Edge-Weighted Personalized PageRank: Breaking a Decade-Old Performance Barrier

W. Xie  D. Bindel  A. Demers  J. Gehrke

12 Aug 2015
PageRank Model

- Random surfer model: \( x^{(t+1)} = \alpha P x^{(t)} + (1 - \alpha) v \) where \( P = AD^{-1} \)
- Stationary distribution: \( Mx = b \) where \( M = (I - \alpha P), b = (1 - \alpha) v \)
Introduce *personalization parameters* $w \in \mathbb{R}^d$ in two ways:

- **Node weights**: $M x(w) = b(w)$
- **Edge weights**: $M(w) x(w) = b$
Node weight personalization is well-studied

- Topic-sensitive PageRank: fast methods based on linearity
- Localized PageRank: fast methods based on sparsity

Some work on edge weight personalization

- ObjectRank/ScaleRank: personalize weights for different edge types
- But lots of work incorporates edge weights *without* personalization

**Our goal:** General, fast methods for edge weight personalization
Model Reduction

Expensive full model $(Mx = b)$

$\approx$

Reduced basis

$\tilde{M}y = \tilde{b}$

Reduced model

Approximation ansatz

Model reduction procedure from physical simulation world:

- **Offline**: Construct reduced basis $U \in \mathbb{R}^{n \times k}$
- **Offline**: Choose $\geq k$ equations to pick approximation $\hat{x} = Uy$
- **Online**: Solve for $y(w)$ given $w$ and reconstruct $\hat{x}$
Reduced Basis Construction: SVD (aka POD/PCA/KL)

\[ \approx U \Sigma V^T \]

Snapshot matrix

\[ x_1 \ x_2 \ \cdots \ x_r \]

\[ w_1 \ w_2 \ \cdots \ w_r \]

Sample points

W. Xie, D. Bindel, A. Demers, J. Gehrke
Approximation Ansatz

Want $r = M U y - b \approx 0$. Consider two approximation conditions:

<table>
<thead>
<tr>
<th>Method</th>
<th>Ansatz</th>
<th>Properties</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bubnov-Galerkin</td>
<td>$U^T r = 0$</td>
<td>Good accuracy empirically</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fast for $P(w)$ linear</td>
</tr>
<tr>
<td>DEIM</td>
<td>$\min | r_I |$</td>
<td>Fast even for nonlinear $P(w)$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Complex cost/accuracy tradeoff</td>
</tr>
</tbody>
</table>

Similar error analysis framework for both (see paper):

Consistency + Stability = Accuracy

- Consistency: Does the subspace contain good approximants?
- Stability: Is the approximation subproblem far from singular?
Bubnov-Galerkin Method

\[ M \begin{bmatrix} U^T & y \end{bmatrix} - b = 0. \]

- Linear case: \( w_i \) = probability of transition with edge type \( i \)

\[
M(w) = I - \alpha \left( \sum_i w_i P^{(i)} \right), \quad \tilde{M}(w) = I - \alpha \left( \sum_i w_i \tilde{P}^{(i)} \right)
\]

where we can precompute \( \tilde{P}^{(i)} = U^T P^{(i)} U \)

- Nonlinear: Cost to form \( \tilde{M}(w) \) comparable to cost of PageRank!
Discrete Empirical Interpolation Method (DEIM)

Equations in $\mathcal{I}$

\[ M - b = 0. \]

- Ansatz: Minimize $\| r_\mathcal{I} \|$ for chosen indices $\mathcal{I}$
- Only need a few rows of $M$ (and associated rows of $U$)
- Difference from physics applications: high-degree nodes!
Interpolation Costs

Consider subgraph relevant to one interpolation equation:

- Really care about weights of edges incident on $I$
  - Need more edges to normalize (unless $A(w)$ is linear)
- High in/out degree are expensive but informative
- **Key question**: how to choose $I$ to balance **cost** vs **accuracy**?
Interpolation Accuracy

- Key: keep $M_{\mathcal{I}}$, far from singular.
- If $|\mathcal{I}| = k$, this is a subset selection over rows of $MU$.
- Have standard techniques (e.g. pivoted QR)
- Want to pick $\mathcal{I}$ once, so look at rows of

$$Z = [M(w_1)U \quad M(w_2)U \quad \ldots]$$

for sample parameters $w^{(i)}$.
- Helps to explicitly enforce $\sum_i \hat{x}_i = 1$
- Several heuristics for cost/accuracy tradeoff (see paper)
Online Costs

