Information Recovery from Pairwise Measurements: A Shannon-Theoretic Approach

Yuxin Chen
Statistics, Stanford University
Email: yxchen@stanford.edu

Changho Suh
EE, KAIST
Email: chsuh@kaist.ac.kr

Andrea J. Goldsmith
EE, Stanford
Email: andrea@wsl.stanford.edu

***Abstract*—This paper is concerned with jointly recovering n node-variables $\{x_1, \cdots, x_n\}$ from a collection of pairwise difference measurements. Specifically, several noisy measurements of $x_i - x_j$ are acquired. This is represented by a graph with an edge set $\mathcal{E}$ such that $x_i - x_j$ is observed only if $(i,j) \in \mathcal{E}$. To accommodate the noisy nature of data acquisition in a general way, we model the measurements by a set of channels with given input/output transition measures. Using information-theoretic tools applied to the channel decoding problem, we develop a *unified* framework to characterize a sufficient and a necessary condition for exact information recovery, which accommodates general graph structures, alphabet sizes, and channel transition measures. In particular, we isolate and highlight a family of *minimum distance measures* underlying the channel transition probabilities, which plays a central role in determining the recovery limits. For a broad class of homogeneous graphs, the recovery conditions we derive are tight up to some explicit constant, which depend only on the graph sparsity irrespective of other second-order graph metrics like the spectral gap.**

I. INTRODUCTION

In various data processing scenarios, one wishes to acquire information about a large collection of objects, but it is infeasible or difficult to directly measure each individual object in isolation. Instead, only certain pairwise relations over a few object pairs can be measured. Partial examples of pairwise relations include relative rotation, pairwise matches, and cluster agreements, as will be discussed later. Fortunately, such pairwise observations often carry a significant amount of information across all objects. As a result, reliable joint information recovery becomes possible as soon as sufficiently many pairwise measurements can be obtained.

This paper explores an important class of pairwise measurements, termed *pairwise difference measurements*. Consider n variables $x_1, \cdots, x_n$, and suppose we obtain independent measurements of the differences $x_i - x_j$ over a few pairs (i,j). This pairwise difference functional can be represented by a graph $\mathcal{G}$ with an edge set $\mathcal{E}$ such that $x_i - x_j$ is observed if and only if $(i,j) \in \mathcal{E}$. To accommodate the noisy nature of data acquisition, we model the noisy measurements y_{ij}'s as the output of the following equivalent channel

$$x_i - x_j \quad \xrightarrow{p(y_{ij} | x_i - x_j)} \quad y_{ij}, \quad \forall (i,j) \in \mathcal{E}, \tag{1}$$

where $p(\cdot \mid \cdot)$ represents the channel transition probability. The distribution of the output y_{ij} is specified solely by the associated channel input $x_i - x_j$. The goal is to recover $\boldsymbol{x} = \{x_1, \cdots, x_n\}$ based on these channel outputs y_{ij}'s. Note that as long as $\mathcal{G}$ is connected, the ground truth $\boldsymbol{x}$ is uniquely determined by the pairwise difference functional $\{x_i - x_j \mid (i,j) \in \mathcal{E}\}$, up to some global offset. Therefore, information recovery is also identical to decoding the input of the equivalent channel (1) based on the y_{ij}'s.

Problems of this kind have received considerable attention across various fields like computer science and social networks. We list a few below; see [1] for a more detailed discussion.

- *Alignment and Synchronization*. Consider n views of a single scene from different positions. The goal is to align all views based on noisy estimates of the *relative* translation and rotation across several pairs of views [2], [3].
- *Joint Graph Matching*. Existing pairwise matching methods allow us to map common features between a pair of shapes. Given multiple shapes, a fundamental problem amounts to how to refine these noisy pairwise matches in order to obtain globally-consistent matches [4], [5].
- *Community Detection*. Various social networks exhibit community structures [6], [7], and the nodes are clustered based on shared features. The aim is to uncover the community structure by observing the similarities between nodes.

Many of these applications have witnessed a flurry of activity in algorithm development, which are primarily motivated by computational concerns. Several of these algorithms are shown to enjoy intriguing recovery guarantees under simple randomized models, although the choice of performance metrics is often studied in a model-specific manner. On the other hand, there have been several information theoretic results in place for a few applications including stochastic block models [7], [8] and synchronization [9]. Despite their intrinsic connections, these results were developed primarily on a case-by-case basis instead of accounting for the most general observation models.

