# Border Correlations, Lattices, and the Subgraph Component Polynomial*

F. Blanchet-Sadri[1] M. Cordier[2] R. Kirsch[3]

July 7, 2016

**Abstract**

We consider the border sets of partial words and study the combinatorics of specific representations of them, called border correlations, which are binary vectors of same length indicating the borders. We characterize precisely which of these vectors are valid border correlations, and establish a one-to-one correspondence between the set of valid border correlations and the set of valid ternary period correlations of a given length, the latter being ternary vectors representing the strong and strictly weak period sets. It turns out that the sets of all border correlations of a given length form distributive lattices under suitably defined partial orderings. We also give connections between the ternary period correlation of a partial word and its refined border correlation which specifies the lengths of all the word's bordered cyclic shifts' minimal borders. Finally, we investigate the population size, that is, the number of partial words sharing a given (refined) border correlation, and obtain formulas to compute it. We do so using the subgraph component polynomial of an undirected graph, introduced recently by Tittmann et al. (European Journal of Combinatorics, 2011), which counts the number of connected components in vertex induced subgraphs.

*This material is based upon work supported by the National Science Foundation under Grants DMS–0754154 and DMS–1060775. The Department of Defense is also gratefully acknowledged. This is a journal version of [8].

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

[2]Department of Mathematical Sciences, Kent State University, P.O. Box 5190, Kent, OH 44242, USA, `mcordie1@kent.edu`

[3]Department of Mathematics, University of Nebraska-Lincoln, 203 Avery Hall, P.O. Box 880130, Lincoln, NE 68588–0130, USA, `rkirsch@math.unl.edu`

*Keywords*: Combinatorics on words; Partial words; Border correlation; Refined border correlation; Ternary period correlation; Population size; Lattice; Subgraph component polynomial.
*AMS Subject Classification*: 68R15; 68Q45; 05A05

## 1 Introduction

*Borders* and *periods* are two fundamental concepts of combinatorics on words that play an important role in several research areas including text compression, computational biology, string searching and pattern matching algorithms (see, e.g., [3, 16, 18, 21, 26]). It is well-known that these two word notions do not exist independently from each other. The length of the maximal border of a word is its length minus the length of its minimal period. This implies that it is unbordered if it has no proper period. Borders and periods are also well-studied concepts in combinatorics on partial words which allow positions to have don't care characters or holes, as well as indeterminate strings which allow positions to have subsets of the alphabet (see, e.g., [1, 2, 6, 9, 11–13, 19, 23, 25, 27]).

The combinatorics of specific representations of the border sets and period sets of (partial) words of length $n$ over a finite alphabet have been studied. Among them are the *period correlations*, which are $n$-bit vectors indicating the periods, and the *border correlations*, which are $n$-bit vectors indicating the presence of borders of a certain length. Guibas and Odlyzko [17] introduced the period correlations, so-called (auto)correlations, provided characterizations of them, asymptotic bounds on their number, and a recurrence for calculating the *population size* of a given period correlation, that is, the number of words sharing it as period correlation. Rivals and Rahmann [24] showed that the set of all period correlations of words of a given length is a lattice under set inclusion, proposed the first efficient algorithm for enumerating them, and improved upon Guibas and Odlyzko's asymptotic lower bounds on their number. They also provided a new recurrence to compute the population size. Harju and Nowotka [20] studied *refined border correlations* which are vectors that specify, for each cyclic shift of the word, the length of a minimal border (if a word can be written as $uv$, then $vu$ is a cyclic shift of the word). Extensions of these results to partial words appear in [7, 10]. In particular, *binary and ternary period correlations*, which are binary and ternary vectors representing the strong and strictly weak period sets of partial words, were considered.

Tittmann et al. [29] introduced the *subgraph component polynomial* $Q(G; x, y)$ of an undirected graph $G$ with $n$ vertices as the bivariate generating function which counts the number of connected components in vertex induced subgraphs, i.e., $Q(G;x,y) = \sum_{j=0}^{n}\sum_{i=0}^{n} q_{ji}(G)x^j y^i$, where $q_{ji}(G)$ is the number of vertex induced subgraphs of $G$ with exactly $j$ vertices and $i$ connected components. They showed that the subgraph component polynomial is more useful for distinguishing graphs than other graph polynomials that appear in the literature (see, e.g., [22] for more information on graph polynomials). Recently, Blanchet-Sadri et al. [5] showed the use of the subgraph component polynomial to count the number of primitive partial words of a given length over an alphabet of a fixed size, which leads to a method for enumerating such partial words.

In this paper, we establish connections between border correlations, lattices, and the subgraph component polynomial. We use the subgraph component polynomial to count the number of partial words of length $n$ over a $k$-letter alphabet that have the same border correlation, or the border correlation's population size. We associate an undirected graph $G_{\mathbf{bc}}$ with any border correlation $\mathbf{bc}$ of length $n$ as follows: the vertices represent the positions $0,\dots,n-1$ and the edges are the pairs $\{i,j\}$ that are $(n-\ell)$-apart, where $\ell$ indicates the length of a border. It turns out that the population size of $\mathbf{bc}$ can be expressed in terms of the $Q(G_{\mathbf{bc}};1,k)$'s. We also associate an undirected graph $H_{\mathbf{rbc}}$ with any refined border correlation $\mathbf{rbc}$ of length $n$ so that the population size of $\mathbf{rbc}$ can be expressed in terms of the $Q(H_{\mathbf{rbc}};1,k)$'s.

Borders are closely related to periods of words, so the study of border correlations of partial words is a relevant topic in combinatorics on words. Our main result gives a formula to determine the size of the set of partial words of a given length over a given alphabet that share a fixed border correlation. We establish this formula by using the subgraph component polynomial, a polynomial of undirected graphs. Thus, we combine two different areas of combinatorics.

The contents of our paper are as follows: In Section 2, we review some basic concepts on partial words such as borders, strong periods and weak periods, and discuss relationships between them. In Section 3, we first recall the fact that the set of binary (respectively, ternary) period correlations of partial words over an arbitrary alphabet of cardinality at least two is the same as the set of binary (respectively, ternary) period correlations of partial words over a binary alphabet. We then introduce border correlations and characterize precisely which vectors are valid border correlations. We establish a one-to-one correspondence between the set of valid border correlations and the set of valid ternary period correlations of a given length giving an algorithm for generating the valid ternary period correlations of a given length. We also prove that all valid border correlations correspond to partial words over the binary alphabet. In Section 4, we explore properties of refined border correlations and ternary period correlations, and we make connections between these two. In Section 5, we look at a better way of counting valid colorings of undirected graphs by using the subgraph component polynomial and looking at different graph structures. In Section 6, we give formulas to calculate the population size of a given border correlation. One approach is based on the previous section and another approach is based on the fact that the sets of all border correlations of a given length form distributive lattices under suitably defined partial orderings. We also give formulas to calculate the population size of a given refined border correlation. Finally in Section 7, we conclude with some open problems.

## 2 Preliminaries

Let $\Sigma$ be a finite and non-empty set of characters, called an *alphabet*. Each element $a$ of $\Sigma$ is referred to as a *letter*, and a sequence of letters from $\Sigma$ as a *word*, or *total word*, over $\Sigma$. A *partial word* over $\Sigma$ is a sequence of characters from $\Sigma_\diamond = \Sigma \cup \{\diamond\}$, where $\diamond$, a new character which is not in $\Sigma$, is the "don't care" or "hole" character (it represents an undefined position). Note that a total word is a partial word with no holes. The *length* of a partial word $w$, denoted by $|w|$, is the number of characters in $w$. For example, if $w = abbac\diamond\diamond c$ then $|w| = 8$. The *empty word* is the word of length zero and we denote it by $\varepsilon$. The set of all words over $\Sigma$ is denoted by $\Sigma^*$. Similarly, the set of all non-empty words over $\Sigma$ is denoted by $\Sigma^+$. We denote by $\Sigma^n$ the set of all words of length $n$ over $\Sigma$.

A partial word $u$ is a *factor* of a partial word $w$ if there exist (possibly empty) partial words $x, y$ such that $w = xuy$. We say that $u$ is a *prefix* of $w$ if $x = \varepsilon$. Similarly, $u$ is a *suffix* of $w$ if $y = \varepsilon$. Starting numbering positions from 0, we denote the character in position $i$ of $w$ by $w[i]$ and the factor of $w$ from position $i$ to position $j$ (inclusive) by $w[i..j]$ and from position $i$ to position $j$ (non-inclusive) by $w[i..j)$. We denote $w$ concatenated with itself $m$ times as $w^m$.

If $u$ and $v$ are partial words of equal length over $\Sigma$, then $u$ is *contained* in $v$, denoted by $u \subset v$, if $u[i] = v[i]$ for all $i$ such that $u[i] \in \Sigma$. 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 $v \subset w$. Given partial words $u$ and $v$ such that $u \uparrow v$, the *least upper bound* of $u$ and $v$ is the partial word $u \vee v$, where $u \subset (u \vee v)$ and $v \subset (u \vee v)$, and if $u \subset w$ and $v \subset w$ then $(u \vee v) \subset w$. For example,

$a\diamond\diamond ab \vee ac\diamond\diamond b = ac\diamond ab$.

A non-empty partial word $w$ is *unbordered* if no non-empty partial words $x_1, x_2, u, v$ exist such that $w = x_1 u = v x_2$ and $x_1 \uparrow x_2$. If such non-empty partial words exist, then $x$ exists such that $x_1 \subset x$ and $x_2 \subset x$ and we call $w$ *bordered* and $x$ a *border* of $w$. It is easy to see that if $w$ is unbordered and $w \subset w'$, then $w'$ is unbordered as well. A border $x$ of $w$ is *minimal* if $|x| > |y|$ implies that $y$ is not a border of $w$ and is *maximal* if $|x| < |y|$ implies that $y$ is not a border of $w$. For example, $a\diamond\diamond ab$ is bordered with borders $ab$ and $aab$, the first one being minimal, while $ab\diamond c$ is unbordered.

A *strong period* of a partial word $w$ over $\Sigma$ is a positive integer $p$ such that $w[i] = w[j]$ whenever $w[i], w[j] \in \Sigma$ and $i \equiv j \bmod p$; $w$ is called strongly $p$-periodic. A *weak period* of $w$ is a positive integer $p$ such that $w[i] = w[i+p]$ whenever $w[i], w[i+p] \in \Sigma$; $w$ is weakly $p$-periodic. A *strictly* weak period is a weak period that is not a strong period. For example, 2 is a strictly weak period of $aba\diamond aca$. The set of all strong periods (respectively, weak periods) of $w$ is denoted by $\mathbf{SP}(w)$ (respectively, $\mathbf{WP}(w)$). The following two lemmas are useful for our purposes. The first one states that a weak period is a strong period if and only if all of its multiples are also weak periods, and the second one establishes relationships between borders, and strong and weak periods. Note that $aba\diamond aca$ has a border of length 5 but does not have strong period 2.

**Lemma 2.1** ( [10]). *Let $w$ be a partial word and let $p \in \mathbf{WP}(w)$. Then $p \in \mathbf{SP}(w)$ if and only if for all $0 < i \leq \frac{|w|}{p}$, $ip \in \mathbf{WP}(w)$.*

Note that given the set of weak periods, the set of strong periods is directly implied by it.

**Lemma 2.2** ( [1]). *Let $w$ be a partial word.*

*(a) If $0 < \ell < |w|$, then $w$ has a border of length $\ell$ if and only if $w$ has weak period $|w| - \ell$.*

*(b) If $0 < \ell \leq \lfloor \frac{|w|}{2} \rfloor$, then $w$ has a border of length $\ell$ if and only if $w$ has strong period $|w| - \ell$.*

## 3 Period and Border Correlations

Binary and ternary period correlations are defined according to the following definition.

