Supplementary Material

In this section, we supply the theoretical justifications that are missing from the main text.

§1. *The regularization map is well defined for $\tau \in (0,1)$:* The regularization problem can be written as

$$\underset{f\in\mathbb{R}^N}{\text{minimize}}\ \|Af-b\|_2^2, \quad A := \begin{bmatrix}\sqrt{1-\tau}L^{1/2}\\ \sqrt{\tau}\Lambda\end{bmatrix}, \quad b := \begin{bmatrix}0\\ \sqrt{\tau}y\end{bmatrix}.$$

Its solutions $\bar{f}$ are characterized by the normal equation $A^*A\bar{f} = A^*b$, i.e., by $\big((1-\tau)L + \tau\Lambda^*\Lambda\big)\bar{f} = \tau\Lambda^* y$. Note that we always make the assumption $\ker(L)\cap\ker(\Lambda) = \{0\}$, otherwise, fixing $f_0 \in \mathcal{K}$ and $e_0 \in \mathcal{E}$ with $\Lambda f_0 + e_0 = y$, the existence of $h \in \mathbb{R}^N\setminus\{0\}$ such that $Lh = 0$ and $\Lambda h = 0$ implies that $f_t := f_0 + th \in \mathcal{K}$, $e_t := e_0 \in \mathcal{E}$, and $\Lambda f_t + e_t = y$ for all $t\in\mathbb{R}$, yielding an infinite local (in turn global) worst-case error for the recovery of $Q = I_N$. This assumption ensures that $(1-\tau)L + \tau\Lambda^*\Lambda$ is positive definite—hence invertible—for any $\tau\in(0,1)$, since

$$\big\langle\big((1-\tau)L + \tau\Lambda^*\Lambda\big)h, h\big\rangle = (1-\tau)\|L^{1/2}h\|_2^2 + \tau\|\Lambda h\|_2^2 \geq 0$$

for all $h\in\mathbb{R}^N$, with equality possible when and only when $h \in \ker(L)\cap\ker(\Lambda) = \{0\}$. This shows that $\bar{f}$ is unique and given by $\bar{f} = \big((1-\tau)L + \tau\Lambda^*\Lambda\big)^{-1}(\tau\Lambda^* y)$. Finally, if the graph G is made of K connected components $C_1,\ldots,C_K$, we observe that

$$\ker(L)\cap\ker(\Lambda) = \left\{h = \sum_{k=1}^K a_k \mathbf{1}_{C_k},\ a\in\mathbb{R}^K,\ h_{|V_\ell} = 0\right\}$$
$$= \left\{\sum_{k=1}^K a_k\mathbf{1}_{C_k},\ a\in\mathbb{R}^K,\ a_k = 0 \text{ when } C_k\cap V_\ell \neq \emptyset\right\},$$

so that $\ker(L)\cap\ker(\Lambda) = \{0\}$ if and only if $C_k \cap V_\ell \neq \emptyset$ for all $k = 1,\ldots,K$, which means that at least one vertex is observed in each connected component.

§2. *The limiting case $\tau\to 0$:* Writing $f_\tau = \Delta_\tau(y)$, if we divide the objective function that f_τ minimizes by $\tau > 0$, we see that

$$f_\tau = \underset{f\in\mathbb{R}^N}{\text{argmin}}\ \frac{1-\tau}{\tau}\|L^{1/2}f\|_2^2 + \|\Lambda f_\tau - y\|_2^2.$$

Intuitively, the limit f_0 of f_τ as $\tau\to 0$ should satisfy $L^{1/2}f_0 = 0$, otherwise $\|L^{1/2}f_\tau\|_2^2 \geq \kappa$ for some $\kappa > 0$ when τ is sufficiently small, and then $\big((1-\tau)/\tau\big)\|L^{1/2}f_\tau\|_2^2$ blows up as $\tau\to 0$, preventing f_τ to be a minimizer of the divided objective function. It suggests—and this can be made precise—that

$$f_0 = \underset{f\in\mathbb{R}^N}{\text{argmin}}\ \|\Lambda f - y\|_2^2 \quad \text{s.to} \quad L^{1/2}f = 0.$$

The constraint $L^{1/2}f = 0$ is equivalent to f taking the form $f = \sum_{k=1}^K a_k\mathbf{1}_{C_k}$ for some $a\in\mathbb{R}^K$, where $C_1,\ldots,C_K$ denote the connected components of the graph G. Under this constraint, we then have

$$\Lambda f - y = f_{|V_\ell} - y = \sum_{k=1}^K \big(a_k\mathbf{1}_{C_k\cap V_\ell} - y_{C_k}\big),$$

and hence, since the summands have disjoint supports,

$$\|\Lambda f - y\|_2^2 = \sum_{k=1}^K \|a_k\mathbf{1}_{C_k\cap V_\ell} - y_{C_k}\|_2^2$$
$$= \sum_{k=1}^K\big(a_k^2|C_k\cap V_\ell| - 2a_k\langle\mathbf{1}_{C_k\cap V_\ell}, y_{C_k}\rangle + \|y_{C_k}\|_2^2\big).$$

This quantity is easily seen to attain its minimal value when $a_k = \langle\mathbf{1}_{C_k\cap V_\ell}, y_{C_k}\rangle/|C_k\cap V_\ell|$ for each $k = 1,\ldots,K$. All in all, this signifies that the component of f_0 on each C_k is equal to the average of y_{C_k}—as announced, the case $\tau\to 0$ is not very interesting!

