Universal Graph Embedding Fine Tuning with Dirichlet Energy Smoothing

Tomi Wójtowicz
Poznan University of Technology
Poznan, Poland
Tomi.Wojtowicz@doctorate.put.poznan.pl

***Abstract*—Traditionally, networks are interpreted as discrete entities with nodes connected by links. In this work we propose to interpret networks as fields describing the distribution of certain properties in the multidimensional space. By following the field interpretation of networks, we introduce a universal fine-tuning of node embeddings using the concept of Dirichlet energy smoothing to obtain desirable properties of node embeddings.**

***Index Terms*—graph embeddings, representation learning, fine-tuning**

I. Introduction

The most common interpretation of networks describes them as discrete constructs consisting of nodes and links, with certain properties assigned to nodes and links. The topological structure of a network is defined only by the pattern of node connections, with no notion of abstract dimensions along which the network exists. Various methods exist to measure the relative importance of nodes, to group nodes into communities, to identify macro- and micro-structures existing in the network, to simulate dissemination processes through the network.

However, there is an alternative interpretation of networks which allows for more computation-based approaches. In science, the *field* is a spatial distribution of some physical property. Usually, we interpret a field as a function which maps every point in the considered space to a certain value. Depending on the type of the assigned value one can distinguish between various types of fields:

- *scalar field*: $F : \mathbb{R}^n \to \mathbb{R}$ assigns to each point of the abstract or physical space a scalar value.
- *vector field*: assigns to each point of the abstract or physical space a vector value,
- *tensor field*: assigns to each point of the abstract or physical space a tensor.

A field is often continuous and/or differentiable. An example of a differentiable operator that can be applied to scalar fields is the *gradient*, which produces the vector field representing the directions of the fastest changes of the values of the scalar field. For vector fields examples of differentiable operators include the divergence (the sum of partial derivatives) or the rotation of the field.

The research has received funding from the EEA Financial Mechanism 2014-2021. Project registration number: 2019/35/J/HS6/03498.

IEEE/ACM ASONAM 2022, November 10-13, 2022
978-1-6654-5661-6/22/$31.00 © 2022 IEEE

According to the field interpretation of networks, each network is a spatial distribution of properties expressed in the abstract space defined by the embedding. Nodes can be regarded as points in space where the measurements have been actually taken. Links help define the mapping of nodes into the abstract space so that the embedding captures both the topology of the network (connected nodes remain close in the embedded space) and the distribution of the property used to fine-tune the embeddings.

In this paper we introduce the idea of fine-tuning of graph embeddings using any network property (hence, universal graph embedding fine-tuning) by incorporating the Dirichlet energy of the field defined by both the embedding and the distribution of the property. This idea is the basis for the PhD project. At this stage, preliminary results are available and the paper serves as the presentation of the idea rather than the report on experimental results. As with most tasks of representation learning, the proper validation of the idea is delegated to a downstream task (e.g., link discovery, node classification, community detection). For now, the paper invites the discussion on the viability and practical usefulness of the presented method.

II. Basic notions

The *gradient* of a scalar-valued function $f(x_1, \ldots, x_n)$ is a vector field ∇f whose value at a given point $\hat{x} = [x_1, \ldots, x_n]$ is the vector of partial derivatives of f at $\hat{x}$:

$$\nabla f(\hat{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1}\hat{x} \\ \vdots \\ \frac{\partial f}{\partial x_n}\hat{x} \end{bmatrix} \tag{1}$$

The *Dirichlet energy* is the measure of the function's variability. For a set $\Omega \subseteq \mathbb{R}$ and a function $f : \Omega \to \mathbb{R}$, the Dirichlet energy of the function f is

$$E[f] = \frac{1}{2}\int_{\Omega} \|\nabla f(x)\|^2 dx \tag{2}$$

Given a graph $G =< V, E >$, $V = \{v_1, \ldots, v_n\}$, $E = \{(v_i, v_j) : v_i, v_j \in V\}$, let the embedding of the graph be the function $\mathbb{E} : V \to \mathbb{R}^k$. We will denote the embedding of a node v_i as $\bar{v_i} = \mathbb{E}(v_i)$.

