Penerbit Akademia Baru

Journal of Advanced Research Design
ISSN (online): 2289-7984 | Vol. 6, No.1. Pages 1-10, 2015

Bio-inspired Computing for Network Modelling

N. Rajaee [*,a], *A. A. S. A Hussaini* [b], *A. Zulkharnain* [c] *and S. M. W Masra* [d]

Faculty of Engineering, Universiti Malaysia Sarawak, 94300 Kota Samarahan, Sarawak, Malaysia.
[a,*] *rnordiana@feng.unimas.my*, [b] *haahmad@frst.unimas.my*, [c] *zazham@frst.unimas.my*, [d] *wmmasnia@feng.unimas.my*

Abstract – *In this paper, we propose an approach to model and simulate a network problem using bio-inspired computing. Information in a network is embedded in DNA sequence strands and in-vitro implementation is proposed via bio-molecular tools such as hybridization, ligation and polymerase chain reaction. The output of the computation is derived from gel electrophoresis process.* ***Copyright © 2015 Penerbit Akademia Baru - All rights reserved.***

Keywords: Bio-inspired computing, network problems, Boolean matrices, DNA

1.0 INTRODUCTION

The diagnosis and solution for many problems such as efficiency control and computerized management systems are used to determine the state of discrete components and are frequently based on the construction of Boolean matrices and determination of minimal coverings for them. It is also possible to use Boolean matrices to construct diagnostic tests for discrete components where the matrices are constructed from tables of fault functions for the components under consideration. The algorithm for obtaining a minimized test requires an approximation of the state of a component by choosing a row with a maximum number of ones in a Boolean matrix [1].

The reverse of a Boolean matrix multiplication is the matrix decomposition. For two Boolean matrices and its product, the decomposition of its product is required to be a factor of the Boolean matrices as well. Matrix decompositions are used for many data mining purposes. One of these purposes is to find a concise but interpretable representation of a given data matrix. Different decomposition formulations have been proposed for this task, many of which assume a certain property of the input data (e.g., non-negativity) and aim at preserving that property in the decomposition. The Boolean matrix decomposition is intuitive when the data represents the presence and absence of items. Firstly, all matrices in the decompositions are binary, which aids the interpretation and helps to preserve the sparsity. Secondly, the Boolean matrix multiplication provides additional advantages. Using OR instead of addition is appealing in many contexts and yields significant improvements on reconstruction accuracy [2].

For example, let S and B be binary matrices of dimensions $m \times k$ and $k \times n$. Let $m \times n$ matrix P denote the Boolean product of S and B.

$$P = S \bullet B \tag{1}$$

Journal of Advanced Research Design
ISSN (online): 2289-7984 | Vol. 6, No.1. Pages 1-10, 2015

The ith row of *P* is the logical *OR* of the rows of *B,* for which the corresponding entry in the ith row of *S* is 1. Intuitively, *S* is the usage matrix, while *B* is the basis vector matrix [2].

Another example of a problem that utilizes Boolean matrices is a network system. A computer network can be defined as a set of devices, which are interconnected with each other. If a computer network system is represented by a graph, the devices are represented by vertices and their interconnectedness corresponds as the edges between the vertices in the graph as shown in Figure 1.

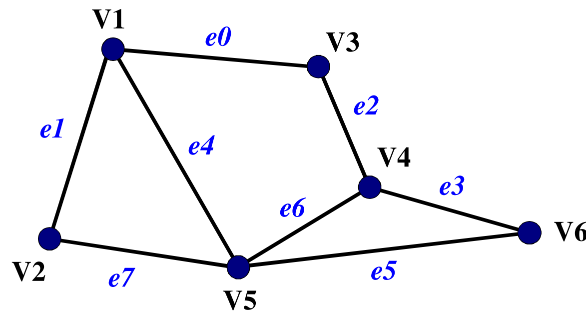

Figure 1: A network consisting of vertices and edges

The vertex-to-vertex relationship in the graph problem is translated into an Adjacency Matrix and the vertex-to-edge relationship is translated into an Incidence Matrix.

