

# On the Information Diffusion Between Web-Based Social Networks

Giannis Haralabopoulos<sup>(✉)</sup> and Ioannis Anagnostopoulos

School of Sciences, Department of Computer Science and Biomedical Informatics, University of Thessaly, Lamia, Greece  
{ghara, janag}@dib.uth.gr

**Abstract.** The topic of Information diffusion has been in the centre of sociology interest for many years. Even before the rise of Online Social Networks (OSNs), social ties and the way the information flows have been studied in traditional real-life social networks. In this paper, we propose the concept of multilayer information flow, by considering every On-line Social Network (OSN) as a separate layer and the information as links connecting these layers. Under such formalization, information is spread from a source layer, while it is diffused in multiple other layers. In order to validate our proposal, experimentations are focused on Reddit and its content, by observing the diffusion of its posts and related content in ImgUr and YouTube.

**Keywords:** Information · Diffusion · On-line social networks

## 1 Introduction – Background

The notion that information flows, from mass media to opinion leaders and later on to a wider population as final consumers, was firstly introduced during the middle of 40's [1]. Almost a decade later, the authors in [2] revisited the subject by proposing their theory of “Two step flow of communication”, while, the authors of the work described in [3] noted that analysis of social networks is suggested as a tool for linking micro and macro levels of sociology theory. Similar conclusions were verified in OSNs after nearly 40 years, while at the same time, the interest for social media analysis has skyrocketed [4, 5].

Viral marketing was introduced in [6], acknowledging the importance of peer communication in social networks. The authors in [7], found that provocative content such as sexuality, humour, violence and nudity are crucial virality factors, in comparison with traditional TV advertising, where emotive content had always been the key. In addition to virality, information diffusion became an important research subject. As a result, the authors in [8], modelled outbreak detection via node selection, while performing a two-fold evaluation of their model by using a water distribution network as well as a blog network. In a work that dealt with word of mouth scenarios, it was noted that 59 % of individuals frequently share online content [9]. Additionally, the authors in [10] observed that positive content is more viral than negative.

Social influence modelling was the scope of [11]. Using old diffusion data to create such models, the authors observed that viral marketing could leverage genuine influence. Similarly, the authors in [12], investigated the social influence in meme social graph.

The authors in [12] noted that content lifespan is very important to virality. For the same OSN, Sakaki et al. [13] study users as social sensors by monitoring information flow and dissemination during earthquake incidents. In [14], authors analyse the entire Twitter graph in order to assess its topological characteristics.

Rajyalakshmi et al. in [15], proposed a stochastic model for the diffusion of several topics. The authors discovered that strong ties play a significant role in virality, having homophily as a major contributor, as observed in [17]. Micro and macro scales of the network become relevant - similarly to [3] - and the authors noted that, acts within groups have a global impact.

Authors of [16], studied the dynamics of the spreading process. Furthermore in [4], Bakshy et al., address the problem of information diffusion in OSNs. The authors ended up with the statement -as seen in [3, 14, 14] that, strong ties are more influential but weak ties are responsible for diffusing novel information.

Guille et al. in [17], addressed the issue of information adoption in Twitter, based on the assumption described in [16] in respect to micro and macro level dynamics. Their results were similar with the ones described in [2].

In this work, we consider every OSN as a layer and the information as links connecting each layer, we propose the concept of multilayer information flow. In this concept, information is spread from a source layer and diffuses in multiple others. To test our proposal, we decided to focus on Reddit and its content.

Reddit is a social news and entertainment site powered by user generated content. Registered users submit content on the form of a descriptive link; namely a post. Posts that acquire a high (vote ratio in a short period after their submission, are moved to the front page. It is apparent that Reddit community, defines the content and determine its "success" or "failure". Across various content posted on Reddit, we focus on posts that link to an external domain and their traffic is measurable.

Content of posts on Reddit can be (amongst others) an image or a set of images hosted in ImgUr, or a video in YouTube. In ImgUr, content is usually created at the same time as the corresponding post in Reddit. While in YouTube, most posts in Reddit are linked to old videos. A short time after their creation and rise in terms of popularity within Reddit, users from different OSNs start mentioning that content, either by citing Reddit or the domain where the content is hosted (ImgUr or YouTube in our case).

For the purposes of our research, we wanted to use famous and heavily visited OSNs. In one hand, Twitter provided us with the ability to fully observe the impact of a front-paged Reddit post. On the other hand, since most content in Facebook and Google Plus is private, we only discovered a fraction of the total references, derived from search through public posts.

