Cross-Pollination of Information in Online Social Media: A Case Study on Popular Social Networks

Paridhi Jain*, Tiago Rodrigues†, Gabriel Magno†, Ponnurangam Kumaraguru*, Virgílio Almeida†
* Indraprastha Institute of Information Technology, Delhi (IIIT-D), India
† Universidade Federal de Minas Gerais (UFMG), Brazil
{paridhij, pk}@iiitd.ac.in
{tiagorm, magno, virgilio}@dcc.ufmg.br

***Abstract*—Owing to the popularity of Online Social Media (OSM), Internet users share a lot of information on and across OSM services every day. Users recommend, comment, and forward information they receive from friends, contributing in spreading the information in and across OSM services. We term this information diffusion process from one OSM service to another as *Cross-Pollination*, and the network formed by users who participate in Cross-Pollination and content produced in the network as *Cross-Pollinated network*. Research has been done about information diffusion within one OSM service, but little is known about Cross-Pollination. We aim at filling this gap by studying how information from three popular OSM services (YouTube, Flickr and Foursquare) diffuses on Twitter, the most popular microblogging service. Our results show that Cross-Pollinated networks follow temporal and topological characteristics of Twitter. Furthermore, popularity of information on source OSM (YouTube, Flickr and Foursquare) does not imply its popularity on Twitter.**

I. Introduction

On Online Social Media (OSM) services, users create and share information with others, in a mechanism termed as *information diffusion*. With users having accounts in different OSM services (e.g. YouTube, Facebook), there is a tendency to exchange information across OSM services [1]. Users usually post URLs on Twitter and Facebook to announce to their friends about a new blog post or a new uploaded video (on YouTube). The information diffusion process across OSM services is analogous to a process in biology, termed as *Cross-Pollination*. In this process, pollen is delivered to a flower from a different plant, with the plants being different in their genesis [2]. Following the same analogy, we term the information diffusion process across OSM services as *Cross-Pollination*. A unit of information is analogous to pollen, and different OSM services are analogous to plants having different genesis.

Studying the dynamics and characteristics of a Cross-Pollination process is important for various reasons. Understanding Cross-Pollination can facilitate marketers to explore the rich environment for advertisement purposes. It can also help social media providers to improve their systems and develop tools to facilitate the information exchange across networks. Literature about information diffusion within one OSM service can be found, but little is known about the process of exchange of information across OSM services. Several important questions are unanswered – (1) What are the characteristics of the Cross-Pollination? (2) Does Cross-Pollination across OSM services help to increase the audience reached by the information diffused? (3) What is the relationship between the OSM services involved and how does it affect the information diffusion process?

In this paper, we study Cross-Pollination of three popular OSM services as *source OSM* – YouTube, the largest video sharing repository; Flickr, one of the largest photo sharing repository; and Foursquare, a popular location-based social networking service – with one another popular OSM service as *diffusion OSM*, Twitter, the largest microblogging service in the world.

We define a basic unit of information as a *meme*. [1] A video on YouTube and a tweet on Twitter are the examples of memes. Memes can be divided into two categories: *foreign* and *local*. We consider all posted URLs embedding meme belonging to another OSM service as a *foreign meme*. URLs embedding YouTube videos or Flickr photos, when shared on Twitter, are examples of foreign memes. We consider all other types of memes generated and diffused within one OSM service as a *local meme*. Hashtags (term starting with # to represent the topic of the tweet, e.g., #BestDad) and mentions (internal link to another user in the form of @username) are examples of local memes on Twitter. Figure 1 illustrates the dynamics of exchange of memes from one OSM service to another one. OSM service where a foreign meme originates is termed as *source OSM* (Flickr, YouTube, and Foursquare in our study), and the OSM service in which the foreign meme diffuses is termed as *diffusion OSM* (Twitter in our study). Twitter is a good medium to study the diffusion of foreign memes as it provides mechanisms that enable fast spreading of information. A network formed by users who participate in Cross-Pollination and content produced in the network is termed as *Cross-Pollinated network*.

