The Anatomy of Large Facebook Cascades

P. Alex Dow
Facebook
adow@fb.com

Lada A. Adamic
Facebook
ladamic@fb.com

Adrien Friggeri
Facebook
friggeri@fb.com

Abstract

When users post photos on Facebook, they have the option of allowing their friends, followers, or anyone at all to subsequently reshare the photo. A portion of the billions of photos posted to Facebook generates cascades of reshares, enabling many additional users to see, like, comment, and reshare the photos. In this paper we present characteristics of such cascades in aggregate, finding that a small fraction of photos account for a significant proportion of reshare activity and generate cascades of non-trivial size and depth. We also show that the true influence chains in such cascades can be much deeper than what is visible through direct attribution. To illuminate how large cascades can form, we study the diffusion trees of two widely distributed photos: one posted on President Barack Obama's page following his reelection victory, and another posted by an individual Facebook user hoping to garner enough likes for a cause. We show that the two cascades, despite achieving comparable total sizes, are markedly different in their time evolution, reshare depth distribution, predictability of subcascade sizes, and the demographics of users who propagate them. The findings suggest not only that cascades can achieve considerable size but that they can do so in distinct ways.

Introduction

Social networks are highly adept at transmitting information. This function has only been enhanced by platforms such as Facebook. It is not only the opportunity to transmit information that has been enhanced by online social networks, but also the ability to study and, specifically, to track such transmission. Of particular interest have been large cascades, which would validate models of viral spread of both information and influence through social networks. However, evidence of such large cascades has been scant. Both explicit and implicit transmission routes have shown cascades to be shallow and highly fragmented in a number of different contexts, ranging from viral marketing (Leskovec, Singh, and Kleinberg 2006; Leskovec, Adamic, and Huberman 2007), to diffusion of virtual goods (Bakshy et al. 2011), and spread of news on platforms such as Twitter (Goel, Watts, and Goldstein 2012; Bhattacharya and Ram 2012). Counterexamples include email chain letters which have been measured to reach depths of hundreds of steps (Liben-Nowell and Kleinberg 2008), although they potentially also have substantial breadth (Golub and Jackson 2010). A specific instance of a copy-and-paste meme on Facebook similarly reached a maximum depth of 40 (Adamic, Lento, and Fiore 2012).

There are therefore what appear to be conflicting findings. On the one hand, large-scale aggregate analyses of content that is propagating in online media shows little or no evidence of viral spread. On the other there are instances where careful tracing of cascades through email headers, content structure, and a known contact graph, has yielded cascades that have achieved a non-trivial depth. What is more, these long cascades align with the common experience of encountering the same content several times over a period of time, suggesting that it has been circulating through social ties through non-trivial paths. What is lacking is an understanding of the extent of viral content and the particular dynamics that govern highly viral spread. Recent experimental work has made considerable advances in understanding how individuals influence one another to propagate information (Bakshy et al. 2012) and the relative importance of demographics, influence and susceptibility in the underlying social network (Aral and Walker 2011; 2012). However, while the individual decisions to share information are now better understood, the global cascades that may or may not result are still obscured.

In this paper we present a large-scale study of reshared photos on Facebook. We show that a small but significant portion of the content is spreading virally. With an in-depth analysis of two large cascades, each containing over a hundred thousand nodes, we demonstrate that the actual information cascade can be much deeper than what is typically measured. Namely, an individual can reshare information directly from the source, appearing to be creating yet another depth 1 branch of the cascade. However, a more careful tracing of how they arrived at the source may place them at a considerably greater depth. We also examine the overall characteristics of such cascades: their growth over time, the nodes' branching factor as a function of cascade depth and time, and the response of users to repeat exposure. We show that the two cascades appealed to different demographics, yet both spread through a substantial portion of the globe within 24 hours. The findings suggest that large global cascades do occur with non-negligible frequency thanks to platforms such as Facebook, and that understanding them is key to understanding how information is propagated in today's hyperconnected environment.

Data description

Our data consists of aggregated, anonymized actions pertaining to a random sample of 1 million photos that were uploaded from 1/13/2013 to 1/20/2013. The actions include reshares, clicks in News Feed, likes, and comments. Photos, as opposed to videos, links, and other shareable content, were selected because of their relative popularity and the fact that they are assigned a unique ID upon upload. This has both advantages and disadvantages. The disadvantage is that an identical photo uploaded anew will be assigned a different ID. The advantage is, however, that it is easier to track cascades originating from a single upload.

In addition to this large data set of photos, we also track two specific memes in detail. These two photos were chosen because of their wide reach and different nature. The first is a photo of American President Barack Obama hugging his wife Michelle, which was posted at 11:15pm ET on election night, November 6, 2012, shortly after the election was called in Obama's favor. The photo was posted on Obama's official Facebook page, which has tens of millions of followers, ensuring that it received immediate and wide exposure. As of February 2013, it had accumulated over 4.4 million likes, and was still being reshared. For the remainder of the paper, we will refer to this photo as OVP, for Obama Victory Photo.

The second is a humorous photo posted by a young Norwegian man, Petter Kverneng, of himself and his friend Cathrine. In the photo he holds up a sign saying that Cathrine will have sex with him if he gets one million likes (we name the meme MLM or million like meme), and asks the viewer to share and like the photo so that he can get "laid." The joke was on the users liking and sharing the meme, because Petter told ABC news (after the photo had received over one million likes): "It started as a joke, and it ended as a joke. Me and Cathrine are just friends" (Milano and Stern 2013). Because the second photo was posted by a user as opposed to a highly popular page, its growth was more organic, and provides an interesting point of comparison for the OVP meme. Table 1 presents aggregate statistics for the two memes in terms of the number of times they were reshared, liked, and commented on.

