Link Prediction in Multi-modal Social Networks

Panagiotis Symeonidis[1] and Christos Perentis[2*]

[1] Aristotle University, Department of Informatics, Thessaloniki 54124, Greece
symeon@csd.auth.gr
[2] Fondazione Bruno Kessler, Trento 38123, Italy
perentis@fbk.eu

Abstract. Online social networks like Facebook recommend new friends to users based on an *explicit* social network that users build by adding each other as friends. The majority of earlier work in link prediction infers new interactions between users by mainly focusing on a single network type. However, users also form several *implicit* social networks through their daily interactions like commenting on people's posts or rating similarly the same products. Prior work primarily exploited both explicit and implicit social networks to tackle the group/item recommendation problem that recommends to users groups to join or items to buy. In this paper, we show that auxiliary information from the user-item network fruitfully combines with the friendship network to enhance friend recommendations. We transform the well-known Katz algorithm to utilize a multi-modal network and provide friend recommendations. We experimentally show that the proposed method is more accurate in recommending friends when compared with two single source path-based algorithms using both synthetic and real data sets.

1 Introduction

Web 2.0 technologies and especially social networking services have gradually allowed users to form different types of interactions, like sharing and rating online items, but primarily to form online friendship networks. For example, online social networks (OSNs) such as Facebook have become popular, since they enable users to share digital content and expand their social circle by recommending new friends, based on their explicit friendship network. Moreover, social rating networks (SRNs) like Epinions and Flixter mainly focus on enabling users to share opinions and rate online items (e.g. posts and movies, respectively), but also to articulate an explicit network of trust. Both OSNs and SRNs constitute multi-modal social networks (MSNs) since they allow people to form simultaneously more than one type of explicit and/or implicit networks. In Figure 1c we demonstrate an example of an MSN, where thick black edges connect users in an explicit friendship social network and thin edges connect users with items in an implicit user-item social network. In MSNs, explicit social relationships among

[*] This work has been conducted during co-author's previous affiliation with Aristotle University[1]

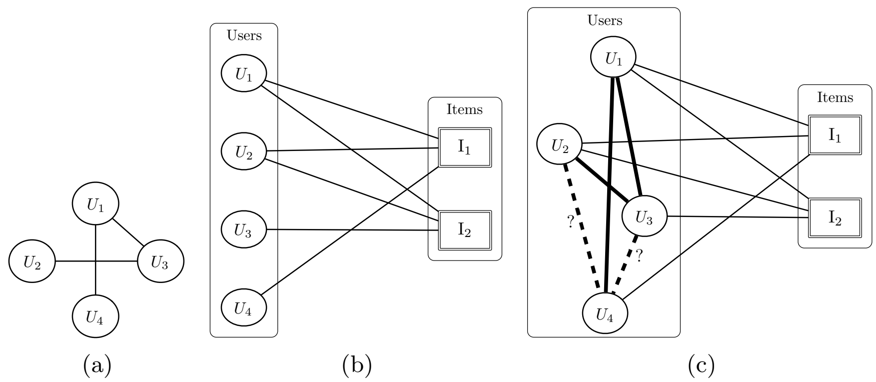

Fig. 1. Example of (a) Unipartite, (b) Bipartite and (c) Multi-modal Social Network

users co-evolve simultaneously with their interactions with several digital items (e.g. co-participating in groups, co-commenting on posts, co-rating on products etc.). MSNs have recently attracted a lot of research attention. For example, an interesting research question is how to recommend new friends to users by combining their existing social circle with the auxiliary information derived from their user-item rating network. The main goal is to enhance the accuracy of the future friendship prediction by using also the user-item rating network. Notice that available information from the bipartite user-item network is crucial due to possible absence of information from the friendship network.

There has been extended research [1, 11, 12] addressing the link prediction problem within the OSNs, by only exploiting single-source information (i.e. the unipartite user-user friendship network). However, little research has focused on exploiting multiple sources of information in predicting links within MSNs. Lu et al. [15] proposed a supervised framework, by incorporating three real implicit networks (i.e. co-author, co-citation and co-reference) to predict links in the co-author network. Vasuki et al. [22] exploited available information derived from both explicit and implicit social networks such as Orkut and Youtube to provide users with group recommendations. They have tackled the group/affiliation recommendation problem by employing both latent factor and graph proximity models, whereas the latter turned out to be the most effective.

In this paper, we propose a framework that aims to boost the friend recommendation task. Unlike previous works that primarily focused on recommending affiliated groups to users [22], we recommend new friends to users. But to do this, we look simultaneously into the user's explicit friendship and user-item implicit network. Our approach, elaborates one combined form of Katz algorithm [11] into an MSN context. We first utilize the unipartite friendship network and consider human chains of varying lengths corresponding to paths of this form $user_i \rightarrow user_j$ and $user_i \rightarrow user_j \rightarrow user_k$ in varying lengths. Then, we expand our approach to an auxiliary bipartite user-item network where we consider paths of this type $user_i \rightarrow item_j \rightarrow user_k$. This combined Katz approach allows us to provide recommendations in a unified level, traversing new paths for users to connect between and through two discrete networks: user-user and user-item. Our experimental evaluation provides evidence that the usage of auxiliary information from the bipartite user-item network succeeds in enhancing the friend recommendation task.

The rest of this paper is organized as follows. Section 2 summarizes the related work, whereas Section 3 briefly reviews preliminaries in graphs and presents a motivating example of the proposed approach. In Section 4, we present the experimental protocol and our results. Finally, in Section 5 we further discuss the proposed approach and possible directions, while Section 6 concludes this paper.

2 Related Work

The research area of link prediction in social networks tries to infer which new interactions among members of a social network are likely to occur in the near future. There are two main approaches [12] that handle the *link prediction* problem. The first approach is based on local topological features of a network, focusing mainly on the structure of the nodes. There is a variety of local similarity measures such as common neighbors, Jaccard's coefficient, Adamic/Adar index [2], Friend of a Friend (FOAF) algorithm [4] and Preferential Attachment [12], which compute the proximity between a potential pair of nodes. These similarity measures employ local features of the network like the number of common neighbors or the total number of connections and several other combinations.

