Quantifying Reciprocity in Social Networks

Huiqi Zhang, Ram Dantu
Department of Computer Science and Engineering
University of North Texas
Denton, TX 76203 US
huiqizhang@my.unt.edu, rdantu@unt.edu,

Joao Cangussu
Department of Computer Science
University of Texas at Dallas
Richardson, TX, 75081
cangussu@utdallas.edu

***Abstract*- In this paper we propose a new reciprocity index for quantifying social relationships based on mobile phone call detail records and Twitter blogs. We use this reciprocity index to measure the level of reciprocity between users. This work is useful for detecting unwanted calls (e.g., spam) and product marketing. For validation of our results, we used actual call logs of 100 users collected at MIT by the Reality Mining Project group for a period of 8 months and Twitter blogs of 460 users collected by the Network Security team at UNT for a period of 12 months. The experimental results show that our model achieves results with high accuracy.**

I. Introduction

In social networks, one of the important relationships between people is reciprocity. Reciprocity can be defined as the action of returning of similar acts [1, 2]. In this study, our interest is to investigate how people utilize technology to construct their social relationships. We focus on the measure of mediated interactions considering the media used to interact. To investigate how people interactively construct their social relationships, we focus on the reciprocity of actions that take place in a social media environment.

Reciprocity plays an important role in economic and social relations. For example, in marketing, sellers sell products to buyers. Buyers receive products and sellers earn money. Furthermore, buyers give the sellers feedback. By the buyers' feedback, the sellers improve their product quality and service. The buyers receive good products and service, and sellers earn more money. Therefore, the reciprocity relation is one of the keys to business success. Similarly, we may enhance detecting unwanted calls (e.g., spam) by reciprocity analysis. For example, the spammers definitely do not receive responses from us. If we are not sure whether the incoming calls come from spammers or not, the system will not let the phone ring and forward the calls to the voice box automatically.

Most social relationship research focuses on the collection of dyads in social networks [1]. In this paper, we propose a new reciprocity index that is different from the previous work on measurements. The experimental results show that our model achieves accurate results.

In Section 2, we briefly review the related work. In Section 3 we describe the model and methodology for a new reciprocity index for quantifying relationships. We performed the experiments with actual call logs and micro blogs, and we discuss the results in Section 4. We describe the validation of our model, conducted by the actual call logs and micro blogs, in Section 5. Finally, we have the conclusions in Section 6.

II. Related Work

A social network is defined as a set of actors (individuals) and the ties (relationships) among them [1]. The study of social networks has been applied in modern sociological studies for some time. The major applications focus on measuring interpersonal relations in groups, describing properties of social structures and individual social environments, etc [1]. There are two fundamental interests in social networks: the relational ties and the actors. In [3] the authors propose an index of mutuality to measure tendency toward mutuality by the probability of a mutual choice between two actors. In [4] the authors propose an index to measure the tendency for mutuality, which compares the observed number of mutual connections to the number expected if choices were randomly made. The formulas for the mean and variance of the number of mutual connections are given. The observed number of mutual connections is then compared to the expected number and a z-score is calculated. In [5] the author finds that if he includes the effect of the reciprocity and the scaling exponent, which are negatively correlated in simulations of growing network, the degree distributions are much closer to those empirically observed. In [6] the authors propose a framework in which the occurrences of mutual links depend on conditional connection probabilities according to their actual degree of correlation between mutual links. In [7] the authors find that the 1-node and 2-node degree correlations are very important to reciprocity in real networks, and the level of correlation contributions to the reciprocity depends on the type of correlations involved. In [8] the authors investigate the lattice reciprocity mechanisms and interpret the onset of lattice reciprocity as a thermodynamic phase transition to enhance evolutionary survival of the cooperative phenomena in social networks. In [9] the author discusses the strength of social relations between two persons, measured by an email conversation. The relationship is strong if email between two persons is exchanged frequently, recently, and reciprocally, and a formula is used for the strength, which is a function of user-determined importance weights and the number of received and sent emails. In [10] the authors find that reciprocity in email behavior is different between multi-recipient and dyadic mail.