Table 1: Aggregate statistics for two large cascades. Comments and likes are aggregated over both the original photo and the reshare posts.

metric	OVP	MLM
reshares	618,015	150,759
reshares by pages	4,496	1,795
likes	7,395,719	1,818,563
comments	574,077	152,276

Photo resharing in aggregate

Only about 5% of uploaded photos are reshared. This is in part because many photos are of interest to a limited audience, e.g. just the users' friends, or are restricted to that audience through privacy settings. The photos that are reshared at least once are reshared an average of 14.8 times, though some enjoy much wider popularity than others (see Figure 1). The top 0.5% of reshared photos, those with over 500 reshares over the span of two weeks, account for 50% of all reshare activity. This small portion of photos still sums up to a large absolute number of photos each week which have spread in what we'll argue is a viral fashion. It is therefore of interest to study the nature of such cascades as they interact with Facebook's social graph.

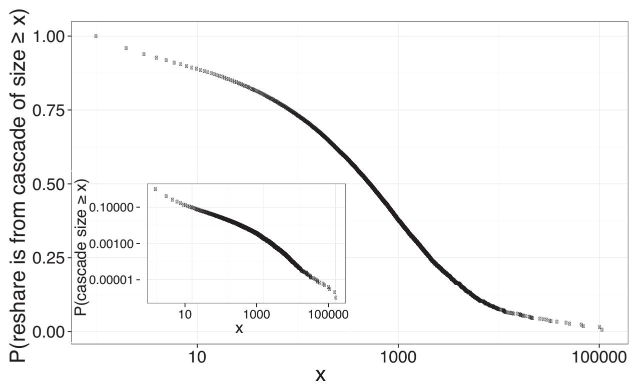


Figure 1: The proportion of reshare activity comprised of photos with at least a given number of reshares. Inset: Cumulative distribution of cascade sizes.

Cascade characteristics

In addition to the overall size of a cascade, we are interested in its shape, and in particular its depth. A truly viral photo would be transmitted from person to person to person, creating long chains. But prior work has rarely uncovered deep cascades. These studies have also typically relied on the direct credit attribution made by the users, typified by the presence of explicit ties – such as "reshare", "retweet" or "via" – between a piece of reshared content and its original instance. This means that unless a user credits the source that alerted them to the content, they may appear as having learned about the content directly from the origin. However, their choice to reshare the original content as opposed to crediting a secondary source may have as much to do with where they find themselves once they have consumed the content and are ready to share it, as with intentional choice of attributing credit. Therefore the use of these user-provided connections confuses the pathway of information diffusion with the social norms of content discovery attribution. Prior work (Gomez-Rodriguez, Leskovec, and Krause 2012) has aimed to infer diffusion cascades from timing information, rather than explicit attribution. However, perhaps because these inferred diffusion cascades are non-exact, and ground truth data has not been available, the shape and characteristics of such cascades have not been measured.

In this study we contrast the explicit credit attribution made by users to inferred information paths. The explicit attribution works as follows. User A uploads a photo. In the simplest case, user A simply adds the photo to his/her Timeline, although they can also attach the photo to a directed post to another user, group, or page. The Timeline is a collection of all the posts by and about the user. A's friend B might see A's photo if they visit A's Timeline, or if a story about A posting the photo appears in B's News Feed. The News Feed is a stream of stories about B's friends' activities, that is curated by an automated ranking algorithm. Most views of photos occur via the News Feed as opposed to Timeline views. When B sees the story (either on A's Timeline or in his/her own News Feed), if the photo is public, B can decide to reshare it. If user C sees the story of user B resharing, again either in her own News Feed or on B's Timeline, user C can click 'share', and a post is generated that optionally credits B with the story which is now also on C's Timeline. This is recorded as a reshare of depth 2, whose parent is B, and whose root node is A.

This explicit attribution information is potentially incomplete, because C may not reveal how she came to learn about A's photo. She could, for example, click on B's reshare, which takes her to A's post with the original photo upload, and from there share the photo directly from A. This places C at depth 1 from the root, making the cascade appear shallower and more fragmented than it really is. When this occurs one can rely on path indicators to reconstruct or "rechain" the cascade. For example if C clicks on B's reshare and shortly therafter reshares A's photo, one can rechain the cascade by moving node C from being directly attached to the root A and instead placing it as a child node of B. This process is naturally not without error. For example if C clicks on multiple reshares, it is unclear which one to attribute to. To minimize error, we look only at reshares occurring within an hour of a click. In subsequent sections, we use rechained diffusion paths to demonstrate that actual information flow often occurs through deeper paths than what is visible through direct attribution on platforms such as Facebook and Twitter.

Reshares that have been posted by users who have clicked on another user's reshare prior to sharing themselves are assigned the latter as a parent. This rechained path reconstructs the chain of influence, since the user encountered the information from a friend before passing it on, rather than independently arriving at the source. Similarly, we rechain shares from Facebook Pages if the user posting as the page – such as the page's owner – has clicked on another share before resharing from the source.

With the rechained data we calculate the depth of each node in the diffusion tree. We observe that in aggregate interest in content dissipates rapidly, both in terms of time elapsed since the photo is first posted and in terms of cascade depth (see Figure 2). However, many cascades achieve remarkable depth, and others sustain interest throughout the two week observation period. The fast fading of interest for most items is consistent with prior findings relating to collective attention in social media, e.g. Digg (Wu and Huberman 2007).

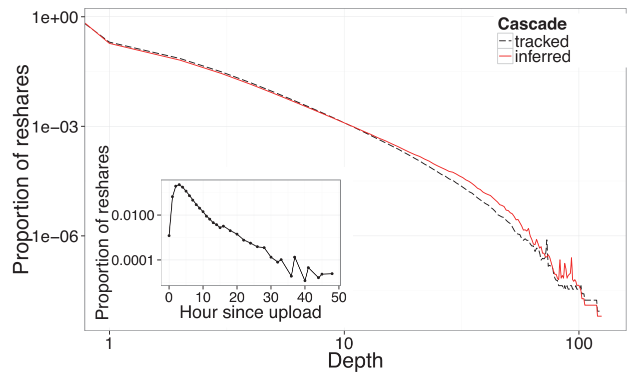


Figure 2: The proportion of reshares recorded at a given depth after applying the described rechaining process. Inset: the proportion of reshares as a function of time elapsed since upload.

