

Social Media Giveth, Social Media Taketh Away: Facebook, Friendships, and APIs

BERNIE HOGAN¹
University of Oxford, UK

The early aughts saw an explosion of interest in social network sites. Many such sites, including Facebook, LinkedIn, and Twitter expanded to become “platforms,” meaning they are both websites and distributors of data. These data would typically be distributed to third parties in privacy-sensitive ways and regulated by the platform. Such access is based on balancing three issues: user privacy, generativity (i.e., capacity for novelty) for third parties, and control for platforms. Platforms appear to be seeking progressively more control at the cost of generativity by severely restricting third-party access to data about the user’s friends. The reductions or outright lack of access means that the insights from our digital traces are no longer as knowable to either third parties or users. This article unpacks this shift by clarifying some of the technical issues involved (particularly APIs, the main means of external data access). The case study of social network visualization is used to exemplify how social network sites seek control at the expense of generativity. The article notes how this shift was done with little oversight.

Keywords: social networks, Facebook, APIs, visualization, privacy

Social network sites (SNSs), by definition, enable users to indicate their relationships to others on the site (boyd & Ellison, 2007). These relationships permit access between users for messaging, sharing, and access to semipublic content such as posts. Social network sites use friendship connections to power key functionality for users, such as the organization of a newsfeed (Zuckerberg, Bosworth, Cox, Sanghvi, & Cahill, 2012). Third parties, such as games, productivity applications, and academics, can leverage friendship connections through regulated, programmatic access to a site’s “social graph,” or the recorded connections between people. For example, on Facebook, a game can learn of a user’s friends and show these to the user so friends can play together (“Game Requests,” 2017). While these relationships are of use to third parties and made by users, it is the platforms that assert sole discretion over who can access these data and how.² In recent years, platforms have restricted access to data about relationships on

Bernie Hogan: bernie.hogan@oii.ox.ac.uk

Date submitted: 2016–12–09

¹ The author wishes to thank the anonymous reviewers for their insightful feedback on the earlier drafts.

² While *platform* and *social network site* tend to be used interchangeably, the platform refers to the controller of all the data that are displayed, linked, or captured, whereas the site is the end-user application that is powered by these data. Facebook as a platform stores all the likes, personal

social network sites. This has had considerable consequences for both industry developers who would leverage a user's social contacts (for activities like job-finding or college advice; Jeon, Ellison, Hogan, & Greenhow 2016) and for researchers who would employ a user's social network in analysis (Brooks, Hogan, Ellison, Lampe, & Vitak, 2014).

Social network sites are premised on the notion of a social graph. The term *social graph* refers to the specific connections among persons and between people and digital entities, like fan pages. Facebook has friends, Twitter has followers, and LinkedIn has contacts. Since the introduction of SNSs, there has been considerable turbulence surrounding how the social graph should be regulated and access permitted. Some of the earliest SNSs, such as Friendster, allowed very generous access, leading to novel visualizations such as Vizster (Heer & boyd, 2005). These visualizations transform the social graph from a list of relationships between people into a visual map that shows how all the relationships are connected. In 2007, Facebook CEO Mark Zuckerberg stated that "third-party applications won't be treated like second-class citizens on Facebook" (Gannes, 2007, para. 5). This led Hogan (2010) as well as others to create network maps of the social graph on Facebook. In the ensuing decade, Facebook as well as other sites have sought increased control over how friends are represented to a user. In 2007, a third party on Facebook could access (but only with a user's permission) a user's basic profile data, a list of the user's friends, and selected data about these friends. In 2015, a standard third party could only access a user's friends if those friends also used the app. This changes data access from being able to represent a person's network to accessing the small fraction of users who add the same applications (a use case that makes sense for games, but not for an overview of a person's friendships). These changes took place primarily through the introduction or deprecation of specific "end points" where a third party could download user data. These end points and the rules for accessing them compose a Web-based application programming interface, or API.

APIs remain a key part of the research process for many academics well as for many commercial enterprises (Lomborg & Bechmann, 2014). Programmatic access through APIs has been highly regulated by the platforms. Twitter, for example, places very clear restrictions on the volume and velocity of Twitter data that a third party can access. For example, one can only access a user's 3,200 most recent tweets, and only in limited batches every 15 minutes ("Rate Limiting," 2017).

To date, there is still no consensus regarding what sort of app should have what sort of access to social graph data. Three competing claims circulate in relation to this issue:

1. Third parties can use the social graph to make innovative products, new representations of data, and generate novel insights that might otherwise be unavailable. This capacity is called "generativity" (Zittrain, 2006). For example, apps can show relationships between friends in novel ways (e.g., a time-line view, sorting photos, showing network maps). We can call the innovative potential related to data about social relationships *relational generativity*.

information, and so forth from a user, whereas Facebook as a social network site is the front-facing page available through <http://www.facebook.com> and the mobile app.

2. Platforms wish to keep users on their site rather than on a competitor's, to accumulate data for more effective advertisements/user monetization, and to test new strategies for reaching key performance indicators. This is platforms seeking control.
3. Users wish to share data with others to whom they are connected in meaningful ways, but not share all data with all other users or all possible third parties. Thus, users have an interest in privacy.

These three issues are in contention because it is seemingly impossible to optimize for all three. A totally private social network site would not allow for the sort of social sharing that is the common use case for such sites (Johns, 2016). Yet a site with no privacy features would suffer from privacy breaches, often referred to as *context collapse* (Marwick & boyd, 2010), as well as self-censorship (Vitak & Kim, 2014), and increased risk of identity fraud or abuse.

A totally generative social network site could lead to a third party poaching all of an individual's friendships or exploiting the server capacity of the platform. A site that is not generative at all would lead to the data on the site being totally managed by the platform, with little to no input from users and third parties who could add value. It would become a monopoly of social life that could introduce hidden biases (Pariser, 2011). A site with total control could change the rules continuously to its benefit, at the expense of users and third parties, with impunity. It could sell a user's data and make decisions that might be illegal, anticompetitive, or discriminatory.

