The MAX-CMO problem

and its dual representation

Paweł Widera
pawel.widera@cs.nott.ac.uk

SUPERVISORS
Natalio Krasnogor, Jonathan Garibaldi

21st European Conference on Operational Research

2006-07-03

Outline

Contact map

A bit of molecular biology

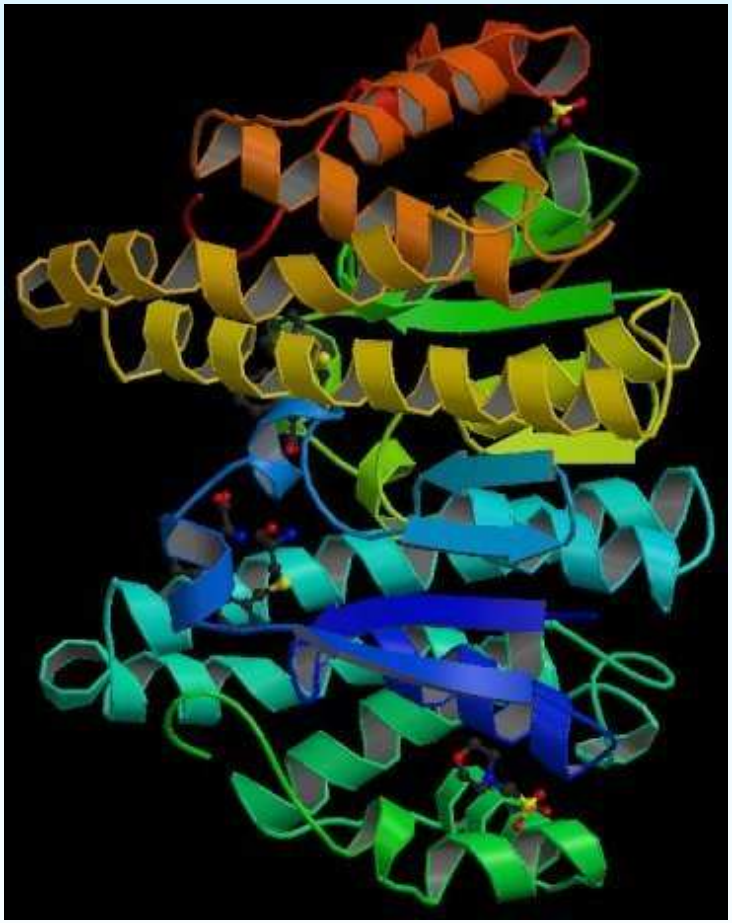


Figure: 10GS (GLUTATHIONE S-TRANSFERASE P1-1)

Protein

- linear arrangement of amino acids (polymer chain)
- amino acid residues linked with peptide bond

Contact map

Mathematical construct capturing the **proximity relation** between residues.

Contact map

A bit of molecular biology

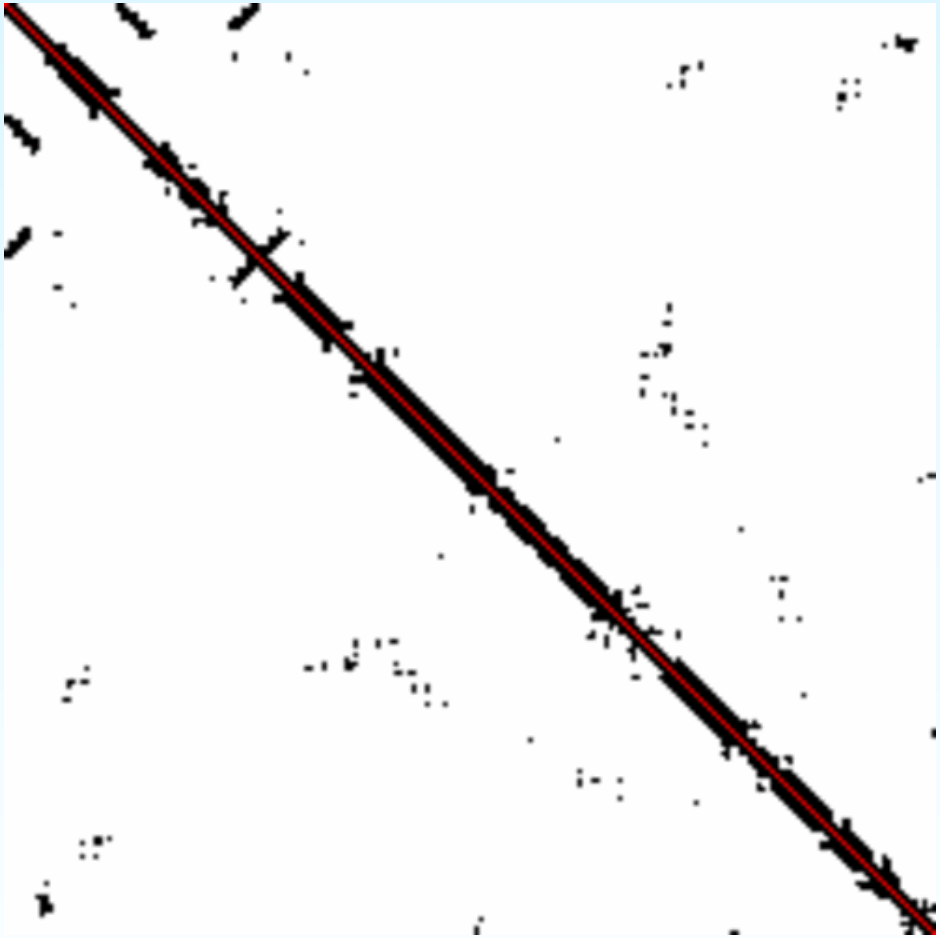


Figure: CM for 10GS:A (210), $\alpha = 0.5\mathring{A}$

Protein

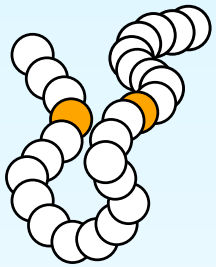
- linear arrangement of amino acids (polymer chain)
- amino acid residues linked with peptide bond

Contact map

Mathematical construct capturing the **proximity relation** between residues.

Contact map

Proximity relation depicted as graph

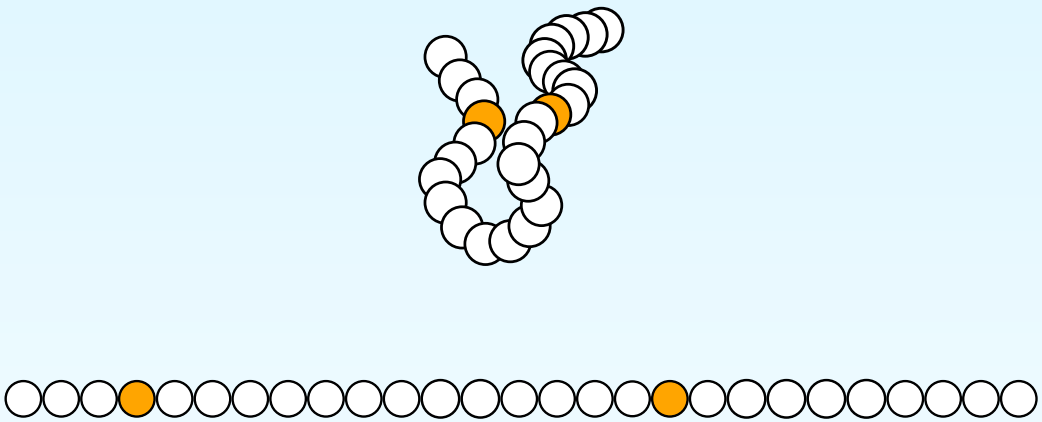


Contact map graph

- $r_i \mathbf{R} r_j \Leftrightarrow \delta(r_i, r_j) \leqslant \alpha$, $\alpha \in [2\text{Å}, 9\text{Å}]$
- node $\leftrightarrows$ residue
- edge $\leftrightarrows$ contact