The second approach is based on global features, detecting the overall path structure in a network. There is a variety of global approaches, such as Random Walk with Restart algorithm [18] and Katz status index [11], SimRank and PageRank [12], which have been used to compute the similarity between a pair of nodes. The Katz status index is a proximity measure that directly sums over the collection of all different length paths that connect two users. An attenuation factor weights the contribution of the paths to the overall similarity according to their length. Symeonidis et al. [20] proposed the FriendTNS algorithm to provide more accurate friend recommendations. They defined a transitive node similarity measure in OSNs by taking into account local and global features of a social graph. Finally, Scholz et al. [19] performed unsupervised random walks for predicting links in user-user networks (i.e. co-author in DBLP).

Besides the aforementioned link prediction algorithms that are based solely on graph structure, there are also other methods that exploit other data sources such as messages among users, co-authored paper and common tagging. For instance, Ido Guy et al. [10], proposed a novel user interface widget for providing users with recommendations of people. Their people recommendations were based on aggregated information collected from various sources across IBM (e.g. common tagging, common link structure, common co-authored papers). Chen et al. [4] evaluated four recommender algorithms (Content Matching, Content-plus-Link, the FOAF algorithm and, SONAR) to help users discover new friends on IBM's OSN. Lo and Lin [13] proposed two algorithms, denoted as *weighted* *minimum message ratio* (WMR) and *weighted information ratio* (WIR), respectively, which generate a friend list based on real-time message interaction among members of an OSN. Cha et al. [3] collected and analyzed large-scale traces of information dissemination in the Flickr social network. They experimentally derived that over 50% of users find their favorite pictures (i.e., pictures they bookmark) from their friends in an OSN. TidalTrust [9] and MoleTrust [16] are also hybrid approaches that combine the rating data of collaborative filtering systems with the link data of trust-based social networks (i.e. Epinions.com) in order to improve the recommendation accuracy.

There has also been research work that uses supervised approaches to address the link prediction problem in multiple data sources. For instance, Lu et al. [15] exploited topological features from four networks and applied a probabilistic model to learn the network dynamics. They showed that supervised approaches can improve link prediction tasks, suggesting that independency assumptions and scaling issues should be further investigated. In addition, Davis et al. [5] introduced a probabilistically weighted extension of the local-based Adamic/Adar measure for heterogenous networks and showed that a supervised approach based on topological features enhances prediction performance. Finally, maximum-likelihood methods have been proposed to deal with the link prediction problem providing insights about network organization that are difficult to obtain from similarity-based approaches [14]. However, these methods presume specific organizing principles of the network structure and suffer from scalability and accuracy issues.

3 Preliminaries in Multi-modal Graphs

In this section, we present the most important notations with the corresponding definitions and a motivating example based on Figure 1 that will be used throughout the rest of the paper. The multi-modal graph of Figure 1c consists of (i) friendships among users of an OSN and (ii) users' affiliations with items shown in Figure 1a and 1b, respectively. For our calculations, we will use well-known representations, such as the adjacency matrix $\mathbf{A}^{u \times u}$ of friendship network, and the user-item matrix $\mathbf{R}^{u \times w}$ of the affiliation network.

3.1 Link Prediction Based on User-User Unipartite Graph

Let $\mathcal{G}$ be a graph with a set of nodes $\mathcal{V}$ and a set of edges $\mathcal{E}$. Every edge is defined by a specific pair of graph nodes (v_i, v_j), where $v_i, v_j \in \mathcal{V}$. We assume that the graph $\mathcal{G}$ is undirected and unweighted, thus the graph edges do not have any weights, plus the order of nodes in an edge is not important. Therefore, (v_i, v_j) and (v_j, v_i) denote the same edge on $\mathcal{G}$. We also assume that the graph $\mathcal{G}$ can not have multiple edges that connect two nodes, thus if two nodes v_i, v_j are connected with an edge of $\mathcal{E}$, then there can not exist another edge in $\mathcal{E}$ also connecting them. Finally, we assume that there can not be self loop edges on $\mathcal{G}$ (i.e. a node can not be connected to itself). A common graph representation is the *adjacency matrix* $\mathbf{A}^{n \times n}$, where $n{=}|\mathcal{V}|$ is the number of nodes in $\mathcal{G}$. Therefore, it has n rows and n columns labelled by the graph nodes. For an unweighted non-multiple graph (such as $\mathcal{G}$), the adjacency matrix values are set as $\mathbf{A}_{ij}{=}1$ if $(v_i, v_j) \in \mathcal{E}$ and $\mathbf{A}_{ij}{=}0$ otherwise. Following all previous assumptions and definitions, the adjacency matrix of an undirected and unweighted graph such as $\mathcal{G}$, is a symmetric matrix with values 0 and 1, if two nodes are neighbors or not, respectively. In addition, as there are no self loop edges, the main matrix diagonal has zero values. The adjacency matrix of the friendship network for our running example is depicted in Figure 2a. As we want to investigate the relations with ?, we can assume that initially are equal to 0 (i.e. there are no connections between the corresponding users). It is obvious from Figure 1a and its corresponding adjacency matrix $\mathbf{A}$ of Figure 2a that U_1 is connected with U_3 and U_4, while U_2 only to U_3. In terms of social networks, U_1 and U_2 have a "mutual" friend U_3, since they are both connected to this user. Let's assume in our running example, that we want to propose new friends to user U_4. There are several global similarity measures [12] (i.e Katz status index, RWR algorithm, SimRank algorithm, etc.) for capturing similarity of nodes in a network, which are path-dependent. We apply the Katz status index, which defines the similarity score between two nodes V_x and V_y, by summing over paths of varying length ℓ connecting V_x to V_y given by Equation 1:

$$Katz_{\beta} = \sum_{\ell=1}^{\infty} \beta^{\ell} |paths^{\ell}_{V_x,V_y}| \tag{1}$$

