

ON UNIVERSAL PARTIAL WORDS FOR WORD-PATTERNS AND SET PARTITIONS

HERMAN Z.Q. CHEN^{1,2,*} AND SERGEY KITAEV^{1,2}

Abstract. Universal words are words containing exactly once each element from a given set of combinatorial structures admitting encoding by words. Universal partial words (u-p-words) contain, in addition to the letters from the alphabet in question, any number of occurrences of a special “joker” symbol. We initiate the study of u-p-words for word-patterns (essentially, surjective functions) and (2-)set partitions by proving a number of existence/non-existence results and thus extending the results in the literature on u-p-words and u-p-cycles for words and permutations. We apply methods of graph theory and combinatorics on words to obtain our results.

Mathematics Subject Classification. 68R15.

Received February 28, 2019. Accepted March 10, 2020.

1. INTRODUCTION

Chung *et al.* [3] introduced the notion of a *universal cycle*, or *u-cycle*, for any class of combinatorial objects that admits encoding by words. A u-cycle is a cyclic word containing each encoded object of fixed size exactly once as a factor (*i.e.* as an interval of consecutive letters). In particular, universal cycles for sets of words are nothing else but the celebrated *de Bruijn sequences* that have found widespread use in real-world applications (see the references in [6]). Examples of objects considered in [3] are *permutations* (that required a slight modification of the notion of a u-cycle) and *set partitions* (that admit encoding in terms of words). U-cycles for *word-patterns* were considered in [1].

A typical approach for constructing u-cycles is trying to find a *Hamiltonian cycle* (*i.e.* a cyclic path going through each vertex in a graph exactly once) in certain directed graphs associated with the objects in question. This, in turn, is often reduced to finding an *Eulerian cycle* (*i.e.* a cyclic walk going through each edge in a graph exactly once) in certain directed graphs in question. It is well known and not difficult to see that proving that an Eulerian cycle in a directed graph G exists is equivalent to showing that G is *balanced* (*i.e.* each vertex has the same in- and out-degrees) and *strongly connected* (*i.e.* there is a directed path from any vertex in the graph to any other vertex). In fact, in a balanced directed graph G without isolated vertices, the condition to be strongly connected is equivalent to the condition that the graph obtained from G by removing orientations is *connected*, that is, that there is a path from any vertex to any other vertex in such

Keywords and phrases: universal word, partial word, set partition, word-pattern, De Bruijn graph, Eulerian path, Hamiltonian path.

¹ School of Statistics and Data Science, Nankai University, PR China.

² Department of Mathematics and Statistics, University of Strathclyde, UK.

* Corresponding author: zqchern@nankai.edu.cn

an undirected graph, which is a useful property. This property is less known, and it is not so easy to find its proof in the literature. Taking into account that the property is crucial for this paper, we sketch a proof of it in Section 1.2.

The non-cyclic version of a u-cycle is known as a *universal word*, or *u-word*. Construction of u-words is often reduced to proving that certain graphs either have a *Hamiltonian path* (defined similarly to a Hamiltonian cycle where the requirement to begin and end at the same vertex is dropped), or an *Eulerian path* (defined as an Eulerian cycle, but dropping the requirement to begin and end at the same vertex). An Eulerian path in a directed graph G exists if and only if (i) each vertex of G has the same in- and out-degrees, except for two vertices, one of which has an extra edge coming in, and the other one has an extra edge coming out, and (ii) the graph obtained from G by removing orientations is connected. In this paper, we will use this characterization to prove the existence of an Eulerian path.

It is worth noting that while constructing u-cycles for set partitions in [3], the authors met an obstacle with the “lifting” an Eulerian cycle in question to a u-cycle, because of the requirement that the beginning of such a u-cycle must be compatible with the end of the u-cycle. In particular, in the case of $n = 3$ the obstacle was non-avoidable, resulting in non-existence of u-cycles for set partitions in this case. We also refer to [5] for recent work on u-cycles for set partitions. In the context of u-words though, which are the subject of interest in this paper, there are no issues related to the “lifting”, as we do not need to worry about compatibility of the beginning and the end of a word in question. Thus, each step in an Eulerian path will result in a straightforward unique prolongation to the right by the minimum available letter of a word to be constructed.

U-cycles and u-words provide an optimal encoding of a set of combinatorial objects in the sense that such an encoding is shortest possible. However, as is discussed in [2] for the case of *de Bruijn sequences*, one can still shorten u-cycles and u-words by using non-deterministic symbols and obtaining, respectively, *u-p-cycles* and *u-p-words* standing for *universal partial cycles* and *universal partial words*. The studies in [2], mainly related to binary alphabets, were extended in [4] to the case of non-binary alphabets. In [6] different ways to shorten u-cycles and u-words for permutations were considered.

This paper deals with natural steps to further develop the theory of u-cycles/u-words and u-p-cycles/u-p-words. We prove a number of existence/non-existence results on u-p-words for word-patterns (essentially, surjective functions) and (2-)set partitions obtained in two different ways depending on the notion of the non-deterministic symbol. In one of the cases such a symbol is context-dependant, in the other case it is context-free.

1.1. Basic definitions and examples

For an alphabet A , we let A^n denote the set of all words of length n over A . A *factor* in a word $w = w_1w_2 \cdots w_n$ is a subword of the form $w_iw_{i+1} \cdots w_j$ for $1 \leq i \leq j \leq n$. We say that a word w *covers* a word u if u occurs in w as a factor. We use $*^i$ to denote a factor of length i with arbitrary letters in A . For a letter x , we let x^n denote the word $x \cdots x$, where x is repeated n times. Assuming an ordered alphabet, the *complement* of a word w is obtained by replacing each i th smallest letter of w by the i th largest letter of w . For example, 31446 is the complement of 46331.

A *word-pattern* is a word containing *each* of the letters in $\{0, 1, \dots, \alpha - 1\}$ for some integer $\alpha \geq 1$ at least once. Thus, a word-pattern of length $k \geq \alpha \geq 1$ is essentially a surjective function from the set $\{1, 2, \dots, k\}$ (representing positions) to the set $\{0, 1, \dots, \alpha - 1\}$. For example, 1130231 is a word-pattern, while 20041421 is not (a 3 is missing). A *set partition* is a grouping of the set’s elements into non-empty subsets, in such a way that every element is included in one and only one of the subsets. Note that any word $w_1w_2 \cdots w_n$ encodes a set partition of $\{1, 2, \dots, n\}$. Indeed, we can let i and j be in the same subset if and only if $w_i = w_j$. Of course, such an encoding is not injective as, for example, 12121 and 21212 encode the same set partition $\{\{1, 3, 5\}, \{2, 4\}\}$. One can define a *canonical encoding* of a set partition to be a word over $\{1, 2, \dots, m\}$ with each letter occurring at least once, such that the leftmost occurrence of a letter i is to the left of the leftmost occurrence of the letter $i + 1$ for $1 \leq i \leq m - 1$. We let $\text{cf}(w_1w_2 \cdots w_n)$ denote the canonical encoding of the set partition encoded by a word $w_1w_2 \cdots w_n$.

