HAL
open science

Fractional graph-based semi-supervised learning

Sarah de Nigris, Esteban Bautista, Patrice Abry, Konstantin Avrachenkov, Paulo Gonçalves

To cite this version:

Sarah de Nigris, Esteban Bautista, Patrice Abry, Konstantin Avrachenkov, Paulo Gonçalves. Fractional graph-based semi-supervised learning. International School and Conference on Network Science, Jun 2018, Paris, France. pp.1. hal-01877500

HAL Id: hal-01877500

https://inria.hal.science/hal-01877500

Submitted on 19 Sep 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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

Graph Semi-supervised learning (gSSL) aims to classify data exploiting two initial inputs: firstly, the data are structured in a network whose edges convey information on the proximity, in a wide sense, of two data points (e.g. correlation or spatial proximity) and, second, there is a partial information on some nodes, which have previously been labelled. Thus, the classification problem is usually a balance between two terms: one diffusing the information from the labelled points to the unlabelled ones through the network and another one that constrains the solution to be similar, on the labelled nodes, to the given labels. In practice, popular SSL methods as Standard Laplacian (SL), Normalized Laplacian (NL) or Page Rank (PR), exploit those operators defined on graphs to spread the labels and, from a random walk perspective, the classification of a given point is given the maximum of the expected number of visits from one class.

Anomalous diffusion can alter the way a graph is "explored" and, therefore, it can alter classification performance. In a nushell, Lévy flights/walks are a way to create *superdiffusive* regimes: the customary rule for their ignition is to allow the walkers to perform non-local jumps, whose length is distributed according to a fat-tailed pdf with diverging second moment. Mathematically speaking, there have been several attempts to convert the Lévy flight phenomenon on networks and, in the context of gSSL, we settled to the use of *fractional operators*. Such operators are defined on the spectrum of the Standard Laplacian matrix by elevating it to some power γ: namely if the Laplacian reads $L_{i,j} = D_{i,j} - A_{i,j}$, where $D_{i,j}$ is the diagonal matrix whose diagonal are the vertexes' degrees and $A_{i,j}$ is the adjacency matrix, then $L^\gamma = Q\Lambda^\gamma Q^T$, where $\Lambda = diag(\lambda_0, \ldots, \lambda_{N-1})$ are the eigenvalues and Q is the matrix whose columns are the eigenvectors. Such operators were investigated in [1, 2] and in such works it was shown that, indeed, they connect to a Lévy Flight type of process on graphs.

Thus, in our SSL context, we cast those operators in the SSL problem in each different incarnation (SL, PR and NL) [3, 4] and we investigated the beneficial effect of such a procedure for classification, as displayed in Fig. 1.

The L^γ operator actually possesses two quite contrasting behaviours in the $0 < \gamma \leq 1$ and $\gamma > 1$ regimes: in the former, it is still a valid stochastic matrix (i.e. the off-diagonal terms are negative) and, therefore, it implies a random walk process. This random walk, in the Lévy Flight style, indeed performs long-range transitions and this enhanced and faster exploration allows for a better classification in cases in which the graph is badly constructed: for instance, in the figure, the existence of edges across the two moons spoils classification with the standard method, but we override this issue through the use of fractional methods.

On the other hand, the $\gamma > 1$ regime, the operator L^γ presents positive off-diagonal terms, so the associated graph, defined from the weights $W^\gamma_{i,j}$, has negative edges. Consequently, it is not possible anymore to draw a random walk process from it but one can still resort, for instance, to an anisotropic diffusion interpretation [5]. In this case, the effect of the operator is evident in traditionally problematic cases for standard methods, like very skewed and unbalanced ones, in which a class possesses more labels or it is denser edge-wise (Fig.1b).

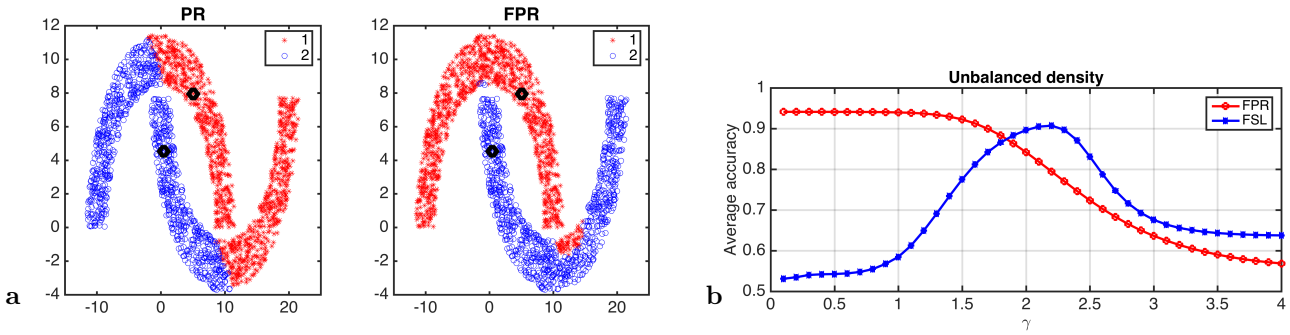

Fig. 1: (a) Classification of the Two Moons dataset with Page Rank and Fractional Page Rank. (b) Accuracy in classification for a dataset composed by connected modules with unbalanced labels.

[1] Alejandro P. Riascos and José L. Mateos. Long-range navigation on complex networks using Lévy random walks. Phys. Rev.E, 86(5):056110, 2012.
[2] Alejandro P. Riascos and José L. Mateos. Fractional dynamics on networks: Emergence of anomalous diffusion and Lévy flights. Phys. Rev. E, 90(3):032809, 2014.
[3] Sarah De Nigris, Esteban Bautista, Patrice Abry, Konstantin Avrachenkov, and Paulo Gonçalves. Fractional graph-based semi-supervised learning. In 25th European Signal Processing Conference (EUSIPCO), 2017. IEEE, 2017.
[4] Esteban Bautista, Sarah De Nigris, Patrice Abry, Konstantin Avrachenkov, and Paulo Gonçalves. Lévy flights for graph based semi-supervised classification. In 26th colloquium GRETSI, 2017.
[5] Andrew V. Knyazev. Signed laplacian for spectral clustering revisited. arXiv preprint arXiv:1701.01394, 2017.