MULTI-CENTRALITY GRAPH SPECTRAL DECOMPOSITIONS AND THEIR APPLICATION TO CYBER INTRUSION DETECTION

Pin-Yu Chen⋆ Sutanay Choudhury† Alfred O. Hero III*, Fellow, IEEE

⋆Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, USA {pinyu, hero}@umich.edu
†Pacific Northwest National Laboratory Sutanay.Choudhury@pnnl.gov

ABSTRACT

Many modern datasets can be represented as graphs and hence spectral decompositions such as graph principal component analysis (PCA) can be useful. Distinct from previous graph decomposition approaches based on subspace projection of a single topological feature, e.g., the Fiedler vector of centered graph adjacency matrix (graph Laplacian), we propose spectral decomposition approaches to graph PCA and graph dictionary learning that integrate multiple features, including graph walk statistics, centrality measures and graph distances to reference nodes. In this paper we propose a new PCA method for single graph analysis, called multi-centrality graph PCA (MC-GPCA), and a new dictionary learning method for ensembles of graphs, called multi-centrality graph dictionary learning (MC-GDL), both based on spectral decomposition of multi-centrality matrices. As an application to cyber intrusion detection, MC-GPCA can be an effective indicator of anomalous connectivity pattern and MC-GDL can provide discriminative basis for attack classification.

1. INTRODUCTION

Many real-world data ranging from physical systems, social interactions, network flows, knowledge graphs to biological and chemical reactions are often represented as graphs, especially for anomaly or community detection [1–9] and graph signal processing [10–14]. Dimensionality reduction methods on graphs allow one to decompose a graph into principal components using a spectral decomposition of the graph adjacency or graph Laplacian matrix. In this paper we propose a general framework to dimensionality reduction based on spectral decomposition of a matrix composed of many different graph centrality statistics. This general framework leads to a single-graph decomposition method that extends graph principal components analysis (PCA) and a graph-ensemble decomposition method that extends dictionary learning. These methods are applicable to both directed and undirected graphs with edge weights and are based on a spectral decomposition, specifically the singular value decomposition (SVD), of a matrix composed of multiple graph centrality statistics. The proposed methods are denoted multi-centrality graph PCA (MC-GPCA) and multi-centrality graph dictionary learning (MC-GDL), respectively. By integrating multiple descriptions of graph centrality, the proposed methods provide graph community detection and graph structure learning that are significantly more robust to noise and variation affecting graph connectivity structures.

In [15], a kind of graph PCA is performed on the distance matrix of average commute time between nodes. In [16], PCA can also be performed on the graph Laplacian matrix of nodal similarities. In [17], the graph Laplacian matrix is used as a smooth regularization function for robust PCA on graphs. In [18, 19], PCA is performed on the matrix of origin-destination traffic. In [20–22], dictionary learning methods for graph signals are proposed based on the graph Laplacian matrix. Dictionary learning, also known as sparse coding, linear unmixing and matrix factorization, has been applied to collections of images, audio, and graph signals to learn low dimensional representations that give a sparse approximation to the entire collection. Dictionary learning finds a low rank factored-matrix approximation to the observation matrix, whose columns span this collection. Many different methods for this approximation problem have been proposed [23, 24]. Among the simplest methods is the K-SVD approach [25] which uses a spectral decomposition to determine the best low rank approximation to the observed matrix. For the purposes of illustration, in this paper we adopt this latter spectral approach for learning a dictionary spanning an ensemble of graphs.

More often graph PCA and graph dictionary learning approaches start with a set of raw multivariate data samples, create a similarity (or dissimilarity) graph of the data samples, and aim to learn a low-dimensional or sparse representation of the original multivariate dataset. When applied to graph data, these methods are often limited to graphs that are weighted, undirected and connected, which may not be feasible for applications such as cyber network data analysis. Furthermore, these methods often accomplish graph decomposition based on a single measure of centrality, e.g., betweenness centrality [26], closeness centrality [27], ego centrality [28], or eigenvector centrality [1]. In this paper we introduce graph spectral decomposition methods that combine multiple centrality measures such as graph walk statistics and graph distances as structural features and apply them to different graph types including weighted, directed and disconnected graphs. The proposed MC-GPCA method decomposes a single graph utilizing multiple centrality features, achieving dimensionality reduction and feature decorrelation of the graph. The proposed MC-GDL performs dictionary learning across a population of graphs using multiple centrality features to learn the atoms of the dictionary and the corresponding coefficients to represent each individual graph in terms of its projection onto the dictionary. Applying our approach to cyber intrusion detection, we use MC-GPCA to define a structural difference score (SDS) that reflects structural variations within a graph and we use MC-GDL to learn discriminative structural atoms for classifying the presence of cyber attacks.
2. STRUCTURAL FEATURE EXTRACTION ON GRAPHS

