Typicality Graphs: Large Deviation Analysis

Ali Nazari, Dinesh Krithivasan, S. Sandeep Pradhan, Achilleas Anastasopoulos
Dept. of Electrical Engineering and computer science
University of Michigan, Ann Arbor
E-mail: {anazari,dineshk,pradhanv,anastas}@umich.edu
Ramji Venkataramanan
Dept. of Electrical Engineering
Stanford University
E-mail: vramji@stanford.edu

Abstract

Let $\mathcal{X}$ and $\mathcal{Y}$ be finite alphabets and P_{XY} a joint distribution over them, with P_X and P_Y representing the marginals. For any $\varepsilon > 0$, the set of n-length sequences x^n and y^n that are jointly typical [1] according to P_{XY} can be represented on a bipartite graph. We present a formal definition of such a graph, known as a *typicality* graph, and study some of its properties.

I. INTRODUCTION

The concept of typicality and typical sequences is central to information theory. It has been used to develop computable performance limits for several communication problems.

Consider a pair of correlated discrete memoryless information sources X[1] and Y characterized by a generic joint distribution p_{XY} defined on the product of two finite sets $\mathcal{X} \times \mathcal{Y}$. An length n X-sequence x^n is typical if the empirical histogram of x^n is close to p_X. A pair of length n sequences $(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n$ is said to be jointly typical if the empirical joint histogram of (x^n, y^n) is close to the joint distribution p_{XY}. The set of all jointly typical sequence pairs is called the typical set of p_{XY}.

Given a sequence length n, the typical set can be represented in terms of the following undirected, bipartite graph. The left vertices of the graph are all the typical X-sequences, and the right vertices are all the typical Y-sequences. From well-known properties of typical sets, there are (approximately) $2^{nH(X)}$ left vertices and $2^{nH(Y)}$ right vertices. A left vertex is connected to a right vertex through an edge if the corresponding X and Y-sequences are *jointly* typical. From the properties of joint typicality, we know that the number of edges in this graph is roughly $2^{nH(X,Y)}$. Further, every left vertex (a typical X-sequence) has degree roughly equal to $2^{nH(Y|X)}$, i.e., it is jointly typical with $2^{nH(Y|X)}$ Y-sequences. Similarly, each right vertex has degree roughly equal to $2^{nH(X|Y)}$.

In this paper we formally characterize the typicality graph and look at some subgraph containment problems. In particular, we answer three questions concerning the typicality graph:

[1]We use the following notation throughout this work. Script capitals $\mathcal{U}$, $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$,... denote finite, nonempty sets. To show the cardinality of a set $\mathcal{X}$, we use $|\mathcal{X}|$. We also use the letters P, Q,... for probability distributions on finite sets, and U, X, Y,... for random variables.

- When can we find subgraphs such that the left and right vertices of the subgraph have specified degrees, say R'_X and R'_Y, respectively ?
- What is the maximum size of subgraphs that are complete, i.e., every left vertex is connected to every right vertex? One of the main contributions of this paper is a sharp answer to this question.
- If we create a subgraph by randomly picking a specified number of left and right vertices, what is the probability that this subgraph has far fewer edges than expected?

These questions arise in a variety of multiuser communication problems. Transmitting correlated information over a multiple-access channel (MAC) [2], and communicating over a MAC with feedback [3] are two problems where the first question plays an important role. The techniques used to answer the second question have been used to develop tighter bounds on the error exponents of discrete memoryless multiple-access channels [4], [5], [6]. The third question arises in the context of transmitting correlated information over a broadcast channel [7]. Moreover, the evaluation of performance limits of a multiuser communication problem can be thought of as characterizing certain properties of typicality graphs of random variables associated with the problem.

The paper is organized as follows. Some preliminaries are introduced in section II. In section III, the typicality graphs are formally defined and some properties about the number vertices, edges, and degree conditions are obtained. The main result of the paper which is obtained in section IV.

II. Preliminaries

In this section, we provide a concise review of some of the results available in the literature on the typical sequences, δ-typical sets and their properties [1].

Definition 1: A sequence $x^n \in \mathcal{X}^n$ is X-typical with constant δ if

1) $|\frac{1}{n}N(a|x^n) - P_X(a)| \leq \delta, \quad \forall a \in \mathcal{X}$
2) No $a \in \mathcal{X}$ with $P_X(a) = 0$ occurs in x^n.

The set of such sequences is denoted $T^n_\delta(P_X)$ or $T^n_\delta(X)$, when the distribution being used is unambiguous.

Definition 2: Given a conditional distribution $P_{Y|X}$, a sequence $y^n \in \mathcal{Y}^n$ is conditionally $P_{Y|X}$-typical with $x^n \in \mathcal{X}^n$ with constant δ if

1) $|\frac{1}{n}N(a,b|x^n,y^n) - \frac{1}{n}N(a|x^n)P_{Y|X}(b|a)| \leq \delta, \quad \forall a \in \mathcal{X}, b \in \mathcal{Y}$.
2) $N(a,b|x^n,y^n) = 0$ whenever $P_{Y|X}(b|a) = 0$.

The set of such sequences is denoted $T^n_\delta(P_{Y|X}|x^n)$ or $T^n_\delta(Y|x^n)$, when the distribution being used is unambiguous.

We will repeatedly use the following results, which we state below as facts:

Fact 1 [1, Lemma 2.10]: (a) If $x^n \in T^n_\delta(X)$ and $y^n \in T^n_{\delta'}(Y|x^n)$, then $(x^n,y^n) \in T^n_{\delta+\delta'}(X,Y)$ and $y^n \in T^n_{(\delta+\delta')|\mathcal{X}|}(Y)$. [2]

(b) If $x^n \in T^n_\delta(X)$ and $(x^n,y^n) \in T^n_\varepsilon(X,Y)$, then $y^n \in T^n_{\delta+\varepsilon}(Y|x^n)$.

[2]The typical sets are with respect to distributions $P_X, P_{Y|X}$ and P_{XY}, respectively.

Fact 2 [1, Lemma 2.13] [3]: There exists a sequence $\varepsilon_n \to 0$ depending only on $|\mathcal{X}|$ and $|\mathcal{Y}|$ such that for every joint distribution $P_X \cdot P_{Y|X}$ on $\mathcal{X} \times \mathcal{Y}$,

$$\begin{aligned} \left|\frac{1}{n}\log|T^n(X)| - H(X)\right| &\le \varepsilon_n \\ \left|\frac{1}{n}\log|T^n(Y|x^n)| - H(Y|X)\right| &\le \varepsilon_n, \quad \forall x^n \in T^n(X). \end{aligned} \tag{1}$$

The next fact deals with the continuity of entropy with respect to probability distributions.

Fact 3 [1, Lemma 2.7] If P and Q are two distributions on X such that

$$\sum_{x\in\mathcal{X}} |P(x) - Q(x)| \le \varepsilon \le \frac{1}{2}$$

then

$$|H(P) - H(Q)| \le -\varepsilon \log \frac{\varepsilon}{|\mathcal{X}|}$$

III. Typicality graphs

Consider any joint distribution $P_X \cdot P_{Y|X}$ on $\mathcal{X} \times \mathcal{Y}$.

Definition 3: For any $\varepsilon_{1n}, \varepsilon_{2n}, \lambda_n \to 0$, the sequence of typicality graphs $G_n(\varepsilon_{1n}, \varepsilon_{2n}, \lambda_n)$ is defined as follows. For every n, G_n is a bipartite graph, with its left vertices consisting of all $x^n \in T^n_{\varepsilon_{1n}}(X)$ and the right vertices consisting of all $y^n \in T^n_{\varepsilon_{2n}}(Y)$. A vertex on the left (say $\tilde{x}^n$) is connected to a vertex on the right (say $\tilde{y}^n$) iff $(\tilde{x}^n, \tilde{y}^n) \in T^n_{\lambda_n}(X, Y)$.

Remark. Henceforth, we will assume that the sequences $\varepsilon_{1n}, \varepsilon_{2n}, \lambda n$ satisfy the 'delta convention' [1, Convention 2.11], i.e.,

$$\varepsilon_{1n} \to 0, \quad \sqrt{n} \cdot \varepsilon_{1n} \to \infty \text{ as } n \to \infty$$

with similar conditions for ε_{2n} and λ_n as well. The delta convention ensures that the typical sets have 'large probability'.

We will use the notation $V_X(.), V_Y(.)$ to denote the vertex sets of any bipartite graph. Some properties of the typicality graph:

1) From Fact 2, we know that for any sequence of typicality graphs $\{G_n(\varepsilon_{1n}, \varepsilon_{2n}, \lambda_n)\}$, the cardinality of the vertex sets satisfies

$$\left|\frac{1}{n}\log|V_X(G_n)| - H(X)\right| \le \varepsilon_n, \quad \left|\frac{1}{n}\log|V_Y(G_n)| - H(Y)\right| \le \varepsilon_n \tag{2}$$

for some sequence $\varepsilon_n \to 0$.

