THEORY OF COMPUTING, Volume 11 (17), 2015, pp. 413–443
www.theoryofcomputing.org

SPECIAL ISSUE: APPROX-RANDOM 2013

Conditional Random Fields, Planted Constraint Satisfaction, and Entropy Concentration

Emmanuel Abbe Andrea Montanari*

Received October 27, 2013; Revised November 30, 2014; Published December 29, 2015

Abstract: This paper studies a class of probabilistic models on graphs, where edge variables depend on incident node variables through a fixed probability kernel. The class includes planted constraint satisfaction problems (CSPs), as well as other structures motivated by coding theory and community detection problems. It is shown that under mild assumptions on the kernel and for sparse random graphs, the conditional entropy of the node variables given the edge variables concentrates. This implies in particular concentration results for the number of solutions in a broad class of planted CSPs, the existence of a threshold function for the disassortative stochastic block model, and the proof of a conjecture on parity check codes. It also establishes new connections among coding, clustering and satisfiability.

ACM Classification: G.3, H.1.1, G.2.2

AMS Classification: 68Q87, 68P30, 05C80

Key words and phrases: planted SAT, planted CPS, clustering, stochastic block model, graph-based codes, entropy

1 Introduction

This paper studies a class of probabilistic models on graphs encompassing models from random combinatorial optimization, coding theory and machine learning. Depending on the context, the class may

A preliminary version of this paper appeared in the Proc. of RANDOM, Berkeley, 2013.

*Supported by NSF CCF-1319979, DMS-11-06627 and AFOSR FA9550-13-1-0036

 DOI: 10.4086/toc.2015.v011a017

be described as a family of planted constrained satisfaction problems (CSPs), encoded channels or conditional random fields. We start by providing motivations in CSPs.

CSPs are key components in the theory of computational complexity as well as important mathematical models in various applications of computer science, engineering and physics. In a CSP, a set of variables $x_1, \dots, x_n$ is required to satisfy a collection of constraints, each involving a subset of the variables. In many cases of interest, the variables are Boolean and the constraints are all of a common type: e. g., in k-SAT, the constraints require the OR of k Boolean variables or their negations to be TRUE, whereas in k-XORSAT, the XOR of the variables or their negations must equal zero. Given a set of constraints and a number of variables, the problem is to decide whether there exists a satisfying assignment. In random CSPs, the constraints are drawn at random from a given ensemble, keeping the constraint density[1] constant. In this setting, it is of interest to estimate the *probability* that a random instance is satisfiable. One of the fascinating phenomena occurring for random instances is the phase transition, which makes the task of estimating this probability much easier in the limit. For a large class of CSPs, and as n tends to infinity, the probability of being satisfiable tends to a step function, jumping from 1 to 0 when the constraint density crosses a critical threshold. For random k-XORSAT the existence of such a critical threshold is proved[2] in For random k-SAT, $k \geq 3$, the existence of a threshold that depends on n is proved in [32]. However it remains open to show that this threshold converges as $n \to \infty$. Upper and lower bounds are known to match up to a term that is of relative order $k 2^{-k}$ as k increases [9, 19]. Phase transition phenomena in other types of CSPs are also investigated in [8, 49, 9]

In planted random CSPs, a "planted assignment" is first drawn, and the constraints are then drawn at random so as to keep that planted assignment a satisfying one. Planted ensembles were investigated in [12, 37, 7, 6, 38, 5], and at high density in [10, 20, 28]. In the planted setting, the probability of being satisfiable is always equal to one by construction, and a more relevant question is to determine the actual *number* of satisfying assignments. One would expect that this problem becomes easier in the limit as $n \to \infty$ due to an asymptotic phenomenon. This paper shows that, indeed, a concentration phenomenon occurs: for a large class of planted CSPs (including SAT, NAE-SAT and XOR-SAT) the normalized logarithm of the number or satisfying assignments concentrates (with respect to the graph of the CSP) to a number. Moreover, we emphasize that this number is independent of n, unlike the SAT threshold.

It is worth comparing the result obtained in this paper for planted CSPs, with the one obtained in [4] for non-planted CSPs. In that case, the number of solutions is zero with positive probability and therefore the logarithm of the number of solution does not have a finite expectation. Technically, standard martingale methods do not allow to prove concentration, even to an n-dependent threshold. In [4] an interpolation method [36] is used to prove the existence of the limit of a "regularized" quantity, namely the logarithm of the number of solutions plus one, divided by the number of variables. A technical consequence of this approach is that the concentration of this quantity to a value that is independent of n can only be proved when the probability of unsatisfiability is known to be $O(1/\log(n)^{1+\varepsilon})$.

This paper shows that in the planted case the concentration around an n-independent value holds unconditionally. We again use the interpolation technique [36, 30, 31, 53, 13, 4] but with an interesting twist. While in all the cited references, the entropy (or log-partition function) is shown to be superadditive,

[1] The ratio of the expected number of constraints and the number of variables.

[2] Reference [24] contains a statement but proofs were omitted from the proceedings. They can be found as appendices in the arXiv version [25].

in the present setting it turns out to be subadditive. This flip from superadditivity to subadditivity has an intriguing information-theoretic interpretation. Considering for instance the application to sparse-graph codes, it corresponds to the fact that more information can be transmitted reliably using—say—a code of block length (n_1+n_2) than two codes of block lengths n_1 and n_2, respectively. While this intuition is folklore among practitioners, the present analysis provides an elegant formalization.

Let us also mention that a fruitful line of work has addressed the relation between planted random CSPs and their non-planted counterparts in the satisfiable phase [5, 40, 58]. These papers show that, when the number of solutions is sufficiently concentrated, planting does not play a critical role in the model. It would be interesting to use these ideas to "export" the concentration result obtained here to non-planted models.

In this paper, we pursue a different type of approach. Motivated by applications,[3] in particular in coding theory and clustering, we consider extensions of the standard planted CSPs to a setting allowing soft probabilistic constraints. Within the setting of soft CSPs, the planted solution is an unknown vector to be reconstructed, and the constraints are regarded as noisy observations of this unknown vector. For instance one can recover the case of planted random k-SAT as follows. Each clause is generated by selecting first k variable indices $i_1,\dots,i_k$ uniformly at random, providing a random hyperedge. Then a clause is drawn uniformly among the ones that are satisfied by the variables $x_{i_1},\dots,x_{i_k}$ appearing in the planted assignment. The clause can hence be regarded as a noisy observation of $x_{i_1},\dots,x_{i_k}$. More generally the formula can be seen as a noisy observation of the planted assignment.

Our framework extends the above to include numerous examples from coding theory, machine learning and statistics. Within LDPC or LDGM codes [55], encoding is performed by evaluating the modulo 2 sum of a random subset of information bits and transmitting it through a noisy communication channel. The selection of the information bits is described by a graph, drawn at random for the code construction, and the transmission of these bits leads to a noisy observation of the graph variables. Similarly, a community detection block model [34] can be seen as a random graph model, whereby each edge is a noisy observation of the community assignments of the incident nodes. Definitions will be made precise in the next section.

The conditional probability of the unknown vector given the noisy observations takes the form of a graphical model, i. e., factorizes according to a hypergraph whose nodes correspond to variables and hyperedges correspond to noisy observations. Such graphical models have been studied by several authors in machine learning [43] under the name of "conditional random fields," and in [48] in the context of LDPC and LDGM codes. The conditional entropy of the unknown vector given the observations is used here to quantify the residual uncertainty of the vector. This is equivalent to considering the mutual information between the node and edge variables. In such a general setting, we prove that the conditional entropy per variable tends to a limit. This framework allows a unified treatment of a large class of random combinatorial optimization problems, raises new connections among them, and opens up to new models. We obtain in particular a proof of a conjecture posed in [42] on low-density parity-check codes, and the existence of a threshold function for the disassortative stochastic block model [23].

[3] Planted models are also appealing to cryptographic application, as hard instances with known solutions provide good one-way functions [35, 16].

2 The model

Let k and n be two positive integers with $n \geq k$.

- Let $V = [n]$ and $g = (V, E(g))$ be a hypergraph with vertex set V and edge set $E(g) \subseteq E_k(V)$, where $E_k(V)$ denotes the set of all possible n^k hyperedges of order k on the vertex set V. We will often drop the prefix "hyper." For the purpose of this paper, one can equivalently work with the model in which the hyperedges as drawn from all $\binom{n}{k}$ hyperedges without replacement of the vertices.
- Let $\mathcal{X}$ and $\mathcal{Y}$ be two finite sets called respectively the input and output alphabets. Let $Q(\cdot|\cdot)$ be a probability transition function (or channel) from $\mathcal{X}^k$ to $\mathcal{Y}$, i. e., for each $u \in \mathcal{X}^k$, $Q(\cdot|u)$ is a probability distribution on[4] $\mathcal{Y}$.
- To each vertex in V, we assign a node-variable taking values in $\mathcal{X}$, and to each edge in $E(g)$, we assign an edge-variable taking values in $\mathcal{Y}$. We now define a type of factor model where the probability of the edge-variables given the node-variables is given by

$$P_{g,Q}(y|x) \equiv \prod_{I \in E(g)} Q(y_I \mid x[I]), \quad x \in \mathcal{X}^V,\, y \in \mathcal{Y}^{E(g)}, \tag{2.1}$$

  where y_I denotes the edge-variable attached to edge I, and $x[I]$ denotes the k node-variables attached to the vertices incident with edge I (where the entries of $x[I]$ are placed in the increasing order of I, although we will only consider kernels that are invariant under this ordering).

The above is a type of factor or graphical model, or a planted constraint satisfaction problem with soft probabilistic constraints. For each $x \in \mathcal{X}^V$, $P_{g,Q}(\cdot \mid x)$ is a product measure on the set of edge-variables. We call $P_{g,Q}$ a *graphical channel with graph g and kernel Q*. We next put the uniform probability distribution on the set of node-variables $\mathcal{X}^V$, and define the a posteriori probability distribution (or reverse channel) by

$$R_{g,Q}(x \mid y) \equiv \frac{1}{S_{g,Q}(y)} P_{g,Q}(y \mid x)|\mathcal{X}|^{-n}, \quad x \in \mathcal{X}^V,\, y \in \mathcal{Y}^{E(g)}, \tag{2.2}$$

where y belongs to the support of $P_{g,Q}(\cdot \mid x)$ and

$$S_{g,Q}(y) \equiv \sum_{x \in \mathcal{X}^V} P_{g,Q}(y \mid x)|\mathcal{X}|^{-n} \tag{2.3}$$

is the marginal distribution of the edge variables y.

Example: planted SAT model. We now show how previous model reduces to planted k-SAT for a specific choice of the kernel. First, we take $\mathcal{X} = \{0,1\}$, i. e., the variables in planted SAT are Boolean, and $x \in \{0,1\}^n$ represents the planted assignment. We want to generate a random k-CNF formula with clauses that keep x a satisfying assignment. Notice that a k-clause is specified by two entities, (i) the variables that are in the clause, i. e., an index set $I \subseteq [n]$ of cardinality $|I| = k$, (ii) the negations that

[4]Note that we will allow $\mathcal{Y}$ to depend on k.

come on some of these variables, which can be encoded by a binary vector $y_I \in \{0,1\}^k$. For example, the 3-clause $u_2 \wedge \bar{u}_7 \wedge \bar{u}_{12}$ (where we use u for the formula's variables, which are dummies in contrast to the planted assignment x) is specified by the index set $(2,7,12)$ and by the negation pattern $(0,1,1)$. Requiring that this clause is satisfied is equivalent to require that $(u_2,u_7,u_{12}) \neq (0,1,1)$ (or equivalently, $u[I] \neq y_I$ in general). This means that the triplet (u_2,u_7,u_{12}) must be different than the triplet $(0,1,1)$, though a strict subset of components can be equal. In the model that we defined above, the edges of the graph play the role of the index sets I, and the output y_I of the kernel Q plays the role of the negation pattern (in particular the output alphabet is $\mathcal{Y} = \{0,1\}^k$). The planted formula is hence defined by the pair (g,y), where g is the graph and $y \in \mathcal{Y}^{|E(g)|}$ are the negation patterns. When working with non-planted SAT, the negations y_I do typically not depend on a specific x and on I, and y_I is uniformly drawn. However, for planted-SAT, y_I depends on the values of the planted assignments $x[I]$, and this dependency is captured by the kernel Q. For a uniform planted SAT model, which we consider in this paper, the following kernel is used

$$Q_{\mathrm{sat}}(y_I \mid x[I]) = \frac{1}{2^k-1} \mathbb{1}(y_I \neq x[I])\,.$$

In words, we draw the negation pattern uniformly at random among all patterns that keep the planted assignment a satisfying one.

