Generation of Large Hypertext and Document Networks from Collective User Retrieval Patterns

Johan Bollen
Old Dominion University
Department of Computer Science

jbollen@cs.odu.edu
http://www.cs.odu.edu/~jbollen
A Personal Introduction
Johan Bollen

   (a) Web site on philosophy of cybernetics
   (b) Designed to represent “knowledge” network, systems science encyclopedia
   (c) Collective effort of 4 editors
   (d) Adaptive Hypertext and models of navigation

   (a) Focus on Distributed Information Systems
   (b) Generation of document proximities for DL recommendation systems
   (c) Application of spreading activation systems

3. **Old Dominion University, Computer Science Department**
   (a) Systems for scientometric analysis
   (b) Large scale applications of spreading activation for cross-resource recommendation
   (c) Temporal analysis of changes in user community preferences
Problem Statement

1. Initially: Hypertext
   (a) **Design and use**: Collection of documents need to be meaningfully linked and user retrieval needs to be made more efficient
   (b) Design:
      i. **Manual Design**: Uneconomical, and user preferences?
      ii. **Taxonomies**: Correspondance to user preferences?

2. More generally: **Digital Libraries**
   (a) Desire to link collections and provide retrieval facilities beyond keyterm matching
   (b) Heterogeneous collections (format, language, etc)
   (c) Scalability issues

   iii. **IR methods**: Dependance on text, and explicit queries
   (c) Use:
      i. Lost in Hyperspace
      ii. Users prefer Boolean searching
Basic Approach

1. Hypertext linking:
   (a) Shift from focus on design to usage
   (b) Patterns of retrieval determine generation of hyperlinks
   (c) Relations to Machine Learning: auto-associators and Hebbian learning

2. Hypertext retrieval:
   (a) Focus on use of network structure rather than document content
   (b) Recurrent spreading activation
   (c) Relations to CLEVER and PAGERANKER

3. Digital Libraries
   (a) Reconstruction of retrieval patterns from server logs
   (b) Generation of large scale document and journal networks
   (c) Network analysis for document and journal impact ranking
Presentation Structure

1. Background: Hypertext linking
   (a) The early experiments
   (b) Results and applications
   (c) Relevancy to general WWW interfaces

2. Digital Libraries
   (a) System for document linking from server logs
   (b) Journal and document impact ranking (as opposed to ISI IF)
   (c) Spreading Activation retrieval
Automated Hypertext Linking: first experiments

1. Replace explicit design or text generated links by user retrieval
2. Frequency of retrieval sequence indicates strength of associative relation
3. Shift from explicit author or designer preferences, to implicit user preferences
Experiment Setup

1. Representation
   (a) Weighted Hyperlinks: weights indicate degree of association
   (b) Directed Weighted Graph Representation of Network Structure
      i. Assumption is asymmetric associations
      ii. Corresponds to word association research and WWW hyperlinks

2. Adaptation
   (a) Reinforcements based on
       Hyperlink Traversals
       (b) Learning Rules Operate on
           Hyperlink Weights
           i. Small additions to link weights for each traversal
           ii. Aaimed at gradual adaptation of structure

3. Interface
   (a) Weighted Ordered Ranking of Hyperlinks
   (b) Dynamic Re-Ordering
   (c) Display “slices” of 10 hyperlinks
Graph Representation
Hypertext Graph Representation (Furner, 1996), (Kleinberg, 1999)

\[ G = (V, L) \]

\[ V = \{v_0, v_1, \cdots, v_{n-1}\} \]

\[ L \subseteq V \times V \]

\((v_i, v_j) \in L\) represents hyperlink \(v_i \in V \rightarrow v_j \in V\)

\(n \times n\) matrix, \(H\), binary values \(h_{ij} \in \{0, 1\}\)

 hyperlink: \((v_i, v_j) \in L : h_{ij} = 1\)

 absence hyperlink: \((v_i, v_j) \notin L : h_{ij} = 0\)
Graph Representation Assumptions

