Graph Connectivity, Partial Words, and a Theorem of Fine and Wilf *

F. Blanchet-Sadri[1] Deepak Bal[2] Gautam Sisodia[3]

October 23, 2007

Abstract

The problem of computing periods in *words*, or finite sequences of symbols from a finite alphabet, has important applications in several areas including data compression, string searching and pattern matching algorithms. The notion of *period* of a word is central in combinatorics on words. There are many fundamental results on periods of words. Among them is the well known and basic periodicity result of Fine and Wilf which intuitively determines how far two periodic events have to match in order to guarantee a common period. More precisely, any word with length at least $p + q - \gcd(p, q)$ having periods p and q has also period the greatest common divisor of p and q, $\gcd(p, q)$. Moreover, the bound $p + q - \gcd(p, q)$ is optimal since counterexamples can be provided for words of smaller length.

Partial words, or finite sequences that may contain a number of "do not know" symbols or *holes*, appear in natural ways in several areas of current interest such as molecular biology, data communication, DNA computing, etc. Any long enough partial word with h holes and having periods p, q has also period $\gcd(p, q)$. In this paper, we give closed formulas for the optimal bounds $L(h, p, q)$ in the case where $p = 2$ and also in the case where q is large. In addition, we give upper bounds when q is small and $h = 3, 4, 5, 6$ or 7. No closed formulas for $L(h, p, q)$ were known except for the cases where $h = 0, 1$ or 2. Our proofs are based on connectivity in graphs associated with partial words. A World Wide Web server interface has been established at

`www.uncg.edu/mat/research/finewilf3`

for automated use of a program which given a number of holes h and two periods p and q, computes the optimal bound $L(h, p, q)$ and an optimal word for that bound (a partial word u with h holes of length $L(h, p, q) - 1$ is optimal if p and q are periods of u but $\gcd(p, q)$ is not a period of u).

Keywords: Words; Partial Words; Fine and Wilf's Theorem; Periods; Graph Connectivity.

1 Introduction

The problem of computing periods in *words*, or finite sequences of symbols from a finite alphabet, has important applications in several areas including data compression, string searching and pattern matching algorithms. The notion of *period* of a word is central in combinatorics on words. There are many fundamental results on periods of words. Among them is the well known periodicity result of Fine and Wilf [28] which intuitively determines how far two periodic events have to match in order to guarantee a common period. More precisely, any word with length at least $p + q - \gcd(p, q)$ having periods p and q has also

*This material is based upon work supported by the National Science Foundation under Grant No. DMS–0452020. We thank the referee of a preliminary version of this paper for his/her very valuable comments and suggestions.

[1]Department of Computer Science, University of North Carolina, P.O. Box 26170, Greensboro, North Carolina 27402–6170, USA, blanchet@uncg.edu.

[2]Department of Mathematics, The Ohio State University, 100 Mathematics Building, 231 West 18th Avenue, Columbus, Ohio 43210–1174, USA.

[3]Department of Mathematics, University of Washington, P.O. Box 354350, Seattle, WA 98195-4350, USA.

period the greatest common divisor of p and q, $\gcd(p, q)$. Moreover, the bound $p + q - \gcd(p, q)$ is optimal since counterexamples can be provided for words of smaller length. This result has been generalized in many ways. For instance, extension to more than two periods are given in [17, 19, 32, 43]. In particular, Constantinescu and Ilie [19] give an extension of Fine and Wilf's theorem for an arbitrary number of periods and prove that their bounds are optimal.

Partial words, or finite sequences that may contain a number of "do not know" symbols or *holes*, appear in natural ways in several areas of current interest such as molecular biology, data communication, DNA computing, etc [35]. Partial words are useful in a new generation of pattern matching algorithms that search for local similarities between sequences. In this area, they are called "spaced seeds" and a lot of work has been dedicated to their influence on the algorithms' performance [16, 27, 34, 37, 38, 39]. In the case of partial words there are two notions of periodicity: one is that of *(strong) period* and the other is that of *weak period*. The original Fine and Wilf's result has been generalized to partial words in several ways:

First, any partial word u with h holes and having weak periods p, q and length at least the so-denoted $l(h, p, q)$ has also period $\gcd(p, q)$ provided u is not (h, p, q)-*special*. This extension was done for one hole by Berstel and Boasson in their seminal paper [2] where the class of $(1, p, q)$-special partial words is empty; for two or three holes by Blanchet-Sadri and Hegstrom [12]; and for an arbitrary number of holes by Blanchet-Sadri [5]. The bounds $l(h, p, q)$ turn out to be optimal. In [14], Blanchet-Sadri, Oey and Rankin further extend these results, allowing an arbitrary number of weak periods. In addition to speciality, the concepts of intractable period sets and interference between periods play a role.

Second, any partial word u with h holes and having (strong) periods p, q and length at least the so-denoted $L(h, p, q)$ has also period $\gcd(p, q)$. The study of the bounds $L(h, p, q)$ was initiated in [40], but no closed formulas were shown except for the cases where $h = 0, 1$ or 2. In this paper, we give closed formulas for the optimal bounds $L(h, p, q)$ in the case where $p = 2$ and also in the case where q is large. We give upper bounds when q is small and $h = 3, 4, 5, 6$ or 7. We obtain results concerning optimal partial words for the bound $L(h, p, q)$ (a partial word u with h holes of length $L(h, p, q) - 1$ is optimal if p and q are periods of u but $\gcd(p, q)$ is not a period of u). In addition, we have established a World Wide Web server interface at

`www.uncg.edu/mat/research/finewilf3`

for automated use of a program which given a number of holes h and two periods p and q, computes the optimal bound $L(h, p, q)$ and an optimal partial word for that bound. Our proofs are based on connectivity in graphs associated with bounds and pairs of periods.

Fine and Wilf's extensions in the framework of partial words are summarized in the following figure:

Periods	Holes	Extended by
2 strong	0	Fine and Wilf [28]
2 weak	1	Berstel and Boasson [2]
	2–3	Blanchet-Sadri and Hegstrom [12]
	h	Blanchet-Sadri [5]
2 strong	h	Shur and Gamzova [40] Blanchet-Sadri, Bal and Sisodia (this paper)
3 strong	0	Castelli, Mignosi and Restivo [17]
n strong	0	Justin [32] Tijdeman and Zamboni [43] Contantinescu and Ilie [19]
n weak	h	Blanchet-Sadri, Oey and Rankin [14]

The contents of our paper are summarized as follows: In Section 2, we review basic concepts on partial words. In Section 3, we discuss the fundamental property of periodicity. We define the set PER_h containing optimal words with h holes of length $L(h, p, q) - 1$ for some periods p and q, and discuss their properties in the cases where $h = 0, 1$ or 2. In Section 4, we describe a way of representing partial words with periods p and q. There, we discuss connectivity in undirected graphs associated with such partial words. In Section 5, we give closed formulas for the optimal bounds $L(h, p, q)$ for the case where $p = 2$, and in Section 6, for the case where q is large. In Section 7, we obtain upper bounds for small q. Finally, Section 8 contains a few concluding remarks.

2 Preliminaries

In this section, we review basic concepts on partial words.

Fixing a nonempty finite set of letters or an *alphabet* A, finite sequences of letters are called *words* over A. The number of letters in a word u, or *length* of u, is denoted by $|u|$. The unique word of length 0, denoted by ε, is called the *empty word*. A word of length n over A can be defined by a total function $u : \{0, \ldots, n-1\} \to A$ and is usually represented as $u = a_0 a_1 \ldots a_{n-1}$ with $a_i \in A$. For any word u, $u[i..j)$ is the *factor* of u that starts at position i and ends at position $j-1$. In particular, $u[0..j)$ is the *prefix* of u of length j and we will sometimes denote it by $\mathrm{pref}_j(u)$. Similarly, $u[|u| - i..|u|)$ is the *suffix* of u of length i and we will sometimes denote it by $\mathrm{suff}_i(u)$. The set of all words over A of finite length (greater than or equal to zero) is denoted by A^*. It is a monoid under the associative operation of concatenation or product of words (ε serves as the identity) and is referred to as the *free monoid* generated by A. Similarly, the set of all nonempty words over A is denoted by A^+. It is a semigroup under the operation of concatenation of words and is referred to as the *free semigroup* generated by A.

A *partial word* u of length n over A is a partial function $u : \{0, \ldots, n-1\} \to A$. For $0 \leq i < n$, if $u(i)$ is defined, then we say that i belongs to the *domain* of u, denoted by $i \in D(u)$, otherwise we say that i belongs to the *set of holes* of u, denoted by $i \in H(u)$. A (*full*) word over A is a partial word over A with an empty set of holes.

For convenience, we will refer to a partial word over A as a word over the enlarged alphabet $A_\diamond = A \cup \{\diamond\}$, where $\diamond \notin A$ represents a "do not know" symbol. So a partial word u of length n over A can be viewed as a total function $u : \{0, \ldots, n-1\} \to A_\diamond$ where $u(i) = \diamond$ whenever $i \in H(u)$. For example, $u = abb\diamond b\diamond cbb$ is a partial word of length 9 where $D(u) = \{0, 1, 2, 4, 6, 7, 8\}$ and $H(u) = \{3, 5\}$. We can thus define for partial words concepts such as concatenation, etc. in a trivial way.

The length of a partial word u over A is denoted by $|u|$, while the set of distinct letters of A occurring in u is denoted by $\alpha(u)$. We denote the reverse of u as $\mathrm{rev}(u)$. For the set of all partial words over A with an arbitrary number of holes we write $A_\diamond^*$. The set $A_\diamond^*$ is a monoid under the operation of concatenation where ε serves as the identity. If $X \subset A_\diamond^*$, then the *cardinality* of X is denoted by $\| X \|$. For example, if $u = abab\diamond cbca$, then $|u| = 9$, $\alpha(u) = \{a, b, c\}$, and $\| \alpha(u) \| = 3$. For partial words, we use the same notions of prefix, suffix and factor as for full ones.

If u and v are two partial words of equal length, then we say that u is *contained in* v, denoted by $u \subset v$, if all elements in $D(u)$ are in $D(v)$ and $u(i) = v(i)$ for all $i \in D(u)$. Partial words u and v are *compatible*, denoted by $u \uparrow v$, if there exists a partial word w such that $u \subset w$ and $v \subset w$. In other words, $u(i) = v(i)$ for every $i \in D(u) \cap D(v)$. Note that for full words, the notion of compatibility is simply that of equality.

