A Study on Compressing Graphical Structures

Basak Guler   Aylin Yener
The Pennsylvania State University
University Park, PA
basak@psu.edu   yener@ee.psu.edu

Prithwish Basu   Carl Andersen
Raytheon BBN Technologies
Cambridge, MA,   Arlington, VA
pbasu@bbn.com   canderse@bbn.com

Ananthram Swami
Army Research Laboratory
Adelphi, MD
a.swami@ieee.org

Abstract—**Many real-world systems such as the WWW, Semantic Web, logical expressions, and social networks can be represented in graphical forms. This paper studies compressing graphical structures from a graph entropy point of view. We show that structural entropy is equal to the chromatic entropy of the characteristic graph, termed the structural characteristic graph. We establish the relation between structural entropy and graph entropy, and investigate the cases when (conditional) graph entropy is equal to (conditional) chromatic entropy.**

I. Introduction

Graph compression is becoming increasingly important in a variety of fields including semantic networks, knowledge representation, logic programming, and social media analyses. In effect, many real-world applications can be represented in graphical forms, including the world wide web, logical expressions, and semantic graphs with RDF descriptions.

The need for efficient representations of graphical structures arises in many different contexts and problems. Shannon identified a graph invariant to study the zero-error capacity of a communication channel in [1], which is upper bounded by the Lovász number [2]. This measures the independence number of a higher-order strong-product of a characteristic graph. Graph entropy has been introduced in [3] to study lossless compression limits for sources with ambiguous alphabets. Since then, it has been applied to areas such as sorting with partial information [4], and bounding the size of Boolean formulae [5], [6]. Conditional graph entropy has emerged as a functional compression counterpart of graph entropy in [7], in which the decoder computes a function of the source symbol and the side information available to the receiver. Various entropy metrics to study graph properties have been studied in [8], [9].

Chromatic entropy is defined in [10] to investigate compression bounds for communicating single and multiple input descriptions with zero error in the presence of side information. Chromatic entropy requires one to determine the minimum entropy coloring of a given characteristic graph. Finding the minimum entropy coloring of a graph is known to be NP-hard in the general case [11]. On the other hand, in a family of graphs studied in [12], minimum entropy coloring can be determined in polynomial time. Graph entropy serves as a lower bound to chromatic entropy [10]. The difference between chromatic and graph entropies can be arbitrarily large [11].

We consider compressing random graphs up to an isomorphism. We define the structural characteristic graph in which each node represents a distinct labeled graph, and two nodes are connected if and only if they correspond to different graph structures. We show that structural entropy is equal to the chromatic entropy of this graph. We show that for graphs with non-overlapping fully connected maximal independent sets, chromatic entropy is equal to graph entropy. Accordingly, coloring a higher-order graph provides the same gains as coloring the first-order graph and then extending the color alphabet.

II. System Model

Consider a random graph $\mathcal{G}$ that entails a probability distribution over graphs with κ vertices. Define the set $\mathcal{G} = \{G_1, G_2, \ldots, G_m\}$ of all distinct labeled graphs, where $m = 2^{\binom{\kappa}{2}}$. Among these m graphs, the probability of graph indexed by i, i.e., G_i, is given by $p(G_i)$. We assume that $p(G_i) > 0$ for $i = 1, \ldots, m$, i.e., each graph has a strictly positive probability.

Definition 1. *(Graph Isomorphism) [13] Two graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ are isomorphic if there exist bijections $f : V_1 \to V_2$ and $h : E_1 \to E_2$ that preserve the adjacency and non-adjacency relations. For every edge $e = (u, v) \in E_1$ such that $u, v \in V_1$, an edge $(f(u), f(v)) = g(e) \in E_2$ exists where $f(u), f(v) \in V_2$. Accordingly, $(u, v) \notin E_1 \to (f(u), f(v)) \notin E_2$ for $u, v \in V_1$. Isomorphic graphs are represented by $G_1 \cong G_2$.*