2) The degree of each each vertex $i \in V_X(G_n)$ and $j \in V_Y(G_n)$ satisfies

$$\text{degree}(x^n) \le 2^{n(H(Y|X)+\varepsilon_n)}, \quad \forall x^n \in V_X(G_n); \qquad \text{degree}(y^n) \le 2^{n(H(X|Y)+\varepsilon_n)}, \quad \forall y^n \in V_Y(G_n) \tag{3}$$

for some $\varepsilon_n \to 0$.

[3] The constants of the typical sets for each n, when suppressed, are understood to be some δ_n with $\delta_n \to 0$ and $\sqrt{n} \cdot \delta_n \to \infty$ (delta convention).

Proof: If $x^n \in T^n_{\varepsilon_{1n}}(X)$ and $(x^n, y^n) \in T^n_{\lambda_n}(X,Y)$, then from Fact 1(b), $y^n \in T^n_{\varepsilon_{1n}+\lambda_n}(Y|x^n)$. From the second part of Fact 2, we know that there exists a sequence $\varepsilon_n \to 0$ such that

$$\left|T^n_{\varepsilon_{1n}+\lambda_n}(Y|x^n)\right| \leq 2^{n(H(Y|X)+\varepsilon_n)} \tag{4}$$

From this we conclude that degree$(x^n) \leq 2^{n(H(Y|X)+\varepsilon_n)}, \forall x^n \in V_X(G_n)$. An identical argument yields degree$(y^n) \leq 2^{n(H(X|Y)+\varepsilon_n)}, \forall y^n \in V_Y(G_n)$.

Property 2 gives upper bounds on the degree of each vertex in the typicality graph. Since we have not imposed any relationships between the typicality constants $\varepsilon_{1n}, \varepsilon_{2n}$ and λ_n, in general it cannot be said that the degree of *every* X-vertex (resp. Y-vertex) is close to $2^{NH(Y|X)}$ (resp. $2^{NH(X|Y)}$). However, such an assertion holds for *almost* every vertex in G_n . Specifically, we can show that the above degree conditions hold for a subgraph with exponentially the same size as G_n.

Proposition 1: Every sequence of typicality graphs $G_n(\varepsilon_{1n}, \varepsilon_{2n}, \lambda_n)$ has a sequence of subgraphs $A_n(\varepsilon_{1n}, \varepsilon_{2n}, \lambda_n)$ satisfying the following properties for some $\delta_n \to 0$.

1) The vertex set sizes $|V_X(A_n)|$ and $|V_Y(A_n)|$, denoted θ^n_X and θ^n_Y, respectively, satisfy

$$\left|\frac{1}{n}\log\theta^n_X - H(X)\right| \leq \delta_n, \quad \left|\frac{1}{n}\log\theta^n_Y - H(Y)\right| \leq \delta_n \quad \forall n$$

2) The degree of each X-vertex x^n, denoted $\theta^{'n}(x^n)$ satisfies

$$\left|\frac{1}{n}\log\theta^{'n}(x^n) - H(Y|X)\right| \leq \delta_n \quad \forall x^n \in V_X(A_n).$$

3) The degree of each Y-vertex y^n, denoted $\theta^{'n}(y^n)$, satisfies

$$\left|\frac{1}{n}\log\theta^{'n}(y^n) - H(X|Y)\right| \leq \delta_n \quad \forall y^n \in V_Y(A_n).$$

Proof: The vertex sets $V_X(G_n)$ and $V_Y(G_n)$ are the ε_{1n}-typical and ε_{2n}-typical sets of P_X and P_Y, respectively. To define the subgraphs A_n, we would like to choose the sequences with type P_X and P_Y, respectively as the vertex sets of the subgraph, with an edge connecting two sequences if they have joint type P_{XY}. However, the values taken by the joint pmfs P_{XY}, P_X, P_Y may be any real number between 0 and 1, whereas the joint type of two n-sequences is always a rational number(with denominator n). So we choose the subgraph A_n as follows:

- For each n, approximate the values of P_{XY} to rational numbers with denominator n to obtain pmf $\tilde{P}_{XY}$, respectively. Clearly $\tilde{P}_{XY}$ is a valid joint type of length n and the maximum approximation error is bounded by $\frac{1}{n}$. In fact, $\forall(x,y)$, we have for all sufficiently large n:

$$|P_{XY}(x,y) - \tilde{P}_{XY}(x,y)| < \frac{1}{n} << \frac{1}{\sqrt{n}} < \lambda_n, \tag{5}$$

  where the last inequality follows from the delta convention. Using Fact 1, we also have

$$|P_X(x) - \tilde{P}_X(x)| < |\mathcal{Y}| \cdot \frac{1}{n} << \frac{1}{\sqrt{n}} < \varepsilon_{1n} \tag{6}$$

$$|P_Y(y) - \tilde{P}_Y(y)| < |\mathcal{X}| \cdot \frac{1}{n} << \frac{1}{\sqrt{n}} < \varepsilon_{2n} \tag{7}$$

- The left vertex set of A_n is $T^n_0(\tilde{P}_X)$, i.e., the set of x^n sequences with type $\tilde{P}_X$. The right vertex set of A_n is $T^n_0(\tilde{P}_Y)$- the set of y^n sequences with type $\tilde{P}_Y$. A vertex in $V_X(A_n)$, say a^n is connected to a vertex in $V_Y(A_n)$, say b^n iff $(a^n, b^n) \in T_0^n(\tilde{P}_{X,Y})$, i.e., (a^n, b^n) have joint type $\tilde{P}_{XY}$.

From (5),(6) and (7), we have

$$T_0^n(\tilde{P}_X) \subset T_{\varepsilon_{1n}}^n(P_X), \quad T_0^n(\tilde{P}_Y) \subset T_{\varepsilon_{2n}}^n(P_Y) \quad \text{and}$$
$$T_0^n(\tilde{P}_{X,Y}) \subset T_{\lambda_n}^n(P_{X,Y}).$$

Hence A_n is a subgraph of G_n, as required.

From [1, Lemma 2.3], we have

$$\left|\frac{1}{n}\log|T_0^n(\tilde{P}_X)| - H(\tilde{P}_X)\right| \leq \delta_{1n}, \quad \left|\frac{1}{n}\log|T_0^n(\tilde{P}_Y)| - H(\tilde{P}_Y)\right| \leq \delta_{2n} \quad \forall n, \tag{8}$$

where $\delta_{1n} = (n+1)^{-|\mathcal{X}|}$ and $\delta_{2n} = (n+1)^{-|\mathcal{Y}|}$. Fact 3 establishes the continuity of entropy with respect to the probability distribution. Using Fact 3 along with (5),(6) and (7), we obtain

$$\left|\frac{1}{n}\log|T_0^n(\tilde{P}_X)| - H(P_X)\right| \leq \delta_{1n}, \quad \left|\frac{1}{n}\log|T_0^n(\tilde{P}_Y)| - H(P_Y)\right| \leq \delta_{2n} \quad \forall n, \tag{9}$$

where we have reused δ_{1n}, δ_{2n} with some abuse of notation. This proves the first property.

We now note that $x^n \in V_X(A_n) = T_0^n(\tilde{P}_X)$ and $y^n \in T_0^n(\tilde{P}_{Y|X}|x^n)$ implies a)$(x^n, y^n) \in T_0^n(\tilde{P}_{X,Y})$ and b)$y^n \in T_0^n(\tilde{P}_Y) = V_Y(A_n)$ (Fact 1). This implies

$$\text{degree}(x^n) \geq |T_0^n(\tilde{P}_{Y|X}|x^n)|, \forall x^n \in V_X(A_n). \tag{10}$$

From [1, Lemma 2.5], we know that

$$|T_0^n(\tilde{P}_{Y|X})| \geq 2^{n(H(\tilde{P}_{Y|X}) - \delta_{3n})} \tag{11}$$

where $\delta_{3n} = |\mathcal{X}||\mathcal{Y}|\frac{\log(n+1)}{n}$. In the above, $H(\tilde{P}_{Y|X})$ stands for $H(Y|X)$ computed under the joint distribution $\tilde{P}_{XY}$. Combining this with (10), we get a lower bound on the degree of each $x^n \in V_X(A_n)$:

$$\text{degree}(x^n) \geq 2^{n(H(\tilde{P}_{Y|X}) - \delta_{3n})} \tag{12}$$

From (5) and (6), one can deduce that $\forall x, y$

$$|P_{Y|X}(y|x) - \tilde{P}_{Y|X}(y|x)| < \gamma_n$$

for some $\gamma_n \to 0$. Combining this with Fact 3, (12) can be written as

$$\text{degree}(x^n) \geq 2^{n(H(P_{Y|X}) - \delta_{3n})}, \tag{13}$$

where we reuse the symbol δ_{3n}.

