



# Computational methods for finding long simple cycles in complex networks



David Chalupa\*, Phininder Balaghan, Ken A. Hawick, Neil A. Gordon

Computer Science School of Engineering and Computer Science, University of Hull, Cottingham Road Hull HU6 7RX, United Kingdom

## ARTICLE INFO

### Article history:

Received 5 August 2016

Revised 24 March 2017

Accepted 29 March 2017

Available online 30 March 2017

### Keywords:

Long simple cycles

Long cycles

Complex networks

Integer linear programming

Graph algorithms

Local search

Hamiltonian cycles

## ABSTRACT

Detection of long simple cycles in real-world complex networks finds many applications in layout algorithms, information flow modelling, as well as in bioinformatics. In this paper, we propose two computational methods for finding long cycles in real-world networks. The first method is an exact approach based on our own integer linear programming formulation of the problem and a data mining pipeline. This pipeline ensures that the problem is solved as a sequence of integer linear programs. The second method is a multi-start local search heuristic, which combines an initial construction of a long cycle using depth-first search with four different perturbation operators. Our experimental results are presented for social network samples, graphs studied in the network science field, graphs from DIMACS series, and protein-protein interaction networks. These results show that our formulation leads to a significantly more efficient exact approach to solve the problem than a previous formulation. For 14 out of 22 networks, we have found the optimal solutions. The potential of heuristics in this problem is also demonstrated, especially in the context of large-scale problem instances.

© 2017 Elsevier B.V. All rights reserved.

## 1. Introduction

With the growing volume of information available as connected data, the need for *efficient algorithms to solve graph problems* has been becoming increasingly important [20].

The *longest simple cycle problem* is one of the classical NP-hard graph problems, in which one simply aims to find the longest simple cycle in an undirected graph. A simple cycle is defined as a connected subgraph of the graph with all vertices having degree 2 and without repetitions in vertices or edges. For simplicity, the term longest cycle will refer to the longest simple cycle hereafter.

A cycle which spans all vertices of the graph is called a Hamiltonian cycle. Hence, the longest cycle problem represents a generalisation of the decision problem for Hamiltonian cycle [26] which is widely studied [6,15,44].

Searching for the longest cycle in a graph has been a subject of study in statistical mechanics, with approaches based on message passing and Monte Carlo procedures being used [33]. However, studying the longest cycle problem in the context of both exact and experimental algorithms, as well as applications to real-world complex networks, has still been somewhat limited.

On the other hand, identifying long cycles finds its applications in *automatic drawing of planar graphs* [40] or in *layout algorithms for metabolic pathways* in bioinformatics [5]. It is also closely related to community structure [31], its hierarchy [11] and propagation processes [50] in real-world networks. To the best of our knowledge, the longest cycle problem has not yet been studied in the context of social networks nor several types of biological networks. It seems that a spectrum of potential applications has not been explored yet.

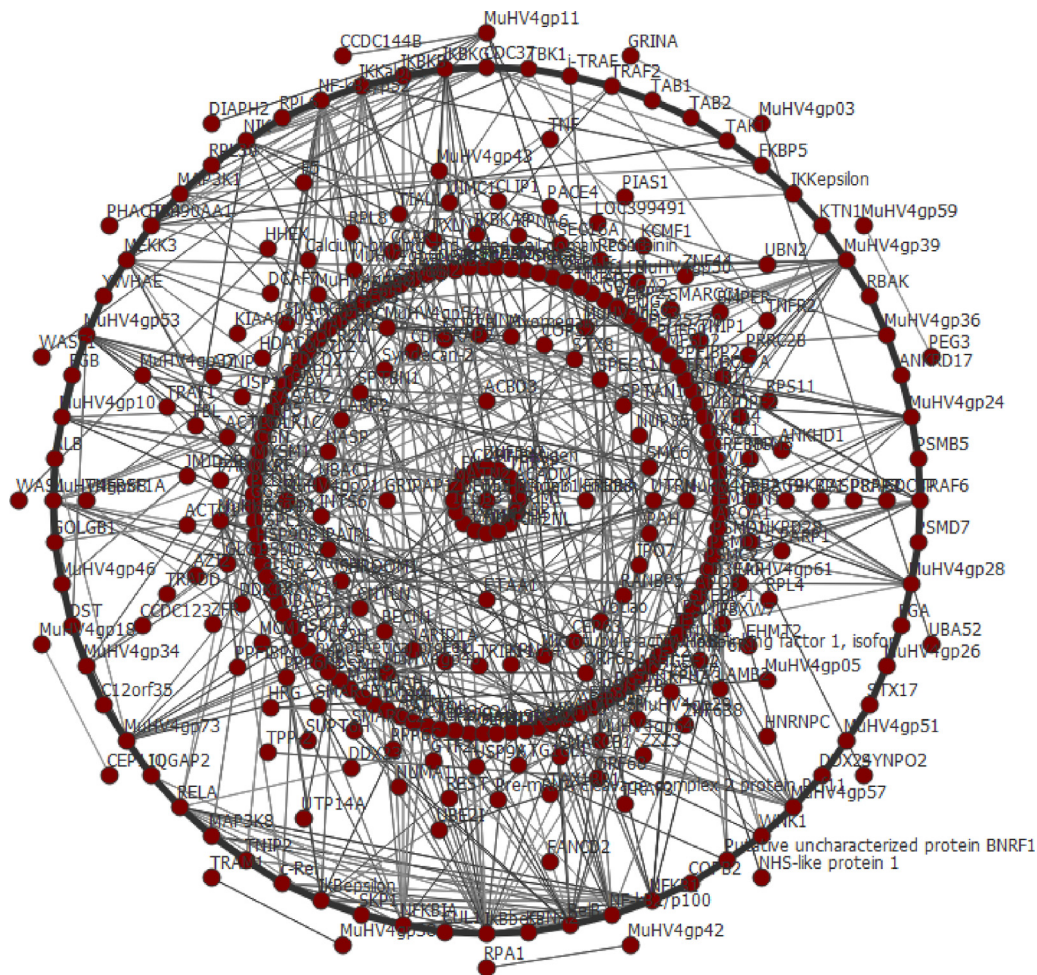
Previous scientific results on the longest cycle problem have been overwhelmingly concerned with its theoretical properties. Complexity and approximability of the problem have been studied in sparse graphs [16,37], bounded degree graphs [12], triangle-free graphs [3], and small graph classes [43].

