MissBiN: Visual Analysis of Missing Links in Bipartite Networks

Jian Zhao*   Maoyuan Sun†   Francine Chen*   Patrick Chiu*

*FX Palo Alto Laboratory   †Northern Illinois University

ABSTRACT

The analysis of bipartite networks is critical in a variety of application domains, such as exploring entity co-occurrences in intelligence analysis and investigating gene expression in bio-informatics. One important task is missing link prediction, which infers the existence of unseen links based on currently observed ones. In this paper, we propose MissBiN that involves analysts in the loop for making sense of link prediction results. MissBiN combines a novel method for link prediction and an interactive visualization for examining and understanding the algorithm outputs. Further, we conducted quantitative experiments to assess the performance of the proposed link prediction algorithm and a case study to evaluate the overall effectiveness of MissBiN.

Keywords: Link prediction, bipartite network, visual analytics.

1 INTRODUCTION

Many real-world systems can be modeled as *bipartite networks* (i.e., two-mode networks). There are two types of nodes in a network and links only exist between different node types. Bipartite relationship analysis has been applied in a variety of application domains, such as studying political leanings with voter-vote networks [5] and investigating gene-expression networks in bioinformatics [23].

One important network analysis task is *link prediction* (i.e., detecting missing links), which infers the existence of implicit relationships between nodes based on currently observed links [21]. Link prediction is extremely useful in practice because real-world data is often noisy or incomplete; for example, guiding biological experiments in checking protein interaction networks [4].

In practice, analysts need to leverage their domain knowledge to examine algorithmic results. This is because the algorithm output is usually a list of scores or probabilities, which may be difficult to interpret. Moreover, these results can be inaccurate. They may ask: why is this link identified as missing with a high score, does it make sense to have a link between these two nodes, and how will the network change by adding one or several missing links? However, there still is a lack of effective tools to allow analysts to browse the results and explore answers to these questions.

To address these issues, we propose MissBiN, a visual analysis tool for detecting and examining missing links in bipartite networks. First, we contribute a novel link prediction approach that leverages the information of bi-cliques, inspired by structural hole theory in social science [3]. The method can be applied with any existing algorithm (e.g., neighbor-based techniques [24]), to both weighted and unweighted networks. Second, we develop an interactive visualization to present detected missing links, allowing for a better understanding of the missing links and their impact. The visualization enables analysts to compare networks with and without specific links interactively added by analysts. To evaluate MissBiN, we conducted quantitative experiments on real-world datasets and an in-depth case study of analyzing entity relationships in documents.

---
*{zhao,chen,chiu}@fxpal.com

†smaoyuan@niu.edu

2 BACKGROUND

2.1 Bipartite Network Analysis

One of the key computational approaches for analyzing a general network (i.e., containing only one type of node) is to calculate the node centrality indices (e.g., betweenness and closeness), which characterizes the importance of a node [2]. These metrics are also applicable for bipartite networks [1]. In addition, any methods for general network analysis can be employed on a projected bipartite network [1], but some information may be lost. Another branch of techniques is to identify special groups of nodes, such as motifs (e.g., chain, star, and clique), clusters, and communities. Due to particular properties of bipartite networks, the motif-based analysis mainly focuses on extracting bi-cliques [40]. Also, biclustering techniques (e.g., [18]) can be applied to simultaneously group two types of nodes, relaxing the criteria of bi-cliques.

Without losing generality, MissBiN supports visual analysis of bipartite networks based on the above two common approaches: metric-based and motif-based. Specifically, MissBiN allows analysts to interactively investigating the influence of particular missing links.

2.2 Missing Link Prediction Algorithms

Common link prediction algorithms roughly fall into two major categories: learning-based and similarity-based [24, 35]. The learning-based methods usually treat link prediction as a binary classification problem and train a machine learning model to predict the class label (i.e., positive for potential linking) for each non-connected node pair. These methods often leverage features extrated from node attributes and structures [22, 27, 36], or probabilistic graph models [11, 39]. These techniques, although effective, are less general, often requiring additional information (e.g., semantic node attributes) of networks. The similarity-based methods attempt to compute a similarity score for every non-connected pair of nodes and rank all these potential links. The similarity metrics can be computed with random-walk based simulation and neighbor-based measures such as the Adamic-Adar coefficient [24, 35]. Researchers have also extended some of the similarity metrics to the bipartite network scenario [6, 37].

