Visualization of Barrier Tree Sequences Revisited

Christian Heine[1] Gerik Scheuermann[1] Christoph Flamm[2]
Ivo L. Hofacker[2] Peter F. Stadler[3]

[1]Image and Signal Processing Group [3]Bioinformatics Group
Department of Computer Science
University of Leipzig

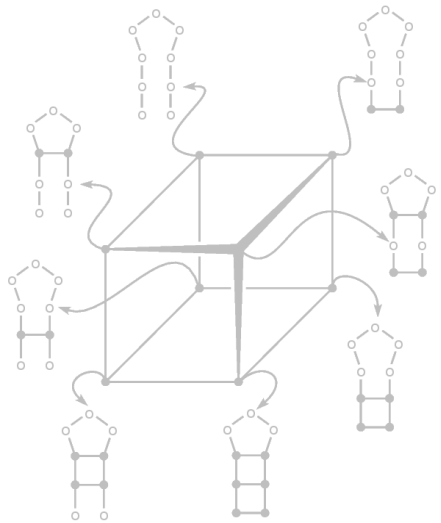
[2]Department of Theoretical Chemistry and Structural Biology
University of Vienna

EMBIO Meeting 2007
2007-05-18, Leipzig, Germany

Outline

Neighborhood of RNA Secondary Structure

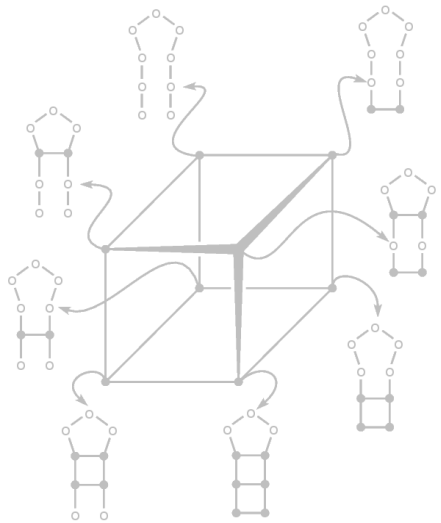
- e.g., all secondary structures that differ by one base pair



neighborhood graph + free energy per vertex = folding landscape

Neighborhood of RNA Secondary Structure

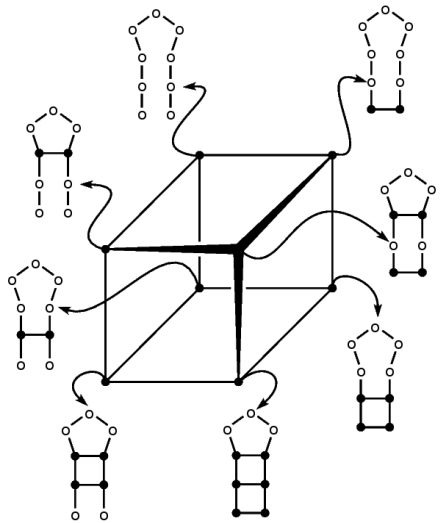
- e.g., all secondary structures that differ by one base pair



neighborhood graph + free energy per vertex = folding landscape

Neighborhood of RNA Secondary Structure

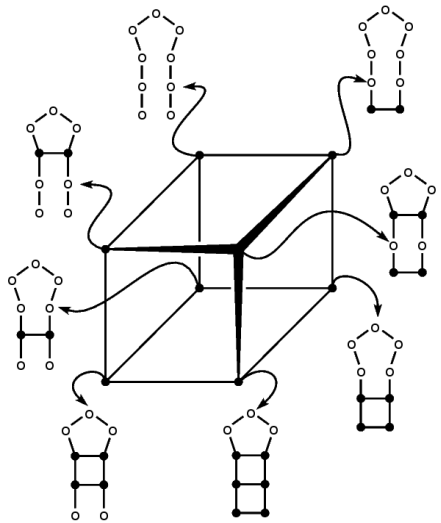
- e.g., all secondary structures that differ by one base pair



neighborhood graph + free energy per vertex = folding landscape

Neighborhood of RNA Secondary Structure

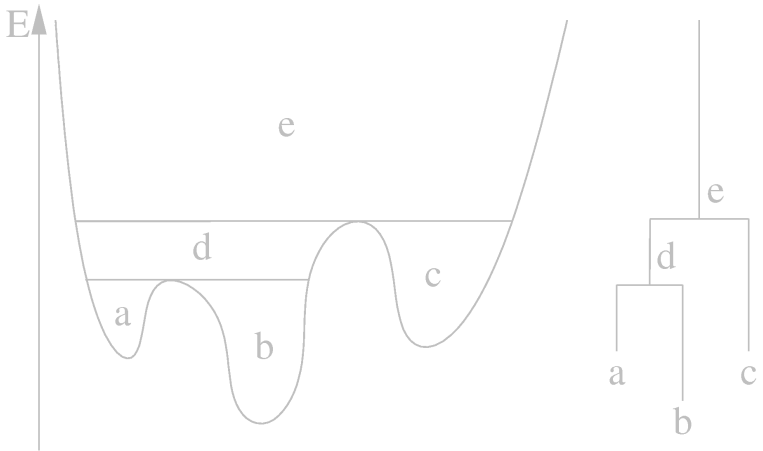
- e.g., all secondary structures that differ by one base pair



neighborhood graph + free energy per vertex = folding landscape

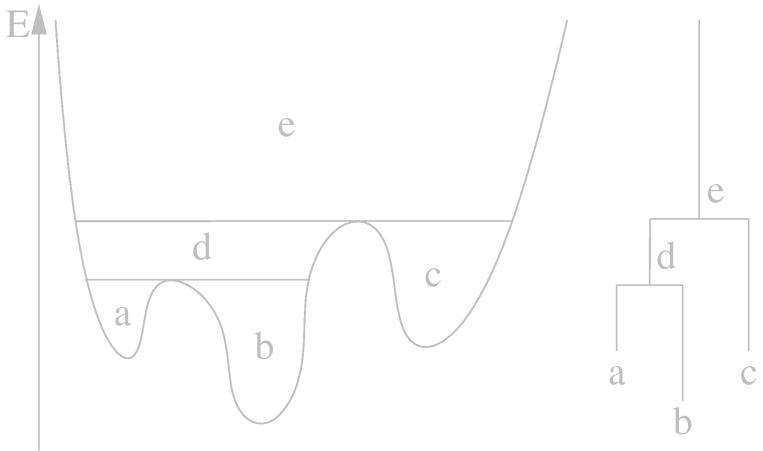
Barrier Trees

- topological simplification of folding landscapes
- reduction of the conformational space for kinetic folding simulation



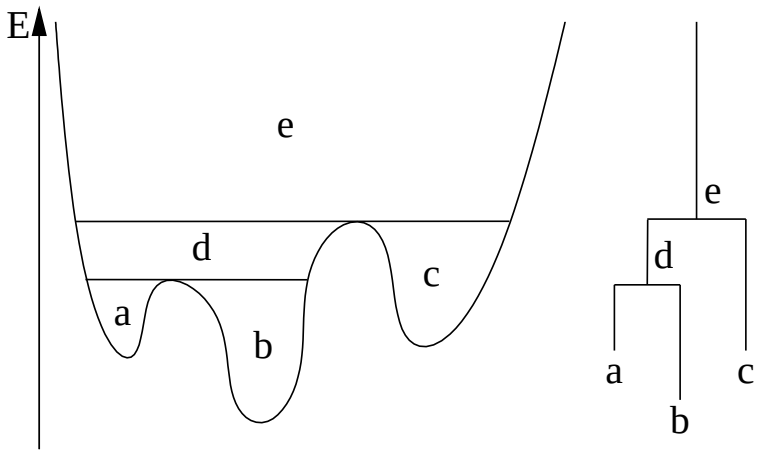
Barrier Trees

- topological simplification of folding landscapes
- reduction of the conformational space for kinetic folding simulation



