ABSTRACT
A new era of data analytics of online social networks promises tremendous high-impact societal, business, and healthcare applications. As more users join online social networks, the data available for analysis and forecast of human social and collective behavior grows at an incredible pace.

The first part of this talk introduces an apparent paradox, where larger online social networks entail more user data but also less analytic and forecasting capabilities [7]. More specifically, the paradox applies to forecasting properties of network processes such as network cascades, showing that in some scenarios unbiased long term forecasting becomes increasingly inaccurate as the network grows but, paradoxically, short term forecasting – such as the predictions in Cheng et al. [2] and Ribeiro et al. [7] – improves with network size. We discuss the theoretic foundations of this paradox and its connections with known information theoretic measures such as Shannon capacity. We also discuss the implications of this paradox on the scalability of big data applications and show how information theory tools – such as Fisher information [3, 8] – can be used to design more accurate and scalable methods for network analytics [6, 8, 10]. The second part of the talk focuses on how these results impact our ability to perform network analytics when network data is only available through crawlers and the complete network topology is unknown [1, 4, 5, 9].

Categories and Subject Descriptors
H.2.8 [Database Management]: Database Applications—Data Mining

Keywords
Social networks; complex networks; network process forecast; information diffusion

Acknowledgements
This work was partially supported by NSF grants CNS-1065133 and ARL Cooperative Agreement W911NF-09-2-0053. The views and conclusions contained in this document are those of the author and should not be interpreted as representing the official policies, either expressed or implied of the NSF, ARL, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

1. REFERENCES