2014 International Symposium on Nonlinear Theory and its Applications
NOLTA2014, Luzern, Switzerland, September 14-18, 2014

Contextual Relationship in Temporal Social Networks: Circle Link

Jing Cui, Yiqing Zhang, and Xiang Li†

Electronic Engineering Department
Fudan University
Shanghai, 200433, China
†Email: lix@fudan.edu.cn

Abstract—Complex network methods have released their talents in social network analyses with statistic topological metrics. In reality, social contacts are dynamic and evolving. Nowadays they can be recorded by ubiquitous electronic information technologies, and generated into temporal social networks to provide new vision in social reality mining. Here, we define *circle link* to quantitatively analyze the contextual relationship in three empirical temporal social networks, and find that two persons having frequent consecutive interactions with a common friend trend to be close. Finally, we present a heuristic link-mining method based on circle link and acquire acceptable results.

1. Introduction

In the past decades, we have witnessed fruitful and exciting advances in studies of complex, large-scale social networks. Many methods have been brought up for predicting social relationships [1] or modeling social network structure [2], based on topological information of network. Recently, the developments in sensing and storage technologies have promoted the appearance of many large-scale human behavior data with high temporal resolution. As a consequence, a new complex network concept, namely *Temporal Networks*, has been presented [3]; new methods using temporal information have been proposed for social network analysis, such as closeness recognition [4] and interactive patterns modeling [5]. A nascent interdisciplinary area, *temporal social network*, is coming to the stage.

Inspired by the work of Song et al. [6] about the predictability of human mobility, Takaguchi et al. [7] quantified the predictability of one's contact partners in face-to-face networks using mutual information. They declare that knowing the current partner decreases the uncertainty about the next partner by a large percentage. Here we define this contextual relationship between the current partner and the next partner as *Circle Link*, implying the potential social circles. Quantitative analyses of three empirical social networks show that there is no universal memory mechanism beneath the contextual relationship, and two persons having frequent consecutive interactions with a common friend trend to be close. Based on the latter, we present a heuristic local link-mining method for the limited prior knowledge circumstance.

2. Datasets

Three datasets are used in our research, where two of them are face-to-face contacts records obtained during the ACM Hypertext Conference in Torino and the *INFECTIOUS: STAY AWAY* art-science exhibition at the Science Gallery in Dublin [8]. The third one is a co-appearance list derived from the Campus Wi-Fi login records in Fudan University [9]. We refer these three networks as HT, SG and WF respectively, and extract first three days from each dataset to generate the corresponding temporal social networks, whose detailed properties are shown in Table 1.

Table 1: Properties of all temporal social networks

Dataset		Nodes	Records	Edges	Sparsity
HT	day1	100	3460	946	5.93e-04
	day2	102	3510	1062	7.14e-04
	day3	97	2895	926	8.39e-04
SG	day1	200	2684	714	7.92e-04
	day2	204	2770	739	7.61e-04
	day3	186	2467	615	7.39e-04
WF	day1	1120	12833	10346	0.0120
	day2	2250	25772	21637	0.0067
	day3	1906	15798	13744	0.0057

Each network can be presented as a list of conversation events, specified by two participants, the start time and the duration. Individual contact list can be selected and ordered by the start time as shown in the left panel of Fig.1. The weight of a link between two person, W_L, is defined as the number of contacts between them (see Fig.1 upper right part).

We apply the definitions [7] of the uncorrelated entropy, $H_i^1 = -\sum_{j\in N_i} P_i(j)\log_2 P_i(j)$, and the conditional entropy, $H_i^2 = -\sum_{j\in N_i} P_i(j)\sum_{l\in N_i} P_i(l|j)\log_2 P_i(l|j)$ in all temporal social networks (where N_i is the set of ego i's partners/neighbors, $P_i(j)$ represents the historical probability that ego i contacts partner j, and $P_i(l|j)$ represents the

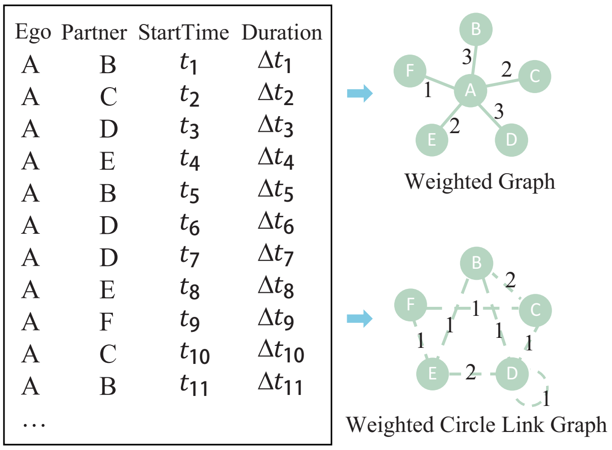

Figure 1: The illustrations of Circle Link

conditional probability that ego i contacts partner l after a contact with partner j). In contrast to the results of [7], the distributions of conditional entropies in SG and WF do not approximate normal distributions (Fig.2).

