Halving Natural Genomes

ILW Master Thesis Award 2022

Leonard Bohnenkämper

Bielefeld University

September 16, 2022

Computational Comparative Genomics

- Differences and similarities of genomes on a high level

Computational Comparative Genomics

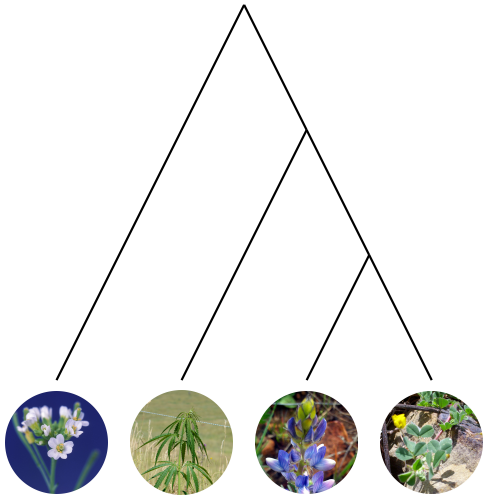
- Differences and similarities of genomes on a high level
  - Evolution

Computational Comparative Genomics

- Differences and similarities of genomes on a high level
  - Evolution
    - Rearrangements

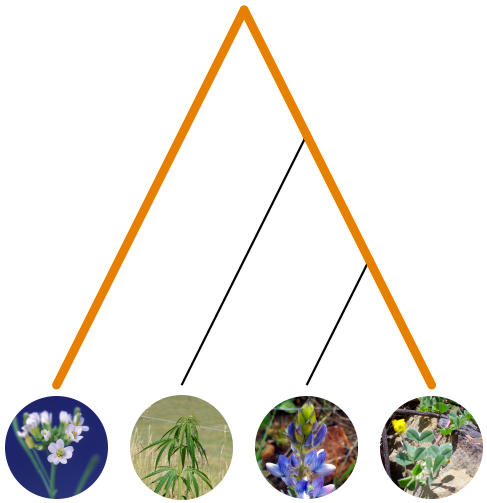
Computational Comparative Genomics

- Differences and similarities of genomes on a high level
  - Evolution
    - Rearrangements
- Classic Problems:



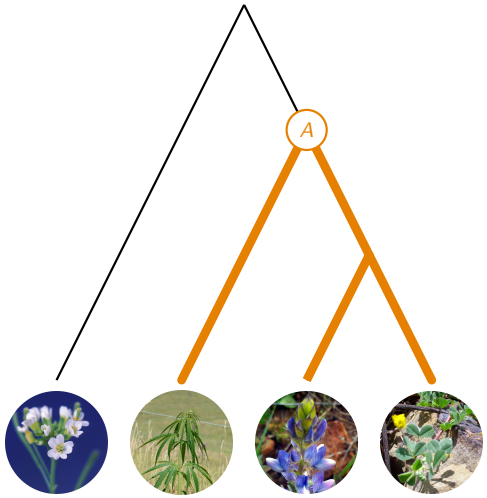
Computational Comparative Genomics

- Differences and similarities of genomes on a high level
  - Evolution
    - Rearrangements
- Classic Problems:
  - Distance



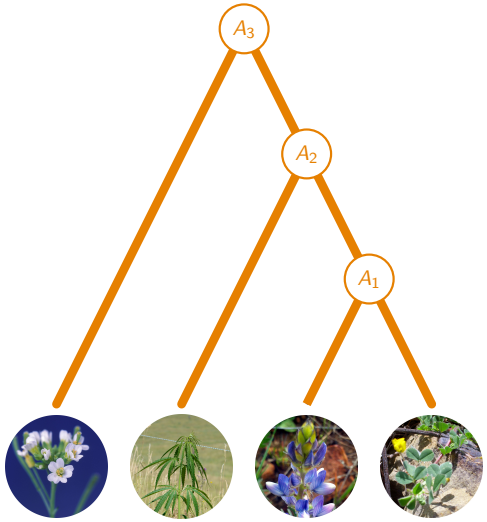
Computational Comparative Genomics

- Differences and similarities of genomes on a high level
  - Evolution
    - Rearrangements
- Classic Problems:
  - Distance
  - Ancestor