We move one step further by integrating important structural information in bipartite networks [3], on top of the link prediction scores generated by existing common approaches. We adopt the similarity-based approach, because the learning-based approach requires dataset-specific information (e.g., node attributes) for training. But our approach can be used with any link prediction algorithms that produce a list of missing link probabilities.

2.3 Bipartite Network Visualization

Similar to visualizing general networks, two main approaches of presenting bipartite networks are: node-link diagrams and matrices. Node-link diagrams are easier to understand and emphasize entities (i.e., nodes), but suffer from increased visual clutters for larger and denser networks [17]. Matrices emphasize relationships (i.e., links) and are more scalable in many network analysis tasks [9].

Jigsaw's List View [29] is a typical example based on node-link diagrams, organizing different types of nodes in separate lists. Variations of this design include Focus+Context lists [28], edge bundling lists [32], employing nested layouts with duplicated nodes but no links [25]. BiSet [31] shows bi-cliques between two lists

of nodes and uses link bundling. Radial layout of nodes has also been applied [7]. Second, bipartite networks can be shown as a bi-adjacency matrix, where rows and columns represent two different types of nodes. Example systems include BiVoc [10], Bicluster viewer [12], Expression Profiler [16], and BicOverlapper 2.0 [26]. An exception is BiDots [41] which organizes bi-cliques in rows to emphasize patterns and places nodes in columns. Hybrid techniques have been proposed to combine the advantages of node-link diagrams and matrices. For example, based on NodeTrix [13], Furby [30] and Bixplorer [8] display each bi-clique as an individual matrix and connect them with links. Xu et al. [38] applied a similar method for the projected bipartite network. Further, Matchmaker [20] and VisBricks [19] augment this visual representation with extensive charts (e.g., heatmaps and parallel coordinates) to display bi-cliques.

While some visual design of MissBiN is inspired by the above systems, none of them has addressed the problems of detecting and visualizing missing links. Particularly, we employ a matrix-based design because links are the focus in our analysis scenarios.

3 MISSING LINK PREDICTION OF MISSBIN

3.1 Problem Definition and Node Similarity

A bipartite network can be formally defined as $G = \langle X, Y, E \rangle$, where X and Y are two non-overlapping sets of nodes and E is the set of links that only exist between X and Y, i.e., $e = \langle x, y \rangle \in E$ where $x \in X$ and $y \in Y$. For a bipartite network, the number of all possible links is $|X| \cdot |Y|$ and we denote this set of links as U.

A link prediction problem is to identify which links are likely missing in the set $U - E$. As discussed in Section 2.2, similarity-based methods first compute the similarity score of every non-connected pair of nodes, and then generate a *ranked list* of missing links with decreasing scores for prediction. These types of methods are more generic because they purely rely on the network topology and are not biased to particular datasets. Thus, we adopt this for development and experiments, but our approach can be used with any link prediction methods that generate link probabilities.

One way to compute the node similarity is via a random walk, and here we employ *random walk with restart (RWR)* [33]. To accommodate bipartite networks, we let the random walker run for an odd number of iterations, such that it always stops at nodes in the other set. Since this measure is not symmetric, we use the average of random walk scores from node x to node y and the opposite.

Another way of measuring similarity is based on comparing the neighborhoods of two nodes. The intuition is that the more similar the topology of two neighborhoods are, the more likely the link connecting the two nodes is missing. For a bipartite network, neighbors of two possible connected nodes must be different. Following the ideas in [6, 37], we define the one-hop neighbors of a node x in a bipartite network as $\Gamma(x)$, and we further define $\gamma(x)$ as the set of the two-hop links of a node x. That is, $\gamma(x) = \{\langle x_i, y \rangle \in E : x_i \neq x, y \in \Gamma(x)\}$. For example, in Figure 1a, $\Gamma(x_4) = \{y_3, y_4\}$ and $\gamma(x_4) = \{\langle x_1, y_3 \rangle, \langle x_5, y_3 \rangle, \langle x_1, y_4 \rangle, \langle x_2, y_4 \rangle, \langle x_3, y_4 \rangle\}$. A number of similarity metrics can be applied to compare the neighbor context of two nodes, $\gamma(x)$ and $\gamma(y)$, such as:

common neighbors: $s_{xy} = |\gamma(x) \cap \gamma(y)|$;

Jaccard coefficient: $s_{xy} = \frac{|\gamma(x) \cap \gamma(y)|}{|\gamma(x) \cup \gamma(y)|}$;

Adamic-Adar coefficient: $s_{xy} = \sum_{<m,n> \in \gamma(x) \cap \gamma(y)} \frac{1}{\log|\Gamma(m)| \cdot |\Gamma(n)|}$;

and *preferential attachment*: $s_{xy} = |\gamma(x)| \cdot |\gamma(y)|$.

