Graph-Based Ranking Algorithms for E-mail Expertise Analysis

Byron Dom[*]
Yahoo! Inc.
701 First Ave.
Sunnyvale, CA, USA
bdom@yahoo-inc.com

Iris Eiron
IBM Almaden Research Center
650 Harry Rd.
San Jose, CA, USA
iriss@us.ibm.com

Alex Cozzi
IBM Almaden Research Center
650 Harry Rd.
San Jose, CA, USA
cozzi@almaden.ibm.com

Yi Zhang[†]
Language Technology Institute
School of Computer Science
Carnegie Mellon University
yiz@cs.cmu.edu

ABSTRACT

In this paper we study graph–based ranking measures for the purpose of using them to rank email correspondents according to their degree of expertise on subjects of interest. While this complete expertise analysis consists of several steps, in this paper we focus on the analysis of digraphs whose nodes correspond to correspondents (people), whose edges correspond to the existence of email correspondence between the people corresponding to the nodes they connect and whose edge directions point from the member of the pair whose relative expertise has been estimated to be higher. We perform our analysis on both synthetic and real data and we introduce a new error measure for comparing ranked lists.

General Terms

Algorithms, Experimentation, Measurement

Keywords

Expert finding, Social network analysis, Digraph node ranking, Ordered list distance

1. INTRODUCTION

A problem faced on occasion by almost everyone is that of locating some desired information or source of knowledge. The latter is desirable when one isn't even sure what to ask for and can usually be satisfied by finding an expert in the topic of interest. This problem of finding an expert is what we address in this paper.

In large enterprises such as companies and government agencies a source of information that can be utilized in this search for experts is the aggregate email of the organization. Email is used today as a major means of communication in business. As such it contains precious information about the activities, interests and priorities of an individual or the organization. The email collection also includes both notes requesting information and notes providing it. It is reasonable to hope, therefore, that appropriate analysis of an organization's aggregate email can identify individuals with a high level of expertise in the topics of interest to the organization as a whole or some sufficiently large fraction of it.

In [3] an approach to performing this expertise analysis of email collections and a system embodying that approach are described. In this paper we concentrate on the third part of the following three–step process. This process is applied to the subset (of the complete collection) of the emails discussing a certain topic.

1. Quantitatively estimate the relative expertise between two people based on all the on–topic emails between them.
2. These pairwise relative–expertise assessments can be seen to implicitly form a weighted directed graph (*digraph*) in which the nodes correspond to people (senders and receivers of emails in the collection). The existence of an edge between nodes i and j indicates that there was at least one email between the associated people. The direction of the edge is from the person of greater expertise to the person of lesser expertise and the edge weight is the quantitive measure of relative expertise.
3. The described digraph is analyzed using a graph–based ranking algorithm whose output is a ranked list of all the nodes (people) according to their absolute expertise level.

The goal of the work described here is to compare various ranking algorithms. Our method for comparison is based on

[*]This research was done while in IBM.
[†]This research was done while on internship in IBM.

applying them to graphs where the ground truth ranking is known and then assessing how well the algorithms output agrees with the true ranking. Both simulations and real–data experiments are presented. We study a simulation of a fully connected "perfect" digraph in which every node pair is connected by a link with the correct direction and weight is formed. This graph is then systematically degraded by removing and/or reversing links. Additionally we analyze graphs based on real data, in which edges account for email exchange. These graphs contain only a small fraction of the possible edges. Furthermore the use of text classification to determine the direction of the edges may further degrade the quality of the graph by assigning incorrect directions and/or weights to the edges.

2. RELATED WORK

Motivated by the prevalence of the WWW, the task of expert finding has lately received growing attention from the research community [6; 4; 8]. For a review of work in this area, both research studies and commercial systems, we refer the reader to [15]. In contrast to our approach, the majority of the systems studies so far focus on connecting people to experts who posses some knowledge on a given topic. Our system in comparison is concerned with ranking people according to their expertise level rather than separating experts from non experts.