where $paths^{\ell}_{V_x,V_y}$ is the set of all length-ℓ paths from node V_x to V_y, which are computed by the adjacency matrix $\mathbf{A}$. Katz status index exploits that raising the adjacency matrix in the power of n produces the number of n-paths connecting one pair of nodes. An attenuation factor β is introduced to efficiently weight the contribution of different lengths of paths to the final similarity score between node pairs. Very low values of β force long paths connecting a pair of nodes to contribute much less to the final similarity score. Thus, it is possible to limit the reach of the similarity measure by weighting higher the shorter paths from node's neighborhood. Both analytical and factorized form of Katz is given by Equation 2 when applied to the adjacency matrix $\mathbf{A}$ of Figure 2a:

$$Katz(\mathbf{A};\beta) = \beta\mathbf{A} + \beta^2\mathbf{A}^2 + \beta^3\mathbf{A}^3 + ... = (I - \beta\mathbf{A})^{-1} - I \tag{2}$$

	U_1	U_2	U_3	U_4
U_1	0	0	1	1
U_2	0	0	1	?
U_3	1	1	0	?
U_4	1	?	?	0

(a)

	U_1	U_2	U_3	U_4
U_1	0	0.16	0.49	0.43
U_2	0.16	0	0.43	0.05
U_3	0.49	0.43	0	0.16
U_4	0.43	**0.05**	**0.16**	0

(b)

Fig. 2. Running Example: (a) Adjacency $\mathbf{A}$ and (b) Similarity Matrix of User-User Unipartite Social Network

The identity matrix $\mathbf{I}_n$ is a $n \times n$ matrix of size n holding ones on the main diagonal and being of the same size n as the adjacency matrix $\mathbf{A}$. The attenuation factor β should take values that can ensure series convergence and allow the computation of the $\mathbf{A}^{-1}$ inverse matrix. Therefore, the β attenuation factor can take values $\beta < 1/\lambda$, where λ is the largest absolute value among any eigenvalue of matrix $\mathbf{A}$ [8, 11]. We choose β equal to $1/(1+K)$, as L.Katz originally introduced [11] and Foster et al. [8] employed for the fast approximation implementation, where K is the maximum row/column sum of $\mathbf{A}$. This choice is sensible satisfying the sufficient condition for the computations to fulfill and adaptive to the matrix size, thus, to each dataset. Back to our running example, we want to recommend new friends to U_4. Thus, we apply Katz algorithm to the unipartite friendship graph $\mathcal{G}$, in order to provide recommendations based on an induced similarity matrix. We compute the Katz status index by applying Equation 2 to the adjacency matrix $\mathbf{A}$ of Figure 2a. The attenuation factor β for matrix $\mathbf{A}$ is $\beta = 1/(1+2)$, equal to 0.33. Notice that Katz calculates similarity between two nodes taking into account paths of length $\ell > 1$.

Firstly, the similarity between U_4 and U_2 is computed based on the unique path that connects them $4 \to 1 \to 3 \to 2$, shown in Figure 1a. This path of length-3 contributes a similarity score of 0.05 given in matrix of Figure 2b. For the similarity between U_4 and U_3, there is only one path of length-2 ($4 \to 1 \to 3$) corresponding to a score of 0.16. The user-user similarity matrix entries of Figure 2b capture the friendship relationships in the unipartite social network and its rows show the "proximity" among users. There is a clear indication from the above similarity matrix that U_3 should be recommended as friend to U_4 instead of U_2, with similarity value $0.16 > 0.05$. Notice that the similarity score of (U_1,U_4) pair is the highest observed matrix entry, but we do not recommend U_1 to U_4, since they are already "friends" and it is not a new link.

3.2 Link Prediction Based on User-Item Bipartite Graph

Users can also form several implicit social networks through their daily interactions like co-commenting on people's posts, co-rating products, and co-tagging people's photos [22]. These implicit relations contain edges between two types of entities (vertices in a graph), such as a user-item bipartite graph. Let $\mathcal{G}' = (\mathcal{V} + \mathcal{W}, \mathcal{E})$ be a bipartite graph with two sets of nodes $\mathcal{V}$ and $\mathcal{W}$, and a set of edges $\mathcal{E}$. Every edge is defined by a specific pair of graph nodes (v_i, w_j), where $v_i \in \mathcal{V}$ denotes users set and $w_j \in \mathcal{W}$ items set. Following the unipartite adjacency matrix notation, we define the biadjacency matrix $\mathbf{R}$ corresponding to bipartite user-item network as a new matrix $\mathbf{B} = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{bmatrix}$ equal to $\begin{bmatrix} 0 & \mathbf{R} \\ \mathbf{R}^T & 0 \end{bmatrix}$, where $R_{v_i,w_j} = 1$ if $(v_i, w_j) \in \mathcal{E}$ and $\mathbf{R}_{v_i,w_j} = 0$ otherwise.

We extend our running example by affiliating users with items, as depicted in Figure 1b and the corresponding biadjacency matrix $\mathbf{R}$ of Figure 3a. Our main task remains the friend recommendation for U_4 by using this time only the bipartite user-item $\mathbf{R}$. Edges of $\mathbf{R}$ represent length-1 paths of from a user U_i ending to an item I_j. By multiplying matrix $\mathbf{R}$ with its transpose $\mathbf{R}^T$, we derive all length-2 paths of this form $U_i \rightarrow I_j \rightarrow U_k$, where users are connected through items. We employ the $\mathbf{B}^{n\times n}$ adjacency matrix of Figure 3b where block $\mathbf{B}^2_{11}(U_i, U_j) = \mathbf{R}(U_i, I_j) \times \mathbf{R}^T(I_j, U_i)$. If $B^2_{11}(U_i, U_j) > 1$, these two users are connected with an *implicit* (i.e co-share, co-like, etc.) relationship with a potential item. Katz algorithm is next applied to adjacency matrix $\mathbf{B}$ using Equation 3 to

