Improved Hidden Clique Detection by Optimal Linear Fusion of Multiple Adjacency Matrices

Himanshu Nayar*, Benjamin A. Miller†, Kelly Geyer†, Rajmonda S. Caceres†, Steven T. Smith†, and Raj Rao Nadakuditi*
*Department of Electrical Engineering and Computer Science, University of Michigan
Ann Arbor, Michigan 48109
†Lincoln Laboratory, Massachusetts Institute of Technology
Lexington, MA 02420

***Abstract*—Graph fusion has emerged as a promising research area for addressing challenges associated with noisy, uncertain, multi-source data. While many ad-hoc graph fusion techniques exist in the current literature, an analytical approach for analyzing the fundamentals of the graph fusion problem is lacking. We consider the setting where we are given multiple Erdős-Rényi modeled adjacency matrices containing a common hidden or planted clique. The objective is to combine them linearly so that the principal eigenvectors of the resulting matrix best reveal the vertices associated with the clique. We utilize recent results from random matrix theory to derive the optimal weighting coefficients and use these insights to develop a data-driven fusion algorithm. We demonstrate the improved performance of the algorithm relative to other simple heuristics.**

I. Introduction

In a wide variety of applications, important data take the form of connections, relationships, or interactions between discrete entities. This relational structure provides additional context, improving situational awareness and enhancing the information that can be inferred from the data. A dataset rich in relational information is represented mathematically as a graph.

One particular problem of interest when working with graph-based data is subgraph detection [1]. Given a graph, the objective is to determine whether the relationships observed are consistent with "normal" behavior in the network, or if there is a subset of vertices—comprising a subgraph within the overall graph—that exhibits a topology that is contrary to the expectation. Several applications are focused on finding denser-than-usual subgraphs within larger graphs, such as detecting communities in social networks [2] and finding highly interactive subsets among proteins [3]. A simplified form of this problem is planted clique detection, where the objective is to detect a fully connected subgraph placed within a larger graph with edges that occur with equal probability. This subset of the more general subgraph detection problem provides mathematical tractability that enables analysis of the fundamental limits of subgraph detectability.

Planted clique detection has a long history within theoretical computer science. Algorithms have been developed using a variety of techniques, including combinatorial search methods [4], spectral methods [5], and statistical query methods [6]. Recent work has focused on deriving analytical bounds for subgraph detectability, including technique-independent bounds [7], and bounds on detection using spectral techniques [8]. Spectral methods are particularly interesting, since they are computationally efficient and yield principled analytical bounds using random matrix theory. Similar bounds for the planted partition model have also been shown to predict a detectability phase transition in other efficient detection methods [9].

While the planted clique problem has traditionally been studied in the context of detection within a single graph, in practice, a network is often the result of fused information from multiple disparate sources. Understanding the implications that the algorithms for a single observation have on multi-source data is extremely important as data analysts are increasingly required to make decisions in this setting.

The objective of this paper is to extend the recent spectral detection bounds to cases where multiple graphs are observed. In this setting, the clique exists in each graph, each of which has a different edge probability. We adopt a linear fusion model in which we analyze a convex combination of the adjacency matrices of the graphs. Within this context, we demonstrate that the optimal fusion method is highly intuitive: weighting the graphs in inverse proportion to their expected background degrees. In a set of experiments on random graphs, we test subgraph detectability at several levels of difficulty, varying the relative density of the background graphs, the clique size, and the weights. In all cases, the optimality of the solution we derive is verified.

The remainder of this paper is organized as follows. In Section II, we define the problem model and formalize the mathematical context in which the graphs are fused. Section III provides a derivation of the formula for optimal fusion within this context. In Section IV, we outline several experiments that empirically validate the derived optimal fusion method.

