Universal Lossless Compression of Graphical Data

Payam Delgosha, *Member, IEEE*, and Venkat Anantharam, *Fellow, IEEE*

***Abstract*— Graphical data is comprised of a graph with marks on its edges and vertices. The mark indicates the value of some attribute associated to the respective edge or vertex. Examples of such data arise in social networks, molecular and systems biology, and web graphs, as well as in several other application areas. Our goal is to design schemes that can efficiently compress such graphical data without making assumptions about its stochastic properties. Namely, we wish to develop a universal compression algorithm for graphical data sources. To formalize this goal, we employ the framework of local weak convergence, also called the objective method, which provides a technique to think of a marked graph as a kind of stationary stochastic processes, stationary with respect to movement between vertices of the graph. In recent work, we have generalized a notion of entropy for unmarked graphs in this framework, due to Bordenave and Caputo, to the case of marked graphs. We use this notion to evaluate the efficiency of a compression scheme. The lossless compression scheme we propose in this paper is then proved to be universally optimal in a precise technical sense. It is also capable of performing local data queries in the compressed form.**

***Index Terms*— Graph compression, sparse graphs, local weak convergence, data compression, universal lossless compression.**

I. Introduction

MODERN data often arrives in a form that is indexed by graphs or other combinatorial structures. This a much richer class of data objects than the traditionally studied time series or multidimensional time series. Examples of such graphical data arise, for instance, in social networks, molecular and systems biology, and web graphs, as well as in several other application areas. An instance of graphical data arising in a social network would be a snapshot view of the network at a given time. Here the graph might describe whether a pair of individuals has ever had an interaction, while the marks on the vertices represent some characteristics of the individuals currently of interest for the data analysis task, e.g. their preference for coffee versus tea, and the marks on the edges the characteristics of their interaction, e.g. whether they are friends or not. Often the graph underlying the data is large. Designing efficient compression schemes to store and analyze the data is therefore of significant importance. It is also desirable that the efficiency guarantee of the compression scheme not be dependent on the presumed accuracy of a stochastic model for it. This motivates the desire for a universal compression scheme for graphical data sources. Moreover, since data analysis is often the ultimate goal, the ability to make a broad class of data queries in the compressed form would be valuable. This paper develops techniques to address these goals and achieves them in a specific technical sense.

Manuscript received September 21, 2019; accepted April 4, 2020. Date of publication May 6, 2020; date of current version October 21, 2020. This work was supported in part by the National Science Foundation (NSF) through the Science and Technology Center under Grant CCF-0939370 titled "Science of Information" and under Grant ECCS-1343398, Grant CNS-1527846, Grant CIF-1618145, and Grant CCF-1901004; and in part by the William and Flora Hewlett Foundation through the Center for Long Term Cybersecurity at Berkeley. This article was presented in part at the 2017 IEEE International Symposium on Information Theory. *(Corresponding author: Payam Delgosha.)*

The authors are with the Department of Electrical Engineering and Computer Sciences, University of California at Berkeley, Berkeley, CA 94720 USA (e-mail: pdelgosha@berkeley.edu; ananth@berkeley.edu).

Communicated by V. Vaishampayan, Associate Editor for Source Coding.

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2020.2991384

A. Prior Work

The literature on compression and on evaluating the information content of graphical data can be divided into two categories based on whether there is a stochastic model for the graphical data or not. To the best of our knowledge, none of the existing works address efficient universal data compression of such data to the extent studied in this paper. Works that do not consider a stochastic model for the data generally propose a scheme and illustrate its performance through some analysis and simulation. For example, Boldi and Vigna proposed the WebGraph framework to encode the web graph, where each node represents a URL, and two nodes are connected if there is a link between them [1]. Later, Boldi et al. proposed a method called layered label propagation as a compression scheme for social networks [2].

Among models making stochastic assumptions, Choi and Szpankowski studied structural compression of the Erdős–Rényi ensemble $\mathcal{G}(n,p)$ [3]. There has been a recent series of works addressing the universal compression of binary trees, see for instance [4], [5], [6], [7]. Basu and Varshney addressed the problem of source coding for deep neural networks [8]. Aldous and Ross studied models of sparse random graphs (i.e. the number of edges is of the order of the number of vertices), with vertex labels [9]. They considered several models on sparse random graphs, and studied the asymptotic behavior of the entropy of such models. They observed that the leading term in these models scales as $n \log n$, where n is the number of vertices in the graph.

B. Our Contributions

We assume that a graph is presented to us where each vertex carries a mark coming from some fixed finite set of possible vertex marks and each edge carries two marks, one towards each of its endpoints, coming from some fixed finite set of possible edge marks. Our goal is to design a lossless compression scheme capable of compressing this marked graph in order to store it in an efficient way. Furthermore, we would like to do this in a universally efficient fashion, meaning that the compressor does not make any prior assumption about the statistical properties of the data, but is nevertheless efficient.

In order to make sense of the question, we employ the framework of local weak convergence, also called the objective method [10]–[12]. This framework views a finite marked graph as a probability distribution on a space of rooted marked graphs. Convergence of a sequence of finite marked graphs is then defined as the convergence of this sequence of probability distributions on rooted marked graphs. Both the pre-limit and the limit can then be thought of as a kind of rooted marked graph valued stationary stochastic process, stationary with respect to changes in the root.

In the case of unmarked graphs, a notion of entropy, which applies for such limits and, crucially, works on a per vertex basis, is defined in [13]. We have generalized this entropy notion so that it works for the limits arising from sequences of marked graphs [14]. Efficient universal compression is then formalized as the ability to compress marked graphs such that one asymptotically pays the minimal entropy cost, the asymptotics being in the number of vertices of the graph, without prior knowledge of the limit. Our main contribution is to design such a universal compression scheme and prove its optimality in a precise technical sense.

The paper is organized as follows. In Section II, we describe the language of local weak convergence. Section III discusses our notion of entropy. In Section IV, we precisely formulate the problem and state the main results. In Section V, we introduce our universal compression scheme and provide the proof of its optimality. We make some concluding remarks in Section VI.

C. Notational Conventions

All the logarithms in this paper are to the natural base unless otherwise stated. We therefore use nats instead of bits as the unit of information. For two sequences $(a_n, n \geq 1)$ and $(b_n, n \geq 1)$ of positive real numbers, we write $a_n = O(b_n)$ if $\sup_n a_n/b_n < \infty$, and we write $a_n = o(b_n)$ if $a_n/b_n \to 0$ as $n \to \infty$. We write $\{0,1\}^* - \emptyset$ for the set of sequences of zeros and ones of finite length, excluding the empty sequence. For $x \in \{0,1\}^* - \emptyset$, we denote the length of the sequence x by $\mathsf{nats}(x)$, which is obtained by multiplying the length of x in bits by $\log 2$. $\mathbb{Z}$ denotes the set of integers. We write $:=$ and $=:$ for equality by definition. Other notation used in this document is introduced at its first appearance.

II. Preliminaries

Let G be a simple graph, i.e. one with no self loops or multiple edges. We consider graphs which may have either a finite or a countably infinite number of vertices. Let Θ and Ξ be finite sets called the set of vertex marks and the set of edge marks respectively. A (simple) marked graph is a simple graph G where vertices carry marks from the vertex mark set Θ, and each edge carries two marks from the edge mark set Ξ, one towards each of its endpoints. We denote the mark of a vertex v in a marked graph G by $\tau_G(v)$, and the mark of an edge (i, j) in G towards vertex j by $\xi_G(i, j)$. For a marked or an unmarked graph G, let $V(G)$ denote the set of vertices in G. For a marked or an unmarked graph G and vertices $i, j \in V(G)$, we write $i \sim_G j$ to denote that i and j are adjacent in G. All graphs in this paper, marked or unmarked, are assumed to be simple. Hence, we may drop the term "simple" when referring to graphs. See Figure 1 for an example. A marked tree is a marked graph T where the underlying graph is a tree.

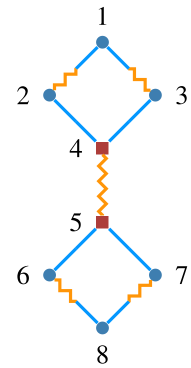

Fig. 1. A marked graph G on the vertex set $\{1, \ldots, 8\}$ where edges carry marks from $\Xi = \{\text{Blue (solid)}, \text{Orange (wavy)}\}$ (e.g. $\xi_G(1,2) = \text{Orange}$ while $\xi_G(2,1) = \text{Blue}$; also, $\xi_G(2,4) = \xi_G(4,2) = \text{Blue}$) and vertices carry marks from $\Theta = \{\bullet, \blacksquare\}$ (e.g. $\tau_G(3) = \bullet$).

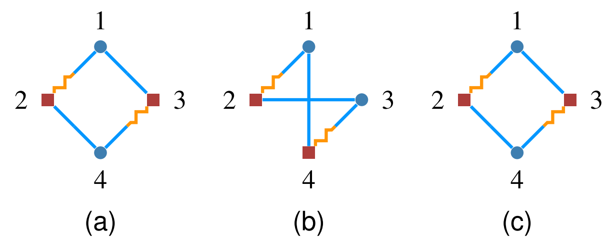

Fig. 2. (a) A marked graph G on the vertex set $\{1, \ldots, 4\}$ with $\Xi = \{\text{Blue (solid)}, \text{Orange (wavy)}\}$ and $\Theta = \{\bullet, \blacksquare\}$; (b) $\pi_1 G$ where $\pi_1 = (3\ 4)$; (c) $\pi_2 G$ where $\pi_2 = (1\ 4)(2\ 3)$. Note that $\pi_2 G = G$.

Two marked graphs G and G' are said to be isomorphic, and we write $G \equiv G'$, if there is a bijection $\phi : V(G) \to V(G')$ such that:

1) $\tau_{G'}(\phi(i)) = \tau_G(i)$ for all $i \in V(G)$;
2) $i \sim_G j$ iff $\phi(i) \sim_{G'} \phi(j)$;
3) For $i \sim_G j$, we have $\xi_{G'}(\phi(i), \phi(j)) = \xi_G(i, j)$.

To better understand this notion, let $\mathcal{S}_n$ denote the permutation group on the set $\{1, \ldots, n\}$. For a permutation $\pi \in \mathcal{S}_n$ and a marked graph G on the vertex set $\{1, \ldots, n\}$, let πG be the marked graph on the same vertex set after the permutation π is applied on the vertices. Namely, vertex $\pi(i)$ has mark $\tau_G(i)$ for $1 \leq i \leq n$, and for each edge (i, j) in G, we place an edge between the vertices $\pi(i)$ and $\pi(j)$ in πG, with mark $\xi_G(i, j)$ towards $\pi(j)$, and mark $\xi_G(j, i)$ towards $\pi(i)$. Then each πG is isomorphic to G and every marked graph that is isomorphic to G is of the form πG for some $\pi \in \mathcal{S}_n$. See Figure 2 for an example.

A path between two vertices v and w in the marked graph G, is a sequence of distinct vertices $v_0, v_1, \ldots, v_k$, such that $v_0 = v$, $v_k = w$ and, for all $1 \leq i \leq k$, we have $v_{i-1} \sim_G v_i$. The length of such a path is defined to be k. The distance between v and w is the length of the shortest path connecting v and w. If there is no such path, the distance is defined to be ∞.

Given a marked graph G, and a subset S of its vertices, the subgraph induced by S is the marked graph comprised of the vertices in S and those edges in G that have both their endpoints in S, with the vertex and edge marks being inherited from G. The connected component of a vertex $v \in V(G)$ is the subgraph of G induced by the vertices that are at a finite distance from v. We write G_v for the connected component of $v \in V(G)$. Note that G_v is a connected marked graph.

For a marked graph G and a vertex $v \in V(G)$, we denote the degree of v, i.e. the number of edges connected to v, by $\deg_G(v)$. Given $x, x' \in \Xi$, we let $\deg_G^{x,x'}(v)$ denote the number of neighbors $w \sim_G v$ such that $\xi_G(w, v) = x$ and $\xi_G(v, w) = x'$. A marked graph G is called locally finite if the degree of every vertex in the graph is finite.

The focus on how a marked graph looks from the point of view of each of its vertices is the key conceptual ingredient in the theory of local weak convergence. For this, we introduce the notion of a rooted marked graph and the notion of isomorphism of rooted marked graphs. Roughly speaking, a rooted marked graph should be thought of as a marked graph as seen from a specific vertex in it and the notion of two rooted marked graphs being isomorphic as capturing the idea that the respective marked graphs as seen from the respective distinguished vertices look the same. Notice that it is natural to restrict attention to the connected component containing the root when making such a definition, because, roughly speaking, a vertex of the marked graph should only be able to see the component to which it belongs.

For a precise definition, consider a marked graph G and a distinguished vertex $o \in V(G)$. The pair (G, o) is called a rooted marked graph. We call two rooted marked graphs (G, o) and (G', o') isomorphic and write $(G, o) \equiv (G', o')$ if $G_o \equiv G'_{o'}$ through a bijection $\phi : V(G_o) \to V(G'_{o'})$ preserving the root, i.e. $\phi(o) = o'$. This notion of isomorphism defines an equivalence relation on rooted marked graphs. Note that in order to determine if two rooted marked graphs are isomorphic (as rooted marked graphs) it is only necessary to examine the connected component of the root in each of the marked graphs. Let $[G, o]$ denote the equivalence class corresponding to (G_o, o). In the sequel, we will only use this notion for locally finite graphs.

For a rooted marked graph (G, o) and integer $h \geq 1$, let $(G, o)_h$ be the subgraph of G rooted at o induced by vertices with distance no more than h from o. If $h = 0$, $(G, o)_h$ is defined to be the isolated root o with mark $\tau_G(o)$. Moreover, let $[G, o]_h$ be the equivalence class corresponding to $(G, o)_h$, i.e. $[G, o]_h := [(G, o)_h]$. Note that $[G, o]_h$ depends only on $[G, o]$.

Finally, note that all the notions introduced in this subsection have the obvious parallels for unmarked graphs. One could simply walk through the definitions while taking each of the mark sets Θ and Ξ to be of cardinality 1.

To close this section we introduce some notation that will only be needed when we develop our compression algorithm for graphical data. For a locally finite graph G and integer Δ, let G^Δ be the graph with the same vertex set that includes only those edges of G such that the degrees of both their endpoints are at most Δ (without reference to their marks). Another way to put this is that to arrive at G^Δ from G we remove all the edges in G that are connected to vertices with degree strictly bigger than Δ. This construction is used as a technical device in the proof of the main result, the main point being that the maximum degree in G^Δ is at most Δ.