In the present paper, we account for the similarities and connections among all these motivating applications, by viewing them as a graph-based functional fed into a collection of general channels. We wish to explore the following questions from an information theoretic perspective:

1) Are there any fundamental distance metrics of the channel transition measures that dictate the success of exact information recovery from pairwise difference measurements?
2) If so, can we characterize the fundamental limits – in terms of these channel distance metrics – such that perfect recovery is feasible only above the limits?

All in all, we aim to gain a *unified* understanding towards the fundamental performance limits that underlie most applications falling in the realm of information recovery from pairwise difference measurements. These fundamental limits will in turn provide a general benchmark for algorithm evaluation and comparison.

A. Main Contributions

The main contribution of this paper is towards a unified characterization of fundamental recovery limits, using both information-theoretic and graph-theoretic tools. In particular, we single out and highlight a family of minimum channel distance measures (i.e. the minimum KL and Hellinger divergence), as well as two graphical metrics (i.e. the minimum cut and the cut-homogeneity exponent), that play central roles in determining the feasibility of exact recovery. Based on these metrics, we develop a sufficient and a necessary condition for information recovery such that the minimax probability of error tends to zero. These results apply to general graphs, any type of input alphabets, and general channel transition measures. Encouragingly, as long as the alphabet size is not super-polynomial in n, these two conditions coincide (modulo some universal constant) for a very broad class of homogeneous graphs, subsuming as special cases Erdős–Rényi models, random geometric graphs, small-world graphs, etc.

Our results reveal that the fundamental limits, presented in terms of the minimum channel distance measures, scale inversely proportional to the minimum cut size. Somewhat surprisingly, for most homogeneous graphs, the fundamental limits depend solely on the vertex degrees, irrespective of other second-order metrics (e.g. the spectral gap) that underlie the sufficient recovery condition derived in some previous work (e.g. [10]).

The unified framework we develop is *non-asymptotic*, in the sense that it accommodates general settings without fixing either the alphabet size or the channel transition probabilities. This allows full characterization of the high-dimensional regime where all parameters are allowed to scale, which has received increasing attention compared to the classical asymptotics where only n is tending to infinity.

B. Related Work

The information-theoretic work most relevant to our results is [10], which characterized the fundamental recovery limit under the Erdős–Rényi model with binary alphabets. A sufficient recovery condition for general graphs has also been derived in [10] under binary alphabets, although it was not guaranteed to be order optimal. Our prior work [9] explored the fundamental recovery limits under general alphabets and graph structures, but was restricted to the simplistic outlier model rather than general channel distributions. In contrast, the fundamental limits derived in the current work allow orderwise tight characterization for general alphabets and channel characteristics.

The pairwise measurement models considered in this paper and [10] can both be treated as a special type of "graphical channel" as defined in [11], which refers to a family of channels whose transition probability factorizes over a set of hyper-edges. This previous work on graphical channels centered on the concentration of the conditional entropy.

Our results are presented in terms of the KL and Hellinger divergence of order α; see [12], [13] for an introduction of these measures. The fundamental lower bounds rely on the generalized Fano inequality [14], [15] as well as an extended version involving more general divergence measures [16].

C. Terminology

Graph Terminology. For any vertex sets $\mathcal{S}_1$ and $\mathcal{S}_2$, we let $e(\mathcal{S}_1, \mathcal{S}_2)$ represent the number of edges with one endpoint in $\mathcal{S}_1$ and another in $\mathcal{S}_2$. An Erdős–Rényi graph of n vertices, denoted by $\mathcal{G}_{n,p_{\mathrm{obs}}}$, is constructed such that each pair of vertices is independently connected by an edge w.p. p_{obs}.

Divergence Measures. For any two measures P and Q, the KL divergence and the Hellinger divergence of order $\alpha \in (0,1)$ [12] between them are defined respectively as

$$\mathsf{KL}\left(P \,\|\, Q\right) := \int \mathrm{d}P \log\left(\frac{\mathrm{d}P}{\mathrm{d}Q}\right); \tag{2}$$

$$\mathsf{Hel}_{\alpha}\left(P \,\|\, Q\right) := \frac{1}{1-\alpha}\left[1 - \int \left(\mathrm{d}P\right)^{\alpha}\left(\mathrm{d}Q\right)^{1-\alpha}\right]. \tag{3}$$

When $\alpha = 1/2$, this reduces to the *squared Hellinger distance*.

