# GRAPH BASED SEMI-SUPERVISED LEARNING USING SPATIAL SEGREGATION THEORY 

FARID BOZORGNIA, MORTEZA FOTOUHI, AVETIK ARAKELYAN AND ABDERRAHIM ELMOATAZ


#### Abstract

In this work we address graph based semi-supervised learning using the theory of the spatial segregation of competitive systems. First, we define a discrete counterpart over connected graphs by using direct analogue of the corresponding competitive system. This model turns out doesn't have a unique solution as we expected. Nevertheless, we suggest gradient projected and regularization methods to reach some of the solutions. Then we focus on a slightly different model motivated from the recent numerical results on the spatial segregation of reaction-diffusion systems. In this case we show that the model has a unique solution and propose a novel classification algorithm based on it. Finally, we present numerical experiments showing the method is efficient and comparable to other semi-supervised learning algorithms at high and low label rates.


## 1. Introduction

In this paper we consider a semi-supervised learning approach which deals with the classification of a large number of unlabeled data when very few labels just are available. In some applications such as medical images, we have few training examples which are labeled. The aim is to find efficient algorithms with good performance with these few labeled examples. In this situation, geometric or topological properties of the unlabeled data has been used to propose and to improve several algorithms.

A common way to use the unlabeled data in semi-supervised learning is to build a graph over the data e.g., in image classification. To start, we requires to construct an adjacency matrix, or weight matrix $W$, for the data set, which encodes the similarities between pairs of date nodes. If our data set consists of $n$ points $\mathcal{X}=\left\{x_{1}, x_{2}, \cdots, x_{n}\right\} \subset \mathbb{R}^{d}$, then the weight matrix $W$ is an $n \times n$ symmetric matrix, where the element $w_{i j}$ represents the similarity between two data points $x_{i}$ and $x_{j}$. The similarity is always nonnegative and should be large when $x_{i}$ and $x_{j}$ are close together spatially, and small (or zero), when $x_{i}$ and $x_{j}$ are far apart. As a choice, the weight matrix can be the Gaussian weights

$$
w_{i j}=\exp \left(-\frac{\left|x_{i}-x_{j}\right|}{2 \sigma^{2}}\right),
$$

where $|\cdot|$ is the Euclidean norm and $\sigma$ is a free parameter that controls the scale at which points are connected by strong similarities $w_{i j}$ in the weight matrix. In fact, the weight matrix $W$ provides the data set with a graph structure, where each pair of points $\left(x_{i}, x_{j}\right)$ is connected by

[^0]an edge with edge weight $w_{i j}$. Other choices of weight matrix are possible, such as the $k$-nearest neighbor graph, see [26].

Recently, many works aim to transpose and adapt Partial Differential Equations (PDEs) on graphs. This reformulation of continuous problems onto a graph is such that the solution behaves analogously to the continuous formulation see [6, 15].

In graph-based semi-supervised learning, we are given a few labeled data on the graph and we aim to extend these labels from a given set to the rest of the nodes in graph in a decisive manner. To model propagating labels in semi-supervised learning, it is assumed that the learned labels vary smoothly and not change fast within high density regions of the graph (smoothness assumptions). Based on this assumption different approaches have been proposed. One of the pioneer methods is Laplace learning, [27]. Later it has been observed that the Laplace learning can give poor results in classification [23]. The results are often poor because the solutions have localized spikes near the labeled points, while being almost constant far from them. To overcome this problem several versions of Laplace learning algorithm have been proposed, for instance Laplacian regularization, [1], weighted Laplacian, [25, 13] and $p$-Laplace learning, [11, 24]. Also, the limiting case in $p$-Laplacian when $p$ tends to infinity is so called Lipschitz learning is studied in [22] and similar to continuum PDEs is related to finding the absolutely minimal lipschitz extension of the training data. Recently, in [10] another approach to increase accuracy of Laplace learning is given and called Poisson learning.

To explain these methods, let as before $X=\left\{x_{1}, \cdots, x_{n}\right\}$ denotes the data points or equivalently vertices in a graph. We assume there is a subset of the nodes $\Gamma=\left\{x_{1}, \cdots, x_{m}\right\} \subset \mathcal{X}$ that their labels are given with a label function $g: \Gamma \rightarrow \mathbb{R}^{k}$. It is further assumed that $y_{i}=g\left(x_{i}\right) \in\left\{e_{1}, \cdots, e_{k}\right\}$ where $e_{i}$ is the standard basis in $\mathbb{R}^{k}$ and represents the $i^{\text {th }}$ class. In graph-based semi-supervised learning, we aim to extend labels to the rest of the vertices $\left\{x_{m+1}, \cdots, x_{n}\right\}$.

In Laplace learning algorithm the labels are extended by finding the minimizer $\mathbf{u}: X \rightarrow \mathbb{R}^{k}$ for the following problem

$$
\left\{\begin{array}{l}
\min J_{n}(\mathbf{u}):=\frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{i j}\left|\mathbf{u}\left(x_{i}\right)-\mathbf{u}\left(x_{j}\right)\right|^{2}  \tag{1}\\
\text { subject to } \mathbf{u}\left(x_{i}\right)=y_{i}, \quad \text { for } i=1,2, \cdots, m .
\end{array}\right.
$$

The minimizer will be a harmonic function satisfying

$$
\begin{cases}\mathcal{L} \mathbf{u}(x)=0, & x \in \mathcal{X} \backslash \Gamma, \\ \mathbf{u}=g, & \text { on } \Gamma,\end{cases}
$$

where $\mathcal{L}$ is the unnormalized graph Laplacian given by

$$
\mathcal{L} \mathbf{u}\left(x_{i}\right)=\sum_{j=1}^{n} w_{i j}\left(\mathbf{u}\left(x_{i}\right)-\mathbf{u}\left(x_{j}\right)\right) .
$$

Let $\mathbf{u}=\left(u_{1}, \cdots, u_{k}\right)$ be a solution of (1), the label of node $x_{i} \in \mathcal{X} \backslash \Gamma$ is dictated by

$$
\underset{j \in\{1, \cdots, k\}}{\arg \max } u_{j}\left(x_{i}\right) .
$$

In $p$-Laplacian algorithm, the object function is replaced with

$$
\min _{\mathbf{u} \in \mathcal{K}} J_{p}(\mathbf{u}):=\sum_{i=1}^{n} \sum_{j=1}^{n} w_{i j}\left|\mathbf{u}\left(x_{i}\right)-\mathbf{u}\left(x_{j}\right)\right|^{p},
$$

or for weighted Laplacian the following object is considered

$$
\min _{\mathbf{u} \in \mathcal{K}} J_{\gamma}(\mathbf{u}):=\sum_{i=1}^{n} \sum_{j=1}^{n} \gamma\left(x_{i}\right) w_{i j}\left|\mathbf{u}\left(x_{i}\right)-\mathbf{u}\left(x_{j}\right)\right|^{2},
$$

where $\gamma(x) \approx \operatorname{dist}(x, \Gamma)^{-\alpha}$ increases the weights of edges adjacent to labels much more than other edges. Using this method encourages the label functions to be flat (more regular) near labels, thus preventing the appearance of spikes (discontinuous solutions). (see [25, 13] for more details).

The authors in 10 have proposed a scheme, called Poisson learning that replaces the label values at training points as sources and sinks, and solves the Poisson equation on the graph as follows:

$$
\mathcal{L} \mathbf{u}\left(x_{i}\right)=\sum_{j=1}^{m}\left(y_{j}-\bar{y}\right) \delta_{i j}, \quad i=1, \cdots, n,
$$

with further condition $\sum_{i=1}^{n} d\left(x_{i}\right) \mathbf{u}\left(x_{i}\right)=0$, where $\bar{y}=\frac{1}{m} \sum_{i=1}^{m} y_{i}$ is the average label vector, $\delta_{i j}$ is Kronecker delta and $d\left(x_{i}\right)=\sum_{j=1}^{n} w_{i j}$ is the degree of vertex $x_{i}$.

A major topic in this strand concerns the continuum limits of these PDEs or functional on graphs, linking between the discrete and continuum perspectives and the study of the consistency of the above methods in the large data limit, we refer the reader to e.g. [13, 17, 19].

Let $x_{1}, x_{2}, \cdots, x_{n}$ be a sequence of independent and identically distributed random variables on $\Omega$ with smooth distribution $\rho(x)$. Define the weight matrix

$$
w_{i j}=\eta_{\varepsilon}\left(\left|x_{i}-x_{j}\right|\right)=\frac{1}{\varepsilon^{d}} \eta\left(\frac{\left|x_{i}-x_{j}\right|}{\varepsilon}\right)
$$

where $\eta$ is a radial kernel $\eta:[0, \infty) \rightarrow[0, \infty)$ which is nonincreasing, continuous at 0 and given by

$$
\eta(t)= \begin{cases}1 & 0 \leq t \leq 1 \\ 0 & t>2\end{cases}
$$

In [16] it has been shown that for $u$ sufficiently smooth, with probability one

$$
\frac{1}{\varepsilon^{2} n^{2}} J_{n}(\mathbf{u}) \rightarrow \sum_{i=1}^{k} \int_{\Omega}\left|\nabla u_{i}(x)\right|^{2} \rho(x) d x:=J_{\infty}(\mathbf{u})
$$

where $\rho$ is density function of a probability measure that data points are generated.
In this paper, we propose a novel classification scheme based on the segregation model. Our motivation for the current work is based on properties of a class of reaction diffusion system with highly comparative rate which yields segregation of different components which means at each point in the domain different components can not coexist. In this model, we solve problem (1) with additional constraint

$$
u_{i}(x) \cdot u_{j}(x)=0, \quad \text { for all } x \in X, 1 \leq i \neq j \leq k
$$

The continuous form corresponding to the segregation model has been studied extensively, for instance [9, 8, 4]. We state the results related to limiting configuration of the following coupled system as parameter $\varepsilon$ tends to zero.

$$
\begin{cases}\Delta u_{i}^{\varepsilon}=\frac{u_{i}^{\varepsilon}}{\varepsilon} \sum_{j \neq i} u_{j}^{\varepsilon}(x) & \text { in } \Omega  \tag{2}\\ u_{i}^{\varepsilon} \geq 0 & \text { in } \Omega \\ u_{i}^{\varepsilon}(x)=\phi_{i}(x) & \text { on } \partial \Omega\end{cases}
$$

for $i=1, \cdots, m$. The boundary values satisfy

$$
\phi_{i}(x) \cdot \phi_{j}(x)=0, \quad i \neq j \text { on the boundary. }
$$

First, for each fixed $\varepsilon$ the exist unique positive solution $\left(u_{1}^{\varepsilon}, \cdots, u_{m}^{\varepsilon}\right)$. Next, to explain the asymptotic behaviour of 2 by construction barrier functions, one can show that the normal derivative of $u_{i}^{\varepsilon}$ is bounded independent of $\varepsilon$, this consequently proves that the $H^{1}$-norm of $u_{i}^{\varepsilon}$ is bounded. Next integrating the equation in 2 over $\Omega$ indicates

$$
\left(u_{i}^{\varepsilon} \sum_{j \neq i} u_{j}^{\varepsilon}(x)\right) \rightarrow 0 \text { as } \varepsilon \text { tends to zero. }
$$