To the best of our knowledge, this is the first work to explicitly study this important and unexplored area of Cross-Pollination. Our main results and contributions are:

[1] A meme is an element of a culture or system of behavior passed from one individual to another by imitation or other non-genetic means, taken from *http://oxforddictionaries.com/*

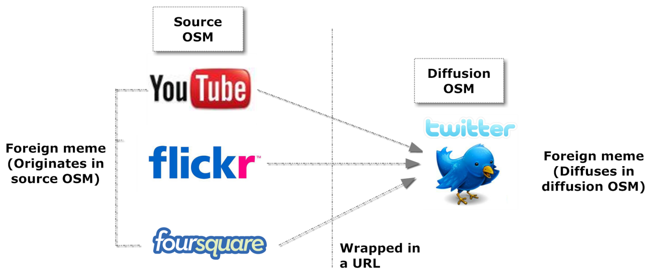


Fig. 1. **Cross-Pollination dynamics. Shows how a foreign meme is created and diffused in a Cross-Pollinated network.**

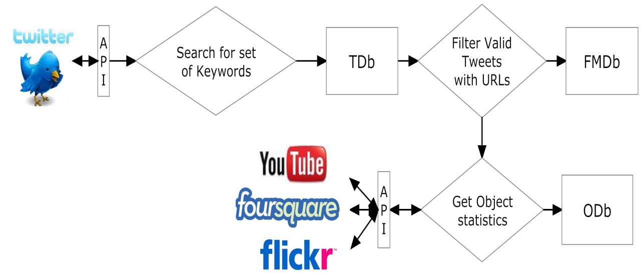


Fig. 2. **Data collection framework. Shows the two phases of the data collection framework which monitors the events to collect data.**

- Cross-Pollinated networks follow temporal and topological characteristics of the diffusion OSM.
- Cross-Pollination helps only a small percentage of foreign memes to reach to large audience.
- Popularity of meme on source OSM does not imply its popularity on diffusion OSM and vice versa.

In the next section, we explain our data collection and methodology; in Section III, we present the analysis and results of our study. We then present the related work in Section IV. Finally, in Section V, we conclude the paper with discussing the implications of our results, future work, and limitations of our research.

II. Methodology

In this section, we describe our data collection framework and provide descriptive characteristics of the datasets used.

A. Data Collection

Our data collection framework is composed of two phases (see Figure 2). In the first phase, we used Twitter Streaming Application Program Interface (API) [3] to collect all tweets periodically, using a set of keywords. This step was part of a research project, developed by a Brazilian Research Institute, [2] which tracks information about important events in several social and traditional media sources, like newspapers, blogs, and online social networks. [3] After this step, we filtered all URLs that appear on the content of the tweets. Due to the usage of URL shorteners like *http://bit.ly/*, [4] we expanded all shortened URLs and filtered all tweets with YouTube videos URLs, Flickr photos URLs, and Foursquare location URLs. We inserted all tweets that contain these types of URLs into *Foreign meme Database* (FMDb). In the second phase, we used YouTube [5], Flickr [6], and Foursquare [7] APIs to collect information about the foreign memes and their uploaders, storing the same in *Objects Database* (ODb).

Out of the most discussed topics on Twitter in 2010 [8], we created a dataset for FIFA World Cup (FWC), a global event. The FWC is an international football competition contested by the senior men's national teams of the members of Fédération Internationale de Football Association (FIFA), the sport's global governing body. The event happens every 4 years and in 2010 it took place in South Africa, from June 11^{th} to July 11^{th}. We monitored the FWC event from June 10^{th} to July 12^{th}, using 112 keywords (e.g. worldcup, FIFA and southafrica) in 7 different languages (like Portuguese, English and Spanish). To ensure no data loss, we used several redundant machines to collect the same data.

