

Beyond Perturbation Stability: Supplementary Material

A Preliminaries Details

Claim A.1. *For Uniform Metric Labeling, we can assume $c(u, i) \geq 0$ and $w(u, v) > 0$ without loss of generality.*

Proof. For problem instances where some node costs are strictly negative, let c_{\min} be the minimum value among all the node costs. Consider a new problem instance where we keep the edge costs the same, but set $c'(u, i) = c(u, i) + |c_{\min}|$ for all $u \in V$ and $i \in [k]$. This new problem instance has all non-negative node costs, and the optimization problem is equivalent, because we added the same constant for all solutions. This reformulation also does not affect the $(2, 1)$ -expansion stability or $(2, 1, \psi)$ -expansion stability of the instance.

Likewise, for problem instances where some edge weights are 0, let E_0 be the set of all edges with 0 edge weight. Consider a new problem instance with $E' = E \setminus E_0$, with $w(u, v)$ unchanged for $(u, v) \in E \setminus E_0$, and identical node costs. The MAP optimization problem remains the same, and the new instance $((V, E'), c, w)$ is equivalent: it has the same MAP solution, and satisfies the stability definitions if and only if the original instance does as well. \square

Claim 3.3. *For a given graph G , every solution $x \in L(G)$ that minimizes $\langle \theta, x \rangle$ for some valid objective vector $\theta = (c, w)$ also belongs to $L^*(G)$. Further, all integer solutions in $L(G)$ also belong to $L^*(G)$.*

Proof of Claim 3.3. Recall the local LP:

$$\min_x \sum_{u \in V} \sum_i c(u, i) x_u(i) + \sum_{(u, v) \in E} w(u, v) \sum_{i \neq j} x_{uv}(i, j) \quad (6)$$

$$\text{subject to: } \sum_i x_u(i) = 1 \quad \forall u \in V \quad (7)$$

$$\sum_i x_{uv}(i, j) = x_v(j) \quad \forall (u, v) \in E, j \in [k] \quad (8)$$

$$\sum_j x_{uv}(i, j) = x_u(i) \quad \forall (u, v) \in E, i \in [k] \quad (9)$$

$$x_{uv}(i, j) \in [0, 1] \quad \forall (u, v), (i, j) \quad (10)$$

$$x_u(i) \in [0, 1] \quad \forall u, i. \quad (11)$$

The feasible region defined by the above constraints is $L(G)$. $L^*(G) \subseteq L(G)$ is the set of points that satisfy the additional constraint that $x_{uv}(i, i) = \min(x_u(i), x_v(i))$ for all $(u, v) \in E$ and $i \in [k]$. For any feasible node variable assignments $\{x_u\}$, $L^*(G)$ is not empty: a simple flow argument² implies that the constraints (8), (9), and (10) are always satisfiable even when we set $x_{uv}(i, i) = \min(x_u(i), x_v(i))$. For all integer feasible solutions in $L(G)$, notice that $x_{uv}(i, j) = 1$ if $x_u(i) = 1$ and $x_v(j) = 1$ or 0 otherwise. Therefore, all integer solutions satisfy this additional constraint. Consider a θ where all edge weights are strictly positive. If x minimizes $\langle \theta, x \rangle$, x must

²For an edge (u, v) , consider the bipartite graph $\tilde{G} = ((U, V), E)$, where $|U| = |V| = k$. We let $x_u(i)$ represent the supply at node i in U , and let $x_v(i)$ represent the demand at node j in V . Because x_u and x_v are both feasible, the total supply equals the total demand. E contains all edges between U and V , so we can send flow from $i \in U$ to $j \in V$ for any (i, j) pair. Let $x_{uv}(i, j)$ represent this flow, and set $x_{uv}(i, i) = \min(x_u(i), x_v(i))$. For every i , this either satisfies the demand at node V_i or exhausts the supply at node U_i . In each case, we can remove that satisfied/exhausted node from the graph. After this choice of $x_u(i, i)$, the total remaining supply equals the total remaining demand ($\sum_i x_u(i) - \min(x_u(i), x_v(i)) = \sum_i x_v(i) - \min(x_u(i), x_v(i))$), all supplies and demands are nonnegative, and the remaining graph \tilde{G}' is a complete bipartite graph (over fewer nodes). This implies that the flow constraints (8), (9), (10) are still feasible.



Figure 3: $(2, 1)$ -expansion stable instance that is not $(2, 1)$ -stable. In the original instance (shown left), the optimal solution labels each vertex with label 1, for an objective of 2.5. The adversarial $(2, 1)$ -perturbation for this instance replaces all the edge weights of $1 + \varepsilon$ with $(1 + \varepsilon)/2$. In this perturbed instance, the optimal solution labels $(u, v, w) \rightarrow (1, 2, 3)$. This has a node cost of 0.5 and an edge cost of $(3 + 3\varepsilon)/2$, for a total of $2 + 3\varepsilon/2 < 2.5$. Since the original solution is not optimal in the perturbed instance, this instance is not $(2, 1)$ -perturbation stable. However, note that the only expansions of the original solution (which had all label 1) that have non-infinite objective are $(u, v, w) \rightarrow (1, 2, 1)$ and $(u, v, w) \rightarrow (1, 1, 3)$. These each have objective $2.5 + \varepsilon$, which is strictly greater than the perturbed objective of the original solution. Therefore, this instance is $(2, 1)$ -expansion stable.

pay the minimum edge cost consistent with its node variables $x_u(i)$. So if we fix the $x_u(i)$ portion of x , we know that the edge variables x_{uv} of x are a solution to:

$$\min_{x \in L(G)} \sum_{(u,v) \in E} w(u,v) \sum_{i \neq j} x_{uv}(i,j).$$

Notice that since we have fixed the node variables $x_u(i)$, there is no interaction between the x_{uv} variables across different edges. So we can minimize this objective by minimizing each individual term $w(u,v) \sum_{i \neq j} x_{uv}(i,j)$. Since $w_{uv} > 0$ for all edges, we need to minimize $\sum_{i \neq j} x_{uv}(i,j)$. Notice that for every edge $(u,v) \in E$, we get that $\sum_i \sum_j x_{uv}(i,j) = 1$ by substituting $x_u(i)$ in constraint 7 with $\sum_j x_{uv}(i,j)$ from constraint 9. Therefore $\sum_{i \neq j} x_{uv}(i,j) = 1 - \sum_i x_{uv}(i,i)$. Thus, minimizing $\sum_{i \neq j} x_{uv}(i,j)$ is the same as maximizing $\sum_i x_{uv}(i,i)$. And the maximizing choice for $x_{uv}(i,i) = \min(x_u(i), x_v(i))$ due to constraints 8 and 9. \square

B Expansion Stability details

Claim B.1. *An instance (G, w, c) is $(2, 1)$ -expansion stable iff the MAP solution \bar{x} is strictly better than all its expansions in the adversarial perturbation θ_{adv} . That is, for all $x \in \mathcal{E}_{\bar{x}}$, $\langle \theta_{adv}, x \rangle > \langle \theta_{adv}, \bar{x} \rangle$ where θ_{adv} has the same node costs c but has weights $w_{adv}(u,v) = \begin{cases} \frac{1}{2}w(u,v) & \bar{x}(u) = \bar{x}(v) \\ w(u,v) & \bar{x}(u) \neq \bar{x}(v). \end{cases}$*

Proof. Consider $\theta' = (c, w')$, any valid $(2, 1)$ -perturbation of $\theta = (c, w)$ i.e. for every edge $(u, v) \in E$, $\frac{w(u,v)}{2} \leq w'(u,v) \leq w(u,v)$. For any valid labeling x , let E_x represent the edges cut by x . Then, for any x which is an expansion of \bar{x} i.e. $x \in \mathcal{E}_{\bar{x}}$,

$$\begin{aligned} \langle \theta', x \rangle - \langle \theta', \bar{x} \rangle &= \sum_{u \in V} c(u, x(u)) - c(u, \bar{x}(u)) + \sum_{(u,v) \in E_x} w'(u,v) - \sum_{(u,v) \in E_{\bar{x}}} w'(u,v) \\ &= \sum_{u \in V} c(u, x(u)) - c(u, \bar{x}(u)) + \sum_{(u,v) \in E_x \setminus E_{\bar{x}}} w'(u,v) - \sum_{(u,v) \in E_{\bar{x}} \setminus E_x} w'(u,v) \\ &= \langle \theta_{adv}, x \rangle - \langle \theta_{adv}, \bar{x} \rangle + \sum_{(u,v) \in E_x \setminus E_{\bar{x}}} w'(u,v) - w_{adv}(u,v) + \sum_{(u,v) \in E_{\bar{x}} \setminus E_x} w_{adv} - w'(u,v) \\ &= \langle \theta_{adv}, x \rangle - \langle \theta_{adv}, \bar{x} \rangle + \sum_{(u,v) \in E_x \setminus E_{\bar{x}}} w'(u,v) - \frac{w(u,v)}{2} + \sum_{(u,v) \in E_{\bar{x}} \setminus E_x} w(u,v) - w'(u,v) \end{aligned}$$

