Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media

# Ladies First: Analyzing Gender Roles and Behaviors in Pinterest

**Raphael Ottoni †** rapha@dcc.ufmg.br
**João Paulo Pesce †** jpesce@dcc.ufmg.br
**Diego Las Casas †** diegolascasas@ufmg.br
**Geraldo Franciscani Jr. †** gfrancis@dcc.ufmg.br
**Wagner Meira Jr. †** meira@dcc.ufmg.br
**Ponnurangam Kumaraguru ∗** pk@iiitd.ac.in
**Virgilio Almeida †** virgilio@dcc.ufmg.br

† Universidade Federal de Minas Gerais, Brazil
∗ Indraprastha Institute of Information Technology, India

### Abstract

Online social networks (OSNs) have become popular platforms for people to connect and interact with each other. Among those networks, Pinterest has recently become noteworthy for its growth and promotion of visual over textual content. The purpose of this study is to analyze this image-based network in a gender-sensitive fashion, in order to understand (i) user motivation and usage pattern in the network, (ii) how communications and social interactions happen and (iii) how users describe themselves to others. This work is based on more than 220 million items generated by 683,273 users. We were able to find significant differences w.r.t. all mentioned aspects. We observed that, although the network does not encourage direct social communication, females make more use of lightweight interactions than males. Moreover, females invest more effort in reciprocating social links, are more active and generalist in content generation, and describe themselves using words of affection and positive emotions. Males, on the other hand, are more likely to be specialists and tend to describe themselves in an assertive way. We also observed that each gender has different interests in the network, females tend to make more use of the network's commercial capabilities, while males are more prone to the role of curators of items that reflect their personal taste. It is important to understand gender differences in online social networks, so one can design services and applications that leverage human social interactions and provide more targeted and relevant user experiences.

## 1 Introduction

Online social networks (*OSN*s) have become popular platforms for people to connect and interact with each other in many ways. Common uses are posting photos, comments, videos, opinions, ideas and thoughts. Essentially users are bringing their private lives and their personalities to the network. According to *comScore* (comscore.com 2011), these networks reach $82\%$ of the world's population of internet users, making *OSN*s a great source of information, not only for research, but also for commercial purposes. More recently, *OSN*s such as Instagram and Tumblr, which are mostly based on pictures, presented a considerable gain in popularity (blog.instagram.com 2012) and that tendency can be confirmed by the fact that Facebook handles 300 million photos uploaded per day (Chan 2012). In this vein, Twitter recently released a mobile service called Vine (blog.twitter.com 2013) that let users share short videos of 6 seconds. Additionally, Instagram expanded the ability of seeing profiles and receiving your news feed directly from the web, without using their application (blog.instagram.com 2013). Pinterest was launched on march of 2010 as an effort to compete in this new trend with an innovative and pioneering paradigm: a pinboard-style image sharing network for people with good taste (Chafkin 2012).

According to a 2012 survey by the Pew Research Center, Pinterest has attracted $15\%$ of internet users to its virtual scrapbooking. Pinterest's users comprise mainly young people, the well-educated, those with higher income, and women. Pew report also indicates that women are about five times as likely to be on the website as men, the largest difference in gender of any social networking website (Duggan and Brenner 2012).

Pinterest stands today with the idea to connect people around the world based on shared tastes and interests through images. It is currently the world's 35th most popular website, the 15th most popular in the United States (alexa.com 2013) and it was the top Google 2012 trending search in Canada (google.com 2013). In social media, Pinterest is the fastest growing social media website in both unique visitors and clicks on search engines (comscore.com 2012), excluding Google+, that pre-created profiles for existing registered users. The network is over-represented by females (alexa.com 2013; Chafkin 2012) and, on average, the monthly usage time per visitor is 98 minutes, which makes Pinterest the second most used *OSN* in terms of time dedication, being only behind Facebook, which averages 405 minutes (statista.com 2012).

A study with nearly 700 million shopping sessions on leading U.S. retailers tried to uncover which channels are driving the most traffic and sales to their website (RichRelevance 2012). It was found that, although Facebook is responsible for $85\%$ of the traffic leading to the stores, and Pinterest only $11,3\%$, the average value of an order coming from Pinterest is much higher: \$168,83 against \$94,70 from Facebook.

Because it is a relatively new *OSN* (it has been around for only two years) and just recently became noteworthy for its impressive growth (comscore.com 2012), there are few references in the literature. Moreover, there is no official public API available for data collection, which makes the process even more laborious. To our knowledge this is the first study to conduct a large scale analysis of Pinterest. Our study focuses on gender-based analysis of user behavior and our contributions are the following:

- We develop a distributed crawler to collect a large dataset from Pinterest. Over a period of 50 days, we collected more than 2 million profiles, which comprise beyond 850 million images and videos *pinned* into more than 20 million boards.
- By analyzing the behaviour of users in the network, we are able to draw relevant conclusions on how different users interact with the service. We find that males and females have distinct motivations when using the *OSN*: women tend to use the website to search and keep a record of items of interest mainly related to products and services, while men tend to act as curators, keeping a collection that reflects their tastes.
- In a network where text is secondary and communication is image based, we study how social interactions are developed. We find that conclusions drawn by social researchers about gender, in which females are more social than males inside *OSNs*, hold true in the form of lightweight interactions such as likes and reciprocity.
- We perform an analysis on how users describe themselves in the network. We find that male users tend to be more assertive by using words associated with work, achievements and money while females tend to use words related to emotional appeal.
- By analyzing attributes that are related with popularity we develop an algorithm to detect self promoters. Furthermore, we found that a high percentage of users who have a website linked with their profiles are, in fact, self promoters.

