Supplementary Material to "Statistical Windows in Testing for the Initial Distribution of a Reversible Markov Chain"

A DISTANCES AND DIVERGENCES BETWEEN DISTRIBUTIONS

We recall the following notions, for two distributions μ, μ' on a finite set $\mathcal{X}$. The total variation distance, denoted by d_{TV}, is defined as

$$d_{\mathsf{TV}}(\mu, \mu') = \frac{1}{2} \sum_{x \in \mathcal{X}} |\mu_x - \mu'_x| = \max_{X \subseteq \mathcal{X}} |\mu(X) - \mu'(X)| \,.$$

The Kullback–Leibler divergence, denoted by KL, is defined as

$$\mathsf{KL}(\mu, \mu') = \sum_{x \in \mathcal{X}} \mu_x \log\left(\frac{\mu_x}{\mu'_x}\right) \,.$$

The chi-square divergence, denoted by D_{χ^2}, is defined as

$$D_{\chi^2}(\mu, \mu') = \sum_{x \in \mathcal{X}} \mu'_x \left(\frac{\mu_x}{\mu'_x} - 1\right)^2 \,.$$

The Hellinger distance, denoted by H, is defined by

$$H^2(\mu, \mu') = \sum_{x \in \mathcal{X}} \mu'_x \left(\sqrt{\frac{\mu_x}{\mu'_x}} - 1\right)^2 \,.$$

We will also make frequent use of the following standard inequalities (see e.g. **?**).

$$\frac{1}{2} H^2(\mu, \mu') \leq d_{\mathsf{TV}}(\mu, \mu') \leq H(\mu, \mu')$$

$$d_{\mathsf{TV}}(\mu, \mu') \leq \sqrt{\frac{1}{2}\mathsf{KL}(\mu, \mu')} \qquad \text{(Pinsker's Inequality)}$$

B PROOFS

B.1 Proofs from Section 3SAMPLE COMPLEXITYsection.3

of Theorem 6theorem.6. Recalling that $\mu_t = \mu P^t$, the following classical result holds, for μ and μ' be distributions over $\mathcal{X}$.

$$\mu = \pi + \sum_{i=2}^{d} \langle u_i, \mu \rangle_\pi u_i \quad \text{and } \mu' = \pi + \sum_{i=2}^{d} \langle u_i, \mu' \rangle_\pi u_i \,,$$

Indeed, μ and μ' are treated as vectors in $\mathbf{R}^d$ and can be expressed in terms of the basis vectors $u_1, \dots, u_d$. The coefficients are given by $\alpha_i = \langle \mu, u_i \rangle_\pi$ (resp. $\alpha'_i = \langle \mu', u_i \rangle_\pi$). As μ and μ' are distributions, Lemma 4theorem.4 give us that $\alpha_1 = \alpha'_1 = 1$.

As the u_i's are eigenvectors, we have $\mu_t = \mu P^t = \pi + \sum_{i=2}^d \alpha_i \lambda_i^t u_i$ and $\mu'_t = \mu' P^t = \pi + \sum_{i=2}^d \alpha'_i \lambda_i^t u_i$. Then, by using the orthonormality of u_i with respect to $\langle \cdot, \cdot \rangle_\pi$, we have:

$$\|\mu_t - \mu'_t\|_\pi^2 = \sum_{i=2}^{d} \lambda_i^{2t} (\langle u_i, \mu\rangle_\pi - \langle u_i, \mu'\rangle_\pi)^2$$

