1-2016

A Comparison of Fundamental Network Formation Principles Between Offline and Online Friends on Twitter

Felicia NATALI
Singapore Management University, felician.2013@phdis.smu.edu.sg

Feida ZHU
Singapore Management University, fdzhu@smu.edu.sg

DOI: https://doi.org/10.1007/978-3-319-28361-6_14

Published in Advances in network science : 12th International Conference and School, NetSci-X 2016, Wroclaw, Poland, January 11-13, 2016, Proceedings, pp. 169-177
https://doi.org/10.1007/978-3-319-28361-6_14

A Comparison of Fundamental Network Formation Principles Between Offline and Online Friends on Twitter

Felicia Natali(✉) and Feida Zhu

School of Information Systems, Singapore Management University, 80 Stamford Rd, 178902 Singapore, Singapore
{felician.2013,fdzhu}@smu.edu.sg

Abstract. We investigate the differences between how some of the fundamental principles of network formation apply among offline friends and how they apply among online friends on Twitter. We consider three fundamental principles of network formation proposed by Schaefer et al.: reciprocity, popularity, and triadic closure. Overall, we discover that these principles mainly apply to offline friends on Twitter. Based on how these principles apply to offline versus online friends, we formulate rules to predict offline friendship on Twitter. We compare our algorithm with popular machine learning algorithms and Xiewei's random walk algorithm. Our algorithm beats the machine learning algorithms on average by 15 % in terms of f-score. Although our algorithm loses 6 % to Xiewei's random walk algorithm in terms of f-score, it still performs well (f-score above 70 %), and it reduces prediction time complexity from $O(n^2)$ to $O(n)$.

Keywords: Network formation · Offline friends · Online friends · Twitter · Social network · Offline friends prediction · Machine learning · Offline online

1 Introduction

Network formation has been studied in both the offline social network and the online social network. Before the emergence of the online social network, researchers investigated the offline social network. They discovered that the formation of the offline social network was characterized by a number of *dependencies* [16], also called *principles* [14]. These principles were by no means arbitrarily generated but were empirically discovered or theoretically formulated in previous studies on social networks [16]. When the online social network emerged, it was seen as a solution to the inconsistency and the high cost of procuring a large real life social networks data [12]. The principles of network formation that were previously discovered in the offline social network are now studied in the online social network. Most of these studies reveal that the principles that apply to the offline social network – such as reciprocity, mutuality, preferential attachment, and homophily – also apply to the online social network [7,9,11]. A provoking question then arises as to whether these similarities between the principles of offline and online network formation happen because "online social networks primarily support pre-existing social relations [3]", particularly the existing offline contacts [5].

To answer the question, we investigate how three fundamental principles of network formation proposed by Schaefer et al. [14] apply among offline pre-existing social relations — referred to as *offline friends* — versus non pre-existing social relations — referred to as *online friends* — on Twitter. In this study, *offline friends* comprises of followers or followees on Twitter whom a user knows in the real world, whereas *online friends* comprises of followers and followees on Twitter whom a user does not know in the real world. As such, the set of offline friends and the set of online friends are mutually exclusive.

Since we only have the ground-truth data of a user's offline and online friends, we are making an assumption that all offline friends are pre-existing social relations, and all online friends are non pre-existing social relations. We believe this is a reasonable assumption to make because people maintain an online social network mainly to keep in touch with existing social relations that they have offline and meet new people online [5].

2 Fundamental Principles of Network Formation Among Offline Versus Online Friends

Social networks are formed through multiple principles. Snijders listed some of the important ones in his work [16], they are: reciprocity, homophily, transitivity, degree differentials (popularity), and hierarchies. Schaefer et al. particularly picked up three principles — reciprocity, popularity, and triadic closure — to study the process of network formation among preschool children [14]. They proposed that these principles were general. Through longitudinal study using the SIENA modeling framework [15], they discovered that reciprocity, popularity and triadic closure shaped the formation of pre-school children's networks. As most children regularly interact with their peers for the first time in preschool, and they do not have prior social experience that might contaminate their motivation in creating social ties with their friends, the principles that govern their network formation are considered fundamental. Therefore, we choose these three principles to investigate in this study.