The Lincoln Laboratory portion of this work is sponsored by the Assistant Secretary of Defense for Research & Engineering under Air Force Contract FA8721-05-C-0002. Opinions, interpretations, conclusions and recommendations are those of the authors and are not necessarily endorsed by the United States Government. The University of Michigan portion of this work was sponsored by the Office of the Assistant Secretary of Defense for Research and Engineering.

Section V concludes the paper with a brief summary and directions for future work.

II. Problem Model

A. Planted Clique Detection

In the planted clique problem, we are given a graph $G = (V, E)$, where V is the set of vertices (representing the entities) and E is the set of edges (representing connections). We will denote $N = |V|$ and $M = |E|$. The degree of a vertex is the number of edges connected to it. We will denote the average degree of the graph by c. A subset of vertices, $V_S \subset V$, comprise the clique, meaning that for all $v, u \in V_S$, there is an edge between v and u in E. We will denote the clique size by $k = |V_S|$. In this paper, we will only consider graphs that are undirected (so edges are unordered pairs of vertices) and unweighted (so edges either exist or not, with no notion of connection strength). If either $v \notin V_S$ or $u \notin V_S$, then an edge exists between v and u with probability p, which is constant across all pairs of vertices and independent of the existence of other edges.

Spectral methods make use of matrix representations of a graph. The most basic matrix representation of a graph is the adjacency matrix. For an unweighted, undirected graph, the adjacency matrix $A = \{a_{ij}\}$ is an $N \times N$ binary matrix where $a_{ij} = 1$ if there is an edge between vertices i and j, and $a_{ij} = 0$ otherwise. (This requires an arbitrary labeling of vertices with integers from 1 to N.) Since we consider undirected graphs, A will be symmetric.

A modification of the adjacency matrix used for community detection is the modularity matrix [10]. The modularity matrix is a residuals matrix: the observed adjacency matrix minus its expected value. Since the planted clique problem assumes a background with equal probability, we use the modularity matrix with respect to the Erdős-Rényi model:

$$B := A - p\mathbf{1}_{N\times N}. \tag{1}$$

This technique cancels out the effects of typical background behavior and allows the detection of deviations from the expectation; in this case the planted clique.

The form of (1) is a rank-1 perturbation of a Wigner matrix, and recent work has defined a sharp threshold for detectability of such a perturbation [11]. This analysis was applied to planted clique detection in [8], where a simple algorithm was used to detect the clique: compute the (unit-normalized) principal eigenvector of B, denoted by u. Since the entries in the principal eigenvector of a Wigner matrix appear normally distributed, u is thresholded, with a false alarm rate based on this distribution, and the estimate of the clique vertices is given by:

$$\hat{V}_S = \left\{ v_i : |\sqrt{N} u_i| > F^{-1}_{\mathcal{N}(0,1)}\left(1 - \frac{\alpha}{2}\right)\right\}, \tag{2}$$

where $F^{-1}_{\mathcal{N}(0,1)}$ is the inverse cumulative density function of a standard normal distribution and α is the desired false alarm probability. Using this algorithm, the following bound was derived.

Claim 2.1 (Nadakuditi [8]): Consider a k-vertex clique planted in an N-vertex graph with edge probability p, where the clique vertices are identified using (2) for a significance level α. Then, for fixed p, as $k, N \to \infty$ such that $k/\sqrt{N} \to \beta \in (0, \infty)$ we have

$$\mathbb{P}(\text{clique discovered}) \xrightarrow{a.s.} \begin{cases} 1 & \text{if } \beta > \beta_{\text{crit.}} := \sqrt{\frac{p}{1-p}} \\ \alpha & \text{otherwise.} \end{cases} \tag{3}$$

