Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO. AD-A109 035 / 3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Probabilistic Analysis of Random Extension-Rotation Algorithms	5. TYPE OF REPORT & PERIOD COVERED Technical Report
	6. PERFORMING ORG. REPORT NUMBER TR-28-81
7. AUTHOR(s) John H. Reif, Paul G. Spirakis	8. CONTRACT OR GRANT NUMBER(s) N00014-80-C-0674, NSF-MCS 77-22974
9. PERFORMING ORGANIZATION NAME AND ADDRESS Harvard University, Cambridge, MA 02138	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research, 800 North Quincy Street, Arlington, VA 22217	12. REPORT DATE October, 1981
	13. NUMBER OF PAGES 48
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) same as above	15. SECURITY CLASS. (of this report)
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

unlimited

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

probabilistic algorithm, random graph, probabilistic analysis, extension rotation algorithm, independence system, Hamilton path, perfect matching.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

see reverse

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601 |

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

403 832 slk

20.

ABSTRACT

We introduce a new random structure generalizing matroids. The *random independence systems* (RIS) allow us to develop general techniques for solving hard combinatorial optimization problems with random inputs. We describe a randomized algorithm for efficiently constructing an independent set of fixed size in an instance of a random independence system. We provide a general method of analysis of the performance of this algorithm, which allows us to derive bounds on the mean, variance and all the moments of the time complexity of the algorithm.

Center for Research in Computing Technology

Aiken Computation Laboratory

Harvard University

Cambridge, Massachusetts 02138

PROBABILISTIC ANALYSIS OF RANDOM EXTENSION-ROTATION ALGORITHMS

John H. Reif and Paul G. Spirakis

(14) TR-28-81

October 1981

PROBABILISTIC ANALYSIS OF RANDOM EXTENSION-ROTATION ALGORITHMS *

John H. Reif and Paul G. Spirakis

Aiken Computation Laboratory
Division of Applied Sciences
Harvard University, Cambridge Massachusetts

A previous draft of this paper, titled "Random Matroids," was presented at the 1980 ACM Symposium on Theory of Computing which was held in Los Angeles, California.

* This work was supported in part by the National Science Foundation Grant NSF-MCS79-21024 and the Office of Naval Research Contract N00014-80-C-0674.

ABSTRACT

We introduce a new random structure generalizing matroids. The *random independence systems* (RIS) allow us to develop general techniques for solving hard combinatorial optimization problems with random inputs. We describe a randomized algorithm for efficiently constructing an independent set of fixed size in an instance of a random independence system. We provide a general method of analysis of the performance of this algorithm, which allows us to derive bounds on the mean, variance and all the moments of the time complexity of the algorithm.

1. Introduction

In a classic paper "On the Abstract Properties of Linear Dependence" of 1935, Whitney provided a set of axioms for a structure called here a *Whitney matroid*. Matroid theory (see [Tutte, 1971], [Lawler, 1976]) has applications to a wide class of combinatorial optimization problems: where we wish to construct a maximal object (a maximum independent set) satisfying a monotone property.

We introduce in this paper (Section 2) the *random independence system* (RIS) which is applicable to a more general class of combinatorial optimization problems with *random inputs*. We define some natural notions, such as "maximal with a given probability." Properties of random independence systems such as the existence of an independent set of given cardinality (with probability 1), the relationship between RIS and Whitney matroids and properties of *intersections* of RIS are discussed in a companion paper [Reif, Spirakis, 81] (see also Section 2 of a previous draft of this paper, [Reif, Spirakis, 80]). In that paper we describe a *nonconstructive proof technique* for determining (with probability 1) the existence of an independent set of given cardinality or given weight in an instance of a random independence system.

In this paper, we develop a randomized algorithm, the extension-rotation (E-R) algorithm, for efficiently constructing an independent set of a given size h_0 in an instance of an RIS. Given an independent set I of size less than h_0, we attempt to *extend* I (by adding a new random element e to I) or else attempt to *rotate* I (by deleting an

element e' of I and adding the new element e). The use of a rotation operation first appeared in Posa's [1976] existence proof for a Hamiltonian path in an undirected random graph of density O(log(n)/n). [Karp, 1976] and [Angluin and Valiant, 1979] consider random algorithms with extensions and rotations.

The introduction of the rotation operation seems necessary for certain independence systems, since the greedy algorithm (which utilizes only extensions) may have arbitrarily bad performance (see [Korte, Hausmann, 78]). We show that the probability density of the number of rotation steps between successive extensions is upper and lower bounded by geometric density functions. From these bounds we derive sufficient conditions (a lower bound on the element density) for the E-R algorithm to succeed, with arbitrarily high probability. Also, we can derive bounds on the probability density function of the total number of steps, and from these density functions derive bounds on the mean, variance and all the moments of the time complexity of the algorithm. Thus we have a *general method for analysis* of the performance of the random extension-rotation algorithm. We view this as the most significant contribution of the paper.

We also give some applications to random graphs $G_{n,p}$ (see Section 2.3 and [Erdös and Spencer, 1974]).

P1 Construct a Hamiltonian path in $G_{n,p}$.

P1' For a graph H of fixed size, construct a subgraph of $G_{n,p}$ homeomorphic to H.

P2 Construct a perfect matching in $G_{n,p}$.

P2' Construct a perfect matching in a random bipartite graph $B_{n,p}$.

(Note that P1' is a generalization of P1.)

The randomized E-R algorithm is applicable to P1, P2 and P2' (and we have an efficient transformation from instances of P1' to instances of P1).

The results of our general method for analysis of the extension-rotation algorithm yield lower bounds for the edge density p to give probability of success $1 - n^{-\alpha}$ for $\alpha > 1$. Previously [Erdös and Rényi, 1959] have considered P1 and [Posa, 1976] considers P1 for undirected graphs. [Angluin and Valiant, 1979] consider P1 and P2 for directed graphs. They derive similar results for a different random graph model $G_{n,N}$ and their results hold for $G_{n,p}$ only under certain conditions as $n \to \infty$.

Our general method also yields significant *new* results for these applications, such as *tight* bounds (within a constant multiple) on the mean and variance of the randomized algorithm's time complexity.

2. Definitions of Random Independence Systems and Their Structure

2.1 Definitions of Random Independence Systems

Let E be a set and let $\mathcal{J}$ be a family of subsets of E. Let p be a real number (the *element's density*) on the interval $[0,1]$. The triple $M = (E, \mathcal{J}, p)$ is a (uniform) *random independence system* (RIS). (Nonuniform random independence systems and weighted random independence systems are defined in [Reif, Spirakis, 1981]). We will frequently write $(E, \mathcal{J}, 1)$ as $(E, \mathcal{J})$. $M = (E, \mathcal{J}, p)$ is a *proper* random independence system if

A1 $\phi \in \mathcal{J}$.

A2 $A \in \mathcal{J} \wedge A' \subseteq A \Rightarrow A' \in \mathcal{J}$.

Intuitively, $\mathcal{J}$ may be considered a property on subsets of E which is *trivially satisfied* (by axiom A1) and *monotone decreasing* (by axiom A2). Let $(E,\mathcal{J})$ be a *Whitney matroid* (a matroid as defined by [Whitney, 1932]) if it satisfies A1, A2 and the additional axiom

A3 For any sets $A,A' \in \mathcal{J}$ of cardinality $h,h+1$ respectively, $\exists e \in A' - A$ such that $A \cup \{e\} \in \mathcal{J}$.

2.2 Instances of Random Independence Systems

An *instance* of a random independence system $M = (E,\mathcal{J},p)$ is a pair $M_0 = (E_0,\mathcal{J}_0)$ where

(i) $E_0 \subseteq E$ is derived by independently choosing each $e \in E$ with probability p.

(ii) $\mathcal{J}_0 = \{I \in \mathcal{J} / I \subseteq E_0\}$.

Note that the *probability* of M_0 is $p^{|E_0|}(1-p)^{|E-E_0|}$. Clearly, any instance $M_0 = (E_0,\mathcal{J}_0)$ of a proper RIS satisfies axioms A1 and A2. (Hence, any instance of a proper RIS is an *independence system*, as defined in [Korte and Hausmann, 1978].)

A set $A \subseteq E_0$ is *independent* in M_0 if $A \subseteq \mathcal{J}_0$ and *dependent* otherwise. An independent set $I \in \mathcal{J}_0$ is *maximum in* M_0 if there does not exist an $I' \in \mathcal{J}_0$ such that $|I'| > |I|$. Let the *rank* of M_0 be the cardinality of a maximum independent set. $I \in \mathcal{J}_0$ is *maximal in* M_0 if there does not exist an $I' \in \mathcal{J}_0$ such that $I' \supset I$. A *minimal dependent set of* M_0 (a *circuit*) has no proper subset which is dependent in M_0. For any $A \subseteq E_0$ let the *rank of* A *in* M_0 be the maximum cardinality of any independent subset of A. It follows from a result of [Korte, Hausmann, 1978] that for any instance M_0 of a proper RIS there exists an integer k and k matroids of which the instance M_0 is an intersection.

2.3 Examples of Random Independence Systems

As an example of an RIS, let Q be a property on graphs and let $G_{n,p}$ be a random undirected graph. (Examples can be given for directed graphs also. For sake of simplicity here we give only examples for random undirected graphs.) The random graph $G_{n,p}$ is a random variable whose instances are graphs with vertices $V = \{1,2,\ldots,n\}$ and each edge chosen independently with probability p from the set $E = \{\{u,v\}/u,v$ are distinct vertices of $V\}$. Let $M = (E,\mathcal{J},p)$ be the (uniform) RIS with E as above and $\mathcal{J} = \{I \subseteq E/Q(V,I)$ holds$\}$. Then any instance $M_0 = (E_0,\mathcal{J}_0)$ of M corresponds to an instance (V,E_0) of the random graph $G_{n,p}$ and $\mathcal{J}_0$ contains precisely those edge sets $I \subseteq E_0$ such that the property Q holds for subgraph (V,I). Note that M is a proper RIS if the graph property Q is

(1) *trivially satisfied*, i.e., Q holds for the graph with no edges and

(2) *decreasing monotone*: $Q(G) \Rightarrow Q(G')$ for all subgraphs G' of G.

We list some graph properties and the corresponding RIS below.

P1 (Hamiltonian paths)

Given a graph $G = (V,E)$, a *simple path* is a path of edges in E containing no cycles, and it is a *Hamiltonian path* if it contains every vertex of V. The property of a "simple path" in a random graph does not yield a proper RIS, since a simple path must be connected (violating axiom A2). However, we can define a proper RIS such that any independent set of cardinality $|V| - 1$ is a Hamiltonian path. We give both formulations here:

Formulation as a non-proper RIS: Let $M = (E, \mathcal{J}, p)$ be the RIS where $\mathcal{J}$ is the set of all simple paths in the complete graph (V,E). Fix an instance $M_0 = (E_0, \mathcal{J}_0)$ of M. Then (V,E_0) has the same probability in random graph $G_{n,p}$ as in M and $\mathcal{J}_0$ is the set of all simple paths in (V,E_0).