The rest of the paper is organised as follows. The next Section provides a detailed description in respect to our methodology. In Sect. 3, we thoroughly present the results derived from our analysis, while in the last two sections (4 and 5), we discuss evaluation issues and conclude our research.

## 2 Methodology – Dataset Description

Our research is focused in Reddit, a social news and entertainment site. Content is generated and ranked by users, based on positive/negative (up/down respectively) votes. Newly created posts with a high enough<sup>1</sup> rank, reach the front page. The submitted content often links to an external domain and varies from simple news posts, political articles, Ask Me Anything sessions, up to entertaining pictures. The view count on their original source is not always available (e.g. Wikipedia, News Sites), while it is also possible for a post to not be linked with any other external domain. Additionally, comments on posts also link to several other domains, yet this link relation is out of context of our research.

Since our initial idea was to be able to count viewership in various parallel social networks, the types of posts, hosted in domains without a view counter, were excluded from our analysis. Subsequently, we were able to monitor pictures in ImgUr and videos in YouTube, since in the front page of Reddit, there are always some posts that link to either one.

Generally, a post in Reddit consists of its title, content and comment section. Title part is self-explanatory; comment section is hosted within Reddit domain, while content is usually hosted in an external domain and rarely in Reddit. We focus on the hosting domain of a post, and more specifically on the content hosted on YouTube and ImgUr domains. We should note that our initial plan included the Quickmeme domain, which was a well-known meme creation site. Unfortunately, the use of Quickmeme in Reddit was banned in June of 2013.

In our work, we need domains that provide viewership counters in order to monitor and examine information diffusion features. In one hand, ImgUr and YouTube count views according to the user's IP address, while on the other hand, Reddit only provides the voting count of a post. As such, the sum of negative and positive votes is used as the Reddit views counter. We were aware that both the absence of an IP based view counter and the method Reddit uses to fuse voting (see Footnote 1) introduces some inaccuracy in our results. However, since our main aim is towards diffusion rather than viewership volume, such kind of inaccuracy does not affect our initial perception over multi-layer information flow.

In some preliminary tests we performed on Facebook and Google Plus, it was observed that both demonstrated a fairly low diffusion, even on viral content. This happens because both APIs, do not provide access to private posts. Thus, only public posts were taken into consideration. Additionally, the network usage of Google Plus is fairly low compared to Facebook. Considering these factors, we decided to only include Facebook in our analysis. Furthermore in order to provide a more detailed picture of how information is spread amongst Facebook users, we not only take into account the post count, but also the "Like" count of each post. This method yields greater numerical results, but still relatively small compared to Twitter mentions. In Twitter we counted the mentions of either the Reddit post, or of the content in its respective domain. Finally, in order to present our results in a consistent format between different OSNs, we address

---

<sup>1</sup> <http://amix.dk/blog/post/19588>.

all kinds of viewership count as “Units of Interest” (UoI). As such, a single Unit of Interest is equal to one content view in its domain (ImgUr or YouTube), one vote in Reddit, one mention in Twitter or one mention/like in Facebook.

During August and September of 2013, we scrapped two categories of Reddit (subreddits), named “new” and “rising”, in an almost hourly-basis. The selection of these two categories was two-fold. Firstly, because we focus on diffusion of newly created content, and secondly, because in both Twitter and Facebook APIs only recent posts can be accessed in their entirety. In both selected subreddits, content is refreshed every 2 min. During the 60 days of our scraping, almost 1 million posts were obtained and further analysed.

These posts are separated into several topics such as “pictures”, “gaming”, “funny”, “news”, “videos”, “music”, etc. As mentioned, not all of them were used in our research. Out of more than 950.000 posts in total, nearly 102.400 met our domain scraping criteria, belonging to either ImgUr or YouTube. Each post was analysed every one hour (in most cases in a less frequent rate), in order to measure the accumulated “Units of Interest” in their hosting domains, Twitter, Facebook and Reddit.

Additionally, in order to discard content that is not gaining attention rapidly, we employed a simple rule (mentioned hereafter as check criterion). In this rule, we examined if “Units of Interest” in Reddit are doubled (in absolute values) on every check, for the first 4 checks after the post creation. We found out that only a small percentage of total posts reached such high viewership counts within such a short time span. Specifically, only 0.66 % of the tested posts passed this UoI check criterion among our dataset (682 out of nearly 102.400 posts).

### 3 Results

In this section, we present the results derived from our scraping procedure and the respective analysis. Data is separated based on their topic, subreddit category and the domain they are hosted in. So “new” and “rising” in every chart denotes the corresponding (subreddit) category, while ImgUr and YouTube defines the hosting domain. Reddit topics of posts are “AdviceAnimals”, “Aww”, “Eathpon”, “Funny”, “Gaming”, “Gifs”, “Movies”, “Music”, “Pics”, “TIL”, “Videos” and “WTF”.

