The Proceedings of the 9th Annual International Digital Government Research Conference

CI-KNOW: Recommendation based on Social Networks

Yun Huang, Noshir Contractor, York Yao
Northwestern University
2145 Sheridan rd, TECH C210
Evanston, IL 60208, U.S.A.
{yun, nosh, yao.gyao}@northwestern.edu

ABSTRACT

Digital media and communication networks have become an important cyberinfrastructure to enable new levels of interactions in organizations and communities. A complicated knowledge network of individuals, documents, data, concepts, and their interconnections forms a virtual knowledge repository. To be more effective in using these resources, knowledge discovery tools are crucial for an organization and individual users to identify the right expertise or knowledge resources from this large "multidimensional network."

Cyberinfrastructure Knowledge Networks on the Web (CI-KNOW) is a suite of Web-based tools that facilitates discovery of resources within communities. CI-KNOW implements a network recommendation system that incorporates social motivations for why we create, maintain, and dissolve our knowledge network ties. The network data is captured by automated harvesting of digital resources using Web crawlers, text miners, tagging tools that automatically generate community-oriented metadata, and scientometric data such as co-authorship and citations. Based on this knowledge network, the CI-KNOW recommender system produces personalized search results through two steps: identify matching entities according to their metadata and network statistics and select the best fits according to requester's perspectives and connections in social networks.

Integrated with community Web portals, CI-KNOW navigation and auditing portlets provide analysis and visualization tools for community members and serves as a research testbed to examine social theories on individuals' motivations for seeking expertise from specific resources (people, documents, datasets, and etc.). As a proof-of-concept, this paper demonstrates how CI-KNOW, integrated with the NCI-supported Tobacco Informatics Grid (TobIG), facilitates knowledge sharing in the tobacco control research community.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval – *search process, selection process;* H.4.m [**Information Systems**]: Miscellaneous*;* J.4 [**Computer Applications**]: Social and Behavior Sciences – *economics, psychology, sociology.*

General Terms

Management, Measurement, Design, Experimentation, Human Factors.

Keywords

collaboration network, community, cyberinfrastructure, scientists' networks, knowledge discovery.

1. INTRODUCTION

Recent advances in Web 2.0 and cyberinfrastructure have enabled new levels of interactions and interconnections among individuals, documents, data, analytic tools and concepts. On the other hand, Web 2.0 applications also take the advantage of the rich sources of information and digital traces of human social interactions and provide recommendation services such as product promotion (e.g. Amazon.com shopping cart recommendations) and online advertising (e.g. Google AdSense).

Many online communities and interactive collaboration spaces (like forums and wikis) eventually became large-scale knowledge networks which are context-dependent and multi-dimensional. In a collaborative environment, users publish, search, and consume knowledge products and communities and Web tools predict and match users' demand. The user engagement and values of communities highly depend on how well they fulfill users' "search."

Knowledge networks and recommender systems are especially important for scientific research communities. Computer-mediated communication (CMC) such as emails, electronic publishing, wikis, and online data sharing has become a major part of the cyberinfrastructure to facilitate academic research and encourage transdisciplinary collaboration. For academic communities to be more effective in using their resources, it is even more crucial that they have tools help them identify the right expertise or knowledge resources from within this large "multidimensional network."

Cyberinfrastructure Knowledge Networks on the Web (CI-KNOW) is a suite of Web-based tools that facilitates Discovery of resources within communities. CI-KNOW facilitates discovery ("If only the community knew what the community knew") by implementing a network recommendation system that incorporates social motivations for why we create, maintain, and dissolve our knowledge network ties.

The major contribution of CI-KNOW recommender system is to personalize the search process and results based on Monge and Contractor's Multi-Theoretical, Multi-Level (MTML)[8] which investigates social drivers for organizing networks in communities with diverse goals such as exploring new ideas and resource, exploiting existing resources and capabilities, social bonding, mobilizing for action, and rapid response (or "swarming"). These drivers may act alone or in concert and in differing amounts within and across communities.

