On Terrain Coverage Optimization by Using a Network Approach for Universal Graph-Based Data Mining and Knowledge Discovery

Michael Preuß[1], Matthias Dehmer[1], Stefan Pickl[1], and Andreas Holzinger[2,3]

[1] Institute for Theoretical Computer Science, Mathematics & Operations Research, University of the German Federal Armed Forces Munich, Germany
michael.preuss@unibw.de, stefan.pickl@unibw.de, matthias.dehmer@umit.at
[2] Research Unit HCI, Institute for Medical Informatics, Medical University Graz
[3] Institute for Information Systems and Computer Media, Graz University of Technology, Austria
a.holzinger@hci4all.at

Abstract. This conceptual paper discusses a graph-based approach for on-line terrain coverage, which has many important research aspects and a wide range of application possibilities, e.g in multi-agents. Such approaches can be used in different application domains, e.g. in medical image analysis. In this paper we discuss how the graphs are being generated and analyzed. In particular, the analysis is important for improving the estimation of the parameter set for the used heuristic in the field of route planning. Moreover, we describe some methods from quantitative graph theory and outline a few potential research routes.

1 Introduction

The on-line terrain coverage problem is very important and can be found in many different real world applications in diverse areas ranging from farming [1] to search and rescue [2].

There are a few research attempts on terrain coverage based on genetic algorithms, in particular ant-robots [3], and in [4] a simultaneous on-line coverage strategy for multi-robots is presented, which assures robust coverage of the surface regardless of the shape of the unknown environment. This is very interesting as ant-robots can cover terrain by leaving "markings" in the terrain, similar as in nature, and these markings can be sensed by all robots and allow them to cover the unknown terrain without direct communication with each other. Such approaches can be used for knowledge discovery and interactive data mining [5, 6].

By means of smart autonomous single agents or a swarm, these applications pursue the main objective to cover an unknown environment without any a priori information. For the multi-agent case this problem is known as *NP*-hard [7]. The coordination of multi-agent systems have been investigated extensively

D. Ślęzak et al. (Eds.): BIH 2014, LNAI 8609, pp. 564–573, 2014.

[8]. There are examples for the coordination of multi-agent systems in biology [9, 10] and physics [11] inspired approaches as well as economic based control models [12, 13].

While the agents visit each location at least once, they create a graph to represent the actual information of the environment. Through continuous sensing and data collection, the graph will change during the whole run time. To optimize the coverage process the agents try to find suitable routes based on the actual graph. There are established heuristics to solve the route planning problem [14–18]. The quality of the determined solutions depends on the used parameter set of the heuristic. Therefore it is necessary to analyze the available graph before. As a result we are able to estimate and adjust the heuristic parameters to find optimized routes.

This paper is structured as follows. Firstly, the general terrain coverage assumptions and an overview of the graph building process are presented. Secondly, we describe the optimization process from the current coverage to Terrain networks through to the routes. Thereafter an introduction of quantitative analysis and the measurement of graphs are presented. The last section summarizes the advantages of a continuous graph analysis in the field of terrain coverage. Furthermore we will outline potential applications.

2 Graph-Based Terrain Coverage Model

The terrain coverage problem can be described with the help of a graph. In general the range of the sensors will determine the size of a cell which is represented by a node. The environment which is blocked by an obstacle is not transferred into the graph. The possible crossings between adjacent cells are represented by the edges. Consequently an undirected graph $G = (V, E)$ with the costs c_{v_i,v_j} represents the environment. In our case the costs are encoded on the considered property like energy consumption. The cost matrix is defined on the edges $(v_i, v_j) \in E$. Furthermore the agents can move between two connected cells within one time step. While the agents are visiting a cell they can sense two different things. They can check all eight adjacent cells for obstacles. Besides the components can sense the cost function for the current cell. Only for two adjacent and sensed vertices we are able to determine the true costs to traverse the referring edge.