#### 3.1 Posts and Counters

An important remark in our analysis was that “rising” category contained less posts linking to ImgUr and YouTube, compared to “new” category. This is perhaps due to the fact that more news and current events usually gain attention in a short timeframe, thus featured in “rising” subreddit. In total, we found nearly 45.000 ImgUr and YouTube posts in “rising” category and nearly 70.000 in “new” category. More specifically, the posts of the “rising” category revealed 40.966 links to ImgUr and 3.084 links to YouTube. Similarly, the distribution for posts of “new” category was 51.984 and 6.366 respectively. We also observed that “new” subreddit content is more prone to become viral, compared to content in “rising” subreddit. Content featured in “new” subreddit was twice as likely to pass our UoI check, while content in “rising” category

presented a higher initial UoI count, but fell short on our criterion (of at least double UoI). Out of 40.966 posts linking to ImgUr in “rising” subreddit, only 200 surpassed our check criterion (0.48 %). As for posts linking to YouTube, out of 3.084 posts in total, only 14 went through our check (0.45 %). On the other hand, in “new” subreddit, out of the 51.984 posts that linked to ImgUr, 431 doubled their UoI in 4 subsequent checks (0.82 %). Similarly, the number of posts that linked to YouTube and successfully passed our UoI criterion, was 37 out of 6.366 (0.58 %).

We must note here that our UoI criterion should not be considered as a virality validation, yet it provides strong evidence that enough viewers, within a small time-frame, have viewed the analyzed posts and the interest is not diminishing. Among 682 highly viewed posts, 631 were linked to ImgUr and 51 were linked to YouTube. The topics spread, according to the subreddit categories and the employed social platforms, are illustrated in Fig. 1 and Table 1.

**Table 1.** Number of posts per topic (after the check criterion)

| Topic         | Rising category |           | New category |           |
|---------------|-----------------|-----------|--------------|-----------|
|               | ImgUr           | YouTube   | ImgUr        | YouTube   |
| AdviceAnimals | 27              | 0         | 73           | 0         |
| Aww           | 18              | 0         | 38           | 0         |
| Eathpon       | 2               | 0         | 5            | 0         |
| Funny         | 84              | 0         | 158          | 0         |
| Gaming        | 25              | 0         | 69           | 0         |
| Gifs          | 4               | 0         | 8            | 0         |
| Movies        | 3               | 2         | 5            | 2         |
| Music         | 0               | 1         | 0            | 3         |
| Pics          | 29              | 0         | 56           | 0         |
| TIL           | 0               | 1         | 0            | 1         |
| Videos        | 0               | 10        | 0            | 29        |
| WTF           | 8               | 0         | 19           | 2         |
| <b>Total</b>  | <b>200</b>      | <b>14</b> | <b>431</b>   | <b>37</b> |

Prior to discussing Fig. 1, let us examine the topic titles. As mentioned in [9, 10], positive and entertaining content was found as the most frequently shared topic. In our case, we could label “AdviceAnimals”, “Funny” and “WTF topics”, as entertaining content. “Aww” and “Eathpon” can be considered as containing emotive content, as described in [7]. “TIL” is mainly informative, while “Pics” and “Videos” characterise various content. Finally, “Gaming”, “Movies” and “Music” contain user-centric and specific entertainment content.

As mentioned, entertaining and positive content is shared more frequently than anything else. Although provocative and controversial content can be created within Reddit and ImgUr, it rarely gets enough votes to appear in the front page. Most of the time, they appear in the form of a “news” or “TIL” post. The number of “gaming” posts that passed our check criterion is descriptive of the interests of Reddit users, especially

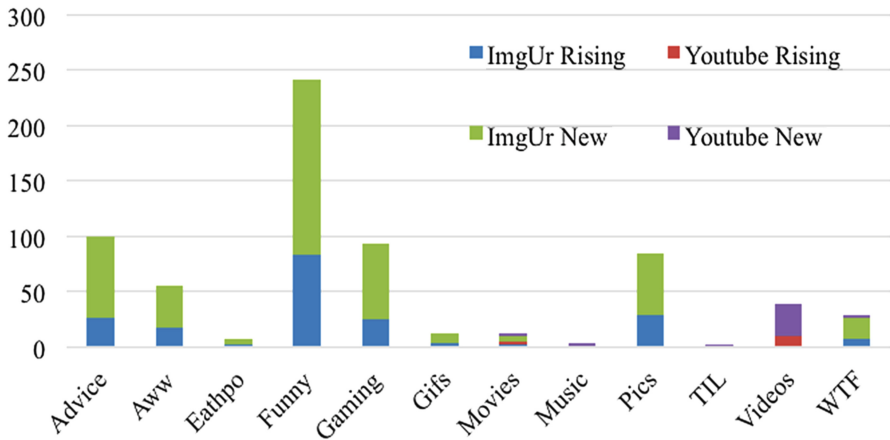


Fig. 1. Distribution of posts per topic (after the check criterion)

