On Takeover Times in Spatially Structured Populations: Array and Ring

Günter Rudolph

Universität Dortmund, Fachbereich Informatik, D–44221 Dortmund / Germany
rudolph@LS11.cs.uni-dortmund.de

Abstract. The takeover time is the expected number of iterations of some selection method until a population consists entirely of copies of the best individual under the assumption that only one best individual is contained in the initial population. This quantity may be used to assess and compare the 'selection pressures' of selection methods used in evolutionary algorithms. Here, the notion is generalized from spatially unstructured to structured populations. Lower bounds are derived for arbitrary connected neighborhood structures, lower and upper bounds for array-like structures, and an exact closed form expression if the neighborhood structure is a ring.

1 Introduction

The notion of the *takeover time* of selection methods used in evolutionary algorithms was introduced by Goldberg and Deb [1]. Suppose that a finite population of size n consists of a single best individual and $n - 1$ worse ones. The takeover time of some selection method is the expected number of iterations of the selection method until the entire population consists of copies of the best individual. Bäck [2] has remarked that the calculations in [1] for spatially unstructured (i.e., panmictic) populations implicitly assume a guaranteed survival of at least one copy of the best individual in order to avoid extinction of the best individual by chance. Without this assumption the takeover time may become infinite for some selection methods. Therefore, Chakraborty et al. [3] have calculated the *takeover probability* of such selection methods numerically via a Markovian base model.

Thierens and Goldberg [4], Bäck [5], as well as Blickle and Thiele [6, 7] determined the *selection intensity* of selection methods, a notion adopted from quantitative genetics [8] and introduced in the field of evolutionary computation by Mühlenbein and Schlierkamp-Voosen [9]. This quantity may be used to derive the takeover time if the initial population's distribution differs from the original definition given above. This approach also neglects extinction by chance.

In case of spatially structured populations Sarma and De Jong [10, 11] postulated that the growth of the number of copies of the best individual obeys a logistic law and they fitted their empirical *growth curves* with a logistic function. Gorges-Schleuter [12] also investigated those growth curves numerically and she approximated the takeover time under the assumption of an infinitely large population. Sprave [13] modeled spatially structured populations by means of hypergraphs and developed a method to estimate growth curves and takeover times by replacing transition probabilities by expected transition rates. Needless to say, neither of these approaches took potential extinction by chance into account.

Here, the analysis focuses on those selection methods in spatially structured populations in which extinction by chance is excluded. Section 2 presents the graph-based model whereas section 3 contains the main results: Some bounds on the takeover time for general population structures, upper bounds and an approximation for array-like population structures, and an exact closed form expression if the population structure is a ring. Section 4 contains a conjecture that points to future work.

2 Spatially Structured Populations

Before the presentation of the model of selection in section 2.2, the basic terminology regarding graphs is briefly recalled in section 2.1 because these terms will be used frequently throughout the paper.

2.1 Graphs

A graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ consists of a finite nonempty set $\mathcal{V}$ of vertices (or nodes) and a collection $\mathcal{E}$ of pairs of vertices from $\mathcal{V}$. Each pair $e \in \mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is termed an edge. If (u, v) is an edge then u and v are adjacent or neighboring vertices and (u, v) is said to be incident to u and v. An edge (v, v) for some $v \in \mathcal{V}$ is called a self-loop. The degree deg(v) of a vertex v is the number of edges that are incident to v. A graph is termed regular if each of its vertices has the same degree.

A path in a graph $\mathcal{G}$ is a finite sequence of distinct vertices $(v_1, v_2, \ldots, v_k)$ such that $(v_1, v_2), (v_2, v_3), \ldots, (v_{k-1}, v_k)$ are edges in $\mathcal{G}$. A graph $\mathcal{G}$ is connected if for each vertex u there is a path $(u \equiv v_1, v_2, \ldots, v_k \equiv v)$ to each other vertex v. The length of a path $(v_1, v_2, \ldots, v_k)$ between v_1 and v_k is $k - 1$, i.e., the number of edges along the path. The distance dist(u, v) between vertices u and v is the length of the shortest path between u and v. The diameter diam$(\mathcal{G}) = \max\{\text{dist}(u, v) | u, v \in \mathcal{V}\}$ of a graph is the maximum of the distances between all pairs of vertices.