The rest of the paper is organized as follows. We begin with a detailed description of the network and its peculiarities in Section 2. In Section 3 we presented our dataset and our metodology. Next we analyze the users behaviours in the network regarding gender in section 4. Finally, we discuss related works in Section 5 and summarize our findings and future work in section 6.

## 2 Pinterest

Pinterest is a pinboard-style image sharing social network, where everything is about photos and videos. Direct communications like private or public messages (Facebook's wall post) are not possible. The only textual interaction feasible is to comment on someone's content.The main idea of the network is to collect and share things users find interesting in an organized and categorized way. Each content - either image or video - is called a *pin*, which in turn, is part of a categorized *board*. There are in total 33 pre-defined categories, varying from "Women's Fashion" and "Hair Beauty" to "Geek" and "Tattoos". Every *pin* posted must have a description and those which are not uploaded must have a direct link to its original source in the web. An interesting fact which helps to understand the Pinterest community is that in its earliest days, the sign up was restricted to invitees only. Many of the invitations were given to groups of design bloggers who Silbermann (CEO and co-founder) himself personally invited, and they were given more invitations urging them to invite only others whose taste they respected. The welcome e-mail had a particularly telling phrase: *You must have good taste!* (Chafkin 2012).

### 2.1 Platform Description

All information in the network is public even to outsiders. The user is able to create a collection of boards, which is summarized in her profile along with a self description, a profile picture, and information about her activity and relationship with other users such as her *pins*, likes, board, users that she follows and follows her, as well the last fifty activities. Recently, but after our crawling period, Pinterest enabled the possibility of creating secret boards: basically, boards which only the owner has access (Milam 2012). Figure 1 shows a typical profile of a user that Silbermann's strategy successfully attracted to the network (Chafkin 2012).

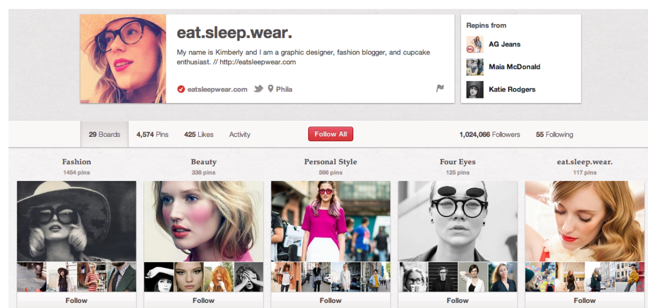


Figure 1: Profile page of a famous female designer/blogger

### 2.2 Social Relationships and Interactions

Relationships in the network are asymmetric. User A can follow user B without asking for B's consent, and this does not imply that B follows A. It is also possible to follow a user as a whole or a subset of her boards (*e.g.* user A is only interested in B's "favorite recipes" *pins*). Following someone means that your news feed (called Following feed) will be updated with content posted by the followed person.

Social interactions are enclosed inside the *pins*, *i.e.* users cannot send private or public direct messages to each other and the only textual interaction possible is to comment on a *pin*. Other forms of interaction are the action of liking a *pin*, *repining* someone's *pin* into one of your boards and sharing a board with other users. In a shared board, invited users can add *pins* and invite other users to be part of it. For this study, we consider as lightweight those actions that let users interact with others without having to spend a lot of time thinking about what to say. Such interactions include likes and *repins* but not comments. It is also possible to mention someone in a comment by adding an @ character followed by the user's ID, but the notification is sent by e-mail rather than specific communication mechanisms of the network.

### 2.3 Linkage with other Social Networks

At the time of our crawling, in order to join the network, the applicant must have either a Facebook or Twitter account to link with Pinterest. It is an opt-out situation, in the sense that once the Pinterest account is created it is possible to remove the association. As our study shows in Section 3, the majority of users do not remove it and, since Pinterest does not endorse textual information, there are few descriptive data about the user. These associated accounts may then act as a source of extra personal information - such as gender, in which this study is based.

## 3 Methodology

There is no official public *API* to gather data from Pinterest, thus we create our own distributed framework based on a client-server model. All gathering is based on HTTP requests, which is a challenge due to how the network was developed. A central element of Pinterest's design is the concept of "infinite scroll", a way to automatically load more content as the user expands the web-browser window horizontally or goes toward the bottom of the page (Chafkin 2012). This implies that our crawler can only gather 50 items per request, which means that, for a user with a million followers, we would need 20,000 requests just to collect her list of followers.

Another problem was the decision on how to sample the network and, since the users unique identification is textual and not numeric, we were unable to collect a random sample, thus we opted to use a breadth-first search (BFS) considering both the followers and followees list. Although the BFS technique is simple and efficient, it exhibits several well-known limitations such as the bias towards sampling high degree nodes, which may affect the degree distribution (Ribeiro and Towsley 2010; Gjoka et al. 2009). We could not find any press release of the network either showing a rank of the top most influencial or popular users in the network, hence we manually selected the most popular user (in number of followers) that fits the profile intended by Silbermann (Chafkin 2012) that we could find.

Our collection process began in August 21th, 2012 and ended in October 9th, 2012. In total we gathered information about 2,031,723 users, 861,566,305 *pins* and 21,890,927 boards. Since there is no press release reporting how many users are in the network, the only information available we know was gathered from third parties. ComScore, reported 10.4 million unique users in January 2012 (comscore.com 2012) and about 19.8 million users were counted by Appdata, on December 2012 (appdata.com 2012). It is important to know that AppData only gathers information through Facebook apps. In our whole database we verified that 90,6% of the users had a link with Facebook, therefore a rough approximation of the entire network is 21.85 million users. Our sample would then represent approximately 9.30% of the network.

