

The Minimum Spanning Tree Problem based on Triple-Stranded DNA Structure Model

Jing Yang^{1,2,*}, Zhi-xiang Yin¹, Ming-qiang Chen² and Kai-feng Huang³

¹ School of Science, Anhui University of Science and Technology, Huainan, Anhui, 232001, China

² School of Earth and Environment, Anhui University of Science and Technology, Huainan, Anhui, 232001, China

³ Department of Information and Electrical engineering, Huainan Vocational & Technical College, Anhui, 232001, China

Received: 29 Jun. 2013, Revised: 8 Nov. 2013, Accepted: 10 Nov. 2013

Published online: 1 Jan. 2014

Abstract: Three-helix structure of the DNA molecule provided by the emergence of a new type of computing model, it has high stability, low rate of wrong solutions, easy separation and so on. Especially during the reaction data pool and the final solution was to come in the form of double-stranded, so avoid the hairpin structure, greatly improved the correct rate of solution. Minimum spanning tree problem is a NP-difficult problems in graph theory, which is looking for the tree contains all vertices and the smallest right of an undirected graph. We will attempt using of triplex DNA computing model for this article to solve the minimum spanning tree problem in graph theory.

Keywords: Minimum spanning tree, triple-stranded DNA, DNA computing

1 Introduction

It is the hot spots of DNA computing research that exploration the new molecular structure of DNA computing. In 1994, Adleman pioneering used DNA computing to solve 7 vertex of a directed graph for Hamilton path problem [1], and it marked the birth of this new research field DNA calculation. Many scholars have subsequently made a variety of DNA computing model, mainly single chains, double chains, molecular beacon and the mix of single and double chain model [2-5], and recently have made a three-chain DNA computing model. Now use of triplex DNA computing model have resolved the problem, 3vertex coloring problem [6], 0-1 integer programming problems [7], working operation problem [8], postman problem [9], traveling salesman problem [10], and so on. Here will try to take advantage of triplex DNA structure to give the minimum spanning tree problem DNA computing model.

2 DNA structure

The two chain of the double-stranded DNA is in the opposite direction, parallel to each other and interaction of

the chain winding through Watson- Crick base pairs between the hydrogen bonds linked. As the base of bias on the side of the screw shaft, the two chains form a large groove and a small groove. For homopurine/homopyrimine triplex structures recognition of the third strand occurs through the major groove by formation of Hoogsteen-type or reversed Hoogsteen-type hydrogen bonds between the bases of the oligopyrimidine or oligopurine third strand and the purine-rich strand of the duplex target. Literature [11] found oligodeoxynucleotides in the presence of *ATP_γS* and RecA protein, with linear double helix DNA can form stable three-chain structure. In the middle of the formation, first oligodeoxynucleotides and RecA protein binding in the presence of *ATP_γS*, and then look homologous sequences at the target double helical DNA. This process is very fast and does not open the DNA double strand. After homologous double chain be found, oligodeoxynucleotides on the RecA protein-mediated with the target DNA double helical DNA form a three-chain. We can make use of this unique structure to study some possible or feasible computing Models of the graph and combinatorial optimization.

* Corresponding author e-mail: jyang@aust.edu.cn

3 Minimum spanning tree problem

Supposed $G = (V(G), E(G), W(G))$ and $T = (V(T), E(T), W(T))$ are the two empowers undirected graph. If $V(T) = V(G)$ and T is a tree (excluding connected graph), we call T is a spanning tree of G . The weight of T is $W(T) = \sum(\omega(e) | e \in E(T))$. If there is a spanning tree T' meeting $W(T \leq W(T'))$, we refer to spanning tree T is the minimum spanning tree of the graph G . Minimum spanning tree problem is to finding a minimum spanning tree problem of empower the undirected graph G . Typical algorithms are Kruskal avoidance circle method, currently broken circle method and anti-circle method. The basic idea of Kruskal algorithm is according to the size of the weight arcs of the network G are first sorted, then starting from the smallest arc selected, after each elect a new arc to determine whether the selected arcs form a ring. If you give up the arc, or selected the arc until the selected number of arcs for vertices number minus 1. Each time algorithm selects the arc, always takes precedence over choice of consider the smallest weight arc, so avoidance circle method is also called the smallest arc optimization.

