2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining

Link Prediction in Human Mobility Networks

Yang Yang *, Nitesh V. Chawla *, Prithwish Basu †, Bhaskar Prabhala ‡, Thomas La Porta ‡
*Dept. of Computer Science, University of Notre Dame
†BBN Technologies, Cambridge MA, USA
‡Dept. of Computer Science, Pennsylvania State University
{yyang1,nchawla}@nd.edu, {pbasu}@bbn.com, {bhaskarprabhala}@gmail.com, {tlp}@cse.psu.edu

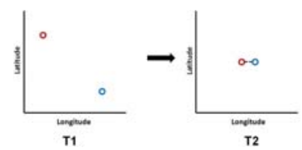
***Abstract*—The understanding of how humans move is a long-standing challenge in the natural science. An important question is, to what degree is human behavior predictable? The ability to foresee the mobility of humans is crucial from predicting the spread of human to urban planning. Previous research has focused on predicting individual mobility behavior, such as the next location prediction problem. In this paper we study the human mobility behaviors from the perspective of network science. In the human mobility network, there will be a link between two humans if they are physically proximal to each other. We perform both microscopic and macroscopic explorations on the human mobility patterns. From the microscopic perspective, our objective is to answer whether two humans will be in proximity of each other or not. While from the macroscopic perspective, we are interested in whether we can infer the future topology of the human mobility network. In this paper we explore both problems by using link prediction technology, our methodology is demonstrated to have a greater degree of precision in predicting future mobility topology.**

I. Introduction

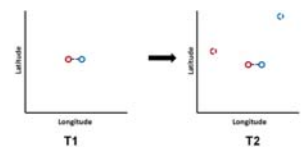
With the development of GPS and mobile technologies, it becomes much easier to monitor human mobility behaviors. The understanding of how humans move has attracted particular interest in recent years, due to the data availability and to the relevance of the topic in various domains. Most of recent research has focused on how predictable people's movements are and how long will people stay in the current place. For instance, one of these hot topics is called next location prediction, which benefits a wide range of communication systems [1] [2], from transportation planning and management to viruses spread prediction. At the same time some other research studies the global law of human mobility, for example, Simini et al. [3] proposed a radiation model which predicts mobility patterns in good agreement with mobility and transport patterns observed in a wide range of phenomena.

In the research of individual mobility prediction, the major task is to detect the period hidden in the observations. The challenges of such a problem reside in the limitations of data collection methods and inherent complexity of periodic behaviors [1] [5]. While in the research of large-scale human mobility system, the complexity in patterns of human mobility, migration and communication has been difficult to unpack, due to the availability of data and lack of sound theories. With the emergence of data on human interactions and mobility, the past couple of years have witnessed a proliferation of empirical studies on these topics [3] [4]. In this paper we take initial attempts to explore the human mobility by using the link prediction technologies.

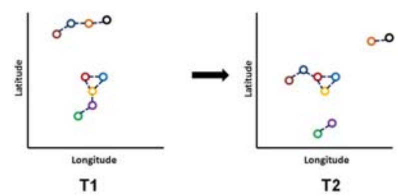
Link Prediction, that is, predicting the formation of links in a network in the future or predicting the missing links in a network, is an active topic of research. In this paper, the human mobility system is constructed as a dynamic network. Individuals in human mobility system are considered to have a link at the time t if they are detected in a proximity to each other (i.e., bluetooth scan between their devices) (Figure 1(c)). Our objective is to infer future links in such a human mobility system without of the corresponding social ties. The problem discussed in this paper is much more difficult than existing work [7] [24], where human mobility information are employed to improve the prediction performance of human social ties. We attempt to predict future proximity topology of human mobility system by only using human mobility information.



(a) Link Formation



(b) Link Dissolution



(c) Proximity Dynamics

Fig. 1. Examples.

A number of important questions remain to be studied. For example, given two humans A and B in the human mobility system, is it possible to predict that whether A and B will have a physical interaction (i.e., physically close to each other) in future (Figure 1(a))? Additionally, if we observe that humans A and B are in a proximity to each other in current time, can we answer that whether this proximity between A and B will hold in next time step (Figure 1(b))? The results of human pairs proximity prediction are informative for the next location prediction problem. Last but not least, based on historical data of human mobility system can we depict the human proximity network topology in future (Figure 1(c))? This can provide us a global view of future human mobility system. In this work we discuss the feasibility of applying link prediction technologies in these problems and develop new methodology that is more applicable in human mobility network.

The rest of the paper is organized as follows. We formally define our problems in Section II and introduce two categories

ASONAM'13, August 25-29, 2013, Niagara, Ontario, CAN

of human mobility networks in Section III. Section IV discusses the link formation prediction problem in the human mobility network. The link dissolution prediction is studied in Section V, in Section VI we present our experimental results of inferring future mobility network topology. In Section VII we conclude our work.

II. Problem Definition and Dataset

In this section, we introduce the definition of human mobility network and three link prediction tasks in this network settings.

A. Human Mobility Network

In this paper individuals are considered to have a link if there is a blue-tooth scan between them or they stay at the same location for long enough time. In such settings we transfer the human mobility system into a network representation. The *human mobility network* discussed in this paper is defined as follows.

Definition (Human Mobility Network) The human mobility network at time t is denoted as $G_t = (V_t, E_t)$, where V_t is the set of humans at time t and E_t is the set of links among humans V_t.

