

# Social Network Analysis on Twitter with Chapel

Ben Albrecht, Michael Ferguson, *Chapel Implementers, Cray Inc.*

With the rise of the information age came a rapidly developing interest in analyzing the flow of data, information, and knowledge. Social network analysis is a popular data analytics technique that utilizes graph theory to study relationships and interactions among users in communities. As Chapel matures into a more featureful and performant language for scalable computing, the natural question arises of how Chapel will fit into the data analytics picture. The work presented here begins the journey to answering this question by investigating the performance of a Twitter workflow social network analysis benchmark implemented in Chapel. The performance of Chapel's implementation is compared to that of Spark, a cluster computing framework well known for its performance in data analytics.

The Twitter workflow benchmark parses JSON-encoded tweets into graph data structures containing data on user IDs and user mentions. From here an iterative label propagation algorithm assigns users to communities based on the total number of mentions among users. The application can be thought of as two computationally distinct phases: the parsing of tweets and the graph-traversal across the Twitter data.

The goal of this project is two-fold: to understand the performance differences between Chapel and Spark, broadly illuminating how Chapel will perform as a data analytics tool; and to pave the road for future Chapel users trying to accomplish similar tasks. This includes building upon the Chapel standard library to ensure an elegant approach exists for scalable I/O and graph analysis. Performance is analyzed on multicore single-locale and multi-locale implementations across different architectures. Preliminary results show the Chapel implementation outperforms the Scala/Spark implementation for the multicore single-locale trials.

This talk will present the analysis of scalability in Chapel compared to Spark for the Twitter workflow benchmark, and describe the new features that were adopted in the standard library to overcome the challenges encountered in development.