Assignment 2

Priyanka Emani - 1001981861
Aravind Kashyap - 1001956591
Shreyas Jagadeep - 1001888859

UNIVERSITY OF
TEXAS
ARLINGTON

CSE: 5470 Bioinformatics

February 25, 2022

1. Simulated Reads (25 points)

a) Write python code to generate a synthetic genome (a string) consisting of random nucleotide sequence of length 10,000.

b) Randomly generate a pool of synthetic "reads" (Pool A) of length 250 such that no reads overlap on the genome from 1.

c) Randomly generate a pool of synthetic "reads" (Pool B) such that every nucleotide in the genome from 1 is in a minimum of 4 reads and each read is guaranteed to overlap with another read by at least 7 nucleotides.

d) Randomly generate a pool of synthetic "reads" (Pool C) such that every nucleotide in the genome from 1 is in a minimum of 30 reads and each read is guaranteed to overlap with another read by at least 14 nucleotides. Use a random number generate to produce these read pools and submit them as text files with the assignment. Also include in your submission a text files containing your synthetic genome. Include the code to generate these in your single code file.

Use a random number generate to produce these read pools and submit them as text files with the assignment. Also include in your submission a text files containing your synthetic genome. Include the code to generate these in your single code file.

Solution

```
#!/usr/bin/env python
# coding: utf-8

# In[2]:


#import random module
from random import choice


# In[3]:


#using a list comprehension to generate a sequence
length=10000
bases=['A','C','G','T']
sequence=[choice(bases) for i in range(length)]


# In[4]:


#convert the list to a string
sequence=''.join(sequence)


# In[5]:


print(sequence)

```

```
# In[6]:


#generating pool A synthetic reads
substring=[]
for i in range(40):
    substring.append(''.join(sequence[i * 250: (i + 1) * 250]))
print(substring)


# In[9]:


#generating pool C synthetic reads
substring=[]
g=0
for i in range(0,len(sequence),250):
    if i==0:
        for k in range(14):
            substring.append(''.join(sequence[0:250]))
    substring.append(''.join(sequence[i - 7 * g: i - 7 * g + 250]))
    g+=1
print(substring)


# In[8]:


#generating pool C synthetic reads
substring=[]
g=0
for i in range(0,len(sequence),250):
    if i==0:
        for k in range(30):
            substring.append(''.join(sequence[0:250]))
    substring.append(''.join(sequence[i - 14 * g: i - 14 * g + 250]))
    g+=1
print(substring)


# In[ ]:
```

2. Greedy SCS (35 points)

Modify the below code such that it is greedy and run on your 3 files with min length=9. Include three separate text files containing the output for running your greedy SCS implementation on each of the three read pools from section 1. Also include your code.

```
import itertools

def overlap(read_a, read_b, min_length=1):
    start = 0  # start all the way at the left
    while True:
        start = read_a.find(read_b[:min_length], start)
        if start == -1:
            return 0
        if read_b.startswith(a[start:]):
            return len(read_a)-start
        start += 1

def shortestCommonSuperstring(string_set):
    shortest_sup = None
    for perm in itertools.permutations(string_set):
        sup = perm[0]
        for i in range(len(string_set)-1):
            olen = overlap(perm[i], perm[i+1], min_length=1)
            sup += perm[i+1][olen:]

        if shortest_sup is None or len(sup) < len(shortest_sup):
            shortest_sup = sup
    return shortest_sup
```

Figure 1:

Worked with:Amnah Abdelrahman, Ragava Vemnuri, Yogesh Kalapala

Solution

```
#!/usr/bin/env python
# coding: utf-8

# In[18]:


def overlap(read_a, read_b, min_length):
    max_possible_overlap_len = min(len(read_a), len(read_b))
    max_overlap_len = 0
    merged_read = ''

    # check suffix of read_a with prefix of read_b
    for i in range(max_possible_overlap_len - min_length + 1):
        checking_overlap_len = max_possible_overlap_len - i
        if (read_b.startswith(read_a[i:])):
            max_overlap_len = checking_overlap_len
```

```
            merged_read = read_a[:i] + read_b
            break

    # check suffix of read_b with prefix of read_a
    for i in range(max_possible_overlap_len - min_length + 1):
        checking_overlap_len = max_possible_overlap_len - i
        if (read_a.startswith(read_b[i:])) and max_overlap_len < checking_overlap_len:
            max_overlap_len = checking_overlap_len
            merged_read = read_b[:i] + read_a
            break

    return (max_overlap_len, merged_read)


# In[19]:


def shortestCommonSuperstring(string_set, min_length=9):
    while len(string_set)>1:
        max_overlap_len = 0
        max_merged_read = ''
        max_overlap_indices = (-1, -1)
        for i in range(len(string_set)):
            for j in range(i+1, len(string_set)):
                overlap_len, merged_read = overlap(string_set[i], string_set[j],
    min_length)
                if overlap_len > max_overlap_len:
                    max_overlap_len = overlap_len
                    max_merged_read = merged_read
                    max_overlap_indices = (i, j)
        if max_overlap_indices[0] == -1:
            string_set[0] = string_set[0] + string_set[1]
            del string_set[1]
        else:
            string_set[max_overlap_indices[0]] = max_merged_read
            del string_set[max_overlap_indices[1]]
    return string_set[0]


# In[29]:


#passing pool-A reads
strings=['
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    ACGGTGTGCGTTAGAAAGCTGCGACCGCACCCCGTGTCAACTTAAGCATTTGCCATAGGCTCCCATCCGCCCATGTGTGGCATACCGCGCCATGTTATTGTCTGATC
    ', '
    TCAATCGTGTATAGTTTGGCATAACCATGTAAGGGCCTTCGAACCGACCCCACTCTTTGCAGTCTGCGAGAAGCAAGAAGCCTTACTCGAAAGGTAAATCGGAGGCT
    ', '
    TACAAAATCAAGGGTAAAACATCCATCGAGAGAACGTTGCGGCTGGAGAGATCTCATTAAGCGTTCAATTGAGGTTATAGAGGTGAAAGCTTCGTCGCGCCCTATAC
    ', '
    CCGAAGTTTGGCAGGTACACACTCTCAGCACCCACCCCAACTAGGGCAGCCGATCCGCATAGGGGCACCAGATGATTCGTGGGCCGACTATTGGCTATGCTACGCCC
    ', '
    ACGGTGGGCACCAGGCAACACCCGATTGCTCTGCGGATACACACTTCTTGCGTTGTATGGGAGAGAATCTGTTTTCTAAACCCATTTAGGCCTCGAGACACCATTTC
    ', '
    ATCTCAGATTGTACCCCGTGTCCCAGTTTAAGGCTAAATCACATATTTCTGCGTACTTCACTAAGCACAACAAAAGTATCACCGTTAATTAACTTTGTCCTCACCGT
    ', '
    CGTATCGAGGTATCAGCATTGACGGCGCTAGTTAGTCTATAAATATTCGTTAGGAAACACAACATATAAGTATTGATATATGCCGCGGCCGAGTAGACAGTAAGAAA
    ', '
    CGAAAGGGAGGTCATGTCAATCCAGTCCACTCCATAGAGAGGTTTATACTCCGTGTAGTGGGACTTAAGACCAGGTCCACGTACTAGACTATCCCATTTTACGACAC
    ', '
```

```
GGGGCATACCCCCAAAGGCATGAGACCAACCACAGCACTCGCATTTCACTACCTTGAGGCCATGGCCGAAGGATATACGGAACTGAAAATCAACGTATAAACTGGGA
', '
TAGATCTTCGTCTCCCATGTAAGCTGGCCCTCGAGCCAGATGCCGTATAACAATGGGAGGACTGGAATCCAGATATGAGCTTCTTGCATCGCGGCGGAGTTGTACGC
', '
ATTATTTTCCGCTCCCGAAGAACCTTTTCTACAGTATTGTGCTGTCGTCAACTTTGTTCGCGCAATTGGCGTAGGGTAATGCCAGCGGTTAAGGGTTCGTACTATAC
', '
GTGAATCACAGGCCCGTCGGTGGTTGGTACCGACGAGTACATGCTAATGGAGCTTACGAGGTCCTCATCCGATACAGAGAGCTGAATTGCTAATCCCCTCCTTCAGT
', '
AGGCGGGGCGGGGGATGCGGATCAAGGCCTTGAGAATAAACTGAAATTACATATCATTTCTAGTGTGGAGAGCTGTGGCTAGTGACTGATCAGACATCAGCCGTTTC
', '
TCTTGCCAGACCCGGCAAAGGCAAGCTATATGCTTTCCACTTTAGAACAAGGCGTTAGGATACCTATCAGTGCTTCAAGTAATTCCATTGACCATTAGGTTATGTTC
', '
TATTGGAAGGGTGGGCTTCCCTTTATTCGCAGCGCGTCGCATGGGTGCATCGATCTGGTAAGCCGCTTTCCAAAGGACCACGGCCAAGACTCCCCTTTTGACATCGC
', '
ATCTATTCCCTTAGGCTTGGAACGTGGAGTCTCTCTATCGGTCATGCGGTGGCTCTCCGCTCAAGAACGCAGTCGGGGGAATGCGCTGCGTCCGACGACTGTTCAGC
', '
AGCTTAACAAACCTCAGGTAGAATTACGGAGTTGATCAACTCCACTACTGTACAGCCCATAAGGGGTGAAGAGTGGCTCATGATTGCGGATCGTTAGCTAACTCGCC
', '
GTGGCCAGGCGACTAAGAAAGGACACAGAAAAATTGCGCTCTACAGTATCTCAGCCGCCCACTAATATTATTAGACGAGCTTCATTGTCTTACAATATTTTGATATC
', '
GCATTGACTCGATAACAATAGGTTGATGCATTCACGCGCGGTCTATTGATCTTCATCCCGATAGTTATGTATGGACGTGTTCTTCTGTGAAAAAACTAGCTCTGGTC
', '
GCCGTAAATACATACCAGCGGACTAGTTGCCAAGGCTGATAGGGCGGTCATCCGCGTTGCTGCACTCCCCTATCTCATGGAGCATGTGAATCTTGGCGTTAGTCACT
', '
CTACGTAAGAAGGTCACTATCCCTGGTGTGTCAGGCAATTGATAGCGATCACGTGAATGCGGCAGTATTTATGCCCTTTCTCGAGTATAGATTTATCCCTGGTTAGC
', '
GGTCACCTAGTGCATAAGTTATAGACGATTTGTTGCAATTGGAATCTACCTATGGCGCTTAGGGAATAGTCACCGACTTAGGCAGAGTCAGTGCTAAACGCACAGGT
', '
TAGCCATTGTAGCGTCTAACCTGTCTACGCAACCAGTCCGGACCTGCGGGAAATATACACATACAGCATTGCGCTGGGCAAATGATACACAAACGTACCGTAGCCTC
', '
AAACCAGCGCTCTGATCATTCCTCGTCTAGGAGGCCTGGAGTTAACACACTCACTCAATTATGAGAGCGAGGGTCGTAACCCGCTCCCGCAGTACTGCACGCGAGTC
', '
CAGAAAACAACGATACTTTCAGCGGACGGCAGTATGTCACATGCCGGTTTCAACATTACTCACTCCGCATTGCCTCCGTCTCCGTAAAGATCTAAGCTTTGTATGTT
', '
AGGGCAAAATATTATCCCTTGTTAGGTAAGTCAGACCTTCATTATTAGGGACCATCTTCAAAAGCACGCTCTAGATAATGTTTGTAAACCGCCTCTCGTCAGCATAT
', '
CCCATGGACAATGCTATTTACGCCGGCGACGACGCTGTTGATAGAATCTTCAGGGCGTCTTACGCTAGGTGCCGTAACGCGGTTTCGGCCAGGACAGCGTTGGCCCC
', '
AGCACCCAATCACGTTAGGTGAAATCTTGCACTTTGTTCATTATCCAGATCGTGTCCCAGTATGTCAAACTAACTAAGGTACGAAACGAGTGCCTGGTTGCAACCAC
', '
CGGTACTGTCGGGTTAATACAATATCCATCCGGTCCAAAAGTACAGATCCCTTAAAGAATGATTGTGTACCTGATAAGGACCTTTAACTGATGCGGACACGTTTGCC
', '
GCGATTACGAGATGACGGAGCCATGGGAGAATAGTCTGTCGCGTCAAGACCTAACCAAAACTCTACTTTTCCCTCCTGTTATAACACTGATTGGTCAAGATTCGCCT
', '
TTACTTATTTTGATAGCGATACAGTTGCACCTATAGTATCATCATTAGTCGGGCTGGGTGTTTAGCGAGTCATTAGAGACGGACGTTCGAATGTGCCCGCAGATTCC
', '
GCCAATTGAGGTCGAGGTACTAAGCGAAGACCTTTCGCCCCAAGGGGCACTATGTGACTCACAACTACCTCCTCCTACAACGAATCAGCGCCAGCGCGAAGGTTGCA
', '
GTTTTATGTTGCGTACCCACTGTCCTCCAAATCCTGGACATCCTTGCAGACAGCAACCATCGCAAATAGTAATAGTGATCGTCTTATTACTTGGGGTAAACTCCAAC
', '
TTCTTCTGTTTGGCAGTCCGCTGTCCGTACGCTTGCATGGTAGAGGCTAGGGCCGGATATATACCTCCGCTTACTCATATGCAGAAGTTTAACGAATGCGTTCACCT
', '
CACGGTACCTGGGCCTAGAACCCTCCTTGACGCGAGGATCGAGCTGACGTCGGCACACGGTCGGTTAGCTTCCTCTGTAGTCAGTGCCGTGGAGTAACCGTCTTCGC
', '
TTAGAGGTCCGAAAGGGTCCGAATAAATCCTCATCGACCCAATGACCACCGCGGTAACCGGTCACGGGTGCGGTCAGTGAAAACAACAGAGCCCTGTTCACCGGGAA
', '
AGTCTCTTCAATGTCAAGTGTTGACGGACTCGCTTTGCTCTAGTCGACCACAAAGCGCCATACAATAATTATGCAAGGTCTCGCCATTTGGTCATACCGGGCGGGTC
', '
GACGGGTAAATGCCTCACAGTTATTCTCTATCCAACTAGGATCAATTAAAGGCGCTAAGTTGTCGAGGAAGGGGCTCAATCAGCCCTTCCCACCATCCGTCTCCAAC
', '
GTCGGGAACGCCGCCTCCTAACCCGCATCCGGCATGCGCACAAACCGCCTAGTTGTGCGAGTTCTGCAAGGTCGTTTCGTATGTGTTAAGAAGATGAGAGGATTATT
']
```

```
# In[30]:


#Shortest Common Superstring for pool-A reads
print((shortestCommonSuperstring(strings,9)))


# In[31]:


#length of shortest common superstring after passing pool-A reads
print(len(shortestCommonSuperstring(strings)))


# In[32]:


#passing pool-B reads
strings=['
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    TGGACAAACGGTGTGCGTTAGAAAGCTGCGACCGCACCCCGTGTCAACTTAAGCATTTGCCATAGGCTCCCATCCGCCCATGTGTGGCATACCGCGCCATGTTATTC
    ', '
    TCGAAGTGTTTGTTTCAATCGTGTATAGTTTGGCATAACCATGTAAGGGCCTTCGAACCGACCCCACTCTTTGCAGTCTGCGAGAAGCAAGAAGCCTTACTCGAAAC
    ', '
    GGCAGCGTGAGAGTCCTTGCTTACAAAATCAAGGGTAAAACATCCATCGAGAGAACGTTGCGGCTGGAGAGATCTCATTAAGCGTTCAATTGAGGTTATAGAGGTGA
    ', '
    TCTCCAGAGCGGAGTCCCCACTAGGAACCCGAAGTTTGGCAGGTACACACTCTCAGCACCCACCCCAACTAGGGCAGCCGATCCGCATAGGGGCACCAGATGATTCC
    ', '
    TCCATGCCTGGTACATACATTAAGTTGATAGTATTACGGTGGGCACCAGGCAACACCCGATTGCTCTGCGGATACACACTTCTTGCGTTGTATGGGAGAGAATCTGT
    ', '
    CCCTGGAGAATTTAATAGTCGTTTTTATACCAGTGGTACAGGATCTCAGATTGTACCCCGTGTCCCAGTTTAAGGCTAAATCACATATTTCTGCGTACTTCACTAAC
```

```
', '
TTATAGTTAAATAGCGTATGATCAAGTTCCACAGGTCGATAGTTGGACTCGTATCGAGGTATCAGCATTGACGGCGCTAGTTAGTCTATAAATATTCGTTAGGAAAC
', '
GTTTTGGTAAGGGAGTACCTATGATGTATTGTTCAAGAAGAGGGCACAGTAATGGACGAAAGGGAGGTCATGTCAATCCAGTCCACTCCATAGAGAGGTTTATACTC
', '
CCAGGATGTTCGATGATATCAGGCGCGAAGTATCGGAAGCGTCATATTAAGCACATATTTAATGGGGCATACCCCCAAAGGCATGAGACCAACCACAGCACTCGCAT
', '
CATTGTTCTTCGGGACTAGTAGAGGTCGTTGTCGTTGCGAACACGTGAGTTGCAAGTGCTCAAGCTTCACTAGATCTTCGTCTCCCATGTAAGCTGGCCCTCGAGCC
', '
CTGCTCTCACCCGCACGCCGGATGTCAGGGCTGGAAGATGGCTATAACTCTTAGCAGACCTGGATGTGTCGTTACCGATTATTTTCCGCTCCCGAAGAACCTTTTCT
', '
ACCATCAGACATACCGAAGCGCCGACTGACAGTTCATTATTTGACGGTGAACAGTACTGGACGCCAGTATTAATGCTGTATCATGTGAATCACAGGCCCGTCGGTGC
', '
AAGAGACCCCGCACGACACAAATGCGTAGCAAGCTGCTTATCCAGCAATGAGGTGTGGTGGTGAGCGGAGTTAACACGGGACGCATCGGAGAGGCGGGGCGGGGAT
', '
TCGATATTATATTCGACTTTAATCCTAATGGACGTCCCAAGGTCACCGGAGTGCGGACGACCGCCACCCCTCGTAAGGTGCCTACAGCACCCGACCCCTCTTGCCAC
', '
GCCTGTAAATTGCAAATAAAGCCATATTCTTGTGAATCGATGCCCCTCATATCACATTATCTACTTCTCGTATGCTAATTGGCAGCGATGAGAGAACCCTCATAGTA
', '
AGGACCGCCGCAAAAATATGGCGCTCACACAAATCGCTGTCATTCATGTGTATGACTTGCTTGTCCCGTTTCGTGCAGATTTTATGATACACTGAACGTCTATTAAC