Definition 1:

Suppose that $G = < V, E >$ is an undirected and simple graph where $| V | = n$, and $v1, v2, \ldots, vn \in V$. The adjacency matrix *A* of *G*, with respect to this listing of the vertices, is the $n \bullet n$ zero-one matrix with 1 as its $(i, j)^{th}$ entry when *vi* and *vj* are adjacent. In other words, if its adjacency matrix is $A = [aij]$, then

$$a_{ij} = \begin{cases} 1 & \text{if } \{ v_i v_j \} \text{ is an edge of G} \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

Definition 2:

Suppose that $G = < V, E >$ is an undirected and simple graph where $| V | = n$, $| E | = m$, $v1, v2, \ldots, vn \in V$ and $e1, e2, \ldots, en \in E$. The incidence matrix *M* of *G*, with respect to this listing of the vertices, is the $n \bullet m$ zero-one matrix with 1 as its $(i, j)^{th}$ entry when *vi* and *ej* are incident, and 0 as its $(i, j)^{th}$ entry when they are not incident. In other words, if its incidence matrix is $M = [aij]$, then

Penerbit Akademia Baru

Journal of Advanced Research Design
ISSN (online): 2289-7984 | Vol. 6, No.1. Pages 1-10, 2015

$$m_{ij} = \begin{cases} 1 & \text{when edge } e_i \text{ is incident with } v_i \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

The adjacency matrix A is given as:

A =	V0	V1	V2	V3	V4	V5
V0	0	1	1	0	1	0
V1	1	0	0	0	1	0
V2	1	0	0	1	0	0
V3	0	0	1	0	1	1
V4	1	1	0	1	0	1
V5	0	0	0	1	1	0

Figure 2: Adjacency Matrix for Network

2.0 DNA COMPUTING

A DNA molecule is a long string consisting of two strands wound around each other to form a double helix. There are four types of organic bases: adenine (A), cytosine (C), guanine (G) and thymine (T). A short single stranded DNA chain, usually less than 30 nucleotides long is called an oligonucleotide. The ends of a DNA strand are chemically polar, with the so called 5' end and the 3' end. Each base has a bonding surface, where the bonding surface of A is complementary to that of T, and that of G is complementary to that of C. This complementary rule is called the Watson-Crick complementary. A single DNA strand can pair with another strand when their sequences of bases are mutually complementary and the chains have opposite polarity. DNA strands are often quoted in 5' – 3' order and the length of a DNA strand is denoted in mer, where one mer represents one DNA oligonucleotides. Two single stranded DNA under certain conditions form a double stranded DNA. The length of a double stranded DNA is denoted in base pairs (b.p.).

DNA computing is a wet-lab process, which includes a number of feasible bio-chemical operations (tools) necessary to execute such computations. In general, it consists of two steps: the reaction, involving the interactions of the molecules to produce potential solution molecules and the extraction, the techniques employed to isolate or otherwise identify any correct solution molecules during the reaction [3].

Basic bio-chemical operations utilized in DNA computation are explained further as follows:

Hybridization is the annealing of complementary single stranded molecules to form a double stranded DNA. This is the basis for initial path formation during the reaction step and is subsequently employed during the extraction phase for the isolation of generated path molecules.

Journal of Advanced Research Design
ISSN (online): 2289-7984 | Vol. 6, No.1. Pages 1-10, 2015

Ligation is a process often invoked after the single stranded DNA is annealed and concatenated to each other. Many single stranded fragments are connected in series and ligase is used as "glue" to seal the covalent bonds between the adjacent fragments.

Denaturation is an in vitro melting process. Double stranded DNA molecules can be separated without breaking the single strands by applying heat to the solution. The double stranded molecules come apart because the hydrogen bonds between complementary nucleotides are much weaker than the covalent bond between the adjacent nucleotides in the same strands.

Cutting process is carried out by using Restriction Enzymes (RE). Restriction enzymes recognize a specific sequence of DNA known as a restriction site. Any DNA that contains the restriction site within its sequence is cut by the enzymes at that point.

Polymerase Chain Reaction (PCR) is an amplification technique widely used in molecular biology. A pair of DNA sequences known as "primers" is used to signal the starting point and ending point for a specific target DNA sequence for amplification. The PCR process is capable of exponentially amplifying a DNA strand into millions of its copies given a site-specific single molecule DNA and the process is usually carried out in three stages of different temperatures.

Gel electrophoresis is a technique used for separation of DNA strands according to their sizes using electric current applied to the gel containing the strands. The size of the DNA strands refers to the weight of the DNA strands, which is proportional to the lengths of their sequences. This technique is based on the fact that DNA molecules are negatively charged. Since DNA molecules have the same charge per unit length, they all migrate at the same speed in aqueous solution. However, if electrophoresis is carried out in the gel, the migration rate is affected by its size causing less weighted strands to migrate faster. Thus, sorting the strands by their sequence lengths is made possible using this technique. The results of gel electrophoresis process can be viewed by staining the gel with fluorescent dye and photographed under UV light [3].

