Fractional Graph-based Semi-Supervised Learning

S. de Nigris[(1)], E. Bautista[(1)], P. Abry[(1)], K. Avrachenkov[(2)], P. Gonçalves[(1)]
[(1)]Univ Lyon, Inria, CNRS, ENS de Lyon, UCB Lyon 1, 15 parvis René Descartes, F-69342, Lyon, FRANCE
[(2)]Inria Sophia Antipolis, 2004 Route des Lucioles, Sophia-Antipolis, France
Email:{sarah.de-nigris,esteban.bautista-ruiz,patrice.abry,paulo.goncalves}@ens-lyon.fr, konstantin.avratchenkov@inria.fr

***Abstract*—Graph-based semi-supervised learning for classification endorses a nice interpretation in terms of diffusive random walks, where the regularisation factor in the original optimisation formulation plays the role of a restarting probability. Recently, a new type of biased random walks for characterising certain dynamics on networks have been defined and rely on the γ-th power of the standard Laplacian matrix $\mathbf{L}^\gamma$, with $\gamma > 0$. In particular, these processes embed long range transitions, the Lévy flights, that are capable of one-step jumps between far-distant states (nodes) of the graph. The present contribution envisions to build upon these volatile random walks to propose a new version of graph based semi-supervised learning algorithms whose classification outcome could benefit from the dynamics induced by the fractional transition matrix.** [1]

I. Introduction

Graph-based semi-supervised learning (G-SSL) has received considerable attention in recent years as an alternative approach to the popular paradigm of supervised learning. While supervised learning methods rely just in labeled data, G-SSL methods aim to exploit both labeled data and the structure imposed by the unlabeled data, naturally captured by a graph, in order to build better classifiers [1]–[3]. This is of utmost importance since nowadays large amounts of unlabeled but relationned data are readily accessible in comparison to labeled data which may be expensive to obtain. G-SSL has seen success in tasks like classification of BitTorrent content and users [4], text categorization [5], medical diagnosis [6], among others. Notably, the popular methods of *Standard Laplacian* (SL) and *PageRank* (PR), on which we will focus, have a closed form solution that allows for a probabilistic interpretation based on the theory of random walks on graphs. Viewed from this angle, the decision rule to classify unlabeled nodes relies on the expected number of visits from each class-specific random walk. In particular, SL can be then interpreted as walkers starting from unlabeled nodes and reaching nodes tagged by an expert, while PR acts conversely. Despite their remarkable performance, in cases of very few labeled points, several questions remain to be addressed such as how to deal with unbalanced scenarios or with graphs that are badly constructed in which usual G-SSL methods are known to perform poorly.

Contributions and Outline: In this article, we propose to build upon the random walk interpretation of G-SSL methods, to extend the classification principle to other types of processes, such as Lévy Flights. The dynamics of these more volatile random walks has been shown to be more efficient/rapid at diffusing over irregular graph structures [7]–[10] and therefore may, in some circumstances, address some of the limitations of G-SSL.
We start recalling in Sec. II-A the generalized formulation of G-SSL, emphasizing the special cases of SL and PR. We present the rationale behind long-range random walks and list different incarnations that lead to Lévy Flights and to biased random walks. In Sec. III, novel definitions of G-SSL using Lévy Flight-based operators are proposed as our main contribution: we call these Fractional Graph based Semi-Supervised Learning (Fractional G-SSL). Theoretical consistency and corresponding analytical solutions of these new objects are devised. Sec. IV presents numerical experiments conducted on synthetic examples, that illustrate the potential of Fractional G-SSL in terms of performance and versatility in challenging settings for which standard G-SSL techniques are known to perform poorly.

II. State of the art and related work

A. Graph-based Semi-Supervised Learning

Problem Statement Graph-based Semi-Supervised Learning techniques aim to provide a classification of data that possess a graph structure and a few pre-known labeled points.

