Boolean matrix factorization *meets* consecutive ones property

Nikolaj Tatti & **Pauli Miettinen**

UNIVERSITY OF HELSINKI

UNIVERSITY OF EASTERN FINLAND

Boolean matrix factorisation

- Given a matrix $\boldsymbol{A}$, find matrices $\boldsymbol{B}$ and $\boldsymbol{C}$ s.t. $\boldsymbol{A} \approx \boldsymbol{B} \circ \boldsymbol{C}^T$
  - SVD, NMF, PCA, ICA, …
- In **Boolean matrix factorisation** (BMF)
  1. the input matrix $\boldsymbol{A}$ is binary
  2. the factor matrices $\boldsymbol{B}$ and $\boldsymbol{C}$ are binary
  3. the algebra is Boolean ($1 + 1 = 1$)
- Sparse factors, good interpretability, sometimes you just need binary factors

Example of BMF

$$
\begin{array}{c} \\ 1\\2\\3\\4\\5\\6 \end{array}
\begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{pmatrix} 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \end{pmatrix} \end{array}
=
\begin{array}{c} \\ 1\\2\\3\\4\\5\\6 \end{array}
\begin{array}{c} \begin{array}{ccc} a & b & c \end{array} \\ \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \end{array}
\circ
\begin{array}{c} \\ d\\e\\f \end{array}
\begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix} \end{array}
$$

Example of BMF

4th column = *a* + *b*

$$\begin{array}{c} \\ 1\\2\\3\\4\\5\\6 \end{array}\begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{pmatrix} 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \end{pmatrix}\end{array} = \begin{array}{c} \\ 1\\2\\3\\4\\5\\6 \end{array}\begin{array}{c} \begin{array}{ccc} a & b & c \end{array} \\ \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}\end{array} \circ \begin{array}{c} \\ d\\e\\f \end{array}\begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix}\end{array}$$

Example of BMF

4th column = *a* + *b*

$$
\begin{array}{c} \\ 1\\2\\3\\4\\5\\6 \end{array}
\begin{array}{c} \begin{matrix} 1 & 2 & 3 & 4 \end{matrix} \\ \begin{pmatrix} 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \end{pmatrix} \end{array}
=
\begin{array}{c} \\ 1\\2\\3\\4\\5\\6 \end{array}
\begin{array}{c} \begin{matrix} a & b & c \end{matrix} \\ \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \end{array}
\circ
\begin{array}{c} \\ d\\e\\f \end{array}
\begin{array}{c} \begin{matrix} 1 & 2 & 3 & 4 \end{matrix} \\ \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix} \end{array}
$$

(2, 4) = 1 = 1 + 1 = (2, *a*) + (2, *b*)

Consecutive ones property

- A binary vector $\boldsymbol{x}$ has the **consecutive ones property** (C1P) if all of its 1s are consecutive
  - e.g. $\boldsymbol{x} = (0, 0, 1, 1, 1, 1, 1, 0)$
- A binary matrix $\boldsymbol{X}$ has the C1P if its rows can be permuted so that all of its columns have C1P simultaneously
- A BMF $\boldsymbol{A} \approx \boldsymbol{B} \circ \boldsymbol{C}^T$ has the C1P if both $\boldsymbol{B}$ and $\boldsymbol{C}$ have the C1P

Cyclic vectors and matrices

- A binary vector $\boldsymbol{x}$ is **cyclic** if it's complement has the consecutive ones property
  - e.g. $\boldsymbol{x} = (1, 1, 0, 0, 0, 0, 0, 1)$
- A binary matrix is **cyclic** if its columns have either the C1P or the cyclic ones property
- A BMF $\boldsymbol{A} \approx \boldsymbol{B} \circ \boldsymbol{C}^T$ is **cyclic** if both $\boldsymbol{B}$ and $\boldsymbol{C}$ are cyclic

