Chapter 3
Eavesdropping on Narrowcast Self-talk and Microchats on Twitter

Shalin Hai-Jew
Kansas State University, USA

ABSTRACT

On Twitter, a range of discourse networks may be extracted showing different types of conversational interactions. While the attention is often on what is trending and large-size high-interactive social graphs, many extracted networks are self-loops and small-group discourse networks based on ad hoc narrowcast conversations. In this exploratory study of microblogging messaging on Twitter, the focus is on microblogging conversations that result in self-loops (self-to-self conversations, individuals microblogging to themselves) and small-group graphs and motifs (one-to-few or few-to-few conversations). This work proposes and tests hypotheses about the various types of seeding #hashtags and keywords that result in different types of ad hoc microblogging microchat network graphs on Twitter.

INTRODUCTION

The popularization of social media platforms came with widely expressed hopes for global-scale real-time informational awareness and pro-social mass-collaborations. The participation of large swaths of the world's people in crowd-sourcing information and interacting on social media has further enabled the extraction of "big data" for novel research and insights. Twitter, one of the world's largest microblogging sites (with 500 million users as of late 2014), is one of the most popular platforms. On their site, they list 288 million monthly active users, with 500 million Tweets (messages) sent daily. Some 77% of social media accounts are outside the U.S., and the platform has support for 35 languages ("About Twitter," 2015). The U.S. Library of Congress has stepped forward to archive all Tweets ("Library of Congress…" Jan. 22, 2013). The popularization of Twitter, now in its 8th year, is impressive. Many have built this "SMS of the Internet" ("short message service of the Internet") into daily lives, work flows, law enforcement, government planning, marketing, and a range of applications.

DOI: 10.4018/978-1-4666-8696-0.ch003

Messages on Twitter are limited to 140 characters but may include images, video, audio, and URLs (uniform resource locators). In the grammar of the tool, user accounts are listed @name, which may be used in “reply,” “retweet,” and “mentions”. Messages may be favorite-d or “liked”. Messaging may be made public or private. Conversations based on particular themes may be labeled with #hashtags, some of which are disambiguated and unique, and others which are generic. In real-world use, hashtags may denote tone or context. They may be used to indicate the particular use of language, as a linguistic marker (#sarcasm, #humor, or #irony); “hashtags are the digital extra-linguistic equivalent of non-verbal expressions that people employ in live interaction when conveying sarcasm” (Kunneman, Liebrecht, van Mulken, & van den Bosch, 2014, p. 1). Other types of related conversations may be identified through the use of keywords, for various keyword-word-sense uses (for multiple theme captures). Interactions around shared information may be understood as Twitter discourse or conversations, what some have termed “conversation-like,” because of the constraints of the 140 characters and the designed structure of Twitter. For many, there is challenge tracking interactions and a sense of partial or truncated messaging. There is a sense of missing contextual detail. (The sparse and minimalist approach of Twitter messaging has encouraged a range of spin-off practices, such as thumb-typed “novels” from mobile devices, 140-character tiny Twitter recipes, and other types of brevity.)

Twitter Inc. makes much of its data usable for research and user self-awareness. Publicly-available account information, the messaging, and various types of interactivity may be captured through data extractions through Twitter’s application programming interfaces (APIs).

Twitter offers two application programming interfaces (APIs) for collecting tweets: one is the search API, which may be used to retrieve past tweets matching a user specified criteria; the other is the streaming API, which may be used to subscribe to a continuing live stream of new tweets matching a user defined criteria and delivered to the user as soon as they become available. Researchers need not define any specific criteria to receive data from the streaming API. They can receive a (free) 1% sample of everything posted each day, as it is posted (Burnap, Rana, Avis, Williams, Houseley, Edwards, Morgan, & Sloan, 2013, p. 3).

A number of software tools have been created that tap Twitter’s APIs. What this means is that, theoretically and practically, every node and network is theoretically highly visible. (This assumes the inclusion of the commercial company and its tools and resources. This company has access to the full databases of Twitter for full data extractions. Publicly available and non-commercial API data extractions are rate-limited and are also content-limited.) There is a lot of information available even given the limits of human attention. With the application of automated data extractions (based on much more expansive machine attention), machine-based text summarization, text frequency counts, and text searches, the capabilities of datamining from the Twitter corpus is enormous and can extend what is knowable.

While some initially saw the brevity requirement of Twitter messages as a negative because of the challenges to conveying complex information, others see the character limits as a barrier-lowering function to encourage people to share (Eyal & Hoover, 2014, p. 70). The ethos of the platform is to provide near-constant status updates to friends, family, and other followers. This sense of over-sharing at the local level may be harnessed as a public good.

Social media accounts on Twitter may be powered by humans, groups, cyborgs (humans and machines), robots, or sensors. For so-called human sensor networks to work, the people within those ad hoc networks have to be constantly capturing and sharing out information, signaling, if you will. By