Lemma 1. *[13] Define $G_1 = (V_1, E_1)$, $G_2 = (V_2, E_2)$ and $G_3 = (V_3, E_3)$ such that $G_1 \cong G_2$ and $G_2 \cong G_3$. Then, $G_1 \cong G_3$.*

Denote an *unlabeled* variant of $\mathcal{G}$ by $\mathcal{S}$, for which the vertices are no longer distinct. $\mathcal{S}$ is called the random structure model of $\mathcal{G}$. Two labeled graphs G_i and G_j with $i \neq j$ have the same structure if their unlabeled versions refer to the same (unlabeled) graph. The probability of a structure S is

$$P(S) = \sum_{\substack{G_i \in \mathcal{G}, G_i \cong S \\ i=1,\ldots,m}} P(G_i). \tag{1}$$

Structural entropy is then defined as $H_S = -\sum_{S \in \mathcal{S}} P(S) \log P(S)$ with $P(S)$ given in (1), and where $\mathcal{S}$ is the set of all distinct structures [14], [15].

III. The Relation Between Structural and Chromatic Entropies

We first introduce a special characteristic graph, [1], [16], to study structural properties.

Definition 2. *(Structural Characteristic Graph) Define a graph $\Gamma = (V, E)$ in which vertex $i \in V$ represents graph $G_i \in \mathcal{G}$ where $i = 1, \ldots, m$. Assign each vertex a probability $p(i) = p(G_i)$. We have edge $(i, j) \in E$ if and only if there exists no structure $S \in \mathcal{S}$ such that $G_i \cong S$ and $G_j \cong S$.*

Each independent set of Γ represents a distinct unlabeled graph, which corresponds to labeled graphs with an identical structure.

Lemma 2. *The maximal independent sets of Γ do not overlap and are fully connected* [2].

Proof. We show by contradiction that maximal independent sets of Γ cannot overlap, using [12]. Define $\mathcal{W}(\Gamma)$ as the set of all maximal independent sets of Γ. Let $w_1, w_2 \in \mathcal{W}(\Gamma)$ denote distinct maximal independent sets with a common element v in both sets. Let u and

---

[2]Our fully connected definition states that every node in a maximal independent set is connected to all the other maximal independent sets.

z be a pair of vertices so that $u \in w_1$ and $z \in w_2$ with an edge in between. Note that since w_1 and w_2 correspond to *distinct*, *maximal* independent sets, we can always find at least one such pair. Suppose $w_1 \cap w_2 \neq \emptyset$, we can always find a node v such that $v \in w_1 \cap w_2$. Let $G_u, G_v, G_z \in \mathcal{G}$ be the graphs represented by nodes $u, v, z \in V$, respectively.

As both u and v are elements of w_1, there exists no edge between the two vertices. From Definition 2, $G_u \cong G_v \cong S$ for some $S \in \mathcal{S}$, i.e., u and v correspond to graphs with the same structure. Similarly, both v and z are in w_2, hence no edge exists between the two in Γ. As a result, v and z represent graphs with the same structure, or $G_v \cong G_z \cong S'$ for some $S' \in \mathcal{S}$. But from Lemma 1, $G_u \cong G_v$ and $G_v \cong G_z$ means that $G_u \cong G_z$, hence S' is an automorphism to S, in other words, the two refer to the same unlabeled graph (structure). Thus, no edge should exists between u and z according to Definition 2. On the other hand, by the choice of the pair u and z described above, there exists an edge between the two vertices. This in turn proves the first part of our statement.

