NCDawareRank
A Novel Ranking Method that Exploits the Decomposable Structure of the Web

Athanasios N. Nikolakopoulos  John D. Garofalakis

Computer Engineering and Informatics Department, University of Patras
Computer Technology Institute and Press “Diophantus”
PageRank Model:

\[ G = \alpha H + (1 - \alpha)E \]

The **Damping Factor Issue**:

- Controls the fraction of importance, propagated through the links.
- The choice of \( \alpha \) has received much attention
  - Picking very small \( \alpha \) \( \Rightarrow \) Uninformative Ranking Vector
  - Picking \( \alpha \) close to 1 \( \Rightarrow \) Computational Problems, Counterintuitive Ranking

We focus on the **Teleportation model** itself!
Enriching the Teleportation Model

Web as a **Nearly Completely Decomposable** System:
- Nested Block Structure
  - Hierarchical Nature $\implies$ **NCD Architecture**
- NCD has been exploited **Computationally**.
- We aim to exploit it **Qualitatively** in order to **Generalize the Teleportation Model**
  - Multiple Levels of Proximity between Nodes
  - **Core Idea**: Direct importance propagation to the NCD blocks that contain the outgoing links.
We partition the Web into **NCD blocks**, \( \{A_1, A_2, \ldots, A_N\} \),

- For every page \( u \) we define \( X_u \) to be its **proximal set** of pages, i.e. the union of the NCD blocks that contain \( u \) and the pages it links to.

- We introduce an **Inter-Level Proximity Matrix** \( M \), designed to propagate a fraction of importance to the proximal set of each page. Matrix \( M \) can be expressed as a product of 2 extremely sparse matrices, \( R \in \mathbb{R}^{n \times N} \) and \( A \in \mathbb{R}^{N \times n} \),

\[
\begin{align*}
\eta \mathbf{H} + \mu \mathbf{M} + (1 - \eta - \mu) \mathbf{E} \\
\mathbf{H} &= [H_{uv}] \triangleq \frac{1}{d_u}, \quad \text{if } v \in \mathcal{G}_u \\
\mathbf{M} &= [M_{uv}] \triangleq \frac{1}{N_u|\mathcal{A}(v)|}, \quad \text{if } v \in \mathcal{X}_u \\
\text{where } \mathcal{X}_u &\triangleq \bigcup_{w \in (u \cup \mathcal{G}_u)} \mathcal{A}(w) \\
\mathbf{E} &= \mathbf{e}v^T
\end{align*}
\]
Theorem (Convergence Rate Bound:)

The subdominant eigenvalue of matrix $P$ involved in the NCDawareRank, is upper bounded by $\eta + \mu$.

Computational Experiments:

<table>
<thead>
<tr>
<th></th>
<th>PageRank</th>
<th>NCDawareRank</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\alpha = 0.85$</td>
<td>$\mu = 0.005$</td>
</tr>
<tr>
<td>cnr-2000</td>
<td>48</td>
<td>47</td>
</tr>
<tr>
<td>eu-2005</td>
<td>42</td>
<td>42</td>
</tr>
<tr>
<td>india-2004</td>
<td>48</td>
<td>47</td>
</tr>
<tr>
<td>indochnina-2004</td>
<td>47</td>
<td>46</td>
</tr>
<tr>
<td>uk-2002</td>
<td>46</td>
<td>45</td>
</tr>
</tbody>
</table>
Experimental Evaluation

Newly Added Pages Bias Problem:

- Methodology:
  - Extract the 90% of the incoming links of a set of randomly chosen pages.
  - Compare the orderings against those induced by the complete graph.

<table>
<thead>
<tr>
<th># New Pages</th>
<th>8000</th>
<th>10000</th>
<th>12000</th>
<th>15000</th>
<th>20000</th>
<th>30000</th>
</tr>
</thead>
<tbody>
<tr>
<td>HyperRank</td>
<td>94.51±0.22</td>
<td>93.26±0.19</td>
<td>92.96±0.21</td>
<td>90.37±0.30</td>
<td>87.72±0.28</td>
<td>82.34±0.30</td>
</tr>
<tr>
<td>LinearRank</td>
<td>93.80±0.48</td>
<td>92.60±0.24</td>
<td>91.23±0.28</td>
<td>89.41±0.47</td>
<td>86.56±0.44</td>
<td>80.69±0.49</td>
</tr>
<tr>
<td>NCDawareRank</td>
<td>96.81±1.06</td>
<td>96.48±1.10</td>
<td>96.64±0.42</td>
<td>95.44±1.39</td>
<td>94.77±0.72</td>
<td>91.49±1.42</td>
</tr>
<tr>
<td>PageRank</td>
<td>93.68±0.59</td>
<td>92.46±0.30</td>
<td>91.04±0.37</td>
<td>89.19±0.55</td>
<td>86.33±0.53</td>
<td>80.26±0.57</td>
</tr>
<tr>
<td>RAPr</td>
<td>94.16±0.37</td>
<td>92.96±0.20</td>
<td>91.64±0.23</td>
<td>89.87±0.49</td>
<td>87.15±0.41</td>
<td>81.47±0.41</td>
</tr>
<tr>
<td>TotalRank</td>
<td>94.15±0.38</td>
<td>92.94±0.21</td>
<td>91.62±0.25</td>
<td>89.84±0.51</td>
<td>87.12±0.43</td>
<td>81.37±0.44</td>
</tr>
</tbody>
</table>

Sparsity:

- Methodology:
  - Randomly select to include 90% – 40% of the links on a new “sparsified” version of the graph
  - Compare the rankings of the algorithms against their corresponding original rankings.

Fig 1. Ranking Stability under Sparseness.
Resistance to Direct Manipulation:

- **Methodology:**
  - Randomly pick a node with small initial ranking and we add a number of \( n \) nodes that funnel all their rank towards it.
  - We run all the algorithms for different values of \( n \) and we compare the spamming node’s rank.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Number of Added Nodes</th>
<th>Spamming Node’s Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>HyperRank</td>
<td>0</td>
<td>0.005</td>
</tr>
<tr>
<td>LinearRank</td>
<td>1000</td>
<td>0.01</td>
</tr>
<tr>
<td>NCDawareRank</td>
<td>2000</td>
<td>0.005</td>
</tr>
<tr>
<td>PageRank</td>
<td>3000</td>
<td>0.005</td>
</tr>
<tr>
<td>RAPr</td>
<td>4000</td>
<td>0.005</td>
</tr>
<tr>
<td>TotalRank</td>
<td>5000</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>6000</td>
<td>0.005</td>
</tr>
</tbody>
</table>

- Test conditions:
  - \( \eta = 0.95, \mu = 0 \)
  - \( \eta / \mu = 5 \)
  - \( \eta / \mu = 1 \)
  - \( \eta / \mu = 1/5 \)
  - \( \eta / \mu = 1/10 \)
  - \( \eta / \mu = 1/30 \)
Conclusions and Future Research

We propose **NCDawareRank**:

- Generalizes PageRank by Enriching the Teleportation Model
- Produces More Stable Ranking Vectors
  - Sparseness Insensitivity
  - Resistance to Manipulation
- Opens new interesting research directions
Thanks!

Q&A