3 Periodicity

In this section, we discuss the fundamental property of periodicity. A *(strong) period* of a partial word u over A is a positive integer p such that $u(i) = u(j)$ whenever $i, j \in D(u)$ and $i \equiv j \bmod p$. In such a case, we call u *p-periodic*. Another notion of periodicity for partial words is that of weak periodicity: a *weak period* of u is a positive integer p such that $u(i) = u(i + p)$ whenever $i, i + p \in D(u)$. In such a case, we call u *weakly p-periodic*. The partial word $abb\diamond bbcbb$ is weakly 3-periodic but is not 3-periodic. In this paper we deal with periods, not weak periods.

First we introduce the concept of what we refer to as *general* or *functional (partial) words*. These words will be the ones that we are concerned with throughout this paper. The *primary general word* u of a certain length and domain set $D(u)$ is the word of that length with letters unique to their position in $D(u)$. To form a *general word*, certain periods are imposed onto a primary general word. A period p is imposed by transforming the general word into a matrix with columns $0, \ldots, p-1$ which represent the congruence classes modulo p. In each column i, the letter of the first non-hole position is placed into each of the other non-hole positions of the column. To impose subsequent periods, every time a letter must be changed, all other instances of that letter throughout the word must also be changed.

Example 3.1. *Suppose we want to impose periods 4 and 7 onto a partial word of length 14 with holes in positions 2 and 11. We start with the word $ab\diamond cdefghij\diamond kl$, form it into a matrix with 4 columns and make the necessary changes:*

$$\begin{array}{cccc} a & b & \diamond & c \\ d & e & f & g \\ h & i & j & \diamond \\ k & l & & \end{array} \longrightarrow \begin{array}{cccc} a & b & \diamond & c \\ a & e & f & g \\ a & i & j & \diamond \\ a & l & & \end{array} \longrightarrow \begin{array}{cccc} a & b & \diamond & c \\ a & b & f & g \\ a & b & j & \diamond \\ a & b & & \end{array} \longrightarrow \begin{array}{cccc} a & b & \diamond & c \\ a & b & f & g \\ a & b & f & \diamond \\ a & b & & \end{array} \longrightarrow \begin{array}{cccc} a & b & \diamond & c \\ a & b & f & c \\ a & b & f & \diamond \\ a & b & & \end{array}$$

We refer to the columns of these matrices as 4-classes*. We then take the resulting word $ab\diamond cabfcabf\diamond ab$ and impose the period 7:*

$$\begin{array}{ccccccc} a & b & \diamond & c & a & b & f \\ c & a & b & f & \diamond & a & b \end{array} \longrightarrow \begin{array}{ccccccc} a & b & \diamond & a & a & b & f \\ a & a & b & f & \diamond & a & b \end{array} \longrightarrow \begin{array}{ccccccc} a & a & \diamond & a & a & a & f \\ a & a & a & f & \diamond & a & a \end{array}$$

$$\longrightarrow \begin{array}{ccccccc} a & a & \diamond & a & a & a & a \\ a & a & a & a & \diamond & a & a \end{array}$$

Each of the columns is a 7-class*. The general word with periods 4 and 7 of length 14 and holes in positions 2 and 11 is $aa\diamond aaaaaaaa\diamond aa$ (up to a renaming of letter).*

The next remark justifies the results of this paper.

Remark 3.2 ([41]). *There exists a smallest integer $L(h, p, q)$, or the* optimal bound *for periods p, q and number of holes h, such that if a partial word u with h holes has periods p, q ($1 < p < q$) and $|u| \geq L(h, p, q)$, then u has period $\gcd(p, q)$. In other words, $L(h, p, q)$ is a* lower bound *and there exists a partial word v with h holes of length $L(h, p, q) - 1$ that has periods p and q, but v does not have period $\gcd(p, q)$.*

Note that the notion of optimal bound makes sense only if $\gcd(p, q) \neq p$.

The essential question is how long the partial word u should be? Fine and Wilf's theorem [28] states that length for $h = 0$ which is $p + q - \gcd(p, q)$. While the bound $p + q - \gcd(p, q)$ is a lower bound, it has also been proved to be an upper bound and thus the optimal bound, that is, there exists a full word v of length $p + q - \gcd(p, q) - 1$ that has periods p and q, but does not have period $\gcd(p, q)$ [18, 36]. For example, the general word $aabaaabaa$ with periods 4 and 7 of length $4 + 7 - \gcd(4, 7) - 1 = 9$ does not have period 1. In the notation of Remark 3.2, $L(0, p, q) = p + q - \gcd(p, q)$. We are interested in this paper in both upper and lower bounds for the length of u when $h > 0$.

Throughout this paper we generally restrict ourselves to cases where periods p and q are co-prime, for if $\gcd(p, q) > 1$, then the problem can be reduced to a case where the two periods are co-prime. Indeed, if u is a partial word with periods p and q such that $\gcd(p, q) = d > 1$, then u can be replaced by a set of partial words $u_0, \ldots, u_{d-1}$ where $u_i = u(i)u(i+d)u(i+2d)\ldots$ has co-prime periods $\frac{p}{d}$ and $\frac{q}{d}$. So each u_i has period 1 if and only if u has period d.

Now let

$$\mathcal{W}_{h,p,q} = \{w \mid w \text{ has periods } p \text{ and } q, \parallel H(w) \parallel = h \text{ and } |w| = L(h,p,q) - 1\}$$

We call the elements of $\mathcal{W}_{h,p,q}$ *optimal*. All words v from Remark 3.2 form a subset of $\mathcal{W}_{h,p,q}$ which we denote here by $\mathcal{V}_{h,p,q}$, that is,

$$\mathcal{V}_{h,p,q} = \{v \mid v \in \mathcal{W}_{h,p,q} \text{ and } v \text{ does not have period } \gcd(p,q)\}$$

The sets PER_h and $\mathcal{V}\mathrm{PER}_h$ are defined as follows:

$$\begin{array}{rcl} \mathrm{PER}_h & = & \bigcup_{\gcd(p,q)=1} \mathcal{W}_{h,p,q} \\ \mathcal{V}\mathrm{PER}_h & = & \bigcup_{\gcd(p,q)=1} \mathcal{V}_{h,p,q} \end{array}$$

It turns out that $\mathcal{V}\mathrm{PER}_0$ has remarkable combinatorial properties [3, 20, 21, 22]. In the next three sections, we discuss properties of PER_h and $\mathcal{V}\mathrm{PER}_h$ in the cases where $h = 0, 1$ or 2.

3.1 The zero-hole case

The following result is a well known property of PER_0.

Theorem 3.3 ([18]). *The set $\mathcal{W}_{0,p,q}$ contains a unique word w (up to a renaming) such that $\|\alpha(w)\| = 2$.*

The set $\mathcal{V}\mathrm{PER}_0$ has a nice characterization, which is a recurrence relation, stated as follows.

Proposition 3.4. *Let $w \in \mathcal{V}_{0,p,q}$.*

- *If $q - p = 1$, then $w = a^{p-1}ba^{p-1}$ (up to a renaming).*
- *If $q - p > 1$, then $w = v\mathrm{suff}_{|v|-q+p+2}(v)$ where $v \in \mathcal{V}_{0,\min(p,q-p),\max(p,q-p)}$.*

Proof. First, suppose that $q - p = 1$, or $q = p + 1$. The word $u = a^{p-1}ba^{p-1}$ has periods p and $p+1$, while it does not have period 1. Also $|u| = p + q - 2$, and thus $w = u$ (up to a renaming).

Now, suppose that $q - p > 1$. We induct on the difference between the two periods p and q where $w \in \mathcal{V}_{0,p,q}$. Set $p = \min(p', q' - p')$ and $q = \max(p', q' - p')$. If $p' \leq q' - p'$, then $p = p'$ and $q = q' - p'$. Here let $u = w\mathrm{suff}_{p'}(w) = w\mathrm{suff}_p(w)$ and show that $u \in \mathcal{V}_{0,p',q'} = \mathcal{V}_{0,p,p+q}$. If $p' > q' - p'$, then $p = q' - p'$ and $q = p'$. Here let $v = w\mathrm{suff}_{p'}(w) = w\mathrm{suff}_q(w)$ and show that $v \in \mathcal{V}_{0,p',q'} = \mathcal{V}_{0,q,p+q}$.

Since $\gcd(p,q) = 1$, we have $\gcd(p, p+q) = 1$ and $\gcd(q, p+q) = 1$. We show the membership for u (the one for v follows much in the same way). Note that $|u| = p + (q+p) - 2$. Also, u does not have period 1, since w does not.

We first show that u has period p. Let $i \in [p+q-2..2p+q-2)$ be a position in u. Since both $\mathrm{suff}_p(u) = \mathrm{suff}_p(w)$ and $\mathrm{pref}_{p+q-2}(u) = w$, $u(i) = w(i-p) = u(i-p)$. Thus, each position of $\mathrm{suff}_p(u)$ belongs to the p-class of its corresponding position in w. Since w has period p, u has period p. We now show that u has period $p+q$. Since $|u| = p + (q+p) - 2 < 2(p+q)$, this is equivalent to showing that for $i \in [p+q..q+2p-2)$, $u(i) = u(i-(p+q))$. From above, we see that $u(i) = u(i-p)$. Also, since $\mathrm{pref}_{p+q-2}(u) = w$ and w is q-periodic and $i - p \in [q..q+p-2)$, we have $u(i-p) = u(i-p-q) = u(i-(p+q))$. Thus, u is $(p+q)$-periodic. □

Example 3.5. *Using this relation, we find the word with periods 9 and 13 in* $\mathcal{V}\mathrm{PER}_0$ *(here* $w_{0,p,q} \in \mathcal{V}_{0,p,q}$*):*

$$\begin{array}{ccc}
(9,13) & & \overbrace{aaabaaabaaa}^{w_{0,4,9}}\overbrace{abaaabaaa}^{w_{0,4,9}[2..11)} \\
\downarrow & & \uparrow \\
(4,9) & & \overbrace{aaabaaa}^{w_{0,4,5}}\overbrace{baaa}^{w_{0,4,5}[3..7)} \\
\downarrow & & \uparrow \\
(4,5) & \rightarrow & aaabaaa
\end{array}$$

The words of $\mathcal{V}\mathrm{PER}_0$ also have another well known property which will be used later and which we prove here for sake of completeness.

