Non-linear Log-Sobolev Inequalities for the Potts Semigroup and Applications to Reconstruction Problems

Communications in

Mathematical Physics

Check for

Yuzhou Gu^D, Yury Polyanskiy

IDSS, LIDS, and Department of EECS, MIT, Cambridge, MA 02139, USA. E-mail: yuzhougu@mit.edu; yp@mit.edu

Received: 13 November 2022 / Accepted: 15 September 2023

Published online: 20 October 2023 – \odot The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

Abstract: Consider the semigroup of random walk on a complete graph, which we call the Potts semigroup. Diaconis and Saloff-Coste (Ann Appl Probab 6(3):695–750, 1996) computed the maximum ratio between the relative entropy and the Dirichlet form, obtaining the constant α_2 in the 2-log-Sobolev inequality (2-LSI). In this paper, we obtain the best possible *non-linear* inequality relating entropy and the Dirichlet form (i.e., *p*-NLSI, $p \ge 1$). As an example, we show $\alpha_1 = 1 + \frac{1+o(1)}{\log q}$. Furthermore, *p*-NLSIs allow us to conclude that for $q \ge 3$, distributions that are not a product of identical distributions can have slower speed of convergence to equilibrium, unlike the case q = 2. By integrating the 1-NLSI we obtain new strong data processing inequalities (SDPI), which in turn allows us to improve results of Mossel and Peres (Ann Appl Probab 13(3):817–844, 2003) on reconstruction thresholds for Potts models on trees. A special case is the problem of reconstructing color of the root of a *q*-colored tree given knowledge of colors of all the leaves. We show that to have a non-trivial reconstruction probability the branching number of the tree should be at least

$$\frac{\log q}{\log q - \log(q-1)} = (1 - o(1))q\log q.$$

This recovers previous results (of Sly in Commun Math Phys 288(3):943–961, 2009 and Bhatnagar et al. in SIAM J Discrete Math 25(2):809–826, 2011) in (slightly) more generality, but more importantly avoids the need for any coloring-specific arguments. Similarly, we improve the state-of-the-art on the weak recovery threshold for the stochastic block model with q balanced groups, for all $q \ge 3$. To further show the power of our method, we prove optimal non-reconstruction results for a broadcasting on trees model with Gaussian kernels, closing a gap left open by Eldan et al. (Combin Probab Comput 31(6):1048–1069, 2022). These improvements advocate information-theoretic methods as a useful complement to the conventional techniques originating from the statistical physics.

Contents

Introduction	770
Non-linear <i>p</i> -Log-Sobolev Inequalities for the Potts Semigroup	779
2.1 Non-linear <i>p</i> -log-Sobolev inequality for $p > 1$	779
2.2 Non-linear 1-log-Sobolev inequality	783
2.3 Input-restricted non-linear SDPI for Potts channels	785
2.4 Behavior for $q \to \infty$	790
Product Spaces	792
3.1 Tensorization	793
3.2 Linear piece	795
3.3 Edge-isoperimetric inequalities	800
Non-reconstruction for Broadcasting on Trees	801
Potts Model on a Tree	805
5.1 Binary symmetric channel	806
5.2 Random coloring	806
Stochastic Block Model	809
6.1 Impossibility of weak recovery via information percolation	809
6.2 Impossibility of weak recovery via Potts model	810
Input-Unrestricted Contraction Coefficient of Potts Channels	813
An Upper Bound for Input-Restricted Contraction Coefficient for Potts Chan-	
nels	817
Non-convexity of Certain Functions	821
Concavity of Log-Sobolev Coefficients	824
Non-reconstruction for Broadcasting with a Gaussian Kernel	825
	IntroductionNon-linear <i>p</i> -Log-Sobolev Inequalities for the Potts Semigroup2.1 Non-linear <i>p</i> -log-Sobolev inequality for $p > 1$ 2.2 Non-linear 1-log-Sobolev inequality2.3 Input-restricted non-linear SDPI for Potts channels2.4 Behavior for $q \rightarrow \infty$ Product Spaces3.1 Tensorization3.2 Linear piece3.3 Edge-isoperimetric inequalitiesNon-reconstruction for Broadcasting on TreesPotts Model on a Tree5.1 Binary symmetric channel5.2 Random coloringStochastic Block Model6.1 Impossibility of weak recovery via information percolation6.2 Impossibility of weak recovery via Potts modelInput-Unrestricted Contraction Coefficient of Potts ChannelsAn Upper Bound for Input-Restricted Contraction Coefficient for Potts ChannelsAn Upper Bound for Input-Restricted Contraction Coefficient for Potts ChannelsNon-convexity of Certain FunctionsConcavity of Log-Sobolev CoefficientsNon-reconstruction for Broadcasting with a Gaussian Kernel

1. Introduction

Log-Sobolev inequalities Log-Sobolev inequalities (LSIs) are a class of inequalities bounding the rate of convergence of a Markov semigroup to its stationary distribution. They upper bound certain relative entropy (KL divergence) functions via a multiple of the Dirichlet form.

Let us introduce some standard notions. Let \mathcal{X} be a finite alphabet and $K : \mathcal{X} \to \mathcal{X}$ be a Markov kernel. For $x, y \in \mathcal{X}$, let K(x, y) denote the transition probability from x to y. Let L = K - I. We consider the semigroup $(T_t)_{t\geq 0}$, where $T_t = \exp(tL)$. Let π be a stationary measure for the semigroup. For $f, g : \mathcal{X} \to \mathbb{R}$, the Dirichlet form is defined by

$$\mathcal{E}(f,g) := -\mathbb{E}_{\pi}[(Lf)g] = -\sum_{x,y\in\mathcal{X}} L(x,y)f(y)g(x)\pi(x).$$
(1)

For non-zero $f : \mathcal{X} \to \mathbb{R}_{\geq 0}$, the relative entropy is defined by

$$\operatorname{Ent}_{\pi}(f) := \mathbb{E}_{\pi}\left[f \log \frac{f}{\mathbb{E}_{\pi}[f]}\right] = \mathbb{E}_{\pi}[f]D(\pi^{(f)} \| \pi)$$
(2)

where $\pi^{(f)}$ is the distribution defined as $\pi^{(f)}(x) = \frac{f(x)\pi(x)}{\mathbb{E}_{\pi}[f]}$, and $D(\mu \| v)$ is the Kullback–Leibler divergence¹

$$D(\mu \| \nu) := \int \log\left(\frac{d\mu}{d\nu}\right) d\mu.$$
(3)

For p > 1, we say the semigroup $(T_t)_{t\geq 0}$ admits *p*-log-Sobolev inequality (*p*-LSI), if for some constant α_p , for all non-zero non-negative real functions f on \mathcal{X} , we have

$$\operatorname{Ent}_{\pi}(f) \leq \frac{1}{\alpha_p} \mathcal{E}\left(f^{\frac{1}{p}}, f^{1-\frac{1}{p}}\right).$$
(4)

For p = 1, we define 1-LSI as

$$\operatorname{Ent}_{\pi}(f) \le \frac{1}{\alpha_1} \mathcal{E}(f, \log f).$$
(5)

The case p = 2 is the standard log-Sobolev inequality, originally studied in Gross [22]. The case p = 1 is studied also under the name "modified log-Sobolev inequality" (e.g. [8,21]).

The relationship between 1-LSI and semigroup convergence can be seen from the following identity

$$\frac{d}{dt}|_{t=0}\operatorname{Ent}_{\pi}(T_t f) = -\mathcal{E}(f, \log f).$$
(6)

Therefore

$$\operatorname{Ent}_{\pi}(T_t f) \le \exp(-\alpha_1 t) \operatorname{Ent}_{\pi}(f) \tag{7}$$

which corresponds to a property of T_t to exponentially fast relax to equilibrium (in the sense of relative entropy).

Polyanskiy and Samorodnitsky [43] introduced non-linear *p*-log-Sobolev inequalities (*p*-NLSI), a finer description of the relationship between relative entropy and Dirichlet forms. They used these inequalities to derive improved hypercontractivity inequalities and the sharp uncertainty principle on the boolean hypercube. For $p \ge 1$, we say the semigroup satisfies *p*-NLSI if for some non-negative function² $\Phi_p : \mathbb{R}_{\ge 0} \to \mathbb{R}_{\ge 0}$, for all non-zero $f : \mathcal{X} \to \mathbb{R}_{\ge 0}$, we have

$$\frac{\operatorname{Ent}_{\pi}(f)}{\mathbb{E}_{\pi}[f]} \leq \Phi_p\left(\frac{\mathcal{E}\left(f^{\frac{1}{p}}, f^{1-\frac{1}{p}}\right)}{\mathbb{E}_{\pi}[f]}\right),\tag{8}$$

where for $p = 1, \mathcal{E}\left(f^{\frac{1}{p}}, f^{1-\frac{1}{p}}\right)$ should be replaced with $\mathcal{E}(f, \log f)$.

Non-linear p-log-Sobolev inequalities imply the ordinary p-log-Sobolev inequalities for

$$\alpha_p = \inf_{x>0} \frac{x}{\Phi_p(x)}.$$
(9)

¹ Throughout this paper, log means natural logarithm.

² [43] requires the function Φ_p to be concave. We do not make this assumption initially, however to extend these inequalities to product semigroups concavification will be necessary – see Sect. 3.

When Φ_p is concave, this can be further simplified to $\alpha_p = (\Phi'_p(0))^{-1}$. Mossel et al. [38] proved that for reversible (π, K) ,

$$\frac{p^2(p'-1)}{(p')^2(p-1)}\alpha_p \le \alpha_{p'} \le \alpha_p \tag{10}$$

for $1 < p' \le p \le 2$. We discuss some general facts about dependence of α_p and Φ_p on p in Appendix D.

Potts semigroup We focus on the simplest Markov semigroup, corresponding to the random walk on a complete graph. The Markov kernel is $K(x, y) = \frac{1}{q-1}\mathbb{1}\{x \neq y\}$, where $q = |\mathcal{X}|$. In the following, we always assume $\mathcal{X} = [q]$. We call it the Potts semigroup, because every operator T_t in the semigroup is a ferromagnetic Potts channel. Its stationary distribution π is uniform on \mathcal{X} and its Dirichlet form is rescaled covariance:

$$\mathcal{E}(f,g) = \frac{q}{q-1} \operatorname{Cov}_{\pi}(f,g).$$
(11)

The Potts channel $PC_{\lambda} : [q] \to [q]$ for $\lambda \in \left[-\frac{1}{q-1}, 1\right]$ is defined by

$$PC_{\lambda}(x, y) = \begin{cases} \lambda + \frac{1-\lambda}{q}, & \text{if } x = y, \\ \frac{1-\lambda}{q}, & \text{if } x \neq y. \end{cases}$$
(12)

We parametrize them by λ because it is the second largest eigenvalue of PC_{λ}. We say the channel is ferromagnetic if $\lambda > 0$; antiferromagnetic if $\lambda < 0$. One can see that T_t in the Potts semigroup is exactly PC_{exp($-\frac{q}{q-1}t$).}

Diaconis and Saloff-Coste [14] computed the 2-log-Sobolev constant

$$\alpha_2 = \frac{q-2}{(q-1)\log(q-1)}.$$
(13)

They observed that the infimum of the ratio $\frac{\mathcal{E}(f,f)}{\operatorname{Ent}_{\pi}(f^2)}$ is achieved at a two-valued function f, i.e., f takes exactly two values. In fact, the infimum is achieved at a function f where f(1) = q - 1 and f(i) = 1 for $i \neq 1$. For $p \neq 2$, it seems hard to give a closed-form expression for α_p . Goel [21] proved that

$$\frac{q}{q-1} \le \alpha_1 \le \left(1 + \frac{4}{\log(q-1)}\right) \frac{q}{q-1},\tag{14}$$

where the upper bound is by using a two-valued function f, where f(1) = q + 1 and f(i) = 1 for $i \neq 1$. Bobkov and Tetali [8] also discussed bounds on α_1 and α_2 , proving that

$$\alpha_1 \ge \frac{q}{q-1} + \frac{2}{\sqrt{q-1}}.$$
(15)

These computations lead to the guess that for all p, the best possible p-LSI constant α_p for the Potts semigroup is achieved at a two-valued function. In Sect. 2, we prove that this is true, and in fact true for p-NLSIs for the Potts semigroup: For fixed $\frac{\text{Ent}_{\pi}(f)}{\mathbb{E}_{\pi}[f]}$, the unique function (up to scalar multiplication) of the form $f(1) \ge f(2) = \cdots = f(q)$

minimizes $\frac{\mathcal{E}\left(f^{\frac{1}{p}}, f^{1-\frac{1}{p}}\right)}{\mathbb{E}_{\pi}[f]}$. As a result we get the sharpest *p*-NLSIs for the Potts semigroup for all $p \ge 1$.

We define a useful function $\psi : [0, 1] \to \mathbb{R}$ as follows.

$$\psi(x) := \log q + x \log x + (1 - x) \log \frac{1 - x}{q - 1}.$$
(16)

Note that $\psi(x)$ is the KL divergence between $\left(x, \frac{1-x}{q-1}, \dots, \frac{1-x}{q-1}\right)$ and $\pi = \text{Unif}([q])$. Simple computation shows that ψ is non-negative, convex, $\psi\left(\frac{1}{q}\right) = 0$, strictly decreasing on $\left[0, \frac{1}{q}\right]$, strictly increasing on $\left[\frac{1}{q}, 1\right]$, and takes value in $[0, \log q]$. For p > 1, define $\xi_p : [0, 1] \to \mathbb{R}$ as

$$\xi_p(x) = \frac{q}{q-1} \left(1 - \frac{1}{q} \left(x^{\frac{1}{p}} + (q-1) \left(\frac{1-x}{q-1} \right)^{\frac{1}{p}} \right) \left(x^{1-\frac{1}{p}} + (q-1) \left(\frac{1-x}{q-1} \right)^{1-\frac{1}{p}} \right) \right). \tag{17}$$

Define $\xi_1 : [0, 1] \to \mathbb{R}$ as

$$\xi_1(x) = \frac{1}{q-1} \left(-\log x - (q-1)\log \frac{1-x}{q-1} + q\left(x\log x + (1-x)\log \frac{1-x}{q-1}\right) \right).$$
(18)

Let $f_x : [q] \to \mathbb{R}_{\geq 0}$ be the two-valued function defined as $f_x(1) = qx$, $f(i) = \frac{q(1-x)}{q-1}$ for $2 \le i \le q$. Then we have $\psi(x) = \operatorname{Ent}_{\pi}(f_x)$, $\xi_p(x) = \mathcal{E}\left(f_x^{\frac{1}{p}}, f_x^{1-\frac{1}{p}}\right)$ for p > 1, and $\xi_1(x) = \mathcal{E}(f_x, \log f_x)$.

For $p \ge 1$, define $b_p : [0, \log q] \to \mathbb{R}$ as³

$$b_p(\psi(x)) = \xi_p(x) \tag{19}$$

for
$$x \in \left[\frac{1}{q}, 1\right]$$
, where ψ is defined in (16). In other words, b_p maps $\operatorname{Ent}_{\pi}(f_x)$ to $\mathcal{E}\left(f_x^{\frac{1}{p}}, f_x^{1-\frac{1}{p}}\right)$ (or $\mathcal{E}(f_x, \log f_x)$ when $p = 1$) for $x \in \left[\frac{1}{q}, 1\right]$.

Theorem 1 (*p*-NLSI for Potts semigroup). Fix $p \ge 1$. The Potts semigroup satisfies *p*-NLSI with $\Phi_p = b_p^{-1}$, where b_p is defined in (19). Furthermore, this is the best possible *p*-NLSI.

In other words, for any $c \in [0, \log q]$, among all functions $f : [q] \to \mathbb{R}_{\geq 0}$ with $\mathbb{E}_{\pi}[f] = 1$, $Ent_{\pi}(f) = c$, there is a unique (up to permuting the alphabet) minimizer of $\mathcal{E}\left(f^{\frac{1}{p}}, f^{1-\frac{1}{p}}\right)$ ($\mathcal{E}(f, \log f)$ for p = 1), and it is of form $\left(qx, \frac{q(1-x)}{q-1}, \cdots, \frac{q(1-x)}{q-1}\right)$ with $x \in \left[\frac{1}{q}, 1\right]$.

³ In the case q = 2, b_p differs from [43] by a constant factor due to a different parametrization of the semigroup.

In particular, we have

$$\alpha_p = \inf_{x \in \left(\frac{1}{q}, 1\right]} \frac{\xi_p(x)}{\psi(x)}.$$
(20)

By tensorizing *p*-NLSIs, we can derive facts about distributions that converge the slowest under the product semigroup $(T_t^{\otimes n})_{t\geq 0}$.

Corollary 2 (Extremal distributions for the product semigroup, informal). *Consider the product semigroup* $(T_t^{\otimes n})_{t\geq 0}$ with invariant distribution $\pi^{\otimes n}$.

- (i) For q = 2 and any $c \in [0, \log q]$, among all distributions with entropy cn, the slowest speed of convergence to equilibrium is achieved at a product of identical distributions.
- (ii) For $q \ge 3$, there exists $c \in [0, \log q]$ such that for n large enough, among all distributions with entropy cn, the slowest speed of convergence to equilibrium is not achieved at any distribution that is a product of identical distributions.

See Prop. 27 for a formal statement. This shows a curious difference between the binary case and the non-binary case.

Furthermore, as a corollary of our 1-NLSI, we derive the second order behavior of α_1 as q goes to ∞ .

Proposition 3. For $q \ge 3$, we have

$$\frac{q}{q-1}\left(1+\frac{1}{\log q}\right) \le \alpha_1 \le \frac{q}{q-1}\left(1+\frac{1+o(1)}{\log q}\right).$$
(21)

Strong data processing inequalities The data processing inequality⁴ states that $I(U; Y) \le I(U; X)$ for any Markov chain $U \to X \to Y$, i.e., we cannot gain information by going through a channel. It is natural to think that when the channel $X \to Y$ is noisy, we should strictly lose information, i.e., I(U; Y) < I(U; X). In fact, we have $I(U; Y) \le \eta I(U; X)$, where the constant η depends only on the channel $P_{Y|X}$ (and possibly also on the distribution of X), but not on the distribution of U. Such inequalities are called strong data processing inequalities (SDPIs). We distinguish between the inequalities that depend on the distribution P_X and that are independent of it. For a Markov kernel W and a distribution ν we define

$$\eta_{\text{KL}}(W) = \sup_{\mu,\nu: 0 < D(\mu \| \nu) < \infty} \frac{D(\mu W \| \nu W)}{D(\mu \| \nu)},$$
(22)

$$\eta_{\text{KL}}(\nu, W) = \sup_{\mu: 0 < D(\mu \| \nu) < \infty} \frac{D(\mu W \| \nu W)}{D(\mu \| \nu)}.$$
(23)

We call $\eta_{\text{KL}}(W)$ the input-unrestricted contraction coefficient and $\eta_{\text{KL}}(v, W)$ the inputrestricted contraction coefficient. It can be shown, e.g. [12,44,46], that we also have alternative characterizations: $\eta_{\text{KL}}(W)$ is the smallest constant such that for any Markov chain $U \to X \to Y$ with $P_{Y|X} = W$, we have

$$I(U;Y) \le \eta_{\mathrm{KL}}(W)I(U;X),\tag{24}$$

⁴ We recall that for a pair of random variables *X*, *Y* the mutual information between *X* and *Y* is defined as $I(X; Y) = D(P_{X,Y} || P_X P_Y)$.

and $\eta_{\text{KL}}(v, W)$ is the smallest constant such that for any Markov chain $U \to X \to Y$ with $P_{Y|X} = W$ and $P_X = v$, we have

$$I(U;Y) \le \eta_{\mathrm{KL}}(\nu, W)I(U;X). \tag{25}$$

Even more generally, some channels can be shown to satisfy (for any Markov chain $U \to X \to Y$ with $P_{Y|X}$ as before, arbitrary U, fixed or arbitrary P_X)

$$I(U;Y) \le s(I(U;X)), \tag{26}$$

for some non-linear function *s*. See Raginsky [46] for more background on SDPIs and their relationship with LSIs.

From (7) and Prop. 3 we obtain

$$\eta_{\mathrm{KL}}(\pi, \mathrm{PC}_{\lambda}) \le \lambda^{\frac{q-1}{q}\alpha_1} = \lambda^{1 + \frac{1+o(1)}{\log q}}$$
(27)

for $\lambda \in [0, 1]$ as $q \to \infty$. It turns out that 1-NLSI can be seen as an infinitesimal version of the non-linear SDPIs (see [43, Theorem 2]). Thus, we can prove the best possible non-linear SDPI for the Potts channels.

Theorem 4 (Non-linear SDPI for Potts channel). *Fix* $\lambda \in \left[-\frac{1}{q-1}, 1\right]$. *Define* s_{λ} : $[0, \log q] \rightarrow \mathbb{R}$ *as*

$$s_{\lambda}(\psi(x)) = \psi\left(\lambda x + \frac{1-\lambda}{q}\right),$$
 (28)

for $x \in \left[\frac{1}{q}, 1\right]$, where ψ is defined in (16). Let \hat{s}_{λ} be the concave envelope of s_{λ} . For any Markov chain $U \to X \to Y$ where X has uniform distribution and $X \to Y$ is the Potts channel PC_{λ} , we have

$$I(U;Y) \le \hat{s}_{\lambda}(I(U;X)).$$
⁽²⁹⁾

In particular, we have

$$\eta_{KL}(\pi, PC_{\lambda}) = \sup_{x \in \left(\frac{1}{q}, 1\right]} \frac{\psi\left(\lambda x + \frac{1-\lambda}{q}\right)}{\psi(x)}.$$
(30)

Furthermore, this is the best possible non-linear SDPI for Potts channels, in the sense that for any $c \in [0, \log q]$, there exists a Markov chain $U \to X \to Y$ where $P_X = Unif([q])$, $P_{Y|X} = PC_{\lambda}$, and I(U; X) = c, such that $I(U; Y) = \hat{s}_{\lambda}(c)$.

In the above theorem, concavification is necessary because one needs concavity when translating from the divergence form to the mutual information form – see Sect. 2.3.

To compare the input-restricted η_{KL} with input-unrestricted one, in Appendix A we compute the exact value of $\eta_{KL}(PC_{\lambda})$, and in Appendix B we prove that

$$\eta_{\mathrm{KL}}(\pi, \mathrm{PC}_{\lambda}) < \eta_{\mathrm{KL}}(\mathrm{PC}_{\lambda}) \tag{31}$$

for $q \ge 3$ and $\lambda \in \left[-\frac{1}{q-1}, 0\right) \cup (0, 1)$. See Sect. 2.4 for discussions on tightness of the bound (27).

In Sect. 3 we extend *p*-NLSIs and SDPIs to product spaces/channels. In these results, the functions \check{b}_p (convexification of b_p) and \hat{s}_{λ} (concavification of s_{λ}) appear naturally. When q = 2, b_p is already convex, and s_{λ} is concave, leading to many good properties for the hypercube and for binary symmetric channels (e.g. Mrs. Gerber's Lemma [51]). However, as shown in Prop. 26, these properties do not hold anymore for $q \ge 3$, implying a different structure of extremal distributions that are the slowest to relax to equilibrium as $t \to \infty$ in $T_t^{\otimes n}$ (see Corollary 2 or Prop. 27).

We have seen in Theorem 1 and Theorem 4 that for the Potts semigroup, the sharpest *p*-NLSIs and non-linear SDPIs are achieved at two-valued functions or distributions. This might be the simplest case of a general phenomenon. Namely, it seems plausible that for the semigroup of random walk on a highly symmetric graph G = (V, E), the sharpest *p*-NLSIs and non-linear SDPIs are achieved at a function *f* of form $f(v) = f_0 (\min_{u \in S} \operatorname{dist}(u, v))$ for some subset *S* of *V* and function $f_0 : \mathbb{Z}_{\geq 0} \to \mathbb{R}_{\geq 0}$, where dist denotes graph distance. We leave this for further research.

