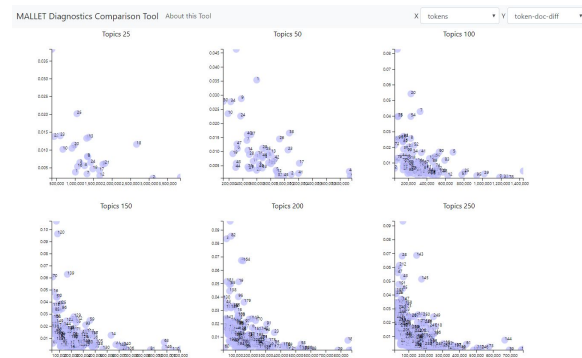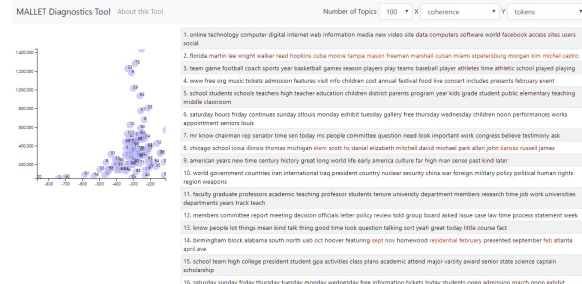# MALLET Diagnostics Visualization Suite

The MALLET diagnostics visualization suite consists of two tools for visualizing the topic model diagnostics files generated by MALLET. Diagnostics files are XML files containing statistical metrics for evaluating topic models. The WE1S tools are modifications of the visualization tool available on the MALLET website. These modifications consist entirely of user interface enhancements. The individual model tool allows the user to view a scatterplot of an individual topic model, along with a table of keywords for each topic. The comparison model tool allows the user to compare scatterplots of different models side by side and provide the keywords in popup dialogs. Both tools convey exactly the same information and merely provide alternative ways of looking at it.

Metrics include number of tokens, document entropy, word length, coherence, uniform distribution, corpus distribution, effective number of words, token/document discrepancy, rank 1 documents, allocation ratio, allocation count, and exclusivity. The MALLET website provides formal discussion of these methods. Shortened explanations are available from within the tools. Further discussion of our experience with these metrics can be found on the Diagnostics Methods Card (M-13).

Screenshot of the individual model tool (above) and the comparison model tool (below). *Click for larger image*

Further Information:

* MALLET Diagnostics Page

WE1S resources:

* Diagnostics Methods Card

Sample Models:

* US Humanities Collection (individual models)
* US Humanities Collection (comparison of models)

10 Feb. 2020; rev. 15 Feb. 2020 (Scott Kleinman)