RANDOM MATROIDS

John H. Reif

Paul G. Spirakis

Harvard University

0. Abstract

We introduce a new random structure generalizing matroids. These *random matroids* allow us to develop general techniques for solving hard combinatorial optimization problems with random inputs.

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) *random matroids* which are applicable to a more general class of combinatorial optimization problems with *random inputs*. We define some natural notions, such as "maximal with a given measure." Section 2 also presents some easy results (which nevertheless we believe are worth mentioning) on the *relationship between random and Whitney matroids*, and on *intersections* of random matroids. In the last part of Section 2 we define *weighted random matroids*.

Section 3 sketches a general *nonconstructive proof technique* for determining the existence (with probability 1) of an independent set of given cardinality in instances of a random matroid; this encompasses various nonconstructive proofs of graph properties in [Erdös and Spencer, 1974] (and complements the next section which is concerned with efficient algorithms for *constructing* independent sets of given size.) We also provide a nonconstructive proof technique for determining the existence of an independent set of given weight in a weighted random matroid.

Section 4 considers a random algorithm for efficiently constructing an independent set of size h_0 in an instance of a random matroid. 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.

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 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.

Section 5 gives some applications to random graphs of type $G_{n,p}$ (see [Erdös and Spencer, 1974]).

P1 Construct a *perfect matching* in $G_{n,p}$.

P2 Construct a *Hamiltonian path* in $G_{n,p}$.

P2' For a graph H homeomorphic to a graph of fixed size, construct a subgraph of $G_{n,p}$ isomorphic to H.

Note that P2' is a generalization of P2.

The random algorithm of Section 4 is applicable to both P1 and P2 (and we have an efficient transformation from instances of P2' to instances of P2).

The results of Section 4 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 P2 for undirected graphs. [Angluin and Valiant, 1979] consider both 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 $n \to \infty$.

Section 4 also yields significant *new* results for these applications, such as *tight* bounds (within a constant multiple) on the mean and variance of the random algorithm's time complexity

2. Definitions of Random Matroids and their Structure

2.1. Definitions of Random Matroids

Let E be a set and let $\mathcal{J}$ be a family of subsets of E. For each element $e \in E$, let p_e be a real number (the *element's density*) on the interval $[0,1]$. The triple $M = (E,\mathcal{J},\{p_e\})$ is a *random matroid*. If for some fixed p, $p_e = p$ for all elements $e \in E$ then M is *uniform* and denoted $(E,\mathcal{J},p)$. We will frequently write $(E,\mathcal{J},1)$ as $(E,\mathcal{J})$. $M = (E,\mathcal{J},\{p_e\})$ is a *proper random matroid* 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 A) 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 Matroids

An *instance* of random matroid $M = (E,\mathcal{J},\{p_e\})$ 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_e

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

Note that the *measure* of M_0 is $(\prod_{e \in E_0} p_e)(\prod_{e \in E - E_0}(1-p_e))$.

Clearly, any instance $M_0 = (E_0,\mathcal{J}_0)$ of a proper random matroid satisfies axioms A1 and A2, but M_0 may not satisfy A3 even if $(\mathcal{J},E)$ is a Whitney matroid. A set $A \subseteq E_0$ is *independent in* M_0 if $A \in \mathcal{J}_0$ and *dependent* otherwise. An independent set $I \in \mathcal{J}_0$ is *maximum in* M_0 if $\neg(\exists I' \in \mathcal{J}_0$ s.t. $|I'| > |I|)$. 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 $\neg(\exists I' \in \mathcal{J}_0$ s.t. $I' > I)$. 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. Also, let the *rank of* M_0 be the size of a maximum independent set of E_0.

2.3. Examples of Random Matroids

As an example of a random matroid, let P be a property on graphs and let $G_{n,p}$ be a random undirected graph. $G_{n,p}$ has instances which are graphs with vertices $V = \{1,2,\ldots,n\}$ and each edge chosen independently with probability p from unordered pairs of distinct vertices in V.

Let $M = (E,\mathcal{J},p)$ be the uniform random matroid with $E = \{\{u,v\} \mid \text{distinct } u,v \in V\}$ and $\mathcal{J} = \{E' \subseteq E \; P(V,E') \text{holds}\}$. 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 $E' \subseteq E_0$ such that the property P holds for subgraph (V,E').

If the graph property P is *trivially satisfied* (P holds for the graph with no edges) and *decreasing monotone* ($P(G) \Rightarrow P(G')$ for all subgraphs G' of G) then M is a proper random matroid.

An (edge) *matching* of a graph is a set of vertex disjoint edges, and is *perfect* if every vertex appears in an edge of the matching. A *simple path* is a path of edges containing no cycles, and is a *Hamiltonian path* if it contains every vertex.

The property of a "matching" in a random graph yields a proper random matroid, but the property of a "simple path" in a random graph does not yield a proper random matroid, since a simple path must be connected (violating axiom A2).

2.4. Maximality in Random Matroids

The definitions of maximum, maximal, and minimal are all standard for monotone properties of *deterministic* combinatorial structures. We extend these notions to independence in random matroids, which is a *random* property.

Let $M = (E,\mathcal{J},\{p_e\})$ be a random matroid and let $A \in \mathcal{J}$. Let A be *maximum with measure* m *in* M if $m = \Pr\{A \text{ is maximum in the same instance} \mid A \text{ appears in an instance}\}$ (All probabilities are defined over the possible instances of M).

Let A be *maximal with measure* m *in* M if $m = \Pr\{A \text{ maximal in the same instance} \mid A \text{ appears in an instance}\}$.

Similarly, let a set $A \in 2^E - \mathcal{J}$ be *minimal* (a *circuit*) *with measure* m *in* M if $m = \Pr\{A$ is a minimal dependent set of the same instance$\mid A$ appears in an instance$\}$. Let *rank* (M) be the random variable giving the rank of instances of M.

For all $m \in [0,1]$, let $\delta_M(m)$ be the minimal $m' \in [0,1]$ such that $\forall A \in \mathcal{J}$: A maximal with measure m in M implies A is maximum with measure $\leq m'$ in M. (It is obvious that $m \geq m'$ and that $\delta_M(m)$ is increasing with p).

