Type Graphs and Small-Set Expansion

Lei Yu
Statistics and Data Science, Nankai University
Tianjin, 300071, China
Email: leiyu@nankai.edu.cn

Venkat Anantharam
EECS, University of California
Berkeley, CA 94720, USA
Email: ananth@berkeley.edu

Jun Chen
ECE, McMaster University
Hamilton, ON L8S 4K1, Canada
Email: chenjun@mcmaster.ca

2021 IEEE International Symposium on Information Theory (ISIT) | 978-1-5386-8209-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/ISIT45174.2021.9517992

***Abstract*—In this paper, we study the *type graph*, namely a bipartite graph induced by a joint type. We study the maximum edge density of induced bipartite subgraphs of this graph having a number of vertices on each side on an exponential scale. This can be seen as an isoperimetric problem. We provide asymptotically sharp bounds for the exponent of the maximum edge density as the blocklength goes to infinity. We also study the biclique rate region of the type graph, which is defined as the set of (R_1, R_2) such that there exists a biclique of the type graph which has respectively e^{nR_1} and e^{nR_2} vertices on the two sides. We provide asymptotically sharp bounds for the biclique rate region as well. We also apply similar techniques to strengthen small-set expansion theorems.**

I. Introduction

Let $\mathcal{X}$ and $\mathcal{Y}$ be two finite sets. Let T_X be an n-type on $\mathcal{X}$, i.e., an empirical distribution of sequences from $\mathcal{X}^n$. (Obviously, any n-type T_{XY} is also a kn-type for $k \geq 1$.) Let $\mathcal{T}_{T_X}^{(n)}$, or simply $\mathcal{T}_{T_X}$, be the type class contained in $\mathcal{X}^n$ with respect to T_X, i.e., the set of sequences in $\mathcal{X}^n$ having the type T_X. Similarly, let T_{XY} be a joint n-type on $\mathcal{X} \times \mathcal{Y}$ and $\mathcal{T}_{T_{XY}}$ the joint type class with respect to T_{XY}. Obviously, $\mathcal{T}_{T_{XY}} \subseteq \mathcal{T}_{T_X} \times \mathcal{T}_{T_Y}$, where T_X, T_Y are the marginal types corresponding to the joint type T_{XY}. In this paper, we consider the undirected bipartite graph $G_{T_{XY}}$ whose vertex set is $\mathcal{T}_{T_X} \cup \mathcal{T}_{T_Y}$ and whose edge set can be identified with $\mathcal{T}_{T_{XY}}$, defined as follows. Consider $\mathbf{x} \in \mathcal{T}_{T_X}$ and $\mathbf{y} \in \mathcal{T}_{T_Y}$ as vertices of $G_{T_{XY}}$. Two vertices $\mathbf{x}, \mathbf{y}$ are joined by an edge if and only if $(\mathbf{x}, \mathbf{y}) \in \mathcal{T}_{T_{XY}}$. $G_{T_{XY}}^{(n)}$ is termed the *graph of* T_{XY} or, more succinctly, a *type graph* [1]. For brevity, when there is no ambiguity, we use the abbreviated notation G for $G_{T_{XY}}^{(n)}$. For subsets $A \subseteq \mathcal{T}_{T_X}, B \subseteq \mathcal{T}_{T_Y}$, we obtain an induced bipartite subgraph $G[A,B]$ of G, whose vertex sets are A and B, and where $\mathbf{x}, \mathbf{y}$ are joined by an edge if and only if they are joined by an edge in G. For the induced subgraph $G[A,B]$, the (edge) density $\rho(G[A,B])$ is defined as

$$\rho(G[A,B]) := \frac{\#\text{ of edges in } G[A,B]}{|A|\,|B|}.$$

Obviously, $\rho(G[A,B]) = \frac{\left|(A\times B)\cap\mathcal{T}_{T_{XY}}\right|}{|A||B|}$. It is interesting to observe that[1] $\rho(G) = \left|\mathcal{T}_{T_{XY}^{(n)}}\right| / \left(\left|\mathcal{T}_{T_X^{(n)}}\right|\left|\mathcal{T}_{T_Y^{(n)}}\right|\right) \doteq e^{-nI_{T^{(n)}}(X;Y)}$ for any sequence of joint types $\left\{T_{XY}^{(n)}\right\}$. Moreover, if we only fix T_X, T_Y, A, and B, then $T_{XY} \in C_n(T_X, T_Y) \mapsto \rho\left(G_{T_{XY}}^{(n)}[A,B]\right)$ forms a probability mass function, i.e., $\rho\left(G_{T_{XY}}^{(n)}[A,B]\right) \geq 0$ and $\sum_{T_{XY}\in C_n(T_X,T_Y)} \rho\left(G_{T_{XY}}^{(n)}[A,B]\right) = 1$, where $C_n(T_X, T_Y)$ denotes the set of joint n-types T_{XY} with marginals T_X, T_Y. We term this distribution a *type distribution*, which, roughly speaking, can be considered as a generalization from binary alphabets to arbitrary finite alphabets of the classic *distance distribution* in coding theory; please refer to [2] for the distance distribution of a single code, and [3] for the distance distribution between two codes.

