5. Proof of Theorem 1

As briefly mentioned in the body of the paper, our proof will proceed by doing a large-deviation analysis of the estimation error, and then integrating it to obtain an upper bound on the expectation. We cannot proceed directly by attempting to take expectation of $\sin^2(\hat{W}, w)$ because the WLSM ends up dividing two random variables (specifically, Eq. (13) has random variables in both the numerator and denominator under the $\arg\min$).

To preview what is to come, our algorithm suffers from three sources of error:

- The difference between p_{ij}, the true probability that i wins a coin toss, and F_{ij} due to randomness of the comparisons.
- The error in taking Taylor expansions.
- The error introduced by replacing v_{ij}, which is proportional to the asymptotic variance of $\log R_{ij}$, by the empirical estimate $\hat{V}_{ij}$.

Our analysis will need to bound the effect of each of these factors.

5.1. Notation

We begin by reiterating all the notation we have introduced:

w_i = true weight of item i

G = the comparison graph, with vertex set $\{1, \ldots, n\}$

E = edge set of the comparison graph G

$\overrightarrow{E}$ = set of directed edges obtained by orienting every edge in E arbitrarily

k = number of comparisons across each edge of G

F_{ij} = proportion of comparisons item i wins against j

p_{ij} = $\dfrac{w_i}{w_i + w_j}$, true probability that item i wins a comparison against item j

$$R_{ij} = \frac{F_{ij}}{F_{ji}}$$

$$\rho_{ij} = \frac{w_i}{w_j}$$

$$v_{ij} = \frac{w_i}{w_j} + \frac{w_j}{w_i} + 2$$

$$\hat{V}_{ij} = \frac{F_{ij}}{F_{ji}} + \frac{F_{ji}}{F_{ij}} + 2$$

We follow the convention that capitalized entries are either random variables or matrices, while lower-case letters correspond to scalars or vectors that are not random. Next, we introduce some new notation:

$$\hat{V} = \operatorname{diag}(\hat{V}_{ij}) \in \mathbb{R}^{|\overrightarrow{E}| \times |\overrightarrow{E}|}$$

$$V = \operatorname{diag}(v_{ij}) \in \mathbb{R}^{|\overrightarrow{E}| \times |\overrightarrow{E}|}$$

M = Edge-vertex incidence matrix of the graph $(\{1, \ldots, n\}, \overrightarrow{E})$, with thus $M \in \mathbb{R}^{n \times |\overrightarrow{E}|}$

$$L_V = MV^{-1}M^T \in \mathbb{R}^{n \times n}$$

$$L_{\hat{V}} = M\hat{V}^{-1}M^T \in \mathbb{R}^{n \times n}$$

X_{ij}^l = Bernoulli random variable describing the outcome of l'th comparison across edge (i, j)

$\mathbf{1}$ = all-ones vector

e_i = i'th basis vector

We take the opportunity to remind the reader of our notational conventions. We will omit the subscripts when we stack the above quantities into vectors. For example, the notation R represents the vector in $\mathbb{R}^{|\overrightarrow{E}|}$ obtained by stacking up the quantities $R_{ij}, (i, j) \in \overrightarrow{E}$. Furthermore, the ordinary graph Laplacian L can be written as $L = MM^T$, and he quantities L_V and $L_{\hat{V}}$ correspond to weighted graph Laplacians, where the edge $(i, j) \in E$ is weighted by v_{ij}^{-1} or $\hat{V}_{ij}^{-1}$, respectively.

Finally, when writing $A = O(B)$ for two expressions A, B, we mean that A is bounded linearly by B for k large enough (with respect to quantities defining the problem). More specifically, there exist an absolute constant K and a function $q(n, b, G, w)$ such that $A \le K.B$ for all $k \ge q(n, b, G, w)$.

We begin by defining an appropriate rescalings of the true weights w to which we can compare $\hat{W}$ defined in Eq. (18). The natural approach is to define w^r to be a rescaling of w such that

$$\prod_{i=1}^{n} (w_i^r)^{w_i^2} = 1, \tag{20}$$

and likewise

$$y_i = w_i \log w_i^r. \tag{21}$$

Observing that for each edge $(i, j) \in E$,

$$\log w_i^r - \log w_j^r = \log \rho_{ij}.$$

we can therefore repeat all the same steps that led to the derivation of Eq. (17) to obtain that

$$y = (\operatorname{diag}(w)^{-1} L_{\hat{V}} \operatorname{diag}(w)^{-1})^{\dagger} \operatorname{diag}(w)^{-1} M\hat{V}^{-1} \log \rho.$$

Putting this together with Eq. (15), we obtain

$$\hat{Y} - y = (\operatorname{diag}(w)^{-1} L_{\hat{V}} \operatorname{diag}(w)^{-1})^{\dagger} \operatorname{diag}(w)^{-1} M\hat{V}^{-1} (\log R - \log \rho). \tag{22}$$

This equation will be the basis of our analysis for the rest of the paper.

Since the sine is unaffected by scalings of the underlying vectors, we note that

$$\sin^2(\hat{W}, w) = \sin^2(\hat{W}^r, w^r).$$

Hence we will analyze $\sin^2(\hat{W}^r, w^r)$ to characterize the WLSM method, even though we do not know either of these vectors, as it turns out that our argument are easier to develop after these rather peculiar scalings.

We conclude this section with an observation which we will use repeatedly in the sequel. Eq. (20) implies that there is at least one i with $w_i \geq 1$ and at least one i with $w_i \leq 1$. Appealing to Eq. (1), we can conclude that $\max_i w_i^r \leq b$ and $\min_i w_i^r \geq b^{-1}$.

5.2. Decomposing the sources of error

We will proceed by obtaining a rate at which the right-hand side of Eq. (22) goes to zero. Our first step is to bound the difference $\log R - \log \rho$. We cite a lemma from the previous literature which derives bound on this quantity by applying Chernoff's inequality.

Proposition 3 (Eq. (13) from (Hendrickx *et al.* , 2019))**.** *Let us write*

$$\log R - \log \rho = V(F - p) + \Delta.$$

Then if $\delta \leq e^{-1}$ and $k = \Omega\left(b \log(n/\delta)\right)$, we have that with probability $1 - \delta$, the vector $\Delta \in \mathbb{R}^{|E|}$ satisfies

$$||\Delta||_\infty \leq O\left(\frac{b \log(n/\delta)}{k}\right).$$

The interpretation of this lemma is as follows. The first term, $V(F - p)$, comes from the linear Taylor expansion of $\log R$ about its limit of $\log \rho$, while the second term, Δ, comes from bounding the rest of the terms in the Taylor expansion. The above lemma shows that $||\Delta||_\infty$ tends to be on the order of $O(1/k)$. As expected, this is a faster decay as compared to the first term: indeed, since $F - p$ is the average of k independent random variables, one for each comparison, by central-limit considerations we expect $F - p$ to be on the order of $O(1/\sqrt{k})$.

Furthermore we remark that $\log R$ could potentially have an infinite entry (this can happen if one node wins every comparison against a neighbor). The above lemma implies that the probability of that is at most δ under the lower bound $k \geq \Omega(b \log(n/\delta))$.

By taking $\delta = n/e^{k^{3/4}}$ in this proposition, which satisfies $\delta \leq e^{-1}$ and $k \geq \Omega(b \log(n/\delta))$ for k large enough, we obtain the following result.

Corollary 4. *Let $f(w, b, G)$ be any function of the weights w, the constant b, and the graph G (and thus of n). Then*

$$P\left(||\Delta||_\infty \geq f(w, b, G)\right) = O\left(ne^{-k^{3/4}}\right).$$

Our next lemma follows up on these observations by decomposing the error from Eq. (22) into three parts.

Lemma 5. *We have*

$$\hat{Y} - y = A + B + C,$$

where

$$\begin{aligned} A &= (\mathrm{diag}(w)^{-1} L_V \mathrm{diag}(w)^{-1})^\dagger \mathrm{diag}(w)^{-1} M(F - p) \\ B &= (\mathrm{diag}(w)^{-1} L_V \mathrm{diag}(w)^{-1})^\dagger \mathrm{diag}(w)^{-1} M V^{-1} \Delta \\ C &= (\mathrm{diag}(w)^{-1} L_{\hat{V}} \mathrm{diag}(w)^{-1})^\dagger \mathrm{diag}(w)^{-1} M \hat{V}^{-1} \\ &\quad -(\mathrm{diag}(w)^{-1} L_V \mathrm{diag}(w)^{-1})^\dagger \mathrm{diag}(w)^{-1} M V^{-1}) \\ &\quad (\log R - \log \rho), \end{aligned}$$

and the vector Δ is defined through Proposition 3.

Before we proceed to the proof, which is quite short, we discuss where each of the three terms comes from. Observe that the first term, A, is obtained by replacing all instances of $\hat{V}^{-1}$ with V^{-1} in Eq. (22) and further replacing $\log R - \log \rho$ by its first order Taylor expansion $V(F - p)$. Naturally, this replacement is going to create errors, and these are handled by adding the terms B and C. The term B corrects the error from replacing $\log R - \log \rho$ by $V(F - p)$, while the term C corrects the error from replacing $\hat{V}^{-1}$ by V^{-1}.

We next give the proof of this lemma.

Proof of Lemma 5. Indeed, beginning from Eq. (22), we can express $\hat{Y} - y$ as

$$\begin{aligned} &(\mathrm{diag}(w)^{-1} L_{\hat{V}} \mathrm{diag}(w)^{-1})^\dagger \mathrm{diag}(w)^{-1} M \hat{V}^{-1} (\log R - \log \rho) \\ &= (\mathrm{diag}(w)^{-1} L_V \mathrm{diag}(w)^{-1})^\dagger \mathrm{diag}(w)^{-1} M V^{-1} (\log R - \log \rho) \\ &+ ((\mathrm{diag}(w)^{-1} L_{\hat{V}} \mathrm{diag}(w)^{-1})^\dagger \mathrm{diag}(w)^{-1} M \hat{V}^{-1} - \\ &(\mathrm{diag}(w)^{-1} L_V \mathrm{diag}(w)^{-1})^\dagger \mathrm{diag}(w)^{-1} M V^{-1}) (\log R - \log \rho) \\ &= (\mathrm{diag}(w)^{-1} L_V \mathrm{diag}(w)^{-1})^\dagger \mathrm{diag}(w)^{-1} M V^{-1} V(F - p) \\ &+ (\mathrm{diag}(w)^{-1} L_V \mathrm{diag}(w)^{-1})^\dagger \mathrm{diag}(w)^{-1} M V^{-1} \Delta \\ &+ ((\mathrm{diag}(w)^{-1} L_{\hat{V}} \mathrm{diag}(w)^{-1})^\dagger \mathrm{diag}(w)^{-1} M \hat{V}^{-1} - \\ &(\mathrm{diag}(w)^{-1} L_V \mathrm{diag}(w)^{-1})^\dagger \mathrm{diag}(w)^{-1} M V^{-1}) (\log R - \log \rho) \\ &= A + B + C. \end{aligned}$$

□

In the next three sections we bound each term in Lemma 5 separately. We will see that B and C decay with k at a faster pace than A, which will thus determine the asymptotic dependence of $\hat{y} - Y$ with k. Our first step is to bound the norm of A.

