Science & Technology Innovation Concept Knowledge-base (STICK): Monitoring, Understanding, and Advancing the (R)Evolution of Science & Technology Innovations

Ping Wang (PI and Presenter, pwang@umd.edu), Yan Qu (Co-PI), Ben Shneiderman (Co-PI)
University of Maryland, College Park

Why do some science and technology innovations come to be highly popular and successful, transforming how we live, work, and play; whereas many other innovations fail? This is a question of great significance to innovation scholars, practitioners, and policy makers, all having employed various methods and developed or applied multiple, sometimes competing theories to answer the question. Despite our sustained efforts, we are still in search of a definitive answer – a comprehensive theory that explains the success and impacts of S&T innovations and coherent policies that effectively guide public and private investments in innovations to maximize economic prosperity and social welfare.

1. The analytical framework of the project, the data involved and the techniques employed

Our project helps innovation research and policy-making overcome theoretical and methodological barriers. **Theoretically**, one stream of innovation research is primarily focused on the **production** of innovations, examining basic and applied research, product development, quality control, and commercialization. The other stream is mainly focused on the **use** of innovations, addressing sense-making, adoption, implementation, and assimilation. Unfortunately, the two streams rarely converge to show the whole picture of innovation supply and demand, limiting the explicability of extant theories. To unify these separate steams, we propose the notion of **innovation community** - a set of organizations and people, such as universities, corporations, venture capitals, media, and regulators, interested in a specific innovation. Such a community emerges to make sense of the innovation and orchestrate material activities for the innovation. As innovations are interrelated, so are their communities. Thus, innovations, communities, and the relationships among innovations, people, and organizations constitute the **innovation ecosystem**.

Methodologically, the project provides much needed data and tools for studying the diverse entities and their complex relationships in the innovation ecosystem. Specifically, we are developing a large-scale, multi-source, longitudinal database, **Science & Technology Innovation Concept Knowledge-base (STICK)**, and a set of visual analytic tools for monitoring and understanding the emergence and revolution/evolution of innovations in three exemplar S&T fields: information technology, biotechnology, and nanotechnology. The knowledge-base captures data about innovations, individual and organizational actors associated with the innovations, and the relationships among the innovations and the actors through a hybrid approach that combines computational analysis of text (e.g., natural language processing-NLP) and social information processing (e.g., social tagging and collaborative writing). State-of-the-art visualization tools are customized for STICK users to visualize and analyze innovation networks and trends.

2. The current status of the research

System development: Using human-supervised automated techniques, we have downloaded published articles from bibliographic databases such as Lexis-Nexis, ProQuest Newspapers, ProQuest Dissertations & Theses, ACM Digital Library, and IEEE Xplore Digital Library, patent data from the U.S. Patent & Trademark Office Patent Full-Text and Image Database, grant data from NSF's Award database, and corporate data from Wharton Research Data Services. While we are still expanding our data collection in other sources, our initial dataset captures the innovation activities undertaken by diverse players in academia, industries, and government agencies.

From the numeric and textual data we have collected, we are extracting innovation entities and

relations that populate the STICK system. The entities include various individual and organizational actors and the relations include numerous relationships among the innovations, between the innovations and the people/organizations, and among the people/organizations. For example, we have customized commercial NLP tools to automatically extract organizations' and people's names from the textual data and then we have employed Amazon Mechanical Turk to correct the errors made by automatic tools (Sayeed et al. 2010a). We have also developed a statistics-based technique to automatically detect the opinions about each innovation expressed by the members of innovation communities (Sayeed et al. 2010b). In addition, we have bundled together several information retrieval techniques (such as co-occurrence analysis and hierarchical clustering) to automatically classify innovations into meaningful categories at any level of abstraction (Tsui 2010). These human-supervised automatic tools provide the scalability we need to process large datasets.

