Classification of de Bruijn-based labeled digraphs

Marta Kasprzak[a,b]

[a] *Institute of Computing Science, Poznan University of Technology, Piotrowo 2, 60-965 Poznan, Poland*
[b] *Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland*

Abstract

Labeled digraphs, thanks to their special properties, are widely used in modeling real-world problems. Starting from de Bruijn graphs, they were used, among others, in modeling communication networks, architecture of parallel computers, or — in the area of bioinformatics — DNA sequencing and assembly problems. One of their most important property is polynomial-time solvability of the Hamiltonian cycle/path problem, which makes these graphs especially useful as computational models. The classification presented here shows relations between subclasses of labeled digraphs, such as de Bruijn graphs, DNA graphs and others, and their connection with adjoints and quasi-adjoint graphs. The most recently defined class of quasi-adjoint graphs has a widest applicability, since it contains as subclasses all the de Bruijn-based labeled digraphs considered in this paper. The current work can be treated as a support in choosing an appropriate combinatorial model, resulting in polynomial time solution of problems related to searching for the Hamiltonian cycle or path, which are strongly NP-hard in general.

Keywords: directed line graph, quasi-adjoint graph, DNA graph, alphabet overlap digraph, Hamiltonian cycle/path problem, DNA sequence assembly

1. Introduction

Labels in graphs can take various forms, here the interest is focused on labels restricted to the form first defined by de Bruijn in 1946 [17]. They are finite strings of equal lengths over an alphabet, attributed to vertices in such a manner that labels of adjacent vertices overlap. Overlapping depends on edge directions — the suffix of length $k-1$ of the predecessor's label is equal to the prefix of the successor's label, where k is the label length — thus graphs examined in this context are directed. All occurrences of such overlaps of vertex labels are represented by arcs in these graphs.

The restriction on composition of labels in a neighborhood entails a special structure of such graphs. Sets of immediate successors of any two vertices in the graph must be the same or disjoint. Further restrictions are present in subclasses of the general class. De Bruijn graphs are one of the subclasses where all possible labels of a given length, constructed over a given alphabet, are associated with vertices. A structure of the de Bruijn graphs has been useful, for example, in modeling communication networks. Other subclasses, DNA graphs and alphabet overlap digraphs, have been used in bioinformatics as models of DNA sequencing and assembly problems. All the labeled graphs and, more broadly, quasi-adjoint graphs are polynomial-time solvable instances of the problem of searching for the Hamiltonian cycle or Hamiltonian path, while in general the problems are computationally hard (strongly NP-hard).

This polynomial time solvability of the Hamiltonian path problem made graphs of such a kind especially suitable for reconstructing linear structures of DNA and RNA. One of the former approaches, sequencing by hybridization, became an inspiration for defining new graph classes and for research in this area. The process of sequencing by hybridization, in its computational part, consists in linear ordering of short sequences of nucleotides (oligonucleotides, k-mers) given at the input in such a way, that a longer sequence is obtained and each two input sequences neighboring in the result overlap on $k-1$ nucleotides. Two classical approaches proposed by Lysov *et al.* [24] and Pevzner [27] showed that, in the context of DNA sequencing, the problems of searching for the Hamiltonian path and for the Eulerian path are equivalent. This fact generated an important question of computational complexity nature and

Email address: marta@cs.put.poznan.pl (Marta Kasprzak)

resulted in a series of papers, e.g. [10, 26, 19, 22, 14, 23, 13]. The graphs used in these approaches were later assigned to the classes of DNA graphs (Lysov graphs) and subgraphs of de Bruijn graphs (Pevzner graphs).

The algorithmic part of sequencing by hybridization is basically very similar to current algorithmic approaches solving the DNA assembly problem. Novel next-generation sequencing methods produce great volumes of reads which can be treated somehow similarly as oligonucleotides in sequencing by hybridization (now on a much larger scale). The reads being decomposed in assembly algorithms into series of shorter k-mers resemble oligonucleotides even more. Therefore, the graph models from the computational part of the sequencing process are now widely applied in sequence assembly. These DNA assembly methods which refer to models from graph theory can be roughly divided into two groups: one exploiting models "reads on vertices", i.e. DNA graphs-like (more popular previously, see e.g. [20]), the second one basing on the Pevzner graphs, with models of type "(decomposed) reads on arcs" (more popular nowadays, see e.g. [25]). Due to the presence of sequencing errors in data and data incompleteness, these models only approximately match the original problems of searching for the Hamiltonian path or the Eulerian path, respectively, in an appropriate graph.