To set our notations, the data is structured on a N nodes graph, encoded by the adjacency matrix W which is positive and symmetric, since the G-SSL framework only holds for undirected graphs. For the sake of simplicity, we take $w_{i,j} = 1$ if i and j are connected, and zero otherwise but the following discussion can flawlessly be extended to $w_{i,j} \geq 0$ [11]. The matrix D is the diagonal matrix whose entries are the nodes' degrees $D = diag(d_1 \dots d_N)$ where $d_i = \sum_j w_{i,j}$. Therefore, the operator $L = D - W$ in the Dirichlet form is, in the original formulation, the Standard (or Combinatorial) Laplacian (SL) and it is diagonalizable according to $L = Q^T \Lambda Q$.

Further, it is assumed that the data on the graph belongs to K classes: $Y \in \mathbb{R}^{N \times K}$ constitute the ground truth labels embedding expert knowledge, concretely $Y_{i,k} = 1$ if node i is known to belong to class k and zero otherwise. Our target, the $F \in \mathbb{R}^{N \times K}$ matrix, denotes the classification functions we look for and finally, the decision rule affects node i to the class k that satisfies $\text{argmax}_k F_{i,k}$.

Our approach is based on a series of works [2]–[4] presenting a generalized expression for G-SSL: this generalization comprehends the different methods, Standard and Normalized

[1]We thank CONACYT and Labex MILyon for the financial support.

ISBN 978-0-9928626-7-1 © EURASIP 2017

Laplacian (SL,NL) and Page Rank (PR) in a unique framework, as we shall display in the following. This generalized expression of the G-SSL [2] reads

$$\min_F \left\{ 2F^T D^{\sigma-1} L D^{\sigma-1} F + \mu \left(F-Y\right)^T D^{2\sigma-1} (F-Y) \right\} . \tag{1}$$

Please note that we have dropped the indices for the sake of notation lightness but, with F and Y, we intend the column vectors $F_{*,k}$ and $Y_{*,k}$. Therefore, as we see in Eq.1, G-SSL aims at finding a classification function F minimizing simultaneously a graph-dependent term, the first, and a label-dependent one. This minimization problem can be solved explicitly and the solution for the F functions takes the form

$$F^T = (1-\alpha) Y^T \left(I - \alpha D^{\sigma-1} W D^{-\sigma} \right)^{-1} , \tag{2}$$

where $\alpha = \frac{2}{2+\mu}$. This result is derived in [12] and we would like to stress that our generalization, in Sec. III, follows theirs closely. In the present context, we focus on two specific cases of the form (1) entailing two widely used G-SSL methods: the Standard Laplacian (SL) and PageRank (PR).

$\sigma = \mathbf{1}$ **- Standard Laplacian** The Standard Laplacian method can be retrieved by setting $\sigma = 1$ in (1). In this case, the minimization problem is then cast as

$$\min_F \left\{ 2F^T L F + \mu \left(F - Y\right)^T D \left(F - Y\right) \right\} , \tag{3}$$

and the expression for the F functions acquires the form

$$F^T = (1-\alpha) Y^T D \left(I - \alpha D^{-1} W \right)^{-1} D^{-1}. \tag{4}$$

$\sigma = \mathbf{0}$ **- PageRank** Another classical formulation of G-SSL, referred to as PageRank (PR), is obtained with $\sigma = 0$. In this setting, the regularization problem of (1) reads

$$\min_F \left\{ 2F^T L F + \mu \left(F - Y\right)^T D^{-1} \left(F - Y\right) \right\} , \tag{5}$$

and leads to the solution F

$$F^T = (1-\alpha) Y^T \left(I - \alpha D^{-1} W \right)^{-1} . \tag{6}$$

Probabilistic Framework From (4) and (6), it clearly emerges the connection to the probabilistic interpretation of G-SSL: the classification function F is, in fact, the result of a random walk (RW) process. Indeed, the $D^{-1}W$ matrix is the *transition matrix* of a RW and the operator $I - \alpha D^{-1} W$, at the core of both solutions, governs the dynamics of a *RW with restart*, the restarts occurring with probability $p_r = 1 - \alpha$, and starting from the ground truth labels Y. Here, it is worth noting that the μ parameter, which calibrates the "strength" of the ground truth knowledge labels Y in the optimization problem, gains a new signification with the random walks appearing in the solutions (4) and (6): the $\alpha = \frac{2}{2+\mu}$ parameter relates to the probability $p_r = 1 - \alpha$ of "refreshing" the initial conditions for the walk, coded by the labels Y [13]. So, a node i will be classified to class k if [12]