Let $\left(u_{1}, \cdots, u_{m}\right)$ be the limiting configuration, then the results in [9] shows that $u_{i}$ are pairwise segregated, i.e., $u_{i}(x) \cdot u_{j}(x)=0$, harmonic in their supports and satisfy the following differential inequalities

- $-\Delta u_{i} \geq 0$,
- $-\Delta\left(u_{i}-\sum_{j \neq i} u_{j}\right) \leq 0$,
- Let $x$ belongs to interface then

$$
\lim _{y \rightarrow x} \nabla u_{i}(y)=-\lim _{y \rightarrow x} \nabla u_{j}(y) \quad \text { Free boundary condition. }
$$

In [14], it has been shown that the limiting solution of (2) minimizes the following functional

$$
\left\{\begin{array}{l}
\min J(\mathbf{u}):=\frac{1}{2} \int_{\Omega} \sum_{i=1}^{k}\left|\nabla u_{i}\right|^{2} d x  \tag{3}\\
\text { subject to } u_{i}=\phi_{i}, \quad \text { on } \partial \Omega \\
\quad u_{i} \geq 0, \quad \text { and } \quad u_{i} \cdot u_{j}=0 \quad \text { in } \Omega
\end{array}\right.
$$

In [8, 7, 2, 3, 4, , the authors have proposed and analyzed the following numerical scheme to solving limiting configuration of system (2) and (3)

$$
u_{i}^{t+1}(x)=\max \left(\bar{u}_{i}^{t}(x)-\sum_{j \neq i} \bar{u}_{j}^{t}(x), 0\right)
$$

where $\bar{v}(x)$ denotes the average of values of $v$ of neighbors of mesh point $x$, and $t$ refers to iterations.

## 2. Calculus on graphs and setting the problem

This section is devoted to review some facts about the calculus on graphs and setting our problem. Let $\mathcal{X}=\left\{x_{1}, \cdots, x_{n}\right\}$ denote the vertices of a graph with the symmetric edge weight $w_{x y}$ between $x, y \in \mathcal{X}$. The degree of a vertex $x$ is given by $d(x)=\sum_{y \in X} w_{x y}$. Let $\ell^{2}(\mathcal{X})$ denote the set of functions $u: X \rightarrow \mathbb{R}$ equipped with the inner product

$$
(u, v)=\sum_{x \in X} u(x) v(x)
$$

for functions $u, v: X \rightarrow \mathbb{R}$. We also define a vector field on the graph to be an antisymmetric function $V: X \times X \rightarrow \mathbb{R}^{2}$, i.e. $V(x, y)=-V(y, x)$ and denote the space of all vector fields by $\ell^{2}\left(X^{2}\right)$. The gradient of a function $u \in \ell^{2}(X)$ is the vector field

$$
\nabla u(x, y)=u(y)-u(x)
$$

For two vector fields $V_{1}, V_{2}$ the inner product is

$$
\left(V_{1}, V_{2}\right)_{\ell^{2}\left(x^{2}\right)}=\frac{1}{2} \sum_{x, y \in X} w_{x y} V_{1}(x, y) V_{2}(x, y)
$$

so the norm of vector field $V$ is $\|V\|_{\ell^{2}\left(X^{2}\right)}=\sqrt{(V, V)_{\ell^{2}\left(X^{2}\right)}}$. The graph divergence of a vector field $V$ is defined by

$$
\operatorname{div} V(x)=\sum_{y \in X} w_{x y} V(x, y)
$$

which satisfies the divergence formula

$$
(\nabla u, V)_{\ell^{2}\left(X^{2}\right)}=-(u, \operatorname{div} V)
$$

The unnormalized graph laplacian $\mathcal{L}$ of a function $u \in \ell^{2}(X)$ is defined as

$$
\mathcal{L} u(x):=-\operatorname{div}(\nabla u)(x)=\sum_{y \in \mathcal{X}} w_{x y}(u(x)-u(y))
$$

The operator $\mathcal{L}$ satisfies

$$
\begin{equation*}
(\mathcal{L} u, v)=(\nabla u, \nabla v)_{\ell^{2}\left(X^{2}\right)} \tag{4}
\end{equation*}
$$

In Appendix, we revisit some important tools for PDE on graphs, such as Poincaré inequality and maximum principle.

We consider a subset of the nodes $\Gamma \subset X$ as the boundary of the graph and define the admissible set

$$
\mathcal{K}:=\left\{\mathbf{u}=\left(u_{1}, \cdots, u_{k}\right) \in\left(\ell^{2}(X)\right)^{k}: u_{i}=\phi_{i} \text { on } \Gamma \text { for } i=1, \ldots, k\right\}
$$

where the boundary data $\phi_{i}$ are known and satisfy the following assumption

$$
\begin{equation*}
\phi_{i} \in\{0,1\} \tag{5}
\end{equation*}
$$

We are going to solve the optimization problem

$$
\begin{align*}
& \min _{\mathbf{u} \in \mathcal{K}} J(\mathbf{u}):=\|\nabla \mathbf{u}\|^{2}=\sum_{i=1}^{k}\left\|\nabla u_{i}\right\|_{\ell^{2}\left(x^{2}\right)}^{2} \\
& \text { subject to: }  \tag{6}\\
& u_{i}(x) \geq 0, \forall x \in X,  \tag{7}\\
& u_{i}(x) \cdot u_{j}(x)=0 \quad \forall x \in X \text { and } i \neq j .
\end{align*}
$$

The following theorem states the existence of solution to problem (6) and describes some properties of the solution.

Theorem 2.1. Problem (6) has a solution. Moreover, the solution satisfies
(i) $\mathcal{L} u_{i}(x) \leq 0$, if $u_{i}(x)=0$, and $x \in X$.
(ii) $\mathcal{L} u_{i}(x)=0$, if $u_{i}(x)>0$, and $x \in X \backslash \Gamma$.
(iii) For every $x \in \mathcal{X} \backslash \Gamma$, there is one component $u_{i}$ such that $u_{i}(x)>0$.

Proof. Consider a minimizing sequence $\mathbf{u}^{n} \in \mathcal{K}$ for problem (6). By Poincaré inequality, Proposition 8.1, we obtain that

$$
\left\|u_{i}^{n}\right\| \leq \frac{1}{\lambda_{1}}\left\|\nabla\left(u_{i}^{n}-\phi_{i}\right)\right\|+\left\|\phi_{i}\right\| \leq \frac{1}{\lambda_{1}}\left(\left\|\nabla u_{i}^{n}\right\|+\left\|\nabla \phi_{i}\right\|\right)+\left\|\phi_{i}\right\| .
$$

Thus for every $i=1, \cdots, k$ the sequence $\left\{u_{i}^{n}\right\}$ is bounded. Hence, there exists a subsequence such that for every components $i$

$$
u_{i}^{n_{j}} \rightarrow u_{i} .
$$

It is obvious that $\mathbf{u}=\left(u_{1}, \cdots, u_{k}\right)$ satisfies the constraints in (6) and is a minimizer.
(i) To prove this part of theorem, notice that if $u_{i}(x)=0$ for some $x \in \mathcal{X}$, then

$$
\mathcal{L} u_{i}(x)=\sum_{y \in X} w_{x y}\left(u_{i}(x)-u_{i}(y)\right)=-\sum_{y \in X} w_{x y} u_{i}(y) \leq 0 .
$$

(ii) Now consider the case $u_{i}(x)>0$ for some fixed node $x \in \mathcal{X} \backslash \Gamma$. Let us define

$$
v_{i}=u_{i}+t \delta_{x}, \quad v_{j}=u_{j} \text { when } j \neq i,
$$

where $\delta_{x}$ is delta function which is $\delta_{x}(y)=0$ for every $y \neq x$ and $\delta_{x}(x)=1$. We also consider some values of $t$ such that $v_{i}(x) \geq 0,(t$ can be negative). Obviously, $\mathbf{v} \in \mathcal{K}$ and satisfies the constraints in (6). Therefore,

$$
\begin{aligned}
0 & \leq\|\nabla \mathbf{v}\|^{2}-\|\nabla \mathbf{u}\|^{2}=\left\|\nabla v_{i}\right\|^{2}-\left\|\nabla u_{i}\right\|^{2} \\
& =\frac{1}{2} \sum_{y, z \in X} w_{y z}\left(\left(v_{i}(z)-v_{i}(y)\right)^{2}-\left(u_{i}(z)-u_{i}(y)\right)^{2}\right) \\
& =\sum_{y \in X} w_{x y}\left(t^{2}+2 t\left(u_{i}(x)-u_{i}(y)\right)\right) \\
& =t^{2} d(x)+2 t \mathcal{L} u_{i}(x) .
\end{aligned}
$$

Since $u_{i}(x)>0$, so parameter $t$ can be take some negative values and when $t \rightarrow 0^{ \pm}$we conclude that $\mathcal{L} u_{i}(x)=0$.
(iii) Let $A:=\left\{x \in \mathcal{X} \backslash \Gamma: u_{1}(x)=\cdots=u_{k}(x)=0\right\}$. We claim that $A=\emptyset$. Otherwise, there is some $x \in A$ such that $w_{x y} \neq 0$ and $u_{i}(y)>0$ for some $i \in\{1, \cdots, k\}$. Thus

$$
\mathcal{L} u_{i}(x)=-\sum_{z \in X} w_{x z} u_{i}(z) \leq-w_{x y} u_{i}(y)<0 .
$$

Now choose the competitor $\mathbf{v}$ with

$$
v_{i}=u_{i}+t \delta_{x}, \quad v_{j}=u_{j} \text { when } j \neq i,
$$

for some $t \geq 0$ and repeat the calculation in the previous part to get

$$
0 \leq\|\nabla \mathbf{v}\|^{2}-\|\nabla \mathbf{u}\|^{2}=t^{2} d(x)+2 t \mathcal{L} u_{i}(x) .
$$

Since $\mathcal{L} u_{i}(x)<0$, we can choose small value for $t$ to arrive at a contradiction.
Remark 2.2. Problem (6) has not necessary a unique solution. For example, in a symmetry model, there are different choices for classification. In a toy example, consider a graph with four vertices $A, B, C$ and $D$. Let $w_{A B}=w_{B C}=w_{C D}=w_{A D}=1$ and $w_{A C}=w_{B D}=0$. Also, $A$ and $C$ are boundary points with boundary data $u_{1}(A)=u_{2}(C)=1$ and $u_{1}(C)=u_{2}(A)=0$. This problem has four solutions
(i) $u_{1}(B)=u_{1}(D)=\frac{1}{2}, \quad u_{2}(B)=u_{2}(D)=0$,
(ii) $u_{2}(B)=u_{2}(D)=\frac{1}{2}, \quad u_{1}(B)=u_{1}(D)=0$,
(iii) $u_{1}(B)=u_{2}(D)=\frac{1}{2}, \quad u_{2}(B)=u_{1}(D)=0$,
(iv) $u_{1}(D)=u_{2}(B)=\frac{1}{2}, \quad u_{1}(B)=u_{2}(D)=0$.