**Definition 3.1** ( [10]). *(a) The* binary period correlation *of a partial word $w$ of length $n$ satisfying $\mathbf{SP}(w) = \mathbf{WP}(w)$ is the binary vector $\mathbf{2pc}_w$ of length $n$ such that $\mathbf{2pc}_w[0] = 1$ and for $1 \leq i < n$:*

$$\mathbf{2pc}_w[i] = \begin{cases} 1 & \textit{if } \ i \in \mathbf{SP}(w), \\ 0 & \textit{otherwise.} \end{cases}$$

*(b) The* ternary period correlation *of a partial word $w$ of length $n$ is the ternary vector $\mathbf{3pc}_w$ of length $n$ such that $\mathbf{3pc}_w[0] = 1$ and for $1 \leq i < n$:*

$$\mathbf{3pc}_w[i] = \begin{cases} 1 & \textit{if } \ i \in \mathbf{SP}(w), \\ 2 & \textit{if } \ i \in \mathbf{WP}(w) \setminus \mathbf{SP}(w), \\ 0 & \textit{otherwise.} \end{cases}$$

Considering the partial word $abaca\diamond\diamond acaba$ which has strong periods 9 and 11 (and 12) and strictly weak period 5, its ternary period correlation vector is 100002000101. For any partial word $w$, note that both $w$ and its reversal, $\text{rev}(w)$, share the same binary and ternary period correlations.

We say that a binary or ternary vector of length $n$ is a *valid* binary or ternary period correlation if it is the binary or ternary period correlation of some partial word of length $n$. The following theorem implies that the sets of all valid binary and ternary period correlations are independent of the alphabet size.

**Theorem 3.2** ( [10]). *If $w$ is a partial word over an alphabet $\Sigma$, $|\Sigma| \geq 2$, then there exists a partial word $w'$ of length $|w|$ over the binary alphabet $\{a, b\}$ such that $\mathbf{SP}(w') = \mathbf{SP}(w)$ and $\mathbf{WP}(w') = \mathbf{WP}(w)$.*

Border correlations are now defined according to the following definition.

**Definition 3.3.** *The* border correlation *of a partial word $w$ of length $n$ is the binary vector $\mathbf{bc}_w$ of length $n$ such that $\mathbf{bc}_w[n-1] = 1$ and for $0 \leq i < n-1$:*

$$\mathbf{bc}_w[i] = \begin{cases} 1 & \textit{if } \ w \ \textit{ has a border of length } \ i+1, \\ 0 & \textit{otherwise.} \end{cases}$$

Considering again the partial word $abaca\diamond\diamond acaba$, its border correlation vector is 101000100001.

The following theorem gives a characterization of the possible border length sets of partial words of arbitrary length.

**Theorem 3.4.** *Given a binary number $q = q_0q_1 \cdots q_{n-1}$ with $q_{n-1} = 1$, the binary partial word $w = w[0..n-1]$ defined by $w[n-1] = b$ and for $0 \le i < n-1$,*

$$w[i] = \begin{cases} \diamond & \text{if } q_i = 1, \\ a & \text{otherwise} \end{cases}$$

*satisfies the equation $\mathbf{bc}_w = q$. In other words, every binary number that ends in 1 is a valid border correlation.*

*Proof.* Let $i$ be such that $1 \le i < n$. Suppose $q_{i-1} = 1$. From the definition of $w$, $w[i-1] = \diamond$, and $w[n-1] = b$, so $w[i-1] \uparrow w[n-1]$. If $i = 1$, then $q_{i-1} = q_0 = 1$, so $w[0] = \diamond$, and $w$ has a border of length 1. If $i > 1$, then from the definition of $w$, $w[0] \cdots w[n-2]$ is unary, so $w[0..i-2] \uparrow w[n-i..n-2]$. Thus, $w[0..i-1] \uparrow w[n-i..n-1]$, and $w$ has a border of length $i$ which means $\mathbf{bc}_w[i-1] = 1$. Conversely, suppose $\mathbf{bc}_w[i-1] = 1$. Then $w$ has a border of length $i$ implying in particular that $w[i-1] \uparrow w[n-1]$. Since $w[n-1] = b$ and $w[i-1] \in \{a, \diamond\}$, we get that $w[i-1] = \diamond$ which means $q_{i-1} = 1$.

Now, suppose $q_{i-1} = 0$. From the definition of $w$, $w[i-1] = a \not\uparrow b = w[n-1]$. Thus, $w$ does not have a border of length $i$ which means $\mathbf{bc}_w[i-1] = 0$. Conversely, suppose $\mathbf{bc}_w[i-1] = 0$. Then $w$ does not have a border of length $i$ implying that $w[0..i-1] \not\uparrow w[n-i..n-1]$. Since $w[0..i-2]$ and $w[n-i..n-2]$ are in $\{a, \diamond\}^*$, we deduce that $w[i-1] \not\uparrow w[n-1]$, so $w[i-1] = a$ and $w[n-1] = b$. This implies that $q_{i-1} = 0$.

Since $q_{i-1} = 1$ if and only if $\mathbf{bc}_w[i-1] = 1$ and $q_{i-1} = 0$ if and only if $\mathbf{bc}_w[i-1] = 0$, and $\mathbf{bc}_w$ and $q$ both end in 1 by definition, $\mathbf{bc}_w = q$. □

Given the binary number 10101, the previous theorem builds the partial word $\diamond a \diamond ab$ having 10101 as its border correlation.

Note that Theorem 3.4 implies that the reverse of a valid binary period correlation is a valid border correlation. However, the converse is not true in general. There are 12 binary period correlations of partial words $w$ with one hole of length 7 over a binary alphabet, while there are 17 border correlations of such partial words. Table 1 shows the valid binary period correlations whose reversals are valid border correlations. However, Table 2 shows the valid border correlations which do not lead to valid binary period correlations when reversed.

**Theorem 3.5.** *There is a one-to-one correspondence between the set of valid border correlations and the set of valid ternary period correlations of a given length.*

| $\mathbf{2pc}_w$ | rev($\mathbf{2pc}_w$) | $\mathbf{2pc}_w$ | rev($\mathbf{2pc}_w$) |
|---|---|---|---|
| 1000000 | 0000001 | 1000110 | 0110001 |
| 1000001 | 1000001 | 1000111 | 1110001 |
| 1000010 | 0100001 | 1001001 | 1001001 |
| 1000011 | 1100001 | 1001011 | 1101001 |
| 1000100 | 0010001 | 1010101 | 1010101 |
| 1000101 | 1010001 | 1111111 | 1111111 |

Table 1: Valid binary period correlations of length 7 whose reversals are valid border correlations.

| $\mathbf{bc}_w$ | rev($\mathbf{bc}_w$) |
|---|---|
| 0000011 | 1100000 |
| 0000101 | 1010000 |
| 0001001 | 1001000 |
| 1000101 | 1010001 |
| 0101001 | 1001010 |

Table 2: Valid border correlations of length 7 whose reversals are not valid binary period correlations.

*Proof.* We show that two partial words of length $n$ have the same border correlation if and only if they have the same ternary period correlation. Assume two partial words have the same border correlation. From Lemma 2.2(a), we know the partial words' set of weak periods. From this set, we can then determine whether a weak period $p$ is a strong period using Lemma 2.1: we check whether the multiples of $p$ are all in the weak period set; if so, we set $\mathbf{3pc}[p] = 1$, and if not, we set $\mathbf{3pc}[p] = 2$. The two partial words both have the unique ternary period correlation determined in this way, $\mathbf{3pc}[0]\,\mathbf{3pc}[1]\cdots\mathbf{3pc}[n-1]$, where $\mathbf{3pc}[0] = 1$.

For the reverse direction, assume two partial words have the same ternary period correlation. From it, we know the partial words' weak period set, and can determine the set of border lengths from Lemma 2.2(a). Set $\mathbf{bc}[n-1] = 1$, $\mathbf{bc}[i] = 1$ if $i+1$ is in the set of border lengths, and $\mathbf{bc}[i] = 0$ otherwise. Both the partial words have the border correlation $\mathbf{bc}[0]\,\mathbf{bc}[1]\cdots\mathbf{bc}[n-1]$. □

The proof of Theorem 3.5 leads to an algorithm that can generate a list of all the valid border correlations of a given length, and their corresponding ternary period correlations. The following example illustrates the algorithm.

**Example 3.6.** *Suppose we want to find the ternary period correlation* $\mathbf{3pc}$ *corresponding to the border correlation* 1111001111*. We first assign* $\mathbf{3pc}[0] = 1$ *by definition. The border correlation gives the border set* $\{1, 2, 3, 4, 7, 8, 9\}$*, which corresponds to the weak period set* $\{9, 8, 7, 6, 3, 2, 1\}$*. Thus, 4 and 5 are not weak periods, so* $\mathbf{3pc}[4]$ *and* $\mathbf{3pc}[5]$ *are zeros. Then we check whether each weak period's multiples are also weak periods. All of the multiples of 9, 8, 7, 6, and 3 are also in the weak period set, so* $\mathbf{3pc}[3]$*,* $\mathbf{3pc}[6]$*,* $\mathbf{3pc}[7]$*,* $\mathbf{3pc}[8]$*, and* $\mathbf{3pc}[9]$ *are all ones. The weak period set does not contain all of the multiples of 2 or of 1, so* $\mathbf{3pc}[1]$ *and* $\mathbf{3pc}[2]$ *are twos. Now* $\mathbf{3pc}$ *has been determined to be* 1221001111*.*

The following result follows from Theorem 3.2 and Theorem 3.5.

**Theorem 3.7.** *If $w$ is a partial word over an alphabet $\Sigma$, $|\Sigma| \geq 2$, then there exists a binary partial word $w'$ such that* $\mathbf{bc}_{w'} = \mathbf{bc}_w$*.*

# 4 Refined Border Correlations

We show some connections between a partial word's ternary period correlation and its refined border correlation. This problem was suggested in [7].

Let $\sigma : \Sigma_\diamond^* \to \Sigma_\diamond^*$, where $\sigma(\varepsilon) = \varepsilon$ and $\sigma(cw) = wc$ for all $w \in \Sigma_\diamond^*$ and $c \in \Sigma_\diamond$, be the *shift function*. Inductively, we define the *i-shift* of a partial word $w$, denoted by $\sigma^i(w)$, by $\sigma^0(w) = w$ if $i = 0$ and by $\sigma^i(w) = \sigma(\sigma^{i-1}(w))$ if $i > 0$. The partial words $u$ and $v$ are said to be *conjugates* if $u = \sigma^i(v)$ for some $i \geq 0$. Refined border correlations are now defined according to the following definition, that is, the refined border correlation is the vector which specifies for the $i$-shift the length of a minimal border.

**Definition 4.1** ( [20])**.** *The* refined border correlation *of a partial word $w$ of length $n$ is the vector $\mathbf{rbc}_w$ of length $n$ such that for $0 \leq i < n$:*

$$\mathbf{rbc}_w[i] = \begin{cases} \textit{length of a minimal border of } \sigma^i(w) & \textit{if } \sigma^i(w) \textit{ is bordered,} \\ 0 & \textit{otherwise.} \end{cases}$$

Note that $\sigma(\mathbf{rbc}_w) = \mathrm{rev}(\mathbf{rbc}_{\mathrm{rev}(w)})$, a fact which is merely a rephrasing of [7, Lemma 4.4].

The next theorem shows that the refined border correlation of a partial word $w$ at position $i$ can be easily calculated from the position of the last non-zero digit of the ternary period correlation of the $i$-shift of $w$.

