A Digital Library for Crowds on the Real-Time Social Web

Jeff McGee and Krishna Y. Kamath

In this project, we want to build a digital library for transient crowds in highly-dynamic social messaging systems like Twitter and Facebook. A crowd is a short-lived ad-hoc collection of users, representing a “hotspot" on the real-time web. For example, an event like superbowl might result in formation of crowds that are discussing various happenings in the game. A set of crowds might be discussing some interesting advertisement while a different set might be discussing a player who might have been involved in an important play. The discovery and tracking of are challenging in comparison to the more static and long-lived group-based membership offered on many social networks (e.g., fan of the LA Lakers on Facebook). Successful detection of these hot-spots can positively impact related research directions in online event detection, content personalization, social information discovery, etc. We will build a framework that allows an analyst or curious user to find interesting crowds and see how they evolve. This framework will have three main parts: a database to store crowds, users, and their messages; a set of crowd detection algorithms and filters; and a tool for searching for crowds by topic, geography, or username.