Empirical findings: We have completed three empirical studies to demonstrate the utility of the STICK system we are building. **First**, from STICK we extracted data about the interactions between eight popular IT innovations (e.g., Enterprise Resource Planning, Customer Relationship Management, and Data Warehouse) and 109 *Fortune* 500 companies in a ten-year period (1994-2003). We found that the impacts of popular IT innovations on corporate reputation and executive compensation were significantly positive but transient, and that the impacts of popular innovations on company performance were initially negative, but positive in the long term (Wang 2010a and 2010b). These findings suggest that technology innovations influence different aspects of organizations in different ways and that organizations may adopt innovations for a reason but benefits for another reason. **Second**, focusing on how organizations realize value from innovations, we analyzed 36 case studies (stored in STICK) on cloud computing, one of today's hottest computer science and IT innovations. We have found that cloud computing helps organizations innovate in four ways: (1) measuring the process and outcome of innovations, (2) experimenting innovative options, (3) sharing technical infrastructure and knowledge resources, and (4) replicating innovative processes and practices in large scales. These findings illustrate how S&T innovations can spawn other valuable innovations. **Third**, we retrieved data from STICK to reconstruct the evolution of three computer science innovations (Shneiderman et al., 2010). The findings add support to those who argue that patents inhibit adoption and diffusion. Ironically, we found that the academically-oriented University of Maryland produced a more broadly adopted and commercialized technology, while the industrially-oriented Xerox PARC produced strong and more high-impact papers. In addition, the novel visualization tools developed in this study have the potential to facilitate scientific discoveries and innovations by identifying and establishing collaborations within and across innovation communities and among academia, industry, and government (Shneiderman 2008).

3. The relevance of the project to one or more specific science policy issues or needs

One important and enduring science policy issue is the constant debate over how federal government should allocate resources to fund different types of S&T innovations. Between basic and applied research, some argue that offshoring R&D would hurt America's competitiveness; others contend that breakthroughs actually come from product development and commercialization, hence heavy investment in basic research may have been overrated. Even within the realm of applied technology innovations, some argue for less direct federal involvement and prefer free market in deciding innovations worthy of investment; others support direct investment by the government serving as a catalyst for cooperation. It is difficult to calculate the returns on investments in different innovations and settle such debate without the comprehensive data and visual analytic tools that this project aims to provide. STICK will help policy makers understand the evolutionary paths of innovations and industries, assess the significance of innovations in rigorously charted terrains, and proactively promote innovations that bring the greatest benefit to our society.

Acknowledgements
This research is supported by the National Science Foundation's Science of Science and Innovation Policy (SciSIP) program under grant SBE-0915645 and Human and Social Dynamics program under grant IIS-0729459.

Our Publications Cited

Sayeed, A., Meyer, T., Nguyen, H., Weinberg, A., and Buzek, O. Crowdsourcing the Evaluation of a Domain-Adapted Named-Entity Recognition System, in *Proceedings of Human Language Technologies: The Annual Conference of the North American Chapter of the Association for Computational Linguistics*, Los Angeles, CA, 2010a, pp. 345–348.

Sayeed, A., Meyer, T., Nguyen, H., and Weinberg, A. "Expresses-an-Opinion-about": Using Corpus Statistics in an Information Extraction Approach to Opinion Mining, in *Proceedings of the 23rd International Conference on Computational Linguistics*, Beijing, China, 2010b.

Shneiderman, B. Science 2.0, *Science* (319:5868), 2008, pp. 1349-1350.

Shneiderman, B., Wang, P., Qu, Y. and Dunne, C. Analyzing Trends in Science & Technology Innovation, in *Proceedings of the 27th Annual Human-Computer Interaction Lab Symposium*, College Park, MD, 2010.

Tsui, C.-j., Wang, P., Fleischmann, K.R., Oard, D.W., and Sayeed, A.B., Exploring the Relationships among ICTs, in *Proceedings of the 43rd Hawai'i International Conference on System Sciences*, Kauai, HI, 2010.

Wang, P., 2010a, Chasing the Hottest IT: Effects of Information Technology Fashion on Organizations, *MIS Quarterly*, 34(1), pp. 63-85.

Wang, P., 2010b, Surprising Impact of Fashions in Information Technology, to appear in *MIT Sloan Management Review*.