II. Problem Formulation

Consider n vertices $\mathcal{V} = \{1, \cdots, n\}$, each represented by a variable x_i over the input alphabet $\mathcal{X} := \{0, 1 \cdots, M-1\}$, where M represents the alphabet size. Let $\boldsymbol{x} := \{x_i\}_{1 \le i \le n}$.

- **Object Representation and Pairwise Difference.** Consider an additive group formed over $\mathcal{X}$ together with an associative addition operation "+". For any $x_i, x_j \in \mathcal{X}$, the pairwise difference operation is defined as

$$x_i - x_j \quad := \quad x_i + (-x_j), \tag{4}$$

  where $-x_j$ stands for the unique additive inverse of x_j. We assume throughout that "+" satisfies the bijective property:

$$\forall x_i \in \mathcal{X}: \quad \begin{cases} x_i + x_j \neq x_i + x_l, & \forall x_l \neq x_j; \\ x_i + x_j \neq x_l + x_j, & \forall x_l \neq x_i. \end{cases} \tag{5}$$

  Partial examples include modular subtraction, relative rotation, and pairwise matching, as discussed in [1].

- **Measurement Graph and Channel**. The pairwise difference measurements are represented by a *measurement graph* $\mathcal{G}$ that comprises an edge set $\mathcal{E}$. Specifically, for each $(i,j) \in \mathcal{E}$ $(i > j)$, the pairwise difference $x_i - x_j$ is independently passed through a noisy channel, whose output y_{ij} follows the distribution such that for any $0 \le l < M$,

$$y_{ij} \;\Big|\; \{x_i - x_j = l\} \quad \sim \quad \mathbb{P}_l, \tag{6}$$

  where $\mathbb{P}_l$ denotes the transition measure that maps a given input l to an output $y_{ij} \in \mathcal{Y}$. The output alphabet $\mathcal{Y}$ can either be continuous or discrete. In contrast to conventional information theory settings, no coding is employed across channel uses.

This paper centers on the fundamental limits for exact information recovery. Note, however, that each input $\boldsymbol{x}$ is unique only up to a *global offset*, because the inputs $\boldsymbol{x}$ and $\boldsymbol{x}+l\cdot 1$ result in an identical output $\boldsymbol{y} := \{y_{ij} \in \mathcal{Y} \mid (i,j) \in \mathcal{E}\}$

from which $\boldsymbol{x}$ and its shifted versions are indistinguishable. Here 1 denotes an all-one vector. In light of this, we introduce the distance modulo a global offset factor as follows

$$\mathrm{dist}\left(\boldsymbol{w},\boldsymbol{x}\right) \quad := \quad 1-\max_{0\leq l<M}\mathbb{I}\left\{\boldsymbol{w}=\boldsymbol{x}+l\cdot 1\right\}. \qquad (7)$$

Equipped with this distance metric, we define, for any recovery procedure $\psi:\ \mathcal{Y}^{|\mathcal{E}|}\mapsto\mathcal{X}^n$, the *probability of error* as

$$P_{\mathsf{e}}\left(\psi\right) \quad := \quad \max_{\boldsymbol{x}\in\mathcal{X}^n}\mathbb{P}\left\{\mathrm{dist}\left(\psi\left(\boldsymbol{y}\right),\boldsymbol{x}\right)\neq 0 \;\middle|\; \boldsymbol{x}\right\}. \qquad (8)$$

We aim to characterize the regime where the minimax probability of error $\inf_{\psi}P_{\mathsf{e}}\left(\psi\right)$ tends to 0.

III. Key Metrics

A. Key Metrics on Probability Distance

There are several information divergence measures that play central roles in the subsequent development of our theory. Specifically, we define the minimum Hellinger and KL divergence with respect to the channel transition measures as

$$\mathsf{Hel}_{\alpha}^{\min}:=\min_{0\leq l,i<M;\; l\neq i}\mathsf{Hel}_{\alpha}\left(\mathbb{P}_l\,\|\,\mathbb{P}_i\right);$$
$$\mathsf{KL}^{\min}:=\min_{0\leq l,i<M;\; l\neq i}\mathsf{KL}\left(\mathbb{P}_l\,\|\,\mathbb{P}_i\right).$$

These measures capture the distinguishability of outputs given minimally separated inputs. Moreover, for any $\epsilon>0$, define

$$m^{\mathsf{kl}}\left(\epsilon\right):=\max_{l}\left|\left\{0\leq i<M\mid i\neq l,\ \mathsf{KL}\left(\mathbb{P}_i\,\|\,\mathbb{P}_l\right)\leq\left(1+\epsilon\right)\mathsf{KL}^{\min}\right\}\right|, \qquad (9)$$

which clearly satisfies $1\leq m^{\mathsf{kl}}<M$. This quantity reflects how many probability measure pairs under study are able to approach the minimum information divergence.