This norm between μ_t and μ'_t can be compared to other notions of distances between distributions. In particular, this can be done for the Hellinger distance between μ_t and μ'_t. Let $\mu_{t,x}/\mu'_{t,x} =: 1 + \gamma_{t,x}$. It then holds that

$$\begin{aligned} H^2(\mu_t, \mu'_t) &= \sum_x \mu'_{t,x}(1 - \sqrt{1+\gamma_{t,x}})^2 \\ &= 2\sum_x \mu'_{t,x}\Big(1 + \frac{\gamma_{t,x}}{2} - \sqrt{1+\gamma_{t,x}}\Big) \\ &\geq \frac{\varepsilon^{3/2}}{8} \sum_x \mu'_{t,x}\gamma_{t,x}^2 \quad \text{(see below)} \\ &= \frac{\varepsilon^{3/2}}{8} \sum_x \frac{(\mu_{t,x} - \mu'_{t,x})^2}{\mu'_{t,x}} \end{aligned}$$

The first inequality can be justified in the following way: Define the function f_c on $[-1, \infty)$ by

$$f_c(t) = 1 + t/2 - ct^2 - \sqrt{1+t}\,.$$

Basic computations yield that for all positive c, $f_c(0) = 0$ and $f'_c(0) = 0$. Furthermore, we have that

$$f''_c(t) = -2c + \frac{1}{4(1+t)^{3/2}}\,.$$

As a consequence, f_c is convex in $t \in [\varepsilon - 1, 1/\varepsilon - 1]$ for $c = \varepsilon^{3/2}/8$, so $f_c(t) \geq 0$ on this interval. We therefore have that $1 + \frac{\gamma_{t,x}}{2} - \sqrt{1+\gamma_{t,x}} \geq \frac{\varepsilon^{3/2}}{8}\gamma_{t,x}^2$ for $\mu_{t,x}/\mu'_{t,x} \in [\varepsilon, 1/\varepsilon]$. As a consequence, it holds that

$$\begin{aligned} H^2(\mu_t, \mu'_t) &\geq \frac{\varepsilon^{5/2}}{8} \sum_x \frac{(\mu_{t,x} - \mu'_{t,x})^2}{\pi_x} \\ &= \frac{\varepsilon^{5/2}}{8} \|\mu_t - \mu'_t\|_\pi^2 \end{aligned}$$

We use this property to control directly the probability of error of the likelihood-ratio test ψ_{LR}

$$\begin{aligned} \max_{\nu \in \{\mu, \mu'\}} \mathbf{P}_{\nu_t}^{\otimes n}(\psi_{\mathsf{LR}} \neq \nu) &\leq \mathbf{P}_{\mu_t}^{\otimes n}(\psi_{\mathsf{LR}} \neq \mu) + \mathbf{P}_{\mu'_t}^{\otimes n}(\psi_{\mathsf{LR}} \neq \mu') \\ &= 1 - \big(\mathbf{P}_{\mu'_t}^{\otimes n}(\psi_{\mathsf{LR}} \neq \mu) - \mathbf{P}_{\mu_t}^{\otimes n}(\psi_{\mathsf{LR}} \neq \mu)\big) \\ &= 1 - d_{\mathsf{TV}}(\mu_t^{\otimes n}, \mu_t'^{\otimes n}) \end{aligned}$$

Indeed, one of the event X realizing the total variation distance between the two distributions is the one on which $\mu_t'^{\otimes n}\big((X_i)_{i\in[n]}\big)$ is greater or equal than $\mu_t^{\otimes n}\big((X_i)_{i\in[n]}\big)$, i.e. where the output of ψ_{LR} is μ'. We have, by properties of the Hellinger distance, that

$$1 - d_{\mathsf{TV}}(\mu_t^{\otimes n}, \mu_t'^{\otimes n}) \leq 1 - \frac{1}{2}H^2(\mu_t^{\otimes n}, \mu_t'^{\otimes n}) = \Big(1 - \frac{1}{2}H^2(\mu_t, \mu'_t)\Big)^n\,.$$

The last equality allows, by tensorization, to relate directly this probability of error to a quantity depending separately on (μ_t, μ'_t), and n. As a consequence, we have

$$\max_{\nu \in \{\mu, \mu'\}} \mathbf{P}_{\nu_t}^{\otimes n}(\psi_{\mathsf{LR}} \neq \nu) \leq \Big(1 - \frac{\varepsilon^{5/2}}{16}\|\mu_t - \mu'_t\|_\pi^2\Big)^n \leq e^{-n\frac{\varepsilon^{5/2}}{16}\|\mu_t - \mu'_t\|_\pi^2}$$