B. Multi-Source Graph Fusion

Our objective is to derive a bound analogous to Claim 2.1 for fusion of multiple graphs. For the multi-source setting, we assume we have m graphs $G_i = (V, E_i)$ for $1 \le i \le m$. Note that the vertex set is the same for all graphs (and, in matrix form, the indices are consistent across observations). The non-clique edges are generated independently in each graph. To analyze the multigraph in the same context as the claim, we combine the adjacency matrices of the graphs into a single matrix, and compute the principal eigenvector of its residuals. We take a linear combination of the adjacency matrices A_i with weights w_i to create a fused adjacency matrix:

$$\tilde{A} = \sum_{i=1}^{m} w_i A_i. \tag{4}$$

We consider only positive weights that sum to 1, meaning that $\tilde{A}$ is a convex combination of the adjacency matrices. Using this convention, the value on the edges between clique vertices remains 1, since these edges exist in each observation. Applying the same weighting to the expected values (p_i being the background edge probability for G_i), we maintain a residuals matrix where the majority of the entries (all of those not part of the clique) have the same zero-mean distribution. The fused residuals matrix is given by:

$$\tilde{B} := \sum_{i=1}^{m} (A_i - p_i \mathbf{1}_{N\times N}) = \tilde{A} - \left(\sum_{i=1}^{m} w_i p_i\right)\mathbf{1}_{N\times N}. \tag{5}$$

By applying the algorithm defined in Section II-A to the principal eigenvector of $\tilde{B}$, we can improve detection performance over what would be possible with a single observation. As demonstrated in the next section, we can derive analytically optimal weights in this setting by minimizing the variance of the entries.

III. Optimal Linear Fusion

We begin by modeling the fused adjacency matrix $\tilde{A}$ in a way that enables random matrix theoretic analysis. Without loss of generality, we can permute the vertex indices so that the clique vertices have indices 1 to k. We have:

$$\tilde{A} = \sum_{i=1}^{m} w_i \begin{bmatrix} \mathbf{1}_{k\times k} & \mathbb{B}_{k\times N'}(p_i) \\ \mathbb{B}_{N'\times k}(p_i) & \mathbb{B}_{N'\times N'}(p_i) \end{bmatrix} \tag{6}$$

$$= \begin{bmatrix} \sum_{i=1}^{m} w_i \mathbf{1}_{k\times k} & \sum_{i=1}^{m} w_i \mathbb{B}_{k\times N'}(p_i) \\ \sum_{i=1}^{m} w_i \mathbb{B}_{N'\times k}(p_i) & \sum_{i=1}^{m} w_i \mathbb{B}_{N'\times N'}(p_i) \end{bmatrix} \tag{7}$$

$$= \begin{bmatrix} \sum_{i=1}^{m} w_i \mathbf{1}_{k\times k} & \sum_{i=1}^{m} w_i p_i \mathbf{1}_{k\times N'} \\ \sum_{i=1}^{m} w_i p_i \mathbf{1}_{N'\times k} & \sum_{i=1}^{m} w_i p_i \mathbf{1}_{N'\times N'} \end{bmatrix} \tag{8}$$

$$+ \begin{bmatrix} \mathbf{0}_{k\times k} & \sum_{i=1}^{m} w_i \mathbb{B}^c_{k\times N'}(p_i) \\ \sum_{i=1}^{m} w_i \mathbb{B}^c_{N'\times k}(p_i) & \sum_{i=1}^{m} w_i \mathbb{B}^c_{N'\times N'}(p_i) \end{bmatrix} \tag{9}$$

$$= \mathbb{E}\left[\tilde{A}\right] + X, \tag{10}$$

where $\mathbb{B}_{N_1\times N_2}(p)$ is an $N_1 \times N_2$ matrix of Bernoulli random variables, each drawn independently with probability p, $\mathbb{B}^c_{N_1\times N_2}(p)$ is the centered version of this matrix, where the random variables have had their expected value subtracted. For convenience, we define X to be the random deviations from the mean from the matrix on line (9). Thus, $\mathbb{E}[\tilde{A}]$ is a rank-2 matrix, and $\tilde{A}$ is a random deviation from this low-rank structure. The random entries in X (i.e., those outside of the clique) have variance $\sum_{i=1}^{m} w_i^2 p_i(1-p_i)$. If k grows more slowly than N, then X will tend toward a Wigner matrix, where the support of the eigenvalue distribution (which determines the noise power, and, thus, the detectability of the clique) scales with the standard deviation of these entries. Subtracting the expected value of the fused matrix, we are left with a rank-1 perturbation of the random matrix. This rank-1 perturbation has values of $1-\sum_{i=1}^{m} w_i p_i$ in the entries where both vertices are part of the clique, and 0 elsewhere. The nonzero eigenvalue of this matrix is:

$$\theta_1 = k\left(1 - \sum_{i=1}^{m} w_i p_i\right).$$

At this point, we have a similar setting as in [8]. In this case, the Wigner matrix has an eigenvalue distribution that tends to a semicircle with radius:

$$R = \sqrt{4N\sum_{i=1}^{m} w_i^2 p_i\,(1-p_i)} = \sqrt{4\sum_{i=1}^{m} w_i^2 c_i\,(1-p_i)}, \tag{11}$$

whereas in the single-source case the radius is $\sqrt{4Np(1-p)}$. The rank-1 perturbation in the new setting is θ_1 rather than simply $k(1-p)$. We can apply similar reasoning to develop a new bound for planted clique detection using the fused modularity matrix with given weights. First, we introduce an approximation that will enhance the interpretability of the result. As graphs grow large, their density tends to decrease. That is, the average degree of the vertices grows slowly, not proportionally to N as it would if p remained constant. Thus, we will focus on a case where $p \longrightarrow 0$ as $N \longrightarrow \infty$. Specifically, we consider the case where the average degree c remains constant. This allows us to express the detectability bound in terms of the average degrees of the graphs, c_i, for large values of N.

The relationship between the maximum eigenvalue of the rank-1 perturbation and the radius of the semicircle in Claim 2.1 is:

$$k(1-p) > \sqrt{Np(1-p)}.$$

Thus, in the similar setting for fused graphs, we want $\theta_1 > R/2$. Since we are assuming c_i remains constant, we can use $(1-p_i) \approx 1$ to approximate the quantities of interest as $\theta_1 \approx k$ and $R/2 \approx \sqrt{\sum_{i=1}^{m} w_i^2 c_i}$. Since the approximate eigenvalue of the rank-1 perturbation is independent of the weights, the objective to maximize detectability is equivalent to that of minimizing R.

To incorporate the constraint that the weights sum to 1, we optimize the Lagrange function:

$$\mathcal{L}(w,\lambda) = \sqrt{\sum_{i=1}^{m} w_i^2 c_i} + \lambda\left(\sum_{i=1}^{m} w_i - 1\right). \tag{12}$$

Setting $\partial\mathcal{L}/\partial w_j$ to zero yields:

$$w_j c_j = -\lambda\sqrt{\sum_{i=1}^{m} w_i^2 c_i}. \tag{13}$$

Since the right hand side of (13) is constant across weights, this critical point exists where each weight is inversely proportional to the average degree of its associated graph. Thus, the optimal weighting scheme for minimizing the support of the eigenvalues is given by:

$$w_j = \frac{1/c_j}{\sum_{i=1}^{m} 1/c_i} = \frac{1}{1+\sum_{i\neq j} c_j/c_i}. \tag{14}$$

This result provides a mathematical justification for what would be an intuitive heuristic: giving the observations that are noisier a proportionally lower weight. For example, if all graphs have equal average degree, then they will be weighted equally. Conversely, if one graph has a much smaller average degree than the others, its weight will be close to 1 while the others will be close to 0. In conjunction with the value of R, this also demonstrates how additional information improves detectability. Indeed, if G_i has substantially lower degree than the other graphs, the detection threshold will remain approximately $\sqrt{c_i}$, since the other information is much noisier and most of the emphasis will be placed on the graph with the sparsest background. On the other hand, if all of the graphs have equal average degree, the detection threshold will be reduced by $\sqrt{m}$. As we show in the next section, the optimality of this weighting scheme is verified by empirical detection performance.

