ON THE SEPARATION OF DOUBLE-OCCURRENCE WORDS

Brad Mostowski

July 25, 2018

Abstract

We develop the notion of a double-occurrence word and also the separation of a double-occurrence word, which describes how "scrambled" a DOW appears. We attempt in this paper to classify for arbitrary n, k, DOWs of size n with separation k. We first show that a DOW of size n always has even separation and is bounded above by $n(n-1)$. We then classify size n DOWs of separation $0, 2, 4$ and $n(n-1)$.

1 Preliminaries

We first begin with some preliminaries. An **alphabet** Σ is a countable set. Elements of an alphabet are called **symbols**.

Definition 1.1. *Let Σ be an alphabet. A* ***word*** *u over Σ is a finite sequence of (not necessarily unique) symbols $\{x_i\}_{i=1}^k$ with $x_i \in \Sigma$ for each i. Elements of u are called* ***letters****. A subsequence v of u is called a* ***subword*** *of u. A subword of $u = \{x_i\}_{i=1}^k$ of the form $v = \{x_a, x_{a+1}, \ldots, x_b\}$ with $1 \leq a \leq b \leq k$ is called a* ***factor*** *of u. The word $u = \{x_i\}_{i=1}^k$ is called a* ***double-occurrence word****, or a* ***DOW****, if for any $a \in \Sigma$, we have the set S_a of all x_i's such that $a = x_i$ is of cardinality either 0 or 2.*

Remark 1. The word $u = \{x_i\}_{i=1}^k$ for convenience is written $u = x_1 x_2 \cdots x_k$ throughout this paper. The set of all words over Σ is denoted Σ^*. Let $u \in \Sigma^*$. The notation $\Sigma[u]$ denotes the set of all symbols in Σ which appear at least once in u. $\Sigma[u]$ may be equivalently defined as the smallest alphabet containing u. The set of all double-occurrence words over Σ is denoted Σ_{DOW}.

Example 1. The word $u = 111232122$ is a word over $\{1, 2, 3\}$. The word $v = 13$ is a subword (but not a factor) of u. The word $w = 2321$ is a factor of u.

Definition 1.2. *Let $u \in \Sigma_{DOW}$. The* ***size*** *of u, written $size(u)$ is the value $|\Sigma[u]|$.*

Remark 2. If u is a word with $2n$ letters, then $size(u) = n$.

Definition 1.3. *The* ***ascending order representation*** *of $u \in \Sigma^*$, sometimes denoted u_a, is the DOW that results from rewriting the ith unique symbol which appears in u as i. The word u is said to be in* ***ascending order*** *if $u = u_a$. Two DOWs $u, v \in \Sigma^*$ are said to be* ***ascending order equivalent*** *if they have the same ascending order representation, in which case we write $u \sim v$.*

Remark 3. Let $u \in \Sigma_1^*$ and $v \in \Sigma_2^*$. Then, $u, v \in \Sigma[u] \cup \Sigma[v]$. Thus, there is no loss of generality in the above definition in assuming u, v to be in the same alphabet. Ascending order equivalence will be referred to throughout this paper simply as equivalence.

Example 2. Let $u = 1232454134 \in \mathbb{N}_{DOW}$. Then, u is in ascending order. The ascending order representation of $u = 3443$ is $u_a = 1221$.

Definition 1.4. *Let $u \in \Sigma_{DOW}$. Then, u is called a **repeat word** (resp. **return word**) of size n if $u \sim 1 \cdots n1 \cdots n$ (resp. $u \sim 1 \cdots nn \cdots 1$).*

Definition 1.5. *Let $u \in \Sigma_{DOW}$ and $[u]$ the set of all double-occurrence words v such that $u \sim v$. Then, $[u]$ is called the **assembly word class** of u.*

Remark 4. The above definition is still valid if "double-occurrence words" is replaced with "words." Many properties that apply to DOWs have obvious analogues for assembly words. For example, if u is irreducible (see section 3), so is any v such that $u \sim v$, whence it naturally makes sense to speak of an assembly word $[u]$ being irreducible.