Example

$$
\begin{array}{c} \\ 1\\2\\3\\4\\5\\6 \end{array}
\begin{array}{c}
\begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\
\begin{pmatrix}
1 & 0 & 1 & 1 \\
1 & 1 & 1 & 1 \\
0 & 1 & 1 & 1 \\
0 & 1 & 1 & 1 \\
0 & 1 & 1 & 0 \\
1 & 0 & 1 & 1
\end{pmatrix}
\end{array}
=
\begin{array}{c} \\ 1\\2\\3\\4\\5\\6 \end{array}
\begin{array}{c}
\begin{array}{ccc} a & b & c \end{array} \\
\begin{pmatrix}
1 & 0 & 0 \\
1 & 1 & 0 \\
0 & 1 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1 \\
1 & 0 & 0
\end{pmatrix}
\end{array}
\circ
\begin{array}{c} \\ d\\e\\f \end{array}
\begin{array}{c}
\begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\
\begin{pmatrix}
1 & 0 & 1 & 1 \\
0 & 1 & 1 & 1 \\
0 & 1 & 1 & 0
\end{pmatrix}
\end{array}
$$

Example

cyclic

$$
\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array}
\begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{pmatrix} 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \end{pmatrix} \end{array}
=
\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array}
\begin{array}{c} \begin{array}{ccc} a & b & c \end{array} \\ \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \end{array}
\circ
\begin{array}{c} d \\ e \\ f \end{array}
\begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix} \end{array}
$$

Example

cyclic C1P

$$
\begin{array}{c|cccc|}
 & 1 & 2 & 3 & 4 \\
1 & 1 & 0 & 1 & 1 \\
2 & 1 & 1 & 1 & 1 \\
3 & 0 & 1 & 1 & 1 \\
4 & 0 & 1 & 1 & 1 \\
5 & 0 & 1 & 1 & 0 \\
6 & 1 & 0 & 1 & 1
\end{array}
=
\begin{array}{c|ccc|}
 & a & b & c \\
1 & 1 & 0 & 0 \\
2 & 1 & 1 & 0 \\
3 & 0 & 1 & 0 \\
4 & 0 & 1 & 0 \\
5 & 0 & 0 & 1 \\
6 & 1 & 0 & 0
\end{array}
\circ
\begin{array}{c|cccc|}
 & 1 & 2 & 3 & 4 \\
d & 1 & 0 & 1 & 1 \\
e & 0 & 1 & 1 & 1 \\
f & 0 & 1 & 1 & 0
\end{array}
$$

Example

$$
\begin{array}{c} \\ 1\\2\\3\\4\\5\\6 \end{array}
\begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{pmatrix} 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \end{pmatrix} \end{array}
=
\begin{array}{c} \\ 1\\2\\3\\4\\5\\6 \end{array}
\begin{array}{c} \begin{array}{ccc} \textbf{cyclic} & \textbf{C1P} & \textbf{C1P} \\ a & b & c \end{array} \\ \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \end{array}
\circ
\begin{array}{c} \\ d\\e\\f \end{array}
\begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix} \end{array}
$$