Proposition 3.6. *If* $v \in \mathcal{V}\mathrm{PER}_0$*, then* v *is a palindrome.*

Proof. This proof is similar to that of Proposition 3.4 in that we induct on the difference between the two periods p and q where $v \in \mathcal{V}_{0,p,q}$.

First, if $q = p + 1$, then $v = a^{p-1}ba^{p-1}$ which is a palindrome.

Now, assume for some periods p and q that $v \in \mathcal{V}_{0,p,q}$ is a palindrome. We must show that $u \in \mathcal{V}_{0,p,p+q}$ and $w \in \mathcal{V}_{0,q,p+q}$ are palindromes.

If $u \in \mathcal{V}_{0,p,p+q}$, then $u = v\mathrm{suff}_p(v)$. Let $u' = \mathrm{rev}(u) = \mathrm{rev}(v\mathrm{suff}_p(v)) = \mathrm{rev}(\mathrm{suff}_p(v))\mathrm{rev}(v) = \mathrm{rev}(\mathrm{suff}_p(v))v$. Now $\mathrm{rev}(\mathrm{suff}_p(v)) = \mathrm{pref}_p(u)$ because v is a palindrome and $\mathrm{pref}_p(u) = \mathrm{pref}_p(v)$. Also, $v[p..p+q-2) = v[0..q-2)$ since v is p-periodic, thus $\mathrm{suff}_{q-2}(v) = \mathrm{pref}_{q-2}(v)$. So $\mathrm{suff}_{q-2}(v)\mathrm{suff}_p(v) = \mathrm{pref}_{q-2}(v)\mathrm{suff}_p(v) = v$. Thus $u = u'$ and u is a palindrome.

Now if $w \in \mathcal{V}_{0,q,p+q}$, then $w = v\mathrm{suff}_q(v)$. Let $w' = \mathrm{rev}(w) = \mathrm{rev}(v\mathrm{suff}_q(v)) = \mathrm{rev}(\mathrm{suff}_q(v))\mathrm{rev}(v) = \mathrm{rev}(\mathrm{suff}_q(v))v$. We see that $\mathrm{rev}(v[p-2..p+q-2)) = w[0..q)$ because v is a palindrome. Also, $v[q..p+q-2) = v[0..p-2)$ since v is q-periodic, thus $\mathrm{suff}_{p-2}(v) = \mathrm{pref}_{p-2}(v)$. So

$$\mathrm{suff}_{p-2}(v)\mathrm{suff}_q(v) = \mathrm{pref}_{p-2}(v)\mathrm{suff}_q(v) = v$$

Thus $w = w'$ and w is a palindrome in this case as well. □

Corollary 3.7. *If* w *is the unique element of* $\mathcal{V}_{0,p,q}$ *of length* $p + q - 2$*, then* $\mathrm{pref}_{p-2}(w)$ *is a palindrome.*

Proof. Since w is q-periodic, $w[0..p-2) = w[q..p+q-2)$, and so $\mathrm{pref}_{p-2}(w) = \mathrm{suff}_{p-2}(w)$. Also w is a palindrome by Proposition 3.6, so

$$\mathrm{pref}_{p-2}(w) = \mathrm{rev}(\mathrm{suff}_{p-2}(w)) = \mathrm{rev}(\mathrm{pref}_{p-2}(w))$$

Hence $\mathrm{pref}_{p-2}(w)$ is a palindrome. □

3.2 The one-hole case

We now turn our attention to the case of partial words with one hole. We start off with a theorem which gives the optimal bound for such partial words.

Theorem 3.8 ([2]). *The equality* $L(1, p, q) = p + q$ *holds.*

Before we give our result concerning partial words with one hole, we need a definition.

Definition 3.9. *Let p, q, and r be integers satisfying $1 < p < q$, $\gcd(p, q) = 1$, and $0 \leq r < p + q - 1$. For $i \neq q - 1$ and $0 \leq i < p + q - 1$, we define the sequence of i relative to p, q, and r as $\mathrm{seq}_{p,q,r}(i) = (i_0, i_1, i_2, \ldots, i_{n-1}, i_n)$ where $i_0 = i$ and*

- *if $r \in \{p - 1, \ldots, q - 1\}$, then $i_n = q - 1$*
- *if $i = r$, then $i_n = q - 1$*
- *if $i \neq r$ and $r \notin \{p - 1, \ldots, q - 1\}$, then $i_n = r$ or $i_n = q - 1$ (whichever comes first)*
- *if $r \notin \{p - 1, \ldots, q - 1\}$, then for $1 \leq j < n$, $i_j \neq r$*
- *for $1 \leq j < n$, $i_j \notin \{i, q - 1\}$*
- *for $1 \leq j \leq n, i_j$ is defined as*

$$i_j = \begin{cases} i_{j-1} + p & \text{if } i_{j-1} < q - 1 \\ i_{j-1} - q & \text{if } i_{j-1} > q - 1 \end{cases}$$

We define $\mathrm{seq}_{p,q,r}(q - 1) = (q - 1)$.

The sequence $\mathrm{seq}_{p,q,r}(i)$ gives a way of visiting elements of $\{0, \ldots, p+q-2\}$ starting at i. For example, if $p = 4$, $q = 11$, and $r = 5$, then

$$\mathrm{seq}_{4,11,5}(3) = (3, 7, 11, 0, 4, 8, 12, 1, 5, 9, 13, 2, 6, 10)$$

Notice that $r \in \{p-1, \ldots, q-1\}$ and we have that all sequences are suffixes of this longest sequence. Now consider the example where $p = 4$, $q = 11$, and $r = 2$. Here $r \notin \{p - 1, \ldots, q - 1\}$ and we have

$$\begin{aligned} \mathrm{seq}_{4,11,2}(2) &= (2, 6, 10) \\ \mathrm{seq}_{4,11,2}(3) &= (3, 7, 11, 0, 4, 8, 12, 1, 5, 9, 13, 2) \end{aligned}$$

Note that $\mathrm{seq}_{4,11,2}(3)$ is the longest sequence ending with 2 and $\mathrm{seq}_{4,11,2}(2)$ is the longest sequence ending in 10. All other sequences are suffixes of these two.

The following theorem gives a uniqueness result for each $\mathcal{W}_{1,p,q}$ with co-prime p, q.

Theorem 3.10. *1. Given a singleton set H satisfying $H \subset \{0, \ldots, p + q - 2\} \setminus \{p - 1, \ldots, q - 1\}$, $\mathcal{W}_{1,p,q}$ contains a unique general word u (up to a renaming) such that $\|\alpha(u)\| = 2$ and $H(u) = H$.*

2. Given a singleton set H satisfying $H \subset \{p - 1, \ldots, q - 1\}$, $\mathcal{W}_{1,p,q}$ contains a unique general word u such that $\|\alpha(u)\| = 1$ and $H(u) = H$.

Proof. The proof is similar to the one of a result in [4]. Let u be a partial word with one hole of length $p + q - 1$ having periods p and q. Set $q = mp + r$ where $0 < r < p$ and $p = nr + s$ where $0 \leq s < r$. Let $ip + j$ where $0 \leq j < p$ be the hole. The proof is divided into six cases: (1) $r = 1$ and $j = p - 1$; (2) $r = 1$ and $0 \leq j < p - 1$; (3) $r \neq 1$ and $j = p - 1$; (4) $r \neq 1$ and $0 \leq j < p - 1$ and $j - r = -1$; (5) $r \neq 1$ and $0 \leq j < p - 1$ and $j - r > -1$; and (6) $r \neq 1$ and $0 \leq j < p - 1$ and $j - r < -1$.

We treat here the fifth case, the others being handled similarly. Note that $s \neq 0$ (otherwise, $\gcd(p, q) \neq 1$ since $p = nr$ and $q = mnr + r$). Also, $\gcd(r, s) = 1$ and $i \leq m$. In addition, if $s \neq 1$, then s is not a factor of r. Set $j = n'r + s'$ where $0 \leq s' < r$. Then $\text{seq}_{p,q,ip+j}(ip + j) =$

$$
\begin{aligned}
&(ip + j, (i + 1)p + j, \ldots, (m - 1)p + j, q + j - r,\\
&j - r, p + j - r, \ldots, (m - 1)p + j - r, q + j - 2r, \ldots,\\
&j - n'r, p + s', \ldots, mp + s', q + p + s' - r,\\
&p + s' - r, 2p + s' - r, \ldots, mp + s' - r, q + p + s' - 2r, \ldots,\\
&p + s' - (n - 1)r, 2p + s' - (n - 1)r, \ldots,\\
&mp + s' - (n - 1)r, q + s + s', s + s', \ldots,\\
&(s + s') \bmod r, \ldots, p + (s + s') \bmod r - r, \ldots,\\
&(2s + s') \bmod r, \ldots, p + ((N - 1)s + s') \bmod r - r, \ldots,\\
&(Ns + s') \bmod r, p + r - 1, \ldots, (m - 1)p + r - 1, q - 1)
\end{aligned}
$$

where $(Ns+s') \bmod r = r-1$ for some $0 \leq N < r$. If $ip+j \in \{p-1, \ldots, q-1\}$, then $\text{seq}_{p,q,ip+j}(p-1) = \text{seq}_{p,q,ip+j}(p + (Ns + s') \bmod r - r) =$

$$
\begin{aligned}
&(p + (Ns + s') \bmod r - r, \ldots,\\
&((N + 1)s + s') \bmod r, \ldots, p + ((N + 1)s + s') \bmod r - r, \ldots,\\
&((r - 2)s + s') \bmod r, \ldots, p + ((r - 2)s + s') \bmod r - r, \ldots,\\
&((r - 1)s + s') \bmod r, \ldots,\\
&j + (n - n')r, p + j + (n - n')r, \ldots,\\
&(m - 1)p + j + (n - n')r, q + j + (n - n' - 1)r,\\
&j + (n - n' - 1)r, \ldots,\\
&j + r, p + j + r, \ldots, (m - 1)p + j + r, q + j,\\
&j, p + j, \ldots, ip + j, \ldots)
\end{aligned}
$$

Here, we have that u is unary. If $ip + j \notin \{p - 1, \ldots, q - 1\}$, then $\text{seq}_{p,q,ip+j}(p - 1)$ ends at $ip + j$. So we can set all the letters of the first sequence (except $ip + j$) to say, a's and all the letters of the second sequence (except $ip + j$) to say, b's and see that u is binary. □

Corollary 3.11. *If $u \in \mathcal{V}_{1,p,q}$, then $H(u) \subset \{0, \ldots, p + q - 2\} \setminus \{p - 1, \ldots, q - 1\}$. Thus, the set $\mathcal{V}_{1,p,q}$ contains $2(p - 1)$ partial words (up to a renaming).*

In other words, there are $2(p - 1)$ partial words (up to a renaming) with one hole of length $p + q - 1$ having periods p and q but not period $\gcd(p, q)$.