1. Hypertext Pages ~ Graph Nodes?
   (a) Homogeneous content
   (b) Singular Concept

2. Hyperlinks ~ Graph Edges
   (a) Singular anchor and target
   (b) Hyperlink Semantics

(c) Directed Connections

Simplification

1. Feature rather than bug
2. Goal is not to construct “best” hypertext representation but study larger class of linked information systems
Learning Rules

\[ G = (V, L) \]

add weight function \( W \),

\[ G = (V, L, W) \]

\[ W : V^2 \rightarrow \mathbb{R}^+, W(v_i, v_j) \in \mathbb{R}^+ \]

User Request Sequence (RS): \( p = (v_0, v_1, \cdots, v_{n-1}), n \in \mathbb{N}^+ \)

RS Sub-path Triplet: \( p_s = (v_i, v_j, v_k), i, j, k \in \mathbb{N}^+ \)

\[ j = i + 1, k = i + 2, i < n - 2 \]
Learning Rules

**Frequency (F) or** $F(p_s, W)$:

- $W(v_i, v_j)_t = W(v_i, v_j)_{t-1} + r_f$
- $W(v_j, v_k)_t = W(v_j, v_k)_{t-1} + r_f$

**Symmetry (S), or** $S(p_s, W)$:

- $W(v_j, v_i)_t = W(v_j, v_i)_{t-1} + r_s$
- $W(v_k, v_j)_t = W(v_k, v_j)_{t-1} + r_s$

**Transitivity (T), or** $T(p_s, W)$:

- $W(v_i, v_k)_t = W(v_i, v_k)_{t-1} + r_t$
Learning rules

1. **Frequency:**
   - (a) Operates on link traversal frequency
   - (b) Reinforce existing connections
   - (c) Hebbian Learning

2. **Transitivity:**
   - (a) Operates on transitive co-retrievals
   - (b) Bridges plausible connections

3. **Symmetry:**
   - (a) Reinforce connection symmetric to those traversed
   - (b) Reinforce plausible connections
Example
Example, cont.
Interface
Link Ranking

Displaying links

1. Display shows groups of 10 links
2. Rank ordered: first slice displays 10 strongest links
3. User can shift from one group to the other

