

Nearly-Linear Time Optimization for Connected Subgraph Detection

Anonymous Authors¹

Abstract

We propose a novel, computationally efficient optimization framework for subgraph detection in graph-structured data, where we aim to discover anomalous patterns present in a connected subgraph of a given graph. This problem arises in many applications such as detection of network intrusions, community detection, detection of anomalous events in surveillance videos or disease outbreaks. Since optimization over connected subgraphs is a combinatorial and computationally difficult problem, we propose a convex relaxation that offers a principled approach to incorporating connectivity and conductance constraints on candidate subgraphs. We develop a novel nearly-linear time algorithm to solve the relaxed problem, establish convergence and consistency guarantees and demonstrate its feasibility and performance with experiments on real networks.

1. Introduction

We consider the problem of connected subgraph detection, motivated by statistical anomaly detection on networks where the aim is to determine whether there exists a set of connected nodes that exhibit anomalous signal values. One example of the anomaly detection problem on networks is disease outbreak detection (Patil et al., 2003), where the nodes are associated with counties and geographical neighborhood characterizes connectivity, and signal values on nodes depict the number of patients related to a disease. In the existence of a disease outbreak that spreads geographically, higher signal values would be present on certain counties that are neighbors of each other, therefore constituting a subgraph structure. Similar problems in different research areas also exist, such as detection

of intrusions in communication or sensor networks, community detection or video surveillance.

The detection or estimation of arbitrary connected subgraphs over graph-structured signals is an example of a structured signal recovery problem and generalizes many useful types of structures such as intervals or paths (Addario-Berry et al., 2010; Arias-Castro et al., 2008). While the existence of structure in terms of connectivity leads to better statistical complexity in detecting or recovering the anomalous sets compared to arbitrary subsets, efficient characterization of sets obeying the connectivity constraint is important for obtaining practical algorithms for detection and estimation. In this paper we aim to characterize the space of arbitrary connected subsets of nodes in a graph in a principled manner through spectral graph theory and propose efficient optimization algorithms that exploit this characterization.

1.1. Related work and contributions

Subgraph detection is a difficult problem since connected subgraphs represent a combinatorial structure and systematic approaches to characterizing the space of connected subgraphs of a given graph are relatively recent. Traditional approaches to this problem usually consider parametric methods, which originate from the scan statistics literature (Glaz et al., 2001) and consider scanning for specific shapes such as rectangles, circles or neighborhood balls on graphs (Kulldorff et al., 2006; Patil et al., 2003; Priebe et al., 2005). More recently nonparametric approaches have been considered for subgraphs with arbitrary shapes on general graphs. The simulated annealing approach of (Duczmal & Assuncao, 2004) is an example of nonparametric methods that can detect complex shapes, however it is a heuristic method without statistical or computational guarantees. There is also a line of work focused on statistical analysis with nonparametric shapes (Addario-Berry et al., 2010; Arias-Castro et al., 2008; 2012), however they consider computationally intractable algorithms.

More recently, (Qian & Saligrama, 2014; Qian et al., 2014) proposed a semidefinite programming (SDP) formulation of the problem, along with a novel linear matrix inequality (LMI) constraint that provably characterizes the connectiv-

¹Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

ity of subsets of nodes exactly in the integer problem. This approach is the most similar to ours where we also consider an SDP relaxation with LMI constraints, however we take a different approach to formulating the problem and its relaxation, with the goal to obtain a convex optimization program that is amenable to more efficient iterative methods. Another notable work in this area is the spectral scan statistic approach proposed by (Sharpnack et al., 2016), which presents a computationally tractable algorithm with consistency guarantees. However it is important to note that this method aims to obtain graph partitions with small conductance and balanced sizes, in contrast to our formulation that guarantees connected subgraphs. In recent work, (Wu et al., 2016) consider nonparametric statistics for signals in addition to nonparametric shapes, obtaining a computationally tractable algorithm by heuristically approximating the underlying graph with trees.

The contribution of this paper is twofold: First, we develop a convex relaxation of the subgraph detection problem that results in an semidefinite optimization formulation, with provable guarantees on the connectivity of the resulting solutions related to the internal conductance of the subgraph. Second, we propose an efficient iterative framework for optimizing the SDP that scales well with large problem sizes, and show computational guarantees. One of the major differences of our formulation to those of (Qian & Saligrama, 2014; Qian et al., 2014) is that prior work enforce a number of constraints that scale with the problem size, whereas ours only considers a constant number of constraints. Also, while aforementioned work utilized generalized convex optimization solvers, our formulation allows us to propose specialized and efficient iterative algorithms.

2. Connected Subgraph Detection

In this section we define the notation and introduce the two statistical models that we consider for the connected subgraph detection problem. Let $G = (V, E)$ denote an undirected unweighted connected graph with n nodes that is provided as input to the problem. For $i \in V$, we write d_i for the degree of vertex i in G and let d be an upper bound on all d_i . For a subset $S \subseteq V$, the notation $\text{Vol}(S)$ indicates the volume measure of S , i.e., $\text{Vol}(S) = \sum_{i \in S} d_i$. We also denote by $G_S = (S, E_S)$ the subgraph induced by the subset S . For an input root vertex $r \in V$, we write $\Lambda_r = \{S \subseteq V : r \in S, G_S \text{ is connected in } G\}$. Indicator vectors with notation 1_S are defined as $n \times 1$ vectors with i -th index 1 if $i \in S$ and zero otherwise.

We consider observations $x_v \in \mathbb{R}^p$ associated with each node $v \in V$ in the graph G . We are concerned with optimization problems of the form

$$\max_{S \in \Lambda_r} c(S), \quad (1)$$

for a cost function $c(\cdot)$ which depends on $x_S = \{x_v\}_{v \in S}$. We remark that this is a difficult problem due to the combinatorial nature of the constraint, in fact variants of the prize-collecting Steiner tree problem which is known to be NP-hard (Johnson et al., 2000) can be reduced to the above formulation. Below we provide two examples of the setup that we consider, namely elevated mean detection and correlation detection.

Elevated mean detection: In this problem, the aim is to detect the existence of a subgraph $S \in \Lambda_r$ comprising of nodes with an elevated mean compared to the other nodes. A simple example is the Gaussian elevated mean model, where $x_v = \mu 1\{v \in S\} + z_v$ for $\mu > 0$ and $z_v \sim \mathcal{N}(0, \sigma^2)$. Another example we consider is the Poisson variant, where $x_v \sim \text{Poisson}((1 + \mu 1\{v \in S\}) \lambda)$.

We consider the optimization (1) with the scan statistic

$$c_1(S) = \frac{1}{\sqrt{|S|}} \sum_{i \in S} x_i, \quad (2)$$

which can be shown to correspond to the generalized likelihood ratio test (GLRT) for the Gaussian detection problem, while also encouraging sparse solutions for estimation of S .

Correlation detection: Another example is the problem of detecting and estimating a subgraph with correlated signal values. The canonical statistical model that has been investigated for this problem is where the signals are jointly Gaussian random variables X_1, \dots, X_n , where $\text{cov}(X_i, X_j) = 1$ if $i = j$, ρ if $i, j \in S$ and zero otherwise (Arias-Castro et al., 2012). Note that while related work consider arbitrary k -sets or special shapes such as intervals for S , we allow arbitrary connected subgraphs containing the root r .

One simple test for detecting or estimating a correlated subgraph induced by set S is (1) with the scan statistic

$$c_2(S) = \frac{1}{|S|} \sum_{i,j \in S} \hat{\Sigma}_{ij}, \quad (3)$$

where $\hat{\Sigma}$ is the estimated covariance matrix which can either be defined by a single observation xx^\top when $p = 1$ or multiple observations for $p > 1$.