Since w' is a valid $(2, 1)$ -perturbation, $w'(u,v) \geq w(u,v)/2$ and $w'(u,v) \leq w(u,v)$. Therefore, for any valid $(2, 1)$ -perturbation θ' , we have

$$\langle \theta', x \rangle - \langle \theta', \bar{x} \rangle \geq \langle \theta_{adv}, x \rangle - \langle \theta_{adv}, \bar{x} \rangle.$$

If the instance is $(2, 1)$ -expansion stable, then certainly $\langle \theta_{adv}, x \rangle > \langle \theta_{adv}, \bar{x} \rangle$ for all $x \in \mathcal{E}_{\bar{x}}$, since θ_{adv} is a valid $(2, 1)$ -perturbation of θ . If the instance is not $(2, 1)$ -expansion stable, there exists a θ' and an $x \in \mathcal{E}_{\bar{x}}$ for which

$\langle \theta', x \rangle - \langle \theta', \bar{x} \rangle \leq 0$. But the above inequality then implies that $\langle \theta_{adv}, x \rangle - \langle \theta_{adv}, \bar{x} \rangle \leq 0$ as well. This gives both directions. \square

This claim shows that to check whether an instance is $(2, 1)$ -expansion stable, it is sufficient to check that the MAP solution is strictly better than all its expansions in the adversarial perturbation θ_{adv} . We don't need to verify that this condition is satisfied in *every* valid $(2, 1)$ -perturbation. Because the optimal expansion of \bar{x} in the instance with objective θ_{adv} can be computed efficiently, this claim also implies that $(2, 1)$ -expansion stability can be efficiently checked once the MAP solution \bar{x} is known.

Claim B.2. $(2, 1)$ -expansion stability is strictly weaker than $(2, 1)$ -perturbation stability.

Proof. Figure 2 gives an instance of uniform metric labeling that is $(2, 1)$ -expansion stable but not $(2, 1)$ -perturbation stable. Here, $0 < \varepsilon < 1/3$. \square

Theorem 4.2 (Local LP is tight on $(2, 1)$ -expansion stable instances). *Let \bar{x} and \hat{x} be the MAP and local LP solutions to a $(2, 1)$ -expansion stable instance (G, c, w) , respectively. Then $\bar{x} = \hat{x}$ i.e. the local LP is tight on (G, c, w) .*

Proof. First, we note that for any $x \in L^*(G)$, the objective value of the local LP can be written in a form that depends only on the node variables x_V . The objective term corresponding to the edges

$$\begin{aligned} \sum_{(u,v) \in E} w(u,v) \sum_{i \neq j} x_{uv}(i,j) &= \sum_{(u,v) \in E} w(u,v) \left(\sum_{i,j} x_{uv}(i,j) - \sum_i x_{uv}(i,i) \right) \\ &= \sum_{(u,v) \in E} w(u,v) \left(1 - \sum_i x_{uv}(i,i) \right) = \sum_{(u,v) \in E} w(u,v) \left(1 - \sum_i \min(x_u(i), x_v(i)) \right) \\ &= \sum_{(u,v) \in E} w(u,v) \left(1 - \sum_i \left(\frac{x_u(i) + x_v(i)}{2} - \frac{|x_u(i) - x_v(i)|}{2} \right) \right) \\ &= \sum_{(u,v) \in E} w(u,v) \left(\frac{1}{2} \sum_i |x_u(i) - x_v(i)| \right) \end{aligned}$$

Here we used the definition of $L^*(G)$ and the facts that $\sum_i x_u(i) = 1$ for all (u, i) and $\sum_j x_{uv}(i, j) = x_u(i)$ for all $(u, v) \in E$, $i \in [k]$.

Thus, for any $x \in L^*(G)$, the objective of the local LP can be written as

$$\sum_{u \in V} \sum_i c(u, i) x_u(i) + \sum_{(u,v) \in E} w(u, v) d(u, v)$$

where $d(u, v) := \frac{1}{2} \sum_i |x_u(i) - x_v(i)|$. This is the objective function of another LP relaxation for uniform metric labeling called the ‘‘metric LP’’, which is equivalent to the local LP (Archer et al., 2004). Note that both \bar{x} and \hat{x} are in $L^*(G)$ by Claim 3.3. Therefore, the objective function can be written in the above form for both of them.

In the next section, we introduce a rounding algorithm and prove some guarantees for the random solutions h output by it. We then use these guarantees to show an upper bound on the expected cost of these random solutions in a perturbed instance of the problem. Finally, we use this upper bound to prove that $\hat{x} = \bar{x}$.

B.1 ε -close rounding:

Given any feasible solution $x \in L(G)$ and a valid labeling \bar{x} , we construct a related feasible solution x' which is ε -close to \bar{x} in the ℓ_∞ -norm i.e. $\|x' - \bar{x}\|_\infty \leq \varepsilon$:

$$x' = \varepsilon x + (1 - \varepsilon) \bar{x}, \tag{12}$$

where $\varepsilon < 1/k$ and we have identified the labeling $\bar{x} : V \rightarrow \mathcal{L}$ with its corresponding vertex of the marginal polytope (a vector in $\{0, 1\}^{nk+mk^2}$). We consider the following rounding algorithm applied to x' , which is a modified version of the ε -close rounding algorithm used in Lang et al. (2018):

Algorithm 1 ε -close rounding

```

1: Choose  $i \in \{1, \dots, k\}$  uniformly at random.
2: Choose  $r \in (0, 1/k)$  uniformly at random.
3: Initialize labeling  $h : V \rightarrow [k]$ .
4: for each  $u \in V$  do
5:   if  $x'_u(i) > r$  then
6:     Set  $h(u) = i$ .
7:   else
8:     Set  $h(u) = \bar{x}(u)$ 
9:   end if
10: end for
11: Return  $h$ 
    
```

Lemma B.3 (Rounding guarantees). *Given any x' constructed using (12), the labeling h output by Algorithm 1 satisfies the following guarantees:*

$$\begin{aligned}
 \mathbb{P}\{h(u) = i\} &= x'_u(i) && \forall u \in V, i \in [k] \\
 \mathbb{P}\{h(u) \neq h(v)\} &\leq 2d(u, v) && \forall (u, v) \in E : \bar{x}(u) = \bar{x}(v) \\
 \mathbb{P}\{h(u) = h(v)\} &= (1 - d(u, v)) && \forall (u, v) \in E : \bar{x}(u) \neq \bar{x}(v),
 \end{aligned}$$

where $d(u, v) = \frac{1}{2} \sum_i |x'_u(i) - x'_v(i)|$ is the edge separation of the constructed feasible point x' .

Proof of Lemma B.3 (rounding guarantees). First, fix $u \in V$ and a label $i \neq \bar{x}(u)$. We output $h(u) = i$ precisely when i is chosen and $0 < r < x'_u(i)$, which occurs with probability $\frac{1}{k} \frac{x'_u(i)}{1/k} = x'_u(i)$ (we used here that $x'_u(i) \leq \varepsilon < 1/k$ for all $i \neq \bar{x}(u)$). Now we output $h(u) = \bar{x}(u)$ with probability $1 - \sum_{j \neq \bar{x}(u)} \mathbb{P}\{h(u) = j\} = 1 - \sum_{j \neq \bar{x}(u)} x'_u(j) = x'_u(\bar{x}(u))$, since $\sum_i x'_u(i) = 1$. This proves the first guarantee.

