04 - Section Relevance Score

In this notebook, we will assess the similarity of each section with the lay summary. Then, we will calculate relevance score (average) for each named section (Abstract, Conclusion, etc)

In [80]:
```
import pandas as pd
from sentence_transformers import SentenceTransformer, util
import numpy as np
import random
from collections import defaultdict
import torch
```

In [2]:
```
file_path = "../data/biolaysumm2024_data/"
file_name = "eLife_train.jsonl"

df = pd.read_json(file_path + file_name,
                  orient="records",
                  lines=True)
df
```

Out[2]:

	lay_summary	article	headings	keywords	id
0	In the USA , more deaths happen in the winter ...	In temperate climates , winter deaths exceed s...	[Abstract, Introduction, Results, Discussion, ...	[epidemiology and global health]	elife-35500-v1
1	Most people have likely experienced the discom...	Whether complement dysregulation directly cont...	[Abstract, Introduction, Results, Discussion, ...	[microbiology and infectious disease, immunolo...	elife-48378-v2
2	The immune system protects an individual from ...	Variation in the presentation of hereditary im...	[Abstract, Introduction, Results, Discussion, ...	[microbiology and infectious disease, immunolo...	elife-04494-v1
3	The brain adapts to control our behavior in di...	Rapid and flexible interpretation of conflicti...	[Abstract, Introduction, Results, Discussion, ...	[neuroscience]	elife-12352-v2
4	Cells use motor proteins that to move organell...	Myosin 5a is a dual-headed molecular motor tha...	[Abstract, Introduction, Results, Discussion, ...	[structural biology and molecular biophysics]	elife-05413-v2
...	...	...	...	...	...
4341	To defend itself against bacteria and viruses ...	Antibodies are critical components of adaptive...	[Abstract, Introduction, Results, Discussion, ...	[structural biology and molecular biophysics, ...	elife-61393-v2
4342	DNA is tightly packaged in a material called c...	RNA polymerase II (PolII) transcribes RNA wi...	[Abstract, Introduction, Results, Discussion, ...	[chromosomes and gene expression]	elife-02042-v1
4343	Associative learning is a simple learning abil...	Gagliano et al . (Learning by association in ...	[Abstract, Introduction, Results, Discussion, ...	[plant biology, short report]	elife-57614-v1
4344	In 1848 , a railroad worker named Phineas Gage...	Activity in prefrontal cortex (PFC) has been...	[Abstract, Introduction, Results, Discussion, ...	[neuroscience]	elife-11945-v2
4345	Medical imaging covers a wide range of techniq...	Medical imaging can visualize characteristics ...	[Abstract, Introduction, Results, Discussion, ...	[computational and systems biology, cancer bio...	elife-23421-v3

4346 rows × 5 columns

In [3]:
```
k = 120 # a random row
item = df.iloc[k]
item
```

Out[3]:
```
lay_summary    Genome editing allows scientists to change an ...
article        The CRISPR-Cas9 targeted nuclease technology a...
headings       [Abstract, Introduction, Results and discussio...
keywords         [chromosomes and gene expression, short report]
id                                                  elife-33761-v5
Name: 120, dtype: object
```

In [4]:
```
# divide by paragraphs
paras = item.article.split("\n")
len(paras)
```

Out[4]:
```
4
```

In [6]:
```
# check with `headings`
print(len(item.headings))
print(item.headings)
```

```
4
['Abstract', 'Introduction', 'Results and discussion', 'Materials\xa0and\xa0method
s']
```

In [7]:
```
len(paras) == len(item.headings)
```

Out[7]:
```
True
```

In [12]:
```
# test sentence embeddings
model = SentenceTransformer("all-MiniLM-L6-v2")
s1 = "This is a paper about t-cell"
s2 = "A new research paper suggesting a new role for t-cell in our body"
v1 = model.encode(s1)
v2 = model.encode(s2)
v1.shape
```

Out[12]:
```
(384,)
```

In [15]:
```
util.cos_sim(v1, v2).item() # should be close to 1
```

Out[15]:
```
0.7994392514228821
```

In [16]:
```
v_lay_summ = model.encode(item.lay_summary)
v_lay_summ.shape
```

Out[16]:
```
(384,)
```

In [21]:
```
for i, heading in enumerate(item.headings):
    print(i, heading)
    v_para = model.encode(paras[i])
    score = util.cos_sim(v_para, v_lay_summ).item()
    print(f"Sim score with lay summary = {score:.2f}")
```

```
0 Abstract
Sim score with lay summary = 0.77
1 Introduction
Sim score with lay summary = 0.80
2 Results and discussion
Sim score with lay summary = 0.64
3 Materials and methods
Sim score with lay summary = 0.36
```