We now define two probability distributions on the hypergraph g, which are equivalent for the purposes of this paper:

- A sparse Erdős-Rényi distribution, where each edge is drawn independently with probability $p = \alpha n/n^k$, where $\alpha > 0$ is the edge density.

- A sparse Poisson distribution, where for each $I \in E_k(V)$, a number of edges m_I is drawn independently from a Poisson distribution of parameter $p = \alpha n/n^k$. Note that m_I takes value in $\mathbb{Z}_+$, hence G is now a multi-edge hypergraph. To cope with this more general setting, we allow the edge-variable y_I to take value in $\mathcal{Y}^{m_I}$, i. e., $y_I = (y_I(1),\ldots,y_I(m_I))$, and define (with a slight abuse of notation)

  $$Q(y_I \mid x[I]) = \prod_{i=1}^{m_I} Q(y_I(i) \mid x[I])\,. \tag{2.4}$$

  This means that for each I, if $m_I \geq 1$, then m_I i. i. d. outputs are drawn from the kernel Q, and if $m_I = 0$, no edge is drawn. We denote by $\mathcal{P}_k(\alpha,n)$ this distribution on (multi-edge) hypergraphs.

Since $p = \alpha n/n^k$, the number of edges concentrates around its expectation given by αn and the two models can be shown to be equivalent for our purposes in the limit as $n \to \infty$.

3 Main Results

3.1 Preliminaries

We start by reviewing some basic definitions of information measures used in this paper. Recall that for a random variable Z having distribution P_Z supported on a finite set $\mathcal{Z}$, the entropy (in bits) of Z or equivalently of P_Z is defined by

$$H(Z) = H(P_Z) = -\sum_{z\in\mathcal{Z}} P_Z(z)\log_2 P_Z(z)\,.$$

In particular, $0 \le H(Z) \le \log_2|\mathcal{Z}|$, where the first inequality is met only for Z deterministic and the second inequality is met only for Z uniform on $\mathcal{Z}$. If W is a random variable having distribution P_W supported on a finite set $\mathcal{W}$, and (W,Z) has a joint probability distribution $P_{W,Z}$ on $\mathcal{W}\times\mathcal{Z}$, then the conditional entropy of Z given that $W = w \in \mathcal{W}$ is defined by

$$H(Z \mid W = w) = H(P_{Z|W=w}) = -\sum_{z\in\mathcal{Z}} P_{Z|W=w}(z \mid w)\log_2 P_{Z|W=w}(z \mid w)\,,$$

where

$$P_{Z|W=w}(z \mid w) = P_{W,Z}(w,z)/P_W(w)$$

is the conditional distribution of Z given that $W = w$. In particular, $0 \le H(Z \mid W = w) \le \log(|\mathcal{Z}|)$ where the first inequality is met only if Z is deterministic given that $W = w$ and the second inequality is met only if Z is uniform on $\mathcal{Z}$ given that $W = w$. The conditional entropy of Z given W is then given by

$$H(Z \mid W) = \sum_{w\in\mathcal{W}} P_W(w)H(Z \mid W = w)$$

and this is a scalar (and not a random variable in contrast to the conditional expectation $E(Z \mid W)$). In particular $0 \le H(Z \mid W) \le \log(\mathcal{Z})$ where the first inequality is met only if Z is a deterministic function of W and the second inequality is met only if Z is independent of W and uniformly distributed on $\mathcal{Z}$. Note that if W_1,W_2 are two random variables jointly distributed with Z under $P_{W_1,W_2,Z}$ on $\mathcal{W}_1\times\mathcal{W}_2\times\mathcal{Z}$, then

$$H(Z \mid W_1, W_2 = w_2) = \sum_{w_1\in\mathcal{W}_1} H(Z \mid W_1 = w_1, W_2 = w_2)P_{W_1|W_2=w_2}(w_1 \mid w_2)\,.$$

Finally, the mutual information between W and Z is given by

$$I(W;Z) = H(W) - H(W \mid Z) = H(Z) - H(Z \mid W)\,.$$

We refer to [22] for more details on these quantities.

In the sequel, we let X be uniformly distributed in $\mathcal{X}^n$, G be a random sparse hypergraph drawn from the $\mathcal{P}_k(\alpha,n)$ ensemble independently of X. For a realization $G = g$, we let Y_g be the output of X through the graphical channel $P_{g,Q}$ defined in (2.1) for a kernel Q. We also define

$$H_g(X \mid Y_g) \equiv H(X \mid Y_g, G = g) = -|\mathcal{X}|^{-n}\sum_{x\in\mathcal{X}^V}\sum_{y\in\mathcal{Y}^{E(g)}} P_{g,Q}(y \mid x)\log R_{g,Q}(x \mid y)\,, \tag{3.1}$$

where $P_{g,Q}$ and $R_{g,Q}$ are defined in (2.1) and (2.2) respectively. Note that $H_g(X \mid Y_g)$ is precisely the conditional entropy of X given Y_g and $G = g$ as defined in the previous paragraph. We use the notation $H_g(X \mid Y_g)$ with the subscript g rather than the notation $H(X \mid Y_g, G = g)$ as it allows us to let G be a random graph and consider $H_G(X \mid Y_G)$ as a random variable (which is a function of G). Instead,

$H(X \mid Y_G, G)$ denotes the expectation of $H(X \mid Y_g, G = g)$ over g as traditionally used in information theory (see previous paragraph), i. e.,

$$H(X \mid Y_G, G) = \mathop{\mathbb{E}}_{G} H_G(X \mid Y_G)\,. \tag{3.2}$$

We will often drop the subscript g (or G) in Y_g (or Y_G) as the Y variable is always implicitly defined from the underlying graph.

Finally, we denote the mutual information between the node and edge variables as

$$I_g(X \mid Y_g) = n \log |\mathcal{X}| - H_g(X \mid Y_g)$$

and the expected value

$$I(X \mid Y, G) = n \log |\mathcal{X}| - H(X \mid Y, G)\,.$$

Example: conditional entropy for planted SAT. We now explain the meaning of the conditional entropy for the planted-SAT model. Recall the definition of planted k-SAT in terms of graphical channels discussed in previous section. The alphabets are $\mathcal{X} = \{0,1\}$, $\mathcal{Y} = \{0,1\}^k$, the input $x \in \mathcal{X}^n$ represents the planted assignment, and for each edge (or index set) $I \in [n]^k$, the kernel Q_{sat} takes the variables $x[I]$ and outputs the negation pattern y_I uniformly at random in $\{0,1\}^k \setminus \{x[I]\}$, so that the I-clause is given by $u[I] \neq y_I$ (where u are the formula's variables) and the formula is specified by (g, y). Assuming that x is drawn uniformly at random in $\mathcal{X}^n$, we can write the reverse distribution of the planted assignment $x \in \mathcal{X}^n$ given a negation pattern $y \in \mathcal{Y}^{|E(g)|}$ on the graph g, which is simply the uniform distribution on all elements of $\mathcal{X}^n$ (the potential planted assignments) that satisfy the planted formula (g, y). This is formally shown by taking the definition of the planted-SAT kernel,

$$Q_{\mathrm{sat}}(y_I \mid x[I]) = \frac{1}{2^k - 1} \mathbb{1}(y_I \neq x[I]) \tag{3.3}$$

and plugging it in the definition of $R_{g,Q}(x \mid y)$ (see (2.2)):

$$R_{g,Q_{\mathrm{sat}}}(x \mid y) = \frac{1}{\sum_u \prod_{e \in g} \mathbb{1}(u[I] \neq y_I)} \prod_{e \in g} \mathbb{1}(x[I] \neq y_I) \tag{3.4}$$

$$= \begin{cases} 0 & \text{if } x \text{ is not a satisfiable assignment of the formula } (g, y), \\ \frac{1}{Z_g(y)} & \text{otherwise,} \end{cases} \tag{3.5}$$

where

$$Z_g(y) = \sum_u \prod_{e \in g} \mathbb{1}(u[I] \neq y_I)$$

is the number of satisfying assignments of the formula (g, y). Hence, for a fixed formula, the conditional entropy $H(X \mid Y_g = y, G = g)$ is simply the logarithm of the number of satisfying assignments. Moreover, for a random graph G (e. g., under the models of previous section), $H(X \mid Y_G, G)$ is the expected logarithm of the number of satisfying assignments of a random planted formula. Since $H(X \mid Y_g = y, G = g) \in [0, n]$, the normalized condition entropy belongs to $[0, 1]$, and captures the exponent at which the number of solutions grows in the exponential scale.

Note that in this paper, we are interested in models that go beyond CSPs, such as for coding or clustering problems, which we cast as soft CSPs induced by a kernel Q that is typically supported on all possible inputs. In such models, the reverse channel is typically not uniform on a subset of "satisfying" inputs as for planted SAT, and all "planted assignments" have non-zero probability. Hence, the number of satisfying assignment is not relevant here (it is the entire space), but we are interested in how likely are the different assignments to a given channel output. In particular, the conditional entropy captures how much information the model output contains about the input: if $H(X \mid Y_G, G)/n$ is close to 0, then the input (i. e., the planted assignment) is determined by the output up to an $o(n)$ entropy, whereas if $H(X \mid Y_G, G)/n$ tends to 1, then the output contains no significant trace (i. e., it only removes an $o(n)$ entropy) about the input. For the case of clustering, this captures how much information the graph tells about the clusters. In particular, the limit of the normalized conditional entropy is strictly less than 1 if and only if the clusters can be detected (i. e., recoverable with better asymptotic accuracy than what a random guess provides [46, 51]). Similarly, for coding problems, the conditional entropy leads directly to the mutual information, which captures how much information the underlying channel carries [22].

We next introduce an operator which plays a crucial role in proving that $H(X \mid Y, G)/n$ admits a limit.

Definition 3.1. We denote by $M_1(\mathcal{X}^l)$ the set of probability measures on $\mathcal{X}^l$. For a kernel Q from $\mathcal{X}^k$ to $\mathcal{Y}$, we define

$$\Gamma_l : M_1(\mathcal{X}^l) \to \mathbb{R}, \tag{3.6}$$

$$\nu \mapsto \Gamma_l(\nu) = \frac{1}{|\mathcal{Y}|} \sum_{u^{(1)},\dots,u^{(l)} \in \mathcal{X}^k} \left[\sum_{y \in \mathcal{Y}} \prod_{r=1}^{l} \big(1 - Q(y \mid u^{(r)})\big) \right] \prod_{i=1}^{k} \nu(u_i^{(1)}, \dots, u_i^{(l)}) . \tag{3.7}$$

Hypothesis H. A kernel Q is said to satisfy Hypothesis H if Γ_l is convex for any $l \geq 1$.

Despite the lengthy expression, it is important to note that the definition of Γ_l depends solely on the kernel Q. The operator can be interpreted as follows. Let $U = (U_1, \dots, U_k)$ be i. i. d. random vectors taking values in $\mathcal{X}^l$ under the distribution ν_l. Each vector U_i for $i \in [k]$ has hence l components denoted by $U_i^{(1)}, \dots, U_i^{(l)}$ and distributed under ν_l. Denote $U^{(l)} = (U_1^{(l)}, \dots, U_k^{(l)})$. For Y uniformly drawn in $\mathcal{Y}$, the operator Γ_l is equivalently defined by the map that sends ν_l to

$$\mathop{\mathbb{E}}_{Y,U} \prod_{r=1}^{l} \big(1 - Q(Y \mid U^{(r)})\big) .$$

The role of Γ_l appears in the proof of Theorem 3.2, to establish that $f(n) = H(X \mid Y, G)$ is a super-additive function, i. e., that $f(n) \leq f(n_1) + f(n_2)$, where $n_1 + n_2 = n$. We provide here a high-level insight. Consider for example the case of Q corresponding to a planted CSP, say k-SAT. In order to show that $f(n) \leq f(n_1) + f(n_2)$, one would like to compare the number of solutions $Z(\alpha, n)$ of a random k-CNF formula with density α on n variables, with the number of solutions of two independent random k-CNF formulae with density α on n_1 and n_2 variables respectively (where $n_1 + n_2 = n$). More precisely, the problem is to determine whether a general inequality holds between $Z(\alpha, n)$ and the product $Z(\alpha, n_1) \cdot Z(\alpha, n_2)$, i. e., between $f(n)$ and $f(n_1) + f(n_2)$. It is delicate to compare these quantities. For example, the inequality does not go in the same direction for k-SAT and planted k-SAT as demonstrated in this paper. At a very high-level, the proof manages to express the super-additivity, which can be thought as a form of convexity property of f with respect to $n = n_1 + n_2$, in terms of a formal convexity property of the operator Γ_l. This operator acts on the empirical distribution of l solutions drawn uniformly at random from a random k-CNF formula and corresponding to the variables $U^{(1)}, \ldots, U^{(l)}$ previously defined. By definition, the empirical distribution counts the occurrence of the $\{0,1\}$-patterns that appear in the l solutions, and this counting can be done separately in the two subsystems of n_1 and n_2 variables, and be added to obtain the counting in the global system of n variables. In other words, denoting by nP the count in the full system and by $n_1 P_1$ and $n_2 P_2$ the counts in the subsystems, the empirical distribution is additive: $nP = n_1 P_1 + n_2 P_2$. The proof expresses the gap $f(n) - f(n_1) - f(n_2)$ in terms of the gaps $n\Gamma_l(P) - n_1\Gamma_l(P_1) - n_2\Gamma_l(P_2)$, for $l \geq 1$, hence allowing us to conclude with a formal convexity argument when Hypothesis H holds.

