A new parallel algorithm to solve one classic water resources optimal allocation problem based on inspired computational model

Zuwen Ji\textsuperscript{a}, Zhaocai Wang\textsuperscript{a,b,*}, Anjun Deng\textsuperscript{a}, Wei Huang\textsuperscript{c}, Tunhua Wu\textsuperscript{d,*}

\textsuperscript{a}State Key Laboratory of Simulation and Regulation of River Basin Water Cycle, China Institute of Water Resources and Hydropower Research, Beijing 100048, China, Tel. +86 10 68786631, Fax +86 10 68786631, email: jzw@iwhr.com, zuwenji70@163.com (Z. Ji), anjdeng@163.com (A. Deng)

\textsuperscript{b}College of Information, Shanghai Ocean University, Shanghai 201306, China, Tel. +86 21 6190062, Fax +86 21 61900624, email: zcwang1028@163.com (Z. Wang)

\textsuperscript{c}College of Mathematics and Information Science, Guiyang University, Guizhou, China, Tel. +86 851 85405556, Fax +86 851 85405556, email: 2533870668@qq.com (W. Huang)

\textsuperscript{d}First Affiliated Hospital, Wenzhou Medical University, Wenzhou 325035, China, Tel. +86 577 86689913, Fax +86 577 86689913, email: appll188@163.com (T. Wu)

Received 23 January 2019; Accepted 10 May 2019

**Abstract**

The shortage of freshwater resources is a serious problem in the process of urbanization in the world, and even become the main constraint factor in some areas. In order to solve the problem, it is imperative to develop an effective, flexible and low-cost water resources management plan. Water resources optimal allocation is one of the hot topics in the field of water resources at present, such as water resources optimal \( k \)-edge cover problem. The \( k \)-edge cover problem aims to find an edge cover set with \( k \) edges in a given undirected graph. The efficient solution of this problem can play an important role in planning and setting up urban water resources network sites. Based on DNA molecular computing, the paper use a new parallel algorithm to solve \( k \)-edge cover problem with \( O(n^2) \) time complexity, which greatly simplifies the computing complexity.

**Keywords:** DNA computation; Water resources optimal allocation problem; The \( k \)-edge cover problem; Adleman-Lipton model; NP-complete problem

1. Introduction

The problem of water resources optimal allocation is theoretically a multi-objective stochastic sequential decision making problem. It is to utilize the effective, fair and sustainable principles, to allocate limited and different forms of water resources scientifically and reasonably through a variety of measures in a specific basin or area. The water resources optimal allocation is the basis of the water resources rational utilization and the fundamental guarantee for the water resources sustainable utilization.

In fact, the water resources rational allocation, in a broad sense, is a study of how to use good water resources, including the development, utilization, protection and management of water resources. It is more urgent to carry out water resources rational allocation. The main reasons are as below: first, the natural spatial and temporal distribution of water resources does not adapt to the distribution of productive forces. Secondly, there have lots of water competitions between the regions and the various water use departments, and thirdly, the water resources development and utilization have caused many ecological environmental problems.

The rational allocation of water resources is a comprehensive system consisting of various measures. Its basic functions include two aspects: by adjusting the industrial structure, building a water-saving society, adjusting the
distribution of productivity, restraining the momentum of water demand, and adapting to the more unfavorable conditions of water resources in demand, and coordinat-
ing competitive water use, strengthening management and changing water resources through engineering mea-
sures in the supply side. Natural space-time distribution is adapted to the distribution of productive forces. The two aspects complement each other in order to promote the sustainable development of the region. Rational allo-
cation is the goal and desire when people allocate scarce resources. Generally speaking, the results of a reasonable allocation are not the best for the benefit or benefit of a certain individual, but the overall benefit or benefit of the whole resource allocation system is the best. The optimal allocation is the way and means people are looking for in reasonable allocation plan.

