Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2015, Article ID 680624, 6 pages
http://dx.doi.org/10.1155/2015/680624

Research Article

Linkage-Based Distance Metric in the Search Space of Genetic Algorithms

Yong-Hyuk Kim[1] and Yourim Yoon[2]

[1]*Department of Computer Science & Engineering, Kwangwoon University, 20 Kwangwoon-ro, Nowon-gu, Seoul 139-701, Republic of Korea*
[2]*Department of Computer Engineering, College of Information Technology, Gachon University, 1342 Seongnam-daero, Sujeong-gu, Seongnam-si, Gyeonggi-do 461-701, Republic of Korea*

Correspondence should be addressed to Yourim Yoon; yryoon@gachon.ac.kr

Received 31 July 2014; Accepted 7 September 2014

Academic Editor: Shifei Ding

Copyright © 2015 Y.-H. Kim and Y. Yoon. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We propose a new distance metric, based on the linkage of genes, in the search space of genetic algorithms. This second-order distance measure is derived from the gene interaction graph and first-order distance, which is a natural distance in chromosomal spaces. We show that the proposed measure forms a metric space and can be computed efficiently. As an example application, we demonstrate how this measure can be used to estimate the extent to which gene rearrangement improves the performance of genetic algorithms.

1. Introduction

Distance metrics are fundamental tools for organizing search spaces, because the introduction of a metric is the simplest way to induce a topology [1]. Different metrics produce different topologies and thus change the shape of the search space. When a space is to be searched by a genetic algorithm (GA), a good distance metric facilitates navigation of the space [2–5] and can also improve the effectiveness of search [6–12]. Hamming distance is a popular metric in a discrete space that is to be searched by a GA. Hamming distance has also been widely used in analyses of solution spaces [13–15].

Fitness distance correlation (FDC), proposed by Jones and Forrest [14], is a measure of the effectiveness of a distance metric in a space to be searched by a GA. An FDC is obtained by measuring the correlation between fitness and the distance to the nearest global optimum for a number of sample solutions. FDC coefficients range from −1 to 1, where higher values suggest increased difficulty in maximizing fitness and decreased difficulty in minimizing fitness. When a GA is hybridized with a local optimization, the population consists entirely of local optima, and it is then more useful to determine FDCs of local-optimum spaces.

In this paper, we propose a new distance measure which takes account of gene interaction and show that it forms a metric space. We use this metric to compute FDCs of search space and show that FDCs obtained in this way have improved correlation with the improvement in GA performance that can be obtained by gene rearrangement. The remainder of this paper is organized as follows. In Section 2, we review gene rearrangement in GAs. In Section 3, we propose a new distance measure for GAs, show that it forms a metric space, and demonstrate an application. Finally, we draw conclusions in Section 4.

2. Gene Rearrangement

Holland's schema theorem [16] shows that schemata (i.e., groups of genes) with high fitness, short defining length, and low order have high probabilities of survival in a standard GA.

These durable schemata are called *building blocks*. They make a major contribution to fitness and have a high degree of mutual interaction. The performance of a GA is strongly dependent on the survival and reproduction of these building blocks.

The survival probability of a gene group through a crossover is strongly affected by the positions of genes in the chromosome. Schemata consisting of genes in scattered positions tend to be too long to survive. Thus, the strategy used for placing genes significantly affects the performance of a GA. Inversion is an operator which changes the location of genes while a GA is running [17], and the process of rearranging genes dynamically to improve performance is called *linkage learning* [18]. Messy GA [19] is an example of a technique that implicitly uses dynamic gene rearrangement.

It has been observed that the performance of GAs on problems with a locus-based encoding can be improved by rearranging the indices of the genes before running the GA. Static gene rearrangement was first suggested by Bui and Moon [20, 21], who rearrange genes within a chromosomal representation to improve the quality of schemata and to help the GA to preserve the better schemata. Many studies on the static rearrangement of gene positions [20–24] have showed performance improvements. However, the improvement in performance achieved in this way has been shown to vary greatly between problem instances. This motivated us to develop a distance metric to improve our ability to estimate how much improvement in the performance of a GA on a particular problem instance can be expected through gene rearrangement.