A. The Framework of Local Weak Convergence

In this section, we review the framework of local weak convergence of graphs, also called the objective method, with our focus being on marked graphs. See [11], [12], [15] for more details. Throughout our discussion, we assume that the set of edge marks, Ξ, and the set of vertex marks, Θ, are finite and fixed.

Let $\bar{\mathcal{G}}_*$ be the space of equivalence classes $[G, o]$ arising from locally finite rooted marked graphs (G, o). We emphasize again that in defining $[G, o]$ all that matters about (G, o) is the connected component of the root. We define the metric $\bar{d}_*$ on $\bar{\mathcal{G}}_*$ as follows: given $[G, o]$ and $[G', o']$, let $\hat{h}$ be the supremum over all integers $h \geq 0$ such that $(G, o)_h \equiv (G', o')_h$, where (G, o) and (G', o') are arbitrary members in equivalence classes $[G, o]$ and $[G', o']$ respectively.[1] If there is no such h (which can only happen if $\tau_G(o) \neq \tau_{G'}(o')$), we define $\hat{h} = 0$. With this, $\bar{d}_*([G, o], [G', o'])$ is defined to be $1/(1 + \hat{h})$. One can check that $\bar{d}_*$ is a metric; in particular, it satisfies the triangle inequality. Moreover, $\bar{\mathcal{G}}_*$ together with this metric is a Polish space, i.e. a complete separable metric space [12].[2] Let $\bar{\mathcal{T}}_*$ be the subset of $\bar{\mathcal{G}}_*$ comprised of the equivalence classes $[G, o]$ arising from some (G, o) where the graph underlying G is a tree. In the sequel we will think of $\bar{\mathcal{G}}_*$ as a Polish space with the metric $\bar{d}_*$ defined above, rather than just a set. Note that $\bar{\mathcal{T}}_*$ is a closed subset of $\bar{\mathcal{G}}_*$.

For a Polish space Ω, let $\mathcal{P}(\Omega)$ denote the set of Borel probability measures on Ω. We say that a sequence of measures μ_n on Ω converges weakly to $\mu \in \mathcal{P}(\Omega)$ and write $\mu_n \Rightarrow \mu$, if for any bounded continuous function on Ω, we have $\int f d\mu_n \to \int f d\mu$. It can be shown that it suffices to verify this condition only for uniformly continuous and bounded functions [16]. For a Borel set $B \subset \Omega$, the ϵ–extension of B, denoted by B^ϵ, is defined as the union of the open balls with radius ϵ centered around the points in B. For two probability measures μ and ν in $\mathcal{P}(\Omega)$, the Lévy–Prokhorov distance $d_{\text{LP}}(\mu, \nu)$ is defined to be the infimum of all $\epsilon > 0$ such that for all Borel sets $B \subset \Omega$ we have $\mu(B) \leq \nu(B^\epsilon) + \epsilon$ and $\nu(B) \leq \mu(B^\epsilon) + \epsilon$. It is known that the Lévy–Prokhorov distance metrizes the topology of weak convergence on the space of probability distributions on a Polish space (see, for instance, [16]). For $x \in \Omega$, let δ_x be the Dirac measure at x.

[1] As all elements in an equivalence class are isomorphic, the definition is invariant under the choice of the representatives.

[2] It is also possible to define a metric on the space of equivalence classes of locally finite rooted marked graphs when the mark space is itself an arbitrary Polish space, by modifying the definition of the distance between two rooted marked graphs to penalize differences in marks according to how close they are to each other. However, this more general definition is not needed for the purposes of this paper.

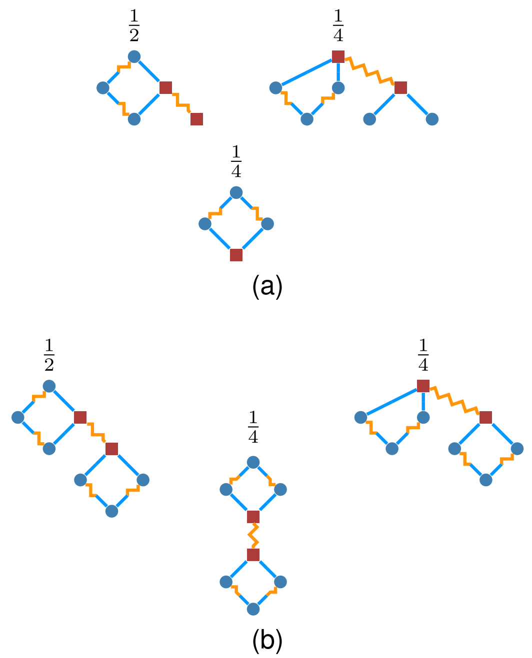

Fig. 3. With G being the graph from Figure 1, (a) illustrates $U_2(G)$, which is a probability distribution on rooted marked graphs of depth at most 2 and (b) depicts $U(G)$, which is a probability distribution on $\bar{\mathcal{G}}_*$.

For a finite marked graph G, define $U(G) \in \mathcal{P}(\bar{\mathcal{G}}_*)$ as

$$U(G) := \frac{1}{|V(G)|} \sum_{o \in V(G)} \delta_{[G,o]}. \tag{1}$$

Note that $U(G) \in \mathcal{P}(\bar{\mathcal{G}}_*)$. In creating $U(G)$ from G, we have created a probability distribution on rooted marked graphs from the given marked graph G by rooting the graph at a vertex chosen uniformly at random. Furthermore, for an integer $h \geq 1$, let

$$U_h(G) := \frac{1}{|V(G)|} \sum_{o \in V(G)} \delta_{[G,o]_h}. \tag{2}$$

We then have $U_h(G) \in \mathcal{P}(\bar{\mathcal{G}}_*)$. See Figure 3 for an example.

We say that a probability distribution μ on $\bar{\mathcal{G}}_*$ is the local weak limit of a sequence of finite marked graphs $\{G_n\}_{n=1}^\infty$ when $U(G_n)$ converges weakly to μ (with respect to the topology induced by the metric $\bar{d}_*$). This turns out to be equivalent to the condition that, for any finite depth $h \geq 0$, the structure of G_n from the point of view of a root chosen uniformly at random and then looking around it only to depth h, converges in distribution to μ truncated up to depth h. This description of what is being captured by the definition justifies the term "local" in local weak convergence.

In fact, $U_h(G)$ could be thought of as the "depth h empirical distribution" of the marked graph G. On the other hand, a probability distribution $\mu \in \mathcal{P}(\bar{\mathcal{G}}_*)$ that arises as a local weak limit plays the role of a stochastic process on graphical data, and a sequence of marked graphs $\{G_n\}_{n=1}^\infty$ could be thought of as being asymptotically distributed like this process when μ is the local weak limit of the sequence.

The degree of a probability measure $\mu \in \mathcal{P}(\bar{\mathcal{G}}_*)$, denoted by $\deg(\mu)$, is defined as

$$\deg(\mu) := \mathbb{E}_\mu \left[\deg_G(o)\right] = \int \deg_G(o) d\mu([G,o]),$$

which is the expected degree of the root. Similarly, for $\mu \in \mathcal{P}(\bar{\mathcal{G}}_*)$ and $x, x' \in \Xi$, let $\deg_{x,x'}(\mu)$ be defined as

$$\deg_{x,x'}(\mu) := \int \deg_G^{x,x'}(o) d\mu([G,o]),$$

which is the expected number of edges connected to the root with mark x towards the root and mark x' towards the other endpoint. We use the notation $\vec{\deg}(\mu) := (\deg_{x,x'}(\mu) : x, x' \in \Xi)$. Moreover, for $\mu \in \mathcal{P}(\bar{\mathcal{G}}_*)$ and $\theta \in \Theta$, let

$$\Pi_\theta(\mu) := \mu(\{[G,o] \in \bar{\mathcal{G}}_* : \tau_G(o) = \theta\}), \tag{3}$$

which is the probability under μ that the mark of the root is θ. We use the notation $\vec{\Pi}(\mu) := (\Pi_\theta(\mu) : \theta \in \Theta)$.

All the preceding definitions and concepts have the obvious parallels in the case of unmarked graphs. These can be arrived at by simply walking through the definitions while restricting the mark sets Θ and Ξ to be of cardinality 1. It is convenient, however, to sometimes use the special notation for the unmarked case that matches the one currently in use in the literature. We will therefore write $\mathcal{G}_*$ for the set of rooted isomorphism classes of unmarked graphs. This is just the set $\bar{\mathcal{G}}_*$ in the case where both Ξ and Θ are sets of cardinality 1. We will also denote the metric on $\mathcal{G}_*$ by d_*, which is just $\bar{d}_*$ when both Ξ and Θ are sets of cardinality 1.

We next present some examples to illustrate the concepts defined so far.

1) Let G_n be the finite lattice $\{-n, \dots n\} \times \{-n, \dots, n\}$ in $\mathbb{Z}^2$. As n goes to infinity, the local weak limit of this sequence is the distribution that gives probability one to the lattice $\mathbb{Z}^2$ rooted at the origin. The reason is that, if we fix a depth $h \geq 0$, then for n large almost all of the vertices in G_n cannot see the borders of the lattice when they look at the graph around them up to depth h, so these vertices cannot locally distinguish the graph on which they live from the infinite lattice $\mathbb{Z}^2$.
2) Suppose G_n is a cycle of length n. The local weak limit of this sequence of graphs gives probability one to an infinite 2–regular tree rooted at one of its vertices. The intuitive explanation for this is essentially identical to that for the preceding example.
3) Let G_n be a realization of the sparse Erdős–Rényi graph $\mathcal{G}(n, \alpha/n)$ where $\alpha > 0$, i.e. G_n has n vertices and each edge is independently present with probability α/n. One can show that if all the G_n are defined on a common probability space then, almost surely, the local weak limit of the sequence is the Poisson Galton–Watson tree with mean α, rooted at the initial vertex. To justify why this should be true without going through the details, note that the degree of a vertex in G_n is the sum of $n-1$ independent Bernoulli random variables, each with parameter α/n. For n large, this approximately has a Poisson distribution with mean α. This argument could be repeated for any of the vertices to which the chosen

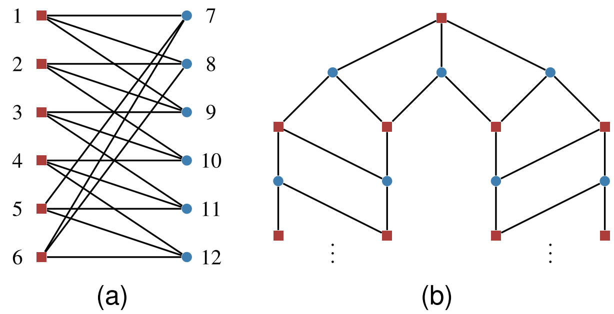

Fig. 4. The graph in Example 4, (a) illustrates the graph G_6 which has 12 vertices and 18 edges. The vertex mark set is $\Theta = \{B(\bullet), R(\blacksquare)\}$, and the edge mark set Ξ has cardinality 1. The local weak limit of G_n is a random rooted graph which gives probability $1/2$ to the rooted marked infinite graph illustrated in (b), and gives probability $1/2$ to a similar rooted marked graph which has a structure identical to (b), but the mark of each vertex is switched from R to B and vice versa.

vertex is connected, which play the role of the offspring of the initial vertex in the limit. The essential point is that the probability of having loops in the neighborhood of a typical vertex up to a depth h is negligible whenever h is fixed and n goes to infinity.

4) Let G_n be a marked bipartite graph on the $2n$ vertices $\{1, \ldots, 2n\}$, the edge mark set having cardinality 1 and the vertex mark set being $\Theta = \{R, B\}$. Suppose $\{1, \ldots, n\}$ is the set of left vertices, all of them having the mark R, and $\{n+1, \ldots, 2n\}$ is the set of right vertices, all of them having the mark B. There are $3n$ edges in the graph, comprised of the edges $(i, n + [\![i]\!])$, $(i, n + [\![i+1]\!])$, and $(i, n + [\![i+2]\!])$ for $1 \leq i \leq n$, where for an integer k, $[\![k]\!]$ is defined to be n if $k \bmod n = 0$, and $k \bmod n$ otherwise, so that $1 \leq [\![k]\!] \leq n$. See Figure 4a for an example. The local weak limit of this sequence of graphs gives probability $\frac{1}{2}$ to the equivalence class of each of the two rooted marked infinite graphs described below. The underlying rooted unmarked infinite graph equivalence class for each of these two rooted marked equivalence classes is the same and can be described as follows: There is a single vertex at level 0, which is the root, three vertices at level 1, and four vertices at each of the levels m for $m \geq 2$. For the purpose of describing the limit (there is no such numbering in the limit), one can number the vertex at level zero as 0, the three vertices at level 1 as $(1, 1)$, $(1, 2)$ and $(1, 3)$, and the four vertices at level m, for each $m \geq 2$, as $(m, 1)$, $(m, 2)$, $(m, 3)$ and $(m, 4)$ such that the edges are the following: Vertex 0 is connected to each of the vertices $(1, 1)$, $(1, 2)$ and $(1, 3)$. Vertex $(1, 1)$ is connected to $(2, 1)$ and $(2, 2)$, vertex $(1, 2)$ is connected to $(2, 2)$ and $(2, 3)$, and vertex $(1, 3)$ is connected to $(2, 3)$ and $(2, 4)$. The edges between the vertices at level k and those at level $k + 1$, for $k \geq 2$, are given by the pattern $((k, 1), (k + 1, 1))$, $((k, 1), (k+1, 2))$, $((k, 2), (k+1, 2))$, $((k, 3), (k+1, 3))$, $((k, 4), (k+1, 3))$, $((k, 4), (k+1, 4))$. There are no other edges. As for the distinction between the two rooted marked equivalence classes which each get probability $\frac{1}{2}$ in the limit, this corresponds to the distinction between choosing the mark R for the root and then alternating between marks B and R as one moves from level to level, or choosing the mark B for the root and then alternating between marks R and B as one moves from level to level. See Figure 4b for an example.

The following lemma gives a useful tool for establishing when local weak convergence holds. This lemma is proved in Appendix A.

Lemma 1: Let $\{\mu_n\}_{n \geq 1}$ and μ be Borel probability measures on $\bar{\mathcal{G}}_*$ such that the support of μ is a subset of $\bar{\mathcal{T}}_*$. Then $\mu_n \Rightarrow \mu$ iff the following condition is satisfied: For all $h \geq 0$ and for all rooted marked trees (T, i) with depth at most h, if

$$A^h_{(T,i)} := \{[G, o] \in \bar{\mathcal{G}}_* : (G, o)_h \equiv (T, i)\}, \quad (4)$$

then $\mu_n(A^h_{(T,i)}) \to \mu(A^h_{(T,i)})$.