Further, (3) gives an upper bound on the degree of each vertex in G_n. Hence we have

$$\left|\frac{1}{n}\log\theta'^n(x^n) - H(Y|X)\right| \leq \max(\delta_{3n}, \varepsilon_n) \quad \forall x^n \in V_X(A_n) \tag{14}$$

Similarly, we can bound the degree of each vertex in $V_Y(A_n)$ as

$$\left|\frac{1}{n}\log\theta'^n(y^n) - H(X|Y)\right| \leq \max(\delta_{4n}, \varepsilon_n) \quad \forall y^n \in V_Y(A_n) \tag{15}$$

Finally, we can set $\delta_n = \max(\delta_{1n}, \delta_{2n}, \delta_{3n}, \delta_{4n}, \varepsilon_n)$ to complete the proof of the proposition. ■

IV. SUB-GRAPHS CONTAINED IN TYPICALITY GRAPHS

In this section, we study the subgraphs contained in a sequence of typicality graphs.

A. Subgraphs of general degree

Definition 4: A sequence of typicality graphs $G_n(\varepsilon_{1n}, \varepsilon_{2n}, \lambda_n)$ is said to contain a sequence of subgraphs Γ_n of rates (R_X, R_Y, R'_X, R'_Y) if for each n, if there exists a sequence $\delta_n \to 0$ such that

1) The vertex sets of the subgraphs have sizes (denoted Δ_X^n and Δ_Y^n) that satisfy

$$\left| \frac{1}{n} \log \Delta_X^n - R_X \right| \le \delta_n, \quad \left| \frac{1}{n} \log \Delta_Y^n - R_Y \right| \le \delta_n, \ \forall n.$$

2) The degree of each vertex x^n in $V_X(\Gamma_n)$, denoted $\Delta'^n(x^n)$ satisfies

$$\left| \frac{1}{n} \log \Delta'^n(x^n) - R'_Y \right| \le \delta_n, \quad \forall x^n \in V_X(\Gamma_n), \ \forall n.$$

3) The degree of each vertex y^n in the $V_Y(\Gamma_n)$, denoted $\Delta'^n(y^n)$ satisfies

$$\left| \frac{1}{n} \log \Delta'^n(y^n) - R'_X \right| \le \delta_n, \quad \forall y^n \in V_Y(\Gamma_n), \ \forall n.$$

The following proposition gives a characterization of the rate-tuple of a sequence of subgraphs in the sequence of typicality graphs of P_{XY}.

Proposition 2: Let $G_n(\varepsilon_{1n}, \varepsilon_{2n}, \lambda_n)$ be a sequence of typicality graphs of P_{XY}. Define

$$\mathcal{R} \triangleq \{(R_X, R_Y, R'_X, R'_Y) : G_n(\varepsilon_{1n}, \varepsilon_{2n}, \lambda_n) \text{ contains subgraphs of rates } (R_X, R_Y, R'_X, R'_Y)\}$$

Then

$$\mathcal{R} \supseteq \{(R_X, R_Y, R'_X, R'_Y) : R_X \le H(X|U), R_Y \le H(Y|U), R'_X \le H(Y|XU), R'_Y \le H(Y|XU) \text{ for some } P_{U|XY}.\} \tag{16}$$

Proof:

Definition of Γ_n. Consider any conditional distribution $P_{U|XY}$. This fixes the joint distribution $P_{XYU} = P_{XY}P_{U|XY}$. We construct Γ_n as follows.

- For each n, approximate the values of P_{UXY} to rational numbers with denominator n to obtain pmf $\tilde{P}_{UXY}$, respectively. Clearly $\tilde{P}_{UXY}$ is a valid joint type of length n and the maximum approximation error is bounded by $\frac{1}{n}$. Marginalizing the joint pmf, we also have $\forall x, y$

$$|P_{XY}(x,y) - \tilde{P}_{XY}(x,y)| < |\mathcal{U}| \cdot \frac{1}{n} << \frac{1}{\sqrt{n}} < \lambda_n, \tag{17}$$

$$|P_X(x) - \tilde{P}_X(x)| < |\mathcal{Y}| \cdot |\mathcal{U}| \cdot \frac{1}{n} << \frac{1}{\sqrt{n}} < \varepsilon_{1n} \tag{18}$$

$$|P_Y(y) - \tilde{P}_Y(y)| < |\mathcal{X}| \cdot |\mathcal{U}| \cdot \frac{1}{n} << \frac{1}{\sqrt{n}} < \varepsilon_{2n}, \tag{19}$$

  where the last inequality in each equation follows from the delta convention. Further $\forall u$

$$|P_U(u) - \tilde{P}_U(u)| < |\mathcal{Y}| \cdot |\mathcal{X}| \cdot \frac{1}{n}. \tag{20}$$

- Pick any length n sequence u^n with type $\tilde{P}_U$, i.e., $u^n \in T_0^n(\tilde{P}_U)$. Consider a bipartite graph Γ_n with X-vertices consisting of all $x^n \in T_0^n(\tilde{P}_{X|U}|u^n)$, Y-vertices consisting of all $y^n \in T_0^n(\tilde{P}_{Y|U}|u^n)$. In other words, having fixed u^n, the X-vertex sets and Y-vertex sets consist of all length n sequences having conditional type $\tilde{P}_{X|U}$ and $\tilde{P}_{Y|U}$, respectively. Vertices $x^n \in V_X(\Gamma_n)$ and $y^n \in V_Y(\Gamma_n)$ are connected in Γ_n iff $(x^n, y^n) \in T_0^n(\tilde{P}_{XY|U}|u^n)$, i.e., if they have the conditional joint type $P_{XY|U}$ given u^n.

Let us verify that Γ_n is a subgraph of G_n. From Fact 1, if $u^n \in T_0^n(\tilde{P}_U)$ and $x^n \in T_0^n(\tilde{P}_{X|U}|u^n)$, then $(x^n, u^n) \in T_0^n(\tilde{P}_{X,U})$. Consequently, $x^n \in T_0^n(\tilde{P}_X)$. Similarly, all $y^n \in T_0^n(\tilde{P}_{Y|U}|u^n)$ belong to $T_0^n(\tilde{P}_Y)$. On the same lines, if $u^n \in T_0^n(\tilde{P}_U)$ and $(x^n, y^n) \in T_0^n(\tilde{P}_{XY|U}|u^n)$, then $(x^n, y^n, u^n) \in T_0^n(\tilde{P}_{X,Y,U})$. This implies $(x^n, y^n) \in T_0^n(\tilde{P}_{X,Y})$. Further, from (17),(18) and (19), we know

$$T_0^n(\tilde{P}_X) \subset T_{\varepsilon_{1n}}^n(P_X) = V_X(G_n), \quad T_0^n(\tilde{P}_Y) \subset T_{\varepsilon_{2n}}^n(P_Y) = V_Y(G_n) \quad \text{and}$$
$$T_0^n(\tilde{P}_{X,Y}) \subset T_{\lambda_n}^n(P_{X,Y}).$$

Hence for all sufficiently large n, Γ_n is a subgraph of the typicality graph G_n.

Properties of $\boldsymbol{\Gamma_n}$. From [1, Lemma 2.3], we have

$$\left|\frac{1}{n}\log|T_0^n(\tilde{P}_{X|U}|u^n)| - H(\tilde{P}_{X|U})\right| \le \delta_{1n}, \quad \left|\frac{1}{n}\log|T_0^n(\tilde{P}_{Y|U}|u^n)| - H(\tilde{P}_{Y|U})\right| \le \delta_{2n} \quad \forall n, \tag{21}$$

where $\delta_{1n} = (n+1)^{-|\mathcal{X}||\mathcal{U}|}$ and $\delta_{2n} = (n+1)^{-|\mathcal{Y}||\mathcal{U}|}$. Using (18), (19) with (20), we know that $\tilde{P}_{X|U}, \tilde{P}_{Y|U}$ are close to $P_{X|U}, P_{Y|U}$, respectively. Using Fact 3, we know that the entropies $H(\tilde{P}_{X|U}), H(\tilde{P}_{Y|U})$ must close to $H(P_{X|U}), H(P_{Y|U})$, respectively. Thus we can write (21) as (reusing δ_{1n}, δ_{2n})

$$\left|\frac{1}{n}\log|T_0^n(\tilde{P}_{X|U}|u^n)| - H(P_{X|U})\right| \le \delta_{1n}, \quad \left|\frac{1}{n}\log|T_0^n(\tilde{P}_{Y|U}|u^n)| - H(P_{Y|U})\right| \le \delta_{2n} \quad \forall n, \tag{22}$$

Thus, the vertex sets of Γ_n have rates $R_X = H(X|U)$ and $R_Y = H(Y|U)$, as required.