$$\sum_{p \sim k} d_p^\sigma q_{pi} > \sum_{s \sim k'} d_s^\sigma q_{si}, \ \forall k' \neq k \tag{7}$$

where q_{pi} is the (ensemble) probability that random walks starting from labeled points p in class k reach the node i, before reinitialization to the absorption state with p_r.

B. *Lévy Flights*

Casting the G-SSL problem into a random walk perspective paves the way into embedding more refined mechanisms that have been developed in random walk theory.

There has been a recent effort to embed Lévy Flights into random walks on graphs. In metric spaces, Lévy Flights have generated an impressive arborescence of research for being a vehicle of efficient and fast exploration [14]. In such spaces, the basic mechanism is very simple: every walker can perform jumps of length ℓ, drawn from a probability distribution $P(\ell)$. Thus, for particular functional choices for the $P(\ell)$, the walker can perform very long jumps and the overall space exploration benefits from these *long-range transitions*.

Now, the generalization of this approach is not straightforward on networks since they are lacking an intrinsic metric to define the "length" of a jump. To this end, various generalizations of the diffusion operator L have been proposed:

Random Walk-like operators In order for the generalization of L to endorse a RW interpretation, it is necessary for the corresponding adjacency matrix to be stochastic, i.e. with non-negative entries and the sum over the rows (or the columns) has to be zero to entail probability conservation. The two following operators satisfy this prescription, being RW operators:

$\diamond$ $\mathbf{L}^\gamma$ **with** $\mathbf{0} < \gamma \leq \mathbf{1}$ **- Fractional Laplacian**

In [8], it has been analytically demonstrated that the fractional powers of L lead, in the $0 < \gamma \leq 1$ regime, to long-range transitions on regular 1-D networks (rings). More specifically, it was demonstrated that, on rings, the transition probability from node i to node j $(\tau_\gamma)_{i \to j} \sim d_{i,j}^{-(1+2\gamma)}$. Furthermore, the long-range nature of the process was shown, by simulation, also on more general random and small-world networks.

$\diamond$ $\tilde{\mathbf{L}} \equiv \tilde{\mathbf{D}} - \tilde{\mathbf{W}}$ **- Laplacian from Biased RW**

In general, with *biased RW*, it is possible to tailor the transition probabilities according to some node property, like the degree, thus introducing a bias in the way the walk is performed [15]. Using this route, it is possible to plug "by hand" long-range transitions: this approach was explored in [7] and it relies on the construction of a transition matrix informed by the *geodesic distances* between nodes. In the first incarnation [7], the geodesic distance was taken into account weighted by a power β: therefore the transition probability from node i to j is imposed *a priori* as $\tau_{i \to j} = d_{i,j}^{-\beta} / \sum_{k \neq i} d_{i,k}^{-\beta}$. We see that this formulation clearly reconnects with the previous one of the Fractional Laplacian, albeit this link could be analytically shown just on regular networks. In [9], the latest evolution of such long range walks implements a Mellin and Laplace-like transform of the transition matrix informed with the geodesic distances, and it is shown that this approach indeed entails a more efficient diffusion as well.

III. Fractional G-SSL

A. Lévy Flight G-SSL

To generalize G-SSL, we proceed to replace in the functional (1) the L operator with its generalization $L^\gamma = Q^T \Lambda^\gamma Q$. We observe that, in order to recast (1) in a proper RW, we need in the fitting term to have some consistent diagonal matrix D_γ. Of course, when we revert to the class of L^γ operators, the matrix D_γ which originally was just the diagonal matrix whose entries were the degrees d_i has to be generalized to $(D_\gamma)_{ii} = (L^\gamma)_{ii}$. If we consider G-SSL with the generalized Laplacian L^γ we have to minimize the new functional $S(F)$

$$S(F) = 2F^T D_\gamma^{\sigma-1} L^\gamma D_\gamma^{\sigma-1} F + \mu (F-Y)^T D_\gamma^{2\sigma-1}(F-Y)$$

