Chapter 2

Platform effects

This is Chapter 2 of:

Momin M. Malik (2018). "Bias and beyond in digital trace data". PhD thesis. Pittsburgh, PA: Carnegie Mellon University. URL: http://reports-archive.adm.cs.cmu.edu/anon/isr2018/abstracts/18-105.html

It is an updated version of a paper previously published as:

Momin M. Malik and Jürgen Pfeffer (2016). "Identifying platform effects in social media data". *Proceedings of the Tenth International AAAI Conference on Web and Social Media*. ICWSM-16, pp. 241–249

If citing, please give references to both the originally published work and to this thesis.

ACM Momin M. Malik and Jürgen Pfeffer. 2016. Identifying Platform Effects in Social Media Data. In *Proceedings of the Tenth International AAAI Conference on Web and Social Media (ICWSM-16)*. 241–249.

APA Malik, M. M., & Pfeffer, J. (2016). Identifying platform effects in social media data. In *Proceedings of the Tenth International AAAI Conference on Web and Social Media (ICWSM-16)* (pp. 241–249).

Chicago Malik, Momin M., and Jürgen Pfeffer. "Identifying Platform Effects in Social Media Data." In *Proceedings of the Tenth International AAAI Conference on Web and Social Media (ICWSM-16)*, 241–249. 2016.

IEEE M. M. Malik and J. Pfeffer, "Identifying platform effects in social media data," *Proc. of the Tenth Int. AAAI Conf. on Web and Social Media (ICWSM-16)*, pp. 241–249, 2016.

MLA Malik, Momin M., and Jürgen Pfeffer. "Identifying Platform Effects In Social Media Data". *Proceedings of the Tenth International AAAI Conference On Web And Social Media (ICWSM-16)*, 2016, pp. 241-249,

BibTeX
```
@inproceedings{malik2016b,
    author = {Malik, Momin M. and Pfeffer, J\"{u}rgen},
    title = {Identifying platform effects in social media data},
    year = {2016},
    booktitle = {Proceedings of the Tenth International
                 AAAI Conference on Web and Social Media},
    pages = {241--249},
    series = {ICWSM-16}
}
```

ACM Momin M. Malik. 2018. Bias and Beyond in Digital Trace Data. Ph.D. Dissertation. Carnegie Mellon University, Pittsburgh, PA. Retrieved from http://reports-archive.adm.cs.cmu.edu/anon/isr2018/abstracts/18-105.html.

APA Malik, M. M. (2018). *Bias and beyond in digital trace data* (Doctoral dissertation, Carnegie Mellon University). Retrieved from http://reports-archive.adm.cs.cmu.edu/anon/isr2018/abstracts/18-105.html.

Chicago Malik, Momin M. "Bias and Beyond in Digital Trace Data." PhD dissertation, Carnegie Mellon University, 2018. http://reports-archive.adm.cs.cmu.edu/anon/isr2018/abstracts/18-105.html.

IEEE M. M. Malik, "Bias and beyond in digital trace data," Ph.D. diss., Inst. Softw. Res., Sch. Comput. Sci., Carnegie Mellon Univ., Pittsburgh, PA, 2018. Available: http://reports-archive.adm.cs.cmu.edu/anon/isr2018/abstracts/18-105.html.

MLA Malik, Momin M. *Bias and Beyond in Digital Trace Data*. 2018. Carnegie Mellon U, PhD dissertation. SCS Technical Report Collection, http://reports-archive.adm.cs.cmu.edu/anon/isr2018/abstracts/18-105.html.

BibTeX
```
@phdthesis{malik2018,
    author = {Malik, Momin M.},
    title = {Bias and beyond in digital trace data},
    year = {2018},
    school = {Carnegie Mellon University},
    address = {Pittsburgh, PA},
    month = {08},
    url = {http://reports-archive.adm.cs.cmu.edu/anon/isr2018/abstracts/18-105.html}
}
```

Summary. In this Chapter, I use the rapid introduction of Facebook's "People You May Know" as a natural experiment by which to observe the causal effect of a recommender system on user behavior. I theoretically frame this as an example of decisions of platform governance having a causal effect on user behavior, which has larger implications for how we think about the data we get from social media platforms.

Compared to the published version, I update all fits to nonparametric quantile fits (previously, I had used one high-order polynomial regression), and have added an extensive additional discussion of literature, including of examples of causal observational inference with social media data that I had previously missed (as well as one nearly simultaneous publication), as well as two important theoretical works I had missed.

2.1 Introduction

In social media data, the design and technical features of a given platform constrain, distort, and shape user behavior on that platform, which I call the *platform effects*. For those inside companies, knowing the effect a particular feature has on user behavior is as simple as conducting an A/B test (i.e., a randomized experiment), and indeed such testing is central to creating platforms that shape user behavior in desirable ways. But external researchers have no access to the propriety knowledge of these tests and their outcomes. This is a serious methodological concern when trying to generalize human behavior from social media data: in addition to multiple other concerns, observed behavior could be artifacts of platform design. This concern has thus far only been raised theoretically (Tufekci, 2014; Ruths and Pfeffer, 2014), and not yet addressed empirically. Even theoretically, the problem is deeper and more subtle than has been appreciated; it is not just a matter of non-embedded researchers having access to the data (Savage and Burrows, 2007; Lazer, Pentland, et al., 2009; Huberman, 2012; boyd and Crawford, 2012), but also that even when researchers have access, without full knowledge of the platform engineering and the decisions and internal research that went into design decisions, the data can be systematically misleading. This topic relates also to discussions in the humanities about the nature of social media platforms as governance and management entities (van Dijck, 2013; Gehl, 2014), and how models have a self-reinforcing property of creating the very reality they purport to describe or explain (Healy, 2015).

One way to study and quantify platform effects as an external researcher is to look for available data that include a significant platform change. Making the assumption that, in absence of the exogenous shock (the change) the previous 'trend' would have remained the same, we can apply the observational inference method of *regression discontinuity design* (Imbens and Lemieux, 2008; Lee and Lemieux, 2010; Li, 2013). While not as certain as experimental design, observational inference methods are the best available way for outside researchers to understand the effects of platform design.

As another theoretical contribution which directly anticipates the call of Lazer and Radford (2017), I argue that *data artifacts*, rather than being incidental or annoyances to be corrected (Roggero, 2012), are a rare place where usual order breaks down, which can provide a glimpse into otherwise inaccessible underlying mechanisms. Here, data artifacts are providing important insights into inner working of platform engineering and management.

I select two data sets: the Facebook New Orleans data collected by Viswanath et al. (2009), and the Netflix Prize data, described by Koren (2009a). The latter is no longer publicly available since the close of the Netflix prize, although the terms of use do not mention any expiration on use for those who have already downloaded it.

In the Netflix Prize data set, Koren (2009a), a member of the team that ultimately won the prize (Koren, 2009b), points out a curious spike in the average ratings in early 2004. As such a change has modeling implications (previous data should be comparable in order to properly use for training purposes), he explores the possible reasons for this, ultimately identifying an undocumented platform effect as the most likely driver. Then, the Facebook New Orleans data contain an identified, and ideal, example of a platform effect: a clear exogenous shock and a dramatic difference after, through the introduction of the "People You May Know" (PYMK) feature on March 26, 2008. This discontinuity is only mentioned in Zignani et al. (2014); the original paper of the data collectors (Viswanath et al., 2009) does not mention it (although, in another example of a platform effect in collected data, they do note that on July 20, 2008, Facebook launched a new site design that allowed users to "more easily view wall posts through friend feeds" which they use to explain a spike in wall posts towards the end of the collected data).

In sum, I re-analyze the Netflix Prize and Facebook New Orleans data to study possible platform effects in the data. The contributions of this paper are:

- To empirically verify previously expressed theoretical concerns about the possible effects of platform design on the generalizability and external validity of substantive (social scientific) conclusions;
- To import into the social media research community a statistical model that allows quantitative estimation of platform effects;
- To quantify two specific cases of common platform effects, the effect on a social network of a triadic closure-based recommender system and the effect of response item wordings on user ratings.

2.2 Background and related work

Authors from multiple disciplines (Tufekci, 2014; Ruths and Pfeffer, 2014) have expressed methodological concerns that the processes found in data derived from social networking sites cannot be generalized beyond their specific platform. Most troublingly, the same things that would cause results to not generalize, such as nonrepresentative samples, idiosyncratic technical constraints on behavior, and partial or uneven data access, are generally unknown and undetectable to an outside researcher (and potentially even to engineers and embedded researchers). Some innovative methods of data comparison have been used to derive demographic information in social media data (Chang et al., 2010; Mislove, Lehmann, et al., 2011; Sloan et al., 2013; Hecht and Stephens, 2014; Longley et al., 2015; Malik, Lamba, et al., 2015) and to identify biases in public APIs (Morstatter, Pfeffer, Liu, and Carley, 2013; Morstatter, Pfeffer, and Liu, 2014), but platform effects remain empirically unaddressed. Part of the problem is that social media platforms are private companies that seek to shape user behavior towards desirable ends, and do so in competition with one another (van Dijck, 2013; Gehl, 2014); thus, the details of features and functionality which successfully guide user behavior are understandably proprietary in ways that representation and data filtering need not be. The results of research experiments, most notably Kramer et al. (2014), deal only indirectly with platform design and engineering. Outside accounting via testing inputs (Diakopoulos, 2014) is an important way of identifying overall effective outcomes, but such cross-sectional audits lack a baseline to know how much a given platform design successfully shapes behavior.