Contact map

Proximity relation depicted as graph

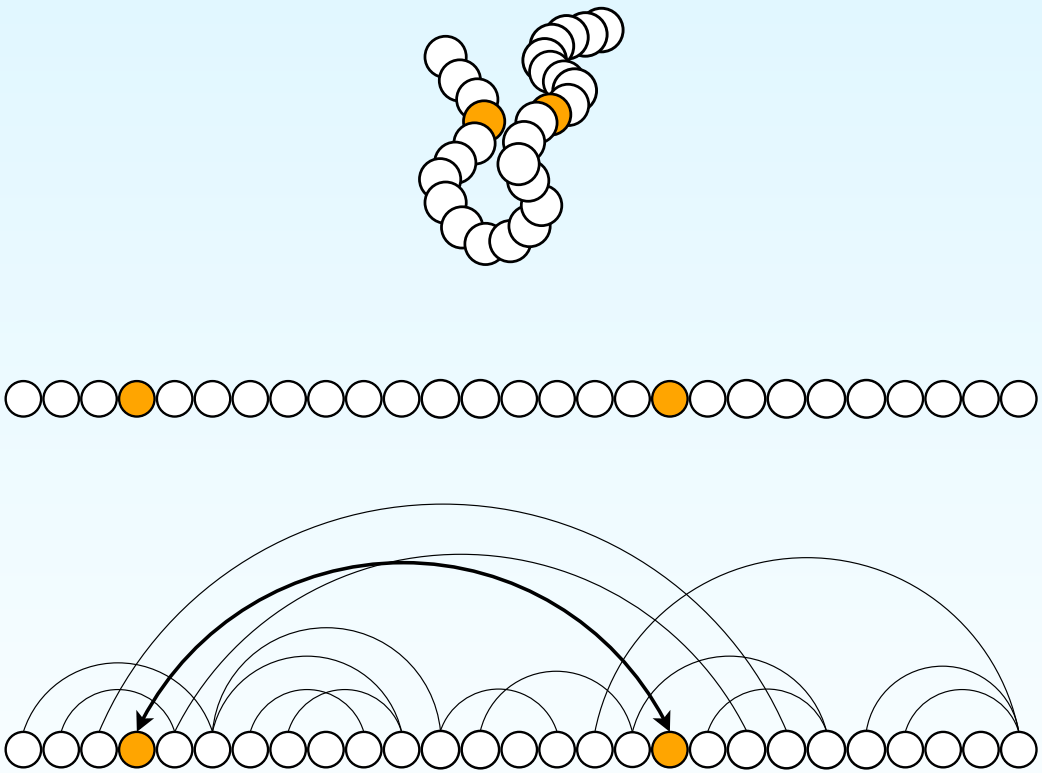


Contact map graph

- $r_i \mathbf{R} r_j \Leftrightarrow \delta(r_i, r_j) \leqslant \alpha$, $\alpha \in [2\text{Å}, 9\text{Å}]$
- node $\leftrightarrows$ residue
- edge $\leftrightarrows$ contact

Contact map

Proximity relation depicted as graph

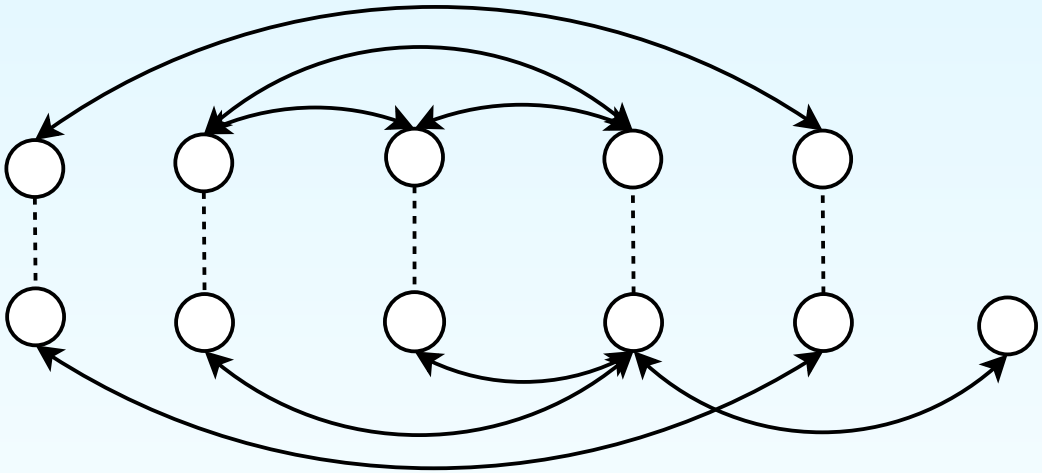


Contact map graph

- $r_i \mathbf{R} r_j \Leftrightarrow \delta(r_i, r_j) \leqslant \alpha$, $\alpha \in [2\text{Å}, 9\text{Å}]$
- node $\leftrightarrows$ residue
- edge $\leftrightarrows$ contact

MAX-CMO problem

Definition and biological background



Definition

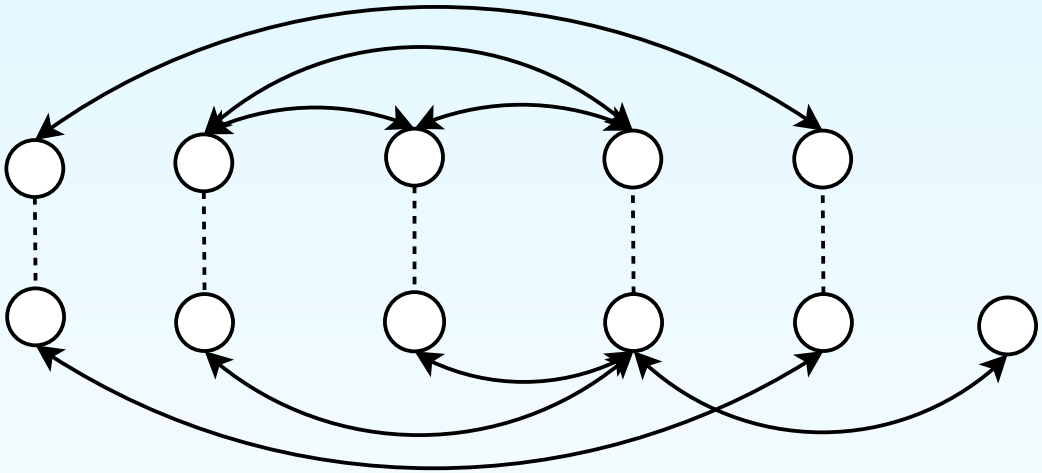
Maximum contact map overlap is an **alignment** of two proteins that **maximises** the structural **similarity**.

Biological application

- proteins structure comparison
- proteins classification

MAX-CMO problem

Definition and biological background



Definition

Maximum contact map overlap is an **alignment** of two proteins that **maximises** the structural **similarity**.

Biological application

- proteins structure comparison
- proteins classification

Graph isomorphism

A bit of graph theory

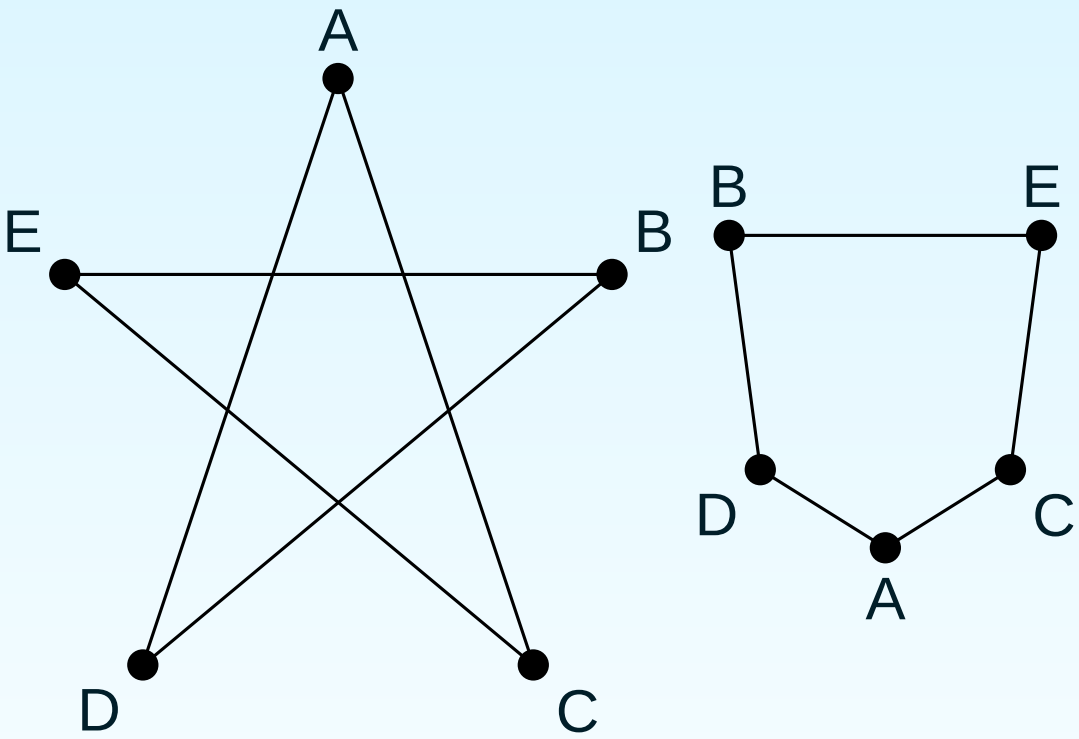


Figure: Isomorphic or not?

