Compression of sparse matrices

November 10th, 2023

Aljaž Jeromel,
Borut Žalik

University of Maribor
Faculty of Electrical Engineering and Computer Science

Institute of Computer Science
Laboratory for Geospatial Modelling, Multimedia and Artificial Intelligence

Presentation contents

- Sparse matrices
- Proposed solution
- Christofides-Serdyukov algorithm
- Compression of chain codes
- Preliminary results
- Future work

Sparse matrices

- 2D arrays
- At least 2/3 of the cells are empty
- It is sensible to only encode non-empty cells
  - How do we efficiently encode the coordinates?

Proposed solution – outline (1 / 3)

- Encode the path between points
- How to generate the path?
  - Travelling salesman problem
  - Complete solution is not feasible, approximation is needed
  - Christofides-Serdyukov algorithm
- Use chain codes to encode the path
- Compress chain codes

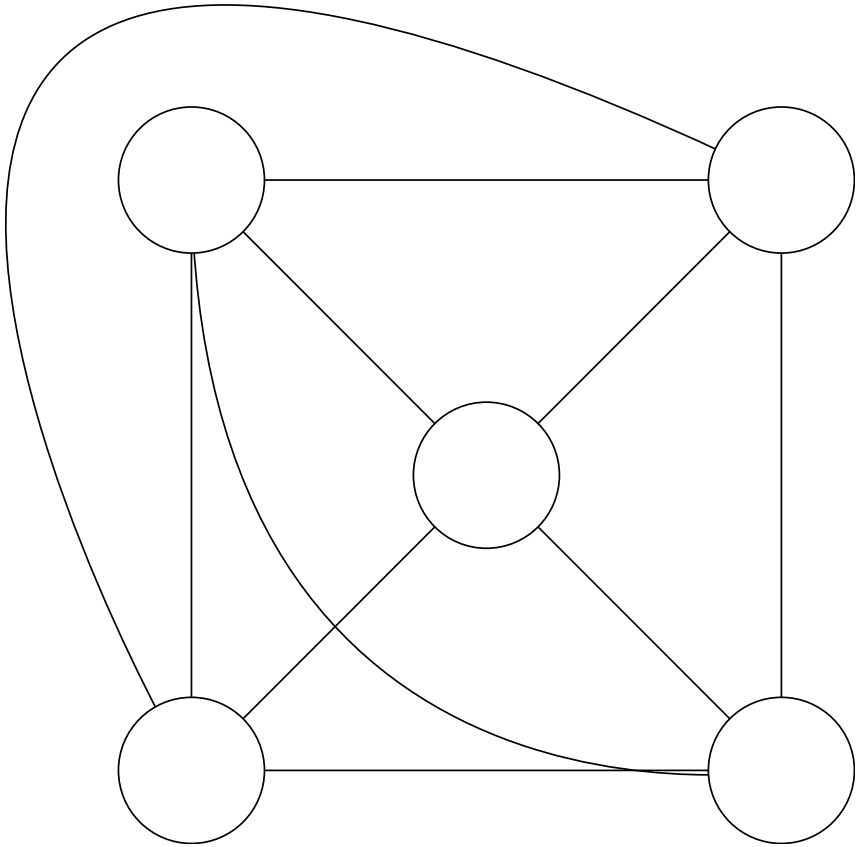
Christofides-Serdyukov algorithm (1/3)

- TSP approximation algorithm
- Used terms:
  - N – the number of vertices
  - Minimum spanning tree: a connected graph with N – 1 edges
  - Perfect matching: connecting the vertices by a set of edges, so that each vertex is connected to exactly one other vertex
  - Eulerian circuit: a circular path on the graph that visits each edge exactly once
  - Hamiltonian circuit: a circular path on the graph that visits each vertex exactly once

