CS4023 Artificial Intelligence

Name: Pavithra Rajan

Roll Number: B190632CS

1. **Heuristics to solve the Caesar Cipher**
    - Utilise the **frequency distribution** of each alphabet in the English language and calculate the score obtained for each alphabet in the ciphertext and each iteration of all possibilities. The iteration with the **maximum** score is chosen.
    - Use the **bigram frequency** distribution of the English language. A bigram is a pair of letters. 'Th' is the most common bigram. Similar to the alphabet frequency heuristic, we can compute the score and choose the iteration with the **maximum** score.
    - As there are only limited alphabets in the English language (26 + 1 for spaces in sentences), we can try out all the possibilities and choose the one that resembles a valid text.

The `cipher.py` consists of the implementation of the aforementioned heuristics. The help menu can be obtained by running the `python` file with `-h` flag.

```
PS C:\Users\Pavithra\Desktop\Caesar-Cipher-Cracking> & C:/Users/Pavithra/AppData/Local/Programs/Python/Python38/
python.exe c:/Users/Pavithra/Desktop/Caesar-Cipher-Cracking/cipher.py -h
usage: cipher.py [-h] -f F -c C [-n N]

optional arguments:
  -h, --help  show this help message and exit
  -f F        Name of the file containing the plaintext to be encrypted
  -c C        The Caesar rotation factor
  -n N        Name to be encoded
```

While running the python file, provide the path to the message to be encrypted via the `-f` flag, the rotation factor with the `-n` flag, and the name to be encrypted following the `-p` flag.

I will provide a rotation factor of 4, and my name 'pavithra'.

