A Generic Visualization Framework for Understanding Missing Links in Bipartite Networks

Jian Zhao
FXPAL
zhao@fxpal.com

Fracine Chen
FXPAL
chen@fxpal.com

Patrick Chiu
FXPAL
chiu@fxpal.com

ABSTRACT

The analysis of bipartite networks is critical in many application domains, such as studying gene expression in bio-informatics. One important task is missing link prediction, which infers the existence of new links based on currently observed ones. However, in practice, analysts need to utilize their domain knowledge based on the algorithm outputs in order to make sense of the results. We propose a novel visual analysis framework, MissBi, which allows for examining and understanding missing links in bipartite networks. Some initial feedback from a management school professor has demonstrated the effectiveness of the tool.

CCS CONCEPTS

• **Human-centered computing** → **Visual analytics**;

KEYWORDS

Missing link prediction, bipartite network, visualization.

1 INTRODUCTION

Many real-world complex systems can be modeled as *bipartite networks*. That is, there are two types of nodes in a network and links only exist between different node types. This kind of bipartite relationship analysis has been applied in many application domains such as investigating gene-expression networks in bioinformatics [4]. One interesting analysis for networks is *link prediction* (i.e., detecting missing links), which infers the existence of new relationships between nodes based on currently observed links [3]. Link prediction of networks is extremely useful in practice because real-world data is usually noisy or incomplete. For example, applying link prediction can guide laboratorial experiments rather than blindly checking all possible interactions between proteins [1].

However, while many link prediction algorithms have been proposed [5], the output is usually a list of scores or probabilities for all potential missing links, which is difficult to interpret; besides, these results can be inaccurate. In practice, analysts need to apply

their domain knowledge to examine the algorithm outputs. They may ask questions such as: why is this link identified as missing with high probability, does it make sense to have a link between these two nodes in the domain, and how will the network change by adding one or many detected missing links?

To address the above issues, we propose a generic visual analysis framework, MissBi, for detecting and examining missing links in bipartite networks. Most prior work in visualization has focused on displaying existing data in networks (e.g., Jigsaw [6] and Bicluster viewer [2]). However, MissBi visually presents detected missing links, allowing for a better understanding of the meaning and influence of missing links. The visualization enables the comparison of networks with and without specific links interactively added by analysts. This comparison is achieved through two of the most common network analysis approaches: metric-based (e.g., computing node betweenness) and motif-based (e.g., detecting cliques).

2 MISSBI FRAMEWORK

The MissBi framework consists of two main parts: an analysis module and a visualization module. The analysis module supports missing link prediction in bipartite networks as well as two of the most common ways for observing networks, including node metrics and sub-network motifs. The visualization module (Figure 1) displays all the outputs from the analysis module and enables analysts to explore the data with user interactions. An analyst can visually investigate the identified missing links, and further examine the influence of particular links by comparing the analytical results of the original network and the one with some missing links added.

MissBi supports the visual exploration of the predicted missing links through two views. The Network View (Figure 1a) displays 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 yellow-green and the missing links are shown in purple. Darker color reflects higher weight/score of a link. Moreover, the Link List View (Figure 1b) simply presents the missing links linearly by probability or score, where each link is visualized in a similar fashion to that in the Network View.

