A Concentration of Measure Approach to Correlated Graph Matching

Farhad Shirani, Siddharth Garg, and Elza Erkip
Electrical and Computer Engineering Department
New York University, NY, USA

Abstract

The graph matching problem emerges naturally in various applications such as privacy, image processing and computational biology. In this paper graph matching is considered under a stochastic model, where a pair of randomly generated graphs with pairwise correlated edges are to be matched with each other. More precisely, given the labeling of the vertices in the first graph, the objective is to recover the labels in the second graph by leveraging the correlation among their edges. The problem is considered under various settings and graph models. In the first step, the Correlated Erdös-Rényi (CER) graph model is considered, where all edge pairs whose vertices have similar labels are generated based on identical distributions and independently of other edges. A matching scheme called the Typicality Matching Scheme is introduced. The scheme operates by investigating the joint typicality of the adjacency matrices of the two graphs. Necessary and sufficient conditions for successful matching are derived based on the parameters of the CER model. In the next step, the results are extended to graph matching in the presence of Community Structure (CS). The CS model is a generalization of the ER model where each vertex in the graph is associated with a community label, which affects its edge statistics. The results are further extended to matching of ensembles of more than two correlated graphs. Lastly, the problem of seeded graph matching is investigated where a subset of the labels in the second graph are known prior to matching. In this scenario, a polytime matching algorithm is proposed. It is shown that successful matching is guaranteed when the number of seeds grows logarithmically in the number of graph vertices. The logarithmic coefficient is shown to be inversely proportional to the mutual information between the edge variables in the two graphs.

This work was supported by NSF grant CCF-1815821. This work was presented in part at IEEE International Symposium on Information Theory (ISIT), July 2018.

I. Introduction

The graph matching problem models problems in a variety of applications including social network de-anonymization, pattern recognition, and computational biology [1], [2]. In this problem, an agent is given a correlated pair of randomly generated graphs: i) an 'anonymized' unlabeled graph, and ii) a 'de-anonymized' labeled graph as shown in Figure 1. The objective is to leverage the correlation among the edges of the graphs to recover the canonical labeling of the vertices in the anonymized graph.

There has been extensive research investigating the fundamental limits of graph matching, i.e. characterizing the necessary and sufficient conditions on graph parameters for successful matching. The problem has been considered under various probabilistic models capturing the correlation among the graph edges. In the *Correlated Erdös-Rényi* (CER) model the edges in the two graphs are pairwise correlated and are generated independently, based on identical distributions. More precisely, in this model, edges whose vertices are labeled identically are correlated through an arbitrary joint probability distribution and are generated independently of all other edges. In its simplest form — where the edges of the two graphs are exactly equal — graph matching is called *graph isomorphism*. Tight necessary and sufficient conditions for successful matching in the graph isomorphism scenario were derived in [3], [4] and polynomial time algorithms were proposed in [5]–[7]. The problem of matching non-identical pairs of correlated Erdös- Rényi graphs was studied in [8]–[14] and conditions for successful matching were derived. The CER model assumes the existence of statistical correlation among edge pairs connecting matching vertices in the two graphs, where the correlation model is based on an identical distribution among all matching edge pairs. Consequently, it does not model the community structure among the graph nodes which manifests in many applications [15], [16]. As an example, in social networks, users may be divided into communities based on various factors such as age-group, profession, and racial background. The users' community memberships affects the probability that they are connected with each other. A matching algorithm may potentially use the community membership information to enhance its performance. In order to take the users' community memberships into account, an extension to the CER model is considered which is called the *Community Structure* (CS) model. In this model, the edge probabilities depend on their corresponding vertices' community memberships. There have been several works studying both necessary and sufficient conditions for graph matching and the design of practical matching

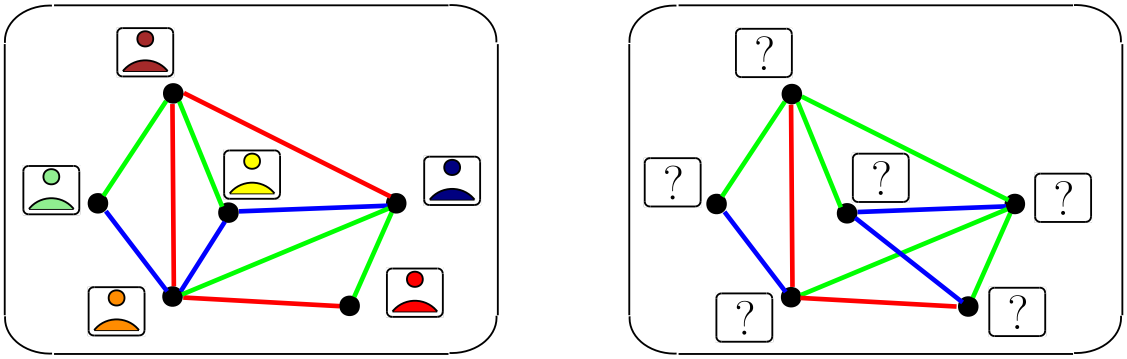

Fig. 1. An instance of the graph matching problem where the anonymized graph on the right is to be matched to the de-anonymized graph on the left.

schemes under the CS model [17], [18]. However, characterizing tight necessary and sufficient conditions for successful matching and designing polytime algorithms which are reliable under these conditions remains an open problem both in the CER and CS settings. A further extension of the problem, called *'seeded graph matching'* has also been investigated in the literature [19]–[28]. Seeded graph matching models applications where the matching agent has access to additional side-information in the form of pre-matched *seeds*. A seed vertex is one whose correct label in both graphs is known prior to the start of the matching process. One pertinent application of seeded graph matching is the de-anonymization of users over multiple social networks. Many web users are members of multiple online social networks such as Facebook, Twitter, Google+, LinkedIn, etc.. Each online network represents a subset of the users' "real" ego-networks. Graph matching provides algorithms to de-anonymize the users by reconciling these online network graphs, that is, to identify all the accounts belonging to the same individual. In this context, the availability of seeds is justified by the fact that a small fraction of individuals explicitly link their accounts across multiple networks. In this case, these linked accounts can be used as seeds in the matching algorithm. It turns out, that in many cases, these connections may be leveraged to identify a very large fraction of the users in the network [20]–[24]. In parallel to the study of fundamental limits of graph matching described above, there has been extensive research on the design of practical low complexity matching algorithms [29]–[31], where reliable matching of real-world networks with up to millions of nodes have been performed.

In this work, we use concentration of measure theorems in order to investigate the fundamental limits of graph matching, and propose the *'typicality matching'* (TM) strategy which operates based on the concept of typicality of sequences of random variables [32], and is applicable under a wide range of graph models including CER,CS and seeded graph matching. In summary, the strategy considers the pair of adjacency matrices corresponding to the two graphs. Each $n \times n$ adjacency matrix may be viewed as an n^2-length sequence of random variables, where n is the number of vertices in the graph. Consequently, one may naturally extend the notion of typicality of sequences of random variables to that of random adjacency matrices. The TM strategy finds a labeling for the vertices in the anonymized graph which results in a pair of jointly typical adjacency matrices for the two graphs, where typicality is defined with respect to the underlying joint edge distribution. The success of the matching algorithm is investigated as the graph size grows asymptotically large. The matching algorithm is said to succeed if the fraction of correctly matched vertices approaches one as the number of vertices goes to infinity. Consequently, the TM algorithm is successful as long as any labeling which leads to a pair of jointly typical adjacency matrices assigns an incorrect label to a negligible fraction of size $o(n)$ vertices in the anonymized graph[1]. In order to study the conditions for the success of the TM strategy, we derive several new bounds on the probability of joint typicality of permutations of sequences of random variables. The bounds may be of independent interest in other research areas as well. The generality of the information theoretic approach allows us to investigate matching under a wide range of statistical models. In addition to deriving new conditions for successful matching under the CER and CS graph models which have been studied in prior works, we also consider weighted graphs, where the graph edges are allowed to have non-binary attributes. We further extend the results to the simultaneous matching of more than two graphs. Additionally, we derive converse results which provide necessary conditions for successful matching based on model parameters. Furthermore, we consider seeded graph matching and derive theoretical guarantees for successful matching as a function of the seed-set size and the parameters of the statistical model. In the case of seeded graph matching, we provide a matching algorithm whose complexity grows polynomially in the number of vertices. We further derive converse results by providing necessary conditions for successful matching as a function of the seed set size.

The rest of the paper is organized as follows: Section II describes the notation used in the paper. Section III provides the problem formulation. Section IV develops the necessary tools

[1] We write $f(x) = o(g(x))$ if $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 0$.

for analyzing the performance of the TM algorithm. Section V studies matching under the CER model.Section VI considers matching under the CS model. Section VII investigates matching collections of more than two graphs. In Section VIII, necessary conditions and converse results for matching of pairs of graphs are investigated. Section IX shows the equivalence of a well-known criterion successful matching, which has been considered in the literature, with the one considered in this work. Section X studies seeded graph matching. Section XI concludes the paper.

II. Notation

We represent random variables by capital letters such as X, U and their realizations by small letters such as x, u. Sets are denoted by calligraphic letters such as $\mathcal{X}, \mathcal{U}$. The set of natural numbers, and the real numbers are represented by $\mathbb{N}$, and $\mathbb{R}$ respectively. The random variable $\mathbb{1}_{\mathcal{E}}$ is the indicator function of the event $\mathcal{E}$. The set of numbers $\{n, n+1, \cdots, m\}, n, m \in \mathbb{N}$ is represented by $[n, m]$. Furthermore, for the interval $[1, m]$, we sometimes use the shorthand notation $[m]$ for brevity. For a given $n \in \mathbb{N}$, the n-length vector $(x_1, x_2, \ldots, x_n)$ is written as x^n.

III. Problem Formulation

A graph $g = (\mathcal{V}, \mathcal{E})$ is characterized by the vertex set $\mathcal{V} = \{v_1, v_2, \cdots, v_n\}$, and the edge set $\mathcal{E}$. We consider weighted graphs, where each edge is assigned an *attribute* $x \in [l]$ and $l \geq 2$. Consequently, the edge set $\mathcal{E}$ is a subset of the set $\{(x, v_i, v_j) | i \neq j, x \in [l]\}$, where for each pair (v_i, v_j) there is a unique attribute x for which $(x, v_i, v_j) \in \mathcal{E}$. The edge attribute models the nature of the connection between the corresponding vertices. For instance in social network graphs, where vertices represent the members of the network and edges capture their connections, an edge may take different attributes depending on whether the members are family members, close friends, or acquaintances. A labeled graph $\tilde{g} = (g, \sigma)$ is a graph equipped with a bijective *labeling function* $\sigma : \mathcal{V} \to [n]$. The labeling represents the identity of the members in the social network. For a labeled graph $\tilde{g}$, the adjacency matrix $G = [g_{i,j}]_{i,j \in [n]}$ captures the edge attributes, where $g_{i,j}$ is the unique value for which $(g_{i,j}, v_i, v_j) \in \mathcal{E}$.

In this work, we consider graphs whose edges are generated stochastically based on an underlying probability distribution. Under the CER and CS models, we consider special instances of the following stochastic graph model.

Definition 1 (**Random Graph**). *A random graph $\tilde{g}$ generated based on $\prod_{i\in[n],j<i} P_{X_{i,j}}$ is an undirected labeled graph, where the edge between $v_i, i \in [n]$ and $v_j, j < i$ is generated according to $P_{X_{\sigma(i),\sigma(j)}}$ independently of the other edges. Alternatively,*

$$P((x, v_i, v_j) \in \mathcal{E}) = P_{X_{\sigma(i),\sigma(j)}}(x), x \in [l], i, j \in [n].$$

In the graph matching problem, we are given a pair correlated graphs $(\tilde{g}^1, \tilde{g}^2)$, where only the labeling for the vertices of the first graph is available. The objective is to recover the labeling of the vertices in the second graph by leveraging the correlation among their edges. A pair of correlated random graphs is defined below.

Definition 2 (**Correlated Random Graph**). *A pair of correlated random graphs $(\tilde{g}^1, \tilde{g}^2)$ generated based on $\prod_{i\in[n],j<i} P_{X^1_{i,j},X^2_{i,j}}$ is a pair of undirected labeled graphs. Let v^1, w^1 and v^2, w^2 be two pairs of similarly labeled vertices in $\tilde{g}^1$ and $\tilde{g}^2$, respectively i.e. $\sigma^1(v^1) = \sigma^2(v^2) = s_1$ and $\sigma^1(w^1) = \sigma^2(w^2) = s_2$. Then, the pair of edges between (v^1, w^1) and (v^2, w^2) are generated according to $P_{X^1_{s_1,s_2},X^2_{s_1,s_2}}$. Alternatively,*

$$P((x^1, v^1_i, w^1_j) \in \mathcal{E}^1, (x^2, v^2_i, w^2_j) \in \mathcal{E}^2) = P_{X^1_{s_1,s_2},X^2_{s_1,s_2}}(x^1, x^2), x \in [l], i, j \in [n].$$

A graph matching strategy takes $(\tilde{g}^1, g^2)$ as its input and outputs $(\tilde{g}^1, \hat{g}^2)$, where g^2 is the graph $\tilde{g}^2$ with its labels removed, and $\hat{g}^2$ is the relabeled graph. The matching strategy is said to succeed if the fraction of correctly matched vertices approaches one as the number of vertices is increased asymptotically. This is formalized below.

Definition 3 (**Matching Strategy**). *Consider a family of pairs of correlated random graphs $\tilde{g}^1_n = (g^1_n, \sigma^1_n)$ and $\tilde{g}^2_n = (g^2_n, \sigma^2_n), n \in \mathbb{N}$, where n is the number of vertices. A matching strategy is a sequence of functions $f_n : (\tilde{g}^1_n, g^2_n) \to (\tilde{g}^1_n, \hat{g}^2_n), n \in \mathbb{N}$, where $\hat{g}^2_n = (g^2_n, \hat{\sigma}^2_n)$ and $\hat{\sigma}^2_n$ is the reconstruction of σ^2. Let I_n be distributed uniformly over $[n]$. The matching strategy is said to succeed if $P\left(\sigma^2(v^2_{I_n}) = \hat{\sigma}^2(v^2_{I_n})\right) \to 1$ as $n \to \infty$.*

The following defines an achievable region for the graph matching problem.

Definition 4 (**Achievable Region**). *For the graph matching problem, a family of sets of distributions $\widetilde{P} = (\mathcal{P}_n)_{n\in\mathbb{N}}$ is said to be in the achievable region if for every sequence of distributions $\prod_{s_1\in[n],s_2<s_1} P^{(n)}_{X^1_{s_1,s_2},X^2_{s_1,s_2}} \in \mathcal{P}_n$,, there exists a successful matching strategy. The maximal achievable*

family of sets of distributions is denoted by $\mathcal{P}^*$.

IV. Permutations of Typical Sequences

In the previous section, we described correlated pairs of random graphs, where the graph edges are generated randomly based on an underlying joint distribution. Alternatively, the adjacency matrices of the graphs are generated according to a joint distribution. Furthermore, as explained in Definition 2, we assume that each edge pair connecting two similarly labeled vertices in the two graphs is generated independently of all other edges based on the distribution $P_{X_{i,j}}$, where i, j are the vertex labels. Consequently, it is expected, given large enough graph sizes, that the adjacency matrices of the graphs look '*typical*' with respect to the joint edge distribution. Roughly speaking, this requires the frequency of joint occurrence of symbols (x^1, x^2) to be close to $\frac{1}{n^2}\sum_{i=1}^n \sum_{j=1}^n P_{X^1_{i,j},X^2_{i,j}}(x^1,x^2)$, where $x^1, x^2 \in [l]$. Based on this observation, in the next sections we propose the typicality matching strategy which operates by finding the labeling for the second graph which results in a jointly typical pair of adjacency matrices. This is analogous to typicality decoding in the channel coding problem in information theory, where the decoder finds the transmitted sequence by searching for a codeword which is jointly typical with the received sequence. In this analogy, the labeled graph $\tilde{g}^2$ is passed through a '*channel*' which outputs g^2, and the '*decode*r' wants to recover $\tilde{g}^2$ using g^2 and the side-information $\tilde{g}^1$. Changing the labeling of g^2 leads to a permutation of its adjacency matrix. Hence, we need to search over permutations of the adjacency matrix and find the one which leads to a typical pair of adjacency matrices. The error analysis of the typicality matching strategy requires investigating the probability of joint typicality of permutations of pairs of correlated sequences.

In this section, we analyze the joint typicality of permutations of collections of correlated sequences of random variables. While the analysis is used in the subsequent sections to derive the necessary and sufficient conditions for successful matching in various graph matching scenarios, it may also be of independent interest in other research areas as well.

We follow the notation used in [33] in our study of permutation groups which is summarized below.

Definition 5 (Permutation). *A permutation on the set of numbers* $[1, n]$ *is a bijection* $\pi : [1, n] \to [1, n]$*. The set of all permutations on the set of numbers* $[1, n]$ *is denoted by* S_n*.*

Definition 6 (Cycle). *A permutation $\pi \in S_n, n \in \mathbb{N}$ is called a cycle if there exists $m \in [1,n]$ and $\alpha_1, \alpha_2, \cdots, \alpha_m \in [1,n]$ such that i) $\pi(\alpha_i) = \alpha_{i+1}, i \in [1, m-1]$, ii) $\pi(\alpha_n) = \alpha_1$, and iii) $\pi(\beta) = \beta$ if $\beta \neq \alpha_i, \forall i \in [1,m]$. The variable m is called the length of the cycle. The element α is called a fixed point of the permutation if $\pi(\alpha) = \alpha$. We write $\pi = (\alpha_1, \alpha_2, \cdots, \alpha_m)$. The permutation π is called a non-trivial cycle if $m \geq 2$.*

Lemma 1 ([33]). *Every permutation $\pi \in S_n, n \in \mathbb{N}$ has a unique representation as a product of disjoint non-trivial cycles.*

Definition 7 (Sequence Permutation). *For a given sequence $y^n \in \mathbb{R}^n$ and permutation $\pi \in S_n$, the sequence $z^n = \pi(y^n)$ is defined as $z^n = (y_{\pi(i)})_{i \in [1,n]}$.*[2]

A. Typicality of Permutations of Pairs of Correlated Sequences

As a first step, we consider typicality of permutations pairs of correlated sequences.

Definition 8 (Strong Typicality). *Let the pair of random variables (X, Y) be defined on the probability space $(\mathcal{X} \times \mathcal{Y}, P_{X,Y})$, where $\mathcal{X}$ and $\mathcal{Y}$ are finite alphabets. The ϵ-typical set of sequences of length n with respect to $P_{X,Y}$ is defined as:*

$$A_\epsilon^n(X, Y) = \left\{(x^n, y^n) : \left|\frac{1}{n} N(\alpha, \beta | x^n, y^n) - P_{X,Y}(\alpha, \beta)\right| \leq \epsilon, \forall (\alpha, \beta) \in \mathcal{X} \times \mathcal{Y}\right\},$$

where $\epsilon > 0$, $n \in \mathbb{N}$, and $N(\alpha, \beta | x^n, y^n) = \sum_{i=1}^n \mathbb{1}((x_i, y_i) = (\alpha, \beta))$.

For a correlated pair of independent and identically distributed (i.i.d) sequences (X^n, Y^n) and an arbitrary permutation $\pi \in S_n$, we are interested in bounding the probability $P((X^n, \pi(Y^n)) \in \mathcal{A}_\epsilon^n(X, Y))$. In our analysis, we make extensive use of the standard permutations defined below.

Definition 9 (Standard Permutation). *For a given $n, m, c \in \mathbb{N}$, and $1 \leq i_1 \leq i_2 \leq \cdots \leq i_c \leq n$ such that $n = \sum_{j=1}^c i_j + m$, an $(m, c, i_1, i_2, \cdots, i_c)$-permutation is a permutation in S_n which has m fixed points and c disjoint cycles with lengths $i_1, i_2, \cdots, i_c$, respectively.*

[2] Note that in Definitions 5 and 7 we have used π to denote both a scalar function which operates on the set $[1,n]$ as well as a function which operates on the vector space $\mathbb{R}^n$.