In [147…

```python
def get_chunks(text, chunk_size=1000, overlap=100):
    """
        split a long text into chunks
    """
    if chunk_size <= overlap:
        return None

    i = 0
    result = []
    while i < len(text):
        result.append(text[i:i+chunk_size])
        i += (chunk_size - overlap)
        # print("New i =", i)

    return result

get_chunks("This is a very very long long long text", 15, 4)
```

Out[147…

```
['This is a very ', 'ery very long l', 'ng long long te', 'g text']
```

In [148…

```python
def get_para_embedding(text):
    """
        return embedding of a paragraph
        TODO: for long text, will get average of n chunks
    """
    result = None
    if len(text) <= 1000:
        result = model.encode(text)
    else: # long text -> split into chunks and average
        chunks = get_chunks(text)
        v_chunks = [model.encode(chk)
                    for chk in chunks
                   ]
        # print(len(v_chunks), v_chunks[0].shape)

        result = np.average(v_chunks, axis=0)

        # print(result.shape)

    return result
```

In [154…

```python
# put into a function
def get_section_score(row_id = 0):
    """
        print section score for row i in the dataset
    """
```

```
    item = df.iloc[row_id]
    item_paras = item.article.split("\n")
    v_lay_summ = model.encode(item.lay_summary)

    result = dict()
    for i, heading in enumerate(item.headings):
        # print(i, heading)
        # v_para = model.encode(item_paras[i])
        v_para = get_para_embedding(item_paras[i])
        # print(v_para.shape)
        score = util.cos_sim(v_para, v_lay_summ).item()
        # print(f"Sim score with lay summary = {score:.2f}")
        result[heading.lower()] = [score]

    return result
```

In [155…
```
get_section_score(123)
```

Out[155…
```
{'abstract': [0.7081866264343262],
 'introduction': [0.7312494516372681],
 'results': [0.6945871710777283],
 'discussion': [0.7358657121658325],
 'materials and methods': [0.5820637345314026]}
```

In [156…
```
# test dictionary update
a = {"intro": [1],
     "background": [2]
    }
b = {"intro": [3],
     "background": [4]
    }
a.update(b)
a
```

Out[156…
```
{'intro': [3], 'background': [4]}
```

In [168…
```
random.seed(42)

n = 1000 # testing for 10 random rows
scores = defaultdict(list)
for i in range(n):
    k = random.randint(0, len(df))
    print(f"i = {i}, k = {k}")
    score = get_section_score(k)
    # print(score)
    # update results
    for k, v in score.items():
        # results.update(score)
        scores[k].extend(v)

# scores
```

```
i = 0, k = 912
i = 1, k = 204
i = 2, k = 2253
i = 3, k = 2006
i = 4, k = 1828
i = 5, k = 1143
i = 6, k = 839
i = 7, k = 712
i = 8, k = 3456
i = 9, k = 260
i = 10, k = 244
i = 11, k = 767
i = 12, k = 1791
i = 13, k = 1905
i = 14, k = 4139
i = 15, k = 217
i = 16, k = 1628
i = 17, k = 3436
i = 18, k = 1805
i = 19, k = 3679
i = 20, k = 2278
i = 21, k = 53
i = 22, k = 1307
i = 23, k = 3462
i = 24, k = 2787
i = 25, k = 2276
i = 26, k = 1273
i = 27, k = 1763
i = 28, k = 2757
i = 29, k = 837
i = 30, k = 759
i = 31, k = 3112
i = 32, k = 792
i = 33, k = 2940
i = 34, k = 2817
i = 35, k = 2166
i = 36, k = 355
i = 37, k = 3763
i = 38, k = 1022
i = 39, k = 3100
i = 40, k = 645
i = 41, k = 2401
i = 42, k = 2962
i = 43, k = 1575
i = 44, k = 569
i = 45, k = 375
i = 46, k = 1866
i = 47, k = 2370
i = 48, k = 653
i = 49, k = 1907
i = 50, k = 827
i = 51, k = 3113
i = 52, k = 2277
i = 53, k = 3714
i = 54, k = 2988
i = 55, k = 1332
```

```
i = 56, k = 3032
i = 57, k = 2910
i = 58, k = 1716
i = 59, k = 2187
i = 60, k = 584
i = 61, k = 1401
i = 62, k = 2005
i = 63, k = 1338
i = 64, k = 3786
i = 65, k = 3108
i = 66, k = 2211
i = 67, k = 1799
i = 68, k = 2656
i = 69, k = 458
i = 70, k = 1876
i = 71, k = 262
i = 72, k = 2584
i = 73, k = 3286
i = 74, k = 2193
i = 75, k = 542
i = 76, k = 1728
i = 77, k = 2577
i = 78, k = 1741
i = 79, k = 4089
i = 80, k = 3241
i = 81, k = 3758
i = 82, k = 1170
i = 83, k = 2169
i = 84, k = 1143
i = 85, k = 2020
i = 86, k = 2152
i = 87, k = 3509
i = 88, k = 3271
i = 89, k = 2965
i = 90, k = 1796
i = 91, k = 1133
i = 92, k = 4174
i = 93, k = 4042
i = 94, k = 744
i = 95, k = 385
i = 96, k = 898
i = 97, k = 1252
i = 98, k = 1310
i = 99, k = 3458
i = 100, k = 520
i = 101, k = 3152
i = 102, k = 3126
i = 103, k = 3834
i = 104, k = 4334
i = 105, k = 2059
i = 106, k = 94
i = 107, k = 938
i = 108, k = 2185
i = 109, k = 2786
i = 110, k = 913
i = 111, k = 2404
```