Focusing specifically on photos that have been reshared at least 100 times, we find that while many have the bulk of their reshares occurring within a single step from the source, Figure 3 shows that there is a significant fraction for which the bulk of the reshares occurs deeper than one level into the cascade. That is, a significant portion are spreading virally, as opposed to being primarily broadcast from a popular source.

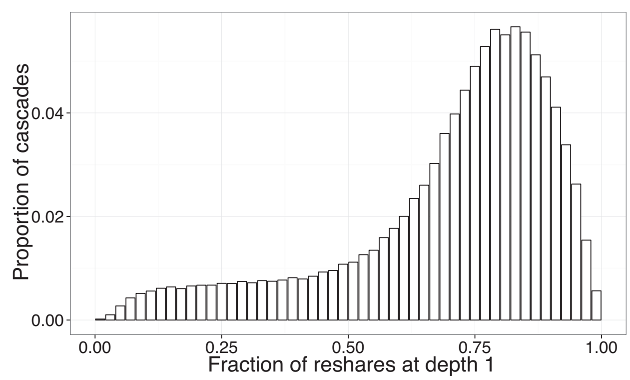


Figure 3: The proportion of reshares occurring at depth 1 for a sample of photos that have been reshared at least 100 times.

Large cascades

After establishing the distribution of cascade sizes and depth, showing that a small but significant fraction achieve wide and deep distribution, we study two large cascades, OVP and MLM individually. By focusing on these two memes, we can gain intuition about the different mechanisms by which a cascade can grow large.

From Figure 4 we see that rechaining for OVP has relatively little effect. As we will discuss in a later section, 50% of those sharing Obama's photo subscribe to his

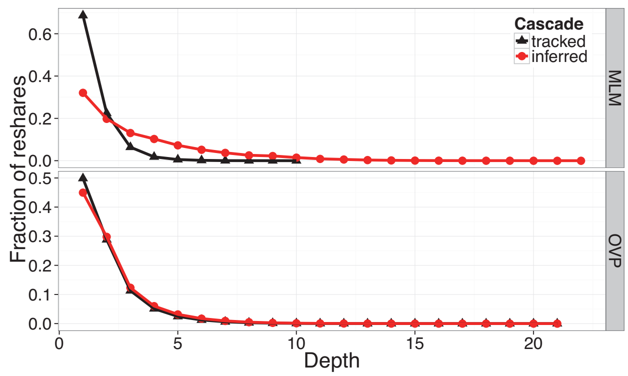


Figure 4: The effect of rechaining on the measured depth of cascades for the two memes

page, so they were likely directly exposed anyway. In contrast, rechaining the MLM meme produces a more dramatic change in the depth distribution, since it is unlikely that tens of thousands of individuals would have independently arrived at a user's Timeline and then decided to reshare the photo. Yet many nodes do remain at this depth after the rechaining procedure. This is because our information traces only cover a subset of all channels via which users can become aware of photos before resharing them. Some users arrive via links shared in email, through Facebook Groups or instant messages. Others arrive from external sources discussing the photo, among them popular news sources and various message boards. The close interaction between a social medium and external sources has previously been observed in Twitter (Myers, Zhu, and Leskovec 2012). Nevertheless, the minority of nodes remaining at depth 1 after rechaining shows that although some users chose to reshare the photo directly, most became acquainted with it through their social networks.

Rechaining not only shows us that one of the photos achieved highly viral spread, but it also serves as a cautionary tale about taking explicit attribution cascades at face value. At first glance, if one were to calculate depth using explicit attribution, one would conclude that MLM is predominantly shared from the root, and that it is an even shallower cascade than OVP. On the contrary, it is OVP that is top-heavy because of the very high connectivity of its root node – the Obama Facebook page. These findings suggest that at least some of the shallow cascades observed in prior studies may be artifacts of missing links in the true chain of influence.

In Figure 6 we visualize both cascades before and after rechaining. As described in Figure 5, reshares are represented by dots – here the color denotes the nature of the resharer, black for users and red for pages – and cascades are laid out radially around the original post. The distance of the dots to the center is a function of the time at which the reshare was posted. Starting with the original post, we divide the space by allocating to each reshare an angle proportional to the fraction of all reshares that are its descendants. We then recursively repeat this process to position every reshare on the plane, processing each child node in chronological order, which causes the layout to adopt a spiral look. For readability's sake and because the layout is already constrained by the cascade's tree structure, we do not explicitly draw the diffusion edges.

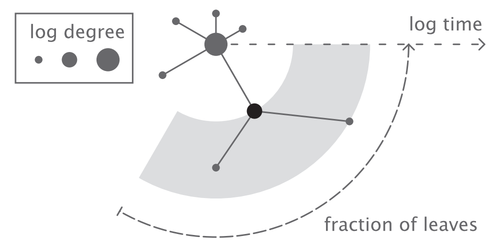


Figure 5: Schematic representation of radial visualization. Each reshare is allocated a distance to the center based on the time it was posted at and an angle proportional to its number of descendants. Its position is then determined by the position of its parent reshare and its siblings.

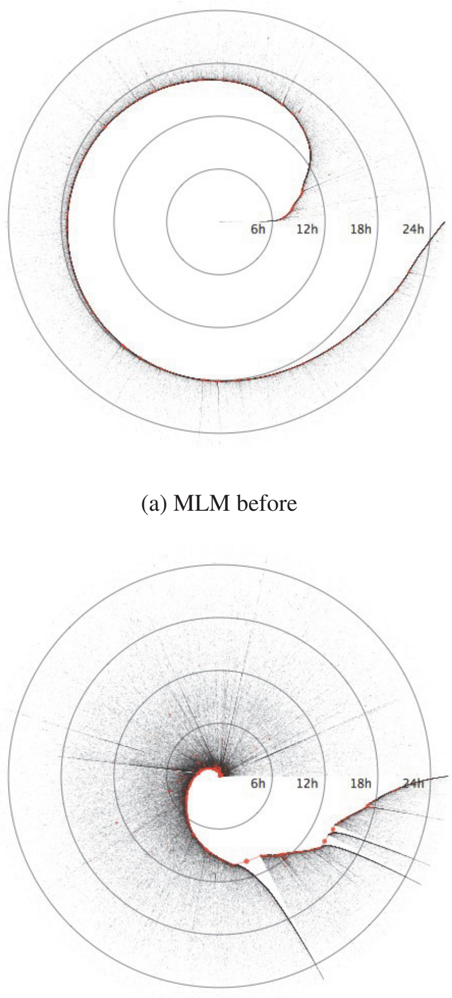
The difference of effect in rechaining is particularly striking when visualizing the cascades before and after the treatment. In both cases, before rechaining, pages lie on the inner spiral which represents the direct reshares of the post. After rechaining, pages are spread out in the MLM cascade, mirroring their increase in depth.