<table>
<thead>
<tr>
<th>weight</th>
<th>rank</th>
<th>hyperlink</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.215</td>
<td>1</td>
<td>period</td>
</tr>
<tr>
<td>0.126</td>
<td>2</td>
<td>moment</td>
</tr>
<tr>
<td>0.109</td>
<td>3</td>
<td>history</td>
</tr>
<tr>
<td>0.084</td>
<td>4</td>
<td>year</td>
</tr>
<tr>
<td>0.075</td>
<td>5</td>
<td>change</td>
</tr>
<tr>
<td>0.074</td>
<td>6</td>
<td>day</td>
</tr>
<tr>
<td>0.038</td>
<td>7</td>
<td>section</td>
</tr>
<tr>
<td>0.028</td>
<td>8</td>
<td>age</td>
</tr>
<tr>
<td>0.026</td>
<td>9</td>
<td>situation</td>
</tr>
<tr>
<td>0.019</td>
<td>10</td>
<td>thought</td>
</tr>
<tr>
<td>0.019</td>
<td>11</td>
<td>man</td>
</tr>
<tr>
<td>0.016</td>
<td>12</td>
<td>end</td>
</tr>
<tr>
<td>0.013</td>
<td>13</td>
<td>experience</td>
</tr>
<tr>
<td>0.013</td>
<td>14</td>
<td>week</td>
</tr>
<tr>
<td>0.012</td>
<td>15</td>
<td>blood</td>
</tr>
<tr>
<td>0.012</td>
<td>16</td>
<td>development</td>
</tr>
<tr>
<td>0.012</td>
<td>17</td>
<td>action</td>
</tr>
<tr>
<td>0.008</td>
<td>18</td>
<td>century</td>
</tr>
<tr>
<td>0.008</td>
<td>19</td>
<td>evening</td>
</tr>
<tr>
<td>0.007</td>
<td>20</td>
<td>light</td>
</tr>
<tr>
<td>0.007</td>
<td>21</td>
<td>mind</td>
</tr>
<tr>
<td>0.007</td>
<td>22</td>
<td>morning</td>
</tr>
<tr>
<td>0.007</td>
<td>23</td>
<td>theory</td>
</tr>
<tr>
<td>0.006</td>
<td>24</td>
<td>part</td>
</tr>
<tr>
<td>0.006</td>
<td>25</td>
<td>movement</td>
</tr>
<tr>
<td>0.006</td>
<td>26</td>
<td>life</td>
</tr>
<tr>
<td>0.004</td>
<td>27</td>
<td>course</td>
</tr>
<tr>
<td>0.004</td>
<td>28</td>
<td>attention</td>
</tr>
<tr>
<td>0.001</td>
<td>29</td>
<td>service</td>
</tr>
<tr>
<td>0.001</td>
<td>30</td>
<td>nature</td>
</tr>
<tr>
<td>0.001</td>
<td>31</td>
<td>sort</td>
</tr>
<tr>
<td>0.001</td>
<td>32</td>
<td>society</td>
</tr>
<tr>
<td>0.001</td>
<td>33</td>
<td>music</td>
</tr>
<tr>
<td>0.001</td>
<td>34</td>
<td>government</td>
</tr>
<tr>
<td>0.001</td>
<td>35</td>
<td>way</td>
</tr>
</tbody>
</table>
Adaptive Hypertext System Construction
Experiments

1. Open Participation
   (a) No Registration of Subjects
   (b) No Data Gathered on Characteristics\Motives of Subjects

2. Instructions to *Browse* Experimental Hypertext System
   (a) Associative Browsing: Position based Hyperlink Selection
   (b) Request not to Apply High-level Strategies

3. Specific Choice of Node Labels
   (a) English nouns
   (b) 150 most frequent

4. Two Experiments
   (a) Frequency + Transitivity ($r_f = 1, r_t = 0.5$)
   (b) Frequency + Transitivity + Symmetry ($r_s = 0.3$)
Nodes
Results

1. Participation
   (a) ±6000 requests
   (b) 600 participants \ experiment

2. Resulting Networks:
   (a) Sensity $\frac{2709}{22350} = 0.121$
   (b) Weight values: $\bar{x} = 3.07$, max = 80
   (c) Clustered structure
   (d) Wide range of generated connections
### Network Matrix

<table>
<thead>
<tr>
<th>number</th>
<th>label</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>act</td>
<td>0</td>
<td>8.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.5</td>
<td>0</td>
<td>0.5</td>
</tr>
<tr>
<td>(2)</td>
<td>action</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(3)</td>
<td>age</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.5</td>
</tr>
<tr>
<td>(4)</td>
<td>amount</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.5</td>
<td>0</td>
</tr>
<tr>
<td>(5)</td>
<td>area</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(6)</td>
<td>art</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.5</td>
</tr>
<tr>
<td>(7)</td>
<td>attention</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(8)</td>
<td>authority</td>
<td>0.5</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(9)</td>
<td>bed</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(10)</td>
<td>blood</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

**Table 1:** Sample of matrix representing final structure of adaptive hypertext network trained in experiment 1
Centrality

Un-normalized Centrality node $i$:

$$c_i = \frac{\sum_{j=0}^{n} w_{ij} + \sum_{j=0}^{n} w_{ji}}{2}$$  \hspace{1cm} (1)

10 most central nodes:

<table>
<thead>
<tr>
<th>label</th>
<th>centrality</th>
</tr>
</thead>
<tbody>
<tr>
<td>knowledge</td>
<td>416.75</td>
</tr>
<tr>
<td>experience</td>
<td>324.75</td>
</tr>
<tr>
<td>development</td>
<td>261</td>
</tr>
<tr>
<td>mind</td>
<td>218.25</td>
</tr>
<tr>
<td>education</td>
<td>210.25</td>
</tr>
<tr>
<td>method</td>
<td>204.5</td>
</tr>
<tr>
<td>thought</td>
<td>191.25</td>
</tr>
<tr>
<td>system</td>
<td>181</td>
</tr>
<tr>
<td>body</td>
<td>178.25</td>
</tr>
<tr>
<td>person</td>
<td>168</td>
</tr>
</tbody>
</table>
Graph structure
# Cluster Analysis

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Documents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>Age, time, century, day, evening, moment, period, week, year</td>
</tr>
<tr>
<td>Space</td>
<td>Place, area, point, stage</td>
</tr>
<tr>
<td>Movement</td>
<td>Action, change, movement, road, car</td>
</tr>
<tr>
<td>Control</td>
<td>Authority, control, power, influence</td>
</tr>
<tr>
<td>Cognition</td>
<td>Knowledge, fact, idea, thought, interest, book, course, development,</td>
</tr>
<tr>
<td></td>
<td>doubt, education, example, experience, language, mind, name, word,</td>
</tr>
<tr>
<td></td>
<td>problem, question, reason, research, result, school, side, situation, story,</td>
</tr>
<tr>
<td></td>
<td>theory, training, use, voice</td>
</tr>
<tr>
<td>Intimacy</td>
<td>Love, family, house, peace, father, friend, girl, hand, body, face, head,</td>
</tr>
<tr>
<td></td>
<td>figure, heart, church, kind, mother, woman, music, bed, wife</td>
</tr>
<tr>
<td>Vitality</td>
<td>Boy, man, wife, health</td>
</tr>
<tr>
<td>Society</td>
<td>Society, state, town, commonwealth</td>
</tr>
<tr>
<td>Office</td>
<td>Building, office, work, room</td>
</tr>
</tbody>
</table>
# Hyperlink Development

<table>
<thead>
<tr>
<th>Rank</th>
<th>Control</th>
<th>Authority</th>
<th>Authority</th>
<th>Authority</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Country</td>
<td>Authority</td>
<td>Authority</td>
<td>Authority</td>
<td>Power</td>
</tr>
<tr>
<td>2</td>
<td>Thing</td>
<td>Society</td>
<td>Rate</td>
<td>Power</td>
<td>Authority</td>
</tr>
<tr>
<td>3</td>
<td>Government</td>
<td>Rate</td>
<td>Society</td>
<td>Effect</td>
<td>Effect</td>
</tr>
<tr>
<td>4</td>
<td>Policy</td>
<td>Theory</td>
<td>Theory</td>
<td>Knowledge</td>
<td>Influence</td>
</tr>
<tr>
<td>5</td>
<td>Film</td>
<td>Problem</td>
<td>Problem</td>
<td>Government</td>
<td>Government</td>
</tr>
<tr>
<td>6</td>
<td>Heart</td>
<td>Job</td>
<td>Government</td>
<td>Society</td>
<td>Law</td>
</tr>
<tr>
<td>7</td>
<td>Group</td>
<td>Service</td>
<td>Job</td>
<td>Money</td>
<td>Knowledge</td>
</tr>
<tr>
<td>8</td>
<td>Room</td>
<td>Money</td>
<td>Money</td>
<td>Father</td>
<td>Society</td>
</tr>
<tr>
<td>9</td>
<td>Paper</td>
<td>Knowledge</td>
<td>Effect</td>
<td>Law</td>
<td>Order</td>
</tr>
<tr>
<td>10</td>
<td>Order</td>
<td>Day</td>
<td>Service</td>
<td>Rate</td>
<td>Money</td>
</tr>
</tbody>
</table>