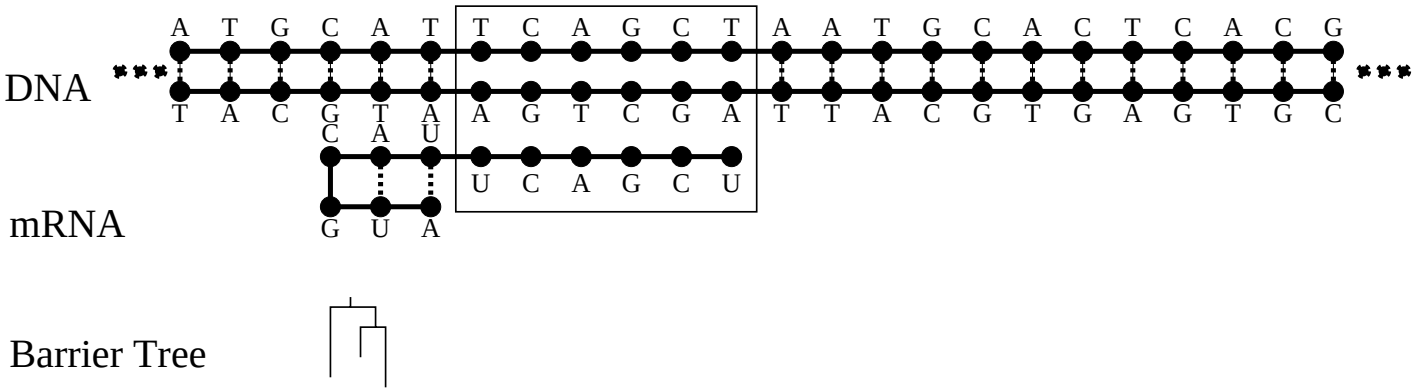
Barrier Trees

- topological simplification of folding landscapes
- reduction of the conformational space for kinetic folding simulation



Barrier Tree Sequences

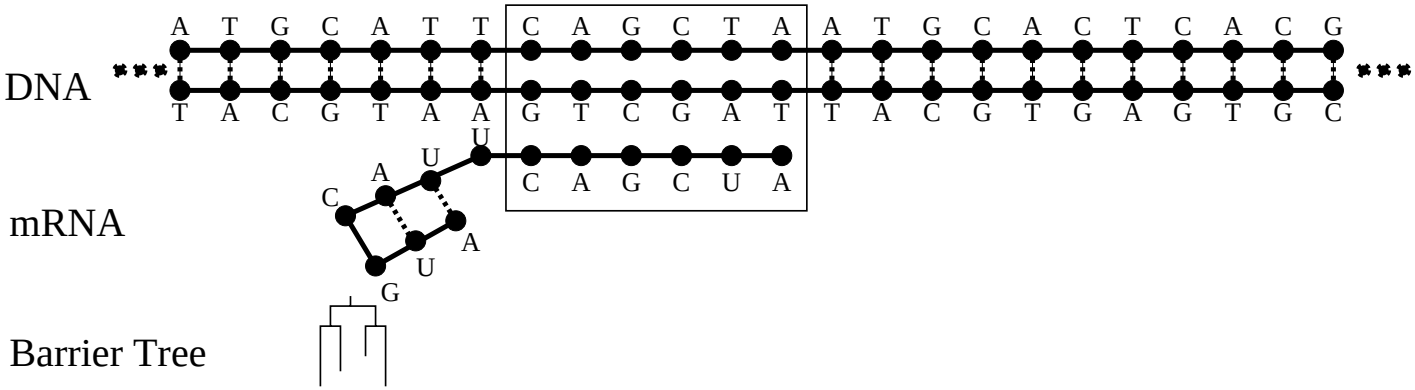
- RNA molecules fold while being transcribed



- assumed to have a strong impact on the spatial structure of short lived RNA molecules
- for each base sequence length there is a folding landscape, these are correlated
- barrier trees are correlated as well

Barrier Tree Sequences

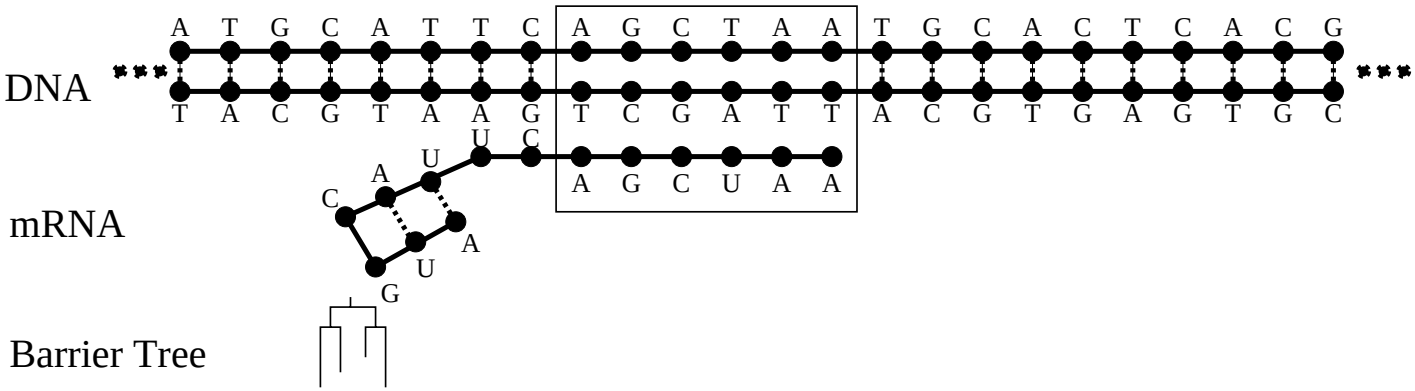
- RNA molecules fold while being transcribed



- assumed to have a strong impact on the spatial structure of short lived RNA molecules
- for each base sequence length there is a folding landscape, these are correlated
- barrier trees are correlated as well

Barrier Tree Sequences

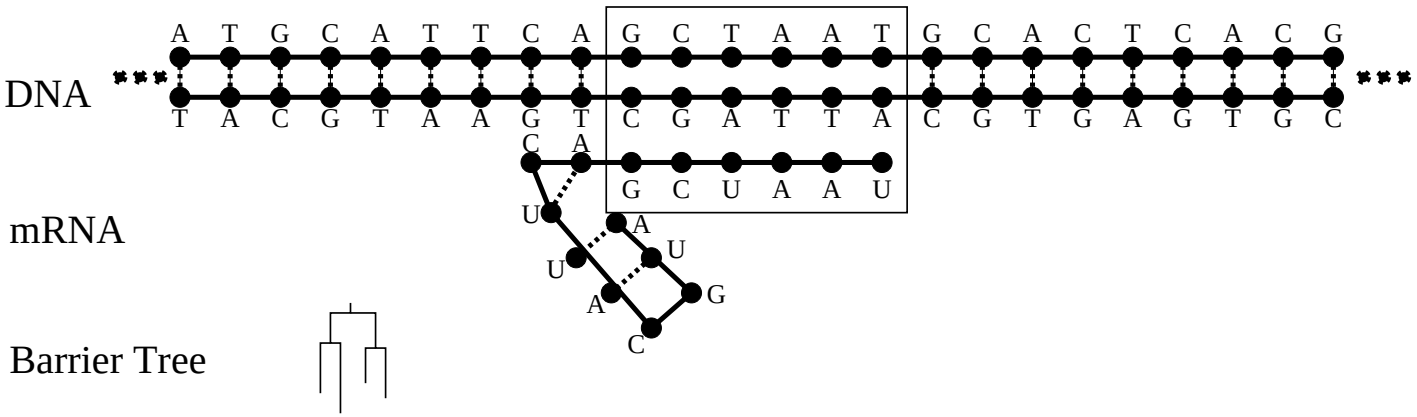
- RNA molecules fold while being transcribed



- assumed to have a strong impact on the spatial structure of short lived RNA molecules
- for each base sequence length there is a folding landscape, these are correlated
- barrier trees are correlated as well

Barrier Tree Sequences

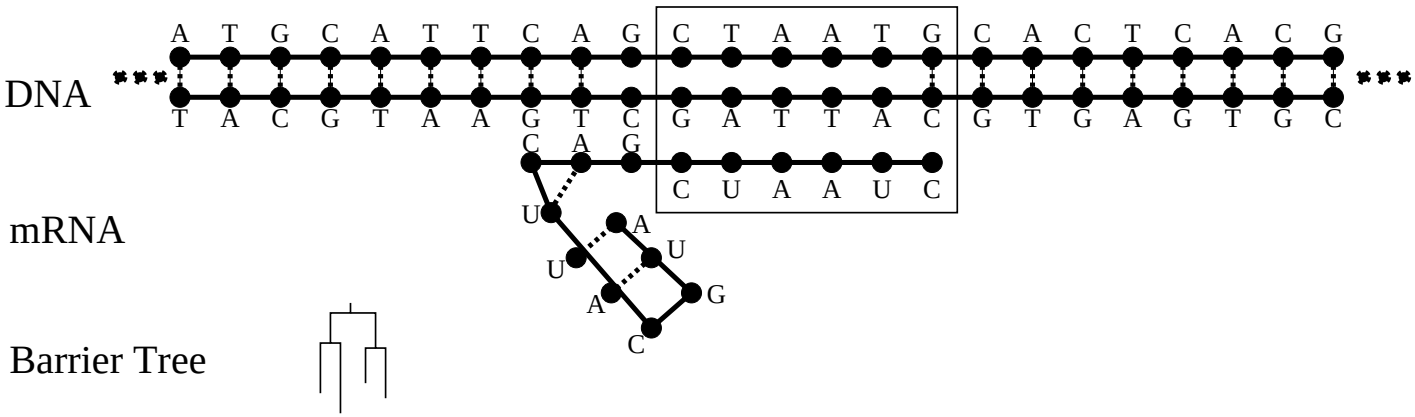
- RNA molecules fold while being transcribed