While the notation $[u]$ is identical to the notation in this paper used to denote the set $[n] = \{1, \ldots, n\}$, in practice this will not cause confusion. Of course, $[u] = [v]$ if and only if $u \sim v$. Every assembly word class $[u]$ contains a unique DOW v such that v is in ascending order, thus as an abuse of notation, we refer to assembly word classes simply as "assembly words" and sometimes identify an assembly word by the unique ascending order DOW v that belongs to it. Thus, the assembly word $[u] = [5665]$ may simply be identified as 1221.

Example 3. The assembly word of 5656 is $[5656] = [1212]$.

2 Separations of DOWs

In this section we define the separation value of a DOW, which gives a natural notion for how "scrambled" a DOW appears. Some basic properties, including a properties proving the separation of an arbitrary DOW to be even and giving an upper bound on the separation of a DOW of a given size, are proved. All words from here on out are assumed to be DOWs unless specified otherwise.

Definition 2.1. *Let $u \in \Sigma^*$ be a word such that each element of Σ appears in u at most twice. Suppose $a \in \Sigma$ appears twice in u, and let $x, y, z \sqsubset u$ such that $u = xayaz$. The **separation of** a **in** u is $sep_u(a) = |y|$. If $b \in \Sigma$ appears at most once in u, then we say that $sep_u(b)$ is zero. The value $sep(u) = \sum_{a \in \Sigma[u]} sep_u(a)$ is called the **separation** of u.*

Remark 5. If $u \in \Sigma^*$ and $v \in \Sigma^*$ are such that either $u \sim v$ or $u \sim v^R$, then $sep(u) = sep(v)$.

Example 4. Let $u = 121233$. Then, the separation of u is 2. Let $v = 123123$. Then, the separation of v is 6.

Lemma 2.2. *Let $u \in \Sigma_{DOW}$. Then, $sep(u)$ is even.*

Proof. Let n be the size of u. Without loss of generality, we may assume that $\Sigma = [n]$. Then, the symmetric group S_{2n} acts on $[n]_{DOW}$ in the following manner:

$$S_{2n} \times [n]_{DOW} \to [n]_{DOW}$$
$$(\sigma, u = u_1 \cdots u_{2n}) \mapsto u_{\sigma(1)} \cdots u_{\sigma(2n)}$$

In the above, $u = u_1 \dots u_{2n}$ is a double-occurrence word with the u_i's letters of u.

Let $u \in [n]_{DOW}$ be such that $sep(u)$ is even. (Such a word always exists; pick $u = 11 \cdots nn$.) We claim for any $\sigma \in S_{2n}$ that σu is even also. To see this, Consider $u' = (i, i+1)u$, $1 \le i \le n-1$. (That is, u' is the result of acting on u with $(i, i+1)$). We have

$$u' = u_1 \cdots u_{i-1} u_{i+1} u_i u_{i+2} \cdots u_{2n}$$

If $u_i = u_{i+1}$ then the separation of u' is precisely that of u and is still even. If $u_i = x$ and $u_{i+1} = y$ are distinct, then

$$\sum_{a \in [n]} sep_{u'}(a) = \sum_{a \in [n] \setminus \{x,y\}} sep_u(a) + \big(sep_{u'}(x) + sep_{u'}(y)\big) \tag{1}$$

There are four different cases to consider on equation 1 based on how $(i, i+1)$ causes $sep_{u'}(x)$ and $sep_{u'}(y)$ to differ from $sep_x(u)$ and $sep_y(u)$. We enumerate the cases as follows:

1. $sep_{u'}(x) = sep_u(x) - 1$ and $sep_{u'}(y) = sep_u(y) - 1$
2. $sep_{u'}(x) = sep_u(x) - 1$ and $sep_{u'}(y) = sep_u(y) + 1$
3. $sep_{u'}(x) = sep_u(x) + 1$ and $sep_{u'}(y) = sep_u(y) - 1$
4. $sep_{u'}(x) = sep_u(x) + 1$ and $sep_{u'}(y) = sep_u(y) + 1$

In each of these cases, we have $sep(u') = sep(u) + k$, $k \in \{-2, 0, 2\}$, in which case $sep(u')$ is even. Transpositions of the form $(i, i+1)$ generate S_{2n}, so σu can be expressed in the form $\sigma_n \dots \sigma_1 u$, with each of the σ_j's a transposition of the form $(i, i+1)$ for some i, in which case $sep(u)$ is even implies $sep(\sigma u)$ is even. The group action of S_{2n} clearly is transitive, so in fact the separation of each double-occurrence word over $[n]$ is even. $\square$

Remark 6. If u is not a double-occurrence word the conclusion does not necessarily hold. (Take $u = 121$.) In particular, factors of a double-occurrence word u don't necessarily have even separations (unless they are double-occurrence words themselves). If u is a DOW, the number of letters $a \in u$ which have odd separation is even.

Definition 2.3. *Let $u \in \Sigma_{DOW}$. The word u is called a* ***permutation word*** *if*

$$u \sim 1 \cdots n \sigma(1) \cdots \sigma(n)$$

for some $\sigma \in S_n$.

Remark 7. The number of permutation words of size n up to ascending order equivalence is given by $n!$.

Example 5. The DOW $u = 1234512345$ is a permutation word and so is $v = 12344213$. Repeat and return words are special examples of permutation words.

Definition 2.4. *Let $w \in \Sigma_{DOW}$. The* ***index mapping*** *of w is a 2-tuple (I_1, I_2) where $I_j : \Sigma[w] \to \mathbb{Z}^+$ (for $j = 1, 2$) is a map where $I_j(a)$ is the position of the jth occurrence of a in w.*

Index mappings give us a convenient way to calculate the separation of a DOW as follows: Let $u \in \Sigma_{DOW}$ have size n, and let (I_1, I_2) its associated index mapping. Then,

$$\begin{aligned}
sep(u) &= \Big(\sum_{a\in\Sigma[u]} I_2(a)\Big) - \Big(\sum_{a\in\Sigma[u]} I_1(a)\Big) - n \\
&= \Big(\sum_{a\in\Sigma[u]} I_2(a)\Big) + \Big(\sum_{a\in\Sigma[u]} I_1(a)\Big) - 2\Big(\sum_{a\in\Sigma[u]} I_1(a)\Big) - n \\
&= 2n^2 + n - 2\sum_{a\in\Sigma[u]} I_1(a) - n \\
&= 2\Big(n^2 - \sum_{a\in\Sigma[u]} I_1(a)\Big)
\end{aligned}$$

We are now ready to give a result which gives an upper bound on the separation of an arbitrary DOW and also completely characterizes the structure of DOWs that achieve the upper bound on separation.

Proposition 2.5. *Let $u \in \Sigma_{DOW}$ and $|u| = 2n$. Then, $sep(u) \leq n(n-1)$. In particular, $sep(u) = n(n-1)$ if and only if u is a permutation word, hence there are $n!$ assembly words of size n and separation $n(n-1)$.*

Proof. Let (I_1, I_2) be the index mapping of u. Previous remarks show that

$$sep(u) = 2\Big(n^2 - \sum_{a\in\Sigma[u]} I_1(a)\Big).$$

$sep(u)$ is maximized when $\sum_{a\in\Sigma[u]} I_1(a)$ is minimized, which occurs if and only if $I_1(a) < n+1$ for all $a \in \Sigma[u]$. This immediately implies

$$\begin{aligned}
sep(u) &\leq 2\Big(n^2 - n(n+1)/2\Big) \\
sep(u) &\leq n^2 - n
\end{aligned}$$

with equality holding if and only if u is a permutation word. □

With the above result, we have classified size n DOWs of separation $n(n-1)$, but We would like to count and classify size n DOWs of separation k for any even $k < n(n-1)$. The next result shows that we can always find a DOW of separation k.

