TECHNIQUES OF INTERSECTION – METRICS AS CONNECTORS AND SEPARATORS IN TWITTER RESEARCH

Carolin Gerlitz
University of Amsterdam

Bernhard Rieder
University of Amsterdam

Social media platforms are characterized by the sheer volume of activity and thus data, whilst at the same time only offering very specific and often limited access possibilities. When studying Facebook or Twitter, questions and techniques of *delimitation*, i.e. the selection of subsets and the use of specific metrics, are particularly relevant. This paper sets out to reflect on metrics in Twitter research as specific techniques that allow to investigate the making of boundaries and connections, while at the same time functioning both as connectors and separators. The question of cutting and slicing has been discussed in relation to sampling techniques in the context of social science research (Uprichard; Bryman; Gilbert) and increasingly in relation to social media research (Rieder; Gerlitz and Rieder), but the kind of analytical "dissections" we perform by selecting and designing metrics are less well understood. Metrics, however, equally need to be seen as epistemic devices that engage in breaking the practices sprawling on social media platforms apart and put them together again in various ways. The objective of the paper is to reflect on the affordances of different metrics in social media research and in particular in Twitter research. We draw on a previous experimental study on 'mining one percent of Twitter' (Gerlitz and Rieder) by expanding our initial reflections on the epistemic work and boundary making capacities of sampling to include the more subtle forms of carving implied by metrics.

Dissecting platform data through features, metrics and samples

A variety of new media research has started to explore the material-technical conditions of platforms (Rogers; Gillespie), drawing attention to the performative capacities of platform protocols to enable and structure specific activities. In the case of Twitter that refers to elements such as tweets, retweets, @replies, favourites, follows, or lists. These features offer blueprints for user activities that are both pre-defined yet underdetermined (Gillespie; Montfort and Bogost; Langlois et al.). They cater to a variety of actors, including users, developers, advertisers, and third party services, and

allow for a variety of distinct use practices to emerge, but also often function as starting points to delineate empirical material for Twitter research.

Using these medium-specific elements as basis for building a collection of tweets, users, etc. to be analyzed has significant epistemic weight: these sampling methods come with specific notions of use scenarios build into them or, as Uprichard suggests, there are certain "a priori philosophical assumptions intrinsic to any sample design and the subsequent validity of the sample criteria themselves" (Uprichard 2). The practice of delimitation is not reserved to sampling, but continues on the level of metrics when we use, for example, hashtags or even words as stand-ins for concepts, ideas, positions or viewpoints. Is a list on Twitter a group? Are retweets a measure of attention? Of agreement? Or a favour amongst friends? Metrics draw together disparate practices, but they also isolate: we do not send hashtags, after all, but tweets. If platform features like hashtags or retweets comprise a variety of practices and meanings, how are these heterogenous practices again re-assembled in the process of research? And how can we use said metrics to trace the various connections and separations between users, topics and practices on Twitter?

Methodology

Following our interest in relations between sampling, metrics and medium-specificity, we turn to *random sampling* and engage Twitter without the commitments - or maybe with *different* commitments? - to particular *a priori* conceptualizations of practices implied by other sampling methods, in order to explore the ways metrics can reveal connections and separations while being themselves caught up in connecting and separating analytically. Using Twitter's Streaming API and the *Digital Methods Initiative Twitter Capture and Analysis Toolset* (DMI-TCAT), we retrieved a one percent sample of all tweets over a period of several weeks. We use traditional and emergent metrics to explore both the datasets and the metrics themselves, including activity patterns, hashtags, retweets, language settings, content shared, interaction chains, co-word associations, user profile data and access devices among others. By starting from a random sample, we can inquire how metrics are both related and separate from each other, and to what extent different metrics can qualify or characterise each other (Do Twitter users writing from different devices use hashtags differently? Are friend/follower relationships indicative of linking diversity? Can we detect language specific practices?). In line with our previous work, the random sample further allows us to expand our analysis of different use practices of platform features. For instance, by drawing on network and co-word metrics, we explore different use cases of hashtags, interaction and retweet patterns. Doing so, we inquire into the connections enabled by such metrics, asking how connections between hashtags, words or users resemble actual use practices or are rather constructed through the analytical work of metrics. In this methodological set-up, samples and metrics are not used to establish data boundaries, but to explore their making and in that sense, the paper contributes to a critical reflection of methodological techniques of intersection.

Making and tracing intersections

The paper concludes by a reflection on the ways in which metrics, operating as epistemic devices, contribute to dynamics of inclusion and exclusion on a methodological and conceptual level. Just as Mackenzie contents in relation to databases that it is not the individual data points that matter but the relations that can be created between them (Mackenzie), the work with metrics involves such bringing into relation of medium-specific objects and activities. Research, we argue, further needs to reflect whether it dissects its datasets alongside issue- oriented or medium-dynamics, and which metrics bring out the dynamics of the medium – such as bot activity or retweets peaks and which allow to zoom in on dynamics specific to topical concerns. Through experimental testing of metrics we hope to initiate a debate on the politics of cuts and connections produced and the possibilities to explore the complexity of Twitter practices within which specific collections and metrics can be situated.

References

Bryman, Alan. *Social Research Methods*. Oxford University Press, (2012).

Gerlitz, Carolin & Rieder, Bernhard. Mining One Percent of Twitter : Collections, Baselines, Sampling. *M/C Journal*, 16(2), pp.1–18, (2013).

Gilbert, Nigel. *Researching Social Life*. SAGE, (2008).

Gillespie, Tarleton. "The Politics of 'Platforms'." *New Media & Society* 12.3 (2010): 347–364.

Langlois, Ganaele et al. "Mapping Commercial Web 2 . 0 Worlds: Towards a New Critical Ontogenesis." Fibreculture 14 (2009): 1–14.

Mackenzie, Adrian. "More Parts Than Elements: How Databases Multiply." *Environment and Planning D: Society and Space* 30.2 (2012): 335 – 350.

Montfort, Nick, and Ian Bogost. *Racing the Beam: The Atari Video Computer System*. MIT Press, (2009).

Rieder, Bernhard. "The Refraction Chamber: Twitter as Sphere and Network." *First Monday* 11. November 5 (2012).

Rogers, Richard. *The End of the Virtual – Digital Methods.* Amsterdam University Press, (2009).

Uprichard, Emma. "Sampling: Bridging Probability and Non-probability Designs." *International Journal of Social Research Methodology* 16.1 (2011): 1–11.