A descriptive analysis of a landscape for a directed graph layout problem

Rémi Lehn, Pascale Kuntz, Bruno Pinaud

A Descriptive Analysis of a Landscape for a Directed Graph Layout Problem

Rémi Lehn * Pascale Kuntz * Bruno Pinaud *

* IRIN/EPUN, University of Nantes
Rue Christian Pauc 44300 Nantes, France
Email: {remi.lehn, pascale.kuntz, bruno.pinaud}@ireste.fr

1 Introduction

Generally speaking, the problem of drawing a graph G is set as a combinatorial optimization problem: producing a layout $L(G)$ of G on a given support according to a drawing convention (e.g. layered drawing) that optimizes some measurable aesthetics (e.g. arc-crossing). Numerous criteria lead to NP-complete problems, and aesthetics often conflict with each other. Hence, various heuristics from classical divide-and-conquer approaches to meta-heuristics have been developed (see the recent surveys in [2], [5]).

In particular, from the first papers of Groves et al. [4] and Kosak et al. [8], applications of genetic algorithms (GA) and evolutionary algorithms to graph drawing have known an increasing interest. Depending on the application fields, they deal either with undirected graphs ([7], [9], [18], [20]) or with directed graphs ([13], [14], [16], [23]). But, as often for GA applications, the interest of these works has been essentially justified by experimental results on test cases and their intrinsic relevance in comparison with other heuristics -in particular with tabu search [12], stochastic hill-climbing [19] and more recently GRASP ([11], [1])- sets still open questions.

To characterize the ability of a GA to solve a problem and to improve the search by introducing problem-based operators, the fitness landscape analysis has become an important research area in combinatorial optimization (e.g. [17], [15]). The properties of the landscape, such as the smoothness or the ruggedness, influence the performances of the search for the optimal solutions and several statistical measures, often based on a correlation analysis, have been developed in the literature to characterize their features (e.g. [24]). These measures have often been studied for standard GA (simple crossover and bitflip operator for mutation) and their calculus and interpretation for more complex problem-based GA operators is not direct. Moreover, it is well-known for some combinatorial optimization problems that the instance characteristics may vary a lot from one to another (e.g. [22]), and that consequently, the heuristic performances closely depend on the processed instances.

Thus, instead of carrying out an additional comparative study, we here favor a descriptive analysis of the fitness landscape. We focus on the arc-crossing minimization of a digraph layered drawing which is known to be a NP-complete problem [3]. This work is only a first stage to better understand the structural characteristics of the problem before applying global statistical measures on large size instances. Our long-term objective is to develop an efficient algorithm for the dynamic version of the problem (i.e. producing at each time t a layout $L(G_t)$ that satisfies the usual readability requirements and which remains close enough to the previous one $L(G_{t-1})$ to preserve the "user's mental map") which is, up to our knowledge, still outstanding, and for which GAs seem promising [10]

In this communication, we analyze the arc-crossing landscape for local transformations of the graph adapted from the so-called Sugiyama-heuristic [21] for drawing directed acyclic graphs. We first define the optima and their attracting sets for all the possible layouts for a given transformation of more than two thousands graphs of small size. Then, we resort to hill-climbing to explore the landscape structures of three hundred graphs of larger size. We describe the results obtained with a set of one thousand hill-climbers. And, at the moment, we are studying the structural properties of the graphs which characterize the landscapes [6].

2 Fitness landscape

We here restrict ourselves to the study of arc-crossing landscapes associated with some local transformations. Let $G = (V, E)$ be an acyclic digraph. In a layered drawing of G, vertices are arranged in given vertical layers on a plane and arcs are represented by line segments joining vertex pairs. Let $\Lambda = \{l_1, ..., l_K\}$ be a set of K layers for the layout and $\Omega^\Lambda(G)$ be the set of all layered drawing of G for a given distribution of the vertex set on these layers. The fitness value $f(L_i^G)$ of any drawing $L_i^G \in \Omega^\Lambda(G)$ is the number of arc-crossings on L_i^G.

