Proceedings of the Twelfth International AAAI Conference on Web and Social Media (ICWSM 2018)

Leveraging Friendship Networks for Dynamic Link Prediction in Social Interaction Networks

Ruthwik R. Junuthula, Kevin S. Xu, Vijay K. Devabhaktuni
EECS Department, University of Toledo
2801 W. Bancroft St. MS 308, Toledo, OH 43606-3390, USA
rjunuth@utoledo.edu, kevin.xu@utoledo.edu, vijay.devabhaktuni@utoledo.edu

Abstract

On-line social networks (OSNs) often contain many different types of relationships between users. When studying the structure of OSNs such as Facebook, two of the most commonly studied networks are friendship and interaction networks. The link prediction problem in friendship networks has been heavily studied. There has also been prior work on link prediction in interaction networks, independent of friendship networks. In this paper, we study the predictive power of *combining friendship and interaction networks*. We hypothesize that, by leveraging friendship networks, we can improve the accuracy of link prediction in interaction networks. We augment several interaction link prediction algorithms to incorporate friendships and predicted friendships. From experiments on Facebook data, we find that incorporating friendships into interaction link prediction algorithms results in higher accuracy, but incorporating predicted friendships does not when compared to incorporating current friendships.

Introduction

Many different types of relationships between people are captured in online social networks (OSNs). For instance, Facebook captures friendships, wall posts, comments, likes, tags, and many other relations. Each type of relation can be used to construct a different type of network over the same set of nodes (people), with edges or links denoting the type of relation. These different types of networks can be categorized into two main types with distinct temporal dynamics:

- *Friendship networks*, where edges denote some form of friendship, acquaintance, family relation, or perhaps simply an expression of interest in a person, i.e. a follow.
- *Interaction networks*, where edges denote some form of interaction between nodes, such as having a conversation on a particular day.

In both cases, edges can be either directed or undirected depending on the type of friendship or interaction.

Dynamic friendship networks evolve slowly over time and are typically growing networks; that is, people add new friends much more often than they remove existing friends so that the networks densify over time (Wilson et al. 2012). On the other hand, dynamic interaction networks are highly variable over time. Two people may interact with each other at a certain time then cease to interact for a variety of reasons while still maintaining their friendship tie. On an OSN, an edge in an interaction network that persists over multiple time snapshots requires repeated interaction over time whereas an edge in a friendship network often requires only a one-time acknowledgment of a friendship or acquaintance.

In this paper, we examine the problem of *using friendship networks to improve predictions of future edges in interaction networks*. Since friendship networks are growing networks, predicting future edges in friendship networks requires only predicting the new edges that may appear in the future. On the other hand, interaction networks are evolving networks where nodes and edges are both added and deleted over time as interactions between people are initiated and dissolved, so predicting future edges in interaction networks requires predicting both the new edges that may appear as well as the current edges that may disappear.

We pose two main research questions in this paper. First, does incorporating the current friendship network lead to a more accurate prediction of the future interaction network? Second, does incorporating a predicted friendship network lead to a more accurate prediction of the future interaction network? We propose several methods of combining friendship and interaction networks to investigate these two questions on a Facebook data set (Viswanath et al. 2009).

We find that either incorporating the current friendship network or the predicted friendship network leads to a more accurate prediction of the future interaction network compared to not using any friendship information at all. We observe this for 4 different interaction link predictors combined with 2 different friendship link predictors. However, we find that incorporating predicted friendships does not improve link prediction accuracy for the interaction network compared to incorporating current friendships. This is due to the predicted friendships adding too many false positives that outweigh the added true positives they contribute.

Related Work

There have been many studies on both the structures of friendship networks, also referred to as social graphs, and interaction networks, also referred to as activity networks, in OSNs. Past examinations of friendship networks have included measurements (Mislove et al. 2007) and models of their growth (Leskovec et al. 2008), while past examinations of interaction networks have focused on persistence of interactions over time (Viswanath et al. 2009). There have also been examinations on the resemblance of the friendship and interaction networks on Facebook (Wilson et al. 2012).

Link Prediction on Friendship Networks

Friendship networks are growing networks that densify over time (Wilson et al. 2012) as many more friends are added than removed. Thus, the "traditional" link prediction setting where the objective is only to predict which new edges will form (Liben-Nowell and Kleinberg 2007) is well-suited for predicting future friendships. The traditional link prediction problem has been extensively studied, and a variety of both supervised and unsupervised algorithms have been proposed; see Lü and Zhou (2011) for a survey of methods.