3.3 The two-hole case

The case of two holes is stated in the following result.

Theorem 3.12 ([40]). *The equality $L(2, p, q) = 2p + q - \gcd(p, q)$ holds.*

The following is a conjecture about $\mathcal{V}\text{PER}_2$.

Conjecture 3.13. *The membership $u \in \mathcal{V}_{2,p,q}$ holds if and only if*

- *$H(u) = \{p - 2, p - 1\}$ or $H(u) = \{q + p - 1, q + p - 2\}$ or $H(u) = \{p - 2, q + p - 1\}$ when $q - p = 1$.*
- *$H(u) = \{p - 2, p - 1\}$ or $H(u) = \{q + p - 1, q + p - 2\}$ or $H(u) = \{p - 2, q + p - 1\}$ or $H(u) = \{p - 1, q + p - 2\}$ when $q - p > 1$.*

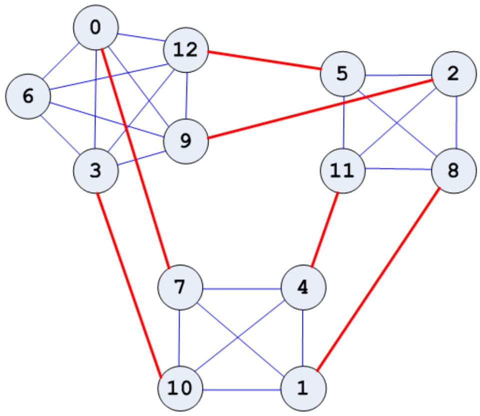

Figure 1: A graph representing a word

4 Representation of partial words

A representation of a partial word u with periods p and q is as an undirected graph $G_{(p,q)}(u) = (V, E)$ defined as follows:

- The vertex set V is $\{0, \ldots, |u| - 1\}$, each vertex representing a position of u.
- The edge set E is $E_p \cup E_q$ where

$$\begin{aligned} E_p &= \{\{i, i + np\} \mid n > 0 \text{ and } i, i + np \in V\} \\ E_q &= \{\{i, i + nq\} \mid n > 0 \text{ and } i, i + nq \in V\} \end{aligned}$$

When we refer to the i^{th} p*-class* (respectively, i^{th} q*-class*), we mean the complete subgraph of $G_{(p,q)}(u)$ consisting of exactly the members of the i^{th} residual class modulo p (respectively, q). We refer to as p*-connections* the edges due to the period p, and q*-connections* the edges due to the period q. An edge is both a p-connection and a q-connection if the positions in the word corresponding to the vertices that it connects are a common multiple of p, q apart.

Example 4.1. *The graph pictured in Figure 1 represents a word of length 13 with periods $p = 3$ and $q = 7$. The p-connections are pictured in thinner lines than the q-connections.*

We say a graph is κ*-connected* or has κ*-connectivity* if it can be disconnected with a suitable choice of κ vertex removals, but cannot be disconnected by any choice of $\kappa - 1$ vertex removals. The graph of Figure 1 is 4-connected.

We are interested in these associated graphs because they provide a way to rephrase our problem in terms of the connectedness of a graph: a word has period 1 if its graph is connected. Similarly, if the removal of vertices corresponding to hole positions results in a graph with multiple connected components, then the graph does not have period 1.

We end this section with a generalization of $G_{(p,q)}(u)$. We give a few definitions that help us formalize some of the graph theoretical arguments that will appear in some of our proofs.

Definition 4.2. *The (p, q)-periodic graph of size l is the graph $G = (V, E)$ with bijection $f : \{0, 1, \ldots, l - 1\} \to V$ such that*

$$\{f(i), f(j)\} \in E \text{ if and only if } i \equiv j \bmod p \text{ or } i \equiv j \bmod q$$

For a vertex v in V, $f^{-1}(v)$ will be referred to as the index of v, and f as the indexing function of G.

Thus, the (p, q)-periodic graph of size l can be thought to represent a word of length l with periods p and q, with the vertices corresponding to positions of the word, and the edges corresponding to equalities between indices' letters of the word forced by one of the periods. Therefore, if the (p, q)-periodic graph of size l is connected, then a word of length l with periods p and q is 1-periodic (when $\gcd(p, q) = 1$). Indeed, there exists a path (a chain of equalities) between every pair of vertices, thus each position's letter in the word must be equal to every other position's letter.

Definition 4.3. *A vertex cut of $G = (V, E)$ is a subset of V, V', such that the removal of V' disconnects G. A k-vertex cut is a vertex cut of size k. If G has at least one pair of distinct nonadjacent vertices, the connectivity $\kappa(G)$ of G is the minimum k such that G has a k-vertex cut; otherwise, $\kappa(G) = \| V \| - 1$.*

Note that a hole in a partial word u of length l with periods p and q corresponds to the removal of the associated vertex from the (p, q)-periodic graph of size l. Thus our search for $L(h, p, q)$ (when $\gcd(p, q) = 1$) can be restated in terms of connectivity: $L(h, p, q)$ is the smallest l such that the (p, q)-periodic graph of length l has connectivity at least $h + 1$. Referring to Figure 1, $L(3, 3, 7) = 13$ since 13 is the smallest l such that the $(3, 7)$-periodic graph of length l has connectivity at least 4.

Definition 4.4. *Let $G = (V, E)$ be the (p, q)-periodic graph of size l with indexing function f. Then the p-class of vertex $f(i)$ is the subset of V, V_i, such that $V_i = \{v \in V : f^{-1}(v) \equiv i \bmod p\}$. A p-connection is an edge $\{v_1, v_2\} \in E$ such that $f^{-1}(v_1) \equiv f^{-1}(v_2) \bmod p$. If an edge $\{v_1, v_2\}$ is a p-connection, then v_1 and v_2 are considered p-connected. Similar statements hold for q-classes and q-connections.*

To illustrate the above definitions, take $p = 3, q = 7$ and $l = 13$. Then the $(3, 7)$-periodic graph of size 13 is $G = (V, E)$ where $V = \{u_0, \ldots, u_{12}\}$ and where the indexing function f of G is given by the bijection $f : \{0, 1, \ldots, 12\} \to V$ with $i \mapsto u_i$. Here $\{u_i, u_j\} \in E$ if and only if $i \equiv j \bmod 3$ or $i \equiv j \bmod 7$. For example, $\{u_4, u_7\} \in E$ since $4 \equiv 7 \bmod 3$. The 3-class of vertex u_5 is $\{u_2, u_5, u_8, u_{11}\}$. The edge $\{u_5, u_8\} \in E$ is an example of a 3-connection since $f^{-1}(u_5) = 5 \equiv 8 = f^{-1}(u_8) \bmod 3$, and so u_5 and u_8 are 3-connected. The set $V' = \{u_1, u_4, u_9, u_{12}\}$ is a 4-vertex cut of G. It can be checked that G has no vertex cut of size smaller than 4, and so the connectivity $\kappa(G)$ of G is 4. Note that V' does not contain a vertex with only 3-connections.

We end this section with a lemma that proves the intuitive idea that a minimum vertex cut of a (p, q)-periodic graph does not contain a vertex with only p-connections. Using this lemma, we can give a new proof of Corollary 3.11.

Lemma 4.5. *Assume that the (p, q)-periodic graph of size l, denoted $G = (V, E)$, has a k-vertex cut $V' \subset V$. If $v \in V'$ has no q-connections, then $V' \setminus \{v\}$ is a $(k - 1)$-vertex cut of G.*

Proof. Suppose V' is a k-vertex cut of G as stated in the lemma. Then $V = V' \cup V_1 \cup V_2$ where V_1 and V_2 are disjoint and nonempty sets, and $v_1 \in V_1$ and $v_2 \in V_2$ imply $\{v_1, v_2\} \notin E$. Thus if $v_1 \in V_1$ and $v_2 \in V_2$, then v_1 and v_2 are not p- or q-connected.

Suppose there exist $v_1 \in V_1$ and $v_2 \in V_2$ such that $\{v_1, v\} \in E$ and $\{v, v_2\} \in E$. Since v has no q-connections, v_1 and v are p-connected, and v and v_2 are p-connected. Then v_1 and v_2 are p-connected, and $\{v_1, v_2\} \in E$ which is a contradiction. Thus if for $v_1 \in V_1$, $\{v_1, v\} \in E$, then $\{v_2, v\} \notin E$ for all $v_2 \in V_2$, and vice versa.

If neither $\{v_1, v\} \in E$ nor $\{v_2, v\} \in E$ for any $v_1 \in V_1$, $v_2 \in V_2$, or $\{v_1, v\} \in E$ for some $v_1 \in V_1$, then $V = (V' \setminus \{v\}) \cup (V_1 \cup \{v\}) \cup V_2$ where $V_1 \cup \{v\}$ and V_2 are disjoint and nonempty sets, and $w_1 \in V_1 \cup \{v\}$ and $w_2 \in V_2$ imply $\{w_1, w_2\} \notin E$. If, on the other hand, $\{v_2, v\} \in E$ for some $v_2 \in V_2$, then $V = (V' \setminus \{v\}) \cup V_1 \cup (V_2 \cup \{v\})$ where V_1 and $V_2 \cup \{v\}$ are disjoint and nonempty sets, and $w_1 \in V_1$ and $w_2 \in V_2 \cup \{v\}$ imply $\{w_1, w_2\} \notin E$. Either way, $V' \setminus \{v\}$ is a $(k - 1)$-vertex cut of G. □

Other proof of Corollary 3.11.