## 3. Gradient projection method

Gradient projection is one method that we use to solve the problem (6). In the sequel, we use the following notation for averaging of a function

$$
\mathcal{A} u(x):=\frac{1}{d(x)} \sum_{y \in X} w_{x y} u(y),
$$

where

$$
d(x)=\sum_{y \in x} w_{x y},
$$

and the admissible set

$$
\mathcal{S}:=\left\{\mathbf{u}=\left(u_{1}, \cdots, u_{k}\right) \in\left(\ell^{2}(X)\right)^{k}: u_{i}=\phi_{i} \text { on } \Gamma, u_{i} \geq 0, u_{i} \cdot u_{j}=0 \text { for } i \neq j\right\} .
$$

We also use the projection $\mathcal{P}:\left(\ell^{2}(X)\right)^{k} \rightarrow \mathcal{S}$ which is defined as follows. For every $\mathbf{v} \in\left(\ell^{2}(X)\right)^{k}$ and every $x \in \mathcal{X} \backslash \Gamma$, first we find

$$
i_{x}:=\arg \max _{1 \leq j \leq k}\left(v_{j}(x)\right)^{+}
$$

and if it has more than one solution we choose the smallest index. $\left(v^{+}:=\max (v, 0)\right.$.) Then we define $\mathbf{u}:=\mathcal{P} \mathbf{v}$ with $u_{i_{x}}(x)=\left(v_{i_{x}}(x)\right)^{+}$and $u_{j}(x)=0$ for $j \neq i_{x}$. For any $x \in \Gamma$, we obviously define $u_{i}(x)=\phi_{i}(x)$.

The following lemma shows why $\mathcal{P}$ is a projection.

Lemma 3.1. Consider $\mathbf{v} \in\left(\ell^{2}(X)\right)^{k}$, then

$$
\|\mathbf{v}-\mathcal{P} \mathbf{v}\| \leq\|\mathbf{v}-\mathbf{w}\|, \quad \text { for all } \mathbf{w} \in \mathcal{S}
$$

Proof. Consider $\mathbf{w} \in \mathcal{S}$ and define the index function $\sigma: \mathcal{X} \rightarrow\{1, \cdots, k\}$ such that $w_{j}(x)=0$ for $j \neq \sigma(x)$. So,

$$
\begin{aligned}
\|\mathbf{v}-\mathbf{w}\|^{2} & =\sum_{x \in X}\left(\left(v_{\sigma(x)}(x)-w_{\sigma(x)}(x)\right)^{2}+\sum_{i \neq \sigma(x)}\left(v_{i}(x)\right)^{2}\right) \\
& =\sum_{x \in X} \sum_{i=1}^{k}\left(v_{i}(x)\right)^{2}+\sum_{x \in X}\left(\left(v_{\sigma(x)}(x)-w_{\sigma(x)}(x)\right)^{2}-\left(v_{\sigma(x)}(x)\right)^{2}\right)
\end{aligned}
$$

Similarly we have,

$$
\|\mathbf{v}-\mathcal{P} \mathbf{v}\|^{2}=\sum_{x \in X} \sum_{i=1}^{k}\left(v_{i}(x)\right)^{2}+\sum_{x \in X}\left(\left(v_{i_{x}}(x)-\left(v_{i_{x}}(x)\right)^{+}\right)^{2}-\left(v_{i_{x}}(x)\right)^{2}\right) .
$$

It is enough to show

$$
\begin{equation*}
\left(v_{i_{x}}(x)-\left(v_{i_{x}}(x)\right)^{+}\right)^{2}-\left(v_{i_{x}}(x)\right)^{2} \leq\left(v_{\sigma(x)}(x)-w_{\sigma(x)}(x)\right)^{2}-\left(v_{\sigma(x)}(x)\right)^{2} \tag{8}
\end{equation*}
$$

for every $x \in \mathcal{X}$. If $v_{i_{x}}(x) \leq 0$, then $v_{\sigma(x)}(x) \leq 0$ by the definition of $i_{x}$. Thus the left hand side of (8) is zero as well as the right hand side is positive (recall that $w_{\sigma(x)}(x) \geq 0$ ). If $v_{i_{x}}(x) \geq 0 \geq v_{\sigma(x)}(x)$, the left hand side of (8) is negative and the right hand side will be positive. If $v_{i_{x}}(x) \geq v_{\sigma(x)}(x) \geq 0$, then (8) will hold trivially.

Our algorithm according to the gradient projection method is as follows:
(1) Choose an initial guess in $\mathcal{S}$. It might be an extension of boundary data as $u_{i, 0}=\phi_{i}$ on $\Gamma$ and $u_{i, 0}=0$ in $X \backslash \Gamma$.
(2) For $t \geq 0$, calculate the gradient of the cost function $J$ at $\mathbf{u}^{t}=\left(u_{1, t}, \cdots, u_{k, t}\right)$. It is equal to

$$
\delta J\left(\mathbf{u}^{t}\right):=\left(\mathcal{L} u_{1, t}, \cdots, \mathcal{L} u_{k, t}\right) .
$$

(3) Update the value of each components by

$$
\mathbf{u}^{t+1}:=\mathcal{P}\left(\mathbf{u}^{t}-\frac{1}{d} \mathcal{L} \mathbf{u}^{t}\right)=\mathcal{P}\left(\mathcal{A} \mathbf{u}^{t}\right)
$$

(4) If $\left\|\mathbf{u}^{t+1}-\mathbf{u}^{t}\right\|$ is small then stop the algorithm, otherwise set $t=t+1$ and iterate the previous steps.
The following proposition proves why the algorithm works.
Proposition 3.2. Assume $\mathbf{u}$ is a solution of problem (6). Consider an arbitrary point $x \in X \backslash \Gamma$ such that $u_{i}(x)>0$, then

$$
u_{i}(x)=\mathcal{A} u_{i}(x) \geq \mathcal{A} u_{j}(x), \quad \text { for all } j \neq i
$$

Proof. For a fixed index $j \neq i$, define a competitor $\mathbf{v} \in \mathcal{S}$

$$
v_{i}:=u_{i}-u_{i}(x) \delta_{x}, \quad v_{j}:=u_{j}+u_{i}(x) \delta_{x}, \quad v_{\ell}:=u_{\ell}, \text { for } \ell \neq i, j
$$

Therefore,

$$
\begin{aligned}
0 & \leq\|\nabla \mathbf{v}\|^{2}-\|\nabla \mathbf{u}\|^{2}=\left\|\nabla v_{i}\right\|^{2}+\left\|\nabla v_{j}\right\|^{2}-\left\|\nabla u_{i}\right\|^{2}-\left\|\nabla u_{j}\right\|^{2} \\
& =\sum_{y} w_{x y}\left(\left(v_{i}(x)-v_{i}(y)\right)^{2}+\left(v_{j}(x)-v_{j}(y)\right)^{2}-\left(u_{i}(x)-u_{i}(y)\right)^{2}-\left(u_{j}(x)-u_{j}(y)\right)^{2}\right) \\
& =\sum_{y \neq x} w_{x y}\left(u_{i}(y)^{2}+\left(u_{i}(x)-u_{j}(y)\right)^{2}-\left(u_{i}(x)-u_{i}(y)\right)^{2}-u_{j}(y)^{2}\right) \\
& =\sum_{y \neq x} 2 w_{x y} u_{i}(x)\left(u_{i}(y)-u_{j}(y)\right)
\end{aligned}
$$

Since $u_{i}(x)>0$, we get

$$
\frac{1}{d(x)} \sum_{y} w_{x y} u_{i}(y) \geq \frac{1}{d(x)} \sum_{y} w_{x y} u_{j}(y)=\mathcal{A} u_{j}(x)
$$

Now apply result (ii) of Theorem 2.1, we obtain that $u_{i}(x)=\frac{1}{d(x)} \sum_{y} w_{x y} u_{i}(y)$.
Define the map $\mathcal{G}:\left(\ell_{+}^{2}(\mathcal{X})\right)^{k} \longrightarrow \mathcal{S}$ with rule $\mathcal{G} \mathbf{u}=\mathbf{v}$, where

$$
v_{i}=\max \left(u_{i}-\sum_{j \neq i} u_{j}, 0\right)
$$

and $\ell_{+}^{2}(X)$ is the set of nonnegative functions. If we replace this map instead of projection $\mathcal{P}$ in the gradient projection algorithm, we will obtain the segregation method. We will study this method in Section 5 ,

Proposition 3.3. Suppose $\mathcal{P}$ is the projection on $\mathcal{S}$ defined in Section 3. Then there is a positive constant $C_{0}$ such that

$$
\|\mathcal{P} \mathbf{u}-\mathbf{u}\| \leq\|\mathcal{G} \mathbf{u}-\mathbf{u}\| \leq C_{0}\|\mathcal{P} \mathbf{u}-\mathbf{u}\|
$$

for every $\mathbf{u} \in\left(\ell_{+}^{2}(X)\right)^{k}$.
Proof. The left inequality will be hold according to Lemma 3.1. For a fixed node $x \in X$, we need to show

$$
\begin{equation*}
\sum_{j=1}^{k}\left|v_{j}(x)-u_{j}(x)\right|^{2} \leq C_{0} \sum_{j=1}^{k}\left|w_{j}(x)-u_{j}(x)\right|^{2} \tag{9}
\end{equation*}
$$

where $\mathcal{G} \mathbf{u}=\mathbf{v}$ and $\mathcal{P} \mathbf{u}=\mathbf{w}$. Let $i:=\arg \max _{1 \leq j \leq k} u_{j}(x)$. If there is more than one index for $i$, then $\mathbf{v}(x)=\mathcal{G}(\mathbf{u})(x)=0$. Thus (9) holds for $C_{0} \geq 2$.