Graph isomorphism

Two graphs are isomorphic if there is a one-to-one correspondence between their vertices and there is an edge between two vertices of one graph if and only if there is an edge between the two corresponding vertices in the other graph.

Subgraph isomorphism

Is G1 isomorphic to a subgraph of G2?

Graph isomorphism

A bit of graph theory

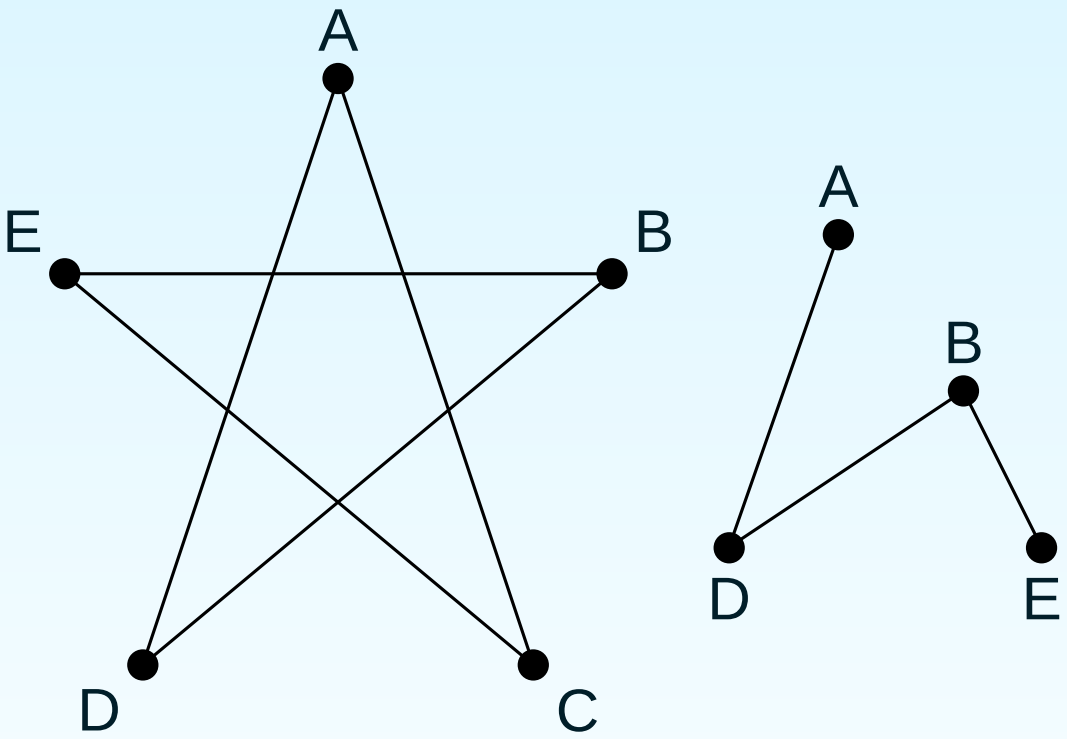


Figure: What about a subgraph?

Graph isomorphism

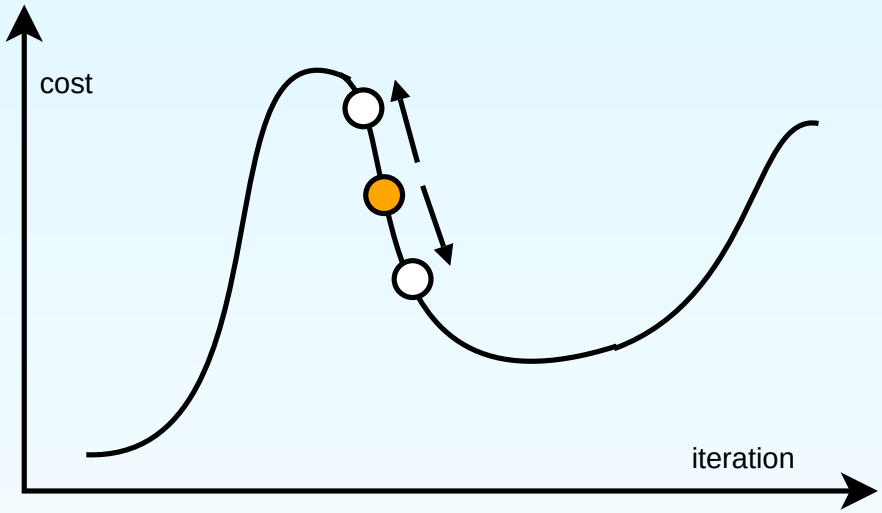
Two graphs are isomorphic if there is a one-to-one correspondence between their vertices and there is an edge between two vertices of one graph if and only if there is an edge between the two corresponding vertices in the other graph.

Subgraph isomorphism

Is G1 isomorphic to a subgraph of G2?

Search strategy

Beyond random walk



Metaheuristic framework

- iterative improvement
- uphill movements acceptance
- based on **neighbourhood** search
- solution **attributes** and **moves**
- global search guidance

