

Trend Makers and Trend Spotters in a Mobile Application

Xiaolan Sha
EURECOM

France
xiaolan.sha@eurecom.fr

Daniele Quercia^{*}
Yahoo! Research Barcelona
Spain
dquercia@yahoo-inc.com

Matteo Dell’Amico
EURECOM
France
della@linux.it

Pietro Michiardi
EURECOM
France
pietro.michiardi@eurecom.fr

ABSTRACT

Media marketers and researchers have shown great interest in what becomes a trend within social media sites. Their interests have focused on analyzing the items that become trends, and done so in the context of Youtube, Twitter, and Foursquare. Here we move away from these three platforms and consider a new mobile social-networking application with which users share pictures of “cool” things they find in the real-world. Besides, we shift focus from items to people. Specifically, we focus on those who generate trends (trend makers) and those who spread them (trend spotters). We analyze the complete dataset of user interactions, and characterize trend makers (spotters) by activity, geographical, and demographic features. We find that there are key characteristics that distinguish them from typical users. Also, we provide statistical models that accurately identify who is a trend maker (spotter). These contributions not only expand current studies on trends in social media but also promise to inform the design of recommender systems, and new products.

ACM Classification Keywords

H.4.m Information Systems Applications: Miscellaneous

Author Keywords

Mobile, Social Marketing

INTRODUCTION

Marketers and researchers have shown great interest in being able to effectively identify trends [7, 2, 3, 19, 30]. That is because being able to identify them translates into effectively managing brands for marketers, and into explaining the general dynamics of opinion spreading for researchers.

^{*}This work was done while at the Computer Laboratory of the University of Cambridge.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CSCW’13, February 23–27, 2013, San Antonio, Texas, USA.

Copyright 2012 ACM 978-1-4503-1209-7/13/02...\$15.00.

As we shall see in related work, there have been studies not only on trends but also on the individuals who create trends. This includes those who generate trends (defined in this paper as trend makers) and those who spread them (trend spotters). However, from these studies, it is not yet clear whether trend makers and trend spotters can be characterized by specific features, and whether these features can be used to easily identify them. To fill this gap, we examine a complete dataset of user interactions in the *iCoolhunt*¹ mobile social networking application and make two main contributions:

- We approach the analysis of who creates trends by defining two distinct classes of individuals: trend spotters and trend makers. We characterize them by combining multiple characteristics including their activity, content, network and geographical features. We find that trend spotters and trend makers differ from typical users, in that, they are more active, show interest in a variety of items, and attract social connections. We then study what differentiates trend spotters from trend makers. We learn that successful trend spotters are adult early adopters who hold interests in very diverse items, while successful trend makers are individuals of any age who focus on specific types of items.
- Using linear regression, we predict the extent to which one is a trend spotter or trend maker. Then, with an existing machine learning algorithm (SVM) and with a logistic regression, we perform a binary classification of whether one is likely to be a trend spotter (trend maker) or not. While linear regression has produced informative results, SVM and logistic regression have returned accurate predictions.

We conclude by discussing the theoretical implications of this work in the literature of opinion spreading, and its practical implications in designing recommender systems, and new products.

RELATED WORK

Previous studies of trends in social media have tried to answer two main questions: 1) what trends are; and 2) who creates trends.

What Trends Are. One way of extracting insights (e.g.,

¹<http://www.icoolhunt.com>

predict large-scale events) from social media is to identify trends [3, 19, 30]. In the literature, there are two types of trends [7]. In a social media site, exogenous trends are those triggered by factors external to the site (e.g., reporting of a piece of news on TV). Endogenous trends are those created by internal factors (e.g., spreading of a piece of news within the site). Exogenous and endogenous trends can be distinguished based on statistical analyses. For example, the two types of trends are distinguishable based on simple temporal patterns in the video sharing site Youtube [7] and based on features that reflect content, interactions among users, and social network structure in Twitter [19].

Who Creates Trends. There are two different views on who creates trends. The first is that *trends are generated by influentials*, which are special kinds of individuals who: are able to spot trends early on [15, 12, 23]; are socially well connected [16]; are able to easily influence others [14]; are considered to be experts [25, 29]; or are celebrities [30]. The second view on who creates trends holds that *trends are generated by coincidences* and that anyone can be influential. As a result, what becomes popular in a network does not depend on the initiators and is thus an accidental process. Duncan Watts uses the terms “accidental influentials” as he considers social epidemics to be “mostly an accident of location and timing” [26], and ideas spread and ultimately become popular only if there is societal willingness to accept them. More recently, researchers have found that there are different classes of individuals who contribute to two parallel processes: early participants start contributing and thus create random seeding, and that contribution spreads then through low threshold individuals [1, 13]. This recent literature has focused on the two parallel processes, while we build upon that by focusing on the individuals who contribute to those processes.

Given what is known in the literature, it is still unclear whether those who generate trends by originally introducing them into social networks and those who spread trends are special individuals or not. To answer this question, we define trend makers (those who generate trends) and trend spotters (those who spread trends). We characterize them by their activity, content, network and geographical features, and quantitatively test whether they have special characteristics comparing to typical users. We show that these features can be used to predict who are trend spotters and trend makers. We do so by going beyond the widely-studied platform of Twitter: we examine a new mobile application, which is described next.

APPLICATION

The application under study is a social application with a mobile phone client in which users can share pictures of “cool items”. Users upload photos of items that they encounter in the real world and consider “cool” (Figure 1). Upon submission, each photo must be tagged with a specific category selected among the five predefined categories (technology, lifestyle, music, design and fashion), and must be textually described. If one enables geolocation, one’s photos are automatically geo-referenced with the locations in which they

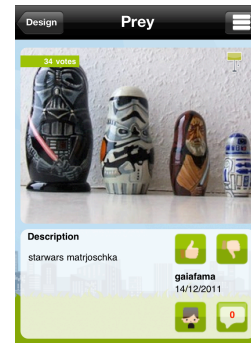


Figure 1. Screenshot of the mobile application.

were uploaded. Users can vote others’ photos with either a *like* or a *dislike* button. Every photo can receive limited-size comments from users, including the uploader. Every user is directly allowed to retrieve the list of popular pictures and latest uploads from any other user. In a way similar to what happens in Twitter, users of the application can follow each other. At the time of data collection, iCoolhunt didn’t provide any “news-feed” of pictures uploaded by followed users: activities from social connections were only accessible by browsing their profile pages.