Proposition 2.6. *Let $n \in \mathbb{N}$. Then, for every even number $0 \leq k \leq n(n-1)$, there exists $u \in [n]_{DOW}$ such that $sep(u) = k$.*

Proof. Proceed by induction on n. The result obviously holds for $n = 1$. Suppose the result holds for all $m < n$. If $v \in [n-1]_{DOW}$ is a DOW of separation k, then the DOW v' obtained from v by affixing nn at the end is a DOW in $[n]_{DOW}$ with separation k. Thus, for every $0 \leq k \leq (n-1)(n-2)$, there exists $u \in [n]_{DOW}$ such that $sep(u) = k$. Now, let $v = v_1 \cdots v_{2(n-1)}$ be a repeat word of size $n-1$ (so $v \in [n-1]_{DOW}$). Then, the DOW

$$v' = v_1 \cdots v_{n-1} n v_n \cdots v_{n+k-1} n v_{n+k} \cdots v_{2(n-1)}$$

is in $[n]_{DOW}$ and has separation $(n-1)(n-2) + 2(k+1)$. Let S be the set of all possible separation values v' could have. Then,

$$S = \{(n-1)(n-2) + 2, (n-1)(n-2) + 4, \ldots, (n-1)(n-2) + 2n\}$$

The result follows. □

3 Irreducibility and Separation

In this section, we define irreducible DOWs and show how they play an important role in determining the separation of a DOW. By developing the notion of a decomposition of a DOW into its irreducible components, we classify DOWs of separation 2 and 4. Every DOW of size $n \geq 3$ and separation $n(n-1) - 2$ is shown to be irreducible. Finally, a conjecture is given on the maximum size of an irreducible DOW of fixed separation.

Definition 3.1. *Let $u \in \Sigma_{DOW}$. If $u = vw$ for some choice of DOW factors v and w, then u is called **reducible**. Otherwise, it is **irreducible**. If u has no DOW factors, then it is **strongly irreducible**.*

Definition 3.2. *Let $u \in \Sigma_{DOW}$. A **decomposition** of u is an n-tuple $(u_1, \ldots, u_n)$ of irreducible DOWs such that $u = u_1 \cdots u_n$.*

Remark 8. As a slight abuse of notation, throughout the rest of this paper, a decomposition $(u_1, \ldots, u_n)$ of u will be written simply as $u_1 \cdots u_n$.

Proposition 3.3. *For every $u \in \Sigma_{DOW}$ there exists a unique decomposition of u.*

Proof. It is obvious that such a decomposition always exists. We now prove the decomposition is unique. Suppose on the contrary $u_1 \cdots u_n$ and $v_1 \cdots v_m$ are two different decompositions of u and let i be the smallest index such that $u_i \neq v_i$. Then, either $|u_i| > |v_i|$ or $|u_i| < |v_i|$. WLOG assume $|u_i| > |v_i|$. Then, $u_i = v_i k$ for some nonempty factor k of u. Since v_i is a DOW, k must be a DOW. This implies u_i is reducible, a contradiction. The result follows. □

Remark 9. Let $u \in \Sigma_{DOW}$ and let $u_1 \cdots u_k$ be the decomposition of u. Then, the following observations are immediate:

1. The set $\{u_i \mid i \in [k]\}$ is precisely the set of all irreducible factors of u.
2. The separation of u is given by $sep(u) = \sum_{i=1}^{k} sep(u_i)$.

3. Given $m, n \in \mathbb{N}$, we have $m(m-1) + n(n-1) \leq (m+n)(m+n-1)$, hence the maximum separation of u is $(n-k+1)(n-k)$.

