



Research Repository UCD

Title	A Framework for Enterprise Social Network Assessment and Weak Ties Recommendation
Authors(s)	Ghaffar, Faisal, Buda, Teodora Sandra, Assem, Haytham, Afsharinejad, Armita, Hurley, Neil J.
Publication date	2018-10-24
Publication information	Ghaffar, Faisal, Teodora Sandra Buda, Haytham Assem, Armita Afsharinejad, and Neil J. Hurley. "A Framework for Enterprise Social Network Assessment and Weak Ties Recommendation," October 24, 2018. https://doi.org/10.1109/ASONAM.2018.8508292 .
Conference details	The 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2018), Barcelona, Spain, 28-31 August 2018
Item record/more information	http://hdl.handle.net/10197/11406
Publisher's statement	© 2018 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.
Publisher's version (DOI)	10.1109/ASONAM.2018.8508292

Downloaded 2025-12-04 23:04:04

The UCD community has made this article openly available. Please share how this access benefits you. Your story matters! (@ucd_oa)



© Some rights reserved. For more information

A framework for enterprise social network assessment and weak ties recommendation

Faisal Ghaffar, Teodora Sandra Buda, and Haytham Assem
IBM Ireland
Email: {faisalgh,haythama,tbuda}@ie.ibm.com

Armita Afsharinejad, Neil Hurley
Insight Center UCD Ireland
Email: {neil.hurley,armita}@ucd.ie

Abstract—Sociological theories of career success provide fundamental principles for the analysis of social links to identify patterns that facilitate career development. Some theories (e.g. Granovetter’s Strength of Weak Ties Theory and Burt’s Structural Hole Theory) have shown that certain types of social ties provide career advantage to individuals by facilitating them to access unique information and connecting them with a diverse range of others in different social cliques. The assessment of link types and prediction of new links in the external social networks such as Facebook and Twitter have been studied extensively. However, this has not been addressed in the enterprise social networks and especially the prediction of weak ties in the context of employee career development. In this paper, we address this problem by proposing an Enterprise Weak Ties Recommendation (EWTR) framework which leverages enterprise social networks, employee collaboration activity streams and the organizational chart. We formulate weak ties recommendation as a link prediction problem. However, unlike any generic link prediction work, we first validated explicit enterprise social network with a set of heterogeneous implicit collaboration interaction networks and show assessment improves the network’s effectiveness in predicting new links. Furthermore, we leverage assessed network for weak ties prediction by incorporating organizational chart information to optimize link prediction methods. We demonstrate that optimization improves prediction accuracy in terms of AUC and precision and our characterization of weak ties to a certain extent aligns with Granovetter’s and Burt’s seminal studies.

I. INTRODUCTION

Online Social Networks (OSNs), e.g. Facebook, twitter and LinkedIn help us in connecting and sharing content with each other in our personal and professional lives. Recently, with the adoption of social networking platforms inside enterprises, known as *Enterprise Social Networks* (ESNs), both employees and employers can benefit from the collaboration and communication data. For example, ESNs can help employees to search for experts [1], share information and gain access to others’ expertise for intrapreneurship [2], [3] in the organization, whereas the employer may leverage ESNs for enhancing customer experience, business performance [4] and forming team of experts for a particular task or project [5]. Examples of such ESNs are IBM Connections¹ and Yammer².

Social links are key not only to the ESNs but also to employees’ professional development in the workplace. Social ties are considered to have an effect on an individual’s career development [6]. A study by [7] found that individuals at

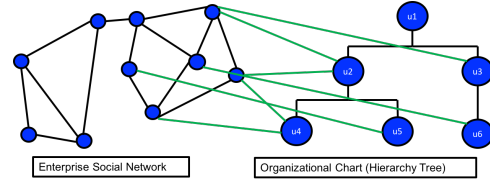


Fig. 1: An example of hybrid aligned enterprise networks

bridging positions in their social networks were more likely to be promoted than individuals who were not at the bridging positions. Similarly, [6] examined the social relationships maintained by individuals in the organization and found that having more contacts in other functions and having more weak ties gives access to unique resources relevant for the instrumental objectives of career success. However, social links in the ESN must be the true representation of social relationships between individuals and must not suffer from noisy links, i.e., links that do not reflect a real relationship which prevent accurate analysis of the network. Hence, a process of assessing the links in the ESN is crucial before leveraging it to derive any meaningful insights.

Granovetter [8] formulated the theory of strong and weak ties, where strong ties correspond to high frequency of interactions and weak ties correspond to acquaintances. In terms of information networks, it is well understood that weak ties are more likely the source than the strong ties in terms of accessing unique information due to their structural ability to connect different parts of the network [9]. According to Granovetter, weak ties are often bridge between densely interconnected circles of cliques. Applying weak ties theory in context of ESN and employee career development, our institution is that assessing employee’s ego network for weak ties with consideration of organizational chart *might* be of more interest to employee in the organization. For example, an employee may not be able to discover a potential collaboration opportunity or new job opening in the other business unit from her close friends, as they tend to possess same information. Instead, she may find connecting with others more appealing, who are similar to her but prefer interdisciplinary collaboration. Following Granovetter’s theory [8] and [6] these link recommendations in the ESN should help individuals connect across functions and form bridging positions.

