Reflection - HDP - Teh
Nicholas Hoernle
<div align="right">**April 11, 2018**</div>

1. Teh targets the problem of sharing of cluster components among distinct groups. Teh builds on the DP to include settings where data has groups and each group has clusters, but the groups are related by having components that can be shared. A clear example is from topic modeling where a corpus might contain documents (groups) with certain topics (clusters). Each document will have a document specific distribution of topics (drawn from potentially infinitely many topics). However, it seems reasonable to assume that the documents all share a number of topics (possibly with different proportions) by virtue of being in the same corpus. The HDP allows this structure where the corpus level structuring of topics might be described by the base draw from a DP ($G_0 \sim DP(\gamma, H)$). Each document is then described by a draw from the document specific DP ($G_j \sim DP(\alpha, G_0)$) with $E[G_j|G_0] = G_0$ describing the interrelated structure of the documents. The DP alone cannot facilitate this hierarchical structure.

2. The Hierarchical Dirichlet Process (HDP) generalizes Latent Dirichlet Allocation (LDA) by not requiring the pre-specification of the number of topics in the corpus. The graphical model relationship is the same in both models but the HDP allows for infinitely many topics.

3. Figure 5 suggests that the HDP is able to borrow statistical strength from the other groups when the group of interest (VS in this case) has a low number of training examples. At 0 training examples, M2 presents better prior information than M1 and M3 but as soon as M3 has one or more group specific example for VS, the HDP structure is able to quickly borrow topic structure from the other groups to decrease its group specific perplexity.