Now assume that $i$ is the unique solution of $i:=\arg \max _{1 \leq j \leq k} u_{j}(x)$. Hence,

$$
w_{i}(x)=u_{i}(x), \text { and } w_{j}(x)=v_{j}(x)=0 \text { for } j \neq i
$$

Therefore, using the definition of $v_{i}$ and applying Cauchy-Schwartz inequality we obtain

$$
\begin{aligned}
\sum_{j=1}^{k}\left|v_{j}(x)-u_{j}(x)\right|^{2} & =\left(\sum_{j \neq i} u_{j}(x)\right)^{2}+\sum_{j \neq i}\left|u_{j}(x)\right|^{2} \\
& \leq((k-1)+1) \sum_{j \neq i}\left|u_{j}(x)\right|^{2}=k \sum_{j=1}^{k}\left|w_{j}(x)-u_{j}(x)\right|^{2}
\end{aligned}
$$

which implies that (9) holds for $C_{0} \geq k$.

## 4. Penalization method

In this section, we apply the penalization method to solve problem (6). Since finding the solution directly is not efficient (the optimization problem (6) is a problem with $(n-m)^{k}$ parameters), we would prefer to solve a PDE instead. In this case, we can just find a PDE for points that $u_{i}>0$ and this subdomain is not known. In fact, we have a free boundary problem and if we know the domain $\left\{u_{i}>0\right\}$, we are able to find the solution. In order to overcome this difficulty, we relax the constraint with a penalty term and try to estimate the solution for the original problem (6).

Indeed, we consider the following problem

$$
\begin{equation*}
\min _{\mathbf{u} \in \mathcal{K}} J_{\varepsilon}(\mathbf{u}):=\sum_{i=1}^{k}\left\|\nabla u_{i}\right\|_{\ell^{2}\left(X^{2}\right)}^{2}+\frac{1}{\varepsilon} \sum_{i \neq j}\left(u_{i}^{2}, u_{j}^{2}\right) \tag{10}
\end{equation*}
$$

Since the energy function is convex, it is straightforward that the problem has a unique solution which satisfies

$$
\begin{equation*}
\mathcal{L} u_{i}+\frac{u_{i}}{\varepsilon} \sum_{j \neq i} u_{j}^{2}=0, \quad \text { in } X \backslash \Gamma \tag{11}
\end{equation*}
$$

Furthermore, we know that the solution is nonnegative due to the maximum principle, Proposition 8.2.

Theorem 4.1. Let $\mathbf{u}^{\varepsilon}$ be the solution of (10) for every $\varepsilon>0$. For any sequence $\varepsilon_{n} \rightarrow 0$, there is a subsequence of $\mathbf{u}^{\varepsilon_{n}}$ which converges to a minimizer of (6).

Proof. Let $\mathbf{v}$ be an arbitrary minimizer of (6), then we have $J_{\varepsilon}(\mathbf{v})=J(\mathbf{v})$ thanks to the constraint in (6). So,

$$
J_{\varepsilon}\left(\mathbf{u}^{\varepsilon}\right) \leq J_{\varepsilon}(\mathbf{v})=J(\mathbf{v})=: \Lambda
$$

Therefore $\left\|\nabla \mathbf{u}^{\varepsilon}\right\| \leq \sqrt{\Lambda}$ is uniformly bounded and then by Poincaré inequality we get $\left\|\mathbf{u}^{\varepsilon}\right\|$ is bounded, since

$$
\lambda_{1}\left\|u_{i}^{\varepsilon}-\phi_{i}\right\| \leq\left\|\nabla\left(u_{i}^{\varepsilon}-\phi_{i}\right)\right\| \leq \sqrt{\Lambda}+\left\|\nabla \phi_{i}\right\|
$$

Hence, toward a subsequence we can assume that $\mathbf{u}^{\varepsilon_{n}} \rightarrow \mathbf{u} \in \mathcal{K}$. We need to show that $\mathbf{u}$ is a minimizer of (6) and satisfies its constraints. First, we have

$$
\frac{1}{\varepsilon_{n}} \sum_{i \neq j}\left(\left(u_{i}^{\varepsilon_{n}}\right)^{2},\left(u_{j}^{\varepsilon_{n}}\right)^{2}\right) \leq J_{\varepsilon_{n}}\left(\mathbf{u}^{\varepsilon_{n}}\right) \leq \Lambda
$$

so,

$$
\sum_{i \neq j}\left(\left(u_{i}^{\varepsilon_{n}}\right)^{2},\left(u_{j}^{\varepsilon_{n}}\right)^{2}\right) \longrightarrow 0
$$

Thus, $\left(u_{i}^{2}, u_{j}^{2}\right)=0$ and taking into account that $u_{i}^{\varepsilon}$ is nonnegative we obtain that the constraint in (6) holds for $\mathbf{u}$. To close the argument, note that

$$
\|\nabla \mathbf{u}\|^{2}=\lim _{\varepsilon_{n} \rightarrow 0}\left\|\nabla \mathbf{u}^{\varepsilon_{n}}\right\|^{2} \leq J_{\varepsilon_{n}}\left(\mathbf{u}^{\varepsilon_{n}}\right) \leq \Lambda
$$

So, $J(\mathbf{u})=\Lambda$ and $\mathbf{u}$ is a minimizer.
In the sequel we introduce an algorithm to solve problem (10) or its equivalent version (11). The later is a nonlinear system of PDEs and is not easy to solve directly. For an explanation of our algorithm, we define the following sequence which converges to the solution 11. First, consider the harmonic extension $u_{i, 0}$ of boundary data given by

$$
\begin{cases}\mathcal{L} u_{i, 0}=0, & \text { in } X \backslash \Gamma \\ u_{i, 0}=\phi_{i} & \text { on } \Gamma\end{cases}
$$

which is a nonnegative function according to the maximum principle. Next, given nonnegative functions $\mathbf{u}_{m}:=\left(u_{1, m}, \cdots, u_{k, m}\right)$, let $u_{i, m+1}$ be the solution of the following system

$$
\begin{cases}\mathcal{L} u_{i, m+1}+\frac{u_{i, m+1}}{\varepsilon} \sum_{j \neq i} u_{j, m}^{2}=0, & \text { in } X \backslash \Gamma \\ u_{i, m+1}=\phi_{i}, & \text { on } \Gamma\end{cases}
$$

The following theorem shows that why our algorithm works for solving problem (11).
Theorem 4.2. Suppose that the boundary data $\phi_{i}$ satisfy (5). Then the sequence $\mathbf{u}_{m}$ makes the following order

$$
\begin{equation*}
1 \geq u_{i, 0} \geq u_{i, 2} \geq \cdots \geq u_{i, 2 m} \geq \cdots \geq u_{i, 2 m+1} \geq \cdots \geq u_{i, 3} \geq u_{i, 1} \geq 0 \tag{12}
\end{equation*}
$$

Moreover, the limit of this sequence is the solution of (11).
Proof. Step 1: We show that $u_{i, m}$ is nonnegative.
This is a matter of maximum principle, Proposition 8.2. For $m=0$, it is obvious due to maximum principle and taking account that $u_{i, 0}$ is a harmonic function. In fact, we consider $p(x) \equiv 0$ in Proposition 8.2. To show $u_{i, m+1} \geq 0$, again apply Proposition 8.2 for nonnegative function $p(x)=\frac{1}{\varepsilon} \sum_{j \neq i} u_{j, m}^{2}$.

Step 2: $\quad u_{i, 0} \leq 1$.
Apply the maximum principle for harmonic function $1-u_{i, 0}$ and recall the assumption (5).
Step 3: In this step we show that $u_{i, m} \leq u_{i, 0}$.
We just need to note that $\mathcal{L} u_{i, m+1} \leq 0=\mathcal{L} u_{i, 0}$. Then maximum principle yields that $u_{i, m+1} \leq$ $u_{i, 0}$.

Step 4: Here, we claim that $u_{i, 2} \geq u_{i, 1}$.
By the result of Step 2, we can write

$$
\begin{aligned}
0 & =\mathcal{L} u_{i, 2}+\frac{u_{i, 2}}{\varepsilon} \sum_{j \neq i} u_{j, 1}^{2} \\
& \leq \mathcal{L} u_{i, 2}+\frac{u_{i, 2}}{\varepsilon} \sum_{j \neq i} u_{j, 0}^{2} .
\end{aligned}
$$

This together with the equation of $u_{i, 1}$, the maximum principle yields that $u_{i, 2} \geq u_{i, 1}$.
Step 5: Now we close the argument with the induction. Assume that

$$
\begin{equation*}
u_{i, 0} \geq u_{i, 2} \geq \cdots \geq u_{i, 2 m} \geq u_{i, 2 m-1} \geq \cdots \geq u_{i, 1} \geq 0 \tag{13}
\end{equation*}
$$

for some $m \geq 1$, we will extend the string for $m+1$. By the following inequality

$$
0=\mathcal{L} u_{i, 2 m+1}+\frac{u_{i, 2 m+1}}{\varepsilon} \sum_{j \neq i} u_{j, 2 m}^{2} \geq \mathcal{L} u_{i, 2 m+1}+\frac{u_{i, 2 m+1}}{\varepsilon} \sum_{j \neq i} u_{j, 2 m-1}^{2}
$$

we can apply Proposition 8.2 for function $\left(u_{i, 2 m}-u_{i, 2 m+1}\right)$ when $p=\frac{1}{\varepsilon} \sum_{j \neq i} u_{j, 2 m-1}^{2}$ to deduce that $u_{i, 2 m} \geq u_{i, 2 m+1}$. Similarly, we have

$$
0=\mathcal{L} u_{i, 2 m+1}+\frac{u_{i, 2 m+1}}{\varepsilon} \sum_{j \neq i} u_{j, 2 m}^{2} \leq \mathcal{L} u_{i, 2 m+1}+\frac{u_{i, 2 m+1}}{\varepsilon} \sum_{j \neq i} u_{j, 2 m-2}^{2}
$$

according the induction assumption $u_{j, 2 m-2} \geq u_{j, 2 m}$. Comparing with the equation for $u_{i, 2 m-1}$ we obtain $u_{i, 2 m+1} \geq u_{i, 2 m-1}$. Now repeat this argument to show that $u_{i, 2 m} \geq u_{i, 2 m+2} \geq u_{i, 2 m+1}$.

