"Woman-Metal-White vs Man-Dress-Shorts": Combining Social, Temporal and Image Signals to Understand Popularity of Pinterest Fashion Boards

Suman Kalyan Maity,[1*] Anshit Chaudhary,[2] Animesh Mukherjee[2]
[1]Northwestern University; [2] Dept. of CSE, IIT Kharagpur, India
suman.maity@kellogg.northwestern.edu, chaudharyanshit@gmail.com, animeshm@cse.iitkgp.ernet.in

Abstract

Pinterest is a popular photo sharing website. Fashion is one the most popular and content generating category on this platform. Most of the popular fashion brands and designers use boards on Pinterest for showcasing their products. However, the characteristics of popular fashion boards are not well-known. These characteristics can be used for predicting popularity of a nascent board. Further, newly formed boards can organize their content in a way similar to the popular fashion boards to garner enhanced popularity. What properties on these fashion boards determine their popularity? Can these properties be systematically quantified? In this paper, we show how *social*, *temporal* and *image* signals can together help in characterizing the popular fashion boards. In particular, we study the sharing/borrowing behavior of pins and the image content characteristics of the fashion boards. We analyze the sharing behavior using social and temporal signals, and propose six novel yet simple metrics: *originality score*, *retention coefficients*, *production coefficients*, *inter-copying time*, *duration of sharing* and *speed coefficients*. We further study the image based content properties by extracting *fashion*, *color* and *gender* terms embedded in the pin images. We observe significant differences across the popular (highly followed or highly ranked by the experts) and the unpopular (less followed) boards. We then use these characteristic features to early predict the popularity of a board and achieve a high correlation of 0.874 with low RMSE value. Our key observation is that likes and repin retention coefficients are the most discriminatory factors of a board's popularity apart from the usage of various color, gender and fashion terms.

Introduction

Pinterest is an image-based online social network which has grown with unprecedented pace attaining a mark of 110 million monthly active users. It was also the fastest site to break the 10 million unique visitors mark[1]. Although Pinterest is fairly new in the social media gamut, it is being heavily used by many big business houses like Etsy, The Gap, Allrecipes, Jettsetter, Nike, Adidas etc. to advertise their products. Further, Pinterest drives more revenue per click than Twitter or Facebook[2]. This stupendous growth makes it interesting to study Pinterest.

Fashion industry and social media

Social media is an amazing marketing tool for the fashion industry. Fashion brands can leverage public perception available on social media over various fashion items. The continuous feedback received by the fashion brands in the form of likes and comments on their social media posts lets them gauge and further viralize their product chain in the market. Image-based social media platforms like Pinterest, Instagram have become popular venues for fashion brand marketing and advertising.

Role of Pinterest in fashion industry

Fashion is an integral part of Pinterest. All the major brands like Nike, Adidas Originals, Dolce Gabbana, Louis Vuitton etc. have their presence on the Pinterest platform. What are the characteristic features of these popular brands? Do they bear certain signatures – social, temporal or image based – that make them distinct from the not-so-popular ones?

In Pinterest, users save images (*pins*) and categorize them on different *boards*. Thus, a board is an important entity in Pinterest, and it has various influences on interest-driven pin propagation or pin sharing. Sharing is an important aspect in social media. If one likes a content, one might tend to use it in the same/modified form. On Pinterest, sharing and borrowing of pins (images) from various boards is a routine phenomena. This motivates us to consider sharing/borrowing behavior in understanding the popularity of the fashion boards. Similarly, we hypothesize that the (image) content of the post should also be a key factor determining the popularity of a fashion board.

There are multiple boards where Pinterest advertises various fashion contents. In this paper, we study the popularity of fashion boards by analyzing their originality, sharing/borrowing behavior, and the characteristics of the image content.

*Most of the work was done when the author was at IIT Kharagpur, India.

[1]https://techcrunch.com/2012/02/07/pinterest-monthly-uniques/

[2]https://en.wikipedia.org/wiki/Pinterest

Research objectives and contributions

In this paper, we analyze a massive Pinterest dataset consisting of pins and boards and make the following contributions.

- We investigate three major factors that can potentially characterize popularity of a fashion board: *social*, *temporal* (i.e., how pins are shared or borrowed across boards) and *image* content characteristics (usage of fashion, color and gender terms etc. in the image) of the pins belonging to a board.
- We observe that *generally popular fashion boards are able to make an existing non-popular pin popular*, whereas less popular fashion boards do not exhibit this characteristic. Note that this is a very non-intuitive finding indicating a non-assortative behavior (the 'popular' pins making the non-popular pins popular) as opposed to what is usually observed in most social networks.
- Another key observation is that *same content in different boards achieve different levels of popularity*. If a pin has originated from a popular board, it achieves higher popularity on the originating board than the subsequent boards to which it gets shared possibly pointing to dampening of the popularity due to re-sharing. In addition, *pins keep getting shared for longer durations in popular boards*.
- We perform extensive image analysis of the pins on the boards and extract fashion, color, and gender terms from them. Popular fashion boards have *more female faces* than the unpopular ones. Further, the popular boards have a rich *collection of pins in which both the gender co-appear*. We also observe significant characteristic differences between popular and unpopular boards in the usage of *color and fashion words*.
- Our characterization further helps us to predict whether a given fashion board would become 'popular' or not. In precise, we attempt to predict the popularity in terms of the future number of followers of the boards. We achieve a very high correlation coefficient of 0.874 with very low RMSE. A post-hoc analysis of the importance of the features indicates that the likes and the repin retention coefficients are the most discriminative ones followed by the color and gender terms embedded in the image.

Lessons for newbie fashion houses

The insights gained from this work can highly impact the new and upcoming fashion brands. For instance, allowing for more female faces or both male and female faces together, certain color terms (white, black, blue, brown, pink etc.) and color combinations (blue-pink, black-pink etc.) can increase the chances of the boards getting popular. Also they could 'engineer' campaigns to promote their boards in such a way that the originating boards are able to retain the 'likes' and 'repins' of their pins in the face of constant sharing of these pins.

Related work

Content characteristics, sharing, and engagement

Content sharing ensures user engagement and commitment in future (Burke, Marlow, and Lento 2009). There are diverse motivations to share content on social media (Lee and Ma 2012). Apart from network structure, the content matter also play important role in sharing, engagement (Maity, Kharb, and Mukherjee 2018; 2017). Users may share useful content to appear knowledgeable or simply to help out others (Wojnicki and Godes 2008). The emotional valence behind content also drive its extent of being shared (Berger and Milkman 2012).

Though there have been various studies on diffusion, sharing, engagement in social media, very less work has been done in the domain of visual analysis of image content. Hochman et al. (2012) show differences in local color usage, cultural production rate, varied hue intensity (blue-gray in New York vs red-yellow in Tokyo) by analyzing images from New York and Tokyo posted on Instagram. Bakhshi et al. (2014) study the engagement characteristic of images containing human faces. They observe that images with human faces in them, have higher chances of receiving likes and comments. Bakhshi and Gilbert (2015) study the role of color in online diffusion of pins in Pinterest. They observe that color significantly impacts the diffusion of images and adoption of content. Red, purple and pink seem to promote diffusion, while green, blue, black and yellow suppress it.