For the second, consider an edge (u, v) not cut by \bar{x} , so $\bar{x}(u) = \bar{x}(v)$. Then (u, v) is cut by h when some $i \neq \bar{x}(u)$ is chosen and r falls between $x'_u(i)$ and $x'_v(i)$. This occurs with probability

$$\frac{1}{k} \sum_{i \neq \bar{x}(u)} \frac{\max(x'_u(i), x'_v(i)) - \min(x'_u(i), x'_v(i))}{1/k} = \sum_{i \neq \bar{x}(u)} |x'_u(i) - x'_v(i)| \leq 2d(u, v).$$

Finally, consider an edge (u, v) cut by \bar{x} , so that $\bar{x}(u) \neq \bar{x}(v)$. Here $h(u) = h(v)$ if some i, r are chosen with $r < \min(x'_u(i), x'_v(i))$. We have $r < \min(x'_u(i), x'_v(i))$ with probability $\frac{\min(x'_u(i), x'_v(i))}{1/k}$. Note that this is still valid if $i = \bar{x}(u)$ or $i = \bar{x}(v)$, since only one of those equalities can hold. So we get

$$\frac{1}{k} \sum_i \frac{\min(x'_u(i), x'_v(i))}{1/k} = \frac{1}{2} \left(\sum_i x'_u(i) + x'_v(i) - |x'_u(i) - x'_v(i)| \right) = 1 - d(u, v),$$

where we used again that $\sum_i x'_u(i) = 1$. □

Given these rounding guarantees, we can relate the expected cost difference between h and \bar{x} in a perturbation of the original instance to the cost difference between x and \bar{x} in the original instance. We are only interested in the case when $x \in L^*(G)$ and so the objective function $f(x) = \sum_{u \in V} \sum_i c(u, i)x_u(i) + \sum_{(u, v) \in E} w(u, v)d(u, v)$.

B.2 Using the rounding guarantees

Lemma B.4. *Given an integer solution \bar{x} , a feasible LP solution $x \in L^*(G)$, and a random output h of Algorithm 1 on an input $x' = \varepsilon x + (1 - \varepsilon)\bar{x}$, define*

$$w'(u, v) = \begin{cases} \frac{1}{2}w(u, v) & \bar{x}(u) = \bar{x}(v) \\ w(u, v) & \bar{x}(u) \neq \bar{x}(v) \end{cases}$$

and let $f'(y) = \sum_{u \in V} \sum_i c(u, i)y_u(i) + \sum_{(u, v) \in E} w'(u, v)d(y, u, v)$ be the objective in the instance with the original costs, but using weights w' . Here $d(y, u, v) = \frac{1}{2} \sum_i |y_u(i) - y_v(i)|$. Let $A_h := f'(h) - f'(\bar{x})$ be the difference in this perturbed objective between h and \bar{x} . Then,

$$\mathbb{E}[A_h] = \mathbb{E}[f'(h) - f'(\bar{x})] \leq \varepsilon \cdot (f(x) - f(\bar{x})).$$

Proof.

$$\begin{aligned} \mathbb{E}[f'(h) - f'(\bar{x})] &= \sum_{u \in V} \sum_i c(u, i) \mathbb{P}\{h(u) = i\} - \sum_{u \in V} c(u, \bar{x}(u)) + \sum_{uv: \bar{x}(u) = \bar{x}(v)} w'(u, v) \mathbb{P}\{h(u) \neq h(v)\} \\ &\quad - \sum_{uv: \bar{x}(u) \neq \bar{x}(v)} w'(u, v) \mathbb{P}\{h(u) = h(v)\} \\ &= \sum_u \sum_i c(u, i) x'_u(i) - \sum_u c(u, \bar{x}(u)) + \sum_{uv: \bar{x}(u) = \bar{x}(v)} 2w'(u, v) d(x', u, v) \\ &\quad - \sum_{uv: \bar{x}(u) \neq \bar{x}(v)} w'(u, v) (1 - d(x', u, v)) \\ &= \sum_u \sum_i c(u, i) x'_u(i) - \sum_u c(u, \bar{x}(u)) + \sum_{uv \in E} w(u, v) d(x', u, v) - \sum_{uv: \bar{x}(u) \neq \bar{x}(v)} w(u, v) \\ &= f(x') - f(\bar{x}). \end{aligned}$$

where the second-to-last equality used the definition of w' (note that w' is identical to the worst-case perturbation w_{adv} for \bar{x}). Because f is convex (in particular, $d(x, u, v)$ is convex in x), we have $f(x') \leq \varepsilon f(x) + (1 - \varepsilon)f(\bar{x})$. Therefore,

$$\mathbb{E}[f'(h) - f'(\bar{x})] \leq \varepsilon f(x) + (1 - \varepsilon)f(\bar{x}) - f(\bar{x}) = \varepsilon(f(x) - f(\bar{x})),$$

which is what we wanted. \square

B.3 Final proof of Theorem 4.2:

Suppose the local LP solution \hat{x} is not the same as the MAP solution \bar{x} i.e. $\hat{x} \neq \bar{x}$. Consider $x' = \varepsilon \hat{x} + (1 - \varepsilon)\bar{x}$ where $0 < \varepsilon < 1/k$ (see equation (12)). Let h be the random integer solution output by using Algorithm 1 on x' . By Lemma B.4, we have

$$\mathbb{E}[f'(h) - f'(\bar{x})] \leq \varepsilon \cdot (f(\hat{x}) - f(\bar{x}))$$

We note that any solution h that we get from rounding x' is either \bar{x} or an expansion move of \bar{x} . This is because we pick only a single label i in step 1 of Algorithm 1 and label all vertices u either i or $\bar{x}(u)$. Therefore, for the i picked in step 1, h is an i -expansion of \bar{x} if $h \neq \bar{x}$.

$$\begin{aligned} \mathbb{E}[f'(h) - f'(\bar{x})] &= \mathbb{E}[f'(h) - f'(\bar{x}) | h \neq \bar{x}] \mathbb{P}[h \neq \bar{x}] + \mathbb{E}[f'(h) - f'(\bar{x}) | h = \bar{x}] \mathbb{P}[h = \bar{x}] \\ &= \mathbb{E}[f'(h) - f'(\bar{x}) | h \neq \bar{x}] \mathbb{P}[h \neq \bar{x}] \end{aligned}$$

Since (G, c, w) is a $(2, 1)$ -expansion stable instance, we know that $f'(h) > f'(\bar{x})$ when $h \neq \bar{x}$ since all h in the support of the rounding (other than \bar{x}) are expansion moves of \bar{x} and we get f' by a valid $(2, 1)$ -perturbation of (G, c, w) . Therefore, $\mathbb{E}[f'(h) - f'(\bar{x}) | h \neq \bar{x}] > 0$. We also have that $\mathbb{P}[h \neq \bar{x}] > 0$ since we assumed that $\hat{x} \neq \bar{x}$. Therefore, $\mathbb{E}[f'(h) - f'(\bar{x})] > 0$. But we know that $f(\hat{x}) - f(\bar{x}) \leq 0$ since \hat{x} is the minimizer of $f(x)$ among all feasible $x \in L(G)$. So Lemma B.4 implies $\mathbb{E}[f'(h) - f'(\bar{x})] \leq 0$. Thus we have a contradiction and so the local LP solution \hat{x} has to be the same as the MAP solution \bar{x} . \square

C Stability and Curvature around MAP solution: details

Theorem 5.2. *Let (G, c, w) be a $(2, 1, \psi)$ -expansion stable instance with MAP solution \bar{x} . Let $\theta = (c, w)$. Then for any $x \in L^*(G)$, the recovery error (see Def. 3.1) satisfies*

$$\frac{1}{2}\|x - \bar{x}\|_1 := \frac{1}{2}\|x_V - \bar{x}_V\|_1 \leq \frac{1}{\psi}|\langle \theta, x \rangle - \langle \theta, \bar{x} \rangle|. \quad (2)$$

Here, we provide two proofs for this theorem, one deals directly with the local LP relaxation and the other uses the dual of the relaxation. The dual proof is more general than the primal proof as it works for all $x \in L(G)$, not just for those in $L^*(G)$.

C.1 Primal-based proof

Proof. For any $x \in L^*(G)$, consider a feasible solution x' which is ε -close to \bar{x} constructed using Equation 12 i.e. $x' = \varepsilon x + (1 - \varepsilon)\bar{x}$. Let h be the random solution output by Algorithm 1 on x' .