Note: In order to correctly decrypt the name via frequency-based heuristics, it is necessary to have a sufficiently long text. To do this, I will concatenate a message along with the name and then extract it in the end after decryption.

```
PS C:\Users\Pavithra\Desktop\Caesar-Cipher-Cracking> & C:/Users/Pavithra/AppData/Local/Programs/Python/Python38/
python.exe c:/Users/Pavithra/Desktop/Caesar-Cipher-Cracking/cipher.py -f .\messages\message.txt -c 4 -n pavithra

------------------------Read Plain Text------------------------
Message: the caesar cipher technique is one of the earliest and simplest methods of encryption technique it is s
imply a type of substitution cipher ie each letter of a given text is replaced by a letter with a fixed number o
f positions down the alphabet for example with a shift of one a would be replaced by b b would become c and so o
n the method is apparently named after julius caesar who apparently used it to communicate with his officials

Encrypted Cipher Text: xlidgeiwevdgmtlivdxiglrmuyidmwdsridsjdxlidievpmiwxderhdwmqtpiwxdqixlshwdsjdirgvbtxmsrdxig
lrmuyidmxdmwdwmqtpbdedxbtidsjdwyfwxmxyxmsrdgmtlivdmidiegldpixxivdsjdedkmzirdxiaxdmwdvitpegihdfbdedpixxivd mxlded
jmaihdryqfivdsjdtswmxmsrwdhs rdxlideptlefixdjsvdiaeqtpid mxldedwlmjxdsjdsrided syphdfidvitpegihdfbdfdfd syphdfig
sqidgderhdwsdsrdxlidqixlshdmwdettevirxpbdreqihdejxivdnypmywdgeiwevd lsdettevirxpbdywihdmxdxsdgsqqyrmgexid mxldlm
wdsjjmgmepwd
------------------------Encoding name------------------------
Ciphertext corresponsding to pavithra: tezmxlve
------------------------Cracking using frequency analysis------------------------
Cracked Cipher Text: the caesar cipher technique is one of the earliest and simplest methods of encryption techn
ique it is simply a type of substitution cipher ie each letter of a given text is replaced by a letter with a fi
xed number of positions down the alphabet for example with a shift of one a would be replaced by b b would becom
e c and so on the method is apparently named after julius caesar who apparently used it to communicate with his
officials
Cracked name: pavithra
The value of n is: 4
------------------------Cracking using bigram analysis------------------------
Cracked Cipher Text: the caesar cipher technique is one of the earliest and simplest methods of encryption techn
ique it is simply a type of substitution cipher ie each letter of a given text is replaced by a letter with a fi
xed number of positions down the alphabet for example with a shift of one a would be replaced by b b would becom
e c and so on the method is apparently named after julius caesar who apparently used it to communicate with his
officials
Cracked name: pavithra
The value of n is: 4
```

```
------------------------Cracking using mono-alphabetic substitution------------------------
Iteration number: 1
Cracked Cipher Text: wkhcfdhvducflskhucwhfkqltxhclvcrqhcricwkhchduolhvwcdqgcvlpsohvwcphwkrgvcrichqfuaswlrqcwhfkq
ltxhclwclvcvlpsoacdcwashcricvxevwlwxwlrqcflskhuclhchdfkcohwwhucricdcjlyhqcwh wclvcuhsodfhgceacdcohwwhuczlwkcdcil
 hgcqxpehucricsrvlwlrqvcgrzqcwkhcdoskdehwciruch dpsohczlwkcdcvkliwcricrqhcdczrxogcehcuhsodfhgceacececzrxogcehfrp
hcfcdqgcvrcrqcwkhcphwkrgclvcdssduhqwoacqdphgcdiwhucmxolxvcfdhvduczkrcdssduhqwoacxvhgclwcwrcfrppxqlfdwhczlwkcklvc
riilfldovc
---------------------------------------------------------------------------------
Type 'y' if it seems valid, else 'n': n
Iteration number: 2
Cracked Cipher Text: vjgbecguctbekrjgtbvgejpkswgbkubqpgbqhbvjgbgctnkguvbcpfbukornguvbogvjqfubqhbgpet rvkqpbvgejp
kswgbkvbkubukorn bcbv rgbqhbuwduvkvwvkqpbekrjgtbkgbgcejbngvvgtbqhbcbikxgpbvgzvbkubtgrncegfbd bcbngvvgtbykvjbcbhk
zgfbpwodgtbqhbrqukvkqpubfqypbvjgbcnrjcdgvbhqtbgzcorngbykvjbcbujkhvbqhbqpgbcbyqwnfbdgbtgrncegfbd bdbdbyqwnfbdgeqo
gbebcpfbuqbqpbvjgbogvjqfbkubcrrctgpvn bpcogfbchvgtblwnkwubecguctbyjqbcrrctgpvn bwugfbkvbvqbeqoowpkecvgbykvjbjkub
qhhkekcnub
---------------------------------------------------------------------------------
Type 'y' if it seems valid, else 'n': n
Iteration number: 3
Cracked Cipher Text: uifadbftbsadjqifsaufdiojrvfajtapofapgauifafbsmjftuaboeatjnqmftuanfuipetapgafodszqujpoaufdio
jrvfajuajtatjnqmzabauzqfapgatvctujuvujpoadjqifsajfafbdiamfuufsapgabahjwfoaufyuajtasfqmbdfeaczabamfuufsaxjuiabagj
yfeaovncfsapgaqptjujpotaepxoauifabmqibcfuagpsafybnqmfaxjuiabatijguapgapofabaxpvmeacfasfqmbdfeaczacacaxpvmeacfdpn
fadaboeatpapoauifanfuipeajtabqqbsfoumzaobnfeabgufsakvmjvtadbftbsaxipabqqbsfoumzavtfeajuaupadpnnvojdbufaxjuiaijta
pggjdjbmta
---------------------------------------------------------------------------------
Type 'y' if it seems valid, else 'n': n
Iteration number: 4
Cracked Cipher Text: the caesar cipher technique is one of the earliest and simplest methods of encryption techn
ique it is simply a type of substitution cipher ie each letter of a given text is replaced by a letter with a fi
xed number of positions down the alphabet for example with a shift of one a would be replaced by b b would becom
e c and so on the method is apparently named after julius caesar who apparently used it to communicate with his
officials
---------------------------------------------------------------------------------
Type 'y' if it seems valid, else 'n': y
Cracked name: pavithra
```

The encryption corresponding to ‘pavithra’ is: `tezmxlve`. It has been successfully decrypted as well.