CI-KNOW captures network data by automated harvesting of digital resources using Web crawlers, text miners, tagging tools that automatically generate community-oriented metadata, and scientometric data such as co-authorship and citations. Based on this knowledge network, the CI-KNOW recommender system produces personalized search results through two steps: identify matching entities according to their metadata and network statistics and select the best fits according to requester's perspectives and connections in the social network.

Integrated with community Web portals, CI-KNOW navigation and auditing portlets provide analysis and visualization tools for community members and serve as a research testbed to test social networks models about individuals' motivations for seeking expertise from specific resources (people documents, datasets, etc.). As a proof-of-concept, we use the NCI-supported Tobacco Informatics Grid (TobIG) to demonstrate how CI-KNOW facilitates knowledge sharing in the tobacco control research community.

2. RELATED WORK

Adomavicius and Tuzhilin [1] present an overview of the field of recommender systems and classify recommendation methods into three main categories: content-based, collaborative, and hybrid recommendation approaches. In this section, we review related research on recommendation in communities and knowledge networks, as well as the literature on the applications of social network analysis in online recommendations.

Collaborative recommender systems (or collaborative filtering systems) explore the similarities among the "relations" between people and rated items. Although better than content-based recommendation approaches in mimicking the human recommendation process, most these human-item relations used by collaborative systems are not "real" social relations because they do not capture users' individual interactions. To extract more information from human connections, some methods integrate recommender systems with social relationship and network statistics.

Expertise Recommender [6] is a system designed to locate expertise within organizations such as a software house. The authors analyze the practices of software development and customer supporting and try to automate the recommendation of programmers and client representatives. The recommender determines coding expertise through the so called "Line 10 Rule": a programmer, who has a problem in changing code, looks into the version control system to see who last modified the code and then approaches the person for help. Client support experts are identified through another heuristic: the support representative scans the records of past support interventions looking for the person solved similar problems before. Using these two relations, as well as friendship and "departmental" relations, the recommender profiles, identifies, and selects experts for a specific problem.

IKNOW (Inquiring Knowledge Networks on the Web) [3] is a knowledge management tool based on social network analysis. In order to exchange knowledge, IKNOW emphasizes both the importance of experts and the perception of where the expertise is. In other words, IKNOW not only recommends the expert of specific knowledge but also enhances user's perception of the distribution of knowledge in an organization or a community. A visualization tool, as Referral Web, helps people discover their structural positions and the spread of expertise in various relation networks. Differently from collaborative systems, IKNOW recommends the linkages between people not only on the basis of similarity but potentially on the basis of different network theoretical mechanisms.

Horting [2] proposes a collaborative filtering method, based on graph analysis, for the generation of recommendations. The system computes two different relations between users: horting and prediction. User 1 "horts" user 2 if user 1 has in common with user 2 over a certain threshold. A horting relation is therefore directed and asymmetrical. User 2 "predicts" user 1 when user 1 horts user 2 and the differences in their ratings are below a certain threshold. When a user queries an item, the recommender will check other nodes he has prediction relations with. If these nodes have not rated the item, the system will try the next level through different paths of prediction relations. When the focal user finally reaches the nodes that have already rated the item, a prediction can be generated. Although it is a collaborative filtering system, Horting method takes transitivity into account, i.e. predictions are not just based on similar users, but based on users similar to similar users.

Graph-model for e-Commerce recommender systems [5] develops a two-layer graph model with the product layer and the customer layer. The product layer is a network of products and their links representing their similarities; the customer layer is a network of users and their links representing the similarity computed based on demographic data or tastes. Interlayer links represent relations between products and users such as purchasing and rating. The product layer represents content-based information, and the customer layer represents information for collaborative filtering system. With the interlayer links, the hybrid system can perform direct retrieval, association mining, and high-degree association retrieval.

GenialChef [9] is a multi-agent system that recommends restaurants. In the system there are different types of agents: service agents and personal agents (PAs). Service agents provide information on the restaurants to the PAs, and help PAs in finding other PAs. PAs provide personalized information to the user. Every user has a PA which reflects his/her user profile and which is in charge of recommending useful restaurants. Whenever a PA needs a recommendation, it asks other PAs for help. PAs communicate only with reliable PAs who recommended good restaurants to the focal PA in the past. A trust network is therefore created among PAs for recommendation.

