**A Split-Merge MCMC Algorithm for the Hierarchical Dirichlet Process**

**Goal**: novel inference method for the Hierarchical Dirichlet Process (HDP), widely used in Bayesian nonparametric modeling and topic modeling

* Exact inference in HDPs is intractable
* Gibbs sampling (standard method) only makes local moves (one variable at a time) -> slow to converge and can cause poor solutions

Synthetic data – Split-Merge MCMC outperforms Gibbs since the second fails to distinguish very similar topics whereas the first doesn’t

Real data – Split-Merge accelerates convergence and finds more accurate topic structures when there are very similar topics

Results of Evaluation

* HDPs representation of Chinese Restaurant Franchise
* Split-merge only at a global topic level, not inside individual documents
* Performance evaluated on synthetic and real data

How

**Split-Merge Markov Chain Monte Carlo (MCMC)** algorithm for HDPs, inspired by similar methods in Dirichlet Process mixtures:

* Larger moves -> splitting one topic into two or merging two topics into one, accepted or rejected by a Metropolis-Hastings Step

Solution

**Conclusion**: Split-merge MCMC algorithm is effective during the burn-in phase of HDP Gibbs sampling on both synthetic and real data.