1-2016

Investigating the influence of offline friendship on Twitter networking behaviors

Young Soo KIM
Singapore Management University, yskim@smu.edu.sg

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

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

Ee-Peng LIM
Singapore Management University, eplim@smu.edu.sg

DOI: https://doi.org/10.1109/HICSS.2016.97

Investigating the Influence of Offline Friendship on Twitter Networking Behaviors

Youngsoo Kim*, Felicia Natali*, Feida Zhu*, and Eepeng Lim*
*School of Information Systems
Singapore Management University, Singapore 178902
Email: yskim@smu.edu.sg, felician.2013@smu.edu.sg, fdzhu@smu.edu.sg, eplim@smu.edu.sg

***Abstract*—We investigate the influence of offline friendship in three specific areas of Twitter networking behaviors: (a) network structure, (b) Twitter content and (c) interaction on Twitter. We observe some interesting findings through the empirical analysis of 2193 pairs of users who are online friends. When these pairs of users know each other offline, they are more likely to (1) respond to the online gesture of friendship from their friend, (2) share mutual online friends, (3) distribute and gather information in their friend's Twitter network, (4) pay attention to their friend's tweets, (5) post tweets that might be of interest to their friend, (6) post tweets similar to their friend's, (7) respond to their friend's tweet, (8) mention their friend in tweets, and (9) distribute their friend's tweets. Overall, offline friendship drives social networking activities on Twitter.**

***Keywords*-networking behavior; social network; online friendship; offline friendship; social network services; Twitter;**

I. Introduction

Social Networking Sites(SNS) provide new avenues for friends, acquaintances, and even strangers to connect. One of the most popular SNS is Twitter. As of May 2015, Twitter has more than 500 million users, out of which more than 302 million are active users[1]. Twitter's popularity and its simple crawling API have drawn much interest from the research community.

Twitter networks create a potential for users to maintain pre-existing relationships and establish new ones. As a result, a Twitter network can consist of offline friends and online friends. Given that in a Twitter network, some friends may know each other offline, we attempt to explore whether and how the existence of offline friendship between two Twitter users influences their Twitter networking behavior.

Social network sites are web-based services that allows individuals to construct a profile, articulate a list of other users with whom they share a connection, and traverse their list of connections and those made by others within the system[2].In social network sites, a user (1) creates connections, (2) generates content, and (3) communicates with his friends. Given this framework, we examine the influence of offline friendship on three specific areas of Twitter as a social networking tool, namely network structure, Twitter content, and interaction on Twitter.

II. Relevant Studies

The study of offline versus online world has drawn interest from various research communities, such as psychology, sociology, computational social science, computer science, data mining, and even physics. Numerous studies have been published on the topic. We will expound them sequentially based on their motivations, namely to validate or invalidate the theories of offline social network, to compare the quality of offline versus online friendship, to investigate the influence of online behavior on offline behavior, to predict one's offline friendship and behavior from one's online social network, and to examine the influence of offline friendship on online behavior. The list is by no means exhaustive, but it gives a general overview on what the research community has done so far.

Before the emergence of the online social network, various social theories have been developed for the offline social network. Therefore, some researchers exerted effort to validate or invalidate these well-known social theories in the online social network. Dunbar et al.(2015) discovered that the structure of online social networks mirrored those in the offline world[3]. They proved that the layered structure corresponding to the frequency of contact in the offline social network also applied in the online social network. Gonçalves et al.(2011) validated Dunbar's number — the theoretical cognitive limit on the number of stable relationship a person can maintain — in Twitter conversation[4]. They discovered that, just like in the offline world, a person can only entertain maximum 100-200 stable relationships on Twitter.

In psychology, and sociology, many of the offline-online friendship studies focused on comparing the difference between offline and online friendship. Chan and Cheng(2004) showed that offline friendships involved more interdependence, breadth, depth, code change, understanding, commitment, and network convergence[5]. Antheunis et al.(2012) observed that offline friendship was higher in quality(i.e., closeness, importance, help, and trust)[6]. Buote et al.(2009) explored the similarities and differences between offline and online friendships in terms of attachment style[7].

Other studies investigated the influence of online behavior on offline behavior. An example is the study by Ellison et al.(2007)[8]. It investigated the influence of one's online social network use on one's self-esteem.

1530-1605/16 $31.00 © 2016 IEEE
DOI 10.1109/HICSS.2016.97

IEEE computer society

In data mining and computer science, many studies predicted one's offline friendship and offline behavior from one's online social network. Heatherly et al.(2009)[9] inferred offline private information, namely political affiliation, from the online social network structure. Dai et al.(2012)[10] investigated how online friendship network structure revealed offline high-risk sexual behavior. Xie et al.(2012) created an algorithm to predict offline friends on Twitter[11]. The algorithm utilized a single variable, i.e. the number of followers, to define the probability of random walk from a user to his friends in his Twitter network. If the probability of random walk to any user was higher than a benchmark, the user was regarded as an offline friend. The process was performed iteratively to discover all offline friends. Backstrom and Kleinberg(2013) formulated a new network measure called *dispersion* — the extent to which two people's mutual friends are not well-connected — to predict family members on Facebook[12].