3.1 Biological algorithm

Step 1: Encode the vertices and arcs of the graph G given on automated DNA synthesizer. Encoding graph arcs to release the test tube;

Step 2: Uses gel electrophoresis to detect the smallest weight arc of DNA fragments, recorded the fragment;

Step 3: Detect vertices and arcs of the fragment. Here, it may note $e_1 = v_1v_2, E_1 = \{e_1\}, n = 2$, and note $V_n = \{v_1, v_2\}$. If $n = |V(G)|$, foolow $T = (V_n, E_{n-1})$ is a minimum spanning tree, otherwise proceed to the next step;

Step 4: Making use of the DNA fragments corresponding to vertices $V_n = \{v_1, v_2\}$ making up chain as a template structure probes, all arcs of the arc set $[V_n, V \setminus V_n]$ are captured. Build a new tube T , and seek the smallest weight arc in arcs set $[V_n, V \setminus V_n]$ by electrophoresis. Transfer to the third step.

3.2 Emulation test

To illustrate the three chain DNA computing model to solve the validity and correctness of the minimum spanning tree problem, we use following simple connected weighted graph G . For example, where the vertex set $V(G) = \{v_1, v_2, v_3, v_4, v_5, v_6\}$, arc set $E(G) = \{e_1, e_2, e_3, e_4, e_5, e_6, e_7, e_8, e_9\}$ and weight set $W(G) = \{5, 6, 1, 2, 7, 5, 3, 4, 4\}$.

Biological step:

Step1: Encode the vertices and arcs of the graph G given on automated DNA synthesizer. Keep the coding

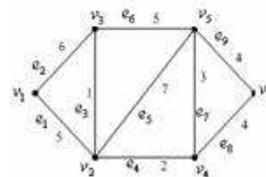


Figure 3. 1: The empowerment undirected graph minimum Spanning Tree Problem

arc of the oligomeric paired nucleotide fragment under W-C rule into double-stranded, mixed together, in order to avoid a hairpin structure or hybrids between the arcs, joined the buffer solution, placed in a test tube T . We plans for any of the vertex v_i , with a length of the 20 oligonucleotide fragments that, and recorded as $M_i, i = 1, 2, \dots, 6$.

$$\begin{aligned}
 v_1 : M_1 &= 5' - CCTATCGAAAGGGAGGAGGA - 3' \\
 \bar{M}_1 &= 5' - GGATAGCTTTCCTCCTCCT - 3' \\
 v_2 : M_2 &= 5' - AAGAGAAAAGGATATTCCT - 3' \\
 \bar{M}_2 &= 5' - TTCTCTTTCCTATAAGGA - 3' \\
 v_3 : M_3 &= 5' - CCTCTAAGGGAGGAGTTGAC - 3' \\
 \bar{M}_3 &= 5' - GGAGATTCCCTCCTCAACTG - 3' \\
 v_4 : M_4 &= 5' - GCCTAGAGGAAAGAAGGAGT - 3' \\
 \bar{M}_4 &= 5' - CGGATCTCCTTTCCTCCTCA - 3' \\
 v_5 : M_5 &= 5' - AAAGGAGTTCAGGAAGAGA - 3' \\
 \bar{M}_5 &= 5' - TTTCTCAAGGTCTTCTCT - 3' \\
 v_6 : M_6 &= 5' - TACCAAATGGTCAAATTGGA - 3' \\
 \bar{M}_6 &= 5' - ATGGTTTACCAGTTAACCT - 3'
 \end{aligned}$$

For encoding, there should be taken into account the weight of arcs. supposed the weight of the arcs based graph G ω_{ij} are positive real numbers, and $D = \max\{\omega_{ij} | i = 1, 2, \dots, n; j = 1, 2, \dots, m\}$, where n is the number of vertices of a graph G , and m is the number of arcs of the graph G . The arc (v_i, v_j) , the use of oligonucleotide fragments M_{ij} that it have three parts: the first part is last 10 base pairs of the M_i make up strand, and the third part is the first 10 base pairs of the M_j make up strand. The second part is weight of arc that can be said that the base $5 \times t_{ij}$, Where t_{ij} denotes the weight value. If the weights are large we can available small real numbers; if the weights of smaller integers we can be used to take values; if the weight of the arcs is a real number, we can used in other encoding [11]. Arc coding is as follows (where the weight to draw a line that represents value):

Step2: Out of the test tube part syrup solution agar gel electrophoresis, ran at the front is the shortest of the DNA chain, which is the same as smallest of arcs in the graph G weight arc. Cut electrophoresis, extraction and PCR amplification of the fragments.