when compared to other forms of established entertainment, such as movies and music. Apart from the number of posts that passed our initial viewership test, we were strongly interested in their diffusion to Twitter and Facebook. Out of the 682 posts, 371 were mentioned in Twitter, 7 in Facebook and 172 in both OSNs. We observed that posts of “funny” topic are the most shared in both Twitter and Facebook. However, a closer look in the percentage of posts shared in these OSNs, reveals that “funny” is not included among the top-shared topics (see Fig. 2 and Table 2). Additionally, almost every category presents high Twitter and low Facebook sharing percentage (a direct result of the Facebook API limitations).

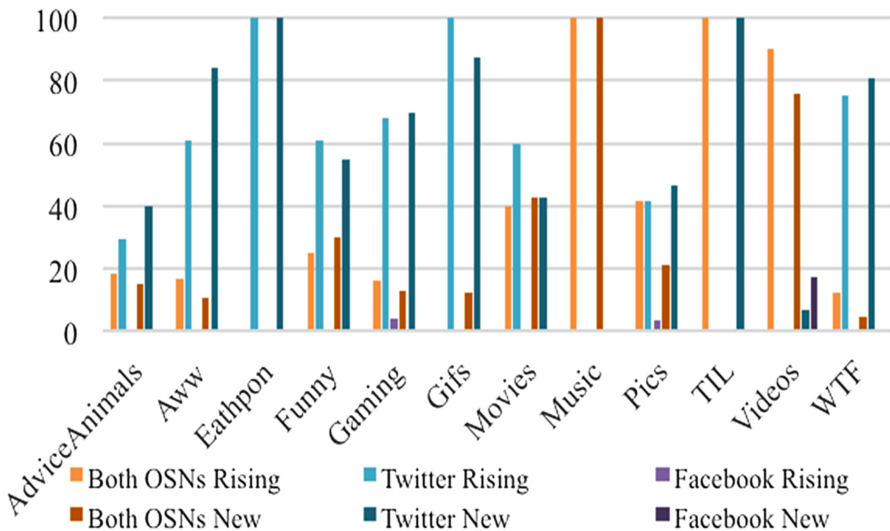


Fig. 2. Percentage of posts that diffused to OSNs

**Table 2.** Number of posts shared in tested OSNs

| Topic         | Rising category      |            |          | New category         |            |          |
|---------------|----------------------|------------|----------|----------------------|------------|----------|
|               | Twitter and Facebook | Twitter    | Facebook | Twitter and Facebook | Twitter    | Facebook |
| AdviceAnimals | 5                    | 8          | 0        | 11                   | 29         | 0        |
| Aww           | 3                    | 11         | 0        | 4                    | 32         | 0        |
| Eathpon       | 0                    | 2          | 0        | 0                    | 5          | 0        |
| Funny         | 21                   | 51         | 0        | 47                   | 87         | 0        |
| Gaming        | 4                    | 17         | 1        | 9                    | 48         | 0        |
| Gifs          | 0                    | 4          | 0        | 1                    | 7          | 0        |
| Movies        | 2                    | 3          | 0        | 3                    | 3          | 0        |
| Music         | 1                    | 0          | 0        | 3                    | 0          | 0        |
| Pics          | 12                   | 12         | 1        | 12                   | 26         | 0        |
| TIL           | 1                    | 0          | 0        | 0                    | 1          | 0        |
| Videos        | 9                    | 0          | 0        | 22                   | 2          | 5        |
| WTF           | 1                    | 6          | 0        | 1                    | 17         | 0        |
| <b>Total</b>  | <b>59</b>            | <b>114</b> | <b>2</b> | <b>113</b>           | <b>257</b> | <b>5</b> |

By further analysing the sharing percentage, we notice that almost all posts of “Music” and “TIL” categories are shared in both OSNs. However, the number of total viral posts in those categories is less than ten. The only categories that present a high number of viral posts diffusing to other OSNs in a percentage level close to 50 %, are “Videos” and “Movies”. This confirms the positive attitude of Internet users towards multimedia content. “Eathpon” posts, with photographs from around the world were 100 % shared in Twitter.