§3. *Proof of Lemma 2:* To prove the inequality, consider $h\in\mathbb{R}^N$ with $\|L^{1/2}h\|_2^2 \leq \varepsilon^2$ and $\|\Lambda h\|_2^2 \leq \eta$. Then define $f_\pm = \pm h \in \mathcal{K}$ and $e_\pm = \mp\Lambda h \in\mathcal{E}$, so that

$$\begin{aligned}
\text{gwce}_Q(\Delta) &\geq \max_\pm \|Q(f_\pm) - \Delta(\Lambda f_\pm + e_\pm)\|_2\\
&= \max_\pm \|Q(\pm h) - \Delta(0)\|_2\\
&\geq \frac{1}{2}\|Q(h) - \Delta(0)\|_2 + \frac{1}{2}\|Q(-h) - \Delta(0)\|_2\\
&\geq \frac{1}{2}\|(Q(h) - \Delta(0)) - (Q(-h) - \Delta(0))\|_2\\
&= \frac{1}{2}\|2Q(h)\|_2 = \|Q(h)\|_2.
\end{aligned}$$

In remains to take the supremum over admissible h to obtain the announced lower bound.

Next, the transformation of the lower bound for the global worst-case error relies on a case of validity of the S-procedure due to Polyak, see Polyak [1998]. We start by writing this (squared) lower bound as

$$\inf_{\gamma\in\mathbb{R}}\ \gamma \quad \text{s.to } \|Q(h)\|_2^2 \leq \gamma \text{ whenever } \|L^{1/2}h\|_2^2 \leq \varepsilon^2,\ \|\Lambda h\|_2^2 \leq \eta^2.$$

Using the S-procedure of Polyak, the above constraint is equivalent to the existence of $c, d \geq 0$ such that, for all $h\in\mathbb{R}^N$,

$$\|Q(h)\|_2^2 - \gamma \leq c(\|L^{1/2}h\|_2^2 - \varepsilon^2) + d(\|\Lambda h\|_2^2 - \eta^2), \quad (5)$$

under the proviso that there exist $\widetilde{h}\in\mathbb{R}^N$ and $\alpha,\beta\in\mathbb{R}$ such that $\|L^{1/2}\widetilde{h}\|_2^2 < \varepsilon^2$, $\|\Lambda\widetilde{h}\|_2^2 < \eta^2$, and $\alpha L + \beta\Lambda^*\Lambda \succ 0$. This proviso is met by taking $\widetilde{h} = 0$ and $(\alpha,\beta) = (1-\tau,\tau)$ for any $\tau\in(0,1)$, see §1 above. Now, the constraint (5) can be written as

$$\langle(cL + d\Lambda^*\Lambda - Q^*Q)h, h\rangle + \gamma - c\varepsilon^2 - d\eta^2 \geq 0$$

for all $h\in\mathbb{R}^N$, which in fact decouples as the two constraints $cL + d\Lambda^*\Lambda - Q^*Q \succeq 0$ and $\gamma - c\varepsilon^2 - d\eta^2 \geq 0$. Under the latter constraint, the minimal value of γ is $c\varepsilon^2 + d\eta^2$ and we arrive at the desired expression. □

§4. *Proof of Lemma 3:* The two additional lemmas below are needed.

Lemma 5. If A, B, C are square matrices of similar size and if $C \succeq 0$, then

$$\left[\begin{array}{c|c} A & 0 \\ \hline 0 & B \end{array}\right] \succeq \left[\begin{array}{c|c} A-C & C \\ \hline C & B-C \end{array}\right].$$

Proof. To prove that the difference of these two matrices is positive semidefinite, we simply write, for any vectors x, y,

$$\begin{aligned}&\left\langle \left[\begin{array}{c|c} C & -C \\ \hline -C & C \end{array}\right] \left[\begin{array}{c} x \\ \hline y \end{array}\right], \left[\begin{array}{c} x \\ \hline y \end{array}\right] \right\rangle \\ &= \langle Cx, x\rangle - \langle Cy, x\rangle - \langle Cx, y\rangle + \langle Cy, y\rangle \\ &= \langle C^{1/2}x, C^{1/2}x\rangle - 2\langle C^{1/2}x, C^{1/2}y\rangle + \langle C^{1/2}y, C^{1/2}y\rangle \\ &= \|C^{1/2}x - C^{1/2}y\|_2^2,\end{aligned}$$

which is obviously nonnegative. □

Lemma 6. If A and B are positive semidefinite matrices of similar size such that $A + B \succ 0$, then $C := A(A+B)^{-1}B$ is positive semidefinite.

