7 Appendix: quasi-Monte Carlo graph random features (q-GRFs)

7.1 On the approximation of the d-regularised Laplacian using GRFs

In this appendix, we demonstrate how to approximate the d-regularised Laplacian $\mathbf{K}_{\text{lap}}^{(d)}$ with GRFs.

Recall that GRFs provide an estimator to the quantity $(\mathbf{I}_N - \mathbf{U})^{-2}$ where $\mathbf{U}$ is a weighted adjacency matrix. Recall also that the matrix elements of the symmetrically normalised Laplacian $\widetilde{\mathbf{L}}$ are given by

$$\widetilde{\mathbf{L}}_{ij} = \begin{cases} 1 & \text{if } i = j, \\ -\frac{\mathbf{W_{ij}}}{\sqrt{\deg_{\mathbf{W}}(i)\deg_{\mathbf{W}}(j)}} & \text{if } i \sim j. \end{cases} \tag{24}$$

where $\deg_{\mathbf{W}}(i) = \sum_{j \in \mathcal{V}} \mathbf{W}_{ij}$ is the weighted degree of the node i. We are typically interested in situations where $\mathbf{W} = \mathbf{A}$, an unweighted adjacency matrix. Now note that

$$\mathbf{K}_{\text{lap}\,ij}^{(2)} = (\mathbf{I}_N + \sigma^2 \widetilde{\mathbf{L}})_{ij}^{-2} = (1+\sigma^2)^{-2} \left(\mathbf{I}_N - \mathbf{U}\right)_{ij}^{-2} \tag{25}$$

where we defined the matrix $\mathbf{U}$ with matrix elements

$$\mathbf{U}_{ij} = \frac{\sigma^2}{1+\sigma^2} \frac{\mathbf{W_{ij}}}{\sqrt{\deg_{\mathbf{W}}(i)\deg_{\mathbf{W}}(j)}}. \tag{26}$$

This is itself a weighted adjacency matrix, as required. It follows that, by estimating $(\mathbf{I}_N - \mathbf{U})^{-2}$ with GRFs, we can trivially estimate $\mathbf{K}_{\text{lap}}^{(2)}$. This was reported in [Choromanski, 2023].

Supposing that we have constructed a low-rank GRF estimator

$$\mathbf{K}_{\text{lap}}^{(2)} = \mathbb{E}\left[\mathbf{C}\mathbf{C}^\top\right] \tag{27}$$

where the matrix $\mathbf{C} \in \mathbb{R}^{N \times N}$ has rows $\mathbf{C}_i := \frac{1}{1+\sigma^2}\phi(i)^\top$, we note that it is straightforward to construct the 1-regularised Laplacian kernel estimator

$$\mathbf{K}_{\text{lap}}^{(1)} = \mathbb{E}\left[\mathbf{C}\mathbf{D}^\top\right] \tag{28}$$

by taking $\mathbf{D} := \left(\mathbf{I}_N + \sigma^2 \widetilde{\mathbf{L}}\right)^\top \mathbf{C}$. It is then trivial to obtain the estimator $\mathbf{K}_{\text{lap}}^{(d)}$ for arbitrary $d \in \mathbb{N}$.

7.2 Derivation of Eq. 15

In this appendix we derive Eq. 15, which gives the expected length of some walk ω_2 given that its antithetic partner ω_1 is of length m: that is, $\mathbb{E}\left(\text{len}(\omega_2)|\text{len}(\omega_1) = m\right)$.

As a warm-up, consider the simpler marginal expected lengths. Note that

$$p\left(\text{len}(\omega) = m\right) = (1-p)^m p. \tag{29}$$

It follows that

$$\mathbb{E}\left(\text{len}(\omega)\right) = \sum_{m=0}^{\infty} m(1-p)^m p = \frac{1-p}{p} \tag{30}$$

where we computed the arithmetic-geometric series. We reported this result in Eq. 14. Meanwhile, the probability of a walk being of length i given that its antithetic partner is of length m is

$$p(\text{len}(\omega_2) = i|\text{len}(\omega_1) = m) = \begin{cases} \left(\frac{1-2p}{1-p}\right)^i \frac{p}{1-p} & \text{if } i < m, \\ 0 & \text{if } i = m, \\ \left(\frac{1-2p}{1-p}\right)^m (1-p)^{i-m-1} p & \text{if } i > m. \end{cases} \tag{31}$$

The analagous sum then becomes

$$\mathbb{E}\left(\text{len}(\omega_2)|\text{len}(\omega_1) = m\right) = \sum_{i=0}^{m} \left(\frac{1-2p}{1-p}\right)^i \frac{p}{1-p} i + \sum_{i=m+1}^{\infty} \left(\frac{1-2p}{1-p}\right)^m (1-p)^{i-m-1} p i. \tag{32}$$

After straightforward but tedious algebra, this evaluates to

$$\mathbb{E}\left(\text{len}(\omega_2)|\text{len}(\omega_1)=m\right)=\frac{1-2p}{p}+2\left(\frac{1-2p}{1-p}\right)^m, \tag{33}$$

as stated in Eq. 15. Note that this is greater than $\mathbb{E}(\text{len}(\omega))$ when m is small and smaller than $\mathbb{E}(\text{len}(\omega))$ when m is large; the two walk lengths are negatively correlated.

7.3 On the superiority of q-GRFs (proof of Theorem 3.2)

Here, we provide a proof of the central result of Theorem 3.2: that the introduction of antithetic termination reduces the variance of estimators of the matrix $(\mathbf{I}_N-\mathbf{U})^{-2}$. From App. 7.1, all our results will trivially extend to the 2-regularised Laplacian kernel $\mathbf{K}_\text{lap}^{(2)}$.

Notation: to reduce the burden of summation indices, we have used Dirac's bra-ket notation from quantum mechanics. $|y\rangle$ can be interpreted as the vector $\boldsymbol{y}$ and $\langle y|$ as $\boldsymbol{y}^\top$.

We will begin by assuming that the graph is d-regular, that all edges have equal weights denoted w, and that our sampling strategy involves the random walker choosing one of its neighbours with equal probability at each timestep. We will relax these assumptions in App. 7.4.

