PROPAGATION OF CHAOS FOR NON-EXCHANGEABLE DIFFUSIONS

DANIEL LACKER, LANE CHUN YEUNG, AND FUZHONG ZHOU

Contents

1. Introduction

Suppose a large number n of particles are initialized at i.i.d. positions and then evolve according to some dynamics. The dynamics of a single particle will consist of a base motion plus pairwise interaction forces exerted by all of the other particles. These interactions immediately correlate the particles' positions. The question we study in this work is: How strongly, or how quickly, does this correlation (or coupling) occur? If the dynamics are invariant with respect to permutations of the particle labels, then the particles are exchangeable, and the problem just posed is known as the *propagation of chaos* for mean field dynamics. The broader context for our work, discussed in detail below, is a recent literature on non-exchangeable extensions of this mean field paradigm. As we will see, the non-exchangeable setting exhibits far richer correlation structures which display an intriguing dependence on the matrix of interaction strengths.

Our concrete setup is as follows, simplified somewhat for the introduction. The n particles X^i indexed by $i \in [n] = \{1, \ldots, n\}$ take values in $\mathbb{R}^d$ and evolve according to

$$dX_t^i = \Big(b_0(X_t^i) + \sum_{j=1}^n \xi_{ij} b(X_t^i, X_t^j)\Big)dt + \sigma dB_t^i, \qquad X_0^i \ \text{ iid} \sim P_0. \tag{1.1}$$

Here B^i are independent Brownian motions, $\sigma > 0$ is constant, and b_0 and b are "self-interaction" and "interaction" functions, respectively, which are assumed for now to be Lipschitz, solely for the simplicity of this introduction. The key feature is the $n \times n$ "interaction matrix" ξ, with nonnegative entries ξ_{ij} representing the influence of particle j on i. We assume zero diagonal entries, $\xi_{ii} = 0$, to separate the self-interaction term b_0.

D.L. is partially supported by the NSF CAREER award DMS-2045328.

1.1. **The exchangeable (mean field) case.** When $\xi_{ij} = 1/(n-1)$ for all $i \neq j$ (or $1/n$), we are in a classical setting of interacting diffusions of mean field type, and there is a well-established sense in which the particles are approximately i.i.d. This makes use of a limiting distribution Q_t, defined by the so-called McKean-Vlasov equation,

$$dY_t = \left(b_0(Y_t) + \int_{\mathbb{R}^d} b(Y_t, y)\, Q_t(dy)\right) dt + \sigma dB_t, \qquad Q_t = \mathrm{Law}(Y_t), \quad Y_0 \sim P_0. \tag{1.2}$$

Precisely, for $t > 0$ and k fixed, the joint law P_t^k of $(X_t^1, \ldots, X_t^k)$ converges weakly to the product measure $Q_t^{\otimes k}$ as $n \to \infty$. The (unnormalized) relative entropy $H_t^k := H(P_t^k \,|\, Q_t^{\otimes k})$ is often used to quantify this convergence and can be thought of as a measure of how strongly correlated k-particle collections are. To prove $H_t^k \to 0$, a popular and tractable method [38, 39] is to estimate the "global" entropy H_t^n, which can often be estimated to be $O(1)$, and which then implies $H_t^k = O(k/n)$ for each $k \leq n$ thanks to a well known subadditivity inequality. We will build on a recent work of the first author [46], which showed that in fact $H_t^k = O((k/n)^2)$ is the optimal order, with a matching lower bound in the Gaussian case where b_0 and b are linear. Given that the global estimate $H_t^n = O(1)$ cannot be improved, this reveals a surprising failure of the subadditivity inequality to capture the correct behavior of H_t^k. The proof in [46] was accomplished by using the so-called *BBGKY hierarchy* to derive a hierarchy of differential inequalities,

$$\frac{d}{dt} H_t^k \leq c_1 \frac{k^2}{n^2} + c_2 k (H_t^{k+1} - H_t^k), \quad k = 1, \ldots, n-1, \qquad \frac{d}{dt} H_t^n \leq c_1. \tag{1.3}$$

where c_1 and c_2 are constants independent of n, k, and t. Further literature review is delayed to Section 1.6.

1.2. **The non-exchangeable setting.** In this work, we adapt the hierarchical approach of [46] to the non-exchangeable setting.

A first challenge of non-exchangeability is the lack of an obvious choice of a reference measure, to replace the Q_t arising from the McKean-Vlasov equation. One way to choose a reference measure would be to identify an alternative large-n limit, to replace the McKean-Vlasov equation. This has been done under asymptotic structural assumptions on ξ, namely that it admits a suitable graphon limit, and we review some of this literature in Section 1.6 below. We instead adopt a non-asymptotic perspective, by working with a particular choice of reference measure termed the *independent projection* in [47], described by the following SDE system:

$$dY_t^i = \left(b_0(Y_t^i) + \sum_{j=1}^n \xi_{ij} \int_{\mathbb{R}^d} b(Y_t^i, y)\, Q_t^j(dy)\right) dt + \sigma dB_t^i, \quad Q_t^i = \mathrm{Law}(Y_t^i), \;\; Y_0^i \text{ iid} \sim P_0. \tag{1.4}$$

The system (1.4) appeared also in [37], in the form of n coupled Fokker-Planck equations, as a means of disentangling two problems which can be seen as separate. The first problem is to approximate the interacting n-particle system one in which the particles are *independent*, with the system (1.4) being a natural choice. Indeed, the paper [47] explains certain senses in which (1.4) can be considered a canonical way to approximate (1.1) in distribution by independent particles. The second problem is exploit this independence to facilitate the analysis of a large-n limit. In this work, we focus entirely on the first of these two problems, because unlike the second it can be studied non-asymptotically, without any need to identify a large-n limit for ξ. Moreover, there is an important class of examples for which the second problem is vacuous, which is when the matrix ξ is *stochastic*:

$$\sum_{j=1}^n \xi_{ij} = 1, \quad \text{for all } i = 1, \ldots, n. \tag{1.5}$$

In this case, the independent projection reduces to the usual McKean-Vlasov equation; that is, $Q_t^i = Q_t$ for all i, where Q_t is given by (1.2). An important special case of (1.5) is the following:

Definition 1.1. We say "the random walk case" to refer to the situation in which we are given a graph with vertex set $[n]$ and no isolated vertices, and we set $\xi_{ij} = 1/\deg(i)$ when (i,j) is an edge and $\xi_{ij} = 0$ when (i,j) is not an edge. Here $\deg(i)$ denotes the degree of vertex i.

In other words, in the random walk case, ξ is the transition matrix of the simple random walk on the graph. Each particle interacts with the average of its neighbors in the graph. Note that when the graph is the complete graph we recover the usual mean field case. A notable sub-case is the following:

Definition 1.2. We say "the m-regular graph case" to refer to the random walk case with the further restriction that the graph is m-regular, i.e., $\deg(i) = m$ is the same for all i.

A second challenge of non-exchangeability is that different choices of k out of the n particles can have different joint laws. For a set of indices $v \subset [n]$, let us write P^v_t for the law of $(X^i_t)_{i\in v}$, and similarly Q^v_t for the law of $(Y^i_t)_{i\in v}$. Let $H_t(v) := H(P^v_t \,|\, Q^v_t)$ denote the corresponding entropy.

1.3. **Summary of our results.** Our main results are bounds on the entropies $H_t(v)$, many of which are sharp, which quantify the degree of approximate independence of the subcollection of particles $v \subset [n]$. An important standing assumption throughout the paper is that the row sums of ξ are bounded:

$$\sum_{j=1}^n \xi_{ij} \le 1, \quad \text{for all } i = 1, \ldots, n. \tag{1.6}$$

The constant 1 here is arbitrary, as any other constant could be absorbed into b. The following parameters are relevant:

$$\delta := \max_{i,j\in[n]} \xi_{ij}, \qquad \delta_i := \max_{j\in[n]} \xi_{ij}. \tag{1.7}$$

(1) **Maximum entropy:** In Theorem 2.8 we estimate the maximum entropy over all k-particle configurations, for each $k \in [n]$:

$$\widehat{H}^k_t = \max_{v\subset[n],\ |v|=k} H_t(v) \lesssim (\delta k + 1)(\delta k)^2, \tag{1.8}$$

where the constant hidden in $\lesssim$ does not depend on k, n, or ξ subject to (1.6). The details of (b_0, b) enter into this hidden constant. If $k = O(1/\delta)$, a Gaussian example (Remark 2.19) shows that this is sharp in the sense that a matching lower bound of order $(\delta k)^2$ can be obtained, for instance in the regular graph case when there exists a clique of size of order k.

This shows that the maximum entropy is controlled by the maximal entry δ of ξ. If we are given a sequence of models with $n \to \infty$, we cannot expect $\widehat{H}^k_t$ to vanish if δ does not, because a pair of particles with interaction strength ξ_{ij} bounded away from zero should not become asymptotically independent.

(2) **Average entropy:** In Theorem 2.9 we estimate the average entropy over all k-particle configurations, for each $k \in [n]$:

$$\overline{H}^k_t = \frac{1}{\binom{n}{k}} \sum_{v\subset[n],\ |v|=k} H_t(v) \lesssim (\delta k + 1)\left(\frac{1}{n}\sum_{i=1}^n (\delta_i k)^4\right)^{1/2}. \tag{1.9}$$

The right-hand side of (1.9) is obviously bounded by that of (1.8). Notably, $\overline{H}^k_t$ can be small even if some of the rows of ξ have large maximal entry, i.e., even if some particles have strong (non-vanishing) connection strengths. In a sequence of models with $n \to \infty$, it can certainly be the case that $\frac{1}{n}\sum_{i=1}^n \delta_i^4$ vanishes while $\delta = \max_{i\in[n]} \delta_i$ does not.

The bound (1.9) is practical in many cases but not sharp. If ξ is symmetric, we obtain in Theorem 2.10 the sharper bound

$$\overline{H}_t^k \lesssim (\delta k+1)\bigg(\frac{k^2}{n^2}\sum_{i,j=1}^n \xi_{ij}^2 + \frac{k}{n}\sum_{i=1}^n\Big(\sum_{j=1}^n \xi_{ij}^2\Big)^2 + \frac{k}{n}\sum_{i,j=1}^n \xi_{ij}^3\bigg). \tag{1.10}$$

In the m-regular graph case, these two bounds (1.9) and (1.10) respectively specialize to $(k/m+1)(k/m)^2$ and $(k/m+1)(k^2/nm+k/m^2)$. The latter can vanish even if k is not $o(m)$. More significant improvements of (1.10) over (1.9) arise when the entries along each row of ξ are heterogeneous. If $k=O(1/\delta)$, we show in Remark 2.16 that the estimate of (1.10) is sharp for a broad class of ξ with not-too-heterogeneous rows, including the regular graph case. This is a consequence of a completely sharp calculation in a Gaussian example, in Theorem 2.15:

$$\overline{H}_t^k \asymp \frac{k(k-1)}{n(n-1)}\sum_{i=1}^n \xi_{ij}^2 + \frac{k(n-k)}{n(n-1)}\sum_{i=1}^n\bigg(\sum_{j=1}^n \xi_{ij}^2\bigg)^2. \tag{1.11}$$

Here $\asymp$ means both $\lesssim$ and $\gtrsim$. We conjecture that the cubic term in (1.10) can be removed.

The above estimates, especially (1.11), reveal a dramatic failure of the *subadditivity inequality* to capture the correct behavior of the average entropy. Indeed, the famous subadditivity inequality of entropy states that

$$\overline{H}_t^k \le (k/n)\overline{H}_t^n, \qquad 1\le k\le n.$$

See [26, Theorem 1] for this level of generality, where exchangeability is not assumed. Applying (1.11) with $k=n$ shows that $\overline{H}_T^n \asymp \sum_{ij}\xi_{ij}^2$. Using subaddivity yields merely $\overline{H}_T^k \lesssim (k/n)\sum_{ij}\xi_{ij}^2$, which completely misses the shape of (1.11).

To elaborate on these bounds, imagine that we are in an asymptotic regime, given a sequence of ξ of size $n\times n$ with $n\to\infty$, though we suppress the dependence of ξ on n. Asymptotic notation like $k=o(n)$ should be interpreted accordingly. The number $k\le n$ of particles is in general allowed to grow with n, except when stated otherwise; this is in the spirit of what is sometimes called the *the size of chaos* or *increasing propagation of chaos* [7].

In the m-regular graph case of Definition 1.2, we have $\delta=\delta_i=1/m$. We deduce propagation of chaos in the sense that $\widehat{H}_t^k\to 0$ as long as $m\to\infty$ and $k=o(m)$. This is means that the sequence of graphs is "dense" in a very mild sense, as there is no restriction on how quickly m diverges. The sparse regime in which m stays bounded is very different, and $\overline{H}_t^k$ will no longer vanish, essentially because the McKean-Vlasov equation is no longer the correct limiting model.

In the random walk case of Definition 1.1, there is an interesting dichotomy of denseness thresholds. Note that $\delta=1/\min_i\deg(i)$ and $\delta_i=1/\deg(i)$, or $\delta_i=0$ if $\deg(i)=0$. For the maximum entropy $\widehat{H}_t^k$ to vanish, the minimal degree $\min_i\deg(i)$ must diverge. This holds almost surely in an Erdös-Rényi random graph $G(n,p)$ even if $p=p_n$ is allowed to vanish, as long as $\liminf np/\log n>1$. For the average entropy $\widehat{H}_t^k$ to vanish, we need only that the "typical degree" diverges in the sense that $(1/n)\sum_i(1\vee\deg(i))^{-4}\to 0$. This is equivalent to saying that $\deg(U)\to\infty$ in probability, where U is a uniformly random vertex. In the Erdös-Rényi graph this is true as long as $np\to\infty$ at any speed. This dichotomy appeared in a recent study [51, Section 2.3.1] of stochastic games on networks.

We lastly mention some setwise estimates, without averages or maxima:

(3) **Setwise entropy:** In Theorem 2.13 we bound the entropy for each configuration $v\subset[n]$:

$$H_t(v) \lesssim (\delta|v|+1)\bigg(\sum_{i,j\in v}\xi_{ij}^2 + \delta\sum_{i,j\in v}(\xi^\top\xi+\xi\xi^\top)_{ij} + \delta^2|v|\bigg). \tag{1.12}$$

Using (1.6), the right-hand side of (1.12) can be bounded by that of (1.8). In the m-regular graph case, this simplifies to

$$H_t(v) \lesssim \left(\frac{|v|}{m} + 1\right)\left(\frac{1}{m^2}p_1(v) + \frac{1}{m^3}p_2(v)\right), \tag{1.13}$$

where we define $p_\ell(v)$ to be the number of paths of length ℓ that start and end in v. The first term $p_1(v)$ can range from 0 to $|v|(|v| \wedge m)$, while the second term $p_2(v)$ can range from $|v|m$ to $|v|^2m$. The smallest values are obtained when v is "disconnected," in the sense that its vertices are nonadjacent and have no common neighbors.

1.4. **From the BBGKY hierarchy to first-passage percolation.** Our analysis proceeds through a natural variant of the BBGKY hierarchy for the non-exchangeable setting. Adapting the methods of [46], we derive the following differential inequalities analogous to (1.3):

$$\frac{d}{dt}H_t(v) \le C(v) + \sum_{j \notin v} \mathcal{A}_{v\to j}(H_t(v \cup \{j\}) - H_t(v)), \quad v \subset [n], \tag{1.14}$$

where for certain constants c_1 and c_2 we define

$$C(v) := c_1 \sum_{i\in v}\Big(\sum_{j\in v}\xi_{ij}\Big)^2, \qquad \mathcal{A}_{v\to j} := c_2 \sum_{i\in v}\xi_{ij}.$$

This hierarchy is indexed by *subsets* rather than *elements* of $[n]$, thanks to non-exchangeability, making it significantly more complex to analyze than (1.3). The key idea driving our analysis is to recognize in (1.14) the infinitesimal generator of a continuous-time Markov process $(\mathcal{X}_t)_{t\ge0}$ taking values in the space of sets, $2^{[n]}$. This process, which we call *the infection process* for reasons explained below, is defined as follows: At each jump time, a single number from $[n]$ is added to $\mathcal{X}_t$, and numbers are never removed. The transition rate from v to $v \cup \{j\}$ is $\mathcal{A}_{v\to j}$, whenever $v \subset [n]$ and $j \notin v$. In other words, the rate matrix (or infinitesimal generator) of $\mathcal{X}$ is given for $u, v \in 2^{[n]}$ by

$$\mathcal{A}(v,u) = \begin{cases} \mathcal{A}_{v\to j} & \text{if } u = v \cup \{j\}, \text{ for some } j \in [n] \setminus v \\ -\sum_{j\notin v}\mathcal{A}_{v\to j} & \text{if } u = v \\ 0 & \text{otherwise.}\end{cases}$$

This is a "rate matrix" in the sense that $\sum_v \mathcal{A}(u,v) = 0$ for each u, and the off-diagonal entries $v \neq u$ are nonnegative. We view $\mathcal{A}$ as an operator acting on functions $F : 2^{[n]} \to \mathbb{R}$, defined naturally by $(\mathcal{A}F)(u) = \sum_v \mathcal{A}(u,v)F(v)$. Let $\mathbb{E}_v[\cdot]$ denote expectation under the initialization $\mathcal{X}_0 = v$. We have the identity

$$\mathbb{E}_v[F(\mathcal{X}_t)] = e^{t\mathcal{A}}F(v), \quad v \subset [n], \quad t > 0.$$

In particular, this formula emphasizes the crucial fact that the semigroup $e^{t\mathcal{A}}$ is monotone with respect to pointwise inequality. With this notation, we may write (1.14) as a pointwise inequality between functions on $2^{[n]}$:

$$\frac{d}{dt}H_t \le C + \mathcal{A}H_t.$$

This implies

$$\frac{d}{dt}\Big(e^{t\mathcal{A}}H_{T-t}\Big) = e^{t\mathcal{A}}\Big(\mathcal{A}H_{T-t} + \frac{d}{dt}H_{T-t}\Big) \ge -e^{t\mathcal{A}}C,$$

for any $T > t > 0$. We deduce for $v \subset [n]$ that

$$\begin{aligned} H_T(v) &\le e^{T\mathcal{A}} H_0(v) + \int_0^T e^{t\mathcal{A}} C(v)\,dt \\ &= \mathbb{E}_v[H_0(\mathcal{X}_T)] + \int_0^T \mathbb{E}_v[C(\mathcal{X}_t)]\,dt, \quad v \subset [n]. \end{aligned} \tag{1.15}$$

This is essentially (an inequality form of) a Feynman-Kac formula for the Markov process $\mathcal{X}$. Note that previously in this introduction we have assumed the initial laws agree, $P_0 = Q_0$, so that $H_0(\cdot) \equiv 0$ in the above inequality, but this can (and will) be easily generalized.

We call the process $\mathcal{X}$ the *infection process* in light of its connection with the Susceptible-Infected (SI) model of epidemiology. Given an initial set $\mathcal{X}_0 = v$ of infected nodes, an uninfected (susceptible) node j then infected at rate $\mathcal{A}_{v\to j}$. The infected set grows but never shrinks (i.e., there is no "recovery" as in the more common SIR model), and $[n]$ is an absorbing state.

The simplest case to understand is when ξ is a scalar multiple of the adjacency matrix of a graph G on vertex set $[n]$. When that the infection currently consists of $\mathcal{X}_0 = v$, one then adds a vertex $j \notin v$ to the infection at rate proportional to the number of neighbors of j in v. In this setting the SI model goes by other names. Its discrete-time counterpart is called the *Eden model* [29], and it is equivalent to *first-passage percolation* (FPP), in the following sense. FPP on a graph G is normally described by equipping each edge e with iid exponential random variables τ_e of rate 1. Any path in the graph is then assigned a weight, determined by summing τ_e over those edges e belonging to the path. The distance between two vertices (i, j) is defined to be the minimal weight over all paths connecting i to j. In this manner, the iid exponentials turn the vertex set of G into a random metric space. Given an initial set $v \subset [n]$, use this random metric to define $\mathcal{B}_t$ as the set of points of distance at most t from v. Then the process $\mathcal{B}_\cdot$ has the same distribution as our process $\mathcal{X}_\cdot$ initialized from $\mathcal{X}_0 = v$. This follows from the memoryless property of the exponential distribution, as was first observed by Richardson [64].

The inequality (1.15) indicates that $\mathbb{E}_v[C(\mathcal{X}_t)]$ is the key quantity that we need to understand. The quantity $C(v)$ measures how strongly the set v is connected internally. In contrast, the transition rates $\mathcal{A}_{v\to j}$ are governed by the strength of connections from v to its complement. Existing results on FPP do not tell us much about the quantity $\mathbb{E}_v[C(\mathcal{X}_t)]$, even in the nicest regular graph settings. The canonical setting of FPP is the lattice $\mathbb{Z}^2$, or more generally $\mathbb{Z}^d$, rather than general graphs. The primary questions of interest pertain to the "limit shape" of $t^{-1}\mathcal{X}_t$ as $t \to \infty$, the fluctuations of passage times, and the existence and structure of geodesics [1]. On the other hand, our setting requires finite-time estimates of the connectivity of $\mathcal{X}_t$. FPP (often under the title of the SI model) has been studied on large (finite) random graphs, though the main results again pertain to the long-time rather than transient behavior, or to the distance between typical vertices, or to the number of edges contained in the shortest path (hopcount).

Because of the lack of applicable prior results on FPP, the most significant technical effort in this paper is in the analysis of $\mathbb{E}_v[C(\mathcal{X}_t)]$. Our approach is to carefully study of the behavior of the semigroup $e^{t\mathcal{A}}$ on C and on simpler functions which upper bound C.

1.5. **A note on the mean field case.** In fact, even in the well-understood mean field case, the infection process perspective yields an alternative proof of $H_t^k = O((k/n)^2)$ from the hierarchy (1.3). This approach is much simpler than the estimates of iterated exponential integrals given in [46, Section 5], or even the simpler induction carried out in [36, Section 3.1]. The argument is so simple that we give it here. Let $(\mathcal{Y}_t)_{t\ge 0}$ denote the *Yule process* process with rate c_2. This is the most classical pure-birth process, the continuous-time Markov chain which transitions from state k to $k+1$ at rate $c_2 k$, for each integer $k \ge 1$. By the same simple argument which leads from (1.14) to (1.15), the hierarchy (1.3) implies for $1 \le k \le n$ that

$$H_t^k \le \frac{c_1}{n^2} \int_0^t \mathbb{E}_k[(\min(\mathcal{Y}_s, n))^2] \, ds \le \frac{c_1}{n^2} \int_0^t \mathbb{E}_k[\mathcal{Y}_s^2] \, ds. \tag{1.16}$$

The distribution of $\mathcal{Y}_t$ given $\mathcal{Y}_0 = k$ is known explicitly to be negative binomial; it is the same as the law of the number of trials until the kth success, when trials of success probability $p = e^{-c_2 t}$ are repeated independently. The second moment of this distribution is known explicitly and bounded by $2k^2/p^2$, and we recover $H_t^k = O((k/n)^2)$. To connect this argument more concretely to our FPP approach, notice in the complete graph case that $\mathcal{A}_{v \to j} = c_2 k/(n-1)$ for $|v| = k$ and $j \notin v$. It follows that the cardinality process $|\mathcal{X}_t|$ is itself Markovian. Its transition $k \to k+1$ occurs at rate $c_2 k(n-k)/(n-1)$, which is smaller than the corresponding transition $c_2 k$ for the Yule process. By scaling the exponential holding times one can therefore couple $\mathcal{Y}$ with $\mathcal{X}$ in such a way that $|\mathcal{X}_t| \le \mathcal{Y}_t$ a.s.

1.6. **Related literature.**

1.6.1. *Relative entropy methods, global and local.* The literature on mean field limits and propagation of chaos for exchangeable systems is vast. For a comprehensive recent review, we refer to the recent survey [19, 20]. For our purposes, it is worth highlighting some recent progress on relative entropy methods, which can be divided roughly into *global* versus *local* methods. Global entropy methods, based on estimating $H_t([n])$ in our notation, were carried out in [7, 38, 41, 45] for non-singular interactions. The breakthrough [39] revealed the power of entropy methods for singular interactions, which appear in many physically relevant models, and it was then developed further in [14] in conjunction with the modulated energy method initiated by Duerinckx [28] and Serfaty [66]. This has lead to significant progress on Riesz and Coulomb-type interactions, and we refer to [24, 23] for recent results and further references. Entropy methods adapt well to the uniform-in-time setting, usually requiring some form of logarithmic Sobolev inequality [35, 65]. See also [18] for a recent probabilistic approach to singular interactions, based on path-space entropy methods though so far yielding mostly non-quantitative results. An interesting recent contribution [43] shows how to derive concentration inequalities from global entropy estimates.

These global entropy methods at best achieve estimates like $H_t^n = O(1)$, which by subadditivity leads only to the suboptimal $H_t^k = O(k/n)$. To show the optimal order of $(k/n)^2$, the local approach summarized above at (1.3) was developed by the first author in [46]. The followup work [48] treated the uniform-in-time case, which was recently improved in [57] via sharper estimates of the log-Sobolev constants along dynamics. Going a step further, the highly original work [36] showed that the n-particle law P_t admits a cumulant-type expansion in powers of $1/n$ around the product measure $Q_t^{\otimes n}$, and they use hierarchical methods to prove optimal L^2 estimates on each term in the expansion. The main advantage of the local approach is that it can achieve the optimal rate, though it did not appear to handle singularities as well as global methods. That said, the recent work [68] showed how to combine the methods of [46] and [39] in order to achieve the optimal entropy estimate for models with singular interaction functions in $W^{-1,\infty}$. Let us mention also [13], which adopted a different local perpsective based on propagating weighted L^p-norm estimates along the BBGKY hierarchy and was able to rigorously derive the singular Vlasov-Poisson-Fokker-Planck equation on short time horizon.

1.6.2. *Nonexchangeable systems.* The literature on interacting particle systems with heterogeneous interactions has exploded in the past decade, motivated by a wide range of disciplines in which network structures play an important role and cannot be reasonably neglected [58]. We focus the subsequent discussion on the mathematical study of continuous-time and mostly stochastic dynamics, of the form (1.1).

An early thread of this literature focused on the question of *universality*: For what sequences $\xi = \xi^n$ of $n \times n$ interaction matrices does the n-particle system (1.1) converge to the usual McKean-Vlasov limit (1.2)? This was answered first in [10, 25, 22] for Erdös-Rényi graphs $G(n,p)$ (and variants), where ξ^n is $1/np$ times the adjacency matrix, culminating with [59] obtaining the minimal denseness condition $np \to \infty$. More generally, if ξ^n is sufficiently dense and has row sums close to 1, we should expect to achieve the usual McKean-Vlasov limit. Our results quantify this, at least when row sums equal 1, because the right-hand sides of (1.9), (1.10), and (1.11) can be interpreted as measuring the denseness of ξ, especially if we recall how they specialized to the regular graph and Erdös-Rényi cases.

For many sequences ξ^n the McKean-Vlasov equation is not the correct limit. Alternative large-n limits have been analyzed using various concepts from the theory of dense graph limits. Some representative papers in this direction include [54, 21, 55, 5, 9], which take advantage of the well-developed theory of graphons [52] and their L^p extensions [11, 12]. Only recently have some papers [6, 15] made this large-n analysis uniform in time, which is more difficult in the absence of exchangeability perhaps due to the lack of a gradient flow structure, though see [62] for an interesting new perspective on the latter point. Nowadays, the large-n limit theories for nonexchangeable systems appear to be evolving hand-in-hand with modern graph limit theories. The recent papers [44, 33] build on operator-theoretic graph limits proposed in [3] which unify dense and sparse regimes. The very recent [2] builds on hypergraphons originating from [30]. The paper [37] even developed its own new notion of extended graphons, tailored to their specific goals, which will be discussed more below.

Sparse interactions such as those induced by graphs with bounded degree behave differently. In situations where ξ is not sufficiently dense, such as when ξ is the adjacency matrix of the cycle graph on n vertices, there is not enough averaging taking place in (1.1), and we cannot expect nearby particles to become asymptotically independent. A completely different phenomenology arises here, and much remains to be understood. The initial work [60, 49] shows how to derive large-n limits using the notion of *local weak convergence* or *Benjamini-Schramm convergence*, a graph limit theory well-suited to sparse settings introduced in [8], and the followup work [50] identifies a new substitute for the McKean-Vlasov equation in the sparse regime. See [63] for a survey including more recent progress.

All of the above perspectives require some asymptotic structural assumption on the interaction matrix ξ^n, in contrast with our decidedly nonasymptotic approach relying on the independent projection (1.4). The recent paper [42] also employs the independent projection for a nonasymptotic analysis of mean field approximations, but in the context of stochastic control problems, and focusing on global estimates.

The important prior work [37] warrants further discussion. As discussed in Section 1.2, we can see [37] as addressing two separate problems. The first problem is the quantitative approximation of the original n-particle system (1.1) by the independent projection (1.4), and the second is to identify large-n limits for the independent projection. The second problem was the main focus of [37], requiring significant ingenuity to achieve under the minimal assumptions imposed on ξ. In contrast, our main goal is a quantitative solution of the first problem, which was only treated suboptimally in [37, Proposition 2.2]. The main assumptions on ξ are different as well. In [37] it was assumed that the row sums *and column sums* of ξ are $O(1)$ and that the maximal entry is $o(1)$. We share their requirement of bounded row sums, but our results like (1.9) show that (quantitative) propagation of chaos can still hold in the averaged sense $\overline{H}_T^k \to 0$ even without bounded column sums or vanishing maximal entry. The proofs in [37] also adopted a hierarchical technique (developed further in [40]) but of a quite different nature, not dealing directly with the marginals of (1.1) but rather using tree-indexed observables modeled on the notion of homomorphism density from graph theory.

1.7. **Outline of the rest of the paper.** The next Section 2 gives a detailed presentation of our most general setting and main results. Section 3 illustrates how they specialize for certain natural classes of interaction matrices ξ. The proofs of the main results occupy the remaining sections. Section 4 proves the main bound (1.15) in which the infection process $\mathcal{X}_t$ first appears. Section 5 then explains how to estimate various expectations of functions of $\mathcal{X}_t$, which are put to use in Section 6 in order to derive our most user-friendly bounds which were summarized in Section 1.3. The final Section 7 carries out the calculations for a Gaussian example presented in Section 2.7.

Acknowledgement. The first author thanks Louigi Addario-Berry for pointing out the connection with first-passage percolation.

2. Main results

2.1. **Notation.** The number n of particles is fixed throughout the paper, as is the dimension d. Let $[n] := \{1, 2, \ldots, n\}$. For $v \subset [n]$, we denote the size of v by $|v|$.

Given any Polish space E, let $\mathcal{P}(E)$ be the space of Borel probability measures on E. For $\mu \in \mathcal{P}(E)$ and measurable function ϕ on E, let $\langle \mu, \phi \rangle$ denote the integral $\int_E \phi \, d\mu$, when well defined. For $Q \in \mathcal{P}(E^n)$, let $Q^v \in \mathcal{P}(E^v)$ denote the marginal law of the coordinates in v. For brevity, when $v = \{j\}$ is a singleton we omit the bracket and write Q^j.

For any $\mu, \nu \in \mathcal{P}(E)$, the relative entropy is defined as usual by

$$H(\nu \,|\, \mu) := \int_{\mathbb{R}^k} \frac{d\nu}{d\mu} \log \frac{d\nu}{d\mu} \, d\mu, \text{ if } \nu \ll \mu, \quad H(\nu \,|\, \mu) = \infty \text{ if } \nu \not\ll \mu.$$

For $\mu, \nu \in \mathcal{P}(\mathbb{R}^k)$, the relative Fisher information between μ and ν is defined as usual by

$$I(\nu \,|\, \mu) := \int_{\mathbb{R}^k} \left| \nabla \log \frac{d\nu}{d\mu} \right|^2 d\nu,$$

where we set $I(\nu \,|\, \mu) := \infty$ if $\nu \not\ll \mu$ or if the weak gradient $\nabla \log d\nu/d\mu$ does not exist in $L^2(\nu)$. The Wasserstein distance is defined by

$$\mathcal{W}_2(\mu, \nu) := \inf_\pi \left(\int_{\mathbb{R}^k \times \mathbb{R}^k} |x - y|^2 \pi(dx, dy) \right)^{1/2},$$

where the infimum is taken over all $\pi \in \mathcal{P}(\mathbb{R}^k \times \mathbb{R}^k)$ with marginals μ and ν.

