

# Expertise Identification using Email Communications

Christopher S. Campbell

Paul P. Maglio

Alex Cozzi

Byron Dom

IBM Almaden Research Center  
650 Harry Rd.

San Jose, CA, USA

ccampbel,pmaglio,cozzi@almaden.ibm.com, bdom@yahoo-inc.com

## ABSTRACT

A common method for finding information in an organization is to use social networks—ask people, following referrals until someone with the right information is found. Another way is to automatically mine documents to determine who knows what. Email documents seem particularly well suited to this task of “expertise location”, as people routinely communicate what they know. Moreover, because people explicitly direct email to one another, social networks are likely to be contained in the patterns of communication. Can these patterns be used to discover experts on particular topics? Is this approach better than mining message content alone? To find answers to these questions, two algorithms for determining expertise from email were compared: a content-based approach that takes account only of email text, and a graph-based ranking algorithm (HITS) that takes account both of text and communication patterns. An evaluation was done using email and explicit expertise ratings from two different organizations. The rankings given by each algorithm were compared to the explicit rankings with the precision and recall measures commonly used in information retrieval, as well as the  $d'$  measure commonly used in signal-detection theory. Results show that the graph-based algorithm performs better than the content-based algorithm at identifying experts in both cases, demonstrating that the graph-based algorithm effectively extracts more information than is found in content alone.

## Keywords

Expertise identification, knowledge management

## 1. INTRODUCTION

Knowledge in an organization is contained in the skill, experience, and expertise of its people. Yet the very problem of discovering who knows what is often challenging. Social networks—relationships among people in an organization—provide the basis for finding experts or for finding answers to questions. People ask others they know to find someone

with a particular skill, experience, or expertise, following pointers until an appropriate person is found. This *social networking* process involves asking a co-worker, manager, or senior employee, which leads to a chain of queries for either the answer to the question or the name of someone who could answer the question. Many times, the chain leads to people who act as contact brokers or mediators, providing pointers to people or groups that might help [3]. Social networks naturally form in a way that turns large groups into “small worlds” [12]. Nevertheless, there are huge costs to following pointers to experts, such as effort repeated by different people looking for the same answers, miscommunication that leads to the wrong expert, and time pressures that lead to taking the advice of not-so-expert experts who happen to be found quickly. How can we take advantage of social networks to find experts more effectively?

Email is a valuable source of expertise. It provides an easy-to-mine repository of communication between people in the social network, and it contains actual demonstrations of expertise (e.g., answering a question on some topic) as well as knowledge of expertise (e.g., decision of who should be asked the question). Both the content of email and the pattern of communication contain information about who knows what in an organization. In this paper, we take seriously communication patterns in email. We describe a system that identifies expertise from email, along with an evaluation of how well two different algorithms for mining expertise compare with human judgments.

### 1.1 How to Find Experts

Knowledge of expertise within organizations is often isolated to a specific team or to a single person. Our goal is to mine knowledge of expertise, store that information, and make it available to the larger organization so that it can be leveraged for problem resolution. In essence, we want to manage the knowledge that people have about who knows what in an organization to facilitate expertise mapping and improve organizational functioning. Because this knowledge of expertise is based on human judgments as coded in the email sent in an organization, our approach mines *perceived expertise* rather than true expertise. Nevertheless, perceived expertise will be valuable, as it reflects judgments of people who are knowledgeable about particular topics.

We define *expertise mapping* as the process of locating and identifying people who have knowledge of a particular topic. By *locating* an expert, we mean finding a person who seems knowledgeable on a topic. By *identifying* an expert, we mean deciding if that person is truly knowledgeable on

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CIKM'03, November 3–8, 2003, New Orleans, Louisiana, USA.  
Copyright 2003 ACM 1-58113-723-0/03/0011 ...\$5.00.

a topic. Because expertise mapping relies on informal social networks, it is not considered a formal business process; rather, it is a tacit process that underlies real organizational functioning. Such tacit processes are increasingly recognized as strategic elements of organizations, and so understanding them provides an opportunity to improve existing processes, develop information technology to enhance tacit processes, reduce risk in organizational change, and improve organizational adaptation.

