VALUENEX
intellectual innovator

FACEBOOK:
A BREACH IN THE WALL

A TECHNOLOGICAL APPROACH TO THE FACEBOOK-CAMBRIDGE ANALYTICA SCANDAL

April 2018

The simple story is this: Facebook exposed data on up to 87 million users to Cambridge Analytica, which used the data for a major political campaign. The longer story is about the explosive research and development of the technology that allowed this to happen to begin with, and what other things these technologies are capable of doing with our data and the length of their reach.

To blow the lid on this from a technological perspective, it makes sense to look at Facebook's technology portfolio, also known as its patent portfolio, as a solid starting point. To retrace the trail back to the technology that opened Pandora's box, we pulled Facebook's 2,826 patent records from the United States Patent Office (USPTO) databases from 2008 (the year of their first granted patent) to April 2018, and ran them through VALUENEX's big data and predictive analytics platform.

With proprietary algorithms incorporating unsupervised machine learning, high-dimensional visualizations and precision clustering, the VALUENEX Radar output allowed for easy-viewing of the 2,826 patents grouped into 524 clusters based on full text semantic similarities precisely plotted on a single map. With the scale of distance representing the difference of technology coverage and contour lines carving out levels of document density, we could quickly gain insight into what Facebook is doing, where their priorities are, and what technology was fundamental in allowing the Cambridge Analytica exploitation to occur.

Before we jump into the results, lets go over some of the key factors: Facebook provides a number of technology tools to software developers, and one of the most popular is Facebook Login, which lets people simply log in to a website or app using their Facebook account instead of creating new credentials. People use it because it's easy, usually involves just one or two taps, and eliminates the need for people to remember several unique username and password combinations.

However, when people use Facebook Login they grant the app's developer a range of information from their Facebook profile, including things such as their name, location, email or friends list. This is what happened in 2015, when a Cambridge University professor named Dr. Aleksandr Kogan created an app called "thisisyourdigitallife" that utilized Facebook's login feature. At least 270,000 people used Facebook Login to create

accounts and therefore permitted the sharing of personal profile data with Dr. Kogan, who later passed the data to Cambridge Analytica.

The spread from 270K to 87M users happened because back in 2015 Facebook also allowed developers to collect some information on the friend networks of people who used Facebook Login. In other words, even though just a single user may have agreed to hand over their data, developers could also access some data about their friends. Regardless of the fact that Facebook phased out this level of access a few years ago, the repercussions have carried on for much longer.

So up to this point, we know that key factors to the scandal are that Facebook has technology allowing third-party access – Facebook Login – and some way of branching out to and collecting data from a user's network accessible to a 3rd party.

Now turning to the Valuenex Radar, let's take the search to the next level. With just a glance at their 2,826 patent landscape, it is immediately apparent that Facebook has some clear focus areas with their R&D and patent filings (fig. 1).