We will see in Section 6 that a large variety of kernels satisfy this hypothesis, including kernels corresponding to parity-check encoded channels, planted SAT, NAE-SAT, XORSAT (for even clause size), and disassortative stochastic block models. There are also natural examples of kernels for which Hypothesis H is not satisfied, for example 3-XORSAT. We do not expect that Hypothesis H is necessary to obtain convergence of the normalized conditional entropy. It appears to be a technical requirement dictated by our proof technique.

3.2 Results

We first show the sub-additivity property for the expected conditional entropy of graphical channels.

Theorem 3.2. *Let Q be a kernel satisfying Hypothesis H. Let $n_1, n_2 \geq k$ and $n = n_1 + n_2$. We denote by G_n a hypergraph drawn from the ensemble $\mathcal{P}_k(\alpha, n)$ and G_{n_i} two hypergraphs drawn independently from the ensembles $\mathcal{P}_k(\alpha, n_i)$, respectively for $i = 1, 2$. Define $f(n) = H(X \mid Y, G_n)$, and $f(n_i) = H(X \mid Y, G_{n_i})$, $i = 1, 2$. Then*

$$f(n) \leq f(n_1) + f(n_2)\,. \tag{3.8}$$

The proof of this theorem is outlined in Section 4. Since f is sub-additive, it follows from Fekete's Lemma (see for example [54]) that it converges (to a finite value since f is lower-bounded by 0).

Corollary 3.3. *Let Q be a kernel satisfying Hypothesis H and G_n be a random hypergraph drawn from the ensemble $\mathcal{P}_k(\alpha, n)$. There exists a constant $C_k(\alpha, Q)$ such that*

$$\lim_{n\to\infty} \frac{1}{n} H(X \mid Y, G_n) = C_k(\alpha, Q)\,. \tag{3.9}$$

The following is obtained using the previous corollary and Azuma-Hoeffding's inequality.

Theorem 3.4. *Let Q be a kernel satisfying Hypothesis H and G_n be a random hypergraph drawn from the ensemble $\mathcal{P}_k(\alpha, n)$. Then, almost surely,*

$$\lim_{n\to\infty} \frac{1}{n} H_{G_n}(X \mid Y) = C_k(\alpha, Q)\,, \tag{3.10}$$

with $C_k(\alpha, Q)$ as in Corollary 3.3.

Note that the almost sure convergence is with respect to the graph G_n. Hence, the result says that with high probability on the drawing of G_n, the conditional entropy $H_{G_n}(X \mid Y)$ is close to a deterministic number $C_k(\alpha, Q)$. The proof is given in Section 5.

In Section 6, we consider three applications of the preceding results:

- In Section 6.1 we consider planted CSPs such as planted k-SAT, k-NAE-SAT and k-XORSAT and show that the normalized logarithm of the number of solutions concentrates with respect to the drawing of the random graph to an n-independent limit.

- In Section 6.2, we consider the stochastic block model with two communities, or equivalently, the planted partition model. We show that for the disassortative case, the normalized conditional entropy of the node variables (the community assignments) given the graph converges to a fixed value. In other words, the uncertainty on the community assignments after observing the graph is asymptotically determined by a fixed valued depending on the connectivity parameters.

- In Section 6.3, we consider LDGM binary codes over symmetric channels and prove the conjecture of [42] on the concentration of the mutual information for k even.

4 Proof of Theorem 3.2: Interpolation method for graphical channels

Notations. Recall that the considered model X is a random uniform vector of dimension n with components valued in $\mathcal{X}$, the graph G (a random graph on the vertex set $[n]$ with edges of order k), and the edge variables Y (a random vector of dimension $|E(G)|$ with components valued in the finite set $\mathcal{Y}$ and indexed by the edges of G). Note that X and G are drawn independently, while Y depends on both G and X, i. e., Y is the output of X on the graphical channel $P_{G,Q}$. While X, G and Y all depend on n, we emphasize the dependency in n only through G in what follows to simplify the notation. We also keep writing Y instead of Y_G even though Y changes if G changes. When carrying expansions where G does not change, we may write $H(X \mid Y)$ instead of $H(X \mid Y, G)$.

For a random variable Z with distribution P_Z, we use the notation

$$\mathop{\mathbb{E}}_{Z} f(Z) = \sum_{z} f(z) P_Z(z)\,, \tag{4.1}$$

and we may simply write $\mathbb{E}\, f(Z)$ when there is no ambiguity about the expectation. For two random variables W, Z with conditional distribution $\mathbb{P}\{W = w \mid Z = z\} = P_{W|Z}(w \mid z)$, we use the notation

$$\mathop{\mathbb{E}}_{W|Z} g(W,Z) = \sum_{w} g(w,Z) P_{W|Z}(w \mid Z)\,, \tag{4.2}$$

note that $\mathbb{E}_{W|Z}\, g(W,Z)$ is a function of Z, hence a random variable.

Our goal is to show the sub-additivity of $H(X \mid Y, G_n)$ as a function[5] of n, namely

$$H(X \mid Y, G_n) \leq H(X \mid Y, G_{n_1}) + H(X \mid Y, G_{n_2}), \tag{4.3}$$

where $n_1 + n_2 = n$. To understand the meaning of this inequality, consider a partition the set of vertices $[n]$ into two disjoint sets of size n_1 and n_2 with $n_1 + n_2 = n$. Let g be a graph on the vertex set $[n]$, and denote by g_1 and g_2 the restriction of g on each of these subsets. Then the following is obtained simply from the fact that conditioning reduced entropy:

$$H(X \mid Y, g) \leq H(X \mid Y, g_1) + H(X \mid Y, g_2). \tag{4.4}$$

Hence, the above is also true for a random graph G drawn from the ensemble $\mathcal{P}_k(\alpha, n)$. However, the random graph obtained by restricting G_n to a subset of n_i vertices is not equivalent to G_{n_i}, which is drawn from the ensemble $\mathcal{P}_k(\alpha, n_i)$, $i = 1, 2$. The issue is that when taking the restriction, the probability of an edge stays at $\alpha n / n^k$, whereas it should be[6] $\alpha n_i / n_i^k$ under the ensemble $\mathcal{P}_k(\alpha, n_i)$. Consequently, the above does not imply

$$H(X \mid Y, G_n) \leq H(X \mid Y, G_{n_1}) + H(X \mid Y, G_{n_2}).$$

To obtain the proper term on the right hand side, one should add the edges lost in the splitting of the vertices (e. g., using a coupling argument), but this gives a lower bound on the right hand side of (4.4) instead of an upper bound. This also shows that it may not be obvious to guess the direction of the inequality (i. e., sub- versus super-additivity). We rely on an interpolation method to compare the right quantities.

The interpolation method was first introduced in [36] for the Sherrington-Kirkpatrick model. This is a model for a spin-glass (i. e., a spin model with random couplings) on a complete graph. It was subsequently shown in [30, 31, 53] that the same ideas can be generalized to models on random sparse graphs, and applications in coding theory and random combinatorial optimization were proposed in [47, 41] and [13, 4]. We next develop an interpolation method to estimate the conditional entropy of general graphical channels. Interestingly, we will see that the planting flips the behaviour of the entropy from super to sub-additive.

Definition 4.1. We define a more general Poisson model for the random graph, where a parameter $\varepsilon_I \geq 0$ is attached to each $I \in E_k(V)$ (where $E_k(V)$ denotes the set of all n^k edges of order k on $V = [n]$), and the number of edges $m_I(\varepsilon_I)$ is drawn from a Poisson distribution of parameter ε_I. This defines a random multi-edge hypergraph whose edge probability is not homogenous but depends on the parameters ε_I. Denoting by $\underline{\varepsilon}$ the collection of all n^k parameters ε_I, we denote this ensemble as $\mathcal{P}_k(\underline{\varepsilon}, n)$. If for all I, $\varepsilon_I = \alpha n / n^k$, then $\mathcal{P}_k(\underline{\varepsilon}, n)$ reduces to $\mathcal{P}_k(\alpha, n)$ as previously defined.

[5] The reader will notice that this corresponds to super-additivity of the mutual information

$$I(X; Y, G_n) \geq I(X; Y, G_{n_1}) + I(X; Y, G_{n_2}).$$

This corresponds to the intuition that more information is conveyed using a graph of size $(n_1 + n_2)$, than two graphs of sizes n_1, and n_2.

[6] When working with the Poisson model, the probabilities $\alpha n_i / n_i^k$ should be multiplied by $\exp(-\alpha n_i / n_i^k)$.

Lemma 4.2. *Let X be uniformly drawn over $\mathcal{X}^n$, $G(\underline{\varepsilon})$ be a random hypergraph drawn from the ensemble $\mathcal{P}_k(\underline{\varepsilon}, n)$ independently of X, and let Y be the output of X through $P_{G(\underline{\varepsilon}),Q}$ defined in (2.1) for the kernel Q. Define also*

$$H(Q) \equiv -2^{-k} \sum_{u \in \mathcal{X}^k, z \in \mathcal{Y}} Q(z \mid u) \log Q(z \mid u) \,.$$

Then

$$\frac{\partial}{\partial \varepsilon_I} H(X \mid Y, G(\underline{\varepsilon})) = H(Q) - H(Y_I' \mid Y, G(\underline{\varepsilon}))\,, \tag{4.5}$$

where Y_I' and Y are independent conditionally on X (i. e., Y_I' is drawn under $Q(\cdot \mid X[I])$ and Y is drawn independently under $R_{G(\underline{\varepsilon}),Q}(\cdot \mid X)$).

Proof. Note that for a random variable Z_ε which is Poisson distributed of parameter ε, and a function f,

$$\frac{\partial}{\partial \varepsilon} \mathbb{E} f(Z_\varepsilon) = \mathbb{E} f(Z_\varepsilon + 1) - \mathbb{E} f(Z_\varepsilon)\,. \tag{4.6}$$

Since

$$\begin{aligned} H(X \mid Y, G(\varepsilon)) &= \sum_{g \in \mathbb{Z}_+^{n^k}} H(X \mid Y, G(\varepsilon) = g) \mathbb{P}\{G(\varepsilon) = g\} && (4.7)\\ &= \sum_{g \in \mathbb{Z}_+^{n^k}} H(X \mid Y, G(\varepsilon) = g) \prod_{I \in E_k(V)} \mathbb{P}\{Z_{\varepsilon_I} = g_I\} && (4.8)\end{aligned}$$

where $Z_{\varepsilon_I} \sim \mathcal{P}(\varepsilon_I)$, hence using (4.6),

$$\frac{\partial}{\partial \varepsilon_I} H(X \mid Y, G(\underline{\varepsilon})) = H(X \mid Y, G(\underline{\varepsilon}), Y_I') - H(X \mid Y, G(\underline{\varepsilon}))\,, \tag{4.9}$$

where Y_I' is an extra output drawn independently from $Y(\underline{\varepsilon})$ but conditionally on the same X. Recall also that for any two random variables A and B, $H(A) - H(A \mid B) = H(B) - H(B \mid A)$. We then have

$$\begin{aligned} \frac{\partial}{\partial \varepsilon_I} H(X \mid Y(\underline{\varepsilon})) &= H(Y_I' \mid X, Y, G(\underline{\varepsilon})) - H(Y_I' \mid Y, G(\underline{\varepsilon})) && (4.10)\\ &= H(Y_I' \mid X[I], Y, G(\underline{\varepsilon})) - H(Y_I' \mid Y, G(\underline{\varepsilon})) && (4.11)\end{aligned}$$

where we used the fact that Y_I' depends only on the components of X indexed by I. Finally, recall the dependency among these variables: X and $G(\varepsilon)$ are drawn independently, Y is drawn from $P_{G,Q}(\cdot \mid X)$ and Y_I' is drawn independently from $Q(\cdot \mid X[I])$. Therefore, Y_I' is independent of $(Y, G(\varepsilon))$ conditionally on $X[I]$ and

$$H(Y_I' \mid X[I], Y, G(\underline{\varepsilon})) = H(Y_I' \mid X[I]) = H(Q) \tag{4.12}$$

where $H(Q) = -2^{-k} \sum_{u \in \mathcal{X}^k, z \in \mathcal{Y}} Q(z \mid u) \log Q(z \mid u)$. □

We define a *path* as a differentiable map $t \mapsto \underline{\varepsilon}(t)$, with $t \in [0,T]$ for some $T \geq 0$. We say that a path is balanced if

$$\sum_{I \in E_k(V)} \frac{\mathrm{d}\varepsilon_I}{\mathrm{d}t}(t) = 0\,. \tag{4.13}$$

We define $G(t) = G(\underline{\varepsilon}(t))$. Since $H(Q)$ in Lemma 4.2 is constant, we have the following.