Characterizing subgraph connectivity: Rather than focusing on exactly characterizing the connectedness of an induced subgraph G_S , we aim to enforce it by lower bounding the *conductance* of cuts within G_S . For a weighted graph $G = (V, E, w)$, the conductance of a cut S is defined as

$$\phi_G(S) = \frac{w(S, V \setminus S)}{\text{Vol}(S)},$$

where $w(S, V \setminus S)$ is the total weight of edges connecting nodes in S to nodes in $V \setminus S$. The graph conductance is the lowest conductance among cuts containing at most half of the volume of the graph, i.e., $\phi_G = \min_{S \subset V: \text{Vol}(S) \leq \text{Vol}(V)/2} \phi_G(S)$. Conductance is a natural graph-partitioning objective because of its intimate connection with the behavior of random walks. It is also widely used in practice in the design of clustering and segmentation algorithms. We can use conductance to ensure subgraph connectivity by imposing the following constraints¹ on the integral solution S :

$$\phi_{G_S}(T) = \frac{w(T, S \setminus T)}{\text{Vol}(T)} \geq \gamma > 0, \quad \forall T \subseteq S \setminus \{r\}. \quad (4)$$

For γ sufficiently small, i.e., $\gamma < 1/\text{Vol}(V)$, this requirement is equivalent to the connectivity condition on G_S . It is useful to notice at this stage that the condition in constraint (4) is stronger than a lower bound on the conductance of the induced subgraph G_S . Indeed, for an unweighted graph the constraint (4) implies

$$\phi_{G_S} = \min_{U \subseteq S: \text{Vol}(U) \leq \frac{\text{Vol}(S)}{2}} \frac{|E(U, S \setminus U)|}{\text{Vol}(U)} \geq \gamma,$$

where $E(U, S \setminus U)$ is the set of edges between sets of nodes U and $S \setminus U$. However, our constraint is stronger than requiring the induced conductance of G_S to be γ , as the bound also holds for subsets $T \subset S$ comprising more than half the volume of G_S .

Comparison with other measures of connectivity To form a better understanding of how conductance arises in this context, consider a (dual) flow formulation of connectedness: $S \in \Lambda_r$ if, for any $i \in S$, there exists a way to route one unit of flow from the root r to i with finite congestion. This can be strengthened by requiring that 1 unit of flow be routed from r to i with congestion $\frac{1}{k}$, yielding the notion of k -edge connectivity. Finally, we can strengthen the flow requirement further, by demanding that (1) d_i units of flow be routed from r to i , and (2) the flows from r to all vertices $i \in S$ be routed *concurrently* with congestion $\frac{1}{\gamma}$. The maxflow-mincut theorem shows that such a flow routing exists if and only if the condition of (4) holds.

While the notions of connectivity and conductance converge to connectedness in the limit as γ goes to 0, they display different behaviors for larger γ . In particular, conductance appears to be a more meaningful in the context of anomaly detection scenarios, where the anomalous set may be constructed by an unspecified diffusion process, such as the epidemic in the example of disease outbreak detection,

¹Note that for technical reasons, the conductance ϕ_{G_S} on the induced graph G_S still employs the definition of volume given by the larger graph G .

which is unlikely to cross low-conductance cuts. Moreover, for noisy input graphs, conductance is a more robust notion than edge connectivity.

3. Relaxation

We next consider a convex relaxation of the objectives of (2) and (3) as a linear functional of positive semidefinite matrix variable M . We remark that in the case that $x \geq 0$ (e.g. the Poisson model), maximizing the scan statistic $c_1(S)$ is equivalent to maximizing its square

$$c_1^2(S) = \frac{1}{|S|} \left(\sum_{i \in S} x_i \right)^2 = \frac{1}{|S|} \sum_{i, j \in S} x_i x_j, \quad (5)$$

which has the same form as the statistic $c_2(S)$.

For this statistic, defining the indicator vector $u = 1_S$ and noting that $u_i = u_i^2$ we can write the quadratic integer program (IP)

$$\max_{u \in \{0,1\}^n, \{i: u_i=1\} \in \Lambda_r} \frac{\sum_{i,j} x_i x_j u_i u_j}{\sum_i u_i^2}. \quad (6)$$

We relax this IP to a semidefinite program (SDP) by turning each element u_i to a vector $v_i \in \mathbb{R}^n$ such that scalar multiplication is transformed to inner product and we have $\langle v_i, v_j \rangle = 1$ if $i, j \in S$ and zero otherwise. Moreover, we also enforce non-negativity by requiring that $\langle v_i, v_j \rangle \geq 0$ for all $i, j \in V$. We then have

$$\max_{v_i \in \mathbb{R}^n, \{i: |v_i|>0\} \in \Lambda_r, \langle v_i, v_j \rangle \geq 0, \forall i, j \in V} \frac{\sum_{i,j} x_i x_j \langle v_i, v_j \rangle}{\sum_i |v_i|^2}. \quad (7)$$

Using the the Gram matrix $M = V^\top V \succeq 0$ instead of the vectors v_i 's and fixing the trace of M , $I \cdot M$ to 1 w.l.o.g. (due to the homogeneity of the ratio in the objective), we obtain the relaxation

$$\max_{M \in \Delta_n, M \succeq 0} C \cdot M \quad \text{s.t.} \quad \{i : |M_{ii}| > 0\} \in \Lambda_r, \quad (8)$$

where $C = xx^\top$ and we define Δ_n to be the spectrahedron of unit trace PSD matrices. The relaxation follows along exactly for $c_2(S)$ with $C = \hat{\Sigma}$. This linear functional formulation is very general and can be adapted to solve subgraph problems with different cost functions, in addition to (2) and (3).

Next, we propose a novel SDP formulation of the connected subgraph constraint $\{i : |M_{ii}| > 0\} \in \Lambda_r$, as a single linear matrix inequality that arises naturally from a spectral relaxation of the integral conductance constraint of (4).

3.1. Spectral Graph Theory

We start by introducing some basic notation and concepts from spectral graph theory. For a weighted graph $H = (V_H, E_H, h)$, we denote its adjacency by A_H and its degree matrix D_H . The Laplacian of H is then defined as $L_H = D_H - A_H$. The $n \times n$ Laplacian matrix for the graph on V consisting only of edge $\{i, j\}$ is $L_{ij} = e_{ii} + e_{jj} - e_{ij} - e_{ji}$, where e_{ij} is an all-zero matrix except for a one at index (i, j) . Notice that $L_H = \sum_{(i,j) \in E_H} h_{ij} L_{ij}$. We omit the subscripts for all graph matrices and sets when referring to the instance graph G . For a subset $S \subset V$, we denote by K_S the complete graph on S , i.e., the graph having an edge of weight $d_i d_j$ between i and j for any $i, j \in S$. The spectral gap λ_S of an induced subgraph G_S of the input graph G is defined as the minimum generalized eigenvalue of L_{G_S} with respect to $\frac{1}{\text{Vol}(S)} L_{K_S}$. Equivalently, the spectral gap λ_S is the largest real λ such that

$$L_{G_S} \succeq \frac{\lambda}{\text{Vol}(S)} L_{K_S}.$$

The star graph $\text{Star}^{(r)}$, rooted at a vertex $r \in V$, is the graph consisting of the $n - 1$ edges of the form (r, i) for all $i \in V$, each with weight d_i . We associate to a solution $M \in \Delta_n$ of (8) two weighted graphs, $G[M]$ and $\text{Star}^{(r)}[M]$, defined by their Laplacians:

$$L_{G[M]} = \sum_{(i,j) \in E} M_{ij} L_{ij} \quad \text{and} \\ L_{\text{Star}^{(r)}[M]} = \sum_{i \in V} d_i M_{ii} L_{ri}.$$

Cheeger's inequality: An important result in spectral graph theory is Cheeger's inequality (Chung, 1997) that relates the conductance of a graph with the spectrum of its Laplacian. An equivalent statement for the subgraph conductance that follows from Cheeger's inequality and relates the spectral gap of G_S to the conductance of G_S can be written as follows.

