On the Hardness and Inapproximability of Identifying Wheeler Graphs

Daniel Gibney
Department of Computer Science, University of Central Florida, Orlando FL, USA
Daniel.Gibney@ucf.edu
https://dblp.uni-trier.de/pers/hd/g/Gibney:Daniel

Sharma V. Thankachan
Department of Computer Science, University of Central Florida, Orlando FL, USA
sharma.thankachan@ucf.edu
http://www.cs.ucf.edu/ sharma/

Abstract
In recent years several variants of Borrows-Wheeler’s transformation based indexes have been introduced. Some of these indexes are used to index structures far more complex than a single string, as was originally done with the FM-index [Ferragina and Manzini, J. ACM 2005]. As such, there has been an increasing effort to better understand under which conditions such an indexing scheme is possible. This has led to the introduction of Wheeler graphs by Gagie et al. in [Theor. Comput. Sci., 2017]. A Wheeler graph is a directed graph with edge labels which satisfies two simple axioms. Importantly, Wheeler graphs can be indexed in a way which is space efficient and allows for the fast traversal of edges. Gagie et al. showed that de Bruijn graphs, generalized compressed suffix arrays, and several other BWT related structures can be represented as Wheeler graphs. However, one may also wish to know if a given graph is a Wheeler graph. Here we answer the open question of whether or not there exists an efficient algorithm for identifying if a graph is a Wheeler graph. We present the following results.

- The problem of identifying whether a given graph is a Wheeler graph is NP-complete for an edge label alphabet of size two or greater, even when the graph is a DAG;
- We define an optimization variant of the problem called Minimum Wheeler Graph Violation, abbreviated WGV, where the aim is to remove the minimum number of edges so as to obtain a Wheeler graph. We then show that WGV is APX-hard. This implies, unless P = NP, there exists a constant $C > 1$ for which there is no $C$-approximation algorithm. This holds even when the graph is a DAG;
- We show that conditioned on the Unique Games Conjecture, for every constant $C \geq 1$, it is NP-hard to find a $C$-approximation to WGV;
- We define the Maximum Wheeler Subgraph problem, abbreviated MWS, where the aim is to find the largest subgraph which is also a Wheeler Graph (the dual of the WGV). We provide a linear time $C$-approximation algorithm demonstrating that MWS is in the complexity class APX;
- We give a $2^{O(n)}$ time algorithm for the identification problem and $2^{O(n+e)}$ time algorithm for both optimization problems where $n$ is the number of vertices, $e$ the number of edges, and the alphabet size is fixed;
- We highlight the relationship between queue number and Wheeler graphs, which reveals that the identification problem can be solved in linear time for an edge alphabet of size one;
- We identify a class of graphs for which the identification problem is solvable in linear time.

2012 ACM Subject Classification Track A: Algorithms, complexity and games

Keywords and phrases Wheeler graph, NP Completeness, queue number

Digital Object Identifier 10.4230/LIPIcs.CVIT.2016.23