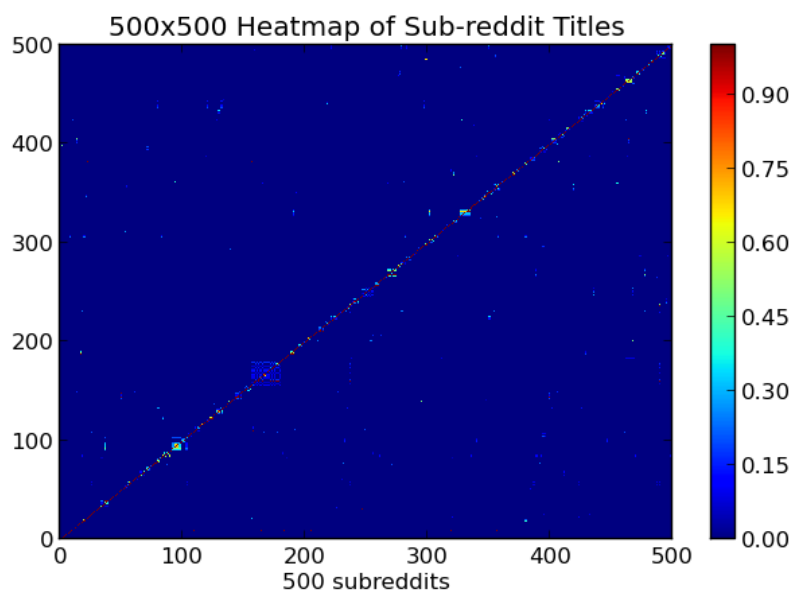
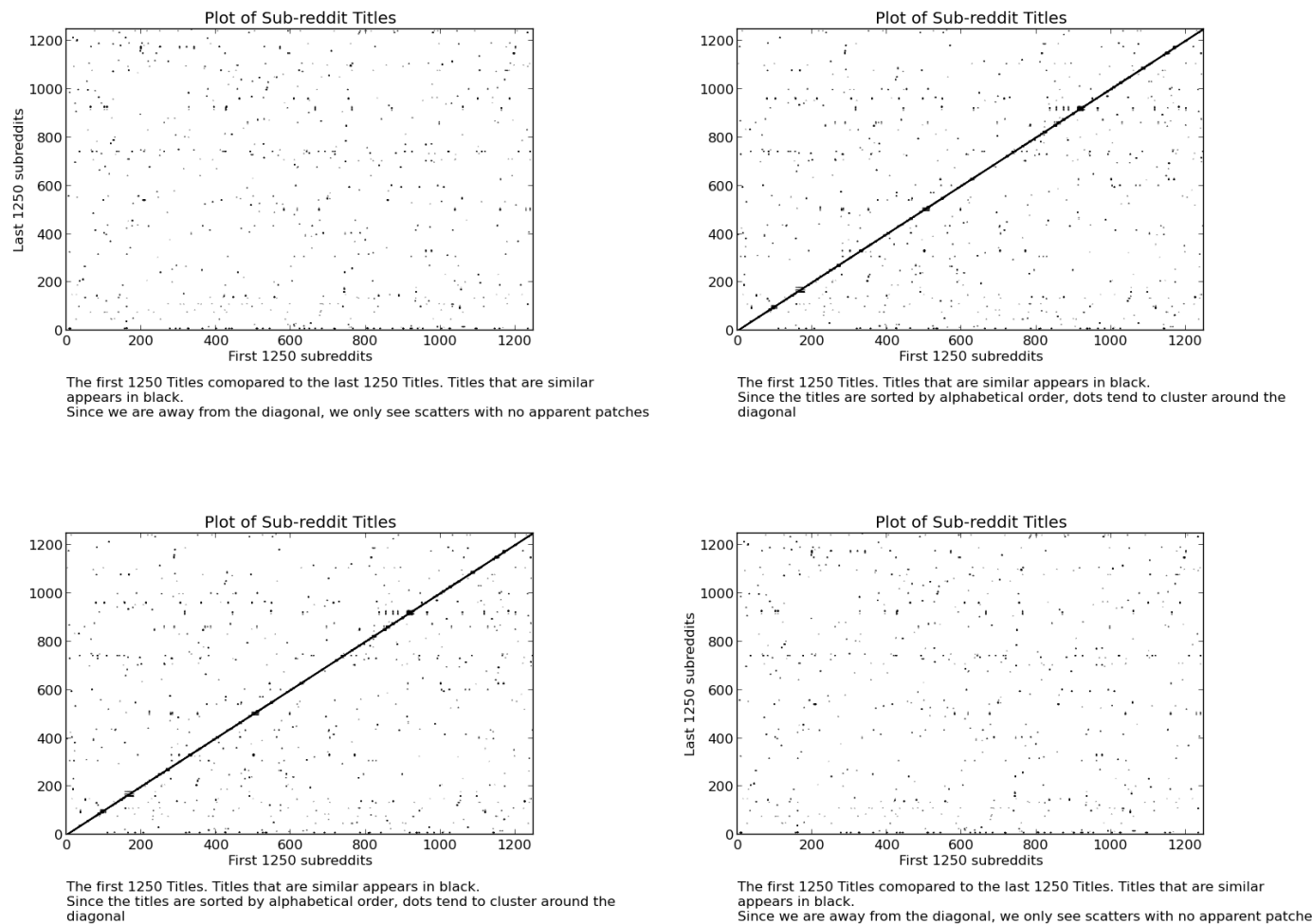