4. $sep(u) = (n-k+1)(n-k)$ if and only if one of u's irreducible factors is the permutation word of size $n-k$ and the rest are each ascending order equivalent to 11.

5. Every size n DOW of separation $(n-1)(n-2) < k \leq n(n-1)$ is irreducible.

In light of the previous proposition, the problem of classifying DOWs of a given separation k boils down to determining irreducible DOWs up to separation k. As long as we know what the irreducible DOWs with separation at most k look like, we can determine precisely what all DOWs with separation up to k look like. Of course, the only irreducible DOWs of separation 2 up to ascending order equivlance are 1221 and 1212, and so with this knowledge we can quickly give a result which counts and classifies DOWs of separation 2.

Proposition 3.4. *Let $u \in \Sigma_{DOW}$ be of size n. Then, $sep(u) = 2$ if and only if $u \sim 11 \cdots kk(v)(k+1)(k+1) \cdots (n-2)(n-2)$, where v is either a size two repeat or return word. There are $2(n-1)$ assembly words of separation 2.*

Proof. The 'if' direction is obvious. To see the 'only if' direction, let $u_1 \cdots u_n$ be the decomposition of u. Then, only one of the u_i's has separation 2. The rest have separation 0. The only irreducible DOWs of separation 2 are the size two repeat and return words. Thus, for some k, we have u_k is a size two repeat or return word, and u_i for $i \neq k$ is the size one word. The number of all assembly words of separation 2 is then counted by the number of choices to insert v into $11 \cdots (n-2)(n-2)$ at an odd index times the number of possibilities for v (up to ascending order equivalence). As there are 2 choices for v and $n-1$ choices to insert v, there must be $2(n-1)$ assembly words u of separation 2. □

Remark 10. The only strongly irreducible assembly word of separation 2 is the size 2 repeat word.

A similar observation to the one deduced above for DOWs of separation 2 can be deduced for DOWs of separation 4 after we characterize irreducible DOWs of separation 4.

Proposition 3.5. *The set of all irreducible assembly words of separation 4 is given by* $\{[122331], [121332], [122313], [121323]\}$

Proof. Let $u \in \Sigma_{DOW}$ and i a symbol that appears in u. Write $u = u_1 ixi u_2$, where x is a (possibly empty) factor of u. The factor x of u is denoted $\phi_u(i)$.

We split the proof into multiple cases based on the possible separation values each symbol in $u \in \Sigma_{DOW}$ could assume:

Case 1: One symbol has separation 4 and the rest 0: It is obvious that [122331] is the only assembly word that fits this case.

Case 2: One symbol has separation 3, one has separation 1, and the rest 0: If u is a DOW satisfying the criteria and k is the symbol of separation 1 in u, then $\phi(k)$ is a single letter j which must have separation 3. It follows that [121332] and [122313] are the only assembly words that can be formed in this case.

Case 3: Two symbols have separation 2 and the rest 0: We show that no assembly word can satisfy this case. Suppose on the contrary u is an irreducible DOW satisfying the given criteria. Let k be one of the symbols in u of separation 2. Then, $\phi(k) \not\sim 11$, for otherwise u is reducible since it takes the form $u_1v_1u_2v_2u_3$ with v_1 and v_2 size 2 return words and the u_i's each (possibly empty) separation 0 words. $\phi(k) \sim 12$ implies, however, that v has three symbols of separation at least 1, which is a contradiction.

Case 4: One symbol has separation 2, two have separation 1, and the rest separation 0: Let u be an irreducible DOW satisfying the given criteria. Let k be the symbol in u of separation 2. We claim $\phi(k) \not\sim 11$. To see this, suppose otherwise. Then, $k\phi(k)k$ is a DOW factor of u. If h, j are the two symbols of u with separation 1, then $\phi(h) = j$, so either $hjhj$ or $jhjh$ is also a DOW factor of u. WLOG we may assume the former case. Then, $u = u_1(hjhj)u_2(k\phi(k)k)u_3$, where the u_i's are each separation 0 words. This implies u is reducible, a contradiction, and the claim is proven. The claim immediately implies both symbols of separation 1 appear exactly once in $\phi(k)$. It follows immediately that $u \sim 121323$.