This article claims that generativity with the social graph has been severely sacrificed by platforms in favor of control. This sacrifice has been accomplished primarily through technical means that make access to data by third parties onerous or impossible. This article will elaborate on this claim by introducing the notion of an API, which is the programmatic means by which a third party can legitimately access a platform's data. It will highlight the use of APIs in the generation of novel visualizations for social network sites and then review the shift in access toward less generativity. In this article, I claim that the arguments about this shift away from generativity are about consolidating control of a social graph and not about greater user privacy. I suggest that greater awareness of how such consolidation occurs as well as what is being lost can help with more effective decision-making. It can help regulators consider whether sites are acting in anticompetitive ways, undermining or biasing scientific research and inhibiting user interests.

The regulation of access to the social graph can be done through contractual and technical means. Contractual means include legal frameworks such as the European General Data Protection Regulation and site-specific terms of use. This article focuses less on the legal and contractual documents than on the technical means of access. Legal and contractual means involve enforcement, take a long time to contest, and can vary by jurisdiction. In contrast, changes in technical means are unambiguous, site-specific, and consequential. Platforms can completely change who can access data and how. Furthermore, where data submitted by a single user are now considered the property of that user, data that signify a relationship between users exist in a state of legal limbo. In the absence of clarity about this situation,

platforms have enacted technical means of regulating access to these social relationships in ways that align with the interests of the platform. Thus, the regulation of the social graph is a clear example of Lessig's (2000) famous notion that, in cyberspace, "code is law."

This discussion has taken on a newfound urgency in recent years. Access to the Facebook social graph is no longer meaningfully available, thereby inhibiting research such as the widely employed MyPersonality project (Kosinski, Stilwell, & Grapel, 2013). For example, we cannot programmatically learn how many friends a user has and whether those friends are connected to each other. We cannot show this information back to a user except as it is currently represented using Facebook's interface. As social media are increasingly embedded in everyday social life, social scientists need the means to understand and evaluate the influence of social media on said life—this includes the influence of social context, something intimately tied to a user's social network.

Codifying Relationships

Social relationships predate humanity, and humans are social creatures by nature. We can remember a large number of faces and names (Kanai, Bahrami, Roylance, & Rees, 2011) and have the cognitive capacity to navigate hundreds of personally known contacts (Dunbar, 1998). Social networks, as understood academically, are an abstraction from this complex web of interactions and situations (Hennig, Brandes, Pfeffer, & Mergel, 2012). Often, social network studies are quantitative, but, most importantly, they articulate specific connections between people as data.

Social network sites, like social network analysis, articulate specific connections between people as data. These connections are usually either "one way," such as a follower–followee relationship, or "symmetric," such as a "friend" relationship on Facebook or a "contact" relationship on LinkedIn (Hogan & Wellman, 2014). We might say that the social network is a posteriori for social network analysis but a priori for social network sites. Social network sites use data as a means to regulate and prioritize access between people. Facebook friends get status updates. LinkedIn contacts get to see more of the pages than what is available by surfing to that page.

When the network that is created focuses on one person and that person's relationships to other people, this is called a *personal network*. Relationships between that person and her network members are referred to as "direct ties," whereas relationships among these network members are referred to as "indirect ties" (Bian, 1997).

In a window between May 2006 and May 2015, Facebook allowed third parties (i.e., Facebook apps) to use the Facebook API to download a user's friends and friendships if an app was authorized by the user to do so. Facebook had initially employed regulatory means to ensure that privacy was protected—that is, in the statement of app rights and responsibilities, one used to be required to delete any data downloaded after 24 hours. This rule was in place until 2010, when it became evident that the 24-hour rule was not being adhered to by some companies. Those who did adhere to it also created bottlenecks by regularly requesting the same data for a user (Popken, 2010). This again highlights how

technical solutions to the regulation of access to friendships appear more absolute and more feasible than contractual means.

In April 2014, Facebook announced it was changing what data were accessed on the site. In a buried footnote suggesting that Facebook were eliminating several “rarely used endpoints” (Spehar, 2014), developers were to discover that Facebook was in fact removing meaningful access to a user’s newsfeed, their friendships, and data about friends (e.g., education, photos, and location). These end points were not rarely used. Whether we look to the million-plus personal networks uploaded to Wolfram Alpha (Wolfram, 2013), job sites such as Job Fusion, and the millions of users of apps that leveraged Facebook’s photo-sharing APIs, it is clear that something else was afoot.

These data were being used at that time in many highly popular applications, to the extent that technology journalism site Mashable Infographic suggested that Facebook platform data were used in seven of the top 10 apps on the Apple iOS app store as of 2012 (Buck, 2012). To this extent, it is clear that Facebook has closed off one of the key tools for helping to make the site a generative platform (Zittrain, 2006). As Rieder, Abdulla, Poell, Woltering, and Zack (2015) note, “What makes APIs important for empirical work is not just the way they jeopardize research, but also how they enable or suggest different directions and methods of analysis” (p. 19). To appreciate this situation in greater detail, I first explain how APIs work.

APIs: The Plugs That Connect Streams of Data

Programmatic access to Facebook, like Twitter, LinkedIn, Instagram, and most every social network site, is regulated by standardized technological practices. The most essential is the API, as it is the means by which data are sent between devices. Second to this is authentication that servers use to identify clients (i.e., third parties that work on behalf of users) and provide personalized access based on which client is requesting what data.

Defining APIs

Here we are using *API* to mean specifically Web-based APIs, as we are referring to the transfer of data over the Web. Other APIs can exist within an operating system, though these would be out of scope. To borrow from API-indexing website Programmable Web, an API is like an electrical socket (Berlind, 2015). Power out of a socket should be at the expected voltage and frequency. The plugs that go in are regularly shaped. It does not matter if one is plugging in a toaster or a television, the socket just sends power in the right format. Similarly, APIs send data in a regular format over the Web when requested. One can use any programming language to receive data from an API as long as it can make and receive Web requests, much like one can use any metal as long as it conducts and the plug is shaped correctly. Each “socket” in this metaphor is an end point. Taking the website Reddit as an example, one end point might be <https://api.reddit.com/.json>. By requesting that specific URL, we are receiving machine-formatted data (specifically in the .JSON format) from Reddit’s front page. The end point

<https://api.twitter.com/1.1/search/tweets.json?q=%40brexit> will perform a search of Twitter for the keyword *Brexit*.³

Virtually all Web apps require some sort of API, especially if they are to return personalized data. This means that one can potentially reverse-engineer an API even if the end points are not publicly available. For example, in early 2015 there was a minor scandal on Reddit and 4chan where pictures and videos from Snapchat were released on to the Web. Snapchat deletes photos after they have been sent to users via a private API. Some users were employing a third-party service that reverse-engineered this API to speak to Snapchat's databases as if it were the Snapchat client and save the snaps for later viewing. These snaps were leaked and distributed through BitTorrent in an event Reddit dubbed "the Snapping" (Stern, 2014).