IV. Simulation Results

In each of the following experiments, we fuse two Erdős-Rényi graphs, each with a planted clique. The graphs each have 10,000 vertices, and we vary the clique size, the average degrees of the backgrounds, and the weights to demonstrate the optimality of the solution derived in Section III. In each case, we set the false alarms rate α from (2) to 0.05, and

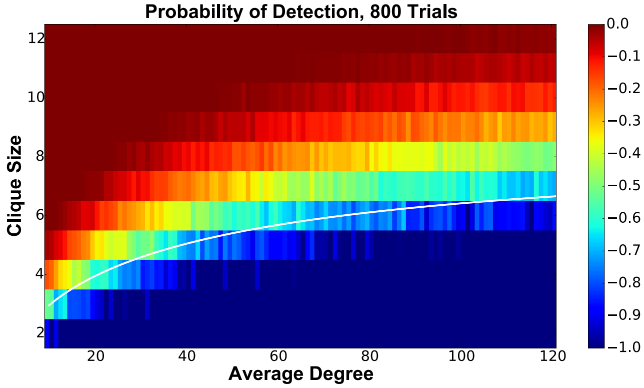

Fig. 1. Probability of detection for a clique of varying size, embedded into one graph with average degree 70 and one with average degree that varies. The optimized weights are used to fuse the graphs. The detection probability is shown on a base-10 logarithmic scale, and the detection threshold is drawn in white.

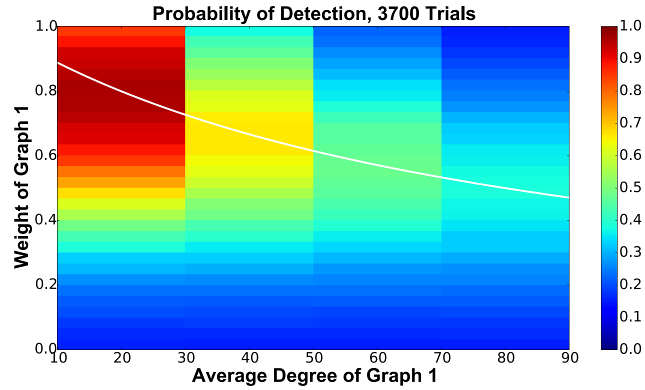

Fig. 2. Probability of detection for a 7-vertex clique, embedded into one graph with average degree 80 and one with average degree that varies. The weights used for fusion are varied to demonstrate optimality of the derived solution, which is drawn in white.

compute the empirical probability of detection of the clique vertices when thresholding the principal eigenvector of $\tilde{B}$ at the corresponding level.

We begin by demonstrating that the improved bound correctly predicts where the clique becomes detectable. In this experiment, we set the average degree of one graph to 70, and vary the average degree of the other graph from 10 to 120. We independently vary the size of the clique from 2 to 12. For each trial, the optimized weights from (14) are used. The empirical probability of detection is shown in Fig. 1. Probability of detection is compared to the detectability threshold:

$$k = \sqrt{w_1^2 c_1 + w_2^2 c_2}.$$

Once k becomes larger than the threshold, its detectability begins to increase until the probability of detection eventually reaches 1. Using only the graph with varying degree, the detection threshold would be the square root of the average degree on the horizontal axis (e.g., a clique of size 10 at an average degree of 100). Using the additional information provided by the other graph, the detection threshold is substantially lowered on the right side of the plot, where the graph with variable average degree is the densest.

Our second experiment demonstrates the optimality of the derived solution. We again fix the average degree of one of the graphs (in this case $c_2 = 80$), and now fix the clique size to $k = 7$. We then vary the average degree of G_1 and its corresponding weight $w_1 = 1 - w_2$. We consider $c_1 \in \{20, 40, 60, 80\}$. Probability of detection is shown in Fig. 2, where the empirical results are compared to the derived optimal weight $w_1 = 1/(1+(c_1/70))$. For all values of c_1, the empirical detection rate is maximized at the derived weight. This is true when c_1 is small, and most of the weight is placed on w_1, and when c_1 is large and the weights are approximately equal.