Given $1 \leq M_1 \leq |\mathcal{T}_{T_X}|, 1 \leq M_2 \leq |\mathcal{T}_{T_Y}|$, define the *maximal density* of subgraphs with size (M_1, M_2) as

$$\Gamma_n(M_1, M_2) := \max_{A\subseteq\mathcal{T}_{T_X}, B\subseteq\mathcal{T}_{T_Y}: |A|=M_1, |B|=M_2} \rho(G[A,B]).$$

Recall that $T_{X|Y}$ and $T_{Y|X}$ denote the conditional types corresponding to the joint type T_{XY}. For a sequence $\mathbf{x} \in \mathcal{T}_{T_X}$, let $\mathcal{T}_{T_{Y|X}}(\mathbf{x})$ denote the corresponding conditional type class. Since $N_1 := \left|\mathcal{T}_{T_{Y|X}}(\mathbf{x})\right|$ is independent of $\mathbf{x} \in \mathcal{T}_{T_X}$, the degrees of the vertices $\mathbf{x} \in \mathcal{T}_{T_X}$ are all equal to the constant N_1. Similarly, the degrees of the vertices $\mathbf{y} \in \mathcal{T}_{T_Y}$ are all equal to the constant $N_2 := \left|\mathcal{T}_{T_{X|Y}}(\mathbf{y})\right|$. Hence we have

$$|B|\rho(G[A,B]) + |B^c|\rho(G[A,B^c]) = N_1,$$

where $B^c := \mathcal{T}_{T_Y} \backslash B$. Thus, over A, B with fixed sizes, maximizing $\rho(G[A,B])$ is equivalent to minimizing $\rho(G[A,B^c])$. In other words, determining the maximal density is in fact an edge-isoperimetric problem. Furthermore, given $A \subseteq \mathcal{T}_{T_X}$ and M_2, $\max_{B\subseteq\mathcal{T}_{T_Y}:|B|=M_2} \rho(G[A,B])$ is attained by B^* such that $\left|A \cap \mathcal{T}_{T_{X|Y}}(\mathbf{y})\right| \geq \left|A \cap \mathcal{T}_{T_{X|Y}}(\mathbf{y}')\right|$ for any $\mathbf{y} \in B^*, \mathbf{y}' \notin B^*$. Hence, $M_2 \mapsto \max_{B\subseteq\mathcal{T}_{T_Y}:|B|=M_2} \rho(G[A,B])$ is nonincreasing, which implies that $\Gamma_n(M_1, M_2)$ is nonincreasing in one parameter given the other parameter.

Let[2] $\mathcal{R}_X^{(n)} := \left\{\frac{1}{n}\log M_1 : M_1 \in [|\mathcal{T}_{T_X}|]\right\}$ and $\mathcal{R}_Y^{(n)} := \left\{\frac{1}{n}\log M_2 : M_2 \in [|\mathcal{T}_{T_Y}|]\right\}$. Given a joint n-type T_{XY}, define the *exponent of maximal density* for a pair $(R_1, R_2) \in \mathcal{R}_X^{(n)} \times \mathcal{R}_Y^{(n)}$ as

$$E_n(R_1, R_2) := -\frac{1}{n}\log\Gamma_n\left(e^{nR_1}, e^{nR_2}\right). \tag{1}$$

[1]Throughout this paper, we write $a_n \doteq b_n$ to denote $a_n = b_n e^{no(1)}$.

[2]We use the notation $[m:n] := \{m, m+1, ..., n\}$ and $[n] := [1:n]$.

If the edge density of a subgraph in a bipartite graph G is equal to 1, then this subgraph is called a biclique of G. Along these lines, we define the *biclique rate region* of T_{XY} as

$$\mathcal{R}_n(T_{XY}) := \big\{(R_1,R_2)\in\mathcal{R}_X^{(n)}\times\mathcal{R}_Y^{(n)}: \Gamma_n\left(e^{nR_1},e^{nR_2}\right)=1\big\}.$$

Observe that any n-type T_{XY} can also be viewed as a kn-type for $k\geq 1$. For an n-type T_{XY}, define the *asymptotic exponent of maximal density* for a pair $(R_1,R_2)\in\mathcal{R}_X^{(n)}\times\mathcal{R}_Y^{(n)}$ as[3]

$$E(R_1,R_2) := \lim_{k\to\infty} -\frac{1}{kn}\log\Gamma_{kn}\left(e^{knR_1},e^{knR_2}\right), \quad (2)$$

and the *asymptotic biclique rate region* as

$$\mathcal{R}(T_{XY}) := \text{closure}\bigcup_{k\geq 1}\mathcal{R}_{kn}(T_{XY}). \quad (3)$$

Han and Kobayashi [4] introduced a concept similar to the asymptotic biclique rate region defined here. However, roughly speaking, their definition is an approximate version of our definition, in the sense that in their definition, for a distribution P_{XY} (not necessarily a type), type classes are replaced with the typical sets with respect to P_{XY}, and the constraint $\Gamma_n\left(e^{nR_1},e^{nR_2}\right)=1$ is replaced with $\Gamma_n\left(e^{nR_1^{(n)}},e^{nR_2^{(n)}}\right)\to 1$ as $n\to\infty$ for a sequence of types $T_{XY}^{(n)}$ converging to P_{XY} and a sequence of pairs $\left(R_1^{(n)},R_2^{(n)}\right)$ converging to (R_1,R_2).

In this paper we are interested in characterizing the limits $E(R_1,R_2), \mathcal{R}(T_{XY})$, and bounding the convergence rates of $E_n(R_1,R_2)$ and $\mathcal{R}_n(T_{XY})$ to these limits as $n\to\infty$.

A. Motivations

Our motivations for studying the type graph have the following three aspects.

1) The method of types is a classic and powerful tool in information theory. In this method, the basic unit is the (joint) type or (joint) type class. To the authors' knowledge, it is not well understood how the sequence pairs are distributed in a joint type class. The maximal density (or the biclique rate region) measures how concentrated are the joint-type sequence pairs by counting the number of joint-type sequence pairs in each "local" rectangular subset. Hence, our study of the type graph deepens the understanding of the distribution (or structure) of sequence pairs in a joint type class. The first study on this topic can be traced back to Han and Kobayashi's work [4], and it was also investigated in [5], [1], [6] recently. However, all these works considered approximate versions of bicliques. In contrast, we consider the exact version.
2) Observe that the type graph can be constructed by permuting two sequences $\mathbf{x},\mathbf{y}$ respectively. Thus, unlike other well-studied large graphs, the type graph is deterministic rather than stochastic. There are relatively few works focusing on deterministic large graphs. Hence, as a purely combinatorial problem, studying the type graph is of independent interest.
3) The maximal and minimal density problems for type graphs are closely related to noninteractive simulation problems and hypercontractivity inequalities. Our results and proof ideas can be applied to prove bounds in noninteractive simulation problems and to strengthen hypercontractivity inequalities.