For the on-line terrain coverage problem there is limited information of the environment. While the agents traverse the environment they collect new data of costs and general connections between the cells. Therefore the individual graph of each agent is changing during the run time in a continuous way. In addition the individual graph is used to exchange the actual information. As a result there is a global graph representing the combined information of all agents.

The swarm of autonomous agents is self-coordinated and organized by an auction based approach [13], [19]. There are different advantages using an auction based model without a central coordinator, called planning and control agent. On the one hand a decentralized and more robust behavior is expected. There is for example no need of direct and permanent communication links. Besides failures by the planning and control agent would affect the swarm in a negative way. In the worst case the whole swarm would collapse. On the other hand for a self-coordinated auction based approach, the agents try to maximize their individual profits in an opportunistic way. Consequently the global efficiency is increased.

The exploration of the swarm can be organized by different approaches [13], [20, 21]. In general there are tasks representing parts of the unknown environment. Each agent uses the currently available and matched graph for determining possible routes. In previous researches [22] a multi-objective ant colony algorithm [23] is used to determine routes considering multiple objectives. For example the agents are able to find good routes which are as safe, short, robust, most informative and economical as possible at the same time. This is a significant improvement for real world applications. For further information on the described terrain coverage approach we recommend [19].

3 Optimization Process

Next we explain our conceptional optimization process for an autonomous multi-agent system. First of all each agent creates an individual graph representing the information about the sensed environment only. As long as an agent does not have a current task, auctions will be initiated by this agent. Every agent within the same communication network participates with a bid. Before the agents determine their routes, they exchange the individual graphs. The auctioneer agent merges the graphs by determining the union of known vertices. Besides some additional costs to traverse the edges can be added to the graph.

Next the matched graph will be used to estimate missing information of vertices and edges/costs for unknown parts of the environment. The transformation from the coverage to the matched Terrain network for two agents at $t = 5$ and $t = 21$ is shown in Figure 1. The estimation is necessary to enlarge the solution space of possible routes. There are different approaches using intervals [24], probability distributions [25] or Fuzzy Logic [26] to describe and determine costs under uncertainties [27]. Already known cells are used to estimate the unknown cells. For this purpose we use the method of diffusion. The closer an unknown cell is to a known cell, the higher is the probability that the costs are similar. Therefore, a diffusion function is introduced which describes the influence of known costs as a function of distance. The diffusion function can be either constant or non-constant. For example interval boundaries for the costs $c_{i,j} \in [\underline{b}_{i,j}, \overline{b}_{i,j}]$ of unknown edges are determined by the weighted mean of all diffused costs starting from known cells. The resulting graph called Terrain network is the basis of further graph analysis. In general a Terrain network is an undirected multigraph which is connected and weighted. In Section 4 we present quantitative graph measurements which can be used for analysis. Because of the analysis results concerning the characteristics, structure and complexity of the graph, the understanding and optimization is facilitated in general. Furthermore the results are used to define a good parameter set for heuristics which solve the route planning problem. In addition the analysis can be used to define an autonomous decision maker for multi-objective optimization. The decision maker evaluates the influences of the different objectives to select one of the compromise solutions. Considered objectives may be the minimization of the energy consumption, directional changes and route length as well as the maximization of the information content described by the number of unknown vertices. For example at the beginning of the coverage process the route length should attach more weight. While the coverage process the weight for routes with higher information content should be increased.

Finally the agents use the optimized parameter set and decision maker to find a good solution for the route planning problem. The agent with the best bid concerning the objective function wins the auction. In Figure 2 we visualize the previous described generic optimization process using networks.

4 Quantitative Analysis of Terrain Networks

Graph analysis is currently ubiquitous and has been inspired by interdisciplinary applications such as the World Wide Web [28, 29] and Network Biology [30–32]. Triggered by the hype dealing with the analysis of complex networks, it turned out that besides exploring random networks the analysis of non-random graphs is crucial too [33–35]. Finally this insight led to the term *complex networks* [33], [36] representing graphs whose network topology is neither regular nor random. Besides investigating the topology of graph classes such as random graphs, small world graphs and various complex networks, the quantitative analysis of networks has been proven useful [37–39]. Instead of only describing structural information of networks [40, 41], quantitative graph theory relates to quantify structural information by using a measurement approach.