A system that attempts to provide ranking of experts by their expertise level is described in [9]. This system is an expert–finder system that exploits technical papers, presentations, resumes, home pages etc. on MITRE's corporate intranet to enable the location of relevant experts. The ranking uses a simple scheme based on the number of mentions of a term/phrase.

Another factor differentiating our system from many other expertise–location systems is the use of email for identifying experts. One of the few systems that are based on email is [13]. However this system does not consider the content of email messages but only flow pattern between the correspondence. Another system that uses email is presented in [14]. This system uses a pre-existing hierarchy of subject areas, characterized by word frequencies, to identify experts in specific areas by analyzing the word frequencies of the email messages written by each individual. However, like many of the systems mentioned above, this system does not deal with the question of the ranking of the identified experts.

3. RANKING MEASURES

As we stated in the introduction, our hope is to use both the texts of email messages and their distribution patterns to determine who the experts are. Because we represent the organization whose email we study as a graph whose nodes correspond to people and whose edges (links) correspond to the existence of email correspondence between the people whose corresponding nodes they connect, we think of and refer to this analysis as "link analysis".

The outcome of our analysis of the email between two people is the determination of which of the two knows more about the topic and a relative–expertise score, which characterizes the magnitude of that assessed difference. We can make an analogy between this comparison and a competition (match, game, etc.) between two sports teams. In this analogy the relative–expertise score corresponds to the difference in score between the two teams. We can further extend this analogy to a correspondence between finding experts and finding a champion by ranking teams, especially in cases where there haven't been competitions between every pair of teams. This problem has been studied extensively (see, for example [1; 2]). We will make use of this analogy in the following descriptions of the ranking approaches we evaluate.

3.1 Affinity (a.k.a Score)

Following our team–ranking analogy, perhaps the simplest way of ranking teams is by the number of wins or (closely related) the total point differential between the team and all of its opponents. We refer to this measure as "affinity" and it was referred to as "score" in [5]. This measure would be fine if every team played every other team or if we had relative–expertise assessments between every pair of people, but when that is not the case, it tends not to work well because it rates wins over all teams equally regardless of the records of those teams and who their opponents were.

3.2 Successor

This is related to the measures described in [5]. Think of directed edges pointing from winners to losers or from greater to lesser expertise. If all these edge directions are correct, all the people "downstream" (reachable by a directed path) from a given node are of lesser expertise. This one measure of expertise is simply the count of such people (nodes).

3.3 PageRank

One of the more popular team–ranking techniques is described in [5]. In this approach the obtained ranking is the principal eigenvector (corresponding to the largest eigenvalue) of the adjacency matrix of the digraph in which edges correspond to matches between teams and the direction of the edge is determined by who won. This approach turns out to be virtually identical to the well known PageRank[12] algorithm for ranking web pages, the only difference being that in PageRank low–weight edges are added between all nodes in both directions. This weight is an adjustable parameter. We refer to this step of adding low–weight links to all nodes as "smoothing", which we also perform.

3.4 Positional Power Function

One of the ranking measures we chose to evaluate is taken from [5]. It is closely related to what they call the "Positional Power Function" (PPF). For PPF the ranks $\{r_i | i = 1, 2, \ldots, n\}$ satisfy the following system of equations.

$$r_i = \sum_{j \in S_i} \frac{1}{n}(r_j + 1), \qquad (1)$$

where S_i denotes the set of *successor* nodes of node i.

A generalization of (1), which we evaluated as part of our experimental characterization is:

$$r_i = \sum_{j \in S_i} [\alpha + (1 - \alpha) r_j], \qquad (2)$$

We solved this system by the usual power–iteration scheme. The corresponding results in our experimental section are labelled "power".

3.5 HITS (authority)

The HITS[7] algorithm was devised with the thought of ranking web pages according to their degree of "authority". It is designed with the structure of the web in mind. In the associated digraph the directed edges correspond to hyperlinks. The HITS algorithm produces two scores/rankings – the "hub" score and the "authority" score. We present only the latter because it gave consistently better results than the hub score as expected. In terms of the digraph adjacency matrix A the authority score is given by the principal eigenvector of the matrix $A^T A$, where A^T represents the *transpose* of A.

4. ACCURACY EVALUATION

In this section we describe the measures we use to characterize the accuracy of ranking by characterizing the degree of agreement/disagreement between the computed ranking r and the correct (ground–truth) ranking ρ. In the analysis of synthetic data (described in Section 5) we used a novel error–distance measure, designed to exhibit qualitative behavior desirable for an expertise ranking accuracy measure. We we refer to this measure as "rnkerr" and describe it in the following section. The real–data analysis (described in Section 6), presents results using both the "rnkerr" measure and recall graphs.

4.1 Rank–Error Distance

This is a measure of the following form:

$$\epsilon(\rho, r) \quad = \quad \sum_{i=1}^{N} \eta(m_i)\, \zeta(\delta_i) \tag{3}$$

where

$$m_i \quad \triangleq \quad \min(\rho_i, r_i) \qquad \text{and} \qquad \delta_i \quad \triangleq \quad |\rho_i - r_i|.$$

The general form of (3) embodies two notions about the desirable qualitative behavior of such a measure. One of these ideas is that the importance of an error (difference in the rank of the i^{th} object) increases with the highest rank the associated object received on either of the lists $(\min(\rho_i, r_i))$. If neither rank is high, the associated error (difference) should contribute very little to the overall score. This notion is captured in the function $\eta(\cdot)$, which decays monotonically with its argument, asymptotically approaching zero.

The second idea is that the magnitude of the error associated with a difference δ in rank should saturate at (asymptotically approach) some value as δ increases. The notion here is that beyond some value, all differences should carry approximately the same weight. This behavior is embodied in the function $\zeta(\delta)$.

In experiments we used the following specific functional forms for η and ζ each characterized by single adjustable parameters – λ and β respectively.

$$\eta(m; \lambda) \quad = \quad \frac{\exp(-\lambda m)}{\sum_{m=1}^{N} \exp(-\lambda m)}$$

and

$$\zeta(\delta; \beta) \quad = \quad \frac{1}{\beta}[1 - \exp(-\beta\delta)]$$

Where the parameter value $\lambda = 0.25$ and $\beta = 0.1$ were used in our experiments.

4.2 Recall

In our *recall*–based analysis we characterize the comparison between two rankings by a *recall* curve (see Figure 2 for example). In the figures depicting this the "query size" (let q represent this) refers to the top q ranked nodes (people) in the computed ranking r and the "recall" is the fraction of the top q nodes of the ground–truth ranking ρ that are contained in the top q members of r.

5. RESULTS FOR SYNTHETIC DATA

As part of our study of the behavior of the various ranking measures we ran them on synthetically generated expertise graphs. The general approach was to initially generate "perfect" graphs and then randomly degrade those graphs by removing and reversing edges. By "perfect" graphs we mean graphs for which there is an edge (link) between every pair of nodes and the direction of these edges correctly reflect the relative expertise of the nodes they connect. In what follows we present results for three such collections of experiments.

1. The first of these corresponds to only removing edges from a perfect 21-node graph.
2. The second corresponds to only reversing edges from a perfect 21-node graph.
3. The third corresponds to both removing and reversing edges from a 15-node perfect graph. That is, a number of edges is first removed and then a number of those remaining are reversed.

In the work presented here all edges have weight 1.

Figures 1(a) and 1(b)) are plots of the error distance ("rnkerr" $= \epsilon(\rho, r)$) described in Section 4 versus the number of nodes either removed or reversed. The individual points are average values from several trials as described below. There are three tables in this section, each table representing an experiments collection. Each table contains three columns for each measure. The following is a description of these columns.

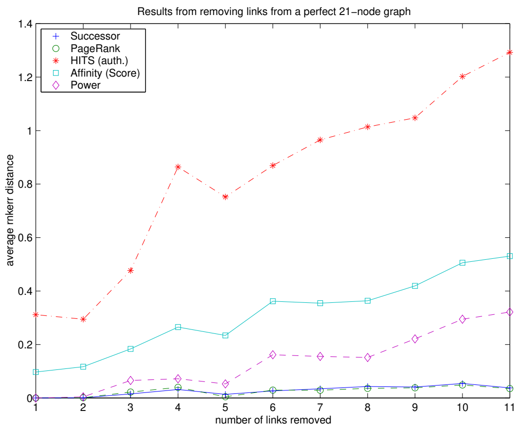
"avg. dist.": This is the average $\epsilon(\rho, r)$ between the ranking produced by the measure and the "ground truth" (i.e. the correct ranking). This average is over all trials in the particular experiment. Smaller values are better; a perfect score is zero.

"no. best": This is the number of trials for which the associated measure produced the best ranking as measured by $\epsilon(\rho, r)$. Larger numbers are better; a perfect score is equal to the total number of trials. The worst possible score is zero.

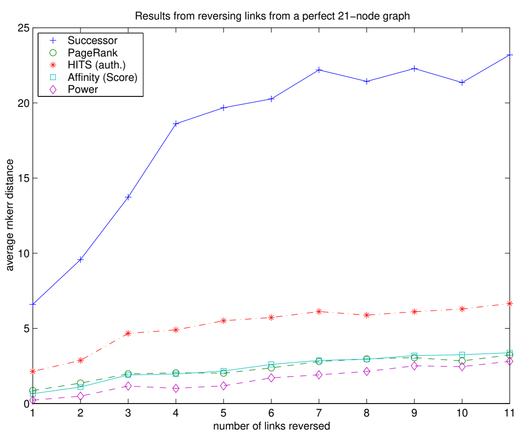
"no. worst": This is the number of trials for which the associated measure produced the worst ranking as measured by $\epsilon(\rho, r)$. Smaller numbers are better; a perfect score is zero. The worst possible score is equal to the total number of trials.

5.1 Removal of Edges from Perfect Graphs

The results presented here are for a set of experiments on 21-node graphs. It is assumed that the associated 21 people have expertise "scores" of $\{1, 2, 3, \ldots, 20, 21\}$. A directed link with weight 1 is placed between every pair of nodes going from higher to lower expertise score. This is the initial "perfect" graph, which has $|E| = \binom{21}{2} = 210$ edges. The number k_{rem} of edges removed from these graphs is varied from 1 through $\lceil n/2 \rceil$, which is 11 in this case. The results of these experiments are plotted in Figure 1(a) and tabulated in Table 1.



(a) Removing links.



(b) Reversing links.

Figure 1: Results from randomly removing/reversing links in a perfect 21-node.

measure	avg. dist.	no. best	no. worst
Successor	0.0342	321	0
PageRank	0.0317	332	0
HITS	0.9711	0	196
Affinity	0.3753	0	1
Power	0.1825	142	0

Table 1: Results of randomly removing links from a perfect 21-node graph.

5.2 Reversal of Edges in Perfect Graphs

In this set of experiments we reverse rather than remove edges in perfect graphs. The number of edges reversed and the scheme for determining that number and selecting the edges is exactly the same as that used in the removal experiments described in Section 5.1. The results of these experiments are plotted in Figure 1(b) and tabulated in Table 2.

measure	avg. dist.	no. best	no. worst
Successor	3.5258	1	369
PageRank	0.4594	62	0
HITS	0.9983	7	0
Affinity	0.4804	27	0
Power	0.3468	256	0

Table 2: Results of randomly reversing links in a perfect 21-node graph.

5.3 Combined Removal and Reversal from Perfect Graphs

The results presented here are for a set of experiments on 15-node graphs. It is assumed that the associated 15 people have expertise "scores" of $\{1, 2, 3, \ldots, 14, 15\}$. A directed link with weight 1 is placed between every pair of nodes going from higher to lower expertise score. This is the initial "perfect" graph, which has $|E| = \binom{15}{2} = 105$ edges. This graph is then degraded in various ways described as follows.

- number of edges removed k_{rem}: Varied from 1 through $\lceil n/2 \rceil$, which is 8 in this case.
- number of random sets of k_{rem} (remaining) edges removed for each value of k_{rem}:

$$\nu(k_{\text{rem}}) = \left\lfloor \log_2 \binom{|E|}{k_{\text{rem}}} \right\rfloor$$

  Let $G_{k_{\text{rem}}}$ be a random variable whose values are graphs formed by deleting random sets of k_{rem} edges from G.
- number of edges reversed k_{rev}: This is varied from 1 through k_{rem} for each $G_{k_{\text{rem}}}$.
- number of random sets of k_{rev} edges reversed for each value of the pair $(G_{k_{\text{rem}}}, k_{\text{rev}})$:

$$\nu(k_{\text{rem}}, k_{\text{rev}}) = \left\lfloor \log_2 \binom{|E| - k_{\text{rem}}}{k_{\text{rev}}} \right\rfloor$$

- total no. graphs analyzed

$$\sum_{k_{\text{rem}}=1}^{\lceil n/2 \rceil} \nu(k_{\text{rem}}) \sum_{k_{\text{rev}}=1}^{k_{\text{rem}}} \nu(k_{\text{rem}}, k_{\text{rev}}) = 19307.$$

The results of this process are tabulated in Table 3.

measure	avg. dist.	no. best	no. worst
Successor	2.2315	457	15109
PageRank	0.6303	6372	14
HITS	1.1722	576	1890
Affinity	0.7225	1324	7
Power	0.5927	9012	16

Table 3: Results of randomly removing and reversing links from a perfect 15-node graph.

6. RESULTS FOR REAL DATA

For the second part of our experiments we collected a data set of internal email communication contributed by individuals in our organization. Fifteen people from a single workgroup contributed email messages from their personal email boxes, containing a total of 13417 messages. From this set we selected, using keyword search, messages related to ten topics. Five of the ten topics were generic, such as LaTeX, Java, and Perl, while the other five were more specific to the type of work done in the workgroup. The full list of people involved in correspondence on each topic was extracted. Human evaluators were asked to rate each person for his/her level of expertise per topic on a scale from 1 to 10. These ratings were then averaged across the evaluators. We used the resulting ranked list for each of the topic as the ground truth for our experiments.

Recall that in the graph representation of an email corpus nodes are associated with the people involved in the correspondence and a directed edge from node i to node j indicates that i is less of an expert on the subject than j is. With an email collection at hand this relative expertise information may be based on the analysis of the email messages exchanged between the parties. In the case where no correspondence exists between a pair of people, the relative expertise level cannot be directly determined and the corresponding edge will not exist in the graph. As a result, graphs based on real email correspondence are much sparser than the ideal fully connected graphs. Table 4 reveals that the number of edges in graphs representing the ten topics is on average only 20% of the number of edges in the equivalent fully connected graphs.

topic	#nodes	#edges	$\frac{\#\text{edges}}{\binom{\#\text{nodes}}{2}}$
java	37	84	12
latex	8	6	21
matlab	12	11	16
perl	12	11	16
xml	26	44	13
cognitive	19	21	12
eye	12	13	19
pong	15	13	12
wbi	28	98	25
wf	13	12	15
avg.	18	31	20

Table 4: Characteristics of real data base graphs.

While in the synthetic experiments (Section 5) we were randomly reversing edges, with an email corpus at hand we assume the direction of an edge is determined by a text classifier which analyzes the content of the messages exchanged between a pair of people. Like any analysis of this kind we cannot expect 100% accuracy from this process. Therefore in the resulting graphs some of the edges may be reversed. We experimented with four different analysis methods. For every pair of people i,j between which correspondence exists the various analysis methods will act as follows:

Ideal Classifier If the expertise level of i is higher than that of j and edge (i, j) is created with weight 1.

Dummy Classifier Creates two parallel edges in opposite directions between i and j, each with weight 1/2.

Random Classifier Flips an unbiased coin. If "tail" — creates an edge i, j with weight 1. Otherwise, creates an edge j, i with weight 1

Maximum Entropy We trained a Maximum Entropy (ME) [11] classifier from the Rainbow package [10]. We ran the classifier on the set of messages sent from i to j. We then create an edge (i, j) with weight p and an edge (j, i) where p is the probability the classifier gave to the event that i has higher expertise than j. If communication exists in both directions we run the classifier separately for each direction. We then assign a weight of $(p + 1 - q)/2$ to the edge (i, j) and $(q+1-p)/2$ to the edge (j, i) where p is as before and q is the output of the classifier when run on message on the opposite direction. In our experiments we used 9 out of the 10 queries to train the classifier and tested on the remaining topic.

	PageRank	Hits	Power	Affinity
Ideal	0.32	0.44	0.42	0.42
Dummy	0.53	0.53	0.53	0.56
Random	0.58	0.58	0.54	0.55
ME	0.56	0.55	0.55	0.57

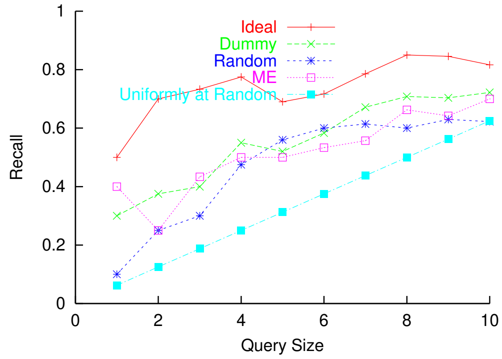
Table 5: Results on on real data based graphs.

A summary of the results of running the four ranking algorithms (excluding Sucessor) perviously mentioned on our email data is brought in table 5. The results are further detailed in Figures 2, 3 and 4.

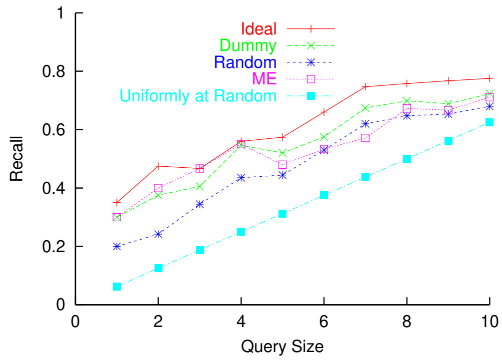
7. DISCUSSION OF RESULTS

In the synthetic results of Section 5 we examined three modes of degradation starting from perfect graphs. In the removal–only results the best (nearly perfect) results were obtained using PageRank and Successor, whereas HITS was significantly worse than the others. In the reversal–only results, on the other hand, the best results were obtained with Power, though the results obtained with PageRank and Affinity were nearly as good. The Successor results were much worse than the others and the HITS results were somewhat worse. Similar results were obtained in the removal–plus–reversal experiments. That is, the best results were obtained with Power, with PageRank and Affinity finishing a relatively close second and third respectively. The HITS algorithm was a more distant fourth and Successor was once again significantly worse than the others. As can be expected, measured accuracies obtained for the removal–only experiment were significantly better than those obtained for reversal–only and removal–plus–reversal results.

In the real data results of Section 6 we examined four modes of degradation starting from sparse graphs. The results obtained with "ideal" classifier support the synthetic removal–only results. PageRank performed significantly better whereas HITS was slightly worse than the other methods. Among the Dummy, Random, and ME classifiers very small differences in overall performance are observed. It should be noted, however, that Power performed slightly better than all the other methods. This is consistent with the results gathered for the removal–plus–reversal and removal–only

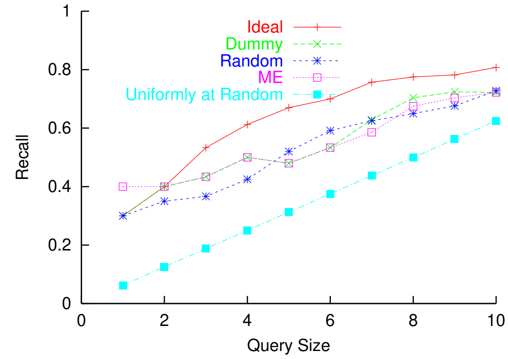


(a) PageRank.

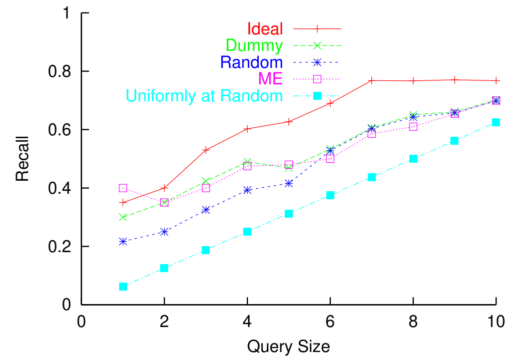


(b) HITS.

Figure 2: Impact of Classification.



(a) Power.



(b) Affinity.

Figure 3: Impact of Classification – contd.

synthetic experiments. Note that the difference in performance between best performing algorithm and worst performing algorithm on synthetic data was much more significant than the difference between those algorithms obtained on real data. These experiments also confirm that using the ideal (i.e. correct) classification to generate the edge directions in the graph yielded significantly better results than those obtained when using the other classification methods. As a calibration point, the graphs presenting real data results (Figures 2,3 and 4) include curves corresponding to picking a ranking uniformly at random. These are the straight–line curves appearing lowest in the four subfigures. Not surprisingly, all examined ranking algorithms perform significantly better than this random–ordering.

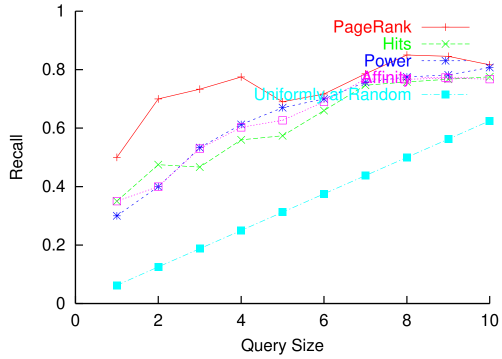
Unfortunately, we discovered that the ME classifier yielded ranking performance that is comparable if not worse than that of the Dummy and Random classifiers. This essentially means that this classifier failed to provide the ranking algorithm with any useful information from the content of the messages. This demonstrates the difficulties of the classification task. Two main factors contribute to the difficulty of this task: (1) Both messages authored by experts and those authored by lesser experts, discuss the same topic, and therefore use similar vocabulary. (2) Email messages vary in length and tend to be shorter than other documents – as short as a single word (e.g. "yes" or "no") – with much of the original context missing. Therefore, not every message contains sufficient information by itself to identify the relative expertise level of the parties involved.

7.1 Conclusion

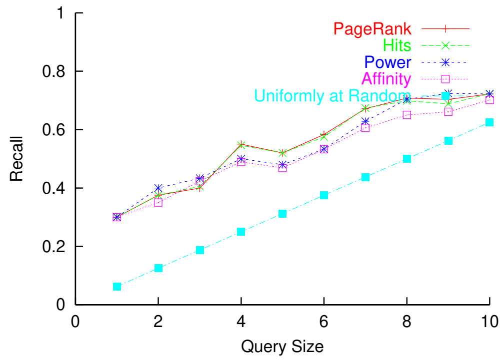
The Successor algorithm, while dealing well with missing edges, is clearly not robust to errors in edge direction. For this reason it was ommited from the real data analysis. The HITS algorithm, while giving results of reasonable accuracy, gave noticeably worse results than the others (except Successor). The inferior results obtained with HITS may be attributable to the fact that it was conceived for the purpose of ranking web pages and is based on the notion of the existence of *hubs* and *authorities* which mutually reinforce each other during the ranking process. While one can hypothesize the existence of certain email correspondents that play a role analogous to that of hubs, it seems unlikely that many exist. A good hub would be someone who corresponds with a large number of experts on a certain topic. It is our feeling that most users of email will tend to have one (or at most a few) person they use as an expert consultant on a given topic.

The tournament–like model implicit in Power and PageRank seems more appropriate to the problem at hand. PageRank performs noticeably better than all other algorithms in the case edges are missing but the ones present in the graph point in the right direction. When noise is introduced in the edge directions Power becomes more advantageous. The final conclusion regarding which of the two is best depends on the accuracy of the available classifier. Affinity also gave reasonable results though somewhat inferior to Power and PageRank.

Examining the performance of the various algorithms operating on a graph constructed with the Dummy classifier



(a) Ideal Classifier.



(b) Dummy Classifier.

Figure 4: Best ranking algorithm.

(Figure 4(b)) reveals that all the examined algorithms perform significantly better than the "random ranking" curve. This result is encouraging because it implies that by only using the patten of the email exchange, independently of the link structure, some significant information was extracted. With a query size of only four we receive 50% recall. If the accuracy of the classifier is to be improved we can expect up to 70% recall with a query size of only two.

8. FUTURE WORK

In the future we may investigate other ranking algorithms and variations of those studied here. In addition we plan more extensive experimental investigations. In the synthetic–data investigations we will use graphs that are much sparser and therefore closer to those encountered in real data. We also plan to include weighted edges (i.e. not only 0/1) in our analysis. In the real–data experiments we will obtain more representative email collections.

Acknowledgments

We would like to thank Christopher Campbell for collecting subjective ranking data, and Nadav Eiron for his continuous support.

9. REFERENCES

[1] Bibliography on college football ranking systems. `http://www.cae.wisc.edu/~dwilson/rsfc/rate/biblio.html`.

[2] Overview of ranking literature from tbeck. `http://www.phys.utk.edu/sorensen/ranking/Data%20Import/overview_of_rank%ing_literature_f.htm`.

[3] A. Cozzi, B. Dom, I. Eiron, Y. Zhang, C. Campbell, and P. Maglio. Identifying experts by analyzing e-mail. submitted, 2003.

[4] L. N. Foner. Entertaining agents: A sociological case study. In *Proceedings of the First International Conference on Autonomous Agents (Agents'97)*, pages 122–129, 1997.

[5] P. J.-J. Herings, G. van der Laan, and D. Talman. Measuring the power of nodes in digraphs. Technical Report 01-096/1, Tinbergen Institute, October 2001.

[6] H. Kautz, B. Selman, and M. Shah. Referral Web: Combining social networks and collaborative filtering. *Communications of the ACM*, 40(3):63–65, 1997.

[7] J. M. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46(5):604–632, 1999.

[8] B. Krulwich and C. Burkey. The ContactFinder agent: Answering bulletin board questions with referrals. In *AAAI-96*, 1996.

[9] D. Mattox, M. Maybury, and D. Morey. Enterprise expert and knowledge discovery. In *Proceedings of the 8th International Conference on Human-Computer Interaction (HCI International'99)*, pages 303–307, 1999.

[10] A. K. McCallum. Bow: A toolkit for statistical language modeling, text retrieval, classification and clustering. http://www.cs.cmu.edu/ mccallum/bow, 1996.

[11] K. Nigam, J. Lafferty, and A. McCallum. Using maximum entropy for text classification, 1999.

[12] L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford Digital Library Technologies Project, 1998.

[13] M. F. Schwartz and D. C. M. Wood. Discovering shared interests using graph analysis. *Communications of the ACM*, 36(8):78–89, 1993.

[14] W. Sihn and F. Heeren. Xpertfinder - expert finding within specified subject areas through analysis of e-mail communication. In *EUROMEDIA 2001 : Sixth Annual Scientific Conference on Web Technology, New Media, Communications and Telematics Theory, Methods, Tools and Applications*, 2001.

[15] D. Yimam and A. Kobsa. Expert finding systems for organizations: Problem and domain analysis and the demoir approach. *"Journal of Organizational Computing and Electronic Commerce"*, 13(1):1–24, 2003.