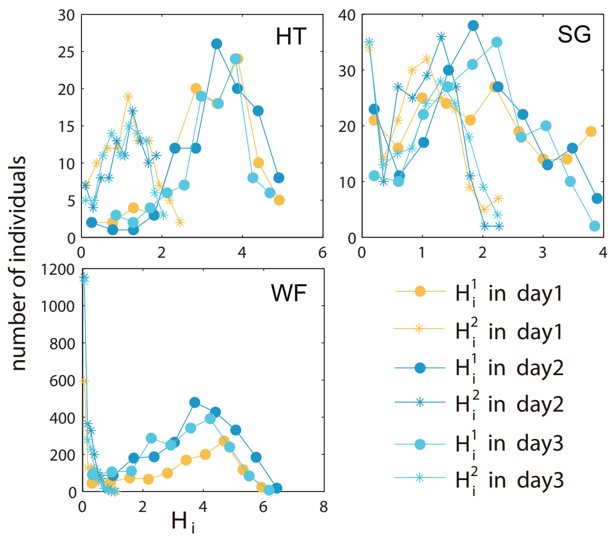

Figure 2: Distribution of entropies in three datasets

3. The Definition of Circle Link

Fig.2 shows that the conditional entropy is smaller than the corresponding uncorrelated entropy in all networks, confirming the conclusion in [7] that knowing the current partner decreases the uncertainty about the next partner. Therefore we define *Circle Link* to further analyze this contextual relationship quantitatively. The weight of a circle link between two persons is defined as the times of consecutive contacts happen with common friend.

$$W^i_{CL}(j,l) \equiv |cl_i(j,l)| \tag{1}$$

As shown in the left panel of Fig.1, since the person A has two consecutive contacts with his or her two partners B and C at the time t_1, t_2 and t_{10}, t_{11}, the weight of the circle link between B and C indicated by $cl_A(B,C)$ is 2 (see the lower right part of Fig.1). We can also define the weight of circle link in network level as follows:

$$W_{CL}(j,l) \equiv \sum_{i \in V, i \neq j,l} W^i_{CL}(j,l) \tag{2}$$

where V is the set of all nodes.

4. Results

4.1. Self Circle Link Phenomenon

The definition of circle link doesn't repel a self-loop phenomenon. An individual can be circle linked to himself or herself, like individual D in the right panel of Fig.1. We define the following *Self Circle Rate* (SCR) to represent the percentage of self circle links among all circle links observed by the ego:

$$SCR_i \equiv \frac{|cl_i(j,j)|}{|cl_i(j,l)|}, \quad j,l \in N_i, \tag{3}$$

where N_i is the set of ego i's partners.

We apply the null hypothesis that an ego contacts his or her partners without a memory mechanism, where $SCR^{null}_i = 1/|N_i|$, and define the ratio m^i_0 between SCR_i and SCR^{null}_i as follows:

$$m^i_0 \equiv \frac{SCR_i}{SCR^{null}_i} = \frac{SCR_i}{1/|N_i|}. \tag{4}$$

We averaged m^i_0 over all nodes and calculated its mean value m_0 in empirical datasets. Table 2 shows that in a conference, people do contact with a memory mechanism ($m_0 > 1$ for HT), while they do not at a gallery setting ($m_0 \approx 1$ for SG). Moreover, in a co-appearance network people tend to repel their current partner when choosing their next one ($m_0 \leq 0.65$ for WF). Furthermore, m_0 is time invariant within a dataset, indicating the memory or inverse-memory mechanism is only determined by the contexts of social contacts.