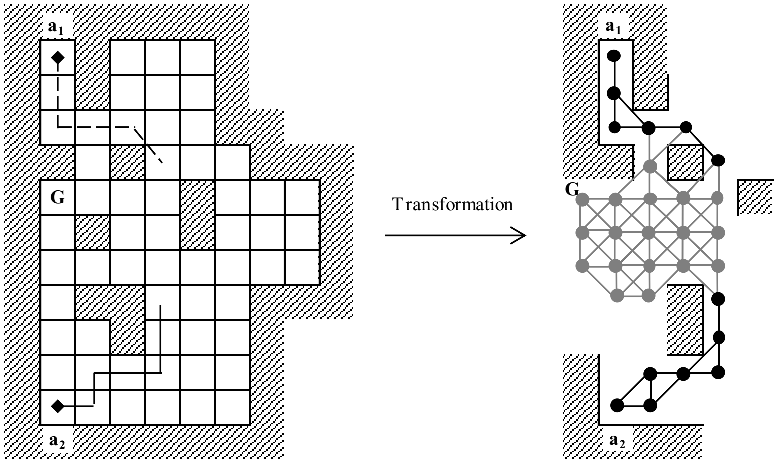
The simplest case is defining graph measures $M : \mathcal{G} \longrightarrow \mathbb{R}$ which capture structural information of the graphs. Those measures are called complexity measures [42, 43] that map graphs to the reals. Examples for simple graph complexity measures are the famous Wiener index and Randić index given by [44]

$$W(G) := \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} d(v_i, v_j) \tag{1}$$

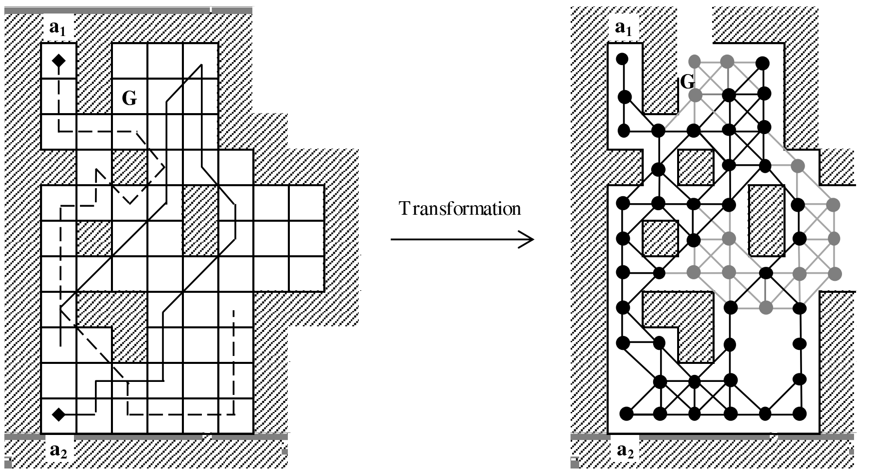
and

$$R(G) := \sum_{(v_i,v_j)\in E} [k_{v_i} k_{v_j}]^{-\frac{1}{2}}, \tag{2}$$

respectively. We define $G = (V, E)$ and $d(v_i, v_j)$ is the shortest distance between the $(v_i, v_j) \in V$. Furthermore k_{v_i} is the vertex degree of v_i.



(a)



(b)

Fig. 1. Terrain network transformation for a) t=5 and b) t=21. Estimated vertices and edges are in gray.