', '
ATGGTAAAGGCTCAAAAAGTACCCGACTTTGGGATTAGACGCGAACCTGGGTACCAGCTATTAAGCATAGTTTTTTCTGAATGGCCACGGGCGGGTGGGGGAGTTTC
', '
CCGGACGATGGCGAGGTGAAGATAGTAAAGTAAGCTGTGGGCGTATATTTATTTAGATAGGACACGGGTAAGTGGTGACTATTGCGTTGTCGGTCGTAAATTTCGCT
', '
ATACGAGGACGTATGGCCGTGATGTCGCTCGTAGCGATTGCTTATGGAGCTTACCCTTGATCATGGTACGCTGTACTTGATATAGAGAAGACTTTTTGACCGGCGCC
', '
GTCACTAAGACAGAATAAAACCGAACGTTGATGCCGAGTGTAACGTATAGATGTTCATCTTATACCTGTGGAGCATGGACTTATAAGATTGATCTCTGGATGTACCC
', '
CACTAATGATCTTTTGGACATTGAGAGTCAGTGCCGATTACCGACTGTCCGTGTACGCGTACTGCAACGGTGGGACTAAATACGGCTACCTCAATTCGCACGCTGGC
', '
CCCTGGTTAGCCACTTGCAGACCGGGCATTATAGGACGGACTATCCGCCCAGAACAAGAGATTGTTATCCTTGCTCATACCACGGATGCACGCCCCCCCAGTCACTT
', '
AGTGCTAAACGCACAGGTGGTGAGGATTACTATAACCGCGGTCGCACACATGGCTAAGTAAGTCCCTCGATCCATAGTTTCACATCCACCCGGACGTAATCCACGAC
', '
TGATACACAAACGTACCGTAGCCTCCCTGCGAAGAGACTCATGCATCAATGAACTAACAAGTCGGCCTGCTATTCTGATATCTCATCCCGTCCTCCCTTAACAGCAT
', '
GTAACCCGCTCCCGCAGTACTGCACGCGAGTGATACAGTACCAGTGAAAGTCTATCGCATGTCAGGTGCAGCTAACACCCCGAACCGAACTACGCGTACGCTGCTCC
', '
ATTGCCTCCGTCTCCGTAAAGATCTAAGCTTTGTATGTTCGGTCAATCCCACCTCTGGCGTGCTTGGTTGATTTTTGCTCTACAGCTCTGGCCAATCGTTGTGCAAC
', '
AAGCACGCTCTAGATAATGTTTGTAAACCGCCTCTCGTCAGCATATTTTTCACAGAGCGGATCTGTGTAAGCTCCGGCTCGGTGCTGTTGAGTCGGATTTTGGCTTC
', '
GCGTCTTACGCTAGGTGCCGTAACGCGGTTTCGGCCAGGACAGCGTTGGCCCCAGTGGCCCCCACAAGTGAGAGCGCGTGGATAAAAACCAGGACATAACGCGCCTC
', '
GATCGTGTCCCAGTATGTCAAACTAACTAAGGTACGAAACGAGTGCCTGGTTGCAACCACCGCCAGATCCGATCATTGGTAAAATCCAGACAGTCCCGGCCCACTTA
', '
GTACAGATCCCTTAAAGAATGATTGTGTACCTGATAAGGACCTTTAACTGATGCGGACACGTTTGCCCAGCCCCGTACTCCGTTTACGGACGCGACATGCCGCAATC
', '
GTCTGTCGCGTCAAGACCTAACCAAAACTCTACTTTTCCCTCCTGTTATAACACTGATTGGTCAAGATTCGCCTAATCCGACGGAGGGGCAAAATCGAGCACAGTTA
', '
GCACCTATAGTATCATCATTAGTCGGGCTGGGTGTTTAGCGAGTCATTAGAGACGGACGTTCGAATGTGCCCGCAGATTCGGGTCTAATATAACAGCCATTTCGGCC
', '
CTAAGCGAAGACCTTTCGCCCCAAGGGGCACTATGTGACTCACAACTACCTCCTCCTACAACGAATCAGCGCCAGCGCGAAGGTTGCAAGGGTAGACGTGCGTACAT
', '
GTACCCACTGTCCTCCAAATCCTGGACATCCTTGCAGACAGCAACCATCGCAAATAGTAATAGTGATCGTCTTATTACTTGGGGTAAACTCCAACGGAGTGAGTCGC
', '
CTGTTTGGCAGTCCGCTGTCCGTACGCTTGCATGGTAGAGGCTAGGGCCGGATATATACCTCCGCTTACTCATATGCAGAAGTTTAACGAATGCGTTCACCTTGACC
', '
ATCACGGTACCTGGGCCTAGAACCCTCCTTGACGCGAGGATCGAGCTGACGTCGGCACACGGTCGGTTAGCTTCCTCTGTAGTCAGTGCCGTGGAGTAACCGTCTTC
', '
CGTTCCGCATTAGAGGTCCGAAAGGGTCCGAATAAATCCTCATCGACCCAATGACCACCGCGGTAACCGGTCACGGGTGCGGTCAGTGAAAACAACAGAGCCCTGTT
```

```
    ', '
    AGCTGGCACAAGAAGAAGTCTCTTCAATGTCAAGTGTTGACGGACTCGCTTTGCTCTAGTCGACCACAAAGCGCCATACAATAATTATGCAAGGTCTCGCCATTTGC
    ', '
    ACGCCGTGTTTTCACGTAACCAGGACGGGTAAATGCCTCACAGTTATTCTCTATCCAACTAGGATCAATTAAAGGCGCTAAGTTGTCGAGGAAGGGGCTCAATCAGC
    ']


# In[33]:


#Shortest Common Superstring for pool-B reads
print((shortestCommonSuperstring(strings,9)))


# In[34]:


#length of shortest common superstring after passing pool-B reads
print(len(shortestCommonSuperstring(strings,9)))


# In[35]:


#passing pool-C reads
strings = ['
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
```

```
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
', '
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
', '
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
', '
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
', '
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
', '
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
', '
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
', '
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
', '
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
', '
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
', '
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
', '
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
', '
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
', '
GGCTACATGGACAAACGGTGTGCGTTAGAAAGCTGCGACCGCACCCCGTGTCAACTTAAGCATTTGCCATAGGCTCCCATCCGCCCATGTGTGGCATACCGCGCCAT
', '
AGCCATACTTACCATCGAAGTGTTTGTTTCAATCGTGTATAGTTTGGCATAACCATGTAAGGGCCTTCGAACCGACCCCACTCTTTGCAGTCTGCGAGAAGCAAGAA
', '
CTCACCAACGCACGGCTTATTGGCAGCGTGAGAGTCCTTGCTTACAAAATCAAGGGTAAAACATCCATCGAGAGAACGTTGCGGCTGGAGAGATCTCATTAAGCGTT
', '
GATTTCGGAGGCATTCCCTACACTTTCCTCTCCAGAGCGGAGTCCCCACTAGGAACCCGAAGTTTGGCAGGTACACACTCTCAGCACCCACCCCAACTAGGGCAGCC
', '
GTAGGCTACGCGTCTTTATGGTCTAATTGCTGACATCCATGCCTGGTACATACATTAAGTTGATAGTATTACGGTGGGCACCAGGCAACACCCGATTGCTCTGCGGA
', '
TTTGACCTCGAGAACCGCACACAGGTTGGTGGATTCTGGCGCCCCTGGAGAATTTAATAGTCGTTTTTATACCAGTGGTACAGGATCTCAGATTGTACCCCGTGTCC
', '
CGGAGGTCCAGAAGATCGTTATAGCATGCTCCAGCGAGATGGAAGACGTTTATAGTTAAATAGCGTATGATCAAGTTCCACAGGTCGATAGTTGGACTCGTATCGAC
', '
ATTCATAGTGGAGAGTATGGCCCGTCGGCATTCCTCGGCGGGTTAACCATTACCTCGTTTTGGTAAGGGAGTACCTATGATGTATTGTTCAAGAAGAGGGCACAGTA
', '
GTCGCCTAGCAAATCCTATTAGGGGCACCTAGCGGGCACCGAGATTCGAGGTGAAGGGTGGCCCCAGGATGTTCGATGATATCAGGCGCGAAGTATCGGAAGCGTCA
', '
ACGGCGCTTTCTAGGATAAAGTACGATGTTCTGTAAGATGTTATGTCTTAACAAGACCCGAAGGCAGAGTCATTGTTCTTCGGGACTAGTAGAGGTCGTTGTCGTTC
', '
GAGTTGTACGCCTCTATGCCCGCGCATGCAGAAATGCCACCCCCACCACTCGGCGGTCGATCTGCTGCTATTGTGGACTGCTCTCACCCGCACGCCGGATGTCAGGC
', '
CAGCGGTTAAGGGTTCGTACTATAGTCGTCTTATGAAGTTTGTACGGATTCCGAGCCTCTCCCCGATCATAAAGGAATGATTTTACCATCAGACATACCGAAGCGCC
', '
CCGATACAGAGAGCTGAATTGCTAATCCCCTCCTTCAGTATTGGTCGTGTATCAACCAAGCCCGCCGGAGGAGCACTCGCAACAACACTTGAAGAGACCCCGCACGA
', '
CATTTCTAGTGTGGAGAGCTGTGGCTAGTGACTGATCAGACATCAGCCGTTTCGAAAGAGGACCTAATGACAGATTAACTAAACAAGTACGCCCTAAGTCGATATTA
', '
TTTAGAACAAGGCGTTAGGATACCTATCAGTGCTTCAAGTAATTCCATTGACCATTAGGTTATGTTGTTACGCGGATTATGTAAGCCGCTCAGTACGCTCATAAGGC
', '
TCGCAGCGCGTCGCATGGGTGCATCGATCTGGTAAGCCGCTTTCCAAAGGACCACGGCCAAGACTCCCCTTTTGACATCGGCATGTGCTGCTCATGTCTTCAATTGT
', '
AGGCTTGGAACGTGGAGTCTCTCTATCGGTCATGCGGTGGCTCTCCGCTCAAGAACGCAGTCGGGGGAATGCGCTGCGTCCGACGACTGTTCAGCGGTTCCGGGTTC
', '
CCAGCTTAACAAACCTCAGGTAGAATTACGGAGTTGATCAACTCCACTACTGTACAGCCCATAAGGGGTGAAGAGTGGCTCATGATTGCGGATCGTTAGCTAACTCC
', '
```

```
    TCGCGATTGGGGTCCGGTGGCCAGGCGACTAAGAAAGGACACAGAAAAATTGCGCTCTACAGTATCTCAGCCGCCCACTAATATTATTAGACGAGCTTCATTGTCTT
    ', '
    CGCGATTCGCCATATGGTGATATCGCTTGGGCATTGACTCGATAACAATAGGTTGATGCATTCACGCGCGGTCTATTGATCTTCATCCCGATAGTTATGTATGGACC
    ', '
    TGGATGTACCGGTTGCTTTGACAAAAGCTTGGCTTTATGGCTTCGCCGTAAATACATACCAGCGGACTAGTTGCCAAGGCTGATAGGGCGGTCATCCGCGTTGCTGC
    ', '
    CTCAATTCGCACGCTGGGAACATCAAGAAGCCTGTCGTGTGCCGAGAGTGATCTATGTCTACGTAAGAAGGTCACTATCCCTGGTGTGTCAGGCAATTGATAGCGAT
    ', '
    CGGATGCACGCCCCCCCAGTCACTTTACATCTTACGCCGAACTGGACGTGAAGGGATGAATAAGGAACGGCCGGTCACCTAGTGCATAAGTTATAGACGATTTGTTC
    ', '
    AGTTTCACATCCACCCGGACGTAATCCACGAGAAAAAGGGCAATTCTTATCACGAAATTCAGCGGAAGCCGTCGAGGTATCAGGTGTAGCCATTGTAGCGTCTAACC
    ', '
    GCTATTCTGATATCTCATCCCGTCCTCCCTTAACAGCATAGGGGCTGACTAGACACCGCGTGTGGCCGACTTCATCGCGATCGGAATAGATATAGTTAATAAACCAC
    ', '
    TCAGGTGCAGCTAACACCCCGAACCGAACTACGCGTACGCTGCTCGCCGCAATTCATTCGTTAAAAAGCACGGGCAGACGGTTAAAGGCTGTTGTATAACCTAGTTC
    ', '
    CTGGCGTGCTTGGTTGATTTTTGCTCTACAGCTCTGGCCAATCGTTGTGCAAGTAGATGACGACTATTGCCTATGGTATGTTATATGCCATTCGGATCACCGGTCTC
    ', '
    TTTCACAGAGCGGATCTGTGTAAGCTCCGGCTCGGTGCTGTTGAGTCGGATTTTGGCTTCACATTTAACTTTCGTCCTCTAAAGTCACTACAAAACATGCTGGTTCC
    ', '
    CAGCGTTGGCCCCAGTGGCCCCCACAAGTGAGAGCGCGTGGATAAAAACCAGGACATAACGCGCCTGGCCGTTCCAGGGTCCAAACTGCTGTTCCCCGTCACGGCCA
    ', '
    ACGAAACGAGTGCCTGGTTGCAACCACCGCCAGATCCGATCATTGGTAAAATCCAGACAGTCCCGGCCCACTTATACTGGCAAGAATCTACTCCAAAACGCCACGGC
    ', '
    GTACCTGATAAGGACCTTTAACTGATGCGGACACGTTTGCCCAGCCCCGTACTCCGTTTACGGACGCGACATGCCGCAATCCACGGGTGCTGAGAGCCTTTCGGACC
    ', '
    AACCAAAACTCTACTTTTCCCTCCTGTTATAACACTGATTGGTCAAGATTCGCCTAATCCGACGGAGGGGCAAAATCGAGCACAGTTAATATAGAAAGGTCTTGTCC
    ', '
    TCATCATTAGTCGGGCTGGGTGTTTAGCGAGTCATTAGAGACGGACGTTCGAATGTGCCCGCAGATTCGGGTCTAATATAACAGCCATTTCGGCGATGCCACAGTGT
    ', '
    CGAAGACCTTTCGCCCCAAGGGGCACTATGTGACTCACAACTACCTCCTCCTACAACGAATCAGCGCCAGCGCGAAGGTTGCAAGGGTAGACGTGCGTACATACAGC
    ', '
    GCGTACCCACTGTCCTCCAAATCCTGGACATCCTTGCAGACAGCAACCATCGCAAATAGTAATAGTGATCGTCTTATTACTTGGGGTAAACTCCAACGGAGTGAGTC
    ', '
    CGGATTCTTCTGTTTGGCAGTCCGCTGTCCGTACGCTTGCATGGTAGAGGCTAGGGCCGGATATATACCTCCGCTTACTCATATGCAGAAGTTTAACGAATGCGTTC
    ', '
    AGCCTACATCGATTGCATCACGGTACCTGGGCCTAGAACCCTCCTTGACGCGAGGATCGAGCTGACGTCGGCACACGGTCGGTTAGCTTCCTCTGTAGTCAGTGCCC
    ', '
    TACCTTCTCGACCCAAGCTCCATCGTTCCGCATTAGAGGTCCGAAAGGGTCCGAATAAATCCTCATCGACCCAATGACCACCGCGGTAACCGGTCACGGGTGCGGTC
    ', '
    ATACCCACCCTTATTCCAAAAACCAGTCTCAGCTGGCACAAGAAGAAGTCTCTTCAATGTCAAGTGTTGACGGACTCGCTTTGCTCTAGTCGACCACAAAGCGCCAT
    ']


# In[36]:


#Shortest Common Superstring for pool-C reads
print((shortestCommonSuperstring(strings,9)))


# In[38]:


#length of shortest common superstring after passing pool-C reads
print(len(shortestCommonSuperstring(strings,9)))


# In[ ]:
```

