UNIVERSIDAD DE BUENOS AIRES
FACULTAD DE CIENCIAS EXACTAS Y NATURALES
DEPARTAMENTO DE COMPUTACIÓN

Secuencias de rachas

Tesis de Licenciatura en Ciencias de la Computación

Tomás Donzis
LU 443/20
donzis.tomy@gmail.com

Directora: Verónica Becher

Buenos Aires, Agosto 2024

Resumen

Fijemos un alfabeto de dos símbolos. Una racha es una secuencia de mismo símbolo seguida por el símbolo contrario. Dado que las rachas son maximales a derecha, cada posición de una secuencia es el inicio de una racha. Una secuencia binaria de longitud N tiene la propiedad de rachas si $1/2^i$ de sus rachas son de longitud i, para $i = 1, ..., \lfloor \log_2 N \rfloor - 1$. En este trabajo definimos una propiedad de rachas más fina que cumple propiedades de balance que surgen al hacer una partición de las posiciones de una secuencia en clases de equivalencia. Dados dos números enteros positivos, k y n, donde k limita las longitudes de las rachas y n es la cantidad de clases de equivalencia, definimos la propiedad de (k, n)-rachas para secuencias binarias de longitud $n2^k$. Damos cotas de la cantidad de secuencias de (k, n)-rachas y caracterizamos a las secuencias de (k, n)-rachas mediante ciclos hamiltonianos en grafos de rachas maximales.

Índice general

1. SECUENCIAS DE RACHAS

Las secuencias pseudo aleatorias se usan en distintas áreas de la computación, como sistemas de comunicación y criptográficos, y hay numerosos trabajos de investigación acerca de cuáles son condiciones necesarias para que una secuencia sea considerada pseudoaleatoria. Estas investigaciones sobre secuencias toman como punto de partida una familia de longitudes conveniente, y luego consideran secuencias finitas de esas longitudes. A cada una de esas secuencias finitas se las puede ver simplemente como una secuencia finita, o como una secuencia infinita periódica que proviene de tomar infinitas copias sucesivas de una secuencia finita, o, lo que es lo mismo, como una secuencias finita circular, para la cual se puede fija una posición de inicio. A las secuencias circulares también se las llama collares.

S. W. Golomb postuló tres propiedades independientes de aleatoriedad: balance, rachas y autocorrelación ideal [4, 5].

El balance de una secuencia requiere que haya esencialmente igual cantidad de apariciones de cada símbolo del alfabeto. Notemos que si la secuencia es impar es imposible tener exactamente la misma cantidad de 0s que de 1s. Si en vez de representar a la secuencia con 0s y 1s lo hacemos con -1s y 1s, entonces la propiedad de balance de una secuencia $a_0 \dots a_{N-1}$ de N símbolos se escribe así

$$-1 \leq \sum_{i=0}^{N-1} a_i \leq 1$$

Una racha es la una seguidilla del mismo símbolo, seguida del símbolo contrario. Por lo tanto cada racha incluye una racha de un símbolo menos, también una de dos símbolos menos, ..., hasta llegar a una rachas de un solo símbolo. La propiedad de rachas requiere que $\frac{1}{2^i}$ de las rachas tengan longitud i, para $0, 1, 2, ..., (\log N) - 1$, donde N es la longitud de la secuencia.

La correlación ideal requiere la similitud de la secuencia con cada una de sus rotaciones. Más precisamente, dada una constante K, un período p, una secuencia $a_1 ... a_p$, y una cantidad de posiciones a rotar τ entre 1 y p,

$$similitud(rotacion(a_0 ... a_{p-1}, \tau)) = \frac{1}{p} \sum_{n=0}^{p-1} a_n a_{n+\tau} = \begin{cases} 1, \text{ si } \tau = 0 \\ K/p, \text{ si } 0 < \tau < p \end{cases}$$

Entonces la secuencia tiene autocorrelación ideal si la función de similitud de cada una de sus rotaciones es bivaluada.

En esta tesis nos concentramos en la propiedad de rachas y de balance, sin prestar atención a la autocorrelación ideal.

Fijemos el alfabeto $\{0, 1\}$.

Definición 1 (racha)**.** En una secuencia, una *racha* es una subsecuencia compuesta por una seguidilla del mismo símbolo, seguida por el símbolo contrario. Una racha de longitud k es una seguidilla del mismo símbolo k veces, seguida por el símbolo contrario.

Si una secuencia tiene longitud ℓ, entonces tiene ℓ rachas, ya que cada una de sus posiciones es el inicio de una racha.

Nos proponemos reforzar la propiedad de rachas de Golomb con fuertes propiedades de balance. Nuestro punto de partida son dos parámetros k y n, que son números enteros positivos. El

parámetro k limita la longitud de las rachas que deben aparecer en la secuencia. El parámetro n determina una cantidad de clases de equivalencia entre las posiciones de la secuencia. Fijamos la longitud de las secuencias en $n2^k$, numeramos las posiciones desde la 0 a la posición $n2^k - 1$. Dividimos las posiciones de la secuencia en n partes, según su congruencia módulo n: las de congruencia 0, las de congruencia 1, ..., las de congruencia $n-1$. Exigiremos que haya no solamente una cantidad determinada de rachas de cada longitud menor que k, sino que además pediremos que haya balance en cada una de las n partes.

Definición 2 (Ocurrencia de una racha)**.** Dada una secuencia $a_0 \dots a_{N-1}$ y dado un valor n, decimos que una racha de i 1s ocurre en una posición congruente a m módulo n si hay una posición p, tal que

- $0 \leq p < N - i$
- $p \mod n = m$
- $a_p \dots a_{p+i-1} = 1^i$.

Lo mismo vale en caso de racha de $0s$. En resumen, tomamos la posición de inicio de la ocurrencia de la racha para decir que la racha ocurre en esa posición. Tomamos luego la congruencia de esa posición modulo n.

Definición 3 (Secuencias de (k, n)-rachas)**.** Sean k y n dos enteros positivos. Una secuencia de (k, n)-rachas es una secuencia de longitud $n2^k$ tal que para cada $i = 1, 2, ..., k - 1$, y para cada $m = 0, .., n - 1$, tiene 2^{k-i} rachas de longitud i en posiciones congruentes a m modulo n, la mitad de 0s y la otra mitad de 1s. Tiene, además, para cada $m = 0, .., n - 1$, 2 rachas de longitud mayor o igual que k en posiciones congruentes a m modulo n, una de 0s y una de 1s.