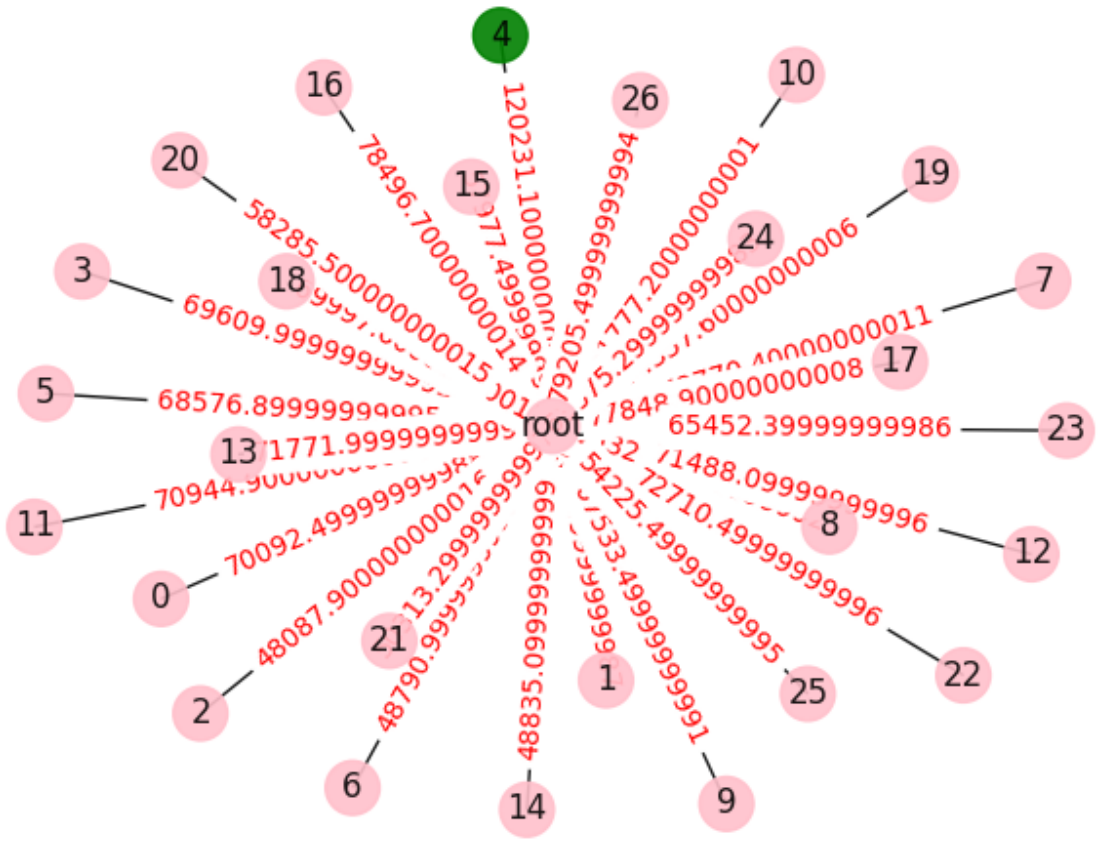
In the **mono-alphabetic substitution** method, I will check until I observe that the cracked text resembles a valid English text.

As we choose the best scores obtained across all iterations for the alphabet frequency and bigram frequency approach, they are **greedy methods.**

Along with the decryption of the ciphertexts, the **state-space graphs** are plotted using the `networkx` library in python.