Using Fact 1, for any $x^n \in V_X(\Gamma_n)$, every $y^n \in T_0^n(\tilde{P}_{Y|XU}|x^n, u^n)$ will satisfy a) $(x^n, y^n) \in T_0^n(\tilde{P}_{X,Y|U}|u^n)$ and b) $y^n \in T_0^n(\tilde{P}_{Y|U}|u^n)$. Hence

$$\text{degree}(x^n) \ge |T_0^n(\tilde{P}_{Y|XU}|x^n, u^n)| \ge 2^{n(H(\tilde{P}_{Y|XU}) - \delta_{3n})}, \tag{23}$$

where $\delta_{3n} = |\mathcal{X}||\mathcal{Y}||\mathcal{U}|\frac{\log(n+1)}{n}$. We can also upper bound the degree of x^n by noting that $x^n \in T_0^n(\tilde{P}_{X|U}|u^n)$ and $(x^n, y^n) \in T_0^n(\tilde{P}_{X,Y|U}|u^n)$ implies $y^n \in T_0^n(\tilde{P}_{Y|XU}|x^n, u^n)$. From [1, Lemma 2.5],

$$|T_0^n(\tilde{P}_{Y|XU}|x^n, u^n)| \le 2^{nH(\tilde{P}_{Y|XU})}.$$

Combining this with (23), we have

$$\left|\frac{1}{n}\log{\Delta'}^n(x^n) - H(\tilde{P}_{Y|XU})\right| \le \delta_{3n}, \quad \forall x^n \in V_X(\Gamma_n), \; \forall n. \tag{24}$$

In a similar fashion, we can show that

$$\left|\frac{1}{n}\log{\Delta'}^n(y^n) - H(\tilde{P}_{X|YU})\right| \le \delta_{4n}, \quad \forall y^n \in V_Y(\Gamma_n), \; \forall n. \tag{25}$$

Since the distributions $\tilde{P}_{Y|XU}$ and $\tilde{P}_{X|YU}$ are close to $P_{Y|XU}$ and $P_{X|YU}$, respectively, Fact 3 enables us to replace $H(\tilde{P}_{Y|XU}), H(\tilde{P}_{X|YU})$ with $H(P_{Y|XU}), H(P_{X|YU})$, respectively in the two preceding equations.

Taking $\delta_n = \max(\delta_{1n}, \delta_{2n}, \delta_{3n}, \delta_{4n})$, we have shown the existence of a sequence of subgraphs Γ_n with rates $(H(X|U), H(Y|U), H(Y|XU), H(X|YU))$. Since we can simply exclude edges from Γ_n to obtain subgraphs with smaller rates, it is clear that all rate tuples characterized by

$$(R_X, R_Y, R'_X, R'_Y) : R_X \le H(X|U), \; R_Y \le H(Y|U), \; R'_X \le H(Y|XU), \; R'_Y \le H(Y|XU)$$

are achievable for every conditional distribution $P_{U|XY}$. ■

B. Nearly complete subgraphs

A complete bipartite graph is one in which each vertex of the first set is connected with every vertex on the other set. We next consider a specific class of subgraphs, namely nearly complete subgraphs. For this class of subgraphs, we have a converse result that fully characterizes the set of nearly complete subgraphs present in any typicality graph.

Definition 5: A sequence of typicality graphs $G_n(\varepsilon_{1n}, \varepsilon_{2n}, \lambda_n)$ is said to contain a sequence of nearly complete subgraphs $\Gamma_n(\varepsilon_{1n}, \varepsilon_{2n}, \lambda_n)$ of rates (R_X, R_Y) if for each n, if there exists a sequence $\delta_n \to 0$ such that

1) The sizes of the vertex sets of the subgraphs, denoted Δ_X^n and Δ_Y^n, satisfy

$$\left|\frac{1}{n}\log\Delta_X^n - R_X\right| \leq \delta_n, \quad \left|\frac{1}{n}\log\Delta_Y^n - R_Y\right| \leq \delta_n, \ \forall n.$$

2) The degree of each vertex x^n in the X-set, denoted ${\Delta'}^n(x^n)$ satisfies

$$\frac{1}{n}\log{\Delta'}^n(x^n) \geq R_Y - \delta_n, \quad \forall x^n \in V_X(\Gamma_n), \ \forall n.$$

3) The degree of each vertex j in the Y-set, denoted ${\Delta'}_j^n$ satisfies for all n

$$\frac{1}{n}\log{\Delta'}^n(y^n) \geq R_X - \delta_n, \quad \forall y^n \in V_Y(\Gamma_n), \ \forall n.$$

Proposition 3: Let $G_n(\varepsilon_{1n}, \varepsilon_{2n}, \lambda_n)$ be a sequence of typicality graphs for P_{XY}. Define

$$\mathcal{R} \triangleq \{(R_X, R_Y) : G_n(\varepsilon_{1n}, \varepsilon_{2n}, \lambda_n) \text{ contains nearly complete subgraphs of rates } (R_X, R_Y)\}$$

Then

1)

$$\mathcal{R} \supseteq \{(R_X, R_Y) : R_X \leq H(X|U),\ R_Y \leq H(Y|U) \text{ for some } P_{U|XY} \text{ s.t. } X - U - Y\}^4 \tag{26}$$

2) For all sequences of nearly complete subgraphs of G_n such that the sequence δ_n (in Definition 5) converges to 0 faster than $1/\log n$ (more precisely, $\delta_n = o(\frac{1}{\log n})$ or $\lim_{n\to\infty}\delta_n \log n = 0$), the rates of the subgraph (R_X, R_Y) satisfy

$$R_X \leq H(X|U),\ R_Y \leq H(Y|U) \text{ for some } P_{U|XY} \text{ s.t. } X - U - Y$$

Proof: The first part of the proposition follows directly from Proposition 2 by choosing $P_{U|XY}$ such that $X - U - Y$ form a Markov chain. We now prove the converse under the stated assumption that the sequence δ_n satisfies $\lim_{n\to\infty}\delta_n \log n = 0$.

[4] X, U, Y form a Markov chain, in that order.

Suppose that a sequence of typicality graphs $G_n(\varepsilon_{1n}, \varepsilon_{2n}, \lambda_n)$ contains nearly complete subgraphs Γ_n of rates R_X, R_Y. The total number of edges in Γ_n can be lower bounded as

$$\begin{aligned} |\text{Edges}(\Gamma_n)| &\geq \Delta_X^n \cdot \text{ minimum degree of a vertex in } V_X(\Gamma_n) \\ &\geq \Delta_X^n \cdot 2^{n(R_Y - \delta_n)} \\ &\geq \Delta_X^n \cdot 2^{n(R_Y - \delta_n)} \Delta_Y^n \cdot 2^{-n(R_Y + \delta_n)} \\ &= \Delta_X^n \cdot \Delta_Y^n \cdot 2^{-2n\delta_n}. \end{aligned} \tag{27}$$

Each of these edges represent a pair (x^n, y^n) that is jointly λ_n-typical with respect to the distribution P_{XY}. In other words, each of these pairs (x^n, y^n) belongs to a joint type[1] that is 'close' to P_{XY}. Since the number of joint types of a pair of sequences of length n is at most $(n+1)^{|\mathcal{X}||\mathcal{Y}|}$, the number of edges belonging to the dominant joint type, say $\bar{P}_{XY}$ satisfies

$$|\text{Edges}(\Gamma_n) \text{ having joint type } \bar{P}_{XY}| \geq \frac{\Delta_X^n \cdot \Delta_Y^n 2^{-2n\delta_n}}{(n+1)^{|\mathcal{X}||\mathcal{Y}|}}. \tag{28}$$

Define a subgraph $\mathcal{A}_n$ of Γ_n consisting only of the edges having joint type $\bar{P}_{XY}$. A word about the notation used in the sequel: We will use i, j to index the vertices in $V_X(\Gamma_n), V_Y(\Gamma_n)$, respectively. Thus $i \in \{1, \ldots, \Delta_X^n\}$ and $j \in \{1, \ldots, \Delta_Y^n\}$. The actual sequences corresponding to these vertices will be denoted $x^n(i), y^n(j)$ etc. Using this notation,

$$\mathcal{A}_n \triangleq \{(i,j) : i \in V_X(\Gamma_n), j \in V_Y(\Gamma_n) \text{ s.t. } (x^n(i), y^n(j)) \text{ has joint type } \bar{P}_{XY} \tag{29}$$

From (28),

$$|\mathcal{A}_n| \geq \frac{\Delta_X^n \cdot \Delta_Y^n 2^{-2n\delta_n}}{(n+1)^{|\mathcal{X}||\mathcal{Y}|}} \tag{30}$$

We will prove the converse result using a series of lemmas concerning $\mathcal{A}_n$. Some of the lemmas are similar to those required to prove in [4, Theorem 1]. We only sketch the proofs of such lemmas, referring the reader to [4] for details.

