Motivation

Link-based ranking methods like PageRank fail for information needs like the following, since the history of the (e.g., web) graph overshadows its recent evolution.

1. **Highest rated Movies to watch (in cinema)**
   - **What we expect:** Mission Impossible III
   - **What we get:** The Godfather

2. **Top-importance-gaining publications in database research**

Idea

Importance (e.g., PageRank) scores of a node $v$ co-evolve with the underlying graph thus forming a **time series** with importance scores $r_i$ at observation times $t_i$:

$$
\{(t_0, r_0), \ldots, (t_n, r_n)\}
$$

BuzzRank analyzes such **time series of importance scores** and **quantifies contained trends** based on a **growth model** of importance scores.

Thus, BuzzRank identifies items that increased their importance significantly in a period of interest, or, caused significant **BUZZ**.

Use Case - Identify Database Research Papers of Interest

Assume a user wants to identify recent publications in database research that are **becoming** important.

**PageRank**

The top-ranked publication up to '00 – “A Relational Model of Data for Large Shared Data Banks” by E. F. Codd (published 1970).

But..., Agrawal et al.’s “Mining Association Rules between Sets of Items in Large Databases” would be a better result if the given information need arises at any time between ‘93 and ‘00.

As can be seen from the above figure, the association rules paper gained relatively more importance at any point since ’93.

BuzzRank identifies this and ranks the paper by Agrawal et al., ahead of its opponent.
BuzzRank

BuzzRank quantifies the buzz created within a time-interval of interest \([t_{\text{begin}}, t_{\text{end}}]\).

Method

The growth of importance scores is modeled by the following generic growth model with \(\alpha_v(t)\) being the rate of importance growth of node \(v\) at time \(t\).

\[
\hat{r}(v, t) = e^{\int_0^t \alpha_v(t)dt}
\]

We assume that the growth rate is time-invariant within the time-interval \([t_{\text{begin}}, t_{\text{end}}]\) giving us the following simplified model with parameters \(\alpha_v\) and \(A_{v, \text{begin}}\).

\[
\hat{r}(v, t) = A_{v, \text{begin}} e^{\alpha_v(t-t_{\text{begin}})} \quad : \quad t_{\text{begin}} \leq t \leq t_{\text{end}}
\]

Optimal parameter values \(\alpha^*_v\) and \(A^*_{v, \text{begin}}\) are estimated using the method of least squares, i.e., minimizing

\[
\sum_{t_{\text{begin}} \leq t \leq t_{\text{end}}} (r(v, t_i) - \hat{r}(v, t_i))^2
\]

Applying a log-transformation to the observed importance scores, this is equivalent to fitting a straight line, so that closed-form solutions exist.

Finally, BuzzRank ranks nodes based on the parameter \(\alpha^*_v\) which is an estimate of the growth rate of a node’s importance in the time-interval \([t_{\text{begin}}, t_{\text{end}}]\).

PageRank Normalization

We use PageRank to assess importance as an input to BuzzRank. Plain PageRank scores, however, are not comparable across graphs.

Therefore, we normalize PageRank scores computed on \(G_t(V_t, E_t)\) (i.e., the graph at time \(t\)) dividing by the lower bound PageRank score that would be assigned to a node without incoming edges

\[
r_{\text{low}, t} = \frac{1}{|V_t|} \left(\epsilon + (1-\epsilon) \sum_{d \in D_t} r_t(d)\right)
\]

It can be shown that the normalized score depends only on the node’s reachability but not on the graph size or the number of dangling nodes \(D_t\).
Experiments

Experiments were conducted on bibliographic DBLP dataset.

**Input:** PageRank rankings computed for years ’89–’99  
**Output:** BuzzRank rankings for two year time intervals [t, t+1]

### Top Buzzing Publications

<table>
<thead>
<tr>
<th>Years</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>[‘89,’90]</td>
<td>The Object-Oriented Database System Manifesto</td>
</tr>
<tr>
<td>[‘90,’91]</td>
<td>CYC: Toward Programs With Common Sense</td>
</tr>
<tr>
<td>[‘92,’93]</td>
<td>Simplifying Decision Trees</td>
</tr>
<tr>
<td>[‘93,’94]</td>
<td>World-Wide Web: The Information Universe</td>
</tr>
<tr>
<td>[‘94,’95]</td>
<td>The Power of Languages for the Manipulation of Complex Values</td>
</tr>
<tr>
<td>[‘96,’97]</td>
<td>Implementing Data Cubes Efficiently</td>
</tr>
<tr>
<td>[‘97,’98]</td>
<td>Modeling Multidimensional Databases</td>
</tr>
<tr>
<td>[‘98,’99]</td>
<td>XML-QL: A Query Language for XML</td>
</tr>
</tbody>
</table>

### Conclusions and Future Work

**BuzzRank…**  
- identifies “hot” authoritative items in given time period  
- is based on time series of precomputed PageRank scores  
- is complementary to PageRank

**In the Future…**  
- Experiments on a variety of datasets: Wikipedia, Web graph,…  
- Extension of the underlying model (e.g., time-varying growth rate)  
- Improvement of scalability