Here we describe three categories of generic structural features that can be extracted from a graph, namely graph walk statistics, centrality measures and internode distances. The utility of the introduced features with respect to different graph types, including weighted, directed and disconnected graphs, is summarized in Table 1. While not investigated in this paper, application-specific features such as website hit rates, social interaction frequency, source-destination traffics can also be leveraged as structural features. Without loss of generality a graph \( G = (V, E) \) can be characterized by two \( n \times n \) matrices \( A \) and \( W \) representing the adjacency and weight matrix, respectively, where \( V \) (\( E \)) is the set of nodes (edges), and \( n \) is the total number of nodes (i.e., graph size). \( A \) is a binary matrix such that its entry \( [A]_{ij} = 1 \) if there is an edge connecting from node \( i \) to node \( j \), and \( [A]_{ij} = 0 \) otherwise. Throughout this paper we consider graphs with nonnegative edge weights such that \( W \) is a nonnegative matrix, where its entry \( [W]_{ij} \geq 0 \) if \( [A]_{ij} = 1 \), and \( [W]_{ij} = 0 \) otherwise.

2.1. Graph walk statistics

Graph walk statistics include commute time and cover time [29], graph diffusion [30], hitting times [31], and hop walks. In this paper we focus on hop walk statistics. An \( h \)-hop walk of a node on a graph is a path starting from the node and traversing through (possibly repeated) \( h \) edges. An \( h \)-hop walk weight is defined as the sum of edge weights of the corresponding path. We consider the number and total weight of \( h \)-hop walks of each node as features since they entail the structural information of nodal reachability relative to its \( h \)-hop vicinity. In principle one should extract graph walk statistics from \( h = 1 \) to at least \( h = \text{graph diameter} \) hops as structural features, where graph diameter is the largest shortest path hop count between any node pairs in all connected components of a graph. We propose an efficient iterative computation method to incrementally computes these two structural features with respect to the hop count number \( h \):

\[ a^{(h+1)} = A \cdot a^{(h)} \]

\[ a^{(h+1)} \equiv A \cdot a^{(h)} \cdot 1_n = A \cdot a^{(h)} \]

- Iterative computation of total \( h \)-hop walk weight

Let \( W^{(h)} \) be an \( n \times n \) matrix such that its entry \( [W^{(h)}]_{ij} \) is the sum of all \( h \)-hop walk weights from node \( i \) to node \( j \). Then we have

\[ [W^{(h+1)}]_{ij} = \sum_{k \in V} [W]_{ik} \cdot [A]_{kj} \cdot [W^{(h)}]_{kj} \cdot A_{ik} \]

\[ = \sum_{k \in V} [W]_{ik} \cdot [A]_{kj} + \sum_{k \in V} [W^{(h)}]_{kj} \cdot A_{ik} \]

\[ = [WA]_{ij} + [AW^{(h)}]_{ij} \]

where we use \( [W]_{ik} \cdot [A]_{kj} = [W]_{ki} \cdot [A]_{kj} \). Let \( w^{(h)} = W^{(h)} \cdot 1_n \) denote a column vector such that its entry \( [w^{(h)}]_{i} \) is the total \( h \)-hop walk weight starting from node \( i \). Then \( w^{(h+1)} \) can be computed by

\[ w^{(h+1)} = [WA] \cdot w^{(h)} \cdot 1_n = w^{(h+1)} \]

Table 1: Utility of the introduced structural features.

<table>
<thead>
<tr>
<th>Feature / Graph Type</th>
<th>Weighted</th>
<th>Directed</th>
<th>Disconnected</th>
</tr>
</thead>
<tbody>
<tr>
<td># of ( h )-hop graph walks</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>total ( h )-hop walk weight</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>degree</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>betweenness</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>closeness</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>eigenvector centrality</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>ego</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>LFVC</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>graph distance</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

2.2. Centrality measures

A centrality measure is a quantity that evaluates the level of importance or influence of a node in a graph and it reflects certain topological characteristics. Here we introduce several centrality measures, which will be used in the sequel to define feature sets associated with a graph or a set of graphs.

- Degree. Degree is defined as the number of edges associated with a node. It can be extended to directed graphs by considering the number of edges connecting to (from) a node as in-degree (out-degree).

- Betweenness [26]. Betweenness is the fraction of shortest paths passing through a node relative to the total number of shortest paths in the graph. It is infeasible for disconnected graphs since it is based on shortest path distance. The betweenness of node \( i \) is defined as

\[ \text{betweenness}(i) = \sum_{k \in V \setminus \{i\}} \sum_{j \in V \setminus \{i\}} \frac{\sigma_{kj}(i)}{\sigma_{kj}} \]

where \( \sigma_{kj} \) is the total number of shortest paths from \( k \) to \( j \) and \( \sigma_{kj}(i) \) is the number of such shortest paths passing through \( i \).