5.3. Part A of the error

Lemma 6. *With probability* $1-\delta$,

$$||A||_2^2 \le O\left(\frac{b}{k}\right) \text{Tr}\left[(\text{diag}(w)^{-1} L_V \text{diag}(w)^{-1})^\dagger\right] \left(1+\log\frac{1}{\delta}\right) \tag{23}$$

Moreover, for any function $f(w,b,G)$,

$$P(||A||_2^2 \ge f(w,b,G)) = O\left(e^{-k/g(w,b,G)}\right), \tag{24}$$

for some function $g(\cdot,\cdot,\cdot)$ *of* w,b,G *that can be determined from* f.

Proof. By definition,

$$\begin{aligned} A &= (\text{diag}(w)^{-1} L_V \text{diag}(w)^{-1})^\dagger \text{diag}(w)^{-1} M(F-p) \\ &= (\text{diag}(w)^{-1} L_V \text{diag}(w)^{-1})^\dagger \text{diag}(w)^{-1} M V^{-1/2} \\ &\quad V^{1/2}(F-p) \end{aligned}$$

Now let X_{ij}^l be the Bernoulli random variable denoting the outcome of l'th toss across edge (i,j) (i.e., 1 if i wins, 0 otherwise). Observe that

$$\text{var}(X_{ij}^l - p_{ij}) = p_{ij}(1-p_{ij}) = \frac{w_i w_j}{(w_i+w_j)^2} = v_{ij}^{-1},$$

and therefore

$$\text{var}\left(\sqrt{v_{ij}}(X_{ij}^l - p_{ij})\right) = 1.$$

After this calculation, remembering that $V = \text{diag}(v_{ij})$, we can write

$$A = \left[(\text{diag}(w)^{-1} L_V \text{diag}(w)^{-1})^\dagger \text{diag}(w)^{-1} M V^{-1/2}\right]\left[V^{1/2}(F-p)\right],$$

where the term in the right brackets is a random vector with zero mean and variance $1/k$. We can rewrite this as

$$\begin{aligned} A = &\left[\frac{1}{\sqrt{k}}(\text{diag}(w)^{-1} L_V \text{diag}(w)^{-1})^\dagger \text{diag}(w)^{-1} M V^{-1/2}\right] \\ &\left[\sqrt{k} V^{1/2}(F-p)\right], \end{aligned} \tag{25}$$

and now the term in brackets has zero mean and unit variance. Let us introduce the notation

$$P = (\text{diag}(w)^{-1} L_V \text{diag}(w)^{-1})^\dagger \text{diag}(w)^{-1} M V^{-1/2}. \tag{26}$$

We have that $\sqrt{k}V^{1/2}(F-p)$ is a subgaussian random variable with subgaussian parameter of $O(\sqrt{b})$ (this follows from observing that the outcome of l'th toss, $X_{ij}^l - p$, has support contained in $[-1,1]$, as well as rules for adding and scaling subgaussian random variables). Via Theorem 2.1 of (Hsu *et al.* , 2012) we have that[4]

$$P\left(||A||_2^2 \ge b\left(\frac{\text{Tr}(PP^T)}{k}(1+4t)\right)\right) \le e^{-t}.$$

Choosing $t = \log(1/\delta)$ and using $\text{Tr}(P^TP) = \text{Tr}(PP^T) = \text{Tr}\left[(\text{diag}(w)^{-1} L_V \text{diag}(w)^{-1})^\dagger\right]$, yields Eq. (23).

To prove Eq. (24), observe that, as a consequence of Eq. (23), for large enough k we must have that $\log(1/\delta)$ has to scale linearly with $k/g(w,b,G)$ in order for $||A||_2^2 > f(w,b,G)$, where we spare ourselves the trouble of writing out the function $g(w,b,G)$ in terms of $f(w,b,G)$. This proves Eq. (23). □

5.4. Part B of the error

We next turn our attention to the second term in Lemma 5, namely the vector B. While Lemma 6 showed that the entries A_i of A effectively decay at an $O(1/\sqrt{k})$ rate, our next lemma shows that B_i decays at the faster $O(1/k)$ rate. The lemma requires a few definitions from electric circuit theory, which we next provide.

Given a weighted undirected graph, we can talk about the effective resistance between any two nodes in the graph by treating every edge (i,j) as if it has a resistor of resistance equal to the weight of that edge. We will define the effective resistance between nodes i and j by $R_{\text{eff}}(i,j)$. We then define

$$R_{\max}(i) = \max_{j=1,\ldots,n} R_{\text{eff}}(i,j), \tag{27}$$

to be the average effective resistance between node i and the rest of the nodes in the graph. For a formal analysis of the electric theory of graphs, we refer the reader to Chapter 4 of (Vishnoi, 2013).

Lemma 7. *Let* $q \in \mathbb{R}^{\overrightarrow{E}}$ *be a positive vector. Then for every* $i = 1,\ldots,n$,

$$\begin{aligned} &\left|\left|\text{diag}(q)^{-1} M^T \text{diag}(w)^{-1}(\text{diag}(w)^{-1} L_q \text{diag}(w)^{-1})^\dagger e_i\right|\right|_1 \\ &\le w_i\sqrt{S_q R_{\max}(i)}, \end{aligned}$$

where

$$S_q = \sum_{(i,j)\in\overrightarrow{E}} q_{ij}^{-1},$$

L_q *is the Laplacian of the weighted graph with weights* q_{ij}^{-1}, *and* $R_{\max}(i)$ *is defined as in Eq. (27) for a graph with the weights* q.

[4]Strictly speaking, the reference (Hsu *et al.* , 2012) only bounds $P(||A||_2^2 \ge u)$ for $A = PZ$ when P is square. In our case, P is rectangular. However, we observe that any concentration bound for A in terms of $\text{tr}(PP^T)$ proved for the case when P is square immediately implies the same bound when P is rectangular. This follows because

$$||A||_2^2 = A^TA = Z^TP^TPZ = ||QZ||_2^2,$$

where Q is the psd square root of P^TP. Thus we can apply the results of (Hsu *et al.* , 2012) to bound $P(||A||_2^2 \ge u) = P(||QZ||_2^2 \ge u)$; and since $\text{tr}(QQ^T) = \text{tr}(PP^T)$, the result will be exactly the same as if we simply ignored the assumption of (Hsu *et al.* , 2012) that P is square.

Proof. The first part of the proof will consist in giving an electrical interpretation to the vector whose norm we want to bound. This will then allow us to apply circuit theory result to establish the Lemma.

Consider turning the comparison graph into a circuit, with edge (i, j) having resistance q_{ij}. In that case, for every vector y whose entries sum to 0, any solution x of

$$L_q x = y$$

can be interpreted as a vector of electric potential consistent with currents y going in and out of nodes in the network. This is in particular true for

$$x^* = \text{diag}(w)^{-1}(\text{diag}(w)^{-1}L_q\text{diag}(w)^{-1})^\dagger \text{diag}(w)^{-1}y. \tag{28}$$

Coming back to the problem of the Lemma, let us adopt the notation

$$f = \text{diag}(q)^{-1}M^T\text{diag}(w)^{-1}(\text{diag}(w)^{-1}L_q\text{diag}(w)^{-1})^\dagger e_i.$$

Since w is in the kernel of $\text{diag}(w)^{-1}L_q\text{diag}(w)^{-1}$, which is symmetric, we have $(\text{diag}(w)^{-1}L_q\text{diag}(w)^{-1})^\dagger w = 0$. Together with the observation that $\text{diag}(w)^{-1}w_i e_i = e_i$ and $\text{diag}(w)^{-1}w^2 = w$, where w^2 means the element-wise square of the entries of w, this allows re-expressing f as

$$\begin{aligned} f =& \text{diag}(q)^{-1}M^T\text{diag}(w)^{-1}(\text{diag}(w)^{-1}L_q\text{diag}(w)^{-1})^\dagger \\ & \text{diag}(w)^{-1}\left(w_i e_i - \frac{w_i}{||w||_2^2}w^2\right), \end{aligned}$$

Observe now that the entries of $\left(w_i e_i - \frac{w_i}{||w||_2^2}w^2\right)$ sum to zero. Hence we can apply the interpretation of (28), which states that

$$\text{diag}(w)^{-1}(\text{diag}(w)^{-1}L_q\text{diag}(w)^{-1})^\dagger \text{diag}(w)^{-1}\left(w_i e_i - \frac{w_i}{||w||_2^2}w^2\right)$$

is the vector of electric potentials when we put w_i units of current into node i and take out $w_i(w_j^2/||w||_2^2)$ out of node j. Moreover, electrical circuit results show f is then the vector of edge currents corresponding to this setup.