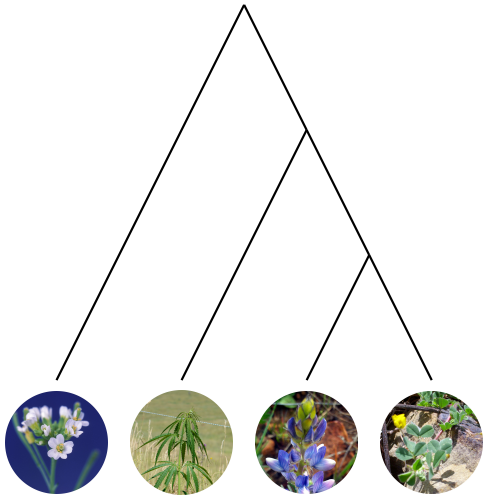
Computational Comparative Genomics

- Differences and similarities of genomes on a high level
  - Evolution
    - Rearrangements
- Classic Problems:
  - Distance
  - Ancestor
  - Phylogeny



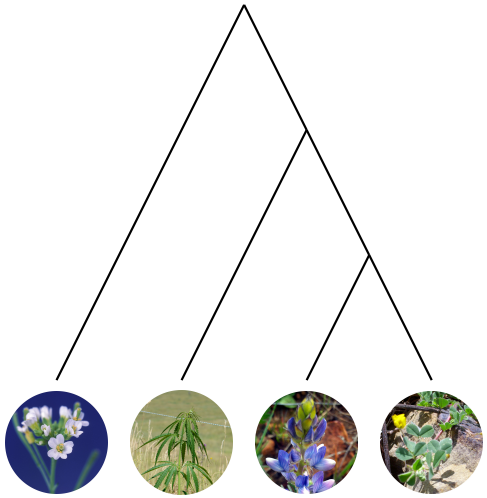
Computational Comparative Genomics

- Differences and similarities of genomes on a high level
  - Evolution
    - Rearrangements
- Classic Problems:
  - Distance
  - Ancestor
  - Phylogeny
  - ...



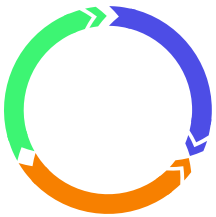
Computational Comparative Genomics

- Differences and similarities of genomes on a high level
  - Evolution
    - Rearrangements
- Classic Problems:
  - Distance
  - Ancestor
  - Phylogeny
  - ...



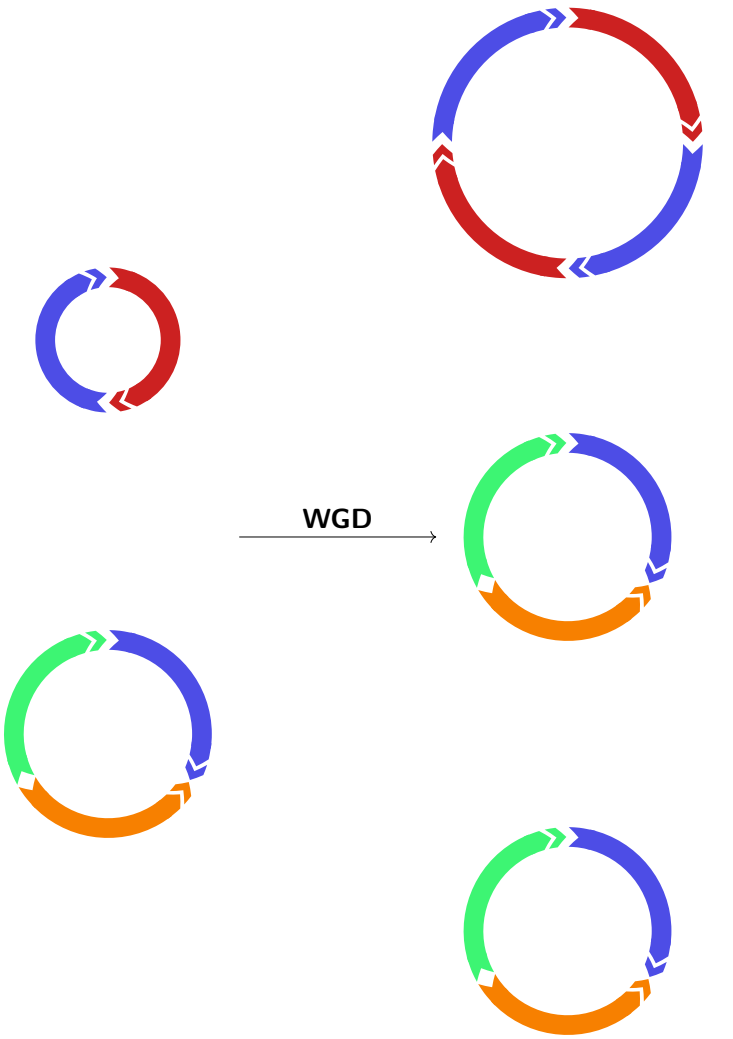
- Abstract from sequence content: Genes or other markers

