

## 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)

- Write python code to generate a synthetic genome (a string) consisting of random nucleotide sequence of length 10,000.
- Randomly generate a pool of synthetic "reads" (Pool A) of length 250 such that no reads overlap on the genome from 1.
- 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.
- 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

```
1 #!/usr/bin/env python
2 # coding: utf-8
3
4 # In[2]:
5
6
7 #import random module
8 from random import choice
9
10
11 # In[3]:
12
13
14 #using a list comprehension to generate a sequence
15 length=10000
16 bases=['A','C','G','T']
17 sequence=[choice(bases) for i in range(length)]
18
19
20 # In[4]:
21
22
23 #convert the list to a string
24 sequence=''.join(sequence)
25
26
27 # In[5]:
28
29
30 print(sequence)
31
```

```

32
33 # In[6]:
34
35
36 #generating pool A synthetic reads
37 substring=[]
38 for i in range(40):
39     substring.append(''.join(sequence[i * 250: (i + 1) * 250]))
40 print(substring)
41
42
43 # In[9]:
44
45
46 #generating pool C synthetic reads
47 substring=[]
48 g=0
49 for i in range(0,len(sequence),250):
50     if i==0:
51         for k in range(14):
52             substring.append(''.join(sequence[0:250]))
53     substring.append(''.join(sequence[i - 7 * g: i - 7 * g + 250]))
54     g+=1
55 print(substring)
56
57
58 # In[8]:
59
60
61 #generating pool C synthetic reads
62 substring=[]
63 g=0
64 for i in range(0,len(sequence),250):
65     if i==0:
66         for k in range(30):
67             substring.append(''.join(sequence[0:250]))
68     substring.append(''.join(sequence[i - 14 * g: i - 14 * g + 250]))
69     g+=1
70 print(substring)
71
72
73 # 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.

```
1  import itertools
2
3  def overlap(read_a, read_b, min_length=1):
4      start = 0 # start all the way at the left
5      while True:
6          start = read_a.find(read_b[:min_length], start)
7          if start == -1:
8              return 0
9          if read_b.startswith(a[start:]):
10             return len(read_a)-start
11         start += 1
12
13 def shortestCommonSuperstring(string_set):
14     shortest_sup = None
15     for perm in itertools.permutations(string_set):
16         sup = perm[0]
17         for i in range(len(string_set)-1):
18             olen = overlap(perm[i], perm[i+1], min_length=1)
19             sup += perm[i+1][olen:]
20
21         if shortest_sup is None or len(sup) < len(shortest_sup):
22             shortest_sup = sup
23     return shortest_sup
```

Figure 1:

Worked with: Amnah Abdelrahman, Ragava Vemnuri, Yogesh Kalapala

## Solution

```
1  #!/usr/bin/env python
2  # coding: utf-8
3
4  # In[18]:
5
6
7  def overlap(read_a, read_b, min_length):
8      max_possible_overlap_len = min(len(read_a), len(read_b))
9      max_overlap_len = 0
10     merged_read = ''
11
12     # check suffix of read_a with prefix of read_b
13     for i in range(max_possible_overlap_len - min_length + 1):
14         checking_overlap_len = max_possible_overlap_len - i
15         if (read_b.startswith(read_a[i:])):
16             max_overlap_len = checking_overlap_len
```

```

17         merged_read = read_a[:i] + read_b
18         break
19
20     # check suffix of read_b with prefix of read_a
21     for i in range(max_possible_overlap_len - min_length + 1):
22         checking_overlap_len = max_possible_overlap_len - i
23         if (read_a.startswith(read_b[i:])) and max_overlap_len < checking_overlap_len:
24             max_overlap_len = checking_overlap_len
25             merged_read = read_b[:i] + read_a
26             break
27
28     return (max_overlap_len, merged_read)
29
30
31 # In[19]:
32
33
34 def shortestCommonSuperstring(string_set, min_length=9):
35     while len(string_set)>1:
36         max_overlap_len = 0
37         max_merged_read = ''
38         max_overlap_indices = (-1, -1)
39         for i in range(len(string_set)):
40             for j in range(i+1, len(string_set)):
41                 overlap_len, merged_read = overlap(string_set[i], string_set[j],
42 min_length)
43                 if overlap_len > max_overlap_len:
44                     max_overlap_len = overlap_len
45                     max_merged_read = merged_read
46                     max_overlap_indices = (i, j)
47             if max_overlap_indices[0] == -1:
48                 string_set[0] = string_set[0] + string_set[1]
49                 del string_set[1]
50             else:
51                 string_set[max_overlap_indices[0]] = max_merged_read
52                 del string_set[max_overlap_indices[1]]
53         return string_set[0]
54
55 # In[29]:
56
57
58 #passing pool-A reads
59 strings=['
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCAATAC
    ',
    'ACGGTGTGCGTTAGAAAAGCTGCGACCGCACCCCGTGTCAACTTAAGCATTGGCCATAGGCTCCCATCCGCCATGTGTGGCATAACCGGCCATGTTATTGTCTBATC
    ',
    'TCAATCGTGTATAGTTTGGCATAACCATGTAAGGGCCTTCGAACCGACCCCACTCTTTGCAGTCTGCGAGAAGCAAGAAGCCTTACTCGAAAGGTAAATCGGAGGCT
    ',
    'TACAAAATCAAGGGTAAAACATCCATCGAGAGAACGTTGCGGCTGGAGAGATCTCATTAAAGCGTTCAATTGAGGTTATAGAGGTGAAAGCTTCGTCGCCCTATAAC
    ',
    'CCGAAGTTTGGCAGGTACACACTCTCAGCACCCACCCCACTAGGGCAGCCGATCCGCATAGGGGCACCAGATGATTTCGTGGGCCGACTATTGGCTATGCTACGCC
    ',
    'ACGGTGGGCACCAGGCAACACCCGATTGCTCTGCGGATACACACTTCTTGCGTTGTATGGGAGAGAATCTGTTTTCTAAACCCATTTAGGCCTCGAGACACCAATT
    ',
    'ATCTCAGATTGTACCCCGTGTCCAGTTTAAGGCTAAATCACATATTTCTGCGTACTTCACTAAGCACAAACAAAGTATCACCGTTAATTAACTTTGTCTCAGCGT
    ',
    'CGTATCGAGGTATCAGCATTGACGGCGCTAGTTAGTCTATAAATATTCGTTAGGAAACACAACATATAAGTATTGATATATCCCGCGGCCGAGTAGACAGTAAGAA
    ',
    'CGAAAGGGAGGTGATGTCAATCCAGTCCACTCCATAGAGAGGTTTATACTCCGTGTAGTGGGACTTAAGACCAGGTCCAGTACTAGACTATCCCATTTTACGACAC
    ',
    '

```

```

GGGGCATACCCCAAAGGCATGAGACCAACCACAGCACTCGCATTTCACTACCTTGAGGCCATGGCCGAAGGATATACGGAAC TGAAATCAACGTATAAACTGGG
|,|,|
TAGATCTTCGTCTCCCATGTAAGCTGGCCCTCGAGCCAGATGCCGTATAACAATGGGAGGACTGGAATCCAGATATGAGCTTCTTG CATCGCGGCGGAGTTGTACG
|,|,|
ATTATTTCCGCTCCCGAAGAACCTTTTCTACAGTATTGTGCTGTCGTCAACTTTGTTCCGCAATTGGCGTAGGGTAATGCCAGCGGTTAAGGGTTCGTACTATAA
|,|,|
GTGAATCACAGGCCCGTCGGTGGTTGGTACCGACGAGTACATGCTAATGGAGCTTACGAGGTCCTCATCCGATACAGAGAGCTGAATTGCTAATCCCTCCTTCAG
|,|,|
AGGCGGGGCGGGGATGCGGATCAAGGCCTTGAGAATAAACTGAAATTACATATCATTCTAGTGTGGAGAGCTGTGGCTAGTGACTGATCAGACATCAGCCGTTT
|,|,|
TCTTGCCAGACCCGCAAGGCAAGCTATATGCTTTCCACTTTAGAACAAGCGTTAGGATACCTATCAGTGCTTCAAGTAATTCATTGACCATTAGGTTATGTT
|,|,|
TATTGGAAGGTGGGCTTCCCTTTATTCGACGCGCTGCGATGGGTGCATCGATCTGGTAAGCCGCTTTCCAAAGGACCACGGCCAAGACTCCCTTTTGACATCG
|,|,|
ATCTATTCCTTAGGCTTGAACGTGGAGTCTCTCTATCGGTCATGCGGTGGCTCTCCGCTCAAGAACGCAGTCGGGGGAATGCGCTGCGTCCGACGACTGTTCA
|,|,|
AGCTTAACAAACCTCAGGTAGAATTACGGAGTTGATCAACTCCACTACTGTACAGCCATAAGGGGTGAAGAGTGGCTCATGATTGCGGATCGTTAGCTAACTCGC
|,|,|
GTGGCCAGGCGACTAAGAAAGGACACAGAAAAATTGCGCTCTACAGTATCTCAGCCGCCACTAATATTATTAGACGAGCTTCATTGTCTTACAATATTTTGATAT
|,|,|
GCATTGACTCGATAACAATAGGTTGATGCATTCACGCGCGGTCTATTGATCTTCATCCCGATAGTTATGTATGGACGTGTTCTTCTGTGAAAAAACTAGCTCTGTC
|,|,|
GCCGTAATAACATACCAGCGGACTAGTTGCCAAGGCTGATAGGCGGTCTACCGCGTTGCTGCACCTCCCTATCTCATGGAGCATGTGAATCTTGCGGTAGTCAC
|,|,|
CTACGTAAGAAGGTCACTATCCCTGGTGTGTCAGGCAATTGATAGCGATCACGTGAATGCGGCAGTATTTATGCCCTTTCTCGAGTATAGATTTATCCCTGGTATG
|,|,|
GGTCACCTAGTGCATAAGTTATAGACGATTTGTTGCAATTGGAATCTACCTATGGCGCTTAGGGAATAGTCACCGACTTAGGCAGAGTCAGTGCTAAACGCACAGG
|,|,|
TAGCCATTGTAGCGTCTAACCTGTCTACGCAACCAGTCCGACCTGCGGGAAATATACACATACAGCATTGCGCTGGGCAATGATACACAAACGTACCGTAGCCT
|,|,|
AAACCAGCGCTCTGATCATTCTCGTCTAGGAGGCCTGGAGTTAACACACTCACTCAATTATGAGAGCGAGGGTCGTAACCCGCTCCCGCAGTACTGCACGCGAGT
|,|,|
CAGAAAAACAACGATACTTTACGCGGACGGCAGTATGTCACATGCCGGTTTCAACATTACTCACTCCGCATTGCCTCCGTCTCCGTAAAGATCTAAGCTTTGTATGT
|,|,|
AGGGCAAAATATTATCCCTTGTTAGGTAAGTCAGACCTTCATTATTAGGGACCATCTTCAAAAGCACGCTCTAGATAATGTTTGTAACCCGCTCTCGTCAGCATAT
|,|,|
CCCATGGACAATGCTATTTACGCCGGGACGACGCTGTTGATAGAATCTTCAGGGCGTCTTACGCTAGGTGCCGTAACCGGGTTTCGGCCAGGACAGCGTTGGCC
|,|,|
AGCACCCAATCACGTTAGGTGAAATCTTGCACTTTGTTTATTATCCAGATCGTGTCCCAGTATGTCAAACCTAACTAAGGTACGAAACGAGTGCCTGGTTGCAACCA
|,|,|
CGGTACTGTGCGGTTAATACAATATCCATCCGGTCCAAAAGTACAGATCCCTTAAAGAATGATTGTGTACCTGATAAGGACCTTTAACTGATGCGGACAGGTTGCG
|,|,|
GCGATTACGAGATGACGGAGCCATGGGAGAATAGTCTGTGCGCTCAAGACCTAACCAAAACTCTACTTTTCCCTCCTGTTATAACACTGATTGGTCAAGATTGCGC
|,|,|
TTACTTATTTTGATAGCGATACAGTTGCACCTATAGTATCATCATTAGTCGGGCTGGGTGTTTAGCGAGTCATTAGAGACGGACGTTCGAATGTGCCCCGAGATT
|,|,|
GCCAATTGAGGTCGAGGTACTAAGCGAAGACCTTTCGCCCCAAGGGGCACTATGTGACTCACAACCTACCTCCTCTACAACGAATCAGCGCCAGCGCGAAGGTGCA
|,|,|
GTTTTATGTTGCGTACCCACTGTCCTCCAAATCCTGGACATCCTTGACAGACGCAACCATCGCAAATAGTAATAGTGATCGTCTTATTACTTGGGGTAAACTCCAA
|,|,|
TTCCTCTGTTTGGCAGTCCGCTGTCCGTACGCTTGATGGTAGAGGCTAGGGCCGGATATATACCTCCGCTTACTCATATGCAGAAGTTTAAACGAATGCGTTACCT
|,|,|
CACGGTACCTGGGCCTAGAACCCTCCTTGACGCGAGGATCGAGCTGACGTGCGCACACGGTCGGTTAGCTTCTCTGTAGTCAGTGCCGTGGAGTAACCGTCTTCG
|,|,|
TTAGAGGTCGGAAGGGTCCGAATAAATCCTCATCGACCCAATGACCACCGCGGTAACCGGTCACGGGTGCGGTCAGTGAAAAACAACAGAGCCCTGTTACCCGGA
|,|,|
AGTCTCTTCAATGTCAAGTGTTGACGGACTCGCTTTGCTCTAGTCGACCACAAAGCGCCATACAATAATTATGCAAGGTCTCGCCATTGGTCATACCGGGCGGTC
|,|,|
GACGGGTAAATGCCTCACAGTTATTCTCTATCCAACCTAGGATCAATTAAGGCGCTAAGTTGTCGAGGAAGGGGCTCAATCAGCCCTTCCCACCATCCGTCTCCA
|,|,|
GTCGGGAACGCCGCTCCTAACCCGCATCCGGCATGCGCACAAACCGCCTAGTTGTGCGAGTTCTGCAAGGTCGTTTCGTATGTGTTAAGAAGATGAGAGGATTAT
|,|,|
']

```