GroupLens on the recommending of research papers [7] tests the ability of various algorithms in recommending citations to a user who needs to build a reference list for a target research paper. They compare six algorithms for selecting citations: Co-citation Matching, User-Item CF, Item-item CF, Naïve Bayesian Classifier, Localized Citation Graph Search, Keyword Search (Google). The authors suggest to create recommender systems that are based different algorithms according to different usage scenarios.

Most recommender systems predict items on the basis of one theoretical mechanism: homophily between user preferences. Monge and Contractor [8] provide a Multi-theoretical multi-level (MTML) to generate predictions in multidimensional networks. The combination of social theories such as Interdependence and Transactive Memory Theory reveals the social drivers of experts and celebrates the values of recommendations for a specific user. The recommendations are based on ratings by experts instead of the same keywords and user's neighborhood.

	Exploring	Exploiting	Mobilizing	Bridging	Bonding	Swarming
Self-Interest	New Unique Scanning		Established Common Multiple Connections	External Diffusion		
Collective Action		High Centrality Internal Diffusion Robustness Prestige Absorptive Capacity	High Centrality Internal Diffusion Robustness Prestige Absorptive Capacity			High Centrality Internal Diffusion Robustness Prestige Absorptive Capacity
Cognition	Inferred Relations	Inferred Relations				Inferred Relations
Balance	Distance > 2 Unconnected		Distance Two	Distance Two	Distance Two	
Exchange		Mutuality Subnetwork of People		Mutuality Subnetwork of People	Mutuality Subnetwork of People	
Contagion	Rate of Access Degree Prestige		Rate of Access Degree Prestige			
Homophily	Different on Attribute			Scanning Different on Attribute	Similar on Attribute	
Proximity	Geographically Topologically Distant		Geographically Topologically Near	Geographically Topologically Distant	Geographically Topologically Near	Geographically Topologically Near

Table 1. Measurements of social drivers in organizational networks based on social theories

3. RECOMMENDER SYSTEM BASED ON MTML

3.1 Measurement of social drivers

The recent study of organizations from a social network perspective identifies the multiple theoretical mechanisms that contribute to the emergence of organizational networks. Contractor et al. [4] empirically test multi-theoretical, multilevel (MTML) hypotheses about the structural tendencies of networks. The study suggests that the social drivers for organizing networks in communities have diverse goals such as exploring new ideas and resource, exploiting existing resources and capabilities, social bonding, mobilizing for action, and rapid response (or "swarming"). These drivers have different levels of impact within and across communities and hence change the mechanisms of recommender systems. For example, looking for a quick solution and producing innovative ideas will result in completely different recommendation systems.

By extending the MTML into the design of recommender systems, CI-KNOW allows users to identify which type of activity (social drivers) they are participating in and then favoring those selection measurements that support recommendations to improve that type of behavior. In Table 1, we propose the measurements of the five social drivers according to different social theories.

3.2 Three-layer model with multiple networks

In order to performance recommendation large-scale multi-dimensional networks, we propose a three-layer model to represent various relations in knowledge networks.

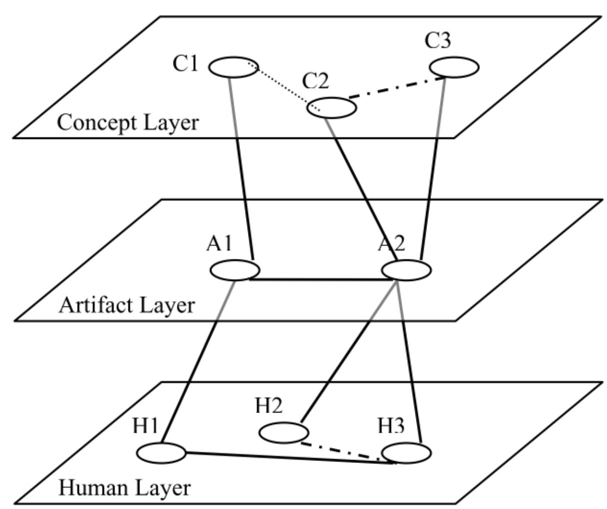

Figure 1. Three-Layer Multi-network Model.