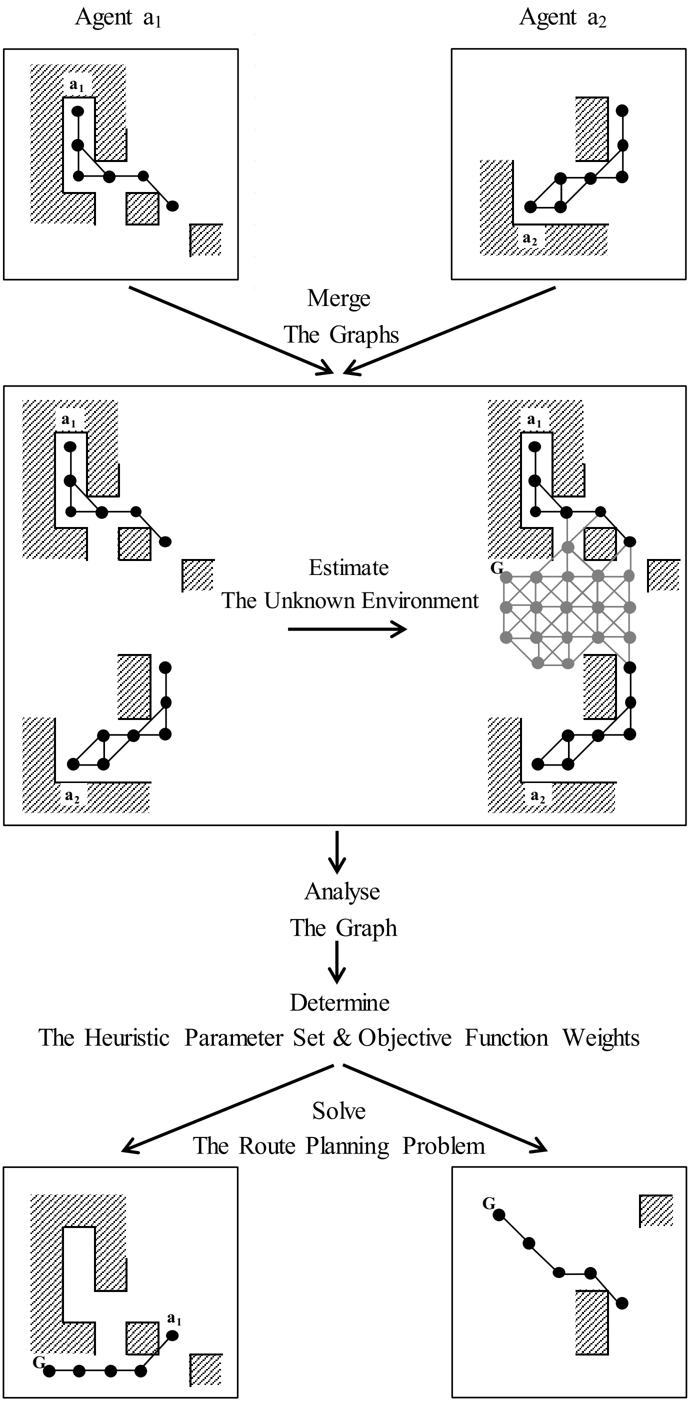


Fig. 2. Generic optimization process by using graph analysis of networks

In the future, we focus on measuring the complexity of Terrain networks by using known complexity measures [44]. For instance, it would be interesting whether these measures can fully discriminate the structure of the networks as it is likely that the networks are non-isomorphic. Note that the discrimination power of information-theoretic and non-information-theoretic graph measures have been investigated extensively [45–48]. Also, we aim to cluster the complexity values by using known techniques [49] and interpret those clusters. Then we get clusters which contain graphs for a given scenario. This leads to results how the graph may interrelate with each other.

Another branch of quantitative graph theory is graph comparison by using graph similarity/distance measures [50–52]. This relates to measure the structural similarity/distance between graphs which can be done by employing several paradigms. Exact graph matching [50–52] relates to determine isomorphic and subgraph isomorphic relations. In case the networks are large, the resulting measures may be inefficient. In case of our Terrain networks, we intend to use inexact graph matching that comprises the well-known graph edit distance (GED) [50] and various other measures, e.g., those which are based on using property strings [53, 54]. The Terrain networks can be classified by using supervised and unsupervised techniques. This would allow defining graph classes for each scenario and to determine their characteristic structural features.