```

1 3
2 TTATAGTTAAATAGCGTATGATCAAGTTCCACAGGTCGATAGTTGGACTCGTATCGAGGTATCAGCATTGACGGCGCTAGTTAGTCTATAAATATTCGTTAGCAAA
3 1
4 3
5 GTTTTGGTAAGGGAGTACCTATGATGTATTGTTCAAGAAGAGGGGCACAGTAATGGACGAAAGGGAGGTTCATGTCAATCCAGTCCACTCCATAGAGAGGTTTATACT
6 1
7 3
8 CCAGGATGTTTCGATGATATCAGGCGCGAAGTATCGGAAGCGTCATATTAAGCACATATTTAATGGGGCATACCCCAAAGGCATGAGACCAACCACAGCACTCGCA
9 1
10 3
11 CATTGTTCTTCGGGACTAGTAGAGGTCGTTGTCGTTGCGAACACGTCGAGTTGCAAGTGCTCAAGCTTCTACTAGATCTTCGCTCTCCCATGTAAGCTGGCCCTCGAGC
12 1
13 3
14 CTGCTCTCACCCGCACGCCGGATGTCAGGGCTGGAAGATGGCTATAACTCTTAGCAGACCTGGATGTGTCGTTACCGATTATTTCCGCTCCCGAAGAACCTTTTC
15 1
16 3
17 ACCATCAGACATACCGAAGCGCCGACTGACAGTTCATTATTTGACGGTGAACAGTACTGGACGCCAGTATTAATGCTGTATCATGTGAATCACAGGCCCGTCGGTG
18 1
19 3
20 AAGAGACCCCGCACGACACAAATGCGTAGCAAGCTGCTTATCCAGCAATGAGGTGTGGTGGTGAGCGGAGTTAACACGGGACGCATCGGAGAGGCGGGGCGGGGAG
21 1
22 3
23 TCGATATTATATTCGACTTTAATCCTAATGGACGTCCCAAGGTACCCGGAGTGGGACGACCGCCACCCCTCGTAAGGTGCCTACAGCACCCGACCCCTCTTGCCA
24 1
25 3
26 GCCTGTAAATTGCAAAATAAGCCATATTCTTGTGAATCGATGCCCTCATATCACATTATCTACTTCTCGTATGCTAATTGGCAGCGATGAGAGAACCCTCATAGT
27 1
28 3
29 AGGACCGCCGCAAAAATATGGCGCTCACACAAATCGCTGTCATTATGTGTATGACTTGCTTGTCGCCGTTTCGTGCAGATTTTATGATACACTGAACGCTATTAA
30 1
31 3
32 ATGGTAAAGGCTCAAAAAGTACCCGACTTTGGGATTAGACGCGAACCTGGGTACCAGCTATTAAGCATAGTTTTTCTGAATGGCCACGGGCGGGTGGGGGAGTTT
33 1
34 3
35 CCGGACGATGGCGAGGTGAAGATAGTAAAGTAAGCTGTGGCGGTATATTTATTTAGATAGGACACGGGTAAGTGGTGACTATTGCGTTGTGCGTTCGTAAATTTCCG
36 1
37 3
38 ATACGAGGACGTATGGCCGTGATGTCGCTCGTAGCGATTGCTTATGGAGCTTACCCTTGATCATGGTACGCTGTACTTGATATAGAGAAGACTTTTGTACCGGCGC
39 1
40 3
41 GTCCTAAGACAGAATAAAACCGAACGTTGATGCCGAGTGTAACGTATAGATGTTTATCTTATACCTGTGGAGCATGGACTTATAAGATTGATCTCTGGATGTACCG
42 1
43 3
44 CACTAATGATCTTTTGGACATTGAGAGTCAGTGCCGATTACCGACTGTCCGTGTACGCGTACTGCAACGGTGGGACTAAATACGGTACCTCAATTTCGACGCGTGG
45 1
46 3
47 CCCTGGTTAGCCACTTTCGAGACCGGGCATTATAGGACGGACTATCCGCCAGAACAAGAGATTGTTATCCTTGCTCATACCACGGATGCACGCCCCCCCAGTCACT
48 1
49 3
50 AGTGCTAAACGCACAGGTGGTGAGGATTACTATAACCGCGGTGCGCACATGGCTAAGTAAGTCCCTCGATCCATAGTTTCACATCCACCCGGACGTAATCCACGA
51 1
52 3
53 TGATACACAAACGTACCGTAGCCTCCCTGCGAAGAGACTCATGCATCAATGAACTAACAAGTCGGCCTGCTATTCTGTATATCTCATCCCGTCCCTCCCTTAACAGCA
54 1
55 3
56 GTAACCCGCTCCCGCAGTACTGCACGCGAGTGATACAGTACCAGTGAAAGTCTATCGCATGTCAGGTGCAGCTAACACCCCGAACCGAACTACGCGTACGCTGCTC
57 1
58 3
59 ATTGCTCCGTCTCCGTAAAGATCTAAGCTTTGTATGTTCCGGTCAATCCACCTCTGGCGTGCTTGGTTGATTTTTGCTCTACAGCTCTGGCCAATCGTTGTCAAC
60 1
61 3
62 AAGCACGCTCTAGATAATGTTTGTAAACCGCCTCTCGTCAGCATATTTTTCACAGAGCGGATCTGTGTAAGCTCCGGCTCGGTGCTGTTGAGTCGGATTTTGCTT
63 1
64 3
65 GCGTCTTACGCTAGGTGCCGTAACGCGGTTTCGGCCAGGACAGCGTTGGCCCCAGTGCCCCCACAAGTGAGAGCGCGTGGATAAAAACAGGACATAACGCGCCT
66 1
67 3
68 GATCGTGTCCAGTATGTCAAATAACTAAGGTACGAAACGAGTGCCTGGTTGCAACCACGCCAGATCCGATCATTGGTAAAAATCCAGACAGTCCCGGCCACTT
69 1
70 3
71 GTACAGATCCCTTAAAGAATGATTGTGTACCTGATAAGGACCTTTAACTGATGCGGACACGTTTGGCCAGCCCCGTACTCCGTTTACGGACGCGACATGCCGCAAT
72 1
73 3
74 GTCTGTGCGCTCAAGACCTAACCAAACTCTACTTTTCCCTCCTGTTATAAACTGATTGGTCAAGATTGCGCTAATCCGACGGAGGGGCAAAATCGAGCACAGTTA
75 1
76 3
77 GCACCTATAGTATCATCATTAGTCGGGCTGGGTGTTTAGCGAGTCATTAGAGACGGACGTTTGAATGTGCCCCGAGATTCCGGTCTAATATAACAGCCATTTCCGGC
78 1
79 3
80 CTAAGCGAAGACCTTTGCCCCAAGGGGCACTATGTGACTCACAACCTCCTCCTACAACGAATCAGCGCCAGCGCGAAGGTTGCAAGGGTAGACGTGCGTACAT
81 1
82 3
83 GTACCCACTGTCCTCCAAATCCTGGACATCCTTGACAGACGCAACCATCGCAAATAGTAATAGTGATCGTCTTATTACTTGGGGTAAACTCCAACGGAGTGAGTCG
84 1
85 3
86 CTGTTTGGCAGTCCGCTGTCCGTACGCTTGATGGTAGAGGCTAGGGCCGGATATATACCTCCGCTTACTCATATGCAGAAGTTTAAAGTAATGCGTTACCTTGAC
87 1
88 3
89 ATCACGGTACCTGGGCCTAGAACCCTCCTTGACGCGAGGATCGAGCTGACGTGGGCACACGGTCGGTTAGCTTCCCTCTGTAGTCAGTGCCGTGGAGTAACCGTCTT
90 1
91 3
92 CGTTCGCATTAGAGGTCCGAAAGGGTCCGAATAAATCCTCATCGACCCAATGACCACCGCGGTAACCGGTACGGGTGCGGTCAGTGAAAAACAACAGAGCCGTGT

```





[illegible]