Our approach is different from the above work. We observe that the structure and transactions in reciprocity are different compared to face-to-face interactions. *The existing approaches measure the tendency of mutual choices for actors (nodes) in a graph. They do not deal with frequency and duration of real time electronic communications between two actors. In real life, the frequency of communication plays an important role for the relationship between persons. To the best of our knowledge no similar work has been reported. We propose a new reciprocity index based on mobile phone call detail records and Twitter blogs.*

III. Model and Methodology

A. Dyads and reciprocity index

The dyadic relationship in a social network is the collection of dyads. A dyad is an unordered pair of nodes (actors) and arcs (ties) between the two nodes. There are the (n × (n-1))/2 dyads in a directed graph with n nodes. A dyad is mutual if both the tie from i to j and the tie from j to i are present. Each of the dyads in the network is assigned to one of three types: mutual (actor i has a tie to actor j and actor j has a tie to actor i), asymmetric (either i has a tie to j or j has a tie to i, but not both), or null (neither the i to j tie nor the j to i tie is present). These are often labeled M, A, and N respectively. The dyad census gives the frequencies of these types.

In [3] the authors propose the index of mutuality, ρ_{kp}. This index focuses on the probability of a mutual choice between two actors i and j:

$$P(i \to j \,\&\, j \to i) = P(i \to j) \times P(j \to i \mid i \to j)$$

The $P(j \to i \mid i \to j)$ can be considered as consisting of two parts: the $P(j \to i)$ and a fraction, denoted by ρ_{kp} of the probability $P(j \xrightarrow{not} i)$ [3]. The ρ_{kp} is 0 if there is no tendency toward mutuality and 1 if there is a maximal tendency toward mutuality. A negative value of index indicates a tendency away from mutuality, toward asymmetry and nulls, referred to as antireciprocation. There are two kinds of ρ_{kp}, fixed choice and free choice. Fixed choice assumes that all actors make the same number of choices, and the estimate of ρ_{kp}^{fixed} is computed by [3]

$$\hat{\rho}_{kp}^{fixed} = \frac{2(n-1)M - nc^2}{nc(n-1-c)}$$

where n is the number of nodes, M is the observed number of mutualities, and c is the number of choices.

Free choice allows different numbers of choices and the estimate of ρ_{kp}^{free} is computed by [3]

$$\hat{\rho}_{kp}^{free} = \frac{2(n-1)^2 M - S^2 + S_2}{S(n-1)^2 - S^2 + S_2}$$

where n is the number of nodes, M is the observed number of mutual connections, $S = \sum x_{i+}$ is the total number of choices and $S = \sum x_{i+}^2$ is the sum of squares of the choices made by each actor.

In the mobile phone social networks, actor i and actor j may call each other multiple times, and the reciprocity reflects their relationship in a period of time. The above mutual index and other existing mutual indices cannot measure this kind of relationship. The existing mutual (reciprocity) indices measure the tendency of mutual choices for actors (nodes) in a graph. They do not deal with frequency of communication. We propose a reciprocity index, $\rho_{a \leftrightarrow b}$ to measure the tendency of reciprocity for actors *a* and *b* in a group.

Fig.1 shows the reciprocity relation between phone user29 and his communication partner 349 where m, h, d and numbers inside the boxes above the arrows indicate minute, hour, day, and call duration respectively.