As far as the topics with the highest number of posts are concerned (“AdviceAnimals”, “Aww”, “Funny”, “Gaming” and “Pics”), we found out that the highest sharing percentage to both OSNs, pertained to “Pics” post from “rising” subreddit (41.38 %), followed by “Funny” posts from “new” subreddit (29.75 %). On the contrary, the lowest percentages appeared to “Aww” and “Gaming” posts from the “new” subreddit (10.53 % and 13.04 % respectively). Posts from “new” subreddit were more often shared in Twitter rather than posts from “rising” subreddit, for every topic except “Funny”. Finally, it is worth noting the extremely low number of posts shared only in Facebook compared to those shared in both OSNs.

### 3.2 Online Social Network Diffusion

As we already mentioned, UoI checks were simultaneously conducted on (at least) hourly-basis for Reddit, Twitter, Facebook and the domain the image or video was hosted (ImgUr or YouTube). Thus, we were able to observe the diffusion of a certain Reddit post and content in both OSNs. We expected to measure a slow-rate information diffusion from its source to both OSNs. In fact, such diffusion pattern was indeed found and was the most common diffusion standard throughout our analyzed posts and topics.

For evaluation purposes we use the mean values for every topic and for each subreddit. When the mean value of OSN diffusion is calculated, we do not take into consideration the posts that presented zero UoI variance in both OSNs. In other words, in our analysis we use the mean values of the posts that presented at least some variance in their UoI.

The time interval between subsequent checks performed in the tested domains (Reddit, ImgUr or YouTube, Twitter and Facebook) is at least one hour and the first check of every post has zero UoI. Each post is monitored for 7 days, yet we noticed that after 14 checks, the interest (in terms of UoI variation) declines significantly. The mean number of checks was 65 per post, while the 14<sup>th</sup> check happens within a 24-h period from the creation of the post. After the 14<sup>th</sup> check, we measured that the percentage level of UoI variation is lower than 0.1 %.

Furthermore, as we mentioned before, all posts are filtered to those that manage to double their UoI on Reddit in an hourly basis, for a 4 h timeframe. Consequently, every post we analyse presents high Reddit UoI variance for the first 5 checks, but only few of them present higher levels of variance after that point. As far as our OSN data scraping is concerned, we only had access on information from the public API of Facebook, while on Twitter we can only check messages within a timespan of a week, due to Twitter API limitations.

The most common pattern observed, corresponds to “AdviceAnimals” posts from the ImgUr domain in the “rising” subreddit. Domain UoI are growing at a high rate, concurrently with Reddit. Twitter is following with relatively small level of interest, while Facebook shows a delayed and low interest. Similar patterns appeared in several other categories; while in most of the cases Facebook presented nearly zero level of interest. This specific pattern was the most common in our research and appeared in 14 out of 18 ImgUr topics. Topics such as “AdviceAnimals”, “Aww”, “Funny”, “Eathpon”, “Gif”, “Pics” and “WTF”, followed this pattern - in both “rising” and “new” subreddits. However in both “Gif” categories, interest variation in Reddit is slightly higher than Domain.<sup>2</sup>

This is a good evidence that the most common diffusion flow of Reddit content, to the tested OSNs. Content hosted in the parent Domain is visible from many external sources and rapidly accumulates interest, whereas Reddit posts grow in the same time with a lower UoI count variance. In our tested OSNs, firstly Twitter increases its - initially high- UoI, presenting low variance, in contrast with Facebook where its UoI grow later with lower initial numbers and with very low variance.

Another interesting pattern was observed in “Movies” topic, in both subreddits. On every monitored post, we noticed interest spikes and, then sudden decreases in the UoI of all tested social media simultaneously, within few hours. This is mainly because the content shared is usually time sensitive (e.g. new trailers or news for upcoming movies). In other words, such posts seem not to be so persistent, since they do not interest users for long periods, but they rapidly diffuse to the tested OSNs. The concept of persistence in our analysis has to do with the ability of a post to retain interest for long periods of time, thus presenting positive UoI variation throughout that period.

---

<sup>2</sup> Every observed pattern can be found, in the appendix of a draft version of this paper, at <http://arxiv.org/abs/1403.1486>.



We should add that “Movies” and “Gaming” posts, as dealing with two of the most popular forms of entertainment, present the highest diffusion levels through OSNs. Posts about movies are generally diffusing faster, but gaming posts appeared to be more persistent.

