Finding Induced Acyclic Subgraphs in Random Digraphs

C.R. Subramanian

The Institute of Mathematical Sciences
Taramani, Chennai - 600 113, India.
crs@imsc.res.in

Submitted: Oct 22, 2001; Accepted : Aug 18, 2003; Published: Dec 4, 2003
MR Subject Classifications : 05C80, 68W40

Abstract

Consider the problem of finding a large induced acyclic subgraph of a given simple digraph $D = (V, E)$. The decision version of this problem is NP-complete and its optimization is not likely to be approximable within a ratio of $O(n^{\epsilon})$ for some $\epsilon > 0$. We study this problem when D is a random instance. We show that, almost surely, any maximal solution is within an $o(\ln n)$ factor from the optimal one. In addition, except when D is very sparse (having $n^{1+o(1)}$ edges), this ratio is in fact $O(1)$. Thus, the optimal solution can be approximated in a much better way over random instances.

1 Introduction

Given a *simple* digraph $D = (V, E)$, we want to find a $V' \subseteq V$ of as large size ($|V'|$) as possible such that the induced subgraph $D[V']$ is acyclic. By "simple", we mean that there is at most one arc (directed edge) between any unordered pair of vertices. Self-loops are not allowed. Throughout, we mean vertex induced subgraphs whenever we use the term subgraphs. The decision version (Is $|V'| \geq k$?) is NP-complete [3]. The optimization version is not polynomial-time approximable even within a multiplicative ratio of $O(n^{\epsilon})$ for some $\epsilon > 0$ unless $P = NP$ [4].

We show that if D is a random digraph (obtained by randomly choosing the arcs), then, the size of any maximal solution is, almost surely, within a ratio of $o(\ln n)$ from the optimal solution. Except for very sparse random digraphs (obtained by choosing the arcs with probability $n^{-1+o(1)}$), the ratio is in fact $O(1)$. Thus, for random digraphs, one can obtain a significant improvement in the quality of approximation.

We assume that $V = \{1, \dots, n\}$ for the rest of the paper. Also, $p \leq 0.5$ is a positive real number. The random model is defined in the following way.

Model $D \in \mathcal{D}(n,p)$ **:** Choose each *undirected* edge $\{u,v\}$ joining elements of V independently with probability $2p$. For each chosen $\{u,v\}$, orient it in any of the two directions $\{u \to v, v \to u\}$ in D with equal probability (=0.5). The choices of arcs are independent for different chosen unordered pairs. This results in a simple digraph where each arc is chosen with probability p. When $p = 0.5$, there is an arc between every pair of vertices resulting in a random tournament.

In the subsequent sections, we analyze this model for finding large acyclic subgraphs and the size of an optimum solution.

2 Analysis of $\mathcal{D}(n,p)$

For an arbitrary H, let $mas(H)$ denote the size of a maximum acyclic subgraph of H. The following two theorems provide respectively an upper and a lower bound on the value of $mas(D)$ for $D \in \mathcal{D}(n,p)$.

Theorem 2.1 *Let $D \in \mathcal{D}(n,p)$, $p \leq 0.5$. Then, $mas(D) \leq \min\left(\lceil \frac{2\,(\ln n)}{-\ln(1-p)} \rceil + 1, n\right)$ almost surely.*

Proof: Without loss of generality, we can assume that $\lceil \frac{2\,(\ln n)}{-\ln(1-p)} \rceil + 1 < n$. Define

$$b = \lceil \frac{2\,(\ln n)}{-\ln(1-p)} \rceil + 2.$$

It is sufficient to prove that, almost surely, the subgraph induced by any A of size b is not acyclic. Fix such a subset A. Let $\sigma : \{1,\ldots,b\} \to A$ denote any linear ordering of the vertices of A. It is well-known that $D[A]$ is acyclic if and only if there exists an ordering σ such that each arc in $D[A]$ is of the form $\sigma(i) \to \sigma(j)$ for some $1 \leq i < j \leq b$. Hence, using Stirling's formula for factorials,

$$\begin{aligned} \mathbf{Pr}(\ D[A] \ \text{ is acyclic }\ ) &\leq b!\ (1-p)^{\binom{b}{2}} \\ &= \left(\frac{b}{e}\right)^b (1-p)^{\binom{b}{2}}\ \sqrt{2\pi b}\ [1+o(1)] \\ &= exp\left(b\left(\ln\left(\frac{b}{e}\right) + \frac{(b-1)\ln(1-p)}{2}\right)\right)\ \sqrt{2\pi b}\ [1+o(1)]. \end{aligned}$$

