Predicting Cooperation Relationships in Heterogeneous Movie Networks

Wei-Chin Hung, Hung-Wei Lin, Yu-Chung Tsao, and Wei-Guang Teng

Abstract—In social network analysis, relationship prediction among people in the interpersonal network is a broadly discussed problem. Nevertheless, when modeling a real network as a heterogeneous information network instead of a homogeneous one, this problem becomes more challenging. In this work, we focus on the movie network constituted by multiple types of entities (e.g., movies, participants, studios, and genres) and multiple types of links among these entities. To clearly represent the semantic meanings in such a movie network, we utilize the meta-path-based prediction model. Advantages of our approach are two-fold. First, the meta-path-based method systematically retrieves topological features in a movie network. Second, we use the supervised method to learn the best weights connected with different topological features in building cooperation relationships. Empirical studies based on the real IMDb dataset show that our approach precisely predicts cooperation relationships in a large-scale movie network.

Index Terms—Social network analysis, link prediction, heterogeneous information network.

I. INTRODUCTION

In social network analysis, social interactions and relationships are usually represented as a connected graph consisting of nodes (or entities, objects) and edges (or links, connections). Among several alternative tasks, link prediction is an important task which attempts to realize interactions among entities. For example, the bibliographic network is typically studied in many relevant works. The bibliographic network comprises rich information for publication, such as papers, authors, venues, topics, and so on. Predicting links in a bibliographic network describes those two authors who have never co-authored a paper may do so in the next few years. Consequently, link prediction is then a challenging problem and becomes increasingly important in several practical applications.

In general, a social network can be represented as a homogeneous information network or a heterogeneous one [1]. Specifically, a homogeneous information network includes only a single entity type and a single link type. On the contrary, a heterogeneous information network includes multiple entity types and multiple link types. In other words, a heterogeneous information network contains rich and complex information. Take the social network service Facebook as an example, there are several entity types (e.g., friends, groups, apps, and pages) and multiple link types among these entities. Nevertheless, it is more challenging to analyze such a heterogeneous information network as compared to a homogeneous one.

To the best of our knowledge, only a few studies have been made to focus on daily-life applications so far. We thus focus on a more entertainment-oriented application, i.e., movie networks, in this work. To model movie networks as a heterogeneous information network, we observe that there are four different entity types in a movie network, i.e., participants (including directors and actors), movies, studios and film genres. Moreover, there are different link types when describing the relationships among these entity types. Also, links are directional and different directions represent different meanings. For example, studios "publish" movies whereas movies are "published by" studios.

It is usually necessary to simplify these complex relationships so as to better understand the structural information embedded in such a heterogeneous network. Consequently, to systematically derive semantic relationships from a heterogeneous information network, the concept of meta-path [2], [3] is introduced in this work. In other words, a meta-path is used to represent a specific relationship in the targeted movie network. For example, there is a path "*Steven Spielberg – m_2 – DreamWorks – m_7 – Tom Hanks*" between the director Steven Spielberg and the actor Tom Hanks. This means that Steven Spielberg and Tom Hanks are indirectly linked together as they participate in movies m_2 and m_7, respectively. Furthermore, these two movies are both published by the same studio, i.e., DreamWorks Studios. Similar relationships between different pairs of participants can then be categorized into one specific meta-path, i.e., "*Participant – Movie – Studio – Movie – Participant*". In addition, more types of meta-paths can be devised to represent different semantic meanings. On the other hand, a number of *topological features* (e.g., path counts) can be extracted to determine the strength of a specific meta-path.

Finally, the targeted problem of predicting cooperation relationships may be simplified as a binary classification problem. Simply speaking, our goal is to predict whether or not two participants (i.e., a director and an actor, two directors, or two actors) cooperate in the same movie in the near future. A supervised method is then used to learn the best weights connected with different topological features so as to build an appropriate classification model.

