**Paper Choice: Bayesian Hierarchical Clustering**

**Outline of Background:**

This paper is to introduce the new Agglomerative Hierarchical Clustering Algorithm from the perspective of Bayesian.

Agglomerative Hierarchical Algorithm is an important method in unsupervised learning: a useful technique for visualizing and discovering groups in a data set by analyzing the data to be a nested tree dendrogram with nodes. A Dirichlet process mixture model builds up this Bayesian-based method. Similar to the traditional hierarchical clustering methods, this new method starts each data point in its own cluster and iteratively merges pair of clusters. The Bayesian Hierarchical Clustering avoids many issues that cannot be handled properly by traditional Agglomerative Hierarchical Algorithm methods. The BHC applies hypothesis testing to decide which data should be merged together or not.

The BHC has more advantages over the traditional algorithm. The traditional Agglomerative Hierarchical Algorithm has limitations. There is no certain rubric to decide the number of clusters to choose: there are generally only four types of linkages to choose from according to similarity measures: complete, single, average and centroid. The “good” or “bad” model can just be decided by intuition. Over-fitting issues are always involved, and there is no proper way to evaluate the degree of over-fitting. Instead, the BHC builds up a probabilistic model by computing marginal probability: the probability that the data merging together. This probabilistic model can be easily applied when evaluate the over-fitting and compute the predictive model. The Bayesian hypothesis testing can be also used to decide to the depth of the dendrogram. The paper provides an example of comparing the BHC and the classical distance-based hierarchical clustering method. BHC does not strongly rely on the assumption. It can also provide good hierarchical structure at high level.