B. Link Formation Prediction

The human mobility system is dynamic and transient, thus the corresponding human mobility network is also changing over time. In this way, to validate the link formation (individuals are in a proximity to each other, Figure 1(a)) mechanism in human mobility network, we employ the link formation prediction problem as our evaluation metric. The associated definition is as follows:

Definition (Link Formation Prediction) In a human mobility network $G_t = (V_t, E_t)$, the link formation prediction task is to predict whether there will be a link between a pair of humans u and v at time $t + \Delta t$, where $u, v \in V_t$ and $e(u, v) \notin E_t$.

C. Link Dissolution Prediction

Different from other genres of networks, such as social networks, the link dissolution phenomenon is negligible (Figure 1(b)). We also include the *link dissolution prediction* task to analyze the human mobility network. The definition is given as follows.

Definition (Link Dissolution Prediction) In a human mobility network $G_t = (V_t, E_t)$, the link dissolution prediction task is to predict whether the link between humans u and v at time t will be in network $G_{t+\Delta t}$.

D. Inferring Network Topology

Besides analyzing the link formation and dissolution mechanisms separately, we also study whether the combination of them can lead to a prediction of future mobility network topology. That is, for the set of humans V in the human mobility system, how precisely can we predict whether there will be a link between any pairs of humans u and v at a specific time step t? The prediction of human mobility network topology is useful for decision making. For example, an accurate prediction of enemy troops mobility network will be informative for the decision of military operations.

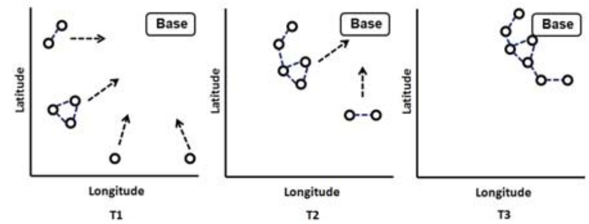


Fig. 2. Explore to Base Toy Example.

E. Datasets

In this paper we examine our approaches and perform our analysis on three social networks and six human mobility networks. The **Condmat** network [8] is extracted from a stream of 19,464 multi-agent events representing condensed matter physics collaborations from 1995 to 2000. Based on the **DBLP** dataset from [9] we attach timestamps for each collaboration and choose 3,215 authors who published at least 5 papers. The **Facebook** dataset is used by Viswanath et al. [10], which contains wall-to-wall post relationship among 11,470 users between 2004.10 and 2009.01. The **Explore** mobility network, **Lakehurst** mobility network and **Dynamic** mobility network are synthetic data generated by the *UMMF* tool [6]. While the **Infocom** dataset [11] includes blue-tooth scan events happening in three days of Infocom conference. The **Reality** dataset [12] contains blue-tooth scan events between 100 individuals between 2004-2005. The **WTD** dataset [22] includes the wireless access and mobility of the freshman students.

III. Human Mobility Network

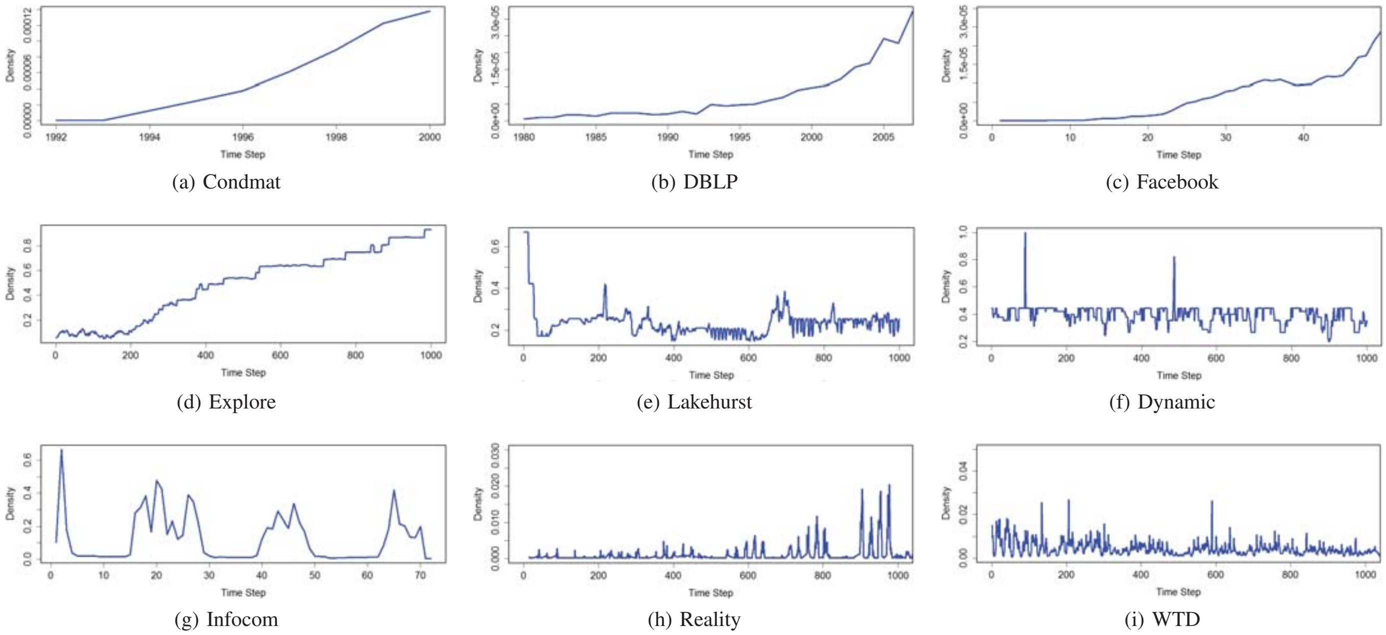
In this section we discuss different categories of human mobility networks and their corresponding network properties.