Step 6: From (12), we know that there are $\bar{u}_{i}$ and $\underline{u}_{i}$ as the limit of

$$
\begin{aligned}
u_{i, 2 m} & \longrightarrow \bar{u}_{i}, \\
u_{i, 2 m+1} & \longrightarrow \underline{u}_{i} .
\end{aligned}
$$

These limits satisfy

$$
\begin{cases}\mathcal{L} \bar{u}_{i}+\frac{\bar{u}_{i}}{\varepsilon} \sum_{j \neq i} \underline{u}_{j}^{2}=0, & \text { in } X \backslash \Gamma  \tag{14}\\ \mathcal{L} \underline{u}_{i}+\frac{u_{i}}{\varepsilon} \sum_{j \neq i} \bar{u}_{j}^{2}=0, & \text { in } X \backslash \Gamma \\ \bar{u}_{i}=\underline{u}_{i}=\phi_{i} & \text { on } \Gamma .\end{cases}
$$

Multiply in inner product $\ell^{2}(X)$ both equations by $\bar{u}_{i}-\underline{u}_{i}$ and subtract to get

$$
\varepsilon\left\|\nabla\left(\bar{u}_{i}-\underline{u}_{i}\right)\right\|_{\ell^{2}\left(x^{2}\right)}^{2}=\left(\underline{u}_{i}\left(\bar{u}_{i}-\underline{u}_{i}\right), \sum_{j \neq i} \bar{u}_{j}^{2}\right)-\left(\bar{u}_{i}\left(\bar{u}_{i}-\underline{u}_{i}\right), \sum_{j \neq i} \underline{u}_{j}^{2}\right) .
$$

It is worthwhile noticing that although the equation holds in $X \backslash \Gamma$, since $\bar{u}_{i}-\underline{u}_{i}=0$ on $\Gamma$ we are able to utilize the relation (4). Now sum over $i$, we obtain

$$
\begin{aligned}
\varepsilon\|\nabla(\overline{\mathbf{u}}-\underline{\mathbf{u}})\|_{\ell^{2}\left(x^{2}\right)}^{2}= & \sum_{i}\left(\underline{u}_{i} \bar{u}_{i}, \sum_{j \neq i}\left(\bar{u}_{j}^{2}+\underline{u}_{j}^{2}\right)\right)-\sum_{i}\left(\underline{u}_{i}^{2}, \sum_{j \neq i} \bar{u}_{j}^{2}\right)-\sum_{i}\left(\bar{u}_{i}^{2}, \sum_{j \neq i} \underline{u}_{j}^{2}\right) \\
= & \sum_{i, j}\left(\underline{u}_{i} \bar{u}_{i}, \bar{u}_{j}^{2}+\underline{u}_{j}^{2}\right)-\sum_{i}\left(\underline{u}_{i} \bar{u}_{i}, \bar{u}_{i}^{2}+\underline{u}_{i}^{2}\right) \\
& -2 \sum_{i, j}\left(\underline{u}_{i}^{2}, \bar{u}_{j}^{2}\right)+2 \sum_{i}\left(\underline{u}_{i}^{2}, \bar{u}_{i}^{2}\right) \\
= & \sum_{x \in X} \sum_{i, j} \underline{u}_{i}(x) \bar{u}_{i}(x)\left(\bar{u}_{j}^{2}(x)+\underline{u}_{j}^{2}(x)\right) \\
& -\sum_{x \in X} \sum_{i} \underline{u}_{i}(x) \bar{u}_{i}(x)\left(\bar{u}_{i}(x)-\underline{u}_{i}(x)\right)^{2} \\
& -2 \sum_{x \in X}\left(\sum_{i}\left(\underline{u}_{i}(x)\right)^{2}\right)\left(\sum_{i}\left(\bar{u}_{i}(x)\right)^{2}\right) \\
\leq & 2 \sum_{x \in X}\left(\sum_{i} \underline{u}_{i}(x) \bar{u}_{i}(x)-\left(\sum_{i}\left(\underline{u}_{i}(x)\right)^{2}\right)\left(\sum_{i}\left(\bar{u}_{i}(x)\right)^{2}\right)\right) \leq 0,
\end{aligned}
$$

where we have used the relation $0 \leq \underline{u}_{i} \leq \bar{u}_{i} \leq 1$ and in the last line the Cauchy-Schwartz inequality has been applied.

Then $\|\nabla(\overline{\mathbf{u}}-\underline{\mathbf{u}})\|_{\ell^{2}\left(X^{2}\right)}^{2} \leq 0$ and so, $\overline{\mathbf{u}}-\underline{\mathbf{u}}$ is constant in $X$. Taking this along to the boundary condition implies that $\overline{\mathbf{u}}=\underline{\mathbf{u}}$ in $X$. Recall (14), $\underline{u}_{i}=\bar{u}_{i}$ is a solution of (11).

## 5. The main algorithm for Clustering

In the previous sections of the paper we considered a minimization problem (6), which unfortunately has no unique solution over connected graphs. In the current section in order to overcome the lack of uniqueness we consider different functional and prove the existence and uniqueness of the minimizer. The definition of a new functional is inspired from the numerical results of the spatial segregation of reaction-diffusion systems (see [2]).
5.1. Existence and uniqueness of a minimizer. We introduce the discrete counterpart of the spatial segregation problem defined on connected graphs. In the rest of the paper the following notation

$$
\hat{z}_{q}=z_{q}-\sum_{j \neq q} z_{j},
$$

for elements $\left(z_{1}, z_{2}, \ldots, z_{k}\right)$, will play a crucial role. Let $\bar{u}_{i}\left(x_{l}\right)$ for $i=1,2 \cdots k$ denote the average value of $u_{i}$ for all neighbor points of $x_{l}$ :

$$
\bar{u}_{i}\left(x_{l}\right)=\frac{1}{\operatorname{deg}\left(x_{l}\right)} \sum_{p \sim l} w_{l p} u_{i}\left(x_{p}\right)
$$

where

$$
\operatorname{deg}\left(x_{l}\right)=\sum_{\left(x_{l}, y\right) \in E} w_{x_{l} y},
$$

and $V$ and $E$ stand for a set of vertices and edges respectively. We will set a graph $X$ to be a tuple $(V, E)$ in the rest of this section.

When $X$ is a connected graph and also consist of discrete and finite number of points, it turns out that we have to consider slightly different functional (see [2, Section 2]). Since $\mathcal{L}$ is a self-adjoint operator, then we set:

$$
\begin{equation*}
J\left(u_{1}, \ldots, u_{k}\right)=\frac{1}{2} \sum_{i=1}^{k}\left\|\nabla u_{i}\right\|_{\ell^{2}\left(X^{2}\right)}^{2}-\sum_{i \neq j}\left(\nabla u_{i}, \nabla u_{j}\right)_{\ell^{2}\left(X^{2}\right)}, \tag{15}
\end{equation*}
$$

over the set

$$
\begin{equation*}
\mathcal{K}=\left\{\mathbf{u}=\left(u_{1}, \cdots, u_{k}\right) \in\left(\ell^{2}(X)\right)^{k}: u_{i}=\phi_{i} \text { on } \Gamma, u_{i} \geq 0, u_{i} \cdot u_{j}=0 \text { for } i \neq j\right\} . \tag{16}
\end{equation*}
$$

Theorem 5.1. The following minimization problem

$$
\begin{equation*}
\inf _{\mathscr{K}} J\left(u_{1}, u_{2}, \ldots, u_{k}\right) \tag{17}
\end{equation*}
$$

has a solution.
Proof. The proof repeats the same lines as in Theorem 2 in [2]. In [2] the functional is defined for standard difference scheme, but it can be easily concluded for the connected graphs as well.

Now, by following the proofs of Proposition 1 and Lemma 2 for $F_{l}(x, s)=0$ in [2], We can observe that the similar results can be obtained for connected graphs instead of finite difference discretization. Although, it is worth to notice that the standard finite difference grid is itself a particular case of connected graphs.

Thus, we conclude the following result:
Theorem 5.2. For every minimizer $\left(u_{1}, \ldots, u_{k}\right) \in \mathcal{K}$, the following properties hold:

- $\mathcal{L} \hat{u}_{i}(x)=0$ whenever $u_{i}(x)>0$.
- $\mathcal{L} \hat{u}_{i}(x) \geq 0$ whenever $x \in \mathcal{X} \backslash \Gamma$.

To prove the uniqueness of the minimizer $\left(u_{1}, \ldots, u_{k}\right) \in \mathscr{K}$ one needs some technical lemmas.
Lemma 5.3. Let $X=(V, E)$ be a connected graph. If any two vectors $\left(u_{1}, u_{2}, \ldots, u_{k}\right)$ and $\left(v_{1}, v_{2}, \ldots, v_{k}\right)$ are minimizers to the (17), then the following equation holds:

$$
\max _{x \in X}\left(\hat{u}_{l}(x)-\hat{v}_{l}(x)\right)=\max _{\left\{x \in X: u_{l}(x) \leq v_{l}(x)\right\}}\left(\hat{u}_{l}(x)-\hat{v}_{l}(x)\right)
$$

for all $l=1,2, \ldots, k$.
Proof. We argue by contradiction. Suppose for some $l_{0}$ we have

$$
\begin{align*}
& \hat{u}_{l_{0}}\left(x_{0}\right)-\hat{v}_{l_{0}}\left(x_{0}\right)=\max _{x \in X}\left(\hat{u}_{l_{0}}(x)-\hat{v}_{l_{0}}(x)\right)= \\
& \quad=\max _{\left\{x \in X: u_{l_{0}}(x)>v_{l_{0}}(x)\right\}}\left(\hat{u}_{l_{0}}(x)-\hat{v}_{l_{0}}(x)\right)>\max _{\left\{x \in X: u_{l_{0}}(x) \leq v_{l_{0}}(x)\right\}}\left(\hat{u}_{l_{0}}(x)-\hat{v}_{l_{0}}(x)\right) . \tag{18}
\end{align*}
$$

It is easy to observe that the following simple chain of inclusions hold:

$$
\begin{equation*}
\left\{u_{l}(x)>v_{l}(x)\right\} \subset\left\{\hat{u}_{l}(x)>\hat{v}_{l}(x)\right\} \subset\left\{u_{l}(x) \geq v_{l}(x)\right\} . \tag{19}
\end{equation*}
$$