Lemma C.1 (Bound for $\mathbb{E}[B_h]$). *For any h in the support of the rounding of $x' = \varepsilon x + (1 - \varepsilon)\bar{x}$, let us define B_h to be the number of vertices which it labels differently from \bar{x} . In other words, it is the number of vertices which are misclassified by h i.e. $B_h := \sum_{u \in V} \mathbb{1}[h(u) \neq \bar{x}(u)]$. Then,*

$$\mathbb{E}[B_h] = \varepsilon \sum_{u \in V} \frac{1}{2}\|x_u - \bar{x}_u\|_1$$

Proof.

$$\begin{aligned} \mathbb{E}[B_h] &= \sum_{u \in V} \mathbb{E}[\mathbb{1}[h(u) \neq \bar{x}(u)]] = \sum_{u \in V} \mathbb{P}\{h(u) \neq \bar{x}(u)\} = \sum_{u \in V} 1 - \mathbb{P}\{h(u) = \bar{x}(u)\} \\ &= \sum_{u \in V} 1 - x'_u(\bar{x}(u)) = \sum_{u \in V} 1 - (\varepsilon x_u(\bar{x}(u)) + (1 - \varepsilon)) = \sum_{u \in V} \varepsilon (1 - x_u(\bar{x}(u))) \\ &= \varepsilon \sum_{u \in V} \frac{1}{2} \left(1 - x_u(\bar{x}(u)) + \sum_{i \neq \bar{x}(u)} x_u(i) \right) = \varepsilon \sum_{u \in V} \frac{1}{2}\|x_u - \bar{x}_u\|_1 \end{aligned}$$

Here, we used the fact that for all $u \in V$, $\bar{x}_u(\bar{x}(u)) = 1$ and $\bar{x}_u(i) = 0 \forall i \neq \bar{x}(u)$ and since x is a feasible solution to the LP, it satisfies $\sum_{i \neq \bar{x}(u)} x_u(i) = 1 - x_u(\bar{x}(u))$ for all $u \in V$. \square

Lemma C.2 (Lower bound for A_h using $(2, 1, \psi)$ -expansion stability). *If (G, w, c) is a $(2, 1, \psi)$ -expansion stable instance, then for any h in the support of the rounding of $x' = \varepsilon x + (1 - \varepsilon)\bar{x}$,*

$$A_h \geq \psi \cdot B_h$$

where $A_h = f'(h) - f'(\bar{x})$ and f' is the objective in the instance (G, c, w') where w' is the worst $(2, 1)$ perturbation for \bar{x} i.e.

$$w'(u, v) = \begin{cases} \frac{1}{2}w(u, v) & \bar{x}(u) = \bar{x}(v) \\ w(u, v) & \bar{x}(u) \neq \bar{x}(v) \end{cases}$$

Proof. Note that A_h here is the same as the one defined for Lemma B.4. Since the instance (G, c, w) is $(2, 1, \psi)$ -expansion stable, we know that (G, c', w) should be $(2, 1)$ -expansion stable for all c' such that $c \leq c' \leq c + \psi \cdot \mathbf{1}$.

Consider the worst c' for \bar{x} i.e. $c'(u, i) = \begin{cases} c(u, i) + \psi & i = \bar{x}(u) \\ c(u, i) & i \neq \bar{x}(u) \end{cases}$. Let f'' be the objective in the instance (G, c', w') .

As discussed in section B.3, we know that any $h \neq \bar{x}$ in the support of the rounding is an expansion move of \bar{x} . Therefore, for any $h \neq \bar{x}$ in the support of the rounding of x' ,

$$\begin{aligned}
 f''(h) - f''(\bar{x}) > 0 &\implies \sum_{u \in V} c'(u, h(u)) - c'(u, \bar{x}(u)) + \sum_{(u,v):h(u) \neq h(v)} w'(u, v) - \sum_{(u,v):\bar{x}(u) \neq \bar{x}(v)} w'(u, v) > 0 \\
 &\implies \sum_{u \in V} c(u, h(u)) - (c(u, \bar{x}(u)) + \psi \cdot \mathbf{1}[h(u) \neq \bar{x}(u)]) + \sum_{(u,v):h(u) \neq h(v)} w'(u, v) - \sum_{(u,v):\bar{x}(u) \neq \bar{x}(v)} w'(u, v) > 0 \\
 &\implies f'(h) - f'(g) > \psi \cdot \sum_{u \in V} \mathbf{1}[h(u) \neq \bar{x}(u)] \implies A_h > \psi \cdot B_h.
 \end{aligned}$$

This is true for all $h \neq \bar{x}$ in the support of the rounding of x' . When $h = \bar{x}$, we have $A_h = B_h = 0$. Therefore for all h in the support of the rounding of x' , we have that $A_h \geq \psi \cdot B_h$. \square

C.2 Final proof of Theorem 5.2:

We use the Lemmas B.4(upper bound for $\mathbb{E}[A_h]$), C.2(lower bound for A_h), and C.1(bound for $\mathbb{E}[B_h]$) to prove Theorem 5.2. For all h in the support of rounding of x , $A_h \geq \psi \cdot B_h$. Also,

$$\mathbb{E}[A_h] \leq \varepsilon (f(x) - f(\bar{x})), \quad \mathbb{E}[B_h] = \varepsilon \sum_{u \in V} \frac{1}{2} \|x_u - \bar{x}_u\|_1$$

Suppose that $\|x - \bar{x}\|_1 > \tau \cdot (f(x) - f(\bar{x}))$. Then,

$$\frac{\mathbb{E}[A_h]}{\mathbb{E}[B_h]} \leq \frac{f(x) - f(\bar{x})}{\sum_{u \in V} \frac{1}{2} \|x_u - \bar{x}_u\|_1} < \frac{2}{\tau}$$

But since $A_h \geq \psi \cdot B_h$ for every h in the rounding of x , we get that $\frac{\mathbb{E}[A_h]}{\mathbb{E}[B_h]} \geq \psi$.

Setting $\tau = \frac{2}{\psi}$, we get a contradiction and thus we get,

$$\frac{1}{2} \|x - \bar{x}\|_1 \leq \frac{1}{\psi} \cdot (f(x) - f(\bar{x})) = \frac{1}{\psi} \cdot (\langle \theta, x \rangle - \langle \theta, \bar{x} \rangle)$$

\square

Corollary F.1 (LP solution is good if there is a nearby stable instance). *Let \hat{x}^{MAP} and \hat{x} be the MAP and local LP solutions to an observed instance (G, \hat{c}, \hat{w}) . Also, let \bar{x} be the MAP solution for a latent $(2, 1, \psi)$ -expansion stable instance (G, \bar{c}, \bar{w}) . If $\hat{\theta} = (\hat{c}, \hat{w})$ and $\bar{\theta} = (\bar{c}, \bar{w})$,*

$$\frac{1}{2} \|\hat{x}_V - \hat{x}_V^{MAP}\|_1 \leq \frac{2d(\hat{\theta}, \bar{\theta})}{\psi} + \frac{1}{2} \|\hat{x}_V^{MAP} - \bar{x}_V\|_1.$$

Proof of Corollary 5.3. First, we note that for the nearby stable instance, the MAP and the local LP solutions are the same due to Theorem 4.2. Therefore, for any feasible solution $x \in L(G)$, $\langle \bar{\theta}, x \rangle \geq \langle \bar{\theta}, \bar{x} \rangle$. In particular, this implies that $\langle \bar{\theta}, \hat{x} \rangle \geq \langle \bar{\theta}, \bar{x} \rangle$ and $\langle \bar{\theta}, \hat{x}^{MAP} \rangle \geq \langle \bar{\theta}, \bar{x} \rangle$ since \hat{x}, \hat{x}^{MAP} are also feasible solutions. Remember that we defined $d(\hat{\theta}, \bar{\theta}) := \sup_{x \in L^*(G)} |\langle \hat{\theta}, x \rangle - \langle \bar{\theta}, x \rangle|$. Therefore,

$$\langle \bar{\theta}, \hat{x} \rangle \leq \langle \hat{\theta}, \hat{x} \rangle + d(\hat{\theta}, \bar{\theta}) \leq \langle \hat{\theta}, \bar{x} \rangle + d(\hat{\theta}, \bar{\theta}) \leq \langle \bar{\theta}, \bar{x} \rangle + 2d(\hat{\theta}, \bar{\theta}).$$

The first and third inequalities hold due to the definition of $d(\hat{\theta}, \bar{\theta})$. The second inequality follows from the fact that \hat{x} is the minimizer for $\langle \hat{\theta}, x \rangle$ among $x \in L(G)$. Thus, $0 \leq \langle \bar{\theta}, \hat{x} \rangle - \langle \bar{\theta}, \bar{x} \rangle \leq 2d(\hat{\theta}, \bar{\theta})$. From Theorem 5.2, we get $\frac{1}{2} \|\hat{x}_V - \bar{x}_V\|_1 \leq \frac{2d(\hat{\theta}, \bar{\theta})}{\psi}$. Thus,

$$\begin{aligned}
 \frac{1}{2} \|\hat{x}_V - \hat{x}_V^{MAP}\|_1 &\leq \frac{1}{2} \|\hat{x}_V - \bar{x}_V\|_1 + \frac{1}{2} \|\hat{x}_V^{MAP} - \bar{x}_V\|_1 \\
 &\leq \frac{2d(\hat{\theta}, \bar{\theta})}{\psi} + \frac{1}{2} \|\hat{x}_V^{MAP} - \bar{x}_V\|_1.
 \end{aligned}$$

\square

C.3 Dual-based proof

Here we provide an alternate proof of the curvature result using the dual of the local LP relaxation. First, we show that the curvature bound is related to the *dual margin* of the instance. Then we show that $(2, 1, \psi)$ -expansion stability implies that the dual margin is at least ψ . Throughout this section, we assume the local LP solution \hat{x} is unique and integral (as guaranteed, for example, by $(2, 1)$ -expansion stability), so $\hat{x} = \bar{x}$.