The function $\delta_M(m)$ gives us a measure with which simple greedy-like algorithms succeed in constructing maximum sets. A similar function may be defined for the measure of success of rotation-greedy algorithms such as in Section 4.

Note that for Whitney matroids $\delta_M(m) = m$.

2.5. The Probability an Instance is a Whitney Matroid

Let us define for random matroid M and $h \geq 0$,

$$\delta_M(h) = \Pr\{M_0 \text{ is a Whitney matroid of rank } h \mid M_0 \text{ is an instance of } M\}$$

It is easy to establish a rough lower bound for $\delta_M(h)$, given $M = (E, \mathcal{I}, p)$ is uniform. Let $\mathcal{I}_h = \{I \mid I \in \mathcal{I} \text{ and } |I| = h\}$.

Proposition 4.1

$$\delta_M(h) \geq |\mathcal{I}_h| \cdot p^h \cdot (1-p)^{|E|-h}$$

For proof, note that for each $E_0 \in \mathcal{I}_h$, $M_0 = (E_0, \{I \in \mathcal{I} \mid I \subseteq E_0\})$ is an instance of M of measure $p^h(1-p)^{|E|-h}$ and M_0 is a Whitney matroid.

2.6. Limiting-Whitney Matroids

Let $M^{(1)}, M^{(2)}, \ldots$ be a sequence of proper random matroids such that $\Pr\{\exists e \in I' - I \text{ s.t. } I + e \text{ independent} \mid I, I' \text{ are independent sets of an instance of } M^{(\ell)} \text{ with } |I'| > |I|\} \to 1$ as $\ell \to \infty$. Such a sequence is called *limiting-Whitney*.

Proposition 2.2

As $\ell \to \infty$,

$\Pr\{I \text{ is maximum} \mid I \text{ is maximal in an instance of } M^{(\ell)}\} \to 1$ and

$\Pr\{I + e \text{ has a unique circuit} \mid I \text{ is maximal in an instance of } M^{(\ell)} \text{ and } e \in E - I\} \to 1$.

These are easy extensions of known results (see Lawler [1977]) for Whitney matroids.

2.7. Intersections of Random Matroids

Let $M^{(1)}, M^{(2)}$ be random matroids with

$M^{(1)} = (E, \mathcal{I}^{(1)}, \{p_e^{(1)}\})$ and

$M^{(2)} = (E, \mathcal{I}^{(2)}, \{p_e^{(2)}\})$. We wish to consider independent sets in both $\mathcal{I}^{(1)}$ and $\mathcal{I}^{(2)}$.

Let $M^{(1)} \wedge M^{(2)}$ be the structure

$M = (E, \mathcal{I}^{(1)} \wedge \mathcal{I}^{(2)}, \{p_e^{(1)} \cdot p_e^{(2)}\})$

It is not difficult to show (by definition of proper matroids)

Proposition 2.3

$M = M^{(1)} \wedge M^{(2)}$ is a proper random matroid if $M^{(1)}$ and $M^{(2)}$ are proper random matroids.

There is no known result relating the complexity of constructing maximum independent sets in random instances of $M^{(1)}$, $M^{(2)}$ to the complexity of constructing a maximum independent set in random instances of $M^{(1)} \wedge M^{(2)}$. Although in practice, we often have that if the efficient algorithm of Section 4 succeeds with high probability on $M^{(1)}$ and $M^{(2)}$ separately, then it succeeds with high probability on $M^{(1)} \wedge M^{(2)}$.

In contrast, *Whitney matroids are not closed under intersection.* The problem of constructing a maximal independent set in the intersection of k Whitney matroids has a polynomial time (in $|E|$) algorithm [Lawler, 1977] for $k = 2$, but is known to be a NP complete problem for any $k \geq 3$.

2.8. Weighted Random Matroids

We now extend our definition of random matroids so that the elements are independently, randomly weighted over given probability distributions. We wish upper and lower bounds on the weight of the maximum independent set. Lueker [1978] considers this problem for graphs with a normal distribution of edge weights and we show his results extend to weighted random matroids with arbitrary uniform distributions.

A *weighted random matroid* M is a triple $(E, \mathcal{I}, \{W_e\})$ where E is a set of elements, $\mathcal{I} \subseteq 2^E$, and for each $e \in E$, W_e is an independent random variable.

An *instance of* M is $M_0 = (E, \mathcal{I}, \{w_e\})$ where the w_e are instances of the W_e for each $e \in E$. M is *uniform* if the W_e have the same distribution.

For all $I \in \mathcal{I}$, let $w(I) = \sum_{e \in I} w_e$. Let $W_{max}(M)$ be the random variable

$$\max\{w(I) \mid I \in \mathcal{I}_{max}\},$$

where $\mathcal{I}_{max}$ is the set of maximum elements of $\mathcal{I}$. Let h_0 = size of maximum elements of $\mathcal{I}$.

Proposition 2.4

$\overline{W_{max}(M)} \leq$ the mean of $w(I)$ over all $I \in \mathcal{I}_{max}$ and instances of M such that $w(I) = W_{max}(M)$. For example, if the $\{W_e\}$ are all normal with mean μ and variance σ^2, then $\overline{W_{max}(M)} \leq h_0 \mu + \sigma\sqrt{2h_0 \log|\mathcal{I}_{max}|}$ as $|E| \to \infty$.

Let $M = (E, \mathcal{I}, W)$ be a uniform weighted random matroid and let F be the probability distribution of W and choose some $p \in (0,1)$. For any instance $M_0 = (E_0, \mathcal{I}_0, w)$ of M, let $M_0' = (E_0', \mathcal{I}_0')$ be derived from M_0 by deleting each element $e \in E_0$ with $w_e < F^{-1}(1-p)$ and let $\mathcal{I}_0' = \{I + \mathcal{I}_0 \mid I \subseteq E_0'\}$. Note that instances of $M' = (E, \mathcal{I}, p)$ have the same measure as the corresponding M_0' instances. Thus,

Proposition 2.5

$\overline{W_{max}(M)} \geq |\mathcal{I}_{max}| F^{-1}(1-p)$ if the mean $\overline{W_{max}(M) \mid \text{rank}(M') < h_0}$ is $o(|\mathcal{I}_{max}| F^{-1}(1-p))$ as $|E| \to \infty$.