i = 112, k = 3561
i = 113, k = 1295
i = 114, k = 3716
i = 115, k = 26
i = 116, k = 2157
i = 117, k = 4100
i = 118, k = 1463
i = 119, k = 4158
i = 120, k = 871
i = 121, k = 2444
i = 122, k = 4158
i = 123, k = 1629
i = 124, k = 1252
i = 125, k = 3063
i = 126, k = 1323
i = 127, k = 4344
i = 128, k = 4
i = 129, k = 2655
i = 130, k = 4002
i = 131, k = 159
i = 132, k = 916
i = 133, k = 2973
i = 134, k = 2519
i = 135, k = 1961
i = 136, k = 474
i = 137, k = 1973
i = 138, k = 645
i = 139, k = 701
i = 140, k = 3981
i = 141, k = 566
i = 142, k = 1030
i = 143, k = 1051
i = 144, k = 3893
i = 145, k = 1352
i = 146, k = 2171
i = 147, k = 4322
i = 148, k = 3466
i = 149, k = 1735
i = 150, k = 1647
i = 151, k = 2553
i = 152, k = 3268
i = 153, k = 3059
i = 154, k = 3588
i = 155, k = 4239
i = 156, k = 3698
i = 157, k = 991
i = 158, k = 2030
i = 159, k = 1840
i = 160, k = 524
i = 161, k = 2769
i = 162, k = 172
i = 163, k = 1885
i = 164, k = 1804
i = 165, k = 58
i = 166, k = 581
i = 167, k = 482

```
i = 168, k = 1875
i = 169, k = 552
i = 170, k = 257
i = 171, k = 2706
i = 172, k = 580
i = 173, k = 4211
i = 174, k = 1949
i = 175, k = 2281
i = 176, k = 3976
i = 177, k = 1755
i = 178, k = 1083
i = 179, k = 3872
i = 180, k = 1990
i = 181, k = 3874
i = 182, k = 3334
i = 183, k = 1559
i = 184, k = 772
i = 185, k = 794
i = 186, k = 3531
i = 187, k = 2902
i = 188, k = 3469
i = 189, k = 3367
i = 190, k = 3825
i = 191, k = 443
i = 192, k = 806
i = 193, k = 496
i = 194, k = 3298
i = 195, k = 2779
i = 196, k = 895
i = 197, k = 2036
i = 198, k = 1569
i = 199, k = 1558
i = 200, k = 3675
i = 201, k = 1148
i = 202, k = 3456
i = 203, k = 1503
i = 204, k = 2281
i = 205, k = 3789
i = 206, k = 2046
i = 207, k = 617
i = 208, k = 3630
i = 209, k = 802
i = 210, k = 414
i = 211, k = 120
i = 212, k = 764
i = 213, k = 1936
i = 214, k = 1362
i = 215, k = 3329
i = 216, k = 3978
i = 217, k = 3943
i = 218, k = 1751
i = 219, k = 3285
i = 220, k = 480
i = 221, k = 1348
i = 222, k = 3104
i = 223, k = 17
```

```
i = 224, k = 3198
i = 225, k = 2172
i = 226, k = 3727
i = 227, k = 2336
i = 228, k = 3465
i = 229, k = 3986
i = 230, k = 1268
i = 231, k = 1555
i = 232, k = 2430
i = 233, k = 1783
i = 234, k = 479
i = 235, k = 499
i = 236, k = 2569
i = 237, k = 468
i = 238, k = 410
i = 239, k = 3905
i = 240, k = 4119
i = 241, k = 1289
i = 242, k = 465
i = 243, k = 4160
i = 244, k = 656
i = 245, k = 1522
i = 246, k = 561
i = 247, k = 556
i = 248, k = 1926
i = 249, k = 3307
i = 250, k = 982
i = 251, k = 2016
i = 252, k = 325
i = 253, k = 671
i = 254, k = 3434
i = 255, k = 4282
i = 256, k = 2591
i = 257, k = 2136
i = 258, k = 1673
i = 259, k = 2573
i = 260, k = 1955
i = 261, k = 2175
i = 262, k = 3242
i = 263, k = 1072
i = 264, k = 2457
i = 265, k = 3745
i = 266, k = 2590
i = 267, k = 594
i = 268, k = 76
i = 269, k = 3754
i = 270, k = 819
i = 271, k = 600
i = 272, k = 1746
i = 273, k = 4144
i = 274, k = 2172
i = 275, k = 1085
i = 276, k = 2859
i = 277, k = 563
i = 278, k = 2001
i = 279, k = 3027
```

```
i = 280, k = 2334
i = 281, k = 1292
i = 282, k = 3589
i = 283, k = 2478
i = 284, k = 4333
i = 285, k = 64
i = 286, k = 2452
i = 287, k = 848
i = 288, k = 1100
i = 289, k = 2166
i = 290, k = 945