5 Conclusions and Future Work

In the future we will further investigate structural features and the complexity of Terrain networks. In particular, we will compare the complexity of Terrain networks with those of other network classes and draw conclusions thereof. This approach is particularly interesting for knowledge discovery and data mining from natural images, e.g. complex biomedical images [55] where multi-agents, e.g. ant-robots can explore the image as an topological landscape and the autonomous robots leave markings on "interesting" spots, where these markings can be sensed by all robots and allow them to cover the unknown terrain without direct communication with each other, hence to discover anomalies, similarities or dissimilarities within such an image. Such approaches can also be useful for overcoming local optima problems in image segmentation, where such an approach takes advantage of random operators and multi-individual search algorithms, so that the best single agent tries to find a global solution [56]. During the autonomous agents covering the unknown terrain they have to make decisions on task allocation and route planning. First results show that there is a need to develop an autonomous process to determine a good parameter set for route planning heuristics, e.g. ant-colony optimization. Particularly, such approaches can be very beneficial when combined with evolutionary algorithms [57] which together have enormous potential in further research on graph-based data mining and knowledge discovery.

References

1. van Evert, F.K., van der Heijden, G.W.A.M., Lotz, L.A.P., Polder, G., Lamaker, A., de Jong, A., Kuyper, M.C., Groendijk, E.J.K., Neeteson, J.J., van der Zalm, T.: A mobile field robot with vision-based detection of volunteer potato plants in a corn crop. Weed Technology 20, 853–861 (2006)
2. Kumar, V., Rus, D., Singh, S.: Robot and sensor networks for first responders. IEEE Pervasive Computing 3, 24–33 (2004)
3. Dorigo, M., Maniezzo, V., Colorni, A.: Ant system: optimization by a colony of cooperating agents. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics 26, 29–41 (1996)
4. Senthilkumar, K.S., Bharadwaj, K.K.: Spanning tree based terrain coverage by multi robots in unknown environments. In: IEEE Annual IEEE INDICON Conference, pp. 120–125 (2008)
5. Holzinger, A., Ofner, B., Dehmer, M.: Multi-touch graph-based interaction for knowledge discovery on mobile devices: State-of-the-art and future challenges. In: Holzinger, A., Jurisica, I. (eds.) Knowledge Discovery and Data Mining. LNCS, vol. 8401, pp. 241–254. Springer, Heidelberg (2014)
6. Holzinger, A., Dehmer, M., Jurisica, I.: Knowledge discovery and interactive data mining in bioinformatics - state-of-the-art, future challenges and research directions. BMC Bioinformatics 15, I1 (2014)
7. Zheng, X., Koenig, S., Kempe, D., Jain, S.: Multirobot forest coverage for weighted and unweighted terrain. Transactions on Robotics 26, 1018–1031 (2010)
8. Olfati-Saber, R., Fax, J.A., Murray, R.M.: Consensus and cooperation in networked multi-agent systems. Proceedings of the IEEE 95, 215–233 (2007)
9. Arkin, R., Balch, T.: Cooperative multiagent robotic systems. In: Artificial Intelligence and Mobile Robots. MIT/AAAI Press (1998)
10. Wagner, I., Bruckstein, A.: From ants to a(ge)nts: A special issue on ant-robotics. Annals of Mathematics and Artificial Intelligence 31, 1–5 (2001)
11. Chevallier, D., Payandeh, S.: On kinematic geometry of multi-agent manipulating system based on the contact force information. In: Proceedings of the 6th International Conference on Intelligent Autonomous Systems (2000)
12. Gerkey, B., Mataric, M.: Sold!: auction methods for multirobot coordination. IEEE Transactions on Robotics and Automation 18, 758–768 (2002)
13. Zlot, R., Stentz, A., Dias, M., Thayer, S.: Multi-robot exploration controlled by a market economy. In: Proceedings of the IEEE International Conference on Robotics and Automation, vol. 3, pp. 3016–3023 (2002)
14. Dijkstra, E.W.: A note on two problems in connexion with graphs. Numerische Mathematik 1, 269–271 (1959)
15. Dorigo, M.: Optimization, Learning and Natural Algorithms. PhD thesis, Dipartimento di Elettronica, Politecnico di Milano, Milan, Italy (1992) (in Italian)
16. Floyd, R.W.: Algorithm 97: Shortest path. Communications of the ACM 5, 345 (1962)
17. Gen, M., Cheng, R., Wang, Q.: Genetic algorithms for solving shortest path problems. In: IEEE International Conference on Evolutionary Computation, pp. 401–406 (1997)
18. Hart, P.E., Nilsson, N.J., Raphael, B.: A formal basis for the heuristic determination of minimum cost paths. IEEE Transactions on Systems Science and Cybernetics 4, 100–107 (1968)