A. Protocol Directed Human Mobility Network

In real-world there are some human mobility systems where each human has designed *target plan*, *steering* and *locomotion* [6]. Basu et al. [6] proposed that individuals in such a system have three mobility building blocks: 1) *target plan*, each individual has its own *target plan* which is responsible for choosing goals and deciding what action plan to follow; 2) *steering*, the *steering* information for each individual includes definitions of how an individual should move and how fast it should travel; 3) *locomotion*, the *locomotion* is the global *mechanism* that defines the rules of movements in a human mobility system. In this paper we denote such kind of human mobility network as *protocol directed human mobility network*, which are very common in military operations. A good example is the circumstance that several groups of soldiers are exploring the location of base and marching towards the base, where each soldier has her/his *target plan - identify the location of base and return base* (Figure 2). This kind of human mobility systems are common in military operations, where three essential elements are well defined.

B. Social Human Mobility Network

Different from the *protocol directed human mobility network*, many human mobility systems in our real life are not necessary to have these three essential building elements discussed above. This kind of human mobility system is influenced by the corresponding social system, which are



(a) Condmat (b) DBLP (c) Facebook

(d) Explore (e) Lakehurst (f) Dynamic

(g) Infocom (h) Reality (i) WTD

Fig. 3. Network Density Time Series.

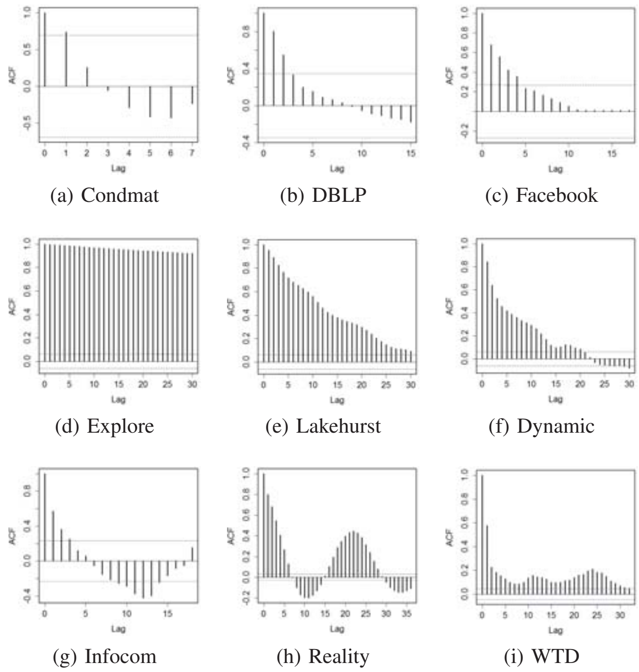
validated by many recent work [4] [7] [24]. We denote such kind of human mobility network as *social human mobility network*. For example, Wang et al. [7] found that individuals' movements strongly correlates with their ties in the social network. Additionally, Cho et al. [4] proposed that the movement information of humans can significantly enhance the link prediction performance of friendship. Generally speaking the social ties between humans have considerable impact on their movement patterns, however it is very difficult to obtaining large-scale data that has both social ties and individuals' movement information. In most cases we either only have social ties or only obtain the physical movement information. In this paper, our objective is to infer humans physical proximity by only using human mobility information. This makes our work different from the work [4], [7] and [24] where the prediction task is to infer social ties. Although the social ties are not available, the *social human mobility network* has many common characteristics that are informative for our prediction tasks.

C. A Case Study of Human Mobility Network

Link prediction is an important task in network analysis, benefiting researchers and organizations in a variety of fields. The link prediction technology is usually applied on the social network analysis, such as friendship recommendation. To better apply the link prediction technology on the human mobility network, we need to identify the differences between these two kinds of networks.

1) Time Series Analysis: The *density* is an important metric that describes the network topology. In Figure 3 we provide the network density time series for social networks and human mobility networks, which demonstrates differences of network dynamics. First, we observe that the network densities of human mobility networks are in the same order of magnitude (10^{-1}), while the network densities of social networks range from 10^{-4} to 10^{-5}. Second, the density of social network gradually increases over time while the human mobility network obviously has different density time series. Among six human mobility networks, only the *Explore* dataset has similar density evolution pattern. Third, the network density time series of three *protocol directed human mobility networks* (*Explore*, *Lakehurst* and *Dynamic*) are significantly different from each other. This is due to the reason that three essential building elements for these three networks are different from each other. At the same time we can observe that three *social human mobility networks* (*Infocom*, *Reality* and *WTD*) have similar network density evolution patterns, i.e. peridodicity or seasonality, which means the evolution of *social human mobility network* is driven by similar factors (influenced by corresponding social ties). Based on the work of [1] the periodicity is one of common phenomena in individual mobility behaviors, it is sensible that the *social human mobility network* is also periodic. This further implies the predictability of the *social human mobility network*.

To further investigate the characteristics of human mobility networks, we also provide the ACF (*auto correlation function*) plots of these time series in Figure 4. Based on our observations in ACF plots, the differences between social networks and human mobility networks are obvious. First, the network density time series of social networks are likely to have small lag order p while the network density time series of human mobility networks have high lag order. This implies that for social network prediction task we can employ recent data for learning and predicting, while for human mobility network the prediction may need longer historical data. Some network density time series of human mobility networks have long-memory properties, their ACF decays slowly to zero at a polynomial rate as the lag increases (such as Figure 4(e)). This means that the historical data has permanent effect in the network density time series, which makes the prediction task become difficult. Second, the differences between *protocol directed human mobility network* and *social human mobility network* are apparent. The *social human mobility network* has significant seasonality in the network density time series as shown in their ACF plots (Figure 4(g,h,i)). Seasonality/periodicity is an important characteristic of data which provides both challenges and opportunities for the forecasting