Lastly, some studies examined the influence of offline friendship on online behavior. So far, we have only discovered one work that did so. Yin et al.(2014) discovered that there was no correlation between the offline and the online interaction in Meetup. Offline interaction(*co-attend event*) did not result in a greater online interaction(*co-join group, co-comment event, online message*)[13]. In terms of motivation, our study is similar to Yin's. We investigate the influence of offline friendship on online networking behavior. However, unlike Yin et al., we do not analyze an *Event-Based Social Network*, a social network that provides a platform for users to engage in various offline activities[14]. Instead, we analyze Twitter, a social platform to communicate and spread news. Besides, Yin et al. did not only focus on investigating the influence of offline interaction on online networking behavior. Therefore, they only examined the influence of offline interaction on one networking behavior, that is, online interaction. We investigate the influence of offline friendship on various networking behaviors, not only interaction. We will cover the theoretical background for the online networking behaviors investigated in this study in Section IV-A.

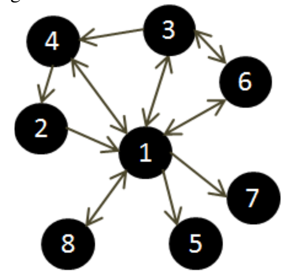
III. Conceptual Framework and Research Questions

In this study, we answer the following research question:

Given a pair of Twitter users who are online friends, how does the presence of offline friendship between them influence their online networking behaviors in comparison to other pairs who are also online friends, but do not know each other offline?

The definition of friendship in this study includes acquaintanceship. Therefore, we define offline friendship as a friendship between two users who know each other in the offline world.

Figure 1. Local Network of User 1



On the other hand, we define online friends as two users who are connected in Twitter, regardless of the connection type. Therefore, a follower in Twitter is also an online friend. We think this is a reasonable assumption to make because our definition of friendship includes acquaintanceship. Although a public user does not have a choice over his followers, he may still get acquainted with his followers through replies or likes that his followers generate for him. In summary, users A and B are online friends in either of the following situations: A follows B, B follows A, or A and B follow each other. When A follows B, A is called B's follower, and B is called A's followee.

We examine the influence of offline friendship on three specific areas of Twitter networking behaviors: network structure, Twitter content similarity, and interactions on Twitter.

In terms of network structure, we examine the influence of offline friendship on reciprocity and the number of mutual friends online because these two variables are closely related to the fundamental principles of social network formation. We also investigate the influence of offline friendship on other network measures — edge betweenness centrality and another two that we develop on our own — because they are related to communication and information distribution and gathering on Twitter. The scope of our analysis for the network structure is a Twitter local network. A Twitter local network consists of a local user and the users to whom the local user is directly connected to, called alters. A local network also includes all the links between all the users in its network (see Figure 1). The local network in Figure 1 is a follow network. Therefore, we draw a link from user i to user j if user i follows user j. A follow link goes against the information flow. Thus, information flows from user j to user i. In Figure 1 user 1 follows user 5. Information flows from user 5 to user 1.

Twitter content is generated in the form of 140-character short messages that can be published along with a picture. These messages are called tweets. Therefore, in terms of Twitter content, we investigate the influence of offline friendship on tweets similarity.

In terms of interaction, we investigate the influence of offline friendship on four different mechanisms of interaction on Twitter: favorite, retweet, reply, and mention. Favorite is liking a friend's status, retweet is reposting a friend's post, reply is responding to a friend's post and mention is mentioning a friend's in a post.

Table I summarizes all the networking behaviors we observe in this study.

IV. Data and Measures

We analyse a sample of 2,193 pairs of online friends on Twitter. This sample is taken from 98 Twitter local networks in 2011. A survey was conducted in 2011 for the 98 Twitter users. Each of the Twitter users was asked whether he knew his Twitter friends in real life. If he did, offline friendship existed between him and his friends.

The pairs that we include in our analysis are the pairs of a local user and his alter who posted English tweets in 2011. We exclude the pairs with any member who has larger than 1,000 followers. There are three reasons for this exclusion. First, in 2010, the users with followers larger than 1,000 made up less than 1% of all Twitter users[15]. In 2013, they made up less than 4% of active Twitter users[16]. Second, 71% of the top spammers have more than 1,000 followers[17]. Third, homophily in terms of popularity and geographic location is not observed between these users and their reciprocal friends on Twitter[15]. Therefore, we can reasonably assume that these users most likely act as a news or business media than a friend to anyone on Twitter. With such an assumption, our research question becomes irrelevant to these online connections as these connections can hardly be defined as friendships.

After the exclusion, we get the sample of 2,193 pairs of online friends for our analysis. Out of these pairs, 873 know each other offline.

We examine the influence of offline friendship on the variables pertaining to: (1) network structure (2) tweets similarity, and (3) interaction on Twitter. All the variables are listed in Table I.

A. Network Structure

Various network measures have been developed to describe a network structure. These basic measures are covered comprehensively by Newman[18]. In this study, we select several measures that are interesting for us. We examine reciprocity and the ratio of overlapping friends as they relate to the fundamental principles of social network formation: reciprocity and triadic closure[19]. We examine edge betweenness centrality as it can quantify a user's ability to facilitate communication on Twitter. We develop two measures based on the number of reachable users, and closeness centrality as they can quantify a user's independence to distribute, and gather information on Twitter.

1) Reciprocity.: Reciprocity is one of the fundamental principles of social network formation[19]. In Twitter, reciprocity means following each other. When a user follows a friend, he will receive his friend's updates on his news feed. Following a user on Twitter is a gesture of friendship from the follower. Through reciprocity, we investigate the influence of offline friendship on responding to an online gesture of friendship.