- assumed to have a strong impact on the spatial structure of short lived RNA molecules
- for each base sequence length there is a folding landscape, these are correlated
- barrier trees are correlated as well

Barrier Tree Sequences

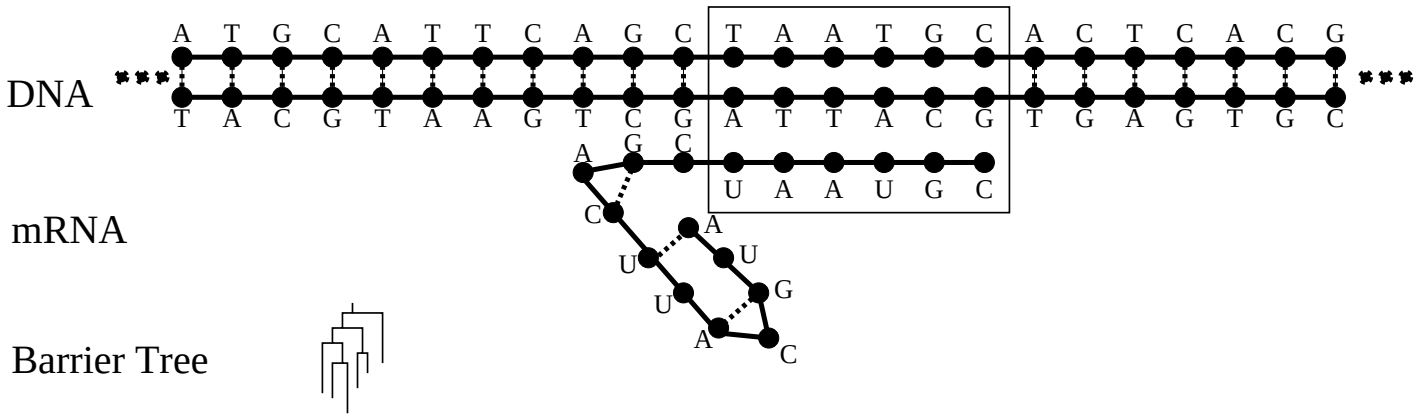
- RNA molecules fold while being transcribed



- assumed to have a strong impact on the spatial structure of short lived RNA molecules
- for each base sequence length there is a folding landscape, these are correlated
- barrier trees are correlated as well

Barrier Tree Sequences

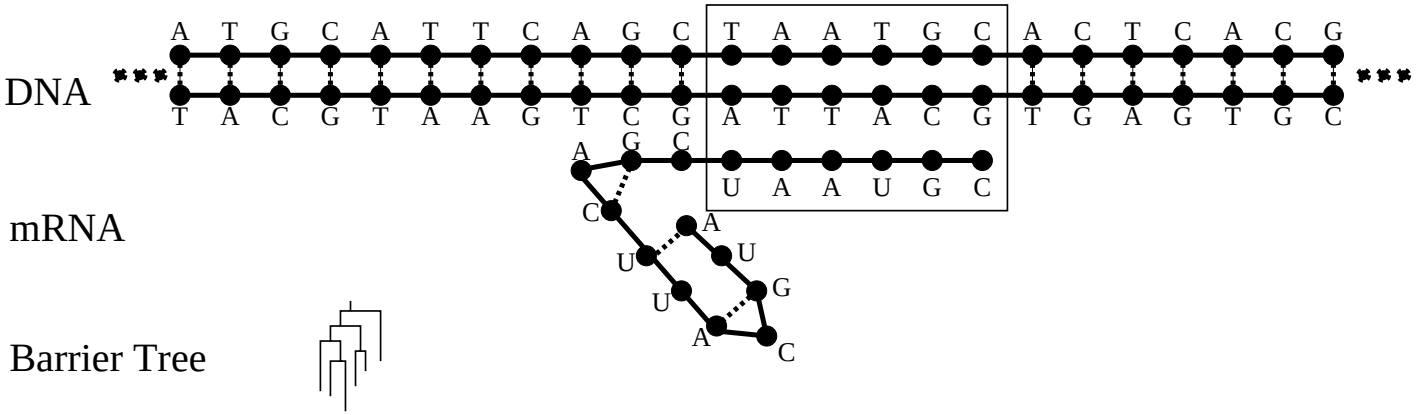
- RNA molecules fold while being transcribed



- assumed to have a strong impact on the spatial structure of short lived RNA molecules
- for each base sequence length there is a folding landscape, these are correlated
- barrier trees are correlated as well

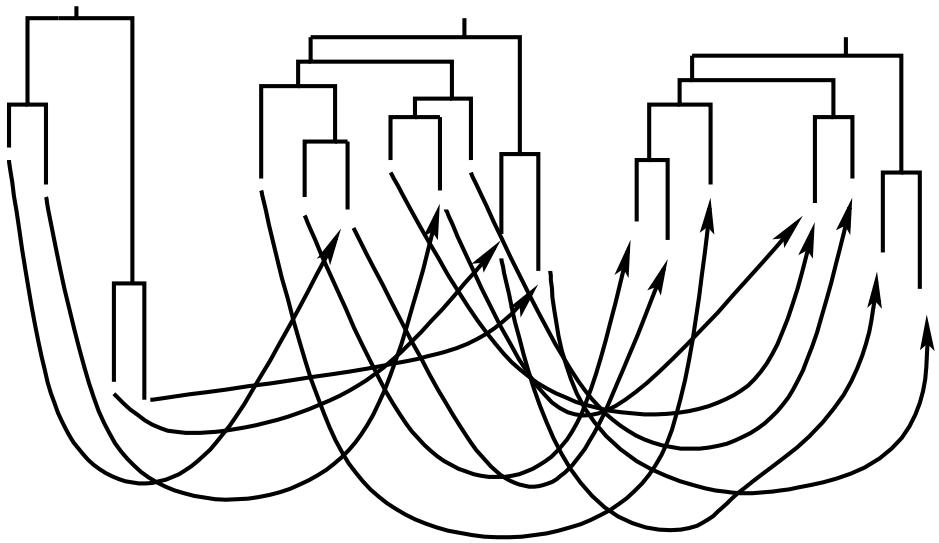
Barrier Tree Sequences

- RNA molecules fold while being transcribed



- assumed to have a strong impact on the spatial structure of short lived RNA molecules
- for each base sequence length there is a folding landscape, these are correlated
- barrier trees are correlated as well

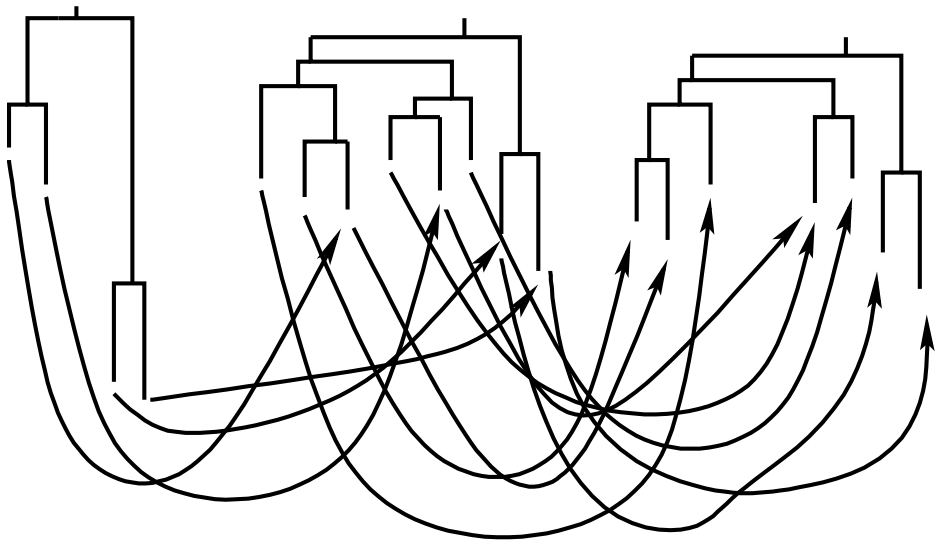
Barrier Tree Animation



. . . is a *dynamic graph drawing* problem

- it is not sufficient to layout and show each *barrier tree* independently
- in general, static aesthetic criteria fight dynamic aesthetic criteria

