Criticality and parallelism in structured SAT instances

Andrea Roli

DEIS – Università degli Studi di Bologna (Italia)
aroli@deis.unibo.it

Abstract. In this work we address the question of whether and how parallel local search exhibits the *criticality and parallelism* phenomenon when performed on structured instances. We experimentally show that also for structured instances there exists an optimal value of parallelism which enables the algorithm to reach the optimal performance and, by analyzing the frequency of node degree of the graphs associated with the SAT instances, we observe that an asymmetric and not regular distribution strongly affects the algorithm performance with respect to the parallelism.

1 Introduction

The phenomenon called *criticality and parallelism* has been observed in the context of local search algorithms applied to combinatorial optimization problems [4, 3], where local search is modified by applying τ local moves in parallel. The most important results observed so far can be summarized in the following properties: (i) the algorithm performance increases as parallelism increases, up to a parallelism degree (τ_{opt}) at which it starts to decrease, (ii) the optimal parallelism depends on a parameter which measures the connectivity of the system, and (iii) when the decrease in performance for $\tau > \tau_{opt}$ is abrupt, the behavior is accompanied by a phase transition.

The same phenomenon has been discovered in parallel local search for SAT [5] and MAXSAT [6], where a parallel version of GSAT [7] (called PGSAT) has been applied. In PGSAT, variables are divided in τ subsets and, for each subset, we flip the variable that, if flipped, decreases the greatest number of unsatisfied clauses. For satisfiable SAT instances it has been experimentally shown that the best global performance (time, iterations, fraction of solved instances) is achieved with an optimal parallelism degree τ_{opt}. Furthermore, τ_{opt} is monotonically non increasing with the connectivity among variables.

Recent studies [9] investigate the different distribution of node degree of the graphs associated with random and structured CSPs and its impact on search. The authors show that real world problems frequently have a connectivity distribution strongly different from that of random instances and they point out how this topological characteristic may affect the search effectiveness.

The aim of this work is to investigate the effects of problem structure on the parallelization of local search for SAT problems, and compare them with previous results concerning random instances [5, 6].

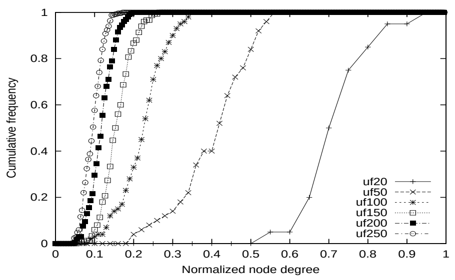

Fig. 1. Cumulative frequency vs. normalized node degree for Uniform Random 3-SAT instances in the threshold region.

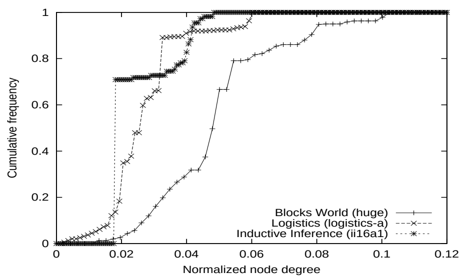

Fig. 2. Cumulative frequency vs. normalized node degree in structured instances: Blocks World (*huge*), Logistics (*logistics-a*) and Inductive Inference (*ii16a1*).

2 Structured SAT instances

In this section we first define the node degree distribution of graphs associated with random and structured SAT instances. Then we show and discuss results of PGSAT, run with different values of τ, on two representative structured SAT instances.

The graph associated with a SAT instance is an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{A})$, where each node $v_i \in \mathcal{V}$ corresponds to a variable and edge $(v_i, v_j) \in \mathcal{A}$ $(i \neq j)$ if and only if variables v_i and v_j appear in a same clause. In an instance with n variables, each node v_i, $i = 1, \ldots, n$, has a degree $q_i \in \{0, 1, \ldots, n-1\}$. We define the *average connectivity* of the instance as the average node degree of the corresponding graph, i.e., $q = \frac{1}{n}\sum_{i=1}^{n} q_i$. Moreover, to make direct comparisons among instances with different number of variables, we also introduce the normalization of q: $\overline{q} = \frac{q}{n-1}$. In order to compare the node degree distribution between instances, we consider the cumulative frequency $Freq(j)$ = frequency of a node connected to not more than j nodes. Fig.1 shows the cumulative frequency vs. the normalized node degree for random 3-SAT instances retrieved from SATLIB [2]. Note that the curves are quite regular and, as the number of variables increases, they converge to a step function located at the average node degree.