As a consequence, for $n \geq 16\varepsilon^{-5/2}\log(1/\delta)/\|\mu_t - \mu'_t\|_\pi^2$, the probability of error is indeed less than δ. $\square$

of Theorem 7theorem.7. We recall from the proof of Theorem 6theorem.6 that

$$\|\mu_t - \mu'_t\|_\pi^2 = \sum_{i=2}^{d} \lambda_i^{2t} (\langle u_i, \mu\rangle_\pi - \langle u_i, \mu'\rangle_\pi)^2 \,.$$

We compare this distance to the Kullback–Leibler divergence $\mathsf{KL}(\mu_t, \mu'_t)$. Let $\mu_{t,x}/\mu'_{t,x} =: 1 + \gamma_{t,x}$; as $\sum_x \mu_{t,x} = \sum_x \mu'_{t,x}$, it follows that $\sum_x \mu'_{t,x}\gamma_{t,x} = 0$. Then consider the following:

$$\begin{aligned}
\mathsf{KL}(\mu_t, \mu'_t) &= \sum_x \mu_{t,x} \ln \frac{\mu_{t,x}}{\mu'_{t,x}} = \sum_x \mu'_{t,x}(1+\gamma_{t,x}) \ln(1+\gamma_{t,x}) \\
&\leq \sum_x \mu'_{t,x}(1+\gamma_{t,x})\gamma_{t,x} && \text{Using the fact that } \ln(1+t) \leq t \\
&= \sum_x \mu'_{t,x}\gamma_{t,x}^2 && \text{As } \sum_x \mu'_x \gamma_x = 0 \\
&= \sum_x \frac{(\mu_{t,x} - \mu'_{t,x})^2}{\mu'_{t,x}} \\
&\leq \frac{1}{\varepsilon} \sum_x \frac{(\mu_{t,x} - \mu'_{t,x})^2}{\pi_x} && \text{As } \frac{1}{\mu'_{t,x}} \leq \frac{1}{\varepsilon \pi_x} \\
&= \frac{1}{\varepsilon} \|\mu_t - \mu'_t\|_\pi^2 \,.
\end{aligned}$$

To give a lower bound on the probability of error, we have

$$\inf_\psi \max_{\nu \in \mu, \mu'} \mathbf{P}_{\nu_t}^{\otimes n}(\psi \neq \nu) \geq \inf_\psi \frac{1}{2}(\mathbf{P}_{\mu_t}^{\otimes n}(\psi \neq \mu) + \mathbf{P}_{\mu'_t}^{\otimes n}(\psi \neq \mu')) \geq \frac{1 - d_{\mathsf{TV}}(\mu_t^{\otimes n}, \mu_t'^{\otimes n})}{2}$$

The above holds by using the definition of total variation distance as the supremum of the difference in probability for all events, and using the event $\psi = \mu$, with an infimum taken over all tests ψ. Furthermore, by Pinsker's inequality and by the tensorization properties of the Kullback–Leibler divergence, we have that

$$d_{\mathsf{TV}}(\mu_t^{\otimes n}, \mu_t'^{\otimes n}) \leq \sqrt{\mathsf{KL}(\mu_t^{\otimes n}, \mu_t'^{\otimes n})/2} = \sqrt{n\mathsf{KL}(\mu_t, \mu'_t)/2} \,.$$

As a consequence, it holds that

$$\inf_\psi \max_{\nu \in \mu, \mu'} \mathbf{P}_{\nu_t}^{\otimes n}(\psi \neq \nu) \geq \frac{1}{2} - \frac{1}{2}\sqrt{\frac{n}{2\varepsilon}\|\mu_t - \mu'_t\|_\pi^2} \,.$$

