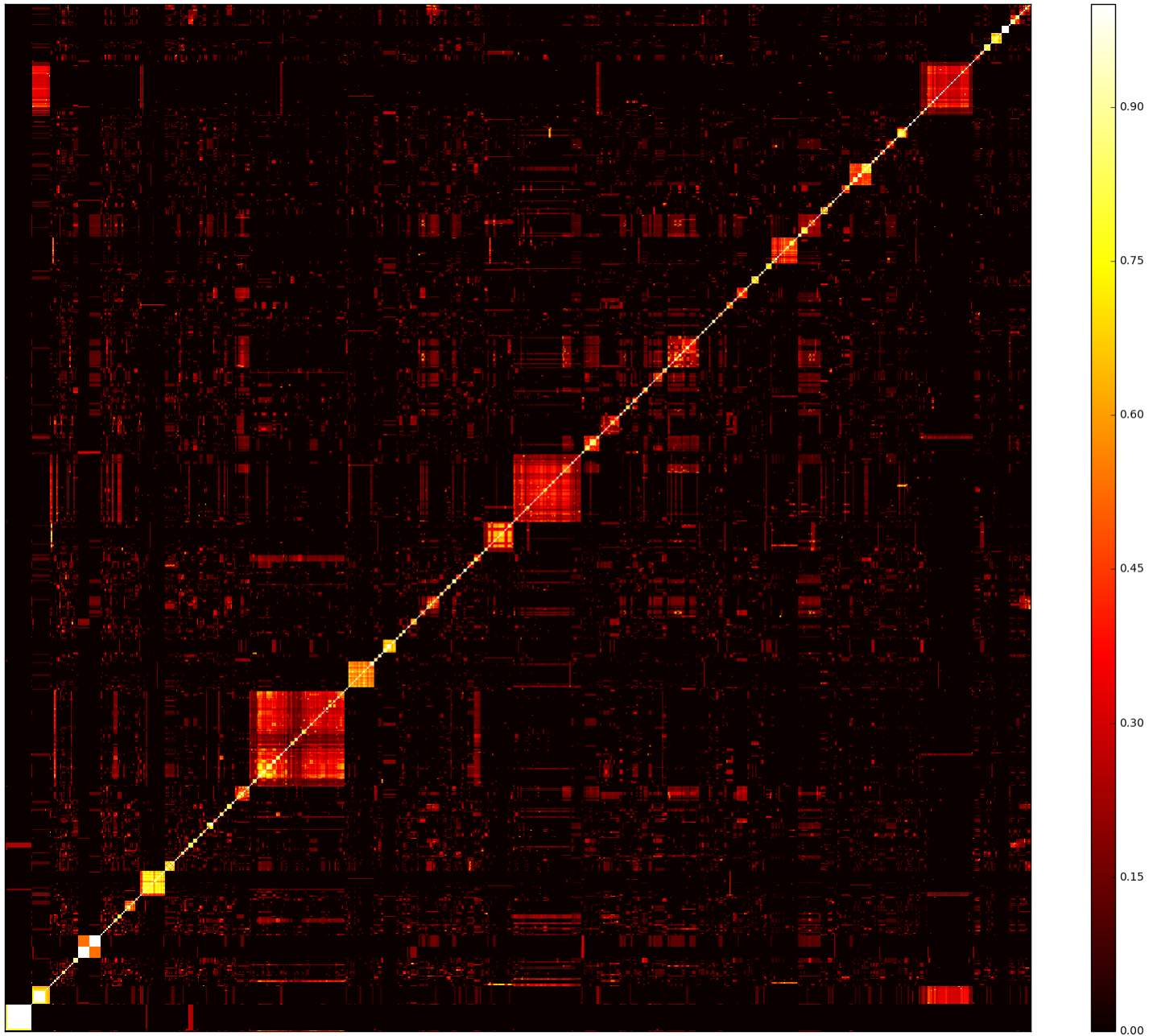


(a) Similarity matrix of the items in the “Online Retail” UCI dataset, using the n-gram method on the descriptions of the items. Based on this similarity, the items are clustered using a hierarchical clustering method. Then, similar items are put near each other to form small clusters.



(b) A larger view of the first 1000 items. As can be seen, the items have formed small clusters based on the similarities between their descriptions.

**Fig. 2.** The similarity (using n-gram) and clustering (using a hierarchical clustering method) results on the “Online Retail” UCI dataset.