Counting Graphs with a Given Degree Sequence: An Information-theoretic Perspective

Shahar Stein Ioushua and Ofer Shayevitz

Abstract—**We revisit the problem of counting the number of directed graphs with a specified degree sequence, which was recently studied and solved by Barvinok using generating functions and convex duality techniques. We describe a systematic information-theoretic approach to this type of problems, based on studying invariant distributions and establishing suitable continuity and concentration properties. Our techniques recover and shed further light on Barvinok's solution, and may be applicable in other similar problems. As a simple example, we also apply our approach to estimating the number of undirected graphs with a given degree sequence. In particular, we show this number is approximately given by the square root of the number of associated directed graphs, whose input and output degree sequences are equal to that of the undirected graph.**

I. Introduction

The classical method-of-types by Csiszár and Körner [1] is a powerful tool with many applications in information theory, combinatorics, and computer science. It is based on the idea of partitioning the space of length-n sequences from a discrete alphabet into type classes of sequences with the same empirical distribution, also called a type. A basic result in the theory of such types is that up to logarithmic terms, the logarithm of a type class size is equal to the entropy of a length-n i.i.d. sample drawn from the empirical distribution. In this paper, we discuss a similar question in a more complex setting where n-sequences are replaced with graphs over n vertices. Specifically, we are interested in counting the number of (directed or undirected) graphs with a prescribed vertex degree sequence, which we call an *edge-type*.

This problem has a fairly long history, which we only briefly highlight. Unlike the sequence case where essentially all possible types are non-empty, in the graph setting not all possible degree sequences correspond to some graph. The famous Gale-Ryser theorem [2], [3] precisely characterizes all non-empty edge-types through majorization conditions. Bounds on the size of edge-type classes for sparse graphs have been obtained by various authors, see e.g. [4]. Recently, the size of a general (dense) edge-type class has been accurately characterized by Barvinok [5] as a solution to a maximum entropy problem, up to terms that are sub-exponential in the number of edges. Bustin and Shayevitz [6] have built on Barvinok's results to bound the optimal lossy compression rate of graphs under an L^{∞} constraint on the degree sequence reconstruction.

In this paper, we revisit the problem of counting the number of graphs with a given edge-type, from an information-theoretic perspective. While Barvinok's results provide a complete characterization (at least in the directed graphs case), his derivations are based on generating functions and convex duality arguments, and as such, shed relatively little light on the operational meaning of the solution. In contrast, our approach is more transparent and also provides a level of abstraction that possibly lends itself to other problems of the same nature. We demonstrate this by applying our techniques to the (relatively simple) case of undirected graphs, which was not handled before (although it is amenable to Barvinok's technique as well).

Let us briefly elaborate on our general approach and the form it takes in the simple sequence case, as well as in the graph case. Given some partition into type classes, we first identify an operation under which the type is *invariant*, and which generates the entire type class. In the sequence case, this operation is trivially a coordinate permutation, and in the graph case it is known as an *interchange*. Now, we find the collection of distributions that are invariant under this operation, which in the sequence case are trivially the i.i.d. distributions, and in the graph case have a more complex structure discussed in the sequel. By construction, these distributions are uniform inside each type. Now, we would like to claim that the logarithm of the type class size is roughly given by the entropy of the invariant distribution whose *expected type* is the correct one. This fact can be easily verified directly in the sequence case, but requires subtle structure in general. Specifically, one needs to show *concentration* around the mean, i.e., that invariant distributions give relatively high probability to their expected type, which in particular requires a *continuity* property to hold, in the sense that a small distance between types implies that their respective classes have similar sizes. Proving these properties for sequences is simple, but is nontrivial in the graph case; we prove this by adapting a technique of Krause [7] used in an alternative proof of the Gale-Ryser Theorem, in conjunction with information-theoretic and majorization-based arguments.

II. Preliminaries

A. Graphs and their edge-types

Let G be a directed graph (*digraph*) with vertex set $[n] = \{1,\dots,n\}$. We write $i \to j$ to denote there exists a directed edge from i to j, and write $i \nrightarrow j$ otherwise. We define $\boldsymbol{r}_G$ and $\boldsymbol{c}_G$ to be the outgoing edge degree (*out-deg*) and ingoing edge degree (*in-deg*) vectors associated with G, i.e.,

$$\boldsymbol{r}_G\left(i\right) \triangleq \sum_{j\in[n]} \mathbb{1}\left(i \to j\right), \quad \boldsymbol{c}_G\left(i\right) \triangleq \sum_{j\in[n]} \mathbb{1}\left(j \to i\right). \tag{1}$$

Note that $\boldsymbol{r}_G$ and $\boldsymbol{c}_G$ are simply the row and column sums of the adjacency matrix $\boldsymbol{A}_G$ associated with the graph, respectively. The Hamming distance between the adjacency matrices of the two digraphs G and G' denoted as $d_H(G,G')$. We define the *edge-type* of G to be the pair $T_G \triangleq (\boldsymbol{r}_G, \boldsymbol{c}_G)$. The *edge-type class* $\Lambda\left(T\right)$ is the set of all digraphs with edge-type T.