Temporal spread

While the two memes we are investigating both spread widely, the rate and manner in which they spread differs markedly. In both cases, the first 24 hours after the photos were posted contained the majority of reshares, 96% for MLM and 90% for OVP. As of February, OVP was still being reshared, enjoying a slight increase in popularity due to Valentine's Day. Shortly after MLM had reached its goal of 1 million likes at 18 hours after upload, the owner made the photo private to himself and his friends, disabling further resharing. Figure 7 illustrates the number of reshares per hour for the two photos in their first 24 hours. OVP enjoyed a burst of attention with a large number of shares in the first hour after it was posted, quickly and steadily falling off after that, except for two spikes where popular Facebook Pages shared the photo with their followers. In contrast, MLM started out slowly, gaining traction before peaking at 17 hours and quickly dropping thereafter.

Repeated exposure

The most common way that users see photos and other content on Facebook is through their News Feed. If a friend or page they follow were to share a photo, the News Feed might contain a story that "[Friend] shared [source]'s photo", along with any comments the friend made, the photo, and the original text accompanying the photo. The user can then share the photo with their own friends by clicking on the "Share" link. Figure 8 shows, for each meme, the percentage of users who shared the photo after seeing it some number of times in their News Feed. For both memes, we see an increase in



(a) MLM before

(b) MLM after

(c) OVP before

(d) OVP after

Figure 6: Cascades before and after rechaining. Black nodes represent users and red nodes represent pages.

share rate from one to two and three impressions, though the increase is greater for MLM. This increase may be due to both influence and homophily (Bakshy et al. 2012): each additional impression gives the user another chance to reshare the photo, but it also indicates that the user is surrounded by individuals susceptible to the meme, and so is more likely to susceptible as well.

Branching factors

Why did one meme fall in popularity from a high initial burst, while the other built up momentum gradually? We can get an idea of how the two memes spread by looking at the branching factors of the reshare trees. Since each node is a reshare, the branching factor of an individual node is the number of subsequent reshares that followed directly from that node. If a cascade were to have a uniform branching factor above 1, it would spread throughout the network as a viral pandemic would, but below 1 it would peter out. Rather than being constant, the branching factor varies both over time and by depth.

Figure 9 shows the average branching factor for reshares (omitting the root shares) of each photo by hour since the original post. For OVP, the mean number of reshares starts below 1 and drops steadily for later shares, with two notable exceptions: at 10 and 14 hours after the photo was posted, two celebrity pages with large numbers of followers (Michelle Obama and Alicia Keys, respectively) shared the photo, each individually generating tens of thousands additional reshares – these reshares are especially notable in the visualized cascades in Figure 6. For MLM, we again see a drop in the first few hours. However, subsequently, the mean branching factor increased steadily until 16 hours, when it surpassed 1, with the meme reaching popular users and pages, at which point it peaked and began dropping

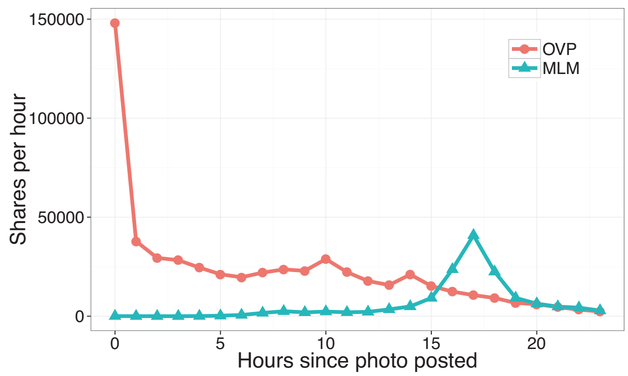


Figure 7: Shares per hour for each meme for the first 24 hours after the respective photo was posted.

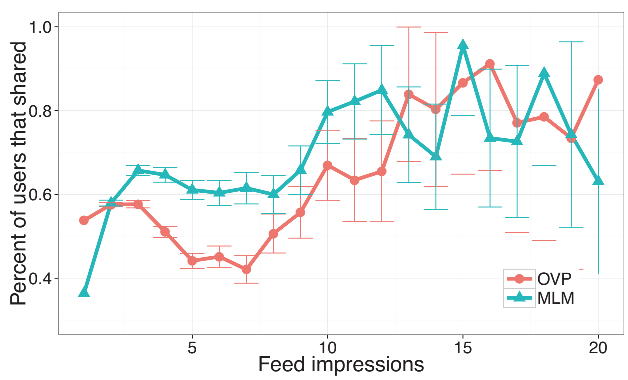


Figure 8: Of the users that saw either photo up to 20 times in their Facebook News Feed, the percentage that then shared it, with 95% confidence intervals.

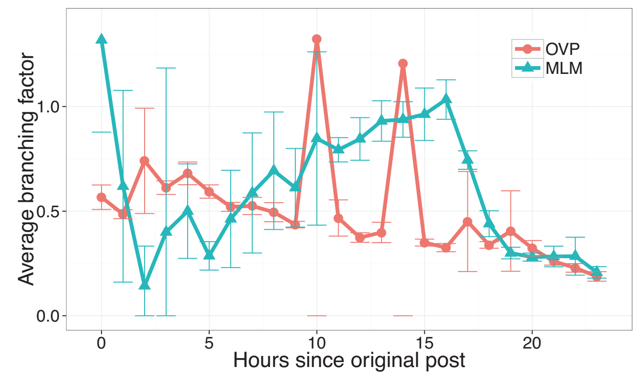


Figure 9: Mean branching factor, with 95% confidence intervals, by hour.

