Graph Cut-Based Semi-Supervised Learning via p-Conductances

Anonymous Author 1*
University of Nowhere
anonymous@nowehere.com

Anonymous Author 2*
Galactic University
anonymous@galactic.edu

Anonymous Author 3
Terra University
anonymous@terra.edu

Abstract

We study the problem of semi-supervised learning in the regime where data labels are scarce or possibly corrupted. We propose a graph-based approach called p-conductance learning that generalizes the p-Laplace and Poisson learning methods by using an affine relaxation of the label constraints. This leads to a family of probability measure mincut programs that balance sparse edge removal with accurate distribution separation and which are connected to well-known variational and probabilistic problems on graphs. Computationally, we develop a semismooth Newton-conjugate gradient algorithm and extend it to incorporate class-size estimates when converting the continuous solutions into label assignments. Empirical results on citation datasets demonstrate that our approach achieves state-of-the-art accuracy in low label-rate, corrupted-label, and partial-label regimes.

1 Introduction

Graph-based semi-supervised learning (SSL) leverages both labeled and unlabeled data to improve prediction when labeled examples are scarce. In the "low label-rate regime," affinity and adjacency among unlabeled points can significantly boost accuracy. Given a weighted graph $G = (V, E, w)$ on n nodes and an initial set of $m \ll n$ binary node labels $y_i \in \{-1, 1\}^m$, a seminal method is Laplace learning [1], which extends the known labels $\{y_i\}_{i=1}^m$ via the minimizer

$$\min_{\phi\in\mathbb{R}^n} \left\{\phi^\top L\phi : \phi_i = y_i,\ 1 \le i \le m\right\}, \tag{1}$$

where L is the graph Laplacian, with predictions $\widehat{y}_i = \operatorname{sign}(\phi_i)$. Recent work has refined this approach to address its known degeneracies and improve accuracy [2–6], e.g., Poisson Learning [3] solves

$$\min_{\phi\in\mathbb{R}^n} \phi^\top L\phi - \sum_{i=1}^{m}(y_i - \bar{y}) \cdot \phi_i. \tag{2}$$

These methods are often analyzed through boundary value problems (BVPs), random walks, and stochastic processes [7], while alternative approaches stem from graph cuts [8].

We propose an affine relaxation of the variational p-Laplacian approach, unifying graph-cut and BVP-based SSL methods within a framework of cuts between probability measures. For binary classification, given label distributions $\mu, \nu \in \mathbb{R}^n$, we solve

$$\mathcal{C}_p(\mu,\nu) = \min_{\substack{\phi\in\mathbb{R}^n\\ \phi^T(\mu-\nu)=1}} \left(\sum_{\{i,j\}\in E} w_{ij}|\phi_i - \phi_j|^p\right)^{1/p}, \tag{$\mathcal{C}_p$}$$

with the $p = \infty$ case given by

$$\mathcal{C}_\infty(\mu,\nu) = \min_{\substack{\phi\in\mathbb{R}^n\\ \phi^T(\mu-\nu)=1}} \max_{\{i,j\}\in E} w_{ij}|\phi_i - \phi_j|. \tag{$\mathcal{C}_\infty$}$$

*Equal contribution.

Anonymous et al., Graph Cut-Based Semi-Supervised Learning via p-Conductances. *Proceedings of the Fourth Learning on Graphs Conference (LoG 2025)*, PMLR 269, Hybrid Event, December 10–12, 2025.