```

TCGCGATTGGGGTCCGGTGGCCAGGCGACTAAGAAAAGGACACAGAAAAATTGCGCTCTACAGTATCTCAGCCGCCCACTAATATTATTAGACGAGCTTCATTGCTCT
',
',
CGCGATTGCCATATGGTGATATCGCTTGGGCATTGACTCGATAACAATAGGTTGATGCATTCACGCGCGGTCTATTGATCTTCATCCGATAGTTATGTATGACG
',
',
TGGATGTACCGGTTGCTTTGACAAAAGCTTGGCTTTATGGCTTCGCCGTAAATACATACCAGCGGACTAGTTGCCAAGGCTGATAGGGCGGTTCATCCGCGTTCTG
',
',
CTCAATTGCACGCTGGGAACATCAAGAAGCCTGTCGTGTGCCGAGAGTGATCTATGTCTACGTAAGAAGGTCACCTATCCCTGGTGTGTCAGGCAATTGATACGAT
',
',
CGGATGCACGCCCCCAGTCACCTTACATCTTACGCCGAACCTGGACGTGAAGGGATGAATAAGGAACGGCCGGTCACCTAGTGCATAAGTTATAGACGATTTGTT
',
',
AGTTTCACATCCACCCGACGTAATCCACGAGAAAAAGGGCAATCTTATCACGAAATTCAGCGGAAGCCGTCGAGGTATCAGGTGTAGCCATTGTAGCGTCTAAC
',
',
GCTATTCTGATATCTCATCCCGTCTCCCTTAACAGCATAGGGGCTGACTAGACACCGCGTGTGGCCGACTTCATCGCGATCGGAATAGATATAGTTAATAAACCA
',
',
TCAGGTGCAGCTAACACCCGAACCGAACTACGCGTACGCTGCTCGCCGAATTCATTCTGTTAAAAAGCACGGGCAGACGGTTAAAGGCTGTTGTATAACCTAGTT
',
',
CTGGCGTGCTTGGTTGATTTTTGCTCTACAGCTCTGGCCAATCGTTGTGCAAGTAGATGACGACTATTGCCTATGGTATGTTATATGCCATTGGATCACCGGTCT
',
',
TTTCACAGAGCGGATCTGTGTAAGCTCCGGCTCGGTGCTGTGAGTCGGATTTTGGCTTCACATTTAACTTTCGTCCTCTAAAGTCACTACAAAACATGCTGGTTC
',
',
CAGCGTTGGCCCCAGTGGCCCCACAAGTGAGAGCGCGTGGATAAAAAACCAGGACATAACGCGCCTGGCCGTTCCAGGGTCCAAACTGCTGTTCCCGTCAACGCG
',
',
ACGAAACGAGTGCCTGGTTGCAACCACCGCCAGATCCGATCATTGGTAAAAATCCAGACAGTCCGGGCCCACTTATACTGGCAAGAATCTACTCCAAAACGCCACGG
',
',
GTACCTGATAAGGACCTTTAACTGATGCGGACACGTTTGCCAGCCCCGTACTCCGTTTACGGACGCGACATGCCGCAATCCACGGGTGCTGAGAGCCTTTCGAC
',
',
AACCAAACTCTACTTTTCCCTCCTGTTATAAACTGATTGGTCAAGATTCCGCTAATCCGACGGAGGGGCAAAATCGAGCACAGTTAATATAGAAAGGTCTTGT
',
',
TCATCATTAGTCGGGCTGGGTGTTTAGCGAGTCATTAGAGACGGACGTTTGAATGTGCCCGCAGATTCGGGTCTAATATAACAGCCATTCGGCGATGCCACAGT
',
',
CGAAGACCTTTTCGCCCCAAGGGGCACTATGTGACTCACAACCTACCTCCTCTACAACGAATCAGCGCCAGCGCGAAGGTTGCAAGGGTAGACGTGCGTACATACG
',
',
GCGTACCCACTGTCTCCAAATCCTGGACATCCTTGACAGACGCAACCATCGCAAATAGTAATAGTGATCGTCTTATTACTTGGGGTAACTCCAACGGAGTCAGT
',
',
CGGATTCTTCTGTTTGGCAGTCCGCTGTCCGTACGCTTGCATGGTAGAGGCTAGGGCCGGATATATACCTCCGCTTACTCATATGCAGAAGTTTAAACGAATGCGT
',
',
AGCCTACATCGATTGCATCACGGTACCTGGGCCTAGAACCCTCCTTGACGCGAGGATCGAGCTGACGTGCGGCACACGGTCGGTTAGCTTCCTCTGTAGTCAGT
',
',
TACCTTCTCGACCCAAGCTCCATCGTTCCGCATTAGAGGTCCGAAAGGGTCCGAATAAATCCTCATCGACCCAATGACCACCGCGGTAACCGGTACGGGTGCGGT
',
',
ATACCCACCCTTATTCAAAAACCAGTCTCAGCTGGCACAAGAAGAAGTCTCTTCAATGTCAAGTGTGACGGACTCGCTTTGCTCTAGTCGACCACAAAGCCCA
']
102
103
104 # In[36]:
105
106
107 #Shortest Common Superstring for pool-C reads
108 print((shortestCommonSuperstring(strings,9)))
109
110
111 # In[38]:
112
113
114 #length of shortest common superstring after passing pool-C reads
115 print(len(shortestCommonSuperstring(strings,9)))
116
117
118 # 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

```
1 #!/usr/bin/env python
2 # coding: utf-8
3
4 # In[3]:
5
6
7 class Node:
8     """ Class Node to represent a vertex in the de bruijn graph """
9     def __init__(self, lab):
10         self.label = lab
11         self.indegree = 0
12         self.outdegree = 0
13
14 class Edge:
15     def __init__(self, lab):
16         self.label = lab
17
18 def read_reads(reads):
19
20     return reads
21
22 def construct_graph(reads, k):
23     """ Construct de bruijn graph from sets of short reads with k length word"""
24     edges = dict()
25     vertices = dict()
26
27     for read in reads:
28         i = 0
29         while i+k < len(read):
30             v1 = read[i:i+k]
31             v2 = read[i+1:i+k+1]
32             if v1 in edges.keys():
33                 vertices[v1].outdegree += 1
34                 edges[v1] += [Edge(v2)]
35             else:
36                 vertices[v1] = Node(v1)
37                 vertices[v1].outdegree += 1
38                 edges[v1] = [Edge(v2)]
39             if v2 in edges.keys():
40                 vertices[v2].indegree += 1
41             else:
42                 vertices[v2] = Node(v2)
43                 vertices[v2].indegree += 1
44                 edges[v2] = []
45             i += 1
46
47     return (vertices, edges)
48
49 def output_contigs(g):
50     """ Perform searching for Eulerian path in the graph to output genome assembly"""
51     V = g[0]
```

```

52     E = g[1]
53     # Pick starting node (the vertex with zero in degree)
54     start = V.keys()
55     for k in V.keys():
56         if V[k].indegree < V[start].indegree:
57             start = k
58
59     contig = start
60     current = start
61     while len(E[current]) > 0:
62         # Pick the next node to be traversed (for now, at random)
63         next = E[current][0]
64         del E[current][0]
65         contig += next.label[-1]
66         current = next.label
67
68     return contig
69
70 def print_graph(g):
71     """ Print the information in the graph to be (somewhat) presentable """
72     V = g[0]
73     E = g[1]
74     for k in V.keys():
75         for e in E[k]:
76             print(e.label)
77
78
79 # In[4]:
80
81
82 #De Bruijn Graph for pool-A for k=5
83 construct_graph(['
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCGAAAGGATGAGGCAATAG
    ',
    ',
    ACGGTGTGCGTTAGAAAAGCTGCGACCGCACCCCGTGTCAACTTAAGCATTGGCCATAGGCTCCCATCCGCCCATGTGTGGCATAACCGGCCATGTTATTGTCTBATC
    ',
    ',
    TCAATCGTGTATAGTTTGGCATAACCATGTAAGGGCCTTCGAACCGACCCCACTCTTTGCAGTCTGCGAGAAGCAAGAAGCCTTACTCGAAAGGTAAATCGGAGGCT
    ',
    ',
    TACAAAATCAAGGGTAAACATCCATCGAGAGAAGCTTGCGGCTGGAGAGATCTCATTAAAGCGTTCAATTGAGGTTATAGAGGTGAAAGCTTCGTCGCGCCCTATA
    ',
    ',
    CCGAAGTTTGGCAGGTACACACTCTCAGCACCCACCCCACTAGGGCAGCCGATCCGCATAGGGGCACCAGATGATTGCGTGGGCCGACTATTGGCTATGCTACGCC
    ',
    ',
    ACGGTGGGCACCAGGCAACACCCGATTGCTCTGCGGATACACACTTCTTGCGTTGTATGGGAGAGAATCTGTTTTCTAAACCCATTTAGGCCTCGAGACACCATTT
    ',
    ',
    ATCTCAGATTGTACCCCGTGTCCAGTTTAAGGCTAAATCACATATTTCTGCGTACTTCACTAAGCACAACAAAAGTATCACCGTTAATTAACTTTGTCTCACC
    ',
    ',
    CGTATCGAGGTATCAGCATTGACGGCGCTAGTTAGTCTATAAATATTCGTTAGGAAACACAACATATAAGTATTGATATATGCCGCGGCCGAGTAGACAGTAAGAA
    ',
    ',
    CGAAAGGGAGGTATGTCAATCCAGTCCACTCCATAGAGAGGTTTATACTCCGTGTAGTGGGACTTAAGACCAGGTCCACGTACTAGACTATCCCATTTTACGACA
    ',
    ',
    GGGGCATACCCCAAAGGCATGAGACCAACCACAGCACTCGCATTTCACTACCTTGAGGCCATGGCCGAAGGATATACGGAAGTAAAAATCAACGTATAAACTGGG
    ',
    ',
    TAGATCTTCGTCTCCCATGTAAGCTGGCCCTCGAGCCAGATGCCGTATAAATGGGAGGACTGGAATCCAGATATGAGCTTCTTGATCGCGGGGAGTTGTACG
    ',
    ',
    ATTATTTTCGCTCCCGAAGAACCTTTTCTACAGTATTGTGCTGTCGTCACACTTTGTTTCGCGCAATTGGCGTAGGTAATGCCAGCGGTTAAGGGTTCGTACTATA
    ',
    ',
    GTGAATCACAGGCCCGTCGGTGGTTGGTACCGACGAGTACATGCTAATGGAGCTTACGAGGTCCTCATCCGATACAGAGAGCTGAATTGCTAATCCCTCCTTCAG
    ',
    ',
    AGGCGGGGCGGGGATGCGGATCAAGGCCTTGAGAATAAACTGAAATTACATATCATTTCTAGTGTGGAGAGCTGTGGCTAGTGACTGATCAGACATCAGCCGTT
    ',
    ',
    TCTTGCCAGACCCGCAAAGGCAAGCTATATGCTTTCACTTTAGAACAAAGGCGTTAGGATACCTATCAGTGCTTCAAGTAATTCATTGACCATTAGGTTATBT
    ',
    ',

```

```

TATTGGAAGGGTGGGCTTCCCTTTATTCGCAGCGCGTCGCATGGGTGCATCGATCTGGTAAGCCGCTTTCCAAAGGACCACGGCCAAGACTCCCCTTTTGACATCGG
',
',
ATCTATTCCCTTAGGCTTGGAACGTGGAGTCTCTCTATCGGTATCGGGTGGCTCTCCGCTCAAGAACGCAGTCGGGGGAATGCGCTGCGTCCGACGACTGTTGAGC
',
',
AGCTTAACAAACCTCAGGTAGAATTACGGAGTTGATCAACTCCACTACTGTACAGCCCATAAGGGGTGAAGAGTGGCTCATGATTGCGGATCGTTAGCTAACTCGC
',
',
GTGGCCAGGCGACTAAGAAAGGACACAGAAAAATTGCGCTCTACAGTATCTCAGCCGCCACTAATATTATTAGACGAGCTTCATTGTCTTACAATATTTTGATATC
',
',
GCATTGACTCGATAACAATAGGTTGATGCATTACGCGCGGTCTATTGATCTTCATCCCGATAGTTATGTATGGACGTGTTCTTCTGTGAAAAAACTAGCTCTGTC
',
',
GCCGTAATACATACCAGCGGACTAGTTGCCAAGGCTGATAGGGCGGTCTCCGCGTTGCTGCACTCCCCTATCTCATGGAGCATGTGAATCTTGGCGTTAGTCAC
',
',
CTACGTAAGAAGGTCACTATCCCTGGTGTGTGAGGCAATTGATAGCGATCACGTGAATGCGGCAGTATTTATGCCCTTTCTCGAGTATAGATTTATCCCTGGTATAG
',
',
GGTCACCTAGTGCATAAGTTATAGACGATTTGTTGCAATTGGAATCTACCTATGGCGCTTAGGGAATAGTCACCGACTTAGGCAGAGTCAGTGTAAACGCACAGG
',
',
TAGCCATTGTAGCGTCTAACCTGTCTACGCAACCAGTCCGACCTGCGGGAAATATACATACAGCATTGCGCTGGGCAATGATACACAAACGTACCGTAGCCTC
',
',
AAACCAGCGCTCTGATCATTCTCGTCTAGGAGGCTGGAGTTAACACACTCACTCAATTATGAGAGCGAGGGTCGTAACCCGCTCCCGCAGTACTGCACGCGAGT
',
',
CAGAAAAACAACGATACTTTACGCGGACGGCAGTATGTCACATGCCGGTTTCAACATTACTCACTCCGCATTGCCTCCGCTCTCCGTAAAGATCTAAGCTTTGTATGT
',
',
AGGGCAAAATATTATCCCTTGTTAGGTAAGTCAGACCTTCATTATTAGGGACCATTCTCAAAGCACGCTCTAGATAATGTTTGTAACCGCCTCTCGTCAGCATAT
',
',
CCCATGGACAATGCTATTTACGCCGGGACGACGCTGTTGATAGAATCTTCAGGGCGTCTTACGCTAGGTGCCGTAACGCGGTTTCGGCCAGGACAGCGTTGCCCG
',
',
AGCACCCAATCACGTTAGGTGAAATCTTGCACTTTGTTTCATTATCCAGATCGTGTCCAGTATGTCAAATACTAAGGTACGAAACGAGTGCCTGGTTGCAACCA
',
',
CGGTACTGTGCGGTTAATACAATATCCATCCGTCCAAAAGTACAGATCCCTTAAAGAATGATTGTGTACCTGATAAGGACCTTTAACTGATGCGGACAGCTTTGC
',
',
GCGATTACGAGATGACGGAGCCATGGGAGAATAGTCTGTGCGCTCAAGACCTAACCAAACTCTACTTTTCCCTCCTGTTATAAACTGATTGGTCAAGATTGCGC
',
',
TTACTTATTTTGATAGCGATACAGTTGCACCTATAGTATCATCATTAGTCGGGCTGGGTGTTTAGCGAGTCATTAGAGACGGACGTTGCAATGTGCCCCGAGATT
',
',
GCCAATTGAGGTCGAGGTACTAAGCGAAGACCTTCGCCCCAAGGGGCACTATGTGACTCACAATACTCCTCCTACAACGAATCAGCGCCAGCGCGAAGGTGCA
',
',
GTTTTATGTTGCGTACCCACTGTCCTCCAAATCCTGGACATCCTTGCGAGACGCAACCATCGCAAATAGTAATAGTGATCGTCTTATTACTTGGGGTAACTCCAA
',
',
TTCTTCTGTTTGGCAGTCGCTGTCCGTACGCTTGATGGTAGAGGCTAGGGCCGATATATACCTCCGCTTACTCATATGCAGAAGTTTAAAGAAATGCGTTACCT
',
',
CACGGTACCTGGGCTAGAACCCTCCTTGACGCGAGGATCGAGCTGACGTGCGCACACGGTCGGTTAGCTTCTCTGTAGTCAGTGCCGTGGAGTAACCGTCTTCG
',
',
TTAGAGGTCGAAAGGGTCCGAATAAATCCTCATCGACCCAATGACCACCGCGGTAACCGGTACAGGGTGCAGTCAGTAAAAACAACAGAGCCCTGTTACCCGGA
',
',
AGTCTCTTCAATGTCAAGTGTTGACGGACTCGCTTTGCTCTAGTCGACCACAAAGCGCCATACAATAATTATGCAAGGTCTCGCCATTGGTCATACCGGGCGGTC
',
',
GACGGGTAAATGCCTCACAGTTATTCTCTATCCAACCTAGGATCAATTAAGGCGCTAAGTTGTGAGGAAGGGCTCAATCAGCCCTTCCACCATCCGTCTCCAA
',
',
GTCGGGAACGCGCCTCCTAACCCGCATCCGGCATGCGCACAAACCGCCTAGTTGTGCGAGTTCTGCAAGGTCGTTTCGTATGTGTTAAGAAGATGAGAGGATTAT
'], 5)