We may further view f as the superposition of n current flows, with the j'th flow f^j obtained by putting $w_j^2 w_i/||w||_2^2$ units of current at i and taking the same amount out of j. We will write this as

$$f = f^1 + \cdots + f^n.$$

The advantage of this representation is that we may apply Thompson's principle (Theorem 4.7 of (Vishnoi, 2013)) to each flow f^j. Slightly rephrased, that theorem states that the effective resistance between nodes i and j satisfies

$$\frac{w_i^2 w_j^4}{||w||_2^4}R_{\text{eff}}(i,j) = \sum_{e\in\overrightarrow{E}} q_e(f^j)_e^2, \tag{29}$$

where for $e = (a, b)$ we use q_e and q_{ab} interchangeably. We may rewrite this as

$$||f^j .* \sqrt{q}||_2 = w_i w_j^2 \frac{\sqrt{R_{\text{eff}}(i,j)}}{||w||_2^2},$$

where we use ".*" to denote the elementwise product of two vectors; note that both f^j and v can be viewed as vectors in $\mathbb{R}^{|\overrightarrow{E}|}$. To conclude, it follows from Cauchy-Schwarz that

$$\begin{aligned} ||f||_1 &\leq \sum_{j=1}^n ||f^j||_1 \\ &= \sum_{j=1}^n ||f^j .* \sqrt{q_j} .* \sqrt{q_j}^{-1}||_1 \\ &\leq \sum_{j=1}^n ||f^j .* \sqrt{q_j}||_2\sqrt{S}, \end{aligned}$$

and applying the bound (29) leads then to

$$\begin{aligned} ||f||_1 &\leq \sum_{j=1}^n w_i w_j^2 \frac{\sqrt{R_{\text{eff}}(i,j)}}{||w||_2^2}\sqrt{S} \\ &\leq w_i\sqrt{SR_{\max}(i)}. \end{aligned}$$

□

With the above definitions in place, we can state our decay bound on the elements B_i of the vector B.

Lemma 8. *If $\delta \leq e^{-1}$ and $k = \Omega(b\log(n/\delta))$, we have that with probability $1-\delta$,*

$$B_i \leq O\left(w_i \frac{b\log(n/\delta)\sqrt{SR_{\text{avg}}(i)}}{k},\right) \tag{30}$$

for all $i = 1, \dots, n$, where

$$S = \sum_{(i,j)\in\overrightarrow{E}} v_{ij}^{-1},$$

and $R_{\max}(i)$ is defined as in Eq. (27) for the graph where the edge (i, j) has resistance v_{ij}. Moreover, for any function $f(w, b, G)$ of the graph G and the weights w, we have that

$$P(B_i \geq f(w,b,G)) \leq O\left(ne^{-k^{3/4}}\right). \tag{31}$$

Proof. By definition we have that

$$\begin{aligned} B_i &= [(\text{diag}(w)^{-1}L_V\text{diag}(w)^{-1})^\dagger\text{diag}(w)^{-1}MV^{-1}\Delta]_i \\ &= e_i^T(\text{diag}(w)^{-1}L_V\text{diag}(w)^{-1})^\dagger\text{diag}(w)^{-1}MV^{-1}\Delta \\ &\leq ||\Delta||_\infty \left|\left|V^{-1}M^T\text{diag}(w)^{-1}(\text{diag}(w)^{-1}L_V\text{diag}(w)^{-1})^\dagger e_i\right|\right|_1 \\ &\leq O\left(\frac{b\log(n/\delta)}{k}\right)\left|\left|V^{-1}M^T\text{diag}(w)^{-1}(\text{diag}(w)^{-1}\right.\right. \\ &\qquad \left.\left. L_V\text{diag}(w)^{-1})^\dagger e_i\right|\right|_1, \end{aligned} \tag{32}$$

where the third inequality used Holder's inequality and the last step used Proposition 3. Now using Lemma 7, the proof of Eq. (30) is concluded.

As for Eq. (31), it follows immediately from Eq. (32) by taking $\delta = n/e^{k^{3/4}}$; for large enough k, we will have both $\delta \leq e^{-1}$ and $k \geq \Omega(b \log(n/\delta))$ required for that equation to hold. □

5.5. Part C of the error

We next turn to the analysis of the final term C in Lemma 5. Unfortunately, this term is the most cumbersome, and will require quite a number of calculations. Our starting point will be to argue that since C_i is by definition

$$e_i^T((\mathrm{diag}(w)^{-1} L_{\hat{V}} \mathrm{diag}(w)^{-1})^\dagger \mathrm{diag}(w)^{-1} M \hat{V}^{-1} - (\mathrm{diag}(w)^{-1} L_V \mathrm{diag}(w)^{-1})^\dagger \mathrm{diag}(w)^{-1} M V^{-1})(\log R - \log \rho)$$

It can be bounded as

$$C_i \leq ||e_i^T((\mathrm{diag}(w)^{-1} L_{\hat{V}} \mathrm{diag}(w)^{-1})^\dagger \mathrm{diag}(w)^{-1} M \hat{V}^{-1} - (\mathrm{diag}(w)^{-1} L_V \mathrm{diag}(w)^{-1})^\dagger \mathrm{diag}(w)^{-1} M V^{-1})||_1 \, || \log R - \log \rho||_\infty \quad (33)$$

We now proceed to bound both of the terms on the right-hand side. We begin with the second term, as it's analysis is the easiest.

Lemma 9. *If $\delta \leq e^{-1}$ and $k \geq \Omega(b \log(n/\delta))$, then with probability $1 - 2\delta$,*

$$||\log R - \log \rho||_\infty \leq O\left(\sqrt{\frac{b \log(n/\delta)}{k}}\right) + O\left(\frac{b \log(n/\delta)}{k}\right), \quad (34)$$

Moreover, if $f(w, b, G)$ is any function of w, b, and G, then

$$P\left(|| \log R - \log \rho||_\infty \geq f(w, b, G)\right) \leq O\left(ne^{-k^{3/4}}\right). \quad (35)$$

Proof. From Proposition 3, we have that

$$\log R - \log \rho = V(F - p) + \Delta, \quad (36)$$

with

$$||\Delta||_\infty \leq O\left(\frac{b \log(n/\delta)}{k}\right) \quad (37)$$

with probability $1 - \delta$. This leads to the second term in the statement of the lemma. For the first term, we will need to bound $||V(F - p)||_\infty$.

To that end, we observe that Lemma 1 in (Hendrickx *et al.*, 2019) proved that if $\delta \leq e^{-1}$ and $k \geq \Omega(b \log(n/\delta))$, then

$$P\left(||F - p||_\infty \geq \sqrt{\frac{O(\log(n/\delta))}{kb}}\right) \leq \delta. \quad (38)$$

Thus with probability $1 - \delta$,

$$||V(F - p)||_\infty \leq O\left(\sqrt{\frac{b \log(n/\delta)}{k}}\right). \quad (39)$$

Putting together the two probability $1 - \delta$ bounds of Eq. (37) and Eq. (39) via the union bound proves Eq. (34). As for Eq. (35), it follows from Eq. (36) and Eq. (39), by taking $\delta = n/e^{k^{3/4}}$ which, for large enough k, satisfies the conditions $\delta \leq e^{-1}$ and $k = \Omega(b \log(n/\delta))$; and Corollary 4. □

Having established this lemma, we have a bound on the second term in Eq. (33); we now turn to analyzing the first term the same equation. Let us introduce the following notation for the first term,

$$\Gamma_i = ||e_i^T((\mathrm{diag}(w)^{-1} L_{\hat{V}} \mathrm{diag}(w)^{-1})^\dagger \mathrm{diag}(w)^{-1} M \hat{V}^{-1} - (\mathrm{diag}(w)^{-1} L_V \mathrm{diag}(w)^{-1})^\dagger \mathrm{diag}(w)^{-1} M V^{-1})||_1, \quad (40)$$

which will make the ensuing discussion more compact. We need to upper bound Γ_i, which is to say we need to upper bound the change that comes from replacing V by $\hat{V}$; what makes things difficult, however, is that the expression involves the pseudoinverses L_V and $L_{\hat{V}}^\dagger$. We will take the "brute force" approach of writing out the derivative of the expression inside the one-norm in Eq. (40) and integrating it over the path between V and $\hat{V}$.

To that end, let us define the function

$$H_i(u) = e_i^T Q_u^\dagger \mathrm{diag}(w)^{-1} M \mathrm{diag}(u),$$

where

$$Q_u = \mathrm{diag}(w)^{-1} M \mathrm{diag}(u) M^T \mathrm{diag}(w)^{-1},$$

a weighted graph Laplacian when $u \in \mathbb{R}^{|\vec{E}|}$ is the vector of weights, scaled left and right by $\mathrm{diag}(w)^{-1}$. As should be clear from matching up dimensions in this equation, H_i maps $\mathbb{R}^{|\vec{E}|}$ into $\mathbb{R}^{|\vec{E}|}$. We will slightly abuse notation by writing expressions like $H_i(v_{ab}^{-1})$, which should be understood to mean H_i applied to the vector in $\mathbb{R}^{|\vec{E}|}$ obtained by stacking up the quantities v_{ab}^{-1} as (a, b) ranges over the edges in $\vec{E}$.

By the definition of the weighted Laplacian, the expression Γ can be written as

$$\Gamma_i = ||H_i(\hat{V}_{ab}^{-1}) - H_i(v_{ab}^{-1})||_1.$$

To make the connection to the underlying coin tosses more explicit, we can write $\hat{V}_{ab}^{-1}$ as a function of the fractions

F_{ab}, and likewise v_{ab}^{-1} can be written as a function of the true probabilities p_{ab}. To spell this out observe that

$$v_{ab}^{-1} = \frac{1}{w_a/w_b + w_b/w_a + 2} = \frac{w_a w_b}{(w_a + w_b)^2} = p_{ab}(1 - p_{ab}),$$

and likewise

$$\hat{V}_{ab}^{-1} = F_{ab}(1 - F_{ab}).$$

Thus defining

$$U(x_1, \ldots, x_{|\overrightarrow{E}|}) \quad = \quad \left(x_1(1 - x_1), \ldots, x_{|\overrightarrow{E}|}(1 - x_{|\overrightarrow{E}|})\right), \tag{41}$$

we can now write

$$\Gamma_i = ||H_i(U(F_{ab})) - H_i(U(p_{ab}))||_1. \tag{42}$$

We thus proceed by upper bounding the gradient of the function $H_i(U(x_{ab}))$ by using the chain rule. Our next lemma takes the first step in this direction by giving an explicit expression for $\partial H_i/\partial u_{ab}$ for some fixed indices $a, b \in \{1, \ldots, n\}$.

