Measuring Polycentricity: A Whole Graph Embedding Perspective

Cheng Fu

Department of Geography, University of Zurich, Zurich, Switzerland
cheng.fu@geo.uzh.ch

Mirco Nanni

Institute of Information Science and Technologies, Italian National Research Council, Pisa, Italy
mirco.nanni@isti.cnr.it

Gevorg Yeghikyan

Scuola Normale Superiore, University of Pisa, Pisa, Italy
gevorg.yeghikyan@sns.it

Robert Weibel

Department of Geography, University of Zurich, Zurich, Switzerland
robert.weibel@geo.uzh.ch

Abstract

Polycentricity is a critical characteristic of the spatial organization of cities. Many indices have been proposed to measure the degree of morphological polycentricity or functional polycentricity. However, selecting a proper set of polycentricity indices for cities in a particular region or country still needs prior expert knowledge. This study demonstrates that whole graph embedding, as a novel and efficient computational tool, can model the city polycentricity in an integrated manner without much prior knowledge. The new method can further support visual analytics and classification very well.

Funding This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 780754.

1 Introduction

Polycentricity is a spatial organization of a city such that the city's socio-economic functions are shared by the traditional central business center (CBD) with subcenters. The spatial organization has complex social and environmental impacts. Highly polycentric cities are found to reduce household CO_2 emission but increase transportation emission in the US [12]. The spatial organization of a city can be conceptualized as a combination of the spatial distribution of its residents and their trip patterns [3]. There have been many indices to measure polycentricity either from the morphological polycentricity perspective, which measures the spatial distribution of the residents, or from the functional polycentricity perspective, measuring e.g. the transportation patterns [9, 13]. Although polycentricity can be interpreted in many different ways, transportation patterns usually form the research focus [6]. Many indices partially utilize graph analysis to model the transportation data and extract graph features, such as network density. However, these indices do not achieve consistent conclusions for the same cities of different countries in empirical studies [2].

Whole graph embedding is a network analysis technique that maps graphs with different sizes and structures into a Euclidean vector space, namely an *embedding space*, so that each graph is represented as a node in the embedding space, namely an *embedding vector*, and similar graphs are located close in the embedding space [5]. As some whole graph embedding methods can represent both edge features and node features in the same representative node, they have the potential to be used as a tool for measuring the city polycentricity combining both the spatial distribution of the residents and the transportation patterns.

This study thus explores the following research question: ***Can whole graph embedding differentiate cities with different polycentricity*?** To answer this research question, we select a proper whole graph embedding model and apply the model to transportation data from artificial cities and a real-world city.

2 Related Work

2.1 Conventional Polycentricity Indices

There are three main groups of polycentricity indices [13]: The first is rooted in social graph analysis, such as using nodality and centrality as the focal metric and using the degree of the standard deviation of a city's metric from a maximum possible standard deviation value as the metric. For example, [9] uses the total volume of inflow traffic to a place as the metric and [13] uses the centrality as the metric. The second group relies on the slope of the size-rank distribution of a specific feature, e.g., population or inflow traffic volume to a place, to measure the polycentricity, such as [14, 4]. The third group compares the real-world observation of a metric to an ideal model to measure the degree of polycentricity, e.g., [1] uses the average commute distance. All these methods use a highly aggregated statistical metric, but each of these metrics can only describe a single perspective of the transportation network, which misses out on many complex patterns. As an empirical study of Polish cities shows, the indices are not consistent, meaning that contradictory conclusions can be made for the same city [2]. [7] further discuss the issue of inconsistencies and propose that further exploration of polycentricity should take multiple critical factors into account, including the concepts of centers, polycentricity, geographical context, and more. Multiple metrics should be applied to see if the results are consistent. That means researchers should already have sufficient knowledge about the dynamics of their study subjects. Therefore the workflow can hardly be applied to a study covering many cities or over several regions or countries that have different geographical contexts.

2.2 Whole Graph Embedding Models

There are several groups of technical solutions to embed a whole graph. The first group is graph-kernel-based, such as graph2vec [15], inspired by doc2vec [11], by combining a shallow neural network with inputs sampled from random walks through the graphs. The second group applies deep learning architectures such as conventional neural networks [10, 16]. The third group utilizes spectral representation [17]. Different whole graph embedding methods are specialized to model directed/undirected or weighted/binary graphs, with/without node properties for different scenarios. Some are designed for downstream classification tasks, meaning that a classifier is required as part of the modeling. In contrast, others can be used for dimension reduction, where no training process is needed.