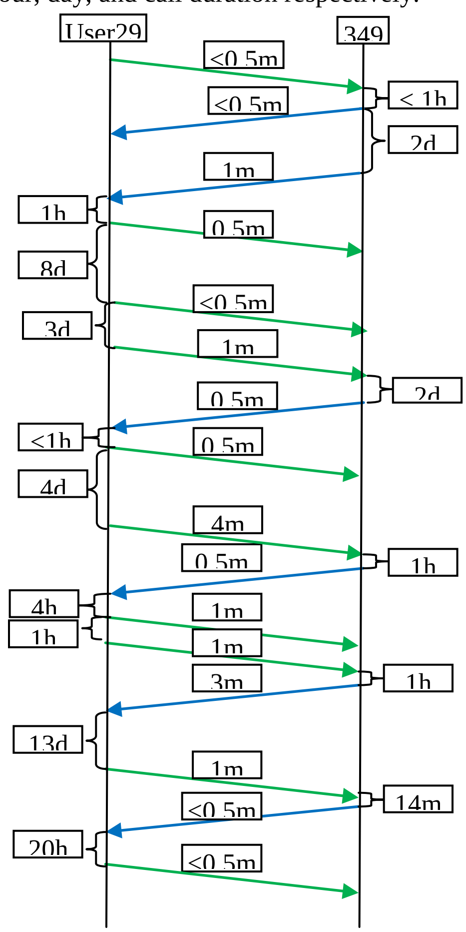

Fig. 1 The event flow chart for phone user29 with his partner 349.

Suppose that the number of phone calls arriving is a Poisson process. Then the probability of no arrivals in the interval [0, t] is given by

$$P(\tau > t) = e^{-\lambda t}$$

where λ is the arrival rate and τ is interarrival time. The occurrence of at least one arrival between 0 and t is given by

$$P(\tau \le t) = 1 - e^{-\lambda t}$$

Considering actor *a* calls actor *b* at time t_i with rate $\lambda_a t$, the probability of actor *b* calling actor *a* back (reciprocity) at a time t_j with rate $\lambda_b t$ can be computed by

$$P(a \to b \,\&\, b \to a) = P(a \to b)P(b \to a \mid a \to b)$$
$$= P(a \to b)[P(b \to a) + \rho_{a \leftrightarrow b} P(b \xrightarrow{not} a)]$$
$$= (1 - e^{-\lambda_a t_i})[(1 - e^{-\lambda_b (t_j - t_i)}) + \rho_{a \leftrightarrow b} e^{-\lambda_b (t_j - t_i)}]$$

The expected value, $E(R \mid \rho_{a \leftrightarrow b})$, of number of reciprocity from *b* to *a* is the total number of calls, S, from *a* to *b* times this probability, i. e.

$$E(R \mid \rho_{a \leftrightarrow b}) = S(1 - e^{-\lambda_a t_i})[(1 - e^{-\lambda_b (t_j - t_i)}) + \rho_{a \leftrightarrow b} e^{-\lambda_b (t_j - t_i)}]$$

After rearranging the terms, we have

$$\rho_{a \leftrightarrow b} = [R - S(1 - e^{-\lambda_a t_i})(1 - e^{-\lambda_b (t_j - t_i)})] / S(1 - e^{-\lambda_a t_i}) e^{-\lambda_b (t_j - t_i)} \quad (1)$$

where R is observed number of reciprocity.

The $\rho_{a \leftrightarrow b}$ is 0 if there is no tendency toward reciprocity and 1 if there is a maximal tendency toward reciprocity.

B. Real-life data sets and parameters

***Real-life traffic profile*:** In this paper, actual call logs and Twitter blogs are used for analysis. These actual call logs were collected at MIT [13] by the Reality Mining Project group for a period of 8 months. Twitter blogs of 460 users were collected by the Network Security team at UNT for a period of 12 months.

The Reality Mining Project group collected mobile phone usage of 100 users, including user ID (unique number representing a mobile phone user), time of call, call direction (incoming and outgoing), incoming call description (missed, accepted), talk time, and tower ID (location of phone user). These 100 phone users are students, professors and staff members. The collection of the call logs was followed by a survey of feedback from participating phone users for behavior patterns such as favorite hangout places; service providers; talk time minutes; and phone users' friends, relatives and parents. We used this extensive dataset for our reciprocity analysis and validation of 10 sample users in this paper. More information about the Reality Mining Project can be found in [13].

Time of day: Everyone has his/her own schedule for working, studying, entertainment, sleeping, traveling and so on. The schedule is mainly based on the time of the day and day of the week.

Call frequency: The call frequency is the number of incoming or outgoing calls in a period of time. The greater the number of incoming or outgoing calls in a period of time, the more socially close the caller and callee relationship.

Call duration: The call duration is how long both caller and callee want to talk to each other. The longer the call duration is in a period of time, the more socially close the caller and callee relationship.