```

```

84
85
86 # In[10]:
87
88
89 #De Bruijn Graph for pool-A for k=9
90 construct_graph(['
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCGAAAGGATGAGGCAAT
    ',
    ',
    ACGGTGTGCGTTAGAAAGCTGCGACCGCACCCCGTGTCAACTTAAGCATTGTCATAGGCTCCCATCCGCCCATGTGTGGCATACCGCGCCATGTTATTGTCTGAT
    ',
    ',
    TCAATCGTGTATAGTTTGGCATAACCATGTAAGGGCCTTCGAACCGACCCCACTCTTTGCAGTCTGCGAGAAGCAAGAAGCCTTACTCGAAAGGTAATTCGGAGGCT

```

```

1 3
2 TACAAAATCAAGGGTAAAACATCCATCGAGAGAACGTTGCGGCTGGAGAGATCTCATTAAAGCGTTCAATTGAGGTTATAGAGGTGAAAGCTTCGTCGCGCCCTATA
3 1
4 CCGAAGTTTGGCAGGTACACACTCTCAGCACCCACCCAACTAGGGCAGCCGATCCGCATAGGGGCACCAGATGATTCGTGGGCCGACTATTGGCTATGCTACGCC
5 1
6 ACGGTGGGCACCAGGCAACCCCGATTGCTCTGCGGATACACACTTCTTGCGTTGTATGGGAGAGAATCTGTTTTCTAAACCCATTAGGCCTCGAGACACCATTTC
7 1
8 ATCTCAGATTGTACCCCGTGTCCCAGTTTAAGGCTAAATCACATATTTCTGCGTACTTCACTAAGCACAAACAAAGTATCACCGTTAATTAACTTTGTCTCTACCG
9 1
10 CGTATCGAGGTATCAGCATTGACGGCGCTAGTTAGTCTATAAATATTGTTAGGAAACACAACATATAAGTATTGATATATGCCGCGGCCGAGTAGACAGTAAGAA
11 1
12 CGAAAGGGAGGTGATGTCAATCCAGTCCACTCCATAGAGAGGTTTATACTCCGTGTAGTGGGACTTAAGACCAGGTCCACGTACTAGACTATCCCATTTTACGACA
13 1
14 GGGGCATACCCCAAAGGCATGAGACCAACCACAGCACTCGCATTTCACTACCTTGAGGCCATGGCCGAAGGATATACGGAACGAAAATCAACGTATAAACTGGG
15 1
16 TAGATCTTCGTCTCCCATGTAAGCTGGCCCTCGAGCCAGATGCCGTATAACAATGGGAGGACTGGAATCCAGATATGAGCTTCTTGTCATCGCGCGGAGTTGTACG
17 1
18 ATTTATTTCCGCTCCCGAAGAACCTTTTCTACAGTATTGTGCTGTGCTCAACTTTGTTGCGCAATTGGCGTAGGGTAATGCCAGCGGTTAAGGGTTTCGTACTATA
19 1
20 GTGAATCACAGGCCCCGTGCGTGGTTGGTACCGACGAGTACATGCTAATGGAGCTTACGAGGTCCCTCATCCGATACAGAGAGCTGAATTGCTAATCCCCTCCTTCAG
21 1
22 AGGCGGGGCGGGGATGCGGATCAAGGCCTTGAGAATAAACTGAAATTACATATCATTCTAGTGTGGAGAGCTGTGGCTAGTGACTGATCAGACATCAGCCGTTTC
23 1
24 TCTTGCCAGACCCGGCAAAGGCAAGCTATATGCTTTCCACTTTAGAACAAGGCGTTAGGATACCTATCAGTGCTTCAAGTAATTCATTGACCATTAGGTTATGTT
25 1
26 TATTGGAAGGTGGGCTTCCCTTTATTGCGAGCGCGTCGCATGGGTGCATCGATCTGGTAAGCCGCTTTCCAAAGGACCACGGCCAAGACTCCCTTTTGACATCG
27 1
28 ATCTATTCCTTAGGCTTGAACGTTGGAGTCTCTCTATCGGTGTCGCGTGGGCTCTCCGCTCAAGAACGCAGTCGGGGGAATGCGCTGCGTCCGACGACTGTTACG
29 1
30 AGCTTAACAAACCTCAGGTAGAATTACGGAGTTGATCAACTCCACTACTGTACAGCCATAAGGGGTGAAGAGTGGCTCATGATTGCGGATCGTTAGCTAACTCGC
31 1
32 GTGGCCAGGCGACTAAGAAAGGACACAGAAAAATTGCGCTCTACAGTATCTCAGCCGCCCCTAATATTATTAGACGAGCTTCATTGTCTTACAATATTTTGATAT
33 1
34 GCATTGACTCGATAACAATAGGTTGATGCATTACGCGCGGTCTATTGATCTTCATCCCGATAGTTATGTATGGACGTGTTCTTCTGTGAAAAAAGTAGCTCTGGT
35 1
36 GCCGTAATAACATACCAGCGGACTAGTTGCCAAGGCTGATAGGGCGGTGTCATCCGCGTTGCTGCACTCCCCTATCTCATGGAGCATGTGAATCTTGGCGTTAGTCA
37 1
38 CTACGTAAAGAAGGTCACTATCCCTGGTGTGTCAGGCAATTGATAGCGATCACGTGAATGCGGCAGTATTTATGCCCTTCTCGAGTATAGATTTATCCCTGGTTAG
39 1
40 GGTACCTAGTGCATAAGTTATAGACGATTTGTTGCAATTGGAATCTACCTATGGCGCTTAGGGAATAGTCACCGACTTAGGCAGAGTCAGTGCTAAACGCAACAG
41 1
42 TAGCCATTGTAGCGTCTAACCTGTCTACGCAACCAGTCCGGACCTGCGGGAAATATACACATACAGCATTGCGCTGGGCAAAATGATACACAAACGTACCGTAGCCT
43 1
44 AAACCAGCGCTCTGATCATTCTCTAGGAGGCCTGGAGTTAACACACTCACTCAATTATGAGAGCGAGGGTCGTAACCCGCTCCCGCAGTACTGCACGCGAGT
45 1
46 CAGAAAAACAACGATACTTTACGCGGACGGCAGTATGTACATGCCGTTTCAACATTACTCACTCCGATTGCGTCCGTCTCCGTAAAGATCTAAGCTTTGTATGT
47 1
48 AGGGCAAAATATTATCCCTTGTAGGTAAGTCAGACCTTCATTATTAGGACCATCTTCAAAGCAGCTCTAGATAATGTTTGTAACCGCCTCTCGTCAGCATAT
49 1
50 CCCATGGACAATGCTATTTACGCGGCGACGACGCTGTTGATAGAATCTTCAGGGCGTCTTACGCTAGGTGCCGTAACGCGGTTTCGGCCAGGACAGCGTTGGCC
51 1
52 AGCACCCCAATCACGTTAGGTGAAATCTTGCACTTTGTTTATTATCCAGATCGTGTCCAGTATGTCAAATAACTAAGGTACGAAACGAGTGCCCTGGTTGCAACCA
53 1
54 CGGTACTGTCGGGTTAATACAATATCCATCCGGTCCAAAAGTACAGATCCCTTAAAGAATGATTGTGTACCTGATAAGGACCTTTAACTGATGCGGACAGTTTGC
55 1
56 GCGATTACGAGATGACGGAGCCATGGGAGAATAGTCTGTGCGTCAAGACCTAACCAAACTCTACTTTTCCCTCCTGTTATAACACTGATTGGTCAAGATTGCGC
57 1
58 TTACTTATTTTGATAGCGATACAGTTGCACCTATAGTATCATCATTAGTCGGGTGGGTGTTTAGCGAGTCATTAGAGACGGACGTTTCAATGTGCCCGCAGATTCC
59 1
60 GCCAATTGAGGTGAGGTACTAAGCGAAGACCTTTGCCCCAAGGGGCACTATGTGACTCACAACCTCCTCCTACAACGAATCAGCGCCAGCGCAAGGTTGCA
61 1
62 GTTTTATGTTGCGTACCCACTGTCCTCCAAATCCTGGACATCCTTGACAGACAGCAACCATCGAAATAGTAATAGTGATCGTCTTATTACTTGGGGTAAACTCCAA

