A Probabilistic Estimator for the Vertex Deletion Problem

C. A. Mandal
Dept. of Info. Sys. & Computing
Brunel University
Uxbridge UB8 3PH, U.K.

P. P. Chakrabarti S. Ghose
Department of Computer Sc. & Engg.
Indian Institute of Technology, Kharagpur
West Bengal 721 302, INDIA

Abstract

Vertex deletion is a well known NP-complete problem. An empirical method for estimating the number of vertices that need to be deleted to make a graph bipartite is presented. The estimator has been developed by modelling the problem using random graphs where an edge is present with a fixed probability p. The method works in $O(n^2)$ time, where n is the number of vertices in the graph. The estimator has been experimented on several randomly generated graphs. The results indicate that the estimator gives accurate results for graphs of various sizes.

Keywords: Vertex Deletion, Probabilistic Estimation, NP-complete Problems.

1 Introduction

Given a graph it is easy to test whether or not the graph is bipartite. If the graph is not bipartite then it may be rendered so by deleting some of its edges or vertices. These are, respectively, the problems of edge deletion and vertex deletion. Both are NP-complete problems [1, 2, 3]. In this paper we present an estimator to estimate the number of vertices that may have to be deleted to render a graph bipartite. Although such an estimator does not give us a solution to the problem, it is nevertheless useful because it does give us an estimate of the cost of the solution. This is useful in some design environments [4] where the quality of a partial design is determined by the solution to a vertex deletion problem and the associated cost. The design may involve making decisions leading to numerous instances of the vertex deletion problem. Since it is expensive to find an actual solution it is useful to be able to estimate the cost of the solution and solve a small number of problems whose cost estimates are favourable.

We model the problem using random graphs. We consider random graphs of the type $G_{n,p}$. $G_{n,p}$ denotes a class of graphs which have n vertices, and the probability that an edge will be present between any two vertices is p. We have used an empirical technique to estimate the minimum number of vertices to be deleted. This based on the number of expected bipartite of random graphs $G_{n,p}$, which is explained in the next section. The construction of the estimate is presented in section 3, followed by the exprimental results and conclusions in the following sections.

2 Expected Number of Bipartite Partitions

By definition the expected number of vertices $E(d)$ to be deleted to render graphs of the type $G_{n,p}$ bipartite is

$$E(d) = \sum_{i=0}^{n-2} iP(i,n,p),$$

where $P(i,n,p)$ is the probability that deletion of i vertices will make $G_{n,p}$ bipartite. We do not attempt to find an analytic expression for $P(i,n,p)$ because of the complexity involved. Instead we take an alternative approach to estimate $E(d)$.

Consider the graph $G(n,p)$, with n vertices and each of the $\binom{n}{2}$ edges being present with probability p. Let $q=(1-p)$ be the probability that the edge is absent. In general this graph will not be bipartite. We wish to consider a subgraph of G containing m of the n vertices, which will be bipartite. This subgraph being bipartite, there will be at least one partition of these m vertices into two independent sets, an independent set being a set of vertices without an edge between any pair of vertices. Consider the possible bi-partitioning of such a set of m vertices of the graph G. Let k vertices be in one partition and $(m-k)$ remaining vertices in the other partition. Let the remaining d vertices, where $m+d=n$, vertices be denoted by the set D. We would like to estimate $d=|D|$ when D represents the minimal number of vertices to be deleted to make $G(n,p)$ bipartite. Unless D is minimal, some of the vertices in D may be easily augmented to one the two sets, and a smaller number of vertices would have to be deleted. This would make the estimate quite useless.

Let $l=\lfloor\frac{m}{2}\rfloor$. Using the probability of an edge being absent between a pair of vertices from both these partitions, we compute two quantities S_1 and S_2 such that:

$$S_1=\sum_{k=1}^{\lfloor\frac{m}{2}\rfloor-1}\binom{m}{k}q^{\frac{k(k-1)}{2}+\frac{(m-k)(m-k-1)}{2}}(1-q^k)(1-q^{(m-k)})^d$$

S_1 represents the expected number of bipartite partitions of m vertices when k are in one partition and $m-k$ are in the other, $k=1..\lfloor\frac{m}{2}\rfloor-1$, and

$$S_2=\binom{m}{l}q^{\frac{l(l-1)}{2}+\frac{(m-l)(m-l-1)}{2}}(1-q^l)(1-q^{(m-l)})^d$$

S_2 represents the expected number of bipartite partitions of m vertices when l are in one partition and $m-l$ are in the other. S_2 is computed separately to avoid double counting when m is even. The expected number of bipartite partitions of $G(m,p)$ is

$$N_{m,p}=\left\{\begin{array}{ll} S_1+S_2 & \text{if m is odd} \\ S_1+\frac{S_2}{2} & \text{if m is even}\end{array}\right\}.$$

For $G(n,p)$ the expected number of bipartite partitions if d vertices are deleted at random is

$$M_{n,p,d}=\binom{n}{d}N_{n-d,p}.$$

3 Estimation of Minimum Number of Vertex Deleted

Consider the minimum value of d for which $M_{n,p,d}=b\geq 1$. Suppose that $b\approx 1$. If d vertices are deleted from a graph $G(n,p)$, on an average the number of bipartite partitions will be b. Some of these graphs will actually have more than b bipartite partitions. Let there be a particular graph G_1, of the type $G(n,p)$, with $L\gg b$ bipartite partitions, when d vertices are deleted from it. The presence of such a graph would tend to drive the expected number of bipartite partitions to some value $b'>b$. To balance this effect there must be approximately $L-1$ graphs of the type $G(n,p)$, for which the number of bipartite partitions is zero, which may be associated with G_1 so that the average number of bipartite partitions is restored to

b. These graphs, which have no bipartite partition, will require more than b vertices to be deleted. On an average, therefore, at least b vertices will have to be deleted from each graph of the type $G(n,p)$. Thus d chosen according to the above criterion serves as a lower bound on the average number of vertices to be deleted to render a graph $G(n,p)$ bipartite. This is a pessimistic estimator, we call this the *lower bound estimate*.