Define random variables X'^n, Y'^n with pmf

$$\Pr((X'^n, Y'^n) = (x^n(i), y^n(j)) = \frac{1}{|\mathcal{A}_n|}, \text{ if } (i,j) \in \mathcal{A}_n. \tag{31}$$

Lemma 1: $I(X'^n; Y'^n) \leq 2n\delta_n + |\mathcal{X}||\mathcal{Y}| \log(n+1)$.

Proof: Follow steps similar to the proof of [4, Lemma 1], using (30) to lower bound the size of $\mathcal{A}_n$.

The next lemma is Ahlswede's version of the 'wringing' technique. Roughly speaking, if it is known that the mutual information between two random sequences is small, then the lemma gives an upper bound on the per-letter mutual information terms (conditioned on some values).

Lemma 2: [8] Let A^n, B^n be RV's with values in $\mathcal{A}^n$, $\mathcal{B}^n$ resp. and assume that

$$I(A^n; B^n) \leq \sigma$$

Then, for any $0 < \delta < \sigma$ there exist $t_1, t_2, ..., t_k \in \{1, ..., n\}$ where $0 \leq k < \frac{2\sigma}{\delta}$ such that for some $\bar{a}_{t_1}, \bar{b}_{t_1}, \bar{a}_{t_2}, \bar{b}_{t_2}, ..., \bar{a}_{t_k}, \bar{b}_{t_k}$

$$I(A_t; B_t | A_{t_1} = \bar{a}_{t_1}, B_{t_1} = \bar{b}_{t_1}, ..., A_{t_k} = \bar{a}_{t_k}, B_{t_k} = \bar{b}_{t_k}) \leq \delta \qquad \text{for } t = 1, 2, ..., n \tag{32}$$

and

$$Pr(A_{t_1} = \bar{a}_{t_1}, B_{t_1} = \bar{b}_{t_1}, ..., A_{t_k} = \bar{a}_{t_k}, B_{t_k} = \bar{b}_{t_k}) \geq (\frac{\delta}{|\mathcal{A}||\mathcal{B}|(2\sigma - \delta)})^k. \tag{33}$$

■

In our case, we will apply Lemma 2 to random variables X'^n and Y'^n. Lemma 1 indicates $\sigma = 2n\delta_n + |\mathcal{X}||\mathcal{Y}|\log(n+1)$, and δ shall be specified later. Hence we have that for some

$$k \leq \frac{2\sigma}{\delta} = \frac{2(n\delta_n + |\mathcal{X}||\mathcal{Y}|\log(n+1))}{\delta},$$

there exist $\bar{x}_{t_1}, \bar{y}_{t_1}, \bar{x}_{t_2}, \bar{y}_{t_2}, ..., \bar{x}_{t_k}, \bar{y}_{t_k}$ such that

$$I(X'_t; Y'_t | X'_{t_1} = \bar{x}_{t_1}, Y'_{t_1} = \bar{y}_{t_1}, ..., X'_{t_k} = \bar{x}_{t_k}, Y'_{t_k} = \bar{y}_{t_k}) \leq \delta \qquad \text{for } t = 1, 2, ..., n. \tag{34}$$

We now define a subgraph of $\mathcal{A}_n$ consisting of all edges (X'^n, Y'^n) that have

$$X'_{t_1} = \bar{x}_{t_1}, Y'_{t_1} = \bar{y}_{t_1}, ..., X'_{t_k} = \bar{x}_{t_k}, Y'_{t_k} = \bar{y}_{t_k}$$

The subgraph denoted as $\bar{\mathcal{A}}_n$ is given by: [5]

$$\bar{\mathcal{A}}_n \triangleq \{(i,j) \in \mathcal{A}_n : X'_{t_1}(i) = \bar{x}_{t_1}, Y'_{t_1}(j) = \bar{y}_{t_1}, ..., X'_{t_k}(i) = \bar{x}_{t_k}, Y'_{t_k}(j) = \bar{y}_{t_k}.\} \tag{35}$$

On the same lines as [4, Lemma 3], we have

$$|\bar{\mathcal{A}}_n| \geq (\frac{\delta}{|\mathcal{X}||\mathcal{Y}|(2\sigma - \delta)})^k |\mathcal{A}_n|. \tag{36}$$

Define random variables $\bar{X}^n$, $\bar{Y}^n$ on $\mathcal{X}^n$ resp. $\mathcal{Y}^n$ by

$$Pr((\bar{X}^n, \bar{Y}^n) = (x^n(i), y^n(j)) = \frac{1}{|\bar{\mathcal{A}}_n|} \text{if } (i,j) \in \bar{\mathcal{A}}_n. \tag{37}$$

If we denote $\bar{X}^n = (\bar{X}_1, ..., \bar{X}_n)$, $Y^n = (\bar{Y}_1, ..., \bar{Y}_n)$, the Fano-distribution of the graph $\bar{\mathcal{A}}_n$ induces a distribution $P_{\bar{X}_t, \bar{Y}_t}$ on the random variables $\bar{X}_t, \bar{Y}_t, t = 1, \ldots, n$. One can show that

$$P(\bar{X}_t = x, \bar{Y}_t = y) = P(X'_t = x, \bar{Y}'_t = y | X'_{t_1}(i) = \bar{x}_{t_1}, Y'_{t_1}(j) = \bar{y}_{t_1}, ..., X'_{t_k}(i) = \bar{x}_{t_k}, Y'_{t_k}(j) = \bar{y}_{t_k}), \ \forall t. \tag{38}$$

Using (38) in Lemma 2, we get the bound $I(\bar{X}_t; \bar{Y}_t) < \delta$. Applying Pinsker's inequality for I-divergences [9], we have

$$\sum_{x,y} |Pr(\bar{X}_t = x, \bar{Y}_t = y) - Pr(\bar{X}_t = x) Pr(\bar{Y}_t = y)| \leq 2\delta^{1/2}, \quad 1 \leq t \leq n. \tag{39}$$

Also define

$$\bar{\mathcal{C}}(i) = \{(i,j) : (i,j) \in \bar{\mathcal{A}}_n, 1 \leq j \leq \Delta_Y^n\}. \tag{40a}$$

$$\bar{\mathcal{B}}(j) = \{(i,j) : (i,j) \in \bar{\mathcal{A}}_n, 1 \leq i \leq \Delta_X^n\}. \tag{40b}$$

We are now ready to present the final lemma required to complete the proof of the converse.

[5]The heirarchy of subgraphs is $G_n \supset \Gamma_n \supset \mathcal{A}_n \supset \bar{\mathcal{A}}_n$

Lemma 3:

$$R_X \leq \frac{1}{n}\sum_{t=1}^{n} H(\bar{X}_t|\bar{Y}_t) + \delta_{1n}$$
$$R_Y \leq \frac{1}{n}\sum_{t=1}^{n} H(\bar{Y}_t|\bar{X}_t) + \delta_{2n}$$
$$R_X + R_Y \leq \frac{1}{n}\sum_{t=1}^{n} H(\bar{X}_t\bar{Y}_t) + +\delta_{3n}$$

for some $\delta_{1n}, \delta_{2n}, \delta_{3n} \to 0$ and the distributions of the RV's are determined by the Fano-distribution on the codewords $\{(x^n(i), y^n(j)) : (i,j) \in \bar{\mathcal{A}}_n\}$.

Proof: We use a strong converse result for non-stationary discrete memoryless channels, found in [10]. Consider a DMC with input A_t and output B_t $(t = 1, \ldots, n)$, with average error probability $\lambda (0 \leq \lambda < 1)$. The result states that the size of the message set M is upper-bounded as

$$\log M < \sum_{t=1}^{n} I(A_t; B_t) + \frac{3}{1-\lambda}|\mathcal{A}|n^{1/2}, \tag{41}$$

where the distributions of the RV's are determined by the Fano-distribution on the codewords.