Instead, one way to study the problem is the econometrics approach of finding cases that can be treated as 'natural experiments' (Angrist and Pischke, 2008; Gelman, 2009). I have located two such instances, the Facebook New Orleans data and the Netflix Prize data, where known or suspected change in the platform led to a shift, documented in publicly available data.

Zignani et al. (2014) used the data of the Facebook New Orleans network (Viswanath et al., 2009), along with data from the Chinese social networking site Renren, to investigate the delay between when it is possible for an edge or triangle to form (respectively, when a node enters the network, and when two nodes are unconnected but share a neighbor) and when it actually forms, which they respectively term *link delay* and *triadic closure delay*. They note that on March 26, 2008, there is a drastic increase in the number of links and triangles (my version of those plots given in figs. 2.1 and 2.2), corresponding to the introduction of Facebook's "People You May Know" (PYMK) functionality. While this was not the central investigation of their paper, they used it as an opportunity to see how an external feature changed their proposed metrics. They find that this increase consists primarily (60%) of links delayed by over 6 months, and also includes many (20%) links delayed by more than a year. They continue to note, "Although the link delay [metric] reveals interesting characteristic in edge creation process, it is not able to capture the reason behind it, i.e., which process causes the observed effects or which algorithms were active in the early rollout of the PYMK feature." However, from their finding that far more triangles were created than edges (based on their fig. 2b, the ratio of new triangles to new edges rose from about 2 before the introduction to about 4 afterwards), it suggests that the created edges were based heavily on triadic closure. They conclude that the external introduction of PYMK manipulated a parameter or parameters of the underlying dynamic network formation process, and furthermore, it did not increase the link creation or triadic closure uniformly, but with bias towards more delayed links and triads. While they say they were able to quantify the effects and impact of the PYMK feature, this did not include estimating the local average treatment effect, which is my specific interest.

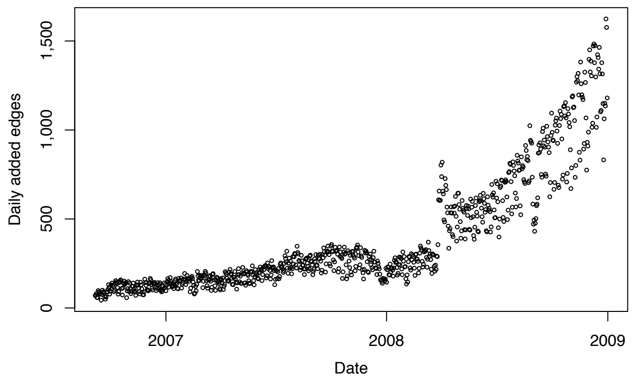

FIGURE 2.1: Observed edges added (friendship ties made) in Facebook New Orleans data.

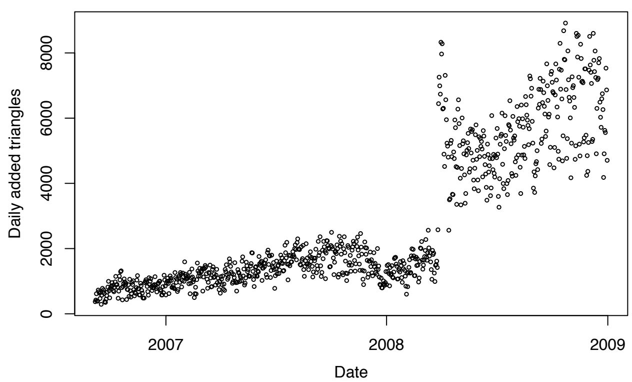

FIGURE 2.2: Triangles created with the added edges in Facebook New Orleans data.

The goal of the Netflix Prize competition was prediction and not explanation (Shmueli, 2010; Breiman, 2001), for which it is not necessary to understand the spike (only to account for it in a model, in order to effectively use past data for training). However, checking for data artifacts is fundamental for any type of data model, and Koren (2009a) devotes some time to investigating an odd spike observed in average ratings in early 2004, about 1500 days into the data set (this plot is recreated in my fig. 2.3). He proposes and explores three hypotheses:

1. Ongoing improvements in Netflix's 'Cinematch' recommendation technology and/or in the GUI led people to watch movies they liked more;
2. A change in the wordings associated with numerical ratings elicited different ratings (e.g., perhaps a rating of 5 was originally explained as "superb movie" and then was changed to "loved it");
3. There was an influx of new users who on average gave higher ratings.

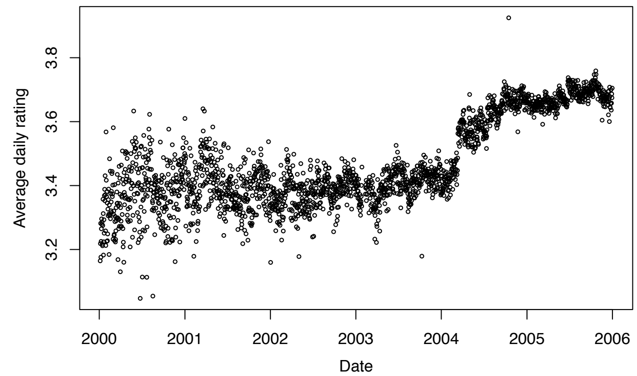

FIGURE 2.3: Observed daily averages for the Netflix Prize data.

By noting that the shift also occurs among users who were present both before and after the observed increase, he rejects the third possibility. He finds some support for the first possibility from a model that decomposes ratings into a baseline effect and a user-movie interaction effect (which corresponds to the extent to which users rate movies "suitable for their own tastes"); the interaction effect shows a smooth increase and the baseline has less variability, but there is is still clearly a sudden jump in the baseline. He writes, "This hints that beyond a constant improvement in matching people to movies they like, something else happened in early 2004 causing an overall shift in rating scale." Note that the change in wordings associated with numerical ratings is Koren's (2009a) guess to what the change was; he specifies that uncovering exactly what the "something else" was "may require extra information on the related circumstances." That such a change in wording *could* produce a shift in ratings is supported by decades of research in survey research into response options (Dillman et al., 2014), but otherwise no further evidence is given.

2.2.1 Causal modeling

Other works have been seeking out cases of natural experiments in social media data. Oktay et al. (2010) appears to be the first, discussing quasi-experimental designs and using Stack Overflow as an example setting. They demonstrate the use of interrupted time series by looking at whether users receiving an 'epic' badge, which is determined by hitting a daily reputation cap 50 times, decreases their daily posts; while only having 54 such examples to consider, it is enough to determine that getting this badge reduces the number of posts. In a case of applying regression discontinuity design, Li (2013) identified Yelp ratings being rounded to the nearest star as appropriate for RD design.

Sharma et al. (2015) provide a systematic statement of the problem of recommender systems: "little is known about how much activity [recommender] systems actually *cause* over and above activity that would have occurred via other means (e.g., search) if recommendations were absent. Although the ideal way to estimate the causal impact of recommendations is via randomized experiments, such experiments are costly and may inconvenience users." Surprisingly, much of the large body of work on recommender systems has not necessarily prioritized this causal aspect, with Sharma et al. (2015) identifying only three papers using experimental designs and only seven using observational data to study the causal effect of recommendations.

They use access to browsing history from opt-in users of the Bing toolbar in Internet Explorers to see whether users going to Amazon pages, whose URLs contain flags for the origin of a click, did so through Amazon's search engine, another product page, other Amazon pages (e.g., wishlists), or from an external website. This by itself is not sufficient, they explain because the causal question is a counterfactual one: users might have still found products through other channels had recommendations not existed (in which case the recommendation merely provide convenient access to item pages). They used browsing data to reconstruct the overall Amazon recommendation graphs, and then used the presence a shock to only one of a pair of co-recommended objects as an instrumental variable. They found 4,774 shocks to 4,126 objects from September 1, 2013 to May 31, 2014. Their final estimate of the causal impact of recommendations is 3%, in line with the experimental findings they reference.

Contemporaneously with the submission of the article version of this chapter (Malik and Pfeffer, 2016), and presented one month prior, is Su et al. (2016), who looked at the effect of recommendations on network structure via the introduction of Twitter's "Who To Follow" feature in 2010. While the discontinuity design with the introduction of the recommendation system is nearly identical to this work, the expected results are different, because Facebook's intended usage (with connections requiring reciprocation to exist) is different than that of Twitter (whose network is built on asymmetric follower relationships). We would expect Facebook to base recommendation systems on processes like (undirected) triadic closure, whereas we would expect Twitter to use cumulative advantage, and possibly *transitive closure* (Ripley et al., 2017), which is if i sends a tie to j and j sends a tie to k, then i will send a tie to k (e.g., k is higher in a hierarchy than j). Note that this is distinct from *two-out-star closure*, which is which i sends a tie to k, and j sends a tie to k, then i will send a tie to j (i.e., i connects with j because they are similar in their following of k).