The application of the first optimality condition $D_F S(F) = 0$ implies, for the solution to be a global minimum, the convexity of the function $S(F)$.
Proof: To ensure that $S(F)$ is a positive or semi-positive definite function, it is convenient to express L^γ in its diagonal form $L^\gamma = Q^T \Lambda^\gamma Q$, where $\Lambda = diag(\lambda_0^\gamma, \ldots, \lambda_{N-1}^\gamma)$. Of course also the generalised laplacian L^γ can be written as a diagonal component, the D_γ diagonal matrix, and a off-diagonal one which corresponds to a new weighted adjacency matrix $(W_\gamma)_{ij} = (L^\gamma)_{ij}$ with $i \neq j$. Starting with the fitting term, we have to ensure the positivity of the generalized degree D_γ and, to this end, we use, for the sake of clarity, $F = e^i$, one vector of the canonical basis. We have

$$F^T(D_\gamma - W_\gamma)F = (d_\gamma)_i - (w_\gamma)_{i,i} = \sum_i Q_{i*}^T \Lambda^\gamma Q_{*i} = \sum_i q_i^2 \lambda_i^\gamma . \quad (8)$$

Since $(w_\gamma)_{i,i} = 0$, we therefore obtain $(d_\gamma)_i \geq 0$. The d_γ being positive, it naturally implies that their $(2\sigma - 1)$ -th power in (8) shall be as well, thus proving the semi-positivity of the fitting term. If we now consider the Dirichelet form $F^T D_\gamma^{\sigma-1} L^\gamma D_\gamma^{\sigma-1} F$ and we apply it to a general function F decomposed on the canonical basis $F = \sum_i a_i e^i$ we obtain

$$F^T D_\gamma^{\sigma-1} L^\gamma D_\gamma^{\sigma-1} F = \sum_i a_i^2 (d_\gamma)_i^{\sigma-1} Q_{i*}^T \Lambda^\gamma (d_\gamma)_i^{\sigma-1} Q_{*i} = \sum_i (a_i q_i (d_\gamma)_i^{\sigma-1})^2 \lambda_i^\gamma \geq 0 \quad \forall F.s \quad (9)$$

Since $S(F)$ is convex, we can proceed further and apply the first optimality condition $D_F S(F) = 0$ in order to obtain the F functions.
Proof The first optimality condition reads

$$2F^T D_\gamma^{\sigma-1} \left(L^\gamma + (L^\gamma)^T \right) D_\gamma^{\sigma-1} + 2\mu (F-Y)^T D_\gamma^{2\sigma-1} = 0$$

Multiplying on the R.H.S. the above equation by $D_\gamma^{1-2\sigma}$

$$2F^T D_\gamma^{\sigma-1} \left(L^\gamma + (L^\gamma)^T \right) D_\gamma^{-\sigma} + 2\mu (F-Y)^T = 0 . \quad (10)$$

Thus, substituting $L^\gamma = D_\gamma - W_\gamma$ into the previous equation

$$F^T D_\gamma^\sigma \left(2I - D_\gamma^{-1} \left(W_\gamma + (W_\gamma)^T \right) + \mu I \right) D_\gamma^{-\sigma} - \mu Y^T = 0.$$

Since W^γ is symmetric we finally arrive to

$$F^T D_\gamma^\sigma \left(2I - 2D_\gamma^{-1} W_\gamma + \mu I \right) D_\gamma^{-\sigma} - \mu Y^T = 0.$$

Therefore, we can conclude that the classification function with the generalized standard Laplacian L^γ takes the form

$$F^T = (1-\alpha) Y^T D_\gamma^\sigma \left(I - \alpha D_\gamma^{-1} W_\gamma \right)^{-1} D_\gamma^{-\sigma}, \quad (11)$$

corresponding to the generalization of the solution for F in (2). As before, the two cases of SL and PR give for F:
$\sigma = 1$ **- Fractional SL:**

$$F^T = (1-\alpha) Y^T D_\gamma \left(I - \alpha D_\gamma^{-1} W_\gamma \right)^{-1} D_\gamma^{-1}. \quad (12)$$