[3] By definition, it is easy to see that $-\frac{1}{kn}\log\Gamma_{kn}\left(e^{knR_1},e^{knR_2}\right)$ is nonincreasing in k. Hence the limit in (2) exists. Moreover, this limit, namely $E(R_1,R_2)$, is only dependent on T_{XY} and is independent of the value of n we attribute to T_{XY}. Similar conclusions can be drawn for the asymptotic biclique rate region defined in (3).

B. Main Contributions

We first completely characterize the asymptotics of the exponent of maximal density and the biclique rate region for any joint type defined on finite alphabets. We observe that, in general, the asymptotic biclique rate region defined by us is a subset (in general, a strict subset) of the approximate one defined by Han and Kobayashi [4]. In fact, their definition for a distribution P_{XY} is equal to the asymptotic rate region of a sequence of n-types $\{T_{XY}^{(n)}\}$ approaching P_{XY}, which satisfy the condition $E_n\left(R_1^{(n)},R_2^{(n)}\right)\to 0$ as $n\to\infty$. Interestingly, our proof for the biclique rate region combines information-theoretic methods and linear algebra, which seems not common in information theory. We also apply similar proof techniques to strengthen the forward and reverse small-set expansion theorems in [7], [8], [9].

C. Notation

For a sequence $\mathbf{x}$, we use $T_{\mathbf{x}}$ to denote the type of $\mathbf{x}$. For an n-length vector or sequence $\mathbf{x}$ and a subset $\mathcal{J}\subseteq[n]$, define $\mathbf{x}_{\mathcal{J}} := (x_j)_{j\in\mathcal{J}}$, i.e., the vector consisting of the components with indices in $\mathbf{x}$. We will also use notations $H_Q(X)$ or $H(Q_X)$ to denote the entropy of $X\sim Q_X$. If the distribution is denoted by P, we sometimes write the entropy $H_P(X)$ as $H(X)$ for brevity. We use $\mathrm{supp}(P_X)$ to denote the support of P_X.

II. Type Graph

In this section, we completely characterize the asymptotic exponent of maximal density and the asymptotic biclique rate region.

Theorem 1. *For any $n\geq 2(|\mathcal{X}||\mathcal{Y}|+2)|\mathcal{X}||\mathcal{Y}|$, any n-type T_{XY}, and $(R_1,R_2)\in\mathcal{R}_X^{(n)}\times\mathcal{R}_Y^{(n)}$, we have*

$$E^*(R_1,R_2)\leq E_n(R_1,R_2)\leq E^*(R_1,R_2)+\varepsilon_n, \quad (4)$$

where $\varepsilon_n := \frac{(|\mathcal{X}||\mathcal{Y}|+2)|\mathcal{X}||\mathcal{Y}|}{n}\log\frac{(n+1)n^6}{|\mathcal{X}|^4|\mathcal{Y}|^4}$, $E^*(R_1,R_2) := R_1+R_2-F(R_1,R_2)$, *and*

$$F(R_1,R_2) := \max_{\substack{P_{XYW}:P_{XY}=T_{XY},\\ H(X|W)\leq R_1, H(Y|W)\leq R_2}} H(XY|W), \quad (5)$$

which we think of as being defined for all nonnegative pairs (R_1,R_2). In particular, $E(R_1,R_2)=E^*(R_1,R_2)$. Without loss of optimality, the alphabet size of W in the definition of F can be assumed to be no larger than $|\mathcal{X}||\mathcal{Y}|+2$.

Remark 1. Obviously, E^* can be also expressed as $E^*(R_1, R_2) = R_1 + R_2 - H_T(XY) + G(R_1, R_2)$, with

$$G(R_1, R_2) := \min_{\substack{P_{XYW}: P_{XY} = T_{XY}, \\ H(X|W) \le R_1, H(Y|W) \le R_2}} I(XY; W) \tag{6}$$

corresponding to the minimum common rate given marginal rates (R_1, R_2) in the Gray-Wyner source coding network [10, Theorem 14.3]. Moreover, for $R_1 = H_T(X), R_2 = H_T(Y)$, we have that (5) or (6) is attained by a constant W, which in turn implies that $E^*(H_T(X), H_T(Y)) = I_T(X; Y)$.

Proof of Theorem 1: The alphabet bound $|\mathcal{W}| \le |\mathcal{X}| |\mathcal{Y}| + 2$ in the definition of F comes from the support lemma in [10]. We next prove the bounds in (4) by using standard information-theoretic techniques.

Upper Bound: For a joint type P_{XYW} such that $P_{XY} = T_{XY}, H(X|W) \le R_1, H(Y|W) \le R_2$ and for a fixed sequence $\mathbf{w}$ with type P_W, we choose A as the union of $\mathcal{T}_{P_{X|W}}(\mathbf{w})$ and a number $e^{nR_1} - |\mathcal{T}_{P_{X|W}}(\mathbf{w})|$ of arbitrary sequences outside $\mathcal{T}_{P_{X|W}}(\mathbf{w})$, and choose B in a similar way, but with $\mathcal{T}_{P_{X|W}}(\mathbf{w})$ replaced by $\mathcal{T}_{P_{Y|W}}(\mathbf{w})$. Then $|A| = e^{nR_1}$ and $|B| = e^{nR_2}$. Observe that

$$\begin{aligned} |(A \times B) \cap \mathcal{T}_{T_{XY}}| &\ge \left|\mathcal{T}_{P_{XY|W}}(\mathbf{w})\right| \\ &\ge e^{n\left(H(XY|W) - \frac{|\mathcal{W}||\mathcal{X}||\mathcal{Y}|\log(n+1)}{n}\right)}, \end{aligned}$$

where the second inequality follows from [11, Lemma 2.5]. Thus we have

$$\begin{aligned} \rho(G[A, B]) &\ge \frac{|(A \times B) \cap \mathcal{T}_{T_{XY}}|}{e^{nR_1} e^{nR_2}} \\ &\ge e^{-n\left(R_1 + R_2 - H(XY|W) + \frac{|\mathcal{W}||\mathcal{X}||\mathcal{Y}|\log(n+1)}{n}\right)}. \end{aligned} \tag{7}$$