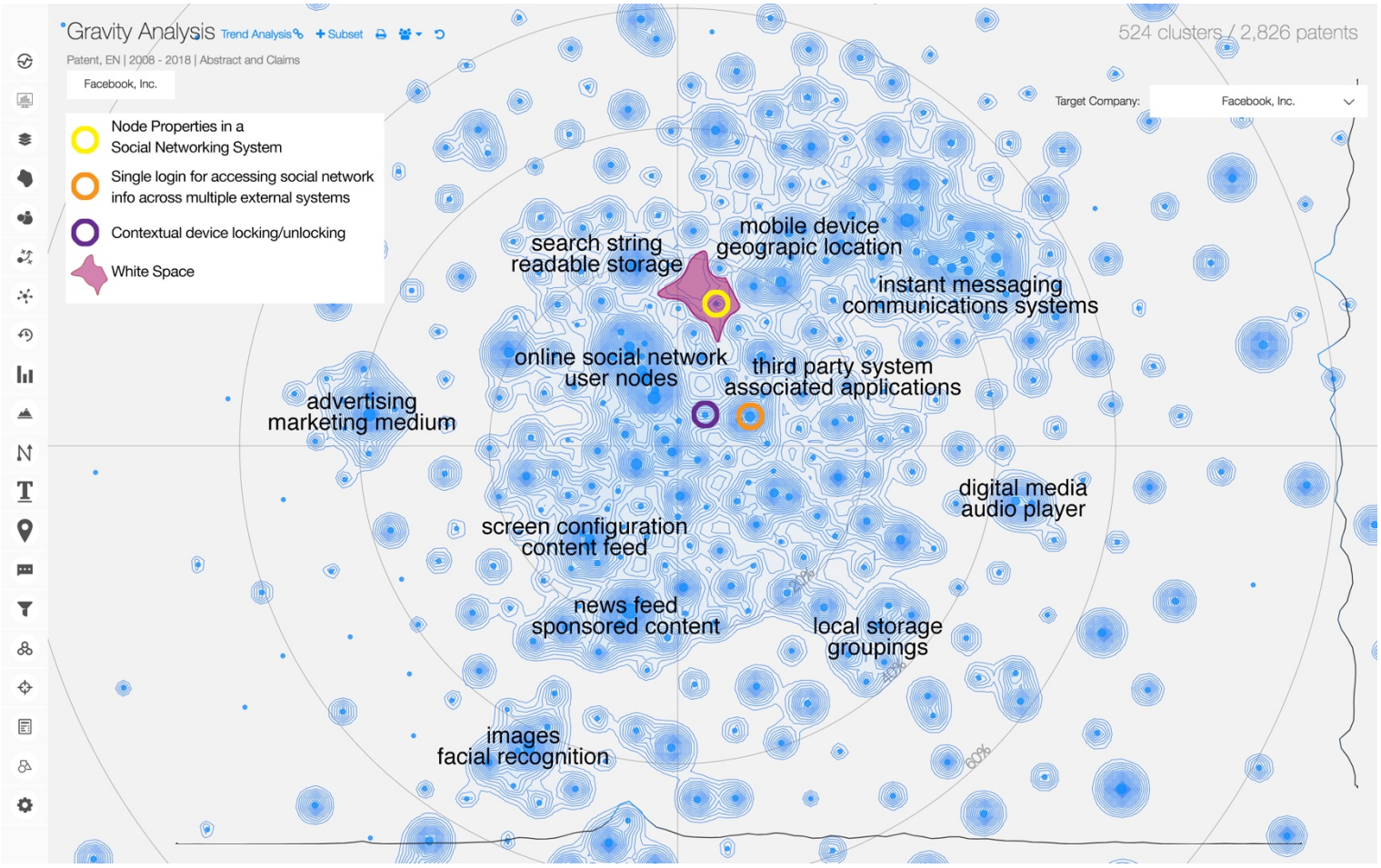


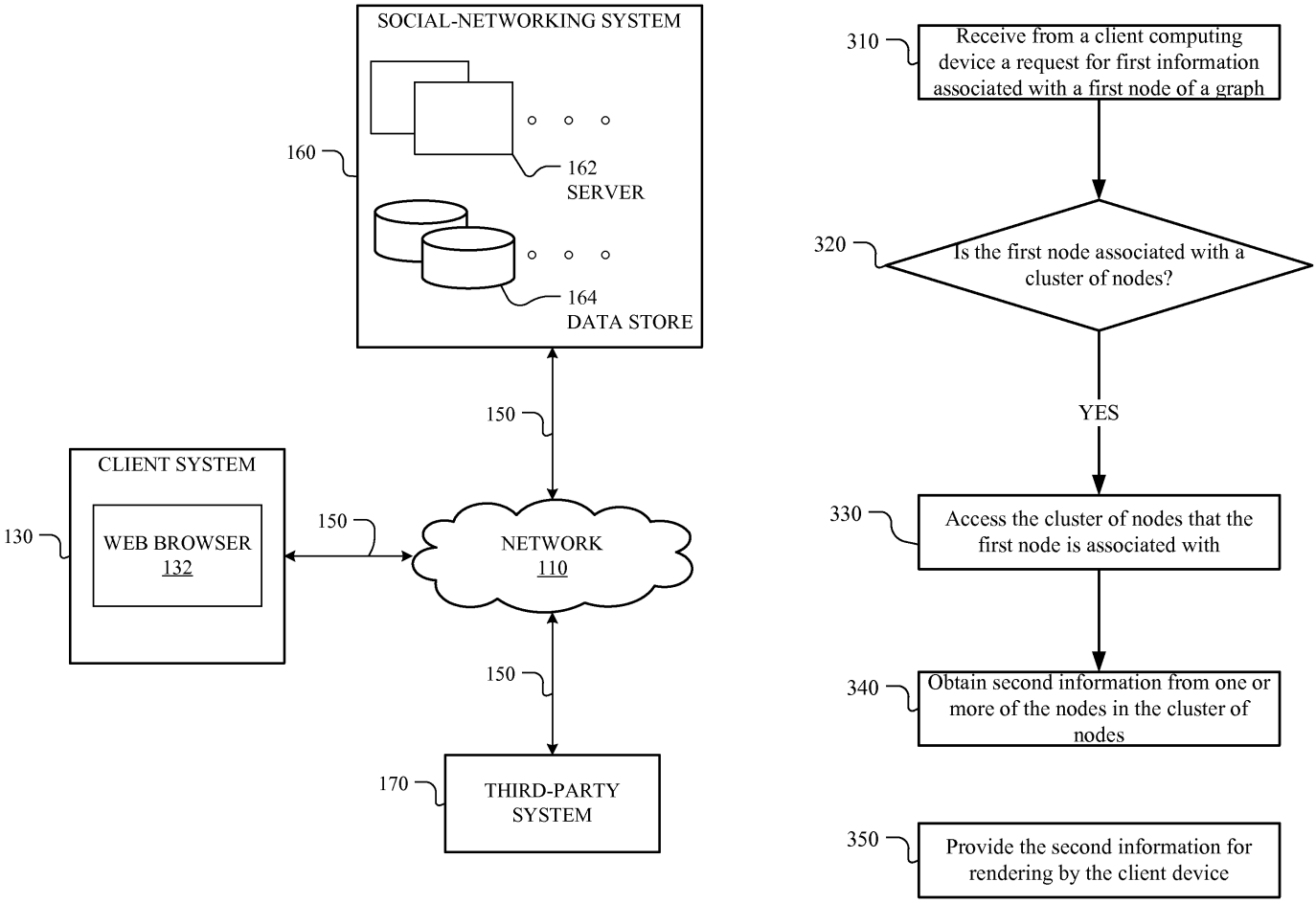
Figure 1 · Facebook Patent Landscape

Mostly as expected, we can see high density areas around advertising, news feeds, facial recognition, digital media, etc. But where things get really interesting is with what is in the *white space* that was automatically detected between high density areas related to online social network, readable storage, geographic location, communications systems, and third party systems. White spaces usually mean there is some gap within a technology portfolio because they don't have the expertise to develop that area, have trade secrets they are trying to hide, or haven't realized this void in their portfolio. On the other hand, if there is at least some level of technology in that space, it can be seen as a 'connector' to bring together the technologies that surround it. In this case, it might be a combination of factors, but the latter seems to be the most likely.

Within the white space was patent 20150113060 titled ***Node properties in a social-networking system*** (fig. 2).

Abstract

In one embodiment, one or more server computing devices receive, from a client computing device, a request for first information associated with a first node of a graph. The one or more server computing devices determine whether the first node is associated with a cluster of nodes. A cluster of nodes includes one or more concept nodes of the graph that are related to each other. When the first node is associated with a cluster of nodes, the one or