Por ejemplo, esta es una secuencia de (k, n)-rachas para $(k, n)= (3, 2)$.

0000100101101111

En cambio,

0000100110110111

no es una secuencia de $(3, 2)$-rachas, porque tiene 2 rachas de 1s de longitud mayor o igual que $k - 1 = 2$ en posiciones impares, cuando la definición exige que haya solo una.

Problema 1. *Para cada par de enteros positivos* (k, n)*, ¿Cuántas secuencias de* (k, n)*-rachas hay?*

Problema 2. *¿Cómo construimos una secuencia de* (k, n)*-rachas?*

En esta tesis daremos una cota para el primer problema y un método de construcción para el segundo problema. Esta cota nos permite concluir que la cantidad de secuencias de (k, n)-rachas es sustancialmente mayor que la cantidad de otras secuencias con propiedades de balance, como las secuencias de De Bruijn de orden k, los collares (k, n)-perfectos y las secuencias de rachas simples de orden k.

Nuestra definición de secuencia de (k, n)-rachas toma como punto de partida el trabajo de Gangsam Kim y Hong-Yeop Song [6] acerca de la propiedad de rachas para secuencias de longitud $2^k - 1$, que desarrollamos en la siguiente sección.

2. RACHAS MAXIMALES Y SECUENCIAS BALANCEADAS

En este capítulo damos definiciones y resultados que usaremos en la solución a los problemas planteados.

2.1. Rachas maximales

Definición 4 (racha maximal)**.** Una racha maximal de i símbolos es una secuencia del mismo símbolo i veces precedida y sucedida por el símbolo contrario.

Tomando el ejemplo anterior,

$$0000100101101111$$

es una secuencia de $(3, 2)$-rachas, por lo tanto tiene longitud $n2^k = 16$.

Sus rachas de 0s son:

Mod. \ Long.	1	2	3	4
0	2	1	0	1
1	2	1	1	0

Mientras que sus rachas maximales de 0s son:

Mod. \ Long.	1	2	3	4
0	1	0	0	1
1	1	1	0	0

Análogamente se pueden definir las rachas, maximales y no maximales, de 1s.

Otro ejemplo. Esta es una secuencia de $(4, 2)$-rachas, por lo que tiene longitud $n2^k = 32$,

$$00000111110110111011010010001001$$

Tiene estas rachas de 0s:

Mod. \ Long.	1	2	3	4	5
0	4	2	1	0	1
1	4	2	1	1	0

Tiene estas rachas maximales de 0s:

Mod. \ Long.	1	2	3	4	5
0	2	1	0	0	1
1	2	1	1	0	0

Un par de tablas similares definen las rachas de 1s.

Es posible definir a las secuencias de (k, n) rachas mediante la cantidad de sus rachas maximales. Exactamente la mitad de las rachas son maximales.

Proposición 1. *Cada secuencia de (k, n)-rachas tiene para cada $m = 0, .., n-1$ en posiciones congruentes a m módulo n,*

- *para cada $i = 1, .., k-2$, 2^{k-i-1} rachas maximales de longitud i ,*
- *2 rachas maximales de longitud mayor o igual que $k-1$. Estas determinan 2 particiones de nk, cada una en n partes entre $k-1$ y $k+n-1$, con posible repetición.*

En cada caso, la mitad son rachas maximales de 0*s y la otra mitad de* 1*s.*

Demostración. Según la definición de la propiedad de (k, n)-rachas, hay $n2^{k-1}$ rachas de longitud 1 y $n2^{k-2}$ rachas de longitud 2. Como en cada posición de la secuencia comienza una racha distinta, entonces podemos afirmar que todas las rachas de longitud 2 contienen como subracha una racha de longitud 1. Entonces hay $n2^{k-1} - n2^{k-2} = n2^{k-2}$ rachas maximales de longitud 1, de las cuales hay 2^{k-2} rachas maximales de longitud 1 en posiciones congruentes a m, para cada $m = 0, 1, .., n$, y la mitad son rachas maximales de 0s, la otra mitad de 1s.

De modo general, para $i = 1, 2, ..., k-1$ por definición hay $n2^{k-i}$ rachas de longitud i, y todas ellas contienen una subracha de longitud $i-1$. Entonces para $i = 1, ..., k-2$ hay $n2^{k-i} - n2^{k-(i+1)} = n2^{k-i-1}$ rachas maximales de longitud i, de las cuales hay 2^{k-i-1} en posiciones congruentes a m modulo n, para cada $m = 0, .., n-1$, y la mitad son de 0, la otra mitad son de 1s. Luego la cantidad de rachas maximales de longitud entre 1 y $k-2$ es

$$\sum_{i=1}^{k-2} n2^{k-1-i} = n2^{k-1} - 2n.$$

¿Cuántas posiciones ocupan las $2n$ rachas maximales de longitud $\geq k-1$?
Dado que la secuencia tiene longitud $n2^k$ y dado que

$$n\sum_{i=1}^{k-2} i2^{k-i-1} = n(2^k - 2k),$$

las rachas maximales de longitud mayor o igual que $k-1$ ocupan $2nk$ símbolos, de los cuales nk son 0s y nk son 1s.

Supongamos que $n-1$ rachas de 0s son de longitud $k-1$, entonces como todas tienen longitud mayor o igual que $k-1$, la más larga puede tener hasta $nk-(n-1)(k-1) = k+n-1$. concluimos que las longitudes de las rachas de $k-1$ o más definen una partición de nk en n partes, cada una entre $k-1$ y $k+n-1$, con posible repetición. □

2.2. Secuencias de De Bruijn

Definición 5. Una secuencia de De Bruijn [3] de orden k binaria circular es una secuencia circular de 0s y 1s de longitud 2^k tal que cada secuencia de longitud k aparece exactamente una vez.

Por ejemplo, para $k = 3$, la secuencia

00010111

es de De Bruijn de orden 3, pero

00100111

no lo es, pues la secuencia 001 aparece dos veces. Dado un entero positivo, k, la cantidad de secuencias circulares binarias de De Bruijn está dada por la fórmula que arroja el teorema BEST [3, 7] que cuenta la cantidad de ciclos eulerianos en el grafo de De Bruijn de palabras de longitud $k-1$,

$$\frac{2^{2^{k-1}}}{2^k}.$$

En la Tabla 3.1 se pueden ver la cantidad de secuencias de De Bruijn para distintos valores de k.