Formulation as a proper RIS: Let $M = (E, \mathcal{J}, p)$ be the RIS with E as above and $\mathcal{J} = \{I \subseteq E / (V,I)$ consists of a set of disjoint simple paths$\}$.

Clearly M satisfies axioms A1, A2. Fix an instance $M_0 = (E_0, \mathcal{J}_0)$ of M. Then (V,E_0) has the same probability in $G_{n,p}$ as in M and $\mathcal{J}_0$ has as elements all different sets of disjoint simple paths in E_0.

In both formulations, if M_0 has an independent set $I \in \mathcal{J}_0$ such that $|I| = n-1$ then (V,I) is a Hamiltonian line in (V,E).

P2 Perfect matchings

An edge *matching* of a graph is a set of vertex disjoint edges, and is *perfect* if every vertex appears in some edge of the matching. To formulate the "perfect matching" problem as an RIS, we assume a complete graph $G = (V,E)$ with $2n$ vertices.

$M = (E, \mathcal{J}, p)$ where $\mathcal{J} = \{I \subseteq E / I$ is a matching$\}$.

Let $M_0 = (E_0, \mathcal{J}_0)$ be an instance of M. Then M_0 has a perfect matching if there is an $I \in \mathcal{J}_0$ such that $|I| = n$. The property of "matching" in a random graph $G_{2n,p}$ yields a proper RIS, since if I is a matching then every $I' \subseteq I$ is a matching.

P2' Bipartite matching

In the following let $V_1 = \{1,\ldots,n\}$, $V_2 = \{n+1,\ldots,2n\}$ be disjoint vertex sets of equal cardinality, and let $E = \{\{u,v\} / u \in V_1, v \in V_2\}$.

A *bipartite graph* $B = (V_1 \cup V_2, E_0)$ has vertex set $V_1 \cup V_2$ and edge set $E_0 \subseteq E$. B is *complete* if $E_0 = E$. A *random bipartite graph* $B_{n,p}$ has instances which are bipartite graphs $(V_1 \cup V_2, E_0)$ where each edge of E_0 is chosen from E with probability p.

An *(edge) matching* of bipartite graph $(V_1 \cup V_2, E_0)$ is a set of vertex disjoint edges $I \subseteq E_0$ and is *perfect* if every vertex of $V_1 \cup V_2$ appears in some edge of I. The *bipartite perfect matching problem* is formulated as a *proper* RIS by assuming a complete bipartite graph $B = (V_1 \cup V_2, E)$. Let $M = (E, \mathcal{J}, p)$ where $\mathcal{J} = \{I \subseteq E / I$ is a (bipartite) matching$\}$. Let M_0 be an instance of M. M_0 has a perfect matching if there is an $I \in \mathcal{J}_0$ such that $|I| = n$.

3. The E-R Algorithm for Constructing Independent Sets

In this section we describe an efficient algorithm for constructing an independent set of fixed size from an instance of a random independence system. This E-R algorithm is a generalization of random graph algorithms which have appeared in [Posa, 1976], [Karp, 1976], and [Angluin and Valiant, 1979]. In Section 5 we develop a *general method of analysis* of the E-R algorithm which provides:

(i) Sufficient conditions for successful termination with probability $1 - |E|^{-\alpha_0}$ for any fixed sufficiently large $\alpha_0 > 1$.

(ii) Upper and lower bounds on the probability density of the time cost of the E-R algorithm, from which the mean, variance and all the moments of the time cost may be derived.

Section 4 gives a simplified discussion of that analysis, which is intended to aid the reader's intuition and lead to the more thorough analysis of Section 5.

Let $M_0 = (E_0, \mathcal{J}_0)$ be an instance of the random independence system $M = (E, \mathcal{J}, p)$. We wish to construct an independent set of size $h_0 > 0$.

For any independent set $I \in \mathcal{J}_0$, let $\mathcal{E}(I) = \{e \in E_0 \mid I \cup \{e\} \in \mathcal{J}_0\}$. Note that if $\mathcal{E}(I) \neq \emptyset$ then we may *extend* I by choosing an $e \in \mathcal{E}(I)$ and substituting $I \cup \{e\}$ for I.

Also, for any independent set $I \in \mathcal{J}_0$, let $\mathcal{R}(I) = \{e \in E_0 \mid I \cup \{e\} \notin \mathcal{J}_0$ but $\exists e' \in I$ with $I \cup \{e\} - \{e'\} \in \mathcal{J}_0\}$. If $\mathcal{R}(I) \neq \emptyset$, we may *rotate* I by choosing an $e \in \mathcal{R}(I)$ and some appropriate $e' \in I$ and substituting $I \cup \{e\} - \{e'\} \in \mathcal{J}_0$ for I.

Actually, in the algorithm below, we choose a *random* element $e \in \mathcal{E}(I) \cup \mathcal{R}(I)$ and first attempt to extend I by e, and else rotate I by e. We call $\mathcal{E}(I)$ the *extension set* of I and $\mathcal{R}(I)$ the *rotation set* of I.

3.1 The E-R Algorithm

INPUT: An instance $M_0 = (E_0, \mathcal{J}_0)$ of a random independence system $M = (E, \mathcal{J}, p)$ and integer $h_0 \geq 0$.

```
INITIALIZATION: I ← ∅;  T ← 0
WHILE |I| < h_0 DO
   BEGIN
      IF E_T(I) ∪ R_T(I) = ∅ THEN FAIL
      choose some random e ∈ E_T(I) ∪ R_T(I)
      IF e ∈ E_T(I) THEN EXTEND:  I ← I ∪ {e}
```

```
            ELSE BEGIN
                 choose  e' ∈ I  with  (I ∪ {e}) - {e'} ∈ 𝒥_0
                 ROTATE: I ← I ∪ {e} - {e'}
                 END
        T ← T + 1
        E_T ← E_{T-1} - {e}
    END
RETURN (I).
```

We define the sets:

$$\mathscr{E}_T(I) = \{e \in E_T \mid I \cup \{e\} \in \mathscr{J}_0\},$$

$$\mathscr{R}_T(I) = \{e \in E_T \mid I \cup \{e\} \notin \mathscr{J}_0 \text{ , but } \exists e' \in I \text{ with } I \cup \{e\} - \{e'\} \in \mathscr{J}_0\}$$

as "macros" which are expanded in-line within the E-R algorithm.

For the problem of perfect matchings in random graphs $G_{n,p}$ the extension and rotation sets are defined as follows: Let $M_0 = (E_0, \mathscr{J}_0)$ be an instance of the matching RIS and $I \in \mathscr{J}_0$. Then

$$\mathscr{E}(I) = \{e \in E - I \mid \text{the vertices of } e \text{ are distributed from the vertices of } I\}$$

and

$$\mathscr{R}(I) = \{e \in E - I \mid \text{one vertex of } e \text{ is an element of } E - I\}.$$

For the bipartite perfect matching problem in bipartite random graphs (V_1, V_2, p) with $|V_1| = |V_2| = n$, the extension and rotation sets are defined as follows: Let $M_0 = (E_0, \mathscr{J}_0)$ be an instance of the bipartite matching RIS and let $I \in \mathscr{J}_0$. Let $V_i(I)$ = set of vertices in V_i which are incident to edges in I, for $i = 1,2$. Fix a $u \in V_1 - V_1(I)$. Then

$$\mathscr{E}_T(I) = \{\{u,v\} \in E_0 / v \notin V_2(I)\}$$

$$\mathscr{R}_T(I) = \{\{u,v\} \in E_0 / v \in V_2(I)\}.$$

In case of an edge e selected from $\mathscr{R}_T(I)$, the rotation is done as follows: Let $e' = \{u',v\}$ be the (unique) edge of I such that e', e

(in the E-R algorithm) have v as a common vertex in V_2. Delete e' from I, add e to I and then set u to u'.

For the Hamiltonian line problem in random graphs we have

(1) For the formulation as a non-proper RIS:

Let $I \in \mathcal{J}_0$ be a non-maximal simple path. We let $V(I)$ be the vertices of I and let $ENDS(I)$ be the vertices of I of valence < 2. Then the extension set is $\mathcal{E}(I) = \{e \in E_0 - I \mid e = \{u,v\}, \ u \in ENDS(I), \ v \in V - V(I)\}$. The rotation set is $\mathcal{R}(I) = \{e \in E - I - \mathcal{E}(I) \mid e = \{u,v\}, \ u \in ENDS(I), \ v \in V(I) - ENDS(I)\}$.

(2) For the formulation as a proper RIS:

Let $I \in \mathcal{J}_0$ be a set of disjoint simple paths which is not maximum. Let $V(I)$ and $ENDS(I)$ be as in (1). Then the extension set is

$$\mathcal{E}(I) = \{e \in E_0 - I \mid e = \{u,v\}, \ u \in ENDS(I), \ v \in V - V(I)\}$$
$$\cup \{e \in E_0 - I \mid e = \{u,v\}, \ u \in ENDS(I), \ v \in ENDS(I) \text{ and } u,v \text{ are in different paths of } I\}.$$

The rotation set is

$$\mathcal{R}(I) = \{e \in E - I - \mathcal{E}(I) \mid e = \{u,v\} \text{ and } (u \in ENDS(I), \ v \in V(I) - ENDS(I)) \text{ or } (u,v \in ENDS(I') \text{ for some path } I' \subseteq I)\}.$$

[Korte, Hausmann, 1978] proved that the greedy algorithm performs as follows in *any* independence system $M = (E, \mathcal{J})$. Let I_g be the output of the greedy and I_{max} the maximum (in cardinality) independent set of M. If M can be written as an intersection of k matroids, then $|I_g| \geq |I_{max}|/k$. For the matching problem, $k = 2$. For the (proper) RIS formulation of the Hamiltonian line problem, $k = 3$. Note that the E-R algorithm has at least as good performance as the greedy algorithm.

As we show in the analysis, if the probability p of the RIS is bigger than a certain value, then rotation succeeds with probability one in finding short augmentation sequences in a random instance of the RIS. Even as a heuristic, E-R constructs bigger maximal independent sets than greedy and has the same worst-case time complexity if a rotation element can be always found in fixed time.

