The Annals of Applied Probability
2020, Vol. 30, No. 3, 1066–1090
https://doi.org/10.1214/19-AAP1523
© Institute of Mathematical Statistics, 2020

AN INFORMATION-PERCOLATION BOUND FOR SPIN SYNCHRONIZATION ON GENERAL GRAPHS

BY EMMANUEL ABBE[1] AND ENRIC BOIX-ADSERÀ[2]

[1]*Electrical Engineering and Applied and Computational Mathematics, Princeton University, emmanuel.abbe@epfl.edu*
[2]*Department of Mathematics, Princeton University, eboix@mit.edu*

This paper considers the problem of reconstructing n independent uniform spins $X_1, \dots, X_n$ living on the vertices of an n-vertex graph G, by observing their interactions on the edges of the graph. This captures instances of models such as (i) broadcasting on trees, (ii) block models, (iii) synchronization on grids, (iv) spiked Wigner models. The paper gives an upper bound on the mutual information between two vertices in terms of a bond percolation estimate. Namely, the information between two vertices' spins is bounded by the probability that these vertices are connected when edges are opened with a probability that "emulates" the edge-information. Both the information and the open-probability are based on the Chi-squared mutual information. The main results allow us to re-derive known results for information-theoretic nonreconstruction in models (i)–(iv), with more direct or improved bounds in some cases, and to obtain new results, such as for a spiked Wigner model on grids. The main result also implies a new subadditivity property for the Chi-squared mutual information for symmetric channels and general graphs, extending the subadditivity property obtained by Evans–Kenyon–Peres–Schulman (*Ann. Appl. Probab.* **10** (2000) 410–433) for trees. Some cases of nonsymmetrical channels are also discussed.

1. Introduction.

The model. We consider the problem of reconstructing n uniform spins $X_1, \dots, X_n \overset{\mathrm{IID}}{\sim} \mathrm{Rad}(1/2)$ living on the vertices of an n-vertex graph G, by observing their interactions on the edges of the graph. Formal definitions are in Section 2. Depending on the choices of the graph and the interaction channel, this captures models such as (i) broadcasting on trees [17, 22] (ii) censored block models [2, 20], (iii) synchronization on grids [3], (iv) spiked Wigner models [13]. Here we refer to these as synchronization problems on different graph/channel models.

To set a running example, consider the case where $G = K_n$ is the complete graph, and where the channel on each edge is BSC_p: a binary symmetric channel with flip probability $p \in [0, 1]$. In other words, for each $1 \le u < v \le n$, we observe the product $Y_{uv} = X_u X_v Z_{uv}$, where $\{Z_{uv}\}_{1 \le u < j \le n}$ are i.i.d. $\mathrm{Rad}(p)$, mutually independent of $\{X_u\}_{u \in [n]}$.

Note that the above model is also related to the Ising model in statistical physics; conditioned on the edge observations, the posterior distribution of the vertex spins is given by an Ising model. However, we will be interested here in the average-case behavior over the edge variables in the model, while results on Ising models (e.g., Dobrushin conditions for correlation decay [14]) typically focus on worst-case behavior over the edge variables.

The problem. Depending on how "rich" the graph is, and how "noisy" the channel is, one may or may not be able to obtain a nontrivial reconstruction of the spins. We focus here on

Received July 2018; revised August 2019.
MSC2010 subject classifications. Primary 62B10, 60K35; secondary 94A12.
Key words and phrases. Synchronization, community detection, spiked Wigner models, stochastic block models, bond percolation, information theory.

understanding when it is information-theoretically impossible to obtain a nontrivial reconstruction. For this purpose, we are interested in conditions for which the mutual information between the spins X_u, X_v of two arbitrary vertices $u, v \in [n]$, given all the edge interaction variables $Y_{E(G)}$, is vanishing as n diverges:

$$I_{\mathrm{KL}}(X_u; X_v \mid Y_{E(G)}) = o(1). \tag{1}$$

For the models mentioned above, this implies in particular that there is no estimator of the spins that solves the so-called weak recovery problem, that is, that gives an asymptotic correlation with the ground truth that is nontrivial; see [1] for discussions on weak recovery.

For the running example, if p is bounded away from $1/2$, then for any pair of vertices, the information on their direct edge suffices to prevent (1) from taking place. If p tends to $1/2$ fast enough, this may break down, but it is not enough to inspect the direct edge as the information may propagate along other paths in the graph.

Known techniques. Different techniques have been developed to upper-bound quantities such as the mutual information of (1). In particular,

- (1) *Upgrading the graph.* This approach was developed for instance for the broadcasting on trees (BOT) problem in [17]. In the BOT model, a random variable is broadcast from the root down the edges of a tree, with each edge potentially flipping the variable, and the goal is to reconstruct the root variable from the leaf variables at infinite depth. See Section 4.1 for formal definitions. One can view this as a synchronization problem using an extra vertex that interacts noiselessly with all the leaf variables; see Section 4.1 for the formal connection. To upper-bound the mutual information (corresponding to (1)) from the root to the leaves in the case of binary variables and symmetric channels, [17] shows a *subadditivity* property of the mutual information over all paths from the root to the leaves, which implies the impossibility part (the “difficult” part) of the Kesten–Stigum threshold. This subadditivity is a crucial component to establish the uniqueness of a threshold in this context, and is proved in [17] using an upgradation of the BOT ensemble on an arbitrary tree to a BOT ensemble on a “stringy” tree, where the branches of the tree are “separated”. One of the open problems/directions mentioned in [17] is to extend such results to more general graphs that contain cycles, finding the right model. Part of the results in this paper can be viewed as such an attempt.
- (2) *Using an oracle to change the graph.* This approach was developed for instance for the stochastic block model in [24]. We consider here the close variant called the censored block model (CBM). Take an Erdős–Rényi random graph in the sparse regime, $G_n \sim G(n, c/n)$, and on edge of the graph, observe the product of the adjacent spins on an independent BSC_ε (as in the running example). This gives an instance of the CBM. It models scenarios where one observes a random measurement that gives positive or negative indication that the two incident “people” are in the same community or not. The model is closely related[1] to the SBM$(n, a/n, b/n)$ (with the parameters $c = (a + b)/2$ and $\varepsilon = b/(a + b)$), where each vertex in the graph is connected by an edge with probability a/n if the adjacent vertices are in the same community, and b/n otherwise. To show that it is not possible to reconstruct the communities in the SBM, [24] upper-bounds (1) with an oracle that reveals the labels of the vertices at small depth from vertex u. Using then the fact that the Erdős–Rényi model is locally tree-like, [24] reduces the problem to the BOT model discussed previously. The

[1] In the SBM, the presence of an edge makes the two incident vertices be in different communities with probability $\varepsilon = b/(a+b)$, and each vertex has an expected number of $(a+b)/2 = c$ neighbors; the difference between the SBM and the CBM is that a nonedge in the SBM carries a slight repulsion probability towards having the incident vertices in the same community, although the latter is negligible in various aspects.

same proof technique applies to the CBM, as also obtained in [23]. Note that this proof technique is particularly helpful in the CBM/SBM because the local neighborhood of a vertex is "simpler," that is, tree-like, allowing us to reduce the model from a loopy graph to known results for trees [17]. Such an approach may not help in the model discussed next.

- (3) *Upgrading the channel.* This approach was used for instance for the synchronization problem on grids in [3]. Consider the case of BSC channels as in the running example, that flip the spins' product with probability p on each edge, and upgrade each channel with an erasure channel that instead erases the product with probability $2p$, revealing otherwise the exact value. This erasure model is clearly an upgradation of the BSC model, since one can always draw a random spin in replacement to an erasure symbol, which gives a BSC of flip probability $2p/2 = p$. As further discussed below, for an erasure model, the mutual information in (1) becomes exactly the probability that u and v are connected in a bond percolation model. In graph models like the grid, this has either a sharp threshold or some known bounds [18], and the overall approach gives a bound for synchronization problems on grids, developed in [3] beyond the case of BSCs. Note however that this approach is unlikely to give a sharp bound, due to the upgradation, but it allows for a direct application of percolation bounds.
- (4) *Interpolation, message-passing and second-moments.* Interpolation techniques take different forms; one consists in establishing a bound between two quantities by parametrizing each quantity with a relevant parameter, typically a notion of signal-to-noise ratio (SNR), establishing the bound for the boundary cases, and interpolating other cases with a "monotonicity" argument (inspecting a derivative). This approach has long been used in different contexts; for example, to establish the "entropy power inequality" in [30]. More closely related to us, it is used in [4] to establish a subadditivity property of the mutual information of graphical channels, where the subadditivity acts on the vertex-set rather than the edge-set as considered here. For the spiked Wigner model with Rademacher inputs, which corresponds to a complete graph with a Gaussian noise channel, one can use the I-MMSE formula from [19] to equate the derivative of the mutual information in (1) to the MMSE, and express the latter using an approximate message passing (AMP) estimate [15]. This allows [13] to establish a limiting expression for the mutual information, and in particular, a tight condition for when the latter vanishes. Similar techniques have been used in various other spiked Wigner models, such as in [8, 25, 26], and block models [10]. It is worth noting that if the goal is to only obtain a condition for when the mutual information vanishes, it may not be necessary to employ such elaborate estimates. In particular, one may rely on second-moment estimates as used in [6, 7] for block models and [8, 26] for Spiked Wigner models. Second-moment estimates typically give conditions on when the distribution of the planted ensemble (where the edge variables depend on the X_i variables) is contiguous to the unplanted ensemble (where the edge variables are independent), and depending on the model, this can be turned into a condition for weak recovery being not solvable, such as in [6–8, 26] (although the implication may not be true in general).

This paper. As apparent in previous discussion, some of the known techniques are fairly graph- and channel-dependent. The goal of this paper is to introduce a general method to upper-bound the mutual information (1) in terms of bond percolation estimates, namely, in terms of the probability that vertices u and v are connected by an "open" path in a model where each edge of G is kept open with some probability.

Note that if the channel on each edge is an erasure channel, that is, if $Y_{uv} = X_u X_v$ with probability q and $Y_{uv} = \star$ (an erasure symbol) with probability $1 - q$, then

$$I_{\mathrm{KL}}(X_u; X_v \mid Y_{E(G)}) = \mathrm{conn}_{G,q}(u, v), \tag{2}$$

where $\mathrm{conn}_{G,q}(u,v)$ denotes the probability that u and v are connected in a bond percolation model on G where each edge is open independently with probability q.

Our main result shows how to turn the previous equality into an inequality beyond the case of erasures, covering a fairly general family of channels that contains models (i)–(iv). The crucial part is to find how to set the openness probabilities on each edge in order to "emulate" the right amount of information, rather than using a degradation argument as discussed in (3) above, that produces loose bounds on models (i)–(iv). For this purpose, we will use an interpolation technique. In a sense, our bound can thus be viewed as an hybrid between the techniques of [4] and [3], as it uses an interpolation technique for a percolation bound.

The main feature of the bound is that it applies to *any* graph. The derived bound subsumes the known results for (1)–(4) (with slight improvements for (3)) and gives also a few new results. These are presented in Section 4. Discussions on how the bound could be extended beyond the binary setting are provided in Section 5. We underline here two aspects of the main results:

- *A Chi-squared bound.* A natural attempt to estimate the information between two vertices in terms of the probability that these vertices are connected in a bond percolation model, is to open each edge with a probability that "emulates" the information of the edge. How should this be formalized?