```

```

    ' ',
    TTCTTCTGTTTGGCAGTCCGCTGTCGCTACGCTTGCATGGTAGAGGCTAGGGCCGGATATATACCTCCGCTTACTCATATGCAGAAGTTTAACGAATGCGTTTACCT
    ' ',
    ' ',
    CACGGTACCTGGGCCTAGAACCCCTCCTTGACGCGAGGATCGAGCTGACGTCGGCACACGGTCGGTTAGCTTCCTCTGTAGTCAGTGCCGTGGAGTAACCGTCTTCGG
    ' ',
    ' ',
    TTAGAGGTCCGAAAGGTCCGAATAAATCCTCATCGACCCAATGACCACCGCGGTAACCGGTACGGGTGCGGTCAAGTAAAAACAACAGAGCCCTGTTACCGCGGA
    ' ',
    ' ',
    AGTCTCTTCAATGTCAAGTGTGACGGACTCGCTTTGCTCTAGTCGACCACAAAGCGCCATACAATAATTATGCAAGGTCTCGCCATTGGTGCATACCGGGCGGGT
    ' ',
    ' ',
    GACGGGTAAATGCCTCACAGTTATTCTCTATCCAAGTATGATCAATTAAGGCGCTAAGTTGTCGAGGAAGGGGCTCAATCAGCCCTTCCACCATCCGTCTCCAA
    ' ',
    ' ',
    GTCGGGAACGCGCCTCCTAACCCGCATCCGGCATGCGCACAAACCGCCTAGTTGTGCGAGTTCTGCAAGGTGTTTTCGTATGTGTTAAGAAGATGAGAGGATTAT
    ' ',
    ' ',
    ']', 9)

91
92
93 # In[12]:
94
95
96 #De Bruijn Graph for pool-A for k=18
97 construct_graph(['
    CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCGAAAGGATGAGGCATA
    ' ',
    ' ',
    ACGGTGTGCGTTAGAAAAGTGCACCGCACCCCGTGCAACTTAAGCATTGGCCATAGGCTCCCATCCGCCCATGTGTGGCATAACCGGCCATGTTATTGTCTGATC
    ' ',
    ' ',
    TCAATCGTGATAGTTTGGCATAACCATGTAAGGGCCTTCGAACCGACCCCACTCTTGCAGTCTGCGAGAAGCAAGAAGCCTTACTCGAAAGGTAAATCGGAGGCT
    ' ',
    ' ',
    TACAAAATCAAGGGTAAACATCCATCGAGAGAAGCTTGGCGGTGGAGAGATCTCATTAAAGCGTTCAATTGAGGTTATAGAGGTGAAAGCTTCGTGCGGCCCTATA
    ' ',
    ' ',
    CCGAAGTTTGGCAGGTACACACTCTCAGCACCCACCCCACTAGGGCAGCCGATCCGCATAGGGGCACCAGATGATTGTTGGGCCGACTATTGGCTATGCTACGCC
    ' ',
    ' ',
    ACGGTGGGCACCAGGCAACACCCGATTGCTCTGCGGATACACACTTCTTGGCTGTATGGGAGAGAATCTGTTTTCTAAACCCATTAGGCTCGAGACCACTTCT
    ' ',
    ' ',
    ATCTCAGATTGTACCCCGTGTCCAGTTTAAGGCTAAATCACATATTTCTGCGTACTTCACTAAGCACAAACAAAGTATCACCGTTAATTAACCTTGTCTCACCG
    ' ',
    ' ',
    CGTATCGAGGTATCAGCATTGACGCGCTAGTTAGTCTATAAATATTCGTTAGGAAACACAACATATAAGTATTGATATATGCCGCGCCGAGTAGACAGTAAGAA
    ' ',
    ' ',
    CGAAAGGGAGGTGATGTCAATCCAGTCCACTCCATAGAGAGGTTTATACTCCGTGTAGTGGGACTTAAGACCAGGTCCACGTACTAGACTATCCCATTTTACGACA
    ' ',
    ' ',
    GGGGCATACCCCAAAGGCATGAGACCAACCACAGCACTCGCATTTCACTACCTTGAGGCCATGGCCGAAGGATATACGGAAGTAAAAATCAACGTATAAACTGGG
    ' ',
    ' ',
    TAGATCTTCGTCTCCCATGTAAGCTGGCCCTCGAGCCAGATGCCGTATAAATGGGAGGACTGGAATCCAGATATGAGCTTCTTGCATCGCGCGGAGTTGTACG
    ' ',
    ' ',
    ATTTATTTCCGCTCCCGAAGAACCTTTTCTACAGTATTGTGCTGTCGTCACACTTGTTCGCGCAATTGGCGTAGGGTAATGCCAGCGGTTAAGGTTTCGTACTATA
    ' ',
    ' ',
    GTGAATCACAGGCCCGTCGGTGGTTGGTACCGACGAGTACATGCTAATGGAGCTTACGAGGTCCTCATCCGATACAGAGAGCTGAATTGCTAATCCCTCCTTCAG
    ' ',
    ' ',
    AGGCGGGGCGGGGATGCGGATCAAGGCCTTGAGAATAAACTGAAATTACATATCATTCTAGTGTGGAGAGCTGTGGCTAGTGACTGATCAGACATCAGCCGTTT
    ' ',
    ' ',
    TCTTGCCAGACCCGGCAAAGGCAAGCTATATGCTTTCCACTTTAGAACAAGGCGTTAGGATACCTATCAGTGCTTCAAGTAATTCATTGACCATTAGGTTATGTT
    ' ',
    ' ',
    TATTGGAAGGTGGGCTTCCTTTATTTCGAGCGCGTCGCATGGGTGCATCGATCTGGTAAGCCGCTTTCCAAAGGACCAGGCCAAGACTCCCTTTTGACATCGC
    ' ',
    ' ',
    ATCTATTCCCTTAGGCTTGGAACGTGGAGTCTCTCTATCGGTCATGCGGTGGCTCTCCGCTCAAGAACGAGTCGGGGGAATGCGCTGCGTCCGACGACTGTTACG
    ' ',
    ' ',
    AGCTTAACAAACCTCAGGTAGAATTACGGAGTTGATCAACTCCACTACTGTACAGCCCATAGGGGTGAAGAGTGCTCATGATTGCGGATCGTTAGCTAACTCGC
    ' ',
    ' ',
    GTGGCCAGGCGACTAAGAAAGGACACAGAAAAATTGCGCTCTACAGTATCTCAGCCGCCCACTAATATTATTAGACGAGCTTCATTGTCTTACAATATTTGATAT
    ' ',
    ' ',
    GCATTGACTCGATAACAATAGGTTGATGCATTACGCGCGGTCTATTGATCTTCATCCCGATAGTTATGTATGGACGTGTTCTTCTGTGAAAAAACTAGCTCTGTC
    ' ',
    ' ',
    GCCGTAAATACATACCAGCGGACTAGTTGCCAAGGCTGATAGGGCGGTTCATCCGCGTTGCTGCACTCCCTATCTCATGGAGCATGTGAATCTTGGCGTTAGTCA
    ' ',
    ' ',
    ' ',

```





```

1 3
2 CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCGGAAAGGATGAGGCGATAG
3 1
4 3
5 CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCGGAAAGGATGAGGCGATAG
6 1
7 3
8 CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCGGAAAGGATGAGGCGATAG
9 1
10 3
11 CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCGGAAAGGATGAGGCGATAG
12 1
13 3
14 CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCGGAAAGGATGAGGCGATAG
15 1
16 3
17 TGGACAAACGGTGTGCGTTAGAAAGCTGCGACCGCACCCCGTGTCAACTTAAGCATTGGCCATAGGCTCCCATCCGCCCATGTGTGGCATACCGCGCCATGTTATTG
18 1
19 3
20 TCGAAGTGTGTTTCAATCGTGTATAGTTTGGCATAACCATGTAAGGGCCTTCGAACCGACCCCACTCTTTCAGTCTGCGAGAAGCAAGAAGCCTTACTCGAAAG
21 1
22 3
23 GGCAGCGTGAGAGTCCTTGCTTACAAAATCAAGGGTAAAACATCCATCGAGAGAACGTTGCGGCTGGAGAGATCTCATTAAAGCGTTCAATTGAGGTTATAGAGGTG
24 1
25 3
26 TCTCCAGAGCGGAGTCCCCACTAGGAACCCGAAGTTTGGCAGGTACACACTCTCAGCACCCACCCCAACTAGGGCAGCCGATCCGCATAGGGGCACCAGATGATTC
27 1
28 3
29 TCCATGCCTGGTACATACATTAAAGTTGATAGTATTACGGTGGGCACCCAGGCAACACCCGATTGCTCTGCGGATACACACTTCTTGCCTGTATGGGAGAGAATCTG
30 1
31 3
32 CCCTGGAGAATTTAATAGTCGTTTTTATACCAGTGGTACAGGATCTCAGATTGTACCCCGTGTCCCAGTTTAAGGCTAAATCACATATTTCTGCGTACTTCACTAA
33 1
34 3
35 TTATAGTTAAATAGCGTATGATCAAGTTCCACAGGTCGATAGTTGGACTCGTATCGAGGTATCAGCATTGACGGCGCTAGTTAGTCTATAAATATTCGTTAGGAAAG
36 1
37 3
38 GTTTTGGTAAGGGAGTACCTATGATGTATTGTTCAAGAAGAGGGGCACAGTAATGGACGAAAGGGAGGTCTGTCAATCCAGTCCACTCCATAGAGAGGTTTATACT
39 1
40 3
41 CCAGGATGTTTCGATGATATCAGGCGCGAAGTATCGGAAGCGTCATATTAAGCACATATTTAATGGGGCATACCCCAAAGGCATGAGACCAACCACAGCACTCGCA
42 1
43 3
44 CATTGTTCTTCGGGACTAGTAGAGGTCGTTGTCGTTGCGAACACGTCAGTTGCAAGTGCTCAAGCTTCTACTAGATCTTCGTCTCCCATGTAAGCTGGCCCTCGAGC
45 1
46 3
47 CTGCTCTCACCCGCACGCCGATGTCAGGGCTGGAAGATGGCTATAACTCTTAGCAGACCTGGATGTGTCGTTACCGATTATTTCCGCTCCCGAAGAACCTTTCT
48 1
49 3
50 ACCATCAGACATACCGAAGCGCCGACTGACAGTTCATTATTTGACGGTGAACAGTACTGGACGCCAGTATTAATGCTGTATCATGTGAATCACAGGCCCGTCGTTG
51 1
52 3
53 AAGAGACCCCGCACGACACAAATGCGTAGCAAGCTGCTTATCCAGCAATGAGGTGTGGTGGTGAGCGGAGTTAAACACGGGACGCATCGGAGAGGCGGGGCGGGAG
54 1
55 3
56 TCGATATTATATTCGACTTTAATCCTAATGGACGTCCCAAGGTCAACCGAGTGGGACGACCGCCACCCCTCGTAAGGTGCCTACAGCACCCGACCCCTCTTGCCA
57 1
58 3
59 GCCTGTAAATTGCAAAATAAGCCATATTCTTGTGAATCGATGCCCTCATATCACATTATCTACTTCTCGTATGCTAATTGGCAGCGATGAGAGAACCCTCATAGT
60 1
61 3
62 AGGACCGCCGCAAAAATATGGCGCTCACACAAATCGCTGTCACTTGTGTATGACTTGCTTGTCCCGTTTCGTGCAGATTTTATGATACACTGAACGTCTATTAA
63 1
64 3
65 ATGGTAAAGGCTCAAAAAGTACCCGACTTTGGGATTAGACGCGAACCTGGGTACCAGCTATTAAGCATAGTTTTTCTGAATGGCCACGGGCGGGTGGGGAGTTT
66 1
67 3
68 CCGGACGATGGCGAGGTGAAGATAGTAAAGTAAGCTGTGGGCGTATATTTATTTAGATAGGACACGGGTAAGTGGTGACTATTGCGTTGTGCGTGTAAATTTCCG
69 1
70 3
71 ATACGAGGACGTATGGCCGTGATGTCGCTCGTAGCGATTGCTTATGGAGCTTACCCTTGATCATGGTACGCTGTACTTGATATAGAGAAGACTTTTGGACGGCGC
72 1
73 3
74 GTCCTAAGACAGAATAAAACCGAACGTTGATGCCGAGTGTAACTATAGATGTTTCTTATACCTGTGGAGCATGGACTTATAAGATTGATCTCTGGATGTACCG
75 1
76 3
77 CACTAATGATCTTTTGGACATTGAGAGTCAGTGCCGATTACCGACTGTCCGTGTACGCGTACTGCAACGGTGGGACTAAATACGGTACCTCAATTGACAGCGTGG
78 1
79 3
80 CCCTGGTTAGCCACTTTCAGACCGGGCATTATAGGACGGACTATCCGCCGAGAACAAGAGATTGTTATCCTTGCTCATACCACGGATGCACGCCCCCCCAGTCACT
81 1
82 3
83 AGTGCTAAACGCACAGGTGGTGAGGATTACTATAACCGCGTGCACACATGGCTAAGTAAGTCCCTCGATCCATAGTTTCACATCCACCGGACGTAATCCACGA
84 1
85 3
86 TGATACACAAACGTACCGTAGCCTCCCTGCGAAGAGACTCATGCATCAATGAACTAACAAAGTCGGCCTGCTATTCTGATATCTCATCCCGTCTCCCTTAACAGCA
87 1
88 3
89 GTAACCCGCTCCCGCAGTACTGCACGCGAGTGATACAGTACCAGTGAAAGTCTATCGCATGTGAGGTGCAGCTAACACCCCGAACCGAACTACGCGTACGCTCTC