The $(m,c,i_1,i_2,\cdots,i_c)$-standard permutation is defined as the $(m,c,i_1,i_2,\cdots,i_c)$-permutation consisting of the cycles $(\sum_{j=1}^{k-1} i_j+1, \sum_{j=1}^{k-1} i_j+2,\cdots,\sum_{j=1}^{k} i_j), k\in[1,c]$. Alternatively, the $(m,c,i_1, i_2,\cdots,i_c)$-standard permutation is defined as:

$$\pi=(1,2,\cdots,i_1)(i_1+1,i_1+2,\cdots,i_1+i_2)\cdots \\ (\sum_{j=1}^{c-1} i_j+1,\sum_{j=1}^{c-1} i_j+2,\cdots,\sum_{j=1}^{c} i_j)(n-m+1)(n-m+2)\cdots(n).$$

Example 1. The $(2,2,3,2)$-standard permutation is a permutation which has $m=2$ fixed points and $c=2$ cycles. The first cycle has length $i_1=3$ and the second cycle has length $i_2=2$. It is a permutation on sequences of length $n=\sum_{j=1}^{c} i_j+m=3+2+2=7$. The permutation is given by $\pi=(123)(45)(6)(7)$. For an arbitrary sequence $\underline{\alpha}=(\alpha_1,\alpha_2,\cdots,\alpha_7)$, we have:

$$\pi(\underline{\alpha})=(\alpha_3,\alpha_1,\alpha_2,\alpha_5,\alpha_4,\alpha_6,\alpha_7).$$

The following proposition shows that in order to find bounds on the probability of joint typicality of permutations of correlated sequences, it suffices to study standard permutations.

Proposition 1. *Let (X^n,Y^n) be a pair of i.i.d sequences defined on finite alphabets. We have:*
i) For an arbitrary permutation $\pi\in S_n$,

$$P((\pi(X^n),\pi(Y^n))\in\mathcal{A}_\epsilon^n(X,Y))=P((X^n,Y^n)\in\mathcal{A}_\epsilon^n(X,Y)).$$

ii) let $n,m,c,i_1,i_2,\cdots,i_c\in\mathbb{N}$ be numbers as described in Definition 9. Let π_1 be an arbitrary $(m,c,i_1,i_2,\cdots,i_c)$-permutation and let π_2 be the $(m,c,i_1,i_2,\cdots,i_c)$-standard permutation. Then,

$$P((X^n,\pi_1(Y^n))\in\mathcal{A}_\epsilon^n(X,Y))=P((X^n,\pi_2(Y^n))\in\mathcal{A}_\epsilon^n(X,Y)).$$

Proof. The proof of part i) follows from the fact that permuting both X^n and Y^n by the same permutation does not change their joint type. For part ii), it is known that there exists a permutation π such that $\pi(\pi_1)=\pi_2(\pi)$ [33]. Then the statement is proved using part i) as follows:

$$P\left((X^n,\pi_1(Y^n))\in\mathcal{A}_\epsilon^n(X,Y)\right)=P\left((\pi(X^n),\pi(\pi_1(Y^n)))\in\mathcal{A}_\epsilon^n(X,Y)\right)$$
$$=P\left((\pi(X^n),\pi_2(\pi(Y^n)))\in\mathcal{A}_\epsilon^n(X,Y)\right)\overset{(a)}{=}P\left(\left(\widetilde{X}^n,\pi_2\left(\widetilde{Y}^n\right)\right)\in\mathcal{A}_\epsilon^n(X,Y)\right)\overset{(b)}{=}P\left((X^n,\pi_2(Y^n))\in\mathcal{A}_\epsilon^n(X,Y)\right),$$

where in (a) we have defined $(\widetilde{X}^n, \widetilde{Y}^n) = (\pi(X^n), \pi(Y^n))$. and (b) holds since $(\widetilde{X}^n, \widetilde{Y}^n)$ has the same distribution as (X^n, Y^n). □

The following theorem provides upper-bound on the probability of joint typicality of permutations of correlated sequences for an arbitrary permutation with $m \in [n]$ fixed points.

Theorem 1. *Let (X^n, Y^n) be a pair of i.i.d sequences defined on finite alphabets $\mathcal{X}$ and $\mathcal{Y}$, respectively. For any permutation π with $m \in [n]$ fixed points, the following holds:*

$$P((X^n, \pi(Y^n)) \in \mathcal{A}_\epsilon^n(X, Y)) \leq 2^{-\frac{n}{4}(D(P_{X,Y}||(1-\alpha)P_X P_Y + \alpha P_{X,Y}) - |\mathcal{X}||\mathcal{Y}|\epsilon + O(\frac{\log n}{n}))}, \qquad (1)$$

where $\alpha = \frac{m}{n}$, and $D(\cdot||\cdot)$ is the Kullback-Leibler divergence.

Proof. Appendix A. □

Remark 1. *The upper bound in Equation* (1) *goes to* 0 *as $n \to \infty$ for any non-trivial permutation (i.e. α bounded away from one) and small enough ϵ, as long as X and Y are not independent.*

The exponent $D(P_{X,Y}||(1 - \alpha)P_X P_Y + \alpha P_{X,Y})$ in Equation (1) can be interpreted as follows: for the fixed points of the permutation (α fraction of indices), we have $\pi(Y_i) = Y_i$. As a result, the joint distribution of the elements $(X_i, \pi(Y_i))$ is $P_{X,Y}$. For the rest of the elements, $\pi(Y_i)$ are permuted components of Y^n, as a result $(X_i, \pi(Y_i))$ are an independent pair of variables since (X^n, Y^n) is a correlated pair of i.i.d. sequences. Consequently, the distribution of $(X_i, \pi(Y_i))$ is $P_X P_Y$ for $(1-\alpha)$ fraction of elements which are not fixed points of the permutation. The average distribution is $(1-\alpha)P_X P_Y + \alpha P_{X,Y}$ which appears as the second argument in the Kullback-Leibler Divergence in Equation (1).

Theorem 1 provides bounds on the probability of joint typicality of X^n and $\pi(Y^n)$ as a function of the number of fixed points m of the permutation $\pi(\cdot)$. A parameter of interest is the number of distinct permutations with a specific number of fixed points and its limiting behavior.

Definition 10 (Derangement). *Let $n \in \mathbb{N}$. A permutation on vectors of length n is called a derangement if it does not have any fixed points. The number of distinct derangements of n-length vectors is denoted by $!n$.*

Lemma 2. *Let $n \in \mathbb{N}$. Let N_m be the number of distinct permutations with exactly $m \in [0, n]$ fixed points. Then,*

$$\frac{n!}{m!(n-m)} \leq N_m = \binom{n}{m}!(n-m) \leq n^{n-m}. \tag{2}$$

Particularly, let $m = \alpha n, 0 < \alpha < 1$. Then, the following holds:

$$\lim_{n\to\infty} \frac{\log N_m}{n \log n} = 1 - \alpha. \tag{3}$$

Proof. Appendix B. □

In the following, we investigate whether the exponent in Equation (1) is tight (i.e. whether the exponent can be improved to arrive at a tighter upper-bound). Previously, we provided the justification for the appearance of the term $D(P_{X,Y}||(1-\alpha)P_XP_Y + \alpha P_{X,Y})$ in the exponent in Equation (1). However, a more careful analysis may yield improvements in the coefficient $\frac{n}{4}$ by focusing on specific classes of permutations as described in the following. As a first step, we only consider permutations consisting of a single non-trivial cycle and no fixed points.

Lemma 3. *Let (X^n, Y^n) be a pair of i.i.d sequences defined on finite alphabets $\mathcal{X}$ and $\mathcal{Y}$, respectively. For any permutation π with no fixed points, and a single cycle (i.e. $m = 0$ and $c = 1$), the following holds:*

$$P((X^n, \pi(Y^n)) \in \mathcal{A}_\epsilon^n(X, Y)) \leq 2^{-\frac{n}{2}(I(X;Y)-\delta)}, \tag{4}$$

where $\delta = 2\sum_{x,y} |\log_2 \frac{P_{X,Y}(x,y)}{P_X(x)P_Y(y)}|\epsilon$ and $\epsilon > 0$.

Proof. Appendix C. □

Remark 2. *Note that Theorem 1 can also be applied to derive a bound on the probability of joint typicality given the permutation considered in Lemma 3. In this case $\alpha = \frac{m}{n} = 0$ and $D(P_{X,Y}||\alpha P_{X,Y} + (1-\alpha)P_XP_Y) = I(X;Y)$ and Theorem 1 yeilds the exponent $\frac{n}{4}I(X;Y)$ for the probability of joint typicality. Hence, Lemma 3 improves the exponent $\frac{n}{4}I(X;Y)$ in Theorem 1 to $\frac{n}{2}I(X;Y)$ for single-cycle permutations with no fixed points.*

The following lemma derives similar results for permutations with a large number of short cycles (e.g. cycles of length two or three) and no fixed points.

Lemma 4. *Let (X^n, Y^n) be a pair of correlated sequences of i.i.d variables defined on finite alphabets $\mathcal{X}$ and $\mathcal{Y}$, respectively. For any $(n, m, c, i_1, i_2, \cdots, i_c)$-permutation π with no fixed points (m=0), where $0 < i_1 < i_2 < \cdots < i_c < s < n$, the following holds:*

$$P((X^n, \pi(Y^n)) \in \mathcal{A}_\epsilon^n(X, Y)) \leq 2^{-\frac{n}{s}(I(X;Y)-\delta)}, \tag{5}$$

where $\delta = \sum_{x,y} |\log_2 \frac{P_{X,Y}(x,y)}{P_X(x)P_Y(y)}|\epsilon$ and $\epsilon > 0$.

Proof. Appendix D. □

Remark 3. *Lemma 4 improves the exponent in Theorem 1 when the maximum cycle length is less than or equal to $s = 3$.*

B. Typicality of Permutations of Collections of Correlated Sequences

In the next step, we consider joint typicality of permutations of collections of correlated sequences.

Definition 11 (Strong Typicality of Collections of Sequences). *Let the random vector X^m be defined on the probability space $(\prod_{j\in[m]} \mathcal{X}_j, P_{X^m})$, where $\mathcal{X}_j, j \in [m]$ are finite alphabets, and $m > 2$. The ϵ-typical set of sequences of length n with respect to P_{X^m} is defined as:*

$$\mathcal{A}_\epsilon^n(X^m) = \left\{(x_{(j)}^n)_{j\in[m]} : \left|\frac{1}{n}N(\alpha^m|x_{(1)}^n, x_{(2)}^n, \cdots, x_{(m)}^n) - P_{X^m}(\alpha^m)\right| \leq \epsilon, \forall \alpha^m \in \prod_{j\in[m]} \mathcal{X}_j\right\},$$

where $\epsilon > 0$, $N(\alpha^m|x_{(1)}^n, x_{(2)}^n, \cdots, x_{(m)}^n) = \sum_{i=1}^n \mathbb{1}\left((x_{(j),i})_{j\in[m]} = \alpha^m\right)$, and $(x_{(j)}^n)_{j\in[m]} = (x_{(1)}^n, \cdots, x_{(m)}^n)$ is a vector of sequences.

In the previous section, in order to investigate the typicality of permutations of pairs of correlated sequences, we introduced standard permutations which are completely characterized by the number of fixed points, number of cycles, and cycle lengths of the permutation. The concept of standard permutations does not extend naturally when there are more than two sequences (i.e. more than one non-trivial permutation). Consequently, investigating typicality of permutations of collections of sequences requires developing additional analytical tools which are described in the following.

Definition 12 (Bell Number). *Consider the set $\mathcal{N} = [1,n]$. Let $\mathsf{P} = \{\mathcal{P}_1, \mathcal{P}_2, \cdots, \mathcal{P}_{b_m}\}$ be the set of all partitions of $\mathcal{N}$, where $\mathcal{P}_k = \{\mathcal{A}_{k,1}, \mathcal{A}_{k,2} \cdots, \mathcal{A}_{k,|\mathcal{P}_k|}\}$. The natural number b_m is the mth Bell number.*

In the following, we define Bell permutation vectors which are analogous to standard permutations for the case when the problem involves more than one non-trivial permutation.

Definition 13 (Partition Correspondence). *Let $m, n \in \mathbb{N}$ and $(\pi_1, \pi_2, \cdots, \pi_m)$ be arbitrary permutations operating on n-length vectors. The index $i \in [n]$ is said to correspond to the partition $\mathcal{P}_k = \{\mathcal{A}_{k,1}, \mathcal{A}_{k,2}, \cdots, \mathcal{A}_{k,|\mathcal{P}_k|}\}$ of the set $[1,m]$ if the following holds:*

$$\forall j, j' \in [m] : \pi_j^{-1}(i) = \pi_{j'}^{-1}(i) \iff \exists r : j, j' \in \mathcal{A}_{k,r}$$

To explain the above definition, let us consider a triple of permutations of n-length sequences, i.e. $m = 3$, and the partition $\mathcal{P}_k = \{\{1,2\},\{3\}\}$. Then an index $i \in [n]$ corresponds to the partition $\mathcal{P}_k$ if the first two permutation map the index to the same integer and the third permutation maps the index to a different integer.

Definition 14 (Bell Permutation Vector). *Let $(i_1, i_2, \cdots, i_{b_m})$ be an arbitrary sequence, where $\sum_{k \in [b_m]} i_k = n, i_k \in [0,n]$, b_m is the mth Bell number, and $n, m \in \mathbb{N}$. The vector of permutations $(\pi_1, \pi_2, \cdots, \pi_m)$ is called an $(i_1, i_2, \cdots, i_{b_m})$-Bell permutation vector if for every partition $\mathcal{P}_k$ exactly i_k indices correspond to that partition. Equivalently:*

$$\forall k \in [b_m] : i_k = |\{i \in [n] : \forall j, j' \in [m] : \pi_j^{-1}(i) = \pi_{j'}^{-1}(i) \iff \exists r : j, j' \in \mathcal{A}_{k,r}\}|.$$

The definition of Bell permutation vectors is further clarified through the following example.

Example 2. Consider 3 permutations (π_1, π_2, π_3) of vectors with length equal to 7, i.e. $m = 3$ and $n = 7$. Then, $b_m = 5$ and we have:

$$\mathcal{P}_1 = \{\{1\},\{2\},\{3\}\}, \quad \mathcal{P}_2 = \{\{1,2\},\{3\}\}, \quad \mathcal{P}_3 = \{\{1,3\},\{2\}\},$$
$$\mathcal{P}_4 = \{\{1\},\{2,3\}\}, \quad \mathcal{P}_5 = \{\{1,2,3\}\}.$$

Let π_1 be the trivial permutation fixing all indices and let $\pi_2 = (135)(24)$, $\pi_3 = (15)(24)(37)$. Then:

$$\pi_1((1,2,\cdots,7)) = (1,2,3,4,5,6,7),$$

$$\pi_2((1,2,\cdots,7)) = (5,4,1,2,3,6,7),$$

$$\pi_3((1,2,\cdots,7)) = (5,4,7,2,1,6,3),$$

Then, the vector (π_1,π_2,π_3) is a $(2,1,0,3,1)$-Bell permutation vector, where the indices $(3,5)$ correspond to the $\mathcal{P}_1$ partition (each of the three permutations map the index to a different integer), index 7 corresponds to the $\mathcal{P}_2$ partition (the first two permutations map the index to the same integer which is different from the one for the third permutation), indices $(1,2,4)$ correspond to the $\mathcal{P}_3$ permutation (the second and third permutations map the index to the same integer which is different from the output of the first permutation), and index 6 corresponds to $\mathcal{P}_5$ (all permutations map the index to the same integer).

Remark 4. *Bell permutation vectors are not unique. In other words, there can be several distinct $(i_1,i_2,\cdots,i_{b_m})$-Bell permutation vectors for given $n,m,i_1,i_2,\cdots,i_{b_m}$. This is in contrast with standard permutations defined in Definition 9, which are unique given the parameters $n,c,i_1,i_2,\cdots,i_c$.*

The following theorem provides bounds on the probability of joint typicality of permutations of collections of correlated sequences:

Theorem 2. *Let $(X^n_{(j)})_{j\in[m]}$ be a collection of correlated sequences of i.i.d random variables defined on finite alphabets $\mathcal{X}_{(j)}, j\in[m]$. For any $(i_1,i_2,\cdots,i_{b_m})$-Bell permutation vector $(\pi_1,\pi_2,\cdots,\pi_m)$, the following holds:*

$$P((\pi_i(X^n_{(i)})_{i\in[m]} \in \mathcal{A}^n_\epsilon(X^m)) \leq 2^{-\frac{n}{m(m-1)b_m}(D(P_{X^m}\| \sum_{k\in[b_m]}\frac{i_k}{n}P_{X_{\mathcal{P}_k}})-\epsilon\prod_{j\in[m]}|\mathcal{X}_j|+O(\frac{\log n}{n}))}, \tag{6}$$

where $P_{X_{\mathcal{P}_k}} = \prod_{l\in[1,|\mathcal{P}_k|]} P_{X_{k_1},X_{k_2},\cdots,X_{k_{|\mathcal{A}_{k,r}|}}}$, $\mathcal{A}_{k,r} = \{l_1,l_2,\cdots,l_{|\mathcal{A}_{k,r}|}\}, k\in[b_m], r\in[1,|\mathcal{P}_k|]$, and $D(\cdot\|\cdot)$ is the Kullback-Leibler divergence.

Proof. Appendix E. □

Note that for permutations of pairs of sequences of random variables, $m=2$ and the second Bell number is $b_2=2$. In this case $m(m-1)b_m=4$, and the bound on the probability of joint typicality given in Theorem 2 recovers the one in Theorem 1. In the following, we provide upper and lower bounds on the number of distinct Bell permutation vectors for a given vector $(i_1,i_2,\cdots,i_{b_m})$.

Definition 15 (r-fold Derangement). *A vector $(\pi_1(\cdot), \pi_2(\cdot), \cdots, \pi_r(\cdot))$ of permutations of n-length sequences is called an r-fold derangement if $\pi_1(\cdot)$ is the identity permutation, and $\pi_l(i) \neq \pi_{l'}(i), l, l' \in [r], l \neq l', i \in [n]$. The number of distinct r-fold derangements of $[n]$ is denoted by $d_r(n)$. Particularly $d_2(n) = !n$ is the number of derangements of $[n]$.*

Lemma 5. *Let $n \in \mathbb{N}$ and $r \in [n]$. Then,*

$$((n - r + 1)!)^{r-1} \leq d_r(n) \leq (!n)^{r-1}.$$

Proof. Appendix F. □

Lemma 6. *Let $(i_1, i_2, \cdots, i_{b_m})$ be a vector of non-negative integers such that $\sum_{k \in [b_m]} i_k = n$. Define $N_{i_1, i_2, \cdots, i_{b_m}}$ as the number of distinct $(i_1, i_2, \cdots, i_{b_m})$-Bell permutation vectors. Then,*

$$\binom{n}{i_1, i_2, \cdots, i_{b_m}} \prod_{k \in [b_m]} d_{|\mathcal{P}_k|}(i_k) \leq N_{i_1, i_2, \cdots, i_{b_m}} \leq \binom{n}{i_1, i_2, \cdots, i_{b_m}} n^{\sum_{k \in [b_m]} |\mathcal{P}_k| i_k - n}. \tag{7}$$

Particularly, let $i_k = \alpha_k \cdot n, n \in \mathbb{N}$. The following holds:

$$\lim_{n \to \infty} \frac{\log N_{i_1, i_2, \cdots, i_{b_m}}}{n \log n} = \sum_{k \in [b_m]} |\mathcal{P}_k| \alpha_k - 1. \tag{8}$$

Proof. Appendix G. □

V. Matching Erdös-Rènyi Graphs

In this section, we consider matching of correlated pairs of Erdös-Rènyi (CPER) graphs. In section III, we described correlated random graphs. A CPER is a special instance of the correlated random graphs defined in Definition 2. We propose the typicality matching strategy and provide sufficient conditions on the joint edge statistics under which the strategy succeeds.