For our analysis, we use the dataset by Xie et al. [17]. This dataset contains the data of 98 Twitter users that includes his ego network in 2011 and the list of his Twitter friends (followers or followees) whom he knows in real life. Overall, the dataset has 20030 Twitter users (ego users and their alters) and 23225 edges labeled as an offline or an online friend. We only use 49 ego networks (9380 users and 10153 labeled edges) for our observation. Based on our observation, we formulate rules to predict offline friendship and use the rest 49 ego networks for our prediction task.

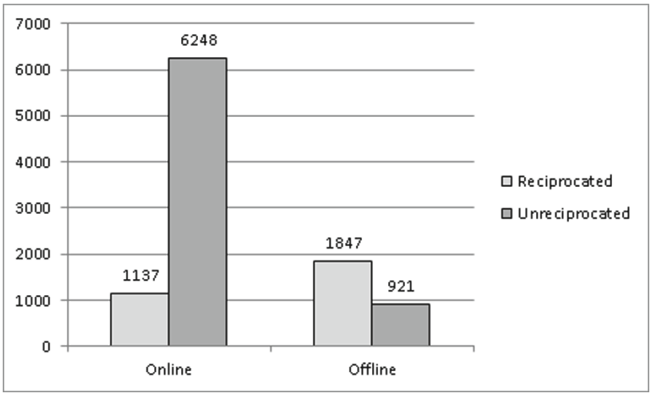


Fig. 1. Reciprocated links among offline and online friends.

2.1 Reciprocity

Reciprocity means requiting a benefit received [8]. Since friends enjoy equality in right, privileges, and obligations [10], reciprocity becomes the basis of friendship. On Twitter, reciprocity can happen when two users reply each other, mention each other, follow each other, etc. In this study, we focus on reciprocity that has a direct impact on a Twitter follow network dependency, that is, reciprocity when two users follow each other. Although reciprocity is one of the basic principles of moral codes in a society which enables social stability [8], it may not necessarily assume such a fundamental role when it comes to online friends in an online society. Therefore, in this study, we answer the following research question:

Research Question 1. *Does reciprocity as the basis of Twitter follow network formation happen as often among online friends as among offline friends?*

Figure 1 shows the distribution of reciprocated links among offline and online friends. To answer the research question, we perform chi-square test of independence to check whether reciprocity depends on the type of friendship (offline or online). Our result shows that reciprocity depends on the type of friendship with odds ratio 11.02 ($\chi^2 = 2553.8$, p-value < 0.001). Offline friends are 11 times more likely to reciprocate on Twitter.

Based on this observation, we create our first rule to predict offline friendship. Given two online friends, A and B, on Twitter,

Rule 1. *IF A and B reciprocate on Twitter THEN A and B are offline friends.*

2.2 Popularity

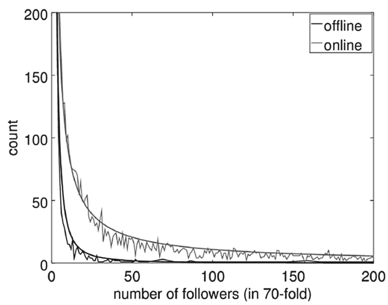
Popularity means the state of having many connections. An individual's popularity increases as the idealized qualities imposed by society increase, e.g. wealth, beauty, and social skill [1]. These idealized qualities increase one's attractiveness and invite connections. As popularity allows a person to access more resources [4], popularity also entails higher popularity. The theoretical account of this phenomenon was elaborated by Price in 1976 [13]. This phenomenon is called *the-rich-get-richer phenomenon*, or *preferential attachment* [2]. Therefore, popularity in itself is also an idealized quality that increases one's attractiveness. On Twitter, the number of followers is the simplest measure of popularity.

