Follow My Friends This Friday!
An Analysis of Human-generated Friendship Recommendations

Ruth Garcia Gavilanes, Neil O'Hare, Luca Maria Aiello, Alejandro Jaimes

Yahoo! Research, Barcelona

Abstract. Online social networks support users in a wide range of activities, such as sharing information and making recommendations. In Twitter, the hashtag *#ff*, or *#followfriday*, arose as a popular convention for users to create contact recommendations for others. Hitherto, there has not been any quantitative study of the effect of such human-generated recommendations. This paper is the first study of a large-scale corpus of human friendship recommendations based on such hashtags, using a large corpus of recommendations gathered over a 24 week period and involving a set of nearly 6 million users. We show that these explicit recommendations have a measurable effect on the process of link creation, increasing the chance of link creation between two and three times on average, compared with a recommendation-free scenario. Also, ties created after such recommendations have up to 6% more longevity than other Twitter ties. Finally, we build a supervised system to rank user-generated recommendations, surfacing the most valuable ones with high precision (0.52 MAP), and we find that features describing users and the relationships between them, are discriminative for this task.

1   Introduction

Social media services have emerged as platforms on which people express opinions, obtain information about topics of interest (e.g. sports, fashion, etc.), discover breaking news, and receive updates from their friends, contacts, and their favorite celebrities. Most social media sites allow users to set up a network of connections (e.g., friends, contacts, celebrities) from which they can receive information. In some networks, such as Twitter or Google+, connections need not be reciprocal, and any user is free to *follow* any other user with a public profile, to be able to see their posts or status updates.

Since users are allowed to follow people they do not know, an important question is who *else* they should follow, in particular people who might be sources for the type of information they are interested in. In response to this need, *Follow Friday* emerged in 2009 as a spontaneous behavior from the Twitter user base, inspired by a blog post of an influential blogger[1]: users post tweets with the *#followfriday* or *#ff* hashtag, and include the usernames of the users they wish to recommend. As the name suggests, by convention these recommendations are made on Fridays. The key idea behind *Follow*

---

[1] http://mashable.com/2009/03/06/twitter-followfriday/

Friday is that people you already follow should be able to suggest new contacts that you will be interested in following.

In 2009 and 2010, in particular, the popularity of these hashtags on Twitter rose considerably, up to the point that the Twitter hashtags *#followfriday* and *#ff* were among the most popular hashtags observed in several large-scale Twitter corpora [25, 20].

Although Twitter now has an automatic recommender system for contacts, the analysis of the dynamics of the *Follow Friday* phenomenon is interesting from multiple perspectives. From the angle of complex systems analysis, measuring the effect that collective recommendation processes have in driving the connectivity choices of individuals is very valuable to quantify the ability of a system to self-organize. Additionally, our analysis identifies features that are most predictive of tie formation in a peer-to-peer link recommendation process. This is useful on the one hand to alleviate the information overload of users receiving recommendations from their peers, by identifying the 'strongest' recommendations among hundreds or even thousands, and on the other hand to improve the design of automatic contact recommendation algorithms.

In this paper we focus on the dynamics of *Follow Friday* as a form of broadcast recommendations, making the following main contributions:

- We analyse for the first time the dynamics of a large-scale human-driven recommendation system and, by comparing it with two baseline conditions, we measure its impact on the process of follower-link creation. We find that recommended users have a chance of being followed that is roughly two or three times higher than a recommendation-free scenario. We also measure how long the recommendation effect lasts, as well as the effect of repeated recommendations and the longevity of the accepted recommendations (i.e. how long these follower links persist).
- We develop a recommender system for ranking the human-generated recommendations received by a user. We evaluate this system against a corpus of known 'accepted' recommendations, identifying the features that are more predictive of link creation. Our recommender achieves a MAP of around 0.52, which is extremely high given the sparsity of the link recommendation problem. To the best of our knowledge, this is the first friend recommender system built and evaluated on human created recommendations.

The rest of this paper is organized as follows. In the next section we summarize related work, followed in Section 3 by a description of the dataset and a summary of key terminology. In Section 4 we analyze the *Follow Friday* phenomenon along a number of dimensions, and quantify the extent to which it has a real effect on users' *following* behavior. We then, in Section 5, propose and evaluate a recommender system for ranking a user's received *Follow Friday* recommendations. Finally, we conclude the paper in Section 6.

2 Related Work

The study of user-generated recommendations based on *Follow Friday* tags lies between two streams of research on recommender systems: recommendation based on user-generated content and social link recommendation.