Finally, we note that the KL divergence and squared Hellinger distance are almost equivalent when two probability measures are close – a regime where two measures are the hardest to distinguish, as recorded in the following fact.

Fact 1. *Suppose that P and Q are two different measures obeying $\frac{\mathrm{d}P}{\mathrm{d}Q}\leq R$ and $\frac{\mathrm{d}Q}{\mathrm{d}P}\leq R$, then one has*

$$\max\left\{2-0.5\log R,\ 1\right\}\leq\frac{\mathsf{KL}\left(P\,\|\,Q\right)}{\mathsf{Hel}_{\frac{1}{2}}\left(P\,\|\,Q\right)}\leq 2+\log R. \qquad (10)$$

B. Key Graphical Metrics

First of all, we denote by mincut, $d_{\min}$, and $d_{\max}$ the size of the minimum cut, the minimum vertex degree, and the maximum vertex degree, respectively. Apart from these common metrics, there are a few other graphical quantities that are crucial in presenting our results, as detailed below.

For any integer m, define

$$\mathcal{N}\left(m\right) \quad := \quad \left\{\mathcal{S}\subset\mathcal{V}:\ e\left(\mathcal{S},\mathcal{S}^{\mathrm{c}}\right)\leq m\right\}, \qquad (11)$$

which consists of the collection of cuts of about the same size. We are particularly interested in the peak growth rate of the cardinality of $\mathcal{N}$ as defined below

$$\tau_k^{\mathrm{cut}}:=\frac{1}{k}\log\left|\mathcal{N}\left(k\cdot\mathsf{mincut}\right)\right|;\quad \tau^{\mathrm{cut}}:=\max_{k>0}\tau_k^{\mathrm{cut}}. \qquad (12)$$

In the sequel, we will term τ^{cut} the ***cut-homogeneity exponent***, as illustrated through the following two examples:

- *Complete graph K_n on n vertices*, where $\mathsf{mincut}=n-1$ and $e\left(\mathcal{S},\mathcal{S}^{\mathrm{c}}\right)=\left|\mathcal{S}\right|\left(n-\left|\mathcal{S}\right|\right)$. A simple combinatorial argument suggests that $\left|\mathcal{N}\left(k\right)\right|\asymp n^{\frac{k}{n}}$ and thus $\tau^{\mathrm{cut}}\asymp\log n$.
- *2 complete graphs $K_{n/2}$ connected by a single bridge*, where $\mathsf{mincut}=1$ due to the existence of a bridge. Since we still have $\left|\mathcal{N}\left(k\right)\right|\lesssim n^{\frac{k}{n}}$ $(k\geq n)$, one has $\tau_k^{\mathrm{cut}}\lesssim\frac{\log n}{n}$.

In many homogeneous graphs (e.g. complete graphs) one has $\mathsf{mincut}\asymp d_{\min}$, and τ^{cut} can be on the order of $\log n$. In contrast, $\mathsf{mincut}=o\left(d_{\min}\right)$ often holds in inhomogeneous graphs, for which τ^{cut} tends to be lower. In light of this, τ^{cut} captures the extent of homogeneity for cut-size distributions.

Interestingly, for various homogeneous graphs (including random geometric graphs, Erdős–Rényi graphs, and other expander graphs with good expansion), one always has [1]

$$\tau^{\mathrm{cut}}\lesssim\log n. \qquad (13)$$

IV. Main Results: General Graphs

Intuitively, faithful decoding is feasible only when (i) $\mathcal{G}$ is sufficiently connected so that we have enough measurements involving each vertex variable, and (ii) the channel output distributions given two different inputs are sufficiently separated. This section formalizes this intuition by developing both sufficient and necessary recovery conditions in terms of the information divergence and graphical metrics, which accommodate general graphs, channel characteristics, and input alphabets. Derivations of all results can be found in [1].

A. Maximum Likelihood Decoding: General Graphs

We start by analyzing the performance of the maximum likelihood (ML) decoder $\psi_{\mathrm{ml}}\left(\boldsymbol{y}\right)$, which seeks a solution that maximizes the likelihood function. Based on ML decoding, we develop a sufficient recovery condition in terms of the minimum Hellinger divergence. The condition is universal and non-asymptotic. It holds for all graph structures, and depends only on the min-cut size and the cut-homogeneity exponent irrespective of other graphical metrics.