We consider the *complete* dataset from February 2010 (the application’s launch) to August 2010.² Within those first six months, 9,316 users uploaded 6,395 photos and submitted 13,893 votes.

The unique characteristics of this dataset fit particularly well our interest in characterizing trend spotters and trend makers. The dataset contains user demographic information, the follower-followee graph, votes, comments, and geographical location of the place where items are uploaded.

To better interpret the results of our data analysis and compare them to the findings in the literature (which are mainly about Twitter and, to a lesser degree, Foursquare), we spell out the similarities and differences between the mobile application and Twitter/Foursquare:

Similarities. In a way similar to Twitter, the mobile application’s users can follow each other and, in a way similar to Foursquare, they can receive “honors” depending on how active they are (these honors are called “guru”, “observer”, “rookie”, and “spotter” based on the number of followers one has and sum of votes his uploads have received).

Differences. There are two main differences with Twitter. The first is about social interactions. Twitter users can *reply*, *retweet* others’ tweets, *mention* specific users, but cannot vote explicitly tweets (although similar information can be inferred from “favorites”). Instead, users of the mobile application can vote others’ items with a “like” or “dislike”, comment items, but can neither forward (“retweet”) items

²The iCoolhunt web application was launched after the end of our study period and, as such, our dataset includes only mobile application users. The format of this dataset cannot be acquired via crawling but directly from the service providers. We did acquire data only until the end of August 2010.

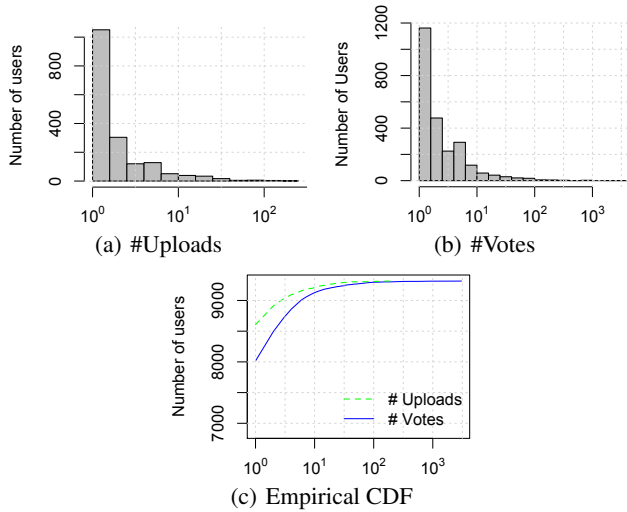


Figure 2. (a) Number of uploads per user; (b) Number of votes per user; (c) Distribution comparison on uploads and votes (log-scale on x-axis)

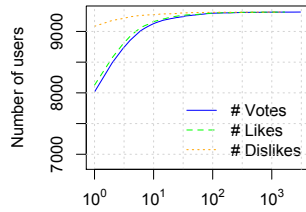


Figure 3. Empirical CDF of the number of votes, likes, and dislikes.

nor *mention* any other user. The second difference is about the user interface. Twitter users exchange status updates with each other, while the mobile application's users have no transparent way of being aware of what others are up to. We will discuss how these differences impact the effectiveness of our mobile application later on.

DATASET

The dataset includes 5,092 males and 4,224 females: 19% of males and 18% of females uploaded items and, out of a total of 13,893 votes, 69% of those were produced by males and 31% by females. This suggests that males are more active in voting than females. To understand how many users are actually using the application, we initially conduct a preliminary analysis of the behavior related to uploading, voting, and managing social relations.

Uploads and Votes

Uploading and voting pictures are the main activities in iCoolhunt, but, as one sees from the distributions in Figure 2(a) and Figure 2(b), only a small portion of the users are active in either uploading or voting. Out of the total 9,316 users, 1,761 (2,463) uploaded (voted) at least once and 710 (1,301) of them uploaded (voted) more than once. However, the minority who have uploaded more than once have contributed 83% pictures, while those who voted more than once have contributed 94% votes. That means that users prefer to vote pictures rather than to upload and, more importantly, that a minority of the users have contributed most of the content.

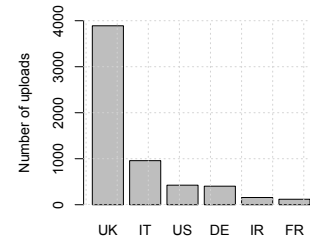


Figure 4. Number of uploads from each country (Top 6)

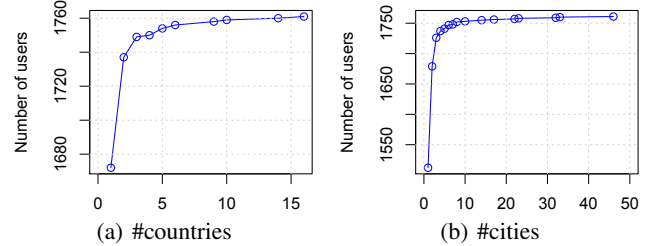


Figure 5. Empirical CDF of number of countries (cities) from where each user has uploaded, with y-axis representing the cumulative number of users.

Users are allowed to vote explicitly with a “like” or “dislike”. Thus, to understand users’ voting behavior, we separate votes into likes and dislikes and also consider any type of vote on aggregate. Figure 3 shows the distribution of the total number of votes per user and distinguishes “likes” from “dislikes”: 584 registered users have submitted “dislike” votes, 2,349 have submitted “like” votes, and 2,463 have simply voted with either a “like” or “dislike”. This suggests that iCoolhunt users are comfortable to express far more positive votes than negative ones.