We have seen that antithetic termination does not modify the walkers' marginal termination behaviour, so the variance of the estimator $\boldsymbol{\phi}(i)^\top\boldsymbol{\phi}(j)$ is only affected via the second-order term $\mathbb{E}\left[(\boldsymbol{\phi}(i)^\top\boldsymbol{\phi}(j))^2\right]$. Writing out the sums,

$$\begin{aligned}(\boldsymbol{\phi}(i)^\top\boldsymbol{\phi}(j))^2=\frac{1}{m^4}\sum_{x,y\in\mathcal{V}}\sum_{k_1,l_1,k_2,l_2=1}^{m}\sum_{\omega_1\in\Omega_{ix}}\sum_{\omega_2\in\Omega_{jx}}\sum_{\omega_3\in\Omega_{iy}}\sum_{\omega_4\in\Omega_{jy}}\frac{\widetilde{\omega}(\omega_1)}{p(\omega_1)}\frac{\widetilde{\omega}(\omega_2)}{p(\omega_2)}\frac{\widetilde{\omega}(\omega_3)}{p(\omega_3)}\frac{\widetilde{\omega}(\omega_4)}{p(\omega_4)}\\ \cdot\,\mathbb{I}(\omega_1\in\bar{\Omega}(k_1,i))\mathbb{I}(\omega_2\in\bar{\Omega}(l_1,j))\mathbb{I}(\omega_3\in\bar{\Omega}(k_2,i))\mathbb{I}(\omega_4\in\bar{\Omega}(l_2,j)).\end{aligned} \tag{34}$$

To remind the reader: the variables x,y sum over the nodes of the graph $\mathcal{V}$. k_1 and l_1 enumerate all the m walks sampled out of node i, whilst k_2 and l_2 enumerate walks from j. The sum over $\omega_1\in\Omega_{ix}$ is over all possible walks between nodes i and x. $\widetilde{\omega}(\omega_1)$ evaluates the product of edge weights traversed by the walk ω_1, which is $w^{\text{len}(\omega_1)}$ in the equal-weights case (with $\text{len}(\omega_1)$ denoting the number of edges in ω_1). $p(\omega_1)$ is the *marginal* probability of the subwalk ω_1, which is equal to $((1-p)/d)^{\text{len}(\omega_1)}$ on a d-regular graph. Lastly, the indicator function $\mathbb{I}(\omega_1\in\bar{\Omega}(k_1,i))$ evaluates to 1 if the k_1th walk out of node i (denoted $\bar{\Omega}(k_1,i)$) contains the walk ω_1 as a subwalk and 0 otherwise.

We immediately note that our scheme only every correlates walks leaving the same node, so walks out of different nodes remain independent. Therefore,

$$\begin{aligned}&\mathbb{E}\left[\mathbb{I}(\omega_1\in\bar{\Omega}(k_1,i))\mathbb{I}(\omega_2\in\bar{\Omega}(l_1,j))\mathbb{I}(\omega_3\in\bar{\Omega}(k_2,i))\mathbb{I}(\omega_4\in\bar{\Omega}(l_2,j))\right]\\&\quad=p(\omega_1\in\bar{\Omega}(k_1,i),\omega_3\in\bar{\Omega}(k_2,i))p(\omega_2\in\bar{\Omega}(l_1,j),\omega_4\in\bar{\Omega}(l_2,j)).\end{aligned} \tag{35}$$

Consider the term in the sum corresponding to one particular set of walks (k_1,l_1,k_2,l_2),

$$\begin{aligned}\sum_{x,y\in\mathcal{V}}\sum_{\omega_1\in\Omega_{ix}}\sum_{\omega_2\in\Omega_{jx}}\sum_{\omega_3\in\Omega_{iy}}\sum_{\omega_4\in\Omega_{jy}}\frac{\widetilde{\omega}(\omega_1)}{p(\omega_1)}\frac{\widetilde{\omega}(\omega_2)}{p(\omega_2)}\frac{\widetilde{\omega}(\omega_3)}{p(\omega_3)}\frac{\widetilde{\omega}(\omega_4)}{p(\omega_4)}\\ \cdot\,p(\omega_1\in\bar{\Omega}(k_1,i),\omega_3\in\bar{\Omega}(k_2,i))p(\omega_2\in\bar{\Omega}(l_1,j),\omega_4\in\bar{\Omega}(l_2,j)).\end{aligned} \tag{36}$$

This object will be of central importance and is referred to as the *correlation term*. In the sum over k_1,k_2,l_1,l_2, there are three possibilities to consider. We stress again that $k_{1,2}$ refers to a pair of walks out of node i and $l_{1,2}$ refers to a pair out of j.

- **Case 1**, **same-same**, $\boldsymbol{k_1=k_2,l_1=l_2}$: the pair of walks out of i are identical and the pair of walks out of j are identical. This term will not be modified by antithetic coupling since the marginal walk behaviour is unmodified and walks out of different nodes remain independent.
- **Case 2**, **different-different**, $\boldsymbol{k_1\neq k_2,l_1\neq l_2}$: the walks out of both i and j differ, and each pair may be antithetic or independent. This term will be modified by the coupling.
- **Case 3**, **same-different**. $\boldsymbol{k_1=k_2,l_1\neq l_2}$: the walks out of i differ – and may exhibit antithetic or independent termination – but the walks out of j are the same. This term will be modified by the coupling. Note that the i and j labels are arbitrary so we have chosen one ordering for concreteness.

If we can reason that the contributions from each of these possibilities $1-3$ either remains the same or is reduced by the introduction of antithetic coupling, then from Eq. 34 we can conclude that the entire sum and therefore the Laplacian kernel estimator variance is suppressed. For completeness, we write out the entire sum from Eq. 34 with the degeneracy factors below:

$$
\begin{aligned}
(\boldsymbol{\phi}(i)^\top \boldsymbol{\phi}(j))^2 = \frac{1}{m^4} \sum_{x,y \in \mathcal{V}} \Big\{ & \\
& \left.\begin{array}{l} m^2 \sum_{\omega_1 \in \Omega_{ix}} \sum_{\omega_2 \in \Omega_{jx}} \sum_{\omega_3 \in \Omega_{iy}} \sum_{\omega_4 \in \Omega_{jy}} \frac{\widetilde{\omega}(\omega_1)}{p(\omega_1)} \frac{\widetilde{\omega}(\omega_2)}{p(\omega_2)} \frac{\widetilde{\omega}(\omega_3)}{p(\omega_3)} \frac{\widetilde{\omega}(\omega_4)}{p(\omega_4)} \\ \cdot \mathbb{I}(\omega_1 \in \bar{\Omega}(k_1,i)) \mathbb{I}(\omega_2 \in \bar{\Omega}(l_1,j)) \mathbb{I}(\omega_3 \in \bar{\Omega}(k_1,i)) \mathbb{I}(\omega_4 \in \bar{\Omega}(l_1,j)) \end{array}\right\} \text{same-same (1)} \\
& \left.\begin{array}{l} + m^2(m-1)^2 \sum_{\omega_1 \in \Omega_{ix}} \sum_{\omega_2 \in \Omega_{jx}} \sum_{\omega_3 \in \Omega_{iy}} \sum_{\omega_4 \in \Omega_{jy}} \frac{\widetilde{\omega}(\omega_1)}{p(\omega_1)} \frac{\widetilde{\omega}(\omega_2)}{p(\omega_2)} \frac{\widetilde{\omega}(\omega_3)}{p(\omega_3)} \frac{\widetilde{\omega}(\omega_4)}{p(\omega_4)} \\ \cdot \mathbb{I}(\omega_1 \in \bar{\Omega}(k_1,i)) \mathbb{I}(\omega_2 \in \bar{\Omega}(l_1,j)) \mathbb{I}(\omega_3 \in \bar{\Omega}(k_2,i)) \mathbb{I}(\omega_4 \in \bar{\Omega}(l_2,j)) \end{array}\right\} \text{different-different (2)} \\
& \left.\begin{array}{l} + m^2(m-1) \sum_{\omega_1 \in \Omega_{ix}} \sum_{\omega_2 \in \Omega_{jx}} \sum_{\omega_3 \in \Omega_{iy}} \sum_{\omega_4 \in \Omega_{jy}} \frac{\widetilde{\omega}(\omega_1)}{p(\omega_1)} \frac{\widetilde{\omega}(\omega_2)}{p(\omega_2)} \frac{\widetilde{\omega}(\omega_3)}{p(\omega_3)} \frac{\widetilde{\omega}(\omega_4)}{p(\omega_4)} \\ \cdot \mathbb{I}(\omega_1 \in \bar{\Omega}(k_1,i)) \mathbb{I}(\omega_2 \in \bar{\Omega}(l_1,j)) \mathbb{I}(\omega_3 \in \bar{\Omega}(k_2,i)) \mathbb{I}(\omega_4 \in \bar{\Omega}(l_1,j)) \\ + m^2(m-1) \sum_{\omega_1 \in \Omega_{ix}} \sum_{\omega_2 \in \Omega_{jx}} \sum_{\omega_3 \in \Omega_{iy}} \sum_{\omega_4 \in \Omega_{jy}} \frac{\widetilde{\omega}(\omega_1)}{p(\omega_1)} \frac{\widetilde{\omega}(\omega_2)}{p(\omega_2)} \frac{\widetilde{\omega}(\omega_3)}{p(\omega_3)} \frac{\widetilde{\omega}(\omega_4)}{p(\omega_4)} \\ \cdot \mathbb{I}(\omega_1 \in \bar{\Omega}(k_1,i)) \mathbb{I}(\omega_2 \in \bar{\Omega}(l_1,j)) \mathbb{I}(\omega_3 \in \bar{\Omega}(k_1,i)) \mathbb{I}(\omega_4 \in \bar{\Omega}(l_2,j)). \Big\} \end{array}\right\} \text{same-different (3)}
\end{aligned}
\tag{37}
$$