For any $n \leq 8\varepsilon\delta^2/\|\mu_t - \mu'_t\|_\pi^2$, the probability of error is at least $1/2 - \delta$. □

B.2 Proofs from Section 3.4Guarantees without likelihood-ratio boundssubsection.3.4

of Theorem 10theorem.10. As in the proof of Theorem 6theorem.6, it holds that

$$\max_{\nu \in \{\mu, \mu'\}} \mathbf{P}_{\nu_t}^{\otimes n}(\psi_{\mathsf{LR}} \neq \nu) \leq 1 - d_{\mathsf{TV}}(\mu_t^{\otimes n}, \mu_t'^{\otimes n}) \leq 1 - \frac{1}{2}H^2(\mu_t^{\otimes n}, \mu_t'^{\otimes n}) = \Big(1 - \frac{1}{2}H^2(\mu_t, \mu'_t)\Big)^n \,.$$

Furthermore, we have that

$$\begin{aligned}
H^2(\tilde\mu_t, \tilde\mu'_t) &= \sum_x \left(\sqrt{\tilde\mu_{t,x}} - \sqrt{\tilde\mu'_{t,x}}\right)^2 \\
&= \sum_x (\sqrt{(1-\eta)\mu_{t,x} + \eta\beta_{t,x}} - \sqrt{(1-\eta)\tilde\mu'_{t,x} + \eta\beta_{t,x}})^2 \\
&\leq (1-\eta) \sum_x \left(\sqrt{\mu_{t,x}} - \sqrt{\mu'_{t,x}}\right)^2 \\
&= (1-\eta) H^2(\mu_t, \mu'_t) \,.
\end{aligned}$$

Indeed the inequality is a consequence of $\sqrt{a+c}-\sqrt{b+c} \leq \sqrt{a}-\sqrt{b}$, for $a \geq b \geq 0$ and $c \geq 0$. As a consequence, we have that

$$\max_{\nu \in \{\mu,\mu'\}} \mathbf{P}_{\nu_t}^{\otimes n}(\psi_{\mathsf{LR}} \neq \nu) \leq \Big(1-\frac{1}{2}H^2(\mu_t,\mu_t')\Big)^n \leq \Big(1-\frac{1}{2(1-\eta)}H^2(\tilde\mu_t,\tilde\mu_t')\Big)^n .$$

By the proof of Theorem 6theorem.6, we therefore have that

$$\max_{\nu \in \{\mu,\mu'\}} \mathbf{P}_{\nu_t}^{\otimes n}(\psi_{\mathsf{LR}} \neq \nu) \leq \Big(1-\frac{1}{2(1-\eta)}H^2(\tilde\mu_t,\tilde\mu_t')\Big)^n \leq \Big(1-\frac{(\eta/3)^{5/2}}{16(1-\eta)}\|\tilde\mu_t-\tilde\mu_t'\|_\pi^2\Big)^n .$$

Indeed, since $\tilde\mu$ and $\tilde\mu'$ have $\eta/3$-bounded likelihood ratios, so do $\tilde\mu_t$ and $\tilde\mu_t'$. Further, by linearity, it holds that

$$\|\tilde\mu_t-\tilde\mu_t'\|_\pi^2 = (1-\eta)^2\|\mu_t-\mu_t'\|_\pi^2 .$$

As a consequence, we finally have

$$\max_{\nu \in \{\mu,\mu'\}} \mathbf{P}_{\nu_t}^{\otimes n}(\psi_{\mathsf{LR}} \neq \nu) \leq \Big(1-\frac{(\eta/3)^{5/2}(1-\eta)}{16}\|\mu_t-\mu_t'\|_\pi^2\Big)^n \leq e^{-cn\|\mu_t-\mu_t'\|_\pi^2} ,$$

with $c > 0$ for any choice of $\eta \in (0,1)$. □

B.3 Proofs from Section 3.5Statistical time guaranteessubsection.3.5

of Theorem 11theorem.11. Let $u_{[i]}$ be the left eigenvector of P corresponding to eigenvalue $\lambda_{[i]}$. Let $\mu = \pi + \alpha u_{[i]}$ and $\mu' = \pi - \alpha u_{[i]}$, where $\alpha > 0$ is sufficiently small so that μ and μ' are valid probability distributions and μ, μ', π pairwise satisfy the bounded likelihood ratio assumption with parameter ε. Let $n_0(\varepsilon, \delta) > 0$ be the sample complexity required to distinguish between μ and μ' with probability greater than $\frac{1}{2}+\delta$.