Challenges and contributions: As weak ties are key for

¹<https://www.ibm.com/uk-en/marketplace/enterprise-social-collaboration>

²<https://www.yammer.com/>

accessing unique information and career success in the organization [6], [8], social link recommendation based on weak ties in the ESNs is significantly different from that in OSNs, and has following unique challenges:

- Socializing at workplace is different than socializing in personal life because of the professional context [10], but social behaviors at the workplace have not been well explored [11]. Individuals connect with each other based on their interactions which are specific to their working environment in the organization.
- In workplace settings, there is unique information available related to an individual's business unit and company's organizational chart. Such information can play a crucial role in assessing an employee's social network for the purpose of identifying weak ties.
- There are multiple sources of social interactions available through employee collaboration which can provide additional patterns of social link formation. Employees could be connected through direct(explicit) as well as indirect(implicit) social relationships based on their common interests or collaboration activities in the organization.

In this paper, we investigate the problem of assessing an explicit social network and identifying workplace weak ties for an employee's career development. More specifically, we propose a Enterprise Weak Ties Recommendation (EWTR) framework to recommend weak ties based on the employees social activity, their interactions in a an enterprise collaboration platform, their position in the organizational hierarchy. The framework involves two phases:

- **Link assessment & weak tie candidates generation**, where we train a set of learning algorithms to categorize social relationships between "real" and "noisy" ties;
- **Candidate refinement and weak ties recommendation**, where we incorporate organizational structure information to link prediction methods and propose a set of topological-based optimized measures for weak ties prediction and ranking.

II. PRELIMINARIES

In this section we first give preliminary definitions of our ESN and *hierarchy tree*. We frame the problem with a set of exogenous interaction networks (e.g., interactions over a blogging application, Communities application, file sharing application on the enterprise collaboration platform), an enterprise social network and a tree of the organization hierarchy.

Enterprise social network: Let $G_{SN} = (V, E)$, $E \subseteq V \times V$ denote an attributed social network in the organization, where V represents a set of employees and E represents a set of explicit social relationships between employees. Let $w : E \rightarrow \mathbb{R}$, be a real-valued edge weight, representing the strength of the link, which we write as $w(x, y)$, where $x, y \in V$ and $(x, y) \in E$. Furthermore, let $\Gamma(x)$ represent the neighborhood of $x \in V$, i.e. all nodes connected to x by an edge.

Implicit Interactions Network: Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{W})$ be an implicit interactions network which is derived from the enterprise applications (e.g. blogging, file sharing, etc.) other

than the explicit social networking application, where $\mathcal{V} \subseteq V$ and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is a set of exogenous interactions between employees; \mathcal{W} is an $|\mathcal{E}| \times d$ attribute matrix associated with edges in \mathcal{E} with each row corresponding to an edge, each column an attribute, and an element x_{ij} denoting the j^{th} attribute of edge e_i .

Organizational hierarchy: The *organizational hierarchy* of a company can be represented as a *rooted tree* [11] such that $\mathcal{HT} = (\mathcal{N}, \mathcal{L}, root)$, where \mathcal{N} is the set of employees, \mathcal{L} represents the set of directed links from manager to subordinate in \mathcal{HT} and root represents the head of the business unit.

A. Learning Task

Given a set of implicit interaction networks $\{\mathcal{G}_1, \mathcal{G}_2, \dots, \mathcal{G}_n\}$, explicit ESN G_{SN} , and organization hierarchy \mathcal{HT} . For a particular employee $v_i \in V$ the goal is to learn a predictive function such that we can infer and suggest weak ties between v_i and other employees in network G_{SN} , i.e.,

$$f(v_i | G_{SN}, \mathcal{G}, \mathcal{HT}) = Y \quad (1)$$

where $Y = \{y_1, y_2, \dots, y_{|V|}\}$ is a set of inferred weak ties between employee v_i and other employees in the the network G_{SN} ; $y_k \in \{0, 1\}$ is binary score indicating whether employee v_k will have a weak tie with v_i at future time t' . The problem formulation is different from existing work on link prediction [12], [13], which focuses only on predicting a link at a point in time in future without classifying whether it is a weak tie or not. It is also different from existing work on the role of weak ties in link prediction [14], [15] in that although these studies show weak ties play an important role in the link prediction but fail to predict the weak ties themselves. Here, we mainly consider employees implicit interaction networks to assess their explicit social network and then determine social plus organizational overlap which controls the similarity between nodes when predicting weak ties.

B. Homophily based similarity metrics

Homophily is the tendency for similar individuals to connect with each other. Similarity between two nodes in a network can be defined by a score S_{xy} which represents the strength of the tendency for a pair of nodes x and y to link. The higher the score, the more likely it is for the pair to form a link. Due to undirected edges in graph G_{SN} , the adjacency matrix A is symmetric, that is, $S_{xy} = S_{yx}$. Ordering unconnected pairs in descending order of the similarity score, gives an ordered list of tie candidates. In the following, weighted versions of a set of most common homophily measures are chosen as similarity scores. All these scores are "common neighborhood" based which means the similarity between nodes x and y will be high if number of common neighbors between them is high.