3. A Linkage-Based Distance Measure

3.1. Second-Order Distance Measure. The most usual first-order distance measure in discrete space is the Hamming distance which is also a natural distance in chromosomal space, although there are other first-order distance measures, such as the quotient metric in redundant encoding [11]. We now define a second-order distance measure derived from first-order distance. Given a problem instance p, consider the unweighted undirected graph G_p representing first-order gene interaction [23], which is the pairwise interaction of genes. For convenience, we will assume that each gene has an interaction with itself, so that $\{g, g\} \in E(G_p)$ for each gene $g \in V(G_p)$. Let A_p be the adjacency matrix of G_p and consider A_p as a binary matrix over $\mathbb{Z}_2$ [25–27].

Definition 1. Suppose that the inverse of A_p exists as a binary matrix over $\mathbb{Z}_2$; that is, $A_p \in GL_n(\mathbb{Z}_2)$. One defines the second-order distance measure $d_p^{(2)}$ as follows:

$$d_p^{(2)}(x, y) := \left\| A_p^{-1}(x \oplus y) \right\|, \tag{1}$$

where $\oplus$ is a vector summation operator, which performs a Boolean XOR (i.e., $0 + 0 = 0$, $0 + 1 = 1$, $1 + 0 = 1$, and $1 + 1 = 0$) in each coordinate, and $\|\cdot\|$ is a norm derived from the first-order distance metric $d^{(1)}$ (i.e., $\|\cdot\| = d^{(1)}(\cdot, 0)$).

Theorem 2. *$d_p^{(2)}$ is a metric.*

Proof. It is enough to show the following four conditions [1].

(i) Nonnegativity: since $d_p^{(2)}(x, y) = d^{(1)}(A_p^{-1}(x \oplus y), 0)$ and $d^{(1)}$ is a metric, $0 \le d_p^{(2)}(x, y) < \infty$ for all x and y in X.

(ii) Identity of indiscernibles: consider

$$\begin{aligned} d_p^{(2)}(x, y) = 0 &\iff \left\| A_p^{-1}(x \oplus y) \right\| = 0 \\ &\iff A_p^{-1}(x \oplus y) = 0 \\ &\iff x \oplus y = 0 \\ &\iff x = y. \end{aligned} \tag{2}$$

(iii) Symmetry: consider

$$\begin{aligned} d_p^{(2)}(x, y) &= \left\| A_p^{-1}(x \oplus y) \right\| \\ &= \left\| A_p^{-1}(y \oplus x) \right\| \\ &= d_p^{(2)}(y, x). \end{aligned} \tag{3}$$

(iv) Triangle inequality: consider

$$\begin{aligned} & d_p^{(2)}(x, y) + d_p^{(2)}(y, z) \\ &= \left\| A_p^{-1}(x \oplus y) \right\| + \left\| A_p^{-1}(y \oplus z) \right\| \\ &\ge \left\| A_p^{-1}(x \oplus y) \oplus A_p^{-1}(y \oplus z) \right\| \\ &= \left\| A_p^{-1}((x \oplus y) \oplus (y \oplus z)) \right\| \\ &= \left\| A_p^{-1}((x \oplus z) \oplus (y \oplus y)) \right\| \\ &= \left\| A_p^{-1}(x \oplus z) \right\| \\ &= d_p^{(2)}(x, z). \end{aligned} \tag{4}$$

□

If the inverse of A_p does not exist, we can extend the scope of the distance metric using the following well-defined formulation:

$$d_p^{(2)}(x, y) := \min \left\| \left(\arg\min_z \left\| (x \oplus y) \oplus A_p z \right\| \right) \right\|. \tag{5}$$

We note that if the inverse of A_p exists, then $z := A_p^{-1}(x \oplus y)$, which implies $(x \oplus y) \oplus A_p z = 0$, and hence $\arg\min_z \|(x \oplus y) \oplus A_p z\| = A_p^{-1}(x \oplus y)$. Our second-order distance and its extension can be computed in $O(n^3)$ by a variant of Gauss-Jordan elimination [28], where n is the number of genes.