The rest of this paper is organized as follows. Related works and preliminaries are reviewed in Section II. The meta-path-based topological features and the prediction

Wei-Chin Hung, Hung-Wei Lin, and Wei-Guang Teng are with the Department of Engineering Science, National Cheng Kung University, Tainan, Taiwan (e-mail: wgteng@mail.ncku.edu.tw).
Yu-Chung Tsao is with the Department of Industrial Management, National Taiwan University of Science and Technology.

model are described in Section III. Empirical studies based on the real IMDb dataset are conducted in Section IV. Conclusions are given in Section V.

II. LITERATURE SURVEY

A. Link Prediction in Social Network Analysis

Social networks usually consist of highly dynamic entities [4]. In general, social networks grow and change quickly over time through the addition of new links which represent the appearance of new interactions in the underlying social structure [5]-[7]. To understand network evolution, the problem of link prediction is crucial and attracts increasing research interests. Usually a snapshot of a social network at time t is given and the goal is to predict the links that will be added to the network during interval from time t to a given future time t' [8]-[10]. Specifically, the problem of link prediction can be described as predicting whether or not the link l_{ij} between any two potentially linked objects o_i and o_j will appear. Therefore, an appropriate network model is usually required to represent social networks that it can support meaningful inferences from observed network data [4], [7], [11], [12].

As link prediction is formulated as a binary classification problem, several supervised methods are presented to combine different attributes with different coefficients [8], [13]. Also, relational languages are utilized to select features for training appropriate prediction models [5].

B. Utilizing the Meta Structures in Movie Networks

In the movie industry, there are many different entities involved in the production of a movie such as directors, actors, actresses, movies, studios, film genres, and so on. As shown in Fig. 1, the corresponding heterogeneous information network tends to be very complex. For simplicity, we categorize all entities into four different types which are participants (i.e., directors, actors, and actresses), movies, studios, and film genres. Note that directors, actors, and actresses are all regarded as participants. Therefore, we may more focus on the interpersonal relationships between each pair of these people as their roles may change over different movies. Moreover, to identify either *direct* or *indirect* relationships between any two participants so as to predict new cooperation relationships in this work, all these four entity types are crucial for further analysis.
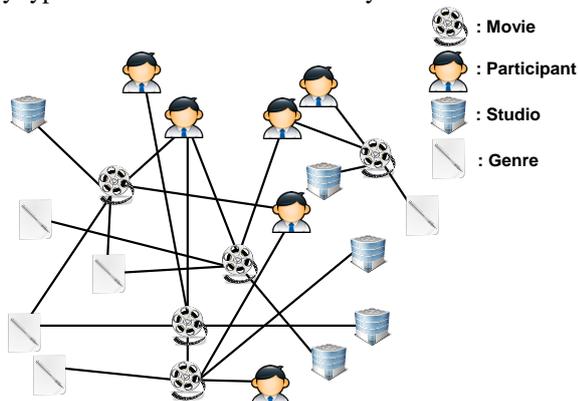

Fig. 1. Complex relationships among different entities in a real movie network.

Fig. 2 shows a directed diagram containing all the four entity types and corresponding link types in a movie network. For simplicity, four symbols are used to represent these entity types, i.e., *P* stands for participants, *M* stands for movies, *S* stands for studios, and *G* stands for film genres. Furthermore, the links are possible relationships between these entities including "participate in" and "participated" (represented as *participate*$^{-1}$) between a participant and a movie, "classify" and "classified as" (represent *classify*$^{-1}$) between a film genres and a movie, "publish" and "published by" (represent *publish*$^{-1}$) between a studio and a movie, and "continue" and "continued with" (represent *continue*$^{-1}$) between two movies, i.e., movie sequels. Note that the simplified diagram in Fig. 2 uses a meta structure to present every possible relationships. In other words, this diagram generalizes meta structures from a heterogeneous movie network.
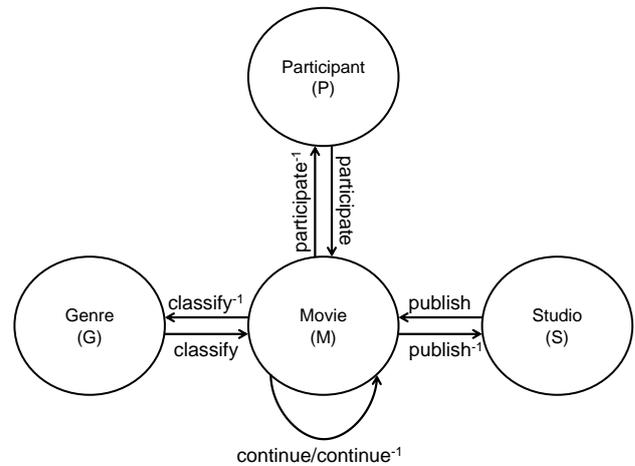