B. Datasets

Table I presents the descriptive statistics of our datasets. A total of 34,306 unique videos URLs were shared on Twitter during the FWC, in a total of 141,118 tweets, posted by 88,231 users. The videos were uploaded by 26,026 YouTube users. Table I also presents statistics about Foursquare and Flickr datasets, which are less popular than YouTube on Twitter (in our dataset), but still have a representative number of URLs to study. We also created a *baseline* dataset, which contains local memes only. In total, the baseline dataset has more than 29 million tweets, created by 3.5 million users. The baseline dataset is used in several of our analysis to contrast the characteristics of Cross-Pollinated networks with the characteristics of Twitter itself. This comparison helps in understanding how the introduction of foreign memes affects the diffusion OSM.

Source OSM (SM)	URLs	Tweets	Twitter Users	SM Users
YouTube	34,306	141,118	88,231	26,026
Foursquare	14,896	23,252	14,401	-
Flickr	1,719	2,560	1,419	711
Baseline	-	29,038,497	3,511,044	-

TABLE I
Descriptive statistics of the datasets. In our dataset, YouTube is more popular on Twitter than Foursquare and Flickr.

In order to verify the representativeness of our datasets, we repeated the analysis using keywords related to another popular event in 2010 on Twitter – the Brazilian Presidential Election. [4] We monitored this event, especially during the candidate's campaign, which started on July 6^{th} and ended on October 31^{st}, the final election day. To monitor this event we used a set of 30 keywords (e.g. dilma, serra and marinasilva)

[2] *Instituto Nacional de Ciência e Tecnologia para a Web*, *http://www.inweb.org.br/*

[3] The *Observatório da Web* Project, *http://observatorio.inweb.org.br/*

[4] Dilma Rouseff, elected president of Brazil, was the second most cited person on Twitter in 2010.

related to the candidates and their political parties. Due to space constraints, we present results only for FWC datasets, but most of our conclusions hold for the Brazilian Presidential Election datasets as well.

III. RESULTS

In this section, we investigate four key questions about Cross-Pollination – (1) What are the characteristics of Cross-Pollination? (2) Does Cross-Pollination across OSM services help to increase the audience reached by the information diffused? (3) If popularity of foreign meme on source/diffusion OSM is a factor affecting its popularity on diffusion/source OSM? (4) What is the role of users, who are present on both OSM, in Cross-Pollination process?

A. *Cross-Pollination characteristics*

Aiming at answering first and second question, we first analyze the temporal and then, topological characteristics of Cross-Pollinated networks.

Sharing Activity: An important temporal characteristic of Cross-Pollination is the volume of tweets generated by foreign memes on a given day during a certain period of time. Figure 3(a) shows the total number of tweets with foreign memes created on each day during the FWC event. For comparison purposes, the figure also shows the total number of tweets with local memes created per day (using baseline dataset). We observe a similar trend during the whole period, for all datasets analyzed. The trend of volume of tweets created due to meme (both foreign and local) sharing is relatively uniform and similar during the whole period, with small peaks occurring on the same days. Hence, foreign meme sharing activity follows local meme sharing activity, although absolute numbers differ significantly (around 10^3 YouTube foreign memes on Twitter and 10^6 local memes on baseline dataset).

User participation: In order to verify whether users contribute equally in the traffic generated by Cross-Pollination on Twitter, we define *User Participation* (UP) as the average number of tweets with memes created per day, for each user. We divided users into bins according to their UP, and then calculated the percentage of users in each bin (see Figure 3(b)). Users contribute equally for the traffic generated by Cross-Pollinated networks; vast majority of users (more than 90%) are in the same bin, with less than 2 tweets with foreign memes created per day. Furthermore, Cross-Pollinated networks follow the diffusion OSM in this aspect, as the vast majority (more than 70%) of users from the baseline dataset is on the same bin. The same observation can be done for Flickr photos and Foursquare locations.