Theorem 1. *For any $\delta>0$ and $0<\alpha<1$, the ML rule achieves*

$$P_{\mathrm{e}}\left(\psi_{\mathrm{ml}}\right)\leq\left(\left(2n\right)^{\delta}-1\right)^{-1},$$

provided that

$$\mathsf{Hel}_{\alpha}^{\min}\geq\frac{8\tau^{\mathrm{cut}}+\left(\delta+8\right)\log\left(2n\right)+4\log M}{\left(1-\alpha\right)\mathsf{mincut}}. \qquad (14)$$

Theorem 1 characterizes a non-asymptotic region of $\mathsf{Hel}_{\alpha}^{\min}$ where $\psi_{\mathrm{ml}}(\cdot)$ admits perfect information recovery. The boundary of the region scales as

$$\Theta\left(\frac{\tau^{\mathrm{cut}}+\log n+\log M}{\mathsf{mincut}}\right). \qquad (15)$$

This scaling reveals that the contribution of τ^{cut} will be negligible if it is below $o\left(\log n\right)$. In general, we are unable to develop a general sufficient recovery condition in terms of the KL divergence, since the KL divergence cannot be controlled for all measures, especially when $\frac{\mathrm{d}\mathbb{P}_l}{\mathrm{d}\mathbb{P}_j}$ $(l\neq j)$ becomes unbounded. In contrast, the Hellinger distance is generally stable and more convenient to analyze.

As will be seen later, this result is most useful when applied to those homogeneous graphs satisfying $\mathsf{mincut} \approx d_{\min} \approx d_{\max}$, for which the fundamental limits decay inversely with the vertex degree. To gain some insights about the metrics, we emphasize that the channel decoding model considered herein differs from the classical information theory setting in that the input is "uncoded". Consequently, the minimum distance between the set of hypotheses (rather than the mutual information in an average sense) presents the fundamental bottleneck for information recovery. Notably, two hypotheses $\boldsymbol{x}$ and $\boldsymbol{w}$ are minimally separated when they differ only by one component v. The resulting pairwise inputs $\boldsymbol{y}$ contain about $\deg(v)$ copies of information for distinguishing $\boldsymbol{x}$ and $\boldsymbol{w}$, where each copy of information can be measured by the information divergence metrics $\mathsf{KL}^{\min}$ or $\mathsf{Hel}_{\alpha}^{\min}$. This offers an intuitive interpretation as to why the fundamental limit scales inversely with the vertex degree.

B. Lower Bounds: General Graphs

Encouragingly, for a broad class of homogeneous graphs, the sufficient recovery condition for the ML decoder matches the fundamental lower limit in an order-wise tight sense, as revealed in the following theorem.

Theorem 2. *Fix $0 < \epsilon \leq \frac{1}{2}$. If the KL divergence satisfies*

$$\begin{aligned}\mathsf{KL}^{\min} \quad \leq \quad & \max\left\{\frac{(1-\epsilon)\max\left\{\tau^{\mathrm{cut}}, \log m^{\mathsf{kl}}\right\} - H(\epsilon)}{\mathsf{mincut}},\right. \\ & \left.\frac{(1-\epsilon)\left(\log n + \log m^{\mathsf{kl}}\right) - H(\epsilon)}{(1+\epsilon)\, d_{\max}}\right\}, \qquad (16)\end{aligned}$$

then $\inf_{\psi} P_{\mathsf{e}}(\psi) \geq \epsilon$. Here, $H(x)$ is the entropy function.

Theorem 2 presents a general fundamental lower limit on $\mathsf{KL}^{\min}$ that allows perfect recovery, which scales as

$$\Theta\left(\frac{\tau^{\mathrm{cut}} + \log m^{\mathsf{kl}}}{\mathsf{mincut}} + \frac{\log n}{d_{\max}}\right).$$

The lower bound is derived in terms of KL divergence, which is almost identical to the Hellinger distance when the transition measures of the graphical channels are close to each other.

C. Implications

The following discussion concentrates on the case where $\frac{\mathrm{d}\mathbb{P}_l}{\mathrm{d}\mathbb{P}_j} = 1 + o(1)$ for all l and j, which essentially corresponds to the most difficult situation for decoding. Recall from Fact 1 that $\mathsf{KL}^{\min} = 2(1+o(1))\,\mathsf{Hel}_{1/2}^{\min}$. Additionally, we suppose that the alphabet size M is *not* super-polynomial in n.