more server computing devices access the cluster of nodes that the first node is associated with, obtain second information from one or more of the nodes in the cluster of nodes that the first node is associated with, and provide the second information for rendering by the client computing device.



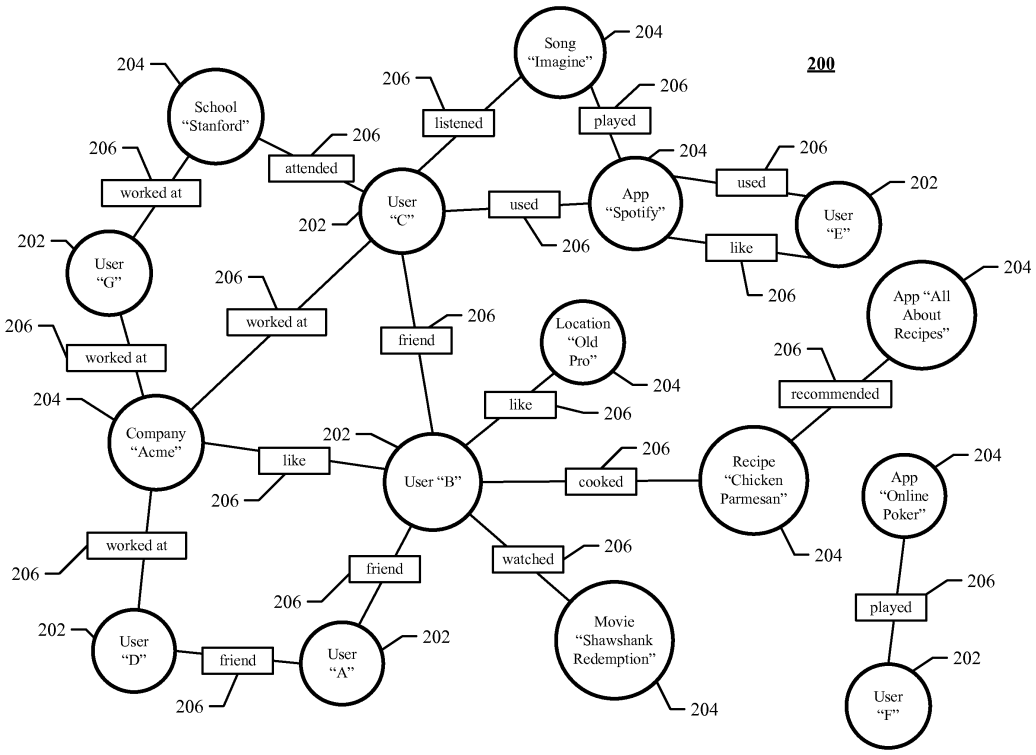


Figure 2 · Patent 20150113060 – images 1-3

So now we have pinpointed (or rather the algorithms did it for us) the technology behind reaching from individual users to their friends and setting the basis for 3rd party access to the data positioned right in the middle of Facebook's core areas – online social network, readable storage, geographic location, communications systems, and third party systems – seeming to connect the capabilities of each.

But what about the Facebook Login app that allowed the exploitation to occur? After filtering out patents based on terms *login – third party – access – nodes*, near the high-density areas of *third party systems* and *user nodes*, several patents emerged including one dating back to 2009, 9306927 **Single login procedure for accessing social network information across multiple external systems** (fig. 3), which summarizes as:

"Social network information can be highly valuable to an external system that does not or cannot maintain such information about its users. Embodiments of the invention allow multiple external systems to access social network information, while also providing a mechanism for managing the login status of a user of these multiple systems.