$e_1 : M_{12} = 5' - \text{CCCTCCTCCTGAGGAGCTTCCGAAAG}$
 $\text{AGCCTTGCTTCTCTTTTC}$
 $e_2 : M_{13} = 5' - \text{CCCTCCTCCTTTTACCGGAAATCCGG}$
 $\text{TTAACCGGTAGCGGAGATTCCC}$
 $e_3 : M_{23} = 5' - \text{CCTATAAGGAGGCATGGAGATTCCC}$
 $e_4 : M_{24} = 5' - \text{CCTATAAGGATTGGACCGGTCGGATCT}$
 CCT
 $e_5 : M_{25} = 5' - \text{CCTATAAGGAGGCCAAGGTTAACCTAC}$
 $\text{AGGCATTCTGAGCTGCATTTCCTCAAG}$
 $e_6 : M_{35} = 5' - \text{TCCTCAACTGTTAGACCGTTAAGCCAG}$
 $\text{GTCTCAGTGACGATTGCCAGATTTCTCAAG}$

 $e_7 : M_{45} = 5' - \text{TTCTTCTCACCTTAAGGCCATGAGTTT}$
 CCTCAAG
 $e_8 : M_{46} = 5' - \text{TTCTTCTCATGAACCTAGTAGAACCTA}$
 TGATGGTTTACC
 $e_9 : M_{56} = 5' - \text{GTCCTTCTCTAAGCTTGCTGCAGTCGT}$
 AGAATGGTTTACC

Step3: Sequencing of the product of step 2, you can detect the shortest sort of bases of the DNA chain, namely the smallest weight arc of the graph G. e_3 corresponds to the vertex v_2, v_3 . v_2, v_3 corresponds to the oligonucleotide fragments make up the strand of DNA single strand of the 5' end are added with poly(A), biotin at the mark. Mix and incubate the strands that make up the vertex strands and RecA protein in a reaction mixture which contain $ATP_{\gamma}S$ under certain conditions. Thus, the nucleoprotein filament will cone into being. Making use of the nucleoprotein filament as a template structure probes P_2, P_3 .

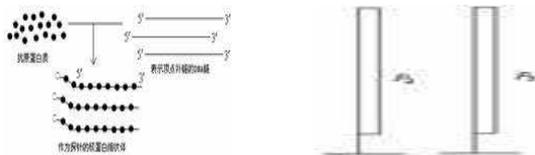


Figure 3. 2: The probes P_2, P_3 of the nucleoprotein filament probes

Step4: Separate arcs containing the vertex v_2, v_3 from the test tube T, and make use of the probe P_2, P_3 . Mix the probe P_2, P_3 to the data pool. Under certain conditions, according to the principle of Hoogsteen pairs this probe will be combined with double-stranded DNA of v_2, v_3 vertices generated three helical structure of DNA. And making use of biological separation operation in operations will no longer generate three-chain double strand DNA removed. The new data pool will be purified. And then on the same method with P_2, P_3 , respectively, containing the vertex v_2 of arcs $\{e_1, e_3, e_4, e_5\}$ and containing the vertex v_3 of arcs $\{e_2, e_3, e_6\}$ are separated out. Take advantage of magnetic bead separation method to restore to a dual chain establishment of test tubes, are still recorded as T. Transfer to the first step. After a second time through the loop from the shortest arc e_4 is found in the arcs $\{e_1, e_3, e_4, e_5, e_6\}$. The above steps will be the product of sequenced (this technology is now mature), and will get all arcs making up the minimum spanning tree, and its

weight summation is $5 + 1 + 2 + 3 + 4 = 15$. Minimum spanning tree is:

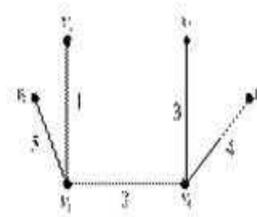


Figure 3. 3: Generate minimum spanning tree

4 Conclusion

Try the three-chain approach in this article for the minimum spanning tree problem, takes advantage of the three-chain model. When you build a data pool, there are the double helix structure of DNA chains, and stability of double-stranded DNA are higher than the stability of single-strand DNA, and do not appear in the data pool of due to coding problems "hairpin" structure, which our biochemical reactions more fully, more efficient. And also for isolation of triplex DNA double strand DNA separation more easily. Thus, for three-chain model calculations wrong solution for lower rates compared with other models. Here we take full advantage of the parallelism to shorten solution time DNA computers, but for actual problem solving, there are biotechnology issues, require further study. When the value for the weight to larger or smaller, coding problems can be solved easily, but too big for the weight difference between the weights of coding is relatively difficult. Yet increased for vertex, encoding is not affected too much. However, the vertex increase in the code but not much. Just vertex increase in the code, the weight to the value of the code is then determined, and it is also possible on the experiment. Next we will combine with some biological technology to further the achievement of triplex DNA model versatility and effectiveness.

Acknowledgement

This work is partially supported by Project supported by CNSF (61170172) and Anhui Province outstanding young talent Fund (2011SQRL198). Thanks for the help.

References

[1] L. Adlenlan, Molecular computation of solutions to combinatorial problems. Science, **26**, 1021-1024 (1994).

- [2] R. J. Lipton, DNA solution of hard computation problem. *Science*, **268**, 542-545 (1995).
- [3] Q. Ouyang, P. D. Kaplan, S. M. Liu, et al. DNA solution of the maximal clique problem. *Science*, **278**, 446-449 (1997).
- [4] T. Head, G. Rozenberg, R. B. Bladergroen, et al. Computing with DNA by operating on plasmids. *Biosystems*, **57**, 87-93 (2000).
- [5] K. Sakamoto, H. Gouzu, K. Komiya, et al. Molecular computation by DNA hairpin formation. *Science*, **288**, 1223-1226 (2000).
- [6] G. Fang, S. M. Zhang, Y. Zhu, et al. The DNA Computing Based on Triple Helix Nucleic Acid. *China Journal of Bioinformatics*, **7**, 181-182 (2009).
- [7] J. Yang, Z. X. Yin, 0-1 Integer Programming Problem Based on RecA-mediated Triple-stranded DNA Structure. *Computer Engineering and Applications*, **44**, 76-79 (2008).
- [8] J. Yang, Z. X. Yin, et al, The Working Operation problem on Triple-stranded DNA Structure model, *Global Congress on Intelligent Systems*, 295-299 (2009).
- [9] J. Yang, Z. X. Yin, et al, The postman problem on Triple-stranded DNA Structure model. *First International Conference on Cellular, Molecular Biology, Biophysics and Bioengineering*, 335-338 (2010).
- [10] J. Yang, Z. X. Yin, K. F. Huang, The Traveling Salesman Problem base on Triple-stranded DNA Structure model. *Informatics in Control, Automation and Robotics*, 769-775 (2011).
- [11] Y. Shigemoriy, M. Oish, Specific cleavage of DNA molecules at RecA-mediated triple-stranded structure. *Nucleic Acids Research*, **32**, 4563-4575 (2004).
- [12] Z. X. Yin. Map and combinatorial optimization of DNA computing. *Science Press*, 47-72(2004).



Jing

Yang was born in 1980, PH.d. Candidate, lecturer, Her research interests include the graph and combinatorial optimization and DNA computer, graph theory, protein structure prediction.



Zhi-xiang Yin was born in 1966, PH.d., professor. His research interests include the graph and combinatorial optimization and DNA computing, graph theory, protein structure prediction.



Ming-qiang

Chen was born in 1964, PH.d., professor, His research interests include development and utilization of biomass energy, coal conversion; in recent years, mainly engaged in the transformation of biomass.



Kai-feng

Huang was born in 1982, PH.d.Candidate, lecturer. His research interests include safety science and engineering, DNA computing.