Structured instances are characterized by the presence of some regularity in their components, for example graph problems based on ring or lattice topology, SAT problems obtained by encoding logic problems (circuit testing, inductive inference, planning, etc.). Fig.2 shows the curve of cumulative frequency for structured SAT instances taken from SATLIB. The plotted curves show apparent differences with those of random SAT problems. They are not as regular as random ones and they have gaps and plateaus, thus structured instances have a more spread and non-uniform connectivity distribution. The instance *ii16a1* has 1650 variables and a normalized average connectivity $\overline{q}_{ii16a1} = 0.0239$. Its frequency of node degree is shown in Fig.3, along with that of a random 3-SAT instance of the same size and normalized connectivity (instance *3sat1650*) in Fig.4. We can note that the node degree frequency of the structured instance is

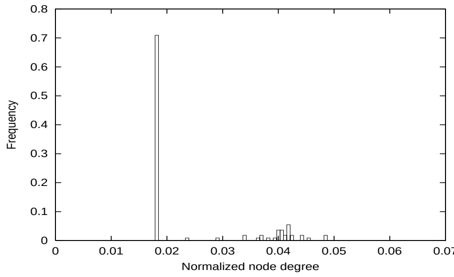

Fig. 3. Frequency vs. normalized node degree in the instance *ii16a1*.

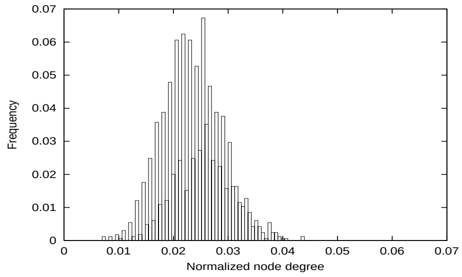

Fig. 4. Frequency vs. normalized node degree in the instance *3sat1650*.

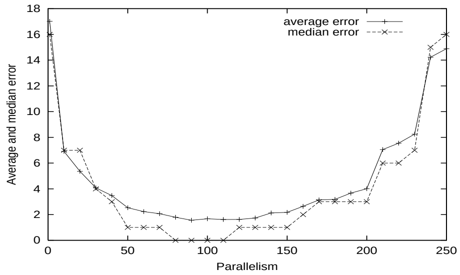

Fig. 5. Average and median error against τ for PGSAT on the instance *ii16a1*.

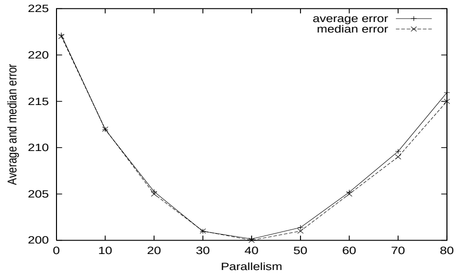

Fig. 6. Average and median error against τ for PGSAT on the instance *3sat1650*.

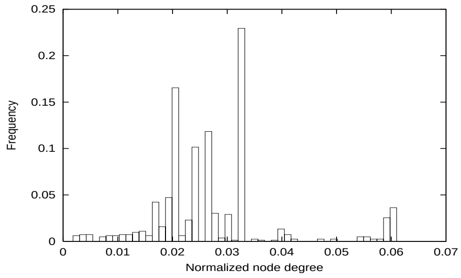

Fig. 7. Frequency vs. normalized node degree in the instance *logistics-a*.

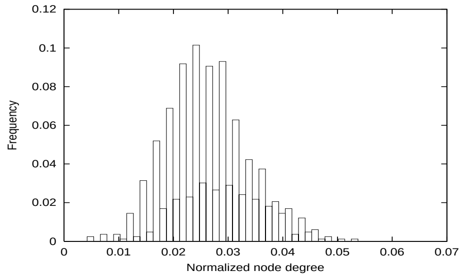

Fig. 8. Frequency vs. normalized node degree in the instance *3sat828*.

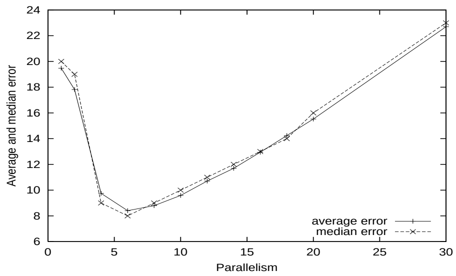

Fig. 9. Average and median error against τ for PGSAT on the instance *logistics-a*.

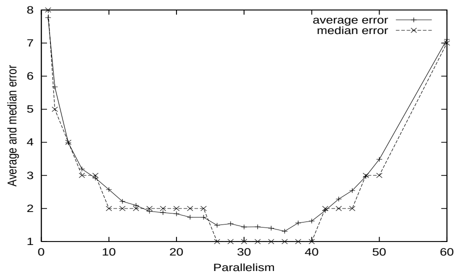

Fig. 10. Average and median error against τ for PGSAT on the instance *3sat828*.