	I_1	I_2
U_1	1	1
U_2	1	1
U_3	0	1
U_4	1	0

(a)

	U_1	U_2	U_3	U_4	I_1	I_2
U_1	0	0	0	0	1	1
U_2	0	0	0	0	1	1
U_3	0	0	0	0	0	1
U_4	0	0	0	0	1	0
I_1	1	1	0	1	0	0
I_2	1	1	1	0	0	0

(b)

	U_1	U_2	U_3	U_4	I_1	I_2
U_1	0	0.18	0.09	0.09	0.36	0.36
U_2	0.18	0	0.09	0.09	0.36	0.36
U_3	0.09	0.09	0	0.01	0.04	0.31
U_4	0.09	**0.09**	**0.01**	0	0.31	0.04
I_1	0.36	0.36	0.04	0.31	0	0.19
I_2	0.36	0.36	0.31	0.04	0.19	0

(c)

Fig. 3. Running Example: (a) User-Item $\mathbf{R}$, (b) Adjacency $\mathbf{B}$ and (c) Similarity Matrix of Bipartite Social Network

obtain a new similarity matrix derived only from the bipartite user-item network.

$$Katz(\mathbf{B}; \beta) = \beta\mathbf{B} + \beta^2\mathbf{B}^2 + \beta^3\mathbf{B}^3 + \beta^4\mathbf{B}^4 \cdots = \sum_{\ell=1}^{\infty} \beta^\ell \mathbf{B}^\ell \tag{3}$$

The odd factors of Equation 3 do not contribute to the similarity among users denoted in $\mathbf{B}_{11}$ block, because they represent paths ending to items (we could exclude them from the equation). Back to the running example, we aim to recommend friends to U_4, thus we calculate its similarity with U_2 and U_3. We apply Katz algorithm to the bipartite graph $\mathcal{G}'$ by applying Equation 3 to the adjacency matrix $\mathbf{B}$ of Figure 3b. The computed similarities are summarized in the matrix of Figure 3c and the attenuation factor for the bipartite network is β=1/(1+3), equal to 0.25.

In the 4th row of similarity matrix of Figure 3c is clearly indicated that user U_2 should be recommended to user U_4 as a friend instead of U_3, with similarity value 0.09>0.01. There is a difference between the produced recommendations when using different information sources, since previously we recommended U_3 to U_4 using only the user-user unipartite social network. The information from user-item bipartite network suggests that we should recommend U_2 to U_4, since more paths through the items connect these two users. Specifically, U_4 and U_2 are connected through one path of length-2 ($U_4 \rightarrow I_1 \rightarrow U_2$) and two paths of length-4 ($U_4 \rightarrow I_1 \rightarrow U_1 \rightarrow I_1 \rightarrow U_2$ and $U_4 \rightarrow I_1 \rightarrow U_1 \rightarrow I_2 \rightarrow U_2$). In contrast, U_4 and U_3 are connected through two paths of length-4 ($U_4 \rightarrow I_1 \rightarrow U_1 \rightarrow I_2 \rightarrow U_3$ and $U_4 \rightarrow I_1 \rightarrow U_2 \rightarrow I_2 \rightarrow U_3$).

In our running example, we produced all the possible similarity scores concerning both the user-user and the user-item relationships, by using the adjacency matrix $\mathbf{B}$ of Figure 3b. We exploit only the information from $\mathbf{B}_{12}$ and

$\mathbf{B}_{21}$ blocks of matrix $\mathbf{B}$ that correspond to the user-item network, in order to capture similarities concerning block $\mathbf{B}_{11}$. We also produced the similarities for the auxiliary item-item network given by block $\mathbf{B}_{22}$ that is not currently used here. In the future this block of the matrix could reveal semantic relationships between items for other recommendation tasks, like cross-domain.

3.3 Proposed Approach: Link Prediction in Multi-modal Graphs

In this section, the approach of combining the heterogeneous multiple sources of the unipartite user-user and the bipartite user-item graphs, is presented. These two graphs are combined in a multi-modal graph of Figure 1c. This approach enables recommendations to be made in a unified way by opening new paths for users to connect among two distinct sets: users and items. Similarity among users results from both the *explicit* user-user friendship and the *implicit* user-item networks. Therefore, in case the friendship network fails to capture similarity between two users, the auxiliary user-item network could be used for this task, and vice versa. The combined adjacency matrix $\mathbf{C}$ of Figure 4a is introduced in the following form of four blocks: $\begin{bmatrix} \mathbf{A} & \mathbf{R} \\ \mathbf{R}^T & 0 \end{bmatrix}$. To obtain the combined similarity matrix of Figure 4b, which uses information from both user-user $\mathbf{A}$ and $\mathbf{RR^T}$, we apply Equation 4 to $\mathbf{C}$:

$$Katz(\mathbf{C};\beta) = \beta\mathbf{C} + \beta^2\mathbf{C}^2 + \beta^3\mathbf{C}^3 + \beta^4\mathbf{C}^4 \cdots = \sum_{\ell=1}^{\infty} \beta^\ell \mathbf{C}^\ell \tag{4}$$

The computed attenuation factor for the multi-modal network is β=1/(1+4), equal to 0.2. Unlike we did previously in the bipartite network where we used only the $\mathbf{B}_{12}$ and $\mathbf{B}_{21}$ blocks of the bipartite network, for the multi-modal we exploit information from blocks $\mathbf{C}_{11}$, $\mathbf{C}_{12}$ and $\mathbf{C}_{21}$. Block $\mathbf{C}_{22}$ holds also for the multi-modal network non observed values. The combined version of Katz constructs

	U_1	U_2	U_3	U_4	I_1	I_2
U_1	0	0	1	1	1	1
U_2	0	0	1	0	1	1
U_3	1	1	0	0	0	1
U_4	1	0	0	0	1	0
I_1	1	1	0	1	0	0
I_2	1	1	1	0	0	0

(a)