Similar to some of my results, Su et al. (2016) find an initial increase in daily numbers of new edges, although they find a decrease afterwards (whereas I find no decrease in edges, but a decrease in the number of triangles created with each added edge). And, they find the fraction of new edges that are reciprocated decrease after the recommendation system's introduction. Overall, they find popular users benefitting disproportionately after the introduction of Who To Follow. They also find an increase in triadic closure, but by looking only at undirected triangles; it would be worthwhile to look at directed triangles, to see if the evidence is for a transitivity closure effect of a two-out-star closure effect which could shed light onto the likely mechanisms of the recommender system and the effect on the flows that are possible within the network. Also interesting is their discussion, where they conclude that a "mismatch between the recommender and the natural network dynamics thus alters the structural evolution of the network." Considering the implications of this, they note that cumulative advantage is often an undesirable property in terms of homogenization of ideas, although they also note that alternatively, there may be latent preferences towards following popular users such that the recommendation system only optimized natural dynamics, not caused unnatural dynamics. While their theoretical consideration of counterfactuals in terms of 'natural dynamics' is interesting, I would note that these systems do not necessarily have analogs to other networks (is it 'natural' to form networks based on communications limited to 140 or 280 characters? or to join a platform out of an interest in following celebrities, as Hargittai and Litt, 2011, find?) such that it may not be meaningful to talk about what is 'natural': we can only talk about comparison to dynamics within another regime of platform design, regulation, and engineered affordances. But the overall lesson is that different recommendation systems, designed to fulfill the goals and purposes of different platforms (respectively for Facebook and Twitter, to maximize connections between pre-existing acquaintances and to encourage more information consumption), will have different effect on platform networks. By identifying goals, we can anticipate what kinds of mechanisms recommendation systems might be designed with, and both the expected, direct effects (more following on Twitter, more friends on Facebook) and potentially indirect, unanticipated, or undesirable effects (rich-get-richer on Twitter, high local clustering on Facebook).

One other recent work is that of Cottica et al. (2017), who consider that online communities are indeed 'managed', and look at how to find the effects of such management. For example, they consider whether onboarding policies have an effect on degree distributions.

2.2.2 Social media networks

This work also relates to the theorization of social media networks. The most comprehensive theorizing is Kane et al. (2014), to which I connect my current investigation. They review Borgatti and Cross's (2003) grouping of social network research into four canonical types: how the network environment exerts influence on members, how resources spread through networks, how network structures benefit and/or constrain individuals, and how nodes use a network to access and benefit from resources. They then note,

> "In a social media context, network content is the digital content contributed by users, which may provide information, influence, or social support... Digital content flows through networks differently than other types of content; a physical object moving through a network occupies only one place at a time, whereas digital resources can be copied, manipulated, aggregated, and searched. While digital content is consistent with [social network analysis], its distinctive characteristics might mean that research on social media networks needs a specialized subset of measures and theories, with adaptations from traditional social network research.
>
> "Social media platforms quantify or formalize relationships or interactions between nodes by explicitly representing them in a formal data structure, operating on a computerized platform. This formalization provides relational capabilities in social media networks that are not present in offline social networks, including the ease of visualizing and analyzing the connections. However, the relational formalism of social media platforms also limits relational capabilities, such as by limiting the amount of nuance people can attribute to labels such as 'friend' or 'follower' (Gilbert and Karahalios, 2009). If people are limited to establishing similar formal connections with diverse sets of others including trusted confidants, casual acquaintances, and family members in their social networks, the platform homogenizes all of these relational connections as being equivalent (e.g., friends, contacts). Thus, while traditional SNA knows what ties mean but has difficulty eliciting these social data and measuring them objectively, social media can objectively measure ties through their digital traces but has trouble articulating the nuanced meanings of ties in a social context."

They note, however, that

> "Traditional SNA proceeds from a natural science paradigm, observing and describing the fundamental components of social networks in ways that best reflect how these networks are observed in the offline world. Social media, however, introduces questions of *design science*, how to implement the fundamental components of the network (i.e., nodes and ties) to achieve particular types of network behaviors (Ren et al., 2007)... In social media networks, tie features are not exclusively a reflection of the underlying social relationships that occur in the network but instead determine in part the nature and characteristics of the relationships that will occur on the platform. These design decisions regarding how relational ties are implemented will enable and constrain users' interactions."

That is, they bring up the causal aspect that platform design has on user behavior:

> "On the one hand, the features of an information system enable and constrain its users in particular ways, resulting in similar behavior among users of the same system. For social media, these features may be technical (e.g., capabilities provided by the platform), normative (e.g., policies and rules of the platform), or economic (e.g., incentives for certain types of use behaviors)."

This is not to say that users have no agency; Kane et al., 2014 note that "users may employ systems in ways that were unintended or unanticipated by designers (Boudreau and Robey, 2005)." But in the case of PYMK, it seems as though the platform succeeded in manipulating user behavior. In itself, this is not negative, as it manipulated it towards potentially desirable ends. Increasing network connectivity potentially increases access to resources:

> "The ability of users to *articulate their relational connections* and *view and navigate those connections* involves a capacity to visualize and manipulate the *network structure*—that is, how people establish and manage the connections between others in a network. Similarly, the ability to *establish a digital profile* and *access and protect content* contributed through the platform primarily involves *network content*, or how digital resources are shared and accessed through a network."

Based on this, they suggest adapted versions of the a 2×2 set of canonical types of social network research for social media, crossing structure and content with homogeneity and heterogeneity. This gives the research topics:

- structural homogeneity induced by the platform, e.g., how different types of ties (e.g., friendship ties, messages, tags) affect behavior and network formation
- content homogeneity induced by the platform, e.g., how available profile features affect behavior
- performance variation in structure from user behavior, e.g., how users or third parties utilize network structure to develop structural capital
- performance variation in content from user behavior, e.g., how people use content access mechanisms to access different resources.

This chapter fits into the first category: looking at how the platform features induce structural homogeneity, in my case how user behavior becomes homogenized in one particular direction by platform design.

Lastly and most important is Healy (2015), a preprint of which was published as early as 2012.[1] Healy discusses the 'performativity thesis', "the claim that parts of contemporary economics and finance, when carried out into the world by professionals and popularizers, reformat and reorganize the phenomena they purport to describe, in ways that bring the world into line with theory". He extends this to argue "that social network analysis is performative in the same sense as the cases studied in this literature."

The performativity thesis has two versions, the more interesting but empirically more difficult to demonstrate strong version, where models *create* or *determine* the reality they purport to describe ("the performative process brings the empirical phenomena into line with the original model…the model helps make itself true, in the sense that before its public appearance the system did not behave in accordance with the model's predictions, whereas subsequently it does"), and a more circumspect and empirically clear weak version, where models merely shape things. With this background, Healy asks,

> "Is there a parallel in the network case? In the previous section we saw a range of web services that put calculative devices in the hands of users in interesting ways. These devices act as 'cognitive prostheses,' in Callon's phrase—they allow users to do things they were unable to do before, such as easily see three or four degrees out of their social network, or discover which of thousands of strangers is most similar to them in their taste in books or music, or quickly locate people with similar financial goals, and so on. It is a relatively short step from here to taking advantage of these tools in ways that bear on actors' conformity to some aspect of network theory. To take a simple but significant example, Facebook uses its data on the structure of social relations to routinely suggest lists of 'people you may know' to users, with the goal of encouraging users to add those people to their network. In this way, the application works automatically to encourage the closure of forbidden triads in people's social networks—something which, in theory, should be the case anyway. This is likely also to increase the degree of measurable homophily in the network. Were a complacent analyst subsequently to acquire some Facebook data and run some standard tests on the network's structure, they would find—to their satisfaction—some confirmatory results about the structure of 'people's social networks.' Moreover, they would be able to claim that these results were plausible partly because of the scale of the data used for the analysis."

In other words, a paper from critical sociology independently proposed the exact same example as I use here, Facebook's People You May Know feature, as a window into how applications of models have a causal effect on human systems. Beyond this shared example, if van Dijck (2013) and Gehl (2014) provide the critiques that I operationalize around social media, on the side of networks, Healy's (2015) critique is effectively what I operationalize. Of course, the idea that the way we discuss the world (and how we act based on framings) can influence the world is a general constructivist insight (see also Hacking's 'dynamic nominalism'; Hacking, 2007), the core idea from which we both drew.

Subsequent work has also followed up on the sociological angle of manipulation, power, and control. Bucher (2017) takes up the critical angle of how Facebook users feel about being managed by Facebook's features, where they are even aware of this. From an STS perspective, Yeung (2017) further theorizes about 'regulation by design', which is how recommender systems are a form of regulatory governance.

[1]I thank Abigail Jacobs for bringing this paper to my attention.

2.3 Data and methods

2.3.1 Facebook New Orleans

Viswanath et al. (2009) detail how they collected the Facebook New Orleans data through a manual crawl of the New Orleans network, starting from a single user and using breadth-first search. Considering that Facebook started as a college-based network, the boundary specification (Laumann, 1973) of users who added themselves to the "New Orleans" network primarily (or those who chose to add it secondarily, perhaps after a college network) may not meaningfully match the college-centric boundaries within which links actually formed (especially since, as the authors point out, regional networks have more lax security than university networks, which require a valid email address from the university's domain). Second, only visible profiles could be accessed: the authors estimate, by comparison with statistics from Facebook, that they collected 52% of the users in the New Orleans network.