We use different alphabets to denote word-patterns and encodings of set partitions by two reasons. First of all, having binary word-patterns over $\{0, 1\}$ we can use verbatim the results in [2], while having encodings of *binary set partitions* (i.e. set partitions with at most two parts) and *2-set partitions* (i.e. set partitions with exactly two parts) over $\{1, 2\}$, we are consistent with [3]. Also, we want word-patterns to look differently from encodings of set partitions to avoid any confusion.

Let A be an alphabet. A *partial word*, or *p-word*, is a word over $A \cup \diamond$, where $\diamond \notin A$. Normally, in the context of universal partial cycles or words, every occurrence of \diamond can be substituted by any letter from A , so we can think of \diamond as a ‘joker’ or ‘wildcard’ symbol. Formally, we define $A_\diamond := A \cup \{\diamond\}$ and we say that a word $v = v_1 v_2 \cdots v_n \in A^n$ appears as a *factor* in a word $u = u_1 u_2 \cdots u_m \in A_\diamond^m$ if there is an integer i such that $u_{i+j} = \diamond$ or $u_{i+j} = v_j$ for all $j = 1, 2, \dots, n$. In the cyclic setting we consider the indices of u in this definition modulo m . For example, for $A = \{0, 1, 2\}$, the word $v = 120$ occurs twice as a factor in the p-word $u = 120\diamond 200$ because of the factors 120 and $\diamond 20$, whereas v does not appear as a factor in $u' = 11\diamond 210\diamond$. Such \diamond s are *context-free* in the sense that they only depend on the alphabet A . However, in this paper, we also deal with *context-dependent* \diamond s, where only certain letters, depending on the neighbouring letters, can be used instead of a \diamond (details of this are given below).

The notions of u-words and u-cycles extend straightforwardly to u-p-words and u-p-cycles, respectively. The simplest example for a u-p-word for A^n is $\diamond^n := \diamond \diamond \cdots \diamond$, the word consisting of n \diamond s, which is called *trivial*. A few more examples/non-examples of u-p-words over the binary alphabet $A = \{0, 1\}$ are as follows. We have that $\diamond \diamond 0111$ is a u-p-word for A^3 , whereas $\diamond \diamond 01110$ is *not* a u-p-word for A^3 , because replacing the first two letters $\diamond \diamond$ by 11 yields the same factor 110 as the last three letters. Similarly, $0\diamond 1$ is *not* a u-p-word for A^2 because the word 10 in A^2 does not appear as a factor (and the word 01 in A^2 appears twice as a factor).

The focus in this paper is u-p-words with a single \diamond (whenever possible, we provide comments for the case of more than one \diamond). Note that applying the complement to a u-p-word/u-word or a u-p-cycle/u-cycle (that inherits \diamond (s) keeping them in the same position(s)) we obtain a u-p-word/u-word or a u-p-cycle/u-cycle, respectively.

Examples of u-p-words for binary word-patterns of length n , for $1 \leq n \leq 5$, are essentially given in Table 1 in [2], because each example there, except for the case $n = 5$ and $k = 14$, contains 1^n after the \diamond that can be replaced by 1^{n-1} obtaining the desired u-p-word for binary word-patterns, and we present the u-p-words in Table 1. For example, in the u-p-word $0\diamond 01100$ from Table 1 (the case of $n = 3$ and $k = 2$) $0\diamond 0$ covers the word-patterns 000 and 010 , $\diamond 01$ covers 001 and 101 , and the word-patterns 011 , 110 and 100 are also covered (note that 111 is not a word-pattern). Using computer, we also found a u-p-word for binary patterns in the case of $n = 5$ and $k = 14$, which is not provided by [2] but can be found in Table 1. On the other hand, examples of u-p-words for binary set partitions and 2-set partitions for $1 \leq n \leq 5$ are given in Table 2 along with references to the respective general theorems. Moreover, in Table 3 we give examples of u-p-words for binary set partitions and 2-set partitions in the case of $n = 6$.

1.2. Strong connectivity versus connectivity in a balanced directed graph

In a balanced directed finite graph G without isolated vertices the condition to be strongly connected (to have a directed path from any node to any other node) is equivalent to the condition to be connected (to have an undirected path from any node to any other node). The forward direction is trivial. For the backward direction, suppose x and y are two vertices in G and we want to prove the existence of a directed path from x to y .

Since x is not isolated, there exists an edge coming out from it. We start our directed walk following this edge. Because the graph is balanced, whenever we come to a vertex, we always can go out from it and continue our walk. If on our way we will meet y , we are done. Otherwise, eventually we will be stuck at x . Suppose our walk visited a set of vertices V_1 and the set of edges E_1 in G . Since G is connected, there must exist a vertex $v \notin V_1$ connected to a vertex $v_1 \in V_1$, and the directed edge between v and v_1 is not in E_1 . If $v_1 \rightarrow v$ is an edge in G , then we begin a new walk over non-visited edges from this edge, and either we will meet y , or we will eventually get stuck in v_1 , thus enlarging the sets V_1 and E_1 . If $v \rightarrow v_1$ is an edge in G , then because v_1 is balanced, there must exist $v_1 \rightarrow v'_1 \notin E_1$ for $v'_1 \notin V_1$. We can start a walk with the edge $v_1 \rightarrow v'_1$, and again,

TABLE 1. Examples of u-p-words for binary word-patterns of length n with a single \diamond at position k from the beginning or end for $1 \leq n \leq 5$ and all possible values of k (u-p-words where the \diamond is closer to the end of the word than to the beginning can be obtained by reversal). A dash indicates that no such u-p-word exists.

n	k	
1	1	\diamond
2	1	$\diamond 01$
	2	–
3	1	$\diamond 0011010$
	2	$0\diamond 01100$
	3	–
	4	–
4	1	$\diamond 0001110100101100$
	2	$0\diamond 01001101110000$
	3	$01\diamond 011100001010$
	4	–
	5	–
	6	$01100\diamond 01110100$
	7	–
	8	$001110\diamond 0010110$
5	1	$\diamond 000011110111001100010110101001000$
	2	$0\diamond 0101100000110100111011110010001$
	3	$01\diamond 01100000100011100101011110100$
	4	$011\diamond 011110000010100100011010110$
	5	–
	6	$00101\diamond 001001110111100000110101$
	7	$010011\diamond 01000001010110111100011$
	8	$0100110\diamond 0100000111001011110110$
	9	$01110010\diamond 011110110100110000010$
	10	$010011011\diamond 01000111100000101011$
	11	$0101000001\diamond 0101111001110110001$
	12	$01010000011\diamond 010110111100010011$
	13	$001001101011\diamond 00101000001111011$
	14	$0010000011011\diamond 0010100111101011$
	15	$01010000010011\diamond 010110111100011$
	16	$001000001101011\diamond 00101001111011$

either we will meet y , or we will eventually get stuck in v_1 , thus enlarging the sets V_1 and E_1 . Repeating this argument, if necessary, because G is finite, we will eventually reach y , thus establishing a directed path from x to y via a standard argument in this context, which we omit here.