Geography

Pictures are geographically tagged while they are uploaded. By tracing the locations of pictures, we are able to infer the number of places each user has been to. Using Google Maps, we are able to classify the coordinates into countries and cities/towns. The pictures have been uploaded from 57 different countries and regions. Among those there are only six countries with more than 100 uploads (Figure 4): United Kingdom (UK), Italy (IT), United States (US), Germany (DE), Ireland (IR) and France (FR). Also, each user could upload from multiple countries: among those who uploaded pictures, 89 users did so from more than one country (Figure 5(a)) and 249 users from at least two different cities/towns (Figure 5(b)).

Following

To cope with information overload, iCoolhunt users define lists of people they know or whose content they like. Each user then preferentially receives pictures from his/her own list of following, and eventually leaves comments and messages on those pictures. Figure 6 shows the number of followers/followees for each user - only few users follow other users, and even fewer users are followed. To then check whether users who upload more also have more followers/followees, we graph the scatter-plots of the number of

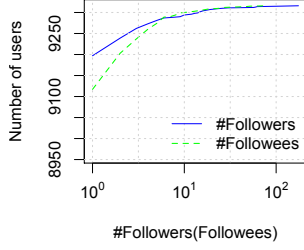


Figure 6. Empirical CDF of the number of followers and followees.

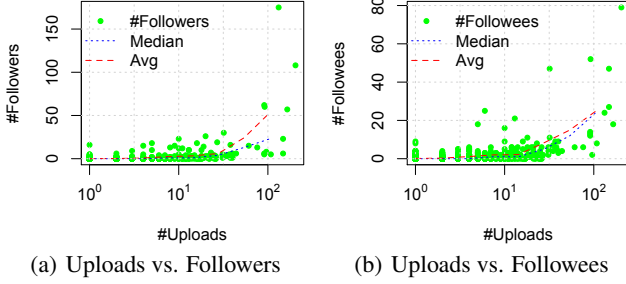


Figure 7. Number of followers(followees) and number of uploads per user

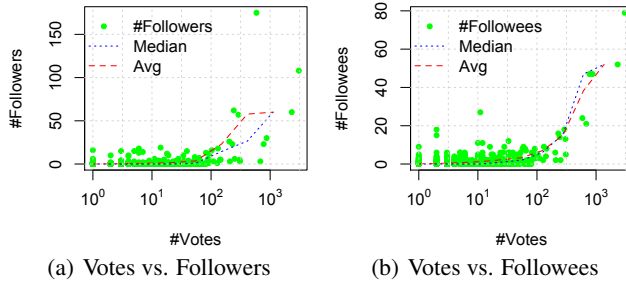


Figure 8. Number of followers(followees) and number of votes per user

followers/followees (y -axis) as a function of the number of uploads and votes for each user. In a way similar to [17], we bin the number of uploads/votes on a log scale and show both of the mean and median of each bin. The relationships are clear: the number of followers increases with the number of uploads (Figure 7(a)) and number of votes (Figure 8(a)); so does the number of followees (Figures 7(b) and 8(b)). In short, people who contribute to the community get followed, those who lurk do not. This makes sense as lurkers are essentially invisible.

To sum up, as one expects, a minority of users have uploaded and voted most of the pictures. Since we cannot get hold of access logs, we are not able to identify lurkers (those who simply browse) among inactive users. What we are able to differentiate though is trend spotters (users who spread trends) from trend makers (those who upload trends), and we do so next.

TREND MAKERS AND SPOTTERS

Our analysis unfolds three steps: identify trend spotters and trend makers; characterize them by conducting a quantitative analysis based on selected features; and build a statistical

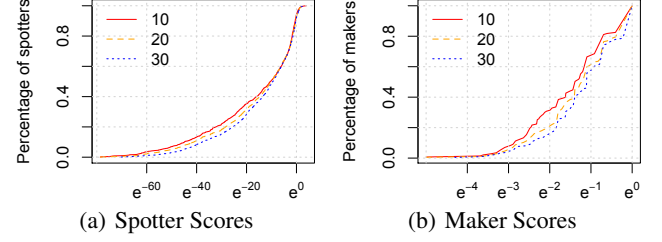


Figure 9. Empirical CDF of spotter (maker) scores (log-transformed) versus top- n ranked items.

model that identifies who is a trend spotter and who is a trend maker.

Defining Trend Spotters and Trend Makers

To identify trend spotters and trend makers, we need to define what a trend is first.

Trends. Trends differ from popular items, in that, they are not necessarily popular but they receive abrupt attention within a short period of time. To identify trends in the dataset, we define a “trend score” metric, which is derived from a simple burst detection algorithm proposed in [19]. At each time unit t (one-week window that incrementally slides every day), we assign to item i a $trendScore(i, t)$ that increases with the number of votes it receives:

$$trendScore(i, t) = \frac{|v_{i,t}| - \mu_i}{\sigma_i} \quad (1)$$

where $|v_{i,t}|$ is the number of votes item i has received within time unit t , μ_i is the mean number of votes it has received per time unit³, and σ_i is the corresponding standard deviation. The higher the item’s $trendScore$, the more votes it has received. For each time unit (each week), we sort items by their trend scores in descending order and select the top- n items to be trends. We have experimented with different $n \in \{10, 20, 30\}$ and found that spotter (maker) scores (defined later on) do not change very much (Figure 9) and, for $n = 10$, the resulting numbers of trend makers (140) and trend spotters (671) were sensible compared to the total number of users who voted (1,301) or uploaded (710) more than once. We thus report the results for $n = 10$.

Trend Spotters. Trend spotters are those who tend to vote items that, after a while, end up being trends. Not all trend spotters are equally good at voting trends. Considering a set of trending items, one’s ability of voting trends depends on three factors: *how many* trending items one has voted, *how early* one has voted them, and *how popular* the voted items turned out to be.