We now address each case $1-3$ in turn.

7.3.1 Case 1: $k_1 = k_2, l_1 = l_2$

Case 1 is trivial. By design, antithetic termination does not affect the marginal walk behaviour (a sufficient condition for the estimator to remain unbiased). This means that it cannot affect terms that consider a single walk out of node i and a single walk out of j, and all terms of case 1 are unchanged by the introduction of antithetic termination.

7.3.2 Case 2: $k_1 \neq k_2, l_1 \neq l_2$

Now we consider terms where both the walks out of node i and the walks out of node j differ. To emphasise, we are considering 4 different random walks: 2 out of i and 2 out of j.

Within this setting, we will need to consider the situations where either i) one or ii) both of the pairs exhibit antithetic termination rather than i.i.d.. Terms of both kind will appear when we use ensembles of antithetic pairs. We need to check that in both cases the result is smaller compared to when both pairs are i.i.d..

To evaluate these terms, we first need to understand how inducing antithetic termination modifies the joint distribution $p(\omega_1 \in \bar{\Omega}(k_1,i), \omega_3 \in \bar{\Omega}(k_2,i))$: namely, the probability that two randomly sampled walks $\bar{\Omega}(k_1,i)$ and $\bar{\Omega}(k_2,i)$ contain the respective subwalks ω_1 and ω_3, given that their termination is either i.i.d. or antithetic. In the i.i.d. case, it is straightforward to convince oneself that

$$
p(\omega_1 \in \bar{\Omega}(1,i), \omega_3 \in \bar{\Omega}(3,i)) = \left(\frac{1-p}{d}\right)^m \left(\frac{1-p}{d}\right)^n, \tag{38}
$$

where m and n denote the lengths of subwalks ω_1 and ω_3, respectively. With antithetic termination, from Eq. 13 it follows that the probability of sampling a walk $\bar{\Omega}_3$ of length j conditioned on sampling an antithetic partner $\bar{\Omega}_1$ of length i is

$$
p(\text{len}(\bar{\Omega}_3) = j | \text{len}(\bar{\Omega}_1) = i) = \begin{cases} \left(\frac{1-2p}{1-p}\right)^j \frac{p}{1-p} & \text{if } j < i, \\ 0 & \text{if } j = i, \\ \left(\frac{1-2p}{1-p}\right)^i (1-p)^{j-i-1} p & \text{if } j > i. \end{cases} \tag{39}
$$

Using these probabilities, it is then straightforward but algebraically tedious to derive the joint probabilities over subwalks

$$
p(\omega_1 \in \bar{\Omega}(1,i), \omega_3 \in \bar{\Omega}(3,i)) = \begin{cases} \frac{1}{d^{m+n}} \left(\frac{1-2p}{1-p}\right)^n (1-p)^m & \text{if } n < m, \\ \frac{1}{d^{2m}} (1-2p)^m & \text{if } n = m, \\ \frac{1}{d^{m+n}} \left(\frac{1-2p}{1-p}\right)^m (1-p)^n & \text{if } n > m, \end{cases} \tag{40}
$$

where m is the length of ω_1, n is the length of ω_3 and i is now the index of a particular node.

To be explicit, we have integrated over the conditional probabilities of *walks* of particular lengths (i, j) to obtain the joint probabilities of sampled walks containing *subwalks* of particular lengths (m, n). Let us consider the case of $n < m$ as an example. Using Eq. 39,

$$\begin{aligned} p(\omega_1 \in \bar{\Omega}(1,i), \omega_3 \in \bar{\Omega}(3,i)) = \frac{1}{d^{m+n}} \sum_{i=m}^{\infty} \Bigg[\sum_{j=n}^{i-1} \left(\frac{1-2p}{1-p}\right)^j \frac{p}{1-p} (1-p)^i p + \\ + \sum_{j=i+1}^{\infty} \left(\frac{1-2p}{1-p}\right)^i (1-p)^{j-i-1} p (1-p)^i p \Bigg], \end{aligned} \tag{41}$$

where the branching factors of d appeared because at every timestep the subwalks have d possible edges to choose from. After we have completed the particular subwalks of lengths m and n we no longer care about *where* the walks go, just their lengths, so we stop accumulating these multiplicative factors. Computing the summations in Eq. 41 (which are all straightforward geometric series), we quickly arrive at the top line of Eq. 40.

