

# Graph Based Approach to Unsupervised Clustering of USA Land Cover

Benjaimin Elfner *CEAS*  
*University of Cincinnati*  
 Cincinnati, USA  
 elfnerbm@mail.uc.edu



**Abstract**—This document is a model and instructions for  $\LaTeX$ . This and the IEEEtran.cls file define the components of your paper [title, text, heads, etc.]. \*CRITICAL: Do Not Use or Math in Paper Title or Abstract.

**Index Terms**—clustering, NLCD, pattern analysis, styling, insert

## 1 INTRODUCTION

Land cover analysis is crucial to forming decisions for tasks such as climate change or land management [1]. Regions can be formed by finding areas of land that exhibit similar spatial patterns of land cover also known as land pattern types (LPTs). These regions are important to many fields since they allow generalizations to be made of the land contained which can speed up analysis [1]. For example, measurements in one part of a region could apply to all related area of land which reduces the work required to collect data.

## 2 UNDERLING DATA MINING PROBLEM

For a given area of classified land cover (represented as a image), form regions of similar land pattern types.

## 3 DATA USED

The data that will be used is the NLCD 2019 Land Cover (CONUS) dataset. This data, collected in 2019, consists of a classification for each 30x30 meter square in the contiguous 48 US states. There are 16 classes that describe contents of the land cover for each square. [2]

## 4 RELATED WORKS

Previously this problem was approached used the formation of motifs, small tiling subsections of the land cover data [1]. A co-occurrence histogram was created for each motif where pixel adjacency are categorized and counted. These histograms are combined into contiguous segments using the similarity of motifs' histogram then those segments were clustered using hierarchical clustering. The benefit of this approach is reduced computational complexity since the data can be split into sections larger than a single

pixel. The downside to this approach is the measure of similarity only considers the connections between two pixel and ignores larger patterns found in the data.

## 5 OUR APPROACH

We will approach this problem by viewing land cover data as a weighted planar graph. Each contiguous segment of land cover that shares a common class will represent the vertices and the edges will be between adjacent segments with a weighting equal to the length of the shared perimeter.

With this representation new approaches can be used to compare vertices of the graph. In this paper we will be comparing three methods. These methods were selected due to their varying methodologies, optimization methods, and experimental performance. Each method produces vector embedding for each vertex of the graph. These vectors will be clustered using hierarchical clustering with ward linkage.

### 5.1 Graph Autoencoders [3]

An autoencoder is a model that is able to take data and convert it to a different representation of the data. Most of the time this new representation has reduced dimensionality or is able to represent the original data in a new format that is more conducive to other data mining tasks. The strength of an autoencoder is judged on its ability to recreate the input from the reduced form. A graph autoencoder is able to convert a vertex to a vector representation.

### 5.2 GraRep [4]

GraRep creates a vector of step representations for each vertex on a graph. The  $k$ th value in the vector encodes information about graph  $k$  steps from the vertex.

### 5.3 LINE [5]

LINE creates vector representations for each vertex but unlike GraRep, the vector only represents a single  $k$  steps from the vertex.

## 6 EVALUATION

The resulting trees for each embedding algorithm can be split such that a desired number of regions are formed. Due to this fact as well as the methods being unsupervised means the resulting regions do not have any human defined meaning nor can the clusters be quantitative compared a ground truth. Therefore, the assessment will be based solely on the usefulness of the clusters. This will be done by examining the patterns found in each cluster qualitatively. The resulting clusters structure will also be compared to regions created by human conducted surveys.

## REFERENCES

- [1] J. Niesterowicz, T. Stepinski, and J. Jasiewicz, "Unsupervised regionalization of the united states into landscape pattern types," *International Journal of Geographical Information Science*, vol. 30, no. 7, pp. 1450–1468, 2016. [Online]. Available: <https://doi.org/10.1080/13658816.2015.1134796>
- [2] (2019) Multi-Resolution Land Characteristics (MRLC) Consortium nlcd 2019 land cover (conus). [Online]. Available: <https://www.mrlc.gov/data/nlcd-2019-land-cover-conus>
- [3] G. Salha-Galvan, R. Hennequin, and M. Vazirgiannis, "Keep it simple: Graph autoencoders without graph convolutional networks," *ArXiv*, vol. abs/1910.00942, 2019.
- [4] S. Cao, W. Lu, and Q. Xu, "Grarep: Learning graph representations with global structural information," *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, 2015.
- [5] J. Tang, M. Qu, M. Wang, M. Zhang, J. Yan, and Q. Mei, "Line: Large-scale information network embedding," *Proceedings of the 24th International Conference on World Wide Web*, 2015.