Fig. 2. A simplified diagram to show all entity types and link types in a heterogeneous movie network.

III. USING META-PATH-BASED PREDICTION MODEL IN HETEROGENEOUS MOVIE NETWORKS

A. Meta-Paths and Corresponding Topological Features

It is noticed that the distance between two people may become closer when their interaction is frequent enough. In other words, it is believed that two participants who are currently close enough are more likely to cooperate in the near future. A real example is that after meeting Matt Damon on the set of the 1997 film Good Will Hunting, director Steven Spielberg cast Damon in the 1998 film Saving Private Ryan. Note that these two films and several previous works of Spielberg and Damon belong to the same film genre, i.e., drama films. This implicitly shows the possibility of their first cooperation in the 1998 film Saving Private Ryan.

For the purpose of link prediction in this work, we then utilize different meta-paths to represent the relationships between two participants. Specifically, indirect relationships between two participants are formulated as several meta-paths of longer length as summarized in Table I. Also, the targeted cooperation relationships are formulated as the meta-path $P - M - P$ which means that two participants both involve the same movie. To reduce the impact of noises, we choose to set a length constraint so that only meta-paths of length (i.e., number of links between two ending nodes) no more than 4

are considered. For example, the length of $P - M - M - P$ is 3 and the length of $P - M - S - M - P$ is 4. All the four qualified meta-paths and their semantic meanings are listed in Table I.

TABLE I: META-PATHS OF LENGTH UP TO 4 BETWEEN TWO MOVIE PARTICIPANTS

Meta-Path	Semantic Meaning of the Relation
$P - M - P$	p_s and p_t cooperate in the same movie (i.e., the target relation)
$P - M - M - P$	p_s and p_t act in movie sequels
$P - M - P - M - P$	p_s and p_t cooperate with the same participant
$P - M - S - M - P$	p_s and p_t act in movies released by the same studio
$P - M - G - M - P$	p_s and p_t act in movies of the same genre

Next, topological features can be extracted from each of the four meta-paths, i.e., not including the targeted relation $P - M - P$ in Table I. In this work, five topological features are adopted including path count, normalized path count, random walk, inverse random walk, and symmetric random walk [14]. These topological features are illustrated as follows.

1) Path count: Given a meta-path, the measure of path count is the number of path instances between two participants p_s, and p_t. The path count is denoted as $PC_R(p_s, p_t)$, where R is the relation to indicate the meta-path. For the example in Fig. 3, $PC_R(p_s, p_t) = 3$ because there are three different paths between p_s and p_t.

2) Normalized path count: Adopting entire connectivity in the network, the measure of normalized path count is to discount the number of paths between two participants. The normalized path count is identified as in (1).

$$NPC_R(p_s, p_t) = \frac{PC_R(p_s, p_t) + PC_{R^{-1}}(p_s, p_t)}{PC_R(p_s, \cdot) + PC_R(\cdot, p_t)} \quad (1)$$

where the inverse relation of R is denoted as R^{-1}, the total number of paths starting from p_s is denoted as $PC_R(p_s, \cdot)$, and the total number of paths ending in p_t is denoted as $PC_R(\cdot, p_t)$. In other words, $PC_R(p_s, \cdot)$ and $PC_R(\cdot, p_t)$ are viewed as degrees of participants p_s and p_t in the network, respectively. Thus, $NPC_R(p_s, p_t) = (3+3) / (5+4) = 2/3$ for the example in Fig. 3.