According to Babai [14] isomorphism of undirected graphs are bijections of vertex sets preserving adjacency as well as non-adjacency. Consequently, automorphisms are $\mathcal{G} \to \mathcal{G}$ isomorphisms. An undirected graph $\mathcal{G}$ is said to be vertex-transitive if for each pair of vertices u and v there is an automorphism α of $\mathcal{G}$ such that $\alpha(u) = v$. Notice that each vertex-transitive graph is regular whereas the converse is wrong in general. Vertex-transitive graphs have the property that "the graph looks the same" viewed from each vertex. Those graphs are frequently used to define the neighborhoods in spatially structured populations.

2.2 Modeling Selection on Graphs

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be an undirected connected graph with $|\mathcal{V}| = n$. Each vertex $i \in \mathcal{V}$ is associated with an individual of the evolutionary algorithm, i.e., individual i may be seen as the value of vertex $i \in \mathcal{V}$. The spatial structure of the population is defined by the set of edges: The neighborhood of individual/vertex i consists of all those individuals/vertices that are adjacent to i. Notice that vertex i may belong to its own neighborhood if self-loops are allowed.

Since there exist numerous selection procedures in the field of evolutionary computation, it is useful to abstract from specific methods. For this purpose assume that the selection procedure is split into two stages: In the first stage an individual is chosen from the neighborhood of each individual according to some method. In the second stage for each individual it is decided whether the previously chosen individual will be adopted or not. Now all selection procedures can be characterized by the following probability values: Let $c_i(k)$ be the probability that an individual with best fitness is chosen from the neighborhood of individual i at step k, whereas $a_i(k)$ denotes the probability that individual i adopts the previously chosen individual if it is worse than individual i. Evidently, if the second stage is not present for some selection procedure then $a_i(k) \equiv 1$ for all $i = 1, \ldots, n$ and $k \geq 1$; this situation occurs frequently since local selection methods that have been actually programmed in some spatially structured evolutionary algorithm are usually simple adaptions of the programmer's favorite global selection method that is only occasionally equipped with an explicit second stage.

For example, one obtains $c_i(k) \in (0, 1)$ for proportional and ranking selection whereas $c_i(k) = 1$ for stochastic universal sampling (SUS) selection. In case of tournament selection (without replacement) one has to distinguish between two cases: If the tournament size is equal to the neighborhood size then $c_i(k) = 1$ and otherwise $c_i(k) \in (0, 1)$.

3 Takeover Time on Graphs

3.1 Definitions

Let $\mathcal{G}$ = ($\mathcal{V}$,E) with $|\mathcal{V}| = n$ be an undirected connected graph representing the neighborhood structure. The random variables $V_i(k) \in \{0, 1\}$ are the value of vertex $i \in \mathcal{V}$ at iteration $k \geq 0$ where the value 1 indicates a copy of the best individual whereas the value 0 indicates a copy of a worse one. Random variable $N_k = \sum_{i=1}^{n} V_i(k)$ denotes the number of copies of the best individual at iteration $k \geq 0$. Initially, $V_i(0) = 1$ for some $i \in \mathcal{V}$ and $V_j(0) = 0$ for all $j \neq i$ such that $N_0 = 1$. Notice that the stochastic process $\{N_k : k \geq 0\}$ is not necessarily Markovian since it is a function of the Markov chain $\{(V_1(k), V_2(k), \ldots, V_n(k)) : k \geq 0\}$. In general, $\{N_k : k \geq 0\}$ has absorbing states at 0 and n. The random time until absorption is called the absorption time $A = \min\{k \geq 1 : N_k = 0 \vee N_k = n\}$. Absorption at state n occurs with probability $\mathsf{P}\{\, N_A = n \,\}$ and absorption at state 0 with probability $\mathsf{P}\{\, N_A = 0 \,\}$. Clearly, $\mathsf{P}\{\, N_A = 0 \,\} + \mathsf{P}\{\, N_A = n \,\} = 1$. In [3] $\mathsf{P}\{\, N_A = n \,\}$ was termed the *takeover probability* and we may call $\mathsf{P}\{\, N_A = 0 \,\}$ the *extinction probability* in the context considered here. If the extinction probability is not zero then the definition of a takeover time is problematic. Moreover, in publications of numerical experiments it is usually undocumented which quantities have been estimated actually: Candidates are the absorption time $\mathsf{E}[\, A \,]$, the absorption time $\mathsf{E}[\, A \mid N_A = n \,]$ conditioned by the event of absorption at state n, and even $\mathsf{P}\{\, N_A = n \,\} \cdot \mathsf{E}[\, A \mid N_A = n \,]$. These quantities are related via

$$\mathsf{E}[\, A \,] = \mathsf{E}[\, A \mid N_A = 0 \,] \cdot \mathsf{P}\{\, N_A = 0 \,\} + \mathsf{E}[\, A \mid N_A = n \,] \cdot \mathsf{P}\{\, N_A = n \,\} \,.$$

