

# **Pinterest Analysis Final Report**

## **Ads Activities within Social Network**

**Project Advisor: Ning Xia**

**Xu Tian, Xiling Sun, Shuangze Tang**

**Northwestern University**

**June 11, 2013**

## **Abstract**

Pinterest is a pinboard-style photo-sharing website that allows users to create and manage theme-based image collections such as events, interests and hobbies. Retail companies have taken advantage of Pinterest for advertising and style trending. The web design provides an ideal layout for "style conscious retailers", where products can easily be visualized within a consumer context. In our Pinterest analysis, we performed a primitive analysis of Pinterest social network. We collected an initial dataset of Pinterest users through their other ONS account such as Facebook and Twitter, which shows several basic network features and large-scale behaviors of users. In addition, we dynamically explored the propagation of information (repin) in Pinterest, studying how the posts move through the social network and learning modeling information diffusion from online social network users.

# Contents

<b>1</b>	<b>Introduction .....</b>	<b>4</b>
1.1	Introductions and Significant Research .....	4
1.2	Project Background.....	5
1.3	Project Objective .....	5
<b>2</b>	<b>Methodology.....</b>	<b>5</b>
2.1	Overall.....	5
2.2	Data Collection.....	5
2.3	Social Analysis.....	6
<b>3</b>	<b>Results and Analysis.....</b>	<b>6</b>
3.1	Connection Between OSNs.....	6
3.2	Pinterest User Overall Analysis.....	7
3.3	Users Activity Analysis.....	7
3.4	Propagation Analysis.....	10
<b>4</b>	<b>Conclusion.....</b>	<b>13</b>
<b>5</b>	<b>References .....</b>	<b>14</b>

# 1 Introduction

## 1.1 Introductions and Significant Research

Online social networks (OSNs) continuously grow in popularity and produce a large source of Internet traffic. Therefore, significant effort has gone into understanding how each of these OSNs work. There are also important analyses performed on the Facebook and Twitter networks, studying the nature of user links, attempting to understand how users post and consume information, etc.

However, since Pinterest is relatively new, not so much research has been made to understand the Pinterest social network. Based on the first-step study, we found that there are several factors that distinguish Pinterest from other OSNs.

First, the Pinterest network derived from both Twitter and Facebook. In the past, the only way for users to login is via Facebook or Twitter. Nowadays, Pinterest allows users to connect their accounts via cross-logins. Pinterest has strong connection with Facebook and Twitter. In this way, we therefore are able to do some research on the migration dynamics between these networks.

Second, Pinterest is significantly designed as a model of sharing. This means that when a user like a posted image and like it, Pinterest encourages them to repost the image to their own boards. In Twitter, users can “retweets”. And in Pinterest, people can “repin”. Therefore, we did research on repins on Pinterest, analyzing the propagation of images. Pinterest is somewhat different from Facebook and Twitter, since the objects of posting, pinning, repinning is image. And Image is an excellent way for advertising. A better understanding about pin’s propagation will help to present a better advertising.

Third, like other research on user activities, in this project we can also do research on issues like: In which period of time in one day be users most active; In which period of day in a week be users most active; etc. These could propose constructive suggestion on advertisements releasing, etc.

## **1.2 Project Background**

It is important to understand the basic format of the Pinterest social network, before the beginning of our analysis. The basic idea of Pinterest derived from the pin board. When a user sees an image on the Internet that they like, they use a simple interface to “pin” the image. This pulls the image into the Pinterest network and associates it with one of the user’s pin boards. Users can follow each other, and are presented with a dashboard- style view of all the pins from the accounts they follow.

When viewing posts, users have several possible actions. They can repin any pins and comment on the pins in which they are interested. The repin is presented as a central component of Pinterest experience, resulting in significant re-sharing of images that is not like Twitter.

## **1.3 Project Objective**

- Understand the OSN users and their composition
- Understand Pinterest user activities
- Modeling information (pin/ads) diffusion from OSN users

# **2 Methodology**

## **2.1 Overall**

- Data Collection
  - Keep data collections
- Social Analysis
  - Graphics related analysis

## **2.2 Data Collection**

We used three different crawlers to get different information from Pinterest Website. We totally got 81, 438 unique user accounts’ profiles as well as users’ past activity, including

pins, repins, likes and comments etc.. As crawling the propagation information is very slow, we now got 49 pin's propagations. More data is coming.

