# Genealogical Patterns in Evolutionary Algorithms

Carlos M. Fernandes Agostinho C. Rosa LARSyS: Laboratory for Robotics and Systems in Engineering and Science, University of Lisbon, Portugal cfernandes@laseeb.org J.L.J. Laredo LITIS, University of Le Havre, Le Havre, France juanlu.jimenez@univ-lehavre.fr

J.J. Merelo Department of Computer Architecture University of Granada, Spain imerelo@geneura.ugr.es

### ABSTRACT

Event Takeover Values (ETV) measure the impact of each individual in the population dynamics of evolutionary algorithms (EA). Previous studies argue that ETV distribution of panmictic EAs fit power laws with exponent between 2.2 and 2.5 and that this property is insensitive to fitness landscapes and design choices of the EAs. One exception is cellular EAs, for which there are deviations of the power law for large values. In this paper, ETVs for structured and panmictic EAs with different population size and mutation probability on several fitness landscapes were computed. Although the ETVs distribution of pamictic EAs is heavy-tailed, the log-log plot of the complementary cumulative distributed function shows no linearity. Furthermore, Vuong's tests on the distributions generated by several instances of the problems conclude that power law models cannot be favored over log-normal models. On the other hand, the tests confirm that cEAs impose significant deviations to the distribution tail.

Keywords: Evolutionary Algorithms, Event Takeover Values.

#### **1. INTRODUCTION**

In spatially structured evolutionary algorithms (EAs), interactions between the individuals are constrained by a network that connects the members of the population. These algorithms can be divided into two generic classes: cellular EAs (cEAs) [1] and island models [2]. It is argued that cEAs provide a better sampling of the search space and therefore improve the performance of nonstructured (panmictic) EAs in multimodal, non-linear and deceptive fitness landscapes [1][5]. The reason for the alleged better performance of cEAs may be the fact that they reduce genetic diversity loss rate during the run: individuals only interact with a restricted number of other individuals. Hence, good solutions diffuse slower through the network, requiring longer takeover times.

Takeover times have some drawbacks: theoretical analysis is difficult when typical working mechanisms (recombination and mutation, for instance) have to be taken into account; they can only be measured in restricted experimental conditions with simplified algorithms; they are computed for a single solution and therefore do not offer much insight on the overall dynamics of the populations. In order to overcome (or complement) the limitations of takeover times, Whitacre *et al.* [6] proposed a new measure of the individuals' influence on the dynamics of populations, called

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from <u>Permissions@acm.org</u>.

*GECCO '19 Companion,* July 13–17, *2019, Prague, Czech Republic* © 2019 Association for Computing Machinery. ACM ISBN 978-1-4503-6748-6/19/07...\$15.00 https://doi.org/10.1145/3319619.3321914 Event Takeover Value (ETV). ETVs evaluate each individual's impact on the population using information from genealogical graphs. In short, an ETV of an individual *id* in generation *t* is the number of descendants of that individual that belong to the population in generation *t*, as described by Eq. 1:

$$ETV_t(id) = \sum_{i=1}^n \sum_{j=1}^{Tobs} \phi_{i,j}$$

$$\phi_{i,j} = \begin{cases} 1 & \text{if } id = M_i(id_j) \\ 0 & \text{otherwise} \end{cases}$$
(1)

where *n* is the population size,  $T_{obs}$  the maximum size of the ascendants list (to compute ETV it is mandatory to keep an ascendants list for each individual) and  $M_i(id_i)$  is the  $j^{th}$  position in the individual *i* ascendants list. To avoid *genetic hitchhiking*, the ETV<sub>t</sub> of an individual is compared with one of its offspring. If the values are equal, the parent's ETV<sub>t</sub> is set to 0. Formally, given two individuals  $id_1$  and  $id_2$ , genetic hitchhiking prevention is described by Equation 2:

$$if ((id_1 = M_i(id_j)) and ((id_2 = M_i(id_{j-1})) and$$
  
and  $ETV_i(id_1) = ETV_i(id_2)$ , then  $ETV_i(id_1) = 0$  (2)

Please note that ETVs are bounded by  $[ETV_{min}, ETV_{max}] = [1, n]$ , where *n* is the population size. For a detailed description of ETV and its calculation procedure please refer to [6].

Whitacre *et al.* claim that ETV distributions of panmictic EAs are fitted by a power law with exponent in the range [2.2, 2.5], while cEAs deviate from power laws for large values. These results are consistent with cEAs takeover times: they show that the probability of the population being taken by a single solution is lower for cEAs. The authors also claim that ETVs' power law distribution is an emergent pattern of panmictic EAs, insensitive to design choices.

This paper investigates and compares ETVs of panmictic and cellular EAs with different edge degree and under several experimental conditions. The main goal is to confirm if ETV distributions fit power laws and are indeed insensitive to population size and other design choices.