A number of studies have been done on friendship recommendations in the context of Twitter, for instance Hannon et al. [11] compared collaborative filtering and content-based recommendation for the purpose of link recommendation on Twitter. Garcia Esparza et al. [9] presented a movie recommendation system that extracts information from a Twitter-like microblog platform for movie reviews. They profile 537 users and 1080 movies according to words and tags, and offer content-based and collaborative-filtering recommendations. Several aspects of user profiles have been studied for recommendations, for example Abel et al. [1] propose a methodology for modeling Twitter user profiles to support personalized news recommendation. They compare profiles constructed from the complete long-term user history with profiles based only on users' most recent tweets.

The task of predicting link formation (or deletion) in social graphs is one of the major challenges in the area of link mining, and has been well studied in the last decade [17, 19, 13]. Approaches have been proposed based on attributes of the nodes [13], structural graph features [26, 18], or both [2]. Unlike most of the work on link prediction that tries to predict future links in balanced sets of positive and negative samples, we are interested in a variant of link prediction, namely *link recommendation*, that is strictly user-centered and aims to provide a list of contacts to a user with the objective of maximizing the acceptance rate. Due to its inherent sparsity, this problem is more difficult than general prediction, and it has received little attention so far [4]. Previous studies also investigated what are the most predictive network and profile features for link formation in Twitter [12].

Despite the previous work in the area, we are not aware of any other attempt at characterizing human-generated recommendations and to leverage them to provide automatic contact suggestions. We also quantify the power of different features in predicting the formation of new links, not just considering structural or profile features of the user accounts, but focusing also on features that are descriptive of the human-driven recommendation process, such as the characterization of the relationship between the different human parties involved: the user who produces the recommendation, the one who receives it, and the one who is recommended.

3 A Dataset of Broadcast Friend Recommendations

Twitter is a social media platform on which users post 140-character messages called *tweets*. Users can follow other users and get notified with the tweets they post. For convenience, we will refer to the follower-followee relationship as a friendship relationship, although strictly speaking this relation only occasionally represents a true friendship: follower links are often not reciprocated [7] and often the followee can be an organization or a celebrity. So, in this sense, it is more correct to think of the followee as an information channel whose updates the user may be interested in subscribing to. In Twitter, a *hashtag* is any sequence of characters, without whitespace, preceded by the # symbol, and a 'mention' consists of any Twitter username preceded by the symbol '@'. We define *Follow Friday* recommendations as broadcast mentions of usernames in tweets containing the hashtag *#followfriday* or *#ff* (case-insensitive)[2]. So, for ex-

---

[2] We use the term *Follow Friday* to refer to the use of either of these *Follow Friday* hashtags.

	Total
Initial seed set	55,000
Receivers	21,270
Recommenders	589,844
Recommended Users	3,261,133
Recommendation Instances	59,055,205
Accepted Recommendation Instances	354,687
Rejected Recommendation Instances	58,700,518

Table 1: Unique # of Receivers, Recommenders, Recommended Users, Recommendation Instances, and Accepted and Rejected Recommendations.

ample, the tweet "`#followfriday @Lula and @Obama for being such great leaders`" recommends people to follow the Twitter users `Lula` and `Obama`.

In March 2011, using the Twitter stream API, we randomly selected a *seed set* of 55,000 users. To remove profiles that are unlikely to be legitimate or active, we follow the approach of Lee et al. [15] and exclude users who have more than 1000, or less than 100, followers or followees. This filter also excludes celebrities, who usually do not interact with other users [14]. We monitored the evolution of the seed users' followees over time by collecting snapshots of the seed users' contact networks during a 24 week period from March 24th, 2011 to September 5th, 2011. The snapshots were taken twice a week, every Thursday and Monday yielding a total of 48 network snapshots. This choice is motivated by the fact that, although the recommendations are mostly broadcast on Fridays (76%), there is still a non-negligible amount of recommendations broadcast on Saturday (14%) and Sunday (3%), therefore Thursday and Monday snapshots can describe the status of the network right before and right after the recommendation takes place.