3) Random walk: The measure of random walk across a meta-path is identified as in (2).

$$RW_R(p_s, p_t) = \frac{PC_R(p_s, p_t)}{PC_R(p_s, \cdot)} \quad (2)$$

Namely, this measure considers one way random walk following R. In other words, $RW_R(p_s, p_t)$ are regarded as the probability from p_s to p_t. Therefore, $RW_R(p_s, p) = 3/5$ for the example in Fig. 3.

4) Inverse Random walk: The measure of inverse random walk across a meta-path is identified as in (3).

$$RW_{R^{-1}}(p_s, p_t) = \frac{PC_{R^{-1}}(p_s, p_t)}{PC_{R^{-1}}(p_t, \cdot)} \quad (3)$$

Namely, this measure considers one way random walk following R and considers another direction from another participant. In other words, $RW_{R^{-1}}(p_s, p_t)$ are regarded as the probability from p_t to p_s. Accordingly, $RW_R(p_s, p_t) = 3/5$ for the example in Fig. 3.

5) Symmetric random walk: The measure of symmetric random walk as identified as in (4) considers the random walk in both two directions across the meta-path.

$$SRW_R(p_s, p_t) = RW_R(p_s, p_t) + RW_{R^{-1}}(p_s, p_t) \quad (4)$$

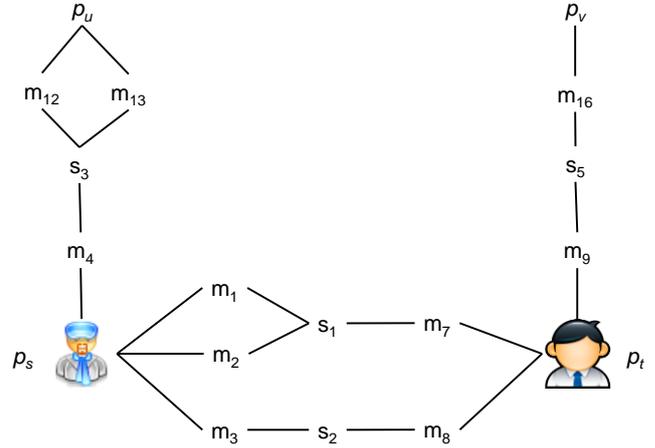Hence, $SRW_R(p_s, p_t) = 3/5 + 3/4 = 27/20$ for the example in Fig. 3.

Fig. 3. An example of the $P–M–S–M–P$ meta-path between two participants.

Namely, given a specific meta-path of two participants, five feature values can be calculated. As there are four meta-paths to be considered, there are twenty features values in total for a specific pair of participants.

B. *Establishing the Cooperation Prediction Model*

Topological features may represent the influencing factors of establishing new cooperation relationships between two participants. Consequently, the targeted problem of predicting cooperation relationships is simplified as a binary classification problem. Moreover, an appropriate supervised technique is required to learn the best weights connected with these topological attributes in predicting forthcoming cooperation relationships.
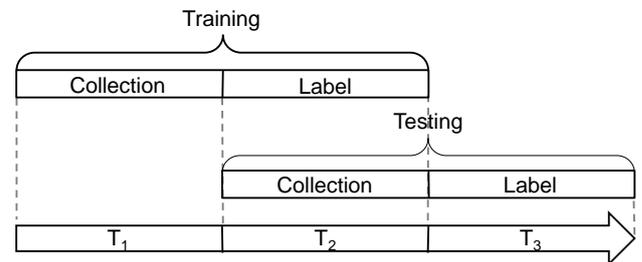

Fig. 4. A supervised learning model containing both the training and the testing frameworks is used to predict cooperation relationships.

The framework of our supervised learning technique is schematized in Fig. 4. In the training stage, given two time interval T_1 and T_2, topological features derived from the aggregated network in T_1 are used to predict the cooperation relationship in T_2, i.e., to learn the prediction model. In the testing stage, this model is usually applied to a future time interval T_3 for evaluating the model performance. Specifically, the set of participant pairs who have never cooperated in T_1

and their corresponding topological features are collected. In addition, whether cooperation relationships appear in T_2 between participants in all pairs are labeled. Note that the pairs with a forthcoming cooperation relationship are defined as *positive training pairs* whereas the rest are *negative training pairs*. Finally, with a sufficient and equal number of these two types of training pairs, an appropriate prediction model is built by learning the best weights associated with each topological feature so as to maximize the likelihood of relationship establishment [14].