Lemma 10. *Let a, b be elements of $\{1, \ldots, n\}$. If the vector $u \in \mathbb{R}^{|\overrightarrow{E}|}$ is elementwise positive, then*

$$\frac{dH_i}{du_{ab}} = e_i^T \left[Q_u^\dagger \mathrm{diag}(w)^{-1}(e_a - e_b)\right] \left[-(e_a - e_b)^T \mathrm{diag}(w)^{-1} Q_u^\dagger \mathrm{diag}(w)^{-1} M \mathrm{diag}(u) + e_{ab}^T\right]$$

Here e_{ab} denotes a column vector in $\mathbb{R}^{|\overrightarrow{E}|}$ with a one in the (a, b) entry and zeros elsewhere. Note that the first expression in brackets is a column vector in $\mathbb{R}^n$ while the second expression in brackets is a row vector in $\mathbb{R}^{|\overrightarrow{E}|}$.

Proof. We first compute

$$\begin{aligned}\frac{dQ_u}{du_{ab}} &= \mathrm{diag}(w)^{-1} M e_{ab} e_{ab}^T M^T \mathrm{diag}(w)^{-1} \qquad (43)\\ &= \mathrm{diag}(w)^{-1}(e_a - e_b)(e_a - e_b)^T \mathrm{diag}(w)^{-1}.\end{aligned}$$

We next use this to find the derivative of $Q_u^\dagger$. We use Theorem 4.3 from (Golub & Pereyra, 1973), which provides an expression for $\partial Q_u/\partial u_{ab}$ in a neighborhood of a point where the rank of Q_u is constant. That formula applies in our case because as long as $u > 0$ and G is a connected graph, we will have that the rank of G equals $n - 1$ (see Section 2.5 of (Brualdi & Ryser, 1991)).

The expression from Theorem 4.3 of (Golub & Pereyra, 1973) is

$$\begin{aligned}\frac{dA^\dagger}{dx} = \quad & -A^\dagger \frac{dA}{dx} A^\dagger + A^\dagger A^{\dagger T} \frac{dA^T}{dx}(I - AA^\dagger)\\ & + (I - A^\dagger A)\frac{dA^T}{dx} A^{\dagger T} A^\dagger \qquad (44)\end{aligned}$$

When we plug in $A = Q_u$, the expression simplifies considerably. Observe indeed that

$$\begin{aligned}(I - Q_u^\dagger Q_u)\frac{dQ_u^T}{du_{ab}} =& (I - Q_u^\dagger Q_u)\mathrm{diag}(w)^{-1}(e_a - e_b)\\ & (e_a - e_b)^T \mathrm{diag}(w)^{-1}\\ =& \left[(I - Q_u^\dagger Q_u)\mathrm{diag}(w)^{-1}(e_a - e_b)\right]\\ & (e_a - e_b)^T \mathrm{diag}(w)^{-1}.\end{aligned}$$

Now $(I - Q_u^\dagger Q_u)$ is the orthogonal projector on the kernel of Q_u, which is $\mathrm{span}\{\mathrm{w}\}$, and the vector $\mathrm{diag}(w)^{-1}(e_a - e_b)$ is orthogonal to w. Hence the expression above, and thus the second term in Eq. (44), is zero. A symmetric argument shows the third term is zero as well. Thus

$$\begin{aligned}\frac{dQ_u^\dagger}{du_{ab}} &= -Q_u^\dagger \frac{dQ_u}{du_{ab}} Q_u^\dagger\\ &= -Q_u^\dagger \mathrm{diag}(w)^{-1}(e_a - e_b)(e_a - e_b)^T \mathrm{diag}(w)^{-1} Q_u^\dagger.\end{aligned}$$

We can now use this to compute the derivative of H using the chain rule. Indeed, we can write $\frac{dH}{du_{ab}}$ as

$$\begin{aligned}& e_i^T \frac{d\left(\mathrm{diag}(w)^{-1} M \mathrm{diag}(u) M^T \mathrm{diag}(w)^{-1}\right)^\dagger}{du_{ab}} \mathrm{diag}(w)^{-1} M \mathrm{diag}(u)\\ & + e_i^T \left(\mathrm{diag}(w)^{-1} M \mathrm{diag}(u) M^T \mathrm{diag}(w)^{-1}\right)^\dagger \mathrm{diag}(w)^{-1} M \frac{d\mathrm{diag}(u)}{du_{ab}}\\ = & -e_i^T Q_u^\dagger \mathrm{diag}(w)^{-1}(e_a - e_b)(e_a - e_b)^T \mathrm{diag}(w)^{-1} Q_u^\dagger\\ & \mathrm{diag}(w)^{-1} M \mathrm{diag}(u)\\ & + e_i^T Q_u^\dagger \mathrm{diag}(w)^{-1}(e_a - e_b) e_{ab}^T.\end{aligned}$$

After some rearranging, this gives the statement of the lemma. ☐

With this explicit expression for the gradient of H in place, we can proceed to upper bound Γ_i. However, we first need the following lemma, which will be used in one of the intermediate steps of the bound.

Lemma 11. *If u is elementwise positive, then for any $i, a, b \in \{1, \ldots, n\}$,*

$$|e_i^T Q_u^\dagger \mathrm{diag}(w)^{-1}(e_a - e_b)| \quad \leq \quad w_i R_{\mathrm{eff}}(a, b) \tag{45}$$

Proof. We prove this by appealing to the electrical interpretation just as we did in the proof of Lemma 8. Using the observations made in the proof of that lemma, we interpret $e_i^T \mathrm{diag}(w)^{-1} Q_u^\dagger \mathrm{diag}(w)^{-1}(e_a - e_b)$ is the potential at node i when a single unit of current is injected at a and taken out at b in the circuit where the resistance of the edge (i, j) is u_{ij}^{-1}. Call this potential v_i. The quantity we are seeking the bound is thus just $w_i v_i$.

Note, however, that the largest potential is at $i = a$ and the smallest potential is at $i = b$. Thus we can instead upper bound $\max(|v_a|, |v_b|)$. Moreover, v_a is positive and v_b is negative because anything in the range of $Q_u^\dagger$ is orthogonal to the positive vector w. Moreover, since by Ohm's law $v_a - v_b = R_{\text{eff}}(a,b) \cdot 1$, we have that

$$\max(|v_a|, |v_b|) \le v_a - v_b = R_{\text{eff}}(a,b).$$

□

We now put all the pieces together and produce a high probability bound on the quantity C_i.

Lemma 12. *When $\delta \le e^{-1}$ and $k = \Omega(b\log(n/b))$, then we have that with probability $1-2\delta$, for all $i = 1, \dots, n$,*

$$C_i \le w_i bn(n-1)\log\left(\frac{n}{\delta}\right)\left(O\left(\frac{1}{k}\right) + O\left(\frac{\sqrt{b\log(n/\delta)}}{k^{1.5}}\right)\right).$$

Moreover, if $f(w,b,G)$ is any function of w, b, and G, then

$$P\left(C_i \ge f(w,b,G)\right) \le O\left(ne^{-k^{3/4}}\right). \tag{46}$$

Proof. Indeed, from Eq. (33) and Eq. (40), we have

$$C_i \le \Gamma_i || \log R - \log \rho ||_\infty.$$

Using the expression form Eq. (42), we can upper bound this as

$$C_i \le ||H_i(U(F_{ij})) - H_i(U(p_{ij}))||_1 || \log R - \log \rho||_\infty.$$

As a consequence of Theorem 4.3 of (Golub & Pereyra, 1973), $H_i(U(u_{ij}))$ is differentiable over the set of $u > 0$; we may therefore invoke the mean value theorem to obtain the bound

$$C_i \le ||\nabla H_i(U(z))(F-p)||_1 || \log R - \log\rho||_\infty,$$

where z is some point lying on the line interval connecting the vectors F and p. Using $||.||_1 \le n||.||_\infty$ and the definition of matrix norm, we can in turn upper bound this as

$$\begin{aligned} C_i \le & n||\nabla H_i(U(z))||_\infty ||F-p||_\infty || \log R - \log\rho||_\infty \\ =& n\left(\max_{j=1,\dots,n} ||e_j^T \nabla H_i(U(z))||_1\right) \\ & ||F-p||_\infty || \log R - \log\rho||_\infty \end{aligned} \tag{47}$$

where we used the standard fact that the infinity norm of a matrix is the largest one-norm of its rows.

The second and third quantities above have been bounded in the previous lemmas; only the quantities $||e_j^T \nabla H_i(U(z))||_1$ needs to be analyzed. However, observe that

$$\begin{aligned} \frac{\partial H_i(U(z))}{\partial z_{ij}} &= \frac{\partial H_i(u)}{\partial u_{ij}} \frac{\partial u_{ij}}{\partial z_{ij}} \\ &= \frac{\partial H_i(u)}{\partial u_{ij}}(1 - 2z_{ij}), \end{aligned}$$

where the last line used Eq. (41). Thus if $z \in [0,1]$, we have that

$$\left|\frac{\partial H_i(U(z))}{\partial z_{ij}}\right| \le \left|\frac{\partial H_i(u)}{\partial u_{ij}}\right|,$$

which in turn implies that

$$||e_j^T \nabla H_i(U(z))||_1 \le ||e_j^T \nabla H_i(u)||_1. \tag{48}$$

Thus what remains to be done is to obtain a bound on the quantity $||e_j^T \nabla H_i(u)||_1$. This can be done using the explicit expression for the partial derivatives of H_i derived in Lemma 10. Indeed, observe that the rows of $\nabla H_i(u)$ are the transposed vectors $\frac{\partial H_i}{\partial u_{ab}}$ from Lemma 10. Thus using that lemma we have

$$\begin{aligned} &\left|\left|\max_{j=1,\dots,n} e_j^T \nabla H(u)\right|\right|_1 \le \sum_{(a,b)\in E} \left|\left|\frac{\partial H_i}{\partial u_{ab}}\right|\right|_\infty \\ &\le \sum_{(a,b)\in E} \left|\left|e_i^T Q_u^\dagger \operatorname{diag}(w)^{-1}(e_a - e_b)\right|\right|_\infty \\ &\quad \left|\left|-(e_a-e_b)^T \operatorname{diag}(w)^{-1} Q_u^\dagger \operatorname{diag}(w)^{-1} M \operatorname{diag}(u) + e_{ab}^T\right|\right|_\infty \\ &\le 2\sum_{(a,b)\in E} w_i R_{\text{eff}}(a,b), \end{aligned} \tag{49}$$

where the last line used Eq. (45) as well as the observation that $(e_a - e_b)^T \operatorname{diag}(w)^{-1} Q_u^\dagger \operatorname{diag}(w)^{-1} M \operatorname{diag}(u) \in [-1,1]^{|E|}$. This last observation follows because each entry of this vector is the current in the graph where the resistance of $(i,j) \in E$ is u_{ij}^{-1} and a single unit of current is injected at a and taken out at b.