3.0 MODELLING WITH DNA COMPUTING

To construct the network in Figure 1, we synthesized single stranded DNA sequences encoded with unique information of each individual device and represent them as vertices in the graph. The length of each single stranded DNA sequence for the vertices is set at 20-mer and comprises of two 10-mer sections. The first 10-mer section is tagged as "row ith" and the latter 10-mer section is tagged as "column j^{th}", where $| i | = | j | = | v | = n$. The first sections for all single strands representing the vertices are defined as the row indicators for the adjacent matrix A. The latter sections for all single strands representing the vertices are defined as the column indicators for the adjacent matrix A. The generated sequences are shown in Table 1. The connections between vertices are shown as edges in the graph. In network connections, the edges are undirected to signal two-way communications between the vertices. The edges are shown as elements of value 1 in the adjacency matrix *A* and the two-way communication is shown by the same edge recurring twice for two corresponding vertices. The adjacency matrix *A* can be rewritten as in Figure 3.

Penerbit Akademia Baru

Journal of Advanced Research Design
ISSN (online): 2289-7984 | Vol. 6, No.1. Pages 1-10, 2015

Table 1: DNA sequences for Vertices

Vertex	DNA Sequences (5'-3')	Length
V_1	tgttccatttgattgcgtcc	20 mer
V_2	ctacatctcgtgcccgttca	20 mer
V_3	agggttgctcttgtctcgga	20 mer
V_4	caggctaagggtgacgggag	20 mer
V_5	ttttcgtctgagtgtttcgc	20 mer
V_6	tatcgtgatttggaggtgga	20 mer

A =

	V1	V2	V3	V4	V5	V6
V1	0	e1	e0	0	e4	0
V2	e1	0	0	0	e7	0
V3	e0	0	0	e2	0	0
V4	0	0	e2	0	e6	e3
V5	e4	e7	0	e6	0	e5
V6	0	0	0	e3	e5	0

Figure 3: Adjacency Matrix for Network (in Edges)

In case of two-way communication between the devices, the total number of DNA strands required to represent all possible edges in the adjacency matrix is N x N. However, to simplify our design for the edge sequences, we constructed only unidirectional edges to show the connections between the vertices. This is achieved by constructing only edges from lower degree vertices to higher degree vertices as shown in Figure 4. The existing edges constructed from the graph are E={e0 (V1→V3), e1 (V1→V2), e2 (V3→V4), e3 (V4→V6), e4 (V1→V5), e5 (V5→V6), e6 (V4→V5), e7 (V2→V5)}. The total number of DNA strands to represent all possible connections in the network is now reduced to half of the elements in the adjacency matrix A.

A =

	V1	V2	V3	V4	V5	V6
V1	0	e1	e0	0	e4	0
V2	e1	0	0	0	e7	0
V3	e0	0	0	e2	0	0
V4	0	0	e2	0	e6	e3
V5	e4	e7	0	e6	0	e5
V6	0	0	0	e3	e5	0

Figure 4: Adjacency Matrix for Network (Directed edges)

Penerbit Akademia Baru
Journal of Advanced Research Design
ISSN (online): 2289-7984 | Vol. 6, No.1. Pages 1-10, 2015

The edge sequences are designed to act as connectors between the vertex strands. Two vertex strands are connected to each other by an edge containing partial complementary sequences of the last 10-mer of the first vertex and the first 10-mer of the second vertex. The concatenation of the two vertex sequences and an edge creates a 40 b.p. (base pair) length elongated strand.

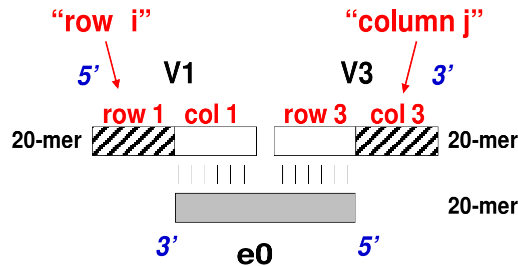

Figure 5: Construction of vertices and edges

Generated DNA sequences for edges are shown in Table 2.

Table 2: Generated DNA Sequence for Edges

Edges	DNA Sequences (5'-3')	Length
e_1	gtcagtacgtggtcagtcag	20 mer
e_2	cgtgggtaacaaatgcgtca	20 mer
e_3	caattgctcgtacaagtagatc	20 mer
e_4	aaatgctgacgtagctaccg	20 mer
e_5	cctgaagtacgtgcttgatcga	20 mer
e_6	ttaagctagctaatgcatggt	20 mer
e_7	tgctggtaagcatggtagta	20 mer
e_8	ttagtggcgtggcaaatagc	20 mer

The process of concatenation occurs simultaneously for all strands and elongated for each vertex-to-vertex connection. The length of the elongated strands depends on the number of vertices connected. For example, a route from V1-V3-V4-V6 has an 80-b.p strand length and has a connection degree of 3 (excluding the starting vertex). To check a connection of degree 1 that involves a direct connection between only two vertices, the elongated strand length is filtered for 40 b.p. All edges shown in the adjacency matrix has a connection degree 1.