Case 5: Four symbols have separation 1 and the rest separation 0: We claim no u satisfies this criteria. Suppose on the contrary such a u does exist. If k is a symbol in u of separation 1, then $\phi(k) = j$ for some letter j in u also of separation 1. It follows that u can be written $u = u_1v_1u_2v_2u_3$ with v_1 and v_2 size 2 repeat words and the u_i's separation 0 DOWs. This implies u is reducible and yields a contradiction. □

Remark 11. The only strongly irreducible assembly word of separation 4 is [121323]. Though every irreducible DOW of separation 4 is of size 3, it is not the case in general for arbitrary k that every irreducible of DOW of separation k has the same size j. Indeed, 123123 and 12133424 both are irreducible of separation 6, but one is of size 3 and the other is of size 4.

We are now ready to give a result with characterizes all assembly words of separation 4.

Proposition 3.6. *Let $u \in \Sigma_{DOW}$ be such that $sep(u) = 4$ and $size(u) = n$. Then, either of the following two cases hold:*

1. *u has two irreducible factors of separation 2, in which case $u = u_1v_1u_2v_2u_3$ where the u_i's are each DOWs of separation 0 and the v_i's are each either repeat or return words of size 2. The number of assembly words of size n and separation 2 that are of this form is $2\binom{n-2}{2} + (n-3)^2 + (n-3) = 2(n-3)(n-2)$.*

2. *u has one irreducible factor v of separation 4, in which case $u = u_1vu_2$, where the u_i's are each DOWs of separation 0. The number of assembly words of size n and separation 4 that are of this form is $4(n-2)$.*

Consequently, the number of assembly words of size n and separation 4 is $2(n-1)(n-2)$.

Remark 12. If $u \in \Sigma_{DOW}$ is of separation 4, then it is of size 3 if and only if it is irreducible.

Proof. By lemma 2.2, either u has two irreducible factors of separation 2 or one irreducible factor of separation 4.

Case 1: u has two irreducible factors of separation 2: In this case we have

$$u = u_1 \cdots u_{i-1}v_1u_i \cdots u_{j-1}v_2u_j \cdots u_{n-4}$$

where the u_i's are ascending order equivalent to 11 and the v_i's are each either repeat or return words of size 2. There are three subcases to consider. If both the v_i's are repeat words, then there are $\binom{n-2}{2}$ possibilities for u up to ascending order equivalence. If both the v_i's are return words, then there are also $\binom{n-2}{2}$ possibilities for u up to ascending order equivalence. We claim if one v_i is a repeat word and the other a return word that there are $(n-3)^2+n-3$ possibilities up to ascending order equivalence. To see this, let S be the set of all assembly words $[u]$ such that u is of size n, separation 4, and contains two irreducible factors v_1, v_2 of separation 2 such that v_1 (resp. v_2) is a repeat (resp. return) word. Then, S is a well-defined set, and the claim follows if we show $|S| = (n-3)^2 + (n-3)$. Now, each 2-tuple (i, j) with both entries in the set $[n-3]$ can be identified with a unique assembly word $[u] \in S$ such that

$$u \sim u_1 \cdots u_{i-1} v_1 u_i \cdots u_{j-1} v_2 u_j \cdots u_{n-4}$$

with

$$u \sim u_1 \cdots u_{i-1} v_1 v_2 u_i \cdots u_{n-4}$$

if $i = j$. This counts $(n-3)^2$ different assembly words in S and in fact counts every assembly word in S except the those such that

$$u \sim u_1 \cdots u_{i-1} v_2 v_1 u_i \cdots u_{n-4}$$

of which there are $n-3$ such words. The claim follows.