Our next step is to use Foster's identity (Foster, 1949; Tetali, 1994)

$$\sum_{(a,b)\in E} R_{\text{eff}}(a,b) u_{ab} = n-1,$$

to conclude from Eq. (49) that

$$\left|\left|\max_{j=1,\dots,n} e_j^T \nabla H(u)\right|\right|_1 \le w_i \frac{2(n-1)}{\min_{(a,b)\in E} u_{ab}}.$$

Plugging this into Eq. (47) we obtain

$$C_i \le w_i \frac{2n(n-1)}{\min_{(a,b)\in E} u_{ab}} ||F-p||_\infty || \log R - \log\rho||_\infty \tag{50}$$

Finally, we observe that, as a consequence of Eq. (38) and some algebra, when $k \ge \Omega(b\log(n/\delta))$, we have that $|F_{ij} - p_{ij}| < 1/(4b)$ with probability $1-\delta$. We can use this to bound the quantity $\min_{(a,b)\in E} u_{ab}$ appearing above. Indeed, $u = U(z)$ where z lies on the path between F_{ij} and p_{ij}, and thus $z_{ij} \ge 1/(4b)$ and $1 - z_{ij} \ge 1/(4b)$ for all $(i,j) \in E$. Then $[U(z)]_{ij} = z_{ij}(1-z_{ij}) \ge \frac{1}{8b}$.

Using this bound in conjunction with Eq. (39) to bound $||F-p||_\infty$ and Eq. (34) to bound $|| \log R - \log\rho||_\infty$ we obtain the first statement of the lemma.

We now turn to proving Eq. (46). Inspecting Eq. (50), we first argue that $1/\min_{(a,b)\in E} u_{ab}$ is $O(b)$ with probability $O(ne^{-k^{3/4}})$. Indeed, as discussed just above, each z_{ij} lies on the path between F_{ij} and p_{ij}; this means that that $\min_{ij} z_{ij} < 1/(2(b+1))$ by with probability $O(ne^{-k^{3/4}})$ by taking $\delta = n/e^{k^{3/4}}$ in Lemma 1 of (Hendrickx *et al.* , 2019); which implies $\max_{(i,j)} 1/u_{ij} = \max_{(i,j)}(1/z_{ij})(1/(1-z_{ij})) = O(b)$ with the same probability.

Thus returning to Eq. (50) and taking into account that $||F-p||_\infty = O(1)$ by definition, we see that in order for C_i to exceed some function $f(w,b,G)$ (and thus of the number n of nodes in G), it must be that $||\log R - \log\rho||_\infty$ exceeds some function of the w and the graph G (and of n); but that has been shown to happen with probability $O(ne^{-k^{3/4}})$ in Eq. (35). In summary, the only way for C_i to exceed a fixed function of w, b, G is one of two events to happen, both of which have probability $O(ne^{-k^{3/4}})$. □

5.6. Recombining the sources of errors

We now put together the bounds we have obtained on A_i, B_i, C_i into several general bounds on the error. Our first step will analyze the worst-case possible scalings. Here we have to consider even the scenario when one node wins all the comparisons to a neighbor, resulting in some F_{ij} which is as small as $O(1/k)$ due to the lines 5 and 7 of our WLSM algorithm.

Lemma 13. *With probability one, we have that for all $i = 1,\ldots,n$,*

$$\begin{aligned} |B_i| &= O_{w,b,G}(\log k) \\ |C_i| &= O_{w,b,G}(k\log k) \\ |\hat{Y}_i| &= O_{w,b,G}(\sqrt{k}\log k) \\ |\hat{W}_i^r| &\leq e^{O_{w,b,G}(\sqrt{k}\log k)}. \end{aligned}$$

where the $O_{w,b,G}(\cdot)$ notation hides factors depending on w, b, G.

Proof. We begin with the upper bound on B_i. Our starting point is the penultimate line of Eq. (32), which implies that $B_i = O_{w,b,G}\left(||\Delta||_\infty\right)$. Using

$$\Delta = \log R - \log\rho - V(F-P),$$

we immediately obtain that $||\Delta||_\infty = O(\log k)$ with probability one, since the smallest F_{ij} will be on the order of $1/k$ due the lines 5 and 7.

Next, we turn to the bound on C_i. Starting from Eq. (50), we use that $||\log R - \log\rho||_\infty = O_{w,b,G}(\log k)$. Since $u_{ij} = z_{ij}(1-z_{ij})$, the quantity z_{ij} is on the path between F_{ij} and p_{ij}, and F_{ij} cannot be smaller than $(1/2)/k$, we have that $1/\min_{(a,b)} u_{a,b} = O(k)$ with probability one. Plugging these observations into Eq. (50) completes the proof for C.

Finally, we turn to the bound on Y_i. Recall that Eq. (17) states that

$$\hat{Y} = \left(\mathrm{diag}(w)^{-1}L_{\hat{V}}\mathrm{diag}(w)^{-1}\right)^{\dagger}\mathrm{diag}(w)^{-1}M\hat{V}^{-1}\log R.$$

Moreover, Lemma 7 shows that

$$\begin{aligned} &\left|\left|\mathrm{diag}(q)^{-1}M^T\mathrm{diag}(w)^{-1}(\mathrm{diag}(w)^{-1}L_q\mathrm{diag}(w)^{-1})^{\dagger}e_i\right|\right|_1 \\ &\leq w_i\sqrt{\sum_{(i,j)} q_{ij}^{-1}R_{\max}(i)}. \end{aligned}$$

Since we have already shown that $\log R = O(\log k)$ with probability one, and since $\sum_{(i,j)} \hat{V}_{ij}^{-1} = O(k)$ with probability one, this proves the bound we need.

Finally, the bound on W_i^r follows from its definition,

$$\log\hat{W}^r = \mathrm{diag}(w)^{-1}\hat{Y},$$

together with the bound on $\hat{Y}_i$.

□

Our next step is to argue that a sufficiently high moment of the quantities B_i, C_i decays fast. We will rely on such moments in the ensuing analysis. It will suffice to use the fourth moment, as in the following lemma.

Lemma 14. *For all $i = 1,\ldots,n$,*

$$\begin{aligned} E[B_i^4] &= O_{w,b,G}\left(\frac{1}{k^4}\right) \\ E[C_i^4] &= O_{w,b,G}\left(\frac{1}{k^4}\right) \end{aligned}$$

where the $O_{w,b,G}(\cdot)$ notation hides all the factors that do not depend on δ and k. We remind that the notation O denotes a linear bound valid for a k large enough with respect to w b and G, and thus, crucially, independently of δ.

Proof. We prove the bound for B_i, and omit the proof of the bound on C_i, which is obtained by a similar development. We will use the identity

$$E[B_i^4] = \int_0^{+\infty} P(B_i^4 \geq u)\, du, \tag{51}$$

and leverage the result of Eq. (32), where we have shown that with probability $1-\delta$,

$$B_i^4 \leq O_{w,b,G}\left(\frac{\log^4(1/\delta)}{k^4}\right). \tag{52}$$

However, this result was shown under the conditions that $k \geq \Omega(b\log(n/\delta))$ and $\delta \leq e^{-1}$, and can thus only be used for certain values of u in the integral in (51). We will therefore decompose that integral in different sub-parts to which we apply different treatments.

Part 1: application of (52)
Given any u, we solve for the δ we must plug into Eq. (52) in order to bound $P(B_i^4 \geq u)$. This yields

$$P(B_i^4 \geq u) \leq e^{-(k^4 u/f_1(w,b,G))^{1/4}},$$

for some function $f_1(w,b,G)$. But this only holds under the conditions $\delta \leq e^{-1}$ and $k \geq \Omega(b\log(n/\delta))$. The former holds when $u \geq f_1(w,b,G)/k^4$; and the latter holds if

$$c\frac{1}{b^4} \geq \frac{\log^4(n/\delta)}{k^4}.$$

for some absolute constant c. Observe that

$$\begin{aligned}\frac{\log^4(n/\delta)}{k^4} &\leq \frac{8\log^4 n + 8\log^4(1/\delta)}{k^4}\\ &\leq \frac{64\log^4 n\log^4(1/\delta)}{k^4}\\ &= \frac{64(\log^4 n)u}{f_1(w,b,G)},\end{aligned}$$

where the second inequality follows from the implication $a \geq \frac{b}{b-1} \Rightarrow ab \geq a+b$, which applies here since $8\log^4 n \geq 1.8$ and $8\log^4(1/\delta) > 8$. We can then ensure that the condition $k \geq \Omega(b\log(n/\delta))$ holds by taking u satisfying

$$u \leq \frac{cf_1(w,b,G)}{64b^4\log^4 n} = c_1 f_1(w,b,G)(b\log n)^{-4}.$$

Hence,

$$\begin{aligned}&\int_{f_1(w,b,G)/k^4}^{c_1 f_1(w,b,G)(b\log n)^{-4}} P(B_i^4 \geq u)\,du\\ &\leq \int_{f_1(w,b,G)/k^4}^{c_1 f_1(w,b,G)(b\log n)^{-4}} e^{-(k^4 u/f_1(w,b,G))^{1/4}}\,du\\ &\leq \int_{f_1(w,b,G)/k^4}^{\infty} e^{-(k^4 u/f_1(w,b,G))^{1/4}}\,du\end{aligned}$$

A substitution $u = (f_1(w,b,G)/k^4)x$ transforms this last integral in

$$\int_1^{+\infty} e^{-x^{1/4}}\,\frac{f_1(w,b,G)}{k^4}dx.$$

Using $\int_1^{+\infty} e^{-x^{1/4}}\,dx = \frac{64}{e}$, we obtain thus

$$\int_{f_1(w,b,G)/k^4}^{c_1 f_1(w,b,G)(b\log n)^{-4}} P(B_i^4 \geq u)\,du = O_{w,b,G}\left(\frac{1}{k^4}\right). \tag{53}$$