Barrier Tree Animation

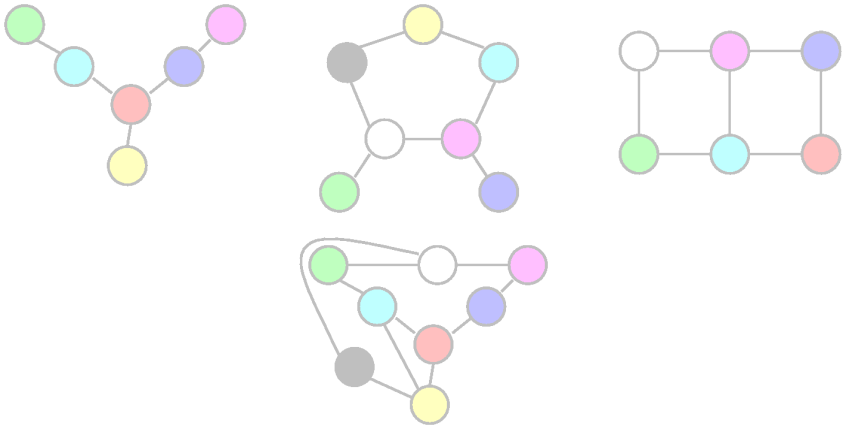


. . . is a *dynamic graph drawing* problem

- it is not sufficient to layout and show each *barrier tree* independently
- in general, static aesthetic criteria fight dynamic aesthetic criteria

Foresight Layout with Tolerance (Diehl and Görg '02)

$G_i = (V_i, E_i)$



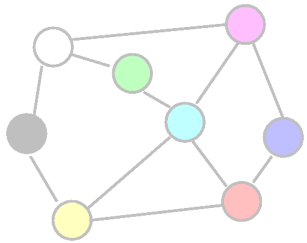
$G = (\bigcup V_i, \bigcup E_i)$

no layout

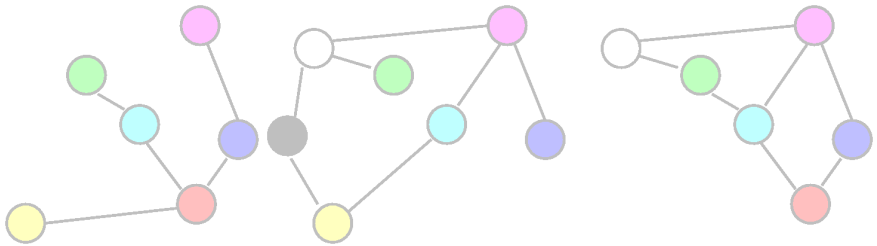
---

with layout

$L = L(G)$

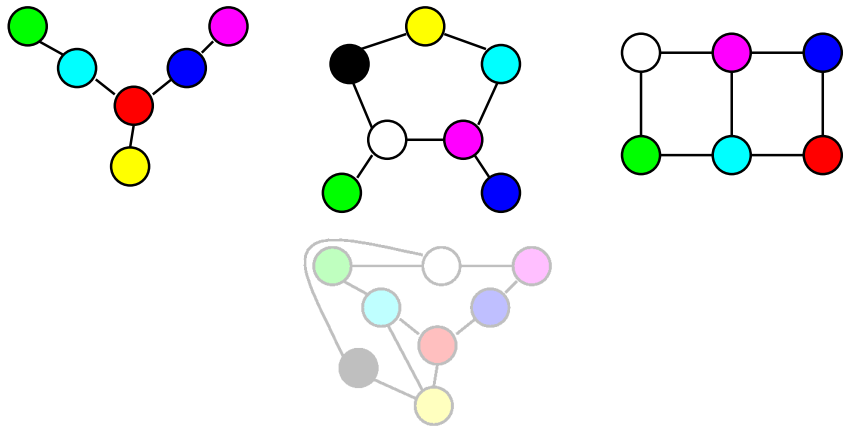


$L(G_i) = L(G)|_{G_i}$



Foresight Layout with Tolerance (Diehl and Görg '02)

$G_i = (V_i, E_i)$



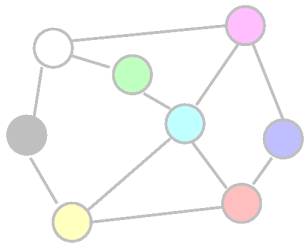
$G = (\bigcup V_i, \bigcup E_i)$

no layout

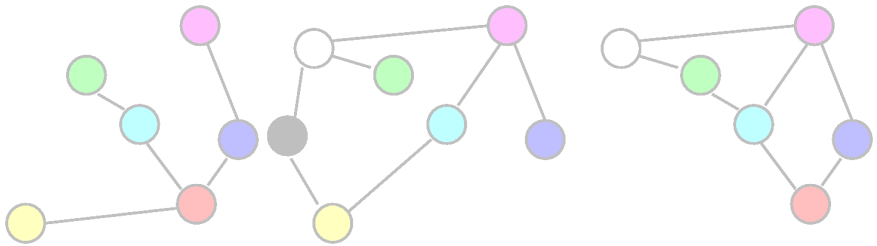
---

with layout

$L = L(G)$

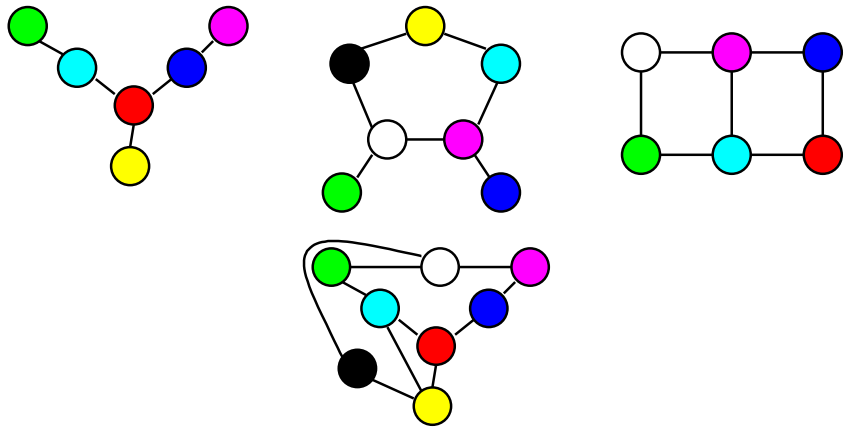


$L(G_i) = L(G)|_{G_i}$



Foresight Layout with Tolerance (Diehl and Görg '02)

$G_i = (V_i, E_i)$



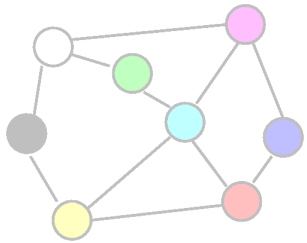
$G = (\bigcup V_i, \bigcup E_i)$

no layout

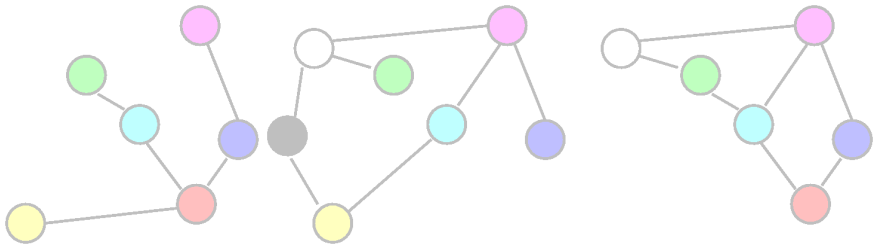
---

with layout

$L = L(G)$

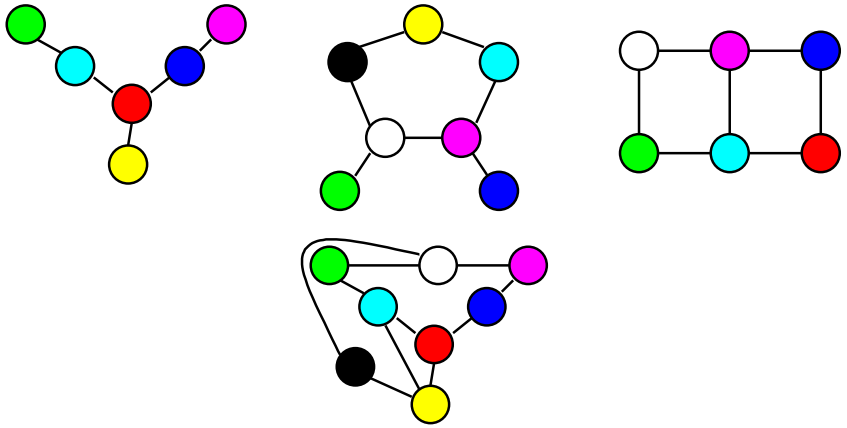


$L(G_i) = L(G)|_{G_i}$



Foresight Layout with Tolerance (Diehl and Görg '02)