We will use some particular notation for probability measures on continuous path space. For $Q \in \mathcal{P}(C([0,\infty); \mathbb{R}^d))$, let $Q_t \in \mathcal{P}(\mathbb{R}^d)$ denote the time-t marginal, i.e., the pushforward of Q by the evaluation map $x \mapsto x_t$. Let $Q_{[t]} \in \mathcal{P}(C([0,t]; \mathbb{R}^d))$ denote the law of the path up to time t, i.e., the pushforward of Q by the restriction map $x \mapsto x|_{[0,t]}$. For $Q \in \mathcal{P}(C([0,\infty); (\mathbb{R}^d)^n))$ and $v \subset [n]$ we will write Q^v_t for the time-t marginal law of the coordinates in v under Q, and we define $Q^v_{[t]}$ similarly.

2.2. **The interacting particle system.** The n-particle system $X_t = (X^1_t, \ldots, X^n_t)$ we study is governed by the following system of stochastic differential equations (SDEs):

$$dX^i_t = \Big(b^i_0(t, X^i_t) + \sum_{j \neq i} \xi_{ij} b^i(t, X^i_t, X^j_t) \Big) dt + \sigma dW^i_t, \quad i = 1, \ldots, n, \tag{2.1}$$

where $W^1, \ldots, W^n$ are independent d-dimensional Brownian motions. Write $P \in \mathcal{P}(C([0,\infty); (\mathbb{R}^d)^n))$ for the law of a weak solution $(X^1, \ldots, X^n)$ of (2.1), started from some given initial distribution $P_0 \in \mathcal{P}((\mathbb{R}^d)^n)$. Here ξ is an $n \times n$ matrix with non-negative entries and zeros on the diagonal, which might be interpreted as the adjacency matrix of a graph on n vertices with weighted edges. The functions $b^i_0 : [0,\infty) \times \mathbb{R}^d \to \mathbb{R}^d$ and $b^i : [0,\infty) \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ are Borel measurable, with more precise assumptions to follow.

Following the terminology of [47], we define the *independent projection* as the solution $Y_t = (Y_t^1, \dots, Y_t^n)$ to the following McKean-Vlasov equation

$$\begin{cases} dY_t^i = \Big(b_0^i(t, Y_t^i) + \sum_{j \neq i} \xi_{ij} \langle Q_t^j, b^i(t, Y_t^i, \cdot) \rangle \Big) dt + \sigma dW_t^i, \quad i = 1, \dots, n \\ Q_t = \mathrm{Law}(Y_t), \quad t \geq 0 \end{cases} \tag{2.2}$$

We write $Q \in \mathcal{P}(C([0,\infty); (\mathbb{R}^d)^n))$ for the law of a weak solution $(Y^1, \dots, Y^n)$ of (2.2), initialized from some product measure $Q_0 = Q_0^1 \otimes \cdots \otimes Q_0^n$. When the SDE (2.2) is well-posed, the coordinates $Y^1, \dots, Y^n$ are independent, because the drift of Y^i depends only on Y^i and not the other coordinates.

Our main results will be estimates on the relative entropies

$$H_t(v) := H(P_t^v \,|\, Q_t^v), \qquad H_{[t]}(v) := H(P_{[t]}^v \,|\, Q_{[t]}^v), \quad v \subset [n], \ t \geq 0. \tag{2.3}$$

Recall that for any $t \geq 0$ and $v \subset [n]$ we write $P_{[t]}^v \in \mathcal{P}(C([0,t]; (\mathbb{R}^d)^v))$ for the law of the path up to time t of the coordinates in v under P; that is, for the law of $(X_s^i)_{s \in [0,t], \, i \in v}$. Similarly, we write P_t^v for the time-t law of $(X_t^i)_{i \in v}$. We write $Q_{[t]}^v$ and Q_t^v for the analogous marginal laws under Q.

2.3. **Assumptions and examples.** Our first set of assumptions will drive our estimates on the path-space entropies $H_{[t]}(v)$, for bounded time intervals. Following [46], rather than making direct assumptions on (b_0, b), we make the following implicit assumptions which emphasize the key ingredients in the method.

Assumption (A). Let $T \in [0, \infty]$.

- (i) Well posedness: The SDEs (2.1) and (2.2) are unique in law from any initial distribution.
- (ii) Square integrability of interaction function:

$$M := \max_{i,j \in [n]} \operatorname*{ess\,sup}_{t \in [0,T)} \int_{(\mathbb{R}^d)^n} \Big| b^i(t, x_i, x_j) - \big\langle Q_t^j, b^i(t, x_i, \cdot) \big\rangle \Big|^2 P_t(dx) < \infty.$$

- (iii) Transport-type inequality: There exists $0 < \gamma < \infty$ such that

$$\Big| \big\langle \nu - Q_t^i, b^i(t, x, \cdot) \big\rangle \Big|^2 \leq \gamma H(\nu \,|\, Q_t^i), \quad \forall i \in [n], \, x \in \mathbb{R}^d, \, \nu \in \mathcal{P}(\mathbb{R}^d), \, t \in [0, T). \tag{2.4}$$

- (iv) The $n \times n$ matrix $\xi = (\xi_{ij})_{i,j=1}^n$ has nonnegative entries, zero diagonal entries $\xi_{ii} = 0$, and bounded row sums:

$$\max_{1 \leq i \leq n} \sum_{j=1}^n \xi_{ij} \leq 1. \tag{rows}$$

Remark 2.1. The row sum bound by 1 condition in (rows) can be relaxed to any other constant, say c. By changing the interaction matrix to ξ/c and the interaction function to cb, we can reduce to the case (rows), with the constants γ and M scaled accordingly by c^2. The restriction that ξ has nonnegative entries is made purely to avoid notational clutter, and it can be removed as long as ξ_{ij} is replaced by $|\xi_{ij}|$ in (rows) and in all of the results to follow.

Example 2.2 (Bounded drift)**.** Suppose all b^i takes the same bounded interaction function b, and b_0^i is such that the SDE $dZ_t^i = b_0^i(t, Z_t^i)dt + \sigma dB_t^i$ is unique in law from any initial position (which holds, e.g., if b_0^i is bounded or Lipschitz). Then Assumption (A) holds. The well-posedness of the independent projection follows from known arguments for McKean-Vlasov equations [45, Theorem 2.5] or [56, Theorem 2]. Condition (ii) holds trivially with $M = 4\||b|^2\|_\infty$, and (iii) follows from Pinsker's inequality, with $\gamma = 2\||b|^2\|_\infty$.

Example 2.3 (Lipschitz drift)**.** Let $T < \infty$. Suppose that b_0^i and b^i are L-Lipschitz and that the initial laws Q_0 and P_0 admit finite second moments. Assume also the following transport inequality: there exists $0 \le \gamma_0 < \infty$ such that

$$\mathcal{W}_2^2(\nu, Q_0^i) \le \gamma_0 H(\nu \,|\, Q_0^i), \quad \forall i \in [n],\ \nu \in \mathcal{P}(\mathbb{R}^d).$$

then Assumption (A) holds. The well-posedness of the independent projection is a straightforward consequence of classical results on McKean-Vlasov equations [47, Proposition 4.1]. It can be shown exactly as in [46, Corollary 2.7] that parts (ii,iii) of Assumption (A) hold, with explicit (n-independent) bounds on γ and M.

Remark 2.4. By no means do examples 2.2 and 2.3 exhaust the scope of Assumption (A). We refer to [46, Section 2B] for further discussion, particularly for the most unusual condition (2.4). In particular, we highlight Remarks 2.12 and 4.5 therein for an explanation of how the arguments extend with minimal effort to kinetic (second-order) models.

Our second and stronger set of assumptions will allow us to obtain uniform-in-time estimates, but (unsurprisingly) only for the time-marginal entropy $H_t(v)$. The following is adapted from [48]:

Assumption (U).

(i) Assumption (A) holds with $T = \infty$.
(ii) Log-Sobolev inequality (LSI): There exists a constant $\eta > 0$ such that

$$H(\nu \,|\, Q_t^i) \le \eta I(\nu \,|\, Q_t^i), \quad \forall \nu \in \mathcal{P}(\mathbb{R}^d),\ i \in [n],\ t \ge 0.$$

(iii) High-temperature/large noise: It holds that $\sigma^2 > 12\eta\gamma$.
(iv) For each $(t, x) \in [0, \infty) \times \mathbb{R}^d$ and $i \in [n]$, we have $b^i(t, x, \cdot) \in L^1(\mathbb{R}^d, Q_t^i)$. The functions b_0^i and $(t, x) \mapsto \langle Q_t^i, b^i(t, x, \cdot)\rangle$ are locally bounded, for each $i \in [n]$. Finally, for each $p, T > 0$,

$$\begin{aligned} \max_{i,j\in[n], i\neq j} \int_0^T \int_{(\mathbb{R}^d)^n} \Big(\big|b^i(t, x^i, x^j)\big|^p + \big|\langle Q_t^j, b^i(t, x^i, \cdot)\rangle\big|^p \Big) P_t(dx) < \infty, \\ \max_{i,j\in[n], i\neq j} \sup_{t\in[0,T)} \int_{(\mathbb{R}^d)^n} \Big(\big|b_0^i(t, x^i)\big|^2 + \big|b^i(t, x^i, x^j)\big|^2 \Big) P_t(dx) < \infty. \end{aligned} \tag{2.5}$$

The essential parts of Assumption (U) are parts (i–iii). As in [48, Assumption (E)], the condition (iv) is purely technical, used only qualitatively to justify an entropy estimate; the values of the integrals play no role in our quantitative bounds. The high-temperature constraint in (iii) is important, as explained in [48, Remark 2.2], and uniform-in-time propagation of chaos can fail for small σ^2. We have not tried to optimize the threshold $12\eta\gamma$, and we certainly do not expect to improve upon [48] in which the threshold was already likely suboptimal, as it could not reach all the way to criticality in the Kuramoto model [48, Example 2.10].

Example 2.5 (Convex potentials)**.** Assume $b_0^i(t, x) = -\nabla U(x)$ and $b^i(t, x, y) = -\nabla W(x - y)$, where U and W are twice continuously differentiable functions satisfying the following:

- W is convex and U is strongly convex , i.e., there exists some $\lambda > 0$ such that $\nabla^2 U \succeq \lambda I$.
- ∇W is bounded, and both ∇U and ∇W are L-Lipschitz for some $L < \infty$.

Suppose the interaction matrix ξ is symmetric, and P_0 admit finite moments of all orders. Assume also the following log-Sobolev inequality: there exists $0 \le \eta_0 < \infty$ such that

$$H(\nu \,|\, Q_0^i) \le \eta_0 I(\nu \,|\, Q_0^i), \quad \forall i \in [n],\ \nu \in \mathcal{P}(\mathbb{R}^d).$$

Then Assumption (U) is satisfied, with $\eta = \max(\eta_0/4, \sigma^2/\lambda)$, $\gamma = 2\||\nabla W|^2\|_\infty$, and $M = 2\gamma$. The proof is a straightforward modification of that of [48, Corollary 2.7], and we give some details in Section A.1. The boundedness condition on ∇W is for convergence, and to relax it would require obtaining bounds on $\max_{i\in[n]} \mathbb{E}|X_t^i|^2$ which are uniform in t and n, which is a delicate task in the presence of heterogeneous interactions.

Example 2.6 (Small interactions on the torus)**.** Suppose the state space is the flat torus $\mathbb{T}^d = \mathbb{R}^d/\mathbb{Z}^d$ instead of $\mathbb{R}^d$. Take $b_0^i \equiv 0$ and $b^i(t,x,y) = K(x-y)$ for some Lipschitz $K : \mathbb{R}^d \to \mathbb{R}^d$. Let $\lambda \geq 1$, and assume Q_0^i admits a smooth density bounded in $[\lambda^{-1}, \lambda]$, for each $i \in [n]$. Finally, assume that $\mathrm{div}K$ is small in the sense that

$$\|\mathrm{div}K\|_\infty \leq 2\sigma^2\pi^2 / \big(1 + \sqrt{2\log\lambda}\big). \tag{2.6}$$

Then Assumption (U) is satisfied. The proof is a modification of that of [48, Corollary 2.9 and Lemma 5.1], and we give the details in Section A.2. We have trivially $\gamma = 2\||K|^2\|_\infty$ and $M = 2\gamma$ (just to be consistent with Example 2.2) by Pinsker's inequality, and the constant η can be taken to be

$$\eta = \frac{\lambda^2}{8\pi^2}\left(1 - \frac{\sqrt{2\log\lambda}\|\mathrm{div}K\|_\infty}{2\sigma^2\pi^2 - \|\mathrm{div}K\|_\infty}\right)^{-1},$$

which is simply $\eta = \lambda^2/8\pi^2$ if K is divergence-free.

2.4. **The general bound via first-passage percolation.** In this section we describe our most general estimates on the relative entropies $H_t(v)$ and $H_{[t]}(v)$ defined in (2.3). It is stated in Theorem 2.7 below in terms of what we call *the infection process* associated with the interaction matrix ξ. We will make use of the following quantities:

$$C(v) := \frac{M}{\sigma^2}\sum_{i\in v}\left(\sum_{j\in v}\xi_{ij}\right)^2, \quad \mathcal{A}_{v\to j} := \frac{\gamma}{\sigma^2}\sum_{i\in v}\xi_{ij}, \qquad \forall v \subset [n],\ j \in [n]\setminus v. \tag{2.7}$$

The infection process is a continuous-time Markov chain $\mathcal{X}$ on the state space $2^{[n]}$ of subsets of $[n]$. Its rate matrix $\mathcal{A}(v,u)$ is defined for $u, v \in 2^{[n]}$ by

$$\mathcal{A}(v,u) = \begin{cases} \mathcal{A}_{v\to j} & \text{if } u = v \cup \{j\},\ \text{for some } j \in [n]\setminus v \\ -\sum_{j\notin v}\mathcal{A}_{v\to j} & \text{if } u = v \\ 0 & \text{otherwise.} \end{cases} \tag{2.8}$$

The key structural feature of $\mathcal{A}$ is that it is a rate matrix, in the sense that $\sum_v \mathcal{A}(u,v) = 0$ for each u, and the off-diagonal entries $v \neq u$ are nonnegative. Let $\mathbb{E}_v[\cdot]$ denote expectation under the initialization $\mathcal{X}_0 = v$. We have the stochastic representation

$$\mathbb{E}_v[F(\mathcal{X}_t)] = e^{t\mathcal{A}}F(v) \quad v \subset [n], \quad t > 0,$$

where we identify a function $F : 2^{[n]} \to \mathbb{R}$ with a vector.

Theorem 2.7.

(1) If Assumption (A) holds for $T < \infty$, then

$$H_{[T]}(v) \leq \mathbb{E}_v[H_0(\mathcal{X}_T)] + \int_0^T \mathbb{E}_v[C(\mathcal{X}_t)]\,dt. \tag{2.9}$$

(2) If Assumption (U) holds, then for all $t > 0$,

$$H_t(v) \leq e^{-\sigma^2 t/4\eta}\mathbb{E}_v\left[H_0(\mathcal{X}_t)\right] + \int_0^t e^{-\sigma^2 s/4\eta}\mathbb{E}_v\left[C(\mathcal{X}_s)\right]ds. \tag{2.10}$$

The remaining, more concrete results presented in this section are all derived from this meta-theorem, via careful estimates of the semigroup $e^{t\mathcal{A}}$. Strictly speaking, the arguments can all be carried out on on purely analytic grounds, without the interpretation of $e^{t\mathcal{A}}$ as the semigroup of a Markov process, but we find this interpretation to be a valuable source of intuition.

2.5. **Concrete bounds.** In this section we give an assortment of more practical bounds on the entropies $H_{[t]}(v)$ and $H_t(v)$ which can be deduced from the general Theorem 2.7. We emphasize results which hold for general matrices ξ, and Section 3 will specialize the results to various classes of ξ. We start with the simplest result to state, for the maximum entropy over sets $v \subset [n]$ of a given size. Define for each $k \in [n]$

$$\widehat{H}^k_{[t]} = \max_{|v|=k} H_{[t]}(v), \qquad \widehat{H}^k_t = \max_{|v|=k} H_t(v).$$

Throughout the section we will make use of the following parameters:

$$\delta := \max_{i,j\in[n]} \xi_{ij}, \qquad \delta_i := \max_{j\in[n]} \xi_{ij} \tag{2.11}$$

Theorem 2.8 (Maximum entropy)**.** *Suppose the following initial chaoticity assumption holds:*

$$\widehat{H}^k_0 \le C_0(\delta k+1)(\delta k)^2, \quad \textit{for all } k \in [n], \tag{2.12}$$

for some constant C_0. If Assumption (A) holds for $T < \infty$, then

$$\widehat{H}^k_{[T]} \le C(\delta k+1)(\delta k)^2, \quad \textit{for all } k \in [n], \tag{2.13}$$

for a constant C depending only on $(C_0, \gamma, M, \sigma, T)$. If Assumption (U) holds, then $\sup_{t\ge0} \widehat{H}^k_t$ is bounded by the same quantity as in (2.13)*, with a constant C depending only on $(C_0, \gamma, M, \sigma, \eta)$.*

In other words, the bound of Theorem 2.8 is $\widehat{H}^k_{[T]} = O((\delta k)^2)$ when $k = O(1/\delta)$. In particular, the size of the maximum entropy is controlled by δ^2, the squared maximal entry of ξ. Remark 2.19 below explains that this cannot be improved in general. It is natural, if we are given a sequence of models with $n \to \infty$, that we cannot expect $\widehat{H}^k_{[T]}$ to vanish if δ does not, because a pair of particles with interaction strength ξ_{ij} bounded away from zero should not become asymptotically independent.

Our next results pertain to the average entropy, which behaves quite differently from the maximum and can be small even when δ is not. For $k \in [n]$, let

$$\overline{H}^k_{[t]} := \frac{1}{\binom{n}{k}} \sum_{|v|=k} H_{[t]}(v), \qquad \overline{H}^k_t := \frac{1}{\binom{n}{k}} \sum_{|v|=k} H_t(v).$$

That is, we are averaging over all $v \subset [n]$ of cardinality k. For some of our results, we will require an additional assumption that the column sums (not just row sums, as in Assumption (A)(iv)) of ξ are bounded by 1:

$$\max_{j\in[n]} \sum_{i=1}^n \xi_{ij} \le 1. \tag{columns}$$

Theorem 2.9 (Average entropy)**.** *Suppose the following initial chaoticity assumption holds:*

$$\widehat{H}^k_0 \le C_0(\delta k+1)k^2\left(\frac{1}{n}\sum_{i=1}^n \delta_i^4\right)^{1/2}, \quad \textit{for all } k \in [n], \tag{2.14}$$

for some finite constant C_0. If Assumption (A) holds for $T < \infty$, then

$$\overline{H}^k_{[T]} \le C(\delta k+1)k^2\left(\frac{1}{n}\sum_{i=1}^n \delta_i^4\right)^{1/2}, \quad \textit{for all } k \in [n], \tag{2.15}$$

for a constant C depending only on $(C_0, \gamma, M, \sigma, T)$. If Assumption (U) holds , then $\sup_{t\ge0} \overline{H}^k_t$ is bounded by the same quantity as in (2.15)*, with a constant C depending only on $(C_0, \gamma, M, \sigma, \eta)$. If we also assume that* (columns) *holds, then both claims remain true with $\left(\frac{1}{n}\sum_{i=1}^n \delta_i^4\right)^{1/2}$ replaced by $\frac{1}{n}\sum_{i=1}^n \delta_i^2$ in both* (2.14) *and* (2.15)*.*

Note that the right-hand side of (2.15) is obviously bounded by that of (2.13). Notably, the former can be small even if some of the rows of ξ have large maximal entry. In a sequence of models with $n \to \infty$, it can certainly be the case that $\frac{1}{n}\sum_{i=1}^n \delta_i^4$ vanishes while $\delta = \max_{i\in[n]} \delta_i$ does not. Intuitively, a few strong connections prevent the maximum entropy $\widehat{H}^k_{[T]}$ from vanishing, but the average entropy $\overline{H}^k_{[T]}$ may still vanish.

In fact, at the price of complexity and assuming (columns), the bound of Theorem 2.9 can be pushed a bit further to sharpen the row-max dependence to certain row-averages:

Theorem 2.10 (Average entropy, without row-max)**.** *Assume* (columns) *holds. Define*

$$p_\xi := \sum_{i,j=1}^n \xi_{ij}^2(\xi_{ij} + \xi_{ji}) + \sum_{i=1}^n \left(\sum_{j=1}^n (\xi_{ij}^2 + \xi_{ji}^2)\right)^2. \tag{2.16}$$

Suppose the following initial chaoticity assumption holds:

$$\widehat{H}^k_0 \le C_0(\delta k + 1)\left(\frac{k^2}{n^2}\sum_{i,j=1}^n \xi_{ij}^2 + \frac{k}{n} p_\xi\right), \quad \text{for all } k \in [n]. \tag{2.17}$$

for some finite constant C_0. If Assumption (A) holds for $T < \infty$, then

$$\overline{H}^k_{[T]} \le C(\delta k + 1)\left(\frac{k^2}{n^2}\sum_{i,j=1}^n \xi_{ij}^2 + \frac{k}{n} p_\xi\right), \quad \text{for all } k \in [n], \tag{2.18}$$

for a constant C depending only on $(C_0, \gamma, M, \sigma, T)$. If Assumption (U) holds, then $\sup_{t\ge 0} \overline{H}^k_t$ is bounded by the same quantity as in (2.18)*, with a constant C depending only on $(C_0, \gamma, M, \sigma, \eta)$.*

Remark 2.11. The bound (2.15) is weaker than (2.18) when ξ is symmetric. Indeed, p_ξ can then be written more concisely as

$$p_\xi = 2\sum_{i,j=1}^n \xi_{ij}^3 + 4\sum_{i=1}^n \left(\sum_{j=1}^n \xi_{ij}^2\right)^2,$$

and we have the following simple estimates for the terms on the right-hand side of (2.18):

$$\begin{aligned}\frac{1}{n^2}\sum_{i,j=1}^n \xi_{ij}^2 \le \frac{1}{n^2}\sum_{i,j=1}^n \delta_i^2 = \frac{1}{n}\sum_{i=1}^n \delta_i^2,&\\ p_\xi \le 2\sum_{i,j=1}^n \delta_i^2 \xi_{ij} + 4\sum_{i=1}^n \left(\sum_{j=1}^n \delta_i \xi_{ij}\right)^2 \le 6\sum_{i=1}^n \delta_i^2,&\end{aligned}$$

with the last step using the row sum bound (rows). Without symmetry of ξ, however, the final term of (2.16) is not controlled by $\sum_i \delta_i^2$, and Theorems 2.9 and 2.10 are not directly comparable.

Remark 2.12. Note by convexity of relative entropy that

$$H\left(\frac{1}{\binom{n}{k}}\sum_{v\subset[n],\ |v|=k} P^v_t \,\middle|\, \frac{1}{\binom{n}{k}}\sum_{v\subset[n],\ |v|=k} Q^v_t\right) \le \overline{H}^k_t.$$

The first measure on the left-hand side is exactly the (exchangeable) law of $(X_t^{\pi(1)}, \ldots, X_t^{\pi(k)})$, where π is a uniformly random permutation of $[n]$, independent of X. Similarly for the second measure. Hence, our bounds on $\overline{H}^k_t$ immediately apply to the symmetrized laws.

We lastly give *setwise* bounds, for each $v \subset [n]$, without taking any average or maximum.

Theorem 2.13 (Setwise entropy)**.** *Define*

$$q_\xi(v) = (\delta|v|+1)\bigg(\sum_{i,j\in v}\xi_{ij}^2 + \delta\sum_{i,j\in v}(\xi^\top\xi+\xi\xi^\top)_{ij} + \delta^2|v|\bigg), \quad v\subset[n]. \tag{2.19}$$

Suppose the following initial chaoticity assumption holds:

$$H_0(v) \le C_0 q_\xi(v), \quad \textit{for all } v\subset[n], \tag{2.20}$$

for some finite constant C_0. If Assumption (A) holds for $T<\infty$, along with (columns)*, then*

$$H_{[T]}(v) \le C q_\xi(v), \quad \textit{for all } v\subset[n], \tag{2.21}$$

for a constant C depending only on (C_0,γ,M,σ,T). If Assumption (U) holds, along with (columns)*, then $\sup_{t\ge 0} H_t(v) \le Cq_\xi(v)$ is bounded by the same quantity as in* (2.21)*, with a constant C depending only on $(C_0,\gamma,M,\sigma,\eta)$. Without assuming* (columns)*, the same conclusions remain true as long as the $\xi^\top\xi$ term is removed from $q_\xi(v)$ in the initial chaoticity assumption* (2.20)*.*

The quantity $q_\xi(v)$ depends not only on the size of v but also on its structure, through the two summations over v. This bound is informative in the regular graph case discussed in Section 3.1 below. It is sharp enough to recover the maximum entropy bounds, in the sense that $q_\xi(v) \lesssim (\delta|v|+1)(\delta|v|)^2$ under assumptions (rows) and (columns), though it is not sharp enough to recover the average entropy bounds. That said, Theorem 2.8 is not a corollary of Theorem 2.13, because the initial chaoticity assumption is stronger in the latter.

Remark 2.14. In certain cases, our entropy bounds transfer immediately to quadratic Wasserstein distance via a Talagrand inequality. For example, in the Lipschitz setting of Example 2.3, the measure $Q^i_{[T]}$ can be shown as in [46] to satisfy the transport inequality $\mathcal{W}_2^2(\cdot,Q^i_{[T]}) \le CH(\cdot\,|\,Q^i_{[T]})$ for a constant independent of i. The quadratic transport inequality tensorizes [34, Proposition 1.9], and so $\mathcal{W}_2^2(\cdot,Q^v_{[T]}) \le CH(\cdot\,|\,Q^v_{[T]})$ for each $v\subset[n]$, with the same constant C. In the uniform-in-time case, by the Otto-Villani theorem [61] (see also [34, Theorem 8.12]), the log-Sobolev inequality of Assumption (U)(ii) implies the quadratic transport inequality $\mathcal{W}_2^2(\cdot,Q^i_t) \le 4\eta H(\cdot\,|\,Q^i_t)$, for all i and t, which tensorizes in the same manner.

2.6. **Reversed entropy.**

Different results can be obtained for $\overleftarrow{H}_{[t]}(v) = H(Q^v_{[t]}\,|\,P^v_{[t]})$, in which the order of the arguments of relative entropy is reversed compared to $H_{[t]}(v)$ defined in (2.3). As in the prior papers [46, 48], the results are somewhat easier to obtain, but only under the stronger assumption that b is bounded; see [46, Remark 4.12] for thoughts on relaxing this assumption. In our setting, under the assumption that b is bounded, the reversed entropy $\overleftarrow{H}_{[t]}(v)$ satisfies all of the same bounds as in Theorems 2.8, 2.9, and 2.10, with the only change being that the prefactor $(\delta k+1)$ is removed (in the time-zero assumptions, as well as the time-T and the uniform-in-time conclusions). The same is true for Theorem 2.13, with the factor $\delta|v|+1$ removed from the definition of $q_\xi(v)$. The proof is somewhat easier, with Remarks 4.2 and 6.2 explaining the differences.

If one is only interested in estimates on the total variation $\|P^v_{[t]} - Q^v_{[t]}\|_{\mathrm{TV}}$, then this can be derived from Pinsker's inequality regardless of the order of arguments in relative entropy. In this sense, the reversed entropy estimate yields a sharper result for total variation, by removing the $\delta k+1$ factor. Of course, this factor is inconsequential when $k=O(1/\delta)$, for instance when k is fixed as $n\to\infty$. In the mean field case where $\xi_{ij}=1/(n-1)$, we have $1/\delta = n-1$, and so it is automatic that $k=O(1/\delta)$. But $k=O(1/\delta)$ is a restriction in general, such as in the m-regular graph case where it requires $k=O(m)$. In other words, we can obtain a larger "size of chaos" by working with reversed entropy.

2.7. **Sharpness, and a Gaussian example.** In this section we discuss a simple Gaussian example, for a which more explicit computations are available. Here, particularly sharp estimates are available, including lower bounds. Consider the following n-particle system with linear drift:

$$dX_t^i = \sum_{j\neq i} \xi_{ij} X_t^j dt + dW_t^i, \quad X_0^i = 0, \quad i \in [n]. \tag{2.22}$$

As usual, ξ is a matrix with non-negative entries and zero diagonal. The law P_t of $(X_t^1, \ldots, X_t^n)$ is the centered Gaussian with covariance matrix

$$\Sigma_t := \int_0^t e^{s\xi} e^{s\xi^\top} ds.$$

The independent projection Y_t must satisfy

$$dY_t^i = \sum_{j\neq i} \xi_{ij} \mathbb{E}[Y_t^j] dt + dW_t^i, \quad Y_0^i = 0, \quad i \in [n]. \tag{2.23}$$

Taking expectations, we find that necessarily $\mathbb{E}[Y_t^j] = 0$, and so $Y^i \equiv W^i$. That is, the law Q_t is the centered Gaussian measure with covariance matrix tI. Thus both P_t and Q_t are two centered Gaussian measures. A well known exact formula for the relative entropy between Gaussians gives

$$H(P_t^v \,|\, Q_t^v) = \frac{1}{2} \mathrm{Tr}\, h(t^{-1}\Sigma_t^v - I),$$

where we define $h(x) = x - \log(1+x)$, and we write A^v for the submatrix of an $n \times n$ matrix A corresponding to those rows and columns indexed by $v \subset [n]$. Noting that $h(0) = h'(0) = 0$, we approximate $h(x)$ to leading order by a quadratic. In particular, letting $\rho = \|\xi\|_{\mathrm{op}}$, we will show

$$H(P_t^v \,|\, Q_t^v) \le \frac{1}{2} e^{8\rho t} \mathrm{Tr}\bigg(\frac{1}{t}\int_0^t (e^{s\xi} e^{s\xi^\top} - I)^v \, ds\bigg)^2. \tag{2.24}$$

For small enough t, specifically $t \le \log(2)/2\rho$, we get a lower bound of the same order,

$$H(P_t^v \,|\, Q_t^v) \ge \frac{1}{6} \mathrm{Tr}\bigg(\frac{1}{t}\int_0^t (e^{s\xi} e^{s\xi^\top} - I)^v \, ds\bigg)^2.$$

We do not consider $t \le \log(2)/2\rho$ to be a significant limitation. Indeed, by the data processing inequality, note that $H(P_{[T]}^v \,|\, Q_{[T]}^v) \ge H(P_t^v \,|\, Q_t^v)$ for $T \ge t \ge 0$. Hence, any lower bound on $H(P_t^v \,|\, Q_t^v)$ for small time applies also to $H(P_{[T]}^v \,|\, Q_{[T]}^v)$ on any longer time horizon.

Without further simplification, the right-hand side of (2.24) admits a network-science interpretation in terms of *communicability* of the set v. Indeed, if ξ is symmetric for simplicity, then expanding out the trace and exponential yields

$$\mathrm{Tr}\bigg(\frac{1}{t}\int_0^t (e^{2s\xi} - I)^v \, ds\bigg)^2 = \sum_{i,j\in v} \bigg(\frac{1}{t}\int_0^t \sum_{\ell=1}^\infty \frac{(2s)^\ell}{\ell!} (\xi^\ell)_{ij} \, ds\bigg)^2.$$

In the language of network science [32], the innermost summation is a measure of the *communicability* of the nodes i and j. The reasoning behind this terminology is that if ξ is the adjacency matrix of a graph, then $(\xi^\ell)_{ij}$ counts the number of length-ℓ paths from i to j, and the power series gives a weighted count over all paths between vertices in v.

It is difficult to simplify the right-hand side of (2.24) in general. However, a simpler bound is available for the *averaged* entropy, at no loss in sharpness. Let us stress that in the following theorem we do not require any bound on the row (or column) sums of ξ, just the spectral norm.

Theorem 2.15. *In the Gaussian setting of this section, we have*

$$\overline{H}_T^k \le 16 e^{10\rho T} \bigg(T^2 \frac{k(k-1)}{n(n-1)} \sum_{i=1}^n \xi_{ij}^2 + \rho^2 T^2 \frac{k(n-k)}{n(n-1)} \sum_{i=1}^n \bigg(\sum_{j=1}^n (\xi_{ij}^2 + \xi_{ji}^2) \bigg)^2 \bigg),$$

for each $T > 0$ and $k \in [n]$. For $T \le \log(2)/2\rho$, we have also the lower bound

$$\overline{H}_T^k \ge \frac{1}{108}\bigg(T^2 \frac{k(k-1)}{n(n-1)} \sum_{i=1}^n \xi_{ij}^2 + T^4 \frac{k(n-k)}{n(n-1)} \sum_{i=1}^n \bigg(\sum_{j=1}^n (\xi_{ij}^2 + \xi_{ij}\xi_{ji})\bigg)^2\bigg). \tag{2.25}$$

Up to a constant factor depending only on ρ and T, the upper and lower bounds in Theorem 2.15 match when the matrix ξ is symmetric (or more generally when it is "not too asymmetric" in the sense that $\xi_{ij} \ge c\xi_{ji}$ holds for a constant c, for all i, j).