Requests: 0 600 1200 2400 4800
Conclusion

1. **Link generation:**
   (a) Spontaneous proliferation of hyperlinks
   (b) Fast development of stable collection of links
   (c) Ranking reflects associative validity

2. **Network structure:**
   (a) Clustering of related items
   (b) In- and out-degree of nodes
   (c) Reflects level of abstraction

3. **Note:**
   (a) Link generation independent from text or format
   (b) No requirement for explicit user preference statements
   (c) Distribution of link weight: few retrievals required
Digital Library Applications

1. Article and Journal Linking
   (a) Citation or IR based
   (b) Static
   (c) Non-specific (User Community)

   (a) No intervention in actual library systems
   (b) Registration of User Requests: server logs
   (c) Reconstruction of User Paths

3. Large Number of Articles and Journals
   (a) Heterogeneous: Validity Atomic Concept representation?
   (b) Scalability?
Methodology

1. Reconstruction of co-retrieval events from DL Server logs
   (a) Use Sequence of Retrieval Requests to Derive Article and Journal Associations
   (b) Integrate access Patterns for Large Range of Applications, Users, etc.

2. Generate Associative Network of Journals
   (a) Every Article appears in a journal
   (b) Journal Relations subsume Article Relations

3. Use Network
   (a) Generate document links without resorting to text-content based IR
   (b) Map Preferences of LANL User Community
   (c) Recommendations Systems that Operate on Network Structure
Test Application

1. Los Alamos National Laboratory Research Library:
   (a) Large digital library with focus on management of own material
   (b) User community: focus on physics and nuclear science
   (c) SciSearch database: +20,000,000 articles
   (d) Downloadable: PDF format after search and metadata retrieval

2. Availability of server logs from 1998 to present
   (a) Download rates: 10,000/mo.
   (b) Server logs used: 2001
   (c) Log file used: 5 months, 40,847 requests, 1829 journals and 20,720 articles
Interface
Implicit learning: news from the front

Axel Cleermans, Arnaud Desrebecq and Maud Boyer

Can we learn without awareness? While the current consensus is most likely to be 'yes', there is, however, considerable ongoing debate about the role that consciousness plays in cognition and about the nature of consciousness itself. In this article, we review recent advances in the field of implicit learning, based on three perspectives: empirical findings (including neuropsychological evidence), methodological issues, and theoretical positions (including computational models). The overall picture that emerges is complex and reflects a field that is very much in flux; while it seems undeniable that cognition involves some form of unconscious processing, it is as yet unclear how to best separate conscious and unconscious influences on learning, and how to best think about the status of the 'cognitive unconscious'. We suggest that implicit learning is best construed as a complex form of priming taking place in continuously learning neural systems, and that the distributional knowledge so acquired can be causally efficacious in the absence of awareness that this knowledge was acquired or that it is currently influencing processing, that is, in the absence of metaknowledge.

Implicit learning (IL) - broadly construed, the ability to learn without awareness - has been under investigation for definitions of IL, learning is implicit when we acquire new information without intending to do so, and it is a way

Interface
Network Generation

Set of co-retrieval events: $E = \{e_1, e_2, \cdots, e_k\}$

Co-retrieval event $e_i$: $e_i = (v_i, v_j, t(v_i, v_j))$

where $t(v_i, v_j)$ represent time in seconds elapsed retrieval requests

for document $v_i$ and $v_j$

$e_i = (v_i, v_j, t(v_i, v_j)) \sim r$, $r = f(e_i)$ reinforcement value so that:

$m_{ij} += r$

$\forall_{i,j} m_{ij} = 0$

for (i=1; i<n+1; i++) {
    $e_i = (v_i, v_j, t(v_i, v_j))$: $m_{ij} += f(e_i)$
}