Proof. Suppose $u \in \mathcal{V}_{1,p,q}$. Let $G = (V, E)$ be the (p, q)-periodic graph of size $|u| = p + q - 1$ with indexing function f whose domain is $\{0, \ldots, p + q - 2\}$. We can show that the q-classes of vertices in $\{f(0), f(1), \ldots, f(p-2)\}$ each have 2 elements, and the q-classes of vertices in $\{f(p-1), \ldots, f(q-1)\}$ each have 1 element. Thus the set of vertices of V that have q-connections is the union of q-classes with 2 elements, which is $\{f(0), f(q)\} \cup \{f(1), f(q+1)\} \cup \cdots \cup \{f(p-2), f(p+q-2)\} = f(\{0, \ldots, p+q-2\} \setminus \{p-1, \ldots, q-1\})$. Then, by Lemma 4.5, $H(u) \subset \{0, \ldots, p+q-2\} \setminus \{p-1, \ldots, q-1\}$. □

5 Optimal bounds for $p = 2$

We now give the optimal bound for the case where $p = 2$.

Theorem 5.1. *$L(h, 2, q) = (2n+1)q + m + 1$ for $h = nq + m$, where $0 \leq m < q$.*

Proof. Throughout the proof, $p = 2$. First, let us show that $(2n+1)q + m + 1$ is a lower bound. Let u be a word of length $(2n+1)q + m + 1$ with periods 2 and q and number of holes $h = nq + m$. We must show that u has period 1 (note that $\gcd(p, q) = 1$). Thus, this is equivalent to showing that the (p, q)-periodic graph of size $(2n+1)q + m + 1$ has connectivity at least $h + 1$, or that a vertex cut of such graph must have at least $h + 1$ elements.

Let $G = (V, E)$ denote such a graph. Note that G has a particular structure. Indeed, each vertex belongs to one of two complete subgraphs representing the p-classes of u, namely the subgraph with vertex set the p-class of vertex 0 (the vertices with even indices) and the subgraph with vertex set the p-class of vertex 1 (the vertices with odd indices). Each p-connection is contained within one of these subgraphs. However, there are a number of q-connections (not all) across these p-classes. Note that in order to disconnect G, all such inter-p-class q-connections must be broken.

Thus a lower bound on the vertex connectivity of G is the sum of the least number of vertex removals required to break all inter-p-class q-connections within a q-class over all q-classes. Let us then consider a single q-class, denoted X. We can think of X as being the union of two sets, namely $Y = X \cap$ (p-class of vertex 0), and $Z = X \cap$ (p-class of vertex 1). Each element in X is q-connected to every other element in X. However, the q-connections within Y or Z are also p-connections; thus the inter-p-class q-connections of X are exactly those connections between Y and Z. Since every element of Y is q-connected with every element of Z, either all of Y or all of Z must be removed from G in order to break all inter-p-class q-connections within X. Also, if all of Y or all of Z is removed from G, then all inter-p-class q-connections within X are broken. Note that $X = \{f(i), f(i+q), f(i+2q), \ldots\}$ where $i \in \{0, 1, .., q-1\}$ is the smallest index in X, and f is the indexing function of G. Since q is odd, the terms of the sequence i, $i+q$, $i+2q$,… alternate between odd and even for integer i. Thus, since Y contains only vertices with even indices and Z contains only vertices with odd indices, the sizes of Y and Z are at most one apart. Then $\left\lfloor \frac{\|X\|}{2} \right\rfloor \leq \|Y\|$ and $\left\lfloor \frac{\|X\|}{2} \right\rfloor \leq \|Z\|$ and at least $\left\lfloor \frac{\|X\|}{2} \right\rfloor$ vertices must be removed from each q-class X in order to disconnect the graph G.

From the size of G, $\|V\| = (2n+1)q + m + 1$, we see that there are q q-classes, $m+1$ of which have $2n+2$ elements and $q - (m+1)$ of which have $2n+1$ elements. Each of the q-classes with $2n+2$ elements require at least $\left\lfloor \frac{2n+2}{2} \right\rfloor = n+1$ vertex removals to break all inter-p-class q-connections within the q-class, and each of the q-classes with $2n+1$ elements require at least $\left\lfloor \frac{2n+1}{2} \right\rfloor = n$ vertex removals to break all inter-p-class q-connections within the q-class. Thus, in all, $(m+1)(n+1) + (q-(m+1))n = mn + m + n + 1 + nq - nm - n = nq + m + 1 = h + 1$ vertex removals are required to disconnect G, thus the connectivity of G is at least $h + 1$.

Now, let us show that $(2n+1)q+m+1$ is an upper bound and thus the optimal bound. Consider the word $u = \diamond^m w(\diamond^q w)^n$ where w is the unique element in $\mathcal{V}_{0,2,q}$ of length q. We will show that u is an optimal word. Note that $|u| = (2n+1)q+m$, u has h holes, and since w is not 1-periodic, we also have that u is not 1-periodic. It is easy to show that u is 2- and q-periodic. □

6 Optimal bounds for large q

In this section, we present our main result which provides a formula for the optimal bound $L(h, p, q)$ when q is large enough.

Define

$$x(p,h) = \begin{cases} p\left(\frac{h}{2}\right) & \text{if } h \text{ is even} \\ p\left(\frac{h+1}{2}\right) & \text{if } h \text{ is odd} \end{cases}$$

and

$$y(h,p,q) = \begin{cases} p\left(\frac{h+2}{2}\right) + q - \gcd(p,q) & \text{if } h \text{ is even} \\ p\left(\frac{h+1}{2}\right) + q & \text{if } h \text{ is odd} \end{cases}$$

Theorem 6.1. *If $q > x(p,h)$, then $L(h,p,q) = y(h,p,q)$.*

The proof of Theorem 6.1 is split into two parts: the part that $y(h,p,q)$ is indeed a lower bound, and the part that this bound is optimal. The former is provided first.

Lemma 6.2. *If $q > x(p,h)$, then $y(h,p,q)$ is a lower bound.*

Proof. We want to show that a partial word u with periods p, q and h holes of length greater than or equal to $l = y(h,p,q)$ also has period $\gcd(p,q)$.

Suppose that $\gcd(p,q) = 1$. First let h be odd. Then we have that $x(p,h) = p\left(\frac{h+1}{2}\right)$. So $q > x(p,h)$ implies that $q = p\left(\frac{h+1}{2}\right) + k$ for some $k > 0$. It is enough to show that if $|u| = l$, then u has period 1 because if $|u| > l$, then all factors of u of length l would have period 1, and so u itself would. To see this, suppose $|u| = l+1$. The prefix of u of length l has periods p and q, and so it has period 1. The same holds for the suffix of u of length l. If u starts or ends with $\diamond$, then the result trivially holds. Otherwise, $u = au'b$ for some u' of length $l-1$ and some $a, b \in A$. There exists an occurrence of the letter b in u' because $D(u') \neq \emptyset$ by the way l is defined. The equality $b = a$ hence holds. Thus, by induction, any word u of length $\geq l$ satisfying our assumptions is 1-periodic. Now, since $|u| = p\left(\frac{h+1}{2}\right) + q$ and $q = p\left(\frac{h+1}{2}\right) + k$, we have that $|u| = (h+1)p + k = 2q - k$.

Consider the graph of u. Since $|u| = 2q-k$, positions of u within $\{q-k, q-k+1, \ldots, q-2, q-1\}$ have no q-connections, and all other elements within $\{0, \ldots, q-k-1\}$ have exactly one q-connection. Therefore, the number of positions of u which have exactly one q-connection is $|u| - k = (h+1)p$. Thus, each p-class has exactly $h+1$ elements with exactly one q-connection and all other elements of the p-class have no q-connections. In each i^{th} p-class, $\frac{h+1}{2}$ elements have q-connections with elements in the $((i+q) \bmod p)^{th}$$p$-class and $\frac{h+1}{2}$ elements have q-connections with elements in the $((i-q) \bmod p)^{th}$$p$-class. Thus, there are at least $\frac{h+1}{2}$ disjoint cycles in the graph that visit all p-classes and contain all the vertices with q-connections. In order to build $\frac{h+1}{2}$ such disjoint cycles, pick the smallest vertex v_0 in the $0^{th} = i_0^{th}$ p-class that has not been visited and that has a q-connection with an element w_1 of the i_1^{th} p-class. Then visit the vertex w_1 followed by the smallest nonvisited vertex v_1 of that i_1^{th} p-class. Go on like this visiting vertices until you visit w_p in the 0^{th} p-class. Then return to v_0. Such cycle has the form $v_0, w_1, v_1, w_2, v_2, \ldots, w_{p-1}, v_{p-1}, w_p, v_0$. Also, for each such cycle, every element of the graph either belongs to the cycle, or is p-connected to a member of the cycle. There are two types of disconnections possible: one that isolates a set of vertices with elements in different p-classes, and one that isolates a set of vertices within a p-class. Thus in order to disconnect the graph, either all $\frac{h+1}{2}$ cycles must be disconnected or all $h+1$ q-connections of a single p-class must be removed. The latter case clearly takes more than h holes, and since two holes are required to disconnect a cycle, we see that at least $h+1$ holes are required to disconnect the graph in the former case. Thus the graph of u is connected and u is 1-periodic.

Now, let h be even. The idea of the proof in this case is similar to that of an odd numbers of holes. When h is even, we must disconnect $\frac{h}{2}$ cycles that each requires two holes to break and one path that requires one hole to break. Hence we require $h+1$ holes to disconnect the graph of length $y(h,p,q)$.

Suppose $\gcd(p,q) = d \neq 1$. Also suppose that h is even; the odd h case follows in much the same way. Thus $|u| = p\left(\frac{h+2}{2}\right) + q - d$. Consider the set of partial words $u_0, \ldots, u_{d-1}$ where $u_i = u(i)u(i+d)u(i+2d)\ldots$. Each of these words has periods $\frac{p}{d}$ and $\frac{q}{d}$ which are co-prime. So if each u_i had period 1, then the word u has period d. Each u_i has length $\frac{p}{d}\left(\frac{h+2}{2}\right) + \frac{q}{d} - 1$ and at most h holes. Thus, by the proof given of this theorem for the case $\gcd(p,q) = 1$, each u_i has period 1, therefore u is d-periodic. □

Lemma 6.3. *If $q > x(p,h)$, then $y(h,p,q)$ is optimal.*

Proof. We will prove this in the case where $\gcd(p,q) = 1$ by giving a word with h holes of length $y(h,p,q) - 1$ which is p-periodic and q-periodic but not $\gcd(p,q)$-periodic.

