How APIs Create Growth by Inverting the Firm

Seth G. Benzell[*], Jonathan Hersh[†], Marshall Van Alstyne [‡]

This draft: August 7, 2021

Abstract

How might technology increase firm value? One method might be to facilitate more efficient use of internal capital. Another method might be to help the firm tap third party capital. This paper uses four unique data sets to measure growth in firm value based on adoption of Application Programming Interfaces (APIs), a technology that lets firms modularize and reconfigure resources for internal use or expose them to third parties for external use. The latter includes apps and services of the platform economy. We perform difference-in-difference and synthetic control analyses of financial outcomes for public firms and find that adopters of externally facing APIs grew an additional 38% over 16 years relative to non-adopters. Internal use cases were inconclusive. Using proprietary data on private APIs, we find that firms with public APIs grew faster after adoption than firms with private APIs. Then, using a Tobin's Q framework, we measure whether API adopting firms grew by lowering capital adjustment costs. Consistent with an inverted firm hypothesis, where value creation moves from inside to outside, we find that using the technology for external value creation explains more firm growth than using it for internal value creation. Finally, we document an important downside of API adoption: increased risk of data breach. Together these facts lead us to conclude that APIs, as the foundation of digital ecosystems, have a large and positive impact on economic growth and do so primarily by enabling external complementors rather than boosting internal productivity.

1 Introduction

In the information age, the value of a firm rests fundamentally on how it stores, shares and processes information.[1] Digital infrastructure is therefore central to a firm's success. For platform businesses, which rely on creating an ecosystem of interactions and capturing a share of the resulting surplus, this truism holds especially strongly. Such systems harness third party resources they do not own (Parker et al., 2017), enable collaboration between actors and integration among resources (Baldwin and Clark, 2006), and feature a modular architecture of remixable resources (Lusch and Nambisan, 2015).

This paper quantitatively investigates the firm-level consequences of a particular type of digital infrastructure. Specifically, we investigate the performance and operations of firms that

[*]Argyros School of Business and Economics, Chapman University, Stanford HAI Digital Economy Lab & MIT Initiative on the Digital Economy, benzell@chapman.edu

[†]Argyros School of Business and Economics, Chapman University

[‡]Boston University& MIT Initiative on the Digital Economy

[1]The value of US corporate intangible assets increased from near zero in the early 1990s to more than \$6.6 Trillion in 2016 (Benzell and Brynjolfsson, 2019) – calculated as US corporate equity and liabilities less financial assets from Federal Reserve series Z.1 and less fixed capital from BEA table 4.1.

implement application programming interfaces (APIs). APIs are a set of tools and protocols that allow computers to communicate directly with each other, as opposed to digital communication mediated through one or more human users. When designing an API, the architect decides how much of the computer system or data to expose to which users. Being accessible on the web, these API endpoints act as a constant conduit to business processes that the firm itself specifies[2]. As a foundation for a digital infrastructure, APIs offer the dual virtues of practical modular design and precise metering of access. Modular architecture allows designers to independently create, subdivide, modify, and remove components without affecting other parts of a larger system (Baldwin and Clark, 2006). In effect, one can "virtualize" business processes in the manner of digitizing computer resources (Iyer and Henderson, 2010). This also facilitates partitioning of decision rights over access to that data (Tiwana et al., 2010). Modularity combines the advantages of standardization typically associated with high volume processes together with the advantages of customization typically associated with bespoke processes (Baldwin and Clark, 2000). APIs also enable precise metering of access permissions to those key resources. Metered access permissions ensure that anyone and anything that consumes system resources adheres to technical and economic policies designed to ensure system health (Jacobson et al., 2011). As an architecture device, APIs provide scalable infrastructure for building platforms. As a regulatory scheme, APIs partition decision rights, augmenting the API controller's ability to govern behavior. In fulfilling these roles – architecture and governance – APIs serve as the foundation for digital platforms (Parker et al., 2016).

We provide empirical evidence that APIs help firms grow, and that they do so primarily by "inverting the firm," i.e. by enlisting third parties to create complements to their products and services. Value creation, moderated and enabled by APIs, moves from inside to outside the firm.

First, we document the development of the API network, which links the business interests of companies together through a matrix of third-party applications that call on one or more APIs. Using summary statistics, difference-in-difference estimates, and synthetic control analyses, we document strong growth in market value among firms that adopt public APIs, an effect driven by firms with consumer facing APIs.

We further show that firms with APIs that achieve high third-party engagement have particularly large gains. Firms with APIs that have more followers, developers, and apps connecting them to other APIs see significantly larger API growth. The same holds for firms with more central APIs. The fourteen public firms with APIs ranked in the top forty by betweenness centrality added $6.6 trillion dollars in market value over the period of interest, a significant

[2] An airline, for example, can provide an API for all of its available flights, allowing users to not just see available flights, but allowing them to process and curate this information via apps or websites that they themselves create.

fraction of the US public equity market's appreciation over that time. The positive effect of API adoption is larger for firms which have used APIs for a longer period of time, consistent with the hypothesis that an API with high third-party engagement increases the growth rate (rather than level) of a firm's market value.

On the other hand, we find weaker evidence for a positive internal productivity effect from use of APIs. Firms who create APIs for exclusively internal use do not see statistically significant growth in market capitalization after API adoption. We also test for the hypothesis that APIs create internal benefits by lowering adjustment costs by looking at the evolution of firm Q (the ratio of market value to book value) after API adoption. If APIs primarily helped firms by reducing adjustment costs, successful API adopting firms should not see large market value growth after controlling for their asset growth. We find the opposite, evidence consistent with the benefits of APIs coming primarily through building a digital ecosystem.

Finally, we investigate an important downside of API adoption, the risk of data breach. We find that firms with public APIs see significantly increased risk of hack events in the years after opening an API. We further find that, for a subset of firms whose API traffic we can observe, hack events cause an increase in testing and login authorization data flows, indicating that firms adapt their API use in order to manage exposure.

This paper proceeds as follows. Section 2 provides background, hypotheses, and literature. Section 3 describe our unique data sets, including (i) proprietary data on private APIs provided by an API services firm, (ii) data on public crowdsourced APIs, (iii) Compustat firm performance data, and (iii) matched breach events from the Privacy Rights Clearinghouse. Section 4 documents the evolution of the public API network over a fifteen year period. Section 5 provides model specification, robustness checks, and synthetic controls. Sections 6.1 and 6.2 test and contrast our main hypotheses, followed by Section 7, which explores whether data breach events implicate APIs and how firms respond. These are followed by a concluding Section 8 and supporting appendices.

2 Background and Hypotheses

An API is a set of routines, protocols, and tools that standardizes building software applications compatible with an associated program or database. APIs fundamentally are code that control access to information. They can also be thought of as contracts (Jacobson et al., 2011). They govern the type and format of calls or communications that any application can make of another associated program. The answering program is agnostic about the source of the call, yet can require access permission, and the calling program need not know anything about the internal workings of the answering program. In the same way that user interfaces on operating systems make personal computers easier to use, interfaces on applications make programs easier to use by machines as well as by people.[3] Firms use APIs to offer network access to distributed data and services. Such services logically represent a business activity, produce a specific result, are self-contained, and are readily recomposable.[4]

It is not obvious that APIs should boost market value. A 2019 analyst report to CIOs observes that, on average, IT investments have led to stagnation.[5] US productivity growth plateaued at 1% after 2010, yet IT investments rose at a rate of 5% over that same period. The report concludes with a warning to CIOs that justifying IT investments on the basis of productivity gains risks jeopardizing their credibility. There is, moreover, considerable downside risk to allowing third parties access to the firm's private data. APIs were implicated in a hack that released compromising and very private photos of celebrities stored on Apple's iCloud.[6] One Facebook API flaw allowed spammers to gain control over user accounts without interacting directly with those users.[7] Another API vulnerability allowed use of nothing more than a license plate to breach an insurance company and learn the name of the owner, all movements of the car, and its position in real time.[8] India banned the social video app TikTok in part for its use of APIs to gather inappropriate data on other apps, all hardware and network data, and the geolocation of the user.[9] The CEO, CIO, and CSO of credit scoring bureau Equifax all stepped down after an API hack released the personally identifiable information of 143 million people.[10] The editor in chief of ProgrammableWeb observes that API Security "is so hard that even the biggest companies with the deepest pockets to hire the best talent make mistakes."[11]

It is not clear when the first API was created, but they clearly predate the Internet. Google's n-gram tool lists usage of the phrase 'application programming interface' as early as 1961. That said, APIs fully came into their own in the Internet era. Many web-pioneers featured APIs as core to their businesses. Salesforce.com included them in their 2000 launch of the world's first 'software-as-a-service' product. Likewise, eBay launched a developer program in 2000 to a select group of partners, encouraging them to create services that drew information from eBay's API. Having created one of the first popular 'open APIs,' eBay's decision led to a virtuous cycle of

[3] Source: David Berlind, Chief Editor ProgrammableWeb.ATT Adds APIs to Help Partners Move More Mobility Products and Services

[4] API data types include documents, images, video, geolocation, news feeds, etc. API functions include ID verification, payment, notification, visualization, language translation, mapping, etc. Wikipedia

[5] Bartels, Andrew "The Shrinking Technology Gains from Technology: CIOs Will Need a New Approach to Justifying Their Technology Investments" *Forrester* May 1, 2019.

[6] https://www.programmableweb.com/news/naked-truth-about-internet-security/analysis/2014/09/17

[7] https://bitredes.top/a-flaw-in-the-facebook-api-allowed-any-account-to-be-hacked/

[8] https://www.andreascarpino.it/posts/how-my-car-insurance-exposed-my-position.html

[9] https://www.programmableweb.com/news/stark-reminder-about-api-security-india-bans-tiktok-over-privacy-concerns/analysis/2020/07/06

[10] https://techbeacon.com/app-dev-testing/post-equifax-why-api-security-should-be-priority

[11] New Snafus Illustrate the Scope of API Security Challenge and Hacker Tenacity

better tools, higher visibility, and more customers.[12]

Rising interest in APIs has gone hand in hand with the rising dominance of platform firms in the economy. In 2019, six of the top ten firms by market value were platforms.[13] APIs simplify the writing and operation of programs that communicate with online services and shared databases. They are essential for powering such systems as Google's documents and maps, Amazon's voice and web services, Apple's online market, and Facebook's authentication services. They mediate economic transactions. Their value is not only determined by the actions of their creators but also by the habits of their users and the strategic choices of third parties who connect systems and reuse components in unanticipated ways.

Perhaps the most iconic effort to place APIs at the center of a firm's strategy was Bezos' 'Big Mandate' of 2002. Frustrated by the haphazard way Amazon solved its digital challenges, and hoping to turn hard won lessons into new sources of revenue, he demanded, among other things, that:

- All teams will henceforth expose their data and functionality through service interfaces...
- There will be no other form of interprocess communication allowed: no direct linking, no direct reads of another team's data store, no shared-memory model, no back-doors whatsoever. The only communication allowed is via service interface calls over the network.
- All service interfaces, without exception, must be designed from the ground up to be externalizable. That is to say, the team must plan and design to be able to expose the interface to developers in the outside world. No exceptions. (Rowan, 2011, citing Yegge)

Since its founding, Amazon has transitioned from reselling printed paper to power selling all sorts of goods and services. How was it that a book seller came to be the world's largest provider of web services? In 'Working Backwards' (2021), Bryar and Carr give an insider's answer to that question. In 2002, Amazon launched the"Amazon Product [Advertising] API." This tool allowed outsiders to build links to Amazon product listings into their apps and websites. Announcing the project launch, Jeff Bezos remarked "We're putting out a welcome mat for developers—this is an important beginning and new direction for us... Developers can now incorporate Amazon.com content and features directly onto their own websites. We can't wait to see how they're going to surprise us." The program attracted over 25,000 users in the first year. One of the biggest surprises was that internal Amazon developers often preferred using resources from the open API to Amazon's internal tools.

The success of the product API led Amazon management to consider other internal strengths they could externalize and monetize, such as data storage and messaging. Small companies needed to buy their own data centers (or hire a partner to do it) as well as manage the software needed to run them – a complex procedure for a company that is not a 'digital native'. Amazon

[12] Op. cit. Berlind

[13] As measured by the presence or absence of external developers. These firms are Apple (1), Google (2), Microsoft (3), Amazon (4), Facebook (5), and Alibaba (8). Source: June 19, 2019. Source: Wikipedia

launched the Amazon S3 API to provide an inexpensive simple storage solution. Amazon's EC2 API, providing elastic cloud computing, quickly followed.

APIs brought results. By 2013, Amazon's marketplace featured more than 2 million third party sellers, accounting for roughly 40% of total sales. In 2020, Amazon Web Services, including S3 storage and EC2 computing, earned over $46 Billion in revenue.[14] Using partner sales data, Amazon has also moved to vertically integrate into 3% of its partners' top selling products (Zhu and Liu, 2018). Amazon's market capitalization has duly expanded. Bezos' gamble that there was more money in managing bytes than managing books succeeded handsomely.

2.1 Hypotheses

Should investments in APIs drive firm value? Theory and evidence offer four reasons why investment might have limited to no effect. First, internally, managers should invest in *any* asset, not just APIs, up to the point where marginal benefit equals marginal cost. APIs have existed since at least 1961, suggesting that consequences from new investment might be strictly marginal. Market capitalization, in particular, should change little as it aggregates across all firm activities. Second, externally, if firms are observed gaining advantage from APIs, then competitors should also invest and compete away that advantage. Competitors' investments restore a balance of "normal" profits. Third, external developers are not employees. Firms that open APIs often have no idea who the developers are. If developers choose not to engage with or use the APIs, then no external value is created. Voluntary third party investment that never materializes could not then drive market value. Fourth, empirical research suggests that IT investments frequently fail to deliver promised productivity gains. The 2019 analyst report to CIOs that highlighted stagnant growth from IT investment also noted that, at the sector level, the relationship between IT investment and growth was often negative.[15] Overlapping our research window, that study impugns any notion that investments in digital transformation, like those in ERP, CRM and others before, unconditionally deliver positive outcomes. Forrester's conclusion is consistent with recent (Brynjolfsson et al., 2018) and early (Brynjolfsson et al., 2002) academic research that IT investments alone can produce negligible or even periods of negative value until they are coupled with complementary investments in "organizational capital" and other intangible assets. Absent complementary investments in new processes, products and business models, we should not expect observable changes in market value. IT investments have a history of not affecting aggregate market value (Tam, 1998). Theory and evidence thus lead to the following null hypothesis.

H0: Implementing APIs does not affect a firm's market capitalization

[14] Amazon Web Services – Wikipedia

[15] Ibid., Forrester.

An alternative thesis suggests lags might separate the time of investment from the time firms realize gains (Brynjolfsson and Hitt, 2000; Brynjolfsson et al., 2018). Costly new investments can require an adjustment period during which the firm hires or trains technical staff, experiments with new capabilities, or customers learn of new offerings. Consistent with the theory of competitive markets, API investments might also grant temporary advantage until such time as competitors invest and catch up. Returns from the investment would then diminish. These arguments lead to an alternate version of **H0** such that gains from APIs might exist but be temporary.

H1: Implementing APIs provides a fixed or decreasing investment benefit.

If APIs were to improve firm value, a natural next question is how. What mechanism provides growth? Answers can be categorized into two possibilities, one internal and one external. APIs deployed internally might drive profits through new products or new sales channels, as in the case of reaching customers via mobile phones (Iyer and Henderson, 2010). Additionally, APIs grant firms metered control over outside access and the ability to capture new data. This can help firms price discriminate among existing products while enabling new kinds of digital services (Tiwana et al., 2010). APIs are more modular than traditional code, potentially increasing data and software access, reuse, and recombination (Yoo et al., 2012; Baldwin and Clark, 2000). The potential to remix resources in new ways creates option value (Baldwin and Clark, 2006). This expands a firms dynamic capabilities by providing low cost variation and selection of business routines (Teece, 1988; Eisenhardt and Martin, 2000). They facilitate the remixing of disconnected resources or pockets of expertise (Purvis et al., 2001), integration of new software into legacy software (Joseph et al., 2016), and speed IT deployment (Iyer and Subramanian, 2015). They help firms raise labor productivity for a given expenditure on programmers (Brynjolfsson and Hitt, 2000). Thus one of the main theories supporting APIs is their ability to lower adjustment costs. Summarizing leads to the following hypothesis.

H2: Implementing APIs increases market value through increasing *internal* efficiency such as the ability to repurpose capital.

The alternate explanation, external value-add, operates by a different mechanism. Public APIs facilitate development of third party complements (Parker and Van Alstyne, 2017). APIs differ in important ways from earlier outsourcing, back-office, and front-office technologies such as electronic data interchange (EDI), enterprise resource planning (ERP), and customer relationship management (CRM). First, target users were either for internal employees or *known* contractors. By contrast, APIs specifically emphasize "permissionless" innovation by *unknown* partners, who generate complements the firm never conceived (Thierer, 2016; Chesbrough and Van Alstyne, 2015; Parker et al., 2017). Salient illustrations of this external value add include the numerous apps sold by Apple, Amazon, and Google but that had nothing to do with these platforms themselves. More permissive licensing, which is enabled by APIs, has been shown to increase complementary device development among handset manufacturers (Boudreau, 2010).

Second, APIs are more than technical plumbing designed to decrease transaction costs or increase efficiency. They *enable* markets. The consequence is not merely a shift from hierarchies to markets or a shift in the 'make-vs-buy' decision (Malone et al., 1987). Instead of entering the market as a more efficient player, the focal firm *becomes* a market, an orchestrator of other firms' transactions. Orchestrating a market gives the platform visibility into the data passing through its systems, which provides insights into competitors' activities, margins, and opportunities (Khan, 2017). This yields a strategic information asymmetry that favors the platform sponsor at the expense of the platform partner (Zhu and Liu, 2018). Advantage born of this asymmetry contributes to antitrust scrutiny of platforms in the EU and the US (Schulze, 2019; Cabral et al., 2021). The strategy of using APIs to orchestrate thirs party value creation which the focal firm can then monetize has been called the "inverted firm" (Parker et al., 2017) where value creation shifts from inside to outside. This shift if reflected in the following hypthesis:

H3: Implementing APIs increases market value through *external* or 3^{rd} party contributions to value.

We can distinguish between **H2** and **H3** based on standard theories in finance. If APIs boost internal efficiency and make it easier to repurpose capital, then it will boost firms' investments. This would show up in the data as a decrease in Q – the ratio of market value to installed capital. Alternatively, if APIs primarily boost value by 'inverting the firm,' and causing third-parties to make investments, then the portion of firm value not explained by its capital stock will increase as a function of API adoption. This would increase Q. We test these alternate hypotheses below. Other ways we distinguish between them include splitting the sample into public vs. internal-use only APIs, and by investigating the relationship between market value growth and third party engagement.

Third party engagement might also be a function of the parties involved in different market segments. Upstream (B2B) and downstream (B2C) opportunities give rise to different API business cases based on supply chain efficiencies or customer access channels. B2B APIs improve order entry, supply chain integration, just-in-time inventory, and on-demand services that cut costs of ownership and improve work flow (Haranis, 2017). Insurance firms have package analytics services based on driver risk scores, roadside assistance, and vehicle telematics for sale to rival insurers, auto manufacturers, and ride share firms (Boulton, 2017). By contrast, B2C APIs externalize internal resources and provide infrastructure to support third parties who build on top to reach or service customers. Walgreens, for example, opened an API for printing color photos at its drugstores. Developers who incorporated Walgreens printing into their apps then drove users to print from their phones, social networks, and cloud accounts, while increasing in store traffic (Iyer and Subramanian, 2015). Facebook uses APIs to allow consumers to authenticate themselves with numerous third parties, capturing useful data on consumer interests. B2B use cases focus on higher value opportunities with a smaller number of partners whereas B2C use cases focus on lower value opportunities with a higher number of partners. Whereas permissioned access with known suppliers favors more traditional technology such as EDI, APIs support metered access with unknown suppliers enabling new forms of permissionless innovation. The larger number of consumers relative to producers together with new forms of permissionless innovation suggests that API impact might be greater for B2C than for B2B. Compared to earlier technology use cases, B2B APIs might be more incremental whereas B2C APIs might be more radical. This lead to the following hypothesis.

H4: Implementing consumer facing (B2C) APIs increases market value more than implementing business facing (B2B) APIs.

Once a firm opens to third parties, the opportunity for interactions among those parties creates new avenues for value creation and value capture. The assumption that network structure influences the resources available to parties embedded in that structure underpins a vast literature spanning decades of research (Simmel, 1922; Moreno and Jennings, 1938; Granovetter, 1973; Baker, 1990; Burt, 1992, 2009; Padgett and Ansell, 1993; Uzzi, 1997; Hansen, 1999; Podolny, 2001; Reagans and Zuckerman, 2001; Aral and Van Alstyne, 2011). The central argument is that structurally diverse networks provide access to diverse resources. Controlling the access points in that structure provides the means to broker opportunities (Granovetter, 1973; Burt, 1992), improve decisions (Hansen, 1999), resolve uncertainty (Podolny, 2001), boost productivity (Aral and Van Alstyne, 2011), innovate (Reagans and Zuckerman, 2001), and extract rents (Burt, 2009). Key measures of structural position include betweenness centrality, which measures the frequency of being on a shortest path (Borgatti, 2005), and effective size, which measures diversity (non-redundancy) and reach among network contacts (Burt, 2009). Our API data allow us to assess these measures in the network of apps that call multiple APIs and pass information back-and-forth among them. Third party developers frequently call APIs of different firms connecting them in larger webs of interaction. We track the evolution of this web of interaction in Section 4. Based on theories of advantage provided by network structure, we posit the following hypothesis.