- **Tightness**. Theorems 1-2 characterize the admissible and inadmissible regions such that as $n \to \infty$,

$$\begin{aligned}&\inf_{\psi} P_{\mathsf{e}} \longrightarrow 0 \quad \text{if} \\ &\qquad \mathsf{Hel}_{\frac{1}{2}}^{\min} > \frac{16\tau^{\mathrm{cut}} + 16\log n + 8\log M}{\mathsf{mincut}}, \\ &\inf_{\psi} P_{\mathsf{e}} \not\longrightarrow 0 \quad \text{if} \\ &\qquad \mathsf{Hel}_{\frac{1}{2}}^{\min} < \frac{1}{2}\max\left\{\frac{\tau^{\mathrm{cut}}}{\mathsf{mincut}}, \frac{\log m^{\mathsf{kl}}}{\mathsf{mincut}}, \frac{\log\left(n m^{\mathsf{kl}}\right)}{d_{\max}}\right\},\end{aligned}$$

  where we ignore $o(1)$ terms. One can then verify that the above recovery conditions are tight up to a factor

$$\mathcal{O}\left(\max\left\{\frac{\log M}{\log m^{\mathsf{kl}}}, \frac{d_{\max}}{\mathsf{mincut}}\right\}\right).$$

  When specialized to the homogeneous graphs satisfying $d_{\max} \asymp \mathsf{mincut}$ (e.g. Erdős–Rényi Graphs, random geometric graphs), the recovery conditions are within a factor of

$$\mathcal{O}\left(\frac{\log M + \log n}{\log m^{\mathsf{kl}} + \log n}\right)$$

  from optimal. Consequently, our results are orderwise tight.

- **First-order v.s. second-order graphical metrics.** For a broad class of homogeneous graphs (e.g. Erdős–Rényi graph, random geometric graphs), one has $\mathsf{mincut} \asymp d_{\max}$ and $\tau^{\mathrm{cut}} \lesssim \log n$ (see [1, Lemma 1]). As a result, the boundary recovery condition for these graphs reduces to

$$\mathsf{Hel}_{\frac{1}{2}}^{\min} = \Theta\left(\log n / d_{\max}\right) \qquad (17)$$

  as long as $M \lesssim \mathrm{poly}(n)$, which depends only on the vertex degree and does not rely on those second-order expansion properties[1] like the Cheeger's constant as given in [10, Theorem 4.3]. This leads to the following observation: for various homogeneous graphs, the fundamental recovery limits are determined solely by *first-order* graphical features (i.e. vertex degrees). This is in stark contrast to the sufficient condition for several classes of tractable algorithms (e.g. spectral methods), whose success typically relies on strong *second-order* expansion properties.

- **A Unified Non-asymptotic Framework.** Our framework accommodates various practical scenarios that respect the high-dimensional regime, where both the number n of objects and the alphabet size M tend to be large. Our problem falls under the category of multi-hypothesis testing in the presence of exponentially many hypotheses, where each hypothesis is not necessarily formed by i.i.d. sequences. Under such a setting, the conventional Sanov bound [17] becomes unwieldy and the Chernoff information is not guaranteed to capture the minimax rate. By comparison, our results build upon alternative divergence measures (particularly the Hellinger-α divergence). This results in a unified framework that allows non-asymptotic characterization of the minimax rate simultaneously for most general settings.

D. Special Case: Erdős–Rényi Graphs

We now turn to the Erdős–Rényi model, the simplest and the most widely adopted random graph model. Specifically, we suppose that the measurement graph $\mathcal{G}$ is drawn from $\mathcal{G}_{n,p_{\mathrm{obs}}}$ for some $p_{\mathrm{obs}} \gtrsim \log n / n$. While our preceding results are sufficient to characterize the fundamental recovery limit in an order-wise optimal sense, we refine our analysis for this special model to obtain tighter universal preconstants. These are stated in the following two theorems.

[1] For instance, random geometric graphs typically have much worse expansion properties than Erdős–Rényi graphs.

