Filtering Information with Imprecise Social Criteria: A FOAF-based backlink model

EUSFLAT 2005 (Barcelona)

Elena García-Barriocanal
elena.garcia@uah.es
Miguel A. Sicilia
msicilia@uah.es
Overview

• Introduction
  – Motivation
  – PageRank background

• A model of social links
  – PeopleRank algorithm
  – Imprecision in computing PeopleRank

• Example implementation

• Conclusions and future work
Motivation

- Web structure metrics has proved to be a very effective paradigm for ranking search results.
  - PageRank
  - Hub-authority
  - ...
- PageRank algorithm uses information which is external to the Web pages themselves – their backlinks, which provide a kind of peer review
  - This entails a subjective value indicator (not always trustworthy- Google boombing)
- Social network models can be used as a weight of reliability.
  - The FOAF project will eventually make available this information.

Examples of google boombing

- Using Google with a criteria like “hell”, it retrieves Microsoft home page in the first position of result list, and using the term “Buffone” it retrieves the Berlusconi’s home page in first position of result list.
PageRank background

- PageRank is one of the methods Google uses to determine a page’s relevance or importance.
- Algorithm calculates a value for every page’s relevance using its backlinks:
  - If page A links out to page B, then page B is said to have a “backlink” from page A.
- The number of links from the source page to other pages is used as correction factor.
  - The fewer links the page has, the more relevance each link has.
PeopleRank Algorithm (i)

• The algorithm definitions and variables:
  – Person A has persons T1..Tn which declare they know him/her (i.e., provide FOAF social pointers to it).
  – The parameter d is a damping factor which can be set between 0 and 1.
  – Also C(A) is defined as the number of (interpreted) declarations going out of A’s FOAF profile.

• The PeopleRank of a person A is given as follows:
  \[ PpR(A) = (1-d) + d \times \left( \frac{PpR(T1)}{C(T1)} + \ldots + \frac{PpR(Tn)}{C(Tn)} \right) \]
  – A model for strength of social ties is needed.
PeopleRank Algorithm (ii)

- Although PageRank and PeopleRank computation structure are similar, redefinitions are needed:

<table>
<thead>
<tr>
<th>PageRank</th>
<th>PeopleRank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intuitively, pages that are well cited from many places around the web are worth looking at.</td>
<td>Intuitively, the trust on the quality of pages is related to the degree of confidence we have on their authors.</td>
</tr>
<tr>
<td>Also, pages that have perhaps only one citation from something like the Yahoo! home-page are also generally worth looking at.</td>
<td>Pages authored or owned by people with a larger positive prestige should somewhat be considered more relevant.</td>
</tr>
</tbody>
</table>
Imprecision in computing
PeopleRank (i)

• The relevance of a social tie is modeled by the general expression:
  \[ r((p_1, p_2)) = s((p_1, p_2)) \times e((p_1, p_2)) \quad , p_2 \neq p_1 \]

• The relevance \( r \) of an edge is determined from:
  – Degree of strength \( s \) (in a scale of fuzzy numbers \([-10, 10]\))
  – Degree of evidence \( e \) about the relationship.

• Degrees of evidence \( e \) support a notion of "external" evidence on the relation that completes the subjectively stated strength.

\( r((p_1, p_2)) \) provides a value for each pair of edges \((p_1, p_2)\) in the directed graph formed by the explicitly declared social relationships.

• The strength \( s(p_1, p_2) \) could be provided by extending recurrent FOAF schema with an additional attribute.
Imprecision in computing
PeopleRank (ii)

- Evidence is based both on the perceptions of "third parties" and on the declaration of common projects.

\[ e(p_1, p_2) = \max(\Phi_{\text{tr}}(p_1, p_2), P(p_1, p_2)) \]

\[ p_i \neq p_2 \neq p_1 , \ i \in U \]

- \( \Phi \) is a simple fuzzy average of strengths of a social tie provided by third parties \( p_i \) in a group of user \( U \).

- \( P(p_1, p_2) \) is the evidence provided by work in common projects in which two persons \( (p_1, p_2) \) collaborate, which are obtained from FOAF declarations:
  - Persons involved in \textit{foaf:currentProject} relations are credited an \textit{(explicit)} amount of 1.
  - Persons involved in \textit{foaf:currentProject} relations are credited an \textit{(explicit)} amount of 0.1.
Example implementation (i)

- The example implementation has used the PageRank class provided by the JUNG libraries.
- The proposed algorithm combines social with standard backlink consideration and has been labeled PageRankSocial
  - Inputs: people graph obtained from the FOAF annotations in Web resources and a graph with documents that those persons own (document graph).
  - Output: a modified document graph, which contains a value for each node that represents its relevance.
Example implementation (ii)

```plaintext
computePageRankSocial(S, D)
  ▷ D is the document graph
  ▷ S is the people graph

1. S ← computeSocialRelevance(S)
2. for each v ∈ vertex(D)
   do
3.   v.source.relevance ←
       NODES(S)[v.source].relevance
4. D ← weightedPageRank(D)
```

- Two steps in the process:
  - A first social-specific computation computeSocialRelevance using expression (1) and a defuzzyfication method to compute fuzzy numbers resulting of that. computeSocialRelevance uses the PeopleRank formulations.
  - An standard use of the weighted page rank provided in JUNG libraries, using the previously computed weights as source-related relevance indicators.

- Line (3) describes an important modification introduced. This line describes the code that assigns as relevance of each document the relevance of the person who owns the document. This value is interpreted by the JUNK implementation of weightedPageRank as a weighted arch in the document graph.
Example implementation (iii)

- Social network view of PeopleRank
  - $r_1 < r_3 < r_2$

Each node in the figure represents an individual in a group of users.
PeopleRank obtains a relevance $r$ for each pair of individuals, in the basis of an explicit degree of strength (represented using FOAF specification) and a computed degree of evidence (using third parties opinions about the specific social tie and objective evidence provided by work in common projects).
Example implementation (iv)

- Resulting document backlink relevance network with social weighting

• The figure shows the document graph resulting of processing the document graph using the algorithm explained before.
• The JUNK weightedPageRank algorithm obtains the new value for each page with consideration of social relevance. The higher the value for a node (page), the higher position the page will appear in the ranking after a search-by-criteria query.
• Node 14 will be listed in a higher ranking position than (for example) node 10 or node 19
Conclusions and future work

- A model of Web page ranking combining link-based metrics with a basic account of social relations has been described.
- The model provides a tentative account for evidence and strength of relationships considering:
  - The opinion of the individuals in the relationship.
  - The external view of other user in the group.
  - A view of co-working as an indicator of evidence about the relationship.
- The model is useful to obtain a more trustworthy ranking of search results.
Conclusions and future work

• Two steps in algorithm:
  – A measure of social relevance is computed
  – Once the measures are defuzzified, they are used as a weighting factor for computing the relevance of documents.

• Future work:
  – To study of other FOAF vocabulary terms that could be subject to an specific and differentiated consideration in ranks that consider social relations.
  – The tentative scheme provided here requires much empirical study to come up with the better computational scheme with regards to reflecting the effects of “prestige” in Web documents.
  – Empirical research are needed to obtain a better fit of algorithm parameters.