

Adam Brown

Justin Owens

Lucas Welton

Matthew Eppard

# NetLSD: Hearing the Shape of a Graph

## Problem to be Considered

The primary focus of this paper is determining the structural similarity between two graph structures, each given by  $G = (V, E)$ , where  $V$  is the set of nodes and  $E$  is the set of edges that connect these nodes. Currently, there does not exist an algorithm to efficiently compare two graph structures that meets the following criteria: **Permutation Invariance**, that is, the ability to interpret the similarity regardless to the order that the nodes are presented in, **Scale Adaptivity**, or the ability to compare characteristics of two graphs at both a local (node) or global (graph) level, and finally **Size Invariance**, the ability determine structural similarities between graphs without regard to their difference in size.

## Algorithm to be Implemented

NetLSD (short for Network Laplacian Spectral Descriptor) aims to rectify this issue by creating an efficient algorithm that fulfills all of the above criteria. The algorithm, based in spectral graph theory, uses a graph representation given by encoded trace signatures (imagine measuring the diffusion of heat or mechanical waves from one node to another) to provide a

distance function between 2 graphs that is *pseudometric*, that is, satisfying both symmetry & the triangle inequality. The algorithm executes in constant time. This distance function can be used to aid in general data mining tasks such as clustering and nearest-neighbor classification.

## **Datasets to be Evaluated**

The authors used a total of 14 data sets: seven focused on bioinformatics, and seven datasets focused on social media. These seven bioinformatics datasets are labelled MUTAG, PTC, PROTEINS, NCI1, NCI109, ENZYMES, and D&D. The social media sets were pulled from 3 social networks: Collab, IMDB, and Reddit. These are extrapolated into COLLAB, IMDB-B, IMDB-M, REDDIT-S, REDDIT-M, REDDIT-L (focused on small, medium, and large populations, respectively), and finally REDDIT-X. We will be evaluating this algorithm based on the same datasets, for maximum accuracy.

## **Data Retrieval Methods**

As this data was originally collected and referenced by the authors of this research paper, we will be contacting the authors to ask if we could use their data. Should this fail, we will utilize social media APIs to collect our own data, as well as finding our own information for bioinformatics.

## **Evaluation Measures**

To collect the data, we will be repeating the process performed by the original researchers, by executing the algorithm 100 times for each data set and averaging the results. We

will then compare the results of executing NetLSD on small datasets to the results of similar graph comparison methods, namely NetSimile and FGSD, to determine the quality of our distance approximation. Finally, as NetSimile and FGSD do not scale to large datasets and therefore are not feasible to compute, we will analyze the results of executing NetLSD on large datasets (REDDIT-L) as standalone data.

## **References**

Anton Tsitsulin, Davide Mottin, Panagiotis Karras, Alex Bronstein, and Emmanuel Müller.

2018. NetLSD: Hearing the Shape of a Graph. KDD 2018. 1-10.

Christos Faloutsos, Danai Koutra, Michele Berlingerio, and Tina Eliassi-Rad. 2012. NetSimile: A

Scalable Approach to Size-Independent Network Similarity. 1-12.

Christos Faloutsos, Danai Koutra, Michele Berlingerio, and Tina Eliassi-Rad. 2013. Network

Similarity via Multiple Social Theories. ASONAM'13. 1-2.

Pinar Yanardag and SVN Vishwanathan. 2015. Deep graph kernels.

KDD 2015. 1365–1374.

Saurabh Verma and Zhi-Li Zhang. 2017. Hunt For The Unique, Stable, Sparse And Fast Feature

Learning On Graphs. NIPS 2017 87–97.