Proof. Writing C as $C = A(A+B)^{-1}(A+B-A)$, i.e., $C = A - A(A+B)^{-1}A$, shows that C is self-adjoint and reveals that we in fact have to prove that $A(A+B)^{-1}A \preceq A$. To see why this is so, we start from $A \preceq A + B$, so that $M := (A+B)^{-1/2}A(A+B)^{-1/2}$ satisfies $M \preceq I$. This implies that $M^2 \preceq M$, which reads

$$(A+B)^{-1/2}A(A+B)^{-1}A(A+B)^{-1/2} \preceq (A+B)^{-1/2}A(A+B)^{-1/2}.$$

Multiplying on the left and on the right by $(A+B)^{1/2}$ yields the desired result. □

Focusing now on the proof of Lemma 3, let us consider $c, d \geq 0$ such that

$$cL + d\Lambda^*\Lambda \succeq Q^*Q \tag{6}$$

and let us set $\tau = d/(c+d)$. From (3), we notice that

$$\begin{aligned}\Delta_\tau\Lambda &= (cL + d\Lambda^*\Lambda)^{-1}d\Lambda^*\Lambda, \\ I - \Delta_\tau\Lambda &= (cL + d\Lambda^*\Lambda)^{-1}cL.\end{aligned}$$

Multiplying (6) on the right by $\left[I - \Delta_\tau\Lambda \,\middle|\, \Delta_\tau\Lambda\right]$, which equals $(cL + d\Lambda^*\Lambda)^{-1}\left[cL \,\middle|\, d\Lambda^*\Lambda\right]$, and on the left by its adjoint, we arrive at

$$\begin{aligned}&\left[\begin{array}{c} cL \\ \hline d\Lambda^*\Lambda \end{array}\right](cL + d\Lambda^*\Lambda)^{-1}\left[cL \,\middle|\, d\Lambda^*\Lambda\right] \\ &\qquad \succeq \left[\begin{array}{c} (I - \Delta_\tau\Lambda)^* \\ \hline (\Delta_\tau\Lambda)^* \end{array}\right] Q^*Q \left[I - \Delta_\tau\Lambda \,\middle|\, \Delta_\tau\Lambda\right].\end{aligned} \tag{7}$$

First, we claim that the left-hand side of (7) takes the form $\left[\begin{array}{c|c} A-C & C \\ \hline C & B-C \end{array}\right]$ with $A = cL$ and $B = d\Lambda^*\Lambda$. To see this, it suffices to observe, e.g., that $A = cL$ is indeed the sum of its upper two blocks, which is clear since these blocks are $cL(cL + d\Lambda^*\Lambda)^{-1}cL$ and $cL(cL + d\Lambda^*\Lambda)^{-1}d\Lambda\Lambda^*$. Second, we claim that C can be written as $C = A(A+B)^{-1}B$, which is also clear—the relation $C = cL(cL + d\Lambda^*\Lambda)^{-1}d\Lambda\Lambda^*$ was just pointed out. Therefore, according to our two additional lemmas, the left-hand side of (7) does not exceed, in the positive semidefinite sense, $\left[\begin{array}{c|c} cL & 0 \\ \hline 0 & d\Lambda\Lambda^* \end{array}\right]$. At this point, we have shown that

$$\left[\begin{array}{c} (I - \Delta_\tau\Lambda)^* \\ \hline (\Delta_\tau\Lambda)^* \end{array}\right] Q^*Q \left[I - \Delta_\tau\Lambda \,\middle|\, \Delta_\tau\Lambda\right] \preceq \left[\begin{array}{c|c} cL & 0 \\ \hline 0 & d\Lambda^*\Lambda \end{array}\right],$$

which is equivalent to

$$\|Q(I - \Delta_\tau\Lambda)f + Q(\Delta_\tau\Lambda)g\|_2^2 \leq c\|L^{1/2}f\|_2^2 + d\|\Lambda g\|_2^2$$

for all $f, g \in \mathbb{R}^N$. The observation map Λ is obviously surjective in the present situation[2], so that any $e \in \mathbb{R}^{n_\ell}$ can be written as $e = \Lambda g$ for some $g \in \mathbb{R}^N$. From here, the desired result follows. □