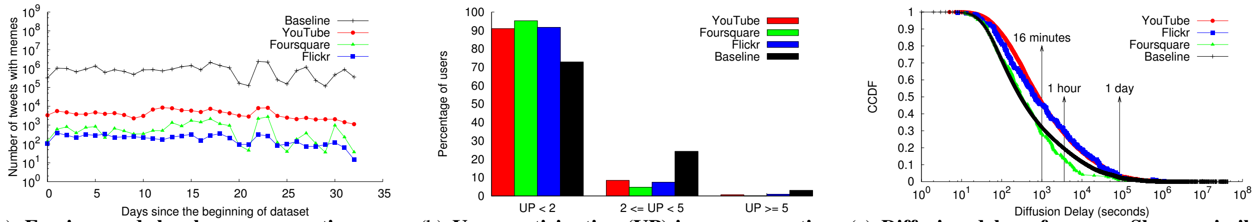
Diffusion delay: We define *diffusion delay* as the time between tweet meme getting created and being retweeted. Figure 3(c) shows the complementary cumulative distribution function (CCDF) for the diffusion delay of the three Cross-Pollinated networks studied and the baseline dataset. On average, 75% of the memes are retweeted in less than 1 hour, and 97% are retweeted within a day. We note from the distributions that YouTube and Flickr memes tend to be retweeted with a slightly higher delay than Foursquare and local memes. For example, around 50% of tweets with YouTube and Flickr memes have a delay larger than 1,000 seconds (around 16 minutes), while 30% of retweets from Foursquare and local memes have a delay larger than 1,000 seconds. Nature of the content is a reasonable explanation for this difference. A user can easily read and quickly respond a direct message (local meme), while a foreign meme becomes an indirect message as the user is expected to view the content of the URL before forwarding it. In this case, Foursquare memes are more similar to local memes because they are usually automatically posted messages which contain the name of the place from where the user "checked in" together with the URL of the location. "I am at DCC, UFMG http://4sq.com/XyZw" is an example of this kind of tweet.

We now analyze topological characteristics of Cross-Pollinated networks.

Diffusion cascades: We now turn our focus to analyze topological characteristics of Cross-Pollinated networks. Diffusion occurs through *originators* and *spreaders*. Originators are users who posted a foreign meme on Twitter, and spreaders are users who forwarded (i.e., either replied or retweeted) that foreign meme posted by an originator. A diffusion cascade is defined as a directed connected graph $G(V, E)$, where nodes represent originators/spreaders and edges represents that a foreign meme tweeted by an originator is forwarded by a spreader. Direction of the edge represents the information diffusion from originator to spreader. Number of nodes present in $G(V, E)$, is the diffusion cascade size. We observe that, only 18% YouTube videos, 10% Flickr photos, and 2% Foursquare locations are forwarded at least once.

Foreign memes diffuse in cascades like star, path, and other connected cascades (see Figure 4), for the three Cross-Pollinated networks. Star cascades are formed when many users forward a single user's foreign meme, resulting in one-to-many diffusion of information. When one user forward many user's foreign meme, information diffuses from many to one user, resulting in many-to-one diffusion cascade. A path cascade is formed when a user forwards an already forwarded foreign meme, resulting in information diffusion from one user to another in a chain. A mixed connected cascade is formed when users involved in diffusion of foreign memes are associated with mixed actions of forwarding. Similar observations were found for the baseline dataset, where only 12% of local memes are forwarded and diffused in path, star, and mixed cascades.