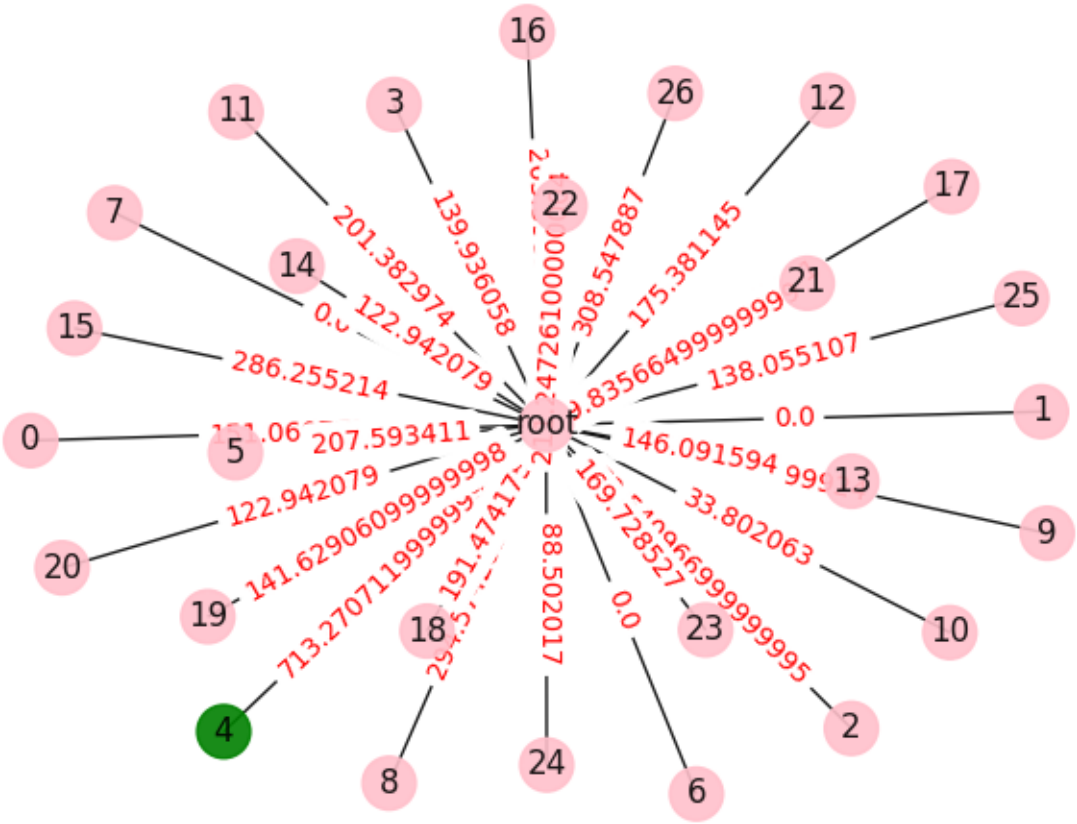
- **Alphabet frequency method**

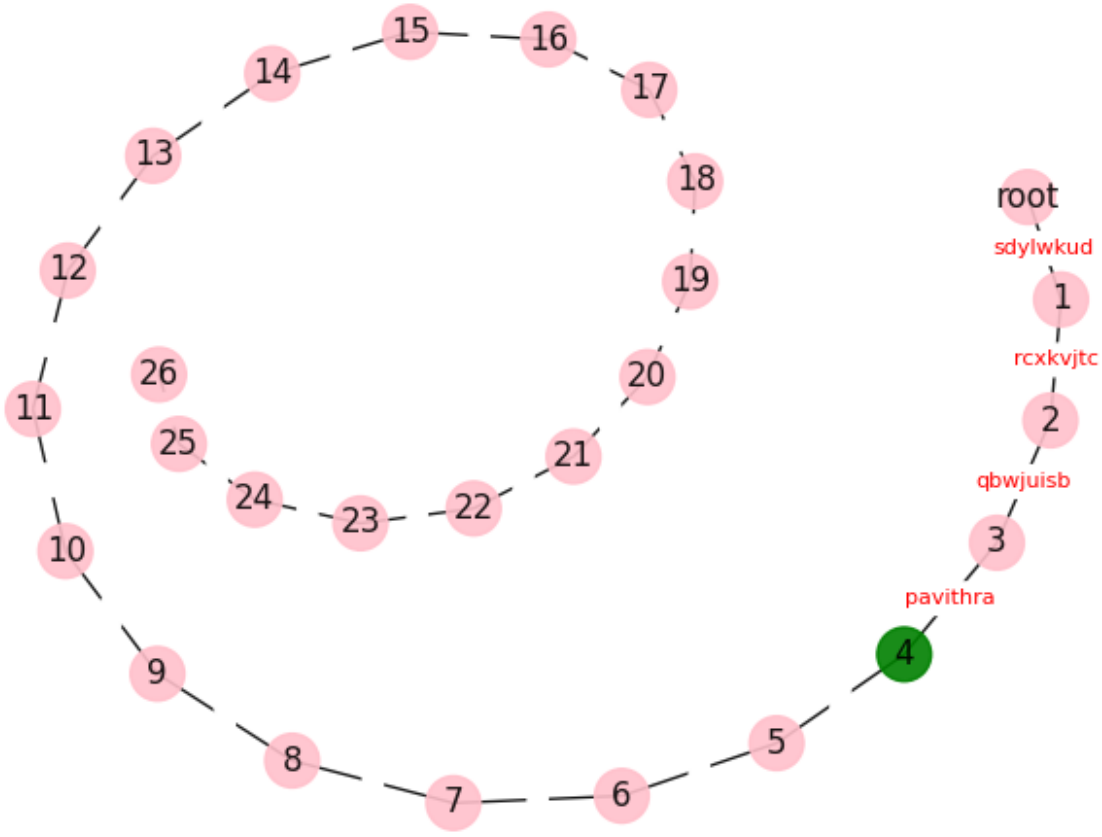
As we can see from the graph, the maximum score is when the rotation factor is 4.



- **Bigram frequency**

As we can see from the graph, the maximum score is when the rotation factor is 4.


root
16
3
11
26
12
22
17
7
14
21
25
15
1
0
5
20
13
9
19
18
23
10
4
8
24
6
2
0.0
207.593411
122.942079
286.255214
138.055107
146.091594
33.802063
88.502017
201.382974
139.936058
308.5478887
175.381145
122.942079


root
sdylwkud
1
rcxkvjtc
2
qbwjuisb
3
pavithra
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26

- **Mono-alphabetic substitution**

Here, we can see all the possible decryptions of my name across each rotation factor till 4.

2. Hill Climbing

In order to show hill climbing, I will use the quadgram frequencies.

The `hill_climb.py` consists of this implementation. The help menu can be obtained by running the `python` file with `-h` flag.