Relaxing the local LP's marginalization constraints in both directions for each edge, we obtain the following Lagrangian for the local LP:

$$L(\delta, x) = \sum_u \sum_i \left(\theta_u(i) + \sum_{v \in N(u)} \delta_{uv}(i) \right) x_u(i) + \sum_{uv} \sum_{ij} (\theta_{uv}(i, j) - \delta_{uv}(i) - \delta_{vu}(j)) x_{uv}(i, j)$$

where each x_u is constrained to be in the $(k-1)$ -dimensional simplex, and each x_{uv} the k^2-1 -dimensional simplex (i.e., the normalization constraints remain). There are no constraints on the dual variables δ . Observe that for any δ and any primal-feasible x , $L(\delta, x) = \langle \theta, x \rangle$. This gives rise to the *reparametrization* view: for a fixed δ , define $\theta_u^\delta(i) = \theta_u(i) + \sum_{v \in N(u)} \delta_{uv}(i)$, and $\theta_{uv}^\delta(i, j) = \theta_{uv}(i, j) - \delta_{uv}(i) - \delta_{vu}(j)$. Then $L(\delta, x) = \langle \theta^\delta, x \rangle$. This will allow us to define equivalent primal problems with simpler structure than the original. $L(\delta, x)$ also gives the dual function:

$$D(\delta) = \min_x L(\delta, x) = \sum_u \min_i \left(\theta_u(i) + \sum_{v \in N(u)} \delta_{uv}(i) \right) + \sum_{uv} \min_{i,j} (\theta_{uv}(i, j) - \delta_{uv}(i) - \delta_{vu}(j)).$$

A dual point δ is a dual *solution* if $\delta \in \operatorname{argmax}_{\delta'} D(\delta')$. Theorem 4.2 implies that the local LP has a unique, integral solution when the instance is $(2, 1, \psi)$ -expansion stable. Sontag et al. (2011, Theorem 1.3) show that this implies the existence of a dual solution δ^* that is *locally decodable* at all nodes u : for each u , $\operatorname{argmin}_i \theta_u^{\delta^*}(i)$ is unique, and moreover, the edge and node dual subproblems agree:

$$\left(\operatorname{argmin}_i \theta_u^{\delta^*}(i), \operatorname{argmin}_j \theta_v^{\delta^*}(j) \right) \in \operatorname{argmin}_{i,j} \theta_{uv}^{\delta^*}(i, j). \quad (13)$$

In this case, the primal solution defined by “decoding” δ^* , $x(u) = \operatorname{argmin}_i \theta_u^{\delta^*}(i)$, is the MAP solution (Sontag et al., 2011).

For locally decodable δ^* , we define the *node margin* $\psi_u(\delta^*)$ at a node u as:

$$\psi_u(\delta^*) = \min_{i \neq \operatorname{argmin}_j \theta_v^{\delta^*}(j)} \theta_u^{\delta^*}(i) - \min_j \theta_v^{\delta^*}(j).$$

This is the difference between the optimal reparametrized node cost at u and the next-smallest cost. Local decodability of δ is the property that $\psi_u(\delta) > 0$ for every u .

Together with (13), the following lemma implies that we need only consider locally decodable dual solutions where the optimal primal solution pays zero edge cost.

Lemma C.3 (Dual edge “removal”). *Given a locally decodable dual solution δ , we can transform it to a locally decodable dual solution δ' that satisfies $\min_{i,j} \theta_{uv}^{\delta'}(i, j) = 0$ and has the same (additive) margin at every node.*

Proof. Fix an edge (u, v) , and consider any pair i^*, j^* in $\operatorname{argmin}_{i,j} \theta_{uv}^\delta(i, j)$. Put $\theta_{uv}^\delta(i^*, j^*) = \theta_{uv}^\delta(i^*, j^*) + \varepsilon$ for $\varepsilon \in \mathbb{R}$. Now define $\delta'_{uv}(i) = \delta_{uv}(i) - \varepsilon$ for all i (or, equivalently, $\delta'_{vu}(j) = \delta_{vu}(j) - \varepsilon$ for all j). Because we changed $\theta_u^\delta(i)$ by a constant for each i , local decodability is preserved and the additive *margin of local decodability* is not changed. We incurred a change of $+\varepsilon$ in the dual objective of δ from the edge term $\min_{i,j} \theta_{uv}^\delta(i, j)$, and a $-\varepsilon$ in the objective from the decrease in the node term $\min_i \theta_u^{\delta'}(i)$, so δ' is still optimal. We can repeat this process for every edge (u, v) . \square

Lemma C.3 implies that when (x^*, δ^*) is a pair of primal/dual optima and δ^* is locally decodable, we can assume that $L(x^*, \delta^*) = \sum_u \theta_u^{\delta^*}(x_u^*)$, where we overload notation to define x_u^* to be the label for which $x_u^*(i) = 1$. That is, the primal optimum pays no edge cost in the problem reparametrized by the dual opt δ^* . Finally, Lemma

C.3 implies that we can always assume that $\theta_{uv}^{\delta^*}(i, j) \geq 0$ for all (u, v) , (i, j) . Therefore, if there is any locally decodable dual solution, and the primal LP solution is integral and unique, we may assume there exists a locally decodable dual solution δ such that $\bar{x}(u) = \operatorname{argmin}_i \theta^\delta(i)$, $(\bar{x}(u), \bar{x}(v)) \in \operatorname{argmin}_{i,j} \theta_{uv}^\delta(i, j)$, $\theta_{uv}^\delta(\bar{x}(u), \bar{x}(v)) = 0$, and $\theta_{uv}^\delta(i, j) \geq 0$.

Lemma C.4 (Dual margin implies curvature around \bar{x}). *For an instance with objective θ and MAP solution \bar{x} , assume there exists a locally decodable dual solution δ such that $\bar{x}(u) = \operatorname{argmin}_i \theta^\delta(i)$, $(\bar{x}(u), \bar{x}(v)) \in \operatorname{argmin}_{i,j} \theta_{uv}^\delta(i, j)$, $\theta_{uv}^\delta(\bar{x}(u), \bar{x}(v)) = 0$, and $\theta_{uv}^\delta(i, j) \geq 0$. Additionally, let $\psi(\delta) = \min_u \psi_u(\delta)$ be the smallest node margin. Note that $\psi(\delta) > 0$ because δ is locally decodable. Then for any $x \in L(G)$,*

$$\frac{1}{2} \|x_V - \bar{x}_V\|_1 \leq \frac{\langle \theta, x - \bar{x} \rangle}{\psi(\delta)}$$

Proof. Let $\Delta = \langle \theta, x - \bar{x} \rangle$. Since x and \bar{x} are both primal-feasible, we have $L(x, \delta) = \langle \theta, x \rangle$ and $L(\bar{x}, \delta) = \langle \theta, \bar{x} \rangle$. Therefore,

$$L(x, \delta) = L(\bar{x}, \delta) + \Delta. \quad (14)$$

Because $\theta^\delta(\bar{x}(u), \bar{x}(v)) = 0$ for all (u, v) , we have

$$L(\bar{x}, \delta) = \sum_u \theta_u^\delta(\bar{x}(u)).$$

Additionally, because $\theta_{uv}^\delta(i, j) \geq 0$,

$$L(x, \delta) = \sum_u \sum_i \theta_u^\delta(i) x_u(i) + \sum_{uv} \sum_{ij} \theta_{uv}^\delta(i, j) x_{uv}(i, j) \geq \sum_u \sum_i \theta_u^\delta(i) x_u(i).$$