1.3. Organization of the paper

This paper is organised as follows. U-p-words in the context of binary word-patterns, binary set partitions and 2-set partitions are considered in Sections 2.1, 2.2 and 2.3, respectively. Note that even though often very similar, the theory of u-p-words for binary set partitions is not equivalent to the theory of u-p-words for 2-set partitions (e.g. for $n = 2$, $\diamond 1$ is a u-p-word for binary set partitions, while no u-p-word for 2-set partitions exists in this case). U-p-words in the context of non-binary word-patterns and non-binary set partitions are considered for the “context-free” and “context-dependent” cases in Sections 3.1 and 3.2, respectively. In Section 4, we provide concluding remarks.

TABLE 2. Examples of u-p-words for binary set partitions and 2-set partitions of $\{1, 2, \dots, n\}$ for $1 \leq n \leq 5$. “Comp.” stands for “obtained by computer”, and “—” stands for no such u-p-word exists. \diamond is considered only in position k where k is less than or equal to the half of the respective u-p-word, because the reverse of a u-p-word gives a u-p-word.

n	k	Binary set partitions	2-set partitions
1	1	\diamond	—
2	1	$\diamond 1$	—
3	1	$\diamond 1121$ (Thm. 2.6)	$\diamond 122$ (Thm. 2.13)
	2	—	—
4	1	$\diamond 111212211$ (Thm. 2.6)	$\diamond 12111221$ (Thm. 2.13)
	2	— (Comp.)	$1\diamond 221211$ (Thm. 2.15)
	3	— (Comp.)	— (Comp.)
	4	$112\diamond 221$ (Comp.)	N/A
5	1	$\diamond 111121112212122111$ (Thm. 2.6)	$\diamond 12122111121112212$ (Thm. 2.13)
	2	$1\diamond 2112121111122212$ (Thm. 2.9)	$1\diamond 221111211212111$ (Thm. 2.15)
	3	$11\diamond 11212111221111$ (Comp.)	$12\diamond 2111121122121$ (Comp.)
	4	— (Comp.)	— (Comp.)
	5	— (Thm. 2.8)	— (Thm. 2.14)
	6	— (Comp.)	— (Comp.)
	7	— (Comp.)	$112212\diamond 2211121$ (Comp.)
	8	— (Comp.)	N/A

2. BINARY WORD-PATTERNS, BINARY SET PARTITIONS AND 2-SET PARTITIONS

2.1. Binary word-patterns

The following theorem is a particular case of Theorem 3.3 given below.

Theorem 2.1. *U-p-words for binary word-patterns of length n exist for any $n \geq 1$.*

Proof. Let u be a u-p-word for words given in [2]. Observe that u cannot have the \diamond be involved in covering both 0^n and 1^n . Indeed, if that were the case, then it is clear that we must have $0^{n-1}\diamond 1^{n-1}$, or the reverse of it as a factor in u . But it is not possible because in this case $0^{n-2}11$ would be covered twice. Thus, either 0^n , or 1^n , is a factor in u . By applying the complement, if necessary, we can assume that 1^n is a factor in u . But then, removing one 1 in 1^n , we obtain a u-p-word for word-patterns. \square

Remark 2.2. Unfortunately, we cannot make any claims for the converse direction, namely, if existence of a u-p-word for binary word-patterns implies existence of a u-p-word for binary words considered in [2]. Indeed, our u-p-word can potentially be of the form $\dots 01^i \diamond 1^{n-2-i} 0 \dots$ (not covering 1^n), and then there is no place in such a u-p-word to insert a 1 to cover 1^n .

For $n = 2$, $1\diamond$ is a u-p-word for word patterns with the single \diamond in position 2, while no u-p-word for words of length n with the single \diamond in position n exists [2]. Thus, the non-existence results in [2] do not automatically imply the respective non-existence results for word-patterns. So, the following theorem needs justification.

Theorem 2.3. *No u-p-word for binary word-patterns of length n exists with the single \diamond in position n for $n \geq 3$.*

Proof. Suppose $u = *^{n-1}\diamond u_1 u_2 \dots u_{n-1} u_n \dots$ is a u-p-word with the \diamond in position n (here we used that $n \geq 3$, so there are at least n letters to the right of the \diamond). First, note that $u_1 \dots u_{n-1} \neq 1^{n-1}$ because we deal with word-patterns so u does not cover 1^n . This implies that $u_1 \dots u_{n-1}$ must appear twice in u in order to cover both $u_1 \dots u_{n-1} u_n$ and $u_1 \dots u_{n-1} \bar{u}_n$. But then, the second copy of $u_1 \dots u_{n-1}$ must be at the very beginning

TABLE 3. Examples of u-p-words for binary set partitions and 2-set partitions of $\{1, 2, \dots, 6\}$ found by computer. Replacing 111111 by 11111 in each u-p-word for $k = 3, 4, 5, 7$ gives respective examples of u-p-words for 2-set partitions. \diamond is considered only in position k where k is less than or equal to the half of the respective u-p-word, because the reverse of a u-p-word gives a u-p-word.

n	k	Binary set partitions (resp., 2-set partitions)
6	1	exists by Thm 2.6 (resp., Thm. 2.13)
	2	exists by Thm 2.9 (resp., Thm. 2.15)
	3	$12\diamond 121122111112111122212121112212$
	4	$112\diamond 11211212122111111212222111221$
	5	$1112\diamond 11121212211211212221111122$
	6	does not exist by Thm 2.8 (resp., Thm. 2.14)
	7	$121112\diamond 12111112121221122221112$
	8	$1111221\diamond 11112111222121211211221$ (resp., $1112112\diamond 111212122211112212112$)
	9	$11112112\diamond 2222111221121112121221$ (resp., $11112112\diamond 111122121112221212112$)
	10	$111121112\diamond 222211211212122112221$ (resp., $111121221\diamond 22221112212121112112$)
	11	$1111211221\diamond 11112221212112111221$ (resp., $1111221121\diamond 2222122211121212212$)
	12	$11112111221\diamond 1111222121211211221$ (resp., $11112121221\diamond 222211122121112112$)
	13	$111121121112\diamond 222211212122112221$ (resp., $111121212212\diamond 11112221222112212$)
	14	$1111211221112\diamond 22221121212212221$ (resp., $1111212221221\diamond 2222111221212112$)
	15	$11112112111221\diamond 1111222121211221$ (resp., $11112111221121\diamond 222211121212212$)
	16	$111121112121221\diamond 111122211221221$ (resp., N/A)

of u , or otherwise, $0u_1 \cdots u_{n-1}$ or $1u_1 \cdots u_{n-1}$ is covered twice. Thus,

$$u = u_1 u_2 \cdots u_{n-1} \diamond u_1 u_2 \cdots u_{n-1} u_n \cdots$$

resulting in $u_1 \cdots u_n$ be covered twice; contradiction. \square

Actually, arguing as in the proof of Theorem 2.1, any u-p-word or u-p-cycle in [2] for binary words with any number of \diamond s that contains 0^n , or 1^n , can be turned into a u-p-word or u-p-cycle, respectively, for binary word-patterns (which extends Theorem 2.1 to more than one \diamond s).