### 2. EXPERIMENTS AND RESULTS

In this paper, cEAs are structured by several regular graphs. Starting from a ring structure (k = 2), k is doubled by linking each individual to the neighbors of its neighbors, thus creating regular graphs with  $k = \{2,4,8,16,32\}$ . Additionally, EAs with k = n - 1 (i.e., panmictic populations), where n is the population size, are tested. Synchronous cEAs are used: offspring are kept in a secondary population that replaces the old population when its



**Figure 1.** MMDP: l = 60; n = 200;  $p_m = 1/l$ . ETV CCDF of panemict EA (k = n - 1) and cEAs with k = 2.

size n' is equal to n. The selection scheme is binary tournament, two-point crossover is the recombination method and bit-flip is the mutation type. In each iteration, each individual (*parent1*) is recombined with one of its neighbors (*parent2*). From the set of two children generated by crossover, one is randomly chosen and replaces *parent1* if its fitness is better. The test set is composed of five functions with different characteristics: onemax, 2-trap, 3trap, 4-trap and MMDP. Each function is tested with different string size l for assessing scalability. Each configuration of each algorithm is executed 30 times in each problem instance. Success rates (number of runs in which the optimum is found) and statistics on best fitness values and number of evaluation to find the best solution are recorded and compared.

The experiments and statistical analysis do not confirm the presence of power laws in the collected data. Figure 1 shows the log-log of the ETV complementary cumulative distributions (CCDF) of a cEA (k = 2) and a panmictic EA in an instance of the MMDP. Population size n = 200 optimizes EAs convergence speed (both algorithms attain the global optimum in every run) in this particular instance of MMDP.

Distributions are clearly not linear through the whole range of ETVs. For the panmictic EA, a power law model was computed and the maximum likelihood used to estimate the exponent, which is  $\alpha = 2.14$ , below the values in [6]. Furthermore, a Vuong's test comparing power law and log-normal models concluded that a log-normal distribution could also produce the ETVs: there is no evidence of power law in the data. The statistical tests were made with the "poweRlaw" R package [4], following guidelines in [3].

ETVs may not fit a power law but, as seen in Figure 1 (and the outcome is similar with other fitness landscapes), structuring the population has a clear effect on the distribution, namely for large sizes. Largest ETVs of cEAs are well below population size n, and in general they are bounded by [1, n/2], while panmictic ETVs very often cover the whole possible range of values [1, n]. As degree k increases, the upper bound also increases and for k = 32 its value is closer to n. Please note that these behavioural patterns were observed in every fitness landscape.

From the large amount of data gathered with the tests, it is possible to identify other behavioural patterns. In general, increasing population size has strong effects on the distribution. Figure 2 shows the CCDFs of panmictic EAs with different population size n on the same instance of MMDP. Every configuration finds the best solution in every run and the EA with n = 200 requires less evaluations to find the optimum (considering median values): n = 200 is therefore the panmictic EA's optimal population size (amongst the n values that were tested) for this instance of the problem.



**Figure 2.** MMDP: l = 60;  $p_m = 1/l$ . Panmictic EAs ETVs CCDFs with different population size. Optimal population: n = 200.

Please note that for n = 50 and n = 100, the distribution is characterized by high probability values when ETVs are close to the upper bound  $ETV_{max} = n$ : i.e, it is highly probable that descents of a good solution takeover the entire population. With optimal size (n = 200), that probability decreases significantly, and above optimal population the distributions do not suffer much deviations from the n = 200 distribution. These results show that ETVs distributions strongly depend on population size and that they can detect below-optimal population size.

#### **3. CONCLUSIONS AND FUTURE WORK**

A statistical analysis of several ETVs distributions generated by panmictic populations on several fitness landscapes show that the presence of power law distribution in data cannot be claimed. These findings, however, do not diminish the importance of ETV in the analysis of EAs dynamic behavior. The results confirm that cEAs generate deviations from the panmictic heavy-tailed distributions, reducing ETVs upper bounds, which mean that they indeed restrict descendants of good solutions from dominating the entire population. Since power law exponents cannot be used to assess the effects of design choices in ETVs, the goal now is to devise other statistical measures to help investigate how population structure, population size and other parameters affect genealogical dynamics and ETVs and how that dynamics can be used understand the mechanisms behind efficient algorithms.

## ACKNOWLEDGEMENTS

Work supported by FCT PROJECT [UID/EEA/50009/2013].

#### References

- [1] Alba, E., Tomassini, M. 2002. Parallelism and evolutionary algorithms. *IEEE Trans. on Evol. Comp.*, 6(5), 443–462.
- [2] Cantú-Paz, E. 2001. Migration Policies, Selection Pressure, and Parallel EAs. *Journal of Heuristics*, 7(4), 311-334.
- [3] Clauset, A., Shalizi, C.R., Newman, M.E.J. 2009. Power-law distributions in empirical data. *SIAM Review*, 51(4), 661-703.
- [4] Gillespie, C.S. 2015. Fitting Heavy Tailed Distributions: The poweRlaw Pa[4]ckage. *Journal of Statistical Software*, 64(2), 1-16
- [5] Tomassini, M. 2005. *Spatially Structured Evolutionary Algorithms*. Springer, Heidelberg.
- [6] Whitacre, J.M., Sarker, R.A., Pham, Q. 2009. Making and Breaking Power Laws in Evolutionary Algorithms Population. *Memetic Computing*, 1(2), 125-137.