§5. *Near optimality under mild overestimation of ε and η:* According to Theorem 1 (and its proof) and using the same notation, we have

$$\inf_{\Delta:\mathbb{R}^{n_\ell}\to\mathbb{R}^n} \mathrm{gwce}_Q(\Delta)^2 = c_\flat\varepsilon^2 + d_\flat\eta^2$$

while $c_\flat L + d_\flat\Lambda^*\Lambda \succeq Q^*Q$. Now suppose that ε and η are not exactly known but overestimated by $\bar\varepsilon$ and $\bar\eta$. Solving the semidefinite program (4) with $\bar\varepsilon$ and $\bar\eta$ provides a parameter $\bar\tau = \bar d/(\bar c + \bar d)$ such that

$$\begin{aligned}&\sup_{\substack{\|L^{1/2}f\|_2\leq\bar\varepsilon \\ \|e\|_2\leq\bar\eta}} \|Q(f) - Q\circ\Delta_{\bar\tau}(\Lambda f + e)\|_2^2 \\ &\qquad = \min\left\{c\bar\varepsilon^2 + d\bar\eta^2 : \; cL + d\Lambda^*\Lambda \succeq Q^*Q\right\}.\end{aligned}$$

Since $\bar\varepsilon \geq \varepsilon$ and $\bar\eta \geq \eta$, we deduce in particular that

$$\sup_{\substack{\|L^{1/2}f\|_2\leq\varepsilon \\ \|e\|_2\leq\eta}} \|Q(f) - Q\circ\Delta_{\bar\tau}(\Lambda f + e)\|_2^2 \leq c_\flat\bar\varepsilon^2 + d_\flat\bar\eta^2.$$

Under the mild overestimations $\bar\varepsilon \leq C\varepsilon$ and $\bar\eta \leq C\eta$, this implies that

$$\begin{aligned}\mathrm{gwce}_Q(Q\circ\Delta_{\bar\tau})^2 &\leq C^2\left[c_\flat\varepsilon^2 + d_\flat\eta^2\right] \\ &= C^2 \inf_{\Delta:\mathbb{R}^{n_\ell}\to\mathbb{R}^n} \mathrm{gwce}_Q(\Delta)^2,\end{aligned}$$

proving that the recovery map $Q\circ\Delta_{\bar\tau}$ is globally near optimal.

§6. *No SDPs to optimal estimate a linear functional:* If $Q = \langle q, \cdot\rangle : \mathbb{R}^N \to \mathbb{R}$ is a linear functional, then solving the semidefinite program (4) and composing the resulting regularization map $\Delta_{\tau_\flat}$ with Q to obtain a globally optimal recovery map is quite wasteful. In such a situation, a globally optimal recovery map can be more directly obtained as $\langle a_\flat, \cdot\rangle$, where $a_\flat \in \mathbb{R}^{n_\ell}$ is a solution to

$$\operatorname*{minimize}_{a\in\mathbb{R}^{n_\ell}} \left[\sup_{\|L^{1/2}f\|_2\leq\varepsilon} |\langle q - \Lambda^*a, f\rangle| \times \varepsilon + \|a\|_2 \times \eta\right]. \tag{8}$$

[2]In general, it is always assumed that $\Lambda : \mathbb{R}^N \to \mathbb{R}^m$ is surjective, as it does not make sense to collect an observation that can be deduced from the others.

This laborious-looking optimization program can be turned into a more manageable one. For instance, if the graph G has connected components $C_1, \ldots, C_K$, then the eigenvalues of L are $0 = \lambda_1 = \cdots = \lambda_K < \lambda_{K+1} \leq \cdots \leq \lambda_N$. Denoting by $(v_1, \ldots, v_N)$ an orthonormal basis associated with these eigenvalues (so that $v_k = \mathbf{1}_{C_k}/\sqrt{|C_k|}$, $k = 1, \ldots, K$), the problem (8) reduces to

$$\underset{a \in \mathbb{R}^{n_\ell}}{\text{minimize}} \left[\sum_{k=K+1}^{N} \frac{\langle q - \Lambda^* a, v_k \rangle^2}{\lambda_k} \right]^{1/2} \times \varepsilon + \|a\|_2 \times \eta$$

$$\text{s.to } \langle q - \Lambda^* a, v_k \rangle = 0, \quad k = 1, \ldots, K.$$

Note that the vector $\Lambda^* a \in \mathbb{R}^N$ appearing above is just the vector $a \in \mathbb{R}^{n_\ell}$ padded with zeros on the unlabeled vertices.

§7. *A graph whose Laplacian is a scaled orthogonal projector:* Suppose that G is an unweighted graph (so that $w_{i,j} \in \{0,1\}$ for all $i,j = 1, \ldots, N$) made of connected components $C_1, \ldots, C_K$ which are all complete graphs on an equal number n of vertices. The adjacency matrix W_k and graph Laplacian L_k of each C_k are

$$W_k = \begin{bmatrix} 0 & 1 & 1 & \cdots & 1 \\ 1 & 0 & 1 & \cdots & 1 \\ 1 & 1 & 0 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{bmatrix},$$

$$L_k = \begin{bmatrix} n-1 & -1 & -1 & \cdots & -1 \\ -1 & n-1 & -1 & \cdots & -1 \\ -1 & -1 & n-1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & -1 \\ -1 & -1 & \cdots & -1 & n-1 \end{bmatrix}.$$

Note that L_k has eigenvalue 0 of multiplicity 1 and eigenvalue n of multiplicity $n-1$. Therefore, the whole graph Laplacian

$$L = \left[\begin{array}{c|c|c|c} L_1 & 0 & \cdots & 0 \\ \hline 0 & L_2 & \ddots & \vdots \\ \hline \vdots & \ddots & \ddots & 0 \\ \hline 0 & \cdots & 0 & L_K \end{array}\right]$$

has eigenvalue 0 of multiplicity K and eigenvalue n of multiplicity $(n-1)K$. This means that the renormalized Laplacian $(1/n)L$ is an orthogonal projector.