Applications One of the implications of NLSIs is improved hypercontractivity inequalities for functions in $[q]^n$ supported on subsets of cardinality $q^{(1-\epsilon)n}$ – this was established generally (for any semigroup) in [43]. Here, we show how NLSIs can be used to close the gap (between functional-analytic proofs and explicit combinatorics of [29]) in the edge-isoperimetric inequality for the $[q]^n$ – see Sect. 3.3.

Similarly, SDPIs have numerous applications. Originally introduced to study certain multi-user data-compression questions in information theory, they have since been adopted in many different scenarios. For example, Evans and Schulman [19] used SD-PIs to investigate fundamental limits of fault tolerant computing. Polyanskiy and Wu [44,45] further developed the idea and related the amount of information transmitted in a directed or undirected graphical model to the percolation probability (existence of an open path) on the same network. Other notable applications include distributed estimation [9,52] and communication complexity [24].

More directly related to our paper is the work of Evans et al. [18] which applied the SDPI (for Potts channels with q = 2) to determine the reconstruction threshold for the Ising model on a tree. We describe a more general class of such questions, which we call the broadcasting on trees (BOT) model.

Consider a tree with a marked root. Each vertex of the tree has a random label in [q], generated as follows: the root label is generated according to some known distribution, and for each vertex, its label is generated from its parent's label, through some channel M. We say reconstruction is possible if given labels of all nodes far away enough from the root, we can guess the root label better than guessing from the initial distribution. Equivalently, reconstruction is possible if the mutual information between the root label and the labels of all nodes distance k away from the root goes to a non-zero limit, as k goes to ∞ . We say reconstruction is impossible otherwise.

Reconstruction problems on trees have been studied for a long time. Kesten and Stigum [27] proved the so-called Kesten–Stigum bound, a reconstruction result based on the second eigenvalue of the channel. Physicists study the problem from a spin glass theoretic perspective (e.g., Mézard and Montanari [35]). Based on careful analysis of the evolution of magnetization, several authors have obtained very tight reconstruction thresholds for various models (e.g., Sly [49,50], Bhatnagar et al. [5], Liu and Ning [30], Mossel et al. [41]).

In this paper, we attack this problem using SDPIs. The method of [18] showed that reconstruction is impossible if $br(T)\eta_{KL}(M) < 1$, where br(T) denotes the branching

number⁵ of *T*. Here we improve their result by considering the input-restricted η_{KL} . We note that the idea of using input-restricted contraction coefficients (but not η_{KL}) has appeared in [20] – see Remark 35 below.

Theorem 5 (Non-reconstruction for BOT). Consider the broadcasting model on a tree T with channel M. Let v be a stationary distribution of M (i.e., vM = v) with full support. Let M^* denote the reverse channel, defined by $v_j M^*(j, i) = v_i M(i, j)$ for all $i, j \in [q]$. If

$$\eta_{KL}(\nu, M^*) br(T) < 1,$$
(32)

then reconstruction is impossible. Here br(T) denotes the branching number of T (Definition 30), and η_{KL} denotes the KL contraction coefficient.

Our method is very simple, is non-asymptotic, works for the branching number, and is often quite tight. Previous results often impose additional restrictions on the tree (e.g., regular tree, or Galton–Watson with Poisson offspring distribution), and some only work when the expected degree is large enough. We discuss in more detail in Sect. 4.

Applying Theorem 5 to the Potts model, we achieve improved non-reconstruction results for Potts models on a tree.

Theorem 6 (Non-reconstruction for Potts models). *Consider the Potts model on a tree T*. *If*

$$\eta_{KL}(\pi, PC_{\lambda}) br(T) < 1, \tag{33}$$

then reconstruction is impossible. Here $\eta_{KL}(\pi, PC_{\lambda})$ is given by (30) and br(T) denotes the branching number of T (Definition 30).

Theorem 6 strictly improves over the explicit bound of Mossel and Peres [39]. In the special case where the channel is the coloring channel ($\lambda = -\frac{1}{q-1}$), Theorem 6 recovers the reconstruction threshold up to the first order, which was previously obtained by Sly [49] and Bhatnagar et al. [6] using more complicated methods. (We note, however, that a major focus of those works was to obtain the lower-order terms.) More detailed analysis is done in Sect. 5.

Last but not least, we consider the problem of weak recovery for the stochastic block model with q communities (q-SBM). In q-SBM, n vertices each receives a uniformly random label in [q], and then a random graph is constructed, such that (1) for two vertices with the same label, there exists an edge with probability $\frac{a}{n}$; (2) for two vertices with different labels, there exists an edge with probability $\frac{b}{n}$. The model is said to have weak recovery, if given the random graph, we can partition the vertices into q parts, such that the partition is correct (up to relabeling the parts) for at least $\left(\frac{1}{q} + \epsilon\right)n$ vertices, for some absolute constant $\epsilon > 0$.

For q = 2, the weak recovery threshold is known: If $(a - b)^2 > 2(a + b)$, weak recovery is possible (Massoulié [34], Mossel et al. [37]); if $(a - b)^2 \le 2(a + b)$, weak recovery is impossible (Mossel et al. [36]). For q = 3, 4, Mossel et al. [41] determined the weak recovery threshold when the expected degree is large enough. For the disassortative regime (a < b), Coja-Oghlan et al. [13] gave a formula for the weak recovery threshold.

⁵ The branching number of a tree roughly measures the growth rate of the tree. For regular trees or Galton–Watson trees, the branching number is almost surely equal to the expected offspring of a vertex. See Definition 30 for a formal definition.

In the assortative regime (a > b), the previous best impossibility result for general q is by Banks et al. [3], which says weak recovery is impossible whenever

$$\frac{(a-b)^2}{a+(q-1)b} < \frac{2q\log(q-1)}{q-1}.$$
(34)

By a standard reduction from the Potts model, we achieve improved impossibility results for weak recovery for the q-stochastic block model.

Theorem 7 (Impossibility of weak recovery for SBM). *Weak recovery for the q-SBM is impossible when*

$$\eta_{KL}(\pi, PC_{\lambda})d < 1, \tag{35}$$

where $d = \frac{a+(q-1)b}{q}$, $\lambda = \frac{a-b}{a+(q-1)b}$, and η_{KL} is given by (30).

For q = 2, we have $\eta_{\text{KL}}(\pi, \text{PC}_{\lambda}) = \lambda^2$ (see Section 5.1), so Theorem 7 implies that weak recovery for the 2-SBM is impossible if $d\lambda^2 = \frac{(a-b)^2}{2(a+b)} < 1$. This matches the weak recovery threshold for the 2-SBM established by Massoulié [34] and Mossel et al. [37]. For general q, using Prop. 42, Theorem 7 implies that weak recovery is impossible when

$$(a-b)^{2} < b \cdot \frac{2q(q-1)\log(q-1)}{q-2} + q(a-b), \qquad a > b$$
(36)

or

$$(a-b)^{2} < a \cdot \frac{2q(q-1)\log(q-1)}{q-2} + (b-a) \cdot \frac{q\log q}{(q-1)(\log q - \log(q-1))}, \quad a < b.$$
(37)

We discuss in more detail, and show that this improves over previous results, in Sect. 6. We also note that this framework can be used to deduce SDPI-based impossibility of weak recovery results for asymmetric SBMs – see Remark 39.

Organization In Sect. 2 we prove the sharpest *p*-NLSIs for the Potts semigroup (Theorem 1), and compute the input-restricted KL divergence contraction coefficients of all Potts channels (Theorem 4).

In Sect. 3 we discuss tensorization of p-NLSIs for the Potts semigroup, and non-linear SDPI for Potts channels.

In Sect. 4 we prove a non-reconstruction result for a general class of broadcast models on trees, based on strong data processing inequalities (Theorem 5).

In Sect. 5 we apply Theorem 5 to the Potts model on a tree (Theorem 6). We show that this improves previous non-reconstruction results. For a special case, the random coloring model on a tree, we obtain non-reconstruction results for arbitrary trees, generalizing previous results which work only for restricted classes of trees.

In Sect. 6, by a standard reduction from the Potts model, we prove impossibility results for weak recovery of the stochastic block model (Theorem 7). This results in improvements for the best known bounds for the q-SBM.

2. Non-linear *p*-Log-Sobolev Inequalities for the Potts Semigroup

In this section, we prove *p*-NLSIs for the Potts semigroup for $p \ge 1$. Because the form of the *p*-LSIs are slightly different for $p \ne 1$ and p = 1, we prove them separately.

Recall our setting. The alphabet is $\mathcal{X} = [q]$ for some positive integer $q \ge 2$. The Potts semigroup is $T_t = \exp(Lt)$ with generator

$$L(x, y) = \begin{cases} -1 & \text{if } x = y, \\ \frac{1}{q-1}, & \text{if } x \neq y. \end{cases}$$
(38)

The stationary distribution is $\pi = \text{Unif}([q])$. The Dirichlet form is

$$\mathcal{E}(f,g) = -\mathbb{E}_{\pi}[(Lf)g] = -\frac{1}{q(q-1)} \left(\sum_{x} f(x)\right) \left(\sum_{y} g(y)\right) + \frac{1}{q-1} \sum_{x} f(x)g(x).$$
(39)

Relative entropy is

$$\operatorname{Ent}_{\pi}(f) = \mathbb{E}_{\pi}\left[f\log\frac{f}{\mathbb{E}_{\pi}[f]}\right].$$
(40)

The non-linear *p*-log-Sobolev inequality says

$$\frac{\operatorname{Ent}_{\pi}(f)}{\mathbb{E}_{\pi}[f]} \le \Phi_p\left(\frac{\mathcal{E}\left(f^{\frac{1}{p}}, f^{1-\frac{1}{p}}\right)}{\mathbb{E}_{\pi}[f]}\right)$$
(41)

for some function Φ_p , where for p = 1, RHS is replaced with $\Phi_p\left(\frac{\mathcal{E}(f,\log f)}{\mathbb{E}_{\pi}[f]}\right)$. Because both sides of the inequality are fixed under scalar multiplication, we can wlog restrict f to be a distribution μ . Then the relative entropy is

$$\operatorname{Ent}_{\pi}(\mu) = \frac{1}{q} D(\mu \| \pi) = \frac{1}{q} (\log q - H(\mu)).$$
(42)

2.1. Non-linear p-log-Sobolev inequality for p > 1 We prove Theorem 1 for p > 1. Before proving the theorem we show the following.

Proposition 8. Fix $r \in (0, 1)$ and $c \in [0, \log q]$. Among all distributions $\mu = (p_1, \ldots, p_q)$ with $H(\mu) = c$, the distribution of form $\mu = \left(x, \frac{1-x}{q-1}, \ldots, \frac{1-x}{q-1}\right)$ with $x \in \left[\frac{1}{q}, 1\right]$ achieves maximum $\sum_i p_i^r$. Furthermore, up to permutation of the alphabet this is the unique maximum-achieving distribution.

Proof. The result for $c \in \{0, \log q\}$ is obvious. In the following, assume that $c \in (0, \log q)$. Write $F(\mu) := \sum_i p_i^r$. The set $\{\mu : H(\mu) = c\}$ is compact, so the maximum value of $F(\mu)$ is achieved at some point $\mu = (p_1, \ldots, p_q)$.

We prove in several steps. In Step 0, we prove that if $p_i = 0$ for some *i*, then there can be at most two different values of p_i 's. In Step 1, we prove that if $p_i > 0$ for all *i*, then there can be at most two different values of p_i 's. In Step 2, we prove that one of the two different values must have multiplicity one, thus finishing the proof of the proposition.

Step 0.

Claim 9. Fix a, b > 0 and $r \in (0, 1)$. Among all solutions $u, v, w \in [0, 1]$ with u + v + w = a and $-u \log u - v \log v - w \log w = b$, the maximum of $u^r + v^r + w^r$ is not achieved at a point where 0 = u < v < w.

Proof. Suppose the maximum is achieved at such a point (u_0, v_0, w_0) where $0 = u_0 < v_0 < w_0$. Extend it to a curve (u, v = v(u), w = w(u)) on $u \in [0, \epsilon)$ for some $\epsilon > 0$, such that u < v < w for all u, satisfying

$$u + v + w = a, \tag{43}$$

$$-u\log u - v\log v - w\log w = b, \tag{44}$$

and

$$v(0) = v_0, w(0) = w_0. \tag{45}$$

We prove that

$$f(u) := u^{r} + v^{r} + w^{r}$$
(46)

decreases as u approaches 0^+ , for small enough u.

By taking derivative of (43) and (44), one can compute that

$$v'(u) = \frac{\log w - \log u}{\log v - \log w}, \quad w'(u) = \frac{\log u - \log v}{\log v - \log w}.$$
(47)

Therefore

$$f'(u) = r \left(u^{r-1} + v^{r-1} v'(u) + w^{r-1} w'(u) \right)$$

= $r \left(u^{r-1} + v^{r-1} \frac{\log w - \log u}{\log v - \log w} + w^{r-1} \frac{\log u - \log v}{\log v - \log w} \right).$ (48)

Because $0 < v_0 < w_0$, the term u^{r-1} dominates the sum, and f'(u) > 0 for small enough u > 0. Therefore the maximum of f is not achieved at u = 0.

By Claim 9, if $p_i = 0$ for some *i*, then there can be at most two different values of p_i 's. **Step 1.**

Claim 10. If $u, v, w \in (0, 1)$ are all different, then

$$\det \begin{pmatrix} 1 \ \log u \ u^{r-1} \\ 1 \ \log v \ v^{r-1} \\ 1 \ \log w \ w^{r-1} \end{pmatrix} \neq 0.$$
(49)

Proof of Claim. Suppose det = 0. Then for some $a, b \in \mathbb{R}$, the equation $x^{r-1} + a \log x = b$ has at least three distinct solutions $x \in (0, 1)$. However

$$\frac{\partial}{\partial x}\left(x^{r-1} + a\log x\right) = (r-1)x^{r-2} + \frac{a}{x}$$
(50)

is smooth on (0, 1), and takes zero at most once. So $x^{r-1} + a \log x$ takes each value at most once on (0, 1). Contradiction.

By Lagrange multipliers, the three vectors

$$\nabla F(\mu) = \left(rp_i^{r-1}\right)_{i \in [q]},\tag{51}$$

$$\nabla H(\mu) = (-1 - \log p_i)_{i \in [q]},$$
 (52)

$$\nabla \sum_{i \in [q]} p_i = \mathbb{1}$$
(53)

should be linear dependent. By Step 0 and Claim 10, there can be at most two different values of p_i 's.

So we can assume that $p_1 = \cdots = p_m = x$, $p_{m+1} = \cdots = p_q = \frac{1-mx}{q-m}$ for some $m \in [q-1], x \in \left(\frac{1}{q}, \frac{1}{m}\right]$.

Step 2. For μ of the above form, we have

$$-H(\mu) = mx \log x + (1 - mx) \log \frac{1 - mx}{q - m},$$
(54)

$$F(\mu) = mx^{r} + (q - m)\left(\frac{1 - mx}{q - m}\right)^{r}.$$
(55)

We smoothly continue both functions so that m can take any real value in [1, q - 1].

Claim 11. For $m \in (1, q - 1]$ and $x \in \left(\frac{1}{q}, \frac{1}{m}\right)$, we have

$$-\frac{\partial}{\partial x}H(\mu) > 0 \tag{56}$$

and

$$\frac{\partial}{\partial x}H(\mu)\frac{\partial}{\partial m}F(\mu) - \frac{\partial}{\partial m}H(\mu)\frac{\partial}{\partial x}F(\mu) > 0.$$
(57)

Proof. We have

$$-\frac{\partial}{\partial x}H(\mu) = m\left(\log x - \log\frac{1-mx}{q-m}\right) > 0,$$
(58)

$$-\frac{\partial}{\partial m}H(\mu) = \frac{1-qx}{q-m} + x\left(\log x - \log\frac{1-mx}{q-m}\right),\tag{59}$$

$$\frac{\partial}{\partial x}F(\mu) = rm\left(x^{r-1} - \left(\frac{1-mx}{q-m}\right)^{r-1}\right),\tag{60}$$

$$\frac{\partial}{\partial m}F(\mu) = x^r + \left(r\frac{1-qx}{1-mx} - 1\right)\left(\frac{1-mx}{q-m}\right)^r.$$
(61)

Let $a = \frac{qx-1}{1-mx}$. Then

$$\begin{aligned} G(\mu) &:= \frac{\partial}{\partial x} H(\mu) \frac{\partial}{\partial m} F(\mu) - \frac{\partial}{\partial m} H(\mu) \frac{\partial}{\partial x} F(\mu) \\ &= (r-1)m \left(x^r - \left(\frac{1-mx}{q-m}\right)^r \right) \left(\log x - \log \frac{1-mx}{q-m} \right) \\ &- rm \frac{1-qx}{q-m} \left(x^{r-1} - \left(\frac{1-mx}{q-m}\right)^{r-1} \right) \\ &= x^{-r} m \left((r-1) \left(1 - (a+1)^{-r} \right) \log(a+1) - r \frac{a}{a+1} \left(1 - (a+1)^{1-r} \right) \right). \end{aligned}$$
(62)

The result then follows from Claim 12.

Claim 12. For all $r \in (0, 1)$ and a > 0 we have

$$(r-1)\left(1-(a+1)^{-r}\right)\log(a+1)-r\frac{a}{a+1}\left(1-(a+1)^{1-r}\right)>0.$$
 (63)

Proof. Let

$$f(a) := (r-1)\left(1 - (a+1)^{-r}\right)\log(a+1) - r\frac{a}{a+1}\left(1 - (a+1)^{1-r}\right).$$
 (64)

Because $\lim_{a\to 0^+} f(a) = 0$, it suffices to prove that f'(a) > 0.

$$f'(a) = (a+1)^{-r-1}(1-r-(a(1-r)+1)\left((a+1)^{r-1}-r\right) - (1-r)r\log(a+1))$$

=: (a+1)^{-r-1}g(a). (65)

Because $\lim_{a\to 0^+} g(a) = 0$, it suffices to prove that g'(a) > 0.

$$g'(a) = \frac{ar(1-r)\left(1-(a+1)^{r-1}\right)}{a+1} > 0.$$
(66)

Now let us return to the proof of Prop. 8. The set of (m, x) where $m \in [1, q - 1]$, $x \in \left(\frac{1}{q}, \frac{1}{m}\right]$, and $H(\mu) = c$ can be parametrized as a curve (m, x = x(m)) for $m \in [1, m_c]$ for some constant m_c . Along the curve, $F(\mu)$ is continuous, and by Claim 11, is decreasing in m. Therefore $F(\mu)$ is maximized at m = 1. This finishes the proof. \Box

Proof of Theorem 1 for p > 1. For a distribution $\mu = (p_1, \ldots, p_q)$, we have

$$\mathcal{E}\left(\mu^{\frac{1}{p}},\mu^{1-\frac{1}{p}}\right) = \frac{1}{q-1}\left(1-\frac{1}{q}\left(\sum_{i}p_{i}^{\frac{1}{p}}\right)\left(\sum_{i}p_{i}^{1-\frac{1}{p}}\right)\right).$$
(67)

By Prop. 8, for fixed value of $\operatorname{Ent}_{\pi}(\mu)$, the unique distribution of the form $\left(x, \frac{1-x}{q-1}, \ldots, \frac{1-x}{q-1}\right)$ with $x \in \left[\frac{1}{q}, 1\right]$ minimizes $\mathcal{E}\left(\mu^{\frac{1}{p}}, \mu^{1-\frac{1}{p}}\right)$. Therefore for any non-zero non-negative f, we have

$$b_p\left(\frac{\operatorname{Ent}_{\pi}(f)}{\mathbb{E}_{\pi}[f]}\right) \leq \frac{\mathcal{E}\left(f^{\frac{1}{p}}, f^{1-\frac{1}{p}}\right)}{\mathbb{E}_{\pi}[f]}.$$
(68)

So *p*-NLSI holds with $\Phi_p = b_p^{-1}$. The statement about optimality is immediate from the above discussions.

2.2. Non-linear 1-log-Sobolev inequality We prove Theorem 1 for p = 1. Before proving the theorem we show the following.

Proposition 13. Fix $0 \le c \le \log q$. Among all distributions $\mu = (p_1, \ldots, p_q)$ with $H(\mu) = c$, the distribution of form $\mu = \left(x, \frac{1-x}{q-1}, \ldots, \frac{1-x}{q-1}\right)$ with $x \in \left[\frac{1}{q}, 1\right]$ achieves maximum $\sum_i \log p_i$. Furthermore, up to permutation of the alphabet this is the unique minimum-achieving distribution.

Proof. The result for $c \in \{0, \log q\}$ is obvious. In the following, assume that $0 < c < \log q$. Write $F(\mu) := \sum_i \log p_i$. The set $\{\mu : H(\mu) = c\}$ is compact, so the maximum value of $F(\mu)$ is achieved at some point $\mu = (p_1, \dots, p_q)$.

We prove in several steps. In Step 0, we prove that $p_i > 0$ for all *i*. In Step 1, we prove that there can be at most two different values of p_i 's. In Step 2, we prove that one of the two different values must have multiplicity one, thus finishing the proof of the proposition.

Step 0. If $p_i = 0$ for some *i*, then $F(\mu) = -\infty$. So $\min_{i \in [q]} p_i > 0$.

Step 1.

Claim 14. If $u, v, w \in (0, 1)$ are all different, then

$$\det \begin{pmatrix} 1 \ \log u \ \frac{1}{u} \\ 1 \ \log v \ \frac{1}{v} \\ 1 \ \log w \ \frac{1}{w} \end{pmatrix} \neq 0.$$
(69)

Proof of Claim. Suppose det = 0. Then for some $a, b \in \mathbb{R}$, the equation $\frac{1}{x} + a \log x = b$ has at least three distinct solutions $x \in (0, 1)$. However, $\frac{\partial}{\partial x} \left(\frac{1}{x} + a \log x\right) = -\frac{1}{x^2} + \frac{a}{x}$ is smooth on (0, 1), and takes zero at most once. So $\frac{1}{x} + a \log x$ takes each value at most once on (0, 1). Contradiction.

By Lagrange multipliers, the three vectors

$$\nabla F(\mu) = \left(\frac{1}{p_i}\right)_{i \in [q]},\tag{70}$$

$$\nabla H(\mu) = (-1 - \log p_i)_{i \in [q]},$$
(71)

$$\nabla \sum_{i \in [q]} p_i = \mathbb{1} \tag{72}$$

should be linear dependent. By Claim 14, there can be at most two different values of p_i 's.

So we can assume that $p_1 = \cdots = p_m = x$, $p_{m+1} = \cdots = p_q = \frac{1-mx}{q-m}$ for some $m \in [q-1], x \in \left(\frac{1}{q}, \frac{1}{m}\right)$.

Step 2. For μ of the above form, we have

$$-H(\mu) = mx \log x + (1 - mx) \log \frac{1 - mx}{q - m},$$
(73)

$$F(\mu) = m \log x + (q - m) \log \frac{1 - mx}{q - m}.$$
(74)

We smoothly continue both functions so that m can take any real value in [1, q - 1].

Claim 15. For
$$m \in (1, q - 1]$$
 and $x \in \left(\frac{1}{q}, \frac{1}{m}\right)$, we have

$$-\frac{\partial}{\partial x}H(\mu) > 0 \tag{75}$$

and

$$\frac{\partial}{\partial x}H(\mu)\frac{\partial}{\partial m}F(\mu) - \frac{\partial}{\partial m}H(\mu)\frac{\partial}{\partial x}F(\mu) > 0.$$
(76)

Proof of Claim. Let $f(x) = \log x - \log \frac{1-mx}{q-m}$. Then we have

$$-\frac{\partial}{\partial x}H(\mu) = mf(x) > 0, \tag{77}$$

$$-\frac{\partial}{\partial m}H(\mu) = \frac{1-qx}{q-m} + xf(x),$$
(78)

$$\frac{\partial}{\partial x}F(\mu) = \frac{m(1-qx)}{x(1-mx)},\tag{79}$$

$$\frac{\partial}{\partial m}F(\mu) = \frac{1-qx}{1-mx} + f(x).$$
(80)

So

$$G(\mu) := \frac{\partial}{\partial x} H(\mu) \frac{\partial}{\partial m} F(\mu) - \frac{\partial}{\partial m} H(\mu) \frac{\partial}{\partial x} F(\mu)$$
$$= m \left(\frac{(1-qx)^2}{x(q-m)(1-mx)} - f(x)^2 \right).$$
(81)

Let $a = \frac{qx-1}{1-mx}$. Then $G(\mu) = m\left(\frac{a^2}{1+a} - \log^2(a+1)\right)$. Because a > 0, we have $G(\mu) > 0$ by Lemma 16.