A. Problem Setup

In order to describe the notation used in this section, we formally define labeled graphs below.

Definition 16 (Labeled Graphs). *A labeling is a bijective function $\sigma : \mathcal{V} \to [1, n]$. The pair $\tilde{g} = (g, \sigma)$ is called an (n, l)-labeled graph. For the labeled graph $\tilde{g}$ the adjacency matrix is defined as $G_\sigma = [g_{\sigma,i,j}]_{i,j \in [1,n]}$ where $g_{\sigma,i,j}$ is the unique value such that $(g_{\sigma,i,j}, v_i, v_j) \in \mathcal{E}_n$, where $(v_i, v_j) = (\sigma^{-1}(i), \sigma^{-1}(j))$. The upper triangle (UT) corresponding to $\tilde{g}$ is the structure $U_\sigma = [G_{\sigma,i,j}]_{i<j}$. The subscript '$\sigma$' is dropped when there is no ambiguity.*

Remark 5. *In the context of Definition 16, an unlabeled graph with binary valued edges is a graph for which $l = 2$. In this case, if the pair $v_{n,i}$ and $v_{n,i}$ are not connected, we write $(0, v_{n,i}, v_{n,j}) \in \mathcal{E}$, otherwise $(1, v_{n,i}, v_{n,j}) \in \mathcal{E}$.*

Remark 6. *Without loss of generality, we assume that for any arbitrary pair of vertices $(v_{n,i}, v_{n,j})$, there exists a unique $x \in [0, l-1]$ such that $(x, v_{n,i}, v_{n,j}) \in \mathcal{E}$.*

Remark 7. *In this work, we often consider sequences of graphs $g^{(n)}, n \in \mathbb{N}$, where $g^{(n)}$ has n vertices. In such instances, we write $g^{(n)} = (\mathcal{V}^{(n)}, C^{(n)}, \mathcal{E}^{(n)})$ to characterize the nth graph in the sequence. The superscript '(n)' is omitted where there is no ambiguity.*

Remark 8. *In this work, we only consider undirected graphs where $(x, v_i, v_j) \in \mathcal{E}$ if and only if $(x, v_j, v_i) \in \mathcal{E}$. The results can be extended to directed graphs in a straightforward manner.*

Any pair of labeling functions are related through a permutation as described below.

Definition 17. *For two labelings σ and σ', the (σ, σ')-permutation is defined as the bijection $\pi_{(\sigma,\sigma')}$, where:*

$$\pi_{(\sigma,\sigma')}(i) = j, \quad if \quad \sigma'^{-1}(j) = \sigma^{-1}(i), \forall i, j \in [1, n].$$

Definition 18 (Correlated Pair of ER Graphs). *Let $P_{X,X'}$ be a conditional distribution defined on $\mathcal{X} \times \mathcal{X}'$, where $\mathcal{X} = \mathcal{X}' = [0, l-1]$. A correlated pair of ER graphs $\underline{\tilde{g}} = (\tilde{g}, \tilde{g}')$ generated according to $P_{X,X'}$ is characterized by: i) the pair of ER graphs (g, g') generated according to P_X and $P_{X'}$, respectively, ii) the pair of labelings (σ, σ') for the unlabeled graphs (g, g'), and iii) the probability distribution $P_{X,X'}$, such that:*
1)The graphs have the same set of vertices $\mathcal{V} = \mathcal{V}'$.
2) For any two edges $e = (x, v_{j_1}, v_{j_2}), e' = (x', v'_{j'_1}, v'_{j'_2}), x, x' \in [0, l-1]$, we have

$$Pr\left(e \in \mathcal{E}, e' \in \mathcal{E}'\right) = \begin{cases} P_{X,X'}(x, x'), & if\ \sigma(v_{j_l}) = \sigma'(v'_{j'_l}) \\ P_X(x)P_{X'}(x'), & Otherwise \end{cases},$$

where $l \in \{1, 2\}$, $v_{j_1}, v_{j_2} \in \mathcal{V}_1 \times \mathcal{V}_2$, and $v'_{j'_1}, v'_{j'_2} \in \mathcal{V}_1 \times \mathcal{V}_2$.

B. The Typicality Matching Strategy for CERs

Given a correlated pair of graphs $\underline{g} = (\tilde{g}^1, g^2)$, where only the labeling for $\tilde{g}^1$ is given, the typicality matching strategy operates as follows. The scheme finds a labeling $\hat{\sigma}^2$, for which the pair of UT's $U^1_{\sigma^1}$ and $U^2_{\hat{\sigma}^2}$ are jointly typical with respect to P_{n,X_1,X_2} when viewed as vectors of length $\frac{n(n-1)}{2}$. The strategy succeeds if at least one such labeling exists and fails otherwise. Alternatively, it finds an element $\hat{\sigma}^2$ in the set:

$$\widehat{\Sigma} = \{\hat{\sigma}^2 | (U^1_{\sigma^1}, U^2_{\hat{\sigma}^2}) \in \mathcal{A}_\epsilon^{\frac{n(n-1)}{2}}(X_1, X_2)\},$$

where $\epsilon = \omega(\frac{1}{n})$. The algorithm declares $\hat{\sigma}^2$ as the correct labeling. Note that the set $\widehat{\Sigma}$ may have more than one element. We will show that under certain conditions on the joint graph statistics, all of the elements of $\widehat{\Sigma}$ satisfy the criteria for successful matching given in Definition 23. In other words, for all of the elements of $\widehat{\Sigma}$ the probability of incorrect labeling for any given vertex is arbitrarily small for large n.

Theorem 3. *For the typicality matching strategy, a given family of sets of distributions $\widetilde{P} = (\mathcal{P}_n)_{n\in\mathbb{N}}$ is achievable, if for every sequence of distributions $P_{n,X_1,X_2} \in \mathcal{P}_n, n \in \mathbb{N}$*

$$8(1-\alpha)\frac{\log n}{n-1} \leq D(P^{(n)}_{X_1,X_2} || (1-\alpha^2)P^{(n)}_{X_1}P^{(n)}_{X_2} + \alpha^2 P^{(n)}_{X_1,X_2}), 0 \leq \alpha \leq \alpha_n, \tag{9}$$

where α_n is a sequence such that $\alpha_n \to 1$ as $n \to \infty$.

Proof. Appendix H. □

Remark 9. *As described in Section IV, the bound in Equation* (9) *can be potentially tightened with the coefficient 8 in the left hand side replaced by 6.*

The Kullback Leibler divergence term $D(P^{(n)}_{X_1,X_2} || (1-\alpha^2)P^{(n)}_{X_1}P^{(n)}_{X_2} + \alpha^2 P^{(n)}_{X_1,X_2})$ in the right hand side of Equation (9) can be interpreted as follows. Let α be the fraction of the vertices which are matched correctly by the typicality matching scheme. Then almost α^2 elements in the two adjacency matrices of the graphs are in the correct position, and the rest are permuted. The elements which are in the correct position are distributed according to the joint distribution $P^{(n)}_{X_1,X_2}$, whereas the permuted elements are distributed according to $P^{(n)}_{X_1}P^{(n)}_{X_1}$, i.e. independently of each other. Consequently, the empirical joint distribution of the elements of the two matrices is close to $(1-\alpha^2)P^{(n)}_{X_1}P^{(n)}_{X_2} + \alpha^2 P^{(n)}_{X_1,X_2}$ with high probability. The typicality matching scheme outputs such a labeling if the resulting adjacency matrix — generated according to $(1-\alpha^2)P^{(n)}_{X_1}P^{(n)}_{X_2} + \alpha^2 P^{(n)}_{X_1,X_2}$

based on the above argument — is typical with respect to $P^{(n)}_{X_1,X_2}$. It is well-known that the error exponent for such a binary hypothesis test is equal to $D(P^{(n)}_{X_1,X_2}||(1-\alpha^2)P^{(n)}_{X_1}P^{(n)}_{X_2}+\alpha^2P^{(n)}_{X_1,X_2})$. Furthermore, the $(1-\alpha)\frac{\log n}{n-1}$ term on the left hand side is the exponent of the total number of permutations with α fraction of fixed points.

VI. Matching Graphs with Community Structure

In this section, we describe the typicality matching scheme for matching graphs with community structure and provide achievable regions for these matching scenarios.

A. Problem Setup

To describe the notation used in the section, consider a graph with $n \in \mathbb{N}$ vertices belonging to $c \in \mathbb{N}$ communities whose edges take $l \geq 2$ possible attributes. It is assumed that the set of communities $C = \{C_1, C_2, \cdots, C_c\}$ partitions the vertex set $\mathcal{V}$. The i^{th} community is written as $C_i = \{v_{j_1}, v_{j_2}, \cdots, v_{j_{n_i}}\}$. The following formally defines a graph with community structure.

Definition 19 (**Graph with Community Structure**). *An $(n, c, (n_i)_{i\in[c]}, l)$-unlabeled graph with community structure (UCS) g is characterized by the triple $(\mathcal{V}, C, \mathcal{E})$, where $n, l, c, n_1, n_2, \cdots, n_c \in \mathbb{N}$ and $l \geq 2$. The set $\mathcal{V} = \{v_1, v_2, \cdots, v_n\}$ is called the vertex set. The family of sets $C = \{C_1, C_2, \cdots, C_c\}$ provides a partition for $\mathcal{V}$ and is called the family of communities. The i^{th} community is written as $C_i = \{v_{j_1}, v_{j_2}, \cdots, v_{j_{n_i}}\}$. The set $\mathcal{E} \subset \{(x, v_{j_1}, v_{j_2})|x \in [0, l-1], j_1 \in [1, n], j_2 \in [1, n]\}$ is called the edge set of the graph. For the edge (x, v_{j_1}, v_{j_2}), the variable 'x' represents the value assigned to the edge between vertices v_{j_1} and v_{j_2}. The set $\mathcal{E}_{i_1,i_2} = \{(x, v_{j_1}, v_{j_2}) \in \mathcal{E}|v_{j_1} \in C_{i_1}, v_{j_2} \in C_{i_2}\}$ is the set of edges connecting the vertices in communities C_{i_1} and C_{i_2}.*

Remark 10. *Single-community graphs, where $c = 1$, have been studied extensively in the graph matching literature. For instance Erdös-Rényi (ER) graphs studied in Section V are single-community graphs.*

We consider graphs generated stochastically based on the community structure model. In this model, the probability of an edge between a pair of vertices is determined by their community memberships. More precisely, for a given vertex set $\mathcal{V}$ and set of communities C, it is assumed that the edge set $\mathcal{E}$ is generated randomly, where the attribute X of the edge between vertices $v_{i_1} \in C_{j_1}$ and $v_{i_2} \in C_{j_2}$ is generated based on the conditional distribution $P_{X|C_{i_1},C_{i_2}}$.

Definition 20 (Random Graph with Community Structure). *Let $P_{X|C_i,C_o}$ be a set of conditional distributions defined on $\mathcal{X} \times C \times C$, where $\mathcal{X} = [0, l-1]$ and C is defined in Definition 19. A random graph with community structure (RCS) g generated according to $P_{X|C_i,C_o}$ is a randomly generated $(n, c, (n_i)_{i\in[c]}, l)$-UCS with vertex set $\mathcal{V}$, community set C, and edge set $\mathcal{E}$, such that*

$$P((x, v_{j_1}, v_{j_2}) \in \mathcal{E}) = P_{X|C_i,C_o}(x|C_{j_1}, C_{j_2}), \forall x \in [0, l-1],$$

where $v_{j_1}, v_{j_2} \in C_{j_1} \times C_{j_2}$, and edges between different vertices are mutually independent.

Remark 11. *Note that for undirected graphs considered in this work, we must have $P_{X|C_i,C_o}(x|C_{j_1}, C_{j_2}) = P_{X|C_i,C_o}(x|C_{j_2}, C_{j_1})$.*

The following provides the notation used to represent the adjacency matrix of labeled graphs with community structure.

Definition 21 (Adjacency Matrix). *For an $(n, c, (n_i)_{i\in[c]}, l)$-UCS $g = (\mathcal{V}, C, \mathcal{E})$, a labeling is defined as a bijective function $\sigma : \mathcal{V} \to [1, n]$. The pair $\tilde{g} = (g, \sigma)$ is called an $(n, c, (n_i)_{i\in[c]}, l)$-labeled graph with community structure (LCS). For the labeled graph $\tilde{g}$ the adjacency matrix is defined as $G_\sigma = [G_{\sigma,i,j}]_{i,j\in[1,n]}$ where $G_{\sigma,i,j}$ is the unique value such that $(G_{\sigma,i,j}, v_i, v_j) \in \mathcal{E}_n$, where $(v_i, v_j) = (\sigma^{-1}(i), \sigma^{-1}(j))$. The submatrix $G_{\sigma,C_i,C_j} = [G_{\sigma,i,j}]_{i,j:v_i,v_j\in C_i\times C_j}$ is the adjacency matrix corresponding to the community pair C_i and C_j. The upper triangle (UT) corresponding to $\tilde{g}$ is the structure $U_\sigma = [G_{\sigma,i,j}]_{i<j}$. The upper triangle corresponding to communities C_i and C_j in $\tilde{g}$ is denoted by $U_{\sigma,C_i,C_j} = [G_{\sigma,i,j}]_{i<j:v_i,v_j\in C_i\times C_j}$. The subscript '$\sigma$' is dropped when there is no ambiguity.*

We consider pairs of correlated RCSs. It is assumed that edges between pairs of vertices in the two graphs with the same labeling are correlated and are generated based on a joint probability distribution, whereas edges between pairs of vertices with different labeling are generated independently. A pair of correlated RCSs is formally defined below.

Definition 22 (Correlated Pair of RCSs). *Let $P_{X,X'|C_{j_1},C_{j_2},C'_{j'_1},C'_{j'_2}}, j_1, j_2, j'_1, j'_2 \in [1, c]$ be a set of conditional distributions defined on $\mathcal{X}\times\mathcal{X}'\times C\times C\times C'\times C'$, where $\mathcal{X} = \mathcal{X}' = [0, l-1]$ and (C, C') are a pair of community sets of size $c \in \mathbb{N}$. A correlated pair of random graphs with community* *structure (CPCS) generated according to $P_{X,X'|C_{j_1},C_{j_2},C'_{j'_1},C'_{j'_2}}$ is a pair $\underline{\tilde{g}} = (\tilde{g},\tilde{g}')$ characterized by: i) the pair of RCSs (g,g') generated according to $P_{X|C_{j_1},C_{j_2}}$ and $P_{X'|C'_{j'_1},C'_{j'_2}}$, respectively, ii) the pair of labelings (σ,σ') for the graphs (g,g'), and iii) the probability distribution $P_{X,X'|C_{j_1},C_{j_2},C'_{j'_1},C'_{j'_2}}$, such that:*

1)The graphs have the same set of vertices $\mathcal{V}=\mathcal{V}'$.

2) For any two edges $e=(x,v_{j_1},v_{j_2}), e'=(x',v'_{j'_1},v'_{j'_2}), x,x'\in[0,l-1]$, we have

$$Pr\left(e\in\mathcal{E}, e'\in\mathcal{E}'\right)=\begin{cases}P_{X,X'}(x,x'), & \text{if } \sigma(v_{j_l})=\sigma'(v'_{j'_l})\\ Q_{X,X'}(x,x'), & \text{Otherwise}\end{cases},$$

where $l\in\{1,2\}$, $v_{j_1},v_{j_2}\in C_{j_1}\times C_{j_2}$, $v'_{j'_1},v'_{j'_2}\in {C'}_{j'_1}\times {C'}_{j'_2}$, the distribution $P_{X,X'}$ is the joint edge distribution when the edges connect vertices with similar labels and is given by $P_{X,X'|C_{j_1},C_{j_2},C'_{j'_1},C'_{j'_2}}$, the distribution $Q_{X,X'}$ is the conditional edge distribution when the edges connect labels with different labels and is given by $P_{X|C_{j_1},C_{j_2}}\times P_{X'|C'_{j'_1},C'_{j'_2}}$.

Remark 12. *In Definition 22, we have assumed that both graphs have the same number of vertices. In other words, the vertex set for both graphs is $\mathcal{V}=\mathcal{V}'=\{v_1,v_2,\cdots,v_n\}$. We further assume that the community memberships in both graphs are the same. In other words, we assume that $v_j\in C_i\Rightarrow v'_{j'}\in C'_i$ given that $\sigma(v_j)=\sigma'(v'_{j'})$ for any $j,j'\in[n]$ and $i\in[c]$. However, the results presented in this work can be extended to graphs with unequal but overlapping vertex sets and unequal community memberships in a straightforward manner.*

Remark 13. *We assume that the size of the communities in the graph sequence grows linearly in the number of vertices. More precisely, let $\Lambda^{(n)}(i)\triangleq|C_i^{(n)}|$ be the size of the i^{th} community, we assume that[3] $\Lambda^{(n)}(i)=\Theta(n)$ for all $i\in[c]$. Furthermore, we assume that the number of communities c is constant in n.*

We consider the matching strategies under two assumptions: i) with side-information, where the strategy uses prior knowledge of vertices' community memberships, ii) without side-information, where the strategy does not use prior knowledge of the vertices' community memberships, rather, it uses the statistics $P_{X,X'|C_i,C_o,C_{i'},C_{o'}}$ and the community sizes $(n_i)_{i\in[c]}$. The matching strategy is said to succeed if the fraction of vertices in the second graph which are labeled correctly approaches one as the number of vertices increases asymptotically.

[3] We write $f(x)=\Theta(g(x))$ if $\lim_{x\to\infty}\frac{f(x)}{g(x)}$ is a non-zero constant.

Definition 23 (Matching Strategy). *A matching strategy is defined under the following two scenarios:*

- ***With Side-information:*** *A matching strategy operating with complete side-information is a sequence of functions* $f_n^{CSI} : (\underline{g}^{(n)}, C^{(n)}, C'^{(n)}) \mapsto \hat{\sigma}'^{(n)}, n \in \mathbb{N}$*, where* $\underline{g}^{(n)} = (\tilde{g}_1^{(n)}, g_2^{(n)})$ *consists of a pair of graphs with CS with n vertices.*
- ***Without Side-information:*** *A matching strategy operating without side-information is a sequence of functions* $f_n^{WSI} : \underline{g}^{(n)} \mapsto \hat{\sigma}'^{(n)}, n \in \mathbb{N}$.

The output of a successful matching strategy satisfies $P\left(\sigma'^{(n)}(v'_{J^{(n)}}) = \hat{\sigma}'^{(n)}(v'_{J^{(n)}})\right) \to 1$ *as* $n \to \infty$*, where the random variable* $J^{(n)}$ *is uniformly distributed over* $[1, n]$ *and* $\sigma'^{(n)}$ *is the labeling for the graph* $g'^{(n)}$ *for which* $(\tilde{g}^{(n)}, \tilde{g}'^{(n)})$ *is a CPCS, where* $\tilde{g}'^{(n)} \triangleq (g'^{(n)}, \sigma'^{(n)})$.

Note that the output of a successful matching strategy $\hat{\sigma}'$ does not necessarily match every vertex correctly, i.e. does not satisfy $\hat{\sigma}' = \sigma'$. In other words, the pair $(\tilde{g}, \hat{g}')$ is not precisely a CPCS, where $\hat{g}' \triangleq (g', \hat{\sigma}')$. Rather, the fraction of mismatched vertices approaches zero as the size of the graph grows asymptotically large. This is in contrast with prior works [20], [26], [29] where a matching scheme is defined to be successful if it matches every vertex correctly, simultaneously, with probability approaching one as the graph grows asymptotically large. This relaxation of the success criteria is essential in application of concentration of measure theorems used in the next sections, and leads to significant simplification of our derivations. However, in Section IX, we show that the necessary and sufficient conditions on graph statistics which guarantee successful matching are equivalent for asymptotically large graphs under both success criteria. Consequently, the conditions derived in this work are applicable under the success criterion considered in prior works as well.