We consider two simple yet effective unsupervised algorithms from the literature: Adamic-Adar (Adamic and Adar 2003), which we abbreviate as AA, and Katz. The link prediction scores of AA and Katz are calculated as follows (Liben-Nowell and Kleinberg 2007):

$$\mathrm{AA}(a,b) = \sum_{c \in \Gamma(a) \cap \Gamma(b)} \frac{1}{\log \Gamma(c)} \tag{1}$$

$$\mathrm{Katz}(a,b) = \sum_{l=1}^{\infty} \beta^l |\, \mathrm{paths}_{a,b}^{<l>} \,| \tag{2}$$

where $\Gamma(c)$ denotes the neighbors of node c, $|\, \mathrm{paths}_{a,b}^{<l>} \,|$ denotes the number of paths of size l, and $\beta \in (0,1)$ is a weight applied to lengths of paths.

Link Prediction on Interaction Networks

Unlike friendship networks, edges are both *added and removed* over time in interaction networks as people may interact for a period of time, cease to interact, and then resume their interactions. Thus, the link prediction problem on interaction networks involves predicting both the new edges that will form and the existing edges that will persist. This more complex problem is often referred to as *dynamic link prediction* (Xu and Hero III 2014).

The dynamic link prediction problem has also gained some recent attention, and most methods fall into one of three categories: univariate time series models, similarity-based methods, and probabilistic generative models (Junuthula, Xu, and Devabhaktuni 2016). We consider several dynamic link prediction algorithms that cover each of the three aforementioned categories: an exponentially-weighted moving average (EWMA), which is a special case of a general ARIMA univariate time series model (Huang and Lin 2009); time series versions of Adamic-Adar (TS-AA) (Güneş, Gündüz-Öğüdücü, and Çataltepe 2015) and Katz (TS-Katz) (Junuthula, Xu, and Devabhaktuni 2016) that apply the EWMA to the AA and Katz scores in (1) and (2), respectively; and the dynamic stochastic block model (DSBM), a probabilistic generative model, combined with the EWMA (Xu and Hero III 2014). The EWMA alone is simply a summary of past interactions and thus does not predict any new interactions, unlike the other three methods.

Link Prediction on Multiple Networks

More recent work on link prediction has involved the use of multiple networks or other data sources. Dong et al. (2015) and Hristova et al. (2016) proposed methods for link prediction across multiple networks, where they predict missing links in one network using links the other network. Gong et al. (2014) considered jointly inferring links between users and user attributes on Google+. Merritt et al. (2013) examined the problem of predicting friendships between players of a multi-player video game using their interactions, such as playing on the same team. None of this work considers the fundamental differences in structure and temporal dynamics between friendship and interaction networks (growth vs. evolution), which we consider in this paper to predict future interactions using both interactions and friendships.

Data Description

We investigate the data set collected by Viswanath et al. (2009) by crawling the Facebook New Orleans regional network. The data set contains friendships and wall posts of over $60,000$ users from September 2006 to January 2009, along with timestamps for all wall posts and for friends added after the initial crawl. We construct a sequence of network snapshots over 90-day intervals, similar to several other analyses of the data (Viswanath et al. 2009; Junuthula, Xu, and Devabhaktuni 2016), from the start of the data trace to the last full snapshot that ends in November 2008, resulting in a total of 9 snapshots. At each snapshot, we create two adjacency matrices, one with edges representing friendships, and one with edges representing interactions (wall posts) occurring during the snapshot.

In this paper, we study a representative sample of the full data set consisting of all users with degree 120 or higher in the aggregated friendship network over the entire data trace, resulting in networks containing $1,911$ nodes. Figure 1a shows that only a small fraction of friends interact via wall posts during any given 90-day time snapshot. This is partially due to the friendship network being much denser than the interaction network. Figure 1b shows that the overwhelming majority of interactions between users occur between friends. Both fractions are only slightly higher in the sample we analyze compared to the full data set.

Methodology

Given the relationship between friendship and interaction networks seen in Figure 1, one might expect friendship networks to be useful for link prediction on interaction networks. Thus, we pose the following research questions:

RQ1 Does incorporating the friendship network at the current time snapshot t lead to a more accurate prediction of the future interaction network (at time $t+1$) compared to just using the interaction networks up to time t?

RQ2 Does incorporating also the predicted friendship network at time $t+1$ (along with the current friendship network) lead to a more accurate prediction of the future interaction network (at time $t+1$)?