§8. *Proof of Theorem 4:* The argument is divided into three parts, namely:

a) there is a parameter $\tau_\natural \in (0,1)$ yielding

$$\|L^{1/2}\Delta_{\tau_\natural}(y)\|_2 = \frac{\varepsilon}{\eta}\|\Lambda\Delta_{\tau_\natural}(y) - y\|_2, \tag{9}$$

and the corresponding reguralizer $\Delta_{\tau_\natural}(y)$ is a solution to

$$\underset{f \in \mathbb{R}^N}{\text{minimize}} \max\left\{\|L^{1/2}f\|_2^2, \frac{\varepsilon^2}{\eta^2}\|\Lambda f - y\|_2^2\right\}; \tag{10}$$

b) the optimization program (10) admits a unique solution $f_\natural$ (hence equal to $\Delta_{\tau_\natural}(y)$);

c) the solution $f_\natural$ to (10) does provide a near optimal local worst-case error.

Justification of **a)**. For any $\tau \in [0,1]$, let f_τ denote $\Delta_\tau(y)$. Recalling that f_0 and f_1 are interpreted as

$$f_0 = \underset{f \in \mathbb{R}^N}{\text{argmin}} \|\Lambda f - y\|_2 \quad \text{s.to } L^{1/2}f = 0,$$

$$f_1 = \underset{f \in \mathbb{R}^N}{\text{argmin}} \|L^{1/2}f\|_2 \quad \text{s.to } \Lambda f = y.$$

we have

$$\|L^{1/2}f_0\|_2 - \frac{\varepsilon}{\eta}\|\Lambda f_0 - y\|_2 = -\frac{\varepsilon}{\eta}\|\Lambda f_0 - y\|_2 < 0,$$

$$\|L^{1/2}f_1\|_2 - \frac{\varepsilon}{\eta}\|\Lambda f_1 - y\|_2 = \|L^{1/2}f_1\|_2 > 0.$$

The continuity of $\tau \mapsto f_\tau = ((1-\tau)L + \tau\Lambda^*\Lambda)^{-1}(\tau\Lambda^* y)$ guarantees that there exists some $\tau_\natural \in (0,1)$ satisfying $\|L^{1/2}f_{\tau_\natural}\|_2 - (\varepsilon/\eta)\|\Lambda f_{\tau_\natural} - y\|_2 = 0$, as announced in (9). We additionally point out that this $\tau_\natural$ is unique, which is a consequence of the facts that $\tau \mapsto \|L^{1/2}f_\tau\|_2$ is strictly increasing and that $\tau \mapsto \|\Lambda f_\tau - y\|_2$ is strictly decreasing. To see the former, say, recall that f_τ is the unique minimizer of $((1-\tau)/\tau)\|L^{1/2}f\|_2^2 + \|\Lambda f - y\|_2^2$. Therefore, given $\sigma < \tau$,

$$\begin{aligned} \left(\frac{1}{\sigma} - 1\right)&\|L^{1/2}f_\sigma\|_2^2 + \|\Lambda f_\sigma - y\|_2^2 \\ &< \left(\frac{1}{\sigma} - 1\right)\|L^{1/2}f_\tau\|_2^2 + \|\Lambda f_\tau - y\|_2^2 \\ &= \left(\frac{1}{\tau} - 1\right)\|L^{1/2}f_\tau\|_2^2 + \|\Lambda f_\tau - y\|_2^2 \\ &\quad + \left(\frac{1}{\sigma} - \frac{1}{\tau}\right)\|L^{1/2}f_\tau\|_2^2 \\ &< \left(\frac{1}{\tau} - 1\right)\|L^{1/2}f_\sigma\|_2^2 + \|\Lambda f_\sigma - y\|_2^2 \\ &\quad + \left(\frac{1}{\sigma} - \frac{1}{\tau}\right)\|L^{1/2}f_\tau\|_2^2. \end{aligned}$$