Theorem 3.1 (Cheeger's Inequality). *For $S \subseteq V$,*

$$\frac{\lambda_S}{2} \leq \phi_{G_S} \leq \sqrt{2\lambda_S}.$$

The right-hand side of this inequality is proved by rounding the generalized eigenvector of L_{G_S} associated with λ_S to a low-conductance cut by using the following lemma.

Lemma 3.1. *Let $y \geq 0$ and $y_r = 0$. Assume that:*

$$y^T L_{G_S} y < \lambda \sum_{i \in S} d_i y_i^2.$$

Then, there exists $\tau > 0$ such that the sweep cut $L_\tau = \{i \in S : y_i \geq \tau\}$ of vector y has $\phi_{G_S}(L_\tau) < \sqrt{2\lambda}$.

3.2. Relaxing the conductance requirement

Our proposed relaxation of the integral conductance constraint (9) with parameter γ is the following:

$$L_{G[M]} \succeq \frac{\gamma^2}{2} L_{\text{Star}^{(r)}[M]} \quad (9)$$

To see how this relaxes the integral constraint, take $M_S = \frac{1}{|S|} 1_S 1_S^T$ to be an integral solution corresponding to a subset $S \subseteq V$. We have:

$$L_{G[M_S]} = \frac{1}{|S|} L_{G_S} \quad \text{and} \quad L_{\text{Star}^{(r)}[M_S]} = \frac{1}{|S|} L_{\text{Star}_S^{(r)}}.$$

Then, our proposed constraint becomes

$$L_{G_S} \succeq \frac{\gamma^2}{2} L_{\text{Star}_S^{(r)}}.$$

We now show that this constitutes a relaxation of constraint (4). This can be seen as a variant of Cheeger's inequality for our relaxed notion of conductance in (9).

Theorem 3.2. *For $S \subseteq V$, if, for all $T \subseteq S, r \notin T$ and $\phi_{G_S}(T) \geq \gamma$, then $L_{G_S} \succeq \frac{\gamma^2}{2} L_{\text{Star}_S^{(r)}}$.*

Proof. We prove the converse. Given a vector x such that

$$x^T L_{G_S} x < \frac{\gamma^2}{2} x^T L_{\text{Star}_S^{(r)}} x,$$

we produce a cut $T \subseteq S, r \notin T$, with $\phi_{G_S}(T) < \gamma$. To do so, consider the embedding $y \in \mathbb{R}^n$ given by $y_i = |x_i - x_r|$. Then, it is easy to check that:

$$y^T L_{G_S} y \leq x^T L_{G_S} x < \frac{\gamma^2}{2} x^T L_{\text{Star}_S^{(r)}} x = \frac{\gamma^2}{2} \sum_{i \in S \setminus \{r\}} d_i y_i^2.$$

By Lemma 3.1, a sweep cut of y yields a cut T such that $\phi_{G_S}(T) < \gamma$. Moreover, because $y_r = 0$, r does not belong to T . \square

Moreover, for a candidate integral solution M_S , in the same way as the integral constraint lower bounds the conductance of the induced subgraph G_S , our relaxation can be shown to lower bound the spectral gap of G_S . To see this, we can use the following lemma, which is the result of a simple algebraic manipulation based on Schur's complementation.

Lemma 3.2. *For $S \subset V$ with $r \in S$, let $y = (\text{Vol}(S) - d_r) e_r - \sum_{j \in S, j \neq r} d_j e_j$. Then,*

$$L_{\text{Star}_S^{(r)}} = \frac{1}{\text{Vol}(S)} [L_{K_S} + yy^T].$$

Applying this lemma to (9), we observe that our constraint, applied to an integral solution M_S , implies a lower bound of $\frac{\gamma^2}{2}$ on the induced spectral gap λ_S through the inequality

$$L_{G_S} \succeq \frac{\gamma^2}{2} L_{\text{Star}_S^{(r)}} \succeq \frac{\gamma^2}{2 \text{Vol}(S)} L_{K_S}.$$

Finally, we prove that if any feasible candidate M that satisfies (9) is rounded to a subset S in a certain way, then the connectivity of subgraph G_S is ensured. This result shows that the inequality constraint (9) is sufficient to ensure connectivity in our framework.

Theorem 3.3. *For any $\gamma > 0$ and $M \succeq 0$ that satisfies (9), the subgraph $G_{\hat{S}}$ for induced subset $\hat{S} = \{i \in V : M_{ii} > 0\}$ is connected.*

We provide the proof of Theorem 3.3 in the appendix, which follows from an alternative formulation of constraint (9) through effective resistance on electrical networks.

3.3. Primal and Dual Formulations

To conclude this section, we restate our relaxation and introduce some shorthand notation for its constraints. Let $Q_\gamma(M) \succeq 0$ be our relaxed connectedness constraint, i.e., $Q_\gamma(M) = L_{G[M]} - \frac{\gamma^2}{2} L_{\text{Star}^{(r)}[M]}$. Then our relaxation is given by

$$\max_{\substack{M \in \Delta_n, M \succeq 0, \\ Q_\gamma(M) \succeq 0}} C \cdot M. \quad (10)$$

We now write the SDP dual of our relaxation. We consider a scalar α as the Lagrange multiplier corresponding to constraint $I \cdot M = 1$, and matrices $Y, Z \in \mathbb{R}^{n \times n}$ corresponding to $Q_\gamma(M) \succeq 0$ and $M \succeq 0$ respectively. Let $P_\gamma(Y) = \sum_{(i,j) \in E} (L_{ij} \cdot Y) e_{ij} - \gamma \sum_{i \in V} d_i (L_{ri} \cdot Y) e_{ii}$ be the transpose of the constraint Q_γ , i.e., $P_\gamma(Y) \cdot M = Q_\gamma(M) \cdot Y$. The dual can then be written as:

$$\begin{aligned} \min \quad & \alpha \\ & C + P_\gamma(Y) + Z \preceq \alpha I \\ & \alpha \geq 0, Y \succeq 0, Z \succeq 0 \end{aligned}$$

At this point, we notice an important feature of our formulation. We have $C \succeq 0$, by definition for the elevated mean problem with nonnegative signal values and with high probability for correlation detection. Then the term $C + P_\gamma(Y)$ in the dual constraint is the sum of a nonnegative matrix plus a diagonal matrix. By the Perron-Frobenius Theorem, the top eigenvector has nonnegative components and we can assume w.l.o.g. that $Z = 0$. We will use the same reasoning in the next section to show that we do not need to explicitly enforce the n^2 element-wise nonnegativity constraints corresponding to $M \succeq 0$, as our dual formulation will automatically yield such solutions.

4. Efficient Optimization via Mirror Descent

We first consider a modification of our original SDP by adding the violation variables $s \geq 0$. For some fixed penalty value $p \geq 0$ we write

$$\max_{M \in \Delta_n, s} C \cdot M - ps \quad \text{s.t.} \quad Q_\gamma(M) + sD \succeq 0. \quad (11)$$

Recalling that D is the degree matrix of G , the term sD provides a measure of how violated the SDP constraint is. If $s \geq \gamma^2$, it is possible to prove that the SDP constraint is trivially satisfied for any M in Δ_n at a cost of ps in the objective (which follows from the fact that $L_{\text{Star}^{(r)}} \preceq 2D$). To avoid such trivial solutions, we set $p \geq \frac{4\text{OPT}}{\gamma^2}$, where OPT is the optimal value for cost function $C \cdot M$. This means that in a solution with nonnegative cost, s can be at most $\frac{\gamma}{4}$. In practice we can replace OPT with $\|C\|$, which is an upper bound to the optimal value.