The Facebook data come in the form of timestamps of added edges between 63,731 unique nodes. About 41.41% of edges do not have a timestamp. On the data download page, Viswanath et al. (2009) write that "the third column is a UNIX timestamp with the time of link establishment (if it could be determined, otherwise it is [blank])" without elaborating on the reasons for missing labels; I make the assumption that these were the edges already present at the start of data collection. However, I find a great deal of repeated edges. Of the 1,545,686 rows of data, there are only 817,090 unique edges (i.e., 52.86% row are unique, 47.14% are redundant). Breaking it down, of the 640,122 rows that have no timestamp, only 481,368 represent unique edges, and of the 905,564 rows that have a timestamp, only 614,796 represent unique edges. 88,494 edges are repeated twice, 728,596 edges are repeated three times, and no edge is repeated more than three times. I make the decision to drop these repeated edges, assuming that repetition was the result of a repeat visit from multiple crawls (and assuming that timestamps were gathered by the time of detection via BFS, rather than extracted from profiles).

To the unlabeled edges I assign the minimum time present among the remaining edges, and for repeated edges I take their first instance only. Using the igraph library (Csárdi and Nepusz, 2006) I take the initial graph and calculate the number of edges, the number of nodes (i.e., non-isolates), the number of triangles, and the transitivity. Since the inter-arrival times are not particularly relevant for my question, I care only about the change in the relative rate over time, I aggregate my analyses by day to create time series: for each day, I add the edges that appeared on that day and recalculate the graph metrics. After, I also calculate the daily density using $2M/(N^2 - N)$ for the number of nodes N and number of edges M. I then difference each of these series, and for each day get the number of edges added, the number of nodes added, the number of new triangles, the change in transitivity, and the change in graph density. (Note that daily aggregation followed by differencing is equivalent to a histogram with day-wide bins, as Zignani et al. (2014) do for the number of triangles and edges.)

2.3.2 Netflix Prize

The Netflix data come in the form of text files for individual movies, with each line being the rating that a given user gave along with the date from 1999-11-11 to 2005-12-31. Following Koren's (2009a) plot, I take the daily average in order to see the sudden jump. Examining the number of ratings (i.e., the number of binned observations) per day, I find that they increase linearly in log scale. However, until 1999-12-31, ratings are not daily and even when present are small, whereas from 2000-01-05 (the next day for which there is data) there are daily ratings in the thousands. I take only the data on and after 2000-01-05.

My own investigation pinpointed the discontinuity as occurring on or around March 12, 2004. I could not find any public record of a platform change at that time nor any clues in press releases around then, and Netflix did not respond to a request for further information.

Statistically, the Netflix data are more straightforward as there is no social network.[2] However, the independence assumptions are more complicated; with a single dynamic network as in the Facebook New Orleans data, I can assume that the network-level rate metrics like the number of added triangles are independent observations across days. If we only consider the average daily rating, we do not take into account multiple ratings by the same individual (and, as Koren (2009a) notes, it is important to correct for different baseline average ratings across users, e.g. making sure an overall 'stingy' user's ratings are comparable to those of an overall 'generous' user). But my interest is not in a full model of user ratings (predictive or explanatory), only a model of the average change to user behavior from a suspected platform effect. That is, we are interested in the marginal effect for which such dependencies are not relevant, and for which we can invoke the random sampling on ratings as a guarantee that my estimate will not have biases in representation.

2.3.3 Causal estimation with discontinuities

Regression discontinuity (RD) design is used to estimate causal effects in cases where there is an arbitrary (and preferably strict) cutoff along one covariate. As shown in Hahn et al. (2001), when the appropriate conditions are met, the treatment is effectively random in the left and right neighborhoods of the cutoff c. Causal effects are defined in terms of counterfactuals Y_{0i} (the value of the response were observation i to not be treated) and Y_{1i} (the value of the response were i to be treated); the point difference between the two at the time of intervention for treated populations is called the *local average treatment effect* (Imbens and Angrist, 1994), α. Given an observed Y_i, this is given by

$$\alpha \equiv E(Y_{1i} - Y_{0i}|X_i = c) = \lim_{x \downarrow c} E(Y_i|X_i = x) - \lim_{x \uparrow c} E(Y_i|X_i = c) \tag{2.1}$$

In the linear univariate case, the model is

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 \mathbf{1}(x_i > c) + \beta_3 x_i \mathbf{1}(x_i > c) + \varepsilon_i \tag{2.2}$$

which effectively fits two separate lines, one for each 'population' before and after the cutoff, with the estimated $\hat{\alpha}$ being the difference between the two fitted lines at the cutoff. The interest is generally in estimating the causal impact, but as a specification test (Imbens and Lemieux, 2008), the joint test for $H_0 : \beta_2 = \beta_3 = 0$ corresponds to a null hypothesis that there is no discontinuity. This model and the corresponding test may be generalized with higher-order polynomial terms. The model also has a natural

[2] Netflix did briefly attempt to add social networking features in late 2004. However these were discontinued in 2010, with part of the justification being that fewer than 2% of subscribers used the service.

nonparametric extension: separately fit the same smoother on either side of the discontinuity to estimate the effect, or, test for the discontinuity by seeing if confidence intervals overlap.

Note that the exemplars of RD design are not temporal, and many standard parts of time series modeling are incompatible with RD design. For example, a discontinuity is necessarily nonstationary, and differencing will destroy it (I fitted ARIMA models, and found that differencing was indeed necessary), and similarly, a one-sided moving average smoother applied to both sides of the discontinuity will leave a gap. I found two alternative methodologies created specifically around time series, 'interrupted time series analysis' (McDowall et al., 1980; Wagner et al., 2002; Taljaard et al., 2014) and 'event studies' (MacKinlay, 1997), but both are essentially less formal versions of RD design and still neither account for temporal features (namely, autocorrelation). I also tried Gaussian Process (GP) regression (Rasmussen and Williams, 2005; MacDonald et al., 2015), as it is able to capture temporal dependencies (Roberts et al., 2012). A squared exponential covariance function gave largely similar results, including posterior intervals about as wide as confidence intervals from other methods (and thus perhaps still not capturing autocorrelation) when fitting separately to either side of the discontinuity. I note that it may be possible in future work to adapt covariance functions that account for 'changepoints' (Garnett et al., 2010) not just to make predictions in the presence of discontinuities, but to do causal inference within the RD framework.

As we are interested in the central tendency rather than on features of the time series, I prioritize the use of the RD framework over time series modeling. To apply RD design, I make the assumption that the respective times at which People You May Know and whatever change took place in Netflix were introduced were effectively random. I use time as the covariate, with the respective cutoffs for the two data sets of 2008-03-26 and 2004-03-12 (i.e., I code for the potential discontinuities starting on those days). I apply nonparametric models, and specifically, local linear regression as is standard in regression discontinuity design (Imbens and Lemieux, 2008) and is also appropriate for time series (Shumway and Stoffer, 2011).

While a nonparametric smoother has the advantages of being able to fit cyclic behavior without including specific cyclic terms, confidence intervals still fail to capture the extent of cyclic variance and so are too optimistic even beyond not accounting for temporal autocorrelation (Hyndman et al., 2002). Prediction intervals are an alternative as they include the overall variance, but are not straightforward to calculate for smoothers. Another alternative, which we use for the Netflix data and for edge counts in the Facebook data, is to use local linear quantile regression (Koenker, 2005) to get tolerance (empirical coverage) intervals, and specifically, using the interval between a fit to 5% and to 95% to get a 90% tolerance interval (I found too much noise for fits at 97.5% and 2.5% to use a 95% tolerance interval). This is analogous to an idea in Taylor and Bunn (1999), who produce forecast errors for an exponential smoother using quantile regression.

For consistency, when I do this I also use quantile regression for the central tendency (i.e., using the median instead of the mean), which is also known as or "robust regression" and has the advantage of being more robust to outliers.

2.4 Results and discussion

2.4.1 Netflix Prize data

First, I note that the number of daily ratings increases over time (fig. 2.4), which corresponds to decreasing variance in the time series plot, suggesting use of weighted least squares. Weighting by the number of daily ratings (so that the days with more ratings are counted more heavily) improved diagnostics across the parametric models I considered; however, I found that the addition of polynomial terms up to and even past 7th order continued to be significant, leading me to prefer the nonparametric approach that can capture the cycles without becoming cumbersome. In fig. (2.5), I show the results of the local linear quantile regression. As we can see, at the cutoff the two 90% tolerance intervals do not overlap, allowing us to reject the null hypothesis that there is no discontinuity at the 0.10 level.