## 2.3 Social Analysis

Based on the data we collected through our crawlers, we analyzed the data from several aspects. Firstly, we figured out the connection between Pinterest and other popular social networks such as Facebook and Twitter (Figure 1). Secondly, we explored patterns of user activity. Then, we got the users' information through their Facebook homepage. Then we compared different users groups of activity. Last but not the least, since Pinterest differs from earlier OSNs in its emphasis on repining as the primary action. The system is built on the idea of sharing images one sees with others, rather than generating original content. While previous platforms have included share (Facebook) and retweet (Twitter) actions, they have not done so with the same emphasis. Thus, we do experiment on the dynamics of a single post in the network could play host to significant complexity, so we focus on how the "repin" action works, namely, we want to know the information propagation of "repin".

## 3 Results and Analysis

### 3.1 Connection Between OSNs

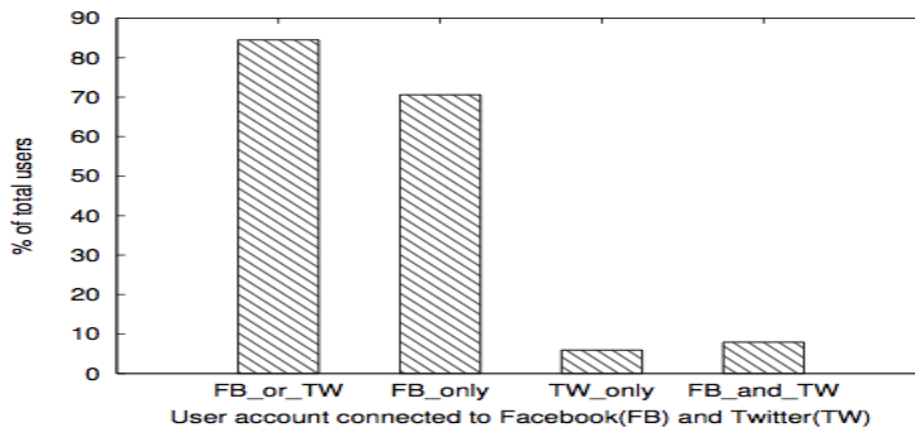


Figure 1. Pinterest user's ONS analysis

Figure 1 suggests that in spite of the structural similarities between Twitter and Pinterest, there is likely more insight to be gained from studying the user overlap with Facebook. The high link rate also provides us with an opportunity to collect additional information about Pinterest users via their Facebook profiles. As we know, Facebook generally has more information of users.

### 3.2 Pinterest User Overall Analysis

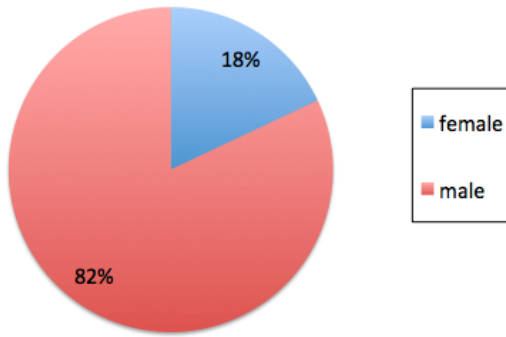


Figure 2. User Gender Analysis

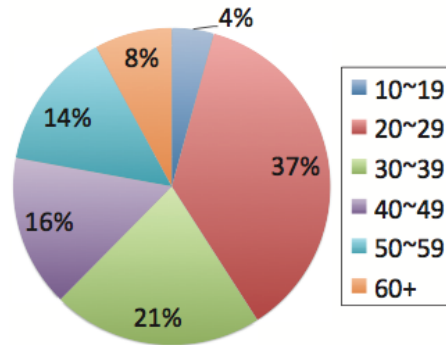


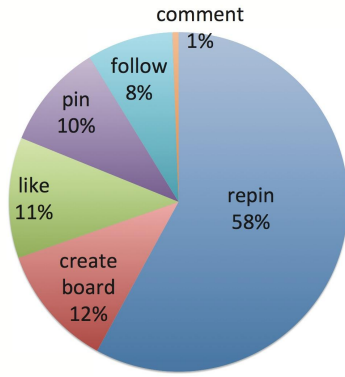
Figure 3. User Age Analysis

We collected user data through their Facebook information and generate Figure 2 and Figure 3. From the Figure 2, we see that Pinterest is mainly dominated by female users, making up 82% of users that our crawler got. The male users only account for 18%. According to Figure 3, we also see that younger users, with 40% of users under 30, mainly dominate the user base. We, however, also surprised with the number of users over 30, given Pinterest's hip Internet nature. Based on the analysis above, young women are dominant among all users.

