Sizes of Relations: a Dynamic Analysis

Danièle Gardy
Université de Versailles Saint-Quentin

February 7, 1994

[summary by Dominique Gouyou-Beauchamps]

1. Introduction

Among the parameters that can be defined on relational databases, the sizes of the relations, either present in the database or computed by application of a relational operator, have long been recognized as important parameters in query optimization.

The basic objects we[1] consider are *relations*, which are sets of (distinct) tuples. They can be seen as tables: a row represents a tuple, and the number of lines is the number of elements of the relation (its *size*); the columns are called the *attributes*. The *projection* of a relation on a subset of the set of attributes is a new relation, obtained by suppressing the corresponding columns, then all the duplicated rows in the resulting table: We keep only one instance of each tuple. We give in Figure 1 an instance of a relation $R[X,Y]$ and its projection denoted $\pi_X(R)$ on the attribute X.

Let us now consider two relations $R[X,Y]$ and $S[X,Z]$ with a common attribute X. The *equijoin* of R and S on their common attribute X, denoted $R \bowtie S$, has three attributes X, Y and Z; it is composed of all triples (x,y,z) such that (x,y) belongs to R and (x,z) belongs to S (see again Figure 1).

The *semijoin* is another operator on relations; although it can be defined using the projection and equijoin: $R[X,Y] \rhd S[X,Z] = \pi_{XY}(R \bowtie S) = R \bowtie \pi_X(S)$ (Figure 1 presents instances of $R \rhd S$ and $S \rhd R$).

We use the terms *initial relation* for the relations to which we apply a relational algebra operator (projection, equijoin or semijoin), and *derived relation* for the relation resulting from the operation.

We gave in former papers [4, 5] conditions which ensure that, in the static cases (i.e., at a given time), the size of a derived relation, obtained by a projection, an equijoin or a semijoin, follows a normal limiting distribution. Our goal here is to extend these results when the database is submitted to a sequence of queries and updates.

In the static case, we study the conditional distribution of the sizes of the derived relations obtained by a projection or by a join, assuming that the sizes of the initial relations are known. In the dynamic case, we want to study the influence of updates and queries on the sizes of initial relations and derived relations. To this effect, we shall use a modelization in terms of urn models.

2. Urn models and databases

We consider a sequence of d urns, each urn being labelled with a distinct value of the attribute X. To each tuple of the relation R, we associate a ball labelled by the value of the tuple on the column X; this ball falls into the corresponding urn. An equivalent way of seeing this phenomenon

[1] The original articles by D. Gardy and G. Louchard can be found in [6, 7].

R

X	Y
x_0	y_0
x_0	y_1
x_1	y_2
x_2	y_3

S

X	Z
x_0	z_0
x_0	z_1
x_1	z_1
x_3	z_2

$\pi_X(R)$

X
x_0
x_1
x_2

$R \bowtie S$

X	Y	Z
x_0	y_0	z_0
x_0	y_0	z_1
x_0	y_1	z_0
x_0	y_1	z_1
x_1	y_2	z_1

$R \rhd S$

X	Y
x_0	y_0
x_0	y_1
x_1	y_2

$S \rhd R$

X	Z
x_0	z_0
x_0	z_1
x_1	z_1

FIGURE 1. Examples of relations *R[X, Y]* and *S[X,Z]* with the projection $\pi_X(R)$ of R on X, the equijoin of R and S on X and the semijoins of R and S, and S and R.

is to consider instead that we have a finite supply of balls, and that we allocate them at random among the d urns, each trial being independent of the others. Each ball then receives the label of the urn it falls into. After coupling all the tuples of the initial relation R with urns, some urns are empty and some contain at least one ball. The number of urns with at least one ball is exactly the number of tuples in the projection of the relation R.

In the rest of the paper, we shall use indifferently the terms *relation size* and *number of balls* or *number of tuples*, and the terms *projection size* and *number of non-empty urns*.

3. Static models

Here, we consider the case of a *free* relation, i.e., there is total independence between the values taken by different tuples. In this framework, we obtain the following theorems [4, 8, 9].