This behavior holds for various clique sizes. Fig. 3 illustrates detection probability for three cases. In each case, two graphs are fused, with a planted clique whose size is at the detection threshold for the sparser graph (i.e., $k = \sqrt{c_1}$, where $c_1 \le c_2$). We choose a case where the optimal weighting has $w_1 = 0.5$, one with $w_1 = 0.7$, and one with $w_1 = 0.9$. By simply averaging the graphs, we would achieve the detection probability at $w_1 = 0.5$, and the only case where this is optimal is the case where the average degrees are equal. In the other cases, while the improvement over only considering the sparser graph (the detection probability achieved at the extreme right of the plot), is more subtle, there is a substantial improvement that is maximized at the theoretically determined weighting.

Finally, we consider a case where we vary the average degrees of both graphs independently, and use the optimal weighting. In this experiment, the size of the planted clique is $k = 9$. Detection probabilities are shown in Fig. 4. First, note that the probability of detection increases as either graph gets sparser. Curves indicating the detection thresholds for cliques of size 5 to 8 are overlayed in the plot. These curves follow the formula:

$$c_2 = \frac{1}{1/k^2 - 1/c_1}.$$

As we expect, at the thresholds for smaller cliques, the 9-vertex clique is more likely to be detected. It is also noteworthy that, along the curve where the threshold is the same, the empirical detection probability remains consistent. This provides additional validation that the optimal weighting is correct: Using the optimal weighting at a given threshold gives a consistent detection probability, regardless of the densities of the individual graphs.

V. Conclusion

This paper extends recent spectral bounds for planted clique detection to cases where multiple graphs are observed. Operating in a multi-source setting has become extremely important in recent years, as correlating observations from multiple datasets has become more common. We demonstrate that the intuitive approach of weighting each graph in inverse proportion to its average degree is, in fact, the optimal technique

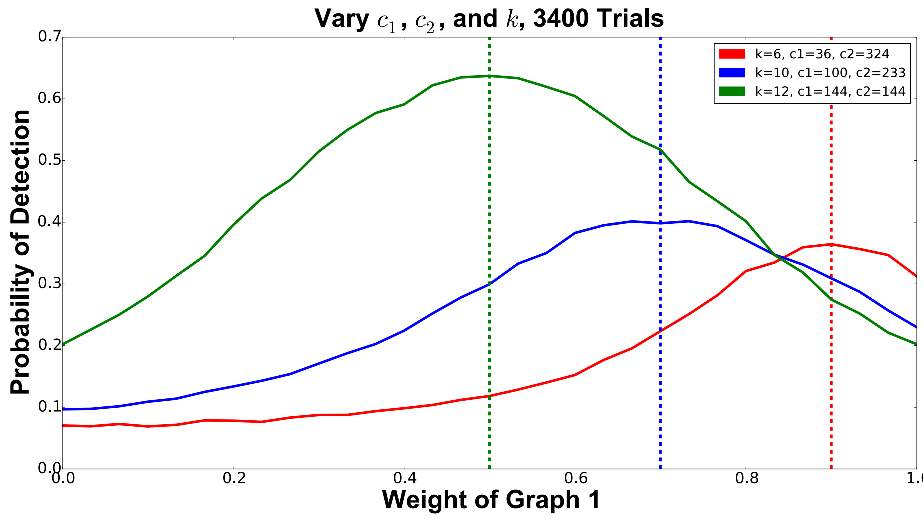

Fig. 3. Probability of detection with respect to the weight of G_1 in three scenarios. In each case, the theoretically optimal weighting is indicated by a vertical dashed line of the same color as the curve.