Authenticating APIs

If API end points are agnostic to where the requests are coming from, then how can they control who sees what data? The answer is that APIs often expect that the request will be accompanied by additional means for authentication. An authenticated request enables the server to determine who is requesting data. This way, a server can regulate the volume, variety, and velocity of data. One common practice (OAuth) requires that a client submit specially generated tokens to signal that a request comes from a known third party and is done on behalf of a specific user. The exact details of this type of authentication are not as important as three consequences of this approach. First, a third party can act on behalf of a user with specific and configurable permissions. Second, a third party can do this without learning the user's password. Third, the third party's interactions with the platform can be monitored and regulated.

Authenticated APIs work as technological gatekeepers for data stored on platforms. If there are any restrictions to be placed on what data are available and to whom, it is commonly through this framework. Authenticated APIs can also be used to restrict the velocity of information—that is, most APIs will "rate limit" in some way. Twitter, for example, places restrictions on the number of times every 15 minutes one can make a specific call for data, such as a call to `/statuses/user_timeline`. This end point allows a third party to make 900 calls for Twitter statuses (i.e., "tweets") every 15 minutes. Each call can return up to 200 statuses up to a maximum of 3,200 per Twitter account. One could similarly make 15 calls to the end point `/friends/IDs` every 15 minutes. Each call returns 5,000 IDs. This means that one could capture up to 75,000 followers every 15 minutes.

Twitter used to allow 200 calls an hour, rather than 60, for standard accounts and 20,000 calls per hour for "whitelisted" accounts, specifically granted to academics. With the introduction of REST API 1.1, announced in 2012 and implemented in 2013 (Cunningham, 2012), however, the new restricted numbers appeared, and whitelists were gone. Twitter actually made other parts of their API more generous—namely, in terms of the number of tweets one could get for free from Twitter's "streaming

³ This URL will only work for applications that have provided the correct credentials; see "Rate limiting" (2017) for more details on how to properly authenticate an API request.

API.” This indicates that Twitter wanted third parties to prioritize streams of content over the graph from which that content was made. Presently, the API would be insufficient for a study such as that by Cha, Haddadi, Benevenuto, and Gummadi (2010), which focused on popular user accounts with many followers and their impact on geopolitical events. Arguably, this sort of work on the follower graph might be possible with circumvention tools, but this is beside the point. The main point is that companies have shifted their APIs away from easy and programmatic means of accessing social graph information. In Twitter’s case, this involved restricting the velocity of information. In Facebook’s case, it has involved restricting the variety of information.

There is little to stop companies from lifting API restrictions for certain third parties, short of privacy legislation that ought to prevent anyone from accessing such data. Different technical restrictions permit certain kinds of representations and not others. In doing so, they make certain kinds of insights knowable and others less knowable. Social graph data in particular have become increasingly less knowable. For some, this is about easing the burden on servers (as Twitter has claimed). For others, this is obvious monetization. LinkedIn, for example, still allows access to its “r_network” end point, but only for premium members who pay a fee. The only exception known to the author is the Socilab visualization, but this visualization is restricted to 500 contacts and prohibits a user from using Socilab to download those contacts.

Of the social network sites, Facebook’s API has seen the most changes as well as access to the greatest amount of social data. As the world’s dominant social graph, it absorbs a huge swath of user data. Facebook’s API is discussed in detail below.

Facebook’s Evolving API

Facebook introduced its developer platform in 2006. At this time, the API was simple in comparison to its current offerings. Facebook still branded itself as an online directory. The documentation page suggested that “you could create applications that utilize the following data: Profiles, Friends, Photos, and Events” (“Facebook Development Platform,” 2006, para. 2). Facebook introduced its API with end points such as “get_friends.”

The next major change in the API happened in April 2010, when Facebook introduced Open Graph API (Warren, 2011). With this new system, every element on Facebook from an event to a newspaper article to a friendship has its own ID that can be queried under the right conditions.

With Open Graph in place, Facebook began work on Graph Search (Olanoff, 2013), a complex technology that allows for queries that relate different parts of Open Graph. One could now use the search bar in Facebook to query for something as complex as “friends of my friends who live in Tokyo and like techno music.” In theory, Graph Search could have been an exemplary form of Facebook as a generative platform. Third parties could leverage this newfound capacity to query data in all sorts of novel ways. For example, one might create an application that asks for “people who are not currently my friends, but know my friends and are going to the same event as me.” This could be a way to introduce people at a

concert through mutual friends. Unfortunately, programmatic access to Graph Search never happened. Instead of rolling Graph Search into its API, Facebook began to make the API more restrictive.

The Open Graph API 2.0 specification indicated that friend lists were still available, albeit in a more restricted way (Constine, 2015). This change would prove to be consequential, although not necessarily being the most privacy sensitive. Prior to Graph API 2.0, any application could download a complete list of a user's friends and all the friendships between these friends. One could also download all the friendships between people in groups that the user has subscribed to (Polonski & Hogan, 2015). Consequently, third parties that had no reason to access these data could still get it for free. With Graph API 2.0, access to friendships was substantially clawed back. An app can now only access a list of the user's friends if those friends also authorize the app. This makes sense for gaming; the app could show the user which friends were also playing. However, if the user wanted to represent friendships as a set, they could no longer do so. Programs such as a time-line tracker or some sort of personal network visualizer would no longer work starting with Graph 2.0. In theory, one could ask each and every Facebook friend to authorize an app so that the user could get a complete view of their Facebook friendships. This sort of request is both unrealistic and impractical. It is unrealistic because users will not want to ask all their friends, as it puts the onus on the user to defend why they ought to have information, even though this information is already accessible through the Facebook Web page, just not programmatically. It is impractical because individuals will want to make use of this information when they add an app, not weeks later when everyone has consented or otherwise ignored the app request. These changes thus practically nullify any capacity for a third party to create or manage a representation of a user and their friends in real time. Some applications, such as Tinder, have gotten around these restrictions. Facebook has not publicly stated how Tinder accesses these data, but it has been reported that this was through a private deal with the company (Seetharaman & Dwoskin, 2015). Figure 1 illustrates the differences among the earlier Graph API, Graph API 2.0, and the special access granted to Tinder (and potentially other apps not known to the author).