$\sigma = 0$ **- Fractional PR:**

$$F^T = (1-\alpha) Y^T \left(I - \alpha D_\gamma^{-1} W_\gamma \right)^{-1}. \quad (13)$$

The solutions (12) and (13) are formally, in clear symmetry with (4) and (6): the labels' attribution seems to rise from a diffusion process, driven by a new generalized "fractional transition matrix" $D_\gamma^{-1} W_\gamma$ and restarted, as before, with probability $p_r = 1 - \alpha$. However, this interpretation just holds in the $0 < \gamma \leq 1$ where, as we detail in Sec. II-B, the W_γ is a stochastic matrix and, therefore, a random walk process is still at play. On the other hand, in the $\gamma > 1$ regime, the nature of L^γ changes. We develop this point in the next Section.

As a closing note, this complete symmetry between the results in Sec. II-A and in our fractional case allows to readily extend the classification criteria of relation (7).

B. From Lévy Flights to fractional differentiation

As we explained in Sec. II-B, the key to a Lévy Flight type of diffusion dynamics on networks is to introduce *long-range transitions* that allow to explore a larger network region than the nearest proximity of a given node.

In our generalization of G-SSL, we have built upon the "fractional approach", in which the γ-th powers of L are at play and the results in Sec. III-A hold in general for $\gamma > 0$. It is worth noting, nevertheless, that our derivation can be flawlessly extended to the biased random walk Laplacian $\tilde{L}$ we defined in Sec. II-B. However, it is important to stress that the operator L^γ with $\gamma > 1$, albeit being a *non-local* operator (as we explain in the following) does no longer imply a random walk process: indeed, the generalized transition matrix W_γ is not a stochastic matrix since "negative transitions" $(W_\gamma)_{i,j} \leq 0$ can occur. Therefore, in the $\gamma > 1$ regime, our approach departs from the random-walk perspective to access a new one, more geared towards regularization.

Differential-like operators The other facet of the Laplacian operator is to naturally act as a differential operator: by application on $f \in R^N$, it gives $(Lf)_i = \sum_j w_{i,j}(f_i - f_j)$, which basically is a measure of the dissimilarity between f_i and its neighbors $f_j \in V_i$, weighted by $w_{i,j}$.

◇ L^γ with $\gamma > 1$ - Integer Powers of the Laplacian
Iterating the application of the Laplace operator thus leads to a kind of (possibly fractional) high-order differential operator. As an example, let us take L^2. Defining $\Delta_i = (Lf)_i$ we

ISBN 978-0-9928626-7-1 © EURASIP 2017

have $(L^2 f)_i = (L(Lf))_i = \sum_j w_{i,j}(\Delta_i - \Delta_j)$; thus we are now comparing Δ_i, which already accounted for node i and its neighbors V_i, to its "extended" neighbors Δ_j, which comprehend node j and its neighborhood V_j.

Furthermore, the Dirichlet form $f^T L^\gamma f$ endorses an interpretation in terms of non necessarily low-pass filters: W_γ containing both positive and negative entries, the functions f that minimize the aforementioned two-form are meant to be rapidly varying, with possibly important differences between f_i and f_j. This setting contrasts with the L operator that favors smooth functions whose differences $f_i - f_j$ are small when nodes i and j are close in the graph.

IV. Numerical Experiments

A. *Lévy Flight G-SSL :* $0 < \gamma < 1$

Experimental setup We perform numerical simulations on two classical toy examples: the *Two Moons* and the *Two Rings* datasets depicted in Fig. 1. The datasets consist of two classes of 1000 points each, with both classes very close in space. For each dataset, we build a graph by means of a radial basis function kernel of the euclidean distance, and we prune down the complete graph to its 20-Nearest Neighbors proxy. These topologies are challenging from a classification perspective because the two classes, albeit easily identifiable from their structures, are strongly intertwined. Moreover, we consider as the ground truth delivered by experts a single labeled point per class (identified by black dots in the Figures).