The classification presented here shows relations between subclasses of labeled digraphs, such as de Bruijn graphs, DNA graphs and others, and their connection with adjoints and quasi-adjoint graphs. Since it provides important information about polynomial time solvability of the Hamiltonian path and cycle problems in these graphs, it is especially useful as a support in the process of choosing an appropriate combinatorial model for a real-world problem.

The graphs systematized in the paper were used in applications in their labeled form, but some of the classes are defined without referring to labels. In fact, these graphs do not need to be labeled and they can be viewed as *labelable* as well.

The organization of the paper is as follows. In Section 2 all the graph classes are defined. Section 3 contains the systematization together with a proof of its correctness. In Section 4 the solvability of the Hamiltonian cycle problem in the mentioned graphs is discussed. A conclusion is given in Section 5.

2. Definitions

The following statements refer to *directed graphs* and 1-graphs (graphs having no multiple arcs), unless stated otherwise. Throughout the paper a standard terminology from graph theory is used, see e.g. [5, 15]. Let V stands for the set of vertices of a graph $G = (V, A)$ and A for its set of arcs, $l(v)$ for the label of vertex v, $\mathrm{suf}_i(s)$ for the suffix of length i of string s, and $\mathrm{pre}_i(s)$ for the prefix of length i of s.

De Bruijn graphs are labeled digraphs, which are complete with respect to the size of an alphabet and the length of labels [17]. For an alphabet of size α and labels of constant length k ($k > 1, \alpha > 0$), de Bruijn graph $B(\alpha, k)$ has α^k vertices, every one labeled by a different word over the alphabet. For all $u, v \in V$, where $|V| = \alpha^k$, $(u \neq v \Rightarrow l(u) \neq l(v))$ and $((u, v) \in A \Leftrightarrow \mathrm{suf}_{k-1}(l(u)) = \mathrm{pre}_{k-1}(l(v)))$.

DNA graphs (Lysov graphs) are graphs used in modeling the problem of DNA sequencing by hybridization. The alphabet is restricted to four letters standing for four nucleotides encoding, within a DNA, genetic information of an organism. Labels of vertices represent short DNA sequences of equal length k and there is the assumption that all sequences are different. The set of input DNA sequences is denoted by S. Arcs show possible overlaps on $k - 1$ letters of the sequences from S and paths in a DNA graph correspond to longer DNA chains. A proper solution to the DNA sequencing problem, if no experimental error is assumed to be present in the instance, is the Hamiltonian path in the graph [24]. In a DNA graph, for all $u, v \in V$, $(u \neq v \Rightarrow l(u) \neq l(v))$ and $((u, v) \in A \Leftrightarrow \mathrm{suf}_{k-1}(l(u)) = \mathrm{pre}_{k-1}(l(v)))$. DNA graphs are vertex-induced subgraphs of de Bruijn graphs with $\alpha = 4$ [10].

Pevzner graphs realize a different model of DNA sequencing by hybridization. There, each k-mer from the input set S is associated with an arc which goes from a vertex labeled by its prefix to a vertex labeled by its suffix, the vertex labels having length $k - 1$. A solution to the problem is the Eulerian path in the graph [27]. In a Pevzner graph, $V = \{\mathrm{pre}_{k-1}(s), \mathrm{suf}_{k-1}(s) \,|\, s \in S\}$ and $A = \{(\mathrm{pre}_{k-1}(s), \mathrm{suf}_{k-1}(s)) \,|\, s \in S\}$. Pevzner graphs are subgraphs of DNA graphs, thus in consequence, subgraphs of de Bruijn graphs (edge-induced). The class of Pevzner graphs does not belong as a whole to the class of labeled/DNA graphs, because there is not 1–1 correspondence between the presence of arcs in such a graph and the overlaps of vertex labels.

As defined in [5] in the context of directed graphs, *adjoint* $G = (V, A)$ of a graph $H = (U, V)$ is a 1-graph whose vertices represent arcs of H, and which has arc (u, v) if and only if the head of arc u in H is the tail of arc v. H is not necessarily a 1-graph. If H is a 1-graph, its adjoint G is a *directed line graph* [10].

1-graph G is an adjoint if and only if the following property is satisfied for all $u, v \in V$:

$$N^+(u) \cap N^+(v) \neq \varnothing \quad \Rightarrow \quad N^+(u) = N^+(v),$$

where $N^+(u)$ is the set of immediate successors of vertex u [5]. 1-graph G is a directed line graph if and only if, for all $u, v \in V$, the following property is satisfied:

$$N^+(u) \cap N^+(v) \neq \varnothing \quad \Rightarrow \quad \begin{array}{ll} N^+(u) = N^+(v) & \wedge \\ N^-(u) \cap N^-(v) = \varnothing, & \end{array}$$

where $N^-(u)$ is the set of immediate predecessors of vertex u [10].