The set of (m, x) where $m \in [1, q-1], x \in \left(\frac{1}{q}, \frac{1}{m}\right]$, and $H(\mu) = c$ can be parametrized as a curve (m, x = x(m)) for $m \in [1, m_c]$ for some constant m_c . Along the curve, $F(\mu)$ is continuous, and by Claim 15, is decreasing in m. Therefore $F(\mu)$ is maximized at m = 1. This finishes the proof.

Lemma 16. For $a \in \mathbb{R}_{>-1}$, we have $\frac{a^2}{1+a} \ge \log^2(a+1)$. Equality holds only when a = 0.

Proof. We start from the well-known fact that $a \ge \log(a+1)$ for $a \in \mathbb{R}_{>-1}$ (and equality holds only when a = 0). Let $f(a) = a(a+2) - (2a+2)\log(a+1)$. We have f(0) = 0 and $f'(a) = 2(a - \log(a+1)) \ge 0$ for $a \in \mathbb{R}_{>-1}$ (and equality holds only when a = 0). So f is negative on (-1, 1) and positive on $(1, \infty)$.

Let $g(a) = \frac{a^2}{1+a} - \log^2(a+1)$. Clearly g(0) = 0. Because $g'(a) = \frac{f(a)}{(a+1)^2}$, g is decreasing on (-1, 1] and increasing on $[1, \infty)$. So $g(a) \ge 0$ for all $a \in \mathbb{R}_{>-1}$, and equality holds only when a = 0.

Proof of Theorem 1 for p = 1. For a distribution $\mu = (p_1, \ldots, p_q)$, we have

$$\mathcal{E}(\mu, \log \mu) = \frac{1}{q-1} \sum_{i \in [q]} p_i \log p_i - \frac{1}{q(q-1)} \sum_{i \in [q]} \log p_i.$$
 (82)

By Prop. 13, for fixed value of $\operatorname{Ent}_{\pi}(\mu)$, the unique distribution of the form $\left(x, \frac{1-x}{q-1}, \ldots, \frac{1-x}{q-1}\right)$ with $x \in \left[\frac{1}{q}, 1\right]$ minimizes $\mathcal{E}(\mu, \log \mu)$. Therefore for any non-zero non-negative f, we have

$$b_1\left(\frac{\operatorname{Ent}_{\pi}(f)}{\mathbb{E}_{\pi}[f]}\right) \le \frac{\mathcal{E}(f, \log f)}{\mathbb{E}_{\pi}[f]}.$$
(83)

So 1-NLSI holds for $\Phi_1 = b_1^{-1}$. The statement about optimality is immediate from the above discussions.

2.3. *Input-restricted non-linear SDPI for Potts channels* In this section, we prove Theorem 4.

Recall that the subset of Potts channels with $\lambda \ge 0$ (i.e., ferromagnetic Potts channels) forms the Potts semigroup, with $T_t = \text{PC}_{\exp\left(-\frac{q}{q-1}t\right)}$. For semigroups, the optimal 1-NLSI is an "infinitesimal version" of the input-restricted non-linear SDPI. Consequently, by integrating the former we can get the latter. This can be formalized as follows.

Proposition 17. Let $\lambda \in [0, 1]$. Fix $0 \le c \le \log q$. Among all distributions μ with $H(\mu) = c$, the distribution of form $\mu = \left(x, \frac{1-x}{q-1}, \ldots, \frac{1-x}{q-1}\right)$ with $x \in \left[\frac{1}{q}, 1\right]$ achieves minimum $H(\mu PC_{\lambda})$. Furthermore, when $\lambda \notin \{0, 1\}$, up to permutation of the alphabet this is the unique minimum-achieving distribution.

Proof. The result for $\lambda \in \{0, 1\}$ or $c \in \{0, \log q\}$ is obvious. In the following assume that $0 < \lambda < 1$ and $0 < c < \log q$.

Let μ and ν be two distributions with $H(\mu) = H(\nu) = c$, where μ is of form $\left(x, \frac{1-x}{q-1}, \dots, \frac{1-x}{q-1}\right)$ for some $x \in \left[\frac{1}{q}, 1\right]$, and ν is not of this form (up to permuting the alphabet). Define $\mu_t = \mu T_t$ and $\nu_t = \nu T_t$, where $(T_t)_{t\geq 0}$ is the Potts semigroup.

We prove that $H(\mu_t) < H(\nu_t)$ for $t \in (0, \infty)$. Suppose this does not hold. Let $u = \inf\{t > 0 : H(\mu_t) \ge H(\nu_t)\}$. Then we have $H(\nu_u) = H(\mu_u)$ by continuity of semigroup. By Prop. 13, we have

$$\frac{\partial}{\partial t}|_{t=u}H(v_t) = \mathcal{E}(v_u, \log v_u) > \mathcal{E}(\mu_u, \log \mu_u) = \frac{\partial}{\partial t}|_{t=u}H(\mu_t).$$
(84)

If u = 0, then for some $\epsilon > 0$, $H(v_t) > H(\mu_t)$ for $t \in (0, \epsilon)$. If u > 0, then for some $\epsilon > 0$, $H(v_t) < H(\mu_t)$ for $t \in (u - \epsilon, u)$. Both cases lead to contradiction with definition of u. So $H(\mu_t) < H(v_t)$ for $t \in (0, \infty)$. This completes the proof. \Box

Surprisingly, the result also extends beyond the semigroup to all of the Potts channels. Namely, we have the following.

Proposition 18. Let $\lambda \in \left[-\frac{1}{q-1}, 1\right]$. Fix $0 \le c \le \log q$. Among all distributions μ with $H(\mu) = c$, the distribution of form $\mu = \left(x, \frac{1-x}{q-1}, \dots, \frac{1-x}{q-1}\right)$ with $x \in \left[\frac{1}{q}, 1\right]$ achieves minimum $H(\mu PC_{\lambda})$. Furthermore, when $\lambda \notin \{0, 1\}$, up to permutation of the alphabet this is the unique minimum-achieving distribution.

The only difference between Prop. 18 and Prop. 17 is that we allow negative λ in Prop. 18.

Proof. The result for $\lambda \in \{0, 1\}$ or $c \in \{0, \log q\}$ is obvious. In the following assume that $\lambda \notin \{0, 1\}$ and $c \notin \{0, \log q\}$. The set $\{\mu : H(\mu) = c\}$ is compact, so the minimum value of $H(\mu \operatorname{PC}_{\lambda})$ is achieved at some point $\mu = (p_1, \ldots, p_q)$.

We prove in several steps. In Step 0, we prove that if $p_i = 0$ for some *i*, then there can be at most two different values of p_i 's. In Step 1, we prove that if $p_i > 0$ for all *i*, then there can be at most two different values of p_i 's. In Step 2, we prove that one of the two different values must have multiplicity one, thus finishing the proof of the proposition.

Step 0.

Claim 19. Fix a, b, d > 0 and $c \in \mathbb{R}_{>-d} \setminus \{0\}$. Among all solutions $u, v, w \in [0, 1]$ with u + v + w = a and $-u \log u - v \log v - w \log w = b$, the maximum of

$$(cu + d)\log(cu + d) + (cv + d)\log(cv + d) + (cw + d)\log(cw + d)$$
(85)

is not achieved at a point where 0 = u < v < w.

Proof. Suppose the maximum is achieved at such a point (u_0, v_0, w_0) where $0 = u_0 < v_0 < w_0$. Extend it to a curve (u, v = v(u), w = w(u)) on $u \in [0, \epsilon)$ for some $\epsilon > 0$, such that u < v < w for all u, satisfying

 $u + v + w = a, \tag{86}$

$$-u\log u - v\log v - w\log w = b, \tag{87}$$

and $v(0) = v_0$, $w(0) = w_0$.

We prove that

$$f(u) := (cu+d)\log(cu+d) + (cv+d)\log(cv+d) + (cw+d)\log(cw+d)$$
(88)

decreases as u approaches 0^+ for small enough u.

By taking derivative of (86) and (87), one can compute that

$$v'(u) = \frac{\log w - \log u}{\log v - \log w}, \quad w'(u) = \frac{\log u - \log v}{\log v - \log w}.$$
(89)

Therefore

$$f'(u) = c \left(\log(cu+d) + \log(cv+d)v'(u) + \log(cw+d)w'(u) \right)$$

= $c \left(\log(cu+d) + \log(cv+d) \frac{\log w - \log u}{\log v - \log w} + \log(cw+d) \frac{\log u - \log v}{\log v - \log w} \right).$ (90)

Because $0 < v_0 < w_0$, terms involving log *u* dominates the sum. The dominating term is

$$-c\log u \frac{\log(cv+d) - \log(cw+d)}{\log v - \log w} > 0.$$
 (91)

Therefore the maximum of f is not achieved at u = 0.

By Claim 19, if $p_i = 0$ for some *i*, then there can be at most two different values of p_i 's. **Step 1.**

Claim 20. If $u, v, w \in (0, 1)$ are all different, then

$$\det \begin{pmatrix} 1 \ \log u \ \log(\lambda u + \frac{1-\lambda}{q}) \\ 1 \ \log v \ \log(\lambda v + \frac{1-\lambda}{q}) \\ 1 \ \log w \ \log(\lambda w + \frac{1-\lambda}{q}) \end{pmatrix} \neq 0.$$
(92)

Proof of Claim. Suppose det = 0. Then for some $a, b \in \mathbb{R}$, the equation

$$\log\left(\lambda x + \frac{1-\lambda}{q}\right) + a\log x = b \tag{93}$$

has at least three distinct solutions $x \in (0, 1)$. However,

$$\frac{\partial}{\partial x} \left(\log \left(\lambda x + \frac{1 - \lambda}{q} \right) + a \log x \right) = \frac{\lambda}{\lambda x + \frac{1 - \lambda}{q}} + \frac{a}{x}$$
(94)

is smooth on (0, 1), and takes zero at most once. So

$$\log\left(\lambda x + \frac{1-\lambda}{q}\right) + a\log x \tag{95}$$

takes each value at most twice on (0, 1). Contradiction.

By Lagrange multipliers, the three vectors

$$\nabla H(\mu \operatorname{PC}_{\lambda}) = \left(-\lambda \log\left(\lambda p_i + \frac{1-\lambda}{q}\right) - \lambda\right)_{i \in [q]},\tag{96}$$

$$\nabla H(\mu) = (-1 - \log p_i)_{i \in [q]},$$
(97)

$$\nabla \sum_{i \in [q]} p_i = \mathbb{1} \tag{98}$$

should be linear dependent. By Claim 20, there can be at most two different values of p_i 's.

So we can assume that $p_1 = \cdots = p_m = x$, $p_{m+1} = \cdots = p_q = \frac{1-mx}{q-m}$ for some $m \in [q-1], x \in \left(\frac{1}{q}, \frac{1}{m}\right].$ **Step 2.** For μ of the above form, we have

$$-H(\mu) = mx \log x + (1 - mx) \log \frac{1 - mx}{q - m},$$
(99)

$$-H(\mu \operatorname{PC}_{\lambda}) = m \left(\lambda x + \frac{1-\kappa}{q}\right) \log \left(\lambda x + \frac{1-\kappa}{q}\right) + (q-m) \left(\lambda \frac{1-mx}{q-m} + \frac{1-\lambda}{q}\right) \log \left(\lambda \frac{1-mx}{q-m} + \frac{1-\lambda}{q}\right).$$
(100)

We smoothly continue both functions so that *m* can take any real value in [1, q - 1]. **Claim 21.** For $m \in (1, q - 1]$ and $x \in \left(\frac{1}{q}, \frac{1}{m}\right)$, we have

$$-\frac{\partial}{\partial x}H(\mu) > 0 \tag{101}$$

and

$$\frac{\partial}{\partial m}H(\mu)\frac{\partial}{\partial x}H(\mu PC_{\lambda}) - \frac{\partial}{\partial x}H(\mu)\frac{\partial}{\partial m}H(\mu PC_{\lambda}) > 0.$$
(102)

Proof of Claim. Let

0

$$f(x) = \log x - \log \frac{1 - mx}{q - m}.$$
 (103)

Then

$$-\frac{\partial}{\partial x}H(\mu) = mf(x) > 0, \tag{104}$$

$$-\frac{\partial}{\partial m}H(\mu) = \frac{1-qx}{q-m} + xf(x),$$
(105)

$$-\frac{\partial}{\partial x}H(\mu \operatorname{PC}_{\lambda}) = \lambda m f\left(\lambda x + \frac{1-\lambda}{q}\right),\tag{106}$$

$$-\frac{\partial}{\partial m}H(\mu \operatorname{PC}_{\lambda}) = \lambda \frac{1-qx}{q-m} + \left(\lambda x + \frac{1-\lambda}{q}\right) f\left(\lambda x + \frac{1-\lambda}{q}\right), \quad (107)$$

and

$$G(\mu) := \frac{\partial}{\partial m} H(\mu) \frac{\partial}{\partial x} H(\mu \operatorname{PC}_{\lambda}) - \frac{\partial}{\partial x} H(\mu) \frac{\partial}{\partial m} H(\mu \operatorname{PC}_{\lambda})$$

$$= \lambda m \frac{1 - qx}{q - m} \left(f\left(\lambda x + \frac{1 - \lambda}{q}\right) - f(x) \right) - m f(x) f\left(\lambda x + \frac{1 - \lambda}{q}\right) \frac{1 - \lambda}{q}.$$
(108)

$$\frac{\partial}{\partial\lambda} \frac{G(\mu)}{m\lambda f(x)f\left(\lambda x + \frac{1-\lambda}{q}\right)} = \frac{1}{q\lambda^2} + \frac{1-qx}{q-m} \frac{\frac{\partial}{\partial\lambda}f\left(\lambda x + \frac{1-\lambda}{q}\right)}{f\left(\lambda x + \frac{1-\lambda}{q}\right)^2}.$$
(109)

Note that

1.

$$\frac{G(\mu)}{m\lambda f(x)f\left(\lambda x + \frac{1-\lambda}{q}\right)}$$
(110)

is continuous for $\lambda \in \left[-\frac{1}{q-1}, 1\right]$, and takes value 0 at $\lambda = 1$; 2. $m\lambda f(x) f\left(\lambda x + \frac{1-\lambda}{q}\right) \ge 0$ for $\lambda \in \left[-\frac{1}{q-1}, 1\right]$. So we only need to prove that

$$\frac{\partial}{\partial \lambda} \frac{G(\mu)}{m\lambda f(x) f\left(\lambda x + \frac{1-\lambda}{q}\right)} \le 0, \tag{111}$$

i.e.,

$$f\left(\lambda x + \frac{1-\lambda}{q}\right)^2 \le \frac{q\lambda^2(qx-1)}{q-m}\frac{\partial}{\partial\lambda}f\left(\lambda x + \frac{1-\lambda}{q}\right).$$
(112)

Let $y = \lambda x + \frac{1-\lambda}{q}$. Then the above inequality can be rewritten as

$$f(y)^{2} \leq \frac{q\lambda^{2}(qx-1)}{q-m} \frac{\partial y}{\partial \lambda} \frac{\partial f(y)}{\partial y} = \frac{(qy-1)^{2}}{(q-m)y(1-my)}.$$
(113)

This is true by Lemma 16, applied to $a = \frac{qy-1}{1-my}$. Equality holds only when $y = \frac{1}{q}$, which cannot happen for $\lambda \neq 0$.

The set of (m, x) where $m \in [1, q - 1]$, $x \in \left(\frac{1}{q}, \frac{1}{m}\right]$, and $H(\mu) = c$ can be parametrized as a curve (m, x = x(m)) for $m \in [1, m_c]$ for some constant m_c . Along the curve, $H(\mu \text{ PC}_{\lambda})$ is continuous, and by Claim 21, is increasing in m. Therefore $H(\mu \text{ PC}_{\lambda})$ is minimized at m = 1. This finishes the proof.

Proof of Theorem 4. Consider a Markov chain $U \to X \to Y$ where X has uniform distribution, and the channel $X \to Y$ is PC_{λ} . Because P_X and P_Y are both uniform, for any u, we have

$$D(P_{X|U=u} \| P_X) = \log q - H(P_{X|U=u}), \tag{114}$$

$$D(P_{Y|U=u} \| P_Y) = \log q - H(P_{Y|U=u}).$$
(115)

So by Prop. 18 we get

$$D(P_{Y|U=u} \| P_Y) \le s_{\lambda}(D(P_{X|U=u} \| P_X)).$$
(116)

Because \hat{s}_{λ} is the concave envelope of s_{λ} , taking expectation over U we get

$$I(U;Y) = D(P_{Y|U} || P_Y | P_U) \le \hat{s}_{\lambda}(D(P_{X|U} || P_X | P_U)) = \hat{s}_{\lambda}(I(U;X)).$$
(117)

Now we prove optimality. Let $c \in [0, \log q]$. Choose $a, b \in [0, \log q]$ and $u \in [0, 1]$ such that c = (1 - u)a + ub and $\hat{s}_{\lambda}(c) = (1 - u)s_{\lambda}(a) + us_{\lambda}(b)$. Choose $\rho, \tau \in [0, 1]$ such that $\operatorname{Cap}(\operatorname{PC}_{\rho}) = a$ and $\operatorname{Cap}(\operatorname{PC}_{\tau}) = b$, where Cap denotes channel capacity. Define random variable U = (V, Z) such that $Z \sim \operatorname{Ber}(u)$, conditioned on Z = 0, $V \sim \operatorname{PC}_{\rho}(X)$, and conditioned on Z = 1, $V \sim \operatorname{PC}_{\tau}(X)$. One can check that

$$I(U; X) = (1 - u)a + ub = c,$$
(118)

$$I(U;Y) = (1-u)s_{\lambda}(a) + us_{\lambda}(b) = \hat{s}_{\lambda}(c).$$
(119)

This finishes the proof.

2.4. Behavior for $q \to \infty$ When should one use *p*-NLSI instead of *p*-LSI? To get some insights, we consider the case of $q \to \infty$. First, we prove Prop. 3 that $\alpha_1 = 1 + \frac{1+o(1)}{\log q}$.

Proof of Prop. 3. Lower bound. By Theorem 1, we need to show that for all $x \in \left(\frac{1}{q}, 1\right]$, we have

$$\frac{q}{q-1}\left(1+\frac{1}{\log q}\right) \le \frac{\xi_1(x)}{\psi(x)}.$$
(120)

Noting that

$$\xi_1(x) = \frac{q}{q-1} \left(\frac{1}{q} \left(-\log x - (q-1)\log \frac{1-x}{q-1} \right) - \log q + \psi(x) \right),$$
(121)

it suffices to prove that

$$f(x) := \frac{\log q}{q} \left(-\log x - (q-1)\log \frac{1-x}{q-1} \right) - \log^2 q - \psi(x) \ge 0.$$
(122)

We have $f\left(\frac{1}{q}\right) = 0$. So it suffices to prove that $f'(x) \ge 0$ for $x \in \left[\frac{1}{q}, 1\right]$.

$$f'(x) = \frac{\log q}{q} \left(-\frac{1}{x} + \frac{q-1}{1-x} \right) - \left(\log x - \log \frac{1-x}{q-1} \right).$$
(123)

We smoothly continue this function to $\{(q, x) \in \mathbb{R}^2 : q \ge 3, x \in \left\lfloor \frac{1}{q}, 1 \right\}$ and prove that it is non-negative in this region.

$$\frac{\partial}{\partial q}f'(x) = \frac{1 - \log q}{q^2} \left(-\frac{1}{x} + \frac{q - 1}{1 - x} \right) + \frac{\log q}{q} \frac{1}{1 - x} - \frac{1}{q - 1}$$
$$= \frac{(q - 1)\log q + q(qx^2 - x - 1) + 1}{q^2(q - 1)x(1 - x)}.$$
(124)

The numerator is a quadratic function in x, and for fixed q, it is minimized at $x = \frac{1}{q}$, leading to

$$\frac{\partial}{\partial q}f'(x) \ge \frac{(q-1)\log q - q + 1}{q^2(q-1)x(1-x)} = \frac{\log q - 1}{q^2x(1-x)} \ge 0.$$
 (125)

So we only need to prove $f'(x) \ge 0$ for minimum q, i.e., $q = \max\{3, \frac{1}{x}\}$. When $q = \frac{1}{x}$, on can verify that f'(x) = 0. So the only remaining case is q = 3. For q = 3, we prove that f is convex in x, i.e., $f''(x) \ge 0$ for $x \in [0, 1]$.

$$f''(x) = \frac{\log q}{q} \left(\frac{1}{x^2} + \frac{q-1}{(1-x)^2} \right) - \left(\frac{1}{x} + \frac{1}{1-x} \right)$$
$$= \frac{\log q \left((1-x)^2 + (q-1)x^2 \right) - qx(1-x)}{qx^2(1-x)^2}$$
$$= \frac{(q\log q + q)x^2 - (q+2\log q)x + \log q}{qx^2(1-x)^2}.$$
(126)

The numerator is a quadratic function in x, and its discriminant is

$$(q+2\log q)^2 - 4(q\log q + q)\log q = q^2 - 4(q-1)\log^2 q.$$
(127)

When q = 3, the above value is < 0. So $f''(x) \ge 0$ for q = 3 and $x \in [0, 1]$. This finishes the proof of the lower bound.

Upper bound. For the upper bound, we need find $x \in \left(\frac{1}{q}, 1\right]$ such that $\frac{f(x)}{\psi(x)} = o(1)$. Because the upper bound to prove is asymptotic, we assume that q is large enough. Take $x = \frac{2}{\log q}$. Then we have

$$\psi(x) = \log q + \frac{2}{\log q} \log \frac{2}{\log q} + \left(1 - \frac{2}{\log q}\right) \log \frac{1 - \frac{2}{\log q}}{q - 1}$$

= $\log q - \left(1 - \frac{2}{\log q}\right) \log(q - 1) + o(1)$
= $2 + o(1)$ (128)

and

$$f(x) = \frac{\log q}{q} \left(-\log \frac{2}{\log q} - (q-1)\log \frac{1 - \frac{2}{\log q}}{q-1} \right) - \log^2 q - \psi(x)$$

= $\frac{\log q}{q} (q-1) \left(\log(q-1) - \log \left(1 - \frac{2}{\log q}\right) \right) - \log^2 q - 2 + o(1)$
= $\log q \cdot \left(1 + O\left(\frac{1}{q}\right)\right) \cdot \left(\log q + O\left(\frac{1}{q}\right) + \frac{2}{\log q} + O\left(\frac{1}{\log^2 q}\right)\right)$
 $-\log^2 q - 2 + o(1)$
= $o(1).$

Remark 22. Numerical computation suggests $\frac{f(x)}{\psi(x)}$ is minimized at a point $x = \frac{2+o(1)}{\log q}$. This guides our proof of the upper bound in Prop. 3, but we have not attempted to prove this fact.

