Summary for BDSS IGERT Research Rotation (November2015)

Host: Dave Hunter (STATs)

Trainee: Cindy Cook (STATS)

Meetings: Wednesdays 9:30am in Thomas

Overview:

We have had successful meetings every Wednesday. So far, I have spent my time understanding latent dirichlet allocation (lda). I began by writing a function in R that will simulate corpuses in the form of a term document matrix based on lda’s generative model. I then used both the *topicmodel* and *lda* packages in R to fit lda on resume data I obtained through my summer externship. Both functions resulted in different solutions. I have since been focusing on how the algorithms are approximating the posterior distribution (i.e. collapsed Gibbs sampling or variational Bayes methods). I want to explore these differences more in depth creating a simulation study. I also want to sit down with Burt to discuss his research in the field of topic modeling and good directions for Dr. Hunter and I to take.