Evidently, if the extinction probability $\mathsf{P}\{ N_A = 0 \}$ is zero such problems disappear and the notion of the takeover time is clear without ambiguity.

Definition 1. Let the random sequence $\{N_k : k \geq 0\}$ with $N_0 = 1$ represent the number of copies of the best individual at iteration $k \geq 0$ of some selection method acting on a population of size n. If the extinction probability is zero then $\mathsf{E}[\, T\,]$ with $T = \min\{k \geq 1 : N_k = n\}$ is called the *takeover time* of the selection method.
In case of spatially structured populations the quantity $\mathsf{E}_i[\, T\,]$, denoting the takeover time if vertex i contains the initial 1, is termed the *takeover time with initial vertex* i. The takeover time is then given by $\mathsf{E}[\, T\,] = \frac{1}{n} \sum_{i \in \mathcal{V}} \mathsf{E}_i[\, T\,]$ assuming a uniformly distributed emergence of the first 1 among all vertices. ■

Finally, it must be specified how the selection method affects the value of each vertex.

Definition 2. Let $c_i(k)$ be the probability that the selection method chooses the best individual in the neighborhood of vertex $i \in \mathcal{V}$ at step $k \geq 0$ and $a_i(k)$ the probability that individual i adopts the previously chosen individual if it is worse than i. Then $\mathsf{P}\{ V_i(k+1) = 1 \,|\, V_i(k) = 0\} = c_i(k)$ is termed the *upgrade probability* of vertex i at step k whereas $\mathsf{P}\{ V_i(k+1) = 0 \,|\, V_i(k) = 1\} = (1 - c_i(k)) \cdot a_i(k)$ is the *downgrade probability*. ■

The theoretical analysis presented here will be restricted to selection methods where **either** $c_i(k) = 1$ **or** $c_i(k) \in (0, 1) \wedge a_i(k) = 0$. This will be the general assumption hereinafter. As a consequence, the downgrade probability is always zero which in turn implies that the extinction probability is zero as well.

3.2 General Results

The derivation of sharp lower bounds on the takeover time is almost trivial:

Proposition 1. Let $\mathcal{G}$ be an undirected connected graph. Then

$$\mathsf{E}[\, T\,] \geq \frac{1}{n} \sum_{i \in \mathcal{V}} \max\{\text{dist}(i, j) : j \in \mathcal{V}\}$$

for all selection methods with downgrade probability zero. If the upgrade probability is 1 then the inequality becomes an equality.
Proof: Evidently, the fastest spread of 1s over arbitrary connected graphs is achieved if the upgrade probability is 1 (which implies a zero downgrade probability). In this case exactly $\max\{\text{dist}(i, j) : j \in \mathcal{V}\}$ iterations are necessary to fill all nodes with 1s from initial vertex i. If the upgrade probability is less than one then additional iterations may be required. ■

This results can be sharpened for a special family of graphs.

Proposition 2. Let $\mathcal{G}$ be a vertex-transitive graph. Then $\mathsf{E}[\, T\,] \geq \text{diam}(\mathcal{G})$ for all selection methods with downgrade probability zero. If the upgrade probability is 1 then the inequality becomes an equality.

Proof: Since $\mathcal{G}$ is vertex-transitive the takeover times are identical regardless of the initial vertex. Again, the fastest spread of 1s is obtained if the upgrade probability is 1. In this case the takeover time is

$$\frac{1}{n}\sum_{i\in\mathcal{V}}\max\{\text{dist}(i,j) : j\in\mathcal{V}\} = \max\{\text{dist}(i,j) : i,j\in\mathcal{V}\} = \text{diam}(\mathcal{G}).$$

If the upgrade probability is less than 1 then additional iterations may be required. ∎