Part 2: small values of u
For values of u below those covered in Eq. (53), we use the trivial bound

$$\begin{aligned}\int_0^{f_1(w,b,G)/k^4} P(B_i^4 \geq u)\,du \leq{}& \frac{f_1(w,b,G)}{k^4}\\ ={}& O_{w,b,G}\left(\frac{1}{k^4}\right).\end{aligned} \tag{54}$$

Part 3: large values of u
We now focus on the last part of the integral Eq. (51), i.e. values of u larger than $c_1 f_1(w,b,G)(b\log n)^{-4}$. It follows from Lemma 13 that $|B_i| = O_{w,b,G}(\log k)$ with probability 1. Besides, the bound of Eq. (31) implies that $P(B_i^4 \geq u) = O\left(ne^{-k^{3/4}}\right)$ for all $u = c_1 f_1(w,b,G)(b\log n)^{-4}$ and therefore for all larger u. These two facts lead to

$$\begin{aligned}&\int_{c_1 f_1(w,b,G)(b\log n)^{-4}}^{\infty} P(B_i^4 \geq u)\,du\\ &\leq \int_{c_1 f_1(w,b,G)(b\log n)^{-4}}^{O_{w,b,G}(\log^4 k)} P(B_i^4 \geq u)\,du\\ &\leq O\left(ne^{-k^{3/4}}\right)O_{w,b,G}(\log^4 k) < O_{w,b,G}\left(\frac{1}{k^4}\right)\end{aligned} \tag{55}$$

The desired bound on $E(B_i^4)$ follows then from Eq. (51) and the combination of Eq. (53), (54) and (55). ☐

As a consequence of the previous lemma, we can bound the fourth moment of the quantity $\hat{Y}_i - y_i$ in the next lemma.

Lemma 15.

$$E(\hat{Y}_i - y_i)^4 = O\left(E[A_i^4]\right) + O_{w,b,G}\left(\frac{1}{k^4}\right)$$

Proof. Indeed,

$$\begin{aligned}E(\hat{Y}_i - y_i)^4 &= E[(A_i + B_i + C_i)^4]\\ &= E\left[O\left(A_i^4 + B_i^4 + C_i^4\right)\right]\\ &= O\left(E[A_i^4]\right) + O_{w,b,G}\left(\frac{1}{k^4}\right),\end{aligned}$$

where the second step follows by Young's inequality and the last step uses Lemma 14. ☐

Taking stock at this point, we are proceeding to bound the fourth moment of the quantity $\hat{Y}_i - y_i = A_i + B_i + C_i$. Our previous lemma reduces this to the fourth moment of A_i, up to terms that decay as fast as $O(1/k^4)$. We thus need to analyze the fourth moment of A_i, which is done in the following lemma.

Lemma 16.

$$\sum_{i=1}^{n} \sqrt{E[A_i^4]} \leq O\left(\frac{\text{Tr}\left[(\text{diag}(w^r)^{-1} L_V \text{diag}(w^r)^{-1})^{\dagger}\right]}{k}\right)$$

Proof. Our starting point is the equations Eq. (25) and Eq. (26). Those equations allow us to write

$$A = \frac{1}{\sqrt{k}} P\left(\sqrt{k} V^{1/2}(F-p)\right).$$

Here the matrix P is defined in Eq. (26). What we need to show is that

$$\sum_{i=1}^{n} \sqrt{E[A_i^4]} = O\left(\frac{\text{Tr}(PP^T)}{k}\right).$$

Equivalently, we need to show that

$$\sum_{i=1}^{n} \sqrt{E[A_i^4]} = O\left(\frac{||P||_F^2}{k}\right),$$

where $||P||_F$ is the Frobenius norm of P. Letting $Z = \sqrt{k} V^{1/2}(F-p)$ we have that

$$A_i = \frac{1}{\sqrt{k}} \sum_{j=1}^{n} P_{ij} Z_j,$$

and therefore

$$\begin{aligned} E[A_i^4] &= \frac{1}{k^2} E \sum_{j,q,l,m=1}^{n} P_{ij} P_{iq} P_{il} P_{im} Z_j Z_q Z_l Z_m \\ &= \frac{1}{k^2} \sum_{q,l=1}^{n} P_{iq}^2 P_{il}^2 O(1), \end{aligned}$$

because $E[Z_m] = 0$ for all $m = 1, \ldots, n$, so that only terms of the form $Z_q^2 Z_l^2$ or Z_q^4 "survive" the expectation. The fact that these are $O(1)$ can be confirmed by expressing each entry $F_{ij} - p_{ij}$ as a sum of k independent centered random-variables supported in $[-1, 1]$ and variance $1/v_{ij}$, (the centered version of the results of each test).

$$\begin{aligned} \sqrt{E[A_i^4]} &= \frac{1}{k} O\left(\sqrt{\sum_{q,l=1}^{N} P_{iq}^2 P_{il}^2}\right) \\ &= \frac{1}{k} O\left(\sqrt{\left(\sum_{j=1}^{n} P_{ij}^2\right)^2}\right) \\ &= \frac{1}{k} O\left(\sum_{j=1}^{n} P_{ij}^2\right). \end{aligned}$$

It follows that

$$\sum_{i=1}^{n} \sqrt{E[A_i^4]} = O\left(\frac{||P||_F^2}{k}\right),$$

and we are done.

□

We are almost ready to put together all the pieces and prove the final theorem. It turns out that we need a technical estimate on how big the ratio $\hat{W}_i^r / w_i^r$ can be; this will be needed to bound various worst-case events. Our next lemma shows that the expectation of the fourth power of this quantity is constant. This will be helpful in the proof of our main theorem, where we will at one point need to interchange these quantities.

Lemma 17. *For large enough k and all $i = 1, \ldots, n$, we have that*

$$E\left[\frac{\max\left((w_i^r)^4, (\hat{W}_i^r)^4\right)}{(w_i^r)^4}\right] = O(1).$$

Proof. Indeed, from

$$\hat{Y} - y = A + B + C,$$

we have

$$\sum_{i=1}^{n} (\hat{Y}_i - y_i)^2 \leq 4 \sum_{i=1}^{n} (A_i^2 + B_i^2 + C_i^2)$$

Using Eq. (18) and Eq. (21) to obtain

$$\hat{Y}_i - y_i = w_i(\log \hat{W}_i^r - \log w_i^r), \tag{56}$$

we obtain

$$\begin{aligned} \sum_{i=1}^{n} \log^2 \frac{\hat{W}_i^r}{w_i^r} &\leq \frac{4}{w_{\text{min}}^2} \sum_{i=1}^{n} (A_i^2 + B_i^2 + C_i^2) \\ &= \frac{4}{w_{\text{min}}^2} (||A||_2^2 + ||B||_2^2 + ||C||_2^2). \end{aligned}$$

In particular, the sum of the $\log^2 \frac{\hat{W}_i^r}{w_i^r}$ is larger then or equal to 16 only if at least one among $|A||_2^2$, $||B||_2^2$ and $||C||_2^2$ exceeds w_{min}^2). This implies that

$$\begin{aligned} P\left(\max_i \frac{\hat{W}_i^r}{w_i^r} \geq e^4\right) &= P\left(\max_i \log \frac{\hat{W}_i^r}{w_i^r} \geq 4\right) \\ &\leq P\left(\max_i \log^2 \frac{\hat{W}_i^r}{w_i^r} \geq 16\right) \\ &\leq P\left(\sum_i \log^2 \frac{\hat{W}_i^r}{w_i^r} \geq 16\right) \\ &\leq P(||A||_2^2 > w_{\text{min}}^2) + P(||B||_2^2 > w_{\text{min}}^2) \\ &\quad + P(||C||_2^2 > w_{\text{min}}^2), \end{aligned}$$

By Eq. (24), Eq. (31), and Eq. (46), we have that

$$P\left(\max_i \left|\frac{\hat{W}_i^r}{w_i^r}\right| \geq e^4\right) = O\left(e^{-k/g(w,b,G)}\right) + O\left(ne^{-k^{3/4}}\right)$$

Moreover, by Lemma 13 we have that with probability one

$$\max_i W_i^r \leq e^{O_{b,w,G}\left(\sqrt{k}\log k\right)}.$$

Putting this all together, we can bound $E\max_i \left|\frac{(\hat{W}_i^r)^4}{(w_i^r)^4}\right|$ by

$$(e^4)^4 + e^{O_{b,w,G}\left(\sqrt{k}\log k\right)}\left[O\left(e^{-k/g(w,b,G)}\right) + O\left(ne^{-k^{3/4}}\right)\right]$$

For large enough k, we therefore have

$$E\max_i \left|\frac{(\hat{W}_i^r)^4}{(w_i^r)^4}\right| \leq e^{16} + 1,$$

which implies the lemma. □

We note that constant such as e^{16} in the proof above are large, but independent of the problem parameters. Smaller constant could have been obtained at the cost of a further increase of the proof complexity.

With all these results in place, we can finally prove our first main result. Our concentration results on B_i, C_i will allow us to argue we can essentially ignore them when k is large enough; and our bound on the expectation of A_i will turn out to give us exactly the expression in Theorem 1.

Proof of Theorem 1. Using the inequality

$$\left|e^a - e^b\right| \leq \max(e^a, e^b)|a - b|,$$

we obtain that

$$|\hat{W}_i^r - w_i^r| \leq \max(w_i^r, \hat{W}_i^r)|\log \hat{W}_i^r - \log w_i^r|. \quad (57)$$

We thus have

$$\begin{aligned} E||\hat{W}^r - w^r||_2^2 &\leq E\sum_{i=1}^n \max(w_i^r, \hat{W}_i^r)^2(\log \hat{W}_i^r - \log w_i^r)^2 \\ &= E\sum_{i=1}^n \frac{\max(w_i^r, \hat{W}_i^r)^2}{(w_i^r)^2}(w_i^r)^2(\log \hat{W}_i^r - \log w_i^r)^2 \\ &\leq \sum_{i=1}^n \sqrt{E\frac{\max(w_i^r, \hat{W}_i^r)^4}{(w_i^r)^4}} \\ &\quad \sqrt{E(w_i^r)^4(\log \hat{W}_i^r - \log w_i^r)^4} \end{aligned}$$

where we have used Cauchy-Schwarz. Using Lemma 17 and the definition of Y, followed by Lemma 15, we then obtain

$$\begin{aligned} E||\hat{W}^r - w^r||_2^2 &\leq \sum_{i=1}^n O(1)\sqrt{E(\hat{Y}_i - y_i)^4} \\ &= O\left(\sum_{i=1}^n \sqrt{E[A_i^4] + O_{w,b,G}\left(\frac{1}{k^4}\right)}\right) \\ &= O\left(\sum_{i=1}^n \sqrt{E[A_i^4]}\right) + O_{w,b,G}\left(\frac{1}{k^2}\right). \end{aligned}$$

We can now apply Lemma 16, which implies that

$$\begin{aligned} E||\hat{W}^r - w^r||_2^2 &\leq O\left(\frac{\mathrm{Tr}\left[(\mathrm{diag}(w^r)^{-1}L_V\mathrm{diag}(w^r)^{-1})^\dagger\right]}{k}\right) \\ &\quad + O_{w,b,G}\left(\frac{1}{k^2}\right) \\ &= O\left(\frac{\mathrm{Tr}\left[(\mathrm{diag}(w^r)^{-1}L_V\mathrm{diag}(w^r)^{-1})^\dagger\right]}{k}\right), \end{aligned}$$

where the last step holds for k large enough; and which further implies that, for k large enough,

$$\begin{aligned} E\left[\sin^2(\hat{W}, w)\right] &\leq E\frac{||W^r - w^r||_2^2}{||w^r||_2^2} \\ &\leq O\left(\frac{\mathrm{Tr}\left[(\mathrm{diag}(w^r)^{-1}L_V\mathrm{diag}(w^r)^{-1})^\dagger\right]}{k||w^r||_2^2}\right) \end{aligned} \quad (58)$$

where we used the following identity about the sine between two vectors:

$$|\sin(x,y)| = \inf_\alpha \frac{||\alpha x - y||_2}{||y||_2} \leq \frac{||x - y||_2}{||y||_2}.$$

Finally, the final expression on the right-hand side of Eq. (58) is unaffected by replacing w^r with w, since both numerator and denominator are scaled by the same scalar. Thus for large enough k,

$$E\left[\sin^2(\hat{W}, w)\right] \leq O\left(\frac{\mathrm{Tr}\left[(\mathrm{diag}(w)^{-1}L_V\mathrm{diag}(w)^{-1})^\dagger\right]}{k||w||_2^2}\right).$$

We complete the proof by observing that $L_\gamma = \mathrm{diag}(w)^{-1}L_v\mathrm{diag}(w)^{-1}$. □

6. Proof of Theorem 2

We will establish that, for large enough k,

$$E\left[\sin^2(w, \hat{w})\right] \geq \Omega\left(\frac{1}{k}\right)\frac{\mathrm{Tr}\left[L_\gamma^\dagger\right]}{||w||_2^2}, \quad (59)$$

for any estimator $\hat{w}$ built from the outcomes of pairwise comparisons. Our proof is a modification of the proof of the lower bound from (Hendrickx *et al.* , 2019), with departures at key steps. As discussed in the main body of the paper, the main departure is to specifically pick out the solution $\hat{W}^r$ in the lower bound analysis.

We first comment on the structure of our proof. We will begin by fixing a vector w and an estimator of the true weights $\hat{w}(\mathbf{Y})$. This estimator $\hat{w}(\mathbf{Y})$ is arbitrary and we may therefore intuitively think that the estimator $\hat{w}(\mathbf{Y})$ "knows" w. We then generate $w_z = w + \frac{1}{\sqrt{k}}R$, where R is a particular random vector whose distribution we'll specify below, and generate k comparisons across each edge according to weights w_z. Once again, because the estimator $\hat{w}(\mathbf{Y})$ is arbitrary, we can also think of it as "knowing" the distribution of w_z. Provided the number of comparisons k satisfies some lower bound depending on w and the graph, we will then prove that

$$E\left[\sin^2(w_z, \hat{w}(\mathbf{Y}))\right] \geq \Omega\left(\frac{\mathrm{Tr}(L_\gamma^\dagger)}{k||w||_2^2}\right). \quad (60)$$

Because the random vector R will be upper bounded with probability one by some function of w and G, this proves Theorem 2.

We now turn to the proof, i.e., to the construction of the random vector R which will allow us to prove Eq. (60). We will use $P_w(y)$ to denote the density on the observation space (consisting of k measurements across each edge of the graph) if w was the vector of true weights. We will use the following lemma [(Hajek & Raginsky, 2019) Chap. 13, Corollary 13.2] to obtain a lower bound on the expectation of the sine-squared:

Lemma 6.1. *Let μ be any joint probability distribution of a random pair (w, w'), such that the marginal distributions of both w and w' are equal to π. Then*

$$\mathbb{E}_{\pi,\mathbf{Y}}[d(w, \hat{w}(\mathbf{Y}))] \geq \mathbb{E}_\mu\left[d(w, w')(1 - \|P_w - P_{w'}\|_{TV}\right]$$

where $||\cdot||_{\mathrm{TV}}$ represents the total-variation distance between distributions and $\mathbf{Y}$ the observations.

It should be clear that under a random choice of w generated according to some distribution π (described later), the expected error is a lower bound on the worst-case estimation error over all possible w. Thus our goal is to massage the right-hand side of Lemma 6.1 to obtain the right-hand side of Eq. (59).

Actually, we need a slight modification of Lemma 6.1: as remarked in (Hendrickx *et al.* , 2019), it is sufficient that $d(w, w')$ satisfies a weak version of triangle inequality, i.e., $\alpha d(w_1, w_2) \leq d(w_1, \hat{w}) + d(w_2, \hat{w})$ for some pre-specified constant α, with the result that the right-hand side in the above lemma is multiplied by α. In particular, our (square) error criterion $\sin^2(\hat{w}, w_z)$ satisfies the weak triangle inequlity with a factor of $\alpha = 1/2$, see Lemma A.1 from (Hendrickx *et al.* , 2019), so we can apply Lemma 6.1 to it with an extra factor of $1/2$ on the right-hand side.

Let v_i be the eigenvectors of the $\mathrm{diag}(w)^{-1}L_V\mathrm{diag}(w)^{-1}$ with corresponding eigenvalues σ_i. In the next paragraph, we will use these eigenvectors to design the distribution for w which we will use to obtain our lower bound. Note that this is the first point where our argument diverges from the proof of (Hendrickx *et al.* , 2019); the introduction of this rescaling by $\mathrm{diag}(w)^{-1}$ here is motivated by Eq. (17) and Eq. (18), where the quantity $\mathrm{diag}(w)^{-1}L_V\mathrm{diag}(w)^{-1}$ appears, and comes from a desire to lower bound the error associated with the regularized solution $\hat{W}^r$.

Let $z_2, \ldots, z_n$ be i.i.d random variable taking values 1 and -1 with equal probability. We then set

$$w_z = w + \delta\sum_{i=2}^{n}\frac{z_i}{\sqrt{\sigma_i}}v_i \quad (61)$$

where, the sum starts at $i = 2$ to omit the eigenvector of $\mathrm{diag}(w)^{-1}L_V\mathrm{diag}(w)^{-1}$ associated with the zero eigenvalue (which is just w), δ is suitably small (to be specified later), and also we set $z_1 = 1$. We remark that we will later choose δ to be on the order of $1/\sqrt{k}$, so that the above expression can be written as $w_z = w + (1/\sqrt{k})R$, where the random vector R depends on w and the underlying graph, and further with probability one R cannot be larger than some function of w and the graph.

Let V be the unitary matrix which has v_i as columns; we can write

$$w_z = V\Lambda z,$$

where this relation defines the entries of Λ (e.g., $\lambda_i = \delta/\sqrt{\sigma_i}$ for $i = 1, \ldots, n$). We note that the norm of w_z's defined this way are equal, i.e.,

$$\begin{aligned}\|w_z\|_2 &= \sqrt{||w||_2^2 + \delta^2\sum_{i=2}^{n}\frac{1}{\sigma_i}} \\ &= \sqrt{||w||_2^2 + \delta^2\mathrm{Tr}\left[(\mathrm{diag}(w)^{-1}L_v\mathrm{diag}(w)^{-1})^\dagger\right]}. \quad (62)\end{aligned}$$

Intuitively the error in estimating w_z should be lower bounded in terms of the errors in estimating z_i, and indeed (Hendrickx *et al.* , 2019) showed that

$$\min_{\hat{w}(\mathbf{Y})}\mathbb{E}_{\pi,\mathbf{Y}}[\rho(w_z, \hat{w}(\mathbf{Y}))] = \sum_{i=2}^{n}\min_{\eta_i(\mathbf{Y})}\frac{\lambda_i^2}{\|w_z\|^2}\mathbb{E}_{\pi,\mathbf{Y}}\left(z_i - \eta_i(\mathbf{Y})\right)^2,$$

where $\mathbf{Y}$ is the vector of outcomes of the comparisons. We are now going to apply Lemma 6.1 to each term on the right hand side individually. Following (Hendrickx *et al.* , 2019), we define the distribution $\mu_i(z, z')$ by keeping z uniformly distributed in $\{-1, 1\}^n$, and flipping the i^{th} bit to obtain z' (formally, $z_i' = -z_i$ and $z_j' = z_j$ for every $j \neq i$). Clearly, $\mathbb{E}_{\pi,\mathbf{Y}}d_i(z, z') = 4$. Moreover, by Pinsker's inequality

$$\begin{aligned}\|P_w^{\otimes k} - P_{w'}^{\otimes k}\|_{\mathrm{TV}}^2 &\leq \frac{1}{2}D_{KL}(P_w^{\otimes k}\|P_w'^{\otimes k}) \quad (63) \\ &\leq O(k\delta^2)\end{aligned}$$

where the proof of the second inequality (which holds for small enough δ) is somewhat involved and is relegated to