We incorporate these three factors in a $spotterScore$ for each user u by dividing the number of trends user u has voted

³The last time unit we consider is that in which the item received the last vote.

$(\sum_{i \in \mathcal{I}_u} g_{u,i})$ by u 's total number of votes (v_u):

$$\text{spotterScore}(u) = \frac{\sum_{i \in \mathcal{I}_u} g_{u,i}}{v_u} \quad (2)$$

In the numerator, $g_{u,i}$ is the gain user u acquires when voting on trending item i and incorporates the three factors of *how many*, *how early*, and *how popular*:

$$g_{u,i} = v_i \times \alpha^{-p_{u,i}} \quad (3)$$

where $\mathcal{I}_u$ is set of trends that u has voted ($\sum_{i \in \mathcal{I}_u}$ reflects the *how many*); v_i is total number of votes item i has received (which reflects the *how popular*), and α is a decay factor (which reflects the *how early*, $\alpha = 2$ in our experiments) whose exponent is the order with which u has voted item i ($p_{u,i}$ means that u is the p^{th} user who voted i). A trend spotter is then anyone with trend spotter score greater than zero.

Trend Makers. Trend makers are those who tend to (not simply vote but) *upload* trending items. So the trend maker score of user u increases with the number of trends u has uploaded. The numerator of the score is $\sum_{i \in \mathcal{I}_u} I(i \text{ is a trend})$, where $\mathcal{I}_u$ is the set of items u has uploaded, and I is the indicator function, which is 1, if the enclosed expression “ i is a trend” is true; 0, otherwise. This numerator is then normalized by u 's total number of uploads ($|\mathcal{I}_u|$) to account for those users who indiscriminately upload a large number of items without any quality control. A trend maker is then anyone with trend maker score greater than zero.

$$\text{makerScore}(u) = \frac{\sum_{i \in \mathcal{I}_u} I(i \text{ is a trend})}{|\mathcal{I}_u|} \quad (4)$$

Typical users. If an active user (i.e., one who uploaded or voted more than once) is not a trend spotter or a trend maker, then he/she is considered to be a typical user. We discover that in our application, there were 1,705 typical users.

Characterizing Trend Spotters and Trend Makers

To characterize trend spotters and trend makers, we need to conduct a quantitative analysis that considers four types of features: activity, content, network, and geographical features.

Activity Features. The first activity feature we consider reflects how long an individual has been actively using the application. We call this feature “lifetime”, and previous studies have identified it to be important as it conveniently identifies “early adopters” [10]. The literature recognizes that early adopters are a special interest group that heavily shapes usage of the application and ultimately determines the social norms within the application [8]. Once social norms are formed, changing them becomes very difficult and might backfire at times [9]. In addition to early adopters, we consider typical users as well. Their activity mainly consists of producing content (uploading pictures) and consuming it (voting pictures). Thus, we add two activity features to “lifetime”: how frequently a user has been

uploading pictures (*daily uploads*) and how many pictures the user has voted (*daily votes*).

Content Features. When users upload pictures, they are able to categorize them by selecting a proper category from the five predefined ones (technology, lifestyle, music, design and fashion). Previous studies on Twitter have linked category diversity to influence. According to [4] and [29], to become influential, one should “stay focused” – one tweet content in a specific category and become the “guru” in it. One may thus wonder whether our trend spotters and trend makers focus on specific categories of pictures, or, rather, whether they diversify consumption and production of content. To answer this question, we adopt a measure of categorical diversity from information theory called Shannon Index:

$$s = - \sum_{i \in \mathcal{C}} (f_i \ln f_i) \quad (5)$$

where $\mathcal{C}$ is the above set of five categories (technology, lifestyle, music, design and fashion) and f_i is the fraction of items (out of the total number of items) that belong to i^{th} category. Using this expression, we measure three types of diversities for each user – *upload diversity*, *vote diversity*, *consumption diversity* (consumption translates into either voting or commenting pictures).

Network Features. Users of our application follow each other in a way similar to what happens in Twitter. Thus, the simplest network measures we could consider are in-degree (*number of followers*) and out-degree (*number of followees*). To then account for local network properties, we also consider the *clustering coefficient* [27] of a user, which is computed from the undirected graph whose nodes are users, and edges link users between whom there is at least one following relation. Clustering reflects the extent to which one's network is densely connected.

Geographical Features. Information propagation faces geographical constraints, caused by the decrease of the probability of a social tie between a pair of individuals with the increase of the geographical distance between the pair [5, 21]. In our application, when users upload pictures of items, these pictures are automatically geo-referenced – they report the longitude-latitude pairs of the items' positions. Thus, we can compute how often and how far users physically move (*wandering*), and we do so using the radius of gyration [6]:

$$r_u = \sqrt{\frac{1}{n} \sum_{i \in \mathcal{I}_u} d_{l_i, c_u}^2} \quad (6)$$

where n is number of u 's uploaded items, $\mathcal{I}_u$ is the set of u 's uploaded items, c_u is user u 's center of mass (which is the “average point” of all geographical locations of u 's items), l_i is the location where item i has been uploaded, and d_{l_i, c_u} is the Euclidean distance between user u 's center of mass and the location of each item i .