3. De Bruijn Graphs (40 points)

Write a de bruijn graph and run it on each of your three read pools from section 1 three times (9 total output files) with k = 1, 9, 18 for each run. Include a text file with the assembly output in your submission as well as your code.

Solution

```
#!/usr/bin/env python
# coding: utf-8

# In[3]:


class Node:
    """ Class Node to represent a vertex in the de bruijn graph """
    def __init__(self, lab):
        self.label = lab
        self.indegree = 0
        self.outdegree = 0

class Edge:
    def __init__(self, lab):
        self.label = lab

def read_reads(reads):

    return reads

def construct_graph(reads, k):
    """ Construct de bruijn graph from sets of short reads with k length word"""
    edges = dict()
    vertices = dict()

    for read in reads:
        i = 0
        while i+k < len(read):
            v1 = read[i:i+k]
            v2 = read[i+1:i+k+1]
            if v1 in edges.keys():
                vertices[v1].outdegree += 1
                edges[v1] += [Edge(v2)]
            else:
                vertices[v1] = Node(v1)
                vertices[v1].outdegree += 1
                edges[v1] = [Edge(v2)]
            if v2 in edges.keys():
                vertices[v2].indegree += 1
            else:
                vertices[v2] = Node(v2)
                vertices[v2].indegree += 1
                edges[v2] = []
            i += 1

    return (vertices, edges)

def output_contigs(g):
    """ Perform searching for Eulerian path in the graph to output genome assembly"""
    V = g[0]
```

```
    E = g[1]
    # Pick starting node (the vertex with zero in degree)
    start = V.keys()
    for k in V.keys():
        if V[k].indegree < V[start].indegree:
            start = k

    contig = start
    current = start
    while len(E[current]) > 0:
        # Pick the next node to be traversed (for now, at random)
        next = E[current][0]
        del E[current][0]
        contig += next.label[-1]
        current = next.label

    return contig

def print_graph(g):
    """ Print the information in the graph to be (somewhat) presentable """
    V = g[0]
    E = g[1]
    for k in V.keys():
        for e in E[k]:
            print(e.label)


# In[4]:


#De Bruijn Graph for pool-A for k=5
construct_graph(['
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    ACGGTGTGCGTTAGAAAGCTGCGACCGCACCCCGTGTCAACTTAAGCATTTGCCATAGGCTCCCATCCGCCCATGTGTGGCATACCGCGCCATGTTATTGTCTGATC
    ', '
    TCAATCGTGTATAGTTTGGCATAACCATGTAAGGGCCTTCGAACCGACCCCACTCTTTGCAGTCTGCGAGAAGCAAGAAGCCTTACTCGAAAGGTAAATCGGAGGCT
    ', '
    TACAAAATCAAGGGTAAAACATCCATCGAGAGAACGTTGCGGCTGGAGAGATCTCATTAAGCGTTCAATTGAGGTTATAGAGGTGAAAGCTTCGTCGCGCCCTATAC
    ', '
    CCGAAGTTTGGCAGGTACACACTCTCAGCACCCACCCCAACTAGGGCAGCCGATCCGCATAGGGGCACCAGATGATTCGTGGGCCGACTATTGGCTATGCTACGCCC
    ', '
    ACGGTGGGCACCAGGCAACACCCGATTGCTCTGCGGATACACACTTCTTGCGTTGTATGGGAGAGAATCTGTTTTCTAAACCCATTTAGGCCTCGAGACACCATTTC
    ', '
    ATCTCAGATTGTACCCCGTGTCCCAGTTTAAGGCTAAATCACATATTTCTGCGTACTTCACTAAGCACAACAAAAGTATCACCGTTAATTAACTTTGTCCTCACCGT
    ', '
    CGTATCGAGGTATCAGCATTGACGGCGCTAGTTAGTCTATAAATATTCGTTAGGAAACACAACATATAAGTATTGATATATGCCGCGGCCGAGTAGACAGTAAGAAA
    ', '
    CGAAAGGGAGGTCATGTCAATCCAGTCCACTCCATAGAGAGGTTTATACTCCGTGTAGTGGGACTTAAGACCAGGTCCACGTACTAGACTATCCCATTTTACGACAC
    ', '
    GGGGCATACCCCCAAAGGCATGAGACCAACCACAGCACTCGCATTTCACTACCTTGAGGCCATGGCCGAAGGATATACGGAACTGAAAATCAACGTATAAACTGGGA
    ', '
    TAGATCTTCGTCTCCCATGTAAGCTGGCCCTCGAGCCAGATGCCGTATAACAATGGGAGGACTGGAATCCAGATATGAGCTTCTTGCATCGCGGCGGAGTTGTACGC
    ', '
    ATTATTTTCCGCTCCCGAAGAACCTTTTCTACAGTATTGTGCTGTCGTCAACTTTGTTCGCGCAATTGGCGTAGGGTAATGCCAGCGGTTAAGGGTTCGTACTATAC
    ', '
    GTGAATCACAGGCCCGTCGGTGGTTGGTACCGACGAGTACATGCTAATGGAGCTTACGAGGTCCTCATCCGATACAGAGAGCTGAATTGCTAATCCCCTCCTTCAGT
    ', '
    AGGCGGGGCGGGGGATGCGGATCAAGGCCTTGAGAATAAACTGAAATTACATATCATTTCTAGTGTGGAGAGCTGTGGCTAGTGACTGATCAGACATCAGCCGTTTC
    ', '
    TCTTGCCAGACCCGGCAAAGGCAAGCTATATGCTTTCCACTTTAGAACAAGGCGTTAGGATACCTATCAGTGCTTCAAGTAATTCCATTGACCATTAGGTTATGTTC
    ', '
```

```
    TATTGGAAGGGTGGGCTTCCCTTTATTCGCAGCGCGTCGCATGGGTGCATCGATCTGGTAAGCCGCTTTCCAAAGGACCACGGCCAAGACTCCCCTTTTGACATCGC
    ', '
    ATCTATTCCCTTAGGCTTGGAACGTGGAGTCTCTCTATCGGTCATGCGGTGGCTCTCCGCTCAAGAACGCAGTCGGGGGAATGCGCTGCGTCCGACGACTGTTCAGC
    ', '
    AGCTTAACAAACCTCAGGTAGAATTACGGAGTTGATCAACTCCACTACTGTACAGCCCATAAGGGGTGAAGAGTGGCTCATGATTGCGGATCGTTAGCTAACTCGCC
    ', '
    GTGGCCAGGCGACTAAGAAAGGACACAGAAAAATTGCGCTCTACAGTATCTCAGCCGCCCACTAATATTATTAGACGAGCTTCATTGTCTTACAATATTTTGATATC
    ', '
    GCATTGACTCGATAACAATAGGTTGATGCATTCACGCGCGGTCTATTGATCTTCATCCCGATAGTTATGTATGGACGTGTTCTTCTGTGAAAAAACTAGCTCTGGTC
    ', '
    GCCGTAAATACATACCAGCGGACTAGTTGCCAAGGCTGATAGGGCGGTCATCCGCGTTGCTGCACTCCCCTATCTCATGGAGCATGTGAATCTTGGCGTTAGTCACT
    ', '
    CTACGTAAGAAGGTCACTATCCCTGGTGTGTCAGGCAATTGATAGCGATCACGTGAATGCGGCAGTATTTATGCCCTTTCTCGAGTATAGATTTATCCCTGGTTAGC
    ', '
    GGTCACCTAGTGCATAAGTTATAGACGATTTGTTGCAATTGGAATCTACCTATGGCGCTTAGGGAATAGTCACCGACTTAGGCAGAGTCAGTGCTAAACGCACAGGT
    ', '
    TAGCCATTGTAGCGTCTAACCTGTCTACGCAACCAGTCCGGACCTGCGGGAAATATACACATACAGCATTGCGCTGGGCAAATGATACACAAACGTACCGTAGCCTC
    ', '
    AAACCAGCGCTCTGATCATTCCTCGTCTAGGAGGCCTGGAGTTAACACACTCACTCAATTATGAGAGCGAGGGTCGTAACCCGCTCCCGCAGTACTGCACGCGAGTC
    ', '
    CAGAAAACAACGATACTTTCAGCGGACGGCAGTATGTCACATGCCGGTTTCAACATTACTCACTCCGCATTGCCTCCGTCTCCGTAAAGATCTAAGCTTTGTATGTT
    ', '
    AGGGCAAAATATTATCCCTTGTTAGGTAAGTCAGACCTTCATTATTAGGGACCATCTTCAAAAGCACGCTCTAGATAATGTTTGTAAACCGCCTCTCGTCAGCATAT
    ', '
    CCCATGGACAATGCTATTTACGCCGGCGACGACGCTGTTGATAGAATCTTCAGGGCGTCTTACGCTAGGTGCCGTAACGCGGTTTCGGCCAGGACAGCGTTGGCCCC
    ', '
    AGCACCCAATCACGTTAGGTGAAATCTTGCACTTTGTTCATTATCCAGATCGTGTCCCAGTATGTCAAACTAACTAAGGTACGAAACGAGTGCCTGGTTGCAACCAC
    ', '
    CGGTACTGTCGGGTTAATACAATATCCATCCGGTCCAAAAGTACAGATCCCTTAAAGAATGATTGTGTACCTGATAAGGACCTTTAACTGATGCGGACACGTTTGCC
    ', '
    GCGATTACGAGATGACGGAGCCATGGGAGAATAGTCTGTCGCGTCAAGACCTAACCAAAACTCTACTTTTCCCTCCTGTTATAACACTGATTGGTCAAGATTCGCCT
    ', '
    TTACTTATTTTGATAGCGATACAGTTGCACCTATAGTATCATCATTAGTCGGGCTGGGTGTTTAGCGAGTCATTAGAGACGGACGTTCGAATGTGCCCGCAGATTCC
    ', '
    GCCAATTGAGGTCGAGGTACTAAGCGAAGACCTTTCGCCCCAAGGGGCACTATGTGACTCACAACTACCTCCTCCTACAACGAATCAGCGCCAGCGCGAAGGTTGCA
    ', '
    GTTTTATGTTGCGTACCCACTGTCCTCCAAATCCTGGACATCCTTGCAGACAGCAACCATCGCAAATAGTAATAGTGATCGTCTTATTACTTGGGGTAAACTCCAAC
    ', '
    TTCTTCTGTTTGGCAGTCCGCTGTCCGTACGCTTGCATGGTAGAGGCTAGGGCCGGATATATACCTCCGCTTACTCATATGCAGAAGTTTAACGAATGCGTTCACCT
    ', '
    CACGGTACCTGGGCCTAGAACCCTCCTTGACGCGAGGATCGAGCTGACGTCGGCACACGGTCGGTTAGCTTCCTCTGTAGTCAGTGCCGTGGAGTAACCGTCTTCGC
    ', '
    TTAGAGGTCCGAAAGGGTCCGAATAAATCCTCATCGACCCAATGACCACCGCGGTAACCGGTCACGGGTGCGGTCAGTGAAAACAACAGAGCCCTGTTCACCGGGAA
    ', '
    AGTCTCTTCAATGTCAAGTGTTGACGGACTCGCTTTGCTCTAGTCGACCACAAAGCGCCATACAATAATTATGCAAGGTCTCGCCATTTGGTCATACCGGGCGGGTC
    ', '
    GACGGGTAAATGCCTCACAGTTATTCTCTATCCAACTAGGATCAATTAAAGGCGCTAAGTTGTCGAGGAAGGGGCTCAATCAGCCCTTCCCACCATCCGTCTCCAAC
    ', '
    GTCGGGAACGCCGCCTCCTAACCCGCATCCGGCATGCGCACAAACCGCCTAGTTGTGCGAGTTCTGCAAGGTCGTTTCGTATGTGTTAAGAAGATGAGAGGATTATT
    '], 5)
84
85
86  # In[10]:
87
88
89  #De Bruijn Graph for pool-A for k=9
90  construct_graph(['
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    ACGGTGTGCGTTAGAAAGCTGCGACCGCACCCCGTGTCAACTTAAGCATTTGCCATAGGCTCCCATCCGCCCATGTGTGGCATACCGCGCCATGTTATTGTCTGATC
    ', '
    TCAATCGTGTATAGTTTGGCATAACCATGTAAGGGCCTTCGAACCGACCCCACTCTTTGCAGTCTGCGAGAAGCAAGAAGCCTTACTCGAAAGGTAAATCGGAGGCT
```

```
', '
TACAAAATCAAGGGTAAAACATCCATCGAGAGAACGTTGCGGCTGGAGAGATCTCATTAAGCGTTCAATTGAGGTTATAGAGGTGAAAGCTTCGTCGCGCCCTATAC
', '
CCGAAGTTTGGCAGGTACACACTCTCAGCACCCACCCCAACTAGGGCAGCCGATCCGCATAGGGGCACCAGATGATTCGTGGGCCGACTATTGGCTATGCTACGCCC
', '
ACGGTGGGCACCAGGCAACACCCGATTGCTCTGCGGATACACACTTCTTGCGTTGTATGGGAGAGAATCTGTTTTCTAAACCCATTTAGGCCTCGAGACACCATTTC
', '
ATCTCAGATTGTACCCCGTGTCCCAGTTTAAGGCTAAATCACATATTTCTGCGTACTTCACTAAGCACAACAAAAGTATCACCGTTAATTAACTTTGTCCTCACCGT
', '
CGTATCGAGGTATCAGCATTGACGGCGCTAGTTAGTCTATAAATATTCGTTAGGAAACACAACATATAAGTATTGATATATGCCGCGGCCGAGTAGACAGTAAGAAA
', '
CGAAAGGGAGGTCATGTCAATCCAGTCCACTCCATAGAGAGGTTTATACTCCGTGTAGTGGGACTTAAGACCAGGTCCACGTACTAGACTATCCCATTTTACGACAC
', '
GGGGCATACCCCCAAAGGCATGAGACCAACCACAGCACTCGCATTTCACTACCTTGAGGCCATGGCCGAAGGATATACGGAACTGAAAATCAACGTATAAACTGGGA
', '
TAGATCTTCGTCTCCCATGTAAGCTGGCCCTCGAGCCAGATGCCGTATAACAATGGGAGGACTGGAATCCAGATATGAGCTTCTTGCATCGCGGCGGAGTTGTACGC
', '
ATTATTTTCCGCTCCCGAAGAACCTTTTCTACAGTATTGTGCTGTCGTCAACTTTGTTCGCGCAATTGGCGTAGGGTAATGCCAGCGGTTAAGGGTTCGTACTATAC
', '
GTGAATCACAGGCCCGTCGGTGGTTGGTACCGACGAGTACATGCTAATGGAGCTTACGAGGTCCTCATCCGATACAGAGAGCTGAATTGCTAATCCCCTCCTTCAGT
', '
AGGCGGGGCGGGGGATGCGGATCAAGGCCTTGAGAATAAACTGAAATTACATATCATTTCTAGTGTGGAGAGCTGTGGCTAGTGACTGATCAGACATCAGCCGTTTC
', '
TCTTGCCAGACCCGGCAAAGGCAAGCTATATGCTTTCCACTTTAGAACAAGGCGTTAGGATACCTATCAGTGCTTCAAGTAATTCCATTGACCATTAGGTTATGTTC
', '
TATTGGAAGGGTGGGCTTCCCTTTATTCGCAGCGCGTCGCATGGGTGCATCGATCTGGTAAGCCGCTTTCCAAAGGACCACGGCCAAGACTCCCCTTTTGACATCGC
', '
ATCTATTCCCTTAGGCTTGGAACGTGGAGTCTCTCTATCGGTCATGCGGTGGCTCTCCGCTCAAGAACGCAGTCGGGGGAATGCGCTGCGTCCGACGACTGTTCAGC
', '
AGCTTAACAAACCTCAGGTAGAATTACGGAGTTGATCAACTCCACTACTGTACAGCCCATAAGGGGTGAAGAGTGGCTCATGATTGCGGATCGTTAGCTAACTCGCC
', '
GTGGCCAGGCGACTAAGAAAGGACACAGAAAAATTGCGCTCTACAGTATCTCAGCCGCCCACTAATATTATTAGACGAGCTTCATTGTCTTACAATATTTTGATATC
', '
GCATTGACTCGATAACAATAGGTTGATGCATTCACGCGCGGTCTATTGATCTTCATCCCGATAGTTATGTATGGACGTGTTCTTCTGTGAAAAAACTAGCTCTGGTC