A directed line graph is such an adjoint which does not contain as a subgraph any of the structures present in the left part of Fig. 1 [10]. These structures, after the transformation to an *original graph* H, are represented by multiple arcs. Therefore, H would be then a multigraph.

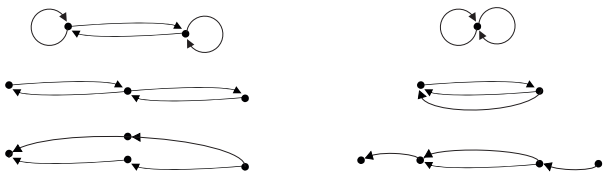

Figure 1: Subgraphs which can be a part of an adjoint but are not present in directed line graphs (on the left) and corresponding structures with parallel arcs in an original graph (on the right)

The graph constructed in the method of Lysov *et al.* is a directed line graph of the Pevzner graph for the same set of input sequences S.

According to definitions in [10], a directed 1-graph G belongs to class $\mathscr{L}_k^\alpha$ (can be (α, k)-*labeled*) if, for all $u, v \in V$, $(u \neq v \Rightarrow l(u) \neq l(v))$ and $((u, v) \in A \Leftrightarrow \mathrm{suf}_{k-1}(l(u)) = \mathrm{pre}_{k-1}(l(v)))$, where $k > 1$ is the length of labels and $\alpha > 0$ is the alphabet size. Labeled graphs (*uniquely labeled graphs*) are these graphs which belong to a class $\mathscr{L}_k^\alpha$ for some α and k. Graphs satisfying the above requirements except the condition that labels must be different are called *non-uniquely labeled graphs*. Graphs from classes $\mathscr{L}_k^4$ are DNA graphs. De Bruijn graphs $B(\alpha, k)$ are 'complete' labeled graphs from $\mathscr{L}_k^\alpha$, for any k and α ('complete' in the sense of containing all possible labels of given parameters).

Properties of graphs from classes $\mathscr{L}_k^\alpha$ $(k > 1, \alpha > 0)$ were studied in a series of papers, see e.g. [10, 8, 26, 19, 22, 23]. The most interesting one from the current point of view is the theorem from [10], that a graph is a directed line graph if and only if it belongs to class $\mathscr{L}_2^\infty$ (i.e. it can be labeled over an alphabet of unbounded size with labels of length 2). Completed with another theorem from that paper, it leads to the conclusion that the classes of directed line graphs and graphs that can be (α, k)-labeled (for all $\alpha > 0$ and $k > 1$) are the same.

In [19] a *self-adjoint* was defined, as a graph isomorphic to its adjoint. In [18] the idea of *alphabet overlap graphs* was introduced. They are undirected graphs similar to de Bruijn graphs in the sense of completeness of their vertex set and overlapping of labels, but undirected edges replace arcs and edges corresponding to shifts of labels greater than one position are allowed. On the basis of this definition, *alphabet overlap digraphs* were proposed as a generalization of de Bruijn graphs [23]. Given three integers, $\alpha \geq 1, k \geq 2$ and $1 \leq i < k$, alphabet overlap digraph $O(\alpha, k; i)$ is defined as a graph labeled with all possible words of length k over an alphabet of size α, where i is a fixed shift in overlaps of vertex labels. For all $u, v \in V$, where $|V| = \alpha^k$, $(u \neq v \Rightarrow l(u) \neq l(v))$ and $((u, v) \in A \Leftrightarrow \mathrm{suf}_{k-i}(l(u)) = \mathrm{pre}_{k-i}(l(v)))$. In [21] the same graphs but with $1 \leq i \leq k$ were defined as lexical graphs and their vertex-induced subgraphs as base-labeled graphs.

In Section 3 the relationship between the above digraphs is shown. The relations described in Section 4 need two additional definitions.

Digraph G is a *quasi-adjoint graph* if and only if, for all $u, v \in V$, the following property holds [14]:

$$N^+(u) \cap N^+(v) \neq \varnothing \quad \Rightarrow \quad \begin{array}{ll} N^+(u) = N^+(v) & \vee \\ N^+(u) \subset N^+(v) & \vee \\ N^+(v) \subset N^+(u). & \end{array}$$

Quasi-adjoint graphs, unlike adjoints, can be multigraphs.

A set of vertex labels of digraph G constitutes the *injective overlap labeling* if and only if they have the same length k and, for all $u, v \in V$, $(u \neq v \Rightarrow l(u) \neq l(v))$ and $((u, v) \in A \Leftrightarrow \mathrm{suf}_{k-i}(l(u)) = \mathrm{pre}_{k-i}(l(v)))$, where $1 \leq i < k$ and i is not fixed within a graph [16].