```



CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCAATAG  
 CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCAATAG  
 TGGACAAACGGTGTGCGTTAGAAAGCTGCGACCGCACCCCGTGTCAACTTAAGCATTTGCCATAGGCTCCCATCCGCCCATGTGTGGCATACCGCGCCATGTTATTG  
 TCGAAGTGTTTGTTCATCGTGTATAGTTTGGCATAACCATGTAAGGGCCTTCGAACCGACCCCACTCTTTGCAGTCTGCGAGAAGCAAGAAGCCTTACTCGAAAG  
 GGCAGCGTGAGAGTCCTTGCTTACAAAATCAAGGGTAAACATCCATCGAGAGAACGTTGCGGCTGGAGAGATCTCATTAAAGCGTTCAATTGAGGTTATAGAGGTGA  
 TCTCCAGAGCGGAGTCCCCACTAGGAACCCGAAGTTTGGCAGGTACACACTCTCAGCACCCACCCCAACTAGGGCAGCCGATCCGCATAGGGGCACCAGATGATTCC  
 TCCATGCCTGGTACATACATTAAGTTGATAGTATTACGGTGGGCACCAGGCAACACCCGATTGCTCTGCGGATACACACTTCTTGCCTGTATGGGAGAGAATCTGT  
 CCCTGGAGAATTTAATAGTCGTTTTTATACCAGTGGTACAGGATCTCAGATTGTACCCCGTGTCCAGTTTAAGGCTAAATCACATATTTCTGCGTACTTCACTAAC  
 TTATAGTTAAATAGCGTATGATCAAGTTCCACAGGTCGATAGTTGGACTCGTATCGAGGTATCAGCATTGACGGCGCTAGTTAGTCTATAAATATTCGTTAGCAAA  
 GTTTTGTAAGGGAGTACCTATGATGTATTGTTCAAGAAGAGGGCAGTAATGGACGAAAGGGAGGTGATGTCAATCCAGTCCACTCCATAGAGAGGTTTATACTG  
 CCAGGATGTTTCGATGATATCAGGCGCGAAGTATCGGAAGCGTCATATTAAGCACATATTTAATGGGGCATACCCCAAAGGCATGAGACCAACCACAGCACTGCGAT  
 CATTGTTCTTCGGGACTAGTAGAGGTGTTGTCGTTGCGAACCGTGAGTTGCAAGTGCTCAAGCTTCACTAGATCTTCGTCTCCCATGTAAGCTGGCCCTCGAGCG  
 CTGCTCTCACCCGCACGCCGATGTCAGGGCTGGAAGATGGCTATAACTCTTAGCAGACCTGGATGTGTCGTTACCGATTATTTCCGCTCCCGAAGAACCTTTTC  
 ACCATCAGACATACCGAAGCGCGACTGACAGTTCATTATTTGACGGTGAACAGTACTGGAGCGCCAGTATTAATGCTGTATCATGTGAATCACAGGCCCGTCGCTG  
 AAGAGACCCCGCACGACACAAATGCGTAGCAAGCTGCTTATCCAGCAATGAGGTGTGGTGGTGAGCGGAGTTAACACGGGACGCATCGGAGAGGCGGGCGGGGAT  
 TCGATATTATATTCGACTTTAATCCTAATGGACGTCCCAAGGTACCCGAGTGCGGACGACCGCCACCCCTCGTAAGGTGCCTACAGACCCGACCCCTCTTCCCA  
 GCCTGTAAATTGCAATAAAGCCATATTCTTGTAATCGATGCCCTCATATCACATTATCTACTTCTCGTATGCTAATTGGCAGCGATGAGAGAACCTCATAGTA  
 AGGACCGCGCAAAATATGGCGCTCACACAAATCGCTGTCAATCATGTGTATGACTTGCTTGTCGCCGTTTCGTGCAGATTTTATGATACACTGAACGTCTATTAA  
 ATGGTAAAGGCTCAAAAAGTACCCGACTTTGGGATTAGACGCGAACCTGGGTACCAGCTATTAAGCATAGTTTTTCTGAATGGCCACGGGCGGGTGGGGAGTTT  
 CCGGACGATGGCGAGGTGAAGATAGTAAAGTAAGCTGTGGGCGTATATTTATTTAGATAGGACACGGGTAAGTGGTGACTATTGCGTTGTGCGTGTAAATTTGCG  
 ATACGAGGACGTATGGCCGTGATGTCGCTCGTAGCGATTGCTTATGGAGCTTACCCTTGATCATGGTACGCTGTACTTGATATAGAGAAGACTTTTTGACCGCGCG  
 GTCACTAAGACAGAATAAACCGAACGTTGATGCCGAGTGTAACGTATAGATGTTTATCTTATACCTGTGGAGCATGGACTTATAAGATTGATCTCTGGATGTACCG  
 CACTAATGATCTTTTGGACATTGAGAGTCAGTGCCGATTACCGACTGTCCGTGTACGCGTACTGCAACGGTGGGACTAAATACGGCTACCTCAATTGCGACGCTGG  
 CCCTGGTTAGCCACTTGACAGACGGGCATTATAGGACGGACTATCCGCCGAGAACAAGAGATTGTTATCCTTGCTCATACCACGGATGCACGCCCCCAGTCACT  
 AGTGCTAAACGCACAGGTGGTGAGGATTACTATAACCGCGGTGCGACACATGGCTAAGTAAGTCCCTCGATCCATAGTTTCACATCCACCCGGACGTAATCCACGA  
 TGATACACAAACGTACCGTAGCCTCCCTGCGAAGAGACTCATGCATCAATGAACATAACAAGTCGGCTGCTATTCTGATATCTCATCCGCTCTCCCTTAACAGCA  
 GTAACCCGCTCCCGCAGTACTGCACGCGAGTGATACAGTACCAGTGAAAGTCTATCGCATGTCAGGTGCAGCTAACACCCGAAACCGAACTACGCGTACGCTGCTC  
 ATTGCTCCGTCTCCGTAAAGATCTAAGCTTTGTATGTTGCGTCAATCCCACCTCTGGCGTGCTTGGTTGATTTTGTCTACAGCTCTGGCCAATCGTTGTGCAAC  
 AAGCACGCTCTAGATAATGTTTGTAACCGCCTCTCGTCAGCATATTTTTCACAGAGCGGATCTGTGTAAGCTCCGGCTCGGTGCTGTTGAGTCGGATTTTGGCTT  
 GCGTCTTACGCTAGGTGCCGTAACGCGGTTTCGGCCAGGACAGCGTTGGCCCCAGTGGCCCCACAAGTGAGAGCGCGTGGATAAAAAACGAGGACATAACGCGCTC  
 GATCGTGTCCAGTATGTCAAATAACTAAGGTACGAAACGAGTGCCTGGTTGCAACCACCGCCAGATCCGATCATTGGTAAAAATCCAGACAGTCCCGGCCACTT

```

GTACAGATCCCTTAAAGAATGATTGTGTACCTGATAAGGACCTTTAACTGATGCGGACACGTTTGCCAGCCCCGTACTCCGTTTACGGACGCGACATGCCGCAATC
',
',
GTCTGTCGCGTCAAGACCTAACCAAACTCTACTTTTCCCTCCTGTTATAAACTGATTGGTCAAGATTGCGCTAATCCGACGGAGGGGCAAAATCGAGCACAGTTA
',
',
GCACCTATAGTATCATCATTAGTCGGGCTGGGTGTTTAGCGAGTCATTAGAGACGGACGTTTGAATGTGCCCCGAGATTCCGGTCTAATATAACAGCCATTTCCGGC
',
',
CTAAGCGAAGACCTTTGCCCCAAGGGGCACTATGTGACTCACAACCTCCTCCTACAACGAATCAGCGCCAGCGCGAAGGTTGCAAGGGTAGACGTGCGTACAT
',
',
GTACCCACTGTCCTCCAATCCTGGACATCCTTGCAGACAGCAACCATCGCAAATAGTAATAGTGATCGTCTTATTACTTGGGGTAACTCCAACGGAGTGAGTCG
',
',
CTGTTTGGCAGTCCGCTGTCCGTACGTTGCATGGTAGAGGCTAGGGCCGGATATATACCTCCGCTTACTCATATGCAGAAGTTTAAAGAAATGCGTTACACCTT
',
',
ATCACGGTACCTGGGCCTAGAACCCTCCTTGACGCGAGGATCGAGCTGACGTGCGCACACGGTCGGTTAGCTTCTCTGTAGTCAGTGCCGTGGAGTAACCGTCTT
',
',
CGTTCGCGATTAGAGGTCCGAAAGGGTCCGAATAAATCCTCATCGACCAATGACCACCGCGGTAAACGGTCAACGGTGCAGTGAACCAACAGAGCCGTGT
',
',
AGCTGGCACAAGAAGAAGTCTCTTCAATGTCAAGTGTGACGGACTCGCTTTGCTCTAGTCGACCACAAAGCCCATACAATAATTATGCAAGGTCTCGCCATTG
',
',
ACGCCGTGTTTTACGTAACCAGGACGGGTAATGCCTCACAGTTATTCTCTATCCAACCTAGGATCAATTAAGGCGCTAAGTTGTCGAGGAAGGGGCTCAATCAG
'], 9)

112
113
114 # In[15]:
115
116
117 #De Bruijn Graph for pool-B for k=18
118 construct_graph(['
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCGAAAGGATGAGGCAAT
',
',
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCGAAAGGATGAGGCAAT
',
',
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCGAAAGGATGAGGCAAT
',
',
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCGAAAGGATGAGGCAAT
',
',
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCGAAAGGATGAGGCAAT
',
',
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCGAAAGGATGAGGCAAT
',
',
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCGAAAGGATGAGGCAAT
',
',
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCGAAAGGATGAGGCAAT
',
',
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCGAAAGGATGAGGCAAT
',
',
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCGAAAGGATGAGGCAAT
',
',
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCGAAAGGATGAGGCAAT
',
',
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCGAAAGGATGAGGCAAT
',
',
TGGACAAACGGTGTGCGTTAGAAAGCTGCGACCGCACCCCGTGTCAACTTAAGCATTGCCATAGGCTCCATCCGCCCATGTGTGGCATACCGCGCCATGTTATT
',
',
TCGAAGTGTTTTGTTTCAATCGTGATAGTTTGGCATAACCATGTAAGGGCCTTCGAACCGACCCCACTCTTTGCAGTCTGCGAGAAGCAAGAAGCCTTACTCGAA
',
',
GGCAGCGTGAGAGTCCTTGCTTACAAAATCAAGGGTAAACATCCATCGAGAGAACGTTGCGGCTGGAGAGATCTCATTAAAGCGTTCAATTGAGGTTATAGAGTGA

```

TCTCCAGAGCGGAGTCCCCACTAGGAACCCGAAGTTTGGCAGGTACACACTCTCAGCACCCACCCCAACTAGGGCAGCCGATCCGCATAGGGGCACCAGATGATTCC  
 TCCATGCCTGGTACATACATTAAAGTTGATAGTATTACGGTGGGCACCAGGCAACACCCGATTGCTCTGCGGATACACACTTCTTGCGTTGTATGGGAGAGAATCTGT  
 CCCTGGAGAATTTAATAGTCGTTTTTATACCAGTGGTACAGGATCTCAGATTGTACCCCGTGTCCCAGTTTAAGGCTAAATCACATATTTCTGCGTACTTCACTAAC  
 TTATAGTTAAATAGCGTATGATCAAGTTCCACAGGTCGATAGTTGGACTCGTATCGAGGTATCAGCATTGACGGCGCTAGTTAGTCTATAAATATTCGTTAGGAAAT  
 GTTTTGGTAAGGGAGTACCTATGATGTATTGTTCAAGAAGAGGGGCACAGTAATGGACGAAAGGGAGGTTCATGTCAATCCAGTCCACTCCATAGAGAGGTTTATACT  
 CCAGGATGTTTCGATGATATCAGGCGCGAAGTATCGGAAGCGTCATATTAAGCACATATTTAATGGGGCATACCCCAAGGCATGAGACCAACCACAGCACTCGCA  
 CATTGTTCTTCGGGACTAGTAGAGGTCGTTGTCTGTTGCGAACAGTGAGTTGCAAGTGCTCAAGCTTCACTAGATCTTCGTCTCCCATGTAAGCTGGCCCTCGAGC  
 CTGCTCTCACCCGCACGCCGATGTCAGGGCTGGAAGATGGCTATAACTCTTAGCAGACCTGGATGTGTCGTTACCGATTATTTCCGCTCCCGAAGAACCTTTCT  
 ACCATCAGACATACCGAAGCGCCGACTGACAGTTCATTATTTGACGGTGAACAGTACTGGACGCCAGTATTAATGCTGTATCATGTGAATCACAGGCCCGTCGTTG  
 AAGAGACCCCGCACGACACAAATGCGTAGCAAGCTGCTTATCCAGCAATGAGGTGTGGTGGTGAGCGGAGTTAACACGGGACGCATCGGAGAGGCGGGGCGGGGAT  
 TCGATATTATATTCGACTTTAATCCTAATGGACGTCCCAAGGTACCCGAGTGCGGACGACGCCACCCCTCGTAAGGTGCCTACAGCACCCGACCCCTCTTCCCA  
 GCCTGTAAATTGCAAAATAAGCCATATTCTTGTAATCGATGCCCTCATATCACATTATCTACTTCTCGTATGCTAATTGGCAGCGATGAGAGAACCCTCATAGT  
 AGGACCGCCGCAAAAATATGGCGCTCACACAAATCGCTGTCTATTGATGTGTATGACTTGCTTGTCGCCGTTTCGTGCAGATTTTATGATACACTGAACGTCTATTAA  
 ATGGTAAAGGCTCAAAAAGTACCCGACTTTGGGATTAGACGCGAACCTGGGTACCAGCTATTAAGCATAGTTTTTCTGAATGGCCACGGGCGGGTGGGGGAGTTT  
 CCGGACGATGGCGAGGTGAAGATAGTAAAGTAAGCTGTGGCGTATATTTATTTAGATAGGACACGGGTAAGTGGTGACTATTGCGTTGTGCGTCGTAAATTTCCG  
 ATACGAGGACGTATGGCCGTGATGTCGCTCGTAGCGATTGCTTATGGAGCTTACCCTTGATCATGGTACGCTGTACTTGATATAGAGAAGACTTTTTGACCGCGC  
 GTCACTAAGACAGAATAAAACCGAAGTTGATGCCGAGTGTAACTGATAGATGTTTATCTTATACCTGTGGAGCATGGACTTATAAGATTGATCTCTGGATGTACCG  
 CACTAATGATCTTTTGACATTGAGAGTCAGTGCCGATTACCGACTGTCCGTGTACGCGTACTGCAACGGTGGGACTAAATACGGCTACCTCAATTCGCACGCTGG  
 CCCTGGTTAGCCACTTGACAGACCGGGCATTATAGGACGGACTATCCGCCAGAACAAGAGATTGTTATCCTTGCTCATACCACGGATGCACGCCCCCCCAGTCACT  
 AGTGCTAAACGCACAGGTGGTGAGGATTACTATAACCGCGGTGCGACACATGGCTAAGTAAGTCCCTCGATCCATAGTTTCACATCCACCCGGACGTAATCCACGA  
 TGATACACAAACGTACCGTAGCCTCCCTGCGAAGAGACTCATGCATCAATGAACAAAGTCGGCCTGCTATTCTGATATCTCATCCCGTCTCCCTTAACAGCAT  
 GTAACCCGCTCCCGCAGTACTGCACGCGAGTGATACAGTACCAGTGAAAGTCTATCGCATGTCAGGTGCAGCTAACACCCCGAACCGAACTACGCGTACGCTGCTC  
 ATTGGCTCCGCTCCTGTAAGATCTAAGCTTTGTATGTTGCGGTCAATCCACCTCTGGCGTGCTTGGTTGATTTTGTCTACAGCTCTGGCCAAATCGTTGTGCAAC  
 AAGCACGCTCTAGATAATGTTTGTAACCGCCTCTCGTCAGCATATTTTTCACAGAGCGGATCTGTGTAAGCTCCGGCTCGGTGCTGTTGAGTCGGATTTTGCTT  
 GCGTCTTACGCTAGGTGCCGTAACGCGGTTTCGGCCAGGACAGCGTTGGCCCCAGTGGCCCCACAAGTGAGAGCGCGTGATAAAAACAGGACATAACGCGCCT  
 GATCGTGTCCAGTATGTCAAATACTAAGGTACGAAACGAGTGCCTGGTTGCAACCACCGCCAGATCCGATCATTGGTAAAAATCCAGACAGTCCCGGCCACTT  
 GTACAGATCCCTTAAAGAATGATTGTGTACCTGATAAGGACCTTTAACTGATGCGGACACGTTTGCCAGCCCCGTACTCCGTTTACGGACGCGACATGCCGCAAT  
 GTCTGTGCGGTCAAGACCTAACCAAACTCTACTTTTCCCTCCTGTTATAACACTGATTGGTCAAGATTGCGCTAATCCGACGGAGGGGCAAAATCGAGCACAGTT  
 GCACCTATAGTATCATCATTAGTCGGGTGGGTGTTTAGCGAGTCATTAGAGACGGACGTTGCAATGTGCGCCGAGATTCCGGTCTAATATAACAGCCATTTTCGGC  
 CTAAGCGAAGACCTTTGCCCCAAGGGGCACTATGTGACTCACAACCTACCTCCTCTACAACGAATCAGCGCCAGCGGAAGGTTGCAAGGGTAGACGTGCGTACAT  
 GTACCCACTGTCCTCCAAATCCTGGACATCCTTGACAGACGAACCATCGCAAATAGTAATAGTGATCGTCTTATTACTTGGGGTAAACTCCAACGGAGTGAGTCG