Popularity: There have been few studies in the domain of Fashion trend and popularity. Sanchis-ojeda et al. (2016) explore various statistical models using clients' temporal reaction to style units change for identification and quantification of linear and cyclical fashion trends. Lee et al. (2017) propose a classifier for identification of fashion-related Twitter accounts whereas we focus on understanding the popular fashion boards. Hessel et al. (2017) propose a relative popularity prediction framework based on content characteristics with minimal influence of other external factors like timing effects, community preferences, and social networks. Wu et al. (2017) study the sequential prediction of popularity for image posts using a deep learning framework by incorporating temporal context and temporal attention into the framework.

Fashion brand marketing

Yamaguchi et al. (2014) study the effects of visual, textual, and social factors on the popularity in a large real-world network focused on fashion. There are several studies that focus on understanding the growing interest in social media marketing (Dubois and Duquesne 1993; Kim and Ko 2012; 2010). Manikonda et al. (2015) study the influence of social media in various behavior of fashion brand marketing. They also analyze fashion brands' audience retention and social engagement.

Color in affective marketing: There are several research works that have studied the role of color in affective marketing. Most of these works focus on various kinds of advertisements, for example, the research on role of specific colors used in magazine ads (Lee and Barnes 1989; Schindler 1986), the efficiency of color ads compared to black and white ads (Meyers-Levy and Peracchio 1995; Sparkman Jr and Austin 1980). There are also studies on understanding the effects of colors on consumer responses (Bellizzi, Crowley, and Hasty 1983; Crowley 1993). This line of research suggests that red backgrounds elicit greater feelings of arousal than blue ones, whereas products presented against blue backgrounds are liked more than products presented against red ones (Bellizzi, Crowley, and Hasty 1983; Middlestadt 1990). Gray et al. (2014) study the relationship between color coordination and 'fashionableness'. They observe that maximum fashionableness is attained by selecting a color combination that is neither completely uniform, nor completely different, i.e., fashionable outfits are those that are moderately matched, not those that are ultra-matched ("matchy-matchy") or zero-matched ("clashing"). This balance of extremes supports Goldilocks principle regarding aesthetic preferences that seeks to balance of simplicity and complexity.

Studies on Pinterest

There have been several works on Pinterest.
Gender Roles: Gilbert et al. (2013) perform analysis focusing on the influence of gender, geography, language usage on Pinterest. They identify features of pins that could predict the activity of the board. They observe that being female on the site leads to more repins while having fewer followers. Our work draws motivation from this work where we study various social, temporal and image content factors as driving factors of popularity. Ottoni et al. (2013) study differences in gender role in platform usage and social interaction on Pinterest. They observe that females invest more effort in reciprocating social links, are more active and generalist in content generation whereas male are more likely to be specialists and tend to describe themselves in an assertive way. Also men and women possess different interests. Chang et al. (2014) study users' topical specialization and homophily.
User interaction and experience: Zarro et al. (2013) investigate professional and personal uses of Pinterest with interview data and observations of online activity. Zhong et al. (2014) perform analysis on how borrowing behavior facilitates social interactivity and experience. Yang et al. (2015) study recommendation of Pinterest boards for the Twitter users. Linder et al. (2014) investigate social and cognitive aspects of creativity that affect the digital curation practices of everyday ideation with Pinterest users. Miller et al. (2015) study perception on Pinterest of the users and non-users and show that there exist differences among these two groups and how exploring Pinterest changes the non-users' experience.
Diffusion, Popularity: Zhong et al. (2016) study the impact of social ties on Pinterest. Lo et al. (2016) study user activity and purchasing behavior on Pinterest for characterization of temporal user purchase intent. Lo et al. (2017) in another paper characterize the growth of Pinterest boards (size) and analyze how initial growth can be used to predict future growth behavior. Han et al. (2017) study popular and viral image diffusion in Pinterest. Deeb-Swihart et al. (2017) study selfie presentation in everyday life on Instagram. You et al. (2017) study various spatio-temporal patterns of Facebook photographs as well as its diffusion pattern via social ties.
The present work: Our study is different from the above ones in that we analyze the popularity aspects of (fashion) boards and attempt to understand its relationship with (i) social and temporal sharing/borrowing behavior and (ii) the gender, fashion and color terms embedded in the images posted on these boards. Our analysis sheds light on strategies and mechanisms an upcoming fashion brand could adopt to make itself popular on Pinterest which can eventually enhance their overall business.

Pinterest terminologies and the dataset

Entities on Pinterest

- **Pin**: A pin is an image (a visual bookmark) which forms the basic building block of Pinterest. The user who posts a pin is known as the 'pinner' and the activity is called 'pinning'. Pins can be liked and shared. Each of these pins has unique pin-id, description, number of likes, number of comments, number of repins, board name, source, and content of the comments. Sharing an already existing pin is referred to as 'repinning'.
- **Board**: A board is a user-generated collection where one saves pins. Boards can be made in secrecy or publicly. One can add collaborators to boards. Each board has a url, a name, a description (optional) and a category (optional, e.g., Art, Celebrities, Food and Drink, Entertainment, Fashion etc.).

Dataset

The dataset used in this study contains information about 0.3 million boards and their 63 million pins. We use Pinterest API v1 to crawl information about boards and pins. Board information constitutes of the following: board description, number of followers, and the creator. Pin information has the following attributes: pin description, number of likes, number of comments, number of repins, board name, and the creator. The data collection process is divided into two parts as follows

Crawling of the massive dataset We crawl a large dataset which should be useful for doing various analysis of the fashion boards.

- *Initial pin collection*: We initiate the data collection process by obtaining the pin-id of 1000 pins from http://www.pinterest.com/popular/ by generating automatic scrolls. Now, each pin-id of these pins are picked and the trailing 6 digits were permuted to generate new pin-ids. About 10 million new pin-ids are generated by this process. Information of all these pins are crawled separately.
- *Massive information collection*: We extract the board-url of each of these pins from the information crawled above. About 0.3 million unique board urls are obtained. Now, for all the board-urls, board information and their individual pin's information are crawled. This results in 59 million unique pins out of a total of 63 million pins. This massive dataset is used to find out the origin of the pins.

Fashion boards dataset We extract names of fashion boards from the following sources: i) Fashion categories on

Pinterest[3] and ii) Expert rankings from Ranker[4], Mashable[5] and Stylecaster[6]. Finally, we could obtain information about $\sim$ 3600 fashion boards. We discard those boards that have very less number of pins. A total of 4 million pins are found on these 3600 boards. For each of these board urls, we crawl the detailed information about the board and its pins from Feb, 2016 till March, 2016.

We then further categorize the 3600 boards above into following two categories of popularity: popular boards and unpopular boards. In popular boards we define two popularity classes - Highly Followed (socially ranked) Boards (HFB) and Expert Ranked Boards (ERB). We denote the unpopular ones as the Less Followed Boards (LFB). From the collection of the fashion boards, we assume the top 20% most followed boards as HFB and the bottom 20% as LFB. We have tried to use other percentage values also but choosing 20% allowed us to have a sizeable data for conducting meaning experiments. The 1200 boards which we obtain from the expert rankings are noted as expert ranked boards (ERB).