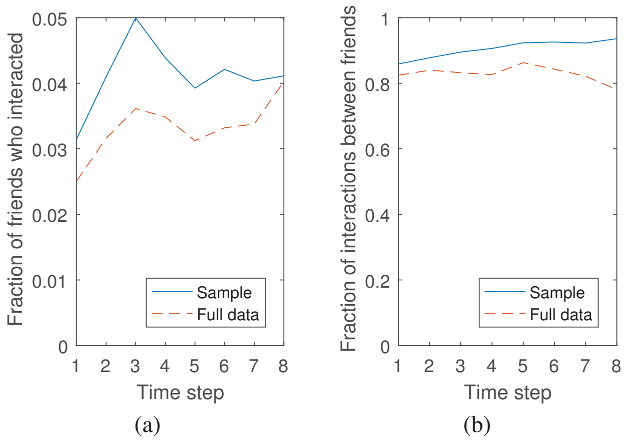

Figure 1: (a) Fraction of friends at time t who interacted at time $t+1$. (b) Fraction of interactions at time $t+1$ that are between friends at time t. In both cases, the fraction computed on the sample we study ($1,911$ nodes) are only slightly higher than on the full data ($> 60,000$ nodes).

Our objective is *not* to develop a new link prediction algorithm for friendship networks or for interaction networks. Rather, we are interested in how we can incorporate information from current or predicted friendship networks into predicting future interaction networks, which would allow us to answer our two research questions.

Using Friendships to Predict Interactions

We consider two approaches to incorporate the friendship network into dynamic link predictions on the interaction network. The first approach is to use *current friendships* (up to time snapshot t) to inform the prediction of future interactions (at time $t+1$). We do this by taking a convex combination of the adjacency matrix of the friendship network at time t and the matrix of link prediction scores obtained from the interactions (using a dynamic link prediction algorithm as described in the section) up to time t[1]. This convex combination is then used as the matrix of link prediction scores for future interactions at time $t+1$. These scores are then compared to the interactions that actually take place at time $t+1$ (see the section for details).

The second approach we consider is to use *predicted friendships* to inform the prediction of future interactions. We do this by running a traditional link prediction algorithm (described in the section) on the friendship network at time t and replacing the zeros in the friendship adjacency matrix with the (normalized) friendship link prediction scores. We then take a convex combination of this matrix with the matrix of dynamic link prediction scores from the interactions.

Evaluation Metrics

Evaluating link prediction accuracy involves comparing a binary label (whether or not an edge occurs) with a real-valued predicted score from the link prediction algorithm. Evaluation in such a setting typically involves computing the area under a threshold curve such as the area under the Receiver Operating Characteristic (ROC) curve, typically referred to just as AUC, or the area under the Precision-Recall (PR) curve, which we denote by PRAUC.

Junuthula, Xu, and Devabhaktuni (2016) studied evaluation metrics for the dynamic link prediction problem and recommended splitting the evaluation into new link prediction using PRAUC and previously observed link prediction using AUC, due to the massive difference in the degree of difficulty of the two problems. These two quantities were then combined into a single balanced metric using the geometric mean (following a correction for chance) denoted by the GMAUC. We adopt these three metrics for evaluating predictions of future interactions.

Results and Discussion

The accuracy metrics for predicting future interaction networks are shown in Table 1. The interaction network link predictors are split into three categories: predictors that do not use friendships, predictors that use only current friendships (appended with FR), and predictors that use predicted friendships (appended with the friendship link predictor used). For predictors that use friendships, we perform a grid search over all possible convex combinations in increments of 0.1 and report the results with the highest GMAUC[2].

We begin by investigating RQ1: whether incorporating current friendships improves predictions of interactions. For each of the four interaction link predictors, we see a substantial improvement in overall link prediction accuracy, as indicated by the GMAUC, by incorporating current friendships compared to no friendships. By examining the PRAUC (new) and AUC (previously observed) values, we see that this is accomplished by trading off accuracy in predicting previously observed edges in order to significantly increase accuracy in predicting new edges. This is a reasonable result because the inclusion of friendships should significantly increase the number of false positives while also leading to a greater number of correctly predicted new edges. Thus, it certainly appears that the answer to RQ1 is yes—*incorporating current friendships does indeed result in a significantly better link predictor for interaction networks.*

Next, we investigate RQ2: whether incorporating predicted friendships improves predictions of interactions. For each of the four interaction link predictors, we do see an improvement in accuracy compared to using no friendships; however, we actually see a *decrease* in accuracy, both in predicting new and previously observed edges, compared to using current friendships, regardless of whether AA or Katz is used as the friendship link predictor. We find this to be due to predicted friendships adding an overwhelming amount of false positives compared to the added true positives, resulting in lower GMAUC compared to adding only current friendships. Thus, the answer to RQ2 appears to be

[1]The scores from the AA and Katz link predictors are normalized to the interval $(0,1)$ before taking the convex combination to put them on the same scale as the friendship adjacency matrix.

[2]Code and data to reproduce our experiment results are available at https://github.com/IdeasLabUT/Friendship-Interaction-Prediction.

Table 1: Accuracy metrics for prediction of interaction links separated into new and previously observed links along with combined GMAUC metric. The first four methods use no friendships, the next four use current friendships (+ FR in the name), and the last eight use predicted friendships (predictor indicated by + AA or + Katz in the name).