Corollary 4.3. *For a balanced path*

$$\frac{\mathrm{d}}{\mathrm{d}t} H(X \mid Y, G(t)) = -\sum_{I \in E_k(V)} H(Y_I' \mid Y, G(t))\, \frac{\mathrm{d}\varepsilon_I}{\mathrm{d}t}(t)\,. \tag{4.14}$$

Given a partition $V = V_1 \sqcup V_2$, we define the associated *canonical path* $\underline{\varepsilon} : t \in [0,1] \to \underline{\varepsilon}(t) \in [0,1]^{E_k(V)}$ as follows. Let $n_i = |V_i|$, $m_i = |E_k(V_i)|$, $i \in \{1,2\}$, and $m = |E_k(V)|$. We define

$$\varepsilon_I(0) \equiv \frac{\alpha n}{m}, \quad \forall I \in E_k(V), \tag{4.15}$$

$$\varepsilon_I(1) \equiv \begin{cases} \frac{\alpha n_1}{m_1} & \text{if } I \in E_k(V_1), \\ \frac{\alpha n_2}{m_2} & \text{if } I \in E_k(V_2), \\ 0 & \text{otherwise,} \end{cases} \tag{4.16}$$

and

$$\underline{\varepsilon}(t) = (1-t)\underline{\varepsilon}(0) + t\underline{\varepsilon}(1)\,. \tag{4.17}$$

Note that the canonical path is balanced. Moreover, at time $t = 0$, $\mathcal{P}_k(\underline{\varepsilon}(0), n)$ reduces to the original ensemble $\mathcal{P}_k(\alpha, n)$, and at time $t = 1$, $\mathcal{P}_k(\underline{\varepsilon}(1), n)$ reduces to two independent copies of the original ensemble on the subset of n_1 and n_2 variables: $\mathcal{P}_k(\alpha, n_1) \times \mathcal{P}_k(\alpha, n_2)$.

Applying Corollary 4.3, and using the chain rule for the derivative, we obtain the following.

Corollary 4.4. *For the canonical path*

$$\frac{\mathrm{d}}{\mathrm{d}t} H(X \mid Y, G(t)) = \alpha n \mathop{\mathbb{E}}_{I} H(Y_I' \mid Y, G(t)) - \alpha n_1 \mathop{\mathbb{E}}_{I_1} H(Y_{I_1}' \mid Y, G(t)) - \alpha n_2 \mathop{\mathbb{E}}_{I_2} H(Y_{I_2}' \mid Y, G(t))\,, \tag{4.18}$$

where I is uniformly drawn in $E_k(V)$, and I_i are drawn uniformly in $E_k(V_i)$, $i = 1,2$, i. e.,

$$\mathop{\mathbb{E}}_{I} H(Y_I' \mid Y, G(t)) = \sum_{I \in E_k(V)} H(Y_I' \mid Y, G(t)) \frac{1}{|E_k(V)|}\,, \tag{4.19}$$

$$\mathop{\mathbb{E}}_{I_1} H(Y_{I_i}' \mid Y, G(t)) = \sum_{I_i \in E_k(V_i)} H(Y_{I_i}' \mid Y, G(t)) \frac{1}{|E_k(V_i)|}\,, \quad i = 1,2. \tag{4.20}$$

Recall that

$$H(Y_I' \mid Y, G(t)) = -\mathop{\mathbb{E}}_{Y,G(t),Y_I'} \log \sum_x Q(Y_I' \mid x[I]) R_{G(t)}(x \mid Y, G(t)) \tag{4.21}$$

$$= -\mathop{\mathbb{E}}_{Y,G(t),Y_I'} \log \mathop{\mathbb{E}}_{X|Y,G(t)} Q(Y_I' \mid X[I]), \tag{4.22}$$

where Y is the output of $P_{G(t),Q}$.

Lemma 4.5.

$$\frac{1}{\alpha|\mathcal{Y}|}\frac{\mathrm{d}}{\mathrm{d}t} H(X \mid Y, G(t)) = -\sum_{l=2}^{\infty} \frac{1}{l(l-1)} \mathop{\mathbb{E}}_{X^{(1)},\dots,X^{(l)}} \left[n\tilde{\Gamma}_l(V) - n_1\tilde{\Gamma}_l(V_1) - n_2\tilde{\Gamma}_l(V_2)\right] \tag{4.23}$$

where

$$\tilde{\Gamma}_l(V) \equiv \mathop{\mathbb{E}}_{I,W_I} \prod_{r=1}^{l} \Big(1 - Q(W_I \mid X^{(r)}[I])\Big), \tag{4.24}$$

I is uniformly drawn in $E_k(V)$, W_I is uniformly drawn in $\mathcal{Y}$, and $X^{(1)},\dots,X^{(l)}$ are drawn under the probability distribution

$$\mathbb{P}\{X^{(1)} = x^{(1)},\dots,X^{(l)} = x^{(l)}\} = \sum_y \prod_{i=1}^{l} R_{G(t),Q}(x^{(i)} \mid y) \sum_u P_{G(t),Q}(y \mid u) 2^{-n}. \tag{4.25}$$

This means that $X^{(1)},\dots,X^{(l)}$ are drawn i. i. d. from the channel $R_{G(t)}$ given a hidden output Y, these are the "replica" variables, which are exchangeable but not i. i. d.

Proof of Lemma 4.5. By definition

$$H(Y_I' \mid Y, G(t)) = -\mathop{\mathbb{E}}_{Y,G(t),Y_I'} \log \mathop{\mathbb{E}}_{X|Y,G(t)} Q(Y_I' \mid X[I]), \tag{4.26}$$

and expanding the logarithm in its power series (the argument in the log is between 0 and 1),

$$\log \mathop{\mathbb{E}}_{X|Y,G(t)} Q(Y_I' \mid X[I]) = -\sum_{l=1}^{\infty} \frac{1}{l} \Big(\mathop{\mathbb{E}}_{X|Y,G(t)} \big(1 - Q(Y_I' \mid X[I])\big)\Big)^l. \tag{4.27}$$

We now introduce the "replicas" $X^{(1)},\dots,X^{(l)}$ distributed as follows. Recall that X is drawn independently of $G(t)$ and Y is drawn from $P_{G(t),Q}(\cdot \mid X)$. The replicas are drawn independently from the reverse channel $R_{G(t),Q}(\cdot \mid Y)$ defined in (2.2). In other words, we have the Markov chain

$$X \overset{P}{-} Y \overset{R}{-} (X(1),\dots,X(l)). \tag{4.28}$$

Defining the kernel $\widetilde{Q} = 1 - Q$, we can rewrite (4.27) as

$$\log \mathop{\mathbb{E}}_{X|Y,G(t)} Q(Y_I' \mid X[I]) = -\sum_{l=1}^{\infty} \frac{1}{l} \mathop{\mathbb{E}}_{X^{(1)},\dots,X^{(l)}|Y,G(t)} \prod_{r=1}^{l} \widetilde{Q}(Y_I' \mid X^{(r)}[I]). \tag{4.29}$$

Collecting terms we have

$$H(Y_I' \mid Y, G(t)) = \mathop{\mathbb{E}}_{X} \mathop{\mathbb{E}}_{Y,G(t)|X} \mathop{\mathbb{E}}_{Y_I'|X} \sum_{l=1}^{\infty} \frac{1}{l} \mathop{\mathbb{E}}_{X^{(1)},\dots,X^{(l)}|Y,G(t)} \prod_{r=1}^{l} \widetilde{Q}(Y_I' \mid X^{(r)}[I])) \tag{4.30}$$

$$= \sum_{l=1}^{\infty} \frac{1}{l} \mathop{\mathbb{E}}_{X,X^{(1)}\dots,X^{(l)}} \mathop{\mathbb{E}}_{Y_I'|X} \prod_{r=1}^{l} \widetilde{Q}(Y_I' \mid X^{(r)}[I]))\,. \tag{4.31}$$

We now use a manipulation that is specific to the planted framework. In the non-planted framework, super-additivity is achieved by showing that each term weighted by $1/l$ is convex in the empirical distribution of the replicas. In the planted framework, this is no longer true from the above expression. We rely on an additional step which will flip the behaviour from super- to sub-additivity. We switch measure in the expectation $\mathbb{E}_{Y_I'|X}$, defining W_I to be uniformly distributed over $\mathcal{Y}$, and write

$$H(Y_I' \mid Y, G(t)) = |\mathcal{Y}| \sum_{l=1}^{\infty} \frac{1}{l} \mathop{\mathbb{E}}_{X,X^{(1)}\dots,X^{(l)}} \mathop{\mathbb{E}}_{W_I} \prod_{r=1}^{l} \widetilde{Q}(W_I \mid X^{(r)}[I]))Q(W_I \mid X[I]))\,. \tag{4.32}$$

Renaming X by $X^{(0)}$, and using the fact that $X^{(0)}, X^{(1)}\dots, X^{(l)}$ are exchangeable (they are i. i. d. conditioned on Y), we can write

$$H(Y_I' \mid Y, G(t)) = |\mathcal{Y}| \sum_{l=1}^{\infty} \frac{1}{l} \mathop{\mathbb{E}}_{X^{(0)},X^{(1)}\dots,X^{(l)}} \left(\mathop{\mathbb{E}}_{W_I} \prod_{r=1}^{l} \widetilde{Q}(W_I \mid X^{(r)}[I])) - \mathop{\mathbb{E}}_{W_I} \prod_{r=0}^{l} \widetilde{Q}(W_I \mid X^{(r)}[I])) \right) \tag{4.33}$$

$$= |\mathcal{Y}| \mathop{\mathbb{E}}_{X^{(1)}} \mathop{\mathbb{E}}_{W_I} \widetilde{Q}(W_I \mid X^{(1)}[I])) - |\mathcal{Y}| \sum_{l=2}^{\infty} \frac{1}{l(l-1)} \mathop{\mathbb{E}}_{X^{(1)}\dots,X^{(l)}} \mathop{\mathbb{E}}_{W_I} \prod_{r=1}^{l} \widetilde{Q}(W_I \mid X^{(r)}[I]))\,. \tag{4.34}$$

Moreover $\mathbb{E}_{X^{(1)}} \mathbb{E}_{W_I} \widetilde{Q}(W_I \mid X^{(1)}[I]))$ does not depend on n. Recalling that from Corollary 4.4,

$$\frac{\mathrm{d}}{\mathrm{d}t} H(X \mid Y, G(t)) = \alpha n \mathop{\mathbb{E}}_{I} H(Y_I' \mid Y, G(t)) - \alpha n_1 \mathop{\mathbb{E}}_{I_1} H(Y_{I_1}' \mid Y, G(t)) - \alpha n_2 \mathop{\mathbb{E}}_{I_2} H(Y_{I_2}' \mid Y, G(t))\,,$$

we can express each term in the above using (4.34), where the first term in (4.34) is the same in the global or subsystems (since it does not depend on n, n_1 or n_2), hence it cancels out in the above sum as $n = n_1 + n_2$. □

Finally, the following corollary is a change of notation from the previous lemma.

Corollary 4.6.

$$\frac{1}{\alpha|\mathcal{Y}|} \frac{\mathrm{d}}{\mathrm{d}t} H(X \mid Y, G(t)) = -\sum_{l=2}^{\infty} \frac{1}{l(l-1)} \mathop{\mathbb{E}}_{X^{(1)}\dots,X^{(l)}} [n\Gamma_l(\nu_l) - n_1\Gamma_l(\nu_{l,1}) - n_2\Gamma_l(\nu_{l,2})] \tag{4.35}$$

where ν_l (respectively $\nu_{l,i}$) is the empirical distribution of $\{X^{(1)}[i], \dots, X^{(l)}[i]\}_{i\in V}$ over $\mathcal{X}^l$ (respectively of $\{X^{(1)}[i], \dots, X^{(l)}[i]\}_{i\in V_i}$ over $\mathcal{X}^l$, $i = 1, 2$), and $\Gamma_l(\cdot)$ is as defined in (3.7) (for which Hypothesis H requires convexity).