  Consider the case of $G = \Pi_n$, a path on the vertices $1, 2, \dots, n$, with a binary symmetric channel (BSC) of flip probability $p = (1-\delta)/2$ on each edge as in the running example. The channel between the first and last vertex (1 and n) is a concatenation of BSCs, each with a flip probability either p or $1-p$ (depending of the value of $Y_{i,i+1}$ for edge $(i, i+1)$). Thus we can explicitly compute the LHS of (1):

  $$I_{\mathrm{KL}}(X_1; X_n \mid Y_{E(\Pi_n)}) = 1 - H\big((1-\delta^{n-1})/2\big). \tag{3}$$

  On the other hand, if we open each edge in the path with probability equal to the mutual information of a BSC_p (or $\mathsf{BSC}_{(1-p)}$), that is, with $q = 1 - H((1-\delta)/2)$, vertex u and v are connected with probability

  $$\mathrm{conn}_{\Pi_n,q}(1,n) = \big(1 - H\big((1-\delta)/2\big)\big)^{n-1}. \tag{4}$$

  Unfortunately, the bound

  $$I_{\mathrm{KL}}(X_1; X_n \mid Y_{E(\Pi_n)}) \geq \mathrm{conn}_{\Pi_n,q}(1,n)$$

  that this gives is in the reverse direction of (1)! Note also that one cannot hope for a general bound in this reverse direction for the mutual information (e.g., one can get a counterexample on a triangle graph[2]).

  In order to obtain a bound that holds for arbitrary finite graphs, we will change our measure of information, using not the KL-divergence but the Chi-squared divergence, that is,

  $$I_2(X;Y) \equiv D_{\chi^2}\big(p_{X,Y} \parallel p_X p_Y\big), \tag{5}$$

  where D_{χ^2} is the Chi-squared f-divergence with $f(t) = (t-1)^2$. In particular, it is easily shown that that for the path example,

  $$I_2(X_1; X_n \mid Y_{E(\Pi_n)}) = \delta^{2(n-1)}. \tag{6}$$

[2] Consider the graph G with two vertices u, v and two parallel edges e, f equipped with BSC_p channels. Then the mutual informations for the individual edges are $q = 1 - H(p)$, so $\mathrm{conn}_{G,q}(u,v) = 2(1-H(p)) - (1-H(p))^2$, but $I_{\mathrm{KL}}(X_u; X_v|Y_e, Y_f) = (p^2 + (1-p)^2)(1 - H(p^2/(p^2+(1-p)^2)))$. One may verify that $I_{\mathrm{KL}}(X_u; X_v|Y_e, Y_f) \leq \mathrm{conn}_{G,q}(u,v)$. To avoid parallel edges and get a triangle graph, simply replace the edge f with two concatenated edges f_1 and f_2, where f_1 has a noiseless channel and f_2 has a BSC_p channel.

Therefore, opening edge with probability equal to the Chi-squared mutual information of a $\mathsf{BSC}_{((1-\delta)/2)}$, that is, δ^2, gives the desired upper bound with equality.

In general, we obtain that for any graph G and for a class of symmetric channels on the edges,

$$I_2(X_u; X_v \mid Y_{E(G)}) \le \mathrm{conn}_{G, I_2(X_e;Y_e)}(u, v) \quad \textit{(main result)}\,, \tag{7}$$

where the RHS is the probability that u and v are connected in a bond percolation model on G where each edge $e \in E(G)$ is open with probability $I_2(X_e; Y_e)$, where X_e denotes the product, $X_i \cdot X_j$, of the spins incident to edge $e = (i, j)$.

Further, one can obtain an upper bound when the LHS is the classical mutual information, since the classical mutual information is upper-bounded by the Chi-squared mutual information (for uniform binary variables); it is however important to keep the Chi-squared mutual information on the RHS. (See Lemma A.9.)

- *Subadditivity for general graphs.* Note that the RHS of (7) can be upper-bounded with the union bound over all paths between u and v, and using (6), we obtain as a corollary the following subadditivity property for general graphs:

$$I_2(X_u; X_v \mid Y_{E(G)}) \le \sum_{P \in \mathcal{P}_G(u,v)} I_2(X_u; X_v \mid Y_{E(P)}), \tag{8}$$

where $\mathcal{P}_G(u, v)$ denotes the set of paths (i.e., self-avoiding walks) from u to v in G. This gives an extension via the synchronization model of the subadditivity obtained for trees in [17] (see point (1) above) to general graphs.

A recent concurrent work of Polyanskiy and Wu [27] also gives an information-percolation bound. We discuss the relationship between our results in detail in Section 3.5.

2. Model.

We begin by defining a "graphical channel" similarly to the definition in [4], but tailored to the case in which vertex labels are binary:

- Let $g = (V, E(g))$ be a finite graph with vertex set $V = [n]$ and edge set $E(g)$.
- For each $e \in E(g)$, let $Q_e(\cdot \mid \cdot)$ be a probability transition function (channel) from the binary input alphabet $\{-1, +1\}$ to an output alphabet $\mathcal{Y}_e$, such that $Q_{e|+}(\cdot) \equiv Q_e(\cdot \mid +1)$ and $Q_{e|-}(\cdot) \equiv Q_e(\cdot \mid -1)$ are probability measures on a measurable space $(\mathcal{Y}_e, \mathcal{A}_e)$.
- Assign a vertex label $x_i \in \{-1, +1\}$ to each vertex $i \in V$. Assign an edge label $y_e \in \mathcal{Y}_e$ to each edge $e \in E(g)$. Then define the channel $P_{g,Q}(\cdot \mid \cdot)$ with input alphabet $\{-1, +1\}^V$ and output alphabet $\mathcal{Y}^{E(g)}$ as follows: for each measurable set $A = \prod_{e \in E(g)} A_e \in \prod_{e \in E(g)} \mathcal{A}_e$, let

$$P_{g,Q}(A \mid x) \equiv \prod_{e \in E(g)} Q_e(A_e \mid x_e),$$

where we use the notation $x_e = x_u \cdot x_v$ for $e = (u, v)$.

DEFINITION 2.1 (Graphical channel, deterministic graph). Let g, Q and $P_{g,Q}$ be as above. We call $P_{g,Q}$ a graphical channel with graph g and channels Q.

DEFINITION 2.2 (Graphical channel, random graph). Let $G = (V, E(G))$ be a random graph with vertex set $V = [n]$, and let Q be a collection of edge channels (as above) so that for any edge e, Q_e is defined if $\mathbb{P}(e \in E(G)) > 0$. Let $P_{G,Q}$ be the random channel with output alphabet $\prod_{e \in E(G)} \mathcal{Y}_e$ and input alphabet $\{-1, +1\}^V$ given by $P_{g,Q}$ for each realization $G = g$.

DEFINITION 2.3 (Binary synchronization instance). Let $P_{G,Q}$ be an n-node graphical channel, and let X be uniformly drawn in $\{-1,+1\}^n$. Let Y be the output of X through the graphical channel $P_{G,Q}$. The pair (X, Y) is an instance of a binary synchronization problem drawn from $P_{G,Q}$.

3. Main results.

In this paper, we provide progress towards answering the following question: given a binary synchronization instance (X, Y) drawn from $P_{G,Q}$, for $u, v \in V$, if we know X_v and we know Y, then when is it impossible to reconstruct X_u?

3.1. *Chi-squared mutual information.*

In particular, we provide an upper bound on the information that X_v and Y give about X_u. This information is quantified by the Chi-squared mutual information

$$I_2(X_u; X_v, Y_{E(G)}),$$

which is the f-mutual information based on the Chi-squared divergence, D_{χ^2}—see the Appendix for a reminder on the definitions and properties of these functionals.

PROPOSITION 3.1. *If (X, Y) is a binary synchronization instance with underlying graph G, and $u \in V(G)$, $S \subseteq V(G)$, then following equality holds*:

$$I_2(X_u; X_S, Y_{E(G)}) = I_2(X_u; X_S \mid Y_{E(G)}). \tag{9}$$

The Chi-squared mutual information takes the following simple expression:

PROPOSITION 3.2. *If (X, Y) is a binary synchronization instance with underlying graph G, and $u, v \in V(G)$, then the following equality holds*:

$$I_2(X_u; X_v \mid Y_{E(G)}) = \mathbb{E}_Y\big[\mathbb{E}_X[X_u \cdot X_v \mid Y]^2\big].$$

The definition of I_2, and the proofs of Propositions 3.1 and 3.2, can be found in the Appendix.

3.2. *Bond percolation.*

In our main result, we bound the Chi-squared mutual information $I_2(X_u; X_v \mid Y_{E(G)})$ by the connection probability between u and v in a bond percolation on the underlying graph, G.

DEFINITION 3.3 (Bond percolation on a graph). Let $G = (V, E(G))$ be a graph, and let $\gamma : E(G) \to [0, 1]$. Then, a bond percolation with open probability γ on G is a random edge-labelling

$$B : E(G) \to \{\text{open}, \text{closed}\},$$

such that each edge label $B(e)$ is assigned independently of the other edge labels, and such that for all e,

$$\mathbb{P}\big[B(e) = \text{open} \mid e \in E(G)\big] = \gamma_e.$$

Let B be a bond percolation on G. If a subgraph $H \subseteq G$ is such that $B(e) = \text{open}$ for all $e \in E(H)$, then we call H an open subgraph.

DEFINITION 3.4 (Connection probability in percolation). Let $S, T \subseteq V(G)$. Then we write their connection probability in a percolation on G with open probability $\gamma : E(G) \to [0, 1]$ as

$$\mathrm{conn}_{G,\gamma}(S, T).$$

This denotes the probability that there is a pair of vertices $u \in S$, $v \in T$, such that u is connected to v by an open path in a bond percolation on G where each edge $e \in E(G)$ is independently open with probability $\gamma(e)$.

3.3. *Symmetric channels.* Our information-theoretic bound for spin synchronization applies to "symmetric" graphical channels defined as follows.

DEFINITION 3.5. A graphical channel $P_{G,Q}$ is symmetric if for each edge $e \in E(G)$ the channel $Q_e(\cdot \mid \cdot)$ is symmetric. An edge channel $Q_e(\cdot \mid \cdot)$ is symmetric if there is a measurable transformation $T_e : \mathcal{Y}_e \to \mathcal{Y}_e$ on the output alphabet of $Q_e(\cdot \mid \cdot)$ such that $T_e = T_e^{-1}$, and such that for all measurable $A \subset \mathcal{A}_e$ we have

$$Q_e(A \mid +1) = Q_e\big(T_e(A) \mid -1\big),$$

and hence

$$Q_e\big(T_e(A) \mid +1\big) = Q_e(A \mid -1).$$

In other words, an edge channel $Q_e(\cdot \mid \cdot)$ is symmetric if "flipping the sign" using T_e of an edge label Y_e with distribution $Q_{e|+}$ gives an edge label $T_e(Y_e)$ with distribution $Q_{e|-}$.

Symmetric graphical channels cover a broad collection of models, discussed in Section 4.

3.4. *Information-percolation bound.*

THEOREM 3.6 (Main percolation bound). *Let $P_{G,Q}$ be a symmetric graphical channel. Let (X, Y) be a binary synchronization instance drawn from $P_{G,Q}$.*

Then for all $u, v \in V$,

$$I_2(X_u; X_v \mid Y_{E(G)}) \le \mathrm{conn}_{G,\gamma}(u, v),$$

where

$$\gamma(i, j) = I_2(X_i; X_j \mid Y_{(i,j)})$$

for all $(i, j) \in E(G)$.