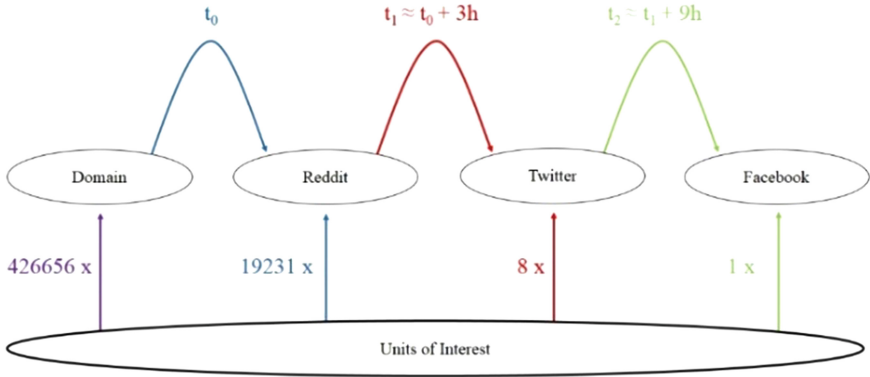
Interestingly enough, in YouTube content, we encountered completely different diffusion patterns. In most cases, domain UoI was not seriously affected by the increased interest in OSNs and Reddit. This is mainly because shared content in YouTube is not usually a newly created video, but an instance of a video that existed for a long time and becomes mildly popular in time. In our evaluation, the videos under study presented a very high initial UoI count, which resulted in a low UoI variance. Apart from that, we observed high UoI variance in Reddit and Twitter, low UoI variance in Facebook, and practically zero UoI variance in the tested domain.

## 4 Discussion

Our analysis verified many observations of previous researches, mainly in respect to the micro and macro effects of information flows [1, 2, 17]. More specifically, a single post (micro effect) in the parent domain connected with a post in Reddit, starts to accumulate views (macro effect) up to a point where the information hops to OSNs (first in Twitter and then in Facebook) flowing through individuals, while eventually the interest dies off. Positive content is confirmed to be the mostly shared content in OSNs, as mentioned in [7, 10, 10]. Similarly to [12], we found out that a Reddit post is mostly popular within the first hours after its creation. This effect is also enhanced by the classification method of Reddit, where new posts need fewer votes (compared to old posts) to move to the first page.

In addition, we have observed that the most shared posts are of positive and entertaining content, also seen in [9, 10]. These categories, would include post from topics such as “AdviceAnimals”, “Funny” and “WTF topics, while we also identified emotive posts, as described in [7], in topics such as “Aww” and “Eathpon”. Unfortunately, not enough viral posts of such content were found, in order to safely identify their diffusion model. “Gaming”, “Movies” and “Music” contain user-centric and very specific entertainment content, with low post appearance but great viral to total posts ratio, while, “TIL” is mainly informative and rarely used in posts linking to either ImgUr or Youtube.

Figure 3 illustrates the information flow scheme of the posts diffused to OSNs, according to our proposed method. Content creation in ImgUr is always tied to the respective post in Reddit occurring in  $t_0$ . Moreover, after approximately 3 h, links to ImgUr or Reddit are shared in Twitter. Finally, after nearly 12 h, the same links appear in public posts of Facebook. This results in the following user interest distribution: for each Facebook view or like, we have 8 Twitter posts, 19.231 Reddit votes and 426.656 views on ImgUr, which is the hosting domain. This reveals that ImgUr UoI are 22 times the UoI of Reddit. This is due to the fact that not every user votes on a post and we only count up and down votes of posts as user actions similar to views. Additionally, another factor is the large number of external sources that link directly to the hosting domain. Furthermore, total Reddit votes were found to be 2.403 more than the

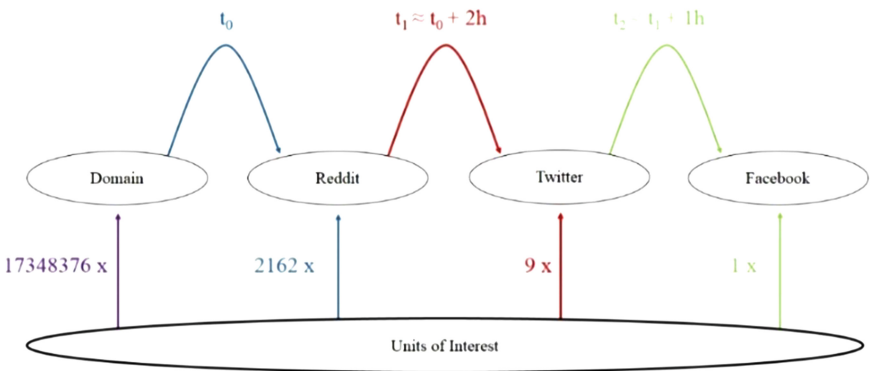


**Fig. 3.** Diffusion time and UoI allotment for ImgUr content of both subreddits

number of Twitter mentions, which is significantly lower than Domain to Reddit views ratio. Although these are mean values, views ratio (Domain to Reddit, Reddit to Twitter etc.) dispersion was fairly low.