3. Classification

The classes of labeled digraphs are aggregated in Fig. 2.

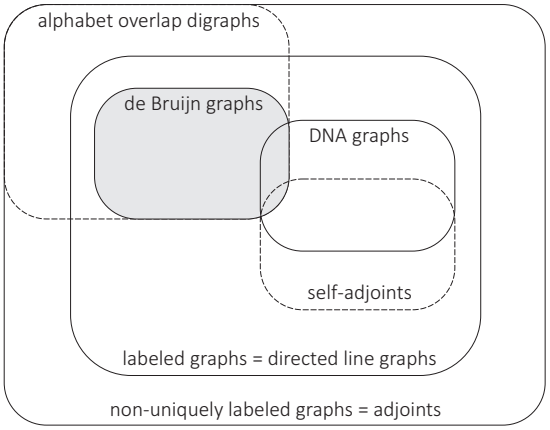

Figure 2: Classes of de Bruijn-based labeled digraphs

Theorem. The relations between the digraph classes: de Bruijn graphs $\mathscr{B}$, DNA graphs $\mathscr{G}_d$, labeled graphs $\mathscr{L}$, directed line graphs $\mathscr{G}_l$, non-uniquely labeled graphs $\widetilde{\mathscr{L}}$, adjoints $\mathscr{G}_a$, self-adjoints $\mathscr{G}_s$, and alphabet overlap digraphs $\mathscr{O}$, are characterized by the following properties (presented together in Fig. 2).

(A) $\mathscr{B} \cap \mathscr{G}_d \neq \varnothing$,
$\mathscr{B} \cap \mathscr{G}_s \neq \varnothing$,
$\mathscr{G}_d \cap \mathscr{G}_s \neq \varnothing$,
$\mathscr{B} \cap \mathscr{G}_d \cap \mathscr{G}_s \neq \varnothing$,
$\mathscr{B} \cap \mathscr{G}_s \subset \mathscr{B} \cap \mathscr{G}_d$,
$\mathscr{B} \cap \mathscr{G}_s \subset \mathscr{G}_d \cap \mathscr{G}_s$.

(B) $\mathscr{L} = \mathscr{G}_l$,
$\mathscr{B} \subset \mathscr{G}_l$,
$\mathscr{G}_d \subset \mathscr{G}_l$,
$\mathscr{G}_s \subset \mathscr{G}_l$.

(C) $\widetilde{\mathscr{L}} = \mathscr{G}_a$,
$\mathscr{G}_l \subset \mathscr{G}_a$,
$\mathscr{O} \subset \mathscr{G}_a$.

(D) $\mathscr{B} \subset \mathscr{O}$,
$\mathscr{O} \cap (\mathscr{G}_d \setminus \mathscr{B}) = \varnothing$,
$\mathscr{O} \cap (\mathscr{G}_s \setminus \mathscr{B}) = \varnothing$.

(E) $(\mathscr{O} \setminus \mathscr{B}) \cap \mathscr{G}_l \neq \varnothing$,
$(\mathscr{O} \setminus \mathscr{B}) \cap (\mathscr{G}_a \setminus \mathscr{G}_l) \neq \varnothing$,
$(\mathscr{G}_a \setminus \mathscr{G}_l) \setminus \mathscr{O} \neq \varnothing$,
$((\mathscr{G}_l \setminus \mathscr{B}) \setminus \mathscr{G}_d) \setminus \mathscr{G}_s \neq \varnothing$.

Proof. **(A)** The common part of the classes of de Bruijn graphs and DNA graphs contains 'complete' (in the de Bruijn sense) labeled graphs with labels over an alphabet of size at most 4. The alphabet size can be less than 4, because a graph belonging to $\mathscr{L}_k^\alpha$ belongs also to $\mathscr{L}_k^\beta$, $\alpha < \beta$ [10]. A de Bruijn graph rarely is a self-adjoint, if and only if $\alpha = 1$ (thus, it is also a DNA graph). It is then a single vertex with a self-loop, where the value of k is arbitrary. In other cases the de Bruijn graph contains (from the definition) more arcs than vertices, therefore, its adjoint cannot be isomorphic to it. An example graph from the intersection of the classes of DNA graphs and self-adjoints, which is not a de Bruijn graph, is shown in Fig. 3, where also other example graphs are visualized.