- *CN index*: The basic idea behind this index is that two individuals are more similar if they have many common neighbors. The weighted form of the CN index between two nodes x and y is defined in [16] as:

$$S_{xy}^{WCN} = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{w(x, z) + w(z, y)}{2} \quad (2)$$

- *Adamic-Adar index (AA)*: This similarity index refines the contribution of each node by taking into account node degree. It assigns less weight to high-degree nodes. Its definition is as follows:

$$S_{xy}^{WAA} = \sum_{\substack{z \in \Gamma(x) \cap \Gamma(y) \\ z \neq x \neq y}} \frac{w(x, z) + w(z, y)}{\log(1 + \sum_{c \in \Gamma(z)} w(z, c))} \quad (3)$$

- *Resource Allocation Index (RA)*: RA is a revision of AA that provides better performance than AA in link prediction [17]. It further reduces the contribution of high-degree nodes and is defined as:

$$S_{xy}^{WRA} = \sum_{\substack{z \in \Gamma(x) \cap \Gamma(y) \\ z \neq x \neq y}} \frac{w(x, z) + w(z, y)}{\sum_{c \in \Gamma(z)} w(z, c)} \quad (4)$$

- *Information Diffusion based Index (INFO DIFF)*: This characterizes the weight of a tie based on the nodes' ability to diffuse information in the network [15]. It was observed that deleting ties with lower value of this index dropped information coverage sharply [15]. It is defined as:

$$S_{xy}^{WINFO} = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{w(x, z) + w(z, y)}{k_x + k_y - 2 - (w(x, z) + w(z, y))} \quad (5)$$

where k_x is the (unweighted) degree of x .

- *Preferential Attachment (PA)*: It indicates that new links will be more likely to connect high-degree nodes than low degree ones. PA based similarity is defined as:

$$S_{xy}^{WPA} = \sum_{x' \in \Gamma(x)} w(x, x') \times \sum_{y' \in \Gamma(y)} w(y, y') \quad (6)$$

- *Hub Promoted Index (HPI)*: This measure finds the similarity between two nodes in networks with a hierarchy structure. Its value is determined by the lower degree of nodes.

$$S_{xy}^{WHPI} = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{w(x, z) + w(z, y)}{2 \times \min(\sum_{x' \in \Gamma(x)} w(x, x'), \sum_{y' \in \Gamma(y)} w(y, y'))} \quad (7)$$

- *Sorensen-Dice Index (SD)*: Besides the size of common neighbors, this measure considers lower degree nodes might have higher link likelihood. It was introduced in ecology to measure association between species [18] and is defined as:

$$S_{xy}^{WSD} = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{w(x, z) + w(z, y)}{\sum_{x' \in \Gamma(x)} w(x, x') + \sum_{y' \in \Gamma(y)} w(y, y')} \quad (8)$$

III. PROPOSED FRAMEWORK

In this section we introduce the Enterprise Weak Ties Recommendation (EWTR) framework in detail. We will exploit the enterprise social network, employee's activities on the collaboration platform, and the organizational hierarchy to understand and model who an employee should connect to within a certain part of the organization such that the

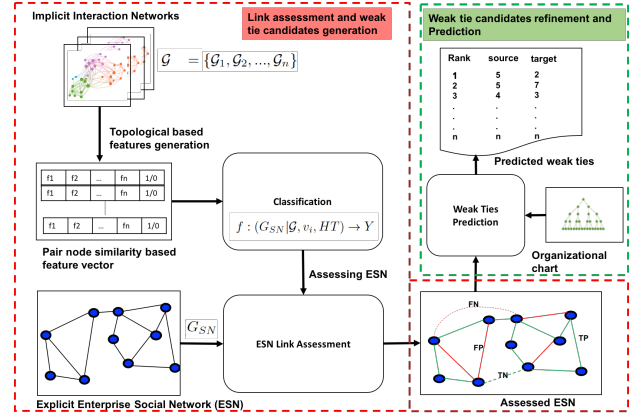


Fig. 2: Enterprise weak ties recommendation framework

recommended link can benefit the employee in her career development.

We formulate the problem as weak ties prediction task and propose a framework for weak ties recommendation. An overall flow of the processes in EWTR is shown in Figure 2, more specifically, our framework has two phases:

A. Link Assessment & Candidates Generation

Our EWTR framework takes the following steps to assess a given ESN and to generate a weak ties candidate list for each member of the ESN:

- 1) Crawl a set of heterogeneous collaboration applications to extract employees social activities as bipartite networks of employees and artifacts (e.g. blogs, wikipages, communities).
- 2) Perform one mode projection on the bipartite networks to construct *implicit affinity networks* $IANs = \{G_1, G_2, G_3, \dots, G_n\}$ among employees such that there is an edge between two employees if they have performed the same activity (e.g., liked/followed/tagged same blog/wikipedia/community). The weight of the edges is determined by the frequency of performing the same action by both nodes.
- 3) Compute the similarity between nodes for each edge in the G_i using the homophily measures defined in equations from (3) to (8).
- 4) Train a set of machine learning models for each of the G_i where the input features to each of the models consists of a vector of similarity measures for each pair of nodes and the prediction target is a binary variable (0/1) which indicates whether there exists a link between nodes in G_i or not. The output of this step will be a network $G_{pred} = (V, E')$ with set of nodes V and edges E' .
- 5) Assess the links in the *explicit* ESN $G = (V, E)$ by comparing it with the predicted network $G_{pred} = (V, E')$. The false-positive links $(E' - E)$, i.e. those links that were observed in G_{pred} and do not exist in G are our potential candidates for weak ties.

