VISUALIZATION OF PROXIMITY AND ROLE-BASED EMBEDDINGS IN A REGIONAL LABOUR FLOW NETWORK

Sara Geremia[1], Fabio Morea[2] and Domenico De Stefano[1]

[1] University of Trieste (e-mail: sara.geremia@phd.units.it)

[2] Area Science Park (e-mail: fabio.morea@areascienepark.it)

[1] University of Trieste (e-mail: ddestefano@units.it)

ABSTRACT: This study uses graph representation learning techniques to analyze a regional labor flow network. The methods employed, VGAE and Role2Vec, reveal community structures and centrality of universities and research institutions in the network. The study demonstrates the potential of such techniques for analyzing complex networks and uncovering hidden structures.

KEYWORDS: graph representation learning; labour flow data; VGAE; Role2Vec

1 Introduction

The mobility of workers creates a network of connections that reflects the interconnectivity between employers. Such network data can reveal insights into the structure of the relationships between employers, which can be used to identify communities of employers that share geographic location, industry, and workforce characteristics (Park *et al.*, 2019). Examining regular patterns in the network is a key step in understanding the role and position of employers in the labour market. The role of large public sector organizations, such as universities, in the economic context under study can be determined by their centrality and relationship with industries employing a high number of experienced professionals (Smallbone *et al.*, 2015).

This work aims to investigate the structure of a labour flow network in Friuli Venezia Giulia (FVG) (Morea & De Stefano, 2022). The labour market flows are collected from Regional data from the Compulsory Communication on Employment (RCCE).

Understanding the structure of a labour flow network involves dealing with graph data containing rich relational information. Traditional machine learning algorithms require hand-engineered feature representation which is labor-intensive and relies on domain-specific knowledge. Representation Learning

(RL) provides an alternative approach to automatically learn to represent graph data using low-dimensional vectors (Hamilton, 2020). The learned embeddings can be used with data visualization techniques to generate representations of graphs useful for discovering communities, hub nodes, and other hidden structures.

The graph RL task can be performed to assess the potential of universities and research institutions as drivers of economic development and innovation in FVG. The interest is in investigating whether exploring the network by looking at both relational proximity and regular patterns yields valuable insights.

2 Methods

In this work two methods for graph RL are employed: Variational Graph Auto-Encoder (VGAE) (Kipf & Welling, 2016), and Role2Vec (Ahmed *et al.*, 2018). The methods differ in their approaches to preserving the structure of the graph, indeed they are based on two different definitions of node structural similarity known as structural and regular equivalence, respectively. Two nodes are structurally equivalent if they are relationally close, while they are regular equivalent if they have similar roles or occupy similar positions in the network.

VGAE is designed to preserve the structural equivalence between nodes, which means that structurally similar nodes should be mapped to similar embeddings. This is achieved by using a graph convolutional neural network as the encoder in the model, that applies convolutional filters to the graph to aggregate information from neighbouring nodes. VGAE share the encoder-decoder structure with standard autoencoders and it is built to learn the generative distribution of data. The decoder is a simple inner product between the latent representations of nodes, that enforces the reconstruction of the original graph from the learned representations. Role2Vec is used to incorporate global regular equivalence information, which means that nodes that share regular patterns in the graph should be mapped to similar embeddings. The latent representations are learned using a feature-based random walk approach, where walks find similar nodes identified by structural properties and higher-order graph features (e.g. triangles, 4-cycles, etc.). Both VGAE and Role2Vec have been shown to achieve state-of-the-art performance on various graph RL tasks, but the choice of method depends on the nature of the dataset and the task at hand. The performance evaluation in this work is conducted without true labels for a supervision task, thus it is based on visualization. The results of the models are evaluated by exploring the network latent representations re-embedded with Uniform Manifold Approximation and Projection (UMAP).

UMAP is a dimensionality reduction technique that takes local structure into account, to increase the data representation quality in terms of clusterability. For data visualization the number of dimensions is set to two.

The RCCE data include science and engineering and information and communication technology occupations over a 8 years period, from 2014 to 2022. To investigate the transfer of experienced professionals (P), weights (W) are assigned to transitions from employer A to B under the assumption that the experience gained by P while working for A is transferred to B (Morea & De Stefano, 2022). The sum of W of adjacent nodes defines the *strength*, which is an attribute included for the interpretation of the results.