We measure reciprocity between users i and j by the following formula:

$$Reciprocity_{ij} = \begin{cases} 1 & \text{if user } i \text{ follows user } j \text{ and vice versa} \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

2) Followers and Followees Overlap.: Followers and followees overlap measure the extent to which two users have the same followers and followees in their Twitter network. People have the tendency to form friendships with those with whom they share multiple mutual friends. This concept is called triadic closure, one of the fundamental principles of social network formation[19]. Through followers and followees overlap, we investigate the influence of offline friendship on forming online friendships with whom one shares mutual friends.

We measure follower overlap between user i and user j by the following formula[20]:

$$FollowerOverlap_{ij} = \frac{\text{\# of common followers of } i \text{ and } j}{\text{\# of unique followers of } i \text{ and } j} \tag{2}$$

We measure followee overlap between user i and user j by the following formula[20]:

$$FolloweeOverlap_{ij} = \frac{\text{\# of common followees of } i \text{ and } j}{\text{\# of unique followees of } i \text{ and } j} \tag{3}$$

3) Network Coverage Ratio.: Network coverage ratio measures how independent a user is in distributing or gathering information to or from the alters in his friend's local network. We develop the measure based on the number of reachable users in the local network. The number of reachable users represent the number of users whom you can distribute information to, or gather information from. We do not only consider direct friends because in online social network, information does not only travel to direct friends.Ye and Wu (2010) observed that a significant portion of messages on Twitter travels far away from the originator and his/her followers[21].

As information flows to followers, in estimating how independent a user is in distributing information in his friend's local network, we consider the number of reachable followers the user can reach:

$$NCR^D_{ij} = \frac{\text{\# user } i\text{'s reachable followers in user } j\text{'s local network, excluding user } j\text{, when tweets cannot flow through user } j}{\text{\# user } i\text{'s reachable followers in user } j\text{'s local network, excluding user } j} \quad (4)$$

The formula above calculates how independent user i is on user j in distributing information in user j's local network. When user i is completely independent from user j, the value of NCR^D_{ij} is 1. On the other hand, if user i is completely dependent on user j, the value is 0.

As information flows from followees, in estimating how independent a user is in gathering information in his friend's local network, we consider the number of reachable followees the user can reach:

$$NCR^G_{ij} = \frac{\text{\# user } i\text{'s reachable followees in user } j\text{'s local network, excluding user } j\text{, when tweets cannot flow through user } j}{\text{\# user } i\text{'s reachable followees in user } j\text{'s local network, excluding user } j} \quad (5)$$

The formula above calculates how independent user i is on user j in gathering information in user j's local network. When user i is completely independent from user j, the value of NCR^G_{ij} is 1. On the other hand, if user i is completely dependent on user j, the value is 0.

We assume that the more independent user i is from user j in distributing or gathering information to or from user j's friends, the more likely user i is in posting or gathering information to or from user j's friends because (1) it is easier for user i to do so, (2) the information coming to or from user j's friends is free from the influence of user j.

For illustration, the following is an example on how to calculate NCR^D_{31}. In Figure 1, there are 4 reachable followers of user 3 in user 1's local network, excluding user 1: user 4 ($4 \rightarrow 1 \rightarrow 3$), user 2 ($2 \rightarrow 1 \rightarrow 3$), user 8 ($8 \rightarrow 1 \rightarrow 3$), and user 6 ($6 \rightarrow 3$). When tweets cannot flow through user 1, there is only 1 reachable follower of user 3: user 6 ($6 \rightarrow 3$). Thus, NCR^D_{31} is $\frac{1}{4}$. NCR^G_{31} can be calculated in the same manner except that we consider the number of reachable followees.

4) Information Flow Efficiency.: Information flow efficiency measures how independent a user is in propagating and gathering information efficiently in his friend's local network. Network study has come up with a measure to measure efficiency, that is, closeness centrality. High closeness centrality translates to the minimum amount of time in spreading messages[22]. Therefore, based on closeness centrality, we develop a measure that we call information flow efficiency to investigate the influence of offline friendship on propagating and gathering information efficiently on Twitter. The formula of closeness centrality that we employ is the one that uses the harmonic shortest distance[18].

$$C_i = \frac{1}{n-1} \sum_{k \neq i} \frac{1}{d_{ik}} \quad (6)$$

C_i is the closeness centrality of user i and d_{ik} is the shortest distance from user i to user k. With Equation 6 as our basis, the formula for information flow efficiency in distributing information is:

$$IFE^D_{ij} = \frac{C_i \text{ in distributing tweets to user } j\text{'s alters in user } j\text{'s local network, when tweets cannot flow through user } j}{C_i \text{ in distributing tweets to user } j\text{'s alters in user } j\text{'s local network}} \quad (7)$$

The formula above calculates how independent user i is on user j in gathering information efficiently in user j's local network. When user i is completely independent from user j, the value of IFE^D_{ij} is 1. On the other hand, if user i is completely dependent on user j, the value is 0.

Meanwhile, the formula for information flow efficiency in gathering information is:

$$IFE^G_{ij} = \frac{C_i \text{ in gathering tweets from user } j\text{'s alters in user } j\text{'s local network, when tweets cannot flow through user } j}{C_i \text{ in gathering tweets from user } j\text{'s alters in user } j\text{'s local network}} \quad (8)$$

The formula above calculates how independent user i is on user j in gathering information efficiently in user j's local network. When user i is completely independent from user j, the value of IFE^G_{ij} is 1. On the other hand, if user i is completely dependent on user j, the value is 0.