Figure 5 shows distribution of number of cascades with cascade size. Out of the cascades formed by foreign memes diffused, most cascades are composed of only one originator and one spreader (i.e., of cascade size 2), which we term as one level of diffusion. There are only few cascades which have large cascade size, reaching many users and users' followers. For foreign memes that were diffused, the level of diffusion remains to only one user. Similar distribution can be seen for local memes. Number of cascades follow 90-10 Pareto distribution with 90% cascades with size ≤ 3 and 10%



(a) **Foreign and local meme creation over time. Show a relatively uniform creation activity during the complete period.** (b) **User participation (UP) in meme creation. Most users have similar participation in meme creation for both local and foreign meme, as most users belong to the same bin.** (c) **Diffusion delay of memes. Show a similar distribution for both foreign and local memes.**

Fig. 3. **Temporal characteristics of Cross-Pollinated networks in Online Social Media.**

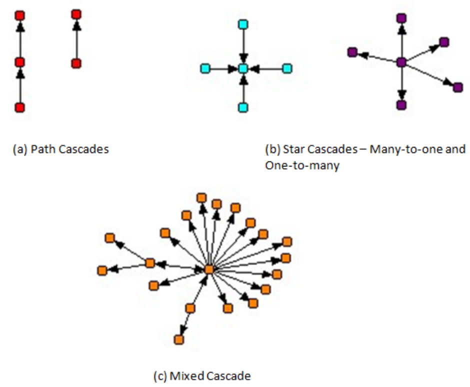


Fig. 4. **Diffusion cascades. Most of the foreign meme and local memes cascades are star-shaped followed by path and mixed directed connected cascades.**

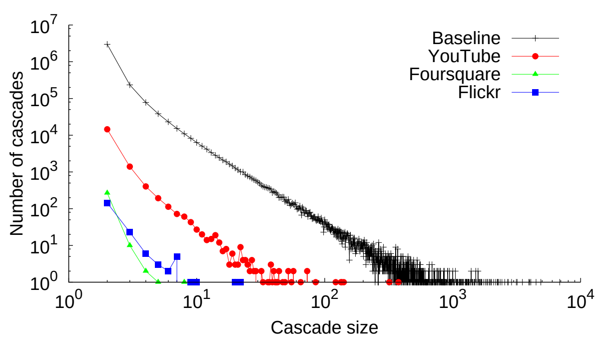


Fig. 5. **Distribution of cascades with cascade size. Shows a large number of foreign and local meme cascades with small size, while a small number of cascades with large size.**

cascades with size ≥ 3, for both Cross-Pollinated network and baseline.

Diffusion cascade statistics: Table II shows a comparison of some graph metrics for diffusion cascades of the three Cross-Pollinated networks in study and the baseline dataset. Twitter users are most attracted towards posting and forwarding YouTube videos than Flickr photos, Foursquare locations and local memes (highest spreaders / meme). Even then, average diffusion cascade size for YouTube remains approximately 2.53, close to the other datasets in study, owing to large number of size-2 cascades, which neutralize larger size cascades. [5] Average in-degree and average out-degree for Cross-Pollinated networks are higher than baseline and close to each other. Hence, Cross-Pollinated networks behave similarly, irrespective of the type of foreign meme diffused.

B. Popularity influence affecting foreign meme popularity on source and diffusion OSM

We turn our focus now to answer the third question. The popularity of a video on YouTube is measured by its view count. On Flickr, the popularity of a photo is also measured by its view count. The popularity of a location on Foursquare is measured by the number of "check-ins." On Twitter, the popularity of a foreign meme is given by the number of tweets with it. We obtained the popularity ranking of foreign memes in both the source OSM and the diffusion OSM. In order to compare two rankings, we used the *Kendall's Tau coefficient* [9], which is a measure of the rank correlation, denoted by τ. A τ of -0.0001 is observed on the Cross-Pollinated network between YouTube and Twitter, which demonstrates that the ranking of popularity of videos in both OSM services are independent. In other words, if a video is popular on YouTube does not mean that it will also be popular on Twitter, and vice-versa. Interestingly, we found same observation for Flickr and Foursquare datasets, as the τ is -0.0013 and -0.0001, respectively. Hence, popularity of foreign meme in source OSM does not influence its popularity in diffusion OSM.