Results and discussion Fig. 1 illustrate the classification results obtained with Lévy Flight ($\gamma = 0.1$) SL and PR on the two datasets, while plots of Fig. 2 compare their classification accuracies with those of standard SL and PR when the restart probability p_r varies from 0 to 1. Results are averaged over 100 independent realizations of randomly chosen labeled points.

As we can see, both standard SL and PR can achieve good classification performance but only for $p_r \to 0$: when the random walks have reached their *stationary regime* which exclusively depends on the graph structure and is totally oblivious of the a priori knowledge ($\mu \to 0$ in form (1)).

On the contrary, Fractional SL and PR methods can achieve comparable performances, but for reinitialization probabilities of the random walks that privilege short duration paths and therefore strong attachment to the a priori knowledge coded in the labeled points.

To interpret these observations, in the regime of heavy reliance on the ground truth labels, the diffusion of the standard SL and PR is impeded since $p_r \to 1$ and the walk is often rebooted. On the other hand, the fractional PR and SL, because of the non-locality of the random walk, are able to ensure an efficient labels diffusion, even in this regime of high restarting probability. Therefore, fractional PR and SL can account for the data structure itself, as standard SL and PR, but they can efficiently cope with cases in which the ground truth labels, provided by experts, have a strong weight, differently from standard methods.

As a final remark, we observe that, in contrast to Fractional SL, Fractional PR is relatively insensitive to the choice of α. This is somehow reminiscent of similar conclusions that were drawn in [12] regarding the behavior of standard SL and PR when applied to less critical datasets.

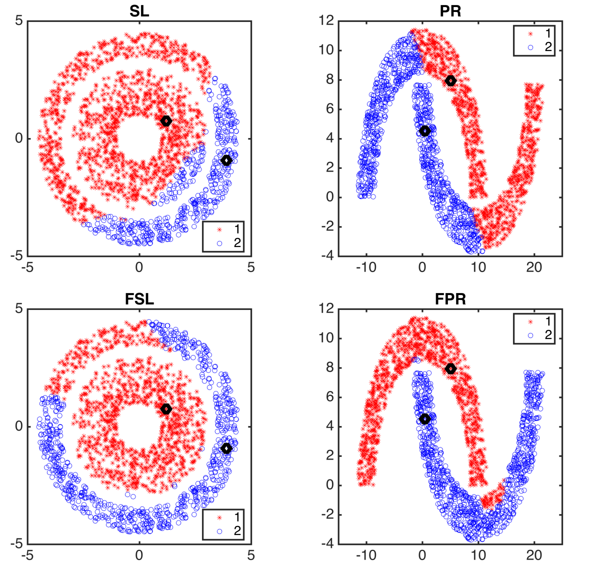

Fig. 1: Two-classes problems with different topologies: Two Rings (left), Two Moons (right). Black dots designate labeled points. Top row: classification with standard G-SSL ($p_r = 0.2$). Bottom row: classification with Fractional G-SSL ($\gamma = 0.1$, $p_r = 0.2$).

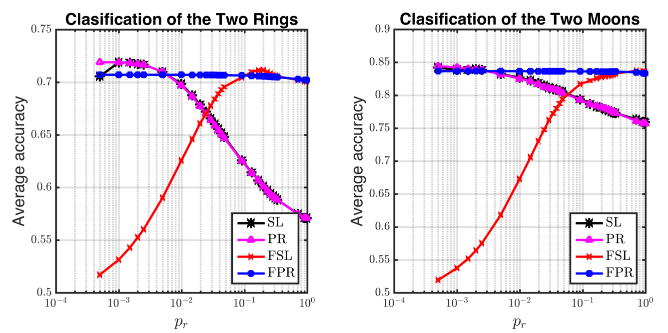

Fig. 2: Comparison of classification accuracies *vs* restart probability on the two datasets ($\gamma = 0.1$ for Fractional G-SSL)

B. *Fractional G-SSL :* $\gamma > 1$

Experimental Setup We generate a planted partition random graph model and use stochastic block models to form two classes of 1000 elements each. We denote P_1 and P_2 the probabilities of links creation within classes 1 and 2, respectively and P_{12} the interclass link probability.