	U_1	U_2	U_3	U_4	I_1	I_2
U_1	0	0.225	0.379	0.332	0.370	0.379
U_2	0.225	0	0.357	0.106	0.307	0.357
U_3	0.379	0.357	0	0.109	0.169	0.392
U_4	0.332	**0.106**	**0.109**	0	0.313	0.1098
I_1	0.370	0.307	0.169	0.313	0	0.169
I_2	0.379	0.357	0.392	0.109	0.169	0

(b)

Fig. 4. Running Example: (a) Adjacency $\mathbf{C}$ and (b) Similarity Matrix of Multi-modal Social Network

multiple paths using both unipartite friendship and bipartite user-item networks by traversing previously unreached paths between users. Generalization of Katz for $\mathbf{C_{11}}$ user-user block is given by Equation 5 showing such form of paths:

$$Katz(\mathbf{C};\beta)_{11} = \beta\mathbf{A} + \beta^2(\mathbf{A}^2 + \mathbf{R}\mathbf{R}^T) + \beta^3(\mathbf{A}^3 + \mathbf{A}\mathbf{R}\mathbf{R}^T + \mathbf{R}\mathbf{R}^T\mathbf{A}) +$$

$$\beta^4(\mathbf{A}^4 + \mathbf{A}^2\mathbf{R}\mathbf{R}^T + \mathbf{R}\mathbf{R}^T\mathbf{A}^2 + \mathbf{A}\mathbf{R}\mathbf{R}^T\mathbf{A} + \mathbf{R}\mathbf{R}^T\mathbf{R}\mathbf{R}^T)\cdots = \sum_{\ell=1}^{\infty}\beta^{\ell}\mathbf{C}_{11}^{\ell} \quad (5)$$

For instance, the $\mathbf{ARR^T}$ factor shown in Equation 5 contains new traversable length-3 paths of this form: $U_i \xrightarrow{A} U_j \xrightarrow{RR^T} U_k$. Finally, the 4th row of the similarity matrix of Figure 4b indicates that U_3 should be recommended to U_4 as a new friend and not U_2, with similarity value 0.109>0.106. One can observe that both unipartite and multi-modal approaches resulted in the same recommendation, but with much smaller difference after the bipartite network was also considered.

4 Experimental Evaluation

In this section, we experimentally compare the performance of the multi-modal link prediction approach with two other single network algorithms. We want to discover in what extent an auxiliary user-item bipartite network contributes to predicting links in the friendship network. Firstly, we evaluate the combined (cKatz) Katz utility for handling more networks, one *user-user* friendship and one *user-item* network. Then, we employ RWR [18, 21] and Katz algorithm [11] for predicting links in single social networks as comparison partners:

RWR is the well-known Random Walk with Restart algorithm [18, 21] taking into account only one single friendship social network for providing recommendations. In general, RWR considers one random walker starting from an initial node V_x and randomly choosing among the available edges with a probability α. Every time, before random walker makes a choice returns back to the initial node with a probability $1-\alpha$. Similarity among nodes is computed by Equation 6:

$$\mathbf{RWR}(\mathbf{P};\alpha) = (1-\alpha)(I-\alpha\mathbf{P})^{-1} \quad (6)$$

where $\mathbf{I}_n$ is the identity and $\mathbf{P}$ the transition-probability matrix.

sKatz is the model proposed in [11], which takes into account only the single friendship social network, and analyzed in Section 3.1. The proposed approach of this paper *cKatz* considers both the unipartite friendship and the bipartite *user-item* auxiliary network, discussed in Section 3.3.

Parameter's values were tuned as described in [8] and Section 3, therefore α and β for both single network algorithms RWR and sKatz, is set at 0.0008 and 0.0003 for xSocial synthetic and Epinions 49K real data set, respectively. For cKatz parameter β is set at 0.0005 and 0.0003. We employ a fast approximate method of Katz introduced by [8] reducing the computational cost to $O(n+m)$, where n is the number of nodes and m the number of edges, since matrix operations require $O(n^3)$ used by the original Katz algorithm. In this implementation, adjacency matrix is normalized by dividing each entry by the row/column degree. Concerning the maximum length of paths that Katz algorithm employs, we denote ℓ equal to infinite in Equation 2, considering all paths until series convergence. Our experiments were performed on a Core 2 Duo processor with 4 GB of memory. All algorithms were implemented in C. To evaluate the examined algorithms, we have generated synthetic data set using the xSocial generator [7] and chosen one real data set from Epinions web site.

4.1 Real World Networks and Data Sets

Recognizing real-network evolution patterns enables us to better understand the human social behavior and capture similarities among people or about their preferences, detect network intrusions or virus propagation and highlight anomalies [6]. There is a range of patterns that have been identified in real life networks, such as power law distributions [7], six degrees [17](small worlds), scale-free and other log-normal distributions [6], which are powerful tools to mimic observed behaviors. Faloutsos et.al [7] classify graph generators models into emergent (e.g. small-world), where the macro network properties emerge from the micro interactions, and generative graph models, which facilitate a utility function performing recursive iterations until the generated networks meet real network properties.

xSocial Synthetic Data *xSocial Generator* proposed by [7], is a multi-modal graph generator that mimics real social networking sites to produce simultaneously a network of friends and a network of their co-participation. In particular, xSocial builds a network with N nodes performing three independent actions at each step (i.e. write a message, add a friend and comment on a message). A node chooses his friends either by their popularity of by the number of messages on which they have commented together, which is determined by a unique preference value. A node can also follow the updated status of his friends by putting comments on the corresponding new written messages. In our experiments we use xSocial generator to produce simultaneously one explicit friendship and one implicit network of co-comments. In particular, we generated a MSN data set[3] with 100K users and 384K edges among pair of users, in which users contributed 233K messages and 467K comments. The derived MSN for xSocial data set consists of 330K user and item nodes with 852K edges. In Figure 5a we calculated several topological properties for xSocial data set revealing a large clustering coefficient (LCC) equal to 0.2 and small average shortest path length value (ASD) equal to 2.1 discovered mostly in small-worlds networks [17]. Such networks hold sub-networks with connections between most pairs of nodes (i.e. high LLC) which are connected by at least one short path (i.e. small ASD).