Combining the above two inequalities with (14) gives:

$$\sum_u \sum_i \theta_u^\delta(i) x_u(i) \leq \sum_u \theta_u^\delta(\bar{x}(u)) + \Delta \quad (15)$$

Because δ is locally decodable to \bar{x} , and the smallest node margin is equal to $\psi(\delta)$, we have that for every u , $\theta_u^\delta(\bar{x}(u)) + \psi(\delta) \leq \theta_u^\delta(i)$ for all $i \neq \bar{x}(u)$. The margin condition implies:

$$\sum_u \theta_u^\delta(\bar{x}(u)) x_u(\bar{x}(u)) + \sum_u \sum_{i \neq \bar{x}(u)} (\theta_u^\delta(\bar{x}(u)) + \psi(\delta)) x_u(i) < \sum_u \sum_i \theta_u^\delta(i) x_u(i),$$

and simplifying using $\sum_i x_u(i) = 1$ gives:

$$\sum_u \theta_u^\delta(\bar{x}(u)) + \psi(\delta) \sum_u \sum_{i \neq \bar{x}(u)} x_u(i) < \sum_u \sum_i \theta_u^\delta(i) x_u(i).$$

Plugging in to (15) gives:

$$\sum_u \sum_{i \neq \bar{x}(u)} x_u(i) < \frac{\Delta}{\psi(\delta)}.$$

The left-hand-side is precisely $\|x_V - \bar{x}_V\|_1/2$. □

Now we show that $(2, 1, \psi)$ -expansion stability implies that there exists a locally decodable dual solution δ with dual margin $\psi(\delta) \geq \psi$.

Lemma C.5 ($(2, 1, \psi)$ -expansion stability gives a lower bound on dual margin). *Let (G, c, w) be a $(2, 1, \psi)$ -expansion stable instance with $\psi > 0$. Then there exists a locally decodable dual solution δ with dual margin $\psi(\delta) \geq \psi$.*

Proof. Define new costs c_ψ as

$$c_\psi(u, i) = \begin{cases} c(u, i) + \psi & \bar{x}(u) = i \\ c(u, i) & \text{otherwise.} \end{cases}$$

By definition, the instance (G, c_ψ, w) is $(2, 1)$ -expansion stable (see Definition 4.1). Theorem 4.2 implies the pairwise LP solution is unique and integral on (G, c_ψ, w) . This implies there exists a dual solution δ^0 that is locally decodable to \bar{x} . The only guarantee on the dual margin of δ^0 is that $\psi(\delta^0) > 0$. But note that δ^0 is also an optimal dual solution for (G, c, w) , since its objective in that instance is the same as the objective of \bar{x} . But in that instance, the dual margin at every node is at least ψ , because $c_\psi(u, \bar{x}(u)) - c(u, \bar{x}(u)) = \psi$. So $\psi(\delta) \geq \psi$. \square

These two lemmas directly imply Theorem 5.2. This dual proof is slightly more general than the primal proof, since the curvature result applies to any $x \in L(G)$.

D Details for Generative model

Definition D.1 (sub-Gaussians and (b, σ) -truncated sub-Gaussians). Suppose $b \in \mathbb{R}, \sigma \in \mathbb{R}_+$. A random variable X with mean μ is sub-Gaussian with parameter σ if and only if $\mathbb{E}[e^{\lambda(X-\mu)}] \leq \exp(\lambda^2\sigma^2/2)$ for all $\lambda \in \mathbb{R}$. The random variable X is (b, σ) -truncated sub-Gaussian if and only if X is supported in (b, ∞) and X is sub-Gaussian with parameter σ .

We remark that the above definition captures many well-studied families of bounded random variables e.g., Rademacher distributions, uniform distributions on an interval etc. We remark that a bounded random variable supported on $[-M, M]$ is also sub-Gaussian with parameter M . However in our setting, it needs to be truncated only on negative side, and the bound M will be much larger than the variance parameter σ ; the bound is solely to ensure non-negativity of edge costs. A canonical example to keep in mind is a truncated Gaussian distribution. We use the following standard large deviations bound for sums of sub-Gaussian random variables (for details, refer to Thm 2.6.2 from Vershynin (2018)). Given independent r.v.s X_1, X_2, \dots, X_n , with X_i drawn from a sub-Gaussian with parameter σ_i we have for $\mu = \sum_{i=1}^n \mathbb{E}[X_i]$ and $\sigma^2 = \sum_{i=1}^n \sigma_i^2$,

$$\mathbb{P}\left[\left|\sum_{i=1}^n X_i - \mu\right| \geq t\right] \leq 2 \exp\left(-\frac{t^2}{2\sigma^2}\right). \quad (16)$$

Lemma 6.1 ($d(\hat{\theta}, \bar{\theta})$ is small w.h.p.). *There exists a universal constant $c < 1$ such that for any instance in the above model, with probability at least $1 - o(1)$,*

$$\sup_{x \in L^*(G)} |\langle \hat{\theta}, x \rangle - \langle \bar{\theta}, x \rangle| \leq c\sqrt{nk} \sqrt{\sum_{u,i} \sigma_{u,i}^2 + \frac{k^2}{4} \sum_{uv} \gamma_{u,v}^2}$$

Proof. As discussed in section B, for any $x \in L^*(G)$, the objective of the local LP can be written as

$$\sum_{u \in V} \sum_i c(u, i) x_u(i) + \sum_{(u,v) \in E} w(u, v) d(u, v)$$

where $d(u, v) := \frac{1}{2} \sum_i |x_u(i) - x_v(i)|$. Let $\hat{f}(x) := \langle \hat{\theta}, x \rangle, \bar{f}(x) := \langle \bar{\theta}, x \rangle$. Then,

$$|\langle \hat{\theta}, x \rangle - \langle \bar{\theta}, x \rangle| = |\hat{f}(x) - \bar{f}(x)| = \left| \sum_{u \in V} \sum_{i \in L} \tilde{c}(u, i) x_u(i) + \sum_{(u,v) \in E} \tilde{w}(u, v) d(u, v) \right|$$

For any feasible LP solution x , consider the following rounding algorithm \mathcal{R} :

Algorithm 2 \mathcal{R} rounding

```

1: for each  $i \in \mathcal{L}$  do
2:   Choose  $r_i \in (0, 1)$  uniformly at random.
3:   for each  $u \in V$  do
4:     if  $x_u(i) > r_i$  then
5:        $\mathcal{R}(x)_u(i) = 1$ .
6:     else
7:        $\mathcal{R}(x)_u(i) = 0$ 
8:     end if
9:   end for
10: end for
    
```

Then, we have

$$\begin{aligned}
 \mathbb{E}[\hat{f}(\mathcal{R}(x)) - \bar{f}(\mathcal{R}(x))] &= \sum_{u \in V} \sum_{i \in L} \tilde{c}(u, i) \mathbb{E}[\mathbb{1}[\mathcal{R}(x)_u(i) = 1]] + \sum_{(u,v) \in E} \frac{\tilde{w}(u,v)}{2} \sum_i \mathbb{E}[\mathbb{1}[\mathcal{R}(x)_u(i) \neq \mathcal{R}(x)_v(i)]] \\
 &= \sum_{u \in V} \sum_{i \in L} \tilde{c}(u, i) \mathbb{P}[x_u(i) > r_i] + \sum_{(u,v) \in E} \frac{\tilde{w}(u,v)}{2} \sum_i \mathbb{P}[\min(x_u(i), x_v(i)) \leq r_i < \max(x_u(i), x_v(i))] \\
 &= \sum_{u \in V} \sum_{i \in L} \tilde{c}(u, i) x_u(i) + \sum_{(u,v) \in E} \frac{\tilde{w}(u,v)}{2} \sum_i |x_u(i) - x_v(i)| \\
 &= \sum_{u \in V} \sum_{i \in L} \tilde{c}(u, i) x_u(i) + \sum_{(u,v) \in E} \tilde{w}(u,v) d(u,v) = \hat{f}(x) - \bar{f}(x)
 \end{aligned}$$

Therefore,

$$\sup_{x \in L^*(G)} |\hat{f}(x) - \bar{f}(x)| = \sup_{x \in L^*(G)} |\mathbb{E}[\hat{f}(\mathcal{R}(x)) - \bar{f}(\mathcal{R}(x))]| \leq \sup_{\hat{x}_V \in \{0,1\}^{nk}} |\hat{f}(\hat{x}_V) - \bar{f}(\hat{x}_V)|$$

Note that for all $x \in L^*(G)$, $\hat{f}(x)$ and $\bar{f}(x)$ only depend on the portion of x restricted to the vertices i.e. x_V . This is why we only need to look at $\hat{x}_V \in \{0, 1\}^{nk}$ for the last inequality.