Theorem 2.15 reveals a dramatic failure of *subadditivity* to capture the correct behavior of $\overline{H}_T^k$. Indeed, the famous subadditivity inequality of entropy states that

$$\overline{H}_T^k \le \frac{k}{n}\overline{H}_T^n, \qquad 1 \le k \le n.$$

See [26, Theorem 1] for this level of generality, where exchangeability is not assumed. Theorem 2.15 applied with $k = n$ shows that $\overline{H}_T^n \asymp \sum_{ij} \xi_{ij}^2$. Using subaddivity, this implies $\overline{H}_T^k \lesssim (k/n)\sum_{ij} \xi_{ij}^2$, which is much weaker than the upper bound of Theorem 2.15.

Remark 2.16 (Sharpness of Theorem 2.10)**.** The lower bound (2.25) indicates that Theorem 2.10 is sharp in certain regimes. Let us focus on the case of symmetric ξ, for simplicity. For $k = o(n)$, the lower bound (2.25) is of the same order as

$$\frac{k^2}{n^2} \sum_{i,j=1}^n \xi_{ij}^2 + \frac{k}{n} \sum_{i=1}^n \bigg(\sum_{j=1}^n \xi_{ij}^2\bigg)^2. \tag{2.26}$$

In the general setup of Theorem 2.10 specialized to symmetric ξ, when $k = O(1/\delta)$ (e.g., if k is a fixed constant), the upper bound (2.18) is of the same order as

$$\frac{k^2}{n^2} \sum_{i,j=1}^n \xi_{ij}^2 + \frac{k}{n} \sum_{i=1}^n \bigg(\sum_{j=1}^n \xi_{ij}^2\bigg)^2 + \frac{k}{n} \sum_{i,j=1}^n \xi_{ij}^3.$$

The only difference is the third term, which we will see is often of the same order as the second term, in many examples of interest. In summary, assuming $k = O(1/\delta)$ and $k = o(n)$, the bound from Theorem 2.10 is sharp in the regime where

$$\sum_{i,j=1}^n \xi_{ij}^3 = O\bigg(\sum_{i=1}^n \bigg(\sum_{j=1}^n \xi_{ij}^2\bigg)^2\bigg). \tag{2.27}$$

The equality (2.27) is easily seen to hold exactly, without the big-O, in the random walk case of Definition 1.1. In the general case, one can interpret (2.27) as meaning that the non-zero entries of ξ are not too heterogeneous. Indeed, note that Littlewood's interpolation inequality (a simple consequence of Hölder's) implies $(\sum_i x_i^2)^2 \le (\sum_i x_i)(\sum_i x_i^3)$ for any vector $x_i \ge 0$; because we assume row sums of ξ to be bounded by 1, this means that the opposite of (2.27) automatically holds. Equality holds in Littlewood's inequality precisely when the non-zero x_i's are all the same.

For the setwise and maximal entropy bounds given in Theorems 2.13 and 2.8, we must restrict the class of ξ further in order to claim sharpness. This will make use of the following lower bound.

Proposition 2.17. *In the Gaussian setting of this section, for $T \le \log(2)/2\rho$ we have*

$$H_T(v) \ge \frac{T^2}{12} \sum_{i,j\in v} \xi_{ij}^2. \tag{2.28}$$

Proposition 2.18. *In the Gaussian setting of this section, if ξ has row sums bounded by 1, then*

$$H_T(v) \le e^{10T}\delta^2|v|^2, \quad \forall v \subset [n], \tag{2.29}$$

where we set $\delta = \max_{i,j\in[n]} \xi_{ij}$ as usual.

Note that the average of the right-hand side of (2.28) over all $v \subset [n]$ with $|v| = k$ is exactly $\frac{T^2}{12}\frac{k(k-1)}{n(n-1)}\sum_{i,j\in[n]}\xi_{ij}^2$, which only recovers the first term in the bounds of Theorem 2.15. Hence, the inequality (2.28) cannot admit a matching upper bound for every $v \subset [n]$. However, it is sharp for well-connected sets v:

Remark 2.19 (Sharpness of Theorem 2.8)**.** Suppose $v \subset [n]$ is such that $\xi_{ij} = \delta$ for all distinct $i, j \in v$. For example, this holds in the regular graph case if v is a clique. Then Proposition 2.17 becomes $H_T(v) \geq (T^2/12)\delta^2|v|(|v|-1)$, which is of the same order as the upper bound of Proposition 2.18. Moving beyond the Gaussian case, in the regime $k = O(1/\delta)$ this matches the upper bound $\widehat{H}^k_{[T]} = O(\delta^2 k^2)$ from Theorem 2.8.

3. Examples of interaction matrices

In this section, we illustrate how the main results in Section 2 specialize in some noteworthy classes of interaction matrix ξ, mostly arising from simple undirected graphs. In the following, for any vertices i and j of a graph, we write $i \sim j$ to mean that they are adjacent to each other. Recall the definitions of $q_\xi(v)$ in (2.19), p_ξ in (2.16), as well as $\delta := \max_{1\leq i,j\leq n}\xi_{ij}$ and $\delta_i = \max_{j=1,\dots,n}\xi_{ij}$.

Throughout this section, we write $a \lesssim b$ to mean that $a \leq Cb$ for some constant C which can depend on the constants from Assumption (A) but not on n, k, or $v \subset [n]$. The constant may also depend on T, except when Assumption (U) holds. While we do not index our matrix ξ by n, in the example in this section we have in mind an asymptotic regime of a sequence of ξ of size $n \times n$ with $n \to \infty$. Asymptotic notation like $k = o(n)$ should be interpreted accordingly. The number $k \leq n$ of particles is in general allowed to grow with n, except when stated otherwise.

In each of the following examples, we take for granted that Assumption (A) holds, except possibly (rows) which we will justify in each case, so that we may focus our attention on the effects of different choices of interaction matrix ξ. For the same reason we shall assume that $P_0 = Q_0$, so the time-zero assumptions such as (2.12) are trivially satisfied with $C_0 = 0$.

3.1. **Regular graphs.**

Recall the m-regular graph case from Definition 1.2. Clearly the row sums are all equal to 1, and $\delta = 1/m$. Applying Theorem 2.8,

$$\widehat{H}^k_{[T]} \lesssim (k/m)^2 + (k/m)^3, \qquad \text{for } 1 \leq k \leq n, \tag{3.1}$$

which is of course $O((k/m)^2)$ when $k \leq m$. Note that the classical exchangeable setting is recovered when $m = n-1$, which yields $\widehat{H}^k_{[T]} \lesssim (k/n)^2$, recovering the main result of [46].

To estimate the average entropy, we get a slightly sharper estimate from Theorem 2.10 than from Theorem 2.9 (as expected from Remark 2.11). In fact, noting that $\delta_i = \delta = 1/m$ for all i, Theorem 2.9 simply bounds $\overline{H}^k_{[T]}$ by the same right-hand side as (3.1), which of course also follows trivially from the inequality $\overline{H}^k_{[T]} \leq \widehat{H}^k_{[T]}$. To use Theorem 2.10, we compute

$$\sum_{i,j=1}^n \xi_{ij}^2 = \frac{n}{m}, \qquad \sum_{i=1}^n\left(\sum_{j=1}^n \xi_{ij}^2\right)^2 = \frac{n}{m^2}, \qquad \sum_{i,j=1}^n \xi_{ij}^3 = \frac{n}{m^2}.$$

Combined with $\delta = 1/m$, applying Theorem 2.10 yields

$$\overline{H}^k_{[T]} \lesssim \left(\frac{k}{m}+1\right)\left(\frac{k^2}{nm}+\frac{k}{m^2}\right). \tag{3.2}$$

For $k = O(1)$, this bound is again of order $1/m^2$, but when k is allowed to grow with n it reveals an interesting new detail compared to the preceding bounds. Specifically, unlike the previous bounds, (3.2) can vanish even in cases where k is larger than m; precisely it vanishes when $k = o(\min(m^{3/2}, (nm^2)^{1/3}))$, for instance when $m = O(n^{2/5})$ and $k = o(m^{3/2})$.

To apply the setwise entropy estimate of Theorem 2.13, it will be helpful to write $\xi = (1/m)A$, where A is the adjacency matrix of the underlying m-regular graph. Then

$$q_\xi(v) = \Big(\frac{|v|}{m} + 1\Big)\Big(\frac{1}{m^2}\sum_{i,j\in v} A_{ij} + \frac{2}{m^3}\sum_{i,j\in v}(A^2)_{ij} + \frac{|v|}{m^2}\Big). \tag{3.3}$$

The two summations on the right-hand side count, respectively, the number of edges in v and the number of paths of length two which start and end in v. The latter is at least $m|v|$, as seen by retaining only the $i = j$ terms in the sum. Thus, the last term $|v|/m^2$ of (3.3) is dominated by the second to last term. Hence, Theorem 2.13 implies

$$H_{[T]}(v) \lesssim q_\xi(v) \lesssim \left(\frac{|v|}{m} + 1\right)\left(\frac{1}{m^2}\sum_{i,j\in v} A_{ij} + \frac{1}{m^3}\sum_{i,j\in v}(A^2)_{ij}\right), \quad v \subset [n]. \tag{3.4}$$

Two extreme cases illustrate the range of values this can take, depending on how "connected" the set v is. If v is "highly disconnected," in the sense that there are no paths of length one or two between distinct vertices in v, then (3.4) becomes

$$H_{[T]}(v) \lesssim \left(\frac{|v|}{m} + 1\right)\frac{|v|}{m^2},$$

which is small as long as $|v| = o(m^{3/2})$. If instead v is highly connected, for instance a clique (which in particular implies $|v| \le m$), then there are $|v|(|v|-1)$ directed edges in v, and (3.4) becomes

$$H_{[T]}(v) \lesssim \frac{|v|^3}{m^3} + \frac{|v|^2}{m^2},$$

which is small if $|v| = o(m)$, and is the same order as the maximal entropy $\widehat{H}^k_{[T]}$ when $|v| = k$. In summary, the size of $H_{[T]}(v)$ is controlled by a tradeoff between the size of v and its connectedness.

3.2. **Random walk on a graph.** Generalizing beyond the regular graph case, consider now a connected graph on vertex set $[n]$, and let ξ denote the transition probability matrix of the simple random walk on the graph. That is, if vertex $i \in [n]$ has m_i neighbors, then $\xi_{ij} = (1/m_i)1_{i\sim j}$. Note that ξ is asymmetric except in the regular graph case. The row sum condition (rows) is clearly satisfied. We have

$$\delta = \frac{1}{m_*}, \text{ where } m_* := \min_{i\in[n]} m_i, \qquad \text{and} \qquad \delta_i = \frac{1}{m_i}.$$

Applying Theorem 2.8, we deduce

$$\widehat{H}^k_{[T]} \lesssim (k/m_*)^2 + (k/m_*)^3,$$

which is of course $O((k/m_*)^2)$ when $k \le m_*$. In other words, the maximal entropy is controlled by the minimum degree. For the average entropy, we apply Theorem 2.9 to find

$$\overline{H}^k_{[T]} \lesssim \Big(\frac{k}{m_*} + 1\Big)\Big(\frac{1}{n}\sum_{i=1}^n \frac{k^4}{m_i^4}\Big)^{1/2}. \tag{3.5}$$

Thus, the average entropy is controlled by the "typical" degree in a certain sense. The average entropy vanishes as long as the typical degree diverges, in contrast with the maximum entropy which vanishes only when the degrees diverge uniformly. Note for a sequence of connected graphs that $\frac{1}{n}\sum_{i=1}^n \frac{1}{m_i^4} \to 0$ if and only if $m_U \to \infty$ in probability, where $U \sim \text{Uniform}[n]$. If we have bounded column sums, which here means that

$$\max_{j\in[n]} \sum_{i\sim j} \frac{1}{m_i} \le 1, \tag{3.6}$$

then we can apply the final claim of Theorem 2.9 to get the sharper bound

$$\overline{H}^k_{[T]} \lesssim \Big(\frac{k}{m_*}+1\Big)\frac{1}{n}\sum_{i=1}^n \frac{k^2}{m_i^2}. \tag{3.7}$$

Note as in Remark 2.1 that if the right-hand side of (3.6) is a constant other than 1, we could change it to 1 by rescaling b in proportion.

An alternative and more complicated estimate of the average entropy arises from Theorem 2.10, which again requires the column sum bound (3.6). We compute $\sum_{i,j}\xi_{ij}^2 = \sum_i \frac{1}{m_i}$ and

$$\begin{aligned} p_\xi &= 2\sum_{i=1}^n \frac{1}{m_i^2} + \sum_{i=1}^n\sum_{j\sim i}\frac{1}{m_j m_i^2} + \sum_{i=1}^n\Big(\sum_{j\sim i}\frac{1}{m_j^2}\Big)^2 + 2\sum_{i=1}^n \frac{1}{m_i}\sum_{j\sim i}\frac{1}{m_j^2} \\ &\le 4\sum_{i=1}^n \frac{1}{m_i^2} + 2\sum_{i=1}^n\Big(\sum_{j\sim i}\frac{1}{m_j^2}\Big)^2, \end{aligned}$$

where the last step uses the assumption (3.6). Then Theorem 2.10 implies

$$\overline{H}^k_{[T]} \lesssim \Big(\frac{k}{m_*}+1\Big)\Big(\frac{k^2}{n^2}\sum_{i=1}^n\frac{1}{m_i} + \frac{k}{n}\sum_{i=1}^n\frac{1}{m_i^2} + \frac{k}{n}\sum_{i=1}^n\Big(\sum_{j\sim i}\frac{1}{m_j^2}\Big)^2\Big) \tag{3.8}$$

For $k=O(1)$, the second term of (3.8) is larger than the first because $\frac{1}{n^2}\sum_{i=1}^n\frac{1}{m_i} \le \frac{1}{n}\sum_{i=1}^n\frac{1}{m_i^2}$. As noted in Remark 2.11, the bounds of (3.7) and (3.8) are not comparable in general.

This example can be applied to many models of random graphs, in the *quenched* sense where the realization of the random graph determines ξ as input to our main theorems. For example, consider the Erdös-Rényi graph with edge probability p. Here p is allowed to depend on n, but we suppress this dependence. As long as $\liminf_{n\to\infty} np/\log n > 1$, we may find a constant β such that

$$\lim_{n\to\infty}\mathbb{P}\Big(\max_{j\in[n]}\sum_{i\sim j}\frac{1}{m_i}\le\beta\Big)=1.$$

Hence, with high probability, the Erdös-Rényi graph leads to ξ with bounded column sums; see Proposition B.1(1) in the appendix, and note that the "high probability" and the constant β could be quantified with more care. Note that the condition $\liminf_{n\to\infty} np/\log n > 1$ is exactly the one required for connectivity of the graph (with high probability), and for the divergence (in probability) of the minimum degree m_*.

3.3. **Scaled adjacency matrix.** As another example, suppose A is the adjacency matrix of a graph G with vertex set $[n]$ and nonempty edge set E. Let $\delta>0$ be a scalar, and set $\xi=\delta A$, so that in particular $\delta=\max_{ij}\xi_{ij}$ as usual. We have two cases in mind:

(I) G is non-random, and $\delta = n/2|E|$ is the reciprocal of the average degree.
(II) G is (a realization of) the Erdös-Rényi graph with edge probability p, and $\delta=1/np$.

This is the natural scaling which ensures that the average row sum is 1, or $(1/n)\sum_{i,j=1}^n\xi_{ij}=1$, in expectation in case (II). Then ξ is symmetric, and its maximal row sum is δm^*, where m^* is the maximal degree of the graph (m^* was used in Section 3.2 to denote the minimal degree. Perhaps we change the max degree to $\widehat{m}$ and the min degree to $\check{m}$? Or $m_{\max}$ and $m_{\min}$?). The bounded row sum assumption (rows), or rather its relaxation in Remark 2.1, is valid as long as $m^*\lesssim 1/\delta$. In case (I) above, this means that the maximal degree is of the same order as the average degree. In case (II), this means that $m^*\lesssim np$, which holds with high probability as $n\to\infty$, even if p is allowed to vanish as $n\to\infty$, as long as $\liminf np/\log n>0$; see Proposition B.1(2) for a precise statement and proof.

Note also that $\max_{i,j}\xi_{ij}=\delta$, and the maximum entropy bound of Theorem 2.8 is easy to apply. In this case, $\max_j\xi_{ij}=\delta$ for any non-isolated vertex i, so the average entropy bound of Theorem

2.9 yields no improvement over Theorem 2.8. To apply Theorem 2.10, we compute with the help of Remark 2.11

$$\sum_{i,j=1}^{n} \xi_{ij}^2 = \delta^2 \sum_{i=1}^{n} m_i, \quad \text{and} \quad p_\xi = 2\delta^3 \sum_{i=1}^{n} m_i + 4\delta^4 \sum_{i=1}^{n} m_i^2,$$

where m_i again denotes the degree of vertex i. Then Theorem 2.10 yields

$$\overline{H}_{[T]}^k \lesssim (\delta k+1)\bigg(\delta^2 \frac{k^2}{n^2} \sum_{i=1}^{n} m_i + \delta^3 \frac{k}{n} \sum_{i=1}^{n} m_i + \delta^4 \frac{k}{n} \sum_{i=1}^{n} m_i^2\bigg). \tag{3.9}$$

Turning to the setwise entropy estimate of Theorem 2.13, assuming $v \subset [n]$ is of size $|v| = O(1/\delta)$ again for simplicity, we have

$$H_{[T]}(v) \lesssim q_\xi(v) \lesssim \delta^2 \sum_{i,j\in v} A_{ij} + \delta^3 \sum_{i,j\in v} (A^2)_{ij} + |v|\delta^2.$$

Note that the first summation equals twice the number of edges in v, and the second summation equals twice the number of paths of length two which start and end in v. Let us summarize how these specialize in the two cases mentioned above:

(I) Letting $\overline{m} = 1/\delta = \frac{1}{n}\sum_{i=1}^n m_i$ denote the average degree, we simplify the above to

$$\widehat{H}_{[T]}^k \lesssim (k/\overline{m})^2 + (k/\overline{m})^3, \qquad \overline{H}_{[T]}^k \lesssim \left(\frac{k}{\overline{m}}+1\right)\bigg(\frac{k^2}{n\overline{m}} + \frac{k}{\overline{m}^4}\frac{1}{n}\sum_{i=1}^{n} m_i^2\bigg). \tag{3.10}$$

The maximal entropy vanishes as long as the average degree diverges and $k = o(\overline{m})$. Curiously, for the average entropy to vanish, the second moment (or variance) of the degrees cannot be too large relative to the first moment. By Jensen's inequality, we can have no better than $\overline{m}^2 = O(\frac{1}{n}\sum_{i=1}^n m_i^2)$, in which case the last term of (3.10) becomes $k/\overline{m}^2$, and the bound on $\overline{H}_{[T]}^k$ then behaves exactly like (3.2) except with m replaced by $\overline{m}$. For fixed k, both bounds are of the same order and vanish when the average degree diverges. The former still vanishes when k grows, as long as $k = o(\overline{m})$, while the latter can accommodate larger k.

(II) In the Erdös-Rényi case, we get $\widehat{H}_{[T]}^k \lesssim (k/np)^2 + (k/np)^3$ with high probability. For $np \ge 1$, the bound (3.9) on $\overline{H}_{[T]}^k$ behaves in expectation exactly like (3.2) except with m replaced by np. Note that to apply our theorems in this case we need to check that the row sums of the (symmetric) matrix $\xi = A/np$ are bounded; this holds with high probability by a standard concentration argument given in Proposition B.1(2) in the appendix.

3.4. **Rank-one matrices.** Suppose $\alpha, \beta \in \mathbb{R}^n$ have nonnegative entries, and $\xi_{ij} = \alpha_i\beta_j$ for $i \neq j$, with $\xi_{ii} = 0$ as usual. Then the corresponding n-particle system is given by

$$dX_t^i = \bigg(b_0^i(t, X_t^i) + \alpha_i \sum_{j\neq i} \beta_j b^i(t, X_t^i, X_t^j)\bigg)dt + \sigma dW_t^i, \quad i = 1,\dots,n.$$

This class of examples arises naturally in the classical n-body problem of masses interacting via gravitational force, where $\alpha_i = \beta_i$ is the mass of the ith body. That said, the true model is second-order (kinetic), typically noiseless, and with a singular interaction kernel, rendering our results not directly applicable.

Write $|\cdot|_p$ for the ℓ_p norm on $\mathbb{R}^n$, for $p \in [1,\infty]$. The row sums of ξ are bounded by 1 (as required by our main theorems) if and only if $|\beta|_1|\alpha|_\infty \le 1$. For the few results of ours which required bounded column sums, we would also require $|\alpha|_1|\beta|_\infty \le 1$.

To apply our main theorems, we compute

$$\delta = |\alpha|_\infty|\beta|_\infty, \qquad \delta_i = \alpha_i|\beta|_\infty.$$

Bound the maximum entropy using Theorem 2.8 by

$$\widehat{H}^k_{[T]} \lesssim k^3|\alpha|^3_\infty|\beta|^3_\infty + k^2|\alpha|^2_\infty|\beta|^2_\infty. \tag{3.11}$$

Bound the average entropy using Theorem 2.9 by

$$\overline{H}^k_{[T]} \lesssim \big(k|\alpha|_\infty|\beta|_\infty + 1\big)k^2|\beta|^2_\infty \frac{|\alpha|^2_4}{\sqrt{n}}. \tag{3.12}$$

If $|\alpha|_1|\beta|_\infty \le 1$ so that the column sum condition is satisfied, the factor $|\alpha|_4^2/\sqrt{n}$ improves to $|\alpha|_2^2/n$. To compare (3.11) and (3.12), let us assume that k is fixed and that $|\beta|_\infty = 1$. Then $\widehat{H}^k_{[T]} \to 0$ as long as $|\alpha|_\infty \to 0$, whereas $\overline{H}^k_{[T]} \to 0$ requires merely that $(1/n)\sum_i \alpha_i^4 \to 0$ (with the 4 replaced by 2 if $|\alpha|_1|\beta|_\infty \le 1$). Theorems 2.10 and 2.13 could be applied here as well, but the results are messier and not do not add much intuition, in our opinion.

An example worth mentioning is when $\alpha_i = 1/n$ for all i, which was studied in [69]. Therein, a mean field limit for the weighted empirical measure $(1/n)\sum_{i=1}^n \beta_i \delta_{X^i_t}$ was shown. Our theorems imply a different sort of result. Note that the row sum $(1/n)\sum_i \beta_i$ does not depend on i, so by rescaling b and recalling the discussion around (1.5), we find that the independent projection is given by i.i.d. copies of the McKean-Vlasov equation. For the row and column sums of ξ to be bounded, we need to assume $|\beta|_\infty = O(1)$, corresponding to the strongest case ($r = \infty$) of [69, Assumption (1.2)]. The maximal matrix entry is then $\delta = |\beta|_\infty/n = O(1/n)$, and the same bounds as in the mean field case are valid, such as $\widehat{H}^k_{[T]} = O((k/n)^2)$.

3.5. **Sequential propagation of chaos.** The recent paper on [27] studies the case where ξ is lower-triangular, so that each particle i in sequence is influenced only by a weighted average over the previous particles $j < i$. A notable special case is where $\xi_{ij} = 1_{j<i}/(i-1)$, so that each particle i views the previous particles $j < i$ as exchangeable and averages over them; the fastest convergence in [27] is in this setting. In this case, the maximum and setwise entropy estimates (Theorems 2.8 and 2.13) are useless, because the maximal entry of this matrix is $\delta = 1$. But we have $\delta_i = 1/(i-1)$ for $i \ge 2$, and Theorem 2.9 yields the following estimate on the average entropy:

$$\overline{H}^k_{[T]} \lesssim k^3 \left(\frac{1}{n}\sum_{i=2}^n \frac{1}{(i-1)^4}\right)^{1/2} \lesssim \frac{k^3}{\sqrt{n}}.$$

4. From the particle system to the infection process

This section is devoted to the proof of Theorem 2.7, which bounds the entropies $H_{[t]}(v)$ and $H_t(v)$ in terms of the infection process. To this end, we first derive in Section 4.1 the hierarchy of differential inequalities satisfied by these entropies, stated in (1.14) in the introduction. Section 4.2 then shows how to deduce Theorem 2.7 from these hierarchies.

The following shorthand notation will be useful: For $v \subset [n]$ and $j \in [n]\backslash v$, let $vj := v \cup \{j\}$.

4.1. **The hierarchy of differential inequalities.** Our first lemma pertains to the path-space entropies $H_{[t]}(v)$, following and adapting the strategy developed in [46] for the exchangeable case; see specifically the proof of Theorem 2.2 therein, up to equation (4-18). Recall in the following the definitions (2.7) related to the infection process.

Lemma 4.1. *Suppose Assumption (A) holds. Let $v \subset [n]$.*

(1) The map $t \mapsto H_{[t]}(v)$ is absolutely continuous, and for a.e. $t \in [0,T]$,

$$\frac{d}{dt}H_{[t]}(v) \le C(v) + \sum_{j \notin v} \mathcal{A}_{v\to j}\left(H_{[t]}(vj) - H_{[t]}(v)\right), \tag{4.1}$$

By convention, the final term of (4.1) is zero if $v = [n]$.

(2) If it holds for some constant h_3 that

$$H_{[T]}(v) \le h_3, \quad \text{for all } v \subset [n] \text{ with } |v| = 3, \tag{4.2}$$

then (4.1) *holds with $C(v)$ replaced by*

$$\widehat{C}(v) := \frac{\sqrt{\gamma M h_3}}{\sigma^2} \sum_{i \in v} \Big(\sum_{j \in v} \xi_{ij} \Big)^2 + \frac{M}{\sigma^2} \sum_{i,j \in v} \xi_{ij}^2. \tag{4.3}$$

At first we will apply part (1). As in [46], after we have a good bound on h_3 from a first pass through the argument, we will apply part (2) and repeat the argument to sharpen the results.

Proof of Lemma 4.1. We begin by treating the case of $v = [n]$ separately, in part for transparency and in part for the technical purpose of implying that $H_{[T]}(v) < \infty$ for all $v \subset [n]$. We will first apply [46, Lemma 4.4(iii)], a well known entropy estimate based on Girsanov's theorem. As a preparation, we first show that the assumptions of [46, Lemma 4.4 (iii)] are satisfied. Thanks to the well-posedness in Assumption (A)(i), we only need to verify the integrability condition in [46, Equation (4.9)], which in our context requires showing

$$\sum_{i=1}^n \int_0^T \int_{(\mathbb{R}^d)^n} \left| \sum_{j \ne i} \xi_{ij} \left(b^i(t, x_i, x_j) - \left\langle Q_t^j, b^i(t, x_i, \cdot) \right\rangle \right) \right|^2 P_t(dx) dt < \infty,$$

$$\sum_{i=1}^n \int_0^T \int_{(\mathbb{R}^d)^n} \left| \sum_{j \ne i} \xi_{ij} \left(b^i(t, y_i, y_j) - \left\langle Q_t^j, b^i(t, y_i, \cdot) \right\rangle \right) \right|^2 Q_t(dy) dt < \infty.$$

The first of these two claims is a straightforward consequence of Assumption (A)(ii). The second follows from the fact that, by [46, Lemma 2.3 and Remark 2.4(i)], our Assumption (A)(iii) implies the following much stronger exponential square-integrability: there exists $\kappa > 0$:

$$\sup_{(t,x) \in [0,T] \times \mathbb{R}^d} \log \int_{\mathbb{R}^d} \exp \left(\kappa \left| b^i(t, y_i, y_j) - \left\langle Q_t^j, b^i(t, y_i, \cdot) \right\rangle \right|^2 \right) Q_t^j(dy_j) < \infty.$$

We may now apply [46, Lemma 4.4, Equation (4.8)] to find

$$\frac{d}{dt} H_{[t]}([n]) = \frac{1}{2\sigma^2} \sum_{i=1}^n \mathbb{E}\Big[\Big| \sum_{j \ne i} \xi_{ij} \left(b^i(t, X_t^i, X_t^j) - \langle Q_t^j, b^i(t, X_t^i, \cdot) \rangle \right) \Big|^2 \Big] \le \frac{M}{2\sigma^2} \sum_{i=1}^n \Big(\sum_{j \ne i} \xi_{ij} \Big)^2,$$

where we used the Cauchy-Schwarz inequality and Assumption (A)(ii) in the last step. Since $H_0([n]) < \infty$, we deduce that $H_{[T]}([n]) < \infty$ as claimed.