$G_i = (V_i, E_i)$



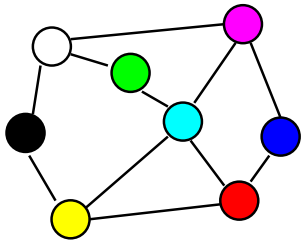
$G = (\bigcup V_i, \bigcup E_i)$

no layout

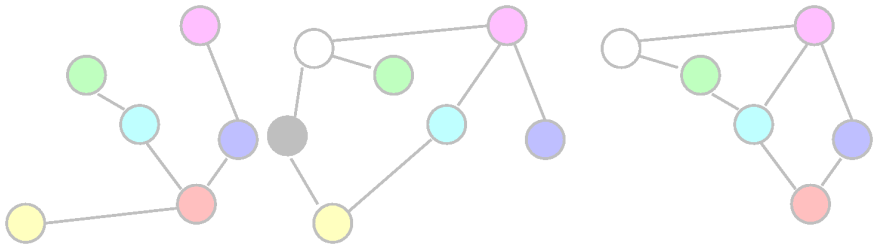
---

with layout

$L = L(G)$

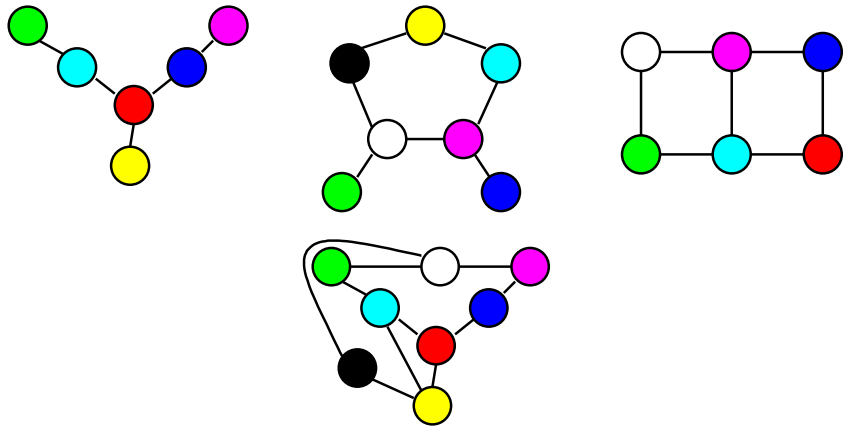


$L(G_i) = L(G)|_{G_i}$



Foresight Layout with Tolerance (Diehl and Görg '02)

$G_i = (V_i, E_i)$



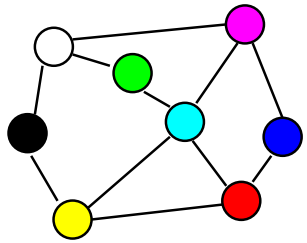
$G = (\bigcup V_i, \bigcup E_i)$

no layout

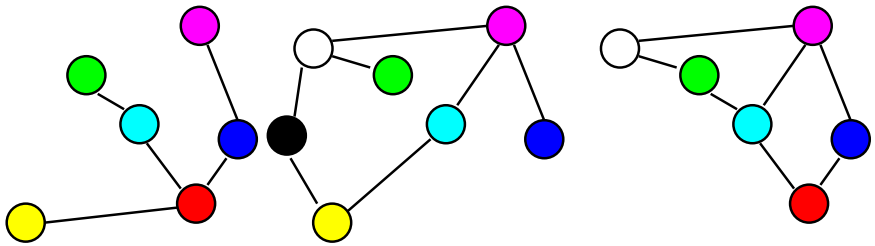
---

with layout

$L = L(G)$



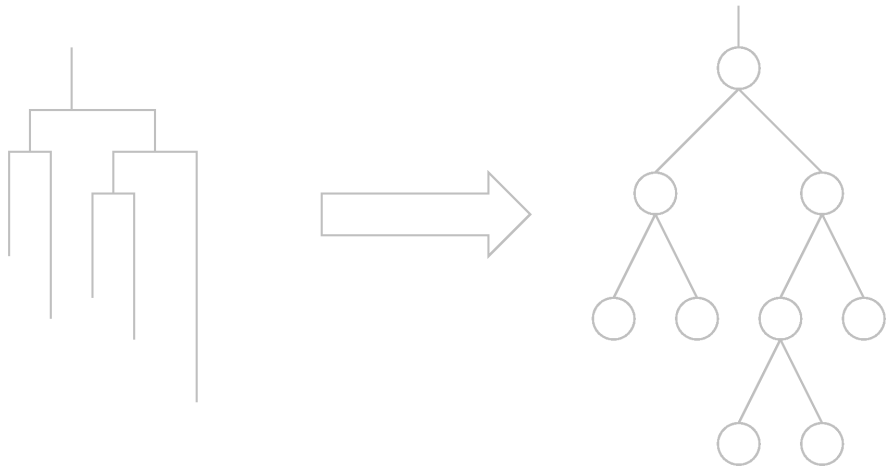
$L(G_i) = L(G)|_{G_i}$



What is wrong with the Supergraph Construction?

algorithm overview:

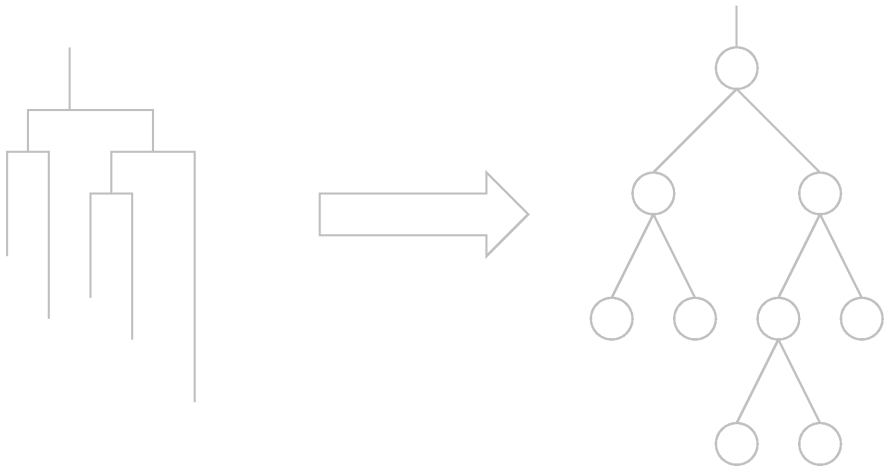
- preserve only father-son-relations of *barrier trees*



What is wrong with the Supergraph Construction?

algorithm overview:

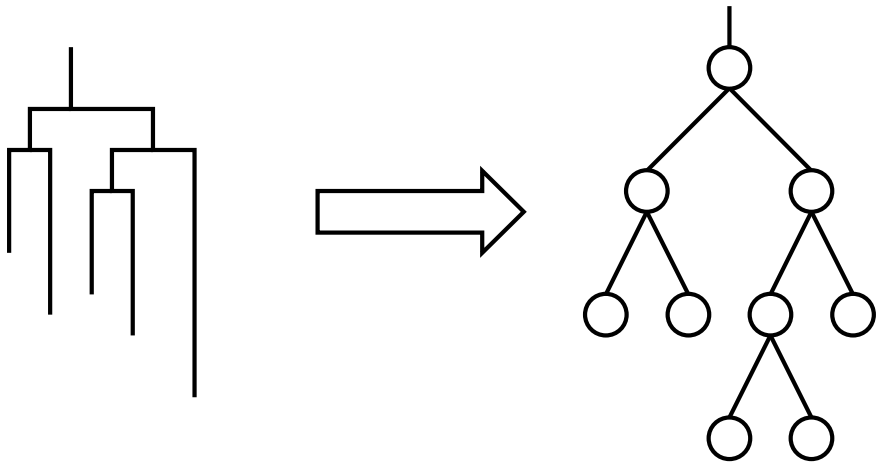
- preserve only father-son-relations of *barrier trees*



What is wrong with the Supergraph Construction?