Without loss of generality, we utilize the logistic regression technique to build the prediction model in this work. Benefits of this technique include: (1) the estimated values are presented in the range between 0 and 1; and (2) an appealing s-shaped description is combined effect of several main factors on the movie network [15]. General concepts of this technique are summarized as follows. For each training pair of participants, let x_i be the $(d+1)$-dimensional vector consisting of the constant 1 and values of d topological attributes (i.e., $d = 20$ in this work) between participants, and y_i be the label of whether participants cooperate in the future (i.e., $y_i = 1$ for positive pairs, and $y_i = 0$ otherwise) which follows binomial distribution with probability p_i. The probability p_i is identified as in (5).

$$p_i = \frac{e^{x_i \beta}}{e^{x_i \beta} + 1} \qquad (5)$$

where β is the $d+1$ coefficient weights associated with the constant and each topological feature.

IV. EMPIRICAL STUDIES

A. Experimental Environments

To conduct our experiments, we firstly crawl the IMDb website (http://www.imdb.com/) to collect the information of numerous movies released during 2001 and 2012. Furthermore, to constitute the $P – M – M – P$ meta-path, we also crawl the Movie Insider website (http://www.movieinsider.com/) to collect the information of movie sequels. The crawled dataset includes 254K movies, 1,227K participants (includes actors, actresses and directors) who play in these movies, 110K production companies, and 20 film genres. Moreover, we consider 3 time intervals, i.e., $T_1 = [2001, 2004]$, $T_2 = [2005, 2008]$, and $T_3 = [2009, 2012]$. In the training stage, two consecutive time intervals (e.g., $T_1 \sim T_2$) are used. For a pair of participants who do not cooperate in the first interval, the feature values and the class label (i.e., yes or no) are obtained in these two intervals, respectively.

TABLE II: STATISTICS OF THE TWO DATASETS

# Entity / Dataset	Movie (M)	Participant (P)	Studio (S)	Genre (G)
16+ movies	31,976	1,459	25,969	20
12+ movies	44,879	3,027	36,519	20

To reduce the impact of noises, we set up a threshold to qualify appropriate participants. In other words, only *highly productive participants* who participate in an enough amount of movies are to be analyzed. In our experiments, we thus

have two different threshold values, i.e., 16 and 12 movies in T_1. Then, two datasets "16+ movies" and "12+ movies" are generated accordingly. Corresponding statistics of these two datasets are shown in Table II.

We adopt the analytical software SAS Enterprise Miner (version 12.1) to perform data analysis of these datasets. There are two measures to evaluate the prediction performance. The first measure is the prediction accuracy rate for binary classification with the cut-off probability $p_i = 0.5$. The second measure is the ROC (receiver operating characteristic) curve is a useful visual tool for comparing prediction models. The participant pairs are sorted by probability values to generate a ROC curve. Furthermore, the AUC (area under curve) reflects the performance of a prediction model. Note that both measures of "accuracy" and "AUC" are ranged from 0 (i.e., the worst) to 1 (i.e., the best.)