[illegible]



```

CGGATGCACGCCCCCAGTCACCTTTACATCTTACGCCGAACCTGGACGTGAAGGGATGAATAAGGAACGGCCGGTCACCTAGTGCATAAGTTATAGACGATTGTTC
',
',
AGTTTCACATCCACCCGGACGTAATCCACGAGAAAAAGGGCAATCTTTATCACGAAATTCAGCGGAAGCCGTCGAGGTATCAGGTGTAGCCATTGTAGCGTCTAAC
',
',
GCTATTCTGATATCTCATCCCGTCTCCCTTAACAGCATAGGGGCTGACTAGACACCGCGTGTGGCCGACTTCATCGCGATCGGAATAGATATAGTTAATAAACCA
',
',
TCAGGTGCAGCTAACACCCCGAACCGAACTACGCGTACGCTGCTCGCCGAATTCATTTCGTTAAAAAGCACGGGCAGACGGTTAAAGGCTGTTGTATAACCTAGTTC
',
',
CTGGCGTGTCTGGTTGATTTTTGCTCTACAGCTCTGGCCAATCGTTGTGCAAGTAGATGACGACTATTGCCTATGGTATGTTATATGCCATTCGGATCACCGGTCTC
',
',
TTTCACAGAGCGGATCTGTGTAAGCTCCGGCTCGGTGCTGTGAGTCGGATTTTGGCTTCACATTTAACTTTCGTCCTCTAAAGTCACTACAAAACATGCTGGTTC
',
',
CAGCGTTGGCCCCAGTGGCCCCACAAGTGAGAGCGCGTGGATAAAAAACCAGGACATAACGCGCCTGGCCGTTCCAGGGTCCAAACTGCTGTTCCCCGTCACG3CC
',
',
ACGAAACGAGTGCCTGGTTGCAACCACCGCCAGATCCGATCATTGGTAAAAATCCAGACAGTCCGGGCCACTTATACTGGCAAGAATCTACTCCAAAACGCCACGG
',
',
GTACCTGATAAGGACCTTTAACTGATGCGGACACGTTTGCCAGCCCCGTA CTCCGTTTACGGACGCGACATGCCGCAATCCACGGGTGCTGAGAGCCTTTCCGAC
',
',
AACCAAACTCTACTTTTCCCTCCTGTTATAAACTGATTGGTCAAGATTCCGCTAATCCGACGGAGGGGCAAAATCGAGCACAGTTAATATAGAAAGGTCTT3TC
',
',
TCATCATTAGTCGGGCTGGGTGTTTAGCGAGTCATTAGAGACGGACGTTTGAATGTGCCCGCAGATTCGGGTCTAATATAACAGCCATTCGGCGATGCCACAGTGT
',
',
CGAAGACCTTTTCGCCCCAAGGGGCACTATGTGACTCACAACCTACCTCCTCTACAACGAATCAGCGCCAGCGCGAAGGTTGCAAGGGTAGACGTGCGTACATACAG
',
',
GCGTACCCACTGTCTCCAAATCCTGGACATCCTTGACAGACGCAACCATCGCAAATAGTAATAGTGATCGTCTTATTACTTGGGGTAACTCCAACGGAGTCAAGT
',
',
CGGATTCTTCTGTTTGGCAGTCCGCTGTCCGTACGCTTGCATGGTAGAGGCTAGGGCCGGATATATACCTCCGCTTACTCATATGCAGAAGTTTAAAGAAATGCG3TT
',
',
AGCCTACATCGATTGCATCACGGTACCTGGGCCTAGAACCCTCCTTGACGCGAGGATCGAGCTGACGTGCGGCACACGGTCCGTTAGCTTCTCTGTAGTCAGT3CC
',
',
TACCTTCTCGACCCAAGCTCCATCGTTCCGCATTAGAGGTCCGAAAGGGTCCGAATAAATCCTCATGACCCAATGACCACCGCGGTAACCGGTACAGGGTGC3GT
',
',
ATACCCACCCTTATTCCAAAACAGTCTCAGCTGGCACAAGAAGAAGTCTCTTCAATGTCAAGTGTGACGGACTCGCTTTGCTCTAGTCGACCACAAAGCCCAT
'],5)

```

126

127

128 # In[17]:

129

130

131 #De Bruijn Graph for pool-C for k=9

132 construct\_graph(['

```

CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCGAAAGGATGAGGCAAT
',
',
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCGAAAGGATGAGGCAAT
',
',
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCGAAAGGATGAGGCAAT
',
',
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCGAAAGGATGAGGCAAT
',
',
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCGAAAGGATGAGGCAAT
',
',
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCGAAAGGATGAGGCAAT
',
',
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCGAAAGGATGAGGCAAT
',
',
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCGAAAGGATGAGGCAAT
',
',
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCGAAAGGATGAGGCAAT
',
',
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCGAAAGGATGAGGCAAT
',
',
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCGAAAGGATGAGGCAAT

```

```

| 3 |
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCGGAAAGGATGAGGCGATAG
| 3 |
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCGGAAAGGATGAGGCGATAG
| 3 |
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCGGAAAGGATGAGGCGATAG
| 3 |
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCGGAAAGGATGAGGCGATAG
| 3 |
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCGGAAAGGATGAGGCGATAG
| 3 |
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCGGAAAGGATGAGGCGATAG
| 3 |
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCGGAAAGGATGAGGCGATAG
| 3 |
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCGGAAAGGATGAGGCGATAG
| 3 |
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCGGAAAGGATGAGGCGATAG
| 3 |
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCGGAAAGGATGAGGCGATAG
| 3 |
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCGGAAAGGATGAGGCGATAG
| 3 |
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCGGAAAGGATGAGGCGATAG
| 3 |
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCGGAAAGGATGAGGCGATAG
| 3 |
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCGGAAAGGATGAGGCGATAG
| 3 |
CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCGGAAAGGATGAGGCGATAG
| 3 |
GGCTACATGGACAAACGGTGTGCGTTAGAAAGCTGCGACCGCACCCCGTGTCAACTTAAGCATTGGCCATAGGCTCCCATCCGCCCATGTGTGGCATACCGCGCCAT
| 3 |
AGCCATACTTACCATCGAAGTGTGTTTCAATCGTGTATAGTTTGGCATAACCATGTAAGGGCCTTCGAACCGACCCCACTCTTTGCAGTCTGCGAGAAGCAAGAA
| 3 |
CTCACCAACGCACGGCTTATTGGCAGCGTGAGAGTCCTTGCTTACAAAATCAAGGGTAAACATCCATCGAGAGAACGTTGCGGCTGGAGAGATCTCATTAAAGCGT
| 3 |
GATTCGGAGGCATTCCCTACACTTTCCTCTCCAGAGCGGAGTCCCCACTAGGAACCCGAAGTTTGGCAGGTACACACTCTCAGCACCCACCCCACTAGGGGAGCG
| 3 |
GTAGGCTACGCGTCTTTATGGTCTAATTGCTGACATCCATGCCTGGTACATACATTAAGTTGATAGTATTACGGTGGGCACCAGGCAACACCCGATTGCTCTCGGA
| 3 |
TTTGACCTCGAGAACCGCACACAGGTTGGTGGATTCTGGCGCCCTGGAGAATTTAATAGTCGTTTTTATACCAGTGGTACAGGATCTCAGATTGTACCCCGTGTCTC
| 3 |
CGGAGGTCCAGAAGATCGTTATAGCATGCTCCAGCGAGATGGAAGACGTTTATAGTTAAATAGCGTATGATCAAGTTCACAGGTCGATAGTTGGACTCGTATCGA
| 3 |
ATTCATAGTGGAGAGTATGGCCCGTCGGCATTCTCGGCGGGTTAACCATACCTCGTTTTTGGTAAGGGAGTACCTATGATGTATTGTTCAAGAAGAGGGCACAGT
| 3 |
GTCGCTAGCAAATCCTATTAGGGGCACCTAGCGGGCACCGAGATTCGAGGTGAAGGGTGGCCCCAGGATGTTTCGATGATATCAGGCGCGAAGTATCGGAAGCGTCA
| 3 |
ACGGCGCTTTCTAGGATAAAGTACGATGTTCTGTAAGATGTTATGTCTTAACAAGACCCGAAGGCAGAGTCATTGTTCTTCGGGACTAGTAGAGTCGTTGTCTTC
| 3 |
GAGTTGTACGCCTCTATGCCCGCGCATGCAGAAATGCCACCCCCACCACTCGGCGGTGCGATCTGCTGCTATTGTGGACTGCTCTCACCCGCGACGCGGATGTGAGG