Results and discussion We start setting, $P_1 = P_2 = 0.3$, $P_{12} = 0.05$ and choosing an unbalanced number of labeled points per class. Results are averaged over 100 realizations of randomly chosen labeled points. Fig. 3(a) displays the classification accuracy obtained with different values of γ for Fractional PR (similar results were obtained with Fractional SL). Clearly, there are two regimes: for the random walk regime, i.e. $0 \le \gamma \le 1$, we know from (7) that the class

ISBN 978-0-9928626-7-1

with more labeled points tends to attract the points from other classes: this behavior persists for Fractional G-SSL.
On the other hand, for $\gamma > 1$, as explained in Sec. III-B, fractional operators can avoid the completely smooth solution that would lead to the spurious dominance of one class, and they manage to distinguish the two classes even in the presence of unbalanced labeled points. For comparison purpose, remind that $\gamma = 1$ coincides with the standard PR method.

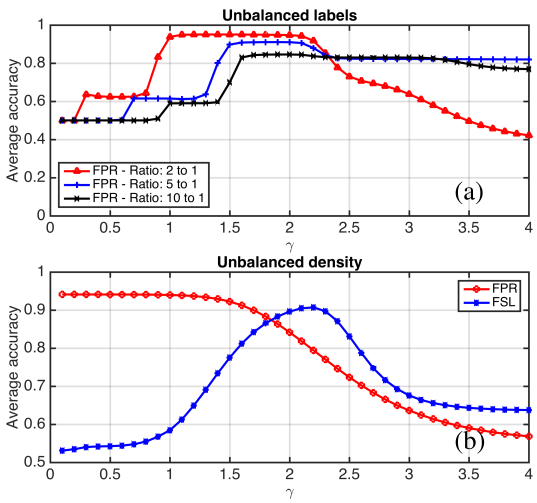

Fig. 3: Classification accuracy *vs* γ: (a) unbalanced number of labels – (b) unbalanced link density for the Planted Partition dataset.

In our second experiment, we set $P_1 = 0.3$, $P_2 = 0.1$ and $P_{12} = 0.05$, yielding unbalanced intraclass connectivities. We chose as unique label point per class, the node with higher degree. Fig. 3(b) stresses the extremely different reactions of the two Fractional methods with respect to γ in this skewed density setting: for Fractional PR, the random walk regime ($0 \leq \gamma \leq 1$) is beneficial, at odds with Fractional SL, for which the denser class dominates, hence, a low accuracy. In contrast, Fractional SL recovers good accuracy for $\gamma \sim 2$, while Fractional PR performance steadily decreases when $\gamma > 1$.

V. Conclusion

In this work, we proposed a natural extension of G-SSL that embeds random walks with *non-local* transitions, allowing for a more efficient exploration of the data structure. The specific case we address, the γ-th powers of the laplacian L, admits a random walk interpretation in the range $0 < \gamma \leq 1$ and, in the $\gamma > 1$ regime, we discussed how this operator leads to a non-smooth solution to the regularization problem.
For the theoretical side, we display in Sec. III-A, that the results on G-SSL keep their validity in our fractional setting. Albeit Fractal G-SSL is not expected to outperform classical G-SSL in most situations in terms of classification performance, our numerical results in Sec. IV clearly point to the capacity of Fractional G-SSL to *override the graph structure*, at least to some extent: thus it is able to cope with topologically skewed situations as non-localized classes or unbalanced configurations, either in terms of connectivity or cardinality of the classes. Moreover, this extra degree of freedom, induced by the fractional operator, can help taming biases due to outliers or spurious edges in the graph construction. Because they foster a more holistic exploration of the graph, the non-local transitions mitigate the classification proneness to obey local, misleading interactions. As a result too, Fractional G-SSL can account for high confidence level in the experts, by sustaining the influence of the labeled points, while guaranteeing an efficient label diffusion.