3 Results and Final Remarks

The RCCE network comprised 1084 nodes and 1641 edges. Figure 1 compares the two-dimensional UMAP visualizations of node embeddings learned with the methods. The distances between nodes in the embedding space reflect structural (left panel) and regular (right panel) equivalence in the original graph. The size of the nodes indicates the strength, thus employers employing a high number of experienced professionals are shown with bigger nodes. The colour of the nodes highlights universities and research institutions in the network. The left plot captures a community structure in which two components clearly detach. It also seizes proximity between coloured nodes and big nodes. The right plot shows the universities very close together in space, along with the research institutes, illustrating the similarity of their roles in the network. The UMAP visualizations reveal community structures and the centrality of universities and research institutions. In particular, the study finds that universities are closer to each other in the embedding space when considering regular equivalence, indicating that although they share qualified employers with different organizations, structurally they play the same role.

Overall, the study highlights the potential of graph RL techniques to analyze complex networks and uncover hidden structures and patterns, which provide new outlooks on economic development and innovation. It also suggests that examining different embedding algorithms tailored to specific tasks would be valuable in addressing different research inquiries.

In encoding large graphs both methods enable the utilization of node attributes, which can provide crucial information regarding a node's community membership and role. In future applications, exploring the impact of incorporating node attributes could yield valuable insights.

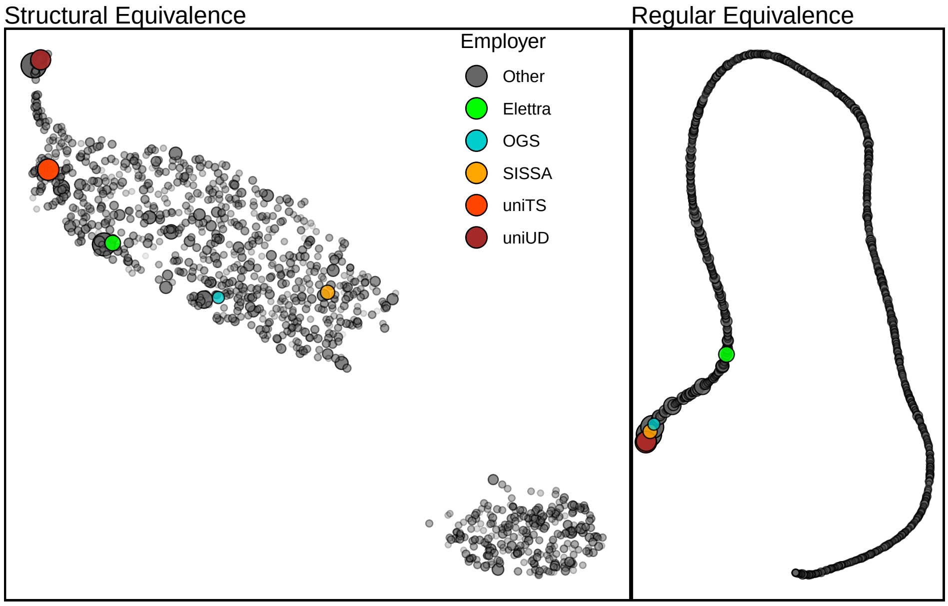

Figure 1. *Two- dimensional UMAP visualization of node embeddings generated from the RCCE network using VGAE (left panel) and Role2Vec (right panel).*

References

AHMED, N. K., ROSSI, R., LEE, J. B., WILLKE, T. L., ZHOU, R., KONG, X., & ELDARDIRY, H. 2018. Learning Role-based Graph Embeddings.

HAMILTON, W. L. 2020. Graph Representation Learning. **14**(3).

KIPF, T. N., & WELLING, M. 2016. Variational Graph Auto-Encoders.

MOREA, F., & DE STEFANO, D. 2022. Innovation patterns within a regional economy through consensus community detection on labour market network.

PARK, J., WOOD, I. B., JING, E., NEMATZADEH, A., GHOSH, S., CONOVER, M. D., & AHN, Y. Y. 2019. Global labor flow network reveals the hierarchical organization and dynamics of geo-industrial clusters. *Nature Communications*, **10**(1).

SMALLBONE, D., KITCHING, J., BLACKBURN, R., & UKCES. 2015. Anchor institutions and small firms in the UK: A review of the literature on anchor institutions and their role in developing management and leadership skills in small firms.