A further shift in Facebook's API process made data access even more restrictive. Now, all applications requesting keys (the means by which they can generate the personalized OAuth tokens discussed above) have to meet Facebook's approval. This gives Facebook considerable power over both access and what constitutes a legitimate program for access. Apps have little leverage in this domain as there is no clear mechanism for arbitration. Within this new context, it would appear that users, developers, and researchers have now lost a form of generativity that we might call *relational generativity*.

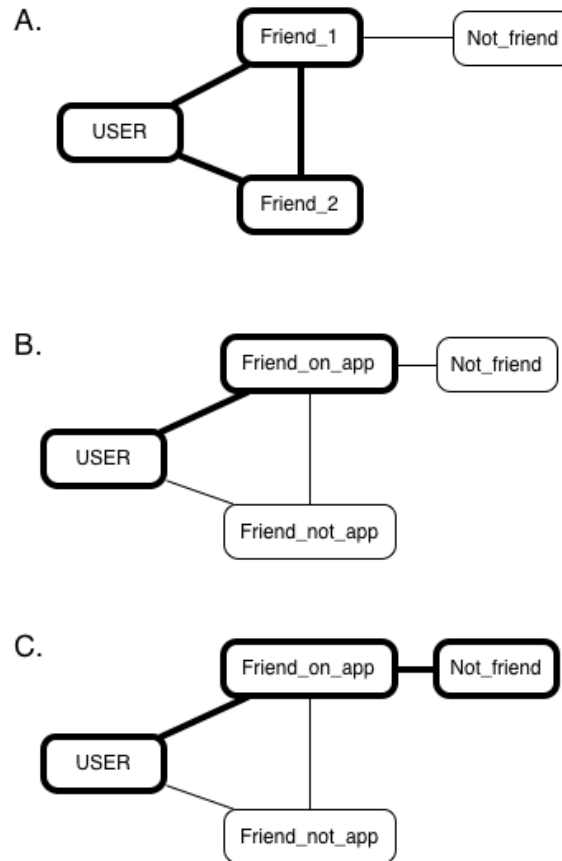


Figure 1. Three forms of access on Facebook. Links denote relationships that are held by Facebook. Bolded links are accessible. (a, Graph API 1.0) An app could request a user's friends as well as mutual ties but not extended ties. (b, Graph API 2.0) An app could only see ties between people who authorized the program and not the complete personal network. (c, Tinder) An app with special hidden access can currently see friends of friends.

Relational Generativity and Social Networks

Relational generativity follows from Zittrain's (2006) definition of generativity as the capacity of a Web technology to foster further innovation and disruption among its users. Classic generative enterprises include Linux and the Raspberry PI. Relational generativity, in this sense, is the capacity for users to represent and interact with their social relations in novel and innovative ways. To codify social networks themselves was a disruptive idea. These networks, however, can only be as generative as our ability to leverage this newly structured information. If those data are under lock and key, then one cannot do

much with them, and certainly nothing on a scale that would allow a novel interface or idea to transfer from obscure hackers to the wider Web.

With the advent of social network software, a new problem arose: how to manage what could be a huge volume of communication that might be relevant to a user. With hundreds of friends, many posting daily or more, a user could, in theory, be subjected to hundreds if not thousands of posts to review every day. In this paradigm, the common response has been to sort and/or filter these posts in a personalized way (Zuckerberg et al., 2012).

Facebook's news feed is a solution to the challenge of sorting information in a way that works with users' needs. Now one does not need to specify who is the closest to the user; Facebook can merely read signals from behavior, network structure, and tastes to infer who ought to be presented in the news feed (Bakshy, Messing, & Adamic, 2015).

Determining the ways in which content is accessed is a precondition for differences in how content is curated and represented. Some representations, such as a time line, might be very different than others, such as a word cloud. These representations are important as users appear to derive considerable benefits from Facebook use (Quan-Haase & Young, 2010; Smock, Ellison, Lampe, & Wohn, 2011), although the true value of Facebook participation has been contested in recent years (Tromholt, 2016). In work going back almost 10 years, it has been reported that those who participate more on Facebook typically claim they feel more connected to the wider world and report greater social capital (Burke, Kraut & Marlow, 2011; Steinfield, Ellison, & Lampe, 2008).⁴

The issue with Facebook regarding relational generativity is the asymmetry of control over relations (Fuchs, 2014). This issue is being exacerbated by recent concerns of fake news, filter bubbles, and collapsed contexts (Marwick & boyd, 2010; Pariser, 2011). It is being exacerbated because if Facebook is the only one to manage content, it is thus the only one that can manage filter bubbles and signals of fake news. Without ways to give alternative representations, we are left to assume that the social media platform is in the best (and often only) position to curate these data for users.

Facebook's company motto is now, "We want to give you the power to share and to make the world more open and connected" ("About Facebook," 2016, para 2.). The restrictions on APIs imply that users can only wield this power to share with the social graph signified on site through interfaces designed by the platform. The dominant interface orders content from one's social network according to its own logic via a proprietary algorithm. This is akin to presenting users with a proprietary route through spaces of social content, but not the map. By contrast, if a third party is given access to the data, it can create a map and enable the user various ways to traverse it. One might view posts from a geographic region, a select set of friends, containing a specific topic, or from a specific time.

⁴ As Appel et al. (2014), note, there remains considerable disagreement that the Williams social capital score measures social capital as traditionally measured, using specific network resources. Unfortunately, it is difficult to compare these measures more fully without further access to the personal network on Facebook.