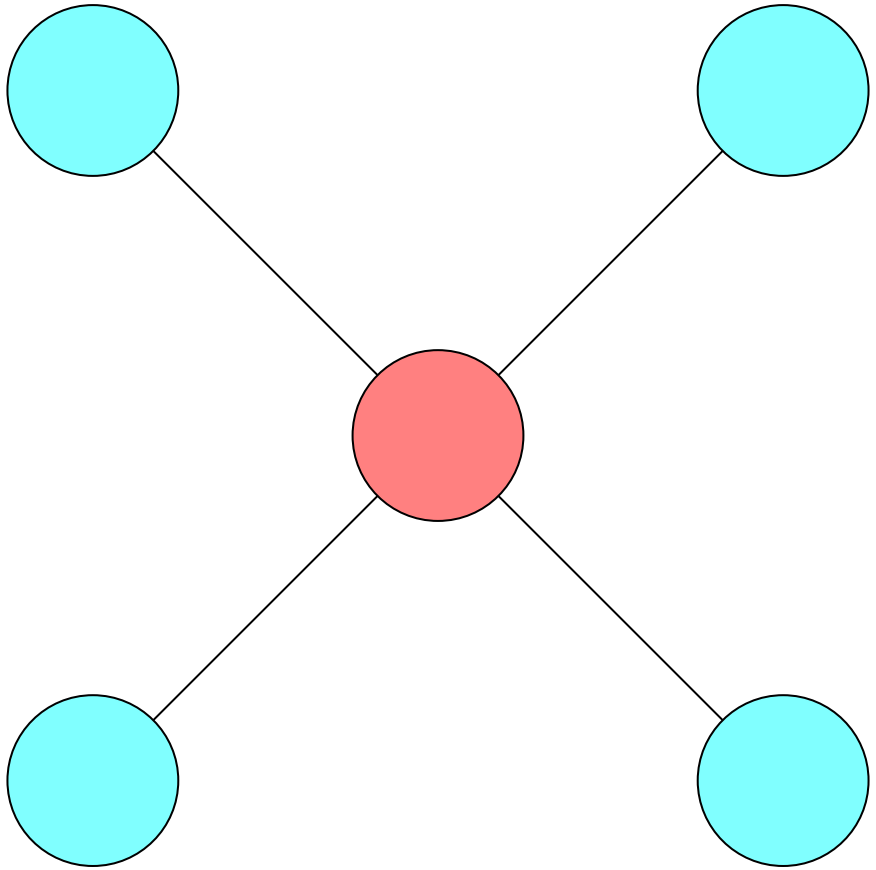
Christofides-Serdyukov algorithm (2/3)

- Create a minimum spanning tree
- Find a minimum-weight perfect matching on vertices of odd degree
- Combine the edges of MST and perfect matching
- Find an Eulerian circuit
- Transform the circuit into a Hamiltonian one

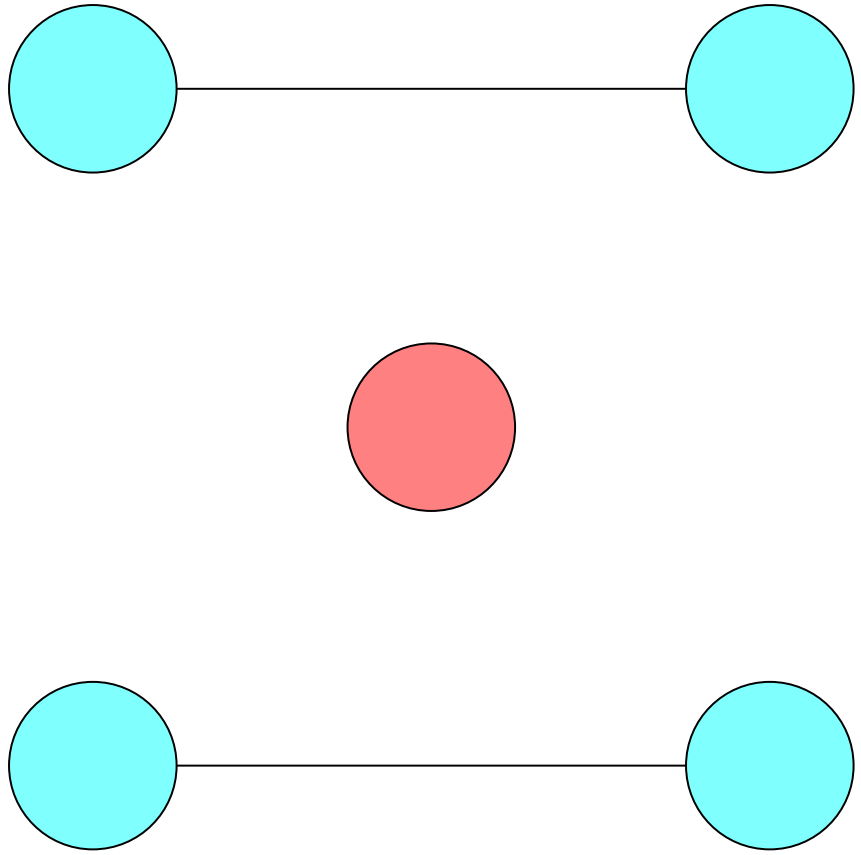
Christofides-Serdyukov algorithm (3/3)