**Figure 1**



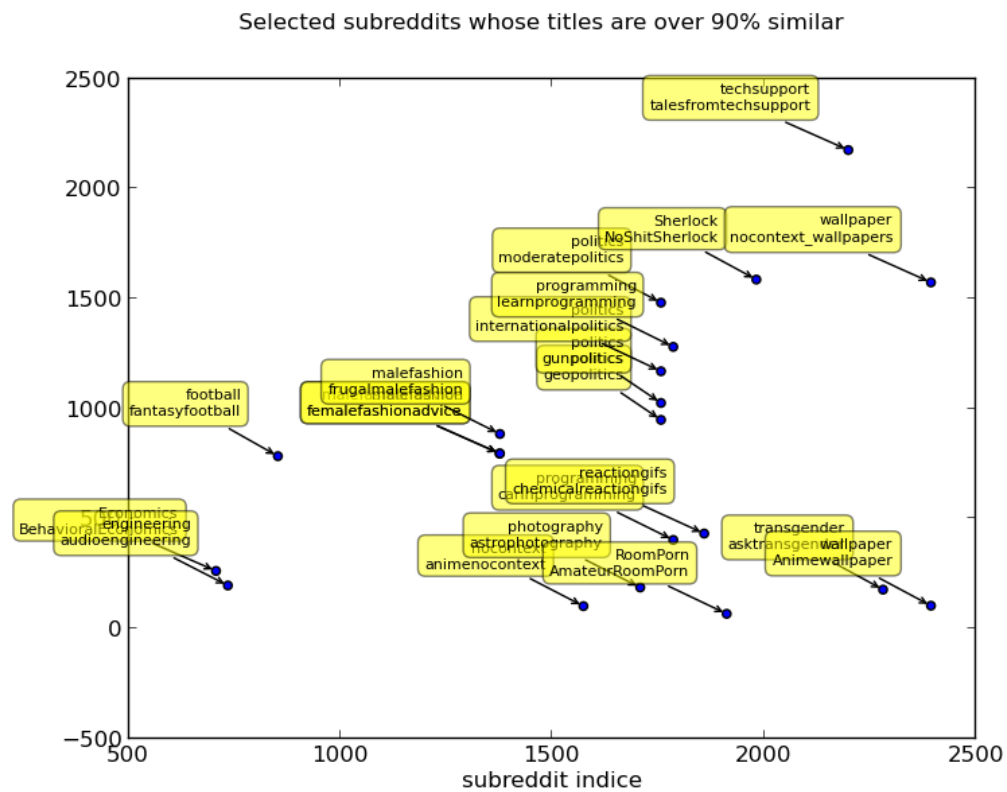
The heatmap of a portion of the 2500x2500 Title similarity matrix (from 500-800).  
There are patches on the diagonal and sparse scatters everywhere else.