Section 6.1 below. Using these facts along with Lemma 6.1, it follows that for every estimator $\eta_i(\mathbf{Y})$ and for such δ,

$$\mathbb{E}_{\pi,\mathbf{Y}}\left(z_i - \eta_i(\mathbf{Y})\right)^2 \geq \frac{1}{2}4\left(1 - \sqrt{O(k\delta^2)}\right),$$

and thus

$$\begin{aligned}\min_{\hat{w}(\mathbf{Y})} \mathbb{E}_{\pi,\mathbf{Y}}[\rho(w, \hat{w}(\mathbf{Y}))] &\geq \sum_{i=2}^{n} \frac{\lambda_i^2}{\|w_z\|^2} 2(1 - \sqrt{O(k\delta^2)}) \\ &\geq \sum_{i=2}^{n} \frac{2\delta^2(1 - \sqrt{O(k\delta^2)})}{\sigma_i S}, \quad (64)\end{aligned}$$

where

$$S = ||w||_2^2 + \delta^2 \mathrm{Tr}\left[\left(\mathrm{diag}(w)^{-1} L_v \mathrm{diag}(w)^{-1}\right)^{\dagger}\right]$$

Now choosing δ such that the $O(k\delta^2)$ term is at most $1/2$ (which involves choosing $\delta^2 = \Theta(1/k)$), and further k is large enough so that

$$||w||^2 + \delta^2 \mathrm{Tr}\left[\left(\mathrm{diag}(w)^{-1} L_V \mathrm{diag}(w)^{-1}\right)^{\dagger}\right] < 2||w||_2^2, \quad (65)$$

Putting all these bounds into Eq. (64) yields

$$\min_{\hat{w}(\mathbf{Y})} \mathbb{E}_{\pi,\mathbf{Y}}[\rho(w, \hat{w}(\mathbf{Y}))] \geq \Omega\left(\frac{1}{k}\right) \frac{\sum_{i=2}^{n} 1/\sigma_i}{||w||_2^2}.$$

Using the definition of σ_i, we can rewrite this as

$$\min_{\hat{w}(\mathbf{Y})} \mathbb{E}_{\pi,\mathbf{Y}}[\rho(w, \hat{w}(\mathbf{Y}))]$$

$$\geq \Omega\left(\frac{1}{k}\right) \frac{\mathrm{Tr}\left[\left(\mathrm{diag}(w)^{-1} L_V \mathrm{diag}(w)^{-1}\right)^{\dagger}\right]}{||w||_2^2}.$$

Finally noting that

$$\mathrm{diag}(w)^{-1} L_V \mathrm{diag}(w)^{-1} = L_\gamma,$$

we have thus proved Eq. (59) (conditionally on Eq. (63) which is considered in the next section).

6.1. Proof of Equation (63)

We will need to begin with several lemmas. Let $B(w_1, w_2)$ be our notation for a Bernoulli that falls on heads with probability of $w_1/(w_1 + w_2)$. We will need some bounds on how the KL-divergence evolves as we perturb w_1, w_2.

Lemma 18. *Fix positive w_1, w_2. For small enough δ,*

$$D_{KL}(B(w_1(1+\delta x_1), w_2(1+\delta x_2))||B(w_1(1-\delta x_1), w_2(1-\delta x_2)))$$
$$\leq \frac{5\delta^2}{V_{12}}(x_1 - x_2)^2$$

where we use the notation $V_{12} = \frac{w_1}{w_2} + 2 + \frac{w_2}{w_1}$ (consistent with our previous usage).

Proof. Let us introduce notations for the probabilities associated with the two Bernoulli distribution, the first corresponding to $z = 1$ and the second to $z = -1$:

$$\begin{aligned} p &= \frac{w_1(1+\delta x_1)}{w_1(1+\delta x_1) + w_2(1+\delta x_2)} \\ p' &= \frac{w_1(1-\delta x_1)}{w_1(1-\delta x_1) + w_2(1-\delta x_2)} \end{aligned}$$

Applying Lemma 7.2 of (Bubeck, 2011), we have the estimate

$$D_{KL}(p||p') \leq \frac{(p-p')^2}{p'(1-p')}$$

We consider what this is like in the limit as $\delta \to 0$, as this leads to a number of simplifications. First, consider the denominator: we have that

$$\lim_{\delta \to 0} \frac{1}{p'(1-p')} = \frac{1}{(w_1/(w_1+w_2))(w_2/(w_1+w_2))} = V_{12}.$$

We conclude that, for δ small enough,

$$D_{KL}(p||p') \leq \left(\frac{5}{4}\right)^{1/3} V_{12}(p-p')^2.$$

Of course, the constant in front of the right-hand side can be chosen to be any number greater than one.

Next, let us consider the difference of the two probabilities:

$$\begin{aligned} p - p' &= \frac{w_1}{w_1 + w_2\frac{1+\delta x_2}{1+\delta x_1}} - \frac{w_1}{w_1 + w_2\frac{1-\delta x_2}{1-\delta x_1}} \\ &= \frac{w_1 w_2\left(\frac{1-\delta x_2}{1-\delta x_1} - \frac{1+\delta x_2}{1+\delta x_1}\right)}{(w_1 + w_2\frac{1+\delta x_2}{1+\delta x_1})(w_1 + w_2\frac{1-\delta x_2}{1-\delta x_1})} \\ &:= C(w_1, w_2, \delta)\left(\frac{1-\delta x_2}{1-\delta x_1} - \frac{1+\delta x_2}{1+\delta x_1}\right) \end{aligned}$$

Observing that

$$\lim_{\delta \to 0} C(w_1, w_2, \delta) = \frac{1}{V_{12}},$$

we can conclude that, for δ small enough,

$$\begin{aligned} D_{KL}(p||p') &\leq \left(\frac{5}{4}\right)^{1/3} V_{12} \left(\frac{5}{4}\right)^{1/3} \frac{1}{V_{12}^2}\left(\frac{1-\delta x_2}{1-\delta x_1} - \frac{1+\delta x_2}{1+\delta x_1}\right)^2 \\ &= \frac{(5/4)^{2/3}}{V_{12}}\left(\frac{1-\delta x_2}{1-\delta x_1} - \frac{1+\delta x_2}{1+\delta x_1}\right)^2 \end{aligned}$$

Finally, observe that the function

$$f(t) = \frac{1+tx_2}{1+tx_1} = \frac{1+tx_1+t(x_2-x_1)}{1+tx_1} = 1 + t\frac{x_2-x_1}{1+tx_1}$$

clearly satisfies $f'(0) = x_2 - x_1$. Consequently, for small enough δ, we have that

$$\left(\frac{1-\delta x_2}{1-\delta x_1} - \frac{1+\delta x_2}{1+\delta x_1}\right)^2 \leq \left(\frac{5}{4}\right)^{1/3}((x_2-x_1)2\delta)^2.$$

Putting it all together, we have that

$$D_{KL}(p||p') \leq \frac{5}{V_{12}}\delta^2(x_2 - x_1)^2.$$

□

Corollary 19. *Fix u_1, u_2 positive and arbitrary a_1, a_2, x_1, x_2. Consider the same situation as Lemma 18 except that the weights of node $j = 1, 2$ are*

$$w_j = u_j + \delta a_j + \delta z x_j,$$

where z is either $+1$ or -1. The KL divergence between the corresponding Bernoulli random variables is upper bounded by

$$\frac{5\delta^2}{V_{12}}\left(\frac{x_1}{u_1} - \frac{x_2}{u_2}\right)^2 + O(\delta^3),$$

with $V_{12} = \frac{u_1}{u_2} + 2 + \frac{u_2}{u_1}$

Proof. The weight of node i can be rewritten as

$$u_j + \delta a_j + \delta z x_j, = u_j\left(1 + \delta\frac{a_j}{u_j}\right)\left(1 + z\delta\frac{x_j/u_j}{1 + \delta a_j/u_j}\right).$$

We can then apply Lemma 18 with the modified (bounded) parameters $\tilde{w}_j = u_j(1 + \delta a_j/u_j)$ and $\tilde{x}_j = \frac{x_j/u_j}{1+\delta a_j/u_j}$, and we obtain that the KL divergence is (with $V_{w,12}$ the variance for the weights $\tilde{w}_j$)

$$\begin{aligned} D_{KL} &= \frac{5\delta^2}{V_{w,12}}(\tilde{x}_1 - \tilde{x}_2)^2 \\ &= \frac{5\delta^2}{V_{w,12}}\left(\frac{x_1/u_1}{1 + \delta a_1/u_1} - \frac{x_2/u_2}{1 + \delta a_2/u_2}\right)^2 \\ &= \frac{5\delta^2}{V_{w,12}}\left(\frac{x_1}{u_1} - \frac{x_2}{u_2} + O(\delta)\right)^2 \\ &= \frac{5\delta^2}{V_{w,12}}\left(\frac{x_1}{u_1} - \frac{x_2}{u_2}\right)^2 + O(\delta^3). \end{aligned} \quad (66)$$

Besides, observe that

$$\begin{aligned} \frac{\tilde{w}_1}{\tilde{w}_2} &= \frac{u_1}{u_2} \cdot \frac{1 + \delta a_1/u_1}{1 + \delta a_2/u_2} \\ &= \frac{u_1}{u_2} + O(\delta). \end{aligned}$$

Hence

$$V_{w,12} = \frac{\tilde{w}_1}{\tilde{w}_2} + 2 + \frac{\tilde{w}_1}{\tilde{w}_2} = V_{u,12} + O(\delta).$$

The result follows then from (66). □

With these facts in place, we can now prove the equation to which this subsection is dedicated.

Proof of Equation (63). We can apply Corollary 19 across each edge, with $u_j = w_j$, $x_j = v_i$, and $a_j = \sum_{j \neq i} \lambda_j v_j$ to argue as follows: $D_{KL}(P_w^{\otimes k} \| P_w'^{\otimes k})$

$$\begin{aligned} &= kD_{KL}(P_w \| P'_w) \\ &\leq \sum_{(a,b)\in E} O\left(\frac{k\delta^2}{V_{ab}}\right)\left(\frac{(v_i)_a}{\sqrt{\sigma_i}w_a} - \frac{(v_i)_b}{\sqrt{\sigma_i}w_b}\right)^2 + O(k\delta^3) \\ &= O(k\delta^2)\frac{1}{\sigma_i}v^T \operatorname{diag}(w)^{-1} L_V \operatorname{diag}(w)^{-1} v + O(k\delta^3) \\ &= O(k\delta^2), \end{aligned}$$

where we used that v_i is an eigenvector of $\operatorname{diag}(w)^{-1} L_v \operatorname{diag}(w)^{-1}$ with eigenvalue σ_i. □