Similarly, Fig. 4 illustrates the respective information flow for the YouTube case. However,  $t_0$  does not correspond to the time of creation in YouTube, but instead is the time at which the post in Reddit appears. In one hand, we found out that Reddit posts with content linking in YouTube are rarely created at the same time as the video itself, since the video achieves a high number of viewers, far after its initial “viral” period. On the other hand, the first post in Twitter and Facebook appears much faster (approximately 2 and 3 h after the Reddit post respectively). For each Facebook view or like, we measured 9 Twitter mentions, 2.162 Reddit votes and almost 1.7 million YouTube views. This difference is because most videos, when used in Reddit posts, are already widely known.

Of course, these figures highlight the analysis of data from our tested domains and OSNs, during our experimentations. However, if we used a slightly different method, results would be significantly different. For example, if we would calculate the

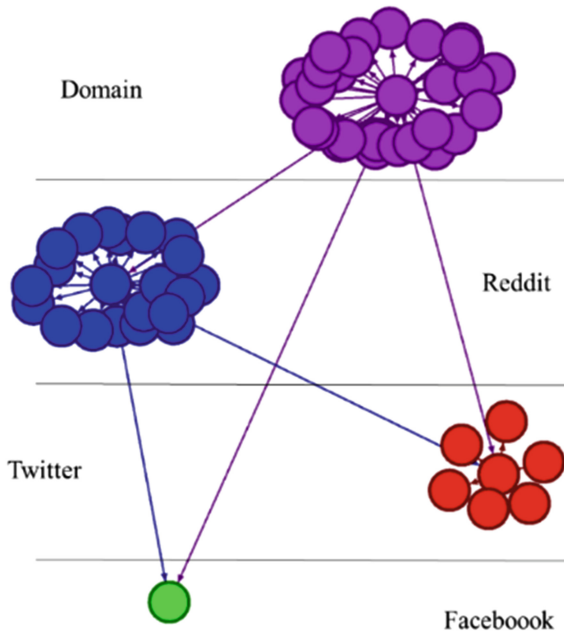


**Fig. 4.** Diffusion time and UoI allotment for YouTube content of both subreddits

exposure on both OSNs, UoI allotment would be smoother. According to “An Exhaustive Study of Twitter Users Across the World”,<sup>3</sup> the mean number of followers in an average Twitter user has been estimated equal to 208. That would mean that Reddit votes and Twitter mentions having a ratio of almost 1 in YouTube content and 0.1 in ImgUr content (down from 240 and 2.403, respectively). Furthermore, if we add the mean number of friends in an average Facebook account, which according to the “Anatomy of Facebook”<sup>4</sup> is equal to 190, we would get slightly higher Twitter to Facebook ratio, but significantly lower Reddit to Facebook and Domain to Facebook ratios. Finally, the fact that only public Facebook can be scraped, is a big factor to those particular UoI allotments.

## 5 Conclusions – Future Work

Information starts within a domain and hops to various other OSNs within minutes. However, upon its initial diffusion to other networks, information flow within the domain does not halt - it just slows down. Positively emotive and entertainment focused content is the most “viral”. Persistence was only found in gaming posts, while posts with movie content were the only ones that spread nearly simultaneously to every domain and OSN.



**Fig. 5.** Information flow through layers

<sup>3</sup> <http://www.beevolve.com/twitter-statistics/>, published results from October 2012.

<sup>4</sup> <https://www.facebook.com/notes/facebook-data-team/anatomy-of-facebook/10150388519243859>.

We should present a model of information flow, one that would accurately describe content sharing in modern OSNs. For example, let us consider a random Reddit post that follows a specific diffusion pattern, which has 30 domain UoI, 20 Reddit UoI, 7 Twitter UoI and 1 Facebook UoI. The UoI count used in this post, is obtained right after its first Facebook UoI appears, long before the 14<sup>th</sup> check. A single thread of information is spread into multiple concurrent users, spanning to various users of different OSNs simultaneously. Then, with a simple clustering the perception of layers (Fig. 5) is revealed. Information flows from the top layer towards the lower one, concurrently with the vertical flow where each layer expands.