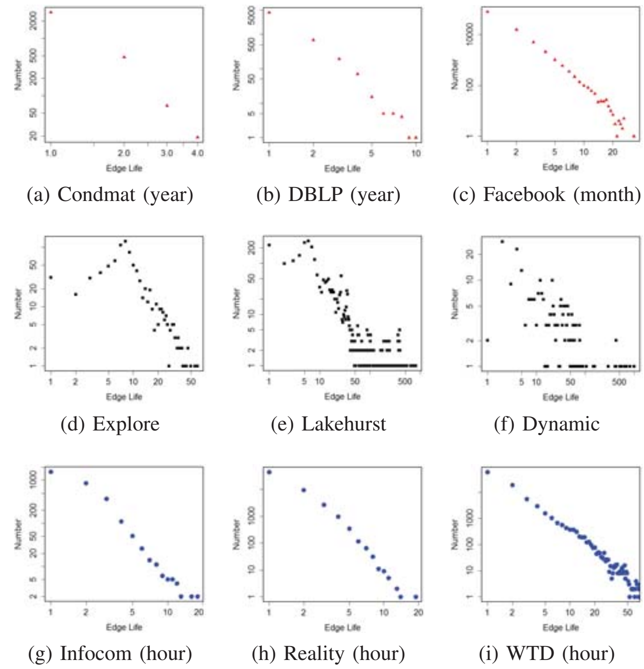


(a) Condmat (b) DBLP (c) Facebook

(d) Explore (e) Lakehurst (f) Dynamic

(g) Infocom (h) Reality (i) WTD

Fig. 4. Auto Correlation Function Plot.



(a) Condmat (year) (b) DBLP (year) (c) Facebook (month)

(d) Explore (e) Lakehurst (f) Dynamic

(g) Infocom (hour) (h) Reality (hour) (i) WTD (hour)

Fig. 5. Edge Life Distribution.

task. The seasonality/periodicity of individual human mobility is well accepted phenomena, this explains well that the *social human mobility network* also has seasonality. In conclusion, the network density time series of *social network*, *protocol directed human mobility network* and *social human mobility network* can be classified into three different categories of *time series models*. The time series of *social network* has better fit in **autoregressive (AR)** model [14], the time series of *protocol directed human mobility network* has better fit in **long-memory** model [13], and the **ARIMA (autoregressive integrated moving average)** model [14] is more appropriate to depict the time series of *social human mobility network*.

2) Edge Life Distribution: To infer the link formation/dissolution, we need to know the stability of links. To measure the stability of links we introduce a measure called *edge life*, if the link $e(u, v)$ exists for k continuous timesteps, then the *edge life* of the link $e(u, v)$ is k. The link $e(u, v)$ can have several different *edge life*. For example, $e(u, v)$ exists for the first m timesteps and at timestep $m + 1$ $e(u, v)$ is disconnected, and then from the timestep $m + 2$ $e(u, v)$ is connected for n continuous timesteps. In such a situation the *edge life* of $e(u, v)$ is $\{m, n\}$. In the context of **DBLP** dataset, if two authors continuously collaborate for 5 years, then their *edge life* is 5 years; while in the context of **Reality** dataset, if two humans are continuously proximate to each other for 3 hours, then their *edge life* is 3 hours. In Figure 5 we provide the *edge life* distributions for all 9 datasets. In our observations the *edge life* of *social networks* and *social human mobility networks* follow the power-law distribution. This further validates the connections between social ties and mobility movements, which implies that the link prediction technologies that are applicable in social networks may also be feasible in *social human mobility networks*. This characteristic of *edge life* distribution ensures that most of links are stable within a fixed duration τ, which implies that the whole network topology is also stable in this time frame. However for the *protocol directed human mobility network*, the power-law distribution does not hold. From the *edge life* distribution plots (Figure 5(d,e,f)) we can see it is difficult to find a fixed duration τ which makes sure that the network topology is stable within τ. This implies the difficulty of the prediction task in *protocol directed human mobility network*. The observations made in Figure 5 define the methodology of testing set construction for our prediction tasks in Section IV.

IV. Link Formation Prediction

In Section III we discuss the similarities and differences between social networks and human mobility networks. On one hand, the observations made in Section III confirm the feasibility of applying link prediction technologies in *social human mobility network*; on the other hand, the differences between these two genres of networks imply that directly applying the same technology will lead to the loss of performance. In this section we first explore the limitations of current link prediction technologies in *human mobility networks*, and then we propose our methodology of applying link prediction technologies in *human mobility networks*.

A. Popularity and Similarity

The principles of 'popularity' and 'similarity' are considered as common explanations for the emergence of scaling in growing networks. The principle 'popularity' is that new links are made preferentially to more popular nodes, while the principle 'similarity' refers that nodes with high similarity are more likely to form links between each other. In the link prediction problem, the predictor *preferential attachment* [16] is based on the principle of 'popularity' while the predictor *common neighbors* [17] is based on the heuristic of 'similarity. In recent work [15] a measure of attractiveness that balanced 'popularity' and 'similarity' was shown to have a better interpretation of the link formation mechanism in growing networks. Besides *preferential attachment* and *common neighbors*, there is another category of link predictors that measure the geodesic proximity between nodes, such as

TABLE I. PERFORMANCE OF PREDICTORS IN HUMAN MOBILITY NETWORKS (AUC)

Networks	Common Neighbors	Preferential Attachment	PropFlow
Explore	0.503	0.307	0.503
Lakehurst	0.536	0.533	0.531
Dynamic	0.516	0.535	0.519
Infocom	0.473	0.477	0.529
Reality	0.664	0.650	0.677
WTD	0.877	0.890	0.876

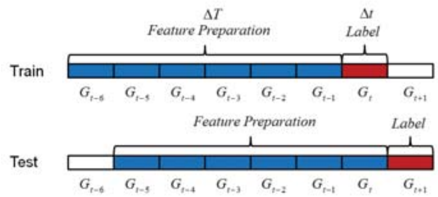


Fig. 6. Supervised Learning Framework.

