

Cyberinfrastructure Knowledge Networks on the Web

A Recommender System for Locating Resources in a Knowledge Network

Noshir Contractor

Professor, Departments of Speech Communication,
Psychology, Coordinated Science Laboratory, and
Graduate School of Library and Information Science

Director, Science of Networks in Communities
Research Group
244 Lincoln Hall

702 S. Wright Street

Urbana, IL 61801
011-217-333-7780

nosh@uiuc.edu

Harold D. Green, Jr.

Research Scientist, Science of Networks in
Communities Research Group
National Center for Supercomputing Applications
1205 W. Clark Street, Room 1008

Urbana, IL 61801
011-217-333-8759

hgreenjr@ncsa.uiuc.edu

ABSTRACT

A Knowledge Network is a multi-dimensional network representing the interactions and interconnections among people, documents, data, analytic tools, and interactive collaboration spaces (like forums and wikis). CI-KNOW is a suite of software tools that leverages automated data collection and social network theories, analysis techniques and algorithms to infer an individual's interests and expertise based on their interactions and activities within a knowledge network. Recent developments in social network theories and methods provide the backbone for a modular system that creates recommendations from relational metadata. The CI-KNOW recommender system is a modular tool that has been integrated into a collaborative research portal to provide the social networking capabilities for tobacco researchers, policy experts, and government representatives engaged in tobacco surveillance research. The CI-KNOW Recommender harvests the knowledge network associated with a communities' use of cyberinfrastructure tools. It combines these data with other relational digital traces about the community such as (co)authoring, (co)citation, funding, and patent data. A network navigation tool allows users to locate colleagues, documents, data or analytic tools in the network and to explore their networks through a visual, step-wise process. An auditing tool offers network leaders, managers, and funders diagnostics to assess the growth and health of the entire Knowledge Network. CI-KNOW development is supported primarily by grants from the National Science Foundation and the National Institutes of Health.

Categories and Subject Descriptors

K.4.3 [Computers and Society]: Organizational Impacts – Computer-supported collaborative work.

H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval – Search process, Selection process.

H.2.8 [Database Management]: Database Applications – Database Mining.

J.3 [Life and Medical Sciences]: Health.

General Terms

Algorithms, Design, Human Factors, Management.

Keywords

Social Network Analysis, Cyberinfrastructure, Recommender System, Tobacco Control, Knowledge Network, CI-KNOW.

1. PROJECT SUMMARY

Multi-disciplinary research and policy communities, such as those engaged in tobacco control and related activities, 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, new, or relevant data sets
- Appropriate analysis tools
- Models or model systems
- Visual analytic tools, workflows, etc

To meet this interest, the Science of Networks in Communities (SONIC) group's social networking technology, Cyberinfrastructure Knowledge Networks on the Web (CI-KNOW), is integrated into portals and other collaborative environments. CI-KNOW is a suite of software tools that leverages automated data collection and network extraction techniques to infer a scientist's or policy maker's interests and expertise based on their interactions and activities within a knowledge network. It implements advanced algorithms to provide network referrals that take into account socio-technical incentives for knowledge sharing.

A knowledge network is a multi-dimensional network created from the interactions and interconnections among people, documents, data, analytic tools, and interactive collaboration spaces (like forums and wikis) associated with a collaborative environment.

This kind of network grows organically as users publish documents, access data, use analysis tools, post to forums, etc. Linkages among items in the knowledge network are stored as network meta-data, which is accessed by CI-KNOW tools. The CI-KNOW recommender system mines the knowledge network associated with a community's use of cyberinfrastructure tools or portal environments, combining that data with information gathered from other sources like bibliographic databases and patent information to extend the knowledge network into relevant conceptual spaces and recommend items that may be of interest to portal users.

The NCI-sponsored Tobacco Informatics Grid, or TobIG, is one such proof-of-concept implementation. TobIG is a portal environment deployed to facilitate the integration of people, data and resources to enable and catalyze research and policymaking in tobacco control. TobIG offers a set of collaborative tools that support synchronous and asynchronous collaboration as well as the development of virtual 'communities of practice' within areas of interest in tobacco research.

CI-KNOW tools are developed as independent web applications that can appear as portlets within a portal or collaborative environment. The CI-KNOW interface resides on each page of the TobIG portal, allowing users to search for recommendations anywhere in their research environment. Additional tools to support collaboration are also provided in the portal. CI-KNOW

continuously retrieves data from the community and their use of these collaboration tools. CI-KNOW processes these data into knowledge network structures, and uses internal network referral algorithms to increase the relevancy and timeliness of its referrals to community members. In general, the functions of CI-KNOW are to provide users with knowledge network visualizations, targeted recommendations, and visual recommendation pathways that clarify the connections between a user and a recommended item. Searches within CI-KNOW provide access to recommended results, information about the 'confidence' associated with a particular recommendation, explanation of the linkages among items in the knowledge network, and connection to external data sources such as PubMed when relevant.

TobIG is unlike traditional grid computing environments in that it extends the concept of "discovery" beyond data and analytic tools by including people and their relationships as a resource that can be accessed and activated. CI-KNOW offers the opportunity to link social network approaches to data, people and resources with other computational infrastructures. The aim is to advance research in the population sciences by reducing the barriers to collaboration often experienced in multi-site, "team science" research endeavors. CI-KNOW is intended to provide a model for how knowledge networks can be leveraged in public and population health to support discovery and innovation.