Next, we identify the dynamics for any subset $v \in [n]$ of particles. For a path $x \in C([0,T]; \mathbb{R}^d)$ write $x_{[t]} = (x_s)_{s \in [0,t]} \in C([0,t]; \mathbb{R}^d)$ for the path up to time $t \le T$, and similarly for $x \in C([0,T]; (\mathbb{R}^d)^v)$. Write $\mathbb{F}^v = (\mathcal{F}_t^v)_{t \in [0,T]}$ for the filtration generated by the particles in v, i.e., $\mathcal{F}_t^v$ is generated by the random variable $X_{[t]}^v$. For any $i \in v, j \notin v$, there exists a progressively measurable function $\widehat{b}_{ij}^v : [0,T] \times C([0,T]; (\mathbb{R}^d)^v) \to \mathbb{R}^d$ such that

$$\widehat{b}_{ij}^v(t, X^v) = \mathbb{E}[b^i(t, X_t^i, X_t^j) \,|\, \mathcal{F}_t^v], \quad a.s., \ a.e. \ t \in [0,T]. \tag{4.4}$$

For any $i \in v$, we compute the conditional expectation of the drift of X_t^i given $\mathcal{F}_t^v$:

$$\mathbb{E}\Big[b_0^i(t, X_t^i) + \sum_{j \ne i} \xi_{ij} b^i(t, X_t^i, X_t^j) \,\Big|\, \mathcal{F}_t^v \Big] = b_0^i(t, X_t^i) + \sum_{j \in v} \xi_{ij} b^i(t, X_t^i, X_t^j) + \sum_{j \notin v} \xi_{ij} \widehat{b}_{ij}^v(t, X^v).$$

By a projection argument [46, Lemma 4.1], we may change the Brownian motions so that this conditional expectation becomes the drift of X^i_t, for each $i \in v$. Precisely, there exist independent $\mathbb{F}^v$-Brownian motions $(\widehat{W}^i)_{i\in v}$ such that

$$dX^i_t = \Big(b^i_0(t,X^i_t) + \sum_{j\in v} \xi_{ij} b^i(t,X^i_t,X^j_t) + \sum_{j\notin v} \xi_{ij} \widehat{b}^v_{ij}(t,X^v)\Big)dt + \sigma d\widehat{W}^i_t, \quad i \in v. \tag{4.5}$$

For the independent projection (2.2), the particles in v solves the SDE system

$$dY^i_t = \Big(b^i_0(t,Y^i_t) + \sum_{j=1}^n \xi_{ij} \big\langle Q^j_t\,,\, b^i(t,Y^i_t,\cdot)\big\rangle\Big)dt + \sigma dW^i_t, \quad i \in v. \tag{4.6}$$

With these dynamics identified, we will apply the entropy identities [46, Lemma 4.4 (ii)] to (4.5) and (4.6). To justify this, note that by the data processing inequality that $H_{[t]}(v) \le H_{[T]}(v) \le H_{[T]}([n])$ holds for any subset $v \subset [n]$, and $H_{[T]}([n])$ is finite as was shown in the first part of the proof. Thus,

$$\begin{aligned}
\frac{d}{dt}H_{[t]}(v) &= \frac{1}{2\sigma^2}\sum_{i\in v}\mathbb{E}\Bigg[\bigg|\sum_{j\in v}\xi_{ij}b^i(t,X^i_t,X^j_t) + \sum_{j\notin v}\xi_{ij}\widehat{b}^v_{ij}(t,X^v) - \sum_{j\neq i}\xi_{ij}\Big\langle Q^j_t, b^i(t,X^i_t,\cdot)\Big\rangle\bigg|^2\Bigg] \\
&\le \frac{1}{\sigma^2}\sum_{i\in v}\mathbb{E}\Bigg[\bigg|\sum_{j\in v}\xi_{ij}\Big(b^i(t,X^i_t,X^j_t) - \big\langle Q^j_t, b^i(t,X^i_t,\cdot)\big\rangle\Big)\bigg|^2\Bigg] \\
&\quad + \frac{1}{\sigma^2}\sum_{i\in v}\mathbb{E}\Bigg[\bigg|\sum_{j\notin v}\xi_{ij}\Big(\widehat{b}^v_{ij}(t,X^v) - \Big\langle Q^j_t, b^i(t,X^i_t,\cdot)\Big\rangle\Big)\bigg|^2\Bigg] \\
&=: \mathrm{I} + \mathrm{II}.
\end{aligned} \tag{4.7}$$

To control term I, we simply use the Cauchy-Schwarz inequality and recall the definition of M from Assumption (A)(ii):

$$\begin{aligned}
\mathrm{I} &\le \frac{1}{\sigma^2}\sum_{i\in v}\bigg(\sum_{j\in v}\xi_{ij}\bigg)\bigg(\sum_{j\in v}\xi_{ij}\mathbb{E}\Big[\Big|b^i(t,X^i_t,X^j_t) - \Big\langle Q^j_t, b^i(t,X^i_t,\cdot)\Big\rangle\Big|^2\Big]\bigg) \qquad (4.8) \\
&\le \frac{M}{\sigma^2}\sum_{i\in v}\bigg(\sum_{j\in v}\xi_{ij}\bigg)^2 = C(v).
\end{aligned}$$

For term II, we introduce some additional notation. Let $P^{j|v}_{t;X^v_{[t]}}(dx^j_t)$ denote a version of the regular conditional law of X^j_t given $X^v_{[t]}$, and let $P^{j|v}_{[t];X^v_{[t]}}(dx^j_{[t]})$ denote a version of the regular conditional law of $X^j_{[t]}$ given $X^v_{[t]}$. Then the assumed transport-type inequality (2.4) implies

$$\begin{aligned}
\big|\widehat{b}^v_{ij}(t,X^v) - \big\langle Q^j_t, b^i(t,X^i_t,\cdot)\big\rangle\big|^2 &= \big|\big\langle P^{j|v}_{t;X^v_{[t]}} - Q^j_t, b^i(t,X^i_t,\cdot)\big\rangle\big|^2 \\
&\le \gamma H\big(P^{j|v}_{t;X^v_{[t]}}\,|\,Q^j_t\big) \le \gamma H\big(P^{j|v}_{[t];X^v_{[t]}}\,|\,Q^j_{[t]}\big), \quad a.s.,
\end{aligned}$$

where we used the data-processing inequality in the last step. Recalling that $\mathbb{E}$ denotes expectation under P, the chain rule for relative entropy implies

$$\mathbb{E}\big[H\big(P^{j|v}_{[t];X^v_{[t]}}\,|\,Q^j_{[t]}\big)\big] = H_{[t]}(vj) - H_{[t]}(v).$$

Therefore, using Cauchy Schwarz inequality again along with the assumption of bounded row sums $\sum_{j\notin v}\xi_{ij} \le \sum_{j=1}^n \xi_{ij} \le 1$, we see that

$$\begin{aligned}
\mathrm{II} &\le \frac{1}{\sigma^2}\sum_{i\in v}\Bigg(\sum_{j\notin v}\xi_{ij}\Bigg)\Bigg(\sum_{j\notin v}\xi_{ij}\mathbb{E}\Big[\big|\widehat{b}^v_{ij}(t,X^v_t) - \big\langle Q^j_t, b^i(t,X^i_t,\cdot)\big\rangle\big|^2\Big]\Bigg) \\
&\le \frac{\gamma}{\sigma^2}\sum_{i\in v}\Bigg(\sum_{j\notin v}\xi_{ij}\Big(H_{[t]}(vj) - H_{[t]}(v)\Big)\Bigg) \\
&= \sum_{j\notin v}\mathcal{A}_{v\to j}\Big(H_{[t]}(vj) - H_{[t]}(v)\Big).
\end{aligned} \tag{4.9}$$

Combining the bounds on terms I and term II, we arrive at (4.1). This last use of Cauchy-Schwarz might be somewhat lossy; see Section **??** below for a discussion of a refined approach from which we were unable, however, to extract any sharper results.

We next prove (2). We improve the bound (4.8) by instead expanding the square to get the identity I=I(a)+I(b), where

$$\mathrm{I(a)} = \frac{1}{\sigma^2}\sum_{i,j,r\in v, j\neq r}\xi_{ij}\xi_{ir}\mathbb{E}\Big[\Big(b^i(t,X^i_t,X^j_t) - \Big\langle Q^j_t, b^i(t,X^i_t,\cdot)\Big\rangle\Big)\cdot\Big(b^i(t,X^i_t,X^r_t) - \big\langle Q^r_t, b^i(t,X^i_t,\cdot)\big\rangle\Big)\Big]$$

$$\mathrm{I(b)} = \frac{1}{\sigma^2}\sum_{i,j\in v}\xi_{ij}^2\mathbb{E}\Big[\Big|b^i(t,X^i_t,X^j_t) - \Big\langle Q^j_t, b^i(t,X^i_t,\cdot)\Big\rangle\Big|^2\Big].$$

Recall that the diagonal entries of ξ are zero, so the terms in the sums vanish if any index i, j, or r is repeated. Using the above notation for conditional measures, we condition on (X^i_t, X^j_t) and use the Cauchy-Schwarz inequality to get, for distinct $i,j,r\in v$,

$$\begin{aligned}
&\mathbb{E}\Big[\Big(b^i(t,X^i_t,X^j_t) - \Big\langle Q^j_t, b^i(t,X^i_t,\cdot)\Big\rangle\Big)\cdot\Big(b^i(t,X^i_t,X^r_t) - \big\langle Q^r_t, b^i(t,X^i_t,\cdot)\big\rangle\Big)\Big] \\
&\quad= \mathbb{E}\Big[\Big(b^i(t,X^i_t,X^j_t) - \Big\langle Q^j_t, b^i(t,X^i_t,\cdot)\Big\rangle\Big)\cdot\Big\langle P^{r|\{i,j\}}_{t;X^{\{i,j\}}_{[t]}} - Q^r_t, b^i(t,X^i_t,\cdot)\Big\rangle\Big] \\
&\quad\le \sqrt{M}\,\mathbb{E}\Big[\Big\langle P^{r|\{i,j\}}_{t;X^{\{i,j\}}_{[t]}} - Q^r_t, b^i(t,X^i_t,\cdot)\Big\rangle^2\Big]^{1/2}.
\end{aligned}$$

Apply the assumption (2.4), followed by the data processing inequality, and the chain rule of relative entropy, we can bound the above further by

$$\sqrt{\gamma M}\,\mathbb{E}\Big[H\big(P^{r|\{i,j\}}_{t;X^{\{i,j\}}_{[t]}}\,|\,Q^r_t\big)\Big]^{1/2} \le \sqrt{\gamma M\Big(H_{[t]}(\{i,j,r\}) - H_{[t]}(\{i,j\})\Big)} \le \sqrt{\gamma M h_3}.$$

Therefore,

$$\mathrm{I(a)} \le \frac{\sqrt{\gamma M h_3}}{\sigma^2}\sum_{i,j,r\in v}\xi_{ij}\xi_{ir} = \frac{\sqrt{\gamma M h_3}}{\sigma^2}\sum_{i\in v}\Bigg(\sum_{j\in v}\xi_{ij}\Bigg)^2.$$

For I(b) we have the simple bound

$$\mathrm{I(b)} \le \frac{M}{\sigma^2}\sum_{i,j\in v}\xi_{ij}^2.$$

Put it together to complete the proof. $\square$

Remark 4.2. In Section 2.6 we mentioned the case of the reversed entropies $\overleftarrow{H}_{[t]}(v) = H(Q^v_{[t]}\,|\,P^v_{[t]})$, under the stronger assumption of bounded b. For the reversed entropies we obtain the same hierarchy (4.1), except with $C(v)$ replaced by $\widetilde{C}(v) = (M/\sigma^2)\sum_{i,j\in v}\xi_{ij}^2$. Indeed, the proof proceeds in the same manner, but with the particles X_t^i replaced throughout by the independent projection Y_t^i, which ultimately results in I=I(b) because I(a) vanishes by independence. See Remark 6.2 for the downstream implications of this.

We next give the analogous result for the time-marginal entropies $H_t(v)$, following the strategy of [48, Section 3.3]. There are many parallels with the proof of Lemma 4.1, to which we will refer repeatedly.

Lemma 4.3. *Suppose Assumption (U) holds. Let $v \subset [n]$.*

(1) For every $t \geq 0$,

$$H_t(v) - H_s(v) \leq C(v)(t-s) + \sum_{j \notin v} \mathcal{A}_{v \to j} \int_s^t \Big(H_u(vj) - H_u(v) \Big) du - \frac{\sigma^2}{4\eta} \int_s^t H_u(v) du. \tag{4.10}$$

By convention, the second-to-last term of (4.10) *is zero if $v = [n]$.*

(2) If it holds for some constant h_3 that

$$\sup_{t \geq 0} H_t(v) \leq h_3, \quad \text{for all } v \subset [n] \text{ with } |v| = 3, \tag{4.11}$$

then (4.10) *holds with C replaced by $\widehat{C}$ defined in* (4.3).

Proof. We first apply a projection argument, to express $X_t^v = (X_t^i)_{i \in v}$ as the solution of a Markovian SDE. At the level of the Fokker-Planck PDEs, this is a marginalization argument exactly like that used in deriving the BBGKY hierarchy. To parallel the previous proof, we favor a stochastic perspective, applying the mimicking theorem [16, Corollary 3.7]. First, let us define the Markovian analogue of (4.4): For any $i \in v, j \notin v$, there exists a Borel function $\widehat{b}_{ij}^v : [0,T] \times (\mathbb{R}^d)^v \to \mathbb{R}^d$ such that

$$\widehat{b}_{ij}^v(t, X_t^v) = \mathbb{E}[b^i(t, X_t^i, X_t^j) \,|\, X_t^v], \quad a.s., \ a.e. \ t > 0.$$

Then, by [16, Corollary 3.7], there exists a weak solution $\widehat{X}^v = (\widehat{X}^i)_{i \in v}$ of the Markovian analogue of the SDE (4.5),

$$d\widehat{X}_t^i = \Big(b_0^i(t, \widehat{X}_t^i) + \sum_{j \in v} \xi_{ij} b^i(t, \widehat{X}_t^i, \widehat{X}_t^j) + \sum_{j \notin v} \xi_{ij} \widehat{b}_{ij}^v(t, \widehat{X}_t^v) \Big) dt + \sigma d\widehat{W}_t^i, \quad i \in v, \tag{4.12}$$

defined on a possibly different probability space with different Brownian motions, and with the crucial property that $\widehat{X}_t^v$ has the same law as X_t^v, for each $t \geq 0$.

We next make use of a well known calculation of the time-derivative of the relative entropy between the laws of two Markovian diffusion processes. To summarize formally how this works, suppose we are given solutions of two different SDEs taking values in some Euclidean space, $dZ_t^i = a^i(t, Z_t^i)dt + \sigma dB_t$, for $i = 1, 2$. Let ρ_t^i be the law of Z_t^i. Then, using the Fokker-Planck equation satisfied by ρ^i, one has the formal computation

$$\begin{aligned} \frac{d}{dt} H(\rho_t^1 \,|\, \rho_t^2) &= \int_{\mathbb{R}} \left((a^1(t,z) - a^2(t,z)) \cdot \nabla \log \frac{d\rho_t^1}{d\rho_t^2}(z) - \frac{\sigma^2}{2} \left| \nabla \log \frac{d\rho_t^1}{d\rho_t^2}(z) \right|^2 \right) \rho_t^1(dz) \\ &\leq \frac{1}{\sigma^2} \int_{\mathbb{R}} |a^1(t,z) - a^2(t,z)|^2 \rho_t^1(dz) - \frac{\sigma^2}{4} I(\rho_t^1 \,|\, \rho_t^2). \end{aligned} \tag{4.13}$$

We refer to [48, Lemma 3.1] for a rigorous version of the integrated form of this inequality, under mild local integrability conditions on a^1 and a^2 of a technical nature. We will apply this with a^1 being the drift of $\widehat{X}^v$ as in (4.12), and with a^2 being the drift of the dynamics for Y^v which was recalled in (4.6). The technical conditions were straightforward to check in [48, Section 3.3], and they are equally straightforward here, so we omit the details. Applying the inequality (4.13) (that is, [48, Lemma 3.1]) then yields

$$H_t(v) - H_s(v) \le \frac{1}{\sigma^2}\int_s^t \sum_{i\in v}\mathbb{E}\Bigg[\bigg| \sum_{j\ne i, j\in v} \xi_{ij} b^i(u, X_u^i, X_u^j) + \sum_{j\notin v} \xi_{ij}\widehat{b}_{ij}^v(t, X_t^v) - \sum_{j\ne i} \xi_{ij}\left\langle Q_u^j, b^i(u, X_u^i, \cdot)\right\rangle \bigg|^2\Bigg] du - \frac{\sigma^2}{4}\int_s^t I(P_u^v \,|\, Q_u^v) du.$$

The expectation term is estimated exactly as in the proof of Lemma 4.1. As for the Fisher information term, Assumption (U)(4) together with tensorization of the log-Sobolev inequality [4, Proposition 5.2.7] implies that $H_u(v) \le \eta I(P_u^v \,|\, Q_u^v)$. Putting it together, we obtain (4.10). Part (2) follows by improving the estimate on I, in exactly the same manner as in the proof of Lemma 4.1(2). ☐

4.2. **Proof of Theorem 2.7.** We essentially repeat here the argument given in Section 1.4. Recall the definition of $\mathcal{A}$ from (2.8), as a matrix indexed by $2^{[n]}$, which can be viewed as an operator on $\mathbb{R}^{2^{[n]}}$ in the natural way,

$$\mathcal{A}F(v) = \sum_{u\in 2^{[n]}} \mathcal{A}(v,u)F(u) = \sum_{j\notin v} \mathcal{A}_{v\to j}(F(vj) - F(v)). \tag{4.14}$$

for $F : 2^{[n]} \to \mathbb{R}$. We may then write the inequality (4.1) in Lemma 4.1 as a pointwise inequality between functions:

$$\frac{d}{dt}H_{[t]} \le C + \mathcal{A}H_{[t]}. \tag{4.15}$$

As mentioned before, $\mathcal{A}$ is the rate matrix of a (continuous-time) Markov process, in the sense that its row sums are zero and its off-diagonal entries are nonnegative. In particular, the associated semigroup $e^{t\mathcal{A}}$ leaves invariant the set of nonnegative functions on $2^{[n]}$. Hence, by reversing time and applying (4.15), we have

$$\frac{d}{dt}\Big(e^{t\mathcal{A}}H_{[T-t]}\Big) = e^{t\mathcal{A}}\Big(\mathcal{A}H_{[T-t]} + \frac{d}{dt}H_{[T-t]}\Big) \ge -e^{t\mathcal{A}}C,$$

Integrate this to find

$$e^{T\mathcal{A}}H_{[0]} \ge H_{[T]} - \int_0^T e^{t\mathcal{A}}C\,dt.$$

Recall the probabilistic expression $e^{t\mathcal{A}}F(v) = \mathbb{E}_v[F(\mathcal{X}_t)]$ for the semigroup, where $\mathcal{X}$ is the infection process and $\mathbb{E}_v$ denotes expectation starting from $\mathcal{X}_0 = v$. Hence, rearranging the previous inequality yields (2.9).

The proof of (2.10) is similar. As a technical point, Lemma 4.3 is not exactly a differential inequality, because we do not know a priori that $t \mapsto H_t(v)$ is differentiable. If it were differentiable, we could write (4.10) in Lemma 4.3 as the following pointwise inequality between functions,

$$\frac{d}{dt}H_t \le C + \mathcal{A}H_t - \frac{\sigma^2}{4\eta}H_t.$$

Hence,

$$\frac{d}{dt}\Big(e^{-(\sigma^2/4\eta)t}e^{t\mathcal{A}}H_{T-t}\Big) = e^{-(\sigma^2/4\eta)t}e^{t\mathcal{A}}\Big(\mathcal{A}H_{T-t} - \frac{\sigma^2}{4\eta}H_{T-t} + \frac{d}{dt}H_{T-t}\Big) \ge -e^{-(\sigma^2/4\eta)t}e^{t\mathcal{A}}C,$$

which we integrate to find

$$e^{-(\sigma^2/4\eta)T}e^{T\mathcal{A}}H_0 \ge H_T - \int_0^T e^{-(\sigma^2/4\eta)t}e^{t\mathcal{A}}C\,dt.$$

In probabilistic notation, this yields (2.10). To address the issue that $t \mapsto H_t(v)$ might not be differentiable, we simply mollify, take limits easily in light of the uniform bound $\sup_{t\in[0,T]} H_t(v) \le H_{[T]}(v) < \infty$ for any $T > 0$. $\square$

As a corollary of the proof, we record here how Theorem 2.7 can modified to be in terms of $\widehat{C}$, instead of C, under the additional assumption on 3-particle entropies stated in (4.2) (and in (4.11) for the uniform-in-time case).

Corollary 4.4.

(1) If Assumption (A) holds for $T < \infty$, and also the condition (4.2) *holds, then* (2.9) *holds with $\widehat{C}$ in place of C.*
(2) If Assumption (U) holds, and also the condition (4.11) *holds, then* (2.10) *holds with $\widehat{C}$ in place of C.*

4.3. **Stochastic control perspective.** May add a short description (or subsection) on stochastic approach here.

5. Expectation estimates for the infection process

We have now completed the proof of Theorem 2.7, which bounds the entropies $H_{[t]}(v)$ and $H_t(v)$ in terms of quantities of the form $\mathbb{E}_v[F(\mathcal{X}_T)]$, with $\mathcal{X}$ being the infection process. Recall that these expectations can be expressed in terms of the semigroup of the infection process,

$$\mathbb{E}_v[F(\mathcal{X}_t)] = e^{t\mathcal{A}}F(v) = \sum_{m=0}^{\infty} \frac{t^m}{m!}\mathcal{A}^m F(v). \tag{5.1}$$

In this section we estimate the expectations for eight functions F. In Section 6, we will put these estimates to use in order to prove the theorems stated in Section 2.5.

The functions F of interest to us are those which arise from bounding C as well as $\widehat{C}$, which were defined respectively in (2.7) and (4.3). For instance, the bound on the maximum entropy in Theorem 2.8 starts by using the crude bound $\xi_{ij} \le \delta$ for all i, j:

$$C(v) = \frac{M}{\sigma^2}\sum_{i\in v}\left(\sum_{j\in v}\xi_{ij}\right)^2 \le \frac{M}{\sigma^2}\delta^2|v|^3,$$

and this leads us to try to estimate $\mathbb{E}_v|\mathcal{X}_t|^3$. Also $\mathbb{E}_v|\mathcal{X}_t|^2$ arises from a similar bound on the second term in $\widehat{C}(v)$.

To study the average entropy, our first Theorem 2.9 starts from the sharper bound

$$C(v) = \frac{M}{\sigma^2}|v|^2\sum_{i\in v}\delta_i^2 = \frac{M}{\sigma^2}|v|^2\langle 1_v, x\rangle, \qquad x = (\delta_1^2, \dots, \delta_n^2),$$

where we recall that $\delta_i = \max_j \xi_{ij}$ is the row-maximum. Throughout this section, we write 1_v to denote the n-vector with ones for the coordinates in $v \subset [n]$ and zeroes otherwise. This leads us to study the quantity $\mathbb{E}_v[|\mathcal{X}_t|^2\langle 1_{\mathcal{X}_t}, x\rangle]$, and also without the square in order to address the second term of $\widehat{C}(v)$.

Our most complex Theorem 2.10 starts from Cauchy-Schwarz inequality,

$$C(v) \le \frac{M}{\sigma^2}|v|\sum_{i,j\in v}\xi_{ij}^2 = \frac{M}{\sigma^2}|v|\langle 1_v, \widehat{\xi}1_v\rangle,$$

where we define $\widehat{\xi}_{ij} = \xi_{ij}^2$ as the entrywise (Hadamard) square of ξ. This leads us to study $\mathbb{E}_v[|\mathcal{X}_t|\langle 1_{\mathcal{X}_t}, \widehat{\xi}1_{\mathcal{X}_t}\rangle]$, and also without the $|\mathcal{X}_t|$ in order to address the second term of $\widehat{C}(v)$.

With these expressions in mind, the following theorem estimates all of the expectation estimates that we will need, and they will be put to use in Section 6. For an $n \times n$ matrix $G = (G_{ij})$, let us also introduce the notation

$$G_{\mathrm{diag}} = (G_{11}, \dots, G_{nn}). \tag{5.2}$$

That is, G_{diag} is the vector formed by the diagonal entries of G.

Theorem 5.1. *Consider a vector $x \in \mathbb{R}^n$ and an $n \times n$ matrix G, both having nonnegative entries. Then we have the following estimates, for any $t \geq 0$ and $v \subset [n]$:*

(i) Polynomial in $|\mathcal{X}_t|$:

$$\begin{aligned} &\text{(a)} \quad \mathbb{E}_v|\mathcal{X}_t| \ \leq e^{\gamma t}|v| \\ &\text{(b)} \quad \mathbb{E}_v|\mathcal{X}_t|^2 \leq 2e^{2\gamma t}|v|^2 \\ &\text{(c)} \quad \mathbb{E}_v|\mathcal{X}_t|^3 \leq 8e^{3\gamma t}|v|^3 \end{aligned}$$

(ii) Polynomial in $|\mathcal{X}_t|$ times linear function of $1_{\mathcal{X}_t}$:

$$\begin{aligned} &\text{(a)} \quad \mathbb{E}_v[\langle 1_{\mathcal{X}_t}, x\rangle] \qquad \leq \langle 1_v, e^{\gamma t\xi}x\rangle \\ &\text{(b)} \quad \mathbb{E}_v[|\mathcal{X}_t|\langle 1_{\mathcal{X}_t}, x\rangle] \ \leq |v|\langle 1_v, e^{\gamma t(I+\xi)}(I+\xi)x\rangle \\ &\text{(c)} \quad \mathbb{E}_v[|\mathcal{X}_t|^2\langle 1_{\mathcal{X}_t}, x\rangle] \leq 2|v|\big\langle 1_v, e^{\gamma t(2I+\xi)}(I+\xi)^2 1_v\big\rangle \end{aligned}$$

(iii) Polynomial in $|\mathcal{X}_t|$ times quadratic function of $1_{\mathcal{X}_t}$: Letting $G_t = e^{\gamma t\xi}Ge^{\gamma t\xi^\top}$,

$$\begin{aligned} &\text{(a)} \quad \mathbb{E}_v[\langle 1_{\mathcal{X}_t}, G1_{\mathcal{X}_t}\rangle] \qquad \leq \langle 1_v, G_t 1_v\rangle + \gamma\int_0^t \big\langle 1_v, \xi e^{\gamma(t-s)\xi}(G_s)_{\mathrm{diag}}\big\rangle\, ds \\ &\text{(b)} \quad \mathbb{E}_v[|\mathcal{X}_t|\langle 1_{\mathcal{X}_t}, G1_{\mathcal{X}_t}\rangle] \leq |v|e^{\gamma t}\langle 1_v, (\xi G_t + G_t\xi^\top + G_t)1_v\rangle \\ &\qquad\qquad + \gamma|v|e^{\gamma t}\int_0^t \langle 1_v, e^{\gamma(t-s)\xi}(I+\xi)\xi(\xi G_s + G_s\xi^\top + 2G_s)_{\mathrm{diag}}\rangle\, ds \end{aligned}$$

The rest of the section is devoted to the proof of Theorem 5.1. Our approach will start from the formula

$$\frac{d}{dt}\mathbb{E}_v[F(\mathcal{X}_t)] = \frac{d}{dt}e^{t\mathcal{A}}F(v) = e^{t\mathcal{A}}\mathcal{A}F(v) = \mathbb{E}_v[\mathcal{A}F(\mathcal{X}_t)]. \tag{5.3}$$

Then, we will try to bound $\mathcal{A}F$ from above in terms of F itself, or other functions for which we have already computed expectations, so that we obtain an estimate of $\mathbb{E}_v[F(\mathcal{X}_t)]$ using Gronwall's inequality. We will use repeatedly the basic formula

$$\mathcal{A}F(v) = \gamma\sum_{j\notin v}\Big(\sum_{i\in v}\xi_{ij}\Big)(F(vj) - F(v)), \tag{5.4}$$

which is just (4.14) combined with the definition of $\mathcal{A}_{v\to j}$ in (2.7). Moreover, a convenient abuse of notation will be to write $\mathcal{A}[F(v)]$ in place of $\mathcal{A}F(v)$. For example, $\mathcal{A}[|v|^2]$ will stand for $\mathcal{A}F(v)$, where $F(v) = |v|^2$.

5.1. **Polynomials.** In this section we prove part (i) of Theorem 5.1. We begin with a more general lemma.

Lemma 5.2. *Let $\ell \geq 1$. Then, for $v \subset [n]$,*

$$\mathcal{A}[|v|^\ell] \leq \gamma|v|\big((|v|+1)^\ell - |v|^\ell\big). \tag{5.5}$$

Proof. Let $F(v) = |v|^\ell$. We have

$$F(vj) - F(v) = (|v|+1)^\ell - |v|^\ell.$$

We then apply (5.4) and recall from Assumption (rows) that row sums of ξ are bounded by 1:

$$\mathcal{A}F(v) = \gamma \sum_{j \notin v} \Big(\sum_{i \in v} \xi_{ij} \Big) \big((|v|+1)^\ell - |v|^\ell \big) \le \gamma |v| \big((|v|+1)^\ell - |v|^\ell \big). \qquad \square$$

Using Lemma 5.2 with $\ell = 1$, we have $\mathcal{A}[|v|] \le \gamma |v|$, and thus from (5.3) we deduce

$$\frac{d}{dt} \mathbb{E}_v |\mathcal{X}_t| \le \gamma \mathbb{E}_v |\mathcal{X}_t|.$$

Since $\mathbb{E}_v|\mathcal{X}_0| = |v|$, from Gronwall we get $\mathbb{E}_v|\mathcal{X}_t| \le e^{\gamma t}|v|$, which is Theorem 5.1(ia). To prove Theorem 5.1(ib), we apply Lemma 5.2 with $\ell = 2$ to get $\mathcal{A}[|v|^2] \le \gamma|v|(2|v|+1)$, which we plug into (5.3) to find

$$\frac{d}{dt} \mathbb{E}_v |\mathcal{X}_t|^2 \le 2\gamma \mathbb{E}_v |\mathcal{X}_t|^2 + \gamma \mathbb{E}_v |\mathcal{X}_t|.$$

Using Gronwall and Theorem 5.1(ia),

$$\begin{aligned} \mathbb{E}_v |\mathcal{X}_t|^2 &\le e^{2\gamma t} \mathbb{E}_v |\mathcal{X}_0|^2 + \gamma \int_0^t e^{2\gamma(t-s)} \mathbb{E}_v |\mathcal{X}_s| \, ds \\ &\le e^{2\gamma t} |v|^2 + \gamma |v| \int_0^t e^{2\gamma(t-s)} e^{\gamma s} \, ds \\ &= e^{2\gamma t} |v|^2 + e^{2\gamma t} (1 - e^{-\gamma t}) |v|. \end{aligned}$$

This proves Theorem 5.1(ib). To prove Theorem 5.1(ic), we apply Lemma 5.2 with $\ell = 3$ to get $\mathcal{A}[|v|^3] \le \gamma|v|(3|v|^2 + 3|v| + 1)$, which we plug into (5.3) to find

$$\frac{d}{dt} \mathbb{E}_v |\mathcal{X}_t|^3 \le 3\gamma \mathbb{E}_v |\mathcal{X}_t|^3 + 3\gamma \mathbb{E}_v |\mathcal{X}_t|^2 + \gamma \mathbb{E}_v |\mathcal{X}_t|.$$

By Gronwall and parts (ia,b),

$$\begin{aligned} \mathbb{E}_v |\mathcal{X}_t|^3 &\le e^{3\gamma t} \mathbb{E}_v |\mathcal{X}_0|^3 + \gamma \int_0^t e^{3\gamma(t-s)} \big(3\mathbb{E}_v |\mathcal{X}_s|^2 + \mathbb{E}_v |\mathcal{X}_s| \big) \, ds \\ &\le e^{3\gamma t} |v|^3 + \gamma \int_0^t e^{3\gamma(t-s)} \big(6 e^{2\gamma s} |v|^2 + e^{\gamma s} |v| \big) \, ds \\ &= e^{3\gamma t} |v|^3 + e^{3\gamma t} \big(6|v|^2 (1 - e^{-\gamma t}) + \tfrac{1}{2} |v| (1 - e^{-2\gamma t}) \big). \end{aligned}$$

Discarding terms yields Theorem 5.1(ic).

Remark 5.3. In the proof of Lemma 5.2, and below, we repeatedly bound $\sum_{j \notin v}$ by $\sum_{j \in [n]}$. Our rough intuition is that this does not lose too much because we view $|v|$ as much smaller than n, so that v^c occupies most of $[n]$. From a practical standpoint, it is hard to imagine obtaining a tractable estimate without using such a bound. This is what lets us close the Gronwall loop in the subsequent argument.

5.2. **Linear functions.** In this section we prove part (ii) of Theorem 5.1, and we again begin with a lemma.

Lemma 5.4. *Let $x \in \mathbb{R}^n$ have nonnegative entries, and let $\ell \ge 0$ be an integer. Let $v \subset [n]$. For $a \ge 0$,*

$$\mathcal{A}[|v|^\ell \langle 1_v, x \rangle] \le \gamma (|v|+1)^\ell \langle 1_v, \xi x \rangle + \gamma |v| \big((|v|+1)^\ell - |v|^\ell \big) \langle 1_v, x \rangle \tag{5.6}$$

$$\tag{5.7}$$

Proof. Let $F(v) = (|v| + a)^\ell \langle 1_v, x\rangle$. We have

$$\begin{aligned} F(vj) - F(v) &= (|v|+1)^\ell \sum_{i\in vj} x_i - |v|^\ell \sum_{i\in v} x_i \\ &= (|v|+1)^\ell x_j + \big((|v|+1)^\ell - |v|^\ell\big) \sum_{i\in v} x_i \end{aligned}$$

Recalling from Assumption (rows) that the row sums of ξ are bounded by 1, we have

$$\begin{aligned} \mathcal{A}F(v) &\le \gamma \sum_{j\notin v} \left(\sum_{i\in v} \xi_{ij}\right)\left((|v|+1)^\ell x_j + \big((|v|+1)^\ell - |v|^\ell\big) \sum_{i\in v} x_i\right) \\ &\le \gamma(|v|+1)^\ell \langle 1_v, \xi x\rangle + \gamma |v| \big((|v|+1)^\ell - |v|^\ell\big)\langle 1_v, x\rangle. \end{aligned}$$

□

Let us now prove Theorem 5.1(iia). Starting from (5.3) and applying Lemma 5.4 with $\ell = 0$,

$$\frac{d}{dt}\mathbb{E}_v[\langle 1_{\mathcal{X}_t}, x\rangle] \le \gamma \mathbb{E}_v[\langle 1_{\mathcal{X}_t}, \xi x\rangle].$$

Applying this with x as basis vectors yields the coordinatewise inequality between vectors,

$$\frac{d}{dt}\mathbb{E}_v[1_{\mathcal{X}_t}] \le \gamma \xi^\top \mathbb{E}_v[1_{\mathcal{X}_t}].$$

Because ξ has nonnegative entries, so does the matrix exponential $e^{s\xi^\top}$ for any $s \ge 0$. Hence, for any $t > s > 0$, we have coordinatewise that

$$\frac{d}{ds}\big(e^{\gamma s \xi^\top}\mathbb{E}_v[1_{\mathcal{X}_{t-s}}]\big) \ge 0.$$

Integrate and rearrange to find

$$\mathbb{E}_v[1_{\mathcal{X}_t}] \le e^{\gamma t \xi^\top}\mathbb{E}_v[1_{\mathcal{X}_0}] = e^{\gamma t\xi^\top} 1_v. \tag{5.8}$$

Taking the inner product with x proves Theorem 5.1(iia).