3.2 Enhancement based on Bi-Clique Information

Based on the above methods, we propose to integrate a key structure in bipartite networks, *maximal bi-cliques*, for link prediction. Formally, a maximal bi-clique is defined as a sub-network, $G' = \langle X', Y', E' \rangle$, where $X' \subseteq X$, $Y' \subseteq Y$, and $E' \subseteq E$, and there exists a link $e = \langle x, y \rangle \in E'$ between every pair of nodes, $x \in X'$ and $y \in Y'$.

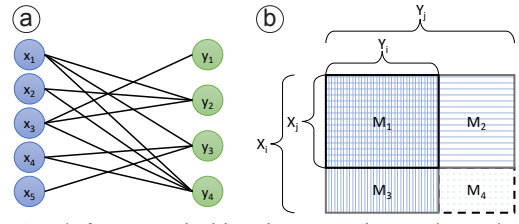Figure 1: a) An example bipartite network $G = \langle X, Y, E \rangle$. b) An illustration of two overlapped bi-cliques, $C_i = \langle X_i, Y_i, E_i \rangle$ and $C_j = \langle X_j, Y_j, E_j \rangle$; each of them is shown as a bi-adjacency matrix.

Algorithm 1: Missing link weighting based on bi-cliques.

Data: A list of bi-cliques, $L = \{C_i = \langle X_i, Y_i, E_i \rangle\}$, detected in a bipartite network $G = \langle X, Y, E \rangle$.
Result: Scores, w, for all non-observed (missing) links in G.
foreach $e \in U - E, U = \{\forall \langle x, y \rangle; x \in X, y \in Y\}$ **do**
    $w_e \leftarrow 0$;
foreach bi-clique pair (C_i, C_j) from L **do**
    $o \leftarrow \frac{|X_i \cap X_j| \cdot |Y_i \cap Y_j|}{|X_i \cup X_j| \cdot |Y_i \cup Y_j|}$;
    if $o < \theta$ **then**
       continue;
    foreach $e \in \{\forall \langle x, y \rangle; x \in (X_i - X_j) \cup (X_j - X_i), y \in (Y_j - Y_i) \cup (Y_i - Y_j)\}$ **do**
       $w_e \leftarrow w_e + \frac{|X_i \cap X_j| \cdot |Y_i \cap Y_j|}{|X_i - X_j| \cdot |Y_j - Y_i| + |X_j - X_i| \cdot |Y_i - Y_j|}$;

Our intuition is that missing links connecting nodes from different bi-cliques to form a larger bi-clique should carry more weight. This is inspired by the *structural hole* theory in social science [3]. That is, a person tends to know what other people in the same community know, and the most beneficial connections come from the *weak ties* that link people from different communities. Also, the literature indicates that it is useful in bipartite network analysis [37]. However, their method cannot fit generally with any node similarity metric.

As is shown in Figure 1b, consider two bi-cliques as two communities that have some nodes in common; each missing link between the non-overlapping nodes (i.e., M_4) from the two communities contributes to the formation of a bigger community that benefits all the nodes. If the two communities have many nodes in common, each of the few missing links that can be added carries more value, as a bigger bi-clique can be formed fairly easily. On the other hand, if the two communities have less in common, then more links need to be added to merge the two bi-cliques into a larger one, and each of the missing links carries less value.

Following this intuition, we develop an algorithm to re-rank the missing link list with scores $s(x, y)$ generated by the above similarity-based methods. As shown in Algorithm 1, it computes scores, w_e, for all missing links (e.g., M_4 in Figure 1b) based on detected bi-cliques. When processing each pair of bi-cliques, w_e is updated with a value determined by the size of two bi-cliques and their overlap. Intuitively, in Figure 1b, the value computed in each iteration corresponds to the area of the intersection M_1 divided by the area of the missing part M_4. We only consider bi-clique pairs with overlap larger than a ratio θ to avoid overly biasing towards non-related bi-cliques. Then, we normalize the scores by the maximum value and generate a new ranked list of predicted missing links with the new scores: $s'(x, y) = w(x, y) \cdot s(x, y)$.

4 VISUAL INTERFACE OF MISSBIN

Algorithms are not always perfect in real-world usage, so some detected missing links may not be meaningful. We design a visual interface of MissBiN (Figure 2) allowing analysts to better make sense of predicted missing links.