Characterization of fashion boards

In this section, we shall discuss the various factors which characterize the popularity of fashion boards. There are several factors that governs the popularity of a board - the originality/novelty, sharing/borrowing behavior as well as the content on the board (the image-characteristics of the pins).

Originality

Originality/novelty of boards is an important aspect. If one observe the creation of pins over the years (see figure 1), one can conclude that the total no. of pins are continuously growing whereas the no. of unique pins have increased only in the early few years but then started decreasing. This indicates that originality in this social media is on a decline over time due to heavy content sharing. Motivated by figure 1, we study board originality as an indicator of popularity.

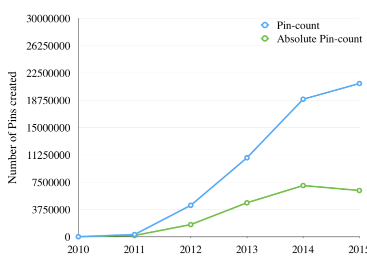

Figure 1: Evolution of unique (i.e., absolute) and total number of pins created per year.

Pin originality: Pins on a board can be classified into two types: *original* pins and *duplicate* pins. If a pin has originated from the board b, then it is called an original pin with respect to the board b whereas, if a pin has not originated from the board b, but is a result of a copy from another board to b, then it is called a duplicate pin with respect to b.

Board originality score: Using the concept of pin originality, we define a measure to compute the originality score of a board. Originality score ($orig_{score}$) of a board (b) can be defined as the ratio of the original pins (o_b) on it to the total number of pins (t_b) on it.

$$orig_{score}(b) = \frac{o_b}{t_b}$$

Originality score of a board lies in interval $[0, 1]$. Boards having originality score close to 1 constitute of mainly original pins, which means that they are content generators. Boards having originality score close to 0 constitute of mainly duplicate pins, which means that they are content copiers/consumers.

In figure 2(a), we observe that the originality scores are highly correlated with follower count. We then group the boards in less followed, highly followed and expert ranked boards and measure the originality scores in these popularity buckets. We observe that originality scores of highly followed and expert ranked boards are high, whereas that of less followed boards are on the lower side. Thus, originality of a board is an important indicator of its popularity.

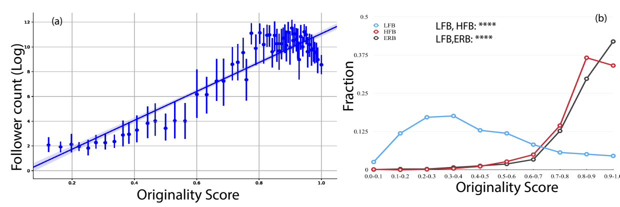

Figure 2: (a) Relationship between originality scores and follower counts of boards. (b) Distribution of originality scores across the less followed, highly followed and expert ranked boards. The K-S test for significance among the relevant distributions are measured. ****,***,**,*, ns denote p-values of significance to be < 0.0001, < 0.001, < 0.01, < 0.05 and non-significant respectively. We have used the same notations for the subsequent figures wherever applicable.

Originality of the top fashion brands We further study the originality scores of the boards corresponding to the top fashion brands. Toward this objective, we consider the top fashion clothing brands[7] and attempt to compute their originality scores. We separately collect the board information and the pins of these top fashion clothing brands. We compute the originality score of these boards and observe that they have highly original content (see table 1).

We then attempt to find the extent of correlation between the originality scores of these boards with their popularity (in terms of the number of followers). The Spearman's rank correlation comes out to be 0.41. This establishes that there

[3]https://www.pinterest.com/categories/

[4]http://www.ranker.com/list/world_s-top-fashion-brands/business-and-company-info

[5]http://mashable.com/2012/08/06/top-fashion-pinterest-accounts

[6]http://stylecaster.com/fashion-pinterest-accounts

[7]http://www.businessinsider.in/The-top-15-clothing-brands-millennials-love/14-Under-Armour/slideshow/51080592.cms

Table 1: Originality scores of top fashion brands

Rank	Brand Name	Originality Score
1	Nike	0.893
2	Target	0.996
3	Adidas	1.0
4	Macy's	0.985
5	JCPenney	0.993
6	Converse	0.877
7	Van's	0.99
8	Ralph Lauren	0.997
9	Forever 21	0.954
10	Victoria's Secret	0.989
11	Levi's	0.945
12	Chanel	0.876
13	Under Armour	0.898
14	Aeropostale	0.916

is a strong positive correlation between originality and popularity of the top fashion brands.

Sharing/borrowing behavior

Sharing/borrowing of pins are very common on the Pinterest platform. We introduce *board retention coefficients* and *board production coefficients* based on the sharing/borrowing behavior dynamics of the pins on a board. On Pinterest, the 'social behavior' of a pin can be measured based on three factors: the *number of likes*, the *number of repins* and the *number of comments* generated by the pin. We however observe that commenting is not practiced extensively in this platform. Hence, we only take number of $likes$ and $repins$ generated by a pin. The two coefficients we define next roughly correspond to the direction and magnitude of flow of information from one board to another. Each of them is independently able to portray meaningful information about sharing.

Retention coefficients *Board retention coefficients* are a novel set of measures concerning the $like/repin$ 'retention' capabilities of a board. It addresses the question that - *how many likes/repins shall a board be able to retain if other boards copy content from it*. We calculate the *likes retention coefficient* using the algorithm 1. Similarly, we also compute the *repins retention coefficient*.

Algorithm 1 Calculation of *likes retention coefficient*.

temp ← [ ]
for each original pin p on board b **do**
$temp.append(\frac{1 + likes\ of\ p\ on\ b}{1 + avg.\ likes\ of\ p\ on\ other\ boards})$
end for
likes retention coefficient of board b = average(temp)

Both the retention coefficient values lie in the interval $(1, \infty)$. If a board b has a higher likes (repins) retention coefficient, then the subsequent boards that copy pins from this board shall be able to garner less likes (repins) than the board b.

A significant fraction of highly followed and expert ranked boards have higher retention coefficients compared to the less followed boards (see figure 3(a) and (b)). Thus, the original pins on highly followed and expert ranked boards are significantly more liked/repinned among their respective duplicate (shared) pins, whereas the original pins on less followed boards are not popular among their duplicate pins. Hence, the likes/repins of the content on highly followed and expert ranked boards do not decline even after they are duplicated through copying.

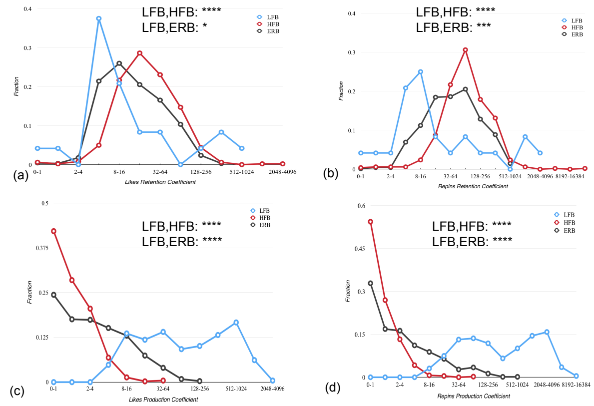

Figure 3: Distribution of a) likes retention coefficient b) repins retention coefficient (c) likes production coefficient b) repins production coefficient for less followed, highly followed and expert ranked boards.