', '
GCCGTAAATACATACCAGCGGACTAGTTGCCAAGGCTGATAGGGCGGTCATCCGCGTTGCTGCACTCCCCTATCTCATGGAGCATGTGAATCTTGGCGTTAGTCACT
', '
CTACGTAAGAAGGTCACTATCCCTGGTGTGTCAGGCAATTGATAGCGATCACGTGAATGCGGCAGTATTTATGCCCTTTCTCGAGTATAGATTTATCCCTGGTTAGC
', '
GGTCACCTAGTGCATAAGTTATAGACGATTTGTTGCAATTGGAATCTACCTATGGCGCTTAGGGAATAGTCACCGACTTAGGCAGAGTCAGTGCTAAACGCACAGGT
', '
TAGCCATTGTAGCGTCTAACCTGTCTACGCAACCAGTCCGGACCTGCGGGAAATATACACATACAGCATTGCGCTGGGCAAATGATACACAAACGTACCGTAGCCTC
', '
AAACCAGCGCTCTGATCATTCCTCGTCTAGGAGGCCTGGAGTTAACACACTCACTCAATTATGAGAGCGAGGGTCGTAACCCGCTCCCGCAGTACTGCACGCGAGTC
', '
CAGAAAACAACGATACTTTCAGCGGACGGCAGTATGTCACATGCCGGTTTCAACATTACTCACTCCGCATTGCCTCCGTCTCCGTAAAGATCTAAGCTTTGTATGTT
', '
AGGGCAAAATATTATCCCTTGTTAGGTAAGTCAGACCTTCATTATTAGGGACCATCTTCAAAAGCACGCTCTAGATAATGTTTGTAAACCGCCTCTCGTCAGCATAT
', '
CCCATGGACAATGCTATTTACGCCGGCGACGACGCTGTTGATAGAATCTTCAGGGCGTCTTACGCTAGGTGCCGTAACGCGGTTTCGGCCAGGACAGCGTTGGCCCC
', '
AGCACCCAATCACGTTAGGTGAAATCTTGCACTTTGTTCATTATCCAGATCGTGTCCCAGTATGTCAAACTAACTAAGGTACGAAACGAGTGCCTGGTTGCAACCAC
', '
CGGTACTGTCGGGTTAATACAATATCCATCCGGTCCAAAAGTACAGATCCCTTAAAGAATGATTGTGTACCTGATAAGGACCTTTAACTGATGCGGACACGTTTGCC
', '
GCGATTACGAGATGACGGAGCCATGGGAGAATAGTCTGTCGCGTCAAGACCTAACCAAAACTCTACTTTTCCCTCCTGTTATAACACTGATTGGTCAAGATTCGCCT
', '
TTACTTATTTTGATAGCGATACAGTTGCACCTATAGTATCATCATTAGTCGGGCTGGGTGTTTAGCGAGTCATTAGAGACGGACGTTCGAATGTGCCCGCAGATTCC
', '
GCCAATTGAGGTCGAGGTACTAAGCGAAGACCTTTCGCCCCAAGGGGCACTATGTGACTCACAACTACCTCCTCCTACAACGAATCAGCGCCAGCGCGAAGGTTGCA
', '
GTTTTATGTTGCGTACCCACTGTCCTCCAAATCCTGGACATCCTTGCAGACAGCAACCATCGCAAATAGTAATAGTGATCGTCTTATTACTTGGGGTAAACTCCAAC
```

```
    ', '
    TTCTTCTGTTTGGCAGTCCGCTGTCCGTACGCTTGCATGGTAGAGGCTAGGGCCGGATATATACCTCCGCTTACTCATATGCAGAAGTTTAACGAATGCGTTCACCT
    ', '
    CACGGTACCTGGGCCTAGAACCCTCCTTGACGCGAGGATCGAGCTGACGTCGGCACACGGTCGGTTAGCTTCCTCTGTAGTCAGTGCCGTGGAGTAACCGTCTTCGC
    ', '
    TTAGAGGTCCGAAAGGGTCCGAATAAATCCTCATCGACCCAATGACCACCGCGGTAACCGGTCACGGGTGCGGTCAGTGAAAACAACAGAGCCCTGTTCACCGGAA
    ', '
    AGTCTCTTCAATGTCAAGTGTTGACGGACTCGCTTTGCTCTAGTCGACCACAAAGCGCCATACAATAATTATGCAAGGTCTCGCCATTTGGTCATACCGGGCGGGTC
    ', '
    GACGGGTAAATGCCTCACAGTTATTCTCTATCCAACTAGGATCAATTAAAGGCGCTAAGTTGTCGAGGAAGGGGCTCAATCAGCCCTTCCCACCATCCGTCTCCAAC
    ', '
    GTCGGGAACGCCGCCTCCTAACCCGCATCCGGCATGCGCACAAACCGCCTAGTTGTGCGAGTTCTGCAAGGTCGTTTCGTATGTGTTAAGAAGATGAGAGGATTATT
    '], 9)

# In[12]:


#De Bruijn Graph for pool-A for k=18
construct_graph(['
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    ACGGTGTGCGTTAGAAAGCTGCGACCGCACCCCGTGTCAACTTAAGCATTTGCCATAGGCTCCCATCCGCCCATGTGTGGCATACCGCGCCATGTTATTGTCTGATC
    ', '
    TCAATCGTGTATAGTTTGGCATAACCATGTAAGGGCCTTCGAACCGACCCCACTCTTTGCAGTCTGCGAGAAGCAAGAAGCCTTACTCGAAAGGTAAATCGGAGGCT
    ', '
    TACAAAATCAAGGGTAAAACATCCATCGAGAGAACGTTGCGGCTGGAGAGATCTCATTAAGCGTTCAATTGAGGTTATAGAGGTGAAAGCTTCGTCGCGCCCTATAC
    ', '
    CCGAAGTTTGGCAGGTACACACTCTCAGCACCCACCCCAACTAGGGCAGCCGATCCGCATAGGGGCACCAGATGATTCGTGGGCCGACTATTGGCTATGCTACGCCC
    ', '
    ACGGTGGGCACCAGGCAACACCCGATTGCTCTGCGGATACACACTTCTTGCGTTGTATGGGAGAGAATCTGTTTTCTAAACCCATTTAGGCCTCGAGACACCATTTC
    ', '
    ATCTCAGATTGTACCCCGTGTCCCAGTTTAAGGCTAAATCACATATTTCTGCGTACTTCACTAAGCACAACAAAAGTATCACCGTTAATTAACTTTGTCCTCACCGT
    ', '
    CGTATCGAGGTATCAGCATTGACGGCGCTAGTTAGTCTATAAATATTCGTTAGGAAACACAACATATAAGTATTGATATATGCCGCGGCCGAGTAGACAGTAAGAAA
    ', '
    CGAAAGGGAGGTCATGTCAATCCAGTCCACTCCATAGAGAGGTTTATACTCCGTGTAGTGGGACTTAAGACCAGGTCCACGTACTAGACTATCCCATTTTACGACAC
    ', '
    GGGGCATACCCCCAAAGGCATGAGACCAACCACAGCACTCGCATTTCACTACCTTGAGGCCATGGCCGAAGGATATACGGAACTGAAAATCAACGTATAAACTGGGA
    ', '
    TAGATCTTCGTCTCCCATGTAAGCTGGCCCTCGAGCCAGATGCCGTATAACAATGGGAGGACTGGAATCCAGATATGAGCTTCTTGCATCGCGGCGGAGTTGTACGC
    ', '
    ATTATTTTCCGCTCCCGAAGAACCTTTTCTACAGTATTGTGCTGTCGTCAACTTTGTTCGCGCAATTGGCGTAGGGTAATGCCAGCGGTTAAGGGTTCGTACTATAC
    ', '
    GTGAATCACAGGCCCGTCGGTGGTTGGTACCGACGAGTACATGCTAATGGAGCTTACGAGGTCCTCATCCGATACAGAGAGCTGAATTGCTAATCCCCTCCTTCAGT
    ', '
    AGGCGGGGCGGGGGATGCGGATCAAGGCCTTGAGAATAAACTGAAATTACATATCATTTCTAGTGTGGAGAGCTGTGGCTAGTGACTGATCAGACATCAGCCGTTTC
    ', '
    TCTTGCCAGACCCGGCAAAGGCAAGCTATATGCTTTCCACTTTAGAACAAGGCGTTAGGATACCTATCAGTGCTTCAAGTAATTCCATTGACCATTAGGTTATGTTC
    ', '
    TATTGGAAGGGTGGGCTTCCCTTTATTCGCAGCGCGTCGCATGGGTGCATCGATCTGGTAAGCCGCTTTCCAAAGGACCACGGCCAAGACTCCCCTTTTGACATCGG
    ', '
    ATCTATTCCCTTAGGCTTGGAACGTGGAGTCTCTCTATCGGTCATGCGGTGGCTCTCCGCTCAAGAACGCAGTCGGGGGAATGCGCTGCGTCCGACGACTGTTCAGC
    ', '
    AGCTTAACAAACCTCAGGTAGAATTACGGAGTTGATCAACTCCACTACTGTACAGCCCATAAGGGGTGAAGAGTGGCTCATGATTGCGGATCGTTAGCTAACTCGCC
    ', '
    GTGGCCAGGCGACTAAGAAAGGACACAGAAAAATTGCGCTCTACAGTATCTCAGCCGCCCACTAATATTATTAGACGAGCTTCATTGTCTTACAATATTTTGATATC
    ', '
    GCATTGACTCGATAACAATAGGTTGATGCATTCACGCGCGGTCTATTGATCTTCATCCCGATAGTTATGTATGGACGTGTTCTTCTGTGAAAAAACTAGCTCTGGTC
    ', '
    GCCGTAAATACATACCAGCGGACTAGTTGCCAAGGCTGATAGGGCGGTCATCCGCGTTGCTGCACTCCCCTATCTCATGGAGCATGTGAATCTTGGCGTTAGTCACT
    ', '
```

```
    CTACGTAAGAAGGTCACTATCCCTGGTGTGTCAGGCAATTGATAGCGATCACGTGAATGCGGCAGTATTTATGCCCTTTCTCGAGTATAGATTTATCCCTGGTTAGC
    ', '
    GGTCACCTAGTGCATAAGTTATAGACGATTTGTTGCAATTGGAATCTACCTATGGCGCTTAGGGAATAGTCACCGACTTAGGCAGAGTCAGTGCTAAACGCACAGGT
    ', '
    TAGCCATTGTAGCGTCTAACCTGTCTACGCAACCAGTCCGGACCTGCGGGAAATATACACATACAGCATTGCGCTGGGCAAATGATACACAAACGTACCGTAGCCTC
    ', '
    AAACCAGCGCTCTGATCATTCCTCGTCTAGGAGGCCTGGAGTTAACACACTCACTCAATTATGAGAGCGAGGGTCGTAACCCGCTCCCGCAGTACTGCACGCGAGTC
    ', '
    CAGAAAACAACGATACTTTCAGCGGACGGCAGTATGTCACATGCCGGTTTCAACATTACTCACTCCGCATTGCCTCCGTCTCCGTAAAGATCTAAGCTTTGTATGTT
    ', '
    AGGGCAAAATATTATCCCTTGTTAGGTAAGTCAGACCTTCATTATTAGGGACCATCTTCAAAAGCACGCTCTAGATAATGTTTGTAAACCGCCTCTCGTCAGCATAT
    ', '
    CCCATGGACAATGCTATTTACGCCGGCGACGACGCTGTTGATAGAATCTTCAGGGCGTCTTACGCTAGGTGCCGTAACGCGGTTTCGGCCAGGACAGCGTTGGCCCC
    ', '
    AGCACCCAATCACGTTAGGTGAAATCTTGCACTTTGTTCATTATCCAGATCGTGTCCCAGTATGTCAAACTAACTAAGGTACGAAACGAGTGCCTGGTTGCAACCAC
    ', '
    CGGTACTGTCGGGTTAATACAATATCCATCCGGTCCAAAAGTACAGATCCCTTAAAGAATGATTGTGTACCTGATAAGGACCTTTAACTGATGCGGACACGTTTGCC
    ', '
    GCGATTACGAGATGACGGAGCCATGGGAGAATAGTCTGTCGCGTCAAGACCTAACCAAAACTCTACTTTTCCCTCCTGTTATAACACTGATTGGTCAAGATTCGCCT
    ', '
    TTACTTATTTTGATAGCGATACAGTTGCACCTATAGTATCATCATTAGTCGGGCTGGGTGTTTAGCGAGTCATTAGAGACGGACGTTCGAATGTGCCCGCAGATTCC
    ', '
    GCCAATTGAGGTCGAGGTACTAAGCGAAGACCTTTCGCCCCAAGGGGCACTATGTGACTCACAACTACCTCCTCCTACAACGAATCAGCGCCAGCGCGAAGGTTGCA
    ', '
    GTTTTATGTTGCGTACCCACTGTCCTCCAAATCCTGGACATCCTTGCAGACAGCAACCATCGCAAATAGTAATAGTGATCGTCTTATTACTTGGGGTAAACTCCAAC
    ', '
    TTCTTCTGTTTGGCAGTCCGCTGTCCGTACGCTTGCATGGTAGAGGCTAGGGCCGGATATATACCTCCGCTTACTCATATGCAGAAGTTTAACGAATGCGTTCACCT
    ', '
    CACGGTACCTGGGCCTAGAACCCTCCTTGACGCGAGGATCGAGCTGACGTCGGCACACGGTCGGTTAGCTTCCTCTGTAGTCAGTGCCGTGGAGTAACCGTCTTCGG
    ', '
    TTAGAGGTCCGAAAGGGTCCGAATAAATCCTCATCGACCCAATGACCACCGCGGTAACCGGTCACGGGTGCGGTCAGTGAAAACAACAGAGCCCTGTTCACCGGGAA
    ', '
    AGTCTCTTCAATGTCAAGTGTTGACGGACTCGCTTTGCTCTAGTCGACCACAAAGCGCCATACAATAATTATGCAAGGTCTCGCCATTTGGTCATACCGGGCGGGTC
    ', '
    GACGGGTAAATGCCTCACAGTTATTCTCTATCCAACTAGGATCAATTAAAGGCGCTAAGTTGTCGAGGAAGGGGCTCAATCAGCCCTTCCCACCATCCGTCTCCAAC
    ', '
    GTCGGGAACGCCGCCTCCTAACCCGCATCCGGCATGCGCACAAACCGCCTAGTTGTGCGAGTTCTGCAAGGTCGTTTCGTATGTGTTAAGAAGATGAGAGGATTATT
    '], 18)


# In[13]:


#De Bruijn Graph for pool-B for k=5
construct_graph(['
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
```

```
', '
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
', '
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
', '
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
', '
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
', '
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
', '
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
', '
TGGACAAACGGTGTGCGTTAGAAAGCTGCGACCGCACCCCGTGTCAACTTAAGCATTTGCCATAGGCTCCCATCCGCCCATGTGTGGCATACCGCGCCATGTTATTC
', '
TCGAAGTGTTTGTTTCAATCGTGTATAGTTTGGCATAACCATGTAAGGGCCTTCGAACCGACCCCACTCTTTGCAGTCTGCGAGAAGCAAGAAGCCTTACTCGAAAC
', '
GGCAGCGTGAGAGTCCTTGCTTACAAAATCAAGGGTAAAACATCCATCGAGAGAACGTTGCGGCTGGAGAGATCTCATTAAGCGTTCAATTGAGGTTATAGAGGTGA
', '
TCTCCAGAGCGGAGTCCCCACTAGGAACCCGAAGTTTGGCAGGTACACACTCTCAGCACCCACCCCAACTAGGGCAGCCGATCCGCATAGGGGCACCAGATGATTCC
', '
TCCATGCCTGGTACATACATTAAGTTGATAGTATTACGGTGGGCACCAGGCAACACCCGATTGCTCTGCGGATACACACTTCTTGCGTTGTATGGGAGAGAATCTGT
', '
CCCTGGAGAATTTAATAGTCGTTTTTATACCAGTGGTACAGGATCTCAGATTGTACCCCGTGTCCCAGTTTAAGGCTAAATCACATATTTCTGCGTACTTCACTAAC
', '
TTATAGTTAAATAGCGTATGATCAAGTTCCACAGGTCGATAGTTGGACTCGTATCGAGGTATCAGCATTGACGGCGCTAGTTAGTCTATAAATATTCGTTAGGAAAC
', '
GTTTTGGTAAGGGAGTACCTATGATGTATTGTTCAAGAAGAGGGCACAGTAATGGACGAAAGGGAGGTCATGTCAATCCAGTCCACTCCATAGAGAGGTTTATACTC
', '
CCAGGATGTTCGATGATATCAGGCGCGAAGTATCGGAAGCGTCATATTAAGCACATATTTAATGGGGCATACCCCCAAAGGCATGAGACCAACCACAGCACTCGCAT
', '
CATTGTTCTTCGGGACTAGTAGAGGTCGTTGTCGTTGCGAACACGTGAGTTGCAAGTGCTCAAGCTTCACTAGATCTTCGTCTCCCATGTAAGCTGGCCCTCGAGCC
', '
CTGCTCTCACCCGCACGCCGGATGTCAGGGCTGGAAGATGGCTATAACTCTTAGCAGACCTGGATGTGTCGTTACCGATTATTTTCCGCTCCCGAAGAACCTTTTCT
', '
ACCATCAGACATACCGAAGCGCCGACTGACAGTTCATTATTTGACGGTGAACAGTACTGGACGCCAGTATTAATGCTGTATCATGTGAATCACAGGCCCGTCGGTGC
', '
AAGAGACCCCGCACGACACAAATGCGTAGCAAGCTGCTTATCCAGCAATGAGGTGTGGTGGTGAGCGGAGTTAACACGGGACGCATCGGAGAGGCGGGGCGGGGGAT
', '
TCGATATTATATTCGACTTTAATCCTAATGGACGTCCCAAGGTCACCGGAGTGCGGACGACCGCCACCCCTCGTAAGGTGCCTACAGCACCCGACCCCTCTTGCCAC
', '
GCCTGTAAATTGCAAATAAAGCCATATTCTTGTGAATCGATGCCCCTCATATCACATTATCTACTTCTCGTATGCTAATTGGCAGCGATGAGAGAACCCTCATAGTA
', '
