The JVWR is an academic journal. As such, it is dedicated to the open exchange of information. For this reason, JVWR is freely available to individuals and institutions. Copies of this journal or articles in this journal may be distributed for research or educational purposes only free of charge and without permission. However, the JVWR does not grant permission for use of any content in advertisements or advertising supplements or in any manner that would imply an endorsement of any product or service. All uses beyond research or educational purposes require the written permission of the JVWR. Authors who publish in the Journal of Virtual Worlds Research will release their articles under the Creative Commons Attribution No Derivative Works 3.0 United States (cc-by-nd) license. The Journal of Virtual Worlds Research is funded by its sponsors and contributions from readers.
Using Google Trends Data to Gauge Interest in Virtual Worlds

Anthony Crider
Elon University

Jessica Torrez-Riley
Elon University

Abstract

Virtual worlds rose and fell in popularity a decade ago, and today's nascent commercially-available virtual reality could repeat this pattern. With sparse data available for gauging interest in technology products, such as virtual worlds or virtual reality, Google Trends search popularity has been used in prior studies as a proxy for global interest. We explore the problems with this approach using data from three virtual worlds: Second Life, Minecraft, and World of Warcraft. We find that Google Trends search volume does not correlate with user purchases or subscriptions, and the single shifted Gompertz function used in prior studies may not be sufficient to model both product user searches and searches driven by media attention.
1. Introduction

1.1 Popularity of Virtual Worlds

Virtual worlds are predicted to see resurgence in popularity with the advent of commercial virtual reality (Perry 2016). Three head-mounted virtual reality units launched commercially in 2016: the HTC Vive in April, the Oculus Rift in May, and the PlayStation VR in October. Also in October of 2016, Mark Zuckerberg revealed a demonstration of the Facebook social VR prototype to the audience at the Oculus Connect conference. Some aspects of the demonstration were integrated with Facebook (e.g. posting of a selfie) but most of the prototype was very different from anything current Facebook users experience. Several other social virtual reality virtual worlds also debuted in 2016: AltspaceVR, vTime, VRChat, Rec Room, and High Fidelity (from Second Life creator, Philip Rosedale). The company behind Second Life, Linden Labs, is also preparing to launch its own VR platform, currently known as Project Sansar. This increased activity has resulted in a new round of media hype and public attention to virtual worlds. But how does public interest relate to actual use of virtual worlds? To answer this, we review the history of social virtual worlds, the attention they garner from users, and their user bases.

Linden Lab released Second Life (SL) to the public in 2003, but the preceding year, Time Magazine had already named the 3D virtual world technology as one of the best inventions of the year (Time, 2002). It was not until the fall of 2007 that social virtual worlds reached into the public consciousness in a significant way. Many Americans saw SL for the first time either in an episode of CBS's crime drama CSI:NY (“Down the Rabbit Hole”) or on the hit NBC sitcom The Office (“Local Ad”). In the decade following, SL experienced a rollercoaster of relevance beginning with this initial boom of excitement, followed by a dramatic peak in usage and a subsequent slow but steady decline in vitality.

Media hype has been blamed in part for SL’s up-and-down history. At first mainstream outlets presented the virtual world as the future of the web praising Second Life’s features and technical capabilities (Marshall, 2011). But once SL became known for illicit user behavior and gambling operations, the media coverage became more negative. The initial promise of Second Life had worn off. The SL story arc concluded as a world once overflowing with new users and lands began to be described as “desolate” (Collins, 2010).

1.2 Gartner Hype Cycle

Originally introduced in 1995, the Gartner Hype Cycle is a pattern that characterizes the lifecycle of many technology products (Fenn & Linden, 2005; Gartner, 2017). The cycle begins with a high peak of inflated expectations, immediately followed by a sharp decline into disillusionment ("the trough of despair") brought about by negative experiences with the technology, and ends with a gently sloping – but longer – period of enlightenment and productivity. Second Life seems to follow this curve closely. "Virtual worlds" first appeared in the Gartner reports in July 2007 at the “Peak of Inflated Expectations” and then sat securely in the “Trough of Disillusionment” from 2008 until 2012. In 2013, the spot for “virtual worlds” was replaced with “virtual reality”, likely due to the debut of the Oculus Rift Developer Kit 1 in March of that year. While virtual worlds and virtual reality are distinct from one another (Bell, 2008; Boellstorff, 2015), this switch is indicative of a general blurring of the lines between the two concepts in the public mind. Gartner slowly moved virtual reality towards the “Slope of Enlightenment” from 2013 to 2016.