We try to identify the gender of the users who have a connected Facebook profile by querying the Facebook API. By doing this, we are able to identify the gender of 97.28% of those users. Because of that, the majority of users whose gender we could not identify either registered using a Twitter account or removed the connection afterwards.

### 3.1 Working Dataset

Although we collected a representative portion of the network, we choose to randomly sample our dataset due to the expressive number of content generated by the users. As we reported in the introduction of Section 3, around 2 million users generated about 900 millions of *pins*. For this reason, we study a working dataset encompassing approximately one fourth of our total *pins*. The summary of this dataset is presented in Table 1.

| | Female | Male | Unknown | Total |
|---|---|---|---|---|
| Users | 550,436 | 29,644 | 103,193 | 683.273 |
| Boards | 5,237,981 | 198,871 | 1,162,398 | 6,599,250 |
| Pins | 174,076,885 | 5,142,786 | 50,441,472 | 229,661,143 |
| Website | 17,534 | 3,727 | 14,881 | 36,142 |
| Facebook | 550,436 | 29,644 | 16,948 | 597,028 |
| Twitter | 27,670 | 4,503 | 40,268 | 72,441 |
| Description | 79,178 | 6,817 | 37,227 | 123,222 |

Table 1: Summary of the working dataset

## 4 User behavior

In this section we present a detailed analyses of user behavior based on gender differences. Our effort is guided to better understand patterns of usage of the service.

### 4.1 Activity

As stated before, Pinterest is all about *pins*, thus our first analysis focuses on the activity of the users. By activity, we mean the amount of content generated. We calculate the complementary cumulative distribution function (CCDF) of boards and *pins* per user by gender (Figure 2) and report that, although the distribution of boards does not vary greatly by gender, females tend to catalog relatively more *pins* thus being more active in the network.

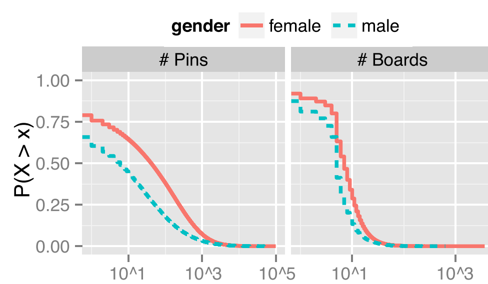


Figure 2: Distribution of the number of boards and pins per user by gender. Note that the x-axis are in log scale.

## 4.2 Social Interactions

Pinterest allows the following social interactions: to follow a profile or a specific board, to like or comment on a *pin*, to *repin* someone's *pin* (*i.e.* to *pin* someone else's content in one of your own boards).

In order to better understand the role of genders, we present the CCDF of different social links (followers, followees and followback) per user by gender (Fig. 3). We define the followback coefficient ($FB$) of each user by:

$$FB(u) = \frac{|FR(u) \bigcap FE(u)|}{|FE(u)|}, \quad (1)$$

where $FR(u)$ is the set of followers from user $u$ and $FE(u)$ is the set of followees of $u$. It represents the reciprocity of links that $u$ gets from her followees in the network.

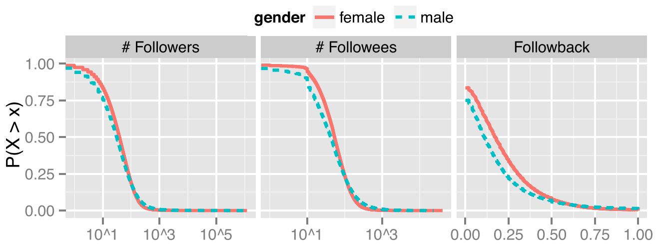


Figure 3: Distributions of followers, followees and followback (Eq. 1) per user by gender. Note that the x-axis are in log scale, except the *Followback* distribution.

Although females in general have more followers and followees, the difference between men and women is not as expressive as it is in the followback distribution (Figure 3). In a network where one way links suggest interest in someone's taste, reciprocity represent a stronger relationship, revealing a mutual content endorsement and a possible identification with each other's taste.

To analyze social interactions we summarized the amount of likes each user gave inside the network, as well the amount of likes, *repins* and comments they received on their *pins* (Figure 4). Females rate higher in all distributions, suggesting they are more prone to establish relations inside the network.

The action of following someone was analyzed in detail. In order to do that, we look at users that follow parts of content of other users. For each user, we calculate the average of the ratio between the followers of a board and her total number of followers. We call this metric *Full Profile Followers* ($FPF$) and it is defined by:

$$FPF(u) = \frac{1}{n} \times \sum_{i=1}^{n} \frac{F_B(u,i)}{F(u)}, \quad (2)$$

where $F(u)$ is the number of followers of user $u$, $n$ is the number of boards of $u$ and $F_B(u,i)$ is the number of followers of $u$'s board $i$.

We aim at answering the question *"If user A creates a new board, how many users will immediately follow and receive updates from that particular board?"*. As shown in figure 4, Pinterest users tend to follow others entirely and this behavior is not mediated by gender.