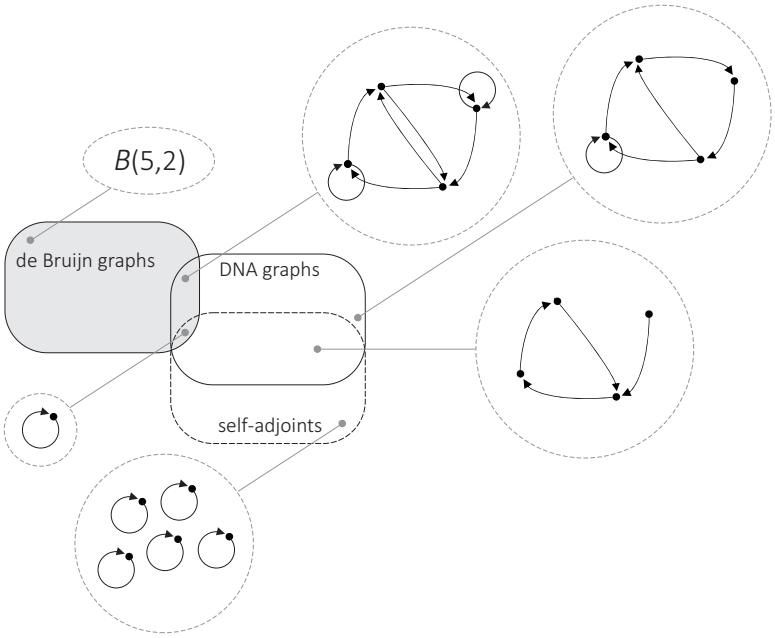

Figure 3: Example graphs from subclasses analyzed in part A of the theorem. De Bruijn graphs outside DNA graphs are the ones built with $\alpha > 4$, with the example $B(5, 2)$ put symbolically due to its size

(B) The equivalence of the classes of directed line graphs and labeled graphs was proven in [10] and quoted in Section 2. Self-adjoints are adjoints, according to the definition. Furthermore, they all must be line graphs; otherwise, they would have multiple arcs and that is inadmissible in adjoints (original graphs of adjoints not being directed line graphs are multigraphs). As de Bruijn graphs and DNA graphs can be uniquely labeled (see definitions), these classes also belong to directed line graphs. An example of a directed line graph outside the classes of de Bruijn graphs, DNA graphs and self-adjoints, is shown in Fig. 4. It cannot be a DNA graph, because it must be labeled over, at least, 5-letter alphabet (it has 5 self-loops). It cannot be a de Bruijn graph, because for the 5-letter alphabet we would expect at least 25 vertices. And it cannot be a self-adjoint, because it has more arcs than vertices. On the other hand, a set of its labels presented in the example below the proof shows that it can be uniquely labeled.

(C) The equivalence of the classes of adjoints and non-uniquely labeled graphs was proven in [10]. An adjoint not being a directed line graph cannot be uniquely labeled, since every one such a graph contains at least one structure resulting in multiple arcs in its original graph H, see Fig. 1. Vertices in these structures will have duplicated labels corresponding to parallel arcs in an original graph. An alphabet overlap digraph always is an adjoint, because with the same value of its parameter i all pairs of vertices in the graph have sets of their immediate successors either disjoint or the same. An example of an adjoint not being an alphabet overlap digraph nor a directed line graph is shown in Fig. 4. Because of a subgraph presented in Fig. 1, it cannot be a directed line graph, and cannot be an alphabet overlap digraph due to too small number of vertices. For a smallest alphabet of size 2, we would expect at least 4 vertices. On the other hand, a set of example labels showing that the graph can be non-uniquely labeled is presented below the proof.

(D) From the respective definitions we see, that de Bruijn graphs $B(\alpha, k)$ constitute subclass $O(\alpha, k; 1)$ of alphabet overlap digraphs. (Let us note that the same graph can be characterized by different parameters as an alphabet overlap