1.3 Lack of Data

How closely does our description of the lifecycle of Second Life mirror the real data about usage and popularity of the product? How closely does the Gartner Hype Cycle description of
Second Life match the quantitative data on virtual world use? In some cases, we are able to examine product usage statistics provided by the company itself, for example from press releases and financial reports. According to Linden Lab’s fourth quarter report in 2008, SL residents had logged nearly 400 million hours into the virtual world that year – a 60 percent growth over 2007. In 2009, the end-of-year numbers revealed a 21 percent increase to 481 million logged hours – however this growth would signal the peak of rapid Second Life usage increase (Linden, 2010b). It is unclear whether these increases were from a user base that was also growing, or whether the same number of users was simply logging more hours in the system. In 2010, the company reported that logged hours fell 10 percent, and plateaued into 2011 (Linden, 2011a). Second Life no longer shares these usage statistics, perhaps indicating that their numbers have only continued to decline (Tateru, 2012).

1.4 Other Virtual Worlds

Part of Second Life’s decline as a social network rests in its inability to grow its user base, something that other social networks overcome by focusing on non-social components (Viswanath, Mislove, Cha, and Gummadi 2009). Virtual worlds with additional, non-social components have also succeeded. One of the most popular video games of all time (second only to the 1984 puzzle game, Tetris) is Minecraft. With many public realms hosted by Mojang or others (e.g. Mineplex), Minecraft clearly meets Bell’s definition (2008) of a “synchronous, persistent network of people, represented as avatars, facilitated by networked computers.” Is it a social virtual world comparable to Second Life? Unlike many other virtual worlds, Minecraft does not strive for graphical realism. Both the world and avatar are composed of large blocks with course textures. The driving mechanisms for the game are survival and construction, not communication (Duncan, 2011). In May 2009, Markus “Notch” Persson began work on the “Cave Game” that would eventually become Minecraft. During the first year of development, Persson began adding multiplayer ability (June) and survival mode (August). In August 2010, he released a survival/multi-player version (1.0.15) that allowed Minecraft to connect to other user-hosted servers (Minecraft Wiki, 2013). While survival mode already existed in single-player format and multiplayer options existed, this new version quickly gained popularity. In Figure 1, we found that during that month, the Google Trends search volume for “minecraft” increased a hundred-fold. This mirrored a similar spike in popularity for Facebook in 2004. In June 2013, the survival mode multiplayer server was by far the most popular. Of the top 40 servers, 39 are in survival mode, 1 is in adventure mode, and none are creative mode (Minecraft Server List, n.d.). The peak number of users on fourteen of the top Minecraft servers (data from Minetrack) is comparable to the ~90,000 daily concurrent users Second Life had at its peak (Tateru, 2012).

A third virtual world frequently studied by researchers is World of Warcraft. While World of Warcraft is not a social virtual world, we examine it here to highlight one other flaw in relating search volume to global interest. Blizzard’s Warcraft series of games has existed since 1994. However, it is World of Warcraft that has dominated this brand and led to early research studies of MMORPGs (Yee, 2006). Launched in 2004 with a monthly subscription, it grew to 12 million concurrent subscribers in October 2010. Even in 2015, over a decade after its launch, it still had 5.5 million subscribers.

2. Methodology

2.1 Google Trends

To examine the relative use and popularity of social virtual worlds when no company statistics are provided, researchers have turned to another source: Google Trends. This tool has been used to make successful predictions in diverse areas such as software engineering (Rech, 2007) and epidemiology (Carneiro & Mylonakis, 2009). Research has shown that Google Trends can outperform surveys in predicting consumer behavior (Vosen & Schmidt, 2011). In this section, we
will critically examine a study by Bauckhage, et al. (2014a, 2014b) that used Google Trends data to gauge the public interest in 175 social media products (e.g. Facebook, YouTube, Twitter), including some social virtual worlds.

2.2 Shifted Gompertz

Everett Rogers’ highly-cited 1962 book, Diffusion of Innovation, qualitatively describes the adoption of new technologies by innovators, early adopters, the early majority, the late majority, and laggards (Rogers, 2010). Influenced by Rogers’ book, Bass (1969) proposed a partial differential equation to quantitatively model the diffusion of consumer durables. The Bass model is expressed as:

$$f_B(t|p, q) = \frac{(p + q)^2}{p} \frac{\exp[-(p + q)t]}{\left(\frac{q}{p} \exp[-(p + q)t] + 1\right)^2}$$

where \(t\) is time, \(p\) models a propensity for innovation and \(q\) models a propensity for imitation. This function has been used to model sales over time and has also been generalized to study the diffusion of things besides durable goods.