19. Preuß, M.: A multi-objective online terrain coverage approach. In: Proceedings of the International Conference on Operations Research. Springer (in print, 2014)
20. Hoog, J., Cameron, S., Visser, A.: Role-based autonomous multi-robot exploration. In: Proceedings of the International Conference on Advanced Cognitive Technologies and Applications (2009)
21. Ghoul, S., Hussein, A., Abdel-Wahab, M., Witkowski, U., Rückert, U.: A modified multiple depth first search algorithm for grid mapping using mini-robots khepera. Journal of Computing Science and Engineering 2, 321–338 (2008)
22. Preuß, M.: Terrain Coverage - Modelle und Algorithmen. Master's thesis, University of the German Federal Armed Forces Munich (2011)
23. Alaya, I., Solnon, C., Ghédira, K.: Ant colony optimization for multi-objective optimization problems. In: Proceedings of the IEEE International Conference on Tools with Artificial Intelligence, pp. 450–457 (2007)
24. Karasan, O., Pinar, M., Yaman, H.: The robust shortest path problem with interval data. Technical report, Bilkent University, Department of Industrial Engineering, Ankara (2001)
25. Bertsekas, D., Tsitsiklis, J.: An Analysis of Stochastic Shortest Path Problems. Mathematics of Operations Research 16 (1991)
26. Yao, J.S., Lin, F.T.: Fuzzy shortest-path network problems with uncertain edge weights. Journal of Information Science and Engineering 19, 329–351 (2003)
27. Sahinidis, N.: Optimization under uncertainty: state-of-the-art and opportunities. Computers & Chemical Engineering 28, 971–983 (2004); FOCAPO 2003 Special issue
28. Adamic, L., Huberman, B.: Power-law distribution of the world wide web. Science 287, 2115a (2000)
29. Chakrabarti, S.: Mining the Web: Discovering Knowledge from Hypertext Data. Morgan Kaufmann, San Francisco (2002)
30. Barabási, A.L., Oltvai, Z.N.: Network biology: Understanding the cell's functional organization. Nature Reviews. Genetics 5, 101–113 (2004)
31. Dehmer, M., Emmert-Streib, F., Graber, A., Salvador, A. (eds.): Applied Statistics for Network Biology. Quantitative and Network Biology. Wiley-Blackwell (2011)
32. Emmert-Streib, F., Dehmer, M. (eds.): Analysis of Microarray Data: A Network-based Approach. Wiley VCH Publishing (2010)
33. Dorogovtsev, S.N., Mendes, J.F.F.: Evolution of Networks. From Biological Networks to the Internet and WWW. Oxford University Press (2003)
34. Erdös, P., Rényi, P.: On the evolution of random graphs. Magyar Tud. Akad. Mat. Kutató Int. Közl 5, 17–61 (1960)
35. Watts, D.J., Strogatz, S.H.: Collective dynamics of 'small-world' networks. Nature 393, 440–442 (1998)
36. Estrada, E.: The Structure of Complex Networks. Theory and Applications. Oxford University Press (2011)
37. Dehmer, M., Emmert-Streib, F.: Quantitative Graph Theory. Theory and Applications. CRC Press (in press, 2014)
38. Mehler, A.: A quantitative graph model of social ontologies by example of wikipedia. In: Mehler, A., Sharoff, S., Rehm, G., Santini, M. (eds.) Genres on the Web: Computational Models and Empirical Studies. Springer (2009) (to appear)
39. Mehler, A.: Social ontologies as generalized nearly acyclic directed graphs: A quantitative graph model of social tagging. In: Dehmer, M., Emmert-Streib, F., Mehler, A. (eds.) Towards an Information Theory of Complex Networks: Statistical Methods and Applications, pp. 259–319. Birkhäuser, Boston/Basel (2011)