Since Adleman’s pioneering work in Hamilton path problem [1], the research about DNA parallel computing has undergone rapid development process. Because of its advantages with parallelism and storage efficiency features, DNA parallel computing has been used to solve optimization NP-complete problems [2–18]. DNA computing refers to the process of solving the problems by producing and detecting a combination of a kind of mathematical process, which is based on DNA molecules parallel biochemical operations and processing techniques.

In general, DNA computing has three steps: (1) Generate DNA molecular data pools representing all possible solutions to the unsolved problem; (2) a series of biochemical experiments are carried out to eliminate the DNA solutions which do not meet the logical requirements of the problem; (3) select the optimal solution chains for the problem. At present, DNA computing research involves many aspects, such as DNA computing capability, models and algorithms.

The k-edge cover problem, aiming to find the cover set with k edges, is a typical NP problem. For G = (V,E), an edge cover means an edge set S of G such that \( \forall v_i \in V \), then have at least one \( e_{ij} \in S \) linking the vertex \( v_i \). The k-edge cover problem is to find edge cover sets with k edges. For example, the undirected graph G in Fig. 1 is such a problem. There has no efficient algorithm to solve the problem by now. Based on the research of Adleman [1] and Lipton [2], we use a new bio-computing procedure to get the solutions of the k-edge cover problem.

The rest of the article is arranged as below. In the Section 2, we introduce the DNA computing Adleman-Lipton model and use the DNA molecular algorithm to solve k-edge coverage problem. Next, we describe feasibility and complexity of the algorithm in Section 3. Then, experimental results of simulated DNA computing are shown in Section 4. We come to the conclusion in Section 5.

2. DNA computing and DNA algorithm for the k-edge cover problem

Although biology and mathematics have their own complexity, in recent years, Mathematical Biology, as an interdisciplinary subject, has been intersecting from the two major fields of biology and mathematics. The operations in Adleman-Lipton model mainly include: Merge \((T_1,T_2)\), Copy \((T_1,T_3)\), Separation \((T_1,T_2,T_3)\), Discard \((T)\), Read \((T)\), Sort \((T_1,T_2,T_3)\) and Append-tail \((T,y)\) [19–22]. Meanwhile, the complexity of these operations is O(1) [25–25].

The symbols \( I, A, B \) \((k = 1,2,...,n)\) denote distinct DNA single strands with same length, say t-mer. We use distinct DNA single strands \( AIB \) \((1 i < j n)\) to denote the edge \( e_{ij} \) with \( AIB \) including in the edge set. In order to get the optimum solutions, we meantime design weight information string \( X \) with t-mer length, that the number of pasting is determined by the weight \( w_i \) of the edge \( e_i \). We use adjacency matrix of graph G denoting the connection relationship between different vertices, which \( a_{ij} = 1 \) means vertex \( v_i \) links with \( v_i \) and \( a_{ij} = 0 \) denotes connectionless. Taking Fig. 1 for example, the adjacency matrix M(G) is as follows:

\[
M(G) = \begin{bmatrix}
0 & 1 & 1 & 0 & 0 & 0 \\
1 & 0 & 1 & 1 & 1 & 0 \\
1 & 1 & 0 & 1 & 1 & 0 \\
0 & 1 & 1 & 0 & 0 & 1 \\
0 & 1 & 1 & 0 & 0 & 1 \\
0 & 0 & 0 & 1 & 1 & 0
\end{bmatrix}
\]

As the simple undirected graph has no self-loops, so \( a_{ii} = 0 \) and \( a_{ij} = a_{ji} \). In progress, we only need to check the numerical value of upper triangular matrix M(G). Let \( T_i = \{ \emptyset \} \).

(1-1) If \( a_{ii} = 1 \) Then
(1-2) Copy \((T_i,T_j)\);
(1-3) Append-tail \((T_i,A1B)\);
(1-4) Merge \((T_i,T_j)\);
(1-5) Discard \((T_j)\);
Else
(1-6) Continue
End for

(2) Each single chain in the tube \( T \) represents a possible set of edges after algorithm (1). As k-edge cover problem requires the edge cover set should contain k edges. So we pick out these strands with the 3kt-mer length since we denote each edge as 3t-mer length.