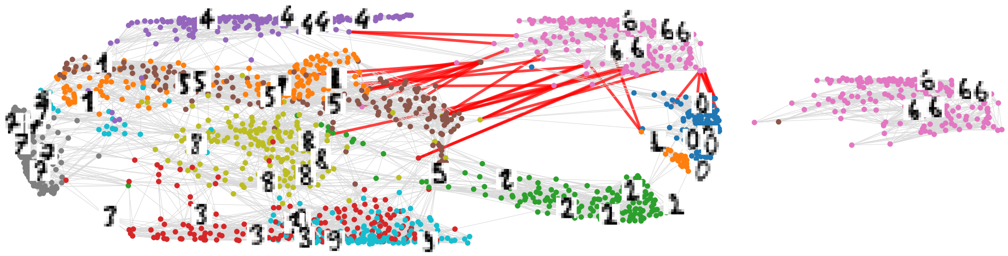

Figure 1: (Left) One-vs-all measure `mincut` on the Sklearn Digits dataset [12]. The initially labeled nodes are shown with images overlaid. μ is given by the five images of the digit six, and ν from all of the other classes. Solving the program ($\mathcal{C}_p$) for $p = 1$, we obtain a sparse solution ϕ such that each $\{i, j\} \in E$ satisfies $|\phi_i - \phi_j| \approx 0$ (light gray) or $|\phi_i - \phi_j| \approx 0.11$ (red). **(Right)** Removing red edges isolates a connected component with near-perfect class separation.

Intuitively, ϕ takes values near $1/2$ on $\operatorname{supp}(\mu)$ and $-1/2$ on $\operatorname{supp}(\nu)$ while minimizing p-Laplacian energy; predictions follow from $\widehat{y} = \operatorname{sgn}(\phi^*)$. For multi-class problems, we adopt a one-vs-all formulation. We call $\mathcal{C}_p(\mu, \nu)$ the measure p-conductance. The parameter p controls solution sparsity: concentrated for $p = 1$, more diffuse as $p \to \infty$.

From a computational perspective, related works include computation of the p-Laplacian and p-effective resistance [9], both of which involve minimizing a weighted p-norm. In [10], Newton methods with homotopy on p were suggested for computing solutions to the p-Laplace equations. Recently, iteratively reweighted least squares were studied in depth in [11]. Notably, the aforementioned methods struggle when $p \in \{1, \infty\}$. In contrast, we describe a globally superlinearly convergent method that works for any $p \in [1, \infty]$.

2 Theoretical analysis of p-Conductances

Here, $G = (V, E, w)$ is a weighted undirected graph on n nodes and N edges with symmetric edge weights w_{ij}. We denote by E' the set of oriented edges, i.e., $E' = \{(i, j), (j, i) : \{i, j\} \in E\}$. We define the degree of each node $i \in V$ by $d_i = \sum_{j=1}^{n} w_{ij}$, and the Laplacian matrix $L = D - W$ where W is the weighted adjacency matrix of G. We denote by $\mathsf{P}(V)$ the probability simplex of V. In practice $\mu, \nu \in \mathsf{P}(V)$ correspond either to sum-normalized indicators of the labelled nodes in each of two classes, or the measures obtained from the one-vs-all setup as described in Remark 3.1.

Theorem 2.1 (Generalized Max Flow - Min Cut Theorem)**.** *Assume G is connected and let $\mu, \nu \in \mathsf{P}(V)$. The program $\mathcal{C}_1(\mu, \nu)$ is realizable as either $\mathtt{mincut}(\mu, \nu)$ or $\mathtt{maxflow}(\mu, \nu)$ and strong duality holds for these programs. Concretely, the former is a linear program in the variables $k = (k_{ij}) \in \mathbb{R}^{2N}$ and $\psi \in \mathbb{R}^n$; and the latter is a linear program in the variables $J = (J_{ij}) \in \mathbb{R}^{E'}$ and $f \in \mathbb{R}$:*

$$\mathtt{mincut}(\mu,\nu) = \begin{cases} \textit{minimize} & \sum_{(i,j)\in E'} w_{ij} k_{ij} \\ \textit{subject to} & k_{ij} \geq 0 \\ & k_{ij} \geq \psi_i - \psi_j \\ & (\mu - \nu)^T \psi \geq 1 \end{cases}, \quad \mathtt{maxflow}(\mu,\nu) = \begin{cases} \textit{maximize} & f \\ \textit{subject to} & f \geq 0 \\ & J_{ij} \geq 0 \\ & J_{ij} \leq w_{ij} \\ & \widetilde{B} J = f(\mu - \nu) \end{cases}. \tag{3}$$

Here, $\widetilde{B} \in \mathbb{R}^{n \times 2N}$ is the oriented incidence matrix according to E'. Finally, strong duality holds.