For any fixed $\hat{x}_V \in \{0, 1\}^{nk}$, since $\tilde{w}(u, v), \tilde{c}(u, i)$ are all mean 0 and sub-Gaussian with parameters $\gamma_{u,v}, \sigma_{u,i}$, we have for any $t > 0$,

$$\mathbb{P} \left[|\hat{f}(\hat{x}_V) - \bar{f}(\hat{x}_V)| > t \right] \leq 2 \exp \left(\frac{-t^2}{2 \left(\sum_{u,i} \sigma_{u,i}^2 + k^2/4 \sum_{uv} \gamma_{u,v}^2 \right)} \right)$$

Taking $t = c\sqrt{nk} \sqrt{\sum_{u,i} \sigma_{u,i}^2 + k^2/4 \sum_{uv} \gamma_{u,v}^2}$, we get that for any fixed $\hat{x}_V \in \{0, 1\}^{nk}$,

$$\mathbb{P} \left[|\hat{f}(\hat{x}) - \bar{f}(\hat{x})| > t \right] \leq 2 \exp(-c^2 nk)$$

Taking a union bound over $\{0, 1\}^{nk}$, we get that

$$\mathbb{P} \left[\sup_{\hat{x}_V \in \{0,1\}^{nk}} |\hat{f}(\hat{x}) - \bar{f}(\hat{x})| > t \right] \leq 2 \exp(nk (\log 2 - c^2))$$

□

Here, c needs to be greater than $\sqrt{\ln 2} \approx 0.83$ to get a high probability guarantee.

Corollary 6.3 (MAP solution recovery for regular graphs). *Suppose we have a d -regular graph G with $\gamma_{u,v}^2 = \gamma^2$ for all edges (u, v) , and $\sigma_{u,i}^2 = \sigma^2$ for all vertex-label pairs (u, i) . Also, suppose only a fraction ρ of the vertices and η of the edges are subject to the noise. With high probability over the random noise,*

$$\frac{\|\hat{x}_V - \bar{x}_V\|_1}{2n} \leq \frac{2ck\sqrt{\rho\sigma^2 + \frac{\eta dk}{8}\gamma^2}}{\psi}$$

Proof. From Theorem 6.2, we have that, with high probability over the random noise

$$\frac{1}{2}\|\hat{x}_V - \bar{x}_V\|_1 \leq \frac{2}{\psi} \cdot c\sqrt{nk} \cdot \sqrt{\sum_{u,i} \sigma_{u,i}^2 + \frac{k^2}{4} \sum_{uv} \gamma_{u,v}^2}$$

In this setting, this leads to

$$\frac{2}{\psi} \cdot c\sqrt{nk} \cdot \sqrt{\sum_{u,i} \sigma_{u,i}^2 + \frac{k^2}{4} \sum_{uv} \gamma_{u,v}^2} = \frac{2}{\psi} \cdot c\sqrt{nk} \cdot \sqrt{\rho nk\sigma^2 + \eta \frac{k^2}{4} \frac{nd}{2} \gamma^2} = \frac{2cnk\sqrt{\rho\sigma^2 + \frac{\eta dk}{8}\gamma^2}}{\psi}$$

since $|V| = n$, $|L| = k$, and $|E| = \frac{nd}{2}$. □

E Algorithm for finding nearby stable instances details

Let \bar{x} be a MAP solution, and let $\mathcal{E}^{\bar{x}}$ be the set of expansions of \bar{x} . We prove that an instance is $(2, 1, \psi)$ -expansion stable if and only if $\langle \theta_{adv}, \bar{x} \rangle \leq \langle \theta_{adv}, x \rangle$ for all $x \in \mathcal{E}^{\bar{x}}$. In other words, it is sufficient to check for stability in the *adversarial* perturbation for \bar{x} . This proves that we need not check every possible perturbation when finding a $(2, 1, \psi)$ -expansion stable instance.

Claim E.1. *Let (G, c, w) be an instance of uniform metric labeling with MAP solution \bar{x} . Define:*

$$w_{adv}(u, v) = \begin{cases} \frac{1}{2}w_{uv} & \bar{x}(u) = \bar{x}(v) \\ w(u, v) & \bar{x}(u) \neq \bar{x}(v) \end{cases}$$

$$c_{adv}(u, i) = \begin{cases} c(u, i) + \psi & \bar{x}(u) = i \\ c(u, i) & \text{otherwise.} \end{cases}$$

Let θ_{adv} be the objective vector in the instance (G, c_{adv}, w_{adv}) . Then

$$\langle \theta', \bar{x} \rangle \leq \langle \theta', x \rangle$$

for all $(2, 1, \psi)$ -perturbations θ' of θ and all $x \in \mathcal{E}^{\bar{x}}$ if and only if:

$$\langle \theta_{adv}, \bar{x} \rangle \leq \langle \theta_{adv}, x \rangle$$

for all $x \in \mathcal{E}^{\bar{x}}$.

Proof. The proof is analogous to that of Claim B.1. If the instance is $(2, 1, \psi)$ -expansion stable, then $\langle \theta', x \rangle - \langle \theta', \bar{x} \rangle \geq 0$ for all $(2, 1, \psi)$ -perturbations θ' and all expansions x of \bar{x} . Because θ_{adv} is a valid $(2, 1, \psi)$ -perturbation, this gives one direction. For the other, note that if the instance is not $(2, 1, \psi)$ -expansion stable, there exists a θ' and an $x \in \mathcal{E}^{\bar{x}}$ for which $\langle \theta', x \rangle - \langle \theta', \bar{x} \rangle < 0$. A direct computation shows that $\langle \theta', x \rangle - \langle \theta', \bar{x} \rangle \geq \langle \theta_{adv}, x \rangle - \langle \theta_{adv}, \bar{x} \rangle$ for all $(2, 1, \psi)$ -perturbations θ' of θ . Then we have $\langle \theta_{adv}, x \rangle - \langle \theta_{adv}, \bar{x} \rangle < 0$. □

This claim justifies (5), which only enforces that \bar{x} is at least as good as all of its expansions in θ_{adv} . The following claim implies that there is always a feasible point of (5) that makes modifications of bounded size to c and w .

Claim E.2. Consider an instance (G, c, w) with a unique MAP solution \bar{x} . Let w' be defined as

$$w'(u, v) = \begin{cases} w(u, v) & \bar{x}(u) \neq \bar{x}(v) \\ 2w(u, v) & \bar{x}(u) = \bar{x}(v), \end{cases}$$

and let c' be defined as

$$c'(u, i) = \begin{cases} c(u, i) - \psi & \bar{x}(u) = i \\ c(u, i) & \bar{x}(u) \neq i. \end{cases}$$

Then the instance (G, c', w') is $(2, 1, \psi)$ -expansion stable with MAP solution \bar{x} .

Proof (sketch). The original MAP solution \bar{x} is also the MAP solution to (G, c', w') . Then the original instance (G, c, w) is obtained from (G, c', w') by performing the adversarial $(2, 1, \psi)$ -perturbation for \bar{x} (see Claim E.1). Because \bar{x} was the unique MAP solution to this instance, it has better objective than all of its expansions. Therefore, (G, c', w') is $(2, 1, \psi)$ -expansion stable, by Claim E.1. \square

(G, c', w') is a “nearby” stable instance to (G, c, w) , but it requires changes to quite a few edges—every edge that is not cut by \bar{x} —and changes the node costs of every vertex. Surprisingly, the stable instances we found in Section 8 were much closer than (G, c', w') —that is, only sparse changes were required to transform the observed instance (G, c, w) to a $(2, 1, \psi)$ -expansion stable instance.