THEOREM 1 ([4]). *Let $R[X,Y]$ be a free relation with a uniform probability distribution on the domain of attribute X. Then the probability distribution of the size of the projection of R on attribute X, conditioned by the size $r = Ad_X + o(d_X)$ of relation R with A a positive constant is asymptotically normal when $d_X \to \infty$. The asymptotic mean and variance are given by $\mu = \mu_0 d_X$ and $\sigma^2 = \sigma_0^2 d_X$ where μ_0 and σ_0^2 are constants that depend on the probability distribution on attribute Y.*

We consider relations where attribute X is a *key*, i.e., in a given instance of the relation the x-value of a tuple uniquely determines its y-value.

THEOREM 2 ([4]). *Let $R[X,Y]$ be a relation with a key X, and $S[X,Z]$ a relation with a key U. We assume that the probability distribution on D_X (the finite domain on which a probability distribution is defined for X) is uniform. The probability distributions on D_Y and D_Z are arbitrary. The sizes r and s of the relations R and S are assumed to satisfy $r < d_X$, $d_X = o(r^{3/2})$, and $s = Bd_X(1 + o(1))$. Then the probability distribution of the size of the semijoin of R and S on attribute X, conditioned by the sizes of R and S, is asymptotically normal. The mean and variance have for asymptotic values $\mu = (1 - e^{-B})r$ and $\sigma^2 = r((e^B - 1)/e^{2B} - rB/d_X e^{2B})$.*

We denote $p_{i,d}$ the probability that the ith element of the domain is selected when choosing at random an element of a finite domain D of size d. Let $\lambda_R(t) = \prod_{1 \le i \le d}((1 + p_{i,d}t)/(1 + p_{i,d}))$ be the

generating function associated with the probabilities of the set of distinct items for relation R. We assume that $\lambda_R(t) \neq (1+t)/2$.

THEOREM 3 ([9, 8]). *Let $R[X,Y]$ (resp. $S[X,Z]$) be a relation with a key Y (resp. Z). The sizes r and s of the relations R and S are assumed to satisfy $r = Ad_X + o(d_X)$ and $s = Bd_X + o(d_X)$. Let $g_R(y)$ (resp. $g_S(z)$) be the function $g_R(y) = y\frac{\lambda'_R}{\lambda_R}(y)$ (resp. $g_S(z) = z\frac{\lambda'_S}{\lambda_S}(z)$). Constants A and B are such that: $lim_{y\to+\infty} g_R(y) > A$ and $lim_{z\to+\infty} g_S(z) > B$. We assume that the probability distribution on D_X is uniform. Then the probability distribution of the size of the equijoin of R and S on attribute X, is asymptotically normal when $d_X \to \infty$. The mean and variance have for asymptotic values $\mu = \sigma^2 = \frac{rs}{d_X} \approx ABd_X$.*

4. Dynamic models

We shall denote by $p_{\mathcal{I}}$, $p_{\mathcal{D}}$ and $p_{\mathcal{Q}}$ the probability of making an insertion, a deletion or a query. We can choose non-equal probabilities for insertion and deletion, as long as the probability of an insertion is at least equal to the probability of a deletion: $p_{\mathcal{I}} \geq p_{\mathcal{D}}$.

If we choose to perform a deletion, the conditional probability of deleting a given ball is $1/n$, n being the number of balls at this time.

Assuming that the urn size is infinite corresponds, in terms of relational database, to a relation with a key on the attribute suppressed in the projection. As we also want to study relations without keys, we need to extend the models to the case where *the urns have a finite capacity* (there are δ places for balls). If we choose to perform an insertion, we must give the conditional probability of inserting a ball into an urn, and this is the place where the infinite and finite models differ. In the infinite urn model, each urn has the same probability of getting the new ball: $1/d$, with d the number of urns. In the finite urn model, we can view each urn as a collection of δ distinguishable cells, and each empty cell, whatever the urn it belongs to, has the same conditional probability of receiving the ball, knowing that we have chosen to perform an insertion. Thus the probability that we put a ball in urn V_i is $v_i/(d\delta - n)$ where v_i is the number of empty cells in V_i and n is the number of balls at this time.

We denote by $\Rightarrow$ the weak convergence of random function in the space of all right-continuous functions having left limits and endowed with the Skorohod metric (see Billingsley [1]). All convergences will be defined for $n \to +\infty$.

We study two related stochastic processes, describing respectively the number of balls denoted by $\mathcal{P}$, and size of the projection (number of non-empty urns), denoted by $\mathcal{Q}$; we shall show that each of these processes has a deterministic component of order n, and a random component of order $\sqrt{n}$.