These new forms of information presentation would enable users to have more control over the way they navigate content from the platform. Alternative ways of navigating these social spaces are thus one way in which the platform could function generatively. Facebook would still host and manage the social graph, but others could repurpose it in novel and imaginative ways. One such repurposing that is no longer feasible is the social network visualization. These visualizations are very much like the map, whereas the news feed is the route.

Visualizing Networks: The Map Rather Than the Route

Sociograms (or conventionally understood "network maps") have intuitive appeal for users and can enlighten individuals about the overall structure of a set of relations. From Moreno's early sociograms to recent big data approaches to representing scientific collaborations (Boyack & Klavans, 2007; Moody, 2004), network maps can help individuals quickly identify patterns, such as clusters of similar people (typically based on mutual friendships); users who link clusters (known as "brokers"; Burt, 1992); users who are highly connected; and ways that other features, such as geography, intervene and shape a social network.

One of the earliest attempts to visualize online social networks was Heer's Vizster (Heer & boyd, 2005). This program was partially a demonstration of the Prefuse visualization toolkit for Java and partially a demonstration of the potential new insights available by leveraging social media data. The name Vizster was play on Friendster, one of the early significant entrants to the social network site genre. Heer and boyd's original article reflects the innocent approach indicative of early work in this field. Vizster was justified as having "potential for fun and engaged social activity" (Heer & boyd, 2005, p. 1). They claim:

On the social side, we attempted to better facilitate the discovery of people, connections, and communities to promote increased awareness of community structure and information exposure, while preserving (and hopefully enhancing) a fun and engaging online space. From a design perspective, this case study explores the mutually informing use of ethnographic techniques and visualization design to craft a domain-specific visualization system in a context as much characterized by play as by analysis. (p. 1)

Embedded within this quote is the suggestion that users can benefit through exploration and an open system of friendship traversal. By viewing one's friendship connections in an intelligible way, it is suggested that Vizster can help individuals foster increased awareness of their social world.

Ten years after Vizster, Jeon et al. (2016) demonstrated that this sort of approach was useful as well as playful. In a within-subjects experimental design, the authors demonstrated that prospective first-generation college students were able to surface useful information in their Facebook network by visualizing their social ties as a social network map and exploring this map. While network maps are not guaranteed to make sense, techniques to arrange the network in meaningful ways have advanced considerably in the past decade. For example, techniques for color-coding nodes based on automatic group detection (formally called "community detection") have now become very accurate in representing real social groups (Lee & Archambault, 2016).

Facebook has never provided an in-house representation of a user's social network. Third parties have tried on Facebook, with mixed outcomes. Early entrants included TouchGraph, a Java applet that used the same underlying technology as the original Vizster, and FriendWheel, a Facebook application that would arrange friends in a circle and draw lines between mutual friends. Later applications such as NameGenWeb (Brooks et al., 2014) and NetVizz (Rieder, 2013) focused on downloading one's social network for use in standard social network applications, such as Gephi or NodeXL. The successor to NameGenWeb, ColledgeConnect, was used in Jeon et al. (2016).

In contrast to Facebook, LinkedIn did provide an in-house social network visualization package called InMaps (LinkedIn Labs, 2011). In keeping with the general norm that platforms are reluctant to reveal data, InMaps did not send friendship data to the browser to be rendered in real time. Instead, it prerendered the data as an image on LinkedIn's servers. This ensured that a user could not recover the underlying relationship structure. Because LinkedIn was monetizing access to these indirect ties, it did not make sense for InMaps to give these away for free. After several years, InMaps was shut down.

For a small period of time, one could replicate the functionality of InMaps using the API and some code (Villedieu, 2015). A year later, however, LinkedIn has restricted their API so that the requisite end point for retrieving network data (the `r_network` end point) is only available to approved partners. This is not to say it is no longer possible. The Socilab project is one such example of an academic project that has been able to get approved partner status for a personal network visualizer (Tutterow, 2016). Socilab no longer exports data to comply with the new API restrictions and makes some considerable limitations on how much data can be accessed. Nevertheless, its existence demonstrates that the drive to prevent programmatic user access to a personal network is social convention based on product strategy rather than any necessary legal or ethical imperative to keep such ties locked away.

Maintaining Privacy: A Red Herring?

If a greater number of actors (and especially third parties) have access to social data, particularly data about friends of friends, this has positive as well as negative consequences for the platform, the user, and the public at large. To be fair to the competing position, it is worth noting that there are legitimate considerations for companies seeking to claw back access. Two in particular stand out. The first is privacy, which is in the user's interest, and the second is control, which is in the platform's interest. A third reason, efficiency, is plausible but contentious. As stated, Facebook originally indicated that it removed friendship permissions because they were rarely used.

Facebook has recently suggested that the ability to see this information has been revoked to maintain user privacy (Constine, 2015). As noted, this information used to be available through the default permissions to everyone. Most people would accept that such information ought to be given only if the user is aware, meaning the app should request special permissions to access friend lists. Nevertheless, it is clear that Facebook still leverages this information extensively on the site. Despite revoking access to third parties, every time someone views the profile of another person, Facebook displays friends in common. As such, this information is not hidden and kept private. It is rather controlled and kept out of the hands of third parties, specifically.

Given that Facebook still exposes this information to the user, it is difficult to suggest that it is kept private. However, one might say that programmatic access is a problem, as the third party (e.g., the researcher) could link multiple networks together in unintended ways. However, Facebook has also undermined its concern for this argument by giving access to the social graph to Tinder, an online dating application, in ways that are more invasive than what was previously available. Although friendship data are not currently available through the API, Tinder still appears to have access to friendship ties (Seetharaman & Dwoskin, 2015; Welch, 2016). This again reinforces the notion that privacy is a red herring, whereas control is tantamount. These data are available, and Facebook has determined how to share these data with a third party in conjunction with the user. It simply has not shared it with most third parties, particularly external researchers.

Social Network Data Post-APIs

Relational generativity concerns both the use of data by third parties for novel user representations and for compelling insights about the user that might not otherwise be available. For example, Stephen Wolfram (2013) has indicated a number of interesting properties of Facebook networks in a long essay from the results of his data collection, such as the fact that as people get older they tend to link to younger friends and family on the site. Other researchers have used programs such as NetVizz and NameGenWeb to capture and analyze these networks (Brooks et al., 2013; Buglass, Binder, Betts & Underwood, 2016; Jeon et al., 2016; Topirceanu, Duma, & Udrescu, 2016). Corporate uses of relationships on Facebook range from the unwanted spam of Zynga to useful apps such as Job Fusion, which leveraged contacts for job postings (Constine, 2015).