If we consider the informal social interactions that happen in our daily lives ultimately lead to the formation of explicit

social links as argued by [19], [20], then we can think of implicit links as weak ties before they get realized and become explicit. Based on this consideration, we build our intuition behind generating potential weak ties such that all implicit links are weak ties unless they become explicit. The candidates generated by this phase of our EWTR framework will have high probability that a randomly chosen activity performed by an employee is also the activity performed by the other connected employee in each link of the set $E' - E$. We evaluate our consideration in the next phase of our framework by leveraging potential candidates for link prediction.

B. Candidate Refinement and weak ties prediction

The link assessment phase of the EWTR framework resulted an assessed network (i.e. network with links removed if they were not observed in implicit interaction networks and added otherwise). The links that were observed in the implicit interactions network but are not present in explicit enterprise social network are potential candidates for weak ties recommendation. In this next phase, our EWTR framework further refines the potential candidates for weak ties by optimizing the basic similarity metrics with a measure that characterizes the link as a weak tie from an information benefits perspective. First it calculates a structural overlap between a pair of nodes by considering Granovetter's Strength of Weak Ties theory [8] and Burt's Structural Hole Theory [9] from the information benefits perspective. Secondly it combines the overlap measure with similarity metrics to predict the likelihood of a pair of nodes being connected via a weak tie.

1) *Social-organizational Overlap*: Granovetter's Strength of Weak Ties suggests that it is the weak ties that are sought for accessing unique information. From the perspective of information benefits, Burt's Structural Hole theory argues that non-redundant contacts (i.e. two nodes which are not directly connected with each other) provide network benefits that are to some degree additive rather than overlapping [9]. Therefore, for information benefits it is desirable for a node to increase its number of non-redundant contacts in the network. From the weak-ties point of view, this means reducing the overlap between contacts of the endpoints of a tie is desirable. However, *homophily* suggests that two nodes have high likelihood of being connected if they have high overlap between their contacts. To resolve this contradicting situation, we propose a measure named social-organization overlap (SOO) for which we leverage the organizational chart along with the enterprise social network to measure the structural overlap between a pair of nodes and their teams in the organization.

The organizational chart is useful in that it represents employees in their teams in the organization and SOA measures the proportion of node's team members that are already connected to the other node in the enterprise social network. It is an average measure for both nodes and in a sense, it indicates the redundancy in contacts of two nodes x and y in the workplace settings by measuring the ratio of their contacts that are also their team members to the size of their teams. Let α be the SOO between nodes i and j $m(i)$ and $m(j)$ denote

the teams of both nodes in the *HT*. If $\Gamma(i), \Gamma(j)$ are neighbors of i and j respectively then the social-organizational overlap α is defined as:

$$\alpha = \frac{2 \times (|\Gamma(i) \cap m(j)| + |\Gamma(j) \cap m(i)|)}{(|m(i)| + |m(j)|)} \quad (9)$$

2) *Optimized topology-based metrics for weak ties prediction*: Granovetter's definition of *Strength of Weak Ties* [8] introduced the concept of a *bridge* as "the only path between two nodes in the network such that there is no alternative route along which information or influence can flow from any contact of one node to any contact of the other node". From this network topological structure based definition of a bridge, it emerges that all *bridges are weak ties* which loosely connect two individuals who belong to different areas of the network. With this, a weak tie can be re-defined from the network community structure point of view as *a link whose terminal nodes belong to two different communities* [21].

The organizational chart divides the employees into well-defined communities in terms of teams. We adapt the definition of weak ties from network community structure point of view as in [21] for our hybrid aligned formal (organizational chart) and informal (enterprise social network) networks. We determine the contribution a common neighbor makes in predicting a weak tie between a pair of nodes depending on the community belonging of the common neighbor with respect to a pair of nodes. Let β denote the contribution of a common neighbor z between a pair of nodes x and y . Then β can be measured by adapting community-based model defined in [22] with teams of nodes x and y as communities. In contrast to [22], we adapt a community-based model to enhance the inter-community links and penalize intra-community links. Hence, highlighting the weak ties as per the definition in [21], it is defined as:

$$\beta = \begin{cases} 2, & \text{if } z, x, y \text{ in same teams,} \\ 1 + \alpha & \text{if } z \text{ in same team with } x \text{ or } y, \\ 2 * \alpha & \text{if } z, x, y \text{ in different teams.} \end{cases} \quad (10)$$

Critically β stands for existing overlap of common neighbor z with team members of x and y . It is used to control the contribution z makes when predicting a weak tie between x and y . With β overlap contribution, optimized topological metrics penalize the z 's contribution if it is in the same team with x and y and there is an overlap, i.e. $\alpha > 1$ and enhances the contribution if z belongs to different team to x and y , i.e. $\alpha < 1$, hence highlighting the inter-teams links as weak ties. The topological metrics degenerates to their base form when respectively when $\beta = 1$. In order to switch a link prediction task to weak tie prediction, we incorporate common neighbor's contribution in terms of it's overlap with endpoint nodes. We optimize the weighted forms of common neighborhood based topological metrics WAA and WRA defined in equations 3 and 4 respectively. Optimized similarity between a pair of nodes named here OWAA and OWRA respectively, are defined as:

$$S_{xy}^{OWAA} = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{w(x, z)^\beta + w(z, y)^\beta}{\log(1 + \sum_{c \in \Gamma(z)} w(z, c))} \quad (11)$$

$$S_{xy}^{ORA} = \sum_{\substack{z \in \Gamma(x) \cap \Gamma(y) \\ z \neq x \neq y}} \frac{w(x, z)^\beta + w(z, y)^\beta}{\sum_{c \in \Gamma(z)} w(z, c)} \quad (12)$$

Similarly rest of the metrics defined in section II-B can be optimized by incorporating the common neighbor's contribution in characterizing the links being predicted as weak ties. In this paper, we leverage only OWAA and OWRA in our experiments for predicting the weak ties.