We have

$$\frac{(b-1)}{2} \geq \frac{(\ln n)}{-\ln(1-p)}\ +\ 0.5$$

and hence

$$\frac{(b-1)\ln(1-p)}{2} \leq -(\ln n) + 0.5\ \ln(1-p).$$

Employing these inequalities, we get

$$\mathbf{Pr}(\ D[A] \ \text{ is acyclic }\ ) \leq exp\left(b\left(\ln\left(\frac{b}{e}\right) - (\ln n) + 0.5\ \ln(1-p)\right)\right)\ \sqrt{2\pi b}\ [1+o(1)].$$

This bound holds for any fixed A of size b. There are

$$\binom{n}{b} \leq \left(\frac{en}{b}\right)^b = exp\left(b \ln\left(\frac{en}{b}\right)\right)$$

sets of size b. This gives us

$$\begin{aligned}
\mathbf{Pr}(\ \exists A,\ |A| = b,\ D[A] \text{ is acyclic }) &\leq \sqrt{2\pi n}\, exp\,(0.5\, b\, (\ln(1-p))) \\
&\leq \sqrt{2\pi n}\, exp\,(-(\ln n) + \ln(1-p)) \\
&= O\left(n^{-0.5}\right).
\end{aligned}$$

This establishes that, almost surely, $mas(D) \leq b$. ∎

Corollary 2.1 *If D is a random tournament ($p = 0.5$), then $mas(D) \leq \lceil \frac{2\,(\ln n)}{\ln 2} \rceil + 1 \leq 2.886\,(\ln n)$ almost surely.*

Let $\mathcal{G}(n,p)$ denote the standard random model for simple undirected graphs defined by including each of the possible edges independently with probability p. The independence number ($\alpha(G)$) (maximum size of an independent set in G) has been well-studied for this model and very surprising concentration results [1, 2] have been obtained for this number. The two quantities $mas(D)$ ($D \in \mathcal{D}(n,p)$) and $\alpha(G)$ ($G \in \mathcal{G}(n,p)$) are related as shown below.

Lemma 2.1 *For any positive integer b,*

$$\mathbf{Pr}(\ mas(D) \geq b\ ) \geq \mathbf{Pr}(\ \alpha(G) \geq b\ ). \tag{1}$$

Proof: Given a linear ordering σ of vertices of D and a subset A of size b, we say that $D[A]$ is consistent with σ if each arc in $D[A]$ is of the form $\sigma(i) \to \sigma(j)$ for some $i < j$.

Let τ denote an arbitrary but fixed ordering of V. Once we fix τ, the spanning subgraph of D formed by arcs of the form $\tau(j) \to \tau(i)$ ($j > i$) is having the same distribution as $\mathcal{G}(n,p)$. Hence, for any A, the event of $D[A]$ being consistent with τ is equivalent to the event of A being independent in $\mathcal{G}(n,p)$. Hence,

$$\begin{aligned}
\mathbf{Pr}(\ mas(D) \geq b\ ) &= \mathbf{Pr}(\ \exists A,\ |A| = b,\ D[A] \text{ is acyclic }) \\
&= \mathbf{Pr}(\ \exists A,\ |A| = b,\ \exists\sigma,\ D[A] \text{ is consistent with } \sigma\ ) \\
&= \mathbf{Pr}(\ \exists\sigma,\ \exists A,\ |A| = b,\ D[A] \text{ is consistent with } \sigma\ ) \\
&\geq \mathbf{Pr}(\ \exists A,\ |A| = b,\ D[A] \text{ is consistent with } \tau\ ) \\
&= \mathbf{Pr}(\ \alpha(G) \geq b\ ).
\end{aligned}$$

∎

Hence it is natural that we have a bigger upper bound for $mas(D)$ than we have for $\alpha(G)$. We use the inequality (1) to get a lower bound on $mas(D)$ when $p = o(1)$. Frieze [2] derives a lower bound on $\alpha(G)$ when $p = o(1)$. Applying this and using (1), we get

Theorem 2.2 *Write $w = np$. For each fixed $\epsilon > 0$, there exists a sufficiently large positive constant w_ϵ such that : If p is such that $w_\epsilon \le w = o(n)$, then, almost surely,*

$$mas(D) \ \ge \ \left(\frac{2}{p}\right)(\ln w - \ln\ln w - \ln 2 + 1 - 0.5\epsilon).$$

3 Finding an acyclic subgraph

We next analyze the following simple heuristic and estimate the size of its solution. The heuristic always outputs a maximal solution. In the description, we use D' both to denote a set of vertices and also to denote the subgraph induced by it.