(2-1) Selection \((T1,3kt,T3)\).
Every single chain in tube $T_i$ denotes a possible $k$-edge set so far. Meanwhile, the edge cover problem requires that arbitrary vertex should connect at least one edge of set. Therefore, we should check whether all sets satisfy the constraints. For $\forall v \in V$ in graph, we should keep the strands with edge symbols $A_j$ and $B_j$ synchronously.

For $i = 1$ to $i = n$

1. Copy ($T_i, T_i$);
2. Separation ($T_i, \{A, 1B\}$, $T_i$);
3. Merge ($T_i, T_i$);
4. Copy ($T_i, T_i$);
5. Append-tail ($T_i, XX\ldots X$);
6. Merge ($T_i, T_i$);
7. Discard ($T_i$);
8. Discard ($T_i$).

End for

The solutions of Algorithm (4) can be found by sorting, which indicate the optimal solutions of the problem.

(3-1) Sort ($T_i, T_i, T_i$);
(3-2) Read ($T_i$).

End for

And we reasonably design the length of $A_j$, $B_j$, $I$ and $X$, for $||A_j|| = ||B_j|| = ||1|| = ||X|| = 1$. We suppose $max\{w_{ij}\} = L$, so

$$|w_{ij}| + |w_{ij}| + |w_{ij}| + |w_{ij}| + |w_{ij}| + |w_{ij}| + \ldots + |w_{ij}| + |w_{ij}| + |w_{ij}| + |w_{ij}| + |w_{ij}| + \ldots + |w_{ij}| + |w_{ij}| + |w_{ij}| + |w_{ij}| + |w_{ij}| + |w_{ij}| + \ldots + |w_{ij}| + |w_{ij}| + |w_{ij}| + |w_{ij}| + |w_{ij}| + \ldots + |w_{ij}| + |w_{ij}| + |w_{ij}| + |w_{ij}| + |w_{ij}| + \ldots + |w_{ij}| + |w_{ij}| + |w_{ij}| + |w_{ij}| + |w_{ij}| + \ldots + |w_{ij}| + |w_{ij}| + |w_{ij}| + |w_{ij}| + |w_{ij}| + \ldots + |w_{ij}| + |w_{ij}| + |w_{ij}| + |w_{ij}| + |w_{ij}| + \ldots + |w_{ij}| + |w_{ij}| + |w_{ij}| + |w_{ij}| + |w_{ij}| + \ldots + |w_{ij}| + |w_{ij}| + |w_{ij}| + |w_{ij}| + |w_{ij}| + \ldots + |w_{ij}| + |w_{ij}| + |w_{ij}| + |w_{ij}| + |w_{ij}| + \ldots + |w_{ij}| + |w_{ij}| + |w_{ij}| + |w_{ij}| + |w_{ij}| + \ldots + |w_{ij}| + |w_{ij}| + |w_{ij}| + |w_{ij}| + |w_{ij}| + \ldots + |w_{ij}| + |w_{ij}| + |w_{ij}| + |w_{ij}| + |w_{ij}| + \ldots + |w_{ij}| + |w_{ij}| + |w_{ij}| + |w_{ij}| + |w_{ij}| + \ldots + |w_{ij}| + |w_{ij}| + |w_{ij}| + |w_{ij}| + |w_{ij}| + \ldots + |w_{ij}| + |w_{ij}| + |w_{ij}| + |w_{ij}| + |w_{ij}| + \ldots + |w_{ij}| + |w_{ij}| + |w_{ij}| + |w_{ij}| + |w_{ij}| + \ldots + |w_{ij}| + |w_{ij}| + |w_{ij}| + |w_{ij}| + |w_{ij}| + \ldots + |w_{ij}| + |w_{ij}| + |w_{ij}| + |w_{ij}| + |w_{ij}| + \ldots + |w_{ij}| + |w_{ij}| + |w_{ij}| + |w_{ij}| + |w_{ij}| + \ldots + |w_{ij}| + |w_{ij}| + |w_{ij}| + |w_{ij}| + |w_{ij}| + \ldots + |w_{ij}| + |w_{ij}| + |w_{ij}| + |w_{ij}| + |w_{ij}| + \ldots + |w_{ij}| + |w_{ij}| + |w_{ij}| + |w_{ij}| + |w_{ij}| + \ldots + |w_{ij}| + |w_{ij}| + |w_{ij}| + |w_{ij}| + |w_{ij}| + \ldots + |w_{ij}| + |w_{ij}| + |w_{ij}| + |w_{ij}| + |w_{ij}| + \ldots + |w_{ij}| + |w_{ij}| + |w_{ij}| + |w_{ij}| + |w_{ij}| + \ldots + |w_{ij}| + |w_{ij}| + |w_{ij}| + |w_{ij}| + |w_{ij}| + \ldots + |w_{ij}| + |w_{ij}| + |w_{ij}| + |w_{ij}| + |w_{ij}| + \ldots + \ldots$$