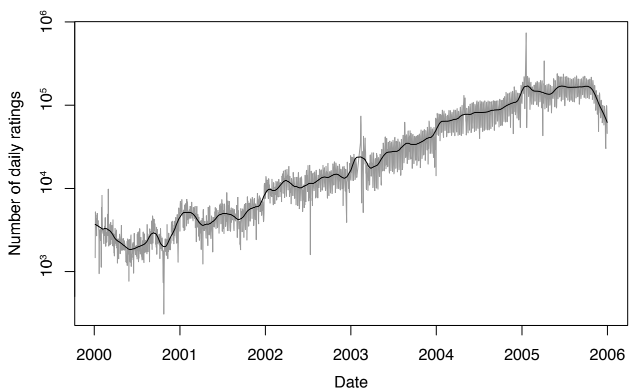

FIGURE 2.4: The number of Netflix ratings increases over time (y-axis shown in log scale); and, we can observe from fig. (2.3), the variance decreases over time, suggesting using the counts as weights. The fitted local linear smoother, which I used for weights, is shown in black. The bandwidth of .026 was selected via 5-fold cross validation.

To test if the model detects jumps at non-discontinuity points, I tried each day as a cutoff. Other than the actual discontinuity, the only points where the tolerance intervals did not overlap were two points before the cutoff I used (March 10th and 11th) and one day after (March 13th). Since I had initially located this date through manual (graphical) investigation, and the choice was not unambiguous within several days, it is unsurprising that the model picks this up as well. While this ambiguity is likely a matter of noise, platform engineers commonly deploy new features gradually to decrease risk, so it is also possible that the ambiguity is a gradual rollout that the model is also detecting.

Sensitivity to the smoothing bandwidth (the tuning parameter which controls the size of the neighborhood used in local fitting) is a concern for estimating the causal effect, so as is recommended, I report the estimates across multiple bandwidths. From 5-fold cross-validation, the optimal bandwidth of 6 (i.e., using kernel $K(x^*, x_i) = \exp\{-.5((x^* - x_i)/6)^2\}$), performed poorly under specification testing, identifying many discontinuities. Larger bandwidths (where the estimator tends towards linear) performed better, but at large bandwidths, again many discontinuities were identified. This is not ideal but unsurprising given the loss function used in quantile regression; quantiles are less swayed by extreme values, such that the non-overlap of tolerance intervals properly capture that there is a discontinuity even far from the actual discontinuity. The estimate of the causal effect may still be good, but with the failure of the specification testing at both low and high bandwidths, I report only within the range that performed well.

I estimate the local average treatment effect, the average amount by which the platform change resulted in a change in user ratings, as 0.118 from a bandwidth of 25 (pictured in fig. 2.5), 0.126 from a bandwidth of 50, 0.124 for a bandwidth of 75, and 0.119 for a bandwidth of 100. Considering the ratings prior to the cutoff had a mean of around 3.44, these amounts are a substantial increase, and are about 3% of the total possible range of ratings (from 1 to 5). This is a less involved case than Facebook, since movie preferences are a relatively low stakes phenomenon, but it shows the application of regression discontinuity. If the cause of the discontinuity is indeed a change in wordings, it shows that, just as in survey research, a change to the format changes the distribution of answers; but unlike in surveys, with large-scale online (streaming) systems, changes become visible as discontinuities in time.

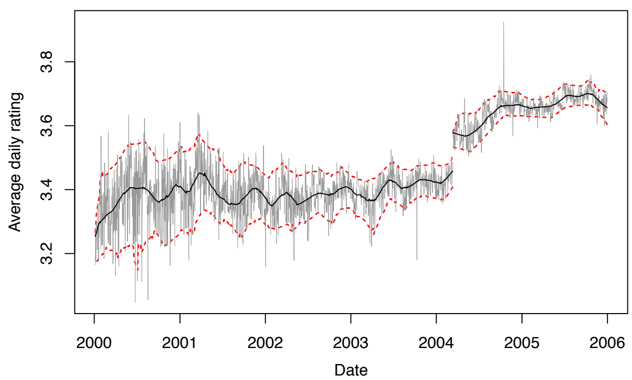

FIGURE 2.5: The solid line shows the local linear fit for the median Netflix ratings. The dashed lines give a fitted 90% tolerance interval, from local linear quantile fits to 5% and 95%. The intervals on both sides of cutoff do not overlap.

2.4.2 Facebook New Orleans data

Fig. (2.6) shows the discontinuity in the Facebook New Orleans data across four graph metrics. In addition to the daily counts of the number of added edges and added triangles as examined by Zignani et al. (2014), the discontinuity is pronounced in the transitivity and the density as well (although the units of these are so small as to not be particularly interpretable, so I do not estimate a local average treatment effect).

For the number of edges, I first used a fifth-order polynomial Poisson regression (not pictured), which had excellent regression diagnostics, from which I estimated a local average treatment effect of 356. This is more than a doubling of the pre-cutoff daily average of 314. However, the confidence intervals from the Poisson regression were very narrow and performed poorly under specification testing (as did bootstrap prediction intervals, which were very wide), in addition again to the problem of relying on higher-order polynomial terms rather than just relying on a nonparametric approach, so I also made fitted tolerance

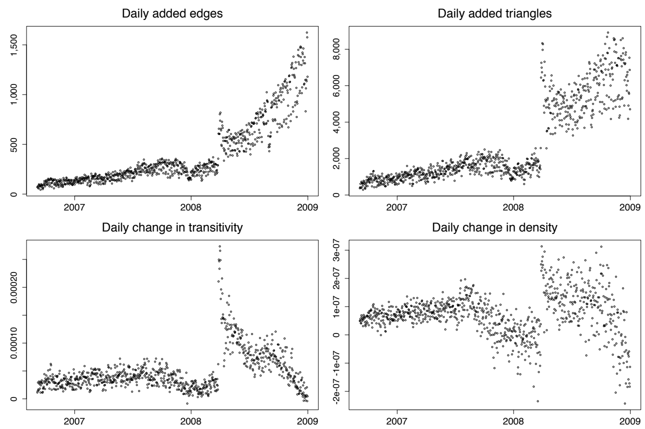

FIGURE 2.6: For the Facebook New Orleans data, the daily added edges and triangles created (top left and right, respectively), and the daily change in transitivity and graph density (bottom left and right, respectively).

intervals using local linear quantile regression as with the Netflix data, shown in fig. (2.7). Again, the optimal bandwidth found from 5-fold cross-validation was small and performed poorly under specification testing, as did large bandwidths (tending towards linear). Reporting within the range that performed well under testing, I estimate the local average treatment effect as 319 from a bandwidth of 25 (pictured), 278 for a bandwidth of 50, 228 for a bandwidth of 75, and 201 for a bandwidth of 100.

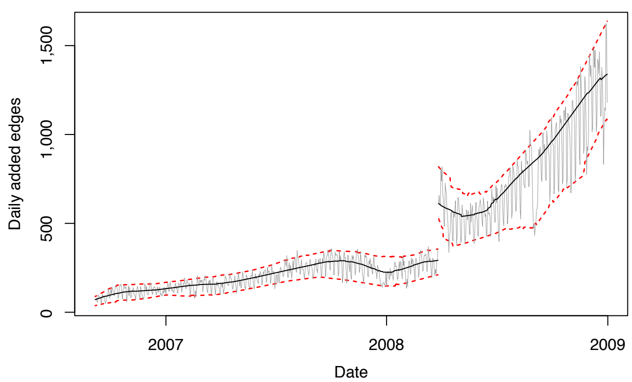

FIGURE 2.7: A local linear fit for the median number of edges added daily in Facebook New Orleans. The dashed lines give a fitted 90% tolerance interval, from local linear quantile fits to 5% and 95%. The intervals on both sides of cutoff do not overlap.

As the number of edges and triangles are closely related (fig. 2.8) and there are enough observations for a ratio to not be a noisy estimation target, I follow Zignani et al. (2014) in taking the ratio of triangles to edges. This represents the average number of triangles created by each added edge, and captures the extent of triadic closure on a scale more interpretable scale than that of changes in transitivity (which are in the ten thousandths). For a parametric model with an indicator for the discontinuity as described in eqn. (2.2), up to fourth-order polynomial terms were significant additions to the model in partial F tests, which is implausible and not parsimonious, so I again prefer a nonparametric fit, shown in fig. (2.9), which estimates a local average treatment effect of 3.86. This is even more dramatic than the effect in Netflix; given that the mean ratio was estimated at 6.25 before the jump, this is an increase of 61.8%.

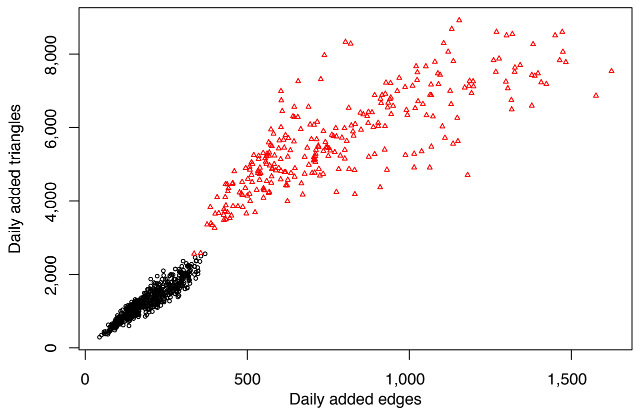

FIGURE 2.8: The daily added edges and triangles have a close relationship in the Facebook data. Black circles are time points before 2008-03-26, and red triangles are time points afterwards.