Production coefficients *Board production coefficients* are a novel set of measures that capture the like/repin production capacities of a board. It addresses the question that - *how many likes/repins shall other boards gain if they copy content from a board*. We compute the *likes production coefficient* using the algorithm 2. Similarly, we also compute *repins production coefficient*.

Algorithm 2 Calculation of *likes production coefficient*.

temp ← []
for each duplicate pin p on board b **do**
$temp.append(\frac{1 + likes\ of\ p\ on\ its\ original\ board}{1 + likes\ of\ p\ on\ b})$
end for
likes production coefficient of board b = average(temp)

Both the production coefficients lie in the interval $(1, \infty)$. If a board b has a lower likes (repins) production coefficient, then the duplicate pins on b shall garner more likes (repins) compared to that on their board of origin.

A vast majority of highly followed and expert ranked boards have lower production coefficients than the less followed boards (see figure 3 (c) and (d)). Thus, the duplicate pins on highly followed and expert ranked boards generate more number of likes/repins compared to that generated on their corresponding boards of origin. On the other hand, duplicate pins on less followed boards generate less likes/repins compared to that in their corresponding boards of origin. Hence, highly followed boards and expert ranked boards are able to make an existing pin more liked/repinned by copying it.

Temporal dynamics of sharing/borrowing

In this section, we introduce two measures based on the temporal aspects of sharing: *inter-copying time* and *duration of sharing*. In addition, we also define *speed coefficients* that indicate the speed at which likes/repins are gained. *Inter-copying time* is a measure defined for the original pins. This is expressed as the average time-gaps between instances of sharing of an original pin on the subsequent boards. For an original pin p, we compute *ICT* as explained in algorithm 3. We now average the value of *ICT*s for all the original pins on board b, and call this as *inter-copying time* of board b.

A significant number of pins belonging to highly followed and expert ranked boards have a higher value of inter-copying time than pins on less followed boards (see figure 4(a) and (b)). This shows that pins on less followed boards have smaller time-gaps between consecutive shares as compared to pins on highly followed and expert ranked boards.

Algorithm 3 Calculation of *inter-copying time* for an original pin p.

$temp \leftarrow []$
for each duplicate pin $p^{'}$ generated from pin p **do**
  $temp$.append(time-stamp of $p^{'}$)
end for
sort $temp$ in non-decreasing order
for each i in range(0, len($temp$)) **do**
  if i == 0 **then**
    $temp[i] \leftarrow 0$
  else
    $temp[i] \leftarrow temp[i]$ - $temp[i-1]$
  end if
end for
ICT for pin $p \leftarrow average(temp)$

Duration of sharing (DoS) Similar to *inter-copying time*, *duration of sharing* is also defined for original pins. It can be interpreted as the life-cycle of sharing of a pin. For an original pin p, we compute *DoS* as explained in algorithm 4. We now average the value of *DoS*s for all original pins on board b, and call this as *duration of sharing* of board b.

Algorithm 4 Calculation of *duration of sharing* for an original pin p.

$temp \leftarrow []$
for each duplicate pin $p^{'}$ generated from pin p **do**
  $temp$.append(time-stamp of $p^{'}$)
end for
sort $temp$ in non-decreasing order
DoS for pin $p \leftarrow temp[len(\text{temp}) - 1] - temp[0]$

A large fraction of pins on highly followed boards and expert ranked boards have high duration of sharing compared to the less followed boards (see figure 4(c) and (d)). Hence, a pin on highly followed and expert ranked boards is likely going to have a longer life span than a pin on the less followed boards.

Speed coefficients In this section, we attempt to combine the likes/repins on a board with its temporal characteristics. Toward this objective, we define likes and repins speed coefficients as follows.

we compute *likes speed coefficient* as explained in algorithm 5. We then average the value of *likes speed coefficient* for all the original pins on a board b, and call this as the *likes speed coefficient* of board b. We similarly calculate *repins speed coefficient*.

Algorithm 5 Calculation of *likes speed coefficient* for an original pin p.

$likes \leftarrow []$
for each duplicate pin $p^{'}$ generated from pin p **do**
  $likes$.append(number of likes on $p^{'}$)
end for
likes speed coefficient of $p \leftarrow \frac{sum(likes)}{DoS(p)}$

Speed coefficients are greater for highly followed and expert ranked boards compared to the less followed boards (see figure 4(e) and (f))). Thus, original pins on highly followed and expert ranked boards gain popularity much more quickly than the original pins on less followed boards.

Image-based content analysis

Pinterest being an image sharing social media, the characteristics of image (pins) also should have an impact on their popularity. In this section, we analyze the content characteristics of images (pins) on the various boards. Toward this objective, we use densecap (Johnson, Karpathy, and Fei-Fei 2016), an image captioning tool that extracts salient regions from an image and describes them in natural language (English). We perform tokenization, stemming, lemmatization and stop-words removal of these generated 'dense' captions to obtain key tokens/phrases demonstrating the image. We further group these tokens into three key types: *gender terms*, *fashion terms* and *color terms*. Gender terms which we analyze are *male* and *female*. We obtain an exhaustive listing of fashion terms from Myvocabulary[8]. A universal set of all color terms are available in Wikipedia[9]. We use the color ***(black, blue, white, brown, green, purple, red, yellow, grey, metal, wooden, pink and silver)*** and fashion terms ***(clothes, tshirt, skin, shirts, jacket, feathers, pillows, sunglasses, buttons, suit, curtains, skirt, leather, pants, trouser, striped, shorts, strap, jeans, pillow, necklace, umbrella, bag, shoe, dress)*** which we find both in the above respective listings and our dataset.

[8]https://myvocabulary.com/word-list/fashion-and-clothing-vocabulary/

[9]https://en.wikipedia.org/wiki/Lists_of_colors

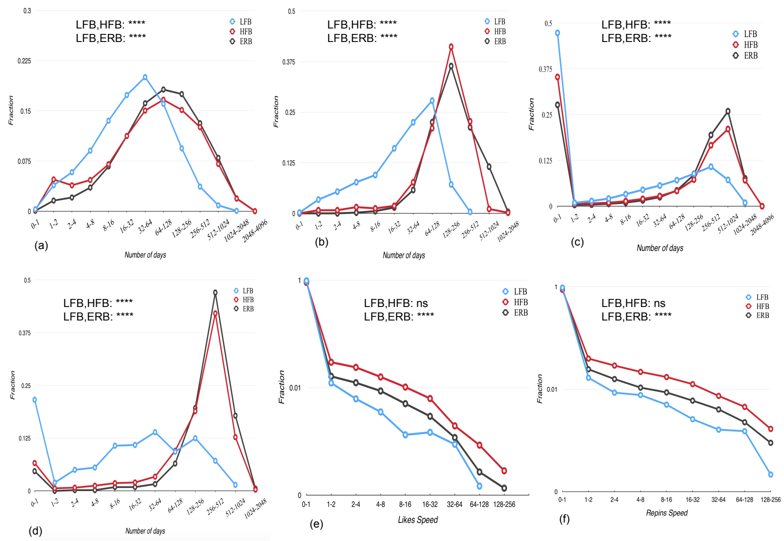

Figure 4: Distribution of a) inter-copying time for pins b) inter-copying time for boards c) duration of sharing for pins d) duration of sharing for boards e) likes speed coefficient f) repins speed coefficient for pins in less followed, highly followed and expert ranked boards.