Note that if the restriction of Proposition 2.5 is satisfied, we have an algorithm which with high likelihood (as $|E| \to \infty$) constructs an independent set with weight $\geq |\mathcal{I}_{max}|F^{-1}(1-p)$ in an instance of M. This idea has been used by Walkup [1977] for discrete distributions of W and by Lueker [1978] for W with normal distributions.

For example, assume W is normal with mean μ and variance σ^2, and $q = \Pr\{\text{rank}(M') = h_o\}$. Then if $q\sqrt{-h_o \log q} = o(h_o\sqrt{-\log q})$, then the mean of $W_{max}(M)$ is $\geq h_o\mu + h_o\sigma\sqrt{-2 \log p}$

3. A General Nonconstructive Existence Theorem

Let $M^{(1)}, M^{(2)}, \ldots$ be a sequence of uniform random matroids. For each $\ell = 1,2,\ldots$ let $M^{(\ell)} = (E^{(\ell)}, \mathcal{I}^{(\ell)}, p)$. Let

$$\mathcal{I}_h^{(\ell)} = \{I \in \mathcal{I}^{(\ell)} \mid |I| = h\} \quad \text{for } h \geq 1.$$

Let the *interdependence ratio for* $M^{(\ell)}$ be

$IR_h^{(\ell)}$ = the mean of $\dfrac{\Pr\{I \text{ Independent} \mid I' \text{ Independent}\}}{\Pr\{I \text{ Independent}\}}$ for $I,I' \in \mathcal{I}_h^{(\ell)}$.

For a fixed $h > 0$, we are interested in a minimum p (the *critical* p) such that as $\ell \to \infty$,

$$\Pr\{\text{rank}(M^{(\ell)}) \geq h\} \to 1$$

or equivalently,

$\Pr\{\exists$ independent set of size h in any instance of $M\} \to 1$.

The following is a generalization of nonconstructive proof technique due to Erdös and Renyi.

Theorem 3.1 If for $\ell \to \infty$, $IR_h^{(\ell)} = 1 + o(1)$ then the critical p is lower bounded by

$$|\mathcal{I}_h^{(\ell)}|^{-1/h} \quad \text{for } |\mathcal{I}_h^{(\ell)}| > 0$$

Proof. Let X be the random variable giving 1 in the event $\text{rank}(M_o^{(\ell)}) = h$ and else $X = 0$, for each instance $M_o^{(\ell)}$ of $M^{(\ell)}$. Then the mean of X is

$$\bar{X} = |\mathcal{I}_h^{(\ell)}|p^h \geq 1 \quad \text{for the given } p$$

The variance of X is

$$\begin{aligned} \text{VAR}[X] &= \bar{X}^2(IR_h^{(\ell)} - 1) \\ &= \bar{X}^2 o(1) \quad \text{as } \ell \to \infty. \end{aligned}$$

By the Chebyshev Inequality,

$$\Pr(X=0) \leq \frac{\text{VAR}[X]}{\bar{X}^2} = o(1) \quad \text{as } \ell \to \infty. \qquad \square$$

In practice (for example, perfect matchings, hamiltonian lines, and cliques in random graphs), the bound $p \geq |\mathcal{I}_h^{(\ell)}|^{-1/h}$ is not sufficient to guarantee $IR_h^{(\ell)} = 1 + o(1)$ as $\ell \to \infty$. To compute $IR_h^{(\ell)}$, we introduce a new random variable $u = |I \wedge I'|$ for randomly chosen $I,I' \in \mathcal{I}_h^{(\ell)}$. Then

$$\begin{aligned} IR_h^{(\ell)} &= \left(\frac{p^{h-u}}{p^h}\right) \\ &= p^{-u} \\ &= \sum_{k=0}^{h} p^{-k}\Pr\{u = k\} \end{aligned}$$

Thus, we must choose p to satisfy also

$$\sum_{k=0}^{h} p^{-k}\Pr\{u = k\} = 1 + o(1)$$

Now we consider random matroids $M = (E^{(\ell)}, \mathcal{I}^{(\ell)}, p)$ for various properties of random (undirected) graphs. $G_{n,p}$ with n vertices V and $\ell = \frac{n(n-1)}{2}$ possible edges $E^{(\ell)} = \{\{u,v\} \mid u,v \in V\}$ each choose with probability p. For *cliques* of v vertices and $h = \frac{v(v-1)}{2}$ edges

$$|\mathcal{I}_h^{(\ell)}| = \binom{n}{v}$$

and the critical p is 1/2 for $h = 2 \log n$. For *perfect matchings* of h edges

$$|\mathcal{I}_h^{(\ell)}| = \binom{n}{h}\binom{n-2h}{h} h!$$

The critical p is $O\left(\frac{\log n}{n}\right)$.

For a Hamiltonian path of h edges,

$$|\mathcal{I}_n^{(\ell)}| = h!\binom{n}{h+1}$$

and the critical p is $O\left(\frac{\log n}{n}\right)$ for $h = n-1$.

The critical p for cliques and perfect matchings [Erdös and Renyi, 1959] can be derived directly from Theorem 3.1 (higher order terms can also be derived). The critical p for Hamiltonian paths was derived by Posa [1976]. Interestingly, his derivation of the critical p for Hamiltonian paths is essentially by a constructive technique generalized in Section 4 and Theorem 3.1 does not seem applicable in this case. On the other hand, there is no known efficient (polynomial time) algorithm for constructing cliques of size 2 log n with probability 1 when the edge density is the critical $p = 1/2$.

Next we describe a nonconstructive existence proof technique for weighted random matroids $M^{(1)}, M^{(2)}, \ldots$, where $M^{(\ell)} = (E^{(\ell)}, \mathcal{I}^{(\ell)}, \{W_e\})$. Let $\mathcal{I}_{max}^{(\ell)}$ be the sets of $\mathcal{I}^{(\ell)}$ of maximum cardinality and let $W_k(M^{(\ell)})$ be the random variable:

$$W_k(M^{(\ell)}) = 1 \quad \text{if } \left(\exists I \in \mathcal{I}_{max} \text{ such that } w(I) \geq k \text{ is an instance of } M \text{ with weighting } w;\right)$$

$$= 0 \quad \text{else.}$$

Let the *weight interdependence ratio* be

$WIR_k^{(\ell)}$ = the mean of $\frac{Pr\{W(I) = k | W(I') = k\}}{Pr\{W(I) = k\}}$ for $I, I' \in \mathcal{J}_{max}^{(\ell)}$.