Retrieval Events
### SciSearch Retrieval Log

<table>
<thead>
<tr>
<th>userID</th>
<th>data/time</th>
<th>Document ID</th>
<th>Service</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>24/Aug/2001:17:12:52</td>
<td>02721716;14;4;69_ddacfrtip</td>
<td>SciServer</td>
</tr>
<tr>
<td>100</td>
<td>24/Aug/2001:17:14:41</td>
<td>01689274;25;4;499_prtuauma</td>
<td>SciServer</td>
</tr>
<tr>
<td>100</td>
<td>24/Aug/2001:17:15:43</td>
<td>00978493;19;2;281_apiaaaortars</td>
<td>SciServer</td>
</tr>
<tr>
<td>101</td>
<td>18/Jun/2001:12:03:04</td>
<td>00207225;38;3;347_otfrim</td>
<td>SciServer</td>
</tr>
<tr>
<td>101</td>
<td>18/Jun/2001:12:04:40</td>
<td>02780062;19;3;211_aotdfmfct</td>
<td>SciServer</td>
</tr>
<tr>
<td>101</td>
<td>18/Jun/2001:13:13:40</td>
<td>08956111;25;2;113_asrgt3tr</td>
<td>SciServer</td>
</tr>
</tbody>
</table>

Table 2: An example of the transformed Science Server log used to reconstruct journal and document co-retrieval events.
SciSearch Retrieval Log Analysis

<table>
<thead>
<tr>
<th></th>
<th>100</th>
<th>101</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(0272-1716, 0168-9274)</td>
<td>(0168-9274, 0097-8493)</td>
</tr>
<tr>
<td></td>
<td>(0020-7225, 0278-0062)</td>
<td>(0278-0062, 0895-6111)</td>
</tr>
</tbody>
</table>

Table 3: Co-retrieval events derived from set of user retrieval requests in table 2.
Analysis application
## Generated Network

<table>
<thead>
<tr>
<th>Journal</th>
<th>01</th>
<th>02</th>
<th>03</th>
<th>04</th>
<th>05</th>
<th>06</th>
<th>07</th>
<th>08</th>
<th>09</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADV COLLOID INTERFAC</td>
<td>01</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ADVAN MATH</td>
<td>02</td>
<td>0</td>
<td>0</td>
<td>1.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1.5</td>
</tr>
<tr>
<td>ANALYT CHIM</td>
<td>03</td>
<td>0</td>
<td>0.3</td>
<td>0</td>
<td>6.9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>ANALYT BIOC</td>
<td>04</td>
<td>0</td>
<td>0</td>
<td>19.6</td>
<td>0</td>
<td>0</td>
<td>0.6</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>ANIMAL BEHAVIOUR</td>
<td>05</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ANN PHYS-NEW YORK</td>
<td>06</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.6</td>
</tr>
<tr>
<td>APPLIED ACOUSTICS</td>
<td>07</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>APPL PHYS LETT</td>
<td>08</td>
<td>0</td>
<td>0</td>
<td>0.3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ARCH RATION MECH AN</td>
<td>09</td>
<td>0</td>
<td>0.3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ARCH BIOCHEM BIOPHYS</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>5.5</td>
<td>10.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ARTIF INTELL</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ASTRONOM J</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>PUBL ASTRON SOC JPN</td>
<td>13</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ASTRON ASTROPHYS</td>
<td>14</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ASTROPHYSICAL JOURNAL</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Generated Network

1. density 0.176%

2. mean link weight = 1.196, standard deviation=0.821 for all \( m_{ij} : m_{ij} > 0 \)
Network Analysis and Comparison to Impact Factor

**Structural Features**

1. Highest link weights
2. Graph-theoretical network characteristics compared to Impact Factor
   (a) Impact Factor: based on citation frequency (ratio of citations over publications for last 2 years)
   (b) Journal Consultation Frequency:
   \[ F(v_i) = \sum_{j=1}^{n} m_{ij} + \sum_{j=1}^{n} m_{ji} \]
   (c) Impact Factor Discrepancy Ratio:
   \[ r_f(i) = \frac{JCF(i)}{IF(i)} \]
## Highest Link Weights