B. Experimental Results

Note that pairs which do not have cooperation relationship in T_1 but do cooperate in T_2 are considered as positive pairs. On the other hand, negative pairs do not cooperate in both T_1 and T_2. Because the amount of negative pairs is larger than positive pairs in our datasets, we randomly choose an identical number of negative pairs to prevent the training bias. In other words, our training datasets all contains the same number of positive pairs and negative pairs.
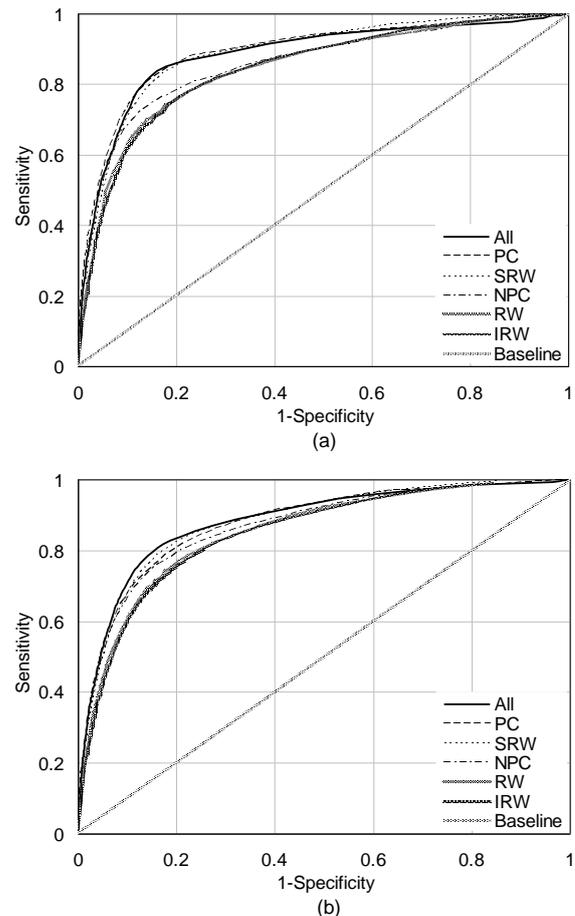

Fig. 5. Prediction accuracy of using different topological features on (a) the "16+ movies" dataset; and (b) the "12+ movies" dataset.

Firstly, we compare the effectiveness of using different sets of topological features, including (1) the path count (denoted

as PC), (2) the normalized path count (denoted as NPC), (3) the random walk (denoted as RW), (4) the inverse random walk (denoted as IRW), (5) the symmetric random walk (denoted as SRW), and (6) the all of topological features include PC, NPC, RW, IRW, and SRW (denoted as All). The comparison results of using different topological features on the two datasets are shown in Fig. 5. Utilizing all the features results in the best prediction performance. Besides, the path count (PC) and the symmetric random walk (SRW) have comparable results and they outperform the other individual topological features.
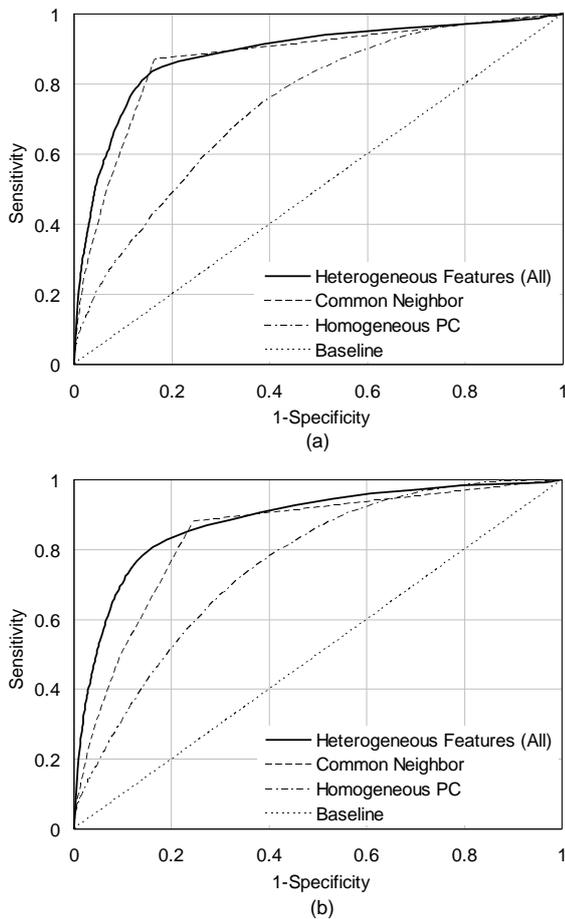

Fig. 6. Prediction accuracy of using heterogeneous features and homogeneous features on (a) the "16+ movies" dataset; and (b) the "12+ movies" dataset.

