TOWARD MATCHED FILTER OPTIMIZATION FOR SUBGRAPH DETECTION IN DYNAMIC NETWORKS

Benjamin A. Miller and Nadya T. Bliss

MIT Lincoln Laboratory
Lexington, Massachusetts 02420
{bamiller, nt}@ll.mit.edu

ABSTRACT

This paper outlines techniques for optimization of filter coefficients in a spectral framework for anomalous subgraph detection. Restricting the scope to the detection of a known signal in i.i.d. noise, the optimal coefficients for maximizing the signal’s power are shown to be found via a rank-1 tensor approximation of the subgraph’s dynamic topology. While this technique optimizes our power metric, a filter based on average degree is shown in simulation to work nearly as well in terms of power maximization and detection performance, and better separates the signal from the noise in the eigenspace.

Index Terms—community detection, dynamic graphs, graph algorithms, matched filtering, signal detection theory

1. INTRODUCTION

A graph $G = (V, E)$ is a pair of sets: a set of vertices, $V$, and a set of edges, $E$, which connect pairs of vertices. Graphs are used in a host of applications in which the data of interest include relationships (the edges) between entities (vertices), including physics, social network analysis, and cyber security. Since connections between entities will vary over time in many applications, a significant amount of recent research has focused on time-varying graphs (e.g., [1]).

While graphs are broadly useful, their non-Euclidean nature complicates the applications of traditional signal processing paradigms. Recent work has focused on developing a detection theory framework for data in the form of graphs, akin to that for Euclidean data [2]. An extension of signal detection theory to network data would be desirable for a variety of disciplines.

The original detection framework of [2] was extended to dynamic graphs via a temporal matched filtering technique in [3]. In this paper, we take steps toward optimizing the filter coefficients in a restricted setting: detection of a subgraph exhibiting known behavior in a temporally independent and identically distributed background graph. One key observation is that, for our chosen power metric, the optimal filter coefficients can be computed via a tensor decomposition. Such decompositions have been used with dynamic graphs [1, 4], but, to our knowledge, this technique has not been applied to subgraph detection before.

The remainder of this paper is organized as follows. In Section 2 we define the subgraph detection problem setting. Section 3 shows that statistics of the noise are restricted by placing a norm constraint on the filter coefficients, and in Section 4 we show that, given this constraint, we can optimally determine the filter coefficients using a rank-1 tensor approximation. Simulation results for signal power maximization and detection performance are provided in Section 5, and in Section 6 we summarize and discuss future research.

2. PROBLEM MODEL

The signal detection problem has many flavors, and in this paper we focus on the detection of a known signal in independent, identically distributed noise. Our observation is an unweighted, undirected, time-varying graph, with $G(n)$ denoting the graph at discrete time step $n$. This consists of a background $G_B(n)$, the noise in our observation, and may or may not include a signal subgraph $G_S(n)$. As in [2,3], we cast the problem as a binary hypothesis test, in which the null hypothesis $H_0$ is that the observation is only noise, $G(n) = G_B(n)$, and the alternative hypothesis $H_1$ is that there is also a signal present, i.e., $G(n) = G_B(n) \cup G_S(n)$. In our formulation, the vertex set $V$ remains constant throughout the time window of interest; only the edge set $E(n)$ changes.

The noise consists of i.i.d. Bernoulli graphs, meaning graphs where edges occur based on the outcome of independent Bernoulli trials. The probabilities are not identical across all pairs of vertices—as in Erdős–Rényi random graphs—but at each individual time step the probability of an edge between vertex $i$ and vertex $j$ is the same, denoted $p_{ij} = p_{ji}$. Thus, under $H_0$, the expected value of the adjacency matrix of the graph at any time instance, $E[A(n)]$, is given by
\( P = \{ p_{ij} \} \), a \(|V| \times |V|\) matrix of edge probabilities.

Under \( H_1 \), a dynamic subgraph that is unlikely to appear under \( H_0 \) is embedded into the background on a randomly selected subset of the vertices, \( V_S \subset V \). In this formulation, we know the subgraph’s temporal evolution pattern, but do not know its location in the background. While this could potentially be solved by a brute-force search, such an approach would be a form of the subgraph isomorphism problem, which is known to be NP-hard.

The temporal integration technique used is outlined in [3]. This method is based on spectral analysis of integrated graph residuals. Let \( \tilde{B}(n) = A(n) - E[A(n)] \), the difference between the adjacency matrix at time \( n \) and its expected value. The first step is to integrate these residuals over a defined time window, computing

\[
\tilde{B}(n) = \sum_{i=0}^{L-1} (A(n-i) - E[A(n-i)]) h_i,
\]

where \( L \) is the length of the time window and \( h \) is a sequence of \( i \) real numbers. This has the form of a classical finite impulse response (FIR) filter. We then consider \( \tilde{B}(n) \) in the space of its two principal eigenvectors. As in [2], we use a chi-squared statistic based on a \( 2 \times 2 \) contingency table for detection of the presence of the subgraph. This table contains the number of vertices (i.e., columns of \( \tilde{B}(n) \)) that are projected into each quadrant, and is maximized over rotation in the 2D plane. In the experiments of [2], under \( H_0 \) this projection was rather radially symmetric, and, thus, the detection statistic was smaller than when the symmetry was skewed by the embedding of an anomalous subgraph.