- Closeness [27]. Closeness is associated with the shortest path distances of a node to all other nodes. Let \( \rho(i, j) \) denote the shortest path distance between node \( i \) and node \( j \) in a connected graph. Then

\[ \text{closeness}(i) = \frac{1}{\sum_{j \in V \setminus \{i\}} \rho(i, j)} \]

- Eigenvector centrality [1]. Eigenvector centrality of node \( i \) is the \( i \)-th entry of the eigenvector associated with the largest eigenvalue of the weight matrix \( W \). It is defined as

\[ \text{eigenvector centrality}(i) = \lambda_{\text{max}}^{-1} \sum_{j \in V} [W]_{ij} \xi_j \]

where \( \lambda_{\text{max}} \) is the largest eigenpair of \( W \).

- Ego centrality [28]. Ego centrality can be viewed as a local version of betweenness that computes the shortest paths between its neighboring nodes. Let \( d_i \) denote the degree of node \( i \), \( W(i) \) denote the \( (d_i + 1) \times (d_i + 1) \) local weight matrix of node \( i \), \( I \) be the identity matrix, and let \( \odot \) denote entrywise matrix product. Ego centrality is defined as

\[ \text{ego}(i) = \sum_{k \in V \setminus \{i\}, j \in V} \frac{1}{[W^2(i) \odot (I - W(i))]_{kj}} \]

- Local Fiedler Vector Centrality (LFVC) [32]. LFVC is a centrality measure that evaluates the structural importance of a node regarding graph connectivity. Let \( y \) denote the eigenvector associated...
with the smallest nonzero eigenvalue of the graph Laplacian matrix. LFVC is defined as

\[ \text{LFVC}(i) = \sum_{j \in \mathcal{N}_i} (|y_i - |y_j|)^2, \]  

where \( \mathcal{N}_i \) is the set of nodes connecting to or from \( i \) (i.e., neighbors).

### 2.3. Graph distances to a set of reference nodes

We propose to use graph distances of each node to a set of reference nodes as structural features that compensate the insufficiency of graph walk statistics and centrality measures when one performs MC-GPCA on graphs with high structural symmetry. For example, consider a star-like graph where the central node is a singleton and each leaf node is an identical clique (i.e., a complete graph). All edges in the graph are undirected and have identical weight. Therefore this graph has high structural symmetry and apparently the nodes of identical structural property (e.g., connected to the central node or not) have the same graph walk statistics and centrality measures. To resolve the ambiguity of graph walk statistics and centrality measures due to high structural symmetry in graphs we use the shortest path distance of each node to the selected \( r \) reference nodes as the \( r \) additional structural features. In the example of the star-like graph with high structural symmetry, if \( r = 1 \) then selecting any but the central node as a reference node can yield distinguishable structural features due to difference in shortest path distance to the reference node. The reference nodes are selected according to a user specified criterion, e.g., the nodes of maximal degrees.

### 3. METHODOLOGY

The extracted centrality features introduced in Sec. 2 can be represented as an \( n \times p \) matrix \( X \), where \( n \) is the graph size, \( p \) is the number of extracted features and each column of \( X \) corresponds to a particular centrality feature that is normalized to have unit norm. The multi-centrality feature matrix \( X \) is then centered by subtracting the row-wise empirical average from each row.

#### 3.1. Multi-centrality graph PCA (MC-GPCA)

In analogy to standard graph PCA, which is applied to the graph Laplacian matrix, MC-GPCA is PCA applied to \( X \). PCA can be formulated as finding an orthonormal transformation \( Q \) on \( X \) such that after transformation the multi-centrality feature matrix \( X \) is represented by an \( n \times q \) matrix \( Y = XQ \) that maximally preserves the total data variance \( \text{trace}(Y^T Y)/n \), where \( \text{trace}(\cdot) \) denotes the sum of diagonal entries of a matrix and \( Q \) is a \( p \times q \) matrix such that \( Q^T Q = I \). Such a matrix \( Q \) can be obtained by solving the \( q \) right singular vectors associated with the \( q \) largest singular values of \( X \), which is denoted by a \( p \times q \) matrix \( V_q \). Moreover, the total variance of \( Y \) is equivalent to the sum of the \( q \) squared largest singular values of \( X \) divided by \( n \). Therefore using MC-GPCA we obtain \( n \)-dimensional coordinates representing structural scores with respect to the \( q \) principal components (i.e., columns of \( V_q \)). The algorithm for MC-GPCA is summarized in Algorithm 1.