MaximalAcyclic($D = (V, E)$**)**

0. If $V = \emptyset$, return V.
1. Pick an arbitrary $u \in V$.
2. Let $D' = D - (\{u\} \cup \{v \ : \ v \to u \in E\})$.
3. Apply **MaximalAcyclic(**D'**)** to get A'.
4. Return $A' \cup \{u\}$.

The following two facts will be used later. For $p = 0$, we take the limits.

$$\frac{p}{-\ln(1-p)} \le 1 \quad \text{and} \quad \frac{-\ln(1-p)}{p} \le 1.5 \quad \text{for } 0 \le p \le 0.5. \tag{2}$$

Theorem 3.1 *Write $p = w/n$. For each fixed positive $\epsilon < 1$, there exists a sufficiently large positive constant w_ϵ such that : If $D \in \mathcal{D}(n,p)$ with $p \le 0.5$ and $w \ge w_\epsilon$, then, with probability $1 - o(1)$,* **MaximalAcyclic(**D**)** *outputs a solution of size at least $\lfloor \frac{\epsilon\,(\ln w)}{-\ln(1-p)} \rfloor$.*

Proof: Define $L = \lfloor \frac{\epsilon\,(\ln w)}{-\ln(1-p)} \rfloor$. Using (2), it follows that $L < n/2$. We prove the theorem for sufficiently large values of n. Define $\alpha = 0.5$. For any $A \subseteq V$, let $X(A)$ be defined by

$$X(A) \doteq \{ \ v \in V - A \mid v \to u \notin E(D), \ \forall u \in A \ \}.$$

For $i \ge 1$, consider the event

$$\mathcal{E}_i : \ \exists A \subseteq V, \ |A| = i, \ |X(A)| \ < \ (1-\alpha)(n-i)(1-p)^i.$$

We will later prove that

$$\mathbf{Pr}(\mathcal{E}_i) \ = \ o(n^{-1}), \ \ \forall i : \ 1 \le i \le L. \tag{3}$$

Hence, with probability $1 - o(1)$, none of the events $\mathcal{E}_i$, $i \le L$, holds.

The function $f(i) = (1-\alpha)(n-i)(1-p)^i$ is a decreasing function of i. Hence, for all $i \le L$,

$$f(i) \ \ge \ f(L) \ \ge \ (1-\alpha)(n-L)e^{-\epsilon(\ln w)} \ \ge \ \frac{n}{4w^\epsilon}.$$

Given that (3) is true, for any $i \leq L$, the following holds with probability $1 - o(1)$: When we come to the recursive call of MaximalAcyclic with recursion depth $i+1$, we have already chosen i vertices to be part of the final solution and these i vertices induce an acyclic subgraph. The input graph for this recursive call has at least $f(i) \geq f(L) \geq n/4w^{\epsilon}$ vertices. Any of these can be added to the partial solution (of i vertices) already constructed without introducing any cycle. This guarantees that the final solution (built in this way) has at least L vertices in it with probability $1 - o(1)$.

To prove (3), consider any $i \leq L$ and any $A \subseteq V$ with $|A| = i$. For any $v \in V - A$,

$$\mathbf{Pr}(\ \forall u \in A : v \rightarrow u \notin D\ ) \ = \ (1-p)^i.$$

Since arcs going from $V - A$ to A are independent, by Chernoff-Hoeffding bounds (see Chapter 4 of [5]), we have

$$\mathbf{Pr}(\ |X(A)| < f(i)\ ) \ \leq \ e^{-\alpha^2 (n-i)(1-p)^i/2}.$$

Hence

$$\mathbf{Pr}(\ \mathcal{E}_i\ ) \ \leq \ \binom{n}{i} e^{-\alpha^2 (n-i)(1-p)^i/2}.$$

Let $g(i)$ denote the term coming to the right of '$\leq$'. $g(i)$ achieves its maximum (for our range of i) at $i = L$. This follows from (a) $L \leq n/2$ and hence $\binom{n}{i}$ is maximum at $i = L$, (b) $(n-i)(1-p)^i$ achieves its minimum at $i = L$. Hence, to prove (3), it is enough to show that $g(L) = o(n^{-1})$. This is true since

$$\begin{aligned} L &\geq \frac{1}{p}\left(\frac{\epsilon\,(\ln w)p}{-\ln(1-p)}\right) - 1 \quad \text{and} \\ g(L) &\leq \left(\frac{en}{L}\right)^L e^{-\alpha^2 (n-L)(1-p)^L/2} \\ &\leq w^L e^{-\alpha^2 n(1-p)^L/4} \quad \text{by using (2) and choosing sufficiently large } w_\epsilon \\ &\leq e^{L(\ln w) - \alpha^2 n/4w^\epsilon}. \end{aligned}$$