First, suppose h is even. Consider the word $u = (\text{pref}_{p-2}(w)\diamond\diamond)^{\frac{h}{2}} w$ where w is the unique element in $\mathcal{V}_{0,p,q}$ of length $p+q-2$. We will show that u is an optimal word. First, note that $|u| = \frac{hp}{2} + p + q - 2 = y(h,p,q) - 1$, u has h holes, and since w is not 1-periodic, we also have that u is not 1-periodic. Now, note that w is p-periodic. Also, $\text{pref}_{p-2}(w)\diamond\diamond$ has length p and since $\text{pref}_{p-2}(w)\diamond\diamond \subset \text{pref}_p(w)$, we see that u is p-periodic. Since $q > x(p,h) = \frac{hp}{2}$, w is of length $q + p - 2 > \frac{hp}{2} + p - 2$. In order to show that u is q-periodic, it is enough to show that

$$\text{pref}_{\frac{hp}{2}+p-2}(u) \uparrow \text{suff}_{\frac{hp}{2}+p-2}(u)$$

Now, $\text{pref}_{\frac{hp}{2}+p-2}(u) = (\text{pref}_{p-2}(w)\diamond\diamond)^{\frac{h}{2}}\text{pref}_{p-2}(w)$, and

$$\text{suff}_{\frac{hp}{2}+p-2}(u) = \text{suff}_{\frac{hp}{2}+p-2}(w) = \text{pref}_{\frac{hp}{2}+p-2}(w)$$

since w is a palindrome by Proposition 3.6. Since w is p-periodic, $\text{pref}_{\frac{hp}{2}+p-2}(w) = (\text{pref}_p(w))^{\frac{h}{2}}\text{pref}_{p-2}(w)$. The desired compatibility relationship follows.

Now, suppose h is odd. We can verify that an optimal word in this case is $u = (\text{pref}_{p-2}(w)\diamond\diamond)^{\frac{h-1}{2}} w\diamond$. □

In the case of no hole, we see that $x(p,0) = 0$ and the formula presented in Theorem 6.1 agrees with $L(0,p,q) = p + q - \gcd(p,q)$. The case of one hole yields $x(p,1) = p$ and once again, our formula gives $L(1,p,q) = p+q$ which corresponds to the expression given in Theorem 3.8.

We end this section with the following result.

Theorem 6.4. *If $u \in \mathcal{V}_{h,p,q}$ and $q > x(p,h)$, then*

$$H(u) \subset \{0, \ldots, x(p,h) - 1\} \cup \{|u| - x(p,h), \ldots, |u| - 1\}$$

Proof. This proof is similar to the one provided for Lemma 4.5. In that proof we mentioned that the way to disconnect a graph was to place holes in positions with q-connections. The same idea holds here. □

7 Upper bounds for small q

In this section, we investigate the bounds $L(h,p,q)$ when $h \geq 2$ and $q \leq x(p,h)$. The word $w_{0,p,q}$ will denote the unique element in $\mathcal{V}_{0,p,q}$ of length $p+q-2$ over the alphabet $\{a,b\}$ starting with a.

7.1 The case of three holes

Define

$$z(3,p,q) = \begin{cases} 2q+p & \text{if } q-p < \frac{p}{2} \\ 4p & \text{if } \frac{p}{2} < q-p < p \\ 2p+q & \text{if } p < q-p \end{cases}$$

Theorem 7.1. *The bound $z(3,p,q)$ is an upper bound.*

Proof. Consider the word v given by

v	if
$a^{2(q-p)-1}\diamond a^{p-1-(q-p)}\diamond a^{(q-p)-1}\diamond a^{p-1}ba^{p-1}$	$q-p<\frac{p}{2}$
$a^{p-1}\diamond a^{p-1-(q-p)}\diamond a^{(q-p)-1}\diamond a^{p-1}ba^{p-1}$	$\frac{p}{2}<q-p<p$
$\text{pref}_{p-2}(w_{0,p,q})\diamond\diamond w_{0,p,q}\diamond$	$p<q-p$

The case where $p < q-p$ was proved earlier. For the other two cases, we can show that v is optimal or that v has three holes, has length $z(3,p,q)-1$, is p-periodic, is q-periodic, and is not 1-periodic. For instance, if $\frac{p}{2} < q-p < p$, then the p-periodicity of v can be checked by noticing that the b is aligned with the underlined $\diamond$'s when we build rows of length p:

$$a^{p-1}\underline{\diamond}a^{p-1-(q-p)}\diamond a^{(q-p)-1}\underline{\diamond}a^{p-1}ba^{p-1}$$

Note that the factor between these underlined positions has length $p-1$, and that the factor $a^{p-1-(q-p)}$ is defined since $p-1-(q-p) \geq 0$ due to the restriction $q-p<p$. Similarly, the q-periodicity of v comes from the fact that the b is aligned with the following underlined $\diamond$ when we build rows of length q:

$$a^{p-1}\diamond a^{p-1-(q-p)}\underline{\diamond}a^{(q-p)-1}\diamond a^{p-1}ba^{p-1}$$

Note that the factor preceding this underlined $\diamond$ has length $3p-q-1$ which is smaller than q due to the restriction $q-p > \frac{p}{2}$. □

7.2 The case of four holes

Define

$$z(4,p,q) = \begin{cases} q+3p-\gcd(p,q) & \text{if } q-p < \frac{p}{2} \\ q+3p & \text{if } \frac{p}{2} < q-p < p \\ q+3p-\gcd(p,q) & \text{if } p < q-p \end{cases}$$

Theorem 7.2. *The bound $z(4,p,q)$ is an upper bound.*

Proof. Consider the following words:

v	if
$w_{0,q-p,p-(q-p)}\diamond\diamond w_{0,p,q}\diamond\diamond w_{0,q-p,p-(q-p)}$	$q-p<\frac{p}{2}$
$a^{(q-p)-1}\diamond a^{p-1}ba^{p-1}\diamond a^{(q-p)-1}\diamond a^{p-1-(q-p)}\diamond a^{p-1}$	$\frac{p}{2}<q-p<p$
$(\text{pref}_{p-2}(w_{0,p,q})\diamond\diamond)^2 w_{0,p,q}$	$p<q-p$

Using Proposition 3.4, we can show that the first item is p-periodic and q-periodic. Moreover, it has 4 holes, has length $q+3p-2$, and does not have period 1. It is therefore optimal when $q-p<\frac{p}{2}$. Similarly to the proof of Theorem 7.1, the other two items can be shown to be optimal. □

7.3 The case of five holes

Define

$$z(5,p,q) = \begin{cases} 3q+p & \text{if } q-p < \frac{p}{3} \\ 5p & \text{if } \frac{p}{3} < q-p < \frac{p}{2} \\ 5p & \text{if } \frac{p}{2} < q-p < \frac{2p}{3} \\ 3q & \text{if } \frac{2p}{3} < q-p < p \\ 6p & \text{if } p < q-p < 2p \\ q+3p & \text{if } 2p < q-p \end{cases}$$

Theorem 7.3. *The bound $z(5,p,q)$ is an upper bound.*

Proof. The words given below are optimal:

v	if
$a^{3(q-p)-1}\diamond a^{p-1-2(q-p)}\diamond a^{2(q-p)-1}\diamond$ $a^{p-1-(q-p)}\diamond a^{(q-p)-1}\diamond a^{p-1}ba^{p-1}$	$q-p<\frac{p}{3}$
$a^{p-1}\diamond a^{p-1-2(q-p)}\diamond a^{2(q-p)-1}\diamond$ $a^{p-1-(q-p)}\diamond a^{(q-p)-1}\diamond a^{p-1}ba^{p-1}$	$\frac{p}{3}<q-p<\frac{p}{2}$
$a^{p-1-(q-p)}\diamond a^{q-p-1}\diamond a^{p-1}ba^{p-1}\diamond$ $a^{q-p-1}\diamond a^{p-1-(q-p)}\diamond a^{p-1}$	$\frac{p}{2}<q-p<\frac{2p}{3}$
$a^{p-1-2(p-(q-p))}\diamond a^{p-1}\diamond a^{p-1-(q-p)}\diamond$ $a^{(q-p)-1}\diamond a^{p-1}ba^{p-1}\diamond a^{(q-p)-1}$	$\frac{2p}{3}<q-p<p$
$a^{p-1}\diamond a^{(2p-1)-(q-p)}\diamond a^{(q-p)-p-1}\diamond a^{p-1}$ $\diamond a^{p-1}ba^{p-1}\diamond a^{p-1}$	$p<q-p<2p$
$(\text{pref}_{p-2}(w_{0,p,q})\diamond\diamond)^2 w_{0,p,q}\diamond$	$2p<q-p$

□

7.4 The case of six holes

Define

$$z(6,p,q) = \begin{cases} 5p & \text{if } q-p < \frac{p}{4} \\ 4q & \text{if } \frac{p}{4} < q-p < \frac{p}{2} \\ 6p & \text{if } \frac{p}{2} < q-p < p \\ 2q+2p & \text{if } p < q-p < \frac{3p}{2} \\ 7p & \text{if } \frac{3p}{2} < q-p < 2p \\ q+4p-\gcd(p,q) & \text{if } 2p < q-p \end{cases}$$

Theorem 7.4. *The bound $z(6,p,q)$ is an upper bound.*

Proof. The words given by

v	if
$a^{p-1-3(q-p)}\diamond a^{3(q-p)-1}\diamond a^{p-1-2(q-p)}\diamond$ $a^{2(q-p)-1}\diamond a^{p-1-(q-p)}\diamond a^{(q-p)-1}\diamond a^{p-1}ba^{p-1}$	$q-p<\frac{p}{4}$
$a^{2(q-p)-1}\diamond a^{p-1-(q-p)}\diamond a^{(q-p)-1}\diamond a^{p-1}$ $ba^{p-1}\diamond a^{(q-p)-1}\diamond a^{p-1-(q-p)}\diamond a^{2(q-p)-1}$	$\frac{p}{4}<q-p<\frac{p}{2}$
$a^{p-1}\diamond a^{p-1-(q-p)}\diamond a^{q-p-1}\diamond a^{p-1}ba^{p-1}$ $\diamond a^{q-p-1}\diamond a^{p-1-(q-p)}\diamond a^{p-1}$	$\frac{p}{2}<q-p<p$
$a^{2((q-p)-p)-1}\diamond a^{p-1}\diamond a^{p-1-((q-p)-p)}\diamond$ $a^{(q-p)-p-1}\diamond a^{p-1}\diamond a^{p-1}ba^{p-1}\diamond a^{p-1}$	$p<q-p<\frac{3p}{2}$
$a^{p-1}\diamond a^{p-1}\diamond a^{p-1-((q-p)-p)}\diamond a^{(q-p)-p-1}$ $\diamond a^{p-1}\diamond a^{p-1}ba^{p-1}\diamond a^{p-1}$	$\frac{3p}{2}<q-p<2p$
$(\text{pref}_{p-2}(w_{0,p,q})\diamond\diamond)^3 w_{0,p,q}$	$2p<q-p$

can be checked to be optimal. □

7.5 The case of seven holes

Define

$$z(7,p,q)=\begin{cases}4q+p & \text{if } q-p<\frac{p}{4}\\ 6p & \text{if } \frac{p}{4}<q-p<\frac{p}{2}\\ 4q & \text{if } \frac{p}{2}<q-p<\frac{2p}{3}\\ q+5p & \text{if } \frac{2p}{3}<q-p<p\\ 7p & \text{if } p<q-p<\frac{4p}{3}\\ 3q & \text{if } \frac{4p}{3}<q-p<\frac{3p}{2}\\ q+5p & \text{if } \frac{3p}{2}<q-p<2p\\ 8p & \text{if } 2p<q-p<3p\\ q+4p & \text{if } 3p<q-p\end{cases}$$