Finally, beyond the classification framework, it is very likely that Fractional G-SSL, in particular their differential-like version ($\gamma > 1$), can complement the recent attempts to apply G-SSL techniques to inpainting problems or more generally to signal recovery [16]. A first step towards this direction would be to cast Fractional G-SSL within the framework of Algebraic Graph Signal Processing as proposed in [17] or [18].

References

[1] X. Mai and R. Couillet, "The counterintuitive mechanism of graph-based semi-supervised learning in the big data regime," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017.
[2] K. Avrachenkov, P. Gonçalves, A. Mishenin, and M. Sokol, "Generalized optimization framework for graph-based semi-supervised learning," in *SIAM Data Mining*, 2012.
[3] K. Avrachenkov, P. Gonçalves, and M. Sokol, "On the choice of kernel and labelled data in semi-supervised learning methods," in *10th WS on Algorithms and Models for the Web Graph*, Harvard U., USA, 2013.
[4] K. Avrachenkov, P. Gonçalves, A. Legout, and M. Sokol, "Classification of content and users in bittorrent by semi-supervised learning methods," in *Int. Wireless Comm. and Mobile Comp. Conf.*, Cyprus, 2012.
[5] A. Subramanya and J. Bilmes, "Soft-supervised learning for text classification," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 2008, pp. 1090–1099.
[6] M. Zhao, R. H. M. Chan, T. W. S. Chow, and P. Tang, "Compact graph based semi-supervised learning for medical diagnosis in alzheimer's disease," *IEEE Signal Processing Letters*, vol. 21, no. 10, pp. 1192–1196, Oct 2014.
[7] A. Riascos and J. L. Mateos, "Long-range navigation on complex networks using lévy random walks," *Phys. Rev.E*, vol. 86, no. 5, p. 056110, 2012.
[8] ——, "Fractional dynamics on networks: Emergence of anomalous diffusion and lévy flights," *Phys. Rev. E*, vol. 90, no. 3, p. 032809, 2014.
[9] E. Estrada, J.-C. Delvenne, N. Hatano, J. L. Mateos, R. Metzler, A. P. Riascos, and M. T. Schaub, "Random multi-hopper model. super-fast random walks on graphs," *arXiv preprint arXiv:1612.08631*, 2016.
[10] T. Weng, J. Zhang, M. Khajehnejad, M. Small, R. Zheng, and P. Hui, "Navigation by anomalous random walks on complex networks," *Sci. Rep.*, vol. 6, 2016.
[11] F. R. Chung, *Spectral graph theory*. Am. Math. Soc., 1997, vol. 92.
[12] M. Sokol, "Graph-based semi-supervised learning methods and quick detection of central nodes," Ph.D. dissertation, Université de Nice, Ecole Doctorale STIC, Inria Sophia Antipolis, Maestro, April 2014.
[13] T. H. Haveliwala, "Topic-sensitive pagerank," in *Proceedings of the 11th international conference on World Wide Web*. ACM, 2002, pp. 517–526.
[14] R. Klages, G. Radons, and I. M. Sokolov, *Anomalous transport: foundations and applications*. John Wiley & Sons, 2008.
[15] R. Lambiotte, R. Sinatra, J.-C. Delvenne, T. S. Evans, M. Barahona, and V. Latora, "Flow graphs: Interweaving dynamics and structure," *Physical Review E*, vol. 84, no. 1, p. 017102, 2011.
[16] S. Chen, A. Sandryhaila, J. Moura, and J. Kovačević, "Signal recovery on graphs: Variation minimization," *IEEE Trans. on Sig. Proc.*, vol. 63, no. 17, 2015.
[17] A. Sandryhaila and J. Moura, "Discrete signal processing on graphs," *IEEE Trans. on Sig. Proc.*, vol. 61, no. 7, 2013.
[18] D. Shuman, S. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Sig. Proc. Mag.*, vol. 30, no. 3, pp. 83–98, 2013.