When only one node is labeled in each class, $\mathtt{mincut}(\mu, \nu)$ recovers the usual node `mincut` problem. We illustrate the measure `mincut` problem on a toy dataset in Figure 1.

The p-conductance problem is also related to a family of optimal transportation metrics on $\mathsf{P}(V)$. Define the p-Beckmann metric (see [13]), denoted $\mathcal{B}_{w,p}(\mu, \nu)$, as the optimal value of the following program, where $1 \leq p < \infty$:

$$\mathcal{B}_{w,p}(\mu,\nu) = \min_{\substack{J \in \mathbb{R}^N \\ BJ = \mu - \nu}} \Big(\sum_{\{i,j\} \in E} w_{ij} |J(i,j)|^p \Big)^{1/p}, \tag{$\mathcal{B}_{w,p}$}$$

with the usual modifications when $p = \infty$. Here, B is the node-edge oriented incidence matrix of G. This class of problems was considered in, for example, [14] as a family of distances that are related to the effective resistance metric of the underlying graph. From this setup we have the following theorem relating $\mathcal{B}_{w,p}(\mu,\nu)$ to $\mathcal{C}_p$.

Theorem 2.2. *Let $\mu,\nu \in \mathsf{P}(V)$ (with $\mu \neq \nu$). Then the optimal values of $\mathcal{C}_p$ and $\mathcal{B}_{w,p}$ are related as:*

$$\mathcal{C}_p(\mu,\nu) = \begin{cases} 1/\mathcal{B}_{\infty,w^{-1}}(\mu,\nu) & \text{if } p = 1, \\ 1/\mathcal{B}_{q,w^{1-q}}(\mu,\nu) & \text{if } p \in (1,\infty) \text{ and } 1/p + 1/q = 1, \\ 1/\mathcal{B}_{1,w^{-1}}(\mu,\nu) & \text{if } p = \infty. \end{cases}$$

This leads to two corollaries relating our program to other well-known optimal transportation and resistance metrics between probability measures on graphs.

Corollary 2.3. *Let $\mu,\nu \in \mathsf{P}(V)$ with $\mu \neq \nu$. Then it holds that $\mathcal{C}_\infty(\mu,\nu) = 1/\mathcal{W}_{1,w^{-1}}(\mu,\nu)$. That, is $\mathcal{C}_\infty(\mu,\nu)$ is the reciprocal of the optimal transportation distance between μ,ν when the ground metric on V is taken to be shortest-path distance weighted by w^{-1}.*

Corollary 2.4. *Assume G is connected and let $\mu,\nu \in \mathsf{P}(V)$ (with $\mu \neq \nu$). Then it holds*

$$\mathcal{C}_2(\mu,\nu)^2 = 1/(\mu-\nu)^T L^\dagger(\mu-\nu). \tag{4}$$

A minimizer ϕ^* for $\mathcal{C}_2(\mu,\nu)$ achieving Equation (4) is given by $\phi^* = L^\dagger(\mu-\nu)/(\mu-\nu)^\top L^\dagger(\mu-\nu)$.

We note that we are also able to establish connections between $\mathcal{C}_p$ and other well-known problems including randomized cuts and normalized cuts.

3 Algorithms for general $p \in [1,\infty]$

In this section we develop a semismooth Newton augmented Lagrangian method (SSNAL) for minimizing ($\mathcal{C}_p$) for $p \in [1,\infty]$. (For $p = 2$, $\mathcal{C}_2(\mu,\nu)$ is computable from $L^\dagger$; see Corollary 2.4.)

Remark 3.1 (One-vs-all). Given measures $\mu_1,\dots,\mu_k$ corresponding to labeled data in each of k classes, we set $R_i = \mu_i - \frac{1}{k-1}\sum_{j\neq i}\mu_j$ and collect $R = [R_1,\dots,R_k]$. For $1 \le p < \infty$,

$$\bar{\mathcal{C}}_p(\mu_1,\dots,\mu_k;R)^p = \min_{\substack{\Phi\in\mathbb{R}^{n\times k} \\ \operatorname{diag}(\Phi^\top R) = \mathbf{1}_k}} \sum_{\{i,j\}\in E} w_{ij}\|\Phi_i - \Phi_j\|_p^p, \tag{5}$$

and for $p = \infty$ replace the sum by $\max_{\{i,j\}\in E} w_{ij}\|\Phi_i - \Phi_j\|_\infty$.