Let $n \geq n_0$ be the sample complexity required to distinguish between distributions μ_t and μ_t'. Using Theorems 6theorem.6 and 7theorem.7, we know that $n \asymp \|\mu_t-\mu_t'\|_\pi^{-2}$ and $n_0 \asymp \|\mu-\mu'\|_\pi^{-2}$. By definition, we have that $\mu-\mu' = 2\alpha u_{[i]}$ and as a result $\|\mu-\mu'\|_\pi = 2\alpha$ and $\|\mu_t-\mu_t'\|_\pi = 2\alpha\lambda_{[i]}^t$. Thus, we have:

$$\frac{n}{n_0} \asymp \left(\frac{\|\mu-\mu'\|_\pi}{\|\mu_t-\mu_t'\|_\pi}\right)^2 = \lambda_{[i]}^{-2t}$$

We can invert the above to express t in terms of n, n_0 and $\lambda_{[i]}$ to get the required result, which is tight up to terms involving only ε and δ. □

B.4 Proofs from Section 4APPLICATIONSsection.4

Proof Sketch of Proposition 4. We write $P = D^{-1}U$, where D is a diagonal matrix with $D_{ii} = \sum_{x \in cX} U_{i,x}$. Writing m_U to be the mean of $U_{i,j}$, which we treat to be a constant, standard concentration inequalities imply that $\|D\| \asymp \|D^{-1}\| \asymp m_U d$. It follows from (**?**, Theorem 1.2) that $|\lambda_{[2]}(P)| \asymp \frac{1}{\sqrt{d}}$.

To establish bounds on $|\lambda_{[d]}|$, we consider $P^{-1} = U^{-1}D$ (it is known that P is non-singular with probability $1-o(1)$), then $|\lambda_{[d]}|^{-1} \leq \|D\| \cdot \|U^{-1}\|$. Using a result of **?**, we know that $\|U^{-1}\| = O(\sqrt{d})$, which gives us $|\lambda_{[d]}^{-1}| = O(m_U d^{3/2})$.

For a lower bound, we have $\|P^{-1}\| \geq \|U^{-1}\| \cdot (\|D^{-1}\|)^{-1}$. By using universality results for random matrices, it is known that $\|U^{-1}\| \geq \sqrt{d}$. (This can be achieved by subtracting a rank one matrix and using interlacing results for eigenvalues, cf. **?**.) This together with the bounds on $\|D^{-1}\|$ establishes that $|\lambda_{[d]}^{-1}| = \Omega(m_U d^{3/2})$ □

of Proposition 6. In the case of two blocks, we simplify the notation a bit and assume that $\Delta_{1,1} = \Delta_{2,2} = a \cdot d$ and $\Delta_{1,2} = \Delta_{2,1} = b \cdot d$ The, degree of every vertex in the graph is $(a+b)d$. Let $V_1 \cup V_2$ be a partition of the set of nodes into the two parts. Consider $v : V \to \mathbf{R}$, where $v(i) = 1$ if $i \in V_1$ and -1 if $i \in V_2$. It is easily checked that v is an eigenvector with eigenvalue $(a-b)/(a+b)$. This shows that $\lambda_{[2]} \geq \frac{a-b}{a+b}$.