4.1 Visual Exploration of Missing Links

MissBiN supports the visual exploration of predicted missing links through two views. First, the Network View (Figure 2a) displays
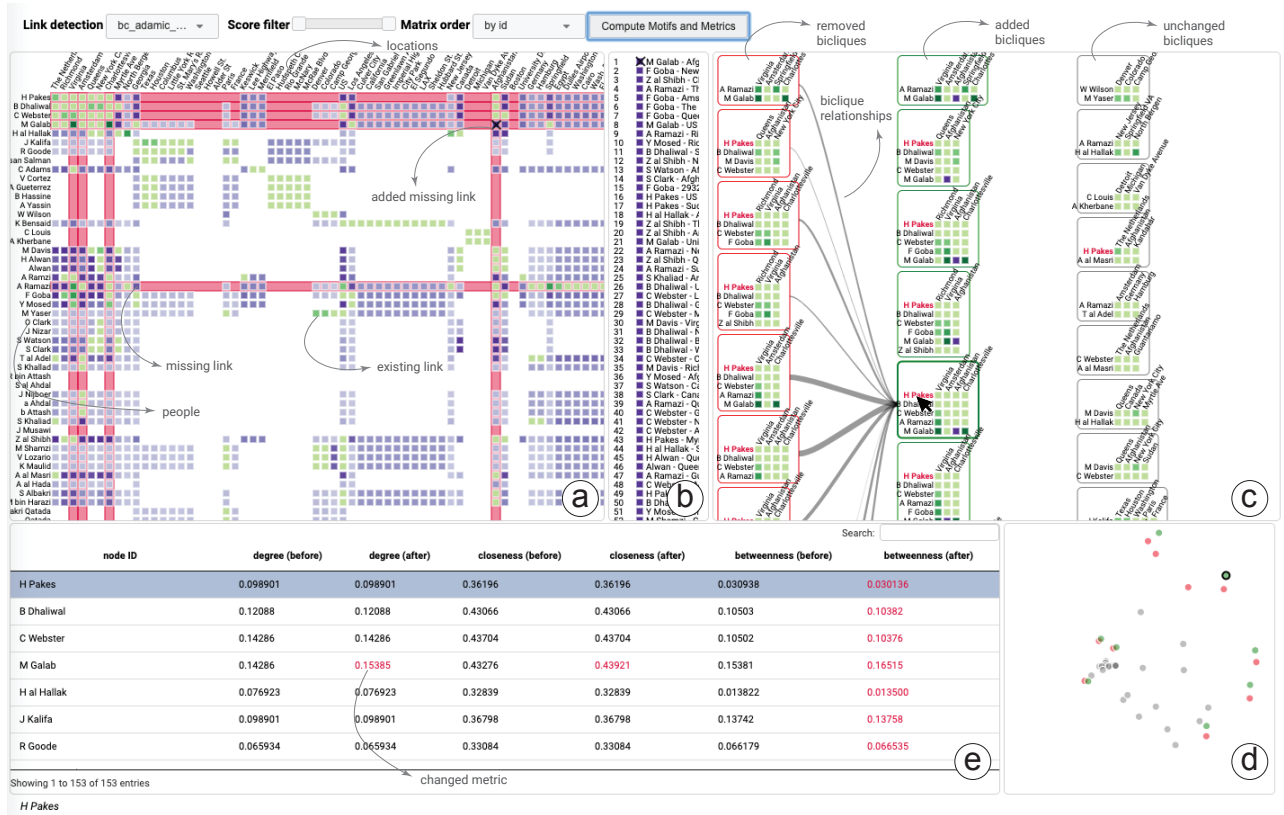
Figure 2: The visual interface of MissBiN: a) Network View, b) Link List View, c) Motifs Detail View, d) Motifs Overview, and e) Metrics View.

the bi-adjacency matrix of a bipartite network, where the row and the column represent two different types of nodes respectively. The links are represented as squares in the intersections of rows and columns. The existing links in the network are shown in a yellow-green colorscale, reflecting the link weight. The predicted links are displayed in a white-purple colorscale with darker color indicating higher score. Second, the Link List View (Figure 2b) shows the computed missing links linearly by prediction score. A number of user interactions are offered, such as reordering rows and columns of the matrix as well as filtering based on the score.

These two views offer an overview of the bipartite network and the performance of the missing link prediction. We choose the matrix-based design for the Network View because our focus is links; matrices utilize much visual space for encoding links and are more scalable for denser networks [9, 17]. Moreover, the color channel cannot precisely reflect the scores and ranks of the missing links, we design the Link List View to augment the Network View for exploring the link prediction. An analyst can further explore the results and hypothetically add certain missing links to examine their influence with visual analysis of motifs and metrics (as described later). The added links are marked as black crosses on the matrix and also displayed at the top of the list (Figure 2a & b).