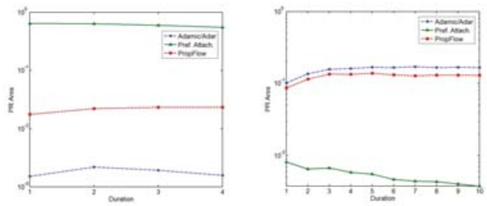
shortest path length and *propflow* [8]. Before applying these predictors in *human mobility networks* directly, we need to verify that whether these heuristics still hold.

In Table I we provide the performance of three representative link predictors in *protocol directed human mobility networks* and *social human mobility networks*. In our observations the performance of three link predictors are close to random guess in *protocol directed human mobility networks*; while in *social human mobility networks* these three predictors still have considerable performance. The performance of the three predictors are not promising in **infocom** due to the reason that there are only three days data available, which makes it difficult for the prediction task. We can observe that the 'popularity' principle (*preferential attachment*) still works in *social human mobility networks*. The explanation is, high 'popularity' of humans implies they are located in location with high population, which makes them more likely to have physical interactions with other people. At the same time the 'similarity' principle is still working in *social human mobility networks*. The heuristic is, similar proximity neighbors of humans indicates that 1) they are physically proximate to each other; 2) they have similar mobility movement patterns. Both of these factors enhance the probability that they are likely to have physical interactions. While the reason that these three predictors do not work well in *protocol directed human mobility networks* is trivial, each human has designed *target plan* and their movements are not spontaneous. In this case, even if two human have similar proximity neighbors they will never meet each other in future. In conclusion, some heuristics applicable in social networks are still working in *social human mobility networks*, however the performance is not generic and stable (Table I (Infocom)); the heuristics in social networks generally do not work well in *protocol directed human mobility networks*. This inspires us to find a more general method for the prediction task.

B. Training and Testing Set Construction

To create training and testing sets for supervised learning, we followed the framework as presented in Figure 6.

There are several important problems in the framework. First, the selection of *feature preparation* duration ΔT has an impact on the final prediction performance. In Figure 7 we can see that in **Condmat** dataset when $\Delta T <= 2$ the performance



(a) Condmat (AUPR) (b) DBLP (AUPR)

Fig. 7. Scaling of Feature Preparation Duration.

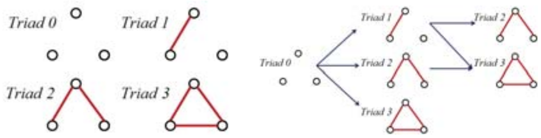
of predictors increases with the increment of ΔT and when when $\Delta T > 2$ the performance decreases with the rise of ΔT. Similar phenomena can be observed in **DBLP** dataset. This implies that an arbitrary selection of ΔT will lead to the loss of performance.

Second, the selection of *label* duration is also crucial for the link prediction task. As we discussed in Section III, *edge life* distributions of *social networks* and *social human mobility networks* follow power-law distribution. Our objective is to infer stable links and network topology within a fixed duration, which is meaningful for our analysis of mobility networks. In this paper we set $\Delta t =$ average(edge life), which ensures that within Δt duration the whole network topology is stable. However for the *protocol directed human mobility networks* the value of Δt is difficult to be selected due to their chaotic *edge life* distribution. The *protocol directed human mobility networks* could be stable for a very long time (Figure 3(d)) or be very transient (Figure 3(e, f)). This makes it extremely difficult to perform prediction tasks in *protocol directed human mobility networks*. Due to this reason in this paper we focus on *social human mobility networks*.

The most important part of the supervised framework configuration is to identify the time window of the training set. For example, as presented in Figure 6 we want to predict the links to be happening in network G_{t+1} with selected ΔT and Δt. Most supervised framework in social networks will select the network snapshots from G_{t-6} to G_{t-1} as the *feature preparation* section and G_t as the *label* section. This kind of configuration is feasible in social networks, however it is not applicable in human mobility networks. In Figure 4(g,h,i) we can observe that the network density time series of *social human mobility networks* have larger lag orders and seasonality, which means if we apply the same configuration, as suggested in Figure 6, there will be performance loss. Considering the seasonality and large lag order in *social human mobility networks*, we introduce an alternate methodology to identify the time window of the training set in the following section.

C. Link Formation Prediction in Mobility Networks

1) Feature Vector Engineering: In Table I we have shown that the three representative link predictors are not working in *protocol directed mobility network*. Additionally these three predictors work well in *social human mobility networks*, however the performance is not stable and generic. Here we introduce a new method without social principle heuristics. Our goal is to explore the collocation profile of nodes pair within triad substructure. In Figure 8 we can see four types of triad substructures and their corresponding evolution relationship. In



(a) Triads (b) Triads Evolution Relation

Fig. 8. Triad and Triad Evolution Relation.

this way a feature vector describing nodes pair's collocation within such triad substructures can lead to an estimation of their link formation likelihood. One advantage of such method is that it is not based on any social principles/heuristics. It will provide us more generic and unbiased description of the link likelihood.