3.2 Parameters of the E-R Algorithm

We wish to analyze the E-R algorithm relative to the "time" index T, which is incremented on each iteration of the algorithm. Note that each "unit time" step from T to $T+1$ may include

(i) a constant number of arithmetic and set operations

(ii) an emptiness test for $\mathscr{E}_T(I) \cup \mathscr{R}_T(I)$

(iii) choice of a random element of $\mathscr{E}_T(I)$

(iv) choice of a "rotation" element $e' \in I$ such that if $e \in \mathscr{R}_T(I)$ then $I \cup \{e\} - \{e'\} \in \mathscr{J}_0$.

(Of course in the applications of Section 5 on a particular machine model such as a RAM, we must determine bounds on the number of machine instructions per "unit time steps" of the algorithm.)

Let H be the size of the independent set I on exit (either by successful termination or by failure). For each $h = 1,2,\ldots,H$ let T_h be the value of T just after I is extended from size $h-1$ to size h. Also, let $T_0 = 0$ and let $T_h = |E_0|$ for $h = H+1,\ldots,h_0$. Note that H and the T_h are random variables which are fixed only for a given execution of the algorithm E-R on a given instance M_0 of the RIS M.

Fix some constant $\alpha > 1$. For each $t = 0,1,\ldots,|E|$ let $\varepsilon_t(h)$, $\hat{\varepsilon}_t(h)$, $\lambda_t(h)$, $\hat{\lambda}_t(h)$ be functions of domain $0 \leq h \leq h_0$ and range $[0,1]$.

We require that for a class $\mathcal{A}_0$ of executions of the Algorithm E-R with total probability $\geq 1 - |E|^{-\alpha}$,

(i) $\varepsilon_t(|I|) \leq \Pr\{$extension of I on step t $|\ \mathcal{E}_t(I) \cup \mathcal{R}_t(I) \neq \emptyset$ and given an execution in $\mathcal{A}_0\}$ $\leq \hat{\varepsilon}_t(|I|)$.

(ii) $\lambda_t(|I|) \leq \Pr\{\mathcal{E}_t(I) \cup \mathcal{R}_t(I) = \emptyset |$ given an execution in $\mathcal{A}_0\}$ $\leq \hat{\lambda}_t(|I|)$.

Also let $\rho_t(h) = (1 - \hat{\lambda}_t(h)) \cdot (1 - \hat{\varepsilon}_t(h))$

and $\hat{\rho}_t(h) = (1 - \lambda_t(h)) \cdot (1 - \varepsilon_t(h))$.

Note that $\rho_t(h)$, $\hat{\rho}_t(h)$ are functions such that except for executions of the E-R algorithm with total measure $\leq |E|^{-\alpha}$,

$$\rho_t(|I|) \leq \Pr\{\text{rotation of } I \text{ on step } t\} \leq \hat{\rho}_t(|I|).$$

The above (somewhat informal) statements can be related t the random variable T_h where $h = |I|$ by:

"extension of I on step t" $\Longleftrightarrow$ "$T_{h+1} = t + 1$"

"rotation of I on step t" $\Longleftrightarrow$ "$T_{h+1} > t + 1$"

"$\mathcal{E}_t(I) \cup \mathcal{R}_t(I) = \emptyset$" $\Longleftrightarrow$ "$T_h = |E_0|$."

Note that the functions $\varepsilon_t(h)$, $\hat{\varepsilon}_t(h)$, $\lambda_t(h)$, $\hat{\lambda}_t(h)$ can always be trivially defined:

$$\varepsilon_t(h) = \lambda_t(h) = 0, \qquad \hat{\varepsilon}_t(h) = \hat{\lambda}_t(h) = 1$$

so they satisfy the above restrictions. In practice, of course, we wish

$$|\hat{\varepsilon}_t(h) - \varepsilon_t(h)| \qquad \text{and} \qquad |\hat{\lambda}_t(h) - \lambda_t(h)|$$

to be minimal, so that the analysis techniques of Section 5 yield tight bounds on the time complexity of the E-R algorithm. In our graph

applications tight $\varepsilon_t(h)$, $\hat{\varepsilon}_t(h)$, $\lambda_t(h)$, $\hat{\lambda}_t(h)$ are obtained in Sections 6 and 7 for matchings and Hamiltonian line problems.

4. A Simplified Probabilistic Analysis of the E-R Algorithm

We describe here a very simplified probabilistic analysis of the E-R algorithm. A much more accurate analysis follows in the next section.

The extension probability is defined as the conditional

$$\text{Prob}\{\text{a random } e, \text{ chosen from } \mathscr{E}_T(I) \cup \mathscr{R}_T(I), \text{ belongs to } \mathscr{E}_T(I)\}$$

and is equal to the ratio $\varepsilon(T,I) = x/(x+y)$, given that $\mathscr{E}_T(I) = x$, $\mathscr{R}_T(I) = y$.

The definition above, suggests that if there are numbers x_{min}, x_{max}, y_{min}, y_{max} (generally depending on $|I|$ such that for some $\alpha > 1$

$$\text{Prob}\{x_{min} \leq |\mathscr{E}_T(I)| \leq x_{max} \quad \text{and} \quad y_{min} \leq |\mathscr{R}_T(I) \leq y_{max}\}$$

is

$$\geq 1 - |E|^{-\alpha} \qquad (*)$$

then

$$\varepsilon_T(h) = \frac{x_{min}}{x_{max} + y_{max}} \leq \varepsilon(T,I) \leq \frac{x_{max}}{x_{min} + y_{min}} = \hat{\varepsilon}_T(h)$$

with probability $\geq 1 - |E|^{-\alpha}$, and we can use these bounds to analyze the E-R algorithm.

The existence of nontrivial x_{min}, x_{max}, y_{min}, y_{max} depends on both the instance of the random independence system given as input to

the algorithm and on the particular random execution of the E-R algorithm on that instance. Hence, $1 - |E|^{-\alpha}$ is the *total* probability of a class of "good" executions on a class of "good" input instances.

Let h be the cardinality of I and N be the biggest $|I|$ for any such set in $\mathcal{J}$. Suppose we could show that property (*) is satisfied with such numbers so that both $x_{min}/(x_{max} + y_{max})$ and $x_{max}/(x_{min} + y_{min})$ are approximately equal to $1 - h/N$. Then the behavior of the E-R algorithm would be modelled by the Markov process of Figure 1, where the numbers in the circles are the possible $|I|$. Thus, we would have transition probabilities

$$\text{Prob}\{|I| = h+1 \text{ at } T+1 / |I| = h \text{ at } T\} = 1 - \frac{h}{N} .$$

(Note that, with the above assumption, this extension probability does not depend on the time T).

Let $p(T,h)$ be the Prob{algorithm E-R achieves an independent set I of size h at time T}. We get by inspection

$$p(T,h) = p(T-1,h-1)\left(1 - \frac{h-1}{N}\right) + p(T-1,h) \cdot \frac{h}{N}$$

and

$$p(0,0) = 1$$

The solution of the above recursion would give the joint probability density of T and h and, consequently, we could easily derive the mean $\bar{T}$ for $h = N$ by

$$\bar{T} = \sum_{T=0}^{|E|} p(T,N) \cdot T \quad .$$

Let $\bar{u}_h$ = mean time the algorithm stays at size h, before extending. By known properties of Markov processes, we have

$$\bar{u}_h \leq \frac{1}{1-h/N} = \frac{N}{N-h} \quad .$$

Thus the mean time of execution of E-R before success is bounded by

$$\bar{T} = \bar{u}_0 + \bar{u}_1 + \cdots + \bar{u}_{N-1} + 1$$

$$\leq N\left(1 + \frac{1}{2} + \cdots + \frac{1}{N}\right) + 1 = O(N \log N) \quad .$$

Note that in most of the applications, $N = |E|^{\beta}$ with $0 < \beta \leq 1$.

The above $\bar{T}$ was produced by the assumption of a "good" class of inputs and executions. In a bad case, the algorithm will fail or stop after time at most $|E|$, hence

$$T_{total} \leq \bar{T}(1 - |E|^{-\alpha}) + |E| \cdot |E|^{-\alpha}$$

and since $\alpha > 1$ we get as $|E| \to \infty$ that

$$\bar{T}_{total} \leq O(N \log N) \quad .$$

This is the phenomenon *approximately* followed in the E-R algorithm. However, in general the extension probabilities depend also on time (the next section takes this dependence into account).

In applications in random graphs $G_{n,p}$ (where usually I is a set of edges) we note that T is equal to the number of edges examined by T, and h is equal to the number of edges successfully extending I by T. Hence, the number of deleted edges by T is T - h and this has to be less than or equal to the number ξ of edges from each vertex of I to all other vertices of I (since, as we shall prove for graph applications, we only delete edges whose vertices stay in I). The average ξ is $ph(h+1)$ and the average T - h is $\leq \bar{T}_{total} - h$. By the above, in order for the algorithm to achieve the maximum size N of h,

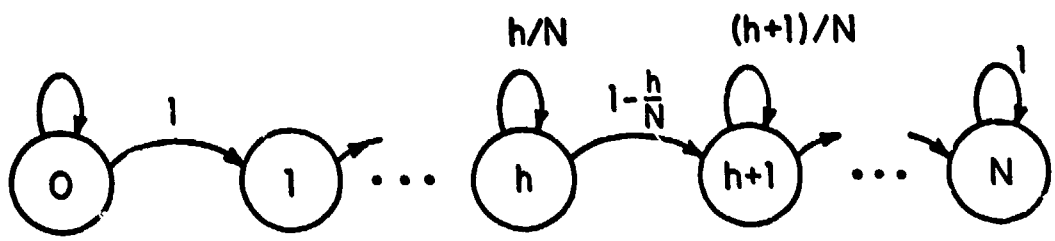

Figure 1

we have

$$\bar{T}_{total} - N \leq pN(N+1)$$

or edge probability

$$p \geq O\left(\frac{\log N}{N}\right) \quad .$$

So, we see that an edge probability of at least $O(\log N/N)$ is necessary for the E-R algorithm to work in graph problems. The constants of multiplication for particular cases follow from the exact analysis given in Section 5 and also in Sections 6 and 7.

5. Rigorous Probabilistic Analysis of the E-R Algorithm

We fix an RIS $M = (E, \mathcal{J}, p)$ throughout this section, and consider a random instance $(E_0, \mathcal{J}_0)$ given as input to the E-R algorithm. All our applications of Sections 6 and 7 satisfy the following *monotonicity restrictions*:

R1 $\varepsilon_t(h)$, $\hat{\varepsilon}_t(h)$ are monotonically decreasing with h but increasing with t.

R2 $\lambda_t(h)$, $\hat{\lambda}_t(h)$ are monotonically increasing with h and t.

Intuitively, assume that the *conditional probability of extension* decreases with $h = |I|$ and that the *probability of failure* increases as I grows and as the elements of E_0 are exhausted.