**Theorem 4.2.** *Let $w$ be a partial word of length $n$. For $0 \leq i < n$, $\mathbf{rbc}_w[i] = (n - k_i) \bmod n$ where $k_i$ is the position of the last non-zero digit of $\mathbf{3pc}_{\sigma^i(w)}$.*

*Proof.* Let us first show that the equality $\mathbf{rbc}_w[0] = (n - k_0) \bmod n$ holds.

Suppose that $w$ is bordered. We prove that the maximal weak period of $w$, the greatest integer that is a weak period, is $n - \mathbf{rbc}_w[0]$. To see this, from Lemma 2.2, $w$ has a weak period $n - \mathbf{rbc}_w[0]$. To show that $n - \mathbf{rbc}_w[0]$ is maximal, assume for the sake of contradiction that $w$ has a weak period $p$ that is greater than $n - \mathbf{rbc}_w[0]$ by a positive integer $q$. Then $p = n - \mathbf{rbc}_w[0] + q = n - (\mathbf{rbc}_w[0] - q)$ is a weak period of $w$, so, from Lemma 2.2, $w$ has a border of length $\mathbf{rbc}_w[0] - q < \mathbf{rbc}_w[0]$. This contradicts our designation of $\mathbf{rbc}_w[0]$ as the length of a minimal border of $w$. Thus $k_0 = n - \mathbf{rbc}_w[0]$ and since $0 < \mathbf{rbc}_w[0] < n$, $\mathbf{rbc}_w[0] = (n - k_0) \bmod n$.

Suppose now that $w$ is unbordered. Then $\mathbf{rbc}_w[0] = 0$ and $w$ does not have a border of length $\ell$ such that $0 < \ell < n$. So, by Lemma 2.2, $w$ does not have a weak period $n - \ell$ such that $0 < \ell < n$. Equivalently, $\mathbf{3pc}[j] = 0$ for all $0 < j < n$. Thus $k_0 = 0$ and $\mathbf{rbc}_w[0] = (n - k_0) \bmod n$.

Therefore, $\mathbf{rbc}_{\sigma^i(w)}[0] = (n - k_i) \bmod n$ for all $0 \leq i < n$. Since $\mathbf{rbc}_{\sigma^i(w)} = \sigma^i(\mathbf{rbc}_w)$, we can equate the first digits of both sides of the equality, so $\mathbf{rbc}_{\sigma^i(w)}[0] = \mathbf{rbc}_w[i]$. Thus, $\mathbf{rbc}_w[i] = (n - k_i) \bmod n$. □

| $i$ | $\sigma^i(w)$ | $\mathbf{3pc}_{\sigma^i(w)}$ | $\mathbf{rbc}_{\sigma^i(w)}$ |
|---|---|---|---|
| 0 | $abba\diamond cdb\diamond bc\diamond$ | 10000000000**1** | **1**01011051121 |
| 1 | $bba\diamond cdb\diamond bc\diamond a$ | **1**00000000000 | **0**10110511211 |
| 2 | $ba\diamond cdb\diamond bc\diamond ab$ | 10000000010**1** | **1**01105112110 |
| 3 | $a\diamond cdb\diamond bc\diamond abb$ | **1**00000000000 | **0**11051121101 |
| 4 | $\diamond cdb\diamond bc\diamond abba$ | 10000000000**1** | **1**10511211010 |
| 5 | $cdb\diamond bc\diamond abba\diamond$ | 10000000000**1** | **1**05112110101 |
| 6 | $db\diamond bc\diamond abba\diamond c$ | **1**00000000000 | **0**51121101011 |
| 7 | $b\diamond bc\diamond abba\diamond cd$ | 1000000**1**0000 | **5**11211010110 |
| 8 | $\diamond bc\diamond abba\diamond cdb$ | 10000000001**1** | **1**12110101105 |
| 9 | $bc\diamond abba\diamond cdb\diamond$ | 10000000001**1** | **1**21101011051 |
| 10 | $c\diamond abba\diamond cdb\diamond b$ | 1000000000**1**0 | **2**11010110511 |
| 11 | $\diamond abba\diamond cdb\diamond bc$ | 10000000000**1** | **1**10101105112 |

Table 3: Illustrating Theorem 4.2 with partial word $w = abba\diamond cdb\diamond bc\diamond$ of length $n = 12$; the bold in the second to last column is the position $k_i$ of the last non-zero digit of $\mathbf{3pc}_{\sigma^i(w)}$, while the bold in the last column is $(n - k_i) \bmod n = \mathbf{rbc}_{\sigma^i(w)}[0] = \mathbf{rbc}_w[i]$.

Table 3 illustrates Theorem 4.2.

We next give some other connections between refined border correlations and ternary period correlations. The ternary period correlation of a given partial word at a specific position, under some conditions, implies lower and upper bounds on the digits of its refined border correlation at a specific range of positions.

**Remark 4.3.** *Let $w$ be a partial word of length $n$. The equation $\mathbf{rbc}_w[0] = 1$ is equivalent to $w[0] \uparrow w[n-1]$, and the equation $\mathbf{rbc}_w[i] = 1$, for any $i$ such that $1 \leq i < n$, is equivalent to $w[i-1] \uparrow w[i]$.*

**Proposition 4.4.** *If the refined border correlation of a partial word $w$ contains exactly one digit, $\mathbf{rbc}_w[\ell]$, not equal to 1, then $\mathbf{rbc}_w[\ell] \neq 0$.*

*Proof.* Let $w$ be a partial word of length $n$, and let $I$ be the set of integers $i$ such that $\mathbf{rbc}_w[i] = 1$. Suppose $I = [0..n-1] \setminus \{\ell\}$. For $i \in I$, let $j = (i-1) \bmod n$. From Remark 4.3, $w[i] \uparrow w[j]$, and the only consecutive characters of $w$ that are not compatible are $w[\ell]$ and $w[\ell-1]$ (if $\ell = 0$, replace $\ell - 1$ with $n-1$). Therefore, if $\ell \neq 0$, then $w[\ell..n-1]w[0..\ell-2] \uparrow w[\ell+1..n-1]w[0..\ell-1]$. Thus, $\sigma^\ell(w) = w[\ell..n-1]w[0..\ell-1]$ is bordered by $w[\ell..n-1]w[0..\ell-2] \vee w[\ell+1..n-1]w[0..\ell-1]$. Similarly, if $\ell = 0$, then

$w[0..n-2] \uparrow w[1..n-1]$, so $\sigma^\ell(w) = w$ is bordered by $w[0..n-2] \vee w[1..n-1]$. Since $\sigma^\ell(w)$ is bordered, $\mathbf{rbc}_w[\ell] \neq 0$. □

Notice that all unary partial words of length $n$ have ternary period correlations and refined border correlations equal to $1^n$, and that all partial words with $1^n$ as a ternary period correlation are unary.

**Proposition 4.5.** *If $w$ is a non-unary partial word of length $n$, then $\mathbf{3pc}_w[1] = 2$ if and only if $\mathbf{rbc}_w[i] = 1$ for all $1 \leq i < n$.*

*Proof.* For the forward implication, suppose that $\mathbf{3pc}_w[1] = 2$. From the definition of ternary period correlations, $w$ has a strictly weak period of 1, so every two consecutive characters of $w$ are compatible; i.e., given $i$ such that $1 \leq i < n$, $w[i-1] \uparrow w[i]$. From Remark 4.3, $\mathbf{rbc}_w[i] = 1$ for all $1 \leq i < n$.

The proof of the converse is similar. Suppose $\mathbf{rbc}_w[i] = 1$ for all $1 \leq i < n$. Then, from Remark 4.3, $w[i-1] \uparrow w[i]$ for all $1 \leq i < n$. All consecutive characters of $w$ are compatible, so $w$ has a weak period of 1; as $w$ is non-unary, it is strictly weak. Thus, $\mathbf{3pc}_w[1] = 2$. □

To illustrate Proposition 4.5, consider the partial word $w = a\diamond bb\diamond c$ which has ternary period correlation 120010 and refined border correlation 211111.

**Proposition 4.6.** *For any partial word $w$ of length $n$, the following hold for $0 < p < n$:*

*(a) If $\mathbf{3pc}_w[p] \neq 0$, then $1 \leq \mathbf{rbc}_w[i] \leq p$ for all $p \leq i \leq n-p$.*

*(b) If $\mathbf{3pc}_w[p] = 1$ and $p$ divides $n$, then $1 \leq \mathbf{rbc}_w[i] \leq p$ for all $0 \leq i < n$.*

*Proof.* We prove the case where $p \geq 2$ (the case where $p = 1$ follows easily from Remark 4.3). For (a), suppose that $\mathbf{3pc}_w[p] = 2$. From the definition of ternary period correlation, $w$ has a strictly weak period of $p$, so, given $j$ such that $p \leq j < n$, $w[j-p] \uparrow w[j]$. The conjugate $\sigma^i(w) = w[i..n-1]w[0..i-1]$ has border $w[i..i+p-1] \vee w[i-p..i-1]$ for $p \leq i \leq n-p$. This border has length $p$, so $1 \leq \mathbf{rbc}_w[i] \leq p$ for $p \leq i \leq n-p$. Now, suppose that $\mathbf{3pc}_w[p] = 1$. It is still true that given $j$ such that $p \leq j < n$, $w[j-p] \uparrow w[j]$, so we reach the same conclusion.

For (b), suppose that $\mathbf{3pc}_w[p] = 1$ and $p$ divides $n$, and let $n = mp$ for some $m \geq 2$. Here $w$ has a (strong) period of $p$, so the factors

$$w[0..p-1], w[p..2p-1], \ldots, w[(m-1)p..mp-1]$$

are pairwise compatible. In other words, there is a partial word $u$ of length $p$, the least upper bound of this set of factors of $w$, such that $w \subset u^m$. We have

$$\begin{aligned}\sigma^i(w) &\subset u[i \bmod p..p-1]u^{m-1}u[0..(i-1) \bmod p] \\ &= (u[i \bmod p..p-1]u[0..(i-1) \bmod p])^m.\end{aligned}$$

Thus, $\sigma^i(w)$ has $u[i \bmod p..p-1]u[0..(i-1) \bmod p]$ as a border, which has length $p$, so $1 \leq \mathbf{rbc}_w[i] \leq p$ for all $0 \leq i < n$. □

To illustrate Proposition 4.6, consider the following example.

**Example 4.7.** *Let $w = aba\diamond\diamond aca$, which has ternary period correlation $\mathbf{3pc}_w = 10220001$ and refined border correlation $\mathbf{rbc}_w = 15211125$. Here $\mathbf{3pc}_w[p] = 2$ for $p = 2$, and $n = 8$, so for $2 \leq i \leq 6$, we have $1 \leq \mathbf{rbc}_w[i] \leq 2$. For $p = 3$, we have a weaker result: for $3 \leq i \leq 5$, $1 \leq \mathbf{rbc}_w[i] \leq 3$. The converse of Proposition 4.6(a) is false, e.g., if $w = \diamond\diamond ab\diamond a$, then $\mathbf{rbc}_w = 111211$. Thus $1 \leq \mathbf{rbc}_w[i] \leq 2$ for all $2 \leq i \leq n-2 = 4$. However, $\mathbf{3pc}_w = 100111$, so $\mathbf{3pc}_w[2] = 0$. The partial word $ab\diamond abaaba$ has ternary period correlation 100100101 and refined border correlation 131123133 satisfying Proposition 4.6(b).*

# 5 Valid Colorings of Undirected Graphs Using the Subgraph Component Polynomial

We adopt the notation $\Sigma_y$ for an arbitrary $y$-letter alphabet. We start with some definitions.