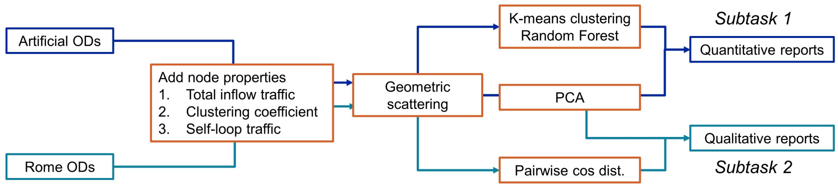

Figure 1 Overall analytical workflow

3 Methodology

To examine if the whole graph embedding method can differentiate cities with different degrees of polycentricity, we applied the method to a set of origin-destination (OD) graphs as artificial cities designed by [2] and the hourly OD graph of Rome. We conducted two subtasks (Figure 1): For the artificial cities, we applied the modeled embedding vectors as input of a classification task and a clustering task, respectively, to quantitatively examine if the embedding vectors of the cities are differentiable, denoted as Subtask 1. For the OD graphs of Rome, we applied visual analytics to investigate if the patterns of embedding vectors fit the existing knowledge and provide new knowledge, denoted as Subtask 2.

3.1 Data Collection and Preprocessing

The *original* artificial cities consist of OD matrices of six cities: three cities have five nodes as places and three cities have ten nodes. The three cities with the same number of nodes are designed as *extremely polycentric*, *intermediate*, and *extremely monocentric*, respectively. *Extended* artificial cities are made by an augmentation process using each original artificial city as a template: For each pair of two nodes, there is a possibility (denoted as PROP) to add an additional traffic volume ranging from one to a maximum volume (denoted as MAX_WEIGHT) to the template city to make a new city. By controlling the value of PROP and MAX_WEIGHT, we were able to generate new artificial cities with a similar polycentricity as their template with slightly different node and edge features. Each template city was augmented into five new cities.

The real-world OD graphs were derived from big vehicle trajectory data recorded in 2017. We only considered the origin and destination of a trip, discarding any passing-by stops. The origins and destinations were tessellated into 1-km grids and aggregated by every hour. The hourly OD traffic volumes of the year were further averaged into 24 OD graphs.

3.2 Whole Graph Embedding and Analytics

The whole graph algorithm used in this study is a geometric-scattering-based algorithm by [8]. Essentially, the algorithm applies cascading multi-layer wavelet transforms to a graph to extract scattering features. This algorithm is able to model the embedding of weighted directed graphs with node properties, which is suitable to model the OD graphs as we would like the model to describe the structural information of the transportation network and the properties of the origin and destination locations such as their demographics holistically. This algorithm is also not specialized for classification tasks as it does not require labeled data. Thus, the algorithm is suitable for exploratory tasks.

In this study, we used the traffic volume of the OD graph as the edge feature. In addition, we used the total volume of inflow traffic, self-loop traffic volume, and the clustering coefficient as the node features. Self-loop traffic volume acts as a proxy for the residential population of the place. The clustering coefficient measures how a place and its neighboring places cluster topologically, which is also used in the original study [8].

For Subtask 1 involving the artificial cities, as we already knew the label of the templates, we explored how the values of PROP and MAX_WEIGHT influence the results of the whole graph embedding. We hypothesized that with increasing values of PROP and MAX_WEIGHT, the extended cities might be more different from their template city in terms of their patterns. Exploring the locations of the extended cities and their corresponding template cities in the embedding space can also examine the sensitivity of the embedding algorithm towards small changes of the cities. We used principal component analysis (PCA) to visually interpret the embedding results. The PCs that cumulatively explain 90 % variability were kept. Then, we passed the embedding vectors to k-means clustering to check if the clustering method can differentiate cities with different polycentricity types, using the silhouette score as the metric. We also passed the embedding vectors as the features for training and used 5-fold cross-validation on random forests to examine the classification results, using average accuracy as the metric. For Subtask 2 on the Rome hourly OD graphs, we applied visual analytics using PCA and the pairwise cosine distance to understand the patterns of the embedding vectors and explored the influence of residents' trip patterns at different times on the embedding outputs given the fixed morphological polycentricity.

