

3400 N. Charles St.
Whitehead Hall 105C
Baltimore, MD 21218
USA

November 22, 2011

TPAMI Transactions Assistant
IEEE Computer Society
PO Box 3014
Los Alamitos, CA 90720-1314
USA

Dear Sir or Madam:

We are delighted to submit to you two closely related manuscripts: (1) “*Shuffled Graph Classification: Theory and Connectome Applications*” and (2) “*Fast Inexact Graph Matching with Applications in Statistical Connectomics*”. These two manuscripts are the final two-thirds of a trilogy of our work on graph classification. This trilogy began with a manuscript currently under review at IEEE TPAMI entitled “*Graph Classification using Signal Subgraphs: Applications in Statistical Connectomics*”.

The representation of data as graphs is gaining in popularity, as estimating connectedness between entities is becoming increasingly popular. This is true in fields as diverse as sociology, economics, as well as neuroscience, which is our motivating application. Specifically, “connectomics” is an emerging field dedicated to studying whole organism neural circuitry. While the data are burgeoning, rigorous statistical analyses of these data remain in its infancy. We are particularly interested in graph classification, that is, predicting the class of a graph based on a collection of training data. To that end, we are developing a suite of tools for such purposes that are (i) amenable to statistical theory and (ii) have improved performance on real world data.

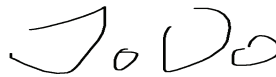
The signal subgraph classifier considers the simplest scenario of interest, building a parametric classifier on graphs with labeled vertices. To complement that work, we also investigate nonparametric classifiers on graphs with vertices lacking labels. These two strategies collectively form the bounds on approaches to classification on graphs.

In the first manuscript, we introduce the notion of a shuffled graph, that is, a graph whose vertices have been shuffled. This is distinct from an unlabeled graph, which we define as the set of isomorphic graphs. Since our data are often shuffled, we investigate via theory the extent to which shuffling degrades classification performance. Moreover, we develop a non-parametric classifier that utilizes a graph matching algorithm. Because graph matching is \mathcal{NP} -

complete, we develop a fast inexact graph matching algorithm, which is the subject of the second manuscript. Utilizing the algorithm developed therein as a subroutine, we were able to classify graphs with polynomial space and time complexity, although more naïve strategies require exponential space and time complexity.

The development of the second manuscript was partially inspired by a very nice manuscript that had recently been published in IEEE TPAMI called “A Path Following Algorithm for the Graph Matching Problem”. That manuscript described a novel algorithm for graph matching with impressive performance on a standard set of benchmarks, with $O(n^3)$ time complexity (besting the previous state of the art on 14 of 16 tests). The algorithm that we developed has the same time complexity, but superior performance on all 16 tests, superior even to the other algorithm that outperformed the PATH algorithm on two tests.

With best regards,

A handwritten signature in black ink, appearing to read 'Jo Vo', with a stylized flourish at the end.

Joshua T. Vogelstein