Let W be the number of balls at some time. We might choose the current number of steps (number of queries or updates) as a measure for the time, which would then belong to the interval $[0, 2n]$. However, we shall study the asymptotic behaviour of W when the time goes to infinity, and it is interesting to change the time scale by choosing a time nt for $t \in [0,2]$, and to normalize the random variable W. For all the models presented below, the number of tuples W has an expectation and a variance of order n, and we can show that, for a suitable function f_1 related to the type of process, and assuming that we start from an empty structure at time 0:

$$\frac{W([nt]) - nf_1(t)}{\sqrt{n}} \Rightarrow X(t), \qquad 0 \leq t \leq 2,$$

where the process X(t) is a Markovian Gaussian process whose covariance is denoted $f_2(s,t)$, $s \leq t$ [6].

THEOREM 4 ([6]). *The size $S([nt])$ of the projection at time nt is asymptotically a non-Markovian Gaussian process such that*

$$\begin{aligned} S([nt]) &\sim nG(t) + \sqrt{n}X_1(t), \\ E[S([nt])] &\sim nG(t), \\ \mathrm{Var}[S([nt])] &\sim n\Phi(t), \end{aligned}$$

where $X_1(t)$ is a non-Markovian Gaussian process whose covariance is denoted $\Psi_R(s,t)$ and where the functions G, Φ and Ψ_R can be given explicitly and depend on urn models (infinite or bounded) and on functions $f_1(t)$ and $f_2(s,t)$. The relative error in the density is $O(1/\sqrt{n})$.

5. Example

The processes can be divided in two families:

(1) the *weighted structure* in the sense of Flajolet *et al.* [3], Louchard [10], with a possibility function given by $pos(\mathcal{D}) = k$ for a k-size structure (there are k ways of deleting an element in a structure composed from k elements!);
(2) the classical *unweighted structure*.

We study *the unweighted structure family* and we consider updates $(\mathcal{I}+\mathcal{D})$ and queries $(\mathcal{Q})$ with arrival at a relation of size $2n\overline{x} + s\sqrt{n}$ at the time $2n$. We assume that we start from an empty structure at time 0. The mean and variance corresponding to one step are given by

$$\overline{x} = p_{\mathcal{I}} - p_{\mathcal{D}}, \qquad \sigma^2 = p_{\mathcal{I}} + p_{\mathcal{D}} - \overline{x}^2$$

and

$$\frac{W([nt]) - n\overline{x}t}{\sqrt{n}} \Rightarrow \sigma BB(t) + \frac{at}{2},$$

with BB a Brownian Bridge. The expectation and covariance are given by

$$f_1(t) = \overline{x}t + \frac{at}{2\sqrt{n}}, \qquad f_2(s,t) = \sigma^2\frac{s(2-t)}{2}, \qquad s \le t.$$

Now, if we assume that the urns have an infinite capacity, we get for the process *size of the projection*:

$$\begin{aligned} S([nt]) &\sim nG(t) + \sqrt{n}X(t), \\ G(t) &= \alpha(1 - e^{-\overline{x}t/\alpha}) + \frac{a}{\sqrt{n}}te^{-\overline{x}t/\alpha} \end{aligned}$$

where $X(t)$ is a non-Markovian Gaussian process whose covariance is

$$\Psi_R(t_1,t_2) = e^{-\overline{x}(t_1+t_2)/\alpha}\left[\alpha\left(e^{\frac{\overline{x}t_1}{\alpha}\left(\frac{t_1}{t_2}\right)^{p_{\mathcal{D}}/\overline{x}}} - 1\right) - \overline{x}t_1\left(\frac{t_1}{t_2}\right)^{p_{\mathcal{D}}/\overline{x}} + \sigma^2\frac{t_1(2-t_2)}{2}\right],$$

where $\alpha = \frac{d}{n}$.

6. Sketch of the proof

The first step is to study the process $\mathcal{P}$ describing the number of tuples in the initial relation. In the cases we are interested in, $\mathcal{P}$ happens to be a Gaussian process with a deterministic part $\mathcal{P}_0$, on which is superimposed a random part $\mathcal{P}_1$:

$$\mathcal{P} = \mathcal{P}_0 + \mathcal{P}_1.$$

The process $\mathcal{P}_0$ follows a deterministic curve $nf_1(t)$; the function f_1 is the expectation of the number of balls (or tuples in the initial relation), and the process $\mathcal{P}_1$ is a Markovian Gaussian process of order $\sqrt{n}$.