We obviously see that $u_{l_{0}}\left(x_{0}\right)>v_{l_{0}}\left(x_{0}\right) \geq 0$ implies $\hat{u}_{l_{0}}\left(x_{0}\right)>\hat{v}_{l_{0}}\left(x_{0}\right)$. On the other hand, Theorem 5.2 gives us

$$
\mathcal{L} \hat{u}_{i}\left(x_{0}\right)=0,
$$

and

$$
\mathcal{L} \hat{v}_{i}\left(x_{0}\right) \geq 0 .
$$

Therefore

$$
\mathcal{L}\left(\hat{u}_{i}-\hat{v}_{i}\right)\left(x_{0}\right) \leq 0 .
$$

Thus,

$$
0<\left(\hat{u}_{l_{0}}\left(x_{0}\right)-\hat{v}_{l_{0}}\left(x_{0}\right)\right) \leq \frac{1}{\operatorname{deg}\left(x_{0}\right)} \sum_{y \in X} w_{x_{0} y}\left(\hat{u}_{l_{0}}(y)-\hat{v}_{l_{0}}(y)\right),
$$

which implies that $\hat{u}_{l_{0}}\left(x_{0}\right)-\hat{v}_{l_{0}}\left(x_{0}\right)=\hat{u}_{l_{0}}(y)-\hat{v}_{l_{0}}(y)>0$, when $w_{x_{0} y} \neq 0$. Due to the chain (19), we apparently have $u_{l_{0}}(y) \geq v_{l_{0}}(y)$. According to our assumption (18), the only possibility is $u_{l_{0}}(y)>v_{l_{0}}(y)$ for all $y \in X$. Now we can proceed the previous steps for all $y \in V$ such that $\left(x_{0}, y\right) \in E$, and then for each one we will get corresponding neighbours with the same strict inequality and so on. Since the graph $X$ is connected, then one can always find the shortest path from a given vertex $y$ to the vertex belonging $\Gamma$. Continuing above procedure along this path we will finally approach to a vertex on $\Gamma$, where as we know $u_{l_{0}}(x)=v_{l_{0}}(x)=\phi_{l_{0}}(x)$ for all $x \in \Gamma$. Hence, the strict inequality fails, which implies that our initial assumption 18) is false. Observe that the same arguments can be applied if we interchange the roles of $u_{l}(x)$ and $v_{l}(x)$. Thus, we also have

$$
\max _{V}\left(\hat{v}_{l}(x)-\hat{u}_{l}(x)\right)=\max _{\left\{v_{l}(x) \leq u_{l}(x)\right\}}\left(\hat{v}_{l}(x)-\hat{u}_{l}(x)\right),
$$

for every $l=1,2, \ldots, m$.
Particularly, for every fixed $l=1,2 \ldots, m$ and $x \in V$ we have

$$
\begin{align*}
& -\max _{\left\{v_{l}(x) \leq u_{l}(x)\right\}}\left(\hat{v}_{l}(x)-\hat{u}_{l}(x)\right)=-\max _{x \in V}\left(\hat{v}_{l}(x)-\hat{u}_{l}(x)\right) \leq \\
& \leq \hat{u}_{l}(x)-\hat{v}_{l}(x) \leq \max _{x \in V}\left(\hat{u}_{l}(x)-\hat{v}_{l}(x)\right)=\max _{\left\{u_{l}(x) \leq v_{l}(x)\right\}}\left(\hat{u}_{l}(x)-\hat{v}_{l}(x)\right) . \tag{20}
\end{align*}
$$

Thanks to Lemma 5.3 in the sequel we will use the following notations:

$$
A:=\max _{l}\left(\max _{x \in V}\left(\hat{u}_{l}(x)-\hat{v}_{l}(x)\right)\right)=\max _{l}\left(\max _{\left\{u_{l}(x) \leq v_{l}(x)\right\}}\left(\hat{u}_{l}(x)-\hat{v}_{l}(x)\right)\right),
$$

and

$$
B:=\max _{l}\left(\max _{x \in V}\left(\hat{v}_{l}(x)-\hat{u}_{l}(x)\right)\right)=\max _{l}\left(\max _{\left\{v_{l}(x) \leq u_{l}(x)\right\}}\left(\hat{v}_{l}(x)-\hat{u}_{l}(x)\right)\right) .
$$

Next lemma we write down without a proof. The proof can be easily adapted from [2] [Lemma 4].

Lemma 5.4. Let $\mathcal{X}=(V, E)$ be a connected graph. Assume given two vectors $\left(u_{1}, u_{2}, \ldots, u_{k}\right)$ and $\left(v_{1}, v_{2}, \ldots, v_{k}\right)$ are minimizers to the (17). For them we set $A$ and $B$ as defined above. If $A>0$ and it is attained for some $l_{0}$, then $A=B>0$ and there exists some $t_{0} \neq l_{0}$, and $y_{0} \in V$, such that

$$
0<A=\max _{\left\{u_{l_{0}}(x) \leq v_{l_{0}}(x)\right\}}\left(\hat{u}_{l_{0}}(x)-\hat{v}_{l_{0}}(x)\right)=\max _{\left\{u_{l_{0}}(x)=v_{l_{0}}(x)=0\right\}}\left(\hat{u}_{l_{0}}(x)-\hat{v}_{l_{0}}(x)\right)=\hat{v}_{t_{0}}\left(y_{0}\right)-\hat{u}_{t_{0}}\left(y_{0}\right) .
$$

Now, we are ready to proof the uniqueness of the minimizer. The following Theorem is true.
Theorem 5.5 (Uniqueness). Let $\mathcal{X}=(V, E)$ be a connected graph. Then there exists a unique minimizer $\left(u_{1}, u_{2}, \ldots, u_{k}\right) \in \mathcal{K}$, to minimization problem (17).

Proof. Let two vectors $\left(u_{1}, u_{2}, \ldots, u_{k}\right)$ and $\left(v_{1}, v_{2}, \ldots, v_{k}\right)$ are minimizers to (17). For these vectors we set the definition of $A$ and $B$. Then, we consider two cases $A \leq 0$ and $A>0$. If we assume that $A \leq 0$, then according to Lemma 5.4, we get $B \leq 0$. But if $A$ and $B$ are non-positive, then the uniqueness follows. Indeed, due to 20 we have the following obvious inequalities

$$
0 \leq-B \leq \hat{u}_{l}(x)-\hat{v}_{l}(x) \leq A \leq 0
$$

This provides for every $l=\overline{1, k}$ and $x \in V$ we have $\hat{u}_{l}(x)=\hat{v}_{l}(x)$, which in turn implies

$$
u_{l}(x)=v_{l}(x)
$$

Now suppose $A>0$. Our aim is to prove that this case leads to a contradiction. Let the value $A$ is attained for some $l_{0} \in \overline{1, k}$, then due to Lemma 5.4 there exist $y_{0} \in V$ and $t_{0} \neq l_{0}$ such that:

$$
\begin{aligned}
0<A=B= & \max _{\left\{u_{l_{0}}(x) \leq v_{l_{0}}(x)\right\}}\left(\hat{u}_{l_{0}}(x)-\hat{v}_{l_{0}}(x)\right) \\
& =\max _{\left\{u_{l_{0}}(x)=v_{l_{0}}(x)=0\right\}}\left(\hat{u}_{l_{0}}(x)-\hat{v}_{l_{0}}(x)\right)=\hat{v}_{t_{0}}\left(y_{0}\right)-\hat{u}_{t_{0}}\left(y_{0}\right) .
\end{aligned}
$$

Thus, along with the fact that $\hat{v}_{t_{0}}\left(y_{0}\right)>\hat{u}_{t_{0}}\left(y_{0}\right)$ implies $v_{t_{0}}\left(y_{0}\right) \geq u_{t_{0}}\left(y_{0}\right)$, we can repeat the same steps as in the proof of Lemma 5.3 to obtain

$$
\left(\hat{v}_{t_{0}}\left(y_{0}\right)-\hat{u}_{t_{0}}\left(y_{0}\right)\right) \leq \frac{1}{\operatorname{deg}\left(y_{0}\right)} \sum_{\left(y_{0}, z\right) \in E} w_{y_{0} z}\left(\hat{v}_{t_{0}}(z)-\hat{u}_{t_{0}}(z)\right)
$$

This implies $A=\hat{v}_{t_{0}}\left(y_{0}\right)-\hat{u}_{t_{0}}\left(y_{0}\right)=\hat{v}_{t_{0}}(z)-\hat{u}_{t_{0}}(z)>0$ for all $\left(y_{0}, z\right) \in E$. The chain 19) provides that for all $\left(y_{0}, z\right) \in E$, we have $v_{t_{0}}(z) \geq u_{t_{0}}(z)$. Since a graph $X$ is connected, then one can always find a shortest path from $y_{0}$ to some vertex $w \in \Gamma$. Assume the vertices along this path are $y_{0} ; y_{1} ; \ldots ; y_{k-1} ; y_{q}=w$. Hence, for every $0 \leq j \leq q-1$, we have $\left(y_{j}, y_{j+1}\right) \in E$, i.e. every vertex $y_{j+1}$ is a closest neighbor for $y_{j}$ and $y_{j+2}$.

According to the above arguments for the neighbor vertex $y_{1} \in V$ we proceed as follows: If $v_{t_{0}}\left(y_{1}\right)>u_{t_{0}}\left(y_{1}\right)$, then obviously

$$
\left(\hat{v}_{t_{0}}\left(y_{1}\right)-\hat{u}_{t_{0}}\left(y_{1}\right)\right) \leq \frac{1}{\operatorname{deg}\left(y_{1}\right)} \sum_{\left(y_{1}, z\right) \in E} w_{y_{1} z}\left(\hat{v}_{t_{0}}(z)-\hat{u}_{t_{0}}(z)\right)
$$

This, as we saw a few lines above, leads to $A=\hat{v}_{t_{0}}\left(y_{1}\right)-\hat{u}_{t_{0}}\left(y_{1}\right)=\hat{v}_{t_{0}}(z)-\hat{u}_{t_{0}}(z)>0$ for all $\left(y_{1}, z\right) \in E$. In particular, $A=\hat{v}_{t_{0}}\left(y_{2}\right)-\hat{u}_{t_{0}}\left(y_{2}\right)>0$.

If $v_{t_{0}}\left(y_{1}\right)=u_{t_{0}}\left(y_{1}\right)$, then due to $\hat{v}_{t_{0}}\left(y_{1}\right)-\hat{u}_{t_{0}}\left(y_{1}\right)=A=B>0$, there exists some $\lambda_{0} \neq t_{0}$, such that

$$
0<A=\hat{v}_{t_{0}}\left(y_{1}\right)-\hat{u}_{t_{0}}\left(y_{1}\right)=\sum_{l \neq t_{0}}\left(u_{l}\left(y_{1}\right)-v_{l}\left(y_{1}\right)\right)=u_{\lambda_{0}}\left(y_{1}\right)-\sum_{l \neq t_{0}} v_{l}\left(y_{1}\right) .
$$

Note that $u_{\lambda_{0}}\left(y_{1}\right)>0$ implies $u_{l}\left(y_{1}\right)=0$ for all $l \neq \lambda_{0}$, and particularly $v_{t_{0}}\left(y_{1}\right)=u_{t_{0}}\left(y_{1}\right)=0$. Following the definition of $A$, we get

$$
A=u_{\lambda_{0}}\left(y_{1}\right)-\sum_{l \neq t_{0}} v_{l}\left(y_{1}\right) \geq \hat{u}_{\lambda_{0}}\left(y_{1}\right)-\hat{v}_{\lambda_{0}}\left(y_{1}\right),
$$

which in turn gives $2 \sum_{l \neq \lambda_{0}} v_{l}\left(y_{1}\right) \leq 0$, and therefore $v_{l}\left(y_{1}\right)=0$ for all $l \neq \lambda_{0}$. Hence

$$
A=u_{\lambda_{0}}\left(y_{1}\right)-\sum_{l \neq t_{0}} v_{l}\left(y_{1}\right)=\hat{u}_{\lambda_{0}}\left(y_{1}\right)-\hat{v}_{\lambda_{0}}\left(y_{1}\right),
$$