Reciprocity: Reciprocity represents the response by one party to calls from another party.

We used the call log data from a data set of four months, the Fall semester, since the communication members were relatively less changed in a semester for students. For the next four months or semester, the social relationship results may be the same as those of before or changed, since some students may graduate and leave, and some new students may come.

Twitter is a free social network and micro blogging service. This service enables users to communicate through the exchange of short messages in Twitter's web interface, SMS (text messages from cell phones), or instant messaging. The text-based message allows up to 140 characters. Twitter allows friends to follow one another with similar interests. Each time one adds an update, all of his/her followers receive a message. They can choose to respond or not. Twitter users are able to send direct messages to specific persons or broadcast to everyone. The relationships among Twitter users for exchanging direct messages are similar to the relationships among phone users. We use the direct messages in the Twitter blogs to compute reciprocity indices. We used the Twitter blog data of 6 months.

We compute the reciprocity indices by formula (1) for the call log data and Twitter blog data. In this paper we define that the reciprocity time interval $t_j - t_i$ is 24 hours, i.e., the returned calls or messages within a 24-hour period are used to compute the reciprocity index. This is only an example to choose $t_j - t_i$. We can adjust this parameter to any reasonable length of time.

IV. Experimental Results and Discussion

Figure 2 shows the reciprocity index results of user29 with his communication partners, where the x-axis indicates the phone numbers and the y-axis indicates the reciprocity index values. User29 has 39 communication partners. For example, the reciprocity index is 0.72 for the communication partner 375 and 0 for the communication partner 7. Figure 3 shows the reciprocity index of user15 with his 29 communication partners.

Figures 4 and 5 show the reciprocity index results of user3713 and user9555 with communication partners for Twitter blog data respectively.

In most cases in our experiments, the higher reciprocity index values reflect a closer relationship between members.

To find the relationships between the reciprocity index and different time intervals of reciprocity, we calculate the reciprocity index for different time intervals $t_j - t_i$ which equals to 1, 2, …, 24 hours respectively.

Figure 6 shows the reciprocity index for user39, where the x-axis indicates the reciprocity time intervals in hours and the y-axis indicates the reciprocity index values. Figure 6 shows decreasing trend of the reciprocity index values when the time intervals increase.

Figures 7 shows the probability of the reciprocity time, where the x-axis indicates the time intervals in hours, the y-axis indicates the probability and the curves are the fitted functions for user39 with his partner 316 who is frequent communication partners. From figures 7 we can see that it is exponential distribution for the reciprocity time and most reciprocity time is within 1 hour. The exponential distributions describe the times between events in a Poisson process which matches our model assumption. We find that the reciprocity distributions follow exponential trends for frequent communication partners. We also find that the case in which the reciprocity time are greater than 10 hours mostly happened when actor *a* called actor *b* at about evening sleeping time and actor *b* called actor *a* back on the next day.

By distribution fitting we have probability density function

$$f(t) = 2.8e^{-1.5t}$$

for user39 with partner 316.