Another kind of social interaction explored was board sharing, for which we calculate the percentage of shared boards among the total amount of boards, separated by gender (Table 2). We also calculate *shared social links* ($SSL$), defined as the number of connections established with other users through sharing a board. This metric is defined by:

$$SSL(u) = \sum_{n=1}^{n} BP(u,n), \quad (3)$$

where $n$ is the number of shared boards of user $u$ and $BP(u,n)$ is the number of participants of $u$'s board $n$. The CCDF of *shared social links* of those with at least 1 shared link is shown in Figure 4. Males tend to share more boards (Table 2) and have more shared links (Figure 4).

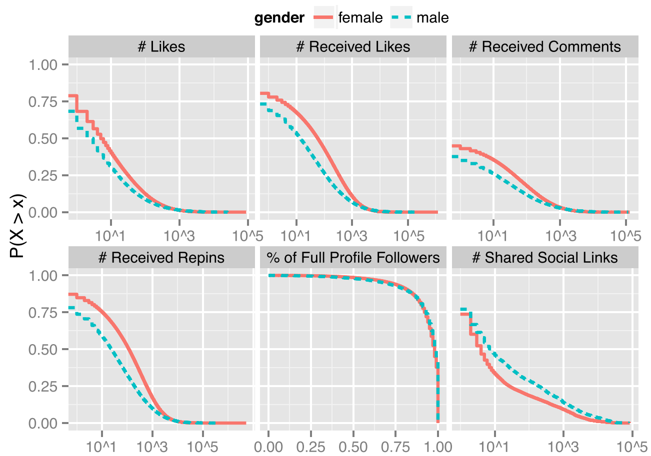


Figure 4: Distributions of social interactions per user by gender. *# Shared Social Links* distribution starts with $x = 1$, *i.e.*, we are only considering profiles with at least one shared board. Note that all x-axis are in log scale except the *% of Full Profile Followers*.

| Gender | # Users | Percentage (%) |
|---|---|---|
| Female | 53,588 | 1.02 |
| Male | 6,575 | 3.31 |
| Unknown | 21,269 | 1.83 |

Table 2: Percentage of users with at least 1 shared board by gender.

## 4.3 Usage Characteristics

Related studies on other social networks reported differences in users motivations and usage of the network (Haferkamp et al. 2012; Mazman and Usluel 2011; Muscanell and

Guadagno 2012; Joinson 2008). In this part of our study we focus on how the network is used by different genders.

**Specialists vs Generalists** First we analyze how specific or general users are, regarding the 33 pre-determined categories in the network. Basically we want to know, for each user, how the content is distributed among the categories. To measure it, we calculate the *Shannon Entropy* ($H$) of each user based on the distribution of their *pins*, which is defined by:

$$H(u) = -\sum_{n=1}^{n} p_i \log(p_i), \qquad (4)$$

where $n$ represents the number of categories present in user's $u$ profile and $p_i$ is the probability that a given *pin* of $u$ will be of category $n$. The most specialized user, will be the one who *pins* only in a single category and the most generalist will be the one who posts the same number of *pins* in all the 33 categories. The closer $H(u)$ is from 0, the more specific the user is.

Figure 5 shows the CCDF distribution of *category entropy*, and reveals that females are more generalists than males. In a similar way, we use this metric to understand the domain source of the *pin* in the web (*e.g.* a picture from www.deviantart.com). We use the same Equation 4, replacing $n$ with the number of different domains used by user $u$ and $p_i$ with the probability of a *pin* to be from that particular domain, to quantify the user's *source entropy*. The results are presented in Figure 5, which reveals once more that females are more generalists in their collection behavior.

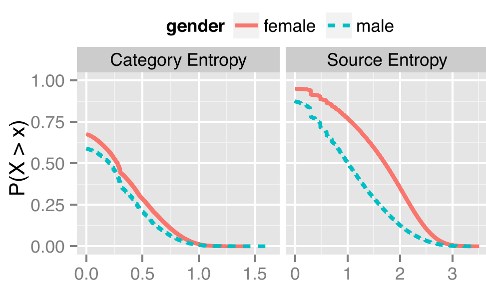


Figure 5: Shannon's Entropy for categories and sources per user by gender

At first sight, the high percentage of users with *category entropy* $H(u) = 0$ (Completely specialized) seems surprising, but it is explained by the amount of users that chose to not categorize their boards. If the board is uncategorized, its category is automatically set to "Null" or "None". In our working dataset $23.45\%$ of the females and $22.21\%$ of the males do not categorize any board, thus the high percentage of completely specialized users.

**Self Promoters** By calculating the source entropy of the users, we are able to determine which is the main source of *pins* from each user, just by observing the domain with the highest probability $p_i$. With this information we propose a simple algorithm to detect users who use Pinterest to promote their external website, hereby called *self promoters*. We do so by verifying if the main domain of a user is also her personal website described in the profile. Surprisingly, we were able to observe that a representative amount of users who have a website in their profiles are in fact self promoters (Table 3).

| | Users with Website | Self Promoters | Percentage (%) |
|---|---|---|---|
| Female | 17,543 | 2,354 | 13.43 |
| Male | 3,727 | 683 | 18.33 |
| Unknown | 14,881 | 3,207 | 21.53 |
| Total | 36,141 | 6,244 | 17.28 |

Table 3: Self Promoters identified by gender

The percentage of self promoters found by this simple approach could be an indicator that in fact Pinterest is becoming more related to e-commerce. Its also goes in line with Gauvin et al. (2010) findings, in which gender neutral profiles in Myspace were found to be usually commercial.

**Curatorship and Commercial Use** Pinterest has an interesting way of dealing with commercial products in the network. It treats dollar signs ($) as a special kind of character which shows that the *pin* is related to a "buyable" product, and clicking on the *pin*, often leads to the store which sells it. Thus, companies and stores can broadcast their products as *pins* to interested users, influential users can act as sellers through their profiles, and all users can compose wish-lists and collections of what they are willing to buy.