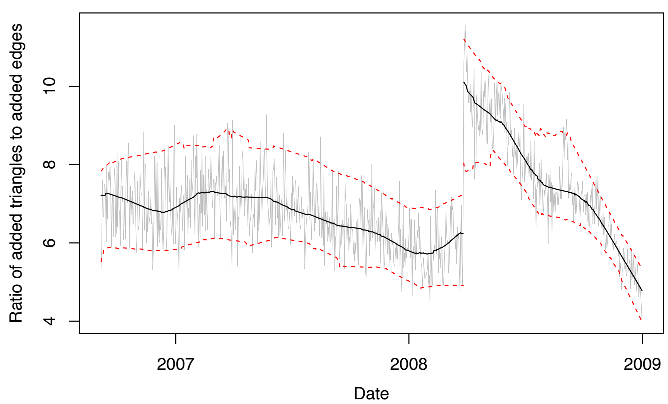

FIGURE 2.9: A local linear fit for the median daily ratio of added triangles to added edges in Facebook New Orleans. The dashed lines give a fitted 90% tolerance interval, from local linear quantile fits to 5% and 95%.

2.5 Conclusion

For much of data analysis, discontinuities (such as from abrupt platform changes in social media) are seen as incidental, or annoyances to be corrected (Roggero, 2012). Indeed, they appear in the literature as curiosities or asides. However, given the theoretical concerns about the nature of social media data, they can give valuable insights. My finding about the change in average Netflix ratings echoes work in survey research about response item wordings in a different setting and with different sort of data, quantifying how much we might expect a platform change to shift a baseline, and the sizes of 3% matches the size of causal estimates of a different process, that of recommendation systems, described in Sharma et al. (2015). For the Facebook New Orleans data, the finding is even more dramatic and widely applicable: we now have a sense that the introduction of a triadic closure-based recommender system can nearly double the rate of link creation. Furthermore, it changes the nature of the created links (focusing on closing triads), which has repercussions for the graph structure, seen for example in the changes in density. This is far above previous estimates of the effect of recommendation systems, which could either mean there are additional confounders, or that applying recommendation systems on a *network* rather than on an individual is something qualitatively different.

This also provides an empirical extension of a concern raised by Schoenebeck (2013) about how variation in technology adoption creates online social networks that differ systematically from the underlying social network: from my results, we see it is not just the process of joining social networking sites that creates observed network properties, but also the ways in which platforms design influences users. Multiple works have considered whether network metrics of large online social networks differ from those of previously studied social networks (Corten, 2012; Quercia et al., 2012; Ugander et al., 2011; Mislove, Marcon, et al., 2007); we can continue to theorize how differences result from platform effects, usage patterns, and demographic representation, rather than from online platforms being a superior way to measure social networks.

There are concerns about what social media ties even represent (Lewis et al., 2008), with some authors pointing to interactions over ties (Viswanath et al., 2009; Romero et al., 2011; Wilson et al., 2012; Jones et al., 2013) as more meaningful than the existence of ties. But my results show that the problem is not just one of ties not being a rich enough measure, but that they a non-naturalistic measure of social relationships, and furthermore, their existence determines visibility and access and thereby what activity happens. As people accept suggested links and begin interacting, the underlying phenomenon (the relationships and the network effects) changes, whether for good (Burke and Kraut, 2014) or ill (Kwan and Skoric, 2013). On Netflix, if changes affect different movies differently, it has consequences for modeling user behavior preferences. Beyond research concerns, there are economic benefits for the creators of movies that benefit from platform changes. Lotan (2015) observed this potentially happening in Apple's App Store, where what appeared to be an (unannounced, undocumented) engineering change in the search results ranking led to changes in app sales.

Regression discontinuity design has a rich literature, and there will likely be many future cases where we can apply RD design or interrupted time series in social media data. In geotags collected from the US in 2014, there was a sudden decrease (fig. 2.10) on September 18th, the same day Twitter released significant updates to profiles on Twitter from iPhone.[3] The recent increase in character limit on Twitter from 140 characters to 280 was, after being trialled with a small number of users, [4] rolled out en masse on November

[3] "A new profile experience on Twitter for iPhone", September 18, 2014, https://blog.twitter.com/2014/a-new-profile-experience-on-twitter-for-iphone, accessed 1/2016.

[4] "Giving you more characters to express yourself", 26 September 2016, https://blog.twitter.com/official/en_us/topics/product/2017/Giving-you-more-characters-to-express-yourself.html, accessed 8/2018.

7, 2017.[5] Indeed, one of the justifications for the change was that the artificiality of the 140 character limit was clear in how 9% of tweets came up against this limit, and the distribution of tweet character length being bimodal; changing the character limit to 280, the announcement on Twitter blog noted, reduced the number of tweets hitting the limit to 1%, and in the distribution of tweet length the bimodality disappeared. But more importantly, people who came up against the 140 character limit were likely adopting conventions of abbreviations, extensively editing, or splitting thoughts into multiple tweets (as also noted in the announcement), and looking past just the distribution of characters per tweet to how certain tweeting conventions become less common would give an idea of the causal impact of platform constraints. It would also be interesting to look at whether only 1% of tweets hitting a 280 character limit is a persistent effect; as people get used to a longer limit, will more and more tweets start coming up against the new 280 character limit? There is also an effect of language; the announcement noted that there would be no change in the 140 character limit for Chinese, Japanese, or Korean, as far fewer tweets in these languages were hitting the 140 character mark. In an example that ties in with Chapter (1), Tasse et al. (2017) note a sharp dropoff in the number of geotagged tweets in May 2015, which they attributed to a change in the user interface that made place-tagging, rather than geotagged, the default. In their survey results, they found geotag tweet users who were unaware about the level of precision of geotags, as their intention was only to provide a general location and not a coordinate. The change in the default behavior brought geotagged tweets more in line with users' understandings of what the platform feature was actually doing. These show how there has begun to be a public body of knowledge about the ways in which platform design are responsible for observed behavior, a body of knowledge that outside researchers can continue to build on. Extensions to regression discontinuity are also relevant, for example in how Porter and Yu (2015) develop specification tests into tests for unknown discontinuities.

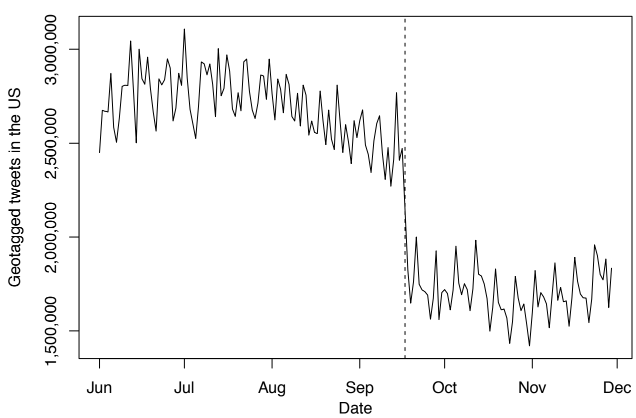

FIGURE 2.10: Another potential discontinuity, seen amidst cyclic behavior in the volume of geotagged tweets collected in the US in 2014.

Social media data have been compared to the microscope in potentially heralding a revolution in social science akin to that following the microscope in biology (Golder and Macy, 2012). This metaphor may have a deeper lesson in a way that its advocates did not expect: history of science has shown (Szekely, 2011)

[5]"Tweeting made easier", 7 November 2017, https://blog.twitter.com/official/en_us/topics/product/2017/tweetingmadeeasier.html, accessed 8/2018.

that it was not a simple process to connect the new instrument, with its multiple shortcomings, to the natural objects it was supposedly being used to study. It took centuries of researchers living with the microscope, improving the instrument but also understanding how to use it (e.g., recognizing the need for staining, or the importance of proper lighting), that microscopes became a firm part of rigorous, cumulative scientific research. I would hope that social media data will not take as long, but at the same time, it is as necessary as ever to question the relationship between the novel instrument and the object of study.

Bibliography

Angrist, Joshua D. and Jörn-Steffen Pischke (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton, NJ: Princeton University Press.

Borgatti, Stephen P. and Rob Cross (2003). "A relational view of information seeking and learning in social networks". *Management Science* 49 (4), pp. 432–445. DOI: 10.1287/mnsc.49.4.432.14428.

Boudreau, Marie-Claude and Daniel Robey (2005). "Enacting integrated information technology: A human agency perspective". *Organization Science* 16 (1), pp. 3–18. DOI: 10.1287/orsc.1040.0103.

boyd, danah m. and Kate Crawford (2012). "Critical questions for Big Data: Provocations for a cultural, technological, and scholarly phenomenon". *Information, Communication & Society* 15 (5), pp. 662–679. DOI: 10.1080/1369118X.2012.678878.

Breiman, Leo (2001). "Statistical modeling: The two cultures (with comments and a rejoinder by the author)". *Statistical Science* 16 (3), pp. 199–231. DOI: 10.1214/ss/1009213726.

Bucher, Taina (2017). "The algorithmic imaginary: Exploring the ordinary affects of Facebook algorithms". *Information, Communication & Society* 20 (1), pp. 30–44. DOI: 10.1080/1369118X.2016.1154086.