Case 2: u has one irreducible factor v of separation 4: We have $u = u_1 \cdots u_{i-1} v u_i \cdots u_{n-3}$, where $u_i \sim 11$ for all i, and $[v] \in \{[122331], [121332], [122313], [121323]\}$ (see the previous proposition). There are $n-2$ choices then for the location of v and hence $4(n-2)$ possibilities for u up to ascending order equivalence in this case. □

The classification of DOWs of arbitrary separation is a daunting task since classifying irreducible DOWs of arbitrary separation is difficult. We turn to a quick observation which may become useful later on for classifying arbitrary DOWs of separation $n(n-1)-2$:

Proposition 3.7. *Let $u \in \Sigma_{DOW}$ be of size n and of separation $n(n-1)-2$. Then, u is irreducible if and only if $n \geq 3$.*

Proof. ($\Rightarrow$): If $n \leq 2$, then the only possibility for u is that u is of size 2 and separation 0, whence u is reducible.

($\Leftarrow$): For simplicity we may assume u has decomposition $u_1 u_2$ (The proof for when u has more irreducible factors is similar). Then, the remarks following proposition 3.3 show that $sep(u) = n(n-1) - 2 \leq (n-1)(n-2)$. The reader may verify quickly that $n \leq 2$. □

Remark 13. Let u be a permutation word. Then, if $\sigma \in S_{2n}$ is a transposition of the form $(i, i+1)$, the above proposition shows that it is always the case that σu is irreducible.

In a previous section we answered the question of the maximum possible separation value that is attainable by an arbitrary DOW of size n. Naturally, given our efforts to classify and count the number of of assembly words of size n and separation k, the similar question arises of what the maximum possible size of an arbitrary DOW of size k. We end this section with a conjecture on the maximum size of an irreducible DOW of separation k. The conjecture can be shown to be true all positive even k up to 6 by exhaustion.

Conjecture 3.8. *The maximum size of a irreducible double-occurrence word* u *of separation* k *is* $j = \frac{k}{2} + 1$.

There is reason to believe this conjecture is true. Let $u \in \mathbb{N}$ be the unique DOW of size j such that $sep_u(1), sep_u(j) = 1$ and $sep_u(k) = 2$ for $1 < k < j$. Then, u is called the tangled cord of size j. For example, the tangled cord of size 5 is 1213243545. If u is the size j tangled cord, then u has separation $2(j-1)$ and is irreducible. In fact, u is strongly irreducible. The only way to increase the size of u without increasing its separation value is by affixing or prefixing u with a DOW $v \sim 11$, and neither vu nor uv are irreducible. Of course, we would like to show in general that every DOW of separation $2(j-1)$ has size at most j.

References: Patterns in words were studied in [4], while the model for DNA rearrangement and introduction of the assembly graphs and asssembly weords was done in [1]. Further studies of the properties of assembly words and DOWs can be found in [3]. Early models on DNA rearrangements are compiled in the book [2].

Acknowledgement: This work has been supported in part by the NSF grants CCF-1526485 and NIH grant R01 GM109459.

References

[1] Angela Angeleska, Nataša Jonoska, and Masahico Saito. Dna recombination through assembly graphs. *Discrete Applied Mathematics*, 157(14):3020 – 3037, 2009.

[2] Andrzej Ehrenfeucht, Tero Harju, and Ion Petre. *Computation in Living Cells: Gene Assembly in Ciliates (Natural Computing Series).* SpringerVerlag, 2004.

[3] Burns; Muche. Counting irreducible double occurrence words. *ArXiv*, 15, May 2011.

[4] Lukas Nabergall, Masahico Saito, and Natasha Jonoska. Patterns and distances in words related to dna rearrangement. *Fundamenta Informaticae*, 154:225–238, 08 2017.