4 Results

As the PCA transformed visualization shows (Figure 2.a), cities with different polycentricity degrees can be differentiated visually: The extremely polycentric cities take the lower-left corner. The cities with extremely monocentric cities take the lower-right, and the intermediate cities take the upper region. For the k-means clustering (Figure 2.b), there will be five clusters using PCA-transformed vectors as input features and using the maximum value of silhouette score as the criterion to pick the best k. The best solution with $k = 5$ is almost the same as the six-type ground truth without prior knowledge of the labels. It can also be observed that the silhouette scores of specific k-values using PCA-transformed vectors are even higher than the score of ground truth labels, which might be due to the noise added during the augmentation. A similar influence of added noise can also be observed in Figure 2.a: some points are far from their templates, such as ip_ten_3 and em_five_0. From the classification results regarding the PROP ranging from 0.1 to 0.7 and the MAX_WEIGHT ranging from 1 to 10 (Figure 2.c and d), it can be observed that the artificial cities are still easy to be differentiated, while the average accuracies are almost all above 0.80. Combining the results of the three methods, it can be summarized that whole graph embedding using geometric scattering can differentiate the degree of polycentricity of the artificial cities.

For the per-hour traffic graphs of Rome, both PCA visualization and the pairwise cosine distance matrix (Figure 3) suggest the existence of three clusters, one ranging from 02:00 to 07:00, a second one from 08:00 to 11:00, and the third one from 14:00 to 18:00. The pairwise cosine distance matrix further suggests that the graphs have a clear temporal autocorrelation with their neighboring hours for most cases. This observation generally fits the common-sense perception of the commuting patterns of a city: The first and third clusters might correspond to more polycentric patterns for diverse mobility, while the second cluster might be more monocentric, corresponding to commuting hours. However, as the per-hour traffic flows do not differentiate between weekdays, weekends, and holidays so far, the patterns still need further validation. In addition, although the embedding vectors cannot provide a simple

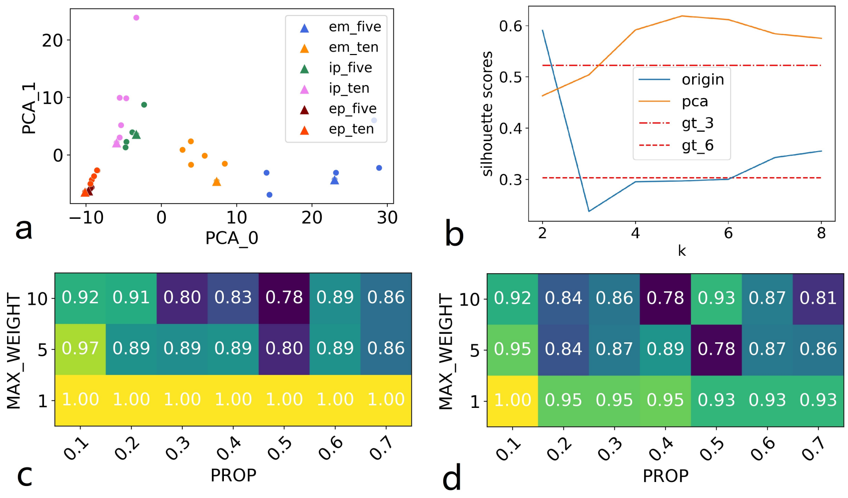

Figure 2 a) The PCA transformed embedding vectors of the six original artificial cities. The first principal component takes 66.2% of the overall variation, and the second principal component takes 20.3%. em: extreme monocentric; ip: intermediate; ep: extreme polycentric. Extended cities are points sharing the same color with their template cities. b) The silhouette scores against different k values. gt_3: using the three city types as the ground truth labels. gt_6: using the combinations of city type and city size (6 labels) as ground truth. c) The average accuracy of 5-fold cross-validation of the classification of the three city types against the PROP and MAX_WEIGHT on the extended artificial cities (36 cities). d) The average accuracy of 5-fold cross-validation of the classification of differentiating the combinations of city type and city size against the PROP and MAX_WEIGHT.

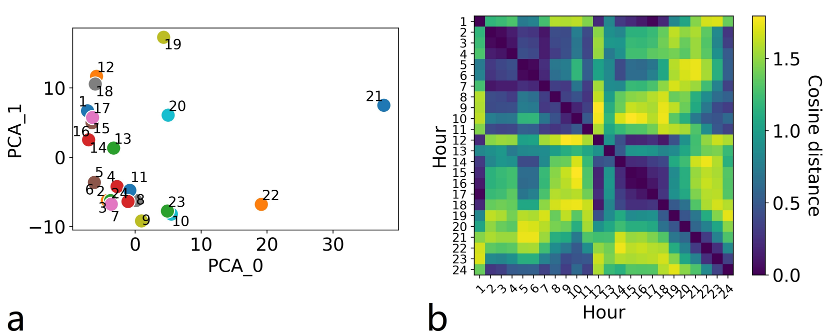

Figure 3 a) The PCA-transformed embedding vectors of per hour traffic graphs of Rome. Point colors are randomly assigned. b) Pairwise cosine distance of the embedding vectors.

and absolute scale for the polycentricity, as other existing statistical metrics do, the relative relationships can still be interpreted from the distance matrix. Nevertheless, the results still show the capacity of whole graph embedding as an analytical tool.