Then the *mean* of $W_k(M^{(\ell)})$ is $\overline{W_k(M^{(\ell)})} = |\mathcal{J}_{max}^{(\ell)}| \cdot Pr\{I \in \mathcal{J}_{max}^{(\ell)}$ has weight $W(I) \geq k\}$

The *variance* of $W_k(M^{(\ell)})$ is

$$VAR[W_k(M^{(\ell)})] = \overline{W_k(M^{(\ell)})}^2 \, (WIR_k^{(\ell)} - 1)$$

Again, by the Chebyshev Inequality,

$$Pr\{W_k(M^{(\ell)}) = 0\} < \frac{VAR[W_k(M^{(\ell)})]}{\overline{W_k(M^{(\ell)})}^2}$$

$$< WIR_k^{(\ell)} - 1$$

So if $\overline{W_k(M^{(\ell)})} \geq 1$ and $WIR_k^{(\ell)} \to 1 + o(1)$ for $\ell \to \infty$ then

$$Pr\{W_k(M^{(\ell)}) = 1\} > 0 \text{ as } \ell \to \infty, \text{ (or equivalently}$$

$$Pr\{\exists I \in \mathcal{J}_{max}^{(\ell)} \text{ with weight } w(I) \geq k\} \to 1.)$$

By the Limit Theorem, we have:

Theorem 3.2. If M is uniform (as so the element weights have uniform probability distribution with mean μ and variance σ) and $\mathcal{J}_{max}^{(\ell)}$ contains maximum sets of size h_0, and

$$k \leq N^{-1}_{h\mu, h\sigma}(|\mathcal{J}_{max}^{(\ell)}|^{-1} - 1) \text{ and } WIR_k^{(\ell)} \to 1 + o(1)$$

as $\ell \to \infty$ then $Pr\{\exists I \in \mathcal{J}_{max}^{(\ell)}$ with weight $w(I) \geq k\} \to 1$ where $N_{h\mu, h\sigma}$ is the normal distribution function with mean $h\mu$ and variance $(h\sigma)^2$.

4. Analysis of an Extension-Rotation Algorithm for Constructing Independent Sets

In part 4.1 of this section we describe an efficient algorithm for constructing an independent set of fixed size from an instance of a random matroid.

This extension-rotation algorithm is a generalization of random graph algorithms which have appeared in [Posa, 1976], [Karp, 1976], and [Angluin and Valiant, 1979]. In parts 4.2 to 4.5 of this section we develop a *general method of analysis* of the extension-rotation algorithm which provides:

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

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

4.1 The Random Extension-Rotation Algorithm

Let $M_o = (E_o, \mathcal{J}_o)$ has an instance of uniform random matroid $M = (E, \mathcal{J}, p)$. We wish to construct an independent set of size $h_o > 0$.

For any independent set $I \in \mathcal{J}_o$, let

$$\mathcal{E}(I) = \{e \in E_o \mid I \cup \{e\} \in \mathcal{J}_o\}.$$

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 $I \in \mathcal{J}_o$, let

$$\mathcal{R}(I) = \{e \in \mathcal{J}_o \mid I \cup \{e\} \notin \mathcal{J}_o \text{ but } \exists e' \in I \text{ with } I \cup \{e\} - \{e'\} \in \mathcal{J}_o\}.$$

If $\mathcal{R}(I) \neq \emptyset$, we may *rotate* I by choosing an $e \leftarrow \mathcal{R}(I)$ and some appropriate $e' \in I$ and substituting $I \cup \{e\} - \{e'\} \in \mathcal{J}_o$ 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.

Algorithm E-R

INPUT: An instance $M_o = (E_o, \mathcal{J}_o)$ of random matroid $M = (E, \mathcal{J}, \{p_e\})$ and integer $h_o \geq 0$.

```
INITIALIZATION: I ← ∅;  T ← 0
WHILE |I| < h_o 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_o
                 ROTATE:  I ← I ∪ {e} - {e'}
              END
      T ← T + 1
      E_T ← E_{T-1} - {e}
   END
RETURN (I)
```

We define the sets:

$$\mathcal{E}_T(I) = \{e \in E_T \mid I \cup \{e\} \in \mathcal{J}_o\}$$

$$\mathcal{R}_T(I) = \{e \in E_T \mid I \cup \{e\} \notin \mathcal{J}_o \text{ but } \exists e' \in I \text{ with } I \cup \{e\} - \{e'\} \in \mathcal{J}_o\}.$$

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

4.2 Parameters of the Algorithm E-R

We wish to analyze the algorithm relative to the "*time*" index T. (Note that each of the "unit time" steps 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 $I \cup \{e\}-\{e'\} \in \mathscr{I}_o$

Of course in applications (see Section 5) or 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 $T_o = 0$ and let $T_h = |E_o|$ for $h = H+1,\ldots,h_o$. 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_o of the random matroid M.

Fix some constant $\alpha > 1$. For each $t = 0,1,\ldots, E$ let $\epsilon_t(h)$, $\hat{\epsilon}_t(h)$, $\lambda_t(h)$, $\hat{\lambda}_t(h)$ be functions of range $0 \le h \le h_o$ and domain $[0,1]$. We require that for a class $\mathscr{A}_o$ of executions of the Algorithm E-R with total measure $\ge 1-|E|^{-\alpha}$,

(i) $\epsilon_t(|I|) <$ Pr{ extension of I on step t | $\mathscr{E}_t(I) \cup \mathscr{R}_t(I) \neq \emptyset$ and given an execution in $\mathscr{A}_o$} $\le \hat{\epsilon}_t(|I|)$

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

Also let $\rho_t(h) = (1 - \lambda_t(h)) \cdot (1 - \hat{\epsilon}_t(h))$ and

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

Note that $\rho_t(h)$, $\hat{\rho}_t(h)$ are functions such that except for executions of Algorithm E-R with total measure $\le |E|^{-\alpha}$, $\rho_t(|I|) \le$ Pr{rotation of I on step t} $\le \hat{\rho}_t(|I|)$.