In the remainder of this paper, we use the following terminology:

- **Receivers (Rcv)**. Users from the initial seed set who accepted at least one *Follow Friday* recommendation at any time during the 24 week period.
- **Recommenders (Rdr)**. The followees of the *receivers* (Rcv) who made at least one *Follow Friday* recommendation during the 24 week period.
- **Recommended users (Rdd)**. The users mentioned after the *Follow Friday* hashtag in the messages of the *recommenders* (Rdr).
- **Recommendation Instance (Rec)**. The tuple $\langle rdd, rdr, rcv, w \rangle$ identifying an instance of a recommended user, made by a recommender, and exposed to a specific receiver in a given week (w). We use lowercase letters to identify elements in the actors sets (e.g., $rdr \in Rdr$)
- **Acceptance**. We consider a *recommendation instance* made at time t to be *accepted* if its receiver becomes a follower of the recommended user between time t and time $t + \Delta$. Unless stated otherwise, the Δ considered is one week. Although we use the term *acceptance*, we cannot be sure about the causal relation between recommendation and acceptance (see discussion in Section 4.1).
- **Rejection**. We consider a recommendation instance made at time t to be *rejected* if the receiver does not follow the recommended user between time t and $t + \Delta$.

Recommended users who are already followees of the rcv are not considered in the analysis.

Table 1 summarizes the quantities of followers, receivers, recommenders, recommended users and recommendation instances in our dataset.

4 Analysis of Broadcast Recommendations

During the 24 weeks captured by our dataset, we have a total of 144,180 *unique* new followees, from 354,687 *Follow Friday* accepted *recommendation instances*: this means that, on average, for accepted *recommendation instances*, the *receiver* got recommendations to follow the same *recommended user* from 2 distinct *recommenders*. Table 2 shows the acceptance rate (the number of accepted recommendations divided by the number recommendation instances) for recommendations under various conditions where one of the actors involved in the *recommendation instance* mentioned one of the others in the previous week (using the '@username' convention). The first column indicates the direction of the mention and the users involved. The case of $rdr \rightarrow rdd$ (recommender mentions recommended) involves all recommendations since this is a necessary condition of a *Follow Friday* recommendation.

We can see that, overall, the acceptance rate is very low at 0.006 (i.e. 0.6% of recommendations are accepted), which is to be expected, since the recommendations are broadcast, as opposed to being personalized, and may not even have been seen by the receiver. When one of the actors mentions another, the acceptance rate tends to increase, which is expected, since these mentions are indicators of an active relationship. When either the *recommended user* (rdd) or the *receiver* (rcv) mention each other, the acceptance rates are roughly 10 times higher than the average (10% to 14% of recommendations accepted), which is not surprising since it shows that there is already a connection between these two users who form the new link. Note that while the acceptance rate for these particular cases is relatively high, the volume is low, indicating that these cases of pre-existing relationships are not typical of *Follow Friday* recommendation acceptances.

4.1 Effect of #FF recommendation

Since *Follow Friday* is a spontaneous recommendation phenomenon, the first question that arises is whether it has an actual impact on the creation of new follower links, and to what extent. In complex social systems, determining the causes of observed evolutionary phenomena is a very challenging task, due to the intrinsic difficulty in disentangling all the factors that produce the events observed in a-posteriori data-driven studies [23]. Even when controlled experiments are performed [3, 5], it is very difficult to know with absolute certainty which factors trigger the observed dynamics.

In our case, the inclusion of a new *recommended user* in the followee list cannot be interpreted directly as a cause-effect sequence, since the adoption may be driven by factors that are not related with the recommendation itself, such as unobserved online interactions or even exogenous events. Nevertheless, when sufficiently extensive

Mentions	Volume	Proportion	Acceptance Rate
rdr→rdd	59,055,205	1.000	0.006
rdd↔rdr	4,667,056	0.079	0.009
rcv→rdr	9,071,311	0.154	0.010
rdr→rcv	9,199,224	0.156	0.011
rdr↔rcv	6,242,059	0.106	0.012
rcv→rdd	205,447	0.003	0.095
rdd→rcv	238,822	0.004	0.097
rcv↔rdd	76,482	0.001	0.145

Table 2: Acceptance Rates for *Follow Friday* Recommendations, under various conditions where the users mention each other in the preceding week. For example 'rdd→rcv' indicates that the recommended mentioned the receiver, and 'rdd↔rcv' indicates that the recommended and receiver both mentioned each other. (rdd→rdr is omitted because it is identical to rdd↔rdr in this dataset: by definition, the recommender mentions the recommended for all recommendations.)

temporal data is available, it is possible to compare the evolution of the system under different conditions, or null models [6, 22], to understand if the the target factor has an effect, distinguishable from the other conditions, on the evolution of the system.