Again, we assume that the more independent user i is from user j in distributing or gathering information efficiently to or from user j's friends, the more likely user i is in posting or gathering information to or from user j's friends because (1) it is easier for user i to do so, (2) the information coming to or from user j's friends is free from the influence of user j.

In the following example on how to calculate IFE^D_{31}, an arrow represents a following link that goes against the information flow. In Figure 1, the shortest distances of user 3 in distributing tweets to the alters in user 1's local network are: 2 steps to user 4 ($4 \rightarrow 1 \rightarrow 3$), 2 steps to user 2 ($2 \rightarrow 1 \rightarrow 3$), 2 steps to user 8 ($8 \rightarrow 1 \rightarrow 3$), and 1 step to user 6 ($6 \rightarrow 3$). Therefore, the denominator is $(3 \times \frac{1}{2} + \frac{1}{1})/(n-1) = 2.5/(n-1)$. When tweets cannot flow through user 1, user 3 can only distribute tweets to user 6 and the shortest distance is 1. Thus, the nominator is $1/(n-1)$. Therefore, IFE^D_{31} is $1/2.5 = \frac{2}{5}$. IFE^G_{31} can be calculated in the same manner, except that we consider the shortest distances in gathering tweets.

5) Edge Betweenness Centrality.: Edge betweenness centrality measures how important a communication between two users is in facilitating all communications in a network. Technically, it measures the extent to which a link in a communication network falls on the shortest path between pairs of other points[23]. In Twitter, a communication link is represented by a following link. Since a following link represents one-to-many instead of one-to-one conversation, in order for communication to flow between two users, a user either has to pay attention to his friend's tweets, or his friend has to post tweets that are of interest to the user. Through the edge betweenness centrality, we investigate the influence of offline friendship on facilitating communication on Twitter, either by paying attention to a friend's tweets or posting tweets that are of interest to a friend.

Mathematically, let $\sigma_{st}(e_{ij})$ be 1 if the communication link from user i to j lies on the shortest path from s to t and 0 if it does not or if there is no such path. Let G_j be the local network of user j, and n be the number of users in G_j. The formula for EBC^D_{ij}, the edge betweenness centrality of the communication link from i to j (i distributes information to j) is:

$$EBC^D_{ij} = \frac{\sum_{st:s,t \in G_j} \frac{\sigma_{st}(e_{ij})}{\sigma_{st}}}{n \times (n-1)} \tag{9}$$

The formula for EBC^G_{ij}, the edge betweenness centrality of the communication link from j to i (i gathers information from j), is:

$$EBC^G_{ij} = \frac{\sum_{st:s,t \in G_j} \frac{\sigma_{st}(e_{ji})}{\sigma_{st}}}{n \times (n-1)} \tag{10}$$

In the following example, we want to calculate EBC^D_{61}, the edge betweenness centrality of the communication link from user 1 to 6 ($6 \rightarrow 1$ in Figure 1). In Figure 1, an arrow represents a following link that goes against the information flow. To calculate the edge betweenness centrality, we first have to calculate the number of shortest paths between all pairs of users. First, let's consider the number of shortest paths between user 6 and all other users. The number of shortest paths from user 6 to all other users are: 1 path to user 1 ($6 \rightarrow 1$), 2 paths to user 2 ($6 \rightarrow 1 \rightarrow 4 \rightarrow 2$,$6 \rightarrow 3 \rightarrow 4 \rightarrow 2$), 1 path to user 3 ($6 \rightarrow 3$), 2 paths to user 4 ($6 \rightarrow 3 \rightarrow 4$,$6 \rightarrow 1 \rightarrow 4$), 1 path to user 5 ($6 \rightarrow 1 \rightarrow 5$), 1 path to user 7 ($6 \rightarrow 1 \rightarrow 7$), and 1 path to user 8 ($6 \rightarrow 1 \rightarrow 8$). Out of these shortest paths, the number of the shortest paths that use the communication link from user 1 to user 6 ($6 \rightarrow 1$) are: 1 path to user 1 ($6 \rightarrow 1$), 1 path to user 2 ($6 \rightarrow 1 \rightarrow 4 \rightarrow 2$), 0 path to user 3, 1 path to user 4 ($6 \rightarrow 1 \rightarrow 4$), 1 path to user 5 ($6 \rightarrow 1 \rightarrow 5$), 1 path to user 7 ($6 \rightarrow 1 \rightarrow 7$), and 1 path to user 8 ($6 \rightarrow 1 \rightarrow 8$). Therefore, we can increase the nominator of EBC^D_{61} by $\frac{1}{1} + \frac{1}{2} + \frac{0}{1} + \frac{1}{2} + \frac{1}{1} + \frac{1}{1} + \frac{1}{1} = 5$. To complete the calculation, we keep increasing the nominator by considering the shortest paths between the other pairs of users in the network in a similar manner that we did for user 6. Dividing the value by $8 \times 7 = 56$, we will get EBC^D_{61}. EBC^G_{61} can be calculated in the same manner, except that we consider the communication link from user 6 to user 1 ($1 \rightarrow 6$).

B. Content Similarity

We measure the similarity of tweets between two users to investigate the influence of offline friendship on tweeting behavior, specifically how likely a user is in posting tweets similar to his friend's.