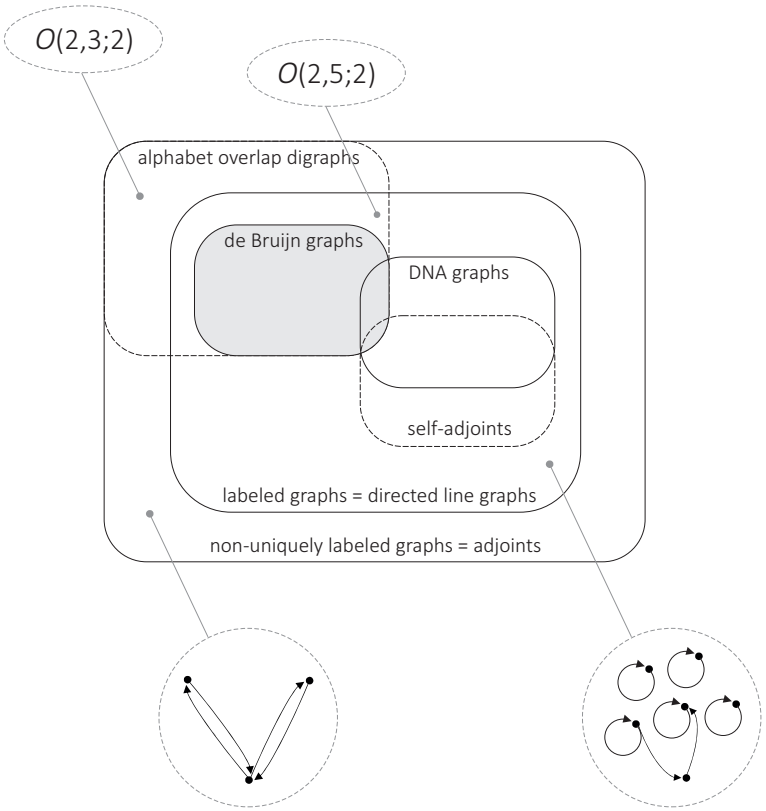

Figure 4: Example graphs from subclasses analyzed in parts B–E of the theorem. Alphabet overlap digraphs $O(2,3;2)$ and $O(2,5;2)$ are put symbolically due to their size

digraph. One of the examples is a single loop, which can be seen as any of the graphs $O(1,k;i)$, for all k and i. Another example is the pair $O(4,2;1)$ and $O(2,4;2)$ resulting in the same structure. Therefore, a de Bruijn graph can be formally written also as an alphabet overlap digraph with parameter $i > 1$.) All graphs $O(\alpha,k;i)$ with $\alpha > 1$ contain more arcs than vertices, thus they cannot be self-adjoints. All alphabet overlap digraphs which can be labeled as DNA graphs (i.e. with labels of adjacent vertices overlapping on $k-1$ positions) are also de Bruijn graphs.

(E) One of the alphabet overlap digraphs not being a de Bruijn graph but belonging to directed line graphs is $O(2,5;2)$. As an alphabet overlap digraph it is an adjoint and does not contain any of the structures from the left part of Fig. 1. One of the alphabet overlap digraphs being an adjoint but not a directed line graph is $O(2,3;2)$. It contains as a subgraph the first structure from the left part of Fig. 1. In Fig. 4 there are shown example elements of other subclasses mentioned in part E of the theorem. □

An example assignment of labels to vertices of graphs presented in Fig. 3 is as follows. The self-loop can be labeled with AA, the set of five loops with AA, BB, CC, DD, and EE. The DNA graph being also a de Bruijn graph can have AA, AC, CC, and CA as labels (with AA and CC at self-loops). The DNA graph being also a self-adjoint can have labels AC, CG, GT, and TC, where AC is assigned to the vertex with zero indegree. The DNA graph being outside the classes of de Bruijn graphs and self-adjoints can be labeled with AA, AC, CT, and TA. The graph containing self-loops in Fig. 4 may have assigned the labels AA, BB, CC, DD, EE, and DE. The remaining graph in Fig. 4 can be labeled with AB, AB, and BA.

4. Solvability of HCP

The Hamiltonian cycle problem (HCP) is frequently studied for its significance in both theoretical and practical branches of computer science. It is one of the best known combinatorial problems, commonly used for modeling more complex and real-life problems. The same concerns the Hamiltonian path problem. Both are strongly NP-complete and it is a strong motivation for discovering classes of graphs being polynomial-time solvable instances of these problems. While many papers were dedicated to classes of undirected graphs easy from the point of view of searching for HCP, analogous classes of directed graphs have not been so widely studied (see e.g. [4, 3]).

For directed graphs, the Hamiltonian cycle/path problem is easily solvable in an adjoint by transforming it into its original graph (H) and then by searching for the Eulerian cycle/path within it. The existence of the Eulerian cycle/path in the original directed graph is the necessary and sufficient condition of the existence of the Hamiltonian cycle/path in its adjoint [10]. What is interesting, this transformation does not work for undirected graphs [6], thus works utilizing this transformation must be restricted to directed graphs.

The widest class of digraphs, found so far, 'easy' for HCP and containing the labeled graphs, is the class of *quasi-adjoint graphs* [14]. It is a superclass of, among others, adjoints and graphs modeling the problem of isothermic DNA sequencing by hybridization. (For description of the latter problem and model see [11], the problem introduced in [9]. Such graphs can be either directed line graphs, adjoints not being directed line graphs, or can be outside adjoints; however, they always are 1-graphs.) The quasi-adjoint graphs are no longer labeled (in the de Bruijn sense) because of lack of the property that, for any pair of vertices, sets of their immediate successors are the same or disjoint (see the definition in Section 2). The recognition whether a graph is a quasi-adjoint graph can be done in $O(n^3)$ time and an exact algorithm solving the Hamiltonian cycle problem within it works in $O(n^2 + m^2)$ time, where n is the number of its vertices, m the number of arcs. The algorithm also utilizes a transformation of a graph G into its original graph H, but this time it is more complicated and includes inserting extra structures (the numbers of vertices before the transformation and arcs after the transformation differ) [14].