3.3 Array

Let $\mathcal{G} = (\mathcal{V},\mathcal{E})$ be an array with $|\mathcal{V}| = n$ and $\mathcal{E} = \{(i, i+1) : i = 1,\ldots,n-1\}$. In this case the upgrade probability is identical for each vertex. Suppose that the initial vertex is $i = 1$. Vertex 2 upgrades from 0 to 1 with (upgrade) probability p, whereas vertices $3,\ldots,n$ will keep the value 0. As soon as vertex 2 has upgraded, vertex 3 will upgrade with probability p, whereas vertices $4,\ldots,n$ will stay at value 0. And so forth until vertex n is upgraded. Thus, the random time required for upgrading vertex $i+1$ if vertex i has value 1 is a geometrically distributed random variable G_i with $\mathsf{E}[\,G_i\,] = 1/p$ and $\mathsf{V}[\,G_i\,] = (1-p)/p^2$. Since $G_1,\ldots,G_{n-1}$ are mutually independent it follows that

$$\mathsf{E}_1[\,T\,] = \sum_{i=1}^{n-1}\mathsf{E}[\,G_i\,] = \frac{n-1}{p} \quad\text{and}\quad \mathsf{V}_1[\,T\,] = \sum_{i=1}^{n-1}\mathsf{V}[\,G_i\,] = \frac{(n-1)\,(1-p)}{p^2}\,.$$

Clearly, $\mathsf{E}_1[\,T\,] = \mathsf{E}_n[\,T\,]$ and $\mathsf{V}_1[\,T\,] = \mathsf{V}_n[\,T\,]$. Now suppose that the initial vertex is $i \in \{2,\ldots,n-1\}$. In this case there are two independent processes starting at i, one process upgrades the vertices to the left until vertex 1, the other one upgrades the vertices to the right until vertex n. If H_1 and H_2 are the independent absorption times of the two processes then $\mathsf{E}_i[\,T\,] = \mathsf{E}_i[\,\max\{H_1,H_2\}\,] \ge \max\{\mathsf{E}_i[\,H_1\,],\mathsf{E}_i[\,H_2\,]\} = \max\{i-1,n-i\}/p$. This leads to the lower bound

$$\mathsf{E}[\,T\,] \ge \begin{cases} \dfrac{3\,n-2}{4\,p} & ,\, n \text{ even} \\[2ex] \dfrac{3\,n+1}{4\,p}\left(1-\dfrac{1}{n}\right) & ,\, n \text{ odd} \end{cases}$$

that reduces to the lower bound in Proposition 1 if $p = 1$. To obtain the exact value for $\mathsf{E}_i[\,\max\{H_1,H_2\}\,]$ one needs the distribution of the random variables H_1 and H_2. Since they are sums of independent geometric random variables they have a negative binomial distribution with

$$\mathsf{P}\{\,H = k\,\} = \binom{k-1}{m-1}\,p^m\,(1-p)^{k-m}$$

where $m \in \{i-1, n-i\}$ denotes the number of necessary upgrades until absorption. If $i-1 = n-i$ then H_1 and H_2 have the same distribution and one may use the result of

Young [15] who derived a formula for the expectation of the minimum of independent and identically distributed negatively binomial random variables. One obtains the expectation of the maximum via the identity $\max\{H_1, H_2\} = H_1 + H_2 - \min\{H_1, H_2\}$. Alternatively, one may use the asymptotical expression developed in [16] for the expectation of the maximum. But in the situation considered here the random variables H_1 and H_2 are generally not identically distributed. Although there is no problem in calculating the exact values via

$$\mathsf{E}[\min\{H_1, H_2\}] = \sum_{k=1}^{\infty} \mathsf{P}\{\min\{H_1, H_2\} \geq k\} = \sum_{k=1}^{\infty} \mathsf{P}\{H_1 \geq k\} \cdot \mathsf{P}\{H_2 \geq k\}$$

a closed form expression is difficult to achieve. Therefore an upper bound will be derived now. Notice that $\mathsf{E}_i[H_1] = (i-1)/p$, $\mathsf{E}_i[H_2] = (n-i)/p$, $\mathsf{V}_i[H_1] = (i-1)(1-p)/p^2$ and $\mathsf{V}_i[H_2] = (n-i)(1-p)/p^2$. Owing to the identity $\max\{a, b\} = (a+b+|a-b|)/2$ the expectation of the random variable $|H_1 - H_2|$ must be bounded. Since