In Figure, the three-layer model represents the main elements of CI-KNOW recommender system. The top concept layer represents the content domain of a knowledge network and consists of all knowledge entities such as keywords, research tops, and products and their semantic networks. The human layer is a network of people and experts in the community. The intermediate artifact layer represents all information entities that can bridge the concept layer and the human layer and establish relations among them. The information entities could be articles, projects, organizations, emails, which connect concepts, users, and other information entities. For example, H1 could be an editor of a journal A1 which focuses on social network analysis (C1). Therefore the information entity A1 establishes a relation between H1 and C1. A2 can be a research paper in the journal A1 coauthored by H2 and H3 with keywords MTML (C2) and recommender systems (C3). The information entity A2 generates many potential relations: coauthorship between H2 and H3, keywords relation between C2 and C3, H2 and H3's expertise indicated by keywords C2 and C3, potential interests of keyword C2 for the journal A1, etc.

The three-layer model is very flexible in preserving various types of relations: semantic networks in the concept layer, social networks in the human layer, and association relations in the artifact layer and between layers. The combination of different relations can produce more information about user behavior and the knowledge domain. For example, emails or wedding invitations as association relations between human and artifact layers can be projected into friendship relations in the human layer. Therefore CI-KNOW is capable to work with multiple networks with complex relations.

3.3 Recommendation procedure

For applications in research communities, we use keywords, persons, and all research articles to indicate the three layers. Using the following processes, CI-KNOW produces internal data structure for calculation.

1. Generate a person-keyword matrix is generated and each cell contains the relative centrality of the concept for the individual's corpus of text.

2. The person-keyword matrix is then dichotomized using average keyword centrality across users as the basis for conversion. The resultant matrix is a person-keyword binary matrix, indicated by (PM2).

3. Generate a person-article binary matrix (PM1).

   a. When two persons are identified as authors of the same article, links are created to the existing article node to avoid article duplication.

   b. Co-authorship information is saved as CR.

4. Abstracts of articles are compared against the keywords list to generate an article-keyword binary matrix (PM3). A "1" or "0" indicates the presence or absence of the keyword in the abstract.

5. These matrices are combined into a single person-article-keyword undirected binary matrix. The resulting partition is shown in Table 2.

Table 2. Person-article-keyword matrix

	Person	Article	Keyword
Person	**CR**	**PM1**	**PM2**
Article	**PM1'**		**PM3**
Keyword	**PM2'**	**PM3'**	

Overall confidence index between user i and recommendation j $\mathbf{C}_{ij}$ is based on structural information about the number of steps between the search keyword k and the recommended item j, the profile similarity (positive matches, structural equivalence) between the search keyword k and the recommended item j, and the number of steps between the user i and the recommended item j.

$$C_{ij} = \frac{[G_{max} - G_{jk}] + [P_{jk}/P_{max}] + [(D_{max} - D_{jk})/(100 * D_{max})] + [1/G_{ij}]}{G_{max} + 1 + 0.01 + 1}$$

This can be interpreted as an initial identification stage based on geodesic distance (coarse identification), positive matches (medium identification) and profile similarity (fine identification) that returns the same scores for all users based on the search keyword and recommended item followed by a selection stage that incorporates information about the relationship between the user and a potential recommendation to arrive at a final score. This score is scaled by the observed global maximum to return confidence measures that are between zero and one. The data matrix that is used for these calculations must be square, but need not be symmetric.

The detail algorithm is in the appendix.

4. CI-KNOW Architecture

In this section, we discuss some implementation details of CI-KNOW.

4.1 Elements of CI-KNOW

Multi-disciplinary research and policy communities have indicated substantial interest in a network referral system that would identify available resources of interest to their members. For example, researchers with similar or complementary expertise, popular and relevant data sets, and appropriate analysis tools.