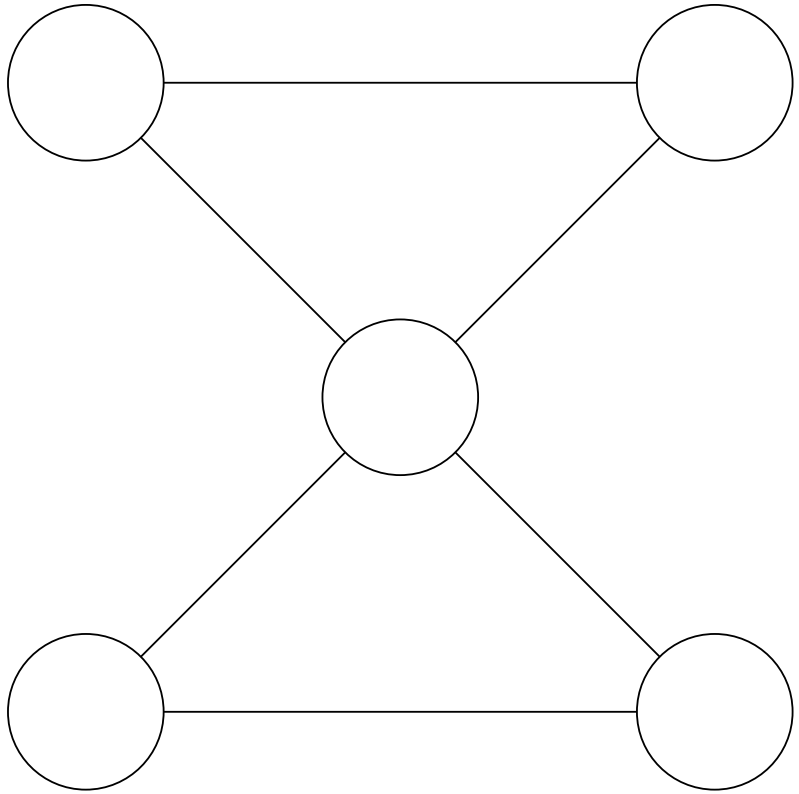
Christofides-Serdyukov algorithm (3/3)



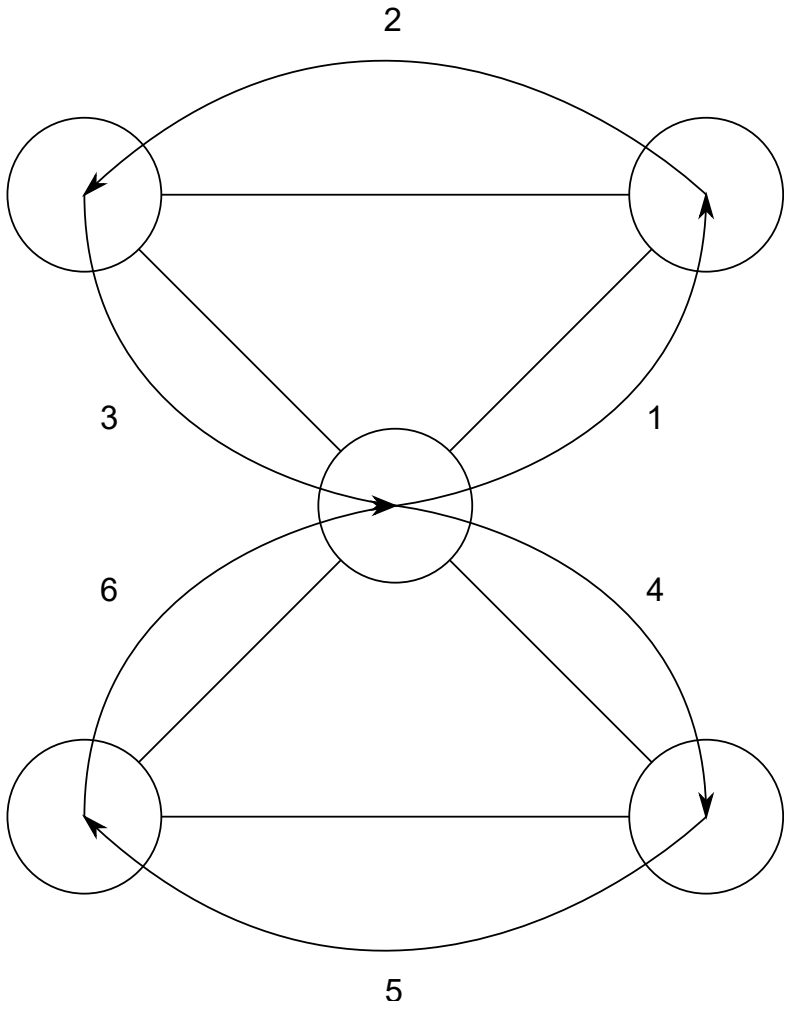
Christofides-Serdyukov algorithm (3/3)