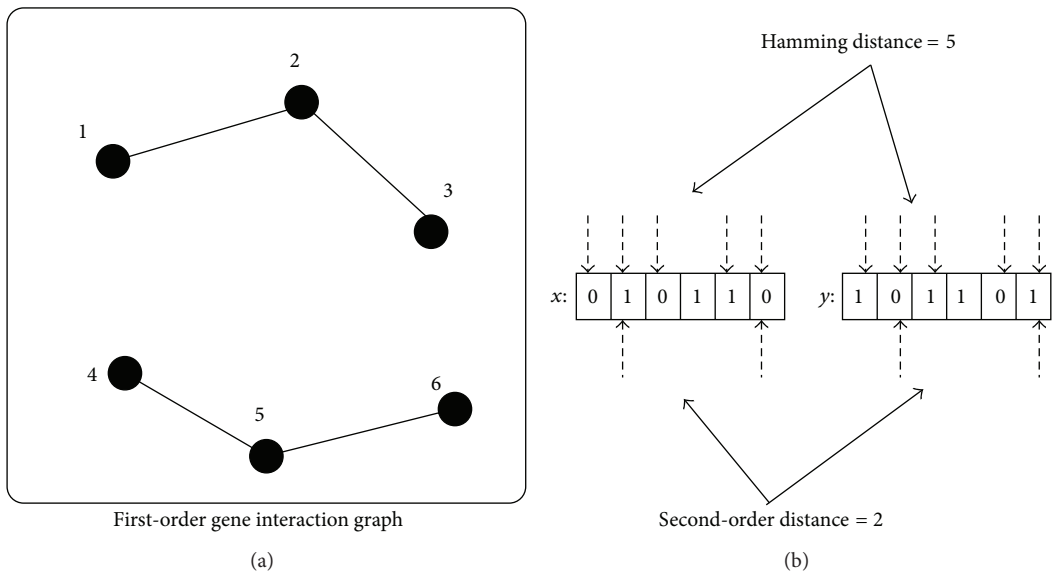


Figure 1: (a) An example of a first-order gene interaction graph G_p and (b) distances between two example chromosomes x and y.

3.2. An Application. Intuitively, our measure of the distance between two chromosomes can be understood as the minimum number of bits that must be changed to transform one chromosome into the other in the genetic process using optimal gene rearrangement.

Given an undirected graph $G = (V, E)$ with edge weights $(w_{ij})_{(i,j)\in E}$, the max-cut problem is that of finding a subset $S \subset V$ which maximizes the sum of the edge weights which traverse the cut $(S, V \setminus S)$ [29–31]. Consider the 6-node max-cut problem instance p, which is to maximize the following expression:

$$x_1 \oplus x_2 + x_2 \oplus x_3 - x_4 \oplus x_5 - x_5 \oplus x_6, \tag{6}$$

where a vertex v_i belongs to the position $x_i \in \{0, 1\}$ and $\oplus$ is the Boolean XOR operator. In this problem instance, edges $\{v_1, v_2\}$ and $\{v_2, v_3\}$ increase the fitness and edges $\{v_4, v_5\}$ and $\{v_5, v_6\}$ reduce the fitness. In the max-cut problem, we can consider that the given graph removing edge weights shows the first-order gene interaction (see, e.g., Figure 1(a)). Figure 1(b) shows an example in which the Hamming and second-order distances between two chromosomes x and y are obtained by optimal gene arrangement of the gene interaction graph G_p. In this example,

$$A_p = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix},$$

$$A_p^{-1} = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 \end{pmatrix}, \tag{7}$$

$x \oplus y = (1\ 1\ 1\ 0\ 1\ 1)^T$, $A_p^{-1}(x \oplus y) = (0\ 1\ 0\ 0\ 0\ 1)^T$, and hence $\|A_p^{-1}(x \oplus y)\| = 2$. If we use the normalized Hamming distance (developed for the 2-grouping problem) [32, 33] as the first-order distance measure, the FDC of this problem is −0.50. But when our second-order distance is used, the FDC becomes −0.95.

Given a graph $G = (V, E)$ and its adjacency matrix $A = (a_{ij})$, the graph bipartitioning problem is that of minimizing the following expression:

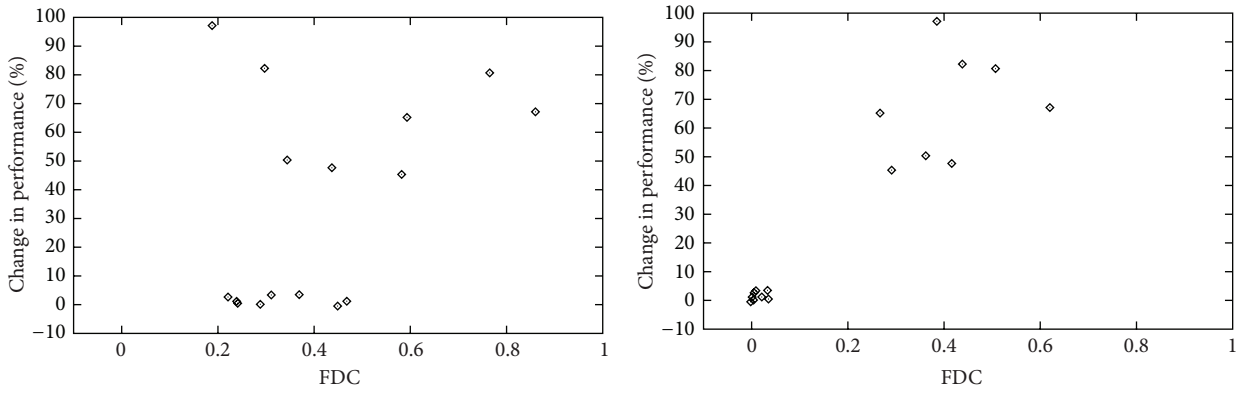
$$\frac{1}{2}\sum_{i=1}^{|V|}\sum_{j=1}^{|V|} a_{ij}\left(x_i \oplus x_j\right) + \gamma\left(\sum_{i=1}^{|V|} x_i - \frac{|V|}{2}\right)^2, \tag{8}$$

where $a_{ij} \in \{0, 1\}$, a vertex v_i belongs to the position $x_i \in \{0, 1\}$, and γ is a positive constant introduced to penalize unbalanced partitions. If we ignore the second balancing term altogether, we can regard the given graph as the first-order gene interaction graph of the given problem instance. Bui and Moon [21] tried gene rearrangement in a GA for graph bipartitioning and obtained dramatic improvements in performance for some graphs. We hypothesized that FDCs calculated using our second-order distance would help identify graphs that could benefit most from gene rearrangement, in terms of GA performance. Figure 2 shows the relationship

Table 1: Effect of gene rearrangement on FDCs computed using first- and second-order distance.

Graph	FDC with $d^{(1)}$	FDC with $d_p^{(2)}$	Improvement (%)† from [21]
G500.2.5	0.369	0.033	3.495
G500.05	0.449	−0.002	−0.487
G500.10	0.221	0.005	2.674
G500.20	0.288	0.004	0.117
G1000.2.5	0.241	0.035	0.469
G1000.05	0.239	0.001	1.167
G1000.10	0.311	0.009	3.362
G1000.20	0.468	0.021	1.201
U500.05	0.297	0.438	82.258
U500.10	0.437	0.416	47.649
U500.20	0.593	0.267	65.198
U500.40	0.860	0.620	67.143
U1000.05	0.188	0.385	97.144
U1000.10	0.344	0.362	50.340
U1000.20	0.582	0.291	45.341
U1000.40	0.765	0.507	80.668

†Change in GA performance obtained by gene rearrangement.



(a) First-order distance $d^{(1)}$ (b) Second-order distance $d_p^{(2)}$

Figure 2: Correlation of gene rearrangement with FDC values computed using first- and second-order distance.

between FDC and the performance improvement of a GA on 16 benchmark graphs (8 random graphs and 8 random geometric graphs) that were used in [34–40].

Here, the performance improvement means the difference in percentage between the average performances of a GA with and without gene rearrangement (data from [21]). The FDC values were approximated from 10,000 randomly generated local optima. When the first-order (normalized Hamming) distance was used, there was little correlation with the change in performance, but our second-order distance provided a clear correlation (see Figure 2(b) and Table 1).

4. Concluding Remarks

In most previous work, distances among chromosomes in GAs have usually been first-order distances, and in particular Hamming distance. We have proposed a second-order distance measure for GAs, which we consider to be more meaningful. We have showed that this distance measure forms a metric space and that it can be computed efficiently.

Using second-order distance allows us to see problem spaces from a different viewpoint. We have demonstrated its value in predicting the effectiveness of gene rearrangement, and we envisage it providing further understanding of the working mechanism of GAs.

Disclosure

A preliminary version of this paper appeared in the *Proceedings of the Genetic and Evolutionary Computation Conference*, pp. 1393–1399, 2005.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

This research was supported by the Gachon University research fund of 2014 (GCU-2014-0121).