We apply the above result to three noiseless DMCs $(B_t = A_t, \lambda = 0)$ as follows. Fix $\bar{Y}^n = y^n(j)$ for some $j \in \bar{\mathcal{A}}_n$ and let the input be $\bar{X}_t,\ t = 1, \cdots, n$. Then, from (41) we have

$$\log|\bar{\mathcal{B}}(j)| \leq \sum_{t=1}^{n} H(\bar{X}_t|\bar{Y}_t = y_t(j)) + 3|\mathcal{X}|n^{1/2}. \tag{42}$$

Similarly,

$$\log|\bar{\mathcal{C}}(i)| \leq \sum_{t=1}^{n} H(\bar{Y}_t|\bar{X}_t = x_t(i)) + 3|\mathcal{Y}|n^{1/2}, \tag{43}$$

$$\log|\bar{\mathcal{A}}_n| \leq \sum_{t=1}^{n} H(\bar{X}_t\bar{Y}_t) + 3|\mathcal{X}||\mathcal{Y}|n^{1/2}. \tag{44}$$

Noting that $Pr(\bar{Y}_t = y) = |\bar{\mathcal{A}}|^{-1}\sum_{(i,j)\in\bar{\mathcal{A}}_n} 1_{\{y_t(j),y\}}$, we can sum both sides of (42) over all $(i,j) \in \bar{\mathcal{A}}_n$ to obtain

$$|\bar{\mathcal{A}}_n|^{-1}\sum_{(i,j)\in\bar{\mathcal{A}}_n} \log|\bar{\mathcal{B}}(j)| \leq \sum_{t=1}^{n} H(\bar{X}_t|\bar{Y}_t) + 3|\mathcal{X}|n^{1/2}. \tag{45}$$

Define

$$B^* \triangleq \frac{2^{-2n\delta_n}}{n}\frac{\Delta_X^n}{(n+1)^{|\mathcal{X}||\mathcal{Y}|}}(\frac{\delta}{|\mathcal{X}||\mathcal{Y}|(2\sigma-\delta)})^k. \tag{46}$$

Then,

$$
\begin{aligned}
|\bar{\mathcal{A}}_n|^{-1} \sum_{(i,j)\in\bar{\mathcal{A}}_n} \log|\bar{\mathcal{B}}(j)| &= |\bar{\mathcal{A}}_n|^{-1} \sum_{j} |\bar{\mathcal{B}}(j)| \log|\bar{\mathcal{B}}(j)| \\
&\geq |\bar{\mathcal{A}}_n|^{-1} \sum_{j:|\bar{\mathcal{B}}(j)|\geq B^*} |\bar{\mathcal{B}}(j)| \log|\bar{\mathcal{B}}(j)| \\
&\geq |\bar{\mathcal{A}}_n|^{-1} \log(B^*) \sum_{j:|\bar{\mathcal{B}}(j)|\geq B^*} |\bar{\mathcal{B}}(j)| \\
&\geq |\bar{\mathcal{A}}_n|^{-1} \log(B^*)(|\bar{\mathcal{A}}_n| - \Delta_Y^n B^*). \qquad (47)
\end{aligned}
$$

Combining (36), (30) and the definition of B^*, we also have

$$
\Delta_Y^n B^* \leq \frac{1}{n}|\bar{\mathcal{A}}_n|. \qquad (48)
$$

Using this in (47), we have

$$
\begin{aligned}
|\bar{\mathcal{A}}_n|^{-1} \sum_{(i,j)\in\bar{\mathcal{A}}_n} \log|\bar{\mathcal{B}}(j)| &\geq |\bar{\mathcal{A}}_n|^{-1} \log(B^*)(|\bar{\mathcal{A}}_n| - \frac{1}{n}|\bar{\mathcal{A}}_n|) \\
&= (1-\frac{1}{n}) \log(\frac{2^{-2n\delta_n}}{n} \frac{\Delta_X^n}{(n+1)^{|\mathcal{X}||\mathcal{Y}|}} (\frac{\delta}{|\mathcal{X}||\mathcal{Y}|(2\sigma-\delta)})^k). \qquad (49)
\end{aligned}
$$

Using (45) in the above we have

$$
\log\Delta_X^n \leq \frac{n}{n-1}(\sum_{t=1}^{n} H(\bar{X}_t|\bar{Y}_t) + 3|\mathcal{X}|n^{1/2}) + 2n\delta_n + \log n + |\mathcal{X}||\mathcal{Y}|\log(n+1) + k\log(\frac{|\mathcal{X}||\mathcal{Y}|2\sigma}{\delta}) \qquad (50)
$$

Analogously,

$$
\log\Delta_Y^n \leq \frac{n}{n-1}(\sum_{t=1}^{n} H(\bar{Y}_t|\bar{X}_t) + 3|\mathcal{Y}|n^{1/2}) + 2n\delta_n + \log n + |\mathcal{X}||\mathcal{Y}|\log(n+1) + k\log(\frac{|\mathcal{X}||\mathcal{Y}|2\sigma}{\delta}) \qquad (51)
$$

Next, we find an upper bound for $\log\Delta_X^n\Delta_Y^n$. From (36), we get

$$
\begin{aligned}
\log|\bar{\mathcal{A}}_n| &\geq \log|\mathcal{A}_n| + k\log(\frac{\delta}{|\mathcal{X}||\mathcal{Y}|(2\sigma-\delta)}) \\
&\geq \log|\mathcal{A}_n| + k\log(\frac{\delta}{|\mathcal{X}||\mathcal{Y}|2\sigma}) \\
&= \log|\mathcal{A}_n| - k\log(\frac{2\sigma}{\delta}) - k\log(|\mathcal{X}||\mathcal{Y}|) \\
&\overset{(a)}{\geq} \log(\Delta_X^n\Delta_Y^n) - |\mathcal{X}||\mathcal{Y}|\log(n+1) - 2n\delta_n - k\log(\frac{|\mathcal{X}||\mathcal{Y}|2\sigma}{\delta}), \qquad (52)
\end{aligned}
$$

where (a) is obtained by using (30). Using (44), the above inequality becomes

$$
\log(\Delta_X^n\Delta_Y^n) \leq \sum_{t=1}^{n} H(\bar{X}_t\bar{Y}_t) + 3|\mathcal{X}||\mathcal{Y}|n^{1/2} + |\mathcal{X}||\mathcal{Y}|\log(n+1) + 2n\delta_n + k\log(\frac{2\sigma}{\delta}) + k\log(|\mathcal{X}||\mathcal{Y}|) \qquad (53)
$$

Using the lower bounds on the sizes of Δ_X, Δ_Y from 5, we can rewrite (50),(51) and (53) as

$$R_X - \delta_n \leq \frac{1}{n-1}\sum_{t=1}^{n} H(\bar{X}_t|\bar{Y}_t) + 3|\mathcal{X}|\frac{n^{1/2}}{n-1} + 2\delta_n + \frac{\log n + |\mathcal{X}||\mathcal{Y}|\log(n+1)}{n} + \frac{k}{n}\log(\frac{2|\mathcal{X}||\mathcal{Y}|\sigma}{\delta}) \tag{54}$$

$$R_Y - \delta_n \leq \frac{1}{n-1}\sum_{t=1}^{n} H(\bar{Y}_t|\bar{X}_t) + 3|\mathcal{Y}|\frac{n^{1/2}}{n-1} + 2\delta_n + \frac{\log n + |\mathcal{X}||\mathcal{Y}|\log(n+1)}{n} + \frac{k}{n}\log(\frac{2|\mathcal{X}||\mathcal{Y}|\sigma}{\delta}) \tag{55}$$

$$R_X + R_Y - 2\delta_n \leq \frac{1}{n}\sum_{t=1}^{n} H(\bar{X}_t\bar{Y}_t) + 3|\mathcal{X}||\mathcal{Y}|\frac{n^{1/2}}{n-1} + |\mathcal{X}||\mathcal{Y}|\frac{\log(n+1)}{n} + 2\delta_n + \frac{k}{n}\log(\frac{2|\mathcal{X}||\mathcal{Y}|\sigma}{\delta}) \tag{56}$$

For our proof we would like all the terms on the right hand side of the above equations (except the entropies) to converge to 0 as $n \to \infty$. This will happen if

$$\frac{k}{n}\log(\frac{2\sigma}{\delta}) \to 0.$$

Recall from Lemma 1 that $\sigma = 2n\delta_n + |\mathcal{X}||\mathcal{Y}|\log(n+1)$ and $k < \frac{2\sigma}{\delta}$. Hence we need to choose δ such that

$$\frac{2\sigma}{n\delta}\log(\frac{2\sigma}{\delta}) \sim \frac{\delta_n + \frac{\log n}{n}}{\delta}\left(\log(n\delta_n + \log n) - \log\delta\right) \to 0. \tag{57}$$

¿From our assumption in the beginning, we have $\delta_n \log n \to 0$. Set

$$\delta = (\delta_n \log n)^{1/2} \tag{58}$$

We see that asymptotically, (57) becomes

$$\frac{\delta_n^{1/2}}{(\log n)^{1/2}}\left[\log(n\delta_n + \log n) - \log(\delta_n^{1/2}) - \log\log n\right] \tag{59}$$