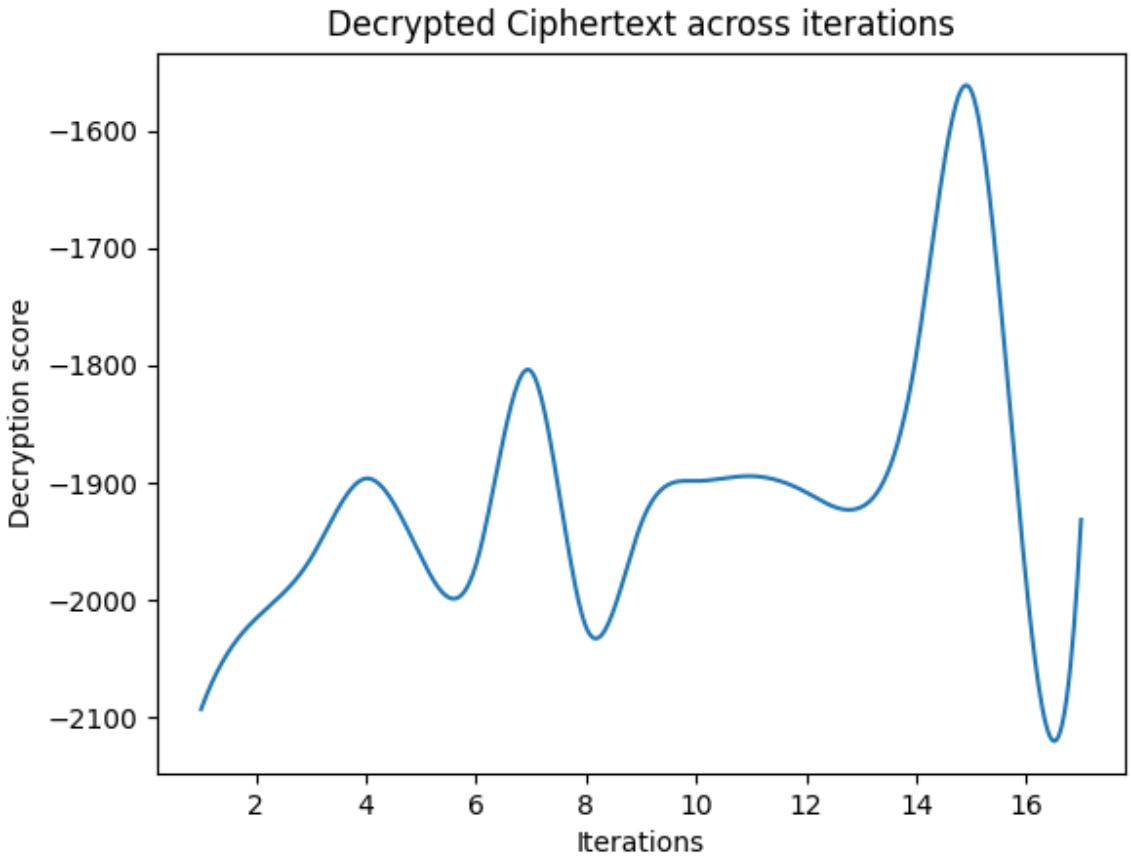
```
PS C:\Users\Pavithra\Desktop\Caesar-Cipher-Cracking> & C:/Users/Pavithra/AppData/Local/Programs/Python/Python38/
python.exe c:/Users/Pavithra/Desktop/Caesar-Cipher-Cracking/hill_climb.py -h
usage: hill_climb.py [-h] -f F

optional arguments:
  -h, --help  show this help message and exit
  -f F        Name of the file containing the ciphertext to be decrypted
```

While running the python file, provide the path to the ciphertext to be decrypted via the `-f` flag.

Here the scores for each iteration will be computed till a **local maximum** is reached and shown to the user. If the text resembles a valid English text, then we can stop.

