



Inferring patterns in mitochondrial DNA sequences through hypercube independent spanning trees



Eduardo Sant'Ana da Silva, Helio Pedrini*

Institute of Computing, University of Campinas, Campinas-SP, 13083-852, Brazil

ARTICLE INFO

Article history:

Received 8 October 2015

Accepted 5 January 2016

Keywords:

DNA sequence

Hypercube

Independent spanning trees

E-cube

Pattern recognition

ABSTRACT

Given a graph G , a set of spanning trees rooted at a vertex r of G is said vertex/edge independent if, for each vertex v of G , $v \neq r$, the paths of r to v in any pair of trees are vertex/edge disjoint. Independent spanning trees (ISTs) provide a number of advantages in data broadcasting due to their fault tolerant properties. For this reason, some studies have addressed the issue by providing mechanisms for constructing independent spanning trees efficiently. In this work, we investigate how to construct independent spanning trees on hypercubes, which are generated based upon spanning binomial trees, and how to use them to predict mitochondrial DNA sequence parts through paths on the hypercube. The prediction works both for inferring mitochondrial DNA sequences comprised of six bases as well as infer anomalies that probably should not belong to the mitochondrial DNA standard.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Mitochondria occupy a unique position among cellular organisms due to their possession of a separate genome and all enzymes for translating genetic information into functional proteins. Early observations suggested that some proteins were also likely to be encoded in mitochondrial DNA (*mtDNA*) [1].

The genetic code is the biochemical system for the expression of the gene. It handles the translation of the information contained in the DNA and RNA molecules into protein sequences. Thus, the genetic code is simultaneously a physiochemical and a communication system.

Dealing with DNA as a communication system, we are interested in efficient mechanisms of message passing and data broadcasting. In this sense, independent spanning trees (ISTs) [2–4] provide several advantages in data broadcasting due to their resilience to faults. Efficient strategies have been proposed to construct independent spanning trees [4–6].

In this paper, we investigate how to infer *mtDNA* sequences formed by six bases through the topological nature of DNA, as well as infer what we refer to as anomalies resulting from longest paths on the hypercube to connect topological neighbors.

The main contribution of the work is to search for patterns on DNA sequences. We have chosen mitochondrial DNA due to its simplicity, where the DNA sequence size is small but fully functional.

From the recognition of the patterns over DNA sequences, it is possible to use them on tasks such as efficient comparison,

compression and reconstruction. Additional to the hypercube pattern sequences on DNA, reported previously in [7], we successfully predicted patterns that should not belong to the mitochondrial DNA. As future steps, we could investigate how those sequences that we considered as anomalies influence subjects in characteristics (organism complexity) such as evolutionary mutations and diseases.

The text is organized as follows. In Section 2, we introduce some preliminaries related to the topic under investigation. In Section 3, we show how to construct independent spanning trees over hypercubes. Section 4 presents the DNA patterns investigated in this work. Section 5 reports and discusses the results obtained through our methodology. Finally, Section 6 presents the conclusions and directions for future work.

2. Preliminaries

Let G be a graph with n nodes and e edges. A set of spanning trees rooted at a vertex r of G is said vertex/edge independent if, for each vertex v of G , $v \neq r$, the paths of r to v in any pair of trees are vertex/edge disjoint. The following conjecture relates a connected graph and independent spanning trees (ISTs).

Conjecture 2.1. For any k -connected graph G , there exist k independent spanning trees (k -ISTs) of G with any vertex v of G as root of the tree [8].

The conjecture was proved by Itai and Rodeh [9] for $k=2$ and independently proved by Zehavi and Itai [8] and Cheriyan and

* Corresponding author.

E-mail address: helio@ic.unicamp.br (E.S.d. Silva).

Maheshwari [10] for $k \leq 4$. However, it is still an open problem for arbitrary graphs when $k > 4$.

Various algorithms have been proposed for some classes of graphs, such as planar graphs [11], product graphs [12], chordal rings [9], hypercubes [4,13], folded hypercubes [14], folded hyperstars [15], De Bruijn and Kautz graphs [16,17], multidimensional torus [18], and circulant recursive graphs [19].

The hypercube Q_k is a k -regular graph with vertex set equal to $\{0, 1, \dots, 2^k - 1\}$ such that there is an edge between two vertices if, and only if, the binary representations of their numbers differ in just one bit position.

Before presenting the developed method, it is necessary to define some terms that will be used throughout this paper, as well as some conventions proposed here.

Let $G = (V, E)$ be a graph with a vertex set $V = \{v_1, \dots, v_n\}$ and an edge set E , the adjacency matrix $M_G = (m_{ij})$, of the same order of G , has non-zero value at the value at coordinate i, j (generally value 1 when the graph is not valued) if, and only if, v_i and v_j have an edge connecting them, otherwise the value at coordinate i, j is 0. All matrix coordinates m_{ij} satisfying the condition $i=j$ have 0 as value.

Fig. 1 shows the adjacency matrices corresponding to the hypercubes Q_1, Q_2 and Q_3 with 2, 4 and 8 vertices, respectively. The number in parentheses after each node index represents its binary representation.

The tensor product of two vector spaces V and W , denoted as $V \otimes W$, is a way to create a new vector space, analogous to the multiplication of integers. For instance

$$\mathbb{R}^n \otimes \mathbb{R}^k \cong \mathbb{R}^{nk} \tag{1}$$

In particular

$$r \otimes \mathbb{R}^n \cong \mathbb{R}^n \tag{2}$$

where r as a scalar.