Proof. Recall that $V = [n]$ and V_1 and V_2 are the subsystems of cardinality n_1 and n_2 respectively.

$$\begin{aligned}
\tilde{\Gamma}_l(V) &= \mathop{\mathbb{E}}_{I,W_I} \prod_{r=1}^{l} \Big(1 - Q(W_I \mid X^{(r)}[I])\Big) && (4.36)\\
&= \frac{1}{|\mathcal{Y}|} E_I \left[\sum_{y\in\mathcal{Y}} \prod_{r=1}^{l} (1 - Q(y \mid X^{(r)}[I]))\right] && (4.37)\\
&= \frac{1}{|\mathcal{Y}|} \sum_{u^{(1)},\dots,u^{(l)}\in\mathcal{X}^k} \left[\sum_{y\in\mathcal{Y}} \prod_{r=1}^{l} (1 - Q(y \mid u^{(r)}))\right] \eta(u^{(1)},\dots,u^{(l)}) && (4.38)
\end{aligned}$$

where the last equality holds by defining $\eta(u^{(1)},\dots,u^{(l)})$ as the empirical distribution of

$$\{X^{(1)}[I],\dots,X^{(l)}[I]\}_{I\in[n]^k}$$

over $\mathcal{X}^{kl}$. Moreover, since the index set I is drawn uniformly at random in $[n]^k$,

$$\eta(u^{(1)},\dots,u^{(l)}) = \prod_{i=1}^{k} \nu_l(u_i^{(1)},\dots,u_i^{(l)}), \tag{4.39}$$

where ν_l is the empirical distribution of $\{X^{(1)}[i],\dots,X^{(l)}[i]\}_{i\in[n]}$ over $\mathcal{X}^l$. Hence

$$\begin{aligned}
\tilde{\Gamma}_l(V) &= \frac{1}{|\mathcal{Y}|} \sum_{u^{(1)},\dots,u^{(l)}\in\mathcal{X}^k} \left[\sum_{y\in\mathcal{Y}} \prod_{r=1}^{l} (1 - Q(y \mid u^{(r)}))\right] \prod_{i=1}^{k} \nu_l(u_i^{(1)},\dots,u_i^{(l)}) && (4.40)\\
&= \Gamma_l(\nu_l).
\end{aligned}$$

□

Proof of Theorem 3.2. Hypothesis H ensures that Γ_l is convex for any distribution on $\mathcal{X}^l$, hence in particular for the empirical distribution of the replicas. Moreover, since $V = V_1 \cup V_2$ and $V_1 \cap V_2 = \emptyset$, the empirical distribution over V can be additively computed from the empirical distribution over V_1 and V_2. Formally,

$$n\nu_l = n_1\nu_{l,1} + n_2\nu_{l,2}, \tag{4.41}$$

for any $l \geq 1$, and the convexity of Γ_l together with Corollary 4.6 imply the theorem. □

5 Proof of Theorem 3.4

Proof. Since we know that $H_G(X \mid Y)/n$ converges in expectation, it is sufficient to show that it concentrates around its expectation. Indeed we claim that there exists $B > 0$ such that

$$\mathbb{P}\big\{|H_G(X \mid Y) - H(X \mid Y)| \geq n\Delta\big\} \leq 2\,e^{-nB\Delta^2}, \tag{5.1}$$

whence our thesis follows from Borel-Cantelli.

The proof of equation (5.1) is a direct application of the Azuma-Hoeffding inequality [11]. We condition on the number $m \in [(\alpha-\delta)n, (\alpha+\delta)n]$ of hyperedges in G and prove that

$$\mathbb{P}\Big\{ |H_G(X \mid Y) - H(X \mid Y)| \geq n\Delta \,\Big|\, |E(G)| = m \Big\} \leq 2\, e^{-nB\Delta^2}. \tag{5.2}$$

The claim follows from summing this inequality over $m \in [(\alpha-\delta)n, (\alpha+\delta)n]$ and noting that $|E(G)| \notin [(\alpha-\delta)n, (\alpha+\delta)n]$ with exponentially small probability by standard concentration of binomial random variables.

In order to prove the bound equation (5.2), we regard $H_G(X \mid Y)$ as a function of the choice of the m hyperedges: $(e_1, e_2, \dots, e_m) \mapsto H_G(X \mid Y)$. Since the the hyper-edges are independent, it is sufficient to prove that this function is Lipshitz continuous. Indeed, we claim that $|H_G(X \mid Y) - H_{G'}(X \mid Y)| \leq 2C$, for some constant C, if G and G' differ only in one of their hyperedges,

In order to prove the last claim, let $G' = G + a$ denote the graph G to which hyperedge $a = (i_1, \dots, i_k)$ has been added. Then, writing explicitly the component of Y corresponding to hyperedge a by Y_a, we need to prove that $|H_{G+a}(X \mid Y, Y_a) - H_G(X \mid Y)| \leq C$. We have, dropping the subscripts and superscript for the sake of simplicity,

$$\begin{aligned} 0 \leq H(X \mid Y) - H(X \mid Y, Y_a) &= H(X \mid Y) - H(X, Y_a \mid Y) + H(Y_a \mid Y) && (5.3)\\ &= H(Y_a \mid Y) - H(Y_a \mid X, Y) && (5.4)\\ &\leq \log_2 |\mathcal{Y}|, && (5.5) \end{aligned}$$

where the last inequality follows from the fact that Y_a takes value in the finite set $\mathcal{Y}$. □

6 Applications to specific models

We next present three applications of the general model and results described in the previous section. While planted CSPs and parity-check codes are directly derived as particular cases of our model, the stochastic block model is obtained with a limiting argument. Note that the general model described in the previous section also allows to generate new hybrid structures. For example, one may consider codes which are not linear but which rely on OR gates as in SAT, community structures whose connectivity rely on collections of k nodes, or network models which have censored edges. The latter case was recently considered in [2, 1].

6.1 Planted constraint satisfaction problems

Definition 6.1. A CSP kernel is given by

$$Q(y \mid u) = \frac{1}{|A(u)|} \mathbb{1}(y \in A(u)), \quad u \in \mathcal{X}^k,\, y \in \mathcal{Y}, \tag{6.1}$$

where $A(u)$ is a subset of $\mathcal{Y}$ containing the variables that y can be assigned to, with the property that $|A(u)|$ is constant (it may depend on k but not on u).

We will next show that a graphical channel with a CSP kernel corresponds to a planted CSP. We derive first a few known examples of CSPs.

- For planted k-SAT, $\mathcal{Y} = \{0,1\}^k$ and $A(u) = \{0,1\}^k \setminus u$. Hence, $|A(u)| = 2^k - 1$. With this kernel, the planted assignment x generates for any selected edge $I \in E_k(V)$ an edge variable y_I in $\{0,1\}^k \setminus x[I]$ uniformly drawn. See Section 2 for an interpretation of this model as the usual planted SAT model.
- For planted k-NAE-SAT, $\mathcal{Y} = \{0,1\}^k$ and $A(u) = \{0,1\}^k \setminus \{u, \bar{u}\}$, where $\bar{u}$ is the vector obtained by flipping each component in u. Hence, $|A(u)| = 2^k - 2$.
- For k-XOR-SAT, $\mathcal{Y} = \{0,1\}$ and $A(u) = \oplus_{i=1}^k u_i$, hence $|A(u)| = 1$.

A graphical channel with graph g and kernel Q as in (6.1) leads to a planted CSP where the constraints are given by $A(u[I]) \ni y_I$ for any $I \in E(g)$. For example, for planted k-SAT, the constraints are equivalent to $u[I] \neq y_I$, whereas for planted k-NAE-SAT, the constraints are equivalent to $u[I] \notin (y_I, \bar{y}_I)$. If y is drawn from the output marginal distribution $S_{g,Q}$ (cf. (2.3)), then there exists a satisfying assignment by construction. If X is drawn uniformly at random, then the posterior distribution is also uniformly distributed on its support, and we have the following lemma.

Lemma 6.2. *For a graphical channel with graph g and CSP kernel Q as in* (6.1)*, and for y in the support of $S_{g,Q}$,*

$$H_g(X \mid Y = y) = \log Z_g(y) \tag{6.2}$$

where $Z_g(y)$ is the number of satisfying assignments of the planted CSP with graph g and constraints specified by y (where the structure of constraints is specified by Q).

Proof of Lemma 6.2. We use the notation $x \sim y$ to say that x is a satisfying assignment for the clause specified by y and the kernel Q. We have

$$\begin{aligned} P_{g,Q}(y \mid x) &= \prod_{I \in E(g)} Q(y_I \mid x[I]) && (6.3)\\ &= \prod_{I \in E(g)} \frac{1}{|A|} \mathbb{1}(y_I \in A(x[I])) && (6.4)\\ &= \begin{cases} \frac{1}{|A|^{|E(g)|}} & \text{if } x \sim y, \\ 0 & \text{otherwise.} \end{cases} && (6.5) \end{aligned}$$

Hence for a given x, $P_{g,Q}(\cdot \mid x)$ is uniform on the set of all y's verifying x, which has cardinality $|A|^{|E(g)|}$. Since X is uniform, for a given y, $R_{g,Q}(\cdot \mid y)$ is a uniform measure on a set of cardinality

$$\sum_{x \in \mathcal{X}^n} \prod_{I \in E(g)} \mathbb{1}(y_I \in A(x[I])) = |\{x \in \mathcal{X}^n : y_I \in A(x[I]), \forall I \in E(g)\}| = Z_g(y) \,. \tag{6.6}$$

Therefore $H_g(X \mid Y = y) = \log Z_g(y)$. □

Corollary 6.3. *For a graphical channel with CSP kernel Q as in (6.1), and for a graph G drawn from the ensemble $\mathcal{P}(\alpha,n)$,*

$$H(X \mid Y,G) = \mathop{\mathbb{E}}_{G,Y} \log Z_G(Y)\,, \tag{6.7}$$

where $Z_G(Y)$ is the number of satisfying assignments of the corresponding random planted CSP.

Remark 6.4. Together with Corollary 3.3, this result gives the convergence of the normalized expected logarithm of the number of solutions for *any* edge density α. This is to be put in contrast with [4], which obtains the convergence of the same quantity (where the number of solutions is shifted by 1 to avoid taking the logarithm of 0) only for a regime of α small enough, essentially where the probability of being unsatisfiable decays faster than $1/\log^{1+\varepsilon}(n)$ for some $\varepsilon > 0$. One should note that an unconditional result of the kind of Corollary 6.3 for the non-planted setting would imply the satisfiability conjecture (the existence of an n-independent threshold for k-SAT). Since [32] provides a n-dependent threshold for k-SAT, the convergence of

$$\frac{1}{n}\, \mathbb{E} \log(ZF_k(n,\alpha))$$

to a limit $\phi(\alpha)$ allows to freeze the limit of Friedgut's threshold, using standard analytical arguments (see the proof of Theorem 2, Section 5 in [4]).

We now verify that several planted CSPs satisfy Hypothesis H.