Proposición 2. *Toda secuencia de De Bruijn de orden k es una secuencia de* $(k, 1)$*-rachas. La recíproca no es cierta.*

Demostración. Por definición de secuencia de De Bruijn de orden k, todas las palabras de longitud k aparecen exactamente una vez cada una. Esto implica que, cada palabra de longitud $k-1$ aparece dos veces, porque será el prefijo de longitud $k-1$ de dos palabras distintas, cada una de ellas aparece una vez. En general, cada palabra de i símbolos es el prefijo de longitud i de 2^{k-i} palabras distintas de longitud k, y cada una de ellas aparece una vez en la secuencia de De Bruijn de orden k. Para $i = 1, .., k-2$, cada racha maximal de i 0s, está precedida y sucedida por un 1, por lo tanto denota una palabra de longitud $i+2$, que aparece en la secuencia de De Bruijn en total $2^{k-(i+2)}$ veces.

¿Cuántas veces aparece 0^{k-1} como racha maximal? En una secuencia de De Bruijn la palabra 0^{k-1} aparece 2 veces, ambas precedidas de un 1. Una aparición está seguida de 0, y la otra aparición está seguida de 1. La que está seguida de 0 no es una racha maximal de longitud $k-1$. La segunda aparición está adentro de una racha maximal de longitud k. Concluimos que para $i = 1, .., k-2$ hay 2^{k-i-2} rachas maximales de i 0s, y no hay rachas maximales de longitud $k-1$, y hay una racha maximal de k 0s.

Lo mismo ocurre para las rachas maximales de 1s, de longitudes $1, .., k$.

Veamos ahora que hay secuencias de $(k, 1)$-rachas que no son de De Bruijn de orden k. Para $k = 4$, $n = 1$, tenemos la secuencia de $(k, 1)$-rachas

$$0000100101101111$$

Esta no es de De Bruijn de orden $k = 4$, pues la subsecuencia de longitud k 1011 aparece dos veces. □

2.3. Collares (k, n)-perfectos

Una secuencia binaria de longitud $n2^k$ es un collar (k, n)-perfecto si cada secuencia de longitud k aparece exactamente n veces como subsecuencia en n posiciones distintas de la secuencia módulo n. Las secuencias de De Bruijn de orden k son exactamente los collares $(k, 1)$-perfectos. Los collares perfectos fueron presentados por primera vez en [1], como variantes de las secuencias de De Bruijn. Ver también [2].

Para $k = n = 2$, la secuencia del orden lexicográfico de palabras de longitud 2

$$00011011$$

es un collar perfecto. De hecho, para todo $k = n$, su respectiva secuencia del orden lexicográfico de k símbolos es (k, k)-perfecta.

Para $k = 3, n = 2$, la secuencia

$$0000100101101111$$

es un collar $(3, 2)$-perfecto, pero

$$0001000101101111$$

no lo es, pues la subsecuencia 001 empieza 2 veces en una posición impar.

En [1] se da una fórmula para contar la cantidad de collares (k, n)-perfectos a partir de la cantidad de ciclos eulerianos en el grafo que proviene de hacer el producto tensorial del grafo de Bruijn de orden $k-1$ con un ciclo simple de longitud n,

Proposición 3 ([1])**.** *La cantidad de collares* (k, n)*-perfectos binarios está dada por la fórmula*

$$\frac{1}{n}\sum_{d_{2,n}|j|n} e(j)\phi(n/j)$$

donde $d_{2,n}$ *es la mayor potencia de 2 que divide a n,* $e(j) = 2^{j2^{k-1}b^{-k}}$*, la cantidad de ciclos eulerianos de orden* $k-1$ *y* ϕ *es la función totiente de Euler.*

En la Tabla 3.2 se pueden ver la cantidad de collares para distintos pares (k, n).

Proposición 4. *Todos los collares* (k, n)*-perfectos son secuencias de* (k, n)*-rachas. La recíproca no es cierta.*

Demostración. La implicación vale como consecuencia directa de la definición los collares (k, n)-perfectos. La definición afirma que para cada módulo $m = 0, ..., n-1$ debe cumplirse la misma propiedad que tienen las secuencias de De Bruijn, en cuanto a apariciones de secuencias de longitud k. Por lo tanto, en cada módulo ocurre lo que ya justificamos para las secuencias de De Bruijn de orden k.

Veamos que hay secuencias de (k, n)-rachas que no son collares (k, n)-perfectos. Para $k = 4, n = 2$, la secuencia

$$00001000010011111011011101101001$$

es una secuencia de (k, n)-rachas pero no es un collar $(4, 2)$-perfecto, pues la subsecuencia 1011, que tiene longitud 4 se repite en dos posiciones congruentes a 1 módulo 2.

□

2.4. Secuencias de rachas de longitud $2^k - 1$

Como ya dijimos, una racha de longitud k es una seguidilla del mismo símbolo, k veces, seguido por el otro. Es decir, una racha es una secuencia del mismo símbolo maximal a derecha. Y una racha es maximal cuando es además maximal a izquierda. Toda racha maximal de longitud k, tiene como subracha una racha de longitud $k-1$, iniciando en la segunda posición, una racha de longitud $k-2$, iniciando en la tercera, hasta llegar a una de longitud 1 iniciando en la última posición de la racha.

Kim y Song [6] definen las secuencias de rachas (en inglés las llaman *run sequences*).

Definición 6 (Secuencias de rachas de orden k)**.** Una secuencia de rachas de orden k es una secuencia de longitud $2^k - 1$ tal que, hay exactamente 2^{k-2-i} rachas de 0s y 2^{k-2-i} rachas de 1s de longitud i para $i \in [1, \dots, k-2]$, exactamente una racha de longitud $k-1$ de 0s y una racha de k 1s.

Un ejemplo de secuencia de racha para $k = 3$ es

$$0011101$$

pues las rachas de 0s son una de longitud 2, que incluye una de longitud 1 y una maximal de longitud 1. Las rachas de 1s son una de longitud 3 que contiene una de longitud 2 y una de longitud 1 y una maximal de longitud 1. Entonces, hay 7 rachas, de las cuales 4 son de longitud 1, 2 de longitud 2, una de las cuales es de 0s y una sola de 3 1s, cumpliendo así con la cantidad de rachas por longitud.