IV. EXPERIMENTS

To evaluate the proposed techniques for enterprise weak ties recommendation, we conducted extensive experiments on a number of implicit and explicit real-world networks obtained from an enterprise social and collaboration platform called IBM Connections³ which was deployed in a large organization for internal use of employees. The platform consists of a number of different applications that supports employee networking and facilitates collaboration among employees. The social networking application allows employees to reciprocally connect (e.g. become friends), tag, mention and follow each other. In addition to direct interactions, the platform also supports employees collaboration through following social applications: a **blogging** application which allows employees to create blog posts, 'like', or comment on others' posts; a **file sharing** application; a **tagging** application that allows employees to tag each other with a descriptive label; and an application to create **wikis** where content can be co-authored by multiple employees. In a way, IBM Connections supports "explicit" and "implicit" social networks through direct and indirect interactions respectively.

A. Data

We conducted our experiments in a large multinational organization using following datasets. Employee identities were aligned across all the datasets and combined dataset was anonymized using non-reversible MD5 transformation.

- 1) *Enterprise social network (ESN)* - This is a "friendship" social graph of employees in the enterprise and it has more than 100k⁴ nodes and millions of edges between them.
- 2) *Collaboration activity streams* - Second dataset contains a large number of activities performed by employees on an *Enterprise collaboration platform - IBM Connections*, for a period of over 2.5 years (Jan 2014 - June 2016). Collaboration activities include "creating/commenting/linking" a blog, "joining/following" a community, "tagging/following" others, "creating/commenting/linking" someone's status update.
- 3) *Organizational Chart* - In addition to social network and the collaboration data, we also had access to employee's organizational chart(HT) from the company's

³<https://www.ibm.com/us-en/marketplace/ibm-connections>

⁴We are not able to reveal actual numbers here and throughout the paper for commercial reasons

TABLE I: The basic topology features of implicit and explicit ESNs. N denotes total number of nodes in the network. M is the number of links in the network. $\langle k \rangle$ represents the mean value of node degrees, C indicates mean clustering coefficient of the network.

Network	N	M	$\langle k \rangle$	C
Blogging (implicit)	67898	4134607	121.78	0.597
Community (implicit)	51130	1063503	41.6	0.034
File Sharing (implicit)	1300995	6221974	9.56	0.023
Wiki-Pages (implicit)	62831	2642472	84.11	0.412
Forum (implicit)	47506	3230957	136.02	0.587
Status Updates (implicit)	72383	1483368	40.98	0.434
Friendship (explicit)	272723	6086478	22.31	0.297

Human Resources department which contained all active employees in the organization till June 2016.

B. Experiment Settings

Train Test dataset splits: The implicit and explicit social networks dataset were divided into two parts: the training set E^{TR} contained social networks for first two years (2014 and 2015) and the testing set E^{TS} consisted of social network for 6 moths of the year 2016 in a such a way that $E = E^{TR} \cup E^{TS}$ and $E^{TR} \cap E^{TS} = \phi$. To further evaluate algorithms, a k -fold cross validation approach was adopted, which randomly divides all links into k subsets () and trains model with $k - 1$ subsets while test it with one left over set. This process is repeated k time, with each of subset used exactly one as the testing set. In this paper, we used 10-fold cross validation.

To evaluate the effectiveness of EWTR framework, we compared our results with baselines for both phases of the framework. The baselines are the traditional weighted forms of link prediction methods described in equations from 2 to 8 of section II-B. We summarize all the comparison methods as follows:

- 1) Baseline scores calculated on explicit social network before the assessment of links are compared with scores calculated after the assessment. This can be used as an indication of the effectiveness of link assessment.
- 2) Comparison of similarity measures optimized for weak ties prediction with the baseline measures.
- 3) Social-organizational overlap compared with performance measures to demonstrate it's effects on prediction.

C. Evaluation Metrics

To quantify the model evaluation, we adopted two measures in this paper, AUC and precision due to the fact that just like many real-world social networks, our ESN is extremely imbalanced in terms of ratio between exiting and non-existing links. AUC can be interpreted as probability of randomly chosen missing link has higher score than randomly chosen non-existing link. AUC can be calculated as defined in [23]:

$$AUC = \frac{n' + 0.5n''}{n} \quad (13)$$

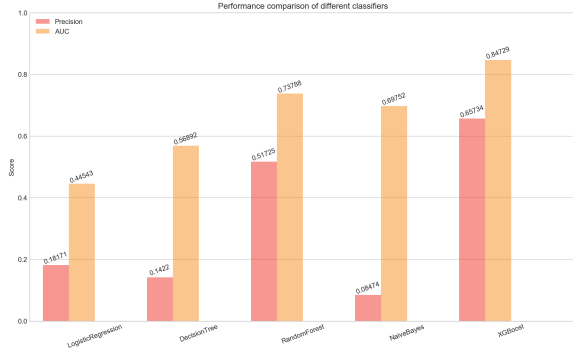


Fig. 3: AUC and precision performance comparison of different classifiers

where n' is number of times missing links have high score, and n'' number of times missing links and non-existent links have same score out of n comparisons.