In an era when the most useful and well-structured friendship graph is no longer available for use, what does such work portend for social network researchers? What solutions exist for trace data other than API access? Following are four scenarios for future data collection:

1. *A return to recall data.* Even Facebook does not have a perfect overlap with an individual's personal social network (Hampton, Goulet, Rainie & Purcell, 2011). Studies interested in this general personal network might now return to recall data. Fortunately, this domain of research is seeing a resurgence as it also adapts to new technological capabilities (D'Angelo, Ryan, & Tubaro, 2016; Hogan et al., 2016).
2. *Working with the platform themselves.* Facebook as well as LinkedIn, Twitter, Tumblr, and others tend to have strong relationships with the academic research community. Working at the companies on sabbatical, secondment, or specific projects is a reasonable strategy for people who wish to work with data that might not be generalizable outside of the platform, but can improve the platform experience. This will predominantly be reasonable for work that does not challenge the ideologies at the companies. Work that promotes Facebook will be doable at Facebook. Work that shows Facebook's potential harm to mental health (e.g., Tromholt, 2016) will most likely not be as welcome.

3. *Appeal to the platform for special external access.* Asking a platform for a special developer key is possible, but it tends to operate through interpersonal relationships and high-powered marketing departments (e.g., Tinder). Some people are lucky and are grandfathered through API changes (e.g., Socilab for LinkedIn); others leverage the intellectual legacy of their academic departments where former colleagues and students now work at a platform's research department. The concern with these strategies is that they lean heavily on social capital and not necessarily on the importance of the research question. There is also far too much uncertainty in this strategy for planning long-term grants.
4. *Covert data access.* Although technologically complicated, it is still possible to access data on-screen using a browser plug-in or some type of screen scraper. This sort of work has already had success (e.g., Gilbert & Karahalios, 2009). For some, *prima facie*, this is not necessarily unethical, but it is not in the spirit of collaboration with the platform that controls the data. API-based work enables the platform to regulate the third party. Programmatic access through an authenticated API, rather than around it, ought to be the fairest way to access data. With authentication, the platform can know who is accessing the data and can ask for what purpose. With a browser plug-in or screen scraper, the platform knows that the user is requesting data but does not know that a third party is also looking at those data.

None of these alternatives are satisfying, as they imply a compromised relationship between the researcher and the platform. Academics have a particularly important role as arbitrators in the court of public opinion. When new features or findings emerge on platforms, academics are often called on to interpret or challenge these findings. If academics fear having their access revoked or having a legal challenge for improper data use, it is possible that they will acquiesce to the ideologies of a platform rather than to the interests of the users when the two come into conflict. This means sidestepping questions that paint the site in a negative light in favor of exploring the restricted set of "win-win" topics. Consequently, analysis on social media data published by company researchers should be treated with the kind of skepticism reserved for pharmaceutical research by companies who predominantly release positive and significant results. Extending this analogy further, one can say that pharmaceutical companies do create value for the public and incur considerable costs in doing so. As with drug manufacture, it is possible to acknowledge that social media are in the public interest when developed, distributed, and consumed appropriately (e.g., in ways that increase social capital and facilitate social support as well as entertain). Yet there are still risks and unknowns involved with privately held massive-scale social graphs that require checks and balances from independent evaluators. Programmatic access to meaningful data about these graphs would appear to be an integral part of such checks and balances.

Conclusion

The future does not look bright for APIs on social network sites. Although this is unfortunate for our sense of the Web as generative, what is especially sad is that the battle for social graph APIs was lost

without a fight. There were no protests when Open Graph API changed. There was barely a mention in the press (Seetharaman & Dwoskin, 2015). The regulation of API access is not a glamorous topic. Furthermore, it is a confusing topic for individuals who feel that Facebook has been continually improving its core business. Following this, it will be a long time before a site, platform, or protocol replaces Facebook. Other competitors have emerged, but they are also not moving toward extensibility and generativity. Snapchat, Tumblr, and Pinterest are all large social network platforms with millions of users. Snapchat's API is closely held by design to ensure that data remain evanescent. Tumblr's API does not have obvious end points for followers and friends. Pinterest's API does not include friendship data.

By providing a historical account of an era in which Facebook potentially offered more than a news feed, it was my goal to disrupt the notion that what is currently offered is in the best interests of users, innovation, and research. Rather, innovation on the social graph must now happen in house at the platform and according to the logic of the platform. This is a logic that prefers single ordered lists (e.g., a news feed) instead of navigable spaces (e.g., network visualization; Hogan, 2015). This comes with a concomitant ideology that giving the platform more data is the key route to a better experience as these data are used to train the algorithms ordering the data. Yet without secure and legitimate programmatic access to the social graph, we are left to the mercy of the values and approaches of the platform. In many reasonable circumstances, these values and approaches may come into conflict with our own as scientific researchers, as developers and as people with our own social networks to which we connect through digital means.

References

- About Facebook. (2016, July 1). *Facebook*. Retrieved from <https://web.archive.org/web/20171014212713/https://www.facebook.com/facebook/about/>
- Appel, L., Dadlani, P., Dwyer, M., Hampton, K. N., Kitzie, V., Matni, Z. . . . Teodoro, R. (2014). Testing the validity of social capital measures in the study of information and communication technologies. *Information, Communication & Society, 17*(4), 398–416. doi:10.1080/1369118X.2014.884612
- Bakshy, E., Messing, S., & Adamic, L. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science, 348*(6239), 1130–1132. doi:10.1126/science.aaa1160
- Berlind, D. (2015, December 3). What are APIs and how do they work? *Programmable Web*. Retrieved from <http://www.programmableweb.com/api-university/what-are-apis-and-how-do-they-work>
- Bian, Y. (1997). Bringing strong ties back in: Indirect ties, network bridges, and job searches in China. *American Sociological Review, 62*, 366–385.
- Boyack, K., & Klavans, D. (2007). Scientific method: Relationships among scientific paradigms. *Seed Magazine, 9*, 36–37.