So the solutions strands of Algorithm (5) can be found in appropriate length range between $3kt$ and $3kt + kl$.

In addition, the calculation of DNA algorithm time complexity $T$ is as below:

- $T(Algorithm (1)) = O(6n(n – 1)/2) = O(n^2)$;
- $T(Algorithm (2)) = O(1)$;
- $T(Algorithm (3)) = O(8n) = O(n)$;
- $T(Algorithm (4)) = O(4n(n – 1)/2) = O(n^2)$;
- $T(Algorithm (5)) = O(2) = O(1)$;
- $T = T(Algorithm (1)) + T(Algorithm (2)) + T(Algorithm (3)) + T(Algorithm (4)) + T(Algorithm (5)) = O(n^2) + O(1) + O(n) + O(n^2) + O(1) = O(n^2)$.

End for

4. Experimental results of simulated DNA computing

DNA-based calculations depend on the biochemical operations of molecules, which may lead to errors in the application of these biochemical operations. Therefore, how to design DNA sequences is an important issue to ensure the reliability of DNA computation. In order to make good performance in hybridization reactions, we learn from the sequence design methods in reference [34–40].

In this paper, we use BioPython, a computational molecular biology tool, as our development platform to generate good DNA sequences suitable for performing our algorithms in the laboratory. The Braich’s programs run on Windows XP systems, with Intel kernel CPU and 8 GB main memory, and Visual C++ compiler. The coding programs are used to generate DNA sequences for solving the $k$-edge cover problem, and construct the DNA sequences of each bit in the library. Meanwhile, when a new DNA sequence is added, the program determines whether the DNA strand meets the constraints listed in the reference [34–41]. If the generated DNA sequence cannot satisfy the constraints, the program will regenerate the new DNA sequence. If the restriction condition is satisfied, the DNA sequence is accepted. When all DNA strands satisfy these constraints, the program stops, and these sequences are the output.

Taking 3-edge cover problem in Fig. 1 for an example, the program generates 4-base random sequences, consisting of $A_j$, $B_j$, $I$ and $X$ shown in Table 1. Edge DNA sequences
generated by Braich’s program are shown in Table 2. And we also calculate the enthalpy, entropy, and free energy for binding of each probe to its corresponding region on a library strand using the Braich’s program, while the energy used is shown in Table 3.