Remark 2.4. Looking at Table 2 in [2], we see that at least for two \diamond s, there always seems to exist u-p-words for words having 1^n as a factor, so that u-p-words for binary word-patterns with two \diamond s should always exist. However, the general construction for two \diamond s in [2], namely, the one beginning with $\diamond 0^{n-1} 1^{n-2} \diamond 10^{n-2}$, is not of help to us, because both 0^n and 1^n are covered involving \diamond s.

Problem 2.5. Do we have the equivalence of the theory of u-p-words for binary words of length n in [2] with our theory of u-p-words for binary word-patterns of length n for $n \geq 3$?

2.2. Binary set partitions

The theory of u-p-words for binary set partitions is not equivalent to the theory of u-p-words for word-patterns. For example, an analogue of Theorem 2.3 does not hold for set partitions as for $n = 4$, $112\Diamond 221$ is a u-p-word for set partitions. However, letting $n \geq 5$, we do obtain an analogue of Theorem 2.3 (see Thm. 2.8). Also, some other properties are shared, like beginning with a single \Diamond . The following theorem is a particular case of Theorem 3.4, so we omit its proof.

Theorem 2.6. *Let $n \geq 1$. In the case of a single \Diamond , there exists a u-p-word for binary set partitions of $\{1, 2, \dots, n\}$ that for $n \geq 3$ begins with $\Diamond 1^{n-1}2$.*

The following lemma gives a structural property of a u-p-word with a single \Diamond for binary set partitions and 2-set partitions.

Lemma 2.7. *Let u be a u-p-word either for binary set partitions, or for 2-set partitions of $\{1, 2, \dots, n\}$ with a single \Diamond in position i such that there are at least n letters in u to the right of \Diamond . If $2 \leq i < n$ then*

$$u = u_1 u_2 \cdots u_{i-1} \Diamond u_1 u_2 \cdots u_n \cdots \text{ or } u = \bar{u}_1 \bar{u}_2 \cdots \bar{u}_{i-1} \Diamond u_1 u_2 \cdots u_n \cdots$$

If $i > n$ then

$$u = u_1 u_2 \cdots u_{n-1} \bar{u}_n \cdots \Diamond u_1 u_2 \cdots u_n \cdots \text{ or}$$

$$u = \bar{u}_1 \bar{u}_2 \cdots \bar{u}_{n-1} u_n \cdots \Diamond u_1 u_2 \cdots u_n \cdots$$

Proof. Suppose $i \geq 2$ and $u = *^{i-1} \Diamond u_1 u_2 \cdots u_{n-1} u_n \cdots$ is a u-p-word with the \Diamond in position i . In order to cover both of the set partitions $u_1 \cdots u_{n-1} u_n$ and $u_1 \cdots u_{n-1} \bar{u}_n$, u must cover, possibly involving the \Diamond , either one more $u_1 u_2 \cdots u_{n-1}$, or $\bar{u}_1 \bar{u}_2 \cdots \bar{u}_{n-1}$ (since $u_1 \cdots u_{n-1} \bar{u}_n$ and $\bar{u}_1 \bar{u}_2 \cdots \bar{u}_{n-1} u_n$ are the same partitions). But then, the second coverage of $u_1 \cdots u_{n-1}$, or coverage of $\bar{u}_1 \bar{u}_2 \cdots \bar{u}_{n-1}$, must be at the very beginning of u not to cover twice either $1u_1 \cdots u_{n-1}$ or $2u_1 \cdots u_{n-1}$ (since the factor $u_1 \cdots u_{n-1}$ is preceded by the \Diamond in u). This proves the lemma. \square

The following theorem is an analogue of Theorem 2.3 for binary set partitions.

Theorem 2.8. *For $n \geq 5$, no u-p-word for binary set partitions of $\{1, 2, \dots, n\}$ exists with the single \Diamond in position n .*

Proof. Since $n \geq 5$, there are at least n elements to the right of the \Diamond in a u-p-word u in question, assuming it exists, as u must be of length $2^{n-1} - 1$ in this case. Thus, by Lemma 2.7 one of the two cases holds:

- $u = u_1 u_2 \cdots u_{n-1} \Diamond u_1 u_2 \cdots u_{n-1} u_n \cdots$ resulting in $u_1 u_2 \cdots u_n$ be covered twice; contradiction.
- $u = \bar{u}_1 \bar{u}_2 \cdots \bar{u}_{n-1} \Diamond u_1 u_2 \cdots u_{n-1} u_n \cdots$ again resulting in $u_1 u_2 \cdots u_n$ be covered twice ($\bar{u}_1 \bar{u}_2 \cdots \bar{u}_n$ is the same set partition); contradiction.

Thus, no such u can exist. \square

Let G_n be the *canonical transition graph* of set partitions of $\{1, 2, \dots, n\}$. That is, each vertex of G_n represents a unique set partition encoded canonically, each set partition is represented by a vertex in G_n , and there is an edge from $\mathbf{cf}(x_1 x_2 \cdots x_n)$ to $\mathbf{cf}(y_1 y_2 \cdots y_n)$ if and only if $\mathbf{cf}(x_2 x_3 \cdots x_n) = \mathbf{cf}(y_1 y_2 \cdots y_{n-1})$. For any set partition π encoded by a word $w_1 w_2 \cdots w_n$, the word “ $\mathbf{cf}(w_1 w_2 \cdots w_{n-1})$ ” is the *signature* of π . The *reduced transition graph* \bar{G}_n is obtained from G_n by placing all set partitions with the same signature into the same *cluster* marked by the signature, and letting the edges in \bar{G}_n be pointing from $\mathbf{cf}(x_1 x_2 \cdots x_n)$ to the cluster with the signature “ $\mathbf{cf}(x_2 x_3 \cdots x_n)$ ” for each $\mathbf{cf}(x_1 x_2 \cdots x_n)$. We refer to Figure 11 in [3] for drawn G_3 and \bar{G}_3 .

Theorem 2.9. *For $n \geq 5$ and $2 \leq i \leq \lfloor \frac{n}{2} \rfloor$, there exists a u-p-word with a single \Diamond for 2-set partitions of $\{1, 2, \dots, n\}$ that begins with $1^{i-1} \Diamond \underbrace{2^i 1^i}_{n-1} \cdots$.*

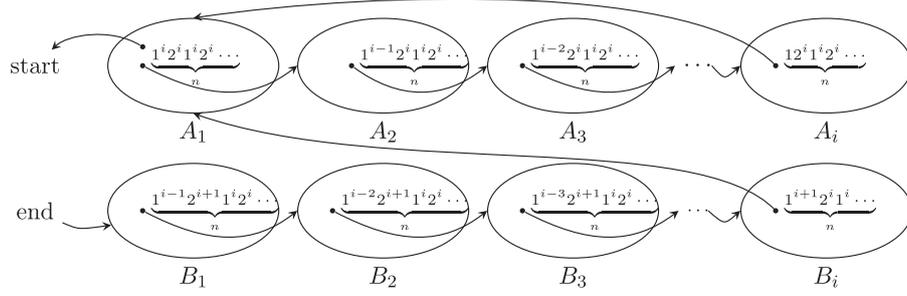


FIGURE 1. To support the proof of Theorem 2.9.