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

"extension of I on step t" <=> "$T_{h+1} = t + 1$"

"rotation of I on step t" <=> "$T_{h+1} > t + 1$"

"$\mathscr{E}_t(I) \cup \mathscr{R}_t(I) = \emptyset$" <=> "$T_h = |E_o|$."

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

$$\epsilon_t(h) = \lambda_t(h) = 0, \quad \hat{\epsilon}_t(h) = \hat{\lambda}_t(h) = 1$$

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

$$|\hat{\epsilon}_t(h) - \epsilon_t(h)| \quad \text{and} \quad |\hat{\lambda}_t(h) - \lambda(h)|$$

to be minimal, so that the analysis techniques of this section yield tight bounds on the time complexity of Algorithm E-R. In our graph applications tight $\epsilon_t(h)$, $\hat{\epsilon}_t(h)$, $\lambda_t(h)$, $\hat{\lambda}_t(h)$ can be obtained by techniques similar to the "equiprobability" and "almost equiprobability" lemmas of [Angluin and Valiant, 1977].

All our applications of Section 5 satisfy the following *monotonicity restrictions*:

R1 $\epsilon_t(h)$, $\hat{\epsilon}_t(h)$ 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_o are exhausted.)

4.3 Sufficient Conditions for Success with High Likelihood

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

$$\Pr\{H = h_o\} \ge 1 - |E|^{-\alpha_o}.$$

(i.e., Algorithm E-R succeeds in constructing an independent set of size h_o with probability $\ge - |E|^{-\alpha_o}$)

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

$$EXT_h = \Pr\{H > h \mid H \ge h,\ t = T_h,\ t' = T_{h+1} - 1$$

and given an execution in $\mathscr{A}_o\}$

Proposition 4.1

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

$$\le \hat{\epsilon}_{t'}(h) \cdot (1 - \lambda_t(h)) \cdot \left[\frac{1 - \hat{\rho}_t(h)^{|E_o| - t + 1}}{1 - \hat{\rho}_t(h)} \right]$$

Unfortunately, we found that a direct derivation of $\Pr\{H = h_o\}$ by use of Proposition 4.1 is intractable, because of the stubborn appearance of the random variables T_h in the conditional probabilities. (Thus Proposition 4.1, as stated, is never used in our analysis of Algorithm E-R.)

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

Lemma 4.1

If M is a uniform random matroid $(E, \mathcal{I}, p)$ (so the elements of E_o are chosen from E with fixed probability p), then

$$\Pr\{p|E|(1-\beta) \le |E_o| \le p|E|(1+\beta)\} \le |E|^{-\alpha}$$

where $\beta = \sqrt{\dfrac{6\,\alpha \log|E|}{p\,|E|}}$.

proof follows from the Chernoff bounds:

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

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

The following two conditions in conjunction imply $\Pr\{H = h_o\} \ge 1 - (1 + c_o)|E|^{-\alpha}$

C1 For some fixed $t_0, t_1, \ldots, t_{h_o}$

$\lambda_t(h) = \hat{\lambda}_t(h) = 0$ for $0 \le h \le h_o$ and $0 \le t \le t_h$,

C2 $\Pr\{T_{h_o} \le t_{h_o} < |E_o|\} \ge 1 - c_o|E|^{-\alpha}$,

for some $c_o > 1$.

Note that C1 does not suffice to imply anything about $\Pr\{H = h_o\}$ since we may frequently fail if the "time index" t exceeds t_h. We now *assume that condition C1 has been verified for some* $t_0, t_1, \ldots, t_h$ and derive bounds on the critical p which insures condition C2 is satisfied.

4.4 Verification of Condition C2

To verify C2, we require upper and lower bounds on the distribution of steps between extensions

We assume here R1, R2, and C1. Let $g(x,q) = q(1-q)^x$ be the *geometric density function*. Let $\mathcal{A}_o$ be the class of executions of Algorithm E - R with measure $1 - |E|^{-\alpha}$, which were used in the definition of the $\epsilon_t(h)$.

Also, let S be the condition:

"$T_{h+1} \le t_h$, $t = T_h < |E_o|$ and given an execution in $\mathcal{A}_o$"

Lemma 4.2

$$\frac{\epsilon_{t+x}(h)}{\hat{\epsilon}_{t_h}(h)}\, g(x,\hat{\epsilon}_{t_h}(h)) \le \Pr\{T_{h+1} - T_h = x+1 \mid S\} \le \frac{\hat{\epsilon}_{t+x}(h)}{\epsilon_t(h)}\, g(x,\epsilon_t(h))$$

Proof By conditions C1 and monotonicity restriction R1,

$$\hat{\rho}_t(h) = (1 - \epsilon_t(h)) \le (1 - \epsilon_{T_h}(h))$$

for $0 \le h \le h_o$ and $T_h \le t \le t_h$.

$$\begin{aligned}\Pr\{T_{h+1} - T_h = x+1 \mid S\} &\le \hat{\epsilon}_{t+x}(h) \prod_{k=t}^{t+x-1} \rho_k(h) \\ &\le \hat{\epsilon}_{t+x}(h)\,(1 - \epsilon_t(h))^x \\ &\le \left\lceil \frac{\hat{\epsilon}_{t+x}(h)}{\epsilon_t(h)} \right\rceil \epsilon_t(h)(1 - \epsilon_t(h))^x\end{aligned}$$

The lower bound derivation is similar. □

We now derive bounds on the steps between extensions. For $h = 0, \ldots, h_o$ and $t = 0, \ldots, t_h$ let $\delta_t(h) = \mathrm{MAX}(h,$

$$\left\lceil \frac{\log\left[(1-\epsilon_{t_h}(h))^{t_h+1} + \dfrac{\epsilon_{t_h}(h)(1-|E|^{-\alpha})}{\hat{\epsilon}_t(h)}\right]}{\log(1-\epsilon_{t_h}(h))} \right\rceil - 1)$$

and let $\hat{\delta}_t(h) = \log\left[1 - \dfrac{\hat{\epsilon}_t(h)(1-|E|^{-\alpha})}{\epsilon_t(h)}\right] \Big/ \log(1-\hat{\epsilon}_t(h))$