We separately consider each of the terms in the equation above

1) If $\log(n\delta_n + \log n) \sim \log(n\delta_n)$ for large n, then

$$\begin{aligned}\frac{\delta_n^{1/2}}{(\log n)^{1/2}}\log(n\delta_n + \log n) &\sim \frac{\delta_n^{1/2}}{(\log n)^{1/2}}\log(n\delta_n) = \frac{\delta_n^{1/2}}{(\log n)^{1/2}}\left[\log n + \log\delta_n\right] \\ &= (\delta_n \log n)^{1/2} + \frac{\delta_n^{1/2}\log\delta_n}{(\log n)^{1/2}} \to 0, \text{ since } \delta_n \to 0.\end{aligned} \tag{60}$$

   If $\log(n\delta_n + \log n) \sim \log(\log n)$ for large n, then

$$\frac{\delta_n^{1/2}}{(\log n)^{1/2}}\log(n\delta_n + \log n) \sim \frac{\delta_n^{1/2}}{(\log n)^{1/2}}\log(\log n) \to 0. \tag{61}$$

2) $\frac{\delta_n^{1/2}}{(\log n)^{1/2}}\log(\delta_n^{1/2}) \to 0$ because $x\log x \to 0$ when $x \to 0$.
3) $\frac{\delta_n^{1/2}}{(\log n)^{1/2}}\log\log n = (\delta_n \log n)^{1/2}\frac{\log\log n}{\log n} \to 0.$

Hence the term in (59) converges to 0 as $n \to \infty$, completing the proof of the lemma. ■

We can rewrite Lemma 3 using new variables $\bar{X}, \bar{Y}, Q$, where $Q = t \in \{1, 2, ..., n\}$ with probability $\frac{1}{n}$ and $P_{\bar{X},\bar{Y}|Q=t} = P_{\bar{X}_t,\bar{Y}_t}$. So we now have (for all sufficiently large n),

$$R_X \leq H(\bar{X}|\bar{Y}, Q) + \delta_{1n} \tag{62}$$

$$R_Y \leq H(\bar{Y}|\bar{X}, Q) + \delta_{2n} \tag{63}$$

$$R_X + R_Y \leq H(\bar{X}, \bar{Y}|Q) + \delta_{3n}, \tag{64}$$

for some $\delta_{1n}, \delta_{2n}, \delta_{3n} \to 0$.

Finally, using (39), we also have

$$\begin{aligned} &|Pr(\bar{X} = x, \bar{Y} = y|Q = t) - Pr(\bar{X} = x|Q = t)Pr(\bar{Y} = y|Q = t)| \\ &= |Pr(\bar{X}_t = x, \bar{Y}_t = y) - Pr(\bar{X}_t = x)Pr(\bar{Y}_t = y)| \\ &\leq 2\delta^{1/2} = 2(\delta_n \log n)^{1/4} \to 0 \text{ as } n \to \infty. \end{aligned} \tag{65}$$

In other words, for all t, $\bar{X}_t, \bar{Y}_t$ are almost independent for large n. Consequently, using the continuity of mutual information with respect to the joint distribution, Lemma 3 holds with for any joint distribution $P_Q P_{\bar{X}|Q} P_{\bar{Y}|Q}$ such that the marginal on $(\bar{X}, \bar{Y})$ is $P_{\bar{X},\bar{Y}}$. Recall that $P_{\bar{X},\bar{Y}}$ is the dominant joint type that is λ_n-close to $P_{X,Y}$. Using suitable continuity arguments, we can now argue that Lemma 3 holds with for any joint distribution $P_Q P_{X|Q} P_{Y|Q}$ such that the marginal on (X, Y) is $P_{X,Y}$, completing the proof of the converse.

C. Nearly Empty Subgraphs

So far, we have discussed properties of subgraphs of the typicality graph $G_n(\varepsilon_{1n}, \varepsilon_{2n}, \lambda_n)$ such as the containment of nearly complete subgraphs and subgraphs of general degree. Now, we turn our attention to the presence of nearly empty subgraphs in the typicality graph. Our approach towards this problem differs slightly from the approach we took in Sections IV-A and IV-B. While in these sections we characterized the subgraphs based on the degrees of their vertices, in this section we would characterize nearly empty subgraphs by the total number of edges present in such graphs. To effect this characterization, we take a different approach than the one used in previous sections and analyze the probability that a randomly chosen subgraph of the typicality graph has far fewer edges than expected. In particular, we focus attention on the case when the random subgraph has no edges.

Consider a pair (X, Y) of discrete memoryless stationary correlated sources with finite alphabets $\mathcal{X}$ and $\mathcal{Y}$ respectively. Suppose we sample 2^{nR_1} sequences from the typical set $T^n_{\varepsilon_{1n}}(X)$ of X independently with replacement and similarly sample 2^{nR_2} sequences from the typical set $T^n_{\varepsilon_{2n}}(Y)$ of Y. The underlying typicality graph $G_n(\varepsilon_{1n}, \varepsilon_{2n}, \lambda_n)$ induces a bipartite graph on these $2^{nR_1} + 2^{nR_2}$ sequences. We provide a characterization of the probability that this graph is sparser than expected. This characterization is obtained through the use of a version of Suen's inequalities [11] and the Lovasz local lemma [12] listed below.

Lemma 4: [11] Let $I_i \in \text{Be}(p_i), i \in \mathcal{I}$ be a family of Bernoulli random variables. Their dependency graph L is formed in the following manner. Denote the random variable I_i by a vertex i and join vertices i and j by an edge if the corresponding random variables are dependent. Let $X = \sum_i \mathbb{E}(I_i)$ and $\Gamma = \mathbb{E}(X) = \sum_i p_i$. Moreover, write $i \sim j$ if (i, j) is an edge in the dependency graph L and let $\Theta = \frac{1}{2}\sum_i \sum_{j \sim i} \mathbb{E}(I_i I_j)$ and $\theta = \max_i \sum_{j \sim i} p_j$. Then, Suen's inequalities state that for any $0 \leq a \leq 1$,

$$P(X \leq a\Gamma) \leq \exp\left\{-\min\left((1-a)^2 \frac{\Gamma^2}{8\Theta + 2\Gamma}, (1-a)\frac{\Gamma}{6\theta}\right)\right\} \tag{66}$$

Putting $a = 0$, this can be further tightened to

$$P(X = 0) \leq \exp\left\{-\min\left(\frac{\Gamma^2}{8\Theta}, \frac{\Gamma}{2}, \frac{\Gamma}{6\theta}\right)\right\} \tag{67}$$

Lemma 5: [12] Let L be the dependency graph for events $\varepsilon_1, \ldots, \varepsilon_n$ in a probability space and let $E(L)$

be the edge set of L. Suppose there exists $x_i \in [0,1], 1 \le i \le n$ such that

$$P(\varepsilon_i) \le x_i \prod_{(i,j)\in E(L)} (1 - x_j). \tag{68}$$

Then, we have

$$P(\cap_{i=1}^n \overline{\varepsilon_i}) \ge \prod_{i=1}^n (1 - x_i). \tag{69}$$

Another version of the local lemma is as given below. Let $\phi(x), 0 \le x \le e^{-1}$ be the smallest root of the equation $\phi(x) = e^{x\phi(x)}$. With definitions of Γ and θ as in Lemma 4 and defining $\tau \triangleq \max_i P(\varepsilon_i)$, we have

$$P\left(\cap_{i=1}^n \overline{\varepsilon_i}\right) \ge \exp\left\{-\Gamma\phi(\theta + \tau)\right\} \tag{70}$$

With these preliminaries, we are ready to state the main result of this section.

Proposition 4: Suppose X and Y are correlated finite alphabet memoryless random variables with joint distribution $p(x,y)$. Let $\varepsilon_{1n}, \varepsilon_{2n}, \lambda_n$ satisfy the 'delta convention' and R_1, R_2 be any positive real numbers such that $R_1 + R_2 > I(X;Y)$. Let $\mathcal{C}_X$ be a collection of 2^{nR_1} sequences picked independently and with replacement from $T^n_{\varepsilon_{1n}}(X)$ and let $\mathcal{C}_Y$ be defined similarly. Let U be the cardinality of the set

$$\mathcal{U} \triangleq \{(x^n, y^n) \in \mathcal{C}_X \times \mathcal{C}_Y : (x^n, y^n) \in T^n_{\lambda_n}(X,Y)\} \tag{71}$$

Assume, without loss of generality that $R_1 \ge R_2$. Then, for any $\gamma \ge 0$, we have

$$\lim_{n\to\infty} \frac{1}{n} \log\log \left[\mathbb{P}\left(\frac{\mathbb{E}(U) - U}{\mathbb{E}(U)} \ge e^{-n\gamma}\right)\right]^{-1} \ge \begin{cases} R_1 + R_2 - I(X;Y) - \gamma & \text{if } R_1 < I(X;Y) \\ R_2 - \gamma & \text{if } R_1 \ge I(X;Y) \end{cases} \tag{72}$$

Setting $\gamma = 0$ in the above equation gives us

$$\lim_{n\to\infty} \frac{1}{n} \log\log \frac{1}{\mathbb{P}(U = 0)} \ge \min\left(R_2, R_1 + R_2 - I(X;Y)\right) \tag{73}$$

This inequality holds with equality when $R_2 \le R_1 \le I(X;Y)$.