La defínción de Kim y Song corresponde a secuencias de longitud $2^k - 1$ que tienen la misma distribución de rachas que las llamadas *m*-secuencias de Golomb [4]. Las *m*-secuencias son secuencias de longitud $2^k - 1$ definidas por una recurrencia lineal y cumplen los postulados de aleatoriedad de Golomb [4], que incluyen la propiedad de rachas que acabamos de definir.

Esta misma secuencia que dimos como ejemplo de secuencia de racha para $k = 3$, es un ejemplo de *m*-secuencia

$$0011101$$

porque está definida por la recurrencia lineal $a_n = a_{n-1} + a_{n-3}$. Todas las *m*-secuencias son secuencias de rachas.

A su vez, todas las *m*-secuencias de longitud $2^k - 1$, son secuencias generadoras de orden k (en inglés, *span sequences*), que son secuencias de De Bruijn de orden k a las que se les quitó un 0 en la ocurrencia de la subsecuencia de n 0s. Todas las secuencias generadoras de orden k son secuencias de rachas. La sucesión de inclusiones es así

m-secuencias son
secuencias generadoras que, a su vez, son
secuencia de rachas

En su trabajo [6], Kim y Song determinan la cantidad exacta de secuencias de rachas de orden k. Prueban que cada secuencia de rachas tiene una correspondencia uno a uno con cada permutación posible de las posiciones de las rachas maximales de 0s y 1s para cada longitud, teniendo en cuenta que las rachas de 0s y las de 1s están alternadas. Kim y Song hacen sus cálculos sobre las secuencias que comienzan con $k-1$ 0s, para cada k, ya que el resto de las secuencias son rotaciones de alguna de estas. Sabiendo que primero tendremos una racha de $k-1$ 0s y luego tendremos una racha de 1s, luego una de 0s, luego una de 1s y así sucesivamente hasta completar la longitud. Cada secuencia de rachas se obtiene a partir de una colocación distinta de las rachas maximales de 0s y de 1s.

Para hacer la cuenta utilizan permutaciones multiconjunto (en inglés *multiset permutation*), que se define como

$$\binom{N}{a_1, a_2, a_3..., a_k} = \frac{N!}{a_1!a_2!a_3!...a_k!}$$

donde

$$\sum_{i=0}^{k} a_i = N.$$

Proposición 5 ([6])**.** *La cantidad de secuencias circulares de rachas de orden k está dada por la fórmula*

$$\frac{1}{2^{k-2}} \binom{2^{k-2}}{2^{k-3}, 2^{k-4}, ..., 2^0, 1}^2.$$

La división por 2^{k-2} se hace para obtener la cantidad de secuencias circulares, y evita contar a la misma secuencia tantas veces como cantidad de rachas maximales, priorizando ubicar la racha maximal de $k-1$ 0s al comienzo de la secuencia. La permutación multiconjunto está elevada al cuadrado para tomar en cuenta las permutaciones de 1s. Para este caso, no debemos agregar

tampoco una racha maximal de longitud $k-1$ ya que la única correspondiente, según la definición de propiedad de rachas, está contenida dentro de la racha maximal de longitud k.

La Tabla 3.1 compara las cantidades entre las secuencias con propiedad de rachas y las secuencias de De Bruijn. Como estas últimas son un subconjunto propio de las primeras, tiene sentido que la cantidad de secuencias de rachas sea mayor.

Proposición 6. *Las secuencias de rachas de orden k con un* 0 *agregado a la racha de* 0*s más larga son exactamente las* $(k,1)$*-rachas.*

Demostración. Dado que todas las posiciones son congruentes a 0 módulo 1, solamente hay que verificar la cantidad de rachas de cada longitud. Se agrega el 0, pues es necesario contar con una racha de longitud k de 0s. □

3. ¿CUÁNTAS SECUENCIAS DE (k, n)-RACHAS HAY?

3.1. Estadísticas

Para generar las secuencias de (k, n)-rachas se construyó un script en PYTHON en el cuál se itera sobre todas las secuencias de longitud $n2^k$ en su representación numérica en base 2. Al conjunto de secuencias se le aplica podas a partir de propiedades que deben cumplir todas las secuencias de (k, n)-rachas, antes de hacer la verificación de que cumple con la propiedad de rachas, con el fin de evitar cálculos innecesarios. Usar la representación numérica binaria nos brindó la posibilidad de hacer operaciones a nivel de bit, en vez de computar la secuencia como un *string*, que es una forma más costosa.

Las podas correspondientes consisten de las siguientes verificaciones:

- El prefijo de la secuencia debe contar con al menos $k - 1$ 0s.
- La secuencia debe terminar en 1.
- La cantidad de 0s y la cantidad de 1s debe ser la misma (balance).
- La cantidad de 0s y la cantidad de 1s debe ser múltiplo de n.
- $k > 1$ y la cantidad de rachas maximales para cada símbolo es igual a $n2^{k-1}$, la mitad de 0s y la otra mitad de 1s.

Si una secuencia pasa estas verificaciones, entonces se evalúa la propiedad de rachas sobre la misma a partir de la definición 3.

Además para evitar secuencias duplicadas, se guardaban las *rotaciones canónicas* de las secuencias de (k, n)-rachas en un conjunto implementado sobre un *Trie* o árbol de prefijos guardado en memoria. Llamamos *rotación canónica* a la rotación que minimiza el valor numérico de la secuencia, es decir, la rotación que tiene más 0s a la izquierda. Para minimizar el cómputo aún más, como el *complemento lógico* de una secuencia de rachas también cumple con la propiedad de rachas, entonces también se guarda la rotación canónica del complemento de la secuencia procesada en el *Trie*, de forma que en cada iteración se consulta si una rotación de la secuencia o de su complemento ya se procesó previamente.

Se lograron generar secuencias de longitud 32 como máximo, pues para secuencias más largas, el espacio de búsqueda era demasiado grande para la capacidad de cómputo de la computadora utilizada, además de que las secuencias largas generaban desbordamiento de memória en el *Trie* y grandes tamaños para los archivos de salida del programa. Esto podría haberse evitado si se utilizaba un lenguaje de programación que permita un manejo de memoria más eficiente, como C o C++.

Por lo tanto, las secuencias más largas que se lograron obtener corresponden con los pares $(k, n) \in \{(3, 4), (4, 2), (5, 1)\}$.

La Tabla 3.2 exhibe algunas cantidades de secuencias de (k, n)-rachas para ejemplos de k y n que fueron calculadas en la etapa de experimentación.