To prove Theorem 5.1(iib), we apply Lemma 5.4 with $\ell = 1$ to get

$$\frac{d}{dt}\mathbb{E}_v[|\mathcal{X}_t|\langle 1_{\mathcal{X}_t}, x\rangle] \le \gamma \mathbb{E}_v[(|\mathcal{X}_t|+1)\langle 1_{\mathcal{X}_t}, \xi x\rangle + |\mathcal{X}_t|\langle 1_{\mathcal{X}_t}, x\rangle].$$

Applying this with x as basis vectors yields the coordinatewise inequality between vectors,

$$\frac{d}{dt}\mathbb{E}_v[|\mathcal{X}_t| 1_{\mathcal{X}_t}] \le \gamma(I + \xi^\top)\mathbb{E}_v[|\mathcal{X}_t|1_{\mathcal{X}_t}] + \gamma\xi^\top \mathbb{E}_v[1_{\mathcal{X}_t}].$$

Integrating this as above and then recalling (5.8) yields

$$\begin{aligned} \mathbb{E}_v[|\mathcal{X}_t|1_{\mathcal{X}_t}] &\le e^{\gamma t(I+\xi^\top)}\mathbb{E}_v[|\mathcal{X}_0|1_{\mathcal{X}_0}] + \gamma \int_0^t e^{\gamma(t-s)(I+\xi^\top)}\xi^\top \mathbb{E}_v[1_{\mathcal{X}_s}]\,ds \\ &\le \left(e^{\gamma t(I+\xi^\top)}|v| + \gamma\int_0^t e^{\gamma(t-s)(I+\xi^\top)}\xi^\top e^{\gamma s\xi^\top}\,ds\right)1_v \\ &= \left(e^{\gamma t(I+\xi^\top)}|v| + e^{\gamma t(I+\xi^\top)}\xi^\top(1-e^{-\gamma t})\right)1_v. \end{aligned} \tag{5.9}$$

Discarding the $1 - e^{-\gamma t}$ term and taking the inner product with x yields Theorem 5.1(iib).

Finally, to prove Theorem 5.1(iic), we apply Lemma 5.4 with $\ell = 2$ to get

$$\begin{aligned} \frac{d}{dt}\mathbb{E}_v[|\mathcal{X}_t|^2\langle 1_{\mathcal{X}_t}, x\rangle] &\le \gamma\mathbb{E}_v[(|\mathcal{X}_t|+1)^2\langle 1_{\mathcal{X}_t}, \xi x\rangle + |\mathcal{X}_t|(2|\mathcal{X}_t|+1)\langle 1_{\mathcal{X}_t}, x\rangle] \\ &= \gamma\mathbb{E}_v[|\mathcal{X}_t|^2\langle 1_{\mathcal{X}_t}, (2I+\xi)x\rangle] + \gamma\mathbb{E}_v[|\mathcal{X}_t|\langle 1_{\mathcal{X}_t}, (I+2\xi)x\rangle] + \gamma\mathbb{E}_v[\langle 1_{\mathcal{X}_t}, \xi x\rangle] \end{aligned}$$

Applying this with x as basis vectors yields the coordinatewise inequality between vectors,

$$\frac{d}{dt}\mathbb{E}_v[|\mathcal{X}_t|^2 1_{\mathcal{X}_t}] \le \gamma(2I+\xi^\top)\mathbb{E}_v[|\mathcal{X}_t|^2 1_{\mathcal{X}_t}] + \gamma(I+2\xi^\top)\mathbb{E}_v[|\mathcal{X}_t| 1_{\mathcal{X}_t}] + \gamma\xi^\top\mathbb{E}_v[1_{\mathcal{X}_t}].$$

Integrate this and plug in (5.8) and (5.9) to get

$$\begin{aligned}
\mathbb{E}_v[|\mathcal{X}_t|^2 1_{\mathcal{X}_t}] &\le e^{\gamma t(2I+\xi^\top)}\mathbb{E}_v[|\mathcal{X}_0|^2 1_{\mathcal{X}_0}] + \gamma\int_0^t e^{\gamma(t-s)(2I+\xi^\top)}\Big((I+2\xi^\top)\mathbb{E}_v[|\mathcal{X}_s|1_{\mathcal{X}_s}] + \xi^\top\mathbb{E}_v[1_{\mathcal{X}_s}]\Big)\,ds \\
&\le e^{\gamma t(2I+\xi^\top)}|v|^2 1_v + \gamma\int_0^t e^{\gamma(t-s)(2I+\xi^\top)}(I+2\xi^\top)e^{\gamma s(I+\xi^\top)}\Big(|v|+\xi^\top\Big)1_v\,ds \\
&\qquad + \gamma\int_0^t e^{\gamma(t-s)(2I+\xi^\top)}\xi^\top e^{\gamma s\xi^\top}1_v\,ds \\
&= e^{\gamma t(2I+\xi^\top)}|v|^2 1_v + (1-e^{-\gamma t})e^{\gamma t(2I+\xi^\top)}(I+2\xi^\top)(|v|I+\xi^\top)1_v \\
&\qquad + \frac{1}{2}(1-e^{-2\gamma t})e^{\gamma t(2I+\xi^\top)}\xi^\top 1_v \\
&\le e^{\gamma t(2I+\xi^\top)}\Big(|v|^2 I + (I+2\xi^\top)(|v|I+\xi^\top) + \xi^\top\Big)1_v \\
&\le |v|^2 e^{\gamma t(2I+\xi^\top)}\Big(I + (I+2\xi^\top)(I+\xi^\top) + \xi^\top\Big)1_v \\
&= 2|v|e^{\gamma t(2I+\xi^\top)}(I+\xi^\top)^2 1_v.
\end{aligned}$$

Take the inner product with x to get Theorem 5.1(iic).

5.3. **Quadratic functions.** We finally prove part (iii) of Theorem 5.1(iii), which is the most involved. We begin with a lemma estimating the action of $\mathcal{A}$ on relevant functions:

Lemma 5.5. *Let G be an $n\times n$ matrix with nonnegative entries, and let $G_{\mathrm{diag}} = (G_{ii})_{i=1,\dots,n}$ denote the vector of diagonal entries. Let $v\subset[n]$.*

(i)
$$\mathcal{A}[\langle 1_v, G1_v\rangle] \le \gamma\langle 1_v, \xi G_{\mathrm{diag}}\rangle + \gamma\langle 1_v, \big(\xi G + G\xi^\top\big)1_v\rangle. \tag{5.10}$$

(ii)
$$\mathcal{A}[|v|\langle 1_v, G1_v\rangle] \le \gamma(|v|+1)\Big[\langle 1_v, \xi G_{\mathrm{diag}}\rangle + \langle 1_v, \big(\xi G + G\xi^\top\big)1_v\rangle\Big] + \gamma|v|\langle 1_v, G1_v\rangle. \tag{5.11}$$

Proof.

(i) Let $F(v) = \langle 1_v, G1_v\rangle = \sum_{i,r\in v} G_{ir}$. For $j\notin v$ we compute

$$F(vj) - F(v) = \sum_{i,r\in vj} G_{ir} - \sum_{i,r\in v} G_{ir} = G_{jj} + \sum_{r\in v}(G_{rj}+G_{jr}).$$

Thus, using (5.4) and the nonnegativity of the entries of ξ,

$$\begin{aligned}
\mathcal{A}F(v) &= \gamma\sum_{j\notin v}\Big(\sum_{i\in v}\xi_{ij}\Big)\Big(G_{jj} + \sum_{r\in v}(G_{rj}+G_{jr})\Big) \\
&\le \gamma\sum_{i\in v}\sum_{j=1}^n \xi_{ij}G_{jj} + \gamma\sum_{i,r\in v}\sum_{j=1}^n(\xi_{ij}G_{rj} + \xi_{ij}G_{jr}) \\
&= \gamma\langle 1_v, \xi G_{\mathrm{diag}}\rangle + \gamma\langle 1_v, \big(\xi G + G\xi^\top\big)1_v\rangle.
\end{aligned}$$

(ii) Set $F(v) = |v|\langle 1_v, G1_v\rangle$. For $j \notin v$ we compute

$$F(vj) - F(v) = (|v|+1)\sum_{\ell,r\in vj} G_{\ell r} - |v|\sum_{\ell,r\in v} G_{\ell r}$$
$$= |v|\Big(G_{jj} + \sum_{r\in v}(G_{rj}+G_{jr})\Big) + \sum_{\ell,r\in vj} G_{\ell r}.$$

Thus

$$\begin{aligned}\mathcal{A}F(v) &= \gamma\sum_{j\notin v}\Big(\sum_{i\in v}\xi_{ij}\Big)\Big(|v|\Big(G_{jj}+\sum_{r\in v}(G_{rj}+G_{jr})\Big)+\sum_{\ell,r\in vj}G_{\ell r}\Big)\\ &\le \gamma|v|\sum_{i\in v}\sum_{j=1}^n \xi_{ij}G_{jj} + \gamma|v|\sum_{i,r\in v}\sum_{j=1}^n\Big(\xi_{ij}G_{rj}+\xi_{ij}G_{jr}\Big) + \gamma\sum_{i\in v}\sum_{j\notin v}\xi_{ij}\sum_{\ell,r\in vj}G_{\ell r}\\ &= \gamma|v|\langle 1_v, \xi G_{\mathrm{diag}}\rangle + \gamma|v|\langle 1_v, \big(\xi G + G\xi^\top\big)1_v\rangle + \gamma\sum_{i\in v}\sum_{j\notin v}\xi_{ij}\sum_{\ell,r\in vj}G_{\ell r}.\end{aligned}$$

For the last term,

$$\begin{aligned}\sum_{i\in v}\sum_{j\notin v}\xi_{ij}\sum_{\ell,r\in vj}G_{\ell r} &= \sum_{i\in v}\sum_{\ell,r=1}^n\sum_{j\notin v}1_{\ell\in vj}1_{r\in vj}\xi_{ij}G_{\ell r}\\ &= \sum_{i\in v}\sum_{\ell,r=1}^n\sum_{j\notin v}(1_{\ell,r\in v}+1_{\ell\in v,r=j}+1_{\ell=j,r\in v}+1_{\ell=r=j})\xi_{ij}G_{\ell r}\\ &\le \sum_{i\in v}\Bigg(\sum_{\ell,r\in v}G_{\ell r}+\sum_{\ell\in v}\sum_{j=1}^n\xi_{ij}G_{\ell j}+\sum_{r\in v}\sum_{j=1}^n\xi_{ij}G_{jr}+\sum_{j=1}^n\xi_{ij}G_{jj}\Bigg)\\ &= |v|\langle 1_v, G1_v\rangle + \langle 1_v, G\xi^\top 1_v\rangle + \langle 1_v, \xi G1_v\rangle + \langle 1_v, \xi G_{\mathrm{diag}}\rangle,\end{aligned}$$

where we used the row sum condition in Assumption (rows) in the inequality step. Combining this with the previous inequality yields

$$\mathcal{A}F(v) \le \gamma(|v|+1)\langle 1_v, \xi G_{\mathrm{diag}}\rangle + \gamma(|v|+1)\langle 1_v, \big(\xi G + G\xi^\top\big)1_v\rangle + \gamma|v|\langle 1_v, G1_v\rangle. \qquad \square$$

Let us now prove Theorem 5.1(iiia). Starting from (5.3) and applying Lemma 5.5(i),

$$\frac{d}{dt}\mathbb{E}_v[\langle 1_{\mathcal{X}_t}, G1_{\mathcal{X}_t}\rangle] \le \gamma\mathbb{E}_v\big[\langle 1_{\mathcal{X}_t}, \xi G_{\mathrm{diag}}\rangle + \langle 1_{\mathcal{X}_t}, (\xi G + G\xi^\top)1_{\mathcal{X}_t}\rangle\big]. \tag{5.12}$$

Let A_t be the $n\times n$ diagonal matrix given by

$$(A_t)_{ij} = \mathbb{E}_v\Bigg[\sum_{r\in\mathcal{X}_t}\xi_{ri}\Bigg]1_{i=j},$$

which is defined in this way so that

$$\langle A_t, G\rangle = \sum_{i=1}^n \mathbb{E}_v\Bigg[\sum_{r\in\mathcal{X}_t}\xi_{ri}G_{ii}\Bigg] = \mathbb{E}_v\big[\langle 1_{\mathcal{X}_t}, \xi G_{\mathrm{diag}}\rangle\big].$$

Defining the symmetric matrix $R_t = \mathbb{E}_v[1_{\mathcal{X}_t}1_{\mathcal{X}_t}^\top]$, we may write (5.12) in duality as

$$\frac{d}{dt}\langle R_t, G\rangle \le \gamma\langle A_t, G\rangle + \gamma\langle R_t\xi + \xi^\top R_t, G\rangle.$$

This holds for every matrix G with nonnegative entries, and we deduce the coordinatewise inequality

$$\frac{d}{dt}R_t \le \gamma\big(A_t + R_t\xi + \xi^\top R_t\big).$$

Because $e^{\gamma s\xi}$ has nonnegative entries, for each $t \geq s \geq 0$ we deduce

$$\frac{d}{ds}\big(e^{\gamma s\xi^\top} R_{t-s} e^{\gamma s\xi}\big) \geq -\gamma e^{\gamma s\xi^\top} A_{t-s} e^{\gamma s\xi}.$$

Integrate to find

$$R_t \leq e^{\gamma t\xi^\top} R_0 e^{\gamma t\xi} + \gamma \int_0^t e^{\gamma(t-s)\xi^\top} A_s e^{\gamma(t-s)\xi}\, ds. \tag{5.13}$$

We next take the inner product on both sides with a matrix G with nonnegative entries. Note that $R_0 = 1_v 1_v^\top$, so that

$$\langle e^{\gamma t\xi^\top} R_0 e^{\gamma t\xi}, G\rangle = \mathrm{Tr}\big(G e^{\gamma t\xi^\top} 1_v 1_v^\top e^{\gamma t\xi}\big) = \langle 1_v, G_t 1_v\rangle,$$

where we recall that $G_t = e^{\gamma t\xi} G e^{\gamma t\xi^\top}$. Recalling the definition of A_s, and

$$\langle e^{\gamma(t-s)\xi^\top} A_s e^{\gamma(t-s)\xi}, G\rangle = \langle A_s, e^{\gamma(t-s)\xi} G e^{\gamma(t-s)\xi^\top}\rangle = \mathbb{E}_v\big[\langle 1_{\mathcal{X}_s}, \xi(G_{t-s})_{\mathrm{diag}}\rangle\big].$$

Hence, if we multiply (5.13) by G and use Theorem 5.1(iia), we get

$$\begin{aligned}\mathbb{E}_v[\langle 1_{\mathcal{X}_t}, G 1_{\mathcal{X}_t}\rangle] = \langle R_t, G\rangle &\leq \langle 1_v, G_t 1_v\rangle + \gamma \int_0^t \mathbb{E}_v\big[\langle 1_{\mathcal{X}_s}, \xi(G_{t-s})_{\mathrm{diag}}\rangle\big]\, ds \\ &\leq \langle 1_v, G_t 1_v\rangle + \gamma \int_0^t \langle 1_v, e^{\gamma s\xi}\xi(G_{t-s})_{\mathrm{diag}}\rangle\, ds.\end{aligned}$$

We finally turn to Theorem 5.1(iiib). We adopt a similar strategy. Starting from (5.3) and applying Lemma 5.5(ii),

$$\begin{aligned}\frac{d}{dt}\mathbb{E}_v[|\mathcal{X}_t|\langle 1_{\mathcal{X}_t}, G 1_{\mathcal{X}_t}\rangle] \leq\ & \gamma \mathbb{E}_v\big[(|\mathcal{X}_t|+1)\big(\langle 1_{\mathcal{X}_t}, \xi G_{\mathrm{diag}}\rangle + \langle 1_{\mathcal{X}_t}, (\xi G + G\xi^\top) 1_{\mathcal{X}_t}\rangle\big)\big] \\ & + \gamma \mathbb{E}_v[|\mathcal{X}_t|\langle 1_{\mathcal{X}_t}, G 1_{\mathcal{X}_t}\rangle].\end{aligned} \tag{5.14}$$

We will translate this into a coordinatewise differential inequality for the matrix $\widetilde{R}_t = \mathbb{E}_v[|\mathcal{X}_t| 1_{\mathcal{X}_t} 1_{\mathcal{X}_t}^\top]$. Define also $R_t = \mathbb{E}_v[1_{\mathcal{X}_t} 1_{\mathcal{X}_t}^\top]$ as above, and define a diagonal matrix $\widetilde{A}_t$ by

$$(\widetilde{A}_t)_{ij} = \mathbb{E}_v\Big[(|\mathcal{X}_t|+1)\sum_{r\in\mathcal{X}_t} \xi_{ri}\Big] 1_{i=j},$$

so that

$$\langle \widetilde{A}_t, G\rangle = \mathbb{E}_v\big[(|\mathcal{X}_t|+1)\langle 1_{\mathcal{X}_t}, \xi G_{\mathrm{diag}}\rangle\big].$$

With these definitions, we can write (5.14) as

$$\frac{d}{dt}\langle \widetilde{R}_t, G\rangle \leq \gamma\langle \widetilde{A}_t, G\rangle + \gamma\langle R_t\xi + \xi^\top R_t, G\rangle + \gamma\langle \widetilde{R}_t\xi + \xi^\top \widetilde{R}_t + \widetilde{R}_t, G\rangle,$$

which means coordinatewise that

$$\frac{d}{dt}\widetilde{R}_t \leq \gamma \widetilde{A}_t + \gamma(R_t\xi + \xi^\top R_t) + \gamma(\widetilde{R}_t\xi + \xi^\top \widetilde{R}_t + \widetilde{R}_t).$$

We may integrate this as in (5.13)

$$\widetilde{R}_t \leq e^{\gamma t} e^{\gamma t\xi^\top} \widetilde{R}_0 e^{\gamma t\xi} + \gamma \int_0^t e^{\gamma(t-s)} e^{\gamma(t-s)\xi^\top}\big(\widetilde{A}_s + R_s\xi + \xi^\top R_s\big) e^{\gamma(t-s)\xi}\, ds.$$

Using $\widetilde{R}_0 = |v| 1_v 1_v^\top$, we take the inner product with G to get

$$\begin{aligned}
\mathbb{E}_v[|\mathcal{X}_t|\langle 1_{\mathcal{X}_t}, G 1_{\mathcal{X}_t}\rangle] &= \langle \widetilde{R}_t, G\rangle \\
&\le e^{\gamma t}|v|\langle e^{\gamma t\xi^\top} 1_v 1_v^\top e^{\gamma t\xi}, G\rangle + \gamma\int_0^t e^{\gamma(t-s)}\big\langle e^{\gamma(t-s)\xi^\top}\big(\widetilde{A}_s + R_s\xi + \xi^\top R_s\big)e^{\gamma(t-s)\xi}, G\big\rangle\, ds \\
&= e^{\gamma t}|v|\langle 1_v, G_t 1_v\rangle + \gamma\int_0^t e^{\gamma(t-s)}\big\langle \widetilde{A}_s + R_s\xi + \xi^\top R_s, G_{t-s}\big\rangle\, ds.
\end{aligned}$$

Using the definition of $\widetilde{A}_s$ and Theorem 5.1(iib) we have

$$\begin{aligned}
\langle \widetilde{A}_s, G_{t-s}\rangle &= \mathbb{E}_v\big[(|\mathcal{X}_s|+1)\langle 1_{\mathcal{X}_s}, \xi(G_{t-s})_{\mathrm{diag}}\rangle\big] \\
&\le 2\mathbb{E}_v\big[(|\mathcal{X}_s|)\langle 1_{\mathcal{X}_s}, \xi(G_{t-s})_{\mathrm{diag}}\rangle\big] \\
&\le 2|v|\langle 1_v, e^{\gamma s(I+\xi)}(I+\xi)\xi(G_{t-s})_{\mathrm{diag}}\rangle.
\end{aligned}$$

Using the definition of R and Theorem 5.1(iiia) we have

$$\begin{aligned}
\langle R_s\xi + \xi^\top R_s, G_{t-s}\rangle &= \langle R_s, \xi G_{t-s} + G_{t-s}\xi^\top\rangle \\
&\le \langle 1_v, (\xi G_{t-s} + G_{t-s}\xi^\top)_s 1_v\rangle + \gamma\int_0^s \langle 1_v, e^{\gamma u\xi}\xi((\xi G_{t-s} + G_{t-s}\xi^\top)_{s-u})_{\mathrm{diag}}\rangle\, du \\
&= \langle 1_v, (\xi G_t + G_t\xi^\top) 1_v\rangle + \gamma\int_0^s \langle 1_v, e^{\gamma u\xi}\xi(\xi G_{t-u} + G_{t-u}\xi^\top)_{\mathrm{diag}}\rangle\, du,
\end{aligned}$$

where we used the simple identity $(\xi G_{t-s})_s = \xi(G_{t-s})_s = \xi G_t$. Putting it together,

$$\begin{aligned}
\mathbb{E}_v[|\mathcal{X}_t|\langle 1_{\mathcal{X}_t}, G 1_{\mathcal{X}_t}\rangle] \le\ & e^{\gamma t}|v|\langle 1_v, G_t 1_v\rangle + 2\gamma|v|\int_0^t e^{\gamma(t-s)}\langle 1_v, e^{\gamma s(I+\xi)}(I+\xi)\xi(G_{t-s})_{\mathrm{diag}}\rangle\, ds \\
&+ \gamma\int_0^t e^{\gamma(t-s)}\langle 1_v, (\xi G_t + G_t\xi^\top)1_v\rangle\, ds \\
&+ \gamma^2\int_0^t e^{\gamma(t-s)}\int_0^s \langle 1_v, e^{\gamma u\xi}\xi(\xi G_{t-u} + G_{t-u}\xi^\top)_{\mathrm{diag}}\rangle\, du\Big)\, ds.
\end{aligned} \tag{5.15}$$

The third term in (5.15) is

$$(e^{\gamma t} - 1)\langle 1_v, (\xi G_t + G_t\xi^\top)1_v\rangle.$$

The second term in (5.15) is

$$2\gamma e^{\gamma t}|v|\int_0^t \langle 1_v, e^{\gamma s\xi}(I+\xi)\xi(G_{t-s})_{\mathrm{diag}}\rangle\, ds.$$

The fourth term in (5.15), after interchanging du and ds, is

$$\begin{aligned}
&\gamma\int_0^t (e^{\gamma(t-u)} - 1)\langle 1_v, e^{\gamma u\xi}\xi(\xi G_{t-u} + G_{t-u}\xi^\top)_{\mathrm{diag}}\rangle\, du \\
&\quad\le \gamma\int_0^t (e^{\gamma(t-u)} - 1)\langle 1_v, e^{\gamma u\xi}(I+\xi)\xi(\xi G_{t-u} + G_{t-u}\xi^\top)_{\mathrm{diag}}\rangle\, du,
\end{aligned}$$

where the last step used nonnegativity of the entries of ξ. Discarding the -1 terms, these somewhat crude bounds let us combine the first and third terms in (5.15), as well as the second and fourth, to get

$$\begin{aligned}
\mathbb{E}_v[|\mathcal{X}_t|\langle 1_{\mathcal{X}_t}, G 1_{\mathcal{X}_t}\rangle] \le\ & |v|e^{\gamma t}\langle 1_v, (\xi G_t + G_t\xi^\top + G_t)1_v\rangle \\
&+ \gamma|v|e^{\gamma t}\int_0^t \langle 1_v, e^{\gamma s\xi}(I+\xi)\xi(\xi G_{t-s} + G_{t-s}\xi^\top + 2G_{t-s})_{\mathrm{diag}}\rangle\, ds.
\end{aligned}$$

6. Proof of the main theorems

This section is devoted to the proofs of the theorems in Section 2.5. We begin by using Theorem 2.7 to prove a lemma that explain how to get a bound on the quantity called h_3 in Lemmas 4.1 and 4.3.

Lemma 6.1. *Suppose there exists $C_0 > 0$ such that $H_0(v) \le C_0\delta^2|v|^3$ for all $v \subset [n]$.*

(1) If Assumption (A) holds for $T < \infty$, then

$$H_{[T]}(v) \lesssim \delta^2, \quad \text{for all } v \subset [n] \text{ with } |v| = 3.$$

where the hidden constant depends only on $(T, C_0, \gamma, M, \sigma^2)$.

(2) If Assumption (U) holds, then

$$\sup_{t\ge 0} H_t(v) \lesssim \delta^2, \quad \text{for all } v \subset [n] \text{ with } |v| = 3.$$

where the hidden constant depends only on $(\eta, C_0, \gamma, M, \sigma^2)$.

Proof.

(1) We begin with the trivial inequality

$$C(v) = \frac{M}{\sigma^2}\sum_{i\in v}\left(\sum_{j\in v}\xi_{ij}\right)^2 \le \frac{M\delta^2}{\sigma^2}|v|^3,$$

where we recall that $\delta = \max_{ij}\xi_{ij}$. Recall from assumption of this lemma that the initial conditions obey $H_0(v) \le C_0|v|^3\delta^2$ for all $v \subset [n]$. Applying Theorem 2.7(i), we have

$$H_{[T]}(v) \le \mathbb{E}_v[H_0(\mathcal{X}_T)] + \int_0^T \mathbb{E}_v[C(\mathcal{X}_t)]\,dt \le C_0\delta^2\mathbb{E}_v|\mathcal{X}_T|^3 + \frac{M\delta^2}{\sigma^2}\int_0^T \mathbb{E}_v|\mathcal{X}_t|^3\,dt.$$

Using Theorem 5.1(ic), we get

$$H_{[T]}(v) \le 8e^{3\gamma T}(C_0 + \frac{M}{3\gamma\sigma^2})\delta^2|v|^3.$$

(2) We proceed exactly as for (1), but using part (ii) of Theorem 2.7 instead of part (i). This yields, with $r = \sigma^2/4\eta$,

$$\begin{aligned} H_T(v) &\le e^{-rT}\mathbb{E}_v[H_0(\mathcal{X}_T)] + \int_0^T e^{-rt}\mathbb{E}_v[C(\mathcal{X}_t)]\,dt \\ &\le C_0\delta^2 e^{-rT}\mathbb{E}_v|\mathcal{X}_T|^3 + \frac{M\delta^2}{\sigma^2}\int_0^T e^{-rt}\mathbb{E}_v|\mathcal{X}_t|^3\,dt \\ &\le 8C_0 e^{(3\gamma-r)T}\delta^2|v|^3 + \frac{8M\delta^2}{\sigma^2}|v|^3\int_0^T e^{(3\gamma-r)t}\,dt \end{aligned}$$

The claim follows because $r > 3\gamma$ by Assumption (U)(iii). □

As much as possible, we will unify the proofs of the estimates on $H_{[T]}(v)$ and on $\sup_{t\ge0} H_t(v)$, with the understanding that, in the uniform-in-time case, Assumption (U) should be imposed instead of Assumption (A), and all hidden constants must be independent of T as well as $(\eta, C_0, \gamma, M, \sigma^2)$.

Let us record a few immediate consequences of Lemma 6.1. Recall here the definition of $\widehat{C}$ from (4.3),

$$\widehat{C}(v) = \frac{\sqrt{\gamma M h_3}}{\sigma^2}\sum_{i\in v}\left(\sum_{j\in v}\xi_{ij}\right)^2 + \frac{M}{\sigma^2}\sum_{i,j\in v}\xi_{ij}^2$$

Here h_3 was a constant bounding the 3-particle entropies, which by Lemma 6.1 can be taken to be $h_3 \lesssim \delta^2$. Hence, we may write

$$\widehat{C}(v) \lesssim \delta \sum_{i \in v} \left(\sum_{j \in v} \xi_{ij} \right)^2 + \sum_{i,j \in v} \xi_{ij}^2. \tag{6.1}$$

where the hidden constant could depend on T if we are using Lemma 6.1(i) or does not depend on T if Lemma 6.1(ii) is used. As a consequence of Lemma 6.1, we may apply Corollary 4.4 to get the following two bounds, which along with (6.1) will be the starting points for the remaining proofs:

- If Assumption (A) holds for $T < \infty$, then

$$H_{[T]}(v) \le \mathbb{E}_v[H_0(\mathcal{X}_T)] + \int_0^T \mathbb{E}_v[\widehat{C}(\mathcal{X}_t)]\, dt. \tag{6.2}$$

- If Assumption (U) holds, then, with $r = \sigma^2/4\eta$,

$$H_T(v) \le e^{-rT}\mathbb{E}_v[H_0(\mathcal{X}_T)] + \int_0^T e^{-rt}\mathbb{E}_v[\widehat{C}(\mathcal{X}_t)]\, dt. \tag{6.3}$$

Remark 6.2. In the case of reversed entropy discussed in Section 2.6, if we apply the previous Remark 4.2 in place of Corollary 4.4, we find that $\overleftarrow{H}_{[T]}(v)$ obeys the same inequality (6.2) except with $\widehat{C}(\cdot)$ sharpened to $\widetilde{C}(v) = (M/\sigma^2)\sum_{i,j\in v} \xi_{ij}^2$. In fact, there is no need for an estimate of the three-particle entropies, and this is the sense in which the case of reversed entropy is easier. By the Cauchy-Schwarz inequality, we have $\widehat{C}(v) \lesssim (\delta|v|+1)\widetilde{C}(v)$, which explains the claim made in Section 2.6 that the reversed entropy bounds save a factor of $(\delta|v|+1)$ compared to the theorems of Section 2.5.

Remark 6.3. We will make repeated use of the fact that $\|\xi\|_{\mathrm{op}} \le 1$. This follows from the Gershgorin circle theorem, after recalling the assumptions that ξ has nonnegative entries, zero diagonal entries, and row sums bounded by 1.

6.1. **Maximum entropy: Proof of Theorem 2.8.** We now prove Theorem 2.8, first proving (2.13). To do this, we use a simple upper bound for (6.1):

$$\widehat{C}(v) \lesssim \delta^3|v|^3 + \delta^2|v|^2.$$

Combining this with (6.2), and the assumption $H_0(v) \lesssim \delta^2|v|^2 + \delta^3|v|^3$, we get

$$H_{[T]}(v) \lesssim \delta^2\mathbb{E}_v|\mathcal{X}_T|^2 + \delta^3\mathbb{E}_v|\mathcal{X}_T|^3 + \int_0^T \left(\delta^3\mathbb{E}_v|\mathcal{X}_t|^3 + \delta^2\mathbb{E}_v|\mathcal{X}_t|^2\right) dt.$$

By Theorem 5.1(i), ignoring the time-dependent constants, we have $\mathbb{E}_v|\mathcal{X}_t|^p \lesssim |v|^p$ for $p = 2, 3$. This yields

$$H_{[T]}(v) \lesssim \delta^2|v|^2 + \delta^3|v|^3,$$

exactly as claimed in (2.13).