Motif analysis is one main approach to understanding the topology of a network. In bipartite networks, a bi-clique is one of the most important structural patterns. MissBi support this with a Detail View (Figure 1c) and an Overview (Figure 1d) for browsing the motifs at different scales. The Detail View shows bi-cliques as small multiples of matrices. The Overview displays all the bi-cliques as dots in a two-dimensional space based on MDS projection. These two views offer the visual exploration of all bi-cliques detected
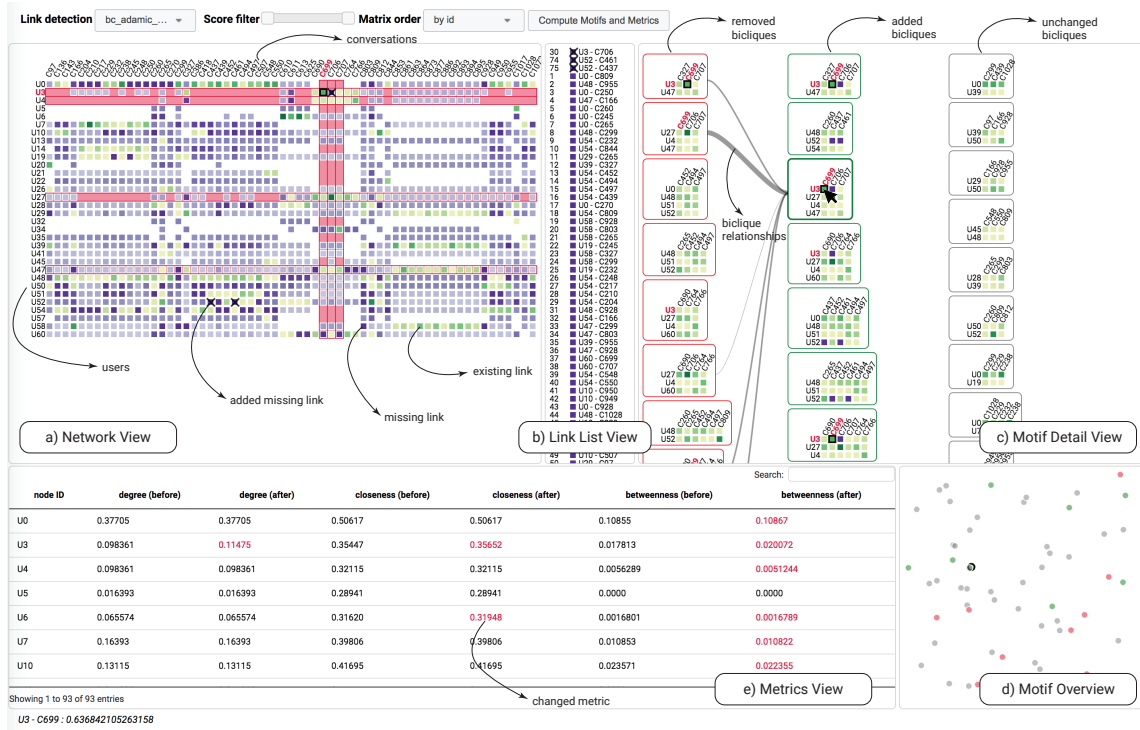
Figure 1: The visualization module of MissBi, which shows chat log data captured in a company.

in the network, and the investigation of the changes of results if certain missing links are added.

Computing node-metrics is a common way of getting a picture of the characteristics of a network in social sciences and other domains. The Metrics View in MissBi (Figure 1e) supports this kind of analysis by presenting a number of metrics in a traditional tabular view, including the degree, closeness, and betweenness centralities of before and after adding certain missing links.

3 APPLICATION ON ORGANIZATIONAL SOCIAL NETWORKS

To evaluate MissBi, we conducted interviews with a professor at the management school of a university. He has extensive research experience on analyzing the social dynamics of people in large organizations and online communities with quantitative methods. For his interview, he explored an employee-conversation network extracted from chat logs of an enterprise communication tool used in a company. That is, the network contains two types of nodes including employee and conversation, and a link represents that an employee participated in a conversation. He advocated that link prediction in such a network is critical, and it implies that some employees could also participate certain conversations.

The user appreciated the Motifs Detail View (Figure 1-c) where he could investigate the effect of missing links from the motifs analysis perspective and had a lower-level view of the network topology. He also liked the support of comparing detected motifs. During the data exploration, he said "*It seems this [missing] link is critical because it dramatically changes the cliques in the network,*" and he further hovered over specific bi-cliques, mentioning that "*This clique is mainly overlapped with these old ones, so not much*

new information is gained." Additionally, he commented that the detected missing links could be explored to verify some of the social science theories in practical datasets, for example, the *structural hole* theory.

4 CONCLUSION

We have presented MissBi, a general visual analysis framework for exploring and understanding missing links in bipartite networks. The framework offers an interactive visualization to present the missing links and investigate their meaning and influence to the network by comparing networks with and without selected missing links. Initial expert feedback has indicated that MissBi is useful and effective in analyzing networks in real-world applications.

REFERENCES

[1] 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), apr 2013.

[2] J. Heinrich, R. Seifert, M. Burch, and D. Weiskopf. Bicluster viewer: A visualization tool for analyzing gene expression data. In G. Bebis, R. Boyle, B. Parvin, D. Koracin, S. Wang, K. Kyungnam, B. Benes, K. Moreland, C. Borst, S. DiVerdi, C. Yi-Jen, and J. Ming, editors, *Proceedings of the International Symposium Advances in Visual Computing*, pages 641–652. Springer Berlin Heidelberg, 2011.

[3] 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, New York, NY, USA, 2003. ACM.

[4] 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, Jan. 2004.

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

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