Optimizing the exponent in (7) over all joint n-types P_{XYW} such that $P_{XY} = T_{XY}, H(X|W) \le R_1, H(Y|W) \le R_2$ yields the upper bound

$$\begin{aligned} E_n(R_1, R_2) \le R_1 + R_2 - F_n(R_1, R_2) \\ + \frac{|\mathcal{W}| |\mathcal{X}| |\mathcal{Y}| \log(n+1)}{n}, \end{aligned} \tag{8}$$

where F_n is defined similarly as F in (5) but with the P_{XYW} in (5) restricted to be a joint type. It is not difficult to remove this restriction at the cost of adding an asymptotically vanishing term, by combining the inequality in [11, Lemma 2.7] and the fact in [12, Lemma 3] that the types are dense in the probability simplex (i.e., any distribution P_{XYW} can be approximated by an n-type within an asymptotically vanishing TV distance). We give the proof detail in [13], and omit it here.

Lower Bound: Let $C := (A \times B) \cap \mathcal{T}_{T_{XY}}$ for some optimal (A, B) attaining $\Gamma_n\left(e^{nR_1}, e^{nR_2}\right)$. Let $(\mathbf{X}, \mathbf{Y}) \sim \mathrm{Unif}(C)$. Then,

$$\Gamma_n\left(e^{nR_1}, e^{nR_2}\right) = \frac{|C|}{|A| |B|} = \frac{e^{H(\mathbf{X}, \mathbf{Y})}}{e^{nR_1} e^{nR_2}},$$
$$\frac{1}{n} H(\mathbf{X}) \le R_1, \ \frac{1}{n} H(\mathbf{Y}) \le R_2.$$

Therefore,

$$\begin{aligned} E_n(R_1, R_2) &= R_1 + R_2 - \frac{1}{n} H(\mathbf{X}, \mathbf{Y}) \\ &= R_1 + R_2 - \frac{1}{n} \sum_{i=1}^n H\left(X_i Y_i | X^{i-1} Y^{i-1}\right) \\ &= R_1 + R_2 - H\left(X_J Y_J | X^{J-1} Y^{J-1} J\right) \end{aligned}$$

where $J \sim \mathrm{Unif}[n]$ is a random time index independent of (X^n, Y^n). On the other hand,

$$H\left(X_J | X^{J-1} Y^{J-1} J\right) \le H\left(X_J | X^{J-1} J\right) = \frac{1}{n} H(\mathbf{X}) \le R_1,$$
$$H\left(Y_J | X^{J-1} Y^{J-1} J\right) \le R_2.$$

Using the notation $X := X_J, Y := Y_J, W := X^{J-1} Y^{J-1} J$, we obtain $(X, Y) \sim T_{XY}$, and

$$\begin{aligned} E_n(R_1, R_2) &\ge \inf_{\substack{P_{XYW}: P_{XY} = T_{XY}, \\ H(X|W) \le R_1, H(Y|W) \le R_2}} R_1 + R_2 - H(XY|W) \\ &= E^*(R_1, R_2). \end{aligned}$$

∎

We next consider the biclique rate region.

Theorem 2. *For any $n \ge 8(|\mathcal{X}| |\mathcal{Y}|)^{7/5}$ and any n-type T_{XY},*

$$\begin{aligned} &(\mathcal{R}^*(T_{XY}) - [0, \varepsilon_{1,n}] \times [0, \varepsilon_{2,n}]) \cap \left(\mathcal{R}_X^{(n)} \times \mathcal{R}_Y^{(n)}\right) \\ &\subseteq \mathcal{R}_n(T_{XY}) \qquad (9) \\ &\subseteq \mathcal{R}^*(T_{XY}) \cap \left(\mathcal{R}_X^{(n)} \times \mathcal{R}_Y^{(n)}\right) \qquad (10) \end{aligned}$$

where "−" is the Minkowski difference (i.e., for $A, B \subseteq \mathbb{R}^m$, $A - B := \bigcap_{b \in B}(A - b)$), $\varepsilon_{1,n} := \frac{|\mathcal{X}||\mathcal{Y}|}{n} \log \frac{n^4(n+1)}{16|\mathcal{X}|}$, $\varepsilon_{2,n} := \frac{|\mathcal{X}||\mathcal{Y}|}{n} \log \frac{n^4(n+1)}{16|\mathcal{Y}|^2}$, and

$$\begin{aligned} \mathcal{R}^*(T_{XY}) := \bigcup_{\substack{0 \le \alpha \le 1, P_{XY}, Q_{XY}: \\ \alpha P_{XY} + (1-\alpha) Q_{XY} = T_{XY}}} &\{(R_1, R_2) : \\ &R_1 \le \alpha H_P(X|Y), R_2 \le (1-\alpha) H_Q(Y|X)\}. \end{aligned}$$

In particular,

$$\mathcal{R}(T_{XY}) = \mathcal{R}^*(T_{XY}), \tag{11}$$

where $\mathcal{R}(T_{XY})$ is the asymptotic biclique rate region, defined in (3).