Using Latent Dirichlet Allocation(LDA)[24], we generate a topic distribution of a user's tweets. We calculate the similarity of tweets between two users by the following formula.

$$Similarity_{ij} = \frac{1}{D_{KL}(i \mid\mid j)} + \frac{1}{D_{KL}(j \mid\mid i)} \tag{11}$$

$D_{KL}(i \mid\mid j)$ is the Kullback-Leibler divergence[25] of the topic distribution from user i's tweets to user j's tweets. $D_{KL}(j \mid\mid i)$ is the Kullback-Leibler divergence of the topic distribution from user j's tweets to user i's tweets.

C. Interaction on Twitter

Twitter, as an online social network provides a platform for users to interact with one another. There are four types of interaction on Twitter: *favorite*, *retweet*, *reply*, and *mention*. *Favorite* is liking a friend's status. *Retweet* is reposting a friend's tweets. *Reply* is replying to a friend's tweets. *Mention* is mentioning a friend in tweets. In this study, we investigate the influence of offline friendship on interacting on Twitter through the following variables.

$$Favorite_{ij} = \text{\# user } j\text{'s tweets that user } i \text{ likes.} \tag{12}$$

$$Retweet_{ij} = \text{\# user } j\text{'s tweets that user } i \text{ retweets.} \tag{13}$$

$$Reply_{ij} = \text{\# user } i\text{'s replies to user } j\text{'s tweets} \tag{14}$$

$$Mention_{ij} = \text{\# user } j\text{'s names in user } i\text{'s tweets.} \tag{15}$$

V. Regression Model Development

We aim to assess whether offline friendship affects behavior measures we developed in this study. We regress a number of possible outcome variables on whether the two users are offline friends.

$$Outcome_{ij} = \beta_0 + \beta_1 RelationshipType_{ij} + \zeta_j + \mu_{ij} \tag{16}$$

The data structure for all the models is a cross-sectional data. The outcome variables include variables pertaining to (a) local network properties, (b) tweets, and (c) interaction on Twitter. We aim to explain the change in $Outcome_{ij}$ with respect to the relationship type: Online versus Offline.

Table I
VARIABLES AND OPERATIONAL DEFINITIONS

Category	Variable	Operational Definitions	Twitter Networking Behaviors Observed
	j	local user index	
	i	alter index	
	$RelationType_{ij}$	the relation type between user i and user j (offline or online)	
Network Structure	1. $Reciprocity_{ij}$	whether the link from user i to user j is reciprocated	responding to online gesture of friendship
	2. $FollowerOverlap_{ij}$, 3. $FolloweeOverlap_{ij}$	followers/followees overlap between user i and user j	forming online friendships with whom one shares mutual friends
	4. NCR^D_{ij} 5. NCR^G_{ij}	user i's network coverage ratio in distributing or gathering information in user j local's network	independence in distributing/gathering information
	6. IFE^D_{ij} 7. IFE^G_{ij}	user i's efficiency in distributing or gathering information in user j's local network	independence in distributing/gathering information efficiently
	8. EBC^D_{ij} 9. EBC^G_{ij}	edge betweenness centrality of information distribution or gathering link between user i and user j in user j's local network	facilitating communication by: (a)posting tweets interesting to a friend, or (b)paying attention to a friend's tweets
Twitter Content Similarity	10. $Similarity_{ij}$	content similarity between user i and user j	posting tweets similar to a friend's
Interactions on Twitter	11. $Favorite_{ij}$	the number of user j's tweets that user i likes	liking a friend's tweets
	12. $Retweet_{ij}$	the number of user j's tweets that user i retweets	retweeting a friend's tweets
	13. $Reply_{ij}$	the number of user i's replies to user j's tweets	replying a friend's tweets
	14. $Mention_{ij}$	the number of user j's names in user i's tweets	mentioning a friend in tweets
D in superscript represents the **D**istribution flow of tweet. G in superscript represents the **G**athering flow of tweet.			

We adopt a fixed effects model ζ_j to control for user j's network structure heterogeneity. The error component, μ_{ij} is an idiosyncratic error term and it varies across i and j. We also assess the correlation between the variables to see whether there are any interesting relationships between the variables.

VI. RESULTS AND DISCUSSION

Table II presents the results of the regression, while Table III presents the results of the correlation.

Reciprocity entails responding to other's gestures of friendship with similar gestures. In Twitter, the gesture of friendship is represented by following a user. When a user follows a friend, he shows an interest in his friend's tweets. Reciprocity is one of the building blocks of a social network[19]. The coefficient of $RelationType_{ij}$ (0.354 $p < 0.001$) for $Reciprocity_{ij}$ shows that offline friendship increases a user's likelihood to respond to another user's gesture of friendship online. The result implies that in the online social world, the percentage of offline friendship might influence the pace of an online social network formation. The lack of responding in kind to the gesture of friendship from an online friend might stop the formation of an online social network. A person who follows you but does not receive a following back might eventually lose interest in you. On the other hand, following a person without receiving a follow back will eventually tire you out of the one-sided relationship. Of course, as Twitter is also a news media[15], one-sided relationships are common as the intention of making a connection online in this type of relationship is to receive news and not to create a friendship. However, a network mainly infused with such a relationship can hardly be called a social network. This network will appear more structurally similar to a news network. Therefore, the formation of a Twitter network as a social network may be driven by the percentage of the offline friends in the network.