References

[1] D. W. Kahn, *Topology: An Introduction to the Point-set and Algebraic Areas*, Dover Publications, New York, NY, USA, 1995.

[2] Y.-H. Kim and B.-R. Moon, "New usage of Sammon's mapping for genetic visualization," in *Proceedings of the Genetic and Evolutionary Computation Conference*, pp. 1136–1147, 2003.

[3] A. Moraglio and R. Poli, "Topological interpretation of crossover," in *Proceedings of the Genetic and Evolutionary Computation Conference*, vol. 1, pp. 1377–1388, 2004.

[4] M. Wineberg and F. Oppacher, "Distance between populations," in *Proceedings of the Genetic and Evolutionary Computation Conference*, pp. 1481–1492, 2003.

[5] Y. Yoon and Y.-H. Kim, "Geometricity of genetic operators for real-coded representation," *Applied Mathematics and Computation*, vol. 219, no. 23, pp. 10915–10927, 2013.

[6] S.-S. Choi and B.-R. Moon, "Normalization in genetic algorithms," in *Proceedings of the Genetic and Evolutionary Computation Conference*, pp. 862–873, 2003.

[7] Y.-H. Kim, A. Moraglio, A. Kattan, and Y. Yoon, "Geometric generalisation of surrogate model-based optimisation to combinatorial and program spaces," *Mathematical Problems in Engineering*, vol. 2014, Article ID 184540, 10 pages, 2014.

[8] Y.-H. Kim, A. Moraglio, Y. Yoon, and B.-R. Moon, "Geometric crossover for multiway graph partitioning," in *Proceedings of the Genetic and Evolutionary Computation Conference*, pp. 1217–1224, July 2006.

[9] A. Moraglio, Y.-H. Kim, Y. Yoon, and B.-R. Moon, "Geometric crossovers for multiway graph partitioning," *Evolutionary Computation*, vol. 15, no. 4, pp. 445–474, 2007.

[10] Y. Yoon and Y.-H. Kim, "An efficient genetic algorithm for maximum coverage deployment in wireless sensor networks," *IEEE Transactions on Cybernetics*, vol. 43, no. 5, pp. 1473–1483, 2013.

[11] Y. Yoon, Y.-H. Kim, A. Moraglio, and B.-R. Moon, "Quotient geometric crossovers and redundant encodings," *Theoretical Computer Science*, vol. 425, pp. 4–16, 2012.

[12] Y. Yoon, Y.-H. Kim, A. Moraglio, and B.-R. Moon, "A theoretical and empirical study on unbiased boundary-extended crossover for real-valued representation," *Information Sciences*, vol. 183, pp. 48–65, 2012.

[13] K. D. Boese, A. B. Kahng, and S. Muddu, "A new adaptive multi-start technique for combinatorial global optimizations," *Operations Research Letters*, vol. 16, no. 2, pp. 101–113, 1994.

[14] T. Jones and S. Forrest, "Fitness distance correlation as a measure of problem difficulty for genetic algorithms," in *Proceedings of the 6th International Conference on Genetic Algorithms*, pp. 184–192, Pittsburgh, Pa, USA, July 1995.

[15] P. Merz and B. Freisleben, "Fitness landscape analysis and memetic algorithms for the quadratic assignment problem," *IEEE Transactions on Evolutionary Computation*, vol. 4, no. 4, pp. 337–352, 2000.

[16] J. Holland, *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Ann Arbor, Mich, USA, 1975.

[17] J. Bagley, *The behavior of adaptive systems which employ genetic and correlation algorithms [Ph.D. thesis]*, University of Michigan, Ann Arbor, Mich, USA, 1967.

[18] G. R. Harik and D. E. Goldberg, "Learning linkage," in *Foundations of Genetic Algorithms*, vol. 4, pp. 247–262, Morgan Kaufmann, San Francisco, Calif, USA, 1996.

[19] D. E. Goldberg, B. Korb, and K. Deb, "Messy genetic algorithms: motivation, analysis, and first results," *Complex Systems*, vol. 3, no. 5, pp. 493–530, 1989.

[20] T. N. Bui and B.-R. Moon, "Hyperplane synthesis for genetic algorithms," in *Proceedings of the 5th International Conference on Genetic Algorithms*, pp. 102–109, July 1993.