Table 2: m_0 in temporal social networks

Dataset	day1	day2	day3
HT	1.8794	1.9596	1.5307
SG	0.8308	0.7573	0.7497
WF	0.6517	0.5587	0.3470

4.2. Strength and Clustering Coefficient of Social Ties Correlated with Circle Link Weights

In the aggregated version of temporal social networks, the weights of edges represent the strengths of social ties. Here we use the Pearson correlation coefficients ρ to characterize the correlation between the strengths of social ties and temporal patterns. Table 3 shows that in all temporal social networks, the Pearson correlation coefficients $\rho_{W_L,W_{CL}}$ between the weight of edges W_L and the weight of circle links W_{CL} have relative high values $\rho_{W_L,W_{CL}} > 0.5$, indicating that two vertices frequently consecutively contact with their common neighbor have dense edges between them. Furthermore, Fig. 3 shows that the weights of circle links are inversely proportional to the corresponding link betweenness centralities, indicating that the vertices with strong circle links are in the dense local network. These are closely related to Granovetter's hypothesis that states that in social networks dense edges have on average higher weights [10]. However, in addition to having higher weights, we find that dense edges are more commonly related to "continuous group talk", temporal patterns involving three individuals.

Furthermore, we apply the definition of edge clustering coefficient [11] $CC_L(i,j) = n_C(i,j)/n_T(i,j)$, where $n_C(i,j)$ is the number of common neighbors of individual i and individual j, $n_T(i,j)$ is the total number of vertices neighbored individual i or individual j. The high edge clustering coefficient represents dense overlaps of the corresponding two vertices' neighborhoods. As shown in Table 3, the Pearson correlation coefficients $\rho_{CC_L,W_{CL}}$ between link clustering coefficients and weights of circle links have relative low values $\rho_{CC_L,W_{CL}} < 0.5$ in two face-to-face networks, but relative high values $\rho_{CC_L,W_{CL}} > 0.5$ in human indoor interaction network, indicating "continuous group talk" involving more than three individuals exits in human indoor interaction network, but not in two face-to-face networks.

Table 3: Pearson correlation coefficient of weight of circle links and other network metrics

		$\rho(W_L, W_{CL})$	$\rho(CC_L, W_{CL})$
HT	day1	0.7390	0.2851
	day2	0.7292	0.2382
	day3	0.5583	0.2248
SG	day1	0.7031	0.4682
	day2	0.7170	0.3971
	day3	0.6993	0.4833
WF	day1	0.5935	0.7532
	day2	0.5079	0.7745
	day3	0.5547	0.7701

5. Relationship Prediction Method

The positive proportion between W_{CL} and W_L reveals the feasibility of link-mining among a person's neighborhood

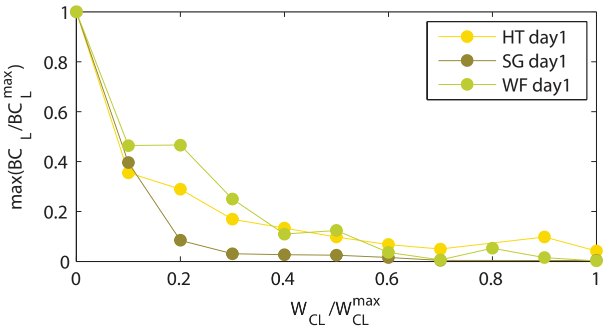

Figure 3: Counter-relationship between W_{CL} and BC_L

according to his or her contact list. Here we use the weights of circle links observed by the ego i, W^i_{CL}, as the predictor of potential links, where $W^i_{CL} > 0$ represents there exists a potential link and vice verse. The weights of edges in weighted social networks W_L are used to testify the predictor, where $W_L > 0$ indicates there exits a contact in real, and vice verse. Therefore, Table 4 shows four classifies of results in our prediction method.

Table 4: The classifies of results in relationship prediction method

	$W_L > 0$	$W_L = 0$
$W^i_{CL} > 0$	True Positive(TP)	False Positive(FP)
$W^i_{CL} = 0$	False Negative(FN)	True Negative(TN)

Precision $(TP/(TP + FP))$ and *recall* $(TP/(TP + FN))$ are used to quantify the exactness and completeness of our method. Table 5 shows that our method performs well in all temporal social networks, giving the evidence that it is possible to observer the structure of a large-scale social network by locating a few sensors and analyzing their temporal interaction data. The well performance of our method can be intuitively contributed to high positive correlation between the predictor W_{CL} and the tester W_L. Moreover, high *precision* is also caused by high positive correlation between the predictor W_{CL} and the edge clustering coefficient CC_L when comparing Table 3 with Table 5. Another possible factor is the clustering coefficient of the network. It has been testified in our previous work [12] that the clustering coefficient of nodes in WF is larger than those of face-to-face networks. We further defined the *Circle Rate* for each node (CR_i), which decides the *precision* of out method

$$CR_i \equiv \frac{\sum_{j,l \in N_i, j \neq l} (W^i_{CL}(j,l) * logical(W_L(j,l)))}{\sum_{j,l \in N_i, j \neq l} W^i_{CL}(j,l)}, \quad (5)$$

where $logical(W_L(j,l))$ is 1 when $W_L(j,l) > 0$, and it is 0 when $W_L(j,l) = 0$. T-test results as shown in Table 6 give the evidence that CR_i and CC_i mostly have same mean within a network, i.e. high *precision* is possibly caused by high clustering phenomenon in temporal social networks. Finally, the *recall* of one face-to-face network (SG) is two times larger than that of the other face-to-face network (HT), which is because that people are more temporally clustered in former network [8].