Another variable that is closely related to one of the fundamental principles of social network formation is the follower overlap and the followee overlap. Mutual friends encourage a principle of network formation called triadic closure. On the other hand, triadic closures increase the number of mutual friends. Triadic closure means two people become friends because they share a mutual friend. There are two mechanisms leading to a triadic closure. First, the increased propinquity of individuals who share a mutual friend[19]. It is not hard to see that the first mechanism applies more to offline friends. Second, the psychological

Table II
FIXED REGRESSION RESULT

$Outcome_{ij}$	β_1	Standard Error	# of Observations	Within R^2
1. $Reciprocity_{ij}$	0.35381^{***}	0.01877	2193	0.1414
2. $FollowerOverlap_{ij}$	0.02211^{***}	0.00144	2193	0.0987
3. $FolloweeOverlap_{ij}$	0.02129^{***}	0.00174	2193	0.0652
4. NCR^D_{ij}	0.18919^{***}	0.01496	2193	0.1663
5. NCR^G_{ij}	0.11289^{***}	0.01514	2193	0.1127
6. IFE^D_{ij}	0.16002^{***}	0.01274	2193	0.2298
7. IFE^G_{ij}	0.06728^{***}	0.01344	2193	0.1396
8. EBC^D_{ij}	0.00083^{***}	0.00010	2193	0.0379
9. EBC^G_{ij}	0.00070^{***}	0.00018	2193	0.0098
10. $Similarity_{ij}$	0.26301^{***}	0.02013	2193	0.0755
11. $Favorite_{ij}$	0.00704	0.00561	2193	0.001
12. $Retweet_{ij}$	0.14519^{*}	0.04817	2193	0.0048
13. $Reply_{ij}$	1.20393^{***}	0.26508	2193	0.0102
14. $Mention_{ij}$	0.38023^{***}	0.09533	2193	0.0075

*significant at p<0.1 **significant at p<0.01 ***significant at p<0.001

Table III
CORRELATIONS BETWEEN THE VARIABLES

Each number represents a variable corresponding to the variable numbered as such in Table I and Table II

Variables	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	1.0000													
2	0.1594	1.0000												
3	0.1654	0.7520	1.0000											
4	0.1203	0.3221	0.3569	1.0000										
5	0.0545	0.2150	0.3119	0.5986	1.0000									
6	0.0381	0.4383	0.4576	0.9461	0.5966	1.0000								
7	-0.0238	0.3143	0.4257	0.6291	0.9432	0.6689	1.0000							
8	0.3389	0.0859	0.0369	-0.0818	-0.0969	-0.1196	-0.0609	1.0000						
9	0.1832	0.0575	0.0070	-0.1184	-0.1844	-0.0842	-0.1953	0.2766	1.0000					
10	0.2101	0.2317	0.2225	0.2333	0.1986	0.2181	0.2076	0.2119	0.0754	1.0000				
11	0.0231	0.0306	0.0163	0.0062	0.0018	0.0114	0.0093	0.0083	0.0014	0.0227	1.0000			
12	0.1319	0.0265	0.0233	0.0270	0.0198	0.0055	-0.0017	0.0436	0.0150	0.1766	0.0091	1.0000		
13	0.1462	0.0517	0.0440	0.0081	0.0045	-0.0048	-0.0078	0.0897	0.0426	0.2101	-0.0043	0.3880	1.0000	
14	0.1310	0.1068	0.0613	0.0150	0.0002	0.0144	-0.0020	0.0913	0.0565	0.1893	-0.0060	0.4090	0.7536	1.0000

need to balance two friends' evaluation on a third party. Such a psychological need requires one another to interact frequently enough[19]. Again, the second mechanism is also more likely to happen between two offline friends. It is, therefore, unsurprising that the result of the regression shows that offline friendship increases one's likelihood to form an online friendship with whom one shares mutual friends online (β_1 0.022 $p < 0.001$ for $FollowerOverlap_{ij}$, β_1 0.021 $p < 0.001$ for $FolloweeOverlap_{ij}$). As triadic closure is also one fundamental principles of social network formation, this result also implies that face-to-face interaction is important for the formation of a Twitter network as a social network. The correlation results show that in Twitter, follower and followee overlap are strongly correlated. Therefore, if a user and his friend have a lot of mutual followers, they are also likely to have a lot of mutual followees.

The coefficients of the regression result for NCR^D_{ij} and NCR^G_{ij} show that offline friendship increases a user's independence to distribute or gather information in his friend's Twitter local network. The independence from his friend to distribute or gather information makes it easier for a user to do so in his friend's local network. Therefore, a user is more likely to distribute or gather information in his friend's Twitter local network if he establishes an offline friendship with his friend.The ramification of information propagation does not only stop at the level of knowledge. Information brings with it influence on one's opinion and emotion. Even a simple information such as price and store name can influence a buyer's product perception[26]. As such, the implication of the results is, in Twitter, a user influences more friends of an offline friend and receives influence from more friends of an offline friend. As receiving and giving influence are motivated by the wish to conform to social demands[27], these findings suggest that Twitter communication is one of the many ways to meet real-life social demands. Through Twitter, one can get an up-to-date information about what's happening and what's popular in one's friend's community. In the regression results for $Reciprocity_{ij}$ and $Follower/FolloweeOverlap_{ij}$ we have seen the importance of face-to-face interactions for the formation of an online social network. In these results, we glimpse into the importance of online interaction to strengthen a relationship offline. In summary, the results so far show that offline and online interactions are complementing each other, instead of substituting each other.

Unlike NCR_{ij}^{D} and NCR_{ij}^{G}, the variables IFE_{ij}^{D} and IFE_{ij}^{G} emphasize the speed at which a message propagates. The results of the regression for these two variables show that offline friendship increases a user's independence to distribute or gather information efficiently in his friend's Twitter local network. Again, the independence from his friend to distribute or gather information efficiently encourages a user to distribute or gather information in his friend's local network.