Rearranging this inequality reads

$$\left(\frac{1}{\sigma} - \frac{1}{\tau}\right)\|L^{1/2}f_\sigma\|_2^2 < \left(\frac{1}{\sigma} - \frac{1}{\tau}\right)\|L^{1/2}f_\tau\|_2^2,$$

i.e., $\|L^{1/2}f_\sigma\|_2 < \|L^{1/2}f_\tau\|_2$, as expected. To finish, we now need to show that $f_{\tau_\natural}$ is a solution to (10). To this end, we remark on the one hand that the objective function of (10) evaluated at $f_{\tau_\natural}$ is

$$\max\left\{\|L^{1/2}f_{\tau_\natural}\|_2^2, \frac{\varepsilon^2}{\eta^2}\|\Lambda f_{\tau_\natural} - y\|_2^2\right\} = \gamma^2,$$

where γ is the common value of both terms in (9). On the other hand,

$$\text{setting} \qquad \tau'_\natural = \frac{(\eta^2/\varepsilon^2)\tau_\natural}{1 - \tau_\natural + (\eta^2/\varepsilon^2)\tau_\natural} \in [0,1],$$

$$\text{so that} \qquad 1 - \tau'_\natural = \frac{1 - \tau_\natural}{1 - \tau_\natural + (\eta^2/\varepsilon^2)\tau_\natural} \in [0,1],$$

the objective function of (10) evaluated at any $f \in \mathbb{R}^N$ satisfies

$$
\begin{aligned}
&\max\left\{\|L^{1/2}f\|_2^2, \frac{\varepsilon^2}{\eta^2}\|\Lambda f - y\|_2^2\right\}\\
&\geq (1-\tau_\natural')\|L^{1/2}f\|_2^2 + \tau_\natural' \frac{\varepsilon^2}{\eta^2}\|\Lambda f - y\|_2^2\\
&= \frac{1}{1-\tau_\natural + (\eta^2/\varepsilon^2)\tau_\natural}\left((1-\tau_\natural)\|L^{1/2}f\|_2^2 + \tau_\natural\|\Lambda f - y\|_2^2\right)\\
&\geq \frac{1}{1-\tau_\natural + (\eta^2/\varepsilon^2)\tau_\natural}\left((1-\tau_\natural)\|L^{1/2}f_{\tau_\natural}\|_2^2 + \tau_\natural\|\Lambda f_{\tau_\natural} - y\|_2^2\right)\\
&= \frac{1}{1-\tau_\natural + (\eta^2/\varepsilon^2)\tau_\natural}\left((1-\tau_\natural)\gamma^2 + \tau_\natural(\eta^2/\varepsilon^2)\gamma^2\right) = \gamma^2.
\end{aligned}
$$

This justifies that $f_{\tau_\natural}$ is a solution to (10). □

Justification of **b)**. Here, we aim at showing that (10) admits a unique minimizer. Let $\widehat{f}$ and μ^2 denote a minimizer and the minimal value of (10), respectively. We first claim that

$$\|L^{1/2}\widehat{f}\|_2 = \frac{\varepsilon}{\eta}\|\Lambda\widehat{f} - y\|_2 = \mu. \tag{11}$$

Indeed, suppose e.g. that $\|L^{1/2}\widehat{f}\|_2 < (\varepsilon/\eta)\|\Lambda\widehat{f} - y\|_2 = \mu$. Pick an $h \in \mathbb{R}^N$ such that $\langle \Lambda\widehat{f} - y, \Lambda h\rangle \neq 0$ (which exists, for otherwise $\Lambda^*(\Lambda\widehat{f} - y) = 0$, hence $\Lambda\widehat{f} - y = \Lambda\Lambda^*(\Lambda\widehat{f} - y) = 0$, and so $\mu = 0$, in which case $\|L^{1/2}\widehat{f}\|_2 < \mu$ cannot occur). Then, considering $\widehat{f}_t := \widehat{f} + th$ for a small enough t in absolute value, we see that

$$\frac{\varepsilon}{\eta}\|\Lambda\widehat{f}_t - y\|_2 = \frac{\varepsilon}{\eta}\left(\|\Lambda\widehat{f} - y\|_2 + t\langle\Lambda\widehat{f} - y, \Lambda h\rangle + o(t)\right)$$

can be made smaller that μ, while $\|L^{1/2}\widehat{f}_t\|_2$ can remain smaller than μ. This contradicts the defining property of $\widehat{f}$ and establishes (11).

Now let $\widehat{f}$ and $\widetilde{f}$ be two minimizers of (10). Applying (11) to $\widehat{f}$, $\widetilde{f}$, and $(\widehat{f} + \widetilde{f})/2$, which is also a minimizer of (10), yields

$$
\begin{aligned}
\left\|\frac{1}{2}L^{1/2}\widehat{f} + \frac{1}{2}L^{1/2}\widetilde{f}\right\|_2 &= \|L^{1/2}\widehat{f}\|_2 = \|L^{1/2}\widetilde{f}\|_2 = \mu,\\
\left\|\frac{1}{2}\Lambda(\widetilde{f} - y) + \frac{1}{2}\Lambda(\widehat{f} - y)\right\|_2 &= \|\Lambda\widehat{f} - y\|_2 = \|\Lambda\widetilde{f} - y\|_2 = \frac{\eta}{\varepsilon}\mu,
\end{aligned}
$$

which forces $L^{1/2}\widehat{f} = L^{1/2}\widetilde{f}$ and $\Lambda\widehat{f} = \Lambda\widetilde{f}$, implying that i.e. $\widehat{f} - \widetilde{f} \in \ker(L^{1/2}) \cap \ker(\Lambda) = \ker(L) \cap \ker(\Lambda) = \{0\}$, i.e., that $\widehat{f} = \widetilde{f}$ is a unique minimizer. □