Real Data We employ the Epinions 49K[4] data set, which is a who-trusts-whom social network. In particular, users of Epinions.com express their Web of Trust,

[3] `http://delab.csd.auth.gr/~symeon/`
[4] `http://www.trustlet.org/wiki/Downloaded\_Epinions\_dataset`

i.e. reviewers whose reviews and ratings they have found to be valuable. In addition, users are enabled to rate a variety of online items (e.g. books, computers, movies, toys) using a 5 star rating scale. Epinions data set contains 49K users and 487K edges among pair of users, constituting one single friendship social network. Apart from that, it offers a *user-item* network with 140K items and 664K ratings as shown in Figure 5b. In our experiments, we use the whole single network and we keep from the *user-item* network only items rated by users with $r \geq 3$, positively affiliating users with items. Keeping all edges is meaningful in rating prediction tasks, but for friend recommendation this binarization process supports the intuition that we should not recommend users who rated differently similar items. After this, the number of ratings, i.e. edges in *user-item* network, is 570K. The MSN for Epinions 49K data set, when combining the trust and rating network, has 189K nodes of users and items with more than 1M edges. The calculated topological features of the Epinions 49K data set shown in Figure 5a characterize also Epinions 49K as a small-world network with LCC equal to 0.26 and ASD equal to 4. Our evaluation considers the division of friends of

TOPOLOGICAL PROPERTIES OF FRIENDSHIP NETWORKS:
N = total number of nodes
E = total number of edges
ASD = average shortest path distance between node pairs
ADEG = average node degree
LCC = average local clustering coefficient
GD = graph diameter (maximum shortest path distance)
GGS = global graph sparsity (number of zeros in adjacency matrix/ N^2)

Data Set	Type	N	E	ASD	ADEG	LCC	GD	GGS
xSocial 100K	undirected	100000	384458	2.10	6.06	0.20	7	99.99%
Epinions 49K	Directed	49288	487183	4.00	19.77	0.26	14	99.96%

(a)

PROPERTIES OF USER-ITEM BIPARTITE NETWORKS:
N = total number of Nodes (users)
R = total number of Ratings
I = total number of Items
MINR = minimum rating value
MAXR = maximum rating value
AVGR = average rating value
GGS = global graph sparsity (zeros in matrix / existing users x items)

Data-Set	N	R	I	MINR	MAXR	GGS
xSocial 100K	100000	467640	233820	0	0	99.99%
Epinions 49K	49288	664824	139738	1	5	99.98%

(b)

Fig. 5. Topological properties of (a) friendship and (b) bipartite user-item networks

each target user into two sets: (i) the training set $\mathcal{E}^T$ is treated as known information and, (ii) the probe set $\mathcal{E}^P$ is used for testing and no information in the probe set is allowed to be used for prediction. It is obvious that, $\mathcal{E} = \mathcal{E}^T \cup \mathcal{E}^P$ and $\mathcal{E}^T \cap \mathcal{E}^P = \oslash$. Therefore, for a target user we generate the recommendations based only on the friends in $\mathcal{E}^T$. Each experiment has been repeated 30 times (each time a different training set is selected at random) and the presented measurements, based on two-tailed t-test, are statistically significant at the 0.05 level. All algorithms predict the friends of the target users in the probe set. We use the classic precision/recall metric as performance measure for friend recommendations. For a test user receiving a list of k recommended friends (top-k list): *precision* is the ratio of the number of relevant users in the top-k list (i.e. those in the top-k list that belong in the probe set $\mathcal{E}^P$ of friends of the target user) to k. *Recall* is the ratio of the number of relevant users in the top-k list to the total number of relevant users (all friends in the probe set $\mathcal{E}^P$ of the target user). *F1-measure* is the normalized harmonic mean of precision and recall, providing the overall performance metric.

4.2 Combined Katz Sensitivity Analysis

In this section, we examine the sensitivity of the combined and single Katz in terms of accuracy performance when we set different density degree of observed items in the user-item network. We want to identify under which circumstances and to what extend the recommendation task is enhanced when we gradually use auxiliary information from an implicit user-item network.

In particular, we test how the performance of cKatz, a multi-modal network approach, is affected when we keep the fraction of observed friend nodes fixed and gradually increase the fraction of observed items as we select user-items edges randomly. We test both in synthetic and real data sets. Firstly, for the synthetic 100K xSocial data set, we set 5 different density cases (i.e. 0.2, 0.4, 0.6, 0.8, 1) by varying the fraction of observed co-comments, as depicted in Figure 6a, while y-axis holds F1-measure at top-1, which is the average performance of the algorithm in terms of both precision and recall when we recommend only one user. Since, sKatz exploits only the friendship network to provide recommendations,

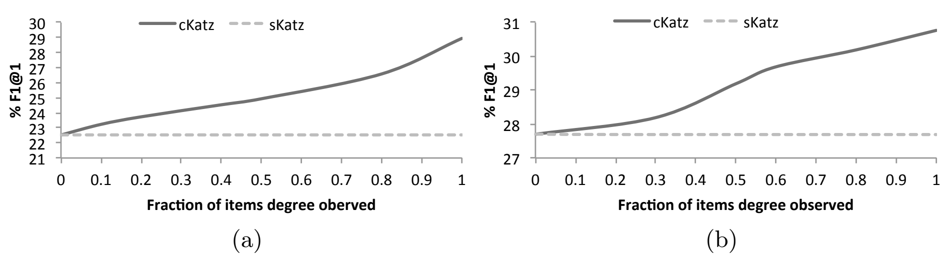

Fig. 6. Comparing cKatz with sKatz Performance in terms of F1-measure at Top-1 vs. fraction of items degree for (a) 100K xSocial synthetic and (b) Epinions 49K data set

increasing the density of the user-item network has no effect in its performance. The fraction of observed friend nodes in the friendship network is fixed to 0.5, where sKatz achieves its best performance. However, cKatz constantly improves its predicting performance as more items from the user-item networks are being observed. We further verify our results in the Epinions 49K real data set shown in Figure 6b. As expected, cKatz improves its overall predicting performance when the fraction of observed items increases. The auxiliary information derived from the affiliation of users with the positively rated items boosts the overall performance, showing that there is fruitful information in the bipartite network. The best performance that sKatz achieves is in 0.5 fraction of observed users, since it does not exploit any auxiliary information and after a certain fraction of friend edges the prediction space of new possible links in the friendship network decreases. Henceforth, we tune the fraction of user nodes observed in 0.5 and this of items observed in 1, for the rest of the experimental evaluation.

