Archival Description and Network Analysis
## Comparison and Preliminary Review

Cory Nimer, Brigham Young University

Introduction

Recent advances in automation, archival authority work, and linked data technologies provide archivists and researchers with new opportunities to visualize finding aid content. Existing methodologies demonstrate the possibilities for exploring archival networks through enhanced archival description. However, the utility and accuracy of such visualizations are limited by the descriptive surrogates themselves, and the processing choices made by archivists. Comparing different datasets on a group of historians of Mormonism highlighted the impact of descriptive practice on resulting network analysis graphs.

Methods
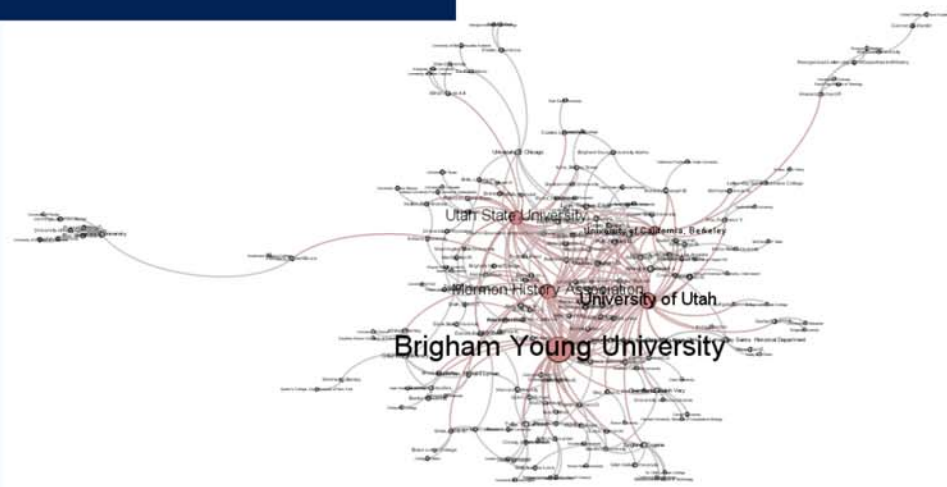
This project focused on a group of 93 academic historians that had published on Mormon history topics, and who began their careers prior to 1980. Authors were selected based on a review of academic catalogs from Utah universities, and their publication record verified using WorldCat. Relationship data about these authors, as well as related creators, was then compiled using the approaches developed by the Social Networks and Archival Context (SNAC) project and Emory University's Belfast Group Poetry Network project. Two additional datasets were then developed for comparison with these archival approaches. The first was composed of bibliographic citation relationships harvested from Google Scholar using the Publish or Perish software application. An expanded index of archival relationships was also developed by expanding the available level of description for locally held collections. Each of these datasets were then entered into Gephi, a network analysis software application, to look at statistical measures and graphical representations of available relationships. The analysis focused on clustering and graph density, as well as statistics for individual nodes.

Results

Reviewing the Gephi results for each of the datasets demonstrated a greater network diameter for the SNAC data than seen in other networks, with two to three times more connections between the most distant nodes in the graph. When comparing the initial finding aid relationships with the set derived from expanded indexing, there was an increase in the average degree of nodes and the diameter of the graph, while the overall graph density was reduced. However, graph density and degree values were generally low.

Bibliographic relationship graph

Thomas G. Alexander

Leonard J. Arrington

Finding aid relationship graph

	Average Degree	Network Diameter	Graph Density	Average Clustering Coefficient	Average Path Length	Number of Nodes	Number of Edges
SNAC	2.172	12	0.004	0.026	4.909	616	669
Bibliographic	3.014	6	0.001	0.099	2.761	9931	29933
Finding Aid	1.131	5	0.001	0.031	2.768	1582	1790
Expanded Indexing	2.238	6	0	0.687	3.172	6942	7767

Comparing the positioning of specific individuals across the datasets generally showed them to be similarly positioned in terms of centrality. In some cases there were significant differences in degree values, though, with some historians either under- or over-represented.
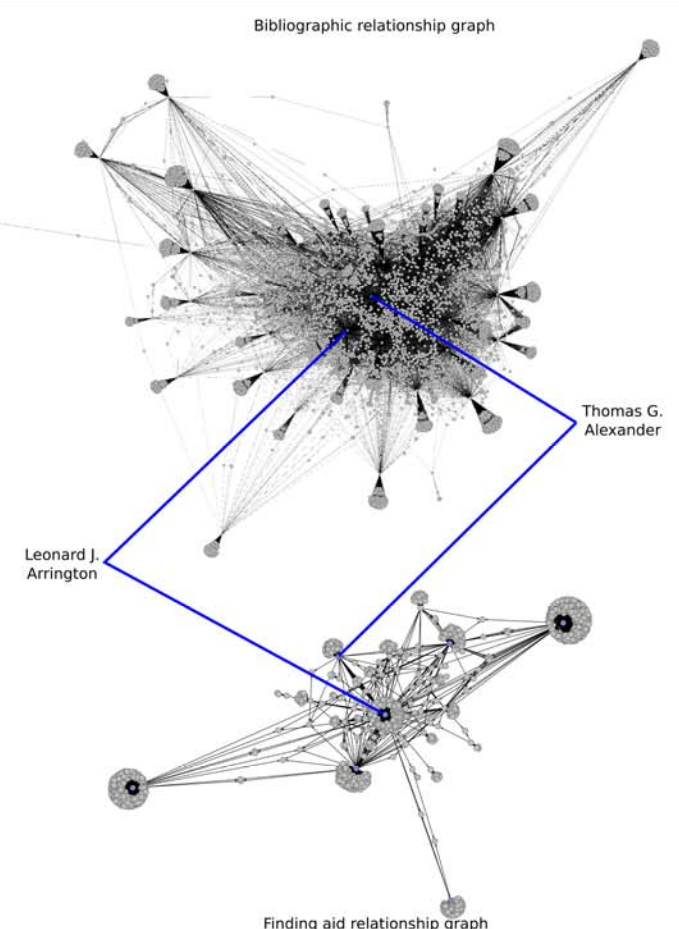
Conclusions

While the network graph based on bibliographic information provided useful measures of notability within the academic community, which might be applied to appraisal projects, graphs based on available archival data generally were not as robust. Reviewing the data, it would appear that discrepancies between the graphs were due to local decisions on appraisal, processing, or level of description. Additional indexing resulted in more closely connected graphs, with measures closer to the bibliographic dataset. The SNAC relationship data remained an outlier, however, perhaps due to the data sources used in its production.