Burke, Moira and Robert E. Kraut (2014). "Growing closer on Facebook: Changes in tie strength through social network site use". *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems*. CHI '14, pp. 4187–4196. DOI: 10.1145/2556288.2557094.

Chang, Jonathan, Itamar Rosenn, Lars Backstrom, and Cameron Marlow (2010). "ePluribus: Ethnicity on social networks". *Proceedings of the Fourth International Conference on Weblogs and Social Media*. ICWSM-10, pp. 18–25. URL: https://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/view/1534.

Corten, Rense (2012). "Composition and structure of a large online social network in the Netherlands". *PLOS ONE* 7 (4), e34760. DOI: 10.1371/journal.pone.0034760.

Cottica, Alberto, Guy Melançon, and Benjamin Renoust (2017). "Online community management as social network design: Testing for the signature of management activities in online communities". *Applied Network Science* 2 (1). DOI: 10.1007/s41109-017-0049-9.

Csárdi, Gábor and Tamás Nepusz (2006). "The igraph software package for complex network research". *InterJournal* Complex Systems (1695). URL: http://igraph.org.

Diakopoulos, Nicholas (2014). "Algorithmic accountability reporting: On the investigation of black boxes". Tow Center for Digital Journalism, Columbia Journalism School. URL: http://towcenter.org/research/algorithmic-accountability-on-the-investigation-of-black-boxes-2/.

Dillman, Don A., Jolene D. Smyth, and Leah Melani Christian (2014). *Internet, phone, mail, and mixed-mode surveys: The tailored design method*. 4th ed. Hoboken, NJ: John Wiley & Sons, Inc.

Garnett, Roman, Michael A. Osborne, Steven Reece, Alex Rogers, and Stephen J. Roberts (2010). "Sequential Bayesian prediction in the presence of changepoints and faults". *The Computer Journal* 53 (9), pp. 1430–1446. DOI: 10.1093/comjnl/bxq003.

Gehl, Robert W. (2014). *Reverse engineering social media: Software, culture, and political economy in new media capitalism*. Philadelphia, PA: Temple University Press.

Gelman, Andrew (2009). "A statistician's perspective on *Mostly harmless econometrics: An empiricist's companion*, by Joshua D. Angrist and Jörn-Steffen Pischke". *Stata Journal* 9 (2), pp. 315–320. URL: http://www.stata-journal.com/article.html?article=gn0046.

Gilbert, Eric and Karrie Karahalios (2009). "Predicting tie strength with social media". *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '09, pp. 211–220. DOI: 10.1145/1518701.1518736.

Golder, Scott and Michael W. Macy (2012). "Social science with social media". *ASA footnotes* 40 (1). URL: http://www.asanet.org/footnotes/jan12/socialmedia_0112.html.

Hacking, Ian (2007). "Kinds of people: Moving targets". *Proceedings of the British Academy*. Vol. 151, pp. 285–318. DOI: 10.5871/bacad/9780197264249.003.0010.

Hahn, Jinyong, Petra Todd, and Wilbert Van der Klaauw (2001). "Identification and estimation of treatment effects with a regression-discontinuity design". *Econometrica* 69 (1), pp. 201–209. DOI: 10.1111/1468-0262.00183.

Hargittai, Eszter and Eden Litt (2011). "The tweet smell of celebrity success: Explaining variation in Twitter adoption among a diverse group of young adults". *New Media & Society* (13), pp. 824–842. DOI: 10.1177/1461444811405805.

Healy, Kieran (2015). "The performativity of networks". *European Journal of Sociology* 56 (2), pp. 175–205. DOI: 10.1017/S0003975615000107.

Hecht, Brent and Monica Stephens (2014). "A tale of cities: Urban biases in volunteered geographic information". *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media*. ICWSM-14, pp. 197–205. URL: http://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/view/8114.

Huberman, Bernardo A. (2012). "Sociology of science: Big data deserve a bigger audience". *Nature* 482 (7385), p. 308. DOI: 10.1038/482308d.

Hyndman, Rob J., Anne B. Koehler, Ralph D. Snyder, and Simone Grose (2002). "A state space framework for automatic forecasting using exponential smoothing methods". *International Journal of Forecasting* 18 (3), pp. 439–454. DOI: 10.1016/S0169-2070(01)00110-8.

Imbens, Guido W. and Joshua D. Angrist (1994). "Identification and estimation of local average treatment effects". *Econometrica* 62 (2), pp. 467–475. DOI: 10.2307/2951620.

Imbens, Guido W. and Thomas Lemieux (2008). "Regression discontinuity designs: A guide to practice". *Journal of Econometrics* 142 (2), pp. 615–635. DOI: 10.1016/j.jeconom.2007.05.001.

Jones, Jason J., Jaime E. Settle, Robert M. Bond, Christopher J. Fariss, Cameron Marlow, and James H. Fowler (2013). "Inferring tie strength from online directed behavior". *PLOS ONE* (8), e52168. DOI: 10.1371/journal.pone.0052168.

Kane, Gerald C., Maryam Alavi, Giuseppe (Joe) Labianca, and Stephen P. Borgatti (2014). "What's different about social media networks? A framework and research agenda". *MIS Quarterly* 38 (1), pp. 274–304.

Koenker, Roger (2005). *Quantile regression*. Econometric Society Monographs. Cambridge, UK: Cambridge University Press.

Koren, Yehuda (2009a). "Collaborative filtering with temporal dynamics". *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '09, pp. 447–456. DOI: 10.1145/1557019.1557072.

– (2009b). "The BellKor solution to the Netflix Grand Prize". URL: http://www.netflixprize.com/assets/GrandPrize2009_BPC_BellKor.pdf.

Kramer, Adam D. I., Jamie E. Guillory, and Jeffrey T. Hancock (2014). "Experimental evidence of massive-scale emotional contagion through social networks". *Proceedings of the National Academy of Sciences* 111 (24), pp. 8788–8790. DOI: 10.1073/pnas.1320040111.

Kwan, Grace Chi En and Marko M. Skoric (2013). "Facebook bullying: An extension of battles in school". *Computers in Human Behavior* 29 (1), pp. 16–25. DOI: 10.1016/j.chb.2012.07.014.

Laumann, Edward O. (1973). *Bonds of pluralism: the form and substance of urban social networks*. New York: John Wiley & Sons, Inc.

Lazer, David, Alex Pentland, Lada Adamic, Sinan Aral, Albert-László Barabási, Devon Brewer, Nicholas Christakis, Noshir Contractor, James Fowler, Myron Gutmann, Tony Jebara, Gary King, Michael Macy, Deb Roy, and Marshall Van Alstyne (2009). "Computational social science". *Science* 323 (5915), pp. 721–723. DOI: 10.1126/science.1167742.

Lazer, David and Jason Radford (2017). "Data ex machina: Introduction to Big Data". *Annual Review of Sociology* 43 (1), pp. 19–39. DOI: 10.1146/annurev-soc-060116-053457.

Lee, David S. and Thomas Lemieux (2010). "Regression discontinuity designs in economics". *Journal of Economic Literature* 48 (2), pp. 281–355. DOI: 10.1257/jel.48.2.281.

Lewis, Kevin, Jason Kaufman, Marco Gonzalez, Andreas Wimmer, and Nicholas Christakis (2008). "Tastes, ties, and time: A new social network dataset using Facebook.com". *Social Networks* 30 (4), pp. 330–342. DOI: 10.1016/j.socnet.2008.07.002.

Li, Xitong (2013). "How does online reputation affect social media endorsements and product sales? Evidence from regression discontinuity design". *The 24th Workshop on Information Systems Economics*. WISE 2013.

Longley, Paul A., Muhammad Adnan, and Guy Lansley (2015). "The geotemporal demographics of Twitter usage". *Environment and Planning A* 47 (2), pp. 465–484. DOI: 10.1068/a130122p. URL: http://www.envplan.com/abstract.cgi?id=a130122p.

Lotan, Gilad (2015). "Apple's App Charts: 2015 Data and Trends ... or how much harder it is to get into the top charts". *i ♥ data* (15 December 2015). URL: https://medium.com/i-data/apple-s-app-charts-2015-data-and-trends-abb95300df57.

MacDonald, Blake, Pritam Ranjan, and Hugh Chipman (2015). "GPfit: An R package for fitting a Gaussian process model to deterministic simulator outputs". *Journal of Statistical Software* 64 (12), pp. 1–23. DOI: 10.18637/jss.v064.i12.

MacKinlay, A. Craig (1997). "Event studies in economics and finance". *Journal of Economic Literature* 35, pp. 13–39.

Malik, Momin M. (2018). "Bias and beyond in digital trace data". PhD thesis. Pittsburgh, PA: Carnegie Mellon University. URL: http://reports-archive.adm.cs.cmu.edu/anon/isr2018/abstracts/18-105.html.

Malik, Momin M., Hemank Lamba, Constantine Nakos, and Jürgen Pfeffer (2015). "Population bias in geotagged tweets". *Papers from the 2015 ICWSM Workshop on Standards and Practices in Large-Scale

Social Media Research. ICWSM-15 SPSM, pp. 18–27. URL: http://www.aaai.org/ocs/index.php/ICWSM/ICWSM15/paper/view/10662.