Next, by contradiction, we show that maximal independent sets are fully connected. Consider two distinct *non-overlapping* maximal independent sets w_1 and w_2. Assume that there exists a node $z \in w_2$ connected to $u \in w_1$ but is not connected to $v \in w_1$. Let G_z, G_u, G_v represent the graphs for nodes z, u, v, respectively. Since $u, v \in w_1$, they have the same structure, i.e., $G_u \cong G_v \cong S$ for some $S \in \mathcal{S}$. As z is not connected to v, G_z and G_v have the same structure, i.e., $G_z \cong G_v \cong S'$ for some $S' \in \mathcal{S}$. From Lemma 1, $G_u \cong G_z$, or $S \cong S'$. However, as z is connected to u, the two cannot have the same structure, leading to a contradiction. □

Corollary 1. *Consider a characteristic graph with non-overlapping fully connected maximal independent sets. Then, each vertex can belong to a single maximal independent set.*

Each labeled graph can belong to a single structure. Let X be a random variable over V with probabilities of the labeled graphs in $\mathcal{G}$. Define $c(\cdot)$ as a coloring function $c : V \to \mathcal{C}$. The entropy of $c(\cdot)$ is

$$H(c(X)) = \sum_{\lambda \in c(\Gamma)} p(c^{-1}(\lambda)) \log \frac{1}{p(c^{-1}(\lambda))} \tag{2}$$

where $c^{-1}(\lambda)$ is the set of vertices assigned to λ. The probability of a set is the sum of the probabilities of its elements [10]. The chromatic entropy of Γ,

$$H_\Gamma^\chi(X) = \min_{c \text{ is a coloring of } \Gamma} H(c(X)), \tag{3}$$

is achieved through the coloring with the minimum entropy [10].

Lemma 3. *Minimum entropy coloring of Γ is obtained by assigning a distinct color to each maximal independent set.*

Proof. Since the maximal independent sets of Γ are non-overlapping and fully connected, a color assigned to a vertex in one maximal independent set cannot be re-used for another vertex from a different maximal independent set, hence the colors used for different maximal independent sets should be distinct. Let w be a maximal independent set colored with a single color λ. Then, the impact of w on the entropy in (2) is given by

$$-p(c^{-1}(\lambda)) \log p(c^{-1}(\lambda)) = -\sum_{v \in w} p(v) \log \left(\sum_{v \in w} p(v) \right). \tag{4}$$

Consider any non-empty k-partitioning of w such that $w = w_1 \cup w_2 \ldots \cup w_k$. Assume that the vertices in $w_1, w_2, \ldots, w_k$ are colored with colors $\lambda_1, \lambda_2, \ldots, \lambda_k$, respectively. Then, the impact of the

vertices in w on the entropy (2) is

$$-\sum_{i=1}^{k} p(c^{-1}(\lambda_i)) \log p(c^{-1}(\lambda_i)) = -\sum_{i=1}^{k} \sum_{v \in w_i} p(v) \log \left(\sum_{v \in w_i} p(v) \right) \tag{5}$$

It follows from (4) and (5) that

$$-\sum_{\substack{v \in w \\ w = w_1 \cup \ldots \cup w_k}} p(v) \log \left(\sum_{v \in w} p(v) \right) = -\sum_{i=1}^{k} \sum_{v \in w_i} p(v) \log \left(\sum_{v \in w} p(v) \right) \tag{6}$$

$$\leq -\sum_{i=1}^{k} \sum_{v \in w_i} p(v) \log \left(\sum_{v \in w_i} p(v) \right) \tag{7}$$

where in the latter we used the fact that $w_1, w_2, \ldots, w_k \subset w$ and

$$\sum_{v \in w} p(v) > \sum_{v \in w_i} p(v) \tag{8}$$

where

$$0 \leq \sum_{v \in w} p(v), \sum_{v \in w_i} p(v) \leq 1 \tag{9}$$

for $i = 1, \ldots, k$. Therefore, minimum entropy is achieved by using a unique color for each maximal independent set. □

Theorem 1. *The structural entropy of $\mathcal{G}$ is equal to the chromatic entropy of Γ*

$$H_S = H_\Gamma^\chi(X). \tag{10}$$

Proof. From Lemma 3, chromatic entropy is achieved by assigning a different color to each of the maximal independent sets. As each maximal independent set represents a unique graph structure, we can define a bijection from the colors in $\mathcal{C}$ to the structures in $\mathcal{S}$. Assume that color $\lambda \in \mathcal{C}$ corresponds to structure $S \in \mathcal{S}$. Then,

$$p(\lambda) = p(c^{-1}(\lambda)) = \sum_{\substack{c(i) = \lambda \\ i \in V}} p(i) = \sum_{G_i \cong S} p(G_i) \tag{11}$$

and

$$p(S) = p(\lambda) = \sum_{G_i \cong S} p(G_i). \tag{12}$$

Then (10) follows from comparing H_S and (2) by using (12) and (11). From Lemma 3, we know that this assignment achieves (3). □