Specifically, we measure the added value of *Follow Friday* by comparing the acceptance rate of *#ff* or *#followfriday* recommendations with two alternative conditions:

(a) **Implicit recommendation model**: all usernames mentioned in any tweet received by users in the receiver set (Rcv) are considered implicit recommendations, based on the assumption that being exposed to the names of some users may increase the probability of adopting them as new followees. The implicit recommendations we consider are all mentions appearing in non-*#ff* tweets during the week before the target week, and that never previously appeared as an explicit *#ff* recommendation (for the same *receiver*) in the 24 week sample.

(b) **Unobserved recommendation model**: for this model, we assume that, due to unobserved factors, the contacts recommended through *#ff* hashtags would have been added by the Rcv set even in absence of any explicit *#ff* recommendation. These unobserved factors could include, for example, the rising popularity of the recommended user or relevance of the topics discussed by the recommended user to external breaking events. To model this condition, we apply a temporal shift: for the set of *#ff* recommendations made at time t, we measure their acceptance rate at time $t - 1$, before the actual recommendation is made, i.e. we measure the acceptance rate in a situation where the external conditions are similar (one week previously), but where no *Follow Friday* recommendation has been made. To keep this model separate from the implicit one, we exclude cases where the receiver received implicit recommendations, up to time $t - 1$, for the same recommended user.

The difference in the acceptance rate between the three models, depicted in Figure 1, shows that *#ff* recommendations lead users to follow a higher proportion of contacts compared to models in which *#ff* is not considered. Apart from an outlier at week
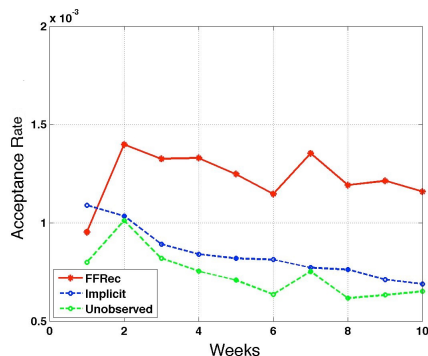
Fig. 1: The acceptance rate for *Follow Friday* recommendations in different weeks, compared with implicit and unobserved recommendation models.

1, the margin between the *#ff* model and the two alternative conditions is large, with *#ff* having an acceptance rate always between two and three times that of the others. By disentangling the role of the presence of the *#ff* tag from other factors that play an important role in the creation of social links, mainly homophily, the comparison with these alternative conditions provides strong evidence that the recommendation has an effect on the probability of link creation.

Whereas homophily may have a role in the selection of a recommended profile among other recommended ones, it seems not to be the main reason for the recommendation acceptance itself. Since the unobserved condition is simulated by performing a one week temporal shift, if we assume that the homophily effect between two users is not likely to change drastically in this one week time frame, then if the probability of acceptance is mainly determined by homophily, the *#ff* and *unobserved* conditions would have similar acceptance rates. The fact that this is not the case is, we believe, strong evidence that the *#ff* recommendation, and not purely the similarity between the profiles, drives the creation of the new link. Of course, there may be cases where the homophily effect changes drastically over the one week time shift, but it seems unlikely that this would explain all of the recommendation acceptances in this very large corpus.

A slight decreasing trend in acceptance rates over time is observed for all conditions, most likely due to the effect of the residual signal of explicit and implicit recommendations from the previous weeks (i.e. due to recommendations made before week 1 of our study, which we have no information about). To further verify this hypothesis, we measure how much the effect of an implicit or explicit recommendation lasts in time by computing the acceptance rate n weeks after the recommendation is made. To do so, we split our 24-week sample in half and observe the percentage of recommendations (from the first 12 weeks) that receivers followed up to 12 weeks after the recommendation was made. We do not consider cases where the recommendation was repeated after the week of the initial recommendation. Figure 2 reveals that the likelihood of subscribing to a recommended profile extends over several weeks and, after an initial substantial drop, fades slowly. We observe that the probability does not seem to stabi-
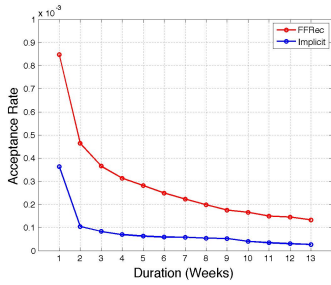
Fig. 2: Acceptance rates n weeks after a recommendation is made.
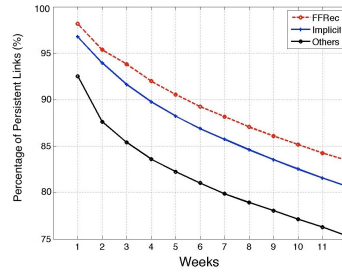