Next, we focus on the combined Katz algorithm and we further investigate its performance sensitivity when we vary the number of k recommended friends in the top-k list. We depict the cKatz precision and recall scores versus the varying number of recommended users when applied to the synthetic xSocial and Epinions 49K data sets in Figure 7a and 7b, respectively. In both synthetic and real data sets, cKatz achieves the most accurate scores when we recommend top-1 user. The precision accuracy of cKatz, as expected, gradually decays when we ask for a higher number k of predictions while recall scores increase. Recall

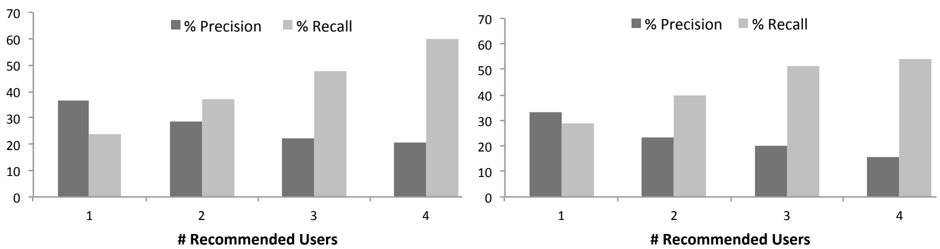

Fig. 7. cKatz Performance in terms of Precision and Recall vs. Top-k for (a) 100K xSocial synthetic and (b) Epinions 49K data set

is the ratio of the number of correct predictions to the number of all the actual friends in the test set. Each user has a different number of actual friends and this indicates the difficulty of getting better predictions as we increase the number of requested recommendations. The average number of friends (ADEG) for xSocial is 6 and for Epinions 49K data set is 19, depicted in Figure 5a for both data sets. Thus, it is more possible that we return correct recommendations in the Epinions 49K data set as we increase k in the top-k list. In Figure 7a and 7b the recall scores versus top-k diagram are depicted with k varying from 1, 2, 3 and 4 for the xSocial and Epinions 49K data set, respectively. In both data sets we observe, as expected, that we get more correct predictions when we ask for more recommendations. When we produce the top-4 list we achieve the best results for both xSocial and Epinions 49K with recall equal to 59,7% and 54,2%, respectively. We would expect that we get better recall scores in the Epinions 49K data set but the average shortest path distance (ASD) is 2 for xSocial and 4 for Epinions 49K, meaning that it is easier to produce more predictions localized in node's neighborhood since we use small values of β.

4.3 Comparison with other Methods

In this section, we conducted the comparison of our multi-modal proposed combined Katz approach with the two other single network comparison partners i.e. sKatz and RWR algorithms, in terms of precision and recall. As the number k of the list varies starting from the top-1 user to top-4, we examine the precision and recall scores. Achieving high recall scores while precision follows with the minimum decline indicates the robustness of the examined algorithm.

For the xSocial synthetic data set, in Figure 8a we visualize the precision vs. recall curve for all three algorithms. As k increases, precision falls while recall increases as expected for all algorithms. cKatz attains the best results achieving the highest precision, outperforming both single network algorithms. This is due the fact that cKatz exploits information from both friendship and the user-item networks. We conduct the same experimental configuration for the Epinions 49K,

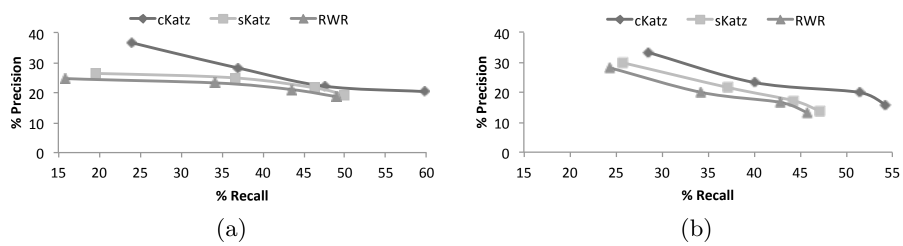

Fig. 8. Comparing cKatz, sKatz, and RWR Performamce in terms of Precision and Recall at Top-k for (a) 100K xSocial synthetic and (b) Epinions 49K data set

shown in Figure 8b real data set to confirm our initial results in the synthetic one. It is clear that cKatz outperforms the two single network partners in terms of both precision and recall, exploiting the user-item auxiliary network. Between the two single network algorithms sKatz performs again better than RWR.

5 Discussion

In this section we discuss several issues concerning the multi-modal network context and our approach. We based our method on path-dependent approaches since they capture the overall structure of the network and can limit their reach to node neighborhood level by using attenuation factors. Furthermore, we understand that weighting strategies are essential to effectively control the contribution of various social networks to the final similarity among users. For us, the main task is to recommend new friends to users by exploiting both explicit and implicit social networks. Therefore, we promote the information derived from the unipartite friendship network and control the contribution of the auxiliary information from the user-item network. In this context, the combined adjacency matrix C takes the following form $C = \begin{bmatrix} A & w\mathbf{R} \\ w\mathbf{R}^T & 0 \end{bmatrix}$, where $w \geq 0$ is the weighting parameter controlling the user-item network contribution to the final similarity.