Expertise mapping can be difficult. For instance, if one is not already knowledgeable on a particular topic, it might be hard to distinguish valuable experts from those who merely talk a lot about a particular topic [7]. For some topics, everyone has an opinion, which might make it even more difficult to identify the real expert. In addition, although working in an organization for years can lead to the development of an efficient expertise map [9], being new to an organization can limit access to informal social networks.

To automatically map expertise in an organization, we need a data source that captures how expertise is communicated in social networks. We believe the best record of this activity is email, which is a primary means of communication in many businesses. It contains precious information about the activities, interests and priorities of an individual or the organization, and because it flows continuously as a part of everyday operations, it naturally captures changing interests, projects, and goals. Email shows who communicates with whom and what those communications are about—effectively providing a window onto informal social networks. Email displays not only demonstrations of expertise but also knowledge of who knows what. The choice of who to send a question to is based on the sender’s knowledge of expertise, as well as knowledge of subject matter. It is reasonable to suppose, therefore, that appropriate analysis of an organization’s aggregate email can identify individuals with a high level of expertise in topics of interest.

## 1.2 Related Work

Schwartz and Wood [10] were first to analyze email flows to identify groups of individuals with common interests. To preserve privacy, they used only email flow and not email content, requiring initial identification of a “distinguished person” to seed the search for others with knowledge of a particular topic. The result was a list of related people with no intrinsic ranking order. The *ContactFinder* system [6] used text and addresses of messages on bulletin boards to find the right person to answer a certain question. *Xpertfinder* uses a pre-existing hierarchy of subject areas, characterized by word frequencies, to identify experts in specific areas by analyzing the word frequencies of email written by each individual. The *ExpertFinder* system uses number of self-published documents containing topic keyword(s) and frequency of person mentions near topic keyword(s) in non-self-published documents to produce expertise scores and ranks [8]. Commercial systems for expert identification include Autonomy’s *IDOL Server*<sup>1</sup>, which identifies employee’s expertise on the basis of the documents they access and submit on the intranet, and Tacit’s *KnowledgeMail*<sup>2</sup> and Lotus’s *Discovery Server*<sup>3</sup>, both of which build interest

profiles by scanning email and matching these to document taxonomies.

In contrast to all these, our approach uses text analysis and network analysis in an integrated system: the text of the messages is used to generate clusters of similar content, and the graph of message exchanges for each cluster is used to compute a ranked list of the individuals involved in the exchanges, sorted according to estimated expertise.

## 2. EMAIL EXPERTISE EXTRACTION

Our system for email expertise extraction (abbreviated  $e^3$ ) relies on a server that stores messages and performs analysis. Users contribute messages by carbon-copying the server’s email address. Users access expertise information through a web interface that enables both browsing by topic and searching by keyword.

In  $e^3$ , topics are generated through unsupervised clustering of message content, and keyword searching is enabled through standard information retrieval techniques run on user-supplied keywords and message content<sup>4</sup>. Given a set of messages generated by topic-clustering or keyword-searching,  $e^3$  identifies which senders and recipients are most knowledgeable by building a weighted directed graph representing the flow of information among the people involved.

More precisely,  $e^3$  embodies a three-step process: (1) collect all email related to a topic, (2) analyze email between every pair of people for whom there was relevant correspondence to build an “expertise graph,” and (3) analyze the expertise graph to obtain ratings for all senders and recipients. For the first step in  $e^3$ , only keyword retrieval and unsupervised clustering have been implemented. In principle, both supervised classification and unsupervised clustering can be done off-line, precomputed and ready when a user searches or browses for experts.

The second step in our process, building an expertise graph, is done by using the *from*’s and *to*’s to determine who is sending information to whom. The directional arrow points from email sender to receiver or receivers. The nodes of the graph correspond to people—or, more precisely, to unique email addresses.

