A LONGITUDINAL STUDY OF TEMPORAL PATTERNS AND ENGAGEMENT WITH TWITTER DURING LE TOUR DE FRANCE (2013 – 2015)

All authors:
Orland Hoeber (corresp), Larena Hoeber, Ryan Snelgrove, Laura Wood

Synopsis:
We will present a longitudinal analysis of the patterns of tweeting during the last two and the upcoming instances of Le Tour de France. The temporal patterns of tweeting and retweeting will be analyzed during the multi-day events, and the stakeholders will be classified based on their level of engagement. Implications for sport management practice and research will be discussed.

Abstract:
While Twitter has been embraced and promoted as a mechanism to enhance communication among sport organizations, athletes, media, and fans (Hambrick, Simmons, Greenhalgh, & Greenwell, 2010), the ability to analyze the content of the tweets and the underlying opinions of the stakeholders has been constrained by the researchers' technical capabilities and the limitations of data analysis software. As a result, many studies resort to sampling methods to reduce the collection to a manageable size, but at the cost of potentially missing important content or losing the temporal relationships inherent in the data. Some of these studies have been critiqued for being primarily descriptive (Hardin, 2014), resulting in a call for better technology to support the qualitative analysis of big datasets (Tinati, Halford, Carr, & Pope, 2014).

In prior work, we have developed a software system called Vista (Visual Twitter Analytics) to support visual exploration and analysis of the temporally changing sentiment (positive, neutral, and negative) within Twitter data. This software follows a visual analytics approach, combining data processing and machine learning technology with information visualization and human-computer interaction methods, for the purpose of supporting data exploration, analytical reasoning, information synthesis, hypothesis development and testing, and human decision making (Keim et al., 2008). This software supports the real-

time collection and storage of tweets matching pre-determined queries, enabling the discovery and analysis of emerging trends in the data.

The purpose of this presentation is to use Vista to examine, in a longitudinal setting, the use of Twitter during Le Tour de France (2013 – 2015). Specifically, we will identify temporal patterns found within the data in each of the three years, and conduct an analysis of the level of stakeholder engagement. Using the official and highly publicized hashtag of the event (#tdf), we have collected data (English language tweets) for the 2013 and 2014 events, and will have collected data for the 2015 event. This data covers the days of the race, as well as the week preceding and the week following. The datasets consist of over 409,000 tweets for the 2013 race, and over 510,000 tweets for the 2014 race. Longitudinal data is beneficial in this case because it allows for an identification of how tweeting behaviour has changed or remained consistent over time and enhances the validity of the results.

Preliminary analysis of the 2013 and 2014 data shows a 24% increase in the number of tweets posted, and reveals a similar temporal pattern in tweeting behaviour. In particular, there are large spikes in tweets at the end of each stage, as well as during exciting moments during the race, illustrating the level of stakeholder engagement. There is a noticeable difference in the number of tweets posted during mountain stages (high) and time trials (low), illustrating the difference in fan engagement between these two aspects of the race. This information may be useful for sponsors to know where to direct their efforts, or for organizers to package sponsorship of the low-engagement stages with high-engagement stages. Furthermore, organizations may wish to tweet content or advertisements during these peak times to engage Le Tour de France fans when they are most likely to be using Twitter.

We will also classify the stakeholders based on their level of engagement throughout the races. Because we have collected all of the tweets that use the official hashtag for the event over the last three years, we can analyze the distinction between different types of stakeholders, such as creators (those who post a large amount of original content), repeaters (those who predominantly re-tweet content), and casual followers (those who post very little original content nor re-tweet). In particular, we will classify the stakeholders along two dimensions: the degree of original tweeting, and the degree of re-tweeting, resulting in a two-dimensional classification. Comparing this across the three years of data will reveal trends in the level of engagement, and will allow us to identify important stakeholders beyond those that can readily be identified by their exceptionally high engagement. As sport and media organizations increasingly promote the use of official Twitter hashtags it will be important for them to understand how the hashtags are being used (e.g., passive re-tweeting of media content or active tweeting of unique content) and how the engagement is changing over time.

References:
Hambrick, M. E., Simmons, J. M., Greenhalgh, G. P., & Greenwell, T. C. (2010). Understanding Professional Athletes' Use of Twitter: A Content Analysis of Athlete Tweets. International Journal of Sport Communication, 3(4), 454–471.
Hardin, M. (2014). Moving Beyond Description: Putting Twitter in (Theoretical)

Context. Communication & Sport, 2(2), 113–116.
doi:10.1177/2167479514527425

Keim, D. A., Andrienko, G., Fekete, J.-D., Görg, C., Kohlhammer, J., & Melançon, G. (2008). Visual Analytics: Definition, Process, and Challenges. In A. Kerren, J. T. Stasko, J.-D. Fekete, & C. North (Eds.), Information Visualization: Human-Centered Issues and Perspectives (pp. 154–175). Springer.

Tinati, R., Halford, S., Carr, L., & Pope, C. (2014). Big Data: Methodological Challenges and Approaches for Sociological Analysis. Sociology, 48(4), 663–681. doi:10.1177/0038038513511561