H5: The network structure of applications that call APIs affects the market value of firms that implement them. Firms with higher API network centrality, more connections, and larger effective network sizes have higher market value.

Because we observe apps that call firm APIs and not just the APIs themselves, we can construct a picture of the network of linkages across firm APIs. This API Network describes how firms' software are interconnected, and how information is digitally shared across firms.

Certainly theory would predict that more popular nodes in a network – that is, those with more connections – will be more valuable (Katz and Shapiro, 1994). Beyond this, we can characterize whether firms positions themselves strategically in the API network in order to capture more value, and the extent to which classical network effects dominate the market. Theoretically, Niculescu et al. (2018) show that incumbents in markets with extremely strong network effects will close the network and make it proprietary; moderate network effects lead the incumbent to open the network. Given that the public API network is open, firms opening public APIs make implicit decisions to cooperate with other firms in the API network.

The API network that we analyze is characterized by several large firms operating the most popular APIs. What then is the best strategy for challengers seeking to enter this market? First, we seek to quantify how strongly firms with more popular APIs are able to capture value. The effect of network degree, or the number of connections to firms' APIs, on market value should reveal this impact. Next we will characterize the extent to which firms with APIs with higher betweenness centrality have higher market value. Betweenness centrality measures the fraction of shortest paths in the network traverse through a given node. Finally, we use a set of theories from Burt's structural holes literature (Burt, 1992, 2009) to characterize how amenable the API network is to competition. Specifically we will use a measure of effective network size to see if firms who bridge distinct sections of the API network are able to capture more value. Effective network size measures number of connections less the number of redundant connections in a network. A firm who is able to act as a bridge between two or more sub-networks will have a high effective network size. If we find a significant impact of effective network size on market value this indicates there is a strategic opportunity if firms can find ways to act as unique bridges between different sections of the network.

H6: Implementing APIs can create security holes. A firm's adoption of externally facing APIs is associated with an increased risk of data breach.

If APIs provide orchestration and innovation benefits, why might firms fail to adopt them? One reason for reluctance is the fear that malicious actors may pose as legitimate users and steal a firm's sensitive data. APIs can facilitate illegitimate access and increase the risk of data breach.

Notable data breaches tied to API flaws are numerous.[16] T-Mobile announced an API data breach had exposed private data of more than 2.3 million users (Spring, 2018). Google shut down Google+, its much maligned social networking venture, after revealing that the private data of more than 52 million users had been exposed to third parties through its APIs (Newman, 2018). Home Depot reported that 53 million email addresses and 56 million payment card numbers were

[16] For example: Tech Beacon article about Equifax data breach

exposed to hackers.[17] As the introduction noted, the CEO, CSO, CIO of Equifax all resigned after an API data breach forced the firm to protect consumers from fraud. Losses had reached more than $1.6 Billion by 2019.[18]

Each of these breaches illustrates a "leaky API," one that is vulnerable to hacking, misuse, or unintended disclosures because third parties are not properly metered or controlled when they request data. Open systems are more susceptible to hacking. Ransbotham (2016) finds that open source software - while often functionally superior - is more likely than closed source software to have zero-day exploits and more obvious avenues of attack. Kamiya et al. (2018) find that cyberattacks are associated with reductions in sales growth, investment, and stock market performance. They reduce reduce CEO bonuses. Makridis and Dean (2018) match data breach reports from the Privacy Rights Clearinghouse Compustat financial data and find a 10% rise in records breached is associated with a .2% fall in firm productivity. Spanos and Angelis (2016) perform a systematic literature review of the impact of information security events on stock market outcomes. They review 45 studies from 37 papers. Over 75% of these studies find a statistically significant effect of digital security events on stock prices. None of these studies, however, use API data to analyze technology adoption per se. This leads to our final hypothesis.

3 Data

Our paper draws on four main sources of data. These are: (1) Compustat data on finances of publicly traded firms; (2) comprehensive data on public APIs their connections from third-party apps (mashups) from the ProgrammableWeb crowdsourced directory; (3) the Privacy Rights Clearinghouse for data breach events as matched to Compustat by Rosati and Lynn (2021); (4) proprietary data on internal API usage from a private provider of API creation tools.

3.1 Financial Outcome

Firms' financial performance is provided by Compustat, which measures market capitalization and other covariates at the quarterly level. Our sample runs from Q1 2007 through Q3 2020.

3.2 Public APIs

Our main data source on firms' API usage comes from the ProgrammableWeb a crowdsourced database of public APIs and the apps that call them. Data used for this analysis was collected

[17] Threat Post article about 2013 Home Depot data breach. These hackers gained access to the Home Depot network by using the legitimate credentials of a third-party vendor, and therefore the event was classified as being by a malicious insider. Once inside the network, they penetrated the point-of-sale system from which they accessed customer credit card records.

[18] https://www.housingwire.com/articles/equifax-expects-to-pay-out-another-100-million-for-data-breach/

in winter 2020. APIs were categorized and matched to the firms that sponsored them by a team of research assistants and checked by the authors. ProgrammableWeb also has data on apps calling one or more APIs, called 'mashups' by ProgrammableWeb, emphasizing the role they play in recombining information from disparate sources. Submitters label these apps with various tags useful for categorization.

ProgrammableWeb data include the dates an APIs was first submitted and the list of apps calling that API. We also collect the number of users who express interest in an API (followers) and those who claim to work on applications using that API (developers), as well as the number of updates the API has undergone. All APIs with at least 15 followers, of which there were 3402, were matched to the firms that own them. The majority, 63.1% were associated with non-public for-profit companies while 19.6% of these APIs were associated with publicly traded firms, 3.2% were associated with governments, and 8.1% were associated with non-profit organizations.[19] Of the 206,411 follows of APIs with at least 15 followers, 33% are of APIs created by public firms. Firms with APIs tend to have higher market value than firms without APIs. We further categorized apps as primarily B2B, B2C, both, or unclassifiable. API orientation is roughly split between B2B and B2C APIs (APIs classified as 'both' are associated with both categories for the purpose of summary statistics and regressions).

Matching ProgrammableWeb data to Compustat allows us to categorize the firms which use APIs by industry. We observe broad trends over time as reported in Figure 1, which plots the fraction of firms by one digit SIC code that have at least one public API over the sample period. APIs grew across all industries, with services and transportation & public utilities growing the fastest. By the end of the sample period, roughly 3.5% of firms matched to Compustat in the services sector have public APIs , and 3% of firms in transportation and public utilities have public APIs. Figure A1 reports the fraction of firms with at least one API by two digit SIC code circa Q3 2020. We see that Air Transportation firms are the most likely to have public facing APIs, followed by firms in apparel, business materials, business services and miscellaneous manufacturing industries.

3.2.1 API Network Statistics

Using ProgrammableWeb's list of of apps that connect to public APIs we trace out the network of APIs connecting firms. For the subset of firms whose APIs connect to this network, we compute a series of network statistics. Each node corresponds to a firm's API and each link connects a firm's APIs via third-party applications.

We compute three network statistics for a given API: betweenness centrality (White and

[19] We have matched many of these non-public firms to Crunchbase information on startups. This is an intriguing database for future research on the financial impact of APIs, and strategic entrepreneurship.

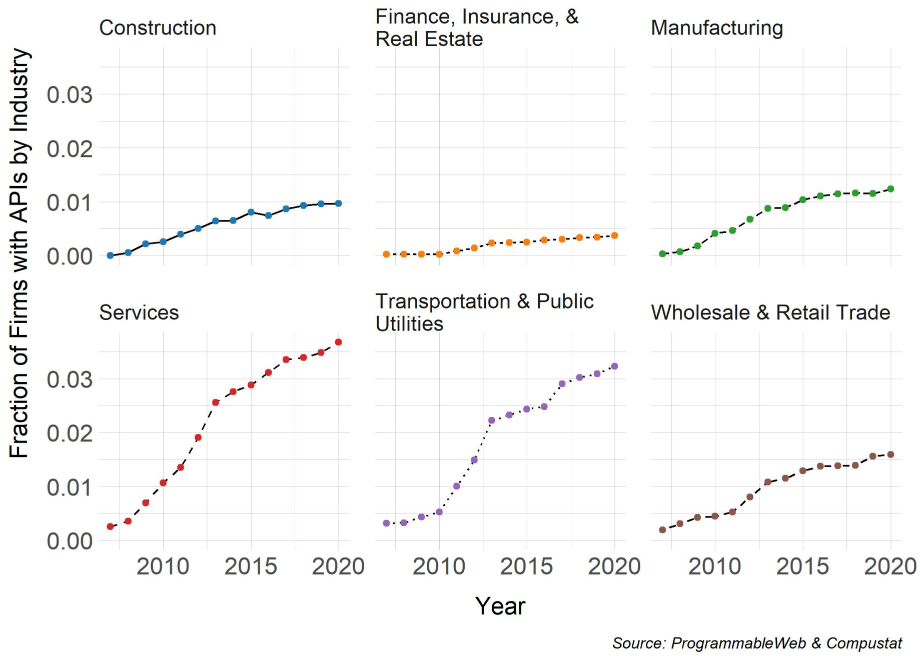

Figure 1: Fraction of firms with APIs by One Digit SIC Code.

Borgatti, 1994), degree (Diestel, 2005), and effective network size (Burt, 1992). Betweenness centrality calculates the share of shortest paths between nodes in the network that pass through a given node. Degree sums the total number of connections between a node and other nodes. Effective network size captures non-redundant connections that each node provides. This measure calculates effective network size as number of connections minus redundant connections between nodes. Network theorists refer to the latter measure as capturing a measure of 'structural holes' within a network (Burt, 2009). Nodes that rank high in effective network size act as structural bridges between sections of the network. Firms may operate several APIs, thus we calculate averages, maximums, and sums of these network statistics across all the APIs a given firm operates for each quarter in which it has at least one operational API. For ease of interpretation, these network statistics have been centered and scaled.

3.3 Private APIs

We received proprietary data from a consulting company that offers API development tools, implements APIs, and offers hosting services on behalf of API adopters. Many of the APIs are not published on ProgrammableWeb, as their use is restricted to actors within a firm. We matched this list of private APIs to our Compustat and ProgrammableWeb data. Some of the firms that operate public APIs also operate internal APIs.

To measure the effect of purely internal APIs, we identify that subset of firms from this proprietary dataset who do not have any APIs reported on ProgrammableWeb. This subsample leaves only internal API use as the treatment. In this sample, private APIs are less popular than public APIs, with approximately 0.7% of firm-quarters representing internal API use. The bulk of APIs, however, are private (Jacobson et al., 2011).

The API management firm also provided us with monthly records of API use for 273 separate accounts. This includes the name of each API used, as well as the number of calls and bytes processed by each API in a given month. Data on calls processed by partner firms' APIs span December 2012 to September 2016. Data on bytes processed span December 2012 to May 2016. We designate the first date that we observe any call to any of a firm's APIs as the API adoption date. Appendix A1 reports the total number of APIs, API calls, and API bytes of data flow that we observe in each month. We have 2,453 firm-months of API usage data. The average firm has 160 million API calls in a given month, as well as 1.98 trillion bytes of data. The average firm in this source of API data has 31.4 APIs.

3.4 Breach Data

The *Privacy Rights Clearinghouse* (PRC) records public breach announcements as matched to Compustat firms by Rosati and Lynn (2021). We collect data for all public and private firms we observe using APIs from 2005 to 2015[20]. PRC distinguishes six different breach types: "PHYS", "PORT", and "STAT" events involve the theft of physical storage media, paper documents, and stationary devices. "INSD" events involve breach events from insiders as well as malicious outsiders who have compromised insider credentials. "DISC" events are unintended disclosures. "HACK" events are incidents of hacking or malware leading to the data breach.

3.5 Data Challenges and Limitations

Our analysis faces several challenges and should be interpreted with caution. First, our ProgrammableWeb data, while hypothetically comprehensive, is crowdsourced and is therefore subject to bias in terms of APIs reporting. Our data from an API tool provision company gives us additional access to purely internal API data, but there are likely many companies deploying internal-use only APIs that we do not observe.

Two more challenges for our paper are endogeneity and anticipatory effects. Using generalized synthetic control methods, we are able to control for certain types of reverse causality. Using firm fixed effects, we are able to control for non-time varying latent factors that might

[20]Data from Rosati and Lynn (2021) end in 2015 so our main results focus on this period. A complementary appendix figure restricts attention to the 78 firms for which we have flow data, extending the PRC data through 2016, however this only increases the number of events by 4.

	Mean	Std. Dev.	Min	Max
Firm Financial Characteristics:				
Log of firm market value (All firms)	5.555	2.51	0	14.5
Log of firm market value (Firms with Public APIs)	9.095	2.03	0.25	14.5
Log of total assets (All firms)	5.962	2.77	0	15.0
Year	2013.0	3.95	2007	2020
Firm's API Characteristics:				
Year of firm's first API	2011.3	3.29	2005	2019
Share of firms ever using an API (public API)	0.0502	0.22	0	1
Total developers for firms' APIs	3.032	92.1	0	4378
Total followers for firms' APIs	20.45	390.5	0	16867
Total changes reported for firm's APIs	0.0989	1.99	0	66
API orientation B2B (business to business)	0.0314	0.17	0	1
API orientation B2C (business to consumer)	0.0331	0.18	0	1
Firm-Quarter Observations (Full sample)		133303		
Firm's API Network Statistics:				
Mean of firm's API network betweenness centralities	0	1.00	-0.44	5.87
Max of firm's API network betweenness centralities	0	1.00	-0.33	6.47
Mean of firm's API degrees	0	1.00	-0.80	7.19
Max of firm's API degrees	0	1.00	-0.54	4.04
Sum of firm's API degrees	0	1.00	-0.33	7.23
Mean of effective network size	0	1.00	-1.44	8.10
Max of effective network size	0	1.00	-0.63	4.33
Sum of effective network size	0	1.00	-0.50	7.06
Firm-Quarter Observations (Sub-sample of firms with APIs connected by apps)		6688		
Data Breaches:				
Any PRC breach event	0.00240	0.049	0	1
Breach via credit card fraud	0.000207	0.014	0	1
Breach via hack or or malware	0.000685	0.026	0	1
Breach via stolen physical device or documents	0.000163	0.013	0	1
Breach via lost or stolen portable device	0.000381	0.020	0	1
Breach via malicious insider	0.000479	0.022	0	1
Log number of records affected	0.00941	0.32	0	18.9
Firm-Quarter Observations (Data breach sample)		91946		
Internal API:				
Share of firms with internal APIs	0.00745	0.086	0	1
Firm-Quarter Observations (Internal APIs and control firms, no Public API firms)		126612		

Table 1: Statistics summarizing merged ProgrammableWeb public API data, Compustat financial outcome data, and Privacy Rights Clearinghouse data. Panel data organized at the firm-quarter level. Compustat data in millions of nominal US dollars. There are 179 public firms matched with APIs on ProgrammableWeb and 78 publicly traded firms for which we have data flow information. Of this 78, 44 are listed with APIs on ProgrammableWeb, leaving 34 observations of firms with purely internal-use APIs.

drive both market value and API adoption. What we cannot control is anticipatory effects: that is, news leakage might bid up a firm's market values in anticipation of a successful API strategy. When utilizing a firm fixed effects model, anticipatory market value increases will tend to muddy the waters and bias point estimates towards zero.

More subtly, to interpret our results properly, one must remember that our results correspond to a treatment effect on the treated group. In no way do we mean to argue that a firm ill-suited to adopting an API strategy will receive large positive results. Firms that adopt APIs successfully are likely predisposed to successful technology adoption along a number of dimensions. Google, Amazon and eBay in the 2000s – even before launching public APIs – had important head-starts in many aspects of the digital economy, including programming talent. In our context, they were able to capitalize on this head-start through development of an API network. However, in an alternate context where APIs never existed, they might have still been positioned to out-perform the market had they devoted resources to their next best digital opportunity. This subtle caveat is analogous to the one raised in Fogel (1962) about the role of railroads in US development. Despite large evidence that railroads played a critical role in regional economic growth, Fogel argues that for most purposes canals might have served as only slightly inferior substitutes. When measuring the contribution of railroads to development, one should be clear about whether the counterfactual is one where no infrastructure is built or one where a slightly inferior substitute is built. In the context of our setting, our estimates should not be taken as identifying the effect of APIs per-se on market value growth, but rather the effect of being the type of firm well suited to launching a successful API strategy at the right time.

4 Evolution of the API Network

This section characterizes evolution of the economy's API network, extending the snapshot provided in Evans and Basole (2016). Figure 2 presents the API network as recorded in the ProgrammableWeb directory through Q3 2020. Nodes in this graph correspond to APIs. Edges connect APIs when an app calls both. Node colors correspond to the company associated with the API. Edges are colored according to the functionality of the app that calls them. For example, DeployPlace is a app, designed as a developer tool. It interacts with Amazon S3 and Gmail APIs, among others. Therefore, there is at least one yellow-green line connecting these two APIs, indicating they are connected by a productivity focused API. Similarly, the 'ecomdash' service, an app involved in eCommerce, calls both the Amazon Product Advertising and eBay APIs. This is visualized by at least one green edge connecting the two nodes.

Several phenomena emerge from visual inspection of Figure 2. First is the relative frequency of companies appearing in the API network. The prevalence of green and orange nodes indicates

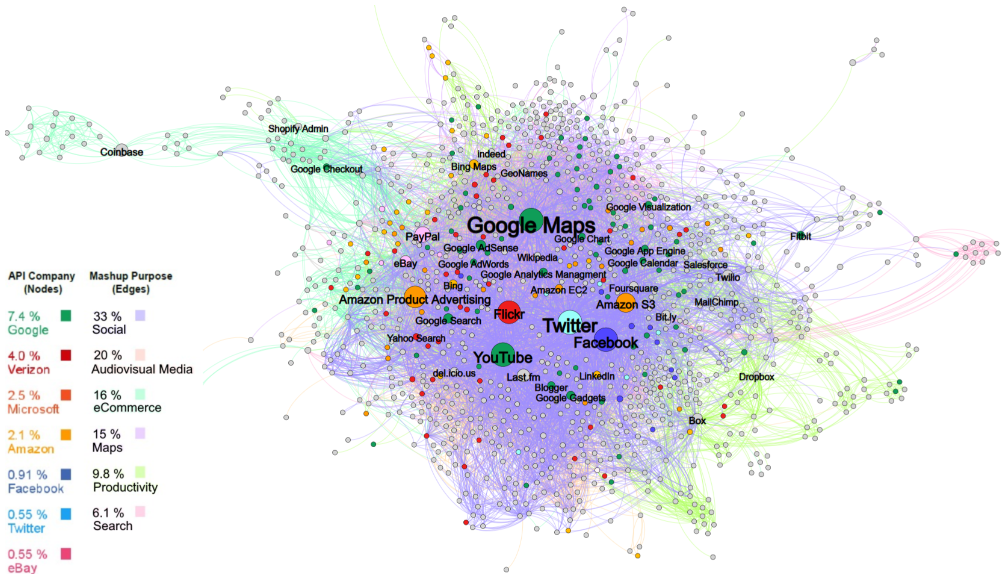

Figure 2: This figure visualizes the network of APIs and the apps that connect them as of Q3 2020. Larger nodes indicate the API has higher centrality. The forty nodes with the highest betweenness centrality are labeled. Nodes colors represent API sponsors (green for Google/Alphabet, dark blue for Facebook, light blue for Twitter, red for Verizon/Yahoo, dark-orange for Amazon, yellow-orange for Microsoft, and pink for eBay/PayPal. All other firms are grey.). Edges exist between any pair of APIs that are called by the same app. Edge color indicates the functionality of the apps calling the APIs (neon-green and yellow-green are eCommerce and Productivity respectively, periwinkle for Social Media, salmon for Maps, pink for Search, orange for Audio-Visual Content).

the network importance of Google/Alphabet and Amazon. Perhaps more surprising is the number of red nodes associated with Verizon/Yahoo. Facebook, Twitter and eBay are also central to the network but with many fewer nodes.

Appendix Exhibit A2 reports the company, degree, betweenness centrality and market capitalization growth for the top 40 APIs by betweenness centrality. The top five APIs ranked by betweenness centrality (Google Maps, Twitter, YouTube, Facebook, and Flickr) are also the highest ranked in-terms of degree. Unsurprisingly, these five APIs are both extremely popular for app calls and also central to the API network. Google Maps provides essential navigation functionality to a wide variety of apps, Twitter and Facebook are go-to social media plugins, while YouTube and Flickr provide popular video and image hosting websites.

Lower on the list, we see that some APIs have centrality ranks much higher than their degree. Firms with high betweenness centrality, whatever their degree, play an important gatekeeping role in that sector of the data economy, which can offer profit opportunities. The API with the most extreme discrepancy between its degree and betweenness ranks is CoinBase, which can be seen in the top left of Figure 2. This API is called by apps that also call several cryptocurrency related APIs (e.g. the Mt. Gox API) and is also connected by apps to several online shopping APIs such as Google Checkout and PayPal. These edges are all related to eCommerce. Absent

API connects to the core of the API network, many cryptocurrencies would be much harder to use in actual transactions. The Shopify API plays a complementary role in the portion of the network devoted to eCommerce, and Dropbox plays a similar role at the nexus of productivity-oriented apps (right side of Figure 2). Sub-networks, organized by purpose, appear in Figure 3. Unsurprisingly, Google Maps is at the center of the mapping network, while YouTube and Flickr are more central in the Audiovisual Media network.

Firms with central APIs saw dramatic increases in market value over our sample period. The 14 publicly traded firms that rank in the top 40 by API betweenness centrality added $6.584 trillion dollars to their market value from 2005 to 2021. This does not count the $179 billion in value that firms created by launching IPOs in that time period. The seven firms in that category that were publicly traded over the entire sixteen year period experienced a market value increase of 580.8%. By comparison, the entire US stock market grew by $16.89 trillion, or 99.3%, from 2005 to 2019.[21] The growth of the 14 firms at the center of the API network represent approximately a third of US market value growth over the time period under consideration.

Creating a top API without large growth in market value is rare. Most top APIs are governed by publicly traded companies. Only two of the top 40 APIs by betweenness centrality (i.e. 5%) are governed by non-profits. These are GeoNames, a location directory, and Wikipedia, an online encyclopedia. Of all ProgrammableWeb APIs with at least 15 followers, 13.3% are produced by governments or non-profits, meaning that for-profit companies are over-represented in the creation of top APIs.

Several notable features stand out concerning API organization. For example, the consumer facing Social (perwinkle) and Search (pink) apps densely connect the heart of the API network. APIs connecting these apps, especially Facebook, Twitter, YouTube, and Google Search, might drive engagement for the apps connecting to them. As Hypotheses **H4** noted, B2C facing APIs may be better at driving network effects than B2B APIs because it is more immediately obvious to third parties how to incorporate consumer facing features into their apps.

Unsurprisingly, the most central APIs in the network are also associated with both search or social media. APIs for Facebook, Twitter, and YouTube are some of the most connected APIs. Dropbox, Box, Salesforce, and Amazon S3 are important to the productivity cluster, yet these also include mapping functionality and are close to Google Maps, Indeed, Bing Maps, and GeoNames. The eCommerce cluster shows high density around the Amazon Product API, as well as the PayPal and eBay APIs.

Appendix Figure A3, which labels nodes for all APIs owned by a given company, gives

[21] See World Bankhttps://data.worldbank.org/indicator/CM.MKT.LCAP.CD?locations=US. According to an alternate source, the total US equity market increased by 30 Trillion in value from 2005 to 2021. See https://siblisresearch.com/data/us-stock-market-value/

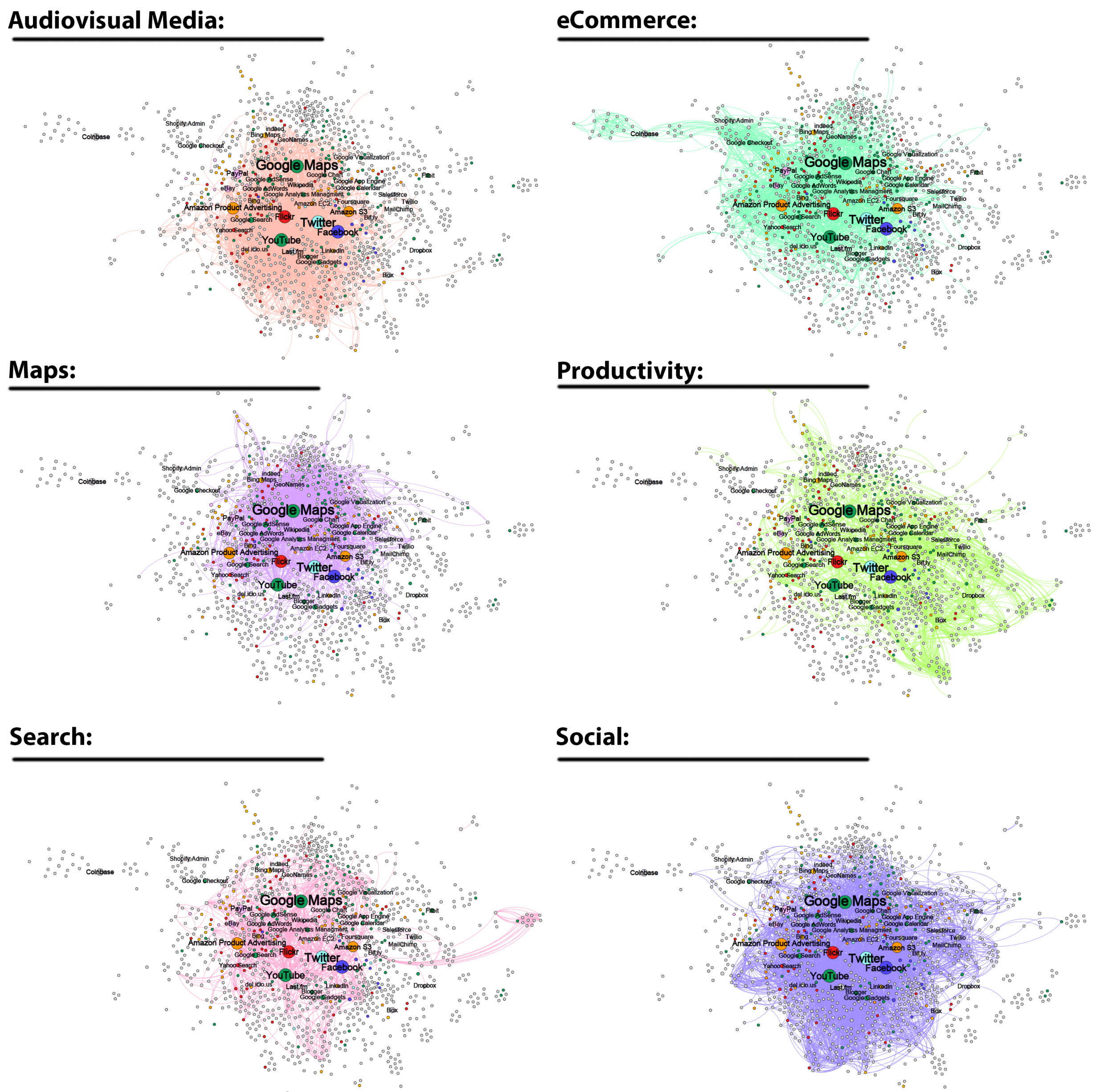

Figure 3: API subnetworks, with only edges of a certain type highlighted. Five notable features of the subnetworks include: (1) Some APIs are highly central to some subsets of the network, despite being of low degree. For example, the Coinbase API is highly central in the eCommerce subnetwork, despite having low degree (2) Dropbox, Box, Salesforce, and Amazon S3 are are more central in the 'productivity' subnetwork (3) Google Maps is central in many subnetworks, but especially the Maps subnetwork (4) Also important in the maps subnetwork is GeoNames, one of the most important non-profit supported APIs (5) Facebook, Twitter, and YouTube are especially central to the social media subnetwork, but they are also central in almost all subnetworks.

another view on how each company fits into the API network. Microsoft's and eBay's nodes are disproportionately located in the top left corner of the network, connected to each other and the cluster of eCommerce oriented APIs. Facebook's nodes are clustered in the bottom right of the figure, located in the heart of the social media sub-graph but also closer to the productivity focused portion of the graph. Apple, despite its huge success as a technology company, is relatively poorly represented. This is perhaps due to the closed nature of Apple's technological ecosystem.

Appendix Figures A4 through A16 visualize the growth of the API network over time. Network density increases substantially in the late 2000s and early 2010s, a period of time when the ProgrammableWeb crowdsourcing was most comprehensive. Note also the early centrality of Flickr, an important early image hosting website. While Flickr has since fallen on hard times (supplanted by Imgur and other close substitutes), it remains central to the API network as we measure it. This occurs because we do not observe deprecated APIs in our data when apps stop using them. Importantly, our measure of the API network at any point in time is cumulative and somewhat backward looking for this reason.

5 Model: Market Value Changes among API Adopters

As shown in Section 4, firms with top APIs have seen tremendous increases in market value over the last fifteen years. This section applies two-way fixed effect, difference-in-difference, and synthetic control approaches to estimate the impact of API adoption on a firm's market value.

We begin by estimating specification (1)

$$\log \text{Market Value}_{i,t} = \beta \cdot \text{API}_{i,t} + \alpha_i + \gamma_t + \epsilon_{i,t} \tag{1}$$

where 'API' is an indicator for whether firm i in period t has an operating API and α and γ correspond to firm and quarter fixed effects. We evaluate (1) for various subsets of public firms. In these specifications we are focused on firms with public APIs, so we use the first date a firm's APIs are submitted to ProgrammableWeb to proxy when the firm initiated a public API strategy.

Table 2 reports the coefficient on Post-API adoption using this specification. As can be seen, API adoption is associated with large increases in market value in all subsets. In the full sample, it is associated with a 38.7% increase in market value. One potential concern with this estimate is that API usage is concentrated in industries, such as communications, computer services, and online retail, where there has been strong market value growth overall. Therefore, columns two and three of Table 2 reports the same specification, but restricting attention to firms in industries with high levels of API usage. Similarly, column four omits all observations

	All Firms	$\geq$ 1% SIC Treated	$\geq$ 5% SIC Treated	Excluding Computer Services Firms	Year API < 2012	Year API $\geq$ 2012
Post x API	0.387***	0.377***	0.516***	0.322***	0.750***	0.260**
	(0.0855)	(0.0856)	(0.100)	(0.0971)	(0.140)	(0.0997)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
R2 Adjusted	0.932	0.932	0.928	0.934	0.929	0.929
Obs	133202	127796	55416	119201	129670	130040
Firms	4647	4478	2140	4060	4556	4561
API Adopters	177	176	140	100	86	91

Notes: Standard errors in parentheses and clustered at the firm level. Outcome variable is the log market value of firm. *Post*x*API* is a binary variable that equals one if a given firm has a public API operating on a given date. + p < 0.10 * p<0.05 ** p<0.01 *** p<0.001

Table 2: Two way fixed effect estimations of the effect of API adoption on log Market Value following equation (1). Column one includes the entire dataset, while the subsequent columns restrict the regressions to various subsets of the data. Table A2 reports the decomposition of the column one estimate into the sources of identification.

of firms (both API adopting and non-adopting) in the computer services industry.

An important concern about difference-in-difference estimates of this form is that if API adoption has an effect on market value growth *rates* rather than *levels*, the estimate of the effect of API adoption will be highly sensitive to the length of the sample. Indeed, columns 5 and 6 of Table 2 show that early adopters of APIs have a greater increase in their market value in the post-adoption period, suggesting the positive effect of API adoption grows with time. Still, most of the effect we identify from API adoption is coming from across-firm decisions, rather than within firm timing. Appendix Figure A2 reports a Bacon decomposition of our column 1 estimate, and finds that 95% of our effect is identified from different decisions to adopt across firms. Hypotheses **H0** and **H1** are firmly rejected.

Because API adoption seems to have an effect on market value growth rates rather than levels, it makes sense to re-analyze our results, separately estimating the effect of API adoption by number of periods since the 'treatment' began. This approach also lets us analyze whether there are pre-trends in the data. We therefore estimate specification (2)

$$\log \text{Market Value}_{i,t} = \sum_{y} \beta_y \text{API}_{i,y,t} + \alpha_i + \gamma_t + \epsilon_{i,t} \tag{2}$$

where y corresponds to the number of periods before or after a firm started using APIs. While in our regression specification we include all leads and lags y for all observed quarters before and after API adoption, here we report only coefficients for the eight quarters immediately

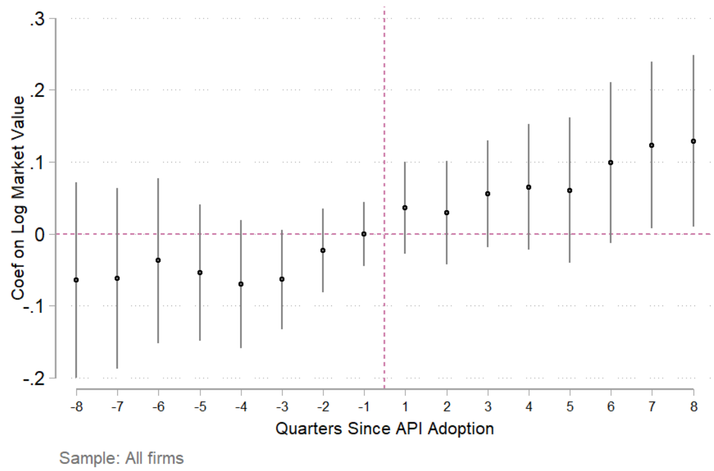

Figure 4: Treatment effect of public API adoption on log market value by quarters since API adoption with 95% confidence intervals. Standard errors clustered at the firm level. Appendix Table A3 reports these same estimates in table form.

preceding and post-API adoption. Figure 4 reports these estimates along with 95% confidence intervals.

Figure 4 shows that firms adopting APIs saw elevated market value growth beginning soon after adoption and significant growth seven periods, or 1.75 years, after ProgrammableWeb received their first API. Eight quarters after adoption, firms have 12.9% higher market values, a very considerable effect. Splitting the sample into firms with at least one B2C oriented API vs. those with B2B APIs, as appendix Figure A17 does, shows that the effect is driven by B2C oriented firms. This is consistent with our finding above that the most important and central APIs tend to be B2C or 'both' oriented (e.g. all APIs in the top 5 by betweenenss centrality in Table A2, are B2C or 'both' oriented). This confirms Hypothesis **H4**.

There is some slight visual evidence of a pre-trend in API adoption beginning half a year before the API announcement date. We believe that this is due to anticipatory market value effects (the stock market can bid up the price of a company before a new technology is implemented) and that some APIs are only posted to ProgrammableWeb after a lag. A lag in posting is certainly consistent with ProgrammableWeb's nature as a crowdsourced dataset. Developers who can take early advantage of a new API may be in the best position to post on ProgrammableWeb after using them. Their private knowledge might motivate these individuals to push information about the new API after they had a chance to exploit that knowledge (Hirshleifer, 1978). Still, this result may lead to concern that our analysis faces a reverse causality problem – in other words, that market value growth causes API adoption rather than vice versa.

To address this concern, we conduct a synthetic control analysis of API adoption. Synthetic control analysis creates composite firms of API non-adopters with the same pre-API adoption

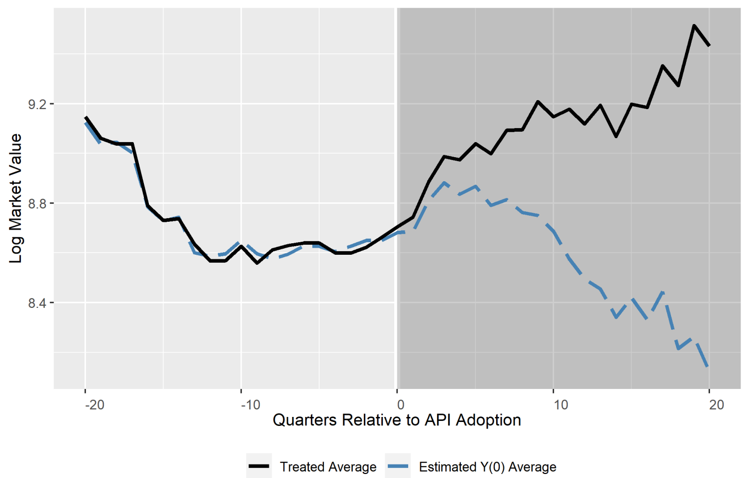

Figure 5: Average market values for API adopting firms and a synthetic control group balanced to match in the twenty quarters before and after adoption. Implied gap is \$8.4B 20 quarters after adoption.

market value growth trend as adopters. If the API adopters and synthetic non-API adopting firm have different outcomes post adoption, then the differential is plausibly attributed to adoption itself and not reverse causality.

Figure 5 reports average log market values for API adopting firms and a composite of synthetic controls for the twenty periods before and after API adoption. The synthetic control was constructed following Xu (2017)'s generalized synthetic control procedure, based on data for API adopters on only the eight quarters prior to adoption. Despite this, both the treated average and synthetic control are virtually identical for the twenty periods prior to adoption. Both the average adopter and synthetic control firm see an increase in market value from about two periods prior to adoption to two periods post adoption. However, after that point, the market value of adopting firms continues to grow rapidly while the synthetic control firms see a large decrease in market capitalization. Because the non-adopters see a decrease in market value, this is consistent with a business stealing effect to the detriment of non-adopters. A further managerial interpretation is that investing in API networks might provide early mover advantage.

Table 3 reports the point estimate and confidence interval for the effect of API adoption, again using generalized synthetic control following Xu (2017). This result should be contrasted with the basic difference in difference result in column 1 of Table 2. The point estimate of the effect is larger than in the baseline estimate and significant at the 5% level.

With evidence in hand that API adopting firms outperform non-adopters, we proceed to investigating the importance of different proposed mechanisms for APIs' positive impact.

Average Treatment on Treated	Std. Err.	CI lower	CI upper	p-value
0.729	0.334	0.075	1.384	0.029

Table 3: Estimated average treatment effect and confidence interval using generalized synthetic control. Appendix figure A18 reports confidence intervals for the treatment effect in each quarter.

6 Why API Adoption Matters: Firm Inversion or Internal Effects?

In our hypotheses and literature review, we pointed to two main classes of mechanisms by which API adoption might help firms. Hypothesis **H3** is "inverting the firm." That is, APIs enable third parties to make complements to the firms' products and services, indirectly boosting the firm's profitability. Hypothesis **H2** is "internal productivity," or that APIs help firms by enhancing internal business processes. This section examines evidence for both hypotheses.

6.1 Evidence for Firm Inversion

If APIs benefit firms primarily through enabling third party complementors, then the extent to which APIs benefit a particular firm should be a function of that firms' APIs level of engagement from the developer community. Appendix Table A4 reports estimates of variations on regression specification (1). These specifications vary from those of Table 2 because in addition to the extensive 'post-API' coefficient we also estimate intensive measures of the API's engagement with a community of third parties.

The three measures of third-party engagement considered in Table A4 are follower count, developer count, and number of API updates. Follower count is a firm level sum of the number of followers of that firms' APIs. Following an API allows a ProgrammableWeb user to easily track updates to those APIs, and is therefore a self reported measure of a ProgrammableWeb user's interest in that API. Likewise, number of developers tracks the interest of self reported developers. Note that it is not uncommon for ProgrammableWeb users to be both followers and developers of the same API. Finally, 'change count' reports the total number updates the firm has made to all of its APIs. It is a firm-side measure of third-party engagement, because APIs are more likely to be updated if they are being used, if third-parties request them, or malicious users force them.

Almost all specifications show the intensity of engagement to be significantly correlated with market value growth, over and above the extensive margin of API adoption. In our basic specification, with firm and quarter fixed effects and no other restrictions, an additional API developer is associated with a 1.75% additional increase in market value. In a parallel specification, an additional API follower is associated with a .13% increase in market value. For context, recall that

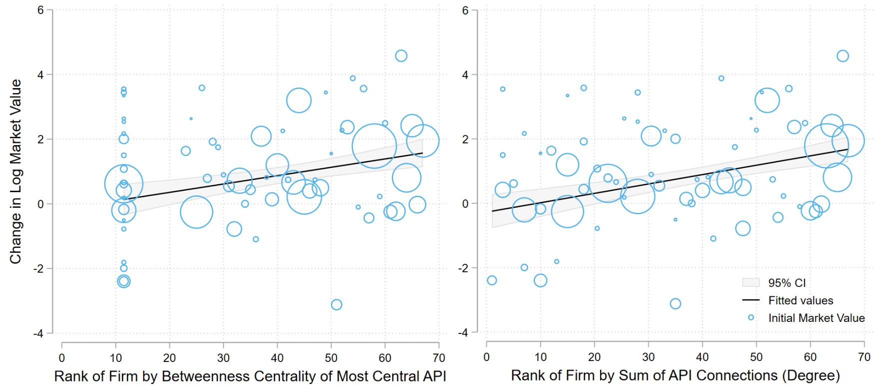

Figure 6: Scatter plots and linear fits, with 95% confidence intervals, of percentage growth in market value on a firm's network rank. Marker sizes are proportional to firms' initial market values. On the left, a firm's network importance is measured by the betweenness centrality of its most central API. There is bunching at this figure at rank 11, as there are 11 firms who have APIs that connect to no more that one other API, and therefore are tied for lowest possible betweenness centrality. On the right, firm importance is measured by the sume of a firm's API degrees. Log market value growth is measured from the date the firm first appeared in Compustat. Firm rank measured for 67 firms as of Q3 2020. Higher rank indicates greater importance. Best fit regressions lines reported in appendix Table A5.

the average number of developers and followers per firm are 20.45 and 3.03 (with s.d.s of 92.1 and 390.5). The managerial implications for labor are large. An increase in outside programmer interest, of a magnitude generating one more self-reported developer on ProgrammableWeb, is associated with an average increase in market value of $4.52 million. This implies that managers need methods to recruit and support outside expertise. Hypothesis **H3** is firmly supported.

Under the firm inversion hypothesis, more intense external use of a firm's resources predicts more value in the API network. APIs that are more frequently integrated into apps are more likely to benefit the API hosting firm. Firms with APIs that are more central to the API network may be more successful as well.

Figure 6 plots percentage growth in platform value as a function of the firm's rank in number of connections in the API network (i.e. the sum of a firm's API's degrees in Figure 2) as well as by the centrality of their most central API in the most recent API network. The figure restricts attention to the 67 firms that have data available for Q3 2020 and have at least one API which is connected to another API.

As can be seen, there is a significant positive relationship between both measures of network importance and market value growth. The effect is large, and approximately the same for both network importance measures. The magnitude is such that a 50 percentile increase in firm rank (e.g. from 25th percentile, at rank 17, to 75th percentile, at rank 50) is associated with about a 90% increase in market value.

According to the 'invert the firm' hypothesis, the nature of a firm's connections are just as important as their abundance. If an API is strategically placed in an 'information bottleneck' this may benefit the API creating firm. APIs with high centrality, especially betweenness centrality, play a more important role in connecting the services of firms that would otherwise not be incorporated to the larger internet economy. As appendix Table A2 shows, API degree and centrality are tightly related. Still there are APIs that 'punch above their weight'. A good example is Coinbase, which has relatively few connections to other APIs (24) but is the 10th most central platform overall, because it is the key API connecting many cryptocurrency APIs to online sales and shopping APIs.

One limitation of the regression described in Figure 6 and Table A5 is that latent firm characteristics may drive both API importance and firm growth. If this latent characteristic is static, it can account for using a two way fixed effect model. Therefore, to further examine the role of network importance on firm outcomes, Table 4 regresses log market value against different measures of the firm's network centrality, with firm and quarter fixed effects. Only public firms with APIs connected by apps to the main API network graph are included in the regressions. For ease of interpretation, all firm-level API importance measures are normalized to have a mean of zero and standard deviation of one.

As can be seen, the sum and maximum of firms' APIs' betweenness centrality, degree, and effective network size are significantly associated with higher log market values. We fail to find evidence that the average importance of a firm's nodes lead to increased market value.

Coefficients in Table 4 can be interpreted as the change in log market value due to a one standard deviation change in the network statistic. A one standard deviation increase in the total number of connections of a firm's APIs is associated with a 16.2% increase in market value. We see a similar magnitude in the coefficient on sum of effective network sizes, which penalizes nodes that are redundant to the network (i.e. connect to nodes that are already connected to each other).

While the characteristics of network importance that aggregate over all a firm's APIs are significantly associated with market value, market value is even more elastic to the importance of a firm's most important API. A firms' log market value increases by more than 30% when it's best API increases its effective size or degree by an order of magnitude. On the other hand, the average importance of a firms' nodes are not significantly associated with larger market values (although point estimates are positive). Together, these facts support Hypothesis **H5** but also may be taken as evidence of a superstar effect. Among APIs, a small subset of extremely important APIs drive most value rather than an array of moderately successful ones. Managerially, this also suggests a cumulative effect of opening APIs early.

Even though we still estimate a significant positive relationship between measures of a firm's importance to the API network and log market value, the effect is noisier and

	1 Log Market Value	2 Log Market Value	3 Log Market Value	4 Log Market Value	5 Log Market Value	6 Log Market Value	7 Log Market Value	8 Log Market Value
Mean Betweenness Centrality	0.210 (0.318)							
Max Betweenness Centrality		0.148^{*} (0.0714)						
Mean Degree			0.402 (0.305)					
Max Degree				0.366^{*} (0.169)				
Sum of Degrees					0.162^{*} (0.0698)			
Mean Effective Network Size						0.188 (0.272)		
Max Effective Network Size							0.313^{*} (0.149)	
Sum Effective Network Size								0.166^{*} (0.0767)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R2	0.365	0.370	0.369	0.377	0.370	0.364	0.375	0.371
Obs	2287	2287	2287	2287	2287	2287	2287	2287
Firms	66	66	66	66	66	66	66	66
Apps Connections	19893	19893	19893	19893	19893	19893	19893	19893

Notes: Standard errors in parentheses, clustered at firm level. Outcome variable log firm market value. + + p<0.10 * p<0.05 ** p<0.01 *** p<0.001

Table 4: Effect of API network measures on firm market value. Standard errors in parentheses and clustered at the firm level. Outcome variable is log market value. Firm and quarter fixed effects included in all specifications. All explanatory variables normalized to mean zero, with a standard deviation of one. The explanatory variable in columns 1, 4, and 6 is the average of a firm's API importance measures – betweenness centrality, degree, and effective size respectively. The explanatory variable in columns 2, 3 and 7 are the maximum importance value taken by any of a firms' APIs in that period, over the same list of importance measures. The explanatory variable in column 8 is the sum of the firm's API's effective network size Burt (1992).

smaller in this panel regression.[22] One reason for this is that rank ordering in the API network does not change much over time. Appendix Figure A21 displays the sum of degree over time for the APIs of five selected firms. Across all five firms, rank increases steadily as new APIs enter until around 2015 and then plateaus, coinciding with the decline of ProgrammableWeb as a comprehensive directory. The relative rank of all five firms hardly changes, excepting Amazon, which inches past Verizon. The lack of variation in firm network rank over time is a challenge for estimation in a two-way fixed effects model. It is also further evidence of a strong first-mover advantage.

6.2 Evidence on Internal Productivity

To distinguish between adopting a public vs. private API, we draw on our second API usage dataset – one from a private API tool provision company. Of the 78 firms who deployed APIs using tools from this company, only 44 are listed as having public APIs available at any point on ProgrammableWeb. Therefore, we can measure the effect of internal APIs by focusing on the effect of API adoption among the remainder. In this data, we measure the date of API adoption as the first date we observe the firm with non-zero data flows through one of their APIs.

Figure 7 reports estimates of the effect, over time, of API adoption on log market value for firms adopting purely internal APIs. The specification used is equation (2), and as in Figure 4 above, while all leads and lags are included in the estimation, only estimates for quarters within two years of adoption are displayed.

Using specification (2), Figure 7 shows there is no clear effect of internal API adoption on firm market value. As an alternative specification we use generalized synthetic controls as above. Again, we fail to find evidence of a positive effect of API adoption, as shown in the counterfactual plot Figure A19 or the synthetic control difference-in-difference estimate Table 5. This test of **H2** shows low confidence in an effect. However, the confidence interval is wide, and consistent with a moderate or even large positive effect.

An alternate mechanism by which APIs are said to boost firms internally is through

[22]To make the estimates of Table 4 directly comparable to those of Figure 6, note that about 1.35 standard deviations separate the 75th and 25th percentile. Therefore, in these regressions, moving from the 25th to 75th percentile in degree sum at the firm level is associated with roughly a 22% increase in market value, compared to about 90% in the pooled, non-panel specification.

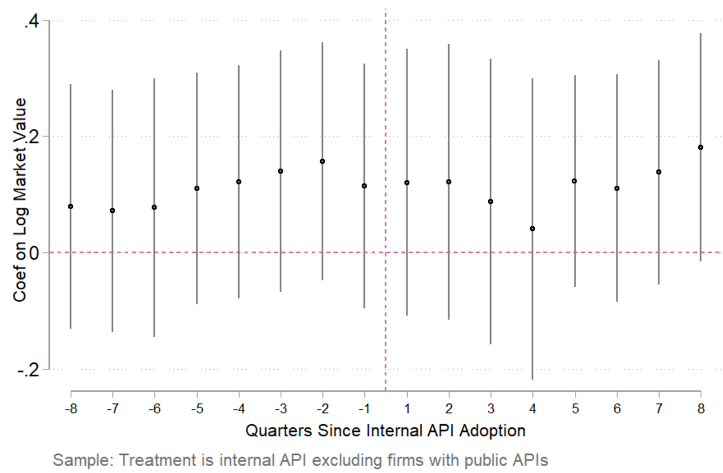

Figure 7: Estimates and 95% confidence intervals for the effect of internal API adoption by number of periods before and after adoption. Equation follows 2, and is comparable to Figure 4 but for internal APIs instead of public.

Average Treatment on Treated	Std. Err.	CI lower	CI upper	p-value
-.055	.426	-0.89	0.781	0.898

Table 5: Estimated effect, and confidence interval, of adopting an internal API using generalized synthetic control following (Xu, 2017).

reducing adjustment costs. This would be consistent with increased dynamic capabilities or options value from remixed resources (Teece and Pisano, 2003; Baldwin and Clark, 2006). If APIs allow firms to more easily integrate new resources or reconfigure old ones, firms should be able to make and capitalize on investments more quickly. This should lead profitable (at the margin) firms to make more investments, boosting their market capitalization. Alternatively, if APIs primarily benefit firms through ‘firm inversion’, the firm itself will not need to make major capital investments in order to grow. Third parties would make them. A typical approach to measuring whether a firm’s investment is limited by capital adjustment costs is Tobin’s Q (Tobin, 1969), the ratio of market capitalization to assets.

To test which theory best explains the growth in market capitalization for API using firms, we run a set of regressions analogous to specification (1) with the addition of log of firm assets. Essentially this means we now estimate the effect of API adoption on Q (log(Q) to be precise). As Table 6 shows, across specifications, API adoption positively predicts market value after controlling for total assets. In the base speci-

	All Firms	≥ 1% SIC Treated	≥ 5% SIC Treated	Excluding Computer Services Firms	Year API < 2012	Year API ≥ 2012
Log of Total Firm Asset	0.738***	0.738***	0.724***	0.747***	0.736***	0.736***
	(0.0164)	(0.0166)	(0.0227)	(0.0184)	(0.0167)	(0.0168)
Post x API	0.135*	0.123*	0.254***	0.156*	0.305***	0.0744
	(0.0607)	(0.0608)	(0.0689)	(0.0716)	(0.0912)	(0.0744)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
R2 Adjusted	0.951	0.951	0.947	0.952	0.948	0.948
Obs	132934	127528	55354	118936	129402	129772
Firms	4645	4476	2138	4058	4554	4559
API Adopters	177	176	140	100	86	91

Notes: Notes: Standard errors in parentheses clustered at firm level. Outcome variable log market value of firm. *Post x API* is a binary variable that equals one if a given firm has a public API operating on a given date. + p<0.10 * p<¡0.05 ** p<0.01 *** p<0.001 + p < 0.10 * p<0.05 ** p<0.01 *** p<0.001

Table 6: Difference in Difference Impact of API Controlling for Firm Assets.

fication, paralleling Table 2 column 1, the effect of API adoption is roughly cut to a third after controlling for growth in assets. This means that while some of the effect of API adoption on market value is mediated by added asset investments, API adoption still increases Q, consistent benefits of API adoption stemming from factors outside the firm. As external value creation explains more market value than internal value creation, managerial attention should shift to harnessing resources outside the firm.

7 API Exposure: Security Challenges & Responses

The usefulness of APIs depends on how well they balance trade-offs. An API is a kind of aperture or membrane that selects which information to diffuse in and out. Too wide an aperture and the firm may give away its data assets. Too narrow or difficult to access and outsiders will struggle to meaningfully engage. As noted above, firms that update their APIs more frequently see larger increases in market value (see appendix Table A4), consistent with the idea that managing details of third party API use is critical.

One dominant decision for trafficking in data is how to defend against data breaches. If APIs increased the risk of major loss or liability, their use would pose an important downside risk. There is a trade-off between an interest in enabling third party innova-

	Any Breach Event	Breach of Credit Card Info	Breach via Malicious Hack	Breach via Stolen Document or Fixed Computer	Breach via Portable Computer	Breach via Malicious Insider	Log Count of Records Exposed
0-2 years post API adoption	1.086	1.482	0.854	5.526+	0.772	6.852**	1.004
	[0.28]	[0.33]	[-0.33]	[1.66]	[-0.22]	[2.63]	[-0.06]
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pseudo R2	0.0887	0.332	0.181	0.355	0.205	0.253	
Log Likelihood	-581.5	-41.21	-165.9	-32.33	-94.29	-104.6	-23811.1
Obs	3878	445	1522	386	1054	987	91946
Event Count	221	19	63	15	35	44	95
R2 Adjusted							0.000151

Notes: T-statistics in parentheses. Exponentiated coefficients presented. Outcome variable is a binary indicator of whether specific type of breach event occurred (first six columns) or log of total amount of records breached (final column). Event count refers to number of distinct breach events of a given type. Last column estimated via panel linear regression. + p < 0.10 * p<0.05 ** p<0.01 *** p<0.001

Table 7: Fixed effect logistic (first six columns) or linear regression (final column) of impact of API adoption on breach events or log total records exposed. *Breach by Malicious Insider* is often misuse of an authorized API key by a malicious unauthorized actor.

tions and an interest in thwarting third party damage or ransom. Opening APIs can have both effects. The trade-off depends in part on the relative mix of benevolent and malicious outsiders, which is hidden information. Ransbotham (2016) has shown this "Paradox of Exposure" to be present in the context of open-source software. This risk is particularly notable given evidence that executives of companies who experience data breaches face negative personal consequences (Kamiya et al., 2018). Even if the ratio of risk to reward is favorable, risk aversion or personal costs to executives may limit investment in API projects. [23] Table 7 reports an increased risk of data breach by insiders in the two years post adoption. Relevant for APIs, this may represent stolen or forged credentials for authorized API keys. Data loss based on physical documents or portable computers show little or no significance.

To explore how firms respond to data breaches, appendix Figure A22 takes advantage of the fact that we observe data flows in our proprietary API data to see how firms respond to data breach events. This data includes all 78 firms who work with the API tool developer, including the 44 which have public APIs. All firms' APIs, in this dataset,

[23] Other hypothetical instances of data 'overexposure', such as intentionally giving away data that later turns out to be key to a firm's competitive advantage are also possible but beyond the scope of the current paper

were classified by purpose based on their names (see section B).

Figure A22 shows that firms who report data breaches see a decrease in API flows in the short run that rebounds over time. The API type that sees the largest reduction after a hack is internal communications, perhaps indicating firms' hesitance to use internal channels after a data breach. On the other hand, the data flows for testing APIs increase dramatically in the months after a data breach is reported. This is consistent with firms taking steps to reduce adverse API exposure in the wake of an unexpected breach.

Substantial differences in API practice separate good firms from bad. Security is challenging and even good programmers exhibit blind spots in coding practice (Oliveira et al., 2018). API practices that distinguish successful from unsuccessful firms include (i) rate limiting data queries and throttling them when rates are exceeded (ii) time limiting queries to curb copycat requests (iii) using well-established standards in preference to custom built (iv) separating the API access tokens name and password credentials (v) never storing plaintext credentials, and (vi) two-factor authentication (Lamba, 2019). Editors at ProgrammableWeb have observed firms shifting from Larger Numbers of Unknown Developers (LNUD) to Smaller Numbers of Known Developers (SNKD).[24] In practice, this strategy balances the benefits of inverting the firm and securing systems from breach.

8 Conclusion

This paper investigates the role of APIs in driving the growth of firm market value. APIs let firms grant insiders and outsiders access to firm resources. Externally, this helps third parties build apps. Internally, APIs help business units share data and reduce adjustment costs within the firm. Clearly, APIs have a positive impact on the third parties who use them in their apps. But is exposing so much of one's internal data and processes to the world good for the API deploying firm itself? We investigate whether it is, and, if so, why and how, as well as assess the risk.

Using data on public APIs from ProgrammableWeb, we first visualized the network of apps calling APIs. We identified the disproportionate role that some APIs play in bridging different clusters. APIs with some of the highest importance, as measured

[24]https://www.programmableweb.com/news/do-you-publish-api-docs-online-availability-to-third-party-developers/analysis/2016/02/18

by betweenness centrality and effective network size, include not only Google Maps, Twitter, YouTube and Facebook but also Shopify, Coinbase, and Dropbox. Firms with successful APIs saw tremendous market value growth. The fourteen publicly traded firms with APIs in the top forty by betweenness centrality saw their total market value increase by $6.6 Trillion dollars from 2005 through 2021, representing a sizable share of total appreciation in the US equity market over that period.

To confirm the role of public APIs in boosting market value, we ran a series of analyses. A difference-in-difference model showed that API adopting firms saw their market value increase by 38% over subsequent quarters, roughly 2% per year. In an event study analysis, with leads and lags, we find the size of the effect grows with the length of time since API adoption. This is consistent with APIs growing in utility as more complementors develop their products. It is also consistent with a first-mover advantage in establishing an API network. While we find some evidence of a pre-trend in market value for API adopting firms, a synthetic control analysis accounts for potential reverse causality (i.e. market value growth causing API adoption) and shows an even larger, still significant effect (of 73%).

We then investigate to what extent the success of API adopting firms is due to enabling third party complementors – the 'inverted firm' hypothesis – or due to internal effects. We first show that, within the subset of firms adopting APIs, those with greater third party engagement, as measured by updates, followers and developers, see greater gains in market value. The same holds for firms with more central and important APIs, a result shown both in pooled and two-way fixed effect panel specifications. Because firms do not have direct control over how much third-party engagement their apps achieve, this is strong evidence that firm-inversion plays an essential role in the positive impact of APIs. These gains increase over time, suggesting a first-mover advantage.

To investigate the alternate thesis that internal productivity gains explain these effects, we use proprietary data from an API tool provision company to replicate our analysis for private APIs. We fail to find evidence of a direct market value effect from internal APIs. That said, those estimates have large confidence intervals, and are consistent with a moderate or even large positive effect. We also test the hypothesis that APIs help firms internally by lowering their capital adjustment costs, which would tend to lower their Q. In a specification controlling for a firm's total assets, we find that

Q rises and there is still a positive effect of API adoption on firm value. This effect is attenuated, however, indicating that some gains derive from internal capital adjustment but much derives from third party use of those assets.

Finally, we investigate one major downside of API adoption – a greater risk of data breach. Panel fixed-effect logistic regressions show an increased risk of data breach in the two years following API adoption. This type of breach can be connected to poorly secured APIs. Further, there are clear adjustments in behavior in the wake of a data breach among those firms we can observe. Consistent with an API a role in these events, firms in the months after a hack decrease their internal API communications APIs and increase their API testing.

Together these results show, quantitatively, that APIs are a critical aspect of the economy's growing digital ecosystem. Firms that use APIs to place themselves at the center of this ecosystem can expect large returns.

References

Aral, Sinan and Marshall Van Alstyne, "The Diversity Bandwidth Trade-off," *American Journal of Sociology*, 2011, *117* (1), 90–171.

Baker, Wayne E, "Market networks and corporate behavior," *American journal of sociology*, 1990, *96* (3), 589–625.

Baldwin, C. and K. Clark, *Design rules: The power of modularity*, Vol. 1, The MIT Press, 2000.

Baldwin, Carliss Y and Kim B Clark, "The architecture of participation: Does code architecture mitigate free riding in the open source development model?," *Management Science*, 2006, *52* (7), 1116–1127.

Benzell, Seth G and Erik Brynjolfsson, "Digital Abundance and Scarce Genius: Implications for Wages, Interest Rates, and Growth," 2019.

Borgatti, Stephen P, "Centrality and network flow," *Social networks*, 2005, *27* (1), 55–71.

Boudreau, K., "Open Platform Strategies and Innovation: Granting Access versus Devolving Control," *Management Science*, 2010, *56* (10), 1849–1872.

Boulton, Clint, "Insurance spin-out rides API-driven strategy," *CIO Magazine*, February 8 2017.

Brynjolfsson, Erik and Lorin M Hitt, "Beyond Computation: Information Technology, Organizational Transformation and Business Performance," *Journal of Economic Perspectives—Volume*, 2000, *14* (4), 23–48.

_ **, Daniel Rock, and Chad Syverson**, "The productivity J-curve: How intangibles complement general purpose technologies," *NBER Working Paper*, 2018.

_ **, Lorin M Hitt, and Shinkyu Yang**, "Intangible assets: Computers and organizational capital," *Brookings papers on economic activity*, 2002, *2002* (1), 137–181.

Burt, Ronald S, *Structural holes*, Harvard university press, 1992.

_ , *Structural holes: The social structure of competition*, Harvard university press, 2009.

Cabral, Luís, Justus Haucap, Geoffrey Parker, Georgios Petropoulos, Tommaso M Valletti, and Marshall W Van Alstyne, "The EU Digital Markets Act: A Report from a Panel of Economic Experts," *Cabral, L., Haucap, J., Parker, G., Petropoulos, G., Valletti, T., and Van Alstyne, M., The EU Digital Markets Act, Publications Office of the European Union, Luxembourg*, 2021.

Chesbrough, Henry and Marshall Van Alstyne, "Permissionless innovation," *Communications of the ACM*, 2015, *58* (8), 24–26.

Diestel, Reinhard, "Graph theory 3rd ed," *Graduate texts in mathematics*, 2005, *173.*

Eisenhardt, Kathleen M and Jeffrey A Martin, "Dynamic capabilities: what are they?," *Strategic management journal*, 2000, *21* (10-11), 1105–1121.

Evans, Peter C and Rahul C Basole, "Revealing the API ecosystem and enterprise strategy via visual analytics," *Communications of the ACM*, 2016, *59* (2), 26–28.

Fogel, Robert William, "A quantitative approach to the study of railroads in American economic growth: a report of some preliminary findings," *The Journal of Economic History*, 1962, *22* (2), 163–197.

Granovetter, Mark S, "The strength of weak ties," *American journal of sociology*, 1973, pp. 1360–1380.

Hansen, Morten T, "The search-transfer problem: The role of weak ties in sharing knowledge across organization subunits," *Administrative science quarterly*, 1999, *44* (1), 82–111.

Haranis, Mark, "Cisco Mounting Massive Open API Offensive," *CRN*, February 17 2017.

Hirshleifer, Jack, "The private and social value of information and the reward to inventive activity," in "Uncertainty in economics," Elsevier, 1978, pp. 541–556.

Iyer, B. and M. Subramanian, "The Strategic Value of APIs," January 2015.

Iyer, Bala and John C Henderson, "Preparing for the Future: Understanding the Seven Capabilities of Cloud Computing," *MIS Quarterly Executive*, 2010, *9* (2), 117–131.

Jacobson, Daniel, Greg Brail, and Dan Woods, *APIs: A strategy guide*, O'Reilly Media, Inc., 2011.

Joseph, Sam, Vivian Ludford, and Bart McAllister, "Plugging In: Enabling the Enterprise for the Platform Economy," Technical Report, Gartner Research Board September 2016.

Kamiya, Shinichi, Jun-Koo Kang, Jungmin Kim, Andreas Milidonis, and René M Stulz, "What is the Impact of Successful Cyberattacks on Target Firms?," 2018.

Katz, Michael L and Carl Shapiro, "Systems competition and network effects," *Journal of economic perspectives*, 1994, *8* (2), 93–115.

Khan, Lina, "Amazon's Antitrust Paradox," *Yale Law Journal*, 2017, *126* (3), 710–805.

Lamba, Anil, "API Design Principles & Security Best Practices," *Cybernomics*, August 2019, *1* (3), 21–25.

Lusch, Robert F and Satish Nambisan, "Service innovation: A service-dominant logic perspective.," *MIS quarterly*, 2015, *39* (1).

Makridis, Christos and Benjamin Dean, "Measuring the Economic Effects of Data Breaches on Firm Outcomes: Challenges and Opportunities," *Available at SSRN 3044726*, 2018.

Malone, Thomas W, Joanne Yates, and Robert I Benjamin, "Electronic markets and electronic hierarchies," *Communications of the ACM*, 1987, *30* (6), 484–498.

Moreno, Jacob L and Helen H Jennings, "Statistics of social configurations," *Sociometry*, 1938, pp. 342–374.

Newman, Lily, "Google+ Exposed Data of 52.5 Million Users and Will Shut Down," *Wired*, December 10 2018.

Niculescu, Marius F, DJ Wu, and Lizhen Xu, "Strategic intellectual property sharing: Competition on an open technology platform under network effects," *Information Systems Research*, 2018, *29* (2), 498–519.

Oliveira, Daniela Seabra, Tian Lin, Muhammad Sajidur Rahman, Rad Akefirad, Donovan Ellis, Eliany Perez, Rahul Bobhate, Lois A DeLong, Justin Cappos, and Yuriy Brun, "{API} Blindspots: Why Experienced Developers Write Vulnerable Code," in "Fourteenth Symposium on Usable Privacy and Security ({SOUPS} 2018)" 2018, pp. 315–328.

Padgett, John F and Christopher K Ansell, "Robust Action and the Rise of the Medici, 1400-1434," *American journal of sociology*, 1993, *98* (6), 1259–1319.

Parker, G., M. Van Alstyne, and X. Jiang, "Platform Ecosystems: How Developers Invert the Firm," *MIS Quarterly*, 2017, *41* (1), 255–266.

Parker, Geoffrey and Marshall Van Alstyne, "Innovation, openness, and platform control," *Management Science*, 2017, *64* (7), 3015–3032.

Parker, Geoffrey G, Marshall W Van Alstyne, and Sangeet Paul Choudary, *Platform Revolution: How Networked Markets Are Transforming the Economy And How to Make Them Work for You*, New York: WW Norton & Company, 2016.

Podolny, Joel M, "Networks as the pipes and prisms of the market," *American journal of sociology*, 2001, *107* (1), 33–60.

Purvis, Russell L, Vallabh Sambamurthy, and Robert W Zmud, "The assimilation of knowledge platforms in organizations: An empirical investigation," *Organization science*, 2001, *12* (2), 117–135.

Ransbotham, Sam, "Open Source Code and the Risk of Attacks after Vulnerability Discovery," http://programme.exordo.com/oui2016/delegates/presentation/140/ 2016. Open and User Innovation Conference.

Reagans, Ray and Ezra W Zuckerman, "Networks, diversity, and productivity: The social capital of corporate R&D teams," *Organization science*, 2001, *12* (4), 502–517.

Rosati, Pierangelo and Theo Lynn, "A dataset for accounting, finance and economics research on US data breaches," *Data in Brief*, 2021, *35*, 106924.

Rowan, Rip, "Steve Yegge's Platform Rant," 2011.

Schulze, Elizabeth, "How Google, Facebook, Amazon and Apple faced EU tech antitrust rules," https://www.cnbc.com/2019/06/07/how-google-facebook-amazon-and-apple-faced-eu-tech-antitrust-rules.html 2019. CNBC.

Simmel, George, *Conflict and the web of group affiliations*, New York Free Press, repreinted 1955, 1922.

Spanos, Georgios and Lefteris Angelis, "The impact of information security events to the stock market: A systematic literature review," *Computers & Security*, 2016, *58*, 216–229.

Spring, Tom, "T-Mobile Alerts 2.3 Million Customers of Data Breach Tied to Leaky API," *ThreatPost*, August 24 2018.

Tam, Kar Yan, "The impact of information technology investments on firm performance and evaluation: evidence from newly industrialized economies," *Information Systems Research*, 1998, *9* (1), 85–98.

Teece, David and Gary Pisano, "The dynamic capabilities of firms," in "Handbook on knowledge management," Springer, 2003, pp. 195–213.

Teece, D.J., "Capturing value from technological innovation: Integration, strategic partnering, and licensing decisions," *Interfaces*, 1988, *18* (3), 46–61.

Thierer, Adam, *Permissionless Innovation: The Continuing Case for Comprehensive Technological Freedom*, Mercatus Center at George Mason University, 2016.

Tiwana, Amrit, Benn Konsynski, and Ashley A Bush, "Research commentary-Platform evolution: Coevolution of platform architecture, governance, and environmental dynamics," *Information Systems Research*, 2010, *21* (4), 675–687.

Tobin, James, "A general equilibrium approach to monetary theory," *Journal of money, credit and banking*, 1969, *1* (1), 15–29.

Uzzi, Brian, "Social structure and competition in interfirm networks: The paradox of embeddedness," *Administrative science quarterly*, 1997, pp. 35–67.

White, Douglas R and Stephen P Borgatti, "Betweenness centrality measures for directed graphs," *Social networks*, 1994, *16* (4), 335–346.

Xu, Yiqing, "Generalized synthetic control method: Causal inference with interactive fixed effects models," *Political Analysis*, 2017, *25* (1), 57–76.

Yoo, Youngin, Richard J. Boland, Kalle Lyytinen, and Ann Majchrzak, "Organizing for Innovation in the Digitized World," *Organization Science*, 2012, *23* (15), 1398–1408.

Zhu, Feng and Qihong Liu, "Competing with Complementors: An Empirical Look at Amazon. com," *Strategic Management Journal*, 2018, *39* (10), 2618–2642.

A Additional Tables and Figures

	Mean	Std Dev	Max	N (Firm Months)
Monthly Calls (Millions)	160	531	6,740	2,453
Monthly Data (Trillions of Bytes)	1.98	10.0	149	1,882
Number of APIs	31.4	46.2	433	2,453

Table A1: Total number of log calls, bytes, and APIs in proprietary API tool provision dataset. Averages by firm-month.

Diff-in-Diff Comparison	Weight	Avg Diff-in-Diff Estimate
Earlier Treated vs Later Control	0.007	-0.111
Later Treated vs Earlier Control	0.013	0.08
Treated vs Never treated	0.952	0.549
Treated vs Already Treated	0.028	0.058
Diff-in-diff estimate:	0.524	

Table A2: Bacon decomposition of baseline difference-in-difference result from equation (1) in Figure 2

	All Firms	≥ 1% SIC Treated	≥ 5% SIC Treated	Excluding Computer Services Firms	Year API < 2012	Year API ≥ 2012
8 quarters until	-0.0643	-0.0632	-0.127	0.0154	-0.512***	0.0826
API	(0.0694)	(0.0695)	(0.0880)	(0.0785)	(0.128)	(0.0794)
7 quarters until	-0.0619	-0.0617	-0.107	-0.0127	-0.407***	0.0838
API	(0.0642)	(0.0643)	(0.0807)	(0.0757)	(0.0984)	(0.0784)
6 quarters until	-0.0372	-0.0368	-0.0597	-0.00206	-0.345***	0.0929
API	(0.0585)	(0.0586)	(0.0747)	(0.0697)	(0.0965)	(0.0705)
5 quarters until	-0.0542	-0.0542	-0.0692	-0.0355	-0.243**	0.0319
API	(0.0482)	(0.0483)	(0.0588)	(0.0560)	(0.0758)	(0.0605)
4 quarters until	-0.0697	-0.0697	-0.0818	-0.0709	-0.258**	0.0201
API	(0.0453)	(0.0453)	(0.0514)	(0.0532)	(0.0894)	(0.0501)
3 quarters until	-0.0636+	-0.0632+	-0.0506	-0.0786+	-0.191**	-0.000166
API	(0.0352)	(0.0352)	(0.0392)	(0.0403)	(0.0733)	(0.0377)
2 quarters until	-0.0232	-0.0228	-0.00494	-0.0433	-0.134*	0.0294
API	(0.0294)	(0.0294)	(0.0320)	(0.0350)	(0.0658)	(0.0293)
1 quarters until	-0.000628	-0.000327	0.0116	-0.00762	-0.0659	0.0335
API	(0.0227)	(0.0227)	(0.0244)	(0.0268)	(0.0453)	(0.0244)
1 quarters since	0.0376	0.0375	0.0601	0.0291	0.0435	0.0316
API	(0.0325)	(0.0325)	(0.0396)	(0.0426)	(0.0490)	(0.0407)
2 quarters since	0.0294	0.0288	0.0650	0.0208	0.0659	0.00725
API	(0.0367)	(0.0367)	(0.0409)	(0.0477)	(0.0513)	(0.0475)
3 quarters since	0.0554	0.0543	0.104*	0.0381	0.0810	0.0401
API	(0.0376)	(0.0376)	(0.0445)	(0.0472)	(0.0536)	(0.0488)
4 quarters since	0.0652	0.0638	0.0889+	0.0530	0.0804	0.0517
API	(0.0442)	(0.0442)	(0.0493)	(0.0541)	(0.0631)	(0.0577)
5 quarters since	0.0603	0.0579	0.116*	0.0226	0.0311	0.0736
API	(0.0514)	(0.0516)	(0.0524)	(0.0644)	(0.0846)	(0.0625)
6 quarters since	0.0988+	0.0959+	0.155*	0.0533	0.157	0.0564
API	(0.0570)	(0.0571)	(0.0640)	(0.0650)	(0.0979)	(0.0660)
7 quarters since	0.123*	0.120*	0.150*	0.0793	0.193*	0.0687
API	(0.0590)	(0.0591)	(0.0675)	(0.0684)	(0.0917)	(0.0733)
8 quarters since	0.126*	0.123*	0.165*	0.0997	0.224*	0.0494
API	(0.0607)	(0.0608)	(0.0678)	(0.0706)	(0.0894)	(0.0775)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
Obs	133303	127892	55473	119285	129769	130141
Firms	4748	4574	2197	4144	4655	4662
API Adopters	179	178	141	102	86	93

Notes: Standard errors in parentheses and clustered at the firm level. Outcome variable is the log market value of firm.*Post x API* is a binary variable that equals one if a given firm has a public API operating on a given date. + $p < 0.10$ * $p<0.05$ ** $p<0.01$ *** $p<0.001$

Table A3: Table version of estimates reported in Figure 4

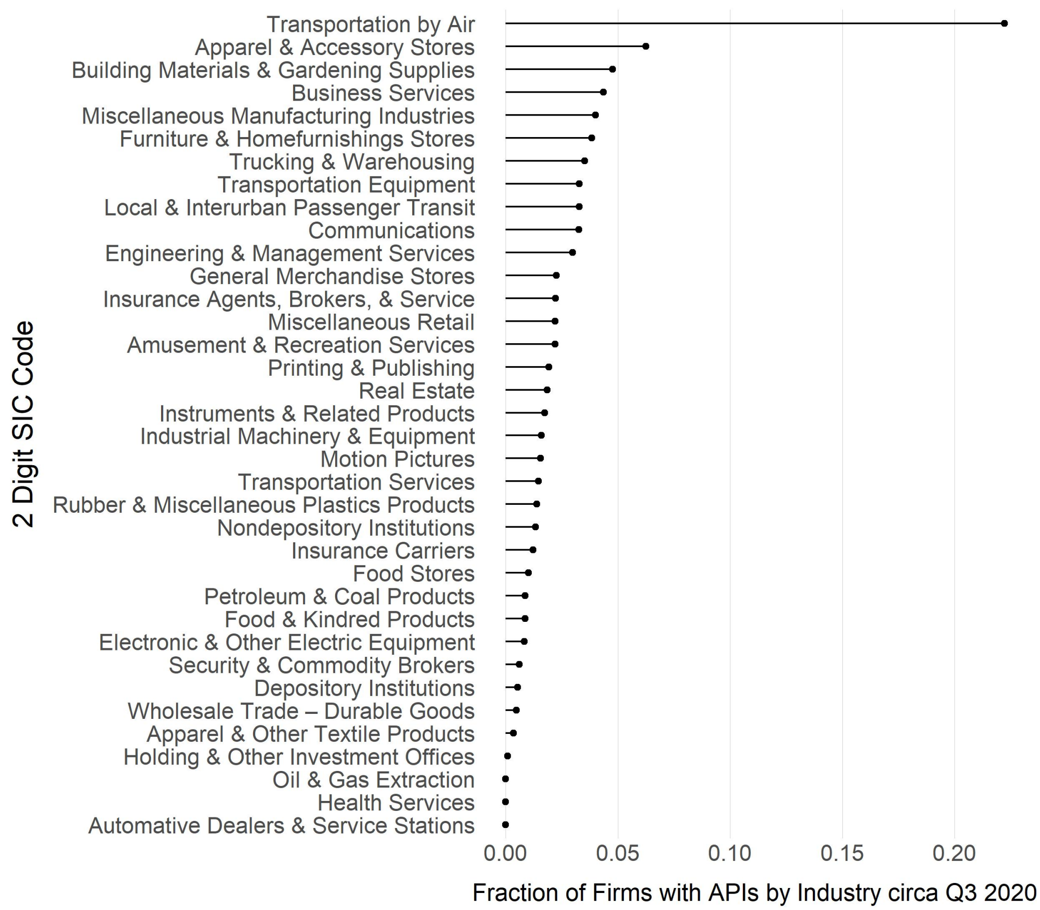

Figure A1: Fraction of firms in a given 2 digit SIC code that operate a public API. Computer services, a major function of many APIs, are included in the Business Services industry.

API Name	Company	Betweeness Centrality	Degree	Degree Rank	Market Cap in Billions June 2005 (* or at IPO)	Market Cap in Billions June 2021	Notes
Google Maps	Google	1	1982	1	363.5	1672.9	
Twitter	Twitter	2	1629	2	22.6*	54.2	IPO in 2013
YouTube	Google	3	1479	3	363.5	1672.9	acquired by google in 2006
Facebook	Facebook	4	1132	5	63.3*	967.9	IPO in 2012
Flickr	Verizon	5	1282	4	95.5	237.4	owned by Yahoo from 2005-2015; Yahoo owned by Verizon from 2017-present
Amazon Product Advertising	Amazon	6	700	6	13.6	1687.0	
Amazon S3	Amazon	7	312	13	13.6	1687.0	
PayPal	eBay	8	118	47	44.7	385.9	numbers for ebay & paypal, paypall spun off in 2015
Last.fm	ViacomCBS	9	587	7	133.4	28.8	
Coinbase	Coinbase	10	24	219	*	58.8	IPO in 2021
eBay	eBay	11	408	8	44.7	385.9	numbers for ebay & paypal, paypall spun off in 2015
Twilio	Twilio	12	266	17	3.1*	66.3	IPO in 2016
Box	Box	13	122	46	0.24*	4.1	IPO in 2015
Google AdSense	Google	14	333	11	363.5	1672.9	
Google Search	Google	15	406	9	363.5	1672.9	
Bit.ly	bitly	16	107	51			
Bing Maps	Microsoft	17	190	26	266	2000.0	
Foursquare	Foursquare	18	331	12			
Google Gadgets	Google	19	99	58	363.5	1672.9	
Bing	Microsoft	20	259	18	266	2000.0	
Google Analytics Managment	Google	21	275	15	363.5	1672.9	
del.icio.us	del.icio.us	22	394	10			private; owned by Yahoo from 2005-2011
Fitbit	Google	23	50	116	363.5	1672.9	
Blogger	Google	24	93	62	363.5	1672.9	acquired by google in 2003
Amazon EC2	Amazon	25	189	27	13.6	1687.0	
Salesforce	Salesforce	26	81	70	2.1	224.0	
GeoNames	non-profit	27	271	16			
MailChimp	MailChimp	28	45	128			
Google Visualization	Google	29	68	82	363.5	1672.9	
Shopify Admin	Shopify	30	42	135	1.8*	183.5	IPO in 2015
Google Chart	Google	31	209	22	363.5	1672.9	
indeed	Recruit Co Ltd.	32	75	75	17.7*	80.4	acquired in 2012 by Recruit Co, which IPO'd in 2015
LinkedIn	Microsoft	33	239	20	266	2000.0	
Google App Engine	Google	34	176	31	363.5	1672.9	
Google Checkout	Google	35	49	119	363.5	1672.9	
Dropbox	Dropbox	36	85	68	12.3*	11.8	IPO in 2018
Google Calendar	Google	37	94	60	363.5	1672.9	
Yahoo Search	Verizon	38	301	14	95.5	237.4	Yahoo owned by Verizon from 2017 to present
Google AdWords	Google	39	137	39	363.5	1672.9	
Wikipedia	non-profit	40	259	19			

Figure A2: This table reports degree, degree rank, and betweenness centrality rank for selected APIs. All APIs in the top 40 for betweenness centrality are displayed. It also reports the company owning the API, and the market value growth of that company since July 2015. Asterisks indicate the company only owned the API for part of the sample period. Company names are colored as in Figure 2. Degree is number of edges connecting the node to other nodes. So a app that connects my API to 4 others gives me 4 extra degree, and a second app that links to the same nodes would count twice. 'Connections to Oneself' (e.g. from apps that only call a single API) are not counted. Market value data source is Companiesmarketcap.com

25

	All Firms	All Firms	All Firms	Excl. Comp. Servs. Firms	Excl. Comp. Servs. Firms	Excl. Comp. Servs. Firms	Year API < 2012	Year API < 2012	Year API < 2012	Year API >= 2012	Year API >= 2012	Year API >= 2012
Post x API	0.337^{***}	0.263^{*}	0.379^{***}	0.267^{**}	0.206^{+}	0.310^{**}	0.689^{***}	0.637^{***}	0.658^{***}	0.234^{*}	0.209	0.254^{*}
	(0.0883)	(0.109)	(0.0867)	(0.0997)	(0.125)	(0.0986)	(0.154)	(0.184)	(0.149)	(0.101)	(0.161)	(0.102)
Post API x API Developers (100s)	1.751^{***}			1.817^{***}			0.854			2.106^{*}		
	(0.473)			(0.483)			(0.761)			(0.895)		
Post API x API Followers (100s)		0.134^{*}			0.124^{*}			0.0657			0.0812	
		(0.0592)			(0.0546)			(0.0620)			(0.171)	
Post API x API Change Count (100s)			0.572			0.715^{+}			9.733^{***}			0.404^{+}
			(0.384)			(0.431)			(2.236)			(0.223)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R2 Adjusted	0.932	0.932	0.932	0.934	0.934	0.934	0.929	0.929	0.929	0.929	0.929	0.929
Obs	133202	133202	133202	119201	119201	119201	129670	129670	129670	130040	130040	130040
Firms	4647	4647	4647	4060	4060	4060	4556	4556	4556	4561	4561	4561
API Adopters	177	177	177	100	100	100	86	86	86	91	91	91

Notes: Standard errors in parentheses and clustered at firm level. Outcome variable log market value of firm. *Post x API* is a binary variable that equals one if a given firm has a public API operating on a given date. *API developers* are the number of developers (in 100s) of an API according to ProgrammableWeb. *API Followers* refers to the number of individuals (in 100s) on ProgrammableWeb who have elected to follow an API. *Change Count* refers to the number of times an API has been updated (in 100s). + $p < 0.10$ * $p<0.05$ ** $p<0.01$ *** $p<0.001$

Table A4: Impact of API usage intensity on market value.

	Log Market Value Increase	Log Market Value Increase
Rank of Firm by API Degree	0.0291*** (0.00633)	
Rank of Firm by Max API Centrality		0.0258** (0.00787)
Constant	-0.272 (0.257)	-0.163 (0.347)
Firms	67	67

Notes: * $p<0.05$ ** $p<0.01$ *** $p<0.001$

Table A5: Regression underlying the lines of best fit in Figure 6. Robust standard errors in parentheses. Outcome variable is increase in log firm market value from begining of sample to Q3 2020. Regressions weighted by firms' initial market value.

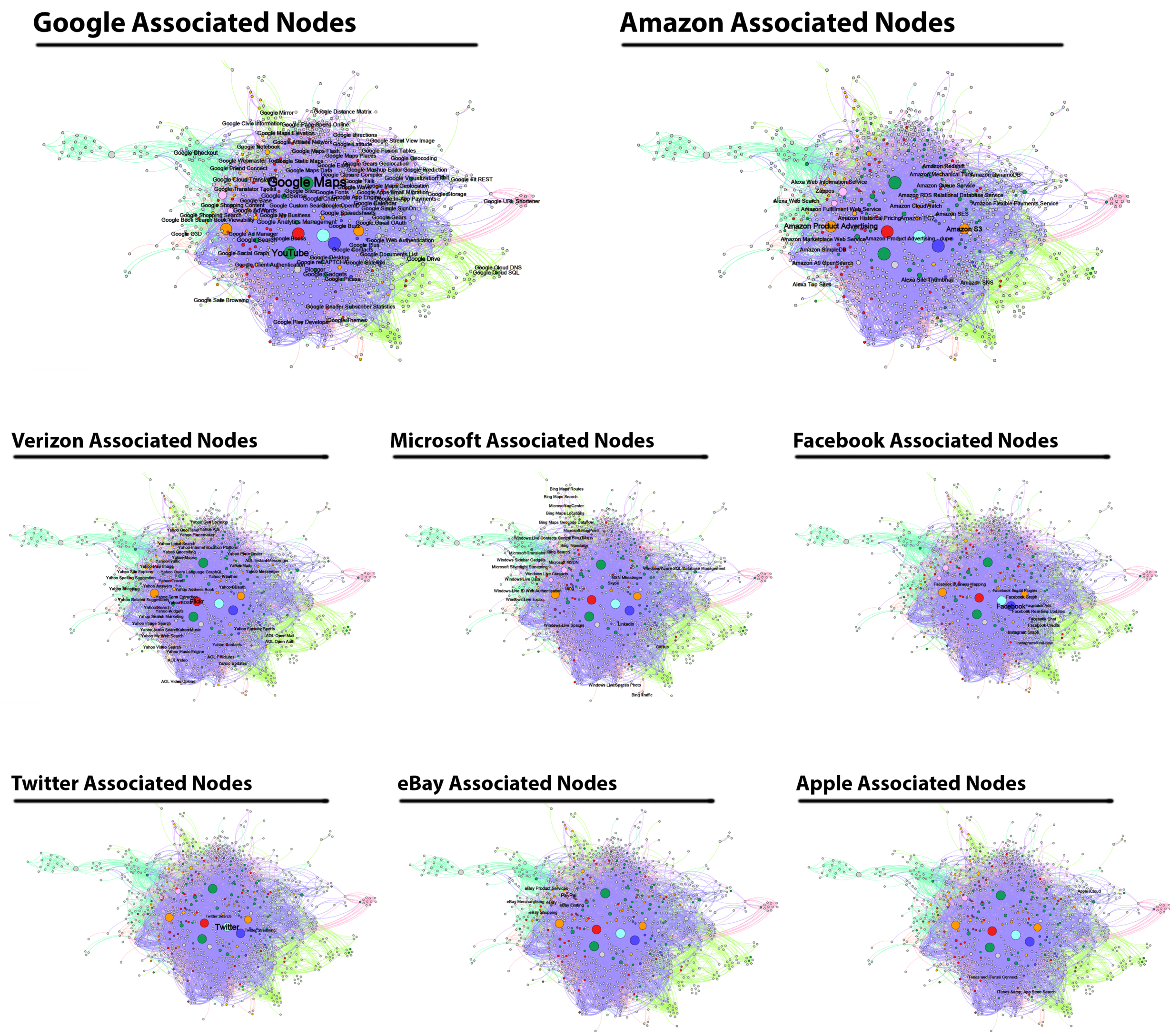

Figure A3: The API network, with nodes owned by different companies highlighted and labeled. Grid format. Notable features include: (1) Google is the host of most nodes in the network (2) Verizon is the second most common source of nodes, in large part due to its acquisition of Yahoo (3) Google's nodes occur everywhere in the API network. On the other hand, eBay's nodes are mostly associated with eCommerce apps, while Twitter and Facebook's nodes are more centrally located, associated with social apps in particular (4) Amazon hosts APIs at opposite ends of the API network, with Amazon's Product Advertising API particularly important to the eCommerce sub-network and Amazon E3 is particularly important to productivity (5) Microsoft hosts nodes important to several subnetworks as well. Many of its APIs are central to the eCommerce network while the LinkedIn API is central to both the social network and productivity subnetworks.

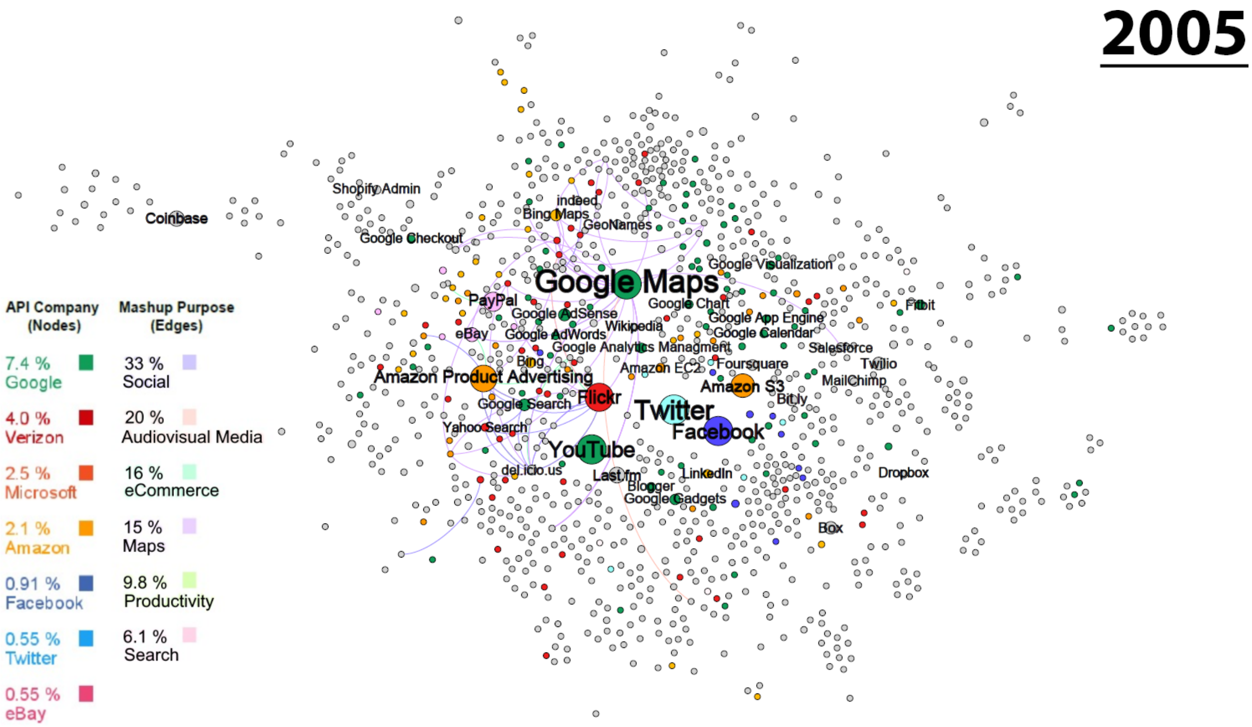

Figure A4: API network as of 2005.

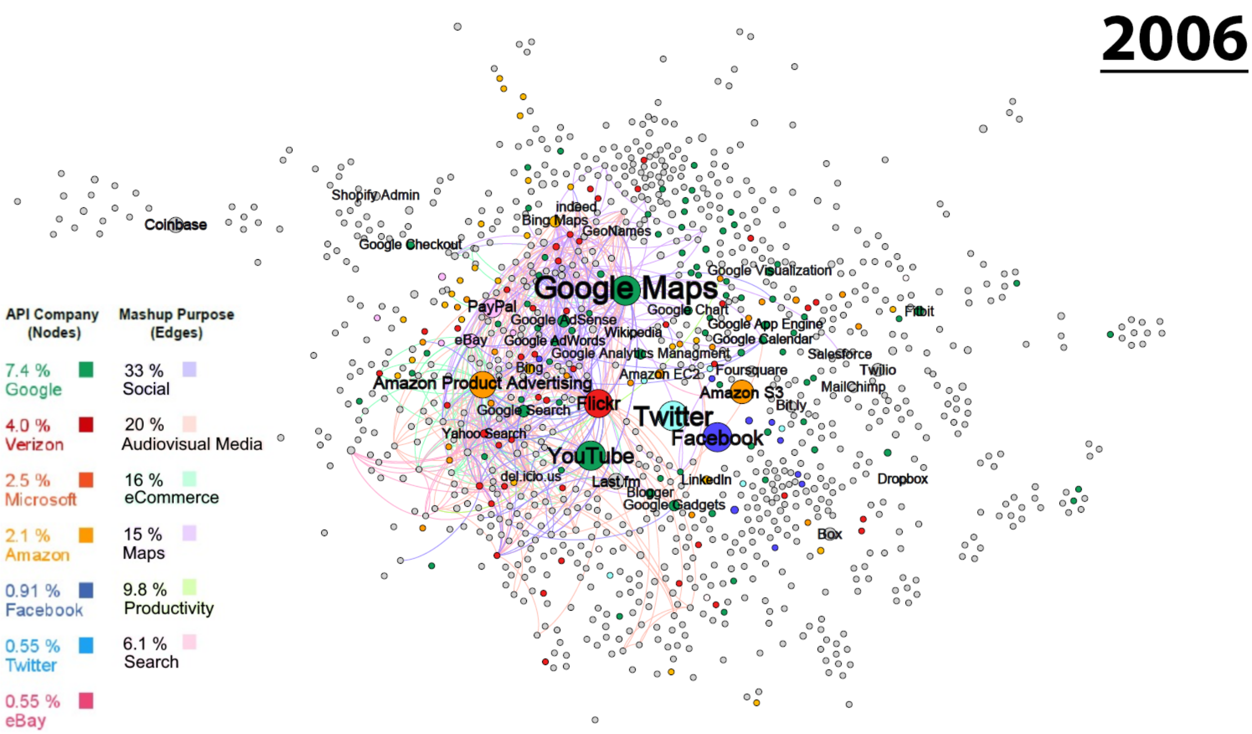

Figure A5: API network as of 2006.

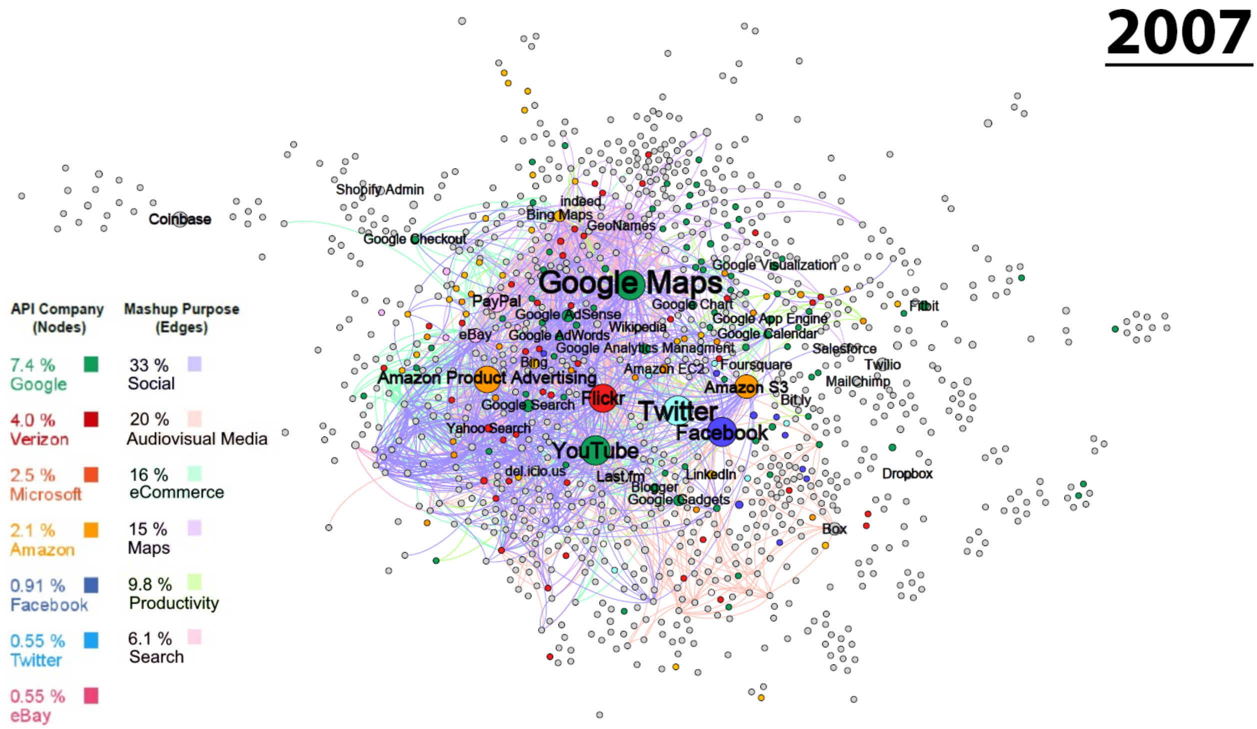

Figure A6: API network as of 2007.

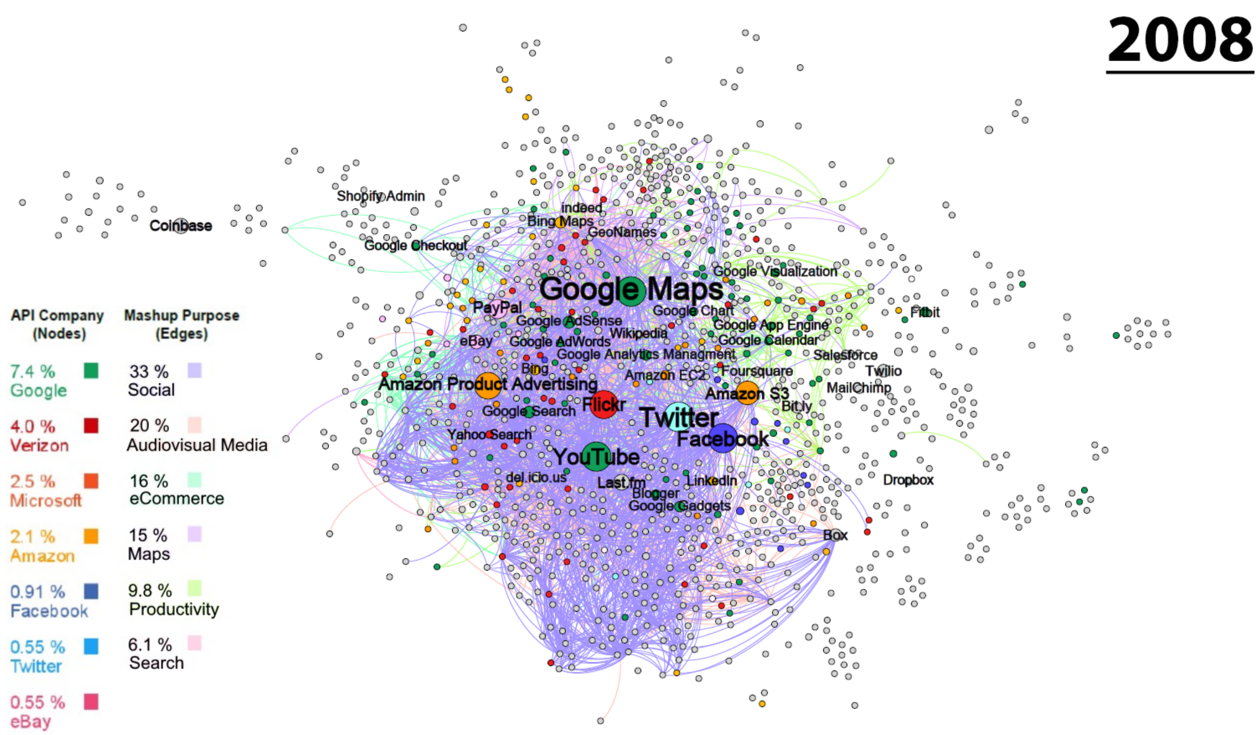

Figure A7: API network as of 2008.

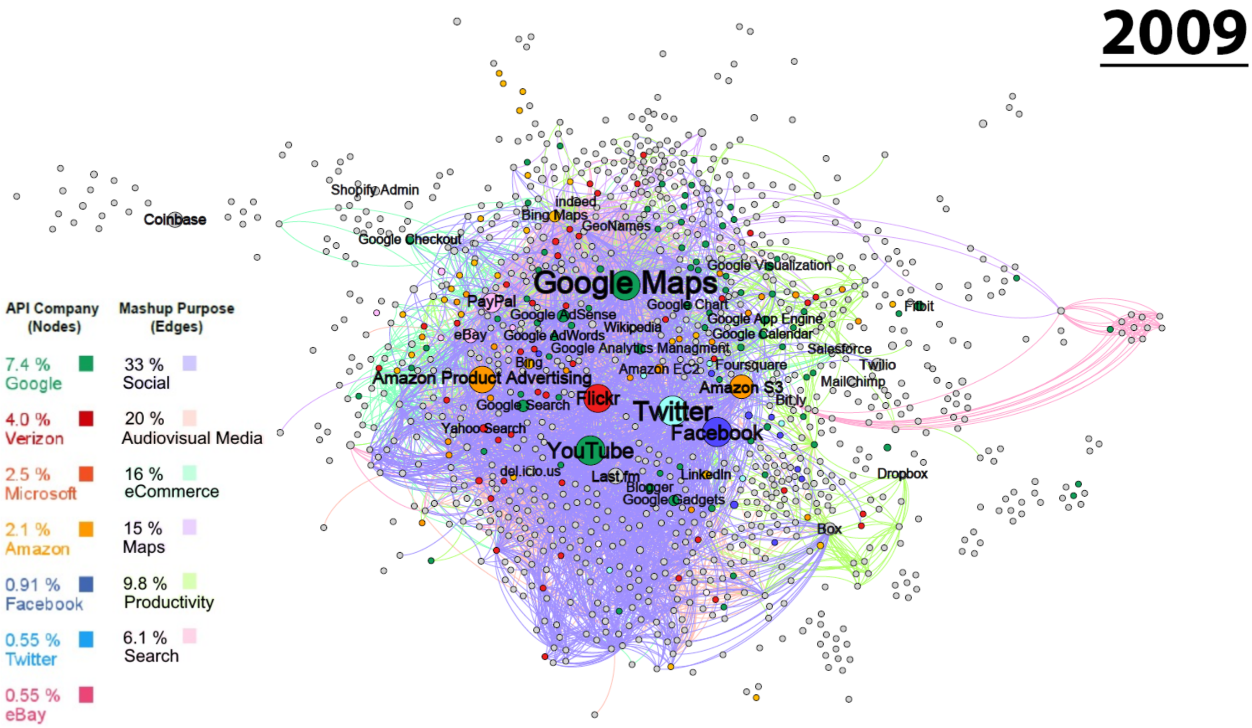

Figure A8: API network as of 2009.

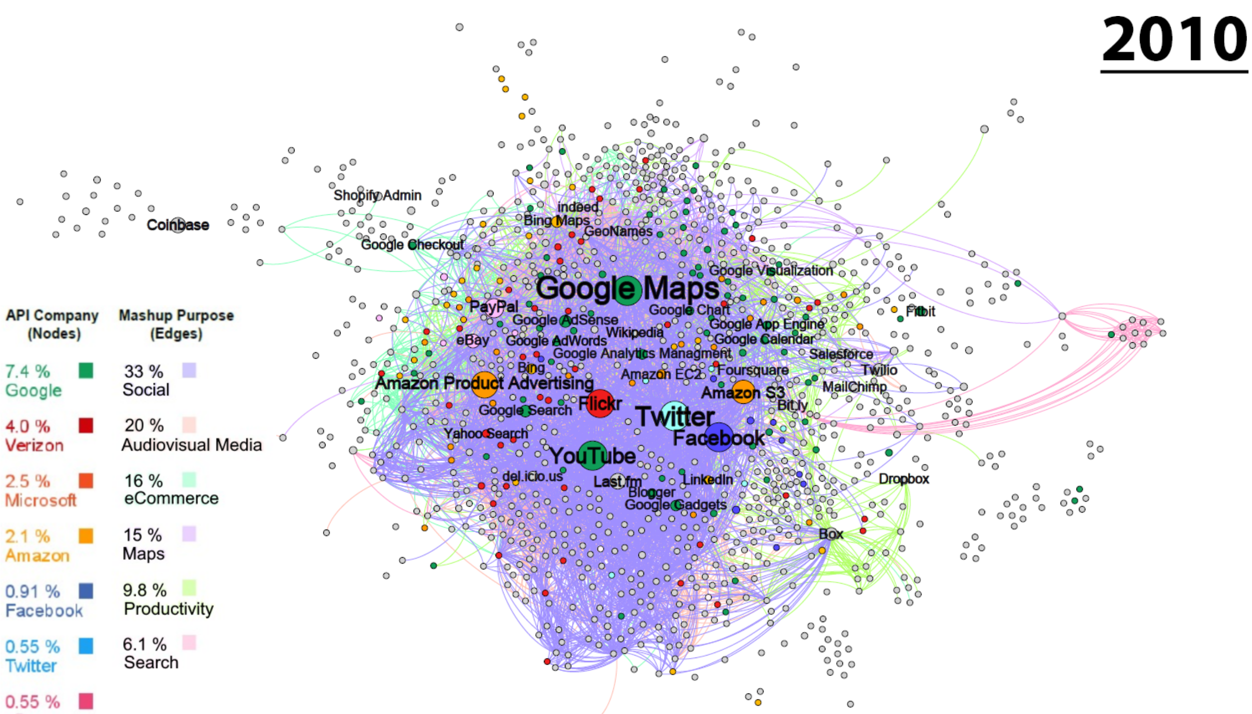

Figure A9: API network as of 2010.

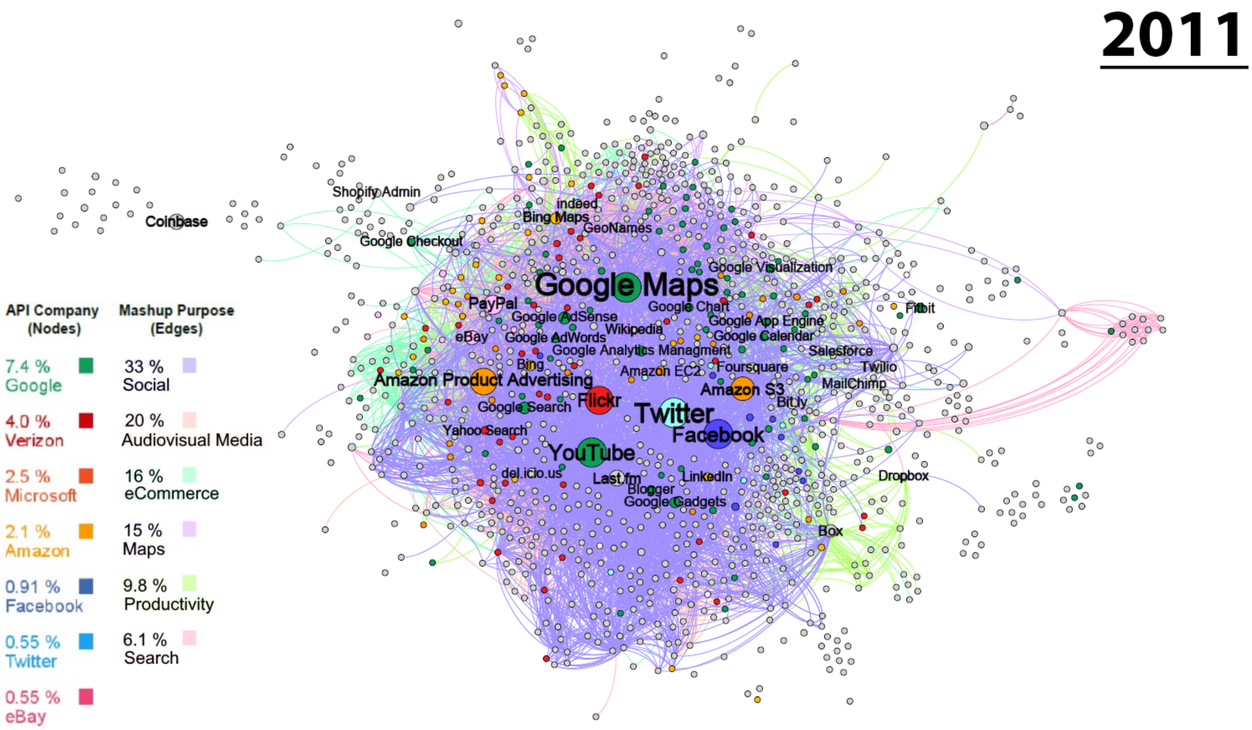

Figure A10: API network as of 2011.

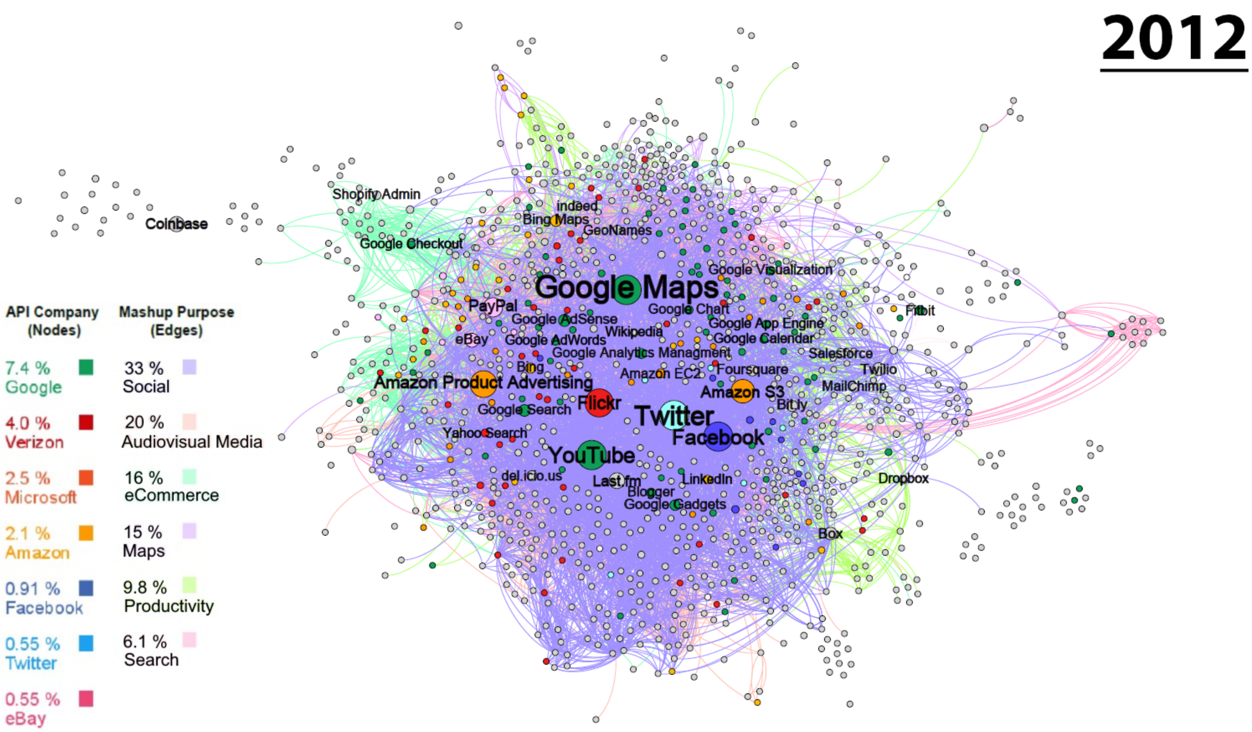

Figure A11: API network as of 2012.

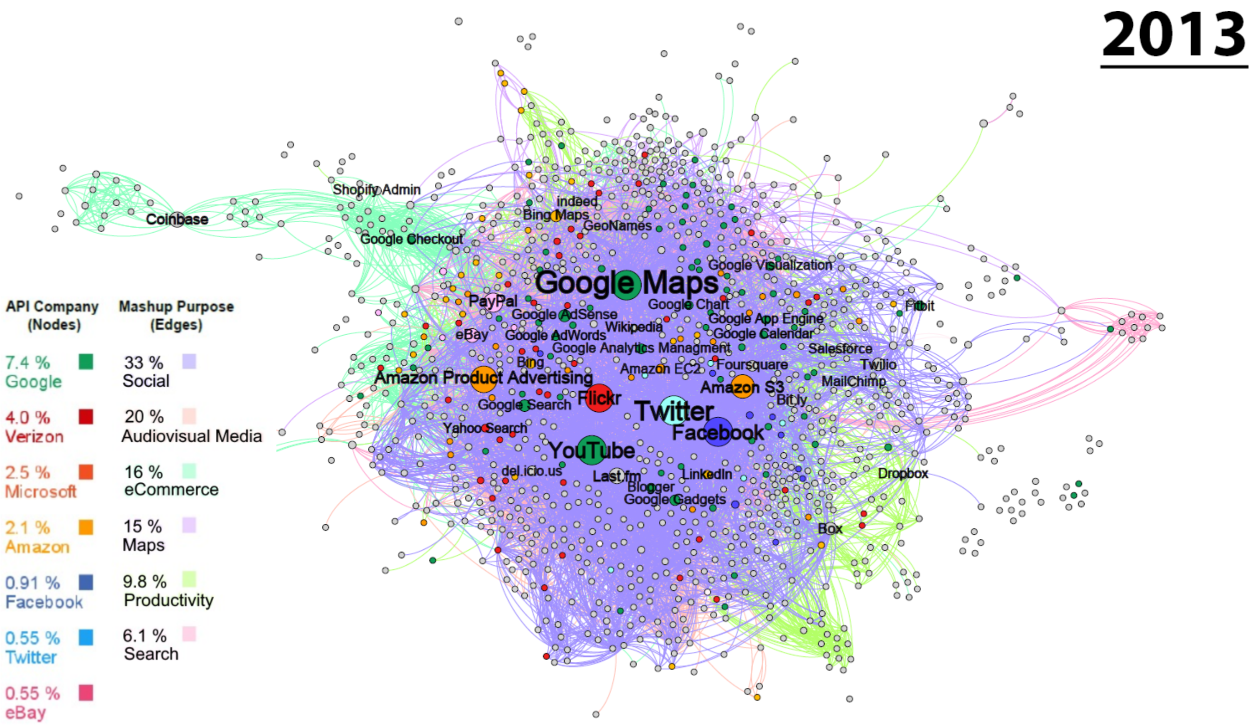

Figure A12: API network as of 2013.

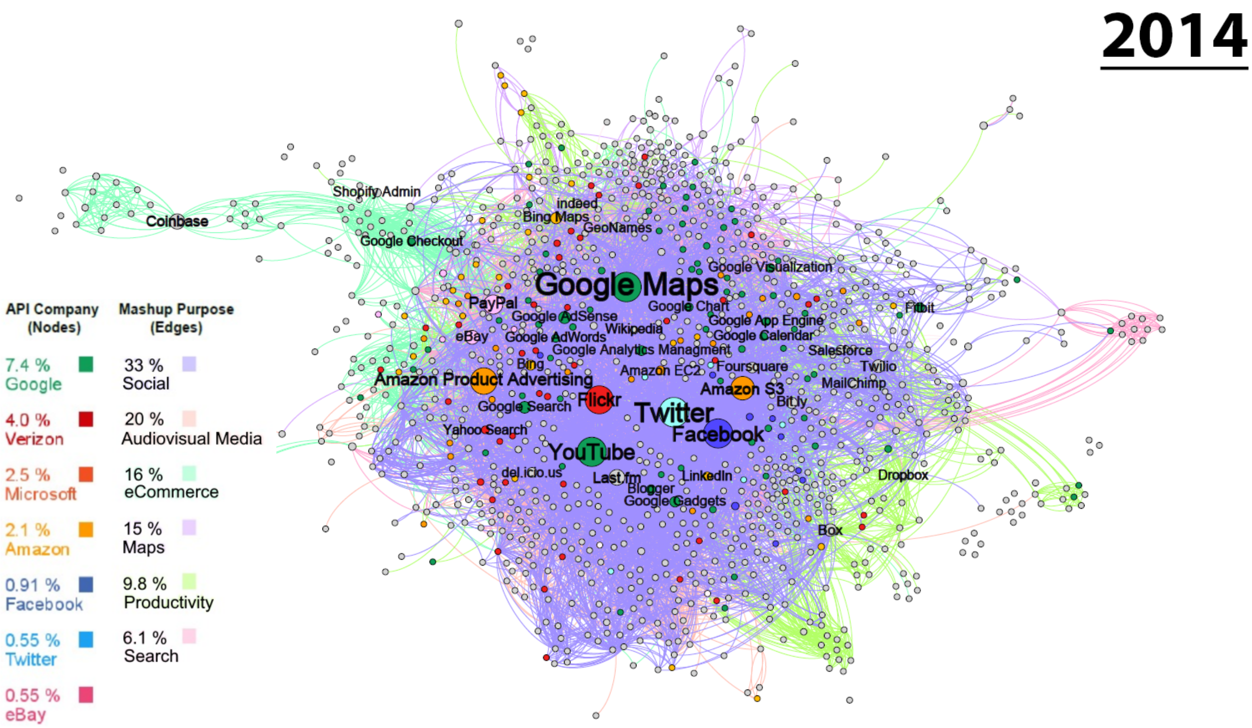

Figure A13: API network as of 2014.

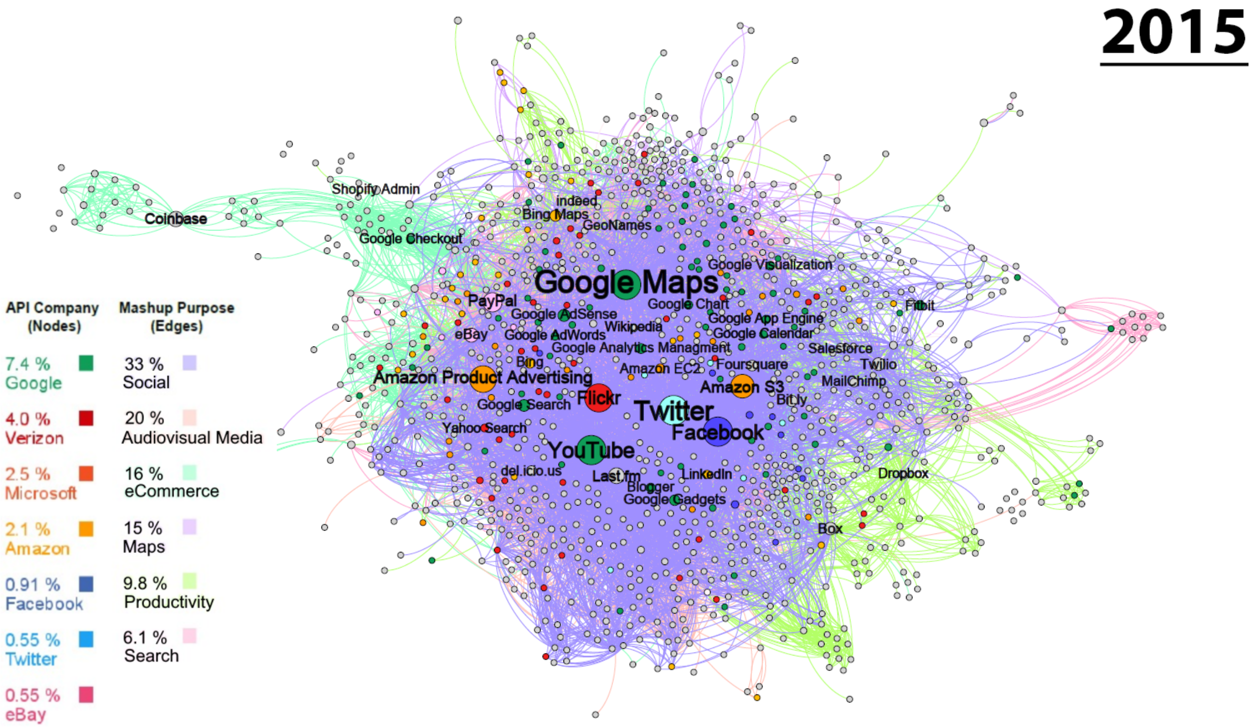

Figure A14: API network as of 2015.

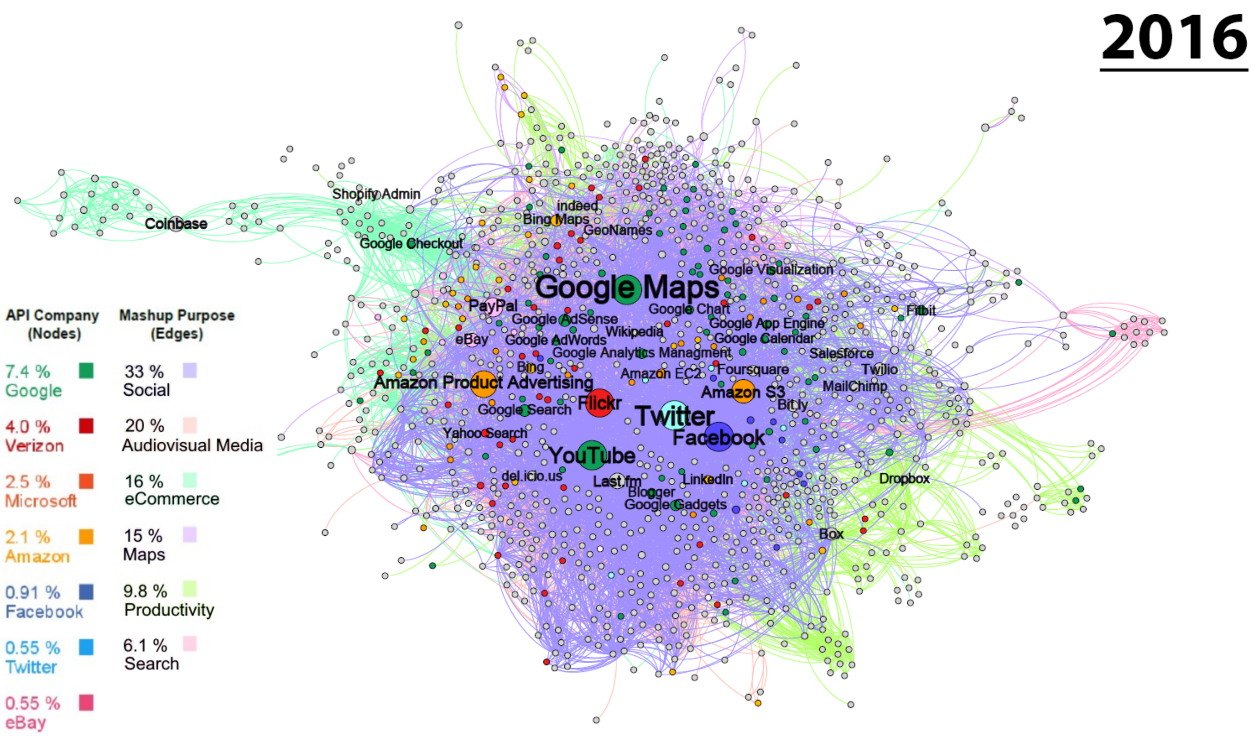

Figure A15: API network as of 2016.

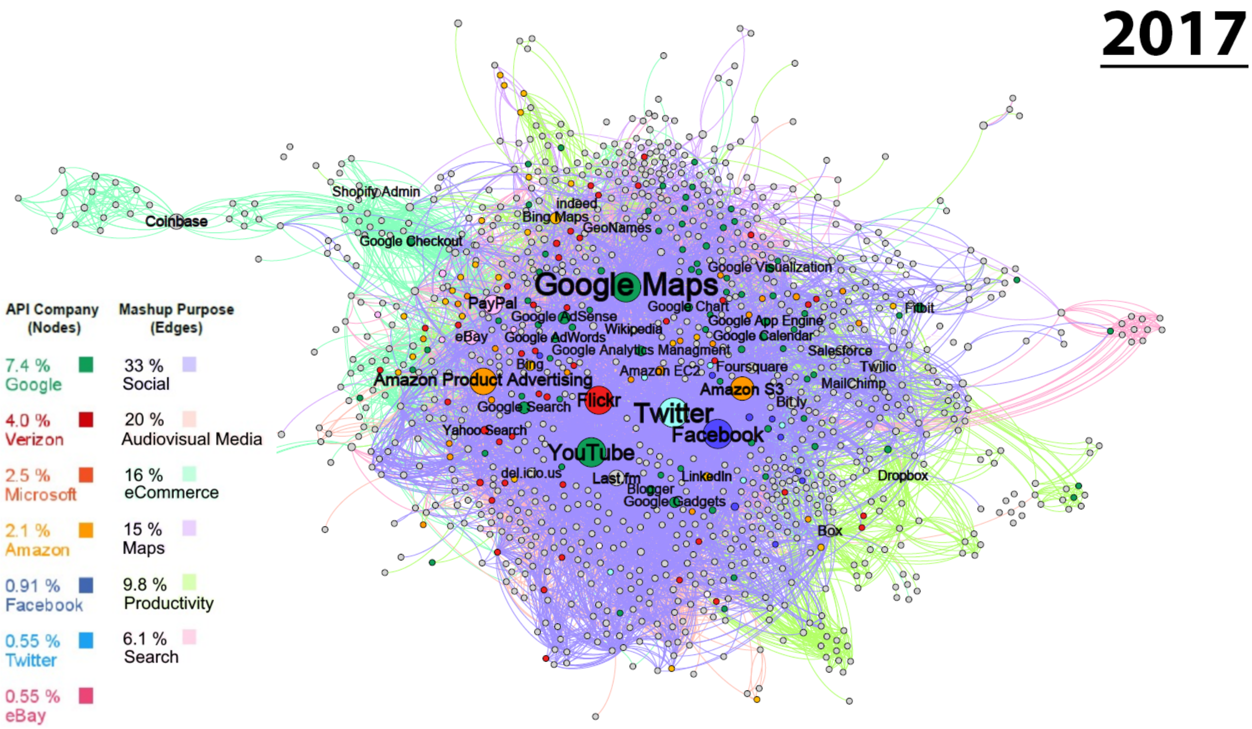

Figure A16: API network as of 2017.

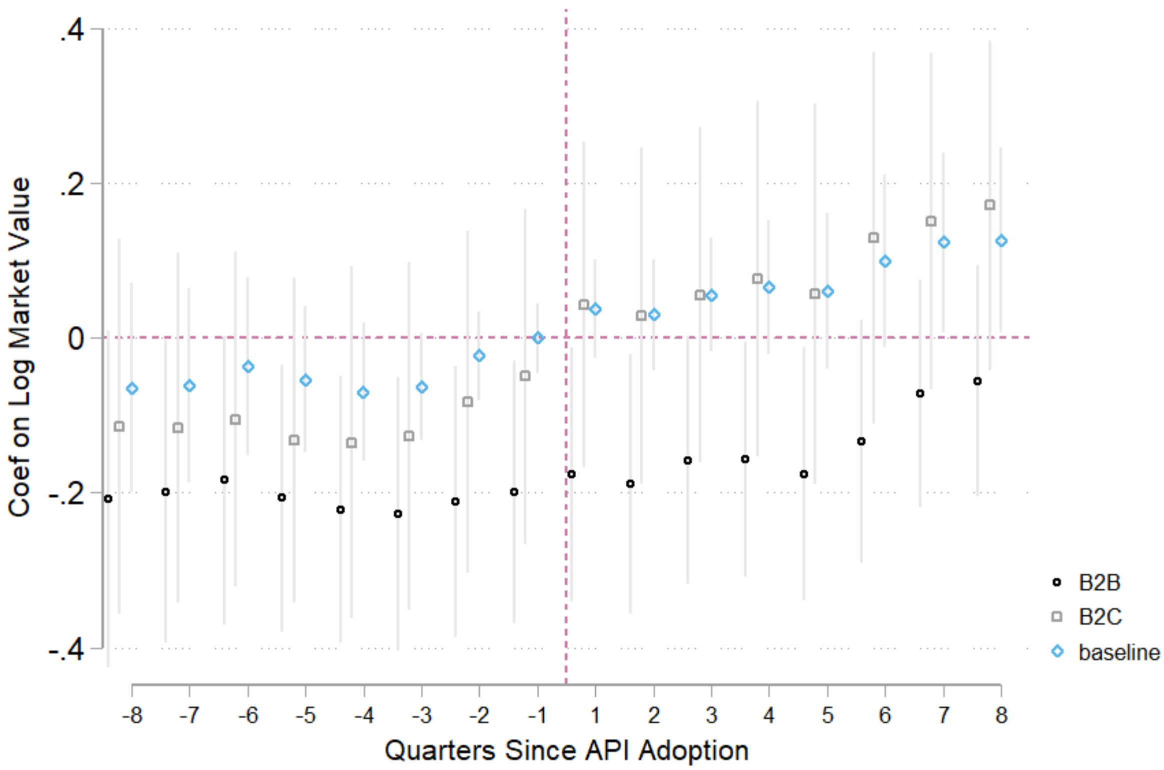

Figure A17: Regression following equation 2 with firms split into bins by whether their APIs are B2C vs. B2B oriented. All leads and lags specified in model, but only eight leads and lags reported. 'Baseline' reports estimates from Figure 4.

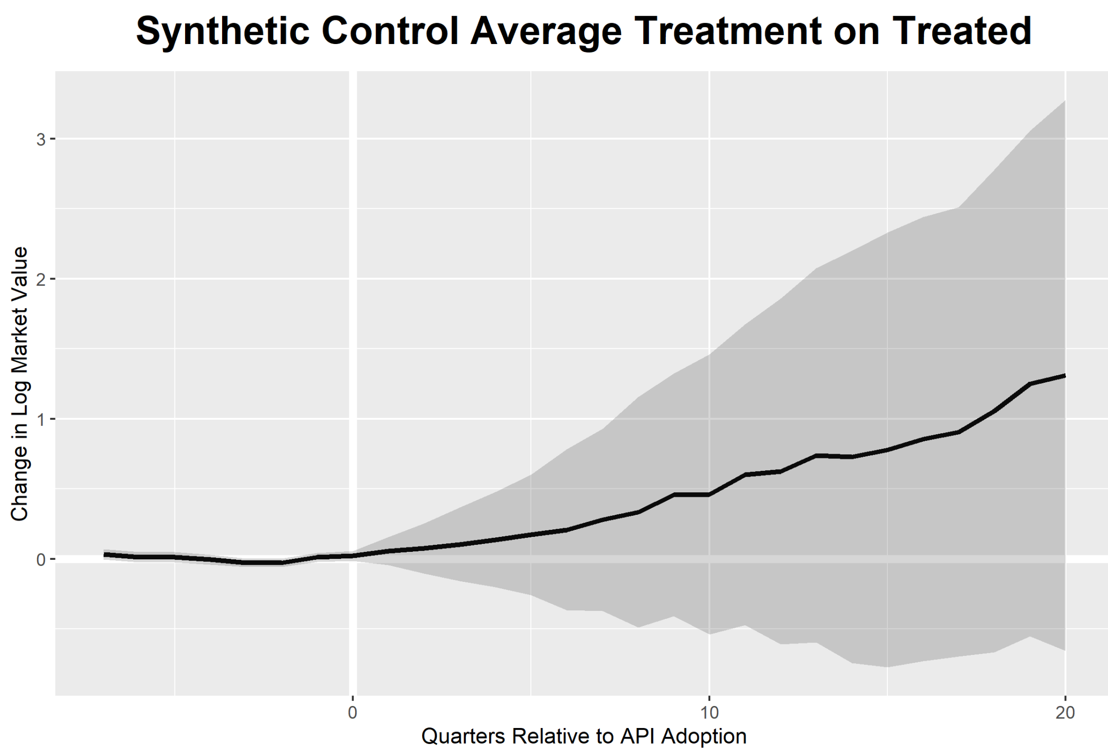

Figure A18: Estimated average treatment on treated and 95% confidence intervals for the effect of API adoption on market values using generalized synthetic control.

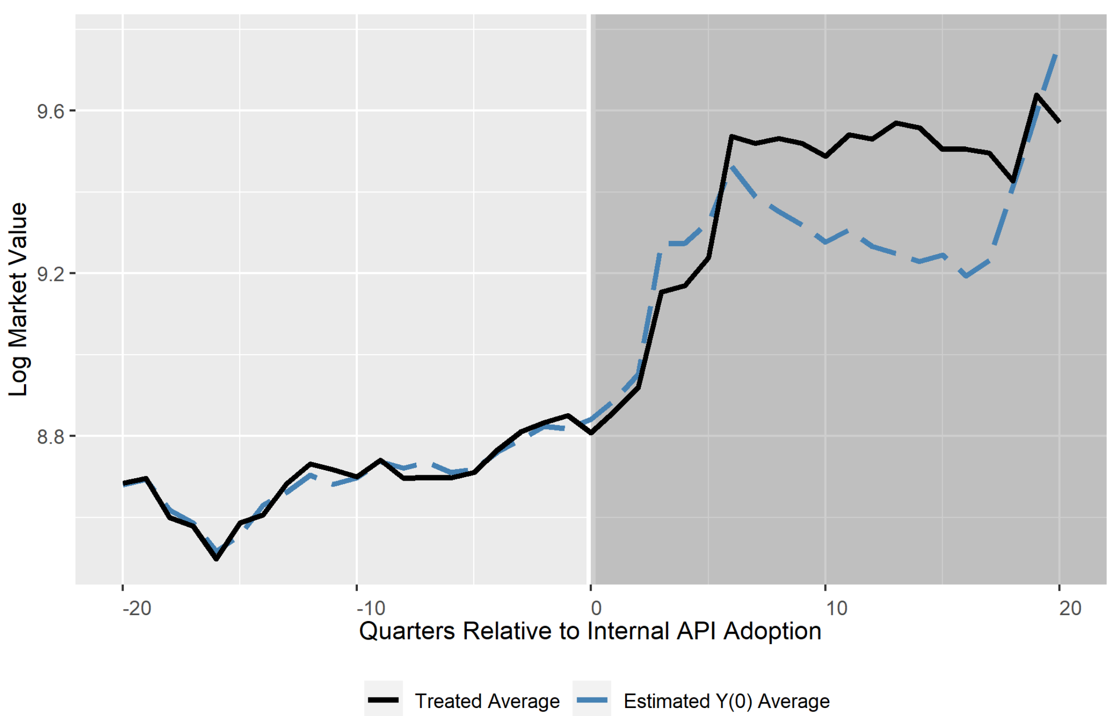

Figure A19: See appendix Figure A20 for confidence interval on difference.

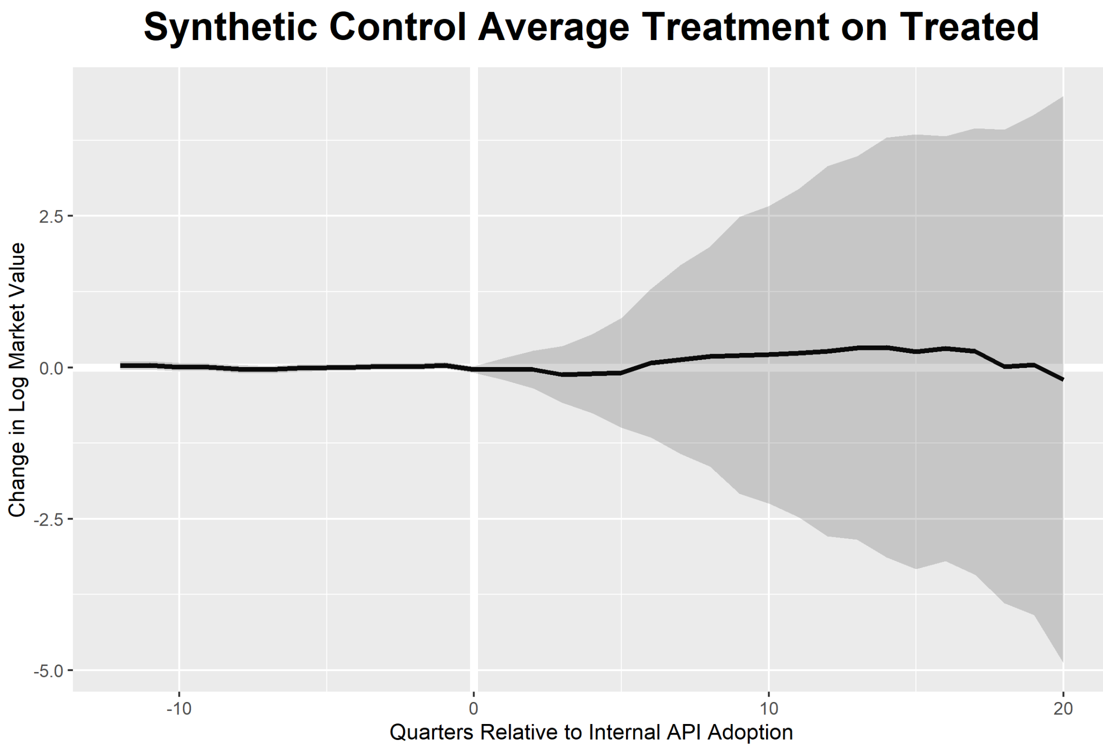

Figure A20: Estimated average treatment on treated and 95% confidence intervals for the effect of purely internal API adoption on market values using generalized synthetic control.

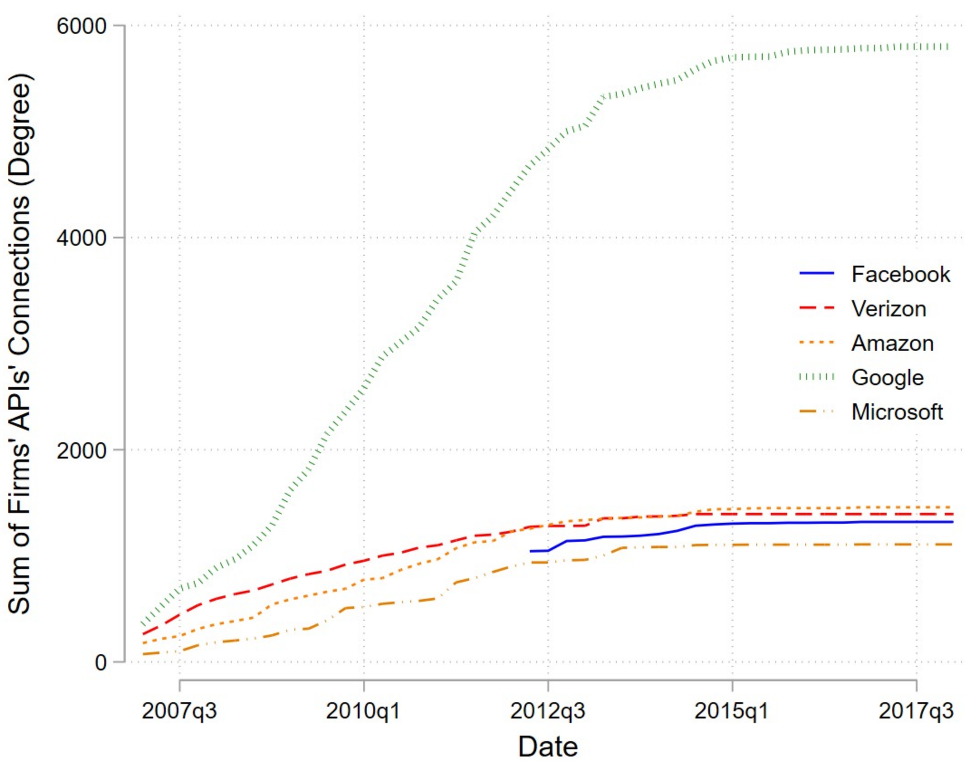

Figure A21: Sum of firms' APIs' degree over time. Five selected firms.

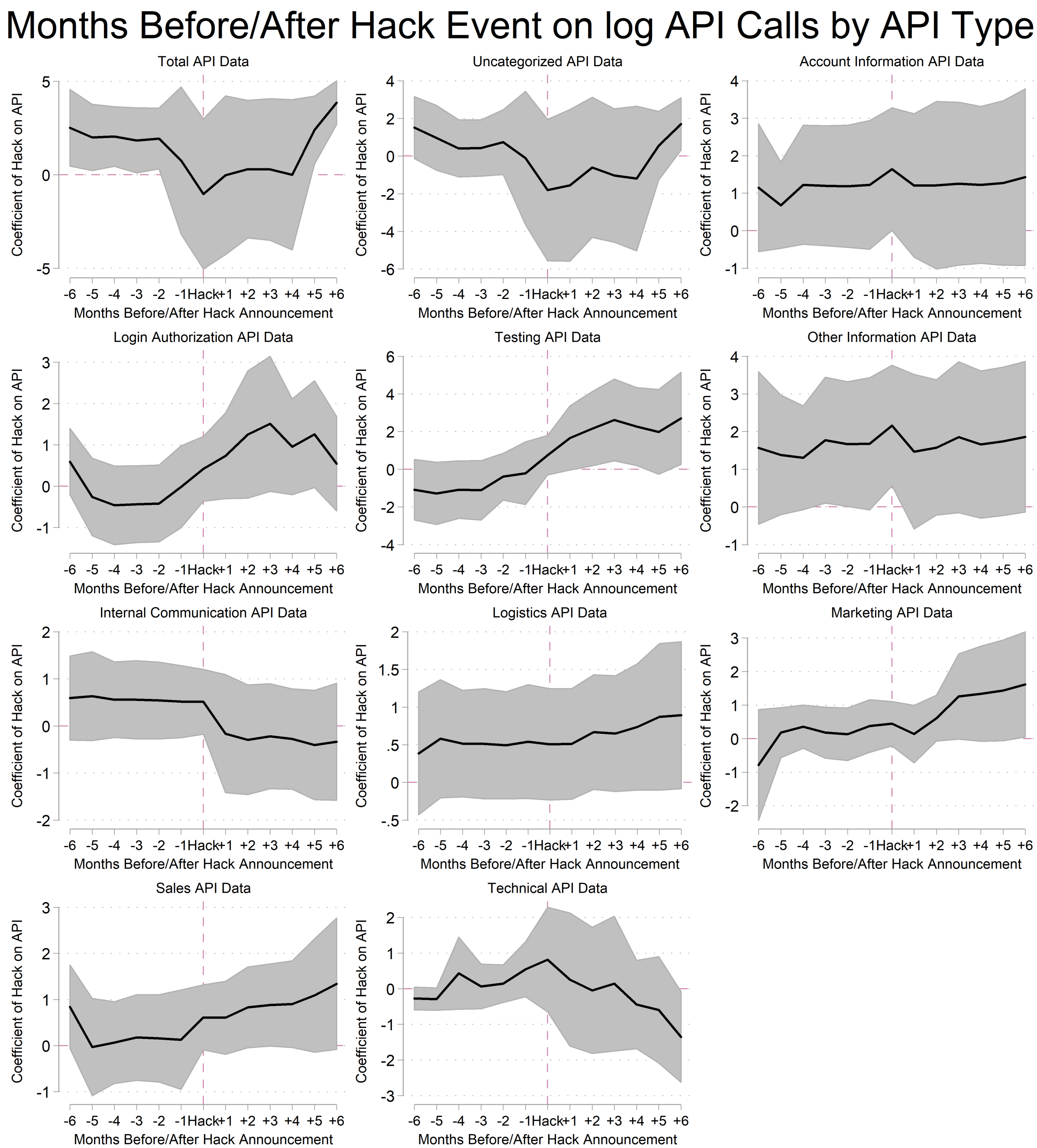

Figure A22: Log API data flows, and 90% confidence by type of API, in the months before and after a data breach event.

B API Functions

Using our proprietary dataset from an API tool provision company, we categorized APIs by their function. We sorted APIs into the following functions:

- Account Information: APIs related to storing, retrieving and displaying users' profiles
- Internal Communication: APIs for internal communication between employees
- Login/Authorization: APIs authenticating users and allowing information to be securely shared with other platforms
- Logistics/Inventory: APIs related to recording, managing and optimizing logistical items and inventory flow such as order delivery
- Maps/Locations: APIs dedicated to maps and GPS platforms, often Google Maps.
- Marketing/Customer Insights/Analytics: APIs related to storing and/or analyzing customer behavior or advertising information
- Media: APIs related to accessing, displaying or linking news or social media content
- Monitoring/Data Traffic Management: APIs related to collecting and managing data traffic
- Other: Identified APIs storing and providing information but unrelated to standard categories
- Sales: APIs related to consumer purchases, especially online shopping
- Test: Any API named a variation on 'test' as well as any other API used for conducting tests of the platform performance
- Technical: APIs performing technical internal function task unrelated to the aforementioned categories
- Uncategorized: APIs whose function could not be discerned from the name, the company developer portal, or Internet search

Many APIs have names which directly point to their functions, such as "sales" or "login" APIs. To determine the function of APIs with unclear or technical names, we did additional research. Internet search of technical API names often revealed their function. There was also often information on a firm's developer portal.

After classifying hundreds of APIs manually, we were able to identify consistent relationships between API names and corresponding functions. Using these relationships, we were able to identify and use certain keywords to partially automate API categorization. All automatic categorizations were double checked by hand.

Occasionally, even after additional research, how an API should be classified remained ambiguous. For example, APIs such as "Pingdom" performed tasks falling in both the Monitoring and Test categories. Similarly, APIs classified as Marketing or Sales could often arguably be placed in the other category. We used our best judgment in the classification of these ambiguous cases.

APIs in the ProgrammableWeb were classified into orientations (B2B, B2C, both, or unknown/neither), and the apps calling them were sorted by function based on their description in the directory and the tags associated.