Penerbit Akademia Baru

Journal of Advanced Research Design
ISSN (online): 2289-7984 | Vol. 6, No.1. Pages 1-10, 2015

		V1	V2	V3	V4	V5	V6
	V1	0	40	40	0	40	0
	V2	40	0	0	0	40	0
A =	V3	40	0	0	40	0	0
	V4	0	0	40	0	40	40
	V5	40	40	0	40	0	40
	V6	0	0	0	40	40	0

Figure 6: Adjacency Matrix for Network (Length in Base Pair)

4.0 EXPERIMENT

All generated sequences for vertices and edges in Table 1 and Table 2 are poured into a single tube T0. The single stranded individual sequences combine and form double stranded sequences via the hybridization-ligation process. Strands are elongated by the concatenation of vertex-to-vertex strands by edge strands and various lengths of elongated strands are formed. The lengths of the strands are dependent on the number of vertices combined in a single elongated strand. To check whether a connection from a vertex to another vertex is established, we filtered the connection strand via Polymerase Chain Reaction method. In bio-molecular terms, filtering is carried out by mass amplifying only the desired strand and reducing the undesired strands to an insignificant quantity. In PCR, a set of sequences known as "primers" are used to signal the "start" and "end" of the sequences to be mass amplified. For checking an established connection between two vertices, the selected primers are used to detect the connection strand, which is from (i) the start of the first vertex sequence to (ii) the end of the second vertex sequence. These correspond to the tagged "row i^{th}" and "column j^{th}" for the elongated strand sequence as in Figure 7. For all connections in the adjacency matrix, the "start" and "end" sites for mass amplification are also the row and column indicators.

	"col 1"	"col 2"	"col3"	"col 4"	"col 5"	"col 6"
"row 1"	0	40	40	0	40	0
"row 2"	40	0	0	0	40	0
"row 3"	40	0	0	40	0	0
"row 4"	0	0	40	0	40	40
"row 5"	40	40	0	40	0	40
"row 6"	0	0	0	40	40	0

Figure 7: Row and Column Indicators for Network

Each row and column indicator intersection represents a cell value and successful amplification of the corresponding strand (which is read out as a highlighted band in gel electrophoresis process) shows an established connection of the vertices. However, connections between vertices vary on their connection degrees, which are based on their

Penerbit Akademia Baru

Journal of Advanced Research Design
ISSN (online): 2289-7984 | Vol. 6, No.1. Pages 1-10, 2015

length of different routes. The availability of an edge connecting two vertices directly is verified by an existence of a 40 b.p. length strand, indicating a connection degree 1.

For the PCR process, the contents of test tube T0 after the initial pool generation process are divided into fifteen test tubes, equivalent to the number of possible edges in the adjacency matrix and each representing the cells in Figure 7. Into each test tube, primers corresponding to the cell's "row" and "column" tags are added as shown in Table 3.

Table 3: Addition of Primers into Test Tubes for PCR

Test Tube	Primer 1	Primer Sequence 5' – 3'	Primer 2	Primer Sequence 5' – 3'
T1	"row 1"	tgttccattt	"col 2"	tgaacgggca
T2	"row 1"	tgttccattt	"col 3"	tccgagacaa
T3	"row 1"	tgttccattt	"col 4"	ctcccgtcac
T4	"row 1"	tgttccattt	"col 5"	gcgaaacact
T5	"row 1"	tgttccattt	"col 6"	tccacctcca
T6	"row 2"	ctacatctcg	"col 3"	tccgagacaa
T7	"row 2"	ctacatctcg	"col 4"	ctcccgtcac
T8	"row 2"	ctacatctcg	"col 5"	gcgaaacact
T9	"row 2"	ctacatctcg	"col 6"	tccacctcca
T10	"row 3"	agggttgctc	"col 4"	ctcccgtcac
T11	"row 3"	agggttgctc	"col 5"	gcgaaacact
T12	"row 3"	agggttgctc	"col 6"	tccacctcca
T13	"row 4"	caggctaagg	"col 5"	gcgaaacact
T14	"row 4"	caggctaagg	"col 6"	tccacctcca
T15	"row 5"	ttttcgtctg	"col 6"	tccacctcca

After the PCR process, the contents of the test tubes are read-out in the gel electrophoresis process. Successfully amplified strands yield highlighted bands indicating existing connections between vertices for the cell. However, only the existence of 40 b.p. highlighted bands indicate connections of degree 1 between the vertices, which is translated as elements of value 1 for cells in the adjacency matrix. The expected read-out from the gel electrophoresis is shown in Figure 8.