Remark 2. Given T_{XY}, $\mathcal{R}^*(T_{XY})$ is a closed convex set. Using the continuity of $H_P(X|Y)$ in P_{XY}, it is straightforward to establish that $\mathcal{R}^*(T_{XY})$ is closed. Convexity follows by the fact that $H_P(X|Y)$ is concave in P_{XY}; see the details in [13].

Remark 3. Theorem 2 can be easily generalized to the k-variables case with $k \ge 3$; see [13].

Proof: Obviously, (11) follows from (9) and (10). It suffices to prove (9) and (10).

Inner Bound: The inner bound proof here uses a standard time-sharing argument. Let d be an integer such that $1 \le d \le n-1$. Let (P_{XY}, Q_{XY}) be a pair of d-joint type and $(n-d)$-joint type on $\mathcal{X} \times \mathcal{Y}$ such that $\frac{d}{n} P_{XY} + \left(1 - \frac{d}{n}\right) Q_{XY} =$

T_{XY}. For a fixed d-length sequence $\mathbf{y}$ with type P_Y and a fixed $(n-d)$-length sequence $\mathbf{x}$ with type Q_X, we choose $A = \mathcal{T}_{P_{X|Y}}(\mathbf{y}) \times \{\mathbf{x}\}$ and $B = \{\mathbf{y}\} \times \mathcal{T}_{Q_{Y|X}}(\mathbf{x})$. Then, from [11, Lemma 2.5], we have $|A| \geq e^{d\left(H_P(X|Y) - \frac{|\mathcal{X}||\mathcal{Y}|\log(d+1)}{d}\right)}$ and similarly $|B| \geq e^{(n-d)\left(H_Q(Y|X) - \frac{|\mathcal{X}||\mathcal{Y}|\log(n-d+1)}{n-d}\right)}$. On the other hand, for this code, $A \times B \subseteq \mathcal{T}_{T_{XY}}$. Hence any rate pair $(R_1, R_2) \in \left(\mathcal{R}_X^{(n)} \times \mathcal{R}_Y^{(n)}\right)$ with

$$R_1 \leq \frac{d}{n}\left(H_P(X|Y) - \frac{|\mathcal{X}||\mathcal{Y}|\log(d+1)}{d}\right),$$
$$R_2 \leq \left(1 - \frac{d}{n}\right)\left(H_Q(Y|X) - \frac{|\mathcal{X}||\mathcal{Y}|\log(n-d+1)}{n-d}\right)$$

is achievable, which in turn implies that a pair of smaller rates $(R_1, R_2) \in \left(\mathcal{R}_X^{(n)} \times \mathcal{R}_Y^{(n)}\right)$ with

$$R_1 \leq \frac{d}{n} H_P(X|Y) - \frac{|\mathcal{X}||\mathcal{Y}|\log(n+1)}{n}, \tag{12}$$
$$R_2 \leq \left(1 - \frac{d}{n}\right) H_Q(Y|X) - \frac{|\mathcal{X}||\mathcal{Y}|\log(n+1)}{n} \tag{13}$$

is achievable.

We next remove the constraint that (P_{XY}, Q_{XY}) are joint types. Due to the space limit, we refer readers to the extended paper [13] for the proof of this.

Outer Bound: We next prove the outer bound by combining information-theoretic methods and linear algebra. Observe that the biclique rate region only depends on the probability values of T_{XY}, rather than the alphabets $\mathcal{X}, \mathcal{Y}$. With this in mind, we observe that we can identify $\mathcal{X}$ and $\mathcal{Y}$ with subsets of $\mathbb{R}$ by one-to-one mappings such that, for any probability distribution P_{XY}, if $(X, Y) \in \mathcal{X} \times \mathcal{Y}$ satisfies $(X, Y) \sim P_{XY}$ we can talk about the expectations $\mathbb{E}_P[X]$, $\mathbb{E}_P[Y]$, the covariance $\mathrm{Cov}_P(X, Y)$, and the correlation $\mathbb{E}_P[XY]$. Translating the choices of $\mathcal{X}$ and/or $\mathcal{Y}$ (as subsets of $\mathbb{R}$) does not change $\mathrm{Cov}_P(X, Y)$, so we can ensure that we make these choices in such a way that $\mathbb{E}_P[XY] = \mathrm{Cov}_P(X, Y) + \mathbb{E}_P[X]\mathbb{E}_P[Y] = 0$.

Let us now choose $\mathcal{X}, \mathcal{Y} \subseteq \mathbb{R}$ in this way, such that for the given joint n-type T_{XY} we have $\mathbb{E}_T[XY] = 0$. Then, for $A \times B \subseteq \mathcal{T}_{T_{XY}}$, we will have $\langle \mathbf{x}, \mathbf{y} \rangle = 0$ for any $(\mathbf{x}, \mathbf{y}) \in A \times B$, where $\mathbf{x}, \mathbf{y}$ are now viewed as vectors in $\mathbb{R}^n$. Let $\overline{A}$ denote the linear space spanned by all the vectors in A, and $\overline{B}$ denote the linear space spanned by all the vectors in B. Hence $\overline{B} \subseteq \overline{A}^{\perp}$, where $\overline{A}^{\perp}$ denotes the orthogonal complement of a subspace $\overline{A}$. As an important property of the orthogonal complement, $\dim(\overline{A}) + \dim(\overline{A}^{\perp}) = n$. Hence $\dim(\overline{A}) + \dim(\overline{B}) \leq n$. ■

We next establish the following exchange lemma. The proof is provided in [13], which is based on the well-known exchange lemma in linear algebra.