B. Matching in Presence of Side-information

First, we describe the matching strategy under the complete side-information scenario. In this scenario, the community membership of the nodes at both graphs are known prior to matching. Given a CPCS $\underline{\tilde{g}}$ generated according to $P_{X,X'|C_{j_1},C_{j_2},C'_{j'_1},C'_{j'_2}}$, $j_1, j_2, j'_1, j'_2 \in [1, c]$, the scheme operates as follows. It finds a labeling $\hat{\sigma}'$, for which i) the set of pairs $(G_{\sigma,C_{j_1},C_{j_2}}, G'_{\hat{\sigma}',C'_{j_1},C'_{j_2}})$, $j_1, j_2 \in [c]$ are jointly typical each with respect to $P_{X,X'|C_{j_1},C_{j_2},C'_{j_1},C'_{j_2}}(\cdot,\cdot|C_{j_1}, C_{j_2}, C'_{j_1}, C'_{j_2})$ when viewed as vectors of length $n_i n_j, i \neq j$, and ii) the set of pairs $(U_{\sigma,C_j,C_j}, U'_{\hat{\sigma}',C'_j,C'_j})$, $j \in [c]$ are jointly typical with respect to $P_{X,X'|C_{j_1},C_{j_2},C'_{j_1},C'_{j_2}}(\cdot,\cdot|C_j,C_j,C'_j,C'_j)$ when viewed as vectors of length $\frac{n_i(n_i-1)}{2}, j \in [c]$. Specifically, it returns a randomly picked element $\hat{\sigma}'$ from the set:

$$\widehat{\Sigma}_{C.C'} = \{\hat{\sigma}'|(U_{\sigma,C_j,C_j}, U'_{\hat{\sigma}',C'_j,C'_j}) \in \mathcal{A}_\epsilon^{\frac{n_j(n_j-1)}{2}}(P_{X,X'|C_j,C_j,C'_j,C'_j}), \forall j \in [c],$$

$$(G_{\sigma,C_i,C_j}, G'_{\hat{\sigma}',C'_i,C'_j}) \in \mathcal{A}_\epsilon^{n_i n_j}(P_{X,X'|C_i,C_j,C'_i,C'_j}), \forall i,j \in [c], i \neq j\},$$

where $\epsilon = \omega(\frac{1}{n})$, and declares $\hat{\sigma}'$ as the correct labeling. We show that under this scheme, the probability of incorrect labeling for any given vertex is arbitrarily small for large n.

Theorem 4. *For the typicality matching scheme, a given family of sets of distributions $\widetilde{P} = (\mathcal{P}^{(n)})_{n\in\mathbb{N}}$ is achievable, if for any constants $\delta > 0$, $\alpha \in [0, 1-\delta]$ and every sequence of distributions $P^{(n)}_{X,X'|C_{j_1},C_{j_2},C'_{j'_1},C'_{j'_2}} \in \mathcal{P}_n, j_1, j_2, j'_1, j'_2 \in [1,c]$, and community sizes $(n_1^{(n)}, n_2^{(n)}, \cdots, n_c^{(n)}), n \in \mathbb{N}$:*

$$4(1-\alpha)\frac{\log n}{n} \leq \min_{[\alpha_i]_{i\in[c]}\in\mathcal{A}_\alpha} \sum_{i,j\in[c],i<j} \frac{n_i^{(n)} n_j^{(n)}}{n^2} \cdot D(P^{(n)}_{X,X'|C_i,C_j}||(1-\beta_{i,j})P^{(n)}_{X|C_i,C_j}P^{(n)}_{X'|C_i,C_j} + \beta_{i,j}P^{(n)}_{X,X'|C_i,C_j})$$

$$+ \sum_{i\in[c]} \frac{n_i^{(n)}(n_i^{(n)}-1)}{2n^2} \cdot D(P^{(n)}_{X,X'|C_i,C_i}||(1-\beta_i)P^{(n)}_{X|C_i,C_i}P^{(n)}_{X'|C_i,C_i} + \beta_i P^{(n)}_{X,X'|C_i,C_i}), \tag{10}$$

as $n \to \infty$, where $\mathcal{A}_\alpha = \{([\alpha_i]_{i\in[c]}) : \alpha_i \leq \frac{n_i^{(n)}}{n}, \sum_{i\in[c]} \alpha_i = \alpha\}$, and $\beta_{i,j} = \frac{n^2}{n_i^{(n)} n_j^{(n)}}\alpha_i\alpha_j, i,j \in [c]$ and $\beta_i = \frac{n\alpha_i(n\alpha_i-1)}{n_i^{(n)}(n_i^{(n)}-1)}, i \in [c]$. The maximal family of sets of distributions which are achievable using the typicality matching scheme with complete side-information is denoted by $\mathcal{P}_{full}$.

Proof. Appendix I. □

Remark 14. *Note that the community sizes $(n_1^{(n)}, n_2^{(n)}, \cdots, n_c^{(n)}), n \in \mathbb{N}$ are assumed to grow in n such that* $\lim_{n\to\infty} \frac{n_i^n}{n} > 0$.

Theorem 3 recovers to the following achievable region for matching of pairs of Erdős-Rènyi graphs derived in Theorem 4.

C. Matching in Absence of Side-information

The scheme described in the previous section can be extended to matching graphs without community memberships side-information. In this scenario, it is assumed that the distribution $P_{X,X'|C_{j_1},C_{j_2},C'_{j'_1},C'_{j'_2}}, j_1, j_2, j'_1, j'_2 \in [1,c]$ is known, but the community memberships of the vertices in the graphs are not known. In this case, the scheme sweeps over all possible possible community membership assignments of the vertices in the two graphs. For each community membership assignment, the scheme attempts to match the two graphs using the method proposed in the complete side-information scenario. If it finds a labeling which satisfies the joint typicality conditions, it declares the labeling as the correct labeling. Otherwise, the scheme proceeds to the next community membership assignment. More precisely, for a given community assignment $(\hat{C}, \hat{C}')$, the scheme forms the following ambiguity set

$$\widehat{\Sigma}_{\hat{C},\hat{C}'} = \{\hat{\sigma}'|(U_{\sigma,\hat{C}_i,\hat{C}_i}, U'_{\hat{\sigma}',\hat{C}'_i,\hat{C}'_i}) \in \mathcal{A}_{\epsilon}^{\frac{n_i(n_i-1)}{2}}(P_{X,X'|\hat{C}_i,\hat{C}_i,\hat{C}'_i,\hat{C}'_i}), \forall i \in [c],$$

$$(G_{\sigma,\hat{C}_i,\hat{C}_j}, \widetilde{G'}_{\hat{\sigma}',\hat{C}'_i,\hat{C}'_j}) \in \mathcal{A}_{\epsilon}^{n_i n_j}(P_{X,X'|\hat{C}_i,\hat{C}_j,\hat{C}'_i,\hat{C}'_j}), \forall i, j \in [c], i \neq j\}.$$

Define $\widehat{\Sigma}_0$ as follows:

$$\widehat{\Sigma}_0 = \cup_{(\hat{C},\hat{C}')\in \mathsf{C}} \widehat{\Sigma}_{\hat{C},\hat{C}'}.$$

where C is the set of all possible community membership assignments. The scheme outputs a randomly and uniformly chosen element of $\widehat{\Sigma}_0$ as the correct labeling. The following theorem shows that the achievable region for this scheme is the same as the one described in Theorem 4.

Theorem 5. *Let $\mathcal{P}_0$ be the maximal family of sets of achievable distributions for the typicality matching scheme without side-information. Then, $\mathcal{P}_0 = \mathcal{P}_{full}$.*

The proof follows similar arguments as that of Theorem 4. We provide an outline. It is enough to show that $|\widehat{\Sigma}_0|$ has the same exponent as that of $|\widehat{\Sigma}_{C.C'}|$. To see this note that the size of the set of all community membership assignments C has an exponent which is $\Theta(n)$:

$$|\mathsf{C}| \leq 2^{cn}.$$

On the other hand,

$$|\widehat{\Sigma}_0| \leq |\mathsf{C}| \cdot |\widehat{\Sigma}_{C.C'}| \leq 2^{nc} \cdot 2^{\Theta(n \log n)} = 2^{\Theta(n \log n)}.$$

The rest of the proof follows by the same arguments as in Theorem 4.

VII. Matching Collections of Graphs

In the previous sections, we considered matching of pairs of correlated graphs. The results can be further extended to problems involving matching of collections of more than two graphs.

In this section, we consider matching collections of more than two correlated graphs, where the first graph is deanonymized and the other graphs are anonymized. For brevity we consider collections of correlated Erdös-Rényi graphs, i.e. single-community random graphs. The results can be further exteneded to correlated graphs with community structure in a straightforward manner. The following formally describes a collection of correlated Erdös-Rényi graphs.

Definition 24 (Correlated Collection of ER Graphs). *Let P_{X^m} be a conditional distribution defined on $\prod_{k\in[m]}\mathcal{X}_i$, where $\mathcal{X}_i=[0,l-1], i\in[m]$ and $m>2$. A correlated collection of ER graphs $\underline{\tilde{g}}=(\tilde{g}^i)_{i\in[m]}$ generated according to P_{X^m} is characterized by: i) the collection of ER graphs $(g^i)_{i\in[m]}$ each generated according to P_{X_i}, ii) the collection of labelings $(\sigma_i)_{i\in[m]}$ for the unlabeled graphs $(g^i)_{i\in[m]}$, and iii) the joint probability distribution P_{X^m}, such that:*
1)The graphs have the same set of vertices $\mathcal{V}=\mathcal{V}_i, i\in[m]$.
2) For any collection of edges $e^i=(x^i,v_{j^i_1},v_{j^i_2}), x^i\in[0,l-1], i\in[m]$, we have

$$Pr\left(e^i\in\mathcal{E}^i, i\in[m]\right)=\begin{cases}P_{X^m}(x^m), & \text{if } \sigma^i(v_{j^i_l})=\sigma^k(v_{j^k_l}), \forall i,k\in[m]\\ \prod_{i\in[m]}P_{X_i}(x_i), & \text{Otherwise}\end{cases},$$

where $l\in\{1,2\}$, and $v_{j^i_1},v_{j^i_2}\in\mathcal{V}_1\times\mathcal{V}_2, i\in[m]$.

Similar to the Typicality Matching Strategy for pairs of correlated graphs described in Section V, we propose a matching strategy based on typicality for collections of correlated graphs. Given a correlated collection of graphs $(g^i)_{i\in[m]}$, where the labeling for $\tilde{g}^1$ is given and the rest of the graphs are ananymized, the typicality matching strategy operates as follows. The scheme finds a collection $\widehat{\Sigma}$ of labelings $\hat{\sigma}^j, j\in[2,m]$, for which the UT's $U^j_{\sigma^j}, j\in[m]$ are jointly typical with respect to P_{n,X^m} when viewed as vectors of length $\frac{n(n-1)}{2}$. The strategy succeeds if at least one such labeling exists and fails otherwise.

Theorem 6. *For the typicality matching strategy, a given family of sets of distributions $\widetilde{P}=(\mathcal{P}_n)_{n\in\mathbb{N}}$ is achievable, if for every sequence of distributions $P_{n,X^m}\in\mathcal{P}_n, n\in\mathbb{N}$ we have*

$$\frac{\log n}{n}\Big(\sum_{k\in[b_m]}|\mathcal{P}_k|\alpha_k-1\Big)\leq\frac{1}{2b_m m(m-1)}D(P_{X^m}\|\sum_{k\in[b_m]}\alpha'_k P_{X_{\mathcal{P}_k}})+O\Big(\frac{\log n}{n}\Big),\tag{11}$$

for all $\alpha_1,\alpha_2,\cdots,\alpha_{b_m}:\sum_{k\in[b_m]}\alpha_k=n, \alpha_{b_m}\in[1,1-\alpha_n]$, where $\alpha'_k=\frac{\alpha_k^2}{2}+\sum_{k',k'':\mathcal{P}_{k',k''}=\mathcal{P}_l}\alpha_{k'}\alpha_{k''}$, $\mathcal{P}_{k',k''}=\{\mathcal{A}'\cap\mathcal{A}'':\mathcal{A}'\in\mathcal{P}_{k'},\mathcal{A}''\in\mathcal{P}_{k''}\}, k',k''\in[b_m]$, and $\mathcal{P}_{b_m}=[1,n]$ is the single-element partition.

Proof. Appendix J. □

Remark 15. *Note that Equation* (11) *recovers the result given in Equation* (9) *for matching of pairs of correlated ER graphs, i.e. $m = 2$.*

VIII. Converse Results

In this section, we provide conditions on the graph parameters under which graph matching is not possible. Without loss of generality, we assume that (σ, σ') are a pair of random labelings chosen uniformly among the set of all possible labeling for the two graphs. Roughly speaking, the information revealed by identifying the realization of σ' is equal to $H(\sigma') \approx n \log n$. Consequently, using Fano's inequality, we show that the information contained in (σ, G, G') regarding σ', which is quantified as the mutual information $I(\sigma'; \sigma, G, G')$, must be at least $n \log n$ bits for successful matching. The mutual information $I(\sigma'; \sigma, G, G')$ is a function of multi-letter probability distributions. We use standard information theoretic techniques to bound $I(\sigma'; \sigma, G, G')$ using information quantities which are functionals of single-letter distributions. The following states the resulting necessary conditions for successful matching.

Theorem 7. *For the graph matching problem under the community structure model with complete side-information, the following provides necessary conditions for successful matching:*

$$\frac{\log n}{n} \leq \sum_{i,j \in [c], i<j} \frac{n_i n_j}{n^2} I(X, X' | C_i, C_j, C'_i C'_j) + \sum_{i \in [c]} \frac{n_i(n_i - 1)}{2n^2} I(X, X' | C_i, C_i, C'_i, C'_i) + O(\frac{\log n}{n}),$$

where $I(X, X'|C_i, C_j, C'_i C'_j)$ is defined with respect to $P_{X,X'|C_i,C_j,C'_iC'_j}$.

Proof. Appendix K. □

For Erdős-Rènyi graphs, the following corollary is a direct consequence of Theorem 7.

Corollary 1. *For the graph matching problem under the Erdős-Rènyi model, the following provides necessary conditions for successful matching:*

$$\frac{2 \log n}{n} \leq I(X, X') + O(\frac{\log n}{n}).$$

IX. Criteria for Successful Matching

In Section III, it was pointed out that the criterion for successful matching defined in Definition 23 requires the fraction of correctly matched vertices to approach one as the size of the graph grows asymptotically large. This is a relaxation of the criterion considered in prior works [20], [26], [29] where a matching scheme is said to be successful if it matches every vertex correctly simultaneously with probability approaching one as the size of the graph grows asymptotically large. In this section, we show that for pairs of Erdös-Rènyi graphs with binary-valued edges, the necessary and sufficient conditions on the edge statistics which guarantee successful matching are equivalent under the two success criteria. The results can be potentially extended to graphs with community structure, non binary-valued edges, and collections of graphs considered in the previous sections. Consequently, the conditions derived in this work are applicable under the scenarios considered in the prior works as well. We use the following proposition to prove the equivalency of the success criteria.

Proposition 2. *Let $\underline{\tilde{g}}$ be a CPER distributed according to $P_{X,X'}$ and let G and G' be the corresponding adjacency matrices. Let $(S^n, S^{'n})$ be the first row in G and G', respectively. Let $S^{''n}$ be the second row in G. Assume that the triple of vectors $(T^n, T^{'n}, T^{''n})$ are such that they are jointly equal to $(S^n, S^{'n}, S^{''n})$ in $(1-\alpha)$ fraction of their elements, where $\alpha \in [0,1]$. Alternatively,*

$$1-\alpha = \frac{1}{n}\sum_{i\in[1,n]} \mathbb{1}((S_i, S'_i, S''_i) = (T_i, T'_i, T''_i)).$$

Then, there exist a binary hypothesis test $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ and function $\zeta : [0,1] \to [0,\infty)$, such that:

$$P(f(T^n, T^{'n}) = 1) \geq \beta_n,$$

$$P(f(T^{''n}, T^{'n}) = 1) \leq 2^{-n\zeta(\alpha)I(X;X')},$$

where $\zeta(\alpha) \to 1$ as $\alpha \to 0$, and $\beta_n \to 1$ as $n \to \infty$.

X. Seeded Graph Matching

So far, we have investigated the fundamental limits of graph matching assuming the availability of unlimited computational resources. In this section, we consider seeded graph matching, and propose a matching algorithm whose complexity grows polynomially in the number of vertices of the graph and leads to successful matching in a wide range of graph matching scenarios. The algorithm leverages ideas from prior work a related problem called *online fingerprinting* which involves matching of correlated bipartite graphs [34].

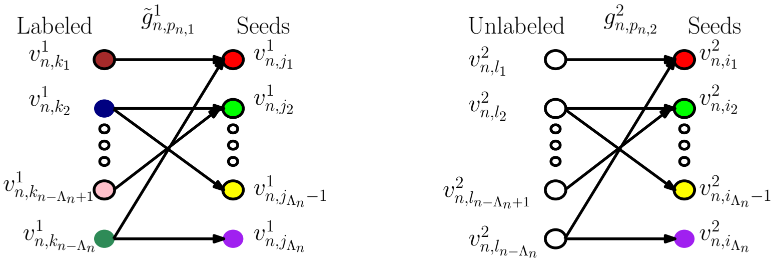

Fig. 2. The matching algorithm constructs the bipartite graph which captures the connections between the unmatched vertices with the seed vertices.

In seeded graph matching, it is assumed that we are given the correct labeling for a subset of the vertices in the anonymized graph prior to the start of the matching process. The subset of pre-matched vertices are called '*seeds*'. The motivation behind the problem formulation is that in many applications of graph matching, the correct labeling of a subset of vertices is known through side-inforamtion. For instance, in social network deanonymization, many users link their social media accounts across social networks publicly. As shown in this section, the seed side-information can be used to significantly reduce the complexity of the matching algorithm.

The proposed graph matching algorithm operates as follows. First, the algorithm constructs the bipartite graph shown in Figure 2 whose edges consist of the connections between the unmatched vertices with the seeded vertices in each graph. The algorithm proceeds in two steps. First, it constructs the '*fingerprint*' vectors for each of the unmatched vertices in the two bipartite graphs based on their connections to the seed vertices. The fingerprint vector of a vertex is the row in the adjacency matrix of the bipartite graph corresponding to the edges between that vertex and the seed vertices. In the second step, the algorithm finds a jointly typical pair of fingerprint vectors in the deanonymized and deanonymized graph adjacency matrices and matches the corresponding vertices, where typicality is defined based on the joint distribution between the edges of the two graphs. Note that the bipartite graphs encompass only a subset of the edges in the original graphs. Hence by restricting the matching process to the bipartite graphs, some of the information which could potentially help in matching is ignored. This leads to more restrictive conditions on successful matching compared to the ones derived in the previous sections. However, the computational complexity of the resulting matching algorithm is considerably improved. In the following, we focus on matching of seeded CPERs. The results can be easily extended to seeded CPCSs similar to the unseeded graph matching in prior sections. A seeded CPER (SCPER) is formally defined below.