Journal of Advanced Research Design
ISSN (online): 2289-7984 | Vol. 6, No.1. Pages 1-10, 2015

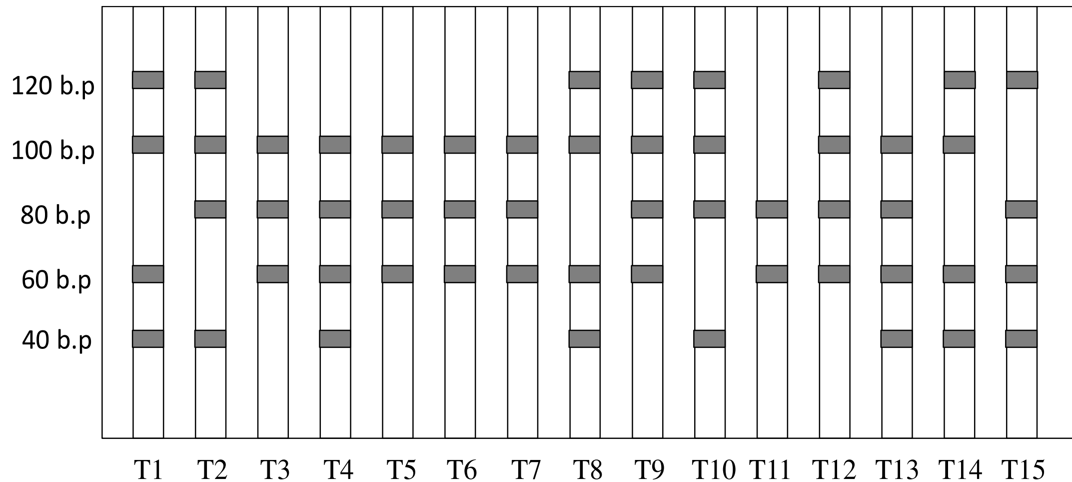

Figure 8: Gel electrophoresis process results

5.0 CONCLUSION

We proposed to model and simulate a network problem by embedding the information in the network into DNA sequence strands. The DNA sequences are designed and synthesized to represent the vertices and edges in the network and the computation are proposed to be implemented via bio-molecular tools. From the output of the computation, which will be visualized from the gel electrophoresis process, all the paths in the network should be visible from the highlighted bands indicating the lengths of the existing paths from one vertex to another. However, only highlighted bands of length 40 b.p indicate the element of value 1 in the row-column of the Adjacent Matrix for the network.

REFERENCES

[1] A.V. Gorish, Boolean Matrices in Problems on Determining the State of Discrete Components in Computerized Measurement Systems, Measurement Techniques 43 (200) 481-485.

[2] P. Miettinen, The Boolean column and column-row matrix decompositions, Data Mining Knowledge Discovery 17 (2008), 39-56.

[3] M. Amos, (2004). Theoretical and Experimental DNA Computation, Natural Computing Series, Springer.

[4] J.Y. Lee, S.Y. Shin, T.H. Park, B.T. Zhang, Solving travelling salesman problems with DNA molecules encoding numerical values, Biosystems 78 (2004) 39-47.

[5] L. Adleman, Molecular computation of solutions to combinatorial problems, Science 266 (1994) 1021-1024.

[6] M. Ogihara, A. Ray, (1996). Simulating Boolean circuits on a DNA computer, Technical Report T631, University of Rochester.

[7] N. Rajaee, O. Ono, Experimental Implementation of Boolean Matrix Multiplication with DNA Computing using Parallel Overlap Assembly, International Journal of Innovative Computing, Information and Control 6 (2010) 591-599.

[8] N. Rajaee, O. Ono, Single stranded DNA and primers for Boolean matrix multiplication with DNA Computing, ICIC-E, International Journal of Research and Surveys (2010) 1067-1072.

[9] J.H. Reif, Parallel molecular computation, Proceedings of Seventh Annual ACM Symposium on Parallel Algorithms and Architectures (1995) 213-223.

[10] R. Agrawal, F. Gustavson, M. Zubair, A high performance matrix multiplication algorithm on a distributed memory parallel computer using overlapped communication, IBM Journal of Research and Development 38 (1994) 673-681.

[11] J.S. Oliver, Matrix multiplication with DNA, Journal of Molecular Evolution 45 (1997) 161-167.