By counting the frequency of *pins* in each board category and ignoring those uncategorized, we were able to rank the most popular categories among genders. We also counted the amount of dollar-signed *pins* by category, which maps to the most commercial categories in the network (Table 4).

| Rank | Female | Male | General ($) |
|---|---|---|---|
| 1 | Food & Drink | Art | Women's Fashion |
| 2 | Women's Fashion | Photography | DIY Crafts |
| 3 | DIY Crafts | Other | Home Decor |
| 4 | Other | Food & Drink | Products |
| 5 | Home Decor | Design | Other |
| 6 | Hair Beauty | Women's Fashion | Weddings |
| 7 | Weddings | Travel | Hair Beauty |
| 8 | Design | Home Decor | Design |
| 9 | Art | Celebrities | Food & Drink |
| 10 | Humor | Film, Music & Books | Kids |

Table 4: Rank of the top categories

In order to compare these ranks encompassing the 33 pre-determined categories, we used the *Kendall rank correlation coefficient*, a measure of correspondence between two rankings. Values close to 1 indicate strong agreement, values close to -1 indicate strong disagreement. We used the tau-b version of Kendall's tau, which is defined by:

$$TauB(x, y) = \frac{(P - Q)}{\sqrt{(P + Q + T) \times (P + Q + V)}}, \qquad (5)$$

where $P$ is the number of concordant pairs and $Q$ is the number of discordant pairs, $T$ and $V$ are the number of ties in $x$ and $y$.

By comparing the female and male ranks with the *General ($)* commercial rank, we observe that the female-related coefficient was $0.62$, which represents a moderate correlation while the male-related was two times lower: $0.31$ (Table 6). This could indicate that female interests are more related with commercial content than male ones. However, the *General ($)* commercial rank could be biased because of the women's over representativeness. Due to this problem, we also rank the categories concerning the use of the dollar sign by gender [Tab. 5].

| Rank | Female ($) | Male ($) |
|---|---|---|
| 1 | Women's Fashion | Women's Fashion |
| 2 | DIY Crafts | Products |
| 3 | Home Decor | DIY Crafts |
| 4 | Products | Other |
| 5 | Other | Design |
| 6 | Weddings | Home Decor |
| 7 | Food & Drink | Art |
| 8 | Design | Technology |
| 9 | Hair Beauty | Men's Fashion |
| 10 | Kids | Geek |

Table 5: Rank of the top comercial categories by gender

The results of the comparisons between the five ranks are shown in Table 6. Given the strong correlation between female's commercial and general commercial *pins*, we can conclude that in fact the general rank is likely to be biased. It is also interesting to notice that there is a moderate correlation between male's commercial and the other commercial trends. For example, the last line of Table 6 shows that comercial *pins* made by males are moderately correlated with *General ($)* and *Female ($)*. This suggests that males commercial interests are, category-wise, similar to females.

There is also a dissociation between commercial and non-commercial *pinning* within genders, especially for males. If all categories were equally likely to have commercial *pins* as they had non-commercial ones, and the only differences were gender-related, then the correlation between, *Males* and *Males ($)* would not be so low (0.19). Females, on the other hand, show almost 3 times greater agreement (0.54). This three-fold increase in agreement suggests that females *pinning* behavior is more coherently related to the commercial use of the network.

| | General ($) | Female | Female ($) | Male | Male ($) |
|---|---|---|---|---|---|
| Female | 0.62 | * | 0.54 | 0.29 | 0.47 |
| Female ($) | 0.82 | 0.54 | * | 0.28 | 0.57 |
| Male | 0.31 | 0.29 | 0.28 | * | 0.19 |
| Male ($) | 0.59 | 0.47 | 0.57 | 0.19 | * |

Table 6: Kendall Tau-B between category ranks

## 4.4 User Portfolio Analysis

To further study user categories, we focus on the portfolio of boards employed by each user. Our starting point is the set of populated boards from each user. We then create two datasets, comprising male and female users. Our characterization strategy is based on the frequent and statistically significant sets of boards. We first determine the frequent sets of boards for each dataset, and then determine, using swap randomization (Gionis et al. 2007) the most significant sets. The first step of swap randomization is to generate random datasets that present the same row and column margins of the original dataset. It then compares the frequent sets from the real data against the frequent sets from the random datasets, determining which sets are random artifacts and which sets stem from positive or negative correlations among boards. The last step contrasts the most significant sets of boards for males and females, highlighting the sets that characterize each gender.

Results are stated in Tables 7 and 8, where we show the top 10 most significant item sets by gender and its position on the other gender's rank. The highlighted item sets are those who represent significantly more their respective gender by not being in the top 10 of the differing rank (*e.g.* Women's Fashion, Travel, Home Decor seems to be strong related with females). It is also possible to recognize item sets that appears to be gender neutral by its strong presence in both ranks (*e.g.* Food & Drink,Diy Crafts).

| Female Rank | Item Set | Male Rank |
|---|---|---|
| 1 | Food & Drink; Home Decor; DIY Crafts | 5 |
| 2 | Food & Drink; DIY Crafts | 3 |
| 3 | **Women's Fashion; Home Decor** | **481** |
| 4 | Food & Drink; Home Decor | 4 |
| 5 | **Food & Drink; Home Decor; Women's Fashion** | **94** |
| 6 | Travel; Food & Drink; Home Decor | 8 |
| 7 | **Women's Fashion; Travel; Home Decor** | **556** |
| 8 | **Food & Drink; Womens Fashion; Travel; Home Decor** | **362** |
| 9 | **Travel; Home Decor** | **20** |
| 10 | DIY Crafts; Home Decor | 9 |