Returning to our main discussion, note that in the d-regular, equal-weights case,

$$\begin{aligned} &\sum_{\omega_1 \in \Omega_{ix}} \sum_{\omega_3 \in \Omega_{iy}} \frac{\widetilde{\omega}(\omega_1)}{p(\omega_1)} \frac{\widetilde{\omega}(\omega_3)}{p(\omega_3)} p(\omega_1 \in \bar{\Omega}(k_1, i), \omega_3 \in \bar{\Omega}(k_2, i)) \\ &= \sum_{\omega_1 \in \Omega_{ix}} \sum_{\omega_3 \in \Omega_{iy}} \left(\frac{wd}{1-p}\right)^{m+n} p(\omega_1 \in \bar{\Omega}(k_1, i), \omega_3 \in \bar{\Omega}(k_2, i)). \end{aligned} \tag{42}$$

The summand depends only on walk lengths m, n but not direction, which invites us to decompose the sum $\sum_{\omega_1 \in \Omega_{ix}} (\cdot)$ over paths between nodes i and x to a sum over path lengths, weighted by the number of paths at each length. Explicitly,

$$\sum_{\omega_1 \in \Omega_{ix}} (\cdot) = \sum_{n=1}^{\infty} (\mathbf{A}^n)_{ix} (\cdot), \tag{43}$$

with $\mathbf{A}$ the (unweighted) adjacency matrix. We have used the fact that $(\mathbf{A}^n)_{ij}$ counts the number of walks of length n between nodes i and x. $\mathbf{A}$ is symmetric so has a convenient decomposition into orthogonal eigenvectors and real eigenvalues:

$$(\mathbf{A}^n)_{ix} = \sum_{k=1}^{N} \lambda_k^n \langle i|k\rangle \langle k|x\rangle \tag{44}$$

where $|k\rangle$ enumerates the N eigenvectors of $\mathbf{A}$ with corresponding eigenvalues λ_k, and $\langle i|$ and $\langle x|$ are unit vectors in the i and x coordinate axes, respectively. We remind the reader that we have adopted Dirac's bra-ket notation; $|y\rangle$ denotes the vector $\boldsymbol{y}$ and $\langle y|$ denotes $\boldsymbol{y}^\top$.

Inserting Eqs 44 and 43 into Eq. 42 and using the probability distributions in Eq. 38 and 40, our all-important variance-determining correlation term from Eq.36 evaluates to

$$\sum_{x,y \in \mathcal{V}} \sum_{k_1,k_2,k_3,k_4=1}^{N} B^{(i)}_{k_1,k_3} B^{(j)}_{k_2,k_4} \langle i|k_1\rangle \langle k_1|x\rangle \langle j|k_2\rangle \langle k_2|x\rangle \langle i|k_3\rangle \langle k_3|y\rangle \langle j|k_4\rangle \langle k_4|y\rangle, \tag{45}$$

where the matrix elements $B^{(i)}_{k_1,k_3}$ and $B^{(j)}_{k_2,k_4}$, corresponding to the pairs of walkers out of i and j respectively, are equal to one of the two following expressions:

$$B_{k_1,k_3} = \begin{cases} C_{k_1,k_3} := \frac{w\lambda_{k_1}}{1-w\lambda_{k_1}} \frac{w\lambda_{k_3}}{1-w\lambda_{k_3}} & \text{if i.i.d.} \\ D_{k_1,k_3} := \frac{w\lambda_{k_1}}{1-w\lambda_{k_1}} \frac{w\lambda_{k_3}}{1-w\lambda_{k_3}} \frac{c(1-w^2\lambda_{k_1}\lambda_{k_3})}{1-cw^2\lambda_{k_1}\lambda_{k_3}} & \text{if antithetic.} \end{cases} \tag{46}$$

Here, c is a constant defined by $c := \frac{1-2p}{(1-p)^2}$ with p the termination probability. These forms are straightforward to compute with good algebraic bookkeeping; we omit details for economy of space.

Eq. 45 can be simplified. Observe that $\sum_{x\in\mathcal{V}} |x\rangle\langle x| = \mathbf{I}_N$ ('resolution of the identity'), and that since the eigenvectors of $\mathbf{A}$ are orthogonal $\langle k_1|k_2\rangle = \delta_{k_1,k_2}$. Applying this, we can write

$$\sum_{k_1,k_3=1}^{N} B^{(i)}_{k_1,k_3} B^{(j)}_{k_1,k_3} \langle i|k_1\rangle \langle j|k_1\rangle \langle i|k_3\rangle \langle j|k_3\rangle \,. \tag{47}$$

Our task is then to determine whether 47 is reduced by conditioning that either one or both of the pairs of walkers are antithetic rather than independent. That is,

$$\sum_{k_1=1}^{N}\sum_{k_3=1}^{N} \left(C_{k_1,k_3} D_{k_1,k_3} - C_{k_1,k_3} C_{k_1,k_3}\right) \langle i|k_1\rangle \langle j|k_1\rangle \langle i|k_3\rangle \langle j|k_3\rangle \overset{?}{\leq} 0, \tag{48}$$

$$\sum_{k_1=1}^{N}\sum_{k_3=1}^{N} \left(D_{k_1,k_3} D_{k_1,k_3} - C_{k_1,k_3} C_{k_1,k_3}\right) \langle i|k_1\rangle \langle j|k_1\rangle \langle i|k_3\rangle \langle j|k_3\rangle \overset{?}{\leq} 0. \tag{49}$$

Define a vector $\boldsymbol{y} \in \mathbb{R}^N$ with entries $y_p := \langle i|k_p\rangle \langle j|k_p\rangle$, such that its pth element is the product of the i and jth coordinates of the pth eigenvector $\boldsymbol{k}_p$. In this notation, Eqs 48 and 49 can be written

$$\sum_{p=1}^{N}\sum_{q=1}^{N} \left(C_{pq} D_{pq} - C_{pq} C_{pq}\right) y_p y_q \overset{?}{\leq} 0, \tag{50}$$

$$\sum_{p=1}^{N}\sum_{q=1}^{N} \left(D_{pq} D_{pq} - C_{pq} C_{pq}\right) y_p y_q \overset{?}{\leq} 0. \tag{51}$$

For Eqs 50 and 51 to be true for arbitrary graphs, it is sufficient that the matrices $\mathbf{E}$ and $\mathbf{F}$ with matrix elements $E_{pq} := C_{pq}D_{pq} - C_{pq}C_{pq}$ and $F_{pq} := D_{pq}D_{pq} - C_{pq}C_{pq}$ are *negative definite*. Our next task is to prove that this is the case.

First, consider $\mathbf{E}$, where just one of the two pairs of walkers is antithetic. Putting in the explicit forms of C_{pq} and D_{pq} from Eq. 46,

$$E_{pq} = -\left(\frac{\bar{\lambda}_p \bar{\lambda}_q}{(1-\bar{\lambda}_p)(1-\bar{\lambda}_q)}\right)^2 \frac{p^2}{(1-p)^2} \frac{1}{1-\frac{1-2p}{(1-p)^2}\bar{\lambda}_p\bar{\lambda}_q} \tag{52}$$

where for notational compactness we took $\bar{\lambda}_p := w\lambda_p$ (the eigenvalues of the *weighted* adjacency matrix $\mathbf{U}$). Taylor expanding,

$$E_{pq} = -\left(\frac{\bar{\lambda}_p \bar{\lambda}_q}{(1-\bar{\lambda}_p)(1-\bar{\lambda}_q)}\right)^2 \frac{p^2}{(1-p)^2} \sum_{m=0}^{\infty} \left(\frac{1-2p}{(1-p)^2}\bar{\lambda}_p\bar{\lambda}_q\right)^m . \tag{53}$$

Inserting this into Eq. 50, we get

$$\sum_{p=1}^{N}\sum_{q=1}^{N} E_{pq} y_p y_q = -\frac{p^2}{(1-p)^2} \sum_{m=0}^{\infty} \left(\sum_{p=1}^{N} \frac{\bar{\lambda}_p^2}{(1-\bar{\lambda}_p)^2} \left(\frac{\sqrt{1-2p}}{1-p}\bar{\lambda}_p\right)^m y_p \right)^2 \leq 0, \tag{54}$$

which implies that $\mathbf{E}$ is indeed negative definite. Note that we have not made any additional assumptions about the values of p and w beyond those already stipulated: namely, $0 < p \leq \frac{1}{2}$ and $\bar{\lambda}_{\text{max}} < 1$.