Lemma 6.5. *For any $k \geq 1$, and for the CSP kernel corresponding to planted k-SAT, the operator Γ_l is convex for any $l \geq 1$.*

Proof of Lemma 6.5. For planted k-SAT, $\mathcal{Y} = \mathcal{X}^k = \{0,1\}^k$,

$$Q(z \mid u) = \frac{1}{2^k - 1} \mathbb{1}(z \neq u) \tag{6.8}$$

and

$$\begin{aligned}
\Gamma_l(\nu_l) &= \frac{1}{2^k}\left(\frac{1}{2^k-1}\right)^l \sum_{u^{(1)},\ldots,u^{(l)} \in \mathcal{X}^k} \left[\sum_{z \in \mathcal{Y}} \prod_{r=1}^{l} \mathbb{1}(z = u^{(r)})\right] \prod_{i=1}^{k} \nu_l(u_i^{(1)},\ldots,u_i^{(l)}) && (6.9)\\
&= \frac{1}{2^k}\left(\frac{1}{2^k-1}\right)^l \sum_{u \in \mathcal{X}^k} \prod_{i=1}^{k} \nu_l(u_i,\ldots,u_i) && (6.10)\\
&= \frac{1}{2^k}\left(\frac{1}{2^k-1}\right)^l \sum_{u_1,\ldots,u_k \in \mathcal{X}} \prod_{i=1}^{k} \nu_l(u_i,\ldots,u_i) && (6.11)\\
&= \frac{1}{2^k}\left(\frac{1}{2^k-1}\right)^l \left(\sum_{u_1 \in \mathcal{X}} \nu_l(u_1,\ldots,u_1)\right)^k, && (6.12)
\end{aligned}$$

which is convex in ν_l for any $k,l \geq 1$. $\square$

Lemma 6.6. *For any $k \geq 1$, and for the CSP kernel corresponding to planted k-NAE-SAT, the operator Γ_l is convex for any $l \geq 1$.*

Proof of Lemma 6.6. For planted k-NAE-SAT, $\mathcal{Y} = \mathcal{X}^k = \{0,1\}^k$,

$$Q(z \mid u) = \frac{1}{2^k - 2} \mathbb{1}(z \notin (u, \bar{u})) \tag{6.13}$$

and

$$\Gamma_l(\nu_l) = \frac{1}{2^k}\left(\frac{1}{2^k-2}\right)^l \sum_{u^{(1)},\ldots,u^{(l)} \in \mathcal{X}^k} \left[\sum_{z \in \mathcal{Y}} \prod_{r=1}^{l} \mathbb{1}(u^{(r)} \in (z, \bar{z}))\right] \prod_{i=1}^{k} \nu_l(u_i^{(1)}, \ldots, u_i^{(l)}) \tag{6.14}$$

$$= \frac{1}{2^k}\left(\frac{1}{2^k-2}\right)^l \sum_{b_1,\ldots,b_l \in \mathcal{X}} \sum_{u \in \mathcal{X}^k} \prod_{i=1}^{k} \nu_l(u_i \oplus b_1, \ldots, u_i \oplus b_l) \tag{6.15}$$

$$= \frac{1}{2^k}\left(\frac{1}{2^k-2}\right)^l \sum_{b_1,\ldots,b_l \in \mathcal{X}} \left(\sum_{u_1 \in \mathcal{X}} \nu_l(u_1 \oplus b_1, \ldots, u_1 \oplus b_l)\right)^k, \tag{6.16}$$

which is convex in ν_l for any $k, l \geq 1$. □

Lemma 6.7. *For any k even, and for the CSP kernel corresponding to planted k-XOR-SAT, the operator Γ_l is convex for any $l \geq 1$.*

Proof of Lemma 6.7. This is a special case of Lemma 6.12 for $s = 1$, $d = -1$. □

Note that for k-XOR-SAT, k odd, Γ_l may not be convex. For example for $k = 3$,

$$\Gamma_2(\nu_l) = 1 + \mathcal{F}(\nu_l)^3(1,1)$$

which is not convex.

Using Theorem 3.4 and the previous lemmas, the following is obtained.

Corollary 6.8. *Let $Z(F_n)$ denote the number of solutions of a random planted formula $F_n = (G_n, Y)$ with graph G_n and k-SAT, k-NAE-SAT, or k-XOR-SAT (k even) kernel as in Definition 6.1. Then*

$$\frac{1}{n} \mathop{\mathbb{E}}_{Y} \log Z(F_n)$$

converges almost surely.

Note that the almost sure convergence is over G_n only, and not on the negation patterns Y which are under the expectation. It remains open to show that concentration holds with respect to both G_n and Y. Note also that the limit depends only on k and α, and is not n-dependent.

6.2 Stochastic block model

The problem of community detection is to divide a set of vertices in a network (or graph) into groups having "similar behavior." This may refer to clustering the nodes into subgroups having higher edge density inside the subgroups (assortative case), or across the subgroups (disassortative case). In this paper, we use the terms clustering and community detection for the same purpose. Community detection is a fundamental problem in many modern statistics, machine learning, and data mining problems with a broad range of applications in population genetics, image processing, biology and social science. A large variety of models have been proposed for community detection problems. We refer to [52, 29, 34] for a survey on the subject.

At an algorithmic level, the problem of finding the smallest cut in a graph with two equally sized groups, i. e., the min-bisection problem, is well-known to be NP-hard [26]. Concerning average-case complexity, various random graphs models have been proposed for community detection. The Erdős-Rényi random graph is typically a very bad model for community structures, since each node is equally likely to be connected to any other nodes and no communities are typically formed. The stochastic block model is a natural extension of an Erdős-Rényi model with a community structure. Although the model is fairly simple (communities emerge but the average degree is still constant[7]) it is a fascinating model with several fundamental questions still open.

We now describe the stochastic block model (SBM) with two groups and symmetric parameters, also called the planted bisection model. Let $V = [n]$ be the vertex set and a, b be two positive real numbers. For a uniformly drawn assignment $X \in \{0,1\}^V$ on the vertices, an edge is drawn between vertex i and j with probability a/n if $X_i = X_j$ and with probability b/n if $X_i \neq X_j$, and each edge is drawn independently conditionally on the vertex variables. We denote this model by $\mathcal{G}(n,a,b)$. Note that the average degree of an edge is $(a+b)/2$, however, a 0-labeled node is connected in expectation with $a/2 - a/n$ nodes that have a 0-label and with $b/2$ nodes that have a 1-label.

This model was studied in [15, 21, 26, 56] for the recovery of the clusters in dense graph regimes and in [17, 57] for sparser regimes with logarithmic degree. The threshold for the logarithmic degree regime was recently obtained in [3], with recovery possible if and only if $\sqrt{\alpha} - \sqrt{\beta} \geq \sqrt{2}$, where $\alpha = pn/\log(n)$, $\beta = qn/\log(n)$ and p (respectively, q) are the intra (respectively, extra) edge probability (with $\alpha > \beta > 0$). The attention to the sparse regime described above was initiated in [18] and later in [23, 50]. In particular, [23] conjectured a phase transition phenomenon, with the detection of clusters (i. e., obtaining a reconstruction positively correlated with the true one) possible if $(a-b)^2 > 2(a+b)$ and impossible otherwise. In [50], a proof of the impossibility part is obtained, and in [46, 51] the conjecture was set.

There are at least two ways to define the SBM as a graphical channel. The direct way is simply to consider G to be the complete graphs, and for each pair of vertices, to use the kernel

$$Q(y_{ij} \mid x_i, x_j) = P(y_{ij} \mid x_i \oplus x_j)\,,$$

where $P(1 \mid 0) = a/n$ and $P(1 \mid 1) = b/n$. With this approach, however, the channel Q depends on the number of vertices n, whereas the edge probability of the graph is constant. We next show how we can "move" the sparsity from the kernel to the graph, obtaining a graphical channel with a sparse graph

[7] Models with corrected degrees have been proposed in [39].

as in the previous section and a fixed (asymmetric) channel. The resulting model will approximate accurately the original model as shown next. Note that this creates a strong connection between coding and clustering, since it expresses the latter problem as a particular type of code (a simple degree 2 LDGM code) on a binary input/output channel which is very noisy.

Definition 6.9. For $a,b \in [0,1]$, an SBM(a,b) kernel is given by $\mathcal{X} = \mathcal{Y} = \{0,1\}$ and

$$Q(1 \mid u_1,u_2) = \begin{cases} a & \text{if } u_1 = u_2, \\ b & \text{if } u_1 \neq u_2, \end{cases} \tag{6.17}$$

for all $u_1,u_2 \in \{0,1\}$.

Lemma 6.10. *There exists $n_0 = n_0(\gamma,a,b)$ and $C = C(a,b)$ such that the following holds true. Let X be uniformly drawn on $\{0,1\}^V$, Y be the output (the graph) of a sparse stochastic block model of parameters a,b, and Y_γ be the output of a graphical channel with graph G_γ drawn from the ensemble $\mathcal{P}(\gamma,n)$ and kernel SBM$(a/\gamma,b/\gamma)$ (cf. Definition 6.9), then, for all $n \geq n_0$*

$$\left|H(X \mid Y) - H(X \mid G_\gamma,Y_\gamma)\right| \leq \frac{Cn}{\gamma}\,. \tag{6.18}$$

Lemma 6.11. *For the SBM kernel given by (6.17), $a \leq b$ (disassortative case) and γ large enough, the operator Γ_l is convex for any $l \geq 1$.*

Proof of Lemma 6.11. This is a special case of Lemma 6.12 below, with $s = (a+b)/\gamma$ and $d = (a-b)/\gamma$. If γ is large enough, then $s \leq 1$ and if $a \leq b$, then $d \geq 0$ and all the coefficients in (6.26) are positive. □

We denote by

$$\mathcal{F}(\nu_l)(w) = \sum_{x\in\mathbb{F}_2^l} (-1)^{x\cdot w}\nu_l(x)$$

the Fourier-Walsh transform of ν_l evaluated at $w \in \mathbb{F}_2^l$ (where $x \cdot w$ denotes the dot product of x and w).

Lemma 6.12. *If $\mathcal{X} = \mathcal{Y} = \{0,1\}$,*

$$Q(y \mid x_1,\dots,x_k) = W\big(y \mid \oplus_{i=1}^k X[I]\big)$$

and W is an arbitrary binary input/output channel, then

$$\Gamma_l(\nu_l) = \frac{1}{2}\sum_{w\in\mathbb{F}_2^l} d^{|w|}\left[s^{l-|w|} + (-1)^{|w|}(2-s)^{l-|w|}\right]\mathcal{F}(\nu_l)^k(w) \tag{6.19}$$

where $s = W(1 \mid 0) + W(1 \mid 1)$, $d = W(1 \mid 0) - W(1 \mid 1)$, $|w| = \sum_{i=1}^l w_i$.

Note that $\mathcal{F}(\nu_l)(w)$ is linear in ν_l, hence

- For $s = 1$, i. e., for symmetric channels,

$$\Gamma_l(\nu_l) = \sum_{w\in\mathbb{F}_2^k:\,|w|\text{ even}} d^{|w|}\mathcal{F}(\nu_l)^k(w) \tag{6.20}$$

  and Γ_l is convex when k is even.

- If $s \geq 1$, $d \geq 0$ or $s \leq 1$, $d \leq 0$, then Γ_l is convex when k is even.

Proof of Lemma 6.12. We have

$$\Gamma_l(\nu_l) = \frac{1}{2} \sum_{u^{(1)},\dots,u^{(l)} \in \mathbb{F}_2^k} \left[\sum_{y \in \mathbb{F}_2} \prod_{r=1}^{l} (1 - P(y \mid u^{(r)})) \right] \prod_{i=1}^{k} \nu_l(u_i^{(1)}, \dots, u_i^{(l)}) \tag{6.21}$$

and using the fact that $P(y \mid u^{(r)}) = W(y \mid \oplus_{i=1}^k u_i^{(r)})$,

$$\Gamma_l(\nu_l) = \frac{1}{2} \sum_{v^{(1)},\dots,v^{(l)} \in \mathbb{F}_2} \left[\sum_{y \in \mathbb{F}_2} \prod_{r=1}^{l} (1 - W(y \mid v^{(r)})) \right] \nu_l^{\star k}(v^{(1)}, \dots, v^{(l)}) \tag{6.22}$$

$$= \frac{1}{2} \sum_{v \in \mathbb{F}_2^l} \gamma(v) \nu_l^{\star k}(v) \tag{6.23}$$

where

$$\gamma(v) \equiv \sum_{y \in \mathbb{F}_2} \prod_{r=1}^{l} (1 - W(y \mid v^{(r)})) = (1-a)^{l-|v|}(1-b)^{|v|} + a^{l-|v|} b^{|v|} \tag{6.24}$$

and $a = W(1 \mid 0)$, $b = W(1 \mid 1)$. Note that the Fourier transform of $a^{l-|v|}b^{|v|}$ is given by

$$a^{l-|v|} b^{|v|} \quad \xrightarrow{\mathcal{F}} \quad (a+b)^{l-|w|}(a-b)^{|w|}, \tag{6.25}$$

hence

$$\mathcal{F}(\gamma)(w) = (2-(a+b))^{l-|w|}(a-b)^{|w|}(-1)^{|w|} + (a+b)^{l-|w|}(a-b)^{|w|}. \tag{6.26}$$

□

Proof of (6.25). To show that we have the Fourier pair

$$\mathbb{F}_2^l \ni v \mapsto \rho^{|v|} \quad \xrightarrow{\mathcal{F}} \quad \mathbb{F}_2^l \ni w \mapsto (1+\rho)^{l-|w|}(1-\rho)^{|w|} \tag{6.27}$$

note that the identity is true when $l = 1$ and assume it to be true for l. Then for $l+1$

$$\sum_{v \in \mathbb{F}_2^{l+1}} \rho^{|v|}(-1)^{|vw|} = \sum_{v \in \mathbb{F}_2^l} \rho^{|v|}(-1)^{|vw_1^l|} + \rho^{|v|+1}(-1)^{|vw_1^l|}(-1)^{w_{l+1}} \tag{6.28}$$

$$= \sum_{v \in \mathbb{F}_2^l} \rho^{|v|}(-1)^{|vw_1^l|}(1+\rho(-1)^{w_{l+1}}) \tag{6.29}$$

$$= (1+\rho)^{l-|w_1^l|}(1-\rho)^{|w_1^l|}(1+\rho(-1)^{w_{l+1}}).$$ □

Corollary 6.13. *For the disassortative SBM, the limit of $H(X \mid Y)/n$ exists and satisfies*

$$\lim_{n\to\infty} \frac{1}{n} H(X \mid Y) = \lim_{\gamma\to\infty} \lim_{n\to\infty} \frac{1}{n} H(X \mid G_\gamma, Y_\gamma). \tag{6.30}$$

In a work in progress, the assortative case is investigated with a different proof technique. The regime of a, b where the above limit is strictly below 1 is expected to reflect the phase transition of the SBM [46, 51].