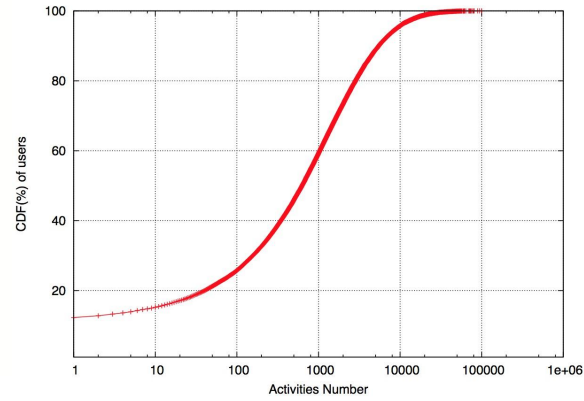
### 3.3 Users Activity Analysis

In this part, we analyze the patterns of user activity. The first (shown in Figure 4) is the fraction of user actions based on each activity. As we know, Pinterest is different from previous OSNs in its dominant action--repinning. Pinterest is built based on the idea of sharing images when users see or feel good rather create their own images. To understand

and give an apparent description of this, we measure the fraction of users actions devoted to each activity.



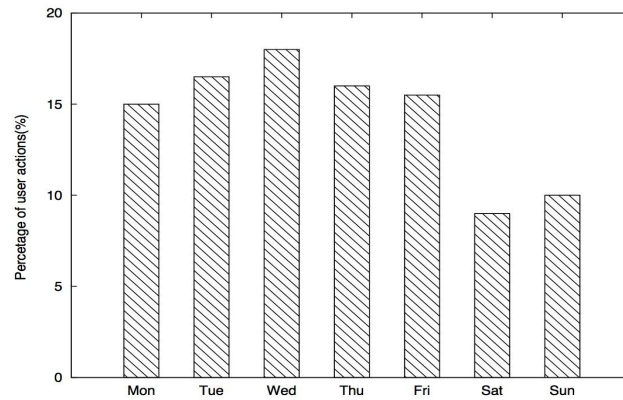
**Figure 4. Fraction of user action**



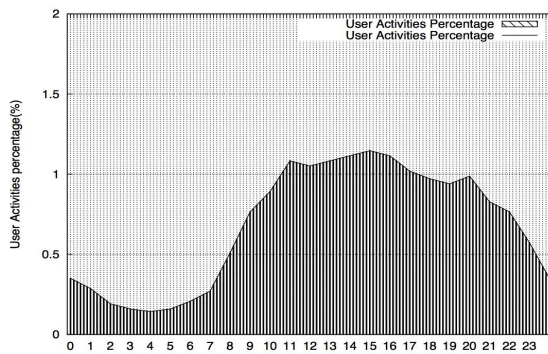
**Figure 5. CDF of Number of activities from each user**

Figure 4 illustrates the percentage of each type of activities. The percentage of each activities are as follows: “Create board” occupies 12%, “Like” occupies 11%, “Pin” occupies 10%, “Follow” occupies “8%”, “Comment” occupies 1% and “Repin” occupies 58%. Obviously, “repin” is the dominant activity. Based on the Figure 4, the percentage of “Create board”, ”Like”, ”Follow”, and “Pin” is almost same. The percentage of “Comment “ is very low, implying that in Pinterest, most users do not like make comments. Figure 5 is the CDF of number of activities from each user. The result is according to common sense. A fair amount of users (almost 20%) having few activities (less than 10). This implies that there are a relative high amount of users who did not use their account too much. The result is in accordance with relative research based on Facebook and Twitter. Another thing we found is that nearly 40% of accounts have more than 1,000 activities. In addition, there are a few fractions of users whose number of activities is very high. Based on Figure 5, 5% of users have more than 10,000 activities. This group is the most active part of all users, which is the foundation of our following part.

