

**NYU Stern School of Business**  
**Department of Information, Operations & Management Sciences**  
**INFORMATION SYSTEMS RESEARCH SEMINAR**

**TOPIC: Talkographics: Using What Viewers Say Online to Calculate Audience Affinity Networks for Social TV-based Recommendations**

**SPEAKER: Shawndra Hill (UPenn)**

**DATE: Thursday, December 5th, 2013**

**TIME: 4:00-5:30pm**

**PLACE: KMC 3-110**

**ABSTRACT**

Viewers of TV shows are increasingly taking to online sites like Facebook and Twitter to comment about the shows they watch as well as to contribute content about their daily lives. We present a novel recommendation system (RS) based on the user-generated content (UGC) contributed by TV viewers via the social networking site Twitter. In our approach, a TV show is represented by all of the tweets of its viewers who follow the show on Twitter. These tweets, in aggregate, enable us to reliably calculate the affinity between TV shows and to describe how and why certain shows are similar in terms of their audiences in a privacy friendly way. This paper's two main contributions are: 1) a new methodology for collecting data from social media — including information about product networks (or how shows are connected through users on a social network), geographic location, and user-contributed text comments — which can be used to generate affinity networks and test them; and 2) a new privacy friendly UGC-based RS that relies on all publicly-available text contributed by viewers, as opposed to only pre-selected keywords extracted from the UGC associated with the shows, a specific ontology or taxonomy, which makes our approach more flexible and generalizable than those used in any prior research. We show that our approach predicts remarkably well the TV shows that Twitter users follow. We also explain why the approach works so well: First, we show that the UGC reflects the demographics, geographic location, and psychographics of viewers, and coin the term talkographics to refer to descriptions of a TV show's viewers — or in general any product's audience — that are revealed by the words used in text messages sent by Twitter-using TV viewers; second, we show that Twitter text can represent many complex nuanced combinations of the demographic, geographic, and psychographic features of the audience; third, we show that we can use talkographic profiles to first calculate similarities between TV shows, then use these similarities reliably in RSs; we also show that our approach can be combined with a product association network approach to achieve even better recommendations; finally, we show that our text-based approach performs best for shows for which there is a demographic bias to the viewing audience compared to those that do not have a demographic bias. To demonstrate that our RS is generalizable, we apply the same approach to followers of clothing and automobile retailers.

## BIO

Shawndra Hill is an Assistant Professor in Operations and Information Management at the Wharton School of the University of Pennsylvania. Generally, she studies data mining, machine learning and statistical relational learning and their alignment with business problems. Specifically, she researches the value to companies of mining data on how consumers interact with each other—for targeted marketing, advertising and fraud detection. Her current research focuses on the interactions between TV content and Social Media ([www.thesocialtvlab.com](http://www.thesocialtvlab.com)). Her past and present industry partners include AT&T Labs Research, Mass Relevance, and Siemens Energy & Automation. Her work appears in IEEE Transactions on Data and Knowledge Engineering, Journal of Computational and Graphical Statistics, SIGKDD Explorations, and Statistical Science. Her research is funded, in part, by the Office of Naval Research, Google, and the National Institutes of Health (NIH). Shawndra holds a B.S. in Mathematics from Spelman College, a B.E.E. from the Georgia Institute of Technology and a Ph.D. in Information Systems from NYU's Stern School of Business.