Table 2: Fraction of pins on boards with various gender combinations.

Gender	LFB	HFB	ERB
Male only	0.45	0.48	0.49
Female only	0.49	**0.57**	**0.61**
Male-Female	0.26	**0.37**	**0.35**

Gender term analysis We analyze the occurrences of both the genders in the pins across all the three categories of boards. In table 2, we report the occurrences of each gender. The cell corresponding to *Male* and *Less Followed Boards* has a value of 0.45. This means that 0.45 fraction of pins belonging to less followed boards have male faces on them. We observe that the number of *female* faces on highly followed boards and expert ranked boards is high, whereas they are significantly lower ($\sim$ **20%** lower) in less followed boards. This shows that having more female faces on a board could increase the popularity of a board. Further, if a board has more female faces, it has more chances of being listed in expert ranked boards. Another very interesting observation is that the more popular boards have a higher fraction of pins containing both male and female faces together on a single pin compared to the less followed boards.

In summary, highly followed boards and expert ranked boards have more female faces than the less followed boards. Moreover, the boards in the former category have a richer collection of pins that together feature faces of the genders.

Fashion term analysis We analyze the occurrences of various fashion terms in the imagery of the pins across all the three categories of boards. We compute number of occurrences of the fashion terms appearing in the pins across the boards. We use the torso of this frequency distribution to extract the most discerning fashion terms. We choose the top 10 fashion terms from the torso to compute results in table 3. Each cell (x, y) in the table represents the fraction of pins belonging to a particular popularity class 'y' having the fashion term 'x'. Therefore, the cell value of 0.318 corresponding to *shirt* and *Less Followed Boards* means that 0.318 fraction of pins belonging to less followed boards have the fashion term *shirt* in them. We observe that 20% of fashion terms have equal distributions in highly followed boards and expert ranked boards whereas their distributions in the less followed boards are quite different.

Table 3: Fraction of pins having top 10 fashion terms which are obtained from the torso of the frequency distribution of all the fashion terms. The boldface values indicate similar distribution among ERB and HFB but a different distribution in LFB. We adopted this convention in the subsequent tables also.

Fashion Terms	LFB	HFB	ERB
shirt	0.318	0.389	0.330
bag	**0.221**	**0.291**	**0.312**
dress	0.221	0.179	0.225
pants	0.210	0.154	0.216
shoe	0.187	0.231	0.201
jacket	0.159	0.148	0.139
umbrella	0.158	0.139	0.146
necklace	0.147	0.134	0.156
pillow	0.143	0.131	0.153
jeans	**0.092**	**0.127**	**0.152**

We further study the co-occurrences of fashion terms in more detail. In table 4, we report the number of occurrences of two co-occurring fashion terms in a pin. Each cell $(x\text{-}z, y)$

Table 4: Fractions of pins having top 10 co-occurring bi-terms, which are obtained from the torso of the distribution of the fashion terms.

Bi-terms	LFB	HFB	ERB
jacket-trouser	**0.0964**	**0.1276**	**0.1198**
bag-umbrella	**0.0753**	**0.1243**	**0.1134**
bag-striped	**0.0683**	**0.1223**	**0.1143**
necklace-strap	0.0879	0.1124	0.0953
jeans-shoe	**0.0762**	**0.1057**	**0.1049**
bag-shorts	0.1032	0.1242	0.0643
bag-trouser	**0.0923**	**0.1243**	**0.1142**
leather-strap	**0.0923**	**0.1214**	**0.1203**
dress-umbrella	**0.0812**	**0.1143**	**0.1043**
dress-skirt	0.1023	0.0854	0.1053

Table 5: Fraction of pins having top 10 co-occurring tri-terms, which are obtained from the torso of the distribution of the fashion terms.

Tri-terms	LFB	HFB	ERB
leather-pillow-shirt	**0.1032**	**0.1343**	**0.1763**
pants-strap-trouser	**0.0913**	**0.1132**	**0.1298**
pants-shoe-skirt	**0.0613**	**0.1232**	**0.1265**
jeans-leather-pants	0.1265	0.0942	0.1135
jeans-shirt-shorts	0.1175	0.0823	0.0732
bag-necklace-skirt	0.0786	0.0974	0.1296
dress-shirt-sunglasses	0.0874	0.0925	0.1145
bag-pants-trouser	**0.0874**	**0.1134**	**0.1341**
bag-shirt-umbrella	**0.0112**	**0.1324**	**0.1142**
dress-pants-shorts	**0.0931**	**0.1121**	**0.1321**

in the table represents the fraction of pins belonging to a particular popularity class 'y' having the co-occurring fashion term 'x-z'. Thus, the value of 0.0964 corresponding to *jacket-trouser* and *Less Followed Boards* means that 0.0964 fraction of pins belonging to *Less Followed Boards* have both *jacket* and *trouser* together in them. We observe that **70%** co-occurring bi-terms have almost similar distributions in highly followed boards and expert ranked boards, whereas their distributions in the less followed boards are very different from the other two.

We also consider the co-occurring tri-terms (three fashion terms together). We observe similar discriminating results for **60%** tri-terms (see table 5). The discrimination becomes more prominent when we use the co-occurring bi-terms and tri-terms. We thus conclude that the collection of fashion terms together used in pins affect the popularity of their boards. Hence, a popularity seeking board can host images having a particular collection of fashion terms from the above analysis.

Color term analysis Colors are a very important factor in fashion (Bakhshi and Gilbert 2015). In this section, we analyze the occurrence of the color terms appearing in the pins across all three categories of boards. We compute the number of occurrences of each color term, co-occurring bi-terms generated from the colors. We choose the top 10 color terms (once again, from the torso of the frequency distribution) in table 6. We observe that 30% color terms have similar distributions in the highly followed boards and expert ranked boards, whereas their distributions in the less followed boards are different from the other two. White, black and blue are found to be the three mostly used color terms whereas purple is the least favored one.

Table 6: Fractions of pins having top 10 color terms which are obtained from the torso of the frequency distribution of all colors terms.

Color	LFB	HFB	ERB
white	0.638	0.664	0.696
black	0.555	0.581	0.538
blue	0.543	0.566	0.492
brown	**0.497**	**0.379**	**0.389**
red	**0.346**	**0.244**	**0.262**
wooden	**0.338**	**0.218**	**0.236**
green	0.224	0.221	0.233
metal	0.256	0.205	0.198
pink	0.122	0.098	0.086
purple	0.012	0.008	0.012

Table 7: Fractions of pins having top 10 co-occurring bi-terms, which are obtained from the torso of the distribution of all possible color bi-terms.