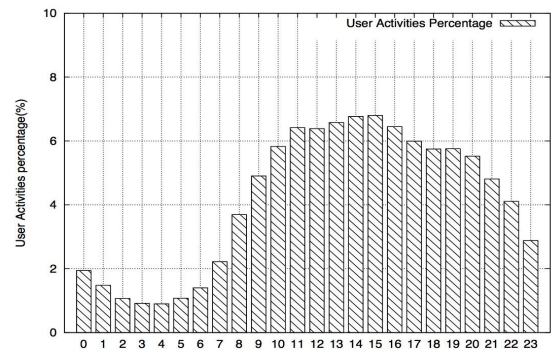




**Figure 6 User Activity Distribution in one week**



**Figure 7 User activities distribution in one day (every 10 minutes)**



**Figure 8 User activities distribution in one day (per hour)**

Figure 6, Figure 7 and Figure 8 show the active period of users' activities. From Figure 6, the most active day in a week is Wednesday. The least active days are weekends. The reason behind this could be that in weekends, users might go outside, with less access to Internet. Figure 7 and Figure 8 show the user activities distribution in different periods in one day every 10 minutes and every hour. According to the two pictures, users' most active period in one day is from 11:00 am to 5:00 pm. The results could provide constructive support for online advertising release.

The next part is for the total activity of sets of user accounts that were created 40, 30, 20, 10 weeks ago, shown in Figure 9. The result is also in accordance with common sense: users are most active when their accounts were created. After that, there is a relative rapid drop off as we can see on Figure 9. Furthermore, after 5 to 10 weeks, the user activity per

week is below 1. Another thing based on our observation is that older users fall off their activities more rapidly than new users. This suggests that users who remain active for longer (in this case 45 weeks) slow to relatively low activity quickly.

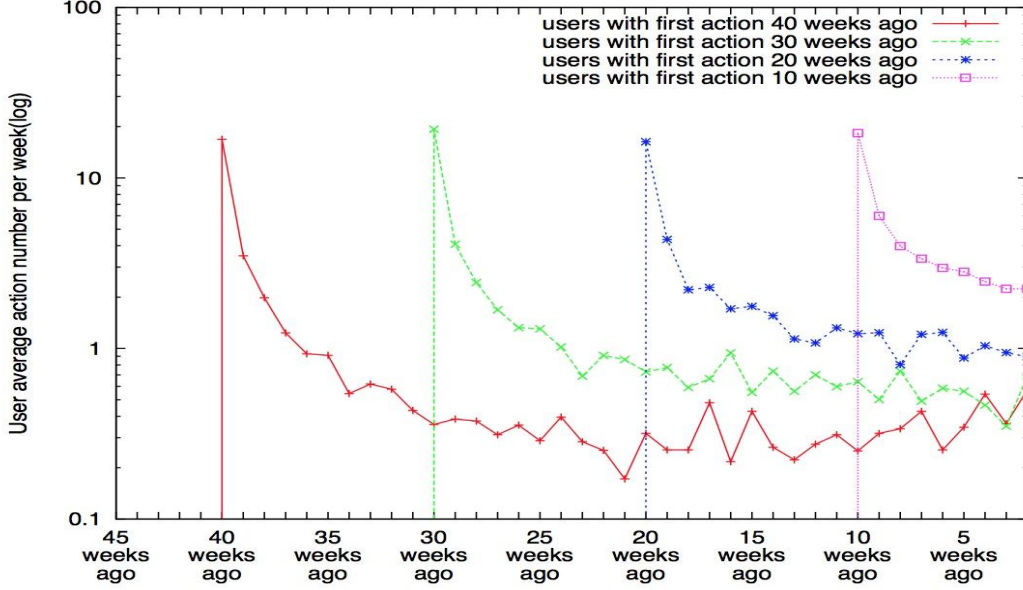
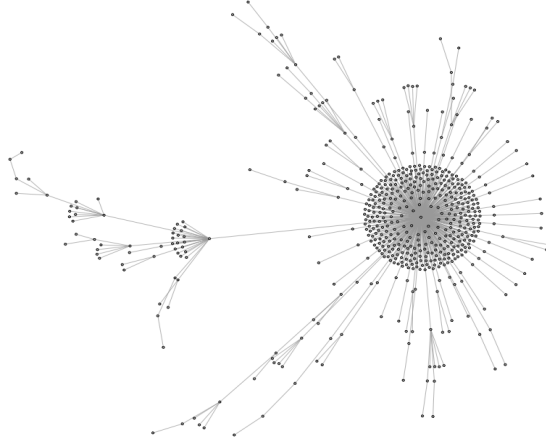


Figure 9 the total number of actions observed for the last 45 weeks.