In studying the diffusion of social media usage, Bauckhage, et al. (2014a, 2014b) assumed that Google Trends popularity was a proxy for “collective attention.” To model and predict changes in collective attention, they tested the effectiveness of three different diffusion models in fitting Google Trends data: the Bass model (Bass, 1969), the shifted Gompertz model (Bemmaor, 1992), and a third function also used in diffusion studies, the Weibull model (Rinne, 2008). They found the shifted Gompertz distribution provided the best goodness-of-fit. The shifted Gompertz function is expressed as:

$$f_{SG}(t|\beta, \eta) = \beta e^{-\beta t} \exp(-\eta e^{-\beta t})(1 + \eta(1 - e^{-\beta t}))$$

where \(t\) is time, \(\beta\) and \(\eta\) are parameters describing respectively the scale and shape of the function, and \(t, \beta, \eta \geq 0\). For this study, we will focus on the shifted Gompertz function, though fitting with any of these should not impact our conclusions.

2.3 Example of Facebook

As an example of both the power and limitations of the Bauckhage, et al. method for understanding the popularity of a social platform, we briefly look at Google Trends data for Facebook. In their papers, Bauckhage, et al. (2014a, 2014b) used the shifted Gompertz model to correctly predict that in 2017 the Google Trends popularity of Facebook would fall to 50% of its 2013 peak value. The Weibull model seriously under-predicted the 2017 popularity (half of the real value) and the Bass model over-predicted the popularity (three times higher than the real value). However, the Bauckhage team's assumption that Google Trends data is a proxy for the “collective attention” to Facebook is easily disproven with additional data. In Figure 1, we show the relative Google Trends popularity of the search term “Facebook” (with a peak at 100%) along with the Monthly Active Users reports in Facebook quarterly reports (Facebook, 2016). While the Google Trends popularity has indeed dropped to 50% of the 2013 value, this drop is only directly correlated with a decline in Facebook web (browser) users. The increase in mobile users of Facebook has continued to drive up the actual collective attention to and popularity of the product.
Figure 1 reveals the first systematic error in equating Google Trends popularity with collective attention; social media products accessed via a web browser may have inflated Google Trends popularity scores. We hypothesize that this is due to users accessing the search term "Facebook" in their browser address bar, and subsequently invoking a web search instead of what they actually intended, which was to visit the site itself. By contrast, users of social media products accessed via mobile apps or non-browser applications will only add to the Google Trends popularity when managing the account via a browser, or doing ancillary searches related to the product.

Another issue that arises when using Google Trends data is the indexed relative “popularity” figures reported by Google. Each Google Trends keyword search yields a distribution with the highest search volume scaled to 100% and all smaller values rounded to the nearest integer. This precludes easy comparison of products with very different user bases, or time-series of analysis of products that substantially change in popularity over time. To address this problem, we introduce pseudo-logarithmically spaced calibration words. Our calibration words show minimal change over time. We use the average value of these words over one decade (2004-2014) to establish baselines values. We choose the average popularity of the word “cats” to be our standard unit measurement of popularity. The table below shows the baseline value of seven words in units of “cats.”
Table 1: Google Trends Search Volume for Calibration Words

<table>
<thead>
<tr>
<th>Calibration Word</th>
<th>Google Trends Search Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>free</td>
<td>32 cats</td>
</tr>
<tr>
<td>sex</td>
<td>11 cats</td>
</tr>
<tr>
<td>cats</td>
<td>1 cats</td>
</tr>
<tr>
<td>cereal</td>
<td>0.17 cat</td>
</tr>
<tr>
<td>trombone</td>
<td>0.052 cats</td>
</tr>
<tr>
<td>follicle</td>
<td>0.020 cats</td>
</tr>
<tr>
<td>cumulonimbus</td>
<td>0.0028 cats</td>
</tr>
</tbody>
</table>

In Figure 2, we show the relative popularity of the search term "Facebook" in units of “cats” plotted on a logarithmic scale. This reveals the 10,000-fold increase in the Google Trends popularity of Facebook over one decade, something that would not be visible with a single Google Trends query. We also compare the relative Google Trends popularity of two other notable virtual worlds: Second Life and Minecraft.