5.1 Sufficient Conditions for Success with High Likelihood

Note that if Q is predicate and A an event on which Q is predicated, we let $\text{Prob}\{Q/A\}$ to denote the conditional probability of Q given that A holds.

Our goal here is to derive sufficient conditions such that for any fixed sufficiently large $\alpha_0 > 1$,

$$\Pr\{H = h_0\} \geq 1 - |E|^{-\alpha_0}$$

(i.e., the E-R algorithm succeeds in constructing an independent set of size h_0 with probability $\geq 1 - |E|^{-\alpha_0}$).

Assuming the above restrictions R1, R2, we can derive bounds for

$$EXT_h = \Pr\{H > h \mid H \geq h,\ t = T_h,\ t' = T_{h+1} - 1 \text{ and given an execution in } \mathcal{A}_0\}$$

PROPOSITION 5.1.

$$\varepsilon_t(h) \cdot (1 - \hat{\lambda}_{t'}(h)) \cdot \left[\frac{1 - \rho_{t'}(h)^{|E_0|-t+1}}{1 - \rho_{t'}(h)}\right] \leq EXT_h$$

$$\leq \hat{\varepsilon}_{t'}(h) \cdot (1 - \lambda_t(h)) \cdot \left[\frac{1 - \hat{\rho}_t(h)^{|E_0|-t+1}}{1 - \hat{\rho}_t(h)}\right] .$$

Unfortunately, we found that a direct derivation of $\Pr\{H = h_0\}$ by use of Proposition 5.1 is intractable, because of the stubborn appearance of the random variables T_h in the conditional probabilities. (Thus Proposition 5.1, as stated, is never used in our analysis of the E-R algorithm.)

To bound the random variable E_0, we may use the following known fact:

LEMMA 5.1. *If* M *is an* RIS $(E, \mathcal{J}, p)$ *and* $(E_0, \mathcal{J}_0)$ *is a random instance of* M, *then*

$$\text{Prob}\{p|E|(1-\beta) \leq |E_0| \leq p|E|(1+\beta)\} \geq 1 - |E|^{-\alpha}$$

where

$$\beta = \sqrt{\frac{6\alpha \log |E|}{p|E|}} \quad .$$

Proof. Recall that the elements of E_0 are chosen from E with fixed probability p. Then this Lemma follows from the Chernoff bounds:

$$\sum_{k=\lceil (1+\beta)|E|p \rceil}^{|E|} \binom{|E|}{k} p^k (1-p)^{E-k} \leq \exp(-\beta^2 |E| p/3)$$

$$\sum_{k=0}^{\lfloor (1-\beta)|E|p \rfloor} \binom{|E|}{k} p^k (1-p)^{|E|-k} \leq \exp(-\beta^2 |E| p/2)$$

The following two conditions in conjunction imply $\text{Prob}\{H = h_0\} \geq 1 - (1 + c_0)|E|^{-\alpha}$

C1 For some fixed $t_0, t_1, \ldots, t_{h_0}$
$\lambda_t(h) = \hat{\lambda}_t(h) = 0$ for $0 \leq h \leq h_0$ and $0 \leq t \leq t_h$,

C2 $\text{Prob}\{T_{h_0} \leq t_{h_0} < |E_0|\} \geq 1 - c_0|E|^{-\alpha}$, for some $c_0 > 1$.

Note that C1 does not suffice to imply anything about $\Pr\{H = h_0\}$ since we may frequently fail if the time t exceeds t_h.

5.2 Verification of Condition C2

We now *assume that conditions* R1, R2 *and* C1 *have been verified for some* $t_0, t_1, \ldots, t_h$ and derive bounds on the critical p which insures condition C2 is satisfied.

To verify C2, we require upper and lower bounds on the distribution of steps between extensions. Let $g(x,q) = q(1-q)^x$ be the *geometric density function*. Let $\mathcal{N}_0$ be the class of executions of algorithm E-R with probability $1 - |E|^{-\alpha}$, which were used in the definition of the $\varepsilon_t(h)$.

Also, let S be the condition

"$T_{h+1} \leq t_h$, $t = T_h < |E_0|$ and given an execution in $\mathcal{N}_0$."

LEMMA 5.2.

$$\frac{\varepsilon_{t+x}(h)}{\hat{\varepsilon}_{t_h}(h)}\, g\left(x,\hat{\varepsilon}_{t_h}(h)\right) \leq \text{Prob}\{T_{h+1} - T_h = x+1 \mid S\} \leq \frac{\hat{\varepsilon}_{t+x}(h)}{\varepsilon_t(h)} \cdot g(x,\varepsilon_t(h))$$

Proof. By conditions C1 and monotonicity restriction R1,

$$\hat{\rho}_t(h) = (1-\varepsilon_t(h)) \leq 1 - \varepsilon_{T_h}(h)$$

for $0 \leq h \leq h_0$ and $T_h \leq t \leq t_n$.

$$\Pr\{T_{h+1} - T_h = x+1 \mid S\} \leq \hat{\varepsilon}_{t+x}(h) \prod_{k=t}^{t+x-1} \rho_k(h)$$

$$\leq \hat{\varepsilon}_{t+x}(h)\,(1-\varepsilon_t(h))^x$$

$$\leq \left[\frac{\hat{\varepsilon}_{t+x}(h)}{\varepsilon_t(h)}\right] \varepsilon_t(h)\,(1-\varepsilon_t(h))^x \quad .$$

The lower bound derivation is similar. □

We now derive bounds on the steps between extensions. For $h = 0,\ldots,h_0$ and $t = 0,\ldots,t_h$ let $\delta_t(h) = MAX(h,h')$ where

$$h' = \left\lfloor \log\left[\left(1-\varepsilon_{t_h}(h)\right)^{t_h+1} + \frac{\varepsilon_{t_h}(h)\left(1-|E|^{-\alpha}\right)}{\hat{\varepsilon}_t(h)}\right] \Big/ \log\left(1-\varepsilon_{t_h}(h)\right)\right\rfloor - 1$$

and let

$$\hat{\delta}_t(h) = \log\left[1 - \frac{\hat{\varepsilon}_t(h)\left(1-|E|^{-\alpha}\right)}{\varepsilon_t(h)}\right] \Big/ \log\left(1-\hat{\varepsilon}_{t_h}(h)\right) \quad .$$

LEMMA 5.3.

$$\Pr\{\delta_t(h) \leq T_{h+1} - T_h \leq \hat{\delta}_t(h) \mid T_{h+1} \leq t_{h+1}, t = T_h\} \geq 1 - 3|E|^{-\alpha} \quad .$$

Proof. Recall that $\Pr\{\text{given an instance in } \mathcal{A}_0\} \geq 1 - |E|^{-\alpha}$ by definition.

It suffices to verify:

$$\Pr\{T_{h+1} - T_h \leq \hat{\delta}(h) \mid S\} = \sum_{x=0}^{\hat{\delta}(h)-1} \Pr\{T_{h+1} - T_h = x+1 \mid S\}$$

$$\geq \frac{\varepsilon_{t+\hat{\delta}(h)-1}(h)}{\hat{\varepsilon}_t(h)} \sum_{x=0}^{\hat{\delta}(h)-1} \varepsilon_t(h)\left(1-\hat{\varepsilon}_t(h)\right)^x$$

by Lemma 5.2

$$= \frac{\varepsilon_{t+\hat{\delta}(h)-1}(h)}{\hat{\varepsilon}_t(h)}\left[1 - \left(1-\hat{\varepsilon}_t(h)\right)^{\hat{\delta}(h)}\right]$$

$$\geq \frac{\varepsilon_t(h)}{\hat{\varepsilon}_t(h)}\left[1 - \left(1-\hat{\varepsilon}_t(h)\right)^{\hat{\delta}(h)}\right] \qquad \text{by R1}$$

$$\geq 1 - |E|^{-\alpha} \qquad \text{by elementary calculations.}$$

Similarly, we can show:

$$\Pr\{T_{h+1} - T_h \geq \delta(h) \mid S\} \geq 1 - |E|^{-\alpha} \ . \qquad \square$$

As a consequence of Lemma 5.3, we may use for $1 \leq h \leq h_0$

$$\Delta(h) = \sum_{i=0}^{h-1} \delta_{\Delta(i)}(i)$$

and

$$\hat{\Delta}(h) = \sum_{i=0}^{h-1} \hat{\delta}_{\hat{\Delta}(i)}(i)$$

to lower and upper bound the time complexity of algorithm E-R with high probability. Let $\Delta(0) = \hat{\Delta}(0) = 0$.

Let $B = p|E|(1 + \sqrt{6\alpha \log|E|/p|E|}$. By Lemma 5.1 B gives an upper bound in the number of elements in an instance of M, which holds with high probability.

THEOREM 5.1. *If* $\Delta(h) \leq t_h$ *then*

$$\text{Prob}\{\Delta(h) \leq T_h \leq \hat{\Delta}(h)\} \geq 1 - a(h)|E|^{-\alpha}$$

where $a(h) = 3h(1 + r) + 1$

with

$$r = \frac{(B - t_h)}{(t_h - \Delta(h) - \hat{\Delta}(h))}$$

Proof. By Lemma 5.1,

$$\text{Prob}\{|E_0| > B\} < |E|^{-\alpha}.$$

By Lemma 5.3,

$$\text{Prob}\{\Delta(h) \leq T_h \leq \hat{\Delta}(h) \,|\, T_h \leq t_h\} \geq 1 - 3h|E|^{-\alpha} \quad .$$

Note that we may assume without loss of generality that $t_h \leq B$. By the monotonicity condition R1, we can show $\Pr\{T_h = k\}$ is unimodular for $k \in \{0,1,\ldots,|E|\}$. Thus

$$\Pr\{T_h > t_0 / |E_0| \leq B\} \leq \text{Prob}\{T_h < \Delta(h) \quad \text{or} \quad \hat{\Delta}(h) < T_h \,|\, T_h \leq t_h\} \cdot r \leq 3hr|E|^{-\alpha}.$$

But

$$\text{Prob}\{T_h > t_h\} \leq \Pr\{T_h > t_h \,\Big|\, |E_0| \leq B\} + |E|^{-\alpha} \leq (3hr + 1)|E|^{-\alpha} \quad .$$

So

$$\begin{aligned} &\text{Prob}\{T_h < \Delta(h) \quad \text{or} \quad \hat{\Delta}(h) < T_h\} \\ &\quad \leq \text{Prob}\{T_h < \Delta(h) \quad \text{or} \quad \hat{\Delta}(h) < T_h \,|\, T_h \leq t_h\} + \text{Prob}\{T_h > t_h\} \\ &\quad < a(h)|E|^{-\alpha}. \end{aligned}$$

□

Note that Theorem 5.1 may be restated:

If $\hat{\Delta}(h) \leq t_h$ *then* $\text{Prob}\{H \geq h\} \geq 1 - |E|^{-\alpha(h)}$ *where*

$$\alpha(h) = \alpha - \left(\frac{\log(1-a(h))}{\log(|E|)}\right) \quad .$$

Furthermore, if we wish

$$\text{Prob}\{H \geq h_0\} \geq 1 - |E|^{-\alpha_0}$$

for any *given* α_0 sufficiently large then we find a minimal $p_0 \in (0,1)$ such that the restrictions of Theorem 5.1 are satisfied and $\alpha_0 = \alpha(h)$. (Note that if $M = (E, \mathcal{J}, p)$ *proper* random independence system and $(E, \mathcal{J})$ has rank $\geq h_0$ then such a p_0 always exists.)