i = 291, k = 876
i = 292, k = 1273
i = 293, k = 2231
i = 294, k = 2308
i = 295, k = 1725
i = 296, k = 2808
i = 297, k = 1667
i = 298, k = 2162
i = 299, k = 4140
i = 300, k = 4002
i = 301, k = 2057
i = 302, k = 416
i = 303, k = 756
i = 304, k = 3469
i = 305, k = 2266
i = 306, k = 361
i = 307, k = 29
i = 308, k = 2732
i = 309, k = 1071
i = 310, k = 2145
i = 311, k = 1323
i = 312, k = 3619
i = 313, k = 3503
i = 314, k = 79
i = 315, k = 916
i = 316, k = 616
i = 317, k = 1221
i = 318, k = 295
i = 319, k = 3024
i = 320, k = 1213
i = 321, k = 3520
i = 322, k = 1044
i = 323, k = 342
i = 324, k = 2525
i = 325, k = 2987
i = 326, k = 326
i = 327, k = 2931
i = 328, k = 1720
i = 329, k = 2044
i = 330, k = 842
i = 331, k = 2897
i = 332, k = 3329
i = 333, k = 1266
i = 334, k = 1939
i = 335, k = 1331
```

```
i = 336, k = 1450
i = 337, k = 3377
i = 338, k = 203
i = 339, k = 1469
i = 340, k = 2721
i = 341, k = 3372
i = 342, k = 2032
i = 343, k = 2185
i = 344, k = 1304
i = 345, k = 885
i = 346, k = 3133
i = 347, k = 317
i = 348, k = 3855
i = 349, k = 1822
i = 350, k = 1634
i = 351, k = 3770
i = 352, k = 2864
i = 353, k = 2500
i = 354, k = 1864
i = 355, k = 1826
i = 356, k = 193
i = 357, k = 1582
i = 358, k = 3264
i = 359, k = 2689
i = 360, k = 2282
i = 361, k = 568
i = 362, k = 2286
i = 363, k = 2876
i = 364, k = 4173
i = 365, k = 3274
i = 366, k = 2712
i = 367, k = 226
i = 368, k = 944
i = 369, k = 2139
i = 370, k = 1462
i = 371, k = 2174
i = 372, k = 313
i = 373, k = 888
i = 374, k = 3559
i = 375, k = 2831
i = 376, k = 2569
i = 377, k = 3574
i = 378, k = 4189
i = 379, k = 947
i = 380, k = 3155
i = 381, k = 1557
i = 382, k = 2086
i = 383, k = 363
i = 384, k = 3572
i = 385, k = 13
i = 386, k = 4259
i = 387, k = 1614
i = 388, k = 2983
i = 389, k = 3533
i = 390, k = 573
i = 391, k = 2704
```

```
i = 392, k = 2571
i = 393, k = 1020
i = 394, k = 2460
i = 395, k = 4154
i = 396, k = 2533
i = 397, k = 3345
i = 398, k = 2672
i = 399, k = 3296
i = 400, k = 2422
i = 401, k = 1042
i = 402, k = 1571
i = 403, k = 3444
i = 404, k = 3105
i = 405, k = 1425
i = 406, k = 2465
i = 407, k = 3326
i = 408, k = 3
i = 409, k = 2489
i = 410, k = 2350
i = 411, k = 1721
i = 412, k = 3521
i = 413, k = 2639
i = 414, k = 3809
i = 415, k = 3619
i = 416, k = 3622
i = 417, k = 1750
i = 418, k = 4187
i = 419, k = 3876
i = 420, k = 1390
i = 421, k = 694
i = 422, k = 2324
i = 423, k = 4222
i = 424, k = 2745
i = 425, k = 765
i = 426, k = 1924
i = 427, k = 2542
i = 428, k = 1840
i = 429, k = 1631
i = 430, k = 1207
i = 431, k = 200
i = 432, k = 378
i = 433, k = 2005
i = 434, k = 3892
i = 435, k = 596
i = 436, k = 3730
i = 437, k = 3395
i = 438, k = 1592
i = 439, k = 3145
i = 440, k = 4049
i = 441, k = 3273
i = 442, k = 1998
i = 443, k = 1208
i = 444, k = 45
i = 445, k = 873
i = 446, k = 3482
i = 447, k = 1792
```

```
i = 448, k = 1440
i = 449, k = 4243
i = 450, k = 3805
i = 451, k = 411
i = 452, k = 2041
i = 453, k = 994
i = 454, k = 3739
i = 455, k = 1092
i = 456, k = 3806
i = 457, k = 2599
i = 458, k = 3625
i = 459, k = 4135
i = 460, k = 3495
i = 461, k = 3652
i = 462, k = 1303
i = 463, k = 3888
i = 464, k = 3686
i = 465, k = 2123
i = 466, k = 2025
i = 467, k = 2271
