Center for Visual & Decision Informatics

Predicting Future Relations: Incremental and Robust Link Prediction

PROJECT ID | CS15-2 **TYPE |** [X] New [ ] Continuing **START DATE |** July 2015

PROJECT LEAD/PARTICIPANTS | Ryan G. Benton, Suresh Golconda, Raju Gottumukkala, Christoph Borst

DESCRIPTION | Many datasets of interest today are best described as a linked collection of interrelated objects or, simply, linked data. In the bibliographic domain, for example, the objects are publications, authors and citations, while, in the biomedical domain, protein networks are an example of linked data. In sales data, the items sold together form a linked graph. Link prediction refers to the problem of discovering links that are expected to occur in future between objects that are currently not directly connected. Examples include discovering potentially new drug applications, potential promotional opportunities for sales, and new relations between users in social media.

The current NSF Fundamental Research Project is permitting CVDI to improve the state-of-the-art Link Prediction methods to incorporate link interestingness and, in some cases, the type of link; it also is allowing for testing of applicability in several domains. However, when finished (July), there will still be work required to move the research into robust practice. This includes permitting easy incorporation of new data, generating multiple models (to reduce single points of failure), visualizing the underlying graphs, and understanding the nature of the prediction. This proposal seeks to address these issues.

EXPERIMENTAL PLAN | During the project period, the team will work to:

- Develop interactive visualization techniques to explore the underlying graphs and to explain the predictions.
- Develop ensemble versions of the link prediction method.
- Incorporate incremental learning into the developed graph prediction methods.
- Identify problems relevant to the IAB that can be expressed as link prediction problems.

RELATED WORK | At present, the current-state-of-the-art method incorporates neighborhood features, random walk features, and semantically enhanced features into the learning process. It also handles the problem of unbalanced data, including how to generate labeled data (for supervised learning) automatically from the underlying graph. Finally, the fundamental research is expanding the link prediction to include link interestingness.

HOW OURS IS DIFFERENT | The current proposal is extending the Fundamental Research effort to include strong visualization capabilities, incorporating updates to the underlying graph without requiring complete retraining of the link prediction model, and creating the ability to merge the predictions of multiple link prediction models. This will improve the ability to (a) understand the predictions and (b) make the prediction more robust.

MILESTONES FOR YEAR |

6 months: Ensemble version of the Link Prediction method along with Interactive Visualization for single graph methods.

12 months: Incremental learning and visualization for Ensembles of graphs.

DELIVERABLES |

- Enhanced link prediction algorithms
- Demonstration system that exploits the CVDI developed Large-Scale Dynamic Graph Visual Analytic Methods
- Visualization tools

BUDGET FOR YEAR |

Students	$52,000
Equipment	$8,000
Travel	$4,800
Overhead	$6,480
Total	**$71,280**

ECONOMICS | A number of problems can be model as graphs, including literature, social media, biological data, drug interaction data, sales and sales/customer data, patient/doctor referrals and so forth. Being able to predict new linkages that will be occurring in the near-future allows organizations to better marshal and plan their resources. Examples include predicting new uses for drugs and identifying new customers for existing products.

POTENTIAL MEMBER COMPANY BENEFITS | The Link Prediction work will (a) provide a more robust set of techniques for determining future potentially actionable relationships, (b) leverage and enhance the IAB investments in Graph-Based Visual Analytic tools, and (c) complement the IAB investments in Visual Gap Analysis. Visual Gap Analysis allows for users to identify what is currently available (and what is missing) and allows for determining the impact of changes. Link Prediction, conversely, predicts what new relationships will occur (or disappear) given the current state.

PROGRESS TO DATE | We have a system that can predict future links for biomedical literature. We are currently exploring link strength prediction and link class prediction, and, by the end of the Fundamental Research Project, will also apply the Link Prediction work outside of the biomedical literature domain. CVDI has also developed a number of Graph Visualization Tools.

KNOWLEDGE TRANSFER TARGET DATE | 12 months