Table 7: Top most frequent item sets for females, comparing with males rank

| Male Rank | Item Set | Female Rank |
|---|---|---|
| 1 | Food & Drink; Home Decor | 4 |
| 2 | **Design; Art** | **49** |
| 3 | Food & Drink; DIY Crafts | 2 |
| 4 | **Film, Music & Books; Celebrities** | **87** |
| 5 | Food & Drink; Home Decor; DIY Crafts | 1 |
| 6 | **Film, Music & Books; Travel** | **31** |
| 7 | **Travel; Celebrities** | **75** |
| 8 | Travel; Food & Drink; Home Decor | 6 |
| 9 | Home Decor; DIY Crafts | 10 |
| 10 | **Film, Music & Books; Travel; Celebrities** | **116** |

Table 8: Top most frequent item sets for males, comparing with females rank

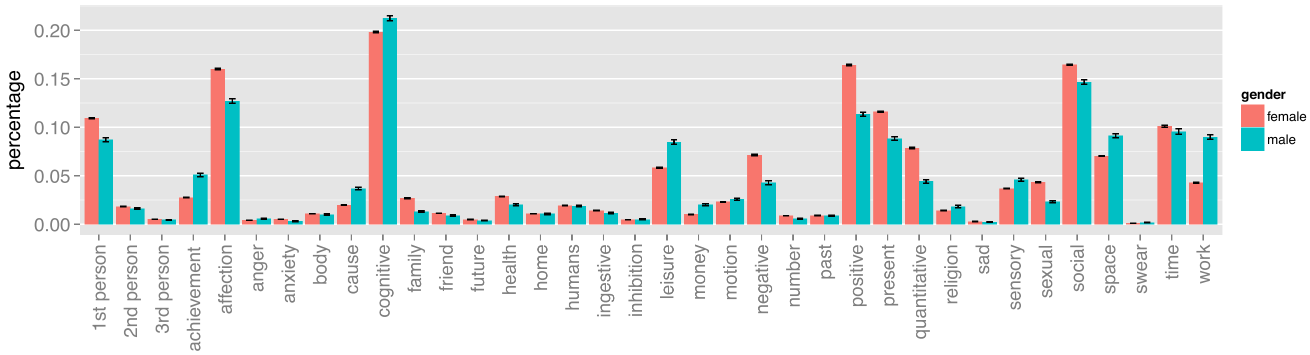


Figure 6: Mean of the percentage of representativity of each language dimension in users description by gender

## 4.5 User's Description Analyses

Computerized text analysis focusing on specific words or classes of words has been broadly used for studying emotional, cognitive, structural and components of individuals language (Tausczik and Pennebaker 2009; Kahn et al. 2007; Veltman 2006). In this section, we present a thorough analysis of the Pinterest users description by gender related to the context. For this purpose, we submitted all user's descriptions to LIWC (Pennebaker, Francis, and Booth 1999; Pennebaker et al. 2007), a system that analyzes text files on a word-by-word basis, calculating the percentage of words that match each of the several language dimensions. In this scenario, we considered 39 of the pre-set dimensions and evaluated the results obtained by detecting meaning of emotionality, social relationships, thinking styles, and individual differences.

After matching the words contained in the users descriptions with the language dimensions in LIWC, we calculate the mean of the percentage and the standard error of each dimension for all users separated by gender (Fig 6). The main goal of those measures is to compare the same dimension between users of different genders so we can verify socio-linguistic findings reported in previous studies (Holmes 1993; Bergvall 1999; Cunha et al. 2012; Coates 2004). In our analyses we confirm that females are more prone to use terms that convey affection (Holmes 1993; Kivran-Swaine et al. 2012) by observing that they use more words of fondness and positive emotions to describe themselves, and males tend to interact in ways that assert to their power and status (Holmes 1993; Bergvall 1999), which in our case, is related to the use of words connected to *work*, *achievements* and *money* to describe themselves.

# 5 Related Work

Gender differences have been widely assessed and debated in the specialized literature for years. Some recent, wide-ranging psychological studies have indeed found consistent patterns of differentiation between genders, (Feingold 1994; Costa, Terracciano, and McCrae 2001; Srivastava et al. 2003), although the extent to which this differentiation goes is still far from consensus (Hyde 2005). In general, men are found to be more assertive and aggressive, whereas women are reported as more extroverted, tender-minded, trustful, but also as more anxious and with a slightly lower self-esteem (Feingold 1994; Hyde 2005). The strength of the differences change with age (Srivastava et al. 2003), and surprisingly, is stronger in developed, egalitarian countries (Schmitt et al. 2008).

It is not clear the extent to which these differences can be attributed to social roles, cognitive functioning or self-perception. It is argued, however, that the observation of gender differences is tightly related to the context of data collection (Hyde 2005). Since we are focusing in the online environment, we will refrain from making offline generalizations. Now we review what kinds of characterizations on gender differences were made online, which may be more comparable to our results.

Cunha et al. (2012) concluded that gender can be a social factor that influences in the choice of hashtags related to voting by analysing over 650 thousands tweets. They observe that females use hashtags that denote more personal involvement, whereas males use more persuasive tags. Also in this line is the work of Kivran-Swaine et al. (2012), who finds that the stream of positive emotional tweets is more likely related to females, especially in female-to-female dyads. We find comparable results in our description analysis, in which females use more words of the *first person*, *present tense* and *positive affect*, and males words related to *achievements* and *cause*.