Color bi-terms	LFB	HFB	ERB
black-yellow	0.1043	0.0745	0.0943
blue-pink	**0.0744**	**0.0935**	**0.1064**
black-pink	**0.0824**	**0.1053**	**0.1034**
metal-red	0.0743	0.0723	0.1053
blue-yellow	0.1024	0.0923	0.0814
blue-silver	0.0908	0.0956	0.0932
pink-red	**0.0824**	**0.1025**	**0.1057**
metal-silver	0.0823	0.0675	0.0723
red-yellow	**0.0923**	**0.0814**	**0.0774**
grey-white	0.0452	0.0423	0.0424

In table 7, we show the occurrence distribution of the top 10 most co-occurring color bi-terms from the torso of the frequency distribution. Once again, we obtain a better discrimination while using bi-terms over single term occurrences (**40%** over **30%**). Black-yellow and blue-yellow are the most co-occurring color terms for the less followed boards whereas black-pink and pink-red are the most dominating color terms occurring together for the highly followed boards in the torso region. For the expert ranked boards, blue-pink and pink-red are the most used color terms whereas blue-white, black-white, blue-black are the most dominant color combinations in all the three categories of boards when we consider the whole distribution. Therefore, we observe that the color composition of images (pins) affect the popularity of their boards. Hence, a popularity seek-

Table 8: Fractions of pins having top 10 gender-based fashion bi-terms.

Gender and Fashion Bigrams	LFB	HFB	ERB
man-bag-jeans	0.1375	0.1323	0.1250
man-dress-shorts	**0.1175**	**0.1442**	**0.1525**
woman-bag-jeans	**0.1275**	**0.1567**	**0.1489**
woman-bag-strap	**0.1325**	**0.1578**	**0.1652**
woman-shirt-striped	0.1450	0.1682	0.1575
man-bag-shoe	**0.1675**	**0.1324**	**0.1434**
woman-bag-shoe	0.1424	0.1550	0.1523
woman-shirt-skirt	**0.1375**	**0.1576**	**0.1503**
woman-necklace-pants	0.1324	0.1425	0.1232
man-shirts-shorts	0.1453	0.1486	0.1502

Table 9: Fractions of pins having top five gender-based co-occurring color bi-terms.

Gender-Color Trigrams	LFB	HFB	ERB
woman-metal-white	0.2853	0.2753	0.2895
woman-pink-white	**0.2514**	**0.2657**	**0.2644**
man-pink-white	0.2425	0.2400	0.2350
woman-black-metal	**0.2850**	**0.2675**	**0.2643**
woman-brown-green	**0.2325**	**0.2184**	**0.2135**

ing board can host images having a particular color composition from the above analysis.

Gender infused fashion analysis In this section, we shall study gender based usage of fashion and color terms across the three board categories. In table 8, we show the gender based usage of the fashion terms (bi-terms). Each cell (x, y) in the table represents the fraction of pins belonging to a particular popularity class 'y' having the gender ('g') based fashion bi-term 'a-b'. Here 'x' corresponds to 'g-a-b'. For example, the cell value corresponding to *man-bag-jeans* and *Less Followed Boards* means that 0.1375 fraction of pins belonging to less followed boards have male, bag and jeans in them. We observe that **50%** of combinations have almost equal distributions in the highly followed boards and expert ranked boards, whereas their distributions in the less followed boards are very different from the other two. Such differences in the combination of gender and fashion terms affect the popularity of boards. It is seen that some combinations increase the popularity of boards, whereas rest decrease it. Another interesting observation we obtain from this analysis is that all the 40% combinations which have *woman* in them correspond to more popular boards.

Gender infused color analysis In table 9, we show the distribution of top five gender-based color bi-terms among pins across the three board categories. We observe that ∼60% combinations have equal distributions in highly followed boards and expert ranked boards, whereas their distributions in less followed boards are different from the other two. Though black-metal is the dominant color combination for female in both less and highly followed boards, white-metal is the prominent color combination in expert ranked boards. In general, metal colors seem to go very well with women.

Prediction model

The previous section demonstrates how several factors serve as indicators of popularity of the fashion boards on Pinterest. In this section, we shall leverage these factors to predict the future popularity of fashion boards. The popularity of a board is governed by the number of followers it has. To prevent any from of data leakage we separately re-crawl the new follower counts of all the fashion boards in our dataset in the month of April, 2017. This follower count statistics therefore is at a distance of 12 months from the training data. [10]

For the prediction task, we shall use the following features each of which is motivated by the analysis in the previous section.

- Originality score;
- Likes retention coefficient;
- Repins retention coefficient;
- Likes production coefficient;
- Repins production coefficient;
- Total number of pins;
- Avg. no. of likes on pins;
- Avg. no. of repins of pins;
- Avg. no. of comments on pins;
- Inter-copying time;
- Duration of sharing;
- Likes speed coefficient;
- Repins speed coefficient;
- Gender counts (2 bins); Gender bi-term count;
- Fashion term count (10 bins); Fashion bi-term count (10 bins); Fashion tri-term count (10 bins);
- Color term count (10 bins); Color bi-term count (10 bins);
- Gender infused fashion bi-term count (10 bins); Gender infused fashion tri-term count (10 bins); Gender infused color count (5 bins).

Predicting the popularity class of the boards

We have seen that the factors we have discussed earlier highly discriminate the unpopular class (LFB) from the two popular classes (HFB and ERB). The factors, however, can only moderately discriminate one of the popular class (HFB) from the other (ERB). We consider equal number of data points for each of the classes and then perform a 10-fold cross validation for generating results. In table 10, we present the classification results for i) HFB vs LFB ii) ERB vs LFB and iii) HFB vs ERB. As evident from the table, we can discriminate both popular (HFB or ERB) from the unpopular class (LFB) very well with a very high accuracy (95.96% for HFB vs LFB and 93.95% for ERB vs LFB) and very high precision, recall and area under ROC curve. Note that we are only able to obtain a moderate accuracy (65.1%) in classifying the two popular classes since the boards belonging to these two classes have very similar characteristics.

[10] Note that we do not use temporal statistics between March 2016 and April 2017 for enhanced robustness of the model; the idea is to make efficient predictions using minimal set of features that can be easily obtainable at any static time point.

Table 10: Performance of various classifiers for classification of i) HFB vs LFB ii) ERB vs LFB ii) ERB vs HFB.

Categories	Classifiers	Accuracy	Precision	Recall	F-Score	ROC Area
HFB vs LFB	SVM	92.51%	0.935	0.925	0.925	0.928
	LR	94.91%	0.95	0.949	0.949	0.981
	RF	**95.96%**	**0.96**	**0.96**	**0.96**	**0.995**
ERB vs LFB	SVM	92.06%	0.921	0.921	0.921	0.92
	LR	93.27%	0.934	0.933	0.933	0.977
	RF	**93.95%**	**0.94**	**0.94**	**0.94**	**0.99**
ERB vs HFB	SVM	61.08%	0.62	0.611	0.603	0.611
	LR	**65.1%**	0.655	**0.651**	**0.649**	**0.684**
	RF	64.81%	**0.672**	0.648	0.636	0.625

Table 11: Regression results.

Method	ρ	Normalized RMSE
10-fold cross-validation	0.8659	0.146
Separate training/testing (4:1 ratio)	**0.8738**	**0.1427**

For the classification task, we have used Support Vector Machines (SVM), Logistic Regression (LR) and Random Forest (RF) classifiers implemented in the Weka Toolkit (Hall et al. 2009). We choose the three classifiers for their diversity since they are known to be able to solve a vast range of different types of classification problems. Each of these classifiers represent different schools of thoughts and have their own set of strengths and advantages[11]. All the classifiers yield similar performance results with Random Forest classifier performing the best.