To meet this interest, CI-KNOW is designed to integrate into portals and other collaborative environments. CI-KNOW is a suite of software tools that takes advantage of automated data collection and network extraction techniques to infer a scientist's interests and expertise based on their interactions and activities within a context-dependent knowledge network. It implements advanced algorithms to provide network referrals that take into account socio-technical incentives.

At the user and administrator levels, CI-KNOW provides tools for:

Knowledge network recommendations and recommendation pathway visualizations

Network neighborhood visualization and navigation

Global network visualization/navigation

Portfolio management and network diagnostics subsystems

4.2 CI-KNOW Metadata Taxonomies

When individuals log in and use a collaborative environment, their behavior is logged within a unique session that is linked to their user ID. Each session stores use logs as Resource Description Framework (RDF) triples that can be called 'metadata' Metadata summarizes objects (whether users or items), locates objects in geographical space, indexes content, stores and defines network ties among objects, and records usage and access information.

A triple is information stored in the form: <subject><predicate><object>, e.g. <user 1><is author of><forum post 1>.

Predicates form the structural linkages between items in the knowledge network. CI-KNOW harvests structural metadata to record connections among entities in the knowledge network. Metadata are classified according to the context from which the data are drawn and the primary descriptive characteristics that the data possess. Figure 2 summarizes all data used for recommendation.

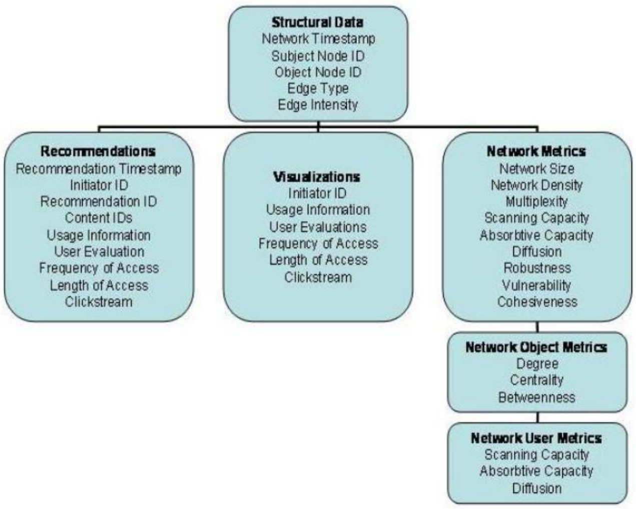

Figure 2. CI-KNOW data structure.

4.3 CI-KNOW Recommender System

CI-KNOW tools are developed as independent, free-standing web applications that can also appear as portlets within a portal or collaborative environment. The network referral system is presented to the community through an intuitive user-interface.

CI-KNOW continuously retrieves data from the community, processes these data, and uses internal network referral algorithms to increase the relevancy and timeliness of its referrals within the community.

Network navigation tools allow users to locate themselves or their colleagues, documents or workspaces in the knowledge network and to explore their networks through a visual, step-wise process. In the near future, users will have the option to explore these networks in an optimal 'topological' space, or to have the networks connected to geographical maps through an approach similar to the one used by Google Maps.

Traditional referral systems do not take into account: (i) multiple relationships that link knowledge network entities; (ii) the multiple socio-technical incentives that influence the effectiveness of a network-based recommendation. Recent developments in Social Network theories and methods provide ways to fill in this gap.

CI-KNOW recommendations emerge from a modular system that allows for the combination and hybridization of multiple mechanisms for creating recommendations from structural metadata. Modules are based on social network analysis theory and methods and can be updated at any time.

Current and Planned Modules:

Structural similarity of users and items to search keywords: what is most alike?

Structural closeness of users and items to search keywords what is closest?

Popularity of an item: what is conventional?

Structural influence: Who or what are key links in the network?

Innovation: what's new?

Burst activities: what's 'hot'?

Exchange: who you previously helped and now "owes you one"?

Proximity: who is geographically near you "at arms length"?

4.4 Portal Interfaces

Rather than providing traditional search results (such as documents that contain a specific keyword), users receive recommendations to entities (datasets, documents, people) that are associated through transactions, interactions, and interconnections.