Example

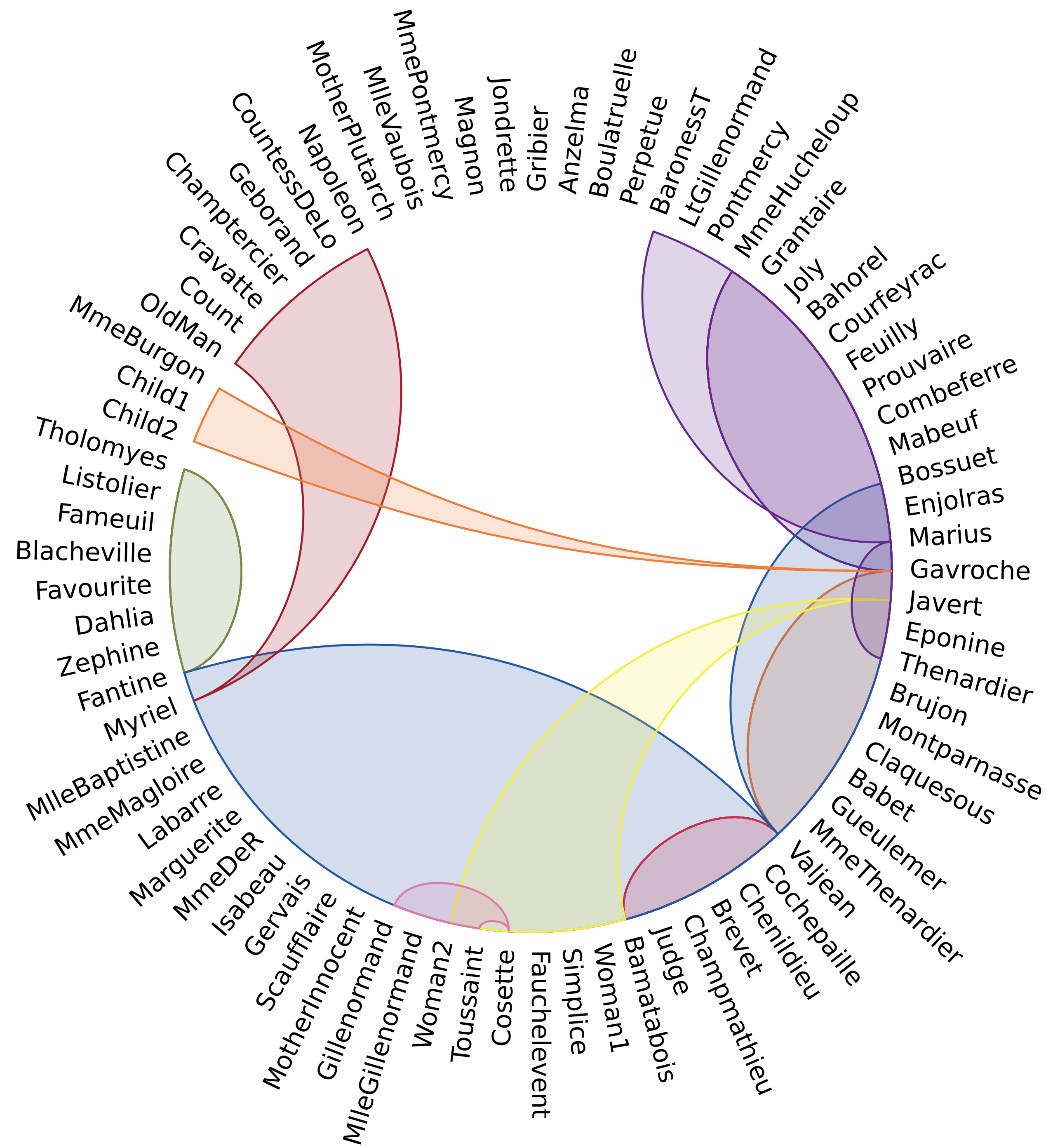
$$
\begin{array}{c|cccc}
 & 1 & 2 & 3 & 4 \\
\hline
1 & 1 & 0 & 1 & 1 \\
2 & 1 & 1 & 1 & 1 \\
3 & 0 & 1 & 1 & 1 \\
4 & 0 & 1 & 1 & 1 \\
5 & 0 & 1 & 1 & 0 \\
6 & 1 & 0 & 1 & 1
\end{array}
=
\begin{array}{c|ccc}
 & \textbf{cyclic} & \textbf{C1P} & \textbf{C1P} \\
 & a & b & c \\
\hline
1 & 1 & 0 & 0 \\
2 & 1 & 1 & 0 \\
3 & 0 & 1 & 0 \\
4 & 0 & 1 & 0 \\
5 & 0 & 0 & 1 \\
6 & 1 & 0 & 0
\end{array}
\circ
\begin{array}{c|cccc|l}
 & 1 & 2 & 3 & 4 & \\
\hline
d & 1 & 0 & 1 & 1 & \textbf{cyclic} \\
e & 0 & 1 & 1 & 1 & \textbf{C1P} \\
f & 0 & 1 & 1 & 0 & \textbf{C1P}
\end{array}
$$

Why?