5.3 Bounds on the Probability Density Function of T_h

We assume here the restrictions given in Theorem 5.1. Actually, we have a much more general result, since we have from Lemma 5.2 bounds on the probability density function of $T_{h+1} - T_h$ for $h = 1,\ldots,h_0 - 1$. By the monotonicity restrictions R1, for $x = 0,\ldots,|E|$

$$\varepsilon_{\Delta(h+1)-1}(h)\,(1 - q(h))^x$$

$$\leqslant \text{Prob}\{T_{h+1} - T_h = x + 1 \,|\, \Delta(h) \leqslant T_h \leqslant \hat{\Delta}(h), \Delta(h+1) \leqslant T_{h+1} \leqslant \hat{\Delta}(h+1)\}$$

$$\leqslant \hat{\varepsilon}_{\hat{\Delta}(h+1)-1}(h)\,(1 - q(h))^x$$

where

$$q(h) = \varepsilon_{\Delta(h)}(h), \qquad \hat{q}(h) = \hat{\varepsilon}_{\hat{\Delta}(h)}(h) .$$

COROLLARY 5.1. *For* $h = 0,\ldots,h_0 - 1$

$$\frac{\varepsilon_{\Delta(h+1)-1}(h)}{\hat{q}(h)}\; g(x,\hat{q}(h)) \cdot |E|^{-\alpha(h+1)} \leqslant \Pr\{T_{h+1} - T_h = x + 1\}$$

$$\leqslant \frac{\hat{\varepsilon}_{\hat{\Delta}(h+1)-1}(h)}{q(h)}\; g(x,q(h)) \cdot |E|^{-\alpha(h+1)} .$$

The Appendix gives the density function of a random variable which is a sum of variables with distinct geometric distributions, and from this and by the bounds of Corollary 5.1, we have upper and lower bounds on the probability density function of the sum:

$$T_h = \sum_{k=0}^{h-1} T_{k+1} - T_k .$$

THEOREM 5.2. *For* $h = 0,\ldots,h_0 - 1$

$$Q(h) - h|E|^{-\alpha(h+1)} \leq \Pr\{T_h = x\} \leq \hat{Q}(h) + h|E|^{-\alpha(h+1)}$$

where

$$Q(h) = w_h \sum_{i=0}^{h-1} g(x,\hat{q}(i))(1-\hat{q}(i))^{h-2} \prod_{\substack{j=1 \\ i\neq j}}^{h-1} \frac{\hat{q}(i)}{\hat{q}(i)-\hat{q}(j)}$$

and

$$w_h = \left(\prod_{k=0}^{h-1} \frac{\varepsilon_{\Delta(k+1)-1}(k)}{\hat{q}(k)} \right)$$

$$\hat{Q}(h) = \hat{w}_h \sum_{i=0}^{h-1} g(x,q(i))(1-q(i))^{h-2} \prod_{\substack{j=1 \\ i\neq j}}^{h-1} \frac{q(i)}{q(i)-q(j)}$$

and

$$\hat{w}_h = \prod_{k=0}^{h-1} \frac{\hat{\varepsilon}_{\hat{\Delta}(k+1)-1}(k)}{q(k)} \quad .$$

Thus, if the restrictions of Theorem 5.1 are satisfied (as they do in our applications in Sections 6 and 7) we can derive by routine methods the mean, variance, and in general any moment of the time cost or algorithm E-R.

6. Applications to Hamiltonian Paths and Subgraph Homeomorphism Problems

6.1 Motivation and Previous Work

Posa [1976] proved a sufficient $p = O(\log n/n)$ for Hamiltonian paths in $G_{n,p}$, previously an open problem in Erdös and Spencer [1974].

Karp [1976] observed that Posa's proof yields a polynomial time algorithm for constructing Hamiltonian paths in a random instance of $G_{n,p}$. Angluin and Valiant [1979] then generalized this Posa-Karp Algorithm to detect Hamiltonian paths in random *directed* graphs.

We can also extend the Posa-Karp Algorithm to the problem of identifying certain classes of isomorphic subgraphs. Consider the problem for a fixed graph H and random graph $G_{n,p}$:

Is H homeomorphic to a subgraph of $G_{n,p}$?

The answer to this problem is very useful for determining the probability of a property characterizable by forbidden subgraphs (e.g., Kuratowski's [1971] forbidden subgraphs for planar graphs, Glover and Hyneke's [1975] forbidden subgraphs for graphs imbedded onto the projective plane, Lekkerkerker and Roland's [1962] forbidden subgraph characterization of interval graphs). Erdös and Spencer [1974] determined the probability that a random graph is planar by forbidden subgraph methods, and Cohen, Komlös and Mueller [1979] found the probability that a random graph is an interval graph by similar methods.

Actually, we can show that a large class of forbbiden subgraph problems on random graphs can be efficiently reduced to the problem of determining a Hamiltonian path. Suppose H is a graph

with k edges. Given an instance G_0 of a random graph $G_{n,p}$ we wish to construct a subgraph G' of G_0 such that G' is homeomorphic to H. (See Figure 2).

We partition the edges of $G_{n,p}$ into k blocks of cardinality n/k, with each block corresponding to an edge of H. Choose these blocks t_0 so that they have a unique "joining vertex" in common just in the case the corresponding edges of H do. Such a partitioning requires only linear time since k is constant. Then we test (by the Posa-Karp Algorithm) if each block of the partitioning has a Hamiltonian path between the "joining vertices" of the block. Each block is considered a random graph with edge probability $p' = p/k$. The application of the Posa-Karp Algorithm then yields the required Hamiltonian paths in each block with probability $\geq 1 - n^{-\alpha}$ for any sufficiently large $\alpha > 1$, if $p > c(k) \frac{\log n}{n}$ and $c(k) > k/2$.

6.2 Analysis of the Posa-Karp Algorithm

We now give a detailed analysis of the Posa-Karp Algorithm for detecting a Hamiltonian path in a random graph $G_{n,p}$. We follow the analysis techniques developed in Section 5.

Step A: *Formulation as an RIS*

We will follow here the formulation as a *non-proper* RIS (see 2.3, Examples of RIS). The extension and rotation operations are described in 3.1 of this paper. The formulation as a *proper* RIS (see 2.3, 3.1) leads to a different algorithm than the algorithm proposed by Karp. A similar analysis to the analysis presented in this chapter can show that

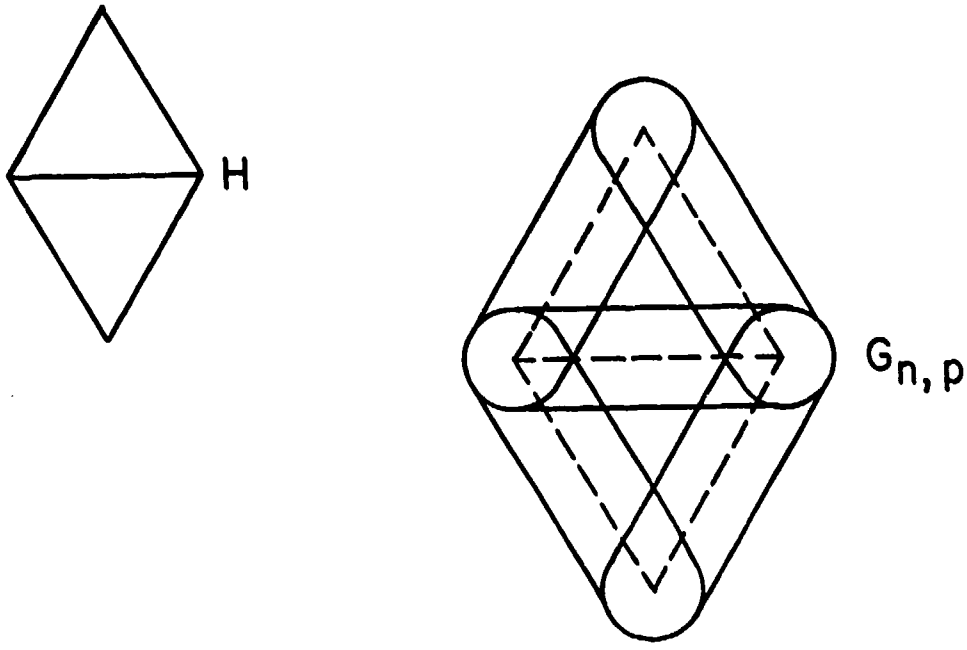

Figure 2

this new algorithm has the same performance and the same probability of success as the Posa-Karp Algorithm.

Step B: *Derivation of the Bounding Parameters:* $\varepsilon_t(h)$, $\hat{\varepsilon}_t(h)$, $\lambda_t(h)$, $\hat{\lambda}_t(h)$

Let V be the set of n vertices of the random graph $G_{n,p}$. Let $(E_0, \mathcal{J}_0)$ be an instance of $(E, \mathcal{J}, p)$ given as input to the E-R algorithm. Let I be an independent set of cardinality h, constructed after t steps of the E-R algorithm. Recall

$$\mathcal{E}_T(I) = \{e \in E_0 / e = \{u,v\}, u \in \text{ENDS}(I), v \in V - V(I)\} ,$$

where $V(I)$ is the vertex set of I. Thus, the structure of $\mathcal{E}_T(I)$ for a particular $V(I)$ depends only on the input instance $(E_0, \mathcal{J}_0)$. The E-R algorithm does not look at any of these edges at times $T' \leq T$, since if E-R examines an edge e at time T' then both vertices of e stay permanently in $V(I)$ for all $T \geq T'$.