Lemma 4.3

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

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

It suffices to verify:

$$\Pr\{T_{h+1} - T_h \le \hat{\delta}(h) \mid S\}$$

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

$$\ge \frac{\epsilon_{t+\hat{\delta}(h)-1}(h)}{\hat{\epsilon}_t(h)} \sum_{x=0}^{\hat{\delta}(h)-1} \epsilon_t(h)(1-\hat{\epsilon}_t(h))^x$$

by Lemma 4.2

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

$$\ge \frac{\epsilon_t(h)}{\hat{\epsilon}_t(h)} \left[1 - (1 - \hat{\epsilon}_t(h))^{\hat{\delta}(h)}\right] \quad \text{by R1}$$

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

Similarly, we can show:

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

As a consequence of Lemma 4.3, we may use for $1 \le h \le h_o$

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

and

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

to bound the time complexity of Algorithm with high probability, when $\Delta(0) = \hat{\Delta}(0) = 0$.

Let M be *uniform* with element density p(i.e., each element is chosen with probability p) and assume R1, R2, and C1. Let

$$B = p|E| \left(1+\sqrt{6 \alpha \log|E|/p|E|}\right.$$

Theorem 4.1 If $\Delta(h) \le t_h$ then

$$\Pr\{\Delta(h) \le T_h \le \hat{\Delta}(h)\} \ge 1 - a(h)|E|^{-\alpha}$$

where

$$a(h) = 3h(1+r)+1 \text{ with}$$

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

Proof By Lemma 4.1,

$$\Pr\{|E_o| > B\} < |E|^{-\alpha}.$$

By Lemma 4.3,

$$\Pr\{\Delta(h) \le T_h \le \hat{\Delta}(h) \mid T_h \le t_h\} \ge 1-3h|E|^{-\alpha}$$

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

Thus $\Pr\{T_h > t_o \mid |E_o| \le B\}$

$$< \Pr\{T_h < \Delta(h) \text{ or } \hat{\Delta}(h) < T_h \mid T_h \le t_h\} \cdot r$$

$$= 3hr|E|^{-\alpha}$$

But $\Pr\{T_h > t_h\} < \Pr\{T_h > t_h \mid |E_o| \le B\} + |E|^{-\alpha}$

$$< (3hr + 1)|E|^{-\alpha}$$

So

$$\Pr\{T_h < \Delta(h) \text{ or } \hat{\Delta}(h) < T_h\}$$
$$< \Pr\{T_h < \Delta(h) \text{ or } \hat{\Delta}(h) < T_h \mid T_h \le t_h\}$$
$$+ \Pr\{T_h > t_h\}$$
$$< a(h)|E|^{-\alpha}. \quad \square$$

Note that Theorem 4.1 may be restated:

If $\hat{\Delta}(h) < t_h$ then

$$\Pr\{H \ge h\} \ge 1 - |E|^{-\alpha(h)}$$

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

Further more, if we wish

$$\Pr\{H \ge h_o\} \ge 1 - |E|^{-\alpha_o}$$

for any *given* α_o sufficiently large then we find a minimal $p_o \in (0,1)$ such that the restrictions of Theorem 4.1 are satisfied and $\alpha_o = \alpha(h)$. (Note that if $M = (E,\mathcal{J},p)$ *proper* random matroid and $(E,\mathcal{J})$ has rank $\ge h_o$, then such a p_o always exists.)

4.5 Bounds on the Probability Density Function of T_h

We assume here the restrictions given in Theorem 4.1. Actually, we have a much more general result, since we have from Lemma 4.2 bounds on the probability density function of $T_{h+1} - T_h$ for $h = 1,\ldots,h_o - 1$. By the monotonicity restrictions R1, for $x = 0,\ldots,|E|$

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

$$\leq \Pr\{T_{h+1} - T_h = x + 1 \mid \Delta(h) \leq T_h \leq \hat{\Delta}(h),\ \Delta(h+1) \leq T_{h+1} \leq \hat{\Delta}(h+1)\}$$

$$\leq \hat{\epsilon}_{\hat{\Delta}(h+1)-1}(h)(1 - q(h))^x$$

where $q(h) = \epsilon_{\Delta(h)}(h)$, $\hat{q}(h) = \hat{\epsilon}_{\hat{\Delta}(h)}(h)$.

Corollary 4.1 for $h = 0,\ldots,h_o-1$

$$\frac{\epsilon_{\Delta(h+1)-1}(h)}{\hat{q}(h)}\, g(x, \hat{q}(h)) - |E|^{-\alpha(h+1)}$$

$$\leq \Pr\{T_{h+1} - T_h = x + 1\}$$

$$\leq \frac{\hat{\epsilon}_{\hat{\Delta}(h+1)-1}(h)}{q(h)}\, g(x,q(h)) + |E|^{-\alpha(h+1)}$$

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 4.1, we have upper and lower bounds on the probability density function of the sum:

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

Theorem 4.2 for $h = 0,\ldots,h_o-1$

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

where

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

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

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

$$\hat{w}_h = \left(\prod_{k=o}^{h-1} \frac{\hat{\epsilon}_{\hat{\Delta}(k+1)-1}(k)}{q(k)} \right)$$

Thus, if the restrictions of Theorem 4.1 are satisfied (as they do in our Applications in Section 5, we can derive by routine methods the mean, variance, and in general any moment of the time cost of Algorithm E-R.

5. Applications

In the previous section, we provided a general method of analysis of random extension rotation algorithms. We discuss here applications to various combinatorial problems for random graph model $G_{n,p}$.

5.1 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.