Precision is the ratio of correctly predicted links to the number of selected links. All the missing links are first ranked in decreasing order according to their predicted score. If l is set of links predicted set out of a sample (let's say L) of top ranked links, then precision is [23]:

$$precision = l/L \quad (14)$$

V. RESULTS AND DISCUSSION

In this section, we first show the performance similarity based learning of new links nodes in our EWTR framework's link assessment phase. Then we compare performance of a set of baseline link prediction methods for explicit social network before and after the link assessment.

A. Effect of link Assessment on link prediction

We train a set of binary classifiers for the interactions network derived from collaboration applications such as blogging, file, community, timeline, where input features to the classifiers are the similarity scores defined in section II-B between pairs of nodes and target is the binary variable representing whether or not there exists a link between nodes in the explicit social network of employees. The classifiers that were trained include Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), Naive Bayes (NB) and XGBoost (XGB). Precision and AUC performance results for each of the classification run are compared for the trained models and comparison for one run with aggregated interaction network is shown in Figure 3. Out of all the classifiers XGB performed better consistently with all the interaction networks on the test dataset. Hence, we selected XGB as our final model.

Our experiments show interesting results for action based enterprise implicit social networks and their ability to predict explicit links among employees. Results demonstrated that certain implicit interaction such as "like/follow" same blog and wikipedia in the *Blog* and *wiki* application have high predictability than others like "share/like/tag" a file in *File* application. Performance results are given in Table II with

TABLE II: AUC and precision performance comparison of various **implicit** interaction networks in assessing links in explicit ESN

Training Implicit Interaction Networks	Testing Network	AUC	Precision
Blogging application	Enterprise Social Network	0.782	0.487
Wiki application		0.708	0.625
File Sharing application		0.624	0.25
Community Application		0.72	0.44
Forum Application		0.67	0.26
Timeline posts		0.605	0.773
Aggregated Network		0.847	0.657

TABLE III: AUC and Precision comparison of baseline methods before and after the link assessment

	ESN before Assessment		ESN after Assessment	
	AUC	Precision	AUC	Precision
AA	0.9061	0.9096	0.9075	0.91107
PA	0.8852	0.8886	0.8847	0.8886
RA	0.9073	0.9107	0.9077	0.9115

bes results in bold across both AUC and precision. The blogging application's accuracy in predicting a link in the *explicit* ESN is 0.782 in terms of AUC whereas it is 0.624 for the *File* application. Similar is true for precision which is 19.2% high in case of blogging application when compared for the same employees in the *File* application. In addition to separate application level implicit interactions, we also used their aggregated implicit networks to assess the links in the *explicit* social network and in contrary to [19], aggregated interaction networks outperformed in our case for both AUC and precision measures. This indicates that a more complete view of employee's collaboration activity can predict explicit link between two employees with higher accuracy.

Furthermore, our experiments identified 2.29% of overall links as "noisy" which are characterized in this paper as the links which exist in explicit ESN but are not observed in implicit social interactions. Results from these experiments were used to correct the links in the *explicit* ESN which was further used in predicting weak ties.

B. Effect of social-organizational overlap on link prediction

The resultant network of previous phase is used by our EWTR framework predict weak ties for an employee in the organization. For this purpose, we incorporated organizational hierarchy information in our link prediction methods. We redefined two of the common neighborhood based link prediction measures (WAA and WRA) by incorporating the *social-organizational overlap* derived from employee's hierarchy tree as defined in section III-B1. The original WAA and WRA measures lack the considerations of existing overlap between common neighbor with professional community (i.e. team in the *HT*) of either of the endpoint nodes.

For community based weak tie, the endpoints of a link must be in different communities. Based on this intuition, we leverage overlap between common friend and the team members of endpoint nodes to penalize the contribution of

the common neighbor for not being in different community(i.e team) to that of endpoints. In summary, our proposed measure of similarity for link prediction enhances the contribution by common neighbor in case of common neighbor is less known to teammates of endpoints, (i.e. $\alpha < 1$) and penalizes the contribution otherwise. In doing so, our framework highlights and predicts weak ties i.e. the inter-team links while acknowledging the existing overlap between employee's connections and teammates.

Figures 4 and 5 present the AUC and precision respectively with respect to organizational overlap varying in range [0,2]. The two curves denote the performance of optimized form of AA (OAA) and optimized form of RA (ORA) respectively. As shown in figures, for both measures a single peak appears when the *overlap* (α) *leq* 1. The precision values are optimal when α is between 0.50 and 0.75. However, optimal AUCs for both OAA and ORA are obtained at 1.1. Besides this, every curve appears descending trend when $\alpha > 1$. The appearance of peak when α is less than one, implying that penalizing the contributions of common neighbors within same team of the nodes and enhancing the contribution of common neighbors in different team with nodes, our OAA and ORA outperform their corresponding baseline forms. In doing so, this phenomenon confirms our hypothesis that the common neighbors that are in different community to those of endpoint nodes contribute more to the connection likelihood for weak-ties. In other words, ORA and OAA measures highlight the importance of links that connect different teams and hence, can be characterized as the weak-ties according to [21].