To understand the case p > 1, let us denote convexification of b_p as \check{b}_p . Then NLSI lower bound, assuming $\mathbb{E}_{\pi}[f] = 1$, gives

$$\mathcal{E}\left(f^{\frac{1}{p}}, f^{1-\frac{1}{p}}\right) \ge \check{b}_p(\operatorname{Ent}_{\pi}(f)).$$
(129)

We see that this improves upon $\alpha_p \cdot \text{Ent}_{\pi}(f)$ the more the larger the entropy. In particular, the maximum improvement happens when $\text{Ent}_{\pi}(f) = \log q$. That is we have for p > 1

$$\alpha_p \le \frac{\ddot{b}_p(x)}{x} \le \frac{1}{\log q}.$$
(130)

Together with (10) and (13), we get $\alpha_p = \Theta\left(\frac{1}{\log q}\right)$ as $q \to \infty$. Numerical computation suggests that $\alpha_p = \frac{1+o(1)}{\log q}$.

At the same time, the improvement given by the 1-NLSI (over 1-LSI) is much stronger, since $b_1(\log q) = \infty$. To summarize, the *p*-NLSI should be preferred for p = 1 or for cases where *q* is small and entropy is large (i.e. functions are highly spiky).

Next, we consider SDPIs and η_{KL} . First, we show that for a fixed $\lambda \ge 0$ we have

$$\eta_{\mathrm{KL}}(\pi, \mathrm{PC}_{\lambda}) = \lambda - \Theta\left(\frac{1}{\log q}\right).$$
 (131)

Indeed, the upper bound is given by (27). For the lower bound we have

$$\eta_{\mathrm{KL}}(\pi, \mathrm{PC}_{\lambda}) \geq \eta_{\min} \coloneqq \frac{\psi\left(\lambda + \frac{1-\lambda}{q}\right)}{\psi(1)} \\ = \frac{\log q + \left(\lambda + \frac{1-\lambda}{q}\right)\log\left(\lambda + \frac{1-\lambda}{q}\right) + \left(1 - \left(\lambda + \frac{1-\lambda}{q}\right)\right)\log\frac{1 - \left(\lambda + \frac{1-\lambda}{q}\right)}{q-1}}{\log q} \\ = \lambda + \frac{\lambda\log\lambda + (1-\lambda)\log(1-\lambda) + o(1)}{\log q}.$$
(132)

On the other hand,

$$\eta_{\mathrm{KL}}(\pi, \mathrm{PC}_{\lambda}) \le \eta_{\mathrm{KL}}(\mathrm{PC}_{\lambda}) \le \eta_{\mathrm{TV}}(\mathrm{PC}_{\lambda}) = \lambda, \tag{133}$$

where η_{TV} is the contraction coefficient for the total variation distance (Dobrushin coefficient, see [12,44,46]).

Notice also that for \hat{s}_{λ} we have generally $\eta_{\min} \leq \frac{\hat{s}_{\lambda}(x)}{x} \leq \eta_{\text{KL}}(\pi, \text{PC}_{\lambda})$. Therefore we have shown that

$$\lim_{q \to \infty} \frac{s_{\lambda}(x)}{x} = \lim_{q \to \infty} \eta_{\text{KL}}(\pi, \text{PC}_{\lambda}) = \lim_{q \to \infty} \eta_{\text{KL}}(\text{PC}_{\lambda}) = \eta_{\text{TV}}(\text{PC}_{\lambda}) = \lambda.$$
(134)

The estimates of information quantities using the more sophisticated tools get improvement over simplistic coupling of at most multiplicative order $\left(1 + \Theta\left(\frac{1}{\log q}\right)\right)$.

Note, however, if λ changes with q (e.g. $\lambda = -\frac{1}{q-1}$), then the improvement over η_{TV} can be as large as a multiplicative factor of $(1 + o(1)) \log q$, as shown in Prop. 33.

3. Product Spaces

In this section we study extensions of *p*-NLSIs and SDPIs to the product semigroup $(T_t^{\otimes n})_{t\geq 0}$ on the product space $[q]^n$ (and product channels $PC_{\lambda}^{\otimes n}$). The general property of tensorization of *p*-NLSI was established in [43, Theorem 1], and thus we only need to concavify the function Φ_p in (8). Similarly, we can show that (non-linear) strong data processing inequalities tensorize if one concavifies function $s(\cdot)$ in (26). Note that both cases, concavification is necessary – see Prop. 23 and Corollary 25.

After showing these extensions to product spaces, we proceed to discussing implications of *p*-NLSI on speed of convergence to equilibrium in terms of $\text{Ent}_{\pi^{\otimes n}}(\nu T_t^{\otimes n})$ and on edge-isoperimetric inequalities.

3.1. Tensorization

Proposition 23. Fix $p \ge 1$. Recall b_p defined in (19). Let \check{b}_p be the convex envelope of b_p . Then p-NLSI holds for the product semigroup $(T_t^{\otimes n})_{t\ge 0}$ with

$$\Phi_{n,p}(x) = n\breve{b}_p^{-1}\left(\frac{x}{n}\right).$$
(135)

Furthermore, for every $c \in [0, \log q]$, there exists a function $f : [q]^n \to \mathbb{R}_{\geq 0}$ with $\frac{Ent_{\pi \otimes n}(f)}{\mathbb{E}_{\pi \otimes n}[f]} = (c + o(1))n \text{ such that } \frac{\mathcal{E}\left(f^{\frac{1}{p}}, f^{1-\frac{1}{p}}\right)}{\mathbb{E}_{\pi \otimes n}[f]} = \left(\check{b}_p(c) + o(1)\right)n \text{ (where for } p = 1,$ $\mathcal{E}\left(f^{\frac{1}{p}}, f^{1-\frac{1}{p}}\right) \text{ should be replaced with } \mathcal{E}(f, \log f)\text{).}$

Proof. The *p*-NLSI follows from Theorem 1 and [43, Theorem 1].

For the second part, choose $a, b \in [0, \log q]$ and $u \in [0, 1]$ such that c = (1-u)a+uband $\check{b}_p(c) = (1-u)b_p(a)+ub_p(b)$. Such a, b, u exist because \check{b}_p is the convex envelope of b_p .

Let $f_a : [q] \to \mathbb{R}_{\geq 0}$ (resp. f_b) be the unique function of form $\left(qx, \frac{q(1-x)}{q-1}, \dots, \frac{q(1-x)}{q-1}\right)$ with $x \in \left[\frac{1}{q}, 1\right]$ satisfying $\frac{\operatorname{Ent}_{\pi}(f)}{\mathbb{E}_{\pi}[f]} = a$ (resp. $\frac{\operatorname{Ent}_{\pi}(f)}{\mathbb{E}_{\pi}[f]} = b$). Note that $\mathbb{E}_{\pi}[f_a] = \mathbb{E}_{\pi}[f_b] = 1$. Let $f : [q]^n \to \mathbb{R}_{\geq 0}$ be defined as

$$f(x) = \left(\prod_{1 \le i \le \lfloor (1-u)n \rfloor} f_a(x_i)\right) \left(\prod_{1 \le i \le \lceil un \rceil} f_b(x_i)\right).$$
(136)

Then $\mathbb{E}_{\pi^{\otimes n}}[f] = 1$ and

$$\operatorname{Ent}_{\pi^{\otimes n}}(f) = \lfloor (1-u)n \rfloor \operatorname{Ent}_{\pi}(f_a) + \lceil un \rceil \operatorname{Ent}_{\pi}(f_b) = (c+o(1))n,$$
(137)

$$\mathcal{E}\left(f^{\frac{1}{p}}, f^{1-\frac{1}{p}}\right) = \lfloor (1-u)n \rfloor \mathcal{E}\left(f_a^{\frac{1}{p}}, f_a^{1-\frac{1}{p}}\right) + \lceil un \rceil \mathcal{E}\left(f_b^{\frac{1}{p}}, f_b^{1-\frac{1}{p}}\right)$$
$$= \left(\check{b}_p(c) + o(1)\right)n, \quad \text{if } p > 1, \tag{138}$$

$$\mathcal{E}(f, \log f) = \lfloor (1 - u)n \rfloor \mathcal{E}(f_a, \log f_a) + \lceil un \rceil \mathcal{E}(f_b, \log f_b)$$
$$= \left(\check{b}_p(c) + o(1)\right)n, \quad \text{if } p = 1.$$
(139)

This finishes the proof.

We show below in Prop. 26 that $\check{b}_p \neq b_p$.

For non-linear SDPI, we first prove a general tensorization result.

Proposition 24. *Fix a probability kernel* $P_{Y|X} : \mathcal{X} \to \mathcal{Y}$ *and a distribution* P_X *on* \mathcal{X} *.*

(i) Suppose for some non-decreasing function $s : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ we have

$$D(Q_Y \| P_Y) \le s(D(Q_X \| P_X)) \tag{140}$$

for all distribution Q_X on \mathcal{X} with $0 < D(Q_X || P_X) < \infty$. Then for all distribution Q_{X^n} on \mathcal{X}^n with $0 < D(Q_{X^n} || P_X^{\otimes n}) < \infty$, we have

$$D(Q_{Y^n} \| P_Y^{\otimes n}) \le n \hat{s} \left(\frac{1}{n} D(Q_{X^n} \| P_X^{\otimes n}) \right), \tag{141}$$

where \hat{s} is the concave envelope of s, and $Q_{Y^n} = P_{Y|X}^{\otimes n} \circ Q_{X^n}$.

(ii) Suppose for some non-decreasing concave function $\hat{s} : [0, \log |\mathcal{X}|] \to \mathbb{R}_{\geq 0}$ we have

$$I(U;Y) \le \hat{s}(I(U;X)) \tag{142}$$

for all Markov chains $U \to X \to Y$ where the distribution of X is P_X . Then for all Markov chains $U \to X^n \to Y^n$ where the distribution of X is $P_X^{\otimes n}$, we have

$$I(U; Y^n) \le n\hat{s}\left(\frac{1}{n}I(U; X^n)\right).$$
(143)

We have separate statements for non-linear SDPI defined via KL divergence and via mutual information, because they are not equivalent in general. It is not hard to show that if KL divergence type non-linear SDPI (140) holds for some function *s*, then mutual information type non-linear SDPI (142) holds for \hat{s} . However, it is not clear what is the best possible KL divergence type non-linear SDPI one can get starting from mutual information type non-linear SDPI. (Note the domain of function *s* would become larger during the translation.)

Proof of Prop. 24. **Proof of (i)**. Perform induction on *n*. The base case n = 1 is trivial. Now consider $n \ge 2$. We have

$$\begin{split} &D(Q_{Y^{n}} \| P_{Y}^{\otimes n}) \\ &= D\left(Q_{Y^{n-1}} \| P_{Y}^{\otimes (n-1)}\right) + D(Q_{Y_{n}|Y^{n-1}} \| P_{Y}|Q_{Y^{n-1}}) \\ &\leq D\left(Q_{Y^{n-1}} \| P_{Y}^{\otimes (n-1)}\right) + D(Q_{Y_{n}|X^{n-1}} \| P_{Y}|Q_{X^{n-1}}) \\ &\leq (n-1)\widehat{s}\left(\frac{1}{n-1}D\left(Q_{X^{n-1}} \| P_{X}^{\otimes (n-1)}\right)\right) + s\left(D(Q_{X_{n}|X^{n-1}} \| P_{X}|Q_{X^{n-1}})\right) \\ &\leq n\widehat{s}\left(\frac{1}{n}D\left(Q_{X^{n-1}} \| P_{X}^{\otimes (n-1)}\right) + \frac{1}{n}D(Q_{X_{n}|X^{n-1}} \| P_{X}|Q_{X^{n-1}})\right) \\ &= n\widehat{s}\left(\frac{1}{n}D\left(Q_{X^{n}} \| P_{X}^{\otimes n}\right)\right). \end{split}$$

The first step is by chain rule. The second step is by Markov chain $Y^{n-1} \to X^{n-1} \to Y_n$ and convexity of KL divergence. For the third step, we bound the first summand using induction hypothesis, and bound the second summand using (140) (note that $Q_{Y_n|X^{n-1}} = P_{Y|X} \circ Q_{X_n|X^{n-1}}$). The fourth step is because \hat{s} is the concave envelope of s. The fifth step is by chain rule. **Proof of** (ii). Perform induction on *n*. The base case n = 1 is trivial. Now consider $n \ge 2$. We have

$$\begin{split} I(U; Y^{n}) &= I(U; Y^{n-1}) + I(U; Y_{n}|Y^{n-1}) \\ &= I(U; Y^{n-1}) + I(U, Y^{n-1}; Y_{n}) \\ &\leq I(U; Y^{n-1}) + I(U, X^{n-1}; Y_{n}) \\ &= I(U; Y^{n-1}) + I(U; Y_{n}|X^{n-1}) \\ &\leq (n-1)\widehat{s} \left(\frac{1}{n-1}I(U; X^{n-1})\right) + \widehat{s} \left(I(U; X_{n}|X^{n-1})\right) \\ &\leq n\widehat{s} \left(\frac{1}{n}I(U; X^{n-1}) + \frac{1}{n}I(U; X_{n}|X^{n-1})\right) \\ &= n\widehat{s} \left(\frac{1}{n}I(U; X^{n})\right). \end{split}$$

The first step is by chain rule. The second step is by chain rule, and that Y_n is independent with Y^{n-1} . The third step is by data processing inequality. The fourth step is by chain rule, and that Y_n is independent with X^{n-1} . For the fifth step, we bound the first summand using induction hypothesis, and bound the second summand using (142). The sixth step is because \hat{s} is concave. The seventh step is by chain rule.

Corollary 25. Recall function s_{λ} defined in Theorem 4. Let Q_{X^n} be a distribution on $[q]^n$ and $Q_{Y^n} = PC_{\lambda}^{\otimes n} \circ Q_{X^n}$. Then we have

$$\frac{1}{n}H(Y^n) \ge \log q - \hat{s}_{\lambda} \left(\log q - \frac{1}{n}H(X^n)\right).$$
(144)

Furthermore, for every $c \in [0, \log q]$, there exists a distribution X^n with $H(X^n) = (c + o(1))n$ such that $\frac{1}{n}H(Y^n) = \log q - \hat{s}_{\lambda}(\log q - c) + o(1)$.

Proof. (144) follows from Prop. 24 and that

$$D(Q_{X^n} \| \pi^{\otimes n}) = n \log q - H(Q_{X^n}).$$
(145)

For the second part, choose $a, b \in [0, \log q]$ and $u \in [0, 1]$ such that c = (1-u)a+uband $\hat{s}_{\lambda}(\log q - c) = (1-u)s_{\lambda}(\log q - a) + us_{\lambda}(\log q - b)$. Such a, b, u exist because \hat{s}_{λ} is the concave envelope of s_{λ} .

Let Q_A (resp. Q_B) be the unique distribution on [q] of form $\left(x, \frac{1-x}{q-1}, \ldots, \frac{1-x}{q-1}\right)$ with $x \in \left[\frac{1}{q}, 1\right]$ and entropy a (resp. entropy b). Now let Q_{X^n} be the distribution $Q_A \times \cdots \times Q_A \times Q_B \times \cdots \times Q_B$, where Q_A appears $\lfloor (1-u)n \rfloor$ times and Q_B appears $\lceil un \rceil$ times. It is easy to see that this distribution satisfies the required properties. \Box

3.2. Linear piece In Prop. 23 and Theorem 4, we make use of convexification of b_p and concavification of s_{λ} . When q = 2, we have $\check{b}_p = b_p$ ([43, Theorem 4 and 6]) and $\hat{s}_{\lambda} = s_{\lambda}$ (known as the Mrs. Gerber's Lemma [51]). However, for $q \ge 3$, the situation is vastly different.

Proposition 26. Recall function $b_p : [0, \log q] \to \mathbb{R}$ defined in Theorem 1 and $s_{\lambda} : [0, \log q] \to \mathbb{R}$ defined in Theorem 4.



Fig. 1. b_p and \check{b}_p for q = 15, p = 2. α_p is achieved at $x^* \approx 2.289$. \check{b}_p is linear for $x \le x^*$ and is equal to b_p for $x \ge x^*$

(*i*) For all
$$q \ge 3$$
 and $p \ge 1$, b_p is not convex near 0.
(*ii*) For all $q \ge 3$, $\lambda \in \left[-\frac{1}{q-1}, 0\right] \cup (0, 1)$, s_{λ} not concave near 0

The proof is deferred to Appendix C. Prop. 26 implies that there is a linear piece near origin in the graph of \tilde{b}_p , $\hat{\Phi}_p$ and \hat{s}_{λ} . See Fig. 1 for an example.

This implies a curious new property distinguishing the Potts semigroup with $q \ge 3$ from its binary cousin and from the Ornstein–Uhlenbeck semigroup. Both of the latter have their *p*-NLSI and SDPI strictly non-linear, which translates into the following fact: among all initial densities f with a given entropy $\text{Ent}_{\pi}(f)$ a product of identical distributions simultaneously maximizes $\text{Ent}_{\pi^{\otimes n}}(T_t^{\otimes n}f)$ for all t. This nice extremal property of product distributions is no longer true for Potts semigroups with $q \ge 3$, due to the fact that b_p is not convex and s_{λ} is not concave.

Proposition 27 (Extremal distributions for the product semigroup). Consider the product semigroup $(T_t^{\otimes n})_{t>0}$ on $[q]^n$ with invariant distribution $\pi^{\otimes n}$.

- (i) If q = 2, then for any $p \ge 1$, $n \ge 1$ and $c \in [0, \log 2]$, among all non-zero functions $f : [q]^n \to \mathbb{R}_{\ge 0}$ with $\frac{Ent_{\pi \otimes n}(f)}{\mathbb{E}_{\pi \otimes n}[f]} = cn$, the minimum of $\frac{\mathcal{E}\left(f^{\frac{1}{p}}, f^{1-\frac{1}{p}}\right)}{\mathbb{E}_{\pi \otimes n}[f]} \left(\frac{\mathcal{E}(f, \log f)}{\mathbb{E}_{\pi \otimes n}[f]}\right)$ for p = 1) is achieved at a function of form $f(x) = \prod_{i \in [n]} g(x_i)$ for some non-zero function $g : [q] \to \mathbb{R}_{\ge 0}$ with $\frac{Ent_{\pi}(g)}{\mathbb{E}_{\pi}[g]} = c$.
- function $g : [q] \to \mathbb{R}_{\geq 0}$ with $\frac{Ent_{\pi}(g)}{\mathbb{E}_{\pi}[g]} = c$. (ii) If $q \geq 3$, then for any $p \geq 1$, there exists $c \in [0, \log q]$ such that for n large enough, among all non-zero functions $f : [q]^n \to \mathbb{R}_{\geq 0}$ with $\frac{Ent_{\pi} \otimes n(f)}{\mathbb{E}_{\pi} \otimes n[f]} = cn$, the

minimum of $\frac{\mathcal{E}\left(f^{\frac{1}{p}}, f^{1-\frac{1}{p}}\right)}{\mathbb{E}_{-\infty n}[f]} \left(\frac{\mathcal{E}(f, \log f)}{\mathbb{E}_{-\infty n}[f]} \text{ for } p = 1\right)$ is not achieved at any functions of form $f(x) = \prod_{i \in [n]}^{n} g(x_i)$, where g is a non-zero function from [q] to $\mathbb{R}_{>0}$.

Consider the product channel $PC_{\lambda}^{\otimes n}: [q]^n \to [q]^n$ with invariant distribution $\pi^{\otimes n}$.

- (iii) If q = 2, then for any $n \ge 1$, $c \in [0, \log 2]$, and $\lambda \in [-1, 1]$, among all distributions v on $[q]^n$ with $D(v \| \pi^{\otimes n}) = cn$, the maximum of $D(v P C_{\lambda}^{\otimes n} \| \pi^{\otimes n})$ is achieved at $v = \mu^{\otimes n}$ for some distribution μ on [q] with $D(\mu \| \pi) = c$.
- (iv) If $q \ge 3$, then for any $\lambda \in \left[-\frac{1}{q-1}, 0\right) \cup (0, 1)$, there exists $c \in [0, \log q]$ such that for n large enough, among all distributions v on $[q]^n$ with $D(v \| \pi^{\otimes n}) = cn$, the maximum of $D(v PC_{\lambda}^{\otimes n} \| \pi^{\otimes n})$ is not achieved at any distributions of form $v = \mu^{\otimes n}$, where μ is a distribution on [q].

Proof. **Proof of** (i). For any non-zero function $f : [q]^n \to \mathbb{R}_{\geq 0}$ with $\frac{\operatorname{Ent}_{\pi \otimes n}(f)}{\mathbb{E}_{\pi \otimes n}[f]} = cn$ we have

$$\frac{\mathcal{E}\left(f^{\frac{1}{p}}, f^{1-\frac{1}{p}}\right)}{\mathbb{E}_{\pi^{\otimes n}}[f]} \ge nb_p(c) = n\widecheck{b}_p(c), \tag{146}$$

where the first step is by Theorem 1 and the second step is by $\check{b}_p = b_p$ ([43, Theorem 4, 6]). (For p = 1 the Dirichlet form should be replaced by $\mathcal{E}(f, \log f)$.) For f of the form $f(x) = \prod_{i \in [n]} g(x_i)$ with $\frac{\text{Ent}_{\pi}(g)}{\mathbb{E}_{\pi}[g]} = c$, equality is achieved.

Proof of (ii). Choose $c \in [0, \log q]$ such that $b_p(c) > \check{b}_p(c)$. Such c exists by Prop. 26. For any function f of the form $f(x) = \prod_{i \in [n]} g(x_i)$, we have

$$\frac{\operatorname{Ent}_{\pi^{\otimes n}}(f)}{\mathbb{E}_{\pi^{\otimes n}}[f]} = n \cdot \frac{\operatorname{Ent}_{\pi}(g)}{\mathbb{E}_{\pi}[g]}, \qquad \frac{\mathcal{E}\left(f^{\frac{1}{p}}, f^{1-\frac{1}{p}}\right)}{\mathbb{E}_{\pi^{\otimes n}}[f]} = n \cdot \frac{\mathcal{E}\left(g^{\frac{1}{p}}, g^{1-\frac{1}{p}}\right)}{\mathbb{E}_{\pi}[g]}.$$
 (147)

(For p = 1 the Dirichlet forms should be replaced by $\mathcal{E}(f, \log f)$ and $\mathcal{E}(g, \log g)$ respectively.) By Theorem 1, if $\frac{\text{Ent}_{\pi}(g)}{\mathbb{E}_{\pi}[g]} = c$, then $\frac{\mathcal{E}\left(g^{\frac{1}{p}}, g^{1-\frac{1}{p}}\right)}{\mathbb{E}_{\pi}[g]} \ge b_{p}(c)$. Therefore for all such product functions f, we have $\frac{\mathcal{E}\left(f^{\frac{1}{p}}, f^{1-\frac{1}{p}}\right)}{\mathbb{E}_{\pi}\otimes n[f]} \ge nb_{p}(c)$. On the other hand, by slightly varying the proof of Prop. 22 for m leaves m

slightly varying the proof of Prop. 23, for *n* large enough, there exists a distribution *f* such that $\frac{\operatorname{Ent}_{\pi\otimes n}(f)}{\mathbb{E}_{\pi\otimes n}[f]} = cn$ and $\frac{\mathcal{E}\left(f^{\frac{1}{p}}, f^{1-\frac{1}{p}}\right)}{\mathbb{E}_{\pi\otimes n}[f]} = \left(\check{b}_{p}(c) + o(1)\right)n.$

Proof of (iii). For any distribution ν on $[q]^n$ with $D(\nu \| \pi^{\otimes n}) = cn$, we have

$$D(\nu \operatorname{PC}_{\lambda}^{\otimes n} \| \pi^{\otimes n}) \le n \widehat{s}_{\lambda}(c) = n s_{\lambda}(c), \tag{148}$$

where the first step is by Theorem 4 and the second step is by $\hat{s}_{\lambda} = s_{\lambda}$. For $\nu = \mu^{\otimes n}$ with $D(\mu \| \pi) = c$, equality is achieved.