Although preferential attachment has been shown to exist in both the online social network [11] and the offline social network [13], we wonder whether the rate at which popularity increases a user's attractiveness among online friends differs from the rate at which it does among offline friends. In this study, we answer the following research question:

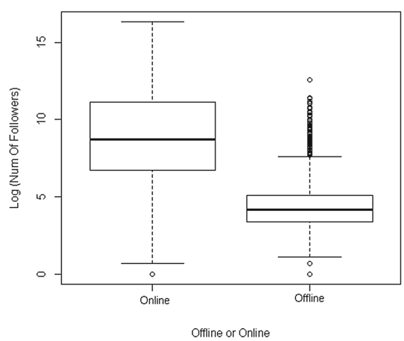
Research Question 2. *On Twitter, does preferential attachment happen among online friends at the same rate as it does among offline friends?*

We plot the distributions of the number of followers of offline friends and online friends. Although in general they follow the power law, there is too much fluctuation in the distributions, thus making it impossible to find the parameters that fit a power law curve closely. Therefore, we try several folds of number of followers and discover that the distributions of the number of followers (in 70-fold) of both offline friends and online friends fit the power law closely ($N = cx^{-\alpha}$ where N is the frequency of users with a specific number of followers, and x is the number of followers in 70-fold), but at different parameters c and α (c is 1482.16 and α is 1.70 among offline friends, c is 769.13 and α is 0.92 among online friends. See Fig. 2a). The power law distributions show that preferential attachment exists [13], and it happens at a faster attachment rate among offline friends judging by the larger α.

A stranger (online friend) has a thicker tail, meaning he has a greater tendency to have a higher number of followers. The next question is, whether there



(a) Distributions of the number of followers of offline and online friends follow the power law.



(b) Boxplot of the number of followers of offline friends and online friends.

Fig. 2. The number of followers of offline and online friends

is a number of followers at which a user is likely to be an online friend to anyone. According to previous studies, there may be. Kwak et al. discovered that homophily was not observed between a user who had more than 1000 followers and his reciprocal friends [9]. Moreover, another study showed that 71 % of top link farmers (users who try to acquire large numbers of follower links to amass influence) on Twitter had more than 1000 followers [6]. Link farmers usually reciprocate even those whom they do not know to amass social capital and promote their Twitter content. As a result, many of the users in their network are strangers. Our boxplot in Fig. 2b also shows that a user who has more than 1000 followers ($\log 1000 = 6.9$) is at around the 87th percentile of all offline friends. Meanwhile, such a user is only at around the 25th percentile of all online friends. Thus, we formulate our second rule to predict offline friendship. Given two online friends A and B on Twitter,

Rule 2. *IF B has more than 1000 followers THEN A and B are not offline friends.*

2.3 Triadic Closure

Triadic closure happens between offline friends because of the increased propinquity and the psychological need for balance between two individuals who share mutual friends [14]. If we assume that a triadic closure in real life translates into a triadic closure online, it is likely that triadic closure happens between offline friends on Twitter. On the other hand, as the pressure towards closure may not be as strong among online friends due to the lack of propinquity, we ask the following research question:

Research Question 3. *Are triadic closures on Twitter as likely to happen among online friends as they are among offline friends?*

We answer the research question by the following logit function:

$$Pr(triadicclosure = 1 | I_1, I_2) = F(\beta_0 + \beta_1 I_1 + \beta_2 I_2) \tag{1}$$

I_1 is 1 if there is 1 offline friendship between any two users in a triad, I_2 is 1 if there are 2 offline friendships between any two users in a triad, and I_1 and I_2 are 0 if there is no offline friendship in a triad. F is the cumulative standard logistic distribution function.

The result shows that when offline friendship does not exist, a triadic closure is unlikely to happen (β_0 -3.36, p-value < 0.0001). When an offline friendship exists, the probability of a triadic closure increases ($\beta_1 = 0.60$, p-value < 0.0001). When two offline friendships exist, the probability increases further ($\beta_2 = 1.41$, p-value < 0.0001). From the result, we expect that when three offline friendships exist in a triad, an online triadic closure is even more likely to happen even though the ground-truth data that we have does not allow us to validate our expectation. In summary, when offline friendships exist in a triad, a triadic closure online is more likely to happen.