Fig. 3: Longevity of accepted recommendations

lize even after 12 weeks. Even though the scenario in which a user remembers a *Follow Friday* recommendation after several weeks is unlikely (especially if the recommendation has not been repeated), the probability decay is evident. The reasons behind such a long-lasting decay are difficult to find, since over such a large time scale many other interconnected events co-occur in the network's evolution. We argue that the effect of the *#ff* recommendation may introduce a perturbation in the network structure that may lead to delayed adoptions.

For instance, a user who received a recommendation before, but did not accept it, may create the link later because other users in his neighboring network accepted it, leading to new opportunities for social triangle closure [16].

To go beyond the acceptance rate of recommendation, we now look at the longevity of the new social ties created as a consequence recommendations. Figure 3 shows the percentage of acceptances that were still in the receivers network after n weeks. The curve labeled as *Others* represents all the users that were followed for reasons not related to the conditions considered in this study. After 12 weeks, we can see that 83% of *#ff* links are still in the receiver's network, versus 80% of links that follow implicit recommendations, and 76% of other follower links. This is an important finding in an environment such as Twitter where social ties have been observed to be very volatile [13].

4.2 Repeated Recommendations

In social sites such as Twitter, it is likely that a single broadcast Tweet may not be seen by many of a user's followers. Repeated recommendations, therefore, are likely to increase the likelihood of a recommendation being accepted, because the follower is more likely to see the recommendation at least once, and because repeated viewings of the recommendation may reinforce it. Figure 4 (a) plots the acceptance rate against recommendation repetitions, where repetitions are counted as recommendations received previously by a user within the time frame covered by the corpus. We consider two cases: when the recommendation is in the form of a *Follow Friday* recommendation only, and when there are only implicit recommendations. The results show that repeated recommendations make a significant difference. We can also see that it takes many implicit recommendations to have a similar effect as even a single *Follow Friday*
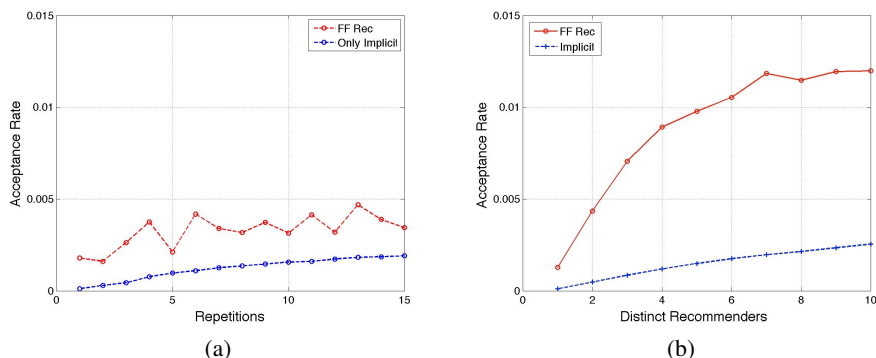
Fig. 4: The effect of repeated recommendations on the acceptance rate. (a) The number of repeated recommendations vs acceptance rate. (b) The number of distinct recommenders vs acceptance rate.

recommendation, with 15 implicit recommendation having a similar acceptance rate as 1 *Follow Friday* recommendation.

Figure 4 (b) plots the acceptance rate versus the number of distinct recommenders who recommend the same recommended user a receiver, and it shows a similar increase in the acceptance rate as the number of distinct recommenders increases, but with a bigger gap between the *Follow Friday* recommendations and the implicit model.

5 Recommender System

In the broadcast recommendation setting given by *Follow Friday*, users are exposed to a large number of friend recommendations every week. In a situation of information overload, the 'good' recommendations are likely to get lost among noisy ones, therefore automated methods are needed to detect the most valuable recommendations. We envision a scenario where all recommendations received by a user in a given week are ranked such that the good recommendations are at the top of the ranking. This essentially corresponds to providing recommender service built on top of the human-generated recommendation system.

In the following, we verify that it is possible to rank Twitter friendship recommendations and surface the most valuable ones, and we evaluate the utility of various features for this task. Secondly, by analysing the predictive value of different features for ranking recommendations, we supplement the analysis of the previous section, giving further insight into features that can predict the creation of a link after a recommendation is made.