Proof of (iv). Choose $c \in [0, \log q]$ such that $s_{\lambda}(c) < \hat{s}_{\lambda}(c)$. Such c exists by Prop. 26. For any distribution ν of the form $\nu = \mu^{\otimes n}$, we have

$$D(\nu \| \pi^{\otimes n}) = n D(\mu \| \pi), \qquad D(\nu \operatorname{PC}_{\lambda}^{\otimes n} \| \pi^{\otimes n}) = n D(\mu \operatorname{PC}_{\lambda} \| \pi)$$
(149)

By Theorem 4, if $D(\mu \| \pi) = c$, then $D(\mu \operatorname{PC}_{\lambda} \| \pi) \leq s_{\lambda}(c)$. Therefore for all such product distributions μ , we have $D(\nu \operatorname{PC}_{\lambda}^{\otimes n} \| \pi^{\otimes n}) \leq ns_{\lambda}(c)$. On the other hand, by slightly varying the proof of Corollary 25, for *n* large enough, there exists a distribution ν on $[q]^n$ such that $D(\nu \| \pi^{\otimes n}) = cn$ and $D(\nu \operatorname{PC}_{\lambda}^{\otimes n} \| \pi^{\otimes n}) = (\widehat{s}_{\lambda}(c) + o(1))n$. \Box

We remark that Prop. 27(i)(iii) hold also for the Ornstein–Uhlenbeck semigroup with extremal distributions of form $\mathcal{N}(0, \sigma^2 I_n)$.

Let us discuss some general implications of non-convexity of b_p and non-concavity of s_{λ} near 0.

Let K be a Markov kernel with stationary distribution π . Consider the tightest possible p-NLSI given by

$$b_p(x) := \inf_{\substack{f: \mathcal{X} \to \mathbb{R}_{\ge 0,} \\ \mathbb{E}_{\pi}[f] = 1, \operatorname{Ent}_{\pi}(f) = x}} \mathcal{E}\left(f^{\frac{1}{p}}, f^{1-\frac{1}{p}}\right).$$
(150)

The *p*-log-Sobolev constant is

$$\alpha_p := \inf_{x>0} \frac{b_p(x)}{x} = \inf_{f: \mathcal{X} \to \mathbb{R}_{\ge 0}, \operatorname{Ent}_{\pi}(f) > 0} \frac{\mathcal{E}\left(f^{\frac{1}{p}}, f^{1-\frac{1}{p}}\right)}{\operatorname{Ent}_{\pi}(f)}.$$
 (151)

We also define the spectral gap

$$\lambda := \inf_{f: \mathcal{X} \to \mathbb{R}_{\ge 0}, \operatorname{Var}(f) > 0} \frac{\mathcal{E}(f, f)}{\operatorname{Var}(f)},$$
(152)

where $\operatorname{Var}(f) = \mathbb{E}_{\pi}(f - \mathbb{E}_{\pi}[f])^2$. For any p > 1, we have

$$\frac{p^2}{2(p-1)}\alpha_p \le \lambda. \tag{153}$$

The case p = 2 is proved in Diaconis and Saloff-Coste [14], and the general case is proved in Mossel et al. [38]. Their proof in fact implies a stronger inequality.

Lemma 28.

$$\limsup_{x \to 0^+} \frac{b_p(x)}{x} \le \frac{2(p-1)}{p^2}\lambda.$$
(154)

In particular, when b_p is strictly concave near 0, we have

$$\alpha_p < \limsup_{x \to 0^+} \frac{b_p(x)}{x} \le \frac{2(p-1)}{p^2}\lambda,\tag{155}$$

and (153) is strict.

Proof. Take any $g : \mathcal{X} \to \mathbb{R}_{\geq 0}$ with $\operatorname{Var}(g) > 0$. Define $f_{\epsilon} = 1 + \epsilon g$. As $\epsilon \to 0$, we have

$$\mathcal{E}\left(f_{\epsilon}^{\frac{1}{p}}, f_{\epsilon}^{1-\frac{1}{p}}\right) = \epsilon^{2} \frac{1}{p} \left(1 - \frac{1}{p}\right) \mathcal{E}(g, g) + o(\epsilon^{2}), \tag{156}$$

$$\operatorname{Ent}_{\pi}(f_{\epsilon}) = \frac{1}{2}\epsilon^{2}\operatorname{Var}(g) + o(\epsilon^{2}).$$
(157)

Because $\operatorname{Ent}_{\pi}(f_{\epsilon}) \to 0$ continuously as $\epsilon \to 0$, we have

$$\limsup_{x \to 0^+} \frac{b_p(x)}{x} \le \lim_{\epsilon \to 0} \frac{\mathcal{E}\left(f_{\epsilon}^{\frac{1}{p}}, f_{\epsilon}^{1-\frac{1}{p}}\right)}{\operatorname{Ent}_{\pi}(f_{\epsilon})} = \frac{2(p-1)}{p^2} \frac{\mathcal{E}(g,g)}{\operatorname{Var}(g)}.$$
 (158)

Lemma then follows because g is arbitrary.

Roughly speaking, existence of a "linear piece" near 0 in \check{b}_p implies that (153) is strict. For the Potts semigroup with $q \ge 3$, b_p is strictly concave near 0 by proof of Prop. 26. So (153) is strict for the Potts semigroup.

The story for non-linear SDPI is very similar. Let W be a channel and v be an input distribution. Consider the tightest possible non-linear SDPI given by

$$s(x) := \sup_{\mu: D(\mu \| v) = x} D(\mu W \| v W).$$
(159)

The input-restricted KL divergence contraction coefficient is

$$\eta_{\mathrm{KL}}(\nu, W) := \sup_{x>0} \frac{s(x)}{x} = \sup_{\mu: 0 < D(\mu \| \nu) < \infty} \frac{D(\mu W \| \nu W)}{D(\mu \| \nu)}.$$
 (160)

We also consider the input-restricted χ^2 -divergence contraction coefficient

$$\eta_{\chi^{2}}(\nu, W) := \sup_{\mu: 0 < \chi^{2}(\mu \| \nu) < \infty} \frac{\chi^{2}(\mu W \| \nu W)}{\chi^{2}(\mu \| \nu)}.$$
(161)

It is known (Ahlswede and Gács [2]) that

$$\eta_{\mathrm{KL}}(\nu, W) \ge \eta_{\chi^2}(\nu, W). \tag{162}$$

Similarly to the *p*-NLSI case, the proof of (162) implies a stronger inequality. **Lemma 29.**

$$\liminf_{x \to 0^+} \frac{s(x)}{x} \ge \eta_{\chi^2}(\nu, W).$$
(163)

In particular, when s is strictly convex near 0, we have

$$\eta_{KL}(\nu, W) > \liminf_{x \to 0^+} \frac{s(x)}{x} \ge \eta_{\chi^2}(\nu, W).$$
(164)

and (162) is strict.

Proof. Fix any distribution μ with $0 < \chi^2(\mu \| \nu) < \infty$. Proof of [44, Theorem 2] constructs a sequence of distributions μ_{ϵ} satisfying

$$D(\mu_{\epsilon} \| \nu) = \epsilon^2 \chi^2(\mu \| \nu) + o(\epsilon^2), \qquad (165)$$

$$D(\mu_{\epsilon} W \| \nu W) = \epsilon^2 \chi^2(\mu W \| \nu W) + o(\epsilon^2), \qquad (166)$$

and $D(\mu_{\epsilon} \| \nu) \to 0$ continuously as $\epsilon \to 0$. Therefore

$$\liminf_{x \to 0^+} \frac{s(x)}{x} \ge \lim_{\epsilon \to 0} \frac{D(\mu_{\epsilon} W \| \nu W)}{D(\mu_{\epsilon} \| \nu)} = \frac{\chi^2(\mu W \| \nu W)}{\chi^2(\mu \| \nu)}.$$
(167)

Because μ is arbitrary, we finish the proof.

Roughly speaking, existence of a "linear piece" near 0 in \hat{s} implies that (162) is strict. For Potts channels PC_{λ} with $\lambda \in \left[-\frac{1}{q-1}, 0\right] \cup (0, 1)$ and $q \ge 3$, s_{λ} is strictly convex near 0 by proof of Prop. 26. So (162) is strict for Potts channels.

П

3.3. Edge-isoperimetric inequalities As a toy application of the NLSIs for the product spaces, we derive an edge-isoperimetric inequality for K_q^n , the graph whose vertex set is $[q]^n$, and edges connect vertex pairs with Hamming distance one. Given a graph G = (V, E), edge-isoperimetric inequalities solve the following combinatorial optimization problem:

$$\Psi_G(N) = \min\{|E(S, S^c)| : |S| = N\},\tag{168}$$

where $|E(S, S^c)| = \#\{e \in E : |e \cap S| = 1\}$. For K_q^n , the edge-isoperimetric problem has been completely solved [4,25,26,29]. Specifically, [29] showed that the optimal *S* minimizing $|E(S, S^c)|$ for a fixed |S| consists of largest elements in $[q]^n$ under a lexicographical order. In particular, we have

$$\Psi_{K_a^n}(q^m) = (n-m)(q-1)q^m.$$
(169)

This was obtained by an explicit combinatorial argument (via a form of shifting/compression). What estimates can be obtained via LSIs and NLSIs?

Let $f = \mathbb{1}_S$ be the indicator function of a set S. Then for any p > 1 we have

$$\frac{\mathcal{E}\left(f^{\frac{1}{p}}, f^{1-\frac{1}{p}}\right)}{\mathbb{E}_{\pi}[f]} = \frac{1}{q-1} \cdot \frac{|E(S, S^{c})|}{|S|} \quad \text{and} \quad \frac{\operatorname{Ent}_{\pi}(f)}{\mathbb{E}_{\pi}[f]} = \log \frac{q^{n}}{|S|}.$$
 (170)

If we relate these two ratios via the 2-LSI (note that from (10), of all p > 1 the p = 2 gives the best result here) and by using the known value of α_2 from (13) we get

$$\Psi_{K_q^n}(q^m) \ge q^m (n-m)(q-2) \frac{\log q}{\log(q-1)}.$$
(171)

Clearly the coefficient in front of $(n - m)q^m$ here is not tight.

The *p*-NLSI allows us to perform a better comparison. First, again via (10) we get the best inequality for p = 2, which results in

$$\Psi_{K_q^n}(q^m) \ge (q-1)q^m n \check{b}_2\left(\frac{n-m}{n}\log q\right).$$
(172)

We know that the function \check{b}_2 is continuous with $\check{b}_2(\log q) = b_2(\log q) = 1$ (from (19)). Thus, for any m = o(n) and $n \to \infty$ we get that (172) implies

$$\Psi_{K_a^n}(q^m) \ge (q-1)q^m(n-m)(1+o(1)), \tag{173}$$

which is tight in this regime. (However, from (19) we can also find that $\tilde{b}'_2(1) = \infty$ and thus, even when m = o(n) the right-hand side of the above inequality is $(q - 1)q^m(n - \omega(m))$, implying the behavior in terms of *m* is not optimal.)

4. Non-reconstruction for Broadcasting on Trees

In this section we prove non-reconstruction results for a general class of broadcasting models on trees, using input-restricted KL divergence SDPI.

Let us formally define the model. Fix an integer $q \ge 2$, a channel $M : [q] \to [q]$, and a stationary distribution $v \in \mathcal{P}([q])$ of M with full support. Let T be a possibly infinite tree with a marked root ρ . The model BOT(T, q, v, M) generates a label $\sigma_v \in [q]$ for every vertex $v \in T$ according to the following process.

- 1. Generate $\sigma_{\rho} \sim \nu$.
- 2. Suppose we have generated a label for a vertex *u*. For every child *v* of *u*, we generate σ_v independently according to

$$\mathbb{P}(\sigma_v = j | \sigma_u = i) = M(i, j).$$
(174)

We often consider the case where *T* is a Galton–Watson tree, meaning that every vertex independently has $t \sim D$ children, where *D* is a distribution on $\mathbb{Z}_{\geq 0}$. We denote the resulting model as BOT(q, v, M, D). An important case is D = Pois(d), the Poisson distribution with mean *d*. When *D* is a singleton at $d \in \mathbb{Z}_{\geq 0}$ we also denote the model as BOT(q, v, M, d).

Let L_k be the set of vertices at distance k to ρ , and σ_{L_k} be the labels of L_k . We say reconstruction is possible if

$$\lim_{k \to \infty} I(\sigma_{\rho}; \sigma_{L_k}, T) > 0, \tag{175}$$

and reconstruction is impossible if the limit is equal to zero.

We recall the branching number br(T) of a tree T defined by Lyons [31].

Definition 30 (*Branching number*). Let *T* be a possibly infinite tree rooted at ρ . Define a flow to be a function $f : V(T) \to \mathbb{R}_{>0}$ such that for every vertex *u*, we have

$$f_u = \sum_{v \in c(u)} f_v, \tag{176}$$

where c(u) denotes the set of children of u. Define br(T) to be the sup of all numbers λ such that there exists a flow f with $f_{\rho} > 0$, and $f_u \leq \lambda^{-d(u,\rho)}$ for all vertices u, where $d(u, \rho)$ is the distance between u and ρ .

Recall Theorem 5, which states that for the model BOT(T, q, ν, M), reconstruction is impossible if $\eta_{\text{KL}}(\nu, M^*)$ br(T) < 1. Now we prove the theorem.

Proof of Theorem 5. For any vertex u, let $L_{u,k}$ denote the set of descendants of u at distance k to ρ . Define

$$a_{u} = H(v)^{-1} \eta_{\mathrm{KL}}(v, M^{*})^{d(u,\rho)} \lim_{k \to \infty} I(\sigma_{u}; \sigma_{L_{u,k}}).$$
(177)

By DPI, $I(\sigma_u; \sigma_{L_{u,k}})$ is non-increasing for $k \ge d(u, \rho)$, so the limit exists.

For any $v \in c(u)$, consider the Markov chain

$$\sigma_{L_{v,k}} \to \sigma_v \xrightarrow{M^*} \sigma_u. \tag{178}$$

Because v is an invariant distribution, the distributions of σ_v and σ_u are both v. By SDPI, we have

$$I(\sigma_{u}; \sigma_{L_{v,k}}) \le \eta_{\mathrm{KL}}(v, M^{*}) I(\sigma_{v}; \sigma_{L_{v,k}}).$$
(179)

Because $(\sigma_{L_{v,k}})_{v \in c(u)}$ are independent conditioned on σ_u , we have

$$I(\sigma_u; \sigma_{L_{u,k}}) \le \sum_{v \in c(u)} I(\sigma_u; \sigma_{L_{v,k}}).$$
(180)

Combine the two inequalities and let $k \to \infty$. We get that

$$a_u \le \sum_{v \in c(u)} a_v. \tag{181}$$

Clearly,

$$a_u \le \eta_{\mathrm{KL}}(\nu, M^*)^{d(u,\rho)} \tag{182}$$

for all vertices *u*. However, *a* is not quite a flow yet. We define a flow *b* from *a*. For a vertex *u*, let $u_0 = \rho, \ldots, u_\ell = u$ be the shortest path from ρ to *u*. Define

$$b_{u} = a_{u} \prod_{0 \le j \le \ell - 1} \frac{a_{u_{j}}}{\sum_{v \in c(u_{j})} a_{v}}.$$
(183)

(If $\sum_{v \in c(u_j)} a_v = 0$ for some *j*, then let $b_u = 0$.) It is not hard to check that

$$b_u = \sum_{v \in c(u)} b_v, \tag{184}$$

and that

$$b_u \le a_u \le \eta_{\text{KL}}(\nu, M^*)^{d(u,\rho)}.$$
 (185)

By definition of branching number, we must have $b_{\rho} = 0$. This means

$$\lim_{k \to \infty} I(\sigma_{\rho}; \sigma_{L_k}) = 0, \tag{186}$$

and non-reconstruction holds.

Remark 31. In the definition of the reconstruction problem, it is not necessary to require σ_{ρ} to have distribution ν . Let $\sigma_{L_k}^i$ denote the leaf labels conditioned on $\sigma_{\rho} = i$. Then Theorem 5 implies that when $\eta_{\text{KL}}(\nu, M^*)$ br(T) < 1, we have

$$\lim_{k \to \infty} \operatorname{TV}(\sigma_{L_k}^i, \sigma_{L_k}^j) = 0,$$
(187)

for $i \neq j \in [q]$.

Theorem 5 directly implies non-reconstruction results for Galton–Watson trees.

Corollary 32. Consider the model BOT(q, v, M, D) with $d = \mathbb{E}_{b \sim D} b$. If

$$\eta_{KL}(\nu, M^*)d < 1, \tag{188}$$

then reconstruction is impossible.

Proof. Let *T* be a Galton–Watson tree with offspring distribution *D*. If *T* extincts, then non-reconstruction obviously hold. Conditioned on non-extinction, we have br(T) = d almost surely by [31], thus Theorem 5 applies.

Polyanskiy and Wu [44,45] proved non-reconstruction results on arbitrary directed acyclic graphs (in particular trees) by reducing to percolation problems on the same graph. In the case of the BOT model, their result says that reconstruction is impossible if

$$\eta_{\mathrm{KL}}(M)\operatorname{br}(T) < 1. \tag{189}$$

For any channel *M*, we have

$$\eta_{\mathrm{KL}}(\nu, M) \le \eta_{\mathrm{KL}}(M),\tag{190}$$

and the inequality is often strict. So for reversible channels (i.e., $M = M^*$), Theorem 5 implies result (189). We do not know, however, how to extend Theorem 5 to general DAGs using input-restricted contraction coefficients.

Külske and Formentin [28] proved a non-reconstruction result very similar to Theorem 5. They considered the symmetrized KL divergence

$$D_{\text{SKL}}(P \| Q) = D(P \| Q) + D(Q \| P), \tag{191}$$

which is the *f*-divergence with $f(x) = (x-1) \log x$. They proved that non-reconstruction holds for a Galton–Watson tree with expected offspring *d* if

$$\eta_{\text{SKL}}(\nu, M^*)d < 1. \tag{192}$$

By slightly modifying the proof of Theorem 5, their result can be strengthened to that reconstruction is impossible for BOT(T, q, π, M) if

$$\eta_{\text{SKL}}(\nu, M^*) \operatorname{br}(T) < 1.$$
 (193)

Proceeding to input-restricted contraction coefficients, we computed both numerically for several binary asymmetric channels and Potts channels.

In Fig. 2, we compare $\eta_{\text{SKL}}(\nu, M)$ and $\eta_{\text{KL}}(\nu, M)$ for the binary asymmetric channel

$$M = \begin{pmatrix} 1-a & a \\ b & 1-b \end{pmatrix}$$
(194)

for a = 0.3 and $b \in [0, 1]$. Simple computation shows that $v = \left(\frac{b}{a+b}, \frac{a}{a+b}\right)$ and $M = M^*$. In Fig. 3, we compare the input-restricted SKL and KL contraction coefficients for

Potts channels PC_{λ} for q = 5 and $\lambda \in \left[-\frac{1}{q-1}, 1\right]$. Because a simplified expression for $\eta_{SKL}(\pi, PC_{\lambda})$ is not known, we use a lower bound $\overline{\eta}_{SKL}(\pi, PC_{\lambda})$, which is defined as the sup of $\frac{D_{SKL}(\mu PC_{\lambda} || \pi)}{D_{SKL}(\mu || \pi)}$ over distributions $\mu = (p_1, \ldots, p_q) \in \mathcal{P}([q])$ with $p_2 = \cdots = p_q$. Clearly $\overline{\eta}_{SKL}(\pi, PC_{\lambda}) \leq \eta_{SKL}(\pi, PC_{\lambda})$. [20] conjectured that $\overline{\eta}_{SKL}(\pi, PC_{\lambda}) = \eta_{SKL}(\pi, PC_{\lambda})$ always holds.

As shown in Figs. 2 and 3, in both cases, we observe

$$\eta_{\mathrm{KL}}(\nu, M^*) \le \eta_{\mathrm{SKL}}(\nu, M^*),\tag{195}$$

which means Theorem 5 yields a stronger non-reconstruction result for these cases.



Fig. 2. Contraction coefficient comparison for binary asymmetric channels with a = 0.3 and varying $b \in$ [0, 1]. The figure shows $\eta_{\text{SKL}}(\nu, M) - \eta_{\text{KL}}(\nu, M)$ is non-negative



Fig. 3. Contraction coefficient comparison for Potts channel with q = 5 and varying $\lambda \in \left[-\frac{1}{q-1}, 1\right]$. The figure shows $\overline{\eta}_{SKL}(\pi, PC_{\lambda}) - \eta_{KL}(\pi, PC_{\lambda})$ is non-negative

We remark that the input-unrestricted KL and SKL contraction coefficients agree. Indeed, the function $x \mapsto (x - 1) \log x$ is operator convex (e.g., [10, Example 3.6]), and thus by [11, Theorem 1], we have

$$\eta_{\rm KL}(M) = \eta_{\rm SKL}(M)$$

for any channel M.

Suppose that for some function g, the g-mutual information satisfies the following subadditivity property: for any Markov chain Y - X - Z, we have

$$I_g(X; Y, Z) \le I_g(X; Y) + I_g(X; Z).$$
 (196)

Then non-reconstruction holds for a tree T with

$$\eta_g(\nu, M^*) \operatorname{br}(T) < 1,$$
 (197)

by modifying the proof of Theorem 5. For mutual information subadditivity is standard. For SKL mutual information, Formentin and Külske [20] proved that

$$I_{\text{SKL}}(X; Y, Z) = I_{\text{SKL}}(X; Y) + I_{\text{SKL}}(X; Z).$$
(198)

An interesting question is, given a pair (v, M), what is the smallest $\eta_g(v, M^*)$ over all subadditive *g*-mutual informations. Solving this question would give the best possible non-reconstruction result that can be achieved by our method.

In Appendix E, we study a broadcasting on trees model with Gaussian kernel considered in Eldan et al. [16], and prove tight non-reconstruction results for this model, closing a gap left in op. cit.

5. Potts Model on a Tree

In this section, we apply Theorem 5 to get non-reconstruction results for Potts models on a tree. In the Potts model, labels propagate through the Potts channel PC_{λ} with invariant distribution $\pi = \text{Unif}([q])$. Because the Potts channels are reversible, Theorem 5 implies non-reconstruction for

$$\eta_{\mathrm{KL}}(\pi, \mathrm{PC}_{\lambda}) \operatorname{br}(T) < 1.$$
(199)

Thus Theorem 6 directly follows from Theorem 5.

Let us briefly discuss previous non-reconstruction results for the Potts channel. Mossel and Peres [39] proved non-reconstruction for

$$\frac{q\lambda^2}{(q-2)\lambda+2}\operatorname{br}(T) < 1.$$
(200)

By Prop. 40 we see that (200) exactly corresponds to using the input-unrestricted KL contraction coefficient $\eta_{\text{KL}}(\text{PC}_{\lambda})$. Therefore, Theorem 6 is strictly stronger than [39]. Martinelli et al. [33] proved non-reconstruction for regular trees for

$$d(1-\epsilon)\frac{q\lambda^2}{(q-2)\lambda+2} < 1$$
(201)

for some $\epsilon = \epsilon(q, d, \lambda) > 0$.

Sly [50] obtained very sharp results for regular trees, including that Kesten–Stigum (KS) threshold is tight for q = 3 and large enough d, and an expression for the reconstruction threshold for larger q and $d \rightarrow \infty$. Mossel et al. [41] improved over [50] and proved that the KS threshold is tight for q = 3, 4 and large enough tree, for Galton–Watson random trees with mild assumptions on the offspring distribution. It is unclear what results can be achieved for small d using their method. It seems that Theorem 6 is not able to give tightness of the KS threshold in these cases.

Formentin and Külske [20] gave non-reconstruction results very similar to ours by using the input-restricted SKL contraction coefficients. As discussed in Sect. 4, numerical computation suggests that Theorem 6 gives better results than theirs in the case of Potts models.

For certain parameters, we can compute the contraction coefficient $\eta_{\text{KL}}(\pi, \text{PC}_{\lambda})$ in closed form. In the following we show two examples.