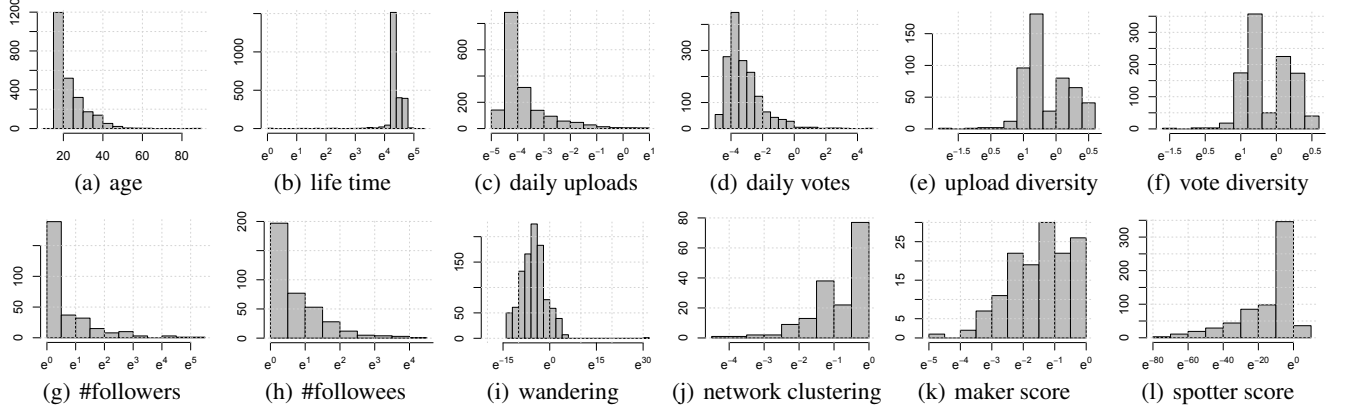


Figure 10. Distributions of features (a-j), trend maker (k) and trend spotter (l) scores with log-transformed values (except for the age feature). The x-axis represents the range of log-transformed features, and the y-axis represents the number of users.

	Content	Result
Spotters vs. Typical	H1.1 Trend spotters are more active than typical users.	✓
	H1.2 Trend spotters tend to be more specialized than typical users in certain category of items.	×
	H1.3 Trend spotters attract more followers than typical users.	✓
Makers vs. Typical	H2.1 Trend makers are more active than typical users.	✓
	H2.2 Trend makers are more specialized than typical users in certain category of items.	×
	H2.3 Trend makers attract more followers than typical users.	✓
Spotters vs. Makers	H3.1 Trend makers upload content more often than trend spotters.	✓
	H3.2 Trend makers vote less often than spotters.	✓
	H3.3 Trend spotters upload more diverse content than trend makers.	*
	H3.4 Trend spotters vote less diverse content than trend makers.	×
	H3.5 Trend makers have more followers than trend spotters.	✓

Table 1. Our Hypotheses (✓: accept hypothesis; ×: accept the alternative hypothesis; *: unknown)

Feature (log-transformed)	S > T	M > T	M > S (if not shown otherwise)
Daily Uploads	0.07 *	0.45 *	0.58 *
Daily Votes	0.66 *	0.18 *	0.57 * (M < S)
Upload Diversity	0.31 *	0.35 *	0.02 (M < S)
Vote Diversity	0.31 *	0.23 *	0.27 * (M < S)
#Followers	0.06 *	0.32 *	0.26 *

Table 2. Summary of Kolmogorov-Smirnov test results of our hypotheses. *D*-values with significance level < 0.05 are highlighted and come with *. M, S and T stand for trend makers, trend spotters and typical users. We test a pair of distributions at a time - e.g., for S > T, we test whether the daily upload distribution for trend spotters is greater than that of typical users, and report the corresponding *D*-value.

Since locations are not only associated with pictures but also with users, we also compute the geographic span of a user’s network [21]:

$$s_u = \frac{1}{m} \sum_{j \in \mathcal{F}_u} d_{u,j} \quad (7)$$

where m is number of u ’s followers, $\mathcal{F}_u$ is the list of u ’s followers, and $d_{u,j}$ is the distance between user u ’s center of mass and each follower j ’s center of mass.

We display the distribution of each feature in Figure 10. Since the distributions of the features are skewed, we show their log-transformed distributions.

Who trend spotters and trend makers are

Having all the features at hand, we are now able to run a comparative analysis. We compare trend spotters, trend makers,

and typical users by testing hypotheses drawn from the literature, which Table 1 collates for convenience. We will now explain these hypotheses one-by-one.

1. Trend spotters (makers) vs. Typical users. Previous studies have shown that, compared to typical Twitter users, influentials tend to be more active, more specialized in specific categories, and be more popular (i.e., attract more followers) [24, 25, 4, 29, 30]. To draw parallels between Twitter influentials and trend spotters (makers), we hypothesize that, compared to typical users, trend spotters (makers) are more active (H1.1 and H2.1 in Table 1), specialized (H1.2 and H2.2), and popular (H1.3 and H2.3). To test these hypotheses, we run Kolmogorov-Smirnov tests (K-S tests [22]) and t-tests, and since both return the same results, we reported only K-S tests in Table 2. The idea is that we consider a pair of distributions, say, those of “daily uploads” for trend spotters (S) and for typical users (T) and compare them - we compare whether the mean of the distribution of trend spotters is greater than that of typical users (i.e., we test $S > T$). We find that, compared to typical users, both trend spotters and trend makers are more active (they upload and vote more) and are more popular (attract more followers). These results are statistically significant, that is, the corresponding p -values are below 0.05. Hence the four hypotheses H1.1, H1.3, H2.1 and H2.3 are confirmed. By contrast, hypotheses H1.2 and H2.2 are not confirmed. When consuming and producing content, trend spotters and trend makers neither focus on specific con-

tent categories nor diversify themselves more than what typical users do. However, by separating what users vote and what they upload, we find that the items voted by trend spotters are more diverse than those uploaded. This preliminary difference between trend spotters and trend makers opens up the way for dwelling on the similarities and differences between these two types of users.