Subsequently, we look forward into analysing a wider range of topics, in as many OSNs as we can. We believe that a predictive algorithm for viral posts will greatly benefit from our scraping findings. However, we would also benefit from such algorithms, because virality is hard to come by in modern OSNs.

In a different note of interest, we observed that various news posts in Reddit, have arisen to the first page before getting reported in dedicated news sites. As such, we could consider a different perspective, where OSNs could serve as a worldwide events station, similarly to what is proposed in [15]. Although we didn't discuss the great personalization options and properties of Reddit, we look forward into researching all these aspects of this modern OSN and its information flows in a future work of ours.

## References

1. Lazarsfeld, P.F., Berelson, B., Gaudet, H.: *The People's Choice: How the Voter Makes up his Mind in a Presidential Campaign*. Columbia University Press, New York (1944)
2. Katz, E., Lazarsfeld, P.F.: *Personal Influence. The Part Played by People in the Flow of Mass Communications*. Transaction Publishers, New Jersey (1970)
3. Granovetter, M.S.: The strength of weak ties. *Am. J. Sociol.* **78**(6), 1360–1380 (1973)
4. Bakshy, E., Itamar, R., Cameron, M., Adamic, L.: The role of social networks in information diffusion. In: *Proceedings of the 21st International Conference on World Wide Web*, pp. 519–528 (2012)
5. Rajyalakshmi, S., Amitabha, B., Soham, D., Tripathy, R.M.: Topic diffusion and emergence of virality in social networks (2012). [arXiv:1202.2215](https://arxiv.org/abs/1202.2215)
6. Jurvetson, S. Draper, T.: *Viral Marketing: Viral Marketing Phenomenon Explained*. DFJ Network News (1997)
7. Porter, L., Golan, G.J.: From subservient chickens to brawny men: a comparison of viral advertising to television advertising. *J. Interact. Advertising* **6**(2), 30–38 (2006)
8. Leskovec, J., Krause, A., Guestrin, C., Faloutsos, C., VanBriesen, J., Glance, N.: Cost-effective outbreak detection in networks. In: *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 420–429 (2007)
9. Allsop, D.T., Bassett, B.R., Hoskins, J.A.: Word-of-Mouth research: principles and applications. *J. Advertising Res.* **47**(4), 398–411 (2007)
10. Berger, J., Milkman, K.: Social transmission, emotion, and the virality of online content. Wharton research paper (2010)
11. Goyal, A., Bonchi, F., VS Lakshmanan, L.: Learning influence probabilities in social networks. In: *Proceedings of the 3rd ACM International Conference on Web Search and Data Mining*, pp. 241–250 (2010)

12. Ienco, D., Bonchi, F., Castillo, C.: The meme ranking problem: maximizing microblogging virality. In: *Data Mining Workshops*, pp. 328–335 (2010)
13. Sakaki, T., Okazaki, M., Matsuo, Y.: Earthquake shakes Twitter users: real-time event detection by social sensors. In: *Proceedings of the 19th International Conference on World Wide Web*, pp. 851–860 (2010)
14. Kwak, H., Lee, C., Park, H., Moon, S.: What is Twitter, a social network or a news media? In: *Proceedings of the 19th International Conference on World Wide Web*, pp. 591–600 (2010)
15. Rajyalakshmi, S., Bagchi, A., Das, S., Tripathy, R.M.: Topic diffusion and emergence of virality in social networks (2012). [arXiv:1202.2215](https://arxiv.org/abs/1202.2215)
16. Guille, A., Hacid, H.: A predictive model for the temporal dynamics of information diffusion in online social networks. In: *Proceedings of the 21st International Conference Companion on World Wide Web*, pp. 1145–1152 (2012)
17. Guille, A., Hacid, H., Favre, C.: Predicting the temporal dynamics of information diffusion in social networks (2013). [arXiv:1302.5235](https://arxiv.org/abs/1302.5235)



<http://www.springer.com/978-3-319-20369-0>

Web Information Systems Engineering - WISE 2014

Workshops

15th International Workshops IWCSN 2014, Org2 2014, PCS

2014, and QUAT 2014, Thessaloniki, Greece, October

12-14, 2014, Revised Selected Papers

Benatallah, B.; Bestavros, A.; Catania, B.; Haller, A.;

Manolopoulos, Y.; Vakali, A.; Zhang, Y. (Eds.)

2015, XX, 251 p. 92 illus., Softcover

ISBN: 978-3-319-20369-0