5.1. Binary symmetric channel The Potts model with q = 2 is also known as the Ising model. In this case, PC_{λ} is the binary symmetric channel BSC_{δ} with $\delta = \frac{1-\lambda}{2}$, which is known (Ahlswede and Gács [2]) to have contraction coefficient $\eta_{KL}(BSC_{\delta}) = \eta_{KL}(\pi, BSC_{\delta}) = (1-2\delta)^2 = \lambda^2$. Therefore Theorem 6 implies non-reconstruction for $(1-2\delta)^2 \operatorname{br}(T) < 1$, which was shown in Bleher et al. [7] (for regular trees) and Evans et al. [18] (for general trees). So for the Ising model on trees our method can give the tight reconstruction threshold.

5.2. Random coloring The random coloring model is a special case of the Potts model with broadcasting channel $\operatorname{Col}_q := \operatorname{PC}_{-\frac{1}{q-1}}$. This channel acts on input $x \in [q]$ by outputting $y \neq x$ uniformly among all q - 1 alternatives.

We compute the exact input-restrict KL contraction coefficient of the coloring channel Col_q .

Proposition 33.

$$\eta_{KL}(\pi, Col_q) = \frac{\log q - \log(q-1)}{\log q}.$$
(202)

Proof. By (30) we have

$$\eta_{\mathrm{KL}}(\pi, \mathrm{Col}_q) = \sup_{x \in \left(\frac{1}{q}, 1\right]} \frac{\log q + \frac{1-x}{q-1} \log \frac{1-x}{q-1} + \frac{q+x-2}{q-1} \log \frac{q+x-2}{(q-1)^2}}{\log q + x \log x + (1-x) \log \frac{1-x}{q-1}}$$
$$= \sup_{x \in \left(\frac{1}{q}, 1\right]} \frac{\log q - \log(q-1) + \frac{1-x}{q-1} \log(1-x) + \frac{q+x-2}{q-1} \log \frac{q+x-2}{q-1}}{\log q + x \log x + (1-x) \log \frac{1-x}{q-1}}.$$
(203)

Taking x = 1, we get

$$\eta_{\mathrm{KL}}(\pi, \mathrm{Col}_q) \ge \frac{\log q - \log(q-1)}{\log q}.$$
(204)

To prove the proposition, we only need to prove that for $x \in \left(\frac{1}{q}, 1\right]$,

$$\frac{\frac{1-x}{q-1}\log(1-x) + \frac{q+x-2}{q-1}\log\frac{q+x-2}{q-1}}{x\log x + (1-x)\log\frac{1-x}{q-1}} \ge \frac{\log q - \log(q-1)}{\log q}.$$
 (205)

(Note that both numerator and denominator in LHS are non-positive.) Define

$$g(x) = (\log q - \log(q - 1))x \log x - \frac{\log q}{q - 1}(1 - x)\log(1 - x),$$
(206)

$$h(x) = g(x) + (q-1)g\left(\frac{1-x}{q-1}\right).$$
(207)

Rearranging (205), we only need to prove that $h(x) \ge 0$ for $x \in \left(\frac{1}{q}, 1\right]$.

We compute that

$$g'(x) = (\log q - \log(q - 1))(1 + \log x) + \frac{\log q}{q - 1}(1 + \log(1 - x)),$$
(208)

$$g''(x) = (\log q - \log(q-1))\frac{1}{x} - \frac{\log q}{q-1}\frac{1}{1-x},$$
(209)

$$g'''(x) = -(\log q - \log(q-1))\frac{1}{x^2} - \frac{\log q}{q-1}\frac{1}{(1-x)^2} < 0.$$
 (210)

Claim 34. h'''(x) < 0 on (0, 1).

Proof.

$$\begin{split} h'''(x) &= g'''(x) - \frac{1}{(q-1)^2} g'''\left(\frac{1-x}{q-1}\right) \\ &= -(\log q - \log(q-1))\frac{1}{x^2} - \frac{\log q}{q-1}\frac{1}{(1-x)^2} \\ &+ \frac{1}{(q-1)^2} \left((\log q - \log(q-1))\frac{1}{\left(\frac{1-x}{q-1}\right)^2} + \frac{\log q}{q-1}\frac{1}{(1-\frac{1-x}{q-1})^2} \right) \\ &= (\log \frac{q}{q-1}) \left(\frac{1}{(1-x)^2} - \frac{1}{x^2}\right) + \frac{\log q}{q-1} \left(\frac{1}{(q-2+x)^2} - \frac{1}{(1-x)^2}\right) \\ &= \frac{1}{(1-x)^2} \left(\left(\log \frac{q}{q-1}\right) \left(1 - \frac{(1-x)^2}{x^2}\right) + \frac{\log q}{q-1} \left(\frac{(1-x)^2}{(q-2+x)^2} - 1\right) \right) \\ &=: \frac{1}{(1-x)^2} (s(x) + t(x)). \end{split}$$

We have

1. s(x) < 0 for $x < \frac{1}{2}$, s(x) > 0 for $x > \frac{1}{2}$; 2. t(x) < 0 for $x \in (0, 1)$; 3. s(x) is increasing for $x \in (0, 1)$; 4. t(x) is decreasing for $x \in (0, 1)$. So h'''(x) < 0 for $x \le \frac{1}{2}$. For $x \ge \frac{1}{2}$, we have

$$s(x) + t(x) < s(1) + t\left(\frac{1}{2}\right) = \log\frac{q}{q-1} + \frac{\log q}{q-1}\left(\frac{1}{(2q-3)^2} - 1\right).$$
 (211)

It is not hard to verify that the last term is < 0 for $q \ge 3$.

By Claim 34, h'(x) is strictly concave. Because $h'\left(\frac{1}{q}\right) = 0$, $h\left(\frac{1}{q}\right) = h(1) = 0$, we get that h(x) > 0 for $x \in (1/q, 1)$. This finishes the proof.

Theorem 5 and Prop. 33 together imply non-reconstruction for

$$br(T) < \frac{\log q}{\log q - \log(q - 1)} = (1 - o(1))q \log q.$$
(212)

This result was previously established by Sly [49] (regular trees and Galton–Watson trees with Poisson offspring distribution), Bhatnagar et al. [6] (regular trees), and Efthymiou [15] (Galton–Watson trees with mild assumptions on the offspring distribution). Our result does not assume any conditions on the offspring distribution other than the expected offspring, and in fact works for arbitrary trees.

We note that [49] achieved more accurate lower order terms, proving that reconstruction is impossible (for regular trees or Galton–Watson trees with Poisson offspring distribution) if

$$d \le q(\log q + \log \log q + 1 - \log 2 - o(1)).$$
(213)

Our method recovers the first order term, but cannot recover the lower order terms. We suggest that the reason is that when combining information coming from subtrees, we use subadditivity of mutual information (see (180)), which is very simple but may be not tight. In contrast, [49] wrote down the exact formula for belief propagation recursion and performed a very careful analysis.

Remark 35. We remark that previous methods based on information contraction do not give the threshold $(1-o(1))q \log q$. The information percolation method [19,44] implies non-reconstruction for

$$\eta_{\mathrm{KL}}(\mathrm{Col}_q)\,\mathrm{br}(T) < 1. \tag{214}$$

By Prop. 40, this gives non-reconstruction for d < q - 1 which is far from being tight. The SKL information contraction method [20] gives non-reconstruction for

$$\eta_{\text{SKL}}(\pi, \text{Col}_q) \operatorname{br}(T) < 1.$$
(215)

If we let $v_{\epsilon} := (1 - \epsilon, \frac{\epsilon}{q-1}, \dots, \frac{\epsilon}{q-1})$, then

$$\eta_{\text{SKL}}(\pi, \text{Col}_q) \ge \lim_{\epsilon \to 0} \frac{D_{\text{SKL}}(\nu_\epsilon \operatorname{Col}_q \| \pi)}{D_{\text{SKL}}(\nu_\epsilon \| \pi)} = \frac{1}{q-1}.$$
(216)

Therefore this method cannot give non-reconstruction results better than for d < q - 1.

$$\Box$$

6. Stochastic Block Model

In this section we study the weak recovery problem for the stochastic block model with q symmetric communities. In this model, there are n vertices, each independently and uniformly randomly assigned one of q labels. Say vertex $v \in [n]$ has lavel $X_v \in [q]$. A random graph is generated, where for any two vertices, there is an edge between them with probability $\frac{a}{n}$ if they have the same labels, and with probability $\frac{b}{n}$ otherwise. We call the resulting model SBM(n, q, a, b).

The goal of weak recovery is to recover a non-trivial fraction of the communities given the unlabeled graph. We say weak recovery is possible if there exists an estimator $\hat{X}(G) \in [q]^V$ such that

$$\lim_{n \to \infty} \frac{1}{n} \mathbb{E} d_H(X, \hat{X}(G)) < 1 - \frac{1}{q},$$
(217)

where
$$d_H(X, Y) := \min_{\tau \in \operatorname{Aut}([q])} \sum_{i \in [n]} \mathbb{1}\{X_i \neq \tau(Y_i)\}.$$
 (218)

Due to symmetry in the labels, we can only expect to recover the labels up to a permutation, so in the definition of d_H we take minimum over all permutations of the labels. If \hat{X} outputs i.i.d. uniformly random labels, then the limit of $\frac{1}{n}\mathbb{E}d_H(X, \hat{X})$ is $1 - \frac{1}{q}$. So the notion of weak recovery indicates whether we can recover the communities better than purely random guessing.

In the following, we show that SDPI-based non-reconstruction results lead to improved impossibility of weak recovery results for the SBM.

6.1. Impossibility of weak recovery via information percolation We first give an impossibility result via an information percolation method of Polyanskiy and Wu [45]. For the stochastic block model SBM(n, q, a, b), they constructed a corresponding broad-casting on trees (BOT) model as follows.

Let $d = (\sqrt{a} - \sqrt{b})^2$. Consider a Galton–Watson tree *T* with offspring distribution Pois(*d*). We independently and uniformly randomly choose a label $\in [q]$ for every vertex. Say vertex *v* has label σ_v . We observe $\omega_{u,v} = \mathbb{1}\{\sigma_u = \sigma_v\}$ for each edge (u, v).

Let ρ denote the root of T and L_k denote the set of vertices at distance k to ρ . Let ω denote the set of all observations. We say reconstruction is impossible if

$$\lim_{k \to \infty} I(\sigma_{\rho}; \sigma_{L_k} | T, \omega) = 0.$$
(219)

Proposition 36 ([45, Prop. 8]). Weak recovery for the model SBM(n, q, a, b) is impossible, if reconstruction is impossible for the above BOT model.

[45] proved that reconstruction is impossible for the above tree model when $d < \frac{q}{2}$ using a coupling argument. We make an improvement using SDPI for the coloring channel Col_q (recall that Col_q is the Potts channel $\operatorname{PC}_{-\frac{1}{2}}$).

Proposition 37. Reconstruction is impossible for the above BOT model if

$$d < \left(\frac{\log q - \log(q-1)}{\log q} \frac{q-1}{q} + \frac{1}{q}\right)^{-1} = q - (1 + o(1))q / \log q.$$
(220)

Proof. The tree model is equivalent to the following top-down process:

- 1. Generate σ_{ρ} uniformly randomly over [q].
- 2. Suppose we have generated a label for a vertex u. For every child v of u, we randomly choose the transition matrix M, which is the identity channel Id with probability $\frac{1}{q}$, and the coloring channel Col_q with probability $1 \frac{1}{q}$. Then we generate v according to $\mathbb{P}(\sigma_v = j | \sigma_u = i) = M_{i,j}$.

For any vertex u, let $L_{u,k}$ denote the set of descendants of u at distance k to ρ . Let v be a child of u.

Note that $\pi = \text{Unif}([q])$ is an invariant distribution for both Id and Col_q , and the two channels are both reversible. We have

$$\mathbb{E}\left[I(\sigma_{u};\sigma_{L_{v,k}}|T,\omega)|\omega_{u,v}=1\right] \leq \eta_{\mathrm{KL}}(\pi,\mathrm{Id})I(\sigma_{v};\sigma_{L_{v,k}}|T,\omega) = I(\sigma_{v};\sigma_{L_{v,k}}|T,\omega)$$
(221)

and, by Prop. 33,

$$\mathbb{E}\left[I(\sigma_{u};\sigma_{L_{v,k}}|T,\omega)|\omega_{u,v}=0\right] \leq \eta_{\mathrm{KL}}(\pi,\mathrm{Col}_{q})I(\sigma_{v};\sigma_{L_{v,k}}|T,\omega)$$
$$=\frac{\log q - \log(q-1)}{\log q}I(\sigma_{v};\sigma_{L_{v,k}}|T,\omega). \tag{222}$$

Taking expectation, we get

$$I(\sigma_u; \sigma_{L_{v,k}} | T, \omega) \le \left(\frac{\log q - \log(q-1)}{\log q} \frac{q-1}{q} + \frac{1}{q}\right) I(\sigma_v; \sigma_{L_{v,k}} | T, \omega).$$
(223)

Rest of the proof is the same as Theorem 5.

Prop. 36 and Prop. 37 together show that weak recovery is impossible for SBM(n, q, a, b) when

$$\left(\sqrt{a} - \sqrt{b}\right)^2 < \left(\frac{\log q - \log(q-1)}{\log q} \frac{q-1}{q} + \frac{1}{q}\right)^{-1}.$$
 (224)

As shown in Fig. 4, for certain parameters, (224) leads to slight improvement over [3].

6.2. *Impossibility of weak recovery via Potts model* We have shown that the information percolation method together with the input-restricted KL contraction coefficients gives a simple yet strong impossibility result for weak recovery of the stochastic block model. The information percolation method can be understood as comparison with the erasure channel. However, the stochastic block model is more closely related to the Potts channel. In this section we prove Theorem 7, an even better impossibility result, via the Potts model on a tree.

The following result says that impossibility of reconstruction for the Potts model implies impossibility of weak recovery for the SBM. The result was first proved by Mossel et al. [36] for the case q = 2, and their proof works for general q with slight modification. For a proof for the general case, see Mossel et al. [41] or Gu [23, Theorem 5.15].

Theorem 38 Consider the model SBM(n, q, a, b). Define

$$d = \frac{a + (q-1)b}{q}, \quad \lambda = \frac{a-b}{a + (q-1)b}.$$
(225)

If reconstruction is impossible for the Potts model $BOT(q, \pi, PC_{\lambda}, Pois(d))$, then weak recovery is impossible for the SBM.

We briefly sketch a proof of Theorem 38.

Proof Sketch. The proof is in two parts. In the first part, we show that for some absolute constant c > 0, there exists a coupling between the $c \log n$ -neighborhood of a vertex in the SBM, and the $c \log n$ -neighborhood of the root in the Potts model BOT $(q, \pi, PC_{\lambda}, Pois(d))$, such that the total variation distance between the two neighborhood (containing labels) is o(1). Proof of this part is by observing that the $c \log n$ -neighborhood in SBM has no cycle with high probability, and can be constructed using a sequence of binomial random variables; on the other hand, the Potts model can be constructed using a sequence of Poisson variables. Then we compare the two sequences of random variables and find that they have very small total variation distance.

In the second part, we show that in the SBM, conditioned on labels on the boundary of the *c* log *n*-neighborhood, labels inside and labels outside are approximately independent. In other words, if *A*, *B*, *C* is a partition of *V* such that $|A \cup B| = o(\sqrt{n})$ and *B* separates *A* and *C*, then

$$\mathbb{P}(X_A | X_{B \cup C}, G) = (1 + o(1)) \mathbb{P}(X_A | X_B, G)$$
(226)

for *G* and *X* with probability 1-o(1). The proof is by writing down the partition function and removing exponents on $1 - \frac{a}{n}$ and $1 - \frac{b}{n}$ that have negligible effect. Combining the two parts, one can prove that for any constant *m* and any vertices

Combining the two parts, one can prove that for any constant *m* and any vertices v_0, \ldots, v_m ,

$$I(X_{v_0}; X_{v_1}, \dots, X_{v_m} | G) = o(1).$$
(227)

This implies impossibility of weak recovery.

Theorem 38 together with Theorem 6 implies Theorem 7 immediately. Figure 4 shows a comparison between the impossibility results for q = 5. Note that (34) is equivalent to

$$\frac{\lambda^2(q-1)}{2\log(q-1)} \cdot d < 1.$$
(228)

Comparing Theorem 7 and (228) using (294), we see that Theorem 7 strictly improves over (34) in the assortative regime.

Remark 39. Theorem 7 can be generalized to asymmetric SBMs. Let $n \ge 1$, $q \ge 2$, v be a distribution on [q] with full support, and A be a symmetric $q \times q$ matrix. We define the general stochastic block model SBM(n, q, v, A) as follows. First every vertex $v \in V$ receives a label X_v independently from v. Then a random graph G is constructed, where every edge (u, v) is added with probability $\frac{1}{n}A_{X_u,X_v}$.

The weak recovery problem for the general SBMs asks to output a set S = S(G) such that

$$\frac{\#\{v \in S : X_v = i\}}{\#\{v \in V : X_v = i\}} - \frac{\#\{v \in S : X_v = j\}}{\#\{v \in V : X_v = j\}} \ge \epsilon$$
(229)



Fig. 4. Impossibility of weak recovery results for SBM for q = 5. In the assortative regime, (224) gives better results than [3] for certain parameters, and Theorem 7 gives the best results among the three

for some $i, j \in [q]$ and absolute constant $\epsilon > 0$. Note that for symmetric SBMs, this agrees with the definition given in the beginning of Sect. 6 (see Abbe and Sandon [1]).

Note that the expected degree of a vertex with label *i* is equal to $(\nu A)_i + o(1)$. If νA is not a multiple of $\mathbb{1}$, then weak recovery can be achieved by simply looking at the degrees. Therefore we assume that $\nu A = d\mathbb{1}$ for some $d \ge 1$ and call such models degree indistinguishable.

Gu [23, Theorem 5.15] proved that for a degree indistinguishable SBM, weak recovery is impossible if reconstruction is impossible for BOT(q, v, M, Pois(d)) (defined in Sect. 4), where M is defined as

$$M(i,j) = \frac{1}{d} A_{i,j} \nu_j, \quad \forall i, j \in [q].$$
(230)

It is easy to verify that M is a Markov kernel, $\nu M = \nu$, and (ν, M) is reversible.

Combining [23, Theorem 5.15] and Theorem 5, we get that weak recovery is impossible if $d\eta_{\text{KL}}(\nu, M) < 1$.

Acknowledgement. The authors are grateful to Jingbo Liu for helpful discussions on reconstruction problems on trees, and to the anonymous reviewers for valuable comments.

Funding This work was supported in part by the MIT-IBM Watson AI Lab, by the Center for Science of Information (CSoI), an National Science Foundation Science and Technology Center, under grant agreement CCF-09-39370, and by the National Science Foundation under Grant No CCF-2131115.

Data Availability This article does not have any external supporting data.

Declarations

Conflict of interest The authors do not have any other competing interests to declare.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

A. Input-Unrestricted Contraction Coefficient of Potts Channels

Computation of (input-restricted or input-unrestricted) contraction coefficients is often a daunting task. Previously, Makur and Polyanskiy [32] obtained lower and upper bounds of input-unrestricted KL divergence contraction coefficients for Potts channels. In this section we compute the exact value of these contraction coefficients.

We remark that after our work, Ordentlich and Polyanskiy [42] proved that the inputunrestricted contraction coefficients are achieved by input distributions of support size at most two, giving an alternative (and simpler) proof for Prop. 40. We include our original proof here for completeness.

Proposition 40.

$$\eta_{KL}(PC_{\lambda}) = \frac{q\lambda^2}{(q-2)\lambda+2}.$$
(231)

Proof. The result is obvious for $\lambda \in \{0, 1\}$. In the following, assume that $\lambda \notin \{0, 1\}$.

We use the following characterization of contraction coefficient using Rényi maximal correlation [47] (see e.g. Sarmanov [48]). For any channel M, we have

$$\eta_{\mathrm{KL}}(M) = \left(\sup_{\mu} \sup_{f,g} \mathbb{E}[f(X)g(Y)]\right)^2$$
(232)

where μ is a distribution on [q], $P_X = \mu$, $P_{Y|X} = M$, $f : \mathcal{X} \to \mathbb{R}$ satisfies $\mathbb{E}_X[f] = 0$ and $\mathbb{E}_X[f^2] = 1$, and $g : \mathcal{Y} \to \mathbb{R}$ satisfies $\mathbb{E}_Y[g] = 0$ and $\mathbb{E}_Y[g^2] = 1$. Specialize to $M = \text{PC}_{\lambda}$. Write $\mu = (p_1, \dots, p_q)$, $f = (f_1, \dots, f_q)$ and $g = (g_1, \dots, g_q)$. Then

$$\mathbb{E}[f(X)g(Y)] = \sum_{i,j} f_i p_i g_j \mathbb{P}[Y=j|X=i] = \lambda \sum f_i p_i g_i.$$
(233)

When $\lambda > 0$, we need to maximize $\sum f_i g_i p_i$. When $\lambda < 0$, we make the transform $f_i \leftarrow -f_i$, and still maximize $\sum f_i g_i p_i$. So we get the following optimization problem.

$$\max \sum f_i g_i p_i$$

s.t. $\sum f_i p_i = 0,$ (234)

$$\sum f_i^2 p_i = 1, \tag{235}$$

$$\sum g_i \left(\lambda p_i + \frac{1 - \lambda}{q} \right) = 0, \tag{236}$$

$$\sum g_i^2 \left(\lambda p_i + \frac{1-\lambda}{q}\right) = 1,$$
(237)

$$p_i \ge 0, \sum p_i = 1. \tag{238}$$

Lower bound. Take

$$\mu = \left(\frac{1}{2}, \frac{1}{2}, 0, \dots, 0\right), \quad f = (1, -1, 0, \dots, 0), \quad g = (u, -u, 0, \dots, 0)$$
(239)

where

$$u = \sqrt{\frac{q}{(q-2)\lambda + 2}}.$$
(240)

Then

$$\sum f_i g_i p_i = u. \tag{241}$$

So

$$\eta_{\mathrm{KL}}(\mathrm{PC}_{\lambda}) \ge (\lambda u)^2 = \frac{q\lambda^2}{(q-2)\lambda+2}.$$
(242)

Upper bound. Let us fix μ and maximize over f and g. Assume for the sake of contrary that $\sum f_i g_i p_i > u$. The space of possible g is bounded; some coordinates of f may be unbounded, but their values do not affect the objective function. By compactness, the maximum value of $\sum f_i g_i p_i$ is achieved at some point f and g. Let us compute the derivatives.

$$\nabla_f \sum f_i g_i p_i = (g_i p_i)_{i \in [q]}, \qquad (243)$$

$$\nabla_f \sum f_i p_i = (p_i)_{i \in [q]}, \tag{244}$$

$$\nabla_f \sum f_i^2 p_i = (2f_i p_i)_{i \in [q]}, \tag{245}$$

$$\nabla_g \sum f_i g_i p_i = (f_i p_i)_{i \in [q]}, \qquad (246)$$

$$\nabla_g \sum g_i \left(\lambda p_i + \frac{1 - \lambda}{q} \right) = \left(\lambda p_i + \frac{1 - \lambda}{q} \right)_{i \in [q]}, \tag{247}$$

$$\nabla_g \sum g_i^2 \left(\lambda p_i + \frac{1-\lambda}{q}\right) = \left(2g_i \left(\lambda p_i + \frac{1-\lambda}{q}\right)\right)_{i \in [q]}.$$
 (248)

By maximality in f, there exist constants A and B such that

$$g_i p_i = A p_i + B f_i p_i \tag{249}$$

for all i. By maximality in g, there exist constants C and D such that

$$f_i p_i = C\left(\lambda p_i + \frac{1-\lambda}{q}\right) + Dg_i\left(\lambda p_i + \frac{1-\lambda}{q}\right)$$
(250)

for all *i*. By (249),

$$\sum f_i g_i p_i = \sum f_i (Ap_i + Bf_i p_i) = B.$$
(251)

By (250),

$$\sum f_i g_i p_i = \sum g_i \left(C \left(\lambda p_i + \frac{1 - \lambda}{q} \right) + D g_i \left(\lambda p_i + \frac{1 - \lambda}{q} \right) \right) = D.$$
 (252)

So B = D > u > 0. For *i* with $p_i \neq 0$, we have $g_i = A + Bf_i$ by (249). If for some *i*, $p_i = 0$, then

$$\frac{1-\lambda}{q}(C+Dg_i) = 0.$$
(253)

This means $#{g_i : p_i = 0} = 1$. So we can choose f_i for such *i* such that

$$g_i = A + Bf_i \tag{254}$$

for all i. From (236), we get

$$0 = \sum g_i \left(\lambda p_i + \frac{1 - \lambda}{q} \right)$$

= $\sum (A + Bf_i) \left(\lambda p_i + \frac{1 - \lambda}{q} \right)$
= $A + B \frac{1 - \lambda}{q} \sum f_i.$ (255)

From (237), we get

$$1 = \sum g_i^2 \left(\lambda p_i + \frac{1-\lambda}{q} \right)$$

= $\sum (A^2 + 2ABf_i + B^2 f_i^2) \left(\lambda p_i + \frac{1-\lambda}{q} \right)$
= $A^2 + 2AB\frac{1-\lambda}{q} \sum f_i + B^2\lambda + B^2\frac{1-\lambda}{q} \sum f_i^2$
= $B^2 \left(\lambda + \frac{1-\lambda}{q} \sum f_i^2 - \left(\frac{1-\lambda}{q} \sum f_i \right)^2 \right).$ (256)

The result then follows from Claim 41 because we have

$$B = \frac{1}{\sqrt{\lambda + \frac{1-\lambda}{q} \left(\sum f_i^2 - \frac{1-\lambda}{q} \left(\sum f_i\right)^2\right)}}$$

$$\leq \frac{1}{\sqrt{\lambda + \frac{1-\lambda}{q} \left(\sum f_i^2 - \frac{1}{q-1} \left(\sum f_i\right)^2\right)}}$$

$$\leq \frac{1}{\sqrt{\lambda + \frac{1-\lambda}{q} \cdot 2}} = u.$$
(257)

The second step is because $0 \le \frac{1-\lambda}{q} \le \frac{1}{q-1}$ for all $\lambda \in \left[-\frac{1}{q-1}, 1\right]$.