This frees the user from the need to maintain authentication information separately for each external system and allows external systems to leverage social network information for various purposes."

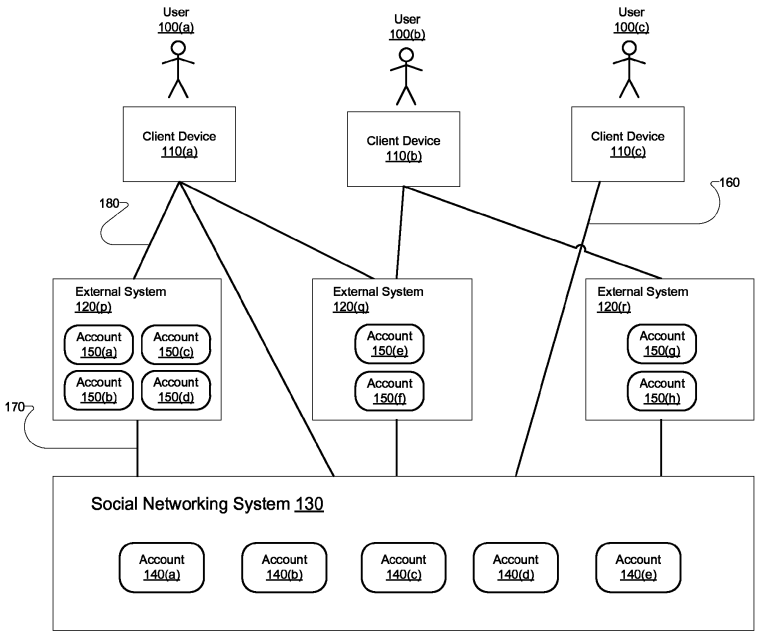


Figure 3 · Patent 9306927 – image 1

Nearby we also found patent 8914875 **Contextual device locking/unlocking** (fig. 4). On the surface this one seems pretty harmless and more about securing users data, but we know there is more to it because it popped up on the radar with the filters applied. Upon further investigation, buried deep in the claims it states:

*"In particular embodiments, a social-networking system **460**, client system **430**, or third-party system **470** may access social graph **500** and related social-graph information for suitable applications. The nodes and edges of social graph **500** may be stored as data objects, for example, in a data store (such as a social-graph database). Such a data store may include one or more searchable or queryable indexes of nodes or edges of social graph **500**."*

Have we found the smoking gun? Seems so. We have uncovered the technology enabling data pull from from a user's network, third party access to the user's data, and third party access to the user's friends' data.

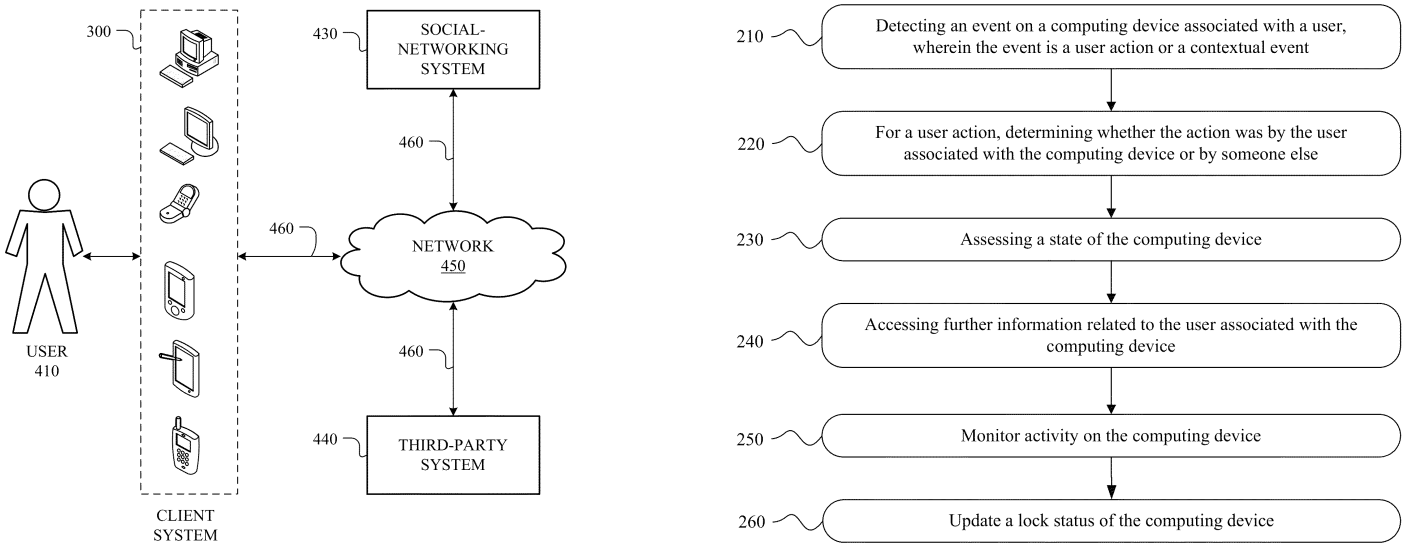


Figure 4 • Patent 8914875 – images 2-3

Taking a little pressure off of Facebook and its business of data sharing, major platforms like iOS and Android allow developers to collect user's contact lists with permission. On top of that, Twitter, Google, and LinkedIn have login features similar to Facebook, potentially meaning users data could be exploitable from there as well.