This suggests us to apply the same approach as above to arrive at

$$
\left(\hat{v}_{\lambda_{0}}\left(y_{1}\right)-\hat{u}_{\lambda_{0}}\left(y_{1}\right)\right) \leq \frac{1}{\operatorname{deg}\left(y_{1}\right)} \sum_{\left(y_{1}, z\right) \in E} w_{y_{1} z}\left(\hat{v}_{\lambda_{0}}(z)-\hat{u}_{\lambda_{0}}(z)\right),
$$

which leads to $A=\hat{u}_{\lambda_{0}}\left(y_{1}\right)-\hat{v}_{\lambda_{0}}\left(y_{1}\right)=\hat{u}_{\lambda_{0}}(z)-\hat{v}_{\lambda_{0}}(z)>0$, for all $\left(y_{1}, z\right) \in E$. In particular, $A=\hat{u}_{\lambda_{0}}\left(y_{2}\right)-\hat{v}_{\lambda_{0}}\left(y_{2}\right)>0$. Thus, combining two cases we observe that for $y_{2} \in V$ there exist an index $1 \leq l_{y_{2}} \leq m$ (in our case $l_{y_{2}}=t_{0}$ or $l_{y_{2}}=\lambda_{0}$ ) such that

$$
\begin{equation*}
\text { either } \hat{u}_{l_{y_{2}}}\left(y_{2}\right)-\hat{v}_{l_{y_{2}}}\left(y_{2}\right)=A \text {, or } \hat{u}_{l_{y_{2}}}\left(y_{2}\right)-\hat{v}_{l_{y_{2}}}\left(y_{2}\right)=-A . \tag{21}
\end{equation*}
$$

It is not hard to understand that the same procedure can be repeated for a vertex $y_{2}$ instead of $y_{1}$ and come to the same conclusion (21) for $y_{3} \in V$ and some index $l_{y_{3}}$ and so on. This allows to claim that for every $y_{j} \in V$ along the path $\left(y_{0}, \ldots, y_{q}\right)$ there exist some $l_{y_{j}}$ such that

$$
\left|\hat{u}_{l_{y_{j}}}\left(y_{j}\right)-\hat{v}_{l_{y_{j}}}\left(y_{j}\right)\right|=A>0 .
$$

But this means that above equality holds also for $y_{q}=w \in \Gamma$, which will lead to a contradiction, because for every $z \in \Gamma$, and $l=1, \cdots, k$ one has $\hat{u}_{l}(z)-\hat{v}_{l}(z)=0$. This completes the proof of uniqueness.
5.2. Semi-supervised learning algorithm. Using definition of graph Laplacian in

$$
\mathcal{L}\left(u_{i}-\sum_{j \neq i} u_{j}\right) \geq 0,
$$

yields

$$
\begin{equation*}
\mathcal{L}\left(u_{i}-\sum_{j \neq i} u_{j}\right)\left(x_{l}\right)=\sum_{s=1}^{n} w_{l s}\left(u_{i}\left(x_{l}\right)-u_{i}\left(x_{s}\right)-\sum_{j \neq i}\left(u_{j}\left(x_{l}\right)-u_{j}\left(x_{s}\right)\right)\right) . \tag{22}
\end{equation*}
$$

To obtain $u_{i}\left(x_{l}\right)$ from (22) we impose the following conditions

$$
u_{i}\left(x_{l}\right) \cdot u_{j}\left(x_{l}\right)=0 \text { and } u_{i}\left(x_{l}\right) \geq 0 .
$$

From these

$$
\operatorname{deg}\left(x_{l}\right) u_{i}\left(x_{l}\right)-\sum_{s=1}^{n} w_{l s} u_{i}\left(x_{s}\right)+\sum_{s=1}^{n} \sum_{j \neq i} w_{l s} u_{j}\left(x_{s}\right)=0
$$

Then

$$
u_{i}\left(x_{l}\right)=\bar{u}_{i}\left(x_{l}\right)-\sum_{j \neq i} \bar{u}_{j}\left(x_{l}\right)
$$

According to the above ideas and following the Theorem 5.2 we can easily check that if $\left(u_{1}, u_{2}, \ldots, u_{k}\right) \in \mathcal{K}$ is a unique minimizer to (17), then it satisfies the following system of equations:

$$
\left\{\begin{array}{l}
u_{1}(x)=\max \left(\bar{u}_{1}(x)-\sum_{p \neq 1} \bar{u}_{p}(x), 0\right), \quad x \in X \backslash \Gamma  \tag{23}\\
u_{2}(x)=\max \left(\bar{u}_{2}(x)-\sum_{p \neq 2} \bar{u}_{p}(x), 0\right), \quad x \in X \backslash \Gamma \\
\ldots \ldots \ldots \ldots \\
u_{k}(x)=\max \left(\bar{u}_{k}(x)-\sum_{p \neq k} \bar{u}_{p}(x), 0\right), \quad x \in X \backslash \Gamma \\
u_{i}(x)=\phi_{i}(x), \quad x \in \Gamma, \text { for all } i=1,2, \ldots, k
\end{array}\right.
$$

Remark 5.6. We remark that the system (23) itself implies the disjointness property, i.e. it is easy to see that if a vector $\left(u_{1}, u_{2}, \ldots, u_{k}\right)$ satisfies the system (23), then $u_{i}(x) \cdot u_{j}(x)=0$, for every $x \in V$ and $i \neq j$.

In order to approximate the solution of system (23) we propose the iterative scheme which is easy to implement as follows: For $i=1, \cdots k$, and $x_{l} \in X \backslash \Gamma$ we set

$$
u_{i}^{(t+1)}\left(x_{l}\right)=\max \left(\bar{u}_{i}^{(t)}\left(x_{l}\right)-\sum_{j \neq i} \bar{u}_{j}^{(t)}\left(x_{l}\right), 0\right)
$$

In the lite of Remark 5.6 it can be seen that for every iteration the disjointness property is preserved. In other words the following lemma is true.

Lemma 5.7. Let $X=(V, E)$ be a connected graph. The above iterative method satisfies

$$
u_{i}^{(t)}(x) \cdot u_{j}^{(t)}(x)=0, \forall x \in V, i \neq j
$$

The label decision for vertex $x_{l}$ is determined by the strictly positive component $u_{i}\left(x_{l}\right)$, i.e. find an index $i_{0}$ such that $u_{i_{0}}\left(x_{l}\right)>0$. Thus, in this case the label corresponding to a vertex $x_{l}$ will be $i_{0}$.

## 6. Experimental Results

In this section we are going to test and compare two well-known semi-supervised learning algorithms to the one we have developed based on segregation theory. We note that our taken dataset for visual implementations will be the random generated half-moons and for the statistic analysis we will use the well-known MNIST. Thus, we will depict the predictions for Laplace learning, Poisson learning and our learning (We call it Segregation learning) algorithms.

We run the learning algorithms for different number of initial label of classes and for different number of classes (basically we will run for 3,4 and 5 classes). For each implementation all the classes have the same number of nodes, i.e, either all classes have 200 or 300 nodes. The reader can also observe the red nodes on every figure. They correspond to the randomly chosen initial known labels. In the figures 115 below one can observe that, when the initial number of labels per class is small, i.e. 2,3 or 5 labels, then the Laplace learning algorithm is performing poorly, whilst both the Poisson and our Segregation learning algorithms are performing much better and have more or less the same accuracy.

When the initial number of labels per class is 10 or 20 labels, then the performance of the Laplace learning becomes more accurate and is getting close to the results depicted for Poisson and Segregation learning algorithms.

Tables 1 and 2 show the average accuracy over all 100 trials for various low and high label rates. The implementations have been done on MNIST dataset only for 3 classes. We see that for low label rates Laplace learning performs poor as we noted in the depicted figures. On the other hand Poisson and Segregation learning perform better and predicted more or less with the same accuracy. For high label rates Laplace learning performs much better and gets close to Poisson and Segregation learning results.


Figure 1. Comparison of Laplace, Poisson and Segregation learning algorithms for 3 classes and initial 2 labels per class.

## 7. Conclusion

In this work we develop several semi-supervised algorithms on graphs. Our main algorithm is a new approach for graph-based semi-supervised learning based on spatial segregation theory. The method is efficient and simple to implement. We presented numerical results showing that Segregation Learning performs more or less as Poisson Learning algorithm not only at high label rates, but also at low label rates on MNIST dataset.


Figure 2. Comparison of Laplace, Poisson and Segregation learning algorithms for 3 classes and initial 3 labels per class.


Figure 3. Comparison of Laplace, Poisson and Segregation learning algorithms for 3 classes and initial 5 labels per class.


Figure 4. Comparison of Laplace, Poisson and Segregation learning algorithms for 3 classes and initial 10 labels per class.

## REFERENCES

[1] R. K. Ando, T. Zhang, Learning on graph with Laplacian regularization, In Advances in Neural Information Processing Systems, pp. 25-32, 2007.
[2] A. Arakelyan, R. Barkhudaryan, A numerical approach for a general class of the spatial segregation of reaction-diffusion systems arising in population dynamics. Computers and Mathematics with Applications 72, (2016), 2823-2838.


Figure 5. Comparison of Laplace, Poisson and Segregation learning algorithms for 3 classes and initial 20 labels per class.


Figure 6. Comparison of Laplace, Poisson and Segregation learning algorithms for 4 classes and initial 2 labels per class.


Figure 7. Comparison of Laplace, Poisson and Segregation learning algorithms for 4 classes and initial 3 labels per class.
[3] A. Arakelyan Convergence of the finite difference scheme for a general class of the spatial segregation of reaction-diffusion systems. Computers and Mathematics with Applications, 75, (2018) 4232-4240.
[4] A. Arakelyan, F. Bozorgnia, On the uniqueness of the limiting solution to a strongly competing system. Electronic Journal of Differential Equations, 96, (2017) 1-8.


Figure 8. Comparison of Laplace, Poisson and Segregation learning algorithms for 4 classes and initial 5 labels per class.


Figure 9. Comparison of Laplace, Poisson and Segregation learning algorithms for 4 classes and initial 10 labels per class.


Figure 10. Comparison of Laplace, Poisson and Segregation learning algorithms for 4 classes and initial 20 labels per class.
[5] M. Belkin, P. Niyogi, Using manifold structure for partially labelled classification. In Advances in Neural Information Processing Systems, 2002.
[6] A. L. Bertozzi, A. Flenner, Diffuse interface models on graphs for classification of high dimensional data. Multiscale Modeling and Simulation, 10(3):1090-1118, 2012.


Figure 11. Comparison of Laplace, Poisson and Segregation learning algorithms for 5 classes and initial 2 labels per class.


Figure 12. Comparison of Laplace, Poisson and Segregation learning algorithms for 5 classes and initial 3 labels per class.