Predicting the followership counts of the boards

To study the robustness of our prediction model, we further try to predict the actual popularity, i.e., the logarithmic values of the followership counts of the boards. Toward this objective, we use Support Vector Regression (SVR) due to non-linearity of the problem. We use sequential minimal optimization (SMO) algorithm for training the SVR. We perform both separate training and testing as well as 10-fold cross validation method. We consider Pearson VII function-based universal kernel (PUK) due to its flexibility and adaptability through adjusting kernel parameter. We set the cost parameter (C) as 1. For evaluating how good the prediction is, we use Pearson correlation coefficient (ρ), normalized root mean square error (RMSE). We achieve high correlation coefficient (0.8738) and low normalized root mean square error (0.1427) which establishes the fact that the features obtained are robust and discriminating in nature (see table 11[12]). Both cross-validation and separate training/testing produces very similar results.

Discriminative features: In order to determine the discriminative power of each feature, we use the $RELIEFF$ feature selection algorithm (Kononenko, Simec, and Robnik-Sikonja 1997) available in the Weka Toolkit. Table 12 shows the rank of the features in terms of their discriminating power for prediction. The rank order clearly indicates that for popularity prediction the sharing/borrowing features, the color terms and some of the fashion terms are important. Likes retention coefficient, repins retention coefficient are the top discriminative features followed by various color term based features. Therefore, color (sometimes in conjunction with fashion and gender term) seems to be one of the most important discriminator for popularity of fashion boards.

Table 12: Top predictive features and their ranks.

Rank	Features
1	LRC
2	RRC
3	bag-striped (fashion)
4	white (color)
5	black (color)
6	blue (color)
7	female
8	male-female
9	brown (color)
10	blue-pink (color)
11	pink (color)
12	black-pink (color)
13	woman-pink-white (gender-color)
14	red (color)
15	man-shirts-shorts (gender-fashion)

Discussions and conclusions

In this section we outline various insights and implications of the current work. We also discuss the generalizability of the current work and finally draw the conclusions.

Insights and implications

Insights: The current study puts forward a lot of insights especially for new and upcoming fashion brands.

- Certain social sharing behavior of users can make boards popular. The most crucial among these are the retention coefficients. Popular boards are able to retain their proportion of 'likes' and 'repins' despite a lot of sharing and re-sharing of pins. Specially, engineered campaigns by the fashion houses can be made to ensure/promote such retentions.
- More female faces or both male and female faces together may be promoted by the fashion houses since that, as we have seen, could lead to enhanced popularity.
- Certain choices of colors (e.g., white, black, blue, pink etc.) and color combinations (e.g., blue-pink, black-pink etc.) may be more advertised to enhance the chances of being more popular.
- Certain fashion items like 'striped bags' seem to be very common in popular brands and could be more promoted by the newbies.
- For male fashion, 'shirts' and 'shorts' are the items that seem to propel popularity and can therefore be more vig-

[11] https://bit.ly/2LkuSf0

[12] We have also tried linear regression model which gives correlation coefficient of 0.7363 and 0.2 as normalized the RMSE value for 10-fold cross validation setting.

orously advertised by the new outlets. Many articles[13], in fact, have noted that shorts like boxers and bathing suits that end above the knee enhance the sex appeal of men.
- For female fashion, colors like pink and white seem to be good indicators of popularity. In fact, pink has been the most favorite color of garments for women for a very long time[14]. These therefore can be items of more focused publicity by the upcoming fashion agencies.

Implications: Our findings make several contributions to existing research. We believe, this research opens new pathway to understand new factors like colors, faces, fashion terms which are influential for understanding popularity. Our work also echoes some of the previous findings on impact of color on diffusion. We also suggest color combinations that makes a board popular. For newbie fashion houses and fashion trend-setters, our findings shed light on how images can be constructed so that they become popular. Pins of certain colors, more female faces or male-female joint faces could be some of the prime suggestions. One could also launch campaigns to promote their boards in such a way that the originating boards are able to retain the 'likes' and 'repins' of their pins in the face of constant sharing of these pins. In fact, Pinterest can make such 'tips-n-tricks' application available in exchange of a small amount of subscription from every newbie. This could potentially be a premium/paid support and could be a business model for the company for possibility of enhanced revenues.

There are several mobile apps which provide users with photo-editing tools. One of the widely used techniques in photo-editing is applying filters to them. These filters can change saturation, brightness, and color distribution of the image. Our findings can be used to design new filters for photo editing. Filters that increase saturation or enhance the warmness of the image will likely increase engagement with the photo.

Generalizability

Though the entire study has been performed on Pinterest, the findings can be generalized in other similar websites focused on images, for example, professional photography site like Flickr, or people-focused website like Instagram. Instagram is also a quite popular website for fashion trend. We believe these findings in the form of importance of color combinations and fashion terms influencing popularity can be generalized to Instagram, Flickr and Tumblr as well, though the popularity figures might vary which is mostly dependent on the website's underlying usage among communities, ranking algorithms etc.

Conclusions

In summary, we study various aspects of fashion boards on Pinterest. Our proposed measures – retention coefficients, production coefficients, inter-copying time and duration of sharing portray the sharing dynamics evident in less followed, highly followed and expert ranked fashion boards.

We observe that generally highly followed and expert ranked fashion boards are able to make an existing non-popular pin popular, whereas less popular fashion boards do not exhibit this characteristic. Further, if a pin has originated from highly followed or expert ranked fashion boards, it would achieve high popularity on this board than the subsequent boards on which it would be shared in future. We also observe that the pins on the highly followed and expert ranked fashion boards keep getting shared for a long time, whereas this happens for a short time for the pins on less followed fashion boards.

Gender, fashion and color terms embedded in images also yield interesting and conclusive results. We observe that both highly followed and expert ranked fashion boards exhibit similar trend in the usage of fashion bi- and tri-terms. We also observe that a large number of pins having female faces are present in highly followed and expert ranked fashion boards; the number of female faces is 20% lower for the less followed boards. Similar trend is observed for pins having both male and female faces. We also study occurrences of gender-based fashion and color terms. We identify combinations which give good discriminatory results across the three board categories. We try to leverage various sharing/borrowing characteristics, image-based content characteristics of fashion boards to predict their future popularity (logarithm of follower count). We achieve a high correlation coefficient of 0.874 and low RMSE.

Limitations: We acknowledge that there is some limitation of the current study. We specifically note the fact that some of the features and outcomes might be influenced by the particulars of the Pinterest ranking algorithms (e.g., what gets featured on the homepage, how personalization affects the probability a pin will be surfaced, etc.). There is no way we can control the internal algorithm promoting pins and boards. However, we believe, the factors we come up with are indeed influential as they strongly correlate with popularity studied on a large-scale data.

[13]https://www.cosmopolitan.com/sex-love/a63297/things-hot-guys-wear/

[14]https://www.racked.com/2015/3/20/8260341/pink-color-history

References

Bakhshi, S., and Gilbert, E. 2015. Red, purple and pink: The colors of diffusion on pinterest. *PloS one* 10(2):e0117148.

Bakhshi, S.; Shamma, D. A.; and Gilbert, E. 2014. Faces engage us: Photos with faces attract more likes and comments on instagram. In *CHI '14*, 965–974. ACM.