To prove the claimed uniform-in-time estimates of Theorem 2.8, we must be more careful and take into account the time-dependence of the estimates of $\mathbb{E}_v|\mathcal{X}_t|^p$. Using (6.3), we have

$$H_T(v) \lesssim e^{-rT}\delta^2\mathbb{E}_v|\mathcal{X}_T|^2 + e^{-rT}\delta^3\mathbb{E}_v|\mathcal{X}_T|^3 + \int_0^T e^{-rt}\left(\delta^3\mathbb{E}_v|\mathcal{X}_t|^3 + \delta^2\mathbb{E}_v|\mathcal{X}_t|^2\right) dt.$$

Using Theorem 5.1(i),

$$H_T(v) \lesssim \delta^2 e^{(2\gamma - r)T}|v|^2 + \delta^3 e^{(3\gamma-r)T}|v|^3 + \int_0^T \left(\delta^3|v|^3 e^{(3\gamma-r)t} + \delta^2|v|^2 e^{(2\gamma-r)t}\right) dt.$$

This is again $\lesssim \delta^2|v|^2 + \delta^3|v|^3$ as long as $3\gamma < r$, which is true by assumption (U)(iii). □

6.2. **Average entropy: Proof of Theorem 2.9.** Here we prove Theorem 2.9. Note that the assumption (2.14) on the initial conditions clearly implies $H_0(v) \lesssim \delta^3|v|^3 \le \delta^2|v|^3$ for all $v \subset [n]$, which allows us to apply Lemma 6.1 and its consequences outlined at the beginning of the section. Recall the notation $\delta_i = \max_j \xi_{ij}$ for the row-maximum. We begin by bounding (6.1) by

$$\widehat{C}(v) \lesssim \delta|v|^2 \sum_{i\in v} \delta_i^2 + |v| \sum_{i \in v} \delta_i^2 = (\delta|v|^2 + |v|)\langle 1_v, x\rangle,$$

where $x = (\delta_1^2, \dots, \delta_n^2)$. Then, using (6.2), we have

$$\begin{aligned} H_{[T]}(v) &\le \mathbb{E}_v[H_0(\mathcal{X}_T)] + \int_0^T \mathbb{E}_v[\widehat{C}(\mathcal{X}_t)]\,dt \\ &\lesssim \mathbb{E}_v\Big[(\delta|\mathcal{X}_T|^3 + |\mathcal{X}_T|^2)\Big]\left(\frac{1}{n}\sum_{i=1}^n \delta_i^4\right)^{1/2} + \int_0^T \mathbb{E}_v[(\delta|\mathcal{X}_t|^2 + |\mathcal{X}_t|)\langle 1_{\mathcal{X}_t}, x\rangle]\,dt, \end{aligned}$$

where we used also the assumption (2.14) on the initial conditions. To control the first term, we apply parts (ib) and (ic) of Theorem 5.1:

$$\mathbb{E}_v|\mathcal{X}_T|^3 \le 8e^{3\gamma T}|v|^3, \qquad \mathbb{E}_v|\mathcal{X}_T|^2 \le 2e^{2\gamma T}|v|^2.$$

To control the second term, we use parts (iib) and (iic) of Theorem 5.1 to get

$$\mathbb{E}_v[|\mathcal{X}_t|\langle 1_{\mathcal{X}_t}, x\rangle] \le |v|\big\langle 1_v, e^{\gamma t(I+\xi)}(I+\xi)x\big\rangle \tag{6.4}$$

$$\mathbb{E}_v[|\mathcal{X}_t|^2\langle 1_{\mathcal{X}_t}, x\rangle] \le 2|v|^2\big\langle 1_v, e^{\gamma t(2I+\xi)}(I+\xi)^2x\big\rangle. \tag{6.5}$$

At this point, we average over all choices of $v \subset [n]$ of cardinality k, which we denote by $\operatorname{avg}_{|v|=k}$. The principle is that for any vector $y \in \mathbb{R}^n$, we have

$$\operatorname*{avg}_{|v|=k} \langle 1_v, y\rangle = \operatorname*{avg}_{|v|=k} \sum_{i=1}^n y_i 1_{i\in v} = \sum_{i=1}^n y_i \operatorname*{avg}_{|v|=k} 1_{i \in v} = \frac{k}{n}\sum_{i=1}^n y_i = \frac{k}{n}\langle 1, y\rangle. \tag{6.6}$$

Indeed, the identity $\operatorname{avg}_{|v|=k} 1_{i\in v} = k/n$ is simply saying that the probability of a fixed $i \in [n]$ belonging to a uniformly random set $v \subset [n]$ of size k is k/n. Apply (6.6) to (6.4) to get

$$\operatorname*{avg}_{|v|=k} \mathbb{E}_v[|\mathcal{X}_t|\langle 1_{\mathcal{X}_t}, x\rangle] \le \frac{k^2}{n}\big\langle 1, e^{\gamma t(I+\xi)}(I+\xi)x\big\rangle, \tag{6.7}$$

where 1 is the all-ones vector. Recall from Remark 6.3 that the row-sum bound implies $\|\xi\|_{\mathrm{op}} \le 1$. Since $|1| = \sqrt{n}$,

$$\operatorname*{avg}_{|v|=k} \mathbb{E}_v[|\mathcal{X}_t|\langle 1_{\mathcal{X}_t}, x\rangle] \le 2\frac{k^2}{\sqrt{n}}e^{2\gamma t}|x| = 2k^2e^{2\gamma t}\left(\frac{1}{n}\sum_{i=1}^n \delta_i^4\right)^{1/2},$$

where the last step plugged in the definition of x. Similarly, (6.5) averages to

$$\operatorname*{avg}_{|v|=k} \mathbb{E}_v[|\mathcal{X}_t|^2\langle 1_{\mathcal{X}_t}, x\rangle] \le 2\frac{k^3}{n}\big\langle 1, e^{\gamma t(2I+\xi)}(I+\xi)^2x\big\rangle \le 8k^3e^{3\gamma t}\left(\frac{1}{n}\sum_{i=1}^n \delta_i^4\right)^{1/2}. \tag{6.8}$$

This completes the proof of the first claim (2.15) of Theorem 2.9.

To prove the uniform-in-time claim, we make minor modifications: Use (6.3) in place of (6.2) to get

$$H_T(v) \lesssim e^{-rT}\mathbb{E}_v\Big[(\delta|\mathcal{X}_T|^3 + |\mathcal{X}_T|^2)\Big]\frac{1}{n}\sum_{i=1}^n \delta_i^2 + \int_0^T e^{-rt}\mathbb{E}_v[(\delta|\mathcal{X}_t|^2 + |\mathcal{X}_t|)\langle 1_{\mathcal{X}_t}, x\rangle]\,dt,$$

for $r = \sigma^2/4\eta$. The largest exponent in the above estimates was $3\gamma t$, which appeared in (6.8). Hence, as long as $r > 3\gamma$, as in assumption (U)(iii), we get the same bound for $\sup_{T>0} \overline{H}_T^k$.

Finally, we prove the last claim, that $\big(\frac{1}{n}\sum_{i=1}^n \delta_i^4\big)^{1/2}$ can be improved to $\frac{1}{n}\sum_{i=1}^n \delta_i^2$ when assumption (columns) holds. The column sum bound rewrites in vector form as the coordinatewise inequality $1^\top \xi \le 1^\top$, which we use instead of using ℓ_2-norms in (6.7). This yields

$$\operatorname*{avg}_{|v|=k} \mathbb{E}_v[|\mathcal{X}_t|\langle 1_{\mathcal{X}_t}, x\rangle] \le 2\frac{k^2}{n}e^{2\gamma t}\langle 1, x\rangle = 2e^{2\gamma t}\frac{k^2}{n}\sum_{i=1}^n \delta_i^2.$$

Argue similarly for the term (6.5). □

6.3. **Average entropy, without row-max: Proof of Theorem 2.10.** Here we prove Theorem 2.10, starting with the bound on $H_{[T]}(v)$ claimed in (2.18). Using the assumption that the row and column sums of ξ are bounded by 1, it is easy to see from the definition that $p_\xi \le 6\delta^2 n$. The assumption (2.17) on the initial condition then implies

$$\begin{aligned} H_0(v) &\lesssim \frac{\delta|v|^3 + |v|^2}{n^2}\sum_{i,j=1}^n \xi_{ij}^2 + \frac{\delta|v|^2 + |v|}{n}p_\xi \\ &\le \frac{\delta|v|^3 + |v|^2}{n^2}\delta^2 n^2 + \frac{\delta|v|^2 + |v|}{n}6\delta^2 n \lesssim \delta^2|v|^3. \end{aligned}$$

This lets us apply Lemma 6.1 along with its consequences described at the beginning of the section. Applying (6.1) and bounding the first term therein using Jensen's inequality, we have

$$\widehat{C}(v) \lesssim \delta|v|\sum_{i,j\in v}\xi_{ij}^2 + \sum_{i,j\in v}\xi_{ij}^2 = (\delta|v|+1)\langle 1_v, \widehat{\xi} 1_v\rangle,$$

where we recall that $\widehat{\xi}_{ij} = \xi_{ij}^2$ is the entrywise (Hadamard) square of ξ. Now, by Lemma 6.1, we may apply (6.2) along with the assumed bound on $H_0(v)$ to get

$$\begin{aligned} H_{[T]}(v) \lesssim\ & \mathbb{E}_v\Big[\delta|\mathcal{X}_T|^3 + |\mathcal{X}_T|^2\frac{1}{n^2}\sum_{i,j=1}^n \xi_{ij}^2 + \delta|\mathcal{X}_T|^2 + |\mathcal{X}_T|\frac{p_\xi}{n}\Big] \\ &+ \int_0^T \mathbb{E}_v\big[(\delta|\mathcal{X}_t|+1)\langle 1_{\mathcal{X}_t}, \widehat{\xi} 1_{\mathcal{X}_t}\rangle\big]\,dt. \end{aligned} \tag{6.9}$$

We next apply Theorem 5.1 to estimate each term. The first expectation is straightforward to bound using Theorem 5.1(i):

$$\mathbb{E}_v\Big[\big(\delta|\mathcal{X}_T|^3 + |\mathcal{X}_T|^2\big)\frac{1}{n^2}\sum_{i,j=1}^n \xi_{ij}^2 + \big(\delta|\mathcal{X}_T|^2 + |\mathcal{X}_T|\big)\frac{p_\xi}{n}\Big] \lesssim \frac{\delta|v|^3 + |v|^2}{n^2}\sum_{i,j=1}^n \xi_{ij}^2 + \frac{\delta|v|^2 + |v|}{n}p_\xi.$$

In particular, averaging over v of size k,

$$\operatorname*{avg}_{|v|=k} \mathbb{E}_v\Big[\big(\delta|\mathcal{X}_T|^3 + |\mathcal{X}_T|^2\big)\frac{1}{n^2}\sum_{i,j=1}^n \xi_{ij}^2 + \big(\delta|\mathcal{X}_T|^2 + |\mathcal{X}_T|\big)\frac{p_\xi}{n}\Big] \lesssim (\delta k + 1)\bigg(\frac{k^2}{n^2}\sum_{i,j=1}^n \xi_{ij}^2 + \frac{k}{n}p_\xi\bigg). \tag{6.10}$$

We next bound the second expectation in (6.9), by applying Theorem 5.1(iii). We do this in two steps.

Step 1. We first show that

$$\operatorname*{avg}_{|v|=k} \mathbb{E}_v\big[\langle 1_{\mathcal{X}_t}, \widehat{\xi} 1_{\mathcal{X}_t}\rangle\big] \lesssim e^{2\gamma t}\bigg(\frac{k^2}{n^2}\sum_{i,j=1}^n \xi_{ij}^2 + \frac{k}{n}p_\xi\bigg). \tag{6.11}$$

Applying Theorem 5.1(iiia) with $G = \widehat{\xi}$, we have

$$\mathbb{E}_v\big[\langle 1_{\mathcal{X}_t}, \widehat{\xi} 1_{\mathcal{X}_t}\rangle\big] \le \langle 1_v, \widehat{\xi}_t 1_v\rangle + \gamma \int_0^t \Big\langle 1_v, \xi e^{\gamma(t-u)\xi}(\widehat{\xi}_u)_{\mathrm{diag}}\Big\rangle \, du, \tag{6.12}$$

where $\widehat{\xi}_u = e^{\gamma u \xi}\widehat{\xi} e^{\gamma u \xi^\top}$. We now average over all choices of $v \subset [n]$ of size k. Starting with the second term, we use the identity (6.6) to get

$$\operatorname*{avg}_{|v|=k} \Big\langle 1_v, \xi e^{\gamma(t-u)\xi}(\widehat{\xi}_u)_{\mathrm{diag}}\Big\rangle = \frac{k}{n}\Big\langle 1, \xi e^{\gamma(t-u)\xi}(\widehat{\xi}_u)_{\mathrm{diag}}\Big\rangle.$$

Then, use the coordinatewise inequality $\xi^\top 1 \le 1$ (column sum bound), recalling that 1 denotes the all-ones vector:

$$\operatorname*{avg}_{|v|=k} \Big\langle 1_v, \xi e^{\gamma(t-u)\xi}(\widehat{\xi}_u)_{\mathrm{diag}}\Big\rangle \le \frac{k}{n} e^{\gamma(t-u)}\Big\langle 1, (\widehat{\xi}_u)_{\mathrm{diag}}\Big\rangle = \frac{k}{n}e^{\gamma(t-u)}\mathrm{Tr}(\widehat{\xi}_u). \tag{6.13}$$

For the first term in (6.12), we use the identity

$$\operatorname*{avg}_{|v|=k} 1_{i,j\in v} = \frac{k(k-1)}{n(n-1)}1_{i\neq j} + \frac{k}{n}1_{i=j} = \frac{k(k-1)}{n(n-1)} + \frac{k(n-k)}{n(n-1)}1_{i=j},$$

valid for $i,j \in [n]$. Indeed, this simply says that the probability of both i and j belonging to a uniformly random set $v \subset [n]$ of size k is $k(k-1)/n(n-1)$, if $i \neq j$, or k/n, if $i = j$. As a consequence, for any $n \times n$ matrix G,

$$\operatorname*{avg}_{|v|=k} \langle 1_v, G 1_v\rangle = \operatorname*{avg}_{|v|=k} \sum_{i,j=1}^n G_{ij} 1_{i,j\in v} = \frac{k(k-1)}{n(n-1)}\sum_{i,j=1}^n G_{ij} + \frac{k(n-k)}{n(n-1)}\mathrm{Tr}(G). \tag{6.14}$$

Apply this to the first term in (6.12) and simplify using the bounds $(k-1)/(n-1) \le k/n$ and $(n-k)/(n-1) \le 1$:

$$\operatorname*{avg}_{|v|=k} \langle 1_v, \widehat{\xi}_t 1_v\rangle \le \frac{k^2}{n^2}\sum_{i,j=1}^n (\widehat{\xi}_t)_{ij} + \frac{k}{n}\mathrm{Tr}(\widehat{\xi}_t). \tag{6.15}$$

The first term on the right-hand side can be controlled using

$$\sum_{i,j=1}^n (\widehat{\xi}_t)_{ij} = \big\langle 1, e^{\gamma t\xi}\widehat{\xi} e^{\gamma t \xi^\top} 1\big\rangle \le e^{2\gamma t}\langle 1, \widehat{\xi} 1\rangle = e^{2\gamma t}\sum_{i,j=1}^n \xi_{ij}^2.$$

Plug this into (6.15), and then plug the result along with (6.13) into (6.12), to get

$$\operatorname*{avg}_{|v|=k} \mathbb{E}_v\big[\langle 1_{\mathcal{X}_t}, \widehat{\xi} 1_{\mathcal{X}_t}\rangle\big] \lesssim e^{2\gamma t}\frac{k^2}{n^2}\sum_{i,j=1}^n \xi_{ij}^2 + \frac{k}{n}\mathrm{Tr}(\widehat{\xi}_t) + \frac{k}{n}\gamma\int_0^t e^{\gamma(t-u)}\mathrm{Tr}(\widehat{\xi}_u)\, du. \tag{6.16}$$

To complete the proof of (6.11), it suffices to show that

$$\mathrm{Tr}(\widehat{\xi}_u) \le 2e^{2\gamma u} p_\xi, \qquad u \ge 0. \tag{6.17}$$

To this end, we make use of a Taylor series formula

$$\widehat{\xi}_t = e^{\gamma t\xi}\widehat{\xi} e^{\gamma t\xi^\top} = \sum_{m=0}^\infty \frac{(\gamma t)^m}{m!}\Gamma_m(\widehat{\xi}), \qquad \Gamma_m(\widehat{\xi}) := \sum_{\ell=0}^m \binom{m}{\ell}\xi^\ell \widehat{\xi}(\xi^\top)^{m-\ell}, \tag{6.18}$$

which is easily derived using a Cauchy product calculation:

$$\begin{aligned} e^{t\xi}Ge^{t\xi^\top} &= \bigg(\sum_{r=0}^\infty \frac{t^r}{r!}\xi^r\bigg)\bigg(\sum_{r=0}^\infty \frac{t^r}{r!}G(\xi^\top)^r\bigg) = \sum_{m=0}^\infty\sum_{r=0}^m \frac{t^m}{r!(m-r)!}\xi^r G(\xi^\top)^{m-r} \\ &= \sum_{m=0}^\infty \frac{t^m}{m!}\Gamma_m(G), \qquad \text{for } t\in\mathbb{R}. \end{aligned} \tag{6.19}$$

The diagonal entries of ξ are zero, which implies that $\mathrm{Tr}(\Gamma_0(\widehat{\xi}))=0$. Hence,

$$\mathrm{Tr}(\widehat{\xi}_t) = \gamma t\,\mathrm{Tr}\big(\Gamma_1(\widehat{\xi})\big) + \sum_{m=2}^\infty \frac{(\gamma t)^m}{m!}\mathrm{Tr}\big(\Gamma_m(\widehat{\xi})\big). \tag{6.20}$$

The $m=1$ term is easily computed via

$$\mathrm{Tr}\big(\Gamma_1(\widehat{\xi})\big) = \mathrm{Tr}\big(\xi\widehat{\xi}\big) + \mathrm{Tr}\big(\widehat{\xi}\xi^\top\big) = \sum_{i,j=1}^n (\xi_{ij}^3 + \xi_{ij}^2\xi_{ji}). \tag{6.21}$$

The $m\ge 2$ terms are estimated as follows. Write

$$\mathrm{Tr}\big(\xi^\ell\widehat{\xi}(\xi^\top)^{m-\ell}\big) = \sum_{i,j=1}^n \xi_{ij}^2\big((\xi^\top)^{m-\ell}\xi^\ell\big)_{ji}.$$

Let $e_1,\ldots,e_n$ denote the standard basis in $\mathbb{R}^n$, and recall that $\|\xi\|_{\mathrm{op}}\le 1$. For $m>\ell>0$, we have

$$\big((\xi^\top)^{m-\ell}\xi^\ell\big)_{ji} \le |\xi^\ell e_i||\xi^{m-\ell}e_j| \le |\xi e_i||\xi e_j| \le \frac{1}{2}(|\xi e_i|^2 + |\xi e_j|^2).$$

If $m=\ell\ge 2$, we have

$$\big((\xi^\top)^{m-\ell}\xi^\ell\big)_{ji} = \langle \xi^\top e_j, \xi^{m-2}\xi e_i\rangle \le |\xi e_i||\xi^\top e_j| \le \frac{1}{2}(|\xi e_i|^2 + |\xi^\top e_j|^2).$$

If $m\ge 2$ and $\ell=0$, we have

$$\big((\xi^\top)^{m-\ell}\xi^\ell\big)_{ji} = \langle \xi e_j, (\xi^\top)^{m-2}\xi^\top e_i\rangle \le |\xi e_j||\xi^\top e_i| \le \frac{1}{2}(|\xi e_j|^2 + |\xi^\top e_i|^2).$$

Hence, for $m\ge 2$, we split off and then recombine the $\ell\in\{0,m\}$ cases to get

$$\begin{aligned} \mathrm{Tr}\big(\Gamma_m(\widehat{\xi})\big) &= \sum_{\ell=0}^m \binom{m}{\ell}\mathrm{Tr}\big(\xi^\ell\widehat{\xi}(\xi^\top)^{m-\ell}\big) \\ &\le \frac{1}{2}\sum_{i,j=1}^n \xi_{ij}^2(|\xi e_i|^2 + |\xi e_j|^2 + |\xi^\top e_i|^2 + |\xi^\top e_j|^2) + \frac{1}{2}\sum_{\ell=1}^{m-1}\binom{m}{\ell}\xi_{ij}^2(|\xi e_i|^2 + |\xi e_j|^2) \\ &\le 2^m \sum_{i,j=1}^n \xi_{ij}^2(|\xi e_i|^2 + |\xi e_j|^2 + |\xi^\top e_i|^2 + |\xi^\top e_j|^2) \\ &= 2^m \sum_{i,j,r=1}^n \xi_{ij}^2(\xi_{ri}^2 + \xi_{rj}^2 + \xi_{ir}^2 + \xi_{jr}^2) \\ &= 2^m \sum_{i=1}^n \bigg(\sum_{j=1}^n (\xi_{ij}^2 + \xi_{ji}^2)\bigg)^2. \end{aligned} \tag{6.22}$$

Plug this and (6.21) into (6.20) to get

$$\mathrm{Tr}(\widehat{\xi}_t) \le \gamma t\sum_{i,j=1}^n (\xi_{ij}^3 + \xi_{ij}^2\xi_{ji}) + \sum_{m=2}^\infty \frac{(2\gamma t)^m}{m!}\sum_{i=1}^n\bigg(\sum_{j=1}^n(\xi_{ij}^2 + \xi_{ji}^2)\bigg)^2 \le 2e^{2\gamma t}p_\xi,$$

completing the proof of (6.17) and thus of Step 1.

Step 2. We next show that

$$\operatorname*{avg}_{|v|=k} \mathbb{E}_v\big[|\mathcal{X}_t|\langle 1_{\mathcal{X}_t}, \widehat{\xi} 1_{\mathcal{X}_t}\rangle\big] \lesssim e^{3\gamma t}\bigg(\frac{k^3}{n^2}\sum_{i,j=1}^n \xi_{ij}^2 + \frac{k^2}{n}p_\xi\bigg). \tag{6.23}$$

Applying Theorem 5.1(iiib) with $G = \widehat{\xi}$, we have

$$\begin{aligned}\mathbb{E}_v\big[|\mathcal{X}_t|\langle 1_{\mathcal{X}_t}, \widehat{\xi} 1_{\mathcal{X}_t}\rangle\big] &\le |v|e^{\gamma t}\Big\langle 1_v, \Big(\xi\widehat{\xi}_t + \widehat{\xi}_t\xi^\top + \widehat{\xi}_t\Big)1_v\Big\rangle \\ &\quad + \gamma|v|e^{\gamma t}\int_0^t \langle 1_v, e^{\gamma(t-s)\xi}(I+\xi)\xi(\xi\widehat{\xi}_s + \widehat{\xi}_s\xi^\top + 2\widehat{\xi}_s)_{\mathrm{diag}}\rangle\, ds.\end{aligned} \tag{6.24}$$

We now average over all choices of $v \subset [n]$ of size k. Starting with the second term of (6.24), we use the identity (6.6) along with the coordinatewise inequality $1^\top\xi \le 1^\top$ (due to the column sums being bounded by 1) to get

$$\begin{aligned}&\operatorname*{avg}_{|v|=k} \gamma|v|e^{\gamma t}\int_0^t \langle 1_v, e^{\gamma(t-s)\xi}(I+\xi)\xi(\xi\widehat{\xi}_s + \widehat{\xi}_s\xi^\top + 2\widehat{\xi}_s)_{\mathrm{diag}}\rangle\, ds \\ &\quad = \frac{k^2}{n}\gamma e^{\gamma t}\int_0^t \langle 1, e^{\gamma(t-s)\xi}(I+\xi)\xi(\xi\widehat{\xi}_s + \widehat{\xi}_s\xi^\top + 2\widehat{\xi}_s)_{\mathrm{diag}}\rangle\, ds \\ &\quad \le \frac{k^2}{n}2\gamma e^{\gamma t}\int_0^t e^{\gamma(t-s)}\langle 1, (\xi\widehat{\xi}_s + \widehat{\xi}_s\xi^\top + 2\widehat{\xi}_s)_{\mathrm{diag}}\rangle\, ds \\ &\quad = \frac{k^2}{n}2\gamma e^{\gamma t}\int_0^t e^{\gamma(t-s)}\mathrm{Tr}(\xi\widehat{\xi}_s + \widehat{\xi}_s\xi^\top + 2\widehat{\xi}_s)\, ds\end{aligned} \tag{6.25}$$

Turning to the first term in (6.24), we start by using (6.14) to get

$$\operatorname*{avg}_{|v|=k} |v|e^{\gamma t}\Big\langle 1_v, \Big(\xi\widehat{\xi}_t + \widehat{\xi}_t\xi^\top + 2\widehat{\xi}_t\Big)1_v\Big\rangle \le e^{\gamma t}\frac{k^3}{n^2}\sum_{i,j=1}^n\Big(\xi\widehat{\xi}_t + \widehat{\xi}_t\xi^\top + 2\widehat{\xi}_t\Big)_{ij} + e^{\gamma t}\frac{k^2}{n}\mathrm{Tr}\Big(\xi\widehat{\xi}_t + \widehat{\xi}_t\xi^\top + 2\widehat{\xi}_t\Big).$$

Using the row and column sum bounds, $\xi^\top 1 \le 1$ and $\xi 1 \le 1$,

$$\begin{aligned}e^{\gamma t}\frac{k^3}{n^2}\sum_{i,j=1}^n\Big(\xi\widehat{\xi}_t + \widehat{\xi}_t\xi^\top + 2\widehat{\xi}_t\Big)_{ij} &= e^{\gamma t}\frac{k^3}{n^2}\Big\langle 1, \big(\xi\widehat{\xi}_t + \widehat{\xi}_t\xi^\top + 2\widehat{\xi}_t\big)1\Big\rangle \le 4e^{\gamma t}\frac{k^3}{n^2}\big\langle 1, \widehat{\xi}_t 1\big\rangle = 4\frac{k^3}{n^2}e^{\gamma t}\big\langle 1, e^{\gamma t\xi}\widehat{\xi}e^{\gamma t\xi^\top}1\big\rangle \\ &\le 4e^{3\gamma t}\frac{k^3}{n^2}\langle 1, \widehat{\xi}1\rangle = 4e^{3\gamma t}\frac{k^3}{n^2}\sum_{i,j=1}^n \xi_{ij}^2.\end{aligned}$$

Plugging this and (6.25) into (6.24), we find

$$\begin{aligned}\operatorname*{avg}_{|v|=k} \mathbb{E}_v\big[|\mathcal{X}_t|\langle 1_{\mathcal{X}_t}, \widehat{\xi} 1_{\mathcal{X}_t}\rangle\big] &\lesssim e^{3\gamma t}\frac{k^3}{n^2}\sum_{i,j=1}^n \xi_{ij}^2 + e^{\gamma t}\frac{k^2}{n}\mathrm{Tr}\Big(\xi\widehat{\xi}_t + \widehat{\xi}_t\xi^\top + 2\widehat{\xi}_t\Big) \\ &\quad + \frac{k^2}{n}2\gamma e^{\gamma t}\int_0^t e^{\gamma(t-s)}\mathrm{Tr}(\xi\widehat{\xi}_s + \widehat{\xi}_s\xi^\top + 2\widehat{\xi}_s)\, ds\end{aligned}$$

Recalling from (6.17) that $\mathrm{Tr}(\widehat{\xi}_s) \le 2e^{2\gamma s}p_\xi$, the proof of (6.23) will be complete once we show that

$$\mathrm{Tr}\Big(\xi\widehat{\xi}_t + \widehat{\xi}_t\xi^\top\Big) \le 2e^{2\gamma t}p_\xi, \tag{6.26}$$

To do so, we will again make use of the Taylor series (6.18), by writing

$$\mathrm{Tr}(\xi\widehat{\xi}_t + \widehat{\xi}_t\xi^\top) = \sum_{m=0}^\infty \frac{(\gamma t)^m}{m!}\mathrm{Tr}\big(\xi\Gamma_m(\widehat{\xi}) + \Gamma_m(\widehat{\xi})\xi^\top\big) = \sum_{m=0}^\infty \frac{(\gamma t)^m}{m!}\mathrm{Tr}\big(\Gamma_{m+1}(\widehat{\xi})\big).$$

The last step used the identity

$$\xi\Gamma_m(G) + \Gamma_m(G)\xi^\top = \sum_{r=0}^m \binom{m}{r}\Big[(\xi)^{r+1}G(\xi^\top)^{m-r} + (\xi)^r G(\xi^\top)^{m-r+1}\Big] = \Gamma_{m+1}(G), \tag{6.27}$$

which follows from a more general fact that for any sequence $\{a_n\}$,

$$\sum_{r=0}^m \binom{m}{r}(a_{r+1} + a_r) = \sum_{r=0}^{m+1}\binom{m+1}{r}a_r. \tag{6.28}$$

Recalling the estimates (6.21) and (6.22), we bound this by

$$\mathrm{Tr}(\xi\widehat{\xi}_t + \widehat{\xi}_t\xi^\top) \le \sum_{i,j=1}^n(\xi_{ij}^3 + \xi_{ij}^2\xi_{ji}) + \sum_{m=1}^\infty \frac{(\gamma t)^m}{m!}2^{m+1}\sum_{i=1}^n\bigg(\sum_{j=1}^n(\xi_{ij}^2 + \xi_{ji}^2)\bigg)^2 \le 2e^{2\gamma t}p_\xi.$$

This proves (6.26), thus completing the proof of Step 2.

With Steps 1 and 2 complete, we now put them together with (6.10) to yield a bound for (6.9). Specifically, adding (6.10) plus (6.11) plus δ times (6.23), we deduce from (6.9) that

$$\overline{H}^k_{[T]} \lesssim (\delta k + 1)\bigg(\frac{k^2}{n^2}\sum_{i,j=1}^n \xi_{ij}^2 + \frac{k}{n}p_\xi\bigg).$$

This proves the claim (2.18) of Theorem 2.10. To prove the uniform-in-time claim, we make minor modifications: Use (6.3) in place of (6.2) to get the following alternative to (6.9):

$$\begin{aligned} H_T(v) &\lesssim e^{-rT}\mathbb{E}_v\big[\delta|\mathcal{X}_T|^3 + |\mathcal{X}_T|^2\big]\frac{1}{n^2}\sum_{i,j=1}^n\xi_{ij}^2 + e^{-rT}\mathbb{E}_v\big[\delta|\mathcal{X}_T|^2 + |\mathcal{X}_T|\big]\frac{p_\xi}{n} \\ &\quad + \int_0^T e^{-rt}\mathbb{E}_v\big[(\delta|\mathcal{X}_t| + 1)\langle 1_{\mathcal{X}_t}, \widehat{\xi}1_{\mathcal{X}_t}\rangle\big]\,dt, \end{aligned}$$

where $r = \sigma^2/4\eta$. In the estimates of the expectations above, although we suppressed many of the time-dependent constants, the largest exponential term was $e^{3\gamma t}$. Hence, because $r > 3\gamma$ in assumption (U)(iii), we end up with the same bound for $\sup_{T>0}H_T(v)$. $\square$

6.4. **Setwise entropy: Proof of Theorem 2.13.** In this section we prove Theorem 2.13. Recall that our assumption therein on the initial condition is that $H_0(v) \le C_0 q_\xi(v)$ for all $v \subset [n]$, where $q_\xi(v)$ can be written as

$$q_\xi(v) = (\delta|v| + 1)\langle 1_v, \widehat{\xi}1_v\rangle + \delta(\delta|v| + 1)\langle 1_v, (\xi^\top\xi + \xi\xi^\top)1_v\rangle + \delta^3|v|^2 + \delta^2|v|, \tag{6.29}$$

where $\widehat{\xi}_{ij} = \xi_{ij}^2$ is the entrywise square of ξ.

We first claim that

$$q_\xi(v) \le 8\delta^2|v|^3, \tag{6.30}$$

which will thus allow us to apply Lemma 6.1 and its consequences outlined at the beginning of the section. Note that (6.30) holds trivially for $v = \emptyset$, so we assume that $|v| \ge 1$ henceforth. The operator norm bound 2.1 $\|\xi\|_{\mathrm{op}} \le 1$ implies

$$\langle 1_v, \xi\xi^\top 1_v\rangle = |\xi^\top 1_v|^2 \le |1_v|^2 = |v|^2.$$

The same bound holds for $\langle 1_v, \xi^\top \xi 1_v\rangle$. Using also $\sum_{i,j\in v} \xi_{ij}^2 \le \delta^2|v|^2$, we deduce

$$q_\xi(v) \le (\delta|v|+1)\delta^2|v|^2 + 2\delta(\delta|v|+1)|v|^2 + \delta^3|v|^2 + \delta^2|v| \le 8\delta^2|v|^3,$$

where the last step just used $|v| \ge 1$ and $\delta \le 1$. This establishes (6.30) for $v \ne \emptyset$..

Bounding the first term in (6.1) using convexity, we have

$$\widehat{C}(v) \lesssim \delta|v|\sum_{i\in v}\xi_{ij}^2 + \sum_{i,j\in v}\xi_{ij}^2 = (\delta|v|+1)\langle 1_v, \widehat{\xi}1_v\rangle.$$

Now, by Lemma 6.1, we may apply (6.2) and (6.29) to get

$$\begin{aligned} H_{[T]}(v) \lesssim \mathbb{E}_v\Big[&(\delta|\mathcal{X}_T|+1)\langle 1_{\mathcal{X}_T}, \widehat{\xi}1_{\mathcal{X}_T}\rangle + \delta(\delta|\mathcal{X}_T|+1)\langle 1_{\mathcal{X}_T}, (\xi^\top\xi+\xi\xi^\top)1_{\mathcal{X}_T}\rangle + \delta^3|\mathcal{X}_T|^2 + \delta^2|\mathcal{X}_T|\Big] \\ &+ \int_0^T \mathbb{E}_v\big[(\delta|\mathcal{X}_t|+1)\langle 1_{\mathcal{X}_t}, \widehat{\xi}1_{\mathcal{X}_t}\rangle\big]\,dt. \end{aligned} \tag{6.31}$$

We next apply Theorem 5.1 to estimate each term. This will be done in five steps.