Secondly, we compare heterogeneous topological features with the homogeneous features. Homogeneous topological features include: (1) the number of common participants (denoted as common neighbor), (2) the number of paths between two participants of length no longer than 4, i.e., the counts of all the four different meta-paths are summed (denoted as homogeneous PC). The comparison results are shown in Fig. 6. In summary, heterogeneous topological features have better results than the other homogeneous topological features do. Furthermore, the results in both Fig. 5 and Fig. 6 are consistent across different datasets, i.e., "16+ movies" and "12+movies." This shows that our approach is robust.

Thirdly, to prove that our model works well in different time intervals, three different settings of training and testing intervals are used to test the "16+ movies" dataset. As shown in Table III, the resulting accuracy and AUC are high in all settings, showing a promising result of our approach. In other words, our model is verified to be general over different time periods. Also, it is noted that the accuracy rate with the cut-off probability $p_i = 0.5$ is underestimated whereas the AUC tends to be a more appropriate measure as all possible cut-offs are considered.

TABLE III: PREDICTION ACCURACY IN DIFFERENT TIME PERIODS USING THE "16+ MOVIES" DATASET

Training Framework	Testing Framework	Accuracy	AUC
$T_1 \sim T_2$	$T_1 \sim T_2$	0.81	0.91
$T_1 \sim T_2$	$T_2 \sim T_3$	0.77	0.88
$T_2 \sim T_3$	$T_2 \sim T_3$	0.80	0.87

Finally, by using the training period of $T_1 \sim T_2$ and the testing period of $T_2 \sim T_3$ on the "16+ movies" dataset, some representative prediction results are extracted and then evaluated. Specifically, the top-10 prediction results of two famous actors, i.e., Jackie Chan and Johnny Depp, for the period of 2009 to 2012 (i.e., T_3) are presented in Table IV. The names in italics are mis-predicted. Note that these prediction results are of high accuracy. For example, Jackie Chan and Fred Tatasciore cooperate in the 2011 movie "Kung Fu Panda 2." Another example is that Johnny Depp and Mark Benton cooperate in the 2009 movie of "The Imaginarium of Doctor Parnassus."

TABLE IV: TOP-10 PREDICTED PARTICIPANTS TO COOPERATE WITH EITHER JACKIE CHAN OR JOHNNY DEPP

Rank	Predicted Participant to Cooperate with	
	Jackie Chan	Johnny Depp
1	Suet Lam	Greg Bennett
2	Andy Lau	Mark Benton
3	Naoto Takenaka	Adolf Hitler
4	Garry Chalk	Jude Law
5	Fred Tatasciore	Marilyn Manson
6	*Kurt Angle*	Santiago Segura
7	Victor Garber	David Wenham
8	*Ralph Herforth*	Mark Zak
9	Brian Drummond	Judi Dench
10	Michael Bell	Stephen Fry

V. CONCLUSIONS

In this work, we have explored the scenario of cooperation relationship prediction in the heterogeneous movie network. To ease this prediction task, we have introduced the concept of meta-path and devised four different types of meta-paths. Furthermore, values of different topological features are calculated from these meta-paths so as to evaluate the possibility for two movie participants to cooperate in the near future. Consequently, we have modeled this prediction task as a binary classification problem and utilized the logistic regression technique to solve the problem. Through empirical studies on real datasets, we have verified that using all the proposed topological features in such a heterogeneous movie network results in the best prediction accuracy. In summary, the proposed approach is effective and general when conducting link prediction in a large-scale heterogeneous information network.

REFERENCES

[1] J. Han, Y. Sun, X. Yan, and P. S. Yu, "Mining knowledge from data: an information network analysis approach," in *Proc. the 28th IEEE International Conference on Data Engineering*, pp. 1214-1217, April 2012.

[2] Y. Sun and J. Han, "Meta-path-based search and mining in heterogeneous information networks," *Tsinghua Science and Technology*, vol. 18, no. 4, pp. 329-338, August 2013.

[3] X. Kong, P. S. Yu, Y. Ding, and D. J. Wild, "Meta path-based collective classification in heterogeneous information networks," in *Proc. the 21st ACM International Conference on Information and Knowledge Management*, pp. 1567-1571, October 2012.