5 Conclusion and Future Work

Whole graph embedding is able to model the complex structure information of graphs and represent the non-Euclidean graphs in a distributed way in a Euclidean space that benefits a lot of analytical tools. With two preliminary experiments of applying a whole graph embedding algorithm to artificial traffic flow graphs and real-world traffic flow graphs, we demonstrated that the selected whole graph embedding algorithm is able to represent a complex traffic graph as an embedding vector. The embedding vectors of the cities can differentiate polycentricity through visual analytics tools and machine learning algorithms.

Although the preliminary results are encouraging, there are still questions remaining that need to be addressed in future work. For example, what node features are the best for modeling polycentricity; and how do morphological polycentricity and functional polycentricity contribute to the embedding vectors, respectively?

References

1. Shlomo Angel and Alejandro M Blei. The spatial structure of American cities: The great majority of workplaces are no longer in CBDs, employment sub-centers, or live-work communities. *Cities*, 51:21–35, 2016.
2. Bartosz Bartosiewicz and Szymon Marcińczak. Investigating polycentric urban regions: Different measures–different results. *Cities*, 105:102855, 2020.
3. Alain Bertaud. The spatial organization of cities: Deliberate outcome or unforeseen consequence? 2004.
4. Monica Brezzi and Paolo Veneri. Assessing polycentric urban systems in the OECD: Country, regional and metropolitan perspectives. *European Planning Studies*, 23(6):1128–1145, 2015.
5. Hongyun Cai, Vincent W Zheng, and Kevin Chen-Chuan Chang. A comprehensive survey of graph embedding: Problems, techniques, and applications. *IEEE Transactions on Knowledge and Data Engineering*, 30(9):1616–1637, 2018.
6. Carey Curtis. Network city: retrofitting the Perth metropolitan region to facilitate sustainable travel. *Urban Policy and Research*, 24(2):159–180, 2006.
7. Ben Derudder, Xingjian Liu, Mingshu Wang, Weiyang Zhang, Kang Wu, and Freke Caset. Measuring polycentric urban development: The importance of accurately determining the 'balance' between 'centers'. *Cities*, 111:103009, 2021.
8. Feng Gao, Guy Wolf, and Matthew Hirn. Geometric scattering for graph data analysis. In *International Conference on Machine Learning*, pages 2122–2131. PMLR, 2019.
9. Nick Green. Functional polycentricity: A formal definition in terms of social network analysis. *Urban studies*, 44(11):2077–2103, 2007.
10. Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.
11. Quoc Le and Tomas Mikolov. Distributed representations of sentences and documents. In *International conference on machine learning*, pages 1188–1196. PMLR, 2014.
12. Sungwon Lee and Bumsoo Lee. The influence of urban form on GHG emissions in the US household sector. *Energy Policy*, 68:534–549, 2014.
13. Xingjian Liu, Ben Derudder, and Kang Wu. Measuring polycentric urban development in china: An intercity transportation network perspective. *Regional Studies*, 50(8):1302–1315, 2016.
14. Evert J Meijers and Martijn J Burger. Spatial structure and productivity in US metropolitan areas. *Environment and planning A*, 42(6):1383–1402, 2010.
15. Annamalai Narayanan, Mahinthan Chandramohan, Rajasekar Venkatesan, Lihui Chen, Yang Liu, and Shantanu Jaiswal. graph2vec: Learning distributed representations of graphs. *arXiv preprint arXiv:1707.05005*, 2017.
16. Mathias Niepert, Mohamed Ahmed, and Konstantin Kutzkov. Learning convolutional neural networks for graphs. In *International conference on machine learning*, pages 2014–2023. PMLR, 2016.
17. Anton Tsitsulin, Davide Mottin, Panagiotis Karras, Alexander Bronstein, and Emmanuel Müller. Netlsd: hearing the shape of a graph. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2347–2356, 2018.