**Definition 5.1.** *Let $G = (V, E)$ be an undirected graph such that $V = [0..n-1]$:*

- *A* valid coloring *over $\Sigma_y$ of $G$ is one in which the colors of adjacent vertices are the same or one is the hole color. Here, the colors are the $y$ letters of the alphabet $\Sigma_y$ as well as the hole color, $\diamond$. The number of* valid colorings *over $\Sigma_y$ of $G$, denoted by $\mathrm{VC}_y(G)$, is the number of partial words $w$ of length $n$ over $\Sigma_y$ such that if $\{i, j\} \in E$, then $w[i] \uparrow w[j]$.*

- *If $s$ is a sequence of pairs of the form $(i, c)$, where $i \in V$ and $c \in \Sigma_y \cup \{\diamond\}$, then the number of $s$-*valid colorings *over $\Sigma_y$ of $G$, denoted by $s$-$\mathrm{VC}_y(G)$, is the number of valid colorings $w$ over $\Sigma_y$ of $G$ subject to the restrictions imposed by $s$, that is, if $(i, c)$ is in $s$, then $w[i] = c$. If $s$ consists of a singleton $(i, c)$, then we simply write $(i, c)$-$\mathrm{VC}_y(G)$.*

**Definition 5.2.** *Let $G = (V, E)$ be an undirected graph such that $V = [0..n-1]$. The* connected component vector *of $G$, denoted by $\mathbf{cc}_G$, is a* *vector of length* $n+1$ *such that for* $0 \le i \le n$, $\mathbf{cc}_G[i]$ *is the number of ways to remove a set of vertices from* $G$ *such that the resulting induced subgraph has* $i$ *connected components.*

The following example illustrates our definitions.

**Example 5.3.** *Consider the graph* $G = (\{0,1,2,3,4\}, \{\{0,2\},\{1,3\},\{2,4\}\})$. *Then* $\mathrm{VC}_2(G) = 119$, *one of the valid colorings over* $\{a,b\}$ *of* $G$ *being* $aa\diamond ab$ *(we show later how to calculate 119). The connected component vector of* $G$ *is* 19(19)300*. For example, the three ways to remove a set of vertices from* $G$ *to get induced subgraphs with three connected components are to remove the sets* $\{2\}$, $\{1,2\}$, *and* $\{2,3\}$, *so* $\mathbf{cc}_G[3] = 3$.

Let $P(G;y) = \sum_{i=0}^{n} y^i \, \mathbf{cc}_G[i]$ be the generating function for the connected component vector $\mathbf{cc}_G$ of an undirected graph $G$ on $n$ vertices. We can obtain this by plugging in $x = 1$ into the *subgraph component polynomial* $Q(G;x,y)$, as defined by Tittmann et al. [29]. The next theorem shows that there are $P(G;y)$ partial words of length $n$ over $\Sigma_y$ that satisfy the required compatibilities to be valid colorings over $\Sigma_y$ of an undirected graph $G$ on $n$ vertices.

**Theorem 5.4.** *If* $G = (V,E)$ *is an undirected graph with* $V = [0..n-1]$, *then*

$$\mathrm{VC}_y(G) = P(G;y) = \sum_{i=0}^{n} y^i \, \mathbf{cc}_G[i].$$

*Proof.* We are counting the number of valid colorings over $\Sigma_y$ of $G = (V,E)$, i.e., the number of partial words $w$ of length $n$ over $\Sigma_y$ such that if $\{i,j\}$ is an edge in $E$, then $w[i]$ and $w[j]$ are compatible. Removing a set of vertices $V'$ from $G$ is equivalent to coloring the vertices in $V'$ with the hole color (and to coloring no other vertex with the hole color); this is also equivalent to putting holes in the positions which correspond to vertices in $V'$ (and to putting holes only in those positions). Suppose that the resulting induced subgraph has $i$ connected components. Each of these components is associated with a color of $\Sigma_y$, thus there are $y^i$ possible color distributions for the $i$ connected components. The coefficient of $y^i$ is the number of ways to remove a set of vertices from $G$ such that the resulting induced subgraph has $i$ connected components, which is $\mathbf{cc}_G[i]$. □

Results on subgraph component polynomials facilitate the computation of connected component vectors. Note that if $G = G_1 \sqcup G_2 \sqcup \cdots \sqcup G_n$ is the disjoint union of the graphs $G_1, G_2, \ldots, G_n$, then $P(G;y) = \prod_{j=1}^{n} P(G_j;y)$.

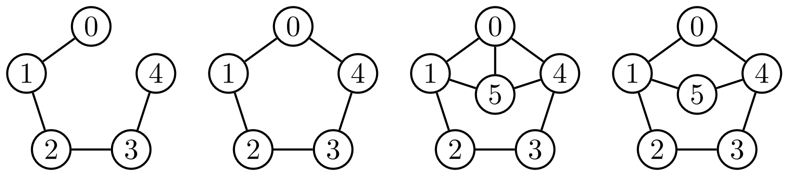


Figure 1: From left to right: the path graph $P_5$, the cycle graph $C_5$, the broken wheel graph $W_{3,2}$, and the broken wheel graph $W_{1,2,1,1}$, also written as $W_{1,1,1,2}$. They are associated with the border correlations 00011, 10011, 110011, and 010011, respectively (see Section 6).

We adopt the following notations for graphs on $n$ vertices: $E_n$ is the empty graph, $K_n$ is the complete graph, $P_n$ is a path graph, and $C_n$ is the cycle graph. A wheel graph (respectively, broken wheel graph) of order $n$ is a graph consisting of a cycle graph on $n$ vertices and an additional vertex, called a hub, that is adjacent to all (respectively, at least one) of the vertices in the cycle graph. We denote a wheel graph of order $n$ by $W_n$ and a broken wheel graph of order $n$ by $W_s$, where $s$ is a sequence of numbers whose sum is $n$. The first number in $s$ is a number of consecutive hub-adjacent vertices, the second number in $s$ is the number of consecutive non-hub-adjacent vertices following the hub-adjacent ones, and so on around the cycle. Figure 1 shows a path, a cycle, and two broken wheels.

Although the subgraph component polynomials of $E_n$, $K_n$, $P_n$, $C_n$, and $W_n$ have been studied in [29] and formulas have been given, we either give new formulas for previously considered cases or give new proofs of known formulas using partial words.

**Proposition 5.5** ( [29])**.** *The following hold:*

*(a)* $P(E_n; y) = (1 + y)^n$*,*

*(b)* $P(K_n; y) = 1 + (2^n - 1)y$*,*

*Proof.* For (a), each of the $n$ isolated vertices can be colored in $1 + y$ ways. For (b), there is one valid coloring in which all of the vertices are $\diamond$'s. In all other valid colorings, at least one vertex has a non-hole color, say $c$. Then all of the other vertices can either be $\diamond$'s or also have the color $c$, because they are adjacent to this vertex. Given that at least one vertex has the non-hole color $c$, there are $2^n - 1$ valid colorings, because we can choose $c$ or $\diamond$ independently for each vertex, and then subtract the one case in which all vertices are $\diamond$'s. There are $y$ possible non-hole colors. $\square$

For path graphs, we can derive the following formulas.

**Proposition 5.6.** *The following equality holds:*

$$P(P_n; y) = 1 + \sum_{h=0}^{n-1} \sum_{i=0}^{h} \binom{n-h-1}{i} \binom{h+1}{h-i} y^{i+1}.$$

*Moreover, let $c$ be one of $y$ non-hole colors, let $s = \langle (0,c), (n-1,c) \rangle$, and let $s' = \langle (0,c) \rangle$. Then the following hold:*

*(a)* $s\text{-}\mathrm{VC}_y(P_n) = 1 + \sum_{h=1}^{n-2} \sum_{i=1}^{h} \binom{n-h-1}{i} \binom{h-1}{h-i} y^{i-1}$,

*(b)* $s'\text{-}\mathrm{VC}_y(P_n) = 1 + \sum_{h=1}^{n-1} \sum_{i=1}^{h} \binom{n-h-1}{i} \binom{h}{h-i} y^{i}$.

*Proof.* We are actually counting the number of weakly 1-periodic partial words of length $n$ over $y$ letters. The number of valid colorings for $P_n$ is the sum of the number of valid colorings for each number of holes $h$. If $h = n$, there is exactly one valid coloring. So let $h \in [0..n-1]$, which is the index of the outer sum. All adjacent vertices must be compatible, so the graph is partitioned into $i+1$ blocks separated by $i$ strings of $\diamond$'s, where $i \in [0..h]$ is the index of the inner sum. Each block is a unary total word having a single non-hole color. There are $i+1$ blocks, and $y$ color choices, thus we will multiply by $y^{i+1}$.

We first separate the $i+1$ blocks by $i$ holes, called "forced" holes because they are required to keep the blocks distinct. We need to have at least one non-hole vertex in each block, so we have $n-h-(i+1)$ remaining non-hole vertices to be placed into any block. There are $n-h-1$ available positions for the remaining $n-h-(i+1)$ non-hole vertices and the $i$ forced holes. Choosing where to place the $n-h-(i+1)$ non-hole vertices is equivalent to choosing the positions of the $i$ forced holes. There are $n-h-1$ positions and $i$ forced holes, so $\binom{n-h-1}{i}$ gives the number of ways to distribute the non-hole vertices. Once the lengths of the blocks are determined, we count the number of ways to distribute the $h-i$ remaining holes among the positions adjacent to existing holes. Along with the remaining holes, we also have to place the $i+1$ blocks. There are $(h-i)+(i+1) = h+1$ available positions for the blocks and remaining holes. We choose $h-i$ positions for the remaining holes from the $h+1$ available ones, so there are $\binom{h+1}{h-i}$ ways to distribute the remaining holes.

We prove (a) (the proof for (b) is similar). The index $i$ in the inner sum is the number of forced holes. If $i = 0$, there is only one block, and from the constraint that $P_n$ begins and ends with the same non-hole color $c$, $P_n$ must be colored with $c$. Notice that $i = 0$ implies $h = 0$, so we add 1 for this case. So, let $h \in [1..n-2]$ and $i \in [1..h]$. The number of ways to distribute the non-hole vertices into $i+1$ blocks is $\binom{n-h-1}{i}$, as before. The number of ways to distribute the $h-i$ remaining holes, the same number of remaining holes as before, is as follows. The number of available positions is decreased by two because we cannot place a remaining hole before the first block or after the last block. Thus, we have $(h-i)+(i+1)-2 = h-1$ available positions, and we are choosing $h-i$ of them. Now that we know the arrangement of the blocks and strings of holes, we must choose a single non-hole color for each block. Our constraint implies that the first and last blocks have the color $c$; thus, there are $i-1$ block color choices and $y$ color choices for each, so we multiply by $y^{i-1}$. □

For cycle graphs, wheel graphs, and broken wheel graphs, we can derive the following formulas.

**Proposition 5.7.** *If $c$ is one of $y$ non-hole colors, then the following hold:*

*(a)* $P(C_n; y) = 2P(P_{n-1}; y) - P(P_{n-2}; y) + (s\text{-}\mathrm{VC}_y(P_n))y$, *where* $s = \langle (0, c), (n-1, c)\rangle$,

*(b)* $P(W_n; y) = P(C_n; y) + 2^n y$,

*(c)* $P(W_{1,n-1}; y) = P(C_n; y) + (s\text{-}\mathrm{VC}_y(P_{n+1}) + P(P_{n-1}; y))y$, *where* $s = \langle (0, c), (n, c)\rangle$,

*(d) If* $n - r > 1$, *then* $P(W_{n-r,r}; y)$ *is*

$$P(C_n; y) + (s\text{-}\mathrm{VC}_y(P_{r+2}) + 2(s'\text{-}\mathrm{VC}_y(P_{r+1})) + P(P_r; y))y2^{n-r-2},$$

*where* $s = \langle (0, c), (r+1, c)\rangle$ *and* $s' = \langle (0, c)\rangle$.