Algebraically the vector space $V \otimes W$ is generated by elements in the form $v \otimes w$, where the following properties are satisfied for any scalar α

$$(v1 + v2) \otimes w = v1 \otimes w + v2 \otimes w \tag{3}$$

$$v \otimes (w1 + w2) = v \otimes w1 + v \otimes w2 \tag{4}$$

$$\alpha(v \otimes w) = (\alpha v) \otimes w = v \otimes (\alpha w) \tag{5}$$

A basic consequence of these expressions is

$$0 \otimes w = v \otimes 0 = 0 \tag{6}$$

The tensor product is a useful tool that allows the combination of vector spaces to form larger vector spaces. The tensor product assists the n -dimensional graph analysis by reducing the scope to a planar shape in two dimensions, that is, the adjacency matrix.

Let $\mathbf{U} \in \mathbb{R}^{m \times n}$ and $\mathbf{V} \in \mathbb{R}^{p \times q}$. The tensor product, also known as Kronecker product [20], of \mathbf{U} and \mathbf{V} is defined in the following equation:

$$\mathbf{U} \otimes \mathbf{V} = \begin{bmatrix} u_{1,1}\mathbf{V} & u_{1,2}\mathbf{V} & \dots & u_{1,n}\mathbf{V} \\ u_{2,1}\mathbf{V} & u_{2,2}\mathbf{V} & \dots & u_{2,n}\mathbf{V} \\ \vdots & \vdots & \ddots & \vdots \\ u_{m,1}\mathbf{V} & u_{m,2}\mathbf{V} & \dots & u_{m,n}\mathbf{V} \end{bmatrix} \in \mathbb{R}^{mp \times nq} \tag{7}$$

$$= \begin{bmatrix} u_{1,1}v_{1,1} \dots u_{1,1}v_{1,q} & u_{1,2}v_{1,1} \dots u_{1,2}v_{1,q} & \dots & u_{1,n}v_{1,1} \dots u_{1,n}v_{1,q} \\ \vdots & \vdots & \ddots & \vdots \\ u_{1,1}v_{p,1} \dots u_{1,1}v_{p,q} & u_{1,2}v_{p,1} \dots u_{1,2}v_{p,q} & \dots & u_{1,n}v_{p,1} \dots u_{1,n}v_{p,q} \\ u_{2,1}v_{1,1} \dots u_{2,1}v_{1,q} & u_{2,2}v_{1,1} \dots u_{2,2}v_{1,q} & \dots & u_{2,n}v_{1,1} \dots u_{2,n}v_{1,q} \\ \vdots & \vdots & \ddots & \vdots \\ u_{2,1}v_{p,1} \dots u_{2,1}v_{p,q} & u_{2,2}v_{p,1} \dots u_{2,2}v_{p,q} & \dots & u_{2,n}v_{p,1} \dots u_{2,n}v_{p,q} \\ \vdots & \vdots & \ddots & \vdots \\ u_{m,1}v_{1,1} \dots u_{m,1}v_{1,q} & u_{m,2}v_{1,1} \dots u_{m,2}v_{1,q} & \dots & u_{m,n}v_{1,1} \dots u_{m,n}v_{1,q} \\ \vdots & \vdots & \ddots & \vdots \\ u_{m,1}v_{p,1} \dots u_{m,1}v_{p,q} & u_{m,2}v_{p,1} \dots u_{m,2}v_{p,q} & \dots & u_{m,n}v_{p,1} \dots u_{m,n}v_{p,q} \end{bmatrix}$$

(8)–(10) illustrate the construction of the Q_3 adjacency matrix using tensor product:

$$A_{Q_k} = \sum_{i=0}^k I^i \otimes R \otimes I^{k-i} \tag{8}$$

$$A_{Q_3} = \sum_{i=0}^3 I^i \otimes R \otimes I^{3-i} \tag{9}$$

$$A_{Q_3} = R \otimes I \otimes I + I \otimes R \otimes I + I \otimes I \otimes R \tag{10}$$

where

$$R = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$R \otimes I = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

$$R \otimes I \otimes I = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

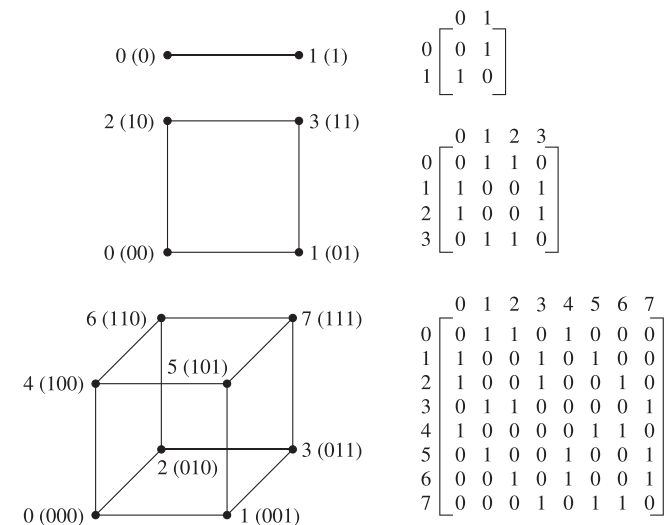


Fig. 1. Hypercubes and their corresponding adjacency matrices.

Download English Version:

<https://daneshyari.com/en/article/6920916>

Download Persian Version:

<https://daneshyari.com/article/6920916>

[Daneshyari.com](https://daneshyari.com)