B. Unimodularity

In order to get a better understanding of the nature of the results proved in this paper, it is important to understand what is meant by a unimodular probability distribution $\mu \in \mathcal{P}(\bar{\mathcal{G}}_*)$. We give the relevant definitions and context in this section.

Since each vertex in G_n has the same chance of being chosen as the root in the definition of $U(G_n)$, this should manifest itself as some kind of stationarity property of the limit μ, with respect to changes of the root. A probability distribution $\mu \in \mathcal{P}(\bar{\mathcal{G}}_*)$ is called sofic if there exists a sequence of finite graphs G_n with local weak limit μ. The definition of unimodularity is made in an attempt to understand what it means for a Borel probability distribution on $\bar{\mathcal{G}}_*$ to be sofic.

To define unimodularity, let $\bar{\mathcal{G}}_{**}$ be the set of isomorphism classes $[G, o, v]$ where G is a marked connected graph with two distinguished vertices o and v in $V(G)$ (ordered, but not necessarily distinct). Here, isomorphism is defined by an adjacency preserving vertex bijection which preserves vertex and edge marks, and also maps the two distinguished vertices of one object to the respective ones of the other. A measure $\mu \in \mathcal{P}(\bar{\mathcal{G}}_*)$ is said to be unimodular if, for all measurable functions $f : \bar{\mathcal{G}}_{**} \to \mathbb{R}_+$, we have

$$\int \sum_{v \in V(G)} f([G, o, v]) d\mu([G, o]) = \int \sum_{v \in V(G)} f([G, v, o]) d\mu([G, o]). \quad (5)$$

Here, the summation is taken over all vertices v which are in the same connected component of G as o. Note that the integrand on the left hand side is $f([G, o, v])$, while the integrand on the right hand side is $f([G, v, o])$. Roughly speaking, this condition ensures that the distribution μ is invariant under switching the root, and it can be considered as a stationarity condition. It can be seen that it suffices to check the above condition for a function f such that $f([G, o, v]) = 0$ unless $v \sim_G o$. This is called *involution invariance* [12]. Let $\mathcal{P}_u(\bar{\mathcal{G}}_*)$ denote the set of unimodular probability measures on $\bar{\mathcal{G}}_*$. Also, since $\bar{\mathcal{T}}_* \subset \bar{\mathcal{G}}_*$, we can define the set of unimodular probability measures on $\bar{\mathcal{T}}_*$ and denote it by $\mathcal{P}_u(\bar{\mathcal{T}}_*)$. A sofic probability measure is unimodular. Whether the other direction also holds is unknown.

An important consequence of unimodularity is that, roughly speaking, every vertex has a positive probability to be the root. The following is a rephrasing of Lemma 2.3 in [12].

Lemma 2 (Everything Shows at the Root): Let $\mu \in \mathcal{P}_u(\bar{\mathcal{G}}_*)$ be unimodular. If for a subset $\Theta_0 \subset \Theta$ the mark at the root is in Θ_0 almost surely (with $[G, o]$ distributed as μ), then the mark at every vertex is in Θ_0 almost surely. Furthermore, if for a subset $A \subset \Xi \times \Xi$ it holds that for every vertex v adjacent to the root o the pair of edge marks $(\xi_G(v, o), \xi_G(o, v))$ on the edge connecting o to v is in A almost surely (with $[G, o]$ distributed as μ), then for every edge (u, w) the pair of edge marks $(\xi_G(u, w), \xi_G(w, u))$ is in A almost surely.

III. A Notion of Entropy for Processes on Random Rooted Marked Graphs

In this section, we introduce a notion of entropy for probability distributions $\mu \in \mathcal{P}(\bar{\mathcal{G}}_*)$ with $0 < \deg(\mu) < \infty$. This is a generalization of the notion of entropy introduced by Bordenave and Caputo in [13] to the marked framework we discuss in this paper, where vertices and edges are allowed to carry marks. This generalization is due to us, and the reader is referred to [14] for more details.

We first need to set up some notation. Throughout the discussion, we assume that the vertex and edge mark sets, Θ and Ξ respectively, are fixed and finite. For a marked graph G, we define the *edge mark count vector* of G by $\vec{m}_G := (m_G(x, x') : x, x' \in \Xi)$ where $m_G(x, x')$ for $x, x' \in \Xi$ is the number of edges in G with the pair of marks x and x', i.e. the number of edges (v, w) such that $(\xi_G(v, w), \xi_G(w, v)) = (x, x')$ or $(\xi_G(w, v), \xi_G(v, w)) = (x, x')$. Also, we define the *vertex mark count vector* of G by $\vec{u}_G := (u_G(\theta) : \theta \in \Theta)$ where $u_G(\theta)$ denotes the number of vertices in G with mark θ.

We define an *edge mark count vector* to be a vector of nonnegative integers $\vec{m} := (m(x, x') : x, x' \in \Xi)$ such that $m(x, x') = m(x', x)$ for all $x, x' \in \Xi$. Likewise, a *vertex mark count vector* is defined to be a vector of nonnegative integers $\vec{u} := (u(\theta) : \theta \in \Theta)$. Since Ξ is finite, we may assume that it is an ordered set. We define $\|\vec{m}\|_1 := \sum_{x \le x' \in \Xi} m(x, x')$ and $\|\vec{u}\|_1 := \sum_{\theta \in \Theta} u(\theta)$.

Given $n \in \mathbb{N}$ together with edge and vertex mark count vectors $\vec{m}$ and $\vec{u}$ respectively, let $\mathcal{G}^{(n)}_{\vec{m},\vec{u}}$ denote the set of marked graphs G on the vertex set $\{1, \dots, n\}$ such that $\vec{m}_G = \vec{m}$ and $\vec{u}_G = \vec{u}$. Note that $\mathcal{G}^{(n)}_{\vec{m},\vec{u}}$ is empty unless $\|\vec{u}\|_1 = n$ and $\|\vec{m}\|_1 \le \binom{n}{2}$.

We define an *average degree vector* to be a vector of nonnegative reals $\vec{d} := (d_{x,x'} : x, x' \in \Xi)$ such that, for all $x, x' \in \Xi$, we have $d_{x,x'} = d_{x',x}$, and also $\sum_{x,x' \in \Xi} d_{x,x'} > 0$.

Definition 1: Given an average degree vector $\vec{d}$ and a probability distribution $Q = (q_\theta : \theta \in \Theta)$, we say that a sequence $(\vec{m}^{(n)}, \vec{u}^{(n)})$, comprised of edge mark count vectors and vertex mark count vectors $\vec{m}^{(n)}$ and $\vec{u}^{(n)}$ respectively, is adapted to $(\vec{d}, Q)$, if the following conditions hold:

1) For each n, we have $\|\vec{m}^{(n)}\|_1 \le \binom{n}{2}$ and $\|\vec{u}^{(n)}\|_1 = n$;
2) For $x \in \Xi$, we have $m^{(n)}(x, x)/n \to d_{x,x}/2$;
3) For $x \neq x' \in \Xi$, we have $m^{(n)}(x, x')/n \to d_{x,x'} = d_{x',x}$;
4) For $\theta \in \Theta$, we have $u^{(n)}(\theta)/n \to q_\theta$;
5) For $x, x' \in \Xi$, $d_{x,x'} = 0$ implies $m^{(n)}(x, x') = 0$ for all n;
6) For $\theta \in \Theta$, $q_\theta = 0$ implies $u^{(n)}(\theta) = 0$ for all n.

If $\vec{m}^{(n)}$ and $\vec{u}^{(n)}$ are sequences such that $(\vec{m}^{(n)}, \vec{u}^{(n)})$ is adapted to $(\vec{d}, Q)$, using Stirling's approximation one can show that

$$\log |\mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}| = \|\vec{m}^{(n)}\|_1 \log n + nH(Q) + n \sum_{x,x' \in \Xi} s(d_{x,x'}) + o(n), \tag{6}$$

where

$$s(d) := \begin{cases} \frac{d}{2} - \frac{d}{2}\log d & d > 0, \\ 0 & d = 0. \end{cases}$$

To simplify the notation, we may write $s(\vec{d})$ for $\sum_{x,x' \in \Xi} s(d_{x,x'})$.

To give the definition of the BC entropy, we first define the upper and the lower BC entropy.

Definition 2: Assume $\mu \in \mathcal{P}(\bar{\mathcal{G}}_*)$ is given, with $0 < \deg(\mu) < \infty$. For $\epsilon > 0$, and edge and vertex mark count vectors $\vec{m}$ and $\vec{u}$ respectively, define

$$\mathcal{G}^{(n)}_{\vec{m},\vec{u}}(\mu, \epsilon) := \{G \in \mathcal{G}^{(n)}_{\vec{m},\vec{u}} : d_{\mathrm{LP}}(U(G), \mu) < \epsilon\}.$$

Fix an average degree vector $\vec{d}$ and a probability distribution $Q = (q_\theta : \theta \in \Theta)$, and also fix sequences of edge and vertex mark count vectors $\vec{m}^{(n)}$ and $\vec{u}^{(n)}$ respectively such that $(\vec{m}^{(n)}, \vec{u}^{(n)})$ is adapted to $(\vec{d}, Q)$. With these, define

$$\overline{\Sigma}_{\vec{d},Q}(\mu, \epsilon)|_{(\vec{m}^{(n)},\vec{u}^{(n)})} := \limsup_{n \to \infty} \frac{\log |\mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(\mu, \epsilon)| - \|\vec{m}^{(n)}\|_1 \log n}{n},$$

which we call the ϵ–upper BC entropy. Since this is increasing in ϵ, we can define the upper BC entropy as

$$\overline{\Sigma}_{\vec{d},Q}(\mu)|_{(\vec{m}^{(n)},\vec{u}^{(n)})} := \lim_{\epsilon \downarrow 0} \overline{\Sigma}_{\vec{d},Q}(\mu, \epsilon)|_{(\vec{m}^{(n)},\vec{u}^{(n)})}.$$

We may define the ϵ–lower BC entropy $\underline{\Sigma}_{\vec{d},Q}(\mu, \epsilon)|_{(\vec{m}^{(n)},\vec{u}^{(n)})}$ similarly as

$$\underline{\Sigma}_{\vec{d},Q}(\mu, \epsilon)|_{(\vec{m}^{(n)},\vec{u}^{(n)})} := \liminf_{n \to \infty} \frac{\log |\mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(\mu, \epsilon)| - \|\vec{m}^{(n)}\|_1 \log n}{n}.$$

Since this is increasing in ϵ, we can define the lower BC entropy $\underline{\Sigma}_{\vec{d},Q}(\mu)|_{(\vec{m}^{(n)},\vec{u}^{(n)})}$ as

$$\underline{\Sigma}_{\vec{d},Q}(\mu)|_{(\vec{m}^{(n)},\vec{u}^{(n)})} := \lim_{\epsilon \downarrow 0} \underline{\Sigma}_{\vec{d},Q}(\mu, \epsilon)|_{(\vec{m}^{(n)},\vec{u}^{(n)})}.$$

Now, we state the following properties of the upper and lower marked BC entropy, which will lead to the definition of the marked BC entropy. The reader is referred to [14] for a proof and more details.

Theorem 1 (Theorem 1 in [14]): Let an average degree vector $\vec{d} = (d_{x,x'} : x, x' \in \Xi)$ and a probability distribution $Q = (q_\theta : \theta \in \Theta)$ be given. Suppose $\mu \in \mathcal{P}(\bar{\mathcal{G}}_*)$ with $0 < \deg(\mu) < \infty$ satisfies any one of the following conditions:

1) μ is not unimodular;
2) μ is not supported on $\bar{\mathcal{T}}_*$;
3) $\deg_{x,x'}(\mu) \neq d_{x,x'}$ for some $x, x' \in \Xi$, or $\Pi_\theta(\mu) \neq q_\theta$ for some $\theta \in \Theta$.

Then, for any choice of the sequences $\vec{m}^{(n)}$ and $\vec{u}^{(n)}$ such that $(\vec{m}^{(n)}, \vec{u}^{(n)})$ is adapted to $(\vec{d}, Q)$, we have $\overline{\Sigma}_{\vec{d},Q}(\mu)|_{(\vec{m}^{(n)},\vec{u}^{(n)})} = -\infty$.

A consequence of Theorem 1 is that the only case of interest in the discussion of marked BC entropy is when $\mu \in \mathcal{P}_u(\bar{\mathcal{T}}_*)$, $\vec{d} = \vec{\deg}(\mu)$, $Q = \vec{\Pi}(\mu)$, and the sequences $\vec{m}^{(n)}$ and $\vec{u}^{(n)}$ are such that $(\vec{m}^{(n)}, \vec{u}^{(n)})$ is adapted to $(\vec{\deg}(\mu), \vec{\Pi}(\mu))$. Namely, the only upper and lower marked BC entropies of interest are $\overline{\Sigma}_{\vec{\deg}(\mu),\vec{\Pi}(\mu)}(\mu)|_{(\vec{m}^{(n)},\vec{u}^{(n)})}$ and $\underline{\Sigma}_{\vec{\deg}(\mu),\vec{\Pi}(\mu)}(\mu)|_{(\vec{m}^{(n)},\vec{u}^{(n)})}$ respectively.

The following Theorem 2 establishes that the upper and lower marked BC entropies do not depend on the choice of the defining pair of sequences $(\vec{m}^{(n)}, \vec{u}^{(n)})$. Further, this theorem establishes that the upper marked BC entropy is always equal to the lower marked BC entropy. The reader is referred to [14] for a proof and more details.