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Start node</th>
<th>End Node</th>
</tr>
</thead>
<tbody>
<tr>
<td>22.0</td>
<td>PHYSICA C</td>
<td>PHYSICA B</td>
</tr>
<tr>
<td>18.0</td>
<td>PHYSICA B</td>
<td>PHYSICA C</td>
</tr>
<tr>
<td>14.0</td>
<td>NUCL INSTRUM METH A</td>
<td>IEEE T NUCL SCI</td>
</tr>
<tr>
<td>12.0</td>
<td>PHYSICA B</td>
<td>J MAGN MAGN MATER</td>
</tr>
<tr>
<td>11.0</td>
<td>J MAGN MAGN MATER</td>
<td>PHYSICA B</td>
</tr>
</tbody>
</table>

Table 4: Five pairs of journals for which highest co-retrieval frequencies have been found.
JCF vs. 2000 IF

<table>
<thead>
<tr>
<th>Frequency</th>
<th>IF</th>
<th>Journal Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>176.00</td>
<td>0.893</td>
<td>PHYSICA B</td>
</tr>
<tr>
<td>134.00</td>
<td>0.964</td>
<td>NUCL INSTRUM METH A</td>
</tr>
<tr>
<td>132.00</td>
<td>1.576</td>
<td>APPL CATAL A-GEN</td>
</tr>
<tr>
<td>112.00</td>
<td>0.897</td>
<td>MAT SCI ENG A-STRUCT</td>
</tr>
<tr>
<td>107.00</td>
<td>0.845</td>
<td>J ALLOY COMPD</td>
</tr>
<tr>
<td>106.00</td>
<td>1.060</td>
<td>IEEE T NUCL SCI</td>
</tr>
<tr>
<td>104.00</td>
<td>1.933</td>
<td>CATAL TODAY</td>
</tr>
<tr>
<td>96.00</td>
<td>3.030</td>
<td>J CATAL</td>
</tr>
<tr>
<td>93.00</td>
<td>2.198</td>
<td>SURF SCI</td>
</tr>
<tr>
<td>92.00</td>
<td>1.241</td>
<td>J NUCL MATER</td>
</tr>
</tbody>
</table>

Table 5: List of 10 most frequent journals in LANL RL Science Server network and 2000 Impact Factors.
Scatterplot JCF vs. ISI IF (2000)

$r_s = 0.13 \ (p < 0.05)$
## Impact Factor Discrepancy Ratio

<table>
<thead>
<tr>
<th>$r_f$</th>
<th>JCF</th>
<th>IF</th>
<th>Journal Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>232.80423</td>
<td>44.00</td>
<td>0.189</td>
<td>ASTROPHYS SPACE SCI</td>
</tr>
<tr>
<td>199.00497</td>
<td>40.00</td>
<td>0.201</td>
<td>NUCL ENG DES</td>
</tr>
<tr>
<td>197.08847</td>
<td>176.00</td>
<td>0.893</td>
<td>PHYSICA B</td>
</tr>
<tr>
<td>189.87341</td>
<td>45.00</td>
<td>0.237</td>
<td>J MATER PROCESS TECH</td>
</tr>
<tr>
<td>189.80170</td>
<td>67.00</td>
<td>0.353</td>
<td>ADV SPACE RES</td>
</tr>
<tr>
<td>172.04301</td>
<td>32.00</td>
<td>0.186</td>
<td>ACTA ASTRONAUT</td>
</tr>
<tr>
<td>153.68852</td>
<td>75.00</td>
<td>0.488</td>
<td>J RADIOANAL NUCL CH</td>
</tr>
<tr>
<td>143.93939</td>
<td>38.00</td>
<td>0.264</td>
<td>PROG NUCL ENERG</td>
</tr>
<tr>
<td>139.00415</td>
<td>134.00</td>
<td>0.964</td>
<td>NUCL INSTRUM METH A</td>
</tr>
<tr>
<td>128.67647</td>
<td>35.00</td>
<td>0.272</td>
<td>MATH COMPUT SIMULAT</td>
</tr>
</tbody>
</table>

