Random Tensors and Planted Cliques

S. Charles Brubaker and Santosh S. Vempala*

Georgia Institute of Technology
Atlanta, GA 30332
{brubaker,vempala}@cc.gatech.edu

Abstract. The r-parity tensor of a graph is a generalization of the adjacency matrix, where the tensor's entries denote the parity of the number of edges in subgraphs induced by r distinct vertices. For $r = 2$, it is the adjacency matrix with 1's for edges and -1's for nonedges. It is well-known that the 2-norm of the adjacency matrix of a random graph is $O(\sqrt{n})$. Here we show that the 2-norm of the r-parity tensor is at most $f(r)\sqrt{n}\log^{O(r)} n$, answering a question of Frieze and Kannan [1] who proved this for $r = 3$. As a consequence, we get a tight connection between the planted clique problem and the problem of finding a vector that approximates the 2-norm of the r-parity tensor of a random graph. Our proof method is based on an inductive application of concentration of measure.

1 Introduction

It is well-known that a random graph $G(n, 1/2)$ almost surely has a clique of size $(2 + o(1)) \log_2 n$ and a simple greedy algorithm finds a clique of size $(1 + o(1)) \log_2 n$. Finding a clique of size even $(1+\epsilon) \log_2 n$ for some $\epsilon > 0$ in a random graph is a long-standing open problem posed by Karp in 1976 [2] in his classic paper on probabilistic analysis of algorithms.

In the early nineties, a very interesting variant of this question was formulated by Jerrum [3] and by Kucera [4]. Suppose that a clique of size p is planted in a random graph, i.e., a random graph is chosen and all the edges within a subset of p vertices are added to it. Then for what value of p can the planted clique be found efficiently? It is not hard to see that $p > c\sqrt{n \log n}$ suffices since then the vertices of the clique will have larger degrees than the rest of the graph, with high probability [4]. This was improved by Alon et al [5] to $p = \Omega(\sqrt{n})$ using a spectral approach. This was refined by McSherry [6] and considered by Feige and Krauthgamer in the more general semi-random model [7]. For $p \geq 10\sqrt{n}$, the following simple algorithm works: form a matrix with 1's for edges and -1's for nonedges; find the largest eigenvector of this matrix and read off the top p entries in magnitude; return the set of vertices that have degree at least $3p/4$ within this subset.

* Supported in part by NSF award CCF-0721503 and a Raytheon fellowship.

I. Dinur et al. (Eds.): APPROX and RANDOM 2009, LNCS 5687, pp. 406–419, 2009.
© Springer-Verlag Berlin Heidelberg 2009

The reason this works is the following: the top eigenvector of a symmetric matrix A can be written as

$$\max_{x:\|x\|=1} x^T Ax = \max_{x:\|x\|=1} \sum_{ij} A_{ij} x_i x_j$$

maximizing a quadratic polynomial over the unit sphere. The maximum value is the spectral norm or 2-norm of the matrix. For a random matrix with $1, -1$ entries, the spectral norm (largest eigenvalue) is $O(\sqrt{n})$. In fact, as shown by Füredi and Komlós [8],[9], a random matrix with i.i.d. entries of variance at most 1 has the same bound on the spectral norm. On the other hand, after planting a clique of size $\sqrt{n}$ times a sufficient constant factor, the indicator vector of the clique (normalized) achieves a higher norm. Thus the top eigenvector points in the direction of the clique (or very close to it).

Given the numerous applications of eigenvectors (principal components), a well-motivated and natural generalization of this optimization problem to an r-dimensional tensor is the following: given a symmetric tensor A with entries $A_{k_1 k_2 \ldots k_r}$, find

$$\|A\|_2 = \max_{x:\|x\|=1} A(x, \ldots, x),$$

where

$$A(x^{(1)}, \ldots, x^{(r)}) = \sum_{i_1 i_2 \ldots i_r} A_{i_1 i_2 \ldots i_r} x^{(1)}_{i_1} x^{(2)}_{i_2} \ldots x^{(r)}_{i_r}.$$

The maximum value is the spectral norm or 2-norm of the tensor. The complexity of this problem is open for any $r > 2$, assuming the entries with repeated indices are zeros.

A beautiful application of this problem was given recently by Frieze and Kannan [1]. They defined the following tensor associated with an undirected graph $G = (V, E)$:

$$A_{ijk} = E_{ij} E_{jk} E_{ki}$$

where E_{ij} is 1 is $ij \in E$ and -1 otherwise, i.e., A_{ijk} is the parity of the number of edges between i, j, k present in G. They proved that for the random graph $G_{n,1/2}$, the 2-norm of the random tensor A is $\tilde{O}(\sqrt{n})$, i.e.,

$$\sup_{x:\|x\|=1} \sum_{i,j,k} A_{ijk} x_i x_j x_k \le C\sqrt{n} \log^c n$$

where c, C are absolute constants. This implied that if such a maximizing vector x could be found (or approximated), then we could find planted cliques of size as small as $n^{1/3}$ times polylogarithmic factors in polynomial time, improving substantially on the long-standing threshold of $\Omega(\sqrt{n})$.

Frieze and Kannan ask the natural question of whether this connection can be further strengthened by going to r-dimensional tensors for $r > 3$. The tensor itself has a nice generalization. For a given graph $G = (V, E)$ the r-parity tensor is defined as follows. Entries with repeated indices are set to zero; any other