In Figure 9(a) we identify four possible *triad collocation elements* for two nodes s and t where $e(s,t)$ does not belong to current network G_t. In this way for two nodes s and t ($e(s,t) \notin G_t$) we can construct a *TCE* (*triad collocation element*) vector for the link formation prediction task,

$$\text{f\_TCE}_{s,t} = \{|\text{f\_TCE\_0}_{s,t}|, |\text{f\_TCE\_1}_{s,t}|, |\text{f\_TCE\_2}_{s,t}|, |\text{f\_TCE\_3}_{s,t}|\}$$

For any disconnected node pairs s and t, $|\text{f\_TCE\_i}_{s,t}|$ is the occurrence times that nodes s and t are collocated in f_TCE_i (triad collocation elements) (Figure 9(a)).

Similarly, we design similar *TCE* vector that describes the link dissolution likelihood of nodes pair u and v if $e(u,v) \in G_t$ (Figure 9(b)),

$$\text{d\_TCE}_{u,v} = \{|\text{d\_TCE\_0}_{u,v}|, |\text{d\_TCE\_1}_{u,v}|, |\text{d\_TCE\_2}_{u,v}|, |\text{d\_TCE\_3}_{u,v}|\}$$

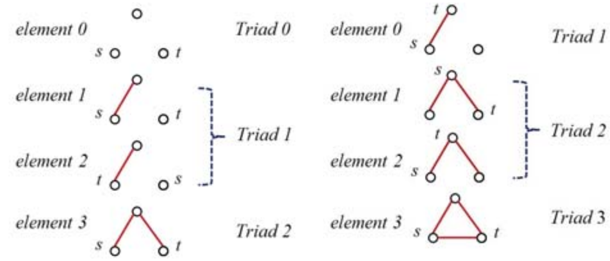
These two *TCE* vectors are included in our feature vectors correspondingly for the link formation and dissolution prediction tasks. These vectors give a multi-dimensional description of the link formation likelihood or link dissolution likelihood.

Additionally in Section III we observe that *social human mobility networks* have periodicity/seasonality, which inspires us to include the periodicity of links into the feature design. We design a measure $linkratio$ to estimate whether the link $e(s,t)$ will show up in the time window (t_0, t_1)

$$\text{link ratio}(s,t) = \frac{|(t_0 - p \times i, t_1 - p \times i), e(s,t) \text{ exists in the time window}|}{|(t_0 - p \times i, t_1 - p \times i)|},$$
$$\text{where } t_0 - p \times i > 0, i \in \{1,2,3,...\}, \text{ p is period length}$$

Additionally we also include *recency* [18] and *activeness* [19] measures for the link prediction task. *Recency* is the length of time elapsed since a node made its last connection, and *activeness* is the number of communications made in last time step. In Table II we provide the features list for baseline method and our method.

2) The Selection of ΔT, Δt and Training Set Position: As we discussed in Secion IV-B, there are three important factors in the link prediction supervised framework: ΔT, Δt and Training Set Position. Based on the observations in Figure 5 we set Δt = average(edge life)). As for the selection ΔT, based on our analysis of network density time series we set $\Delta T = order(\text{time series}, MLE)$, which is the lag order estimated from the network density time series by using *Maximum Likelihood Estimation* method and *Akaike Information Criterion* [20]. This is designed to ensure that all historical data having effect on current G_t are included in the *feature preparation* set (Figure 6).



(a) TCE for Link Formation (b) TCE for Link Dissolution

Fig. 9. Triad Collocation Elements.

TABLE II. LINK FORMATION FEATURES LIST

Features	Baseline	Triad-Period
Common Neighbors	√	√
Preferential Attachment	√	√
PropFlow	√	
Adamic/Adar [21]	√	
Jaccard Coefficient [17]	√	
f_TCE$_{s,t}$		√
link ratio		√
recency		√
activeness		√

TABLE III. PERFORMANCE COMPARISONS BETWEEN BASELINE AND TRIAD-PERIOD (AUC & AUPR) (LINK FORMATION)

Networks	Basline (AUC)	Triad-Period (AUC)	Basline (AUPR)	Triad-Period (AUPR)
Infomcom	0.476	**0.675**	0.132	**0.265**
Reality	0.661	**0.923**	0.133	**0.261**
WTD	0.835	**0.865**	0.047	**0.195**

Considering the differences of network density time series between social networks and *social human mobility networks*, we can not use the same methodology to identify the training set as described in Figure 6. Here we employ the network density time series information to identify the most similar time window to the test section (Figure 10).

3) Experimental Results: In this section we compare our methodology with the baseline method, the feature vector of *baseline* method is given in Table II. The feature vector of our *triad-period* method is also presented in Table II. For all methods, we use Bagging with *WEKA* [25] Logistic Regression as the supervised learning model. For *baseline* method we employ the traditional strategy to construct training and testing set (Figure 6), while for the *triad-period* method we use the optimized framework designed for *social human mobility networks* (Figure 10). In Table III we provide the performance of both methods. In our observation our *triad-period* method outperforms *baseline* method in terms of AUC and AUPR, which demonstrates that combining network substructure profile with network periodicity information can achieve much better performance than the traditional link prediction methodology.

V. LINK DISSOLUTION PREDICTION

As we discussed in Section III the link dissolution is significant in the evolution of *social human mobility networks*. Thus in this section we explore the link dissolution prediction problem. In Table IV we present the features list to be used

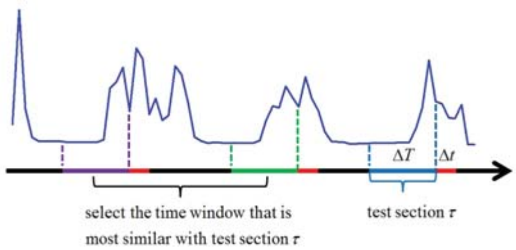


Fig. 10. Optimized Supervised Learning Framework.

