

Untangling Dark Webs: Theories, Methods, and Models for a Computational Social Science of Clandestine Networks

Half-day tutorial

Brian Keegan^{*}, Muhammad Ahmad[†], Sophia Sullivan^{*}, Dmitri Williams[^], Jaideep Srivastava[†], Noshir Contractor^{*}

^{*} {bkeegan, nosh}@northwestern.edu, School of Communication,
Northwestern University 2240 Campus Drive, Evanston, IL 60201

[†] {mahmad, srivasta}@cs.umn.edu, Department of Computer Science and Engineering, University of
Minnesota 200 Union St., Minneapolis, MN 55455 612-625-4002

[^] dmitri.williams@usc.edu, Annenberg School of Communication,
University of Southern California 3502 Watt Way, Los Angeles, CA 90089

The explosion of digital trace data in from computer-mediated communication technologies, online communities, and social media offers exciting new ways to understand large-scale social phenomena [1]. A key insight for understanding social behavior in both online and offline domains is situating it in larger and more complex networks of user relationships and interactions. These networks emerge, stabilize, and dissolve in response to a variety of endogenous and exogenous processes operating at multiple levels of analysis [2]. Extracting these multi-relational data for network analyses requires novel computational approaches for not only dealing with petascale data, but also imputing relationships from other types of digital trace data. The emerging fields of computational social science, social computing, and web science demand interdisciplinary approaches to developing theories and methods for describing and modeling social behavior so to better answer fundamental questions about general social dynamics.

Massively-multiplayer online games (MMOGs) like *World of Warcraft*, *EverQuest II*, and *EVE Online* have tens of thousands of players who have complex social interactions and organizations within virtual worlds. To the extent that individuals in online games operate under similar motivations and constraints as is encountered in offline contexts, their behavior in online worlds can potentially be generalized and mapped to better understand offline social behavior [3]. MMOGs not only have complex market economies where players barter and exchange for items and services, but illicit markets for other goods and services also exist. “Gold farming” is one example in which players pay real money to outside third-party vendor to supply them with in-game currency which allows players to accelerate their progress through the game. The player community and game developers oppose this practice for a variety of economic, legal, and ludic reasons. Thus the agents who procure these virtual currencies and items within the game face the challenge of how to organize themselves to both avoid detection while providing a very profitable service.

Clandestine and deviant behavior is still fundamentally socially-mediated and the fields of social computing and computational social science need to develop theories, methods, and metrics for characterizing and understanding them. These clandestine networks of gold farmers thus provide a model case for understanding larger issues about how clandestine organizations in both online and offline contexts emerge, stabilize, and dissolve. The intrinsically computer-mediated nature of MMOGs allows unobtrusive and exhaustive collection of data about how these organizations structure themselves and evolve over time. Integrating network analysis, statistical models, and machine learning tools with this large-scale data could allow the development of algorithms which improve detection methods, predictive

models of how organizations change over time, and metrics for assessing and evaluating the reach and influence of these organizations.

In this tutorial, we will review our prior work which has developed predictive models of likely gold farmers [4], characterized the structure and resilience of their trade networks [5], the ways in which these trade networks emerge from complex trust networks [6], and the ways in which these networks are structured similarly to – and thus potentially generalizable to – offline drug trafficking networks [5, 7]. We discuss how to integrate methods developed for other contexts to a dark/clandestine network context, how to reconcile these methods with the validity concerns about mapping between very different contexts, extending these approaches to other information/communication technology and social media contexts, and the implications for both privacy and “dual use” implementations of this line of research.

- [1] D. Lazer, *et al.*, "Life in the network: the coming age of computational social science," *Science (New York, NY)*, vol. 323, pp. 721-723, 2009.
- [2] P. R. Monge and N. S. Contractor, *Theories of communication networks*. New York: Oxford University Press, 2003.
- [3] D. Williams, "The mapping principle and a research framework for virtual worlds," *Communication Theory*, vol. 20, pp. 451-470, 2010.
- [4] M. A. Ahmad, *et al.*, "Mining for Gold Farmers: Automatic Detection of Deviant Players in MMOGs," in *Proceedings of IEEE SocialCom 2009*, Vancouver, Canada, 2009, pp. 340-345.
- [5] B. Keegan, *et al.*, "Dark Gold: Statistical Properties of Clandestine Networks in Massively Multiplayer Online Games," *IEEE Conference on Social Computing*, 2010.
- [6] M. A. Ahmad, *et al.*, "Trust Amongst Rogues? A Hypergraph Approach for Comparing Clandestine Trust Networks in MMOGs," in *International Conference on Weblogs and Social Media*, Barcelona, Spain, 2011.
- [7] B. Keegan, *et al.*, "Sic Transit Gloria Mundi Virtuali? Promise and Peril at the Intersection of Computational Social Science and Online Clandestine Organizations," presented at the 3rd International Conference on Web Science, Koblenz, Germany, 2011.

Tutorial Outline

- [0:00-0:15] Defining clandestine behavior
 - Social and organizational behavior of clandestine organizations
 - Clandestine behavior in different contexts
 - Goals: analyze behavior, improve detection, identify undetected nodes
- [0:15-0:45] Modeling clandestine organizations and behaviors as networks
 - Properties of complex networks – small world, power law degree dist., etc.
 - Multiple types of attributes – demographic, psychological, system-specific
 - Multiple types of ties – trust, communication, affiliation, exchange, etc.
 - Dynamics – changes and stability of nodes, links, attributes over time
 - Data collection problems – limits of traditional methods, problems with censoring, sampling
 - Pros and Cons of using large-scale computer-mediated behavioral data
- [0:45-1:00] Methodological approaches
 - Deductive approaches: goals of hypothesis testing, benefits and limitations
 - Inductive approaches: goals of machine learning, benefits and limitations
 - Integrating deductive and inductive approaches, review P, R, F
 - Revisit goals to analyze behavior and identify/predict undetected nodes
- [1:00-1:15] Introduction to MMOG, gold farming, and EQ2 data
 - What's an MMOG and what do people do in it
 - What's gold farming, why is it a problem, how is it like other forms of trafficking
What kind of data do we have and use, what analogues available offline?
- [1:15-1:30] Break
- [1:30-2:15] Descriptive network properties
 - Differences in centrality types, clustering
 - Assortativity
 - Attack tolerance
 - Mixed-Membership Networks: trust, mentoring, item transaction
 - A multi-network multi-level view of clandestine networks
- [2:15-2:55] Predictive modeling
 - Machine Learning approaches
 - Structural Signatures based approaches
 - Mixed Membership dynamic modeling
 - Label Propagation Approaches – similarity based, network propagation approaches
 - Link Prediction problems in Clandestine Networks
- [2:55-3:00] Conclusion