The same algorithm can be used for solving the Hamiltonian path problem. If we do not know the first and last vertices of the path, we simply run the algorithm at most $O(n^2)$ times with successive pairs of vertices taken as the ends of the path. In one such run, all arcs incoming to the first vertex are removed from the graph, as well as all arcs outgoing from the last vertex, and the arc from the last vertex to the first one is added. If the initial graph is a quasi-adjoint graph, after this modification it still belongs to that class. Sometimes a graph not being a quasi-adjoint graph becomes such a graph due to the modification, but it does not change the answer to the Hamiltonian path problem (only these arcs are removed, which will never appear in any solution).

It should be noted that there are works somehow related to the graph transformation considered here. Blais and Laporte dealed with a transformation for directed, undirected, and mixed graphs, related to the generalized routing problem [7]. The transformation replaces vertices, arcs and edges of the initial graph by vertices in a new complete weighted digraph, so it is carried out inversely to the one for adjoints or quasi-adjoint graphs. This approach preserves computational hardness of the problems before and after the transformation. Apollonio and Franciosa defined and analyzed *partial directed line graphs* in [1]. These graphs are subgraphs of directed line graphs and constitute a superclass of the latter. There, a minimum completion algorithm was proposed which transforms a partial directed line graph to a directed line graph. However, it does not keep the property of the existence of the Hamiltonian cycle in the graph before and after the completion. This concept was continued in *path partial directed line graphs* involved in the theory of directed path families [2].

Digraphs possessing the *injective overlap labeling* (the definition in Section 2) do not have such a relation of sets of successors as adjoints and quasi-adjoint graphs have. There, overlaps are allowed to be variable in size within one digraph. A vertex labeled, for example, with ABC can have arcs outgoing to BCD and CEF, while from BAC arcs can go to CEF and ACD. Thus, the useful transformation vertex-arc, which changes the computational complexity of HCP, cannot be applied to such graphs. The length k of labels in the injective overlap labeling was investigated by Chikhi and co-authors in [16], where they showed that the parameter readability (which is the smallest possible k) for digraphs has the lower and upper bounds expressed by $\Omega(n)$ and $O(2^n)$, respectively. They posed open questions, one of them being "given a digraph, is it NP-hard to compute its readability?". Somewhat similar questions were solved for graphs from classes $\mathscr{L}_k^\alpha$. In [26] three problems were proven to be NP-hard: to decide whether a digraph belongs to $\mathscr{L}_k^\alpha$ for a given α and any fixed $k \geq 3$; to decide whether a digraph belongs to $\mathscr{L}_k^\alpha$ for a given k and any fixed $\alpha \geq 3$; to decide whether a digraph belongs to $\mathscr{L}_\infty^\alpha$ for any fixed $\alpha \geq 3$. A result on a maximum length of labels

for unbounded alphabet was given in [8], where it was proven that $G \in \mathscr{L}_{2n}^{\infty} \Leftrightarrow G \in \mathscr{L}_{k}^{\infty}$ for every $k \geq 2$. That is, the length of labels has no upper bound if the graph can be labeled with $k = 2n$, which is verified in polynomial time.

In [12] a systematization of several classes of digraphs with reference to HCP solvability was provided. The currently analyzed classes from Fig. 2, as subclasses of the quasi-adjoint graphs, are instances of HCP solvable in polynomial time.

5. Conclusion

There are a number of classes of graphs labelable in the de Bruijn sense with a practical application, among others, in bioinformatics. These graphs are not only nice theoretical models of real-world problems, but, first and foremost, they provide exact polynomial-time solutions of the problems recalling the Hamiltonian cycle or path. The class of quasi-adjoint graphs, as the widest class considered in this context, is potentially most useful. The classification presented here organizes subclasses of the quasi-adjoint graphs and helps researchers to choose an appropriate combinatorial model.

Acknowledgement

The research has been supported by grant No. 2012/05/B/ST6/03026 from the National Science Centre, Poland.

References

[1] N. Apollonio and P.G. Franciosa, A characterization of partial directed line graphs, *Discrete Mathematics* 307 (2007) 2598–2614.

[2] N. Apollonio and P.G. Franciosa, On the complexity of recognizing directed path families, *Discrete Applied Mathematics* 157 (2009) 2525–2535.