The following corollary follows by a union bound:

COROLLARY 3.7 (Subadditivity of I_2). *Let $P_{G,Q}$, (X, Y) be as in Theorem 3.6.*

Then for all $u, v \in V$, I_2 is subadditive over paths, that is,

$$I_2(X_u; X_v \mid Y_{E(G)}) \le \sum_{P \in \mathcal{P}_G(u,v)} I_2(X_u; X_v \mid Y_{E(P)}),$$

where $\mathcal{P}_G(u, v)$ is the set of paths (i.e., self-avoiding walks) from u to v in G.

Moreover, the theorem can be extended to bound the mutual information between a vertex and a set of vertices.

COROLLARY 3.8 (Mutual information between label and set of labels). *Let $P_{G,Q}$, (X, Y), and γ be as in Theorem 3.6. Then for all $u \in V$, $S \subseteq V$,*

$$I_2(X_u; X_S \mid Y_{E(G)}) \le \mathrm{conn}_{G,\gamma}(u, S).$$

Finally, we note that our upper bounds on $I_2(X_u; X_S \mid Y_{E(G)})$ imply upper bounds on $I_{\mathrm{KL}}(X_u; X_S \mid Y_{E(G)})$.

COROLLARY 3.9. *Let X, Y, u, S, and γ be as in Corollary 3.8. Then*

$$I_{\mathrm{KL}}(X_u; X_S \mid Y_{E(G)}) \le \mathrm{conn}_{G,\gamma}(u, S).$$

PROOF. This follows from Theorem 3.6 and the upper bound of Lemma A.9 in the Appendix. □

3.5. *General* (*asymmetric*) *channels and comparison to* [27]. An information-percolation bound analogous to Theorem 3.6 is obtained by Polyanskiy and Wu in a concurrent work [27]. In particular, Polyanskiy and Wu are inspired by their prior results for strong data-processing inequalities for channels and Bayesian networks [28], and their original motivation was to obtain a simple proof of a result from here.

In the model of [27], there is a bipartite graph $G = (V, W, F)$ with parts V, W and edge set F, and vertex labels $\{X_v\}_{v \in V}$ and $\{Y_w\}_{w \in W}$ on a discrete alphabet such that the labels $\{Y_w\}_{w \in W}$ are independent conditioned on X_V, and each $Y_w \sim P_{Y_w|X_{N(w)}}$, where $N(w) \subset V$ denotes the set of neighbors of w. This gives a generalization of our model since our model corresponds to the case where (i) the labels $\{X_v\}_{v \in V}$ are independent and distributed as $\mathrm{Rad}(1/2)$, (ii) each vertex $w \in W$ has degree 2, and (iii) each observation Y_w with $N(w) = \{i, j\}$ is distributed as $Y \sim P_{Y_w|X_i \cdot X_j}$.

Under this model, Polyanskiy and Wu prove two theorems that generalize Theorem 3.6. The first result of [27] states that if the labels $\{X_v\}_{v \in V}$ are independent, then for any subsets $S_1, S_2 \subset V$,

$$I_{\mathrm{KL}}(X_{S_1}; X_{S_2} \mid Y_W) \leq \sup_{v \in V} H(X_v) \cdot \sum_{u \in S_1} \mathrm{siteconn}_{G,\eta_{\mathrm{KL}}}(u, S_2). \tag{10}$$

Here $\mathrm{siteconn}_{G,\eta_{\mathrm{KL}}}(u, S_2)$ denotes the probability that u is connected to a vertex in S_2 in a *site* percolation on G, where each vertex $w \in W$ is included with independent probability $\eta_{\mathrm{KL}}(P_{Y_w|X_{N(w)}})$, the Chi-squared Strong Data Processing Inequality (SDPI) constant of the channel. Applying (10) to the model of this paper yields Corollary 3.9 when the edge channels Q_e are symmetric—because the SDPI constant is equal to the I_2 information in this case. The bound of [27] also applies to asymmetric channels, but with the SDPI constant instead of the I_2 information for the edge openness probability in the percolation. Specifically, applied to our model, inequality (10) yields for any $S_1, S_2 \subset V$:

$$I_{\mathrm{KL}}(X_{S_1}; X_{S_2} \mid Y_{E(G)}) \leq \sum_{u \in S_1} \mathrm{conn}_{G,\eta_{\mathrm{KL}}}(u, S_2). \tag{11}$$

The proofs of the first result of [27] and of this paper's result are similar in that both proceed by induction on the number of observations Y_w, and both use the linearity of the connection probability in the percolation model (as a function of the probability of opening an edge when all other edge probabilities are fixed). Moreover, both proofs bound the increase in information each time an extra edge observation is added. While this basic structure is similar, [27] benefits from the chain rule for KL-mutual information. We note that a similar inductive proof is already used in the prior work of Polyanskiy and Wu [28].

The second result of [27] applies to the case where the labels $\{X_v\}_{v \in V}$ are not required to be independent. This result uses the tensorization of the less-noisy relation proved previously by Polyanskiy and Wu in [28] in order to show that given labels $\{Y_w\}_{w \in W}$ and $\{\tilde{Y}_w\}_{w \in W}$ distributed as $Y_w \sim P_{Y_w|X_{N(w)}}$ and $\tilde{Y}_w \sim Q_{\tilde{Y}_w|X_{N(w)}}$ where the channels $Q_{\tilde{Y}_w|X_{N(w)}}$ are less-noisy than the channels $P_{Y_w|X_{N(w)}}$, then for all subsets $S_1, S_2 \subset V$,

$$I_{\mathrm{KL}}(X_{S_1}; Y_W | X_{S_2}) \leq I_{\mathrm{KL}}(X_{S_1}; \tilde{Y}_W | X_{S_2}). \tag{12}$$

Applied to our model, the inequality (12) also yields Corollary 3.8 in the case of symmetric edge channels. And, similarly to their first result (10), their second result also yields the bound (11) for general asymmetrical channels.

Nevertheless, we now show that the SDPI approach of [27] can be loose on asymmetric cases. We first provide a bound that applies to asymmetric channels and that is based on the Chi-squared mutual information approach derived here. The bound gives up some generality on the base graph, but it applies to any edge channels:

THEOREM 3.10. *Let $P_{G,Q}$ be a graphical channel where G is a series-parallel graph with terminals u and v, and the edge channels Q are arbitrary (potentially asymmetric). Let (X, Y) be a binary synchronization instance drawn from $P_{G,Q}$. Then*

$$I_2(X_u; X_v \mid Y_{E(G)}) \le \sum_{P \in \mathcal{P}_G(u,v)} I_2(X_u; X_v \mid Y_{E(P)}).$$

Here $\mathcal{P}_G(u, v)$ is the set of paths (self-avoiding walks) from u to v in G.

We next given an example where the bound of [27] is looser than the one above.

Example and comparison to (11). Suppose we construct G by taking a d-ary tree T of height h with root u, and adding a vertex v adjacent to all of the leaves of T. Then G is a series-parallel graph with terminals u and v, and we may apply Theorem 3.10 to graphical channels on G. In particular, suppose all of the edges $(i, j) \in E(T) \subset E(G)$ have channels Q_{ij} defined by $Q_{ij}(\cdot \mid +1) \sim \mathrm{Ber}(a/d)$ and $Q_{ij}(\cdot \mid -1) \sim \mathrm{Ber}(b/d)$ for some constants $a, b \ge 0$, and suppose the the edge channels Q_{lv} for leaves $l \in V(T)$ are noiseless. Applying Theorem 3.10, one sees that $I_2(X_u; X_v \mid Y_{E(G)})$ vanishes (with increasing height h) as soon as

$$\frac{(a-b)^2}{2(a+b)} < 1,$$

while the bound (11) of [27] requires

$$(\sqrt{a} - \sqrt{b})^2 < 1.$$

Since $\frac{(a-b)^2}{2(a+b)} < (\sqrt{a} - \sqrt{b})^2$ unless $a = b$, the bound of Theorem 3.10 is tighter than (11) for this case.

This theorem follows directly from (i) a multiplicative property of the Chi-squared mutual information on paths (see Proposition A.5), and (ii) a subadditivity property of the Chi-squared mutual information on depth-1 trees (see Lemma A.6). Interestingly, (i) does not hold for the classical mutual information, I_{KL}, making the Chi-squared mutual information, I_2, a natural choice for this proof.

PROOF OF THEOREM 3.10. In the following, we implicitly use $I_2(X_i X_j; Y) = I_2(X_i; X_j \mid Y)$, by Proposition A.4.

The proof is by induction on $|E(G)|$. The base case, $|E(G)| = 1$, is trivial. For the inductive step, one of two cases holds:

(Case 1) G is the series composition of H_1 which is series-parallel with terminals u, w, and H_2, which is series-parallel with terminals w, v.

$$\begin{aligned}
&I_2(X_u X_v; Y_{E(G)}) \\
&\quad = I_2\big((X_u X_w) \cdot (X_w X_v); Y_{E(H_1)}, Y_{E(H_2)}\big) \\
\text{(Prop. A.5)} &\quad = I_2(X_u X_w; Y_{E(H_1)}) I_2(X_w X_v; Y_{E(H_2)}) \\
&\quad \le \sum_{\substack{P_1 \in \mathcal{P}_{H_1}(u,w) \\ P_2 \in \mathcal{P}_{H_2}(w,v)}} I_2(X_u X_w; Y_{E(P_1)}) I_2(X_w X_v; Y_{E(P_2)}) \\
\text{(Prop. A.5)} &\quad = \sum_{\substack{P_1 \in \mathcal{P}_{H_1}(u,w) \\ P_2 \in \mathcal{P}_{H_2}(w,v)}} I_2(X_u X_v; Y_{E(P_1)}, Y_{E(P_2)}) \\
&\quad = \sum_{P \in \mathcal{P}_G(u,v)} I_2(X_u X_v; Y_{E(P)}).
\end{aligned}$$

The inequality is by the inductive hypothesis.

(Case 2) G is the parallel composition of H_1 and H_2 both series-parallel, with terminals u, v. Then,

$$\begin{aligned} & I_2(X_u X_v; Y_{E(G)}) \\ & \quad = I_2(X_u X_v; Y_{E(H_1)}, Y_{E(H_2)}) \\ \text{(Lemma A.6)} \qquad & \quad \le I_2(X_u X_v; Y_{E(H_1)}) + I_2(X_u X_v; Y_{E(H_2)}). \end{aligned}$$

The inductive step follows by the inductive hypothesis, since $\mathcal{P}_G(u,v) = \mathcal{P}_{H_1}(u,v) \sqcup \mathcal{P}_{H_2}(u,v)$. □

The proofs of the auxiliary propositions can be found in the Appendix.

4. Applications. Many common edge channels enjoy the symmetry property of Definition 3.5. We discuss here some important examples.

Binary symmetric channel. One example is the binary symmetric channel with flip probability ε (BSC_ε, for short). This channel has input and output alphabet $\{-1, +1\}$, and is given by

$$\mathsf{BSC}_\varepsilon(y \mid x) = \begin{cases} 1-\varepsilon, & x = y, \\ \varepsilon, & x \neq y. \end{cases}$$

This channel is symmetric in the sense of Definition 3.5, because the transformation $T(y) = -y$ satisfies both $T^2 = 1$ and $\mathsf{BSC}_\varepsilon(T(y) \mid x) = \mathsf{BSC}_\varepsilon(y \mid -x)$.