We also checked whether the diffusion of foreign memes in the diffusion OSM helps in increasing the traffic (popularity of memes) in the source OSM. Although our datasets do not have information of how many clicks each URL received on Twitter, some URL shorteners provide APIs with statistics of access to their links. One of the most popular services is *http://bit.ly/*, which provides an API [10]. We used the API to analyze how many clicks each meme shortened in a *bit.ly* URL received from Twitter. In order to do this analysis, we collected data for all *bit.ly* URLs of our dataset, and then checked how many clicks they received from the referrer *twitter.com*. In total, we have 13,158 videos from YouTube dataset (38.4% of total) and

[5] Spreaders can belong to different cascades for different foreign meme.

Source OSM	Degree	In-degree	Out-degree	Path Length	# cascades / meme	Cascade Size	# spreaders / meme
YouTube	1.06	1.17	1.12	0.37	2.81	2.53	7.08
Flickr	1.11	1.06	1.48	0.43	1.11	2.69	2.97
Foursquare	1.03	1.09	1.06	0.48	1.02	2.13	2.18
Baseline	1.07	0.53	0.53	-	1.00	2.78	2.78

TABLE II

Diffusion cascade statistics for three Cross-Pollinated networks and baseline. All numbers are averages. Cascade characteristics are similar across Cross-Pollinated networks, implying its independence of the source OSM and are higher than baseline dataset.

1,719 photos from Flickr dataset (38.2% of total) shortened with *bit.ly*. We did not consider Foursquare dataset as we had only 37 *bit.ly* URLs, which is not representative. [6]

We then analyze the fraction of views [7] that each foreign meme (videos and photos) received from Twitter. We observe low fractions of views from Twitter, for many foreign memes tweeted. About 97% of the videos received no more than 1% of their views from Twitter, and almost 59% of the photos received no more than 1% of their views from Twitter. Twitter does not seem to be effective in increasing the popularity of foreign meme on the source OSM. By including only clicks from *bit.ly* URLs, we have a lower bound of the fraction of views that came from Twitter. There can be various other sources contributing to number of views for a foreign meme which we did not analyze here.

We observe, from the above analysis, that popularity of information on one OSM does not imply/affect its popularity on other OSM, and vice versa.

C. Users role in Cross-Pollination process

Another interesting aspect to analyze is the importance of the users present on both networks. First, to estimate such user presence, we checked, for all Twitter users who shared foreign memes, if their usernames exist on YouTube and Flickr. We also checked, for all YouTube and Flickr users who created videos and photos, respectively, if their usernames exist on Twitter. We refer to these users with presence in both the source OSM and the diffusion OSM as *carriers*, as they might have brought foreign memes from the source OSM to the diffusion OSM. Table III shows the number of Twitter users on each Cross-Pollinated network, the number of creators of foreign memes in the source OSM, and the number of carriers in each case (i.e., Twitter users and source OSM users). We did not include Foursquare in this analysis because a location is not associated to an owner.

Source OSM	Users	Carriers	Twitter users	Twitter Carriers
YouTube	28,721	1,207 (4.2%)	88,231	57,620 (65.3%)
Flickr	711	143 (20.1%)	4,028	403 (10.0%)

TABLE III

Statistics about carriers of information in Cross-Pollinated networks. Note that we have two types of carries – creators of content in source OSM; and users who shared a URL with foreign memes on Twitter.

[6] Foursquare has its own URL shortener *4sq.com*, which might be the reason for a small number of *bit.ly* URLs in our dataset.

[7] We consider that each click on the URL represents one view in the source OSM.

One limitation of this analysis is that a real person might use different usernames on different OSM services, but we believe that most part tend to use the same username as it will be easier to be found by friends. Another limitation is that a Twitter user might have access to the content of a source OSM without having an account on it.