Theorem 2 (Theorem 2 in [14]): Assume that an average degree vector $\vec{d} = (d_{x,x'} : x, x' \in \Xi)$ together with a probability distribution $Q = (q_\theta : \theta \in \Theta)$ are given. For any $\mu \in \mathcal{P}(\bar{\mathcal{G}}_*)$ such that $0 < \deg(\mu) < \infty$, we have

1) The values of $\overline{\Sigma}_{\vec{d},Q}(\mu)|_{(\vec{m}^{(n)},\vec{u}^{(n)})}$ and $\underline{\Sigma}_{\vec{d},Q}(\mu)|_{(\vec{m}^{(n)},\vec{u}^{(n)})}$ are invariant under the specific choice of the sequences $\vec{m}^{(n)}$ and $\vec{u}^{(n)}$ such that $(\vec{m}^{(n)}, \vec{u}^{(n)})$ is adapted to $(\vec{d}, Q)$. With this, we may simplify the notation and unambiguously write $\overline{\Sigma}_{\vec{d},Q}(\mu)$ and $\underline{\Sigma}_{\vec{d},Q}(\mu)$.
2) $\overline{\Sigma}_{\vec{d},Q}(\mu) = \underline{\Sigma}_{\vec{d},Q}(\mu)$. We may therefore unambiguously write $\Sigma_{\vec{d},Q}(\mu)$ for this common value, and call it the marked BC entropy of $\mu \in \mathcal{P}(\bar{\mathcal{G}}_*)$ for the average degree vector $\vec{d}$ and a probability distribution $Q = (q_\theta : \theta \in \Theta)$. Moreover, $\Sigma_{\vec{d},Q}(\mu) \in [-\infty, s(\vec{d}) + H(Q)]$.

From Theorem 1 we conclude that unless $\vec{d} = \vec{\deg}(\mu)$, $Q = \vec{\Pi}(\mu)$, and μ is a unimodular measure on $\bar{\mathcal{T}}_*$, we have $\Sigma_{\vec{d},Q}(\mu) = -\infty$. In view of this, for $\mu \in \mathcal{P}(\bar{\mathcal{G}}_*)$ with $0 < \deg(\mu) < \infty$, we write $\Sigma(\mu)$ for $\Sigma_{\vec{\deg}(\mu),\vec{\Pi}(\mu)}(\mu)$. Likewise, we may write $\underline{\Sigma}(\mu)$ and $\overline{\Sigma}(\mu)$ for $\underline{\Sigma}_{\vec{\deg}(\mu),\vec{\Pi}(\mu)}(\mu)$ and $\overline{\Sigma}_{\vec{\deg}(\mu),\vec{\Pi}(\mu)}(\mu)$, respectively. Note that, unless $\mu \in \mathcal{P}_u(\bar{\mathcal{T}}_*)$, we have $\overline{\Sigma}(\mu) = \underline{\Sigma}(\mu) = \Sigma(\mu) = -\infty$.

We are now in a position to define the marked BC entropy.

Definition 3: For $\mu \in \mathcal{P}(\bar{\mathcal{G}}_*)$ with $0 < \deg(\mu) < \infty$, the marked BC entropy of μ is defined to be $\Sigma(\mu)$.

IV. MAIN RESULTS

Let $\bar{\mathcal{G}}_n$ be the set of marked graphs on the vertex set $\{1, \dots, n\}$, with edge marks from Ξ and vertex marks from Θ. Our goal is to design a compression scheme, comprised of compression and decompression functions f_n and g_n for each n, such that f_n maps $\bar{\mathcal{G}}_n$ to $\{0,1\}^* - \emptyset$ and g_n maps $\{0,1\}^* - \emptyset$ to $\bar{\mathcal{G}}_n$, with the condition that $g_n \circ f_n(G) = G$ for all $G \in \bar{\mathcal{G}}_n$. Motivated by the notion of entropy introduced in the previous section, we want our compression scheme to be universally optimal in the following sense: if $\mu \in \mathcal{P}(\bar{\mathcal{T}}_*)$ is unimodular and $G^{(n)}$ is a sequence of marked graphs with local weak limit μ, then, with $\vec{m}^{(n)} := \vec{m}_{G^{(n)}}$, we have

$$\limsup_{n\to\infty} \frac{\mathsf{nats}(f_n(G^{(n)})) - \|\vec{m}^{(n)}\|_1 \log n}{n} \leq \Sigma(\mu). \quad (7)$$

In Section V, we design such a universally optimal compression scheme and prove its optimality. This is stated formally in the next theorem.

Theorem 3: There is a compression scheme that is optimal in the above sense for all $\mu \in \mathcal{P}_u(\bar{\mathcal{T}}_*)$ such that $\deg(\mu) \in (0, \infty)$.

We also prove the following converse theorem, which justifies the claim of optimality for compression schemes that satisfy the condition in (7).

Theorem 4: Assume that a compression scheme $\{f_n, g_n\}_{n=1}^\infty$ is given. Fix some unimodular $\mu \in \mathcal{P}_u(\bar{\mathcal{T}}_*)$ such that $\deg(\mu) \in (0, \infty)$ and $\Sigma(\mu) > -\infty$. Moreover, fix a sequence $\vec{m}^{(n)}$ and $\vec{u}^{(n)}$ of edge mark count vectors and vertex mark count vectors respectively, such that $(\vec{m}^{(n)}, \vec{u}^{(n)})$ is adapted to $(\vec{\deg}(\mu), \vec{\Pi}(\mu))$. Then, there exists a sequence of positive real numbers ϵ_n going to zero, together with a sequence of independent graph–valued random variables $\{G^{(n)}\}_{n=1}^\infty$ defined on a joint probability space, with $G^{(n)}$ being uniform in $\mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(\mu, \epsilon_n)$, such that with probability one

$$\liminf_{n\to\infty} \frac{\mathsf{nats}(f_n(G^{(n)})) - \|\vec{m}^{(n)}\|_1 \log n}{n} \geq \Sigma(\mu).$$

Proof: First note that any marked graph on n vertices can be represented with $O(n^2)$ bits. Hence, without loss of generality, we may assume that, for some finite positive constant c, we have $\mathsf{nats}(f_n(G^{(n)})) \leq cn^2$ for all $G^{(n)}$ on n vertices. Consequently, by adding a header of size $O(\log n^2) = O(\log n)$ to the beginning of each codeword in f_n, in order to describe its length, we can make f_n prefix–free. Thus, without loss of generality, we may assume that f_n is prefix–free for all n.

From the definition of $\underline{\Sigma}(\mu)$, one can find a sequence of positive numbers ϵ_n going to zero, such that

$$\underline{\Sigma}(\mu) = \liminf_{n\to\infty} \frac{\log |\mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(\mu, \epsilon_n)| - \|\vec{m}^{(n)}\|_1 \log n}{n}.$$

From Theorem 2, we have $\underline{\Sigma}(\mu) = \Sigma(\mu)$, and since $\Sigma(\mu) > -\infty$ by assumption, $\mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(\mu, \epsilon_n)$ is nonempty once n is large enough. Using Kraft's inequality, we have

$$\sum_{G \in \mathcal{G}^{(n)}_{\vec{m}^{(n)}, \vec{u}^{(n)}}(\mu, \epsilon_n)} e^{-\mathsf{nats}(f_n(G))} \leq 1.$$

With $G^{(n)}$ being uniform in $\mathcal{G}^{(n)}_{\vec{m}^{(n)}, \vec{u}^{(n)}}(\mu, \epsilon_n)$, the Markov inequality then implies that

$$\mathbb{P}\left(\mathsf{nats}(f_n(G^{(n)})) < \log |\mathcal{G}^{(n)}_{\vec{m}^{(n)}, \vec{u}^{(n)}}(\mu, \epsilon_n)| - 2 \log n \right) \leq \frac{1}{n^2}.$$

From this, using the Borel-Cantelli lemma, we have $\mathsf{nats}(f_n(G^{(n)})) \geq \log |\mathcal{G}^{(n)}_{\vec{m}^{(n)}, \vec{u}^{(n)}}(\mu, \epsilon_n)| - 2 \log n$ eventually. Therefore, with probability 1, we have

$$\begin{aligned}
&\liminf_{n \to \infty} \frac{\mathsf{nats}(f_n(G^{(n)})) - \|\vec{m}^{(n)}\|_1 \log n}{n} \\
&\geq \liminf_{n \to \infty} \frac{\log |\mathcal{G}^{(n)}_{\vec{m}^{(n)}, \vec{u}^{(n)}}(\mu, \epsilon_n)| - \|\vec{m}^{(n)}\|_1 \log n}{n} \\
&= \underline{\Sigma}(\mu) = \Sigma(\mu),
\end{aligned}$$

which completes the proof. □

Remark 1: Note that the existence of a sequence of graph–valued random variables $\{G^{(n)}\}_{n=1}^{\infty}$ for which with probability one the normalized codeword length is asymptotically no less than the BC entropy $\Sigma(\mu)$, as is implied by Theorem 4 above, in particular implies the existence of a sequence of deterministic graphs for which the normalized codeword length is asymptotically no less than the BC entropy. This draws a connection between our converse setup of Theorem 4 and the achievability result of Theorem 3 in which a sequence of deterministic graphs is considered which is convergent in the local weak limit sense.

V. The Universal Compression Scheme

In this section, we propose our compression scheme. First, in Section V-A, we introduce our compression scheme under certain assumptions. Then, in Section V-B, we relax these assumptions.

A. A First–Step Scheme

We first give an outline of the compression scheme, then illustrate it via an example, and finally formally describe it and prove its optimality. Fix two sequences of integers k_n and Δ_n as design parameters, which will be specified in Section V-B. Given a marked graph $G^{(n)}$ on n vertices, with maximum degree no more than Δ_n, we first encode its depth–k_n empirical distribution, i.e. $U_{k_n}(G^{(n)})$ (defined in (2)). We do this by counting the number of times each marked rooted graph with depth at most k_n and maximum degree at most Δ_n appears in the graph $G^{(n)}$. Then, in the set of all graphs which result in these counts, we specify the target graph $G^{(n)}$. Figure 5 illustrates an example of this procedure. In this example, the marked graph on $n = 4$ vertices in Figure 5a is given and the design parameters $k_n = 1$ and $\Delta_n = 2$ are chosen. We then list all the rooted marked graphs with depth at most $k_n = 1$ and maximum degree at most $\Delta_n = 2$, and count the number of times each of these patterns appears in the graph, as depicted in Figure 5b. Finally, we consider all the graphs that would result in the same counts if we run this procedure on them (shown in Figure 5c for this example), and specify the input graph within this collection of graphs. In principle, this scheme is similar to the conventional universal coding for sequential data in which we first specify the type of a given sequence and then specify the sequence itself among all the sequences that have this type.

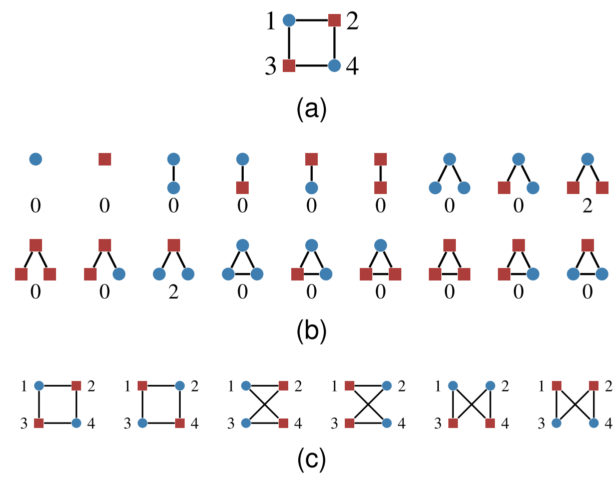

Fig. 5. An example of encoding via the compression function associated to our compression scheme with the parameter $k = 1$ and the graph $G^{(4)}$ on $n = 4$ vertices, with vertex mark set $\Theta = \{●, ■\}$ and edge mark set Ξ with cardinality 1, shown for (a) acting as the input. (b) depicts all members in the set $\mathcal{A}_{1,2}$ with the corresponding number of times each of them appears in the graph, i.e. the vector $(|\psi^{(4)}_{G^{(4)}}([G, o])|, [G, o] \in \mathcal{A}_{1,2})$ and (c) illustrates all the graphs with the same count vector, i.e. W_4.

Before formally explaining the compression scheme, we need some definitions. For integers k and Δ, let $\mathcal{A}_{k,\Delta}$ be the set of equivalence classes of rooted marked graphs $[G, o] \in \bar{\mathcal{G}}_*$ with depth at most k and maximum degree at most Δ. Note that since k and Δ are finite and the mark sets are also finite sets, $\mathcal{A}_{k,\Delta}$ is a finite set.

For a marked graph $G^{(n)}$ on the vertex set $\{1, \dots, n\}$, with maximum degree at most Δ_n, and for $[G, o] \in \mathcal{A}_{k_n, \Delta_n}$, we denote the set $\{1 \leq i \leq n : [G^{(n)}, i]_{k_n} = [G, o]\}$ by $\psi^{(n)}_{G^{(n)}}([G, o])$. This is the set of vertices in $G^{(n)}$ with local structure $[G, o]$ up to depth k_n. Recall that $[G^{(n)}, i]_{k_n} = [G, o]$ means that $G^{(n)}$ rooted at i is isomorphic to (G, o) up to depth k_n. Note that when the maximum degree in $G^{(n)}$ is no more than Δ_n, $[G^{(n)}, i]_{k_n}$ is a member of $\mathcal{A}_{k_n, \Delta_n}$, for all $1 \leq i \leq n$. Therefore, the subsets $\psi^{(n)}_{G^{(n)}}([G, o])$, as $[G, o]$ ranges over $\mathcal{A}_{k_n, \Delta_n}$, form a partition of $\{1, \dots, n\}$.

We encode a marked graph $G^{(n)}$ with vertex set $\{1, \dots, n\}$ and maximum degree no more than Δ_n as follows:

1) Encode the vector $(|\psi^{(n)}_{G^{(n)}}([G, o])|, [G, o] \in \mathcal{A}_{k_n, \Delta_n})$. Since we have $|\psi^{(n)}_{G^{(n)}}([G, o])| \leq n$ for all $[G, o] \in \mathcal{A}_{k_n, \Delta_n}$, this can be done with at most $|\mathcal{A}_{k_n, \Delta_n}|(1 + \lfloor \log_2 n \rfloor) \log 2$ nats.
2) Let W_n be the set of marked graphs G on the vertex set $\{1, \dots, n\}$ with degrees bounded by Δ_n such that

$$|\psi^{(n)}_G([G', o'])| = |\psi^{(n)}_{G^{(n)}}([G', o'])|, \ \forall \ [G', o'] \in \mathcal{A}_{k_n, \Delta_n}. \quad (8)$$

Specify $G^{(n)}$ among the elements of W_n by sending $(1+\lfloor \log_2 |W_n| \rfloor)\log 2$ nats to the decoder.

See Figure 5 for an example of the running of this procedure.

Remark 2: The vector $(\psi^{(n)}_{G^{(n)}}([G,o]) : [G,o] \in \mathcal{A}_{k_n,\Delta_n})$ is directly compressed in the above scheme; therefore, we are capable of making local queries in the compressed form without going through the decompression process. An example of such a query is "how many triangles exist in the graph?"