Recommendations are based on algorithms that consider the lengths of the paths between relevant entities, any common or shared connections between entities, and other indicators that proxy the social motivations and incentives for knowledge sharing.

The CI-KNOW portlet window "returns" a folded list of recommendations to users while they are working inside the portal. The recommendation itself is an active link to the contact information for the user or to the particular forum post, document, or data that is recommended (Figure 2).

Clicking on the "Why" button in the portlet provides more detail about why a user has been given a recommendation. CI-KNOW provides a textual description and a network visualization of the path between the user, the search keyword and the recommendation.

The 'Tell Me Why' page presents information about the shortest path between the user, the search keyword and the recommendation. With one click, the user can show all the shortest paths, hide node labels, and access more tools for Knowledge Network exploration

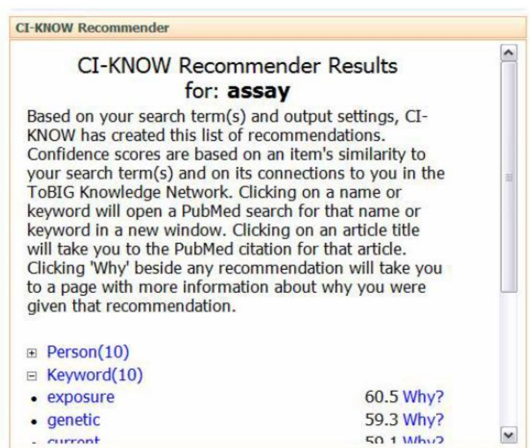

Figure 3. CI-KNOW recommendation results.

5. CASE STUDY – The Tobacco Informatics Grid (TobIG)

The CI-KNOW recommender system has been integrated to provide the collaborative backbone of the NCI-supported tobacco Informatics Grid (TobIG). Created by the SONIC group to support the activities of tobacco researchers and policy makers, the TobIG is a proof-of-concept integration of CI-KNOW with other collaboration and analysis tools to provide knowledge network services within a grid computation framework.

The Tobacco Informatics Grid (TobIG) is a portal environment deployed to facilitate the integration of people, data and resources to enable and catalyze discovery in tobacco control. TobIG is unlike traditional grid computing environments in that it extends the concept of "discovery" beyond dataset and analysis toolkits and toward the inclusion of people and their relationships as a "resource" that can be located and activated by users of cyberenvironments.

TobIG offers a set of collaborative tools that support synchronous and asynchronous collaboration as well as the development of virtual sub communities of practice within areas of interest in tobacco research. TobIG, with its proof-of-concept implementation of CI-KNOW, offers the opportunity to link social network approaches to data, people and resources with computational infrastructures and analytic tools to advance research in the population sciences. It is expected to reducing the barriers to collaboration often experienced in multi-site, "team science" research endeavors.

The CI-KNOW interface can be found on each page of TobIG to allow users to search for recommendations from anywhere in the portal.

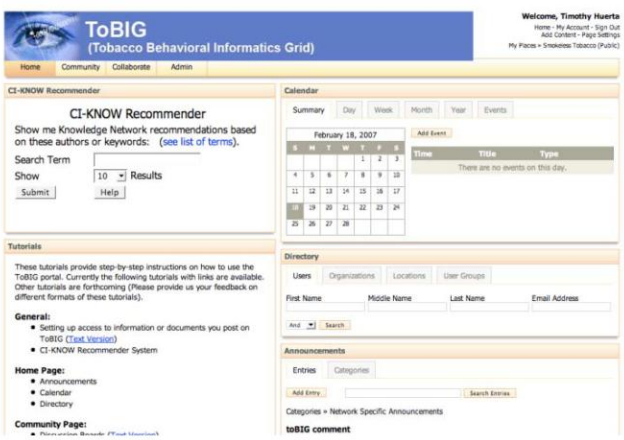

Figure 4. TobIG portal and CI-KNOW recommender.

CI-KNOW recommender illustrates knowledge networks in TobIG (Figure 5), provides recommendations with confidence levels, visualizes recommendation context (Figure 6), and, when relevant, links to external data sources such as PubMed.