Additive white Gaussian noise channel. Another example is the Gaussian noise channel AWGN_λ, whose output distribution $\mathsf{AWGN}_\lambda(\cdot \mid x)$ is the distribution of the random variable

$$Y_x = \sqrt{\lambda} x + Z,$$

where $Z \sim \mathcal{N}(0, 1)$ is independent Gaussian noise with mean 0 and variance 1. This channel is also symmetric in the sense of Definition 3.5, because the transformation $T(y) = -y$ satisfies $T^2 = 1$ and $\mathsf{AWGN}_\lambda(T(\cdot) \mid x) = \mathsf{AWGN}_\lambda(\cdot \mid x)$, since $-Y_x = -\sqrt{\lambda} x - Z \sim \sqrt{\lambda}(-x) + Z = Y_{-x}$, because $Z \sim -Z$.

Tables 1 and 2 give examples of information-theoretic thresholds that can be obtained as a direct consequence of Theorem 3.6. In all of these cases, our bounds either match or improve the previously-known bounds.[3] The tables also give a few new results.

4.1. *Broadcasting on trees.* In the "broadcasting on trees" problem, each vertex $v \in V(T)$ of an infinite tree T has a binary hidden label σ_v. The hidden labels are assigned by letting the root ρ have spin $\sigma_\rho \sim \mathrm{Rad}(1/2)$, and by defining edge labels $\{\eta_e\} \overset{\text{i.i.d.}}{\sim} \mathrm{Rad}(\varepsilon)$, and letting

$$\sigma_v = \sigma_\rho \prod_e \eta_e,$$

where the product is over the edges in the path from ρ to v.

In [17], it is proved that for $(1-2\varepsilon)^2 < p_c(T)$,

$$I_{\mathrm{KL}}\big(\sigma_\rho; (\sigma_v)_{\{v: d(\rho,v)=t\}}\big) \to 0 \quad \text{as } t \to \infty,$$

[3] Note that [3] does not attempt to obtain the tightest bound, but rather the existence of a positive lower bound on the threshold.

TABLE 1
Regimes in which weak recovery/reconstruction is impossible for BSC_ε *edge channels*

	BSC_ε	
Graph	Known bound	Our bound
Tree T	$(1-2\varepsilon)^2 \le p_c(T)$ Broadcasting on Trees [17]	$(1-2\varepsilon)^2 \le p_c(T)$ Section 4.1
Erdős–Rényi$(n, c/n)$	$(1-2\varepsilon)^2 \le 1/c$ Censored Block Model [23, 24]	$(1-2\varepsilon)^2 \le 1/c$ Section 4.2
Grid $\mathbb{L}^2$	$(1-2\varepsilon)^2 \le 1/4$ Grid Synchronization [3]	$(1-2\varepsilon)^2 \le 1/2$ Section 4.3
Complete K_n		$(1-2\varepsilon)^2 < 1/n$

where $d(v, w)$ denotes for the distance between two vertices v and w in T and $p_c(T)$ denotes the critical probability for bond percolation on T.[5] In other words, for ε too close to $\frac{1}{2}$, the information given by the depth-n vertex labels about the root goes to 0, and hence reconstruction of the root label from the leaf labels becomes impossible (by an analogue of Proposition A.7). In fact, [17] showed this bound on the mutual information is tight: reconstruction is possible for $(1-2\varepsilon)^2 > p_c(T)$, which was already known from [22] in some cases. But we will only concern ourselves with the impossibility result of the paper.

EXAMPLE 4.1. We rederive the impossibility result of [17] by applying Corollary 3.8.

PROOF. The proof follows by constructing a group synchronization problem that is equivalent to the broadcasting problem.

Let $\{X_v\}_{v \in V(T)\setminus\rho} \overset{\text{i.i.d.}}{\sim} \mathrm{Rad}(1/2)$. Let $X_\rho = \sigma_\rho$. For each $e = (i, j) \in E(T)$ define

$$Y_{ij} = X_i \cdot X_j \cdot \eta_e.$$

TABLE 2
Regimes in which weak recovery/reconstruction is impossible for AWGN_λ *edge channels*

	AWGN_λ	
Graph	Known bound	Our bound[4]
Tree T		$f(\lambda) \le p_c(T)$
Erdős–Rényi$(n, c/n)$		$f(\lambda) \le 1/c$
Grid $\mathbb{L}^2$		$f(\lambda) \le 1/2$
Complete K_n	$\lambda \le c/n$ for $c < 1$	$\lambda \le c/n$ for $c < 1$
Complete K_n	Spiked Wigner [13]	Section 4.4

[4] Where $f(\lambda) = I_2(X_1; X_2 \mid Y^{(\lambda)})$ for $Y^{(\lambda)} = \sqrt{\lambda} X_1 X_2 + Z$, and $X_1, X_2 \overset{\text{i.i.d.}}{\sim} \mathrm{Rad}(1/2)$, $Z \sim \mathcal{N}(0, 1)$. As calculated in [4], $f(\lambda) = \mathbb{E}[\tanh(\lambda + \sqrt{\lambda} Z)^2]$.

[5] Formally, $0 \le p_c(T) \le 1$ is the critical value such that if $p < p_c(T)$ then the open connected components of a bond percolation on T (with edge openness probability p) are finite a.s., but if $p > p_c(T)$, then a.s. there are open components of infinite size.

Then (X, Y) is a binary synchronization instance drawn from $P_{T,Q}$, where Q_e is BSC_ε for each edge $e \in E(T)$. Notice that

$$I_2\big(\sigma_\rho; (\sigma_v)_{\{v:d(\rho,v)=t\}}\big) \le I_2\big(X_\rho; (X_v)_{\{v:d(\rho,v)=t\}}, Y\big) \tag{13}$$

$$\le \mathbb{P}\begin{bmatrix}\text{There is a length-}t\text{ path from }\rho\text{ in}\\ \text{a bond percolation on }T\text{ with open}\\ \text{probability }(1-2\varepsilon)^2\end{bmatrix}, \tag{14}$$

where (13) follows by the data-processing inequality, and (14) follows by Theorem 3.8. For $(1-2\varepsilon)^2 \le p_c(T)$, the probability of a length-t path from the root in the $(1-2\varepsilon)^2$-bond percolation vanishes as $t \to \infty$, proving the impossibility result by application of Proposition A.7. □

4.2. *Clustering in the censored block model.* Another application arises in the domain of graph clustering and community detection. Our bound applies to the Censored Block Model (CBM). This model is defined in [2] for general graphs G when the edge channel consists of BSCs, that is,

$$Y_{ij} = X_i \cdot X_j \cdot Z_{ij},$$

for each $(i, j) \in E(G)$, where $Z_{ij} \sim \text{Rad}(\varepsilon)$ is independent noise. In the language of our paper, (X, Y) is a binary synchronization instance on G, and all the edge channels are BSC_ε.

EXAMPLE 4.2. Suppose G is distributed as an Erdös–Rényi random graph $G(n, \frac{c}{n})$. Weak recovery is impossible in a censored block model on G with flip probability ε if

$$c \le 1/(1-2\varepsilon)^2.$$

PROOF. For all distinct $u, v \in V(G)$, by Theorem 3.6,

$$\begin{aligned} I_2(X_u; X_v \mid Y) &\le \text{conn}_{G,(1-2\varepsilon)^2}(u, v) \\ &= \text{conn}_{K_n,(c(1-2\varepsilon)^2/n)}(u, v) \to 0 \end{aligned}$$

if $c \le 1/(1-2\varepsilon)^2$, since the largest component of $G(n, c/n)$ is of size $O(n^{2/3}) = o(n)$ in this regime (by [16]).

Suppose by contradiction that we are given an algorithm $\hat{X}(Y)$ that solves weak recovery for the CBM with parameters c and ε. In particular, there is a pair of distinct $u, v \in V(G)$ such that

$$\mathbb{P}[X_u \cdot X_v = \hat{X}_u \cdot \hat{X}_v] = \mathbb{P}[X_v \cdot \hat{X}_u \cdot \hat{X}_v = X_u] > \frac{1}{2} + d$$

for some $d > 0$ independent of n. So by Proposition 3.1, the data-processing inequality, and Proposition A.7

$$I_2(X_u; X_v \mid Y) = I_2(X_u; X_v, Y) \ge I_2(X_u; X_v, \hat{X}) \ge I_2(X_u; X_v \cdot \hat{X}_u \cdot \hat{X}_v) > d',$$

for $d' > 0$ independent of n. This contradicts the fact that $I_2(X_u; X_v \mid Y) \to 0$ as $n \to \infty$. □

This rederives a threshold conjectured in [20] and proved in [23]. The proof is analogous to the proof of [24] that establishes nonreconstruction for the two-community symmetric Stochastic Block Model SBM$(n, a/n, b/n)$ when $(a-b)^2 \le 2(a+b)$. While [23] does not establish the impossibility of reconstruction at the critical threshold, it is straightforward to extend the argument at the threshold. Note also that this gives a tight threshold, that is, it is proved that reconstruction (a.k.a. weak recovery) is possible above this threshold [9, 29].

4.3. *Grid synchronization.* The proof of [23] which implies impossibility of reconstruction in the censored block model on the Erdös–Rényi random graphs $G(n, c/n)$ relies crucially on the fact that for constant c, most small neighborhoods of vertices in $G(n, c/n)$ are trees.

However, the method of coupling with trees would no longer apply if we were to work with the Censored Block Model on a grid, because grids have many small cycles. In this case, our bound still goes through, and is in fact stronger than the previously-known bound of [3] for binary synchronization. Supposing the edge channels were binary symmetric channels with flip probability ε, The previous bound required $(1-2\varepsilon)^2 \le \frac{1}{4}$ for impossibility of synchronization, while ours only requires $(1-2\varepsilon)^2 \le \frac{1}{2}$:

EXAMPLE 4.3. Let $\mathbb{L}^2$ be the two-dimensional lattice with vertices $V(\mathbb{L}^2) = \mathbb{Z}^2$ and edges given by the Hamming distance. Let $v_1, \dots, v_k, \dots$ be a sequence of vertices such that v_k is at distance k from 0. Let (X, Y) be a binary synchronization instance drawn from $P_{\mathbb{L}^2,Q}$, where all the edge channels are BSC_ε. Then, if $(1-2\varepsilon)^2 \le \frac{1}{2}$, we have $I_2(X_0; X_{v_k} \mid Y) \to 0$ as $k \to \infty$. Impossiblity of reconstruction follows by Proposition A.7.

PROOF. By Theorem 3.6

$$I_2(X_0; X_{v_k} \mid Y) \le \mathrm{conn}_{\mathbb{L}^2,(1-2\varepsilon)^2}(0, v_k) \tag{15}$$

$$\to 0 \quad \text{as } k \to \infty. \tag{16}$$

Line (16) follows because $(1-2\varepsilon)^2 \le 1/2$, which is the critical bond percolation constant of $\mathbb{L}^2$. And it is known that the probability that there is an open length-k path containing the origin in a critical or sub-critical bond percolation on $\mathbb{L}^2$ vanishes as $k \to \infty$. A reference for this is [18].

Notice that in (15) we have applied Theorem 3.6 in the case of an infinite graph, although we have technically proved the theorem only for finite graphs. We may do this by the monotone convergence of the information and of the connection probability in the percolation. □

4.4. *Spiked Gaussian Wigner model.* In the spiked Wigner model with Rad(1/2) priors, we are given an $n \times n$ matrix

$$Y_\lambda = \sqrt{\frac{\lambda}{n}} X X^T + W,$$

where X is uniform in $\{-1, +1\}^n$, and W is an independent Gaussian Wigner matrix (real, symmetric, the entries are distributed as unit Gaussians $\mathcal{N}(0, 1)$ and are all independent except for the symmetry constraint).