- We aim at finding cyclic or C1P BMF
  - All problems are NP-hard
  - The extra constraints typically make the reconstruction less accurate
  - But they can also make the factors easier to understand
    - E.g. rows are time points $\Rightarrow$ C1P = something starts, lasts, and ends, but doesn't resurrect
- And we can draw graphs

Edge bundle plots

- Edge bundles are a way to visualise graphs
  - Which edges to bundle?
  - How to order the vertices?
- Cyclic or C1P BMF:
  - Factors define the edges
  - The order is given by the constraint



How?

Typical BMF algorithm:

1. construct a set of candidate vectors
2. **repeat**
   1. add a new factor from the candidates based on what reduces the error most
3. **until** k factors are selected

But how to satisfy the constraints?

PQ-trees

- A PQ-tree is a data structure form the ’70s that encodes permutations
  - the leaves are the indices
  - the children of P-nodes can be ordered arbitrarily
  - the children of Q-nodes must stay in order (or reverse)

Theorem (Booth & Lueker, ’76). There exists a PQ-tree that encodes exactly the row permutations where a matrix has the C1P (or cyclic) property and this tree can be updated in polynomial time.

Using PQ-trees

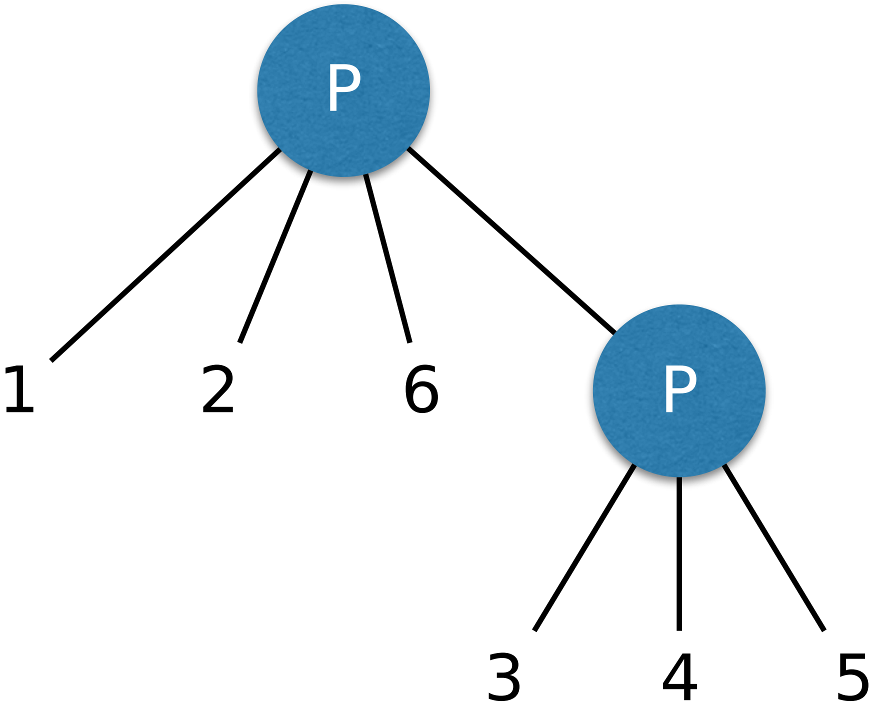
- We keep a PQ-tree for both factor matrices
- Every candidate column introduces new constraints
  - Update the PQ-tree with these constraints
    $\Rightarrow$ will show if constraints can't be satisfied
- Select the best of the possible candidates
  - Quality can be calculated quickly
- One factor can be added in $O(\mathrm{nnz}(\boldsymbol{A}) + m + n)$ time

Example

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array} \begin{array}{c} a \\ \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \end{array}$$