Table 5: The precision and recall of our method in all temporal social networks*

Dataset		Precision	Recall
HT	day1	0.5023	0.2650
	day2	0.4781	0.2374
	day3	0.4413	0.2563
SG	day1	0.5322	0.5669
	day2	0.5501	0.5806
	day3	0.5681	0.6483
WF	day1	0.8484	0.2537
	day2	0.7824	0.2491
	day3	0.7788	0.2896

*The precision and recall are averaged over all nodes.

Table 6: T-test of the null hypothesis that CR_i and CC_i are independent random samples from normal distributions with equal means and equal but unknown variances*

p	day1	day2	day3
HT	0.4022	0.0683	0.7784
SG	0.2189	0.4761	0.3425
WF	1.8e-07	0.1886	0.1190

*$p < 0.05$ indicates a rejection of the null hypothesis at the 5% significance level. $p \geq 0.05$ indicates a failure to reject the null hypothesis at the 5% significance level.

6. Conclusions

In this work, We defined a new term *Circle Link* to measure the contextual relationship of ego and help predict potential relationship between ego's partners. The empirical analyses confirmed that the memory mechanism is not universal in all social contacts. Furthermore, the tendency of close friends having frequent continuous interaction with their common friend can be seen as an extension of Granovetter's hypothesis to temporal social networks. Finally, we presented a heuristic method of using contextual information to excavate potential relationship within ego's neighborhoods and discuss main influence factors. We believe future amelioration of this method would help to implement larger-scale data collection of temporal social networks with limited sensors.

Acknowledgments

The authors acknowledged the SocioPatterns project for sharing their data on human face-to-face proximity contact, the Informatization Office of Fudan University for the WiFi Data collection. This work was partly supported by the National Key Basic Research and Development Program (No.2010CB731403), the NCET program (No.NCET-09-0317), and the National Natural Science Foundation (No.61273223) of China.

References

[1] L. Lü and T. Zhou. Link prediction in complex networks: A survey. *Physica A*, 390(6):1150–1170, 2011.

[2] D. J. Watts and S. H. Strogatz. Collective dynamics of small-worldnetworks. *Nature*, 393(6684):440–442, 1998.

[3] P. Holme and J. Saramäki. Temporal networks. *Phys. Rep.*, 519(3):97–125, 2012.

[4] N. Eagle, A. Pentland, and D. Lazer. Inferring friendship network structure by using mobile phone data. *Proc. Natl. Acd. Sci. USA*, 106(36):15274–15278, 2009.

[5] N. Perra, B. Gonçalves, R. Pastor-Satorras, and A. Vespignani. Activity driven modeling of time varying networks. *Sci. Rep.*, 2, 2012.

[6] C. Song, Z. Qu, N. Blumm, and A.-L. Barabási. Limits of predictability in human mobility. *Science*, 327(5968):1018–1021, 2010.

[7] T. Takaguchi, M. Nakamura, N. Sato, and et al. Predictability of conversation partners. *Phys. Rev. X*, 1(1):011008, 2011.

[8] L. Isella, J. Stehlé, A. Barrat, and et al. What's in a crowd? analysis of face-to-face behavioral networks. *J. Theor. Biol.*, 271(1):166–180, 2011.

[9] Y.-Q. Zhang and X. Li. Temporal dynamics and impact of event interactions in cyber-social populations. *Chaos*, 23(1):013131, 2013.

[10] M. Granovetter. The strength of weak ties. *Am. J. Sociol.*, 78(6):1360–1380, 1973.

[11] S. Pajevic and D. Plenz. The organization of strong links in complex networks. *Nat. Phys.*, 8(5):429–436, 2012.

[12] J. Cui, Y.-Q. Zhang, and X. Li. On the clustering coefficients of temporal networks and epidemic dynamics. In *Circuits and Systems (ISCAS), 2013 IEEE International Symposium on*, pages 2299–2302. IEEE, 2013.