- boyd, d. m., & Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer Mediated Communication*, 13(1), 210–230. doi:10.1111/j.1083-6101.2007.00393.x
- Brooks, B., Hogan, B., Ellison, N., Lampe, C., & Vitak, J. (2014). Assessing structural correlates to social capital in Facebook ego networks. *Social Networks*, 38(1), 1–15. doi:10.1016/j.socnet.2014.01.002
- Buck, S. (2012, May 24). The history of Facebook's developer platform [infographic]. *Mashable*. Retrieved from <http://mashable.com/2012/05/24/facebook-developer-platform-infographic/>
- Buglass, S. L., Binder, J. F., Betts, L. R., & Underwood, J. D. M. (2016). When "friends" collide: Social heterogeneity and user vulnerability on social network sites. *Computers in Human Behavior*, 54, 62–72. doi:10.1016/j.chb.2015.07.039
- Burke, M., Kraut, R., & Marlow, C. (2011). Social capital on Facebook: Differentiating uses and users. *Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems* (pp. 571–580). New York, NY: ACM. doi:10.1145/1978942.1979023
- Burt, R. S. (1992). *Structural holes: The structure of competition*. Cambridge, MA: Harvard University Press.
- Cha, M., Haddadi, H., Benevenuto, F., & Gummadi, K. P. (2010). Measuring user influence in Twitter: The million-follower fallacy. *Proceedings of the 4th International AAAI Conference on Weblogs and Social Media (ICWSM)* (pp. 10–17). Menlo Park, CA: AAAI. Retrieved from <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/download/1538/1826>
- Constine, J. (2015, April 28). Facebook is shutting down its API for giving your friends' data to apps. *Tech Crunch*. Retrieved from <https://techcrunch.com/2015/04/28/facebook-api-shut-down/>
- Cunningham, A. (2012, August 17). New API severely restricts third-party Twitter applications. *Ars Technica*. Retrieved from <https://arstechnica.com/information-technology/2012/08/new-api-severely-restricts-third-party-twitter-applications/>
- D'Angelo, A., Ryan, L., & Tubaro, P. (2016). Visualization in mixed-methods research on social networks. *Sociological Research Online*, 21(2), 15. doi:10.5153/sro.3996
- Dunbar, R. I. M. (1998). The social brain hypothesis. *Evolutionary Anthropology: Issues, News, and Reviews*, 6(5), 178–190. doi:10.1002/(SICI)1520-6505(1998)6:5<178::AID-EVAN5>3.0.CO;2-8
- Facebook development platform. (2006). *Internet Archive*. Retrieved from <https://web.archive.org/web/20060820131607/https://developers.facebook.com/>
- Fuchs, C. (2014). *Social media: A critical introduction*. London, UK: SAGE Publications.

- Game requests. (2017). *Internet Archive*. Retrieved from <https://web.archive.org/web/20170727092437/https://developers.facebook.com/docs/games/services/gamerequests>
- Gannes, L. (2007, May 24). Live at the Facebook launch. *Gigaom*. Retrieved from <https://gigaom.com/2007/05/24/live-at-the-facebook-launch/>
- Gilbert, E., & Karahalios, K. (2009). Predicting tie strength with social media. In *Proceedings of the 27th Annual SIGCHI Conference on Human Factors in Computing Systems* (pp. 211–220). New York, NY: ACM. doi:10.1145/1518701.1518736
- Hampton, K. N., Goulet, L. S., Rainie, L., & Purcell, K. (2011). Social networking sites and our lives. *Pew Internet and American Life Project*. Retrieved from <http://www.pewinternet.org/2011/06/16/social-networking-sites-and-our-lives/>
- Heer, J., & boyd, d. (2005). Vizster: Visualizing online social networks. *IEEE Symposium on Information Visualization, 2005* (pp. 32–39). Minneapolis, MN: IEEE. doi:10.1109/INFVIS.2005.1532126
- Hennig, M., Brandes, U., Pfeffer, J., & Mergel, I. (2012). *Studying social networks: A guide to empirical research*. Berlin, Germany: Campus Verlag.
- Hogan, B. (2010). Visualizing and interpreting Facebook networks. In D. Hansen, M. A. Smith, & B. Shneiderman (Eds.), *Analyzing social media networks with NodeXL* (pp. 165–180). Burlington, MA: Morgan Kaufmann.
- Hogan, B. (2015). From invisible algorithms to interactive affordances: Data after the ideology of machine learning. In E. Bertino & S. Matei (Eds.), *Roles, trust, and reputation in social media knowledge markets* (pp. 103–117). Cham, Switzerland: Springer.
- Hogan, B., Melville, J., Phillips, G., II, Janulis, P., Contractor, N., Mustanski, B., . . . Birkett, M. (2016). Evaluating the paper-to-screen translation of participant-aided sociograms with high-risk participants. In *Proceedings of the 2016 Conference on Human Factors in Computing—CHI'16* (pp. 5360–5371). New York, NY: ACM. doi:10.1145/2858036.2858368
- Hogan, B., & Wellman, B. (2014). The relational self-portrait: Selfies meet social networks. In M. Graham & W. H. Dutton (Eds.), *Society and the Internet: How networks of information and communication are changing our lives* (pp. 53–66). Oxford, UK: Oxford University Press.
- Jeon, G. Y., Ellison, N. B., Hogan, B., & Greenhow, C. (2016, February–March). First-generation students and college: The role of Facebook networks as information sources. *Proceedings of the 2016 ACM Conference on Computer-Supported Cooperative Work and Social Computing*, San Francisco, CA. doi:10.1145/2818048.2820074