Now we show the optimality of this compression scheme under an assumption on k_n and Δ_n that allows us to bound the size of the set $\mathcal{A}_{k_n,\Delta_n}$. Lemma 7, which is proved in Appendix B, shows that the assumptions made in Proposition 1 below are not vacuous. To avoid confusion, we use $\tilde{f}_n$ for the compression function in this section and f_n for that of Section V-B (which does not require any a priori assumed bound on the maximum degree of the graph on n vertices, $G^{(n)}$).

Proposition 1: Fix the parameters k_n and Δ_n such that $|\mathcal{A}_{k_n,\Delta_n}| = o(\frac{n}{\log n})$, and $k_n \to \infty$ as $n \to \infty$. Assume that a sequence of marked graphs $\{G^{(n)}\}_{n=1}^{\infty}$ is given such that $G^{(n)}$ is on the vertex set $\{1,\dots,n\}$, the maximum degree in $G^{(n)}$ is no more than Δ_n, and $\{G^{(n)}\}_{n=1}^{\infty}$ converges in the local weak sense to a unimodular $\mu \in \mathcal{P}_u(\bar{\mathcal{T}}_*)$ as $n \to \infty$. Furthermore, assume that $\deg(\mu) \in (0,\infty)$.

Then we have

$$\limsup_{n\to\infty} \frac{\mathsf{nats}(\tilde{f}_n(G^{(n)})) - \|\vec{m}^{(n)}\|_1 \log n}{n} \le \Sigma(\mu), \qquad (9)$$

where $\vec{m}^{(n)} := \vec{m}_{G^{(n)}}$.

Before proving this proposition, we need the following tools. Lemma 3 is stated in a way which is stronger than what we need here, but this stronger form will prove useful later on. The proofs of Lemmas 3, 4, and 5 are given in Appendix B.

Lemma 3: Let G and G' be marked graphs on the vertex set $\{1,\dots,n\}$. For a permutation $\pi \in \mathcal{S}_n$ and an integer $h \ge 0$,

let L be the number of vertices $1 \le i \le n$ such that $(G,i)_h \equiv (G',\pi(i))_h$. Then, we have

$$d_{\mathrm{LP}}(U(G),U(G')) \le \max\left\{\frac{1}{1+h}, 1-\frac{L}{n}\right\}.$$

Lemma 4: If, for integers n and $0 \le m \le \binom{n}{2}$, $\mathcal{G}_{n,m}$ denotes the set of simple unmarked graphs on the vertex set $\{1,\dots,n\}$ with precisely m edges, we have

$$\log|\mathcal{G}_{n,m}| = \log\left|\binom{\binom{n}{2}}{m}\right| \le m\log n + ns\left(\frac{2m}{n}\right),$$

where $s(d) := \frac{d}{2} - \frac{d}{2}\log d$ for $d>0$, and $s(0) := 0$. Moreover, since $s(x) \le 1/2$ for all $x>0$, we have, in particular,

$$\log|\mathcal{G}_{n,m}| \le m\log n + \frac{n}{2}.$$

Lemma 5: Assume that a unimodular $\mu \in \mathcal{P}_u(\bar{\mathcal{T}}_*)$ is given such that $\deg(\mu) \in (0,\infty)$. Moreover, assume that sequences of edge and vertex mark count vectors $\vec{m}^{(n)}$ and $\vec{u}^{(n)}$ respectively are given such that

$$\liminf_{n\to\infty} \frac{m^{(n)}(x,x')}{n} \ge \deg_{x,x'}(\mu), \qquad \forall x \ne x' \in \Xi; \quad (10a)$$

$$\liminf_{n\to\infty} \frac{m^{(n)}(x,x)}{n} \ge \frac{\deg_{x,x}(\mu)}{2}, \qquad \forall x \in \Xi; \quad (10b)$$

$$\lim_{n\to\infty} \frac{u^{(n)}(\theta)}{n} = \Pi_\theta(\mu), \qquad \forall \theta \in \Theta. \quad (10c)$$

Then, for any sequence ϵ_n of positive reals converging to zero, we have

$$\limsup_{n\to\infty} \frac{\log|\mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(\mu,\epsilon_n)| - \left\|\vec{m}^{(n)}\right\|_1 \log n}{n} \le \Sigma(\mu). \qquad (11)$$

Proof of Proposition 1: For our compression scheme, we have

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{\mathsf{nats}(\tilde{f}_n(G^{(n)})) - \|\vec{m}^{(n)}\|_1 \log n}{n} \\
&\quad \le \limsup_{n\to\infty} \frac{|\mathcal{A}_{k_n,\Delta_n}|(\log 2 + \log n)}{n} \\
&\qquad + \frac{\log 2 + \log|W_n| - \|\vec{m}^{(n)}\|_1 \log n}{n} \\
&\quad = \limsup_{n\to\infty} \frac{\log|W_n| - \|\vec{m}^{(n)}\|_1 \log n}{n},
\end{aligned} \qquad (12)$$

where the last equality employs the assumption $|\mathcal{A}_{k_n,\Delta_n}| = o(n/\log n)$. We now show that

$$W_n \subseteq \mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}\left(\mu, \epsilon_n + \frac{1}{1+k_n}\right), \qquad (13)$$

where $\epsilon_n := d_{\mathrm{LP}}(U(G^{(n)}),\mu)$ and $\vec{u}^{(n)} := \vec{u}_{G^{(n)}}$.

For this, let $G \in W_n$. By definition, for all $[G',o'] \in \mathcal{A}_{k_n,\Delta_n}$, we have $|\psi^{(n)}_{G^{(n)}}([G',o'])| = |\psi^{(n)}_{G}([G',o'])|$. Hence there exists a permutation π on the set of vertices $\{1,\dots,n\}$ such that $(G^{(n)},i)_{k_n} \equiv (G,\pi(i))_{k_n}$ for all $1 \le i \le n$. Using Lemma 3 above with $h = k_n$ and $L = n$, we have

$$d_{\mathrm{LP}}(U(G^{(n)}),U(G)) \le \frac{1}{1+k_n}.$$

Consequently,

$$d_{\mathrm{LP}}(U(G),\mu) < \epsilon_n + 1/(1+k_n). \qquad (14)$$

We claim that for $G \in W_n$ we have $\vec{m}_G = \vec{m}^{(n)}$ and $\vec{u}_G = \vec{u}^{(n)}$. To see this, note that for $\theta \in \Theta$ we have

$$\begin{aligned}
u_G(\theta) &= \sum_{i=1}^{n} \mathbb{1}\left[\tau_G(i) = \theta\right] \\
&= \sum_{[G',o'] \in \mathcal{A}_{k_n,\Delta_n}} \sum_{i \in \psi^{(n)}_G([G',o'])} \mathbb{1}\left[\tau_G(i) = \theta\right].
\end{aligned}$$

Note that for $i \in \psi^{(n)}_G([G',o'])$ we have $\tau_G(i) = \tau_{G'}(o')$. Therefore,

$$u_G(\theta) = \sum_{[G',o'] \in \mathcal{A}_{k_n,\Delta_n} : \tau_{G'}(o') = \theta} |\psi^{(n)}_G([G',o'])|.$$

A similar argument implies

$$u^{(n)}(\theta) = \sum_{[G',o'] \in \mathcal{A}_{k_n,\Delta_n} : \tau_{G'}(o')=\theta} |\psi^{(n)}_{G^{(n)}}([G',o'])|.$$

Hence $u^{(n)}(\theta) = u_G(\theta)$. Likewise, for $G \in W_n$ and $x \neq x' \in \Xi$, we can write, for n large enough,

$$\begin{aligned} m_G(x,x') &= \sum_{i=1}^n \deg_G^{x,x'}(i) \\ &= \sum_{[G',o'] \in \mathcal{A}_{k_n,\Delta_n}} \deg_{G'}^{x,x'}(o') |\psi^{(n)}_G([G',o'])| \\ &= \sum_{[G',o'] \in \mathcal{A}_{k_n,\Delta_n}} \deg_{G'}^{x,x'}(o') |\psi^{(n)}_{G^{(n)}}([G',o'])| \\ &= m^{(n)}(x,x'). \end{aligned}$$

The proof of $m_G(x,x) = m^{(n)}(x,x)$ for $x \in \Xi$ is similar. This, together with (14), implies that $G \in \mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(\mu, \epsilon_n + 1/(1+k_n))$ which completes the proof of (13).

Note that, for fixed $t > 0$ and $x, x' \in \Xi$, the mapping $[G,o] \mapsto \deg_G^{x,x'}(o) \wedge t$ is bounded and continuous. Therefore, for $x \neq x' \in \Xi$, we have

$$\begin{aligned} \frac{m^{(n)}(x,x')}{n} &= \int \deg_G^{x,x'}(o) dU(G^{(n)})([G,o]) \\ &\geq \int (\deg_G^{x,x'}(o) \wedge t) dU(G^{(n)})([G,o]) \\ &\xrightarrow{n\to\infty} \int (\deg_G^{x,x'}(o) \wedge t) d\mu. \end{aligned}$$

Sending t to infinity, we get

$$\liminf_{n\to\infty} \frac{m^{(n)}(x,x')}{n} \geq \deg_{x,x'}(\mu).$$

Similarly, for $x \in \Xi$, we have $\liminf_{n\to\infty} m_n(x,x)/n \geq \deg_{x,x}(\mu)/2$. On the other hand, for $\theta \in \Theta$, the mapping $[G,o] \mapsto \mathbb{1}[\tau_G(o) = \theta]$ is bounded and continuous. This implies that

$$\lim_{n\to\infty} u^{(n)}(\theta) = \Pi_\theta(\mu).$$

Thus, substituting (13) into (12), using the fact that $\epsilon_n + 1/(1+k_n) \to 0$, and using Lemma 5 above, we get

$$\begin{aligned} &\limsup_{n\to\infty} \frac{\mathsf{nats}(\tilde{f}_n(G^{(n)})) - \|\vec{m}^{(n)}\|_1 \log n}{n} \\ &\leq \limsup_{n\to\infty} \frac{\log \left| \mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}} \left(\mu, \epsilon_n + \frac{1}{1+k_n}\right) \right| - \|\vec{m}^{(n)}\|_1 \log n}{n} \\ &\leq \Sigma(\mu), \end{aligned}$$

which completes the proof. □

B. Step 2: The General Compression Scheme

In Section V-A we introduced a compression scheme which achieves the BC entropy of μ by focusing on the depth k_n empirical distribution of the graph $G^{(n)}$ in the sequence of graphs $\{G^{(n)}\}_{n=1}^\infty$, under the assumption that the maximum degree of $G^{(n)}$ is bounded above by Δ_n which does not grow too fast, in the sense that $|\mathcal{A}_{k_n,\Delta_n}| = o(n/\log n)$. In principle, we can choose the design parameter k_n, but we have no control over the maximum degree Δ_n. In order to overcome this and drop the restriction on the compression scheme in Section V-A, we first choose k_n and Δ_n and then trim the input graph by removing some edges to make its maximum degree no more than Δ_n. Then, we encode the resulting trimmed graph by the compression function in Section V-A. Finally, we encode the removed edges separately. More precisely, we encode a graph $G^{(n)} \in \bar{\mathcal{G}}_n$ as follows:

1) Define $\Delta_n := \log\log n$.
2) Let $\widetilde{G}^{(n)} := \left(G^{(n)}\right)^{\Delta_n}$ be the trimmed graph obtained by removing each edge connected to any vertex with degree more than Δ_n. Moreover, define
$$\begin{aligned} R_n := \{1 \leq i \leq n : &\deg_{G^{(n)}}(i) > \Delta_n \\ &\text{or } \deg_{G^{(n)}}(j) > \Delta_n \text{ for some } j \sim_{G^{(n)}} i\}, \end{aligned}$$
which consists of the endpoints of the removed edges.
3) Encode the graph $\widetilde{G}^{(n)}$ by the compression function introduced in Section V-A, with $k_n = \sqrt{\log\log n}$.
4) Encode $|R_n|$ using at most $(1 + \lfloor \log_2 n \rfloor) \log 2$ nats.
5) Encode the set R_n using at most $(1+\lfloor \log_2 \binom{n}{|R_n|} \rfloor) \log 2$ nats.
6) Let $\vec{m}^{(n)} = \vec{m}_{G^{(n)}}$ and $\vec{\widetilde{m}}^{(n)} = \vec{m}_{\widetilde{G}^{(n)}}$. Note that the edges present in $G^{(n)}$ but not in $\widetilde{G}^{(n)}$ have both endpoints in the set R_n. So we can first encode $m^{(n)}(x,x') - \widetilde{m}^{(n)}(x,x')$ for all $x, x' \in \Xi$ by $|\Xi|^2(1 + \lfloor \log_2 n^2 \rfloor) \log 2 \leq 2|\Xi|^2(1 + \lfloor \log_2 n \rfloor) \log 2$ nats and then encode these removed edges using
$$\sum_{x \leq x' \in \Xi} \left(1 + \left\lfloor \log_2 \binom{\binom{|R_n|}{2}}{m^{(n)}(x,x') - \widetilde{m}^{(n)}(x,x')} \right\rfloor \right) \log 2$$
nats by specifying the removed edges of each pair of marks separately.

Now we show that this general compression scheme asymptotically achieves the upper BC entropy rate, as was stated in Theorem 3. Before this, we need the results of the following lemmas. We postpone the proofs of these lemmas to Appendix B.

Lemma 6: Assume that $\{G^{(n)}\}_{n=1}^\infty$ is a sequence of marked graphs with local weak limit $\mu \in \mathcal{P}(\bar{\mathcal{T}}_*)$, where $G^{(n)}$ is on the vertex set $\{1, \dots, n\}$.

If Δ_n is a sequence of integers going to infinity as $n \to \infty$, μ is also the local weak limit of the trimmed sequence $\{(G^{(n)})^{\Delta_n}\}_{n=1}^\infty$.

Lemma 7: If $\Delta_n \leq \log\log n$ and $k_n \leq \sqrt{\log\log n}$, then $|\mathcal{A}_{k_n,\Delta_n}| = o(n/\log n)$.