The correlation matrix shows that follower overlap is moderately correlated with the independence at which one distributes information efficiently in a friend's network. Meanwhile, followee overlap is moderately correlated with the independence at which one gathers information efficiently in a friend's network.

The correlation matrix also shows that NCR_{ij}^{D} is very strongly correlated with IFE_{ij}^{D} (corr: 0.9461), while NCR_{ij}^{G} is very strongly correlated with IFE_{ij}^{G} (corr: 0.9432). It is not surprising to see such a strong correlation since the concept of information flow efficiency is closely related to the concept of network coverage ratio. In network coverage ratio, we consider the number of reachable users. In information flow efficiency, we do not consider only the number of reachable users, but also the distance of the reachable users. The rest of the combination of these 4 variables, namely NCR_{ij}^{D}-NCR_{ij}^{G}, NCR_{ij}^{D}-IFE_{ij}^{G}, NCR_{ij}^{G}-IFE_{ij}^{D}, and $IFE_{ij}^{D} - IFE_{ij}^{G}$ are moderately correlated (corr: 0.4-0.7). The correlations imply that in Twitter, influence is seemingly two-way. If one is likely to give influence, then one is also likely to receive influence. However, in this study, we have removed users who are likely to be famous people, news media, business media or spammers (users with #followers $>$ 1000). These users are more likely to give influence than to receive influence.

The regression results for EBC_{ij}^{D} and EBC_{ij}^{G} show that offline friendship increases the importance of the online communication between two friends to facilitate all communications in a Twitter local network. Online communication, represented by the following link, means the act of giving and receiving information. It does not necessarily entail exclusive communication between two people because Twitter is after all, a one-to-many communication platform. However, it entails one posting a topic that might be of interest to another, or one pays special attention to another's post. By virtue of their importance to facilitate all communications, a user and his offline friend may pay special attention to one another's tweets or post tweets that are of interest to one another due to the online peer pressure. Of course the peer pressure to do so only happens if a Twitter local network is an active network where its users frequently communicate and actively receive and propagate information.

In terms of Twitter content, when $Similarity_{ij}$ is regressed on $RelationType_{ij}$, the resulting coefficient reports a greater tweets similarity between offline friends (β_1 0.26 $p < 0.001$). We do not know whether this result is due to homophily — offline friends are sharing similar interest — or due to social influence — offline friends are influencing each other. It could be due to both. We only know that offline friendship increases one's tendency to post tweets similar to one's friend. Although intuitively online friends connect due to similar interest, this shared interest between online friends apparently does not translate into a relatively greater tweets similarity. There could be several reasons for this. It could be that a user pays more attention to the tweets of his offline friends, or he receives more influence from his offline friends.

The regression results for $Retweet_{ij}$, $Reply_{ij}$, $Mention_{ij}$ show that offline friendship increases one's frequency to reply to a friend, to mention a friend, and to retweet a friend. Only $Favorite_{ij}$ — liking a friend's status — is not influenced by offline friendship. $Reply_{ij}$ is strongly correlated with $Mention_{ij}$, meaning, if one replies a lot to a friend, one also mentions the friend in his tweets often. These results imply that offline interaction may propel online interaction. This implication is different from the result of the previous research on Event-Based Social Network that showed offline interaction did not translate into a greater online interaction[13]. The difference may be due to the functional differences between an Event-Based Social Network and Twitter. An event-based social network consists mainly of offline acquaintances. People who engage in an event-based social network have similar interests or participate in similar events. They meet several times, but friendship does not develop further. In an event-based social network, people rarely post about their life in general. Communication is mainly about the next upcoming events or social events just attended. Communication on Twitter is different from communication on an event-based social network. While Twitter has been shown to be mostly used in a passive way (i.e., reading or following)[28] or as a news media[15], many tweets are phatic in nature[29], serving to maintain social bonds[30]. Therefore, compared to an event-based social network, a Twitter network may consist of closer friends.

The regression result for $Reply_{ij}$ especially has an interesting impact on the interpretation of the previous studies on social network. These previous studies validated existing social network theories in the online world. The first study proved that the layered structure corresponding to the frequency of interaction in the offline world also exists in the online word, specifically Facebook and Twitter[3]. The second study showed that the Dunbar's number also applied on Twitter[4]. In the first study, the users in the network analysis were the users who replied each other. In the second study, interaction strength was quantified by the number of replies. In our study, we have seen that users who reply each other are more likely to be offline friends. Therefore, it is very likely that the theories of social network that the previous studies have validated in the online social network only apply to offline friends. Future social network studies can investigate whether this is indeed the case.

VII. Conclusion

In the digital world we live in, along with the proliferation of online social networking sites, the number of online friends surges and the presence of online friends is gaining more importance. Nevertheless, through our examination of Twitter network, we conclude that offline friendship that provides face-to-face interaction still plays an important role even in online communication. Overall, our results show that offline friendship propels online networking activities on Twitter.

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).

References

[1] P. Quintaro. (2015, Apr.) Twitter mau were 302m for q1, up 18% yoy. [Online]. Available: http://www.benzinga.com/news/earnings/15/04/5452400/twitter-mau-were-302m-for-q1-up-18-yoy#ixzz3aOd1VDpf

[2] D. M. Boyd and N. B. Ellison, "Social network sites: Definition, history, and scholarship," *Journal of Computer-Mediated Communication*, vol. 13, no. 1, pp. 210–230, October 2007.