### 3.4 Propagation Analysis

In this part, we want to model information (pin or ads) diffusion from Pinterest’s users. For this project, we did 49 experiments which one experiment related to only one original pin id. At last we got 49 information propagation ways. After analyzing the first five hops for all 49 information propagation ways, we found that the first hop of propagation decides the scale of the information diffusion. We also analyzed the time distribution of the first hop and found that most of the repins happened at the first 25 weeks.

Figure 10 is one example of pin propagation. And Table 1 is the first five hops of this pin. This pin has 563 repins. According to Figure 10, we can observe this pin’s propagation way very clearly. This pin has a huge number of first hops, some third hops and few fifth hops. These observations are identical with Table 1’s analytical result. According to Table 1, this pin has 403 first hops, 102 second hops, 46 third hops, 10 fourth hops and 2 fifth hops.

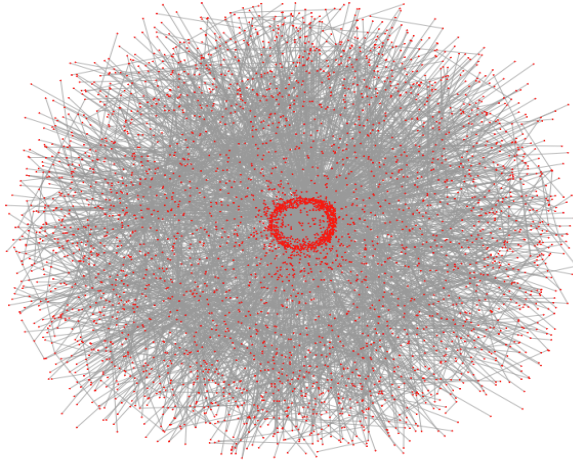


**Figure 10**

	# of nodes
1st Hop	403
2nd Hop	102
3rd Hop	46
4th Hop	10
5th Hop	2

**Table 1**

Figure 11 and Table 2 is another big pin example, which has 1575 repins. The information diffusion way of this example has an explosion effect. We can hardly figure out the propagation ways by eyes. According to our analysis, this picture has 663 first hops, 598 second hops and 314 third hops. This picture does not have any fourth hops and fifth hops at all.



**Figure 11**

	# of nodes
1st Hop	663
2nd Hop	598
3rd Hop	314
4th Hop	0
5th Hop	0

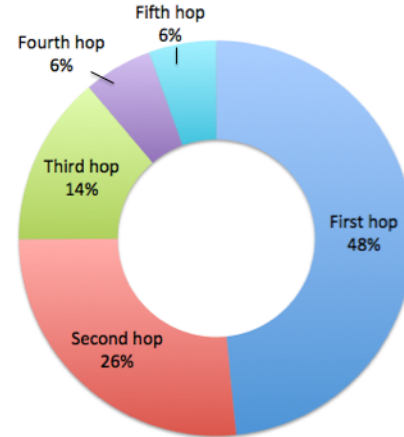
**Table 2**

These above two examples give us the similar observations, which we can conclude from all 49 propagation experiments. We analyze the first five hops for all 49 information diffusion ways. As shown in table 3, the average nodes number of the first hops is 201.80. And the first hop is nearly twice than the second hop. Other hops are even smaller than the second hop. Figure 12 could give us an exact analysis about the first five hops: first

hop makes up 48.40% of all nodes in propagation ways. Second hop makes up 26.53% of all nodes in propagation ways. Third hop, fourth hop and fifth hop make up the remaining 25.07% nodes.

	Avg # of nodes	Median # of nodes
1st Hop	201.80	92
2nd Hop	110.61	39
3rd Hop	58.02	28
4th Hop	23.45	10
5th Hop	23.08	2