Lemma 1. *Let V_1, V_2 be mutually orthogonal linear subspaces of $\mathbb{R}^n$ with dimensions, denoted as n_1, n_2, satisfying $n_1 + n_2 = n$. Then there always exists a partition $\{\mathcal{J}_1, \mathcal{J}_2\}$ of $[n]$ such that $|\mathcal{J}_i| = n_i$ and $\mathbf{x} = f_i(\mathbf{x}_{\mathcal{J}_i}), \forall \mathbf{x} \in V_i, i = 1, 2$ for some deterministic linear functions $f_i : \mathbb{R}^{n_i} \to \mathbb{R}^n$, where $\mathbf{x}_{\mathcal{J}_i} := (x_j)_{j \in \mathcal{J}_i}$.*

Remark 4. The condition "mutually orthogonal linear subspaces of $\mathbb{R}^n$" can be replaced by "mutually (linearly) independent linear subspaces of $\mathbb{R}^n$", or more generally, "affine subspaces that are translates of mutually independent linear subspaces of $\mathbb{R}^n$".

Remark 5. In other words, under the assumption in this lemma there always exists a permutation σ of $[n]$ such that $\mathbf{x}^{(\sigma)} = f_1\left(\mathbf{x}^{(\sigma)}_{[1:n_1]}\right), \forall \mathbf{x} \in V_1$ and $\mathbf{x}^{(\sigma)} = f_2\left(\mathbf{x}^{(\sigma)}_{[n_1+1:n]}\right), \forall \mathbf{x} \in V_2$ for some deterministic functions $f_i : \mathbb{R}^{n_i} \to \mathbb{R}^n$, where $\mathbf{x}^{(\sigma)}$ is obtained by permuting the components of $\mathbf{x}$ using σ.

Proof: Let d denote $\dim(\overline{A})$, so we have $\dim(\overline{A}^{\perp}) = n - d$. Let $\mathbf{X} \sim \mathrm{Unif}(A), \mathbf{Y} \sim \mathrm{Unif}(B)$ be two independent random vectors, i.e., $(\mathbf{X}, \mathbf{Y}) \sim P_{\mathbf{X},\mathbf{Y}} := \mathrm{Unif}(A)\,\mathrm{Unif}(B)$. Now we set $V_1 = \overline{A}, V_2 = \overline{A}^{\perp}$ in Lemma 1. Then there exists a partition $\{\mathcal{J}, \mathcal{J}^c\}$ of $[n]$ such that $|\mathcal{J}| = d$ and both $\mathbf{x} = f_1(\mathbf{x}_{\mathcal{J}}), \forall \mathbf{x} \in \overline{A}$, and $\mathbf{y} = f_2(\mathbf{y}_{\mathcal{J}^c}), \forall \mathbf{y} \in \overline{A}^{\perp}$ for some deterministic functions $f_1 : \mathbb{R}^d \to \mathbb{R}^n, f_2 : \mathbb{R}^{n-d} \to \mathbb{R}^n$. We further have $\mathbf{X} = f_1(\mathbf{X}_{\mathcal{J}}), \mathbf{Y} = f_2(\mathbf{Y}_{\mathcal{J}^c})$ since $\mathbf{X}, \mathbf{Y}$ are respectively defined on $\overline{A}, \overline{A}^{\perp}$. By this property, on the one hand we have

$$\begin{aligned}
R_1 &= \frac{1}{n} H(\mathbf{X}) = \frac{1}{n} H(\mathbf{X}|\mathbf{Y}) = \frac{1}{n} H(\mathbf{X}_{\mathcal{J}}|\mathbf{Y}) \\
&\leq \frac{1}{n} H(\mathbf{X}_{\mathcal{J}}|\mathbf{Y}_{\mathcal{J}}) \leq \frac{1}{n} \sum_{j \in \mathcal{J}} H(X_j|Y_j) \\
&= \frac{d}{n} H(X_J|Y_J J) \leq \frac{d}{n} H(X_J|Y_J) = \frac{d}{n} H\left(\widetilde{X}|\widetilde{Y}\right)
\end{aligned}$$

where $J \sim \mathrm{Unif}(\mathcal{J})$, $\widetilde{X} := X_J, \widetilde{Y} := Y_J$, with J being independent of $(\mathbf{X}, \mathbf{Y})$. Similarly,

$$\begin{aligned}
R_2 &= \frac{1}{n} H(\mathbf{Y}) = \frac{1}{n} H(\mathbf{Y}|\mathbf{X}) = \frac{1}{n} H(\mathbf{Y}_{\mathcal{J}^c}|\mathbf{X}) \\
&\leq \frac{1}{n} H(\mathbf{Y}_{\mathcal{J}^c}|\mathbf{X}_{\mathcal{J}^c}) \leq \frac{1}{n} \sum_{j \in \mathcal{J}^c} H(Y_j|X_j) \\
&= \left(1 - \frac{d}{n}\right) H\left(Y_{\widehat{J}}|X_{\widehat{J}}\widehat{J}\right) \leq \left(1 - \frac{d}{n}\right) H\left(Y_{\widehat{J}}|X_{\widehat{J}}\right) \\
&= \left(1 - \frac{d}{n}\right) H\left(\widehat{Y}|\widehat{X}\right),
\end{aligned}$$

where $\widehat{J} \sim \mathrm{Unif}(\mathcal{J}^c)$, $\widehat{X} := X_{\widehat{J}}, \widehat{Y} := Y_{\widehat{J}}$, with $\widehat{J}$ being independent of $(\mathbf{X}, \mathbf{Y}, J)$. On the other hand,

$$\begin{aligned}
\frac{d}{n} P_{\widetilde{X}\widetilde{Y}} + \left(1 - \frac{d}{n}\right) P_{\widehat{X}\widehat{Y}} &= \frac{1}{n} \sum_{j \in \mathcal{J}} P_{X_j Y_j} + \frac{1}{n} \sum_{j \in \mathcal{J}^c} P_{X_j Y_j} \\
&= \frac{1}{n} \sum_{j=1}^{n} P_{X_j Y_j} = \mathbb{E}_{(\mathbf{X},\mathbf{Y})}[T_{\mathbf{X}\mathbf{Y}}] = T_{XY},
\end{aligned}$$

where $T_{\mathbf{X}\mathbf{Y}}$ denotes the joint type of a random pairs $(\mathbf{X}, \mathbf{Y})$ which is hence also random. This completes the proof for the outer bound. ■

We next study when the asymptotic biclique rate region is a triangle region. We obtain the following necessary and sufficient condition. The proof is provided in [13].