[21] T. N. Bui and B. R. Moon, "Genetic algorithm and graph partitioning," *IEEE Transactions on Computers*, vol. 45, no. 7, pp. 841–855, 1996.

[22] T. N. Bui and B. R. Moon, "New genetic approach for the Traveling salesman problem," in *Proceedings of the 1st IEEE Conference on Evolutionary Computation*, pp. 7–12, June 1994.

[23] Y.-H. Kim, Y.-K. Kwon, and B.-R. Moon, "Problem-independent schema synthesis for genetic algorithms," in *Proceedings of the Genetic and Evolutionary Computation Conference*, pp. 1112–1122, Chicago, Ill, USA, July 2003.

[24] B.-R. Moon and C. K. Kim, "A two-dimensional embedding of graphs for genetic algorithms," in *Proceedings of the International Conference on Genetic Algorithms*, pp. 204–211, 1997.

[25] Y.-H. Kim and K. Seo, "Two congruence classes for symmetric binary matrices over $\mathbb{F}_2$," *WSEAS Transactions on Mathematics*, vol. 7, no. 6, pp. 339–343, 2008.

[26] Y.-H. Kim and Y. Yoon, "Effect of changing the basis in genetic algorithms using binary encoding," *KSII Transactions on Internet and Information Systems*, vol. 2, no. 4, pp. 184–193, 2008.

[27] Y. Yoon and Y.-H. Kim, "A mathematical design of genetic operators on $GL_n(\mathbb{Z}_2)$," *Mathematical Problems in Engineering*, vol. 2014, Article ID 540936, 8 pages, 2014.

[28] M. Anderson and T. Feil, "Turning lights out with linear algebra," *Mathematics Magazine*, vol. 71, no. 4, pp. 300–303, 1998.

[29] S.-H. Kim, Y.-H. Kim, and B.-R. Moon, "A hybrid genetic algorithm for the max cut problem," in *Proceedings of the Genetic and Evolutionary Computation Conference*, pp. 416–423, 2001.

[30] K. Seo, S. Hyun, and Y.-H. Kim, "A spanning tree-based encoding of the MAX CUT problem for evolutionary search," in *Proceedings of the International Conference on Parallel Problem Solving from Nature*, vol. 7491 of *Lecture Notes in Computer Science*, pp. 510–518, 2012.

[31] K. Seo, S. Hyun, and Y.-H. Kim, "An edge-set representation based on spanning tree for searching cut space," *IEEE Transactions on Evolutionary Computation*, 2014.

[32] Y.-H. Kim and B.-R. Moon, "Investigation of the fitness landscapes and multi-parent crossover for graph bipartitioning," in *Genetic and Evolutionary Computation—GECCO 2003*, vol. 2723 of *Lecture Notes in Computer Science*, pp. 1123–1135, Springer, Berlin, Germany, 2003.

[33] Y.-H. Kim and B.-R. Moon, "Investigation of the fitness landscapes in graph bipartitioning: an empirical study," *Journal of Heuristics*, vol. 10, no. 2, pp. 111–133, 2004.

[34] I. Hwang, Y.-H. Kim, and B.-R. Moon, "Multi-attractor gene reordering for graph bisection," in *Proceedings of the 8th Annual Genetic and Evolutionary Computation Conference*, pp. 1209–1215, July 2006.

[35] D. S. Johnson, C. R. Aragon, L. A. McGeoch, and C. Schevon, "Optimization by simulated annealing: an experimental evaluation, part I. Graph partitioning," *Operations Research*, vol. 37, no. 6, pp. 865–892, 1989.

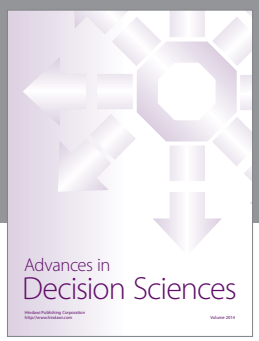
[36] Y.-H. Kim, "An enzyme-inspired approach to surmount barriers in graph bisection," in *Proceedings of the International Conference on Computational Science and Its Applications*, vol. 5072 of *Lecture Notes in Computer Science*, pp. 841–851, 2008.