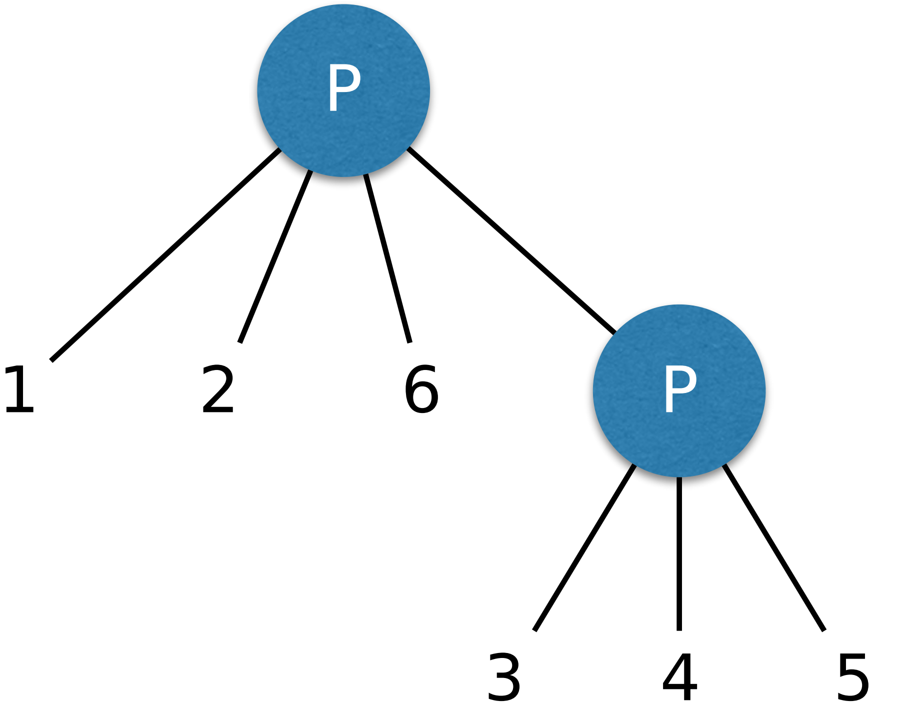
Example

$$
\begin{array}{c}
\\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6
\end{array}
\begin{array}{c}
a \\
\begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}
\end{array}
$$