Predictor	PRAUC (new)	AUC (previous)	GMAUC
EWMA	0.001	0.699	0
TS-AA	0.011	0.577	0.040
TS-Katz	0.012	0.600	0.046
DSBM	0.004	0.627	0.025
EWMA + FR	0.028	0.698	0.103
TS-AA + FR	0.042	0.567	0.075
TS-Katz + FR	0.047	0.588	0.091
DSBM + FR	0.034	0.602	0.115
EWMA + AA	0.027	0.696	0.102
EWMA + Katz	0.026	0.695	0.098
TS-AA + AA	0.037	0.572	0.072
TS-AA + Katz	0.036	0.567	0.068
TS-Katz + AA	0.037	0.567	0.070
TS-Katz + Katz	0.036	0.561	0.065
DSBM + AA	0.031	0.562	0.061
DSBM + Katz	0.029	0.553	0.054

no—*incorporating predicted friendships offers an improvement compared to no friendships but not compared to using current friendships*.

For almost all of the predictors, we find that assigning a weight of 0.9 to interaction predictors and 0.1 to friendships (current or predicted) results in the highest GMAUC. Hence, the structures of current and past interaction networks appear to produce a much better predictor of future interactions than the structures of friendship networks; however, our results demonstrate that incorporating friendship networks can definitely improve accuracy of interaction link predictions. Another interesting observation is that the EWMA and DSBM benefit more from the addition of friendships than TS-AA and TS-Katz. We believe that this is due to the EWMA and DSBM lacking mechanisms for predicting triadic closure, which is a key element of link prediction (for both interactions and friendships) in OSNs. Adding the friendship network, which includes many triangles, provides them some mechanism to favor triadic closure, which greatly improves the accuracy for both predictors.

References

Adamic, L. A., and Adar, E. 2003. Friends and neighbors on the web. *Social Networks* 25(3):211–230.

Dong, Y.; Zhang, J.; Tang, J.; Chawla, N. V.; and Wang, B. 2015. CoupledLP: Link prediction in coupled networks. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 199–208.

Gong, N. Z.; Talwalkar, A.; Mackey, L.; Huang, L.; Shin, E. C. R.; Stefanov, E.; Shi, E. R.; and Song, D. 2014. Joint link prediction and attribute inference using a Social-Attribute Network. *ACM Transactions on Intelligent Systems and Technology* 5(2):1–20.

Güneş, İ.; Gündüz-Öğüdücü, C.; and Çataltepe, Z. 2015. Link prediction using time series of neighborhood-based node similarity scores. *Data Mining and Knowledge Discovery* 147–180.

Hristova, D.; Noulas, A.; Brown, C.; Musolesi, M.; and Mascolo, C. 2016. A multilayer approach to multiplexity and link prediction in online geo-social networks. *EPJ Data Science* 5(1):24.

Huang, Z., and Lin, D. K. J. 2009. The time-series link prediction problem with applications in communication surveillance. *INFORMS Journal on Computing* 21(2):286–303.

Junuthula, R. R.; Xu, K. S.; and Devabhaktuni, V. K. 2016. Evaluating link prediction accuracy on dynamic networks with added and removed edges. In *Proceedings of the 9th IEEE International Conference on Social Computing and Networking*, 377–384.

Leskovec, J.; Backstrom, L.; Kumar, R.; and Tomkins, A. 2008. Microscopic evolution of social networks. In *Proceeding of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 462–470.

Liben-Nowell, D., and Kleinberg, J. 2007. The link-prediction problem for social networks. *Journal of the American Society for Information Science and Technology* 58(7):1019–1031.

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

Merritt, S.; Jacobs, A. Z.; Mason, W.; and Clauset, A. 2013. Detecting friendship within dynamic online interaction networks. In *Proceedings of the 7th International AAAI Conference on Weblogs and Social Media*.

Mislove, A.; Marcon, M.; Gummadi, K. P.; Druschel, P.; and Bhattacharjee, B. 2007. Measurement and analysis of online social networks. In *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement*, 29–42.

Viswanath, B.; Mislove, A.; Cha, M.; and Gummadi, K. P. 2009. On the evolution of user interaction in Facebook. In *Proceedings of the 2nd ACM Workshop on Online Social Networks*, 37–42.

Wilson, C.; Sala, A.; Puttaswamy, K. P. N.; and Zhao, B. Y. 2012. Beyond social graphs: User interactions in online social networks and their implications. *ACM Transactions on the Web* 6(4):1–31.

Xu, K. S., and Hero III, A. O. 2014. Dynamic stochastic blockmodels for time-evolving social networks. *IEEE Journal of Selected Topics in Signal Processing* 8(4):552–562.