Joinson (2008) recruited Facebook users in a effort to understand the motivations of use this particularly *OSN*, his findings showed that posting and sharing pictures is one of the main reasons for females to visit the service. Recently Haferkamp et al. (2012), by performing an online survey with users of StudiVZ (German Social network for students, similar to Facebook), concluded that females tend to use *OSNs* more likely to search for information and compare themselves with others while males, on the other hand, tend to use it to find friends. We find that females are more active on generating content and are much more representative inside Pinterest, which is coherent with these findings.

In 2009, Bond (2009) conducted a study focused on how gender could be a factor influencing self-disclosure on social networks, and found that females tend to reveal themselves in a wider variety of topics than men. Also females were marginally more likely to report being sexually expressive on their profiles. We find similar results, since females were substantially more likely to be sexually expressive on their descriptions (Figure 6) and were more generalist, category-wise, on their content (Figure 5).

More recently, Szell and Thurner (2013) performed a similar study on how differently users of distinct genders organize their social network in an online-game society about 300 thousand players. They reported that females have more communication partners and are more active in positive actions (making friends, send messages and communicate in general). In our study we showed it is also true in Pinterest's network in the form of lightweight interactions (Figure 4). They also found that exist a strong homophily-related effect for the female network as they invest more effort in reciprocating social links than males, which we also varify with the followback coefficient showed in Figure 3.

Most existing works focus in popular social networks where direct interactions and textual communication are of paramount importance. To our knowledge, no research paper focused on user behavior, based on gender, was conducted in a heavy image-based network such as Pinterest. This work provides a first look into this matter on a relative new social network that grew quickly in importance and became one of the most popular and peculiar social network.

## 6 Conclusion

In this paper we study users behaviors, social interactions and characteristics in a network recognized by its appeal to visual content and lack of direct social communications. Our analyses are particularly focused on gender differences and based on a data set of more than 220 million items generated by 683,273 users. Pinterest does not publish demographic information about its users but reveals connections to either Facebook or Twitter. In our working dataset, $87.38\%$ of all users have a direct link to their Facebook accounts, thus using the Facebook public API we are able to identify her/his gender.

In this study we report significant findings. For instance, even with Pinterest restrictions for social communication, we show that such communication still happens in the form of lightweight interactions such as likes and *repins*. Moreover, we report that females make more use of this kind of interaction and are more active in terms of content generation. Our study also corroborates the findings of Szell and Thurner (2013), by showing that females invest more efforts in reciprocating social links, in the form of reciprocity in follower/followee relation. We report that females tend to be more generalist than males by calculating the entropy based on both the category of the board and the domain of the *pin*. We also execute a swap randomization test with the categories of boards of each user by gender and concluded that there are item sets significantly related with gender (*e.g. Women's Fashion*, *Travel* and *Home Decor* for females) and gender neutral ones (*e.g.* Food Drink and DIY Crafts). By calculating the most used *pin* source of each user and comparing it with the user's personal website, available in her description, we are able to identify self-promoters, people who use Pinterest to promote their outside webpage. By this approach we conclude that roughly two in each ten users who have a webpage linked in their profiles are in fact self-promoters, and the proportion of males in this situation is higher than females.

We also show that females tend to make more use of the network's commercial capabilities by comparing the categories rank in frequency for males and females, and then crossing this comparison with the frequency of *pins* with dollar signs ($). In addition we use the same metric to correlate the categories ranks of interest of each gender and concluded that they have substantial difference of interest inside the network. With a socio-linguistic analyses on how users describe themselves using LIWC we corroborate with conclusions drawn by the literature were females are more prone to use words of fondness and affection while males tend to describe themselves in an assertive way (Holmes 1993; Kivran-Swaine et al. 2012; Bergvall 1999).

**Future Work** There are several interesting directions for future research. First we plan to use the bag of features paradigm for image classification and texture recognition, believing that the *pin*'s core, the image itself, holds important information about the users interest. Although our algorithm to detect self promoters presented good results, there are possible optimizations. Gauvin et al. (2010) study suggest that gender neutral profiles tend to be commercial, thus we could use their behavior to improve our algorithm. We are also interested in socio-linguistic studies, for instance we could use LIWC on *pin*'s description in order to map a linguistic pattern for each pre determined category and then try to reveal the underlying category of those uncategorized boards.

By analyzing *pins* containing a dollar sign ($) in its description, we show that females tend to use more the commercial aspect of the network. However, we can not say whether these patterns arise from different *pinning* styles - with some users contributing much more to commercial *pinning* (*Shoppers*) and others focusing in non-commercial collections (*Curators*) - or from different commercial possibilities inside each category. If the first hypothesis is true, then the *shoppers* will be a subset of users who use Pinterest differently from the others, either for selling or for buying content. If, otherwise, the later is true, then there is a subset of categories with commercially-appealing content - and these are *not* the ones users *pin* the most.

**Acknowledgements** This work was partially supported by CNPq, Capes, Fapemig, and InWeb – National Institute of Science and Technology for the Web.

## References

alexa.com. 2013. Statistics summary for pinterest.com. http://bit.ly/pXz8LL.

appdata.com. 2012. Pinterest. http://bit.ly/rmGWWJ.