Next, consider $\mathbf{F}$, where both pairs of walkers are antithetic. Again inserting Eqs 46, we find that

$$F_{pq} = \left(\frac{\bar{\lambda}_p \bar{\lambda}_q}{(1-\bar{\lambda}_p)(1-\bar{\lambda}_q)}\right)^2 \left[\left(\frac{c - c\bar{\lambda}_p\bar{\lambda}_q}{1-c\bar{\lambda}_p\bar{\lambda}_q}\right)^2 - 1\right] \tag{55}$$

where we remind the reader that $c = \frac{1-2p}{(1-p)^2}$. The Taylor expansion in $\bar{\lambda}_p\bar{\lambda}_q$ is

$$\begin{aligned} F_{pq} &= \left(\frac{\bar{\lambda}_p \bar{\lambda}_q}{(1-\bar{\lambda}_p)(1-\bar{\lambda}_q)}\right)^2 \left[\sum_{i=0}^{\infty} (\bar{\lambda}_p\bar{\lambda}_q)^i (c-1) c^i (1 + c + i(c-1))\right] \\ &= w^4 (\lambda_p\lambda_q)^2 \sum_{i,j,k=0}^{\infty} (\lambda_p\lambda_q)^{i+j+k} w^{2i+j+k} (c-1) c^i (1+c+i(c-1))(j+1)(k+1). \end{aligned} \tag{56}$$

In fact, $\mathbf{F}$ is *not* generically negative definite, but will be at sufficiently small p or w. Write $\mathbf{F} = w^4(\mathbf{G}+\mathbf{H})$, with

$$G_{pq} := (\lambda_p\lambda_q)^2 \left(c^2-1\right), \tag{57}$$

$$H_{pq} := (\lambda_p\lambda_q)^2 \sum_{i,j,k=0\backslash\{i=j=k=0\}}^{\infty} (\lambda_p\lambda_q)^{i+j+k} w^{2i+j+k}(c-1)c^i(1+c+i(c-1))(j+1)(k+1). \tag{58}$$

$\mathbf{G}$ is manifestly negative definite because $c<1$ but $\mathbf{H}$ may not be. Treat $\mathbf{H}$ as a perturbation to $\mathbf{G}$.

Recalling that the spectral radius of $\mathbf{H}$ is defined

$$\rho(\mathbf{H}) := \max_{\|\boldsymbol{x}\|_2=1} \mathbf{H}\boldsymbol{x}, \tag{59}$$

it is clear that the spectral radius of $\mathbf{H}$ approaches 0 smoothly as $w\to 0$ since all its matrix elements vanish. Recall also an important corollary of Weyl's perturbation inequality: any perturbed eigenvalue of $\mathbf{F}+\mathbf{G}$ will be within one spectral radius $\rho(\mathbf{G})$ of the original eigenvalue of $\mathbf{F}$. This means that, by reducing w, we can shrink the spectral radius of $\mathbf{G}$ until $\rho(\mathbf{G}) < (\lambda_p\lambda_q)^2\left(1-c^2\right)$, at which point we are guaranteed that $\mathbf{F}$ will be negative definite. Hence, at sufficiently small w, correlation terms with both pairs antithetic are suppressed as required.

Taylor expanding in $c\to 1$ (which corresponds to $p\to 0$) instead of $\lambda_p\lambda_q$, we can make exactly analogous arguments to find that $\mathbf{F}$ is also guaranteed to be negative definite with when p is sufficiently small. Briefly: let $c = 1-\delta$ with $\delta = \left(\frac{p}{1-p}\right)^2$. Then we have that

$$\begin{aligned} F_{pq} &= \left(\frac{\bar{\lambda}_p\bar{\lambda}_q}{(1-\bar{\lambda}_p)(1-\bar{\lambda}_q)}\right)^2 \left((1-\delta)^2\left(\frac{1-\bar{\lambda}_p\bar{\lambda}_q}{1-\bar{\lambda}_p\bar{\lambda}_q+\delta\bar{\lambda}_p\bar{\lambda}_q}\right)^2 - 1\right) \\ &= \left(\frac{\bar{\lambda}_p\bar{\lambda}_q}{(1-\bar{\lambda}_p)(1-\bar{\lambda}_q)}\right)^2 \left(\frac{-2\delta}{1-\bar{\lambda}_p\bar{\lambda}_q} + \mathcal{O}(\delta^2)\right). \end{aligned} \tag{60}$$

Taylor expanding $\frac{1}{1-\bar{\lambda}_p\bar{\lambda}_q}$, it is easy to see that the operator defined by the $\mathcal{O}(\delta)$ term of Eq. 60 is negative definite. This part will dominate over higher order terms (which are *not* in general negative definite) when δ is sufficiently small, guaranteeing the effectiveness of our mechanism on these terms.

As an aside, we also note that Taylor expanding about $c=0$ (which corresponds to $p\to\frac{1}{2}$) yields

$$F_{pq} = \left(\frac{\bar{\lambda}_p\bar{\lambda}_q}{(1-\bar{\lambda}_p)(1-\bar{\lambda}_q)}\right)^2 \left(-1+\mathcal{O}(c^2)\right) \tag{61}$$

which is manifestly negative definite at small enough c. Hence, intriguingly, the $k_1\neq k_2$ variance contributions are also suppressed in the $p\to\frac{1}{2}$ limit.

This concludes our study of variance contributions in Eq. 36 where $k_1\neq k_2$, $l_1\neq l_2$. We have found that these correlation terms are indeed suppressed by antithetic termination when p or $\rho(\mathbf{U})$ is small enough (or when p is sufficiently close to $\frac{1}{2}$).