$$\begin{aligned}
\mathsf{E}_i[\,|H_1 - H_2|\,]^2 \;\leq\; \mathsf{E}_i[(H_1 - H_2)^2] &= \mathsf{V}_i[H_1 - H_2] + \mathsf{E}_i[H_1 - H_2]^2 \\
&= \mathsf{V}_i[H_1] + \mathsf{V}_i[H_2] + (\mathsf{E}_i[H_1] - \mathsf{E}_i[H_2])^2 \\
&= \frac{(n-1)(1-p)}{p^2} + \left(\frac{2i-n-1}{p}\right)^2
\end{aligned}$$

one immediately obtains

$$\begin{aligned}
\mathsf{E}_i[\max\{H_1, H_2\}] &\leq \frac{n-1}{2p} + \frac{1}{2p}\sqrt{(n-1)(1-p) + (2i-n-1)^2} \\
&\leq \frac{n-1}{2p} + \frac{1}{2p}\sqrt{n(n-1)} \\
&\leq \frac{n-1}{2p} + \frac{n-1/2}{2p}
\end{aligned}$$

and finally $\mathsf{E}[T] \leq (4n-3)/(4p)$.

3.4 Ring

If the spatial structure is a ring then the upgrade probabilities are identical for each vertex. Since a ring is vertex-transitive the takeover time does not depend on the initial vertex.

Proposition 3. Let $\mathcal{G}$ be a ring of size $n \geq 2$. The takeover time of a selection method with upgrade probability $p \in (0,1)$ and downgrade probability zero is recursively determinable via

$$\mathsf{E}[T_{n+1}] = \frac{1}{p(2-p)} + \frac{2(1-p)}{2-p}\mathsf{E}[T_n] + \frac{p}{2-p}\mathsf{E}[T_{n-1}] \tag{1}$$

for $n \geq 3$ where $\mathsf{E}[T_2] = 1/p$ and $\mathsf{E}[T_3] = (3-2p)/(2p-p^2)$.

Proof: If $n = 2$ the only empty vertex is reached with probability p in one step. Thus, T_2 is geometrically distributed with $\mathsf{E}[\,T_2\,] = 1/p$. Let $n = 3$. At the beginning, exactly one empty vertex is reached in one step with probability $2\,p\,(1-p)$ whereas both empty vertices are reached simultaneously in one step with probability p^2. Thus, the probability to reach at least one empty vertex in one step is $2\,p\,(1-p) + p^2 = p\,(2-p)$. This happens after $1/(2\,p - p^2)$ steps on average. Assume the event has happened. Then a single empty vertex has been reached with probability $2\,p\,(1-p)/(2\,p\,(1-p)+p^2) = 2\,(1-p)/(2-p)$. The mean time to reach the last empty vertex is again $1/p$ so that

$$\mathsf{E}[\,T_3\,] = \frac{1}{p\,(2-p)} + \frac{2\,(1-p)}{2-p} \cdot \frac{1}{p} = \frac{3-2\,p}{p\,(2-p)}\,.$$

The argumentation for $\mathsf{E}[\,T_3\,]$ is easily generalized: $\mathsf{E}[\,T_{n+1}\,]$ consists of the time to "leave" the initial vertex plus either the time $\mathsf{E}[\,T_n\,]$ if only one vertex has adopted the 1 or the time $\mathsf{E}[\,T_{n-1}\,]$ if two vertices have adopted the 1. The first event happens with probability $2\,(1-p)/(2-p)$ and the second one with probability $1-2\,(1-p)/(2-p) = p/(2-p)$. Putting altogether one immediately arrives at the equation (1) and the proof is completed. ■

The recurrence in equation (1) can be solved by means of generating functions (see e.g. [17], ch. 7.3). The result itself is easily verified by induction.

Proposition 4. Let $\mathcal{G}$ be a ring of size $n \geq 2$. The takeover time of a selection method with upgrade probability $p \in (0,1)$ and downgrade probability zero is

$$\mathsf{E}[\,T_n\,] = \frac{n}{2\,p} - \frac{1}{4}\left[1 - (-1)^{n-2}\left(\frac{p}{2-p}\right)^{n-2}\right]. \tag{2}$$

Proof: (by induction)
Insertion of $n = 2$ and $n = 3$ in equation (2) yields the same values for the takeover times that have been proven in Proposition 3. Assume that the hypothesis is true for $n > 3$. Insertion of the hypothesis (2) into the recursion (1) proves that the result is true for $n + 1$ and hence for all $n \geq 2$. ■

4 A Conjecture ... and Future Work

The results developed so far seem to indicate that the takeover time depends to a smaller extent on the selection method itself than on the diameter of the underlying neighborhood structure. A careful examination of this conjecture is desirable.