2. Trend spotters vs. Trend makers. Since no previous study has compared the characteristics of trend spotters and trend makers, we need to start with some initial hypotheses based on our intuition. So we initially consider that trend makers tend to upload items, while trend spotters tend to vote items. More specifically, we hypothesize that, compared to trend spotters, trend makers upload more content (H3.1), vote less (H3.2), upload less diverse content (H3.3), vote more diverse content (H3.4), and are more popular (H3.5). After running Kolmogorov-Smirnov tests (Table 2), we find that trend makers *upload* more frequently than trend spotters who, by contrast, *vote* more frequently. That confirms both H3.1 and H3.2. By then considering what users upload/vote, we find that trend makers “stay focused” (i.e., they upload and vote items in specific categories), while trend spotters vote items belonging to a variety of categories. So trend makers act in a way similar to the content contributors discussed in [20, 18] who tended to have special care in producing quality content. In a similar way, our trend spotters tend to upload items in the few categories they are more familiar with, while they vote on items of different categories, suggesting a wide spectrum of interests. Finally, trend makers tend to be more popular (are followed more) than trend spotters. To recap, trend spotters preferentially engage in voting and do so across a broad range of categories, trend makers engage uploading within a limited number of categories. Both of them are popular, but trend makers are followed more than trend spotters.

PREDICTING TREND MAKERS AND SPOTTERS

By considering four types of features, we have been able to find statistically significant similarities and differences among trend spotters, trend makers, and typical users. Now we study to which extent these features are potential predictors of whether users are trend spotters (makers), and do so in two steps:

1. We model trend spotter (maker) score as a linear combination of the features.
2. We predict trend spotter (maker) using a logistic regression and a machine learning model: Support Vector Machines (SVM).

Upon the set of 140 trend makers, 671 trend spotters and 1,705 typical users (identified in the previous section), we now run our predictions.

Regression Models

Before running the regression, we compute the (Pearson) correlation coefficients between each pair of predictors (Table 5). As one expects, we find that different types of activities are correlated (i.e., high positive correlation between the *number*

(a) Logistic Regression

Features	I(Score > 0)	
	Spotters	Makers
Age	2e-04	0.001
Life Time	0.006 *	0.001 *
Daily Votes (Daily Uploads)	0.007 *	0.16 *
Vote Diversity (Upload Diversity)	0.38 *	0.14 *
Wandering	-6e-15	-7e-15
#Followers	2e-05	0.009 *
Network Clustering	0.08	0.28 *

(b) Linear Regression

Features	log(Score)	
	Spotters	Makers
Age	0.36 *	0.01
Life Time	0.19 *	0.0001
Daily Votes (Daily Uploads)	0.16	-1.03 *
Vote Diversity (Upload Diversity)	7.28 *	-1.09 *
Wandering	-2.1e-13	-1.4e-15
#Followers	-0.06	0.01 *
Network Clustering	2.75	-0.64 *
R^2	0.15	0.65
Adjusted R^2	0.14	0.64

Table 3. Coefficients of the linear regression. A correlation coefficient within 2 standard errors is considered statistically significant. We highlight and mark them with *.

of followers, daily uploads, daily votes, and content diversity). Attracting followers is correlated more with uploading content (i.e., positive correlation between the *number of followers* and *daily uploads*) rather than voting content (i.e., no significant high correlation between the *number of followers* and *daily votes*).

Next, we perform both logistic and linear regressions on input of the following predictors that tend not to be strongly correlated with each other: *age*, *life time*, *daily votes*, *daily uploads*, *votes diversity*, *upload diversity*, *wandering*, *number of followers* and *network clustering*.

Regressions. We model trend spotter (maker) score as a combination of the features in two steps, as it is commonly done [11]. In the first, we use a logistic regression to model whether a user has trend spotter (maker) score greater than zero or not:

$$Pr(score_u > 0) = \text{logit}^{-1}(\alpha + \sum_{i \in \mathcal{V}} \beta_i \mathcal{U}_{u,i}). \quad (8)$$

In the second step, we take only those users with trend spotter (maker) scores greater than zero, and predict their scores with a linear regression of the form:

$$\log(score_u) = \alpha' + \sum_{i \in \mathcal{V}} \beta'_i \mathcal{U}_{u,i}, \quad (9)$$

In Equation 8 and 9, $\mathcal{V}$ is a set of predictors, and $\mathcal{U}_{u,i}$ refers to user u 's value of predictor i .

The results of the logistic regression (coefficients in Table 2(a)) suggest that trend spotters tend to be early-adopters

who vote often and are interested in diverse items, and trend makers tend to be early adopters who upload often and also upload items from different categories, moreover, they tend to attract followers and have a dense connected network.

Considering then only the users who have trend spotter (maker) score greater than zero, we focus on the features that can potentially predict how successful a trend spotter (maker) is. The results of the linear regression (β coefficients in Table 2(b)) suggest that successful trend spotters are adult early-adopters who vote items from various categories. By contrast, successful trend makers are users of any age who upload items belonging to specific categories (they “stay focused”) and tend to attract social followers from different communities.

The goodness of fit of a linear regression model is indicated by R^2 . In our case, the adjusted R^2 is very similar to R^2 , which is 0.15 for trend spotter score and 0.65 for trend maker. So one is able to explain 15% variability in trend spotter score and 65% in trend maker score. The difference in these two results might be explained by either: 1) the idea that trend spotters might well be “accidental influentials” [26] and, as such, trend spotters are harder to identify than trend makers; or 2) the fact that our predictors simply encapsulate complex phenomena and, as such, their explanatory power is limited. Next, we test whether trend makers and trend spotters can be predicted by a machine learning model that has shown good performance in similar learning settings – that is, by SVM.

Support Vector Machines (SVM)

We formulate the task of predicting trend spotters (makers) as a binary classification problem, where the response variable is whether a user’s trend spotter (maker) score is greater than or equal to zero. To our sample of 671 trend spotters and 140 trend makers, we add an equal number of typical users (those 1,705 users have been identified in the previous section). By construction, the resulting sample is balanced (the response variable is split 50-50), and interpreting the results becomes now easy as the accuracy of a random prediction model would be 50%.

We split randomly each set of samples into two subsets, 80% of them are used for training and 20% for testing. We apply SVM on the input of the same seven features previously used in the regressions to predict trend spotter and trend maker scores. We compare the results with those obtained by the previous logistic regression model. We compare prediction performance by ROC (Receiver Operating Characteristic) curve (Figure 11), AUC (area under the ROC curve), and accuracy (Table 4).