Now, consider a vector $v : V \to \mathbf{R}$, with $v(i) = \frac{1}{\sqrt{2}}$, $v(j) = -\frac{1}{\sqrt{2}}$ and $v(l) = 0$ for $l \neq i, j$. Note that $\|v\|_2 = 1$. Let $u = Pv$, then, we have:

$$u(l) = \begin{cases} \frac{1}{d(a+b)\sqrt{2}} & \text{if } \{l,i\} \in E \wedge \{l,j\} \notin E \\ -\frac{1}{d(a+b)\sqrt{2}} & \text{if } \{l,j\} \in E \wedge \{l,i\} \notin E \\ 0 & \text{otherwise} \end{cases}$$

If we let $N(i)$ and $N(j)$ denote the neighborhoods of i and j respectively, we have that,

$$\|u\|^2 = \frac{|N(i)\Delta N(j)|}{2d^2(a+b)^2} \qquad |S\Delta T| \text{ denotes the symmetric difference}$$

We have that $\|Pu\|^2 \geq |\lambda_{[d]}|$. If we set $\Delta_{1,1} = ad = d/2 - O(1)$, there exist i, j such that $|N(i)\Delta N(j)| = O(1)$; further, if we set $b = o(1)$, then the result follows. □

of Proposition 15theorem.15. The matrix P is symmetric, one can check directly that the given vectors are indeed eigenvectors with corresponding eigenvalues, i.e. that $Pv_i = \lambda_i v_i$. Finally, we have for $2 \leq k \leq r$,

$$\gamma_k - \gamma_{k+1} = 2\beta_{r+1-k} - \beta_{r+2-k} > 0\,,$$

by order of the β_ℓ. For $k = 1$, we have $\gamma_1 - \gamma_2 = 2\beta_r$ and $\gamma_{r+1} = \beta_0 - \beta_1 > 0$. □

C FURTHER APPLICATIONS

C.1 Random Walk on the Line Graph

The random walk on the line graph with d nodes is very similar to that on the cycle. In fact, the walk can be viewed as a projection of the random walk on a cycle with $2(d-1)$ nodes.

Definition 1 (Random Walk on d-Line)**.** *Let $\mathcal{X} = \{0, \ldots, d-1\}$ be the d nodes on a line. Let P be the Markov Chain on $\mathcal{X}$, where,*

$$\mathbf{P}(X_s = i | X_{s-1} = j) = \begin{cases} \frac{1}{2} & \text{for } i \in j-1, j+1, j \in \{1, \ldots, d-2\} \\ 1 & \text{if } j = 0 \wedge i = 1, \text{ or } j = d-1 \wedge i = d-2 \\ 0 & \text{otherwise} \end{cases}$$

As in the case of the cycle, the spectrum is explicitly known (cf. (**?**, Chap. 12.3)). The result is stated as the following lemma; this implies a statistical window of at least d^{2t}.

Lemma 2. *For any $d \geq 3$, the eigenvalues of P are given by $\cos(\pi i/(d-1))$ for $i \in \{0, \ldots, d-1\}$; the right eigenvector $u_i = (u_{i,0}, \ldots, u_{i,d-1})$ corresponding to eigenvalue $\cos(\pi i/(d-1))$, is given by $u_{i,k} = \cos(\pi i k/(d-1))$.*

C.2 Random Markov chains

The notion of random Markov chains, and in general of random walks in random media, has been thoroughly studied. We consider here the case of random reversible Markov chains, as studied by **??**.

Definition 3. *Consider a finite connected undirected graph $G = (\mathcal{X}, E)$; for every $\{i, j\} \in E$, let $U_{i,j}$ be drawn in an i.i.d. manner from a distribution on the positive part of the real line with bounded second moments; we set $U_{i,j} = U_{j,i}$. (An edge $\{i\}$ is allowed, which corresponds to a self-loop at i.) The coefficients of P are obtained by normalization to a stochastic matrix*

$$P_{i,j} = U_{i,j} / \sum_{x \in \mathcal{X}} U_{i,x}\,.$$

Understanding the spectra of random (symmetric) matrices has been intensely studied in recent years (see e.g. **????**). Below, we use results from this literature to understand the behavior of $\lambda_{[2]}(P)$ and $\lambda_{[d]}(P)$; again, this yields a statistical window of d^{2t}.