Figure 2: Google Trends popularity of Facebook plotted to show large changes.

3. Results

3.1 Second Life

Despite the decline we described earlier, Second Life is likely the most popular social virtual world of the last decade, excluding Minecraft. To confirm this, we compared its Google Trends popularity to several other social virtual worlds. While Google Trends allows international searches, to study search terms that vary rapidly over time (e.g. “Facebook”) we must calibrate using relatively constant terms (e.g. our list of calibration words). To limit impacts due to multiple possible linguistic interpretations of the calibration words, for this portion of the study we only use Google Trends data from the United States. Our search volumes for terms of interest were scaled with respect to the word “cat” since it showed minimal variation over the span of a decade (6%).

Using these words to examine the peak search volume of several direct competitors shows the relative prominence of Second Life in the social virtual world market (Table 2). Since Google Trends provides relative but not absolute search volumes, we use the 2004-2014 average search volume for “cats” as our standard unit of measure. Given the low search volume for social virtual world keywords, in Table 2, we use millicats or 1/1000th the average search volume for “cats” using Google Trends. For comparison, “minecraft” has a peak search volume of 8,300 millicats and “warcraft” has a peak search volume of 2,900 millicats. From this data, we conclude that indeed none of these exclusively social virtual worlds surpassed Second Life in interest.

<table>
<thead>
<tr>
<th>Virtual World</th>
<th>Google Trends Search Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>Second Life</td>
<td>290 millicats</td>
</tr>
<tr>
<td>Sims Social</td>
<td>100 millicats</td>
</tr>
<tr>
<td>OurWorld</td>
<td>68 millicats</td>
</tr>
<tr>
<td>Habbo Hotel</td>
<td>46 millicats</td>
</tr>
<tr>
<td>Google Lively</td>
<td>27 millicats</td>
</tr>
<tr>
<td>Blue Mars</td>
<td>8 millicats</td>
</tr>
<tr>
<td>Active Worlds</td>
<td>3 millicats</td>
</tr>
<tr>
<td>OpenSim</td>
<td>3 millicats</td>
</tr>
</tbody>
</table>

Bauckhage, et al. (2014a, 2014b) focused on global differences in adoption and diffusion times for various social media products. These international differences are clearly evident when examining Google Trends data for Second Life. Figure 3 shows the global search popularity in three countries: Italy, Brazil, and the United States. While the data for Italy mimics the global search volume, the Second Life curve in the United States appears to be one broad, smooth curve plus a “hype phase” in 2007. The small international spikes in February 2007 are prominent in Brazil and non-existent in the US, corresponding to a push by Linden Labs and two Brazilian partners to launch a Portuguese-language client (Marque, 2007).
Figure 4 shows changes in the Google Trends popularity of Second Life over time within the United States. Bauckhage, et al. (2014a, 2014b) used a single shifted Gompertz function to model this data. While they assumed that different countries would have different diffusion curves, they do not explore how a single country might also have multiple diffusion curves. We found that two shifted Gompertz functions were far superior in fitting the Google Trends popularity of Second Life within the United States. One corresponds to a slow rise and slow fall in popularity over 13 years, with a fitted $t_0 = \text{May 2002}$. This is halfway after the creation of the first user within SL in March 2002 and the launch of the SL public beta in October 2002. The peak of this component is 2009 which corresponds to the peak in reported concurrent SL usage (Totetur). The second shifted Gompertz function shows a rapid rise in mid-September 2006 followed by a rapid decline over two years. This component coincides with the attention from media, including the Gartner reports mentioned above. For reference, a widely circulated article about the first Second Life millionaire, Anshe Chung, appeared on the cover of Business Week on May 1, 2006. We conclude that a single shifted Gompertz does not accurately portray the collective attention to Second Life.
Figure 4: Google Trends (GT) popularity of Second Life within the United States.

Two summed shifted Gompertz functions are fit to the data.
The first component ($\beta = 0.32$, $\eta = 7.7$, $t_0 = 2002.40$) shows a slow rise and slow decline in popularity.
The second component shows a rapid rise and rapid fall from mid-September 2006 to 2009.

3.2 Minecraft

Examining the Google Trends data for Minecraft in Figure 5, we see that a single Gompertz can fit six years of data (2011-2017) for “minecraft” searches. This might lead to the incorrect conclusion that interest in Minecraft peaked in 2013. However, in Figure 6 both the sales and the derivative of the sales show this is not true.