Graph entropy [3] is a lower bound for the lossless compression of graph structures

$$H_\Gamma(X) = \min_{X \in W \in \mathcal{W}(\Gamma)} I(W; X) \tag{13}$$

W is a random variable defined over $\mathcal{W}(\Gamma)$. $X \in W$ indicates that the joint distribution $q(w, x)$ on $\mathcal{W}(\Gamma) \times \mathcal{X}$ such that $q(w, x) > 0$ implies $x \in w$.

Lemma 4. *[10] Let Γ^n denote the nth-order OR-product of Γ. Then,*

$$\lim_{n \to \infty} \frac{1}{n} H_{\Gamma^n}^\chi(X) = H_\Gamma(X). \tag{14}$$

where the OR-product $\Gamma^n = (V^n, E^n)$ has a vertex set V^n. Consider two vertices $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{x}' = (x_1', \ldots, x_n')$ such that $\mathbf{x}, \mathbf{x}' \in V^n$. An edge $(\mathbf{x}, \mathbf{x}') \in E^n$ is defined if and only if there exists an edge (x_i, x_i') such that $x_i, x_i' \in V$ for some $i = 1, \ldots, n$.

IV. CASES WHERE GRAPH ENTROPY IS EQUAL TO CHROMATIC ENTROPY

In this section, we establish the relation between chromatic entropy and graph entropy and study the cases when the two are equal.

Theorem 2. *Consider a characteristic graph G_X with a random variable $X \sim p(x)$ defined over a set of vertices $\mathcal{X}$. If the maximal independent sets of G_X are non-overlapping and fully connected, its chromatic entropy is equal to its graph entropy. Hence, coloring G_X can achieve the same compression performance as coloring its nth-order OR-product counterpart G_X^n.*

Proof. We first consider from [3]

$$I(W; X) = \sum_{x \in \mathcal{X}} \sum_{w:x \in w \in \mathcal{W}(G_X)} q(x, w) \log \frac{q(x, w)}{p(x)q(w)} \quad (15)$$

where $q(x, w) > 0$ implies that $x \in w$ such that

$$\sum_{w:x \in w} q(x, w) = p(x). \quad (16)$$

The marginal distribution of w is

$$q(w) = \sum_{x:x \in w} q(x, w). \quad (17)$$

We know from Lemma 1 that, each vertex in a characteristic graph with non-overlapping fully connected maximal independent sets can belong to a single maximal independent set. Let x belong to $w^* \in \mathcal{W}(G_X)$. Then, $\{w : x \in w \in \mathcal{W}(G_X)\} = w^*$. It follows from (16) that

$$\sum_{w:x \in w} q(x, w) = q(x, w^*) = p(x) \quad (18)$$

whereas the marginal distribution of w^* is

$$q(w^*) = \sum_{x:x \in w^*} q(x, w^*) = \sum_{x:x \in w^*} p(x). \quad (19)$$

From Lemma 2, maximal independent sets define a partition of the vertices in $\mathcal{X}$. Moreover, the random variable representing each maximal independent set has the joint and marginal probability distributions (18) and (19). Then (15) can be written as

$$I(W; X) = -\sum_{x \in \mathcal{X}} \sum_{w:x \in w \in \mathcal{W}(G_X)} q(x, w) \log \frac{p(x)q(w)}{q(x, w)} \quad (20)$$

$$= -\sum_{x \in \mathcal{X}} p(x) \log \frac{p(x)(\sum_{x':x' \in w} p(x'))}{p(x)} \quad (21)$$