But what about other types of major companies that have troves of data related to modern life? Based on the books you buy, hygiene products you order, where your shipments are delivered, or how you interact with their hands-free smart speaker *Echo*, you can be sure that Amazon knows your demographics and psychographics better than your neighbor does.

To see what Amazon might be doing and where their technology lies compared to Facebooks', we added Amazon's technology portfolio of 11,108 patents from 2008 to April 2018 to the analysis, creating a combined total of 1,617 clusters precisely grouped and plotted based on full text semantic similarities (fig. 5). With Amazon's patents colored in purple and Facebook in red, we can quickly see that Amazon is active in areas that Facebook isn't, namely electromagnetic energy, audio signals, financial payments, DNS servers, and inventory management systems. This all make sense based on the nature of their business, but there clearly are several areas of overlap in the user node and geographic location areas.

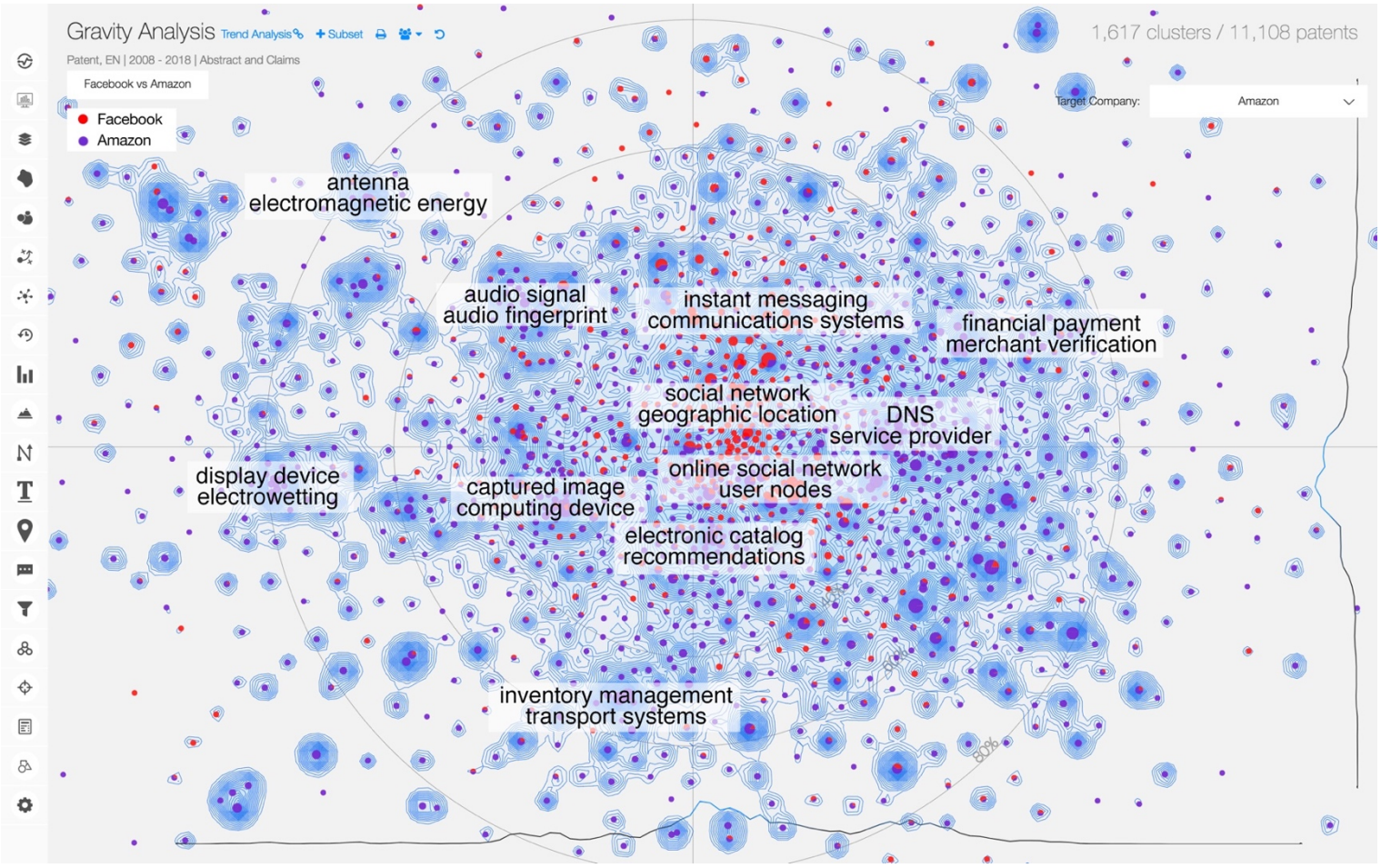


Figure 5 · Facebook and Amazon portfolios - analyzed together

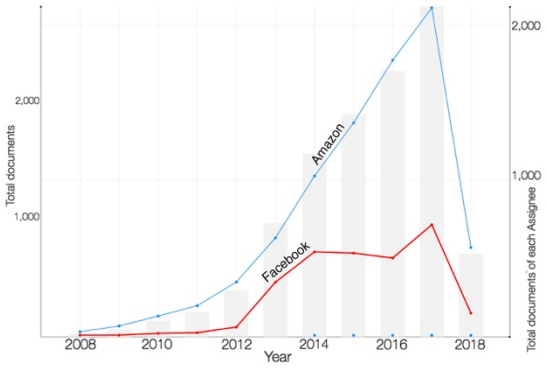


Figure 6 · Patents from Jan. 2008 to Apr. 2018

Because of the expansiveness of Amazon's portfolio and total patent volume compared to Facebook (fig. 6), we were hoping to cut to the chase by pinpointing Facebook's patents that are guilty of facilitating the data exploitation and see what Amazon has in that same area (fig. 7). However, while some Amazon patents were placed in or positioned near the targeted clusters, they were more related to Detection of and response to network attacks, Containerized examination of sensitive data, and Techniques for capturing data sets.

So far no dirt.