4.2 Visual Analysis of Network Motifs and Node Metrics

Motif analysis is one key approach to understanding the topology of a network. MissBiN provides a Detail View (Figure 2c) and an Overview (Figure 2d) for browsing the motifs at different scales. In the Motifs Detail View, bi-cliques are displayed as small multiples of matrices in similar visual encodings to the Network View (Figure 2a). We employ a similar matrix-based design as the Network View because a bi-clique is essentially a portion of the bi-adjacency matrix of the entire network, keeping visual consistency. Usually many bi-cliques can be identified in a network, and a lot of them have overlapping nodes and links. To reveal their relationships, we deisgn

the Motifs Overview that displays all the bi-cliques as dots in a 2D space based on MDS, and thus distance reflects bi-clique similarity.

These two views not only support the visual exploration of bi-cliques, but also the investigation of the impact when certain missing links are added, answering "what if" questions. To support comparing two sets of bi-cliques from the networks with and without added links by an analyst, the Motifs Detail View organizes bi-cliques in three columns: *removed* (due to merging) bi-cliques in red, *newly-added* ones in green, and *unchanged* ones in gray, compared to the bi-clique set of the original network. Similarly, the Motifs Overview encodes these bi-cliques in the same colors, also revealing bi-clique clusters and changing trends. Although we currently only allow for one type of motif (i.e., bi-clique), other forms of motif analysis can be supported with a similar design.

Computing node-metrics is another important way of learning a big picture of network characteristics quantitatively. The Metrics View (Figure 2e) supports this type of analysis by presenting a number of metrics in an interactive tabular view: degree, closeness, and betweenness centralities of before and after adding certain missing links, etc. Changes of metric values are highlighted in red, revealing the effect of added links.

5 EVALUATION OF MISSING LINK PREDICTION

5.1 Experimental Datasets

We used two real-world datasets to conduct experiments for testing our algorithm. The first is a person-place network extracted from the Atlantic Storm corpus containing 111 intelligence reports [15]. We identified 207 person nodes, 165 place nodes, and 1,718 links between them, from these reports. The second dataset is a weighted user-conversation bipartite network detected from Slack communication messages of a group within an IT company. The weight of a link is based on the number of words that a user contributed to a conversation. The resulting bipartite network contains 41 users, 61 conversations, and 258 links.

		Atlantic Storm			Slack communication		
R-Precision	JA	.395	.428	.032	.332	.341	.009
	CN	.359	.580	.221	.202	.339	.138
	AA	.440	.455	.016	.203	.218	.015
	PA	.021	.585	.564	.096	.432	.336
	RW	.467	.531	.064	.398	.407	.009
AUC PR	JA	.398	.451	.053	.225	.249	.024
	CN	.305	.607	.302	.133	.285	.152
	AA	.398	.429	.031	.153	.177	.025
	PA	.008	.566	.557	.055	.406	.352
	RW	.435	.516	.082	.279	.308	.029

Table 1: For each condition (i.e., in a table cell), the three numbers indicate: the average metric of the baseline, the average metric of the proposed method, and the improvement over the five runs.
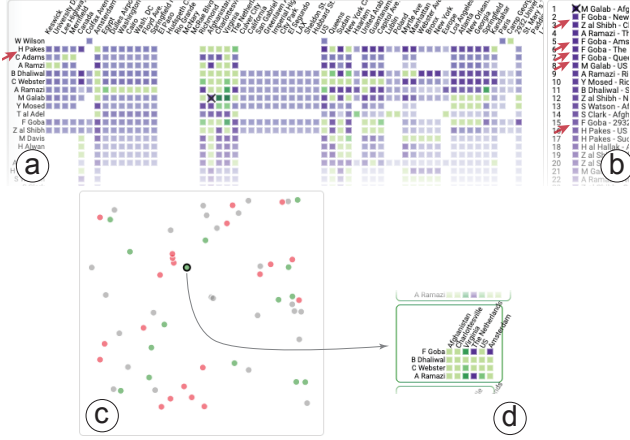

Figure 3: Analyzing *The Sign of the Crescent* using MissBiN.