Introducing the Lagrange multiplier $Y \succeq 0$ corresponding to the constraint $Q_\gamma(M) + sD \succeq 0$, we then obtain the saddle point problem

$$\max_{M \in \Delta_n, s} \min_{Y \succeq 0} C \cdot M - ps + Y \cdot (Q_\gamma(M) + sD),$$

from which we obtain the dual

$$\min_{Y \in \Delta_n^{p,D}} f(Y), \quad \text{where} \quad f(Y) = \max_{M \in \Delta_n, M \succeq 0} (C + P_\gamma(Y)) \cdot M,$$

and we defined $\Delta_n^{p,D}$ to be the scaled spectrahedron $\{X \succeq 0 : D \cdot X = p\}$. For this dual optimization over Y we utilize the mirror descent method, which is the optimal optimization algorithm for non-smooth functions in the blackbox model (Ben-Tal & Nemirovski, 2015). Mirror descent can also be interpreted as a generalization of sub-gradient methods to non-Euclidean spaces. To overcome the non-smoothness of the objective, mirror descent makes use of a strongly convex differentiable mirror map function $\omega(\cdot)$. We refer the reader to Section 5.3 of (Ben-Tal & Nemirovski, 2015) for more details. As it is standard in the spectrahedron setup $\max_{X \in \Delta_n} f(X)$, we will use the negative von Neumann entropy as our mirror map. Finally, to apply mirror descent, we need access to the gradient of f at $Y^{(t)}$. By Danskin's Theorem, this is given by:

$$\nabla_Y f(Y^{(t)}) = Q_\gamma(M^{(t)}),$$

$$M^{(t)} = \arg \min_{M \in \Delta_n, M \succeq 0} (C + P_\gamma(Y^{(t)})) \cdot M.$$

Hence, computation of the gradient requires finding $M^{(t)}$, which plays the role of the primal update at time t . However, this is just the rank-1 matrix given by the projection over the top eigenvector of $C + P_\gamma(Y^{(t)})$, where $M \succeq 0$ is once again ensured by Perron-Frobenius. Using the definition of the mirror map, for a step size η , we obtain the

Algorithm 1 Mirror descent for connected subgraph detection

Input: $C, p, r, \gamma, \eta, \epsilon$
Output: \hat{M}, \hat{S}
 $Y^{(0)} \leftarrow \frac{p}{\text{Tr}(D)} I_n$
 $G^{(0)} \leftarrow 0$
for $t = 1, \dots, T$ **do**
 $v \leftarrow \text{eig}(C + P_\gamma(Y^{(t-1)}))$
 $M^{(t)} \leftarrow vv^\top$
 $G^{(t)} \leftarrow G^{(t-1)} + Q_\gamma(M^{(t)})$
 $Y^{(t)} \leftarrow \exp(-\eta G^{(t)})$
 $Y^{(t)} \leftarrow \frac{p}{D \cdot Y^{(t)}} Y^{(t)}$
end for
 $\hat{M} \leftarrow \frac{1}{T} \sum_{t=1}^T M^{(t)}$
 $\hat{S} \leftarrow \{i : \hat{M}_{ii} > \epsilon\}$

well-known multiplicative update rule for Y , (see (Kivinen & Warmuth, 1997))

$$Y^{(t+1)} \propto \exp\left(\log Y^{(t)} - \eta Q_\gamma(M^{(t)})\right),$$

where $Y^{(t+1)}$ is normalized so that $D \cdot Y^{(t+1)} = p$ and $Y^{(0)} = \frac{p}{\text{Tr}(D)} I_n$. We can also avoid computing the matrix logarithm and unwind the recursion to get

$$Y^{(t+1)} \propto \exp\left(-\eta \sum_{\tau=1}^t Q_\gamma(M^{(\tau)})\right),$$

to directly compute $Y^{(t+1)}$ from a running sum of $Q_\gamma(M^{(\tau)})$ matrices. We formally present the resulting algorithm in Algorithm 1, where $\text{eig}(\cdot)$ operator returns the eigenvector of the operand corresponding to the largest eigenvalue.

Using standard results on the convergence of mirror descent (Ben-Tal & Nemirovski, 2015), which are applicable thanks to our careful formulation of the problem, we obtain the following convergence bound.

Theorem 4.1. *Setting $\eta = \frac{1}{4\epsilon OPT}$, Algorithm 1 converges to an ϵ -multiplicative approximation of optimal in $T = O\left(\frac{\log n}{\gamma^2 \epsilon^2}\right)$ steps.*

Moreover, each iteration consists of computing the top eigenvector of a non-negative matrix and the matrix exponential of a the sum of a Laplacian and a rank-1 term (cf. Lemma 3.2). Thanks to recent theoretical results, both of these objects can be approximated sufficiently closely in nearly-linear-time (Cohen et al., 2016; Orecchia et al., 2012; Vladu, 2016-02-03) quickly by exploiting well-known dimensionality reduction techniques and numerical algebra results. In practice, existing iterative solvers, combined with the use of the Johnson-Lindenstrauss Lemma to

keep a low-dimensional sketch of the matrix exponential, already provide a very efficient computational approach to this problem.

Future work: In this work, we did not perform a theoretically study of the approximation guarantees achievable in rounding our relaxation to an integral solution in the worst-case, as this is a much more challenging mathematical task. For this paper, we empirically tested a number of different rounding techniques, including random projections and truncating the diagonal entries.

5. Detection and Statistical Guarantees

In this section we consider an analysis of the statistical detection power of our method on certain types of graphs. We specifically consider the elevated mean detection problem with Poisson variables, such that for some $\mu > 1$ the null and the composite alternative hypotheses are given by

$$H_0 : x_i \sim \text{Poisson}(\lambda), \quad i \in V,$$

$$H_1 : x_i \sim \text{Poisson}((1 + (\mu - 1) 1\{i \in S\})\lambda), \quad \text{for some } S \subseteq V, i \in V.$$

We investigate the sufficient conditions on the SNR parameter μ for asymptotic separability of the two hypotheses. Namely, we show that the scores $xx^\top \cdot M$ for M returned by the algorithm are separable in the cases of H_0 and H_1 with high probability, as the problem size $n \rightarrow \infty$. This implies that there exists a threshold α_n such that thresholding the optimization score $xx^\top \cdot M \leq \alpha_n$ is a reliable test for detection of anomalous subgraphs with elevated mean.

For the Poisson setup above we write the following theorem.

Theorem 5.1. *For a line graph with n nodes, the hypotheses H_0 and H_1 are asymptotically separable if $\mu = \Omega(\log K \sqrt{\log n})$ for $|S| = K = \Omega(\log n)$.*

The theorem is proved in the supplementary material. To prove separability, we first lower bound the expected primal optimization value for H_1 with feasible solution $M = \frac{1}{|S|} 1_S 1_S^\top$. We then upper bound the optimal value under H_0 by considering the dual of the optimization problem and computing the expected dual value for a feasible dual solution. The theorem then follows by finding the value of μ such that the two bounds are asymptotically separable and applying standard concentration inequalities.