F Experiment details

In this section, we give more details for the numerical examples for which we evaluate our curvature bound from Theorem 5.2. We studied instances for stereo vision, where the input is two images taken from slightly offset locations, and the desired output is the disparity of each pixel location between the two images (this disparity is inversely proportional to the depth of that pixel). We used the models from Tappen & Freeman (2003) on three images from the Middlebury stereo dataset (Scharstein & Szeliski, 2002). In this model, G is a grid graph with one node corresponding to each pixel in one of the images (say, the one taken from the left), the costs $c(u, i)$ are set using the Birchfield-Tomasi matching costs (Birchfield & Tomasi, 1998), and the edge weights $w(u, v)$ are set as:

$$w(u, v) = \begin{cases} P \times s & |I(u) - I(v)| < T \\ s & \text{otherwise.} \end{cases}$$

Here $I(u)$ is the intensity of one of the images (again, say the left image) at pixel location u , and we set $(P, T, s) = (2, 50, 4)$. This is a Potts model. The `tsukuba`, `cones`, and `venus` images were 120×150 , 125×150 , and 383×434 , respectively. These models had $k = 7$, $k = 5$, and $k = 5$, respectively.

To generate Table 1, we ran the algorithm in (5) using Gurobi (Gurobi Optimization, 2020) for the L1 distance. For each observed instance (G, \hat{c}, \hat{w}) , this output a nearby $(2, 1, \psi)$ -stable instance (G, \bar{c}, \bar{w}) . In all of our experiments, we used $\psi = 1$. Additionally, we always set the target MAP solution x^t in (5) to be equal to the observed MAP solution \hat{x}^{MAP} . To evaluate our recovery bound, we compared the objective of the observed LP solution \hat{x} to the \hat{x}^{MAP} in (G, \bar{c}, \bar{w}) . That is, if $\bar{\theta}$ is the objective for (G, \bar{c}, \bar{w}) , we computed $\langle \bar{\theta}, \hat{x} \rangle - \langle \bar{\theta}, \hat{x}^{MAP} \rangle = \langle \bar{\theta}, \hat{x} \rangle - \langle \bar{\theta}, \bar{x} \rangle$, where the second equality is because we set the target solution x^t to be equal to \hat{x}^{MAP} , so $\hat{x}^{MAP} = \bar{x}$. Because $\psi = 1$, the difference between these two objectives is precisely the value of our curvature bound. In particular, Theorem 5.2 guarantees that

$$\frac{1}{2n} \|\hat{x}_V - \hat{x}_V^{MAP}\|_1 \leq \frac{1}{n} (\langle \bar{\theta}, \hat{x} \rangle - \langle \bar{\theta}, \hat{x}^{MAP} \rangle).$$

The right-hand-side is shown for these instances in the “Recovery error bound” column of Table 1, and the true value of $\frac{1}{2n} \|\hat{x}_V - \hat{x}_V^{MAP}\|_1$ (i.e., the true recovery error) is shown in the identically titled column of Table 1. On these instances, $\frac{1}{n} (\langle \bar{\theta}, \hat{x} \rangle - \langle \bar{\theta}, \hat{x}^{MAP} \rangle)$ is close to 0, so our curvature bound “explains” a large portion of \hat{x} ’s recovery of \hat{x}^{MAP} . These instances are close to $(2, 1, \psi)$ -stable instances where \hat{x} and \hat{x}^{MAP} have close objective, and this implies by Theorem 5.2 that \hat{x} approximately recovers \hat{x}^{MAP} .

However, this result relies on a property of the LP solution \hat{x} : that it has good objective in the stable instance discovered by the procedure (5). Compare this to Corollary 5.3, which only depends on properties of the observed

instance $\hat{\theta}$ and the stable instance $\bar{\theta}$ (in particular, some notion of “distance” between them). Given an observed instance $\hat{\theta}$ and stable instance $\bar{\theta}$, we can try to compute $d(\bar{\theta}, \hat{\theta})$ from Corollary 5.3 to give a bound that does not depend on \hat{x} . Unfortunately, this distance can be large, leading to a bound that can be vacuous (i.e., normalized Hamming recovery > 1). The following refinement of Corollary 5.3 gives much tighter bounds.

Corollary F.1 (LP solution is good if there is a nearby stable instance, refined). *Let \hat{x}^{MAP} and \hat{x} be the MAP and local LP solutions to an observed instance (G, \hat{c}, \hat{w}) . Also, let \bar{x} be the MAP solution for a latent $(2, 1, \psi)$ -expansion stable instance (G, \bar{c}, \bar{w}) . If $\hat{\theta} = (\hat{c}, \hat{w})$ and $\bar{\theta} = (\bar{c}, \bar{w})$, define*

$$d(\bar{\theta}, \hat{\theta}) := \sup_{x \in L^*(G)} \langle \bar{\theta}, x \rangle - \langle \bar{\theta}, \bar{x} \rangle - (\langle \hat{\theta}, x \rangle - \langle \hat{\theta}, \bar{x} \rangle).$$

Note that while we still use the name $d(\cdot, \cdot)$, evoking a metric, d is not symmetric. Then

$$\frac{1}{2} \|\hat{x}_V - \hat{x}_V^{MAP}\|_1 \leq \frac{d(\bar{\theta}, \hat{\theta})}{\psi} + \frac{1}{2} \|\hat{x}_V^{MAP} - \bar{x}_V\|_1.$$

Proof. Note:

$$\frac{1}{2} \|\hat{x}_V - \hat{x}_V^{MAP}\|_1 \leq \frac{1}{2} \|\hat{x}_V - \bar{x}_V\|_1 + \frac{1}{2} \|\hat{x}_V^{MAP} - \bar{x}_V\|_1.$$

By the definition of d , for any $x \in L^*(G)$,

$$\langle \bar{\theta}, x \rangle - \langle \bar{\theta}, \bar{x} \rangle \leq d(\bar{\theta}, \hat{\theta}) + (\langle \hat{\theta}, x \rangle - \langle \hat{\theta}, \bar{x} \rangle).$$

Now if we set $x = \hat{x}$, the LP solution to the observed instance, we have $\langle \hat{\theta}, \hat{x} \rangle - \langle \hat{\theta}, \bar{x} \rangle \leq 0$, so

$$\langle \bar{\theta}, \hat{x} \rangle - \langle \bar{\theta}, \bar{x} \rangle \leq d(\bar{\theta}, \hat{\theta}).$$

Theorem 5.2 then implies $\frac{1}{2} \|\hat{x}_V - \bar{x}_V\|_1 \leq d(\bar{\theta}, \hat{\theta})/\psi$, which gives the claim. \square

Given an observed instance $\hat{\theta}$ and a $(2, 1, \psi)$ -expansion stable instance $\bar{\theta}$ output by (5) with $\bar{x}^{MAP} = \hat{x}^{MAP}$, we can upper bound $d(\bar{\theta}, \hat{\theta})$ by computing

$$d_{up}(\bar{\theta}, \hat{\theta}) := \sup_{x \in L(G)} \langle \bar{\theta}, x \rangle - \langle \bar{\theta}, \bar{x} \rangle - (\langle \hat{\theta}, x \rangle - \langle \hat{\theta}, \bar{x} \rangle),$$

which is a linear program in x because we relaxed $L^*(G)$ to $L(G)$. Corollary F.1 then implies that the recovery error of \hat{x} is at most $d_{up}(\bar{\theta}, \hat{\theta})/\psi$, which we can compute. Table 2 shows the results of this procedure on two of the same instances from Table 1 in the “Unconditional bound” column. While the values of this bound are much larger than the “Curvature bound” of Theorem 5.2, they are much more theoretically appealing, since they only depend on the difference between $\hat{\theta}$ and $\bar{\theta}$ rather than on a property of the LP solution \hat{x} to $\hat{\theta}$. For Table 2, we did a grid search for ψ over $\{1, \dots, 10\}$; $\psi = 4$ gave the optimal unconditional bound for both instances. The difference in ψ explains the slight differences between the other columns of Tables 1 and 2.

Table 2: Results from the output of (5) on three stereo vision instances. Curvature bound shows the bound obtained from Theorem 5.2, which depends on the observed LP solution \hat{x} . Unconditional bound shows the bound from the refined version of Corollary 5.3, which depends *only* on the observed instance and the stable instance. This “unconditional” bound explains a reasonably large fraction of the LP solution’s recovery for these instances: because the instance is close to a stable instance, the LP solution approximate recovers the MAP solution.

Instance	Costs changed	Weights changed	Curvature bound	Unconditional bound	$\ \hat{x}_V - \hat{x}_V^{MAP}\ _1/2n$
tsukuba	4.9%	2.8%	0.0173	0.4878	0.0027
cones	2.81%	2.31%	0.0137	0.2819	0.0022