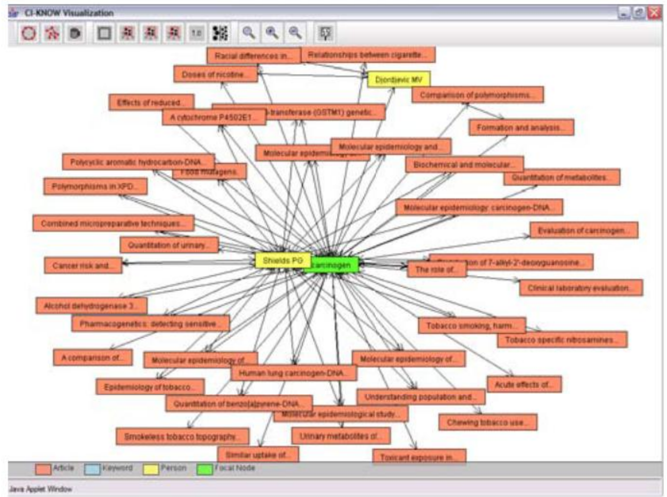

Figure 5. Visualize a network of parsons, articles and keywords in TobIG.

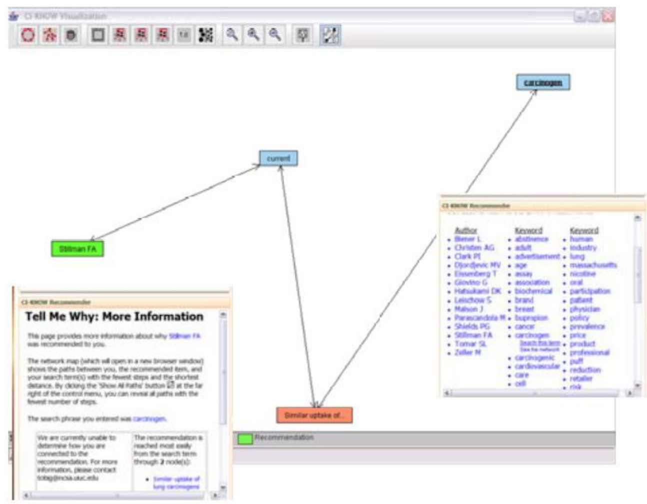

Figure 6. Visualize recommendation results.

6. SUMMARY AND FUTUREWORK

In this paper, we have described the design and implementation of CI-KNOW, a recommender system for locating resources in a knowledge network. The key advantage of CI-KNOW is its ability to work with multidimensional networks and different organizational goals. Based on MTML and various social theories, CI-KNOW explores the relations among users and contents, performances recommendations based on different social drivers, and visualizes recommendation results.

There are three major limitations of current CI-KNOW implementation. First, because the concept layer is finite and pre-determined, a search keyword must match an item that is recognized by the system. We will incorporate some content-based data mining techniques to automatically build new knowledge entities from object entities. Second, the presence of multiple shortest paths is also important and will be incorporated as an improvement to the current recommender algorithm. Third, because of sparsity problems, some areas of our primitive data matrix are undefined. We will adopt a squared Euclidean distance calculation that allows us to weight those undefined areas by zero and remove them from the distance calculation.

7. ACKNOWLEDGMENTS

CI-KNOW development is supported by NSF grants BES-0414259, BES-0533513, IIS-0535214, SBE-0555115, and SCI-0525308 and Office of Naval Research grant N00014-04-1-0437.

8. REFERENCES

[1] Adomavicius, G., and Tuzhilin, A. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. Transactions on Knowledge and Data Engineering, 17:6 (June 2005), 734-749.

[2] Aggarwal, C.C., Wolf, J.L., Wu, K-L., and Yu, P.S. Horting Hatches an Egg: A New Graph-Theoretic Approach to Collaborative Filtering. Knowledge Discovery and Data Mining (1999), 201-212.

[3] Contractor, N., Zink, D., and Chan, M. KNOW: A Tool to Assist and Study the Creation, Maintenance, and Dissolution of Knowledge Networks. Lecture Notes In Computer Science; Vol. 1519 (1998) Springer-Verlag, 201-217.