Figure 5: Google Trends popularity of Minecraft worldwide.

A single shifted Gompertz functions fits the rough shape of the data ($\beta = 0.45$, $\eta = 2.4$, $t_0 = 2011.04$).
Figure 6: Minecraft sales (in green) for PC/Mac worldwide.

There has been nearly linear growth from 2011 to 2017. The derivative of the sales (in gray) should correlate with changes in interest level. However, it is flat.

3.3 World of Warcraft

Figure 7 shows the Google Trends data for “warcraft” from 2004 to 2015. The search volume for “world of warcraft” is similar but lower. We have omitted the year 2016 from our analysis since a World of Warcraft movie in the summer of 2016 led to a short spike in searches. Curiously, while the peak number of subscriptions was in 2010, the data shows a near linear decline in search volume worldwide after the launch of World of Warcraft in 2004. This points towards another clear difference between search volume and global interest. Even though World of Warcraft is played on a PC, its search volume is clearly not dominated by its user base, who may play this game ~25 hours per week (Billeux et al., 2013). We conclude that, even for PC/Mac-only virtual worlds, there is no direct correlation between search popularity and paid subscriptions, even though either of these may be considered metrics of “global interest.”
Figure 7: Google Trend popularity from 2004 to 2015 of the search term “warcraft” worldwide (blue line). The search peak coincides with the launch of the product and decreases nearly linearly over time. However, the number of World of Warcraft paid subscribers (green circles) peaked in 2010.

4. Discussion

4.1 Summary

What can be learned from about social virtual worlds based on their search popularity? Bauckhage, Kersting, and Restegarpanah (2014) argued that Google Trends data was a proxy for “collective attention” to 175 social media products, including the virtual world of Second Life. They also found a shifted Gompertz distribution, sometimes used to model diffusion of technology, could fit the Google Trends search popularity. We applied their methodology to three popular virtual worlds: Second Life, Minecraft, and World of Warcraft, and found two primary problems with their study:

- A single shifted Gompertz distribution does not account for spikes of interest or rapid rises which can dominate the time series, even within a single country, and
- Search popularity does not correlate with other measures of “collective attention” including purchased units or paid subscriptions.

Each of our examples illustrates that search popularity may be related to, but not correlated with, actual user interest. In the case of Second Life, there are at least two Gompertz distributions in the relatively smooth USA data. In the case of Minecraft, neither the purchased units (which increased linearly) nor the time derivative of purchases (which was nearly constant) mirrored the rise and fall of the Google Search popularity. In the case of World of Warcraft, there is a steady decline in search popularity over the lifetime of the product even though the peak in paid subscriptions occurred several years after product launch. In all cases, we conclude that while Google Trends data may contain valuable information, this data does not relate directly to either purchases or usage of the virtual worlds we examined.
4.2 Popularity of Social Virtual Worlds in Virtual Reality

While there is currently much hype about virtual reality adoption, it is still quite early in the adoption cycle for the three head-mounted units released in 2016. In spite of the findings listed above, it is tempting to use Google Trends data to compare, contrast, and predict the popularity of these devices. Looking at the 2016 and 2017 data (up until May 2017) shows all three have comparable Google Trends values with spikes for the PlayStation VR release date (October 13, 2016) and the Oculus Rift CV pre-order and shipment dates (January 6, 2016, and March 25, 2016). The significantly cheaper Samsung Gear VR, designed to work with a Samsung smartphone rather than serve as a stand-alone HMU, has a similar Google Trends popularity. However, the adoption of these platforms has been drastically different. In 2016, sales of the PlayStation VR were three times that of either the Oculus Rift or the HTC Vive. Sales of the Samsung Gear VR were almost six times higher than the PlayStation VR (Raskin & McQueen 2017). Thus, Google Trends does not adequately mirror the sales of virtual reality head-mounted units.

Looking instead at the most popular social virtual worlds in virtual reality, Google Trends popularities for four have already peaked and are in decline. VRChat peaked in February 2017, Rec Room in December 2016, AltspaceVR in April 2016 and vTime back in January 2016. All four may rise again (as Minecraft did in 2009-2010) or continue to decline (as Second Life has since 2008). This is especially true for the worlds with the most recent peaks. However, as we found in our long-term study of virtual worlds in the past decade, even once several years of Google Trends data are available, they may not reflect the “collective attention” for each social virtual world. This research study suggests that Google Trends data can supplement, but not replace, user data provided by software companies.
References