5.2 Experimental Setup and Results

Because there is no ground truth for missing links, we followed a commonly used method [37]: randomly remove a certain number of links from an original network, apply the link prediction algorithms on this new network, and measure the performance by comparing the predicted links with the removed links (i.e., the ground truth). We evaluated our bi-clique oriented link prediction approach by comparing it with five corresponding baseline algorithms, including common neighbors (CN), Jaccard coefficient (JA), Adamic-Adar coefficient (AA), preferential attachment (PA), and random walk (RW) [24, 35]. For each algorithm, we randomly removed 1%, 2%, 5%, 10%, and 15% of links. For each condition, we performed the experiment five times to reduce sampling bias.

We used two metrics to measure the performance of the algorithms: R-Precision (the ratio between all the relevant items retrieved until the rank that equals the number of relevant items in the collection), and AUC PR (the area under the precision-recall curve). Table 1 shows the average performance of each condition. Based on the results, we can observe that the bi-clique oriented methods enhance the baselines in all the conditions with different levels of improvement on both metrics. Some of the performance gain is substantial, where the maximum improvement appears with the PA condition of the Atlantic Storm dataset.

6 CASE STUDY

To demonstrate the usefulness of MissBiN, in this section, we walk through a case study of investigative analysis with *The Sign of the Crescent* [15] dataset. It contains 41 fictional reports regarding three coordinated terrorist plots in the US, of which 24 are relevant to the plot. Using name-entity detection techniques, we analyze a person-location bipartite network containing 49 persons and 104 locations as well as 328 connections between them. Examining missing links in such an analysis can be greatly helpful, because the information is

often incomplete. We developed a simple Document View on top of the MissBiN interface, which displays all the related reports based on user-selected entities or relationships. Our analyst's task is to identify suspicious persons and activities from these reports. She launches MissBiN, which visualizes the person-location bipartite network (Figure 2). First, she adjusts the threshold to only show predicted missing links with probability higher than 0.7, because there are too many purple squares in the Network View, which is a bit overwhelming. Among all the missing links, *M. Galab - Afghanistan* has the highest probability (Figure 2b). She thus adds this potential link by clicking it and re-computes the motifs and node metrics. The Motifs Detail View then displays the removed, added, and unchanged bi-cliques of the network. Of the eight newly formed bi-cliques, a person named *H. Pakes* appears the most, which is in six bi-cliques (Figure 2c). Navigating back to the Network View, she finds that *H. Pakes* connects with many locations with high probability. She further sorts the nodes in decreasing order based on average missing link probability, and confirms this observation since *H. Pakes* is ranked the second (Figure 3a). The first person node has too few connections with the locations, which seems an isolated node. Thus, she focuses on *H. Pakes* and reads a few reports regarding him. She then finds that *H. Pakes* is a person carrying a forged Dutch passport and his actual given name is *Abu al Masri*, a member of the terrorist organization *Al Qaeda*. The reports reveal that *H. Pakes* was involved in shipping explosive materials from *Holland* to the US.

From the Link List View, she discovers that a particular person named *F. Goba* appears frequently in missing links with high probability (Figure 3b). For example, its connections to *New York City*, *Amsterdam*, and *Queens* are ranked second, fifth, and seventh, respectively. She reads the reports related to these locations and connects them with the information obtained before. She finds that *F. Goba* was involved in attacking a train to *New York City* with a bomb made with the explosive materials shipped by *H. Pakes*. Next, she adds the missing links between *F. Goba* and *New York City*, *Amsterdam*, and *Queens*, which results in a lot of newly-formed and removed bi-cliques. Then, she uses the Motifs Overview to explore them (Figure 3c). One interesting pattern catches her eyes: a cluster of four red dots close to a green dot, indicating that these (red) bi-cliques are quite similar and may be merged together into one new (green) bi-clique after adding a few missing links. By exploring the new bi-clique in the Motifs Detail View (Figure 3d), she finds two names: *B. Dhaliwal* and *C. Webster*. She further investigates these two people from the reports and learns that *B. Dhaliwal* is also a fake name, who carries an Indian passport, and his actual name is *S. Aloakri*, a Pakistani who served *Taliban* in *Afghanistan*. More investigation of the reports reveals that *B. Dhaliwal* was likely related to the previously identified train attack.

7 CONCLUSION AND FUTURE WORK

We have presented MissBiN, a visual analysis tool for exploring and understanding missing links in bipartite networks. It offers a novel approach for missing link prediction by using the information of bi-cliques in networks. Moreover, MissBiN provides an interactive visualization to present computed missing links and support the investigation of their meaning and influence by comparing networks with and without selected missing links. Quantitative experimental results and a case study indicate promising applications of MissBiN in practice. However, more evaluation such as deployment studies needs to be conducted to investigate its usage in real-world settings and in other domains. Also, we plan to extend the scalability of both the algorithm and visualization of MissBiN. We want to employ faster, but approximated, bi-clique detection algorithms [40], and enhance the matrix view for large networks with multi-scale [34] and focus+context techniques [14].

