**Integrated Social Media Analytics Platform**

For this project, you will be constructing a general purpose framework to enable scientists to rapidly run a standard battery of tests on social media datasets. The framework will integrate the functionality of several existing tools into a single data processing pipeline and provide visualizations of the data for the researchers. Python is the preferred development language for the project. Data to be processed will be stored in a mixture of csv, MySQL, and mongoDB files.

The first part of the analysis will consist of constructing social networks showing the connections between the different people in the dataset. These networks will be multiplex networks, consisting of multiple types of links between people. The framework should be able to compute standard network centrality metrics and visualize them in a file format supported by Gephi (<http://gephi.org>).

The second part of the analysis will consist of evaluating the performance of different supervised and unsupervised machine learning classifiers at predicting variables in the dataset. These classifiers will be implemented using the Weka toolkit (<http://www.cs.waikato.ac.nz/ml/weka/>), and the framework will merely provide a wrapper around Weka.

The third part of the analysis will be in creating an API for researchers to run their own customized algorithms. We will provide several example community detection algorithms for detecting tightly connected subgroups of nodes.

The demonstration dataset for this project was extracted from the Travian multiplayer strategy game by Drs. Rolf Wigand and Nitin Agrawal from the University of Arkansas.

This project is sponsored by the Intelligent Agents Lab (<http://ial.eecs.ucf.edu/>), directed by Dr. Gita Sukthankar. Sandia National Labs is the project partner and undergrads with US citizenship may have the option to continue to a summer internship.