En las Figuras 3.1 y 3.2 se brinda una comparación visual entre la cantidad de collares (k, n)-perfectos y la cantidad de Secuencias de (k, n)-rachas contra la cantidad total de secuencias de longitud $n2^k$.

k	#Secuencias de De Bruijn	#Secuencias de rachas
2	1	1
3	2	2
4	16	36
5	2.048	88.200
6	67.108.864	7.304.587.290.000

Tabla 3.1: Cantidad de secuencias de De Bruijn y cantidad de secuencias de rachas, distintos valores de k.

k	n	#Collares (k, n)-perfectos	#Secuencias de (k, n)-rachas
2	1	1	1
2	2	2	2
2	3	6	6
2	4	16	16
2	5	52	52
3	1	2	2
3	2	16	16
3	3	172	172
3	4	2.048	2.048
4	1	16	36
4	2	2.048	10.368
5	1	2.048	88.200

Tabla 3.2: Cantidad de collares (k, n)-perfectos, y secuencias de (k, n)-rachas.

3.2. Cotas para la cantidad de secuencias de (k, n)-rachas

En esta sección damos cotas sobre la cantidad de Secuencias de (k, n)-rachas para cada par (k, n). Vamos a llamar a este número $r_{k,n}$. Todos los cálculos se hacen considerando la clase de equivalencia de las secuencias que comienzan con $k - 1$ 0s consecutivos y finalizan con un 1, ya que el resto de las secuencias que cumple la propiedad son rotaciones de alguna instancia de esta clase.

Por la Proposición 5, cuando $n = 1$ tenemos el número exacto de Secuencias de (k, n)-rachas que es la cuenta que hicieron Kim y Song [6].

Teorema 1. *Para todo par* (k, n)*,*

$$\#\text{collares } (k,n)\text{-perfectos} \leq r_{k,n} \leq \binom{n2^k - k}{n2^{k-1} - k + 1}.$$

Demostración. Comenzando por la cota inferior, según la Proposición 4, todos los collares (k, n)-perfectos cumplen con la propiedad de rachas, pero no todas las secuencias de (k, n)-rachas son collares perfectos. Por lo tanto, como todo collar perfecto es una secuencia de rachas, el número de secuencias de rachas es al menos tan grande como la cantidad de collares perfectos para un mismo par (k, n).

Luego, para la cota superior, la cuenta considera únicamente la cantidad de combinaciones de posiciones posibles en las que podemos poner los 0s restantes en una secuencia balanceada

k	n	#Collares (k,n)-perfectos	#Secuencias de (k,n)-rachas	Cota superior
2	1	1	1	2
2	2	2	2	20
2	3	6	6	252
2	4	16	16	3.432
2	5	52	52	48.620
3	1	2	2	10
3	2	16	16	1.716
3	3	172	172	352.716
3	4	2.048	2.048	77.558.760
4	1	16	36	792
4	2	2.048	10.368	37.442.160
5	1	2.048	88.200	17.383.860

Tabla 3.3: Cantidad de Secuencias de (k,n)-rachas y cotas propuestas.

teniendo en cuenta que las primeras $k-1$ posiciones de la secuencia están ocupadas por 0s y la última posición está ocupada por un 1. Las Secuencias de (k,n)-rachas son un subconjunto propio de las secuencias balanceadas de longitud $n2^k$. Esto quiere decir que hay ejemplares de secuencias balanceadas que no cumplen con la propiedad de rachas. □

En la Figura 3.1 se puede observar una comparación entre $r_{k,n}$, la cantidad de secuencias de longitud $n2^k$ y las cotas propuestas.

3.3. ¿Podemos mejorar la cota superior?

Dar una fórmula cerrada para la cantidad de secuencias de (k,n)-rachas es un problema de combinatoria y partición de un número con restricciones. No sabemos como resolverlo.

Las cotas provistas en el Teorema 1 encasillan al valor de $r_{k,n}$. La cota superior cuenta por demás porque solamente toma en cuenta la propiedad de balance entre 0s y 1s de las secuencias. Es decir, ignora la restricción de cantidad de rachas por módulo.

En esta sección consideramos una forma de dar una cota superior más ajustada que la que dimos en la sección anterior. Según la Proposición 1, para cada símbolo la cantidad de rachas maximales de longitud i en cada módulo m, para $i \leq k-2$, es 2^{k-i-2}. Debemos determinar de qué longitudes podrían ser las $2n$ rachas maximales restantes, sólo sabemos que tienen longitud mayor o igual a $k-1$. Para esto definimos el conjunto de configuraciones compatibles de rachas maximales de 0s y 1s.

Definición 7 (Configuración compatible)**.** Las rachas maximales de longitud i, con $i \in [1, k+n-1]$ en el módulo m se representan con números enteros no negativos $a_{i,m}$ para los 0s y $b_{i,m}$ para los 1s. Una configuración de rachas $(\{a_{1,0}, a_{2,0}, ..., a_{k+n-1,n-1}\}, \{b_{1,0}, b_{2,0}, ..., b_{k+n-1,n-1}\})$ es compatible si cumple que para todo módulo m, $m = 0, 1, .., n-1$

$$\sum_{i=1}^{k+n-1} \sum_{m=0}^{n-1} i\, a_{i,m} = n2^{k-1}, \qquad \sum_{i=1}^{k+n-1} \sum_{m=0}^{n-1} i\, b_{i,m} = n2^{k-1},$$

$$\sum_{i=1}^{k+n-1} a_{i,m} = 2^{k-2}, \qquad \sum_{i=1}^{k+n-1} b_{i,m} = 2^{k-2}.$$