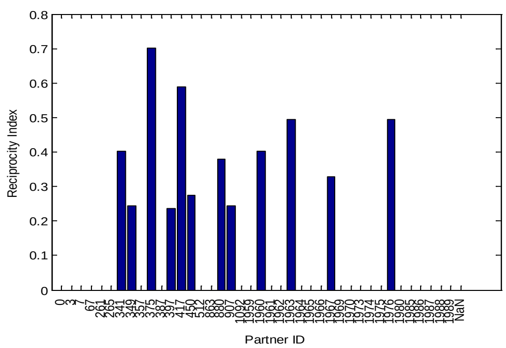

Fig. 2 The reciprocity indices for phone user29

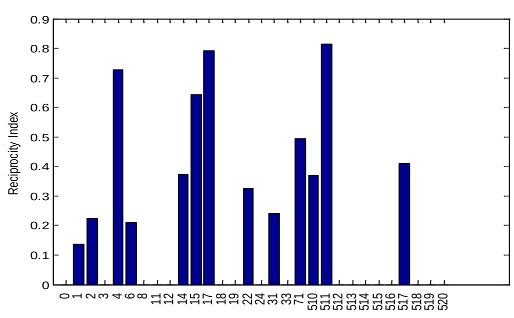

Fig. 3 The reciprocity indices for phone user15

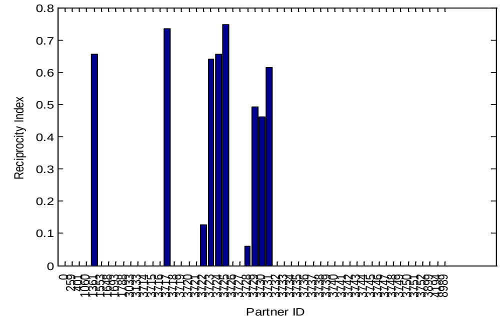

Fig. 4 The reciprocity indices for Twitter user3713

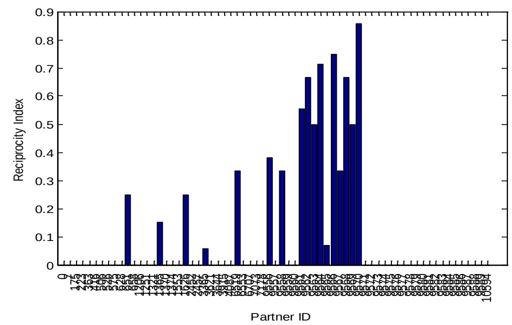

Fig. 5 The reciprocity indices for Twitter user9555

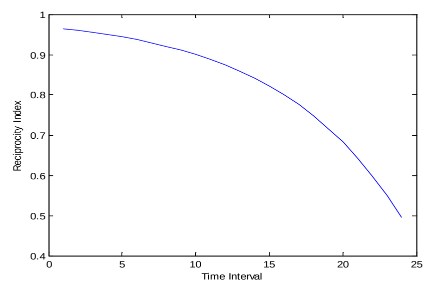

Fig. 6 Reciprocity index versus response time for phone user39

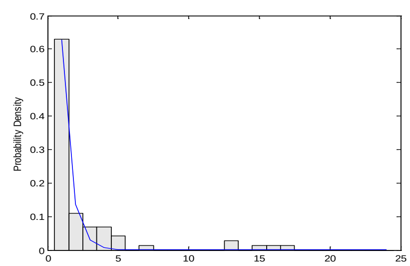

Fig. 7 The probabilities of response for phone user39 with his partner 316

V. Validation

To evaluate the accuracy of our model, we used actual call logs of 100 phone users and 460 Twitter blog users, and we randomly chose 10 phone users and 10 Twitter blog users. The phone users included students, professors and staff members. The best way to validate the results is to contact the users to get feedback. However, because of privacy issues it is almost impossible to use this method. Thus we use quantitatively hand-labeling methods to validate our model. We have used the four-month call log data and six-month micro blog data to quantify reciprocity. Note that we cannot use the data of the next four months to validate our model since social relationships may change with time. We hand-labeled the communication members based on the number of calls, duration of calls in the period, history of call logs, location, time of arrivals, and other humanly intelligible factors. We hand-labeled Twitter blog data in the same way. We compare the computed reciprocity index value of actor *a* to actor *b* with the ratio of the number of reciprocal calls or messages from actor *b* to actor *a* to the number of calls or messages from actor *a* to actor *b* in the real datasets. If they have the same trend, it is hit, otherwise, it is fail. If we cannot decide that they have the same trend or not, it is unsure (i.e., manually checking if the computed reciprocity values indeed correlate with response times)

Tables 1 and 2 describe the experimental results for reciprocity indices for 10 phone users and 10 Twitter blog users respectively. Cases appear in the "fail" or "unsure" column when the number of calls is very few. However, these kinds of cases seldom happened in our experiments. Our reciprocity index model achieves good performance with high accuracy of 91% and 89% for phone users and Twitter blog users respectively.