i = 468, k = 4270
i = 469, k = 3969
i = 470, k = 1959
i = 471, k = 2249
i = 472, k = 3603
i = 473, k = 634
i = 474, k = 2340
i = 475, k = 1920
i = 476, k = 2225
i = 477, k = 2751
i = 478, k = 2619
i = 479, k = 660
i = 480, k = 1133
i = 481, k = 1235
i = 482, k = 1894
i = 483, k = 3137
i = 484, k = 1251
i = 485, k = 1752
i = 486, k = 526
i = 487, k = 3398
i = 488, k = 3339
i = 489, k = 2710
i = 490, k = 3816
i = 491, k = 3406
i = 492, k = 510
i = 493, k = 1694
i = 494, k = 3441
i = 495, k = 3190
i = 496, k = 160
i = 497, k = 3116
i = 498, k = 3907
i = 499, k = 48
i = 500, k = 2881
i = 501, k = 2446
i = 502, k = 3194
i = 503, k = 3432
```

```
i = 504, k = 1806
i = 505, k = 3999
i = 506, k = 1797
i = 507, k = 2235
i = 508, k = 3570
i = 509, k = 3978
i = 510, k = 237
i = 511, k = 3185
i = 512, k = 2753
i = 513, k = 3312
i = 514, k = 1352
i = 515, k = 3828
i = 516, k = 1045
i = 517, k = 220
i = 518, k = 3227
i = 519, k = 222
i = 520, k = 687
i = 521, k = 3511
i = 522, k = 1111
i = 523, k = 3782
i = 524, k = 1488
i = 525, k = 411
i = 526, k = 2131
i = 527, k = 3105
i = 528, k = 2681
i = 529, k = 1733
i = 530, k = 3724
i = 531, k = 2677
i = 532, k = 2764
i = 533, k = 3105
i = 534, k = 2279
i = 535, k = 3453
i = 536, k = 2066
i = 537, k = 670
i = 538, k = 3852
i = 539, k = 158
i = 540, k = 426
i = 541, k = 2866
i = 542, k = 1836
i = 543, k = 562
i = 544, k = 329
i = 545, k = 254
i = 546, k = 2025
i = 547, k = 1633
i = 548, k = 166
i = 549, k = 1248
i = 550, k = 1954
i = 551, k = 1034
i = 552, k = 3879
i = 553, k = 937
i = 554, k = 1785
i = 555, k = 3809
i = 556, k = 2099
i = 557, k = 3021
i = 558, k = 1374
i = 559, k = 938
```

```
i = 560, k = 1341
i = 561, k = 2548
i = 562, k = 885
i = 563, k = 210
i = 564, k = 2555
i = 565, k = 3074
i = 566, k = 3249
i = 567, k = 1624
i = 568, k = 622
i = 569, k = 1989
i = 570, k = 834
i = 571, k = 2470
i = 572, k = 991
i = 573, k = 336
i = 574, k = 2844
i = 575, k = 3509
i = 576, k = 3035
i = 577, k = 564
i = 578, k = 4144
i = 579, k = 2795
i = 580, k = 103
i = 581, k = 3441
i = 582, k = 4015
i = 583, k = 864
i = 584, k = 3551
i = 585, k = 2967
i = 586, k = 3766
i = 587, k = 1253
i = 588, k = 3567
i = 589, k = 1442
i = 590, k = 4274
i = 591, k = 2212
i = 592, k = 3960
i = 593, k = 3808
i = 594, k = 3568
i = 595, k = 2198
i = 596, k = 2640
i = 597, k = 2011
i = 598, k = 709
i = 599, k = 2284
i = 600, k = 3692
i = 601, k = 1997
i = 602, k = 3806
i = 603, k = 3104
i = 604, k = 2755
i = 605, k = 235
i = 606, k = 4049
i = 607, k = 2662
i = 608, k = 1489
i = 609, k = 3994
i = 610, k = 1737
i = 611, k = 2906
i = 612, k = 2116
i = 613, k = 2788
i = 614, k = 2290
i = 615, k = 2263
```

```
i = 616, k = 83
i = 617, k = 4232
i = 618, k = 1565
i = 619, k = 701
i = 620, k = 1977
i = 621, k = 3329
i = 622, k = 4002
i = 623, k = 1968
i = 624, k = 3900
i = 625, k = 4020
i = 626, k = 3671
i = 627, k = 141
i = 628, k = 762
i = 629, k = 2410
i = 630, k = 1815
i = 631, k = 3312
i = 632, k = 1993
i = 633, k = 2508
i = 634, k = 3023
i = 635, k = 3876
i = 636, k = 2816
i = 637, k = 3485
i = 638, k = 2709
i = 639, k = 2882
i = 640, k = 3717
i = 641, k = 2219
i = 642, k = 2511
i = 643, k = 2059
i = 644, k = 1888