As a metric of signal and noise power, we use the spectral norm, i.e., the absolute value of the largest eigenvalue, denoted by \( \| \cdot \| \). To best detect the presence of the anomalous subgraph, our goal is to maximize signal power while restricting noise power, that is, to use coefficients

\[
h^* = \arg \max_h \left\| \sum_{i=0}^{L-1} A_S(n-i) h_i \right\|
\]

subject to \( \left\| \sum_{i=0}^{L-1} (A_B(n) - E[A_B(n)]) h_i \right\| = \eta \).

Here \( A_S(n) \) is the \(|V_S| \times |V_S|\) adjacency matrix of the dynamic foreground only, and \( A_B(n) \) is the adjacency matrix of the background alone. In the next two sections, we focus on coefficient optimization in this problem setting.

### 3. NOISE REDUCTION

To restrict the noise power after integration, we use the property that

\[
\| \tilde{B}(n) \| = \max_{\| u \|_2 = 1} \left| u^T \tilde{B}(n) u \right|.
\]

Rather than truly limit the maximum eigenvalue, we will restrict the variance of \( \tilde{B}(n) \) in any 1-dimensional subspace of \( \mathbb{R}^{|V|} \). The analysis in this section assumes knowledge of the probability matrix \( P \). We will analyze the moments of the quantity \( u^T \tilde{B}(n) u \), and assume an arbitrary, fixed \( u \) of unit magnitude. The first, simple observation is that, since \( \tilde{B}(n) \) is a random variable minus its expected value, \( E[u^T \tilde{B}(n) u] = 0 \), i.e., the distribution of \( u^T \tilde{B}(n) u \) is centered at the origin. The second-order moment of this quantity is given by

\[
E \left( u^T \tilde{B}(n) u \right)^2 = E \left( \sum_{i=0}^{L-1} u^T (A_B(n-i) - P) u h_i \right)^2
\]

\[
= E \sum_{i=0}^{L-1} u^T B(n-i) u h_i \sum_{j=0}^{L-1} u^T B(n-j) u h_j
\]

\[
= \sum_{i=0}^{L-1} h_i^2 E \left( u^T B(n-i) u \right)^2
\]

\[
= \sum_{i=0}^{L-1} \sum_{j=1}^{|V|} \sum_{k=1}^{|V|} \sum_{k=1}^{|V|} 2u_i^2 u_j^2 ((a_{jk} - p_{jk})^2) - \sum_{j=1}^{|V|} u_j^4 [(a_{jj} - p_{jj})^2]
\]

Regardless of the direction of \( u \), the variance of \( u^T \tilde{B}(n) u \) scales linearly with the sum of the squares of the filter coefficients. To restrict the expected noise power, therefore, we will fix the \( \ell_2 \) norm of the vector of filter coefficients to be 1.

### 4. SIGNAL MAXIMIZATION

In this section, we determine coefficients that solve the optimization problem as stated in equation (1), and consider two other formulations. As discussed in Section 3, we restrict the noise by setting \( \sum_{i=0}^{L-1} h_i^2 = 1 \), so the focus is on finding

\[
h^* = \arg \max_{h, \| h \|_2 = 1} \sum_{i=0}^{L-1} A_S(n-i) h_i.
\]

This can be rewritten as

\[
h^* = \arg \max_{h, \| h \|_2 = 1, \| u \| = 1} \sum_{i=0}^{L-1} h_i u^T A_S(n-i) u.
\]

Let \( A \) be a 3-way tensor in which \( A(i,j,k) \) contains the value from the \( j \)th row and \( k \)th column of \( A_S(n-i) \). For symmetric (undirected) subgraphs, (3) is equivalent to maximizing

\[
\sum_{i=0}^{L-1} \sum_{j=1}^{|V|} \sum_{k=1}^{|V|} \sum_{k=1}^{|V|} A(i,j,k) u_i u_j w_k,
\]

with the \( \ell_2 \) norms of \( h, u \) and \( w \) all constrained to be 1. This can be solved by finding the rank-1 approximation of \( A \), i.e.,
to compute \( h, u \) and \( w \), and a scalar \( \lambda \), such that

\[
A \approx \lambda (h \circ u \circ w),
\]

where \( \circ \) denotes the 3-way tensor outer product, with the \((i, j, k)\)th entry of \( h \circ u \circ w \) equal to \( h_i u_j w_k \). This is analogous to approximating a matrix \( M \) by the scaled outer product of its principal left and right singular vectors \( u \) and \( w \), which also maximize the quantity \( u^T M w = \sum_{ij} m_{ij} u_i w_j \). We can, thus, solve (2) by computing the rank-1 tensor approximation of \( A \) (via a CP decomposition using the Matlab Tensor Toolbox\(^1\)) and use the resulting vector \( h \) as the filter coefficients.