We remark that the detection bound above asymptotically matches the bound for the Poisson model given in Theorem 5 of (Qian & Saligrama, 2014) and is within a logarithmic factor in K of the lower bound provided in (Qian & Saligrama, 2014). However our result does not yet generalize to finite and asymptotic 2-D lattice graphs, for which we plan to extend our analysis in future work.

6. Experiments

In this section we present experiments on a disease outbreak dataset on a real world geographical network. We compare the statistical detection performance of our mirror descent (MD) algorithm with subgraph detection methods from related work.

The geographical map and corresponding network that we consider are illustrated in Figures 1b and 1a respectively, with 129 nodes representing counties in the northeastern United States and average degree 4.7. The ground truth cluster of 16 nodes for the anomalous case and the chosen anchor node are also illustrated. Note that this dataset has also been considered in (Qian et al., 2014) and (Qian & Saligrama, 2014), although the network connectivity is slightly different and (Qian et al., 2014) considers a different ground truth cluster and does not consider detection performance.

Following (Patil et al., 2003) and (Qian & Saligrama, 2014; Qian et al., 2014), we consider an elevated mean Poisson formulation for modeling the diseased population, where the number of disease cases y_i for a county i is given by $y_i \sim \text{Poisson}(N_i \lambda_0)$ where N_i is the population of the county, whereas for anomalous counties we have $y_i \sim \text{Poisson}(N_i \lambda_1)$. We consider $\lambda_0 = 5 \times 10^{-5}$ for the base disease rate and different $\frac{\lambda_1}{\lambda_0}$ ratios $\{1.1, 1.3, 1.5\}$ corresponding to different SNR values. As our test statistic we consider the disease rate per person $x_i = \frac{y_i}{N_i}$. One sample realization for the anomaly case with high SNR $\frac{\lambda_1}{\lambda_0} = 4$ is illustrated in Figure 1c.

To compare the performance with MD as proposed in Algorithm 1, we consider several other methods in the related literature, including the LMI-test (LMIT) method of (Qian & Saligrama, 2014), simulated annealing (SA) of (Duczmal & Assuncao, 2004) and the nearest-ball test (NB), which is a parametric method that scans over nearest-neighbor balls for different nodes. For the MD method we consider the optimization value $xx^\top \cdot M$ as the scan statistic, with $T = 100$ iterations, $\eta = 5$ and different γ values to quantify the size and conductance of the anomalous graph

For LMIT we use the same anchor node as MD, anomaly size $|S| = 16$ corresponding to the ground truth and consider scan statistic $x^\top \text{diag}(M)$. We search over a range of values for parameter γ . For SA and NB we consider the test statistic $\frac{\sum_{i \in S} x_i}{\sqrt{|S|}}$. We initialize SA with the result from NB and run for 40 restarts.

To quantify detection performance, we threshold the scan statistics given by the algorithms with various threshold values and compute missed detection and false positive rates over a number of samples (50 for MD and LMIT, 25 for SA and NB) generated from both H_0 and H_1 . We

then compute the area under the curve (AUC) generated by the pairs of missed detection and false positive rates corresponding to threshold values.

Sensitivity of MD to the choice of γ : We first investigate the sensitivity of detection performance of MD to γ , which serves the purpose of parameterizing the internal conductance of candidate subgraphs. We note that unlike (Qian & Saligrama, 2014; Qian et al., 2014), we do not explicitly specify or search over different cluster sizes, but size information is also implicitly incorporated in γ . We run MD on the range of values 0.01 to 5 in 10 logarithmic intervals. We illustrate the obtained AUC values for different SNR's in Figure 2.

We observe that while optimal γ values differ slightly with different SNR levels, the 0.3–0.7 value range is mostly optimal in all cases. This is in accordance with our expectations, since the size and conductance of the ground truth anomalies do not change across SNR levels.

Comparison to related methods: We also compare AUC performance of MD to aforementioned methods and tabulate the results in Table 1. For MD we use a γ value of 0.7 and for LMIT we use a γ value of 0.3 which we observed to perform best empirically. We see that MD performs relatively similar to LMIT, with better performance at some SNR levels. This is expected since the LMI connectivity constraints in both methods are very similar, even though the relaxation to the space of matrices M differ. On the other hand SA and NB perform worse, with SA not significantly improving upon the results of NB. It is also notable that the performance of these three methods seem low when compared to their performance in (Qian & Saligrama, 2014) which considered a very similar experimental setup. However this can be partially explained by the different scan statistics used: Aforementioned work utilized the Poisson likelihood test whereas we used the simpler linear form we specified above to be better in line with the scan statistic of MD. It is also possible that the performance of SA can be improved with a larger number of restarts.

AUC	λ_1/λ_0			Runtime
	1.1	1.3	1.5	
MD	0.74	0.79	0.92	0.8s
LMIT	0.65	0.81	0.86	3s
SA	0.57	0.67	0.72	~3m
NB	0.57	0.67	0.68	5s

Table 1. AUC performance of various algorithms with different SNR values.

Run-time comparison: We also provide the average run-time for the recovery methods for a single set of measure-

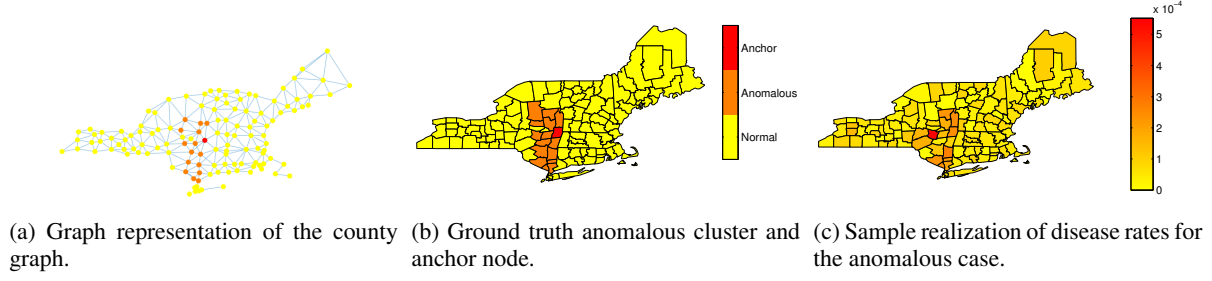
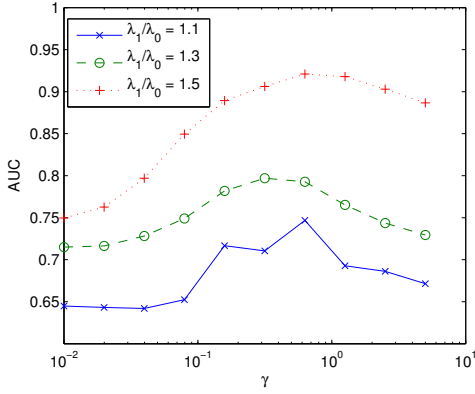


Figure 1. Disease outbreak detection in northeastern United States counties.


 Figure 2. Performance of MD algorithm for different γ values.

ments in Table 1, where the experiments were run on MATLAB on a computer with an Intel i5 4590 processor. We expect that the MD method scale better with larger problems than LMIT that uses cvx/sedumi, which is an off-the-shelf solver for convex problems. We also have not implemented any dimensionality reduction schemes for solving the MD iterations more efficiently as we remarked in Section 4.

7. Conclusion

We investigated the problem of connected subgraph detection on graph-structured data and proposed an efficient and compact convex relaxation that provably characterizes connectivity and conductance. We then developed a novel iterative optimization method that is provably nearly-linear time in the number of nodes in the graph. We demonstrated its statistical performance on experiments with real world networks, outperforming state-of-the-art methods on the problem while being computationally efficient and scalable.