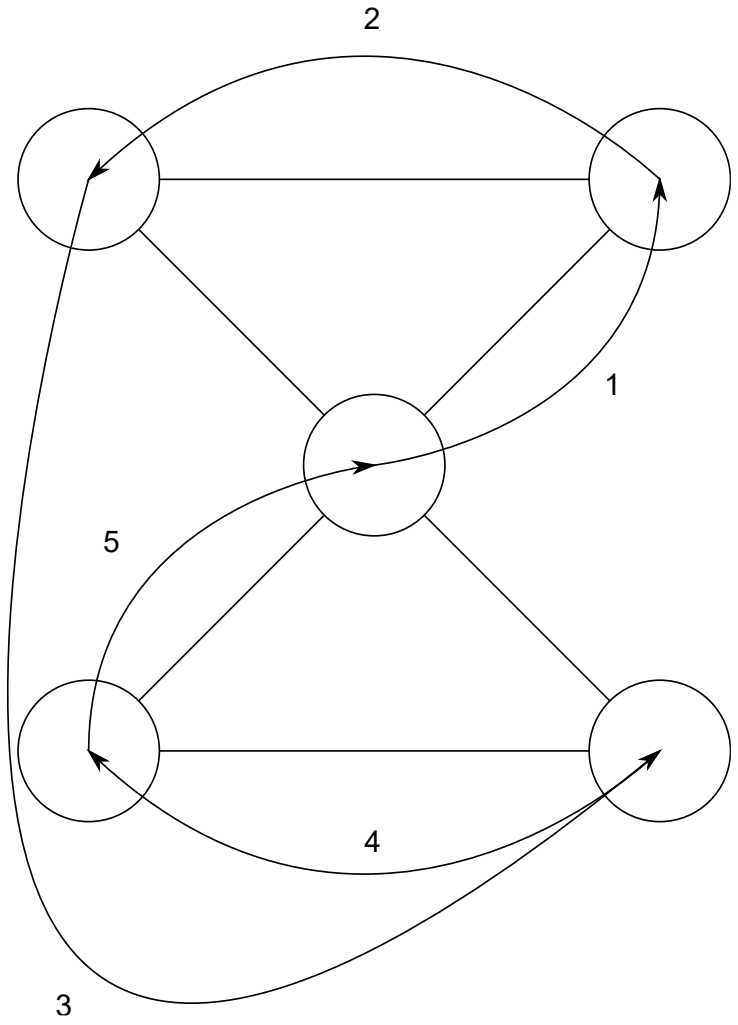
Christofides-Serdyukov algorithm (3/3)



Christofides-Serdyukov algorithm (3/3)



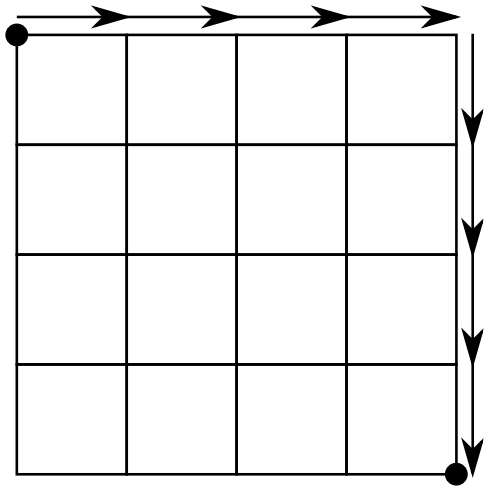
Christofides-Serdyukov algorithm (3/3)



Proposed solution – details (2/3)

- Split the matrix into four quadrants
- Apply Christofides-Serdyukov algorithm on each quadrant separately
- Encode the TSP circuit in each quadrant using chain codes
- Concatenate the chain codes
- Compress the chain codes
- Compress the matrix data (not yet a part of this research)

Proposed solution – chain code traversal (2/3)



- Firstly, move along X-axis
- Then move along Y-axis
- F4 – point position is marked by outputting reverse of the previous direction
  (0, 0, 0, 0, 3, 3, 3, 3, 1)
- Relative F4 – only three codes used to encode movements, fourth code used for marking points
  (0, 0, 0, 0, 1, 0, 0, 0, 3)
- 3OT, VCC – not tested yet

Compression of chain codes

- Testing different combinations of encoding
- String transformations:
  - BWT
  - MTF
- Entropy coding:
  - Arithmetic coder
  - Interpolative coder
  - ANS coder
  - RLE
  - Binary coder (PAQ8L)

Preliminary results

- Best results are obtained with PAQ8L entropy coder
- The method has not yet been tested extensively
- Best preliminary results on a black and white image with 2% white pixels:
  - Relative F4 + BWT + PAQ8L: 0,0479 bpp
  - Relative F4 + BWT + MTF + PAQ8L: 0,0483 bpp
  - Relative F4 + PAQ8L: 0,0494 bpp

Future work

- Efficiency evaluation on a larger dataset
- Test different chain codes
- Test another rasterizing method
- Evaluation of method with a higher number of dimensions
- Research compression of matrix/image values