Representation issues

Definition of a move

1 2 3 4 5 6 7

1 2 6 4 5 3 7

$\{ 1, 2, 3, 4, 5 \} \rightarrow \{ 1, 2, 3 \}$

$\{ 1, 2, 3, 4, 5 \} \rightarrow \{ 1, 2, 3, 4 \}$

$\{ 1, 2, 3, 4, 5 \} \rightarrow \{ 1, 2, 4 \}$

Move as a transition

- permutation
- n-opt

Move as a building block

- construction vs. destruction
- oscillation strategy

Representation issues

Definition of a move

1 2 3 4 5 6 7

1 2 6 4 5 3 7

$\{ 1, 2, 3, 4, 5 \} \rightarrow \{ 1, 2, 3 \}$

$\{ 1, 2, 3, 4, 5 \} \rightarrow \{ 1, 2, 3, 4 \}$

$\{ 1, 2, 3, 4, 5 \} \rightarrow \{ 1, 2, 4 \}$

Move as a transition

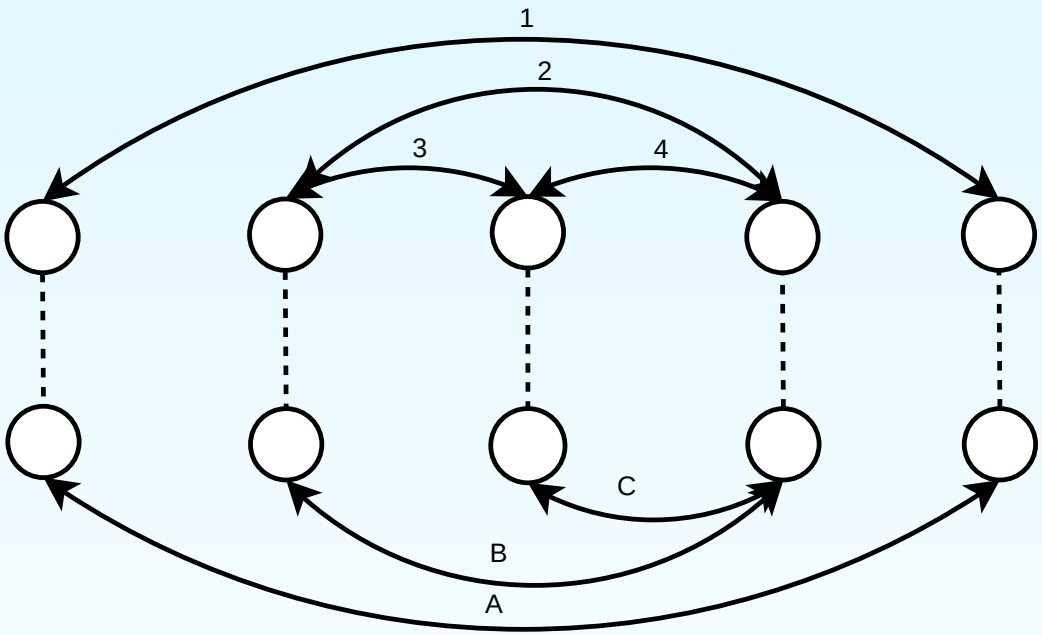
- permutation
- n-opt

Move as a building block

- construction vs. destruction
- oscillation strategy

Classical approach

Domain of a problem

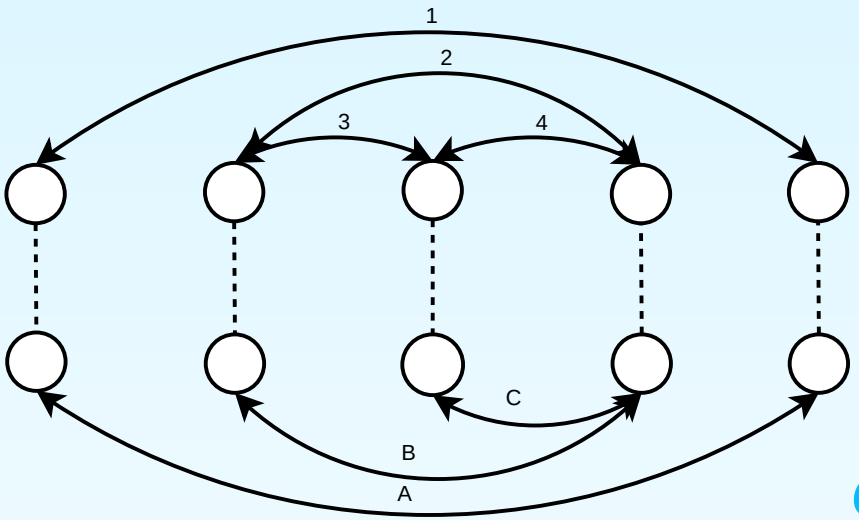


Representation

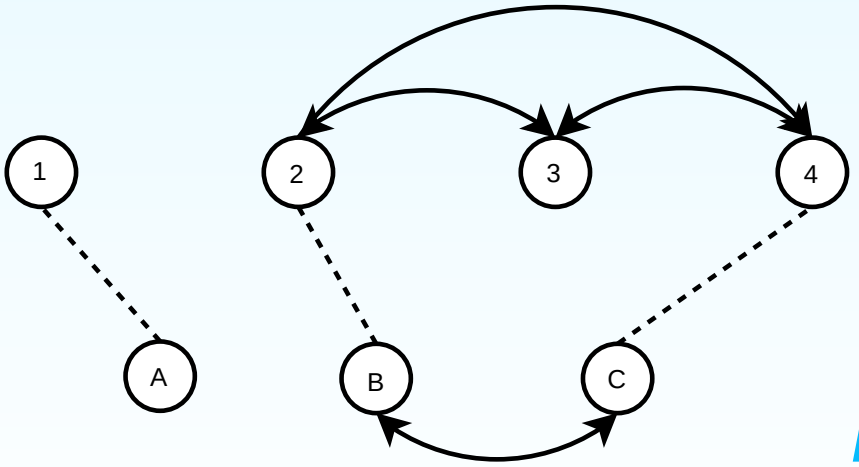
- protein alignment
- nodes matching

Line graph approach

Domain of a solution



G



$L(G)$

Line graph

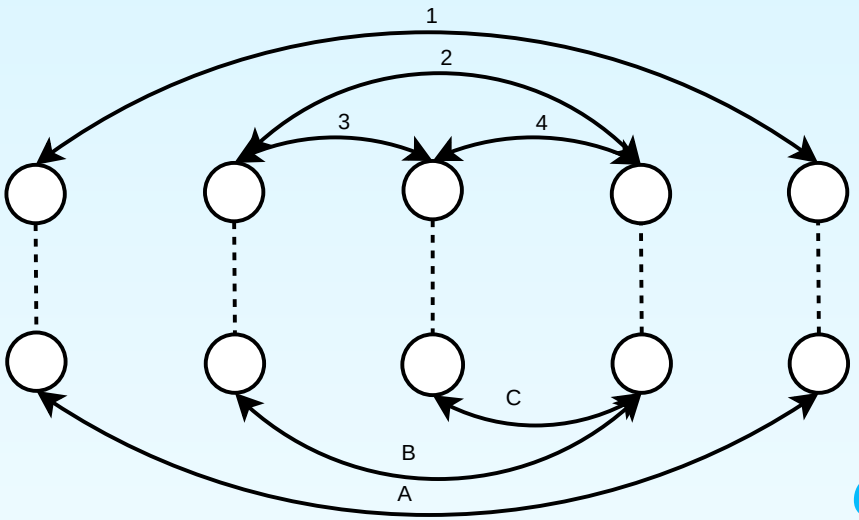
- a node of L(G) represents an edge of G
- two nodes of L(G) are adjacent if edges in G share common node

