Social Genome: Putting Big Data to Work for Population Informatics

In the digital era, our daily activities from birth until death leave digital crumbs all over cyberspace. Collectively, these digital traces form a population's social genome, the collective footprints of our society. And we have the technology to gather and analyze these crumbs to obtain crucial insights into the most challenging problems in society such as healthcare. Population informatics is the burgeoning field at the intersection of SBEH (Social, Behavioral, Economic, & Health) sciences, computer science, and statistics that applies quantitative methods and computational tools to answer fundamental questions about human populations. Social genome data can tell us about how people live, work, respond to change, and make decisions, and most importantly the collective impact of these individual decisions. Such insights help us understand the root causes of social and public health problems, predict the downstream effects of different policy options, and allocate our collective resources for the greatest impact. Just as bioinformatics has revolutionized biological research, population informatics could catalyze significant advances in our understanding of trends in society, health, and human behavior. However, issues around privacy, confidentiality, access, and data integration have slowed progress. Here, we propose a new paradigm that regards person level data as valuable but hazardous research material. Integrated data about populations can hold the key to transforming SBEH sciences to a new level of evidence. Yet, when handled improperly, there is the potential for serious privacy violations that can undermine the public trust in research. This talk presents the new paradigm and the privacy-by-design approach to protection that focuses on building a safe environment for population informatics. We will cover basics of big data, data science, population informatics, information privacy, and how scientists can safely handle social genome data in their research.

Dr. Hye-Chung Kum is an associate professor at the School of Public Health at Texas A&M. She holds a joint appointment in the Department of Computer Science & Engineering and the Department of Industrial Systems & Engineering. She received her Ph.D. (2004) in Computer Science and MSW (1998) in Policy and Management from University of North Carolina at Chapel Hill. She is the founder of the Population Informatics Research Group which applies informatics, data science, and computational methods to the increasingly large digital traces available about people to advance public health, SBEH sciences, and population research by bringing together domain experts and computer scientists. Her vision paper on population informatics and social genome was published in the IEEE Computer Special Outlook Issue in January 2014.

Time and Date | 1:50 – 2:40 pm, April 22, 2016
1005 Emerging Technologies Building (ETB)
Faculty host: Dr. Yu Ding
Phone: 979-458-2343, Email: yuding@tamu.edu