SVM and logistic regression show comparable performance (for both, $AUC = 0.77$; accuracy is 71.52% for the regression, and 71.85% for SVM). SVM only slightly outperforms the logistic regression in identifying trend makers. This suggests that one is able to effectively identify trend spotters and trend makers even with a simple logistic regression. Also, SVM might not have shown considerable prediction gain simply because of our (limited) dataset’s size.

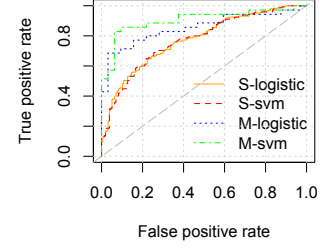


Figure 11. ROC curve of logistic regression and SVM model (S: trend spotters; M: trend makers).

	Spotters		Makers	
	AUC	Accuracy	AUC	Accuracy
Logistic	0.77	71.52%	0.85	82.09%
SVM	0.77	71.85%	0.90	88.06%

Table 4. AUC and best accuracy of each predictive model.

DISCUSSION

We have characterized trend spotters and trend makers based on four types of features (i.e., activity, content, network, and geographical features) and proposed a statistical model to accurately identify them. This work has both theoretical and practical implications.

Theoretical Implications

We show that trend spotters and trend makers are similar only to a certain extent. Compared to typical users, both of them: are more active in uploading/voting; attract more social connections; and upload/consume more diverse content. Yet, when they are compared not with typical users but with each other, differences emerge:

1. Trend spotters prefer voting more than uploading, and when they vote, they do so in very diverse categories. By contrast, trend makers act in a way similar to the content contributors in [18] who have special care in producing quality content and “stay focused” – they upload and vote items in very specific categories.
2. Successful trend spotters are early adopters who are attracted by diverse *items*, while successful trend makers attract diverse social *relations* (they tend to be followed by users from different social clusters).

These notable differences between trend spotters and trend makers would call for a rethink of current studies in opinion spreading.

In iCoolhunt, geographical features seem not to matter. That might suggest that, to be successful, trend spotters or trend makers do not necessarily need to move often or travel around. However, based on further analysis, we have learned that, while the application was originally designed to let users share items on the move, some users have started to assume unexpected behaviors – for example, some have started to post content (e.g., images from the Web) that was not explicitly related with the location from which it was uploaded. Given such behaviors, to make more grounded claims, a longitudinal analysis would be required.

	Age	Life Time	Daily Uploads	Daily Votes	Upload Diversity	Vote Diversity	Wandering	Follower Geo Span	#Followers	#Followees
Life Time	0.21									
Daily Uploads	0.02	-0.12								
Daily Votes	0.05	-0.09	0.47 *							
Upload Diversity	0.02	0.09	0.40 *	0.08						
Vote Diversity	0.04	0.08	0.22	0.08	0.42 *					
Wandering	0.004	0.13	0.16	0.11	0.06	0.05				
Follower Geo Span	0.05	0.12	0.16	0.10	0.12	0.11	0.23			
#Followers	0.03	0.23	0.37 *	0.14	0.22	0.16	0.44	0.16		
#Followees	0.05	0.17	0.52 *	0.31 *	0.29 *	0.22	0.56 *	0.21	0.64 *	
Network Clustering	0.03	0.13	0.22	0.04	0.24	0.23	-0.001	0.27 *	0.08	0.22
Spotter Score	0.07	0.18	0.03	0.01	0.05	0.10	0.04	0.07	0.13	0.11
Maker Score	0.07	0.10	0.06	0.01	0.07	0.06	0.02	0.12	0.12	0.09

Table 5. Pearson Correlation coefficients between each pair of predictors. Coefficients greater than ± 0.25 with statistical significant level < 0.05 are marked with a *.

Practical Implications

The ability of identifying trend spotters and trend makers has implications in designing recommender systems, marketing campaigns, new products, privacy tools, and user interfaces.

Recommender Systems. Every user has different interests and tastes and, as such, might well benefit from personalized suggestions of content. These suggestions are automatically produced by so-called “recommender systems”. Typically, these systems produce recommendations people might like by *equally* weighting all user ratings. Given that trend spotters are effective social filters, one could imagine to weight their ratings more than those from typical users to construct a new recommender system.

New Products. Some web services (*e.g.*, 99designs⁴) provide a platform to crowd source design work, where clients submit their requests and designers try to fulfill them. Since trend spotters and trend makers are “fashion leaders”, soliciting their early feedbacks might result into avoiding mistakes when designing new products. Often, at design stage, costs of correcting minor mistakes are negligible, while, at production stage, they become prohibitive.

User Interfaces. Trend spotters and trend makers do not connect to as many users as one would expect. That is likely because it is hard for iCoolhunt users to be aware of what others are up to. The user interface does not come with clear-cut “social features” that create a sense of connection and awareness among users as much as Facebook or Twitter sharing features do (as we have detailed in the Application section).

We are currently working on a recommender system that exploits the ability to distinguish between trend makers and trend spotters to suggest highly-dynamic content in a timely fashion.

⁴<http://www.99designs.com>

CONCLUSION

A community is an emergent system. It forms from the actions of its members who are reacting to each other’s behavior. Here we have studied a specific community of individuals who are passionate about sharing pictures of items (mainly fashion and design items) using a mobile phone application. This community has a specific culture in which a set of habits, attitudes and beliefs guide how its members behave. In it, we have seen and quantified the importance of early adopters. In general, these individuals are those who initially set the unwritten rules that other community members learn (from observing those around them), internalize, and follow. In our case, early adopters tend to be successful trend spotters who like very diverse items. Trend makers, by contrast, tend to be highly organized individuals who focus on specific items. Understanding the characteristics of “the many” – of regular individuals with specific interests (trend makers) connected to early adopters with very diverse interests (trend spotters) – turned out to be more important than trying to find the “special few”. At least, it has been so for our social application, and for a variety of (more) complex networks [1, 13, 28].