Introduce $u \in \mathbb{R}^N$ indexed by oriented edges and the incidence matrix $B \in \mathbb{R}^{n\times N}$. Let $s(u) = \sum_{\{i,j\}\in E} w_{ij}|u_{ij}|^p$ for $1 \le p < \infty$ (with the usual modification for $p = \infty$). We define the augmented Lagrangian (penalties $\sigma_1,\sigma_2 > 0$):

$$\begin{aligned} \mathcal{L}_\sigma(\phi,u;y,z) = {} & s(u) + \langle y, \phi^\top(\mu-\nu) - 1\rangle + \langle z, B^\top\phi - u\rangle \\ & + \tfrac{\sigma_1}{2}\|B^\top\phi - u\|^2 + \tfrac{\sigma_2}{2}\|\phi^\top(\mu-\nu) - 1\|^2. \end{aligned} \tag{6}$$

Next for fixed $(\tilde{y},\tilde{z})$, minimize

$$F(\phi,u) := \mathcal{L}_\sigma(\phi,u;\tilde{y},\tilde{z}). \tag{7}$$

Eliminating u via the proximal map gives

Algorithm 1 Semismooth augmented Lagrangian

---

while not converged **do**
  $(\phi_{k+1},u_{k+1}) \leftarrow \operatorname{argmin}_{\phi,u}\mathcal{L}_\sigma(\phi,u;z_k,y_k)$
  $z_{k+1} \leftarrow z_k + (\sigma_1)_k(B^\top\phi_{k+1} - u_{k+1})$
  $y_{k+1} \leftarrow y_k + (\sigma_2)_k(\phi^\top(\mu-\nu) - 1)$
  Update $\sigma_{k+1} \le \infty$
end while

Algorithm 2 Semismooth Newton-CG

---

while $||\nabla f(\phi)|| \le \epsilon$ **do**
  Compute $H\Delta\phi \approx -\nabla f(\phi)$.
  (Armijo rule: $r > 0$, $\sigma,\eta \in (0,1)$) Set $\zeta = \eta^m r$, m is the first integer ≥ 0 s.t.

$$f(\phi_t) - f(\phi_t + \eta^m r\Delta\phi_t) \ge \sigma\eta_t s\langle\nabla f(\phi_t),\Delta\phi_t\rangle$$

  Set $\phi_{t+1} = \phi_t + \zeta\Delta\phi$
end while

$$\bar{u} = \operatorname{prox}_{s/\sigma_1}\big(B^\top \phi + \tfrac{1}{\sigma_1}\tilde{z}\big), \qquad \bar{\phi} = \arg\min_{\phi} f(\phi) := \inf_{u} F(\phi, u),$$

with gradient $\nabla f(\phi) = B\tilde{z} + \sigma_1 B(B^\top \phi - \bar{u}) + \tilde{y}(\mu - \nu) + \sigma_2(\phi^\top(\mu - \nu) - 1)(\mu - \nu)$. The equations $\nabla f(\phi) = 0$ are solved using Newton-CG with a generalized Hessian

$$H(\phi) \in \hat{\partial}^2 f(\phi) = \sigma_1 BB^\top - \sigma_1 B\, \partial\bar{u}(\phi)\, B^\top + \sigma_2(\mu - \nu)(\mu - \nu)^\top,$$

where $\partial\bar{u}(\phi)$ is any element of the generalized Jacobian of the prox mapping.

In practice, $H(\phi)$ is sparse and (under connectivity) positive definite, so CG matrix-vector products cost $O(|E|)$, making SSNCG efficient on large graphs; outer SSNAL updates enforce feasibility of $B^\top \phi = u$ and $\phi^\top(\mu - \nu) = 1$.

