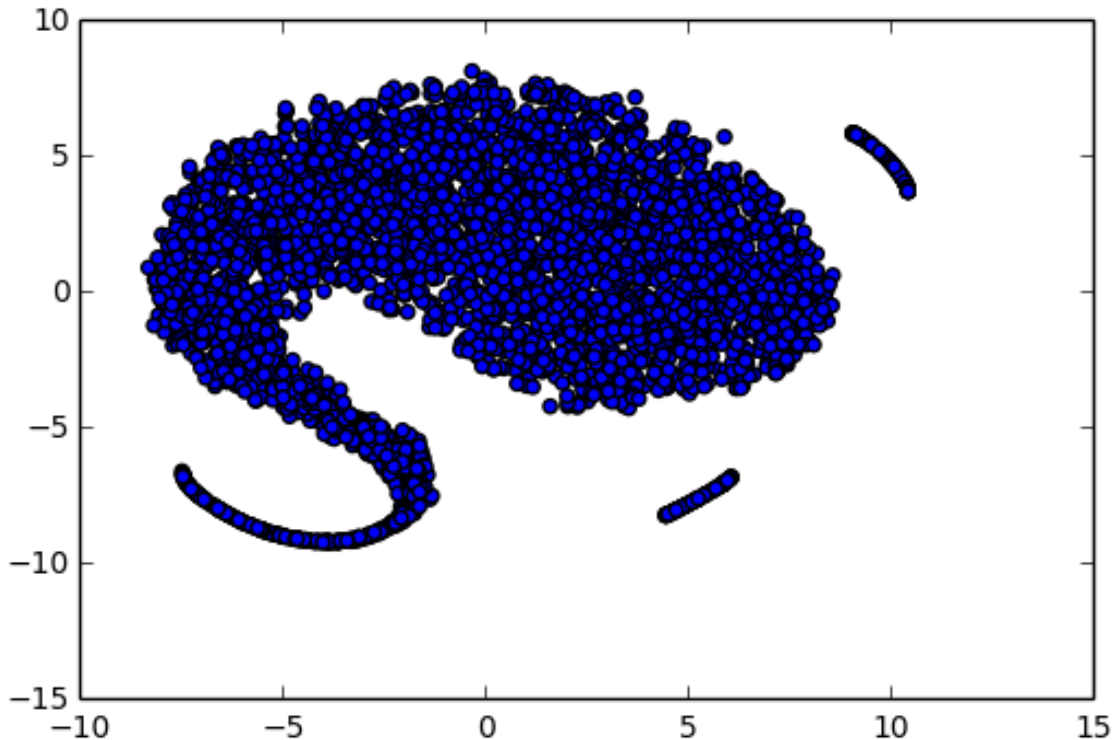


Shakespeare's Sonnets and Word2Vec Report



The above graph is a visualization of the word vectors from Shakespeare's Sonnets. Using word2vec and TSNE, word vectors were extracted from Shakespeare's vocabulary and displayed in this graph where the vectors' dimensionalities are reduced to 2. Although this graph isn't quite clear on why these particular clusters are being formed (as can be seen there is tadpole structure which is made up of most the words while there are two smaller segments around the main structure), I was able to extract interesting analogies using word2vec.

Shakespeare's sonnets can be divided into two major parts. Sonnets 1-126 address a fair youth. This is believed to be a young man whom Shakespeare expresses love towards. Sonnets 127-154 on the other hand are directed to a dark lady, who Shakespeare is drawn to sexually. However, he resents this woman as the fair youth has also fallen to her lustful charm. In searching for analogies from this text, I wanted to draw out what types of words may be similar to the key words of "man", "woman", "dark", "youth", etc. These are some of the results:

- Woman + dark:
 - (u'hate', 0.9914706945419312)
 - (u'1', 0.9912949800491333)
 - (u'This', 0.9909372925758362)

Most of the words that were most similar for 'woman' + 'dark' were common words like 'this', however the most common word is 'hate' which may demonstrate Shakespeare's contempt towards the dark lady.

- Man + youth:
 - (u'Then', 0.9954723715782166)
 - (u'men', 0.9954501986503601)
 - (u'any', 0.9952216148376465)

Like the analogy before many of these words are common. However although not in the top 3 similar words, the words 'art' and 'love' came in with 0.994820177555084 and 0.9947765469551086. This demonstrates the affection Shakespeare expressed towards the young man.

- Woman + youth
 - (u'was', 0.995618462562561)
 - (u'donations', 0.9948531985282898)
 - (u'hate', 0.9947900176048279)

I chose to add these particular vectors together because I thought there may be a chance that pairing an arguably positive word such as 'youth' with 'man' while pairing a negative word such as 'dark' with 'woman' would be the cause of my results before. However even when pairing 'woman' and 'youth' we see that 'hate' still appears in the top three most similar words.

- Woman + black + dark - man
 - (u'unkind', 0.9840575456619263)
 - (u'water', 0.9776075482368469)
 - (u'head', 0.9726619124412537)

Next I wanted to see what types of words I would get when adding the words 'woman', 'black', 'dark' (all of which are meant to describe the dark lady) and then subtracting 'man'. Not surprisingly, the first word that shows up is 'unkind.'

- Man - dark - woman - black
 - (u'belongs', 0.6486631631851196)
 - (u'engraft', 0.5399505496025085)
 - (u'dignified', 0.4954431653022766)

This is the same as the previous analogy except reversed ('man' is now a positive value while the other terms were subtracted). This again shows Shakespeare's positivity towards the young man as 'belongs' and 'dignified' are two of the top words for this analogy.