REFERENCES

[1] A. S. Asratian, T. M. J. Denley, and R. Hggkvist. *Bipartite Graphs and their Applications*. Cambridge Tracts in Mathematics. Cambridge University Press, 1998.

[2] P. Bonacich. Power and centrality: A family of measures. *American Journal of Sociology*, 92(5):1170–1182, 1987.

[3] R. S. Burt. *Structural Holes: the Social Structure of Competition*. Harvard University Press, 1992.

[4] C. V. Cannistraci, G. Alanis-Lobato, and T. Ravasi. From link-prediction in brain connectomes and protein interactomes to the local-community-paradigm in complex networks. *Scientific Reports*, 3(1), 2013.

[5] C. Carrubba, M. Gabel, and S. Hug. Legislative voting behavior, seen and unseen: A theory of roll-call vote selection. *Legislative Studies Quarterly*, 33(4):543–572, 2008.

[6] Y.-J. Chang and H.-Y. Kao. Link prediction in a bipartite network using wikipedia revision information. In *Proceedings of the Conference on Technologies and Applications of Artificial Intelligence*. IEEE, 2012.

[7] M. Dumas, J. Robert, and M. J. Mcguffin. Alertwheel: radial bipartite graph visualization applied to intrusion detection system alerts. *IEEE Network*, 26(6):12–18, 2012.

[8] P. Fiaux, M. Sun, L. Bradel, C. North, N. Ramakrishnan, and A. Endert. Bixplorer: Visual analytics with biclusters. *Computer*, (8):90–94, 2013.

[9] M. Ghoniem, J.-D. Fekete, and P. Castagliola. On the readability of graphs using node-link and matrix-based representations: A controlled experiment and statistical analysis. *Information Visualization*, 4(2):114–135, 2005.

[10] G. A. Grothaus, A. Mufti, and T. Murali. Automatic layout and visualization of biclusters. *Algorithms for Molecular Biology*, 1(1):1–15, 2006.

[11] D. Heckerman, C. Meek, and D. Koller. *Probabilistic Entity-Relationship Models, PRMs and Plate Models*, pages 201–239. MIT Press, 2004.

[12] J. Heinrich, R. Seifert, M. Burch, and D. Weiskopf. Bicluster viewer: A visualization tool for analyzing gene expression data. In *Proceedings of the International Symposium Advances in Visual Computing*, pages 641–652. Springer Berlin Heidelberg, 2011.

[13] N. Henry, J.-D. Fekete, and M. McGuffin. NodeTrix: a hybrid visualization of social networks. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1302–1309, 2007.

[14] Y. Hu and L. Shi. Visualizing large graphs. *Wiley Interdisciplinary Reviews: Computational Statistics*, 7(2):115–136, 2015.

[15] F. Hughes and D. Schum. Discovery, proof, choice: the art and science of the process of intelligence analysis-preparing for the future of intelligence analysis. *Washington, DC: Joint Military Intelligence College*, 2003.

[16] M. Kapushesky, P. Kemmeren, A. C. Culhane, S. Durinck, J. Ihmels, C. Korner, M. Kull, A. Torrente, U. Sarkans, J. Vilo, and A. Brazma. Expression profiler: next generation–an online platform for analysis of microarray data. *Nucleic Acids Research*, 32:W465–W470, 2004.

[17] R. Keller, C. M. Eckert, and P. J. Clarkson. Matrices or node-link diagrams: Which visual representation is better for visualising connectivity models? *Information Visualization*, 5(1):62–76, 2006.

[18] Y. Kluger. Spectral biclustering of microarray data: Coclustering genes and conditions. *Genome Research*, 13(4):703–716, 2003.

[19] A. Lex, H.-J. Schulz, M. Streit, C. Partl, and D. Schmalstieg. VisBricks: Multiform visualization of large, inhomogeneous data. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2291–2300, 2011.

[20] A. Lex, M. Streit, C. Partl, K. Kashofer, and D. Schmalstieg. Comparative analysis of multidimensional, quantitative data. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1027–1035, 2010.

[21] D. Liben-Nowell and J. Kleinberg. The link prediction problem for social networks. In *Proceedings of the Twelfth International Conference on Information and Knowledge Management*, CIKM '03, pages 556–559. ACM, 2003.