ACKNOWLEDGEMENTS

The work was funded by the French National Research Agency (ANR) under the PROSE project (ANR-09-VERS-007), and by an industrial scholarship from PlayAdz⁵.

REFERENCES

1. Aral, S. and Walker, D. Identifying Influential and Susceptible Members of Social Networks. *Science*, 337, 2012.
2. H. Becker, M. Naaman, and L. Gravano. Learning Similarity Metrics for Event Identification in Social Media. In *Proceedings of the 3rd International Conference on Web Search and Data Mining (WSDM)*, 2010.

⁵<http://www.playadz.com>

3. H. Becker, M. Naaman, and L. Gravano. Beyond Trending Topics: Real-world Event Identification on Twitter. In *Proceedings of the 5th International AAAI Conference on Weblogs and Social Media (ICWSM)*, 2011.
4. M. Cha, H. Haddadi, F. Benevenuto, and K. Gummadi. Measuring User Influence in Twitter: The Million Follower Fallacy. In *Proceedings of the 4th International AAAI Conference on Weblogs and Social Media (ICWSM)*, 2010.
5. Z. Cheng, J. Caverlee, and K. Lee. You are Where You Tweet: A Content-based Approach to Geo-locating Twitter Users. In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management (CIKM)*, 2010.
6. Z. Cheng, J. Caverlee, K. Lee, and D. Sui. Exploring Millions of Footprints in Location Sharing Services. In *Proceedings of the 5th International AAAI Conference on Weblogs and Social Media (ICWSM)*, 2011.
7. R. Crane and D. Sornette. Robust Dynamic Classes Revealed by Measuring the Response Function of a Social System. In *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, volume 105, October 2008.
8. danah boyd. The Future of Privacy: How Privacy Norms Can Inform Regulation. Invited Talk at the 32nd International Conference of Data Protection and Privacy Commissioners, October 2010.
9. danah boyd. Designing for Social Norms (or How Not to Create Angry Mobs). <http://www.zephoria.org/thoughts/archives/2011/08/05/design-social-norms.html>, August 2011.
10. C. Droge, M. Stanko, and W. Pollitte. Lead Users and Early Adopters on the Web: The Role of New Technology Product Blogs. *Journal of Product Innovation Management*, 27(1), 2010.
11. A. Gelman and J. Hill. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, 2006.
12. M. Gladwell. *The Tipping Point: How Little Things Can Make a Big Difference*. Little, Brown and Company, 2000.
13. S. González-Bailón, J. Borge-Holthoefer, A. Rivero, and Y. Moreno. The Dynamics of Protest Recruitment through an Online Network. *Scientific reports*, 1, 2011.
14. A. Goyal, F. Bonchi, and L. V. Lakshmanan. Learning Influence Probabilities in Social Networks. In *Proceedings of the 3rd International Conference on Web Search and Data Mining (WSDM)*, 2010.
15. E. Katz and P. Lazarsfeld. Personal Influence: The Part Played by People in the Flow of Mass Communications. 1955.
16. D. Kempe, J. Kleinberg, and E. Tardos. Influential Nodes in a Diffusion Model for Social Networks. In *Automata, Languages and Programming*, volume 3580. Springer Berlin / Heidelberg, 2005.
17. H. Kwak, C. Lee, H. Park, and S. Moon. What is Twitter, a Social Network or a News Media? In *Proceedings of the 19th ACM Conference on World Wide Web (WWW)*, 2010.
18. M. Muller, N. S. Shami, D. R. Millen, and J. Feinberg. We are all Lurkers: Consuming Behaviors among Authors and Readers in an Enterprise File-Sharing Service. In *Proceedings of the 16th ACM International Conference on Supporting Group Work (GROUP)*, 2010.
19. M. Naaman, H. Becker, and L. Gravano. Hip and Trendy: Characterizing Emerging Trends on Twitter. *Journal of the American Society for Information Science and Technology*, 65, May 2011.
20. C. Neustaedter, A. Tang, and J. K. Tejjinder. The Role of Community and Groupware in Geocache Creation and Maintenance. In *Proceedings of the 28th International Conference on Human Factors in Computing Systems (CHI)*, 2010.
21. J.-P. Onnela, S. Arbesman, M. C. González, A.-L. Barabási, and N. A. Christakis. Geographic Constraints on Social Network Groups. *PloS One*, 6, 2011.
22. M. Panik. *Advanced Statistics From an Elementary Point of View*. Academic Press, 2005.
23. D. Saez-Trumper, G. Comarela, V. Almeida, R. Baeza-Yates, and F. Benevenuto. Finding Trendsetters in Information Networks. In *Proceedings of the 18th ACM International Conference on Knowledge Discovery and Data Mining (KDD)*, 2012.
24. C. Steinfield, N. B. Ellison, and C. Lampe. Social Capital, Self-esteem, and Use of Online Social Network Sites: A Longitudinal Analysis. *Journal of Applied Developmental Psychology*, 29(6), 2008.
25. J. Tang, J. Sun, C. Wang, and Z. Yang. Social Influence Analysis in Large-scale Networks. In *Proceedings of the 15th International Conference on Knowledge Discovery and Data Mining (KDD)*, 2009.
26. D. Watts. Challenging the Influentials Hypothesis. *Measuring Word of Mouth*, 3, 2007.
27. D. Watts and S. Strogatz. Collective Dynamics of Small-world Networks. *Nature*, 393(6684), 1998.
28. D. J. Watts. *Everything Is Obvious: *Once You Know the Answer*. Crown Business, March 2011.
29. J. Weng, E.-P. Lim, J. Jiang, and Q. He. TwitterRank: Finding Topic-sensitive Influential Twitterers. In *Proceedings of the 3rd ACM International Conference on Web Search and Data Mining (WSDM)*, 2010.
30. L. Yu, S. Asur, and B. A. Huberman. What Trends in Chinese Social Media. *The 5th Workshop on Social Network Mining and Analysis (SNA-KDD)*, 2011.