test_words = ['man', 'boston', 'save', '226', 'doctrine', 'he', 'what', 'about', 'other', ‘sounds']

newsgroup similarity with tf_idf weighing:

```
['soc.religion.christian', 'rec.autos', 'talk.politics.misc', 'rec.sport.hockey', 'alt.atheism', 'rec.sport.baseball', 'talk.politics.mideast', 'rec.motorcycles',
 'talk.politics.guns', 'talk.religion.misc']
```

newsgroup similarity without tf_idf weighing:

```
['soc.religion.christian', 'rec.autos', 'talk.politics.misc', 'rec.sport.hockey', 'alt.atheism', 'rec.sport
 .baseball', 'talk.politics.mideast', 'rec.motorcycles', 'talk.politics.guns', 'talk.religion.misc']
```

3-similarity functions with term-newsgroup matrix (raw count):

```
['speaking', '152', 'call', 'interference', 'christ', 'him', 'other', 'how', 'all', 'anyone']
['however', 'division', 'knows', 'hated', 'eternal', 'but', 'all', 'would', 'only', 'add']
['however', 'division', 'knows', 'hated', 'eternal', 'but', 'all', 'would', 'only', 'add']
```

3-similarity functions with term-newsgroup matrix (tf_idf):

```
['complete', 'anyways', 'call', 'facing', 'biblical', 'his', 'other', 'how', 'what', 'significant']
['however', 'record', 'knows', 'dig', 'biblical', 'who', 'all', 'would', 'only', 'usually']
['however', 'record', 'knows', 'dig', 'biblical', 'who', 'all', 'would', 'only', 'usually']
```

3-similarity functions with term-context matrix (raw count):

```
['great', 'detroit', 'move', '111', 'part', 'it', 'how', 'to', 'and', 'looks']
['person', 'toronto', 'move', '158', 'catholic', 'but', 'if', 'so', 'which', 'looks']
['person', 'toronto', 'move', '158', 'catholic', 'but', 'if', 'so', 'which', 'looks']
```

3-similarity functions with term-context matrix (ppmi):

```
['and', 'bruins', 'to', '61', 'catholic', 'it', 'you', 'that', 'are', 'it']
['and', 'chicago', 'and', '227', 'catholic', 'they', 'that', 'that', 'and', 'it']
['and', 'chicago', 'and', '227', 'catholic', 'they', 'that', 'that', 'and', 'it']
```

3-similarity functions with Word2Vec matrix:

```
['woman', 'detroit', 'steal', '118', 'concept', 'she', 'why', 'exactly', 'different', 'looks']
['soviet', 'soviet', 'soviet', 'soviet', 'greek', 'soviet', 'soviet', 'soviet', 'soviet', 'soviet']
['firearms', 'brian', 'states', 'normal', 'teach', 'generally', 'cwru', 'cleveland', 'town', 'congressional']
```

Conclusion:

- For all the similarity functions, the similarity of words generated by term – context matrix is much better than term-document matrix and word2vector matrix.
- Word2vector matrix has a good performance on cosine similarity, sometimes it can provide a more relevant words than all others. However, it performed bad on other similarity functions are bad.
- Term-document matrix doesn't give a clear result, although I changed the similarity function, the result didn't improve a lot, compared with term-context matrix.
- Jaccard and dice similarity give the same result for term-document matrix and term-context matrix.
- For document similarity, Tf_Idf doesn't give me any impressive improvement, the results between with and without tf_idf are very close. It might due to the words I chose are too few.
- PPMI didn't improve much on term-context matrix's in similarity between words result.