Claim 41. For any distribution μ and any f satisfying (234) and (235), we have

$$\sum f_i^2 - \frac{1}{q-1} \left(\sum f_i \right)^2 \ge 2.$$
(258)

Proof. Let us first prove the result for f with support size two. WLOG assume that $f_1 > 0, f_2 < 0, f_3 = \cdots = f_q = 0$. One can compute that

$$f_1 = \sqrt{\frac{p_2}{p_1(p_1 + p_2)}}, \quad f_2 = -\sqrt{\frac{p_1}{p_1(p_1 + p_2)}}.$$
 (259)

Then

$$f_1^2 + f_2^2 - \frac{1}{q-1}(f_1 + f_2)^2$$

$$\geq f_1^2 + f_2^2 - (f_1 + f_2)^2$$

$$= \frac{1}{p_1 + p_2} \left(\frac{p_2}{p_1} + \frac{p_1}{p_2} - \left(\sqrt{\frac{p_2}{p_1}} - \sqrt{\frac{p_1}{p_2}} \right)^2 \right)$$

$$= \frac{2}{p_1 + p_2} \geq 2.$$
(260)

Let us define

$$S(\mu) := \left\{ f : \sum f_i p_i = 0, \sum f_i^2 p_i = 1 \right\}$$
(261)

$$U(f) := \sum f_i^2 - \frac{1}{q-1} \left(\sum f_i\right)^2.$$
 (262)

Now suppose that for some μ and $f \in S(\mu)$ we have U(f) < 2. The space $S(\mu)/\{\pm\}$ is continuous as a subsapce of $(\mathbb{R}^q \setminus \{0\})/\{\pm\}$ (with quotient topology), and there exists $f \in S(\mu)$ with $U(f) \ge 2$ (e.g., f with support size two), so for sufficiently small $\epsilon > 0$ there exists $f \in S(\mu)$ such that $U(f) \in (2 - \epsilon, 2)$.

Let $\lambda = -\frac{1}{q-1}$. Take ϵ small enough so that $\lambda + \frac{1-\lambda}{q}(2-\epsilon) > 0$ and choose $f \in S(\mu)$ with $U(f) \in (2-\epsilon, 2)$. Define

$$B = \frac{1}{\sqrt{\lambda + \frac{1-\lambda}{q}U(f)}} > u, \qquad (263)$$

$$A = -B\frac{1-\lambda}{q}\sum f_i,\tag{264}$$

$$g_i = A + B f_i \forall i. \tag{265}$$

One can check that g satisfies (236) and (237), and

$$\sum f_i g_i p_i = B > u. \tag{266}$$

By (232) and (233), this implies

$$\eta_{\text{KL}}\left(\text{PC}_{-\frac{1}{q-1}}\right) > \frac{1}{q-1}.$$
 (267)

However, we have

$$\eta_{\mathrm{KL}}\left(\mathrm{PC}_{-\frac{1}{q-1}}\right) \le \eta_{\mathrm{TV}}\left(\mathrm{PC}_{-\frac{1}{q-1}}\right) = \frac{1}{q-1}.$$
(268)

Contradiction.

B. An Upper Bound for Input-Restricted Contraction Coefficient for Potts Channels

In this section we prove an upper bound for the input-restricted KL divergence contraction coefficient for ferromagnetic Potts channels.

Proposition 42. *Fix* $q \ge 3$ *. For all* $\lambda \in [0, 1]$ *, we have*

$$\eta_{KL}(\pi, PC_{\lambda}) \le \frac{\lambda^2}{(1-\lambda)\frac{2(q-1)\log(q-1)}{q(q-2)} + \lambda}.$$
(269)

For all $\lambda \in [-\frac{1}{q-1}, 0]$, we have

$$\eta_{KL}(\pi, PC_{\lambda}) \le \frac{\lambda^2}{(1 + (q - 1)\lambda)\frac{2(q - 1)\log(q - 1)}{q(q - 2)} - \lambda \frac{\log q}{(q - 1)(\log q - \log(q - 1))}}.$$
 (270)

We first prove a lemma.

Lemma 43. $\frac{(qx-1)^2}{\psi(x)}$ is concave in $x \in [0, 1]$, where $\psi : [0, 1] \to \mathbb{R}$ is defined in (16).

Proof. Let $f(x) = \frac{(qx-1)^2}{\psi(x)}$.

$$f'(x) = \frac{2q(qx-1)}{\psi(x)} - \frac{(qx-1)^2\psi'(x)}{\psi^2(x)}.$$
(271)

$$f''(x) = \frac{2q^2}{\psi(x)} - \frac{4q(qx-1)\psi'(x)}{\psi^2(x)} - \frac{(qx-1)^2\psi''(x)}{\psi^2(x)} + \frac{2(qx-1)^2(\psi')^2(x)}{\psi^3(x)}$$

$$= \frac{2}{\psi^3(x)}(q\psi(x) - (qx-1)\psi'(x))^2 - \frac{(qx-1)^2\psi''(x)}{\psi^2(x)}.$$
(272)

Therefore it suffices to prove that

$$g(x) := \psi^{3}(x) f''(x) = 2(q\psi(x) - (qx - 1)\psi'(x))^{2} - (qx - 1)^{2}\psi(x)\psi''(x)$$
(273)

is non-positive for $x \in [0, 1]$. Note that $g\left(\frac{1}{q}\right) = 0$. So we only need to prove that $g'(x) \ge 0$ for $x \in \left[0, \frac{1}{q}\right]$ and $g'(x) \le 0$ for $x \in \left[\frac{1}{q}, 1\right]$.

$$g'(x) = -4(qx-1)\psi''(x)(q\psi(x) - (qx-1)\psi'(x)) - 2q(qx-1)\psi(x)\psi''(x) - (qx-1)^2\psi'(x)\psi''(x) - (qx-1)^2\psi(x)\psi'''(x) = (qx-1)(-6q\psi(x)\psi''(x) + (qx-1)(3\psi'(x)\psi''(x) - \psi(x)\psi'''(x))).$$
(274)

Therefore we would like to prove that

$$u(q,x) := -6q\psi(x)\psi''(x) + (qx-1)(3\psi'(x)\psi''(x) - \psi(x)\psi'''(x))$$
(275)

is non-positive. We enlarge the domain of u and prove that $u(q, x) \le 0$ for real q > 1 and $x \in (0, 1)$.

We fix $x \in (0, 1)$ and consider $u_x(q) := u(q, x)$. We have $u_x(\frac{1}{x}) = 0$. So it suffices to prove that u_x is concave in q. We have

$$\psi'(x) = \log x - \log \frac{1-x}{q-1},$$
(276)

$$\psi''(x) = \frac{1}{x} + \frac{1}{1-x},\tag{277}$$

$$\psi^{\prime\prime\prime}(x) = \frac{1}{(1-x)^2} - \frac{1}{x^2},$$
(278)

$$\frac{\partial}{\partial q}\psi(x) = \frac{1}{q} - \frac{1-x}{q-1},\tag{279}$$

$$\frac{\partial}{\partial q}\psi'(x) = \frac{1}{q-1},\tag{280}$$

$$\frac{\partial}{\partial q}\psi''(x) = \frac{\partial}{\partial q}\psi'''(x) = 0.$$
(281)

So

$$u'_{x}(q) = -6\psi(x)\psi''(x) - 6q\left(\frac{1}{q} - \frac{1-x}{q-1}\right)\psi''(x) + x(3\psi'(x)\psi''(x) - \psi(x)\psi'''(x))$$

$$+ (qx-1)\left(3\frac{1}{q-1}\psi''(x) - \left(\frac{1}{q} - \frac{1-x}{q-1}\right)\psi'''(x)\right).$$
(282)
$$= -12\left(\frac{1}{q} - \frac{1-x}{q-1}\right)\psi''(x) - 6q\left(-\frac{1}{q} + \frac{1-x}{q-1}\right)\psi''(x)$$

$$u_x''(q) = -12\left(\frac{1}{q} - \frac{1-x}{q-1}\right)\psi''(x) - 6q\left(-\frac{1}{q^2} + \frac{1-x}{(q-1)^2}\right)\psi''(x) + 6x\frac{1}{q-1}\psi''(x) - 2x\left(\frac{1}{q} - \frac{1-x}{q-1}\right)\psi'''(x) + (qx-1)\left(-3\frac{1}{(q-1)^2}\psi''(x) - \left(-\frac{1}{q^2} + \frac{1-x}{(q-1)^2}\right)\psi'''(x)\right) = \frac{(qx-1)^2(1-2q+(q-2)x)}{q^2(q-1)^2x^2(1-x)^2} \le 0.$$
(283)

We are done.

Proof of Prop. 42. For $x \in [0, 1]$ and $\lambda \in \left[-\frac{1}{q-1}, 1\right]$ we define

$$f_x(\lambda) := \frac{\lambda^2 \psi(x)}{\psi\left(\lambda x + \frac{1-\lambda}{q}\right)}.$$
(284)

(Value of $f_x(0)$ is defined using continuity.) Note that

$$\eta_{\mathrm{KL}}(\pi, \mathrm{PC}_{\lambda}) = \sup_{x \in (0,1]} \frac{\lambda^2}{f_x(\lambda)}.$$
(285)

So to compute an upper bound for $\eta_{\text{KL}}(\pi, \text{PC}_{\lambda})$, it suffices to lower bound $f_x(\lambda)$. Because

$$f_x(\lambda) = \frac{\psi(x)}{(qx-1)^2} \cdot \frac{\left(q\left(\lambda x + \frac{1-\lambda}{q}\right) - 1\right)^2}{\psi\left(\lambda x + \frac{1-\lambda}{q}\right)},\tag{286}$$

by Lemma 43, for fixed x, $f_x(\lambda)$ is concave for $\lambda \in \left[-\frac{1}{q-1}, 1\right]$. Therefore by computing lower bounds of $f_x(\lambda)$ for $\lambda = -\frac{1}{q-1}, 0, 1$, we can get lower bounds on $f_x(\lambda)$ for all $\lambda \in \left[-\frac{1}{q-1}, 1\right].$ By Prop. 33, we have

$$f_x\left(-\frac{1}{q-1}\right) \ge \frac{\log q}{(q-1)^2(\log q - \log(q-1))}.$$
 (287)

By L'Hôpital's rule,

$$f_{x}(0) = \psi(x) \lim_{\lambda \to 0} \frac{2\lambda}{\left(x - \frac{1}{q}\right)\psi'\left(\lambda x + \frac{1 - \lambda}{q}\right)}$$
$$= \psi(x) \lim_{\lambda \to 0} \frac{2}{\left(x - \frac{1}{q}\right)^{2}\psi''\left(\lambda x + \frac{1 - \lambda}{q}\right)}$$
$$= \frac{2(q - 1)\psi(x)}{(qx - 1)^{2}}.$$
(288)

By Lemma 43, $g(x) := \frac{(qx-1)^2}{\psi(x)}$ is concave in x. Also

$$g'\left(1-\frac{1}{q}\right) = \frac{2q(q-2)}{\frac{1}{q}(q-2)\log(q-1)} - \frac{(q-2)^2 \cdot 2\log(q-1)}{\left(\frac{1}{q}(q-2)\log(q-1)\right)^2} = 0.$$
 (289)

So

$$g(x) \le g\left(1 - \frac{1}{q}\right) = \frac{q(q-2)}{\log(q-1)}$$
 (290)

and

$$f_x(0) \ge \frac{2(q-1)\log(q-1)}{q(q-2)}.$$
(291)

It is easy to see that

$$f_x(1) \ge 1. \tag{292}$$

Because $f_x(\lambda)$ is concave in λ , (269) follows from (291) and (292), and (270) follows from (287) and (291).

Proof of Prop. 42 implies the first order limit behavior of $\eta_{\text{KL}}(\pi, \text{PC}_{\lambda})$ as $\lambda \to 0$.

$$\lim_{\lambda \to 0} \frac{\eta_{\text{KL}}(\pi, \text{PC}_{\lambda})}{\lambda^2} = \frac{q(q-2)}{2(q-1)\log(q-1)}.$$
(293)

For all $q \ge 3$ and $\lambda \in (0, 1]$, we have

$$\eta_{\text{KL}}(\pi, \text{PC}_{\lambda}) \leq \frac{\lambda^{2}}{(1-\lambda)\frac{2(q-1)\log(q-1)}{q(q-2)} + \lambda} \leq \lambda^{2}(1-\lambda)\frac{q(q-2)}{2(q-1)\log(q-1)} + \lambda^{3} < \lambda^{2}\frac{q(q-2)}{2(q-1)\log(q-1)} < \lambda^{2}\frac{q-1}{2\log(q-1)},$$
(294)

where the second step is by Cauchy inequality.

For comparison with input-unrestricted contraction coefficient

$$\eta_{\rm KL}(\rm PC_{\lambda}) = \frac{q\lambda^2}{(q-2)\lambda+2},\tag{295}$$

we note that $\frac{\lambda^2}{\eta_{KL}(PC_{\lambda})}$ is linear in $\lambda,$ and

$$\frac{1}{q-1} < \frac{\log q}{(q-1)^2 (\log q - \log(q-1))},$$
(296)

$$\frac{2}{q} < \frac{2(q-1)\log(q-1)}{q(q-2)}.$$
(297)

So Prop. 42 implies (31).

C. Non-convexity of Certain Functions

In this section we prove Prop. 26. Let us first prove a lemma.

Lemma 44. Let g be a strictly increasing smooth function from $[x_0, x_1]$ to $[y_0, y_1]$, and f be a smooth function from $[x_0, x_1]$ to \mathbb{R} . Assume that $g'(x_0) = f'(x_0) = 0$ and $(g'' f''' - f'' g''')(x_0) > 0$. Then the function $h = f \circ g^{-1} : [y_0, y_1] \to \mathbb{R}$ is not concave near y_0 .

Proof. Derivatives of *h* are

$$h'(x) = \frac{f'(g^{-1}(x))}{g'(g^{-1}(x))},$$

$$h''(x) = \left(\frac{f''}{g'} - \frac{f'g''}{(g')^2}\right) \left(g^{-1}(x)\right) \frac{1}{g'(g^{-1}(x))}$$

$$= \left(\frac{f''}{(g')^2} - \frac{f'g''}{(g')^3}\right) \left(g^{-1}(x)\right).$$
(298)
(298)
(298)

So it suffices to study the sign of g'f'' - f'g'' for x near x_0 . Let u = g'f'' - f'g''. We have $u(x_0) = 0$. Let us compute the derivatives.

$$u' = g' f''' - f' g''', (300)$$

$$u'' = g'f^{(4)} + g''f''' - f''g''' - g'g^{(4)}.$$
(301)

So $u'(x_0) = 0$ and $u''(x_0) = (g'' f''' - f'' g''')(x_0) > 0$. So *u* is positive near x_0 . *Proof of Prop. 26.* We apply Lemma 44 to $g = \psi$, $x_0 = \frac{1}{q}$, $x_1 = 1$, $y_0 = 0$, $y_1 = \log q$, and various *f*. We have

$$\psi'\left(\frac{1}{q}\right) = 0,\tag{302}$$

$$\psi^{\prime\prime}\left(\frac{1}{q}\right) = \frac{q^2}{q-1},\tag{303}$$

$$\psi^{\prime\prime\prime}\left(\frac{1}{q}\right) = -\frac{q^3(q-2)}{(q-1)^2}.$$
(304)

Part (i), p = 1. For b_1 , take

$$f(x) = -(q-1)\xi_1(x) = \log x + (q-1)\log \frac{1-x}{q-1} + q(\psi(x) - \log q).$$
(305)

Then

$$f'(x) = \frac{1}{x} - \frac{q-1}{1-x} + q\psi'(x), \tag{306}$$

$$f''(x) = -\frac{1}{x^2} - \frac{q-1}{(1-x)^2} + q\psi''(x),$$
(307)

$$f'''(x) = \frac{2}{x^3} - \frac{2(q-1)}{(1-x)^3} + q\psi'''(x).$$
(308)

So

$$f'\left(\frac{1}{q}\right) = 0,\tag{309}$$

$$f''\left(\frac{1}{q}\right) = -\frac{2q^3}{q-1},$$
(310)

$$f'''\left(\frac{1}{q}\right) = \frac{3(q-2)q^4}{(q-1)^2}.$$
(311)

We have

$$(\psi''f''' - f''\psi''')\left(\frac{1}{q}\right) = \frac{q^2}{q-1} \cdot \frac{3(q-2)q^4}{(q-1)^2} - \left(-\frac{2q^3}{q-1}\right)\left(-\frac{q^3(q-2)}{(q-1)^2}\right)$$
$$= \frac{q^6(q-2)}{(q-1)^3} > 0.$$
(312)

So Lemma 44 applies.

Part (i), p > 1. For b_p with p > 1, take

$$f(x) = q - (q - 1)\xi_p(x)$$

= $\left(x^{\frac{1}{p}} + (q - 1)\left(\frac{1 - x}{q - 1}\right)^{\frac{1}{p}}\right)\left(x^{1 - \frac{1}{p}} + (q - 1)\left(\frac{1 - x}{q - 1}\right)^{1 - \frac{1}{p}}\right).$ (313)

For simplicity, write $r = \frac{1}{p}$ and let $u_r(x) = x^r + (q-1)\left(\frac{1-x}{q-1}\right)^r$. Then $f(x) = u_r(x)u_{1-r}(x)$. Let us compute derivatives of u_r .

$$u'_{r}(x) = r\left(x^{r-1} - \left(\frac{1-x}{q-1}\right)^{r-1}\right),$$
(314)

$$u_r''(x) = r(r-1)\left(x^{r-2} + \frac{1}{q-1}\left(\frac{1-x}{q-1}\right)^{r-2}\right),\tag{315}$$

$$u_r''(x) = r(r-1)(r-2)\left(x^{r-3} - \frac{1}{(q-1)^2}\left(\frac{1-x}{q-1}\right)^{r-3}\right).$$
 (316)

So

$$u_r\left(\frac{1}{q}\right) = q^{1-r},\tag{317}$$

$$u_r'\left(\frac{1}{q}\right) = 0,\tag{318}$$

$$u_r''\left(\frac{1}{q}\right) = r(r-1)\frac{q}{q-1}\left(\frac{1}{q}\right)^{r-2},$$
(319)

$$u_r'''\left(\frac{1}{q}\right) = r(r-1)(r-2)\frac{q(q-2)}{(q-1)^2}\left(\frac{1}{q}\right)^{r-3}.$$
(320)

Now we compute derivatives of f.

$$f'(x) = u'_r(x)u_{1-r}(x) + u_r(x)u'_{1-r}(x),$$
(321)

$$f''(x) = u_r''(x)u_{1-r}(x) + 2u_r'(x)u_{1-r}'(x) + u_r(x)u_{1-r}''(x),$$
(322)

$$f'''(x) = u_r'''(x)u_{1-r}(x) + 3u_r''(x)u_{1-r}'(x) + 3u_r'(x)u_{1-r}''(x) + u_r(x)u_{1-r}''(x).$$
(323)

So

$$f'\left(\frac{1}{q}\right) = 0,$$
(324)
$$f''\left(\frac{1}{q}\right) = r(r-1)\frac{q}{q-1}\left(\frac{1}{q}\right)^{r-2} \cdot q^r + (1-r)(-r)\frac{q}{q-1}\left(\frac{1}{q}\right)^{-r-1} \cdot q^{1-r}$$

$$= 2r(r-1)\frac{q^3}{(q-1)},$$
(325)

$$f'''\left(\frac{1}{q}\right) = r(r-1)(r-2)\frac{q(q-2)}{(q-1)^2}\left(\frac{1}{q}\right)^{r-3} \cdot q^r + (1-r)(-r)(-r-1)\frac{q(q-2)}{(q-1)^2}\left(\frac{1}{q}\right)^{-r-2} \cdot q^{1-r},$$
$$= -3r(r-1)\frac{q^4(q-2)}{(q-1)^2}.$$
(326)

So

$$(\psi'' f''' - f'' \psi''') \left(\frac{1}{q}\right) = \frac{q^2}{q-1} \left(-3r(r-1)\frac{q^4(q-2)}{(q-1)^2}\right) -2r(r-1)\frac{q^3}{(q-1)} \left(-\frac{q^3(q-2)}{(q-1)^2}\right) = r(1-r)\frac{q^6(q-2)}{(q-1)^3} > 0.$$
(327)

So Lemma 44 applies. **Part** (ii). For s_{λ} , take

$$f(x) = \psi\left(\lambda x + \frac{1-\lambda}{q}\right).$$
(328)

Then

$$f'(x) = \lambda \psi'\left(\lambda x + \frac{1-\lambda}{q}\right),\tag{329}$$

$$f''(x) = \lambda^2 \psi''\left(\lambda x + \frac{1-\lambda}{q}\right),\tag{330}$$

$$f'''(x) = \lambda^3 \psi'''\left(\lambda x + \frac{1-\lambda}{q}\right). \tag{331}$$

So

$$f'\left(\frac{1}{q}\right) = 0,\tag{332}$$

$$f''\left(\frac{1}{q}\right) = \lambda^2 \psi''\left(\frac{1}{q}\right) = \lambda^2 \frac{q^2}{q-1},\tag{333}$$

$$f'''\left(\frac{1}{q}\right) = \lambda^3 \psi'''\left(\frac{1}{q}\right) = -\lambda^3 \frac{q^3(q-2)}{(q-1)^2}.$$
 (334)

We have

$$(\psi''f''' - f''\psi''')\left(\frac{1}{q}\right) = \frac{q^2}{q-1}\left(-\lambda^3 \frac{q^3(q-2)}{(q-1)^2}\right) - \lambda^2 \frac{q^2}{q-1}\left(-\frac{q^3(q-2)}{(q-1)^2}\right)$$
$$= \frac{q^5(q-2)}{(q-1)^3}(\lambda^2 - \lambda^3) > 0.$$
(335)

So Lemma 44 applies.