LEMMA 6.1. *For every* β, $0 < \beta < 1$ *and for any* $p \geq c \frac{\log n}{n}$ *with* $c > 0$ *we have*

$$(1 - \beta) 2p(n - h) \leq |\mathcal{E}_T(I)| \leq (1 + \beta) 2p(n - h)$$

with probability

$$\geq 1 - 2n^{-\left(1 - \frac{h}{n}\right)\beta^2 c/3} .$$

Proof. We have observed that $|\mathcal{E}_T(I)|$ does not depend on the random variable T of the algorithm. It only depends on $|I| = h$. By definition of the $G_{n,p}$ model

$$\text{Prob}\{|\mathcal{E}_T(I)| = j\} = \binom{2(n-h)}{j} p^j (1-p)^{2(n-h)-j} .$$

The Lemma then follows by the Chernoff bounds. □

By Lemma 6.1, the mean value of $|\mathcal{E}_T(I)|$ is $2p(n-h)$.

In the following lemma, $|E| = n(n-1)/2$

LEMMA 6.2. *Let* $E_{min} = (1-\beta)2p(n-h)$ *and* $E_{max} = (1+\beta)2p(n-h)$ *with* $0 < \beta < 1$ *and* $p \geq c\frac{\log n}{n}$ *where* $c > 6/\beta^2$. *Then there is an* $\alpha > 1$ *such that*

$$\text{Prob}\{E_{min} \leq |\mathcal{E}_T(I)| \leq E_{max}\} \geq 1 - |E|^{-\alpha}$$

for $T = 0, \ldots, |E|$.

Proof. By Lemma 1 we can get $\alpha = \beta^2 c/6$ so that $\alpha > 1$ if $c > 6/\beta^2$. □

In the following, we consider edges examined by the algorithm but not added to I to be *deleted*.

LEMMA 6.3. *Then the mean number of deleted edges per vertex of* I *is the same for every* $v \in V(I)$ *and is equal to* $t/h - 1$.

Proof. Since the algorithm examines an edge at each time step and since we got up to h edges at time t, the number of deleted edges is $t-h$. These edges have their vertices in I (as previously noted). So, it is enough to show that the mean number of visits of the E-R algorithm to each vertex of I by t is the same. This follows by symmetry and since the algorithm selects at random an edge e from $\mathcal{E}_t(I) \cup \mathcal{R}_t(I)$ before each extension or rotation. Hence we get that the

mean number of deleted edges per vertex of $V(I)$ is $t-h/h$ proving the lemma. Note that this holds for any value of p. □

COROLLARY 6.3. *For any* $\beta \in (0,1)$ *there exist constants* $c > 1/2$ *and* $\alpha > 1$ *such that if* $p \geq c\frac{\log n}{n}$ *and* $m =$ number of deleted edges per vertex of $V(I)$ by time $t \leq O(n \log n)$ *then*

$$(1-\beta)\left(\frac{t}{h}-1\right) \leq m \leq (1+\beta)t/h$$

with probability $\geq 1 - |E|^{-\alpha}$.

Proof. We will first observe that for any numbers k, A and $p \geq c\frac{\log n}{n}$ with $c \geq 2(A+k+1)$ we have

$$\text{Prob}\{\text{every vertex in } G_{n,p} \text{ has } \geq k \log n \text{ edges}\} \geq 1 - O(n^{-A}) .$$

To see this, if v is any node then the probability that v has $<k \log n$ neighbors can be bounded by $O(n^{-A-1})$ by the Chernoff bounds. The result follows by summing over the n choices of v (see also Sociability Lemma of [Angluin, Valiant, 1979]).

We shall also utilize the bottleneck lemma [Angluin, Valiant, 1979] which can be described as follows:

Let us have a rooted tree of depth m and uniform S-way branching. Let Y be a set of paths from the root to certain of the leaves of the tree. Let us color green all nodes in these paths. Assume that along each path of Y there exist k nodes (called *bottlenecks*) such that at the i-th such node the probability of drawing a green successor is at most p_i. Then,

BOTTLENECK LEMMA. *The probability that a bottleneck will cut our random path to a green leaf is* $\leq p_1 \cdot p_2 \cdot \ldots \cdot p_k$.

We can now complete the Proof of the Corollary: Let the tree above be the tree of possible executions of E-R. Any vertex of I visited less than $(1-\beta)(t/h-1)$ times or more than $(1+\beta)t/h$ times can be considered as a bottleneck and this event would be bounded by the sum of the probabilities

$$\left[\left(n-(1-\beta)\left(\frac{t}{h}-1\right)\right)^{-k\log n}+\left(n-(1+\beta)\left(\frac{t}{h}-1\right)\right)^{-k\log n}\right]$$

for all possible vertices, which is $O(n^{-\alpha})$ for suitable values of k.□

LEMMA 6.4. *For any* $\beta\in(0,1)$ *there are constants* $\alpha>1$ *and* $c=c(\beta,\alpha)>0$ *such that*

$$(1-\beta)2ph-(1+\beta)\frac{t}{h}\leq|\mathcal{R}_t(I)|\leq(1+\beta)2ph-\left(\frac{t}{h}-1\right)(1-\beta)$$

holds with probability $\geq 1-|E|^{-\alpha}$.

Proof. Let A_1 be the number of edges from endpoints of I to vertices of I at time $t=0$ ($V(I)$ is fixed here) and let A_2 be the number of edges deleted from the endpoints of I up to time t. Then $|\mathcal{R}_t(I)|=A_1-A_2$. By the Lemma 6.2 and Corollary 6.3 we get the result.

Applying Property (*) at the beginning of Chapter 4 and Lemmas 6.2, 6.4 we get that

$$\varepsilon_t(h)=\frac{(n-h)(1-\beta)}{n(1+\beta)-\frac{t}{2ph}(1-\beta)}\quad\text{and}\quad\hat{\varepsilon}_t(h)=\frac{(n-h)(1+\beta)}{n(1-\beta)-\frac{t}{2ph}(1+\beta)}$$

are bounds on the conditional extension probability of the E-R algorithm:

$$\varepsilon_t(h) \leq \frac{|\mathscr{E}_t(h)|}{|\mathscr{E}_t(h)| + |\mathscr{R}_t(h)|} \leq \hat{\varepsilon}_t(h)$$

for executions in a class $\mathscr{A}_0$ of total probability $\geq 1 - |E|^{-\alpha}$.

Observe that

$$\frac{\partial \varepsilon_t(h)}{\partial t} > 0, \quad \frac{\partial \hat{\varepsilon}_t(h)}{\partial t} > 0, \quad \frac{\partial \hat{\varepsilon}_t(h)}{\partial h} < 0 \quad \text{and} \quad \frac{\partial \hat{\varepsilon}_t(h)}{\partial h} < 0$$

so the monotonicity condition R1 is satisfied.

Note that Lemma 5.1 fixes $\beta = \sqrt{6\alpha \log|E|/|E|\cdot p}$ where $|E| = n(n-1)/2$. Note also that since $2ph$ is the mean value of the number of edges from the endpoints of I to the other vertices of I at the beginning of an execution of E-R and since $2t/h$ is the mean number of deleted edges from the endpoints of I by the time t, we must have (in order for E-R not to stop at t) that

$$t/h \leq ph \qquad \text{or} \qquad t \leq ph^2 .$$

For $\bar{T} = O(n \log n)$ and $h = n-1$ this again implies $p \geq O(\frac{\log n}{n})$ for the E-R algorithm to be able to construct a Hamiltonian line. (Compare with the general statement at the end of Section 4.)

Restriction R2 can be readily verified for it is obvious that

$$\text{Prob}\{\mathscr{E}_t(I) \cup \mathscr{R}_t(I) = \phi\}$$

monotonically increases with t and $h = |I|$.

To satisfy condition C1 we set $t_h = 2pnh(1-\beta)$. Then for executions in $\mathscr{A}_0$ and $0 \leq t \leq t_h$,

$$\mathscr{E}_t(I) \cup \mathscr{R}_t(I) \neq \phi \quad .$$

Step C: *Verification of C2*

We now must verify condition C2 to insure the algorithm succeeds with high probability. For simplicity, we proceed with the *asymptotic analysis as* $n \to \infty$ (although the techniques of Section 5 allow analysis for any fixed n as well). Note that as $n \to \infty$, $\beta \to 0$ so

$$\varepsilon_t(h) \sim \hat{\varepsilon}_t(h) \sim \frac{n-h}{n-\frac{t}{2ph}}$$

so in the asymptotic case the bounding parameters are identical.

Also,

$$\hat{\delta}_t(h) \sim \frac{\alpha \log \ell}{\log(1-\varepsilon_t(h))} \qquad \text{as} \quad n \to \infty ,$$

where

$$\ell = \frac{n(n-1)}{2} .$$

We must determine

$$\hat{\Delta}(h+1) = \hat{\Delta}(h) + \hat{\delta}_{\hat{\Delta}(h)}(h) .$$

Let

$$k_1 = \frac{pn}{\log n} .$$

We now show by induction on h that $\hat{\Delta}(h) \leq k_2 h \log n$ where $k_2 = \frac{2\alpha k_1}{2\alpha + k_1}$.

LEMMA 6.5. *Assume* $p \geq c \frac{\log n}{n}$. *Then* $\hat{\Delta}(x) \leq k x \log n$, *where* $k \geq \frac{2\alpha c}{2\alpha + c}$ *and* α *is the constant appearing in* $\hat{\delta}_t(h)$.

Proof. We have from the definition of $\hat{\Delta}(h)$ that

$$\hat{\Delta}(h) = \sum_{i=0}^{h-1} \hat{\delta}_{\hat{\Delta}(i)}(i)$$

and $\hat{\Delta}(0) = 0$. It follows that $\hat{\Delta}(h) = \hat{\Delta}(h-1) + \hat{\delta}_{\hat{\Delta}(h-1)}(h-1)$. Also from

$$\hat{\delta}_t(h) \simeq \frac{-\alpha \log n}{\log(1-\varepsilon_t(h))}$$

as $n \to \infty$ we get

$$\hat{\delta}_t(h) \quad \frac{-2\alpha \log n}{\log(1-\varepsilon_t(h))}$$

as $n \to \infty$.

Basis:

Since at 0 edges, E-R will increase the size of I with certainty in the first attempt, we must have $\hat{\delta}_{\hat{\Delta}(0)}(0) = 1$. Then $\hat{\Delta}(1) = \hat{\Delta}(0) + \hat{\delta}_{\hat{\Delta}(0)}(0) = 1 \leq k \log n$ for large n.

Induction Hypothesis:

Assume that for $k \geq \frac{2\alpha c}{2\alpha + c}$ and all j in $\{0,1,\ldots,x-1\}$ it is true that $\hat{\Delta}(j) \leq kj \log n$.