Lemma 8: Assume that $\{G^{(n)}\}_{n=1}^\infty$ is a sequence of marked graphs with local weak limit $\mu \in \mathcal{P}(\bar{\mathcal{T}}_*)$, where $G^{(n)}$ is on the vertex set $\{1, \dots, n\}$. Let $\{\Delta_n\}_{n=1}^\infty$ be a sequence of integers such that $\Delta_n \to \infty$ and define

$$\begin{aligned} R_n := \{1 \leq i \leq n : &\deg_{G^{(n)}}(i) > \Delta_n \\ &\text{or } \deg_{G^{(n)}}(j) > \Delta_n \text{ for some } j \sim_{G^{(n)}} i\}. \end{aligned}$$

Then $|R_n|/n \to 0$ as n goes to infinity.

Proof of Theorem 3: Let $\tilde{f}_n$ be the compression function of the scheme in Section V-A.

We have

$$\begin{aligned}\mathsf{nats}(f_n(G^{(n)})) &\leq \mathsf{nats}(\tilde{f}_n(\widetilde{G}^{(n)})) + \log n + \log \binom{n}{|R_n|} \\ &\quad + 2|\Xi|^2 \log n \\ &\quad + \sum_{x \leq x' \in \Xi} \log \binom{\binom{|R_n|}{2}}{m^{(n)}(x,x') - \widetilde{m}^{(n)}(x,x')} \\ &\quad + C \log 2,\end{aligned}$$

where $C = 2 + 3|\Xi|^2$. Using the inequality $\binom{r}{s} \leq (re/s)^s$ and Lemma 4 above, we have

$$\begin{aligned}\mathsf{nats}(f_n(G^{(n)})) &\leq \mathsf{nats}(\tilde{f}_n(\widetilde{G}^{(n)})) + (1 + 2|\Xi|^2) \log n \\ &\quad + |R_n| \log \frac{ne}{|R_n|} \\ &\quad + (\|\vec{m}^{(n)}\|_1 - \|\vec{\widetilde{m}}^{(n)}\|_1) \log |R_n| \\ &\quad + \frac{|R_n||\Xi|^2}{2} + C \log 2.\end{aligned}$$

Using the fact that $|R_n| \leq n$, this gives

$$\begin{aligned}\mathsf{nats}(f_n(G^{(n)})) &\leq \mathsf{nats}(\tilde{f}_n(\widetilde{G}^{(n)})) + (1 + 2|\Xi|^2) \log n \\ &\quad + |R_n| \log \frac{ne}{|R_n|} \\ &\quad + (\|\vec{m}^{(n)}\|_1 - \|\vec{\widetilde{m}}^{(n)}\|_1) \log n \\ &\quad + \frac{|R_n||\Xi|^2}{2} + C \log 2.\end{aligned}$$

Hence,

$$\begin{aligned}&\limsup_{n\to\infty} \frac{\mathsf{nats}(f_n(G^{(n)}) - \|\vec{m}^{(n)}\|_1 \log n}{n} \\ &\leq \limsup_{n\to\infty} \frac{\mathsf{nats}(\tilde{f}_n(\widetilde{G}^{(n)}) - \|\vec{\widetilde{m}}^{(n)}\|_1 \log n}{n} \qquad (15) \\ &\quad + \limsup_{n\to\infty} \frac{|R_n||\Xi|^2}{2n} + \limsup_{n\to\infty} \frac{|R_n|}{n} \log \frac{ne}{|R_n|}.\end{aligned}$$

Now, we claim that the conditions of Proposition 1 hold for the sequence $\widetilde{G}^{(n)}$ and the parameters k_n and Δ_n defined above. To show this, note that both k_n and Δ_n go to infinity by definition. Lemma 6 then implies that μ is also the local weak limit of the sequence $\widetilde{G}^{(n)}$. Moreover, by Lemma 7, $|\mathcal{A}_{k_n,\Delta_n}| = o(n/\log n)$. On the other hand, the maximum degree in $\widetilde{G}^{(n)}$ is at most Δ_n. Therefore, all the conditions of Proposition 1 are satisfied and

$$\limsup_{n\to\infty} \frac{\mathsf{nats}(\tilde{f}_n(\widetilde{G}^{(n)})) - \|\vec{\widetilde{m}}^{(n)}\|_1 \log n}{n} \leq \Sigma(\mu).$$

Furthermore, all the other terms in (15) go to zero, since, by Lemma 8, $|R_n|/n \to 0$, and the function $\delta \mapsto \delta \log \delta$ goes to zero as $\delta \to 0$. Therefore,

$$\begin{aligned}&\limsup_{n\to\infty} \frac{\mathsf{nats}(f_n(G^{(n)})) - \|\vec{m}^{(n)}\|_1 \log n}{n} \\ &\leq \limsup_{n\to\infty} \frac{\mathsf{nats}(\tilde{f}_n(\widetilde{G}^{(n)}) - \|\vec{\widetilde{m}}^{(n)}\|_1 \log n}{n} \\ &\leq \Sigma(\mu),\end{aligned}$$

which completes the proof. □

Remark 3: From Lemma 8 above, for typical graphs, $|R_n| = o(n)$. Hence, similar to our discussion in Remark 2, we are capable of answering local queries with an error of $o(n)$ without needing to go through the decompression process.

VI. Conclusion

We introduced a notion of stochastic process for graphs, using the language of local weak convergence. Besides, we discussed a generalized notion of entropy for such processes. Using this, we formalized the problem of efficiently compressing graphical data without assuming prior knowledge of its stochastic properties. Finally, we proposed a universal compression scheme which is asymptotically optimal in the size of the underlying graph, as characterized using the discussed notion of entropy, and is capable of performing local data queries in the compressed form, with an error negligible compared to the number of vertices.

Appendix A
Proof of Lemma 1

We first prove that if the condition mentioned in Lemma 1 is satisfied, then $\mu_n \Rightarrow \mu$. Let $f : \bar{\mathcal{G}}_* \to \mathbb{R}$ be a uniformly continuous and bounded function. Since f is uniformly continuous, for fixed $\epsilon > 0$ there exists $\delta > 0$ such that $|f([G,o]) - f([G',o'])| < \epsilon$ for all $[G,o]$ and $[G',o']$ such that $\bar{d}_*([G,o],[G',o']) < \delta$. For this δ, choose k such that $1/(1+k) < \delta$. Note that since Ξ and Θ are finite there are countably many locally finite rooted trees with marks in Ξ and Θ and depth at most k. Therefore, one can find countably many rooted trees $\{(T_j,i_j)\}_{j=1}^\infty$ with depth at most k such that $A^k_{(T_j,i_j)} \cap \bar{\mathcal{T}}_*$ partitions $\bar{\mathcal{T}}_*$. On the other hand, as μ is a probability measure with its support being a subset of $\bar{\mathcal{T}}_*$, one can find finitely many of these (T_j,i_j), which we may index without loss of generality by $1 \leq j \leq m$, such that $\sum_{j=1}^m \mu(A^k_{(T_j,i_j)}) \geq 1 - \epsilon$. To simplify the notation, we use A_j for $A^k_{(T_j,i_j)}$, $1 \leq j \leq m$. Note that if $[G,o] \in A_j$, $\bar{d}_*([G,o],[T_j,i_j]) \leq \frac{1}{1+k} < \delta$. Hence, if $\mathcal{A}$ denotes $\cup_{j=1}^m A_j$, we have

$$\begin{aligned}&\left| \int f d\mu - \sum_{j=1}^m f([T_j,i_j])\mu(A_j) \right| \\ &\quad \leq \sum_{j=1}^m \left| \int_{A_j} f d\mu - f([T_j,i_j])\mu(A_j) \right| + \|f\|_\infty \, \mu(\mathcal{A}^c) \\ &\quad \leq \epsilon(1 + \|f\|_\infty),\end{aligned}$$

where the last inequality uses the facts that $\mu(\mathcal{A}^c) \leq \epsilon$, $1/(1+k) < \delta$, and $|f([G,o]) - f([T_j,i_j])| < \epsilon$ for $[G,o] \in A_j$. Similarly, we have

$$\begin{aligned}&\left| \int f d\mu_n - \sum_{j=1}^m f([T_j,i_j])\mu(A_j) \right| \\ &\quad \leq \left| \int f d\mu_n - \sum_{j=1}^m f([T_j,i_j])\mu_n(A_j) \right| \\ &\qquad + \sum_{j=1}^m |f([T_j,i_j])||\mu_n(A_j) - \mu(A_j)|\end{aligned}$$

$$\leq \|f\|_\infty \left(1 - \sum_{j=1}^m \mu_n(A_j)\right) + \epsilon + \|f\|_\infty \sum_{j=1}^m |\mu_n(A_j) - \mu(A_j)|.$$

Combining the two preceding inequalities, we have

$$\left|\int f d\mu_n - \int f d\mu\right| \leq \|f\|_\infty \left(1 - \sum_{j=1}^m \mu_n(A_j)\right) + \|f\|_\infty \sum_{j=1}^m |\mu_n(A_j) - \mu(A_j)| + \epsilon(2 + \|f\|_\infty).$$

Now, as n goes to infinity, $\mu_n(A_j) \to \mu(A_j)$ by assumption and also $1 - \sum_{j=1}^m \mu_n(A_j) \to \mu(\mathcal{A}^c) \leq \epsilon$. Thus,

$$\limsup_{n\to\infty} \left|\int f d\mu_n - \int f d\mu\right| \leq 2\epsilon(1 + \|f\|_\infty).$$

Since $\|f\|_\infty < \infty$ and $\epsilon > 0$ is arbitrary, $\int f d\mu_n \to \int f d\mu$, whereby $\mu_n \Rightarrow \mu$.

For the converse, fix an integer $h \geq 0$ and a rooted marked tree (T, i) with depth at most h. Since $\mathbb{1}_{A^h_{(T,i)}}([G, o]) = \mathbb{1}_{A^h_{(T,i)}}([G', o'])$ when $\bar{d}_*([G, o], [G', o']) < 1/(1 + h)$, we see that $\mathbb{1}_{A^h_{(T,i)}}$ is a bounded continuous function. This immediately implies that

$$\mu_n(A^h_{(T,i)}) = \int \mathbb{1}_{A^h_{(T,i)}} d\mu_n \to \int \mathbb{1}_{A^h_{(T,i)}} d\mu = \mu(A^h_{(T,i)}),$$

which completes the proof.

Appendix B
Proofs for Section V

Proof of Lemma 3: Let $\mathcal{A}$ be the set of $1 \leq i \leq n$ such that $[G, i]_h = [G', \pi(i)]_h$. Then, for any Borel set $B \subset \bar{\mathcal{G}}_*$, we have

$$U(G)(B) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}\left[[G, i] \in B\right] \leq \frac{1}{n} \sum_{i\in\mathcal{A}} \mathbb{1}\left[[G, i] \in B\right] + 1 - \frac{L}{n}.$$

Note that if for some $i \in \mathcal{A}$ we have $[G, i] \in B$ then, since $(G, i)_h \equiv (G', \pi(i))_h$, we have $d_*([G, i], [G', \pi(i)]) \leq \frac{1}{1+h}$. This means that, for such i, $[G', \pi(i)] \in B^{\delta+1/(1+h)}$ for arbitrary $\delta > 0$. Continuing the chain of inequalities, we have

$$\begin{aligned} U(G)(B) &\leq \frac{1}{n} \sum_{i\in\mathcal{A}} \mathbb{1}\left[[G', \pi(i)] \in B^{\delta+1/(1+h)}\right] + 1 - \frac{L}{n} \\ &\leq \frac{1}{n} \sum_{i=1}^n \mathbb{1}\left[[G', i] \in B^{\delta+1/(1+h)}\right] + 1 - \frac{L}{n} \\ &= U(G')(B^{\delta+1/(1+h)}) + 1 - \frac{L}{n}. \end{aligned}$$

Changing the order of G and G', we have

$$d_{\mathrm{LP}}(U(G), U(G')) \leq \max\left\{\frac{1}{1+h} + \delta, 1 - \frac{L}{n}\right\}.$$

We get the desired result by sending δ to zero. □

Proof of Lemma 4: Using the classical upper bound $\binom{r}{s} \leq (re/s)^s$, we have

$$\begin{aligned} \log\left|\binom{\binom{n}{2}}{m}\right| &\leq m \log \frac{n^2\ e}{2m} \\ &= m \log n + m \log \frac{ne}{2m} = m \log n + ns(2m/n), \end{aligned}$$

which completes the proof of the first statement. Also, it is easy to see that $s(x)$ is increasing for $x < 1$, decreasing for $x > 1$ and attains its maximum value $1/2$ at $x = 1$. Therefore, $s(x) \leq 1/2$. This completes the proof of the second statement. □

Next, we prove Lemma 5. Before that, we state and prove the following lemmas which will be useful in our proof. For a marked graph G on a finite or countably infinite vertex set, let $\mathsf{UM}(G)$ denote the unmarked graph which has the same set of vertices and edges as in G, but is obtained from G by removing all the vertex and edge marks. Given a probability distribution $\mu \in \mathcal{P}(\mathcal{G}_*)$ on the space of isomorphism classes of rooted unmarked graphs, for $\epsilon > 0$ and integers n and m, let $\mathcal{G}_{n,m}(\mu, \epsilon)$ denote the set of unmarked graphs G on the vertex set $\{1, \ldots, n\}$ with m edges such that $d_{\mathrm{LP}}(U(G), \mu) < \epsilon$.

Lemma 9: For $[G, o]$ and $[G', o']$ in $\bar{\mathcal{G}}_*$, we have $d_*([\mathsf{UM}(G), o], [\mathsf{UM}(G'), o']) \leq \bar{d}_*([G, o], [G', o'])$.

Proof: By definition, for $\epsilon > 0$, the condition $\bar{d}_*([G, o], [G', o']) < \epsilon$ means that for some k with $1/(1 + k) < \epsilon$, we have $[G, o]_k \equiv [G', o']_k$. This implies $[\mathsf{UM}(G), o]_k \equiv [\mathsf{UM}(G'), o']_k$, which in particular means that $d_*([\mathsf{UM}(G), o], [\mathsf{UM}(G'), o']) \leq 1/(1 + k) < \epsilon$. □

Lemma 10: Assume $\mu \in \mathcal{P}(\bar{\mathcal{G}}_*)$ is given. Let $\tilde{\mu} \in \mathcal{P}(\mathcal{G}_*)$ be the law of $[\mathsf{UM}(G), o]$ when $[G, o]$ has law μ. Then, given an integer n, edge and vertex mark count vectors $\vec{m}^{(n)}$ and $\vec{u}^{(n)}$ respectively, and $\epsilon > 0$, for all $G \in \mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(\mu, \epsilon)$, we have $\mathsf{UM}(G) \in \mathcal{G}_{n,m_n}(\tilde{\mu}, \epsilon)$ where $m_n := \|\vec{m}^{(n)}\|_1$.