[22] R. N. Lichtenwalter and N. V. Chawla. Vertex collocation profiles: Subgraph counting for link analysis and prediction. In *Proceedings of the 21st International Conference on World Wide Web*, WWW '12, pages 1019–1028, New York, NY, USA, 2012. ACM.

[23] S. C. Madeira and A. L. Oliveira. Biclustering algorithms for biological data analysis: A survey. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 1(1):24–45, 2004.

[24] V. Martínez, F. Berzal, and J.-C. Cubero. A survey of link prediction in complex networks. *ACM Computing Surveys*, 49(4):1–33, 2016.

[25] C. Partl, A. Lex, M. Streit, H. Strobelt, A.-M. Wassermann, H. Pfister, and D. Schmalstieg. Contour: Data-driven exploration of multi-relational datasets for drug discovery. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1883–1892, 2014.

[26] R. Santamaria, R. Theron, and L. Quintales. BicOverlapper 2.0: visual analysis for gene expression. *Bioinformatics*, 30(12):1785–1786, 2014.

[27] S. Scellato, A. Noulas, and C. Mascolo. Exploiting place features in link prediction on location-based social networks. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '11, pages 1046–1054, New York, NY, USA, 2011. ACM.

[28] H.-J. Schulz, M. John, A. Unger, and H. Schumann. Visual Analysis of Bipartite Biological Networks. In C. Botha, G. Kindlmann, W. Niessen, and B. Preim, editors, *Eurographics Workshop on Visual Computing for Biomedicine*. The Eurographics Association, 2008.

[29] J. Stasko, C. Görg, and Z. Liu. Jigsaw: Supporting investigative analysis through interactive visualization. *Information visualization*, 7(2):118–132, 2008.

[30] M. Streit, S. Gratzl, M. Gillhofer, A. Mayr, A. Mitterecker, and S. Hochreiter. Furby: fuzzy force-directed bicluster visualization. *BMC Bioinformatics*, 15(Suppl 6):S4, 2014.

[31] M. Sun, P. Mi, C. North, and N. Ramakrishnan. Biset: Semantic edge bundling with biclusters for sensemaking. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):310–319, 2016.

[32] M. Sun, J. Zhao, H. Wu, K. Luther, C. North, and N. Ramakrishnan. The effect of edge bundling and seriation on sensemaking of biclusters in bipartite graphs. *IEEE Transactions on Visualization and Computer Graphics*, 2019 (In Press).

[33] H. Tong, C. Faloutsos, and J. yu Pan. Fast random walk with restart and its applications. In *Proceedings of International Conference on Data Mining*. IEEE, 2006.

[34] F. van Ham. Using multilevel call matrices in large software projects. In *IEEE Symposium on Information Visualization*. IEEE, 2003.

[35] P. Wang, B. Xu, Y. Wu, and X. Zhou. Link prediction in social networks: the state-of-the-art. *Science China Information Sciences*, 58(1):1–38, 2015.

[36] T. Wohlfarth and R. Ichise. Semantic and event-based approach for link prediction. In T. Yamaguchi, editor, *Practical Aspects of Knowledge Management*, pages 50–61, Berlin, Heidelberg, 2008. Springer Berlin Heidelberg.

[37] S. Xia, B. Dai, E.-P. Lim, Y. Zhang, and C. Xing. Link prediction for bipartite social networks: The role of structural holes. In *Proceedings of International Conference on Advances in Social Networks Analysis and Mining*. IEEE, 2012.

[38] P. Xu, N. Cao, H. Qu, and J. Stasko. Interactive visual co-cluster analysis of bipartite graphs. In *Proceedings of IEEE Pacific Visualization Symposium*. IEEE, 2016.

[39] K. Yu and W. Chu. Gaussian process models for link analysis and transfer learning. In J. C. Platt, D. Koller, Y. Singer, and S. T. Roweis, editors, *Advances in Neural Information Processing Systems*, pages 1657–1664. Curran Associates, Inc., 2008.

[40] Y. Zhang, C. A. Phillips, G. L. Rogers, E. J. Baker, E. J. Chesler, and M. A. Langston. On finding bicliques in bipartite graphs: a novel algorithm and its application to the integration of diverse biological data types. *BMC Bioinformatics*, 15(1):110, 2014.

[41] J. Zhao, M. Sun, F. Chen, and P. Chiu. BiDots: Visual exploration of weighted biclusters. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):195–204, 2018.