We also calculate the mean and standard deviation of enthalpy, entropy and free energy in all probe/library inter-

Table 3
The energies for of binding each probe to its corresponding region on a library strand

<table>
<thead>
<tr>
<th>Edges</th>
<th>Enthalpy energy H</th>
<th>Entropy energy S</th>
<th>Free energy G</th>
</tr>
</thead>
<tbody>
<tr>
<td>A&lt;sub&gt;1&lt;/sub&gt;B&lt;sub&gt;2&lt;/sub&gt;</td>
<td>108.7</td>
<td>102.1</td>
<td>21.1</td>
</tr>
<tr>
<td>A&lt;sub&gt;1&lt;/sub&gt;B&lt;sub&gt;3&lt;/sub&gt;</td>
<td>104.6</td>
<td>102.1</td>
<td>22.5</td>
</tr>
<tr>
<td>A&lt;sub&gt;2&lt;/sub&gt;B&lt;sub&gt;3&lt;/sub&gt;</td>
<td>103.5</td>
<td>102.1</td>
<td>23.7</td>
</tr>
<tr>
<td>A&lt;sub&gt;2&lt;/sub&gt;B&lt;sub&gt;4&lt;/sub&gt;</td>
<td>101.3</td>
<td>102.1</td>
<td>22.5</td>
</tr>
<tr>
<td>A&lt;sub&gt;2&lt;/sub&gt;B&lt;sub&gt;5&lt;/sub&gt;</td>
<td>107.4</td>
<td>102.1</td>
<td>24.5</td>
</tr>
<tr>
<td>A&lt;sub&gt;3&lt;/sub&gt;B&lt;sub&gt;5&lt;/sub&gt;</td>
<td>99.7</td>
<td>102.1</td>
<td>23.6</td>
</tr>
<tr>
<td>A&lt;sub&gt;3&lt;/sub&gt;B&lt;sub&gt;6&lt;/sub&gt;</td>
<td>105.8</td>
<td>102.1</td>
<td>23.9</td>
</tr>
<tr>
<td>A&lt;sub&gt;4&lt;/sub&gt;B&lt;sub&gt;6&lt;/sub&gt;</td>
<td>107.1</td>
<td>102.1</td>
<td>24.2</td>
</tr>
<tr>
<td>A&lt;sub&gt;4&lt;/sub&gt;B&lt;sub&gt;5&lt;/sub&gt;</td>
<td>102.2</td>
<td>102.1</td>
<td>22.8</td>
</tr>
</tbody>
</table>

Table 4
The energies over all probe/library strand interactions

<table>
<thead>
<tr>
<th></th>
<th>Enthalpy energy H</th>
<th>Entropy energy S</th>
<th>Free energy G</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>104.478</td>
<td>265.267</td>
<td>23.489</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>3.0429</td>
<td>9.0128</td>
<td>0.8388</td>
</tr>
</tbody>
</table>

actions. The energy level is shown in Table 4. Table 5 gets solutions strands of the 3-edge cover problem.

5. Conclusions

This paper introduces a new DNA computing algorithm to solve k-edge cover problem. The DNA algorithm has two merits. First of all, the aforementioned algorithm has a lower hybrid error rate, as we utilize reasonable DNA sequences to produce solutions. Secondly, the DNA algorithm can deal with the k-edge cover problem having n vertices in the O(n²) time complexity and compare exponential time complexity with the electronic computer, such as 2O(n^n) [29], O(2.3147^n) [30], O(n^n) [32] and so on. Many scholars believe that the study of DNA computing lays the foundation for human development of molecular computers. The intelligent system based on DNA computing will set up a bridge between DNA computing and intelligent system research, and will play a stepping stone role in DNA intelligent computers and other related research. The integration of DNA computing and soft computing will provide a good method to realize DNA intelligent computers. DNA intelligent computer can completely solve the intelligent computing function that the existing computer can’t achieve [26–28].

Acknowledgements

It was supported by National Key Research and Development Program of China (grant No. 2017YFC0405501) and the National natural science foundation of China (grant No. 51879282) and the Open Research Fund of State Key Laboratory of Simulation and Regulation of Water Cycle in River Basin, China Institute of Water Resources and Hydropower Research (grant No. SKL2018ZY07, IWHR-SKL-KF201905). The project was also financially supported by the Education Department of Guizhou province science and technology talents support programs (KJH KY[2016]088).

References