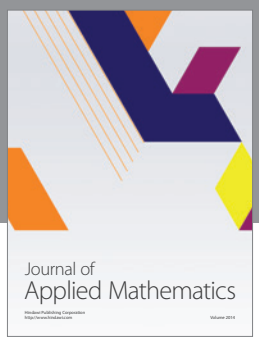
[37] Y.-H. Kim and B.-R. Moon, "A hybrid genetic search for graph partitioning based on lock gain," in *Proceedings of the Genetic and Evolutionary Computation Conference*, pp. 167–174, 2000.

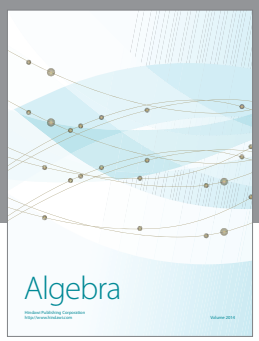
[38] Y.-H. Kim and B.-R. Moon, "Lock-gain based graph partitioning," *Journal of Heuristics*, vol. 10, no. 1, pp. 37–57, 2004.

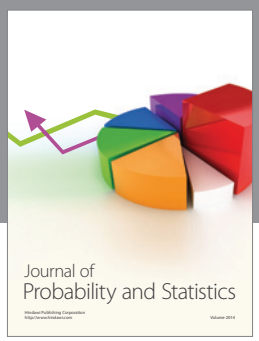
[39] Y. Yoon and Y.-H. Kim, "New bucket managements in iterative improvement partitioning algorithms," *Applied Mathematics and Information Sciences*, vol. 7, no. 2, pp. 529–532, 2013.

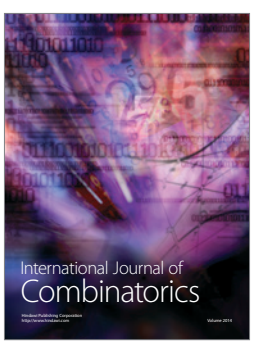
[40] Y. Yoon and Y.-H. Kim, "Vertex ordering, clustering, and their application to graph partitioning," *Applied Mathematics and Information Sciences*, vol. 8, no. 1, pp. 135–138, 2014.

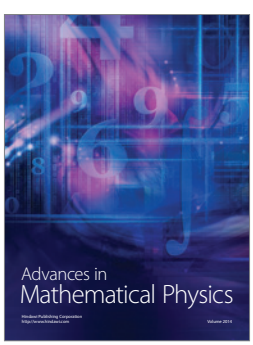


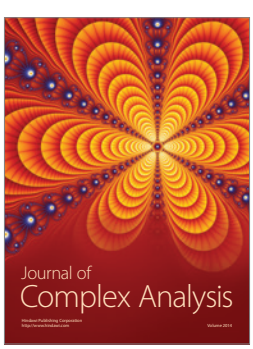


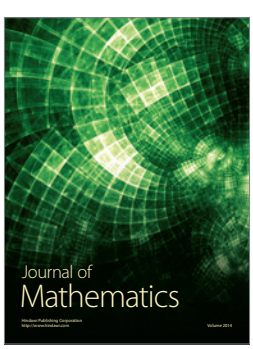


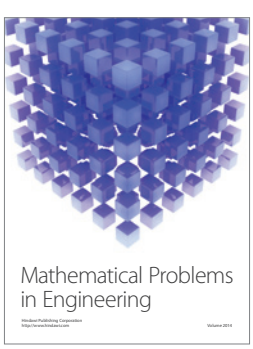


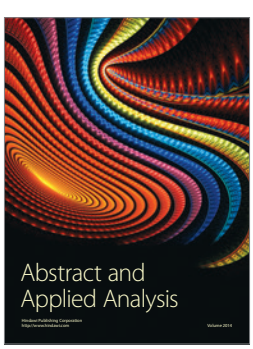


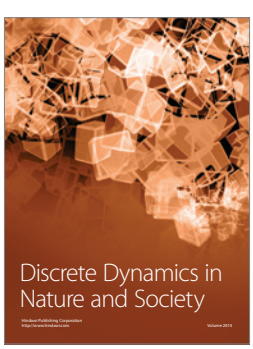


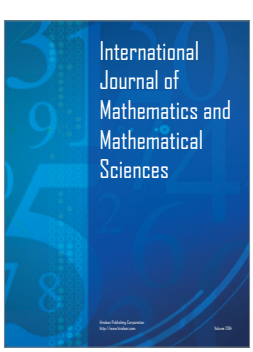












Hindawi

Submit your manuscripts at
http://www.hindawi.com