Chapter 19
Manually Profiling Egos and Entities across Social Media Platforms: Evaluating Shared Messaging and Contents, User Networks, and Metadata

Shalin Hai-Jew
Kansas State University, USA

ABSTRACT
Social media accounts on various social media platforms represent the public-facing Web presences of egos (individuals) and entities (groups). On the surface, these may be understood based on their profiles, their shared contents and postings, and their interactions with other user accounts online. A number of software tools and analytical techniques enable further analyses of these accounts through network analysis, content analysis, machine-based text summarization, and other approaches. This chapter describes some of the capabilities of “manual” or semi-automated (vs. fully automated) remote profiling of social media accounts for insights that would not generally be attainable by other means.

INTRODUCTION
Social media platforms are considered to be interactive spaces where individuals and groups congregate, socialize, share, intercommunicate, interact, and otherwise engage. An “ego” is understood as a person, a persona, or some type of agent with awareness, perspective, and will. An ego is understood to have preferences or “biases.” An “entity” is understood as a group, with an identity, purpose, impetus, resources, and will. When people want to “check out” each other online, they will visit their respective social media accounts online, peruse shared contents, explore messaging and images, see how others interact with that individual online, and often call it good. In some cases, the curious may have to create an account in order to check out others’ identities on a particular closed social media platform (and often inadvertently leave a signature on an online guestbook to record the visit). Some may actually engage with their target, by posting messages and eliciting responses.

DOI: 10.4018/978-1-5225-0559-4.ch019
In general, they pursue information that is readily available, maybe with links appearing in the top few pages of a Google search. Beyond the Surface Web, they may tap some portals to the Deep Web (websites not easily trawled using classic web browsers built to access http-based web pages) to check out people’s government records, legal records, property records, marital status, pay information, and other data. They may look up property records and map those to a location. The exploratory actions may be an expression of social, professional, personal, or other interest(s); such explorations may be an extension of their own social performances in engaging with others—by building up an online network that demonstrates clout or attractiveness or wealth or some other socially desirable feature. In general, people pursue information that is readily available. They pursue information that is generally processed (vs. raw). What they are leaving untapped is latent (hidden, non-obvious) information that is not so directly readily available. They are often not exploiting data leakage from unintended revelations. They are not exploiting structural trace data (created from people’s interactions with a system). They are not using metadata or information about information.

A more technologically-based approach, which enables the collection of a wider range of open-source intelligence (OSINT), involves applying a semi-automated way of remote profiling egos and entities across social media platforms. This is considered a semi-automated approach because this is a human-supervised data extraction, a “manual” vs. a “fully-automated” (and “unsupervised” machine learning) approach. This approach involves analysis of three main areas of information:

1. Messaging and shared (multimedia) contents (for content analysis),
2. Trace data (for link analysis), and
3. Metadata (for network analysis, for categorical analysis, for spatiality analysis).

“Content data” on social media sites may include a variety of textual messages (such as “short message service” or “SMS” texts, microblogging messages, emails, and others), images, memes, audio files, video files, slideshows, and other contents. “Trace data” generally refers to “log data” or records of interactions (the interacting user accounts, the times of interactions, and other data). “Metadata,” broadly speaking, is labeling data about information. The applied analytical approaches for (1) content data involves content analysis (writ large), and machine-based text summarization; for (2) or trace data, it involves electronic social network analysis (e-SNA); and (3) for metadata, it involves related tags network analysis.

All three types of data analyses methods include extracted data visualizations (the mapping of the extracted data to graphs, maps, and other visuals). A fourth data type is human- and machine-coding, which may aid in the extraction of themes and other insights from the collected data. (This fourth type is beyond the purview of this chapter but is understood as an important element in data analysis.)

The intuition behind this chapter is that being aware of how an ego or entity manifests across social media platforms offers a deeper and more holistic understanding of the target at three main levels: the target node, its ego neighborhood, and its broader standing in a social media context. This is not to suggest that this approach reveals all, but a social media analysis approach is complementary to other ways of on-ground knowing. Another general insight is that social media platforms involve digital files in a variety of file format types. While textual communications are basic to virtual communities of all stripes and seminal in their “evolution, growth, and sustenance” (Bagozzi & Dholakia, 2002, p. 4), there is a lot of social media data that is heterogeneous and multimedia-based, which requires some additional work to enable information extraction (such as transcription and coding).