$$= -\sum_{w \in \mathcal{W}(G_X)} \left(\sum_{x:x \in w} p(x) \right) \log \left(\sum_{x':x' \in w} p(x') \right) \quad (22)$$

$$= -\sum_{\lambda \in \mathcal{C}} p(\lambda) \log p(\lambda) \quad (23)$$

where (22) since there is only a single w per x, and (22) follows from the fact that $\{x : x \in w\}$ is a disjoint set for each $w \in \mathcal{W}(G_X)$. In (23), we assigned a distinct color to every maximal independent set in (22), with each color having a probability equal to the sum of the vertex probabilities of the maximal independent set. Let λ be the color assigned to w^*. Then,

$$p(\lambda) = \sum_{x:c(x)=\lambda} p(x) = \sum_{x:x \in w^*} p(x) \quad (24)$$

where $|\mathcal{C}| = |\mathcal{W}(G_X)|$. We know from Lemma 3 that assigning a distinct color to each maximal independent set, i.e., (23), achieves the chromatic entropy in (3). Since each vertex belongs to a single maximal independent set, it can be observed that (22) is equal to (13). Therefore, chromatic entropy is equal to graph entropy. $\square$

Graphs with non-overlapping fully connected maximal independent sets are complete k-partite graphs where k is the number of maximal independent sets. Graph entropy of a complete k-partite graph is

known to be equal to the graph entropy of a complete graph on k vertices, in which each vertex represents a maximal independent set [17]. As the new graph is complete, its graph entropy is equal to the entropy of a random variable whose distribution consists of probabilities equal to those of the new vertices. Because graph entropy is achieved by assigning a distinct color to each maximal independent set, for complete k-partite graphs, graph entropy is equal to chromatic entropy. Hence, minimum entropy coloring can be done over the first-order graph instead of the product graph.

Theorem 3. *For the lossless compression of graphical structures, graph entropy is equal to chromatic entropy and structural entropy.*

Proof. The proof follows the lines in Theorem 2 by letting G_X denote Γ. Note that each vertex in Γ represents a labeled graph. The receiver wants to determine the structure of x, which corresponds to vertices that need to be distinguished from one another. $\square$

V. CONDITIONAL GRAPH AND CHROMATIC ENTROPIES OF COMPLETE MULTIPARTITE CHARACTERISTIC GRAPHS

This section considers the characteristic graphs for functional compression, and studies the relationship between conditional graph entropy and conditional chromatic entropy. Consider $X, Y \sim p(x, y)$ and a function $f(x, y)$ such that $x \in \mathcal{X}$ and $y \in \mathcal{Y}$. Define $G_X = (V, E)$ with a vertex set $V = \mathcal{X}$, in which an edge $(x, x') \in E$ exists if and only if there exists $y \in \mathcal{Y}$ such that $p(x, y), p(x', y) \geq 0$ and $f(x, y) \neq f(x', y)$. We show that for the scenarios discussed in [12], conditional chromatic entropy of G_X,

$$H_{G_X}^{\chi}(X|Y) = \min_{c \text{ is a coloring of } G_X} H(c(X)|Y), \quad (25)$$

is equal to its conditional graph entropy $H_{G_X}(X|Y)$. We will need the following lemma to prove (25).

Lemma 5. *Conditional chromatic entropy of a k-partite graph G_X is achieved by assigning a single distinct color to each partition.*

Proof. Every pair of vertices in different partitions of G_X is adjacent, hence a color that is assigned to any vertex in one partition cannot be reused to color vertices from other partitions. Therefore, each partition should receive at least one distinct color. Let $c(\cdot)$ be an arbitrary valid coloring of G_X with a conditional entropy of

$$H(c(X)|Y) = \sum_{y \in \mathcal{Y}} \sum_{\lambda \in c(G_X)} p(y)p(c^{-1}(\lambda)|y) \log \frac{1}{p(c^{-1}(\lambda)|y)} \quad (26)$$