Step 1. Using Theorem 5.1(ia, ib), we have

$$\mathbb{E}_v|\mathcal{X}_T| \le e^{\gamma T}|v|, \quad \mathbb{E}_v|\mathcal{X}_T|^2 \le 2e^{2\gamma T}|v|^2. \tag{6.32}$$

Step 2. We next show that

$$\mathbb{E}_v\big[\langle 1_{\mathcal{X}_t}, \widehat{\xi}1_{\mathcal{X}_t}\rangle\big] \lesssim e^{2\gamma t}\Big(\langle 1_v, \widehat{\xi}1_v\rangle + \delta\langle 1_v, (\xi\xi^\top+\xi^\top\xi)1_v\rangle + |v|\delta^2\Big). \tag{6.33}$$

By Theorem 5.1(iiia) with $G = \widehat{\xi}$, we have

$$\mathbb{E}_v\big[\langle 1_{\mathcal{X}_t}, \widehat{\xi}1_{\mathcal{X}_t}\rangle\big] \le \langle 1_v, \widehat{\xi}_t 1_v\rangle + \gamma\int_0^t \Big\langle 1_v, \xi e^{\gamma(t-u)\xi}(\widehat{\xi}_u)_{\mathrm{diag}}\Big\rangle\,du,$$

where $\widehat{\xi}_t = e^{\gamma t\xi}\widehat{\xi}e^{\gamma t\xi^\top}$. To estimate this, we write

$$\widehat{\xi}_t = \widehat{\xi} + (e^{\gamma t\xi}-I)\widehat{\xi} + e^{\gamma t\xi}\widehat{\xi}(e^{\gamma t\xi^\top}-I).$$

Using the Cauchy-Schwarz inequality, $\|\xi\|_{\mathrm{op}} \le 1$, and the simple coordinatewise inequality $\widehat{\xi} \le \delta\xi$,

$$\begin{aligned} \langle 1_v, (e^{\gamma t\xi}-I)\widehat{\xi}1_v\rangle = \gamma\int_0^t\langle 1_v, \xi e^{\gamma u\xi}\widehat{\xi}1_v\rangle\,du &\le \gamma te^{\gamma t}|\xi^\top 1_v||\widehat{\xi}1_v| \le \delta\gamma te^{\gamma t}|\xi^\top 1_v||\xi 1_v| \\ &\le \frac{1}{2}\delta\gamma te^{\gamma t}\big(|\xi^\top 1_v|^2 + |\xi 1_v|^2\big). \end{aligned}$$

Similarly,

$$\begin{aligned} \langle 1_v, e^{\gamma t\xi}\widehat{\xi}(e^{\gamma t\xi^\top}-I)1_v\rangle &= \gamma\int_0^t\langle 1_v, e^{\gamma t\xi}\widehat{\xi}e^{\gamma u\xi^\top}\xi^\top 1_v\rangle\,du \\ &\le \delta\gamma\int_0^t\langle 1_v, e^{\gamma t\xi}\xi e^{\gamma u\xi^\top}\xi^\top 1_v\rangle\,du \\ &\le \delta\gamma te^{2\gamma t}|\xi^\top 1_v|^2. \end{aligned}$$

Combining the above three displays,

$$\langle 1_v, \widehat{\xi}_t 1_v\rangle \lesssim \langle 1_v, \widehat{\xi}1_v\rangle + e^{2\gamma t}\delta\langle 1_v, (\xi\xi^\top+\xi^\top\xi)1_v\rangle. \tag{6.34}$$

Finally, letting 1 denote the vector of all ones, note the coordinatewise inequality

$$(\widehat{\xi}_u)_{\mathrm{diag}} = \big(e^{\gamma u\xi}\widehat{\xi}e^{\gamma u\xi^\top}\big)_{\mathrm{diag}} \le \delta^2\big(e^{\gamma u\xi}11^\top e^{\gamma u\xi^\top}\big)_{\mathrm{diag}} \le \delta^2 e^{2\gamma u}\big(11^\top\big)_{\mathrm{diag}} = \delta^2 e^{2\gamma u}1,$$

where the second inequality used the fact that $\xi 1 \le \xi$. Hence,

$$\left\langle 1_v, \xi e^{\gamma(t-u)\xi} (\widehat{\xi}_u)_{\mathrm{diag}} \right\rangle \le \delta^2 e^{2\gamma u} \left\langle 1_v, \xi e^{\gamma(t-u)\xi} 1 \right\rangle \le \delta^2 e^{2\gamma t} \left\langle 1_v, 1 \right\rangle = |v| \delta^2 e^{2\gamma t}.$$

This proves (6.33).

Step 3. We next show that

$$\mathbb{E}_v\big[|\mathcal{X}_t| \langle 1_{\mathcal{X}_t}, \widehat{\xi} 1_{\mathcal{X}_t} \rangle\big] \lesssim |v| e^{3\gamma t} \Big(\langle 1_v, \widehat{\xi} 1_v \rangle + \delta \langle 1_v, (\xi\xi^\top + \xi^\top \xi) 1_v \rangle + |v| \delta^2 \Big). \tag{6.35}$$

By Theorem 5.1(iiib) with $G = \widehat{\xi}$, we have

$$\begin{aligned} \mathbb{E}_v\big[|\mathcal{X}_t| \langle 1_{\mathcal{X}_t}, \widehat{\xi} 1_{\mathcal{X}_t} \rangle\big] &\le |v| e^{\gamma t} \Big\langle 1_v, \Big(\xi \widehat{\xi}_t + \widehat{\xi}_t \xi^\top + \widehat{\xi}_t \Big) 1_v \Big\rangle \\ &\quad + \gamma |v| e^{\gamma t} \int_0^t \langle 1_v, e^{\gamma(t-s)\xi} (I + \xi) \xi (\xi \widehat{\xi}_s + \widehat{\xi}_s \xi^\top + 2 \widehat{\xi}_s)_{\mathrm{diag}} \rangle \, ds \end{aligned} \tag{6.36}$$

The last term is $\lesssim e^{3\gamma t} |v|^2 \delta^2$, by using again the coordinatewise inequality $(\widehat{\xi}_s)_{\mathrm{diag}} \le \delta^2 e^{2\gamma s} 1$ justified in the previous step, which is also valid for $(\xi \widehat{\xi}_s)_{\mathrm{diag}}$ and $(\widehat{\xi}_s \xi^\top)_{\mathrm{diag}}$ by the same reasoning. To estimate the first term in (6.36), we reuse the estimate (6.34), and we note also that

$$\begin{aligned} \langle 1_v, \xi \widehat{\xi}_t 1_v \rangle &= \langle 1_v, \xi e^{\gamma t \xi} \widehat{\xi} (e^{\gamma t \xi^\top} - I) 1_v \rangle + \langle 1_v, \xi e^{\gamma t \xi} \widehat{\xi} 1_v \rangle \\ &= \gamma \int_0^t \langle 1_v, \xi e^{\gamma t \xi} \widehat{\xi} e^{\gamma u \xi^\top} \xi^\top 1_v \rangle \, du + \langle 1_v, \xi e^{\gamma t \xi} \widehat{\xi} 1_v \rangle \\ &\le \delta \gamma t e^{2\gamma t} |\xi^\top 1_v|^2 + \delta e^{\gamma t} |\xi^\top 1_v| |\xi 1_v|, \end{aligned}$$

and similarly

$$\begin{aligned} \langle 1_v, \widehat{\xi}_t \xi^\top 1_v \rangle &= \gamma \int_0^t \langle 1_v, \xi e^{\gamma u \xi} \widehat{\xi} e^{\gamma t \xi^\top} \xi \xi^\top 1_v \rangle \, du + \langle 1_v, \widehat{\xi} e^{\gamma t \xi^\top} \xi \xi^\top 1_v \rangle \\ &\le \delta \gamma e^{2\gamma t} |\xi^\top 1_v|^2 + \delta e^{\gamma t} |\xi^\top 1_v|^2. \end{aligned}$$

By Young's inequality, $2|\xi^\top 1_v||\xi 1_v| \le \langle 1_v, (\xi\xi^\top + \xi^\top \xi) 1_v \rangle$, we see that the first inner product in (6.36) is bounded by the same right-hand side as in (6.34). This proves (6.35).

Step 4. Similarly to Step 2, we will next show that

$$\mathbb{E}_v\big[\langle 1_{\mathcal{X}_t}, (\xi^\top \xi + \xi \xi^\top) 1_{\mathcal{X}_t} \rangle\big] \lesssim e^{2\gamma t} \Big(\langle 1_v, (\xi\xi^\top + \xi^\top \xi) 1_v \rangle + |v| \delta \Big). \tag{6.37}$$

By Theorem 5.1(iiia) with $G = \xi^\top \xi + \xi \xi^\top$, we have

$$\mathbb{E}_v\big[\langle 1_{\mathcal{X}_t}, (\xi^\top \xi + \xi \xi^\top) 1_{\mathcal{X}_t} \rangle\big] \le \langle 1_v, G_t 1_v \rangle + \gamma \int_0^t \Big\langle 1_v, \xi e^{\gamma(t-u)\xi} (G_u)_{\mathrm{diag}} \Big\rangle \, du, \tag{6.38}$$

where $G_t = e^{\gamma t \xi} G e^{\gamma t \xi^\top}$. To estimate this, we write

$$G_t = G + (e^{\gamma t \xi} - I) G + G (e^{\gamma t \xi^\top} - I) + (e^{\gamma t \xi} - I) G (e^{\gamma t \xi^\top} - I).$$

Using the Cauchy-Schwarz inequality and $\|\xi\|_{\mathrm{op}} \le 1$,

$$\begin{aligned} \langle 1_v, (e^{\gamma t \xi} - I) G 1_v \rangle &= \gamma \int_0^t \langle 1_v, \xi e^{\gamma u \xi} G 1_v \rangle \, du \le \gamma t e^{\gamma t} |\xi^\top 1_v| |G 1_v| \\ &\le \gamma t e^{\gamma t} |\xi^\top 1_v| (|\xi^\top 1_v| + |\xi 1_v|) \end{aligned}$$

Similarly,

$$\langle 1_v, G (e^{\gamma t \xi^\top} - I) 1_v \rangle = \gamma \int_0^t \langle 1_v, G e^{\gamma u \xi^\top} \xi^\top 1_v \rangle \, du \le \gamma t e^{\gamma t} |\xi^\top 1_v| (|\xi^\top 1_v| + |\xi 1_v|).$$

Lastly, using $\|G\|_{\mathrm{op}} \le 2$, we get

$$\langle 1_v, (e^{\gamma t\xi} - I)G(e^{\gamma t\xi^\top} - I)1_v\rangle = \gamma^2 \int_0^t \int_0^t \langle 1_v, \xi e^{\gamma u\xi} G e^{\gamma u'\xi^\top} \xi^\top 1_v\rangle \, du \, du' \le 2\gamma^2 t^2 e^{2\gamma t} |\xi^\top 1_v|^2.$$

Combining the above four displays and again using $2|\xi^\top 1_v||\xi 1_v| \le \langle 1_v, (\xi\xi^\top + \xi^\top\xi)1_v\rangle$, we find

$$\begin{aligned}\langle 1_v, G_t 1_v\rangle &\le \langle 1_v, G1_v\rangle + 2\gamma t e^{\gamma t}|\xi^\top 1_v|(|\xi^\top 1_v| + |\xi 1_v|) + 2\gamma^2 t^2 e^{2\gamma t}|\xi^\top 1_v|^2 \\ &\lesssim e^{2\gamma t}\langle 1_v, (\xi^\top\xi + \xi\xi^\top)1_v\rangle. \end{aligned} \tag{6.39}$$

Next, combining the entrywise inequalities $\xi \le \delta 11^\top$, $\xi 1 \le 1$, and $1^\top \xi \le 1^\top$, we find

$$G_t = e^{\gamma t\xi}(\xi^\top\xi + \xi\xi^\top)e^{\gamma t\xi^\top} \le \delta e^{\gamma t\xi}(11^\top\xi + \xi 11^\top)e^{\gamma t\xi^\top} \le 2\delta e^{2\gamma t}11^\top, \tag{6.40}$$

and thus $(G_t)_{\mathrm{diag}} \le 2\delta e^{2\gamma t}1$. Using also $\langle 1_v, 1\rangle = |v|$, we deduce

$$\gamma \int_0^t \left\langle 1_v, \xi e^{\gamma(t-u)\xi}(G_u)_{\mathrm{diag}}\right\rangle du \le 2\delta\gamma \int_0^t \left\langle 1_v, \xi e^{\gamma(t-u)\xi} e^{2\gamma u} 1\right\rangle du \le 2\delta\gamma t e^{3\gamma t}|v|.$$

Plug this and (6.39) into (6.38) to deduce (6.37).

Step 5. Similarly to Step 3, we will next show that

$$\mathbb{E}_v\big[|\mathcal{X}_t|\langle 1_{\mathcal{X}_t}, (\xi^\top\xi + \xi\xi^\top)1_{\mathcal{X}_t}\rangle\big] \lesssim |v|e^{3\gamma t}\Big(\langle 1_v, (\xi\xi^\top + \xi^\top\xi)1_v\rangle + |v|\delta\Big). \tag{6.41}$$

Let us again use the notation G and G_t from Step 4. By Theorem 5.1(iiib), we have

$$\begin{aligned}\mathbb{E}_v\big[|\mathcal{X}_t|\langle 1_{\mathcal{X}_t}, G1_{\mathcal{X}_t}\rangle\big] &\le |v|e^{\gamma t}\Big\langle 1_v, \Big(\xi G_t + G_t\xi^\top + G_t\Big)1_v\Big\rangle \\ &\quad + \gamma|v|e^{\gamma t}\int_0^t \langle 1_v, e^{\gamma(t-s)\xi}(I+\xi)\xi(\xi G_s + G_s\xi^\top + 2G_s)_{\mathrm{diag}}\rangle \, ds, \end{aligned} \tag{6.42}$$

The last term is $\lesssim e^{3\gamma t}|v|^2\delta$, by using again the coordinatewise inequality $(G_u)_{\mathrm{diag}} \le 2\delta e^{2\gamma u}1$ justified in the previous step, and similarly for $(\xi G_u)_{\mathrm{diag}}$ and $(G_u\xi^\top)_{\mathrm{diag}}$. To estimate the first term in (6.42), we reuse the estimate (6.39). Note also that, since $\|G\|_{\mathrm{op}} \le 2$,

$$\begin{aligned}\langle 1_v, \xi G_t 1_v\rangle &= \langle 1_v, \xi e^{\gamma t\xi} G(e^{\gamma t\xi^\top} - I)1_v\rangle + \langle 1_v, \xi e^{\gamma t\xi} G 1_v\rangle \\ &= \gamma\int_0^t \langle 1_v, \xi e^{\gamma t\xi} G e^{\gamma u\xi^\top}\xi^\top 1_v\rangle \, du + \langle 1_v, \xi e^{\gamma t\xi}(\xi^\top\xi + \xi\xi^\top)1_v\rangle \\ &\le 2\gamma t e^{2\gamma t}|\xi^\top 1_v|^2 + e^{\gamma t}|\xi^\top 1_v|(|\xi^\top 1_v| + |\xi 1_v|), \end{aligned}$$

and similarly

$$\begin{aligned}\langle 1_v, G_t\xi^\top 1_v\rangle &= \gamma\int_0^t \langle 1_v, \xi e^{\gamma u\xi} G e^{\gamma t\xi^\top}\xi^\top 1_v\rangle \, du + \langle 1_v, (\xi^\top\xi + \xi\xi^\top)e^{\gamma t\xi^\top}\xi^\top 1_v\rangle \\ &\le 2\gamma t e^{2\gamma t}|\xi^\top 1_v|^2 + e^{\gamma t}|\xi^\top 1_v|(|\xi^\top 1_v| + |\xi 1_v|). \end{aligned}$$

Using Young's inequality and putting it together yields (6.41).

Step 6. In this step we put together Steps 1–5 to produce a bound for (6.31). Indeed, note that the bound (6.35) from Step 3 is $|v|$ times the bound (6.33) from Step 2, and similarly the bound (6.41) from Step 5 is $|v|$ times the bound (6.37) from Step 4. Keeping track of the factors of δ in (6.31), we get

$$\begin{aligned} H_{[T]}(v) &\lesssim \delta^2|v| + \delta^3|v|^2 + (\delta|v| + 1)\Big(\langle 1_v, \widehat{\xi} 1_v\rangle + \delta\langle 1_v, (\xi\xi^\top + \xi^\top\xi)1_v\rangle + |v|\delta^2\Big) \\ &\quad + \delta(\delta|v| + 1)\Big(\langle 1_v, (\xi\xi^\top + \xi^\top\xi)1_v\rangle + |v|\delta\Big). \end{aligned}$$

Combining terms, the right-hand side is $\lesssim q_\xi(v)$, and the proof of the first claim of Theorem 2.13 is complete.

Step 7. Next, we explain the uniform-in-time part of Theorem 2.13. This requires only some minor adaptations of the above arguments, most importantly keeping track of exponents. Using (6.3) instead of (6.2), we get the following analogue of (6.31), with $r = \sigma^2/4\eta$:

$$\begin{aligned} H_T(v) \lesssim e^{-rT}\mathbb{E}_v\Big[(\delta|\mathcal{X}_T|+1)\langle 1_{\mathcal{X}_T}, \widehat{\xi}1_{\mathcal{X}_T}\rangle + \delta(\delta|\mathcal{X}_T|+1)\langle 1_{\mathcal{X}_T}, (\xi^\top\xi + \xi\xi^\top)1_{\mathcal{X}_T}\rangle + \delta^2|\mathcal{X}_T|\Big] \\ + \int_0^T e^{-rt}\mathbb{E}_v\big[(\delta|\mathcal{X}_t|+1)\langle 1_{\mathcal{X}_t}, \widehat{\xi}1_{\mathcal{X}_t}\rangle\big]\,dt. \end{aligned} \tag{6.43}$$

This is the same as the right-hand side of (6.31) aside from the exponential terms. Checking through Steps 2–5 above, the largest exponential factor was $e^{3\gamma t}$, and thus the resulting bound on (6.43) is uniform in $T > 0$ because $r > 3\gamma$ by Assumption (U)(iii).

Step 8. It remains to justify the last claim of Theorem 2.13 regarding the removal of the column sum assumption. Indeed, (columns) was only used in Step 4 to obtain the bound the $\xi^\top\xi$ term in (6.40). If $\xi^\top\xi$ was removed from $q_\xi(v)$ in the initial chaoticity assumption, then tracing through the proof, the $\xi^\top\xi$ term would not be present in the left-hand side of (6.37), and thus also in (6.40). $\square$

7. Proof for Gaussian case

In this section we prove Theorem 2.15 and Proposition 2.17. Let us write $\lambda_{\max}(A)$ and $\lambda_{\min}(A)$ for the largest and smallest eigenvalues of a square symmetric matrix A. We first give an upper and lower bound for relative entropy between two centered Gaussian measures, which essentially performs a lead-order (quadratic) Taylor expansion of the entropy in terms of the covariance matrices. In the following, we write $\alpha_- = \max(-a, 0)$ for the negative part of a number α.

Proposition 7.1. *Consider two centered nondegenerate Gaussian measures γ_0 and γ_1 on $\mathbb{R}^k$ with covariance matrices Σ_0 and Σ_1.*

(1) If $-1 < \alpha \le \lambda_{\min}(\Sigma_0^{-1}\Sigma_1 - I)$, we have

$$H(\gamma_1 \,|\, \gamma_0) \le \Big(\frac{1}{2} + \frac{\alpha_-}{3(1+\alpha)^3}\Big)\mathrm{Tr}((\Sigma_0^{-1}\Sigma_1 - I)^2)$$

(2) If $-1 < \lambda_{\min}(\Sigma_0^{-1}\Sigma_1 - I) \le \lambda_{\max}(\Sigma_0^{-1}\Sigma_1 - I) \le 1$,

$$H(\gamma_1 \,|\, \gamma_0) \ge \frac{1}{6}\mathrm{Tr}((\Sigma_0^{-1}\Sigma_1 - I)^2)$$

Proof of Proposition 7.1. We make use of the following basic fact: If $f, g : \mathbb{R} \to \mathbb{R}$ are measurable functions satisfying $f(x) \le g(x)$ for all $x \in [-\rho, \rho]$, then

$$\mathrm{Tr}(f(A)) \le \mathrm{Tr}(g(A)) \tag{7.1}$$

for any symmetric matrix A with eigenvalues contained in $[-\rho, \rho]$. We start from the following well known explicit formula:

$$\begin{aligned} H(\gamma_1|\gamma_0) &= \frac{1}{2}\left[\mathrm{Tr}\left(\Sigma_0^{-1}\Sigma_1\right) - k + \log\frac{\det(\Sigma_0)}{\det(\Sigma_1)}\right] \\ &= \frac{1}{2}\Big(\mathrm{Tr}(\Sigma_0^{-1}\Sigma_1 - I) - \log\det(\Sigma_0^{-1}\Sigma_1)\Big) \\ &= \frac{1}{2}\mathrm{Tr}\,h(\Sigma_0^{-1}\Sigma_1 - I), \end{aligned}$$

where we used $\log\det = \operatorname{Tr}\log$, and the scalar function h is defined by $h(x) := x - \log(1+x)$. Note that $h(0) = h'(0) = 0$. With a bit of calculus, we have the following upper and lower bounds on h. For $-1 < \alpha \le x$, we have

$$h(x) \le x^2\Big(\frac{1}{2} + \frac{\alpha_-}{3(1+\alpha)^3}\Big). \tag{7.2}$$

Using the fact that the fourth derivative of h is positive, we have for $1 \ge x > -1$ that

$$h(x) \ge \frac{1}{2}x^2 - \frac{1}{3}x^3 = x^2\left(\frac{1}{2} - \frac{1}{3}x\right) \ge \frac{1}{6}x^2. \tag{7.3}$$

Combine these inequalities with (7.1) completes the proof. $\square$

Recall that the laws P_T and Q_T of the SDE systems (2.22) and (2.23) are the centered Gaussian measures on $\mathbb{R}^n$ with covariance matrices Σ_T and TI, respectively, where $\Sigma_T := \int_0^T e^{t\xi}e^{t\xi^\top}\,dt$. Recall the notation $\rho = \|\xi\|_{\mathrm{op}}$. The identity

$$\frac{1}{T}\Sigma_T - I = \frac{1}{T}\int_0^T (e^{t\xi}e^{t\xi^\top} - I)\,dt \tag{7.4}$$

implies that

$$\lambda_{\min}(T^{-1}\Sigma_T - I) \ge e^{-2\rho T} - 1, \qquad \lambda_{\max}(T^{-1}\Sigma_T - I) \le e^{2\rho T} - 1.$$

Marginalizing, the law P_t^v is the centered Gaussian with covariance matrix denoted Σ_t^v; in general, for an $n \times n$ matrix A, we write A^v for the $|v| \times |v|$ principal submatrix of A corresponding to indices in v. By Cauchy's interlacing theorem, we have

$$\lambda_{\min}(T^{-1}\Sigma_T^v - I) \ge e^{-2\rho T} - 1, \qquad \lambda_{\max}(T^{-1}\Sigma_T^v - I) \le e^{2\rho T} - 1, \tag{7.5}$$

for any $v \subset [n]$. Note that $\lambda_{\max}(T^{-1}\Sigma_T^v - I) \le 1$ when $T \le \log(2)/2\rho$. Hence, applying Proposition 7.1 with $\alpha = e^{-2\rho T} - 1$, and using $\frac{1}{2} + \frac{\alpha_-}{3(1+\alpha)^3} =\le \frac{1}{2} + (1 - e^{-2\rho T})/3e^{-6\rho T} \le e^{6\rho T}$, we have

$$H(P_T^v \,|\, Q_T^v) \le e^{6\rho T}\operatorname{Tr}\Big(\Big(\frac{1}{T}\Sigma_T^v - I\Big)^2\Big), \quad \forall T \ge 0, \tag{7.6}$$

$$H(P_T^v \,|\, Q_T^v) \ge \frac{1}{6}\operatorname{Tr}\Big(\Big(\frac{1}{T}\Sigma_T^v - I\Big)^2\Big), \qquad \forall T \le \log(2)/2\rho. \tag{7.7}$$

These inequalities will be the starting point for the proofs below. We will also make use of a Taylor expansion, used also within the proof of Theorem 2.10 (see (6.19)):

Lemma 7.2. *We have*

$$\frac{1}{T}\Sigma_T - I = \sum_{m=1}^\infty \frac{T^m}{(m+1)!}\Gamma_m, \qquad \textit{where} \quad \Gamma_m := \sum_{r=0}^m \binom{m}{r}\xi^r(\xi^\top)^{m-r}, \; m \in \mathbb{N}.$$

Proof. We have the Cauchy product identity

$$e^{t\xi}e^{t\xi^\top} = \left(\sum_{r=0}^\infty \frac{t^r}{r!}\xi^r\right)\left(\sum_{r=0}^\infty \frac{t^r}{r!}(\xi^\top)^r\right) = \sum_{m=0}^\infty \frac{t^m}{m!}\Gamma_m,$$

for $t \in \mathbb{R}$. Thus, using $\Gamma_0 = I$ and Fubini,

$$\frac{1}{T}\Sigma_T - I = \frac{1}{T}\int_0^T (e^{t\xi}e^{t\xi^\top} - I)\,dt = \frac{1}{T}\int_0^T \sum_{m=1}^\infty \frac{t^m}{m!}\Gamma_m\,dt = \sum_{m=1}^\infty \frac{T^m}{(m+1)!}\Gamma_m. \qquad \square$$

7.1. **Proof of Proposition 2.17.** Starting from (7.7) and applying Lemma 7.2,

$$H(P_T^v \,|\, Q_T^v) \geq \frac{1}{6}\mathrm{Tr}\bigg(\bigg(\sum_{m=1}^{\infty} \frac{T^m}{(m+1)!}\Gamma^v_m\bigg)^2\bigg) \geq \frac{T^2}{24}\mathrm{Tr}\big((\Gamma^v_1)^2\big),$$

where the second inequality follows from the fact that all entries of Γ_m are nonnegative. Using $\Gamma_1 = \xi + \xi^\top$,

$$\mathrm{Tr}\big((\Gamma^v_1)^2\big) = \sum_{i,j\in v}(\xi_{ij}+\xi_{ji})^2 \geq 2\sum_{i,j\in v}\xi_{ij}^2,$$

where we again used $\xi_{ij} \geq 0$. □

7.2. **Proof of Theorem 2.15.** We start from a general calculation for any symmetric $n\times n$ matrix A, where we write $\mathrm{avg}_{|v|=k}$ to denote the average over all choices of $v \subset [n]$ of cardinality k. That is, $\mathrm{avg}_{|v|=k} = \frac{1}{\binom{n}{k}}\sum_{v\subset[n]:|v|=k}$. Note for any indices $i,j\in[n]$ that

$$\mathop{\mathrm{avg}}_{|v|=k} 1_{i,j\in v} = \frac{k(k-1)}{n(n-1)}1_{i\neq j} + \frac{k}{n}1_{i=j} = \frac{k(k-1)}{n(n-1)} + \frac{k(n-k)}{n(n-1)}1_{i=j}.$$

This equation was used before. Cite instead of rewrite? This implies

$$\begin{aligned}\mathop{\mathrm{avg}}_{|v|=k} \mathrm{Tr}((A^v)^2) &= \mathop{\mathrm{avg}}_{|v|=k}\sum_{i,j\in v}A_{ij}^2 = \sum_{i,j=1}^n A_{ij}^2\big(\mathop{\mathrm{avg}}_{|v|=k} 1_{i,j\in v}\big) \\ &= \frac{k(k-1)}{n(n-1)}\mathrm{Tr}(A^2) + \frac{k(n-k)}{n(n-1)}\sum_{i=1}^n A_{ii}^2. \qquad (7.8)\end{aligned}$$

Using (7.6) and (7.7), we deduce that

$$\mathop{\mathrm{avg}}_{|v|=k} H(P_T^v \,|\, Q_T^v) \leq e^{6\rho T}\bigg(\frac{k(k-1)}{n(n-1)}\mathrm{Tr}((T^{-1}\Sigma_T - I)^2) + \frac{k(n-k)}{n(n-1)}\sum_{i=1}^n (T^{-1}(\Sigma_T)_{ii}-1)^2\bigg), \qquad (7.9)$$

$$\mathop{\mathrm{avg}}_{|v|=k} H(P_T^v \,|\, Q_T^v) \geq \frac{1}{6}\bigg(\frac{k(k-1)}{n(n-1)}\mathrm{Tr}((T^{-1}\Sigma_T - I)^2) + \frac{k(n-k)}{n(n-1)}\sum_{i=1}^n (T^{-1}(\Sigma_T)_{ii}-1)^2\bigg). \qquad (7.10)$$

It remains to express the right-hand sides in terms of ξ.

We start with the upper bound for the trace term. Let $(e_1,\ldots,e_n)$ denote the standard basis in $\mathbb{R}^n$. Using Lemma 7.2,

$$\begin{aligned}\mathrm{Tr}((T^{-1}\Sigma_T - I)^2) &= \sum_{i=1}^n |(T^{-1}\Sigma_T - I)e_i|^2 = \sum_{i=1}^n \bigg|\sum_{m=1}^\infty \frac{T^m}{(m+1)!}\Gamma_m e_i\bigg|^2 \qquad (7.11)\\ &\leq \sum_{i=1}^n\bigg(\sum_{m=1}^\infty \frac{T^m}{(m+1)!}|\Gamma_m e_i|\bigg)^2,\end{aligned}$$

To bound $|\Gamma_m e_i|^2$, we note first that for $0 \leq r \leq m$,

$$|\xi^r(\xi^\top)^{m-r}e_i| \leq \rho^{m-1}\big(|\xi^\top e_i|1_{r<m} + |\xi e_i|1_{r=m}\big).$$

Discarding the indicators, we find for $m \geq 1$ that

$$|\Gamma_m e_i| \leq \sum_{r=0}^m \binom{m}{r}\rho^{m-1}\big(|\xi^\top e_i| + |\xi e_i|\big) \leq 2^m\rho^{m-1}\big(|\xi^\top e_i| + |\xi e_i|\big).$$

Thus,

$$\begin{aligned}\mathrm{Tr}((T^{-1}\Sigma_T - I)^2) &\le \sum_{i=1}^n \bigg(\sum_{m=1}^\infty \frac{T^m}{(m+1)!} 2^m \rho^{m-1} \big(|\xi^\top e_i| + |\xi e_i| \big) \bigg)^2 \\ &\le 4T^2 e^{4\rho T} \sum_{i=1}^n \big(|\xi^\top e_i| + |\xi e_i| \big)^2 \\ &\le 16 T^2 e^{4\rho T} \sum_{i,j=1}^n \xi_{ij}^2. \end{aligned} \tag{7.12}$$

The lower bound for the trace term is similar: Using nonnegativity of the entries of Γ_m and ξ, from (7.11) we deduce

$$\begin{aligned}\mathrm{Tr}((T^{-1}\Sigma_T - I)^2) &\ge \sum_{i=1}^n \frac{T^2}{4} |\Gamma_1 e_i|^2 = \frac{T^2}{4} \sum_{i=1}^n |(\xi + \xi^\top) e_i|^2 \\ &\ge \frac{T^2}{4} \sum_{i=1}^n \big(|\xi e_i|^2 + |\xi^\top e_i|^2 \big) = \frac{T^2}{2} \sum_{i,j=1}^n \xi_{ij}^2. \end{aligned} \tag{7.13}$$

Let us next turn to upper bounding the $(\Sigma_T)_{ii}$ term in (7.9). We start from

$$\sum_{i=1}^n (T^{-1}(\Sigma_T)_{ii} - 1)^2 = \sum_{i=1}^n \bigg(\sum_{m=2}^\infty \frac{T^m}{(m+1)!} (\Gamma_m)_{ii} \bigg)^2 \tag{7.14}$$

where we note that the inner summation starts at $m = 2$ because $\Gamma_1 = \xi + \xi^\top$ is zero on the diagonal. For each i, $m \ge 2$, and $0 \le r \le m$, we have by Young's inequality

$$\begin{aligned}|\langle e_i, \xi^r (\xi^\top)^{m-r} e_i \rangle| &\le \begin{cases} \rho^{m-1} |\xi^\top e_i|^2, & \text{if } 0 < r < m, \\ \rho^{m-1} |\xi e_i| |\xi^\top e_i|, & \text{if } r \in \{0, m\} \end{cases} \\ &\le \rho^{m-1} \big(|\xi e_i|^2 + |\xi^\top e_i|^2 \big). \end{aligned}$$

Thus,

$$|(\Gamma_m)_{ii}| \le \sum_{r=0}^m \binom{m}{r} \rho^{m-1} \big(|\xi e_i|^2 + |\xi^\top e_i|^2 \big) = 2^m \rho^{m-1} \sum_{j=1}^n (\xi_{ij}^2 + \xi_{ji}^2).$$

This yields

$$\begin{aligned}\sum_{i=1}^n (T^{-1}(\Sigma_T)_{ii} - 1)^2 &\le \sum_{i=1}^n \bigg(\sum_{m=2}^\infty \frac{T^m}{(m+1)!} 2^m \rho^{m-1} \sum_{j=1}^n (\xi_{ij}^2 + \xi_{ji}^2) \bigg)^2 \\ &\le 4\rho^2 T^2 e^{4\rho T} \sum_{i=1}^n \bigg(\sum_{j=1}^n (\xi_{ij}^2 + \xi_{ji}^2) \bigg)^2. \end{aligned} \tag{7.15}$$

The lower bound for the $(\Sigma_T)_{ii}$ term is similar: Using nonnegativity of the entries of Γ_m and ξ, we discard all of the $m > 2$ terms in (7.14), and note that $\Gamma_2 = \xi^2 + 2\xi\xi^\top + (\xi^\top)^2$:

$$\sum_{i=1}^n (T^{-1}(\Sigma_T)_{ii} - 1)^2 \ge \sum_{i=1}^n \frac{T^4}{36} (\Gamma_2)_{ii}^2 = \frac{T^4}{18} \sum_{i=1}^n \big(\xi\xi^\top + \xi^2 \big)_{ii}^2 = \frac{T^4}{18} \sum_{i=1}^n \bigg(\sum_{j=1}^n (\xi_{ij}^2 + \xi_{ij}\xi_{ji}) \bigg)^2. \tag{7.16}$$

To complete the proof of the theorem, we plug (7.12) with (7.15) into (7.9) to get the upper bound, and we combine (7.13) with (7.16) into (7.10) to get the lower bound. $\square$

7.3. **Proof of Theorem 2.18.** As noted in Remark 6.3, the bounded row sum assumption implies $\rho \le 1$. Use (7.6) and Lemma 7.2 to write

$$H(P_T^v \,|\, Q_T^v) \le e^{6T} \sum_{i,j\in v} \bigg(\sum_{m=1}^\infty \frac{T^m}{(m+1)!} (\Gamma_m)_{ij} \bigg)^2.$$

We first note that every entry of ξ^r is bounded by δ, for each $r \in \mathbb{N}$. Indeed, this is true for $r = 1$ by definition of δ, and if we assume it is true for some r then we prove it for $r + 1$ by using the assumption that row sums of ξ are bounded by 1:

$$(\xi^{r+1})_{ij} = \sum_{k=1}^n \xi_{ik}(\xi^r)_{kj} \le \delta \sum_{k=1}^n \xi_{ik} \le \delta.$$

Similarly, every entry of $\xi^r(\xi^\top)^{m-r}$ is bounded by δ, for any integers $m \ge r \ge 0$ with $m \ge 1$. We deduce that $(\Gamma_m)_{ij} \le \delta 2^m$ for $m \ge 1$. Thus,

$$H(P_T^v \,|\, Q_T^v) \le e^{6T} \sum_{i,j\in v} \bigg(\sum_{m=1}^\infty \frac{T^m}{(m+1)!} \delta 2^m \bigg)^2 \le e^{10T}\delta^2|v|^2.$$

Appendix A. Proofs for Examples

A.1. **Convex potentials: Example 2.5.** Recall in this setting that $b_0^i(t,x) = -\nabla U(x)$ and $b^i(t,x,y) = -\nabla W(x-y)$ for all i. We need to check that Assumption (U) holds, which includes Assumption (A) in particular. Assumption (A)(i), on the wellposedness of the main SDE systems (2.1) and (2.2), follows from the Lipschitz continuity of $(\nabla U, \nabla W)$, the independent projection being discussed in [47, Proposition 4.1]. Assumption (A)(ii,iii) follow trivially from boundedness of ∇W, with $\gamma = 2\||\nabla W|^2\|_\infty$ and $M = 2\gamma$.