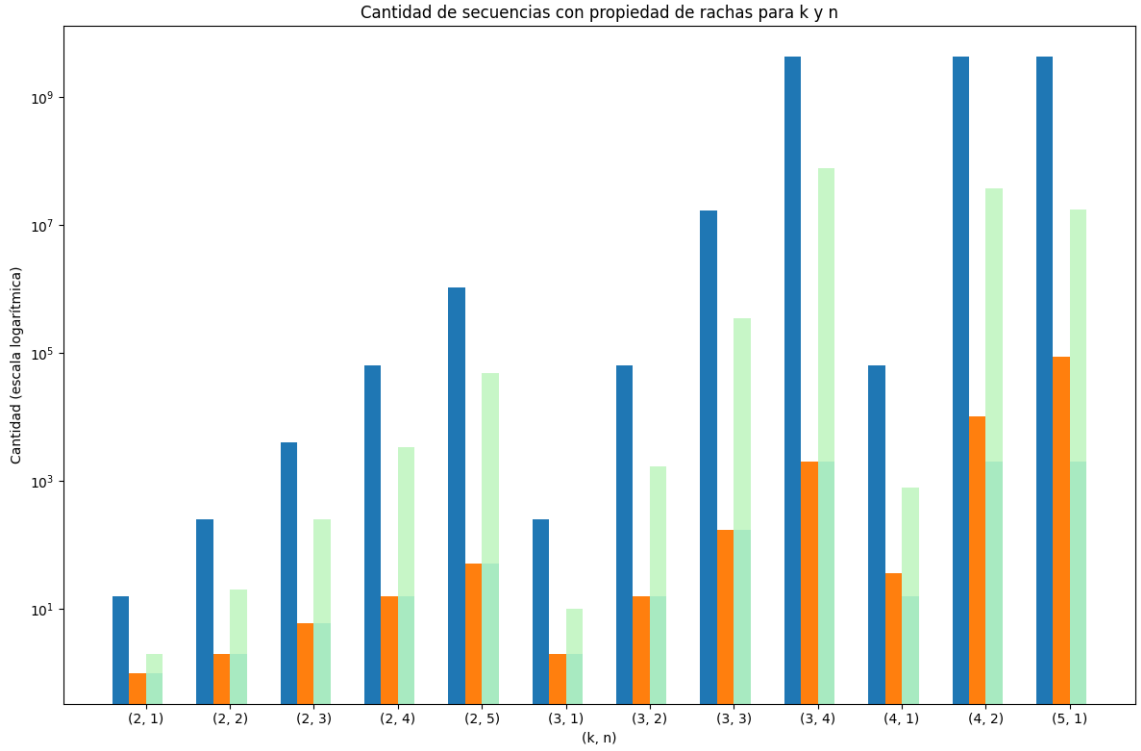

Figura 3.1: En naranja: cantidad de secuencias de (k, n)-rachas.
En turquesa: cantidad de collares (k, n)-perfectos.
En verde: cota superior propuesta en el Teorema 1.
En azul: cantidad total de secuencias de longitud $n2^k$ para distintos k y n.

Proposición 7. *Sean k y n enteros positivos con $k > 1$,*

$$r_{k,n} \leq \sum_{a,b \text{ compatibles}} \frac{1}{2^{k-2}} \prod_{m=0}^{n-1} \binom{2^{k-2}}{a_{1,m}, ..., a_{k+n-1,m}} \binom{2^{k-2}}{b_{1,m}, ..., b_{k+n-1,m}}.$$

Demostración. La fórmula indica que para cada configuración compatible se deben considerar todas las formas en las que se pueden ordenar las rachas maximales según el módulo en el que aparecen para esta configuración teniendo en cuenta la restricción de los $k - 1$ 0s iniciales. Si bien a y b son compatibles, no se tiene en cuenta el orden en que se eligen las posiciones iniciales de cada racha maximal, por lo que una racha maximal de una secuencia podría no quedar alineada a su módulo asignado. Por lo tanto estamos contando a aquellas secuencias que sí cumplen con la propiedad de rachas, pero también estamos considerando secuencias que no son válidas para esta definición. Por lo tanto el número obtenido es mayor o igual que la cantidad de secuencias de (k, n)-rachas.

□

No basta con considerar el número de configuraciones compatibles para obtener el valor de los multicombinatorios, sino que hay que saber efectivamente cuáles son todas las configuraciones y cómo se conforman para poder dar un número exacto para la cota.

A modo de ejemplo, utilizamos los casos computados para las cotas del Teorema 1 para determinar las combinaciones de longitudes de rachas de 0s y 1s. Queda como pregunta abierta y

k	n	#Secuencias de (k, n)-rachas	Cota superior anterior	Cota superior mejorada
2	1	1	2	1
2	2	2	20	2
2	3	6	252	6
2	4	16	3.432	16
2	5	52	48.620	52
3	1	2	10	2
3	2	16	1.716	40
3	3	172	352.716	672
3	4	2.048	77.558.760	9.472
4	1	16	792	36
4	2	2.048	37.442.160	31.104
5	1	2.048	17.383.860	88.200

Tabla 3.4: Cotas superiores contra la cantidad de Secuencias de (k, n)-rachas para distintos k y n.

trabajo a futuro el problema de partición con restricciones para poder dar una fórmula cerrada.

En la Figura 3.2 se pueden comparar $r_{k,n}$ contra la cantidad de secuencias de longitud $n2^k$, la cota inferior propuesta anteriormente y la cota superior mejorada.