The superlinear convergence of inexact ALM is ensured whenever the following criterion holds, for a summable sequence of nonnegative stopping criteria corresponding to the error of the Newton equations $\{\epsilon_k\}$.

$$\operatorname{dist}(0, \partial F(\phi_{k+1}, u_{k+1})) \le \epsilon_k / \max\{1, \sqrt{\sigma_k}\} \tag{8}$$

Theorem 3.2 (Convergence of SSNAL)**.** *Let (ϕ_k, u_k, z_k, y_k) be the sequence generated by SSNAL with stopping criterion (8). Then, the sequence $\{\phi_k\}$ is bounded and converges to the unique optimal solution of (5), and $||B^\top \phi - u||$ and $||\phi^\top(\mu - \nu) - 1||$ converge to zero.*

4 Experimental results

Our method achieves state-of-the-art accuracy on several benchmark data classification tasks, including citation networks, image datasets MNIST [15], Fashion-MNIST [16] and CIFAR-10 [17], as well as in settings where the labels have been partially corrupted. We highlight two experiments: classification accuracy on the PubMed dataset (Tab. Table 1), as well as accuracy on the CIFAR-10 dataset (Fig. Figure 2) when the labels have been perturbed to varying degrees and the input measures have been smoothed on the graph via the diffusion kernel e^{-tL} (using Poisson learning as a benchmark). This provides experimental evidence for the robustness of our method when label noise is present in practice. We also investigate the setting of partially labelled learning, which is a weakly supervised learning problem where each training instance is equipped with a *set* of candidate labels, including the ground-truth label.

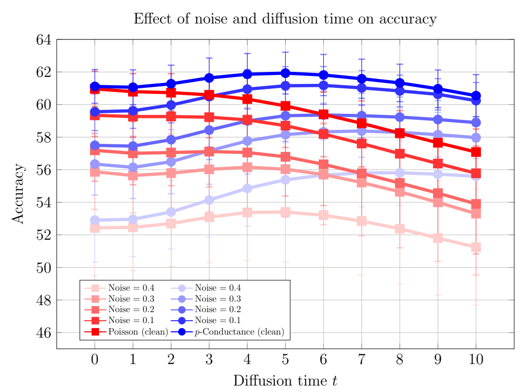

Figure 2: Accuracy on CIFAR-10 for various diffusion times. Shades denote noise level. Label rate is 10 labels per class.

Table 1: Average accuracy over 100 trials with standard deviation in brackets. Best is bolded.

PUBMED # LABELS PER CLASS	**1**	**3**	**5**	**10**	**100**
LAPLACE/LP [1]	34.6 (8.8)	35.7 (8.2)	36.9 (8.1)	39.6 (9.1)	74.9 (3.6)
SPARSE LP [18]	32.4 (4.7)	33.0 (4.8)	33.6 (4.8)	33.9 (4.8)	43.2 (4.1)
P-LAPLACE [10]	44.8 (11.2)	58.3 (9.1)	61.6 (7.7)	66.2 (4.7)	74.3 (1.1)
P-EIKONAL [2]	44.3 (11.8)	55.6 (10.0)	58.4 (9.1)	65.1 (5.8)	74.9 (1.5)
POISSON [3]	55.1 (11.3)	66.6 (7.4)	68.8 (5.6)	71.3 (2.2)	75.7 (0.8)
p-CONDUCTANCE ($p = 1, \epsilon = n$)	39.6 (0.3)	39.6 (0.3)	39.6 (0.3)	40.3 (0.3)	41.2 (0.3)
p-CONDUCTANCE ($p = 2, \epsilon = n$)	58.0 (12.1)	67.5 (7.5)	70.8 (4.9)	**72.4 (2.5)**	77.6 (0.6)
p-CONDUCTANCE ($p = \infty, \epsilon = n$)	48.0 (8.1)	56.5 (5.2)	57.0 (8.2)	62.6 (3.0)	72.9 (1.3)
POISSONMBO [3]	54.9 (11.4)	65.3 (7.8)	68.2 (5.3)	69.9 (3.0)	74.8 (1.0)
p-CONDUCTANCE ($p = 2, \epsilon = 0$)	**58.7 (11.5)**	**67.5 (7.5)**	**70.8 (4.8)**	**72.4 (2.6)**	**77.8 (0.6)**