Proof. For $1 \leq j \leq i$, let A_j (resp., B_j) denote the cluster in \bar{G}_n with the signature $\underbrace{1^{i+1-j} 2^j 1^i 2^i \dots}_{n-1}$ (resp., $\underbrace{1^{i-j} 2^{i+1} 1^i 2^i \dots}_n$), where 1^0 is the empty word.

None of A_j s and B_j s can have a loop (a vertex in a cluster pointing to the same cluster) because $i \geq 2$, but clusters with a loop must have one of the signatures 1^{n-1} or $1212\dots$. Further note that for $1 \leq j \leq i-1$, there is an edge pointing from the vertex $\underbrace{1^{i+1-j} 2^j 1^i 2^i \dots}_n$ in A_j to A_{j+1} . Also, there is an edge pointing from the vertex $\underbrace{12^i 1^i 2^i \dots}_n$ in A_i to A_1 . Similarly, for $1 \leq j \leq i-1$, there is an edge pointing from the vertex $\underbrace{1^{i-j} 2^{i+1} 1^i 2^i \dots}_n$ in B_j to B_{j+1} . Also, there is an edge pointing from the vertex $\underbrace{1^{i+1} 2^i 1^i 2^i \dots}_n$ in B_i to B_1 . See

Figure 1 for the respective part of \bar{G}_n .

Clearly, A_j s are distinct, and B_j s are distinct for $1 \leq j \leq i$. Moreover, since $2i \leq n$, each of B_j s has the factor of the form x^{i+1} , but none of A_j s has such a factor. Thus, all A_j s and B_j s are distinct. Finally, note that the vertices presented in Figure 1 are *all* the words of length n covered using \diamond in a p-word that begins with $1^{i-1} \diamond \underbrace{2^i 1^i \dots}_{n-1}$. Removing all these vertices, together with the edges coming out from them, from \bar{G}_n , we obtain

\bar{G}'_n which is balanced except the cluster A_1 now has an extra out-degree, and B_1 has an extra in-degree. Since the connectivity of \bar{G}_n is not affected by the removed vertices and edges (we removed exactly one vertex from each of A_j s and B_j s), which is to be proved below, we conclude that \bar{G}'_n has an Eulerian path starting at the remaining vertex in A_1 and ending at the remaining vertex in B_1 . Such a path gives us a way to extend the p-word beginning with $1^{i-1} \diamond \underbrace{2^i 1^i \dots}_{n-1}$ to a desired u-p-word, because immediately to the right of \diamond we have a factor matching the signature of A_1 .

To complete the proof, we show that \bar{G}'_n is connected (which will be equivalent to strong connectivity of \bar{G}'_n by Section 1.2). Let us call the partitions/vertices that are covered by using the \diamond *bad*, and all other partitions/vertices *good*. Our goal is to show that for any good vertex p , either there is a directed path avoiding bad vertices from p to the good vertex $p^* = 1212\dots$, or there is a directed path avoiding bad vertices from the vertex p^* to p .

The most generic form of p is $1^{x_1} 2^{y_1} 1^{x_2} 2^{y_2} \dots z_{n-2} z_{n-1} z_n$, where $x_1 > 0$ and all other $x_i, y_j \geq 0$ so that if $x_i = 0$ then $y_i = 0$ and $x_j = y_j = 0$ for all $j > i$, and if $y_i = 0$ then, again, $x_j = y_j = 0$ for all $j > i$. The $z_{n-2} z_{n-1} z_n$ records the last three letters of p . If the vertex $p' = 21^{x_1} 2^{y_1} 1^{x_2} 2^{y_2} \dots z_{n-2} z_{n-1}$ is good, then there is an edge coming from this vertex to p 's cluster, and we can obtain a sequence of good vertices by appending to the left 1 and removing z_{n-1} , then appending to the left 2 and removing z_{n-2} , etc, until we reach the good vertex p^* (this uses the property that no bad vertex can contain the factor 121 or 212 since $i \geq 2$). This gives us a directed path from p^* to p , as desired.

In the worst case scenario, the vertex $p' = 21^{x_1}2^{y_1}1^{x_2}2^{y_2} \cdots z_{n-2}z_{n-1}$ is bad. We conclude that the canonical form of p' is either

- (a) $12^{i+1}1^i2^i \cdots \bar{z}_{n-2}\bar{z}_{n-1}$, or
- (b) $12^i1^i2^i \cdots \bar{z}_{n-2}\bar{z}_{n-1}$

because only the vertices in the clusters A_i and B_1 begin with a single 1, where \bar{x} is the complement of a letter x . If (a) is the case, then p is $1^{i+1}2^i \cdots z_{n-2}z_{n-1}z_n$, so we can redefine p' to be obtained from p by appending 1 to the left, so that $p' = 1^{i+2}2^i \cdots z_{n-2}z_{n-1}$ is good. Appending now 2 to the left and removing z_{n-1} , then appending 1 to the left and removing z_{n-2} , etc, will clearly result in a sequence of good vertices leading to p^* (first two steps of the process use the property that no bad vertex contains the factor 1^{i+2} or 2^{i+2} , and the remaining steps use the property that no bad vertex can contain the factor 121 or 212 since $i \geq 2$). This gives us a directed path from p^* to p , as desired.

If (b) is the case, then p is $1^i2^i1^i2^i \cdots z_{n-2}z_{n-1}z_n$, so that z_n breaks the pattern using which bad vertices are constructed. We consider the following possible endings of p and explain how to find a directed path from p to the vertex p^* in each of the cases.

- (c) $p = 1^i2^i1^i2^i \cdots 1^{i+1}$. Consider the following directed path in \bar{G}_n : $p \rightarrow 1^{i-1}2^i1^i2^i \cdots 1^{i+2} \rightarrow 1^{i-2}2^i1^i2^i \cdots 1^{i+2}2 \rightarrow z_1z_2z_3 \cdots 1^{i+2}21 \rightarrow z_2z_3 \cdots 1^{i+2}212 \rightarrow z_3 \cdots 1^{i+2}2121 \rightarrow \cdots \rightarrow p^*$ where $z_1z_2z_3$ is an initial factor of the respective vertex. All of the vertices involved are good, which is guaranteed by the presence of 1^{i+2} in the first two steps, and then the presence of 121.
- (d) $p = 1^i2^i1^i2^i \cdots 2^{i+1}$. Same as (c) but interchanging 1s and 2s.
- (e) $p = 1^i2^i1^i2^i \cdots 12^x1$ where $1 \leq x < i$. In this case, each of the vertices, but possibly p , in the following path contains 212 and/or 121, so is good: $p \rightarrow 1^{i-1}2^i1^i2^i \cdots 12^x12 \rightarrow 1^{i-2}2^i1^i2^i \cdots 12^x121 \rightarrow \cdots \rightarrow p^*$.
- (f) $p = 1^i2^i1^i2^i \cdots 21^x2$ where $1 \leq x < i$. Same as (e) but interchanging 1s and 2s.