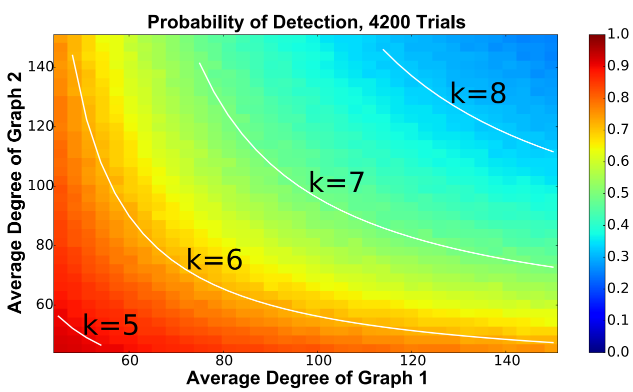

Fig. 4. Probability of detection for a 9-vertex clique, embedded into two graphs of varying degree. Optimized weights are used to fuse the graphs. Detection thresholds for smaller cliques (5 to 8 vertices) are drawn in white.

under a linear weighting scheme. Empirical results confirm the theory in a wide range of settings.

There are many possible future directions for this work. A simple extension of the result to dense subgraphs, rather than cliques, follows rather directly. Extending results to more complicated background models—such as Chung-Lu models, with arbitrary average degree, or stochastic blockmodels, with inherent community structure—would be another natural progression. This could also apply to cases where the subgraph changes over the observations, as in [12]. As data are often correlated across sources, extending to a setting where there are dependencies across the observations would also be beneficial.

References

[1] B. A. Miller, M. S. Beard, P. J. Wolfe, and N. T. Bliss, "A spectral framework for anomalous subgraph detection," *IEEE Trans. Signal Process.*, vol. 63, no. 16, pp. 4191–4206, Aug 2015.

[2] M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Phys. Rev. E*, vol. 69, no. 2, 2004.

[3] D. Bu, Y. Zhao, L. Cai, H. Xue, X. Zhu, H. Lu, J. Zhang, S. Sun, L. Ling, N. Zhang, G. Li, and R. Chen, "Topological structure analysis of the protein–protein interaction network in budding yeast," *Nucleic Acids Research*, vol. 31, no. 9, pp. 2443–2450, 2003.

[4] M. Jerrum, "Large cliques elude the Metropolis process," *Random Structures & Algorithms*, vol. 3, pp. 347–359, 1992.

[5] N. Alon, M. Krivelevich, and B. Sudakov, "Finding a large hidden clique in a random graph," in *Proc. ACM-SIAM Symp. Discrete Algorithms*, 1998, pp. 594–598.

[6] V. Feldman, E. Grigorescu, L. Reyzin, S. Vempala, and Y. Xiao, "Statistical algorithms and a lower bound for detecting planted clique," 2015, preprint: arXiv.org:1201.1214.

[7] E. Arias-Castro and N. Verzelen, "Community detection in random networks," 2013, preprint: arXiv.org:1302.7099.

[8] R. R. Nadakuditi, "On hard limits of eigen-analysis based planted clique detection," in *Proc. IEEE Statistical Signal Process. Workshop*, 2012, pp. 129–132.

[9] R. R. Nadakuditi and M. E. J. Newman, "Graph spectra and the detectability of community structure in networks," *Phys. Rev. Lett.*, vol. 108, p. 188701, May 2012. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevLett.108.188701

[10] M. E. J. Newman, "Finding community structure in networks using the eigenvectors of matrices," *Phys. Rev. E*, vol. 74, no. 3, 2006.

[11] F. Benaych-Georges and R. R. Nadakuditi, "The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices," *Advances in Math.*, vol. 227, no. 1, pp. 494–521, May 2011.

[12] B. A. Miller and N. T. Bliss, "Toward matched filter optimization for subgraph detection in dynamic networks," in *Proc. IEEE Statistical Signal Process. Workshop*, 2012, pp. 113–116.