The Genome Halving Problem

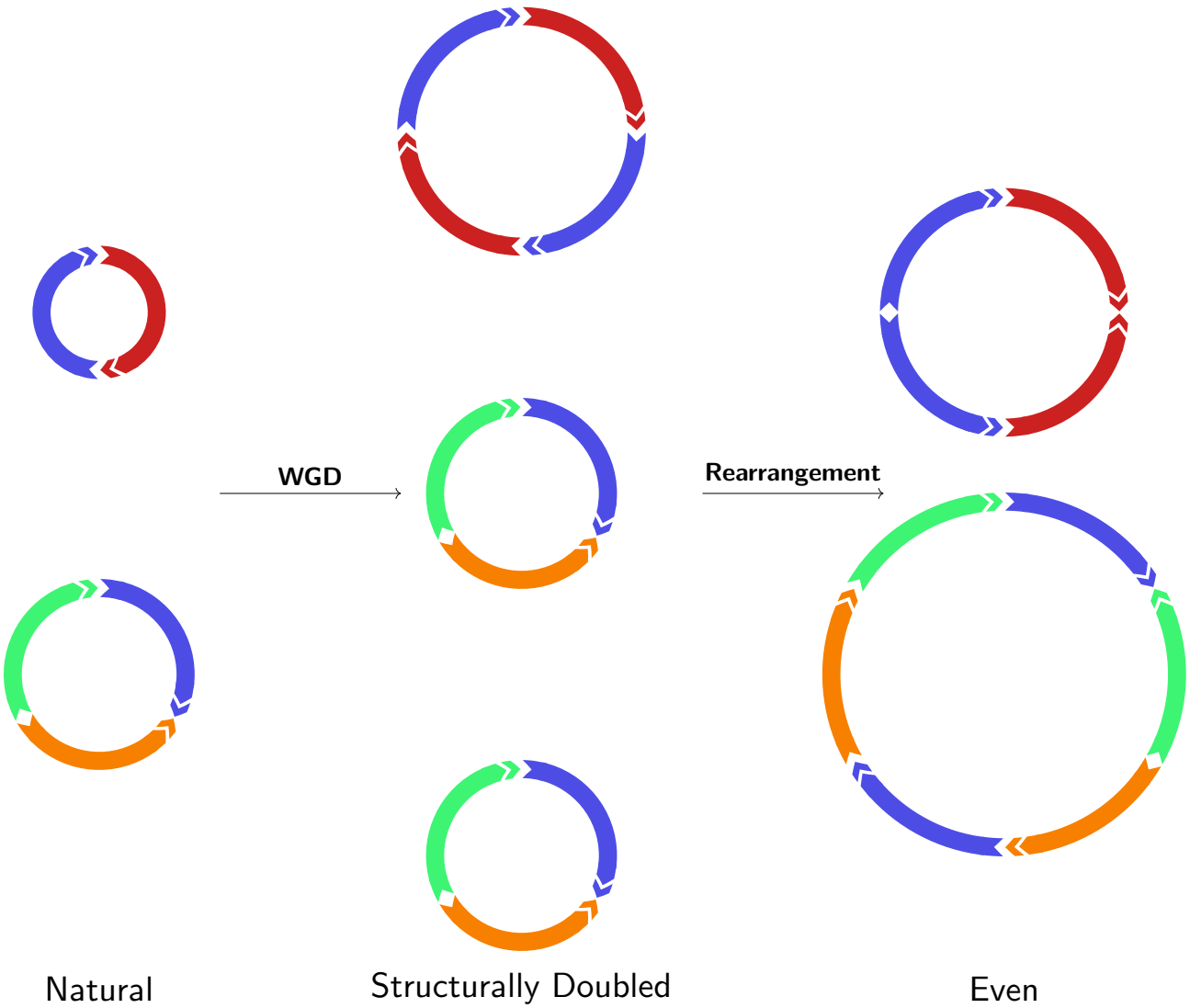