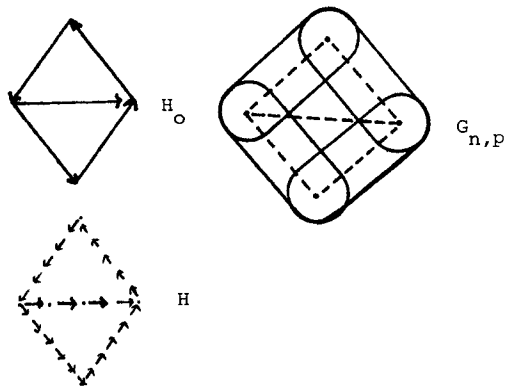
We can further extend the Posa-Karp Algorithm to the problem of identifying certain classes of isomorphic subgraphs. Consider the problem for a fixed graph H and random graph $G_{n,p}$:

Is H isomorphic 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., K. Kuratowski's [1971] forbidden subgraphs for planar graphs, Glover and Hyneke's [1975] forbidden subgraphs for graphs imbedded onto the projective plane, Lekkerkerker and Boland'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 forbidden subgraph problems or random graphs can be efficiently reduced to the problem of determining a Hamiltonian path. Suppose H is a fixed graph of size O(h) but homeomorphic to a graph H_o of size O(n) for some constant $k > 0$. We wish to determine a subgraph G' of an instance of random graph $G_{n,p}$ such that G' is isomorphic to H.

Let k be the number of edges of H_0. By assumption, k is constant. We partition the edges of $G_{n,p}$ into k blocks of size $n/k + 0(1)$, with each block corresponding to an edge of H_o. Choose these blocks t_o so that they have a unique "joining vertex" in common just in the case the corresponding edges of H_o 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. A slight modification of the Posa-Karp algorithm then yields the required Hamiltonian paths with probability $\geq 1-n^{-\alpha}$ for any sufficiently large $\alpha > 1$.



5.2 Analysis of the Posa-Karp Algorithm

We now give a detailed analysis of the Posa-Karp algorithm for delecting a Hamiltonian path in a random graph $G_{n,p}$. We follow the analysis techniques developed in Section 4.

Step A Formulation as a random matroid

First, we formulate the "simple path" (we require that simple paths contain no cycles, but allow empty simple paths) property for random graph $G_{n,p}$ as a random matroid, just as described in Section 2.4. Let V be a set of n vertices. Let $M = (E,\mathcal{I},p)$ the uniform random matroid:

(i) $E = \{\{u,v\} \mid \text{distinct } u,v, \in V\}$

(ii) $\mathcal{I}$ = all simple (acyclic) paths in the complete graph (V,E).

Instances of M correspond to instances of the random graph $G_{n,p}$ as described in Section 2.3. Fix an instance

$$M_o = (E_o,\mathcal{I}_o,p) \text{ of } M.$$

Then (V,E_o) has the same measure in $G_{n,p}$ as in M, and $\mathcal{I}_o$ is the set of all simple paths in (V,E_o).

We wish to determine a maximum independent set $I \in \mathcal{I}_o$ of cardinality $h_o = n - 1$, i.e. a Hamiltonian path.

To describe *extension* of a non-maximal simple path $I \in \mathcal{I}_o$, we let $V(I)$ be the vertices of I and let $ENDS(I)$ be the set of extremal vertices of I (i.e. the first and last vertices, if they exist).

Then the extension set is

$$\mathcal{E}(I) = \{e \in E_o - I \mid e = \{u,v\}$$
$$u \in ENDS(I),\ v \in V - V(I)\}$$

The rotation set is

$$\mathcal{R}(I) = \{e \in E - I - \mathcal{E}(I) \mid e = \{u,v\}$$
$$u \in ENDS(I),\ v \in V(I) - ENDS(I)\}$$

Step B Derivation of the bounding parameters:

$$\epsilon_t(h),\ \hat{\epsilon}_t(h),\ \lambda_t(h),\ \hat{\lambda}_t(h)$$

By Lemma 4.1,

$$Pr\{\, p\ell(1-\beta) \leq |E_o| \leq p\,\ell(1+\beta)\}$$

when $\beta = \sqrt{6\,\alpha \log \ell/p\ell}$,

$$\ell = |E| = \frac{n(n-1)}{2},$$

and $\alpha > 1$ is an arbitrary constant.

We can show using symmetry arguments as in Angluin and Valiant [1979] that there exists a class of executions $\mathcal{A}_o$ of the algorithm E-R of total measure $1 - |E|^{-\alpha}$ such that for $t = 0,\ldots,|E|$ $b_h p(n-h)(1-\beta) \leq |\mathcal{E}_t(h)| \leq b_h p(n-h)(1+\beta)$ and $b_h ph(1-\beta) - t/h \leq |\mathcal{R}_t(h)| \leq b_h ph(1+\beta) - t/h$ for $h = 0,1,\ldots,h_o-1$, where $b_h = 1$ if $h = 0$ and else $b_h = 2$ (for the executions in $\mathcal{A}_o$ we also have $|E_o| \geq p\ell(1-\beta)$, by by appealing to Lemma 4.1.)

Therefore we may define as bounds on the conditional extension probability:

$$\epsilon_t(h) = \frac{(n-h)(1-\beta)}{n(1+\beta) - \frac{t}{phb_h}}$$

$$\hat{\epsilon}_t(h) = \frac{(n-h)(1+\beta)}{n(1-\beta) - \frac{t}{phb_h}}$$

so that $\epsilon_t(h) \leq \dfrac{|\mathcal{E}_t(h)|}{|\mathcal{E}_t(h)| + |\mathcal{R}_t(h)|} \leq \hat{\epsilon}_t(h)$

for executions in $\mathcal{A}_o$ of the Algorithm E-R.

Observe that $\frac{\partial \epsilon_t(h)}{\partial t} > 0$, $\frac{\partial \hat{\epsilon}_t(h)}{\partial t} > 0$, $\frac{\partial \epsilon_t(h)}{\partial h} < 0$ and $\frac{\partial \hat{\epsilon}_t(h)}{\partial h} < 0$ so the monotonicity condition R1 is satisfied.

Restriction R2 can be readily verified for it is obvious that

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

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

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

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

Step C Verification of C2

We now must verify condition C2 to insure the algorithm succeeds with high probability. For simplicity, we proceed with an *asymptotic analysis as* $n \to \infty$ (although the techniques of Section 4 allow analysis for any fixed n as well).

Note that as $n \to \infty$, $\beta \to 0$ so

$$\epsilon_t(h) \sim \hat{\epsilon}_t(h) \sim \frac{n-h}{n - \frac{t}{ph\, b_h}}$$

so in the asymptotic case the bounding parameters are identical.