6.3 Parity-check encoded channels

Shannon's coding theorem states that for a discrete memoryless channel W from $\mathcal{X}$ to $\mathcal{Y}$, the largest rate at which reliable communication can take place is given by the capacity $I(W) = \max_X I(X;Y)$, where $I(X;Y)$ is the mutual information of the channel W with a random input X. To show that rates up to capacity are achievable, Shannon used random codebooks, relying on a probabilistic argument. Shortly after, Elias [27] showed[8] that random linear codes allow to achieve capacity, reducing the encoding complexity from exponential to quadratic in the code dimension. However, Berlekamp, McEliece, and Van Tilborg showed in [14] that the maximum likelihood decoding of unstructured linear codes is NP-hard.

In order to reduce the complexity of the decoder, Gallager proposed to the use sparse linear codes [33], giving birth to the LDPC codes, with sparse parity-check matrices, and LDGM codes, with sparse generator matrices. Various types of LDPC/LDGM codes depend on various types of row and column degree distributions. Perhaps the most basic class of such codes is the LDGM code with constant right degree, which corresponds to a generator matrix with column having a fixed number k of ones. This means that each codeword is the XOR of k uniformly selected information bits. In other words, this is a graph based code drawn from an Erdős-Rényi or Poisson ensemble $\mathcal{P}_k(\alpha,n)$. The code can also be seen as a planted k-XOR-SAT formula. The dimension of the code is $m = \alpha n$ and the rate is $r = 1/\alpha$.

Despite the long history of research on the LDPC and LDGM codes, and their success in practical applications of communications, there are still many open questions concerning the behaviour of these codes. In particular, even for the simple code described above, it is still open to show that the mutual information $(1/n)I(X^n;Y^m)$ concentrates, with the exception of the binary erasure channel for which much more is known [44, 45]. In the case of dense random codes, standard probability arguments show that concentration occurs with a transition at capacity for any discrete memoryless channels. But for sparse codes, the traditional arguments fail. Recently, the following was conjectured in [42] for constant right degree LDGM codes G and binary input symmetric output channels,[9]

$$\mathbb{P}_G\left\{\frac{1}{m} I_G(X;Y) < I(W)\right\} \to \begin{cases} 0 & \text{if } \alpha < C_k(W), \\ 1 & \text{if } \alpha > C_k(W), \end{cases} \tag{6.31}$$

where $C_k(W)$ is a constant depending on k and W.

We provide next a concentration result for this model, which implies the above conjecture for even degrees.

Definition 6.14. An encoded symmetric kernel is given by

$$Q(z \mid u) = W(z \mid \oplus_{i=1}^{k} u_i)\,, \tag{6.32}$$

where W is a binary input symmetric output (BISO) channel from $\mathcal{X}$ to $\mathcal{Y}$.

Note that this corresponds to the output of a BISO W when the input to the channel is the XOR of k information bits. This corresponds also to the constant right-degree LDGM codes considered in the conjecture of [42].

[8] The result of Elias is originally for binary-input channels.

[9] This means that the channel output can be obtained by drawing at random a BSC among a finite list of given BSCs and then drawing an output from the selected BSC. In terms of the matrix representing the channel, it has the property that each column is either constant or each column comes in pair with another column where the top and bottom components are interchanged.

Lemma 6.15. *For an encoded symmetric kernel with k even, the operator Γ_l is convex for any $l \geq 1$.*

Proof. We represent the channel W as a $2 \times |\mathcal{Y}|$ stochastic matrix. By definition of BISO channels, this matrix can be decomposed into pairs of columns which are symmetric as

$$\begin{pmatrix} c & d \\ d & c \end{pmatrix} \tag{6.33}$$

with $c, d \geq 0$, or into single columns which have constant values. Let us assume that W contains m such matrices and s such constant columns. We have

$$\begin{aligned}
\Gamma_l(\nu_l) &= \frac{1}{|\mathcal{Y}|} \sum_{u^{(1)},\ldots,u^{(l)} \in \mathbb{F}_2^k} \left[\sum_{y \in \mathcal{Y}} \prod_{r=1}^{l} (1 - P(y \mid u^{(r)})) \right] \prod_{i=1}^{k} \nu_l(u_i^{(1)}, \ldots, u_i^{(l)}) && (6.34) \\
&= \frac{1}{|\mathcal{Y}|} \sum_{v^{(1)},\ldots,v^{(l)} \in \mathbb{F}_2} \left[\sum_{y \in \mathcal{Y}} \prod_{r=1}^{l} (1 - W(y \mid v^{(r)})) \right] \nu_l^{\star k}(v^{(1)}, \ldots, v^{(l)}) && (6.35) \\
&= \frac{1}{|\mathcal{Y}|} \sum_{v \in \mathbb{F}_2^l} g(v) \nu_l^{\star k}(v) && (6.36) \\
&= \frac{1}{|\mathcal{Y}|} \sum_{w \in \mathbb{F}_2^l} \mathcal{F}(g)(w) \mathcal{F}(\nu_l)^k(w) && (6.37)
\end{aligned}$$

where

$$g(v) = \sum_{i=1}^{m} \left(C_i^{l-|v|} D_i^{|v|} + D_i^{l-|v|} C_i^{|v|} \right) + \sum_{i=1}^{s} E_i^l \,, \tag{6.38}$$

for some positive constants C_i, D_i, $i \in [m]$, E_i, $i \in [s]$. Moreover, using (6.25),

$$\begin{aligned}
C^{l-|v|} D^{|v|} + D^{l-|v|} C^{|v|} \quad &\xrightarrow{\mathcal{F}} \quad (C+D)^{l-|w|}(C-D)^{|w|} + (C+D)^{l-|w|}(D-C)^{|w|} && (6.39) \\
&= (C+D)^{l-|w|}(C-D)^{|w|}(1 + (-1)^{|w|}) \,, && (6.40)
\end{aligned}$$

and $\mathcal{F}(g)(w)$ has only positive coefficients since only the terms with $|w|$ even survive. Hence Γ_l is convex when k is even. □

Corollary 6.16. *Let X be uniformly drawn in $\{0,1\}^n$, $U = XG$ be the output of a k-degree LDGM code G of dimension αn, and Y be the output of U on a BISO channel W. (Note that we use G for both the graph and the generator matrix. This abuse of notation is however explained by the fact that the generator matrix is indeed the incidence matrix of the graph G.) Then*

$$\frac{1}{n} I_G(X;Y)$$

converges almost surely to a constant $C_k(\alpha, W)$.

Note that for any realization of G, and dropping the subscript G, we have

$$\frac{1}{m} I(X;Y) = \frac{1}{m} \big(H(Y) - H(Y \mid X)\big),$$

where $H(Y \mid X)$ is the conditional entropy of the forward channel. Since the forward channel factorizes over the m edges, we have $H(Y \mid X) = mH(W)$, where $H(W)$ denotes the conditional entropy of the kernel W. Hence

$$\frac{1}{m} I(X;Y) < 1 - H(W) \equiv \frac{1}{m} H(Y) < 1.$$

Since $(1/m)H(Y)$ converges from previous corollary, and since the limit must be decreasing in α (increasing in r), the conjecture (6.31) follows.

7 Open problems

Some technical requirements resulting from Hypothesis H, such k being even for XORSAT or the disassortativity for the stochastic block model, should be removed. It would also be interesting to compute the value of the limit for various models such as SAT, LDGM codes and the stochastic block model (even though it may not be an explicit formula). Finally, it would be interesting to obtain concentration of the number of solutions for planted CSPs with respect to both the drawing of the random graph and of the clause variables (e. g., the negation variables in SAT).

References

[1] EMMANUEL ABBE, AFONSO S. BANDEIRA, ANNINA BRACHER, AND AMIT SINGER: Decoding binary node labels from censored edge measurements: Phase transition and efficient recovery. *To appear in the IEEE Transactions on Network Science and Engineering*, 2014. [arXiv:1404.4749] 429

[2] EMMANUEL ABBE, AFONSO S. BANDEIRA, ANNINA BRACHER, AND AMIT SINGER: Linear inverse problems on Erdös-Rényi graphs: Information-theoretic limits and efficient recovery. In *IEEE Internat. Symp. on Information Theory (ISIT 2014)*, pp. 1251–1255, 2014. [doi:10.1109/ISIT.2014.6875033] 429

[3] EMMANUEL ABBE, AFONSO S. BANDEIRA, AND GEORGINA HALL: Exact recovery in the stochastic block model. 2014. [arXiv:1405.3267] 433

[4] EMMANUEL ABBE AND ANDREA MONTANARI: On the concentration of the number of solutions of random satisfiability formulas. *Random Structures Algorithms*, 45(3):362–382, 2014. [doi:10.1002/rsa.20501, arXiv:1006.3786v1] 414, 423, 431

[5] DIMITRIS ACHLIOPTAS: Algorithmic barriers from phase transitions in graphs. In *Graph Theoretic Concepts in Computer Science*, volume 6410 of *LNCS*, pp. 1–1. Springer, 2010. Preliminary version in FOCS'08. [doi:10.1007/978-3-642-16926-7_1] 414, 415

[6] DIMITRIS ACHLIOPTAS, HAIXIA JIA, AND CRISTOPHER MOORE: Hiding satisfying assignments: Two are better than one. *J. Artif. Intell. Res. (JAIR)*, 24:623–639, 2005. Preliminary version in AAAI'04. [doi:10.1613/jair.1681] 414

[7] DIMITRIS ACHLIOPTAS, HENRY KAUTZ, AND BART SELMAN CARLA GOMES: Generating satisfiable problem instances. In *Proc. AAAI*, pp. 256–261, 2000. Available at AAAI. 414

[8] DIMITRIS ACHLIOPTAS, JEONG HAN KIM, MICHAEL KRIVELEVICH, AND PRASAD TETALI: Two-coloring random hypergraphs. *Random Structures Algorithms*, 20(2):249–259, 2002. [doi:10.1002/rsa.997] 414

[9] DIMITRIS ACHLIOPTAS, ASSAF NAOR, AND YUVAL PERES: Rigorous Location of Phase Transitions in Hard Optimization Problems. *Nature*, 435(7043):759–764, 2005. [doi:10.1038/nature03602] 414

[10] FABRIZIO ALTARELLI, REMI MONASSON, AND FRANCESCO ZAMPONI: Can rare SAT formulas be easily recognized? On the efficiency of message passing algorithms for k-SAT at large clause-to-variable ratios. 2006. [arXiv:cs/0609101] 414

[11] KAZUOKI AZUMA: Weighted sums of certain dependent random variables. *Tohoku Mathematical Journal*, 19(3):357–367, 1967. [doi:10.2748/tmj/1178243286] 429

[12] WOLFGANG BARTHEL, ALEXANDER K. HARTMANN, MICHELE LEONE, FEDERICO RICCI-TERSENGHI, MARTIN WEIGT, AND RICCARDO ZECCHINA: Hiding solutions in random satisfiability problems: A statistical mechanics approach. *Phys. Rev. Lett.*, 88(18):188701, 2002. [doi:10.1103/PhysRevLett.88.188701] 414

[13] MOHSEN BAYATI, DAVID GAMARNIK, AND PRASAD TETALI: Combinatorial approach to the interpolation method and scaling limits in sparse random graphs. *Ann. Probab.*, 41(6):4080–4115, 2013. Preliminary version in STOC'10. [doi:10.1214/12-AOP816] 414, 423

[14] ELWYN R. BERLEKAMP, ROBERT J. MCELIECE, AND HENK C.A. VAN TILBORG: On the inherent intractability of certain coding problems (corresp.). *IEEE Trans. Inform. Theory*, 24(3):384–386, 1978. [doi:10.1109/TIT.1978.1055873] 436

[15] PETER J. BICKEL AND AIYOU CHEN: A nonparametric view of network models and newman-girvan and other modularities. *Proceedings of the National Academy of Sciences*, 106(50):21068–21073, 2009. [doi:10.1073/pnas.0907096106,] 433