The set $N(L_i^G)$ of the neighbors of L_i^G is composed of all layouts of $\Omega^\Lambda(G)$ which are deduced from L_i^G by applying a local operator $\mathbf{O}$ once, and we denote by $\Omega_\mathbf{O}(G)$ the set of all the layouts deduced from one another with $\mathbf{O}$. We here only retain local transformations which improve the fitness. A *local optimum* $L_i^G \in \Omega_\mathbf{O}(G)$ is s.t. $f(L_i^G) \leq f(L_j^G)$ for $L_j^G \in N(L_i^G)$ and a *global optimum* $\widehat{L}^G \in \Omega_\mathbf{O}(G)$ is s.t. $f(\widehat{L}^G) \leq f(L_i^G)$ for $L_i^G \in \Omega_\mathbf{O}(G)$. The *basin of attraction* $b(\widehat{L}^G)$ of $\widehat{L}^G$ is the set of layouts of $\Omega_\mathbf{O}(G)$ which can be reached from $\widehat{L}^G$ by a descent method: $b(\widehat{L}^G) = \{L_0^G; \exists\, L_1^G, ..., L_{n-1}^G \in \Omega_\mathbf{O}(G) \text{ with } L_i^G \in N(L_{i+1}^G) \text{ and } f(L_i^G) \leq f(L_{i+1}^G)\ \forall i\}$ where $L_n^G = \widehat{L}^G$. The relative height of a local optimum $h(L_i^G)$ is $1 - \frac{f(L_i^G) - f(\widehat{L}_G)}{f(L_w^G) - f(\widehat{L}_G)}$ where L_w^G is the worst layout. Intuitively, if $h(L_i^G)$ is very close to 1, then the local optimum L_i^G can be considered as a "good" solution.

We study three local operators which can be used for local optimization in stochastic or multi-start hill-climbing, as well as in GA combined with a bitflip operator to introduce a local improvement in the mutation phase. The *greedy switching operator* switchs adjacent pairs of vertices on a layer l_k when this exchange makes crossing number decrease. The *barycenter operator* repositions a vertex at the average position of its neighbors on the adjacent layers l_{k-1} and l_{k+1}. The average position $avg(v)$ (see [10] for details) is computed for each vertex v of l_k and the local transformation is the new vertex ordering σ'_k on l_k obtained after sorting the average values: $\sigma'_k(u) > \sigma'_k(v)$ if $avg(u) > avg(v)$ for u, v on l_k. The *median operator* acts similarly to the barycenter: it repositions a vertex at the median position of its neighbors on the adjacent layers l_{k-1} and l_{k+1}.

3 Experimental results

We here present several main features of the optima distribution on different graph sets.

3.1 Exhaustive exploration for small graphs

Let D be the set of acyclic digraphs different from trees s.t. each arc is incident to vertices on adjacent layers and s.t. the vertex number n_k on each layer l_k is bounded by $\prod_{k=1}^{K} n_k! \leq 2000$. For 2021 graphs from D, we have analyzed the fitness landscapes, associated with the greedy switching operator, for all possible layouts. The average number of explored layouts for each graph is 917.

Even for these small graphs, the fitness landscape is highly multimodal: 70% of the cases have a local optimum and the mean number of local (resp. global) optima is equal to 25.2 (resp. 92.9). There is no correlation between these two values ($\rho = -0.16$) and the global optimum number partly results from the numerous symmetries of the graphs. The local optima are relatively high: 58% have a height between 0.7 and 0.99.

Our results confirm the importance of the instance characteristics in combinatorial optimization. Here the distribution of the optima greatly varies from graph to graph: the standard deviation of the local (resp. global) optimum number is equal to 31.9 (resp. 264.5). We find large disparities again when studying the basin of attractions: 30% (resp. 70%) of the graphs with at least one local optima have a relative basin of attraction size less (resp. greater or equal) than 70%.

3.2 Exploration with multi-start hill-climbers

As the computation time quickly becomes prohibitive for an exhaustive exploration of larger graphs, we resort to an extended multi-start hill-climbing. In a first experiment, a set of one thousand layouts has been randomly generated, and each layout has been improved by an iterative application of an operator **O** (here the successive application on each layer of the three operators quoted above): O is applied on each layer of a random layout $L_i^G \in \Omega_{\mathbf{O}}^G$ taken one after the other and this loop starts again until the fitness stabilizes. The experiment is concerned with 300 graphs of various size (between 30 and 100 vertices and a maximum of 20 layers).

There are some differences with the previous results due to the graph sizes but the main tendencies are confirmed. Here, 72.14% of the graphs have at least a "local optimum" (i.e. a layout on which the hill-climbers converge different from the best reached solution), and on average, among the 1000 climbers a little more than the half reach a local optimum and a little less than the find a best solution. However, there are again very important differences between graphs: the distribution of the proportion of hill-climbers reaching a global optimum is spread out.

The number of hill-climbers chosen for this study has been empirically fixed after some experiments. The mean time computation is equal to 33mns.14s (wallclock time on a single workstation). In order to confirm or inflect the results, we are studying the variation of the results with an increasing set of hill-climbers. The objective is to give an approximation of the probability of reaching a global optimum depending on the landscape structures.

References

[1] CC.Ribeiro and M.G.C. Resende. Algorithm 797: Fortran subroutines for approximate solution of graph planarization problems using grasp. *ACM Trans. on Mathematical Software*, 25, 1999.