Around 65% of the Twitter users who shared YouTube videos have an account on YouTube; while only 10% of the Twitter users who shared Flickr photos have an account on Flickr (see Table III). A feasible impact of this difference is the higher popularity of YouTube videos shared on Twitter, but we let for future work to confirm this observation. Interestingly, only 4.2% of YouTube users, creators of videos, have an account on Twitter, while 20.1% of owners of photos on Flickr have an account on Twitter as well. On manual inspection, we found that most YouTube videos are of general interest, like comedies and music clips, while most part of Flickr photos are of personal interest. So, theoretically, YouTube videos attract interest of a higher number of users, who watch and share them on Twitter most frequently than Flickr photos, which are mostly shared by their own creators to a limited number of their friends.

Another interesting question about user presence is whether creators of contents in a source OSM make use of Cross-Pollination as an attempt to increase the traffic to their contents. For example, users might upload a video to YouTube and share a tweet with its link to announce to friends about the new video. We note that only 0.7% of YouTube videos and only 7.4% of Flickr photos were first tweeted by creators of contents in the source OSM, which implies that the audience (users who receive and watch videos and photos) play an important role in carrying the information from the source OSM to the diffusion OSM.

Hence, we answer the fourth question that users that are present on both OSMs and are creators of the information, are not, actually the source of information diffusion in other media. This implies, users are not needed to be present on both OSM, for a Cross-Pollination process.

IV. RELATED WORK

In recent years, researchers contributed significantly in understanding several aspects of OSM services, including diffusion of information. Blogs [11], YouTube [12], Facebook [13], and Digg [14] are some of the OSM services which have been extensively studied. Blogging and micro-blogging networks are shown to have temporal and topological patterns which largely exhibit power law behavior [11], [15], [16]. Krishnamurthy *et al.* presented a detailed characterization of Twitter [17], and

Choudhury *et al.* analyzed how user similarities (homophily) along various attributes can affect the information diffusion process on Twitter [18]. Cha *et al.* analyzed the blogging network structure and information diffusion patterns within the network [19]. Liben-Nowell and Kleinberg [20] reconstructed the propagation of massively circulated Internet chain letters and showed that their diffusion proceeds in a narrow but very deep-like pattern. Broxton *et al.* analyzed the diffusion of viral video popularity in social media, but focused only on how the popularity of a video varies with its introduction in social media. They concluded that viral videos gain popularity faster on OSM than through any other referring source or itself (e.g., search engines, etc), and that viral video popularity on Twitter is at a higher rate than in any other OSM website, but without analyzing the underlying network structure affecting the higher rate [1]. Researchers have also studied information diffusion process on Facebook through News Feeds [21].

Most of the studies on information diffusion in OSM are limited only to diffusion of local memes. Some studies analyze the diffusion of foreign memes, but they did not aim at understanding how the OSM services involved are related. Our research intends to fill this gap, analyzing the diffusion of local and foreign memes on Twitter, as well as the relationship between Twitter and the source OSM services (YouTube, Flickr, and Foursquare).

V. Discussion

In this paper we studied some properties of Cross-Pollinated networks. We believe that understanding these properties can help OSM service providers to improve or introduce new effective ways of sharing information across OSM services; to comprehend user involvement in information diffusion across OSM services; and to help users to chose a diffusion OSM in which they should share information, in an attempt to make it spread fast and effectively. Understanding Cross-Pollination also enhances an understanding of evolving information diffusion process across OSM services, which can be used for business perspectives.

To the best of our knowledge, this work is the first step towards studying an emerging phenomenon in social media environments, with several future opportunities for researchers. We plan to provide a generalization of the Cross-Pollination and develop a formal model for the diffusion process in a Cross-Pollinated network. We also plan to extend some of our analysis. Firstly, we analyzed the Cross-Pollination considering the diffusion only in Twitter. Results involving different social media environments can present different characteristics. Secondly, we studied the relationship between the source OSM and the diffusion OSM only, but, for example, a video created on YouTube can be brought to Twitter by users who watched it on Facebook.