Malik, Momin M. and Jürgen Pfeffer (2016). "Identifying platform effects in social media data". *Proceedings of the Tenth International AAAI Conference on Web and Social Media*. ICWSM-16, pp. 241–249.

McDowall, David, Richard McCleary, Errol E. Meidinger, and Richard A. Hay Jr. (1980). *Interrupted time series analysis*. Quantitative applications in the social sciences. Thousand Oaks, CA: SAGE. DOI: 10.4135/9781412984607.

Mislove, Alan, Sune Lehmann, Yong-Yeol Ahn, Jukka-Pekka Onnela, and J. Rosenquist (2011). "Understanding the demographics of Twitter users". *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*. ICWSM-11, pp. 554–557. URL: https://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2816.

Mislove, Alan, Massimiliano Marcon, Krishna P. Gummadi, Peter Druschel, and Bobby Bhattacharjee (2007). "Measurement and analysis of online social networks". *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement*. IMC '07, pp. 29–42. DOI: 10.1145/1298306.1298311.

Morstatter, Fred, Jürgen Pfeffer, and Huan Liu (2014). "When is it biased? Assessing the representativeness of Twitter's Streaming API". *Companion to the Proceedings of the 23rd International Conference on World Wide Web*. WWW Companion '14, pp. 555–556. DOI: 10.1145/2567948.2576952.

Morstatter, Fred, Jürgen Pfeffer, Huan Liu, and Kathleen Carley (2013). *Is the sample good enough? Ccomparing data from Twitter's streaming API with Twitter's firehose*. URL: http://aaai.org/ocs/index.php/ICWSM/ICWSM13/paper/view/6071/6379.

Oktay, Hüseyin, Brian J. Taylor, and David D. Jensen (2010). "Causal discovery in social media using quasi-experimental designs". *Proceedings of the First Workshop on Social Media Analytics*. SOMA '10, pp. 1–9. DOI: 10.1145/1964858.1964859.

Porter, Jack and Ping Yu (2015). "Regression discontinuity designs with unknown discontinuity points: Testing and estimation". *Journal of Econometrics* 189 (1), pp. 132–147. DOI: 10.1016/j.jeconom.2015.06.002.

Quercia, Daniele, Licia Capra, and Jon Crowcroft (2012). "The social world of Twitter: Topics, geography, and emotions". *Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media*. ICWSM-12, pp. 298–305. URL: http://www.aaai.org/ocs/index.php/ICWSM/ICWSM12/paper/view/4612.

Rasmussen, Carl Edward and Christopher K. I. Williams (2005). *Gaussian processes for machine learning*. Adaptive Computation and Machine Learning. Cambridge, MA: The MIT Press.

Ren, Yuqing, Robert E. Kraut, and Sara Kiesler (2007). "Applying common identity and bond theory to design of online communities". *Organization Studies* 28 (3), pp. 377–408. DOI: 10.1177/0170840607076007.

Ripley, Ruth M., Tom A. B. Snijders, Zsófia Boda, András Vörös, and Paulina Preciado (2017). "Manual for RSiena". University of Oxford: Department of Statistics; Nuffield College, and University of Groningen: Department of Sociology. URL: http://www.stats.ox.ac.uk/~snijders/siena/RSiena_Manual.pdf.

Roberts, S., M. Osborne, M. Ebden, S. Reece, N. Gibson, and S. Aigrain (2012). "Gaussian processes for time-series modelling". *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 371 (1984). DOI: 10.1098/rsta.2011.0550.

Roggero, M. (2012). "Discontinuity detection and removal from data time series". *VII Hotine-Marussi Symposium on Mathematical Geodesy*. Ed. by Nico Sneeuw, Pavel Novák, Mattia Crespi, and Fernando

Sansò. International Association of Geodesy Symposia. Berlin, Heidelberg: Springer, pp. 135–140. DOI: 10.1007/978-3-642-22078-4_20.

Romero, Daniel, Brendan Meeder, Vladimir Barash, and Jon Kleinberg (2011). "Maintaining ties on social media sites: The competing effects of balance, exchange, and betweenness". *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*. ICWSM-11, pp. 606–609. URL: https://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2860/3253.

Ruths, Derek and Jürgen Pfeffer (2014). "Social media for large studies of behavior". *Science* 346 (6213), pp. 1063–1064. DOI: 10.1126/science.346.6213.1063.

Savage, Mike and Roger Burrows (2007). "The coming crisis of empirical sociology". *Sociology* 41 (5), pp. 885–899. DOI: 10.1177/0038038507080443.

Schoenebeck, Grant (2013). "Potential networks, contagious communities, and understanding social network structure". *Proceedings of the 22nd International Conference on World Wide Web*. WWW '13, pp. 1123–1132. DOI: 10.1145/2488388.2488486.

Sharma, Amit, Jake M. Hofman, and Duncan J. Watts (2015). "Estimating the causal impact of recommendation systems from observational data". *Proceedings of the Sixteenth ACM Conference on Economics and Computation*. EC '15, pp. 453–470. DOI: 10.1145/2764468.2764488.

Shmueli, Galit (2010). "To explain or to predict?" *Statistical Science* 25 (3), pp. 289–310. DOI: 10.1214/10-STS330.

Shumway, Robert H. and David S. Stoffer (2011). *Time series analysis and its applications with R examples*. 3rd ed. Springer Texts in Statistics. New York: Springer. DOI: 10.1007/978-1-4419-7865-3.

Sloan, Luke, Jeffrey Morgan, William Housley, Matthew Williams, Adam Edwards, Pete Burnap, and Omer Rana (2013). "Knowing the tweeters: Deriving sociologically relevant demographics from Twitter". *Sociological Research Online* (18). DOI: 10.5153/sro.3001.

Su, Jessica, Aneesh Sharma, and Sharad Goel (2016). "The effect of recommendations on network structure". *Proceedings of the 25th International Conference on World Wide Web*. WWW '16, pp. 1157–1167. DOI: 10.1145/2872427.2883040.

Szekely, Francisc (2011). "Unreliable observers, flawed instruments, 'disciplined viewings': Handling specimens in early modern microscopy". *Parergon* 28 (1), pp. 155–176. DOI: 10.1353/pgn.2011.0032.

Taljaard, Monica, Joanne E. McKenzie, Craig R. Ramsay, and Jeremy M. Grimshaw (2014). "The use of segmented regression in analysing interrupted time series studies: An example in pre-hospital ambulance care". *Implementation Science* 9 (77). DOI: 10.1186/1748-5908-9-77.

Tasse, Dan, Zichen Liu, Alex Sciuto, and Jason I. Hong (2017). "State of the geotags: Motivations and recent changes". *Proceedings of the Eleventh International AAAI Conference on Web and Social Media*. ICWSM 2017, pp. 250–259. URL: https://aaai.org/ocs/index.php/ICWSM/ICWSM17/paper/view/15588.

Taylor, James W. and Derek W. Bunn (1999). "A quantile regression approach to generating prediction intervals". *Management Science* 45 (2), pp. 225–237. DOI: 10.1287/mnsc.45.2.225.

Tufekci, Zeynep (2014). "Big questions for social media big data: Representativeness, validity and other methodological pitfalls". *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media*. ICWSM-14, pp. 505–514. URL: https://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/view/8062.

Ugander, Johan, Brian Karrer, Lars Backstrom, and Cameron Marlow (2011). "The anatomy of the Facebook social graph". eprint: https://arxiv.org/abs/1111.4503.

van Dijck, José (2013). *The culture of connectivity: A critical history of social media*. New York, NY: Oxford University Press.

Viswanath, Bimal, Alan Mislove, Meeyoung Cha, and Krishna P. Gummadi (2009). "On the evolution of user interaction in Facebook". *Proceedings of the 2nd ACM SIGCOMM Workshop on Social Networks*. WOSN '09, pp. 37–42. DOI: 10.1145/1592665.1592675.

Wagner, A. K., S. B. Soumerai, F. Zhang, and D. Ross-Degnan (2002). "Segmented regression analysis of interrupted time series studies in medication use research". *Journal of Clinical Pharmacy and Therapeutics* 27 (4), pp. 299–309. DOI: 10.1046/j.1365-2710.2002.00430.x.

Wilson, Christo, Alessandra Sala, Krishna P. N. Puttaswamy, and Ben Y. Zhao (2012). "Beyond social graphs: User interactions in online social networks and their implications". *ACM Transactions on the Web* 6 (4), 17:1–17:31. DOI: 10.1145/2382616.2382620.

Yeung, Karen (2017). "'Hypernudge': Big Data as a mode of regulation by design". *Information, Communication & Society* 20 (1), pp. 118–136. DOI: 10.1080/1369118X.2016.1186713.

Zignani, Matteo, Sabrina Gaito, Gian Paolo Rossi, Xiaohan Zhao, Haitao Zheng, and Ben Zhao (2014). "Link and triadic closure delay: Temporal metrics for social network dynamics". *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media*. ICWSM-14, pp. 564–573. URL: http://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/view/8042.