[4] D. Liben-Nowell and J. Kleinberg, "The link-prediction problem for social networks," *Journal of the American Society for Information Science and Technology*, vol. 58, no. 7, pp. 1019-1031, March 2007.

[5] A. Popescul, R. Popescul, and L. H. Ungar, "Statistical relational learning for link prediction," in *Proc. the Workshop on Learning Statistical Models from Relational Data (in conjunction with IJCAI-2003)*, August 2003.

[6] J. O'Madadhain, P. Smyth, and L. Adamic, "Learning predictive models for link formation," in *Proc. the International Sunbelt Social Network Conference*, February 2005.

[7] J. O'Madadhain, J. Hutchins, and P. Smyth, "Prediction and Ranking Algorithms for Event-based Network Data," *ACM SIGKDD Explorations Newsletter*, vol. 7, no. 2, pp. 23-30, December 2005.

[8] R. N. Lichtenwalter, J. T. Lussier, and N. V. Chawla, "New Perspectives and Methods in Link Prediction," in *Proc. the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 243-252, July 2010.

[9] L. Getoor, "Link mining: a new data mining challenge," *ACM SIGKDD Explorations Newsletter*, vol. 5, no. 1, pp. 84-89, July 2003.

[10] M. J. Rattigan and D. Jensen, "The case for anomalous link discovery," *ACM SIGKDD Explorations Newsletter*, vol. 7, no. 2, pp. 41-47, December 2005.

[11] Y. Sun, J. Han, C. C. Aggarwal, and N. V. Chawla, "When will it happen? relationship prediction in heterogeneous information networks," in *Proc. the 5th ACM International Conference on Web Search and Data Mining*, pp. 663-672, February 2012.

[12] L. Getoor, N. Friedman, D. Koller, and B. Taskar, "Learning probabilistic models of link structure," *Journal of Machine Learning Research*, vol. 3, pp. 679-707, March 2003.

[13] C. Wang, V. Satuluri, and S. Parthasarathy, "Local probabilistic models for link prediction," in *Proc. the 7th IEEE International Conference on Data Mining*, pp. 322-331, October 2007.

[14] Y. Sun, R. Barber, M. Gupta, C. C. Aggarwal, and J. Han, "Co-author relationship prediction in heterogeneous bibliographic networks," in *Proc. the International Conference on Advances in Social Networks Analysis and Mining*, pp. 121-128, July 2011.

[15] D. G. Kleinbaum and M. Klein, "Introduction to logistic regression," *Logistic Regression: A Self-Learning Text*, Springer New York, pp. 1-39, 2010.

Wei-Chin Hung received his B.S. degree, double major in computer science and finance, from the Providence University, Taichung, Taiwan in 2006. He is currently a M.S. student in the Department of Engineering Science, National Cheng Kung University, Tainan, Taiwan. His research interests include data mining and social network analysis.

Hung-Wei Lin received his B.S. degree from the Department of Engineering Science at the National Cheng Kung University, Tainan, Taiwan in 2013. He is currently a M.S. student in the same department. His research interests include recommendation systems and embedded systems.

Yu-Chung Tsao is currently an associate professor in the Department of Industrial Management at National Taiwan University of Science and Technology. Prior to his current position, he was an associate professor in the Sino-US Global Logistics Institute at Shanghai Jiao Tong University and in the Department of Business Management at Tatung University. He was a visiting scholar in the School of Industrial and Systems Engineering at Georgia Institute of Technology. His research interests are in the areas of decision sciences, optimization application, and business analytics.

Wei-Guang Teng received B.S. and Ph.D. degrees from the Department of Electrical Engineering at the National Taiwan University, Taipei, Taiwan in 1998 and 2004, respectively. Since August 2005, he has been on the faculty at the National Cheng Kung University, Tainan, Taiwan where he is currently an associate professor in the Department of Engineering Science. He was a senior researcher at Groundhog Technologies from 2004 to 2005. His research interests include data mining, multimedia networking, and mobile computing.