Bergvall, V. L. 1999. Toward a comprehensive theory of language and gender. *Language in Society* 28(2):273–293.

blog.instagram.com. 2012. Year in review: 2011 in numbers. http://bit.ly/vPJzXb.

blog.instagram.com. 2013. Introducing your instagram feed on the web. http://bit.ly/WMC1mX.

blog.twitter.com. 2013. Vine: A new way to share video. http://bit.ly/Wi995S.

Bond, B. J. 2009. He posted, she posted: Gender differences in self-disclosure on social network sites. *Rocky Mountain Communication Review* 6(2):29–37.

Chafkin, M. 2012. Can ben silbermann turn pinterest into the world's greatest shopfront? http://bit.ly/OgmHYl.

Chan, C. 2012. Introducing your instagram feed on the web. http://bit.ly/UHyjJg.

Coates, J. 2004. *Women, men and language : a sociolinguistic account of gender differences in language*, volume 3rd ed. Longman.

comscore.com. 2011. It's a social world: Top 10 need-to-knows about social networking and where i's headed. http://bit.ly/THm9zN.

comscore.com. 2012. State of us internet in q1 2012. http://bit.ly/V9L10Y.

Costa, Paul, J.; Terracciano, A.; and McCrae, R. R. 2001. Gender differences in personality traits across cultures: Robust and surprising findings. *Journal of Personality and Social Psychology* 81(2):322–331.

Cunha, E.; Magno, G.; Almeida, V.; Gonçalves, M. A.; and Benevenuto, F. 2012. A gender based study of tagging behavior in twitter. *Proceedings of the 23rd ACM conference on Hypertext and social media - HT '12* 323.

Duggan, M., and Brenner, J. 2012. The demographics of social media users. http://bit.ly/XOWHJq.

Feingold, a. 1994. Gender differences in personality: a meta-analysis. *Psychological bulletin* 116(3):429–56.

Gauvin, W.; Ribeiro, B.; Towsley, D.; Liu, B.; and Wang, J. 2010. Measurement and gender-specific analysis of user publishing characteristics on myspace. *Network, IEEE* 24(5):38–43.

Gionis, A.; Mannila, H.; Mielikäinen, T.; and Tsaparas, P. 2007. Assessing data mining results via swap randomization. *TKDD* 1(3).

Gjoka, M.; Kurant, M.; Butts, C. T.; and Markopoulou, A. 2009. A walk in facebook: Uniform sampling of users in online social networks. *CoRR* abs/0906.0060.

google.com. 2013. 2012 search trends:canada. http://bit.ly/XSzvgG.

Haferkamp, N.; Eimler, S. C.; Papadakis, A.-M.; and Kruck, J. V. 2012. Men are from mars, women are from venus examining gender differences in self-presentation on social networking sites. *Cyberpsychology behavior and social networking* 15(2):91–8.

Holmes, J. 1993. Women's talk: The question of sociolinguistic universals. *Australian Journal of Communication* 20(3):125–49.

Hyde, J. S. 2005. The gender similarities hypothesis. *The American psychologist* 60(6):581–92.

Joinson, A. N. 2008. Looking at, looking up or keeping up with people?: motives and use of facebook. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, 1027–1036. ACM.

Kahn, J. H.; Tobin, R. M.; Massey, A. E.; and Anderson, J. A. 2007. Measuring emotional expression with the linguistic inquiry and word count. *The American journal of psychology* 120(2):263–286.

Kivran-Swaine, F.; Brody, S.; Diakopoulos, N.; and Naaman, M. 2012. Of joy and gender: emotional expression in online social networks. CSCW '12, 139–142. ACM.

Mazman, S., and Usluel, Y. 2011. Gender Differences in Using Social Networks. *Turkish Online Journal of Educational Technology* 10(April):133–139.

Milam, E. 2012. Announcing secret boards for the holidays! http://bit.ly/UnElBn.

Muscanell, N. L., and Guadagno, R. E. 2012. Make new friends or keep the old: Gender and personality differences in social networking use. *Computers in Human Behavior* 28(1):107 – 112.

Pennebaker, J. W.; Chung, C. K.; Ireland, M.; Gonzales, A.; and Booth, R. J. 2007. The development and psychometric properties of liwc2007 the university of texas at austin. *Development* 1(2):1–22.

Pennebaker, J. W.; Francis, M. E.; and Booth, R. J. 1999. *Linguistic Inquiry and Word Count*. Lawrence Erlbaum.

Ribeiro, B. F., and Towsley, D. F. 2010. Estimating and sampling graphs with multidimensional random walks. *CoRR* abs/1002.1751.

RichRelevance. 2012. Richrelevance shopping insights study reveals how social networks stack up for retailers. http://bit.ly/P7HJKY.

Schmitt, D. P.; Realo, A.; Voracek, M.; and Allik, J. 2008. Why can't a man be more like a woman? Sex differences in Big Five personality traits across 55 cultures. *Journal of personality and social psychology* 94(1):168–82.

Srivastava, S.; John, O. P.; Gosling, S. D.; and Potter, J. 2003. Development of personality in early and middle adulthood: Set like plaster or persistent change? *Journal of Personality and Social Psychology* 84(5):1041–1053.

statista.com. 2012. Pinterest u.s. user engagement growth from may 2011 to january 2012, based on average minutes per visitor. http://bit.ly/W8M5oO.

Szell, M., and Thurner, S. 2013. How women organize social networks different from men. *Scientific Reports* 3(i):20–22.

Tausczik, Y. R., and Pennebaker, J. W. 2009. The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of Language and Social Psychology* 29(1):24–54.

Veltman, B. R. 2006. *Linguistic analysis of the semantic content of the Rorschach Inkblot Test.* Ph.D. Dissertation.