[3] J. Bang-Jensen and G. Gutin, On the complexity of hamiltonian path and cycle problems in certain classes of digraphs, *Discrete Applied Mathematics* 95 (1999) 41–60.

[4] J. Bang-Jensen, G. Gutin, and A. Yeo, A polynomial algorithm for the Hamiltonian cycle problem in semicomplete multipartite digraphs, *Journal of Graph Theory* 29 (1998) 111–132.

[5] C. Berge, *Graphs and Hypergraphs*, North-Holland Publishing Company, London, 1973.

[6] A.A. Bertossi, The edge Hamiltonian path problem is NP-complete, *Information Processing Letters* 13 (1981) 157-159.

[7] M. Blais and G. Laporte, Exact solution of the generalized routing problem through graph transformations, *Journal of the Operational Research Society* 54 (2003) 906-910.

[8] J. Blazewicz, P. Formanowicz, M. Kasprzak, and D. Kobler, On the recognition of de Bruijn graphs and their induced subgraphs, *Discrete Mathematics* 245 (2002) 81–92.

[9] J. Blazewicz, P. Formanowicz, M. Kasprzak, and W.T. Markiewicz, Sequencing by hybridization with isothermic oligonucleotide libraries, *Discrete Applied Mathematics* 145 (2004) 40–51.

[10] J. Blazewicz, A. Hertz, D. Kobler, and D. de Werra, On some properties of DNA graphs, *Discrete Applied Mathematics* 98 (1999) 1–19.

[11] J. Blazewicz and M. Kasprzak, Computational complexity of isothermic DNA sequencing by hybridization, *Discrete Applied Mathematics* 154 (2006) 718–729.

[12] J. Blazewicz and M. Kasprzak, Reduced-by-matching graphs: toward simplifying Hamiltonian circuit problem, *Fundamenta Informaticae* 118 (2012) 225–244.

[13] J. Blazewicz and M. Kasprzak, Complexity issues in computational biology, *Fundamenta Informaticae* 118 (2012) 385–401.

[14] J. Blazewicz, M. Kasprzak, B. Leroy-Beaulieu, and D. de Werra, Finding Hamiltonian circuits in quasi-adjoint graphs, *Discrete Applied Mathematics* 156 (2008) 2573–2580.

[15] G. Chartrand and L. Lesniak, *Graphs and Digraphs*, Wadsworth & Brooks/Cole, Pacific Grove, 1986.

[16] R. Chikhi, P. Medvedev, M. Milanic, and S. Raskhodnikova, On the readability of overlap digraphs, *Discrete Applied Mathematics* 205 (2016) 35-44.

[17] N.G. de Bruijn, A combinatorial problem, *Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen* 49 (1946) 758–764.

[18] A.P. Godbole, D. Knisley, R. Norwood, Some properties of alphabet overlap graphs, http://arxiv.org/abs/math/0510094, 2005.

[19] J. Hao, The adjoints of DNA graphs, *Journal of Mathematical Chemistry* 37 (2005) 333–346.

[20] J.D. Kececioglu, E.W. Myers, Combinatorial algorithms for DNA sequence assembly, *Algorithmica* 13 (1995) 7-51.

[21] A. Kozak, T. Glowacki, P. Formanowicz, On a generalized model of labeled graphs, *Discrete Applied Mathematics* 161 (2013) 1818-1827.

[22] X. Li, H. Zhang, Characterizations for some types of DNA graphs, *Journal of Mathematical Chemistry* 42 (2007) 65–79.

[23] X. Li, H. Zhang, Embedding on alphabet overlap digraphs, *Journal of Mathematical Chemistry* 47 (2010) 62–71.

[24] Yu.P. Lysov, V.L. Florentiev, A.A. Khorlin, K.R. Khrapko, V.V. Shik, A.D. Mirzabekov, Determination of the nucleotide sequence of DNA using hybridization with oligonucleotides. A new method, *Doklady Akademii Nauk SSSR* 303 (1988) 1508–1511.

[25] P. Medvedev, S. Pham, M. Chaisson, G. Tesler, P. Pevzner, Paired de Bruijn graphs: a novel approach for incorporating mate pair information into genome assemblers, *Lecture Notes in Computer Science* 6577 (2011) 238–251.

[26] R. Pendavingh, P. Schuurman, G.J. Woeginger, Recognizing DNA graphs is difficult, *Discrete Applied Mathematics* 127 (2003) 85–94.

[27] P.A. Pevzner, *l*-tuple DNA sequencing: computer analysis, *Journal of Biomolecular Structure and Dynamics* 7 (1989) 63–73.