In [3], the filter coefficients used were proportional to the largest eigenvalues (in magnitude) of the associated adjacency matrices, i.e., the instantaneous signal power. While this provided adequate integration gain in the simulations, it is only the optimal solution for (2) when the principal eigenvector of \( A_S(n) \) is constant across \( n \). We provide results with such a filter as a point of comparison.

Finally, if the task requires not only detection of anomalous activity but also localization, i.e., determining which vertices are exhibiting the activity of interest, then maximizing the largest eigenvalue may not be optimal. In this case, it may be ideal to emphasize the cross section of the integrated residuals space that points equally in the direction of all subgraph vertices. To do this, we maximize the quantity

\[
\sum_{i=0}^{L-1} h_i \frac{1}{|V_S|} A_S(n-i) \frac{1}{|V_S|} - \sum_{i=0}^{L-1} h_i \text{Vol}(G_S(n-i)),
\]

where \( 1_N \) is a column vector of \( N \) ones and \( \text{Vol}(\cdot) \) is the volume of the graph (the sum of the vertex degrees). Thus, a filter based on the subgraph’s average degree will most emphasize the portion of the residuals space aligned with the subgraph.

### 5. RESULTS

We ran several Monte Carlo simulations to demonstrate the filter optimization techniques. In each experiment, we consider a time window of 32 samples. At each sample, the background consists of an independent R-MAT Kronecker graph [5] with 1024 vertices, a typical average degree of about 10, and a base probability matrix

\[
P_{\text{RMAT}} = \begin{bmatrix}
0.5 & 0.125 \\
0.125 & 0.25 
\end{bmatrix}.
\]

We run the R-MAT algorithm for a fixed number of iterations, so it is a Bernoulli graph as discussed in Section 3.

The foreground in these simulations consists of a 20-vertex subgraph, divided into 2 portions. The behavior of the subgraph involves one subset of the vertices densifying over the first half of the window, with edges then shifting to the other portion over the second half, e.g., a community forming, then bringing in new members as others leave. Two subsets \( V_1, V_2 \subset V_S \) both have 12 vertices, with 4 of them overlapping. The subgraph starts with no edges and, over the first half of the time window, adds edges within \( V_1 \) until it reaches a density \( d \). In the second half of the window, edges are removed from \( V_1 \) and added to \( V_2 \), while maintaining the total number of edges, until edges only exist within \( V_2 \).

We first demonstrate the difference in the spectral norm of the integrated signal subgraph using the three filters specified in Section 4. The results of a 10,000-trial Monte Carlo simulation are summarized in Fig. 1. In each trial, a different foreground is randomly generated. As expected, the tensor decomposition always provides the greatest spectral norm, and this figure shows the relative magnitudes using the other techniques. The filter based on average degree performs quite well, always achieving a norm of at least 96.5% of the optimal value. The filter based on maximum eigenvalues does not always underperform the average degree filter, obtaining a larger norm in just over 4% of the trials, but does frequently take on significantly smaller values.

When embedding the subgraph into the R-MAT background, we fixed the foreground across all experiments for each density to maintain consistency of the filter coefficients. As an estimate for \( E[A] \), we simply use the average of the adjacency matrices over the time window. (This does preclude using a uniform filter, as it would result in \( B \) being a zero matrix.) For each filter, we computed chi-squared statistics (as described in Section 2) for 10,000 graphs under \( H_0 \) and \( H_1 \). Receiver operating characteristic curves are presented in Fig. 2, with detection performance in all cases increasing as the maximum subgraph density increases from 20% to 35%. The plots confirm that the filter using average degree yields substantially similar results to the optimal tensor decomposition, while both outperforming the filter based on maximum eigenvalues. Indeed, the equal error rates for when using filters derived from the average degree and the tensor decomposition differ by at most 2%.

While the tensor decomposition maximizes signal power, other metrics are worth considering as well. In Fig. 3, we

---

*Available online at http://www.sandia.gov/~tgolds/TensorToolbox/*.
**Fig. 2.** ROC curves for detection of the deterministic foreground in i.i.d. R-MAT backgrounds. Performance is shown for filters derived using maximum eigenvalue (left), average degree (center) and the tensor decomposition (right).

Fig. 3. Scatterplots of the principal 2-dimensional subspace of $\tilde{B}$. While the tensor decomposition (top) obtains the greatest signal power, the average degree filter (bottom) provides better foreground/background separation in this space.

**6. SUMMARY**

This paper provides techniques for optimization of filter coefficients in a spectral framework for subgraph detection. The optimal coefficients are shown to be found via a rank-1 tensor approximation of the subgraph, but a filter based on average degree is shown in simulation to work nearly as well, and does a better job separating the signal from the noise in the eigenspace. Since these results depend on the assumption that the noise is i.i.d., determining a noise whitening process for models such as [6] will be useful future work. Expanding this analysis to non-Bernoulli graphs and random signals with known distributions will also be of significant interest.

**7. REFERENCES**