7.3.3 Case 3: $k_1 = k_2, l_1 \neq l_2$

We now consider terms where $k_1=k_2$ and $l_1\neq l_2$. We are considering a total of 3 walks: just 1 out of node i but a pair (which may be antithetic or i.i.d.) out of node j. We inspect the term

$$\sum_{\omega_1\in\Omega_{ix}}\sum_{\omega_3\in\Omega_{iy}} \left(\frac{wd}{1-p}\right)^{m+n} p(\omega_1\in\bar{\Omega}(k_1,i),\omega_3\in\bar{\Omega}(k_1,i)), \tag{62}$$

where m denotes the length of ω_1 and n denotes the length of ω_3. What is the form of $p(\omega_1\in\bar{\Omega}(k_1,i),\omega_3\in\bar{\Omega}(k_1,i))$? It is the probability that a *single* walk out of node i, $\bar{\Omega}(k_1,i)$, contains walks ω_1 between nodes i and x and ω_3 between i and y as subwalks. Such a walk must pass through all three nodes i, x and y. After some thought,

$$p(\omega_1,\omega_3\in\bar{\Omega}(k_1,i)) = \begin{cases} \left(\frac{1-p}{d}\right)^m & \text{if } \omega_1=\omega_3, \\ \left(\frac{1-p}{d}\right)^m & \text{if } \omega_3\in\omega_1, \\ \left(\frac{1-p}{d}\right)^n & \text{if } \omega_1\in\omega_3, \\ 0 & \text{otherwise.} \end{cases} \tag{63}$$

Here, $\omega_1 \in \omega_3$ means ω_1 is a strict subwalk of ω_3, so the sequence of nodes traversed is $i \to x \to y$. Likewise, $\omega_3 \in \omega_1$ implies a path $i \to y \to x$. Summing these contributions,

$$\begin{aligned}\sum_{\omega_1 \in \Omega_{ix}} \sum_{\omega_3 \in \Omega_{iy}} \left(\frac{wd}{1-p}\right)^{m+n} p(\omega_1 \in \bar{\Omega}(l_1, i), \omega_3 \in \bar{\Omega}(l_1, i)) \\ = \underbrace{\sum_{\omega_1 \in \Omega_{ix}} \left(\frac{wd}{1-p}\right)^{2\text{len}(\omega_1)} p(\omega_1 \in \bar{\Omega}(k_1, i))\delta_{xy}}_{\omega_1=\omega_3, i\to x=y} \\ + \underbrace{\sum_{\omega_1 \in \Omega_{ix}} \left(\frac{wd}{1-p}\right)^{2\text{len}(\omega_1)} p(\omega_1 \in \bar{\Omega}(k_1, i)) \sum_{\omega_\delta \in \Omega_{xy}} \left(\frac{wd}{1-p}\right)^{\text{len}(\omega_\delta)} p(\omega_\delta \in \bar{\Omega}(k_1, x))}_{\omega_1 \in \omega_3, i\to x\to y} \\ + \underbrace{\sum_{\omega_3 \in \Omega_{iy}} \left(\frac{wd}{1-p}\right)^{2\text{len}(\omega_3)} p(\omega_3 \in \bar{\Omega}(k_1, i)) \sum_{\omega_\delta \in \Omega_{yx}} \left(\frac{wd}{1-p}\right)^{\text{len}(\omega_\delta)} p(\omega_\delta \in \bar{\Omega}(k_1, y))}_{\omega_3 \in \omega_1, i\to y\to x}.\end{aligned} \tag{64}$$

We introduced ω_δ for the sum over paths between nodes x and y, and $p(\omega_\delta \in \bar{\Omega}(k_1, x))$ is the probability of some particular subwalk $x \to y$, equal to $\left(\frac{1-p}{d}\right)^{\text{len}(\omega_\delta)}$ in the d-regular case. ω_3 is a dummy variable so can be relabelled ω_1. The variance-determining correlation term from Eq. 36 becomes

$$\begin{aligned}\sum_{x,y\in\mathcal{V}} \Bigg[\sum_{\omega_1 \in \Omega_{ix}} \left(\frac{wd}{1-p}\right)^{2\text{len}(\omega_1)} p(\omega_1 \in \bar{\Omega}(k_1, i))\delta_{xy} \\ + \sum_{\omega_1 \in \Omega_{ix}} \left(\frac{wd}{1-p}\right)^{2\text{len}(\omega_1)} p(\omega_1 \in \bar{\Omega}(k_1, i)) \sum_{\omega_\delta \in \Omega_{xy}} \left(\frac{wd}{1-p}\right)^{\text{len}(\omega_\delta)} p(\omega_\delta \in \bar{\Omega}(k_1, x)) \\ + \sum_{\omega_1 \in \Omega_{iy}} \left(\frac{wd}{1-p}\right)^{2\text{len}(\omega_1)} p(\omega_1 \in \bar{\Omega}(k_1, i)) \sum_{\omega_\delta \in \Omega_{yx}} \left(\frac{wd}{1-p}\right)^{\text{len}(\omega_\delta)} p(\omega_\delta \in \bar{\Omega}(k_1, y))\Bigg] \\ \cdot \sum_{k_2=1}^{N} \sum_{k_4=1}^{N} B_{k_2,k_4}^{(j)} \langle j|k_2\rangle \langle k_2|x\rangle \langle j|k_4\rangle \langle k_4|y\rangle .\end{aligned} \tag{65}$$

where $B_{k_2,k_4}^{(j)}$ depends on whether the coupling of the pair of walkers out of node j is i.i.d. or antithetic, as defined in Eq. 46. x and y are dummy variables so can also be swapped, and the sum over the paths ω_δ is computed via the usual sum over path lengths and eigendecomposition of $\mathbf{A}$. Using the resolution of the identity and working through the algebra, we obtain the correlation term

$$\begin{aligned}\sum_{x\in\mathcal{V}} \Bigg[\sum_{\omega_1 \in \Omega_{ix}} \left(\frac{wd}{1-p}\right)^{2\text{len}(\omega_1)} p(\omega_1 \in \bar{\Omega}(k_1, i))\Bigg] \\ \cdot \sum_{k_2,k_4=1}^{N} \left(\frac{1-w^2\lambda_{k_2}\lambda_{k_4}}{(1-w\lambda_{k_2})(1-w\lambda_{k_4})}\right) B_{k_2,k_4}^{(j)} \langle x|k_2\rangle \langle k_2|j\rangle \langle x|k_4\rangle \langle k_4|j\rangle .\end{aligned} \tag{66}$$

Now observe that the prefactor in square brackets is positive for any node x since it is the expectation of a squared quantity. This means that, for the sum in Eq. 66 to be suppressed by antithetic coupling, it is sufficient for the summation in its lower line to be reduced. Defining a vector $\boldsymbol{y} \in \mathbb{R}^N$ with elements $y_p := \langle x|k_p\rangle \langle k_p|j\rangle$, it becomes clear that we require that the operator $\mathbf{J}$ with matrix elements

$$J_{pq} := \left(\frac{1-w^2\lambda_p\lambda_q}{(1-w\lambda_p)(1-w\lambda_q)}\right)(D_{pq} - C_{pq}) \tag{67}$$

is negative definite. Using the forms in Eq. 46,

$$J_{pq} = -\frac{w^2\lambda_p\lambda_q}{(1-w\lambda_p)^2(1-w\lambda_q)^2} \frac{\frac{p^2}{(1-p)^2}(1-w^2\lambda_p\lambda_q)}{1-\frac{1-2p}{(1-p)^2}w^2\lambda_p\lambda_q}. \tag{68}$$

Making very similar arguments to in Sec. 7.3.2 (namely, Taylor expanding and appealing to Weyl's perturbation inequality), we can show that, whilst this operator is not generically negative definite, it will be at sufficiently small p or w.