$$= -\sum_{y \in \mathcal{Y}} p(y) \sum_{\lambda \in c(G_X)} p(c^{-1}(\lambda)|y) \log p(c^{-1}(\lambda)|y) \quad (27)$$

where

$$p(c^{-1}(\lambda)|y) = \sum_{x:c(x)=\lambda} p(x|y). \quad (28)$$

Each of the k vertex partitions in a complete k-partite graph is equal to a maximal independent set $w \in \mathcal{W}(G_X)$. Let w denote such a partition. Assume that every $x \in w$ is assigned to color λ_w. Then the contribution of λ_w on the conditional entropy in (27) is

$$-p(c^{-1}(\lambda_w)|y) \log p(c^{-1}(\lambda_w)|y)$$
$$= -\sum_{x:c(x)=\lambda_w} p(x|y) \log \left(\sum_{x:c(x)=\lambda_w} p(x|y) \right). \quad (29)$$

Note that $w = \{x : c(x) = \lambda_w\}$ is the set of vertices assigned to λ_w. Consider any partitioning of the vertices in w. Assume that the number of partitions is r, where $1 \leq r \leq |w|$ such that $w =$

$w_1 \cup w_2 \cup \ldots \cup w_r$ where w_i denotes the ith partition. Assign a distinct color λ_i to each partition w_i. The impact of this partitioning on the conditional entropy in (27) is

$$-\sum_{i=1}^{r} p(c^{-1}(\lambda_i)|y) \log p(c^{-1}(\lambda_i)|y)$$

$$= -\sum_{i=1}^{r} \sum_{x \in w_i} p(x|y) \log \left(\sum_{x \in w_i} p(x|y) \right). \tag{30}$$

From (29), we have

$$- \sum_{x:c(x)=\lambda_w} p(x|y) \log \left(\sum_{x:c(x)=\lambda_w} p(x|y) \right) \tag{31}$$

$$= - \sum_{x \in w_1 \cup w_2 \cup \ldots \cup w_r} p(x|y) \log \left(\sum_{x \in w_1 \cup w_2 \cup \ldots \cup w_r} p(x|y) \right) \tag{32}$$

$$= - \sum_{i=1}^{r} \sum_{x \in w_i} p(x|y) \log \left(\sum_{x \in w_1 \cup w_2 \cup \ldots \cup w_r} p(x|y) \right) \tag{33}$$

$$\leq - \sum_{i=1}^{r} \sum_{x \in w_i} p(x|y) \log \left(\sum_{x \in w_i} p(x|y) \right) \tag{34}$$

which follows from the fact that $w_i \subset w$ and

$$\sum_{x \in w} p(x|y) \geq \sum_{x \in w_i} p(x|y) \tag{35}$$

where $0 \leq \sum_{x \in w} p(x|y) \leq 1$ and $0 \leq \sum_{x \in w_i} p(x|y) \leq 1$ for all $i = 1, \ldots, r$. Since (34) is true for every $w \in \mathcal{W}(G_X)$,

$$H_{G_X}^{\mathcal{X}}(X|Y) = \min H(c(X)|Y) \tag{36}$$

$$= -\sum_{y \in \mathcal{Y}} p(y) \sum_{w \in \mathcal{W}(G_X)} \sum_{x:c(x)=\lambda_w} p(x|y) \log \left(\sum_{x:c(x)=\lambda_w} p(x|y) \right) \tag{37}$$

which is achieved by assigning a single distinct color to every maximal independent set, from which Lemma 5 follows. □

Theorem 4. *Conditional chromatic entropy of a complete multipartite characteristic graph is equal to its conditional graph entropy*

$$H_{G_X}(X|Y) = H_{G_X}^{\mathcal{X}}(X|Y). \tag{38}$$

For a functional compression problem with a complete multipartite characteristic graph G_X, one can consider the minimum entropy coloring of G_X instead of G_X^n for which $n \to \infty$.