*Proof.* For (a), a valid coloring of $C_n$ is equivalent to a valid coloring of $P_n$ with the property that either the first and last vertices have the same non-hole color, or at least one of them is a hole. If the first vertex is a hole, then the number of valid colorings is $P(P_{n-1}; y)$, and similarly if the last vertex is a hole. These counts include the valid colorings in which both the first and last vertices are holes; there are $P(P_{n-2}; y)$ of these which me must subtract. We must also add the number of valid colorings where the first and last vertices have the same non-hole color, say $c$, which is $s\text{-}\mathrm{VC}_y(P_n)$, given in Proposition 5.6(a), and there are $y$ non-hole colors.

For (b), when the hub of $W_n$ is a hole, it imposes no restrictions on the coloring of the other $n$ vertices, so the number of valid colorings is $P(C_n; y)$.

When the hub is any of the $y$ non-hole colors, say $c$, each other vertex can be either $c$ or $\diamond$, so there are $2^n$ valid colorings.

For (c), when the hub is a hole, there are $P(C_n; y)$ valid colorings of $W_{1,n-1}$. When the hub is the non-hole color $c$, the single vertex $v$ adjacent to the hub can be either a hole or $c$. If it is a hole, then it imposes no restrictions on the remaining $n-1$ vertices in the cycle, which form a path graph $P_{n-1}$ and can be colored in $P(P_{n-1}; y)$ many ways. If it has the non-hole color $c$, then the colorings of the remaining vertices have the same restrictions as if they were the $n-1$ vertices of degree two in a path graph $P_{n+1}$ where the vertices of degree one are $c$. Thus the number of valid colorings in this case is $s\text{-}\mathrm{VC}_y(P_{n+1})$. The total number of colorings when the hub is $c$ is $s\text{-}\mathrm{VC}_y(P_{n+1}) + P(P_{n-1}; y)$, and there are $y$ non-hole colors.

For (d), when the hub is a hole, there are $P(C_n; y)$ valid colorings of the broken wheel. When the hub is $c$, the vertices adjacent to the hub can be either holes or $c$. If the vertices adjacent to the hub are labeled $0, \ldots, n-r-1$, then the colorings of the vertices $1, \ldots, n-r-2$ do not affect the colorings of the vertices $n-r, \ldots, n-1$. The vertices $0$ and $n-r-1$ affect the colorings of the vertices $n-r, \ldots, n-1$, though, so we consider the four possible colorings of $0$ and $n-r-1$ (each of them can be either $c$ or $\diamond$). If they are both $c$, then the vertices $n-r, \ldots, n-1$ are colored like the $r$ vertices of degree two in $P_{r+2}$, in $s\text{-}\mathrm{VC}_y(P_{r+2})$ ways. If one of them is $c$ and the other is a hole, then the vertices $n-r, \ldots, n-1$ are colored like $r$ consecutive vertices of $P_{r+1}$, where the remaining vertex is $c$, in $s'\text{-}\mathrm{VC}_y(P_{r+1})$ ways. If both $0$ and $n-r-1$ are holes, then the vertices $n-r, \ldots, n-1$ can be colored like $P_r$, in $P(P_r; y)$ ways. The total number of valid colorings when the hub is $c$ is $s\text{-}\mathrm{VC}_y(P_{r+2}) + 2(s'\text{-}\mathrm{VC}_y(P_{r+1})) + P(P_r; y)$, multiplied by $2^{n-r-2}$ for the independent choices of $c$ or $\diamond$ for each of the vertices $1, \ldots, n-r-2$. Finally, we multiply this number by $y$ for the total number of valid colorings when the hub is a non-hole color. $\square$

The graph $G$ may not have any of the above forms, but its complement $\overline{G}$ may have.

**Proposition 5.8.** *If $G = (V, E)$ is an undirected graph with $V = [0..n-1]$, then $\mathrm{VC}_y(G) = P(G; y) = 1 + (2^n - 1)y + \mathrm{UC}_y(\overline{G})$, where $1 + (2^n - 1)y$ counts the number of valid unary colorings of $G$, i.e., those with at most one non-hole color, and $\mathrm{UC}_y(\overline{G})$ counts the number of valid non-unary colorings of $G$, i.e., those with at least two and at most $y$ non-hole colors.*

We give formulas when $\overline{G} = P_n$ or $\overline{G} = C_n$.

**Proposition 5.9.** *Let $G = (V, E)$ be an undirected graph such that $V = [0..n-1]$. If $\overline{G} = P_n$, then $\mathrm{UC}_y(\overline{G}) = y(y-1)(2n-3)$.*

*Proof.* We claim that the only valid non-unary colorings of $G$ are as follows:

- Two vertices that form a $P_2$ in $\overline{G}$ have the non-hole colors $c_0$ and $c_1$, and all other vertices in $\overline{G}$ have the hole color.
- Three vertices that form a $P_3$ in $\overline{G}$ have the non-hole colors $c_0$, $c_1$, and $c_0$, respectively, and all other vertices in $\overline{G}$ have the hole color.

To prove our claim, first, suppose there are exactly two non-hole vertices in $\overline{G}$. If these vertices are adjacent in $\overline{G}$, they are not adjacent in $G$, so color one vertex $c_0$ and the other $c_1$, which is a valid coloring of $G$. If these vertices are not adjacent in $\overline{G}$, if we color one vertex with $c_0$ and color the other with $c_1$, then this is not a valid coloring of $G$ because they are adjacent in $G$. Next, suppose there are exactly three non-hole vertices in $\overline{G}$. All possible distributions of the three non-hole vertices have already been considered except when these three vertices form a $P_3$ inside $\overline{G}$. Say the three vertices are $i_1$, $i_2$, and $i_3$, which can only be colored with $c_0$, $c_1$, and $c_0$, respectively, because $i_1$ and $i_3$ are adjacent in $G$. Finally, suppose there are four or more non-hole vertices in $\overline{G}$. We cannot have four consecutive non-hole vertices $i_1, i_2, i_3, i_4$ in $\overline{G}$: vertices $i_1$ and $i_2$ must have distinct non-hole colors because they are adjacent in $\overline{G}$, but their colors must be compatible with that of $i_4$, because $i_1$ and $i_2$ are both adjacent to $i_4$ in $G$.

Our claim follows, so we only need to consider the number of non-hole colorings of $P_2$ in $\overline{G}$ and the number of non-hole colorings of $P_3$ in $\overline{G}$. There are $n-1$ of the former and $n-2$ of the latter. The two cases are mutually exclusive, so their sum, $2n-3$, is the total number of possibilities. We are choosing two non-hole colors, and we have $y$ non-hole colors, so we have $y(y-1)$ choices for distinct colorings. □

**Proposition 5.10.** *Let $G = (V, E)$ be an undirected graph such that $V = [0..n-1]$. If $\overline{G} = C_n$, then*

$$\mathrm{UC}_y(\overline{G}) = \begin{cases} y(y-1)(2n+y-2) & \text{if } n = 3, \\ y(y-1)(2n+1) & \text{if } n = 4, \\ 2y(y-1)n & \text{otherwise.} \end{cases}$$

*Proof.* We claim that the only valid non-unary colorings of $G$ are as follows:

- Two vertices that form a $P_2$ in $\overline{G}$ have the non-hole colors $c_0$ and $c_1$, and all other vertices in $\overline{G}$ have the hole color.

- Three vertices that form a $P_3$ in $\overline{G}$ have the non-hole colors $c_0$, $c_1$, and $c_0$ respectively, and all other vertices in $\overline{G}$ have the hole color.

- If $n = 3$ and $y \geq 3$, then the vertices in $\overline{G}$ have the non-hole colors $c_0$, $c_1$ and $c_2$.

- If $n = 4$, then the vertices $0, 1, 2, 3$ in $\overline{G}$ have the non-hole colors $c_0$, $c_1$, $c_0$, and $c_1$ respectively.

To prove our claim, the argument is similar to that of Proposition 5.9 for the first two cases. For the third case, it is possible to color all the vertices distinct non-hole colors (assuming $y \geq 3$), because none of the vertices are adjacent in $G$. For the fourth case, it is possible to color the vertices $0, 1, 2, 3$ with $c_0, c_1, c_0, c_1$, because all of the distinctly colored non-hole vertices are adjacent in $\overline{G}$. We notice that these are the only special cases.

Our claim follows. For any values of $n$ and $y$, the number of $P_2$'s in $\overline{G}$ is $n$, and the number of $P_3$'s in $\overline{G}$ is also $n$. We add the numbers of valid non-unary colorings because they are mutually exclusive, giving us $2n$. For these colorings we assign two of the $y$ colors and there are $y(y-1)$ ways to do this. Thus, the number of valid colorings is $2y(y-1)n$. For the case when $n = 3$, there are $y(y-1)(y-2)$ valid colorings with three distinct non-hole colors and we add this number to $2y(y-1)n$ to get $y(y-1)(2n+y-2)$. Finally for the case when $n = 4$, we need to count the number of valid colorings when the vertices $0, 1, 2, 3$ are colored $c_0, c_1, c_0, c_1$, respectively. Assigning two of the $y$ colors to these vertices yields $y(y-1)$ additional colorings. Thus, adding this number to $2y(y-1)n$ gives the total $y(y-1)(2n+1)$.

Finally, note that $y(y-1)(2n+y-2) = 2y(y-1)n$ when $n = 3$ and $y = 2$. □

If a graph $G$ has only one connected component and $\overline{G}$'s connected components are only path graphs, cycle graphs, and vertices of degree zero, we can calculate the number of valid non-unary colorings of $G$ by looking at each connected component of $\overline{G}$, ignoring the vertices of degree zero, and then add the results because the non-unary colorings of these components are mutually exclusive by the proofs of Propositions 5.9 and 5.10.

The following example illustrates our results.

**Example 5.11.** *Returning to Example 5.3, we calculate* $\mathrm{VC}_y(G)$ *as follows. First, consider* $P_2$ *on vertices 1 and 3, and* $P_3$ *on vertices 0, 2 and 4. We have* $P(P_2; y) = 1 + 3y$ *and* $P(P_3; y) = 1 + 6y + y^2$ *by using Proposition 5.6. Applying Theorem 5.4, we obtain* $\mathrm{VC}_y(G) = P(P_2 \sqcup P_3; y) =$

$P(P_2; y)P(P_3; y) = 1 + 9y + 19y^2 + 3y^3$, *as is also computed from the connected component vector of* $G$. *If* $y = 2$, *we get* $\mathrm{VC}_2(G) = 119$. *We also obtain* $\mathrm{VC}_y(\overline{P_2 \sqcup P_3}) = 1 + (2^5 - 1)y + \mathrm{UC}_y(P_2 \sqcup P_3) = 1 + (2^5 - 1)y + \mathrm{UC}_y(P_2) + \mathrm{UC}_y(P_3) = 1 + 27y + 4y^2$ *by using Proposition 5.8 and Proposition 5.9.*

We can modify Definitions 5.1 and 5.2 to count valid colorings when restricting the number of holes.

**Definition 5.12.** *Let* $G = (V, E)$ *be an undirected graph such that* $V = [0..n - 1]$, *and let* $h$ *be an integer in* $[0..n]$*:*

- *A* $h$-valid coloring *over* $\Sigma_y$ *of* $G$ *is one in which no vertices colored with two different colors from* $\Sigma_y$ *are adjacent and in which exactly* $h$ *vertices are colored with the hole color. The number of* $h$-valid colorings *over* $\Sigma_y$ *of* $G$, *denoted by* $\mathrm{VC}_{h,y}(G)$, *is the number of partial words* $w$ *of length* $n$ *with* $h$ *holes over* $\Sigma_y$ *such that if* $\{i, j\} \in E$, *then* $w[i] \uparrow w[j]$.
- *The* $(n - h)$-connected component vector *of* $G$, *denoted by* $\mathbf{cc}_{n-h,G}$, *is a vector of length* $n + 1$ *such that for* $0 \leq i \leq n$, $\mathbf{cc}_{n-h,G}[i]$ *is the number of ways to remove a set of exactly* $h$ *vertices from* $G$ *such that the resulting induced subgraph has* $i$ *connected components.*

Referring to the subgraph component polynomial $Q(G; x, y)$, we use the notation $Q_{n-h}(G; x, y)$ when we restrict to induced subgraphs with exactly $n - h$ vertices, and similarly for $P_{n-h}(G; y)$.

