

All measurements are an average taken over 10 trials.

Table1. Comparing Bregman's Hard Clustering With K-Means Clustering

Homogeneity	Completeness	V-Measure	Adj Rand Score	Norm Mutual Score	Fowlkes Mallows
Bigram Bregman's Hard Clustering					
0.50	0.14	0.23	0.07	0.23	0.37
Trigram Bregman's Hard Clustering					
0.56	0.16	0.24	0.08	0.24	0.37
Bigram K-Means Clustering					
0.78	0.21	0.33	0.12	0.33	0.39
Trigram K-Means Clustering					
0.59	0.15	0.24	0.07	0.24	0.34

When comparing the Bigram Clustering of BHC compared to K-means we can notice our largest differences, this is particularly noticeable with homogeneity where the homogeneity of K-Means is much better than that of Bigram Bregman's Hard Clustering. The same holds true for completeness as well as the majority of the other measurements excluding Fowlkes Mallows, which was only a slight improvement when using K-Means. On the other hand, the Trigram Bregman's Hard Clustering showed improvements over the Bigram format, and also was much more on par with the results seen from Trigram K-Means Clustering. For the most part, between the Trigram K-Means Clustering and the Trigram Bregman's Hard Clustering, the results difference was minimal and as such not substantial. Therefore in this case, we could consider the Trigram Bregman's Hard Clustering to be on par with the Trigram K-Means Clustering.