i = 645, k = 988
i = 646, k = 1577
i = 647, k = 2584
i = 648, k = 979
i = 649, k = 1516
i = 650, k = 1569
i = 651, k = 1772
i = 652, k = 3966
i = 653, k = 2265
i = 654, k = 4297
i = 655, k = 2318
i = 656, k = 823
i = 657, k = 1590
i = 658, k = 2426
i = 659, k = 1863
i = 660, k = 2956
i = 661, k = 1469
i = 662, k = 2476
i = 663, k = 115
i = 664, k = 1036
i = 665, k = 2247
i = 666, k = 372
i = 667, k = 446
i = 668, k = 2393
i = 669, k = 1034
i = 670, k = 4021
i = 671, k = 840
```

```
i = 672, k = 100
i = 673, k = 2329
i = 674, k = 3845
i = 675, k = 3921
i = 676, k = 3608
i = 677, k = 2791
i = 678, k = 1510
i = 679, k = 420
i = 680, k = 2068
i = 681, k = 3913
i = 682, k = 934
i = 683, k = 535
i = 684, k = 3282
i = 685, k = 4028
i = 686, k = 606
i = 687, k = 439
i = 688, k = 1242
i = 689, k = 1222
i = 690, k = 2489
i = 691, k = 697
i = 692, k = 2033
i = 693, k = 970
i = 694, k = 3409
i = 695, k = 1848
i = 696, k = 4280
i = 697, k = 3116
i = 698, k = 3690
i = 699, k = 3626
i = 700, k = 2435
i = 701, k = 3512
i = 702, k = 2501
i = 703, k = 493
i = 704, k = 812
i = 705, k = 1702
i = 706, k = 1728
i = 707, k = 2167
i = 708, k = 665
i = 709, k = 1286
i = 710, k = 1964
i = 711, k = 1423
i = 712, k = 614
i = 713, k = 1282
i = 714, k = 21
i = 715, k = 3346
i = 716, k = 3690
i = 717, k = 3849
i = 718, k = 2385
i = 719, k = 267
i = 720, k = 1896
i = 721, k = 2360
i = 722, k = 2316
i = 723, k = 3719
i = 724, k = 583
i = 725, k = 1912
i = 726, k = 2167
i = 727, k = 1620
```

```
i = 728, k = 3482
i = 729, k = 940
i = 730, k = 1841
i = 731, k = 1220
i = 732, k = 2176
i = 733, k = 1165
i = 734, k = 584
i = 735, k = 488
i = 736, k = 1359
i = 737, k = 2519
i = 738, k = 2364
i = 739, k = 3597
i = 740, k = 1018
i = 741, k = 3839
i = 742, k = 2491
i = 743, k = 3297
i = 744, k = 2230
i = 745, k = 4099
i = 746, k = 4045
i = 747, k = 3586
i = 748, k = 658
i = 749, k = 326
i = 750, k = 3539
i = 751, k = 2640
i = 752, k = 2051
i = 753, k = 211
i = 754, k = 748
i = 755, k = 1875
i = 756, k = 169
i = 757, k = 2207
i = 758, k = 329
i = 759, k = 1435
i = 760, k = 3854
i = 761, k = 4251
i = 762, k = 3622
i = 763, k = 2278
i = 764, k = 1486
i = 765, k = 3570
i = 766, k = 4028
i = 767, k = 747
i = 768, k = 3850
i = 769, k = 2850
i = 770, k = 3345
i = 771, k = 2730
i = 772, k = 2630
i = 773, k = 856
i = 774, k = 1317
i = 775, k = 2701
i = 776, k = 3372
i = 777, k = 4058
i = 778, k = 2361
i = 779, k = 3280
i = 780, k = 300
i = 781, k = 3725
i = 782, k = 721
i = 783, k = 2576
```

```
i = 784, k = 2067
i = 785, k = 2648
i = 786, k = 949
i = 787, k = 3311
i = 788, k = 4215
i = 789, k = 9
i = 790, k = 3784
i = 791, k = 3385
i = 792, k = 444
i = 793, k = 1536
i = 794, k = 4247
i = 795, k = 2963
i = 796, k = 4083
i = 797, k = 3621
i = 798, k = 422
i = 799, k = 1667
i = 800, k = 2187
i = 801, k = 1073
i = 802, k = 2359
i = 803, k = 3589
i = 804, k = 3970
i = 805, k = 994
i = 806, k = 236
i = 807, k = 1960
i = 808, k = 1297
i = 809, k = 2545
i = 810, k = 112
i = 811, k = 3342
i = 812, k = 763
i = 813, k = 1840
i = 814, k = 929
i = 815, k = 3780
i = 816, k = 962
i = 817, k = 1261
i = 818, k = 4082
i = 819, k = 2390
i = 820, k = 4168
i = 821, k = 2239
i = 822, k = 3403
