Combining Content and Quality Indicators in Ranking Ambiguous Query Results On Flickr

Hao Zhong **Ying Zhou**

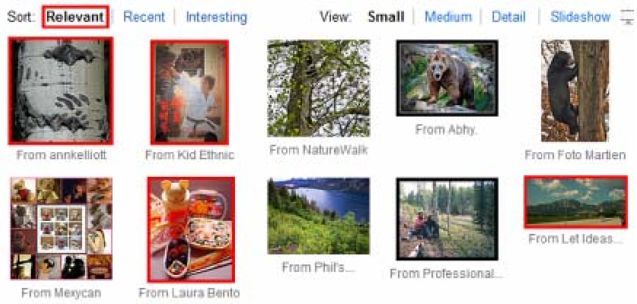
School of Information Technologies,
The University of Sydney
NSW 2006
Email: {hzho2468,ying.zhou}@sydney.edu.au

Abstract

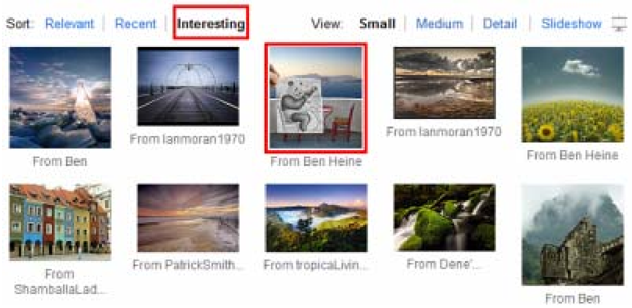
When a user submits a text based query to content sharing sites like Flickr, a list of ranked result with limited refinement options are normally provided. Typical options would allow a user to rank the results in different ways such as relevancy, time, or quality respectively. The downside of such approach is that relevant results might not be of high quality while high quality results are often irrelevant. Possible ambiguity of query terms makes it even more difficult to get high quality and relevant results. In this paper we apply link based analysis to combine content and quality indicators for ranking query results in Flickr. Experiment show that our approach are able to identify high quality photos that match a query user's intention and put them at the top of the list. The precision is better than original quality based ranking and possible query expansion results. Our approach relies on a set of seed users representing content and quality preference. We prove experimentally that the ranking is not sensitive to seed user selection, which makes it very practical.
Keywords: Social Networks, Link Analysis, Ambiguous Query, Flickr

1 Introduction

With the advance of various new web technologies, sharing content with friends and other people becomes a major part of the online social activity for many Internet users. The users are willing to and are able to publish all sorts of resources including many non-textual resources on social network based content sharing sites. Most non-textual resources such as images or videos are described by textual information such as titles or tags. Such textual information are often short and noisy by nature. Yet, they are the basis for text based queries for such resources, which poses lots of new challenges for the design of the query engine and the ranking algorithms. One of the most explored area is the disambiguation of query terms. Ambiguity of query terms is a very common problem in non-textual resources retrieval because the content range is huge while the textual information involved is limited and the vocabulary is uncontrolled. Terms that are not considered ambiguous otherwise may refers to several underlying concepts in those systems. Figure 1.(a) shows an example of query result for term "bear". The result is sorted by "relevance". In addition to animal bear we can see certain "surprising" bear related content such as bear claws, bear karte, bear-theme meal and a mountain named "bear peak". It also includes several unrelated photos.



(a) "Relevance" ranking on Flickr



(b) "Interestingness" ranking on Flickr

Figure 1: Flickr "bear" query results

The common disambiguation approach is to apply text mining and clustering technique to organize the results into several groups (Cai et al. 2004, Lee et al. 2009, Sadikov et al. 2010). This approach relies heavily on the relationship between individual terms as expressed in the textual description or as from external knowledge base such as wordNet. It can effectively create content based groups within the query results. However, it cannot differentiate the quality of resources if they have similar textual description.

In traditional search engine design, the quality factor is controlled by hyper links coming in and going out of the web pages (Kleinberg 1999, Page et al. 1999). The link based approach cannot be applied directly in ranking user-published content as most content does not have links among each other. However, most sites hosting user-published content have mechanisms for users to express their judgments on a particular contents. For instance, youtube users can indicate if they "like" or "dislike" a particular video, flickr users can specify a photo as "favorite". Such user generated judgment, together with other factors can be used to provide quality based ranking. For instance, flickr provides rank by *interestedness* option. Yet such quality based rank largely sacrifices content relevancy. Figure 1.(b) shows the same "bear" query ranked by *interestedness*. Only one out of the top ten photos is related with the query term "bear".

Quality is a relatively subjective measure. One user's favorite might not be another user's favorite if they have very different tastes. In recommendation system, a user's taste is represented as a profile established by analyzing this user's previous judgments. A group of users with similar profiles can be used to recommend items to each other. Apart from the implicit profile, explicit social network also serves as an good indicator of a user's special tastes.

In this paper we apply link analysis on social network formed by users to re-rank text based photo query result in flickr. The link based analysis combines the content and quality indicators to provide relevant and high quality results matching a query user's preference. Our work has the following contributions:

1. We conduct a study on social network formed by contact relations in Flickr and its impact on the quality indicators of photos. We observe that contact relation can reflect users' shared preference in terms of content and and quality.

2. We apply link analysis combining Personalized PageRank(PPR) and Weighted HITS to re-rank query result. Experiment shows that our approach achieves higher precision compared with original results and possible query expansion results. In particular, we show that using only a few seed users for a general category is sufficient enough to generate accurate results for all topics in that category and the result is not sensitive to seed user selection.

In section 2, we briefly describe flickr social network and quality factors involved. Section 3 describe the link analysis approach and its application in reranking flickr result. In section 4, we describe the experiments and analyze the results. We describe some related work in section 5. Section 6 concludes the paper.

2 Social network and a user's preference on content and quality in flickr

Flickr is an online photo sharing website that allows users to upload, share, organize and view photos. Thousands of photos are uploaded every minute. Flickr users can form various relationships. A flickr user can assign other users as *friends* and give them permission to view photos that are not public. Friends may have social ties off line. A flickr user can also assign other users as *contacts*. The *contact* relation creates a shortcut for one user to easily view all contacts' public photos. Moreover, a user will be notified about any new photos published by her contacts. The friends network is relatively small and private, it is more of an access level control mechanism. The contact network is a public and forms a large social network among flickr users. Assigning contact is a way of showing interest on another user's photo. *Contact* is a one way relation. User A assigns user B as contact does not mean user A is also a contact of B(Lerman & Jones 2006). Lerman & Jones (2006) uses several data sets to study the social browsing behavior in flickr. It discovers strong correlation between a photos view number and it owner's reverse contact number(the number of users assign the owner as contact). It also discovers that a large proportion of a photo's comments are from its owner's reverse contact. This indicates that the social network formed by contact relation helps to reveal a user's preference and viewing habit.

Flickr also provides mechanisms for users to explicitly specify their opinion or preferences on a particular photos. They can write comments on photos, display photos in their galleries and vote a photo as favorite. Favorite voting is the most commonly used rating feature in Flickr, the voting summary of a photo is public to the viewer and shown on the photo's main page. Voting as favorite can be seen as a direct and positive feedback to the photo.

Gallery is a relatively new feature, it can be seen as an advanced version of *favorite*. A user can create a few galleries to feature her favorite photos. Galleries can reflect one's preference more properly since they are more organizable. A photo can be included in multiple galleries of a user. In particular, favorite voting and galleries are independent. A photo can appear both in a user's favorite list and galleries; or it can appear just in one place.

Commenting feature is one of the oldest interactive mechanisms used in many websites. It is also used very extensively by Flickr users. However, a user may leave positive, negative or off-topic comments on photos. Language processing is required to understand the nature of comments. Therefore, we will particularly look into the favorite voting and galleries in our research and use them as photo's quality indicators.

Lerman & Jones (2006) observes that users do view and leave comments on their contacts' photo. We argue that a user may share preference in terms of content and quality with her contacts. To assess the content similarity, we examine the meta data of photos uploaded by a user and by her contacts. To assess the quality preference, we examine if a user has a high proportion of photos as favorite or in gallery from her contacts' uploads.

The content similarity is computed using the *vector space model*(Salton et al. 1975) by mapping a user as a document consisting of tags from all her uploaded photos. We select 4 sets of target users differ in the number of contacts(around 50, 100, 200, 300 respectively). A target user and her contacts form the contact collection C. We also select a set of random users(more likely to be "strangers") for each target user to form a random collection R. The size of R is determined by the size of C. We compare the average similarity between a user and her contacts with the average similarity between a user and random users. The result shows that a user is more "similar" with her contacts than with strangers(37.73% greater as shown in table 1), which means a user is more likely having shared preference with her contacts in terms of their uploaded photos. We do not see very high similarity values between user and her contacts because most users' content preference are diverse and it is not likely for two users to match 100% in terms of content.

We take a sample of around 2000 users from flickr and collect their contacts, favorite photos and favorite photo owners. We compare a user's contacts with the favorite photo owners to find percentage of overlapping. In particular, we compute the percentage of a user's favorite photos that are owned by her contacts (foc), the percentage a user's contacts that has at least one photo voted as favorite by th user (caf) and the percentage of a user's favorite photo owners that are also in her contact list(fic). Table 2 shows the

Table 1: Comparison of average user similarities by percentage

size	avg. sim in C	avg. sim in R	+%
±50	0.341668	0.249671	36.84%
±100	0.2697	0.202808	33.17%
±200	0.379732	0.288658	31.55%
±300	0.319567	0.246634	29.57%
total avg.	0.32749	0.24674075	**32.73%**

result. The first row shows the overall averages of the important measures. Depending on the total number of contacts a user has and the total number of favorite photos she has voted, the overlapping percentages are slightly different. We show the stratified averages in this table.

Table 2 clearly shows that a large proportion of a user's favorite photos comes from her contacts. When a user has 100+ contacts, more than half of the favorite photos are from the contacts. There is also a large proportion of a user's contacts has at least one photo being voted as favorite. When a user has 1000+ favorite photos, more than half of the contacts has a photo voted as favorite.

Gallery serves a similar purpose with favorite voting from the perspective of photo quality, we are interested to see if they overlap a lot. We take a sample of photos that have non-zero favorite counts and have at least been featured in one gallery. The photos are classified into 4 groups depending on their favorite counts. We obtain each photo's gallery counts, gallery owners and favorite count and compute average values for each group. Table 3 shows the result. Although the gallery count increases with the favorite count, the large gap between the two indicates gallery featuring is not as popular as favorite voting. Discrepancy between gallery counts and gallery owners shows that a user may include a photo more than once in her galleries. We compute the number of overlapping users who both votes a photo as favorite and includes it in galleries. The result shows only 30.94% gallery users vote their photo as favorite, which means gallery and favorite features are relatively independent. Therefore, it is worthwhile to use gallery featuring as one of the quality indicators of photos.

These observations show that the contact network in flickr is a good indicator of shared taste in content and photo quality. In other words, it is safe to say that we can infer a user's content and quality preference from its contacts. It can be used as a trust network in terms of content and quality.

3 Link Analysis and Query Result Ranking

Link based object ranking has long been used in the context of web information retrieval with two notable algorithms PageRank and HITS(Getoor & Diehl 2005). The *query independent algorithm* PageRank(Page et al. 1999) computes importance scores for all web pages. The scores can be biased towards a topic computed using a variation of the algorithm called Personalized Page Rank. In contrast, HITS(Hyperlink-Induced Topic Search)(Kleinberg 1999) computes a hub and authority score for each page in a broad topic query results. HITS often operates on a sub-graph constructed from the query result and is considered as *query dependent algorithm*. There are quite a few efforts trying to apply, to extend or to combine this two approaches for better and more resilient ranking(Zheng et al. 2009, Ding et al. 2002, Lempel & Moran 2000). In particular, Dell'Amico et al. Dell'Amico & Capra (2008) propose SOFIA (SOcial FIltering Algorithm) to make social recommendation in a robust way. SOFIA operates on two networks: a social network among users and a judgment network between users and objects. It applies Personalized PageRank to quantify trust intention and use the trust intention to compute subjective HITS scores for recommending objects to a target user.

We adopt the idea of trust network and judgment network from SOFIA. Since the query user can be an anonymous Internet surfer or any registered user, it is not practical to provide fine tuned results for each individual user. We use pre-defined categories (such as *animal*, *architecture*) as preference indicators. For each category, a few seed users who have uploaded photos related with that category are selected as filters for it. We have proven experimentally in section 4 that general category together with query term can effectively pick high quality photos in many noises and the results are not sensitive to seed user selection.

3.1 Trust Network and PPR computation

Personalized PageRank (Page et al. 1999) expresses interest or topic preference as a set of predefined web pages call seed. In addition to following hyper links in a page, a random walker has a predefined probability (computed based on a given damping factor) to jump to any of the seed pages. The PPR can be applied in computing a particular user's trust vector by setting this user as the seed. Mathematically, a vector E is used to represent the seed. E has a value 1 at the position corresponding to the seed user and 0 in all other positions.

For a trust network $G \in (V, E)$, let O be the number of out-links (one's contacts); I be the number of in-links; for a user $u \in G$, t_u represents the trust vector. The trust value a user u imposes on a user $v \in G$ can be defined as below:

$$t_u(v) = (1-d) \sum_{q \in I_v} \frac{t_q}{O_q} + dE_u(v)$$

d is a damping factor defined in (Page et al. 1999), setting to a value representing the probability of deviating from random walk along hyper links. The recommended value for d is 0.15.

Not all users have contacts. A user with no contact becomes dangling nodes in the social network. Page et al. (1999) suggests to remove dangling pages before the computation. However, this approach is impractical to here because users without contacts are still deserved to possess a trust value as long as they are reachable from the seed user. Ng et al. (2001) and Lempel & Moran (2000) make another assumption that each dangling page should be a special page that points to all the pages; In the random surf model, this can be explained as the surfer has a equal chance to jump to any other page when he/she is about to stop at a dangling page. This assumption is convincing when dealing with the dangling pages in PageRank computation. Yet in terms of the trust propagation, we should not allow trust distributes to a random node.

Another approach to address the dangling problem is to create a "dummy" node(Bianchini et al. 2005). The dummy node is pointed by all the dangling node from the graph and has a self link.

Table 2: Comparison of a user's contacts and favorite photos and favorite photo owners

	#contacts	#fOwners	#fPhotos	fic	caf	foc
Overall	252.55	407.09	1143.61	24.77%	38.23%	39.62%
contacts[1,100)	32.30	193.02	432.94	15.53%	35.49%	25.80%
contacts[100,500)	231.99	547.01	1572.86	32.48%	44.33%	**53.37**%
contacts[500,1000)	672.63	818.30	1834.58	45.13%	40.47%	**66.13**%
contacts[1000,-)	2319.54	1314.88	2712.15	49.48%	23.78%	**65.85**%
fPhotos[1,100)	89.95	26.59	36.84	25.48%	17.49%	30.70%
fPhotos[100,500)	173.63	158.32	264.34	24.29%	37.70%	37.30%
fPhotos[500, 1000)	278.32	360.82	717.69	25.83%	47.37%	45.57%
fPhotos[1000,5000)	431.04	889.52	2152.75	24.51%	**57.00**%	48.67%
fPhotos[5000,-)	959.54	2499.81	10350.91	21.85%	**68.96**%	**56.88**%

Table 3: Overlap between favorite and gallery photo

	gallery_count	gallery_owner	fav_count	overlap	%
fPhotos[0,100)	4.33	4.20	59.02	1.17	27.86%
fPhotos[100,500)	10.23	9.70	212.48	3.00	30.93%
fPhotos[500,1K)	26.27	25.71	664.02	8.10	31.51%
fPhotos[1K,-)	47.20	45.85	1696	17.20	37.51%
overall	9.00	**8.63**	190.83	**2.67**	**30.94**%

This approach eliminates dangling nodes yet produces a heavy weighted dummy node. In fact, such a "dummy" node occupies a large proportion of ranking score in our PPR approach, because values from all dangling nodes aggregates at "dummy" node and the self link prevents them from being distributed again.

To overcome the dangling node issue properly, we simply combine the ideas from above: for each dangling node in our trust network, we create a back link to the seed node. This ensures the sum of trust scores on all nodes equals to 1.

3.2 Ranking based on Judgment Network

With any set of photos P, we can construct a bipartite graph with links from a user set U to the photo set P. We add a link from user u to photo p if u votes p as favorite; adds p in her galleries; or if p is simply uploaded by u. It is easy to see that such link represents certain positive judgments from a user to a photo. We call the bipartite graph a judgment network. The weight of the link w_{up} is computed by counting the number of judgments. For instance, if u votes p as favorite and add p in two of her galleries, then $w_{up} = 3$. We aim to compute a rank score for p using HITS algorithm. The computation is based on collective judgments from users and the trusts a query user placed on those users.

In our judgment network, users are pure *hubs* and photos are pure *authorities*. Therefore the *authority* values can be used as the ranking scores. (Lempel & Moran 2000) propose the idea of subjective HITS. (Dell'Amico & Capra 2008) further develop an algorithm to include trust value in the *backward* stage of computing hub value based on authority values. We adopt the similar idea with some modifications. For each user $u \in U$ and photo $p \in P$, the subjective HITS algorithm will generate the hub H and authority A iteratively through following 2 operations:

$$A(p) = \sum_{u:w_{up}>0} \frac{w_{up}}{\sum_{q:w_{uq}>0} w_{uq}} \cdot H(u) \qquad (1)$$

$$H(u) = \sum_{p:w_{up}>0} \frac{T(u) \cdot w_{up}}{\sum_{v:w_{vp}>0} T(v) \cdot w_{vp}} \cdot A(p) \qquad (2)$$

We start the computation by initializing the hub value to $\frac{1}{|P|}$ for each user. This ensures a total score of 1 to be divided among authorities and hubs in each forward and backward steps respectively. Equation 2 represents the backward step. For each photo p, we distribute the hub value to all pointing users based on their trust value. As this algorithm runs iteratively, the authority value of a photo will also be affected by the trust of all users who place positive judgments on it. Algorithm 1 gives the steps of computing hub and authority scores iteratively.

Algorithm 1: trust weighted HITS

Input: $G = (U, P, E)$: a judgment network, T: a trust vector
Output: A: Authority values for all $p \in P$; H: Hub values for all $u \in U$
Let z denote the vector $(\frac{1}{|P|}, \frac{1}{|P|}, \cdots, \frac{1}{|P|}) \in R^{|P|}$
Initialize A to z
while *not converged* **do**
  foreach u *in* U **do**
    apply Equation 2 to compute H(u)
  foreach p *in* P **do**
    apply Equation 1 to compute A(p)
return H, A

4 Experiment and Results

4.1 Constructing social network

A list of typical ambiguous query terms have been used in many experiments in IR field(Cai et al. 2004, Goldberger et al. 2006, Liu et al. 2009). We are interested in not only individual terms, but also the top category those terms come from. One of the top categories where many ambiguous query terms come from is *animal* since animal names are used in various places. For instance, widely recognized ambiguous animal terms include *tiger*, which also refers to a baseball team (Detroit tiger), a golf player(tiger wood), a type of flower (tiger lily), apple operation system and so on; *jaguar* which also refers to a car model; *Pluto* which can refer to a planet; *raptor* which also refers to a type or aircraft. We choose a few top categories containing many ambiguous terms to start our experiment data collection. These include *animal*, *flower*, *car* and a few others. We then locate a few large groups in that category using flickr's group search function. We download all users from each group and use those as starting point to build a contact network by progressively crawl each level of contacts. We have obtained 72.9M contact links among 3.45M users, in which 300K users have at least 1 contact and over 120K of them have less than 50 contacts. These form the underlying social network.

4.2 Experiment design and metrics

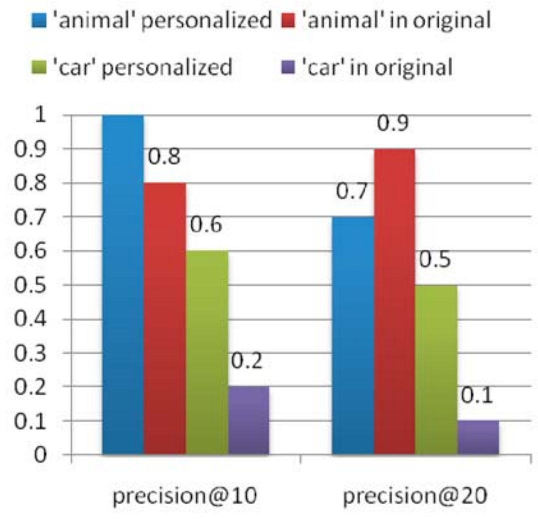
We use raw Flickr query results as test collection. Each query may represent one or many underlying topics belonging to various top categories. We also select seed users for topic or category by matching their preferences with the target category. For simplicity, each user's preference is represented by top tags in her photo collection. We run experiments to test the personalization power of our approach and compare it with simple query expansion approach.

Any web search returns a large number of results. Most people care only the top few results. Precision at (top)k is a typical measure used in ranked query to evaluate the quality of results(Manning et al. 2008). We adopt this measure in our experiments. We set k=10 and 20 respectively. The relevancy is judged by human evaluators based on seed user preference and the photo content.

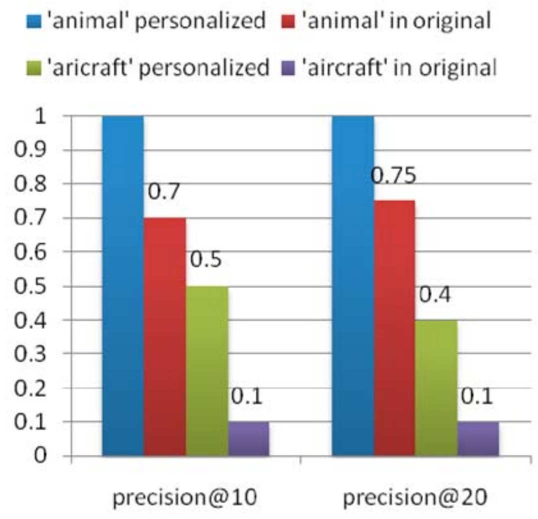
4.3 Results and Discussion

First, we run a set of experiments to see if users representing a general top category can effectively pick out photos in that category from an ambiguous query result. We choose two ambiguous terms *jaguar* and *raptor* to construct the test collections by querying flickr to obtain the *top200* results ranked by relevancy and by interestedness. Both terms return results sets containing two major topics: jaguar animal vs. jaguar car and raptor animal vs. raptor aircraft. We select users representing *animal*, *car* and *aircraft* as seed to re-rank the results.

The ranking result for each query is shown in figure 2. Results from our approach are denoted as "personalized" while original flickr ranking are denoted as "original". The personalized results show improvement in *jaguar animal* on precision@10 and *raptor animal/aircraft* on both precision@10 and @20. We can that in the original results, photos of *animal* category from both queries have high precisions(at least 0.7), which means they are the dominant topic for the respective ambiguous terms. Our approach performs well for less dominant topics(e.g. for *jaguar car*, the precision is increased from 0.2 to 0.6 at top10 and from 0.1 to 0.5 at top20). The performance of dominant topics is also promising (e.g. for *jaguar animal*, the precision is increased to 1 at top10 and for *raptor animal* the precisions are increased to 1 at top10 and top20). In certain cases, the popular topic's personalized result might not be as good as the original one because our test collection has a fixed number of photos and there is little room for improvement.



(a) jaguar

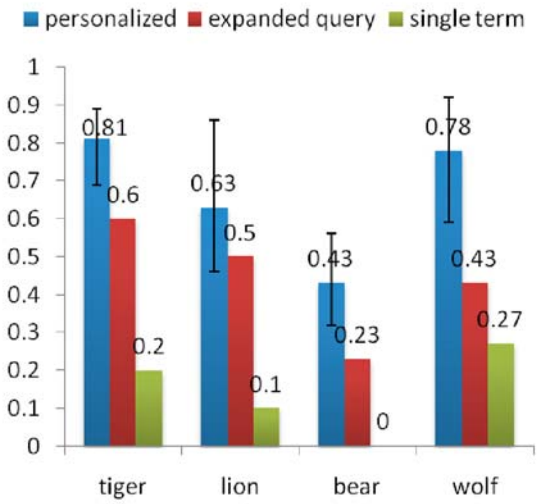


(b) raptor

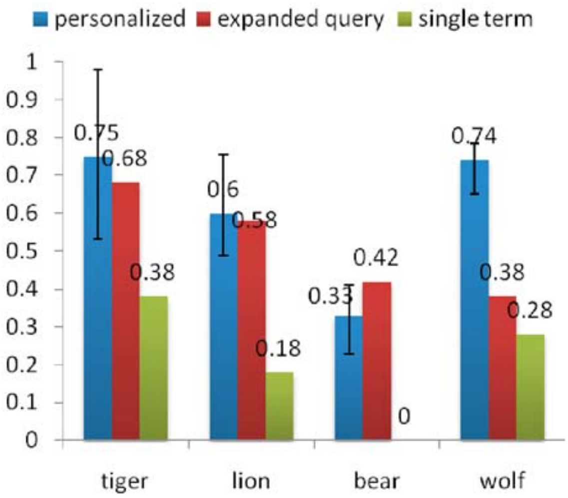
Figure 2: Precision comparison on input queries

When query results are not desirable, it is not uncommon for a user to expand the original query with additional keywords. Such simple expansion may help to filter out unrelated results. For example, the original *bear* query contains lots of noise including photos about flowers and landscape(as shown in figure1. A user interested only in bear animal may have a second attempt by expanding the query to specify its top category such as *bear animal*. However, simple query expansion does not always improve the quality of results. Flickr *interestingness* result for a expended query *bear animal* contains only 3 out of top10 relevant results.

The next set of experiments compare our approach with simple query expansion approach. It also examines the sensitivity of seed users selection. We first select two top categories *animal* and *flower*; for each category, we choose several topics such as *tiger, lion, wolf, bear* in *animal* and *rose, lotus, pansy, peony* in *flower*. For each topic, we run a single term query and an expanded query to obtain two sets of results from Flickr. Expanded queries are formed by adding the category name to the original term. Next we find six seed users for each general category. We ensure that seed users have common interest in the top category, while each may have preference in certain topics within that category.



(a) Precision@10 in category *animal*



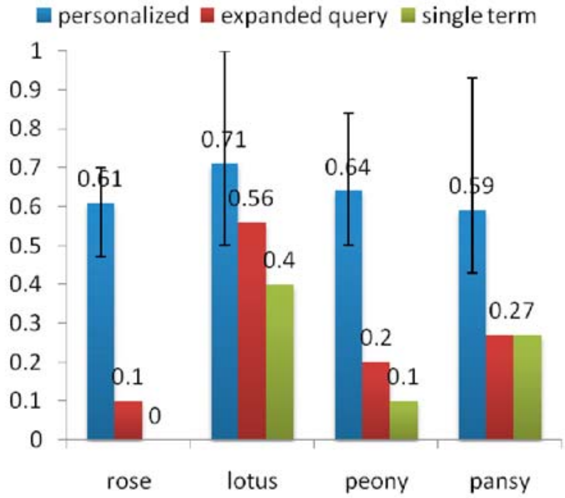
(b) Precision@20 in category *animal*

Figure 3: Precision comparison for category *animal*

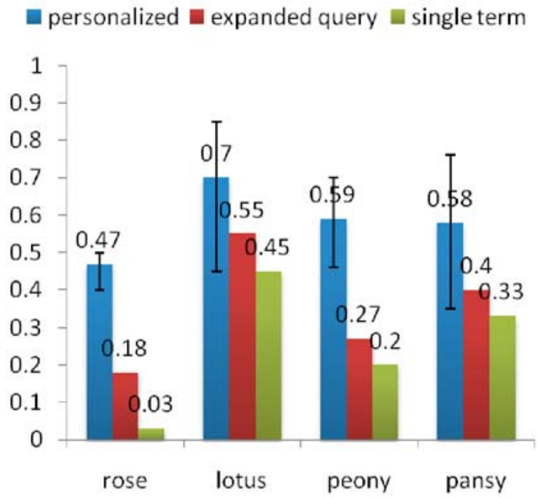
We apply trust weighted HITS algorithms on single term query results for each seed user and compare the precision with that of original single-term and expanded query result. Figure 3 and 4 give the detailed precision comparison over the two general categories. For the personalized result, the figure shows the average precision of all 6 users with the error bar showing the maximum and minimum individual precision.

Single-term query result in general has the worst precision. Expanded query performs better than single term query in all cases. The average precision of all 6 seed users are higher than that of expanded query result, but use some seed users may not get better results than simple query expansion. For instance, in figure 3 the lowest precision@10 from personalized result for query *lion* is lower than the precision achieved in expanded query. This is more obvious in precision@20 chart.

An interesting observation is that the precision values of personalized results are not sensitive to seed



(a) Precision@10 in category *flower*

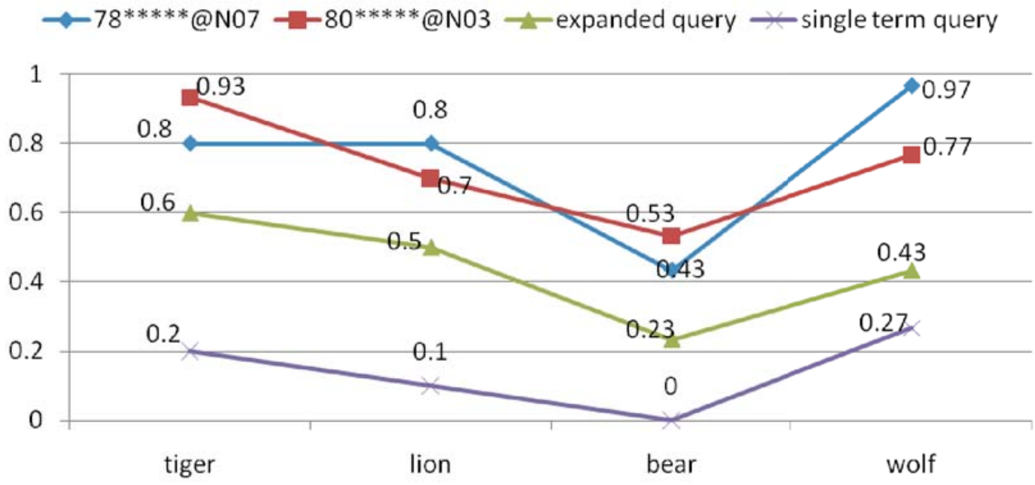


(b) Precision@20 in category *flower*

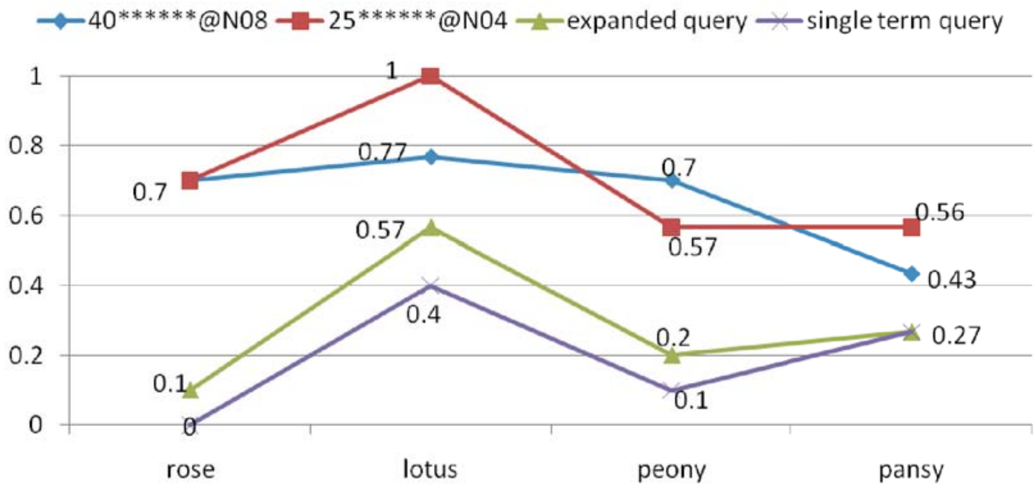
Figure 4: Precision comparison for category *flower*

users selection. The error bar shows the range of precision among all users. There are some big ranges in certain sub-topics. However, all seed users achieve much better precision than single query results; majority of the users achieve better precision than expanded query result. Such insensitivity is a good indicator for the practical value of our approach. It implies that we only need to precompile a few seed users for some top categories.

The variation of precision are caused by preference differences among the users representing the same top categories. If a seed user has preference on one or two particular sub-topic. The personalized rank based on that user would achieve the highest precision for queries of that particular sub-topic. From the *animal* category, we extract user 78*****@N07 who prefers *lion* and user 80*****N03 who prefers *tiger* and plot the precision against those from original and expanded queries in Figure 5.(a). We can see the relatively high quality personalized results of user 78*****@N07 and 80*****@N03 gained at their special interests(0.8 and 0.93 respectively). For the *flower* category, we extract user 25******@N04 who prefers in general flowers and user 40******@N08 who has a special interest in rose and any pink flowers and



(a) Precision@10 of two users over all 4 topics in *animal*



(b) Precision@10 of two users over all 4 topics in *flower*

Figure 5: Partial personalized results compared with results obtained via single-term query and query expansion

plot the precision comparison in Figure 5.(a).

However, such biased preference towards a specialized topic does not prevent a seed user from being a good filter for other sub-topics. Sometimes the broad interests of a user's contacts compensates the effect. In Figure 5 (a), it is clear that both users perform better than original and expanded queries across all four sub-topics. After evaluating the preferences of their contacts, we found that most contacts of user 80*****@N03 are interested in *animal*, but not being restricted to *tiger*. Similarly, in category *flower*, none of the two users are particularly interested in topic *lotus*, but our approach can generate quality results for *lotus* query with precision above 0.7.

5 Related Work

We adopt the idea of trust and judgment network as presented in SOFIA(Dell'Amico & Capra 2008). However, SOFIA is focused on recommendation and the presence of seed user and its judgment in the judgment network is important for predicting judgment on other objects in the collection. We focus more on re-ranking query results based on content preference and quality. Our algorithm applies different initialization and weight distribution operations to achieve subjective HITS scores. It does not rely on a seed user's judgment links. More importantly, the ranking is not sensitive to seed user selection hence it is very practical.

It is impossible to compute a PPR vector from scratch at query time. Jeh & Widom (2003) gives an efficient way of computing PPR vectors on the fly using precomputed partial vectors and *hub skeleton*. We adopt the idea in our implementation.

6 Conclusion

In this paper we apply link analysis methods to rerank ambiguous query results based on seed users' social contacts and quality indicators of non-textual resources. The heuristic behind our approach is that contact relationship between users can reflect their shared preference, which is confirmed by observations on sample data from Flickr. We further define the behavior of adding contacts in the social network as a way of expressing trust and quantify such trust via Personalized PageRank(PPR). The ranking is performed on a judgment network based on original query results and associated users. Particularly to Flickr data, we use three types of quality indicators to quantify the judgment between users and photos: ownership, favorite voting and gallery featuring. We apply a trust weighted HITS on such judgment network to generate re-rank based on the preference of the seed user.

Experiments are carried out on Flickr datasets. We compare the precision again original and simple query expansion results. Our approach can generate better ranking based on a seed user and its social contacts. In particular, our approach only require a small set of seed users representing top categories, the experiment shows that such seed users are able to re-rank topics within the top category. The results are not sensitive to the selection of seed users in terms of the size of their social network and the judgment links they may have. Results may vary for different seed users but in general, the precision is higher than that of original and expanded query results.Although the algorithm and experiment are mainly based on Flickr network. The general idea presented can be used in other content sharing sites with an underlying social network.

References

Bianchini, M., Gori, M. & Scarselli, F. (2005), 'Inside pagerank', *ACM Trans. Internet Technol.* **5**, 92–128.

Cai, D., He, X., Li, Z., Ma, W.-Y. & Wen, J.-R. (2004), Hierarchical clustering of www image search results using visual, textual and link information, *in* 'Proceedings of the 12th annual ACM international conference on Multimedia', MULTIMEDIA '04, ACM, New York, NY, USA, pp. 952–959.

Dell'Amico, M. & Capra, L. (2008), Sofia: Social filtering for robust recommendations, *in* Y. Karabulut, J. Mitchell, P. Herrmann & C. Jensen, eds, 'Trust Management II', Vol. 263 of *IFIP Advances in Information and Communication Technology*, Springer Boston, pp. 135–150. 10.1007/978-0-387-09428-1_9.

Ding, C., He, X., Husbands, P., Zha, H. & Simon, H. D. (2002), Pagerank, hits and a unified framework for link analysis, *in* 'Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval', SIGIR '02, ACM, New York, NY, USA, pp. 353–354.

Getoor, L. & Diehl, C. P. (2005), 'Link mining: a survey', *SIGKDD Explor. Newsl.* **7**, 3–12.

Goldberger, J., Gordon, S. & Greenspan, H. (2006), 'Unsupervised image-set clustering using an information theoretic framework', *Image Processing, IEEE Transactions on* **15**(2), 449 –458.

Jeh, G. & Widom, J. (2003), Scaling personalized web search, *in* 'Proceedings of the 12th international conference on World Wide Web', WWW '03, ACM, New York, NY, USA, pp. 271–279.

Kleinberg, J. M. (1999), 'Authoritative sources in a hyperlinked environment', *J. ACM* **46**, 604–632.

Lee, J., Hwang, S.-w., Nie, Z. & Wen, J.-R. (2009), Query result clustering for object-level search, *in* 'Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining', KDD '09, ACM, New York, NY, USA, pp. 1205–1214.

Lempel, R. & Moran, S. (2000), 'The stochastic approach for link-structure analysis (salsa) and the tkc effect', *Computer Networks* **33**(1-6), 387 – 401.

Lerman, K. & Jones, L. (2006), 'Social browsing on flickr', *CoRR*.

Liu, D., Hua, X.-S., Yang, L., Wang, M. & Zhang, H.-J. (2009), Tag ranking, *in* 'Proceedings of the 18th international conference on World wide web', WWW '09, ACM, New York, NY, USA, pp. 351–360.

Manning, C., Raghavan, P. & Schtze, H. (2008), *Introduction to Information Retrieval*, Cambridge University Press New York, NY, USA.

Ng, A., Zheng, A. & Jordan, M. (2001), Link analysis, eigenvectors and stability, *in* 'International Joint Conference on Artificial Intelligence', Vol. 17, Citeseer, pp. 903–910.

Page, L., Brin, S., Motwani, R. & Winograd, T. (1999), The pagerank citation ranking: Bringing order to the web., Technical Report 1999-66, Stanford InfoLab. Previous number = SIDL-WP-1999-0120.

Sadikov, E., Madhavan, J., Wang, L. & Halevy, A. (2010), Clustering query refinements by user intent, *in* 'Proceedings of the 19th international conference on World wide web', WWW '10, ACM, New York, NY, USA, pp. 841–850.

Salton, G., Wong, A. & Yang, C. S. (1975), 'A vector space model for automatic indexing', *Commun. ACM* **18**, 613–620.

Zheng, Y., Zhang, L., Xie, X. & Ma, W.-Y. (2009), Mining interesting locations and travel sequences from gps trajectories, *in* 'Proceedings of the 18th international conference on World wide web', WWW '09, ACM, New York, NY, USA, pp. 791–800.