From this observation, we formulate the following rule to predict offline friendship. Given A-B-C, an online closed triad on Twitter,

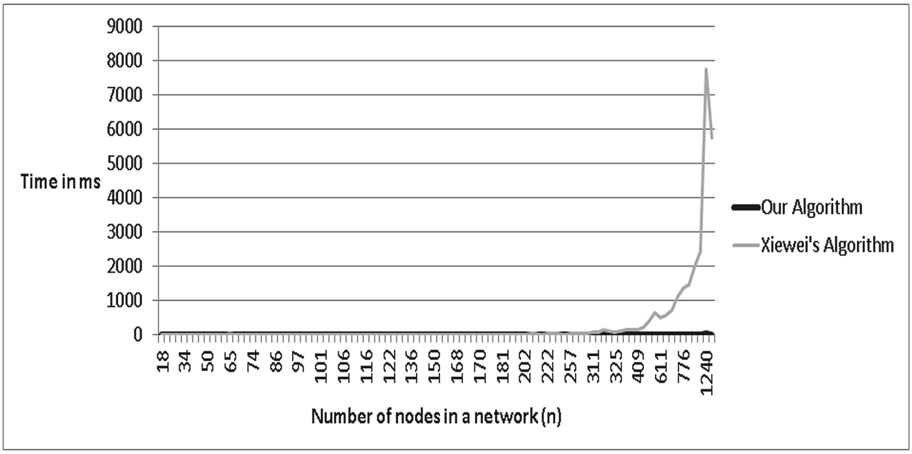


Fig. 3. Milliseconds required to perform prediction

Table 1. Prediction Results

Algorithm		Precision	Recall	F-score
Our algorithm		0.78	0.74	0.76
Machine learning	Logistic regression	0.73	0.52	0.61
	Naive bayes	0.47	0.81	0.60
	Support vector machine	0.78	0.36	0.50
	Artificial neural network	0.72	0.72	0.72
Xiewei's random walk algorithm		0.77	0.88	0.82

Rule 3. *IF A and B are offline friends AND B and C are offline friends, THEN A and C are offline friends.*

3 Practical Application: Predicting Offline Friendship on a Twitter Network

A hands-on practical application from the above observation is the formulation of rules for offline friendship prediction on a Twitter network which we will investigate in this work. We predict a user's offline friends on Twitter based on the three rules we formulate above (Algorithm 1). We compare the results with Xiewei's random walk algorithm and several popular machine learning algorithms. Xiewei's algorithm [17] creates a matrix of a user's ego network and assigns a probability of walk from a user to his Twitter followers that decreases polynomially as a user's number of followers increases. Therefore, a user who has 1000 followers has a lower probability of walk to anyone than a user who has 100 followers. When the probability of walk to a friend is higher than the probability of walk to another friend who has the median number of followers, the friend is regarded as an offline friend. The process is performed iteratively to include offline friends of offline friends as offline friends. For the machine learning algorithms, we extract various features on Twitter as predictors such as tweets LDA-topic similarity, the number of replies, the number of mentions, various centrality measures, follower overlap, followee overlap, the type of following link, etc.

The prediction result is shown in Table 1. Overall, our algorithm performs well and beats the machine learning algorithms. Although its predictive accuracy loses to Xiewei's, our algorithm reduces the time complexity from $O(n^2)$ to $O(n)$ (See Fig. 3).

```
Data: a Twitter user, u_i
Result: u_i's offline friends, C_i
u_i has a set of friends on Twitter S_i where S_i = {f_1, f_2, f_3...};
Let C_i be the set of u_i's offline friends;
for each friend f_j ∈ S_i do
    Apply Rule 1: If u_i and f_j reciprocates on Twitter then f_j ∈ C_i;
    for each friend f_j ∈ C_i do
        Apply Rule 2: If f_j has a number of followers larger than 1000
        then f_j ∉ C_i
    end
end
Apply Rule 3: Offline friends of an offline friend are offline friends;
temp = {u_i};
while temp.size != 0 do
    for each friend f_j ∈ C_i do
        Let S_j be the set of f_j's friends on Twitter where S_j ⊂ S_i;
        Let C_j be the set of f_j's offline friends where C_j ⊂ S_i;
        for each friend f_g ∈ S_j do
            Apply Rule 1: If f_j and f_g reciprocates on Twitter then
            f_g ∈ C_j;
            for each friend f_g ∈ C_j do
                Apply Rule 2: If f_g has a number of followers larger than
                1000 then f_g ∉ C_j
            end
        end
        temp = {temp ∪ C_j};
    end
    temp = temp \ {C_i, u_i} ;
    C_i = {C_i ∪ temp};
end
```