Definition 25 (Correlated Pair of Seeded ER Graphs). *An SPCER is a triple* $(\tilde{g}, \tilde{g}', \mathcal{S})$*, where* $\underline{\tilde{g}} = (\tilde{g}, \tilde{g}')$ *is a CPER generated according to* $P_{X,X'}$*, and* $\mathcal{S} \subseteq \mathcal{V}$ *is the seed set.*

Let $\mathcal{S} = \{v_{i_1}, v_{i_2}, \cdots, v_{i_\Lambda}\}$ and define the reverse seed set $\mathcal{S}^{-1} = \{v_{j_1}, v_{j_2}, \cdots, v_{j_\Lambda}\}$, where $\sigma(v_{j_k}) = \sigma'(v_{i_k}), k \in [1, \Lambda]$. The algorithm is given the correct labeling of all the vertices in the first graph $\sigma : \mathcal{V} \to [1, n]$ and the seed vertices in the second graph $\sigma'|_{\mathcal{S}} : \mathcal{S} \to [1, n]$. The objective is to find the correct labeling of the rest of the vertices in the second graph $\hat{\sigma}_n : \mathcal{V}; \to [1, n]$ so that the fraction of mislabeled vertices is negligible as the number of vertices grows asymptotically large, i.e. $P(\hat{\sigma}' = \sigma') \to 1$ as $n \to \infty$. To this end, the algorithm first constructs a fingerprint for each vertex in each of the graphs. For an arbitrary vertex v_i in g_{P_X}, its fingerprint is defined as the vector $\underline{F}_i = (F_i(1), F_i(2), \cdots, F_i(\Lambda))$. which indicates its connections to the reverse seed elements:

$$F_i(l) = \begin{cases} 1 \quad \text{if} \qquad (v_i, v_{j_l}) \in \mathcal{E} \\ 0 \quad \text{Otherwise} \end{cases}, \qquad l \in [1, \Lambda].$$

The fingerprint of a vertex v_i in the second graph is defined in a similar fashion based on connections to the elements of the seed set $\mathcal{S}$. Take an unmatched vertex $v_i \notin \mathcal{S}$. The algorithm matches v_i in g to a vertex v_j in g' if it is the unique vertex such that the fingerprint pair $(\underline{F}_i, \underline{F}'_j)$ are jointly ϵ-typical with respect to the distribution $P_{X,X'}$, where[4] $\epsilon = \omega(\frac{1}{\sqrt{\Lambda}})$:

$$\exists! i : (\underline{F}_i, \underline{F}'_j) \in \mathcal{A}_\epsilon^n(X, X') \Rightarrow \hat{\sigma}(v_i) = \sigma'(v_j),$$

where $\mathcal{A}_\epsilon^n(X, X')$ is the set of jointly ϵ-typical set sequences of length n with respect to $P_{X,X'}$. If a unique match is not found, then vertex v_i is added to the ambiguity set $\mathcal{L}$. Hence, $\mathcal{V} \backslash \mathcal{L}$ is the set of all matched vertices. In the next step, these vertices are added to the seed set and the expanded seed set is used to match the vertices in the ambiguity set. The algorithm succeeds if all vertices are matched at this step and fails otherwise. We call this strategy the Seeded Typicality Matching Strategy (STMS).

Theorem 8. *Define the family of sets of pairs of distribution and seed sizes* $\widetilde{\mathcal{P}}$ *as follows:*

$$\widetilde{\mathcal{P}} = \Big\{(\mathcal{P}_n, \Lambda_n)_{n\in\mathbb{N}} \Big| \forall P_{n,X,X'} \in \mathcal{P}_n : \frac{2\log n}{I(X, X')} \leq \Lambda_n, I(X; X') = \omega\left(\sqrt{\frac{1}{\Lambda_n}}\right)\Big\}.$$

[4] Alternatively, $\lim_{n\to\infty} \frac{\epsilon}{\sqrt{|S|}} = \infty$.

Any family of SCPERs with parameters chosen from $\widetilde{\mathcal{P}}$ is matchable using the STMS.

The proof of Theorem 8 which is provided in Appendix M uses the following lemma on the cardinality of $\mathcal{L}$.

Lemma 7. *The following holds:*

$$P(|\mathcal{L}| > \frac{2n}{\Lambda\epsilon^2}) \to 0, \text{ as } n \to \infty,$$

Proof. Appendix L. □

XI. Conclusion

We have considered matching of collections of correlated graphs. We have studied the problem under the Erdös-Rènyi model as well as the more general community structure model. The derivations apply to graphs whose edges may take non-binary attributes. We have introduced a graph matching scheme called the Typicality Matching scheme which relies on tools such as concentration of measure and typicality of sequences of random variables to perform graph matching. We further provide converse results which lead to necessary conditions on graph parameters for successful matching. We have investigated seeded graph matching, where the correct labeling of a subset of graph vertices is known prior to the matching process. We have introduced a matching algorithm for seeded graph matching which successfully matches the graphs in wide range of matching problems with large enough seeds and has a computational complexity which grows polynomially in the number of graph vertices.

Appendix A
Proof of Theorem 1

Define the following partition for the set of indices $[1,n]$:

$$\mathcal{A}_0 = \{1, i_1+1, i_1+i_2+1, \cdots, \sum_{j=1}^{r-1} i_j + 1\}, \quad \mathcal{A}_1 = \{k|\text{k is even}, \ \& \ k \notin \mathcal{A}_0, \ \& \ k \leq \sum_{i=1}^{r} i_j\},$$

$$\mathcal{A}_2 = \{k|\text{k is odd}, \ \& \ k \notin \mathcal{A}_0, \ \& \ k \leq \sum_{i=1}^{r} i_j\}, \quad \mathcal{A}_3 = \{k|k > \sum_{i=1}^{r} i_j\}.$$

The set $\mathcal{A}_1$ is the set of indices at the start of each cycle in π, the sets $\mathcal{A}_2$ and $\mathcal{A}_3$ are the sets of odd and even indices which are not start of any cycles and $\mathcal{A}_4$ is the set of fixed points of π. Let $Z^n = \pi(Y^n)$. It is straightforward to verify that $(X_i, Z_i), i \in \mathcal{A}_j, j \in [3]$ are three sequences of independent and identically distributed variables which are distributed according to $P_X P_Y$. The reason is that the standard permutation shifts elements of a sequence by at most one position, whereas the elements in the sequences $(X_i, Z_i), i \in \mathcal{A}_j, j \in [3]$ are at least two indices apart and are hence independent of each other (i.e. $Z_i \neq Y_i$). Furthermore, $(X_i, Z_i), i \in \mathcal{A}_4$ is a sequence of independent and identically distributed variables which are distributed according to $P_{X,Y}$ since $Z_i = Y_i$. Let $\underline{T}_j, j \in [4]$ be the type of the sequence $(X_i, Z_i), i \in \mathcal{A}_j, j \in [4]$, so that $T_{j,x,y} = \frac{\sum_{i \in \mathcal{A}_j} \mathbb{1}(X_i = x, Z_i = y)}{|\mathcal{A}_j|}, j, x, y \in [4] \times \mathcal{X} \times \mathcal{Y}$. We are interested in the probability of the event $(X^n, Z^n) \in \mathcal{A}_\epsilon^n(X, Y)$. From Definition 8 this event can be rewritten as follows:

$$
\begin{aligned}
&P\left((X^n, Z^n) \in \mathcal{A}_\epsilon^n(X, Y)\right) = P\left(\underline{T}(X^n, Y^n) \doteq P_{X,Y}(\cdot, \cdot) \pm \epsilon\right) \\
&= P(\alpha_1 \underline{T}_1 + \alpha_2 \underline{T}_2 + \alpha_3 \underline{T}_3 + \alpha_4 \underline{T}_4 \doteq P_{X,Y}(\cdot, \cdot) \pm \epsilon),
\end{aligned}
$$

where $\alpha_i = \frac{|\mathcal{A}_i|}{n}, i \in [4]$, we write $a \doteq x \pm \epsilon$ to denote $x - \epsilon \leq a \leq x_\epsilon$, and addition is defined element-wise. We have:

$$
P((X^n, Z^n) \in \mathcal{A}_\epsilon^n(X, Y)) = \sum_{(\underline{t}_1, \underline{t}_2, \underline{t}_3, \underline{t}_4) \in \mathcal{T}} P(\underline{T}_i = \underline{t}_i, i \in [4]),
$$

where $\mathcal{T} = \{(\underline{t}_1, \underline{t}_2, \underline{t}_3, \underline{t}_4) : \alpha_1 \underline{t}_1 + \alpha_2 \underline{t}_2 + \alpha_3 \underline{t}_3 + \alpha_4 \underline{t}_4 \doteq n(P_{X,Y}(\cdot, \cdot) \pm \epsilon)\}$. Using the property that for any set of events, the probability of the intersection is less than or equal to the geometric average of the individual probabilities, we have:

$$
P((X^n, Z^n) \in \mathcal{A}_\epsilon^n(X, Y)) \leq \sum_{(\underline{t}_1, \underline{t}_2, \underline{t}_3, \underline{t}_4) \in \mathcal{T}} \sqrt[4]{\Pi_{i \in [4]} P(\underline{T}_i = \underline{t}_i)}.
$$

Since the elements $(X_i, Z_i), i \in \mathcal{A}_j, j \in [4]$ are i.i.d, it follows from standard information theoretic arguments [32] that:

$$
P(\underline{T}_i = \underline{t}_i) \leq 2^{-|\mathcal{A}_i|(D(\underline{t}_i \| P_X P_Y) - |\mathcal{X}||\mathcal{Y}|\epsilon)}, i \in [3], \quad P(\underline{T}_4 = \underline{t}_4) \leq 2^{-|\mathcal{A}_4|(D(\underline{t}_4 \| P_{X,Y}) - |\mathcal{X}||\mathcal{Y}|\epsilon)}.
$$

We have,

$$
\begin{aligned}
&P((X^n, Z^n) \in \mathcal{A}_\epsilon^n(X, Y)) \\
&\leq \sum_{(\underline{t}_1, \underline{t}_2, \underline{t}_3, \underline{t}_4) \in \mathcal{T}} \sqrt[4]{2^{-n(\alpha_1 D(\underline{t}_1 \| P_X P_Y) + \alpha_2 D(\underline{t}_2 \| P_X P_Y) + \alpha_3 D(\underline{t}_3 \| P_X P_Y) + \alpha_4 D(\underline{t}_4 \| P_{X,Y}) - |\mathcal{X}||\mathcal{Y}|\epsilon)}}
\end{aligned}
$$

$$\overset{(a)}{\leq} \sum_{(\underline{t}_1,\underline{t}_2,\underline{t}_3,\underline{t}_4)\in\mathcal{T}} \sqrt[4]{2^{-n(D(\alpha_1\underline{t}_1+\alpha_2\underline{t}_2+\alpha_3\underline{t}_3+\alpha_4\underline{t}_4||(\alpha_1+\alpha_2+\alpha_3)P_XP_Y+\alpha_4P_{X,Y})-|\mathcal{X}||\mathcal{Y}|\epsilon)}}$$

$$= |\mathcal{T}|\sqrt[4]{2^{-n(D(P_{X,Y}||(1-\alpha)P_XP_Y+\alpha P_{X,Y})-|\mathcal{X}||\mathcal{Y}|\epsilon)}}$$

$$\overset{(b)}{\leq} 2^{-\frac{n}{4}(D(P_{X,Y}||(1-\alpha)P_XP_Y+\alpha P_{X,Y})-|\mathcal{X}||\mathcal{Y}|\epsilon+O(\frac{\log n}{n}))},$$

where the (a) follows from the convexity of the divergence function and (b) follows by the fact that the number of joint types grows polynomially in n. □

Appendix B
Proof of Lemma 2

First, we prove Equation (2). Note that

$$N_m = \binom{n}{m}!(n-m) \leq \binom{n}{m}(n-m)! = \frac{n!}{m!} \leq n^{n-m}.$$

This proves the right hand side of the equation. To prove the left hand side, we first argue that the iterative inequality $!n \geq !(n-1)(n-1)$ holds. In other words, the number of derangements of numbers in the interval $[n]$ is at least $n-1$ times the number of derangements of the numbers in the interval $[n-1]$. We prove the statement by constructing $!(n-1)(n-1)$ distinct derangements of the numbers $[n]$. Note that a derangement $\pi(\cdot)$ of $[n]$ is characterized by the vector $(\pi(1), \pi(2), \cdots (n))$. There are a total of $n-1$ choices for $\pi(1)$ (every integer in $[n]$ except for 1). Once $\pi(1)$ is fixed, the rest of the vector $(\pi(2), \pi(3), \cdots, \pi(n))$ can be constructed using any derangement of the set of numbers $[n]-\{\pi(1)\}$. There are a total of $!(n-1)$ such derangements. So, we have constructed $(n-1)!(n-1)$ distinct derangements of $[n]$. Consequently. $!n \geq !(n-1)(n-1)$. By induction, we have $!n \geq (n-1)!$. So,

$$N_m = \binom{n}{m}!(n-m) \geq \binom{n}{m}(n-m-1)! = \frac{n!}{m!(n-m)}.$$

Next, we prove that Equation (3) holds. Note that from the right hand side of Equaation (2) we have:

$$\lim_{n\to\infty} \frac{\log N_m}{n\log n} \leq \lim_{n\to\infty} \frac{\log n^{n-m}}{n\log n} = \lim_{n\to\infty} \frac{n-m}{n} = 1-\alpha.$$

Also, from the left hand side of Equation (3), we have:

$$\lim_{n\to\infty} \frac{\log N_m}{n\log n} \geq \lim_{n\to\infty} \frac{\log \frac{n!}{m!(n-m)}}{n\log n} = \lim_{n\to\infty} \frac{\log \frac{n!}{m!}}{n\log n} - \frac{\log(n-m)}{n\log n}.$$

The second term in the last inequality converges to 0 as $n \to \infty$. Hence,

$$\lim_{n\to\infty} \frac{\log N_m}{n\log n} \geq \lim_{n\to\infty} \frac{\log \frac{n!}{m!}}{n\log n}$$
$$\overset{(a)}{\geq} \lim_{n\to\infty} \frac{\log \frac{n!}{m^m}}{n\log n} \geq \lim_{n\to\infty} \frac{\log n!}{n\log n} - \frac{\log m^m}{n\log n} \overset{(b)}{\geq} \lim_{n\to\infty} \frac{n\log n - n + O(\log n)}{n\log n} - \frac{\log m^m}{n\log n}$$
$$= \lim_{n\to\infty} \frac{n\log n}{n\log n} - \frac{\alpha n\log \alpha n}{n\log n} = 1 - \alpha,$$

where in (a) we have used the fact that $m! \leq m^m$, and (b) follows from Stirling's approximation. This completes the proof. □

Appendix C

Proof of Lemma 3

The proof builds upon some of the techniques developed in [35]. Let $\mathcal{A} = \{(x,y) \in \mathcal{X} \times \mathcal{Y} \big| P_X P_Y(x,y) < P_{X,Y}(x,y)\}$. Let $Z^{\{(x,y)\}}_{(\pi),i} = \mathbb{1}(X_i, Y_{\pi(i)} = (x,y))$. We have:

$$P((X^n, \pi(Y^n)) \in \mathcal{A}^n_\epsilon(X,Y)) \leq$$
$$P\Big(\Big(\bigcap_{(x,y)\in\mathcal{A}} \{\frac{1}{n}\sum_{i=1}^{n} Z^{\{(x,y)\}}_{(\pi),i} > P_{X,Y}(x,y) - \epsilon\}\Big)\bigcap\Big(\bigcap_{(x,y)\in\mathcal{A}^c} \{\frac{1}{n}\sum_{i=1}^{n} Z^{\{(x,y)\}}_{(\pi),i} < P_{X,Y}(x,y) + \epsilon\}\Big)\Big)$$

For brevity let $\alpha_{x,y} = \frac{1}{n}\sum_{i=1}^{n} Z^{\{(x,y)\}}_{(\pi),i}$, and $t_{x,y} = \frac{1}{2}\log_e \frac{P_{X,Y}(x,y)}{P_X(x)P_Y(y)}, x, y \in \mathcal{X}$. Then,

$$Pr\Big(\Big(\bigcap_{(x,y)\in\mathcal{A}} \{n\alpha_{x,y} > nP_{X,Y}(x,y) - n\epsilon\}\Big)\bigcap\Big(\bigcap_{(x,y)\in\mathcal{A}^c} \{n\alpha_{x,y} < nP_{X,Y}(x,y) + n\epsilon\}\Big)\Big)$$
$$= Pr\Big(\bigcap_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \{e^{nt_{x,y}\alpha_{x,y}} > e^{nt_{x,y}P_{X,Y}(x,y)+n\epsilon_{x,y}}\}\Big),$$

where $\epsilon_{x,y} = t_{x,y}(1 - 2\mathbb{1}(x,y \in \mathcal{A}))\epsilon$ and we have used the fact that by construction:

$$\begin{cases} t_{x,y} > 0 \quad \text{if} \qquad (x,y) \in \mathcal{A} \\ t_{x,y} < 0 \quad \text{if} \qquad (x,y) \in \mathcal{A}^c. \end{cases} \tag{12}$$

So,

$$P\Big(\Big(\bigcap_{(x,y)\in\mathcal{A}} \{n\alpha_{x,y} > nP_{X,Y}(x,y) - n\epsilon\}\Big)\bigcap\Big(\bigcap_{(x,y)\in\mathcal{A}^c} \{n\alpha_{x,y} < nP_{X,Y}(x,y) + n\epsilon\}\Big)\Big)$$
$$\overset{(a)}{\leq} P\Big(\prod_{(x,y)\in\mathcal{X}\times\mathcal{Y}} e^{nt_{x,y}\alpha_{x,y}} > \prod_{(x,y)\in\mathcal{X}\times\mathcal{Y}} e^{nt_{x,y}P_{X,Y}(x,y)-n\epsilon_{x,y}}\Big) \tag{13}$$
$$\overset{(b)}{\leq} e^{-\sum_{x,y} n(t_{x,y}P_{X,Y}(x,y)-\epsilon_{x,y})}\mathbb{E}(\prod_{x,y} e^{nt_{x,y}\alpha_{x,y}}) = e^{-\sum_{x,y} n(t_{x,y}P_{X,Y}(x,y)-\epsilon_{x,y})}\mathbb{E}(e^{\sum_{i=1}^{n}\sum_{x,y} t_{x,y}Z^{\{(x,y)\}}_{(\pi),i}}) \tag{14}$$

$$\overset{(c)}{\leq} e^{-\sum_{x,y} n(t_{x,y}P_{X,Y}(x,y)-\epsilon_{x,y})}\mathbb{E}^{\frac{1}{2}}(e^{\sum_{i\in\mathcal{O}}\sum_{x,y} 2t_{x,y}Z^{\{(x,y)\}}_{(\pi),i}})\mathbb{E}^{\frac{1}{2}}(e^{\sum_{i\in\mathcal{E}}\sum_{x,y} 2t_{x,y}Z^{\{(x,y)\}}_{(\pi),i}}) \tag{15}$$

$$= e^{-\sum_{x,y} n(t_{x,y}P_{X,Y}(x,y)-\epsilon_{x,y})}\prod_{i\in\mathcal{O}}\mathbb{E}^{\frac{1}{2}}(e^{\sum_{x,y} 2t_{x,y}Z^{\{(x,y)\}}_{(\pi),i}})\prod_{i\in\mathcal{E}}\mathbb{E}^{\frac{1}{2}}(e^{\sum_{x,y} 2t_{x,y}Z^{\{(x,y)\}}_{(\pi),i}}), \tag{16}$$

where $\mathcal{O}$ and $\mathcal{E}$ are the odd and even indices in the set $[1,n]$. In (a) we have used the fact that the exponential function is increasing and positive, (b) follows from the Markov inequality and (c) follows from the Cauchy-Schwarz inequality. Note that:

$$\mathbb{E}(e^{\sum_{x,y} 2t_{x,y}Z^{\{(x,y)\}}_{(\pi),i}}) \overset{(a)}{=} \sum_{x,y} P_X(x)P_Y(y)e^{2t_{x,y}} = \sum_{x,y} P_X(x)P_Y(y)e^{\log_e \frac{P_{X,Y}(x,y)}{P_X(x)P_Y(y)}} = \sum_{x,y} P_{X,Y}(x,y) = 1,$$

where in (a) we have used the fact that X_i and $Y_{\pi(i)}$ are independent since the permutation does not have any fixed points. Consequently, we have shown that:

$$Pr((X^n,\pi(Y^n))\in\mathcal{A}^n_\epsilon(X,Y)) \leq e^{-\sum_{x,y} n(t_{x,y}P_{X,Y}(x,y)-\epsilon_{x,y})} = e^{-\sum_{x,y} n(\frac{1}{2}P_{X,Y}(x,y)\log_e\frac{P_{X,Y}(x,y)}{P_X(x)P_Y(y)}-\epsilon_{x,y})} = 2^{-\frac{1}{2}n(I(X;Y)-\delta)}.$$

This completes the proof.