Representation

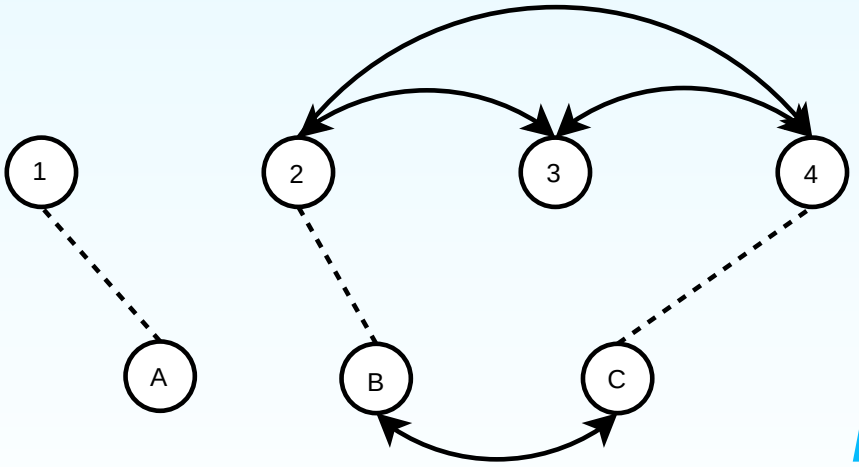
- graph isomorphism
- edges matching

Line graph approach

Domain of a solution



G



$L(G)$

Line graph

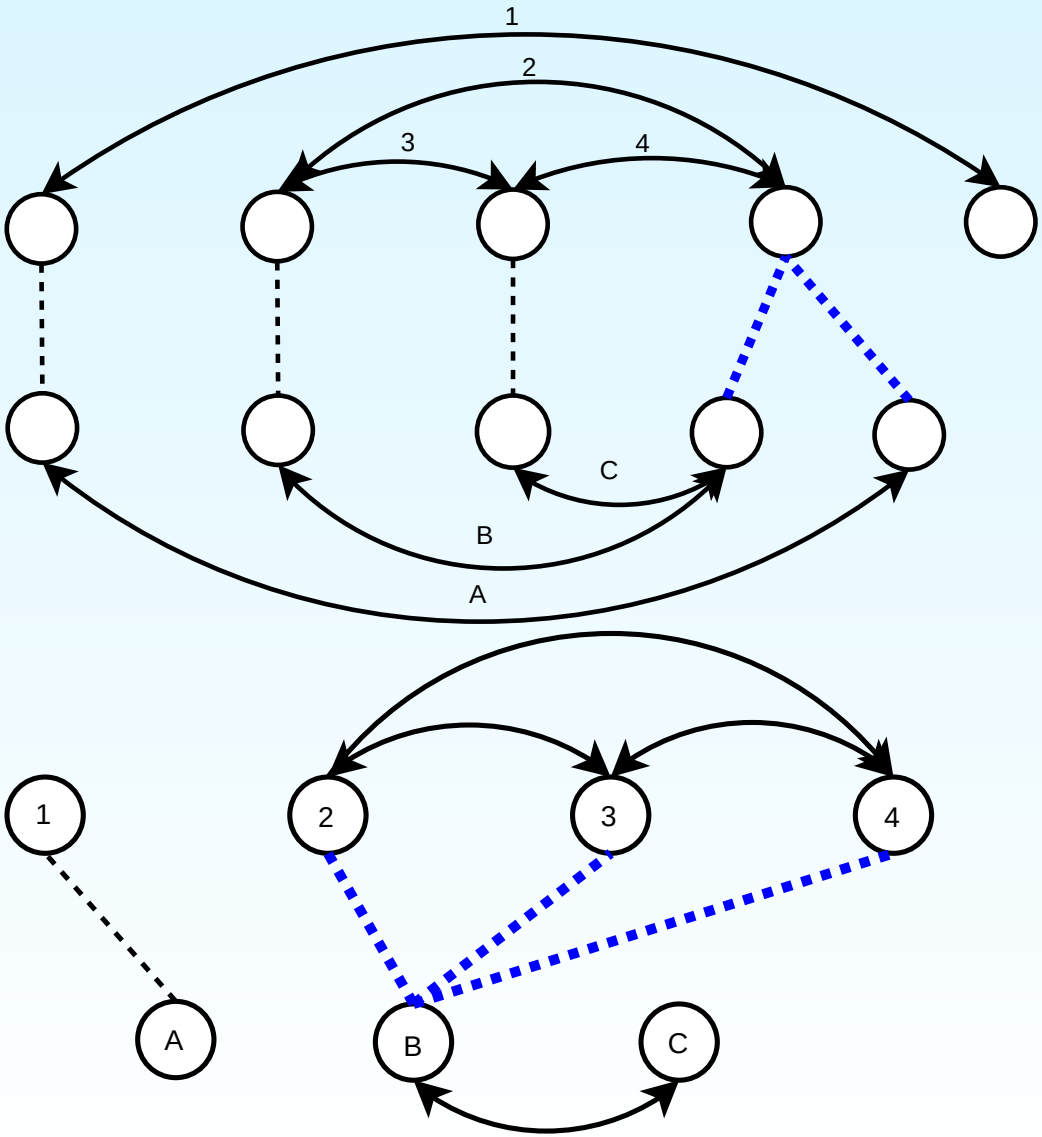
- a node of L(G) represents an edge of G
- two nodes of L(G) are adjacent if edges in G share common node

Representation

- graph isomorphism
- edges matching

Representation comparison

Construction of a neighbourhood

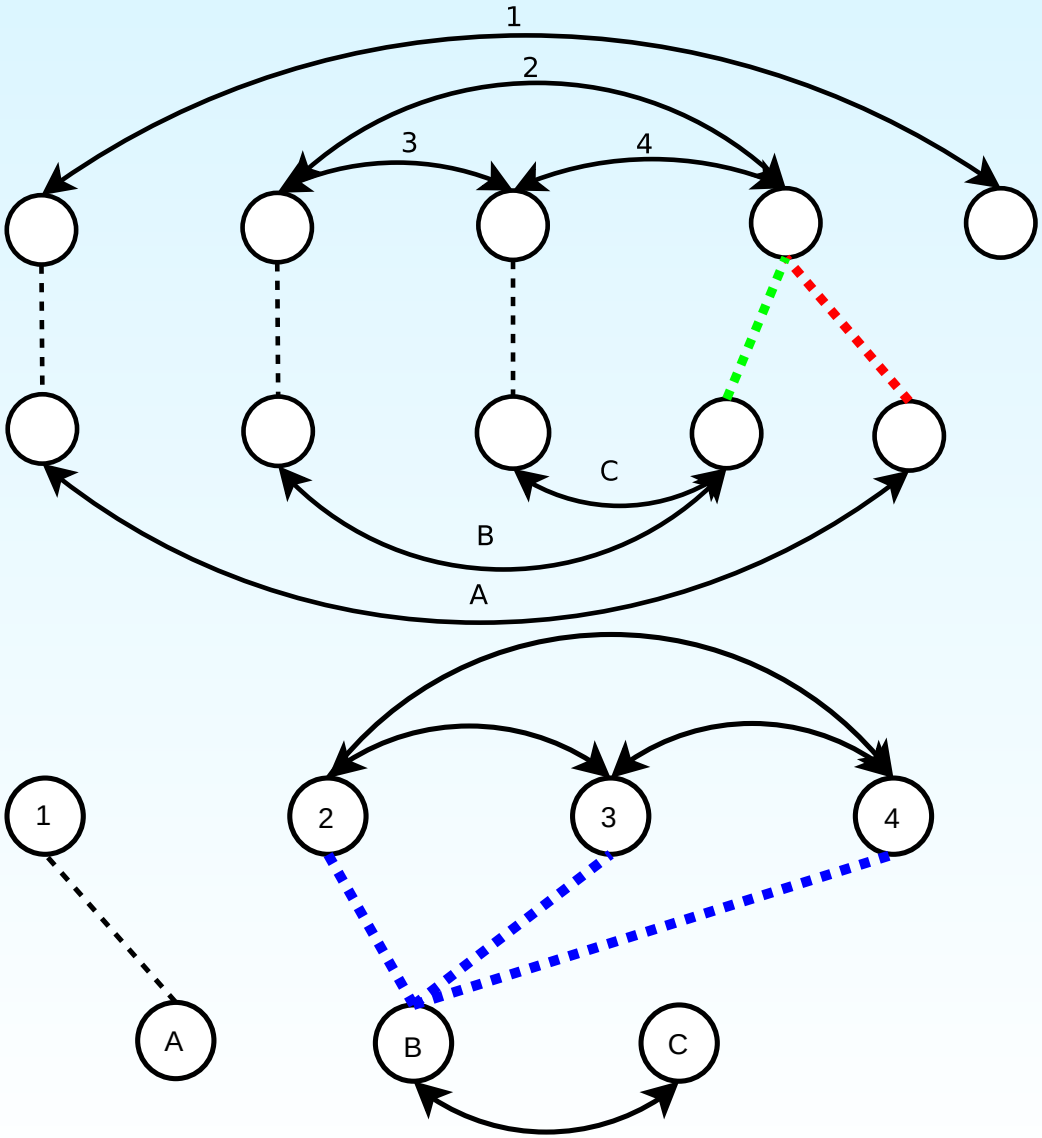


Complexity

- single - $O(1)$
- whole set - $O(n^2)$

Representation comparison

Solution evaluation

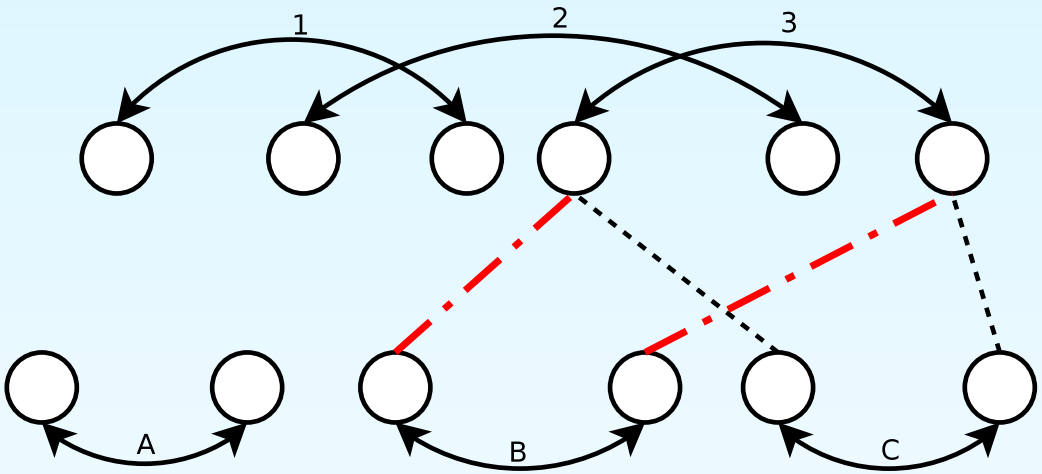


Complexity

- classic
  - $O(n * d(V))$ – full
  - $O(d(V))$ – only new
- dual
  - $O(n)$ – full
  - $O(1)$ – only new