5.1 Features for Ranking Recommendations

For each recommendation instance $\langle rdd, rdr, rcv, w \rangle$ we calculate a number of features, and group them into 3 main types: *user-*, *relation-*, and *format-based*.

User-Based features. These features describe an individual Twitter user, whether it be a *receiver*, a *recommender* or a *recommended user*. We identify two types of user-based features, attention-based and activity-based:

(a) *Attention-Based features* are related to the level of attention given to the user by other users. We measure *popularity* ($followers/(followers + followees)$), the number of times the user has been *mentioned*, the number of people mentioning the user, the number of times the user has been recommended with a *Follow Friday* hashtag, and the number of distinct recommenders.
(b) *Activity Based features* describe the level of activity of a user. We count the number of *new followees* of a given user for a given week, the *average tweets per day* of the user (over the entire history of the account), the number of *recommendations accepted* by receivers, and the number of distinct recommenders the a receiver has accepted recommendations from. Finally, we also count *mentions*, the number of distinct Twitter accounts mentioned by the user.

Relation-Based features. These features describe the relation between pairs of users in the $\langle rdd, rdr, rcv \rangle$ triple, based on either profile similarity or communication patterns.

(a) *Communication-based features* describe the level of communication between two users. *Conversational mentions* count the number of times a user mentions another user, and is calculated separately for each pair of users involved in a recommendation. The *number of Follow Friday recommendations*, (the number of times a user recommended another user) is calculated for each pair of users. We also count the number of previous acceptances between the *receiver* and the *recommender*, based on *Follow Friday* recommendations and on *implicit* recommendations. Last, we measure the friendship duration between the receiver and recommender by number of weeks.
(b) *Similarity-based features* describe the similarity of users. Separate *content-based similarity* features calculate similarity between all the tweets of two users, hashtags only, mentions only, and urls only. All these features use the weighted Jaccard similarity coefficient, as in Sudhof et al. [24]. *Geograpical similarity* is a binary feature, set to 1 if actors are in the same country or 0 if not. The location is parsed from the users's declared location using the Yahoo! PlaceMaker API[3].

Format-based features. These features describe a recommendation with information of the profile of the users, based either on the context or the format of the recommendation.

(a) The *repetitions* counts the number times the recommendation has been repeated in a *receiver's* timeline, the number of distinct *recommenders* who made the recommendation, and the number of previous weeks in which the recommendation was received.
(b) *Context* features describe the format or the context of the tweets containing the recommendation. We consider the *day of week* on which the recommendation was made, we record whether the recommendation was made in a retweet or not, the

---

[3] http://developer.yahoo.com/geo/placemaker/

number of other users appearing *together* with the recommended user in the *Follow Friday* tweet(s), and the *length of recommendation tweet* (the number of tokens in the tweet, excluding *#ff* hashtags and *@mentions*). Finally, we count the number of urls in the recommendation tweets. Since many of these measures can have more than one value for a given user (i.e. they receive the same recommendation from different people) we calculate both the maximum and minimum for all of them.

Most of the features are calculated over a temporal window prior to the recommendation. For all such features, we calculate two versions of the feature: (1) based on the one week period prior to the recommendation (to capture recent activity, similarity, etc), (2) based on all previous weeks in the corpus (to capture longer-term activity).

5.2 Evaluation Methodology

We consider all the unique *Follow Friday recommendation instances* that a given receiver is exposed to at week t and rank them with the aim of putting the 'best' recommendations at the top of the ranking. For the recommender, we set $\Delta = 2$, meaning that we consider the acceptances within two weeks of the recommendation, based on the ground truth of known acceptances. Recommendations accepted after two weeks are considered as not accepted.

Using the *acceptance* information as a ground truth, we evaluate our method by its ability to place the accepted recommendations towards the top of the ranking.

We split the data into training and test sets based on time, with data from weeks 1 to 16 used for training, and weeks 17 to 23 used for testing. We do not test against data from week 24, as we do not have details of the evolution of the followee network one week later.

We use two distinct methods to rank the *Follow Friday* recommendations received by a user in a given week: (1) based on a linear combination of the normalized scores from each feature, and (2) using the confidence score from a supervised classifier trained to classify recommendations as acceptances or rejections. To build the supervised classifier, we take a random subset of recommendations from the training set, ensuring that this subset contains a balanced set of acceptances and rejections. We train a binary classifier on this data using the Rotation Forest algorithm [21] as implemented in the WEKA library [10]. The Rotation Forest method constructs an ensemble of decision trees using random subspaces and principal components transformation applied to the input data [21]. For the linear combination of features, we normalize each list of recommendations by dividing by the feature's maximum value.