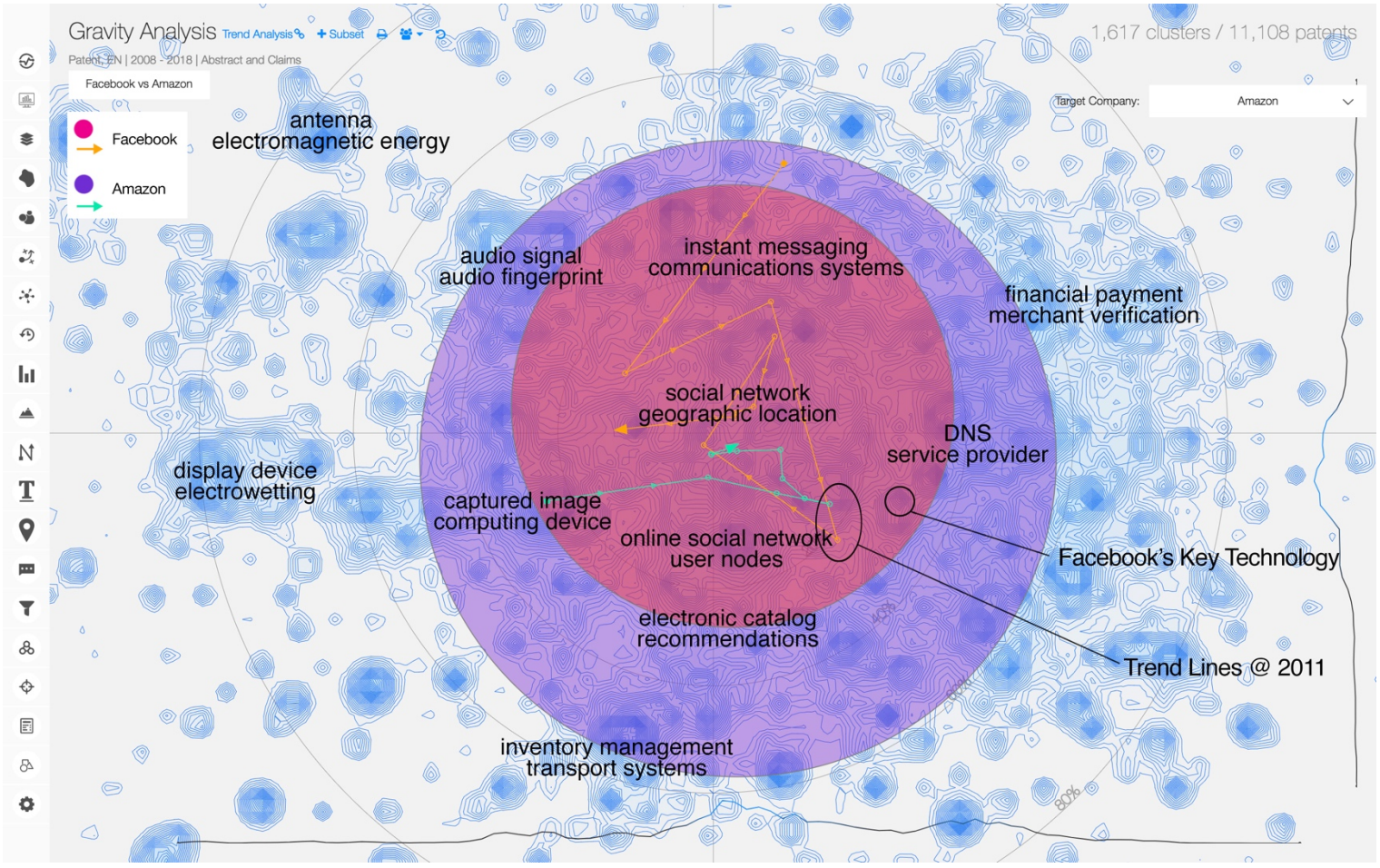


Figure 7 · Centers of Gravity and Trend Lines

Approaching from another angle, we removed the clusters and activated the Center of Gravity function to show the scope of core technology overlap of the two portfolios, which shows Facebook completely encompassed within Amazon's reach – not too surprising after seeing the overall diversity and breadth of Amazon's patent coverage. Then we added Trend Lines to map the movement of their core activity from 2008-2018 to reveal where their R&D paths might have crossed (fig. 7-8). After extracting the relevant patents where their coverage and paths *nearly* converged – mainly from 2011 – Amazon still came out clean. In fact, Amazon's trend line never got very close to the area containing Facebook's technology of focus.

Not stopping there, we filtered out and checked dozens of patents based on key word related to data sharing, 3rd party data access, node access, unrestricted access, network pull, related users and the like. But once again, no significant or threatening patents connected to Amazon came up.

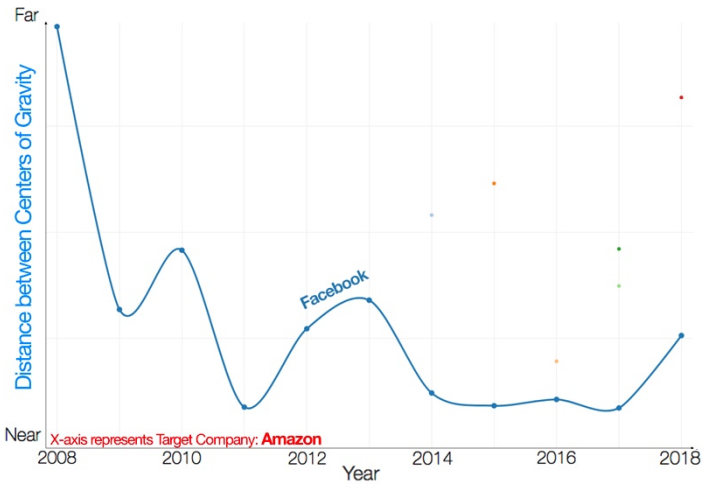