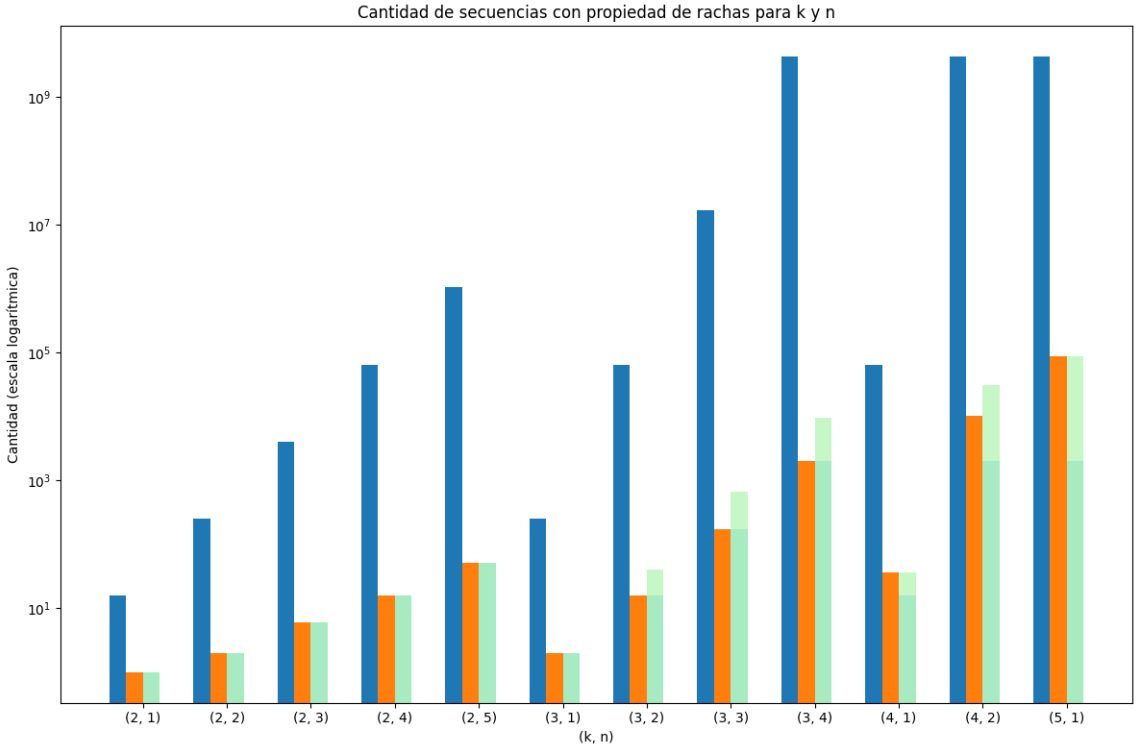

Figura 3.2: En naranja: cantidad de secuencias de (k, n)-rachas.
En turquesa: cantidad de collares (k, n)-perfectos.
En verde: una cota superior más ajustada propuesta en la Proposición 7.
En azul: cantidad total de secuencias de longitud $n2^k$.

4. SECUENCIAS DE (k, n)-RACHAS EN GRAFOS

Usar la noción de rachas maximales y considerar a una secuencia de (k, n)-rachas como una racha maximal de 0s seguida por una de 1s y así sucesivamente, como lo hacen Kim y Song [6], nos sugiere una forma de construir los collares. Sin embargo hay una diferencia importante entre la definición de Kim y Song y la nuestra: ellos tienen determinada la cantidad exacta de rachas maximales de cada longitud, y nosotros no.

Damos a continuación una caracterización de las secuencias de (k, n)-rachas mediante grafos, cuyos vértices son todas las rachas maximales. Necesitamos indicar las longitudes de todas las rachas maximales de 0s y las de 1s. Como ya dijimos, la definición de secuencia de (k, n)-rachas no especifica exactamente cuántas rachas hay de cada longitud mayor o igual que $k - 1$. Sabemos que hay una de 0s y una de 1s para cada módulo $m = 0, .., n - 1$, que tiene longitud entre $k - 1$ y $k + n - 1$. Tanto las rachas de 0s como las de 1s tienen longitudes que suman nk.

Definición 8 (Arreglo de particiones)**.** Definimos el par de arreglos $P = \{p_0, p_1\}$, donde p_0, p_1 son arreglos de longitud n de enteros no negativos tales que

- $k - 1 \leq p_0[i] \leq k + n - 1$ para todo $i \in [0, ..., n - 1]$
- $k - 1 \leq p_1[i] \leq k + n - 1$ para todo $i \in [0, ..., n - 1]$
- $\sum_{i=0}^{n-1} p_0[i] = \sum_{i=0}^{n-1} p_1[i] = nk$.

Definición 9 (Grafo de rachas maximales)**.** Dado el par de arreglos P, el grafo de rachas maximales $G_P(k, n) = (V, E)$ es aquel donde

$$V = \begin{cases} (0^s, m) & : s \in [1, ..., k-2], m \in [0, ..., n-1] \text{ con multiplicidad } 2^{k-s-2}, \\ (0^{p_0[i]}, i) & : i \in [0, ..., n-1], \\ (1^t, m) & : t \in [1, ..., k-2], m \in [0, ..., n-1] \text{ con multiplicidad } 2^{k-t-2}, \\ (1^{p_1[i]}, i) & : i \in [0, ..., n-1] \end{cases}$$

$$E = \begin{cases} ((0^s, m), (1^t, (m + s) \bmod n)), \\ ((1^t, m), (0^s, (m + t) \bmod n)) \\ \text{para todo } (0^s, m), (1^t, m) \text{ en } V. \end{cases}$$

La cantidad de vértices de $G_P(k, n)$ es la cantidad de rachas maximales de una secuencia de (k, n) rachas. Por lo tanto, $|V| = n2^{k-1}$.

Proposición 8. *$G_P(k, n)$ es bipartito y regular, es decir, todos los vértices tiene el mismo grado de entrada que de salida. Sin embargo el grafo no siempre es conexo.*

En la Figura 4.1 ambos grafos de rachas maximales son conexos. Pero en la Figura 4.2 vemos dos casos en que son disconexos.

Definición 10. Llamamos $G(k, n)$ a la familia de los grafos $G_P(k, n)$, tales que P son pares de arreglos de particiones dados en Definición 8.

Teorema 2. *Sean k y n enteros no negativos. Las secuencias de (k, n)-rachas se corresponden con todos los ciclos hamiltonianos sobre cada uno de los grafos de la familia de grafos $G(k, n) = \bigcup_P G(k, n)$ para el par (k, n), variando todos los posibles pares de arreglos de particiones P.*

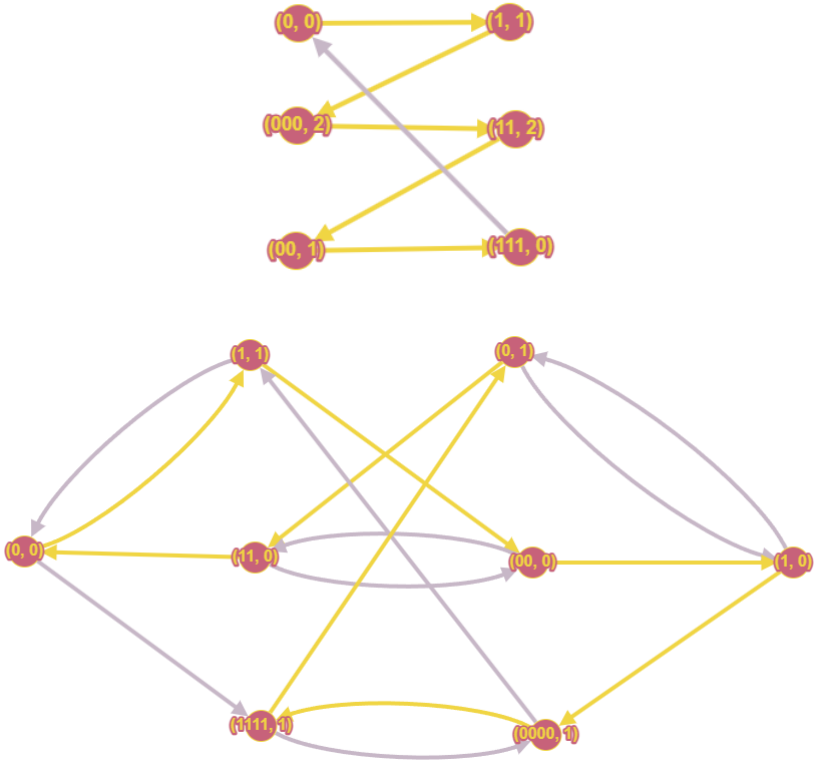

Figura 4.1: Arriba: $G_P(2, 3)$ para $p_0 = [1, 2, 3]$ y $p_1 = [3, 1, 2]$.
Define la secuencia de (k, n)-rachas 000110011101.
Abajo: $G_P(3, 2)$ con $p_0 = p_1 = [2, 4]$.
Define la secuencia de (k, n)-rachas 0000111101101001.