The proof of the following theorem is similar to that of Theorem 5.4 but we remove sets of exactly $h$ vertices.

**Theorem 5.13.** *If* $G = (V, E)$ *is an undirected graph with* $V = [0..n - 1]$ *and* $h$ *is an integer in* $[0..n]$, *then*

$$\mathrm{VC}_{h,y}(G) = P_{n-h}(G; y) = \sum_{i=0}^{n} y^i \, \mathbf{cc}_{n-h,G}[i].$$

## 6 Population Size

We derive formulas to compute the *population size* over $\Sigma_k$, an arbitrary $k$-letter alphabet, of a given border correlation $\mathbf{bc}$, i.e., the number of partial words over $\Sigma_k$ sharing $\mathbf{bc}$ as their border correlation, which we denote by $\mathrm{PS}_k(\mathbf{bc})$. We also consider refined border correlations. Given $m$ undirected graphs $G_1 = (V, E_1), \ldots, G_m = (V, E_m)$ with the same vertex set $V$, we let $\cup(G_1, \ldots, G_m) = (V, E_1 \cup \cdots \cup E_m)$.

### 6.1 Border correlations

Recall that by Theorem 3.5, the population size of a border correlation coincides with the population size of its corresponding ternary period correlation. We first describe a graphical approach based on the subgraph component polynomial to calculate their population size.

We define two types of graph collections that are associated with a border correlation.

**Definition 6.1.** *Let* $\mathbf{bc} = \mathbf{bc}[0..n)$ *be a border correlation. For each* $\ell \in [1..n)$*, associate an undirected graph whose set of vertices consists of* $[0..n-1]$ *and whose set of edges consists of the pairs* $\{i, j\}$ *such that* $|i - j| = n - \ell$*.*

- *Let* $\mathbf{G}_C(\mathbf{bc})$ *be the collection of all graphs associated with integers* $\ell$ *such that* $\mathbf{bc}[\ell - 1] = 1$*. Each graph in this collection is called a* compatibility graph*. If* $\mathbf{G}_C(\mathbf{bc}) = \{G_1, \dots, G_m\}$*, then let* $G_{\mathbf{bc}}$ *be* $\cup(G_1, \dots, G_m)$*.*
- *Let* $\mathbf{G}_I(\mathbf{bc})$ *be the collection of all graphs associated with integers* $\ell$ *such that* $\mathbf{bc}[\ell - 1] = 0$*. Each graph in this collection is called an* incompatibility graph*. If* $\mathbf{G}_I(\mathbf{bc}) = \{G_1, \dots, G_m\}$*, then let* $\overline{G}_{\mathbf{bc}}$ *be* $\cup(G_1, \dots, G_m)$ *(note that* $\overline{G}_{\mathbf{bc}}$ *is the complement graph of* $G_{\mathbf{bc}}$*).*

Note that the graph associated with $\ell$ records the pairs of positions that are $(n - \ell)$-apart. A compatibility graph associated with $\ell$ encodes the compatibilities a partial word $w$ must satisfy to have a border of length $\ell$, i.e., it must satisfy $w[0..\ell - 1] \uparrow w[n - \ell..n - 1]$, while an incompatibility graph associated with $\ell$ has a set of edges such that at least one edge must correspond to an incompatibility in order for $w$ not to have a border of length $\ell$. Figure 2 gives an example.

Note that the graphs $G_{\mathbf{bc}}$ and $\overline{G}_{\mathbf{bc}}$ may have the forms we discussed in Section 5. For example, $G_{000001} = E_6$, $G_{111111} = K_6$, $G_{100101} = P_6$, $G_{100011} = C_6$, $G_{010011} = W_{1,2,1,1}$, $G_{110011} = W_{3,2}$, $\overline{G}_{111101} = P_6$, $\overline{G}_{011101} = C_6$, and so on, but $G_{110101}$ is not from the previous forms. Letting $\overline{0} = 1$ and $\overline{1} = 0$, we have the following proposition.

**Proposition 6.2.** *Let* $\mathbf{bc} = \mathbf{bc}[0..n)$ *be a border correlation. Let* $\mathbf{bc}' = \mathbf{bc}'[0..n)$ *be the border correlation such that* $\mathbf{bc}'[i] = \overline{\mathbf{bc}[i]}$ *for* $0 \leq i < n - 1$*. Then* $\overline{G}_{\mathbf{bc}} = G_{\mathbf{bc}'}$*.*

The graph $G_{000100010001000100010001}$ is the disjoint union of four $K_6$'s; note that $G_{111111} = K_6$. This observation generalizes to the following proposition.

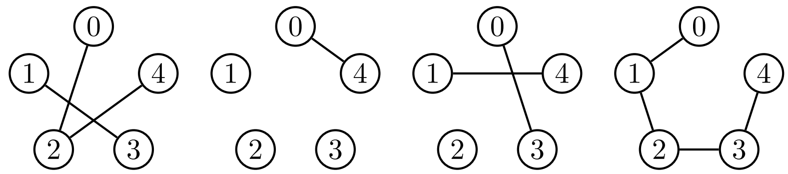


Figure 2: The first graph is the compatibility graph of border correlation 00101 that corresponds to a border of length 3. The three other graphs are the incompatibility graphs of border correlation 00101 that correspond to a border of length 1, 2, and 4, respectively.

**Proposition 6.3.** *Let* $\mathbf{bc} = [0..n)$ *be a border correlation and let* $p$ *be a positive integer. Let* $\mathbf{bc}' = [0..pn)$ *be the border correlation such that* $\mathbf{bc}'[j] = \mathbf{bc}[i]$ *for* $j = ip+p-1, 0 \leq i < n$ *and* $\mathbf{bc}'[j] = 0$ *otherwise. Then* $G_{\mathbf{bc}'}$ *is the disjoint union of* $p$ *graphs isomorphic to* $G_{\mathbf{bc}}$ *and* $P(G_{\mathbf{bc}'}; y) = (P(G_{\mathbf{bc}}; y))^p$.

*Proof.* Let $G_{\mathbf{bc}'} = (V, E)$. We can partition $V$ into $p$ equivalence classes as follows: vertex $j$ belongs to the class of vertex $i$ if $i - j$ is a multiple of $p$. So $G_{\mathbf{bc}'}$ can be written as the disjoint union of $p$ graphs $G_0, \ldots, G_{p-1}$, where the set of vertices of $G_i$ is the class of vertex $i$. Each $G_i$ contains $n$ vertices, and any two vertices $j_1 = i + k_1 p$ and $j_2 = i + k_2 p$ are connected by an edge in $G_i$ if and only if $k_1$ and $k_2$ are connected by an edge in $G_{\mathbf{bc}}$.

To see this, suppose that $j_1 = i + k_1 p$ and $j_2 = i + k_2 p$ are connected by an edge in $G_i$, so they are connected by an edge in $G_{\mathbf{bc}'}$. First note that $k_1, k_2 \in [0..n)$ since $j_1, j_2 \in [0..pn)$. By the definition of $G_{\mathbf{bc}'}$, let $\ell$ be such that $\mathbf{bc}'[\ell - 1] = 1$ and $|j_1 - j_2| = pn - \ell$. Then by the definitions of $\mathbf{bc}$ and $\mathbf{bc}'$, we have $\ell = p\ell'$ for some $\ell'$ such that $\mathbf{bc}[\ell' - 1] = 1$. Thus, $|(i + k_1 p) - (i + k_2 p)| = pn - p\ell' = p(n - \ell')$ and so $|k_1 - k_2| = n - \ell'$. By the definition of $G_{\mathbf{bc}}$, this implies that $k_1$ and $k_2$ are connected by an edge in $G_{\mathbf{bc}}$. The converse is similar.

Thus $G_i$ is isomorphic to $G_{\mathbf{bc}}$. The formula can now be derived. □

We now have the necessary definitions to count the number of partial words sharing a given border correlation.

**Theorem 6.4.** *Let* $\mathbf{bc}$ *be a border correlation. Let* $\mathbf{G}_I(\mathbf{bc}) = \{G_1, G_2, \ldots, G_m\}$ *and let* $G'_i = \cup(G_{\mathbf{bc}}, G_i)$ *for* $1 \leq i \leq m$. *Then,*

$$\mathrm{PS}_k(\mathbf{bc}) = \mathrm{VC}_k(G_{\mathbf{bc}}) + \sum_{j=1}^{m} (-1)^j \sum_{\{i_1,\ldots,i_j\}} \mathrm{VC}_k(\cup(G'_{i_1}, \ldots, G'_{i_j})),$$

*where* $\{i_1, \ldots, i_j\}$ *is a subset of* $j$ *distinct elements of* $\{1, \ldots, m\}$.

*Proof.* We start with $\mathrm{VC}_k(G_{\mathbf{bc}})$, which counts the number of partial words satisfying the required compatibilities. Every $G'_i$ is a graph that is associated with partial words satisfying the required compatibilities, i.e., the ones embodied by $G_{\mathbf{bc}}$, but also satisfying compatibilities that prevent them from having border correlation $\mathbf{bc}$, i.e., the ones embodied by $G_i$. Thus, we subtract the number of partial words that satisfy compatibilities they should not satisfy, i.e., we subtract

$$\sum_{\{i_1\}} \mathrm{VC}_k(G'_{i_1}) = \sum_{i=1}^{m} \mathrm{VC}_k(G'_i).$$

The result follows by using the inclusion-exclusion principle because for each subset $\{i_1, \ldots, i_j\}$ of $j$ distinct elements of $\{1, \ldots, m\}$, we have

$$\mathrm{VC}_k(G'_{i_1}) + \cdots + \mathrm{VC}_k(G'_{i_j}) = \mathrm{VC}_k(\cup(G'_{i_1}, \ldots, G'_{i_j})).$$

□

The following example illustrates Theorem 6.4.

**Example 6.5.** *Referring to Figure 2, consider the border correlation* 00101. *By Example 5.11,* $\mathrm{VC}_k(G_{00101}) = 1 + 9k + 19k^2 + 3k^3$. *We obtain*

$$\begin{array}{rclcl}
\mathrm{VC}_k(G'_1) & = & \mathrm{VC}_k(P_2 \sqcup C_3) & = & 1 + 10k + 21k^2 \\
\mathrm{VC}_k(G'_2) & = & \mathrm{VC}_k(\overline{C_5}) & = & 1 + 21k + 10k^2 \\
\mathrm{VC}_k(G'_3) & = & \mathrm{VC}_k(\overline{P_4}) & = & 1 + 26k + 5k^2 \\
\mathrm{VC}_k(\cup(G'_1, G'_2)) & = & \mathrm{VC}_k(\overline{P_5}) & = & 1 + 24k + 7k^2 \\
\mathrm{VC}_k(\cup(G'_1, G'_3)) & = & \mathrm{VC}_k(\overline{P_2 \sqcup P_2}) & = & 1 + 29k + 2k^2 \\
\mathrm{VC}_k(\cup(G'_2, G'_3)) & = & \mathrm{VC}_k(\overline{P_2}) & = & 1 + 30k + k^2 \\
\mathrm{VC}_k(\cup(G'_1, G'_2, G'_3)) & = & \mathrm{VC}_k(K_5) & = & 1 + 31k.
\end{array}$$

*Thus,* $\mathrm{PS}_k(00101) = 4k - 7k^2 + 3k^3$.