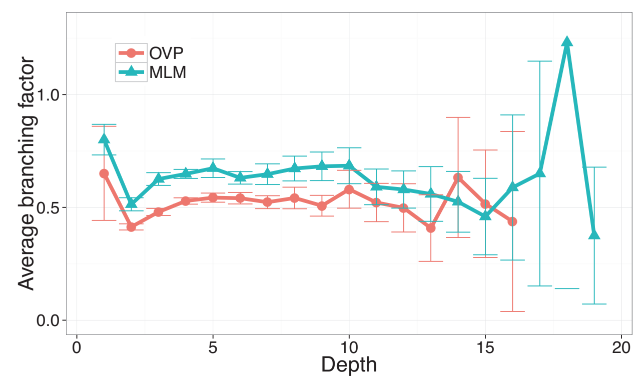


Figure 10: Mean branching factor, with 95% confidence intervals, by depth.

again. The drop is likely due to the photo reaching its goal of 1 million likes at 17.5 hours (see Figure 7), depressing the branching factor of any nodes posting shortly before the goal was reached, since many seeing those reshares would likely also see that there was no need to reshare further.

While the relationship between the reshare time and branching factor is clearly different for the two memes, Figure 10 shows that both memes have similar branching factor at a given tree depth. For both memes, there was a drop from nodes at depth 1 to nodes at depth 2. This is likely due to most popular Facebook Pages sharing directly from the source. From depths 2 to 5 the mean branching factor increases at which point it stays fairly flat until later depths where there are fewer nodes and, thus, more noise.

Implicit in our discussion so far is the idea that the branching factor depends on the number of individuals any given node could influence to reshare. These include friends and followers of users, and followers of pages. As discussed previously, these users have a high likelihood of seeing a story about the reshare in their own News Feed, although they and other users may become aware through other paths. Figure 11 shows the average branching factor per audience member for nodes in the OVP and MLM cascades, bucketed by ranges of audience sizes. Though branching factor increases with audience size, each added audience member contributes mostly the same or less to the branching factor. In the following section, we examine just how much explanatory power variables such as audience size and time elapsed since first post have on a node's branching factor.

Explaining spread

We modeled two measures of a node's influence: the number of other nodes who directly shared from it (its branching factor), and the size of the entire subcascade it generated. Explanatory variables we looked at included the depth of the reshare in the cascade, time elapsed since the original post, time of day, whether a node corresponds to a Facebook user or a Facebook Page, and its audience size. For a page, the audience size is the number of followers, and for users it is the number of friends and subscribers. We also looked at a number of demographic variables for users, including age, gender, and country.

Table 2 shows the results of linear regression models fit with least squares, separated by meme and outcome variable. Of all the independent variables we considered, audience size yielded the greatest explanatory power, with other variables explaining less than 1% of the variance. We restrict

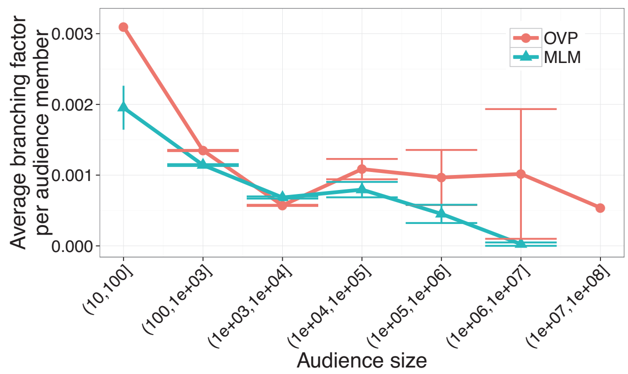


Figure 11: Mean branching factor, with 95% confidence intervals, by audience size.

the data used for the MLM regression to only include those nodes where the share occurred within 15 hours of the original post. This is to avoid including nodes which posted late enough to be affected by the drop in interest after the photo achieved 1 million likes (as we saw in Figures 7 and 9). Without this restriction, the regression is able to explain substantially less variance than reported in Table 2. This points to the need to know external and internal driving factors to fully understand the temporal dynamics of cascades.

In Table 2 we see that, for both outcomes, the audience size explains more variance for the OVP meme than it does for MLM. This is likely due to having more skew in the OVP data than in MLM. In particular, OVP includes a small number of data points with very large branching factors and subcascades. By fitting these points well, the OVP models look like a better fit. To remove this skew, we also fit regressions after applying a logarithmic transform to variables. For these regressions the variance explained was much less, though it was more comparable between the two memes. The modest correlation between a node's audience and its cascade influence is consistent with prior findings on Twitter (Cha et al. 2010; Romero et al. 2011).

Table 2: Regressing the branching factor and subcascade size on the audience size. For MLM only nodes that shared in the first 15 hours are included.

	branching factor		subcascade	
	OVP	MLM	OVP	MLM
intercept	0.08	0.40	0.41	2.34
audience * 10^{-4}	8.17	5.18	11.33	13.60
$\overline{R}^2$	0.49	0.30	0.45	0.10

As expected, for both memes, the audience size explains more of the variance in direct influence (branching factor) than it does when including indirect influence (subcascade size), though the difference is more dramatic for MLM than it is for OVP. This is likely due to the fact that the ratio of branching factor to subcascade size is smaller for MLM. For OVP there is little difference between branching factor and subcascade size, thus the regressions are not very different.

Demographics and susceptibility

In the previous section we saw that some of the variance in the spread of memes can be explained simply by the connectivity of the nodes sharing the content. The better connected the node, the more additional shares are likely to result. However, this leaves a large portion of the response to the information unexplained. Naturally, some of that variance may be due to how much the content matches the interests of the audience. For example, one might expect Alicia Keys' post of the OVP to garner intense interest because of the millions of subscribers to her Facebook Page. But the response might also depend on Keys' involvement with the Obama campaign; she recorded Obama's campaign theme song. As well, her own gender and ethnicity (she is of African American descent), and potentially that of her followers, are those of likely Obama supporters.