The spiked Wigner model, and spiked matrix models in general, have been studied in various contexts: for example, in order to evaluate statistical methods such as PCA that estimate low-rank information from noisy data, or as variants of the stochastic block model ([5, 21, 26]). For Y_λ as above, [13] proved that there is a phase transition in the problem of weak recovery at exactly the critical threshold $\lambda_c = 1$. The impossibility part of this phase transition was later rederived in a more general setting by [26].

The impossibility of recovery for $\lambda < 1$ is a direct consequence of Theorem 3.6:

EXAMPLE 4.4. Let Y_λ be defined as above. Then, for $\lambda < 1$, $I_2(X_u; X_v \mid Y_\lambda) \to 0$ for all $u \ne v$, and hence it is impossible to weakly recover X from Y_λ.

PROOF. (X, Y_λ) is distributed as a binary synchronization instance drawn from a graphical channel on K_n, in which each edge channel $Q_{(i,j)}$ is given by

$$Y_{\lambda,ij} = \sqrt{\frac{\lambda}{n}} X_i \cdot X_j + Z_{ij},$$

where $Z \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$. Notice that the edge channels are symmetric (with the transformation $y \mapsto -y$). Analogously to the case of the censored block model on $G(n, c/n)$, it suffices to show that

$$I_2(X_i; X_j \mid Y_{\lambda,ij}) = \frac{\lambda}{n} + o(1/n).$$

This is done by explicit calculation. Writing $a = \sqrt{\lambda/n}$,

$$\begin{aligned} I_2(X_i; X_j \mid Y_{\lambda,ij}) &= \mathbb{E}\big[\mathbb{E}[X_i \cdot X_j \mid Y_{\lambda,ij}]^2\big] \\ &= 2 \int_{-\infty}^{+\infty} \frac{e^{-(x-a)^2/2}}{2\sqrt{2\pi}} \left(\frac{e^{-(x-a)^2/2} - e^{-(x+a)^2/2}}{e^{-(x-a)^2/2} + e^{-(x+a)^2/2}} \right)^2 dx \\ &\le 2 \int_{-\infty}^{+\infty} \frac{e^{-(x-a)^2/2}}{2\sqrt{2\pi}} (ax)^2 \, dx \qquad (17) \\ &= a^2\big(a^2 + 1\big) \\ &= \frac{\lambda}{n} + o(1/n). \end{aligned}$$

Line (17) is a standard Gaussian integral. □

5. Additional results and future directions.

Relationship with correlation decay. As mentioned in the Introduction, fixing the edge observations and applying the Ising model correlation decay conditions yields an impossibility result for reconstruction. However, the bounds that we achieve with this method are not as strong as those we proved in this paper, because the techniques in our paper allow us to deal with the average-case edge observations, while fixing the edge observations and applying the Dobrushin conditions requires us to work with the worst-case edge observations. It would nonetheless be interesting to elaborate on this connection.

Extensions to our results. Various natural extensions can be considered for the information-percolation bound, Theorem 3.6. For example, one may consider more general edge channels, such as nonbinary input alphabets and nonsymmetrical channels.

We provide below a more general condition on the edge channel that would suffice for the current proof technique to work in these more general settings, without giving explicit examples. In the theorem below, the vertex labels are uniformly random members of some finite group $\mathcal{G}$ (not necessarily $\{+1, -1\}$), and the edge labels, $Y_{(i,j)}$, are noisy observations of the differences of the endpoints, $X_i \cdot X_j^{-1}$. The proof of Theorem 5.1 is analogous to the proof of Theorem 3.6.

THEOREM 5.1. *Let $G = (V, E)$ be a finite graph with vertex set V and edge set E. For every $\gamma \in [0, 1]$, let Q^γ be a collection of edge channels for G, with input alphabet a finite group $\mathcal{G}$. For any $\Gamma \in [0, 1]^E$, let Q^Γ be the collection of edge channels $\{Q_e^{\Gamma(e)}\}_{e \in E}$, and let (X^Γ, Y^Γ) be a group-$\mathcal{G}$ synchronization instance drawn from P_{G,Q^Γ}.*

1. *Suppose that*

$$I_2(X_e^0; Y_e^0) = 0$$

for all $e \in E$.

2. *Suppose also that for every $e \in E$, $u, v \in V$, $\Gamma \in [0,1]^E$, $I(\gamma)$ is continuous for all $\gamma \in [0,1]$ and*

$$\frac{\partial}{\partial \gamma} \frac{I(\gamma) - I(0)}{\gamma} \ge 0,$$

for all $\gamma \in (0,1)$, where

$$I(\gamma) \equiv I_2\big(X_u^{\Gamma_{e,\gamma}}; X_v^{\Gamma_{e,\gamma}} \mid Y^{\Gamma_{e,\gamma}}\big),$$

and $\Gamma_{e,\gamma}$ denotes the function in $[0,1]^E$ such that $\Gamma_{e,\gamma}(e) = \gamma$ and $\Gamma_{e,\gamma}(f) = \Gamma(f)$ for all $f \neq e$.

Then, for any $u, v \in V$,

$$I_2\big(X_u^\Gamma; X_v^\Gamma \mid Y^\Gamma\big) \le \big(|\mathcal{G}| - 1\big) \cdot \mathrm{conn}_{G,\Gamma}(u,v).$$

The concurrent work [27] applies to some of these more general cases, although we discussed with Theorem 3.10 how the resulting bound can be loose on these. In order to further investigate the tightness of [27], it would be useful to determine whether the subadditivity inequality of Theorem 3.10 holds on arbitrary graphs. We conjecture that it does.

A percolation lower bound. When the edge channels of the graphical channel are symmetric, Theorem 3.6 is tight on trees, so one cannot open the edges with lower probability in general. Is there a converse to Theorem 3.6: that is, is the mutual information lower-bounded by the connection probability on some nontrivial percolation? For example, for some bounded-degree graphs?

6. Proofs of Theorem 3.6 and Corollaries 3.7 and 3.8.

We first prove a version of Theorem 3.6 for the special case in which all of the edge channels are binary symmetric. We will then extend this specific result to general symmetric channels.

THEOREM 6.1. *Let $P_{G,Q}$ be a graphical channel, where each edge channel Q_e is a binary symmetric channel. Let (X,Y) be a binary synchronization instance drawn from $P_{G,Q}$. Then for all $u, v \in V$,*

$$I_2(X_u; X_v \mid Y_{E(G)}) \le \mathrm{conn}_{G,\gamma}(u,v),$$

where

$$\gamma(i,j) := I_2(X_i; X_j \mid Y_{(i,j)})$$

for all $(i,j) \in E(G)$.

PROOF. Suppose we can prove the theorem for the case in which the graph is deterministic. Then, since the graph G is a deterministic function of the edge observations Y,

$$\begin{aligned} I_2(X_u; X_v \mid Y) &= \mathbb{E}_G\big[I_2(X_u; X_v \mid Y)\big] \\ &\le \mathbb{E}_G\big[\mathrm{conn}_{G,\gamma}(u,v)\big] \\ &= \mathrm{conn}_{G,\gamma}(u,v), \end{aligned}$$

as desired. Therefore, we may assume that G is deterministic.

Let the flip probability of Q_e be $\varepsilon(e)$, and define $\delta(e) = (1 - 2\varepsilon(e))$. We can assume that $\delta(e) \in [0, 1]$, because we lose no information by flipping edge labels deterministically. Also, by direct calculation $\gamma(e) = \delta(e)^2 = I_2(X_i; X_j | Y_{ij})$ for each edge $e = (i, j)$.

The proof goes by induction on $|S_\delta|$, where

$$S_\delta := \{e \in E(G) : \delta(e) \notin \{0, 1\}\}.$$

In the base case, $|S_\delta| = 0$, so all edge observations are completely noiseless or completely noisy. Hence, $I_2(X_u, X_v|Y) = 1$ if there is a path P from u to v whose edges are all noiseless. If there is no such path, then $I_2(X_u, X_v|Y) = 0$. This is exactly the statement $I_2(X_u, X_v|Y) = \mathrm{conn}_{G,\gamma}(u, v)$.

For the inductive step, assume the theorem when the BSC channel flip probabilities are given by $\delta' : E(G) \to [0, 1]$ with $|S_{\delta'}| < |S_\delta|$. Pick an arbitrary edge $f \in S_\delta$. We will now interpolate between the case in which $\delta(f) = 0$, and the case in which $\delta(f) = 1$, with the other edge channels held fixed. For any $t \in [0, 1]$, let $\delta_t : E(G) \to [0, 1]$ be given by $\delta_t(e) = \delta(e)$ for $e \neq f$, and $\delta_t(f) = t$. Define corresponding spin synchronization instances (X_t, Y_t), and also $\gamma_t = \delta_t^2$. Write

$$I(t) := I_2(X_{t,u}; X_{t,v}|Y_t) \quad \text{and} \quad C(t) := \mathrm{conn}_{G,\gamma_t}(u, v).$$

In order to prove that $I(t) \le C(t)$ for all $t \in [0, 1]$, we need the following claim:

CLAIM 6.2. *There is nondecreasing $h : [0, 1] \to \mathbb{R}$ such that*

$$I(t) = I(0) + \big(I(1) - I(0)\big) \cdot t^2 \cdot h(t).$$

Assume the claim is true. Then, $h(1) = 1$, and since $h(t)$ is nondecreasing, $h(t) \le 1$. Hence,

$$\begin{aligned} I(t) &\le I(0) + \big(I(1) - I(0)\big) \cdot t^2 \\ &= I(0) \cdot \big(1 - t^2\big) + I(1) \cdot t^2 \\ &\le C(0) \cdot \big(1 - t^2\big) + C(1) \cdot t^2 = C(t). \end{aligned}$$

The inequality of the last line follows because $I(0) \le C(0)$ and $I(1) \le C(1)$ by the inductive hypothesis. The equality of the last line follows by the linearity of the connection probability in the parameter $\gamma_t(f) = t^2$.

It only remains to prove the claim. Write $E' = E(G) \setminus f$. Also write $f = (i, j)$, $A_t = X_{t,u} \cdot X_{t,v}$, and $B_t = X_{t,i} \cdot X_{t,j}$. By Proposition A.3, and because $Y_{t,E'}$ is a subset of Y_t,

$$I(t) = \mathbb{E}\big[\mathbb{E}[A_t|Y_t]^2\big] = \mathbb{E}\big[\mathbb{E}\big[\mathbb{E}[A_t|Y_t]^2|Y_{t,E'}\big]\big].$$

Since the only edge channel to change with t is Q_f, we can couple $X_0 = X_t$, and $Y_{0,E'} = Y_{t,E'}$. Hence, it suffices to prove that the function

$$h(t; Y_{0,E'}) := \frac{1}{t^2}\big(\mathbb{E}\big[\mathbb{E}[A_t|Y_t]^2|Y_{t,E'}\big] - \mathbb{E}\big[\mathbb{E}[A_t|Y_0]^2|Y_{0,E'}\big]\big)$$

is nondecreasing in t, since then we can set $h(t) = \sum_{\sigma \in \{-1,+1\}^{E'}} h(t; \sigma) \cdot \mathbb{P}[Y_{0,E'} = \sigma]$, which will also be nondecreasing in t.