□

Appendix D
Proof of Lemma 4

The proof follows by similar arguments as that of Lemma 3. Following similar steps, we have

$$P((X^n,\pi(Y^n))\in\mathcal{A}^n_\epsilon(X,Y)) = Pr\Big(\bigcap_{(x,y)\in\mathcal{X}\times\mathcal{Y}}\{e^{\frac{n}{s}t_{x,y}\alpha_{x,y}} > e^{\frac{n}{s}t_{x,y}P_{X,Y}(x,y)+n\epsilon_{x,y}}\}\Big)$$

$$\leq P\Big(\prod_{(x,y)\in\mathcal{X}\times\mathcal{Y}} e^{\frac{n}{s}t_{x,y}\alpha_{x,y}} > \prod_{(x,y)\in\mathcal{X}\times\mathcal{Y}} e^{\frac{n}{s}t_{x,y}P_{X,Y}(x,y)-\frac{n}{s}\epsilon_{x,y}}\Big)$$

$$\leq e^{-\sum_{x,y}\frac{n}{s}(t_{x,y}P_{X,Y}(x,y)-\epsilon_{x,y})}\mathbb{E}(\prod_{x,y} e^{\frac{n}{s}t_{x,y}\alpha_{x,y}})$$

$$= e^{-\sum_{x,y}\frac{n}{s}(t_{x,y}P_{X,Y}(x,y)-\epsilon_{x,y})}\prod_{j\in[1,c]}\mathbb{E}(e^{\frac{1}{s}\sum_{x,y}\sum_{k=1}^{i_j} t_{x,y}Z^{\{(x,y)\}}_{(\pi),i}}). \tag{17}$$

We need to investigate $\mathbb{E}(e^{\frac{1}{s}\sum_{x,y}\sum_{k=1}^{i_j} t_{x,y}Z^{\{(x,y)\}}_{(\pi),i}})$. Define $T^{(x,y)}_j = \sum_{k=1}^{i_j} Z^{\{(x,y)\}}_{(\pi),i}, j\in[1,c], x,y\in\mathcal{X}\times\mathcal{Y}$ as the number of occurrences of the pair (x,y) in the jth cycle. Note that by definition, we have $\sum_{x,y}\sum_{k=1}^{i_j} Z^{\{(x,y)\}}_{(\pi),i} = \sum_{x,y} T^{(x,y)}_j = i_j$. Define $S^{(x,y)}_j = \frac{1}{s}T^{(x,y)}_j, j\in[1,c], x,y\in\mathcal{X}\times\mathcal{Y}$. Let $\mathcal{B} = \{(s^{(x,y)}_j)_{j\in[1,c],x,y\in\mathcal{X}\times\mathcal{Y}} : \sum_{x,y} s^{(x,y)}_j = \frac{i_j}{s}, j\in[1,c]\}$ be the set of feasible values for the vector $(S^{(x,y)}_j)_{j\in[1,c],x,y\in\mathcal{X}\times\mathcal{Y}}$. We have:

$$\mathbb{E}(e^{\frac{1}{s}\sum_{x,y}\sum_{k=1}^{i_j} t_{x,y}Z^{\{(x,y)\}}_{(\pi),i}}) = \mathbb{E}(e^{\sum_{x,y} t_{x,y}\frac{1}{s}\sum_{k=1}^{i_j} Z^{\{(x,y)\}}_{(\pi),i}}) = \mathbb{E}(e^{\sum_{x,y} t_{x,y}S^{(x,y)}_j})$$

$$= \sum_{(s_j^{\{(x,y)\}})_{j\in[1,c],x,y\in\mathcal{X}\times\mathcal{Y}}\in\beta} P((s_j^{\{(x,y)\}})_{j\in[1,c],x,y\in\mathcal{X}\times\mathcal{Y}})e^{\sum_{x,y} t_{x,y}s_j^{\{(x,y)\}}}.$$

For a fixed vector $(s_j^{\{(x,y)\}})_{j\in[1,c],x,y\in\mathcal{X}} \in \beta$, let $V^{(x,y)}$ be defined as the random variable for which $P(V^{(x,y)} = t_{(x,y)}) = s_j^{\{(x,y)\}}, x, y \in \mathcal{X}$ and $P(V^{(x,y)} = 0) = 1 - \frac{i_j}{s}$ (note that P_V is a valid probability distribution). We have:

$$\mathbb{E}(e^{\frac{1}{s}\sum_{x,y}\sum_{k=1}^{i_j} t_{x,y}Z^{\{(x,y)\}}_{(\pi),i}}) = \sum_{(s_j^{\{(x,y)\}})_{j\in[1,c],x,y\in\mathcal{X}\times\mathcal{Y}}\in\beta} P((s_j^{\{(x,y)\}})_{j\in[1,c],x,y\in\mathcal{X}\times\mathcal{Y}})e^{\sum_{x,y} t_{x,y}s_j^{\{(x,y)\}}}$$

$$= \sum_{(s_j^{\{(x,y)\}})_{j\in[1,c],x,y\in\mathcal{X}\times\mathcal{Y}}\in\beta} P((s_j^{\{(x,y)\}})_{j\in[1,c],x,y\in\mathcal{X}\times\mathcal{Y}})e^{\mathbb{E}(V^{(x,y)})} \leq \sum_{(s_j^{\{(x,y)\}})_{j\in[1,c],x,y\in\mathcal{X}\times\mathcal{Y}}\in\beta} P((s_j^{\{(x,y)\}})_{j\in[1,c],x,y\in\mathcal{X}\times\mathcal{Y}})\mathbb{E}(e^{V^{(x,y)}}),$$

where we have used Jensen's inequality in the last equation. Note that by construction, we have $\mathbb{E}(e^{V^{(x,y)}}) = 1 - \frac{i_j}{s} + \sum_{x,y} s_j^{(x,y)}e^{t_{x,y}}$. Consequently:

$$\mathbb{E}(e^{\frac{1}{s}\sum_{x,y}\sum_{k=1}^{i_j} t_{x,y}Z^{\{(x,y)\}}_{(\pi),i}}) \leq \sum_{(s_j^{\{(x,y)\}})_{j\in[1,c],x,y\in\mathcal{X}\times\mathcal{Y}}\in\beta} P((s_j^{\{(x,y)\}})_{j\in[1,c],x,y\in\mathcal{X}\times\mathcal{Y}})(1 - \frac{i_j}{s} + \sum_{x,y} s_j^{(x,y)}e^{t_{x,y}})$$

$$= 1 - \frac{i_j}{s} + \sum_{x,y} e^{t_{x,y}}\mathbb{E}(S_j^{(x,y)}) = 1 - \frac{i_j}{s} + \sum_{x,y} e^{t_{x,y}}\mathbb{E}(\frac{1}{s}\sum_{k=1}^{i_j} Z^{\{(x,y)\}}_{(\pi),i})$$

$$= 1 - \frac{i_j}{s} + \frac{1}{s}\sum_{x,y}\sum_{k=1}^{i_j} e^{t_{x,y}}\mathbb{E}(Z^{\{(x,y)\}}_{(\pi),i}) = 1 - \frac{i_j}{s} + \frac{1}{s}\sum_{x,y}\sum_{k=1}^{i_j} e^{t_{x,y}}P_X(x)P_Y(y)$$

$$= 1 - \frac{i_j}{s} + \frac{1}{s}\sum_{x,y}\sum_{k=1}^{i_j} P_{X,Y}(x,y) = 1.$$

Setting $\mathbb{E}(e^{\frac{1}{s}\sum_{x,y}\sum_{k=1}^{i_j} t_{x,y}Z^{\{(x,y)\}}_{(\pi),i}}) \leq 1$ in Equation (17), we get:

$$P((X^n, \pi(Y^n)) \in \mathcal{A}^n_\epsilon(X,Y)) \leq e^{-\sum_{x,y}\frac{n}{s}(t_{x,y}P_{X,Y}(x,y)-\epsilon_{x,y})} = 2^{-\frac{n}{s}(I(X;Y)-\epsilon_{x,y})}.$$

□

Appendix E
Proof of Theorem 2

The proof builds upon the arguments provided in the proof of Theorem 1. Let $Y^n = \pi_j(X^n_{(j)})_{j\in[m]}$. First, we construct a partition $\mathsf{D} = \{C_{k,l} : k \in [b_m], l \in [m(m-1)]\}$ such that each sequence of vectors $(Y_{(j),C_{k,l}})_{j\in[m]}$ is an collection of independent vectors of i.i.d variables, where $Y_{(j),C_{k,l}} = (Y_{(j),c})_{c\in C_{k,l}}$. Loosely speaking, this partitioning of the indices 'breaks' the multi-letter correlation among the sequences induced due to the permutation and allows the application of standard information theoretic tools to bound the probability of joint typicality. The partition is constructed in two steps. We first construct a *coarse* partition $\mathsf{C} = \{C_1, C_2, \cdots, C_{b_m}\}$ of the indices $[1, n]$ for which the sequence of vectors $(Y_{(j),C_k}), j \in [m]$ is identically distributed but not necessarily independent. The set $C_k, k \in [b_m]$ is defined as the set of indices corresponding to partition $\mathcal{P}_k$, where correspondence is defined in Definition 13. Clearly, $\mathsf{C} = \{C_1, C_2, \cdots, C_{b_m}\}$ partitions $[1, n]$ since each index corresponds to exactly one partition $\mathcal{P}_k$. To verify that the elements of the sequence $(Y_{(j),C_k}), j \in [m]$ are identically distributed let us consider a fixed $k \in [b_m]$ and an arbitrary index $c \in C_k$. Then the vector $(Y_{(1),c}, Y_{(2),c}, \cdots, Y_{(m),c})$ is distributed according to $P_{X_{\mathcal{P}_k}}$. To see this, note that:

$$P_{Y_{(1),c},Y_{(2),c},\cdots,Y_{(m),c}} = P_{X_{(1),(\pi_1^{-1}(c))},X_{(2),(\pi_2^{-1}(c))},\cdots,X_{(m),(\pi_m^{-1}(c))}}$$

From the assumption that the index c corresponds to the partition $\mathcal{P}_k$, we have that $\pi_j^{-1}(c) = \pi_{j'}^{-1}(c)$ if and only if $j, j' \in \mathcal{A}_{k,r}$ for some integer $r \in [|\mathcal{P}_k|]$. Since by the theorem statement $(X_{(j)}^n)_{j\in[m]}$ is an i.i.d. sequence of vectors, the variables $X_{(j),\pi_j^{-1}(c)}$ and $X_{(j'),\pi_{j'}^{-1}(c)}$ are independent of each other if $\pi_j^{-1}(c) \neq \pi_{j'}^{-1}(c)$. Consequently,

$$P_{Y_{(1),c},Y_{(2),c},\cdots,Y_{(m),c}} = \prod_{r\in[|\mathcal{P}_k|]} P_{X_{l_1},X_{l_2},\cdots,X_{l_{|\mathcal{A}_{k,r}|}}} = P_{X_{\mathcal{P}_k}}.$$

This proves that the sequences $(Y_{(j),C_k}), j \in [m]$ are identically distributed with distribution $P_{X_{\mathcal{P}_k}}$. In the next step, we decompose the partition C to arrive at a finer partition $\mathsf{D} = \{C_{k,l} : k \in [b_m], l \in [m(m-1)]\}$ of $[1, n]$ such that $(Y_{(j),C_{k,l}})_{j\in[m]}$ is an i.i.d sequence of vectors. Let $C_k = \{c_1, c_2, \cdots, c_{|C_k|}\}, k \in [b_m]$. The previous step shows that the sequence consists of identically distributed vectors. In order to guarantee independence, we need to ensure that for any $c, c' \in C_{k,l}$, we have $\pi_j^{-1}(c) \neq \pi_{j'}^{-1}(c'), \forall j, j' \in [m]$. Then, independence of $(Y_{(j),c})_{j\in[m]}$ and $(Y_{(j),c'})_{j\in[m]}$ is guaranteed due to the independence of the sequence of vectors $(X_{(j)}^n)_{j\in[m]}$. To this end we assign the indices in C_k to the sets $C_{k,l}, l \in [m(m-1)]$ as follows:

$$c_1 \in C_{k,1}, \tag{18}$$

$$c_i \in C_{k,l} : l = \min\{l' | \nexists c' \in C_{k,l'}, j, j' \in [m] : \pi_j^{-1}(c_i) = \pi_{j'}^{-1}(c')\}, i > 1. \tag{19}$$

Note that the set $C_{k,l}$ defined in Equation (19) always exists since for any given $j \in [m]$, the value $\pi_j^{-1}(c)$ can be the same for at most m distinct indices c since each of the m permutations maps one index to $\pi_j^{-1}(c)$. Furthermore, since j takes m distinct values, there are at most $m(m-1)-1$ indices c' not equal to c for which there exists $j, j' \in [m]$ such that $\pi_j(c) = \pi_{j'}(c')$. Since there are a total of $m(m-1)$ sets $C_{k,l}$, by the Pigeonhole Principle, there exists at least one set for which there is no element c' such that $\pi_j(c) = \pi_{j'}(c')$ for any value of j, j'. Consequently, $(Y_{(j),C_{k,l}})_{j\in[m]}$ is an i.i.d. sequence with distribution $P_{X_{\mathcal{P}_k}}$.

Let $\underline{T}_{k,l}, k \in [b_m], l \in [m(m-1)]$ be the type of the sequence of vectors $(Y_{(j),C_{k,l}})_{j\in[m]}$, so that $T_{k,l,x^m} = \frac{\sum_{c\in C_{k,l}} \mathbb{1}((Y_{(1),c}, Y_{(2),c}, \cdots, Y_{(m),c}) = x^m)}{|C_{k,l}|}, x^m \in \mathcal{X}^m$. We are interested in the probability of the event $(Y^n_{(j)})_{j\in[m]} \in \mathcal{A}^n_\epsilon(X^m)$. From Definition 11 this event can be rewritten as follows:

$$
\begin{aligned}
P\left(\left(Y^n_{(j)}\right)_{j\in[m]}\right) \in \mathcal{A}^n_\epsilon(X^m)\Big) &= P\left(T((Y^n_{(j)})_{j\in[m]}, x^m) \doteq P_{X^m}(x^m) \pm \epsilon, \forall x^m\right) \\
&= P(\sum_{k,l} \alpha_{k,l} T_{k,l,x^m} \doteq P_{X^m}(x^m) \pm \epsilon, \forall x^m),
\end{aligned}
$$

where $\alpha_{k,l} = \frac{|C_{k,l}|}{n}, k \in [b_m], l \in [m(m-1)]$, we write $a \doteq x \pm \epsilon$ to denote $x - \epsilon \leq a \leq x_\epsilon$, and addition is defined element-wise. We have:

$$
P\left(\left(Y^n_{(j)}\right)_{j\in[m]}\right) \in \mathcal{A}^n_\epsilon(X^m)\Big) = \sum_{(\underline{t}^{b_m, m(m-1)})\in\mathcal{T}} P(\underline{T}_{k,l} = \underline{t}_{k,l}, k \in [b_m], l \in [m(m-1)]),
$$

where $\mathcal{T} = \{(\underline{t}^{b_m,m(m-1)} : \sum_{k,l} \alpha_{k,l} T_{k,l,x^m} \doteq P_{X^m}(x^m) \pm \epsilon, \forall x^m\}$. Using the property that for any set of events, the probability of the intersection is less than or equal to the geometric average of the individual probabilities, we have:

$$
P((Y^n_{(j)})_{j\in[m]} \in \mathcal{A}^n_\epsilon(X^m)) \leq \sum_{(\underline{t}^{b_m, m(m-1)})\in\mathcal{T}} \sqrt[m(m-1)b_m]{\prod_{i\in[k,l]} P(\underline{T}_{k,l} = \underline{t}_{k,l})}.
$$

Since the elements $(Y_{(j),C_{k,l}}), k \in [b_m], l \in [m(m-1)]$ are i.i.d by construction, it follows from standard information theoretic arguments [32] that:

$$
P(\underline{T}_{k,l} = \underline{t}_{k,l}) \leq 2^{-|C_{k,l}|(D(\underline{t}_i \| P_{X_{\mathcal{P}_k}}) - \prod_{j\in[m]} |\mathcal{X}_j| \epsilon)}, k \in [b_m], l \in [m(m-1)].
$$

We have,

$$
\begin{aligned}
P((Y^n_{(j)})_{j\in[m]} \in \mathcal{A}^n_\epsilon(X^m)) \leq& \sum_{(\underline{t}^{b_m, m(m-1)})\in\mathcal{T}} \sqrt[m(m-1)b_m]{\prod_{i\in[k,l]} 2^{-|C_{k,l}|(D(\underline{t}_i \| P_{X_{\mathcal{P}_k}}) - \prod_{j\in[m]} |\mathcal{X}_j| \epsilon)}} \\
\overset{(a)}{\leq}& \sum_{(\underline{t}^{b_m, m(m-1)})\in\mathcal{T}} \sqrt[m(m-1)b_m]{2^{-n(D(\sum_{k,l} \alpha_{k,l} \underline{t}_{k,l} \| \sum_k P_{X_{\mathcal{P}_k}}) - \prod_{j\in[m]} |\mathcal{X}_j| \epsilon)}} \\
\overset{(b)}{\leq}& 2^{-\frac{n}{m(m-1)b_m}(D(P_{X,Y} \| \sum_{k\in[b_m]} \frac{|C_k|}{n} P_{X_{\mathcal{P}_k}}) - \epsilon \prod_{j\in[m]} |\mathcal{X}_j| + O(\frac{\log n}{n}))}.
\end{aligned}
$$

where the (a) follows from the convexity of the divergence function and (b) follows by the fact that the number of joint types grows polynomially in n.