Proof: Fix $G \in \mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(\mu, \epsilon)$. Note that $\mathsf{UM}(G)$ has m_n edges, and we only need to show that $d_{\mathrm{LP}}(U(\mathsf{UM}(G)), \tilde{\mu}) < \epsilon$. Let $\delta := d_{\mathrm{LP}}(U(G), \mu)$. This means that for all $\delta' > \delta$, and for all Borel sets A in $\bar{\mathcal{G}}_*$, we have $(U(G))(A) \leq \mu(A^{\delta'}) + \delta'$ and $\mu(A) \leq (U(G))(A^{\delta'}) + \delta'$, where, $A^{\delta'}$ denotes the δ'–extension of A. Define $T : \bar{\mathcal{G}}_* \to \mathcal{G}_*$ that maps $[G, o] \in \bar{\mathcal{G}}_*$ to $[\mathsf{UM}(G), o] \in \mathcal{G}_*$. Lemma 9 above implies that T is continuous and in fact 1–Lipschitz. It is easy to see that $U(\mathsf{UM}(G))$ is the pushforward of $U(G)$ under the mapping T. Also, $\tilde{\mu}$ is the pushforward of μ under T. Using the fact that T is 1–Lipschitz, it is easy to see that for any Borel set B in $\mathcal{G}_*$, and any $\zeta > 0$, we have $(T^{-1}(B))^\zeta \subset T^{-1}(B^\zeta)$. Putting these together, for $\delta' > \delta$ and a Borel set B in $\mathcal{G}_*$, we have

$$\begin{aligned} U(\mathsf{UM}(G))(B) &= U(G)(T^{-1}(B)) \leq \mu((T^{-1}(B))^{\delta'}) + \delta' \\ &\leq \mu(T^{-1}(B^{\delta'})) + \delta' = \tilde{\mu}(B^{\delta'}) + \delta'. \end{aligned}$$

Similarly,

$$\begin{aligned} \tilde{\mu}(B) &= \mu(T^{-1}(B)) \leq (U(G))((T^{-1}(B))^{\delta'}) + \delta' \\ &\leq (U(G))(T^{-1}(B^{\delta'})) + \delta' \\ &= (U(\mathsf{UM}(G)))(B^{\delta'}) + \delta'. \end{aligned}$$

Since this holds for any $\delta' > \delta$ and any Borel set B in $\mathcal{G}_*$, we have $d_{\mathrm{LP}}(U(\mathsf{UM}(G)), \tilde{\mu}) \le \delta = d_{\mathrm{LP}}(U(G), \mu) < \epsilon$. Consequently, we have $\mathsf{UM}(G) \in \mathcal{G}_{n,m_n}(\tilde{\mu}, \epsilon)$ and the proof is complete. □

Now, we are ready to prove Lemma 5.

Proof of Lemma 5: To simplify the notation, for $\epsilon > 0$ define

$$a_n(\epsilon) := \frac{\log |\mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(\mu,\epsilon)| - \|\vec{m}^{(n)}\| \log n}{n}.$$

Note that there exists a subsequence $\{n_k\}$ such that $\limsup_{n\to\infty} a_n(\epsilon_n) = \lim_{k\to\infty} a_{n_k}(\epsilon_{n_k})$. Moreover, there is a further subsequence n_{k_r} such that for all $x, x' \in \Xi$, there exists $\bar{d}_{x,x'} \in [0,\infty]$ where

$$\frac{m^{(n_{k_r})}(x,x')}{n_{k_r}} \to \bar{d}_{x,x'}, \qquad x \neq x'; \tag{16a}$$

$$\frac{2m^{(n_{k_r})}(x,x)}{n_{k_r}} \to \bar{d}_{x,x}. \tag{16b}$$

Observe that since $a_{n_k}(\epsilon_{n_k})$ is convergent, it suffices that we focus on the subsequence $\{n_{k_r}\}$ and show that $\lim a_{n_{k_r}}(\epsilon_{n_{k_r}}) \le \Sigma(\mu)$. Note that due to conditions (10a) and (10b), we have $\bar{d}_{x,x'} \ge \deg_{x,x'}(\mu)$ for all $x, x' \in \Xi$. Therefore, there are two possible cases: either $\bar{d}_{x,x'} = \deg_{x,x'}(\mu)$ for all $x, x' \in \Xi$, or there exist $x, x' \in \Xi$ such that $\bar{d}_{x,x'} > \deg_{x,x'}(\mu)$. To simplify the notation, without loss of generality, we may assume that the subsequence n_{k_r} is the whole sequence, i.e. $a_n(\epsilon_n)$ is convergent, and (16a) and (16b) hold for the whole sequence.

<u>Case 1:</u> $\bar{d}_{x,x'} = \deg_{x,x'}(\mu)$ for all $x, x' \in \Xi$. We define edge and vertex mark count vectors $\vec{\widetilde{m}}^{(n)} = (\widetilde{m}^{(n)}(x,x') : x, x' \in \Xi)$ and $\vec{\widetilde{u}}^{(n)} = (\widetilde{u}^{(n)}(\theta) : \theta \in \Theta)$ as follows. For $x, x' \in \Xi$, define $\widetilde{m}^{(n)}(x,x')$ to be $m^{(n)}(x,x')$ if $\deg_{x,x'}(\mu) > 0$ and 0 otherwise. Also, fix some $\theta_0 \in \Theta$ such that $\Pi_{\theta_0}(\mu) > 0$. For $\theta \in \Theta$, define

$$\widetilde{u}^{(n)}(\theta) := \begin{cases} 0, & \Pi_\theta(\mu) = 0, \\ u^{(n)}(\theta), & \Pi_\theta(\mu) > 0, \theta \neq \theta_0, \\ u^{(n)}(\theta_0) + \sum_{\theta' : \Pi_{\theta'}(\mu) = 0} u^{(n)}(\theta') & \theta = \theta_0. \end{cases}$$

Note that, by construction and from (16a) and (16b), the sequences $\vec{\widetilde{m}}^{(n)}$ and $\vec{\widetilde{u}}^{(n)}$ are adapted to $(\vec{\deg}(\mu), \vec{\Pi}(\mu))$. Also, $m^{(n)}(x,x') \ge \widetilde{m}^{(n)}(x,x')$ for all n and all $x, x' \in \Xi$.

Now, fix $\epsilon > 0$, and pick an integer h such that $1/(1+h) < \epsilon$. Define B to be the set of $[G,o] \in \bar{\mathcal{G}}_*$ such that either for some $x, x' \in \Xi$ with $\deg_{x,x'}(\mu) = 0$ there exists an edge in $[G,o]_h$ with pair of marks x, x', or, for some $\theta \in \Theta$ with $\Pi_\theta(\mu) = 0$, there exists a vertex in $[G,o]_h$ with mark θ. Then, from Lemma 2, we have $\mu(B) = 0$. On the other hand, for n large enough that $\epsilon_n < 1/(1+h)$, we have $B^{\epsilon_n} = B$. Hence, for large enough n and $G \in \mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(\mu, \epsilon_n)$, we have $U(G)(B) \le \epsilon_n$. For such G, we construct a marked graph $\widetilde{G}$ which is obtained from G by removing all edges which have their pair of marks x, x' with $\deg_{x,x'}(\mu) = 0$. Moreover, if a vertex v in G has mark θ with $\Pi_\theta(\mu) = 0$, we change its mark to θ_0 in $\widetilde{G}$, with the θ_0 defined above. Note that $\widetilde{G} \in \mathcal{G}^{(n)}_{\vec{\widetilde{m}}^{(n)},\vec{\widetilde{u}}^{(n)}}$. Furthermore, since $U(G)(B) \le \epsilon_n$, the number of vertices v in G such that $(G,v)_h \equiv (\widetilde{G},v)_h$ is at least $n(1-\epsilon_n)$. Consequently, using Lemma 3, when n is large enough that $\epsilon_n < \epsilon$, we have $d_{\mathrm{LP}}(G,\widetilde{G}) \le \max\{1/(1+h), \epsilon_n\} < \epsilon$. This means that $\widetilde{G} \in \mathcal{G}^{(n)}_{\vec{\widetilde{m}}^{(n)},\vec{\widetilde{u}}^{(n)}}(\mu, \epsilon_n + \epsilon) \subset \mathcal{G}^{(n)}_{\vec{\widetilde{m}}^{(n)},\vec{\widetilde{u}}^{(n)}}(\mu, 2\epsilon)$.

Motivated by this discussion, for n large enough, we have

$$\begin{aligned} |\mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(\mu,\epsilon_n)| &\le |\mathcal{G}^{(n)}_{\vec{\widetilde{m}}^{(n)},\vec{\widetilde{u}}^{(n)}}(\mu,2\epsilon)| \\ &\times \prod_{\substack{x \le x' \in \Xi \\ \deg_{x,x'}(\mu)=0}} |\mathcal{G}_{n, m^{(n)}(x,x') - \widetilde{m}^{(n)}(x,x')}| \times 2^{m^{(n)}(x,x') - \widetilde{m}^{(n)}(x,x')} \\ &\times \prod_{\theta \in \Theta} \binom{n}{|u^{(n)}(\theta) - \widetilde{u}^{(n)}(\theta)|}. \end{aligned} \tag{17}$$

Here we have assumed that, since Ξ is finite, it is an ordered set. For $x \le x' \in \Xi$ with $\deg_{x,x'}(\mu) = 0$, using Lemma 4, we have

$$\begin{aligned} &\log\left(|\mathcal{G}_{n, m^{(n)}(x,x') - \widetilde{m}^{(n)}(x,x')}| \times 2^{m^{(n)}(x,x') - \widetilde{m}^{(n)}(x,x')} \right) \\ &\quad \le (m^{(n)}(x,x') - \widetilde{m}^{(n)}(x,x')) \log n \\ &\qquad + ns\left(\frac{2(m^{(n)}(x,x') - \widetilde{m}^{(n)}(x,x'))}{n} \right) \\ &\qquad + (m^{(n)}(x,x') - \widetilde{m}^{(n)}(x,x')) \log 2. \end{aligned}$$

Note that, for all $x, x' \in \Xi$, $\frac{1}{n}(m^{(n)}(x,x') - \widetilde{m}^{(n)}(x,x')) \to 0$. Also, for all $\theta \in \Theta$, $\frac{1}{n}|u^{(n)}(\theta) - \widetilde{u}^{(n)}(\theta)| \to 0$. Additionally, $s(y) \to 0$ as $y \to 0$. Using these in (17) and simplifying, we get

$$\begin{aligned} &\limsup_{n\to\infty} \frac{\log |\mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(\mu,\epsilon_n)| - \|\vec{m}^{(n)}\|_1 \log n}{n} \\ &\quad \le \limsup_{n\to\infty} \frac{\log |\mathcal{G}^{(n)}_{\vec{\widetilde{m}}^{(n)},\vec{\widetilde{u}}^{(n)}}(\mu,2\epsilon)| - \|\vec{\widetilde{m}}^{(n)}\|_1 \log n}{n} \\ &\quad \le \overline{\Sigma}_{\vec{\deg}(\mu),\vec{\Pi}(\mu)}(\mu, 2\epsilon)|_{\vec{\widetilde{m}}^{(n)},\vec{\widetilde{u}}^{(n)}}, \end{aligned}$$

where the last inequality employs the fact that, by construction, $\vec{\widetilde{m}}^{(n)}$ and $\vec{\widetilde{u}}^{(n)}$ are adapted to $(\vec{\deg}(\mu), \vec{\Pi}(\mu))$. The above inequality holds for all $\epsilon > 0$. Therefore, from Theorem 2, as $\epsilon \to 0$, the right hand side converges to $\Sigma(\mu)$. This completes the proof for this case.

<u>Case 2:</u> $\bar{d}_{x,x'} > \deg_{x,x'}(\mu)$ for some $x, x' \in \Xi$. Let $\bar{d} := \sum_{x,x' \in \Xi} \bar{d}_{x,x'}$. Note that $\bar{d} > \deg(\mu)$. First, assume that $\bar{d} = \infty$. Observe that

$$\begin{aligned} |\mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(\mu,\epsilon)| &\le |\mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}| \\ &\le |\Theta|^n \prod_{x \le x' \in \Xi} |\mathcal{G}_{n, m^{(n)}(x,x')}| 2^{m^{(n)}(x,x')}. \end{aligned}$$

Using Lemma 4,

$$|\mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(\mu,\epsilon)| \leq n\log|\Theta| + \|\vec{m}^{(n)}\|_1 \log n + n\sum_{x\leq x'\in\Xi}\left(s\left(\frac{2m^{(n)}(x,x')}{n}\right) + m^{(n)}(x,x')\log 2\right) \tag{18}$$

$$= n\log|\Theta| + \|\vec{m}^{(n)}\|_1\log n + 2n\sum_{x\leq x'\in\Xi} s\left(\frac{m^{(n)}(x,x')}{n}\right).$$

Since we have assumed $\bar{d}=\infty$, there exist $\bar{x}\leq\bar{x}'\in\Xi$ such that $\bar{d}_{\bar{x},\bar{x}'}=\infty$. Therefore, (16a) and (16b) imply that $m^{(n)}(\bar{x},\bar{x}')/n\to\infty$. On the other hand, $s(y)\to-\infty$ as $y\to\infty$. Using these in (18), we get $\limsup_{n\to\infty} a_n(\epsilon_n) = -\infty$ which completes the proof. Therefore, it remains to consider the case $\bar{d}<\infty$.