Proof. A multipartite graph, for which every pair of vertices from different independent sets are connected by an edge, is k-partite for some k. Let G_X be a complete k-partite graph, and V_i denote the set of vertices in partition i for $i = 1, \ldots, k$. The vertex set can be represented as $V = V_1 \cup V_2 \cup \ldots \cup V_k$. It follows that these partitions are equal to the maximal independent sets of G_X, as by definition every pair of vertices in different partitions are adjacent to each other, hence an additional vertex violates the independence rule for any partition. Then $\mathcal{W}(G_X) = \{V_1, \ldots, V_k\}$ and

$$H_{G_X}(X|Y) = \min_{\substack{W-X-Y \\ X \in W \in \mathcal{W}(G_X)=\{V_1,V_2,\ldots,V_k\}}} I(W;X|Y) \tag{39}$$

for which W is a random variable distributed over the set of all partitions such that

$$\sum_{w:x \in w \in \mathcal{W}(G_X)} p(w|x) = 1, \tag{40}$$

where by definition of a complete k-partite graph, each vertex belongs to a single maximal independent set. Let $g(x) \in \mathcal{W}(G_X)$ denote the

independent set x is assigned to. Then,

$$p(w|x) = \begin{cases} 1 & \text{if} \quad w = g(x) \\ 0 & \text{o.w.} \end{cases} \tag{41}$$

for all $w \in \mathcal{W}(G_X)$. Note that (41) is the only valid probability distribution that satisfies (40). Hence, the search space of (39) is restricted to the membership relations satisfying (41), which assigns each vertex to the independent set corresponding to its partition. Then,

$$I(W;X|Y)$$

$$= \sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}} \sum_{w:x \in w \in \mathcal{W}(G_X)} p(y)p(x,w|y) \log \frac{p(x,w|y)}{p(x|y)p(w|y)} \tag{42}$$

$$= \sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}} \sum_{w:x \in w \in \mathcal{W}(G_X)} p(y)p(w|x,y)p(x|y) \log \frac{p(w|x,y)p(x|y)}{p(x|y)p(w|y)} \tag{43}$$

$$= \sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}} \sum_{w:x \in w \in \mathcal{W}(G_X)} p(y)p(w|x)p(x|y) \log \frac{p(w|x)}{\sum_{x' \in \mathcal{X}} p(w,x'|y)} \tag{44}$$

$$= \sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}} \sum_{w:x \in w \in \mathcal{W}(G_X)} p(y)p(w|x)p(x|y) \log \frac{p(w|x)}{\sum_{x' \in \mathcal{X}} p(w|x')p(x'|y)} \tag{45}$$

From (41), we find that

$$\sum_{x' \in \mathcal{X}} p(w|x')p(x'|y) = \sum_{x' \in w} p(x'|y) \tag{46}$$

Then we can transform (49) into

$$I(W;X|Y)$$

$$= \sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}} \sum_{w:x \in w \in \mathcal{W}(G_X)} p(y)p(w|x)p(x|y) \log \frac{p(w|x)}{\sum_{x' \in w} p(x'|y)} \tag{47}$$

$$= \sum_{y \in \mathcal{Y}} p(y) \sum_{x \in \mathcal{X}} \sum_{w:x \in w \in \mathcal{W}(G_X)} p(w|x)p(x|y) \log \frac{p(w|x)}{\sum_{x' \in w} p(x'|y)} \tag{48}$$

$$= \sum_{y \in \mathcal{Y}} p(y) \sum_{x \in \mathcal{X}} \sum_{w:g(x)=w \in \mathcal{W}(G_X)} p(x|y) \log \frac{1}{\sum_{x' \in w} p(x'|y)} \tag{49}$$

$$= -\sum_{y \in \mathcal{Y}} p(y) \sum_{w \in \mathcal{W}(G_X)} \left(\sum_{x \in w} p(x|y) \right) \log \left(\sum_{x' \in w} p(x'|y) \right) \tag{50}$$

where (49) follows from the fact that each x belongs to a single w and (50) holds since vertices $\{x : x \in w\}$ form a disjoint set for each $w \in \mathcal{W}(G_X)$. Since (50) is achieved by the only valid probability distribution for assigning vertices to maximal independent sets,

$$H_{G_X}(X|Y) = -\sum_{y \in \mathcal{Y}} p(y) \sum_{w \in \mathcal{W}(G_X)} \sum_{x \in w} p(x|y) \log \left(\sum_{x' \in w} p(x'|y) \right) \tag{51}$$