i = 823, k = 3952
i = 824, k = 3868
i = 825, k = 1996
i = 826, k = 3741
i = 827, k = 1184
i = 828, k = 3142
i = 829, k = 1561
i = 830, k = 4163
i = 831, k = 1118
i = 832, k = 571
i = 833, k = 2263
i = 834, k = 3399
i = 835, k = 2784
i = 836, k = 4159
i = 837, k = 2188
i = 838, k = 21
i = 839, k = 2317
```

```
i = 840, k = 2445
i = 841, k = 4011
i = 842, k = 1217
i = 843, k = 3658
i = 844, k = 3967
i = 845, k = 2827
i = 846, k = 2723
i = 847, k = 3090
i = 848, k = 3730
i = 849, k = 2636
i = 850, k = 1545
i = 851, k = 1956
i = 852, k = 3137
i = 853, k = 1913
i = 854, k = 3365
i = 855, k = 357
i = 856, k = 2606
i = 857, k = 3874
i = 858, k = 3123
i = 859, k = 3162
i = 860, k = 1246
i = 861, k = 4057
i = 862, k = 303
i = 863, k = 1034
i = 864, k = 4114
i = 865, k = 2719
i = 866, k = 822
i = 867, k = 3606
i = 868, k = 816
i = 869, k = 4308
i = 870, k = 3743
i = 871, k = 125
i = 872, k = 1180
i = 873, k = 3358
i = 874, k = 1264
i = 875, k = 612
i = 876, k = 3846
i = 877, k = 2171
i = 878, k = 2773
i = 879, k = 3256
i = 880, k = 657
i = 881, k = 2691
i = 882, k = 3113
i = 883, k = 2594
i = 884, k = 3997
i = 885, k = 294
i = 886, k = 560
i = 887, k = 1923
i = 888, k = 2354
i = 889, k = 1863
i = 890, k = 740
i = 891, k = 3555
i = 892, k = 806
i = 893, k = 823
i = 894, k = 3634
i = 895, k = 1362
```

```
i = 896, k = 2453
i = 897, k = 237
i = 898, k = 376
i = 899, k = 2657
i = 900, k = 459
i = 901, k = 2403
i = 902, k = 2936
i = 903, k = 3070
i = 904, k = 3528
i = 905, k = 1192
i = 906, k = 2000
i = 907, k = 3375
i = 908, k = 1475
i = 909, k = 1392
i = 910, k = 1434
i = 911, k = 646
i = 912, k = 3133
i = 913, k = 1972
i = 914, k = 4076
i = 915, k = 1172
i = 916, k = 1902
i = 917, k = 3777
i = 918, k = 2080
i = 919, k = 3764
i = 920, k = 2091
i = 921, k = 76
i = 922, k = 3811
i = 923, k = 2356
i = 924, k = 1294
i = 925, k = 605
i = 926, k = 3618
i = 927, k = 2830
i = 928, k = 2450
i = 929, k = 3475
i = 930, k = 2048
i = 931, k = 3742
i = 932, k = 2474
i = 933, k = 1631
i = 934, k = 3151
i = 935, k = 3958
i = 936, k = 873
i = 937, k = 1943
i = 938, k = 3124
i = 939, k = 2940
i = 940, k = 2423
i = 941, k = 2418
i = 942, k = 179
i = 943, k = 3242
i = 944, k = 2248
i = 945, k = 66
i = 946, k = 401
i = 947, k = 4069
i = 948, k = 2344
i = 949, k = 1885
i = 950, k = 2886
i = 951, k = 1794
```

```
i = 952, k = 1557
i = 953, k = 2053
i = 954, k = 1120
i = 955, k = 795
i = 956, k = 322
i = 957, k = 2530
i = 958, k = 3611
i = 959, k = 273
i = 960, k = 2988
i = 961, k = 1076
i = 962, k = 738
i = 963, k = 2417
i = 964, k = 2676
i = 965, k = 3403
i = 966, k = 1438
i = 967, k = 1644
i = 968, k = 1082
i = 969, k = 2997
i = 970, k = 4110
i = 971, k = 2232
i = 972, k = 1347
i = 973, k = 2105
i = 974, k = 3947
i = 975, k = 2417
i = 976, k = 2774
i = 977, k = 943
i = 978, k = 3836
i = 979, k = 616
i = 980, k = 1153
i = 981, k = 1848
i = 982, k = 3255
i = 983, k = 2996
i = 984, k = 739
i = 985, k = 3232
i = 986, k = 114
i = 987, k = 2166
i = 988, k = 1012
i = 989, k = 3725
i = 990, k = 3019
i = 991, k = 2147
i = 992, k = 3121
i = 993, k = 3043
i = 994, k = 887
i = 995, k = 1915
i = 996, k = 3862
i = 997, k = 205
i = 998, k = 2687
i = 999, k = 1813
```