#### 3.2. Structural difference score (SDS)

We use these structural coordinates (i.e., each row of \( Y = XV_q \)) to define a structural difference score (SDS) for each node in a graph. The SDS of node \( i \) is associated with the total squared Euclidean distance to its neighboring nodes \( \mathcal{N}_i \) and its number of edges (i.e., degree \( d_i \)), which is defined as

\[ \text{SDS}(i) = \frac{\sum_{j \in \mathcal{N}_i} ||\text{row}_i(Y) - \text{row}_j(Y)||^2}{d_i + 1}. \]

where \( \text{row}_i(Y) \) denotes the \( i \)-th row of \( Y \), \( \| \cdot \| \) denotes Euclidean distance, and the denominator \( d_i + 1 \) is such that the SDS of a singleton node is well-defined.

#### 3.3. Multi-centrality graph dictionary learning (MC-GDL)

Consider the case where a set of graphs \( \{G_i\}_{i=1}^\ell \) is available, each possibly being of different graph size and connectivity pattern, e.g., data from a cyber network at different time instances. Multi-centrality graph dictionary learning (MC-GDL) is proposed to learn a sparse structure representation of \( \{G_i\}_{i=1}^\ell \) by finding a dictionary \( D \) consisting of \( K \) atoms (columns of \( D \)) and an associated sparse coefficient matrix \( C \in \mathbb{R}^{K \times g} \) such that the representation error \( \| Z - DC \|_F \) is minimized while satisfying the column-wise sparsity constraints on \( C \) that the number of nonzero entries of each column can not exceed a specified value \( S \), where the columns in \( Z \) are structural features of \( \{G_i\}_{i=1}^\ell \) and \( \| \cdot \|_F \) denotes the Frobenius norm. Many different methods exist for solving the dictionary learning problem of estimating \( D \) and \( C \), often called the sparse coding problem [23, 24]. In this paper, we focus on a spectral method (K-SVD) of dictionary learning introduced in [25]. The proposed MC-GDL selects the \( z \) highest SDS from each graph as columns of \( Z \) and applies K-SVD to find the dictionary and the corresponding coefficient matrix. The algorithm is summarized in Algorithm 2.

### 4. EXPERIMENTS AND CYBER INTRUSION DETECTION

#### 4.1. Illustration of sensitivity to structural changes on graphs

Here we consider four similar graphs with different structural characteristics as displayed in Fig. 1 (a). From top to bottom, these four graphs represent high structural symmetry, reduced structural symmetry due to edge removal, increase of the weight of edge (3,4), and change in edge direction. The extracted multi-centrality features are 1) graph walk statistics from 1 to 4 hops, and 2) the graph distance to node 1 (the reference node). It can be observed from Fig. 1 (b)
that MC-GPCA can reflect structural perturbations, and total data variance is explained by one or two principal components. Moreover, the first principal component is shown to complete describe the network flow pattern for the directed example graph. Fig. 1 (c) shows that the graph distance feature adds discrimination power as the MC-GPCA scores are better differentiated.

### 4.2. Cyber intrusion detection

The UNB intrusion detection evaluation dataset [33] described in Table 2 is a collection of directed cyber network graphs where each node is a host (machine) in a cyber system and an edge indicates the existence of communication between hosts. No information beyond graph topology is used for analysis. The extracted multi-centrality features are 1) graph walk statistics from 1 to 20 hops, 2) all centrality measures introduced in Sec. 2.2 (edge directions are omitted for computing LFVC), and 3) graph distances to 10 reference nodes of highest degree, resulting in $p = 56$ features (columns of $X$). Fig. 2 (a) shows that the proposed SDS statistic (Eqn. (9)) with $q = 2$ principal components from MC-GPCA. The SDS statistics are similar over days without attacks, whereas they are significantly higher in days under attacks that induce anomalous connectivity patterns (i.e. Days 3, 4 and 5). On the other hand degree statistic (Fig. 2 (b)) fails to be a valid indicator of cyber attacks. The SDS statistic fails to detect the SSH attack (Day 7) since it is a password attack that takes place only between a single host and a single server.

We applied MC-GDL to the entire UNB database of graphs to learn a dictionary that spans the dataset. For this implementation of MC-GDL we select $K = 2$ atoms, $z = 300$ SDS features and $S = 2$ sparsity level. The two learned structural atoms in Fig. 2 (c) can be interpreted as a normal activity atom consisting of identical SDS features except for one spike accounting for the main router and an attack activity atom of higher variance in SDS features. The corresponding coefficients in Fig. 2 (d) reflect the mixture portion of these atoms and they can be used for attack classification. For instance, $K$-means clustering with 2 clusters identifies Days 3, 4 and 5 as being anomalous and thus under attack.

### 5. CONCLUSION

This paper proposes PCA and dictionary learning graph decomposition methods that are based on multi-centrality features of the graph. The proposed methods can reflect structural perturbations in graph symmetry, edge weight and edge direction. When applied to cyber intrusion detection, our experiments show that MC-GPCA and MC-GDL can effectively detect attacks on the network.
6. REFERENCES