D. Concavity of Log-Sobolev Coefficients

Let *K* be a Markov kernel with stationary distribution π . Define Dirichlet form $\mathcal{E}(\cdot, \cdot)$ and entropy form $\text{Ent}_{\pi}(\cdot)$ as in Sect. 1. For $r \in \mathbb{R}$, we consider the tightest $\frac{1}{r}$ -log-Sobolev inequality, corresponding to

$$\widetilde{b}_{\frac{1}{r}}(x) := \inf_{\substack{f: \mathcal{X} \to \mathbb{R}_{\ge 0}, \\ \mathbb{E}_{\pi}[f]=1, \operatorname{Ent}_{\pi}(f)=x}} \mathcal{E}(f^r, f^{1-r}),$$
(336)

$$\widetilde{\Phi}_{\frac{1}{r}}(y) := \sup_{\substack{f: \mathcal{X} \to \mathbb{R}_{\geq 0}, \\ \mathbb{E}_{\pi}[f] = 1, \mathcal{E}(f^{r}, f^{1-r}) = y}} \operatorname{Ent}_{\pi}(f).$$
(337)

The $\frac{1}{r}$ -log-Sobolev constant is

$$\widetilde{\alpha}_{\frac{1}{r}} := \inf_{x>0} \frac{b_{\frac{1}{r}}(x)}{x} = \inf_{y>0} \frac{y}{\Phi_{\frac{1}{r}}(y)}.$$
(338)

Remark 45. When r = 0, the fraction $\frac{1}{r}$ should be understood as a formal symbol, and by definition we have $\tilde{b}_{\frac{1}{0}}(x) = \tilde{b}_1(x)$ and $\tilde{\Phi}_{\frac{1}{0}}(y) = \tilde{\Phi}_1(y)$ whenever they are defined. For $r \in (0, 1)$, $\tilde{\Phi}_{\frac{1}{r}}(y) = \Phi_{\frac{1}{r}}(y)$ where $\Phi_{\frac{1}{r}}$ the (pointwise) smallest function satisfying (8), and $\tilde{\alpha}_{\frac{1}{r}} = \alpha_{\frac{1}{r}}$ where $\alpha_{\frac{1}{r}}$ is defined in (9). However, in general $\tilde{\alpha}_1$ is not equal to α_1 . We use $\tilde{\cdot}$ to emphasize the difference.

The following result says that log-Sobolev constants are concave in r.⁶

Proposition 46. We have

- (i) For fixed x, $\tilde{b}_{\underline{1}}(x)$ is concave in r.
- (*ii*) $\widetilde{\alpha}_{\underline{1}}$ *is concave in r*.

Furthermore, if (π, K) is reversible, then

(iii) For fixed x, $\tilde{b}_{\frac{1}{r}}(x)$ is maximized at $r = \frac{1}{2}$. (iv) $\tilde{\alpha}_{\frac{1}{2}}$ is maximized at $r = \frac{1}{2}$.

Proof. Because inf of concave functions is still concave, it suffices to prove that for any $f : \mathcal{X} \to \mathbb{R}_{\geq 0}$ with $\mathbb{E}_{\pi}[f] = 1$, $\mathcal{E}(f^r, f^{1-r})$ is concave in r.

$$\begin{aligned} \frac{d}{dr^2} \mathcal{E}(f^r, f^{1-r}) &= \frac{d}{dr^2} \sum_{x, y \in \mathcal{X}} (I - K)(x, y) f(y)^r f(x)^{1-r} \pi(x) \\ &= \sum_{x, y \in \mathcal{X}} (I - K)(x, y) f(y)^r f(x)^{1-r} \pi(x) (\log f(y) - \log f(x))^2 \\ &= \sum_{x \neq y \in \mathcal{X}} -K(x, y) f(y)^r f(x)^{1-r} \pi(x) (\log f(y) - \log f(x))^2 \\ &\leq 0. \end{aligned}$$

When the Markov chain is reversible, we have $\mathcal{E}(f, g) = \mathcal{E}(g, f)$. So $\tilde{b}_{\frac{1}{r}}(x) = \tilde{b}_{\frac{1}{1-r}}(x)$ and by concavity, $\tilde{b}_{\frac{1}{2}}(x)$ is maximized at $r = \frac{1}{2}$.

E. Non-reconstruction for Broadcasting with a Gaussian Kernel

In this section, we prove optimal non-reconstruction results for a BOT model with continuous alphabet considered in Eldan et al. [16], using our method developed in Sect. 4.

Definition 47 (*Broadcasting on trees with a Gaussian kernel*). In this model, we are given a (possibly) infinite tree T with a marked root ρ . The state space \mathcal{X} is the unit circle $S^1 = \mathbb{R}/2\pi\mathbb{Z}$. Let $\pi = \text{Unif}(S^1)$ be the uniform distribution. Let t > 0 be a parameter. The transfer kernel is M_t , defined as $Y = X + Z_t$ where $Z_t \sim \mathcal{N}(0, t)$, where X is the input and Y is the output.

Now for each vertex $v \in T$, we generate a label $\sigma_v \in \mathcal{X}$ according to the following process:

⁶ An earlier version of the paper incorrectly stated that $\Phi_{\frac{1}{r}}(y)$ is convex in *r* for fixed *y*. The incorrect statement was not used elsewhere in the paper.

- 1. Generate $\sigma_{\rho} \sim \pi$.
- 2. Suppose we have generated a label for vertex *u*. For every child *v* of *u*, we generate *v* according to $\sigma_v \sim M_t(\cdot | \sigma_u)$.

Let L_k denote the set of vertices at distance k to ρ . We say reconstruction is impossible if and only if

$$\lim_{k \to \infty} I(\sigma_{\rho}; \sigma_{L_k}) = 0.$$
(339)

Let $\lambda(M_t)$ denote the second largest eigenvalue of M_t . [16] proved that for the above BOT model on a regular tree with offspring *d*, reconstruction holds when $d\lambda(M_t)^2 > 1$, and non-reconstruction holds for $d\lambda(M_t) < 1$. Note that there is a $\lambda(M_t)$ factor gap between the reconstruction result and the non-reconstruction result. In the following, we prove that non-reconstruction holds as long as $d\lambda(M_t)^2 < 1$, closing the gap.

We remark that Mossel et al. [40] studied a different BOT model with Gaussian broadcasting channels, and deteremined the reconstruction threshold for their model (which happened to also coincide with the Kesten–Stigum threshold). While sharing some similarities, their and our models do not seem to be directly comparable with each other.

Theorem 48 (Non-reconstruction for Gaussian BOT model). *Consider the BOT model defined in Definition 47.*

Let T be an infinite rooted tree with bounded maximum degree. Then reconstruction is impossible when

$$br(T)\lambda(M_t)^2 < 1. \tag{340}$$

Let T be a Galton-Watson tree with expected offspring d. Then reconstruction is impossible when

$$d\lambda (M_t)^2 < 1. \tag{341}$$

The proof idea is to upper bound the input-restricted KL contraction coefficient by $\lambda(M_t)^2$, then use a tree recursion similar to that of Theorem 5. However, because we are working in a continuous space, we must be careful about what we mean by contraction coefficients.

We would like an inequality of form

$$I(\sigma_u; \sigma_{L_{v,k}}) \le \widetilde{\eta}_{\mathrm{KL}}(\pi, M_t) I(\sigma_v; \sigma_{L_{v,k}}) \tag{342}$$

where $u \in V(T)$, v is child of u, $L_{v,k}$ is the set of descendants of v at distance k to ρ , and $\tilde{\eta}_{\text{KL}}(\pi, M_t)$ is a continuous version of contraction coefficient $\eta_{\text{KL}}(\pi, M_t)$. We have

$$I(\sigma_u; \sigma_{L_{v,k}}) = \mathbb{E}_{\sigma_{L_{v,k}}} D(P_{\sigma_u | \sigma_{L_{v,k}}} \| P_{\sigma_u})$$
(343)

$$= \mathbb{E}_{\sigma_{L_{v,k}}} D(M_t \circ P_{\sigma_v | \sigma_{L_{v,k}}} \| \pi).$$
(344)

Let us consider the distribution $P_{\sigma_u | \sigma_{L_{u,k}}}$. If $k = d(v, \rho)$, then $P_{\sigma_u | \sigma_{L_{u,k}}}$ is a point measure. However, as long as $k > d(u, \rho)$, pdf of $P_{\sigma_u | \sigma_{L_{u,k}}}$ is smooth on \mathcal{X} by an induction using belief propagation equation. Therefore we make the following definition.

Definition 49 (Smooth contraction coefficient). We define

$$\widetilde{\eta}_{\mathrm{KL}}(\pi, M_t) := \sup_{f \in \mathcal{C}} \frac{\mathrm{Ent}_{\pi}(M_t f)}{\mathrm{Ent}_{\pi}(f)},\tag{345}$$

$$\mathcal{C} := \{ f : \mathcal{X} \to \mathbb{R}_{\geq 0} | f \text{ smooth}, \mathbb{E}_{\pi}[f] = 1 \}.$$
(346)

where $\operatorname{Ent}_{\pi}(f)$ is defined in (2).

Lemma 50.

$$\widetilde{\eta}_{KL}(\pi, M_t) \le \exp(-t).$$
(347)

Proof. Note that $(M_t)_{t\geq 0}$ forms a semigroup. Therefore it suffices to prove that for all $f \in C$, we have

$$\frac{d}{dt}|_{t=0}\operatorname{Ent}_{\pi}(f_t) \le -\operatorname{Ent}_{\pi}(f)$$
(348)

where $f_t = M_t f$. We have

$$\begin{aligned} \frac{d}{dt}|_{t=0} \operatorname{Ent}_{\pi}(f_{t}) \\ &= \mathbb{E}\left[\frac{d}{dt}|_{t=0}(f_{t} \log f_{t})\right] \\ &= \mathbb{E}\left[(1 + \log f)\frac{d}{dt}|_{t=0}f_{t}\right] \\ &= \mathbb{E}\left[(\log f)\frac{d}{dt}|_{t=0}f_{t}\right] \\ &= \frac{1}{2}\mathbb{E}\left[f'' \log f\right] & \text{heat equation} \\ &= -\frac{1}{2}\mathbb{E}\left[\frac{(f')^{2}}{f}\right] & \text{integration by parts} \\ &\leq -\operatorname{Ent}_{\pi}(f). \end{aligned}$$

This finishes the proof.

Now we are ready to prove Theorem 48.

Proof of Theorem 48. By Lemma 50, we have

$$\widetilde{\eta}_{\mathrm{KL}}(\pi, M_t) \le \exp(-t) = \lambda (M_t)^2,$$
(349)

where the value of $\lambda(M_t)$ is proved in e.g., [16]. Therefore we only need to prove that $br(T)\tilde{\eta}_{KL}(\pi, M_t) < 1$ implies non-reconstruction. Note that the channel M_t is reversible.

Bounded degree case: For $u \in V(T)$, define

$$r_u := \lim_{k \to \infty} I(\sigma_u; \sigma_{L_{u,k}}).$$
(350)

By data processing inequality, $I(\sigma_u; \sigma_{L_{u,k}})$ is non-increasing for $k \ge d(u, \rho)$, so the limit always exists. Because *T* has bounded maximum degree, we have

$$r_u \le I(\sigma_u; \sigma_{L_{u,d(u,\rho)+1}}). \tag{351}$$

So there exists a constant C > 0 such that $r_u \le C$ for all $u \in v(T)$.

Now define

$$a_u = C^{-1} \widetilde{\eta}_{\mathrm{KL}}(\pi, M_t)^{d(u,\rho)} r_u.$$
(352)

Let c(u) be the set of children of u. For any $v \in c(u)$, by Markov chain

$$\sigma_{L_{v,k}} \to \sigma_v \to \sigma_u \tag{353}$$

and discussion before Lemma 50, we have

$$I(\sigma_{u}; \sigma_{L_{v,k}}) \leq \widetilde{\eta}_{\mathrm{KL}}(\pi, M_{t}) I(\sigma_{v}, \sigma_{L_{v,k}}).$$
(354)

Because $(\sigma_{L_{v,k}})_{v \in c(u)}$ are independent conditioned on σ_u , we have

$$I(\sigma_u; \sigma_{L_{u,k}}) \le \sum_{v \in c(u)} I(\sigma_u; \sigma_{L_{v,k}}).$$
(355)

Combining the two inequalities and let $k \to \infty$, we get

$$a_u \le \sum_{v \in c(u)} a_v. \tag{356}$$

Furthermore, we have $a_u \leq \widetilde{\eta}_{\text{KL}}(\pi, M_t)^{d(u,\rho)}$.

Now define a flow *b* as follows. For any $u \in V(T)$, let $u_0 = \rho, \ldots, u_\ell = u$ be the shortest path from ρ to *u*. Define

$$b_{u} = a_{u} \prod_{0 \le j \le \ell - 1} \frac{a_{u_{j}}}{\sum_{v \in c(u_{j})} a_{v}}.$$
(357)

(If $\sum_{v \in c(u_j)} a_v = 0$ for some *j*, then let $b_u = 0$.) Then we have

$$b_u = \sum_{v \in c(u)} b_v, \tag{358}$$

and that

$$b_u \le a_u \le \widetilde{\eta}_{\mathrm{KL}}(\pi, M_t)^{d(u,\rho)}.$$
(359)

By definition of branching number, we must have $b_{\rho} = 0$. Therefore $r_{\rho} = 0$ and non-reconstruction holds.

Galton–Watson tree case: Let D be the offspring distribution. We have

$$\begin{split} I(\sigma_{\rho}; \sigma_{L_{k}}|T) &\leq \mathbb{E}_{c(\rho)} \sum_{v \in c(\rho)} I(\sigma_{\rho}; \sigma_{L_{v,k}}|T) \\ &\leq \mathbb{E}_{c(\rho)} \sum_{v \in c(\rho)} \widetilde{\eta}_{\mathrm{KL}}(\pi, M_{t}) I(\sigma_{v}; \sigma_{L_{v,k}}|T_{v}) \\ &= \widetilde{\eta}_{\mathrm{KL}}(\pi, M_{t}) \mathbb{E}_{c(\rho)} \sum_{v \in c(\rho)} I(\sigma_{v}; \sigma_{L_{v,k}}|T_{v}) \\ &= \widetilde{\eta}_{\mathrm{KL}}(\pi, M_{t}) \mathbb{E}_{b \sim D} \left[bI(\sigma_{\rho}; \sigma_{L_{k-1}}|T) \right] \\ &= \widetilde{\eta}_{\mathrm{KL}}(\pi, M_{t}) dI(\sigma_{\rho}; \sigma_{L_{k-1}}|T). \end{split}$$

Here T_v denotes the subtree rooted at v. Because $I(\sigma_\rho; \sigma_{L_1}|T) < \infty$, when $d\tilde{\eta}_{\text{KL}}(\pi, M_t) < 1$, we have

$$\lim_{k \to \infty} I(\sigma_{\rho}; \sigma_{L_k} | T) = 0.$$
(360)

This finishes the proof.

References

- 1. Abbe, E., Sandon, C.: Achieving the KS threshold in the general stochastic block model with linearized acyclic belief propagation. Adv. Neural Inform. Process. Syst. 29, (2016)
- Ahlswede, R., Gács, P.: Spreading of sets in product spaces and hypercontraction of the Markov operator. Ann. Probab. 925–939 (1976)
- Banks, J., Moore, C., Neeman, J., Netrapalli, P.: Information-theoretic thresholds for community detection in sparse networks. In: Conference on Learning Theory, pp. 383–416. PMLR (2016)
- Bernstein, A.J.: Maximally connected arrays on the n-cube. SIAM J. Appl. Math. 15(6), 1485–1489 (1967)
- Bhatnagar, N., Sly, A., Tetali, P.: Reconstruction threshold for the hardcore model. In: Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques: 13th International Workshop, APPROX 2010, and 14th International Workshop, RANDOM 2010, Barcelona, Spain, September 1-3, 2010. Proceedings, pp. 434–447. Springer (2010)
- Bhatnagar, N., Vera, J., Vigoda, E., Weitz, D.: Reconstruction for colorings on trees. SIAM J. Discret. Math. 25(2), 809–826 (2011)
- Bleher, P.M., Ruiz, J., Zagrebnov, V.A.: On the purity of the limiting Gibbs state for the Ising model on the Bethe lattice. J. Stat. Phys. 79, 473–482 (1995)
- Bobkov, S.G., Tetali, P.: Modified logarithmic Sobolev inequalities in discrete settings. J. Theor. Probab. 19, 289–336 (2006)
- Braverman, M., Garg, A., Ma, T., Nguyen, H.L., Woodruff, D.P.: Communication lower bounds for statistical estimation problems via a distributed data processing inequality. In: Proceedings of the Forty-Eighth Annual ACM Symposium on Theory of Computing, pp. 1011–1020 (2016)
- Chansangiam, P.: Operator monotone functions: characterizations and integral representations (2013). arXiv:1305.2471
- Choi, M.-D., Ruskai, M.B., Seneta, E.: Equivalence of certain entropy contraction coefficients. Linear Algebra Appl. 208, 29–36 (1994)
- Cohen, J., Kempermann, J.H.B., Zbaganu, G.: Comparisons of Stochastic Matrices with Applications in Information Theory, Statistics, Economics and Population. Springer, New York (1998)
- Coja-Oghlan, A., Krzakala, F., Perkins, W., Zdeborová, L.: Information-theoretic thresholds from the cavity method. In: Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, pp. 146–157 (2017)
- Diaconis, P., Saloff-Coste, L.: Logarithmic Sobolev inequalities for finite Markov chains. Ann. Appl. Probab. 6(3), 695–750 (1996)
- Efthymiou, C.: Reconstruction/non-reconstruction thresholds for colourings of general Galton–Watson trees (2014). arXiv:1406.3617

- Eldan, R., Mikulincer, D., Pieters, H.: Community detection and percolation of information in a geometric setting. Comb. Probab. Comput. 31(6), 1048–1069 (2022)
- Émery, M., Yukich, J.E.: A simple proof of the logarithmic Sobolev inequality on the circle. Séminaire de probabilités de Strasbourg 21, 173–175 (1987)
- Evans, W., Kenyon, C., Peres, Y., Schulman, L.J.: Broadcasting on trees and the ising model. Ann. Appl. Probab. 410–433 (2000)
- Evans, W.S., Schulman, L.J.: Signal propagation and noisy circuits. IEEE Trans. Inf. Theory 45(7), 2367–2373 (1999)
- Formentin, M., Külske, C.: On the purity of the free boundary condition Potts measure on random trees. Stoch. Process. Appl. 119(9), 2992–3005 (2009)
- Goel, S.: Modified logarithmic Sobolev inequalities for some models of random walk. Stoch. Process. Appl. 114(1), 51–79 (2004)
- 22. Gross, L.: Logarithmic Sobolev inequalities. Am. J. Math. 97(4), 1061–1083 (1975)
- Gu, Y.: Channel Comparison Methods and Statistical Problems on Graphs. PhD thesis, Massachusetts Institute of Technology (2023)
- Hadar, U., Liu, J., Polyanskiy, Y., Shayevitz, O.: Communication complexity of estimating correlations. In: Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing, pp. 792–803 (2019)
- Hueston Harper, L.: Optimal assignments of numbers to vertices. J. Soc. Ind. Appl. Math. 12(1), 131–135 (1964)
- 26. Hart, S.: A note on the edges of the n-cube. Discret. Math. 14(2), 157-163 (1976)
- Kesten, H., Stigum, B.P.: Additional limit theorems for indecomposable multidimensional Galton–Watson processes. Ann. Math. Stat. 37(6), 1463–1481 (1966)
- Külske, C., Formentin, M.: A symmetric entropy bound on the non-reconstruction regime of Markov chains on Galton–Watson trees. Electron. Commun. Probab. 14, 587–596 (2009)
- 29. John, H., Lindsey, I.I.: Assignment of numbers to vertices. Am. Math. Mon. 71(5), 508–516 (1964)
- Liu, W., Ning, N.: Large degree asymptotics and the reconstruction threshold of the asymmetric binary channels. J. Stat. Phys. 174, 1161–1188 (2019)
- 31. Lyons, R.: Random walks and percolation on trees. Ann. Probab. 18(3), 931-958 (1990)
- Makur, A., Polyanskiy, Y.: Comparison of channels: criteria for domination by a symmetric channel. IEEE Trans. Inf. Theory 64(8), 5704–5725 (2018)
- 33. Martinelli, F., Sinclair, A., Weitz, D.: Fast mixing for independent sets, colorings, and other models on trees. Random Struct. Algor. **31**(2), 134–172 (2007)
- Massoulié, L.: community detection thresholds and the weak Ramanujan property. In: Proceedings of the Forty-Sixth Annual ACM Symposium on Theory of Computing, pp. 694–703 (2014)
- Mézard, M., Montanari, A.: Reconstruction on trees and spin glass transition. J. Stat. Phys. 124, 1317– 1350 (2006)
- Mossel, E., Neeman, J., Sly, A.: Reconstruction and estimation in the planted partition model. Probab. Theory Relat. Fields 162, 431–461 (2015)
- Mossel, E., Neeman, J., Sly, A.: A proof of the block model threshold conjecture. Combinatorica 38(3), 665–708 (2018)
- Mossel, E., Oleszkiewicz, K., Sen, A.: On reverse hypercontractivity. Geom. Funct. Anal. 23(3), 1062– 1097 (2013)
- 39. Mossel, E., Peres, Y.: Information flow on trees. Ann. Appl. Probab. 13(3), 817–844 (2003)
- 40. Mossel, E., Roch, S., Sly, A.: Robust estimation of latent tree graphical models: inferring hidden states with inexact parameters. IEEE Trans. Inf. Theory **59**(7), 4357–4373 (2013)
- 41. Mossel, E., Sly, A., Sohn, Y.: Exact phase transitions for stochastic block models and reconstruction on trees (2022). arXiv:2212.03362
- Ordentlich, O., Polyanskiy, Y.: Strong data processing constant is achieved by binary inputs. IEEE Trans. Inf. Theory 68(3), 1480–1481 (2021)
- Polyanskiy, Y., Samorodnitsky, A.: Improved log-Sobolev inequalities, hypercontractivity and uncertainty principle on the hypercube. J. Funct. Anal. 277(11), 108280 (2019)
- 44. Polyanskiy, Y., Wu, Y.: Strong data-processing inequalities for channels and Bayesian networks. In: Convexity and Concentration, pp. 211–249. Springer (2017)
- 45. Polyanskiy, Y., Yihong, W.: Application of the information-percolation method to reconstruction problems on graphs. Math. Stat. Learn. **2**(1), 1–24 (2020)
- Raginsky, M.: Strong data processing inequalities and φ-Sobolev inequalities for discrete channels. IEEE Trans. Inf. Theory 62(6), 3355–3389 (2016)
- 47. Rényi, A.: On measures of dependence. Acta Math. Hungar. 10(3-4), 441-451 (1959)
- Sarmanov, O.V.: Maximum correlation coefficient (nonsymmetric case). Select. Transl. Math. Stat. Probab. 2, 207–210 (1963)
- 49. Sly, A.: Reconstruction of random colourings. Commun. Math. Phys. 288(3), 943–961 (2009)

- 50. Sly, A.: Reconstruction for the Potts model. Ann. Probab. 39(4), 1365–1406 (2011)
- Wyner, A.D., Ziv, J.: A theorem on the entropy of certain binary sequences and applications: Part I. IEEE Trans. Inf. Theory 19(6), 769–772 (1973)
- Xu, A., Raginsky, M.: Converses for distributed estimation via strong data processing inequalities. In: 2015 IEEE International Symposium on Information Theory (ISIT), pp. 2376–2380. IEEE (2015)

Communicated by J. Ding