A. Project Summary

NeTS: Medium: Collaborative Research: Towards Building Time Capsule for Online Social Activities

Online social networking sites such as Facebook, LinkedIn, and Twitter allow users to seek out friends or colleagues and interact with them in novel ways. At the same time, user-generated content has mushroomed in a collaborative manner in multiple contexts, e.g., online reviews of services/products, Wikipedia, and open software development. It is extremely easy today to share personal likes/dislikes (of news, products, etc.) with friends. New technology platforms are often used in ways not anticipated by the engineers who built them, e.g., micro-finance, telemedicine, and trading cell phone minutes as a form of currency. This new ecosystem of content generation, sharing, consumption, and innovation has completely reshaped the web by making it more of a massive online social system. This has huge implications on information propagation, consumer behavior and market economics (e.g., online rating/recommendations), as well as political behavior (e.g., 2011-12 uprisings in the Middle East). As more and more user social interactions become 'online', coupled with the advancement in large-scale measurements, data mining, and graph analytics, a unique opportunity arises. For the first time, rich empirical user data are being collected and analyzed through rigorous measurement methodology to study and test hypothesis derived from multiple disciplines (e.g., economics, history, political sciences, and social science). However, many of such studies require longitudinal data from multiple sources, which may become a challenge to store and process.

Intellectual Merit: Our grand vision is to build a *time capsule* that can condense pertinent information about *online social activities* (derived from massive data from multiple sources) as user interactions evolve through time. Such *digests* should be compact enough to store and process, yet contain rich enough information to support a wide range of (some yet unknown) queries or hypothesis testings in future. Towards this end, we propose to capture and analyze the *online social activities* over multiple online social networks/platforms by studying the associated *user activity graphs (UAGs)*, in which a node represents a user, and an edge represents a specific action. Specifically, we propose to:

- **Task 1:** Model dynamic growth/evolution of UAGs both at the macro-scopic and micro-scopic level. These models, along with the parameters, provide an alternative to sharing/storing massive, privacy-sensitive data sets while providing insights into evolving user interactions on social media. At the micro-scopic level, we will investigate the presence (or lack of) different types of *influential* nodes.
- **Task 2:** Develop quantitative models to study cascades and co-evolution (competition as well as co-operation) across multiple networks that may share nodes, be connected by edges, or have multiple dependencies between them.
- **Task 3:** Investigate semi-streaming algorithm solutions for answering queries and testing hypotheses encountered in Task 1 and 2; and ultimately for compressing and archiving UAG data.
- **Task 4:** Evaluate/validate our models and algorithms using a wide variety of empirical user activity data from OSN platforms/applications.

Broader Impacts: The tangible outputs of the proposed project include publication of novel quantitative models and measurement algorithms that will provide a strong fundamental basis for studying key interaction behavior of other complex systems/networks (e.g., work flows in open software systems). The project also provides training of postdocs, graduate, and undergraduate students in multi-disciplinary team environment, which is becoming increasingly important for future work force. All PIs are committed to outreach efforts across all levels of the educational spectrum and to broaden the participation of under-represented groups, e.g., K-14 Outreach Center, NSF funded RET project, & Women in Engineering program at UC Davis and Undergraduate Research Opportunities in Computing at Georgia Tech.

Keywords: Network growth model; dynamic evolution; online social interactions; cascades; graph streaming