Towards a Graph of (American) Tech Companies: A Prototype Visualization Tool for Research on Technology and Users

Niel Chah
University of Toronto, Faculty of Information
Toronto, Ontario, Canada
niel.chah@mail.utoronto.ca

ABSTRACT

A number of large technology companies, or so-called "tech giants", such as Alphabet/Google, Amazon, Apple, Facebook, and Microsoft, are increasingly dominant in people's daily lives, and critically studied in fields such as Science, Technology and Society (STS) studies, with an emphasis on technology, data, and privacy. This project aims to contribute to research at the intersection of technology and society with a prototype visualization tool that shows the vast spread and scope of these large technology companies. In this paper, a prototype graph visualization of notable American technology companies, their acquisitions, and services is presented. The potential applications and limitations of the visualization tool for research are explored. This is followed by a discussion of applying the visualization tool to research on personal data and privacy concerns and possible extensions. In particular, difficulties of data collection and representation are emphasized.

KEYWORDS
Science and technology studies; technology companies; graph visualization; user studies

CHI'19 Extended Abstracts, May 4–9, 2019, Glasgow, Scotland UK

ACM ISBN 978-1-4503-5971-9/19/05.
https://doi.org/10.1145/3290607.3313058

INTRODUCTION

In recent years, certain technology companies, such as Alphabet/Google, Amazon, Apple, Facebook, and Microsoft, have become increasingly prominent in people's everyday technology usage at a global level. Products and services such as Amazon, Facebook, Gmail, Instagram, MS Office, Siri, WhatsApp, and YouTube to name only a few have become household names with millions of daily users. However, as the international reach of these so-called "tech giants" increases, there is relatively little research on the full extent of their global scope using interactive information visualization tools.

Technology companies and their implications for people and society are the subject of intense study in fields such as Science and Technology Studies (STS) [3, 13], Internet and marketing research [6, 7], and media studies [8, 16] among others. These representative works consider the impact of these tech giants on society, democracy, privacy and other concerns from a critical perspective.

Previous studies into the public perception of technology companies have found that a significant portion of American users were not aware that Facebook owns Instagram [6] or WhatsApp [7] and that Google owns YouTube [6] or Waze [7]. This knowledge gap can become problematic at times when certain products and services are suggested as alternatives to users based on privacy concerns (e.g. suggesting Instagram as an alternative to Facebook) although both products are owned by the same parent company. Future research should benefit from the use of large-scale visualization tools that map the privacy and security implications of the most widely used products and services.

Prior works in visualizing the graph-like network of corporate ownership have been done for the "Big 10" food and consumer goods companies [9, 10], American media companies [17], American financial institutions, car manufacturers, and even beer [5]. In the information and communications technology (ICT) domain, while prior research has focused on particular platforms, such as Google and Facebook, there has been relatively little work done with visualizations to capture the full scope of the technology companies. This gap in the literature and current tooling is understandable considering the heterogeneous characteristics of the "tech giant" companies where "differences between these organisations and their services outnumber their similiarities" [13].

GRAPH OF (AMERICAN) TECH COMPANIES

The design of the visualization tool was motivated by the Visual Information Seeking Mantra to "overview first, zoom and filter, [and] then view details-on-demand" [18]. The interactive graph visualization tool is available at http://nchah.github.io/tech-graph/. The prototype web application is built using the D3.js library with the following components and features. Browser compatibility was tested on Chrome and Firefox, with optimal performance on Chrome.

Overall Architecture. *(1) Data Collection.* The data on the technology companies and their acquisitions and services were gathered from the company's English Wikipedia pages and external

Figure 1: Cluster for Alphabet.

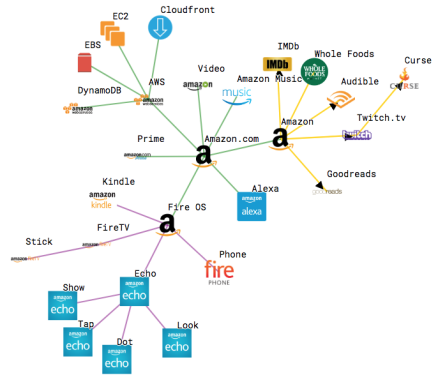

Figure 2: Cluster for Amazon.

web resources. This method of data collection was required in the absence of a dedicated data set that specifically covered the scope of American tech giants. In particular, data collection focused on currently supported products and services. *(2) Pre-processing.* The node and link arrangements were pre-processed with a prototype *ontology* [15] to structure the data into Company, Product, Service typed entities linked through "owns", "acquired", and "transitioned to" relationships. *(3) Web Application.* The tool is built as a web application using HTML, CSS, and the D3.js library to render the visualization. The data consumed by D3.js is in a JSON format that contains the technology company's products, services, acquisitions, descriptions, and relationships with other nodes. The entire application is hosted for free on the GitHub Pages platform.

Graph Visualization. The visualization displays the technology companies in a graph format with an interactive element for users to explore and learn about the technology networks. The center root node of each graph represents a distinct technology company with additional nodes representing the products, services, and acquisitions. The D3.js force-directed layout is used so that the layout is automatically generated by specifying the link distance between nodes (70), the "charge" between nodes to add repelling forces (-250), and node size (32px).

Rearrange and Zoom. As an interactive web application that uses the D3.js library, users can zoom in or out, and move nodes with their pointer by left-clicking, holding, and dragging the node. Once perturbed, the graph rearranges automatically in the target location. This mode of interaction allows the user to rearrange the visualization to suit their needs [20] and according to their mental model (e.g. certain products the user is interested in are moved to one side of the page).

Contextual Tooltips. The tooltip feature allows users to gain further contextual information by hovering their mouse pointer over the different nodes. The tooltip displays a brief description of the entity from the English Wikipedia page. With this feature, users could learn about the different products, services, and acquisitions as they explored the sprawling technology networks.

Double-click. By double-clicking on a node, a new browser tab is opened to the relevant English Wikipedia page for the entity. Similar to the tooltip, this feature allows the user to gain further information as they interact with the visualization tool.

FINDINGS

The resulting graphs for each examined technology company produced interesting visualizations that highlight each organization's unique structure, strategic acquisitions, and product or service characteristics. Notable aspects of each tech giant's graph will be briefly described. In total, over 170 nodes were included in the current implementation of the visualization. It should also be emphasized that the specific relationships in the tool are structured according to a schema or *ontology* [15] to structure the entities in a certain way.

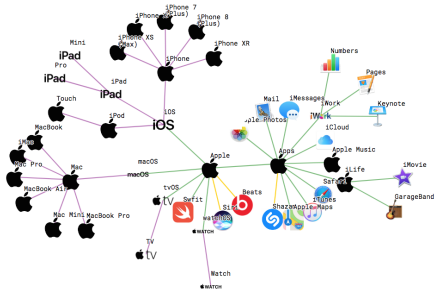

Figure 3: Cluster for Apple.

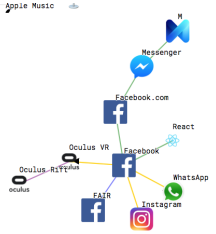

Figure 4: Cluster for Facebook.

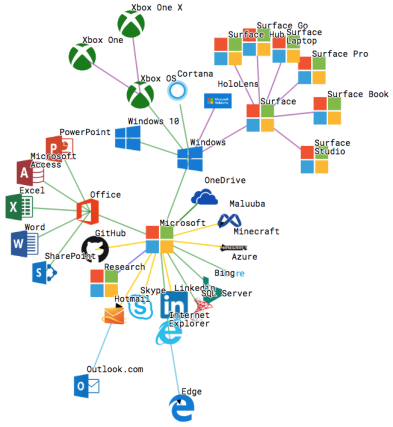

Figure 5: Cluster for Microsoft.

Alphabet/Google. The Alphabet cluster shows that Google is unsurprisingly the dominant subsidiary in comparison to the many off-shoots of Alphabet, from the biotech venture Calico to the moonshot lab X. From the Google node, the company's cloud offerings, Android platform, and Chrome web browser suite in particular make up substantial portions of the graph. Many graph traversals are not even needed to see connections to other significant products such as Maps, Gmail, and YouTube. Notably, the graph also highlights how strategic acquisitions contributed to a large part of Alphabet's current products from Android, DeepMind, Maps, Nest, Waze, and YouTube.

Amazon. For the Amazon cluster, a distinction is made between Amazon, the company, and Amazon.com (or other localized URL), where users shop for products. Similar to Alphabet's heavy focus on Google, many of Amazon's products and services are in some way related to Amazon.com: the cloud services Amazon Web Services (AWS), the virtual assistant Alexa, and the Fire OS based Echo products. Amazon also owns other entities such as IMDb, Twitch, and Whole Foods.

Apple. In contrast to the diversified domains of other tech giants, the Apple cluster shows the company's focus on producing its own hardware products and various bundled software applications. To structure this cluster, the first hops out of the central node are for the software entities such as the different operations systems (OS) and applications. From the iOS, macOS, tvOS, and watchOS nodes, the many related hardware products are represented. Other nodes in the graph are software products, such as the many apps from Apple Maps and Safari to the recently acquired Shazam.

Facebook. Although the Facebook cluster has fewer nodes than other tech company clusters, each of the social platforms, Facebook.com, Instagram, and WhatsApp, are used by billions of users [12]. Other notable entities that fall in this cluster include the widely used React JavaScript framework, the Oculus virtual reality company, and the artificial intelligence research department.

Microsoft. The Microsoft cluster is similar to Apple in its focus on development for a particular operating system, Windows, while also similar to Alphabet in its diversification through acquisitions of various companies. The nodes for Windows and Office software link to many of the iconic products by the company. Other acquisitions of Github, LinkedIn, Skype and even the game Minecraft are connected directly to the central node.

IMPLICATIONS FOR RESEARCH

Based on the initial findings from this prototype tool, the relevant implications of this project for future research are considered. In particular, applications as a teaching tool for general users unfamiliar with the technology landscape on privacy and security issues, and data exports in a linked data format (e.g. RDF) for academic researchers are proposed.

Presentation of Scale with Interactive Visualizations. A graph visualization of the tech giants acted as a general reference for users that facilitated a top-down view of large-scale relationships. With an interactive bird's-eye view, users could (1) view industry-wide scale of the various products

and services (e.g. understanding the diversity of products owned by a single company) and (2) gain awareness of the potential privacy implications (e.g. understanding which "alternative" products are actually operated by the same parent company). Extensions to previous research [6, 7] that study general user's technology knowledge on a larger range of products and services are possible with this tool. In addition, the visualization may be expanded to cover other international "tech giants".

Studies on Privacy and Technology Literacy. Another extension for future research is conducting user studies with the visualization tool on individual technology usage, personal data and information sharing issues, and privacy concerns [1] (e.g. exploring user privacy preferences across these widely used technology products and services with a customized tooltip that displays privacy policies). Such studies should be done noting that users have different technology familiarity, particularly for under-served and under-resourced communities [4]. A further direction is to increase the research applications of this tool by supporting data exports in a linked data format [2] that can be joined with other datasets to explore additional questions and educate users on additional dimensions [11, 19]. This allows for other communities of researchers, such as those based in STS and platform studies, to analyze the companies from critical perspectives.

DISCUSSION AND FUTURE WORK

This paper presented a prototype tool to visualize a subset of the many products, services, and acquisitions by notable American technology companies. In creating the graphs for each company, a number of difficulties were faced in (1) selecting the specific acquisitions, products, and services that were represented, and (2) constructing the structure (schema or ontology [15]) of the linkages between entities. In its current state as a prototype, it is important to consider what future development and features may improve the maturity of the tool and its applications for general users and research.

A grounded theory approach [14] was used to construct the graphs for each technology company, starting with the Wikipedia pages for each entity and navigating to the linked products and services by examining the infoboxes and freeform text. This approach filled in the gaps in data coverage of existing data sources such as DBpedia [11] and WikiData [19]. One recurring pattern that arose from the data was the distinction between hardware and software products across many of the companies. For the most part, the graphs were constructed to show that a company "owns" (or maintains) a software product or service which in turn "is implemented as hardware" in (or run on) a specific hardware product (e.g. Apple's iOS is run on an iPhone X).

The difficulties of implementing an ontology to represent these diverse technology companies should be emphasized. For example, in the case of Amazon Alexa, the virtual assistant could be linked to the Echo product line or more directly to the main Amazon.com entity as it constitutes a virtual assistant used across many devices. By exporting the structured data of this application, other researchers and users can adjust the relationships and schema for specific uses.

Further development of features is also crucial. For instance, to visualize statistics on market value, the node sizes could be scaled according to market capitalization. Temporal elements (i.e. showing acquisitions and products over time) could also be added with an interactive slider.

REFERENCES

[1] Louise Barkhuus. 2012. The mismeasurement of privacy: using contextual integrity to reconsider privacy in HCI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 367–376.

[2] Christian Bizer, Tom Heath, and Tim Berners-Lee. 2011. Linked data: The story so far. In *Semantic services, interoperability and web applications: emerging concepts*. IGI Global, 205–227.

[3] Jean Burgess and Joshua Green. 2018. *YouTube: Online video and participatory culture.* John Wiley & Sons.

[4] Nicola Dell and Neha Kumar. 2016. The ins and outs of HCI for development. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 2220–2232.

[5] Jesus Diaz. 2014. Fascinating graphics show who owns all the major brands in the world. https://gizmodo.com/fascinating-graphic-shows-who-owns-all-the-major-brands-1599537576.

[6] DuckDuckGo. 2018. A Majority of Americans Don't Know That Facebook Owns Instagram. https://spreadprivacy.com/facebook-instagram/.

[7] DuckDuckGo. 2018. Most Americans aren't aware that Facebook owns WhatsApp, or that Google owns Waze. https://spreadprivacy.com/facebook-whatsapp/.

[8] Tarleton Gillespie. 2018. *Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media*. Yale University Press.

[9] Beth Hoffman. 2013. *Behind the Brands: Food justice and the'Big 10'food and beverage companies*. Vol. 166. Oxfam.

[10] Beth Hoffman. 2013. PepsiCo Makes Quaker Oats Chewy Bars (And Other Surprising Brand-Product Associations). https://www.forbes.com/sites/bethhoffman/2013/02/26/who-owns-your-favorite-food-brands/.

[11] Jens Lehmann, Robert Isele, Max Jakob, Anja Jentzsch, Dimitris Kontokostas, Pablo N Mendes, Sebastian Hellmann, Mohamed Morsey, Patrick Van Kleef, Sören Auer, et al. 2015. DBpedia–a large-scale, multilingual knowledge base extracted from Wikipedia. *Semantic Web* 6, 2 (2015), 167–195.

[12] Rani Molla. 2018. WhatsApp is now Facebook's second-biggest property, followed by Messenger and Instagram. https://www.recode.net/2018/2/1/16959804/whatsapp-facebook-biggest-messenger-instagram-users.

[13] Martin Moore. 2016. Tech giants and civic power. *Centre for the Study of Media, Communication and Power. King's College London* (2016).

[14] Michael J Muller and Sandra Kogan. 2010. Grounded theory method in HCI and CSCW. *Cambridge: IBM Center for Social Software* (2010), 1–46.

[15] Natalya F Noy, Deborah L McGuinness, et al. 2001. Ontology development 101: A guide to creating your first ontology.

[16] Jean-Christophe Plantin, Carl Lagoze, Paul N Edwards, and Christian Sandvig. 2018. Infrastructure studies meet platform studies in the age of Google and Facebook. *New Media & Society* 20, 1 (2018), 293–310.

[17] Columbia Journalism Review. 2018. Who Owns What. https://www.cjr.org/resources.

[18] Ben Shneiderman. 2003. The eyes have it: A task by data type taxonomy for information visualizations. In *The Craft of Information Visualization*. Elsevier, 364–371.

[19] Denny Vrandečić and Markus Krötzsch. 2014. Wikidata: a free collaborative knowledgebase. *Commun. ACM* 57, 10 (2014), 78–85.

[20] Ji Soo Yi, Youn ah Kang, and John Stasko. 2007. Toward a deeper understanding of the role of interaction in information visualization. *IEEE transactions on visualization and computer graphics* 13, 6 (2007), 1224–1231.