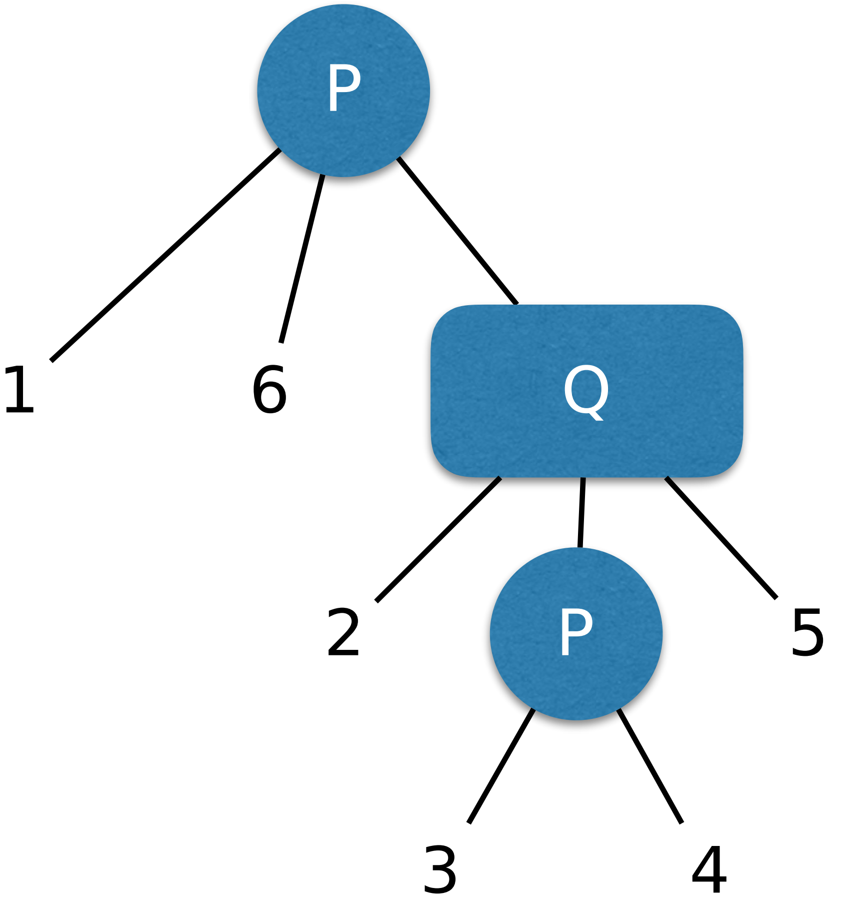
Example

	a	b
1	1	0
2	1	1
3	0	1
4	0	1
5	0	0
6	1	0

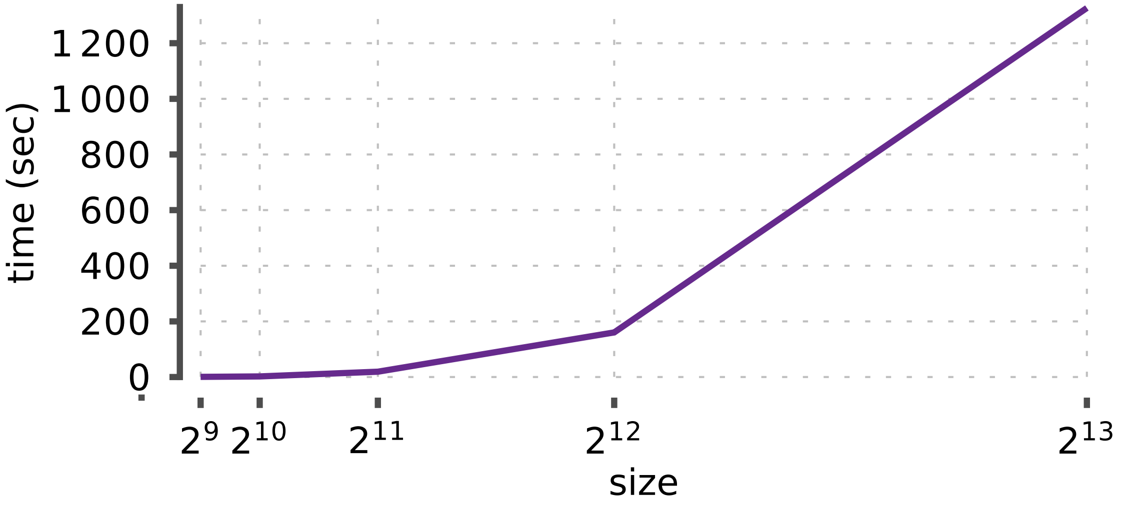


Example

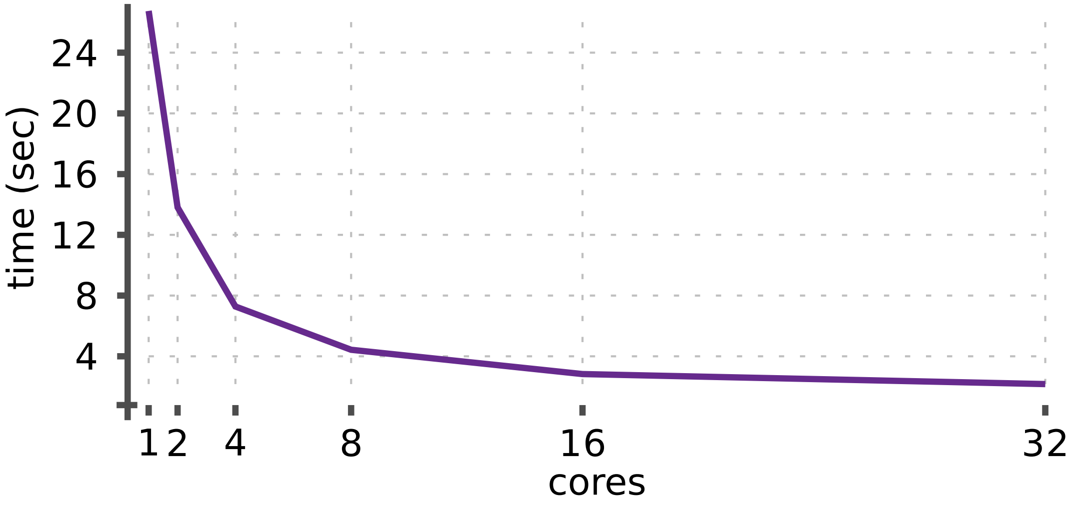
	a	b
1	1	0
2	1	1
3	0	1
4	0	1
5	0	0
6	1	0