Fix $\sigma \in \{-1, +1\}^{E'}$ such that $\mathbb{P}[Y_{t,E'} = \sigma] > 0$, and let $P_{\alpha\beta} = \mathbb{P}[(A_t, B_t) = (\alpha, \beta) \mid Y_{t,E'} = \sigma]$. Set $a = P_{1,1}$, $b = P_{1,-1}$, $c = P_{-1,1}$, $d = P_{-1,-1}$. Since $Y_{t,f} \perp\!\!\!\perp A_t, Y_{t,E'}|B_t$, one can explicitly calculate

$$\begin{aligned}&\mathbb{E}[\mathbb{E}[A_t|Y_t]^2|Y_{t,E'}]\\&\quad=\biggl(\frac{((a(1-t)+b(1+t))-(c(1-t)+d(1+t)))^2}{2((a(1-t)+b(1+t))+(c(1-t)+d(1+t)))}\\&\qquad+\frac{((a(1+t)+b(1-t))-(c(1+t)+d(1-t)))^2}{2((a(1+t)+b(1-t))+(c(1+t)+d(1-t)))}\biggr).\end{aligned}$$

Plugging this in and simplifying, if $b=d=0$ or $a=c=0$, then $h(t;\sigma)=0$, which is nondecreasing because it is constant. Otherwise we get $h(t;\sigma)=\frac{16(ad-bc)^2}{1-t^2(a-b+c-d)^2}$, which is nondecreasing on [0, 1] because $(a-b+c-d)^2<(a+b+c+d)^2=1$. This proves the claim. □

In order to see the relationship between Theorem 6.1 and Theorem 3.6, we define the "absolute value" of the output of a symmetric edge channel:

DEFINITION 6.3. Given a symmetric edge channel Q with output alphabet $\mathcal{Y}$ and symmetry transformation $T:\mathcal{Y}\to\mathcal{Y}$, we define the absolute value $|\cdot|_T:\mathcal{Y}\to 2^{\mathcal{Y}}$ by

$$|y|_T=\{y,T(y)\}.$$

The definition is motivated by viewing T as a sign-flipping transformation (which it is in the BSC and AWGN cases). Notice that since $T^2=\mathrm{id}$, $|y|_T=|T(y)|_T$ for all $y\in\mathcal{Y}$.

We are now ready to prove Theorem 3.6.

PROOF OF THEOREM 3.6. For each edge channel $Q_e(\cdot|\cdot)$, let T_e be the symmetry transformation such that $Q_e(T_e(\cdot)|-1)=Q_e(\cdot|+1)$, and $T_e^2=1$. Define $Z_e=|Y_e|_T$. We claim that by the symmetry property of the edge channels, $X\perp\!\!\!\perp Z$. Because of this, $\mathcal{L}(X,Y|Z)$, the law of (X,Y) conditioned on Z, is almost surely the law of a spin synchronization instance (X',Y') on G, where each of the channels Q'_e is binary-valued (either Y_e or $T_e(Y_e)$).[6] Explicitly, for $z\in Z_e$,

$$Q'_e(z|+1)=\frac{dQ_e(z|+1)}{d(Q_e(z|+1)+Q_e(T_e(z)|+1))},$$

and by the symmetry property this equals

$$\frac{dQ_e(T_e(z)|-1)}{d(Q_e(T_e(z)|-1)+Q_e(z|-1))}=Q'_e(T_e(z)|-1).$$

So Q'_e is a binary symmetric channel. Hence, proving Theorem 3.6 when all the edge channels are BSC yields the general bound:

$$\begin{aligned}&(\text{Since } Z \text{ is a function of } Y)\quad I_2(X_u;X_v|Y)\\&\qquad=\mathbb{E}_Z[I_2(X_u;X_v|Y,Z)]\\&(\text{Replacing } Q \text{ with } Q' \text{ for each realization of } Z)\\&\qquad=\mathbb{E}_Z[I_2(X'_u;X'_v|Y',Z)]\end{aligned}$$

[6] If for some edge $e=(i,j)$, we have the corner case $Y_e=T_e(Y_e)$, then the output channel Q'_e is single-valued. However, in this case the channel Q'_e gives no information about $X_i\cdot X_j$, and therefore we may equivalently view it as a binary symmetric channel with flip probability $1/2$.

$$\begin{aligned} &\text{(By the BSC edge channel case, Theorem 6.1)} && \\ & && \leq \mathbb{E}_Z[\mathrm{conn}_{G,\gamma_Z}(u,v)] \\ &\text{(Since } \gamma = \mathbb{E}_Z[\gamma_Z]\text{)} && = \mathrm{conn}_{G,\gamma}(u,v). \end{aligned}$$

Here, $\gamma_Z(i,j) = I_2(X_i, X_j | Y_{ij}, Z_{ij})$. In particular, for the last equality, we can let P be a bond percolation on G such that each edge e is independently open with probability $\gamma_Z(e)$. Then the probability that u and v are connected by an open path is $\mathbb{E}_Z[\mathrm{conn}_{G,\gamma_Z}(u,v)]$. We can also calculate this probability in a different way, noticing that each edge (i,j) in P is in fact independently open with probability $\mathbb{E}_Z[\gamma_Z(i,j)] = I_2(X_i; X_j \mid Y_{ij}, Z_{ij}) = \gamma(i,j)$, where the independence occurs because the entries of Z are all independent, since they are independent of each other given X, and Z is independent of X. Hence, there is an open path in P connecting u and v with probability $\mathrm{conn}_{G,\gamma}(u,v)$. This proves that $\mathbb{E}_Z[\mathrm{conn}_{G,\gamma_Z}(u,v)] = \mathrm{conn}_{G,\gamma}(u,v)$.

So, to conclude the argument, it suffices to prove the claim that X is independent of Z. For this, it is sufficient to prove that for each $e = (i,j) \in E(G)$, $X_i \cdot X_j$ is independent of Z_{ij}. This is true, because given $z \in Z_{ij}$ the relative likelihood that $X_i \cdot X_j = 1$ versus $X_i \cdot X_j = -1$ is

$$\frac{d(Q_e(z|+1) + Q_e(T_e(z)|+1))}{d(Q_e(z|-1) + Q_e(T_e(z)|-1))} = \frac{d(Q_e(z|+1) + Q_e(T_e(z)|+1))}{d(Q_e(T_e(z)|+1) + Q_e(z|+1))} = 1,$$

by the symmetry property of T_e. □

We now prove the corollaries to the theorem. In order to prove subadditivity over paths (Corollary 3.7), we will need the following.

LEMMA 6.4. *Suppose (X,Y) is a spin synchronization instance on a path P with endpoints u and v. Then*

$$\mathrm{conn}_{P,\gamma}(u,v) = I_2(X_u; X_v | Y_{E(P)}).$$

PROOF.

$$\begin{aligned} & && \mathrm{conn}_{P,\gamma}(u,v) = \prod_{(i,j)\in E(P)} I_2(X_i; X_j | Y_{ij}) \\ &\text{(Prop. A.4)} && = \prod_{(i,j)\in E(P)} I_2(X_i \cdot X_j; Y_{ij}) \\ &\text{(Prop. A.5)} && = I_2\Bigl(\prod_{(i,j)\in E(P)} X_i \cdot X_j; Y_{E(P)}\Bigr) \\ & && = I_2(X_u \cdot X_v; Y_{E(P)}) \\ &\text{(Prop. A.4)} && = I_2(X_u; X_v | Y_{E(P)}). \end{aligned}$$

□

PROOF OF COROLLARY 3.7. The corollary follows from Theorem 3.6, the union bound $\mathrm{conn}_{G,\gamma}(u,v) \leq \sum_{P \in \mathcal{P}_G(u,v)} \mathrm{conn}_{P,\gamma}(u,v)$, and Lemma 6.4. □

We also extend Theorem 3.6 to bound the information that the edge labels Y and a set X_S of vertex label give about another vertex label X_u:

PROOF OF COROLLARY 3.8. Create a "virtual" vertex w and construct the graph G' with $V(G') = V(G) \cup w$ and $E(G') = E(G) \cup \{(v,w) : v \in S\}$. Let Q' be edge channels such that $Q'_e = Q_e$ for all $e \in E(G)$, and

$$Q'_{(v,w)}(y \mid x) = \delta(x = y)$$

for all $v \in S$. Draw (X', Y') from $P_{G',Q'}$. Since $P_{G',Q'}$ is a symmetric graphical channel, by Theorem 3.6,

$$\text{(Theorem 3.6)} \qquad I_2(X'_u; X'_w \mid Y') \le \operatorname{conn}_{G',\gamma'}(u, w)$$

$$(18) \qquad = \operatorname{conn}_{G,\gamma}(u, S),$$

where $\gamma' : E(G') \to [0, 1]$ is defined by $\gamma'(i, j) = I_2(X'_i; X'_j \mid Y'_{(i,j)})$ for all $(i, j) \in E(G')$, and $\gamma : E(G) \to [0, 1]$ is analogous. Line (18) follows because $\gamma(e) = \gamma'(e)$ for all $e \in E(G)$, and $\gamma'(e) = 1$ for all $e \in E(G') \setminus E(G)$. Finally, note that since Q_e is noiseless for each $e \in E(G') \setminus E(G)$, by data-processing

$$I_2(X'_u; X'_w \mid Y') = I_2(X'_u; X'_S \mid Y') = I_2(X_u; X_S \mid Y). \qquad \square$$

APPENDIX: CHI-SQUARED MUTUAL INFORMATION

In this appendix, we define the Chi-squared mutual information, I_2, and prove Propositions 3.1.

A.1. *f*-divergences and *f*-mutual informations.

f-divergences. Given two probability distributions μ and ν over a probability space Ω such that $\mu \ll \nu$ (i.e., μ is absolutely continuous with respect to ν), and given convex $f : (0, \infty) \to \mathbb{R}$ such that $f(1) = 0$ and f is strictly convex at 1, we may define the f-divergence

$$D_f(\mu \parallel \nu) \equiv \int_\Omega f\left(\frac{d\mu}{d\nu}\right) d\nu.$$

Here $\frac{d\mu}{d\nu}$ denotes the Radon–Nikodym derivative. (When Ω is finite, $\frac{d\mu}{d\nu}(x) = \frac{\mu(x)}{\nu(x)}$ for all $x \in \Omega$.) f-divergences were introduced in [12].

f-mutual informations. Given variables A, B with joint distribution $\nu_{A,B}$ on $\mathcal{A} \times \mathcal{B}$, and marginal distributions ν_A on $\mathcal{A}$, ν_B on $\mathcal{B}$, the f-mutual information between them is given by

$$I_f(A; B) \equiv D_f(\nu_{A,B} \parallel (\nu_A \times \nu_B)).$$

I_f is nonnegative, and zero if and only if A and B are independent. Thus, we can take it as a measure of the degree of independence of the variables A and B: the higher the mutual information, the more "correlated" the variables are, and the more information they give about each other.

Moreover, the f-mutual information also has the following well-known "data-processing" property (see [11], e.g.):

PROPOSITION A.1. *for A, B, C such that A is independent of C given B,*

$$(19) \qquad I_f(A; C) \le I_f(A; B).$$

In particular, if C is a deterministic function of B, then $I_f(A; C) \le I_f(A; B)$.

A.2. Definition and basic properties of I_2.

DEFINITION A.2. The Chi-squared mutual information, I_2, is the f-mutual information, I_f, with $f(t) = (t - 1)^2$.