Let $\tilde{\mu}\in\mathcal{P}(\bar{\mathcal{T}}_*)$ be the law of $[\mathsf{UM}(T),o]$ when $[T,o]$ has law μ, and let $m_n := \|\vec{m}^{(n)}\|_1$. From Lemma 10, if $G\in\mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(\mu,\epsilon_n)$, we have $\mathsf{UM}(G)\in\mathcal{G}_{n,m_n}(\tilde{\mu},\epsilon_n)$. Moreover, by finding an upper bound on the number of possible ways to mark vertices and edges for an unmarked graph in $\mathcal{G}_{n,m_n}$, we have

$$|\mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(\mu,\epsilon_n)| \leq |\mathcal{G}_{n,m_n}(\tilde{\mu},\epsilon_n)|\times|\Theta|^n \times\frac{m_n!}{\prod_{x\leq x'} m^{(n)}(x,x')!}\times 2^{m_n}. \tag{19}$$

Note that $m_n/n\to\bar{d}/2<\infty$, and $m^{(n)}(x,x')/n$ converges to $\bar{d}_{x,x'}/2$ when $x=x'$, and $\bar{d}_{x,x'}$ when $x\neq x'$. Hence,

$$\lim_{n\to\infty}\frac{1}{n}\log\left(|\Theta|^n\times\frac{m_n!}{\prod_{x\leq x'}m^{(n)}(x,x')!}\times 2^{m_n}\right) = \log|\Theta| + \sum_{x<x'\in\Xi}\bar{d}_{x,x'}\log\frac{\bar{d}}{\bar{d}_{x,x'}} + \sum_{x\in\Xi}\frac{\bar{d}_{x,x}}{2}\log\frac{2\bar{d}}{\bar{d}_{x,x}} =: \alpha.$$

Note that, as $\bar{d}<\infty$, α is a bounded real number. Also, since $\epsilon_n\to 0$, for each $\epsilon>0$ fixed we have $\epsilon_n<\epsilon$ for n large enough. Putting these in (19), we get

$$\limsup_{n\to\infty} a_n(\epsilon_n) \leq \alpha + \limsup_{n\to\infty}\frac{\log|\mathcal{G}_{n,m_n}(\tilde{\mu},\epsilon)| - m_n\log n}{n}. \tag{20}$$

Note that $\deg(\tilde{\mu}) = \deg(\mu)$. Moreover, $m_n/n\to\bar{d}>\deg(\mu)=\deg(\tilde{\mu})>0$. Furthermore, our notion of marked BC entropy reduces to the unmarked BC entropy of [13] when Θ and Ξ have cardinality one. Therefore, since $\bar{d}\neq\deg(\tilde{\mu})$, from part 3 of Theorem 1, (or equivalently from part 3 of Theorem 1.2 in [13]), the right hand side of (20) goes to $-\infty$ as $\epsilon\to 0$. Therefore, $\limsup_{n\to\infty}a_n(\epsilon_n)=-\infty$, which completes the proof. □

Proof of Lemma 6: Let μ_n and $\tilde{\mu}_n$ denote $U(G^{(n)})$ and $U((G^{(n)})^{\Delta_n})$ respectively. For an integer $k\geq 0$ and a marked rooted tree (T,i) with depth at most k, define

$$A^k_{(T,i)} := \{[G,o]\in\bar{\mathcal{G}}_* : (G,o)_k \equiv (T,i)\},$$

as in (4). From Lemma 1 in Section II-A, in order to show $\tilde{\mu}_n\Rightarrow\mu$, it suffices to show that $\tilde{\mu}_n(A^k_{(T,i)})\to\mu(A^k_{(T,i)})$ for all such k and (T,i). We will now do this. Fix some integer $k\geq 0$

throughout the rest of the discussion. For an integer Δ, define

$$B^\Delta := \{[G,o]\in\bar{\mathcal{G}}_* : \deg_G(j)>\Delta \text{ for some } j \text{ with distance at most } k+1 \text{ from } o\}.$$

With this, we have

$$\begin{aligned}
\lim_{\Delta\to\infty}\mu((B^\Delta)^c) &= \mu\left(\bigcup_{\Delta=1}^\infty (B^\Delta)^c\right)\\
&= \mu\Big(\deg_G(j)<\infty \text{ for all } j \text{ with distance at most } k+1 \text{ from } o\Big)\\
&= 1,
\end{aligned} \tag{21}$$

where the last equality comes from the fact that all graphs in $\bar{\mathcal{G}}_*$ are locally finite. Next, define

$$\begin{aligned}
C_n &:= \{i\in V(G^{(n)}) : \deg_{G^{(n)}}(j)\leq\Delta_n \text{ for all nodes } j \text{ in } G^{(n)} \text{ with distance at most } k+1 \text{ from } i\}\\
&= \{i\in V(G^{(n)}) : [G^{(n)},i]\in(B^{\Delta_n})^c\}.
\end{aligned}$$

Now, since $V(G^{(n)}) = \{1,\dots,n\}$, we have

$$\begin{aligned}
\tilde{\mu}_n(A_{(T,i)}) &= \frac{1}{n}\sum_{j=1}^n \mathbb{1}\left[\left(\left(G^{(n)}\right)^{\Delta_n},j\right)_k\equiv(T,i)\right]\\
&= \frac{1}{n}\sum_{j\in C_n}\mathbb{1}\left[\left(\left(G^{(n)}\right)^{\Delta_n},j\right)_k\equiv(T,i)\right] + \frac{1}{n}\sum_{j\in C_n^c}\mathbb{1}\left[\left(\left(G^{(n)}\right)^{\Delta_n},j\right)_k\equiv(T,i)\right]\\
&= \frac{1}{n}\sum_{j\in C_n}\mathbb{1}\left[(G^{(n)},j)_k\equiv(T,i)\right] + \frac{1}{n}\sum_{j\in C_n^c}\mathbb{1}\left[\left(\left(G^{(n)}\right)^{\Delta_n},j\right)_k\equiv(T,i)\right].
\end{aligned}$$

Comparing this to

$$\mu_n(A_{(T,i)}) = \frac{1}{n}\sum_{j=1}^n\mathbb{1}\left[(G^{(n)},j)_k\equiv(T,i)\right],$$

we realize that

$$|\tilde{\mu}_n(A_{(T,i)}) - \mu_n(A_{(T,i)})| \leq \frac{1}{n}|C_n^c| = \mu_n(B^{\Delta_n}).$$

Now, fix an integer $\Delta>0$. Since $\Delta_n\to\infty$, we have $\Delta<\Delta_n$ for n large enough. Moreover, as B^Δ is closed,

$$\limsup_{n\to\infty}|\tilde{\mu}_n(A_{(T,i)}) - \mu_n(A_{(T,i)})| \leq \limsup_{n\to\infty}\mu_n(B^\Delta) \leq \mu(B^\Delta).$$

This is true for all $\Delta > 0$; therefore, sending Δ to infinity and using (21), we have $|\tilde{\mu}_n(A_{(T,i)}) - \mu_n(A_{(T,i)})| \to 0$. On the other hand, we have assumed that $\mu_n \Rightarrow \mu$. Therefore, Lemma 1 implies that $\mu_n(A_{(T,i)}) \to \mu(A_{(T,i)})$. This means that $\tilde{\mu}_n(A_{(T,i)}) \to \mu(A_{(T,i)})$. Since this is true for all k and (T,i), Lemma 1 implies that $\tilde{\mu}_n \Rightarrow \mu$, which completes the proof. $\square$

Proof of Lemma 7: In order to count $|\mathcal{A}_{k_n,\Delta_n}|$, note that, for $\Delta_n \geq 2$, a rooted graph of depth at most k_n and maximum degree at most Δ_n has at most

$$1 + \Delta_n + \Delta_n^2 + \cdots + \Delta_n^{k_n} \leq \Delta_n^{k_n+1},$$

many vertices, each of which has $|\Theta|$ many choices for the vertex mark. On the other hand, such a graph can have at most $\Delta_n^{2(k_n+1)}$ many edges, each of which can be present or not, and, if present, has $|\Xi|^2$ many choices for the edge mark. Consequently,

$$|\mathcal{A}_{k_n,\Delta_n}| \leq (1 + |\Xi|^2)^{\Delta_n^{2(k_n+1)}} |\Theta|^{\Delta_n^{k_n+1}}.$$

Therefore,

$$\begin{aligned}\log|\mathcal{A}_{k_n,\Delta_n}| &\leq \Delta_n^{2(1+k_n)} \log(1+|\Xi|^2) + \Delta_n^{1+k_n} \log|\Theta| \\ &\leq \Delta_n^{2(1+k_n)} \log(|\Theta|(1+|\Xi|^2)) \\ &\leq \Delta_n^{4k_n} \log(|\Theta|(1+|\Xi|^2)),\end{aligned}$$

where the last inequality holds for n large enough that $k_n \geq 1$. Note that in order to show $|\mathcal{A}_{k_n,\Delta}| = o(n/\log n)$, it suffices to show that $\log|\mathcal{A}_{k_n,\Delta_n}| - \log(n/\log n) \to -\infty$. Motivated by the above inequality, we observe that this is satisfied if $\Delta_n^{4k_n} = O(\sqrt{\log n})$. Suppose now that $\Delta_n \leq \log\log n$ and $k_n \leq \sqrt{\log\log n}$. For n large enough, we have

$$\begin{aligned}\log(\Delta_n^{4k_n}) &\leq 4\sqrt{\log\log n}\log\log\log n \\ &\leq \frac{1}{2}\sqrt{\log\log n}\sqrt{\log\log n} \\ &= \frac{1}{2}\log\log n.\end{aligned}$$

This means that for n large enough we have $\Delta_n^{4k_n} \leq \sqrt{\log n}$. This completes the proof. $\square$

Proof of Lemma 8: For $\Delta > 0$, define $B_\Delta \subset \bar{\mathcal{G}}_*$ as

$$\begin{aligned}B_\Delta := \{[G,o] \in \bar{\mathcal{G}}_* : \deg_G(o) \leq \Delta \text{ and} \\ \deg_G(i) \leq \Delta \text{ for all } i \sim_G o\}.\end{aligned}$$

Since all graphs in $\bar{\mathcal{G}}_*$ are locally finite, we have $\mu(B_\Delta) \to 1$ as $\Delta \to \infty$. On the other hand,

$$\frac{|R_n|}{n} = U(G^{(n)})(B^c_{\Delta_n}).$$

Since $\Delta_n \to \infty$, we have $B_\Delta \subseteq B_{\Delta_n}$ for n large enough, for any value of Δ. Moreover, B_Δ is both open and closed. Therefore,

$$\frac{|R_n|}{n} = U(G^{(n)})(B^c_{\Delta_n}) \leq U(G^{(n)})(B^c_\Delta) \to \mu(B^c_\Delta).$$

But this is true for all Δ, and $\mu(B_\Delta) \to 1$ as $\Delta \to \infty$. Consequently, $|R_n|/n \to 0$, and the proof is complete. $\square$

References

[1] P. Boldi and S. Vigna, "The Webgraph framework I: Compression techniques," in *Proc. 13th Conf. World Wide Web (WWW)*. New York, NY, USA: ACM, 2004, pp. 595–602.

[2] P. Boldi, M. Rosa, M. Santini, and S. Vigna, "Layered label propagation: A multiresolution coordinate-free ordering for compressing social networks," in *Proc. 20th Int. Conf. World Wide Web (WWW)*. New York, NY, USA: ACM, 2011, pp. 587–596.

[3] Y. Choi and W. Szpankowski, "Compression of graphical structures: Fundamental limits, algorithms, and experiments," *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 620–638, Feb. 2012.

[4] J. C. Kieffer, E.-H. Yang, and W. Szpankowski, "Structural complexity of random binary trees," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2009, pp. 635–639.

[5] J. Zhang, E.-H. Yang, and J. C. Kieffer, "A universal grammar-based code for lossless compression of binary trees," *IEEE Trans. Inf. Theory*, vol. 60, no. 3, pp. 1373–1386, Mar. 2014.

[6] A. Magner, K. Turowski, and W. Szpankowski, "Lossless compression of binary trees with correlated vertex names," *IEEE Trans. Inf. Theory*, vol. 64, no. 9, pp. 6070–6080, Sep. 2018.

[7] M. Ganardi, D. Hucke, M. Lohrey, and L. Seelbach Benkner, "Universal tree source coding using grammar-based compression," *IEEE Trans. Inf. Theory*, vol. 65, no. 10, pp. 6399–6413, Oct. 2019.

[8] S. Basu and L. R. Varshney, "Universal source coding of deep neural networks," in *Proc. Data Compress. Conf. (DCC)*, Apr. 2017, pp. 310–319.

[9] D. J. Aldous and N. Ross, "Entropy of some models of sparse random graphs with vertex-names," *Probab. Eng. Inf. Sci.*, vol. 28, no. 2, pp. 145–168, Apr. 2014.

[10] I. Benjamini and O. Schramm, "Recurrence of distributional limits of finite planar graphs," *Electron. J. Probab.*, vol. 6, no. 23, p. 13, 2001, doi: 10.1214/EJP.v6-96.

[11] D. Aldous and J. M. Steele, "The objective method: Probabilistic combinatorial optimization and local weak convergence," in *Probability on Discrete Structures*. Berlin, Germany: Springer, 2004, pp. 1–72.

[12] D. Aldous and R. Lyons, "Processes on unimodular random networks," *Electron. J. Probab.*, vol. 12, no. 54, pp. 1454–1508, 2007.

[13] C. Bordenave and P. Caputo, "Large deviations of empirical neighborhood distribution in sparse random graphs," *Probab. Theory Rel. Fields*, vol. 163, nos. 1–2, pp. 149–222, Oct. 2015.

[14] P. Delgosha and V. Anantharam, "A notion of entropy for stochastic processes on marked rooted graphs," 2019, *arXiv:1908.00964*. [Online]. Available: http://arxiv.org/abs/1908.00964

[15] I. Benjamini and O. Schramm, "Recurrence of distributional limits of finite planar graphs," in *Selected Works of Oded Schramm*. New York, NY, USA: Springer, 2011, pp. 533–545.

[16] P. Billingsley, *Convergence of Probability Measures*. Hoboken, NJ, USA: Wiley, 2013.

Payam Delgosha (Member, IEEE) received the B.Sc. degree in electrical engineering and pure mathematics and the M.Sc. degree in electrical engineering from the Sharif University of Technology, Tehran, Iran, in 2012 and 2014, respectively. He is currently pursuing the Ph.D. degree with the Department of Electrical Engineering and Computer Sciences (EECS), University of California at Berkeley.

Venkat Anantharam (Fellow, IEEE) received the B.Tech. degree in electronics from the Indian Institute of Technology Madras (IIT-M) in 1980, and the M.A. and C.Phil. degrees in mathematics and the M.S. and Ph.D. degrees in electrical engineering from the University of California at Berkeley (UCB) in 1983, 1984, 1982, and 1986, respectively.

From 1986 to 1994, he was on the Faculty of the School of Electrical Engineering, Cornell University, Ithaca, NY, USA. Since 1994, he has been with the Department of Electrical Engineering and Computer Sciences, UCB. He received the Philips India Medal and the President of India Gold Medal from IIT-M in 1980, and the NSF Presidential Young Investigator Award from 1988 to 1993. He also received the Distinguished Alumnus Award from IIT-M in 2008. He was a co-recipient of the 1998 Prize Paper Award of the IEEE Information Theory Society and the 2000 Stephen O. Rice Prize Paper Award of the IEEE Communications Theory Society.