

[15] C. De Simone and J. Körner, "On the odd cycles of normal graphs", Proceedings of the Third International Conference on Graphs and Optimization, GO-III (Leukerbad, 1998), *Discrete Appl. Math.* **94** (1999), no. 1 - 3, 161-169.

[16] D. R. Fulkerson, "Blocking and anti-blocking of polyhedra", *Math. Programming* **1**, (1971), 168-194.

[17] Chris Godsil, and Gordon Royle, Algebraic Graph Theory, Graduate Texts in Mathematics-Springer, 2001.

[18] J. Kahn and J. H. Kim, "Entropy and Sorting", *Proc. 24th Annual ACM Symposium on the Theory of Computing*, pp. 178-187.

[19] J. Körner, "Coding of an information source having ambiguous alphabet and the entropy of graphs," in *Transactions of the 6th Prague Conference on Information Theory*, Academia, Prague, (1973), 411-425.

[20] J. Körner, "An extension of the class of perfect graphs", *Studia Sci. Math. Hung.*, **8** (1973), 405-409.

[21] J. Körner, "Fredman-Komlós Bounds and Information Theory," in *SIAM J. Algebraic and Discrete Methods*, **7** (1986), 560-570.

[22] J. Körner and G. Longo, "Two-step encoding of finite memoryless sources", *IEEE Trans. Information Theory*, **19** (1973), 778-782.

[23] J. Körner, G. Simonyi, and Zs. Tuza, "Perfect couples of graphs," in *Combinatorica*, **12** (1992), 179-192.

[24] J. Körner, and K. Marton, "Graphs that split entropies", *SIAM J. Discrete Math.*, **1** (1988), 71-79.

[25] J. Körner, and K. Marton, "New bounds for perfect hashing via information theory", *European J. of Combinatorics*, **9** (1988), 523-530.

[26] L. Lovász, "Perfect graphs", in *More Selected Topics in Graph Theory*, (L. W. Beineke, R. J. Wilson, eds), Academic Press, New York-London, 1983, 55-87.

[27] L. Lovász, "Normal hypergraphs and the perfect graph conjecture", in *Discrete Mathematics* **2** (3), 1972a: 253267.

[28] L. Lovász,"A characterization of perfect graphs", in *Journal of Combinatorial Theory, Series B* **13** (2), 1972b: 9598.

[29] L. Lovász, "On the Shannon Capacity of a Graph", in *IEEE Trans. on Information Theory*, **25** (1979), pp. 1-7.

[30] L. Lovász and M. D. Plummer, Matching Theory, North-Holland Publishing Co., Amsterdam, 1986.

[31] Katalin Marton, "On the Shannon Capacity of Probabilistic Graphs", in *Journal of Combinatorial Theory Series B*, **57**, pp. 183-195 (1993).

[32] Zsolt Patakfalvi, "Line-graphs of cubic graphs are normal", in *Discrete Mathematics* **308** (2008), pp. 2351-2365.

[33] J. Radhakrishnan,"Entropy and counting", in the book Computational Mathematics, Modelling and Algorithms 2001 (Ed. J.C. Misra). Narosa Publishers, New Delhi.

[34] Alexander Schrijver, Combinatorial Optimization, Springer-Verlag Berlin Heidelberge 2003.

[35] G. Simonyi, "Graph Entropy: A Survey," in *Combinatorial Optimization*, W. Cook, L. Lovas, and P.Seymour ed., DIMACS Series in Discrete Mathematics and Computer Science, Vol. 20, AMS, 399-441.

[36] G. Simonyi, "Perfect graphs and graph entropy. An updated survey," Perfect Graphs, John Wiley and Sons (2001) pp. 293-328.

[37] L. E. Trotter, jr., "A Class of Facet Producing Graphs for Vertex Packing Polyhedra", in *Discrete Math.* **12** (1975) pp. 373-388.

[38] Daniel Ullman and Edward Scheinerman, Fractional Graph Theory : A Rational Approach to the Theory of Graphs, Wiley-Interscience; 1 edition (August 25, 1997).

[39] Annegret K. Wagler, "The Normal Graph Conjecture is true for Circulants", in *Graph Theory in Paris Trends in Mathematics* 2007, pp. 365-374.

[40] Douglas B. West, Introduction to Graph Theory, Second Edition, Prentice-Hall, Inc., 2001.

Index