The process $\mathcal{P}$: *number of tuples* determines another process $\mathcal{Q}$: *size of the projection*. Before considering $\mathcal{Q}$, we shall study another process $\mathcal{Q}_0$, defined as the size of the projection of a relation R, when the size of R is given by the process $\mathcal{P}_0$ (which is a first-order approximation of $\mathcal{P}$). To this effect, we define two random variables, say Y_1 and Y_2, which are simply the size of the projection at different times t_1 and t_2. The covariance $\mathrm{Cov}(Y_1, Y_2)$ will allow us to characterize $\mathcal{Q}_0$ as a process composed of a deterministic part $G(t)$ and a random part $\sqrt{n}V(t)$.

We then consider the process $\mathcal{P}$ obtained by superimposing $\mathcal{P}_1$ on $\mathcal{P}_0$. We can again define two random variables *size of the projection* at the times t_1 and t_2; let us call them S_1 and S_2. It is possible to write their covariance as

$$\mathrm{Cov}(S_1, S_2) = \mathrm{Cov}(Y_1, Y_2) + \gamma(t_1)\gamma(t_2)f_2(t_1, t_2)$$

for a suitable function $\gamma(t)$, $f_2(t_1, t_2)$ being the covariance of the process $\mathcal{P}_1$ taken at different times t_1 and t_2. The covariance of Y_1 and Y_2 thus characterizes the "static" part, and the term added to it to get the covariance of S_1 and S_2 comes from the fact that the number of tuples $\mathcal{P}$ is itself a Gaussian process.

Once we have the covariance of the sizes of the derived relation at times t_1 and t_2, the next part is to show that the final process $\mathcal{Q}$ is still asymptotically a Gaussian process. More precisely, we show that $\mathcal{Q}$ has a part $\mathcal{Q}_0$ of order n coming from $\mathcal{P}_0$, on which is added a random part $\mathcal{Q}_1$ of order $\sqrt{n}$ coming from $\mathcal{P}_0$ and from $\mathcal{P}_1$:

$$\mathcal{Q} = \mathcal{Q}_0 + \mathcal{Q}_1.$$

6.1. $\mathrm{Cov}(Y_1, Y_2)$ **for a non-random static structure.** For each urn model (bounded or unbounded), there exist two functions $F(x)$ and $\Psi_{NR}(t_1, t_2)$ such that, if we consider the size of the projection of a relation, itself of size $nf_1(t)$, the asymptotic values of its expectation at time t_1, $E(Y_1)$, and of its covariance at distinct times t_1 and t_2, $\mathrm{Cov}(Y_1, Y_2)$, are:

$$E(Y_1) \sim nF(f_1(t)),$$
$$\mathrm{Cov}(Y_1, Y_2) \sim n\Psi_{NR}(t_1, t_2).$$

6.1.1. Unbounded urns

The one ball survival probability between t_1 and t_2 is denoted by $ps_{1,2}(t_1, t_2)$ ($= 1$ if $t_1 = t_2$).

Then, we obtain with $\mathcal{P}_0$ a deterministic process

$$E(\mathcal{P}_0) = nf_1(t),$$
$$F(X) = \alpha\left(1 - e^{-X/\alpha}\right),$$
$$\Psi_{NR}(t_1, t_2) = \alpha e^{-\frac{f_1(t_1)}{\alpha}}\left(e^{-\frac{f_3(t_1,t_2)}{\alpha}} - e^{-\frac{f_1(t_2)}{\alpha}}\right) - f_1(t_1)ps_{1,2}e^{-\frac{f_1(t_1)+f_1(t_2)}{\alpha}}.$$

6.1.2. Bounded urns

We can view the content of an urn with $\nu(t)$ balls as a population of $\nu(t)$ type 1 (balls) individuals and $\delta - \nu(t)$ type 2 (empty places) individuals. Let

$$p_{i,j}(t_1, t_2) = \Pr[\text{individual of type } i \text{ at time } t_1 \text{ is of type } j \text{ at time } t_2].$$

Then, we obtain:

$$F(X) = \alpha \left(1 - \left(1 - \frac{X}{\beta}\right)^{\sigma}\right),$$

$$\Psi_{NR}(t_1, t_2) = \left(1 - \frac{f_1(t_1)}{\beta}\right)^{\delta} \left[\alpha p_{2,2}^{\delta} - \left(1 - \frac{f_1(t_2)}{\beta}\right)^{\delta - 1} \left(\alpha p_{2,2} + (1 - \frac{1}{\delta}) f_1(t_1) f_7\right)\right],$$

where $\beta = \alpha\delta$.