Theorem 3. *Fix $\delta > 0$ and $0 < \alpha < 1$. For any constant $\zeta > 0$, there exist some universal constants C, $c_0, c_1 > 0$ such that if $p_{\mathrm{obs}} \geq c_0 \log n / n$ and*

$$\mathsf{Hel}_{\alpha}^{\min} \quad \geq \quad \frac{(1+\delta)\log(2n) + 2\log(M-1)}{(1-\zeta)(1-\alpha)\, p_{\mathrm{obs}} n}, \quad (18)$$

then the ML decoder achieves

$$P_{\mathrm{e}}(\psi_{\mathrm{ml}}) \leq \frac{1}{(2n)^{\max\left\{\frac{3}{4}\delta - \frac{1}{4}\delta^2,\, \frac{\delta-1}{2}\right\}} - 1} + \frac{3}{n^{10}} + Cn^{-c_1 \delta n}.$$

Theorem 4. *Fix $\epsilon > 0$, and suppose that $p_{\mathrm{obs}} > c \log n / n$ for some sufficiently large constant $c > 0$.*

(a) One has $\inf_{\psi} P_{\mathsf{e}} \geq \epsilon - \frac{1}{n^{10}}$ in the regime where

$$\mathsf{KL}^{\min} \quad \leq \quad \frac{(1-\epsilon)\left(\log n + \log m^{\mathsf{kl}}\right) - \log 2}{(1+\epsilon)^2\, n p_{\mathrm{obs}}}, \quad (19)$$

(b) For any $\alpha \leq (1+\epsilon)^{-1}$ and any small constant $\zeta > 0$, one has $\inf_{\psi} P_{\mathsf{e}} \geq n^{-\epsilon} - \frac{1}{n^{10}}$, provided that

$$\mathsf{Hel}_{\alpha}^{\min} < \frac{\epsilon \alpha \log n}{(1+\zeta)(1-\alpha)\, n p_{\mathrm{obs}}} - r_{\epsilon} \quad (20)$$

for some residual $r_{\epsilon} \lesssim (n p_{\mathrm{obs}})^{-1}$.

These two theorems tighten the preconstants for both sufficient and necessary recovery conditions, as discussed below.

1) Consider the first-order convergence, that is, the regime where $\inf_{\psi} P_{\mathrm{e}} \to 0$ as $n \to \infty$. Combining Theorems 3 and 4(a) suggests that as $n \to \infty$,

$$\inf_{\psi} P_{\mathsf{e}}(\psi) \quad \longrightarrow \quad 0 \quad \text{if } \mathsf{Hel}_{\frac{1}{2}}^{\min} > \frac{2\log n + 4\log M}{p_{\mathrm{obs}} n};$$

$$\inf_{\psi} P_{\mathsf{e}}(\psi) \quad \not\longrightarrow \quad 0 \quad \text{if } \mathsf{KL}^{\min} < \frac{\log n + \log m^{\mathsf{kl}}}{p_{\mathrm{obs}} n}.$$

For the challenging case where $\frac{\mathrm{d}\mathbb{P}_l}{\mathrm{d}\mathbb{P}_j} = 1 + o(1)$ for all $l \neq j$, these taken collectively with Fact 1 reduce to

$$\inf_{\psi} P_{\mathsf{e}}(\psi) \quad \longrightarrow \quad 0 \quad \text{if } \mathsf{Hel}_{\frac{1}{2}}^{\min} > \frac{2\log n + 4\log M}{p_{\mathrm{obs}} n};$$

$$\inf_{\psi} P_{\mathsf{e}}(\psi) \quad \not\longrightarrow \quad 0 \quad \text{if } \mathsf{Hel}_{\frac{1}{2}}^{\min} < \frac{\log n + \log m^{\mathsf{kl}}}{2 p_{\mathrm{obs}} n}.$$

In general, our minimum distance bounds are tight to within a factor of

$$(1 + o(1))\, \frac{4\log n + 8\log M}{\log n + \log m^{\mathsf{kl}}}.$$

2) Consider the second-order convergence and, particularly, the regime where $\inf_{\psi} P_{\mathrm{e}} \to n^{-1}$ ($n \to \infty$). Putting Theorems 3 and 4(b) together implies that

$$\inf_{\psi} P_{\mathsf{e}} \quad > \quad \frac{1}{n} \quad \text{if } \mathsf{Hel}_{\frac{1}{2}}^{\min} < \log n / (p_{\mathrm{obs}} n),$$

$$\inf_{\psi} P_{\mathsf{e}} \quad < \quad \frac{1}{n} \quad \text{if } \mathsf{Hel}_{\frac{1}{2}}^{\min} > \frac{8\log n + 4\log M}{p_{\mathrm{obs}} n},$$

which holds for *all settings*. This implies that our results are within a multiplicative gap of

$$(1 + o(1))\left(8 + \frac{4\log M}{\log n}\right)$$

from optimal, uniformly over all possible alphabet sizes and channel transition measures.