1) *Comparison with baselines*: Experiments with optimized AA and RA, demonstrate that overall AUC and precision were improved with OAA and ORA as shown in Table IV (improved results are in bold) as compared to baselines with the exception of precision for ORA. All the AUCs and precisions were averaged across under 10 independent runs with different random data set divisions.

The results presented in Figure 4-5 and Table IV have shown that by considering the overlap contribution of common neighbors in predicting can improve the accuracy of different topological measurements on the real-world networks. Furthermore, optimized measures can highlight the links that are crucial for control and diffusion of information in the network. There are two reasons why optimized forms have higher accuracy and predict weak-ties. First, by penalizing the common neighbors belonging to same communities with nodes, EWTR framework differentiates some common neighbors contributions. Second, the overlap factor α is crucial in that it is agnostic to network's topological structure and can flexibly adapt to different structures by identifying the underlying communities and resulting in better predictive adaption than baselines.

VI. CONCLUSIONS AND FUTURE WORK

In this paper, we presented a framework for recommending weak ties to an employee at workplace using enterprise social networks, employee collaboration activity streams and the

TABLE IV: Performance comparison of optimized AA and RA with baseline

Evaluation Metric	AA	OAA	RA	ORA
AUC	0.9075	0.9139	0.9077	0.9096
Precision	0.9110	0.9107	0.9115	0.9065

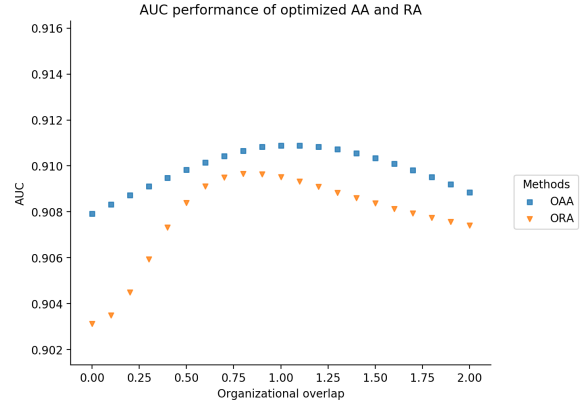


Fig. 4: Effects of organizational overlap in AUC for AA and RA

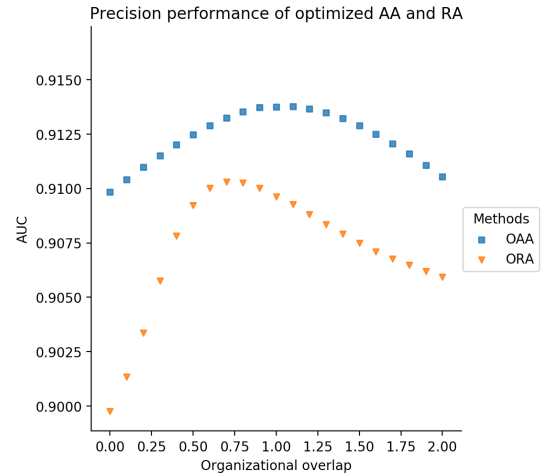


Fig. 5: Effects of organizational overlap in Precision for AA and RA

organizational hierarchy. We formulated the problem as a link prediction problem. However, unlike any generic link prediction work, we first validated explicit enterprise social network with a set of heterogeneous implicit collaboration interaction networks. Extensive experiments on both networks demonstrated that evidence from multiple sources of interactions other than social network can enhance the quality of the social network which further leads to more accurate results in deriving insights from the network such as predicting new links.

Furthermore, our framework leverages the assessed ESN in weak ties prediction for an employee. We characterize weak ties from information benefits perspective and define

them as link between employees who belong to different teams and have minimal exiting overlap from the social and organizational structure's point of view. To this end, we introduce a "social-organizational overlap" factor which we used to optimize common-neighborhood based link prediction methods for weak ties prediction. This overlap factor penalizes the common neighbors which are in the same team and encourages the common neighbors in different communities than the endpoints when predicting a link between two nodes. Our experiments confirm that link prediction accuracy increases when organizational hierarchy information was included in link prediction. Experiment results demonstrated high AUC and precision for a predicted link with endpoints in the different teams (i.e. when the "social-organization overlap is low") and vice versa. Hence, we highlight and predict the weak ties for an employee as links that span across different teams high. Our institution for weak ties aligns with multiple works such as community based re-definition of Granovetter's weak tie theory [8] by [21] and notion of *structural equivalence* for identifying "structural holes" as proposed by [9].

In future, we aim to leverage the EWTR framework proposed in this paper for the assessment of "social capital" of employees for their career development at work. We will leverage weak ties as characterized in this paper and the notion of "structural hole" to highlight the horizontal (at same level of organization hierarchy as the employee) as well vertical (levels above the current level of a particular employee in the hierarchy) social capital of career development. Furthermore, we aim to enhance our link assessment strand of work by first improving our existing work by utilizing additional features that are not common neighborhood based as well as by implementing strategy for handling imbalanced labels and incorporating transfer learning techniques for implicit interaction networks to learn additional information from social science theories perspective in inferring explicit social links.

