Ranking and Sparsifying a Connection Graph

Fan Chung¹,² and Wenbo Zhao²

¹ Department of Mathematics
² Department of Computer Science and Engineering
University of California, San Diego
La Jolla, CA 92093, USA
{fan,w3zhao}@ucsd.edu

Abstract. Many problems arising in dealing with high-dimensional data sets involve connection graphs in which each edge is associated with both an edge weight and a d-dimensional linear transformation. We consider vectorized versions of the PageRank and effective resistance which can be used as basic tools for organizing and analyzing complex data sets. For example, the generalized PageRank and effective resistance can be utilized to derive and modify diffusion distances for vector diffusion maps in data and image processing. Furthermore, the edge ranking of the connection graphs determined by the vectorized PageRank and effective resistance are an essential part of sparsification algorithms which simplify and preserve the global structure of connection graphs.

1 Introduction

In this paper, we consider a generalization of graphs, called connection graphs, in which each edge of the graph is associated with a weight and also a “rotation” (which is a linear orthogonal transformation acting on a d-dimensional vector space for some positive integer d). The adjacency matrix and the discrete Laplace operator are acting on the space of vector-valued functions (instead of the usual real-valued functions) and therefore can be represented by matrices of size \(dn \times dn\) where \(n\) is the number of vertices in the graph.

Connection graphs arise in numerous applications, in particular for data and image processing involving high-dimensional data sets. To quantify the affinities between two data points, it is often not enough to use only a scalar edge weight. For example, if the high-dimensional data set can be represented or approximated by a low-dimensional manifold, the patterns associated with nearby data points are likely to related by certain rotations [29]. There are many recent developments of related research in cryo-electron microscopy [15,28], angular synchronization of eigenvectors [11,27] and vector diffusion maps [29]. In many areas of machine learning, high-dimensional data points in general can be treated by various methods, such as the Principle Component Analysis [18], to reduce vectors into some low-dimensional space and then use the connection graph with rotations on edges to provide the additional information for proximity. In computer vision, there has been a great deal of recent work dealing with trillions

© Springer-Verlag Berlin Heidelberg 2012
of photos that are now available on the web [2]. Feature matching techniques [24] can be used to derive vectors associated with the images. Then information networks of photos can be built which are exactly connection graphs with rotations corresponding to the angles and positions of the cameras in use. The use of connection graphs can be further traced to earlier work in graph gauge theory for computing the vibrational spectra of molecules and examining the spins associated with vibrations [9].

Many information networks arising from massive data sets exhibit the small world phenomenon. Consequently the usual graph distance is no longer very useful. It is crucial to have the appropriate metric for expressing the proximity between two vertices. Previously, various notions of diffusion distances have been defined [29] and used for manifold learning and dimension reduction. Here we consider two basic notions, the connection PageRank and the connection resistance, (which are generalizations of the usual PageRank and effective resistance). Both the connection PageRank and connection resistance can then be used to define correlations between vertices in the connection graph. To illustrate the usage of both metrics, we derive edge ranking using the connection PageRank and the connection resistance. In the applications to cryo-electron microscopy, the edge ranking can help eliminate the superfluous or erroneous edges that appear because of various “noises”.

The notion of PageRank was first introduced by Brin and Page [7] in 1998 for Google’s Web search algorithms. Although the PageRank was originally designed for the Web graph, the concepts work well for any graph for quantifying the correlations of pairs of vertices (or pairs of subsets) in any given graph. There are very efficient and robust algorithms for computing and approximating PageRank [3, 6, 17]. In this paper, we further generalize the PageRank for connection graphs and give efficient and sharp approximation algorithms for computing the connection PageRank.

The effective resistance plays a major role in electrical network theory and can be traced back to the classical work of Kirchhoff [22]. Here we consider a generalized version of effective resistance for the connection graphs. To illustrate the usage of connection resistance, we examine a basic problem on graph sparsification. Graph sparsification was first introduced by Benczúr and Karger [5, 19, 20, 21] for approximately solving various network design problems. The heart of the graph sparsification algorithms is the sampling technique for randomly selecting edges. The goal is to approximate a given graph \( G \) on \( n \) vertices by a sparse graph \( \tilde{G} \), called a sparsifier, with fewer edges on the same set of vertices such that every cut in the sparsifier \( \tilde{G} \) has its size within a factor \((1 \pm \epsilon)\) of the size of the corresponding cut in \( G \) for some constant \( \epsilon \). Spielman and Teng [30] constructed a spectral sparsifier with \( O(n \log^c n) \) edges for some large constant \( c \). In [33], Spielman and Srivastava gave a different sampling scheme using the effective resistances to construct an improved spectral sparsifier with only \( O(n \log n) \) edges. In this paper, we will construct the connection sparsifier using the weighted connection resistance.