Justification of **c)**. Since the original signal $f \in \mathbb{R}^N$ that we try to recover satisfies $\|L^{1/2}f\|_2 \leq \varepsilon$ and $\|\Lambda f - y\|_2 \leq \eta$, it is clear that the minimizer $\widehat{f}$ of (10) satisfies $\|L^{1/2}\widehat{f}\|_2 \leq \varepsilon$ and $\|\Lambda\widehat{f} - y\|_2 \leq \eta$, too. In other words, it is model- and data-consistent, which always leads to near optimality of the local worst case error with a factor 2. Indeed, considering the set $\{Q(f) \colon f \in \mathcal{K}, e \in \mathcal{E}, \Lambda f + e = y\}$, let $f^\star$ denote its Chebyshev center (in our situation, it exists and is unique, see Garkavi [1962]). Then, for any $f \in \mathcal{K}$ and $e \in \mathcal{E}$ with $\Lambda f + e = y$, we have

$$
\begin{aligned}
\|Q(f) - Q(\widehat{f})\|_2 &\leq \|Q(f) - Q(f^\star)\|_2 + \|Q(\widehat{f}) - Q(f^\star)\|_2\\
&\leq \mathrm{lwce}_Q(y, f^\star) + \mathrm{lwce}_Q(y, f^\star)\\
&= 2\inf_{z\in\mathbb{R}^N} \mathrm{lwce}_Q(y, z).
\end{aligned}
$$

Taking the supremun over the admissable $f \in \mathcal{K}$ and $e \in \mathcal{E}$ gives $\mathrm{lwce}_Q(y, \widehat{f}) \leq 2\inf_{z\in\mathbb{R}^N} \mathrm{lwce}_Q(y, z)$, as desired. □

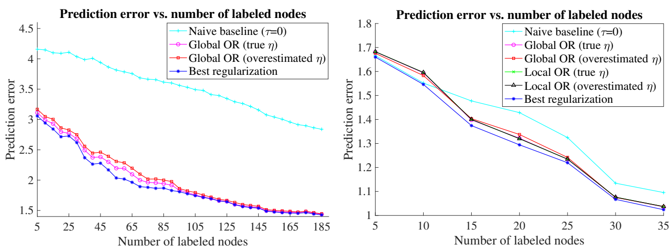
§9. *Implementation details and additional experiments:* We consider the following well-known graph datasets: adjnoun (112 nodes, 425 edges) [Newman, 2006], Netscience (379 nodes, 914 edges) [Girvan and Newman, 2002], polbooks (105 nodes, 441 edges) [Krebs], lesmis (77 nodes, 254 edges) [], and dolphins (62 nodes, 159 edges) [Lusseau et al., 2003]. All of these can be downloaded from the Suitesparse Matrix Collection [Davis and Hu, 2011]. When generating synthetic signals, we follow an approach similar to Equation (15) in [Dong et al., 2019]. Let $L = \chi D \chi^T$ be an eigendecomposition of the graph Laplacian, let $D^\dagger$ be the pseudoinverse of D, and let $c \sim \mathcal{N}(0, D^\dagger)$ be a Gaussian vector. The ground truth labels are then given by $f = \chi c$. The main difference with [Dong et al., 2019] is that f is not assumed to be corrupted simply by Gaussian noise, but we consider different additive noise vectors satisfying $\|e\|_2 \leq \eta$. In the main text, the plots were shown for a noise vector that is generated by taking a uniform random noise vector and subtracting the mean, and before scaling to ensure that $\|e\|_2 \leq \eta$. Here, to illustrate results of a more deterministic flavor, we show results for noise of magnitude proportional to the node degree (Figure 4) and to the inverse degree (Figure 5).

Keeping the same parameters as those used in the main text, we test several optimal recovery methods on different graphs and with different error models. Figures 3-5 support our conclusions that a mild overestimation of η does not lead to bad prediction error and that the prediction errors attached to the locally/globally optimal recovery maps are close to the smallest prediction error possible for any choice of regularization parameter. This confirms that these methods provide a suitable way to choose regularization parameters.