Induction Step:

We have, by replacing $\hat{\delta}_{\hat{\Delta}(x-1)}(x-1)$ in the equation for $\hat{\Delta}(x)$

$$\hat{\Delta}(x) = \hat{\Delta}(x-1) + \frac{-2\alpha \log n}{\log\left(1 - \frac{2p(n-x+1)}{2pn-2\hat{\Delta}(x-1)/(x-1)}\right)}$$

By the induction assumption we may substitute $\hat{\Delta}(x-1) \leq (x-1) \cdot k \log n$ and by using also $p \geq c \log n/n$ we get by elementary manipulations that

$$\hat{\Delta}(x) \leq k(x+x') \log n$$

where

$$x' = \frac{2\alpha/k}{\log[2c-2k] - \log\left[\frac{2c(x-1)}{n} - 2k\right]} - 1 \quad .$$

But $x' < 0$ for $k \geq \frac{2\alpha c}{2\alpha + c}$ and $x \leq n-1$ as assumed.

Thus $\hat{\Delta}(x) \leq k \times \log n$. □

Thus for $c \geq \frac{k + 2\alpha}{2\alpha - 1}$ we have $\hat{\Delta}(h) \leq t_{n-1}$ and we conclude that the E-R algorithm outputs a Hamiltonian path with probability $\geq 1 - |E|^{-\alpha_0}$ where $\alpha_0 < \alpha - 1/2$.

Step D: *Bounds on the Mean and Variance of* T_h

We have from Corollary 5.1 that

$$\text{Prob}\{T_{h+1} - T_h = x + 1\} \leq \hat{s}_h q(x, q(h)) + |E|^{-\alpha(h+1)}$$

where

$$\hat{s}_h = \frac{\hat{\varepsilon}_{\hat{\Delta}(h+1)-1}(h)}{q(h)} \qquad \text{and} \qquad q(h) = \varepsilon_{\Delta(h)}(h) \quad .$$

This requires calculation of the lower bound $\Delta(n)$, which in this application is trivial: $\Delta(h) = h$. But $s_h \sim 1/(1 - k_2/k_1)$ is constant for $p = \theta(\log n/n)$. Also, for $\alpha(h+1) > 0$, $|E|^{-\alpha(h+1)} \to 0$ as $|E| \to \infty$.

D.a: *Upper Bound on the Mean of* T_{h_0} *for* $h_0 = n-1$

From the Lemma 6.5 we remark that the upper bound of the mean must be $\leq kn \log n$. To analytically derive a more tight bound, we have:

$$\hat{\varepsilon}_{\hat{\Delta}(h+1)}(h) = \frac{n-h}{n - \frac{\hat{\Delta}(h+1)}{2ph}} \leq \frac{n-h}{n - \frac{k(h+1)\log n}{2ph}}$$

(by the fact $\hat{\Delta}(x) \leq k \times \log n$).

So, we get

$$\hat{\varepsilon}_{\hat{\Delta}(h+1)}(h) \leq \frac{n-h}{n - \frac{kn}{2c} \cdot \frac{h+1}{h}} \qquad \text{(using } pn \geq c \log n\text{)}$$

So,

$$\hat{\varepsilon}_{\hat{\Delta}(h+1)}(h) \leq \frac{n-h}{n - \frac{kn}{c}} = \left(\frac{1}{1 - \frac{k}{c}}\right)\left(\frac{n-h}{n}\right)$$

$$\leq \left(\frac{1}{1 - k/c}\right) \cdot \; q(h)$$

where

$$q(h) = 1 - \frac{h}{n} \quad .$$

Let us define a constant

$$d' = \frac{1}{1 - \frac{k}{c}} \quad .$$

Then

$$\hat{s}_h = \frac{\hat{\varepsilon}_{\hat{\Delta}(h+1)}(h)}{q(h)} = d' \quad .$$

Then, from Corollary 5.1 and the Appendix we get (by taking means) that

$$\text{mean}(T_{h+1} - T_h) \leq \hat{s}_h \cdot \frac{1 - q(h)}{q(h)} = d' \frac{h}{n-h} \quad .$$

So

$$\text{mean}(T_{h_0}) = \sum_{h=0}^{n-1} \text{mean}(T_{h+1} - T_h) \leq d' \int_0^{n-1} \frac{h}{n-h} \, dh$$

$$\leq d'[n \log n - (n-1)] \quad .$$

D.b: *Lower Bounds on the Mean of* T_{h_0}

Again we do an asymptotic analysis as $n \to \infty$. We have

$$\varepsilon_{\Delta(h+1)}(h) = \frac{2p(n-h)}{2pn - \frac{\Delta(h+1)}{h}}$$

$$\simeq \frac{1 - h/n}{1 - (h+1)/2phn} , \qquad \text{by using } \Delta(h) = h.$$

Since $pn \geq c \log n$,

$$\varepsilon_{\Delta(h+1)}(h) \simeq 1 - \frac{h}{n} \qquad \text{as} \quad n \to \infty .$$

Also,

$$s(h) = \frac{\varepsilon_{\Delta(h+1)}(h)}{\hat{q}(h)} \simeq \frac{1 - h/n}{(1 - h/n)\left(\frac{1}{1 - k/2c}\right)} \simeq \frac{1}{d}$$

with $d = (1 - k/2c)^{-1}$.

By Corollary 5.1 and Appendix

$$\text{mean}(T_{h+1} - T_h) \geq s(h) \frac{1 - \hat{q}(h)}{\hat{q}(h)} \geq f(h)$$

where

$$\hat{q}(h) = d\left(1 - \frac{h}{n}\right) \qquad \text{and} \qquad f(h) = \frac{n - d(n-h)}{d^2(n-h)}$$

So,

$$\text{mean}(T_{h_0}) \geq \sum_{h=0}^{n-1} \text{mean}(T_{h+1} - T_h) = \sum_{h=0}^{n-1} f(h) ,$$

then

$$\text{mean}(T_{h_0}) > \int_0^{n-1} f(h)\, dh - f(0) > \frac{n \log n}{d^2} - \frac{n-1}{d} - \frac{1-d}{d^2} . \qquad \square$$

As $n \to \infty$, the obtained lower and upper bounds are tight within a constant factor

$$d'd^2 = \frac{c}{c-k}\left(\frac{2c}{2c-k}\right)^2 \qquad \text{if} \quad p = \theta\left(\frac{\log n}{n}\right) .$$

Thus,

COROLLARY

$$\text{mean}(T_{h_0}) = \theta(n \log n) \qquad \textit{for} \quad p = \theta\left(\frac{\log n}{n}\right) .$$

D.c: *Upper Bounds on the Second Moment of* T_{h_0}

From the Appendix

$$\text{mean}(Y^2) = \sum_{i=1}^{m} p_i D_i \left[r_i \frac{\partial h}{\partial r_i} + r_i^2 \frac{\partial^2 h}{\partial r_i^2} \right]$$

where Y is a sum of m truncated geometrics of parameters p_i and r_i and

$$h(r_i) = \frac{r_i^{s+1} - 1}{r_i - 1} , \qquad s = m\, n_0 ,$$

n_0 the truncation point, and

$$D_i = r_i^{n_0 - 1} \cdot \prod_{j \neq i} \frac{p_j}{p_i - p_j} .$$

In our case, $p_h = q(h) = n-h/(n-1)$. So,

$$(1 - p_i)^{n-1} = \left(\frac{i-1}{n-1}\right)^{n-1} \leq \exp(i-n)$$

and by noting that

$$\prod_{i \neq 1} \frac{n-j}{j-1} = \frac{i}{n-i} (-1)^{i-1}$$

we get

$$D_i \leq \frac{i}{n-i} \cdot \exp(i-n).$$

By the Appendix

$$\text{mean}(T_{n-1}^2) \simeq \sum_{i=1}^{n-1} D_i \left(\frac{2}{p_i^2} - \frac{3}{p_i} + 2 \right)$$

$$\leq \sum_{i=1}^{n-1} \exp(i-n) \frac{i}{n-i} \left(\frac{2}{p_i^2} - \frac{3}{p_i} + 2 \right) .$$

Using $p_i \simeq 1 - i/n$ for large n and replacing the above sum by an integral, we get

$$\text{mean}(T_{n-1}^2) \leq (n^3 + 5n^2 + 5n) \cdot \frac{c}{\exp(1)} \qquad \text{as} \quad n \to \infty .$$

By using the lower bound for the mean and the upper bound on the second moment we can get an upper bound on the variance as follows:

$$\text{var}(T_{n-1}) = \text{mean}(T_{n-1}^2) - \text{mean}^2(T_{n-1}) .$$

So

$$\text{var}(T_{n-1}) \leq \frac{c}{\exp(1)} (n^3 + 5n^2 + 5n) - \left(\frac{n \log n}{d^2} - \frac{n-1}{d} - \frac{1-d}{d^2} \right)^2 .$$

D.d: *Lower Bounds on* T_{n-1}^2

For the lower bound, we use $p_i = \hat{q}(i) = d(1 - \frac{i}{n})$ in the formula for D_i. Let

$$B = \left[1 + \frac{di}{n(1-d)} \right]^{n-1} (1-d)^{n-1} .$$

Then

$$D_i = B \frac{i}{n-i} (-1)^{n-i} .$$

We can prove by an easy induction that $B \geq \exp(d(i-n))$. So we get

$$\text{mean}(T_{n-1}^2) \geq \sum_{i=1}^{n-1} u(i)$$

where

$$u(i) = \exp(d(i-n)) \frac{i}{n-i} (-1)^{n-i} A_i$$

and

$$A_i = \frac{2}{p_i^2} - \frac{3}{p_i} + 2$$

or

$$\text{mean}(T_{n-1}^2) \geq \int_{i=1}^{n-1} u(i)\, di - u(0) .$$

A calculation of this integral gives us

$$\text{mean}(T_{n-1}^2) \geq \exp(-d)\left[\frac{3}{4} d^2 n^3 + \left(\frac{1}{2d} - \frac{3}{4}\right) n\right] .$$

A lower bound on $\text{var}(T_{n-1})$ follows immediately from our bounds on $\text{mean}(T_{n-1}^2)$ and $\text{mean}(T_{n-1})$. Hence,

LEMMA 6.6

$$\frac{3}{4} e^{-d} n^3 + \Omega(n) \leq \text{mean}(T_{n-1}^2) \leq \frac{c}{e} n^3 + O(n^2)$$

and $\text{var}(T_{h_0}) = \theta(n^3)$, *if* $\exp(d)c$ *is constant.*

This completes the analysis of the Posa-Karp Algorithm.

7. Applications to Matchings

7.1 The E-R Algorithm for Matchings in Random Bipartite Graphs

Step A: *Formulation as an RIS*

We will follow here the formulation as an RIS given in 2.3 (Examples of RIS). The extension and rotation operations are described in 3.1. Let G_0 be an instance of the random graph $B_{n,p}$ and let I be an independent set of size h, obtained after t steps of the E-R algorithm.