Appendices

A. Proof of Theorem 3.3 and an alternative formulation using effective resistance

In this section we offer an alternative approach to obtain the inequality constraint (9) through electrical networks and the concept of effective resistance. Using this formulation, we then prove Theorem 3.3 that shows that the inequality constraint enforces connected through a simple rounding of M .

We shortly introduce the concept of effective resistance in the electrical network interpretation of graphs. In contrast to s - t flow that interprets edge weights on a graph as flow capacities, electrical flow considers edge weights w_{ij} as the conductance (inverse resistance) $\frac{1}{r_{ij}}$ between two nodes i, j .

Define the pseudoinverse of a Laplacian $L^+ = \sum_{i=2}^n \frac{1}{\lambda_i} v_i v_i^\top$ where $L = \sum_{i=2}^n \lambda_i v_i v_i^\top$ is its eigendecomposition, and we note that the minimum eigenvalue λ_1 corresponding to all 1 eigenvector is zero. Also note that $L_{G[M]}$ denotes the Laplacian of the subgraph with adjacency matrix $A \odot M$ (\odot is the elementwise/Hadamard matrix product)² and let $L_{G[M]}^+$ denote its pseudoinverse.

Defining a vector of directional current flows into/out of nodes with f (e.g. where positive elements are currents into the node and vice versa), the relationship between the vector of voltages v and f in an electrical circuit graph with resistances r_{ij} are given by the relation $v = L^+ f$. This fact follows from Kirchoff's current and voltage laws (Vishnoi, 2012). The effective resistance R_{ab} between any two nodes a, b is then defined by the voltage difference $v_a - v_b = v^\top (e_a - e_b)$ when unit current is flowing between two nodes such that $f = e_a - e_b$ (or equivalently, inverse of the current flow with unit voltage difference between a and b). We then have the identity $R_{ab} = (e_a - e_b)^\top v =$

²Corresponding to the original graph with edge weights scaled by M_{ij} .

$$(e_a - e_b)^\top L^+(e_a - e_b).$$

Given a root/anchor node $r \in V$ that is assumed to be contained in the subgraph S , consider the case where a current flow of $d_i M_{ii}$ is present between r and $i \in V$ on the induced graph with Laplacian $L_{G[M]}$. Letting $m_i = d_i M_{ii}(e_r - e_i)$, the voltage vector v^i corresponding to these inputs is given by $v^i = L_{G[M]}^+ m_i$. We then obtain the resistance between nodes r and i for input current $d_i M_{ii}$ with the identity $v_r^i - v_i^i = (e_r - e_i)^\top L_{G[M]}^+ m_i = \frac{1}{d_i M_{ii}} m_i^\top L_{G[M]}^+ m_i$, which we intend to be finite for $i \in S$ (and thus $M_{ii} > 0$) if G_S is connected. For a conductance threshold τ , we would thus like to impose the constraint

$$m_i^\top L_{G[M]}^+ m_i \leq \frac{d_i M_{ii}}{\tau}, \quad (12)$$

for all $i \in V$, where the scaling with M_{ii} serves to restrict the constraint to only the nodes $i \in S$. We also remark that the constraint (12) is independent of uniform multiplicative scaling of M , thus constraining M to unit trace does not affect the choice of τ .

We next unify the n different resistance constraints to a single PSD constraint. Define the $2n \times 2n$ matrix \mathcal{A}_M as

$$\mathcal{A}_M = \begin{pmatrix} L_{G[M]} & L_{\text{Star}^{(r)}[M]} \\ L_{\text{Star}^{(r)}[M]} & \frac{1}{\tau} L_{\text{Star}^{(r)}[M]} \end{pmatrix},$$

for which we define our connectivity constraint to be $\mathcal{A}_M \succeq 0$. For $i \neq r$ consider the submatrix \mathcal{A}_i formed by the top-left $n \times n$ submatrix (i.e. $L_{G[M]}$) and the $(n+i)$ -th row and column, which results in

$$\mathcal{A}_i = \begin{pmatrix} L_{G[M]} & m_i^\top \\ m_i & \frac{d_i M_{ii}}{\tau} \end{pmatrix}.$$

Since $\mathcal{A}_M \succeq 0$ it implies that any such submatrix satisfies $\mathcal{A}_i \succeq 0$ and through the Schur complement condition that is equivalent to condition (12). We have thus shown that the condition $\mathcal{A}_M \succeq 0$ encapsulates the pairwise effective resistance constraints of the form (12). We note that we obtain a similar condition for the root node, where with currents $m_r \triangleq \sum_{j \neq r} m_j$ applied to the nodes and a convex combination of voltage differences, $m_r^\top L_{G[M]}^+ m_r$ is being compared to $\frac{\sum_{j \neq r} M_{jj}}{\tau}$.

Finally, we can again utilize the Schur complement condition for positive semidefiniteness and simplify the PSD condition $\mathcal{A}_M \succeq 0$ on the $2n \times 2n$ matrix to a PSD condition on an $n \times n$ matrix, which directly leads to inequality (9) when we replace τ with $\frac{\gamma^2}{2}$.

Now we prove Theorem 3.3 using the above reformulation of (9). Assume there exists $i \in \hat{S}$ such that there exists no path between r and i , i.e., the subgraph $G_{\hat{S}}$ is disconnected. Since $Q_\gamma(M) \succeq 0$, it follows that $\mathcal{A}_M \succeq 0$ and

thus $\mathcal{A}_i \succeq 0$ for $\tau = \frac{\gamma^2}{2}$. Through the Schur complement lemma and (12) we have that $v_r^i - v_i^i \leq \frac{1}{\tau}$, i.e., that the voltage difference between nodes r and i is finite when a current flow of $d_i M_{ii}$ is applied that is non-zero (since $M_{ii} > 0$). This voltage difference is computed for the graph G with edge weights given by M_{ij} . It is easy to see that if this voltage difference is finite, the voltage difference in the original graph G is also finite since M_{ij} are bounded. However a contradiction arises because there exists no path between i and r in G and thus the effective resistance $R_{ri} = (e_r - e_i)^\top L^+(e_r - e_i)$ is infinite.

B. Proof of Theorem 5.1

Consider the candidate primal solution $M = \frac{1}{K} 1_S 1_S^\top$ under the hypothesis H_1 where $|S| = K$. For an n -node line graph it is easy to see that this solution is primal feasible for $\gamma \leq \frac{2}{K^2}$. Then for the Poisson model we have a feasible primal value $E[xx^\top \cdot M] = \frac{1}{K} E[(\sum_{i \in S} x_i)^2] = \mu\lambda + \mu^2\lambda^2 K$.