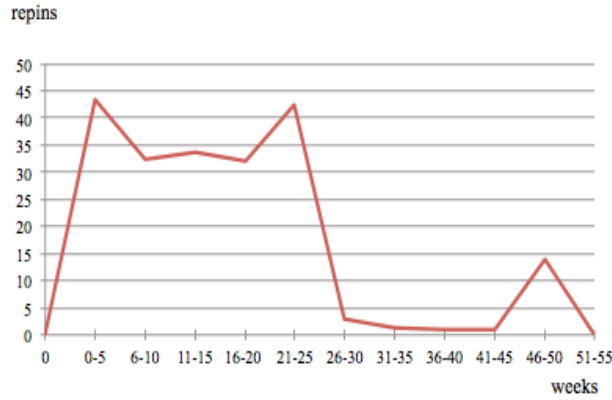
**Table 3**



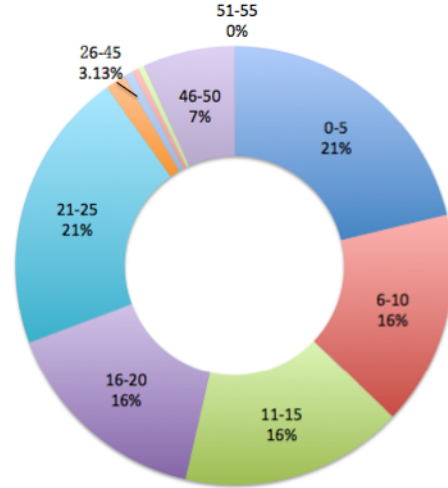
**Figure 12**

From the above analysis, we can get the conclusion that the first hop of propagation decides the scale of the information diffusion. Therefore, if we want to do a better advertising, we need focus on the first hop of our advertising in order to do a better advertisement.

As the first hop is so important for picture's (or advertising's) propagation, we analyzed the time distribution of the first hop for all 49 experiment results. We tried to find a time factor during all these propagations. According to Figure 13, we found that most of the repins happened at the first 25 weeks. After 25 weeks, there is almost no repins happened at all. For example, indicted by Figure 13, during 0-5 weeks after the picture was posted on Pinterest, there are more than 40 repins happened. The same situation happened during 21-25 weeks after the picture was posted on Pinterest. But after 25 weeks, there is almost no more than 5 repins happened for each 5-week interval. Figure 14 proved our observations. 90.1% repins happened in the first 25 weeks. After 25 weeks, there are only 9.9% repins happened.



**Figure 13**



**Figure 14**

According to the above analysis for time distributions of the first hop, we can conclude that the lifecycle of advertising could be as much as 25 weeks. The first 25 weeks is the diffusion time for advertising. And we can also ensure if the advertising is good or not after 25 weeks of its propagation.

## 4 Conclusion

In this report, we have undertaken a preliminary analysis of the Pinterest social network. Based on our objective, we explored the linking with other OSNs, demonstrating that Pinterest is highly linked to Facebook. And we also found that young female is dominant among all users. We were further able to gain insight into patterns of users and their activities. Finally, we explored the propagation way of Pinterest. After analysis for the whole pin propagation path, we found that the first hop of propagation decides the scale of the information diffusion and most of the repins happened at the first 25 weeks.

For future work, as we already have enough dataset, we plan to use some model to analyze user activities and information propagation. For example, we may use models in physics to analyze the propagation path of pictures and draw the heat map to show the energy flow (pictures propagation) on Pinterest. We also want to study the relationship between Peer Pressure and Authority Pressure during information propagation. We plan to write a paper based on our work in this project and submit it to a conference. Please

refer our research idea report (What is Pinterest, a Social Network or an Advertising Media?) for the details about our research plan for Pinterest.

## **5 References**

- [1] Gilbert, E., Bakhshi, S., Chang, S. and Terveen, L. 2013. "I Need to Try This!": A Statistical Overview of Pinterest. ACM Conference on Human Factors in Computing Systems.
- [2] Hall, C. and Zarro, M. 2012. Social Curation on the Website Pinterest.com. American Society for Information Science and Technology 75th Annual Meeting (ASIS&T).
- [3] Constine, J. 2012. "Pinterest Hits 10 Million U.S. Monthly Uniques Faster Than Any Standalone Site Ever -comScore." Retrieved 2/10/2013, from <http://techcrunch.com/2012/02/07/pinterest-monthly-unique/>.
- [4] Pinterest And Quora: Why Facebook's "Like" Button Is So Popular 2011. Forbes: 2012.  
<http://www.forbes.com/sites/victoriabarret/2012/01/04/pinterest-and-quora-why-facebooks-like-button-is-so-2011/>. Accessed: 2012-01-15.
- [5] Owen, Diana (2013) "Tumblr, Vimeo, Pinterest, Wikis, Mashups: Eloquentia Perfecta and the New Media Landscape," *Conversations on Jesuit Higher Education*: Vol. 43, Article 7.