CHAPTER 8

CONCLUSION

8.1 OVERVIEW

Active research has been into the area of finding influential individuals in the social, citation and collaboration networks as this domain has got many potential applications useful to mankind. Though many solutions do exist, they are not effective in modeling the behavior of individuals in real world and do not study the evolving nature of these networks. This being the motivation for this work, a sincere attempt has been made in devising a more comprehensive solution mechanism by studying the small world characteristics inherent in these networks and their growth dynamics. Certain small world properties are effective in reducing the computational complexity involved in analyzing the patterns of interaction that occur among group of individuals. Further, the study on the network dynamics can elevate the chances of locating elite individuals that the user would be looking forward in a real world scenario.

8.2 COMMUNITY STRUCTURE AND THE ROLE OF WEAK TIES

In order to study the nature of interaction that occur among group of individuals in a social network, this proposed work adopts the game theoretic framework. The strength that an individual adds to a team is computed using Shapley value, the solution space of co-operative game theory. Shapley value provides a fair means of dividing the gains among the
individuals proportional to their contribution. However, the problem of finding the most influential nodes using naïve Shapley value is intractable. The proposed strategy attempts to reduce the computational complexity by exploring the community structure inherent in the social networks and the weak ties that connect the communities together. Initially, communities are identified based on the influencing ability of the individuals and thus individuals within a community can influence others in their community. To increase the influence spread in the whole network, bridge nodes that connect communities together are identified within each community. The influential nodes are then identified only among the bridge nodes based on the refined Shapley value approach, thus reducing the computational complexity.

8.3 TEMPORAL CO-CITATION AND TOPIC SPECIFIC PAGERANK

Effective techniques to track the changes that occur due to the evolving nature of the citation networks are proposed in this research work. A novel co-citation counting technique has been devised to include temporal dimension in its analysis. This proposed strategy in the domain of author co-citation analysis is thus found to be effective in locating influential as well as active nodes in the citation network. PageRank, a link analysis algorithm is adapted to include topic-specific information and temporal dimension in its analysis. Due to this, significance measures of the links are extended to include similarity between documents and the publication time period of the documents connected by these links. The proposed strategy in the domain of link analysis is thus found to be effective in locating influential and active nodes in the knowledge domain pertaining to the given topic.
8.4 DYNAMIC KNOWLEDGE MODEL

To capture the learning ability of the individuals in a time evolving co-author collaboration network, a dynamic knowledge model is proposed as part of a referral system in this research work. The proposed model is based on a learning factor that incrementally updates the knowledge of the researchers to avoid global expertise re-computation. The proposed system also intends to maintain the privacy of the users by passing the query to only a few immediate neighbors selected based on their proficiency instead of flooding the same to all of its neighbors.

8.5 MULTIPLE EXPERTISE IDENTIFICATION SYSTEM

The proposed strategy in the content analysis solution space attempts to find influential nodes with multiple expertise in an academic domain. This is one of the pressing needs in the real world scenario as many research works carried out are interdisciplinary in nature. TV tree a data structure for representing high dimensional vector space documents is used to represent the researchers profile and effectively retrieve documents related to the user query. An appropriate author score is also proposed to find the relative expertise of the researchers in various domains based on these set of retrieved documents.

8.6 SUMMARY AND LIMITATIONS

In this thesis, solutions to locate influential nodes in the various solution spaces are suggested. A game theoretic framework based on small-world characteristics in the context of maximizing the influence spread is proposed and evaluated with real time datasets. Methods to analyze the evolving nature of the citation and scientific collaboration were evaluated with ACM TODS, KDD datasets and ACM CCS. Effectiveness of the
proposed methods has been tested in terms of the number of nodes influenced, number of active researchers found, entropy of the author clusters and a comparative analysis was also executed to investigate the efficiency of the work.

However, the real time datasets used in this thesis for assessing the game theoretic and author co-citation solution spaces are limited in terms of the topology size. Edge weights that represent the influence measure of connected individuals are assigned randomly and the link structure is assumed to undirected in the game theoretic study. This undirected link approach enforces a constraint that two nodes mutually influence each other, which is an unlikely real world scenario.

In the link analysis approach, investigation on the various parameters of the LDA model that has significant impact in the number of topics retrieved is not carried out. Further the referral systems and the content analysis approaches were evaluated with the fixed size topic taxonomy.

8.7 FUTURE WORK

As a future work, the analysis could be carried out in an online social networking environment to monitor the real time interaction among the individuals. The efficiency of the independent cascade model in the information diffusion process and comparative analysis with the linear threshold model used in this thesis can be done to study the suitability of these models in different contexts.

Experiments on the other solution concepts of the cooperative game theory such as core and nucleolus that possibly has important implications for discovering influential nodes could be carried out. Edge weights in the game theoretic approach could effectively be assigned based on
the private information that the node may have about its neighbors. Such scenarios can effectively be modeled using mechanism design in game theory.

Visualization techniques can be studied to effectively portray the author clusters and their interactions in the citation analysis and content analysis method.

Comparative analysis of the efficiency of the LDA with other state-of-art topic models could be carried out. Generalized term vector models that can accommodate variable number of topics can eliminate the fixed size topic constraint in the referral systems and content analysis method.