Having constructed the expertise graph, the third step relies on a modified version of the HITS (Hyperlink-Induced Topic Search) graph-based ranking algorithm [5]. HITS associates two non-negative scores with each node in the graph: the *repute* score (“authority” in [5]), and the *resourcefulness* score (“hub score” in [5]). To provide an expertise rating for each person,  $e^3$  uses the *repute* score of each node. The HITS algorithm has been used successfully to support world-wide web searching by leveraging the link structure between websites [1]. Because people often take the time to add links to websites that are useful, interesting, or authoritative, the link structure reflects deliberate choices people make about what information is useful, interesting, and authoritative. The same logic can be applied to expertise location based on patterns of email communication. Because people’s knowledge of expertise determines who they send questions to, there is likely to be more email about certain topics sent to those who provide good information about that topic [11]. This is exactly the sort of knowledge we hope to mine from email using the HITS algorithm (see also [2]).

<sup>1</sup>www.autonomy.com

<sup>2</sup>www.tacit.com

<sup>3</sup>www.lotus.com/kd

<sup>4</sup>jakarta.apache.org/lucene/docs/index.html

### 3. EVALUATION

Our central question is whether the information contained in the pattern of email communication—taking account of who sends messages to whom—is useful for discovering expertise in an organization. Put differently, does our graph-based approach do a better job of finding experts than the simpler content-based approach? To answer this question, the HITS algorithm was compared against a query-term frequency approach (simple algorithm) on email corpora from two different organizations. Both algorithms were compared to expertise ratings explicitly solicited from the individuals in the two organizations for a set of topics automatically extracted from the messages. The comparisons were made using the recall and precision measures that are standard in the information-retrieval literature, as well as the  $d'$  measure that is standard in signal-detection theory [4]. These two different analyses were done to provide a complete measure of performance and ensure reliability.

#### 3.1 Method

Email was collected from a research organization (*OrgA*), and a software development organization (*OrgB*). Messages were collected from members of a single department or group within each organization. The groups were both fairly cohesive in that most members interacted with every other member, but the groups still maintained distinct sub-groups representing different projects, some of which were headed by different managers. For *OrgA*, 15 people (including 3 managers) selected email from their personal email databases. The resulting database contained 13,417 messages spanning about four years. For *OrgB*, 9 members of the department (2 managers) submitted their entire email databases without prescreening. The resulting database contained 15,928 messages spanning about two years.

Topics were obtained by (a) considering the type of work done in each group, and (b) reading through the messages for topics of interest. For both *OrgA* and *OrgB*, candidate experts were found for each topic through a keyword search. Candidates were then ranked by the number of email messages sent containing that keyword. The top 30 candidates in each category were used. The raters selected were those that had a lot of interaction with many different candidates.

The surveys included lists of candidates organized by topic. Raters were asked to rate each candidate for their level of expertise on a scale from 1 to 10, with 1 indicating no expertise and 10 indicating very high expertise. Raters were instructed to not rate people they did not know and to add names and ratings for any people not already listed. The instructions to raters defined expertise as the level of skill, formal training, and experience of a person for a topic, as well as the likelihood of seeking out a person to answer a question about a topic.

The HITS algorithm was compared against a simple content-based algorithm. The content-based algorithm counted the number of emails about a given topic sent by each person. An email was considered to be “about” a topic if it contained one or more keywords or search terms that described the topic. The resulting list of people was rank ordered according to who had sent the most email.