Although this research presented some interesting results about information diffusion across OSM services, it has some limitations. We only have data of tweets that contain some keyword used in the search. In order to reduce the impact of this limitation, a wide range of keywords related to the event were used, in several languages. Identifying if the tweet is related with the event is another limitation, which was addressed by the research project from which we received the data.

Acknowledgements

This work was supported by the Indo-Brazil Science Council. The authors would like to thank everyone who helped in collecting the data. The authors acknowledge all members of PreCog research group at IIIT-Delhi for their support.

References

[1] T. Broxton, Y. Interian, J. Vaver, and M. Wattenhofer, "Catching a Viral Video," in *Workshop on Social Interactions Analysis and Services Providers (SIASP)*, 2010.
[2] D. Charles, *The effects of cross and self fertilisation in the vegetable kingdom.* J. Murray, London, 1900. [Online]. Available: http://www.biodiversitylibrary.org/item/59301
[3] "Twitter Application Program Interface," http://dev.twitter.com/pages/streaming_api.
[4] "How to Shorten Links URLs," http://support.twitter.com/articles/78124-how-to-shorten-links-urls.
[5] "YouTube Application Program Interface," http://code.google.com/apis/youtube/overview.html.
[6] "Flickr Application Program Interface," http://www.flickr.com/help/api/.
[7] "Foursquare Application Program Interface," https://developer.foursquare.com/.
[8] "Top Twitter Trends in 2010," http://yearinreview.twitter.com/trends/.
[9] M. G. Kendall, "A New Measure of Rank Correlation," *Biometrika*, 1938.
[10] "*Bit.ly* Application Program Interface," http://code.google.com/p/bitly-api/wiki/ApiDocumentation.
[11] J. Leskovec, M. Mcglohon, C. Faloutsos, N. Glance, and M. Hurst, "Cascading behavior in large blog graphs," in *SIAM Conference on Data Mining (SDM)*, 2007.
[12] M. Cha, H. Kwak, P. Rodriguez, Y.-Y. Ahn, and S. Moon, "I Tube, You Tube, Everybody Tubes: Analyzing the World's Largest User Generated Content Video System," in *Internet Measurement Conference (IMC)*, 2007.
[13] A. Nazir, S. Raza, and C. Chuah, "Unveiling facebook: a measurement study of social network based applications," in *Internet Measurement Conference (IMC)*, 2008.
[14] G. Szabo and B. A. Huberman, "Predicting the popularity of online content," *Communications of the ACM*, 2010.
[15] W. Galuba, D. Chakraborty, K. Aberer, Z. Despotovic, and W. Kellerer, "Outtweeting the Twitterers - Predicting Information Cascades in Microblogs," in *Workshop on Online Social Networks (WOSN)*, 2010.
[16] H. Kwak, C. Lee, H. Park, and S. Moon, "What is Twitter, a Social Network or a News Media?" in *World Wide Web Conference (WWW)*, 2010.
[17] B. Krishnamurthy, P. Gill, and M. Arlitt, "A Few Chirps About Twitter," in *Workshop on Online Social Networks (WOSN)*, 2008.
[18] M. D. Choudhury, H. Sundaram, A. John, D. D. Seligmann, and A. Kelliher, "Birds of a Feather: Does User Homophily Impact Information Diffusion in Social Media?" *Computing Research Repository (CoRR)*, 2010.
[19] M. Cha, J. A. N. Perez, and H. Haddadi, "Flash Floods and Ripples: The Spread of Media Content through the Blogosphere," in *Conference on Weblogs and Social Media (ICWSM) Data Challenge Workshop*, 2009.
[20] D. Liben-Nowell and J. Kleinberg, "Tracing Information Flow on a Global Scale using Internet Chain-Letter Data," *Proceedings of the National Academy of Sciences (PNAS)*, 2008.
[21] E. Sun, I. Rosenn, C. Marlow, and T. Lento, "Gesundheit! Modeling Contagion Through Facebook News Feed," in *Conference on Weblogs and Social Media (ICWSM)*, 2009.