

# ACMS Statistics Seminar

**Glenn Strycker**  
Conversant, LCC

**Tuesday, October 8, 2019**  
**154 Hurley Hall**  
**3:30 - 4:30 PM**



## **Stable, Robust Community Detection on Weighted, Undirected, Dynamic Graphs at Scale**

Conversant is a marketing tech company that provides personalized online digital advertising on behalf of large retail clients and other verticals. Our goal is to identify real people anonymously and assign online touchpoints to individual profiles as accurately as possible, without the use of PII. To accomplish this, we have a proprietary "Identity Graph" consisting of various anonymized IDs and online touchpoints that belong to individual, real-life persons. By constructing profiles from clusters in this graph, we are able to send personalized ads to individuals, based on their purchase histories and browsing behaviors, without ever knowing names or addresses.

My team's goal is to extend our capabilities into inferential edge networks and group IDs into household-level profiles. Analogous to constructing anonymous individual profiles, the challenge is that without PII, we lack ground truth. We must infer correctness and measure our accuracy through proxies and experiments. To make matters more complicated, inferred networks are typically over-connected and contain large connected components. We need to cleanse our data, break up large components into clusters of reasonable size, and finally prove that these are 1:1 with households. These clusters must be stable over time, despite edges entering and leaving the graph. The amount of noise and dynamism in inferred networks makes this a tough problem.

This problem maps well to "community detection on weighted, undirected, dynamic graphs", and the big issues for us are scale and consistency. We have BILLIONS of IDs and need to cluster these into roughly 100s of millions of real-life households, and we need these clusters to be consistent over time. Our algorithms need to run within hours or days at most, and on a distributed computing platform. Our solution space is heavily constrained by computability and business goal requirements, making standard algorithms difficult to use out-of-the-box.

We are looking to academia to see if this is an interesting unsolved problem, and if there are existing algorithms and past research that could be recommended. Additionally, we want to help academic researchers understand the real-world issues facing industry use of graphs.

The Department of Applied and Computational  
Mathematics and Statistics

Please visit [acms.nd.edu](http://acms.nd.edu) to view the full list of speakers.