AGGACCGCCGCAAAAATATGGCGCTCACACAAATCGCTGTCATTCATGTGTATGACTTGCTTGTCCCGTTTCGTGCAGATTTTATGATACACTGAACGTCTATTAAC
', '
ATGGTAAAGGCTCAAAAAGTACCCGACTTTGGGATTAGACGCGAACCTGGGTACCAGCTATTAAGCATAGTTTTTTCTGAATGGCCACGGGCGGGTGGGGGAGTTTC
', '
CCGGACGATGGCGAGGTGAAGATAGTAAAGTAAGCTGTGGGCGTATATTTATTTAGATAGGACACGGGTAAGTGGTGACTATTGCGTTGTCGGTCGTAAATTTCGCT
', '
ATACGAGGACGTATGGCCGTGATGTCGCTCGTAGCGATTGCTTATGGAGCTTACCCTTGATCATGGTACGCTGTACTTGATATAGAGAAGACTTTTTGACCGGCGCC
', '
GTCACTAAGACAGAATAAAACCGAACGTTGATGCCGAGTGTAACGTATAGATGTTCATCTTATACCTGTGGAGCATGGACTTATAAGATTGATCTCTGGATGTACCC
', '
CACTAATGATCTTTTGGACATTGAGAGTCAGTGCCGATTACCGACTGTCCGTGTACGCGTACTGCAACGGTGGGACTAAATACGGCTACCTCAATTCGCACGCTGGC
', '
CCCTGGTTAGCCACTTGCAGACCGGGCATTATAGGACGGACTATCCGCCCAGAACAAGAGATTGTTATCCTTGCTCATACCACGGATGCACGCCCCCCCAGTCACTT
', '
AGTGCTAAACGCACAGGTGGTGAGGATTACTATAACCGCGGTCGCACACATGGCTAAGTAAGTCCCTCGATCCATAGTTTCACATCCACCCGGACGTAATCCACGAC
', '
TGATACACAAACGTACCGTAGCCTCCCTGCGAAGAGACTCATGCATCAATGAACTAACAAGTCGGCCTGCTATTCTGATATCTCATCCCGTCCTCCCTTAACAGCAT
', '
GTAACCCGCTCCCGCAGTACTGCACGCGAGTGATACAGTACCAGTGAAAGTCTATCGCATGTCAGGTGCAGCTAACACCCCGAACCGAACTACGCGTACGCTGCTCC
```

```
    ', '
    ATTGCCTCCGTCTCCGTAAAGATCTAAGCTTTGTATGTTCGGTCAATCCCACCTCTGGCGTGCTTGGTTGATTTTTGCTCTACAGCTCTGGCCAATCGTTGTGCAAC
    ', '
    AAGCACGCTCTAGATAATGTTTGTAAACCGCCTCTCGTCAGCATATTTTTCACAGAGCGGATCTGTGTAAGCTCCGGCTCGGTGCTGTTGAGTCGGATTTTGGCTTC
    ', '
    GCGTCTTACGCTAGGTGCCGTAACGCGGTTTCGGCCAGGACAGCGTTGGCCCCAGTGGCCCCCACAAGTGAGAGCGCGTGGATAAAAACCAGGACATAACGCGCCTC
    ', '
    GATCGTGTCCCAGTATGTCAAACTAACTAAGGTACGAAACGAGTGCCTGGTTGCAACCACCGCCAGATCCGATCATTGGTAAAATCCAGACAGTCCCGGCCCACTTA
    ', '
    GTACAGATCCCTTAAAGAATGATTGTGTACCTGATAAGGACCTTTAACTGATGCGGACACGTTTGCCCAGCCCCGTACTCCGTTTACGGACGCGACATGCCGCAATC
    ', '
    GTCTGTCGCGTCAAGACCTAACCAAAACTCTACTTTTCCCTCCTGTTATAACACTGATTGGTCAAGATTCGCCTAATCCGACGGAGGGGCAAAATCGAGCACAGTTA
    ', '
    GCACCTATAGTATCATCATTAGTCGGGCTGGGTGTTTAGCGAGTCATTAGAGACGGACGTTCGAATGTGCCCGCAGATTCGGGTCTAATATAACAGCCATTTCGGCC
    ', '
    CTAAGCGAAGACCTTTCGCCCCAAGGGGCACTATGTGACTCACAACTACCTCCTCCTACAACGAATCAGCGCCAGCGCGAAGGTTGCAAGGGTAGACGTGCGTACAT
    ', '
    GTACCCACTGTCCTCCAAATCCTGGACATCCTTGCAGACAGCAACCATCGCAAATAGTAATAGTGATCGTCTTATTACTTGGGGTAAACTCCAACGGAGTGAGTCGC
    ', '
    CTGTTTGGCAGTCCGCTGTCCGTACGCTTGCATGGTAGAGGCTAGGGCCGGATATATACCTCCGCTTACTCATATGCAGAAGTTTAACGAATGCGTTCACCTTGACC
    ', '
    ATCACGGTACCTGGGCCTAGAACCCTCCTTGACGCGAGGATCGAGCTGACGTCGGCACACGGTCGGTTAGCTTCCTCTGTAGTCAGTGCCGTGGAGTAACCGTCTTC
    ', '
    CGTTCCGCATTAGAGGTCCGAAAGGGTCCGAATAAATCCTCATCGACCCAATGACCACCGCGGTAACCGGTCACGGGTGCGGTCAGTGAAAACAACAGAGCCCTGTT
    ', '
    AGCTGGCACAAGAAGAAGTCTCTTCAATGTCAAGTGTTGACGGACTCGCTTTGCTCTAGTCGACCACAAAGCGCCATACAATAATTATGCAAGGTCTCGCCATTTGC
    ', '
    ACGCCGTGTTTTCACGTAACCAGGACGGGTAAATGCCTCACAGTTATTCTCTATCCAACTAGGATCAATTAAAGGCGCTAAGTTGTCGAGGAAGGGGCTCAATCAGC
    '], 5)


# In[14]:


#De Bruijn Graph for pool-B for k=9
construct_graph(['
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
```

```
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
', '
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
', '
TGGACAAACGGTGTGCGTTAGAAAGCTGCGACCGCACCCCGTGTCAACTTAAGCATTTGCCATAGGCTCCCATCCGCCCATGTGTGGCATACCGCGCCATGTTATTC
', '
TCGAAGTGTTTGTTTCAATCGTGTATAGTTTGGCATAACCATGTAAGGGCCTTCGAACCGACCCCACTCTTTGCAGTCTGCGAGAAGCAAGAAGCCTTACTCGAAAC
', '
GGCAGCGTGAGAGTCCTTGCTTACAAAATCAAGGGTAAAACATCCATCGAGAGAACGTTGCGGCTGGAGAGATCTCATTAAGCGTTCAATTGAGGTTATAGAGGTGA
', '
TCTCCAGAGCGGAGTCCCCACTAGGAACCCGAAGTTTGGCAGGTACACACTCTCAGCACCCACCCCAACTAGGGCAGCCGATCCGCATAGGGGCACCAGATGATTCC
', '
TCCATGCCTGGTACATACATTAAGTTGATAGTATTACGGTGGGCACCAGGCAACACCCGATTGCTCTGCGGATACACACTTCTTGCGTTGTATGGGAGAGAATCTGT
', '
CCCTGGAGAATTTAATAGTCGTTTTTATACCAGTGGTACAGGATCTCAGATTGTACCCCGTGTCCCAGTTTAAGGCTAAATCACATATTTCTGCGTACTTCACTAAC
', '
TTATAGTTAAATAGCGTATGATCAAGTTCCACAGGTCGATAGTTGGACTCGTATCGAGGTATCAGCATTGACGGCGCTAGTTAGTCTATAAATATTCGTTAGGAAAC
', '
GTTTTGGTAAGGGAGTACCTATGATGTATTGTTCAAGAAGAGGGCACAGTAATGGACGAAAGGGAGGTCATGTCAATCCAGTCCACTCCATAGAGAGGTTTATACTC
', '
CCAGGATGTTCGATGATATCAGGCGCGAAGTATCGGAAGCGTCATATTAAGCACATATTTAATGGGGCATACCCCCAAAGGCATGAGACCAACCACAGCACTCGCAT
', '
CATTGTTCTTCGGGACTAGTAGAGGTCGTTGTCGTTGCGAACACGTGAGTTGCAAGTGCTCAAGCTTCACTAGATCTTCGTCTCCCATGTAAGCTGGCCCTCGAGCC
', '
CTGCTCTCACCCGCACGCCGGATGTCAGGGCTGGAAGATGGCTATAACTCTTAGCAGACCTGGATGTGTCGTTACCGATTATTTTCCGCTCCCGAAGAACCTTTTCT
', '
ACCATCAGACATACCGAAGCGCCGACTGACAGTTCATTATTTGACGGTGAACAGTACTGGACGCCAGTATTAATGCTGTATCATGTGAATCACAGGCCCGTCGGTGC
', '
AAGAGACCCCGCACGACACAAATGCGTAGCAAGCTGCTTATCCAGCAATGAGGTGTGGTGGTGAGCGGAGTTAACACGGGACGCATCGGAGAGGCGGGGCGGGGAT
', '
TCGATATTATATTCGACTTTAATCCTAATGGACGTCCCAAGGTCACCGGAGTGCGGACGACCGCCACCCCTCGTAAGGTGCCTACAGCACCCGACCCCTCTTGCCAC
', '
GCCTGTAAATTGCAAATAAAGCCATATTCTTGTGAATCGATGCCCCTCATATCACATTATCTACTTCTCGTATGCTAATTGGCAGCGATGAGAGAACCCTCATAGTA
', '
AGGACCGCCGCAAAAATATGGCGCTCACACAAATCGCTGTCATTCATGTGTATGACTTGCTTGTCCCGTTTCGTGCAGATTTTATGATACACTGAACGTCTATTAAC
', '
ATGGTAAAGGCTCAAAAAGTACCCGACTTTGGGATTAGACGCGAACCTGGGTACCAGCTATTAAGCATAGTTTTTTCTGAATGGCCACGGGCGGGTGGGGGAGTTTC
', '
CCGGACGATGGCGAGGTGAAGATAGTAAAGTAAGCTGTGGGCGTATATTTATTTAGATAGGACACGGGTAAGTGGTGACTATTGCGTTGTCGGTCGTAAATTTCGCT
', '
ATACGAGGACGTATGGCCGTGATGTCGCTCGTAGCGATTGCTTATGGAGCTTACCCTTGATCATGGTACGCTGTACTTGATATAGAGAAGACTTTTTGACCGGCGCC
', '
GTCACTAAGACAGAATAAAACCGAACGTTGATGCCGAGTGTAACGTATAGATGTTCATCTTATACCTGTGGAGCATGGACTTATAAGATTGATCTCTGGATGTACCC
', '
CACTAATGATCTTTTGGACATTGAGAGTCAGTGCCGATTACCGACTGTCCGTGTACGCGTACTGCAACGGTGGGACTAAATACGGCTACCTCAATTCGCACGCTGGC
', '
CCCTGGTTAGCCACTTGCAGACCGGGCATTATAGGACGGACTATCCGCCCAGAACAAGAGATTGTTATCCTTGCTCATACCACGGATGCACGCCCCCCCAGTCACTT
', '
AGTGCTAAACGCACAGGTGGTGAGGATTACTATAACCGCGGTCGCACACATGGCTAAGTAAGTCCCTCGATCCATAGTTTCACATCCACCCGGACGTAATCCACGAC
', '
TGATACACAAACGTACCGTAGCCTCCCTGCGAAGAGACTCATGCATCAATGAACTAACAAGTCGGCCTGCTATTCTGATATCTCATCCCGTCCTCCCTTAACAGCAT
', '
GTAACCCGCTCCCGCAGTACTGCACGCGAGTGATACAGTACCAGTGAAAGTCTATCGCATGTCAGGTGCAGCTAACACCCCGAACCGAACTACGCGTACGCTGCTCC
', '
ATTGCCTCCGTCTCCGTAAAGATCTAAGCTTTGTATGTTCGGTCAATCCCACCTCTGGCGTGCTTGGTTGATTTTTGCTCTACAGCTCTGGCCAATCGTTGTGCAAC
', '
AAGCACGCTCTAGATAATGTTTGTAAACCGCCTCTCGTCAGCATATTTTTCACAGAGCGGATCTGTGTAAGCTCCGGCTCGGTGCTGTTGAGTCGGATTTTGGCTTC
', '
GCGTCTTACGCTAGGTGCCGTAACGCGGTTTCGGCCAGGACAGCGTTGGCCCCAGTGGCCCCCACAAGTGAGAGCGCGTGGATAAAAACCAGGACATAACGCGCCTC
', '
GATCGTGTCCCAGTATGTCAAACTAACTAAGGTACGAAACGAGTGCCTGGTTGCAACCACCGCCAGATCCGATCATTGGTAAAATCCAGACAGTCCCGGCCCACTTA
', '
```

```
    GTACAGATCCCTTAAAGAATGATTGTGTACCTGATAAGGACCTTTAACTGATGCGGACACGTTTGCCCAGCCCCGTACTCCGTTTACGGACGCGACATGCCGCAATC
    ', '
    GTCTGTCGCGTCAAGACCTAACCAAAACTCTACTTTTCCCTCCTGTTATAACACTGATTGGTCAAGATTCGCCTAATCCGACGGAGGGGCAAAATCGAGCACAGTTA
    ', '
    GCACCTATAGTATCATCATTAGTCGGGCTGGGTGTTTAGCGAGTCATTAGAGACGGACGTTCGAATGTGCCCGCAGATTCGGGTCTAATATAACAGCCATTTCGGCC
    ', '
    CTAAGCGAAGACCTTTCGCCCCAAGGGGCACTATGTGACTCACAACTACCTCCTCCTACAACGAATCAGCGCCAGCGCGAAGGTTGCAAGGGTAGACGTGCGTACAT
    ', '
    GTACCCACTGTCCTCCAAATCCTGGACATCCTTGCAGACAGCAACCATCGCAAATAGTAATAGTGATCGTCTTATTACTTGGGGTAAACTCCAACGGAGTGAGTCGC
    ', '
    CTGTTTGGCAGTCCGCTGTCCGTACGCTTGCATGGTAGAGGCTAGGGCCGGATATATACCTCCGCTTACTCATATGCAGAAGTTTAACGAATGCGTTCACCTTGACC
    ', '
    ATCACGGTACCTGGGCCTAGAACCCTCCTTGACGCGAGGATCGAGCTGACGTCGGCACACGGTCGGTTAGCTTCCTCTGTAGTCAGTGCCGTGGAGTAACCGTCTTC
    ', '
    CGTTCCGCATTAGAGGTCCGAAAGGGTCCGAATAAATCCTCATCGACCCAATGACCACCGCGGTAACCGGTCACGGGTGCGGTCAGTGAAAACAACAGAGCCCTGTT
    ', '
    AGCTGGCACAAGAAGAAGTCTCTTCAATGTCAAGTGTTGACGGACTCGCTTTGCTCTAGTCGACCACAAAGCGCCATACAATAATTATGCAAGGTCTCGCCATTTGC
    ', '
    ACGCCGTGTTTTCACGTAACCAGGACGGGTAAATGCCTCACAGTTATTCTCTATCCAACTAGGATCAATTAAAGGCGCTAAGTTGTCGAGGAAGGGGCTCAATCAGC
    '], 9)


# In[15]:


#De Bruijn Graph for pool-B for k=18
construct_graph(['
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    TGGACAAACGGTGTGCGTTAGAAAGCTGCGACCGCACCCCGTGTCAACTTAAGCATTTGCCATAGGCTCCCATCCGCCCATGTGTGGCATACCGCGCCATGTTATTC
    ', '
    TCGAAGTGTTTGTTTCAATCGTGTATAGTTTGGCATAACCATGTAAGGGCCTTCGAACCGACCCCACTCTTTGCAGTCTGCGAGAAGCAAGAAGCCTTACTCGAAAC
    ', '
    GGCAGCGTGAGAGTCCTTGCTTACAAAATCAAGGGTAAAACATCCATCGAGAGAACGTTGCGGCTGGAGAGATCTCATTAAGCGTTCAATTGAGGTTATAGAGGTGA
```

```
', '
TCTCCAGAGCGGAGTCCCCACTAGGAACCCGAAGTTTGGCAGGTACACACTCTCAGCACCCACCCCAACTAGGGCAGCCGATCCGCATAGGGGCACCAGATGATTCC
', '
TCCATGCCTGGTACATACATTAAGTTGATAGTATTACGGTGGGCACCAGGCAACACCCGATTGCTCTGCGGATACACACTTCTTGCGTTGTATGGGAGAGAATCTGT
', '
CCCTGGAGAATTTAATAGTCGTTTTTATACCAGTGGTACAGGATCTCAGATTGTACCCCGTGTCCCAGTTTAAGGCTAAATCACATATTTCTGCGTACTTCACTAAC
', '
TTATAGTTAAATAGCGTATGATCAAGTTCCACAGGTCGATAGTTGGACTCGTATCGAGGTATCAGCATTGACGGCGCTAGTTAGTCTATAAATATTCGTTAGGAAAC
', '
GTTTTGGTAAGGGAGTACCTATGATGTATTGTTCAAGAAGAGGGCACAGTAATGGACGAAAGGGAGGTCATGTCAATCCAGTCCACTCCATAGAGAGGTTTATACTC
', '
CCAGGATGTTCGATGATATCAGGCGCGAAGTATCGGAAGCGTCATATTAAGCACATATTTAATGGGGCATACCCCCAAAGGCATGAGACCAACCACAGCACTCGCAT
', '
CATTGTTCTTCGGGACTAGTAGAGGTCGTTGTCGTTGCGAACACGTGAGTTGCAAGTGCTCAAGCTTCACTAGATCTTCGTCTCCCATGTAAGCTGGCCCTCGAGC
', '
CTGCTCTCACCCGCACGCCGGATGTCAGGGCTGGAAGATGGCTATAACTCTTAGCAGACCTGGATGTGTCGTTACCGATTATTTTCCGCTCCCGAAGAACCTTTTCT
', '
ACCATCAGACATACCGAAGCGCCGACTGACAGTTCATTATTTGACGGTGAACAGTACTGGACGCCAGTATTAATGCTGTATCATGTGAATCACAGGCCCGTCGGTGC
', '
AAGAGACCCCGCACGACACAAATGCGTAGCAAGCTGCTTATCCAGCAATGAGGTGTGGTGGTGAGCGGAGTTAACACGGGACGCATCGGAGAGGCGGGGCGGGGAT
', '
TCGATATTATATTCGACTTTAATCCTAATGGACGTCCCAAGGTCACCGGAGTGCGGACGACCGCCACCCCTCGTAAGGTGCCTACAGCACCCGACCCCTCTTGCCAC
', '
GCCTGTAAATTGCAAATAAAGCCATATTCTTGTGAATCGATGCCCCTCATATCACATTATCTACTTCTCGTATGCTAATTGGCAGCGATGAGAGAACCCTCATAGTA