TABLE IV. LINK DISSOLUTION FEATURES LIST

Features	Baseline	Triad-Period
Common Neighbors	√	√
Preferential Attachment	√	√
PropFlow	√	
Adamic/Adar [21]	√	
Jaccard Coefficient [17]	√	
d_TCE$_{s,t}$		√
link ratio		√
recency		√
activeness		√

TABLE V. PERFORMANCE COMPARISONS BETWEEN BASELINE AND TRIAD-PERIOD (AUC & AUPR) (LINK DISSOLUTION)

Networks	Basline (AUC)	Triad-Period (AUC)	Basline (AUPR)	Triad-Period (AUPR)
Infomcom	0.632	**0.701**	0.763	**0.856**
Reality	0.488	**0.720**	0.528	**0.755**
WTD	0.568	**0.817**	0.742	**0.896**

in the supervised learning model. For the *baseline* method we are using the same features in the link formation prediction problem, while for the *triad-period* method we replace the f_TCE$_{s,t}$ vector with the d_TCE$_{s,t}$ vector.

By using the same experimental settings in link formation prediction task, we perform our link dissolution prediction. The results of both methods are provided in Table V.

From the table we can see that *triad-period* method yields better results than the *baseline* method in all three datasets. To note that, due to the reason that the link dissolution prediction problem does not own the same imbalance problem as the link formation prediction task, we do not oversample or undersample the training and testing set. In conclusion, based on the results of link formation prediction and link dissolution prediction we demonstrate the efficiency and stability of our methodology in *social human mobility networks*.

VI. INFERRING MOBILITY NETWORK TOPOLOGY

Our observations and results in link formation prediction and link dissolution prediction inspire us that the combination of link formation and link dissolution results can lead to an inference of future network structure. Due to the availability of historical mobility information and the periodicity of the *social human mobility network*, we can estimate the links to be formed or to be disconnected in a given time window. In this paper we can calculate the average numbers of formed links and disconnected links in a given periodic time window, we denote them as $avg\_formation$ and $avg\_dissolution$.

A. Construction of Future Network Topology

In order to predict future network topology, we need to combine the results generated by the link formation prediction and link dissolution prediction. In the results of the link formation prediction, we select top $avg\_formation$ most likely to be connected nodes pairs (these nodes pairs are unconnected in G_t), this set of nodes pairs is denoted as P. While in the results of the link dissolution prediction, we select top $avg\_dissolution$ most likely to be disconnected links (the set of links in G_t is denoted as S), the top $avg\_dissolution$ disconnected links is denoted as D. Thus the links in our predicted future network structure will be $P \cup (S - D)$.

TABLE VI. NETWORK AGREEMENT

Comparison Pairs	GDD Agreement [23]	Degree Distr. Correlation
G_{t+1} vs. Baseline (Infocom)	0.546	0.360
G_{t+1} vs. G_{t+1-p} (Infocom)	0.582	-0.37$\times e^{-18}$
G_{t+1} vs. Triad-Period (Infocom)	**0.610**	**0.435**
G_{t+1} vs. Baseline (Reality)	0.593	0.348
G_{t+1} vs. G_{t+1-p} (Reality)	0.601	NaN
G_{t+1} vs. Triad-Period (Reality)	**0.696**	**0.869**
G_{t+1} vs. Baseline (WTD)	0.560	-0.055
G_{t+1} vs. G_{t+1-p} (WTD)	0.596	0.436
G_{t+1} vs. Triad-Period (WTD)	**0.616**	**0.849**

B. Evaluation of the Predicted Network Topology

In order to evaluate the predicted network structure, we employ two metrics that are frequently used in network comparison research for evaluation. The first one is *graphlet degree distribution agreement* [23], which is frequently used in network alignment and network comparison evaluation. And the second metric is *degree distribution correlation* [23], which is also commonly used for evaluating the network comparison task.

In the *social human mobility network*, if we are asked to predict the network structure at time $t+1$, first based on historical data (network topology before time t) we employ link prediction technologies to generate a predicted network structure, denoted as G'_{t+1}. In order to evaluate the prediction, we extract the links happening at time $t+1$ (denoted as G_{t+1}, the ground-truth network) and calculate *graphlet degree distribution agreement* and *degree distribution correlation* between G'_{t+1} and G_{t+1}.

C. Network Comparison

In Table VI we provide the performance of *baseline* method and *triad-period* method. In order to demonstrate the efficiency of prediction, we also include the comparison between G_{t+1} and G_{t+1-p}, where p is the period of the network.

In Table VI we can see the *triad-period* method outperforms other methods in terms of both *graphlet degree distribution agreement* and *degree distribution correlation*, which gives a more accurate prediction of future network topology. Additionally besides the quantitative evaluation for the network comprison, we also provide the visualized comparison between networks in Figure 11. In the visualizations of **WTD** network, we can see that *triad-period* method gives a better prediction of future network topology.