Figure 2



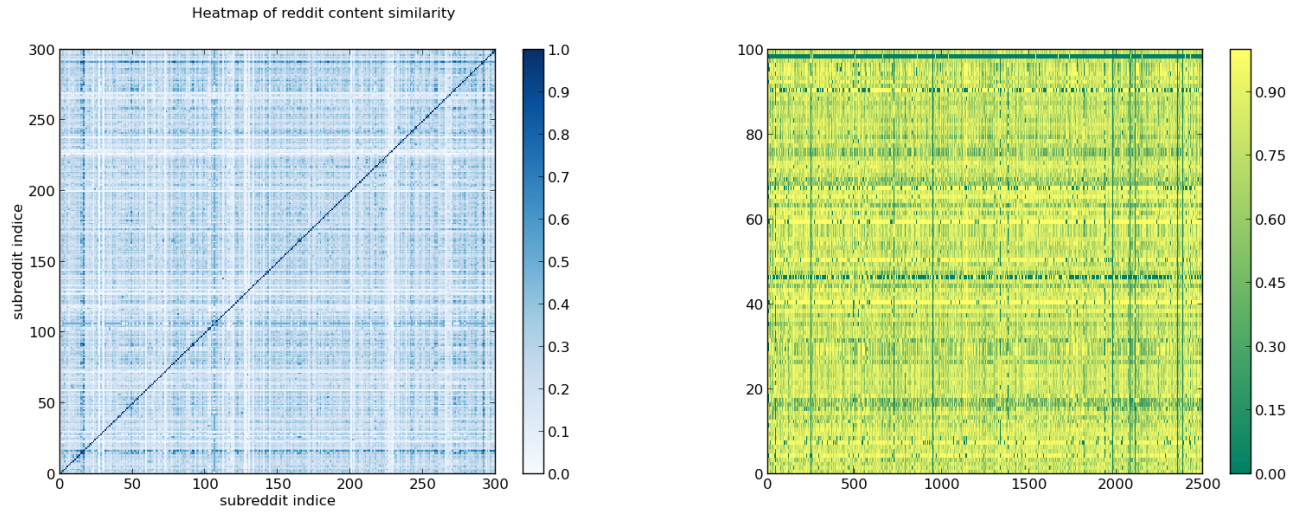
A full view of the title similarity matrix, divided into four regions.

Figure 3



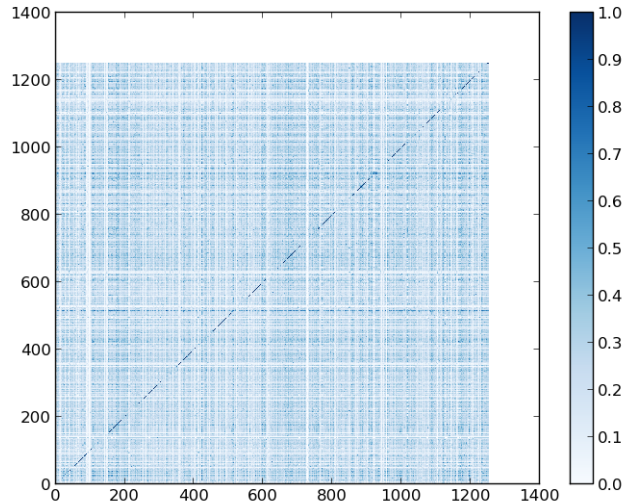
Subreddits that are more than 90 percent similar to each others, with corresponding names attached.