It is worth pointing out that the globally optimal recovery map is linear since the regularization parameter does not depend on the observation vector y. In contrast, the locally near optimal recovery map is nonlinear since the unique parameter $\tau_\natural$ satisfying

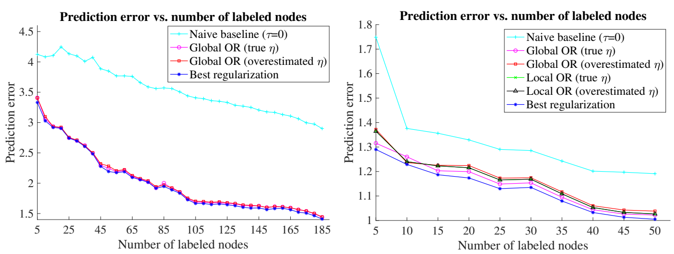
$$\|L^{1/2}\Delta_\tau(y)\|_2 = \frac{\varepsilon}{\eta}\|\Lambda\Delta_\tau(y) - y\|_2$$

does depend on the observation vector y. When implementing globally optimal recovery maps, we compute the globally optimal regularization parameter $\tau_\flat$ for each n_ℓ once and make a prediction when receiving different observation vectors y. For locally optimal recovery maps, we have to recompute the locally near optimal parameter $\tau_\natural$ when receiving new observation vectors y. Therefore, it is recommended to opt for the globally optimal recovery map in order to reduce computational complexity, see e.g. Figures 3(a), 4(a), and 5(a) where the locally near optimal recovery map is not executed. However, dealing with large graphs may result in semidefinite programs that cannot be run, so it can be better to implement the locally near optimal recovery map by using the bisection method to find the near optimal parameter $\tau_\natural$.



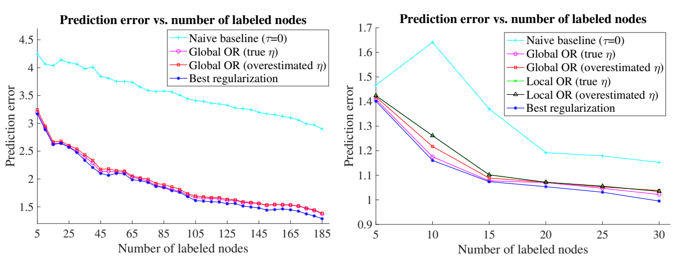
(a) (b)

Fig. 3. Prediction errors vs. number of labeled nodes on two different graphs with additive noise generated uniformly: (a) Netscience (b) lesmis.



(a) (b)

Fig. 4. Prediction errors vs. number of labeled nodes on two different graphs with additive noise proportional to degree: (a) Netscience (b) polbooks.



(a) (b)

Fig. 5. Prediction errors vs. number of labeled nodes on two different graphs with additive noise proportional to the inverse of degree: (a) Netscience (b) dolphins.

§10. *Numerical computation of the upper bound:* Given $y \in \mathbb{R}^{n_\ell}$, the square of the local worst-case error $\mathrm{lwce}_Q(y,z)$ for the estimation of Q by $z \in \mathbb{R}^n$ is

$$\sup_{f \in \mathbb{R}^N} \|Q(f) - z\|_2^2 \quad \text{s.to } \|L^{1/2} f\|_2^2 \le \epsilon^2, \|\Lambda f - y\|_2^2 \le \eta^2.$$

Introducing a slack variable γ, we write the above optimization program as

$$\inf_{\gamma} \gamma \quad \text{s.to } \|Q(f) - z\|_2^2 \le \gamma \quad \text{whenever } \|L^{1/2} f\|_2^2 \le \epsilon^2, \|\Lambda f - y\|_2^2 \le \eta^2.$$

The constraint is a consequence of (but is not equivalent to) the existence of $c, d \ge 0$ such that

$$\|Q(f) - z\|_2^2 - \gamma \le c(\|L^{1/2} f\|_2^2 - \epsilon^2) + d(\|\Lambda f - y\|_2^2 - \eta^2)$$

for all $f \in \mathbb{R}^N$. The latter can be also reformulated as the condition that, for all $f \in \mathbb{R}^N$,

$$\langle (cL + d\Lambda^*\Lambda - Q^*Q) f, f \rangle - 2 \langle Q^* z - \Lambda^* y, f \rangle + \gamma - \|z\|_2^2 - c\epsilon^2 + d(\|y\|_2^2 - \eta^2) \ge 0,$$

or, more succinctly, that

$$\left[\begin{array}{c|c} cL + d\Lambda^*\Lambda - Q^*Q & Q^*z - d\Lambda^* y \\ \hline (Q^*z - d\Lambda^* y)^* & \gamma - \|z\|_2^2 - c\epsilon^2 + d(\|y\|_2^2 - \eta^2) \end{array}\right] \succeq 0. \quad (12)$$

We conclude from the above considerations that the squared local worst-case error is upper-bounded by the optimal value of a semidefinite program, namely

$$\mathrm{lwce}_Q(y,z)^2 \le \inf_{\substack{\gamma \\ c,d \ge 0}} \gamma \quad \text{s.to (12)}.$$