Figure 8 · Center of Gravity Transition - Horizontal View

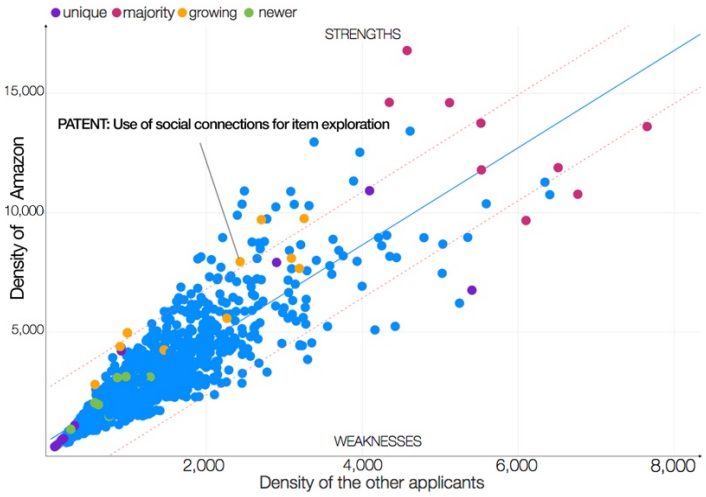


Figure 9 · Scatter Graph of Amazon and Facebook

In the final approach, a look into the scatter graph of Amazon - showing the correlation of coverage density of Amazon's technology compared to Facebook - didn't reveal what we were digging around for (fig. 9), but we did come across a patent that shows Amazon is also well aware of, and tapping into, the commercial value of accessing a user's friends within their social network: 9092816 titled **Use of social connections for item exploration** (fig. 10). The claims state:

"In one embodiment, a user having a user account with an electronic commerce system may have developed one or more relationships with various contacts and/or friends through a social networking account. As such, the user's social

networking account may be accessed in order to obtain information about each of the user's contacts and/or friends."

"For instance, a user having a Facebook® account may have hundreds of friends by submitting friend requests to other users having a Facebook® account.