**Figure 4**



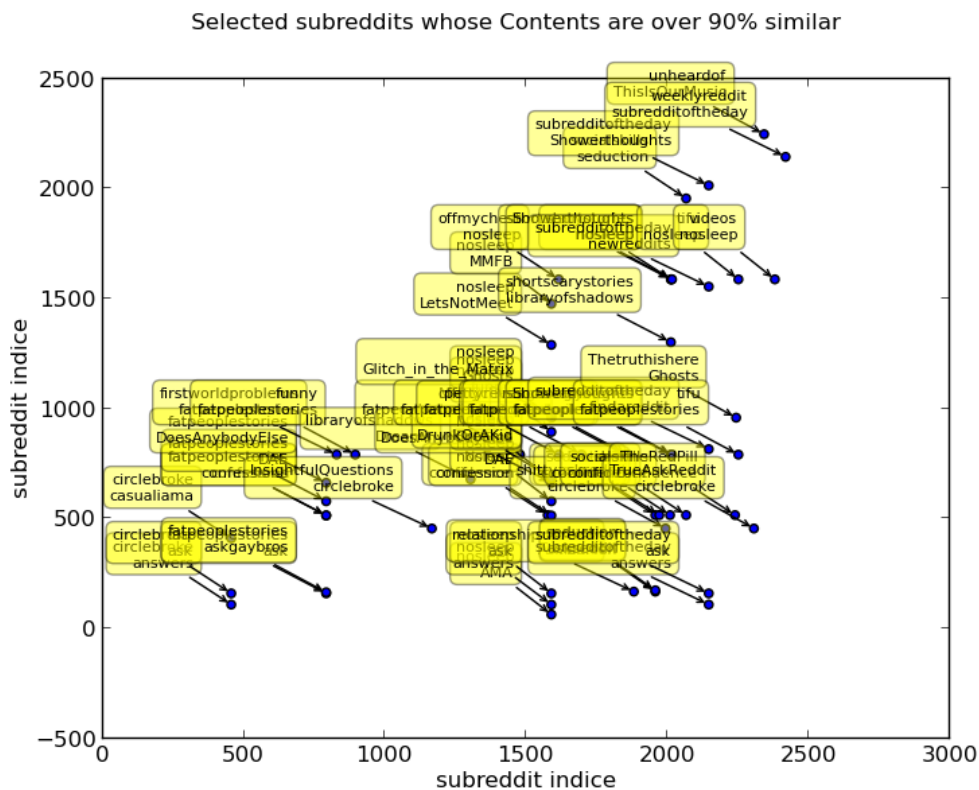
Left: a 300x300 region of the word-count similarity matrix. It has a large area of lightblue, only dark patches along the diagonal, implying most subreddits are distinct.

Right: a 100x1250 strip of the TFIDF-similarity matrix that aims to compare the similarity between any two given subreddits. It shows a high portion of yellow, indicating subreddits are quite alike.



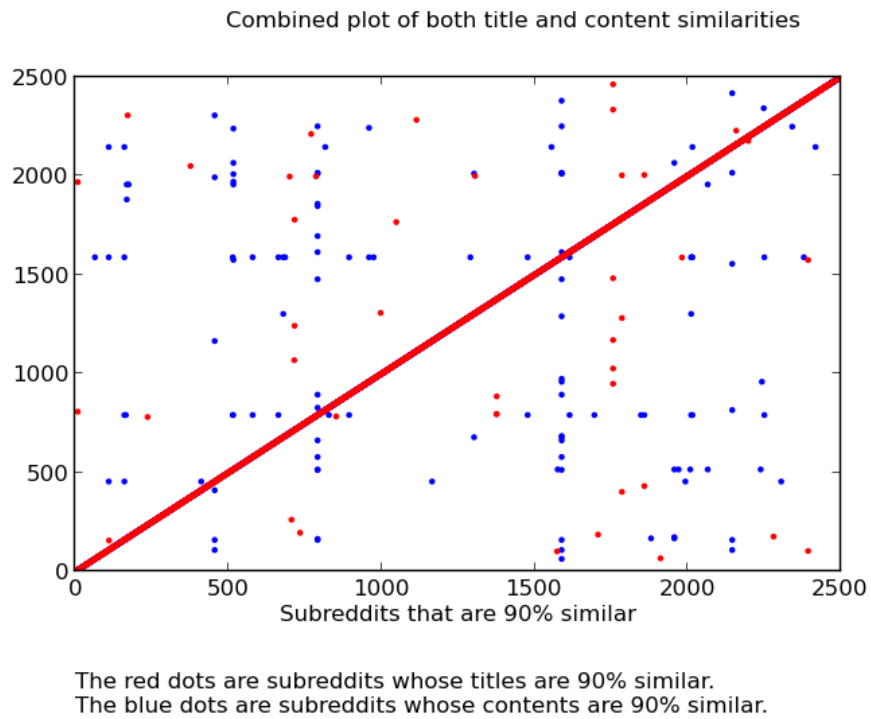
The same heatmap of content similarity computed base on word counts, at a large scale (1250x1250).

Figure 5



A plot of subreddits whose contents are more than 90 percent similar to each other. Note that several subreddits are similar to more than one other subreddits, therefore, appearing more than once.

Figure 6



Subreddits who are 90 percent alike in title and 90 percent alike in content plotted on top of each other.

Subreddits who are similar in title tend to be similar in content, but subreddits that are similar in content may not have the same title as there are more blue dots than red dots.