□

Appendix F
Proof of Lemma 5

The upper-bound follows by the fact that for r-fold derangement $(\pi_1(\cdot), \pi_2(\cdot), \cdots, \pi_m(\cdot))$, the first permutation is $\pi_1(\cdot)$ is the identity permutation, and the rest of derangements with respect to $\pi_1(\cdot)$, so by the counting principle there are at most $(!n)^{r-1}$ choices for $(\pi_1(\cdot), \pi_2(\cdot), \cdots, \pi_m(\cdot))$. Next we prove the lower bound. Note that $\pi_1(\cdot)$ is the identity permutation. By the same arguments as in the proof of Lemma 2, there are at least $(n-1)!$ choices of distinct $\pi_2(\cdot)$, and for any fixed $\pi_2(\cdot)$ there are at least $(n-2)!$ distinct $\pi_3(\cdot)$. Generally, for fixed $\pi_2(\cdot), \pi_3(\cdot), \cdots, \pi_j(\cdot)$, there are at least $(n-j+1)!$ choices of distinct $\pi_{j+1}(\cdot)$. By the counting principle, there are at least $\prod_{j\in[r]}(n-j+1)! \geq ((n-r+1)!)^r$ distinct $(\pi_1(\cdot), \pi_2\cdot, \cdots, \pi_r(\cdot))$. This completes the proof. □

Appendix G
Proof of Lemma 6

First, we prove the upper-bound in Equation (7). As an initial step, we count the number of distinct allocations of partition correspondence to indices $i \in [1, n]$. Since we are considering $(i_1, i_2, \cdots, i_{b_m})$-Bell permutation vectors, there are a total of i_k indices corresponding to $\mathcal{P}_k$ for $k \in [b_m]$. So, there are $\binom{n}{i_1, i_2, \cdots, i_{b_m}}$ allocations of partition correspondence to different indices. Now assume that the i^{th} index corresponds to the kth partition. Then, we argue that there are at most $n^{|\mathcal{P}_k|}$ possible values for the vector $(\pi_j(i) : j \in [m])$. The reason is that by definition, for any two $\pi_j(i)$ and $\pi_{j'}(i)$, their value are equal if and only if $j, j' \in \mathcal{A}_{k,r}$ for some integer $r \in [|\mathcal{P}_k|]$. So, the elements of $(\pi_j(i) : j \in [m])$ take $|\mathcal{P}_k|$ distinct values among the set $[1, n]$. Consequently $(\pi_j(i) : j \in [m])$ takes at most $n^{|\mathcal{P}_k|}$ distinct values. By the counting principle, the sequence of vectors $(\pi_j(i) : j \in [m]), i \in [n]$ takes at most $n^{\sum_{k\in[b_m]}|\mathcal{P}_k|i_k - n}$ distinct values given a specific partition correspondence, since $\pi_1(\cdot)$ is assumed to be the identity permutation. Since there are a total of $\binom{n}{i_1, i_2, \cdots, i_{b_m}}$ partition correspondences, we have:

$$N_{i_1, i_2, \cdots, i_{b_m}} \leq \binom{n}{i_1, i_2, \cdots, i_{b_m}} n^{\sum_{k\in[b_m]}|\mathcal{P}_k|i_k - n}.$$

Next, we prove the lower-bound in Equation (7). The proof follows by constructing enough distinct $(i_1, i_2, \cdots, i_{b_m})$-Bell permutation vectors. First, we choose a partition correspondence for the indices $i \in [n]$ similar to the proof for the lower-bound. There are $\binom{n}{i_1, i_2, \cdots, i_{b_m}}$ distinct ways of allocating the partition correspondence. We argue that for every fixed partition correspondence, there are at least $\prod_{k\in[b_m]]} d_{|\mathcal{P}_k|}(i_k)$ permutations which are $(i_1, i_2, \cdots, i_{b_m})$-Bell permutation vectors. To see this, without loss of generality, assume that the first i_1 indices $[1, i_1]$ correspond to $\mathcal{P}_1$, the next i_2 indices $[i_1+1, i_1+i_2]$ correspond to $\mathcal{P}_2$, and in general the indices $[\sum_{t=1}^{k-1} i_t + 1, \sum_{t=1}^{k} i_t]$ correspond to $\mathcal{P}_k$. Let $(\pi'_{1,k}, \pi'_{2,k}, \cdots, \pi'_{|\mathcal{P}_k|,k})$ be vectors of $|\mathcal{P}_k|$-fold derangements of $[\sum_{t=1}^{k-1} i_t + 1, \sum_{t=1}^{k} i_t]$, where $k \in [b_m]$. Then, the following is an $(i_1, i_2, \cdots, i_{b_m})$-Bell permutation vector.

$$\pi_j([\sum_{t=1}^{k-1} i_t + 1, \sum_{t=1}^{k} i_t]) = \pi'_{l,k}([\sum_{t=1}^{k-1} i_t + 1, \sum_{t=1}^{k} i_t]), \quad \text{if } j \in \mathcal{A}_{l,k}, l \in [|\mathcal{P}_k|], k \in [b_m].$$

There are a total of $d_{|\mathcal{P}_k|(i_k)}$ choices of $(\pi'_{1,k}, \pi'_{2,k}, \cdots, \pi'_{|\mathcal{P}_k|,k})$. So, by the counting principle, there are a total of $\prod_{k\in[b_m]} d_{|\mathcal{P}_k|}(i_k)$ choices of $(\pi_1(\cdot), \pi_2(\cdot), \cdots, \pi_m(\cdot))$ for a fixed partition correspondence. As argued previously, there are a total of $\binom{n}{i_1, i_2, \cdots, i_{b_m}}$ distinct choices for partition correspondence. Consequently we have shown that,

$$\binom{n}{i_1, i_2, \cdots, i_{b_m}} \prod_{k\in[b_m]} d_{|\mathcal{P}_k|}(i_k) \leq N_{i_1, i_2, \cdots, i_{b_m}}.$$

This completes the proof of Equation (7). We proceed with to prove Equation (8). Note that from the right hand side of Equation (7), we have:

$$\lim_{n\to\infty} \frac{\log_e N_{i_1,i_2,\cdots,i_{b_m}}}{n \log_e n} \leq \lim_{n\to\infty} \frac{\log_e \binom{n}{i_1,i_2,\cdots,i_{b_m}} n^{(\sum_{k\in[b_m]} |\mathcal{P}_k| i_k - n)}}{n \log_e n} = \lim_{n\to\infty} \frac{\log_e n^{(\sum_{k\in[b_m]} |\mathcal{P}_k| i_k - n)}}{n \log_e n} + \lim_{n\to\infty} \frac{\log_e \binom{n}{i_1,i_2,\cdots,i_{b_m}}}{n \log_e n}$$

$$= \lim_{n\to\infty} \frac{(\sum_{k\in[b_m]} |\mathcal{P}_k| i_k - n)}{n} + \lim_{n\to\infty} \frac{\log_e 2^n}{n \log_e n} = \sum_{k\in[b_m]} |\mathcal{P}_k| \alpha_k - 1.$$

On the other hand, from the left hand side of Equation (7), we have:

$$\lim_{n\to\infty} \frac{\log N_{i_1,i_2,\cdots,i_{b_m}}}{n \log n} \geq \lim_{n\to\infty} \frac{\log \binom{n}{i_1,i_2,\cdots,i_{b_m}} \prod_{k\in[b_m]} d_{|\mathcal{P}_k|}(i_k)}{n \log n}$$

$$\overset{(a)}{\geq} \lim_{n\to\infty} \frac{\log 2^n \prod_{k\in[b_m]} d_{|\mathcal{P}_k|}(i_k)}{n \log n} \overset{(b)}{\geq} \lim_{n\to\infty} \frac{\log \prod_{k\in[b_m]} ((i_k - |\mathcal{P}_k| + 1)!^{|\mathcal{P}_k| - 1})}{n \log n}$$

$$= \lim_{n\to\infty} \frac{\sum_{k\in[b_m]} (|\mathcal{P}_k| - 1) \log (i_k - |\mathcal{P}_k| + 1)!}{n \log n}$$

$$\overset{(c)}{=} \lim_{n\to\infty} \frac{\sum_{k\in[b_m]} (|\mathcal{P}_k| - 1)((i_k - |\mathcal{P}_k| + 1) \log (i_k - |\mathcal{P}_k| + 1) - (i_k - |\mathcal{P}_k| + 1) + O(\log (i_k - |\mathcal{P}_k| + 1)))}{n \log n}$$

$$= \sum_{k\in[b_m]} |\mathcal{P}_k| \alpha_k - 1,$$

where (a) follows from the fact that $\binom{n}{i_1,i_2,\cdots,i_{b_m}} \leq 2^n$, (b) follows from Lemma 5, and in (c) we have used Stirling's approximation. □

Appendix H
Proof of Theorem 3

First, note that for the correct labeling the two UTs are jointly typical with probability approaching one as $n \to \infty$:

$$P((U^1_{\sigma^1}, U^2_{\sigma^2}) \in \mathcal{A}_\epsilon^{\frac{n(n-1)}{2}}(X_1, X_2)) \to 1 \quad \text{as} \quad n \to \infty.$$

So, $P(\widehat{\Sigma} = \phi) \to 0$ as $n \to \infty$ since the correct labeling is a member of the set $\widehat{\Sigma}$. We will show that the probability that a labeling in $\widehat{\Sigma}$ labels $n(1-\alpha_n)$ vertices incorrectly goes to 0 as $n \to \infty$. Define the following:

$$\mathcal{E} = \{\sigma'^2 \Big| ||\sigma^2 - \sigma'^2||_0 \geq n(1-\alpha_n)\},$$

where $|| \cdot ||_0$ is the L_0-norm. The set $\mathcal{E}$ is the set of all labelings which match more than $n\alpha_n$ vertices incorrectly. We show the following:

$$P(\mathcal{E} \cap \widehat{\Sigma} \neq \phi) \to 0, \qquad \text{as} \qquad n \to \infty.$$

Note that:

$$P(\mathcal{E} \cap \widehat{\Sigma} \neq \phi) = P\left(\bigcup_{\sigma'^2:||\sigma^2-\sigma'^2||_0 \geq n(1-\alpha_n)} \{\sigma'^2 \in \widehat{\Sigma}\} \right) \overset{(a)}{\leq} \sum_{i=0}^{n\alpha_n} \sum_{\sigma'^2:||\sigma^2-\sigma'^2||_0 = n-i} P(\sigma'^2 \in \widehat{\Sigma})$$

$$\overset{(b)}{=} \sum_{i=0}^{n\alpha_n} \sum_{\sigma'^2:||\sigma^2-\sigma'^2||_i = n-i} P((U^1_{\sigma^1}, \Pi_{\sigma^2,\sigma'^2}(U^2_{\sigma^2})) \in \mathcal{A}_\epsilon^{\frac{n(n-1)}{2}})$$

$$\overset{(c)}{\leq} \sum_{i=0}^{n\alpha_n} \sum_{\sigma'^2:||\sigma^2-\sigma'^2||_i = n-i} 2^{-\frac{n(n-1)}{8}(D(P_{X,Y}||(1-\frac{i(i-1)}{n(n-1)})P_XP_Y+\frac{i(i-1)}{n(n-1)}P_{X,Y})-|\mathcal{X}||\mathcal{Y}|\epsilon+O(\frac{\log n}{n^2}))}$$

$$\overset{(d)}{=} \sum_{i=0}^{n\alpha_n} \binom{n}{i} (!(n-i)) 2^{-\frac{n(n-1)}{8}(D(P_{X,Y}||(1-\frac{i(i-1)}{n(n-1)})P_XP_Y+\frac{i(i-1)}{n(n-1)}P_{X,Y})-|\mathcal{X}||\mathcal{Y}|\epsilon+O(\frac{\log n}{n^2}))}$$

$$\leq \sum_{i=0}^{n\alpha_n} n^{n-i} 2^{-\frac{n(n-1)}{8}(D(P_{X,Y}||(1-\frac{i(i-1)}{n(n-1)})P_XP_Y+\frac{i(i-1)}{n(n-1)}P_{X,Y})-|\mathcal{X}||\mathcal{Y}|\epsilon+O(\frac{\log n}{n^2}))}$$

$$\leq \sum_{i=0}^{n\alpha_n} 2^{(n-i)\log n-\frac{n(n-1)}{8}(D(P_{X,Y}||(1-\frac{i(i-1)}{n(n-1)})P_XP_Y+\frac{i(i-1)}{n(n-1)}P_{X,Y})-|\mathcal{X}||\mathcal{Y}|\epsilon+O(\frac{\log n}{n^2}))}.$$

where (a) follows from the union bound, (b) follows from the definition of $\widehat{\Sigma}$, in (c) we have used Theorem 1 and the fact that $||\sigma^2 - \sigma'^2||_0 = n - i$ so that Π_{σ^2,σ'^2} has $\frac{i(i-1)}{2}$ fixed points, in (d) we have denoted the number of derangement of sequences of length i by $!i$. Note that the right hand side in the last inequality approaches 0 as $n \to \infty$ as long as:

$$(n-i)\log n \leq \frac{n(n-1)}{8}(D(P_{X,Y}||(1-\frac{i(i-1)}{n(n-1)})P_XP_Y + \frac{i(i-1)}{n(n-1)}P_{X,Y}) - |\mathcal{X}||\mathcal{Y}|\epsilon + O(\frac{\log n}{n^2})), i \in [0, n\alpha_n]$$

$$\leftrightarrow (1-\alpha)\log n \leq \frac{(n-1)}{8}(D(P_{X,Y}||(1-\alpha^2)P_XP_Y + \alpha^2 P_{X,Y}) - |\mathcal{X}||\mathcal{Y}|\epsilon + O(\frac{\log n}{n^2})), \alpha \in [1, 1-\alpha_n],$$

where we have defined $\alpha = \frac{i}{n}$. The last equation is satisfied by the theorem assumption for small enough ϵ. □

Appendix I
Proof of Theorem 4

Let $\epsilon_n = O(\frac{\log n}{n})$ be a sequence of positive numbers. Fix $n \in \mathbb{N}$ and let $\epsilon = \epsilon_n$. For a given labeling σ'', define the event $\mathcal{B}_{\sigma''}$ as the event that the sub-matrices corresponding to each community pair are jointly typical:

$$\mathcal{B}_{\sigma''} : (U_{\sigma,C_i,C_i}, U'_{\sigma'',C'_i,C'_i}) \in \mathcal{A}_\epsilon^{\frac{n_i(n_i-1)}{2}}(P_{X,X'|C_i,C_i,C'_i,C'_i}), \forall i \in [c],$$

$$(G_{\sigma,C_i,C_j}, \widetilde{G}'_{\sigma'',C'_i,C'_j}) \in \mathcal{A}_\epsilon^{n_i \cdot n_j}(P_{X,X'|C_i,C_j,C'_i,C'_j}), \forall i, j \in [c], i \neq j\},$$

Particularly, $\beta_{\sigma'}$ is the event that the sub-matrices are jointly typical under the canonical labeling for the second graph. From standard typicality arguments it follows that:

$$P(\mathcal{B}_{\sigma'}) \to 1 \quad \text{as} \quad n \to \infty.$$

So, $P(\widehat{\Sigma}_{C.C'} = \phi) \to 0$ as $n \to \infty$ since the correct labeling is a member of the set $\widehat{\Sigma}_{C.C'}$. Let $(\lambda_n)_{n\in\mathbb{N}}$ be an arbitrary sequence of numbers such that $\lambda_n = \Theta(n)$. We will show that the probability that a labeling in $\widehat{\Sigma}_{C.C'}$ labels λ_n vertices incorrectly goes to 0 as $n \to \infty$. Define the following:

$$\mathcal{E} = \{\sigma'^2 \Big| ||\sigma^2 - \sigma'^2||_1 \geq \lambda_n\},$$

where $|| \cdot ||_1$ is the L_1-norm. The set $\mathcal{E}$ is the set of all labelings which match more than λ_n vertices incorrectly.

We show the following:

$$P(\mathcal{E} \cap \widehat{\Sigma}_{C.C'} \neq \phi) \to 0, \qquad \text{as} \qquad n \to \infty.$$

We use the union bound on the set of all permutations along with Theorem 1 as follows:

$$P(\mathcal{E} \cap \widehat{\Sigma}_{C.C'} \neq \phi) = P(\bigcup_{\sigma'':||\sigma'-\sigma''||_1 \geq \lambda_n} \{\sigma'' \in \widehat{\Sigma}_{C.C'}\}) \overset{(a)}{\leq} \sum_{k=\lambda_n}^{n} \sum_{\sigma'':||\sigma'-\sigma''||_1=k} P(\sigma'' \in \widehat{\Sigma}_{C.C'})$$

$$\overset{(b)}{=} \sum_{k=\lambda_n}^{n} \sum_{\sigma'':||\sigma'-\sigma'''||_1=k} P(\beta_{\sigma''}) \overset{(c)}{\leq} \sum_{k=\lambda_n}^{n} \sum_{\sigma'^2:||\sigma^2-\sigma'^2||_0=k} 2^{O(nlogn))}\times$$

$$
\begin{aligned}
&\prod_{i,j\in[c],i<j} 2^{-\frac{n_i\cdot n_j}{4}(D(P_{X,X'|C_i,C_j,C'_i,C'_j}||(1-\beta_{i,j})P_{X|C_i,C_j}P_{X'|C'_i,C'_j}+\beta_{i,j}P_{X,X'|C_i,C_j,C'_i,C'_j}))}\\
&\times\prod_{i\in[c]}2^{-\frac{n_i(n_i-1)}{8}(D(P_{X,X'|C_i,C_i,C'_i,C'_i}||(1-\beta_i)P_{X|C_i,C_i}P_{X'|C'_i,C'_i}+\beta_iP_{X,X'|C_i,C_i,C'_i,C'_i})}\\
&\overset{(d)}{\leq}\sum_{k=\lambda_n}^{n}\binom{n}{k}(!k)\max_{[\alpha_i]_{i\in[c]}\in\mathcal{A}}(2^{-\frac{n^2}{4}(\Phi([\alpha_i]_{i\in[c]})+O(\frac{\log n}{n}))})\\
&\leq\max_{\alpha\in[0,1-\frac{\lambda_n}{n}]}\max_{[\alpha_i]_{i\in[c]}}(2^{-\frac{n^2}{4}(-(1-\alpha)\frac{\log n}{n}+\Phi([\alpha_i]_{i\in[c]})+O(\frac{\log n}{n}))}),
\end{aligned}
$$

where $\mathcal{A}=\{([\alpha_i]_{i\in[c]}):\alpha_i\leq\frac{n_i}{n},\sum_{i\in[c]}\alpha_i=\frac{n-\lambda_n}{n}\}$ and

$$
\begin{aligned}
\Phi([\alpha_i]_{i\in[c]})=&\sum_{i,j\in[c],i<j}n_in_j\cdot D(P_{X,X'|C_i,C_j,C'_i,C'_j}||(1-\beta_{i,j})P_{X|C_i,C_j}P_{X'|C'_i,C'_j}+\beta_{i,j}P_{X,X'|C_i,C_j,C'_i,C'_j})\\
&+\sum_{i\in[c]}\frac{n_i(n_i-1)}{2}D(P_{X,X'|C_i,C_i,C'_i,C'_i}||(1-\beta_i)P_{X|C_i,C_i}P_{X'|C'_i,C'_i}+\beta_iP_{X,X'|C_i,C_i,C'_i,C'_i}),
\end{aligned}
$$

and $\beta_{i,j}=\frac{n^2}{n_in_j}\alpha_i\alpha_j$ and $\beta_i=\frac{n\alpha_i(n\alpha_i-1)}{n_i(n_i-1)}$. Here, α_i is the number of fixed points in the i^{th} community divided by n, and β_i is the number of fixed points in U'_{σ'',C'_i,C'_i} divided by $\frac{n_i(n_i-1)}{2}$, and $\beta_{i,j}$ is the number of fixed points in U'_{σ'',C'_i,C'_j} divided by n_in_j. Inequality (a) follows from the union bound, (b) follows from the definition of $\widehat{\Sigma}_{C.C'}$, in (c) we have used Theorem 1, in (d) we have denoted the number of derangement of sequences of length i by $!i$. Note that the right hand side in the (d) goes to 0 as $n\to\infty$ as long as (10) holds. □

,

Appendix J
Proof of Theorem 6

The proof build upon the arguments used in the proof of Theorem 3. First, note that for the correct labeling the UTs are jointly typical with probability approaching one as $n\to\infty$. So, $P(\widehat{\Sigma}=\phi)\to0$ as $n\to\infty$ since the correct labeling is a member of the set $\widehat{\Sigma}$. On the other hand, the probability that a labeling in $\widehat{\Sigma}$ labels $n(1-\alpha_n)$ vertices incorrectly goes to 0 as $n\to\infty$. Define the following:

$$\mathcal{E}=\{(\sigma'_i)_{i\in[m]}\Big|||(\sigma'_i)_{i\in[m]}-(\sigma_i)_{i\in[m]}||_0\geq n(1-\alpha_n)\},$$

where $||\cdot||_0$ is the L_0-norm. Without loss of generality, we assume that the labeling for the denonymized graph is the trivial labeling, i.e. $\sigma_1=\sigma'_1=id(\cdot)$, where $id(\cdot)$ is the identity function. The set $\mathcal{E}$ is the set of all labelings which match more than $n\alpha_n$ vertices incorrectly. We show the following:

$$P(\mathcal{E}\cap\widehat{\Sigma}\neq\phi)\to 0, \qquad \text{as} \qquad n\to\infty.$$

We partition the set $\mathcal{E}$ into subsets of Bell permutation vectors with the same parameters. We define the following:

$$\mathcal{E}_{i_1,i_2,\cdots,i_{b_m}}=\{(\sigma'_i)_{i\in[m]}\Big|(\sigma'_i)_{i\in[m]} \text{ is a } (i_1,i_2,\cdots,i_{b_m})\text{-Bell permutation vector}\},$$

where $i_1,i_2,\cdots,i_{b_m}\in[0,n]$ and $\sum_{k\in[b_m]} i_{\mathcal{P}_k}=n$. Then the family of sets $\{\mathcal{E}_{i_1,i_2,\cdots,i_{b_m}}:\sum_{k\in[b_m]} i_{\mathcal{P}_k}=n\}$ partitions the set $\mathcal{E}$.