When w=0, the bipartite social network does not offer any information in the computation of the similarity between users. In this case, the combined Katz behaves like the single version of Katz, sKatz. Earlier in our running example, we observed from the similarity matrix of Figure 2b that when we use information only from the unipartite friendship network we recommend U_3 to U_4. The same result is acquired when using the MSN and matrix of Figure 4b, where U_3 is again recommended to U_4, but with much smaller similarity difference from U_2. However, when we exploit information only from the bipartite user-item network, U_2 is recommended to U_4 as seen in matrix of Figure 3c. Therefore, we understand that the friendship network is in any case important for providing friend recommendations within the friendship domain. However, the contribution of the user-item network could be proven both fruitful, but in some cases also noisy. Parameter w is a factor that could be tuned by either learning the dynamics of the network, or following a specific range according to the recommendation domain, or being adjusted by the user.

Concerning computational issues, our approach is based on a fast approximation of Katz algorithm introduced by [8], who reduce the computational cost to $O(n+m)$ where n is the number of nodes and m the number of edges, since matrix operations require $O(n^3)$ used by the original Katz algorithm [11]. Concerning the maximum length of paths that Katz algorithm employs, we set ℓ equal to infinite taking into account all the paths until the convergence of the series. Nevertheless, wisely tuning ℓ could potentially improve the proposed approach in terms of efficiency by not traversing very long paths. Truncated versions of Katz can reduce the computational cost, but can also improve the efficacy of the recommendations by learning how to avoid uninformative paths [15]. Notice that Katz algorithm can also handle directed graphs.

6 Conclusions and Future Work

In this paper, we presented an extended framework exploiting multi-modal social networks to provide friend recommendations. We experimentally showed that implicit information can be proven fruitful for the friend recommendation task. In the future, MSNs will allow us to perform more cross-domain recommendation tasks, but will also raise challenges like scaling, the effective weighting of multiple information sources and the exploitation of semantic information.

References

1. Adamic, L., Adar, E.: Friends and neighbors on the web. Social Networks 25(3), 211–230 (2003)
2. Adamic, L., Adar, E.: How to search a social network. Social Networks 27(3), 187–203 (2005)
3. Cha, M., Mislove, A., Gummadi, K.: A measurement-driven analysis of information propagation in the flickr social network. In: 18th International World Wide Web Conference (WWW). pp. 721–730. Madrid, Spain (2009)
4. Chen, J., Geyer, W., Dugan, C., Muller, M., Guy, I.: Make new friends, but keep the old: recommending people on social networking sites. In: 27th International Conference on Human Factors in Computing Systems (CHI). pp. 201–210 (2009)

5. Davis, D., Lichtenwalter, R., Chawla, N.V.: Multi-relational link prediction in heterogeneous information networks. In: Advances in Social Networks Analysis and Mining (ASONAM), 2011 International Conference on. pp. 281–288. IEEE (2011)
6. Du, N., Faloutsos, C., Wang, B., Akoglu, L.: Large human communication networks: patterns and a utility-driven generator. In: 15th ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 269–278 (2009)
7. Du, N., Wang, H., Faloutsos, C.: Analysis of large multi-modal social networks: patterns and a generator. In: Machine Learning and Knowledge Discovery in Databases, pp. 393–408. Springer (2010)
8. Foster, K.C., Muth, S.Q., Potterat, J.J., Rothenberg, R.B.: A faster katz status score algorithm. Computational & Mathematical Organization Theory 7(4), 275–285 (2001)
9. Golbeck, J.: Personalizing applications through integration of inferred trust values in semantic web-based social networks. In: Semantic Network Analysis Workshop at the 4th International Semantic Web Conference. vol. 16, p. 30 (2005)
10. Guy, I., Ronen, I., Wilcox, E.: Do you know?: recommending people to invite into your social network. In: 13th International Conference on Intelligent User Interfaces (IUI). pp. 77–86 (2009)
11. Katz, L.: A new status index derived from sociometric analysis. Psychometrika 18(1), 39–43 (1953)
12. Liben-Nowell, D., Kleinberg, J.: The link-prediction problem for social networks. In: Conference on Information and Knowledge Management. pp. 556–559 (2003)
13. Lo, S., Lin, C.: Wmr: a graph-based algorithm for friend recommendation. In: IEEE/ACM International Conference on Web Intelligence. pp. 121–128. Hong Kong, China (2006)
14. Lü, L., Zhou, T.: Link prediction in complex networks: A survey. Physica A: Statistical Mechanics and its Applications 390(6), 1150–1170 (2011)
15. Lu, Z., Savas, B., Tang, W., Dhillon, I.S.: Supervised link prediction using multiple sources. In: IEEE 10th International Conference on Data Mining. pp. 923–928 (2010)
16. Massa, P., Avesani, P.: Trust-aware collaborative filtering for recommender systems. In: Federated Int. Conference On The Move to Meaningful Internet: CoopIS, DOA, ODBASE. pp. 492–508 (2004)
17. Milgram, S.: The small world problem. Psychology Today 22, 61–67 (1967)
18. Pan, J., Yang, H., Faloutsos, C., Duygulu, P.: Automatic multimedia cross-modal correlation discovery. In: 10th ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 653–658 (2004)
19. Scholz, C., Atzmueller, M., Barrat, A., Cattuto, C., Stumme, G.: New insights and methods for predicting face-to-face contacts. In: 7th International AAAI Conference on Weblogs and Social Media (ICWSM) (2013)
20. Symeonidis, P., Tiakas, E., Manolopoulos, Y.: Transitive node similarity for link prediction in social networks with positive and negative links. In: 4th ACM conference on Recommender systems. pp. 183–190. ACM (2010)
21. Tong, H., Faloutsos, C., Pan, J.Y.: Fast random walk with restart and its applications. In: 6th International Conference on Data Mining. pp. 613–622. ICDM '06, IEEE Computer Society, Washington, DC, USA (2006)
22. Vasuki, V., Natarajan, N., Lu, Z., Savas, B., Dhillon, I.: Scalable affiliation recommendation using auxiliary networks. ACM Trans. Intell. Syst. Technol. 3(1), 3:1–3:20 (2011)