[16] ANDREJ BOGDANOV AND YOUMING QIAO: On the security of Goldreich's one-way function. *Comput. Complexity*, 21(1):83–127, 2012. Preliminary version in RANDOM'09. [doi:10.1007/s00037-011-0034-0] 415

[17] RAVI B. BOPPANA: Eigenvalues and graph bisection: An average-case analysis. In *Proc. 28th FOCS*, pp. 280–285. IEEE Comp. Soc. Press, 1987. [doi:10.1109/SFCS.1987.22] 433

[18] AMIN COJA-OGHLAN: Graph partitioning via adaptive spectral techniques. *Combin. Probab. Comput.*, 19(2):227–284, 2010. [doi:10.1017/S0963548309990514] 433

[19] AMIN COJA-OGHLAN: The asymptotic k-SAT threshold. In *Proc. 46th STOC*, pp. 804–813. ACM Press, 2014. ACM DL. 414

[20] AMIN COJA-OGHLAN, MICHAEL KRIVELEVICH, AND DAN VILENCHIK: Why almost all satisfiable k-CNF formulas are easy. In *Proc. 13th Int. Conf. Analysis of Algorithms (AofA'07)*, pp. 89–102, 2007. Available at DMTCS. 414

[21] ANNE CONDON AND RICHARD M. KARP: Algorithms for graph partitioning on the planted partition model. *Random Structures Algorithms*, 18(2):116–140, 2001. Preliminary version in RANDOM'99. [doi:10.1002/1098-2418(200103)18:2<116::AID-RSA1001>3.0.CO;2-2] 433

[22] THOMAS M. COVER AND JOY A. THOMAS: *Elements of Information Theory*. Wiley Interscience, New York, 1991. 418, 420

[23] AURELIEN DECELLE, FLORENT KRZAKALA, CRISTOPHER MOORE, AND LENKA ZDEBOROVÁ: Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Phys. Rev.*, 84(6):066106, 2011. [doi:10.1103/PhysRevE.84.066106] 415, 433

[24] MARTIN DIETZFELBINGER, ANDREAS GOERDT, MICHAEL MITZENMACHER, ANDREA MONTANARI, RASMUS PAGH, AND MICHAEL RINK: Tight thresholds for cuckoo hashing via XORSAT. In *Proc. 37th Internat. Colloq. on Automata, Languages and Programming (ICALP'10)*, pp. 213–225. Springer, 2010. [doi:10.1007/978-3-642-14165-2_19, arXiv:0912.0287] 414

[25] MARTIN DIETZFELBINGER, ANDREAS GOERDT, MICHAEL MITZENMACHER, ANDREA MONTANARI, RASMUS PAGH, AND MICHAEL RINK: Tight thresholds for cuckoo hashing via XORSAT. 2010. [arXiv:0912.0287v1] 414

[26] MARTIN E. DYER AND ALAN M. FRIEZE: The solution of some random NP-hard problems in polynomial expected time. *J. Algorithms*, 10(4):451–489, 1989. Preliminary version in FOCS'86. [doi:10.1016/0196-6774(89)90001-1] 433

[27] PETER ELIAS: Coding for noisy channels. *IRE Convention Record*, 4:37–46, 1955. 436

[28] URIEL FEIGE, ELCHANAN MOSSEL, AND DAN VILENCHIK: Complete convergence of message passing algorithms for some satisfiability problems. *Theory of Computing*, 9(19):617–651, 2013. Preliminary version in RANDOM'06. [doi:10.4086/toc.2013.v009a019] 414

[29] SANTO FORTUNATO: Community detection in graphs. *Physics Reports*, 486(3-5):75–174, 2010. [doi:10.1016/j.physrep.2009.11.002, arXiv:0906.0612] 433

[30] SILVIO FRANZ AND MICHELE LEONE: Replica bounds for optimization problems and diluted spin systems. *J. Stat. Phys.*, 111(3-4):535, 2003. [doi:10.1023/A:1022885828956] 414, 423

[31] SILVIO FRANZ, MICHELE LEONE, AND FABIO LUCIO TONINELLI: Replica bounds for diluted non-Poissonian spin systems. *J. Phys. A*, 36(43):10967, 2003. [doi:10.1088/0305-4470/36/43/021] 414, 423

[32] EHUD FRIEDGUT: Sharp thresholds of graph properties, and the k-sat problem. *J. Amer. Math. Soc.*, 12:1017–1054, 1999. Appendix by Jean Bourgain. [doi:10.1090/S0894-0347-99-00305-7] 414, 431

[33] ROBERT G. GALLAGER: *Low-Density Parity-Check Codes*. MIT Press, Cambridge, Massachussetts, 1963. 436

[34] ANNA GOLDENBERG, ALICE X. ZHENG, STEPHEN E. FIENBERG, AND EDOARDO M. AIROLDI: A survey of statistical network models. *Foundations and Trends in Machine Learning*, 2(2):129–233, 2010. [doi:10.1561/2200000005] 415, 433

[35] ODED GOLDREICH: In *Candidate one-way functions based on expander graphs*, volume 6650 of *LNCS*, pp. 76–87. Springer, 2011. [doi:10.1007/978-3-642-22670-0_10] 415

[36] FRANCESCO GUERRA AND FABIO LUCIO TONINELLI: The thermodynamic limit in mean field spin glasses. *Commun. Math. Phys.*, 230(1):71–79, 2002. [doi:10.1007/s00220-002-0699-y] 414, 423

[37] HARRI HAANPÄÄ, MATTI JÄRVISALO, PETTERI KASKI, AND ILKKA NIEMELÄ: Hard satisfiable clause sets for benchmarking equivalence reasoning techniques. *Journal on Satisfiability, Boolean Modeling and Computation*, 2(1-4):27–46, 2005. Available at JSAT. 414

[38] HAIXIA JIA, CRISTOPHER MOORE, AND DOUG STRAIN: Generating hard satisfiable formulas by hiding solutions deceptively. *J. Artif. Intell. Res. (JAIR)*, 28:107–118, 2007. Preliminary version in AAAI'05. [doi:10.1613/jair.2039, arXiv:cs/0503044] 414

[39] BRIAN KARRER AND MARK E.J. NEWMAN: Stochastic blockmodels and community structure in networks. *Phys. Rev. E*, 83(1):016107, 2011. [doi:10.1103/PhysRevE.83.016107] 433

[40] FLORENT KRZAKALA AND LENKA ZDEBOROVÁ: Hiding quiet solutions in random constraint satisfaction problems. *Phys. Rev. Lett.*, 102(23):238701, 2009. [doi:10.1103/PhysRevLett.102.238701, arXiv:0901.2130] 415

[41] SHRINIVAS KUDEKAR AND NICOLAS MACRIS: Sharp bounds for optimal decoding of Low-Density Parity-Check codes. *IEEE Trans. Inform. Theory*, 55(10):4635–4650, 2009. [doi:10.1109/TIT.2009.2027523, arXiv:0807.3065] 423

[42] RAJ KUMAR, KRISHNA KUMAR, PAYAM PAKZAD, AMIR HESAM SALAVATI, AND MOHAMMAD AMIN SHOKROLLAHI: Phase transitions for mutual information. In *Turbo Codes and Iterative Information Processing (ISTC), 2010 6th International Symposium on*, pp. 137–141, 2010. 415, 422, 436

[43] JOHN D. LAFFERTY, ANDREW MCCALLUM, AND FERNANDO C. N. PEREIRA: Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proc. 18th Int. Conf. on Machine Learning (ICML'01)*, pp. 282–289, 2001. ACM DL. 415

[44] MICHAEL LUBY, MICHAEL MITZENMACHER, MOHAMMAD AMIN SHOKROLLAHI, AND DANIEL A. SPIELMAN: Efficient erasure correcting codes. *IEEE Trans. Inform. Theory*, 47(2):569–584, 2001. [doi:10.1109/18.910575] 436

[45] MICHAEL LUBY, MICHAEL MITZENMACHER, MOHAMMAD AMIN SHOKROLLAHI, DANIEL A. SPIELMAN, AND VOLKER STEMANN: Practical loss-resilient codes. In *Proc. 29th STOC*, pp. 150–159. ACM Press, 1997. [doi:10.1145/258533.258573] 436

[46] LAURENT MASSOULIÉ: Community detection thresholds and the weak Ramanujan property. In *Proc. 46th STOC*, pp. 694–703. ACM Press, 2014. ACM DL. [arXiv:1311.3085] 420, 433, 435

[47] ANDREA MONTANARI: Tight bounds for LDPC and LDGM codes under MAP decoding. *IEEE Trans. Inform. Theory*, 51(9):3221–3246, 2005. [doi:10.1109/TIT.2005.853320, arXiv:cs.IT/0407060] 423

[48] ANDREA MONTANARI: Estimating random variables from random sparse observations. *European Transactions on Telecommunications*, 19(4):385–403, 2008. ACM DL. [arXiv:0709.0145] 415

[49] ANDREA MONTANARI, RICARDO RESTREPO, AND PRASAD TETALI: Reconstruction and Clustering in Random Constraint Satisfaction Problems. *SIAM J. Discrete Math.*, 25(2):771–808, 2011. [doi:10.1137/090755862, arXiv:0904.2751] 414

[50] ELCHANAN MOSSEL, JOE NEEMAN, AND ALLAN SLY: Stochastic Block Models and Reconstruction. *Prob. Theor. Rel. Fields*, 2012. To appear. [arXiv:1202.1499] 433

[51] ELCHANAN MOSSEL, JOE NEEMAN, AND ALLAN SLY: A proof of the block model threshold conjecture. 2014. [arXiv:1311.4115] 420, 433, 435

[52] MARK E. J. NEWMAN: Communities, modules and large-scale structure in networks. *Nature Physics*, 8(1):25–31, 2011. [doi:10.1038/nphys2162] 433

[53] DMITRY PANCHENKO AND MICHEL TALAGRAND: Bounds for diluted mean-field spin glass models. *Prob. Theor. Rel. Fields*, 130(3):319–336, 2004. [doi:10.1007/s00440-004-0342-2] 414, 423

[54] GEORGE POLYA AND GABOR SZEGÖ: *Problems and Theorems in Analysis I*. Springer, Berlin, 1998. 421

[55] TOM RICHARDSON AND RÜDIGER URBANKE: *Modern Coding Theory*. Cambridge Univ. Press, Cambridge, 2008. 415

[56] TOM A.B. SNIJDERS AND KRZYSZTOF NOWICKI: Estimation and Prediction for Stochastic Blockmodels for Graphs with Latent Block Structure. *Journal of Classification*, 14(1):75–100, 1997. [doi:10.1007/s003579900004] 433

[57] ENDRE SZEMERÉDI: Regular partitions of graphs. *Colloq. Internat. CNRS: Problèmes combinatoires et théorie des graphes*, 260:399–401, 1978. 433

[58] LENKA ZDEBOROVÁ AND FLORENT KRZAKALA: Quiet planting in the locked constraint satisfaction problems. *SIAM J. Discrete Math.*, 25(2):750–770, 2011. [doi:10.1137/090750755, arXiv:0902.4185] 415

AUTHORS

Emmanuel Abbe
Assistant professor
Princeton University, Princeton, NJ
eabbe@princeton.edu
http://www.princeton.edu/~eabbe

Andrea Montanari
Associate professor
Stanford University, Stanford, CA
montanari@stanford.edu
http://web.stanford.edu/~montanar

ABOUT THE AUTHORS

EMMANUEL ABBE received his Ph. D. from the EECS department at M.I.T., under the supervision of Lizhong Zheng, and his M. Sc. degree from the Mathematics Department at EPFL. He is currently an assistant professor in the Department of Electrical Engineering and in the Program for Applied and Computational Mathematics at Princeton University. His research interests are in coding theory, random graphs, and in the interplay between those fields.

ANDREA MONTANARI is an associate professor in the Departments of Electrical Engineering and of Statistics, Stanford University. He received the Laurea degree in physics in 1997, and the Ph. D. in theoretical physics in 2001, both from Scuola Normale Superiore, Pisa, Italy. He has been a Postdoctoral Fellow with the Laboratoire de Physique Théorique of Ecole Normale Supérieure (LPTENS), Paris, France, and the Mathematical Sciences Research Institute, Berkeley, CA. From 2002 to 2010 has been Chargé de Recherche at LPTENS. In September 2006, he joined the faculty of Stanford University. Dr. Montanari was co-awarded the ACM SIGMETRICS Best Paper Award in 2008. He received the CNRS Bronze Medal for Theoretical Physics in 2006, the National Science Foundation CAREER award in 2008, the Okawa Foundation Research Grant in 2013, and the Best Publication Award of the Applied Probability Society in 2015. His research focuses on algorithms on graphs, graphical models, statistical inference and estimation.