[2] G. Di-Battista, P. Eades, R. Tamassia, and I.-G. Tollis. *Graph drawing – Algorithms for the visualization of graphs.* Prentice-Hall, 1999.

[3] P. Eades and N. Wormald. Edge crossings in drawings of bipartite graphs. *Algorithmica*, 11:379–403, 1994.

[4] L.-J. Groves, Z. Michalewicz, P.-V. Elia, and C.-Z. Janikow. Genetic algorithms for drawing directed graphs. In *Proceedings of the 5^{th} Int. Symp. on Methodologies for Intelligent Systems*, pages 268–276. Elsevier, 1990.

[5] I. Herman, G. Melançon, and S. Marshall. Graph visualization and navigation in information visualization: a survey. *IEEE Trans. on Visualization and Computer Graphics*, 6(1):24–43, 2000.

[6] T. Jones and S. Forrest. Genetic algorithms and heuristic search. In *Santa Fe Institute Tech. Report 95-02-021.* Santa Fe Institute, 1995.

[7] D. Kobler and A. Tettamanzi. Recombination operators for evolutionary graph drawing. In *Proc. of Parallel Problem Solving from Nature PPSN-V*, volume 1498 of *Lect. Notes in Comp. Sc.*, pages 988–997. Springer-Verlag, 1998.

[8] C. Kosak, J. Marks, and S. Shieber. A parallel genetic algorithm for network-diagram layout. In *Proceedings of the 4^{th} International Conference on Generic Algorithms*, pages 458–465. Morgan-Kaufmann, 1991.

[9] C. Kosak, J. Marks, and S. Shieber. Automating the layout of network diagrams with specified visual organization. In *IEEE Transactions on System, Man and Cybernetics*, volume 24, pages 440–454, 1994.

[10] P. Kuntz, R. Lehn, and H. Briand. Dynamic rule graph drawing by genetic search. In *Proc. of the IEEE Int. Conf. on System Man and Cybernetics*, 2000.

[11] M. Laguna and R. Marti. Grasp and path relinking for 2-layer straight line crossing minimization. *INFORMS J. on Computing*, 11(1), 2001.

[12] M.R. Laguna, R. Marti, and V. Vals. Arc crossing minimization in hierarchical digraphs with tabu search. *Comp. and Operations Res.*, 24(12):1165–1186, 1997.

[13] E. Makinen and M. Sieranta. Genetic algorithms for drawing bipartite graphs. *Int. J. of Computer Mathematics*, 53(3-4):157–166, 1994.

[14] T. Masui. Graphic object layout with interactive genetic algorithms. In *Proc. of the 1992 IEEE Workshop on Visual Languages*, pages 74–80. IEEE Comp. Soc. Press, 1992.

[15] P. Merz and B. Freisleben. Fitness landscape and memetic algorithm design. In D. Corne, M. Dorigo, and F. Glover, editors, *New Ideas in Optimization*. McGraw-Hill, 1999.

[16] A. Ochoa-Rodrìguez and A. Rosete-Suàrez. Automatic graph drawing by genetic search. In *Proc. of the 11^{th} Int. Conf. on CAD, CAM, Robotics and Manufactories of the Future*, pages 982–987, 1995.

[17] C.R. Reeves. Landscapes, operators and heuristic search. *Annals of Operations Res.*, 86:473–490, 1999.

[18] A. Rosete and A. Ochoa. Genetic graph drawing. In *Proc. of the 13^{th} Int. Conf. on Appl. of Artificial Intelligence in Engineering*, 1998.

[19] A. Rosete-Suárez, A. Ochoa-Rodríguez, and M. Sebag. Automatic graph drawing and stochastic hill climbing. In *Proc. of GECCO'99*, volume 2, pages 1699–1706. Morgan Kaufmann, 1999.

[20] A. Rosete-Suarez, M. Sebag, and A. Ochoa-Rodriguez. A study of evolutionary graph drawing. Rapport de Recherche 1228, LRI, Université Paris-Sud XI, 1999.

[21] K. Sugiyama, S. Tagawa, and M. Toda. Methods for visual understanding of hierarchical systems. *IEEE Trans. Sys. Man and Cybernetics*, 11(2):109–125, 1981.

[22] E. Taillard. Comparison of iterative searches for the quadratic assignment problem. *Location Science*, 3:87–105, 1995.

[23] J. Utech, J. Branke, H. Schmeck, and P. Eades. An evolutionary algorithm for drawing directed graphs. In *Proc. of the Int. Conf. on Imaging Science, Systems and Technology*, pages 154–160. CSREA Press, 1998.

[24] V.K. Vassilev. *Fitness Landscapes and Search in the Evolutionary Design of Digital Circuits*. PhD thesis, Napier Univ., UK, 2000.