Also, $\hat{\delta}_t(h) \sim \frac{\alpha \log \ell}{\log(1-\epsilon_t(h))}$ as $n \to \infty$

We must determine

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

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

We can show by induction or h that $\hat{\Delta}(h) \le k_2 h \log n$ where $k_2 = \frac{2\alpha k_1}{2\alpha + k_1}$. Thus for $k_1 > \frac{2\alpha}{2\alpha-1}$ we have: $\hat{\Delta} \le t_{h_o}$ and we conclude the Algorithm E-R outputs a Hamiltonian path with probability $\ge 1 - |E|^{-\alpha_o}$ where $\alpha_o < \alpha - \frac{1}{2}$.

Step D Bounds on the Mean and Variance of T_h.

We have from Corollary 4.1 that

$$\Pr\{T_{h+1} - T_h = x+1\} \le s_h q(x, q(h)) + |E|^{-\alpha(h+1)}$$

where $s_h = \frac{\hat{\epsilon}_{\hat{\Delta}(h+1)-1}(h)}{q(h)}$ and $q(h) = \epsilon_{\Delta(h)}(h)$

This requires calculation of the lower bound $\Delta(h)$, which in this application is trivial: $\Delta(h) = h$.

But $s_h \sim c$ where $c = \frac{1}{1 - k_2/k_1}$ is constant for $p = \theta(\frac{\log n}{n})$. Also, for sufficiently large $\alpha(h+1)$, $|E|^{-\alpha(h+1)} \to 0$ as $|E| \to \infty$.

Let X_h be a random variable with geometric distribution $q(x, q(h))$. By the Appendix,

$$\bar{X}_h \sim \frac{(1-q(h))}{q(h)} \sim \frac{h}{n-h}$$

So
$$\bar{T}_{h_o} \le \sum_{h=o}^{h_o-1} s_h \bar{X}_h \le c \int_o^{h_o-1} \frac{h}{n-h}\, dh \le cn \log n + O(n) \quad \text{for } h_o = n-1.$$

A similar calculation yields a lower bound on the mean of T_{h_o}:

$$\bar{T}_{h_o} \ge \frac{n \log n}{d^2} + \Omega(n),$$

where $d = \frac{1}{1 - k_2/(2k_1)}$

Note that as $n \to \infty$ these bounds are tight within a factor of cd^2, which we may assume constant if $p = \theta(\frac{\log n}{n})$. Thus, $\bar{T}_{h_o} = \theta(n \log n)$.

We also can apply the formulas of the Appendix to bound the second moment of T_{h_o}:

$$\frac{3}{4}e^{-d}n^3 + \Omega(h) \le \bar{T}^2_{h_o} \le \frac{c^2}{e} n^3 + O(n^2)$$

and so we have tight bounds for the variance

$$\mathrm{VAR}[T_{h_o}] = \bar{T}^2_{h_o} - \bar{T}^2_{h_o} = \theta(n^3).$$

in the case $e^d c^2$ is constant.

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 $\log n$ (for the bounds on the mean) and $(\log n)^2$ (for the bounds on the variance).

5.3. Analysis of an Algorithm for Matching in Random Graphs

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

To formulate the "perfect matching" problem as a random matroid, we assume a complete graph $G=(V,E)$ with n nodes (n is assumed even). Let $M = (E,\mathcal{J}, p)$ when

$$\mathcal{J} = \{E' \subseteq E \mid E' \text{ is a matching}\}.$$

Let $M_o = (E_o,\mathcal{J}_o)$ be an instance of M.

Note that for $I \in \mathcal{J}_o$ the extension set is

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

and

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

Let

$$a(h) = p\binom{2-2h}{2}$$

$$a'(h) = 2ph(n-2h)$$

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

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

Again we may use symmetry arguments and Lemma 4.1 to bound the cardinalities of $\mathcal{E}_t(I)$ and $\mathcal{R}_t(I)$ and $|E_o|$ for a class of executions $\mathcal{A}_o$ with measure $\geq 1-|E|^{-\alpha}$.

For executions in $\mathcal{A}_o$,

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

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

Let $t_h = (1-\beta)(a(h) + a'(h)) - f_t(h) - f_t'(h)$. Then $|\mathcal{E}_t(I)| + |\mathcal{R}_t(I)| > 0$ for $t \leq t_h$ in executions of $\mathcal{A}_o$, verifying condition C1.

We may let

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

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

so we have

$$\epsilon_t(h) \leq \frac{|\mathcal{E}_t(I)|}{|\mathcal{E}_t(I)| + |\mathcal{R}_t(I)|} \leq \hat{\epsilon}_t(h)$$

for $h = |I|$ and executions in $\mathcal{A}_o$.

By taking partial derivatives of $\epsilon_t(h)$ 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: $\epsilon_t(h) \sim \hat{\epsilon}_t(h)$. By the routine calculations, described in Section 4, the reader may verify that $\hat{\Delta}(h_o) \leq t_{h_o}$, where $h_o = \frac{n}{2}$, so the Algorithm E-R outputs a perfect matching with probability $\geq 1-|E|^{-\alpha(h_o)}$. We also leave the reader to calculate tight bounds on the mean and variance of T_h:

$$\bar{T}_{h_o} = \theta(n \log n) \text{ and } \bar{T}^2_{h_o} = \theta(n^3)$$

by applying Corollary 4.1 (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.

Acknowledgments

We wish to thank Andy Langer, Allen Emerson, and Christos Papadimitriou for their helpful suggestions and spirited discussions 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.

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.

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 distrubuted 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

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

The density function of $X_1 + X_2$ is for $0 \le k \le 2n_o$,

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

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

By applying induction, we derive the density function of $Y = \sum_{i=1}^{m} X_i$

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

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

The moments of Y are given by

$$\overline{Y^t} = \sum_{k=o}^{s} k^t (g_1 * \ldots * g_m)(k)$$

where $s = mn_o$.

Mean of Y : $\bar{Y} = \sum_{i=1}^{m} \bar{X}_i$

$$\overline{X_i} = \frac{r_i}{p_i} \left[1 - r_i^{n_o} (n_o p_i + 1) \right]$$

Variance of Y: $\mathrm{VAR}(Y) = \bar{Y}^2 - \bar{Y}^2$

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

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

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

Asymptotic Analysis

Note that as $s \to \infty$

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

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

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