- Johns, N. (2016). *The age of sharing*. Cambridge, UK: Polity Press.
- Kanai, R., Bahrami, B., Roylance, R., & Rees, G. (2011). Online social network size is reflected in human brain structure. *Proceedings of the Royal Society B: Biological Sciences*, 279(1732). doi:10.1098/rspb.2011.1959
- Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, 110(15), 2–5. doi:10.1073/pnas.1218772110
- Lee, A., & Archambault, D. (2016). Communities found by users—Not algorithms: Comparing human and algorithmically generated communities. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 2396–2400). New York, NY: ACM. doi:10.1145/2858036.2858071
- Lessig, L. (2000). *Code and other laws of cyberspace*. Toronto, Canada: HarperCollins Canada.
- LinkedIn Labs. (2011, February 16). LinkedIn maps: Your professional world visualized. *Internet Archive*. Retrieved from <http://web.archive.org/web/20110216200850/http://inmaps.linkedinlabs.com/>
- Lomborg, S., & Bechmann, A. (2014). Using APIs for data collection on social media. *The Information Society*, 30(4), 256–265. doi:10.1080/01972243.2014.915276
- Marwick, A. E., & boyd, d. (2010). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society*, 13(1), 114–133. doi:10.1177/1461444810365313
- Moody, J. (2004). The structure of a social science collaboration network: Disciplinary cohesion from 1963 to 1999. *American Sociological Review*, 69(2), 213–238. doi:10.1177/000312240406900204
- Olanoff, D. (2013, January 15). Facebook announces its third pillar “Graph Search” that gives you answers, not links like Google. *TechCrunch*. Retrieved from <https://techcrunch.com/2013/01/15/facebook-announces-its-third-pillar-graph-search/>
- Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you*. New York, NY: Penguin.
- Polonski, V. W., & Hogan, B. (2015). Assessing the structural correlates between friendship networks and conversational agency in Facebook groups. In *Proceedings of the 9th International AAAI Conference on Weblogs and Social Media* (pp. 674–677). Palo Alto, CA: AAAI Press.
- Popken, B. (2010, April 21). Facebook used to make partners delete your data after 24 hours. No longer. *The Consumerist*. Retrieved from <https://consumerist.com/2010/04/21/facebook-used-to-make-partners-delete-your-data-after-24-hrs-no-longer/>

- Quan-Haase, A., & Young, A. L. (2010). Uses and gratifications of social media: A comparison of Facebook and instant messaging. *Bulletin of Science, Technology & Society, 30*(5), 350–361.
doi:10.1177/0270467610380009
- Rate limiting. (2017). *Twitter*. Retrieved from <https://dev.twitter.com/rest/public/rate-limiting>
- Rieder, B. (2013). Studying Facebook via data extraction: The Netvizz application. In *Proceedings of WebSci '13, the 5th Annual ACM Web Science Conference* (pp. 346–355). New York, NY: ACM.
doi:10.1145/2464464.2464475
- Rieder, B., Abdulla, R., Poell, T., Woltering, R., & Zack, L. (2015). Data critique and analytical opportunities for very large Facebook pages: Lessons learned from exploring “We are all Khaled Said.” *Big Data & Society, 2*(2). doi:10.1177/2053951715614980
- Seetharaman, D., & Dwoskin, E. (2015, September 21). Facebook’s restrictions on user data cast a long shadow. *The Wall Street Journal*. Retrieved from <http://www.wsj.com/articles/facebooks-restrictions-on-user-data-cast-a-long-shadow-1442881332>
- Smock, A. D., Ellison, N. B., Lampe, C., & Wohn, D. Y. (2011). Facebook as a toolkit: A uses and gratification approach to unbundling feature use. *Computers in Human Behavior, 27*(6), 2322–2329. doi:10.1016/j.chb.2011.07.011
- Spehar, J. (2014). The new Facebook login and Graph API 2.0. *Facebook Developer News*. Retrieved from <https://developers.facebook.com/blog/post/2014/04/30/the-new-facebook-login>
- Steinfeld, C., Ellison, N., & Lampe, C. (2008). Social capital, self-esteem, and use of online social network sites: A longitudinal analysis. *Journal of Applied Developmental Psychology, 29*(6), 434–445.
doi:10.1016/j.appdev.2008.07.002
- Stern, M. (2014, October 13). “The Snapping” is real: 90,000 private photos and 9,000 hacked Snapchat videos leak online. *The Daily Beast*. Retrieved from <http://www.thedailybeast.com/articles/2014/10/13/the-snapping-is-real-90k-private-photos-and-9k-videos-hacked-and-leaked-online.html>
- Topirceanu, A., Duma, A., & Udrescu, M. (2016). Uncovering the fingerprint of online social networks using a network motif based approach. *Computer Communications, 73*, 167–175.
doi:10.1016/j.comcom.2015.07.002
- Tromholt, M. (2016). The Facebook experiment: Quitting Facebook leads to higher levels of well-being. *Cyberpsychology, Behavior, and Social Networking, 19*(11), 661–667.
doi:10.1089/cyber.2016.0259

Tutterow, S. (2016, October 27). LinkedIn network visualization and analysis. *Internet Archive*. Retrieved from <http://web.archive.org/web/20161027105030/http://www.socilab.com/>

Villedieu, J. (2015, August 27). LinkedIn InMaps discontinued: How to visualize your professional network now? *Internet Archive*. Retrieved from <http://web.archive.org/web/20160516020332/https://linkurio.us/linkedin-inmaps-discontinued-visualize-network-now/>

Vitak, J., & Kim, J. (2014). "You can't block people offline": Examining how Facebook's affordances shape the disclosure process. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work* (pp. 461–474). New York, NY: ACM. doi:10.1145/2531602.2531672

Warren, C. (2011, April 13). Facebook Open Graph: What it means for privacy. *Mashable*. Retrieved December from <http://mashable.com/2010/04/21/open-graph-privacy/>

Welch, C. (2016, April 27). Tinder's new "Social" feature reveals which Facebook friends are swiping. *The Verge*. Retrieved from <https://www.theverge.com/2016/4/27/11518034/tinder-social-reveals-swiping-facebook-friends>

Wolfram, S. (2013, April 24). Data science of the Facebook world [Web log post]. Retrieved from <http://blog.wolfram.com/2013/04/24/data-science-of-the-facebook-world/>

Zittrain, J. L. (2006). The generative Internet. *Harvard Law Review*, 119(7), 1974–2040.

Zuckerberg, M., Bosworth, A., Cox, C., Sanghvi, R., & Cahill, M. (2012). *Communicating a newsfeed of media content based on a member's interactions in a social network environment*. U.S. Patent No. US8171128 B2. Washington, DC: U.S. Patent and Trademark Office.