Table 6: List of 10 journals for which highest ratio of JCF and ISI IF has been found.
Spreading Activation

1. Model Facilitated, Automatic Retrieval from human memory.
3. Concept Activation spreads through network, modulated by connection weights.
Spreading Activation for Information Retrieval

1. Promises
   (a) Text and Language Independent
   (b) Cue-based retrieval or retrieval by example
   (c) Use of Network structure: incomplete structure can still lead to complete retrieval

2. Problems
   (a) Establishment of Document Networks
   (b) Questions regarding efficiency (Salton, 1988)
   (c) Main Application in Concept Networks, Query Expansion
Spreading Activation on Journal Titles

1. Features
   (a) Transformation of initial keyterm query to network initial activation values
   (b) Recommendation Refinement by “Query by example”
   (c) Network interface (Servlet), marked-up output

2. Implementation
   (a) Java Servlet

   i. Third-party service for LANL DL

   (b) Packed Query

   (c) Link to LANL RL journal pages
      i. Where available
      ii. Added link to LinkSeeker service

3. Prototype:
Implementation

Servlet

Client

Modules
- Query to Activation Vector
- Title Matcher
- Data Parser
- HTML Conversion

Data
- List of Titles and Labels
- Associative Data

Client

Query
Parameters
Refinement Query

Result Page
Use the following form to generate SpreadIt journal recommendations.
Input keywords, ISSN numbers (remove dash) or full journal titles (replace spaces by _).
The parameters have been set for optimal results.
Extend the generated list by clicking on the little SpreadIt icon. Retrieve the recommended journal itself by clicking on its hyperlinked title.

New Query

Submit Query

Window: 12
Iterations: 6
Applet threshold: 0.04
Retention: 0.05
Applet: 1
Query Expansion

```
SpRedit

Recommendations for:
- TECTONOPHYSICS
- earth

Direct Matches
- EARTH AND PLANETARY SCIENCE LETTERS
- EARTH SCIENCE REVIEWS
- PHYSICS OF THE EARTH AND PLANETARY INTERIORS
- TECTONOPHYSICS

Recommendations
Order Spread:
0  EARTH AND PLANETARY SCIENCE LETTERS
1  JOURNAL OF VOLCANOLOGY AND GEOTHERMAL RESEARCH
2  JOURNAL OF ATMOSPHERIC AND SOLAR TERRESTRIAL PHYSICS
3  CONTRIBUTIONS TO MINERALOGY AND PETROLOGY
4  BULLETIN OF VOLCANOLOGY
5  CHEMICAL GEOLOGY
6  MUTATION RESEARCH ENVIRONMENTAL MUTAGENESIS AND RELATED SUBJECTS
7  GEOCHIMICA ET COSMOCHIMICA ACTA
8  INTERNATIONAL JOURNAL OF CLINICAL PRACTICE
9  APPLIED GEOCHEMISTRY
10  ASTRONOMICAL JOURNAL
11  FUEL

New Query: 
Submit Query
```
Dynamic Network representation

Design and implemented by
Johan Bollen and Herbert Veldakowski
Conclusion

1. What have we got?
   (a) Methodology to link documents across formats, content and languages.
   (b) Method to analyse networks and obtain information on preferences and structure of user community
   (c) Generation of alternatives for citation based impact factors
   (d) Spreading Activation Retrieval

2. Issues:
   (a) Applications to other information resources
   (b) Scalability
   (c) Sufficient server logs for network training
   (d) Hebbian learning assumption (invalid labeling, user relevance)