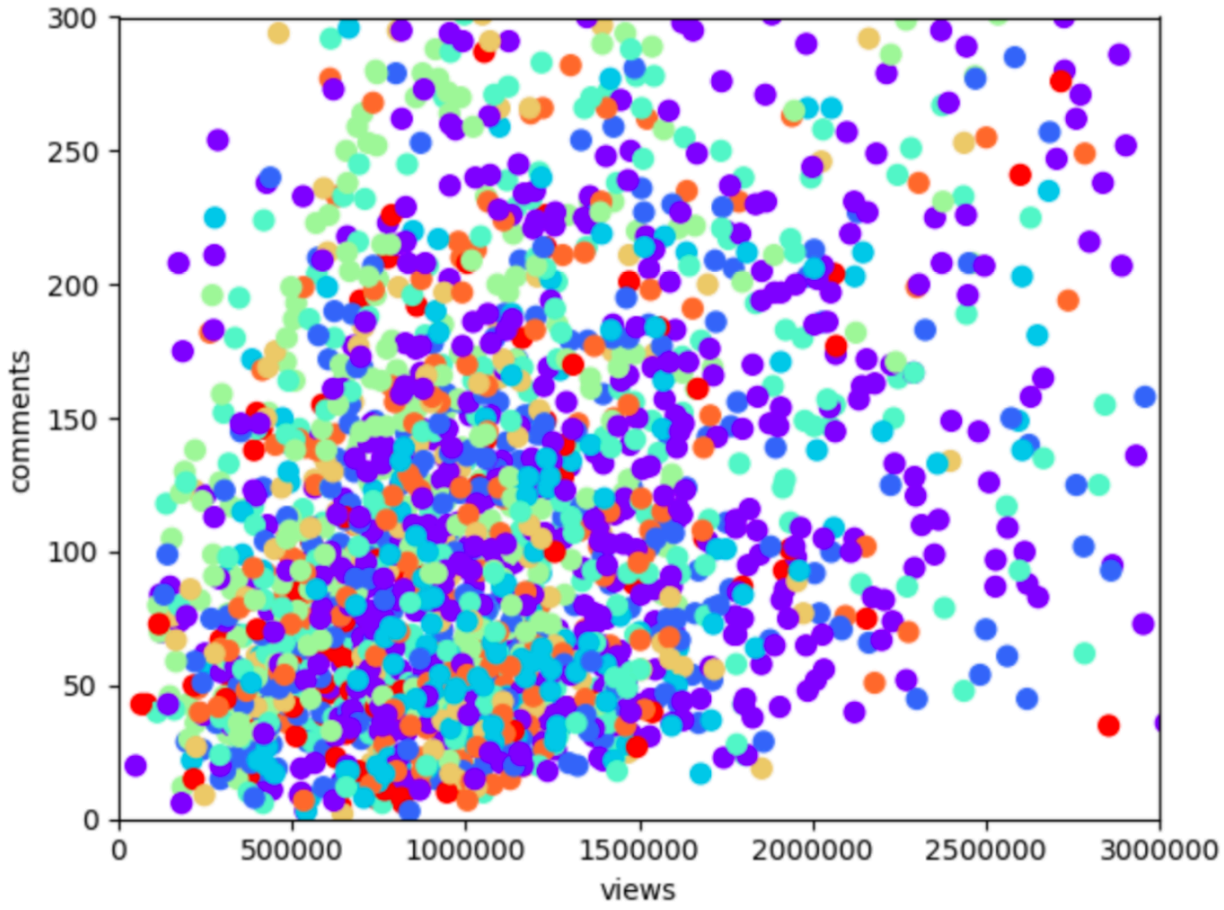
Clustering with Silhouette

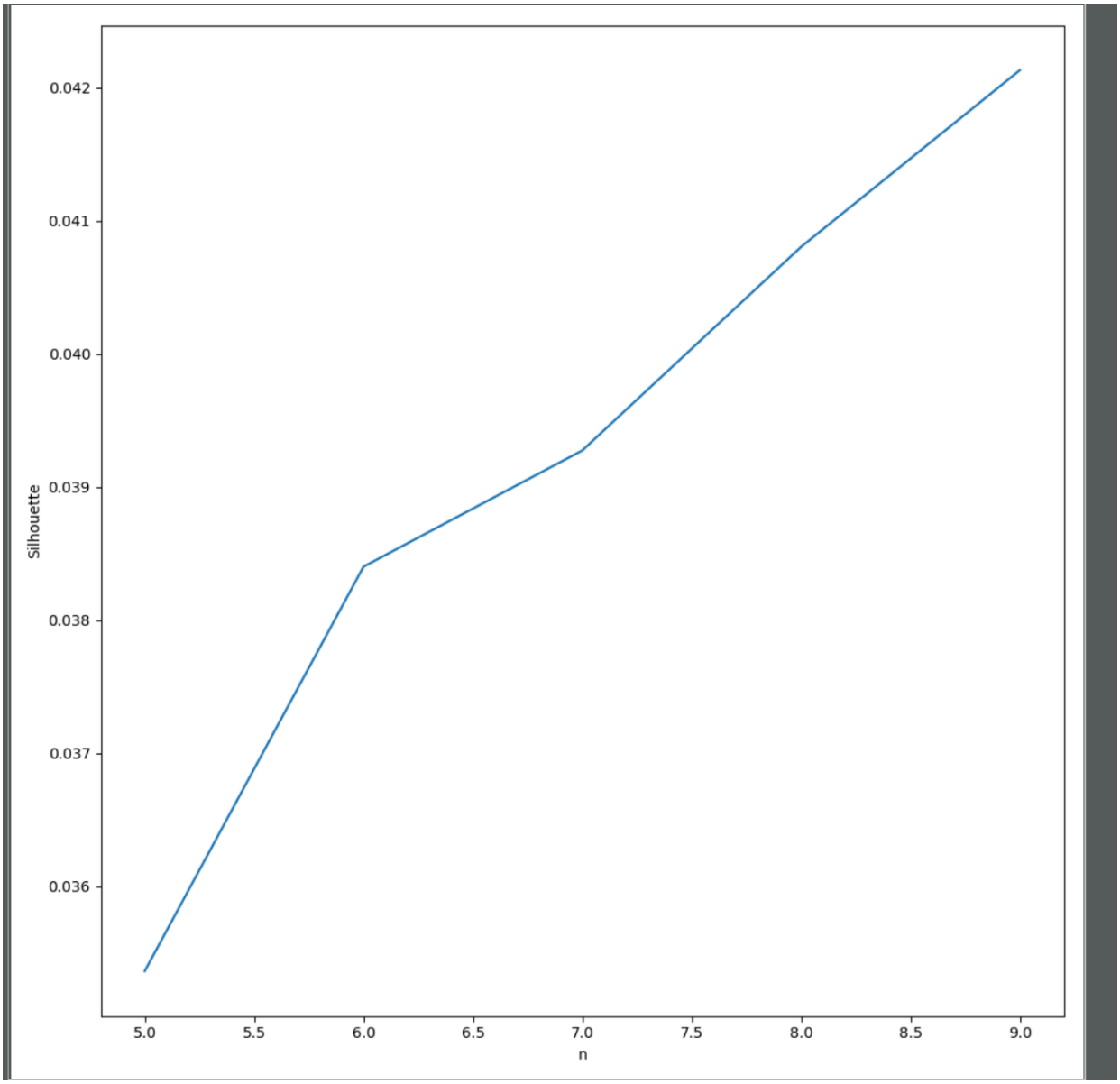
1. Clustering

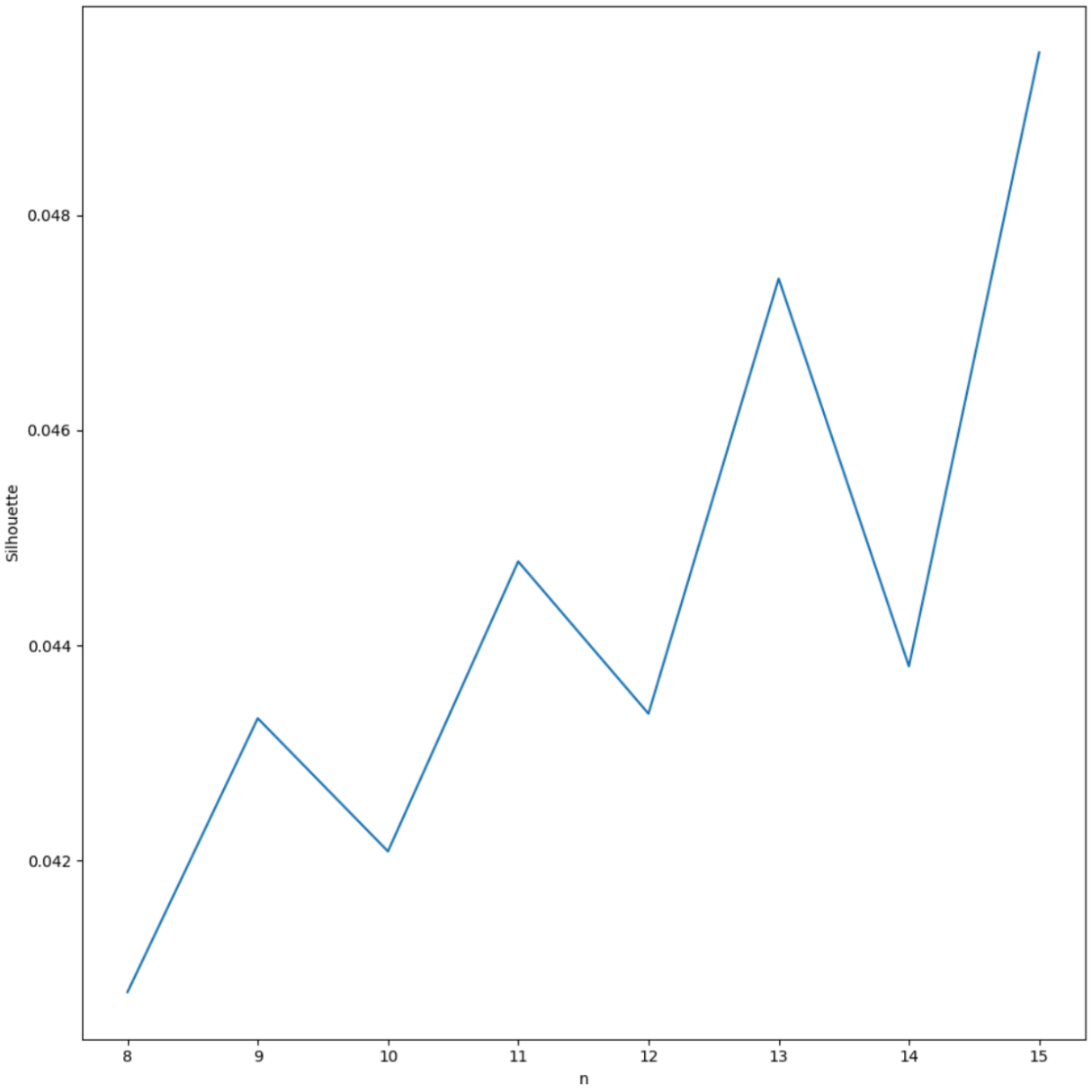
We've calculated the TF-IDF matrix of the ‘tags’ attribute and used it as a clustering feature to using the clustering algorithm, in order to get a rough classification based on the theme. According to the amount of comments as the ordinate, the amount of play is the abscissa, each color represents a classification, we get the following image:

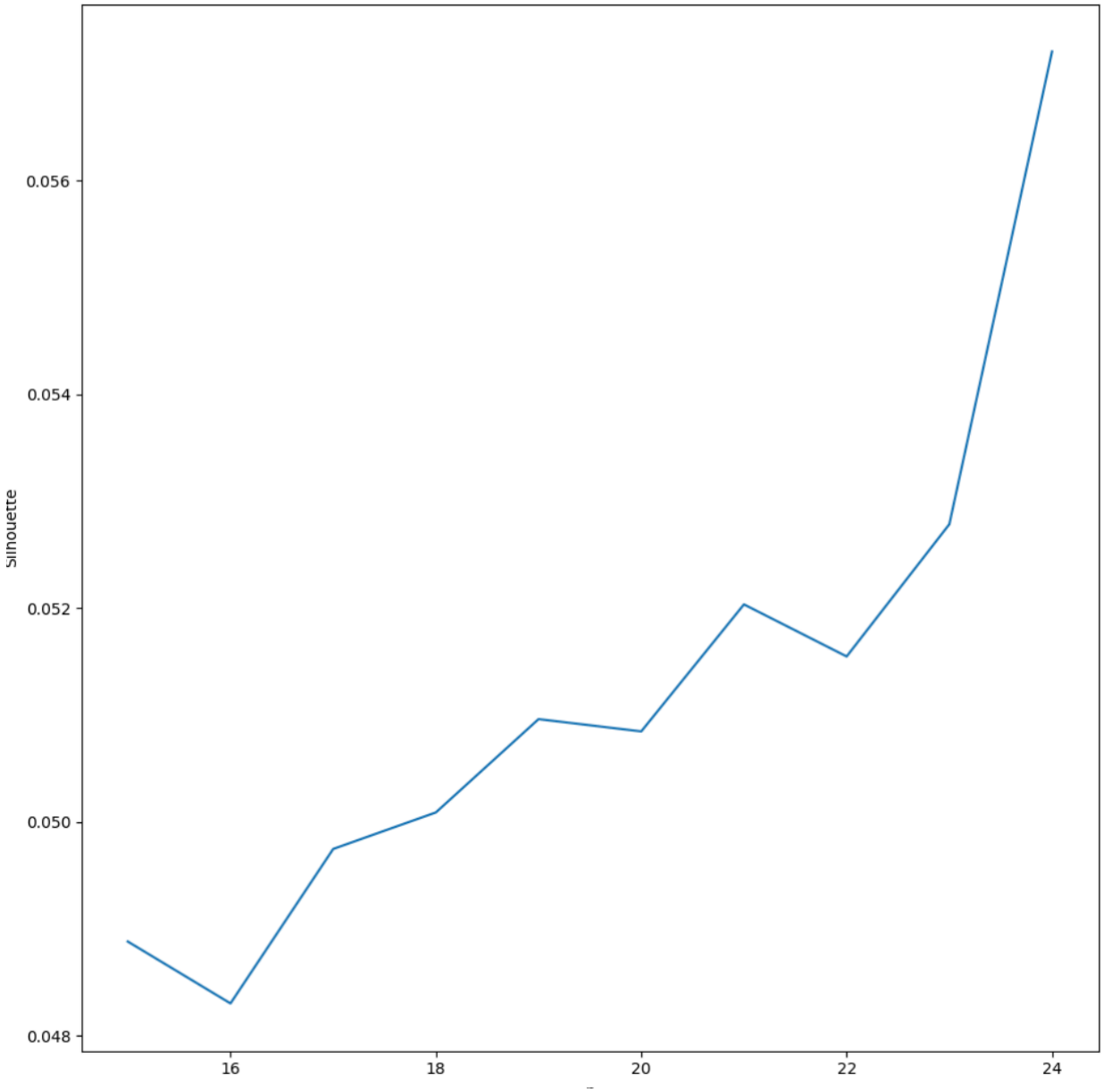


2. Silhoutte

However, when Silhoutte calculates the clustering quality, we find that the Silhoutte value is very low. Almost every value makes the Silhoutte coefficient approach 0. This means that our clustering effect is very bad, maybe we need to change a feature to do clustering.


0.042
0.041
0.040
0.039
0.038
0.037
0.036
Silhouette
5.0
5.5
6.0
6.5
7.0
7.5
8.0
8.5
9.0
n


0.048
0.046
0.044
0.042
Silhouette
8
9
10
11
12
13
14
15
n


Silhouette
0.056
0.054
0.052
0.050
0.048
16
18
20
22
24