□

Remark 2.10. It is easy to check that the results of Theorem 2.9 are not true for the case of $n = 4$, because there will be no possibility for such a u-p-word to cover the set partition 1111. This happens because 1111 is in the same cluster B_2 as 1112, and 1112 was covered using the \diamond , so the Eulerian path will never visit B_2 to be able to cover 1111. For $n \geq 5$, this is never a problem as $1 \cdots 1$ is never in one of the clusters $A_1, \dots, A_i, B_1, \dots, B_i$. Equally important is that, for $n \geq 5$, the set partition $1212 \cdots$ (giving a loop in the reduced transition graph like $1 \cdots 1$) is also never in one of the clusters $A_1, \dots, A_i, B_1, \dots, B_i$, so the Eulerian path will visit it. Finally, we have used $\lfloor n/2 \rfloor$ instead of $\lceil n/2 \rceil$ by the same reason: if $n = 5$, then $\lceil n/2 \rceil = 3$, so choosing $i = 3$ we obtain that $1^{i+1}2$ and 11111 are in the same cluster B_5 .

Theorem 2.11. *For $n \geq 6$, there exists a u-p-word with a single \diamond for binary set partitions of $\{1, 2, \dots, n\}$ that begins with $12^{n-3}\diamond 12^{n-3}12$.*

Proof. It can be checked that $1222\diamond 12221211211221212111112222$ is a u-p-word for binary set partitions in the case $n = 6$. Thus, we can assume that $n \geq 7$, which is to be used by us when proving the connectivity below.

In our proof we mimic the steps in the proof of Theorem 2.9, but using Figure 2 instead. It is easy to see that none of the clusters in Figure 2 corresponding to the vertices covered by \diamond has a loop (whose signature would have to be either 1^{n-1} or $1212 \cdots$), and also the clusters are distinct. Removing all the vertices obtained using \diamond , together with the edges coming out from them, from the reduced transition graph \bar{G}_n , we obtain \bar{G}'_n which is balanced except the cluster with signature " $12^{n-3}1$ " has an extra out-degree, and the cluster with signature " $12^{n-2}1$ " has an extra in-degree. As in the proof of Theorem 2.9, it remains to show that \bar{G}'_n is connected, which we do in a similar way to Theorem 2.9 by again introducing "good" (*i.e.* not covered using \diamond) and "bad" (*i.e.* covered using \diamond) vertices and showing how to find a directed path from any good vertex p to the good vertex $p^* = 1212 \cdots$, or from p^* to p .

The most generic form of p is $1^{x_1}2^{y_1}1^{x_2}2^{y_2} \cdots z_{n-3}z_{n-2}z_{n-1}z_n$, where $x_1 > 0$ and all other $x_i, y_j \geq 0$ so that if $x_i = 0$ then $y_i = 0$ and $x_j = y_j = 0$ for all $j > i$, and if $y_i = 0$ then, again, $x_j = y_j = 0$ for all $j > i$. The

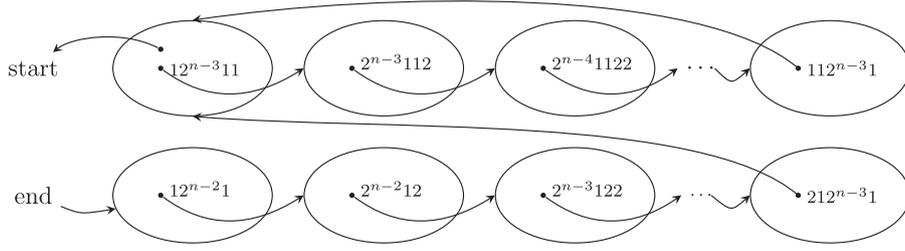


FIGURE 2. To support Theorem 2.11.

$z_{n-3}z_{n-2}z_{n-1}z_n$ records the last four letters of p . If the vertex $p' = 21^{x_1}2^{y_1}1^{x_2}2^{y_2} \dots z_{n-2}z_{n-1}$ is good, then there is an edge coming from this vertex to p 's cluster, and we have two subcases to consider depending on whether the vertex $p'' = 121^{x_1}2^{y_1}1^{x_2}2^{y_2} \dots z_{n-3}z_{n-2}$ is good or bad (note that there is an edge coming from p'' to the cluster containing p').

- If p'' is good, then we can obtain a sequence of good vertices by appending to the left 2 and removing z_{n-2} , then appending to the left 2 and removing z_{n-3} , etc., until we reach the good vertex p^* (this uses the property that no bad vertex can contain the factor 1212 or 2121). This gives us a directed path from p^* to p , as desired.
- If p'' is bad, then we must have $p'' = 121^{n-3}2$, which is the vertex in the last cluster in the bottom row in Figure 2. But then, because $21^{n-3}22$ is bad, we must have $p' = 21^{n-3}21$ (p' is assumed to be good and the edge coming from p'' goes to the cluster of p'). Further, because $1^{n-3}211$ is a bad vertex, we must have $p = 1^{n-3}212$ (p is assumed to be good and the edge coming from p' goes to the cluster of p). But then it is straightforward to find a directed path from p to p^* :

$$p \rightarrow 1^{n-4}2121 \rightarrow 1^{n-5}21212 \rightarrow \dots \rightarrow p^*$$

which uses the property that no bad vertex contains 1212 or 2121 as a factor.

If $p' = 21^{x_1}2^{y_1}1^{x_2}2^{y_2} \dots z_{n-2}z_{n-1}$ is bad, then p' is one of the partitions in the set $\{12^{n-3}11, 2112^{n-3}, 12^{n-2}1, 212^{n-3}1\}$. We consider four subcases.

- (a) If $p' = 12^{n-3}11$. Then we must have $p = 2^{n-3}111$, because the vertex $2^{n-3}112$ is bad (p is assumed to be good). But then, taking into account that $n \geq 7$, none of the vertices in the following directed path from p to p^* is bad:

$$p \rightarrow 2^{n-4}1112 \rightarrow 2^{n-5}11121 \rightarrow 2^{n-6}111212 \rightarrow \dots \rightarrow p^*.$$

- (b) If $p' = 2112^{n-3}$, redefine $p' := 1112^{n-3}$, which is the other vertex with the property that there is an edge from p' to the cluster of p . But then, again taking into account that $n \geq 7$, we have the following directed path from p^* to p involving only good vertices:

$$p^* \rightarrow \dots \rightarrow 2121112^{n-6} \rightarrow 121112^{n-5} \rightarrow 21112^{n-4} \rightarrow p' \rightarrow p.$$

- (c) If $p' = 12^{n-2}1$, redefine $p' := 2^{n-1}1$, which is the other vertex with the property that there is an edge from p' to the cluster of p . In this case we have the following directed path from p^* to p involving only good vertices:

$$p^* \rightarrow \dots \rightarrow 1212^{n-3} \rightarrow 212^{n-2} \rightarrow 12^{n-1} \rightarrow p' \rightarrow p.$$

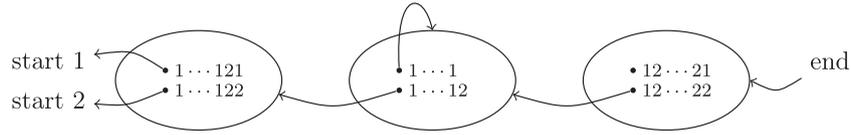


FIGURE 3. To support Theorem 2.13.