Figure 13. Comparison of Laplace, Poisson and Segregation learning algorithms for 5 classes and initial 5 labels per class.
[7] F. Bozorgnia, A. Arakelyan, Numerical algorithms for a variational problem of the spatial segregation of reaction-diffusion systems. Applied Mathematics and Computation 219 (17), 8863-8875.
[8] F. Bozorgnia, Numerical algorithms for the spatial segregation of competitive systems. SIAM J. Sci. Comput, 31, (2009) 3946-3958.


Figure 14. Comparison of Laplace, Poisson and Segregation learning algorithms for 5 classes and initial 10 labels per class.


Figure 15. Comparison of Laplace, Poisson and Segregation learning algorithms for 5 classes and initial 20 labels per class.

Table 1. Average accuracy scores over 100 trials for 3 classes on MNIST dataset.

| Labels per class | $\mathbf{2}$ | $\mathbf{3}$ | $\mathbf{4}$ | $\mathbf{5}$ | $\mathbf{1 0}$ |
| :---: | :--- | :--- | :--- | :--- | :--- |
| Laplace learning | 31.3 | 45.4 | 58.2 | 67.7 | 83.4 |
| Poisson learning | 93.6 | 94.5 | 94.9 | 95.3 | 96.7 |
| Segregation learning | $\mathbf{9 0 . 3}$ | $\mathbf{9 2 . 1}$ | $\mathbf{9 3 . 5}$ | $\mathbf{9 5 . 4}$ | $\mathbf{9 6 . 2}$ |

Table 2. Average accuracy scores over 100 trials for 3 classes on MNIST dataset.

| Labels per class | $\mathbf{2 0}$ | $\mathbf{4 0}$ | $\mathbf{8 0}$ | $\mathbf{1 0 0}$ | $\mathbf{1 2 0}$ |
| :---: | :--- | :--- | :--- | :--- | :--- |
| Laplace learning | 88.6 | 91.3 | 93.7 | 95.2 | 97.6 |
| Poisson learning | 95.6 | 97.2 | 97.9 | 98.3 | 99.4 |
| Segregation learning | $\mathbf{9 4 . 3}$ | $\mathbf{9 6 . 7}$ | $\mathbf{9 8 . 2}$ | $\mathbf{9 8 . 8}$ | $\mathbf{9 9 . 2}$ |

[9] L. Caffarelli, F. Lin, Singularly perturbed elliptic systems and multi-valued harmonic functions with free boundaries. J. Amer. Math. Soc. 21, no. 3, (2008) 847-862.
[10] J. Calder, B. Cook, M. Thorpe, and D. Slepčev. Poisson Learning: Graph based semi-supervised learning at very low label rates. Proceedings of the 37th International Conference on Machine Learning, PMLR,

119:1306-1316, 2020.
[11] J. Calder, The game theoretic p-Laplacian and semisupervised learning with few labels. Nonlinearity, 32 (1), 2018.
[12] J. Calder, Consistency of Lipschitz learning with infinite unlabeled data and finite labeled data. SIAM Journal on Mathematics of Data Science, 1:780-812, 2019.
[13] J. Calder, J. and D. Slepčev, Properly-Weighted graph Laplacian for semi-supervised learning. Applied mathematics and optimization, 82(3), 1111-1159, 2020.
[14] M. Conti, S. Terracini, S., and G. Verzini, A variational problem for the spatial segregation of reactiondiffusion systems. Indiana University mathematics journal, 779-815, 2005.
[15] X. Desquesnes, A. Elmoataz, O.Lezoray Eikonal equation adaptation on weighted graphs: fast geometric diffusion process for local and non-local image and data processing. Journal of Mathematical Imaging and Vision, Springer Verlag, 2013, 46 (2), pp.238-257. ff10.1007/s10851-012- 0380-9ff. ffhal-00932510f
[16] A. El Alaoui, X. Cheng, A. Ramdas, M. J. Wainwright, and M. I. Jordan. Asymptotic behavior of lp-based Laplacian regularization in semi-supervised learning. In 29th Annual Conference on Learning Theory, pages 879-906, 2016.
[17] I. El Bouchairi, J. Fadili, A. Elmoataz, Continuum limit of p-Laplacian evolution problems on graphs: $L^{q}$ graphons and sparse graphs. https://arxiv.org/pdf/2010.08697.pdf.
[18] H. Ennaji, Y. Quéau, A. Elmoataz, Tug-of-War games and PDEs on graphs: simple image and high dimensional data processing. 2022, https://hal.archives-ouvertes.fr/hal-03675971.
[19] N. García Trillos, M. Gerlach, M. Hein, and Slepčev, Error Estimates for Spectral Convergence of the Graph Laplacian on Random Geometric Graphs Toward the Laplace-Beltrami Operator. In: Foundations of Computational Mathematics 20 (2020),pp. 827-887. DOI: 10.1007/s10208-019-09436-w.
[20] C. Garcia-Cardona, E. Merkurjev, A. L. Bertozzi, A. Flenner, and A. G. Percus, Multiclass data segmentation using diffuse interface methods on graphs, IEEE Trans. Pattern Anal. Mach. Intell. 36 (2014), 1600-1613.
[21] K. Ghasedi, X. Wang, C. Deng, and H. Huang, Balanced self-paced learning for generative adversarial clustering network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4391-4400, 2019.
[22] R. Kyng, A. Rao, S. Sachdeva, D. A. Spielman, Algorithms for Lipschitz learning on graphs. In Conference on Learning Theory, pp. 1190-1223, 2015.
[23] B. Nadler, N. Srebro, and X. Zhou, Semi-supervised learning with the graph Laplacian: The limit of infinite unlabelled data. Advances in Neural Information Processing Systems, 22:1330-1338, 2009.
[24] Slepčev, D. and Thorpe, M., Analysis of p-Laplacian regularization in semisupervised learning, SIAM Journal on Mathematical Analysis, 51(3), pp.2085-2120, 2019.
[25] Shi, Z., Osher, S. and Zhu, W., Weighted nonlocal laplacian on interpolation from sparse data. Journal of Scientific Computing, 73(2), pp.1164-1177, 2017.
[26] Ulrike von Luxburg. A tutorial on spectral clustering. Statistics and computing, Vol. 17, 4 (2007),pp. 395-416.
[27] X. Zhu, Z. Ghahramani, and J. D. Lafferty, Semisupervised learning using Gaussian fields and harmonic functions. In Proceedings of the 20th International Conference on Machine learning (ICML-03), pp. 912-919, 2003.

## 8. Appendix

In this section, we record two important statements in graphs, Poincaré inequality and maximum principle for superharmonic functions.

Proposition 8.1 (Poincaré inequaity). Assume the graph $X$ is connected. For every $\Gamma \subset \mathcal{X}$, there exists constant $\lambda_{1}>0$, the first eigenvalue of Laplacian, such that

$$
\lambda_{1}\|u\|_{\ell^{2}(X)} \leq\|\nabla u\|_{\ell^{2}\left(X^{2}\right)}
$$

for all $u \in \ell^{2}(X)$ satisfying $u=0$ on $\Gamma$.
Proof. By the contradiction, we may assume that there exist the sequence $u_{n}$ such that

$$
\left\|\nabla u_{n}\right\|_{\ell^{2}\left(X^{2}\right)} \leq \frac{1}{n}\left\|u_{n}\right\|_{\ell^{2}(X)}, \quad u_{n}=0 \text { on } \Gamma
$$

Let $\hat{u}_{n}:=u_{n} /\left\|u_{n}\right\|$, then $\left\|\nabla \hat{u}_{n}\right\| \rightarrow 0$. The sequence $\left\{\hat{u}_{n}(x)\right\}_{n}$ is uniformly bounded $\left(\left|\hat{u}_{n}(x)\right| \leq\right.$ $\left.\left\|\hat{u}_{n}\right\|=1\right)$ and so there is a subsequence $\hat{u}_{n_{i}}$ and the limit function $u$ such that $\hat{u}_{n_{i}}(x) \rightarrow u(x)$ for every $x \in \mathcal{X}$. Hence, $\nabla \hat{u}_{n_{i}} \rightarrow \nabla u$, and so $\nabla u=0$ which yields that $u$ is a constant function on $\mathcal{X}$. On the other hand, from the boundary data $\hat{u}_{n_{i}}=0$ on $\Gamma$ we obtain that $u=0$ on $\Gamma$ and so $u=0$ on all of the graph. This contradicts the condition $\|u\|=\lim _{n_{i} \rightarrow \infty}\left\|\hat{u}_{n_{i}}\right\|=1$.

Proposition 8.2 (Maximum principle). Assume the graph $X$ is connected. Let $p(x)$ be a nonnegative function on $\mathcal{X}$, and $u$ satisfies $\mathcal{L} u+p(x) u \geq 0$ in $X \backslash \Gamma$. If $u \geq 0$ on $\Gamma$, then $u \geq 0$ in X.

Proof. Define $A^{+}:=\{x \in \mathcal{X}: u(x) \geq 0\}$ and $A^{-}:=\{x \in \mathcal{X}: u(x)<0\}$. Let $v:=\max (-u, 0)$ and multiply by the equation to get

$$
\begin{aligned}
0 \leq(\mathcal{L} u+p u, v)= & (\nabla u, \nabla v)_{\ell^{2}\left(x^{2}\right)}+(p u, v) \\
= & \frac{1}{2} \sum_{x, y \in X} w_{x y}(u(x)-u(y))(v(x)-v(y))+\sum_{x \in X} p(x) u(x) v(x) \\
= & -\frac{1}{2} \sum_{x, y \in A^{-}} w_{x y}(u(x)-u(y))^{2}-\sum_{x \in A^{-}, y \in A^{+}} w_{x y}(u(x)-u(y)) u(x) \\
& -\sum_{x \in A^{-}} p(x)(u(x))^{2} \leq 0
\end{aligned}
$$

If $A^{-} \neq \emptyset$, we can find an edge between $A^{+}$and $A^{-}$due to the connectedness of $G$. (Note that the boundary condition ensures that $A^{+} \neq \emptyset$.) But the second term in the above calculation yields that $w_{x y}=0$ for every $x \in A^{-}$and $y \in A^{+}$. Therefore, we must have $A^{-}=\emptyset$.

CAMGSD, Department of Mathematics, Instituto Superior Técnico, Lisbon, Portugal Email address: faridb.bozorgnia@tecnico.ulisboa.pt
Department of Mathematical Sciences, Sharif University of Technology, Tehran, Iran.
Email address: fotouhi@sharif.edu
Institute of Mathematics, NAS of Armenia, 0019 Yerevan, Armenia
Email address: arakelyanavetik@gmail.com
University of Caen Normandy, GREYC UMR CNRS 6072, France
Email address: abderrahim.elmoataz@unicaen.fr


[^0]:    Key words and phrases. Free boundary, Semi-supervised learning, Laplace learning.
    F. Bozorgnia was supported by the FCT fellowship SFRH/BPD/33962/2009 and by Marie Skłodowska-Curie grant agreement No. 777826 (NoMADS).