Theorem 7.5. *The bound $z(7,p,q)$ is an upper bound.*

Proof. The words given by

v	if
$a^{4(q-p)-1}\diamond a^{p-1-3(q-p)}\diamond a^{3(q-p)-1}\diamond a^{p-1-2(q-p)}\diamond a^{2(q-p)-1}\diamond$ $a^{p-1-(q-p)}\diamond a^{(q-p)-1}\diamond a^{p-1}ba^{p-1}$	$q-p<\frac{p}{4}$
$a^{p-1}\diamond a^{p-1-3(q-p)}\diamond a^{3(q-p)-1}\diamond a^{p-1-2(q-p)}\diamond a^{2(q-p)-1}\diamond$ $a^{p-1-(q-p)}\diamond a^{(q-p)-1}\diamond a^{p-1}ba^{p-1}$	$\frac{p}{4}<q-p<\frac{p}{3}$
$a^{p-1}\diamond a^{p-1-2(q-p)}\diamond a^{2(q-p)-1}\diamond a^{p-1-(q-p)}\diamond a^{(q-p)-1}\diamond$ $a^{p-1}ba^{p-1}\diamond a^{(q-p)-1}\diamond a^{p-1-(q-p)}$	$\frac{p}{3}<q-p<\frac{p}{2}$
$a^{(q-p)-1}\diamond a^{p-1}ba^{p-1}\diamond a^{(q-p)-1}\diamond a^{p-1-(q-p)}\diamond$ $a^{p-1}\diamond a^{2q-3p-1}\diamond a^{4p-2q-1}\diamond a^{3q-4p-1}$	$\frac{p}{2}<q-p<\frac{2p}{3}$
$a^{(q-p)-1}\diamond a^{p-1}ba^{p-1}\diamond a^{(q-p)-1}\diamond a^{p-1-(q-p)}\diamond$ $a^{p-1}\diamond a^{2q-3p-1}\diamond a^{4p-2q-1}\diamond a^{p-1}$	$\frac{2p}{3}<q-p<p$
$a^{p-1-2((q-p)-p)}\diamond a^{2((q-p)-p)-1}\diamond a^{p-1}\diamond a^{p-1-((q-p)-p)}\diamond$ $a^{(q-p)-p-1}\diamond a^{p-1}\diamond a^{p-1}ba^{p-1}\diamond a^{p-1}$	$p<q-p<\frac{4p}{3}$
$a^{2((q-p)-p)-1}\diamond a^{p-1}\diamond a^{p-1-((q-p)-p)}\diamond$ $a^{(q-p)-p-1}\diamond a^{p-1}\diamond a^{p-1}ba^{p-1}\diamond a^{p-1}\diamond a^{(q-p)-p-1}$	$\frac{4p}{3}<q-p<\frac{3p}{2}$
$a^{p-1}\diamond a^{p-1}\diamond a^{p-1-((q-p)-p)}\diamond$ $a^{(q-p)-p-1}\diamond a^{p-1}\diamond a^{p-1}ba^{p-1}\diamond a^{p-1}\diamond a^{(q-p)-p-1}$	$\frac{3p}{2}<q-p<2p$
$a^{p-1}\diamond a^{(3p-1)-(q-p)}\diamond a^{(q-p)-2p-1}\diamond a^{p-1}\diamond a^{p-1}$ $\diamond a^{p-1}ba^{p-1}\diamond a^{p-1}\diamond a^{p-1}$	$2p<q-p<3p$
$(\mathrm{pref}_{p-2}(w_{0,p,q})\diamond\diamond)^3 w_{0,p,q}\diamond$	$3p<q-p$

can be checked to be optimal. □

Referring to Sections 7.1, 7.2, 7.3, 7.4 and 7.5, we conjecture that our bounds are optimal for $h = 3, 4, 5, 6$ and 7.

Conjecture 7.6. *The equality $L(h,p,q) = z(h,p,q)$ holds for $h = 3,4,5,6$ and 7.*

Figure 2 summarizes our conjectures when $\gcd(p,q) = 1$.

From Figure 2, some bounds can be generalized. Indeed, from Cases 1, 3, 5 and 7, we provide in the following theorem a general form for $z(h,p,q)$ in case h is odd and $q-p < \frac{2p}{h+1}$. The optimal words for that bound turm out to be rather elegant.

Theorem 7.7. *For $h \geq 1$ odd, define $z(h,p,q) = (\frac{h+1}{2})q + p$ if $q-p < \frac{2p}{h+1}$. Then the bound $z(h,p,q)$ is an upper bound in that case.*

Proof. Define $u_1 = a^{(q-p)-1}\diamond a^{p-1}ba^{p-1}$, and for $h > 0$ define

$$u_{2h+1} = a^{(h+1)(q-p)-1}\diamond a^{p-1-h(q-p)}\diamond u_{2h-1}$$

By induction on h, we can show that u_{2h+1} has $2h+1$ holes, has length $(h+1)q+p-1$, is p-periodic, is q-periodic, and is not 1-periodic. The p-periodicity can be checked by noticing that the b is aligned with $\diamond$'s when we build rows of length p. Indeed, the b is aligned with every other $\diamond$ starting with the $\diamond$ preceding it and ending with the first $\diamond$. Note that the factor preceding the first $\diamond$ has length $(h+1)(q-p)-1$ which is smaller than $p-1$ due to the restriction $q-p < \frac{p}{h+1}$. In addition, the q-periodicity can be checked by noticing that the b is aligned with $\diamond$'s when we build rows of length q. Indeed, the b is aligned with every other $\diamond$ starting with the second $\diamond$ preceding it and ending with the second $\diamond$. The factor preceding the second $\diamond$ has length $q-1$. □

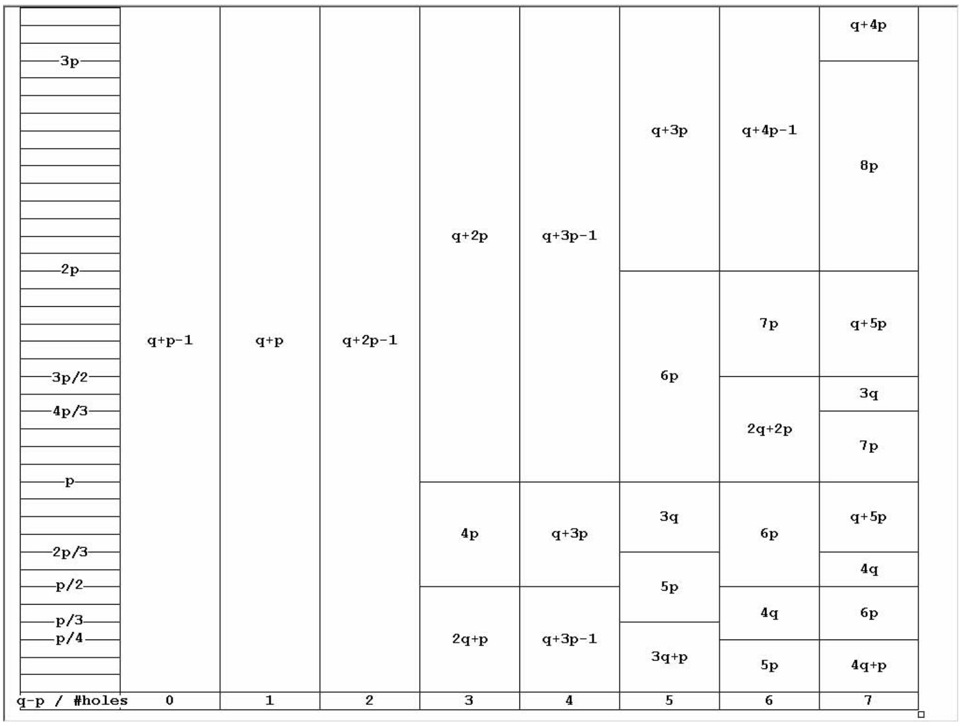

Figure 2: Bounds $z(h, p, q)$ where $\gcd(p, q) = 1$

Conjecture 7.8. *If $h \geq 1$ is odd and $q - p < \frac{2p}{h+1}$, then $L(h, p, q) = (\frac{h+1}{2})q + p$.*

From Figure 2, we can conjecture the following across the number of holes.

Conjecture 7.9. *If $h \geq 3$ is odd and $\frac{2p}{h+1} < q - p < \frac{4p}{h+1}$, then $L(h, p, q) = (\frac{h+5}{2})p$.*

Finding general optimal words for Conjecture 7.9 is challenging due in part to the fact that the optimal words take different shapes in subintervals (see for instance the case of $h = 7$ where the interval $\frac{p}{4} < q - p < \frac{p}{2}$ gets split into the two subintervals $\frac{p}{4} < q - p < \frac{p}{3}$ and $\frac{p}{3} < q - p < \frac{p}{2}$).

8 Conclusion

In this paper, we connected the problem of finding optimal bounds for Fine and Wilf's generalizations to partial words with that of finding the vertex connectivity of certain graphs. Many algorithms for the computation of vertex connectivity in graphs have been developed over the years [1, 23, 24, 25, 26, 29, 30, 31, 33, 42]. While such computation can often be reduced to solving a number of max-flow problems, it can also be computed using other methods such as randomised algorithms. Algorithms have also been developed for deciding whether a graph is k-vertex connected, some of which are max-flow based while some are not.