Algorithm 1. Offline friendship prediction

4 Conclusion

We have shown that some of the fundamental principles of social network formation, namely reciprocity, popularity, and triadic closure apply mainly to offline friends on Twitter. The results suggest that using an online social network as a substitute for a real life social network requires careful consideration as the dynamics that apply to the offline social network does not necessarily apply to the online friends in the online social network. We also use the results of our observation to create an efficient algorithm for offline friendship prediction. Future work can be directed to assess the applicability of the algorithm across various social networks in a larger dataset.

Acknowledgment. This research is supported by the Singapore National Research Foundation under its International Research Centre @ Singapore Funding Initiative and administered by the IDM Programme Office, Media Development Authority(MDA). It is also partly supported by Pinnacle Lab at Singapore Management University(SMU).

References

1. Adler, P.A., Kless, S.J., Adler, P.: Socialization to gender roles: popularity among elementary school boys and girls. Sociol. Educ. **65**(3), 169–187 (1992)
2. Barabási, A.-L., Albert, R.: Emergence of scaling in random networks. Science **286**(5439), 509–512 (1999)
3. Boyd, D.M., Ellison, N.B.: Social network sites: definition, history, and scholarship. J. Comput. Mediated Commun. **13**(1), 210–230 (2007)
4. Coie, J.D., Dodge, K.A.: Continuities and changes in children's social status: a five-year longitudinal study. Merrill-Palmer Q. **29**, 261–282 (1983)
5. Ellison, N.B., Steinfield, C., Lampe, C.: The benefits of Facebook friends: social capital and college students use of online social network sites. J. Comput. Mediated Commun. **12**(4), 1143–1168 (2007)
6. Ghosh, S., Viswanath, B., Kooti, F., Sharma, N.K., Korlam, G., Benevenuto, F., Ganguly, N., Gummad, K.P.: Understanding and combating link farming in the twitter social network. In: WWW 2012 Proceedings of the 21st international conference on World Wide Web, pp. 61–70. ACM (2012)
7. Golder, S., Yardi, S.: Structural predictors of tie formation in twitter: transitivity and mutuality. In: 2010 IEEE Second International Conference on Social Computing (SocialCom), pp. 88–95. IEEE (2010)
8. Gouldner, A.W.: The norm of reciprocity: a preliminary statement. Am. Sociol. Rev. **25**(2), 161–178 (1960)
9. Kwak, H., Lee, C., Park, H., Moon, S.: What is twitter, a social network or a news media? In: WWW 2010 Proceedings of the 19th International Conference on World Wide Web, pp. 591–600. ACM (2010)
10. Laursen, B., Hartup, W.W.: The origins of reciprocity and social exchange in friendships. New Dir. Child Adolesc. Dev. **2002**(95), 27–40 (2002)
11. Leskovec, J., Backstrom, L., Kumar, R., Tomkins, A.: Microscopic evolution of social networks. In: KDD 2008 Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 462–470. ACM (2008)

12. Newman, M.E.J.: The structure and function of complex networks. Soc. Ind. Appl. Math. (SIAM) Rev. **45**(2), 167–256 (2003)
13. Price, D.D.S.: A general theory of bibliometric and other cumulative advantage processes. J. Am. Soc. Inf. Sci. **27**(5), 292–306 (1976)
14. Schaefer, D.R., Light, J.M., Fabes, R.A., Hanish, L.D., Martin, C.L.: Fundamental principles of network formation among preschool children. Soc. Netw. **32**, 61–71 (2010)
15. Snijders, T.A.B.: The statistical evaluation of social network dynamics. Sociol. Methodol. **31**(1), 361–395 (2001)
16. Snijders, T.A.: Statistical models for social networks. Ann. Rev. Sociol. **37**, 131–153, 469 (2011)
17. Xie, W., Li, C., Zhu, F., Lim, E., Gong, X.: When a friend in twitter is a friend in life. In: WebSci 2012 Proceedings of the 4th Annual ACM Web Science Conference, pp. 344–347. ACM (2012)