Proof: Let $X^n(i)$ and $Y^n(j)$ denote the ith and jth codewords in the random codebooks $\mathcal{C}_X$ and $\mathcal{C}_Y$ respectively. For $1 \le i \le 2^{nR_1}$ and $1 \le j \le 2^{nR_2}$, define the indicator random variables

$$U_{ij} \triangleq \begin{cases} 1 & \text{if } (X^n(i), Y^n(j)) \in T^n_{\lambda_n}(X,Y) \\ 0 & \text{else} \end{cases} \tag{74}$$

The cardinality of the set $\mathcal{U}$ is then

$$U = \sum_{i=1}^{2^{nR_1}} \sum_{j=1}^{2^{nR_2}} U_{ij} \tag{75}$$

We derive upper bounds on the probability of the lower tail of U using Suen's inequality. To do this, we first set up the dependency graph of the indicator random variables U_{ij}. The vertex set of the graph is indexed by the ordered pair $(i,j), 1 \le i \le 2^{nR_1}, 1 \le j \le 2^{nR_2}$. From the nature of the random experiment, it is clear that the indicator random variables U_{ij} and $U_{i'j'}$ are independent if and only if $i \ne i'$ and $j \ne j'$. Thus, each vertex (i,j) is connected to exactly $2^{nR_1} + 2^{nR_2} - 2$ vertices of the form $(i,j'), j' \ne j$ or $(i',j), i' \ne i$. If vertices (i,j) and (k,l) are connected, we denote it by $(i,j) \sim (k,l)$.

In order to estimate Γ, Θ and θ as defined in Lemma 4, define the following quantities. Let $\alpha_{ij} \triangleq \mathbb{P}(U_{ij} = 1)$ and $\beta_{\{ij\}\{kl\}} \triangleq \mathbb{E}(U_{ij}U_{kl})$ where $(i,j) \sim (k,l)$. Using Facts 1 and 2, uniform bounds can be derived for these quantities as

$$\alpha \triangleq 2^{-n(I(X;Y)+\varepsilon_{3n})} \leq \alpha_{ij} \leq 2^{-n(I(X;Y)-\varepsilon_{3n})} \triangleq \alpha^{'} \tag{76}$$

where ε_{3n} is a continuous positive function of $\varepsilon_{1n}, \varepsilon_{2n}$ and λ_n that goes to 0 as $n \to \infty$. Similarly, a uniform bound on $\beta_{\{ij\}\{kl\}}$ can be derived as

$$2^{-2n(I(X;Y)+2\varepsilon_{4n})} \leq \beta_{\{ij\}\{kl\}} \leq 2^{-2n(I(X;Y)-2\varepsilon_{4n})} \triangleq \beta \tag{77}$$

where ε_{4n} is a continuous positive function of $\varepsilon_{1n}, \varepsilon_{2n}$ and λ_n that goes to 0 as $n \to \infty$.

The quantities involved in Suen's inequality can now be estimated.

$$\Gamma \triangleq \mathbb{E}(U) \geq 2^{n(R_1+R_2)}\alpha \tag{78}$$

$$\Theta \triangleq \frac{1}{2}\sum_{(i,j)}\sum_{(k,l)\sim(i,j)} \mathbb{E}(U_{ij}U_{kl}) \leq \frac{1}{2}2^{n(R_1+R_2)}(2^{nR_1}+2^{nR_2}-2)\beta \tag{79}$$

$$\theta \triangleq \max_{(i,j)} \sum_{(k,l)\sim(i,j)} \mathbb{E}(U_{kl}) \leq (2^{nR_1}+2^{nR_2}-2)\alpha^{'} \tag{80}$$

Substituting these bounds into equations (67) and (66) proves the claims made in equations (72) and (73) of Proposition 4.

A lower bound to the probability of the induced random subgraph being empty can be derived by employing the Lovasz local lemma on the $2^{n(R_1+R_2)}$ events $\{U_{ij}=1\}, 1 \leq i \leq 2^{nR_1}, 1 \leq j \leq 2^{nR_2}$. Symmetry considerations imply that all x_i can be set identically to x in Lemma 5. Then the local lemma states that if there exists $x \in [0,1]$ such that $\alpha \leq P(U_{ij}=1) \leq x(1-x)^{(2^{nR_1}+2^{nR_2}-2)}$, then $P(U=0) \geq (1-x)^{2^{n(R_1+R_2)}}$. It is easy to verify that for such an x to exist, we need $R_2 \leq R_1 < I(X;Y)$ and if so, $x = 2^{-nR_1}$ satisfies the condition. Therefore, we have

$$P(U=0) \geq \exp\left(-\left(2^{nR_2}+1\right)\right) \qquad R_2 \leq R_1 < I(X;Y) \tag{81}$$

We can derive a similar bound using the second version of the local lemma given in Lemma 5. While Γ and θ are same as estimated earlier, $\tau = \max_{(i,j)} P(U_{ij}=1)$ is upper bounded by $\alpha^{'}$ as defined in equation (76). Hence,

$$P(U=0) \geq \exp\left(-\Gamma\phi(\theta+\tau)\right). \tag{82}$$

Under the same assumption $R_2 \leq R_1 < I(X;Y)$, $\theta+\tau \leq (2^{nR_1}+2^{nR_2}-2)\alpha^{'} \to 0$ as $n \to \infty$ and hence $\phi(\theta+\tau) \to 1$. Combining equations (81) and (82), taking logarithms and letting $n \to \infty$, we get

$$\lim_{n\to\infty}\frac{1}{n}\log\log\frac{1}{P(U=0)} \leq \min\left(R_2, R_1+R_2-I(X;Y)\right). \tag{83}$$

Comparing this to equation (73) shows that this expression is asymptotically tight in the regime $R_2 \leq R_1 < I(X;Y)$. ■

References

[1] I. Csiszar and J. Korner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic Press,, 1981.

[2] S. S. Pradhan, S. Choi, and K. Ramchandran, "A graph-based framework for transmission of correlated sources over multiple-access channels," *IEEE Trans Inf Theory*, vol. 53, no. 12, pp. 4583–4604, 2007.

[3] R. Venkataramanan and S. S. Pradhan, "A new achievable rate region for the discrete memoryless multiple-access channel with feedback," in *Proc. IEEE Int. Symp. on Inf. Theory*, June 2009.

[4] A. Nazari, S. S. Pradhan, and A. Anastasopoulous, "A new sphere-packing bound for maximal error exponent for multiple-access channels," in *Proc. IEEE Int. Symp. Inf. Theory*, July 2008. Online: http://arxiv.org/abs/0803.3645.

[5] A. Nazari, S. S. Pradhan, and A. Anastasopoulous, "New bounds on the maximal error exponent for multiple-access channels," in *Proc. IEEE Int. Symp. Inf. Theory*, July 2009.

[6] A. Nazari, S. S. Pradhan, and A. Anastasopoulous, "A new upper bound for the average error exponent for discrete memoryless multiple-access channels," in *Online: http://arxiv.org*, Dec 2009.

[7] S. Choi and S. S. Pradhan, "A graph-based framework for transmission of correlated sources over broadcast channels," *IEEE Transactions on Information Theory*, vol. 54, no. 7, pp. 2841–2856, 2008.

[8] R. Ahlswede, "An elementary proof of the strong converse theorem for the multiple-access channel," 1982.

[9] A. A. Fedotov, F. Topsoe, and P. Harremoes, "Refinements of pinsker's inequality," *IEEE Trans. Inf. Theory*, vol. 49, pp. 1491–1498, June 2003.

[10] U. Agustin, "Gedachtnisfreie kannale for diskrete zeit," *Z. Wahrscheinlichkelts theory verw*, pp. 10–61, 1966.

[11] S. Janson, "New versions of Suen's correlation inequality," *Random Structures Algorithms*, vol. 13, pp. 467–483, 1998.

[12] N. Alon and J. Spencer, *The Probabilistic Method*. John Wiley and Sons, Inc., 1992.