PROPOSITION A.3. *Let A, U be jointly-distributed random variables, with $U \in \{-1, +1\}$. Then*

$$I_2(A; U) = \frac{\mathrm{Var}[\mathbb{E}[U|A]]}{\mathrm{Var}[U]}.$$

In particular, if $U \sim \mathrm{Rad}(1/2)$, then

$$I_2(A; U) = \mathbb{E}[\mathbb{E}[U|A]^2].$$

PROOF. Letting ν_Z denote the distribution of Z, and Ω denote the sample set of A,

$$\begin{aligned}
I_2(A; U) &= \int_{\Omega \times \{-1,+1\}} \left(\frac{d(\nu_{A,U})}{d(\nu_A \times \nu_U)} - 1\right)^2 d(\nu_A \times \nu_U) \\
&= \int_\Omega \sum_{u \in \{-1,+1\}} \nu_U(u) \left(\frac{1}{\nu_U(u)} \cdot \frac{d(\nu_{A,U}(\cdot, u))}{d\nu_A(\cdot)} - 1\right)^2 d\nu_A \\
&= \int_\Omega \sum_{u \in \{-1,+1\}} \frac{1}{\nu_U(u)} \left(\frac{d(\nu_{A,U}(\cdot, u))}{d\nu_A(\cdot)} - \nu_U(u)\right)^2 d\nu_A.
\end{aligned}$$

So, since

$$\begin{aligned}
\frac{d(\nu_{A,U}(\cdot, 1))}{d\nu_A(\cdot)} - \nu_U(1) &= \left(1 - \frac{d(\nu_{A,U}(\cdot, -1))}{d\nu_A(\cdot)}\right) - (1 - \nu_U(-1)) \\
&= -\left(\frac{d(\nu_{A,U}(\cdot, -1))}{d\nu_A(\cdot)} - \nu_U(-1)\right)
\end{aligned}$$

ν_A-almost everywhere, we have

$$\begin{aligned}
I_2(A; U) &= \left(\sum_{u \in \{-1,+1\}} \frac{1}{\nu_U(u)}\right) \cdot \int_\Omega \left(\frac{d(\nu_{A,U}(\cdot, 1))}{d\nu_A(\cdot)} - \nu_U(1)\right)^2 d\nu_A \\
&= \frac{4}{\mathrm{Var}[U]} \cdot \int_\Omega (\mathbb{P}[U = 1|A] - \mathbb{P}[U = 1])^2 d\nu_A \\
&= \frac{4\,\mathrm{Var}[\mathbb{P}[U = 1|A]]}{\mathrm{Var}[U]} = \frac{4\,\mathrm{Var}[\mathbb{E}[(U/2)|A]]}{\mathrm{Var}[U]} = \frac{\mathrm{Var}[\mathbb{E}[U|A]]}{\mathrm{Var}[U]}.
\end{aligned}$$

When $U \sim \mathrm{Rad}(1/2)$, we have $\mathbb{E}[U] = 0$ and $\mathrm{Var}[U] = 1$, so $I_2(A; U) = \mathbb{E}[\mathbb{E}[U|A]^2]$. □

PROOF OF PROPOSITION 3.1. Let (X, Y) be a synchronization instance. We wish to show that $I_2(X_u; X_W, Y) = I_2(X_u; X_W|Y)$.

$$\begin{aligned}
&\text{(Prop. A.3)} & I_2(X_u; X_W, Y) &= \mathbb{E}[\mathbb{E}[X_u|X_W, Y]^2] \\
& & &= \mathbb{E}[\mathbb{E}[\mathbb{E}[X_u|X_W, Y]^2|Y]] \\
&\text{(Prop. A.3)} & &= I_2(X_u; X_W|Y)
\end{aligned}$$

The last step uses $X_u|Y \sim \mathrm{Rad}(1/2)$, because $X_u \perp\!\!\!\perp Y$. □

PROPOSITION A.4. *Let (X, Y) be a spin synchronization instance. Then $I_2(X_u X_v; Y) = I_2(X_u; X_v, Y)$.*

PROOF.

$$\begin{aligned} &\text{(Data-processing)} && I_2(X_u; X_v, Y) = I_2(X_u X_v; X_v, Y) \\ &\text{(Since } X_u X_v, Y \perp\!\!\!\perp X_v) && \qquad = I_2(X_u X_v; Y). \end{aligned}$$ □

PROOF OF PROPOSITION 3.2. By Proposition 3.1 it is equivalent to show that

$$I_2(X_u; X_v, Y) = \mathbb{E}_Y\big[\mathbb{E}_X[X_u \cdot X_v \mid Y]^2\big].$$

This is true by Propositions A.3 and A.4. □

A.3. Series multiplicativity and parallel subadditivity of I_2.

PROPOSITION A.5. *Let $U, V, W \overset{i.i.d.}{\sim} \mathrm{Rad}(1/2)$. Let A be the output of a channel on UW, and let B be the output of a channel on WV. Then*

$$I_2(UV; A, B) = I_2(UW; A) I_2(WV; B).$$

PROOF.

$$\begin{aligned} & && I_2(UV; A, B) \\ &\text{(Prop. A.3)} && = \mathbb{E}\big[\mathbb{E}[UV|A, B]^2\big] \\ &\text{(Since } W^2 = 1) && = \mathbb{E}\big[\mathbb{E}[UWWV|A, B]^2\big] \\ &\text{(Using } UW \perp\!\!\!\perp WV|A, B) && = \mathbb{E}\big[\mathbb{E}[UW|A, B]^2 \mathbb{E}[WV|A, B]^2\big] \\ &\text{(Using } UW \perp\!\!\!\perp B|A \text{ and } WV \perp\!\!\!\perp A|B) && = \mathbb{E}\big[\mathbb{E}[UW|A]^2 \mathbb{E}[WV|B]^2\big] \\ &\text{(Using } A \perp\!\!\!\perp B.) && = \mathbb{E}\big[\mathbb{E}[UW|A]^2\big] \mathbb{E}\big[\mathbb{E}[WV|B]^2\big] \\ &\text{(Prop. A.3)} && = I_2(UW; A) I_2(WV; B). \end{aligned}$$ □

LEMMA A.6. *Let $U \sim \mathrm{Rad}(1/2)$. Let A and B be the outputs of two independent channels on U. Then $I_2(U; A, B) \leq I_2(U; A) + I_2(U; B)$.*

PROOF. Let $\nu_{A,B,U}$ denote the joint distribution of A, B, U. For simplicity, we prove the lemma when A, B are discrete. For any two random variables C, D with joint distribution $\nu_{C,D}$, $I_2(C; D) = D_{(1-1/t)}(\nu_C \nu_D || \nu_{C,D})$, where $D_{(1/1-t)}$ is the $(1 - 1/t)$-divergence, so

$$\begin{aligned} & I_2(U; A, B) - \big(I_2(U; A) + I_2(U; B)\big) \\ &\quad = \mathbb{E}\bigg[\bigg(\frac{\nu_{A,B,U}}{\nu_{A,B}\nu_U} - 1\bigg) - \bigg(\frac{\nu_{A,U}}{\nu_A \nu_U} - 1\bigg) - \bigg(\frac{\nu_{B,U}}{\nu_B \nu_U} - 1\bigg)\bigg] \end{aligned}$$

$$= \mathbb{E}\bigg[\bigg(\frac{\nu_{U|A,B}}{\nu_{U|A}} - 1\bigg)\bigg(\frac{\nu_{A,U}}{\nu_A \nu_U} - 1\bigg)\bigg] \tag{20}$$

$$+ \mathbb{E}\bigg[\bigg(\frac{\nu_{U|A,B}}{\nu_{U|A}} - 1\bigg) - \bigg(\frac{\nu_{B,U}}{\nu_B \nu_U} - 1\bigg)\bigg]. \tag{21}$$

We claim Terms (20) and (21) are ≤ 0, which implies the lemma statement.

We rewrite Term (20), using

$$\frac{\nu_{A,U}(a,u)}{\nu_A(a)\nu_U(u)} - 1 = 2\nu_{U|A}(u|a) - 1 = \mathbb{E}[U|A=a]\cdot u,$$

$$\begin{aligned}(20) &= \mathbb{E}\left[\left(\frac{\nu_{U|A,B}}{\nu_{U|A}} - 1\right)\cdot \mathbb{E}[U|A]\cdot U\right]\\ &= \mathbb{E}\left[\mathbb{E}[U|A]\cdot\left(U\cdot\frac{\nu_{U|A,B}}{\nu_{U|A}} - U\right)\right]\\ &= \mathbb{E}\left[\mathbb{E}[U|A]\cdot\left(U\cdot\frac{\nu_{U|A,B}}{\nu_{U|A}} - \mathbb{E}[U|A]\right)\right].\end{aligned}$$

Define

$$t_a := \sum_b \frac{\nu_{B,U|A}(b,1|a)\nu_{B,U|A}(b,-1|a)}{\nu_{B,U|A}(b,1|a)+\nu_{B,U|A}(b,-1|a)}.$$

Note

$$0 \le t_a \le \nu_{U|A}(1|a)\nu_{U|A}(-1|a) \tag{22}$$

by the subadditivity of $f(x,y) = xy/(x+y)$ for $x, y \ge 0$. (In particular, for all $a, b, c, d \ge 0$, $f(a,b) + f(c,d) \le f(a+c, b+d)$.) So

$$\begin{aligned}&\mathbb{E}\left[U\cdot\frac{\nu_{U|A,B}}{\nu_{U|A}}\middle|A\right]\\ &\quad= \sum_u \frac{u}{\nu_{U|A}}\sum_b \frac{\nu_{B,U|A}\nu_{B,U|A}}{\nu_{B|A}}\\ &\quad= \sum_u \frac{u}{\nu_{U|A}}\sum_b\left(\frac{(\nu_{B|A} - (\nu_{B|A} - \nu_{B,U|A}))\nu_{B,U|A}}{\nu_{B|A}}\right)\\ &\quad= \sum_u \frac{u}{\nu_{U|A}}\left(-t_A + \sum_b \nu_{B,U|A}\right) = \sum_u \frac{u}{\nu_{U|A}}(\nu_{U|A} - t_A)\\ &\quad= -\sum_u u\frac{t_A}{\nu_{U|A}}\\ &\quad= -\frac{t_A}{\nu_{U|A}(1|A)} + \frac{t_A}{\nu_{U|A}(-1|A)}\\ &\quad= \frac{t_A(\nu_{U|A}(1|A) - \nu_{U|A}(-1|A))}{\nu_{U|A}(1|A)\nu_{U|A}(-1|A)}\\ &\quad= \frac{t_A}{\nu_{U|A}(1|A)\nu_{U|A}(-1|A)}\cdot\mathbb{E}[U|A]\\ &\quad= c_A\mathbb{E}[U|A],\end{aligned}$$

for some $0 \le c_A \le 1$ by (22). Thus, $(20) = \mathbb{E}[U|A]^2(c_A - 1) \le 0$, as desired.

