PROBABILISTIC ANALYSIS OF RANDOM EXTENSION-ROTATION ALGORITHMS

John H. Reif and Paul G. Spirakis

TR-28-81

October 1981

Technical Report of Division of Applied Science, Harvard University TR-28-81, 1980.

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.

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.)

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.

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 $E = \{\{u,v\} / u \in V_1, v \in V_2\}$.

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: $I \leftarrow \emptyset$; $T \leftarrow 0$
WHILE $|I| < h_0$ DO
  BEGIN
    IF $\mathcal{E}_T(I) \cup \mathcal{R}_T(I) = \emptyset$ THEN FAIL
    choose some random $e \in \mathcal{E}_T(I) \cup \mathcal{R}_T(I)$
    IF $e \in \mathcal{E}_T(I)$ THEN EXTEND: $I \leftarrow I \cup \{e\}$

(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.

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 to 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

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

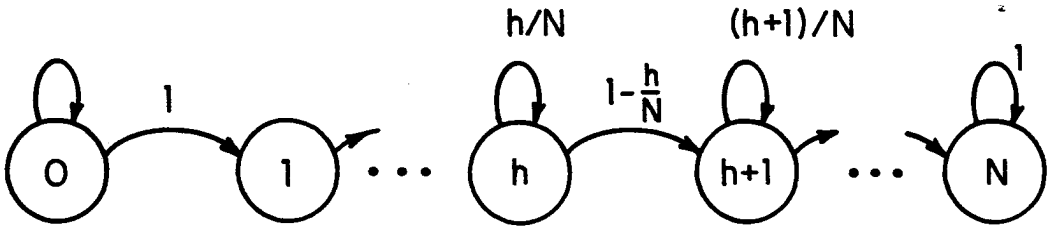

Figure 1

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*

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 $\mathcal{A}_0$ 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

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

LEMMA 5.2.

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

Proof. By conditions C1 and monotonicity restriction R1,

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

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

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

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

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

The lower bound derivation is similar. □

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:

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

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,

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

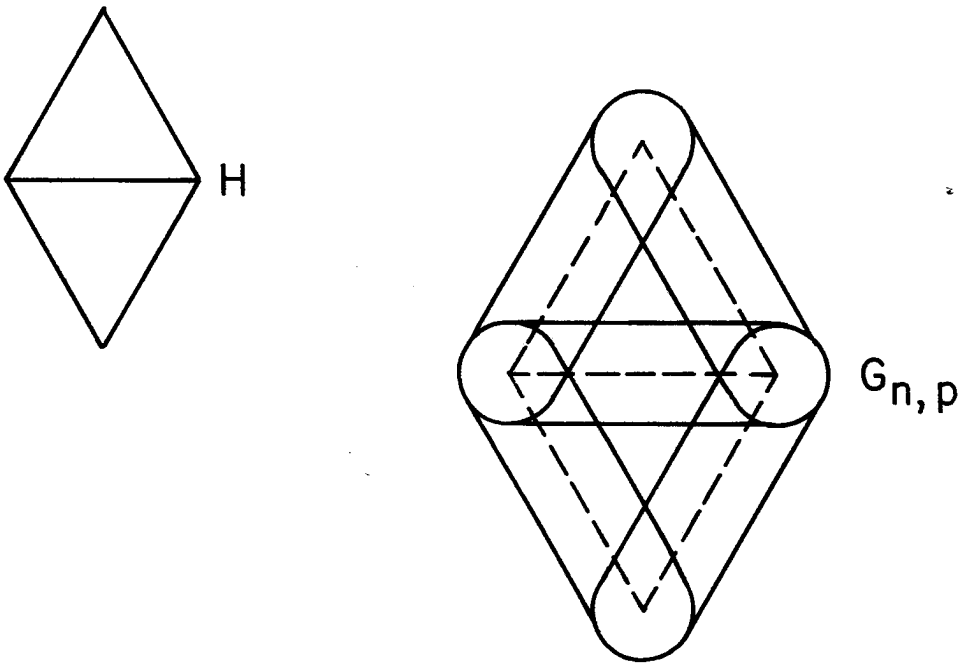

Figure 2

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

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

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*

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

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

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 $|\mathscr{R}_t(I)|=A_1-A_2$. 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:

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

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

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

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 $\Delta(h)$, which in this application is trivial: $\Delta(h) = h$. But $s_h \sim 1/(1 - k_2/k_1)$ is constant for $p = \vartheta(\log n/n)$. 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 $\hat{\Delta}(x) \leq k \times \log n$).

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

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

we get

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

By the Appendix

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

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

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

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

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

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

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

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 $|\mathcal{E}_T(I)| = p(n-h)$. Lemma 6.1 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

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

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.

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.

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