- (d) If $p' = 212^{n-3}1$, redefine $p' := 112^{n-3}1$, which is the other vertex with the property that there is an edge from p' to the cluster of p . Note that the edge coming out from the good vertex $p'' = 1112^{n-3}$ comes to the cluster of p' . But in (b) we gave a directed path of good vertices from p^* to p'' (in (b) the vertex 1112^{n-3} is called p'), which gives a desired path from p^* to p in this case.

The connectivity of \bar{G}_n is proved, which proves the existence of a u-p-word in question. \square

Remark 2.12. Note that Theorem 2.11 is not true for $n = 5$ because then the \diamond will be used to cover the set partition 12211 twice.

To extend the results in Theorems 2.9 and 2.11, we provide the following easy to check facts for 2-set partitions:

2.3. 2-set partitions

Recall that 2-set partitions are the same as binary set partitions without the partition of the form $1 \cdots 1$.

We have the following counterpart of Theorem 2.6 for 2-set partitions, where note that clearly no u-p-word for $n = 1, 2$ exists.

Theorem 2.13. *Let $n \geq 3$. In the case of a single \diamond , there exist u-p-words for 2-set partitions of $\{1, 2, \dots, n\}$ that begin with $\diamond 1^{n-2}21$ and $\diamond 1^{n-2}22$, respectively.*

Proof. Note that in both cases of beginning a partial word in the statement of the theorem, the \diamond covers the set partitions $1^{n-1}2$ and $12^{n-2}1$ thus removing two vertices and the respective edges in the reduced transition graph \bar{G}_n ; see Figure 3 for illustration of the situation. This makes the cluster with the signature “ 1^{n-1} ” isolated. However, we deal with 2-set partitions, so that the vertex 1^n is not to be covered by us anyway. We see that in the graph \bar{G}_n with the two vertices removed, every cluster is balanced except the cluster with the signature “ 1^{n-2} ” has one extra out-going edge, and the cluster with the signature 12^{n-2} has one extra in-coming edge. Thus, in such a graph, an Eulerian path exists giving us the respective u-word for 2-set partitions. Finally, we note that there are two possible choices to begin such a path, as indicated in Figure 3, which gives two desired possibilities to begin the respective u-p-words. \square

The following theorem is an analogue of Theorem 2.8, and its proof is identical to the proof of Theorem 2.8.

Theorem 2.14. *For $n \geq 5$, no u-p-word for 2-set partitions of $\{1, 2, \dots, n\}$ exists with the single \diamond in position n .*

For 2-set partitions we have the following counterpart of Theorem 2.9 where the only difference is $n \geq 5$ instead of $n \geq 4$.

Theorem 2.15. *For $n \geq 4$ and $2 \leq i \leq \lfloor \frac{n}{2} \rfloor$, there exists a u-p-word with a single \diamond for 2-set partitions of $\{1, 2, \dots, n\}$ that begins with $1^{i-1} \diamond \underbrace{2^i 1^i}_{n-1} \dots$.*

Proof. Note that for $n = 4$, the p-word $1 \diamond 221211$ is universal. Assuming $n \geq 5$ we can literally follow all arguments in the proof of Theorem 2.9 to obtain the desired result, because the presence/absence of the set

partition 1^n does not play any role based on Remark 2.10 (also, proving the connectivity of \tilde{G}'_n we do not use 1^n). \square

Finally, the presence/absence of the set partition 1^n does not play any role in the proof of Theorem 2.11, because 1^n is never in the same cluster as a set partition covered using the \diamond , which automatically gives us the following result.

Theorem 2.16. *For $n \geq 6$, there exists a u-p-word with a single \diamond for 2-set partitions of $\{1, 2, \dots, n\}$ that begins with $12^{n-3}\diamond 12^{n-3}12$.*

3. NON-BINARY WORD-PATTERNS AND SET PARTITIONS

3.1. Non-binary word-patterns and set partitions with a single \diamond

Theorem 3.1. *No u-p-word exists for word-patterns over $\{0, 1, \dots, \alpha - 1\}$ for $\alpha \geq 3$.*

Proof. A word-pattern of length n can be built on at most n distinct letters. Thus, if $n < \alpha$, then letting the \diamond take the value of $\alpha - 1$ will definitely result in a non-word pattern being covered. This argument actually works for any number of \diamond s in a prospective u-p-word in question.

If $n \geq \alpha$, clearly, without loss of generality, we can assume that a u-p-word u in question is of the form

$$\dots \diamond x_1 x_2 \dots x_n \dots$$

But then, $\omega = x_1 x_2 \dots x_{n-1}$ must be a word over $\{0, 1, \dots, \alpha - 2\}$ to accommodate the substitution of the \diamond by $\alpha - 1$, so ω has at least two distinct letters, and thus it must have at least three possible extensions to the right, say ωx_n , ωa and ωb . Since $\omega \diamond$ cannot be a factor in u (or else, u covers ωx_n twice), u must contain at least two more occurrences of ω , only one of which can be at the beginning of u . Thus, u must contain a factor $y\omega$ for some y , which is covered twice (additionally by $\diamond\omega$); contradiction. No such u can exist. \square

Next, we consider set partitions of $\{1, 2, \dots, n\}$ having at most n parts for $n \geq 3$. It is straightforward to check that $\diamond 1222$ is a u-p-word in the case of $n = 3$.

Theorem 3.2. *No u-p-word with a single \diamond for all set partitions of $\{1, 2, \dots, n\}$ exists for $n \geq 4$.*

Proof. Suppose such a u-p-word u exists. Reading u in the reverse order, if necessary, we can assume that u has a factor $v = \diamond x_1 x_2 \dots x_{n-1}$. Then the letters in $x_1 x_2 \dots x_{n-1}$ must be distinct, or else u covers a set partition among $1x_1 x_2 \dots x_{n-1}$, $2x_1 x_2 \dots x_{n-1}, \dots, nx_1 x_2 \dots x_{n-1}$ twice. Thus, without loss of generality, we can assume that $v = \diamond 12 \dots (n-1)$. Now, if v is preceded by a letter x in u , that is, if $v' = x \diamond 12 \dots (n-1)$ is a factor in u , then x cannot be n or $n-1$ since then v' covers the set partition $12 \dots n$ already covered by v . On the other hand, arguing as in the case of v , we see that if $1 \leq x \leq n-2$ then v' covers the same set partition twice. Thus, u must begin with v , so that the \diamond is the leftmost element in u , and we assume that $u = \diamond 12 \dots (n-1)x \dots$ for some letter x .