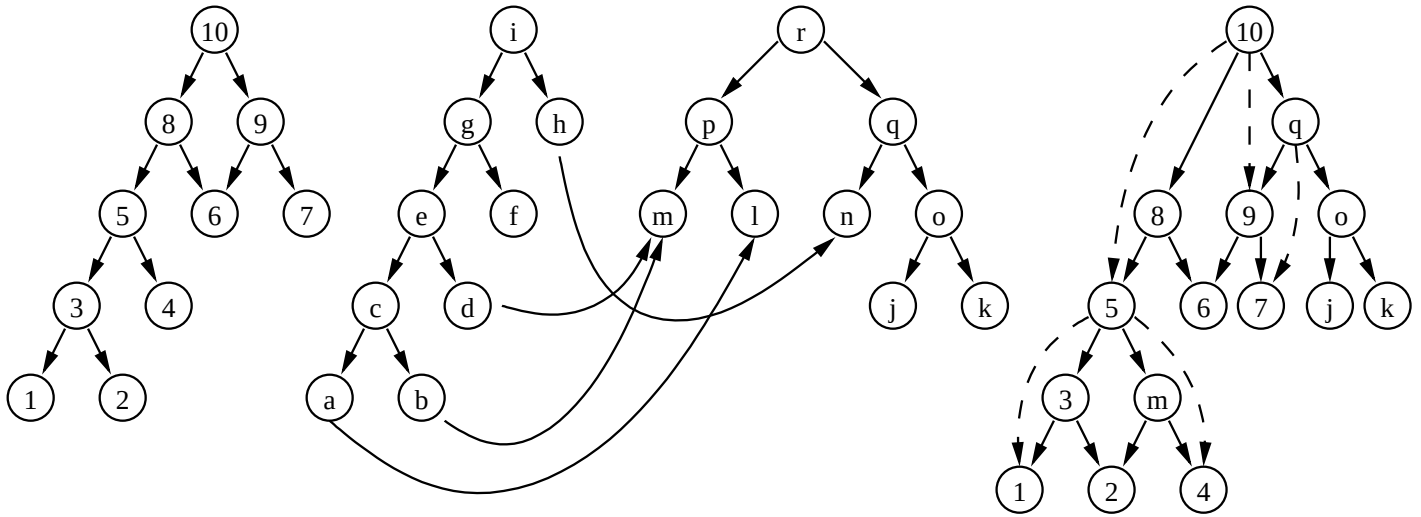
algorithm overview:

- preserve only father-son-relations of *barrier trees*



Supergraph Construction

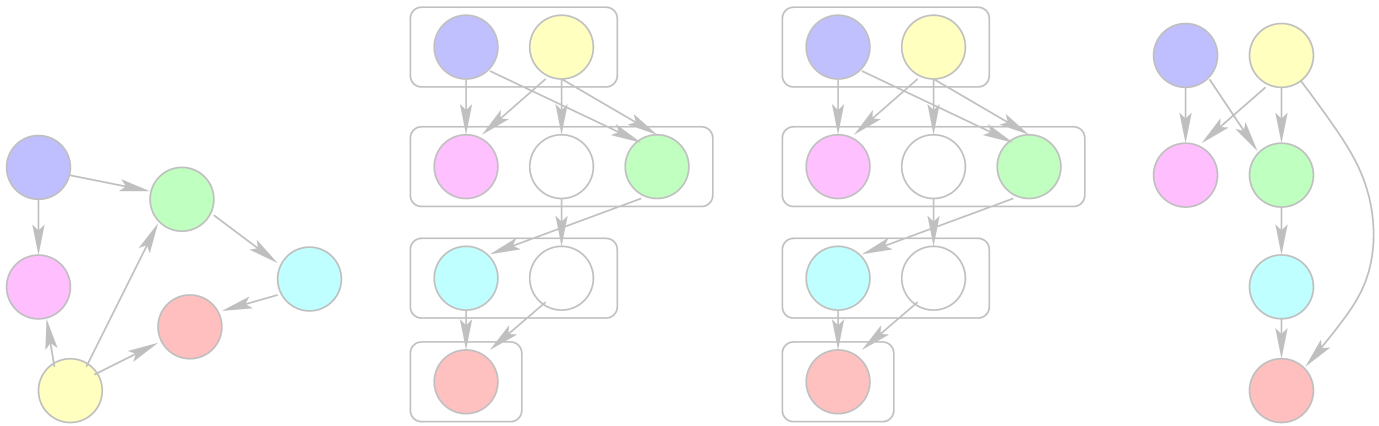
- "set of leaves" heuristic (see Heine, Scheuermann, Flamm, Hofacker, and Stadler, *Visualization of Barrier Tree Sequences*, IEEE Transactions on Visualization and Computer Graphics, 12(5), 2006, 781-788.



supergraph n tree n tree $n+1$ supergraph $n+1$

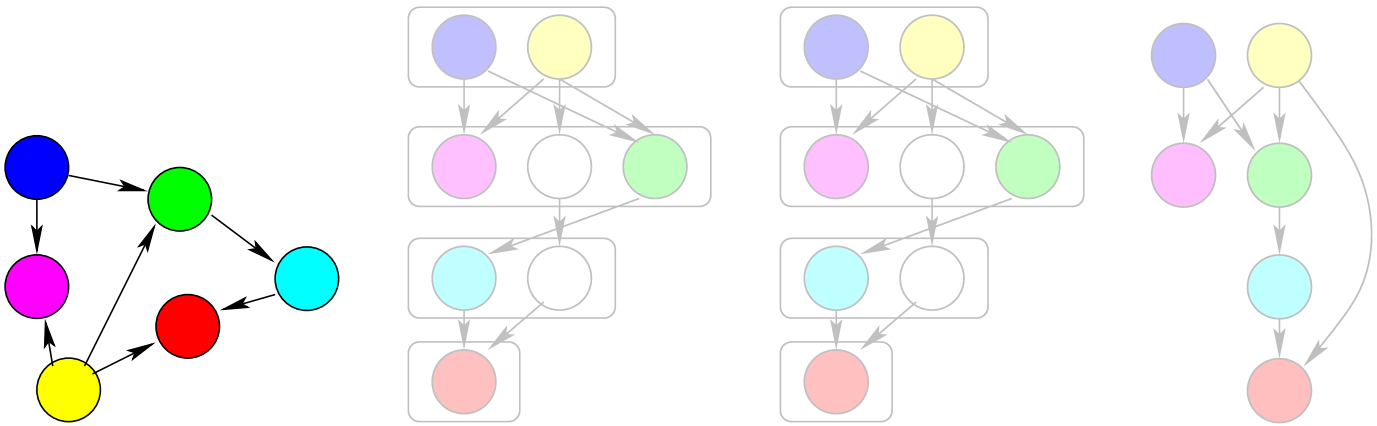
Layout

Supergraph layout uses the DOT algorithm (Gansner et al. ’93)
Sugiyama type DAG Layout



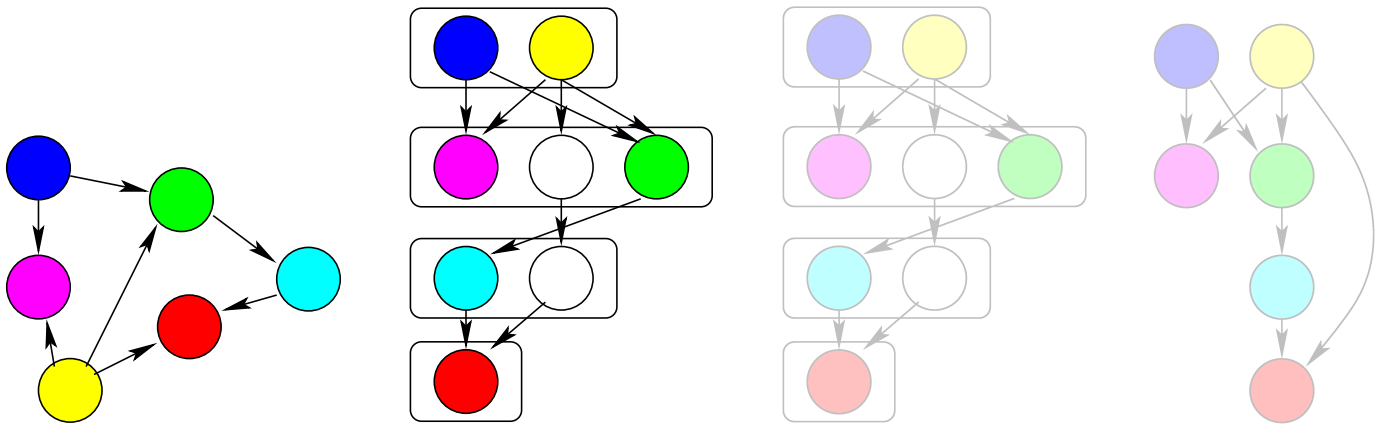
Layout

Supergraph layout uses the DOT algorithm (Gansner et al. ’93)
Sugiyama type DAG Layout