Therefore, we next investigate whether some portions of the underlying population on which each of the two memes were spreading were more susceptible than others. In principle, highly viral memes need not be uniformly appealing to the entire internet population, as long as the susceptible population is sufficiently interconnected.

In fact, we find that the two memes appealed to broad but distinct populations. There is further differentiation between users in terms of who chooses to click, like, comment and reshare. Each of these actions requires a different amount of effort. Clicking and liking are just a single action. Resharing can be just two button clicks, but the user also frequently types in a description or comment about what they are resharing. Comments require typing in text. Beyond the effort required in each of these interactions, users consider the signal they are sending by engaging in them. A "like" is typically considered an endorsement, and in the case of the MLM, a like has the additional feature of helping the poster move toward his goal. A comment can be supportive, but it can also engage with the poster in other ways, e.g. expressing disapproval. While clicks are not visible to their Facebook friends, a friend might notice a like and comment. Finally, reshares are the strongest endorsement, as they broadcast the information content to the user's selected audience.

We therefore examine not just demographic differences in resharing, but compare them against rates of exposure and other types of activities, using as a baseline the exposure and actions related to content overall.

Unsurprisingly, the lighthearted sexual subject matter of MLM appealed primarily to young males. As Figure 12(a) shows, nearly as many women as men were exposed to MLM in their News Feed, though they were less likely than men to like it and commented on it even less frequently. Their reshare rate was significantly lower, with only 19% of the reshares being made by women. That is, the greater the amount of effort, and the greater the endorsement of the meme, the less likely it was for the action to be taken by a woman. Nevertheless, because the Facebook network is relatively well mixed, that is, there is little gender homophily in the network, the exposure is relatively even. In terms of age, this meme appealed to younger than average users, as shown in Figure 13. It skews even younger for those users liking the photo. Resharers are typically older than the aver-

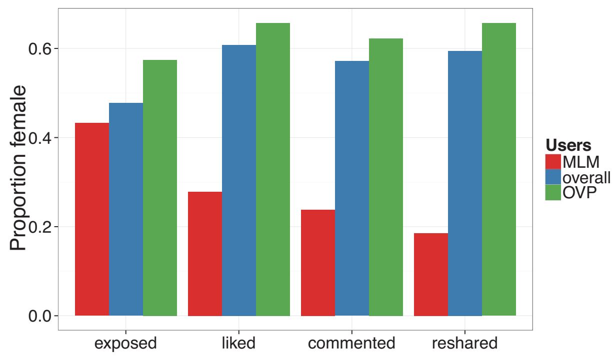


Figure 12: Gender differences in exposure and actions by users to two specific memes and to content overall.

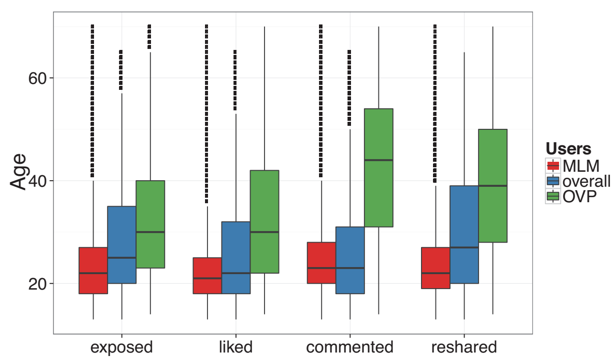


Figure 13: Age differences in exposure and actions by users to two specific memes and to content overall.

age user, but for this meme the resharers were younger than average, again showing a higher interest among the younger population in the meme.

In contrast, the OVP meme skewed toward older users. This was true of everything ranging from exposure to liking, commenting and sharing. Interestingly, the highest skew was seen in those who are commenting. We also observe that women are more actively engaging with OVP than with other content on average, which could be due to the presence of Michelle Obama in the photo or the support Obama enjoyed among women voters relative to Romney. A second photo of Obama by himself on election night had 61.5% female resharers relative to 65.7% for OVP.

Even more interesting than simple age and gender demographics were the political affiliations of the users propagating and reacting to OVP. About 37% of Facebook users post about politics (Rainie and Smith 2012), and the OVP post could be considered political. We looked at political preferences in two ways. The first was through users' explicit statement about the candidate that they support, namely by examining which if either of the official Obama and Romney Facebook Pages they "liked". Overall, by the time of the election, 33.9 million users had liked the Obama page, and 11.9 million users had liked the Romney page. Of the ones

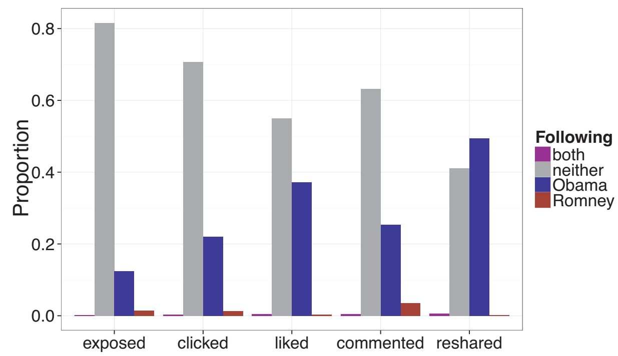


Figure 14: Users who were exposed to and engaged with OVP, segmented by whether they follow the Obama and Romney Facebook Pages.

who were active on Facebook during election week, 27% of Obama page likers saw OVP, contrasted to 8.75% of Romney page likers. As Figure 14 shows, the majority of those exposed had followed neither candidate on Facebook. However, roughly half of those who reshared OVP were followers of the Obama page. This indicates that there is strong selectivity in sharing the photo for those individuals who were already supporting Obama. We also note that although Romney supporters were far less likely to engage with OVP in a way that would endorse it, i.e. by liking or resharing it, the ratio of comments from conservative as opposed to liberal users is 1:4, not as low as it could be. Comments need not express support, e.g. a common theme in disapproving comments seen publicly on the Barack Obama page was "i don't think this country can really stand 4 more years" in answer to the photo's tagline "Four more years."