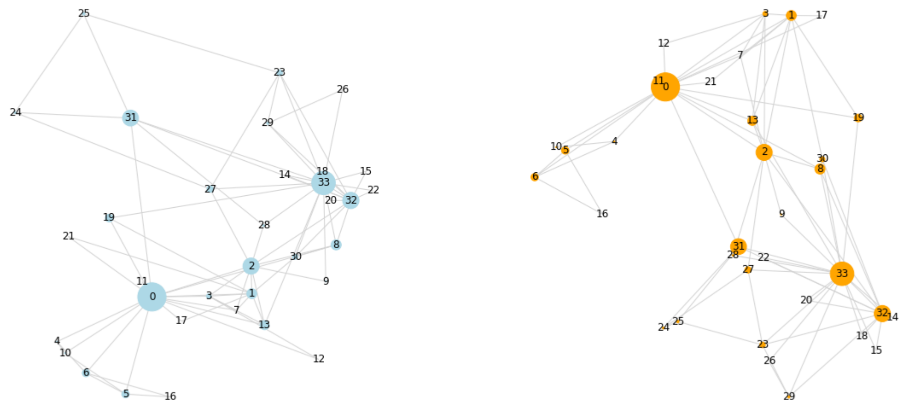

Fig. 1. Two-dimensional embeddings of the Zachary karate club network with the betweenness as the tuning function (left) without Dirichlet energy smoothing (right) with Dirichlet energy smoothing

III. Universal Fine-Tuning of Graph Embeddings

Any measure describing traditional nodes can be used as the tuning function for our fine-tuning procedure. The choice of a particular measure influences the final embedding, and thus, impacts the interpretation and downstream usability of the embedding. Without the loss of generality, let us assume that a *tuning function* $f_T : V \to \mathbb{R}$ is given which assigns a scalar value to every node $v \in V$. This tuning function defines the vector field ∇f_T such that

$$\nabla f_T(\bar{v_i}) = \bar{v_j} - \bar{v_i} \tag{3}$$

where

$$\bar{v_j} = \arg\max_k f_T(v_k) - f_T(v_i) : (v_i, v_k) \in E \tag{4}$$

In other words, the gradient ∇f_T at point $\bar{v_i}$ points to the neighbour node of v_i with the largest value of the tuning function f_T. Having defined the vector field ∇f_T we can define the objective function for training the node embeddings as the minimization of the Dirichlet energy of the gradient vector field:

$$\mathbb{E}[f_T] = \frac{1}{2} \sum_v \|\nabla f_T(v)\|^2 \tag{5}$$

This general framework can be applied to any method for graph embedding computation. For instance, when training `node2vec` embeddings one can add the Dirichlet energy as the regularization term to the traditional cross-entropy loss function to generate embeddings which smooth the distribution of the tuning function for the neighbouring nodes.

To illustrate this idea in Figure 1 we present the well-known Zachary karate club network with two-dimensional `node2vec` embeddings. The tuning function is node betweenness (the size of nodes is proportional to their betweenness). The left figure presents the traditional embeddings, and the figure on the right presents the embeddings fine-tuned with the Dirichlet energy computed over the gradients of betweenness. Of course, this figure does not prove the viability of the method in any meaningful way, the idea needs a solid experimental verification. As we have stated in the introduction, this paper serves as the invitation to discuss the idea further.

IV. Related Work

Representation learning on graphs and networks is attracting a lot of research attention, as witnessed by the plethora of embedding methods introduced over the last years [1]–[3]. Our method is best suited to be used in combination with traditional graph embedding methods, such as `node2vec` [4] or `DeepWalk` [5].

Recently, the idea of using Dirichlet energy to constrain the learning process in Graph Neural Networks has been presented in [6]. The only other proposal to adapt Dirichlet energy to neural network training that we are aware of has been presented in [7].

References

[1] H. Cai, V. W. Zheng, and K. C.-C. Chang, "A comprehensive survey of graph embedding: Problems, techniques, and applications," *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 9, pp. 1616–1637, 2018.

[2] P. Goyal and E. Ferrara, "Graph embedding techniques, applications, and performance: A survey," *Knowledge-Based Systems*, vol. 151, pp. 78–94, 2018.

[3] J. Zhou, G. Cui, S. Hu, Z. Zhang, C. Yang, Z. Liu, L. Wang, C. Li, and M. Sun, "Graph neural networks: A review of methods and applications," *AI Open*, vol. 1, pp. 57–81, 2020.

[4] A. Grover and J. Leskovec, "node2vec: Scalable feature learning for networks," in *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 855–864, 2016.

[5] B. Perozzi, R. Al-Rfou, and S. Skiena, "Deepwalk: Online learning of social representations," in *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 701–710, 2014.

[6] K. Zhou, X. Huang, D. Zha, R. Chen, L. Li, S.-H. Choi, and X. Hu, "Dirichlet energy constrained learning for deep graph neural networks," *Advances in Neural Information Processing Systems*, vol. 34, pp. 21834–21846, 2021.

[7] J. Chen, Y. Wang, C. Bodnar, P. Liò, and Y. G. Wang, "Dirichlet energy enhancement of graph neural networks by framelet augmentation,"