References

[1] Xiaojin Zhu, Zoubin Ghahramani, and John Lafferty. Semi-supervised learning using Gaussian fields and harmonic functions. In *Proceedings of the Twentieth International Conference on International Conference on Machine Learning*, ICML'03, page 912–919. AAAI Press, 2003. ISBN 1577351894. 1, 4

[2] Jeff Calder and Mahmood Ettehad. Hamilton-jacobi equations on graphs with applications to semi-supervised learning and data depth. *J. Mach. Learn. Res.*, 23(1), January 2022. ISSN 1532-4435. 1, 4

[3] Jeff Calder, Brendan Cook, Matthew Thorpe, and Dejan Slepčev. Poisson learning: Graph based semi-supervised learning at very low label rates. In *Proceedings of the 37th International Conference on Machine Learning*, ICML'20. JMLR.org, 2020. 1, 4

[4] Xiaojin Jerry Zhu. Semi-supervised learning literature survey. Technical report, University of Wisconsin-Madison Department of Computer Sciences, 2005.

[5] Xueyuan Zhou and Mikhail Belkin. Semi-supervised learning by higher order regularization. In Geoffrey Gordon, David Dunson, and Miroslav Dudík, editors, *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, volume 15 of *Proceedings of Machine Learning Research*, pages 892–900, Fort Lauderdale, FL, USA, 11–13 Apr 2011. PMLR.

[6] Chester Holtz, Pengwen Chen, Zhengchao Wan, Chung-Kuan Cheng, and Gal Mishne. Continuous partitioning for graph-based semi-supervised learning. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. URL https://openreview.net/forum?id=hCOuip5Ona. 1

[7] Bruno Klaus de Aquino Afonso and Lilian Berton. Analysis of label noise in graph-based semi-supervised learning. In *Proceedings of the 35th Annual ACM Symposium on Applied Computing*, pages 1127–1134, 2020. 1

[8] Avrim Blum and Shuchi Chawla. Learning from labeled and unlabeled data using graph mincuts. 2001. 1

[9] Shota Saito and Mark Herbster. Multi-class graph clustering via approximated effective p-resistance. In *International Conference on Machine Learning*, pages 29697–29733. PMLR, 2023. 2

[10] Mauricio Flores, Jeff Calder, and Gilad Lerman. Analysis and algorithms for ℓp-based semi-supervised learning on graphs. *Applied and Computational Harmonic Analysis*, 60:77–122, 2022. ISSN 1063-5203. doi: https://doi.org/10.1016/j.acha.2022.01.004. URL https://www.sciencedirect.com/science/article/pii/S1063520322000045. 2, 4

[11] Deeksha Adil, Richard Peng, and Sushant Sachdeva. Fast, provably convergent irls algorithm for p-norm linear regression. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper_files/paper/2019/file/46c7cb50b373877fb2f8d5c4517bb969-Paper.pdf. 2

[12] Dheeru Dua and Casey Graff. UCI machine learning repository, 2017. URL http://archive.ics.uci.edu/ml. 2

[13] Martin Beckmann. A continuous model of transportation. *Econometrica: Journal of the Econometric Society*, 1952. 2

[14] Sawyer Robertson, Zhengchao Wan, and Alexander Cloninger. All you need is resistance: On the equivalence of effective resistance and certain optimal transport problems on graphs. *arXiv preprint arXiv:2404.15261*, 2024. 3

[15] Li Deng. The MNIST database of handwritten digit images for machine learning research. *IEEE Signal Processing Magazine*, 29(6):141–142, 2012. 4

[16] Han Xiao, Kashif Rasul, and Roland Vollgraf. Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. *ArXiv*, abs/1708.07747, 2017. 4

[17] Alex Krizhevsky and Geoffrey Hinton. Learning multiple layers of features from tiny images. (0), 2009. 4

[18] Alexander Jung, Alfred O. Hero III, Alexandru Mara, and Saeed Jahromi. Semi-supervised learning via sparse label propagation, 2016. 4