We turn next to Assumption (U)(iv). The assumed boundendess of ∇W immediately implies that $\nabla W(x-\cdot) \in L^1(Q_t^j)$ for all $(t,x) \in [0,\infty)\times\mathbb{R}^d$ and $j \in [n]$, as well as the local boundendess of $(t,x) \mapsto \langle Q_t^j, \nabla W(x-\cdot)\rangle$. Finally, for the integrability requirements (2.5), note that the assumed LSI for Q_0 implies that Q_0 has finite moments of every order. It was shown in [47, Proposition 4.1] that Lipschitz coefficients finite moments at time zero lead to the moment bound $\sup_{t\in[0,T]} \mathbb{E}|Y_t^j|^p < \infty$ for any $p \ge 1$, $T > 0$ and $j \in [n]$. Similarly, the Lipschitz continuity of $(\nabla U, \nabla W)$ and the assumption that P_0 has finite moments of all orders implies the moment bound $\sup_{t\in[0,T]} \mathbb{E}|X_t^j|^p < \infty$ for any $p \ge 1$, $T > 0$ and $j \in [n]$. The the integrability requirements (2.5) are then consequences of the linear growth of $(\nabla U, \nabla W)$.

We lastly explain why the LSI of Assumption (U)(ii) holds. The independent projection (2.2) can be written as

$$dY_t^i = \Big(-\nabla U(Y_t^i) - \sum_{j\neq i} \xi_{ij} \nabla W * Q_t^j(Y_t^i) \Big) dt + \sigma dW_t^i, \quad i \in [n]. \tag{A.1}$$

Fix $i \in [n]$. The drift of Y_t^i at time t is the gradient of the function

$$\Psi_t(x) = U(x) + \sum_{j\neq i} \xi_{ij} W * Q_t^j(x), \qquad x \in \mathbb{R}^d,$$

which is easily checked to satisfy $\nabla^2\Psi_t(x) \ge \nabla^2 U(x) \succeq \lambda I$, using the assumed λ-convexity of U and convexity of W. This verifies the curvature condition of [53, Proposition 3.12] and we can follow the arguments therein to deduce that Q_t^i satisfies a LSI with constant

$$\frac{\sigma^2}{\lambda}(1 - e^{-4\lambda t/\sigma^2}) + \frac{\eta_0}{4} e^{-4\lambda t/\sigma^2} \le \max(\eta_0/4, \sigma^2/\lambda) =: \eta.$$

A.2. **Models on the torus: Example 2.6.** Checking Assumption (U) in this example is almost the same as in the proof of [48, Corollary 2.9], and we just sketch the main differences. The well-posedness Assumption (A)(i) is straightforward, as are Assumption (A)(ii,iii) and (U)(iv) by the boundedness of K. The only changes are in checking the LSI, Assumption (U), and mainly identifying the constant therein. To this end, we give the following lemma, adapted from [48, Corollary 2.9], which in turn borrowed key ideas from the proofs of [17, Proposition 3.1] and [35, Theorem 2].

Lemma A.1. *For each $t > 0$ and $i \in [n]$, the density of Q^i_t is C^2 and obeys the pointwise bound*

$$\frac{1}{\lambda e^r} \le Q^i_T(x) \le \frac{\lambda}{1 - re^r}, \quad \text{where} \quad r = \frac{\sqrt{2\log\lambda}\|\mathrm{div}K\|_\infty}{2\sigma^2\pi^2 - \|\mathrm{div}K\|_\infty}.$$

Moreover, it holds that $r < 1/2$, and $Q_t = Q^1_t \otimes \cdots \otimes Q^n_t$ satisfies the LSI

$$H(\cdot\,|\,Q^i_t) \le \eta I(\cdot\,|\,Q^i_t), \quad \text{where} \quad \eta := \lambda^2(1-2r)^{-1}(8\pi^2)^{-1}. \tag{A.2}$$

$$\begin{aligned}\eta &= \frac{\lambda^2}{8\pi^2}\left(1 - \frac{\sqrt{2\log\lambda}\|\mathrm{div}K\|_\infty}{2\sigma^2\pi^2 - \|\mathrm{div}K\|_\infty}\right)^{-1} \\ &= \frac{\lambda^2(2\sigma^2\pi^2 - \|\mathrm{div}K\|_\infty)}{8\pi^2\big(2\sigma^2\pi^2 - (1+\sqrt{2\log\lambda})\|\mathrm{div}K\|_\infty\big)}\end{aligned}$$

Proof.
Step 1. We first adapt the argument of [17, Proposition 3.1] to show that for each $t > 0$ and $i \in [n]$

$$H(Q^i_t\,|\,\mathbf{1}) \le e^{-2ct}\log\lambda, \quad c := 2\sigma^2\pi^2 - \|\mathrm{div}K\|_\infty. \tag{A.3}$$

Note that Q^i_t satisfies the Fokker-Planck equation $\partial_t Q^i = -\mathrm{div}(b^i Q^i) + (\sigma^2/2)\Delta Q^i$ with $b^i_t = \sum_j \xi_{ij} K * Q^j_t$. Letting $\mathbf{1}$ denote the uniform (Lebesgue) measure on $\mathbb{T}^d$, a standard computation followed by integration by parts yields

$$\frac{d}{dt}H(Q^i_t\,|\,\mathbf{1}) = -\int Q^i_t\,\mathrm{div}b^i_t - \frac{\sigma^2}{2}\int Q^i_t|\nabla\log Q^i_t|^2.$$

Using the log-Sobolev inequality for the uniform measure on $\mathbb{T}^d$, we have

$$\int Q^i_t|\nabla\log Q^i_t|^2 \ge 8\pi^2 H(Q^i_t\,|\,\mathbf{1}). \tag{A.4}$$

Indeed, see [31] for proof of this LSI in dimension $d = 1$, which tensorizes to general dimension. Using the form of b,

$$\begin{aligned}-\int Q^i_t\,\mathrm{div}b^i_t &= -\sum_{j\neq i}\xi_{ij}\int Q^i_t\,\mathrm{div}K * Q^j_t = -\sum_{j\neq i}\xi_{ij}\int (Q^i_t - \mathbf{1})\mathrm{div}K * (Q^j_t - \mathbf{1}) \\ &\le \|Q^i_t - \mathbf{1}\|_{\mathrm{TV}}\|\mathrm{div}K\|_\infty\sum_{j\neq i}\xi_{ij}\|Q^j_t - \mathbf{1}\|_{\mathrm{TV}}.\end{aligned}$$

Combining the three previous displays and using Gronwall's inequality,

$$e^{4\sigma^2\pi^2 t}H(Q^i_t\,|\,\mathbf{1}) \le H(Q^i_0\,|\,\mathbf{1}) + \|\mathrm{div}K\|_\infty\int_0^t e^{4\sigma^2\pi^2 s}\|Q^i_s - \mathbf{1}\|_{\mathrm{TV}}\sum_j \xi_{ij}\|Q^j_s - \mathbf{1}\|_{\mathrm{TV}}\,ds.$$

Letting $\widehat{H}_t = \max_{i\in[n]} H(Q^i_t\,|\,\mathbf{1})$ and using Pinsker's inequality along with $\sum_j \xi_{ij} \le 1$, we deduce

$$e^{4\sigma^2\pi^2 t}\widehat{H}_t \le \widehat{H}_0 + 2\|\mathrm{div}K\|_\infty\int_0^t e^{4\sigma^2\pi^2 s}\widehat{H}_s\,ds.$$

Applying Gronwall's inequality again, along with the assumption $Q_0^i \le \lambda$, we find

$$\widehat{H}_t \le e^{(2\|\mathrm{div}K\|_\infty - 4\sigma^2\pi^2)t}\widehat{H}_0 \le e^{(2\|\mathrm{div}K\|_\infty - 4\sigma^2\pi^2)t}\log\lambda.$$

Step 2. We next prove the pointwise bound on Q_t^i. Fix $T > 0$ and $x_i \in \mathbb{T}^d$, and let $(Z^i)_{t\in[0,T]}$ be unique strong solution of the SDE system

$$dZ_t^i = -\sum_{j\neq i}\xi_{ij}K * Q_{T-t}^j(Z_t^i)dt + dW_t^i, \quad Z_0^i = x_i.$$

Using Ito's formula and the Fokker-Planck equation for Q^i, and taking expectations, we have

$$\mathbb{E}\left[Q_{T-t}^i(Z_t^i)\right] = Q_T^i(x) + \sum_j \xi_{ij}\mathbb{E}\int_0^t Q_{T-s}^i(Z_s^i)\,\mathrm{div}K * Q_{T-s}^j(Z_s^i)ds, \quad t\in[0,T]. \tag{A.5}$$

Noting that $\mathrm{div}K * \mathbf{1} \equiv 0$, we have for any $u\in[0,T]$ that

$$\|\mathrm{div}K * Q_u^j\|_\infty \le \|\mathrm{div}K * (Q_u^j - \mathbf{1})\|_\infty \le \|\mathrm{div}K\|_\infty\|Q_u^j - \mathbf{1}\|_{\mathrm{TV}} \le \sqrt{2\log\lambda}\|\mathrm{div}K\|_\infty e^{-cu},$$

where the last step uses Pinsker's inequality and (A.3). Setting $a = \sqrt{2\log\lambda}\|\mathrm{div}K\|_\infty$, and using $\sum_j \xi_{ij} \le 1$, this implies

$$\mathbb{E}\left[Q_{T-t}^i(Z_t^i)\right] \le Q_T^i(x) + a\int_0^t e^{-c(T-s)}\mathbb{E}\left[Q_{T-s}^i(Z_s^i)\right]ds.$$

By Gronwall's inequality, we obtain for $t\in[0,T]$

$$\mathbb{E}\left[Q_{T-t}^i(Z_t^i)\right] \le Q_T^i(x)\exp\left(ae^{-cT}\int_0^t e^{cs}ds\right) \le Q_T^i(x)e^{a/c}.$$

Setting $t = T$ and using the lower bound $Q_0^i \ge \lambda^{-1}$ yields

$$Q_T^i(x) \ge e^{-a/c}\mathbb{E}\left[Q_0^i(Z_t^i)\right] \ge e^{-a/c}\lambda^{-1}.$$

Similarly, using (A.5), we can deduce

$$\begin{aligned} Q_T^i(x) &\le \mathbb{E}\left[Q_0^i(Z_t^i)\right] + a\int_0^T e^{-c(T-s)}\mathbb{E}[Q_{T-s}^i(Z_s^i)]ds \\ &\le \lambda + aQ_T^i(x)e^{a/c}\int_0^T e^{-c(T-s)}\,ds. \end{aligned}$$

Therefore,

$$Q_T^i(x) \le \lambda\big/\big(1 - (a/c)e^{a/c}\big).$$

Combining gives us the claimed bounds on the density. It was assumpted in (2.6) that $\|\mathrm{div}K\|_\infty < 2\sigma^2\pi^2/(1+\sqrt{2\log\lambda})$, which ensures that $r < 1/2$. Lastly, by noting that

$$\frac{\sup Q_t^i}{\inf Q_t^i} \le \frac{\lambda^2 e^r}{1 - re^r} \le \frac{\lambda^2}{1-2r}. \tag{A.6}$$

we have by Holley-Stroock [4, Proposition 5.1.6] that Q_t^i satisfies the claimed LSI. □

Remark A.2. The above proof corrects two small errors in the argument of [48, Corollary 2.9]. First, the constant c (and thus the denominator of r) was missing the factor of 2, carrying forward a typo from [17, Equation (3.3)] in which the constant in the LSI (A.4) was misquoted as $4\pi^2$ instead of $8\pi^2$. Second, the factor $(8\pi^2)^{-1}$ was missing from η in [48, Corollary 2.9], due to a misapplication of Holley-Stroock at the end of the proof.

Appendix B. Concentration bounds for Erdös-Rényi graphs

The following proposition was referenced in Sections 3.2 and 3.3, which discussed choices of interaction matrix ξ derived from realizations of Erdös-Rényi graphs. The following proposition shows how to justifies the claims in those sections, that with high probability those examples of ξ satisfy the assumption of having row and column sums which are bounded with respect to n.

Proposition B.1. *Let A^n be the adjacency matrix of an Erdös-Rényi graph on n vertices with edge probability p_n.*

(1) If $\liminf_{n\to\infty} np_n/\log n > 1$, then there exists a constant β such that

$$\lim_{n\to\infty} \mathbb{P}\left(\max_{j\in[n]} \sum_{i=1}^n \frac{A^n_{ij}}{\sum_{k\neq i} A^n_{ik}} > \beta\right) = 0. \tag{B.1}$$

By convention, we set $A^n_{ij}/\sum_{k\neq i} A^n_{ik} = 0$ if the denominator is zero.

(2) If $\liminf_{n\to\infty} np_n/\log n > 0$, then there exists constant β such that

$$\lim_{n\to\infty} \mathbb{P}\bigg(\max_{i\in[n]} \sum_{j=1}^n \frac{A^n_{ij}}{(n-1)p_n} > \beta\bigg) = 0. \tag{B.2}$$

Proof. The proofs will make use of the multiplicative form of Chernoff's inequalities [67, Theorem 2.3.1, Exercise 2.3.2]: For $\delta > 0$ and for $X \sim$ Binomial(n,p), we have

$$\mathbb{P}(X > np(1+\delta)) \le e^{-nph_+(\delta)}, \qquad h_+(\delta) := (1+\delta)\log(1+\delta) - \delta, \tag{B.3}$$

$$\mathbb{P}(X < np(1-\delta)) \le e^{-nph_-(\delta)}, \qquad h_-(\delta) := (1-\delta)\log(1-\delta) + \delta. \tag{B.4}$$

(1) Let $0 < \varepsilon < 1$ and $\beta > 1/(1-\varepsilon)$, so that $\delta := (1-\varepsilon)\beta - 1$ is positive. Define the events

$$S_i := \Big\{\sum_{k=1}^n A^n_{ik} < (n-1)p_n(1-\varepsilon)\Big\}, \qquad S := \bigcup_{i=1}^n S_i,$$

$$B_j := \Big\{\sum_{i=1}^n \frac{A^n_{ij}}{\sum_{k=1}^n A^n_{ik}} > \beta\Big\}, \qquad B := \bigcup_{j=1}^n B_j.$$

Note that $\sum_{k\neq i} A^n_{ik} \sim$ Binomial($n-1$,p_n). Using the Chernoff bound (B.4), we have

$$\mathbb{P}(S_i) \le \exp\big(-(n-1)p_n h_-(\varepsilon)\big).$$

Using (B.3) and $(1-\varepsilon)\beta = 1+\delta$, for each $j \in [n]$ we have

$$\mathbb{P}(B_j \cap S^c) \le \mathbb{P}\Big(\sum_{i=1}^n A^n_{ij} > (n-1)p_n(1-\varepsilon)\beta\Big) \le \exp\big(-(n-1)p_n h_+(\delta)\big).$$

Using the union bound, we then deduce

$$\begin{aligned}\mathbb{P}(B) &\le \mathbb{P}(S) + \sum_{j=1}^n \mathbb{P}(B_j \cap S^c)\\ &\le n\exp\big(-(n-1)p_n h_-(\varepsilon)\big) + n\exp\big(-(n-1)p_n h_+(\delta)\big).\end{aligned}$$

Let $\eta > 0$. For sufficiently large n we have $(n-1)p_n \ge (1+\eta)\log n$ and thus

$$\mathbb{P}(B) \le n^{1-(1+\eta)h_-(\varepsilon)} + n^{1-(1+\eta)h_+(\delta)}.$$

Since $h_-(\varepsilon) \to 1$ as $\varepsilon \to 1$, we may choose ε so that $h_-(\varepsilon) > 1/(1+\eta)$. Once ε is chosen, we may take β (and hence δ) as large as we want, to make $h_+(\delta)$ as large as desired, and we finally deduce that $\mathbb{P}(B) \to 0$.

(2) Let $a > 0$. Applying union bound and (B.3),

$$\mathbb{P}\Big(\max_{i\in[n]}\sum_{j=1}^{n}\frac{A_{ij}^{n}}{(n-1)p_n} > 1+a\Big) \le \sum_{i=1}^{n}\mathbb{P}\Big(\sum_{j=1}^{n}\frac{A_{ij}^{n}}{(n-1)p_n} > 1+a\Big) \le n\exp\big(-(n-1)p_n h_+(a)\big).$$

Noting that $h_+(a)$ can be made arbitrarily large by choosing a large, the claim follows.

□

References

1. Antonio Auffinger, Michael Damron, and Jack Hanson, *50 years of first-passage percolation*, vol. 68, American Mathematical Soc., 2017.
2. Nathalie Ayi, Nastassia Pouradier Duteil, and David Poyato, *Mean-field limit of non-exchangeable multi-agent systems over hypergraphs with unbounded rank*, arXiv preprint arXiv:2406.04691 (2024).
3. Ágnes Backhausz and Balázs Szegedy, *Action convergence of operators and graphs*, Canadian Journal of Mathematics **74** (2022), no. 1, 72–121.
4. Dominique Bakry, Ivan Gentil, Michel Ledoux, et al., *Analysis and geometry of markov diffusion operators*, vol. 103, Springer, 2014.
5. Erhan Bayraktar, Suman Chakraborty, and Ruoyu Wu, *Graphon mean field systems*, The Annals of Applied Probability **33** (2023), no. 5, 3587–3619.
6. Erhan Bayraktar and Ruoyu Wu, *Stationarity and uniform in time convergence for the graphon particle system*, Stochastic Processes and their Applications **150** (2022), 532–568.
7. G. Ben Arous and O. Zeitouni, *Increasing propagation of chaos for mean field models*, Annales de l'Institut Henri Poincare (B) Probability and Statistics, vol. 35, Elsevier, 1999, pp. 85–102.
8. Itai Benjamini and Oded Schramm, *Recurrence of distributional limits of finite planar graphs*, Selected Works of Oded Schramm (2011), 533–545.
9. Gianmarco Bet, Fabio Coppini, and Francesca Romana Nardi, *Weakly interacting oscillators on dense random graphs*, Journal of Applied Probability **61** (2024), no. 1, 255–278.
10. Shankar Bhamidi, Amarjit Budhiraja, and Ruoyu Wu, *Weakly interacting particle systems on inhomogeneous random graphs*, Stochastic Processes and their Applications **129** (2019), no. 6, 2174–2206.
11. Christian Borgs, Jennifer Chayes, Henry Cohn, and Yufei Zhao, *An l^p theory of sparse graph convergence I: Limits, sparse random graph models, and power law distributions*, Transactions of the American Mathematical Society **372** (2019), no. 5, 3019–3062.
12. Christian Borgs, Jennifer T. Chayes, Henry Cohn, and Yufei Zhao, *An L^p theory of sparse graph convergence II: LD convergence, quotients and right convergence*, The Annals of Probability **46** (2018), no. 1, 337 – 396.
13. Didier Bresch, Pierre-Emmanuel Jabin, and Juan Soler, *A new approach to the mean-field limit of Vlasov-Fokker-Planck equations*, arXiv preprint arXiv:2203.15747 (2022).
14. Didier Bresch, Pierre-Emmanuel Jabin, and Zhenfu Wang, *Mean field limit and quantitative estimates with singular attractive kernels*, Duke Mathematical Journal **172** (2023), no. 13, 2591–2641.
15. Pierre Le Bris and Christophe Poquet, *A note on uniform in time mean-field limit in graphs*, arXiv preprint arXiv:2211.11519 (2022).
16. Gerard Brunick and Steven Shreve, *Mimicking an Itô process by a solution of a stochastic differential equation*, The Annals of Applied Probability **23** (2013), no. 4, 1584 – 1628.
17. JA Carrillo, RS Gvalani, GA Pavliotis, and A Schlichting, *Long-time behaviour and phase transitions for the mckean–vlasov equation on the torus*, Archive for Rational Mechanics and Analysis **235** (2020), no. 1, 635–690.
18. Patrick Cattiaux, *Entropy on the path space and application to singular diffusions and mean-field models*, arXiv preprint arXiv:2404.09552 (2024).
19. Louis-Pierre Chaintron and Antoine Diez, *Propagation of chaos: a review of models, methods and applications. II. Applications*, arXiv preprint arXiv:2106.14812 (2021).
20. ———, *Propagation of chaos: a review of models, methods and applications. I. Models and methods*, arXiv preprint arXiv:2203.00446 (2022).
21. Hayato Chiba and Georgi S Medvedev, *The mean field analysis for the kuramoto model on graphs i. the mean field equation and transition point formulas*, arXiv preprint arXiv:1612.06493 (2016).
22. Fabio Coppini, Helge Dietert, and Giambattista Giacomin, *A law of large numbers and large deviations for interacting diffusions on Erdős–Rényi graphs*, Stochastics and Dynamics **20** (2020), no. 02, 2050010.
23. Antonin Chodron de Courcel, Matthew Rosenzweig, and Sylvia Serfaty, *The attractive log gas: Stability, uniqueness, and propagation of chaos*, arXiv preprint arXiv:2311.14560 (2023).

24. Antonin Chodron de Courcel, Matthew Rosenzweig, and Sylvia Serfaty, *Sharp uniform-in-time mean-field convergence for singular periodic Riesz flows*, Annales de l'Institut Henri Poincaré C (2023).
25. Sylvain Delattre, Giambattista Giacomin, and Eric Luçon, *A note on dynamical models on random graphs and Fokker–Planck equations*, Journal of Statistical Physics **165** (2016), 785–798.
26. Amir Dembo, Thomas M Cover, and Joy A Thomas, *Information theoretic inequalities*, IEEE Transactions on Information theory **37** (1991), no. 6, 1501–1518.
27. Kai Du, Yifan Jiang, and Xiaochen Li, *Sequential propagation of chaos*, arXiv preprint arXiv:2301.09913 (2023).
28. Mitia Duerinckx, *Mean-field limits for some Riesz interaction gradient flows*, SIAM Journal on Mathematical Analysis **48** (2016), no. 3, 2269–2300.
29. Murray Eden, *A two-dimensional growth process*, Dynamics of fractal surfaces **4** (1961), 223–239.
30. Gábor Elek and Balázs Szegedy, *A measure-theoretic approach to the theory of dense hypergraphs*, Advances in Mathematics **231** (2012), no. 3-4, 1731–1772.
31. Michel Émery and Joseph E Yukich, *A simple proof of the logarithmic Sobolev inequality on the circle*, Séminaire de probabilités de Strasbourg **21** (1987), 173–175.
32. Ernesto Estrada and Naomichi Hatano, *Communicability in complex networks*, Physical Review E **77** (2008), no. 3, 036111.
33. Marios Antonios Gkogkas and Christian Kuehn, *Graphop mean-field limits for Kuramoto-type models*, SIAM Journal on Applied Dynamical Systems **21** (2022), no. 1, 248–283.
34. Nathael Gozlan and Christian Léonard, *Transport inequalities. a survey*, arXiv preprint arXiv:1003.3852 (2010).
35. Arnaud Guillin, Pierre Le Bris, and Pierre Monmarché, *Uniform in time propagation of chaos for the 2D vortex model and other singular stochastic systems*, Journal of the European Mathematical Society (2024), 1–28.
36. Elias Hess-Childs and Keefer Rowan, *Higher-order propagation of chaos in l^2 for interacting diffusions*, arXiv preprint arXiv:2310.09654 (2023).
37. Pierre-Emmanuel Jabin, David Poyato, and Juan Soler, *Mean-field limit of non-exchangeable systems*, arXiv preprint arXiv:2112.15406 (2021).
38. Pierre-Emmanuel Jabin and Zhenfu Wang, *Mean field limit and propagation of chaos for Vlasov systems with bounded forces*, Journal of Functional Analysis **271** (2016), no. 12, 3588–3627.
39. ———, *Quantitative estimates of propagation of chaos for stochastic systems with $w^{-1,\infty}$ kernels*, Inventiones mathematicae **214** (2018), 523–591.
40. Pierre-Emmanuel Jabin and Datong Zhou, *The mean-field limit of sparse networks of integrate and fire neurons*, arXiv preprint arXiv:2309.04046 (2023).
41. Jean-Francois Jabir, *Rate of propagation of chaos for diffusive stochastic particle systems via Girsanov transformation*, arXiv preprint arXiv:1907.09096 (2019).
42. Joe Jackson and Daniel Lacker, *Approximately optimal distributed stochastic controls beyond the mean field setting*, arXiv preprint arXiv:2301.02901 (2023).
43. Joe Jackson and Antonios Zitridis, *Concentration bounds for stochastic systems with singular kernels*, arXiv preprint arXiv:2406.02848 (2024).
44. Christian Kuehn and Chuang Xu, *Vlasov equations on digraph measures*, Journal of Differential Equations **339** (2022), 261–349.
45. Daniel Lacker, *On a strong form of propagation of chaos for McKean-Vlasov equations*, Electronic Communications in Probability **23** (2018), 1 – 11.
46. Daniel Lacker, *Hierarchies, entropy, and quantitative propagation of chaos for mean field diffusions*, 2022.
47. Daniel Lacker, *Independent projections of diffusions: Gradient flows for variational inference and optimal mean field approximations*, arXiv preprint arXiv:2309.13332 (2023).
48. Daniel Lacker and Luc Le Flem, *Sharp uniform-in-time propagation of chaos*, Probability Theory and Related Fields (2023), 1–38.
49. Daniel Lacker, Kavita Ramanan, and Ruoyu Wu, *Local weak convergence for sparse networks of interacting processes*, The Annals of Applied Probability **33** (2023), no. 2, 843–888.
50. ———, *Marginal dynamics of interacting diffusions on unimodular Galton–Watson trees*, Probability Theory and Related Fields **187** (2023), no. 3, 817–884.
51. Daniel Lacker and Agathe Soret, *A case study on stochastic games on large graphs in mean field and sparse regimes*, Mathematics of Operations Research **47** (2022), no. 2, 1530–1565.
52. László Lovász, *Large networks and graph limits*, vol. 60, American Mathematical Soc., 2012.
53. Florient Malrieu, *Logarithmic sobolev inequalities for some nonlinear pde's*, Stochastic processes and their applications **95** (2001), no. 1, 109–132.
54. Georgi S Medvedev, *The nonlinear heat equation on dense graphs and graph limits*, SIAM Journal on Mathematical Analysis **46** (2014), no. 4, 2743–2766.
55. ———, *The continuum limit of the kuramoto model on sparse random graphs*, arXiv preprint arXiv:1802.03787 (2018).

56. Yuliya Mishura and Alexander Veretennikov, *Existence and uniqueness theorems for solutions of McKean–Vlasov stochastic equations*, Theory of Probability and Mathematical Statistics **103** (2020), 59–101.
57. Pierre Monmarché, Zhenjie Ren, and Songbo Wang, *Time-uniform log-Sobolev inequalities and applications to propagation of chaos*, arXiv preprint arXiv:2401.07966 (2024).
58. Mark Newman, *Networks*, Oxford university press, 2018.
59. Roberto I Oliveira and Guilherme H Reis, *Interacting diffusions on random graphs with diverging average degrees: Hydrodynamics and large deviations*, Journal of Statistical Physics **176** (2019), no. 5, 1057–1087.
60. Roberto I. Oliveira, Guilherme H. Reis, and Lucas M. Stolerman, *Interacting diffusions on sparse graphs: hydrodynamics from local weak limits*, Electronic Journal of Probability **25** (2020), 1 – 35.
61. Felix Otto and Cédric Villani, *Generalization of an inequality by talagrand and links with the logarithmic sobolev inequality*, Journal of Functional Analysis **173** (2000), no. 2, 361–400.
62. Jan Peszek and David Poyato, *Heterogeneous gradient flows in the topology of fibered optimal transport*, Calculus of Variations and Partial Differential Equations **62** (2023), no. 9, 258.
63. Kavita Ramanan, *Interacting stochastic processes on sparse random graphs*, arXiv preprint arXiv:2401.00082 (2023).
64. Daniel Richardson, *Random growth in a tessellation*, Mathematical Proceedings of the Cambridge Philosophical Society, vol. 74, Cambridge University Press, 1973, pp. 515–528.
65. Matthew Rosenzweig and Sylvia Serfaty, *Modulated logarithmic Sobolev inequalities and generation of chaos*, arXiv preprint arXiv:2307.07587 (2023).
66. Sylvia Serfaty, *Mean field limit for Coulomb-type flows*, Duke Mathematical Journal **169** (2020), no. 15, 2887 – 2935.
67. Roman Vershynin, *High-dimensional probability: An introduction with applications in data science*, vol. 47, Cambridge university press, 2018.
68. Songbo Wang, *Sharp local propagation of chaos for mean field particles with $w^{-1,\infty}$ kernels*, arXiv preprint arXiv:2403.13161 (2024).
69. Zhenfu Wang, Xianliang Zhao, and Rongchan Zhu, *Mean-field limit of non-exchangeable interacting diffusions with singular kernels*, arXiv preprint arXiv:2209.14002 (2022).

Department of Industrial Engineering & Operations Research, Columbia University