EDGE SEARCH IN GRAPHS AND HYPERGRAPHS OF BOUNDED RANK: AVERAGE SEARCH LENGTHS IN THE WEIGHTED CASE

Ingo Althöfer
Fakultät für Mathematik
Universität Bielefeld
Postfach 100 131
33501 Bielefeld
Germany

Abstract

Consider a finite undirected graph $G = (X, E)$ with a probability distribution $P = (p_1, p_2, \ldots, p_m)$ on the set E of its edges. What is the average case complexity $\overline{L}(G, P)$ of finding an unknown edge $e^* \in E$ which is distributed according to P if the following tests are admitted: For any $Y \subset X$ we may test whether $e^* \subset Y$ or not.

We prove the existence of a constant c_2 such that $H(p_1, \ldots, p_m) \leq \overline{L}(G, P) \leq H(p_1, \ldots, p_m) + c_2$ for all pairs (G, P). Here $H(P)$ denotes the entropy of the distribution P. The upper bound stems from a construction which is presented in this note.

The result is generalized to hypergraphs of bounded rank r. A hypergraph is said to have rank r if each hyperedge in it contains at most r vertices.

1. Introduction

Suppose $G = (X, E)$ is a finite undirected graph without loops or multiple edges. In [Aig 86], Aigner introduced the following search problem on graphs which generalizes group testing problems with two defectives:

The search domain is the edge set E of G. Admissible tests are the following: For each subset $Y \subset X$ we may test whether the unknown edge e^* is in the induced subgraph $G[Y]$ or not. Let $L(G)$ denote the worst case complexity of this search problem.

The information theoretic lower bound is $L(G) \geq \lceil \log_2 |E| \rceil$, and there are graphs G with $L(G) = \lceil \log_2 |E| \rceil + 1$.

Recently Damaschke proved [Dam 93] that for all graphs G

$$L(G) \leq \lceil \log_2 |E| \rceil + 1 \ .$$

In this note we investigate the *average* search costs in case of a probability distribution $P = (p_1, \ldots, p_m)$ on the set E of edges. For a fixed search scheme and every edge $e \in E$ let $\mathrm{cost}(e)$ denote the number of questions required in case of $e^* = e$. We want to find a scheme that minimizes $\sum_{e \in E} p(e)\mathrm{cost}(e)$.

For given G and P , let $\overline{L}(G, P)$ denote the minimum possible value for this sum. The information theoretic lower bound is

$$\overline{L}(G, P) \geq H(p_1, \ldots, p_m),$$

where $H(P) = H(p_1, \ldots, p_m) = -\sum_{i=1}^{m} p_i \log_2 p_i$ is the entropy of P .

Theorem 1: *For every graph G with a probability distribution P on its edge set we give a search scheme which yields $cost(e) \leq -\log_2 p(e) + c_2$ for every edge $e \in E$. The absolute constant c_2 does not depend on G or P .*

Corollary 2: $\overline{L}(G, P) \leq H(p_1, \ldots, p_m) + c_2$ *for every graph G and every probability distribution P on its edge set.*

This paper is organized as follows. In Section 2 we present the search scheme on which Theorem 1 is based. Section 3 contains the proof of Theorem 1. In Section 4 we sketch a straightforward generalisation of our search scheme to hypergraphs of bounded rank. A hypergraph is said to have rank r if each hyperedge in it contains at most r vertices.

Theorem 3: *For every hypergraph H of bounded rank r with a probability distribution P on its set E of hyperedges there is a search scheme which yields $cost(e) \leq -\log_2 p(e) + c_r$ for every $e \in E$. The absolute constant $c_r > 0$ depends on the rank bound r , but not on the hypergraph H and P .*

Taking the equal distribution on the set of hyperedges, our arguments give a new and simpler proof of the Hypergraph Theorem 3.4 in [AlT 93, p. 8].

Corollary 4: $\overline{L}(H, P) \leq H(P) + c_r$ *for every hypergraph H of bounded rank r and every probability distribution on its set of hyperedges.*

Section 5 contains a short discussion and several open problems.

We refer the reader to [Aig 88] for an introduction to combinatorial search theory. In [AW 87] search problems are discussed in a more general setting.