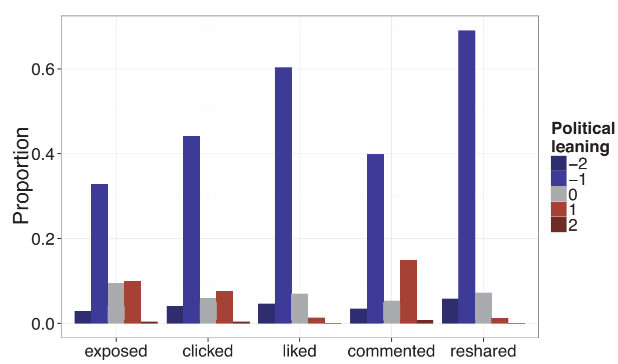


Figure 15: Political leanings of US users interacting with the OVP meme.

The analysis of the followers of the Obama page confounds two things: the likelihood of direct exposure because a user follows the Obama page and therefore was likely to have seen the photo directly in their News Feed from Obama, and the overall inclination that someone who supports Obama would reshare it. We therefore also looked at a second signal about a user's political affiliation and that is the free-form political affiliation users can fill out on their profiles. Limiting ourselves just to users in the United States, we manually mapped the top 100 occurring terms on a -2 to +2 scale, -2 being very liberal and +2 being very conservative. For example, 'Democrat' maps to -1, 'Tea Party Republican' maps to +2, and 'Moderate' maps to 0. Of the 2,122,145 users who saw OVP and had a political affiliation that we could map on a -2 to 2 scale, 882,040 liked the Obama page, and 242,807 liked the Romney page.

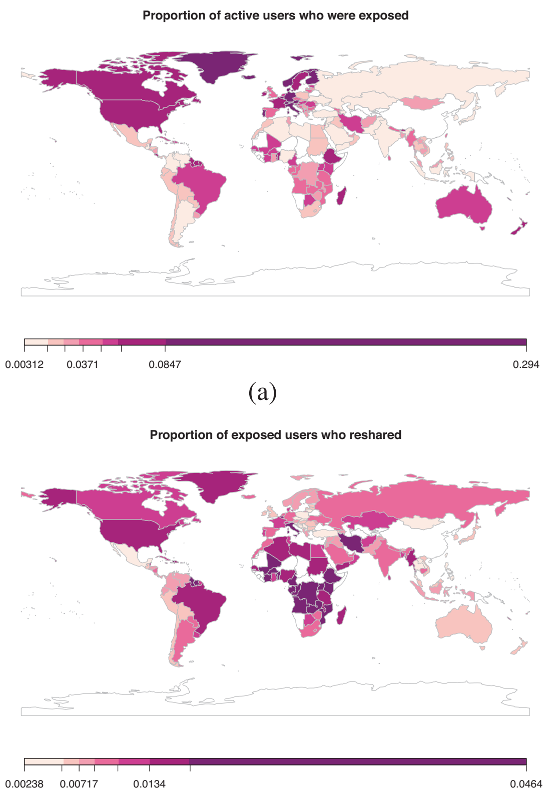


Figure 16: Geographical distribution of (a) exposures relative to active country users populations and (b) reshare rate per exposure for OVP

As Figure 15 shows, there is less skew in exposure and interaction based on political affiliation alone. This is because the number of liberal and conservative users on Facebook is roughly balanced (whereas the popularity of Obama and Romney is not). So even though about 15.2% of users who identify as liberal in their profile were exposed, relative to 5.2% percent of those who identify as conservative, the relative proportions of liberals and conservatives interacting with the meme, though still skewed, are more balanced. Interestingly, the very liberal (12.8%) and very conservative (4.7%) were slightly less likely to be exposed than their more moderate counterparts.

Finally, we examine the geographic distribution of the two memes. Even though both memes' primary activity spanned little over a single day, their reach throughout the world is expansive. While one might expect that a meme pertaining to the election outcome in the United States would primarily be of interest to US users, we find instead that a majority of the impressions (70.3%), likes (60.1%), comments (63.2%) and reshares (62.9%) occurred *outside* of the US. These statistics are an interesting proxy of the interest in US politics across the world, and are also suggestive of the rapid, global diffusion that Facebook enables. MLM, posted by a Norwegian user in English, quickly spread across continents.

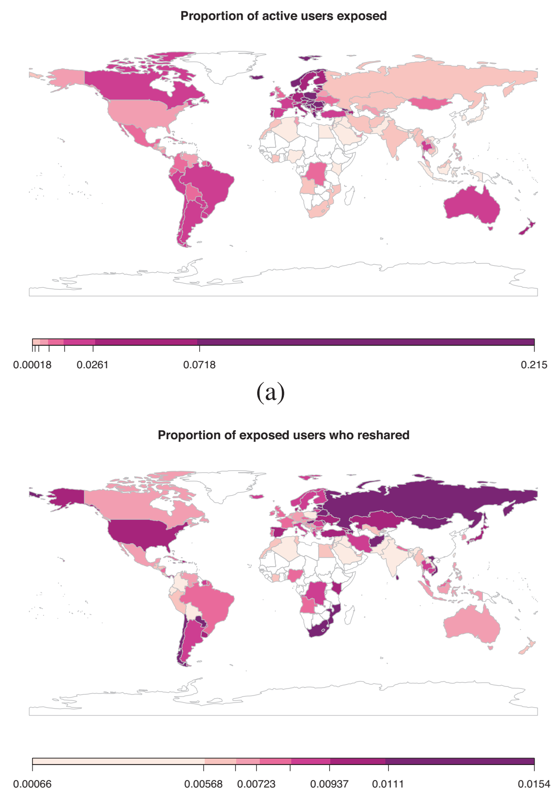


Figure 17: Geographical distribution of (a) exposures relative to active country users populations and (b) reshare rate per exposure for MLM

Conclusion

In this paper we took a first step toward understanding the distribution and diffusion of viral content on one large social networking platform. We showed that while most content is not viral, a small minority is. This is likely a natural result of the limited bandwidth of users. With a large number of photos uploaded daily, only a very small fraction of photos could be sustained as viral without users having to reshare an unreasonable number of photos every day. By examining two specific instances, we showed that photos reshared hundreds of thousands of times are seen by millions of users, affecting a significant fraction of the user population and confirming anecdotal accounts of seemingly ubiquitous memes. We saw that there are different ways in which memes can achieve such ubiquity. In the case of OVP, the source was itself a large hub, and half of the shares were made by followers of the source. In the case of MLM, the growth was largely organic, although the meme may not have spread nearly as far without the help of the high branching factors of pages which reshared it.