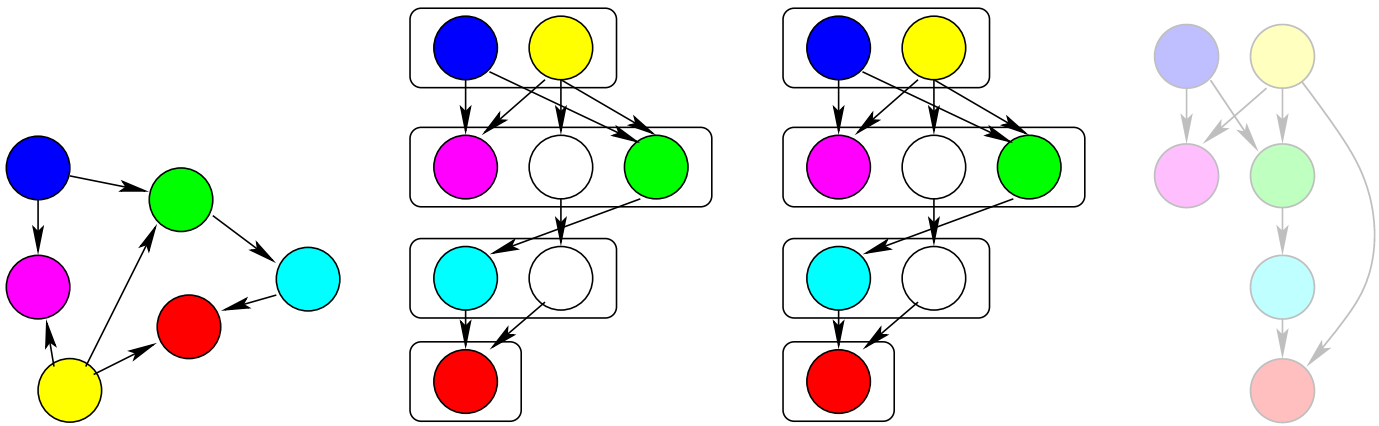
Layout

Supergraph layout uses the DOT algorithm (Gansner et al. ’93)
Sugiyama type DAG Layout



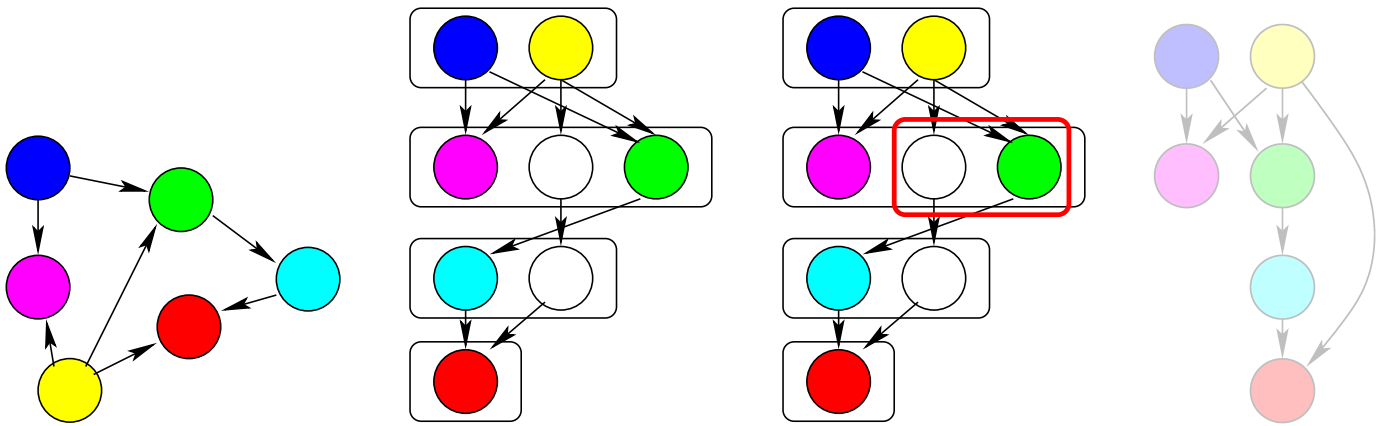
Layout

Supergraph layout uses the DOT algorithm (Gansner et al. ’93)
Sugiyama type DAG Layout



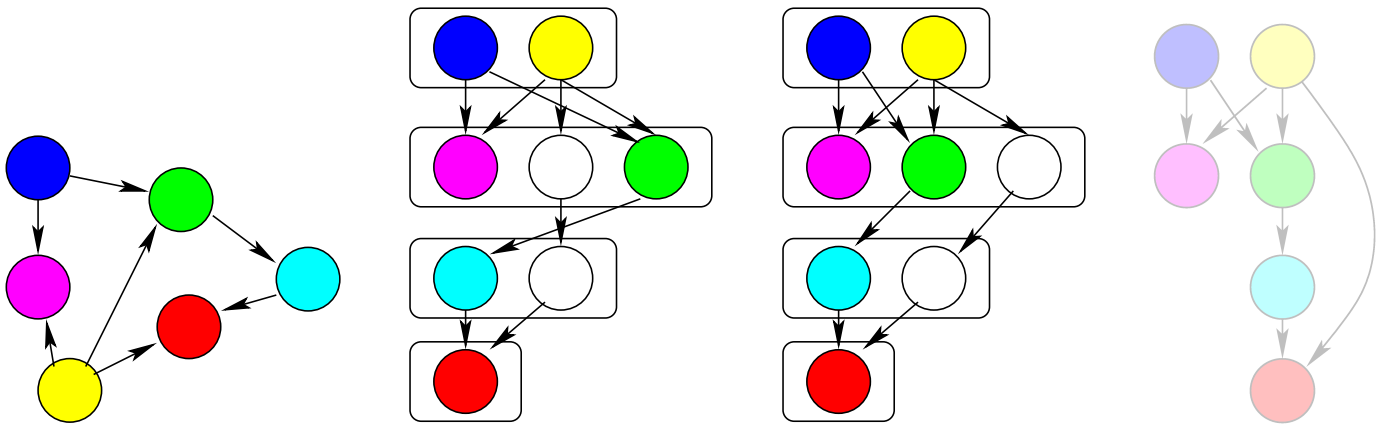
Layout

Supergraph layout uses the DOT algorithm (Gansner et al. ’93)
Sugiyama type DAG Layout



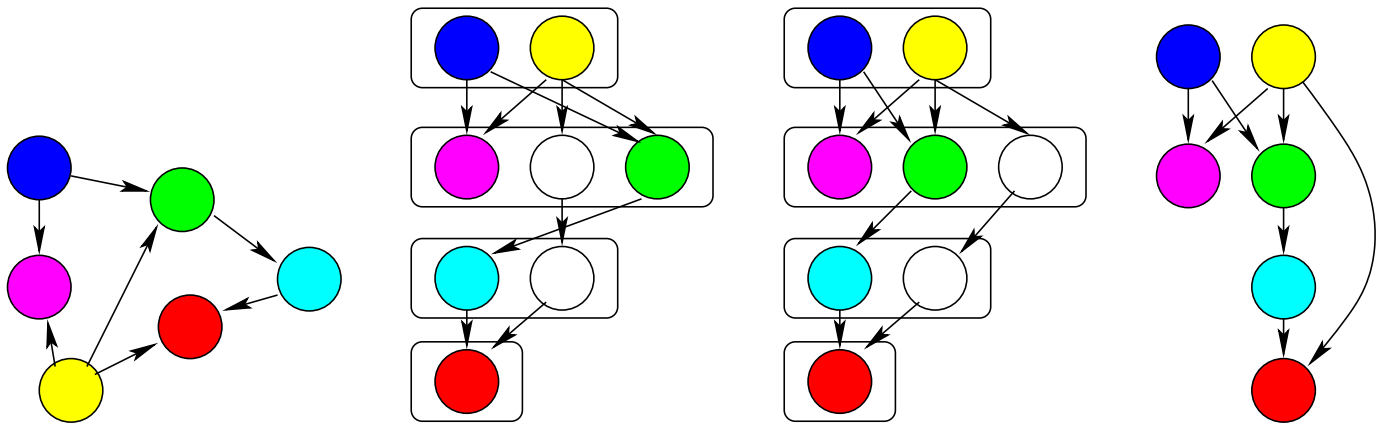
Layout

Supergraph layout uses the DOT algorithm (Gansner et al. ’93)
Sugiyama type DAG Layout