Representation comparison

Solution feasibility

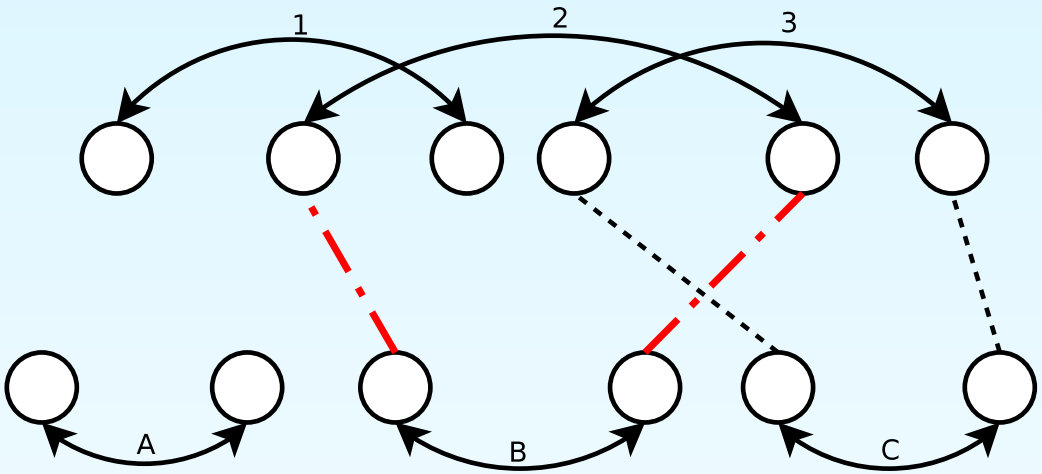


Complexity

- classic - $O(n)$
- dual - ?

Representation comparison

Solution feasibility

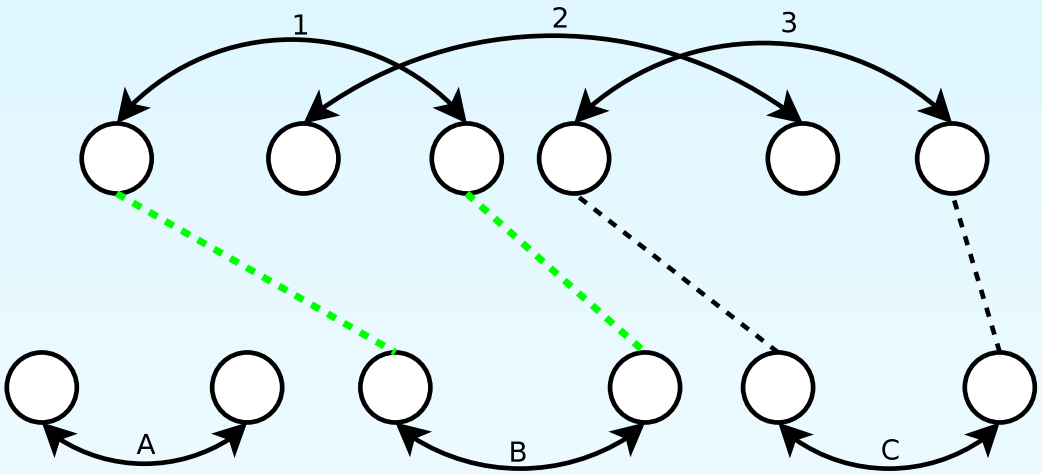


Complexity

- classic - $O(n)$
- dual - ?

Representation comparison

Solution feasibility

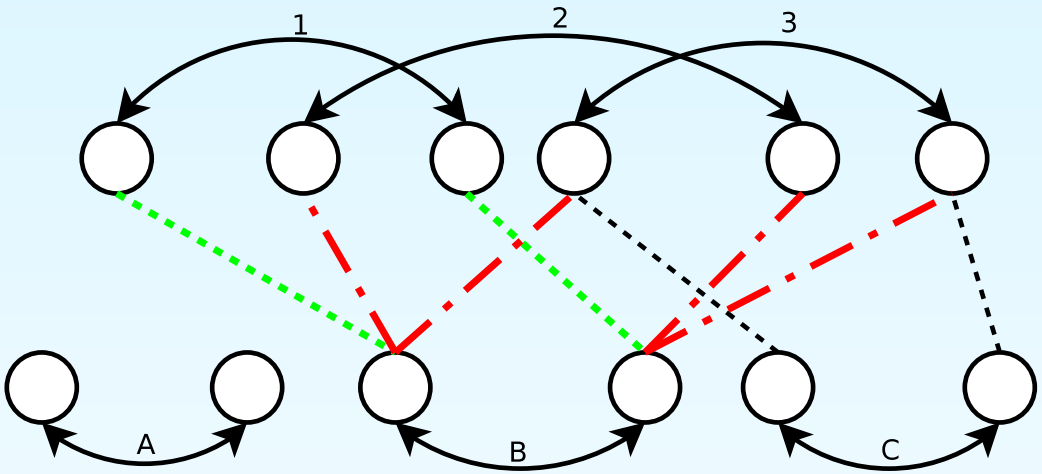


Complexity

- classic - $O(n)$
- dual - ?

Representation comparison

Solution feasibility

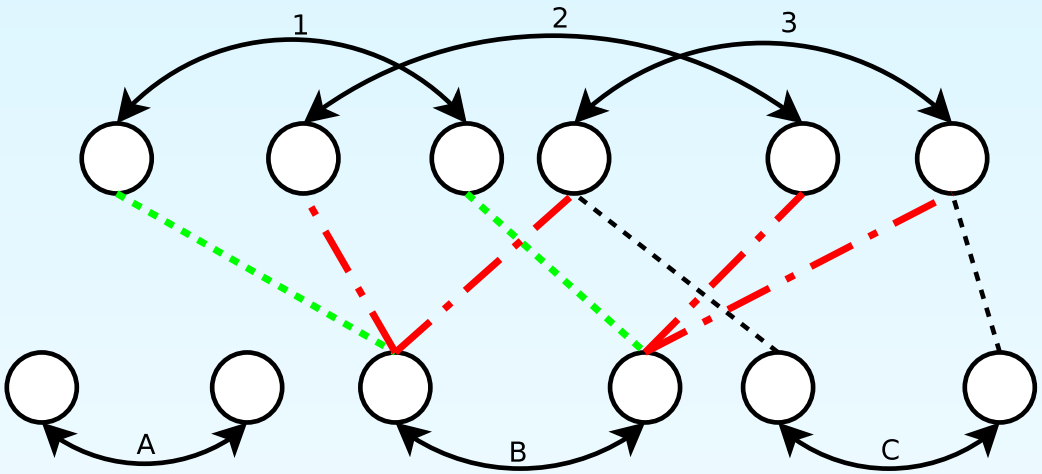


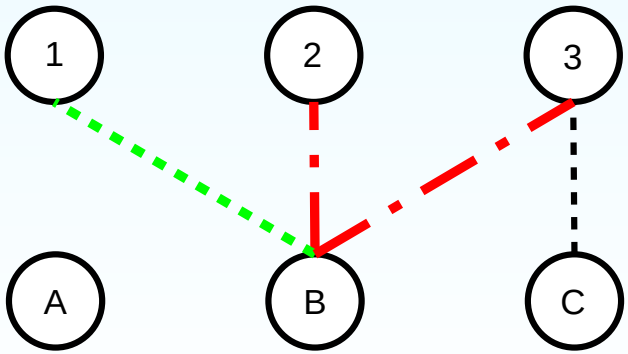
Complexity

- classic - $O(n)$
- dual - ?