If $\ell = \# \text{PR components needed}$, online costs are:

- Form $\tilde{M}$: $O(dk^2)$ for B-G
- More complex for DEIM
- Factor $\tilde{M}$: $O(k^3)$
- Solve for $y$: $O(k^2)$
- Form $Uy$: $O(k\ell)$

Online costs do not depend on graph size!
(unless you want the whole PR vector)
Example Networks

DBLP (citation network)
- 3.5M nodes / 18.5M edges
- Seven edge types $\Rightarrow$ seven parameters
- $P(w)$ linear
- Competition: ScaleRank

Weibo (micro-blogging)
- 1.9M nodes / 50.7M edges
- Weight edges by topical similarity of posts
- Number of parameters $=$ number of topics (5, 10, 20)

(Studied global and local PageRank – see paper for latter.)
Singular Value Decay

\[ 10^{-1}, 10^0, 10^1, 10^2, 10^3, 10^4, 10^5, 10^6 \]

Value

\[ 0, 50, 100, 150, 200 \]

\( i^{th} \) Largest Singular Value

DBLP-L
Weibo-S5
Weibo-S10
Weibo-S20

\( r = 1000 \) samples, \( k = 100 \)
DBLP Accuracy

![Bar chart showing accuracy metrics for different methods: Galerkin, DEIM-100, DEIM-120, DEIM-200, and ScaleRank. The x-axis represents the methods, and the y-axis shows accuracy on a logarithmic scale.]
W. Xie, D. Bindel, A. Demers, J. Gehrke

KDD2015

12 Aug 2015
Weibo Accuracy

![Bar graph showing Kendall@100 and Normalized L1 accuracy for different numbers of parameters (5, 10, 20). The x-axis represents the number of parameters, and the y-axis represents the accuracy on a logarithmic scale. The graph shows an increase in accuracy as the number of parameters increases.]

- **Kendall@100**
  - 5 parameters: $10^{-4}$
  - 10 parameters: $10^{-3}$
  - 20 parameters: $10^{-2}$

- **Normalized L1**
  - 5 parameters: $10^{-4}$
  - 10 parameters: $10^{-3}$
  - 20 parameters: $10^{-2}$

Authors: W. Xie, D. Bindel, A. Demers, J. Gehrke

KDD2015, 12 Aug 2015
Weibo Running Times (All Nodes)

![Bar Chart]

**Coefficients**
- 0
- 0.1
- 0.2
- 0.3
- 0.4
- 0.5

**Construction**
- 5
- 10
- 20

**Running time (s)**

W. Xie, D. Bindel, A. Demers, J. Gehrke
KDD2015 12 Aug 2015
Application: Learning to Rank

Goal: Given $T = \{(i_q, j_q)\}_{q=1}^{T}$, find $w$ that mostly ranks $i_q$ over $j_1$.
(c.f. Backstrom and Leskovec, WSDM 2011)

- Standard: Gradient descent on full problem
  - One PR computation for objective
  - One PR computation for each gradient component
  - Costs $d + 1$ PR computations per step

- With model reduction
  - Rephrase objective in reduced coordinate space
  - Use factorization to solve PR for objective
  - Re-use same factorization for gradient
DBLP Learning Task

(8 papers for training + 7 params)
Test case: DBLP, 3.5M nodes, 18.5M edges, 7 params

Cost per Iteration:

<table>
<thead>
<tr>
<th>Method</th>
<th>Standard</th>
<th>Bubnov-Galerkin</th>
<th>DEIM-200</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time (sec)</td>
<td>159.3</td>
<td>0.002</td>
<td>0.033</td>
</tr>
</tbody>
</table>
In the paper (but not the talk)

- Selecting interpolation equations for DEIM
- Localized PageRank experiments (Weibo and DBLP)
- Comparison to BCA for localized PageRank
- Quasi-optimality framework for error analysis

Room for future work! Analysis, applications, systems, ...
Questions?

Edge-Weighted Personalized PageRank: Breaking a Decade-Old Performance Barrier

Wenlei Xie, David Bindel, Johannes Gehrke, and Al Demers

KDD 2015, paper 117

Sponsors:

- NSF (IIS-0911036 and IIS-1012593)
- iAd Project from the National Research Council of Norway