Now we bound Term (21).

$$\text{(Using } B \perp\!\!\!\perp A|U) \qquad (21) = \mathbb{E}\left[\left(\frac{\nu_{B,U}}{\nu_{B|A}\nu_U} - 1\right) - \left(\frac{\nu_{B,U}}{\nu_B\nu_U} - 1\right)\right]$$

$$\left(\text{Since } \mathbb{E}\left[\frac{\nu_B}{\nu_{B|A}} - 1\right] = 0\right) \qquad = \mathbb{E}\left[\left(\frac{\nu_A\nu_B}{\nu_{A,B}} - 1\right)\left(\frac{\nu_{B,U}}{\nu_B\nu_U} - 1\right)\right]$$

$$= \sum_{a,b,u} \nu_{A,B,U}\left(\frac{\nu_B}{\nu_{B|A}} - 1\right)\left(\frac{\nu_{B|U}}{\nu_B} - 1\right).$$

For compactness, write $\alpha_a = \nu_{A|U}(a|1)$, $\beta_a = \nu_{A|U}(a|-1)$, $\gamma_b = \nu_{B|U}(b|1)$, $\delta_b = \nu_{B|U}(b|-1)$:

$$\begin{aligned}(21) = \sum_{a,b} & \frac{\alpha_a\gamma_b}{2}\left(\frac{(\gamma_b+\delta_b)/2}{(\alpha_a\gamma_b+\beta_a\delta_b)/(\alpha_a+\beta_a)} - 1\right)\left(\frac{\gamma_b}{(\gamma_b+\delta_b)/2} - 1\right) \\ & + \frac{\beta_a\delta_b}{2}\left(\frac{(\gamma_b+\delta_b)/2}{(\alpha_a\gamma_b+\beta_a\delta_b)/(\alpha_a+\beta_a)} - 1\right)\left(\frac{\delta_b}{(\gamma_b+\delta_b)/2} - 1\right) \\ = & \sum_b\left(-\frac{(\gamma_b-\delta_b)^2}{4(\gamma_b+\delta_b)}\right)\sum_a\left((\alpha_a-\beta_a)\frac{\alpha_a\gamma_b-\beta_a\delta_b}{\alpha_a\gamma_b+\beta_a\delta_b}\right).\end{aligned}$$

We conclude by using

$$\sum_a(\alpha_a-\beta_a) = \sum_a \nu_{A|U}(a|1) - \nu_{A|U}(a|-1) = 1 - 1 = 0,$$

so

$$\begin{aligned}\sum_a(\alpha_a-\beta_a)\left(\frac{\alpha_a\gamma_b-\beta_a\delta_b}{\alpha_a\gamma_b+\beta_a\delta_b}\right) &= \sum_a(\alpha_a-\beta_a)\left(\frac{\alpha_a\gamma_b-\beta_a\delta_b}{\alpha_a\gamma_b+\beta_a\delta_b} + 1\right) \\ &= \sum_a(\alpha_a-\beta_a)\left(\frac{2\alpha_a\gamma_b}{\alpha_a\gamma_b+\beta_a\delta_b}\right) \\ \text{(The inequality is term-wise)} \qquad &\geq \sum_a(\alpha_a-\beta_a)\left(\frac{2\gamma_b}{\gamma_b+\delta_b}\right) \\ &= 0.\end{aligned}$$

Since $(-\frac{(\gamma_b-\delta_b)^2}{4(\gamma_b+\delta_b)}) \leq 0$ for all b, this proves that Term (21) ≤ 0. □

We also note the following fact, used throughout Section 4:

PROPOSITION A.7. *Let $(A_n, U_n)_{n=1}^\infty$ be a sequence of jointly-distributed random variables, where each $U_n \sim \mathrm{Rad}(1/2)$. If $I_2(A_n; U_n) \to 0$, then for any estimators $\hat{U}_n(A_n) \in \{+1, -1\}$, we have $\mathbb{P}[U_n = \hat{U}_n] \to 1/2$.*

PROOF. By the data-processing inequality (Proposition A.1) and Proposition A.3,

$$\begin{aligned} I_2(A_n; U_n) &\geq I_2(\hat{U}_n; U_n) \\ &= \mathbb{E}_{\hat{U}_n}\big[\mathbb{E}_{U_n}[U_n \mid \hat{U}_n]^2\big] \\ &= \sum_{i\in\{-1,+1\}}\left(\frac{(\mathbb{P}[U_n = \hat{U}_n = i] - \mathbb{P}[U_n \neq \hat{U}_n = i])^2}{\mathbb{P}[\hat{U}_n = i]}\right) \\ &\geq \sum_{i\in\{-1,+1\}}(\mathbb{P}[U_n = \hat{U}_n = i] - \mathbb{P}[U_n \neq \hat{U}_n = i])^2.\end{aligned}$$

Since $I_2(A_n; U_n) \to 0$, we have $\mathbb{P}[U_n = \hat{U}_n = i] - \mathbb{P}[U_n \neq \hat{U}_n = i] \to 0$ for $i \in \{+1, -1\}$, and adding these up $\mathbb{P}[U_n = \hat{U}_n] - \mathbb{P}[U_n \neq \hat{U}_n] \to 0$, as desired. □

A.4. I_2 versus I_{KL}.

DEFINITION A.8. The KL-mutual information, I_{KL}, is the f-mutual information, I_f, with $f(t) = t\log_2 t$.

LEMMA A.9 (I_2 vs. I_{KL}). *Let A, U be jointly-distributed random variables, where $U \sim$ Rad$(1/2)$. Then*

$$\frac{1}{2} I_2(A; U) \le I_{\mathrm{KL}}(A; U) \le I_2(A; U).$$

Lemma A.9 is well known (e.g., see [17]), and follows from the inequalities

$$\frac{x^2}{2} \le \frac{1+x}{2}\lg(1-x) + \frac{1-x}{2}\lg(1+x) \le x^2.$$

Acknowledgments. The first author was supported in part by NSF CAREER Award CCF-1552131 and in part by NSF Center for the Science of Information, CCF-0939370.

The second author was supported in part by NSF Center for the Science of Information, CCF-0939370.

REFERENCES

[1] ABBE, E. (2017). Community detection and stochastic block models: Recent developments. *J. Mach. Learn. Res.* **18** Paper No. 177, 86. MR3827065

[2] ABBE, E., BANDEIRA, A. S., BRACHER, A. and SINGER, A. (2014). Decoding binary node labels from censored edge measurements: Phase transition and efficient recovery. *IEEE Trans. Netw. Sci. Eng.* **1** 10–22. MR3349181 https://doi.org/10.1109/TNSE.2014.2368716

[3] ABBE, E., MASSOULIÉ, L., MONTANARI, A., SLY, A. and SRIVASTAVA, N. (2017). Group Synchronization on Grids. Available at arXiv:1706.08561.

[4] ABBE, E. and MONTANARI, A. (2015). Conditional random fields, planted constraint satisfaction, and entropy concentration. *Theory Comput.* **11** 413–443. MR3446024 https://doi.org/10.4086/toc.2015.v011a017

[5] ALAOUI, A. E. and KRZAKALA, F. (2018). Estimation in the spiked Wigner model: A short proof of the replica formula.

[6] BANKS, J. and MOORE, C. (2016). Information-theoretic thresholds for community detection in sparse networks. Available at arXiv:1601.02658.

[7] BANKS, J., MOORE, C., NEEMAN, J. and NETRAPALLI, P. (2016). Information-theoretic thresholds for community detection in sparse networks. *Proc. of COLT*.

[8] BANKS, J., MOORE, C., VERSHYNIN, R., VERZELEN, N. and XU, J. (2018). Information-theoretic bounds and phase transitions in clustering, sparse PCA, and submatrix localization. *IEEE Trans. Inform. Theory* **64** 4872–4994. MR3819345 https://doi.org/10.1109/tit.2018.2810020

[9] CHIN, P., RAO, A. and VU, V. (2015). Stochastic block model and community detection in the sparse graphs: A spectral algorithm with optimal rate of recovery. Available at arXiv:1501.05021..

[10] COJA-OGHLAN, A., KRZAKALA, F., PERKINS, W. and ZDEBOROVÁ, L. (2018). Information-theoretic thresholds from the cavity method. *Adv. Math.* **333** 694–795. MR3818090 https://doi.org/10.1016/j.aim.2018.05.029

[11] COVER, T. M. and THOMAS, J. A. (1991). *Elements of Information Theory. Wiley Series in Telecommunications*. Wiley, New York. MR1122806 https://doi.org/10.1002/0471200611

[12] CSISZÁR, I. (1967). Information-type measures of difference of probability distributions and indirect observations. *Studia Sci. Math. Hungar.* **2** 299–318. MR0219345

[13] DESHPANDE, Y., ABBE, E. and MONTANARI, A. (2017). Asymptotic mutual information for the balanced binary stochastic block model. *Inf. Inference* **6** 125–170. MR3671474 https://doi.org/10.1093/imaiai/iaw017

[14] DOBRUŠIN, R. L. (1968). The problem of uniqueness of a Gibbsian random field and the problem of phase transitions. *Funct. Anal. Appl.* **2** 302–312. MR0250631

[15] DONOHO, D. L., MALEKI, A. and MONTANARI, A. (2009). Message-passing algorithms for compressed sensing. *Proc. Natl. Acad. Sci. USA* **106** 18914–18919.

[16] ERDŐS, P. and RÉNYI, A. (1960). On the evolution of random graphs. *Magy. Tud. Akad. Mat. Kut. Intéz. Közl.* **5** 17–61. MR0125031

[17] EVANS, W., KENYON, C., PERES, Y. and SCHULMAN, L. J. (2000). Broadcasting on trees and the Ising model. *Ann. Appl. Probab.* **10** 410–433. MR1768240 https://doi.org/10.1214/aoap/1019487349

[18] GRIMMETT, G. (1999). *Percolation*, 2nd ed. *Grundlehren der Mathematischen Wissenschaften* **321**. Springer, Berlin. MR1707339 https://doi.org/10.1007/978-3-662-03981-6

[19] GUO, D., SHAMAI, S. and VERDÚ, S. (2005). Mutual information and minimum mean-square error in Gaussian channels. *IEEE Trans. Inform. Theory* **51** 1261–1282. MR2241490 https://doi.org/10.1109/TIT.2005.844072

[20] HEIMLICHER, S., LELARGE, M. and MASSOULIÉ, L. (2012). Community detection in the labelled stochastic block model. Available at arXiv:1209.2910.

[21] JAVANMARD, A., MONTANARI, A. and RICCI-TERSENGHI, F. (2016). Phase transitions in semidefinite relaxations. *Proc. Natl. Acad. Sci. USA* **113** E2218–E2223. MR3494080 https://doi.org/10.1073/pnas.1523097113

[22] KESTEN, H. and STIGUM, B. P. (1966). A limit theorem for multidimensional Galton–Watson processes. *Ann. Math. Stat.* **37** 1211–1223. MR0198552 https://doi.org/10.1214/aoms/1177699266

[23] LELARGE, M., MASSOULIÉ, L. and XU, J. (2015). Reconstruction in the labelled stochastic block model. *IEEE Trans. Netw. Sci. Eng.* **2** 152–163. MR3453283 https://doi.org/10.1109/TNSE.2015.2490580

[24] MOSSEL, E., NEEMAN, J. and SLY, A. (2012). Stochastic block models and reconstruction. Available at arXiv:1202.1499 [math.PR].

[25] PERRY, A., WEIN, A. S. and BANDEIRA, A. S. (2016). Statistical limits of spiked tensor models. Available at arXiv:1612.07728.

[26] PERRY, A., WEIN, A. S., BANDEIRA, A. S. and MOITRA, A. (2018). Optimality and sub-optimality of PCA I: Spiked random matrix models. *Ann. Statist.* **46** 2416–2451. MR3845022 https://doi.org/10.1214/17-AOS1625

[27] POLYANSKIY, Y. and WU, Y. (2018). Application of information-percolation method to reconstruction problems on graphs. Available at arXiv:1806.04195.

[28] POLYANSKIY, Y. and WU, Y. (2017). Strong data-processing inequalities for channels and Bayesian networks. In *Convexity and Concentration. IMA Vol. Math. Appl.* **161** 211–249. Springer, New York. MR3837272

[29] SAADE, A., KRZAKALA, F., LELARGE, M. and ZDEBOROVÁ, L. (2015). Spectral detection in the censored block model. Available at arXiv:1502.00163.

[30] STAM, A. J. (1959). Some inequalities satisfied by the quantities of information of Fisher and Shannon. *Inf. Control* **2** 101–112. MR0109101