Superpolylogarithmically long cycle detection [17], treewidth-based approximation [2,9], and matrix-based approximation [34] have been explored as avenues for efficient finding of approximate solutions to the problem.

Exact methods for the problem include enumerative techniques [21,23,41] and a branch-and-bound algorithm, which is based on an integer linear programming (ILP) formulation of the problem [14]. The enumerative techniques allow one to explore the distribution of the cycle numbers, as a function of the cycle length. This idea has successfully been explored for small worlds, as well as the Kauffman networks [22,28]. However, these algorithms are naturally computationally demanding. Therefore, for large graphs,

\* Corresponding author.

E-mail addresses: [d.chalupa@hull.ac.uk](mailto:d.chalupa@hull.ac.uk) (D. Chalupa), [p.balaghan@hull.ac.uk](mailto:p.balaghan@hull.ac.uk) (P. Balaghan), [k.a.hawick@hull.ac.uk](mailto:k.a.hawick@hull.ac.uk) (K.A. Hawick), [n.a.gordon@hull.ac.uk](mailto:n.a.gordon@hull.ac.uk) (N.A. Gordon).



**Fig. 1.** The longest cycle identified for a high-throughput human protein-protein interaction network *Hsapi20160114HT* from the UCLA database of interacting proteins [38,47–49]. Both of our approaches have found this cycle which consists of 64 vertices.

it is interesting to search for the location of the far end of this distribution using the identification of the longest cycle.

For the Hamiltonian cycle problem, approximation algorithms for maximal planar graphs have been explored [36]. Heuristics for the problem are a subject of interest, too. These include an ant-inspired heuristic [44] or an interior point heuristic [15]. Hamiltonian cycles in scale-free networks have also been previously explored [6]. The longest cycle problem is also often linked with the longest path problem, for which efficient algorithms [42] and approximations [25] have been studied. Some other studies have also been concerned with directed longest cycles [8].

Fig. 1 shows the longest cycle in a high-throughput human protein-protein interaction network, taken as one of the test instances for this paper from the UCLA database of interacting proteins [38,47–49]. This cycle consists of 64 vertices. This visualisation arranges the proteins of the longest cycle as an outer ring of the drawing. This shows the long cyclic dependency of the protein interactions. These are valuable in a reconstruction of the metabolic pathways [13], in which long cycles are used in layout algorithms for their visualisation [5].

In this present article we focus on computational methods for finding long cycles in real-world complex networks. These include social networks, protein-protein interactions, as well as networks from several other domains. Our approaches include both an *exact method* based on our own ILP formulation of the problem, as well as a heuristic combining *depth-first search* with a *local improvement procedure* using clustering properties of real-world networks.

Our main contribution is in proposing a new ILP formulation the longest cycle problem, as well as designing of a pipeline for efficient exact approach to find the longest cycle. Additionally, we propose a hybrid heuristic, which combines construction of a long cycle using depth-first search with four local search operators to improve this initial cycle. Our computational results are presented for a diverse set of real-world networks, indicating that both the size and the structure of the graph seem to influence the difficulty of the problem.

Our article is structured as follows. In Section 2 we review the existing ILP formulation of the longest cycle problem with fixed initial vertex and introduce our own ILP formulation of the same problem. We then prove the correctness of our formulation and propose a pipeline for the exact approach which can be used with both formulations. In Section 3 we describe our heuristic based on depth-first search and four local search operators. In Section 4 we present the computational results of the exact approach and compare it to the approach based on the previous formulation. We also show the result of the heuristic on a real-world network data set. In Section 5 we discuss the obtained results and their interpretation. We offer several conclusions and identify the open problems in Section 6.

## 2. Exact approach to the longest simple cycle problem

In this section, we first review the current ILP formulation of the longest simple cycle problem with fixed initial vertex [14].

Download English Version:

<https://daneshyari.com/en/article/4946287>

Download Persian Version:

<https://daneshyari.com/article/4946287>

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