Representation comparison

Solution feasibility



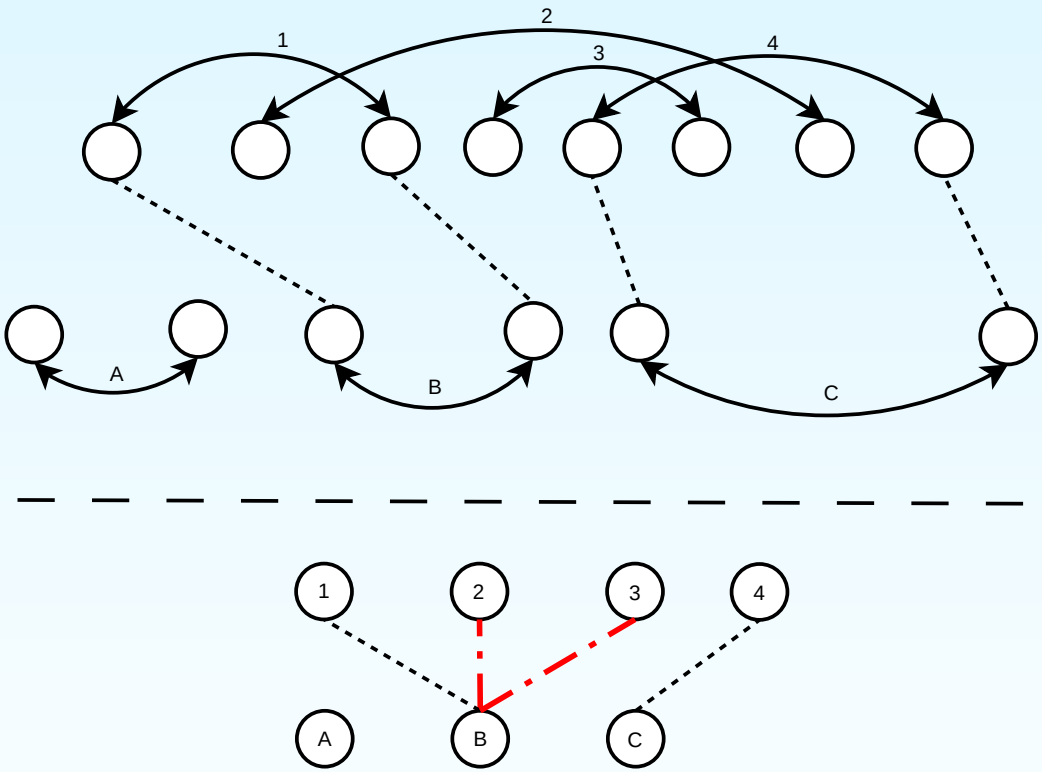


Complexity

- classic - $O(n)$
- dual - ?

Feasibility issues

Interesting features of a line graph



Complexity

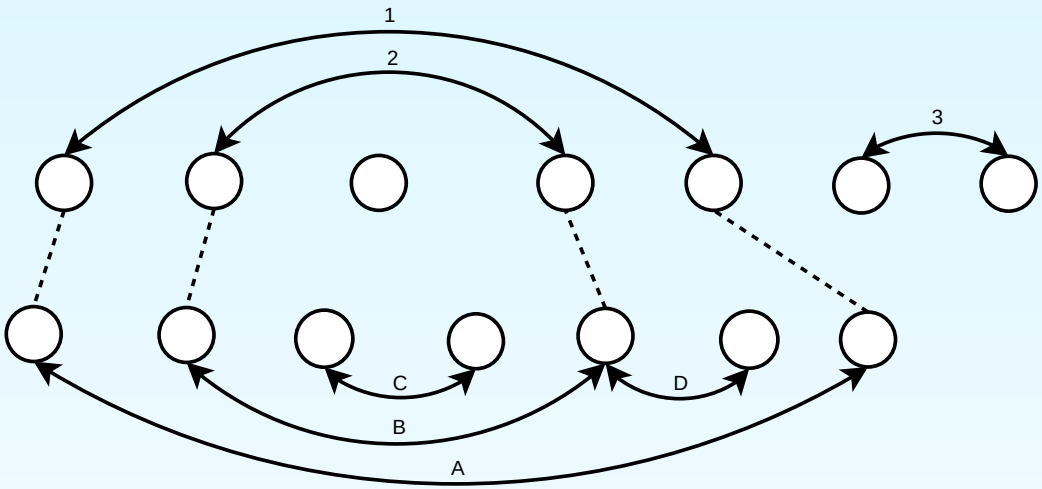
- crossmatching
- superedge boundary
- match consistency

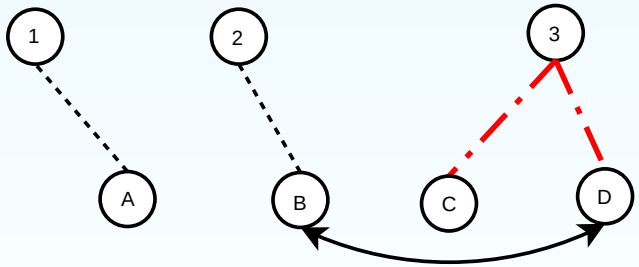
Complexity

- dual - $O(d(V))$?

Feasibility issues

Interesting features of a line graph





Complexity

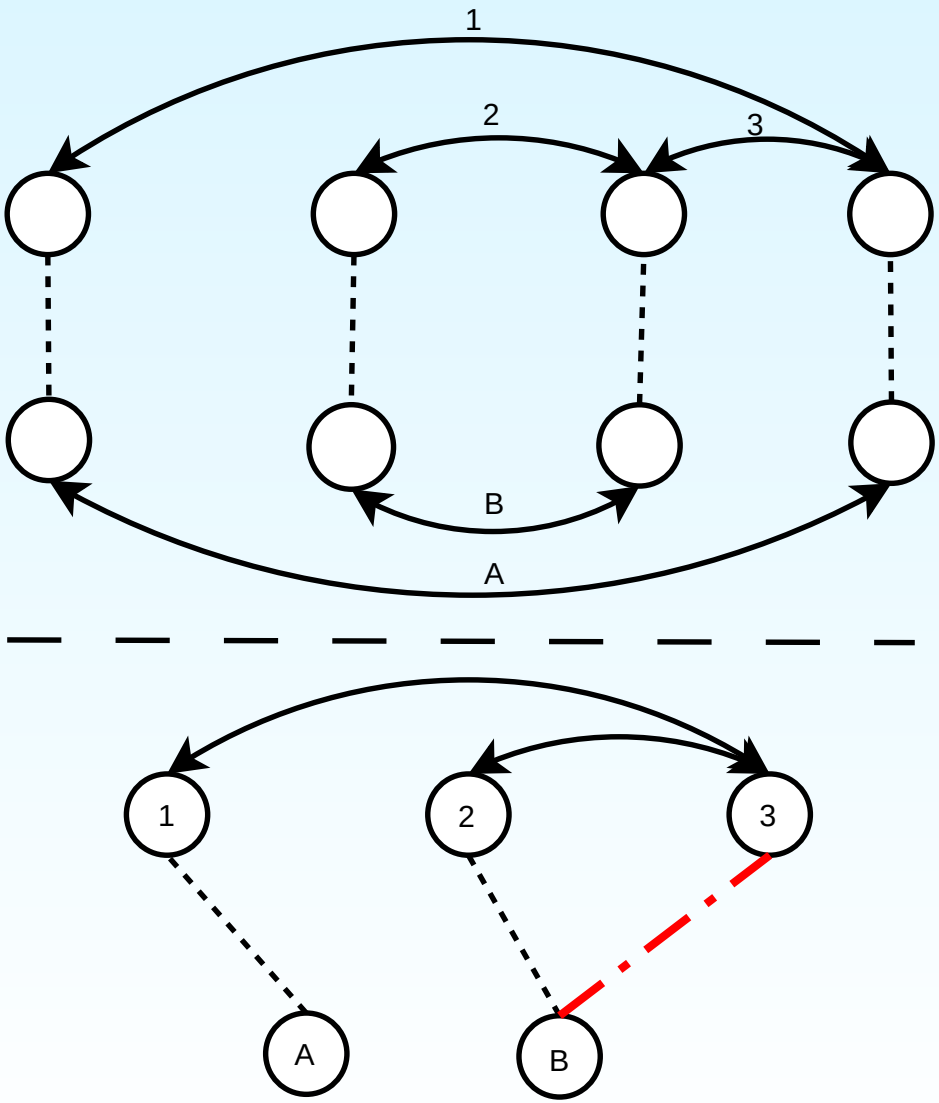
- crossmatching
- superedge boundary
- match consistency

Complexity

- dual - $O(d(V))$?