Note that similar to the proof of Theorem 3, we have:

$$P(\mathcal{E}\cap\widehat{\Sigma}\neq\phi)=P\left(\bigcup_{(\sigma'_i)_{i\in[m]}:\|\sigma^2-\sigma'^2\|_0\geq n(1-\alpha_n)}\{\sigma'^2\in\widehat{\Sigma}\}\right)$$

$$P(\mathcal{E}\cap\widehat{\Sigma}\neq\phi)=P\left(\bigcup_{\substack{(i_1,i_2,\cdots,i_{b_m}):\\ \sum_{k\in[b_m]} i_k=n}}\ \bigcup_{\substack{(\sigma'_i)_{i\in[m]}\in\mathcal{E}_{i_1,i_2,\cdots,i_{b_m}}:\\ \|\sigma^2-\sigma'^2\|_0\geq n(1-\alpha_n)}}\{\sigma'^2\in\widehat{\Sigma}\}\right)$$

$$\leq\sum_{\substack{(i_1,i_2,\cdots,i_{b_m}):\\ \sum_{k\in[b_m]} i_k=n}}\ \sum_{\substack{(\sigma'_i)_{i\in[m]}\in\mathcal{E}_{i_1,i_2,\cdots,i_{b_m}}:\\ \|\sigma^2-\sigma'^2\|_0\geq n(1-\alpha_n)}} P\left(\{\sigma'^2\in\widehat{\Sigma}\}\right)$$

$$\leq\sum_{\substack{(i_1,i_2,\cdots,i_{b_m}):\\ \sum_{k\in[b_m]} i_k=n}}\ \sum_{\substack{(\sigma'_i)_{i\in[m]}\in\mathcal{E}_{(i_1,i_2,\cdots,i_{b_m})}:\\ \|\sigma^2-\sigma'^2\|_0\geq n(1-\alpha_n)}} 2^{-\frac{\frac{n(n-1)}{2}}{m(m-1)b_m}(D(P_{X^m}\|\sum_{k\in[b_m]}\frac{i'_k}{\frac{n(n-1)}{2}}P_{X_{\mathcal{P}_k}})-\epsilon\prod_{j\in[m]}|\mathcal{X}_j|+O(\frac{\log n}{n^2}))}$$

$$=\sum_{\substack{(i_1,i_2,\cdots,i_{b_m}):\\ \sum_{k\in[b_m]} i_k=n, i_{b_m}\geq n(1-\alpha_n)}} N_{i_1,i_2,\cdots,i_{b_m}}2^{-\frac{\frac{n(n-1)}{2}}{m(m-1)b_m}(D(P_{X^m}\|\sum_{k\in[b_m]}\frac{i'_k}{\frac{n(n-1)}{2}}P_{X_{\mathcal{P}_k}})-\epsilon\prod_{j\in[m]}|\mathcal{X}_j|+O(\frac{\log n}{n^2}))}$$

$$\leq\sum_{\substack{(i_1,i_2,\cdots,i_{b_m}):\\ \sum_{k\in[b_m]} i_k=n, i_{b_m}\geq n(1-\alpha_n)}} 2^{n\log n(\sum_{k\in[b_m]}|\mathcal{P}_k|\frac{i_k}{n}-1)+O(n\log n)}2^{-\frac{\frac{n(n-1)}{2}}{m(m-1)b_m}(D(P_{X^m}\|\sum_{k\in[b_m]}\frac{i'_k}{\frac{n(n-1)}{2}}P_{X_{\mathcal{P}_k}})-\epsilon\prod_{j\in[m]}|\mathcal{X}_j|+O(\frac{\log n}{n^2}))}$$

$$\leq\sum_{\substack{(i_1,i_2,\cdots,i_{b_m}):\\ \sum_{k\in[b_m]} i_k=n, i_{b_m}\geq n(1-\alpha_n)}} 2^{n\log n(\sum_{k\in[b_m]}|\mathcal{P}_k|\frac{i_k}{n}-1)+O(n\log n)-\frac{\frac{n(n-1)}{2}}{m(m-1)b_m}(D(P_{X^m}\|\sum_{k\in[b_m]}\frac{i'_k}{\frac{n(n-1)}{2}}P_{X_{\mathcal{P}_k}})-\epsilon\prod_{j\in[m]}|\mathcal{X}_j|+O(\frac{\log n}{n^2}))},$$

where $i'_k=\frac{i_k(i_k-1)}{2}+\sum_{k',k'':\mathcal{P}_{k',k''}=\mathcal{P}_l} i_{k'}i_{k''}$, and $\mathcal{P}_{k',k''}=\{\mathcal{A}'\cap\mathcal{A}'':\mathcal{A}'\in\mathcal{P}_{k'},\mathcal{A}''\in\mathcal{P}_{k''}\}, k',k''\in[b_m]$. Note that the right hand side in the last inequality approaches 0 as $n\to\infty$ as long as:

$$n\log n(\sum_{k\in[b_m]}|\mathcal{P}_k|\alpha_k-1)\leq\frac{n(n-1)}{2b_m m(m-1)}(D(P_{X^m}\|\sum_{k\in[b_m]}\alpha'_k P_{X_{\mathcal{P}_k}})-\epsilon\prod_{j\in[m]}|\mathcal{X}_j|+O(\frac{\log n}{n^2}))$$

$$\leftrightarrow \log n(\sum_{k\in[b_m]} |\mathcal{P}_k|\alpha_k - 1) \leq \frac{(n-1)}{2b_m m(m-1)}(D(P_{X^m}\| \sum_{k\in[b_m]} \alpha'_k P_{X_{\mathcal{P}_k}}) - \epsilon \prod_{j\in[m]} |\mathcal{X}_j| + O(\frac{\log n}{n^2})),$$

for all $\alpha_1, \alpha_2, \cdots, \alpha_{b_m} : \sum_{k\in[b_m]} \alpha_k = n, \alpha_{b_m} \in [1, 1-\alpha_n]$, where we have defined $\alpha'_k = \frac{i'_k}{\frac{n(n-1)}{2}}$. The last equation is satisfied by the theorem assumption for small enough ϵ. □

Appendix K
Proof of Theorem 7

Let $n \in \mathcal{N}$, and G and G' be the adjacency matrices of the two graphs under a pre-defined labeling. Let $\hat{\sigma}$ be the output of the matching algorithm. Let $\mathbb{1}_C$ be the indicator of the event that the matching algorithm mislabels at most ϵ_n fraction of the vertices with probability at least P_e, where $\epsilon_n, P_e \to 0$ as $n \to \infty$. Note that $\hat{\sigma}$ is a function of σ', G, G'. So:

$$0 = H(\hat{\sigma}|\sigma, G, G') \overset{(a)}{=} H(\sigma', \hat{\sigma}, \mathbb{1}_C|\sigma, G, G') - H(\sigma', \mathbb{1}_C|\hat{\sigma}, \sigma, G, G') = H(\sigma', \hat{\sigma}, \mathbb{1}_C|\sigma, G, G') -$$

$$H(\sigma'|\mathbb{1}_C, \hat{\sigma}, \sigma, G, G') - H(\mathbb{1}_C|\hat{\sigma}, \sigma, G, G') \overset{(b)}{\geq} H(\sigma', \hat{\sigma}, \mathbb{1}_C|\sigma, G, G') - H(\sigma'|\mathbb{1}_C, \hat{\sigma}, \sigma, G, G') - 1$$

$$= H(\sigma', \hat{\sigma}, \mathbb{1}_C|\sigma, G, G') - P(\mathbb{1}_C = 1)H(\sigma'|\mathbb{1}_C = 1, \hat{\sigma}, \sigma, G, G') - P(\mathbb{1}_C = 0)H(\sigma'|\mathbb{1}_C = 0, \hat{\sigma}, \sigma, G, G') - 1$$

$$\overset{(c)}{\geq} H(\sigma', \hat{\sigma}, \mathbb{1}_C|\sigma, G, G') - \epsilon_n n \log n - P_e n \log n - 1 \overset{(d)}{\geq} H(\sigma'|\sigma, G, G') - (\epsilon_n + P_e) n \log n - 1,$$

where in (a) we have used the chain rule of entropy, in (b) we have used the fact that $\mathbb{1}_C$ is binary, in (c) we define the probability of mismatching more than ϵ_n fraction of the vertices by P_e, and (d) follows from the fact that entropy is non-negative. As a result,

$$H(\sigma'|\sigma, G, G') \leq (\epsilon_n + P_e) n \log n + 1.$$

Consequently,

$$n \log n \overset{(a)}{=} \log n! + n + O(\log n) = H(\sigma') \overset{(b)}{=} I(\sigma'; \sigma, G, G') + O(n \log n),$$

where in (a) we have used Stirling's approximation, and in (b) we have used the fact that $\epsilon, P_e \to 0$ as $n \to \infty$. We have:

$$n \log n \leq I(\sigma'; \sigma, G, G') + O(n \log n)$$

$$= I(\sigma'; G') + I(\sigma'; \sigma, G|G') + O(n \log n) \overset{(a)}{=} I(\sigma'; \sigma, G|G') + O(n \log n)$$

$$= I(\sigma'; G|G') + I(\sigma'; G|G', \sigma) + O(n \log n) \overset{(b)}{=} I(\sigma'; G|G', \sigma) + O(n \log n)$$

$$\overset{(c)}{\leq} I(\sigma', G'; G|\sigma) + O(n \log n) \overset{(d)}{=} I(G'; G|\sigma, \sigma')$$
$$\overset{(e)}{=} \sum_{i,j\in[c],i<j} n_i n_j I(X, X'|C_i, C_j, C'_i C'_j) + \sum_{i\in[c]} \frac{n_i(n_i-1)}{2} I(X, X'|C_i, C_i, C'_i, C'_i) + O(n \log n),$$

where (a) follows from $\sigma' \perp\!\!\!\perp G'$, (b) follows from the fact that $\sigma' \perp\!\!\!\perp G, G'$, (c) is true due to the non-negativity of the mutual inforamtion, (d) follows from $\sigma, \sigma' \perp\!\!\!\perp G$, and (e) follows from the fact that the edges whose vertices have different labels are independent of each other given the labels.

□

Appendix L
Proof of Lemma 7

The ambiguity set $\mathcal{L}$ is defined as:

$$\mathcal{L} = \{v_j \big| \nexists ! i : (\underline{F}_i, \underline{F}'_j) \in \mathcal{A}^n_\epsilon(X, X')\}.$$

From the Chebychev inequality, we have:

$$P\left(|\mathcal{L}| > 2\mathbb{E}(|\mathcal{L}|)\right) = P\left(\big||\mathcal{L}| - \mathbb{E}(|\mathcal{L}|)\big| > \mathbb{E}(|\mathcal{L}|)\right) \leq \frac{Var(|\mathcal{L}|)}{\mathbb{E}^2(|\mathcal{L}|)}. \tag{20}$$

Let B_j be the event that $v_j \in \mathcal{L}$, then

$$\mathbb{E}(|\mathcal{L}|) = \mathbb{E}\left(\sum_{j=1}^{n} \mathbb{1}(v_j \in \mathcal{L})\right) = \sum_{j=1}^{n} P(vj \in \mathcal{L}) = \sum_{j=1}^{n} P(B_j)$$
$$Var(|\mathcal{L}|) = \sum_{j=1}^{n} P(B_j) + \sum_{i\neq j} P(B_i, B_j) - \left(\sum_{j=1}^{n} P(B_j)\right)^2$$
$$= \sum_{j=1}^{n} P(B_j) - \sum_{j=1}^{n} P^2(B_j) \leq \mathbb{E}(|\mathcal{L}|).$$

So, from (20), we have:

$$P\left(|\mathcal{L}| > 2\mathbb{E}(|\mathcal{L}|)\right) \leq \frac{1}{\mathbb{E}(|\mathcal{L}|)},$$

which goes to 0 as $n \to \infty$ provided that $\mathbb{E}(|\mathcal{L}|) \to \infty$ (otherwise the claim is proved since $\mathbb{E}(|\mathcal{L}|)$ is finite.). It remains to find an upper bound on $\mathbb{E}(|\mathcal{L}|)$. Let C_j be the event that the fingerprint $\underline{F}'_j$ is not typical with respect to $P_{X'}$ and let $D_{i,j}$ be the event that there exists $i \in [1, n]$ such that $\underline{F}_i$ and $\underline{F}'_j$ are jointly typical with respect to $P_{X,X'}$. Then,

$$P(B_j) \leq P\left(C_j \bigcup \left(\bigcup_{i\neq j} D_{i,j}\right)\right) \leq P(C_j) + \sum_{i\neq j} P(D_{i,j}|C_j^c) \overset{(a)}{\leq} \frac{1}{\Lambda\epsilon^2} + n2^{-\Lambda(I(X;X')-\epsilon)},$$

where (a) follows from the standard information theoretic arguments (e.g proof of Theorem 3 in [34]). So,

$$\mathbb{E}(|\mathcal{L}|) \leq \frac{n}{\Lambda\epsilon^2} + n^2 2^{-\Lambda(I(X;X')-\epsilon)}.$$

From $\Lambda > \frac{2\log n}{I(X;X')}$, we conclude that the second term approaches 0 as $n \to \infty$. This completes the proof. □

Appendix M
Proof of Theorem 8

Let H_1 be the event the algorithm fails and H_2 the event that $|\mathcal{L}| > \frac{2n}{\Lambda\epsilon^2}$. Then:

$$P(H_1) \leq P(H_2) + P(H_1|H_2^c).$$

From Claim 1, we know that $P(H_2) \to 0$ as $n \to \infty$. For the second term, let $\mathcal{L}'$ be the set of vertices which are not matched in the second iteration. The algorithm fails if $\mathcal{L}' \neq \phi$. However, by a similar argument as in the proof of claim 1, we have:

$$P(|\mathcal{L}'| > \frac{1}{2}\Big||\mathcal{L}| < \frac{2n}{\Lambda\epsilon^2}) \to 0 \text{ as } n \to \infty.$$

So, $P(|\mathcal{L}'| = 0) \to 1$ as $n \to \infty$. This completes the proof.

□

References

[1] Donatello Conte, Pasquale Foggia, Carlo Sansone, and Mario Vento. Thirty years of graph matching in pattern recognition. *International journal of pattern recognition and artificial intelligence*, 18(03):265–298, 2004.

[2] Frank Emmert-Streib, Matthias Dehmer, and Yongtang Shi. Fifty years of graph matching, network alignment and network comparison. *Information Sciences*, 346:180–197, 2016.

[3] Paul Erdos and Alfréd Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci*, 5(1):17–60, 1960.

[4] Edward M Wright. Graphs on unlabelled nodes with a given number of edges. *Acta Mathematica*, 126(1):1–9, 1971.

[5] László Babai, Paul Erdos, and Stanley M Selkow. Random graph isomorphism. *SIAM Journal on computing*, 9(3):628–635, 1980.

[6] Béla Bollobás. Random graphs. 2001. *Cambridge Stud. Adv. Math*, 2001.

[7] Tomek Czajka and Gopal Pandurangan. Improved random graph isomorphism. *Journal of Discrete Algorithms*, 6(1):85–92, 2008.

[8] Ehsan Kazemi. Network alignment: Theory, algorithms, and applications. 2016.

[9] Lyudmila Yartseva and Matthias Grossglauser. On the performance of percolation graph matching. In *Proceedings of the first ACM conference on Online social networks*, pages 119–130. ACM, 2013.

[10] Pedram Pedarsani, Daniel R Figueiredo, and Matthias Grossglauser. A bayesian method for matching two similar graphs without seeds. In *2013 51st Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 1598–1607. IEEE, 2013.

[11] Shouling Ji, Weiqing Li, Mudhakar Srivatsa, and Raheem Beyah. Structural data de-anonymization: Quantification, practice, and implications. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, pages 1040–1053. ACM, 2014.

[12] Daniel Cullina and Negar Kiyavash. Exact alignment recovery for correlated erdos renyi graphs. *arXiv preprint arXiv:1711.06783*, 2017.

[13] Vince Lyzinski. Information recovery in shuffled graphs via graph matching. *arXiv preprint arXiv:1605.02315*, 2016.

[14] Daniel Cullina, Negar Kiyavash, Prateek Mittal, and H Vincent Poor. Partial recovery of erd\h {o} sr\'{e} nyi graph alignment via *k*-core alignment. *arXiv preprint arXiv:1809.03553*, 2018.

[15] Michelle Girvan and Mark EJ Newman. Community structure in social and biological networks. *Proceedings of the national academy of sciences*, 99(12):7821–7826, 2002.

[16] Santo Fortunato and Claudio Castellano. Community structure in graphs. *Computational Complexity: Theory, Techniques, and Applications*, pages 490–512, 2012.

[17] Shirin Nilizadeh, Apu Kapadia, and Yong-Yeol Ahn. Community-enhanced de-anonymization of online social networks. In *Proceedings of the 2014 acm sigsac conference on computer and communications security*, pages 537–548. ACM, 2014.

[18] Kushagra Singhal, Daniel Cullina, and Negar Kiyavash. Significance of side information in the graph matching problem. *arXiv preprint arXiv:1706.06936*, 2017.

[19] Efe Onaran, Siddharth Garg, and Elza Erkip. Optimal de-anonymization in random graphs with community structure. In *Signals, Systems and Computers, 2016 50th Asilomar Conference on,*, pages 709–713. IEEE, 2016.

[20] Daniel Cullina and Negar Kiyavash. Improved achievability and converse bounds for erdos-renyi graph matching. *SIGMETRICS Perform. Eval. Rev.*, 44(1):63–72, June 2016.

[21] Ehsan Kazemi, S Hamed Hassani, and Matthias Grossglauser. Growing a graph matching from a handful of seeds. *Proceedings of the VLDB Endowment*, 8(10):1010–1021, 2015.

[22] Carla-Fabiana Chiasserini, Michele Garetto, and Emilio Leonardi. Social network de-anonymization under scale-free user relations. *IEEE/ACM Transactions on Networking*, 24(6):3756–3769, 2016.

[23] Vince Lyzinski, Donniell E Fishkind, and Carey E Priebe. Seeded graph matching for correlated erdös-rényi graphs. *Journal of Machine Learning Research*, 15(1):3513–3540, 2014.

[24] Marcelo Fiori, Pablo Sprechmann, Joshua Vogelstein, Pablo Musé, and Guillermo Sapiro. Robust multimodal graph matching: Sparse coding meets graph matching. In *Advances in Neural Information Processing Systems*, pages 127–135, 2013.

[25] Farhad Shirani, Siddharth Garg, and Elza Erkip. Seeded graph matching: Efficient algorithms and theoretical guarantees. In *2017 51st Asilomar Conference on Signals, Systems, and Computers*, pages 253–257. IEEE, 2017.

[26] E. Kazemi, L. Yartseva, and M. Grossglauser. When can two unlabeled networks be aligned under partial overlap? In *2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 33–42, Sept 2015.

[27] Elchanan Mossel and Jiaming Xu. Seeded graph matching via large neighborhood statistics. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1005–1014. SIAM, 2019.

[28] Donniell E Fishkind, Sancar Adali, Heather G Patsolic, Lingyao Meng, Digvijay Singh, Vince Lyzinski, and Carey E Priebe. Seeded graph matching. *Pattern Recognition*, 87:203–215, 2019.

[29] Vince Lyzinski and Daniel L Sussman. Matchability of heterogeneous networks pairs. *arXiv preprint arXiv:1705.02294*, 2017.

[30] Si Zhang and Hanghang Tong. Final: Fast attributed network alignment. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1345–1354. ACM, 2016.

[31] Mark Heimann, Haoming Shen, Tara Safavi, and Danai Koutra. Regal: Representation learning-based graph alignment. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, pages 117–126. ACM, 2018.

[32] I. Csiszár and J. Korner. *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Academic Press Inc. Ltd., 1981.

[33] I Martin Isaacs. *Algebra: a graduate course*, volume 100. American Mathematical Soc., 1994.

[34] F. Shirani, S. Garg, and E. Erkip. An information theoretic framework for active de-anonymization in social networks based on group memberships. In *2017 55rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Sept 2017.

[35] Xinjia Chen. Concentration inequalities for bounded random vectors. *arXiv preprint arXiv:1309.0003*, 2013.