#### 3.2 Results

Survey results were averaged across ratings for each candidate and topic. Only candidates who received three or

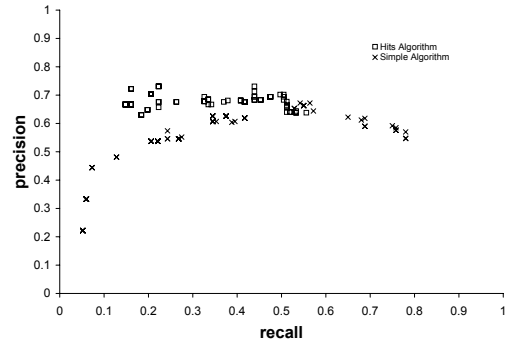


Figure 1: Precision and recall of each algorithm across a range of thresholds for *OrgB*.

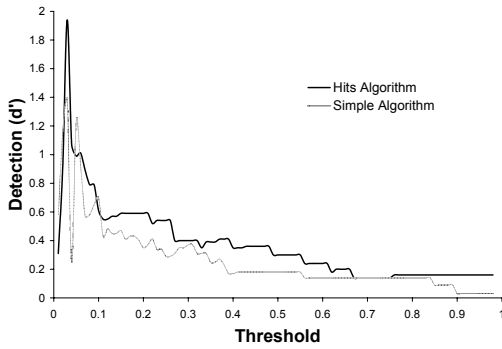
more ratings were retained. Candidates with average ratings above 7 were marked *expert* and below were marked *non-expert*. A range of thresholds was tested but all produced essentially the same results. The lists were then compared to lists produced by the algorithms at differing thresholds by matching pairs of people. Those who could not be matched were excluded from further analysis.

We first examined whether the expertise ratings provided by the raters were consistent by computing inter-rater reliability—the correlation between all pairs of raters on each topic. A lack of consistency in the ratings suggests (a) there may be no real knowledge about expertise in the workgroup for this topic or (b) this topic is poorly defined in the minds of the raters. For *OrgA*, reliability was high (i.e.,  $\geq 0.50$ ) for 9 of 11 topics. For *OrgB*, we also had high (i.e.,  $\geq 0.50$ ) inter-rater reliability for 9 of 13 topics.

Results were calculated by performing a precision and recall analysis at .01 increments from a threshold of .01 to .99, producing 98 data points for each algorithm. For *OrgA*, HITS maintained a higher precision than the simple algorithm but was somewhat limited in obtaining higher recall. The simple algorithm did well at recall but at the price of lower precision. Overall, the simple algorithm was more vulnerable to the precision/recall tradeoff than HITS. For *OrgB*, HITS again maintained a higher precision but did not seem to obtain the high peak levels of recall produced by the simple algorithm (see Figure 1). Precision also dropped as recall increased for the simple algorithm but did not change very much for HITS.

Signal detection theory’s  $d'$  is used here to determine how well each algorithm can detect expertise. There are four possible outcomes: correct identification (hit), incorrect identification (false alarm), incorrect rejection (miss), and correct rejection.  $d'$  combines all four measures into a single value between 0 (no detection) and about 3.0 (excellent detection).

The signal detection analysis across a range of thresholds for *OrgA* shows that HITS was better at detecting expertise than the simple algorithm at almost all thresholds (see Figure 2). For most of the thresholds, however, detection was quite low (i.e.,  $< 1.0$ ). Detection was best for both HITS and the simple algorithm at a threshold of about .05. Here, HITS showed good detection with a  $d'$  of about 2.0 and the simple algorithm also produced good detection with a  $d'$  of about 1.4. Similar results are seen for *OrgB*. HITS produced higher  $d'$ s across almost all thresholds with a peak at



**Figure 2: Detection ( $d'$ ) is higher for HITS across most thresholds for OrgA.**

	OrgA		OrgB	
	HITS	Simple	HITS	Simple
Percent Correct	38	44	33	31
Percent False Alarms	35	71	55	64
Precision	52	38	67	50
$d'$	.39	.28	.63	.44

**Table 1: Results for all topics and thresholds.**

a threshold of .15. The peak for the simple algorithm occurs at a threshold of .2. Compared to OrgA, the  $d'$ s for OrgB were generally higher indicating that both algorithms were better able to detect expertise for OrgB.