In a analogous manner, we also consider undirected graphs and write $i \sim j$ (resp. $i \nsim j$) to denote an edge (resp. no edge) between i and j. While not essential for our results, we assume for simplicity that the graphs have no self-loops. We define $\boldsymbol{e}_G$ to be the degree vector associated with G, i.e., $\boldsymbol{e}_G\left(i\right) \triangleq \sum_{j\in[n]} \mathbb{1}(i \sim j)$, which is the row (or column) sum of the (symmetric) adjacency matrix of G. The (symmetric) edge-type of G is now defined to be $S_G \triangleq \boldsymbol{e}_G$, and the (symmetric) edge-type class $\Lambda(S)$ is defined to be the set of all undirected graphs with edge-type S.

The authors are with the Department of EE–Systems, Tel Aviv University, Tel Aviv, Israel {steinioushua@mail.tau.ac.il, ofersha@eng.tau.ac.il}. This work has been supported by an ISF grant no. 1495/18.

The number of distinct digraphs (resp. graphs) is clearly 2^{n^2} (resp. $2^{n(n-1)/2}$). Using the fact that the in-deg and out-deg of each vertex can take at most $n+1$ distinct values (n values in the undirected case, since we assumed no self-loops), it is easy to see that the number of different edge-type classes is significantly smaller.

Lemma 1. *The number of edge-type classes (resp. symmetric edge-type classes) is at most $2^{2n\log(n+1)}$ (resp. $2^{n\log(n)}$).*

B. Interchanges

Our approach is based on identifying distributions over graph edges, that are uniform inside edge-types. To that end, it is important to study operations under which the edge-type is preserved. One such general operation is the following. Let G be a digraph over $[n]$, and let (i,j,k,ℓ) be any vector of (not necessarily distinct) vertices in G, such that $i\to k$, $i\nrightarrow \ell$, and $j\to\ell$, $j\nrightarrow k$. Let G' be a digraph obtained from G by "crisscrossing" these edges, namely such $i\nrightarrow k$, $i\to\ell$, and $j\nrightarrow\ell$, $j\to k$, and where all other edges remain intact (copied from G). This operation is called an *interchange*. It is easy to check that an interchange preserves the edge type, i.e., that $T_G=T_{G'}$.

The importance of interchanges stems from the following result proved by Ryser, which we adapt to our lingo.

Lemma 2 (Theorem 3.1 [3])**.** *Let $G,G'\in\Lambda(T)$. Then G can be transformed into G' via a finite number of interchanges.*

Sketch of proof. Let $T=(\boldsymbol{r},\boldsymbol{c})$ be a nonempty edge-type (a majorization condition for non-emptiness is given in [3]). Then one can construct a "'canonical" graph in $\Lambda(T)$ in a greedy way, by connecting the $r(1)$ outgoing edges of vertex 1 to the $r(1)$ vertices in $[n]$ with maximal ingoing edge degree, then decreasing 1 from the corresponding entries in $\boldsymbol{c}_G$, and repeating the process with $r(2)$, and so on. Ryser showed that this process indeed terminates in a graph in $\Lambda(T)$, and proved by induction that any graph $G\in\Lambda(T)$ can be transformed into the canonical graph via interchanges. □

C. Random graphs

To count the number of graphs with a given edge-type, we will also make use of random graphs. Specifically, we write that $\mathcal{G}\sim\{p_{ij}\}$ to mean that the random digraph $\mathcal{G}$ contains the directed edge $i\to j$ with probability p_{ij}, and that the directed edges are drawn statistically independent of each other. The associated graph type $T_{\mathcal{G}}$ is hence a random variable, and we write $\mathbb{E}T_{\mathcal{G}}\triangleq(\mathbb{E}\boldsymbol{r}_{\mathcal{G}},\mathbb{E}\boldsymbol{c}_{\mathcal{G}})$ for its expectation, noting also that $\mathbb{E}\boldsymbol{r}_{\mathcal{G}}(i)=\sum_j p_{ij}$, and $\mathbb{E}\boldsymbol{c}_{\mathcal{G}}(i)=\sum_j p_{ji}$. We also write $P_{\mathcal{G}}$ to denote the probability distribution of $\mathcal{G}$, i.e., $P_{\mathcal{G}}(G)=\Pr(\mathcal{G}=G)$. When considering undirected random graphs, $\mathcal{G}\sim\{p_{ij}\}_{i>j}$ means that $i\sim j$ with probability p_{ij} independently over edges. The random edge-type $S_{\mathcal{G}}$ and its expectation $\mathbb{E}S_{\mathcal{G}}$ are defined similarly. We write $H(\mathcal{G})=\sum_{ij}h(p_{ij})$ for the entropy of the distribution $P_{\mathcal{G}}$, with $h(p)\triangleq -p\log p-(1-p)\log(1-p)$ the binary entropy function, and $D(\mathcal{G}\|\mathcal{F})$ for the KL divergence between the distributions $P_{\mathcal{G}}$ and $P_{\mathcal{F}}$.

III. Main results

In this section, we provide an exponentially-tight expression for the size of edge-type classes, which we proceed to prove in the section V. To that end, we are interested in random digraphs that are *interchange-invariant*, namely such that $P_{\mathcal{G}}(G)=P_{\mathcal{G}}(G')$ for any two digraphs G,G' where G' is obtained from G via an interchange. Interchange-invariance plays a major role in our proofs due to the following simple fact, which follows immediately from Lemma 2, and allows us to obtain entropy bounds on the size of edge-type classes.

Corollary 1. *$\mathcal{G}$ is interchange-invariant if and only if it is uniformly distributed inside each nonempty edge-type class.*

Next, we completely characterize the family of all interchange-invariant random digraphs (with independent edges).

Lemma 3. *Let $\mathcal{G}\sim\{p_{ij}\}$ where $p_{ij}\in(0,1)$. Then $\mathcal{G}$ is interchange-invariant if and only if*

$$p_{ij}=\frac{x_iy_j}{1+x_iy_j}\qquad(2)$$

for some positive numbers x_i,y_j.

Proof. Fix any graph G, and suppose G' is obtained from G by an interchange replacing the edges $i\to j$ and $k\to\ell$ with the edges $i\to\ell$ and $k\to j$. Now assume (2) holds. Let

$$q=\frac{1}{p_{ij}p_{kl}(1-p_{il})(1-p_{kj})}\prod_{t\to s}p_{ts}\cdot\prod_{t'\nrightarrow s'}(1-p_{t's'})\qquad(3)$$

be the probability associated with all the edges / missing edges not participating in the interchange (identical for both G and G'). Here the edges are taken with respect to G. Then

$$\begin{aligned}P_{\mathcal{G}}(G)&=q\cdot\frac{x_iy_j}{1+x_iy_j}\cdot\frac{x_ky_\ell}{1+x_ky_\ell}\frac{1}{1+x_iy_\ell}\cdot\frac{1}{1+x_ky_j}&(4)\\&=\frac{q\cdot x_iy_jx_ky_\ell}{(1+x_iy_j)(1+x_iy_\ell)(1+x_ky_j)(1+x_ky_\ell)}&(5)\\&=q\cdot\frac{x_iy_\ell}{1+x_iy_\ell}\cdot\frac{x_ky_j}{1+x_ky_j}\cdot\frac{1}{1+x_iy_j}\frac{1}{1+x_ky_\ell}&(6)\\&=P_{\mathcal{G}}(G').&(7)\end{aligned}$$

Conversely, suppose that the random digraph $\mathcal{G}$ is interchange-invariant. Consider an interchange over distinct vertices. Then $P_{\mathcal{G}}(G)=P_{\mathcal{G}}(G')$ implies that

$$p_{ij}p_{k\ell}(1-p_{i\ell})(1-p_{kj})=p_{i\ell}p_{kj}(1-p_{ij})(1-p_{k\ell}),\qquad(8)$$

or equivalently, that

$$\lambda_{ij}\lambda_{k\ell}=\lambda_{i\ell}\lambda_{kj}.\qquad(9)$$

where $\lambda_{ij}\triangleq\frac{p_{ij}}{(1-p_{ij})}$. Now, consider the matrix $\boldsymbol{L}$ with entries $\{\lambda_{ij}\}$. Since (9) holds for any distinct $\{i,j,k,l\}$, it follows that all 2×2 minors of $\boldsymbol{\lambda}$ vanish, which in turn implies that $\boldsymbol{L}$ is a rank-one matrix. Therefore, it can be written as $\boldsymbol{L}=\boldsymbol{x}\boldsymbol{y}^T$, and we get

$$\lambda_{ij}=x_iy_j.\qquad(10)$$

Since λ_{ij} are all positive, x_i,y_j must have equal signs for all i,j, and can be chosen to be all positive. The relation (2) now easily follows from (10) and the definition of λ_{ij}. □

We are now ready to state the bounds on the cardinality of edge-type classes. We prove this Theorem in Section V.

Theorem 1. *Let T be an edge-type. The size of the edge-type class $\Lambda(T)$ is bounded by*[1]

$$H(\mathcal{G}^*)-O(n^{\frac{3}{2}}\log^{\frac{3}{2}}n)\le\log|\Lambda(T)|\le H(\mathcal{G}^*),\qquad(11)$$

where $\mathcal{G}^*\sim\{p_{ij}\}$ is the unique interchange-invariant digraph (i.e., satisfying (2)) such that $\mathbb{E}T_{\mathcal{G}^*}=T$.

[1]The gap between the bounds is in fact of order $O(n\log n)$. However, we prefer to give the slightly looser bound above since it admits a much more transparent proof.

For undirected graphs, the size of type classes admits a similar form, which is tightly related to the directed case.

Theorem 2. *Let $S = \boldsymbol{e}$ be a symmetric edge-type. The size of the edge-type class $\Lambda(S)$ is bounded by*

$$\frac{1}{2}H\left(\mathcal{G}^*\right) - O(n^{\frac{3}{2}}\log^{\frac{3}{2}} n) \leq \log\left|\Lambda\left(S\right)\right| \leq \frac{1}{2}H\left(\mathcal{G}^*\right), \quad (12)$$

where $\mathcal{G}^* \sim \{p_{ij}\}$ is the unique interchange-invariant digraph (i.e., satisfying (2)) such that $\mathbb{E}T_{\mathcal{G}^*} = (\boldsymbol{e}, \boldsymbol{e})$.

In other words, we conclude that the size of the class associated with a symmetric edge-type $S = \boldsymbol{e}$ is roughly the square root of the class size of its asymmetric counterpart $T = (\boldsymbol{e}, \boldsymbol{e})$, i.e., $|\Lambda(S)| \approx \sqrt{|\Lambda(T)|}$.

IV. Comparison with Barvinok's results

Let us juxtapose Theorem 1 with results obtained using generating functions and convex duality techniques by Barvinok [5]. Adapted to the lingo of our paper, his bound has the same form as ours, where in his case $\mathcal{G}^*$ is given by a solution to a maximum entropy convex optimization problem

$$\mathcal{G}^* = \max_{\mathcal{F}} H\left(\mathcal{F}\right) \quad \text{s.t.} \quad \mathbb{E}T_{\mathcal{F}} = T, \quad (13)$$

and the maximization is taken over *all* random digraphs $\mathcal{F}$ with independent edges. Furthermore, from convex duality he shows that the solution to (13) has the form (2), which here we identify as the unique interchange-invariance property. In the course of our proof, we will see exactly why the two forms of the bound are identical. However, it is instructive to explain this equivalence (less rigorously) from a lossless compression perspective.

Pick any deterministic digraph F. Then F can be represented by indicating its edge-type class T_F using $O(n\log n)$ bits (in light of Lemma 1), and then indicating its index within $\Lambda(T_F)$ using $H(\mathcal{G}_F^*)$ bits, where $\mathcal{G}_F^*$ is the unique interchange-invariant random digraph with $\mathbb{E}T_{\mathcal{G}_F^*} = T_F$ (in light of Theorem 1). Using this representation to compress a general random digraph $\mathcal{F} \sim \{p_{ij}\}$, yields an average code-length of $H(\mathcal{G}_{\mathcal{F}}^*|\mathcal{F}) + O(n\log n)$. This code-length cannot be lower than the entropy of $\mathcal{F}$, hence we obtain the inequality

$$H(\mathcal{F}) \leq H(\mathcal{G}_{\mathcal{F}}^*|\mathcal{F}) + O(n\log n). \quad (14)$$

Now, one can apply standard concentration results to show that $\frac{1}{n}\|T_{\mathcal{F}} - \mathbb{E}T_{\mathcal{F}}\|_\infty = o(1)$ with probability $1 - o(1)$. Hence by continuity of $H(\mathcal{G}^*)$ (which we later discuss) we obtain that $H(\mathcal{G}_{\mathcal{F}}^*|\mathcal{F}) \approx H(\mathcal{G}^*)$, where $\mathcal{G}^*$ is the unique interchange-invariant random digraph with $\mathbb{E}T_{\mathcal{G}^*} = \mathbb{E}T_{\mathcal{F}}$. Thus, we roughly have the inequality (up to negligible factors)

$$H(\mathcal{F}) \lessapprox H(\mathcal{G}^*), \quad (15)$$

which is clearly saturated by interchange-invariant random digraphs, hence the max-entropy form (13) essentially follows.

To summarize, While our bounds coincide with Barvinok's, our approach is markedly different and admits some additional merits. First, we provide an elementary information theoretic proof that sheds light on the structure and symmetries of the problem, in contrast to the more technical and obfuscated proof in [5]. In particular, our proof reveals the fact that the maximal entropy graph associated with an edge-type is in fact the unique interchange-invariant graph with the correct expected type behavior. Finally, we present a systematic approach of identifying distributions that are type-invariant and analyzing their entropy, which on the one hand naturally extends the traditional method-of-types of the vector case, and on the other hand can arguably lend itself to other similar problems, e.g., of analyzing types that depend on more elaborate local structure of graphs.

V. Proof of Main Results

We begin by providing a simple information-theoretic expression for the probability distribution of interchange-invariant random digraphs, adapting and generalizing a result in [6].

Lemma 4 (Theorem 3 in [6])**.** *Let $\mathcal{F} \sim \{p_{ij}\}$ be interchange-invariant. Then for any digraph G*

$$P_{\mathcal{F}}\left(G\right) = 2^{-H(\mathcal{G}) - D(\mathcal{G}\|\mathcal{F})}, \quad (16)$$

where $\mathcal{G} \sim \{q_{ij}\}$ is any random digraph satisfying $\mathbb{E}T_{\mathcal{G}} = T_G$. Specifically, $P_{\mathcal{F}}(F) = 2^{-H(\mathcal{F})}$ for any F with $T_F = \mathbb{E}T_{\mathcal{F}}$.

Proof. Recall that interchange-invariance implies that $p_{ij} = x_i y_j/(1 + x_i y_j)$ for some positive x_i, y_j. We find it useful to equivalently write

$$\frac{p_{ij}}{1 - p_{ij}} = 2^{u_i + v_j}, \quad (17)$$

for some u_i, v_j. Write $a_{ij} = \mathbb{1}\left(i \to j\right)$ and let $T_G = (\boldsymbol{r}, \boldsymbol{c})$ be the edge-type of G. Note that since $\mathbb{E}T_{\mathcal{G}} = T_G$, we have that

$$\sum_j q_{ij} = \sum_j a_{ij} = r_i \quad (18)$$

$$\sum_i q_{ij} = \sum_i a_{ij} = c_j. \quad (19)$$

Now write

$$\begin{aligned}
&\log P_{\mathcal{F}}\left(G\right) && (20)\\
&= \sum_{ij} a_{ij}\log p_{ij} + (1 - a_{ij})\log(1 - p_{ij}) && (21)\\
&= \sum_{ij} a_{ij}\log\frac{p_{ij}}{1 - p_{ij}} + \log(1 - p_{ij}) && (22)\\
&= \sum_{ij} a_{ij}(u_i + v_j) + \log(1 - p_{ij}) && (23)\\
&= \sum_i u_i\sum_j a_{ij} + \sum_j v_j\sum_i a_{ij} + \sum_{ij}\log(1 - p_{ij}) && (24)\\
&= \sum_i u_i\sum_j q_{ij} + \sum_j v_j\sum_i q_{ij} + \sum_{ij}\log(1 - p_{ij}) && (25)\\
&= \sum_{ij} q_{ij}(u_i + v_j) + \log(1 - p_{ij}) && (26)\\
&= \sum_{ij} q_{ij}\log p_{ij} + (1 - q_{ij})\log(1 - p_{ij}) && (27)\\
&= -H\left(\mathcal{G}\right) - D\left(\mathcal{G} \parallel \mathcal{F}\right). && (28)
\end{aligned}$$

□

Before we proceed, we need the following fact, whose proof is relegated to the Appendix.

Lemma 5. *For any edge-type T there exists a unique interchange-invariant digraph $\mathcal{G}$ satisfying $\mathbb{E}T_{\mathcal{G}} = T$.*

We can now easily prove the upper bound in (11), in a way similar to the vector case. $\mathcal{G}$ be interchange-invariant such that $\mathbb{E}T_{\mathcal{G}} = T$. Then

$$1 \geq P_{\mathcal{G}^*}(\Lambda(T)) = \sum_{G\in\Lambda(T)} P_{\mathcal{G}^*}(G) = |\Lambda(T)| \cdot 2^{-H(\mathcal{G}^*)}, \quad (29)$$

where we have used Lemma 4 in the last equality, and the upper bound follows.

We proceed to handle the lower bound in (11). To that end, we prove the following concentration lemma, which says that the probability given to an edge-type by its associated interchange-invariant distribution is relatively high.

Lemma 6 (concentration)**.** *Let T be an edge-type, $\mathcal{G}^* \sim \{p_{ij}\}$ an interchange-invariant digraph with $\mathbb{E}T_{\mathcal{G}^*} = T$. Then*

$$-\log P_{\mathcal{G}^*}(\Lambda(T)) = O(n^{\frac{3}{2}} \log^{\frac{3}{2}} n) \quad (30)$$

Note that once this lemma is proved, the lower bound in (11) easily follows from $\log|\Lambda(T)| = H(\mathcal{G}^*) + \log P_{\mathcal{G}^*}(\Lambda(T))$. In order to prove this lemma, we first need to establish a continuity result, showing that if two edge-types are close in L^1 then the sizes of the respective edge-type classes are not too far apart.

Lemma 7 (continuity)**.** *Let $T = (\boldsymbol{r}_1, \boldsymbol{c}_1)$ and $T_2 = (\boldsymbol{r}_2, \boldsymbol{c}_2)$ be two edge-types and suppose that*

$$\|\boldsymbol{r}_1 - \boldsymbol{r}_2\|_1 \le m, \quad \|\boldsymbol{c}_1 - \boldsymbol{c}_2\|_1 \le m, \quad (31)$$

for some integer m. Then

$$\left|\log \frac{|\Lambda(T_1)|}{|\Lambda(T_2)|}\right| \le 36m \log n. \quad (32)$$

We prove Lemma 7 in Section VI.

Proof of Lemma 6. Let T' be the edge-type with maximal probability under $P_{\mathcal{G}^*}$, and $\mathcal{F} \sim \{q_{ij}\}$ an interchange-invariant digraph with $\mathbb{E}T_{\mathcal{F}} = T'$. Then

$$(n+1)^{-2n} \le P_{\mathcal{G}^*}(\Lambda(T')) = P_{\mathcal{F}}(\Lambda(T')) \cdot 2^{-D(\mathcal{F}\|\mathcal{G}^*)} \quad (33)$$

where we have used Lemma 1 and Lemma 4. Hence we conclude that

$$D(\mathcal{F} \parallel \mathcal{G}^*) \le 2n \log(n+1). \quad (34)$$

It follows that

$$\sqrt{2n\log(n+1)} \ge \sqrt{\sum_{ij} D(q_{ij} \parallel p_{ij})} \underset{(*)}{\ge} \sqrt{\frac{2}{\ln 2}\sum_{ij} |q_{ij} - p_{ij}|^2}$$

$$\underset{(**)}{\ge} \sqrt{\frac{2}{n^2 \ln 2}} \sum_{ij} |q_{ij} - p_{ij}| \ge \sqrt{\frac{2}{n^2 \ln 2}} \sum_{i} |\tilde{r}_i - r_i| \quad (35)$$

where $(*)$ is due to Pinsker's inequality and $(**)$ follows since $\|\boldsymbol{x}\|_1 \le \sqrt{m}\|\boldsymbol{x}\|_2$ for any $\boldsymbol{x} \in \mathbb{R}^m$. Therefore, we get

$$\max\{\|\tilde{\boldsymbol{c}} - \boldsymbol{c}\|_1, \|\tilde{\boldsymbol{r}} - \boldsymbol{r}\|_1\} \le n^{\frac{3}{2}} \sqrt{\log(n+1)} \cdot \sqrt{\ln 2}, \quad (36)$$

and hence Lemma 7 gives the estimate

$$\log \frac{|\Lambda(T')|}{|\Lambda(T)|} = O(n^{\frac{3}{2}} \sqrt{\log n})$$

Now, write

$$\begin{aligned}
-\log P_{\mathcal{G}^*}(\Lambda(T)) &= H(\mathcal{G}^*) - \log|\Lambda(T)| && (37)\\
&\le H(\mathcal{G}^*) - \log|\Lambda(T')| + O(n^{\frac{3}{2}}\sqrt{\log n}) && (38)\\
&\le H(\mathcal{G}^*) - (H(\mathcal{F}) + D(\mathcal{F}\|\mathcal{G}^*) && (39)\\
&\quad - 2n\log(n+1)) + O(n^{\frac{3}{2}}\sqrt{\log n}) &&\\
&\le H(\mathcal{G}^*) - H(\mathcal{F}) + O(n^{\frac{3}{2}}\sqrt{\log n}) && (40)\\
&\le \max\{H(\mathcal{G}) - H(\mathcal{F})\} + O(n^{\frac{3}{2}}\sqrt{\log n}) && (41)
\end{aligned}$$

where the maximization is over all $\mathcal{G}, \mathcal{F}$ such that $\|\mathcal{G} - \mathcal{F}\|_1 = O(n^{\frac{3}{2}}\sqrt{\log n})$. Let $\mathcal{G} \sim \{g_{ij}\}$ and $\mathcal{F} \sim \{f_{ij}\}$ be the maximizers. Assume without loss of generality that all $g_{ij} < \frac{1}{2}$, hence since the binary entropy function $h(\cdot)$ is increasing in $[0, 1/2]$ and symmetric around $1/2$, $f_{ij} \le g_{ij}$ must hold. Now consider $\mathcal{G}' \sim \{g'_{ij}\}$ where $g'_{ij} = g_{ij} - f_{ij}$, and the associated $\mathcal{F}'$ with $f'_{ij} = 0$. By concavity of $h(\cdot)$, it holds that $H(\mathcal{G}) - H(\mathcal{F}) \le H(\mathcal{G}') - H(\mathcal{F}') = H(\mathcal{G}')$, yet $\|\mathcal{G} - \mathcal{F}\|_1 = \|\mathcal{G}' - \mathcal{F}'\|_1$. Therefore, we can equivalently maximize $H(\mathcal{G})$ subject to $\|\mathcal{G}\|_1 = O(n^{\frac{3}{2}}\sqrt{\log n})$. In this case it is easy to see, again due to concavity of $h(\cdot)$, that the entropy is maximized for an i.i.d. distribution, i.e., where $g_{ij} = O(\log n/\sqrt{n})$. This results in $H(\mathcal{G}) = O(n^{\frac{3}{2}} \log^{\frac{3}{2}} n)$. The proof now follows by substituting the above as an upper bound into (41). □

Before proceeding to the undirected case, we mention two immediate implications of Theorem 1.

Corollary 2. *The unique interchange-invariant graph $\mathcal{G}$ satisfying $\mathbb{E}T_{\mathcal{G}} = T$ is the maximum entropy digraph with this expected type, that is, the solution to the problem*

$$\max_{\mathcal{G}} \; H(\mathcal{G}) \quad s.t \quad \mathbb{E}T_{\mathcal{G}} = T. \quad (42)$$

Proof. This proof follows directly from [5, Lemma 1.3] which shows that the solution to the maximum entropy problem (42) is given by (53). Therefore, this solution is the unique interchange-invariant graph with the given expected type. □

Corollary 3. *Let $T = (\boldsymbol{r}, \boldsymbol{r})$ be an edge-type for which the in-deg and out-deg vectors are identical. Then, the solution $\mathcal{G}^* \sim \{p_{ij}\}$ to (53) is symmetric, i.e. $p_{ij} = p_{ji}$.*

Proof. Since (53) is strictly convex, and in the case where $\boldsymbol{r} = \boldsymbol{c}$, it is also symmetric in $\boldsymbol{u}$ and $\boldsymbol{v}$, the unique solution must maintain $u_i^* = v_i^*, 1 \le i \le n$. It is then immediately follows from (52) that $p_{ij} = p_{ji}$ □

Next, we apply our approach to undirected graphs with no self-loops, to prove Theorem 2. First, we note that the edge-types are invariant to undirected edge interchange, and that one can traverse the edge-type via such interchanges, as in the directed case. Next, we characterize a collection of distributions that are invariant to undirected interchanges, and are therefore uniform inside each edge-types.

Let $\mathcal{G} \sim \{p_{ij}\}_{i>j}$ be a random undirected graph where

$$p_{ij} = \begin{cases} \frac{x_i x_j}{1+x_i x_j} & i \ne j \\ 0 & i = j \end{cases}, \quad (43)$$

for some $\{x_i > 0\}$. It is easy to check that $\mathcal{G}$ is invariant to undirected interchanges, and it can be shown that it satisfies a concentration property, i.e., it gives relatively high probability to the expected type $\mathbb{E}S_{\mathcal{G}} = \boldsymbol{e}$, hence we can conclude that $H(\mathcal{G})$ is equal to $\log|\Lambda(\mathbb{E}S_{\mathcal{G}})|$ up to negligible factors, as in the directed case, Moreover, it can be easily shown that

$$H(\mathcal{G}) = \frac{1}{2} H(\mathcal{G}'), \quad (44)$$

where $\mathcal{G}' \sim \{q_{ij}\}$ is the (symmetric) digraph with $q_{ij} = q_{ji} = p_{ij}$. The claim of Theorem 2 then follows by noting that the expected type of G' is $T = (\boldsymbol{e}, \boldsymbol{e})$.

VI. Proof of Continuity Lemma 7

Our approach is to show that if two edge-types T and T' are close in L^1, then for any graph $G \in \Lambda(T)$ there exists another graph $G' \in \Lambda(T')$ such that the Hamming distance (between the adjacency matrices) $d_H(G, G')$ is small. To show this, we adapt a proof by Krause [7] of the classical Gale-Ryser Theorem [2], [3], to upper bound the minimal number of edge additions / removals that are required in order to move from one edge-type to another that is majorized by it. In the general case where no majorization relation between the edge-types exists, we construct another edge-type that is majorized by both, and whose distance to both remains small, following the approach in [8]. Once these results are proved, our bound follows from a simple counting argument.

Lemma 8. *Let $T=(\boldsymbol{r},\boldsymbol{c}), T'=(\boldsymbol{r}',\boldsymbol{c}')$ be edge-types where $\boldsymbol{r},\boldsymbol{c},\boldsymbol{r}',\boldsymbol{c}'$ are all arranged in descending order. Assume that*

$$\max\left\{\|\boldsymbol{r}-\boldsymbol{r}'\|_1, \|\boldsymbol{c}-\boldsymbol{c}'\|_1\right\} \le m. \tag{45}$$

Then for each $G\in\Lambda(T)$ there exists $G'\in\Lambda(T')$ such that

$$d_H\left(G,G'\right) \le 9m. \tag{46}$$

Proof. Let us assume for now that the total number of edges in both types is the same, and handle the case where it is not at the end. First we show, using the technique introduced in [7], that if $\boldsymbol{c}=\boldsymbol{c}'$ and $\boldsymbol{r}'\preceq\boldsymbol{r}$, then for each $G\in\Lambda(T)$ there exists $G'\in\Lambda(T')$ such that $d_H\left(G,G'\right)\le m$. Let $G\in\Lambda\left(T\right)$, and let i be the minimal index such that $r_i>r_i'$, and $r_{i+1}<r_i$. Further let j be the minimal index such that $r_j<r_j'$. Note that $\boldsymbol{r}'\preceq\boldsymbol{r}$ guarantees that these indices indeed exist and satisfy $i<j$. Now, construct a new vector $\tilde{\boldsymbol{r}}$ as

$$\tilde{r}_l = \begin{cases} r_\ell & \ell\neq i,j,\\ r_i-1 & \ell=i,\\ r_j+1 & \ell=j. \end{cases} \tag{47}$$

It is striaghforward to check that $\tilde{\boldsymbol{r}}$ is in descending order, satisfies the majorization relation $\boldsymbol{r}'\preceq\tilde{\boldsymbol{r}}$, and also $\|\tilde{\boldsymbol{r}}-\boldsymbol{r}'\|_1=\|\boldsymbol{r}-\boldsymbol{r}'\|_1-2$. Since $r_j<r_j'\le r_i'<r_i$, there is at least one vertex k in G such that $k\to i$ and $k\nrightarrow j$. Changing the position of the edge such that $k\nrightarrow i$ and $k\to j$ yields a new graph $\tilde{G}\in\tilde{T}\triangleq(\tilde{\boldsymbol{r}},\boldsymbol{c})$ with $d_H(G,\tilde{G})=2$. Repeating this procedure at most $m/2$ times, we end up with a graph $G'\in\Lambda(T')$ such that $d_H(G,G')\le m$.

Moving on to the general case, where we borrow an idea from [8] to construct a new edge-type $\tilde{T}=(\tilde{\boldsymbol{r}},\tilde{\boldsymbol{c}})$, where

$$\tilde{r}_i=\min\left(\sum_{j=1}^{i} r_j, \sum_{j=1}^{i} r_j'\right)-\min\left(\sum_{j=1}^{i-1} r_j, \sum_{j=1}^{i-1} r_j'\right), \tag{48}$$

and similarly for $\tilde{\boldsymbol{c}}$. These in-deg and out-deg vectors are majorized by those of T and T' respectively, and $\max\{\|\tilde{\boldsymbol{r}}-\boldsymbol{r}\|_1,\|\tilde{\boldsymbol{r}}-\boldsymbol{r}'\|_1,\|\tilde{\boldsymbol{c}}-\boldsymbol{c}\|_1,\|\tilde{\boldsymbol{c}}-\boldsymbol{c}'\|_1\}\le m$. Let $\tilde{T}_1=(\boldsymbol{r},\tilde{\boldsymbol{c}})$, $\tilde{T}_2=(\boldsymbol{r}',\tilde{\boldsymbol{c}})$, and note that the existence of $\tilde{T},\tilde{T}_1,\tilde{T}_2$ follows directly from the Gale-Ryser Theorem Then, applying the process described in the first part of this proof we can move between the types $T,\tilde{T}_1,\tilde{T},\tilde{T}_2,T'$ in this order, with a Hamming cost of at most m for each transition, hence by the triangle inequality, $d_H\left(G,G'\right)\le 4m$.

To conclude, we need to handle the case where the total number of edges in T and T' is not the same. In this case, it is simple to check that we can always create a graph $\tilde{G}$ with the same number of edges as G, such that $d_H(\tilde{G},G)\le m$, and where $T_{\tilde{G}}=(\tilde{\boldsymbol{r}},\tilde{\boldsymbol{c}})$ are in descending order and $\max\left\{\|\tilde{\boldsymbol{r}}-\boldsymbol{r}'\|_1,\|\tilde{\boldsymbol{c}}-\boldsymbol{c}'\|_1\right\}\le 2m$. We can now apply the previous reasoning to create the desired $G'\in\Lambda(T')$, and from the triangle inequality it follows that $d_H(\tilde{G},G')\le 8m$. □

Lemma 9. *Let $T=(\boldsymbol{r},\boldsymbol{c})$ and $T'=(\boldsymbol{r}',\boldsymbol{c}')$ be edge-types satisfying $\max\{\|\boldsymbol{r}-\boldsymbol{r}'\|_1,\|\boldsymbol{c}-\boldsymbol{c}'\|_1\}\le m$. Then,*

$$\frac{|\Lambda\left(T\right)|}{|\Lambda(T')|}\le\binom{n^2}{18m} \tag{49}$$

Proof. First, note that without loss of generality we can assume that the in-deg and out-deg vectors in T and T' are already in descending order, since the size of a type class is invariant to permutation of the in-deg and out-deg vectors.

Now, let $f:\Lambda\left(T\right)\to\Lambda\left(T'\right)$ be a function that maps graphs in $\Lambda\left(T\right)$ to their closest neighbor in Hamming distance in $\Lambda\left(T'\right)$, with ties broken arbitrary. By construction, and in light of Lemma 8, for any $F\in f(\Lambda(T))\subseteq\Lambda(T')$ in the image, and any two graphs $G,G'\in f^{-1}\left(F\right)$ in the inverse image of F, it must hold that $d\left(G,G'\right)\le 18m$. Hence it follows that

$$\left|f^{-1}\left(F\right)\right|\le\binom{n^2}{18m}. \tag{50}$$

Therefore

$$|\Lambda(T)|=\sum_{F\in f(\Lambda(T))}|f^{-1}(F)|\le\binom{n^2}{18m}\cdot|\Lambda(T')|. \tag{51}$$

□

Lemma 7 now follows from Lemma 9 since $\log\binom{n^2}{18m}\le 36m\log n$.

Appendix A

Proof of Lemma 5. To prove this, we show that any interchange-invariant digraph $\mathcal{G}\sim\{p_{ij}\}$ with expected type $\mathbb{E}T_{\mathcal{G}}=T=(\boldsymbol{r},\boldsymbol{c})$ can be written as

$$p_{ij}=\frac{e^{u_i^*}e^{v_j^*}}{1+e^{u_i^*}e^{v_j^*}} \tag{52}$$

where $u_i^*,v_j^*,\ 1\le i,j\le n$ are solution to the convex problem

$$\inf_{\boldsymbol{u},\boldsymbol{v}}\left[-\sum_i u_i r_i-\sum_j v_j c_j-\sum_{ij}\log\left(1+e^{u_i}e^{v_j}\right)\right]. \tag{53}$$

Then, since this problem is convex and has a unique solution, we deduce that there is a single such digraph.

Consider the problem

$$\begin{aligned}&\inf_{\boldsymbol{x},\boldsymbol{y}}\log\Big(\prod_i x_i^{-r_i}\prod_j y_j^{-c_j}\prod_{ij}\left(1+x_iy_j\right)\Big). \\ &s.t.\ \ 0<x_i,\ 0<y_i,\end{aligned} \tag{54}$$

Solving the KKT conditions for (54) yields

$$\sum_i\frac{x_i^*y_j^*}{1+x_i^*y_j^*}=c_j,\quad\sum_j\frac{x_i^*y_j^*}{1+x_i^*y_j^*}=r_i. \tag{55}$$

Therefore, any interchange-invariant digraph $\mathcal{G}$ satisfying $\mathbb{E}T_{\mathcal{G}}=T$ is a solution to the problem. Substituting $x_i=e^{u_i},y_j=e^{v_j}$ in (54) leads to (53), which is strictly convex. Hence the solution $\mathcal{G}$ is unique. □

References

[1] I. Csiszar and J. Korner, *Information theory: coding theorems for discrete memoryless systems*. Orlando, FL, USA: Academic Press, Inc., 1982.

[2] D. Gale, “A theorem on flows in networks,” *Pacific J. Math*, vol. 7, no. 2, pp. 1073–1082, 1957.

[3] H. J. Ryser, “Combinatorial properties of matrices of zeros and ones,” *Canad. J. Math*, vol. 9, pp. 371–377, 1957.

[4] C. Greenhill, B. D. McKay, and X. Wang, “Asymptotic enumeration of sparse 0-1 matrices with irregular row and column sums,” *Journal of Combinatorial Theory, Series A*, vol. 113, pp. 291–324, Feb. 2006.

[5] A. Barvinok, “On the number of matrices and a random matrix with prescribed row and column sums and 0-1 entries,” *Advances in Mathematics*, vol. 224, pp. 316–339, May 2010.

[6] R. Bustin and O. Shayevitz, “On lossy compression of binary matrices,” in *Information Theory (ISIT), 2017 IEEE International Symposium on*, pp. 1573–1577, IEEE, 2017.

[7] M. Krause, “A simple proof of the Gale-Ryser theorem,” *The American Mathematical Monthly*, vol. 103, p. 335, Apr. 1996.

[8] F. Cicalese and U. Vaccaro, “Supermodularity and subadditivity of the entropy on the majorization lattice,” *IEEE Transactions on Information Theory*, vol. 48, pp. 933–938, 2002.