Feasibility issues

Interesting features of a line graph



Complexity

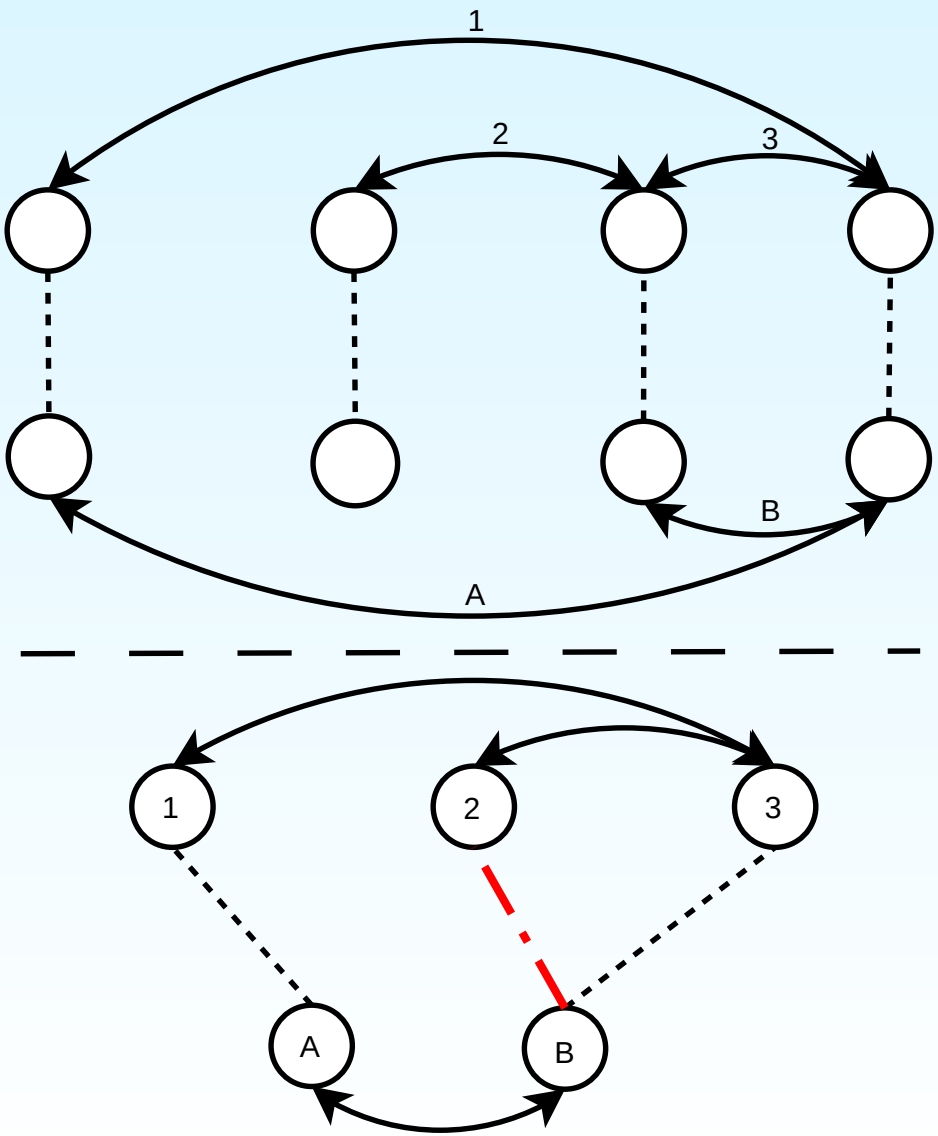
- crossmatching
- superedge boundary
- match consistency

Complexity

- dual - $O(d(V))$?

Feasibility issues

Interesting features of a line graph



Complexity

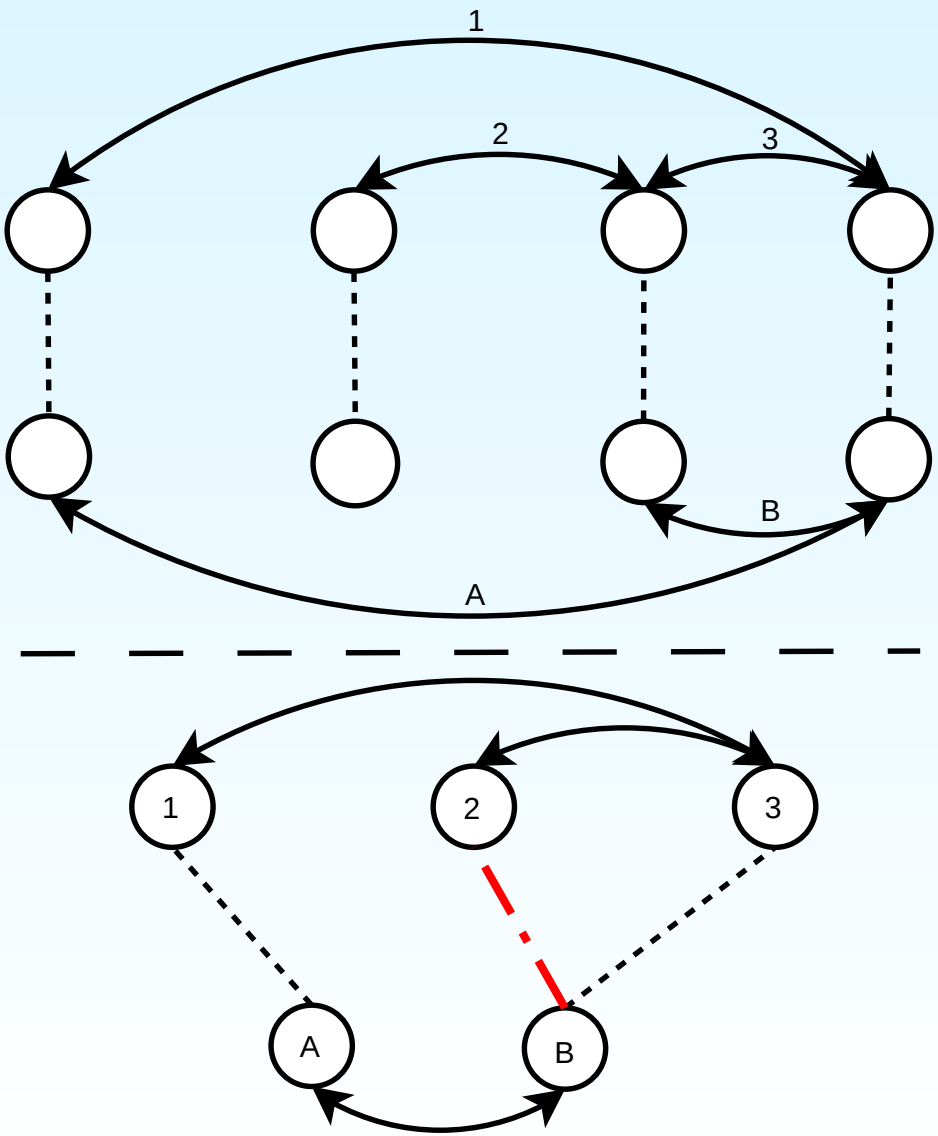
- crossmatching
- superedge boundary
- match consistency

Complexity

- dual - $O(d(V))$?

Feasibility issues

Interesting features of a line graph



Complexity

- crossmatching
- superedge boundary
- match consistency

Complexity

- dual - $O(d(V))$?

Further work

Work in progress

- implementation of a basic version of TS
- frequency based intensification/diversification with Elite List
- empirical proof of correctness

Ideas to consider

- dynamic length of tabu list (reactive search)
- intensification with path relinking

Further work

Work in progress

- implementation of a basic version of TS
- frequency based intensification/diversification with Elite List
- empirical proof of correctness

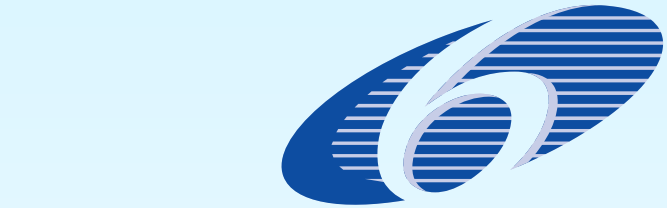
Ideas to consider

- dynamic length of tabu list (reactive search)
- intensification with path relinking

Thank you!

Acknowledgements

This work was supported by **Marie Curie Action** MEST-CT-2004-7597 under the **Sixth Framework Programme** of the European Community.



Contact

pawel.widera@cs.nott.ac.uk