We next describe a lattice approach to calculate the population size of a given border correlation and its corresponding ternary period correlation. We denote the set of all partial word border correlations of length $n$ by $\mathbf{BC}_n$. For $\mathbf{bc} \in \mathbf{BC}_n$, define $\mathcal{B}(\mathbf{bc}) = \{i \mid 0 \leq i < n \text{ and } \mathbf{bc}[i] = 1\}$, and for $\mathbf{bc}, \mathbf{bc}' \in \mathbf{BC}_n$, define $\mathbf{bc} \leq \mathbf{bc}'$ if $\mathcal{B}(\mathbf{bc}) \subseteq \mathcal{B}(\mathbf{bc}')$. We use the symbolism $\mathbf{bc} < \mathbf{bc}'$ to denote $\mathbf{bc} \leq \mathbf{bc}'$ and $\mathbf{bc} \neq \mathbf{bc}'$. Referring to Theorem 3.4, the pair $(\mathbf{BC}_n, \leq)$ is a distributive lattice.

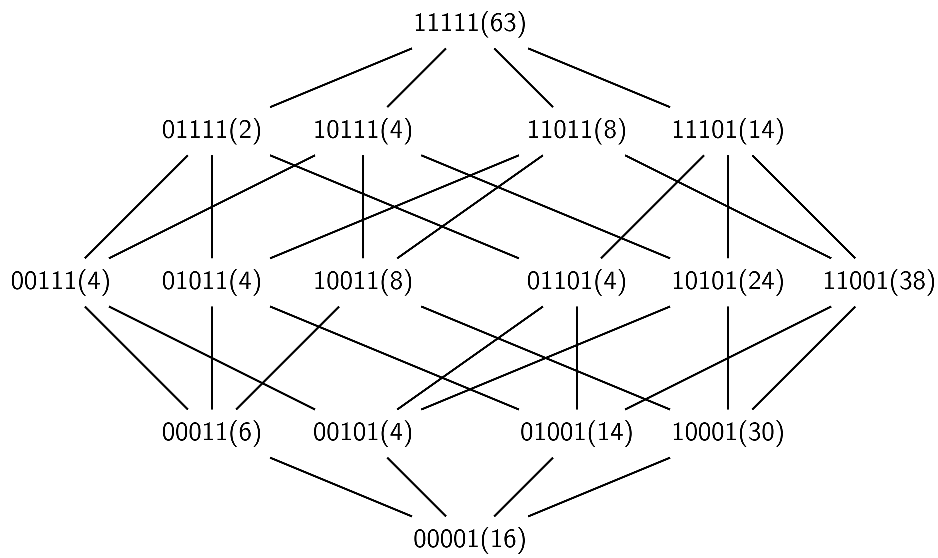


Figure 3: The distributive lattice $(\mathbf{BC}_5, \leq)$ along with the population size over the binary alphabet of each of its border correlations (population size written in parentheses).

Let $\mathbf{bc}$ be a border correlation. A partial word $w$ satisfying $\mathbf{bc}_w = \mathbf{bc}$ corresponds to a valid coloring of the vertices of $G_{\mathbf{bc}}$. Each valid coloring of $G_{\mathbf{bc}}$ corresponds to a partial word having a border correlation $\mathbf{bc}'$ such that $\mathbf{bc} \leq \mathbf{bc}'$. The number of valid colorings of $G_{\mathbf{bc}}$ is then the sum of the population sizes of all border correlations $\mathbf{bc}'$ such that $\mathbf{bc} \leq \mathbf{bc}'$. To find the population size of $\mathbf{bc}$, we subtract the population sizes of all border correlations $\mathbf{bc}'$ such that $\mathbf{bc} < \mathbf{bc}'$ from the number of valid colorings of $G_{\mathbf{bc}}$.

**Theorem 6.6.** *If* $\mathbf{bc} \in \mathbf{BC}_n$*, then*

$$\begin{aligned} \mathrm{PS}_k(\mathbf{bc}) &= \mathrm{VC}_k(G_{\mathbf{bc}}) - \sum_{\mathbf{bc}' \in \mathbf{BC}_n, \mathbf{bc} < \mathbf{bc}'} \mathrm{PS}_k(\mathbf{bc}'), \\ &= \mathrm{VC}_k(\overline{G}_{\mathbf{bc}}) - \sum_{\mathbf{bc}' \in \mathbf{BC}_n, \mathbf{bc} < \mathbf{bc}', \mathbf{bc}' \neq 1^n} \mathrm{PS}_k(\mathbf{bc}'). \end{aligned}$$

Figure 3 illustrates the lattice $(\mathbf{BC}_5, \leq)$ along with the population size of each of its border correlations, calculated over a binary alphabet using our results. The following example gives details for the computation of $\mathrm{PS}_2(01011)$.

**Example 6.7.** *The graph* $\overline{G}_{01011}$ *contains two connected components,* $P_2$ *and* $C_3$*. Thus, the number of valid colorings of* $G_{01011}$ *with at most two non-hole colors is* $1 + (2^5 - 1)2 + \mathrm{UC}_2(P_2) + \mathrm{UC}_2(C_3) = 63 + 2 + 12$*. To find the population size of the border correlation 01011, we must subtract the population sizes of the border correlations* $\mathbf{bc}'$ *such that* $01011 < \mathbf{bc}'$*. Thus, we must subtract* $\mathrm{PS}_2(01111) + \mathrm{PS}_2(11011) + \mathrm{PS}_2(11111) = 2 + 8 + 63$*, so* $\mathrm{PS}_2(01011) = 4$*.*

We can also calculate the population size when restricting the number of holes. Given a border correlation $\mathbf{bc}$ of length $n$, denote by $\mathrm{PS}_{h,k}(\mathbf{bc})$ the number of partial words of length $n$ with $h$ holes over $\Sigma_k$ sharing $\mathbf{bc}$ as their border correlation. Referring to Theorems 5.13 and 6.4, we obtain the following.

**Theorem 6.8.** *Let* $\mathbf{bc} = \mathbf{bc}[0..n)$ *be a border correlation, let* $h$ *be an integer in* $[0..n]$*, let* $\mathbf{G}_I(\mathbf{bc}) = \{G_1, G_2, \ldots, G_m\}$*, and let* $G'_i = \cup(G_{\mathbf{bc}}, G_i)$ *for* $1 \leq i \leq m$*. Then,*

$$\mathrm{PS}_{h,k}(\mathbf{bc}) = \mathrm{VC}_{h,k}(G_{\mathbf{bc}}) + \sum_{j=1}^{m}(-1)^j \sum_{\{i_1,\ldots,i_j\}} \mathrm{VC}_{h,k}(\cup(G'_{i_1}, \ldots, G'_{i_j})),$$

*where* $\{i_1, \ldots, i_j\}$ *is a subset of* $j$ *distinct elements of* $\{1, \ldots, m\}$*.*

## 6.2 Refined border correlations

We obtain a formula for the population size of a refined border correlation. We define two types of graph collections that are associated with it.

**Definition 6.9.** *Let* $\mathbf{rbc} = \mathbf{rbc}[0..n)$ *be a refined border correlation. For each* $i \in [0..n)$*:*

- *If* $\mathbf{rbc}[i] = \ell \neq 0$*, then associate a* compatibility graph*, i.e., an undirected graph whose set of vertices consists of* $[0..n-1]$ *and whose set of edges consists of the pairs*

$$\{i, (i + n - \ell) \bmod n\}, \ldots, \{(i + \ell - 1) \bmod n, (i + n - 1) \bmod n\}.$$

- *If* $\mathbf{rbc}[i] = \ell \neq 1$*, then associate for each* $j \in [1..(n + \ell - 1) \bmod n]$ *an* incompatibility graph*, i.e., an undirected graph whose set of vertices consists of* $[0..n-1]$ *and whose set of edges consists of the pairs*

$$\{i, (i + n - j) \bmod n\}, \ldots, \{(i + j - 1) \bmod n, (i + n - 1) \bmod n\}.$$

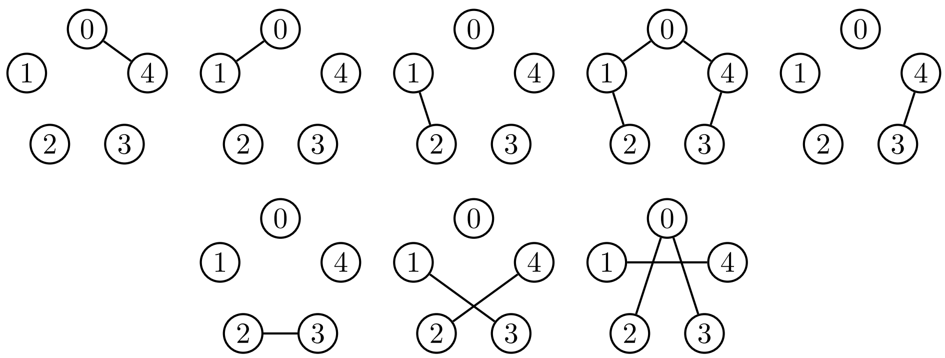

Figure 4: The top five graphs are the compatibility graphs of refined border correlation 11141 that correspond to a minimal border of length 1 for the 0-shift, 1 for the 1-shift, 1 for the 2-shift, 4 for the 3-shift, and 1 for the 4-shift, respectively. The three bottom graphs are the incompatibility graphs of refined border correlation 11141 that correspond to the 3-shift for not having a border of length 1, 2, or 3, respectively.

- *Let* $\mathbf{H}_C(\mathbf{rbc})$ *be the collection of all compatibility graphs. If* $\mathbf{H}_C(\mathbf{rbc}) = \{H_1, \ldots, H_m\}$*, then let* $H_{\mathbf{rbc}}$ *be* $\cup(H_1, \ldots, H_m)$*.*
- *Let* $\mathbf{H}_I(\mathbf{rbc})$ *be the collection of all incompatibility graphs. If* $\mathbf{H}_I(\mathbf{rbc}) = \{H_1, \ldots, H_m\}$*, then let* $\overline{H}_{\mathbf{rbc}}$ *be* $\cup(H_1, \ldots, H_m)$ *(note that* $\overline{H}_{\mathbf{rbc}}$ *is the complement graph of* $H_{\mathbf{rbc}}$*).*

A compatibility graph associated with $i$ encodes the compatibilities the $i$-shift of a partial word $w$ must satisfy to have a border of length $\ell$, while an incompatibility graph associated with $i$ has a set of edges such that at least one edge must correspond to an incompatibility in order for the $i$-shift of $w$ not to have a border of length $j$ for some $1 \leq j < \ell$. Since there is an incompatibility graph for each $1 \leq j < \ell$, this ensures that the border length $\ell$ is minimal. Figure 4 gives an example.

We now have the necessary definitions to count the number of partial words sharing a given refined border correlation.

**Theorem 6.10.** *Let* $\mathbf{rbc}$ *be a refined border correlation. Let* $\mathbf{H}_I(\mathbf{rbc}) = \{H_1, H_2, \ldots, H_m\}$ *and let* $H'_i = \cup(H_{\mathbf{rbc}}, H_i)$ *for* $1 \leq i \leq m$*. Then,*

$$\mathrm{PS}_k(\mathbf{rbc}) = \mathrm{VC}_k(H_{\mathbf{rbc}}) + \sum_{j=1}^{m}(-1)^j \sum_{\{i_1,\ldots,i_j\}} \mathrm{VC}_k(\cup(H'_{i_1}, \ldots, H'_{i_j})),$$

*where* $\{i_1, \ldots, i_j\}$ *is a subset of* $j$ *distinct elements of* $\{1, \ldots, m\}$*.*

The following example illustrates Theorem 6.10.