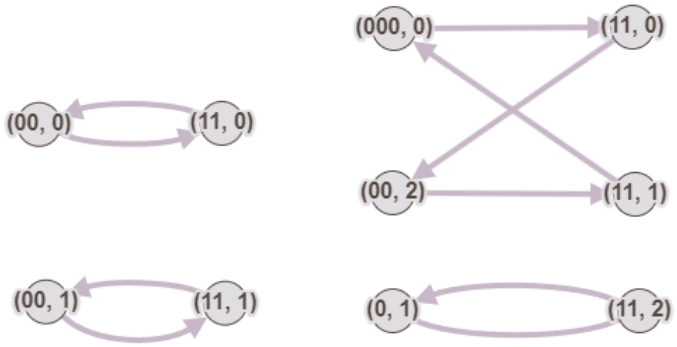

Figura 4.2: Izquierda: $G_P(2, 2)$ para $p_0 = p_1 = [2, 2]$.
Derecha: $G_P(2, 3)$ con $p_0 = [3, 1, 2]$ y $p_1 = [2, 2, 2]$.
En ninguno de los dos grafos se puede formar un ciclo hamiltoniano,
por lo que no describen ninguna secuencia de (k, n)-rachas.

Demostración. Por definición, una secuencia de (k, n)-rachas alterna rachas maximales de 0s y de 1s. Los ciclos hamiltonianos de un grafo $G_P(k, n)$ son aquellos que pasan por todos sus vértices.

Esto ocurrirá en caso de que P esté dando un grafo $G_P(k,n)$ que sea conexo. La sucesión de vértices de cada ciclo hamiltoniano en cada $G_P(k,n)$ es una secuencia de (k,n)-rachas.

Sin embargo, dos grafos distintos $G_P(k,n)$ y $G_{P'}(k,n)$ pueden dar origen a ciclos hamiltonianos distintos que denotan la misma secuencia de (k,n)-rachas. Se puede ver un ejemplo para $k=2$, $n=2$ en la Figura 4.3. □

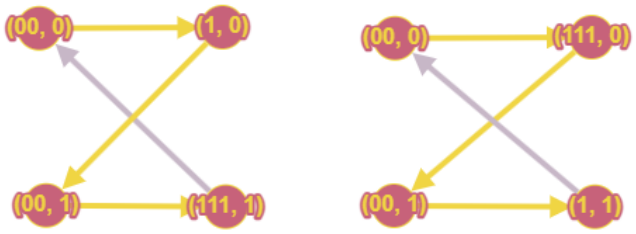

Figura 4.3: Izquierda: $G_P(2,2)$ para $p_0=[2,2]$ y $p_1=[1,3]$.
Derecha: $G_P(2,2)$ con $p_0=[2,2]$ y $p_1=[3,1]$.
Si bien los grafos son distintos, ambos ciclos hamiltonianos definen 00100111.

El Teorema 2 nos da una forma de construir secuencias de (k,n)-rachas. Dado un grafo conexo de la familia $G(k,n)$, debemos dar un ciclo hamiltoniano, y eso nos da una secuencia de (k,n)-rachas, concatenando los vértices que lo conforman en el orden indicado por los arcos. Este método permite construir todas las secuencias de (k,n)-rachas.

Por otro lado, la caracterización del Teorema 2 nos da una manera de contar cuántas secuencias de (k,n)-rachas hay. Notemos que en la cota superior presentada en la sección anterior, no teníamos en cuenta las congruencias disponibles para cada una de las rachas maximales, por lo que contábamos de más. En contraste, ahora sí estamos controlando que la sucesión de rachas maximales comience en posiciones que provengan de una partición de nk que cumpla con las restricciones dadas en la Definición 8. Podemos dar el siguiente enunciado, que aún está lejos de dar un valor numérico satisfactorio. Por un lado, no tenemos fórmula para la cantidad de pares de arreglos de particiones P que dejan al grafo conexo. Por el otro, para cada grafo conexo $G_P(k,n)$ no sabemos contar cuántos ciclos hamiltonianos tiene.

Sin embargo, a partir del Teorema 2 podemos determinar el siguiente resultado:

Corolario 1.

$$r_{k,n} \leq \sum_P \textit{cantidad de ciclos hamiltonianos en } G_P(k,n).$$

Bibliografía

[1] N. Álvarez, V. Becher, P. Ferrari, and S. Yuhjtman. Perfect necklaces. *Advances in Applied Mathematics*, 80:48 – 61, 2016.

[2] V. Becher and O. Carton. Normal numbers and computer science. In V. Berthé and editors M. Rigo, editors, *Sequences, Groups, and Number Theory*, Trends in Mathematics Series, pages 233–269. Birkhäuser/Springer, 2018.

[3] N. G. de Bruijn. A combinatorial problem. *Koninklijke Nederlandse Akademie v.Wetenschappen*, 49:758–764, 1946. Indagationes Mathematicae 8 :461-467, 1946.

[4] S. W. Golomb. *Shift register sequences*. Holden-Day, Inc., San Francisco, Calif.-Cambridge-Amsterdam, 1967. With portions co-authored by Lloyd R. Welch, Richard M. Goldstein, and Alfred W. Hales.

[5] T. Helleseth. Golomb's randomness postulates. In *Encyclopedia of Cryptography and Security*, pages 516–517. Springer, 2011.

[6] G. Kim and H. Y. Song. Statistical span property of binary run sequences. *IEEE transactions on information theory*, 69:2713–2721, 2023.

[7] W. T. Tutte. *Graph theory*, volume 21 of *Encyclopedia of Mathematics and its Applications*. Addison-Wesley Publishing Company, Advanced Book Program, Reading, MA, 1984.