Overall algorithm performance across topics and threshold is shown in Table 1. For OrgA, the simple algorithm had a higher percent correct identifications, but made many more errors which is why the percent of false alarms was so high and precision was better for HITS than for simple. Overall, HITS achieved higher  $d'$ s due to many fewer errors. The results for OrgB are much the same; though, percent correct was about the same and the false alarms were not quite as bad for the simple algorithm as they were for OrgA. However, the simple algorithm still made many more errors, had lower precision, and therefore, had lower  $d'$ s.

### 3.3 Discussion

These results suggest that unlike the simple algorithm, HITS makes more specific or targeted predictions about who are the experts. This conservative criterion is given by the lower recall but higher precision of HITS. The simple algorithm seems to be less targeted, having a more liberal criterion for who is an expert, and therefore makes many more errors. The detection measure,  $d'$ , bears this out, showing that the high recall of the simple algorithm is not worth it given the false alarms. In other words, HITS is more sensitive to expertise than the simple algorithm.

Overall, detection rates and correct identifications are low but still comparable to results from other expertise location systems using larger databases containing a variety of documents. For the ExpertFinder system, for instance, average precision was 41% and recall was 29% [8]. Results from HITS was actually better with precision for OrgA at 52% and OrgB 67%. Recall for HITS was 38% and 33% respectively. This shows that email data does contain useful

information about expertise. However, the marginal expertise identification rates indicate that (a) better algorithms are needed to capture more information from email or (b) email just has a limited amount of expertise information to capture.

## 4. CONCLUSION

We have shown that for a relatively small sample of email gathered from each of two different organizations, a graph-based algorithm that takes account of communication patterns does a better job of identifying who knows most about specific topics than a content-based algorithm when compared with explicit human judgments of expertise. Thus, the flow of email within an organization can serve as a source of expertise information that can be mined effectively. It seems that social networks are in fact buried in the patterns of email communication and can be recovered and used much as social networks are used by people in organizations: to discover answers to questions by following pointers to people with specific knowledge, skill, or experience.

## 5. REFERENCES

- [1] S. Chakrabarti, B. E. Dom, D. Gibson, J. Kleinberg, R. Kumar, P. Raghavan, S. Rajagopalan, and A. Tomkins. Hypersearching the web. *Scientific American*, June 1999.
- [2] B. Dom, I. Eiron, A. Cozzi, and Y. Zhang. Graph-based ranking algorithms for e-mail expertise analysis. In *8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery*, 2003.
- [3] K. Ehrlich and D. Cash. Turning information into knowledge. In *Proceedings of Digital Libraries*, pages 119–125, 1994.
- [4] D. M. Green and J. A. Swets. *Signal detection theory and psychophysics*. Wiley, New York, 1966.
- [5] J. M. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46(5):604–632, 1999.
- [6] B. Krulwich and C. Burkey. The ContactFinder agent: Answering bulletin board questions with referrals. In *AAAI-96*, 1996.
- [7] G. E. LittlePage and A. L. Mueller. Recognition and utilization of expertise in problem-solving groups: Expert characteristics and behavior. *Group Dynamics: Theory, Research, and Practice*, 1:324–328, 1997.
- [8] D. Mattox, M. Maybury, and D. Morey. Enterprise expert and knowledge discovery. Technical report, 1999.
- [9] D. W. McDonald and M. S. Ackerman. Just talk to me: A field study of expertise location. In *Computer Supported Cooperative Work*, pages 315–324, 1998.
- [10] M. F. Schwartz and D. C. M. Wood. Discovering shared interests using graph analysis. *Communications of the ACM*, 36(8):78–89, 1993.
- [11] E. W. Stein. Social and individual characteristics of human experts. *International Journal of Expert Systems: Research and Applications*, 8(2):121–143, 1995.
- [12] S. Wasserman and K. Faust. *Social Network Analysis: Methods and Applications*. Cambridge University Press, Cambridge, 1994.