Step B: *Derivation of the Bounding Parameters*

By the definition of the rotation and extension, we note that as soon as an edge e is examined by the algorithm, both its vertices stay at I for subsequent time steps. Hence, $|\mathcal{E}_T(I)|$ follows the same distribution (as in Lemma 6.1) with mean $|\mathcal{E}_T(\cdot)| = p(n-h)$. Lemma 1 also holds here (since it depends only on the cardinality of I) and Corollary 6.3 can be proved by similar arguments. For $p \geq c \log n/n$ we get exactly the same values of x_{min}, x_{max}, y_{min}, y_{max} and the same asymptotic expressions for $\varepsilon_t(h)$, $\hat{\varepsilon}_t(h)$.

Steps C and D:

The analysis is the same as in the corresponding steps of the analysis of the Posa-Karp algorithm. So, we get:

If $p \geq c \log n/n$, the algorithm E-R constructs a perfect matching I with $|I| = n$ in the random bipartite graph $B_{n,p}$, in average time $\text{mean}(T_n) = \theta(n \log n)$, with probability of success $\geq 1 - n^{-2\alpha}$, $\alpha > 1$. The constant c depends on α as in the Posa-Karp case. The

second moment again satisfies $\text{mean}(T_n^2) = \theta(n^3)$, so $\text{var}(T_n) = \theta(n^3)$.

7.2 The E-R Algorithm for Matchings in Random Graphs $G_{2n,p}$

Previously, Angluin and Valiant [1979] and Walkup [1977] have described algorithms for detecting perfect matchings in a random graph $G_{2n,p}$ with $p \geq c(\log n)/n$. We now briefly sketch an analysis of the performance of the extension-rotation algorithm for perfect matching.

Step A: *Formulation as an RIS*

We will follow the formulation given in 2.3 and use the extension and rotation as in 3.1.

Step B: *Derivation of* $\varepsilon_t(h)$, $\hat{\varepsilon}_t(h)$

Let

$$a(h) = (n-h)(2n-2h-1)$$

$$a'(h) = 4ph(n-h)$$

$$f_t(h) = t(n-h-1/2)(n-h)/n^2$$

$$f_t'(h) = ht(n-h)/n^2 .$$

Again, we may use symmetry arguments and Lemma 5.1 to bound the cardinalities of $\mathscr{E}_t(I)$, $\mathscr{R}_t(I)$ and $|E_0|$ for a class of executions $\mathscr{A}_0$ with probability $\geq 1 - |E|^{-\alpha}$. Let $h = |I|$.

For executions in $\mathscr{A}_0$,

$$(1-\beta)a(h) \leq |\mathscr{E}_t(I)| + f_t(h) \leq (1+\beta)a(h)$$

and

$$(1-\beta)a'(h) \leq |\mathscr{R}_t(I)| + f_t'(h) \leq (1+\beta)a'(h) .$$

Let

$$t_h = (1-\beta)(a(h) + a'(h)) - f_t(h) - f'_t(h) \quad .$$

Then $|\mathscr{E}_t(I)| + |\mathscr{R}_t(I)| > 0$ for $t \leq t_h$ in executions of $\mathscr{A}_0$, verifying condition C1.

We may let

$$\varepsilon_t(h) = \frac{(1-\beta)a(h) - f_t(h)}{(1+\beta)(a(h) + a'(h)) - f_t(h) - f'_t(h)}$$

$$\hat{\varepsilon}_t(h) = \frac{(1+\beta)a(h) - f_t(h)}{t_h}$$

so we have

$$\varepsilon_t(h) \leq \frac{|\mathscr{E}_t(I)|}{|\mathscr{E}_t(I)| + |\mathscr{R}_t(I)|} \leq \hat{\varepsilon}_t(h)$$

for executions in $\mathscr{A}_0$.

By taking partial derivatives of $\varepsilon_t(h)$ with respect to t and h, we can again show the monotonicity condition R1 is satisfied. It is also obvious that monotonicity condition R2 holds.

As $n \to \infty$, the asymptotic bounds on the conditional extension probability is again tight: $\varepsilon_t(h) \sim \hat{\varepsilon}_t(h)$. By the routine calculations, described in Section 5, the reader may verify that $\hat{\Delta}(n) \leq t_n$, so the E-R algorithm outpus a perfect matching with probability $\geq 1 - |E|^{-\alpha(n)}$. We also leave the reader to calculate tight bounds on the mean and variance of T_n:

$$\text{mean}(T_n) = \theta(n \log n) \quad \text{and} \quad \text{mean}(T_n^2) = \theta(n^3)$$

by applying Corollary 5.1 (which bounds the probability density of $T_{h+1} - T_h$ by geometric density functions) and using the formulas of the Appendix to calculate the moments, as we did in the Hamiltonian path applications.

Angluin and Valiant [1979] show that each "unit time" step of Algorithm E-R for this application requires $\theta(\log n)$ instructions on a RAM machine. Thus, the above mean and variance bounds must be multiplied by a constant multiple of $\log n$ and $(\log n)^2$, respectively.

Acknowledgments

We wish to thank Andy Langer, Allen Emerson, and Christos Papadimitriou for their helpful suggestions on these topics.

Bibliography

Angluin, D. and L. Valiant, "Fast probabilistic algorithms for Hamiltonian circuits and matchings," *J. Computer System Sciences*, 18, 1979.

Cohen, J., J. Komlös, and T. Mueller, "The probability of an interval graph and why it matters," *Proc. Symposia in Pure Mathematics*, 34, 1979.

Erdös, P. and A. Renyi, "On random graphs," *Publicationes Mathematicae*, 6, 1959, pp. 290-297.

Erdös, P. and A. Renyi, "On the evolution of random graphs," *Publ. Math. Inst. Hung. Acad. Sci.*, 5A, 1960, pp. 17-61.

Erdös, P. and J. Spencer, *Probabilistic Methods in Combinatorics*, Academic Press, New York, 1974.

Feller, W., *An Introduction to Probability Theory and Its Applications*, Vol. 1, Third Edition, John Wiley and Sons, New York, 1968.

Grimmet, G.S. and C.J. McDiarmid, "On coloring random graphs," *Math. Proc. Camb. Phil. Soc.*, 77, 1975, pp. 313-324.

Glover, H. and J.P. Huneke, "Cubic irreducible graphs for the projective plane," *Discrete Mathematics*, 13, 1975, pp. 341-355.

Karp, R.M., "The probabilistic analysis of some combinatorial search algorithms," *Algorithms and Complexity: New Directions and Recent Results*, J.F. Traub, ed., Academic Press, New York, 1976, pp. 1-19.

Korte, R. and D. Hausmann, "An analysis of the greedy heuristic for independence systems," *Annals of Discrete Mathematics* 2, 1978, pp. 65-74.

Kuratowski, K., "Sur le problème des courbes gauches en topologie," *Fund. Math.* 15, 1930, pp. 217-283.

Lawler, E.L., *Combinatorial Optimization: Networks and Matroids*, Holt, Rinehard and Winston, 1976.

Lekkerkerker, C.G. and J.C. Boland, "Representation of a finite graph by a set of intervals on the real line," *Fund. Math. Polska Akad.*

Lueker, G.S., "Maximization on graphs with edge weights chosen from a normal distribution," *Proc. Tenth Annual Symposium on Theory of Computing*, San Diego, California, 1978.

Matula, D.W., "On the complete subgraphs of a random graph," *Proc. 2nd Chapel Hill Conference on Combinatorial Math. and Its Applications*, University of North Carolina, Chapel Hill, May 1970, pp. 356-369.

Papoulis, A., *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, 1965.

Posa, L., "Hamiltonian circuits in random graphs," *Discrete Mathematics*, 14, 1976, pp. 359-364.

Reif, J.H. and P.G. Spirakis, "Random Independence Systems," to appear in 1981, also appearing in preliminary form as the 2nd and 3rd Chapter of "Random Matroids," STOC, 1980.

Tutte, W.T., *Introduction to the Theory of Matroids*, American Elsevier, New York, 1971.

Walkup, D.W., "On the expected value of a random assignment problem," draft, December, 1977.

Walkup, D.W., "Matchings in random regular bipartite graphs," draft, December, 1977.

Whitney, H., "On the abstract properties of linear dependence," *American J. Mathematics*, 57, 1935, pp. 509-533.

APPENDIX

We consider a random variable Y which is a sum

$$Y = X_1 + \ldots X_m$$

of geometrically distributed variables $X_1,\ldots,X_m$. This Appendix provides formulas for the mean, variance and some low order moments of Y.

For each $i = 1,\ldots,m$ we assume X_i has truncated geometric density with parameter $p_i \in [0,1]$. Let $r_i = 1 - p_i$ and

$$g_i(k) = p_i r_i^k, \qquad k = 0,1,\ldots,n_0$$
$$= 0 \quad \text{else}$$

The density function of $X_1 + X_2$ is for $0 \leq k \leq 2n_0$,

$$g_1 * g_2(k) = \sum_{j=0}^{k} g_1(j) g_2(k-j)$$

$$= \frac{p_1 p_2}{p_2 - p_1} \left[r_1^{k+1} - r_2^{k+1} \right] .$$

By applying induction, we derive the density function of

$$Y = \sum_{i=1}^{m} X_i$$

$$f(k) = (g_1 * \ldots * g_m)(k)$$

$$= \sum_{i=1}^{m} g_i(k) r_i^{m-1} \prod_{\substack{j=1 \\ j \neq i}}^{m} \frac{p_j}{r_i - p_j} .$$

The t-th moment of Y is given by

$$\text{mean}(Y^t) = \sum_{k=0}^{s} k^t (g_1 * \ldots * g_m)(k)$$

when $s = mn_0$.

Mean of Y:

$$\text{mean}(Y) = \sum_{i=1}^{m} \text{mean}(X_i)$$

$$\text{mean}(X_i) = \frac{r_i}{p_i}\left[1 - r_i^{n_0}(n_0 p_i + 1)\right]$$

Variance of Y:

$$\text{mean}(Y^2) = \sum_{i=1}^{m} p_i D_i \left[r_i \frac{\partial h}{\partial r_i} + r_i^2 \frac{\partial^2 h}{\partial 2 r_i}\right]$$

where

$$h(r_i) = \frac{r_i^{s+1} - 1}{r_i - 1}$$

$$D_i = r_i^{n_0 - 1} \prod_{j \neq 1} \frac{p_j}{p_i - p_j}$$

Asymptotic Analysis:

Note that as $s \to \infty$

$$\frac{\partial h}{\partial r_i} \to \frac{1}{p_i^2}$$

$$\frac{\partial^2 h}{\partial r_i^2} \to \frac{1}{p_i^3}$$

$$\text{mean}(Y^2) \to \sum_{i=1}^{m} D_i \left(\frac{2}{p_i^2} - \frac{3}{p_i} + 2 \right) \quad .$$