6.2. $\mathrm{Cov}(S_1, S_2)$. Y_1 and Y_2 denote the size of the projection of a relation R at the times t_1 and t_2, when the number of tuples of R is given by the process $\mathcal{P}_0$. S_1 and S_2 denote the same quantities when the number of tuples of R is given by the process $\mathcal{P}$. We first compute the variation of $\mathrm{Cov}(Y_1, Y_2)$ introduced by assuming that the numbers of tuples are no longer fixed, but Gaussian random variables; this gives $n\Psi_C(t_1, t_2)$. Then we compute the actual covariance of S_1 and S_2 and we show that it is of the type $n\Psi_R(t_1, t_2)$; we also prove that the size of the projection is then a Gaussian process.

As the process $\mathcal{P}$ is obtained by adding a process $\mathcal{P}_1$ of order $\sqrt{n}$ to the process $\mathcal{P}_0$, itself of order n, the number n_1 of balls at time t_1 is given by:

$$n_1 = n \left(f_1(t_1) + \frac{\theta_1}{\sqrt{n}}\right) + O(1),$$

where θ_1 is a Gaussian random variable with mean 0 and covariance $f_2(s, t)$.

Setting $\gamma(t) = F'(f_1(t))$, we obtain:

$$E[Y_1] \sim n \left(F(f_1(t_1)) + \frac{\theta_1}{\sqrt{n}} \gamma(t_1)\right),$$

$$\Psi_C(t_1, t_2) = \Psi_{NR}(t_1, t_2) + \overline{\varphi}_1(t_1, t_2) \frac{\theta_1}{\sqrt{n}} + \overline{\varphi}_2(t_1, t_2) \frac{\theta_2}{\sqrt{n}} + O\left(\frac{1}{n}\right),$$

for some $\overline{\varphi}_1$ and $\overline{\varphi}_2$.

We know from previous work [4] that for a known size of the initial relation R at times t_1 and t_2 (static case), the projection size Y_1 and Y_2 are asymptotically Gaussian. Then for any ξ_1 and ξ_2,

$$E\left[e^{i(\xi_1 Y_1 + \xi_2 Y_2)}\right] \sim E\left(\exp[i(\xi_1 E[Y_1] + \xi_2 E[Y_2]) - \frac{1}{2}(\xi_1^2 \sigma^2(Y_1) + 2\xi_1 \xi_2 \,\mathrm{Cov}(Y_1, Y_2) + \xi_2^2 \sigma^2(Y_2))]\right).$$

Plugging the modified values for $E[Y_1]$ and $E[Y_2]$ into this equation, and substituting $\mathrm{Cov}(Y_1, Y_2)$ by $n\Psi_C(t_1, t_2)$ (and similarly for $\sigma^2(Y_1)$ and $\sigma^2(Y_2)$), we obtain:

$$E\left[e^{i(\xi_1 S_1 + \xi_2 S_2)}\right] \sim e^{A(t_1, t_2)} E\left[e^{B(t_1, t_2)}\right],$$

where the term $B(t_1, t_2)$ contains all the contribution from the Gaussian random variables θ_1 and θ_2 and is of the form $B(t_1, t_2) = i(\zeta_1 \theta_1 + \zeta_2 \theta_2)$.