', '
AGGACCGCCGCAAAAATATGGCGCTCACACAAATCGCTGTCATTCATGTGTATGACTTGCTTGTCCCGTTTCGTGCAGATTTTATGATACACTGAACGTCTATTAAC
', '
ATGGTAAAGGCTCAAAAAGTACCCGACTTTGGGATTAGACGCGAACCTGGGTACCAGCTATTAAGCATAGTTTTTTCTGAATGGCCACGGGCGGGTGGGGGAGTTTC
', '
CCGGACGATGGCGAGGTGAAGATAGTAAAGTAAGCTGTGGGCGTATATTTATTTAGATAGGACACGGGTAAGTGGTGACTATTGCGTTGTCGGTCGTAAATTTCGCT
', '
ATACGAGGACGTATGGCCGTGATGTCGCTCGTAGCGATTGCTTATGGAGCTTACCCTTGATCATGGTACGCTGTACTTGATATAGAGAAGACTTTTTGACCGGCGCC
', '
GTCACTAAGACAGAATAAAACCGAACGTTGATGCCGAGTGTAACGTATAGATGTTCATCTTATACCTGTGGAGCATGGACTTATAAGATTGATCTCTGGATGTACCC
', '
CACTAATGATCTTTTGGACATTGAGAGTCAGTGCCGATTACCGACTGTCCGTGTACGCGTACTGCAACGGTGGGACTAAATACGGCTACCTCAATTCGCACGCTGGC
', '
CCCTGGTTAGCCACTTGCAGACCGGGCATTATAGGACGGACTATCCGCCCAGAACAAGAGATTGTTATCCTTGCTCATACCACGGATGCACGCCCCCCCAGTCACTT
', '
AGTGCTAAACGCACAGGTGGTGAGGATTACTATAACCGCGGTCGCACACATGGCTAAGTAAGTCCCTCGATCCATAGTTTCACATCCACCCGGACGTAATCCACGAC
', '
TGATACACAAACGTACCGTAGCCTCCCTGCGAAGAGACTCATGCATCAATGAACTAACAAGTCGGCCTGCTATTCTGATATCTCATCCCGTCCTCCCTTAACAGCAT
', '
GTAACCCGCTCCCGCAGTACTGCACGCGAGTGATACAGTACCAGTGAAAGTCTATCGCATGTCAGGTGCAGCTAACACCCCGAACCGAACTACGCGTACGCTGCTCC
', '
ATTGCCTCCGTCTCCGTAAAGATCTAAGCTTTGTATGTTCGGTCAATCCCACCTCTGGCGTGCTTGGTTGATTTTTGCTCTACAGCTCTGGCCAATCGTTGTGCAAC
', '
AAGCACGCTCTAGATAATGTTTGTAAACCGCCTCTCGTCAGCATATTTTTCACAGAGCGGATCTGTGTAAGCTCCGGCTCGGTGCTGTTGAGTCGGATTTTGGCTTC
', '
GCGTCTTACGCTAGGTGCCGTAACGCGGTTTCGGCCAGGACAGCGTTGGCCCCAGTGGCCCCCACAAGTGAGAGCGCGTGGATAAAAACCAGGACATAACGCGCCTC
', '
GATCGTGTCCCAGTATGTCAAACTAACTAAGGTACGAAACGAGTGCCTGGTTGCAACCACCGCCAGATCCGATCATTGGTAAAATCCAGACAGTCCCGGCCCACTTA
', '
GTACAGATCCCTTAAAGAATGATTGTGTACCTGATAAGGACCTTTAACTGATGCGGACACGTTTGCCCAGCCCCGTACTCCGTTTACGGACGCGACATGCCGCAATC
', '
GTCTGTCGCGTCAAGACCTAACCAAAACTCTACTTTTCCCTCCTGTTATAACACTGATTGGTCAAGATTCGCCTAATCCGACGGAGGGGCAAAATCGAGCACAGTTA
', '
GCACCTATAGTATCATCATTAGTCGGGCTGGGTGTTTAGCGAGTCATTAGAGACGGACGTTCGAATGTGCCCGCAGATTCGGGTCTAATATAACAGCCATTTCGGCC
', '
CTAAGCGAAGACCTTTCGCCCCAAGGGGCACTATGTGACTCACAACTACCTCCTCCTACAACGAATCAGCGCCAGCGCGAAGGTTGCAAGGGTAGACGTGCGTACAT
', '
GTACCCACTGTCCTCCAAATCCTGGACATCCTTGCAGACAGCAACCATCGCAAATAGTAATAGTGATCGTCTTATTACTTGGGGTAAACTCCAACGGAGTGAGTCGC
```

```
    ', '
    CTGTTTGGCAGTCCGCTGTCCGTACGCTTGCATGGTAGAGGCTAGGGCCGGATATATACCTCCGCTTACTCATATGCAGAAGTTTAACGAATGCGTTCACCTTGACC
    ', '
    ATCACGGTACCTGGGCCTAGAACCCTCCTTGACGCGAGGATCGAGCTGACGTCGGCACACGGTCGGTTAGCTTCCTCTGTAGTCAGTGCCGTGGAGTAACCGTCTTC
    ', '
    CGTTCCGCATTAGAGGTCCGAAAGGGTCCGAATAAATCCTCATCGACCCAATGACCACCGCGGTAACCGGTCACGGGTGCGGTCAGTGAAAACAACAGAGCCCTGTT
    ', '
    AGCTGGCACAAGAAGAAGTCTCTTCAATGTCAAGTGTTGACGGACTCGCTTTGCTCTAGTCGACCACAAAGCGCCATACAATAATTATGCAAGGTCTCGCCATTTGC
    ', '
    ACGCCGTGTTTTCACGTAACCAGGACGGGTAAATGCCTCACAGTTATTCTCTATCCAACTAGGATCAATTAAAGGCGCTAAGTTGTCGAGGAAGGGGCTCAATCAGC
    '], 18)


# In[16]:


#De Bruijn Graph for pool-C for k=5
construct_graph(['
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
```

```
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
', '
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
', '
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
', '
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
', '
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
', '
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
', '
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
', '
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
', '
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
', '
GGCTACATGGACAAACGGTGTGCGTTAGAAAGCTGCGACCGCACCCCGTGTCAACTTAAGCATTTGCCATAGGCTCCCATCCGCCCATGTGTGGCATACCGCGCCAT
', '
AGCCATACTTACCATCGAAGTGTTTGTTTCAATCGTGTATAGTTTGGCATAACCATGTAAGGGCCTTCGAACCGACCCCACTCTTTGCAGTCTGCGAGAAGCAAGAA
', '
CTCACCAACGCACGGCTTATTGGCAGCGTGAGAGTCCTTGCTTACAAAATCAAGGGTAAAACATCCATCGAGAGAACGTTGCGGCTGGAGAGATCTCATTAAGCGTT
', '
GATTTCGGAGGCATTCCCTACACTTTCCTCTCCAGAGCGGAGTCCCCACTAGGAACCCGAAGTTTGGCAGGTACACACTCTCAGCACCCACCCCAACTAGGGCAGCC
', '
GTAGGCTACGCGTCTTTATGGTCTAATTGCTGACATCCATGCCTGGTACATACATTAAGTTGATAGTATTACGGTGGGCACCAGGCAACACCCGATTGCTCTGCGGA
', '
TTTGACCTCGAGAACCGCACACAGGTTGGTGGATTCTGGCGCCCCTGGAGAATTTAATAGTCGTTTTTATACCAGTGGTACAGGATCTCAGATTGTACCCCGTGTCC
', '
CGGAGGTCCAGAAGATCGTTATAGCATGCTCCAGCGAGATGGAAGACGTTTATAGTTAAATAGCGTATGATCAAGTTCCACAGGTCGATAGTTGGACTCGTATCGAC
', '
ATTCATAGTGGAGAGTATGGCCCGTCGGCATTCCTCGGCGGGTTAACCATTACCTCGTTTTGGTAAGGGAGTACCTATGATGTATTGTTCAAGAAGAGGGCACAGTA
', '
GTCGCCTAGCAAATCCTATTAGGGGCACCTAGCGGGCACCGAGATTCGAGGTGAAGGGTGGCCCCAGGATGTTCGATGATATCAGGCGCGAAGTATCGGAAGCGTCA
', '
ACGGCGCTTTCTAGGATAAAGTACGATGTTCTGTAAGATGTTATGTCTTAACAAGACCCGAAGGCAGAGTCATTGTTCTTCGGGACTAGTAGAGGTCGTTGTCGTTC
', '
GAGTTGTACGCCTCTATGCCCGCGCATGCAGAAATGCCACCCCCACCACTCGGCGGTCGATCTGCTGCTATTGTGGACTGCTCTCACCCGCACGCCGGATGTCAGGC
', '
CAGCGGTTAAGGGTTCGTACTATAGTCGTCTTATGAAGTTTGTACGGATTCCGAGCCTCTCCCCGATCATAAAGGAATGATTTTACCATCAGACATACCGAAGCGCC
', '
CCGATACAGAGAGCTGAATTGCTAATCCCCTCCTTCAGTATTGGTCGTGTATCAACCAAGCCCGCCGGAGGAGCACTCGCAACAACACTTGAAGAGACCCCGCACGA
', '
CATTTCTAGTGTGGAGAGCTGTGGCTAGTGACTGATCAGACATCAGCCGTTTCGAAAGAGGACCTAATGACAGATTAACTAAACAAGTACGCCCTAAGTCGATATTA
', '
TTTAGAACAAGGCGTTAGGATACCTATCAGTGCTTCAAGTAATTCCATTGACCATTAGGTTATGTTGTTACGCGGATTATGTAAGCCGCTCAGTACGCTCATAAGGC
', '
TCGCAGCGCGTCGCATGGGTGCATCGATCTGGTAAGCCGCTTTCCAAAGGACCACGGCCAAGACTCCCCTTTTGACATCGGCATGTGCTGCTCATGTCTTCAATTGT
', '
AGGCTTGGAACGTGGAGTCTCTCTATCGGTCATGCGGTGGCTCTCCGCTCAAGAACGCAGTCGGGGGAATGCGCTGCGTCCGACGACTGTTCAGCGGTTCCGGGTTC
', '
CCAGCTTAACAAACCTCAGGTAGAATTACGGAGTTGATCAACTCCACTACTGTACAGCCCATAAGGGGTGAAGAGTGGCTCATGATTGCGGATCGTTAGCTAACTCC
', '
TCGCGATTGGGGTCCGGTGGCCAGGCGACTAAGAAAGGACACAGAAAAATTGCGCTCTACAGTATCTCAGCCGCCCACTAATATTATTAGACGAGCTTCATTGTCTT
', '
CGCGATTCGCCATATGGTGATATCGCTTGGGCATTGACTCGATAACAATAGGTTGATGCATTCACGCGCGGTCTATTGATCTTCATCCCGATAGTTATGTATGGACC
', '
TGGATGTACCGGTTGCTTTGACAAAAGCTTGGCTTTATGGCTTCGCCGTAAATACATACCAGCGGACTAGTTGCCAAGGCTGATAGGGCGGTCATCCGCGTTGCTGC
', '
CTCAATTCGCACGCTGGGAACATCAAGAAGCCTGTCGTGTGCCGAGAGTGATCTATGTCTACGTAAGAAGGTCACTATCCCTGGTGTGTCAGGCAATTGATAGCGAT
', '
```

```
    CGGATGCACGCCCCCCCAGTCACTTTACATCTTACGCCGAACTGGACGTGAAGGGATGAATAAGGAACGGCCGGTCACCTAGTGCATAAGTTATAGACGATTTGTTC
    ', '
    AGTTTCACATCCACCCGGACGTAATCCACGAGAAAAAGGGCAATTCTTATCACGAAATTCAGCGGAAGCCGTCGAGGTATCAGGTGTAGCCATTGTAGCGTCTAACC
    ', '
    GCTATTCTGATATCTCATCCCGTCCTCCCTTAACAGCATAGGGGCTGACTAGACACCGCGTGTGGCCGACTTCATCGCGATCGGAATAGATATAGTTAATAAACCAC
    ', '
    TCAGGTGCAGCTAACACCCCGAACCGAACTACGCGTACGCTGCTCGCCGCAATTCATTCGTTAAAAAGCACGGGCAGACGGTTAAAGGCTGTTGTATAACCTAGTTC
    ', '
    CTGGCGTGCTTGGTTGATTTTTGCTCTACAGCTCTGGCCAATCGTTGTGCAAGTAGATGACGACTATTGCCTATGGTATGTTATATGCCATTCGGATCACCGGTCTC
    ', '
    TTTCACAGAGCGGATCTGTGTAAGCTCCGGCTCGGTGCTGTTGAGTCGGATTTTGGCTTCACATTTAACTTTCGTCCTCTAAAGTCACTACAAAACATGCTGGTTCC
    ', '
    CAGCGTTGGCCCCAGTGGCCCCCACAAGTGAGAGCGCGTGGATAAAAACCAGGACATAACGCGCCTGGCCGTTCCAGGGTCCAAACTGCTGTTCCCCGTCACGGCCA
    ', '
    ACGAAACGAGTGCCTGGTTGCAACCACCGCCAGATCCGATCATTGGTAAAATCCAGACAGTCCCGGCCCACTTATACTGGCAAGAATCTACTCCAAAACGCCACGGC
    ', '
    GTACCTGATAAGGACCTTTAACTGATGCGGACACGTTTGCCCAGCCCCGTACTCCGTTTACGGACGCGACATGCCGCAATCCACGGGTGCTGAGAGCCTTTCGGACC
    ', '
    AACCAAAACTCTACTTTTCCCTCCTGTTATAACACTGATTGGTCAAGATTCGCCTAATCCGACGGAGGGGCAAAATCGAGCACAGTTAATATAGAAAGGTCTTGTCC
    ', '
    TCATCATTAGTCGGGCTGGGTGTTTAGCGAGTCATTAGAGACGGACGTTCGAATGTGCCCGCAGATTCGGGTCTAATATAACAGCCATTTCGGCGATGCCACAGTGT
    ', '
    CGAAGACCTTTCGCCCCAAGGGGCACTATGTGACTCACAACTACCTCCTCCTACAACGAATCAGCGCCAGCGCGAAGGTTGCAAGGGTAGACGTGCGTACATACAGC
    ', '
    GCGTACCCACTGTCCTCCAAATCCTGGACATCCTTGCAGACAGCAACCATCGCAAATAGTAATAGTGATCGTCTTATTACTTGGGGTAAACTCCAACGGAGTGAGTC
    ', '
    CGGATTCTTCTGTTTGGCAGTCCGCTGTCCGTACGCTTGCATGGTAGAGGCTAGGGCCGGATATATACCTCCGCTTACTCATATGCAGAAGTTTAACGAATGCGTTC
    ', '
    AGCCTACATCGATTGCATCACGGTACCTGGGCCTAGAACCCTCCTTGACGCGAGGATCGAGCTGACGTCGGCACACGGTCGGTTAGCTTCCTCTGTAGTCAGTGCCC
    ', '
    TACCTTCTCGACCCAAGCTCCATCGTTCCGCATTAGAGGTCCGAAAGGGTCCGAATAAATCCTCATCGACCCAATGACCACCGCGGTAACCGGTCACGGGTGCGGTC
    ', '
    ATACCCACCCTTATTCCAAAAACCAGTCTCAGCTGGCACAAGAAGAAGTCTCTTCAATGTCAAGTGTTGACGGACTCGCTTTGCTCTAGTCGACCACAAAGCGCCAT
    '],5)


# In[17]:


#De Bruijn Graph for pool-C for k=9
construct_graph(['
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
```

```
', '
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
', '
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
', '
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
', '
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
', '
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
', '
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
', '
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
', '
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
', '
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
', '
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
', '
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
', '
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
', '
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
', '
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
', '
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
', '
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
', '
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
', '
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
', '
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
', '
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
', '