TABLE I

VALIDATION OF RECIPROCITY INDICES FOR PHONE USERS

User ID	Total # of members	Hit	Fail	Unsure
29(student)	39	39	0	0
41(professor)	39	37	1	1
21(student)	20	18	0	1
74(student)	13	12	0	0
88(staff)	66	63	1	2
33(staff)	31	31	0	0
15(student)	29	25	1	1
49(student)	18	16	0	1
50(student)	63	61	1	1
95(professor)	8	8	0	0

TABLE II

VALIDATION OF RECIPROCITY INDICES FOR TWITTER USERS

User ID	Total # of members	Hit	Fail	Unsure
User9555	67	61	4	2
User3713	53	46	3	4
User4467	70	61	4	5
User2691	61	53	6	2
User9413	79	69	4	6
User11153	67	59	3	5
User10509	42	38	3	1
User508	181	159	14	8
User8015	302	271	17	14
User7531	46	41	2	3

VI. CONCLUSION

In this paper we propose a reciprocity index to measure the level of reciprocity between users based on mobile phone call detail records and Twitter blogs.

The best way to validate the results is to contact the users to get feedback, but because of privacy issues it is almost impossible to use this method. Thus we use hand-labeling to validate our model.

We are able to quantify relationships for short-term period (e.g., a month), as well as long-term periods (a year or more), using our model by adjusting the parameters. Errors do appear when the number of calls is very small. However, these are few in our experiments.

This work is useful for detecting unwanted calls, e.g., spam marketing, etc. The experimental results show that our model achieves good performance with high accuracy. In our future work we plan to analyze the reciprocity index evolution and study the social network dynamics.

ACKNOWLEDGMENT

We would like to thank Nathan Eagle and Massachusetts Institute of Technology for providing us the call logs of the Reality Mining dataset.

This work is supported by the National Science Foundation under grants CNS- 0627754, CNS-0516807, CNS-0619871 and CNS-0551694. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

[1] S. Wasserman and K. Faust. Social Network Analysis: Methods and Applications. Cambridge University Press, 1994.

[2] A. W. Gouldner. "The norm of reciprocity: A preliminary statement". *American Sociological Review*, 25, 161-178, 1960.

[3] L. Katz and J. Powell. "Measurement of the tendency toward reciprocation of choice." *Sociometry*, vol. 18, pp. 659-665, 1955.

[4] L. Katz and T. Wilson. "The variance of the number of mutual choices in sociometry." *Psychometrika*, vol. 21, pp. 299-304, 1956.

[5] M. Schnegg. "Reciprocity and the Emergence of Power Laws in Social Networks". *Int. J. Mod. Phys.* C 17, issue 8, 2006.

[6] D. Garlaschelli, M. I. Loffredo. "Patterns of link reciprocity in directed networks". *Phys. Rev. Lett*. 93, 268701, 2004.

[7] G. Zamora-López, V. Zlatić, C. Zhou, Hrvoje Štefančić and J. Kurths. "Reciprocity of networks with degree correlations and arbitrary degree sequences". *Phys. Rev*. E **77**, 016106, 2008.

[8] L. M. Floría, C. Gracia-Lázaro, J. Gómez-Gardeñes, and Y. Moreno. "Social network reciprocity as a phase transition in evolutionary cooperation". *Phys. Rev*. E 79, 026106, 2009.

[9] H. Ogatha. "Computer Supported Social Networking for Augmenting Cooperation" *Computer Supported Cooperative Work* 10, 2001, pp. 189-209, Kluwer Academic Publishers.

[10] B. Hogan, D. Fisher. "Email reciprocity and personal communication bias". *NetSci2006*, May 22–25, 2006, Bloomington, IN, USA.

[11] G. Kossinets, D. Watts. "Empirical analysis of an evolving social network". *Science*, 311:88 – 90, 2006.

[12] G. Palla, A. Barabasi, T. Vicsek. "Quantifying social group evolution". *Nature*. vol 446, pp. 664-667, 2007.

[13] Massachusetts Institute of Technology: Reality Mining. http://reality.media.mit.edu/ 2009.

[14] H. Zhang, R. Dantu. "Quantifying the presence for phone users". In *Proceeding of Fifth IEEE Consumer Communications & Networking Conference*. pp. 883-887, 2008.