Proposition 4. *Let P be the transition matrix of a random Markov chain as defined in Defn. 3 with $G(\mathcal{X}, E)$ being the complete graph. As $d \to \infty$ d, it holds with probability going to 1 that*

$$|\lambda_{[2]}(P)| \asymp 1/\sqrt{d}\,, \quad \text{and} \quad |\lambda_{[d]}(P)| \asymp 1/d^{3/2}\,,$$

C.3 Random Walk on the Regular Block Model

We use a variant of the stochastic blockmodel (**?**) where the graph is regular (as opposed to approximately regular). Note that the model is completely deterministic.

Definition 5 (Regular Blockmodel)**.** *A regular blockmodel with k blocks on d nodes with degrees $(\Delta_{i,j})$ for $1 \le i,j \le k$ and $\Delta_{i,j} = \Delta_{j,i}$ is defined as follows: The vertex set V is partitioned as $V = V_1 \cup V_2 \cup \cdots \cup V_k$, with $|V_i| = d/k$. The induced subgraph $G_i = (V_i, E(V_i))$ is a Δ_i regular graph for each i, and the subgraph $G_{i,j} = (V_i \cup V_j, E(V_i \cup V_j) \setminus (E(V_i) \cup E(V_j)))$ is a $\Delta_{i,j}$ regular bipartite graph for all i, j, $i \neq j$.*

Proposition 6. *There exist regular block models with $k = 2$ blocks on d nodes, satisfying*

$$|\lambda_{[2]}(P)| = 1 - o(1)\,, \quad \textit{and} \quad |\lambda_{[d]}(P)| \asymp 1/d$$

For blockmodels with $k = 2$, the eigenvector corresponding to $\lambda_{[2]} = \lambda_2$, correlates strongly (in fact for the regular blockmodel with equal sized blocks, exactly), with the block structure. Thus, if μ and μ' start off with significantly different probability mass on the two blocks, the statistical problem remains easy essentially until mixing time. On the other hand, if they have the same probability mass on the individual blocks (even though the distributions may differ on the blocks significantly), in typical cases, the statistical problem becomes hard quickly, e.g. if each block is an expander.

Remark 7. *A special case of regular graphs is the class of Ramanujan graphs (**?**); the eigenvalues of the transition matrix of a random walk on a Ramanujan graph are ± 1 or satisfy $|\lambda| = O(1/\sqrt{d})$. In the non-bipartite case, there is no guarantee that the statistical window is large.*

C.4 Random Walk on the Hypercube

The hypercube on $d = 2^k$ nodes can be viewed as a graph with the node set denoted by $\{-1,1\}^k$. We first consider the standard random walk on the hypercube which is defined below; for $x, x' \in \{-1,1\}^k$, let $|x - x'|_H := \frac{1}{2}\|x - x'\|_1$ denote the Hamming distance between x and x'.

Definition 8. *Let $\mathcal{X} = \{-1,1\}^k$ be the d nodes of the hypercube. Let P be the Markov chain on $\mathcal{X}$, where*

$$\mathbf{P}(X_s = x | X_{s-1} = x') = \begin{cases} \frac{1}{d} & \textit{if } |x - x'|_H = 1 \\ 0 & \textit{otherwise} \end{cases}$$

The spectral properties of P are summarized by the following lemma (c.f. **?**, Chap 12.4); this implies a statistical window of at least d^{2t}.