Proposition 1. *Let T_{XY} be a joint n-type such that $H_T(X|Y), H_T(Y|X) > 0$. Then, the asymptotic biclique rate region $\mathcal{R}(T_{XY})$ is a triangle region, i.e.,*

$$\mathcal{R}(T_{XY}) = \mathcal{R}_{\triangle}(T_{XY}) := \bigcup_{0\le\alpha\le1} \big\{(R_1, R_2) : R_1 \le \alpha H_T(X|Y), R_2 \le (1-\alpha) H_T(Y|X)\big\},$$

if and only if T_{XY} satisfies that $T_{X|Y}(x|y)^{1/H_T(X|Y)} = T_{Y|X}(y|x)^{1/H_T(Y|X)}$ for all x, y.

The condition in Proposition 1 is satisfied by the joint n-types T_{XY} which have marginals $T_X = \mathrm{Unif}(\mathcal{X}), T_Y = \mathrm{Unif}(\mathcal{Y})$ and satisfy at least one of the following two conditions: 1) $|\mathcal{X}| = |\mathcal{Y}|$; 2) X, Y are independent under the distribution T_{XY}. Hence, Proposition 1 implies that the asymptotic biclique rate region is a triangle region if the joint n-type T_{XY} is a doubly symmetric binary source (DSBS).

The Han and Kobayashi [4] approximate version of the asymptotic biclique rate region equals

$$\mathcal{R}^{**}(T_{XY}) := \bigcup_{P_{XYW}:P_{XY}=T_{XY}, X\leftrightarrow W\leftrightarrow Y} \big\{(R_1, R_2) : R_1 \le H(X|W), R_2 \le H(Y|W)\big\}.$$

By Theorem 1, this is $\{(R_1, R_2) : E^*(R_1, R_2) = 0\}$ and hence $\mathcal{R}^*(T_{XY}) \subseteq \mathcal{R}^{**}(T_{XY})$. In fact, it is a strict subset if the joint n-type T_{XY} is a DSBS or $\mathrm{Unif}(\mathcal{X}\times\mathcal{Y})$. This difference is caused by the "type overflow" effect, crystallized by the first author and Tan in [14]. Let (R_1, R_2) be a pair such that $E^*(R_1, R_2) = 0$. Let (A, B) be an optimal pair of subsets attaining $E^*(R_1, R_2)$. All the sequences in A have type T_X, and all the sequences in B have type T_Y. However, in general, the joint types of $(\mathbf{x}, \mathbf{y}) \in A \times B$ might "overflow" from the target joint type T_{XY}. The number of non-overflowed sequence pairs (i.e., $|(A \times B) \cap \mathcal{T}_{T_{XY}}|$) has exponent $R_1 + R_2$, since $E^*(R_1, R_2) = 0$. This means that not too many sequence pairs have overflowed. However, if type overflow is forbidden, then we must reduce the rates of A and B to satisfy this requirement.

III. Strong Small-Set Expansion Theorem

In this section, we study the noninteractive simulation problem with $(\mathbf{X}, \mathbf{Y}) \sim P_{XY}^n$, where P_{XY} is a joint distribution defined on $\mathcal{X}\times\mathcal{Y}$. We still assume that $\mathcal{X}, \mathcal{Y}$ are the supports of P_X, P_Y, and moreover, are finite. Let $E_{1,\max} := -\log(\min_x P_X(x))$, $E_{2,\max} := -\log(\min_y P_Y(y))$. For $E_1 \in [0, E_{1,\max}], E_2 \in [0, E_{2,\max}]$, define

$$\underline{\Theta}_n(E_1, E_2) := -\frac{1}{n}\log \max_{\substack{A\subseteq\mathcal{X}^n, B\subseteq\mathcal{Y}^n:\\ P_X^n(A)\le e^{-nE_1},\\ P_Y^n(B)\le e^{-nE_2}}} P_{XY}^n(A\times B). \quad (14)$$

Define $\overline{\Theta}_n(E_1, E_2)$ similarly by replacing the maximization with minimization. Denote their limits as $n \to \infty$ as $\underline{\Theta}(E_1, E_2), \overline{\Theta}(E_1, E_2)$. Ordentlich, Polyanskiy, and Shayevitz [15] studied $\underline{\Theta}(E_1, E_2), \overline{\Theta}(E_1, E_2)$ for binary symmetric distributions $P_{X,Y}$ In this section, we consider an arbitrary distribution $P_{X,Y}$ and prove the following two theorems, which improve the forward and reverse small-set expansion theorems in [7], [8], [9]. The proofs are in [13], and are based on coupling techniques and information-theoretic techniques similar to those used in the proof of Theorem 1.