A brief note: Taylor expanding in c,

$$J_{pq} = -\frac{w^2 \lambda_p \lambda_q}{(1-w\lambda_p)^2(1-w\lambda_q)^2}(1-w^2\lambda_p\lambda_q) + \mathcal{O}(c), \tag{69}$$

which is only negative definite when we also simultaneously take $w \to 0$. Interestingly, in contrast to case 2, these terms are *not* suppressed by $p \to \frac{1}{2}$ on its own; we need to control the spectral radius of $\mathbf{U}$.

This concludes the section of the proof addressing terms $k_1 = k_2$ and $l_1 \neq l_2$ (case 3). Again, these variance contributions are always suppressed by antithetic termination at sufficiently small p or $\rho(\mathbf{U})$.

Having now considered all the possible variance contributions enumerated by cases $1-3$ and shown that each is either reduced or unmodified by the imposition of antithetic termination, we can finally conclude that our novel mechanism does indeed suppress the 2-regularised Laplacian kernel estimator variance for a d-regular graph of equal weights at sufficiently small p or $\rho(\mathbf{U})$. □

As mentioned in the main body of the manuscript, these conditions tend not to be very restrictive in experiments. Intriguingly, small $\rho(\mathbf{U})$ with $p = \frac{1}{2}$ actually works very well.

Our next task is to generalise these results to broader classes of graphs.

7.4 Extending the results to arbitrary graphs and sampling strategies (Theorem 3.2 cont.)

Throughout Sec. 7.3, we considered the simplest setting of a d-regular graph where all edges have equal weight. We have also taken a basic sampling strategy, with the walker choosing one of its current node's neighbours at random at every timestep. Here we relax these assumptions, showing that our results remain true in more general settings.

7.4.1 Relaxing d-regularity

First, we consider graphs whose vertex degrees differ. It is straightforward to see that the terms in case 2 (Sec. 7.3.2) are unmodified because taking $d^m \to \prod_{i=1}^m d_i$ in $p(\omega_1)$ and $d^n \to \prod_{i=1}^n d_i$ in $p(\omega_3)$ is exactly compensated by the corresponding change in in joint probability $p(\omega_1 \in \bar{\Omega}(k_1, i), \omega_3 \in \bar{\Omega}(k_2, i))$. Our previous arguments all continue to hold.

Case 3 (Sec. 7.3.3) is only a little harder. Now the prefactor in square parentheses in the top line of Eq. 66 evaluates to

$$\left[\sum_{\omega_1 \in \Omega_{ix}} \left(\frac{w}{1-p}\right)^{2\text{len}(\omega_1)} \left(\prod_{i=1}^{\text{len}(\omega_1)} d_i^2\right) p(\omega_1 \in \bar{\Omega}(k_1, i))\right] \tag{70}$$

which is still positive for any node x. The lower line of Eq. 66 is unmodified because once again the change $d^m \to \prod_{i=1}^m d_i$ exactly cancels in the marginal and joint probabilities, so $\mathbf{J}$ is unchanged and our previous conclusions prevail.

7.4.2 Weighted graphs

Now we permit edge weights to differ across the graph. Once again, case 2 (Sec. 7.3.2) is straightforward: instead of Eq. 43, we take

$$\sum_{\omega_1 \in \Omega_{ix}} \widetilde{\omega}(\omega_1)\,(\cdot) = \sum_{n=1}^{\infty} (\mathbf{U}^n)_{ix}\,(\cdot)\,, \tag{71}$$

where $\mathbf{U}$ is the *weighted* adjacency matrix. We incorporate the product of each walk's edge weights into the combinatorial factor, then sum over path lengths as before. In downstream calculations we drop all instances of w and reinterpret λ as the eigenvalues of the $\mathbf{U}$ instead of $\mathbf{A}$, but our arguments are otherwise unmodified; these variance contributions will be suppressed if $\rho(\mathbf{U})$ or p is sufficiently small.

Case 3 (Sec. 7.3.3) is also easy enough; the bracketed prefactor of 66 becomes

$$\left[\sum_{\omega_1 \in \Omega_{ix}} \left(\frac{1}{1-p} \right)^{2\text{len}(\omega_1)} \left(\prod_{i=1}^{\text{len}(\omega_1)} w_{i\sim i+1}^2 d_i^2 \right) p(\omega_1 \in \bar{\Omega}(k_1, i)) \right] \tag{72}$$

which is again positive. Here, $w_{i\sim i+1}$ denotes the weight associated with the edge between the i and $i+1$th nodes of the walk. Therefore, it is sufficient that the matrix $\mathbf{J}$ with matrix elements

$$J_{pq} = -\frac{\lambda_p \lambda_q}{(1-\lambda_p)^2 (1-\lambda_q)^2} \frac{\frac{p^2}{(1-p)^2}(1-\lambda_p\lambda_q)}{1 - \frac{1-2p}{(1-p)^2}\lambda_p \lambda_q} \tag{73}$$

is negative definite, with λ_p now the pth eigenvalue of the *weighted* adjacency matrix $\mathbf{U}$. Following the same arguments as in Sec. 7.3.3, this will be the case at small enough p or $\rho(\mathbf{U})$.

7.4.3 Different sampling strategies

Finally, we consider modifying the sampling strategy for random walks on the graph. We have previously assumed that the walker takes successive edges at random (i.e. with probability $\frac{1}{d_i}$), but the transition probability can also be a function of the edge weights. For example, if all the edge weights are positive, we might take

$$p(i \to j | \bar{s}) = \frac{w_{ij}}{\sum_{k\sim i} w_{ik}} \tag{74}$$

for the probability of transitioning from node i to j at a given timestep (with $w_{ij} \coloneqq \mathbf{U}_{ij}$), given that the walker does not terminate. This strategy increases the probability of taking edges with bigger weights and which therefore contribute more to $(\mathbf{I}_N - \mathbf{U})^{-2}$ – something that empirically suppresses the variance on the estimator of the 2-regularised Laplacian kernel. Does antithetic termination reduce it further?

Case 2 (Sec. 7.3.2) is again easy; the w-dependent modifications to $p(\omega_1)$ and $p(\omega_3)$ are exactly compensated by adjustments to $p(\omega_1 \in \bar{\Omega}(k_1, i), \omega_3 \in \bar{\Omega}(k_2, i))$. To wit, Eq. 40 becomes

$$p(\omega_3 \in \bar{\Omega}(3, i), \omega_1 \in \bar{\Omega}(1, i)) = \begin{cases} \frac{\widetilde{\omega}(\omega_1)\widetilde{\omega}(\omega_3)}{\gamma(\omega_1)\gamma(\omega_3)} \left(\frac{1-2p}{1-p}\right)^n (1-p)^m & \text{if } n < m \\ \frac{\widetilde{\omega}(\omega_1)^2}{\gamma(\omega_1)^2}(1-2p)^m & \text{if } n = m \\ \frac{\widetilde{\omega}(\omega_1)\widetilde{\omega}(\omega_3)}{\gamma(\omega_1)\gamma(\omega_3)} \left(\frac{1-2p}{1-p}\right)^m (1-p)^n & \text{if } n > m. \end{cases} \tag{75}$$

where we defined a new function of a a walk,

$$\gamma(\omega) \coloneqq \prod_{i\in\omega} \sum_{k\sim i} w_{ik}. \tag{76}$$

γ computes the sum of edge weights connected to each node in the walk ω (excluding the last), then takes the product of these quantities. It is straightforward to check that, when all the graph weights are equal, $\frac{\widetilde{\omega}(\omega)}{\gamma(\omega)} = \frac{1}{d^m}$ with m the length of ω. Meanwhile, $p(\omega_1)$ becomes

$$p(\omega_1) = \frac{(1-p)^m \widetilde{\omega}(\omega_1)}{\gamma(\omega_1)} \tag{77}$$

such that these modifications cancel out when we evaluate Eq. 36.

Case 3 (Sec. 49) is also straightforward. The prefactor in square brackets is equal to 72 and is again positive for any valid sampling strategy $p(\omega_1 \in \bar{\Omega}(k_1, i))$ and $\mathbf{J}$ does not change, so our arguments still hold and these variance contributions are reduced by antithetic coupling.

We note that these arguments will generalise straightforwardly to any weight-dependent sampling strategy and are not particular to the linear case. $\widetilde{\omega}/\gamma$ can be replaced by some more complicated variant that defines a valid probability distribution $p(\omega_1 \in \bar{\Omega}(k_1, i))$ and antithetic termination will still prove effective.

7.4.4 Summary

In Sec. 7.4, our theoretical results for antithetic termination have proved robust to generalisations such as relaxing d-regularity and changing the walk sampling strategy. A qualitative explanation for this is as follows: upon making the changes, the ratio of the joint to marginal probablities

$$\frac{p(\omega_1, \omega_3)}{p(\omega_1)p(\omega_3)} \tag{78}$$

is unmodified. This is because *we know* how we are modifying the probability over walks and construct the estimator to compensate for it. Meanwhile, the correlations between walk *lengths* are insensitive to the walk directions, so in every case they continue to suppress the kernel estimator variance. The only kink is the terms described in Sec. 7.3.3 which require a little more work, but the mathematics conspires that our arguments are again essentially unmodified, though perhaps without such an intuitive explanation.

7.5 Beyond antithetic coupling (proof of Theorem 3.4)

Our final theoretical contribution is to consider random walk behaviour when TRVs are offset by *less* than p, $\Delta < p$. Unlike antithetic coupling, it permits simultaneous termination. Eqs 13 become

$$\begin{gathered} p(s_1) = p(s_2) = p, \quad p(\bar{s}_1) = p(\bar{s}_2) = 1-p, \quad p(s_2|s_1) = \frac{p-\Delta}{p}, \\ p(\bar{s}_2|s_1) = \frac{\Delta}{p}, \quad p(s_2|\bar{s}_1) = \frac{\Delta}{1-p}, \quad p(\bar{s}_2|\bar{s}_1) = \frac{1-p-\Delta}{1-p}. \end{gathered} \tag{79}$$

The probability of two antithetic walks $\bar{\Omega}(1,i)$ and $\bar{\Omega}(3,i)$ containing subwalks ω_1 and ω_3 becomes

$$p(\omega_3 \in \bar{\Omega}(3,i), \omega_1 \in \bar{\Omega}(1,i)) = \begin{cases} \frac{1}{d^{m+n}} \left(\frac{1-p-\Delta}{1-p}\right)^n (1-p)^m & \text{if } n < m \\ \frac{1}{d^{2m}} (1-p-\Delta)^m & \text{if } n = m \\ \frac{1}{d^{m+n}} \left(\frac{1-p-\Delta}{1-p}\right)^m (1-p)^n & \text{if } n > m, \end{cases} \tag{80}$$

which the reader might compare to Eq. 40. In analogy to Eq. 46, this induces the matrix

$$D^{\Delta}_{k_1,k_3} := \frac{w^2 \lambda_{k_1} \lambda_{k_3} \frac{1-p-\Delta}{(1-p)^2}}{1 - w^2 \lambda_{k_1} \lambda_{k_3} \frac{1-p-\Delta}{(1-p)^2}} \left(\frac{1 - w^2 \lambda_{k_1} \lambda_{k_3}}{(1-w\lambda_{k_1})(1-w\lambda_{k_3})} \right). \tag{81}$$

We can immediately observe that this is exactly equal to C_{k_1,k_3} when $\Delta = p(1-p)$, so for a pair of walkers with this TRV offset the variance will be identical to the i.i.d. result. Replacing D by D^{Δ} in E_{pq} and F_{pq} and J_{pq} and reasoning about negative definiteness via their respective Taylor expansions (as well as the new possible cross-term $D_{k_1,k_3} D^{\Delta}_{k_1,k_3}$), it is straightforward conclude that variance is suppressed compared to the i.i.d. case provided $\Delta > p(1-p)$ and $\rho(\mathbf{U})$ or p is sufficiently small. The $p \to 0$ limit demands a slightly more careful treatment: in order to stay in the regime $p(1-p) < \Delta < p$ we need to simultaneously take $\Delta \to 0$, e.g. by defining $\Delta(p) := p(1-p) + ap^2$ with the constant $0 < a < 1$. □

This result was reported in Theorem 3.4 of the main text.

7.6 What about diagonal terms?

The alert reader might remark that all derivations in Sec. 7.3 have taken $i \neq j$, considering estimators of the off-diagonal elements of the matrix $(\mathbf{I}_N - \mathbf{U})^{-2}$. In fact, estimators of the diagonal elements $\phi(i)^\top \phi(i)$ will be biased for both GRFs and q-GRFs if $\phi(i)$ is constructed using the same ensemble of walkers because each walker is manifestly correlated with, rather than independent of, itself. This is rectified by taking *two* ensembles of walkers out of each node, each of which may exhibit antithetic correlations among itself, then taking the estimator $\phi_1(i)^\top \phi_2(i)$. It is straightforward to convince oneself that, in this setup, the estimator is unbiased and q-GRFs will outperform GRFs. In practice, this technicality has essentially no effect on (q-)GRF performance and doubles runtime so we omit further discussion.

7.7 Further experimental details: compute, datasets and uncertainties

The experiments in Secs. 4.1, 4.2 and 4.4 were carried out on an Intel® Core™ i5-7640X CPU @ 4.00GHz × 4. Each required ~ 1 CPU hour. The experiments in Sec. 4.3 were carried out on a 2-core Xeon 2.2GHz with 13GB RAM and 33GB HDD. The computations for the largest considered graphs took ~ 1 CPU hour.

The real-world graphs and meshes were accessed from the repositories [Ivashkin, 2023] and [Dawson-Haggerty, 2023], with further information about the datasets available therein. Where we were able to locate them, the original papers presenting the graphs are: [Zachary, 1977, Lusseau et al., 2003, Newman, 2006, Bollacker et al., 1998, Leskovec et al., 2007].

All our experiments report standard deviations on the means, apart from the clustering task in Sec. 4.3 because running kernelised k-means on large graphs is expensive.