Natural

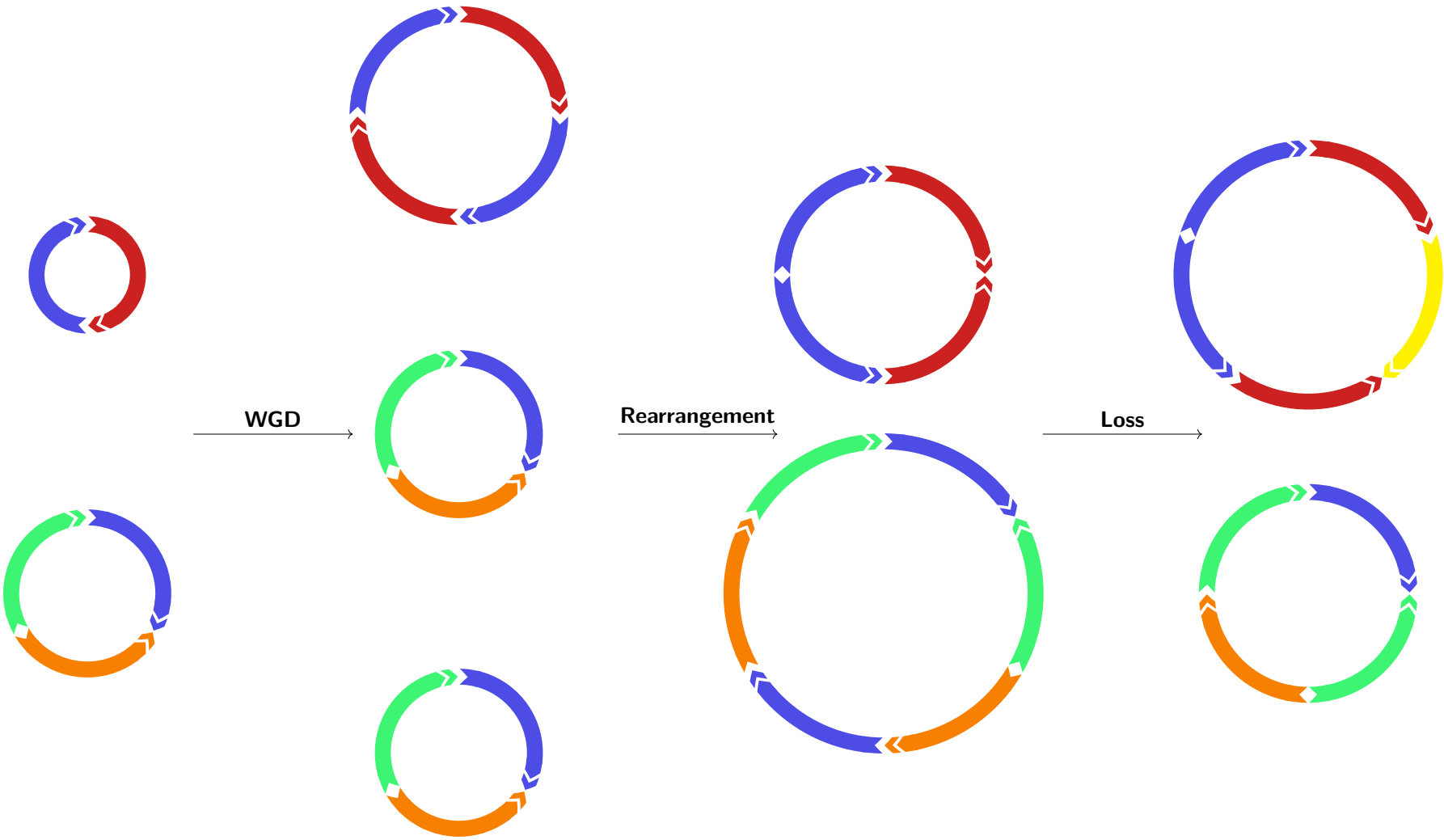
The Genome Halving Problem