Lemma 9. *For $d = 2^k$ (with $k \ge 1$), the eigenvalues of P are given by $1 - 2j/k$, $j = 0, \ldots, k$. The eigenvalue $1 - 2j/k$ appears with multiplicity $\binom{k}{j}$; the corresponding eigenvectors u are given by considering sets of size j. A set $S \subseteq \{1, \ldots, k\}$, $|S| = j$, yields the eigenvector $u(x) = \prod_{i \in S} x_i$, for $x \in \{-1,1\}^k$. (These are the so-called parity functions.)*

Product Distributions on the Hypercube. One can consider other product distributions on the hypercube. Consider a two state Markov chain, with states denoted by $\mathcal{X}_2 = \{-1,1\}$ and transitions given as:

Definition 10. *Let $\mathcal{X}_2 = \{-1,1\}$, $0 < p, q \le 1$, define $P : \mathcal{X}_2 \times \mathcal{X}_2 \to [0,1]$ as $P(-1,-1) = 1 - p$, $P(-1,1) = p$, $P(1,-1) = q$, $P(1,1) = 1 - q$. The Markov chain on $\mathcal{X}_2$ defined by P is given by,*

$$\mathbf{P}(X_s = x | X_{s-1} = x') = P(x, x')$$

The eigenvectors of P are easiest to express as functions from $\mathcal{X}_2 \to \mathbf{R}$. The right eigenvector corresponding to eigenvalue 1 is given by $u(x) = 1$ for $x \in \mathcal{X}_2$. The stationary distribution is given by $\pi(-1) = q/(p+q)$ and $\pi(1) = p/(p+q)$, let $\xi = \mathbf{E}_{x \sim \pi}[x] = (p-q)/(p+q)$. The second eigenvalue is $1 - (p+q)$ and the corresponding eigenvector is given by $u(x) = (x - \xi)/\sqrt{1 - \xi^2}$.

Definition 11 (Product Chain on the Hypercube)**.** *Let $d = 2^k$, $\mathcal{X} = \{-1,1\}^k = \mathcal{X}_2 \otimes \cdots \otimes \mathcal{X}_2$, let $P^{(1)}, \ldots, P^{(k)}$ be transition matrices of the chain on $\mathcal{X}_2$ defined in Defn. 10 with parameters $(p^{(1)}, q^{(1)}), \ldots, (p^{(k)}, q^{(k)})$, and let $w_1, \ldots, w_k$ be positive weights such that $\sum_i w_k = 1$. Then, for $x, x' \in \mathcal{X}$, we have the following Markov chain:*

$$\mathbf{P}(X_s = x | X_{s-1} = x') = \begin{cases} \sum_{j=1}^{k} w_j P^{(j)}(x_j, x'_j) & \textit{if } |x - x'|_H = 1 \\ 0 & \textit{otherwise} \end{cases}$$

The eigenvectors and eigenvalues of the product Markov chain on the hypercube are easily defined through the eigenvectors and eigenvalues of Markov chain defined on $\mathcal{X}_2$. The following lemma follows from results stated in (**?**, Chap 12.4).

Lemma 12. *Let P be the transition matrix of the product Markov chain obtained using the transition matrices $P^{(1)}, \dots, P^{(k)}$ of chains on $\mathcal{X}_2$. Let $u^{(i)}$ denote the eigenvector of $P^{(i)}$ with eigenvalue $1 - (p^{(i)} + q^{(i)})$, then for each subset $S \subseteq \{1, \dots, k\}$, define $u_S : \mathcal{X} \to \mathbf{R}$ as follows:*

$$u_S(x) = \prod_{i \in S} u^{(i)}(x_i) = \prod_{i \in S} \frac{x_i - \xi^{(i)}}{\sqrt{1 - (\xi^{(i)})^2}}$$

where $\xi^{(i)} = (p^{(i)} - q^{(i)})/(p^{(i)} + q^{(i)})$. Then u_S is an eigenvector of P with eigenvalue $1 - \sum_{i \in S} w_i(p^{(i)} + q^{(i)})$.

Remark 13. *It is easily observed that if we set $p^{(i)} = q^{(i)} = 1$ in all the chains and $w_i = \frac{1}{k}$ for each i, then we get exactly the standard random walk on the hypercube with $d = 2^k$ vertices.*