**Example 6.11.** *Referring to Figure 4, consider the refined border correlation* 11141*. By Proposition 5.6,* $\mathrm{VC}_k(H_{11141}) = \mathrm{VC}_k(P_5) = 1+15k+15k^2+k^3$*. We obtain*

$$\begin{array}{rcccl} \mathrm{VC}_k(H'_1) & = & \mathrm{VC}_k(C_5) & = & 1+21k+10k^2 \\ \mathrm{VC}_k(H'_2) & = & \mathrm{VC}_k(\overline{P_2 \sqcup C_3}) & = & 1+26k+4k^2+k^3 \\ \mathrm{VC}_k(H'_3) & = & \mathrm{VC}_k(\overline{P_4}) & = & 1+26k+5k^2 \\ \mathrm{VC}_k(\cup(H'_1, H'_2)) & = & \mathrm{VC}_k(\overline{P_2 \sqcup P_3}) & = & 1+27k+4k^2 \\ \mathrm{VC}_k(\cup(H'_1, H'_3)) & = & \mathrm{VC}_k(\overline{P_2 \sqcup P_2}) & = & 1+29k+2k^2 \\ \mathrm{VC}_k(\cup(H'_2, H'_3)) & = & \mathrm{VC}_k(\overline{P_2}) & = & 1+30k+k^2 \\ \mathrm{VC}_k(\cup(H'_1, H'_2, H'_3)) & = & \mathrm{VC}_k(K_5) & = & 1+31k. \end{array}$$

*We conclude that* $\mathrm{PS}_k(11141) = -3k+3k^2$*.*

**Corollary 6.12.** *Let* $\mathbf{rbc} = \mathbf{rbc}[0..n)$ *be a refined border correlation and let* $i \in [0..n)$*. Let* $\mathbf{rbc}'$ *be the refined border correlation that is the* $i$*-shift of* $\mathbf{rbc}$*. Then,* $\mathrm{PS}_k(\mathbf{rbc}) = \mathrm{PS}_k(\mathbf{rbc}')$*. Moreover, the partial words sharing* $\mathbf{rbc}'$ *as their refined border correlation are the* $i$*-shifts of the partial words sharing* $\mathbf{rbc}$ *as their refined border correlation.*

We can also calculate the population size when restricting the number of holes.

**Theorem 6.13.** *Let* $\mathbf{rbc} = \mathbf{rbc}[0..n)$ *be a refined border correlation, let* $h$ *be an integer in* $[0..n]$*, let* $\mathbf{H}_I(\mathbf{rbc}) = \{H_1, H_2, \ldots, H_m\}$*, and let* $H'_i = \cup(H_{\mathbf{rbc}}, H_i)$ *for* $1 \leq i \leq m$*. Then,*

$$\mathrm{PS}_{h,k}(\mathbf{rbc}) = \mathrm{VC}_{h,k}(H_{\mathbf{rbc}}) + \sum_{j=1}^{m} (-1)^j \sum_{\{i_1,\ldots,i_j\}} \mathrm{VC}_{h,k}(\cup(H'_{i_1}, \ldots, H'_{i_j})),$$

*where* $\{i_1, \ldots, i_j\}$ *is a subset of* $j$ *distinct elements of* $\{1, \ldots, m\}$*.*

# 7 Conclusion and Open Problems

Our method for computing population sizes of border correlations also applies to computing population sizes of ternary period correlations that record the strong and weak period sets, as defined in Definition 3.1, through the one-to-one correspondence of Theorem 3.5. A World Wide Web server interface has been established at www.uncg.edu/cmp/research/bordercorrelations2

for automated use of a program that implements the algorithm in the proof of Theorem 3.5, which lists the $2^{n-1}$ valid border correlations of length $n$ (they are the binary vectors that end in 1) and their corresponding ternary period correlations. There are several open problems involving binary period correlations, the binary vectors that record the strong period sets of partial words that do not have strictly weak periods: listing the valid binary period correlations of a given length, finding their number, and computing their population size.

The concept of refined border correlation, as defined in Definition 4.1, is the same as that defined in [20] (denoted by $\beta'$ in that paper), but the concept of border correlation, as defined in Definition 3.3, is different from the concept of border correlation defined in [20] (denoted by $\beta$ in that paper and defined by $\beta[i] = 0$ if $\sigma^i(w)$ is unbordered and by $\beta[i] = 1$ otherwise). Our method of computing population sizes can be applied to that different concept of border correlation as well. We expect that there are more connections between refined border correlations and ternary period correlations than those we have discussed in Section 4. It would be nice to find a characterization of all vectors that can be refined border correlations for partial words.

We have established connections between border correlations and undirected graphs. Recent papers have established connections between border/prefix arrays and undirected graphs (see [4, 15]). We suggest to find faster methods of computing the population size of a border correlation or to find efficient ways to determine the connected component vector. We also suggest to find a characterization of the graphs $G_{\mathbf{bc}}$ associated with border correlations **bc** in terms of graph structures such as the ones discussed in this paper: complete graphs, path graphs, cycle graphs, wheel graphs, broken wheel graphs, etc. Our method has the advantage that it also applies to indeterminate strings; in that context, we can perhaps encounter structures not discussed in this paper such as ladder graphs. We refer the reader to the well-known Kuratowski's theorem in graph theory [14, 28], which is a mathematical characterization via forbidden minors, which could give some insights into what a characterization should be like.

In [1], bordered partial words of length $n$ with $h$ holes over $\Sigma_k$ have been counted by so-called critical positions. A formula was given for the parameter $h = 1$, but no general formula was given. Our method here gives formulas for any $h \geq 1$, i.e., we can calculate $\binom{n}{h}k^{n-h} - \mathrm{PS}_{h,k}(0^{n-1}1)$. So we can link the population size of the border correlation $0^{n-1}1$, or the ternary period correlation $10^{n-1}$ through Theorem 3.5, to the counting of bordered partial words with the above mentioned parameters. If we do not restrict the number of holes, the number of unbordered partial words of length $n$ over $\Sigma_k$ is $\mathrm{PS}_k(0^{n-1}1)$, which can be calculated with our formulas in Theorem 6.6 in particular.

## Acknowledgements

We thank Michelle Bodnar and Joe Hidakatsu for their valuable suggestions. They suggested to count the number of unbordered partial words of length $n$ with $h$ holes by utilizing incompatibility graphs and only considering the subgraphs that have $n-h$ vertices. We thank Emily Clader and Olivia Simpson for the data they collected on refined border correlations. We also thank the referees for their comments and suggestions which have helped improve the exposition of the paper.

A research assignment from the University of North Carolina at Greensboro for the first author is gratefully acknowledged. Some of this assignment was spent at the IRIF: Institut de Recherche en Informatique Fondamentale of Université Paris-Diderot–Paris 7, France.

## References

[1] E. Allen, F. Blanchet-Sadri, C. Byrum, M. Cucuringu, and R. Mercaş. Counting bordered partial words by critical positions. *The Electronic Journal of Combinatorics*, 18:P138, 2011.

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

[3] F. Blanchet-Sadri. *Algorithmic Combinatorics on Partial Words*. Chapman & Hall/CRC Press, Boca Raton, FL, 2008.

[4] F. Blanchet-Sadri, M. Bodnar, and B. De Winkle. New bounds and extended relations between prefix arrays, border arrays, undirected graphs, and indeterminate strings. In E. Mayr and N. Portier, editors, *STACS 2014, 31st International Symposium on Theoretical Aspects of Computer Science, Lyon, France*, volume 25 of *LIPIcs Schloss Dagstuhl-Leibniz-Zentrum für Informatik*, pages 162–173, Germany, 2014. Schloss Dagstuhl.

[5] F. Blanchet-Sadri, M. Bodnar, J. Nikkel, J. D. Quigley, and X. Zhang. Squares and primitivity in partial words. *Discrete Applied Mathematics*, 185:26–37, 2015.

[6] F. Blanchet-Sadri and A. Chriscoe. Local periods and binary partial words: an algorithm. *Theoretical Computer Science*, 314:189–216, 2004.

[7] F. Blanchet-Sadri, E. Clader, and O. Simpson. Border correlations of partial words. *Theory of Computing Systems*, 47:179–195, 2010.

[8] F. Blanchet-Sadri, M. Cordier, and R. Kirsch. Border correlations, lattices, and the subgraph component polynomial. In J. Kratochvíl et al., editors, *IWOCA 2014, 25th International Workshop on Combinatorial Algorithms, Duluth, Minnesota*, volume 8986 of *Lecture Notes in Computer Science*, pages 62–73, Switzerland, 2015. Springer International Publishing.

[9] F. Blanchet-Sadri, C. D. Davis, J. Dodge, R. Mercaş, and M. Moorefield. Unbordered partial words. *Discrete Applied Mathematics*, 157:890 – 900, 2009.

[10] F. Blanchet-Sadri, J. Fowler, J. D. Gafni, and K. H. Wilson. Combinatorics on partial word correlations. *Journal of Combinatorial Theory, Series A*, 117:607–624, 2010.

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

[12] F. Blanchet-Sadri, T. Mandel, and G. Sisodia. Periods in partial words: An algorithm. *Journal of Discrete Algorithms*, 16:113–128, 2012.

[13] F. Blanchet-Sadri, R. Mercaş, A. Rashin, and E. Willett. Periodicity algorithms and a conjecture on overlaps in partial words. *Theoretical Computer Science*, 443:35–45, 2012.

[14] J. A. Bondy and U. S. R. Murty. *Graph Theory*. Graduate Texts in Mathematics 244, Springer, 2008.

[15] M. Christodoulakis, P. J. Ryan, W. F. Smyth, and S. Wang. Indeterminate strings, prefix arrays & undirected graphs. *Theoretical Computer Science*, 600:34–48, 2015.

[16] M. Crochemore, C. Hancart, and T. Lecroq. *Algorithms on Strings.* Cambridge University Press, 2007.

[17] L. J. Guibas and A. M. Odlyzko. Periods in strings. *Journal of Combinatorial Theory, Series A*, 30:19–42, 1981.

[18] D. Gusfield. *Algorithms on Strings, Trees, and Sequences.* Cambridge University Press, Cambridge, 1997.

[19] V. Halava, T. Harju, and T. Kärki. Interaction properties of relational periods. *Discrete Mathematics and Theoretical Computer Science*, 10:87–112, 2008.

[20] T. Harju and D. Nowotka. Border correlation of binary words. *Journal of Combinatorial Theory, Series A*, 108:331–341, 2004.

[21] M. Lothaire. *Applied Combinatorics on Words.* Cambridge University Press, Cambridge, 2005.

[22] J. A. Makowsky. From a zoo to a zoology: Towards a general study of graph polynomials. *Theory of Computing Systems*, 43:542–562, 2008.

[23] F. Manea, R. Mercaş, and C. Tiseanu. An algorithmic toolbox for periodic partial words. *Discrete Applied Mathematics*, 179:174–192, 2014.

[24] E. Rivals and S. Rahmann. Combinatorics of periods in strings. *Journal of Combinatorial Theory, Series A*, 104:95–113, 2003.

[25] A. M. Shur and Y. V. Gamzova. Partial words and the interaction property of periods. *Izvestiya Rossiiskoi Akademii Nauk. Seriya Matematicheskaya*, 68:191–214, 2004.

[26] W. F. Smyth. *Computing Patterns in Strings.* Pearson Addison-Wesley, 2003.

[27] W. F. Smyth and S. Wang. A new approach to the periodicity lemma on strings with holes. *Theoretical Computer Science*, 410:4295–4302, 2009.

[28] C. Thomassen. Kuratowski's theorem. *Journal of Graph Theory*, 5:225-241, 1981.

[29] P. Tittmann, I. Averbouch, and J. A. Makowsky. The enumeration of vertex induced subgraphs with respect to number of components. *European Journal of Combinatorics*, 32:954–974, 2011.