Since $\epsilon < 1$ is a constant, $w \geq w_\epsilon$ is sufficiently large and using (2),

$$\alpha^2 n/4w^\epsilon - L(\ln w) \ \geq \ \frac{n}{16w^\epsilon} - \frac{n\epsilon(\ln w)^2}{w}\left(\frac{p}{-\ln(1-p)}\right) \ = \ \Omega\left(\frac{n}{w^\epsilon}\right) \ = \ \Omega\left(n^{1-\epsilon}\right).$$

Hence (3) follows. ■

Note : (i) The proof actually shows that, almost surely, the size of *any* maximal acyclic subgraph is at least $\lfloor\frac{\epsilon\,(\ln w)}{-\ln(1-p)}\rfloor$. **(ii)** The failure probability of the algorithm is exponentially small, that is, $O(e^{-n^\gamma})$ for some constant $\gamma > 0$.

Combining Theorems 2.1 and 3.1, we get

Corollary 3.1 *Let $D \in \mathcal{D}(n,p)$ with $p = w(n)/n$ where $w(n) \to \infty$ is arbitrary. Then, given $D \in \mathcal{D}(n,p)$, with probability $1 - o(1)$,* **MaximalAcyclic(D)** *outputs a solution whose size is within a multiplicative ratio of $O((\ln n)/(\ln np)) = o(\ln n)$ from the optimal solution. For $p \geq n^{-1+a}$ for any fixed $a > 0$, the ratio is $O(1)$. For $p = 0.5$, the ratio is $\leq 2 + \delta$ for any fixed $\delta > 0$.*

4 Conclusions

The guarantee stated in Theorem 3.1 is not tight. For example, after building a solution of guaranteed size, the remaining graph still has at least $\frac{n}{4w^{\epsilon}}$ vertices in it. By working the analysis till D' becomes empty, one can possibly improve the guarantee to $\lfloor \frac{\ln w}{-\ln(1-p)} \rfloor$. But it seems unlikely that one can obtain a guarantee better than this. It would be of interest to design and analyze heuristics which actually find solutions better than the guarantee given in Theorem 3.1. As a beginning, the following problem could be looked at.

Problem : Design a polynomial time algorithm (deterministic or randomized) which, given $D \in \mathcal{D}(n,p)$, almost surely (over the input) finds an induced subgraph A of size at least $\lfloor \frac{(1+\epsilon)\ln w}{-\ln(1-p)} \rfloor$ for some positive constant ϵ.

However, it is possible that solving this problem might be difficult. A similar problem of designing a polynomial time algorithm for finding (given a $G \in \mathcal{G}(n, 1/2)$) an independent set of size at least $(1+\epsilon)$ times the size guaranteed by the greedy heuristic remains unsolved for more than two decades.

Similarly, the upper bound of Theorem 2.1 does not seem tight and there is scope for asymptotic improvement of that bound, particularly for very small p, that is, for $p = n^{-1+o(1)}$. We conjecture that even for such small p, almost surely, $mas(D) = \Theta(\frac{\ln w}{-\ln(1-p)})$. In fact, we believe in the following stronger statement.

Conjecture : There exist positive constants c_1 and c_2 such that : If $p = w/n$ with $c_1 n^{-1} \leq p \leq 0.5$ and $D \in \mathcal{D}(n,p)$, then, almost surely, $mas(D) = \frac{c_2(\ln w)}{-\ln(1-p)}[1 \pm o(1)]$. Here, $o(1)$ is with respect to n.

References

[1] B. Bollobás, "The chromatic number of random graphs", *Combinatorica* **8**, 49-55, 1988.

[2] A.M. Frieze, "On the independence number of random graphs", *Discrete Mathematics* **81**, 171-176, 1990.

[3] M.R. Garey and D.S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*, W.H. Freeman, San Francisco, 1978.

[4] C. Lund and M. Yannakakis, "The Approximation of Maximum Subgraph Problems", *Proceedings of the 20th International Colloquium on Automata, Languages and Programming (ICALP'93)*, LNCS 700, pp. 40–51.

[5] R. Motwani and P. Raghavan, *Randomized Algorithms*, Cambridge University Press, 1995.