Although we studied photo reshares in aggregate and two very large cascades individually, a fair portion of activity is actually centered around moderately-sized memes, those being reshared dozens to hundreds of times. It would be interesting to see whether such memes are specific to particular interests and communities (Macskassy and Michelson 2011), and whether they may actually be relatively successful in the sense that they permeate their target audience. If some memes are specific to certain communities, it would be of interest to infer this from their initial spread. Prior work has shown that cascades may die out due to a combination of high clustering among the susceptible population and a lesser probability of resharing content upon repeat exposure (Greg Ver Steeg et al. 2011).

The particular cascade shape may depend on the topic the item relates to (Romero, Meeder, and Kleinberg 2011). Cascades may also be competing for attention and affecting one another's spread (Weng et al. 2012; Myers and Leskovec 2012). For example, the second OVP, showing just the president standing alone, received half the likes and only a sixth of the reshares. Potentially its spread was affected by the OVP that preceded it. We leave these and other questions for future work.

Acknowledgments

The authors would like to thank Cameron Marlow and Jon Kleinberg for insightful discussions.

References

Adamic, L.; Lento, T.; and Fiore, A. 2012. How you met me. In *Proc. of the 6th International AAAI Conference on Weblogs and Social Media (ICWSM)*.

Aral, S., and Walker, D. 2011. Creating social contagion through viral product design: A randomized trial of peer influence in networks. *Management Science* 57(9):1623–1639.

Aral, S., and Walker, D. 2012. Identifying influential and susceptible members of social networks. *Science* 337(6092):337–341.

Bakshy, E.; Hofman, J. M.; Mason, W. A.; and Watts, D. J. 2011. Everyone's an influencer: Quantifying influence on Twitter. In *Proc. of the 3rd ACM Conference on Web Search and Data Mining (WSDM)*.

Bakshy, E.; Rosenn, I.; Marlow, C.; and Adamic, L. 2012. The role of social networks in information diffusion. In *Proc. of the 21st International Conference on World Wide Web (WWW)*, 519–528. ACM.

Bhattacharya, D., and Ram, S. 2012. Sharing news articles using 140 characters: A diffusion analysis on Twitter. In *Proc. Advances in Social Networks Analysis and Mining (ASONAM)*, 966 –971.

Cha, M.; Haddadi, H.; Benevenuto, F.; and Gummadi, K. P. 2010. Measuring user influence in Twitter: The million follower fallacy. In *Proc. of the 4th International AAAI Conference on Weblogs and Social Media (ICWSM)*, volume 14, 8.

Goel, S.; Watts, D.; and Goldstein, D. 2012. The structure of online diffusion networks. In *Proc. of the 13th ACM Conference on Electronic Commerce (EC)*, 623–638. ACM.

Golub, B., and Jackson, M. 2010. Using selection bias to explain the observed structure of Internet diffusions. *Proceedings of the National Academy of Sciences* 107(24):10833–10836.

Gomez-Rodriguez, M.; Leskovec, J.; and Krause, A. 2012. Inferring networks of diffusion and influence. *ACM Transactions on Knowledge Discovery from Data (TKDD)* 5(4):21.

Greg Ver Steeg; Ghosh; Rumi; Lerman; and Kristina. 2011. What stops social epidemics? In *Proc. of the 5th International AAAI Conference on Weblogs and Social Media (ICWSM)*.

Leskovec, J.; Adamic, L.; and Huberman, B. 2007. The dynamics of viral marketing. *ACM Transactions on the Web (TWEB)* 1(1):5.

Leskovec, J.; Singh, A.; and Kleinberg, J. 2006. Patterns of influence in a recommendation network. *Advances in Knowledge Discovery and Data Mining* 380–389.

Liben-Nowell, D., and Kleinberg, J. 2008. Tracing information flow on a global scale using internet chain-letter data. *Proceedings of the National Academy of Sciences* 105(12):4633–4638.

Macskassy, S. A., and Michelson, M. 2011. Why do people retweet? Anti-homophily wins the day! In *Proc. of the 5th International AAAI Conference on Weblogs and Social Media (ICWSM)*, 209–216.

Milano, D., and Stern, J. 2013. 'One million likes' on Facebook for a puppy, clothing, sex. http://abcnews.go.com/Technology/facebook-million-likes-puppy-bunny-sex-clothing-meme/story?id=18295476. ABC News.

Myers, S. A., and Leskovec, J. 2012. Clash of the contagions: Cooperation and competition in information diffusion. In *IEEE International Conference on Data Mining (ICDM)*, 539–548. IEEE.

Myers, S. A.; Zhu, C.; and Leskovec, J. 2012. Information diffusion and external influence in networks. In *Proc. of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 33–41.

Rainie, L., and Smith, A. 2012. Social networking sites and politics. *Washington, DC: Pew Internet & American Life Project. Retrieved June* 12:2012.

Romero, D.; Galuba, W.; Asur, S.; and Huberman, B. 2011. Influence and passivity in social media. *Machine Learning and Knowledge Discovery in Databases* 18–33.

Romero, D. M.; Meeder, B.; and Kleinberg, J. 2011. Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on Twitter. In *Proc. of the 20th International Conference on World Wide Web (WWW)*, 695–704. ACM.

Weng, L.; Flammini, A.; Vespignani, A.; and Menczer, F. 2012. Competition among memes in a world with limited attention. *Nature Scientific Reports* 2(355).

Wu, F., and Huberman, B. A. 2007. Novelty and collective attention. *Proceedings of the National Academy of Sciences* 104(45):17599–17601.