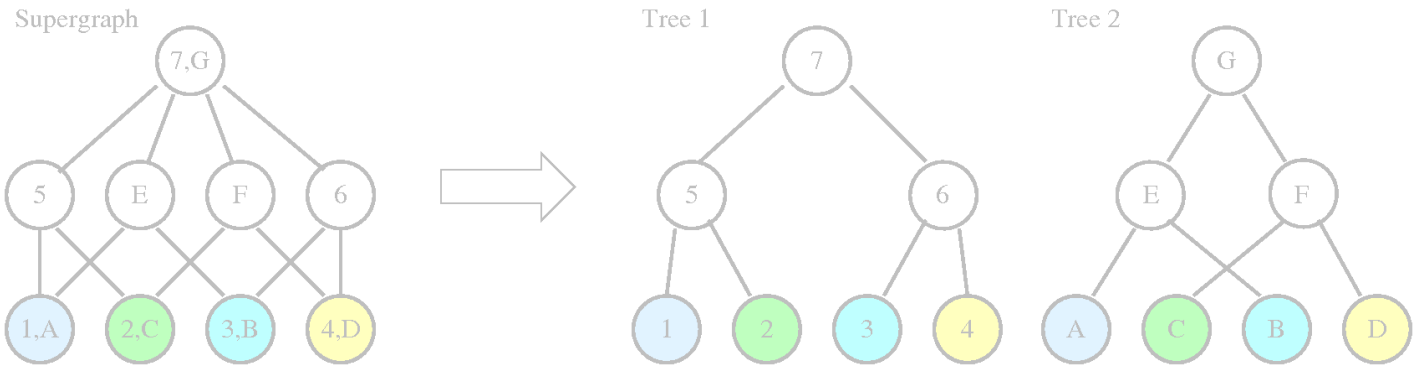


Layout

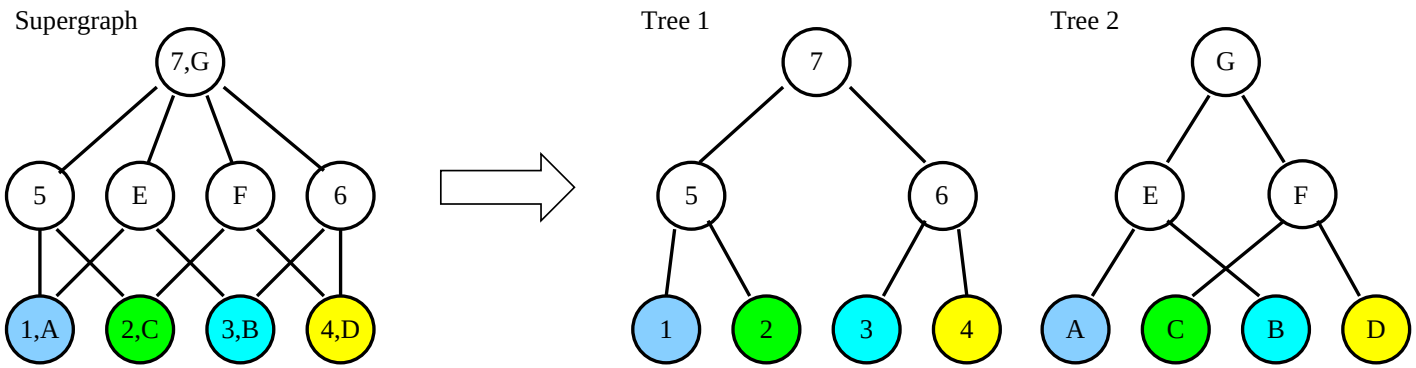
Supergraph layout uses the DOT algorithm (Gansner et al. ’93)
Sugiyama type DAG Layout



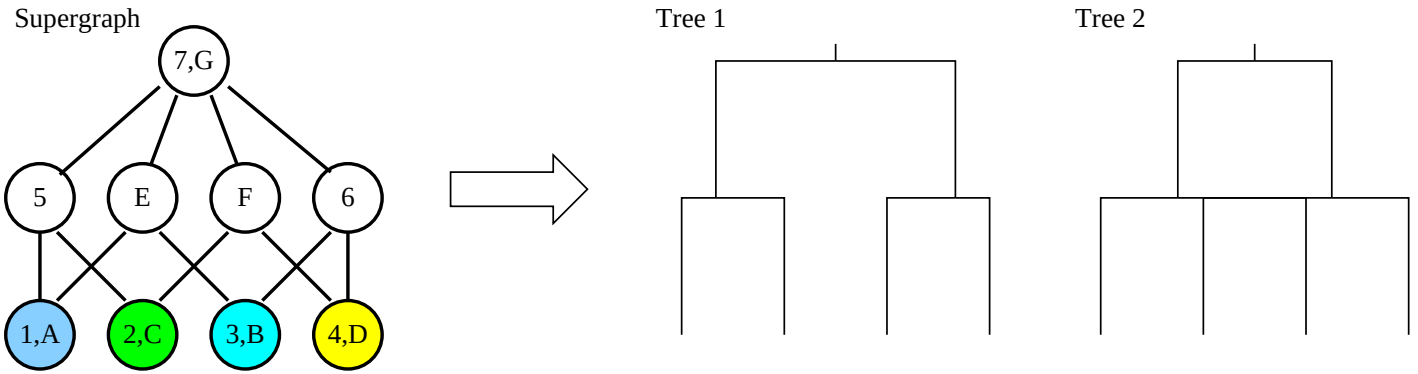
Edge Crossings



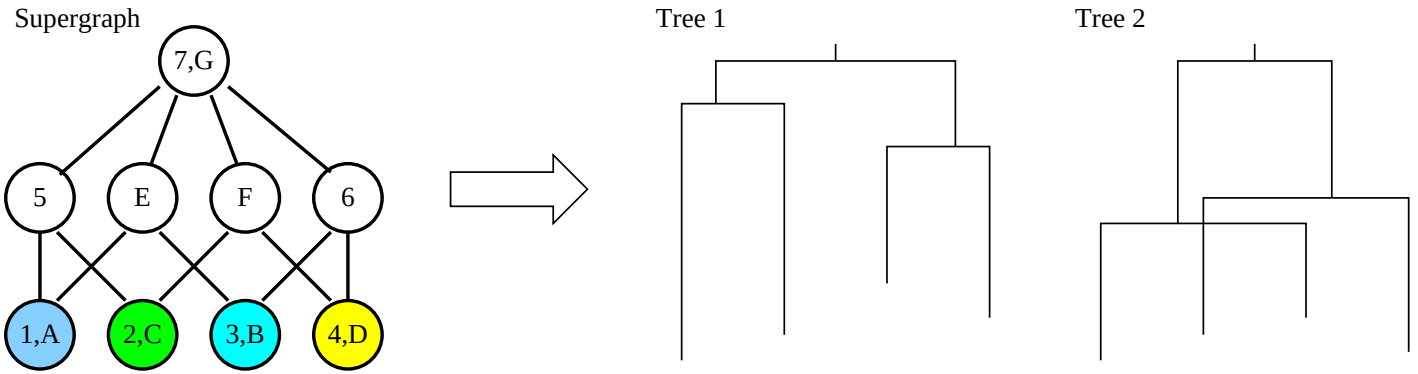
Edge Crossings



Edge Crossings



Edge Crossings



New Layout Algorithm

- identification of equivalent vertices using “set of leaves” approach
- find a permutation of the equivalence classes that makes all trees of the sequence “nice”, formally:

$$\sigma = \underset{O}{\mathit{argmin}} \sum_{i=1}^{N} \left(\alpha \cdot \mathit{crossings}(T_i, O_{V_i}) + \beta \cdot \mathit{localorder}(T_i, O_{V_i}) \right)$$

New Layout Algorithm

- find the solution by repetition of the following:
  - randomly switching two vertices in the permutation,
  - reevaluation cost function $C(\sigma)$
  - keep new permutation with

$$p = \begin{cases} 1 & \Delta C < 0 \\ \frac{1}{1+exp(\Delta C\, T_t^{-1})} & \Delta C \geq 0 \end{cases} \qquad T_t = \frac{n_t - t}{t}$$

- just reducing crossings is not enough
- nice side effect: distribute vertices more uniformly on the screen

Results

- runtime and quality improved greatly for HOK and LEPTO datasets

	equivalence classes	runtime old (1000 iterations)	runtime new (100000 iteration)
ATT	381	37s	27s
LEPTO	1531	3462s	79s
HOK	3793	16790s	182s

Conclusions and Future Work

Conclusion

- strong improvement quality- and runtime-wise over old method

Future Work

- determine optimal number of iterations automatically

more Information:

Christian Heine, Gerik Scheuermann, Christoph Flamm, Ivo L. Hofacker, and Peter F. Stadler, *Visualization of Barrier Tree Sequences Revisited*, to appear in Proceedings of Visualization in Medicine and Life Sciences (VLMS), 2007

Thank You!