```

```

133
134
135 # In[18]:
136
137
CAGCGGTTAAGGGTTCGTACTATAGTCGTCTTATGAAGTTTGTACGGATTCCGAGCCTCTCCCCGATCATAAAGGAATGATTTTACCATCAGACATACCGAAGCGG
CCGATACAGAGAGCTGAATTGCTAATCCCCTCCTTCAGTATTGGTCGTGTATCAACCAAGCCCGCGGAGGAGCACTCGCAACAACACTTGAAGAGACCCCGOACG
CATTCTAGTGTGGAGAGCTGTGGCTAGTGACTGATCAGACATCAGCCGTTTCGAAAGAGGACCTAATGACAGATTAACATAACAAGTACGCCCTAAGTCGATATT
TTTAGAACAAAGGCGTTAGGATACCTATCAGTGCTTCAAGTAATTCATTGACCATTAGGTTATGTTGTTACGCGGATTATGTAAGCCGCTCAGTACGCTCATAAGG
TCGCAGCGCGTCGCATGGGTGCATCGATCTGGTAAGCCGCTTTCGAAAGGACCACGGCCAAGACTCCCCTTTTGACATCGGCATGTGCTGCTCATGTCTTCAATTG
AGGCTTGGAACGTGGAGTCTCTCTATCGGTCATGCGGTGGCTCTCCGCTCAAGAACGCAGTCGGGGGAATGCGCTGCGTCCGACGACTGTTACGCGGTTCCGGGTT
CCAGCTTAAACAACCTCAGGTAGAATTACGGAGTTGATCAACTCCACTACTGTACAGCCCATAAGGGGTGAAGAGTGGCTCATGATTGCGGATCGTTAGCTAACTC
TCGCGATTGGGGTCCGGTGGCCAGGCGACTAAGAAAAGGACACAGAAAAATTGCGCTCTACAGTATCTCAGCCGCCCACTAATATTATTAGACGAGCTTCATTGCTT
CGCGATTGCGCCATATGGTGATATCGCTTGGGCATTGACTCGATAACAATAGGTTGATGCATTCACGCGCGGTCTATTGATCTTCATCCCGATAGTTATGTATGAC
TGGATGTACCGGTTGCTTTGACAAAAGCTTGGCTTTATGGCTTCGCCGTAATACATACCAGCGGACTAGTTGCCAAGGCTGATAGGGCGGTCTCCGCGTTGCTG
CTCAATTCGCACGCTGGGAACATCAAGAAGCCTGTGCTGTGCCGAGAGTGATCTATGTCTACGTAAGAAGGTCACTATCCCTGGTGTGTACAGGCAATTGATACGGA
CGGATGCACGCCCCCAGTCACTTACATCTTACGCCGAACCTGGACGTGAAGGGATGAATAAGGAACGCGCCGTCACCTAGTGCATAAGTTATAGACGATTTGTT
AGTTTCACATCCACCCGGACGTAATCCACGAGAAAAAGGGCAATTCTTATCACGAAATTCAGCGGAAGCCGTCGAGGTATCAGGTGTAGCCATTGTAGCGTCTAAC
GCTATTCTGATATCTCATCCCGTCTCCCTTAACAGCATAGGGGCTGACTAGACACCCGCTGTGGCCGACTTCATCGCGATCGGAATAGATATAGTTAATAAACCA
TCAGGTGCAGCTAACACCCCGAACCGAACCTACGCGTACGCTGCTCGCCGCAATTCATTGTTAAAAAGCACGGGCAGACGGTTAAAGGCTGTTGTATAACCTAGTT
CTGGCGTGCTTGTTGATTTTTGTCTCTACAGCTCTGGCCAATCGTTGTGCAAGTAGATGACGACTATTGCCTATGGTATGTTATATGCCATTCCGATCACCGGTCT
TTTCACAGAGCGGATCTGTGTAAGCTCCGGCTCGGTGCTGTTGAGTCGGATTTTGGCTTCACATTTAACTTTCGTCTCTAAAGTCACTACAAAACATGCTGTTCC
CAGCGTTGGCCCCAGTGGCCCCACAAGTGAGAGCGCGTGGATAAAAAACAGGACATAACGCGCCTGGCCGTTCCAGGGTCCAAACTGCTGTTCCCCGTACAGGCC
ACGAAACGAGTGCCTGGTTGCAACCACCGCCAGATCCGATCATTGGTAAAAATCCAGACAGTCCCGGCCCACTTATACTGGCAAGAATCTACTCCAAAACGCCACGG
GTACCTGATAAGGACCTTTAACTGATGCGGACACGTTTGCCAGCCCCGTACTCCGTTTACGGACGCGACATGCCGCAATCCACGGGTGCTGAGAGCCTTTCCGAC
AACCAAAACTCTACTTTTCCCTCCTGTTATAAACTGATTGGTCAAGATTTCGCCTAATCCGACGGAGGGGCAAAATCGAGCACAGTTAATATAGAAAGGTCTTGT
TCATCATTAGTCGGGCTGGGTGTTTAGCGAGTCATTAGAGACGGACGTTTCAAGTGTGCCCCGAGATTTCGGGTCTAATATAACAGCCATTTCCGGCGATGCCACAGT
CGAAGACCTTTGCCCCAAGGGGCACATGTGACTCACAACCTCCTCCTTACAACGAATCAGCGCCAGCGCGAAGGTTGCAAGGGTAGACGTGCGTACATACAG
GCGTACCCACTGTCCTCCAATCCTGGACATCCTTGACAGACAGCAACCATCGCAAAATAGTAATAGTGATCGTCTTATTACTTGGGGTAAACTCCAACGGAGTGAGT
CGGATTCTTCTGTTTGGCAGTCCGCTGTCCGTACGCTTGCATGGTAGAGGCTAGGGCCGGATATATACCTCCGCTTACTCATATGCAGAAGTTTAAACGAATGCGTT
AGCTACATCGATTGCATCACGGTACCTGGGCCTAGAACCTCCTTGACGCGAGGATCGAGCTGACGTGCGCACACGGTCGGTTAGCTTCCTCTGTAGTCAGTCCG
TACCTTCTCGACCCAAGCTCCATCGTTCCGCATTAGAGGTCGAAAGGGTCCGAATAAATCCTCATCGACCCAATGACCACCGCGGTAACCGGTACGGGTGCGGT
ATACCCACCCTTATTCCAAAACAGTCTCAGCTGGCACAAGAAGAAGTCTCTTCAATGTCAAGTGTGACGGACTCGCTTTGCTCTAGTCGACCACAAAGCCCA
'],9)

```

```
139 construct_graph([
```

[illegible]

CAGAGGTAATGAGAAGCCTATCCGTGGATGCTACGTTACGGCATACTTGTATCCGCTAGATAGTGAGACAACCCAAACATTGGATATTCCGAAAGGATGAGGCAATAG  
 GGCTACATGGACAAACGGTGTGCGTTAGAAAGCTGCGACCGCACCCCGTGTCAACTTAAGCATTGTCCATAGGCTCCCATCCGCCCATGTGTGGCATAACCGGCCAT  
 AGCCATACTTACCATCGAAGTGTGTTTCAATCGTGATAGTTTGGCATAACCATGTAAGGGCCTTCGAACCGACCCCACTCTTTGCAGTCTGCGAGAAGCAAGAT  
 CTCACCAACGCACGGCTTATTGGCAGCGTGAGAGTCCTTGCTTACAAAATCAAGGGTAAACATCCATCGAGAGAACGTTGCGGCTGGAGAGATCTCATTAAACCGT  
 GATTCGGAGGCATTCCCTACACTTTCCTCTCCAGAGCGGAGTCCCCACTAGGAACCCGAAGTTTGGCAGGTACACACTCTCAGCACCCACCCCACTAGGGCAGCG  
 GTAGGCTACGCGTCTTTATGGTCTAATTGCTGACATCCATGCTGGTACATACATTAAGTTGATAGTATTACGGTGGGCACCAGGCAACACCCGATTGCTCTCGGA  
 TTTGACCTCGAGAACCGCACACAGTTGGTGGATTCTGGCGCCCTGGAGAATTTAATAGTCGTTTTTATACCAGTGGTACAGGATCTCAGATTGTACCCCGTGTCTC  
 CGGAGGTCCAGAAGATCGTTATAGCATGCTCCAGCGAGATGGAAGACGTTTATAGTTAAATAGCGTATGATCAAGTTCACAGGTCGATAGTTGGACTCGTATCGAG  
 ATTCATAGTGGAGAGTATGGCCCGTCGGCATTCTCGGCGGGTTAACCATACCTCGTTTTTGGTAAGGGAGTACCTATGATGTATTGTTCAAGAAGAGGGCACAGT  
 GTCGCTTAGCAAATCCTATTAGGGGCACCTAGCGGGCACCGAGATTGAGGTGAAGGGTGGCCCCAGGATGTTTCGATGATATCAGGCGCGAAGTATCGGAAGCGTCA  
 ACGGCGCTTTCTAGGATAAAGTACGATGTTCTGTAAGATGTTATGTCTTAACAAGACCCGAAGGCAGAGTCATTGTTCTTCGGGACTAGTAGAGGTGTTGTCTTC  
 GAGTTGTACGCCTCTATGCCCAGCATGCAGAAATGCCACCCCACTCGGCGGTGATCTGCTGCTATTGTGGACTGCTCTCACCCGACAGCCGGATGTGAGG  
 CAGCGGTTAAGGGTTGCTACTATAGTCGCTTATGAAGTTGTACGGATTCCGAGCCTCTCCCGATCATAAAGGAATGATTTTACCATCAGACATACCGAAGCGCG  
 CCGATACAGAGAGCTGAATTGCTAATCCCTCCTTCAGTATTGGTCGTGTATCAACCAAGCCCGCGGAGGAGCACTCGCAACAACACTTGAAGAGACCCCGCAGCA  
 CATTTCTAGTGTGGAGAGCTGTGGCTAGTGAAGTATCAGACATCAGCCGTTTCGAAAGAGGACCTAATGACAGATTAATAAACAAGTACGCCCTAAGTCGATATT  
 TTTAGAACAAAGCGTTAGGATACCTATCAGTGCTTCAAGTAATTCATTGACCATTAGGTTATGTTGTTACGCGGATTATGTAAGCCGCTCAGTACGCTCATAAGG  
 TCGCAGCGCGTCGCATGGGTGCATCGATCTGTAAGCCGCTTTCCAAAGGACCACGGCCAAGACTCCCTTTTGACATCGGCATGTGCTGCTCATGTCTTCAATTG  
 AGGCTTGAAACGTGGAGTCTCTCTATCGGTCTATGCGGTGCTCTCCGCTCAAGAACGCAGTCGGGGGAATGCGCTGCGTCCGACGACTGTTTCAGCGGTTCCGGGTT  
 CCAGCTTAACAAACCTCAGGTAGAATTACGGAGTTGATCAACTCCACTACTGTACAGCCATAAGGGGTGAAGAGTGGCTCATGATTGCGGATCGTTAGCTAACTC  
 TCGCGATTGGGTGCGGTGGCCAGGCGACTAAGAAAGGACACAGAAAAATTGCGCTCTACAGTATCTCAGCCGCCCACTAATATTATTAGACGAGCTTCATTGCT  
 CGCGATTGCCATATGGTGATATCGCTTGGGCATTGACTCGATAACAATAGGTTGATGCATTACGCGCGGTCTATTGATCTTCATCCCGATAGTTATGTATGAC  
 TGGATGTACCGTTGCTTTGACAAAAGCTTGGCTTTATGGCTTCGCGTAAATACATACCAGCGGACTAGTTGCCAAGGCTGATAGGGCGGTCTATCCGCGTTCTG  
 CTCAATTGCGACGCTGGGAACATCAAGAAGCCTGTCGTGTGCCGAGAGTGATCTATGTCTACGTAAGAAGGTCACCTATCCCTGGTGTGTCAGGCAATTGATAGCGA  
 CGGATGCACGCCCCCAGTCACCTTACATCTTACGCCGAAGTGGACGTGAAGGGATGAATAAGGAACGCGCGGTACCTAGTGCATAAGTTATAGACGATTGTTT  
 AGTTTCACATCCACCCGACGTAATCCACGAGAAAAAGGGCAATTCCTATCACGAAATTCAGCGGAAGCCGTCGAGGTATCAGGTGTAGCCATTGTAGCGTCTAAC  
 GCTATTCTGATATCTCATCCCGTCTCCTTAAACAGCATAGGGGCTGACTAGACACCGCGTGTGGCCGACTTCATCGCGATCGGAATAGATATAGTTAATAAACCA  
 TCAGGTGCAGCTAACACCCCGAACCGAACTACGCGTACGCTGCTCGCCGAATTCATTGTTAAAAAGCACGGGCAGACGGTTAAAGGCTGTTGTATAACCTAGTT  
 CTGGCGTGCTTGGTTGATTTTTGCTCTACAGCTCTGGCCAATCGTTGTGCAAGTAGATGACGACTATTGCTATGGTATGTTATATGCCATTCCGATCACCGGTC  
 TTTACAGAGCGGATCTGTGTAAGCTCCGGCTCGGTGCTGTTGAGTCGATTTTGGCTTACATTTAACTTTTCGTCCTCTAAAGTCACTACAAAACATGCTGGTTCC  
 CAGCGTTGGCCCCAGTGGCCCCACAAGTGAGAGCGGTGGATAAAAAACGAGGACATAACGCGCCTGGCCGTTCCAGGGTCCAAACTGCTGTTCCCGTTCACGCGCA  
 ACGAAACGAGTGCTTGGTTGCAACCACCGCCAGATCCGATCATTGGTAAAAATCCAGACAGTCCCGGCCCACTTATACTGGCAAGAATCTACTCCAAACGCCACGG

```
142 # In[ ]:
```