By comparing (37) with (50), we have

$$H_{G_X}(X|Y) = H_{G_X}^{\mathcal{X}}(X|Y) \tag{52}$$

hence conditional graph entropy is equal to conditional chromatic entropy. □

VI. CONCLUSION

We studied graphical structures through a characteristic graph model. We established the relations between structural entropy, graph entropy and chromatic entropy. We investigated the cases in which (conditional) graph entropy is equal to (conditional) chromatic entropy. Future work includes lossy compression of graphical structures through graph similarity metrics.

REFERENCES

[1] C. E. Shannon, "The zero error capacity of a noisy channel," *IRE Transactions on Information Theory*, vol. 2, no. 3, pp. 8–19, 1956.

[2] L. Lovász, "On the shannon capacity of a graph," *IEEE Transactions on Information Theory*, vol. 25, no. 1, pp. 1–7, 1979.

[3] J. Körner, "Coding of an information source having ambiguous alphabet and the entropy of graphs," in *6th Prague Conference on Information Theory*, 1973, pp. 411–425.

[4] J. Kahn and J. H. Kim, "Entropy and sorting," *Journal of Computer and System Sciences*, vol. 51, no. 3, pp. 390–399, 1995.

[5] I. Newman, P. Ragde, and A. Wigderson, "Perfect hashing, graph entropy, and circuit complexity," in *Proc. of Fifth Annual Structure in Complexity Theory Conference*, 1990, pp. 91–99.

[6] J. Radhakrishnan, "Better bounds for threshold formulas," in *Proc. of 32nd Annual Symposium on Foundations of Computer Science (FOCS'91)*, 1991, pp. 314–323.

[7] A. Orlitsky and J. R. Roche, "Coding for computing," in *Proc. of IEEE International Symposium on Information Theory (ISIT'95)*, 1995, p. 451.

[8] A. Mowshowitz, "Entropy and the complexity of graphs: I. an index of the relative complexity of a graph," *The Bulletin of Mathematical Biophysics*, vol. 30, no. 1-4, 1968.

[9] M. Dehmer and A. Mowshowitz, "A history of graph entropy measures," *Information Sciences*, vol. 181, no. 1, pp. 57–78, 2011.

[10] N. Alon and A. Orlitsky, "Source coding and graph entropies," *IEEE Transactions on Information Theory*, vol. 42, pp. 1329–1339, 1995.

[11] J. Cardinal, S. Fiorini, and G. Joret, "Minimum entropy coloring," in *Algorithms and Computation*. Springer, 2005, pp. 819–828.

[12] S. Feizi and M. Médard, "Cases where finding the minimum entropy coloring of a characteristic graph is a polynomial time problem," in *Proc. of IEEE International Symposium on Information Theory (ISIT'10)*, 2010, pp. 116–120.

[13] J. A. Bondy and U. S. R. Murty, *Graph theory with applications*. Macmillan London, 1976, vol. 6.

[14] Y. Choi and W. Szpankowski, "Compression of graphical structures," in *Proc. of IEEE International Symposium on Information Theory (ISIT'09)*. IEEE, 2009, pp. 364–368.

[15] ——, "Compression of graphical structures: Fundamental limits, algorithms, and experiments," *IEEE Transactions on Information Theory*, vol. 58, no. 2, pp. 620–638, 2012.

[16] H. Witsenhausen, "The zero-error side information problem and chromatic numbers," *IEEE Transactions on Information Theory*, vol. 22, no. 5, pp. 592–593, 1976.

[17] G. Simonyi, "Graph entropy: a survey," *Combinatorial Optimization*, vol. 20, pp. 399–441, 1995.