An algorithm that computes the minimum of maximum flows between all non-adjacent vertices of an associated digraph can be described as follows. The justification is based on Menger's Theorem and the Maximum Flow-Minimum Cut Theorem, and the algorithm for finding the maximum flow is due to Ford and Fulkerson (See Reference [15] where Menger's theorem is on page 46, and a detailed description of the algorithm by Ford and Fulkerson is given on pages 198–202).

We have shown that $L(h, p, q)$ is the smallest l such that the (p, q)-periodic graph of size l has connectivity at least $h + 1$. This leads to an efficient algorithm for determining $L(h, p, q)$. To find the smallest l

such that G, the (p, q)-periodic graph of size l, has connectivity at least $h + 1$, we iterate over l (note that $p + q - 1 \leq L(h, p, q)$), checking the connectivity of each G. Suppose we wish to find the connectivity of $G = (V, E)$ for a certain size l. By Menger's Theorem, this is equivalent to finding the minimum of the maximum number of vertex-disjoint paths between pairs of non-adjacent vertices in G. To find the maximum number of vertex-disjoint paths between a pair of non-adjacent vertices $v_1, v_2 \in V$, we first produce a digraph $D = (V', A')$ from G as follows:

1. Add v_1 and v_2 to V',
2. For each vertex $v \in V \setminus \{v_1, v_2\}$, add two vertices v' and v'' to V' and arc (v', v'') to A', and
3. For each edge $(v, u) \in E$, add the two arcs (v'', u') and (u'', v') to A'.

We see that directed paths between v_1 and v_2 in D correspond directly to paths between v_1 and v_2 in G. Furthermore, two directed paths between v_1 and v_2 in D are arc-disjoint if and only if the corresponding paths in G are vertex-disjoint. We make D a network by giving each arc unit capacity and setting v_1 to be the source and v_2 to be the sink. Then from Lemma 11.4 (see page 203 of Reference [15]), the maximum number of arc-disjoint directed paths from v_1 to v_2 is equal to the value of a maximum flow in D. To find a maximum flow in D, we use the algorithm of Ford and Fulkerson known as the labelling method, which starts with a known flow on D (say, the zero flow), and recursively increments the flow, terminating with the maximum flow.

Vertex connectivity in the (p, q)-periodic graphs needs to be further studied in order to prove the conjectures of Section 7 and to give bounds for any number of holes. This becomes complicated as the number of holes increases, since the number of cases increases as well.

References

[1] M. Becker, W. Degenhardt, J. Doenhardt, S. Hertel, G. Kaninke and W. Keber, A probabilistic algorithm for vertex connectivity of graphs, *Information Processing Letters* **15** (1982) 135–136.

[2] J. Berstel and L. Boasson, Partial words and a theorem of Fine and Wilf, *Theoretical Computer Science* **218** (1999) 135–141.

[3] J. Berstel and A. de Luca, Sturmian words, Lyndon words and trees, *Theoretical Computer Science* **178** (1997) 171–203.

[4] F. Blanchet-Sadri, A periodicity result of partial words with one hole, *Computers and Mathematics with Applications* **46** (2003) 813–820.

[5] F. Blanchet-Sadri, Periodicity on partial words, *Computers and Mathematics with Applications* **47** (2004) 71–82.

[6] F. Blanchet-Sadri, Codes, orderings, and partial words, *Theoretical Computer Science* **329** (2004) 177–202.

[7] F. Blanchet-Sadri, Primitive partial words, *Discrete Applied Mathematics* **148** (2005) 195–213.

[8] F. Blanchet-Sadri and Arundhati R. Anavekar, Testing primitivity on partial words, *Discrete Applied Mathematics* **155** (2007) 279–287 (`www.uncg.edu/mat/primitive`).

[9] F. Blanchet-Sadri, D. Dakota Blair and Rebeca V. Lewis, Equations on partial words, in R. Królovic and P. Urzyczyn (Eds.), *MFCS 2006, 31st International Symposium on Mathematical Foundations of Computer Science, August 28–September 1, 2006, Stará Lesná, Slovakia*, Lecture Notes in Computer Science, Vol. 4162 (Springer-Verlag, Berlin, Heidelberg, 2006) 167–178 (`www.uncg.edu/mat/research/equations`).

[10] F. Blanchet-Sadri and Ajay Chriscoe, Local periods and binary partial words: an algorithm, *Theoretical Computer Science* **314** (2004) 189–216 (`www.uncg.edu/mat/AlgBin`).

[11] F. Blanchet-Sadri and S. Duncan, Partial words and the critical factorization theorem, *Journal of Combinatorial Theory, Series A* **109** (2005) 221–245 (`www.uncg.edu/mat/cft`).

[12] F. Blanchet-Sadri and Robert A. Hegstrom, Partial words and a theorem of Fine and Wilf revisited, *Theoretical Computer Science* **270** (2002) 401–419.

[13] F. Blanchet-Sadri and D.K. Luhmann, Conjugacy on partial words, *Theoretical Computer Science* **289** (2002) 297–312.

[14] F. Blanchet-Sadri, Taktin Oey and Tim Rankin, Partial words and generalized Fine and Wilf's theorem for an arbitrary number of weak periods (`www.uncg.edu/mat/research/finewilf2`).

[15] J.A. Bondy and U.S.R. Murty, *Graph Theory with Applications* (Elsevier Science Publishing Co., New York, 1982) (`www.ecp6.jussieu.fr/pageperso/bondy/books/gtwa/gtwa.html`).

[16] J. Buhler, U. Keich and Y. Sun, Designing seeds for similarity search in genomic DNA, *Journal of Computer and System Sciences* **70** (2005) 342–363.

[17] M.G. Castelli, F. Mignosi and A. Restivo, Fine and Wilf's theorem for three periods and a generalization of Sturmian words, *Theoretical Computer Science* **218** (1999) 83–94.

[18] C. Choffrut and J. Karhumäki, Combinatorics of Words, in G. Rozenberg and A. Salomaa (Eds.), *Handbook of Formal Languages, Vol. 1, Ch. 6* (Springer-Verlag, Berlin, 1997) 329–438.

[19] S. Constantinescu and L. Ilie, Generalised Fine and Wilf's theorem for arbitrary number of periods, *Theoretical Computer Science* **339** (2005) 49–60.

[20] A. de Luca, Sturmian words: Structure, combinatorics and their arithmetics, *Theoretical Computer Science* **183** (1997) 45–82.

[21] A. de Luca, On the combinatorics of finite words, *Theoretical Computer Science* **218** (1999) 13–39.

[22] A. de Luca and F. Mignosi, Some combinatorial properties of Sturmian words, *Theoretical Computer Science* **136** (1994) 361–385.

[23] E.A. Dinic, Algorithm for solution of a problem of maximum flow in a network with power estimation, *Soviet Math. Dokl.* **11** (1970) 1277–1280.

[24] A.H. Esfahanian and S.L. Hakimi, On computing the connectivities of graphs and digraphs, *Networks* (1984) 355-366.

[25] S. Even, An algorithm for determining whether the connectivity of a graph is at least k, *SIAM J. Computing* **4** (1975) 393–396.

[26] S. Even and R.E. Tarjan, Network flow and testing graph connectivity, *SIAM J. Computing* **4** (1975) 507–518.

[27] M. Farach-Colton, G.M. Landau, S.C. Sahinalp and D. Tsur, Optimal spaced seeds for approximate string matching, In L. Caires, G.F. Italiano, L. Monteiro, C. Palanidessi and M. Yung (eds.): *ICALP 2005*, Lecture Notes in Computer Science, Vol. 3580, (Springer-Verlag, Berlin, 2005) 1251–1262.

[28] N.J. Fine and H.S. Wilf, Uniqueness theorems for periodic functions, *Proceedings of the American Mathematical Society* **16** (1965) 109–114.

[29] Z. Galil and G.F. Italiano, Fully dynamic algorithms for 2-edge connectivity, *SIAM J. Computing* **21** (1992) 1047–1069.

[30] M.R. Henzinger, S. Rao and H.N. Gabow, Computing vertex connectivity: new bounds from old techniques, *Proc. 37th IEEE FOCS* (1996) 462–471.

[31] J. Hopcroft and R.E. Tarjan, Dividing a graph into triconnected components, *SIAM J. Computing* **2** (1973) 135–158.

[32] J. Justin, On a paper by Castelli, Mignosi, Restivo, *Theoretical Informatics with Applications* **34** (2000) 373–377.

[33] A. Kanevsky and V. Ramachandran, Improved algorithms for graph fourconnectivity, *J. Computer and System Sciences* **42** (1991) 288–306.

[34] U. Keich, M. Li, B. Ma and J. Tromp, On spaced seeds for similarity search, *Discrete Applied Mathematics* **138** (2004) 253–263.

[35] P. Leupold, Partial words for DNA coding, Lecture Notes in Computer Science, Vol. 3384 (Springer-Verlag, Berlin, 2005) 224–234.

[36] M. Lothaire, *Combinatorics on Words*, Vol. 17 of Encyclopedia of Mathematics and Applications (Addison-Wesley, Reading, Mass., 1983) (Reprinted in the Cambridge Mathematical Library (Cambridge University Press, Cambridge, UK, 1997)).

[37] B. Ma, J. Tromp and M. Li, PatternHunter: faster and more sensitive homology search, *Bioinformatics* **18** (2002) 440–445.

[38] F. Nicolas and E. Rivals, Hardness of optimal spaced seed design, *CPM 2005, 16th Annual Symposium on Combinatorial Pattern Matching* (2005) 144–155.

[39] L. Noé and G. Kucherov, Improved hit criteria for DNA local alignment, *BMC Bioinformatics* **5** (2004).

[40] A.M. Shur and Y.V. Gamzova, Partial words and the periods' interaction property, *Izvestiya RAN* **68** (2004) 199–222.

[41] A.M. Shur and Y.V. Konovalova, On the periods of partial words, Lecture Notes in Computer Science, Vol. 2136 (2001) 657–665.

[42] R.E. Tarjan, Depth first search and linear graph algorithms, *SIAM J. Computing* **1** (1972) 146–160.

[43] R. Tijdeman and L. Zamboni, Fine and Wilf words for any periods, *Indag. Math. (N.S.)* **14** (2003) 135–147.