We do not normalize the similarity features based on the weighted Jaccard index, since those features are already normalized.

Since a receiver can receive many recommendations in a given week, and can accept one or more of them, we evaluate our various ranking approaches using the standard Information Retrieval measure *Mean Average Precision* (MAP). MAP evaluates a ranking by averaging the precision at the rank position where each relevant item is retrieved [8]. In the evaluation of friendship recommendation, an accepted recommendation is analogous to a relevant item, and a recommendation that is not accepted is non-relevant.

Ranking	MAP
Rotation Forest	0.4959
Linear Combination	0.0565
Random	0.0368

Table 3: Recommendation Mean Average Precision using all features

Features	MAP
All	0.4959
User-based	0.0741
Relation-based	0.3976
Format-based	0.0615
User + Relation	0.5176
User + Format	0.0790
Relation + Format	0.3787

Table 4: Recommendation performance for subsets of features (Rotation Forest)

5.3 Results

Table 3 shows performance of the Rotation Forest classifier, compared against the linear combination and a random baseline. The linear combination performs very poorly, while the Rotation Forest gives encouraging performance, with a MAP of almost 0.5, showing that machine learning approaches can give good results for this task.

In Table 4, we show the results when using various subsets of features, according to the grouping of features described in Section 5.1. The *relation-based* features are the most discriminative for friend recommendation, while the *format-based* features are not useful at all, and always harm performance. Finally, the *user-based* features, while they do not perform particularly well on their own, improve performance when combined them with the *relation-based* features. Overall, the best performing set of features is *user-based + relation-based* (i.e. ignoring the format-based features), with a MAP of almost 0.52.

Due to space we do not show detailed results for individual features, but the single best performing individual feature is the previous behaviour of the *receiver* in accepting recommendations from the *recommender*. Other relation-based features based on similarity (not communication) are also important, however, and the results in Table 4 show that optimal performance is achieved when we also consider user-based features.

6 Conclusions

In this paper, we describe the first study of the *Follow Friday* phenomenon, which aims to better understand the dynamics of a large scale collective process of human-generated link recommendations, and to understand the features and conditions that may predict the creation of new social links.

Furthermore, in contrast with other studies of link prediction in social media, we use a direct and reliable ground-truth of acceptances and rejections, based on real user be-

havior. We compare acceptance rates of *Follow Friday* recommendations with baseline conditions where (a) another user is mentioned, without being explicitly recommended, and (b) we simulate a condition where there is no observed (explicit or implicit) recommendation made via Twitter. Through this comparison, we show that explicit *Follow Friday* recommendations have a large, measurable, effect on who users choose to follow on Twitter. We also show that the effect of a recommendation (explicit or implicit) lingers for a number of weeks, that repeating recommendations has a strong effect, and that ties formed after *Follow Friday* recommendations tend to have more longevity than other ties, an important finding in Twitter, where social ties are quite volatile.

To surface more valuable recommendations above others, we propose an automated recommender system based on a number of features, which we group into three distinct categories: *user-based*, *relation-based* and *format-based*. We show that the most discriminative features for friendship recommendation are those features based on communication and similarity between users. In particular, past behavior in following recommendations coming from a given recommender is the most predictive feature of future recommendation acceptance.

7 Acknowledgments

References

1. Analyzing user modeling on twitter for personalized news recommendations. In *User Modeling, Adaption and Personalization*, volume 6787 of *Lecture Notes in Computer Science*, pages 1–12. 2011.
2. Luca Maria Aiello, Alain Barrat, Rossano Schifanella, Ciro Cattuto, Benjamin Markines, and Filippo Menczer. Friendship prediction and homophily in social media. *ACM Trans. Web*, 6, June 2012.
3. Luca Maria Aiello, Martina Deplano, Rossano Schifanella, and Giancarlo Ruffo. People are Strange when you're a Stranger: Impact and Influence of Bots on Social Networks. In *Proceedings of the 6th AAAI International Conference on Weblogs and Social Media (ICWSM)*, 2012.
4. Lars Backstrom and Jure Leskovec. Supervised random walks: predicting and recommending links in social networks. In *Proceedings of the fourth ACM international conference on Web search and Data Mining (WSDM)*, 2011.
5. Eytan Bakshy, Itamar Rosenn, Cameron Marlow, and Lada Adamic. The role of social networks in information diffusion. In *Proceedings of the 21st international conference on World Wide Web*, WWW '12, New York, NY, USA, 2012. ACM.
6. Alain Barrat, Marc Barthlemy, and Alessandro Vespignani. *Dynamical Processes on Complex Networks*. Cambridge University Press, New York, NY, USA, 1st edition, 2008.
7. Meeyoung Cha, Hamed Haddadi, Fabricio Benevenuto, and P. Krishna 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.

