Available online at www.sciencedirect.com

ScienceDirect

Simulation Modelling Practice and Theory 16 (2008) 1561–1570

SIMULATION MODELLING PRACTICE AND THEORY

www.elsevier.com/locate/simpat

A DNA-based algorithm for arranging weighted cliques

Ikno Kim [a,*], Junzo Watada [a], Witold Pedrycz [b]

[a] *Graduate School of Information, Production and Systems, Waseda University, 2-7 Hibikino, Wakamatsu-ku, Kitakyushu 808-0135, Japan*
[b] *Department of Electrical and Computer Engineering, University of Alberta, Edmonton, Canada*

Received 21 December 2006; received in revised form 26 April 2007; accepted 3 November 2007
Available online 17 November 2007

Abstract

A fundamental idea and realization of networks arises in a variety of areas of science and engineering. Their theoretical underpinnings stem from graph theory where numerous fundamental concepts being formulated and solved there have become of immediate interest at the applied side. In this study, our focus is on the weighted maximum clique problem, a highly challenging problem in graph theory. The essence of the problem is to find the nodes with the maximum total of weights in a graph where an edge connects every pair of nodes, meaning every node connects to every other node. We propose an algorithm to find all the weighted cliques as well as the weighted maximum clique in order of size using the framework of DNA computing.

Keywords: DNA oligonucleotide; Graph theory; Weighted clique; Weighted graph

---

1. Introduction

The current technology and application of information technology are concerned quite commonly with weighted clique problems. For instance, they are inherently present when dealing with subnetworks which are connected in a clique with some specific weights. The importance of weighted clique problems is emphasized in usefulness through information technology. Unfortunately, several key problems existing in graph theory cannot be efficiently solved given the computing resources that currently available. One of the most common and difficult problems existing in graph theory concern finding the largest complete graph. Known to be NP-complete, it is also referred as the maximal clique problem (MCP) [2]. In addition, the weighted maximum clique problem (WMCP) in which is given an undirected graph weighted on the nodes, is more adaptive, flexible, and wide-ranging than MCP in real situations, but is also a very intractable NP-complete problem [4].

Several approaches have proposed to solve the WMCP with the use of various heuristics. For instance, a branch-and-bound algorithm based on error-correcting codes was offered by Ostergard [8], who proposed a set of weighted graphs that can be subsequently used for evaluation of weighted maximum clique algorithms. The

---

[*] Corresponding author.
E-mail addresses: octoberkim@akane.waseda.jp (I. Kim), watada@waseda.jp (J. Watada), pedrycz@ee.ualberta.ca (W. Pedrycz).

1569-190X/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.simpat.2007.11.003

use of a Hopfield network for finding the near-optimal WMCP solution was proposed by Jain et al. [5]. On the side of computational biology side, Butenko et al. [3] described clique detection models based on biochemistry. Ouyang et al. [9] proposed an algorithm based on a DNA optimization for solving the maximal clique problem, yet its weighted version has not been addressed. Consequently, in this paper, we augment these concepts and develop an algorithm to find the weighted maximum clique. We also present how all weighted cliques as well as the weighted maximum clique can be arranged in order of size through the use of DNA oligonucleotides. An efficient solution of finding the weighted maximum clique which can be obtained by molecular computation is studied in detail. We also examine the effectiveness of molecular computation with this regard. The paper is organized in the following manner. Section 2 formally describes graphical models of two theories and show how they are related to clique problems from a theoretical point of view. Section 3 introduces the clique problems in both unweighted and weighted version. In Section 4, we propose the algorithm through the application of DNA oligonucleotides. Section 5 elaborates on the experimental method based on DNA computing. Finally, in Section 6, we cover some concluding remarks.

2. Theoretical background

Clique problems could be made more understandable by casting them in the setting of probability theory and graph theory. Given this, in this section, we bring some essential background.

In general, graphical models discussed here are based on probability theory [14] and graph theory [6]. Graphical models in probability theory generally provide a mathematical basis for relating data to hypotheses when uncertainty is present, and a consistent calculus for uncertain inference, so the output of the system is always unambiguous. Graphical models in graph theory provide both an intuitively appealing interface by which humans can model highly-interacting sets of variables and a data structure that naturally lends itself to the design of algorithms for an efficient purpose. Furthermore, graphical models can be divided into both directed and undirected models. Basically, a graphical model is a graph where nodes present the concepts of random variables and a pattern of the edge represents a set of conditional independence assumptions made among the random variables. For example, undirected graph models induce different sets of conditional independence assumptions in three undirected models which correspond to a model where every conditional independence assumption holds. A model where some conditional independence assumptions hold and some are not satisfied, and a model with no conditional independence assumptions, leading to some undirected complete graph.

When a subset of nodes X isolates the set of nodes Y from the set of nodes Z, in the formalism of graphical model, we say that Y and Z are conditionally independent from X, this is $\Pr(Y, Z|X) = \Pr(Y|X)\Pr(Z|X)$. Graphical models that usually induce different sets and can represent positive probability distributions can be simplified by imposing assumptions of conditional independence assumptions. A fundamental result in graphical models is the Hammersley-Clifford theorem [14]. Assume that $\Pr(x_1, x_2, \ldots, x_n) > 0$. Then,

$$\Pr(x) = \frac{1}{Z}\exp\left(\sum_{c\in \mathrm{Sc}} \Phi_c(x_c)\right), \tag{1}$$

center the equations where c is a maximal clique, Sc is the set of all maximal cliques, x_c is the restriction of x to the clique c, and the positive function $\Phi_c(x_c)$ corresponds to the potential function. Z is a normalizing constant (partition function) that is

$$Z = \sum_x \exp\left(-\sum_{c\in Sc} \Phi_c(x_c)\right) < \infty. \tag{2}$$

It also renders

$$\sum_x \Pr(x) = 1 \text{ for } \quad \Pr(x) > 0. \tag{3}$$

Thus, finding a maximal clique and the set of all maximal cliques are easy to specify dependencies between random variables, and probability distribution is a connectivity pattern for graphical models related to clique problems from this theorem. This maximal clique is either a clique or an independent set that is not a proper