GGCTACATGGACAAACGGTGTGCGTTAGAAAGCTGCGACCGCACCCCGTGTCAACTTAAGCATTTGCCATAGGCTCCCATCCGCCCATGTGTGGCATACCGCGCCAT
', '
AGCCATACTTACCATCGAAGTGTTTGTTTCAATCGTGTATAGTTTGGCATAACCATGTAAGGGCCTTCGAACCGACCCCACTCTTTGCAGTCTGCGAGAAGCAAGAA
', '
CTCACCAACGCACGGCTTATTGGCAGCGTGAGAGTCCTTGCTTACAAAATCAAGGGTAAAACATCCATCGAGAGAACGTTGCGGCTGGAGAGATCTCATTAAGCGTT
', '
GATTTCGGAGGCATTCCCTACACTTTCCTCTCCAGAGCGGAGTCCCCACTAGGAACCCGAAGTTTGGCAGGTACACACTCTCAGCACCCACCCCAACTAGGGCAGCC
', '
GTAGGCTACGCGTCTTTATGGTCTAATTGCTGACATCCATGCCTGGTACATACATTAAGTTGATAGTATTACGGTGGGCACCAGGCAACACCCGATTGCTCTGCGGA
', '
TTTGACCTCGAGAACCGCACACAGGTTGGTGGATTCTGGCGCCCCTGGAGAATTTAATAGTCGTTTTTATACCAGTGGTACAGGATCTCAGATTGTACCCCGTGTCC
', '
CGGAGGTCCAGAAGATCGTTATAGCATGCTCCAGCGAGATGGAAGACGTTTATAGTTAAATAGCGTATGATCAAGTTCCACAGGTCGATAGTTGGACTCGTATCGAC
', '
ATTCATAGTGGAGAGTATGGCCCGTCGGCATTCCTCGGCGGGTTAACCATTACCTCGTTTTGGTAAGGGAGTACCTATGATGTATTGTTCAAGAAGAGGGCACAGTA
', '
GTCGCCTAGCAAATCCTATTAGGGGCACCTAGCGGGCACCGAGATTCGAGGTGAAGGGTGGCCCCAGGATGTTCGATGATATCAGGCGCGAAGTATCGGAAGCGTCA
', '
ACGGCGCTTTCTAGGATAAAGTACGATGTTCTGTAAGATGTTATGTCTTAACAAGACCCGAAGGCAGAGTCATTGTTCTTCGGGACTAGTAGAGGTCGTTGTCGTTC
', '
GAGTTGTACGCCTCTATGCCCGCGCATGCAGAAATGCCACCCCCACCACTCGGCGGTCGATCTGCTGCTATTGTGGACTGCTCTCACCCGCACGCCGGATGTCAGGC
```

```
    ', '
    CAGCGGTTAAGGGTTCGTACTATAGTCGTCTTATGAAGTTTGTACGGATTCCGAGCCTCTCCCCGATCATAAAGGAATGATTTTACCATCAGACATACCGAAGCGC
    ', '
    CCGATACAGAGAGCTGAATTGCTAATCCCCTCCTTCAGTATTGGTCGTGTATCAACCAAGCCCGCCGGAGGAGCACTCGCAACAACACTTGAAGAGACCCCGCACGA
    ', '
    CATTTCTAGTGTGGAGAGCTGTGGCTAGTGACTGATCAGACATCAGCCGTTTCGAAAGAGGACCTAATGACAGATTAACTAAACAAGTACGCCCTAAGTCGATATTA
    ', '
    TTTAGAACAAGGCGTTAGGATACCTATCAGTGCTTCAAGTAATTCCATTGACCATTAGGTTATGTTGTTACGCGGATTATGTAAGCCGCTCAGTACGCTCATAAGGC
    ', '
    TCGCAGCGCGTCGCATGGGTGCATCGATCTGGTAAGCCGCTTTCCAAAGGACCACGGCCAAGACTCCCCTTTTGACATCGGCATGTGCTGCTCATGTCTTCAATTGT
    ', '
    AGGCTTGGAACGTGGAGTCTCTCTATCGGTCATGCGGTGGCTCTCCGCTCAAGAACGCAGTCGGGGGAATGCGCTGCGTCCGACGACTGTTCAGCGGTTCCGGGTTC
    ', '
    CCAGCTTAACAAACCTCAGGTAGAATTACGGAGTTGATCAACTCCACTACTGTACAGCCCATAAGGGGTGAAGAGTGGCTCATGATTGCGGATCGTTAGCTAACTCC
    ', '
    TCGCGATTGGGGTCCGGTGGCCAGGCGACTAAGAAAGGACACAGAAAAATTGCGCTCTACAGTATCTCAGCCGCCCACTAATATTATTAGACGAGCTTCATTGTCTT
    ', '
    CGCGATTCGCCATATGGTGATATCGCTTGGGCATTGACTCGATAACAATAGGTTGATGCATTCACGCGCGGTCTATTGATCTTCATCCCGATAGTTATGTATGGACC
    ', '
    TGGATGTACCGGTTGCTTTGACAAAAGCTTGGCTTTATGGCTTCGCCGTAAATACATACCAGCGGACTAGTTGCCAAGGCTGATAGGGCGGTCATCCGCGTTGCTGC
    ', '
    CTCAATTCGCACGCTGGGAACATCAAGAAGCCTGTCGTGTGCCGAGAGTGATCTATGTCTACGTAAGAAGGTCACTATCCCTGGTGTGTCAGGCAATTGATAGCGAT
    ', '
    CGGATGCACGCCCCCCCAGTCACTTTACATCTTACGCCGAACTGGACGTGAAGGGATGAATAAGGAACGGCCGGTCACCTAGTGCATAAGTTATAGACGATTTGTTC
    ', '
    AGTTTCACATCCACCCGGACGTAATCCACGAGAAAAAGGGCAATTCTTATCACGAAATTCAGCGGAAGCCGTCGAGGTATCAGGTGTAGCCATTGTAGCGTCTAACC
    ', '
    GCTATTCTGATATCTCATCCCGTCCTCCCTTAACAGCATAGGGGCTGACTAGACACCGCGTGTGGCCGACTTCATCGCGATCGGAATAGATATAGTTAATAAACCAC
    ', '
    TCAGGTGCAGCTAACACCCCGAACCGAACTACGCGTACGCTGCTCGCCGCAATTCATTCGTTAAAAAGCACGGGCAGACGGTTAAAGGCTGTTGTATAACCTAGTTC
    ', '
    CTGGCGTGCTTGGTTGATTTTTGCTCTACAGCTCTGGCCAATCGTTGTGCAAGTAGATGACGACTATTGCCTATGGTATGTTATATGCCATTCGGATCACCGGTCTC
    ', '
    TTTCACAGAGCGGATCTGTGTAAGCTCCGGCTCGGTGCTGTTGAGTCGGATTTTGGCTTCACATTTAACTTTCGTCCTCTAAAGTCACTACAAAACATGCTGGTTCC
    ', '
    CAGCGTTGGCCCCAGTGGCCCCCACAAGTGAGAGCGCGTGGATAAAAACCAGGACATAACGCGCCTGGCCGTTCCAGGGTCCAAACTGCTGTTCCCCGTCACGGCCA
    ', '
    ACGAAACGAGTGCCTGGTTGCAACCACCGCCAGATCCGATCATTGGTAAAATCCAGACAGTCCCGGCCCACTTATACTGGCAAGAATCTACTCCAAAACGCCACGGC
    ', '
    GTACCTGATAAGGACCTTTAACTGATGCGGACACGTTTGCCCAGCCCCGTACTCCGTTTACGGACGCGACATGCCGCAATCCACGGGTGCTGAGAGCCTTTCGGACC
    ', '
    AACCAAAACTCTACTTTTCCCTCCTGTTATAACACTGATTGGTCAAGATTCGCCTAATCCGACGGAGGGGCAAAATCGAGCACAGTTAATATAGAAAGGTCTTGTCC
    ', '
    TCATCATTAGTCGGGCTGGGTGTTTAGCGAGTCATTAGAGACGGACGTTCGAATGTGCCCGCAGATTCGGGTCTAATATAACAGCCATTTCGGCGATGCCACAGTGT
    ', '
    CGAAGACCTTTCGCCCCAAGGGGCACTATGTGACTCACAACTACCTCCTCCTACAACGAATCAGCGCCAGCGCGAAGGTTGCAAGGGTAGACGTGCGTACATACAGC
    ', '
    GCGTACCCACTGTCCTCCAAATCCTGGACATCCTTGCAGACAGCAACCATCGCAAATAGTAATAGTGATCGTCTTATTACTTGGGGTAAACTCCAACGGAGTGAGTC
    ', '
    CGGATTCTTCTGTTTGGCAGTCCGCTGTCCGTACGCTTGCATGGTAGAGGCTAGGGCCGGATATATACCTCCGCTTACTCATATGCAGAAGTTTAACGAATGCGTTC
    ', '
    AGCCTACATCGATTGCATCACGGTACCTGGGCCTAGAACCCTCCTTGACGCGAGGATCGAGCTGACGTCGGCACACGGTCGGTTAGCTTCCTCTGTAGTCAGTGCCC
    ', '
    TACCTTCTCGACCCAAGCTCCATCGTTCCGCATTAGAGGTCCGAAAGGGTCCGAATAAATCCTCATCGACCCAATGACCACCGCGGTAACCGGTCACGGGTGCGGTC
    ', '
    ATACCCACCCTTATTCCAAAAACCAGTCTCAGCTGGCACAAGAAGAAGTCTCTTCAATGTCAAGTGTTGACGGACTCGCTTTGCTCTAGTCGACCACAAAGCGCCAT
    '],9)
133
134
135 # In[18]:
136
137
```

```
#De Bruijn Graph for pool-C for k=18
construct_graph(['
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
    ', '
```

```
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCATAC
', '
GGCTACATGGACAAACGGTGTGCGTTAGAAAGCTGCGACCGCACCCCGTGTCAACTTAAGCATTTGCCATAGGCTCCCATCCGCCCATGTGTGGCATACCGCGCCAT
', '
AGCCATACTTACCATCGAAGTGTTTGTTTCAATCGTGTATAGTTTGGCATAACCATGTAAGGGCCTTCGAACCGACCCCACTCTTTGCAGTCTGCGAGAAGCAAGAA
', '
CTCACCAACGCACGGCTTATTGGCAGCGTGAGAGTCCTTGCTTACAAAATCAAGGGTAAAACATCCATCGAGAGAACGTTGCGGCTGGAGAGATCTCATTAAGCGTT
', '
GATTTCGGAGGCATTCCCTACACTTTCCTCTCCAGAGCGGAGTCCCCACTAGGAACCCGAAGTTTGGCAGGTACACACTCTCAGCACCCACCCCAACTAGGGCAGCG
', '
GTAGGCTACGCGTCTTTATGGTCTAATTGCTGACATCCATGCCTGGTACATACATTAAGTTGATAGTATTACGGTGGGCACCAGGCAACACCCGATTGCTCTGCGGA
', '
TTTGACCTCGAGAACCGCACACAGGTTGGTGGATTCTGGCGCCCCTGGAGAATTTAATAGTCGTTTTTATACCAGTGGTACAGGATCTCAGATTGTACCCCGTGTCC
', '
CGGAGGTCCAGAAGATCGTTATAGCATGCTCCAGCGAGATGGAAGACGTTTATAGTTAAATAGCGTATGATCAAGTTCCACAGGTCGATAGTTGGACTCGTATCGAC
', '
ATTCATAGTGGAGAGTATGGCCCGTCGGCATTCCTCGGCGGGTTAACCATTACCTCGTTTTGGTAAGGGAGTACCTATGATGTATTGTTCAAGAAGAGGGCACAGTA
', '
GTCGCCTAGCAAATCCTATTAGGGGCACCTAGCGGGCACCGAGATTCGAGGTGAAGGGTGGCCCCAGGATGTTCGATGATATCAGGCGCGAAGTATCGGAAGCGTCA
', '
ACGGCGCTTTCTAGGATAAAGTACGATGTTCTGTAAGATGTTATGTCTTAACAAGACCCGAAGGCAGAGTCATTGTTCTTCGGGACTAGTAGAGGTCGTTGTCGTTC
', '
GAGTTGTACGCCTCTATGCCCGCGCATGCAGAAATGCCACCCCCACCACTCGGCGGTCGATCTGCTGCTATTGTGGACTGCTCTCACCCGCACGCCGGATGTCAGGC
', '
CAGCGGTTAAGGGTTCGTACTATAGTCGTCTTATGAAGTTTGTACGGATTCCGAGCCTCTCCCCGATCATAAAGGAATGATTTTACCATCAGACATACCGAAGCGCC
', '
CCGATACAGAGAGCTGAATTGCTAATCCCCTCCTTCAGTATTGGTCGTGTATCAACCAAGCCCGCCGGAGGAGCACTCGCAACAACACTTGAAGAGACCCCGCACGA
', '
CATTTCTAGTGTGGAGAGCTGTGGCTAGTGACTGATCAGACATCAGCCGTTTCGAAAGAGGACCTAATGACAGATTAACTAAACAAGTACGCCCTAAGTCGATATTA
', '
TTTAGAACAAGGCGTTAGGATACCTATCAGTGCTTCAAGTAATTCCATTGACCATTAGGTTATGTTGTTACGCGGATTATGTAAGCCGCTCAGTACGCTCATAAGGC
', '
TCGCAGCGCGTCGCATGGGTGCATCGATCTGGTAAGCCGCTTTCCAAAGGACCACGGCCAAGACTCCCCTTTTGACATCGGCATGTGCTGCTCATGTCTTCAATTGT
', '
AGGCTTGGAACGTGGAGTCTCTCTATCGGTCATGCGGTGGCTCTCCGCTCAAGAACGCAGTCGGGGGAATGCGCTGCGTCCGACGACTGTTCAGCGGTTCCGGGTTC
', '
CCAGCTTAACAAACCTCAGGTAGAATTACGGAGTTGATCAACTCCACTACTGTACAGCCCATAAGGGGTGAAGAGTGGCTCATGATTGCGGATCGTTAGCTAACTCC
', '
TCGCGATTGGGGTCCGGTGGCCAGGCGACTAAGAAAGGACACAGAAAAATTGCGCTCTACAGTATCTCAGCCGCCCACTAATATTATTAGACGAGCTTCATTGTCTT
', '
CGCGATTCGCCATATGGTGATATCGCTTGGGCATTGACTCGATAACAATAGGTTGATGCATTCACGCGCGGTCTATTGATCTTCATCCCGATAGTTATGTATGGACC
', '
TGGATGTACCGGTTGCTTTGACAAAAGCTTGGCTTTATGGCTTCGCCGTAAATACATACCAGCGGACTAGTTGCCAAGGCTGATAGGGCGGTCATCCGCGTTGCTGC
', '
CTCAATTCGCACGCTGGGAACATCAAGAAGCCTGTCGTGTGCCGAGAGTGATCTATGTCTACGTAAGAAGGTCACTATCCCTGGTGTGTCAGGCAATTGATAGCGAT
', '
CGGATGCACGCCCCCCCAGTCACTTTACATCTTACGCCGAACTGGACGTGAAGGGATGAATAAGGAACGGCCGGTCACCTAGTGCATAAGTTATAGACGATTTGTTC
', '
AGTTTCACATCCACCCGGACGTAATCCACGAGAAAAAGGGCAATTCTTATCACGAAATTCAGCGGAAGCCGTCGAGGTATCAGGTGTAGCCATTGTAGCGTCTAACC
', '
GCTATTCTGATATCTCATCCCGTCCTCCCTTAACAGCATAGGGGCTGACTAGACACCGCGTGTGGCCGACTTCATCGCGATCGGAATAGATATAGTTAATAAACCAC
', '
TCAGGTGCAGCTAACACCCCGAACCGAACTACGCGTACGCTGCTCGCCGCAATTCATTCGTTAAAAAGCACGGGCAGACGGTTAAAGGCTGTTGTATAACCTAGTTC
', '
CTGGCGTGCTTGGTTGATTTTTGCTCTACAGCTCTGGCCAATCGTTGTGCAAGTAGATGACGACTATTGCCTATGGTATGTTATATGCCATTCGGATCACCGGTCTC
', '
TTTCACAGAGCGGATCTGTGTAAGCTCCGGCTCGGTGCTGTTGAGTCGGATTTTGGCTTCACATTTAACTTTCGTCCTCTAAAGTCACTACAAAACATGCTGGTTCC
', '
CAGCGTTGGCCCCAGTGGCCCCCACAAGTGAGAGCGCGTGGATAAAAACCAGGACATAACGCGCCTGGCCGTTCCAGGGTCCAAACTGCTGTTCCCCGTCACGGCCA
', '
ACGAAACGAGTGCCTGGTTGCAACCACCGCCAGATCCGATCATTGGTAAAATCCAGACAGTCCCGGCCCACTTATACTGGCAAGAATCTACTCCAAAACGCCACGGC
', '
```

```
    GTACCTGATAAGGACCTTTAACTGATGCGGACACGTTTGCCCAGCCCCGTACTCCGTTTACGGACGCGACATGCCGCAATCCACGGGTGCTGAGAGCCTTTCGGACC
    ', '
    AACCAAAACTCTACTTTTCCCTCCTGTTATAACACTGATTGGTCAAGATTCGCCTAATCCGACGGAGGGGCAAAATCGAGCACAGTTAATATAGAAAGGTCTTGTCC
    ', '
    TCATCATTAGTCGGGCTGGGTGTTTAGCGAGTCATTAGAGACGGACGTTCGAATGTGCCCGCAGATTCGGGTCTAATATAACAGCCATTTCGGCGATGCCACAGTGT
    ', '
    CGAAGACCTTTCGCCCCAAGGGGCACTATGTGACTCACAACTACCTCCTCCTACAACGAATCAGCGCCAGCGCGAAGGTTGCAAGGGTAGACGTGCGTACATACAGC
    ', '
    GCGTACCCACTGTCCTCCAAATCCTGGACATCCTTGCAGACAGCAACCATCGCAAATAGTAATAGTGATCGTCTTATTACTTGGGGTAAACTCCAACGGAGTGAGTC
    ', '
    CGGATTCTTCTGTTTGGCAGTCCGCTGTCCGTACGCTTGCATGGTAGAGGCTAGGGCCGGATATATACCTCCGCTTACTCATATGCAGAAGTTTAACGAATGCGTTC
    ', '
    AGCCTACATCGATTGCATCACGGTACCTGGGCCTAGAACCCTCCTTGACGCGAGGATCGAGCTGACGTCGGCACACGGTCGGTTAGCTTCCTCTGTAGTCAGTGCCC
    ', '
    TACCTTCTCGACCCAAGCTCCATCGTTCCGCATTAGAGGTCCGAAAGGGTCCGAATAAATCCTCATCGACCCAATGACCACCGCGGTAACCGGTCACGGGTGCGGTC
    ', '
    ATACCCACCCTTATTCCAAAAACCAGTCTCAGCTGGCACAAGAAGAAGTCTCTTCAATGTCAAGTGTTGACGGACTCGCTTTGCTCTAGTCGACCACAAAGCGCCAT
    '],18)
140
141
142 # In[ ]:
```