Bellizzi, J. A.; Crowley, A. E.; and Hasty, R. W. 1983. The effects of color in store design. *Journal of retailing*.

Berger, J., and Milkman, K. L. 2012. What makes online content viral? *Journal of marketing research* 49(2):192–205.

Burke, M.; Marlow, C.; and Lento, T. 2009. Feed me: Motivating newcomer contribution in social network sites. In *CHI '09*, 945–954.

Chang, S.; Kumar, V.; Gilbert, E.; and Terveen, L. G. 2014. Specialization, homophily, and gender in a social curation site: Findings from pinterest. CSCW '14.

Crowley, A. E. 1993. The two-dimensional impact of color on shopping. *Marketing letters* 4(1):59–69.

Deeb-Swihart, J.; Polack, C.; Gilbert, E.; and Essa, I. 2017. Selfie-presentation in everyday life: A large-scale characterization of selfie contexts on instagram. In *ICWSM*, 42–51.

Dubois, B., and Duquesne, P. 1993. The market for luxury goods: Income versus culture. *European Journal of Marketing* 27(1):35–44.

Forte, M. Z. C. H. A.; Hall, C.; and Zarro, M. 2013. Wedding dresses and wanted criminals: Pinterest. com as an infrastructure for repository building. ICWSM '13, 650–658.

Gilbert, E.; Bakhshi, S.; Chang, S.; and Terveen, L. 2013. "i need to try this"?: A statistical overview of pinterest. CHI '13, 2427–2436.

Gray, K.; Schmitt, P.; Strohminger, N.; and Kassam, K. S. 2014. The science of style: In fashion, colors should match only moderately. *PloS one* 9(7):e102772.

Hall, M.; Frank, E.; Holmes, G.; Pfahringer, B.; Reutemann, P.; and Witten, I. H. 2009. The weka data mining software: An update. *SIGKDD Explor. Newsl.* 11(1):10–18.

Han, J.; Choi, D.; Joo, J.; and Chuah, C.-N. 2017. Predicting popular and viral image cascades in pinterest. In *ICWSM*, 82–91.

Hessel, J.; Lee, L.; and Mimno, D. 2017. Cats and captions vs. creators and the clock: Comparing multimodal content to context in predicting relative popularity. In *WWW '17*, 927–936.

Hochman, N., and Schwartz, R. 2012. Visualizing instagram: Tracing cultural visual rhythms. In *SocMedVis in conjunction with ICWSM '12*, 6–9.

Johnson, J.; Karpathy, A.; and Fei-Fei, L. 2016. Densecap: Fully convolutional localization networks for dense captioning. CVPR '16.

Kim, A. J., and Ko, E. 2010. Impacts of luxury fashion brand's social media marketing on customer relationship and purchase intention. *Journal of Global Fashion Marketing* 1(3):164–171.

Kim, A. J., and Ko, E. 2012. Do social media marketing activities enhance customer equity? an empirical study of luxury fashion brand. *Journal of Business Research* 65(10):1480–1486.

Kononenko, I.; Simec, E.; and Robnik-Sikonja, M. 1997. Overcoming the myopia of inductive learning algorithms with relieff. *Applied Intelligence* 7:39–55.

Lee, S., and Barnes, J. H. 1989. Using color preferences in magazine advertising. *Journal of Advertising Research*.

Lee, C. S., and Ma, L. 2012. News sharing in social media: The effect of gratifications and prior experience. *Computers in Human Behavior* 28(2):331 – 339.

Lee, D. J.-L.; Han, J.; Chambourova, D.; and Kumar, R. 2017. Identifying fashion accounts in social networks.

Linder, R.; Snodgrass, C.; and Kerne, A. 2014. Everyday ideation: All of my ideas are on pinterest. CHI '14, 2411–2420.

Lo, C.; Cheng, J.; and Leskovec, J. 2017. Understanding online collection growth over time: A case study of pinterest. WWW '17 Companion, 545–554.

Lo, C.; Frankowski, D.; and Leskovec, J. 2016. Understanding behaviors that lead to purchasing: A case study of pinterest. KDD '16, 531–540.

Maity, S. K.; Kharb, A.; and Mukherjee, A. 2017. Language use matters: Analysis of the linguistic structure of question texts can characterize answerability in quora. In *ICWSM*.

Maity, S. K.; Kharb, A.; and Mukherjee, A. 2018. Analyzing the linguistic structure of question texts to characterize answerability in quora. *IEEE Transactions on Computational Social Systems* (99):1–13.

Manikonda, L.; Venkatesan, R.; Kambhampati, S.; and Li, B. 2015. Trending chic: Analyzing the influence of social media on fashion brands. *arXiv preprint arXiv:1512.01174*.

Meyers-Levy, J., and Peracchio, L. A. 1995. How the use of color in advertising affects attitudes: The influence of processing motivation and cognitive demands. *Journal of Consumer Research* 22(2):121–38.

Middlestadt, S. E. 1990. The effect of background and ambient color on product attitudes and beliefs. *ACR North American Advances*.

Miller, H. J.; Chang, S.; and Terveen, L. G. 2015. "i love this site!" vs. "it's a little girly": Perceptions of and initial user experience with pinterest. CSCW '15, 1728–1740.

Ottoni, R.; Pesce, J. P.; Las Casas, D. B.; Franciscani Jr, G.; Meira Jr, W.; Kumaraguru, P.; and Almeida, V. 2013. Ladies first: Analyzing gender roles and behaviors in pinterest. ICWSM '13.

Sanchis-Ojeda, R.; Sibley, D.; and Massimi, P. 2016. Detection of fashion trends and seasonal cycles through the analysis of implicit and explicit client feedback. In *KDD Fashion Workshop*.

Schindler, P. S. 1986. Color and contrast in magazine advertising. *Psychology & Marketing* 3(2):69–78.

Sparkman Jr, R., and Austin, L. M. 1980. The effect on sales of color in newspaper advertisements. *Journal of Advertising* 9(4):39–42.

Wojnicki, A., and Godes, D. 2008. Word-of-mouth as self-enhancement. *HBS Marketing Research Paper*.

Wu, B.; Cheng, W.-H.; Zhang, Y.; Huang, Q.; Li, J.; and Mei, T. 2017. Sequential prediction of social media popularity with deep temporal context networks. In *IJCAI '17*.

Yamaguchi, K.; Berg, T. L.; and Ortiz, L. E. 2014. Chic or social: Visual popularity analysis in online fashion networks. MM '14, 773–776.

Yang, X.; Li, Y.; and Luo, J. 2015. Pinterest board recommendation for twitter users. MM '15, 963–966.

You, Q.; García-García, D.; Paluri, M.; Luo, J.; and Joo, J. 2017. Cultural diffusion and trends in facebook photographs. ICWSM '17.

Zhong, C.; Salehi, M.; Shah, S.; Cobzarenco, M.; Sastry, N.; and Cha, M. 2014. Social bootstrapping: How pinterest and last.fm social communities benefit by borrowing links from facebook. WWW '14, 305–314.

Zhong, C.; Kourtellis, N.; and Sastry, N. R. 2016. Pinning alone? a study of the role of social ties on pinterest. In *ICWSM*, 751–755.