

COMBINATORICS ON WORDS

MATH CIRCLE (HS1) 1/11/2015

Warmup Problems

1) Suppose you have to guess a 3 digit binary (i.e. 0's and 1's) code on a keypad.

a) How many different codes are possible?

b) Suppose that the door opens as soon as the 3 digit codes is entered. For example, if the code is 000, the door opens after 1000 is entered. Try to come up with the shortest binary sequence that is guaranteed to open the door.

For example, if we had a 2 digit code, the sequence 00110 works.

c)* Start exploring codes of length 4, length 5, etc.

2) Below are two *directed graphs* (A, B). (Note a vertex can have an edge to itself.)

Graph A consists of 4 vertices (0, 1, 2, 3) and 6 edges (a, b, c, d, e, f).

Graph B consists of 5 vertices (0, 1, 2, 3, 4) and 8 edges (a, b, c, d, e, f, g, h).

A *Eulerian cycle* is a path through ALL of the edges in a graph (using each only once) which starts and ends at the same vertex. For example, $aedcb$ is an Eulerian cycle of graph A .

a) Find all of the Eulerian cycles of graph A . Why is this not as hard as it seems?

b) Find 3 different Eulerian cycles of graph B , all starting with a . Argue that there are at least 24 different Eulerian cycles of graph B . (In fact, there are exactly 24 different Eulerian cycles of graph B .)

Words

Suppose we have an *alphabet* A (made up of *letters*). Today we will be mainly concerned with binary (with letters 0, 1) or ternary (with letters 0, 1, 2) alphabets.

A *word* w (in the alphabet A), is simply a sequence of letters. For example, 001, 101, and 111 are all binary words, and 102 is a ternary word. Note, we'll allow words of length 0, denoted by ε .

A word v is a *subword* of a word w if v is contained in w . For example, $\varepsilon, 0, 1, 00, 01, 001$ are all subwords of 001.

We let $p_w(n)$ denote the number of subwords of w with length n . For example, if $w = 001$, we have

$$p_w(0) = 1, p_w(1) = 2, p_w(2) = 2, p_w(3) = 1.$$

If u, v are two words, then we use the natural notation uv to denote the word u followed by the word v . For example, if $u = 01$ and $v = 10$, then $uv = 0110$.

1) Let $u = 101001$ and $v = 012202$.

a) Calculate, for $n = 0, 1, 2, \dots, 6$, $p_u(n)$.

b) Calculate, for $n = 0, 1, 2, \dots, 6$, $p_v(n)$.

2) Suppose we have an alphabet A of size k . Prove the following facts (where w is any word with letters from A):

(1) $p_w(n) \leq k^n$.

(2) $p_w(n) \geq p_w(n-1) - 1$.

(3) $p_w(n) \leq k \cdot p_w(n-1)$.

3) The *Champernowne word* of order m (denoted by c_m) is obtained by writing (successively) the binary representations of the natural numbers $0, 1, \dots, 2^m - 1$. For example,

$$c_1 = 0, c_2 = 011011, c_3 = 011011100101110111.$$

a) Let $w = c_3$. Calculate $p_w(0), p_w(1), p_w(2)$, and $p_w(3)$.

b) Let $w = c_m$. Prove that $p_w(n) = 2^n$ for $n < m$. Also prove that $p_w(m) < 2^m$.

4) The *Fibonacci word* of order m (denoted by f_m) is defined recursively as follows

$$f_0 = 0, f_1 = 01, f_n = f_{n-1}f_{n-2}.$$

a) Write out f_3, f_4, f_5 .

b) Let $w = f_5$. Calculate $p_w(n)$ for $n = 0, 1, 2, 3, 4, 5$.

c)* Prove by induction that the length of f_n is the $(n+2)$ Fibonacci number (the Fibonacci numbers are 0, 1, 1, 2, 3, ...).

de Bruijn Words

1) Understand the following statement: “The Champernowne word of order m is an answer to the warmup problem c) with a code of length $m - 1$.”

It is natural to want to come up with an optimal (i.e. shortest) solution to the warmup question. We will call an optimal solution to a binary code of length n a *de Bruijn* word of order n .

In other words, w is a de Bruijn word if $p_w(n) = 2^n$ and there is no shorter word v with $p_v(n) = 2^n$.

2) Let L_n be the length of a de Bruijn word of order n . Prove the following (easy) bounds:

- (1) $L_n \leq 4^n$.
- (2) $L_n \geq 2^n + n - 1$

Our aim now is to show that $L_n = 2^n + n - 1$.

Suppose we have a word w . Call u the *prefix* of w if u is all of w except for the last letter. Similarly, v is the *suffix* of w if v is all of w except for the first letter. For example, 0101 has prefix 010 and suffix 101; 1 has prefix and suffix equal to ε .

3) A de Bruijn graph of order n , denoted G_n is constructed as follows. The vertices of the graph are all binary words of length $n - 1$. There is an edge from vertex v to vertex w if the suffix of v is equal to the prefix of w . We label this edge by the last letter of w . As examples, the order 2 and order 3 de Bruijn graphs are below.

a) Draw G_4 .

b) Show that G_n has 2^{n-1} vertices and 2^n edges, each vertex has 2 edges going in and going out from it, and that there are the same number of edges labelled 0 as there are edges labelled 1.

4) We can now construct de Bruijn words of order n as follows: Step 1) Construct G_n . Step 2) Find an Eulerian cycle of G_n . Step 3) The word w consisting of the starting vertex of our Eulerian cycle followed by the Eulerian cycle is now a de Bruijn word!

a) Find de Bruijn words of order 2, 3, 4.

b) Argue that we actually are constructing de Bruijn words. That is, if we use G_n to construct a word w using the method above we have: i) the length of w is $2^n + n - 1$ and ii) $p_w(n) = 2^n$.