[4] Contractor, N., Wasserman, S., and Faust, K. Testing multi-theoretical multilevel hypotheses about organizational networks: An analytic framework and empirical example. Academy of Management Review, 31:3 (2006), 681-703

[5] Huang, Z., Chung, W., and Chen, H. A Graph Model for E-Commerce Recommender Systems. Journal of the American Society for Information Science & Technology 55:3 (2004), 259-274.

[6] McDonald, D.W., and Ackerman, M.S. Expertise recommender: a flexible recommendation system and architecture, Proceedings of the 2000 ACM conference on Computer supported cooperative work (Philadelphia PA, 2000), ACM Press, 231-240.

[7] McNee, S.M., Albert, I., Cosley, D., Gopalkrishnan, P., Lam, S.K., Rashid, A.M., Konstan, J.A. and Riedl, J. On the Recommending of Citations for Research Papers. In Proc. of ACM CSCW 2002, ACM Press (2002), 116-125.

[8] Monge, P. R., and Contractor, N. Theories of Communication Networks. New York: Oxford University Press. 2003.

[9] Palau, J., Montaner, M., López, B., and Rosa, J.L. Collaboration Analysis in Recommender Systems Using Social Networks. Proceedings of CIA 2004 (Erfurt, Germany, September 2004), 137-151.

9. APPENDIX – CONFIDENCE CALCULATION

A. For structural information between keyword k and recommended item j ($\mathbf{G}_{max}$- $\mathbf{G}_{jk}$):

1. Assume that the matrix represents an actor by actor matrix
2. Assume that the shortest path (geodesic) between two items represents a measure of structural 'reachability'
3. Calculate geodesics between j and k from original data, $\mathbf{G}_{jk}$.
4. Smaller geodesics represent greater reachability

To favor similar items in our calculation, we subtract the geodesic distance between the search keyword and the recommended item ($\mathbf{G}_{jk}$)from the global observed maximum geodesic distance ($\mathbf{G}_{max}$)

B. For positive matches [P_{jk}/P_{max}]:

1. Assume that the data matrix represents an actor by variable matrix
2. The search keyword identifies a row in the data matrix (k)
3. Compare rows (j,k) by counting the number of positive matches between the two vectors.
   a. Computationally this is done by taking changing the difference in the standardized Euclidean distance formula into a product.
   b. Variables are standardized to account for scaling differences.
   c. Standardization is completed by multiplying the original data matrix by a diagonal matrix that contains the reciprocal of the variance for each variable.
4. Larger counts represent greater similarity
   a. To return this measure to a scale similar to those in the geodesic calculations, we standardize by the observed global maximum count of positive matches.

C. For profile similarity [($\mathbf{D}_{max}$-$\mathbf{D}_{jk}$)/100*$\mathbf{D}_{max}$]

1. Assume that the data matrix represents an actor by variable matrix
2. The search keyword identifies a row in the data matrix (k).
3. Compare rows (j,k) by calculating standardized Euclidean distance, $\mathbf{D}_{jk}$.
   a. Distance is standardized by variable to account for possible differences in scale across variables, not within cases.
   b. Standardization is completed by multiplying the original data matrix by a diagonal matrix that contains the reciprocal of the variance
4. Smaller distances represent greater similarity in profiles
   a. To favor similar items in our calculation, we subtract the distance between the search keyword and the recommended item ($\mathbf{D}_{jk}$)from the maximum Standardized Euclidean distance ($\mathbf{D}_{max}$).
   b. To return this measure to a scale similar to those in the geodesic calculations, we standardize by 100 times the observed global maximum standardized Euclidean distance.

D. For structural information between user i and recommended item j (1/$\mathbf{G}_{ij}$):

1. Assume that the matrix represents an actor by actor matrix
2. Assume that the shortest path (geodesic) between two items represents a measure of structural 'reachability'
3. Calculate geodesics between i and j from original data, $\mathbf{G}_{ij}$.
4. Smaller geodesics represent greater reachability
   a. To favor closer items but not overpower selection criteria, take the reciprocal of the geodesic [1/ $\mathbf{G}_{ij}$]