It is desirable to have a value of d which will serve as a more pragmatic estimate of the expected number of vertices to be deleted to make a graph $G(n,p)$ bipartite. Such a pragmatic estimate has been designed empirically as follows: Let $B_1 = M_{n,p,d}$, for the minimum value of d such that $B_1 \geq 1$. Let $B_2 = M_{n,p,d}$, for the value of d which maximizes B_2. Let $B_P = \sqrt{B_1 B_2}$. Find the maximum value of $d = d_P$ for which $M_{n,p,d} \leq B_P$. This has been used as an empirical evaluation of $E(d)$. We call this the *pragmatic estimate*. It has been experimentally observed that this estimate closely matches the cost of the actual solutions returned by the GA.

It is evident that the computation of $M_{n,p,d}$ takes $O(n)$ time. For both the estimators this has to be computed for several values of d, which itself is bounded by n. Hence the time complexity for computing these estimators is $O(n^2)$.

4 Experimentation

The estimate has been tested for graphs of various sizes and for various edge probabilities. Vertex deletion being an NP-complete problem it has been necessary to use a non-exact algorithm for finding the actual result to compare with the estimate. A genetic algorithm for vertex deletion [4] referred to here as ND-GA has been used. This algorithm itself has been observed to yield excellent results, finding optimal results for small graphs for up to fifteen vertices, with test cases. Random graphs of the type $G_{n,p}$ have been used for the experimentation. The test results for graphs of 10, 20, 30, 40 and 60 vertices and edge probabilities 0.082, 0.107, 0.154, 0.250 0.400 and 0.500 have been presented in table 1. The first three probabilities have been chosen to reveal situations where the estimate performs most badly. In this table the first column shows the number of vertices $|V|$, in the graph and the second column the edge probability p. For each $< |V|, p >$ combination table thirty random graphs of the type $G(n,p)$ were generated. The third column is the lower bound (pessimistic) estimate while the fourth column is the proposed (pragmatic) estimate. The last column is the actual deletion by ND-GA. We observe that the proposed pragmatic estimate comes very close to the deletion by ND-GA.

5 Conclusion

An empirical method has been used, using the model of random graphs, to estimate the number of vertices that must be deleted to make a graph bipartite. Based on this method two estimators have been developed, a pessimistic and a pragmatic estimator. Both the estimators have been subject to considerable experimentation. While the number of vertices to be deleted suggested by the pessimistic estimator is usually less compared to the number of vertices required to be deleted, the estimates generated by the pragmatic estimator are very close to the acutal number of deletions. This suggests that the pragmatic estimator provides reliable estimates for the minimum number of vertices to deleted to render a graph bipartite. It may be noted that the acutal vertices to be deleted are determined by a genetic algorithm (ND-GA) [4] developed for the vertex deletion problem.

It would, therefore be advisable to use the pragmatic estimator in conjunction with ND-GA, in "choose and solve" environments, where it is desirable to choose a set of vertex deletion

no. of. vertices (n)	edge prob. (p)	l.b. est.	Proposed estimator	deletion ND-GA [a]
10	0.082	0	0	0.067
10	0.107	0	0	0.290
10	0.154	0	0	0.548
10	0.250	0	0	0.870
10	0.400	2	2	2.290
10	0.500	2	3	2.935
10	0.650	4	4	4.419
20	0.082	0	0	0.516
20	0.107	0	1	1.677
20	0.154	1	2	2.709
20	0.250	4	5	5.516
20	0.400	7	8	8.870
20	0.500	9	10	10.161
20	0.650	11	12	12.226
30	0.082	0	1	2.612
30	0.107	1	4	4.161
30	0.154	4	7	6.935
30	0.250	9	11	11.129
30	0.400	14	16	16.032
30	0.500	17	18	18.064
30	0.650	20	21	20.935
40	0.082	2	4	5.806
40	0.107	4	7	8.451
40	0.154	9	12	12.064
40	0.250	16	18	18.322
40	0.400	23	24	24.000
40	0.500	26	27	26.741
40	0.650	29	30	29.712
60	0.082	8	13	14.903
60	0.107	13	18	18.838
60	0.154	21	25	25.129
60	0.250	31	34	33.419
60	0.400	40	42	41.193
60	0.500	44	45	44.870
60	0.650	48	49	48.667

[a] The deletion in each line has been reported as the average obtained by running ND-GA on 30 individual random graphs of the type $G_{n,p}$.

Table 1: Comparison of the estimator against the number of vertices deleted by ND-GA.

problems from a possibly large set of problem instances and then solve them to determine the actual set of vertices to be deleted.

References

[1] M. Yannakakis, "The node deletion problem for heriditary properties," Tech. Rep. TR-240, Princeton University, N.J., Computer Science Lab., 1988.

[2] M. Yannakakis, "Node and edge deletion np-complete problems," in *Proceedings, 10th ACM Symp. on Theory of Computation*, pp. 253–264, 1978.

[3] M. Yannakakis, "Node-deletion problems on bipartite graphs," *SIAM Journal of Computing*, vol. 10, pp. 310–327, 1981.

[4] C. A. Mandal, P. P. Chakrabarti, and S. Ghose, "Port assignment for dual and triple port memories using a genetic approach," in *Proceedings of The Third Asia Pacific Conference on Hardware Description Languages (APCHDL) 96*, pp. 60–64, 1996.