40. Halin, R.: Graphentheorie, Berlin, Germany. Akademie Verlag (1989)
41. Harary, F.: Graph Theory, Reading, MA, USA. Addison Wesley Publishing Company (1969)
42. Bonchev, D., Rouvray, D.H.: Complexity in Chemistry, Biology, and Ecology, New York, NY, USA. Mathematical and Computational Chemistry. Springer (2005)
43. Mowshowitz, A.: Entropy and the complexity of the graphs I: An index of the relative complexity of a graph. Bull. Math. Biophys. 30, 175–204 (1968)
44. Todeschini, R., Consonni, V., Mannhold, R.: Handbook of Molecular Descriptors, Weinheim, Germany. Wiley-VCH (2002)
45. Bonchev, D., Mekenyan, O., Trinajstić, N.: Isomer discrimination by topological information approach. J. Comp. Chem. 2, 127–148 (1981)
46. Dehmer, M., Emmert-Streib, F., Grabner, M.: A computational approach to construct a multivariate complete graph invariant. Inf. Sci. 260, 200–208 (2014)
47. Dehmer, M., Grabner, M., Varmuza, K.: Information indices with high discriminative power for graphs. PLoS One 7, e31214 (2012)
48. Konstantinova, E.V., Skorobogatov, V.A., Vidyuk, M.V.: Applications of information theory in chemical graph theory. Indian Journal of Chemistry 42, 1227–1240 (2002)
49. Jain, A.K., Dubes, R.C.: Algorithms for clustering data. Prentice-Hall, Inc., Upper Saddle River (1988)
50. Bunke, H.: Graph matching: Theoretical foundations, algorithms, and applications. In: Proceedings of Vision Interface 2000, pp. 82–88 (2000)
51. Sobik, F.: Graphmetriken und Klassifikation strukturierter Objekte. ZKI-Informationen, Akad. Wiss. DDR 2, 63–122 (1982)
52. Zelinka, B.: On a certain distance between isomorphism classes of graphs. Časopis pro p̌est. Mathematiky 100, 371–373 (1975)
53. Dehmer, M., Emmert-Streib, F.: Comparing large graphs efficiently by margins of feature vectors. Applied Mathematics and Computation 188, 1699–1710 (2007)
54. Dehmer, M., Mehler, A.: A new method of measuring similarity for a special class of directed graphs. Tatra Mountains Mathematical Publications 36, 39–59 (2007)
55. Holzinger, A., Malle, B., Bloice, M., Wiltgen, M., Ferri, M., Stanganelli, I., Hofmann-Wellenhof, R.: On the generation of point cloud data sets: the first step in the knowledge discovery process. In: Holzinger, A., Jurisica, I. (eds.) Interactive Knowledge Discovery and Data Mining in Biomedical Informatics. LNCS, vol. 8401, pp. 57–80. Springer, Heidelberg (2014)
56. Kasaiezadeh, A., Khajepour, A.: Multi-agent stochastic level set method in image segmentation. Computer Vision and Image Understanding 117, 1147–1162 (2013)
57. Holzinger, K., Palade, V., Rabadan, R., Holzinger, A.: Darwin or lamarck? future challenges in evolutionary algorithms for knowledge discovery and data mining. In: Holzinger, A., Jurisica, I. (eds.) Knowledge Discovery and Data Mining. LNCS, vol. 8401, pp. 35–56. Springer, Heidelberg (2014)