Journal of Mathematical Chemistry Vol. 42, No. 1, July 2007 (© 2006)
DOI: 10.1007/s10910-006-9099-4

Characterizations for some types of DNA graphs

Xianyue Li and Heping Zhang*
School of Mathematics and Statistics, Lanzhou University, Lanzhou, Gansu 730000, P.R. China
E-mails: lixiany03@st.lzu.edu.cn, zhanghp@lzu.edu.cn

Received 22 January 2006; revised 12 February 2006

Vertex induced subgraphs of directed de Bruijn graphs with labels of fixed length k and over α letter alphabet are (α, k)-labelled. DNA graphs are $(4, k)$-labelled graphs. Pendavingh et al. proved that it is NP-hard to determine the smallest value $\alpha_k(D)$ for which a directed graph D can be $(\alpha_k(D), k)$-labelled for any fixed $k \geqslant 3$. In this paper, we obtain the following formulas: $\alpha_k(C_n) = \lceil \sqrt[k-1]{n} \rceil$ and $\alpha_k(P_n) = \lceil \sqrt[k-1]{n+1} \rceil$ for cycle C_n and path P_n. Accordingly, we show that both cycles and paths are DNA graphs. Next we prove that rooted trees and self-adjoint digraphs admit a (Δ, k)-labelling for some positive integer k and they are DNA graphs if and only if $\Delta \leq 4$, where Δ is the maximum number in all out-degrees and in-degrees of such digraphs.

KEY WORDS: DNA graph, de Bruijn graph, (α, k)-labelling

1. Introduction

Błażewicz et al. [4] introduced DNA graphs, which have vertices labelled in a special way by words over an alphabet $\{A, C, G, T\}$ corresponding to the four nucleotides of DNA chains: adenine, cytosine, guanine and thymine. Such graphs are used in the computational and reconstruction phase of DNA chain sequencing by hybridization (SBH) [1].

For a directed graph D with vertex-set $V(D)$ and arc-set $A(D)$, we assign every vertex v a label with length k as $(l_1(v), \ldots, l_k(v))$ such that every $l_i(v)$ belongs to the set $\{1, \ldots, \alpha\}$. Such a labelling is called an (α, k)-*labelling* if the distinct vertices of D have different labels, and for any arc (u, v) of H, $l_i(u) = l_{i-1}(v)$ for $i = 2, \ldots, k$ and vice versa. For given $k > 1$ and $\alpha > 0$, if D has an (α, k)-labelling, we call that D can be (α, k)-*labelled*. Figure 1 shows a digraph D with a $(3, 3)$-labelling. Hence D is $(3, 3)$-labelled.

Let $D = (V, A)$ be a digraph. For any arc $e = (u, v)$ of D, u is called the *tail* of e and v the *head* of e. For any given vertex v of D, a vertex w of D is an *in-neighbour* or *out-neighbour* of v according as (w, v) or (v, w) is an arc of D. The number of in-neighbors of v is called the *in-degree* of v, denoted by

*Corresponding author.

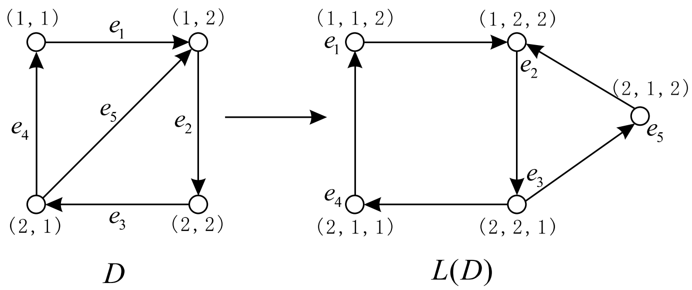

Figure 1. A digraph D with a (3, 3)-labelling.

$d^-(v)$. Similarly, the *out-degree* $d^+(v)$ of v is the number of out-neighbours of v. The *maximum out-degree* and *maximum in-degree* of D are defined, respectively, as $\Delta^+(D) = \max\{d^+(v) : v \in V(D)\}$ and $\Delta^-(D) = \max\{d^-(v) : v \in V(D)\}$. Put $\Delta(D) := \max\{\Delta^+(D), \ \Delta^-(D)\}$. If no confusion can arise, we write Δ, Δ^+, and Δ^- instead of $\Delta(D)$, $\Delta^+(D)$, and $\Delta^-(D)$, respectively. The other concepts of digraphs not given here can be found in [2].

For a directed graph $D = (V, A)$, its *line digraph* $L(D)$ has vertex-set $V(L(D)) = A(D)$ such that there is an arc from x to y in $L(D)$ if and only if the head of arc x in D is the tail of arc y in D. A digraph H is a line digraph if there is a digraph D such that $H \cong L(D)$. Błażewicz et al. [4] showed that if a digraph D can be (α, k)-labelled for some integers $\alpha > 0$ and $k > 1$, then D is a line digraph.

A digraph D is a *DNA graph* if and only if there exists an integer $k > 1$ such that D admits a $(4, k)$-labelling. Recently, Pendavingh et al. [8] showed that it is a NP-hard problem to decide whether a given digraph is a DNA graph. If a digraph D can be (α, k)-labelled for some integers $k > 1$ and $\alpha > 0$, then D also can be $(\alpha + 1, k)$-labelled. Let $\alpha_k(D)$ be the smallest integer α such that D can be (α, k)-labelled for fixed integer $k > 1$. Pendavingh et al. [8] also showed that it is NP-hard to decide whether a given digraph has an (α, k)-labelling for any fixed integer $k \geq 3$ and an input parameter α. Hence, it is also NP-hard [3] to determine $\alpha_k(D)$ for any given digraph D and for any fixed integer $k \geq 3$ (this problem is polynomial-solved for $k = 2$ [4]).

An (α, k)-labelled graph can be described by an induced digraph of the directed de Bruijn graph $B(\alpha, k)$. $B(\alpha, k)$ [5] is a directed graph with α^k vertices labelled by the words of length k over a certain alphabet of cardinality α: there is an arc from a vertex v labelled by $(v_1, v_2, \ldots, v_k)$ to a vertex w labelled by $(w_1, w_2, \ldots, w_k)$ if and only if $v_i = w_{i-1}$ for $i = 2, \ldots, k$. The out-degree and in-degree of each vertex are both equal to α.

In this paper, we first introduce a novel labelling of a digraph called quasi-(α, k)-labelling, and establish a relationship between such two labellings and other useful lemmas. In section 3, by using the pancyclicity of directed de Bruijn graphs, we obtain simple formulas to compute $\alpha_k(D)$ for both cycle and path D: $\alpha_k(C_n) = \lceil \sqrt[k-1]{n} \rceil$ and $\alpha_k(P_n) = \lceil \sqrt[k-1]{n+1} \rceil$, where $\lceil x \rceil$ denotes the least integer with no less than number x. Accordingly, both cycles and paths are DNA graphs. In section 4, we show that every out-tree T_s^+ (in-tree T_s^-) can be $(\Delta, k)-$labelled for large k by applying Δ-nary numeral system. Then we obtain that an out-tree T_s^+ (in-tree T_s^-) is a DNA graph if and only if $\Delta \leqslant 4$. Finally, we show that a connected self-adjoint digraph, i.e. a digraph obtained from a unique cycle C by generating simultaneously an out-tree (resp. in-tree) at each vertex, is a DNA graph if and only if $\Delta \leqslant 4$.

2. Quasi-(α, k)-labelling l^*

To study an (α, k)-labelling l of a digraph D, we introduce a novel labelling of D as follows. For a directed graph $D = (V, A)$, let $l^* : V \to \{1, \ldots, \alpha\}^k$, i.e. every vertex v of D is assigned a label $l^*(v) = (l_1^*(v), \ldots, l_k^*(v))$ with every $l_i^*(v) \in \{1, \ldots, \alpha\}$. We call l^* a *quasi-(α, k)-labelling* of D, if

(i) for any two distinct vertices u and v, their labels are different, i.e. $l^*(u) \neq l^*(v)$, and

(ii) if (u, v) is an arc in D, then $l_i^*(u) = l_{i-1}^*(v)$ for $i = 2, \ldots, k$.

For given integers $k > 1$ and $\alpha > 0$, if D has a quasi-(α, k)-labelling, we say D can be *quasi-(α, k)-labelled*. For example, figure 2 shows a digraph D with a quasi-(3, 2)-labelling l, which is indeed not a (3,2)-labelling since $l_2(v_3) = l_1(v_2)$, but (v_3, v_2) is not arc of D.

Notice that if D is an induced subgraph of $B(\alpha, k)$, then D can be (α, k)-labelled; if D is a subgraph of $B(\alpha, k)$, then D can be quasi-(α, k)-labelled. The next lemma gives a relation between such two labellings.

Lemma 2.1. Let D be a digraph. If D is quasi-$(\alpha, k-1)$-labelled, then its line digraph $L(D)$ is (α, k)-labelled.

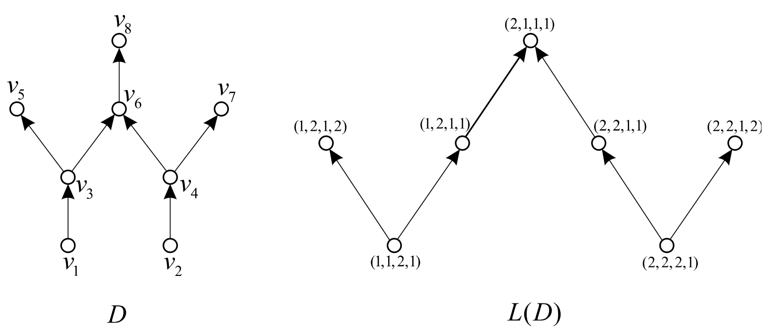

Figure 2. A digraph D with a quasi-(3, 2)-labelling.

Proof. Let l^* be a quasi-$(\alpha, k-1)$-labelling of D. Let v be any vertex of $L(D)$ corresponding to an arc (v_1, v_2) in D. An (α, k)-labelling l of $L(D)$ is defined as

$$\begin{aligned} l(v) &= (\, l_1(v), l_2(v)\,, \ldots, l_{k-1}(v)\,, l_k(v)\,) &(2.1)\\ &= (\, l_1^*(v_1), l_2^*(v_1), \ldots, l_{k-1}^*(v_1), l_{k-1}^*(v_2)\,) &(2.2)\\ &= (\, l_1^*(v_1), l_1^*(v_2), \ldots, l_{k-2}^*(v_2), l_{k-1}^*(v_2)\,). &(2.3) \end{aligned}$$

Clearly, for each i, $l_i(v) \in \{1, 2, \ldots, \alpha\}$. For any two distinct vertices u and v of $L(D)$, corresponding to arcs (u_1, u_2) and (v_1, v_2), respectively, we have $l(u) \neq l(v)$. Otherwise, by equations (2.1)–(2.3), $l(u) = l(v)$ implies that $l^*(u_1) = l^*(v_1)$ and $l^*(u_2) = l^*(v_2)$. Hence $u_1 = v_1$ and $u_2 = v_2$, contradicting $u \neq v$.

Further, if (u, v) is an arc of $L(D)$, then $u_2 = v_1$ in D, and

$$\begin{aligned} l(u) &= (\, l_1^*(u_1), l_1^*(u_2), \ldots, l_{k-2}^*(u_2), l_{k-1}^*(u_2)\,),\\ l(v) &= (\, l_1^*(v_1), l_2^*(v_1), \ldots, l_{k-1}^*(v_1), l_{k-1}^*(v_2)\,). \end{aligned}$$

So we have $(l_2(u), \ldots, l_k(u)) = (l_1(v), \ldots, l_{k-1}(v))$. Conversely, suppose that $(l_2(u), \ldots, l_k(u)) = (l_1(v), \ldots, l_{k-1}(v))$. By equations (2.1) – (2.3) again, we have

$$l^*(u_2) = (l_1^*(u_2), \ldots, l_{k-1}^*(u_2)) = (l_1^*(v_1), \ldots, l_{k-1}^*(v_1)) = l^*(v_1).$$

Since l^* is a quasi-$(\alpha, k-1)$-labelling of D, $u_2 = v_1$. Hence $(u, v) \in A(L(D))$. □

Lemma 2.1 is exemplified in figure 3.

Note that the converse of lemma 2.1 is not true. A counterexample is shown in figure 4. In fact, $L(D)$ can be $(2, 4)$-labelled, but D cannot be quasi-$(2, k)$-labelled for any integer $k > 1$. Suppose to the contrary that D has a quasi-$(2, k)$-labelling l^* for some $k > 1$. Let $l^*(v_6) = (\bar{a}, b)$, where $\bar{a} \in \{1, 2\}^{k-1}$ and $b \in \{1, 2\}$. Then $l^*(v_3) = (a_1, \bar{a})$ and $l^*(v_4) = (a_2, \bar{a})$, where $a_1, a_2 \in \{1, 2\}$ and $a_1 \neq a_2$. Further, $l^*(v_5) = (\bar{a}, b_1)$ and $l^*(v_7) = (\bar{a}, b_2)$. Since b_1, b_2 and

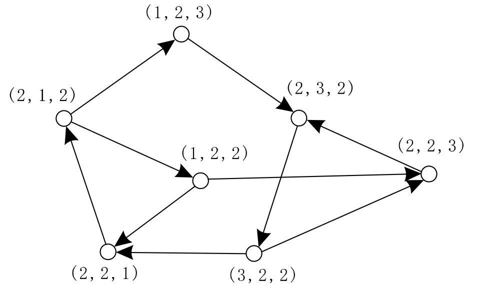

Figure 3. A quasi-(2, 2)-labelling of D is transformed into a (2,3)-labelling of $L(D)$.

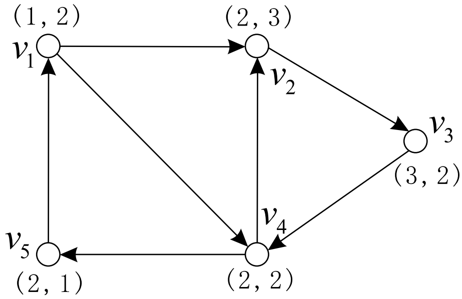

Figure 4. A counterexample to the converse of lemma 2.1.

b belong to $\{1, 2\}$, two of them have the same values. This implies that two in v_5, v_6 and v_7 are assigned the same labels under l^*, a contradiction.

Next, we give some lemmas which will be used repeatedly later in this paper.

Lemma 2.2. If a digraph D is (α, k)-labelled, then $\alpha \geq \Delta$.

Proof. Let v be a vertex of D such that $d^+(v) = \Delta^+$. For any (α, k)-labelling l of D, let $l(v) = (l_1(v), \ldots, l_k(v))$. For every out-neighbour u of v, we have $l(u) = (l_2(v), \ldots, l_k(v), a)$. Since v has Δ^+ out-neighbours and any two distinct out-neighbours have different labels, we use at least Δ^+ words. Hence we have $\alpha \geqslant \Delta^+$. Similarity, we can see that $\alpha \geqslant \Delta^-$. So $\alpha \geqslant \Delta = \max\{\Delta^+, \Delta^-\}$. □

The *converse* of a directed graph D is a new digraph obtained from D by reversing the direction of every arc of D, denoted by D^c. Clearly, $(D^c)^c = D$.

Lemma 2.3. If a digraph D can be (α, k)-labelled, then the converse D^c also can be (α, k)-labelled.

Proof. Let l be an (α, k)-labelling of D. A labelling l' of D^c is defined: for each vertex $v \in V(D^c)$, let $l'_i(v) := l_{k+1-i}(v)$ $(i = 1, \ldots, k)$. We shall verify that l' is an (α, k)-labelling of D^c. For any two distinct vertices u and v of D^c, we have $l'_i(v) \in \{1, \ldots, \alpha\}$ and $l'(u) \neq l'(v)$.

If (u, v) is an arc of D^c, then (v, u) is an arc of D. Hence we have

$$\begin{aligned}(l'_2(u), \ldots, l'_k(u)) &= (l_{k-1}(u), \ldots, l_1(u))\\ &= (l_k(v), \ldots, l_2(v))\\ &= (l'_1(v), \ldots, l'_{k-1}(v)).\end{aligned}$$

Conversely, suppose that $(l_2'(u), \ldots, l_k'(u)) = (l_1'(v), \ldots, l_{k-1}'(v))$. We have

$$\begin{aligned}(l_1(u), \ldots, l_{k-1}(u)) &= (l_k'(u), \ldots, l_2'(u)) \\ &= (l_{k-1}'(v), \ldots, l_1'(v)) \\ &= (l_2(v), \ldots, l_k(v)).\end{aligned}$$

Since l is an (α, k)-labelling of D, (v, u) is an arc of D. Hence (u, v) is an arc of D^c. □

3. Computing $\alpha_k(C_n)$ and $\alpha_k(P_n)$

A cycle C_n is a digraph (V, A):

$$V = \{v_1, v_2, \ldots, v_n\}, \quad \text{and} \quad A = \{(v_1, v_2), (v_2, v_3), \ldots, (v_{n-1}, v_n), (v_n, v_1)\}.$$

A path $P_n = (V, A)$ is a digraph with

$$V = \{v_1, v_2, \ldots, v_n\}, \quad \text{and} \quad A = \{(v_1, v_2), (v_2, v_3), \ldots, (v_{n-1}, v_n)\}.$$

In particular, C_1 is a loop and P_1 is a single vertex. We can see that if a digraph D without loops, then $\alpha_k(D) \geqslant 2$ for any integer $k > 1$. Otherwise, there exists a $(1, k)$-labelling l of D for some integer $k > 1$. Then for any vertex v of D, $l(v) = (1, 1, \ldots, 1)$ and there is a loop at v, a contradiction. Hence $\alpha_k(C_1) = 1$ and $\alpha_k(P_1) = 2$. From now on, we suppose that $n \geqslant 2$.

Lemma 3.1. If l is an $(\alpha, k)-$labelling of C_n or P_n, for any $1 \leqslant i < j \leqslant n$, we have

$$(l_1(v_i), \ldots, l_{k-1}(v_i)) \neq (l_1(v_j), \ldots, l_{k-1}(v_j)).$$

Proof. Suppose to the contrary that there exist two vertices v_i and v_j $(1 \leqslant i < j \leqslant n)$ such that

$$(l_1(v_i), \ldots, l_{k-1}(v_i)) = (l_1(v_j), \ldots, l_{k-1}(v_j)).$$

Considering arc (v_{j-1}, v_j), we have

$$\begin{aligned}(l_2(v_{j-1}), \ldots, l_k(v_{j-1})) &= (l_1(v_j), \ldots, l_{k-1}(v_j)) \\ &= (l_1(v_i), \ldots, l_{k-1}(v_i)).\end{aligned}$$

By the definition of (α, k)-labelling, there exists an arc from v_{j-1} to v_i. This implies $i = j$, a contradiction. □

A directed graph of order n is *pancyclic* if it has cycles of all length $3, 4, \ldots, n$. Every directed de Bruijn graph $B(\alpha, k)$ is pancyclic (cf. Refs. [7] and [2, pp. 308]).

Theorem 3.2. $\alpha_k(C_n) = \lceil \sqrt[k-1]{n} \rceil$.

Proof. We first show that C_n can be $(\lceil \sqrt[k-1]{n} \rceil, k)$-labelled. Let $\alpha := \lceil \sqrt[k-1]{n} \rceil$. Then $n \leqslant \alpha^{k-1}$. By the pancyclicity of directed de Bruijn graph $B(\alpha, k-1)$, C_n is a subgraph of $B(\alpha, k-1)$. Hence, the $(\alpha, k-1)$-labelling of $B(\alpha, k-1)$ induces a quasi-$(\alpha, k-1)$-labelling of C_n. By lemma 2.1, $C_n \cong L(C_n)$ can be (α, k)-labelled. Hence $\alpha_k(C_n) \leqslant \lceil \sqrt[k-1]{n} \rceil$. It remains to show that $\alpha_k(C_n) \geqslant \alpha$. Let $\beta := \alpha_k(C_n)$. If $n > \beta^{k-1}$, for any (β, k)-labelling l of C_n, there exist two vertices v_i and v_j, such that

$$(l_1(v_i), \ldots, l_{k-1}(v_i)) = (l_1(v_j), \ldots, l_{k-1}(v_j)),$$

which contradicts lemma 3.1. So $n \leqslant \beta^{k-1}$, i.e. $\alpha_k(C_n) \geqslant \lceil \sqrt[k-1]{n} \rceil$. □

Corollary 3.3. Any cycle C_n is a DNA graph.

Proof. For $k \geqslant \lceil \log_4 n \rceil + 1$, $\alpha_k(C_n) \leqslant 4$. Accordingly, C_n can be $(4, k)$-labelled. □

In the remainder of this section, we compute the $\alpha_k(P_n)$. Since $P_n = C_{n+1} - v_{n+1}$, any (α, k)-labelling of C_{n+1} gives an (α, k)-labelling of P_n. Hence, theorem 3.2 implies the following result.

Lemma 3.4.

$$\alpha_k(P_n) \leqslant \alpha_k(C_{n+1}) = \lceil \sqrt[k-1]{n+1} \rceil.$$

Theorem 3.5.

$$\alpha_k(P_n) = \lceil \sqrt[k-1]{n+1} \rceil.$$

Proof. By lemma 3.1, similar to the proof of theorem 3.2, we have $\alpha_k(P_n) \geqslant \lceil \sqrt[k-1]{n} \rceil$. Let $\alpha := \lceil \sqrt[k-1]{n} \rceil$. Then $(\alpha - 1)^{k-1} < n \leqslant \alpha^{k-1}$. If $n < \alpha^{k-1}$, then $(\alpha - 1)^{k-1} < n + 1 \leqslant \alpha^{k-1}$. By lemma 3.4, we have

$$\alpha_k(P_n) \leqslant \lceil \sqrt[k-1]{n+1} \rceil = \alpha = \lceil \sqrt[k-1]{n} \rceil \leqslant \alpha_k(P_n). \tag{3.1}$$

So the all equalities in (3.1) hold.

If $n = \alpha^{k-1}$, then $\alpha + 1 \geqslant \sqrt[k-1]{n+1} > \alpha$. Hence $\lceil \sqrt[k-1]{n+1} \rceil = \alpha + 1$. If P_n admits an $(\alpha, k)-$labelling, by lemma 3.1, there are exactly $n = \alpha^{k-1}$ pairwise different sequences $(l_1(v_i), \ldots, l_{k-1}(v_i)), i = 1, \ldots, n$. Hence there exists v_i $(1 \leqslant i \leqslant n)$ such that

$$(l_1(v_i), \ldots, l_{k-1}(v_i)) = (l_2(v_n), \ldots, l_k(v_n)).$$

This implies that P_n has an arc from v_n to v_i, a contradiction. So

$$\alpha_k(P_n) \geqslant \alpha + 1 = \lceil \sqrt[k-1]{n+1} \rceil. \tag{3.2}$$

Hence the theorem follows from (3.2) and lemma 3.4. □

For $k \geqslant \lceil \log_4(n+1) \rceil + 1$, $\alpha_k(P_n) \leqslant 4$. So we arrive in the following corollary.

Corollary 3.6. Any path is a DNA graph.

4. Rooted trees

A digraph T is an *out-tree* (*in-tree*) if T is an oriented tree and T has only one vertex s of in-degree (out-degree) zero. Then s is called the root of T. Let T_s^+ and T_s^- denote out-trees and in-trees rooted at s, respectively. Since the converse of an in-tree is an out-tree and vice versa, by lemma 2.3, a rooted tree can be (α, k)-labelled if and only if its converse can be (α, k)-labelled. So in the section, we only consider out-trees.

For any non-trivial out-tree T_s^+, $\Delta^+ \geqslant \Delta^- = 1$. Hence $\Delta = \Delta^+$. For any vertex v of T_s^+, the *layer number* of v is defined as the distance from s, and the *height* of T_s^+ is defined as the maximum value of layer numbers of all vertices of T_s^+. Let $T_s^+(\Delta, p)$ denote an out-tree rooted at s with height p and out-degree Δ for any vertex of layer number less than p. Clearly, if T_s^+ has height p and maximum out-degree Δ, then T_s^+ is an induced subgraph of $T_s^+(\Delta, p)$ (figure 5). Since $T_s^+(1, p)$ is a path, we assume that $\Delta \geqslant 2$ in the following.

If $T_s^+(\Delta, p)$ can be (α, k)-labelled, then its all induced subgraphs also can be (α, k)-labelled for the same α and k. Clearly, if $T_s^+(\Delta, p)$ can be (α, k)-labelled, by lemma 2.2, we have $\alpha \geqslant \Delta$. The next thoerem shows that $T_s^+(\Delta, p)$ can be $(\Delta, p+2)$-labelled.

Theorem 4.1. $T_s^+(\Delta, p)$ can be $(\Delta, p+2)$-labelled for $\Delta \geqslant 2$.

Proof. Let $T'(\Delta, p)$ be another out-tree obtained from $T_s^+(\Delta, p)$ by adding a new vertex t together with an arc from t to s. For example, $T'(3, 3)$ is shown in figure 6. Clearly, $L(T'(\Delta, p)) \cong T_s^+(\Delta, p)$. By lemma 2.1, it is sufficient to give a quasi-$(\Delta, p+1)$-labelling of $T'(\Delta, p)$.

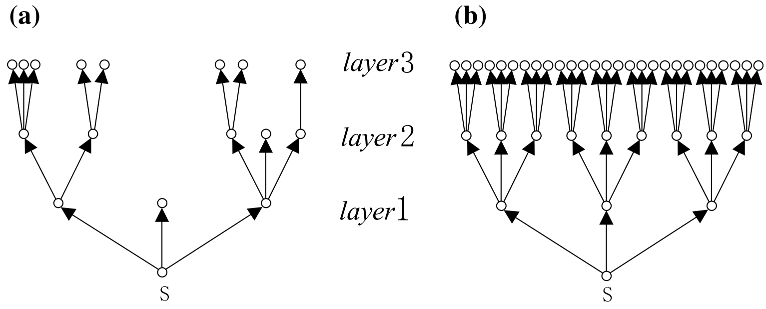

Figure 5. (a) An out-tree T_s^+ with $p = 3$ and $\Delta = 3$, and (b) a complete out-tree $T_s^+(3, 3)$.

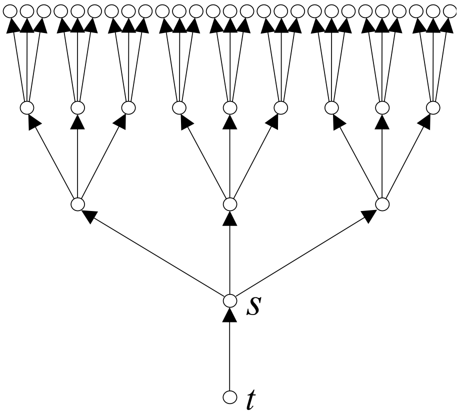

Figure 6. $T'(3, 3)$.

For convenience, we replace alphabet $\{1, \ldots, \Delta\}$ by $\{0, \ldots, \Delta-1\}$. The vertices of $T'(\Delta, p)$ are marked by pairs of non-negative integers in accordance with the following rules:

(i) t and s are denoted by $v_{0,0}$ and $v_{0,1}$, respectively, and

(ii) for a vertex already marked by $v_{i,t}$ with $(i, t) \neq (0, 0)$ and $i<p$, all the out-neighbours of $v_{i,t}$ are denoted by $v_{i+1,(t-1)\Delta+1}$, $v_{i+1,(t-1)\Delta+2}$, $\ldots$, $v_{i+1,t\Delta}$, respectively.

Clearly, different vertices have different marks. We claim that

$$1 \leqslant n \leqslant \Delta^i, \text{ for each vertex } v := v_{i,n} \text{ with } 1 \leqslant i \leqslant p. \tag{4.1}$$

To prove this claim, we use induction on i. For $i = 1$, since v is an out-neighbour of $s := v_{0,1}$, $1 \leqslant n \leqslant \Delta$. Now let $i \geqslant 2$. Suppose that the assertion is true for smaller i. Let u be the unique in-neighbour of v, marked by $v_{i-1,t}$.

By the marked rules,

$$(t-1)\Delta + 1 \leqslant n \leqslant t\Delta. \tag{4.2}$$

By the induction hypothesis, we have

$$1 \leqslant t \leqslant \Delta^{i-1}. \tag{4.3}$$

Hence claim (4.1) follows from (4.2) and (4.3).

Now we define an integer-valued function f on the vertex-set of $T'(\Delta, p)$ as follows:

$$f(v_{i,n}) = \Delta^i + n - 1 \tag{4.4}$$

for each vertex $v_{i,n}$ of $T'(\Delta, p)$. By (4.1) and (4.4), we have

$$f(v_{i,n}) \leqslant \Delta^p + n - 1 \leqslant \Delta^p + \Delta^p - 1 \leqslant \Delta^{p+1} - 1. \tag{4.5}$$

For any two distinct vertices $u := v_{i_1,n_1}$ and $v := v_{i_2,n_2}$, we have

$$f(u) \neq f(v). \tag{4.6}$$

If $i_1 = i_2$, then $n_1 \neq n_2$ and

$$f(u) = \Delta^{i_1} + n_1 - 1 \neq \Delta^{i_2} + n_2 - 1 = f(v).$$

Otherwise, suppose that $i_2 > i_1$. By (4.1) and (4.4), we have

$$f(u) = \Delta^{i_1} + n_1 - 1 \leqslant \Delta^{i_1} + \Delta^{i_1} - 1 = 2\Delta^{i_1} - 1, \text{ and}$$
$$f(v) = \Delta^{i_2} + n_2 - 1 \geqslant \Delta^{i_2} = \Delta^{i_2 - i_1} \cdot \Delta^{i_1} \geqslant 2\Delta^{i_1}.$$

So (4.6) follows.

Let $l^*(v_{i,n})$ be the representation of $f(v_{i,n})$ by Δ-nary numeral system with $(p+1)$-digit. More precisely, let

$$f(v_{i,n}) = a_p\Delta^p + a_{p-1}\Delta^{p-1} + \cdots + a_1\Delta + a_0,$$

where $0 \leqslant a_j \leqslant \Delta - 1$ is an integer for each $0 \leqslant j \leqslant p$. Then $l^*(v_{i,n}) = (a_p, a_{p-1}, \ldots, a_1, a_0)$. In the following, we show that l^* is a quasi-$(\alpha, p+1)$-labelling of $T'(\Delta, p)$.

By (4.5) and (4.6), we can see that it is sufficient to show if (v_1, v_2) is an arc of $T'(\Delta, p)$, then $(l_2^*(v_1), \ldots, l_{p+1}^*(v_1)) = (l_1^*(v_2), \ldots, l_p^*(v_2))$. If $v_1 = t$ and $v_2 = s$, $f(v_1) = 0$ and $f(v_2) = 1$. Then $l^*(u) = (0, \ldots, 0, 0)$ and $l^*(v) = (0, \ldots, 0, 1)$.

So suppose that $v_1 = v_{i,n_1} \neq t$. As (v_1, v_2) is an arc, we have $v_2 = v_{i+1,n_2}$, $n_2 = (n_1 - 1)\Delta + m$ $(1 \leqslant m \leqslant \Delta)$ and

$$\begin{aligned} f(v_2) &= \Delta^{i_2} + n_2 - 1 \\ &= \Delta^{i_1+1} + (n_1 - 1)\Delta + m - 1 \\ &= (\Delta^{i_1} + n_1 - 1)\Delta + m - 1 \\ &= f(v_1)\Delta + m - 1. \end{aligned}$$

If $f(v_1) = a_p\Delta^p + a_{p-1}\Delta^{p-1} + \cdots + a_1\Delta + a_0$, $0 \leqslant a_i < \Delta, i = 0, \ldots, p$, then $f(v_2) = a_{p-1}\Delta^p + \cdots + a_0\Delta + m - 1$. This implies that $l^*(v_1) = (0, a_{p-1}, \ldots, a_1, a_0)$ and $l^*(v_2) = (a_{p-1}, \ldots, a_1, a_0, m - 1)$. Accordingly,

$$(l_2^*(v_1), \ldots, l_{p+1}^*(v_1)) = (l_1^*(v_2), \ldots, l_p^*(v_2)).$$

As above, l^* is a quasi-$(\Delta, p + 1)$-labelling of $T'(\Delta, p)$. □

From the above proof, we can see that for every $k \geqslant p + 1$, if $l^*(v)$ is denoted by the representation of $f(v)$ by Δ-nary numeral system with k-digit, l^* will be a quasi-(α, k)-labelling of $T'(\Delta, p)$. Hence $T_s^+(\Delta, p)$ can be (Δ, k)-labelled for $k \geqslant p + 2$.

Corollary 4.2. If T_s^+ is an out-tree with height p and maximum out-degree Δ, then $\alpha_k(T_s^+) = \Delta$ for any integer $k \geqslant p + 2$.

Proof. Let k be any integer such that $k \geqslant p + 2$. By lemma 2.2, we have $\alpha_k(T_s^+) \geqslant \Delta$. Since T_s^+ is an induced subgraph of $T_s^+(\Delta, p)$, any (Δ, k)-labelling of $T_s^+(\Delta, p)$ induces a (Δ, k)-labelling of T_s^+. By theorem 4.1, we have that $T_s^+(\Delta, p)$ has a (Δ, k)-labelling. So $\alpha_k(T_s^+) \leqslant \Delta$. Hence $\alpha_k(T_s^+) = \Delta$. □

Corollary 4.3. An out-tree T_s^+ is a DNA graph if and only if $\Delta(T_s^+) \leqslant 4$.

Proof. Let T_s^+ be an out-tree with height p and maximum out-degree Δ. If $\Delta \leqslant 4$, by theorem 4.1, T_s^+ can be $(\Delta, p + 2)-$labelled, and T_s^+ is thus a DNA graph. Conversely, if T_s^+ is a DNA graph, by lemma 2.2, we have $\Delta \leqslant 4$. □

5. Self-adjoint digraphs

A digraph D is *self-adjoint* if D is isomorphic to its line digraph $L(D)$. Let $\mathcal{A}$ be the set of all digraphs A, for which there exists a digraph sequence $A_0, \ldots, A_m = A$ satisfying $A_0 = C_n$ and each A_{i+1} arises from A_i by adding some new vertices $v_1, \ldots, v_t$ and arcs $(v, v_1), \ldots, (v, v_t)$ where v is a vertex of A_i. Define $\mathcal{A}^c = \{A^c | A \in \mathcal{A}\}$. We can see that a digraph $A \in \mathcal{A}$ $(\mathcal{A}^c)$ if and only if A is the union of a cycle C_n and n pairwise disjoint out-trees (in-trees)

$T_1, \ldots, T_n$ such that each T_i has the root v_i lying in C_n. Hao [6] showed that a connected digraph A is self-adjoint if and only if $A \in \mathcal{A}$ or $\mathcal{A}^c$.

By lemma 2.3, if A can be (α, k)-labelled, then A^c also can be (α, k)-labelled. So we only consider $\mathcal{A}$ in the following. For every digraph $A \in \mathcal{A}$, it is easy to see $d^-(v) = 1$ for every $v \in V(A)$. Recall that $\Delta = \Delta^+$ is the maximum out-degree of A and n the length of the unique cycle in A. Let $p := \max(p_1, \ldots, p_n)$, where p_i is the height of T_i for each $1 \leqslant i \leqslant n$.

We define a sign system as $v_i^t(a_1, \ldots, a_t)$ $(1 \leqslant i \leqslant n, 0 \leqslant a_j \leqslant \Delta - 1, 0 \leqslant t \leqslant p)$ to represent the vertices of A. First, we denote the vertices of the unique cycle C_n by $v_1^0, \ldots, v_n^0$ in a way. Next let $u_1, \ldots, u_k$ be the out-neighbours of v_i^0 in T_i and mark u_j by $v_i^1(j)$, $j = 1, 2, \ldots, k$. Then for every already marked vertex $v_i^t(a_1, \ldots, a_t)$, $t \geqslant 1$, its all out-neighbours $w_0, \ldots, w_{k'}$ are marked by $v_i^{t+1}(a_1, \ldots, a_t, a_{t+1})$, $a_{t+1} = 0, 1, \ldots, k'$. For example, figure 7 gives such a sign system of vertices of a self-adjoint graph A_1 with $\Delta = 2, n = 3$ and $p = 3$. Since A is a cycle for $\Delta = 1$, assume that $\Delta \geqslant 2$ in the following.

Theorem 5.1. Let $A \in \mathcal{A}$, n, p, $\Delta \geqslant 2$ be defined as above. Then A can be $(\Delta, k+1)$-labelled, where $k = (\lceil \frac{p}{n} \rceil + 1)n$.

Proof. We first construct a self-adjoint digraph $A' \in \mathcal{A}$ from A by adding some new vertices and arcs: every T_i of A' is rooted at v_i^0 with height p, $d_{T_i}^+(v_i^0) = \Delta - 1$ and any other vertex has out-degree Δ if its layer number is less than p. So both A and A' have the same height, and A is an induced subgraph of A'. By lemma 2.1, if A' can be quasi-(Δ, k)-labelled, then A' can be $(\Delta, k+1)$-labelled and A can be $(\Delta, k+1)$-labelled. So it is sufficient to give a quasi-(Δ, k)-labelling of A' in the following.

For convenience, we replace alphabet $\{1, \ldots, \Delta\}$ with $\{0, \ldots, \Delta - 1\}$. Let $k_1 := \lceil \frac{p}{n} \rceil$, $k := (k_1 + 1)n$ and $r = k_1 n - p$. Then $0 \leqslant r < n$ and $p = k_1 n - r$. First, we label the vertices of the cycle C_n as follows. Let $l(v_i^0) := (\underbrace{S_{n+1-i}, \ldots, S_{n+1-i}}_{k_1+1})$, where $S_i = (0, \ldots, 0, 1, 0, \ldots, 0)$ is a sequence of length n such that the ith position is 1 and the others are 0, for $i = 1, \ldots, n$. Obviously, this is a quasi-$(2, k)$-labelling of C_n. Hence for any two integers $i < j$, we have that

$$l(v_j^0) = (l_{1+j-i}(v_i^0), l_{2+j-i}(v_i^0), \ldots, l_k(v_i^0), l_1(v_i^0), \ldots, l_{j-i}(v_i^0)). \qquad (5.1)$$

Next we label any other vertex v of A. If $v = v_i^1(j)$, then v is an out-neighbour of v_i^0. Let $l(v_i^1(j)) := (l_2(v_i^0), \ldots, l_k(v_i^0), a)$, where $a = 1 - l_1(v_i^0)$, if $j = 1$; $a = j$, otherwise. If $v = v_i^t(a_1, \ldots, a_t)$ for $2 \leqslant t \leqslant p$, then v is an out-neighbour of $u = v_i^{t-1}(a_1, \ldots, a_{t-1})$. Let $l(v_i^t(a_1, \ldots, a_t)) := (l_2(u), \ldots, l_k(u), a_t)$.

We claim that

$$l_j(v_i^t(a_1,\dots,a_t)) = \begin{cases} l_{j+t}(v_i^0), & j \leqslant k-t, \\ 1-l_1(v_i^0), & j=k-t+1 \text{ and } a_1=1, \\ a_{j+t-k}, & \text{otherwise.} \end{cases} \tag{5.2}$$

To prove this claim, for each fixed i we proceed by induction on t. For $t=0$ and 1, it is trivially true. So let $t \geqslant 2$ and suppose that the claim is true for smaller t. Let $v := v_i^t(a_1,\dots,a_t)$ and let u be the unique in-neighbour of v. Then $u = v_i^{t-1}(a_1,\dots,a_{t-1})$.

For $j=1,2,\dots,k-1$, $l_j(v) = l_{j+1}(u)$. Further, if $j \leqslant k-t$, then $j+1 \leqslant k-(t-1)$ and

$$l_j(v) = l_{j+1}(u) = l_{j+1+(t-1)}(v_i^0) = l_{j+t}(v_i^0)$$

by the induction hypothesis; if $j=k-t+1$ and $a_1=1$, then $j+1=k-(t-1)+1$ and

$$l_j(v) = l_{j+1}(u) = 1-l(v_i^0).$$

Otherwise, if $j \leqslant k-1$,

$$l_j(v) = l_{j+1}(u) = a_{j+1+(t-1)-k} = a_{j+t-k}.$$

If $j=k$, $l_j(v) = a_t$ from the above labelling method. So claim (5.2) follows.

Finally, we show that this labelling l is a quasi-(Δ,k)-labelling of A'. It is sufficient to verify that distinct vertices of A' have different labels. Suppose not, there exist two distinct vertices u and v of A' such that $l(u) = l(v)$. There are two cases to be considered.

Case 1. u, v lie in the same tree T_i.

Let $u = v_i^t(b_1,\dots,b_t)$ and $v = v_i^q(c_1,\dots,c_q)$. If $t=q$, then there exists $j \in \{1,2,\dots,t\}$ such that $b_j \neq c_j$ since $u \neq v$. Hence, we have $l_{j+k-t}(u) \neq l_{j+k-t}(v)$ by equation (5.2). So suppose $q>t$. By simple computation and comparison we obtain $n \leqslant k-q \leqslant k-t$. Since $(l_1(u),\dots,l_n(u)) = (l_1(v),\dots,l_n(v))$, by equation (5.2) we can see that

$$(l_{1+t}(v_i^0),\dots,l_{n+t}(v_i^0)) = (l_{1+q}(v_i^0),\dots,l_{n+q}(v_i^0)), \tag{5.3}$$

which is equivalent to

$$(l_1(v_{i+t}^0),\dots,l_n(v_{i+t}^0)) = (l_1(v_{i+q}^0),\dots,l_n(v_{i+q}^0))$$

from equation (5.1); that is, $S_{n+1-(t+i)} = S_{n+1-(q+i)}$. Then we have $q \equiv t \pmod n$. Since $q>t$, there exists an integer $g \geqslant 1$ such that $q = t+gn$. But by equation (5.2), we have

$$l_{k-q+1}(v) = 1-l_1(v_i^0) \text{ or } c_1 \ (c_1 \geqslant 2), \text{ and}$$
$$l_{k-q+1}(u) = l_{k-t-gn+1}(u) = l_{k-gn+1}(v_i^0) = l_{(k_1-g+1)n+1}(v_i^0) = l_1(v_i^0).$$

Hence $l_{k-q+1}(u) \neq l_{k-q+1}(v)$, a contradiction.
Case 2. u, v lie in different trees T_i and $T_j (i \neq j)$.

Let $u = v_i^t(b_1, \ldots, b_t)$ and $v = v_j^q(c_1, \ldots, c_q)$. Without loss of generality, suppose that $q \geqslant t$ and $i = 1$. Considering the first n digits of $l(u)$ and $l(v)$, by equation (5.1), we have

$$(l_1(u), \ldots, l_n(u)) = (l_{1+t}(v_1^0), \ldots, l_{n+t}(v_1^0)) = (l_1(v_{1+t}^0), \ldots, l_n(v_{1+t}^0)) = S_{n+1-(1+t)},$$
$$(l_1(v), \ldots, l_n(v)) = (l_{1+q}(v_j^0), \ldots, l_{n+q}(v_j^0)) = (l_1(v_{j+q}^0), \ldots, l_n(v_{j+q}^0)) = S_{n+1-(j+q)}.$$

Hence, $1 + t \equiv q + j \pmod{n}$. Since $q \geqslant t$, there exists an integer $g \geqslant 1$ such that $q + j = 1 + t + gn$. By equation (5.2), we have

$$l_{k-q+1}(v) = 1 - l_1(v_j^0) \text{ or } c_1 \quad (c_1 \geqslant 2),$$
$$l_{k-q+1}(u) = l_{k-t-gn+j}(u) = l_{k-gn+j}(v_1^0) = l_{(k_1-g+1)n+j}(v_1^0) = l_j(v_1^0) = l_1(v_j^0).$$

So $l_{k-q+1}(u) \neq l_{k-q+1}(v)$, a contradiction.

By cases 1 and 2, we can see that all vertices of A' have pairwise different labels. Accordingly, l is a quasi-(Δ, k)-labelling of A'. □

As an example we now give a (2, 7)-labelling of a self-disjoint digraph A_1 from the method used in the proof of theorem 5.1. The sign system of vertices of A_1 were previously given in figure 7. We first construct another self-adjoint graph A_1' as shown in figure 8(a) such that A_1 is an an induced subgraph of A_1'. Then, a quasi-(2,6)-labelling l of A_1' is constructed in figure 8(a). Finally, l can

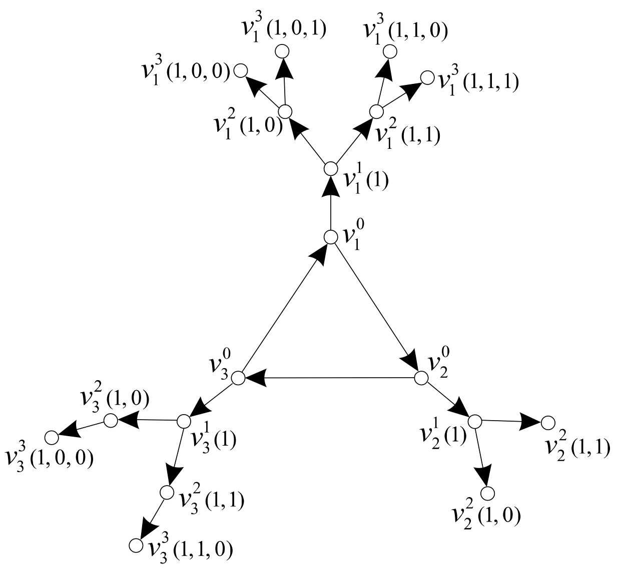

Figure 7. A self-adjoint graph A_1 with $\Delta = 2, n = 3, p = 3$ and a sign system of vertices.

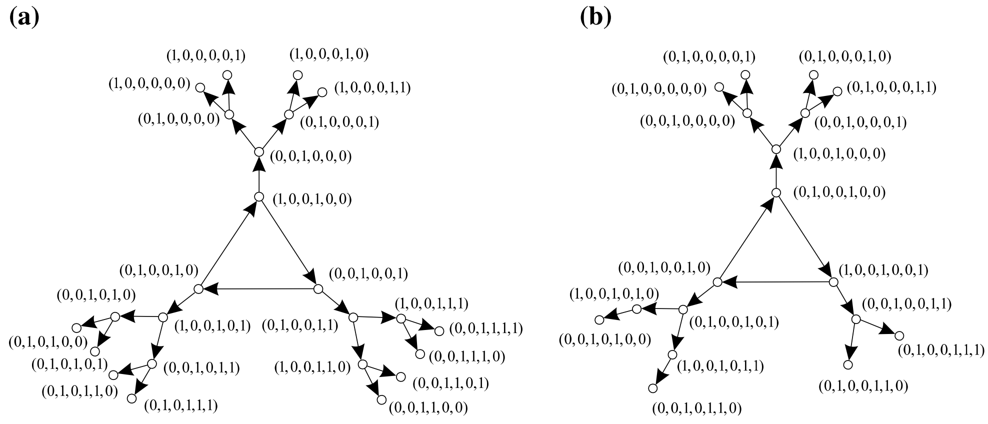

Figure 8. (a) A quasi-(2,6)-labelling of A_1' with $\Delta = 2$, $p = 3$, and (b) a (2,7)-labelling of A_1.

be transformed into a (2,7)-labelling of A_1' by lemma 2.1, which induces a (2,7)-labelling of A_1 as shown in figure 8(b).

Combining theorem 5.1 with lemmas 2.2 and 2.3, we can obtain the following result.

Corollary 5.2. A connected self-adjoint digraph A is a DNA graph if and only if $\Delta \leqslant 4$.

Acknowledgment

Research supported by NSFC and TRAPOYT.

References

[1] W. Bains and G.C. Smith, A novel method for nucleic acid sequence determination, J. Theoret. Biol. 135 (1988) 303–307.
[2] J. Bang-Jensen and G. Gutin, *Digraphs: Theory, Algorithms and Applications*, (Springer-Verlag, London, 2001).
[3] J. Błażewicz, P. Formanowicz, M. Kasprzak and D. Kobler, On the recognition of de Bruijn graphs and their induced subgraphs, Discrete Math. 245 (2002) 81–92.
[4] J. Błażewicz, A. Hertz, D. Kobler and D. de Werra, On some properties of DNA graphs, Discrete Appl. Math. 98 (1999) 1–19.
[5] N.G. de Bruijn, A combinatorial problem, Nederl. Akad. Wetensch. 49 (1946) 758–764 (Indagationes Math. 8 (1946) 461–467).
[6] J. Hao, The adjoints of DNA graphs, J. Math. Chem. 37 (2005) 333–346.
[7] M. Imori, M. Matsumoto and H. Yamada, The line digraph of a regular and pancircular digraph is also regular and pancircular, Graphs Combin. 4 (1988) 235–239.
[8] R. Pendavingh, P. Schuurman and G.J. Woeginger, Recognizing DNA graphs is difficult, Discrete Appl. Math. 127 (2003) 85–94.