*To this end, a user may wish to link their user account **140** to their social networking account. To do so, a user may submit a request to the electronic commerce system **115** through network **109**. In one embodiment, the user may provide security information in the request that may be used in order to access their social networking account, such as, for instance, a username, passwords, and so on. In another embodiment, such security information may have been previously provided and stored in association with a user's user account **140**."*

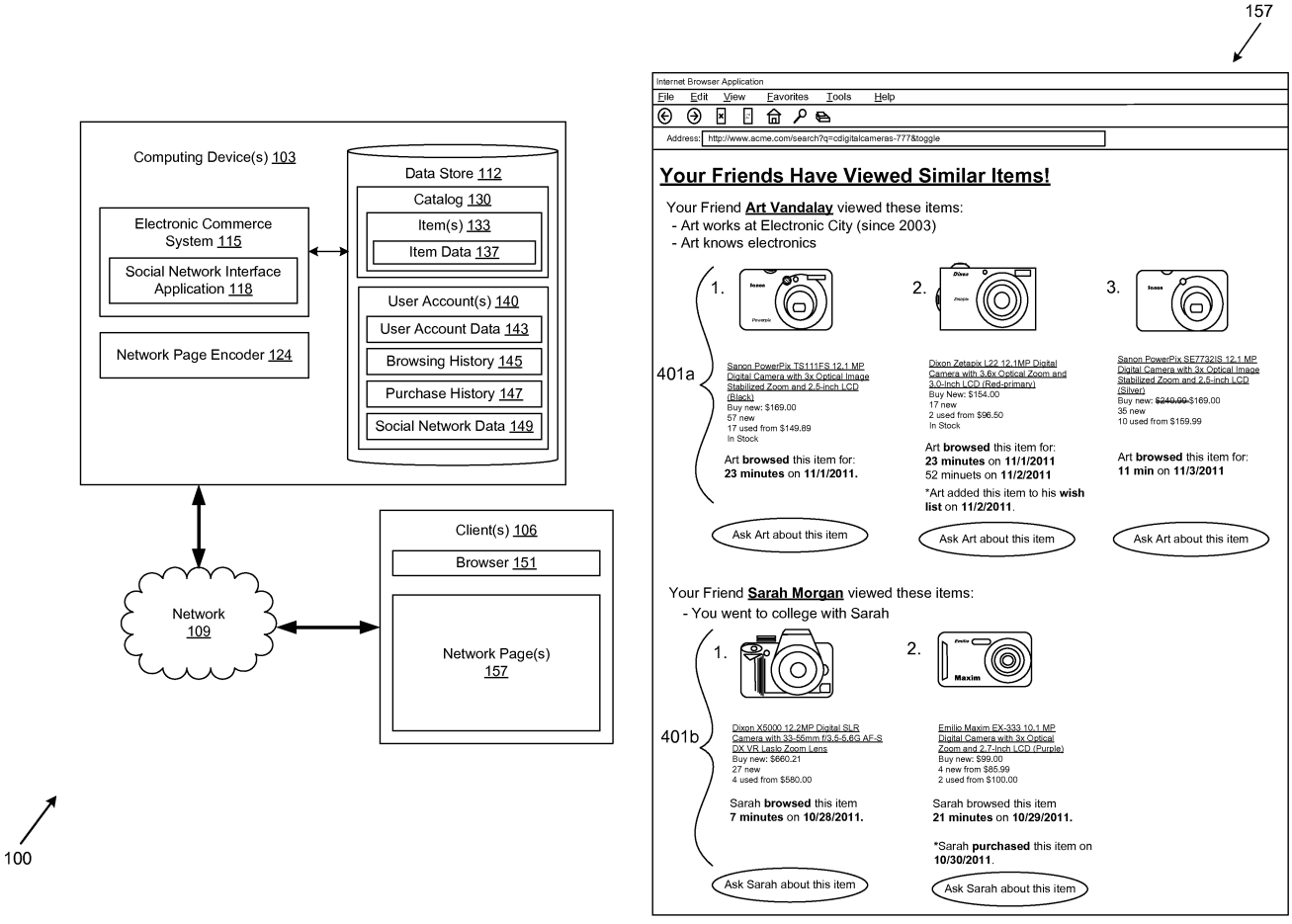


Figure 10 • Patent 9092816 – image 1

In short, this allows for Amazon to access details about friends of a user on Facebook (if that user has used Facebook Login), and then tie that information to their own user network in order to promote products. And seeing as how this patent dates back to 2011, Amazon has been actively engaging in extended social network access for their own benefit for some time.

So even though Amazon seems to be keeping its user data locked tight, they are also one of the 3rd parties taking advantage of data from Facebook users' personal networks for their own gain.

In the end, we now understand the technology that enabled the Facebook – Cambridge Analytica scandal, but where does that leave us? The straight answer is that because the tech is already out there and is so intertwined with the way we access and manage our personal data, other measures need to be taken to ensure that it doesn't happen again. Simply deleting all of your social network, online shopping, and instant messaging accounts is not an option in modern day life. Facebook and the other numerous companies that have up to thousands of data points on each one of us need to bolster their R&D and tech around data security, especially related to interconnected networks holding everything from our bank and hospital records to shopping and charity habits. If they can create it to begin with then they can surely develop methods to control it effectively, providing a guarantee to the fundamental right to privacy and not treat it like a tradable commodity, all backed by tangible technology. Until then, we can only hope that the worst of the Facebook data breach has passed, and that this has been a wake-up call for all of us to be more vigilant with our online lives.

VALUENEX, INC.

Tel: +1 650 843 9010
Email: customer@valuenex.com
www.valuenex.net

April 2018