We next find a feasible dual solution and compute its value under hypothesis H_0 . We first consider the anchored problem with anchor node $r = 1$. For simplicity we consider $D = I$ when defining the constraint $Q_\gamma(M)$; however this is not relevant to the asymptotic bound since nodes of the graph have bounded constant degree. Letting α be the Lagrange multiplier for $I \cdot M = 1$ and Y be the one for $Q_\gamma(M) \succeq 0$, we can write the dual of the optimization problem (10) without the $M \geq 0$ constraint:

$$\max_{M \succeq 0} \min_{\alpha, Y \succeq 0} \alpha(1 - I \cdot M) + Y \cdot Q_\gamma(M),$$

which can be rewritten as

$$\max_{M \succeq 0} \min_{\alpha, Y \succeq 0} \alpha + \left(xx^\top - \alpha I + \sum_{(i,j) \in E} (Y \cdot L_{ij}) e_{ij} - \gamma \sum_{j \neq r} (Y \cdot L_{rj}) e_{jj} \right) \cdot M$$

Then the dual problem to the original optimization can be written as

$$\min_{\alpha, Y \succeq 0} \alpha \quad \text{s.t.} \quad xx^\top + \sum_{(i,j) \in E} (Y \cdot L_{ij}) e_{ij} - \gamma \sum_{j \neq r} (Y \cdot L_{rj}) e_{jj} \preceq \alpha I, \quad (13)$$

Furthermore considering dual variables of the form $Y = vv^\top$, a more constrained version of the above problem is given by

$$\min_{\alpha, v} \alpha \quad \text{s.t.} \quad xx^\top + \sum_{(i,j) \in E} (v_i - v_j)^2 e_{ij} - \gamma \sum_{j=1}^n (v_r - v_j)^2 e_{jj} \preceq \alpha I$$

Assume G is a line graph on n nodes, $r = 1$ and let $v_i = \Delta(i-1)$ be the linearly increasing ‘‘voltages’’ on the nodes. Then, for the dual constraint we have

$$xx^\top + \Delta^2 A_{\text{line}} - \gamma \Delta^2 \text{diag}([0, 1, \dots, n-1]) \preceq \alpha I,$$

where A_{line} is the adjacency matrix of the line graph G . Note that $A_{\text{line}} = A_{\text{line}} - D_{\text{line}} + D_{\text{line}} = D_{\text{line}} - L_{\text{line}} \preceq D_{\text{line}} = dI$ where $d = 2$ is the upper bound on the node degrees. Then a sufficient condition for the above constraint to be satisfied is

$$xx^\top - \gamma\Delta^2 \text{diag}([0, 1, \dots, n-1]) \preceq (\alpha - d\Delta^2)I.$$

We are now interested in the maximum eigenvalue λ_{\max} of the left-hand side of the constraint, as that will give a lower bound $\alpha \geq d\Delta^2 + \lambda_{\max}$ which immediately gives a feasible α value along with a feasible $Y = vv^\top$ with the defined voltages. This in turn provides an upper bound on the optimal dual value of (13) and thus on the optimal value of the primal (10).

We note that the left-hand side is a matrix that is a rank-one matrix xx^\top plus a diagonal matrix $-\gamma\Delta^2 \text{diag}([0, 1, \dots, n-1])$. The eigenvalues of such a matrix must satisfy the following equality (Golub, 1973):

$$f_\Delta(\lambda) \triangleq \sum_{i=0}^{n-1} \frac{x_{i+1}^2}{\lambda + \gamma\Delta^2 i^2} = 1.$$

Note that $f_\Delta(\lambda)$ is a decreasing function of λ for positive values and we can assume that $\lambda_{\max} > 0$ (the converse leads to a trivial bound with $\alpha = 0$ feasible). This implies that if there exists $\lambda_0 > 0$ such that $f_\Delta(\lambda_0) \leq 1$, then $\lambda_{\max} \leq \lambda_0$.

Note that we can upper bound $f_\Delta(\lambda)$ with

$$\begin{aligned} f_\Delta(\lambda) &\leq \frac{x_{\max}^2}{\gamma\Delta^2} \sum_{i=0}^{n-1} \frac{1}{\lambda_\Delta + i^2} \leq \frac{x_{\max}^2}{\gamma\Delta^2} \int_0^n \frac{1}{\lambda_\Delta + z^2} dz \\ &= \frac{x_{\max}^2}{\gamma\Delta^2 \sqrt{\lambda_\Delta}} \arctan \frac{n}{\sqrt{\lambda_\Delta}} = \frac{x_{\max}^2}{\Delta \sqrt{\gamma\lambda}} \arctan \frac{n\Delta\sqrt{\gamma}}{\sqrt{\lambda}}, \\ &\leq \frac{\pi x_{\max}^2}{2\Delta\sqrt{\gamma\lambda}}, \end{aligned}$$

where we defined $\lambda_\Delta \triangleq \frac{\lambda}{\Delta^2\gamma}$ and used the fact that $\arctan(\cdot) \leq \frac{\pi}{2}$. Then it is clear that $f_\Delta(\lambda_0) \leq 1$ for $\lambda_0 = \frac{\pi^2 x_{\max}^4}{4\Delta^2\gamma}$.

Finally, we have that

$$\alpha = d\Delta^2 + \frac{\pi^2 x_{\max}^4}{4\Delta^2\gamma} = \pi x_{\max}^2 \sqrt{\frac{d}{\gamma}},$$

as a feasible solution to the dual by choosing $\Delta^2 = \frac{\pi x_{\max}^2}{2\sqrt{d\gamma}}$.

We next improve upon the previous bound on α . Assume $K \log^2 n < n$, $K \geq \log n$, let $K_n \triangleq K \log^2 n$ for brevity and \bar{x} such that $\max_{i=1, \dots, n} x_i^2 \leq \bar{x}^2 \log^2 n$ w.h.p. Also let

$\Delta^2 = \bar{x}^2 K \log^2 K$ and $\gamma = \frac{2}{K^2}$. Note that we can write

$$f_\Delta(\lambda) = \sum_{i=0}^{K_n-1} \frac{x_{i+1}^2}{\lambda + \gamma\Delta^2 i^2} + \sum_{i=K_n}^{n-1} \frac{x_{i+1}^2}{\lambda + \gamma\Delta^2 i^2}.$$

We first analyze the second term, for which we have

$$\begin{aligned} \sum_{i=K_n}^{n-1} \frac{x_{i+1}^2}{\lambda + \gamma\Delta^2 i^2} &\leq \frac{\bar{x}^2 \log^2(n - K_n)}{\gamma\Delta^2} \int_{K_n}^n \frac{1}{\lambda_\Delta + x^2} dx \\ &= \frac{\bar{x}^2 \log^2(n - K_n)}{\gamma\Delta^2 \sqrt{\lambda_\Delta}} \left(\arctan \left(\frac{n}{\sqrt{\lambda_\Delta}} \right) - \arctan \left(\frac{K_n}{\sqrt{\lambda_\Delta}} \right) \right) \\ &= \frac{\bar{x}^2 \log^2(n - K_n)}{\gamma\Delta^2} \left(\frac{1}{K_n} - \frac{1}{n} - O \left(\frac{1}{K_n^3} \right) \right), \end{aligned}$$

where the last equality assumes $\frac{K_n}{\sqrt{\lambda_\Delta}} \rightarrow \infty$, which we will argue with our choice of λ . Replacing the quantities with chosen scalings, we then have

$$\begin{aligned} \sum_{i=K_n}^{n-1} \frac{x_{i+1}^2}{\lambda + \gamma\Delta^2 i^2} &\leq \frac{K \log^2(n - K_n)}{2 \log^2 K} \left(\frac{1}{K_n} - \frac{1}{n} - O \left(\frac{1}{K_n^3} \right) \right) \\ &= O \left(\frac{K \log^2(n - K_n)}{K_n \log^2 K} \right) = O \left(\frac{1}{\log^2 K} \right). \end{aligned}$$

The analysis of the first term follows along the lines of the previous bound, with n replaced by K_n , where we have the upper bound

$$\begin{aligned} \sum_{i=0}^{K_n-1} \frac{x_{i+1}^2}{\lambda + \gamma\Delta^2 i^2} &\leq \frac{\bar{x}^2 \log^2 K_n}{\gamma\Delta^2} \int_0^{K_n} \frac{1}{\lambda_\Delta + z^2} dz \\ &\leq \frac{\bar{x}^2 \log^2 K_n}{\sqrt{\gamma\Delta^2\lambda}} \frac{\pi}{2} = \frac{\bar{x} \sqrt{K} \log^2 K_n}{\log K \sqrt{\lambda}} \frac{\pi}{2\sqrt{2}}. \end{aligned}$$