We can see below that the scores progressively increase (becomes less negative). The graph depicting the scores across each iteration is also generated.

```
python.exe c:/Users/Pavithra/Desktop/Caesar-Cipher-Cracking/hill_climb.py -f .\ciphers\cipher.txt
Iteration:  1
Best score so far: -2092.737928584506
Cracked Cipher Text: aprderseydtopryardpltvmrtsilrinaprreyutrsaelcstfoursafrapicsinrldyboatilardpltvmrtatsstfoub
eaborinsmhsatamatildtoprytrredpuraaryinextgrlarwatsyrouedrchbeuraaryktapentwrclmfhryinoistatilsciklapreuopehrani
yrwefourktapesptnainilrekimuchryrouedrchbhhkimuchrdifrdelcsiilaprfrapictseooeyrlaublefrcenaryjmutmsderseykpieooe
yrlaubmsrctaaidiffmltdearktapptsinntdteus

Iteration:  2
Best score so far: -2015.4807267745393
Cracked Cipher Text: rdsneshetnaldstrsndoakcsahiosimrdssetyashreofhaplyshrpsrdifhimsontblraiorsndoakcsarahhaplybe
rblsimhcuhrarcraionaldstassendysrrstimejaxsorsgrahtslyensfubeysrrstwardemagsfocpustimliharaiohfiwordseyldeusrmits
geplyswardehdamrimiosewicyfustslyensfubuuwicyfusnipsneofhiiordspsrdifahelletsoryboepsfemrstvcyachneshetwdielletso
rybchsfarrinippcoanerswarddahimmanaeyh

Iteration:  3
Best score so far: -1963.9762491613449
Cracked Cipher Text: themaenarmishertemhligueinclecdtheearpientalonikspentkethconcdelmrysticltemhligueitinnikspya
tysecdnufntituticlmisherieeamhpettercdawizeltevtinrespameofyapetterbithadiveolukfercdscniticlnocbltheapshafetdcre
vakspebithanhidtcdcleabcupoferespameofyffbcupofemckemaloncclthekethcoinassareltpylakeoadterjupiunmaenarbhcassarel
tpyuneoittcmckkulimatebithhincddimiapn

Iteration:  4
Best score so far: -1896.3335920806883
Cracked Cipher Text: thepierinpodhentephsofyeorasealtheeinmoertiscrobdmertbethacralespnudtoastephsofyeotorrobdmui
tudealrykrtotytoaspodhenoeeiphmettenaligojestevtornedmipeckuimettenwothilovecsybkenaldarotoasrcawstheimdhiketlane
vibdmewothirholtalaseiwaymckenedmipeckukkwaymckepabepiscraasthebethacoriddinestmusibeciltenzymoyrpierinwhaiddines
tmuyrecottapabbysopitewothhorallopoimr

Iteration:  7
Best score so far: -1805.0793899297735
Cracked Cipher Text: thedaesaldicheltedhnixueisoneoftheealkiestanrsimckestmethorsofendlpctiontedhnixueitissimckpa
tpceofsubstitutiondichelieeadhkettelofawizentevtisleckaderbpakettelyithafivernumbelofcositionsroyntheakchabetfole
vamckeyithashiftofoneayoukrbeleckaderbpbbyoukrbedomedanrsoonthemethorisaccalentkpnamerafteljukiusdaesalyhoaccalen
tkpuserittodommunidateyithhisoffidiaks

Iteration:  14
Best score so far: -1792.0096152034484
Cracked Cipher Text: thecaesaycirheytechnikpeisoneoftheeayliestandsimrlestmethodsofencyurtiontechnikpeitissimrlua
tureofspbstitptioncirheyieeachletteyofajiwentevtisyerlacedbualetteygithafivednpmbeyofrositionsdogntheralrhabetfoye
vamrlegithashiftofoneagopldbeyerlacedbubbgopldbecomecandsoonthemethodisarrayentlunamedafteyxplipscaesayghoarrayen
tlupsedittocommpnicategithhisofficials

Iteration:  15
Best score so far: -1566.0091345202923
Cracked Cipher Text: thecaesarciphertechniqueisoneoftheearliestandsimplestmethodsofencryptiontechniqueitissimplya
typeofsugstitutioncipherieeachletterofakizentevtisreplacedgyaletterwithafivednumgerofpositionsdownthealphagetfore
vamplewithashiftofoneawouldgereplacedgyggwouldgecomecandsoonthemethodisapparentlynamedafterbuliuscaesarwhoapparen
tlyuseditocommunicatewithhisofficials

End of decryption
```


Decrypted Ciphertext across iterations
Decryption score
−1600
−1700
−1800
−1900
−2000
−2100
2
4
6
8
10
12
14
16
Iterations