In [169…

```python
# calculate average and sd
for section in scores.keys():
    print("Section =", section)
    score_avg = np.average(scores[section])
    score_std = np.std(scores[section])
    print(f"Count = {len(scores[section])}, Average = {score_avg:.2f}, std = {score
    print("-----------")
```

```
Section = abstract
Count = 1000, Average = 0.68, std = 0.09
-----------
Section = introduction
Count = 993, Average = 0.73, std = 0.08
-----------
Section = results
Count = 922, Average = 0.62, std = 0.10
-----------
Section = discussion
Count = 927, Average = 0.69, std = 0.09
-----------
Section = materials and methods
Count = 887, Average = 0.49, std = 0.10
-----------
Section = materials and methods
Count = 64, Average = 0.49, std = 0.09
-----------
Section = material and methods
Count = 22, Average = 0.46, std = 0.08
-----------
Section = methods
Count = 5, Average = 0.58, std = 0.07
-----------
Section = results and discussion
Count = 55, Average = 0.66, std = 0.08
-----------
Section = acknowledgments
Count = 1, Average = 0.11, std = 0.00
-----------
Section = materials and methods
Count = 3, Average = 0.49, std = 0.07
-----------
Section = main text
Count = 3, Average = 0.74, std = 0.03
-----------
Section = materials
Count = 3, Average = 0.49, std = 0.05
-----------
Section = set-up of the quorum-sensing model
Count = 1, Average = 0.77, std = 0.00
-----------
Section = results of numerical simulations
Count = 1, Average = 0.62, std = 0.00
-----------
Section = results of mathematical analysis
Count = 1, Average = 0.60, std = 0.00
-----------
Section = conclusion
Count = 3, Average = 0.70, std = 0.10
-----------
Section = analysis
Count = 1, Average = 0.54, std = 0.00
-----------
Section = accession numbers
Count = 1, Average = 0.19, std = 0.00
```

```
-----------
Section = result
Count = 1, Average = 0.67, std = 0.00
-----------
Section = materials and methods
Count = 2, Average = 0.50, std = 0.00
-----------
Section = model
Count = 3, Average = 0.65, std = 0.03
-----------
Section = conclusions
Count = 3, Average = 0.67, std = 0.08
-----------
Section = supplemental information
Count = 1, Average = 0.04, std = 0.00
-----------
Section = results and conclusions
Count = 1, Average = 0.65, std = 0.00
-----------
Section = m
Count = 1, Average = 0.43, std = 0.00
-----------
```

In [ ]: