```python
"""Hello :)"""
from collections import Counter
import numpy as np
from tqdm import tqdm

with open("movies-pp.txt", "r", encoding="utf-8") as source_file:
    full_lines: list = source_file.readlines()
full_reviews: list   = [line.split() for line in full_lines] #list[list[words]]
small_reviews: list  = full_reviews[0:100] # Used reduced corpus of first 100 documents for development set
```

```python
"""Parameters"""
n_topics: int   = 20
iterations: int = 500
alpha: float    = 0.02
beta: float     = 0.1
```

```python
"""Full corpus setup and initialization"""
count_document_topic = [Counter() for _ in range(n_topics)] # initialize counters
count_word_topic     = [Counter() for _ in range(n_topics)]
count_topic          = Counter()
topic_assignment     = list() # z[d][i] - topic assignment of ith word in document d

for d, document in enumerate(full_reviews):
    topic_assignment.append(np.random.randint(n_topics, size = len(document))) # randomly initialize z
    for i, word in enumerate(document): # increment counters' initial counts
        topic = topic_assignment[d][i]
        count_document_topic[topic][d] += 1
        count_word_topic[topic][word]  += 1
        count_topic[topic]             += 1

vocabulary: set = {word for document in full_reviews for word in document}
W: int          = len(vocabulary)
P: list         = np.zeros(n_topics)
T: list         = list(range(n_topics))
w_beta: float   = (W * beta) # just wanted to move this outside the inner loop
```

```python
"""Gibbs Sampler algorithm"""
for n in tqdm(range(iterations)):
    for d, document in enumerate(full_reviews):
        for i, word in enumerate(document):# for i = 0 -> N-1 do

            topic = topic_assignment[d][i] # topic assigned to ith word of dth document
            count_document_topic[topic][d] -= 1 # w(-i) - count ignoring current instance
            count_word_topic[topic][word]  -= 1
            count_topic[topic]             -= 1
            denom_2 = len(document)-1 + (n_topics * alpha) # just wanted to move this outside the inner loop

            for t in range(n_topics): # Building probability distribution based on expected counts
                """P(z_i = t | z_-i, w) =ish    (n^wi_-i,t + Beta)  * (n^di_-i,t + alpha) /
                                                (n^._-i,t + W*Beta) * (n^di_-i,. + T*alpha)"""
                P[t] = ((count_word_topic[t][word] + beta) * (count_document_topic[t][d]  +  alpha)) \
                       / ((count_topic[t]     +   w_beta) * (denom_2))
                if P[t] < 0: # prevent negative probabilities that can occur if count < 0
                    P[t] = 0

            P = P/sum(P) # normalise distribution for sampling
            topic = np.random.choice(T, p=P) # topic <- sample from p(z|.)
            topic_assignment[d][i] = topic # z[d][i] <- topic - reassign topic based on sample
            count_document_topic[topic][d] += 1
            count_word_topic[topic][word]  += 1
            count_topic[topic]             += 1

    if n % 100 == 0: # Interesting to see how it evolves between iterations
        print("Glimpse of topics at iteration number ", n)
        for j, counter in enumerate(count_word_topic):
            print(f"Topic {j} top 10: ", counter.most_common(10))
```

```python
for i, counter in enumerate(count_word_topic):
    print(f"Topic {i} top 10: ", counter.most_common(10))
```

Discussion (full output for topic top 10 words below)

Here's a pretty cohesive topic sample from the run with the parameters suggested in the assignment:

Topic 18 top 10: [('alien', 328), ('aliens', 201), ('space', 154), ('science', 147), ('earth', 134), ('mars', 134), ('planet', 118), ('effects', 113), ('sci-fi', 106), ('computer', 98)]

And here's a much less cohesive topic from the same run:

Topic 1 top 10: [('film', 7914), ('movie', 4523), ('one', 4509), ('like', 2892), ('even', 2133), ('would', 1910), ('good', 1910), ('time', 1855), ('much', 1723), ('story', 1546)]

Even in categories that are otherwise quite cohesive, there are still question mark words:

Topic 12 top 10: [('jackie', 170), ('chan', 146), ('action', 88), ('martial', 72), ('chinese', 71), ('hong', 69), ('kong', 64), ('arts', 61), ('bulworth', 59)???, ('nbsp', 58)???]

From run 2:

With 25 topics, and swapping the alpha and beta values, we still ended up with some pretty cohesive-seeming categories, like this one which seems to be mostly people's names:

Topic 7 top 10: [('truman', 163), ('show', 125), ('murphy', 125), ('carrey', 125), ('eddie', 99), ('west', 99), ('jim', 98), ('williams', 96), ('patch', 78), ('wild', 69)]

And then there's this utterly bizaare specimen:

Topic 18 top 10: [('tom', 100), ('frank', 77), ('park', 59), ('scale', 56), ('0', 54), ('willis', 53), ('cole', 53),('wife', 51), ('baseball', 49), ('chicken', 47)]

Conclusion:

Over all, I think both sets of parameters came up with some fairly reasonable topic groupins, though some are just first name + last name pairs or movie titles, and there are definitely some strange words in each topic for both runs. It seems the suggested alpha and beta values converged on somewhat more logical groupings, but that's hard to quantify. As well, some topics seem to have two distinct groupings within them, namely this one from run 1, which appears to lump movie ratings in with proper names:

Topic 3 top 10: [('allen', 69), ('nick', 62), ('8/10', 59), ('zero', 56), ('7/10', 54), ('5/10', 51), ('jerry', 47), ('alice', 47), ('sonny', 42), ('critique', 42)]

First run final iteration

```
"""Parameters"""
n_topics: int = 20
iterations: int = 500
alpha: float = 0.02
beta: float = 0.1

Topic 0 top 10:  [('simon', 110), ('shakespeare', 84), ('oscar', 64), ('bobby', 57), ('dog',
52), ('ghost', 47), ('queen', 39), ('jack', 38), ('elizabeth', 38), ('melvin', 37)]
Topic 1 top 10:  [('film', 7914), ('movie', 4523), ('one', 4509), ('like', 2892), ('even',
2133), ('would', 1910), ('good', 1910), ('time', 1855), ('much', 1723), ('story', 1546)]
Topic 2 top 10:  [('girls', 94), ('flynt', 80), ('max', 75), ('spawn', 71), ('larry', 60),
('spice', 58), ('crystal', 49), ('kudrow', 39), ('krippendorf', 38), ('tribe', 35)]
Topic 3 top 10:  [('allen', 69), ('nick', 62), ('8/10', 59), ('zero', 56), ('7/10', 54),
('5/10', 51), ('jerry', 47), ('alice', 47), ('sonny', 42), ('critique', 42)]
Topic 4 top 10:  [('house', 73), ('blair', 66), ('witch', 63), ('nights', 56), ('boogie',
49), ('54', 49), ('porn', 46), ('scary', 45), ('haunting', 43), ('hill', 41)]
Topic 5 top 10:  [('tarantino', 111), ('jackie', 95), ('apes', 72), ('burton', 69), ('de',
```

66), ('pulp', 62), ('snake', 61), ('fiction', 58), ('brown', 58), ('planet', 53)]
Topic 6 top 10: [('action', 217), ('batman', 145), ('arnold', 97), ('vampire', 95), ('van', 89), ('vampires', 75), ('damme', 74), ('blade', 72), ('robin', 70), ('schwarzenegger', 64)]
Topic 7 top 10: [('disney', 249), ('truman', 143), ('animated', 135), ('animation', 124), ('voice', 98), ('mulan', 96), ('children', 87), ('toy', 85), ('carrey', 84), ('family', 72)]
Topic 8 top 10: [('star', 238), ('wars', 140), ('trek', 125), ('lucas', 98), ('phantom', 79), ('jedi', 77), ('menace', 73), ('effects', 62), ('special', 60), ('jar', 47)]
Topic 9 top 10: [('harry', 125), ('carter', 74), ('cop', 71), ('bacon', 58), ('lebowski', 56), ('police', 55), ('dude', 53), ('murder', 53), ('derek', 49), ('wild', 47)]
Topic 10 top 10: [('movie', 1144), ('film', 980), ('one', 787), ('comedy', 666), ('like', 648), ('funny', 566), ('get', 461), ('even', 417), ('good', 404), ('two', 355)]
Topic 11 top 10: [('bond', 84), ('dvd', 70), ('evil', 69), ('wild', 65), ('austin', 63), ('dr', 57), ('powers', 53), ('wrestling', 50), ('original', 47), ('west', 46)]
Topic 12 top 10: [('jackie', 170), ('chan', 146), ('action', 88), ('martial', 72), ('chinese', 71), ('hong', 69), ('kong', 64), ('arts', 61), ('bulworth', 59), ('nbsp', 58)]
Topic 13 top 10: [('war', 250), ('godzilla', 109), ('battle', 88), ('men', 82), ('troopers', 77), ('army', 76), ('ryan', 74), ('soldiers', 73), ('military', 71), ('starship', 65)]
Topic 14 top 10: [('annie', 64), ('judd', 47), ('husband', 43), ('jones', 40), ('ford', 40), ('fugitive', 37), ('tom', 37), ('carry', 33), ('grace', 33), ('horse', 32)]
Topic 15 top 10: [('scream', 219), ('horror', 142), ('school', 97), ('killer', 80), ('2', 77), ('julie', 76), ('smith', 75), ('summer', 70), ('teen', 69), ('high', 67)]
Topic 16 top 10: [('joe', 148), ('tarzan', 112), ('lynch', 64), ('species', 47), ('jane', 40), ('granger', 35), ('paulie', 35), ('gorilla', 29), ('jungle', 29), ('patrick', 29)]
Topic 17 top 10: [('titanic', 95), ('ship', 84), ('wedding', 58), ('angels', 53), ('kate', 51), ('julia', 50), ('love', 47), ('rose', 47), ('cameron', 46), ('barrymore', 36)]
Topic 18 top 10: [('alien', 328), ('aliens', 201), ('space', 154), ('science', 147), ('earth', 134), ('mars', 134), ('planet', 118), ('effects', 113), ('sci-fi', 106), ('computer', 98)]
Topic 19 top 10: [('life', 691), ('film', 548), ('story', 383), ('love', 352), ('man', 294), ('one', 279), ('world', 274), ('family', 272), ('character', 232), ('young', 230)]

Second run final iteration

```
"""Parameters"""
n_topics: int = 25
iterations: int = 500
alpha: float = 0.1
beta: float = 0.02
```

Topic 0 top 10: [('war', 242), ('action', 172), ('jackie', 160), ('chan', 147), ('men', 141), ('spielberg', 114), ('fight', 108), ('army', 94), ('chinese', 92), ('films', 91)]
Topic 1 top 10: [('american', 170), ('political', 151), ('government', 107), ('america', 94), ('president', 85), ('black', 82), ('war', 81), ('even', 78), ('hollywood', 78), ('make', 69)]
Topic 2 top 10: [('sex', 258), ('scene', 171), ('man', 148), ('way', 118), ('camera', 116), ('violence', 116), ('kevin', 106), ('women', 100), ('look', 98), ('girl', 94)]
Topic 3 top 10: [('ship', 183), ('godzilla', 126), ('titanic', 115), ('cameron', 86), ('water', 76), ('effects', 71), ('disaster', 71), ('new', 67), ('monster', 65), ('york', 63)]
Topic 4 top 10: [('joe', 235), ('harry', 150), ('town', 118), ('grace', 93), ('derek', 66), ('man', 53), ('horse', 51), ('new', 50), ('country', 48), ('york', 47)]
Topic 5 top 10: [('king', 102), ('musical', 98), ('girls', 96), ('shakespeare', 95), ('spice', 80), ('music', 76), ('songs', 74), ('love', 69), ('band', 67), ('queen', 61)]
Topic 6 top 10: [('star', 343), ('alien', 226), ('wars', 179), ('effects', 172), ('series', 156), ('trek', 150), ('special', 131), ('aliens', 110), ('lucas', 105), ('menace', 98)]
Topic 7 top 10: [('truman', 163), ('show', 125), ('murphy', 125), ('carrey', 125), ('eddie', 99), ('west', 99), ('jim', 98), ('williams', 96), ('patch', 78), ('wild', 69)]
Topic 8 top 10: [('disney', 276), ('animated', 156), ('story', 144), ('family', 141), ('animation', 133), ('children', 124), ('voice', 119), ('tarzan', 118), ('kids', 116), ('toy', 101)]
Topic 9 top 10: [('movie', 5671), ('film', 5068), ('one', 3799), ('like', 3365), ('even', 2478), ('good', 2147), ('get', 1919), ('would', 1820), ('time', 1816), ('see', 1638)]
Topic 10 top 10: [('mr', 165), ('sequence', 87), ('lynch', 68), ('opening', 52), ('upon', 51), ('minutes', 49), ('original', 49), ('ms', 48), ('segment', 46), ('jolie', 46)]

```
Topic 11 top 10:  [('nick', 101), ('bond', 93), ('gibson', 90), ('sam', 77), ('bill', 67),
('alice', 65), ('character', 64), ('mel', 60), ('wild', 59), ('james', 50)]
Topic 12 top 10:  [('world', 199), ('city', 132), ('life', 115), ('science', 101),
('reality', 100), ('death', 98), ('dark', 96), ('angels', 82), ('human', 75), ('game', 74)]
Topic 13 top 10:  [('comedy', 189), ('romantic', 180), ('love', 154), ('wedding', 123),
('ben', 92), ('julia', 90), ('life', 89), ('two', 88), ('ryan', 88), ('sandler', 77)]
Topic 14 top 10:  [('comedy', 398), ('funny', 395), ('jokes', 184), ('humor', 166),
('laughs', 127), ('smith', 119), ('bob', 109), ('brothers', 104), ('hilarious', 103),
('comic', 102)]
Topic 15 top 10:  [('evil', 139), ('action', 106), ('powers', 86), ('arnold', 84), ('dr',
80), ('lee', 73), ('jones', 73), ('austin', 69), ('agent', 68), ('bruce', 68)]
Topic 16 top 10:  [('action', 273), ('effects', 132), ('mission', 119), ('van', 113),
('vampire', 103), ('special', 97), ('team', 94), ('blade', 92), ('computer', 90), ('john',
85)]
Topic 17 top 10:  [('family', 312), ('life', 265), ('mother', 220), ('father', 205),
('love', 196), ('relationship', 147), ('son', 142), ('daughter', 127), ('husband', 124),
('child', 116)]
Topic 18 top 10:  [('tom', 100), ('frank', 77), ('park', 59), ('scale', 56), ('0', 54),
('willis', 53), ('cole', 53), ('wife', 51), ('baseball', 49), ('chicken', 47)]
Topic 19 top 10:  [('school', 244), ('high', 188), ('big', 106), ('teen', 106), ('club',
91), ('john', 84), ('dude', 83), ('mike', 81), ('girls', 74), ('girl', 66)]
Topic 20 top 10:  [('earth', 231), ('planet', 184), ('effects', 118), ('space', 112),
('alien', 109), ('human', 86), ('humans', 82), ('deep', 79), ('special', 79), ('troopers',
79)]
Topic 21 top 10:  [('crime', 150), ('cop', 146), ('tarantino', 117), ('jackie', 110),
('police', 104), ('de', 99), ('action', 93), ('drug', 93), ('pulp', 91), ('jack', 90)]
Topic 22 top 10:  [('film', 4235), ('one', 1711), ('story', 1364), ('character', 990),
('characters', 928), ('also', 891), ('two', 683), ('life', 653), ('however', 650),
('director', 642)]
Topic 23 top 10:  [('horror', 332), ('scream', 246), ('killer', 164), ('original', 155),
('film', 140), ('2', 133), ('scary', 116), ('house', 105), ('sequel', 98), ('first', 94)]
Topic 24 top 10:  [('batman', 195), ('robin', 113), ('cage', 80), ('spawn', 80), ('comic',
57), ('schumacher', 57), ('mr', 49), ('tim', 48), ('michael', 47), ('joel', 45)]
```

Run 2 data every 100 iterations:

Glimpse of topics at iteration number 0:

```
Topic 0 top 10:  [('one', 220), ('film', 211), ('movie', 191), ('could', 150), ('even',
141), ('get', 135), ('first', 122), ('much', 118), ('actors', 117), ('like', 114)]
Topic 1 top 10:  [('film', 510), ('one', 254), ('movie', 217), ('characters', 202),
('would', 195), ('like', 162), ('get', 159), ('also', 146), ('first', 142), ('seems', 132)]
Topic 2 top 10:  [('film', 416), ('movie', 416), ('like', 248), ('one', 178), ('sex', 161),
('every', 132), ('also', 117), ('something', 111), ('know', 107), ('kind', 107)]
Topic 3 top 10:  [('film', 410), ('movie', 224), ('time', 224), ('get', 182), ('one', 159),
('like', 140), ('two', 127), ('made', 120), ('even', 119), ('good', 112)]
Topic 4 top 10:  [('film', 342), ('movie', 187), ('one', 186), ('like', 165), ('man', 134),
('although', 122), ('black', 110), ('comes', 105), ('york', 99), ('looks', 97)]
Topic 5 top 10:  [('film', 311), ('movie', 255), ('one', 227), ('like', 178), ('much', 160),
('really', 127), ('man', 116), ('also', 116), ('story', 108), ('even', 101)]
Topic 6 top 10:  [('film', 419), ('one', 255), ('movie', 249), ('films', 144), ('well',
143), ('story', 112), ('back', 109), ('scene', 108), ('like', 105), ('comedy', 105)]
Topic 7 top 10:  [('film', 376), ('movie', 304), ('one', 201), ('like', 180), ('would',
159), ('character', 148), ('good', 133), ('us', 127), ('way', 125), ('even', 123)]
Topic 8 top 10:  [('film', 432), ('one', 385), ('good', 179), ('like', 178), ('story', 124),
('back', 120), ('original', 119), ('gets', 114), ('movie', 113), ('john', 109)]
Topic 9 top 10:  [('film', 303), ('movie', 299), ('one', 276), ('like', 160), ('plot', 156),
('two', 154), ('even', 142), ('seen', 136), ('time', 134), ('get', 132)]
Topic 10 top 10:  [('film', 332), ('movie', 297), ('one', 193), ('characters', 163),
('story', 129), ('plot', 120), ('really', 117), ('two', 115), ('way', 112), ('many', 104)]
Topic 11 top 10:  [('film', 501), ('one', 296), ('movie', 225), ('story', 154), ('like',
146), ('time', 119), ('bad', 106), ('gives', 106), ('carrey', 104), ('good', 103)]
Topic 12 top 10:  [('film', 312), ('movie', 280), ('characters', 147), ('great', 146),
('one', 144), ('best', 125), ('story', 123), ('big', 120), ('better', 117), ('see', 113)]
```

```
Topic 13 top 10:  [('film', 386), ('movie', 285), ('like', 237), ('one', 203), ('still',
169), ('character', 156), ('time', 145), ('films', 118), ('two', 115), ('story', 102)]
Topic 14 top 10:  [('film', 401), ('one', 169), ('also', 146), ('like', 141), ('would',
134), ('story', 131), ('even', 114), ('character', 111), ('first', 110), ('get', 104)]
Topic 15 top 10:  [('film', 456), ('one', 389), ('movie', 236), ('good', 160), ('character',
153), ('movies', 151), ('time', 138), ('right', 116), ('well', 111), ('almost', 103)]
Topic 16 top 10:  [('film', 354), ('one', 248), ('movie', 242), ('even', 180), ('like',
167), ('could', 131), ('times', 113), ('well', 112), ('story', 108), ('plot', 97)]
Topic 17 top 10:  [('film', 359), ('movie', 246), ('one', 212), ('like', 181), ('get', 146),
('world', 139), ('love', 126), ('story', 125), ('even', 116), ('scene', 116)]
Topic 18 top 10:  [('film', 338), ('one', 268), ('movie', 224), ('first', 170), ('good',
162), ('films', 139), ('action', 116), ('time', 111), ('like', 106), ('story', 103)]
Topic 19 top 10:  [('movie', 225), ('one', 213), ('film', 190), ('like', 150), ('two', 142),
('funny', 129), ('would', 126), ('also', 120), ('make', 111), ('far', 109)]
Topic 20 top 10:  [('film', 273), ('movie', 149), ('good', 146), ('director', 122), ('time',
118), ('know', 116), ('also', 114), ('one', 112), ('plot', 111), ('would', 111)]
Topic 21 top 10:  [('film', 502), ('one', 240), ('like', 183), ('see', 151), ('movie', 146),
('make', 123), ('even', 119), ('would', 114), ('tarantino', 112), ('characters', 97)]
Topic 22 top 10:  [('film', 450), ('movie', 213), ('man', 163), ('one', 160), ('another',
144), ('us', 129), ('well', 128), ('tarzan', 117), ('see', 112), ('story', 101)]
Topic 23 top 10:  [('film', 480), ('movie', 287), ('one', 175), ('see', 165), ('character',
163), ('time', 156), ('even', 155), ('well', 138), ('bad', 136), ('director', 109)]
Topic 24 top 10:  [('film', 379), ('one', 217), ('much', 157), ('like', 148), ('movie',
129), ('trek', 118), ('never', 109), ('another', 109), ('star', 109), ('years', 102)]
```

Glimpse of topics at iteration number 100:

```
Topic 0 top 10:  [('war', 277), ('men', 230), ('american', 170), ('chan', 142), ('movie',
134), ('one', 133), ('ryan', 124), ('spielberg', 103), ('battle', 99), ('army', 98)]
Topic 1 top 10:  [('film', 898), ('story', 418), ('two', 335), ('could', 292), ('one', 281),
('characters', 273), ('would', 264), ('even', 250), ('much', 223), ('many', 207)]
Topic 2 top 10:  [('film', 1251), ('like', 634), ('one', 618), ('way', 429), ('time', 377),
('scene', 376), ('make', 358), ('man', 339), ('john', 304), ('something', 295)]
Topic 3 top 10:  [('make', 157), ('new', 150), ('godzilla', 132), ('ship', 131), ('cameron',
122), ('titanic', 121), ('jack', 120), ('disaster', 111), ('one', 108), ('first', 89)]
Topic 4 top 10:  [('joe', 205), ('one', 180), ('life', 159), ('harry', 158), ('black', 147),
('town', 144), ('man', 137), ('new', 126), ('old', 125), ('small', 112)]
Topic 5 top 10:  [('music', 124), ('wedding', 86), ('spice', 73), ('king', 72), ('band',
72), ('rock', 69), ('singer', 61), ('world', 58), ('songs', 57), ('like', 53)]
Topic 6 top 10:  [('star', 345), ('alien', 206), ('series', 199), ('effects', 192), ('wars',
179), ('trek', 151), ('special', 150), ('aliens', 116), ('lucas', 97), ('new', 96)]
Topic 7 top 10:  [('movie', 883), ('good', 475), ('script', 347), ('end', 332), ('make',
316), ('like', 301), ('film', 295), ('even', 273), ('big', 267), ('character', 258)]
Topic 8 top 10:  [('disney', 276), ('story', 219), ('family', 188), ('animated', 155),
('children', 142), ('little', 141), ('animation', 138), ('voice', 133), ('kids', 124),
('tarzan', 118)]
Topic 9 top 10:  [('movie', 3882), ('film', 2940), ('one', 2219), ('like', 1753), ('even',
1390), ('would', 1291), ('get', 1266), ('good', 1214), ('really', 1161), ('time', 1103)]
Topic 10 top 10:  [('film', 291), ('one', 162), ('though', 154), ('television', 137), ('tv',
130), ('however', 121), ('mr', 118), ('young', 116), ('screen', 113), ('show', 111)]
Topic 11 top 10:  [('nick', 125), ('character', 113), ('role', 100), ('gibson', 94),
('bond', 86), ('mel', 82), ('however', 77), ('director', 73), ('sam', 72), ('bacon', 67)]
Topic 12 top 10:  [('world', 413), ('life', 198), ('city', 196), ('human', 167), ('new',
156), ('reality', 128), ('science', 120), ('may', 113), ('effects', 109), ('audience', 105)]
Topic 13 top 10:  [('love', 241), ('comedy', 239), ('romantic', 202), ('one', 191), ('like',
139), ('time', 133), ('character', 133), ('little', 129), ('never', 128), ('characters',
120)]
Topic 14 top 10:  [('comedy', 341), ('funny', 337), ('humor', 196), ('two', 147), ('jokes',
146), ('bob', 125), ('like', 122), ('simon', 119), ('laugh', 107), ('laughs', 103)]
Topic 15 top 10:  [('action', 162), ('evil', 161), ('king', 104), ('movie', 103), ('lee',
102), ('powers', 92), ('hero', 89), ('dr', 87), ('make', 79), ('austin', 71)]
Topic 16 top 10:  [('action', 251), ('effects', 162), ('mission', 156), ('crew', 155),
('film', 145), ('mars', 145), ('van', 138), ('computer', 123), ('vampire', 114), ('team',
112)]
```

Topic 17 top 10: [('life', 377), ('family', 296), ('love', 273), ('mother', 247), ('father', 235), ('son', 180), ('man', 174), ('husband', 164), ('woman', 155), ('wife', 146)]
Topic 18 top 10: [('story', 129), ('tom', 124), ('one', 120), ('park', 99), ('new', 84), ('wife', 82), ('willis', 79), ('reeves', 70), ('run', 70), ('horse', 60)]
Topic 19 top 10: [('high', 239), ('school', 208), ('girls', 194), ('sex', 146), ('big', 122), ('teen', 102), ('around', 101), ('john', 94), ('dude', 92), ('american', 91)]
Topic 20 top 10: [('action', 295), ('earth', 198), ('film', 175), ('planet', 171), ('plot', 137), ('alien', 123), ('dr', 102), ('aliens', 100), ('space', 92), ('military', 92)]
Topic 21 top 10: [('jackie', 182), ('crime', 155), ('cop', 131), ('robert', 122), ('jack', 122), ('police', 122), ('drug', 120), ('tarantino', 116), ('de', 114), ('plot', 110)]
Topic 22 top 10: [('film', 2835), ('one', 885), ('character', 706), ('also', 677), ('story', 648), ('characters', 531), ('films', 503), ('best', 495), ('great', 475), ('performance', 440)]
Topic 23 top 10: [('horror', 337), ('scream', 247), ('film', 240), ('killer', 214), ('first', 146), ('original', 145), ('2', 137), ('scary', 133), ('movie', 107), ('last', 106)]
Topic 24 top 10: [('batman', 188), ('one', 164), ('like', 158), ('robin', 109), ('cage', 106), ('world', 95), ('george', 88), ('life', 85), ('film', 83), ('frank', 78)]

Glimpse of topics at iteration number 200:

Topic 0 top 10: [('war', 286), ('men', 153), ('american', 151), ('chan', 147), ('jackie', 143), ('battle', 107), ('spielberg', 105), ('army', 99), ('world', 84), ('ryan', 82)]
Topic 1 top 10: [('would', 196), ('role', 185), ('new', 185), ('hollywood', 180), ('story', 174), ('many', 173), ('could', 161), ('one', 149), ('old', 120), ('even', 118)]
Topic 2 top 10: [('one', 481), ('man', 448), ('like', 385), ('film', 301), ('end', 271), ('way', 265), ('life', 247), ('scene', 243), ('something', 226), ('sex', 224)]
Topic 3 top 10: [('ship', 148), ('godzilla', 132), ('titanic', 120), ('new', 82), ('disaster', 80), ('cameron', 78), ('york', 70), ('rose', 58), ('studio', 57), ('jack', 56)]
Topic 4 top 10: [('joe', 228), ('town', 180), ('harry', 165), ('man', 103), ('black', 86), ('small', 85), ('old', 84), ('white', 70), ('derek', 67), ('grace', 57)]
Topic 5 top 10: [('music', 106), ('musical', 99), ('girls', 93), ('king', 89), ('spice', 80), ('band', 79), ('songs', 72), ('rock', 70), ('love', 67), ('singer', 67)]
Topic 6 top 10: [('star', 314), ('alien', 240), ('wars', 179), ('effects', 163), ('series', 161), ('special', 139), ('trek', 139), ('aliens', 136), ('lucas', 103), ('menace', 92)]
Topic 7 top 10: [('big', 273), ('script', 266), ('end', 239), ('man', 231), ('director', 212), ('show', 204), ('new', 203), ('role', 180), ('lot', 178), ('character', 173)]
Topic 8 top 10: [('disney', 277), ('story', 160), ('animated', 148), ('family', 137), ('animation', 137), ('children', 132), ('voice', 129), ('tarzan', 118), ('kids', 109), ('original', 109)]
Topic 9 top 10: [('movie', 5489), ('film', 4972), ('one', 3974), ('like', 2932), ('even', 2338), ('would', 1954), ('get', 1922), ('good', 1884), ('time', 1800), ('really', 1555)]
Topic 10 top 10: [('film', 213), ('mr', 174), ('show', 121), ('series', 116), ('new', 109), ('sequence', 102), ('tv', 100), ('television', 93), ('release', 92), ('audience', 86)]
Topic 11 top 10: [('nick', 110), ('bond', 106), ('gibson', 89), ('mel', 73), ('henry', 71), ('species', 63), ('wild', 58), ('sam', 58), ('lee', 58), ('sex', 50)]
Topic 12 top 10: [('world', 343), ('city', 139), ('human', 132), ('science', 121), ('reality', 115), ('life', 100), ('effects', 97), ('dark', 92), ('special', 81), ('death', 78)]
Topic 13 top 10: [('romantic', 202), ('comedy', 194), ('love', 185), ('wedding', 112), ('ryan', 97), ('sweet', 89), ('ben', 81), ('tom', 78), ('romance', 78), ('julia', 76)]
Topic 14 top 10: [('funny', 375), ('comedy', 370), ('humor', 182), ('jokes', 161), ('laughs', 126), ('bob', 124), ('comic', 116), ('laugh', 113), ('max', 86), ('jay', 80)]
Topic 15 top 10: [('evil', 172), ('action', 103), ('powers', 96), ('lee', 85), ('dr', 77), ('jones', 73), ('austin', 72), ('king', 70), ('movie', 62), ('wrestling', 55)]
Topic 16 top 10: [('action', 318), ('mars', 139), ('mission', 126), ('effects', 124), ('van', 117), ('vampire', 103), ('crew', 99), ('team', 96), ('special', 90), ('computer', 83)]
Topic 17 top 10: [('life', 370), ('family', 343), ('mother', 264), ('father', 254), ('love', 245), ('son', 185), ('people', 183), ('woman', 151), ('wife', 148), ('relationship', 147)]
Topic 18 top 10: [('tom', 109), ('frank', 90), ('willis', 76), ('park', 63), ('horse', 59), ('run', 55), ('reeves', 52), ('sixth', 50), ('chicken', 49), ('new', 47)]

```
Topic 19 top 10:  [('school', 239), ('high', 205), ('sex', 111), ('girls', 100), ('big',
98), ('teen', 97), ('dude', 90), ('comedy', 72), ('characters', 71), ('mike', 66)]
Topic 20 top 10:  [('earth', 215), ('planet', 171), ('action', 161), ('effects', 98),
('space', 90), ('alien', 88), ('troopers', 80), ('aliens', 77), ('human', 72), ('special',
71)]
Topic 21 top 10:  [('cop', 149), ('crime', 142), ('jackie', 128), ('tarantino', 117),
('police', 112), ('drug', 112), ('de', 100), ('jack', 96), ('robert', 93), ('fiction', 91)]
Topic 22 top 10:  [('film', 3718), ('one', 944), ('story', 929), ('character', 873),
('characters', 867), ('also', 734), ('however', 636), ('best', 623), ('films', 620),
('performance', 534)]
Topic 23 top 10:  [('horror', 347), ('scream', 247), ('killer', 193), ('2', 152), ('film',
141), ('scary', 141), ('original', 137), ('house', 98), ('first', 97), ('know', 94)]
Topic 24 top 10:  [('batman', 182), ('robin', 124), ('george', 112), ('spawn', 82), ('cage',
81), ('mr', 71), ('clooney', 59), ('schumacher', 57), ('joel', 51), ('wife', 49)]
```

Glimpse of topics at iteration number 300:

```
Topic 0 top 10:  [('war', 264), ('jackie', 160), ('chan', 147), ('men', 137), ('spielberg',
118), ('american', 108), ('action', 102), ('army', 91), ('battle', 86), ('chinese', 83)]
Topic 1 top 10:  [('even', 161), ('american', 160), ('new', 146), ('hollywood', 145),
('political', 144), ('world', 139), ('could', 132), ('would', 129), ('government', 114),
('history', 110)]
Topic 2 top 10:  [('man', 368), ('like', 263), ('sex', 249), ('way', 236), ('one', 235),
('scene', 212), ('film', 190), ('violence', 186), ('look', 179), ('something', 148)]
Topic 3 top 10:  [('ship', 158), ('godzilla', 125), ('titanic', 113), ('new', 79),
('cameron', 79), ('disaster', 75), ('effects', 73), ('york', 64), ('water', 57), ('studio',
56)]
Topic 4 top 10:  [('joe', 234), ('harry', 166), ('town', 141), ('grace', 90), ('man', 73),
('derek', 65), ('country', 53), ('small', 52), ('men', 52), ('new', 49)]
Topic 5 top 10:  [('girls', 103), ('musical', 96), ('king', 92), ('band', 77), ('music',
77), ('spice', 76), ('rock', 70), ('songs', 69), ('love', 61), ('joan', 57)]
Topic 6 top 10:  [('star', 344), ('alien', 244), ('wars', 179), ('series', 173), ('effects',
160), ('trek', 146), ('aliens', 136), ('special', 128), ('lucas', 101), ('menace', 87)]
Topic 7 top 10:  [('big', 206), ('script', 179), ('man', 160), ('truman', 159), ('new',
148), ('show', 143), ('end', 140), ('director', 139), ('murphy', 138), ('carrey', 125)]
Topic 8 top 10:  [('disney', 277), ('animated', 158), ('story', 147), ('animation', 135),
('kids', 123), ('voice', 120), ('tarzan', 118), ('children', 115), ('allen', 108),
('family', 106)]
Topic 9 top 10:  [('movie', 5661), ('film', 4998), ('one', 3719), ('like', 3241), ('even',
2384), ('would', 2021), ('good', 1997), ('get', 1907), ('time', 1857), ('really', 1550)]
Topic 10 top 10:  [('film', 213), ('mr', 199), ('sequence', 105), ('tv', 92), ('series',
81), ('lynch', 70), ('new', 65), ('show', 64), ('feature', 62), ('television', 61)]
Topic 11 top 10:  [('nick', 97), ('bond', 94), ('gibson', 89), ('mel', 74), ('wild', 66),
('alice', 56), ('character', 53), ('things', 52), ('sam', 52), ('sex', 51)]
Topic 12 top 10:  [('world', 276), ('city', 135), ('life', 113), ('human', 111), ('science',
108), ('reality', 107), ('angels', 82), ('dark', 80), ('death', 69), ('new', 60)]
Topic 13 top 10:  [('romantic', 182), ('love', 168), ('comedy', 166), ('wedding', 111),
('ryan', 107), ('julia', 93), ('ben', 89), ('tom', 79), ('sweet', 78), ('romance', 75)]
Topic 14 top 10:  [('funny', 418), ('comedy', 382), ('humor', 189), ('jokes', 173), ('bob',
123), ('hilarious', 107), ('laughs', 102), ('laugh', 102), ('comic', 94), ('max', 93)]
Topic 15 top 10:  [('evil', 137), ('action', 126), ('powers', 84), ('arnold', 80), ('jones',
68), ('austin', 67), ('dr', 64), ('lee', 64), ('king', 60), ('bruce', 56)]
Topic 16 top 10:  [('action', 299), ('van', 128), ('vampire', 102), ('effects', 99),
('mission', 98), ('blade', 93), ('computer', 90), ('team', 87), ('damme', 81), ('fight',
79)]
Topic 17 top 10:  [('life', 389), ('family', 346), ('mother', 257), ('love', 233),
('father', 209), ('son', 161), ('home', 158), ('relationship', 151), ('daughter', 142),
('people', 133)]
Topic 18 top 10:  [('tom', 87), ('willis', 76), ('frank', 75), ('scale', 60), ('0', 53),
('park', 52), ('cole', 50), ('wife', 47), ('team', 47), ('baldwin', 44)]
Topic 19 top 10:  [('school', 223), ('high', 193), ('big', 120), ('john', 98), ('teen', 93),
('sex', 92), ('dude', 87), ('girls', 84), ('comedy', 83), ('characters', 72)]
Topic 20 top 10:  [('earth', 219), ('planet', 172), ('effects', 124), ('space', 120),
('alien', 93), ('action', 89), ('human', 85), ('mars', 84), ('troopers', 81), ('starship',
```

73)]
Topic 21 top 10: [('cop', 157), ('crime', 146), ('drug', 117), ('tarantino', 117), ('jackie', 110), ('police', 97), ('fiction', 94), ('l', 94), ('action', 91), ('pulp', 91)]
Topic 22 top 10: [('film', 3955), ('one', 1613), ('characters', 1038), ('character', 1030), ('story', 1006), ('also', 830), ('best', 669), ('performance', 663), ('however', 643), ('end', 642)]
Topic 23 top 10: [('horror', 349), ('scream', 240), ('killer', 189), ('scary', 142), ('original', 129), ('2', 122), ('first', 116), ('sequel', 98), ('house', 97), ('blair', 90)]
Topic 24 top 10: [('batman', 184), ('robin', 125), ('george', 81), ('spawn', 80), ('mr', 73), ('cage', 68), ('comic', 57), ('schumacher', 57), ('hand', 47), ('clooney', 45)]

Glimpse of topics at iteration number 400:

Topic 0 top 10: [('war', 220), ('jackie', 155), ('chan', 147), ('action', 140), ('men', 140), ('spielberg', 109), ('fight', 105), ('battle', 102), ('army', 97), ('chinese', 95)]
Topic 1 top 10: [('american', 177), ('political', 132), ('hollywood', 120), ('government', 110), ('black', 108), ('president', 96), ('war', 91), ('world', 90), ('america', 86), ('history', 79)]
Topic 2 top 10: [('sex', 262), ('man', 218), ('like', 162), ('scene', 154), ('violence', 138), ('one', 137), ('life', 135), ('women', 129), ('white', 126), ('director', 112)]
Topic 3 top 10: [('ship', 171), ('titanic', 120), ('godzilla', 119), ('cameron', 84), ('water', 82), ('disaster', 81), ('effects', 74), ('york', 68), ('new', 67), ('boat', 55)]
Topic 4 top 10: [('joe', 243), ('harry', 146), ('town', 116), ('grace', 102), ('man', 71), ('new', 59), ('derek', 53), ('lake', 46), ('city', 44), ('country', 43)]
Topic 5 top 10: [('music', 112), ('king', 111), ('girls', 103), ('musical', 102), ('shakespeare', 96), ('spice', 81), ('band', 76), ('love', 76), ('songs', 72), ('queen', 61)]
Topic 6 top 10: [('star', 339), ('alien', 240), ('series', 188), ('wars', 179), ('effects', 163), ('trek', 156), ('aliens', 134), ('special', 128), ('lucas', 103), ('menace', 99)]
Topic 7 top 10: [('truman', 163), ('show', 158), ('murphy', 126), ('carrey', 125), ('director', 102), ('west', 96), ('jim', 92), ('smith', 88), ('eddie', 87), ('big', 84)]
Topic 8 top 10: [('disney', 276), ('animated', 154), ('animation', 134), ('story', 133), ('family', 133), ('kids', 128), ('voice', 126), ('children', 119), ('tarzan', 118), ('toy', 104)]
Topic 9 top 10: [('movie', 5670), ('film', 4988), ('one', 4004), ('like', 3327), ('even', 2483), ('good', 2250), ('get', 1922), ('would', 1896), ('time', 1893), ('see', 1632)]
Topic 10 top 10: [('mr', 185), ('sequence', 78), ('new', 76), ('series', 74), ('lynch', 65), ('feature', 59), ('original', 56), ('ms', 53), ('segment', 52), ('note', 51)]
Topic 11 top 10: [('nick', 109), ('gibson', 96), ('bond', 94), ('mel', 82), ('sam', 69), ('alice', 65), ('wild', 60), ('things', 54), ('species', 54), ('sex', 52)]
Topic 12 top 10: [('world', 236), ('city', 128), ('science', 117), ('reality', 117), ('dark', 103), ('life', 92), ('human', 90), ('death', 88), ('angels', 85), ('new', 77)]
Topic 13 top 10: [('love', 206), ('romantic', 186), ('comedy', 168), ('wedding', 120), ('ryan', 110), ('gets', 89), ('julia', 88), ('gay', 77), ('roberts', 71), ('relationship', 71)]
Topic 14 top 10: [('comedy', 435), ('funny', 411), ('jokes', 173), ('humor', 161), ('laughs', 128), ('bob', 121), ('smith', 113), ('brothers', 104), ('comic', 98), ('hilarious', 87)]
Topic 15 top 10: [('evil', 147), ('action', 113), ('lee', 84), ('powers', 84), ('arnold', 82), ('dr', 81), ('end', 72), ('austin', 68), ('jones', 65), ('bruce', 57)]
Topic 16 top 10: [('action', 315), ('effects', 141), ('van', 110), ('mission', 105), ('vampire', 104), ('special', 103), ('computer', 91), ('team', 90), ('blade', 88), ('damme', 81)]
Topic 17 top 10: [('life', 320), ('family', 300), ('mother', 251), ('love', 204), ('father', 197), ('son', 169), ('relationship', 151), ('husband', 137), ('home', 137), ('children', 130)]
Topic 18 top 10: [('tom', 84), ('frank', 82), ('willis', 67), ('park', 62), ('0', 54), ('scale', 51), ('baseball', 50), ('horse', 49), ('annie', 49), ('cole', 49)]
Topic 19 top 10: [('school', 242), ('high', 205), ('teen', 98), ('characters', 92), ('dude', 92), ('club', 91), ('girl', 89), ('girls', 85), ('john', 83), ('mike', 81)]
Topic 20 top 10: [('earth', 214), ('planet', 160), ('effects', 132), ('special', 105), ('alien', 105), ('space', 98), ('action', 90), ('troopers', 81), ('mars', 78), ('seagal', 75)]
Topic 21 top 10: [('crime', 153), ('cop', 145), ('tarantino', 117), ('jackie', 116),

('police', 108), ('drug', 100), ('michael', 99), ('money', 92), ('fiction', 91), ('pulp', 89)]
Topic 22 top 10: [('film', 4243), ('one', 1434), ('story', 1296), ('character', 1076), ('characters', 935), ('also', 807), ('two', 756), ('end', 715), ('life', 659), ('performance', 630)]
Topic 23 top 10: [('horror', 336), ('scream', 234), ('film', 212), ('killer', 194), ('scary', 137), ('2', 126), ('original', 113), ('sequel', 95), ('summer', 90), ('house', 88)]
Topic 24 top 10: [('batman', 191), ('robin', 124), ('george', 87), ('spawn', 82), ('cage', 76), ('mr', 60), ('schumacher', 57), ('comic', 55), ('clooney', 51), ('tom', 45)]

Glimpse of topics at iteration number 500:

Topic 0 top 10: [('war', 242), ('action', 172), ('jackie', 160), ('chan', 147), ('men', 141), ('spielberg', 114), ('fight', 108), ('army', 94), ('chinese', 92), ('films', 91)]
Topic 1 top 10: [('american', 170), ('political', 151), ('government', 107), ('america', 94), ('president', 85), ('black', 82), ('war', 81), ('even', 78), ('hollywood', 78), ('make', 69)]
Topic 2 top 10: [('sex', 258), ('scene', 171), ('man', 148), ('way', 118), ('camera', 116), ('violence', 116), ('kevin', 106), ('women', 100), ('look', 98), ('girl', 94)]
Topic 3 top 10: [('ship', 183), ('godzilla', 126), ('titanic', 115), ('cameron', 86), ('water', 76), ('effects', 71), ('disaster', 71), ('new', 67), ('monster', 65), ('york', 63)]
Topic 4 top 10: [('joe', 235), ('harry', 150), ('town', 118), ('grace', 93), ('derek', 66), ('man', 53), ('horse', 51), ('new', 50), ('country', 48), ('york', 47)]
Topic 5 top 10: [('king', 102), ('musical', 98), ('girls', 96), ('shakespeare', 95), ('spice', 80), ('music', 76), ('songs', 74), ('love', 69), ('band', 67), ('queen', 61)]
Topic 6 top 10: [('star', 343), ('alien', 226), ('wars', 179), ('effects', 172), ('series', 156), ('trek', 150), ('special', 131), ('aliens', 110), ('lucas', 105), ('menace', 98)]
Topic 7 top 10: [('truman', 163), ('show', 125), ('murphy', 125), ('carrey', 125), ('eddie', 99), ('west', 99), ('jim', 98), ('williams', 96), ('patch', 78), ('wild', 69)]
Topic 8 top 10: [('disney', 276), ('animated', 156), ('story', 144), ('family', 141), ('animation', 133), ('children', 124), ('voice', 119), ('tarzan', 118), ('kids', 116), ('toy', 101)]
Topic 9 top 10: [('movie', 5671), ('film', 5068), ('one', 3799), ('like', 3365), ('even', 2478), ('good', 2147), ('get', 1919), ('would', 1820), ('time', 1816), ('see', 1638)]
Topic 10 top 10: [('mr', 165), ('sequence', 87), ('lynch', 68), ('opening', 52), ('upon', 51), ('minutes', 49), ('original', 49), ('ms', 48), ('segment', 46), ('jolie', 46)]
Topic 11 top 10: [('nick', 101), ('bond', 93), ('gibson', 90), ('sam', 77), ('bill', 67), ('alice', 65), ('character', 64), ('mel', 60), ('wild', 59), ('james', 50)]
Topic 12 top 10: [('world', 199), ('city', 132), ('life', 115), ('science', 101), ('reality', 100), ('death', 98), ('dark', 96), ('angels', 82), ('human', 75), ('game', 74)]
Topic 13 top 10: [('comedy', 189), ('romantic', 180), ('love', 154), ('wedding', 123), ('ben', 92), ('julia', 90), ('life', 89), ('two', 88), ('ryan', 88), ('sandler', 77)]
Topic 14 top 10: [('comedy', 398), ('funny', 395), ('jokes', 184), ('humor', 166), ('laughs', 127), ('smith', 119), ('bob', 109), ('brothers', 104), ('hilarious', 103), ('comic', 102)]
Topic 15 top 10: [('evil', 139), ('action', 106), ('powers', 86), ('arnold', 84), ('dr', 80), ('lee', 73), ('jones', 73), ('austin', 69), ('agent', 68), ('bruce', 68)]
Topic 16 top 10: [('action', 273), ('effects', 132), ('mission', 119), ('van', 113), ('vampire', 103), ('special', 97), ('team', 94), ('blade', 92), ('computer', 90), ('john', 85)]
Topic 17 top 10: [('family', 312), ('life', 265), ('mother', 220), ('father', 205), ('love', 196), ('relationship', 147), ('son', 142), ('daughter', 127), ('husband', 124), ('child', 116)]
Topic 18 top 10: [('tom', 100), ('frank', 77), ('park', 59), ('scale', 56), ('0', 54), ('willis', 53), ('cole', 53), ('wife', 51), ('baseball', 49), ('chicken', 47)]
Topic 19 top 10: [('school', 244), ('high', 188), ('big', 106), ('teen', 106), ('club', 91), ('john', 84), ('dude', 83), ('mike', 81), ('girls', 74), ('girl', 66)]
Topic 20 top 10: [('earth', 231), ('planet', 184), ('effects', 118), ('space', 112), ('alien', 109), ('human', 86), ('humans', 82), ('deep', 79), ('special', 79), ('troopers', 79)]
Topic 21 top 10: [('crime', 150), ('cop', 146), ('tarantino', 117), ('jackie', 110), ('police', 104), ('de', 99), ('action', 93), ('drug', 93), ('pulp', 91), ('jack', 90)]

```
Topic 22 top 10:  [('film', 4235), ('one', 1711), ('story', 1364), ('character', 990),
('characters', 928), ('also', 891), ('two', 683), ('life', 653), ('however', 650),
('director', 642)]
Topic 23 top 10:  [('horror', 332), ('scream', 246), ('killer', 164), ('original', 155),
('film', 140), ('2', 133), ('scary', 116), ('house', 105), ('sequel', 98), ('first', 94)]
Topic 24 top 10:  [('batman', 195), ('robin', 113), ('cage', 80), ('spawn', 80), ('comic',
57), ('schumacher', 57), ('mr', 49), ('tim', 48), ('michael', 47), ('joel', 45)]
```

In [ ]:
```python
"""Reduced corpus for development"""
# """Gibbs Sampler setup - Small Reviews"""
# count_document_topic = [Counter() for _ in range(n_topics)] # initialize counters
# count_word_topic    = [Counter() for _ in range(n_topics)]
# count_topic         = Counter()
# topic_assignment    = list() # z[d][i] - topic assignment of ith word in document d
#
# for d, document in enumerate(small_reviews):
#     topic_assignment.append(np.random.randint(n_topics, size = len(document))) # randomly initialize z
#     for i, word in enumerate(document): # increment counters
#         topic = topic_assignment[d][i]
#         count_document_topic[topic][d] += 1
#         count_word_topic[topic][word]  += 1
#         count_topic[topic]             += 1
#
# vocabulary: set = {word for document in small_reviews for word in document}
# W: int = len(vocabulary)
# P:list = np.zeros(n_topics)
# T:list = list(range(n_topics))
```

In [ ]:
```python
"""Reduced corpus for development"""
# """Gibbs Sampler algorithm - Small Reviews"""

# for _ in tqdm(range(iterations)):
#     for d, document in enumerate(small_reviews):
#         for i, word in enumerate(document):# for i = 0 -> N-1 do
#             topic = topic_assignment[d][i] # topic assigned to ith word of dth document
#             count_document_topic[topic][d] -= 1
#             count_word_topic[topic][word]  -= 1
#             count_topic[topic]             -= 1
#             for t in range(n_topics):
#                 """P(z_i = t | z_-i, w) =ish    (n^wi_-i,t + Beta)  * (n^di_-i,t + alpha) /
#                                                 (n^._-i,t + W*Beta) * (n^di_-i,. + T*alpha)"""
#                 P[t] = ((count_word_topic[t][word] + beta) * (count_document_topic[t][d] + alpha)) \
#                     / ((count_topic[t]       +       (W * beta)) * (len(document)-1 + (n_topics * alpha)))
#             P = P/sum(P) # normalise distribution for sampling
#             topic = np.random.choice(T, p=P) # topic <- sample from p(z|.)
#             topic_assignment[d][i] = topic # z[d][i] <- topic - reassign topic based on sample
#             count_document_topic[topic][d] += 1
#             count_word_topic[topic][word]  += 1
#             count_topic[topic]             += 1

# for i, counter in enumerate(count_word_topic):
#     print(f"Topic {i} top 5: ", counter.most_common(5))
```