Clustering Sparse Matrices with Information from Both Numerical Values and Pattern

Daniel Ruiz\(^1\)
with I. Duff, P.A. Knight, S. Mouysset, D. Ruiz, B. Uçar

\(^1\)University of Toulouse

17th SIAM Conference on Parallel Processing for Scientific Computing
Paris, April 2016
Motivation

- How do we balance size and structure of nonzeros in a matrix to draw out key features?
- Can we make effective use of spectral information?
Motivation

- Preprocessing for large sparse matrices
- Linear algebra
- Community detection

- Genomics: biclustering on microarray data
Doubly stochastic scaling

- Assume that $A \in \mathbb{R}^{n \times n}, A \geq 0$.
- We want to satisfy $DAFe = e$ and $FA^TDe = e$.
- Motivated by desire to find scaling such that $\kappa(DAF) \ll \kappa(A)$ (Bauer, 1963).
- Fast balancing: Newton method
- $AA^T$ and $A^TA$ are also doubly stochastic
- Relies on both numerical values and pattern.
Existence and Uniqueness of Scaling

- $DAF$ doubly stochastic, $\text{diag}(D), \text{diag}(F) > 0$.
- Solution exists if $A$ contains sufficient nonzero entries.
- $A$ fully indecomposable.
- $A$ is symmetric and has positive main diagonal.
- Solution is unique (up to scaling).
Spectral Properties

- \( P = DAF \) with \( A \geq 0 \).
- We can extend Fiedler/Perron–Frobenius theory.
- If \( P \) has \( k \) disjoint components principal singular value is repeated \( k \) times.
- \( \sigma_1 = 1 \).
- Typical singular vector: permutation of 
  \[
  \begin{bmatrix}
    1 & \ldots & 1 & 0 & \ldots & 0 \\
  \end{bmatrix}^T.
  \]
- We can infer block entire block structure from a single singular vector.
- For symmetric matrices we can ensure \( D = F \) and work with eigenvectors.
**Algorithm**

1. Preprocess and balance.
2. Calculate singular vector(s).
3. Split vector(s) to identify blocks.
4. Quality measure and cluster amalgamation.

and we can iterate on the last three steps to define clusters recursively.
Preprocessing

- We want to avoid scaling a matrix if it is not fully indecomposable.
- Initialise by looking for BTF and work on biggest diagonal blocks.
- After roughly balancing matrix, remove strongly dominant parts.
Fast Balancing

- Suppose $A$ is symmetric and $DADe = e$ where $D = \text{diag}(x)$.
- Rewrite: $Ax - \text{diag}(x)^{-1}e = 0$.
- Solve using Newton method. Newton step solved approximately with CG.
- Easily adapted to nonsymmetric $A$.
- Typically requires a small number of matrix-vector products using $A$ and $A^T$. 
Computing Singular Vectors

- We compute \( p = 1, 2, 3, 4, 5, \ldots \) singular vectors with svds or eigs.
- Output dependent on \( p \) and initial guess.
- We project out contribution of \( e \).
- We can use information from \( p \)th vector to further project to enhance \((p + 1)\)th.
Splitting a vector

- Reorder components of vector by size.
- Identify jumps with an edge detecting algorithm (Canny filter).
  - 1D signal to suppress noise on step signal / 2D image processing
  - good localization, good detection and one detected peak per window
  - optimal by considering the convolution product of the signal and the first derivative of Gaussian function.
Splitting a vector

- Reorder components of vector by size.
- Identify jumps with an edge detecting algorithm (Canny filter).
- Jumps resolved at multiple levels.
- Parameter free determination of $k$ blocks.
Example: Matrix Blocks, first left-right singular vector
Example: Matrix Blocks, second left-right singular vector

and after 5 recursive spectral clustering steps,
Amalgamation and quality measure (1/2)

How to qualify the merging clusters?

▶ use (an adapted variant of) modularity measure

\[
Q_r = \frac{1}{m} \sum_{k=1}^{r_c} \left( v_k^T A A^T v_k - \frac{1}{m} |J_k|^2 \right)
\]

(1)

▶ \(0 \leq Q_r \leq 1 - \frac{1}{r_c}\)

▶ test for pairwise amalgamation that maximise the increase of the quality measure

▶ loop until local maximum is reached.

▶ no pairwise amalgamation improvement implies no improvement by any type of amalgamation on a current state.
Amalgamation and quality measure (2/2)

<table>
<thead>
<tr>
<th>Clustering step</th>
<th>$r_c$</th>
<th>$Q_r$</th>
<th>$c_c$</th>
<th>$Q_c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1</td>
<td>4</td>
<td>0.49971</td>
<td>5</td>
<td>0.40197</td>
</tr>
<tr>
<td>Step 2 (before amalg.)</td>
<td>12</td>
<td>0.50007</td>
<td>20</td>
<td>0.30859</td>
</tr>
<tr>
<td>Step 2 (after amalg.)</td>
<td>7</td>
<td>0.55605</td>
<td>5</td>
<td>0.49542</td>
</tr>
<tr>
<td>Step 3 (before amalg.)</td>
<td>28</td>
<td>0.42255</td>
<td>25</td>
<td>0.31697</td>
</tr>
<tr>
<td>Step 3 (after amalg.)</td>
<td>7</td>
<td>0.55757</td>
<td>6</td>
<td>0.49898</td>
</tr>
<tr>
<td>Step 4 (before amalg.)</td>
<td>16</td>
<td>0.50657</td>
<td>10</td>
<td>0.48335</td>
</tr>
<tr>
<td>Step 4 (after amalg.)</td>
<td>7</td>
<td>0.55757</td>
<td>6</td>
<td>0.49898</td>
</tr>
<tr>
<td>Step 5 (before amalg.)</td>
<td>16</td>
<td>0.50657</td>
<td>10</td>
<td>0.48335</td>
</tr>
<tr>
<td>Step 5 (after amalg.)</td>
<td>7</td>
<td>0.55757</td>
<td>6</td>
<td>0.49898</td>
</tr>
</tbody>
</table>
Application: Network Clusters with Zakary karate club (34 nodes, 78 edges)
Example: The Need for Preprocessing
Future Work

- We aim to provide a block structure amenable to factorisation and/or preconditioning.
- Preprocessing: there is no point in trying to scale a matrix if it is not fully indecomposable. We guard against this but would like to do better.
- Bi-clustering: algorithm can work with rectangular input.
- Perturbation theory for singular vectors of nearly block matrices is missing.
- We use a measure of cluster quality in reconstructing blocks...are we using the right one?
- Need to fully understand role of diagonal dominance in the substructures.