Letting $\lambda_0 = 4.5\pi^2 \bar{x}^2 K \log^2 K$, we see that the first term can be upper bounded by $\frac{\log^2 K_n}{9\sqrt{2} \log^2 K} \leq \frac{1}{\sqrt{2}}$ with the assumption that $K \geq \log n$ thus $\frac{\log(K \log^2 n)}{\log K} \leq 3$. We also observe that $\lambda_\Delta = \frac{\lambda_0}{\gamma\Delta^2} = 2.25\pi^2 K^2$, thus satisfying the condition $\frac{K_n}{\sqrt{\lambda_\Delta}} \rightarrow \infty$ that we used while analyzing the second term. Thus we have that

$$f_\Delta(\lambda_0) \leq \frac{1}{\sqrt{2}} + O \left(\frac{1}{\log^2 K} \right) < 1,$$

for the above choice of λ_0 , which leads to

$$\alpha = d\Delta^2 + \lambda_0 = (d + 4.5\pi^2) \bar{x}^2 K \log^2 K,$$

as a feasible dual value.

Finally note that under the H_0 hypothesis, $\bar{x}^2 \leq \lambda + \lambda^2$ due to the concentration of the Poisson variables. Considering

1100	a constant λ , we then have an upper bound to the dual op-	1155
1101	timal value under H_0 , $\alpha = O(K \log^2 K)$. It then follows	1156
1102	from standard concentration results that $\mu = \Omega(\log K)$ is	1157
1103	sufficient to obtain asymptotic separability between the pri-	1158
1104	mal optimal under H_1 and primal optimal under H_0 . It then	1159
1105	follows directly from union bounding over different anchor	1160
1106	nodes $r = 1, \dots, n$ that $\mu = \Omega(\log K \sqrt{\log n})$ is sufficient	1161
1107	in the anchor-agnostic case.	1162
1108		1163
1109		1164
1110		1165
1111		1166
1112		1167
1113		1168
1114		1169
1115		1170
1116		1171
1117		1172
1118		1173
1119		1174
1120		1175
1121		1176
1122		1177
1123		1178
1124		1179
1125		1180
1126		1181
1127		1182
1128		1183
1129		1184
1130		1185
1131		1186
1132		1187
1133		1188
1134		1189
1135		1190
1136		1191
1137		1192
1138		1193
1139		1194
1140		1195
1141		1196
1142		1197
1143		1198
1144		1199
1145		1200
1146		1201
1147		1202
1148		1203
1149		1204
1150		1205
1151		1206
1152		1207
1153		1208
1154		1209

References

- Addario-Berry, Louigi, Broutin, Nicolas, Devroye, Luc, Lugosi, Gábor, et al. On combinatorial testing problems. *The Annals of Statistics*, 38(5):3063–3092, 2010.
- Arias-Castro, Ery, Candès, Emmanuel J, Helgason, Hannes, and Zeitouni, Ofer. Searching for a trail of evidence in a maze. *The Annals of Statistics*, pp. 1726–1757, 2008.
- Arias-Castro, Ery, Bubeck, Sébastien, Lugosi, Gábor, et al. Detection of correlations. *The Annals of Statistics*, 40(1):412–435, 2012.
- Ben-Tal, Aharon and Nemirovski, Arkadi. Lectures on modern convex optimization. 2015. URL http://www2.isye.gatech.edu/~nemirovs/Lect_ModConvOpt.pdf.
- Chung, F. R. K. *Spectral Graph Theory*, volume 92. American Mathematical Society, 1997.
- Cohen, Michael B., Kelner, Jonathan A., Peebles, John, Peng, Richard, Sidford, Aaron, and Vladu, Adrian. Faster algorithms for computing the stationary distribution, simulating random walks, and more. *CoRR*, abs/1608.03270, 2016. URL <http://arxiv.org/abs/1608.03270>.
- Duczmal, Luiz and Assuncao, Renato. A simulated annealing strategy for the detection of arbitrarily shaped spatial clusters. *Computational Statistics & Data Analysis*, 45(2):269–286, 2004.
- Glaz, Joseph, Naus, Joseph, and Wallenstein, Sylvan. *Scan Statistics*. Springer Science & Business Media, 2001.
- Golub, Gene H. Some modified matrix eigenvalue problems. *Siam Review*, 15(2):318–334, 1973.
- Johnson, David S., Minkoff, Maria, and Phillips, Steven. The prize collecting steiner tree problem: Theory and practice. In *Proceedings of the Eleventh Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '00, pp. 760–769, Philadelphia, PA, USA, 2000. Society for Industrial and Applied Mathematics. ISBN 0-89871-453-2. URL <http://dl.acm.org/citation.cfm?id=338219.338637>.
- Kivinen, J and Warmuth, MK. Exponentiated Gradient Versus Gradient Descent for Linear Predictors. *Inf. Comput.*, 1997. URL <http://www.sciencedirect.com/science/article/pii/S0890540196926127>.
- Kulldorff, Martin, Huang, Lan, Pickle, Linda, and Duczmal, Luiz. An elliptic spatial scan statistic. *Stat. Med.*, 25(22):3929–3943, 2006.
- Orecchia, Lorenzo, Sachdeva, Sushant, and Vishnoi, Nisheeth K. Approximating the exponential, the lanczos method and an $\tilde{O}(m)$ -time spectral algorithm for balanced separator. In *Proceedings of the 44th symposium on Theory of Computing - STOC '12*, volume 2, pp. 1141–1160, New York, New York, USA, November 2012. ACM Press. ISBN 9781450312455. doi: 10.1145/2213977.2214080. URL <http://dl.acm.org/citation.cfm?doid=2213977.2214080>[http://arxiv.org/abs/1111.1491\\$%5Cdelimiter%22026E30F\\$%5Bhttp://dl.acm.org/citation.cfm?doid=2213977.2214080](http://arxiv.org/abs/1111.1491$%5Cdelimiter%22026E30F$%5Bhttp://dl.acm.org/citation.cfm?doid=2213977.2214080).
- Patil, GP, Taillie, C, et al. Geographic and network surveillance via scan statistics for critical area detection. *Statist. Sci.*, 18(4):457–465, 2003.
- Priebe, Carey E, Conroy, John M, Marchette, David J, and Park, Youngser. Scan statistics on Enron graphs. *Computational & Mathematical Organization Theory*, 11(3): 229–247, 2005.
- Qian, J. and Saligrama, V. Efficient minimax signal detection on graphs. In *Advances in Neural Information Processing Systems (NIPS)*, pp. 2708–2716. Curran Associates, Inc., 2014.
- Qian, J., Saligrama, V., and Chen, Y. Connected sub-graph detection. In *Proc. of the Seventeenth Int. Conf. on Artificial Intelligence and Statistics (AISTATS)*, pp. 796–804, Reykjavik, Iceland, April 2014.
- Sharpnack, James, Rinaldo, Alessandro, and Singh, Aarti. Detecting anomalous activity on networks with the graph Fourier scan statistic. *IEEE Transactions on Signal Processing*, 64(2):364–379, 2016.
- Vishnoi, Nisheeth K. Laplacian solvers and their algorithmic applications. *Theoretical Computer Science*, 8(1-2): 1–141, 2012.
- Vladu, Adrian. personal communication, 2016-02-03.
- Wu, Nannan, Chen, Feng, Li, Jianxin, Zhou, Baojian, and Ramakrishnan, Naren. Efficient nonparametric subgraph detection using tree shaped priors. In *Thirtieth AAAI Conference on Artificial Intelligence*, 2016.