Theorem 3 (Forward and Reverse Strong Small-Set Expansion Theorems)**.** *For $E_1 \in [0, E_{1,\max}], E_2 \in [0, E_{2,\max}]$,*

$$\underline{\Theta}(E_1, E_2) = \underline{\Theta}^*(E_1, E_2) := \min_{\substack{Q_{XYW}:D(Q_{X|W}\|P_X|Q_W)\ge E_1,\\ D(Q_{Y|W}\|P_Y|Q_W)\ge E_2}} D\left(Q_{XY|W}\|P_{XY}|Q_W\right), \quad (15)$$

$$\overline{\Theta}(E_1, E_2) = \begin{cases} \overline{\Theta}^*(E_1, E_2) & \text{if } E_1, E_2 > 0;\\ E_1 & \text{if } E_2 = 0;\\ E_2 & \text{if } E_1 = 0, \end{cases} \quad (16)$$

where for $E_1 \in [0, E_{1,\max}], E_2 \in [0, E_{2,\max}]$,

$$\overline{\Theta}^*(E_1, E_2) := \max_{\substack{Q_W, Q_{X|W}, Q_{Y|W}:\\ D(Q_{X|W}\|P_X|Q_W)\le E_1,\\ D(Q_{Y|W}\|P_Y|Q_W)\le E_2}} \min_{Q_{XY|W}\in C(Q_{X|W}, Q_{Y|W})} D\left(Q_{XY|W}\|P_{XY}|Q_W\right). \quad (17)$$

Moreover, $\underline{\Theta}_n(E_1, E_2) \ge \underline{\Theta}^*(E_1, E_2)$ and $\overline{\Theta}_n(E_1, E_2) \le$ the RHS of (16) for any $n \ge 1$ and $E_1 \in [0, E_{1,\max}], E_2 \in [0, E_{2,\max}]$. Without loss of optimality, the alphabet size of W in both (15) *and* (17) *can be assumed to be no larger than* 3*.*

IV. Concluding Remarks

The minimal density and the independent-set rate region are also of interest. Given $1 \le M_1 \le |\mathcal{T}_{T_X}|, 1 \le M_2 \le |\mathcal{T}_{T_Y}|$, the *minimal density* of subgraphs with size (M_1, M_2) is

$$\underline{\Gamma}_n(M_1, M_2) := \min_{A\subseteq\mathcal{T}_{T_X}, B\subseteq\mathcal{T}_{T_Y}:|A|=M_1, |B|=M_2} \rho(G[A, B]).$$

Given n and T_{XY}, define the *independent-set rate region* as

$$\underline{\mathcal{R}}_n(T_{XY}) := \big\{(R_1, R_2) \in \mathcal{R}_X^{(n)} \times \mathcal{R}_Y^{(n)} : \underline{\Gamma}_n\left(e^{nR_1}, e^{nR_2}\right) = 0\big\}.$$

It still remains open to determine the asymptotics of the independent-set rate region.

In addition, it is worth nothing that in the extended version of this conference paper [13], our results and proof ideas here are also applied to strengthen some classical hypercontractivity inequalities [16], [17], [18].

Acknowledgements

Research of the first two authors was supported by NSF grants CNS-1527846, CCF-1618145, CCF-1901004, CIF-2007965, the NSF Science & Technology Center grant CCF-0939370 (Science of Information), and the William and Flora Hewlett Foundation supported Center for Long Term Cybersecurity at Berkeley.

References

[1] A. Nazari, S. S. Pradhan, and A. Anastasopoulos. New bounds on the maximal error exponent for multiple-access channels. In *2009 IEEE International Symposium on Information Theory*, pages 1704–1708. IEEE, 2009.

[2] F. J. MacWilliams and N. J. A. Sloane. *The Theory of Error-Correcting Codes*, volume 16. Elsevier, 1977.

[3] L. Yu and V. Y. F. Tan. On non-interactive simulation of binary random variables. *IEEE Trans. Inf. Theory*, 2021.

[4] T. S. Han and K. Kobayashi. Maximal rectangular subsets contained in the set of partially jointly typical sequences for dependent random variables. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 70(1):15–32, 1985.

[5] D. Krithivasan and S. S. Pradhan. On large deviation analysis of sampling from typical sets. In *Proc. Workshop on Information Theory and Applications (ITA)*. Citeseer, 2007.

[6] A. Nazari, D. Krithivasan, S. S. Pradhan, A. Anastasopoulos, and R. Venkataramanan. Typicality graphs and their properties. In *2010 IEEE International Symposium on Information Theory*, pages 520–524. IEEE, 2010.

[7] J. Kahn, G. Kalai, and N. Linial. The influence of variables on Boolean functions. In *29th Annual Symposium on Foundations of Computer Science*, pages 68–80. IEEE, 1988.

[8] E. Mossel, R. O'Donnell, O. Regev, J. E. Steif, and B. Sudakov. Non-interactive correlation distillation, inhomogeneous Markov chains, and the reverse Bonami-Beckner inequality. *Israel Journal of Mathematics*, 154(1):299–336, 2006.

[9] R. O'Donnell. *Analysis of Boolean Functions*. Cambridge University Press, 2014.

[10] A. El Gamal and Y.-H. Kim. *Network Information Theory*. Cambridge University Press, 2011.

[11] I. Csiszar and J. Körner. *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Cambridge University Press, 2011.

[12] L. Yu and V. Y. F. Tan. Rényi resolvability and its applications to the wiretap channel. *IEEE Trans. Inf. Theory*, 65(3):1862–1897, 2018.

[13] L. Yu, V. Anantharam, and J. Chen. Graphs of joint types, noninteractive simulation, and stronger hypercontractivity. *arXiv preprint arXiv:2102.00668*, Feb. 2021. [Online]. Available: https://arxiv.org/abs/2102.00668.

[14] L. Yu and V. Y. F. Tan. On exact and ∞-Rényi common informations. *IEEE Trans. Inf. Theory*, 66(6):3366–3406, 2020.

[15] O. Ordentlich, Y. Polyanskiy, and O. Shayevitz. A note on the probability of rectangles for correlated binary strings. *IEEE Trans. Inf. Theory*, 2020.

[16] E. Nelson. The free markoff field. *Journal of Functional Analysis*, 12(2):211–227, 1973.

[17] R. Ahlswede and P. Gács. Spreading of sets in product spaces and hypercontraction of the markov operator. *The Annals of Probability*, pages 925–939, 1976.

[18] C. Borell. Positivity improving operators and hypercontractivity. *Mathematische Zeitschrift*, 180(3):225–234, 1982.