ACKNOWLEDGMENT

The authors would like to thank IBM for access to anonymized dataset. This work has received funding from the European Unions Horizon 2020 research and innovation programme through the DEVELOP project, under grant agreement No 688127.

REFERENCES

- [1] K. Ehrlich, C.-Y. Lin, and V. Griffiths-Fisher, "Searching for experts in the enterprise: combining text and social network analysis," in *Proceedings of the 2007 international ACM conference on Supporting group work*. ACM, 2007, pp. 117–126.
- [2] A. Richter and K. Riemer, "The contextual nature of enterprise social networking: A multi case study comparison," in *ECIS*, 2013, p. 94.
- [3] J. Huang, J. Zhang, Y. Li, and Z. Lv, "Business value of enterprise micro-blogging: Empirical study from weibo.com in sina," *J. Glob. Inf. Manage.*, vol. 22, no. 3, pp. 32–56, Jul. 2014. [Online]. Available: <http://dx.doi.org/10.4018/jgim.2014070102>
- [4] J. Spiess, Y. T. Joens, R. Dragnea, and P. Spencer, "Using Big Data to Improve Customer Experience and Business Performance," *Bell Labs Technical Journal*, vol. 18, no. 4, pp. 3–17, 2014.
- [5] T. Lappas, K. Liu, and E. Terzi, "Finding a team of experts in social networks," in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '09. New York, NY, USA: ACM, 2009, pp. 467–476. [Online]. Available: <http://doi.acm.org/10.1145/1557019.1557074>
- [6] S. E. Seibert, M. L. Kraimer, and R. C. Liden, "A social capital theory of career success," *Academy of Management Journal*, vol. 44, no. 2, pp. 219–237, 2001.
- [7] P. S. Adler and S.-w. Kwon, "Prospects for a New Concept," *The Academy of Management Review*, vol. 27, no. 1, pp. 17–40, 2002.
- [8] M. S. Granovetter, "The strength of weak ties," in *Social networks*. Elsevier, 1977, pp. 347–367.
- [9] R. S. Burt, "The social structure of competition," in *Explorations in Economic Sociology*, Russell Sage Foundation. Citeseer, 1993.
- [10] A. Wu, J. M. DiMicco, and D. R. Millen, "Detecting professional versus personal closeness using an enterprise social network site," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '10. New York, NY, USA: ACM, 2010, pp. 1955–1964. [Online]. Available: <http://doi.acm.org/10.1145/1753326.1753622>
- [11] J. Zhang and P. S. Yu, "Enterprise Social Link Recommendation Categories and Subject Descriptors," *CIKM 2015: Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, pp. 841–850, 2015.
- [12] M. Al Hasan, V. Chaoji, S. Salem, and M. Zaki, "Link prediction using supervised learning," in *SDM06: workshop on link analysis, counter-terrorism and security*, 2006.
- [13] D. Liben-Nowell and J. Kleinberg, "The link-prediction problem for social networks," *Journal of the Association for Information Science and Technology*, vol. 58, no. 7, pp. 1019–1031, 2007.
- [14] L. Lü and T. Zhou, "Role of weak ties in link prediction of complex networks," in *Proceedings of the 1st ACM international workshop on Complex networks meet information & knowledge management*. ACM, 2009, pp. 55–58.
- [15] J. Zhao, J. Wu, and K. Xu, "Weak ties: Subtle role of information diffusion in online social networks," *Physical Review E*, vol. 82, no. 1, p. 016105, 2010.
- [16] T. Murata and S. Moriyasu, "Link prediction of social networks based on weighted proximity measures," in *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence*, ser. WI '07. Washington, DC, USA: IEEE Computer Society, 2007, pp. 85–88. [Online]. Available: <http://dx.doi.org/10.1109/WI.2007.71>
- [17] T. Zhou, L. Lü, and Y.-C. Zhang, "Predicting missing links via local information," *The European Physical Journal B*, vol. 71, no. 4, pp. 623–630, 2009.
- [18] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.
- [19] M. Abufouda and K. A. Zweig, "Link classification and tie strength ranking in online social networks with exogenous interaction networks," *arXiv preprint arXiv:1708.04030*, 2017.
- [20] M. Smith, C. Giraud-Carrier, and N. Purser, "Implicit affinity networks and social capital," *Information Technology and Management*, vol. 10, no. 2-3, pp. 123–134, 2009.
- [21] E. Ferrara, P. De Meo, G. Fiumara, and A. Provetti, "The role of strong and weak ties in facebook: a community structure perspective," *Preprint at* <http://arXiv.org/abs/1203.0535>, 2012.
- [22] Z. Yang, J. Song, Z. Huang, X. Zhu, and H. Tian, "A community-structure based adaptively optimized link prediction algorithm," *Proceedings - 4th IEEE International Conference on Big Data and Cloud Computing, BDCloud 2014 with the 7th IEEE International Conference on Social Computing and Networking, SocialCom 2014 and the 4th International Conference on Sustainable Computing and C*, pp. 463–469, 2015.
- [23] H. Liu, Z. Hu, H. Haddadi, and H. Tian, "Hidden link prediction based on node centrality and weak ties," *EPL (Europhysics Letters)*, vol. 101, no. 1, p. 18004, 2013.