Scalability

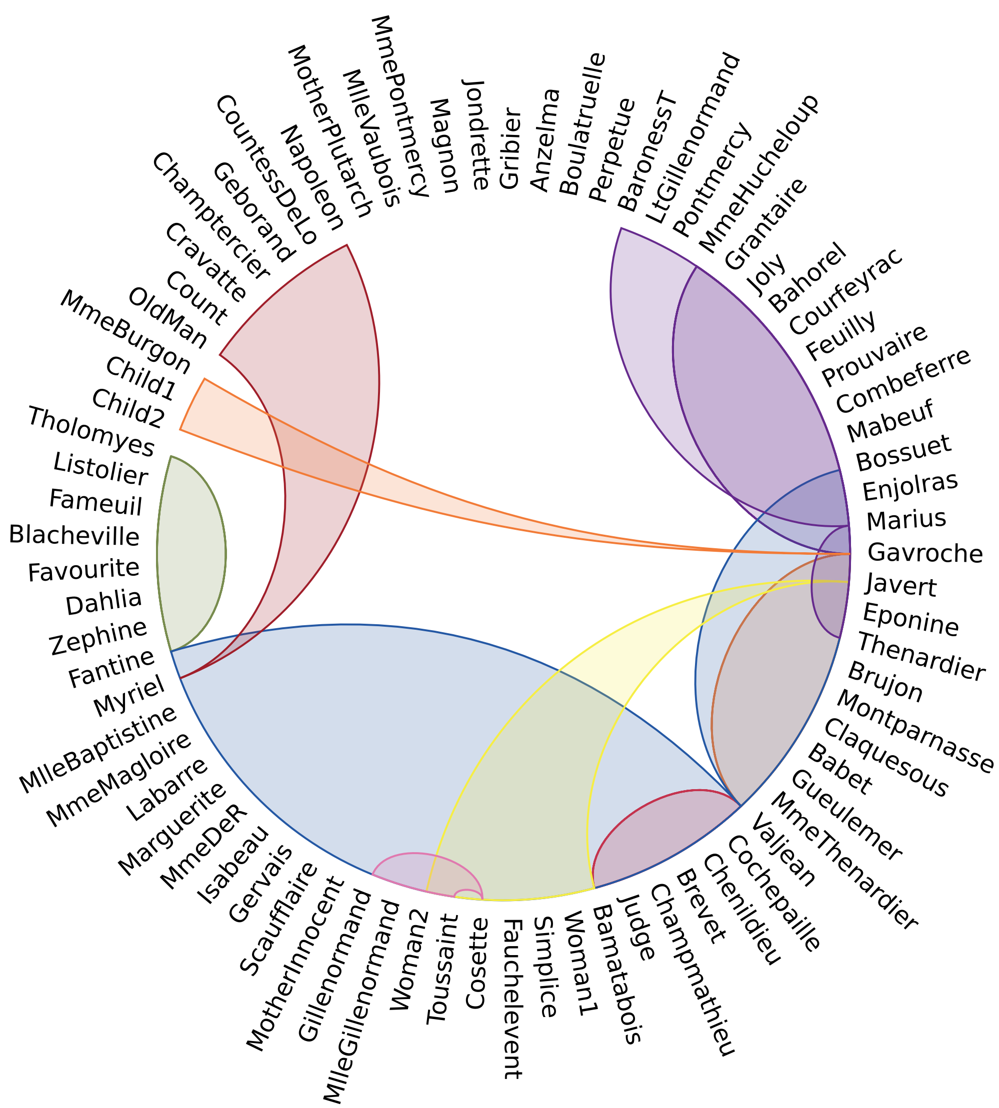


Scales well with input size

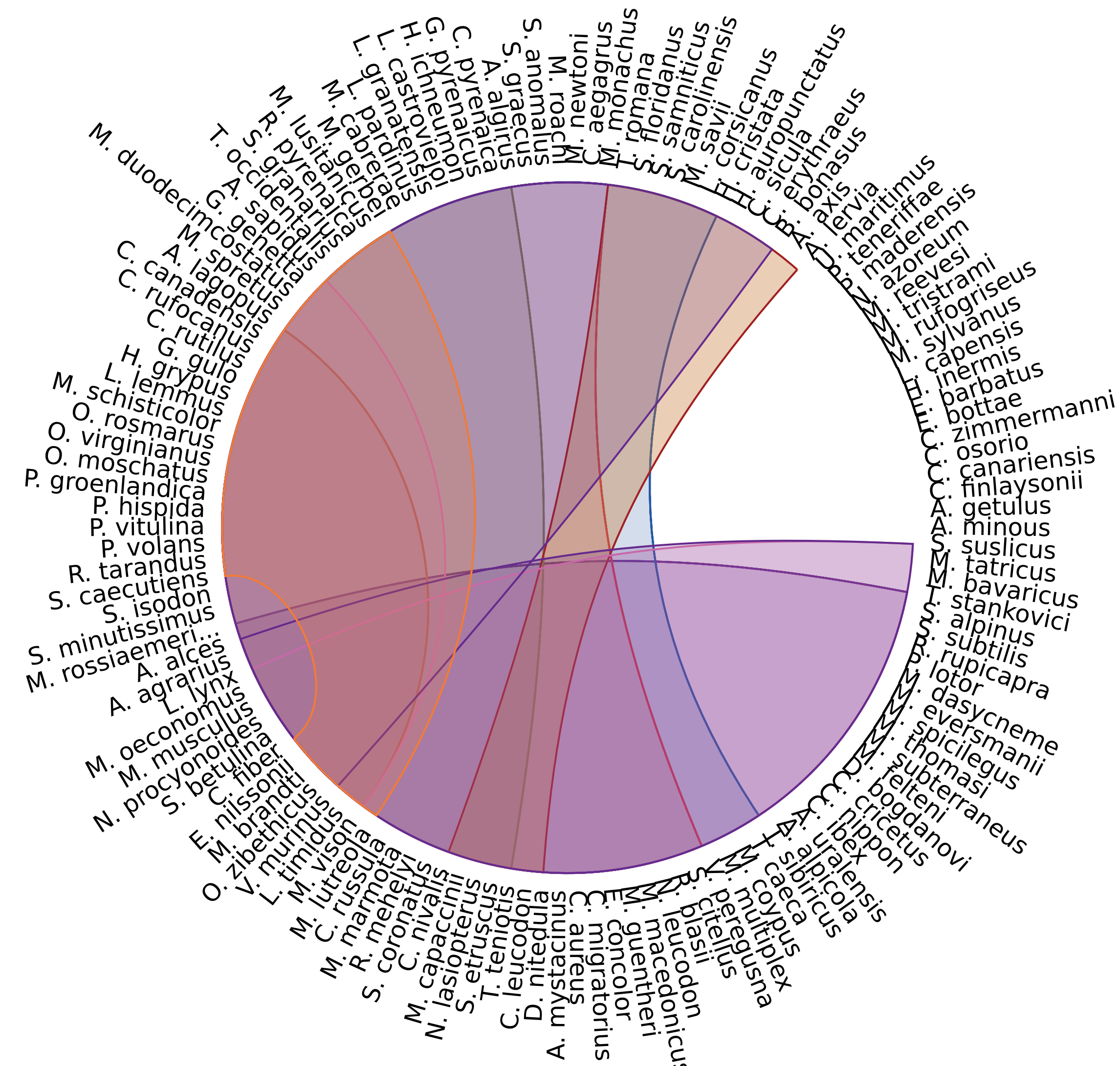


Parallelises well to many cores

More edge bundle plots



Les Misérables characters
cyclic factors



Land mammal species
C1P factors

Summary

- C1P and cyclic factors further improve the interpretability of BMF factors
  - Natural for edge bundle plots
- PQ-trees facilitate the efficient testing of compatible factors
- Possible extensions to other algebras, time series, …

Summary

- C1P and cyclic factors further improve the interpretability of BMF factors
  - Natural for edge bundle plots
- PQ-trees facilitate the efficient testing of compatible factors
- Possible extensions to other algebras, time series, …

Thank You!