Acknowledgments

This work was supported by the *Deutsche Forschungsgemeinschaft (DFG)* as part of the *Collaborative Research Center "Computational Intelligence" (SFB 531).*

References

1. D. E. Goldberg and K. Deb. A comparative analysis of selection schemes used in genetic algorithms. In G. J. E. Rawlins, editor, *Foundations of Genetic Algorithms*, pages 69–93. Morgan Kaufmann, San Mateo (CA), 1991.
2. T. Bäck. Selective pressure in evolutionary algorithms: A characterization of selection mechanisms. In *Proceedings of the First IEEE Conference on Evolutionary Computation, Vol. 1*, pages 57–62. IEEE Press, Piscataway (NJ), 1994.
3. U. Chakraborty, K. Deb, and M. Chakraborty. Analysis of selection algorithms: A Markov chain approach. *Evolutionary Computation*, 4(2):133–167, 1996.
4. D. Thierens and D. E. Goldberg. Convergence models of genetic algorithm selection schemes. In Y. Davidor, H.-P. Schwefel, and R. Männer, editors, *Parallel Problem Solving from Nature — PPSN III*, pages 119–129. Springer, Berlin and Heidelberg, 1994.
5. T. Bäck. Generalized convergence models for tournament and (μ, λ)–selection. In L. J. Eshelman, editor, *Proceedings of the Sixth International Conference on Genetic Algorithms*, pages 2–8. Morgan Kaufmann, San Fransisco (CA), 1995.
6. T. Blickle and L. Thiele. A mathematical analysis of tournament selection. In L. J. Eshelman, editor, *Proceedings of the Sixth International Conference on Genetic Algorithms*, pages 9–16. Morgan Kaufmann, San Fransisco (CA), 1995.
7. T. Blickle and L. Thiele. A comparison of selection schemes used in evolutionary algorithms. *Evolutionary Computation*, 4(4):361–394, 1996.
8. M. G. Bulmer. *The Mathematical Theory of Quantitative Genetics*. Clarendon Press, Oxford, 1980.
9. H. Mühlenbein and D. Schlierkamp-Voosen. Predictive models for the breeder genetic algorithm I: Continuous parameter optimization. *Evolutionary Computation*, 1(1):25–49, 1993.
10. J. Sarma and K. A. De Jong. An analysis of the effects of neighborhood size and shape on local selection algorithms. In H.-M. Voigt, W. Ebeling, I. Rechenberg, and H.-P. Schwefel, editors, *Parallel Problem Solving From Nature — PPSN IV*, pages 236–244. Springer, Berlin, 1996.
11. J. Sarma and K. A. De Jong. An analysis of local selection algorithms in a spatially structured evolutionary algorithm. In T. Bäck, editor, *Proceedings of the 7th International Conference on Genetic Algorithms*, pages 181–186. Morgan Kaufmann, San Francisco (CA), 1997.
12. M. Gorges-Schleuter. An analysis of local selection in evolution strategies. In W. Banzhaf et al., editors, *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO '99), Vol. 1*, pages 847–854. Morgan Kaufmann, San Fransisco (CA), 1999.
13. J. Sprave. A unified model of non-panmictic population structures in evolutionary algorithms. In *Proceedings of the 1999 Congress on Evolutionary Computation, Vol. 2*, pages 1384–1391. IEEE Press, Piscataway (NJ), 1999.
14. L. Babai. Automorphism groups, isomorphism, reconstruction. In R. Graham, M. Grötschel, and L. Lovász, editors, *Handbook of Combinatorics, Vol. II*, pages 1447–1540. Elsevier Science, Amsterdam, 1995.
15. D. H. Young. The order statistics of the negative binomial distribution. *Biometrika*, 57:181–186, 1970.
16. P. J. Grabner and H. Prodinger. Maximum statistics of N random variables distributed by the negative binomial distribution. *Combinatorics, Probability & Computing*, 6(2):179–183, 1997.
17. R. L. Graham, D. E. Knuth, and O. Patashnik. *Concrete Mathematics*. Addison-Wesley, Reading (MA), 4th edition, 1990.