highly asymmetric and has a peak close to 0.018, corresponding to the large gap in the cumulative frequency. Therefore, *ii16a1* has a very large number of nodes with lower connectivity than the average. Conversely, the node degree frequency of the random instance is regular and the highest peak in frequency is very close to the mean. Fig.5 and Fig.6 show the average and median error (number of unsatisfied clauses) respectively on *ii16a1* and *3sat1650* for PGSAT with different values of τ. Results are averaged over 500 trials. We first observe that also for the structured instance there exists an optimal value of τ. Nevertheless, despite the fact that the two instances have the same average connectivity, the optimal parallelism is higher for *ii16a1* than for *3sat1650*. We conjecture that the high number of nodes with low degree present in the instance *ii16a1* is the cause of higher optimal parallelism. We performed the same kind of experiments on the *logistics-a* instance, by comparing it with a random 3-SAT instance of the same size (828 variables) and normalized average connectivity $\overline{q} = 0.0275$ (instance *3sat828*). Fig.7 and Fig.8 show the respective node degree frequency. Note that the distribution of *logistics-a* is not regular and has a high peak at normalized node degree 0.0326. The results of PGSAT performance with respect to τ are plotted in Fig.9 and Fig.10. The results are analogous, but dual, to the previous ones on *ii16a1*. We can observe that also for this structured instance there is a value of τ leading to a minimum average error, but in this case the optimal parallelism for *logistics-a* is lower than that of the random instance. This difference may be explained by observing that the highest peak in *logistics-a* node degree frequency corresponds to a node degree higher than the average value, which characterizes the random instance.

We can conclude this section by considering that the results obtained clearly show that an optimal parallelism value exists also for structured instances. Nevertheless, not surprisingly, this value is strongly affected by the asymmetric frequency of node degree. In particular, we observe that the highest peaks location seem to be the most relevant characteristic influencing the optimal parallelism.

3 Morphing from random to structure

The results of the previous section show that for structured instances, characterized by asymmetric frequency distribution, the optimal parallelism is mostly affected by the highest peaks of node degree distribution. To investigate more deeply this difference, we generated instances with a controlled amount of *structure*, by means of a technique called *morphing* [1]. We generated instances gradually morphing from a satisfiable random 3-SAT instance to a structured one (namely, *ii16a1*). We observed[1] that the more similar the node degree frequency distributions of two instances (especially peaks location), the closer is their optimal parallelism degree.

[1] Due to lack of space, graphs are omitted. A longer version of this paper is available as LIA Technical report at `www.lia.deis.unibo.it/research/TechReport.html`.

4 Conclusion and future work

In this work we have compared the connectivity distribution of random and structured SAT instances and observed its impact on the parallelization of local search. We first showed that also in the case of structured instances there is a critical value of parallelism τ_{opt} that optimizes the average solution quality. Secondly, we observed that in structured instances τ_{opt} is strongly affected by the highest peaks in the node degree distribution. The definition of an automatic procedure to set τ to a near-optimal value is subject of ongoing work, along with the use of parallelization as a mechanism to adapt the balance between intensification and diversification.

We are currently exploring different ways of parallelizing local search, for example by probabilistic flips, such that each variable flips with a probability which depends on the increase of number of satisfied clauses. Moreover, we are studying the relations between SAT/MAXSAT instance connectivity and properties of the search landscape and we are configuring experiments to check whether a phase transition exists also in the case of PGSAT on SAT/MAXSAT problems.

There are several open issues to explore, like the exploitation of graph properties to define the subsets of strongly connected variables (instead of dividing variables at random), the investigation of the relation between criticality and parallelism and the introduction of noise in local search, the study of other graph characteristics, like diameter, clustering and *small world* properties [8].

Acknowledgments I thank Michela Milano for her support, helpful discussions and suggestions. I also thank the anonymous reviewers for useful comments.

References

1. I. P. Gent, H. H. Hoos, P. Prosser, and T. Walsh. Morphing: Combining structure and randomness. In *Proceedings of AAAI99*, pages 654–660, 1999.
2. H. Hoos and T. Stützle. SATLIB: An online resource for research on SAT. In I. Gent, H. van Maaren, and T. Walsh, editors, *SAT2000: Highlights of Satisfiability Research in the Year 2000*, pages 283– 292. IOS Press, 2000.
3. S.A. Kauffman and W. Macready. Technological evolution and adaptive organizations. *Complexity*, 26(2):26–43, March 1995.
4. W.G. Macready and S.A. Kauffman A.G. Siapas. Criticality and parallelism in combinatorial optimization. *Science*, 271:56–59, January 1996.
5. A. Roli. Criticality and parallelism in GSAT. *Electronic Notes in Discrete Mathematics*, 9, 2001.
6. A. Roli and C. Blum. Critical Parallelization of Local Search for MAX–SAT. In *AI*IA2001: Advances in Artificial Intelligence*, volume 2175 of *Lecture Notes in Artificial Intelligence*, pages 147–158. Springer, 2001.
7. B. Selman, H.J. Levesque, and D. Mitchell. A new method for solving hard satisfiability problems. In *Proceedings of the Tenth National Conference on Artificial Intelligence*, pages 440–446, 1992.
8. T. Walsh. Search in a small world. In *proceedings of IJCAI99*, pages 1172–1177, 1999.
9. T. Walsh. Search on high degree graphs. In *Proceedings of IJCAI-2001*, 2001.