VII. CONCLUSION

In this paper we discuss the differences between *social networks* and *human mobility networks*, additionally we

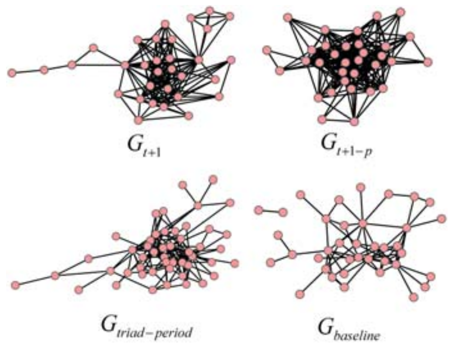


Fig. 11. Network Comparison in WTD Dataset.

demonstrate that *human mobility networks* can be categorized into two types: *protocol human mobility network* and *social human mobility network*. We show the connections between *social network* and *social human mobility network*, such as both of their *edge life* distributions follow *power-law*. And we also find that similar to individual mobility pattern, the *social human mobility network* also has periodicity which inspires our method design. The characteristics of *social human mobility network* differentiate it from the *protocol directed human mobility network*, which make it predictable. While for the *protocol directed human mobility*, three essential elements (i.e., *target plan*) are difficult to be identified from the network topology. We further explore the feasibility of link prediction technologies in *social human mobility network*, and propose a more general method (triad-period) to depict the link likelihood of link formation and dissolution. Based on the experiments on real-world datasets, we demonstrated the efficiency of our methods in the link formation prediction and link dissolution prediction problems. Besides these two link prediction problems, we integrated the predicted results to further infer *social human mobility network* future topology. By using the network comparison metrics we empirically validate the efficiency and accuracy of our methodology comparing with *baseline* method.

Future work includes extending the current framework to more *human mobility networks*. Also, for *social human mobility networks* we may further explore the interactions between *social network* and its corresponding *human mobility network*. The social information may be considered, such as cellphone call and sms information, which further empowers our framework to be more efficient in predicting links and network future topology.

Acknowledgment

Research was sponsored by the Army Research Laboratory under Cooperative Agreement Number W911NF-09-2-0053, and by the grant FA9550-12-1-0405 from the U.S. Air Force Office of Scientific Research (AFOSR) and the Defense Advanced Research Projects Agency (DARPA).

References

[1] Z. Li, J. Wang, and J. Han, *Mining Periodicity for Sparse and Incomplete Event Data*, in Proceedings of KDD, 2012.

[2] F. Calabrese, G. D. Lorenzo, and C. Ratti, *Human Mobility Prediction based on Individual and Collective Geographical Preferences*, Intelligent Transportation Systems (ITSC), 2010.

[3] F. Simini, M. C. Gonzlez, A. Maritan, and A. L. Barabsi, *A universal model for mobility and migration patterns*, Nature 2012.

[4] E. Cho, S. A. Myers, and J. Leskovec, *Friendship and Mobility: User Movement In Location-Based Social Networks*, in Proceedings of KDD'11, 2011.

[5] T. M. T. Do and D. G. Perez, *Contextual Conditional Models for Smartphone-based Human Mobility Prediction*, in Proceedings of UbiComp'12, 2012.

[6] P. Basu, A. Medina, O. Pikalo, and C. Santivanez, *Universal Mobility Modeling for MANETS*, in Proceedings of International Technology Alliance Collaboration System, 2008.

[7] D. Wang, D. Pedreschi, C. Song, F. Giannotti, and A. L. Barabasi, *Human Mobility, Social Ties, and Link Prediction*, in Proceedings of KDD'12, 2012.

[8] R. N. Litchenwalter, J. T. Lussier, and N. V. Chawla, New Perspectives and Methods in Link Prediction, in Proceedings of KDD'10, 2010.

[9] H. Deng, J. Han, B. Zhao, Y. Yu and C. Xide Lin, *Probabilistic topic models with biased propagation on heterogeneous information networks*, in Proceedings of KDD'11, 2011.

[10] B. Viswanath, A. Mislove, M. Cha, and K. P. Gummadi, *On the evolution of user interaction in Facebook*, in Proceedings of the 2nd ACM SIGCOMM Workshop on Social Networks, 2009.

[11] A. Chaintreau, A. Mtibaa, L. Massouli, and C. Diot, *Diameter of Opportunistic Mobile Networks*, in Proceedings of ACM Sigcomm CoNext, 2007

[12] N. Eagle and A. Pentland, *Reality Mining: Sensing Complex Social Systems*, Journal of Personal and Ubiquitous Computing, 2005.

[13] J. R. M. Hosking, *Fractional Differencing*, Biometrika, 1981.

[14] T. C. Mills, *Time Series Techniques for Economists*, Cambridge University Press, 1990.

[15] F. Papadopoulos, M. Kitsak, M. A. Serrano, M. Boguna and D. Krioukov, *Popularity versus Similarity in Growing Networks*, Nature 2012.

[16] A. L. Barabasi, H. Jeong, Z. Neda, E. Ravasz, A. Schubert and T. Vicsek, *Evolution of the Social Network of Scientific Collaboration*, Physica A, 2002.

[17] M. Mitzenmacher, *A Brief History of Lognormal and Power Law Distributions*, in Proceedings of the Allerton Conference on Communication, Control, and Computing, 2001.

[18] A. Potgieter, K. April, R. Cooke and I. Osunmakinde, *Temporality in link prediction: Understanding social complexity*, Sprouts: Working Papers on Info. Sys. 2007.

[19] Z. Huang and D. Lin, *The Time-Series Link Prediction Problem with Applications in Communication Surveillance*, INFORMS Journal on Computing, 2009.

[20] H. Akaike, Information Theory and An Extension of the Maximum Likelihood Principle, 2nd International Sysmposium on Information Theory, 1973.

[21] L. Adamic and E. Adar, *Friends and Neighbors on the Web*, Social Networks, 2001

[22] M. McNett and G. M. Voelker, *Access and Mobility of Wireless PDA Users*, Mobile Computing Communications Review, 2005.

[23] N. Prulj, *Biological network comparison using graphlet degree distribution*, Bioinformatics, 2007.

[24] S. Scellato, A. Noulas, C. Mascolo, *Exploiting place features in link prediction on location-based social networks*, KDD'11, 2011.

[25] M.k Hall, E. Frank, G. Holmes, B.d Pfahringer, P. Reutemann, and I.n H. Witten, *The WEKA Data Mining Software: An Update*, SIGKDD Explorations.