This leads to:

$$E\left[e^{i(\xi_1 S_1+\xi_2 S_2)}\right] \sim \exp(i(\xi_1 nG(t_1)+\xi_2 nG(t_2)) - \frac{1}{2}(\xi_1^2 n\Psi_R(t_1,t_2)+2\xi_1\xi_2 n\Psi_R(t_1,t_2)+\xi_2^2 n\Psi_R(t_1,t_2)) + \textit{ cubic terms in } \xi_1,\ \xi_2 + O\left(\frac{1}{\sqrt{n}}\right)).$$

Now remember that we are actually interested in the normalized process $S'([nt]) = (S([nt]) - nG(t))/\sqrt{n}$. Substituting ξ_1 by $\xi_1'/\sqrt{n}$ and ξ_2 by $\xi_2'/\sqrt{n}$, we get:

$$E\left[e^{i(\xi_1' S_1'+\xi_2' S_2')}\right] \sim \exp\left[-\frac{1}{2}(\xi_1'^2\Psi_R(t_1,t_2)+2\xi_1'\xi_2'\Psi_R(t_1,t_2)+\xi_2'^2\Psi_R(t_1,t_2)+O\left(\frac{1}{\sqrt{n}}\right)\right].$$

Then we state a new version, more precise, of theorem 4:

THEOREM 5 ([6]). *In the projection model, the size $S([nt])$ of the projection at time nt is asymptotically a non-Markovian Gaussian process with*

$$\begin{aligned}
S([nt]) &\sim nG(t)+\sqrt{n}X_1(t),\\
E[S([nt])] &\sim nG(t), && \textit{with} \quad G(t)=F(f_1(t)),\\
\mathrm{Cov}(S([nt_1]),S([nt_2])) &\sim n\Psi_R(t_1,t_2), && \textit{with} \quad \Psi_R(t_1,t_2)=\Psi_{NR}(t_1,t_2)+f_2(t_1,t_2)\gamma(t_1)\gamma(t_2),\\
\mathrm{Var}[S([nt])] &\sim n\Phi(t), && \textit{with} \quad \Phi(t)=\Psi_R(t,t)=\Psi_{NR}(t,t)+\gamma^2(t)f_2(t,t),
\end{aligned}$$

where $X_1(t)$ is a non-Markovian Gaussian process whose covariance is denoted $\Psi_R(s,t)$ and where the functions G, Φ and Ψ_R can be given explicitly and depend of urn models (infinite or bounded) and on functions $f_1(t)$ and $f_2(s,t)$. The relative error in the density is $O(1/\sqrt{n})$.

7. Projection maximum

We have shown that the process $\mathcal{Q}$ happens to be a Gaussian process $X(t)$ superimposed on a deterministic process $G(t)$:

$$S([nt]) = G(t)+X(t).$$

If we look for its maximum $m=\max\{G(t)+X(t)\}$, and for the time $t^\star$ at which this maximum occurs, it is equivalent to searching for the hitting time of $X(t)$ to the absorbing boundary $m-G(t)$. A theorem of Daniels [2] leads to the result.

Bibliography

[1] Billinsley (Patrick). – *Convergence of Probability Measures.* – John Wiley & Sons, 1968.

[2] Daniels (H. E.). – The maximum of a Gaussian process whose mean path has a maximum, with an application to the strength of bundles of fibres. *Advances in Applied Probability*, vol. 21, 1989, pp. 315–333.

[3] Flajolet (P.), Puech (C.), and Vuillemin (J.). – The analysis of simple list structures. *Information Sciences*, vol. 38, 1986, pp. 121–146.

[4] Gardy (D.). – Normal limiting distribution for projection and semijoin sizes. *SIAM Journal on Discrete Mathematics*, vol. 5, n° 2, 1992, pp. 219–248.

[5] Gardy (D.). – Join sizes, urn models and normal limiting distributions. *Theoretical Computer Science*, vol. 131, n° 2, August 1994.

[6] Gardy (D.) and Louchard (G.). – *Dynamic analysis of some relational data base parameters. I: Projections.* – Research Report n° 94-6, PRISM, University of Versailles, February 1994.

[7] Gardy (D.) and Louchard (G.). – *Dynamic analysis of some relational data base parameters. II: Equijoins and semijoins.* – Research Report n° 94-7, PRISM, University of Versailles, February 1994.

[8] Gardy (D.) and Puech (C.). – On the effect of joins operations on relation sizes. *ACM Transactions on Database Systems*, vol. 14, n° 4, 1989, pp. 574–603.

[9] Gardy (Danièle). – *Bases de données et allocations aléatoires: Quelques analyses de performance.* – Doctorate in Sciences, Université de Paris-Sud, 1989.

[10] Louchard (G.). – Random walks, Gaussian processes and list structures. *Theoretical Computer Science*, vol. 53, 1987, pp. 99–124.