Note that v covers the partitions $12 \dots n$, $1123 \dots (n-1)$, $12134 \dots (n-1)$, $123145 \dots (n-1), \dots, 12 \dots (n-1)1$ (encoded in the canonical form). Among these partitions only two begin with $12 \dots (n-1)$, namely, the first one and the last one. However, since $n \geq 4$, there are at least two more partitions, namely, $12 \dots (n-1)2$ and $12 \dots (n-1)(n-1)$ to be covered by u . Only one of these partitions can be covered by $12 \dots (n-1)x$ staying next to the \diamond . If $y_1 y_2 \dots y_n$ is a factor of u covering the other partition, and $yy_1 y_2 \dots y_n$ is a factor of u for some letter y , then we obtain a contradiction with the partition $yy_1 y_2 \dots y_{n-1}$ being covered twice (the second time by $\diamond 12 \dots (n-1)$). Thus, no such u can exist. \square

3.2. Context-dependent non-binary word-patterns and set partitions

As we have seen in Section 3.1, usage of a “context-free” \diamond is not so useful in shortening u-words for word-patterns and set partitions in the non-binary case. By “context-free” we mean that \diamond can be substituted by

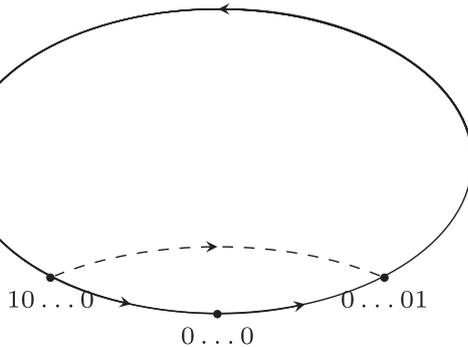


FIGURE 4. Hamiltonian cycle in the proof of Theorem 3.3.

any letter from the available range. This brought us to the idea of introducing a “context-dependent” \diamond , which actually allows to shorten u-words in many situations. By “context-dependent” here we mean that if a factor containing \diamond s has t distinct letters, then there are $t + 1$ possible substitutions for such \diamond s, which are using one of the letters, or using a different letter (in the context of word-patterns, this letter must be $m + 1$, while in the context of set partitions, its value is not important). For example, the u-p-word $\diamond 1222$ for $n = 3$ discussed in Section 3.1 covers all partitions, namely, 112, 121, 123, 122 and 111, in both senses, while the u-p-word $\diamond 11213$ also covers all the partitions in the context-dependent sense (when \diamond can only be 1 or 2), but covers twice the partition 122 in the context-free sense (letting the \diamond be 2 or 3 yields the same result).

The following two theorems give examples of existence results in the case of word-patterns and set partitions, respectively, when context-dependant \diamond s are applied.

Theorem 3.3. *In the case of a single context-dependent \diamond , there exists a u-p-word for word-patterns over $\{0, 1, \dots, \alpha - 1\}$, $\alpha \geq 1$, of length $n \geq \alpha$ that begins with $\diamond 0^{n-1}1$.*

Proof. It is known [1] that there exists a u-cycle for word-patterns of length n over $\{0, 1, \dots, \alpha - 1\}$ for any $\alpha \geq 1$ and $n \geq \alpha$. Such a u-cycle corresponds to a Hamiltonian cycle in the respective transition graph (which is an induced subgraph of the de Bruijn graph and is called in [1] the *graph of pattern overlaps*). Such a Hamiltonian cycle, shown schematically in Figure 4, must have the subpath $10^{n-1} \rightarrow 0^n \rightarrow 0^{n-1}1$. But then, if we remove the vertices 10^{n-1} and 0^n from the graph, we would obtain a Hamiltonian path (not cycle!) in it that begins at $0^{n-1}1$, follows the original Hamiltonian cycle, and ends at the vertex from which the Hamiltonian cycle got to the removed 10^{n-1} . Such a Hamiltonian path (beginning with $0^{n-1}1$), together with the initial factor of $\diamond 0^{n-1}$ covering 10^{n-1} and 0^n (recall that \diamond is context-dependent), gives the desired u-word. \square

Theorem 3.4. *Let $n \geq 1$. In the case of a single context-dependent \diamond , there exists a u-p-word for all set partitions of $\{1, 2, \dots, n\}$ that for $n \geq 3$ begins with $\diamond 1^{n-1}2$.*

Proof. Clearly \diamond and $\diamond 1$ are u-p-words for all set partitions of $\{1\}$ and $\{1, 2\}$, respectively. Also, the statement is true for $n = 3$ by the example $\diamond 11213$ given above. Thus, we can assume that $n \geq 4$, and thus there exists a u-cycle for all set partitions of $\{1, 2, \dots, n\}$ by [3] (no such u-cycle exists for $n = 3$). Such a u-cycle defines a Hamiltonian graph in the respective transition graph. The rest of the proof essentially copies the proof of Theorem 3.3.

Such a Hamiltonian cycle must have the subpath $21^{n-1} \rightarrow 1^n \rightarrow 1^{n-1}2$. But then, if we remove the vertices 21^{n-1} and 1^n from the graph, we would obtain a Hamiltonian path in it that begins at $1^{n-1}2$, follows the original Hamiltonian cycle, and ends at the vertex from which the Hamiltonian cycle got to the removed 21^{n-1} . Such a Hamiltonian path (beginning with $1^{n-1}2$), together with the initial factor of $\diamond 1^{n-1}$ covering 21^{n-1} and 1^n (recall that \diamond is context-dependent), gives the desired u-word. \square

4. CONCLUDING REMARKS

This paper initiates the study of universal partial words (u-p-words) for word-patterns and (2-)set partitions. One of the most significant ideas we introduce in this paper is usage of context-dependent \diamond s in the context of u-p-words, which can be seen as a brother of the restricted \diamond s studied in the context of permutations in [6]. We hope our studies will raise interest, and more results in this direction will be obtained, which would classify existence of u-p-words for word-patterns and (2-)set partitions with a single \diamond and/or give some existence results when more than one \diamond is used.

Acknowledgements. The authors are grateful to the anonymous referee for providing their numerous comments helping to improve the presentation of the paper. The work of the first author was supported by the National Natural Science Foundation of China (Grant Numbers 11901319) and the Fundamental Research Funds for the Central Universities (Grant Number 63191349).

REFERENCES

- [1] A. Burstein and S. Kitaev, On Unavoidable Sets of Word Patterns. *SIAM J. Discr. Math.* **19** (2005) 371–381.
- [2] H.Z.Q. Chen, S. Kitaev, T. Mütze and B.Y. Sun, On universal partial words. *Discr. Math. Theor. Comp. Sci.* **19** (2017) 16.
- [3] F.R.K. Chung, P. Diaconis and R. Graham, Universal cycles for combinatorial structures. *Discr. Math.* **110** (1992) 43–59.
- [4] B. Goeckner, C. Groothuis, C. Hettle, B. Kell, P. Kirkpatrick, R. Kirsch and R. Solava, Universal Partial Words over Non-Binary Alphabets. *Theor. Comp. Sci.* **713** (2018) 56–65.
- [5] Z. Higgins, E. Kelley, B. Sieben, A. Godbole. Universal and near-universal cycles of set partitions. *Electron. J. Combin.* **22** (2015) 4, Paper 4.48, 15pp.
- [6] S. Kitaev, V.N. Potapov and V. Vajnovszki, On shortening u-cycles and u-words for permutations. *Discr. Appl. Math.* **260** (2019) 203–213.