8. Bruce Croft, Donald Metzler, and Trevor Strohman. *Search Engines: Information Retrieval in Practice*. Addison-Wesley Publishing Company, USA, 1st edition, 2009.

9. Sandra Garcia Esparza, Michael P. O'Mahony, and Barry Smyth. On the real-time web as a source of recommendation knowledge. In *Proceedings of the fourth ACM conference on Recommender systems (RecSys)*, 2010.

10. Jan E. Gewehr, Martin Szugat, and Ralf Zimmer. BioWeka - extending the Weka framework for bioinformatics. *Bioinformatics/computer Applications in The Biosciences*, 2007.

11. John Hannon, Mike Bennett, and Barry Smyth. Recommending twitter users to follow using content and collaborative filtering approaches. In *Proceedings of the fourth ACM conference on Recommender systems*, RecSys '10, New York, NY, USA, 2010. ACM.

12. C.J. Hutto, Sarita Yardi, and Eric Gilbert. A longitudinal study of follow predictors on twitter. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 821–830, New York, NY, USA, 2013. ACM.

13. Haewoon Kwak, Hyunwoo Chun, and Sue Moon. Fragile online relationship: a first look at unfollow dynamics in twitter. In *Proceedings of the 2011 ACM Annual Conference on Human factors in computing systems (CHI)*, 2011.

14. Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue Moon. What is Twitter, a Social Network or a News Media? In *Proc. 19th ACM International Conference on World Wide Web (WWW)*, 2010.

15. Kyumin Lee, Brian Eoff, and James Caverlee. Seven months with the devils: A long-term study of content polluters on twitter. In *Proceedings of the Fifth International Conference on Weblogs and Social Media (ICWSM)*. AAAI.

16. Jure Leskovec, Lars Backstrom, Ravi Kumar, and Andrew Tomkins. Microscopic evolution of social networks. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge Discovery and Data Mining (KDD)*, New York, NY, USA, 2008. ACM.

17. David Liben-Nowell and Jon Kleinberg. The link prediction problem for social networks. In *Proceedings of the twelfth international conference on Information and knowledge management*, CIKM '03, pages 556–559, New York, NY, USA, 2003. ACM.

18. Linyuan Lu, Ci-Hang Jin, and Tao Zhou. Effective and efficient similarity index for link prediction of complex networks. In *arXiv:0905.3558 2009.*

19. Linyuan Lü and Tao Zhou. Link prediction in complex networks: A survey. *Physica A*, 2011.

20. Saša Petrović, Miles Osborne, and Victor Lavrenko. Rt to win! predicting message propagation in twitter. In *Proceedings of the Fifth AAAI International Conference on Weblogs and Social Media (ICWSM)*.

21. Juan J. Rodríguez, Ludmila I. Kuncheva, and Carlos J. Alonso. Rotation forest: A new classifer ensemble method. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 2006.

22. Rossano Schifanella, Alain Barrat, Ciro Cattuto, Benjamin Markines, and Filippo Menczer. Folks in folksonomies: social link prediction from shared metadata. In *Proceedings of the Third ACM international conference on Web search and Data Mining (WSDM)*, 2010.

23. Cosma Rohilla Shalizi and Andrew C. Thomas. Homophily and contagion are generically confounded in observational social network studies. *Sociological Methods & Research*, 2011.

24. Moritz Sudhof. Politics, twitter, and information discovery: Using content and link structures to cluster users based on issue framing. 11, 2012.

25. Bongwon Suh, Lichan Hong, Peter Pirolli, and Ed H. Chi. Want to be retweeted? large scale analytics on factors impacting retweet in twitter network. In *SocialCom/PASSAT'10*, 2010.

26. Tao Zhou, Linyuan Lu, and Yi-Cheng Zhang. Predicting Missing Links Via Local Information. In *European Physical Journal B*, volume 71(4) of *Special Issue: The Physics Approach to Risk: Agent-Based Models and Networks.*, 2009.