[3] R. Dunbar, V. Arnaboldi, M. Conti, and A. Passarella, "The structure of online social networks mirrors those in the offline world," *Social Networks*, vol. 43, pp. 39–47, 2015.

[4] B. Gonçalves, N. Perra, and A. Vespignani, "Validation of dunbars number in twitter conversations," *PLoS ONE*, vol. 6, no. 8, 2011.

[5] D. K. S. Chan and G. H. L. Cheng, "A comparison of offline and online friendship qualities at different stages of relationship development," *Journal of Social and Personal Relationships*, vol. 21, no. 3, pp. 305–320, 2004.

[6] M. L. Antheunis, P. M. Valkenburg, and J. Peter, "The quality of online, offline, and mixed-mode friendships among users of a social networking site," *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, vol. 6, no. 3, 2012.

[7] V. M. Buote, E. Wood, and M. Pratt, "Exploring similarities and differences between online and offline friendships: The role of attachment style," *Computers in Human Behavior*, vol. 25, no. 2, pp. 560–567, 2009.

[8] N. B. Ellison, C. Steinfield, and C. Lampe, "The benefits of facebook "friends:" social capital and college students use of online social network sites," *Journal of Computer-Mediated Communication*, vol. 12, p. 11431168, 2007.

[9] R. Heatherly, M. Kantarcioglu, and B. Thuraisingham, "Preventing private information inference attacks on social networks," *Knowledge and Data Engineering*, vol. 25, no. 8, pp. 1849–1862, 2012.

[10] L. Dai, J. der Luo, X. Fu, and Z. Li, "Predicting offline behaviors from online features - an ego-centric dynamical network approach," in *ACM HotSocial'12*. ACM, 2012.

[11] W. Xie, C. Li, F. Zhu, E. Lim, and X. Gong, "When a friend in twitter is a friend in life," in *ACM WebSci 2012*. ACM, 2012.

[12] L. Backstrom and J. Kleinberg, "Romantic partnership and the dispersion of social ties: A network analysis of relationship status on facebook," in *Proc. 17th ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW)*. ACM, 2014.

[13] P. Yin, Q. He, X. Liu, and W. Lee, "It takes two to tango: Exploring social tie development with both online and offline interactions," in *SIAM International Conference on Data Mining 2014*. SIAM, 2014.

[14] X. Liu, Q. He, Y. Tian, W.-C. Lee, J. McPherson, and J. Han, "Event-based social networks: Linking the online and offline social worlds," in *ACM Conference on Knowledge Discovery and Data Mining (KDD 2012)*. ACM, 2012.

[15] H. Kwak, C. Lee, H. Park, and S. Moon, "What is twitter, a social network or a news media," in *International World Wide Web Conference (WWW 2010)*, 2010.

[16] N. Zipkin. (2013, Dec.) Have 1,000 followers? you're in the 96th percentile of twitter users. [Online]. Available: http://www.entrepreneur.com/article/230487

[17] S. Ghosh, B. Viswanath, F. Kooti, N. K. Sharma, G. Korlam, F. Benevenuto, N. Ganguly, and K. P. Gummadi, "Understanding and combating link farming in the twitter social network," in *WWW '12 Proceedings of the 21st international conference on World Wide Web*. ACM, 2012, pp. 61–70.

[18] M. Newman, *Networks: An Introduction*. New York: Oxford, 2010.

[19] R. A. F. L. D. H. C. L. M. David R. Schaefer, John M. Light, "Fundamental principles of network formation among preschool children," *Social Networks*, vol. 32, pp. 61–71, 2010.

[20] D. Easley and J. Kleinberg, *Networks, Crowds, and Markets: Reasoning about a Highly Connected World*. Cambridge University Press, 2010.

[21] S. Ye and S. F. Wu, "Measuring message propagation and social influence on twitter.com," in *The International Conference on Social Informatics (SocInfo) 2010*. Springer-Verlag Berlin Heidelberg, 2010, pp. 216–231.

[22] A. Bavelas, "Communication patterns in task oriented groups," *Journal of the Acoustical Society of America*, vol. 22, pp. 271–282, 1950.

[23] M. Girvan and M. Newman, "Community structure in social and biological networks," *PNAS*, vol. 99, pp. 7821–7826, 2002.

[24] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of Machine Learning Research*, vol. 3, pp. 993–1022, 2003.

[25] S. Kullback and R. A. Leibler, "On information and sufficiency," *Annals of mathematical statistics*, vol. 22, pp. 79–86, 1951.

[26] W. B. Dodds, "In search of value: How price and store name information influence buyers product perceptions," *Journal of Services Marketing*, vol. 5, no. 3, pp. 27–36, 1991.

[27] R. B. Cialdini and N. J. Goldstein, "Social influence: Compliance and conformity," *Annual Review of Psychology*, vol. 55, pp. 591–613, 2004.

[28] V. A. Matthias Hofer, "Perceived bridging and bonding social capital on twitter: Differentiating between followers and followees," *Computers in Human Behavior*, vol. 29, pp. 2134–2142, 2013.

[29] V. Miller, "New media networking and phatic culture," *Convergence*, vol. 4, pp. 387–400, 2008.

[30] K. Crawford, "These foolish things: On intimacy and insignificance in mobile media," in *Mobile Technologies: From Telecommunications to Media*. New York: Routledge, 2009.