V. Concluding Remarks

This paper investigates simultaneous recovery of multiple vertex-variables based on noisy graph-based measurements, under the pairwise difference model. The problem formulation spans numerous applications including community detection, image registration, and also computational biology [18], [19]. We develop a unified framework in understanding all problems of this kind based on representing the available pairwise measurements as a graph, and then representing the noise on the measurements using a general channel with a given input/output transition measure. This framework accommodates large alphabets, general channel transition probabilities, and general graph structures in a non-asymptotic manner. Our results underscore that the feasibility of information recovery is captured by the minimum information divergence measures. We expect that such fundamental limits will provide a general benchmark for evaluating the performance of practical algorithms over many applications.

Acknowledgments

Y. Chen would like to thank E. Abbe, A. Bandeira, A. Dembo, A. Singer, and J. Xu for inspiring discussion. This work is supported in part by the NSF CSoI, AFOSR under MURI Grant FA9550-12-1-0215, and ICT R&D program of MSIP/IITP [B0101-15-1272].

References

[1] Y. Chen, C. Suh, and A. J. Goldsmith, "Information recovery from pairwise measurements: A Shannon-theoretic approach," 2015. [Online]. Available: http://arxiv.org/abs/1504.01369

[2] L. Wang and A. Singer, "Exact and stable recovery of rotations for robust synchronization," *arXiv preprint arXiv:1211.2441*, 2012.

[3] A. Bandeira, M. Charikar, A. Singer, and A. Zhu, "Multireference alignment using semidefinite programming," *ITCS*, pp. 459–470, 2014.

[4] Q. Huang and L. Guibas, "Consistent shape maps via semidefinite programming," in *SGP*, vol. 32, no. 5, 2013, pp. 177–186.

[5] Y. Chen, L. Guibas, and Q. Huang, "Near-optimal joint object matching via convex relaxation," *International Conference on Machine Learning (ICML)*, pp. 100–108, June 2014.

[6] S. Fortunato, "Community detection in graphs," *Physics Reports*, 2010.

[7] E. Abbe, A. S. Bandeira, and G. Hall, "Exact recovery in the stochastic block model," *arXiv preprint arXiv:1405.3267*, 2014.

[8] E. Mossel, J. Neeman, and A. Sly, "Consistency thresholds for binary symmetric block models," *arXiv:1407.1591*, 2014.

[9] Y. Chen and A. J. Goldsmith, "Information recovery from pairwise measurements," *ISIT*, pp. 2012–2016, June 2014.

[10] E. Abbe, A. S. Bandeira, A. Bracher, and A. Singer, "Decoding binary node labels from censored edge measurements: Phase transition and efficient recovery," *arXiv preprint arXiv:1404.4749*, 2014.

[11] E. Abbe and A. Montanari, "Conditional random fields, planted constraint satisfaction and entropy concentration," in *RANDOM*, 2013, pp. 332–346.

[12] F. Liese and I. Vajda, "On divergences and informations in statistics and information theory," *IEEE Trans Inf. Theory*, vol. 52, no. 10, 2006.

[13] F. Topsoe, "Some inequalities for information divergence and related measures of discrimination," *IEEE Trans Inf. Theory*, vol. 46, 2000.

[14] A. B. Tsybakov and V. Zaiats, *Introduction to nonparametric estimation*. Springer, 2009, vol. 11.

[15] T. S. Han and S. Verdu, "Generalizing the Fano inequality," *IEEE Trans. on Inf. Theory*, vol. 40, no. 4, pp. 1247–1251, 1994.

[16] A. Guntuboyina, "Lower bounds for the minimax risk using-divergences, and applications," *IEEE Trans on Inf Theory*, vol. 57, no. 4, 2011.

[17] I. Csiszár, "Sanov property, generalized I-projection and a conditional limit theorem," *The Annals of Probability*, pp. 768–793, 1984.

[18] H. Si, H. Vikalo, and S. Vishwanath, "Haplotype assembly: An information theoretic view," in *ITW*. IEEE, 2014, pp. 182–186.

[19] G. M. Kamath, E. Şaşoğlu, and D. Tse, "Optimal haplotype assembly from high-throughput mate-pair reads," *arXiv:1502.01975*, 2015.