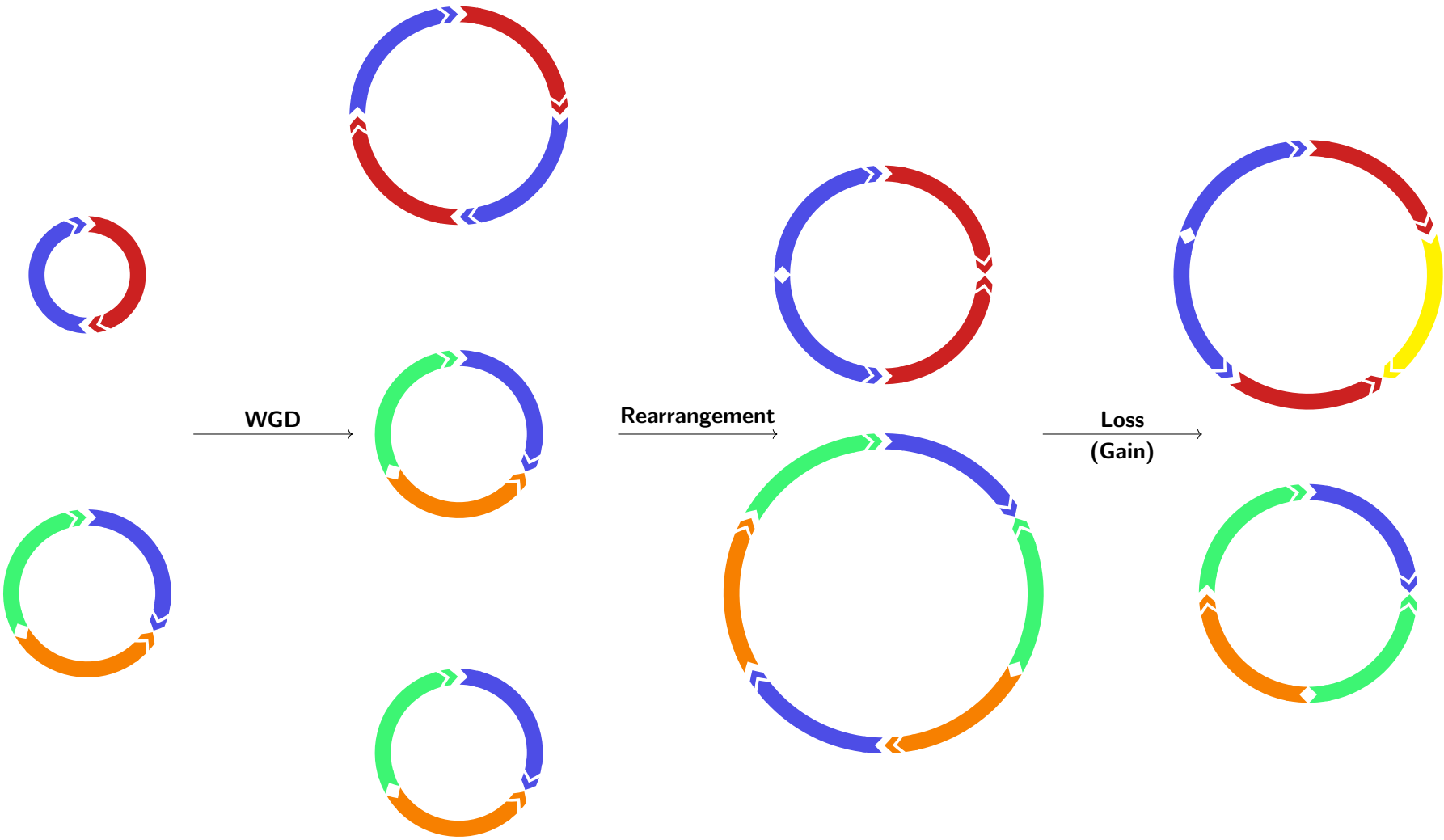
The Genome Halving Problem



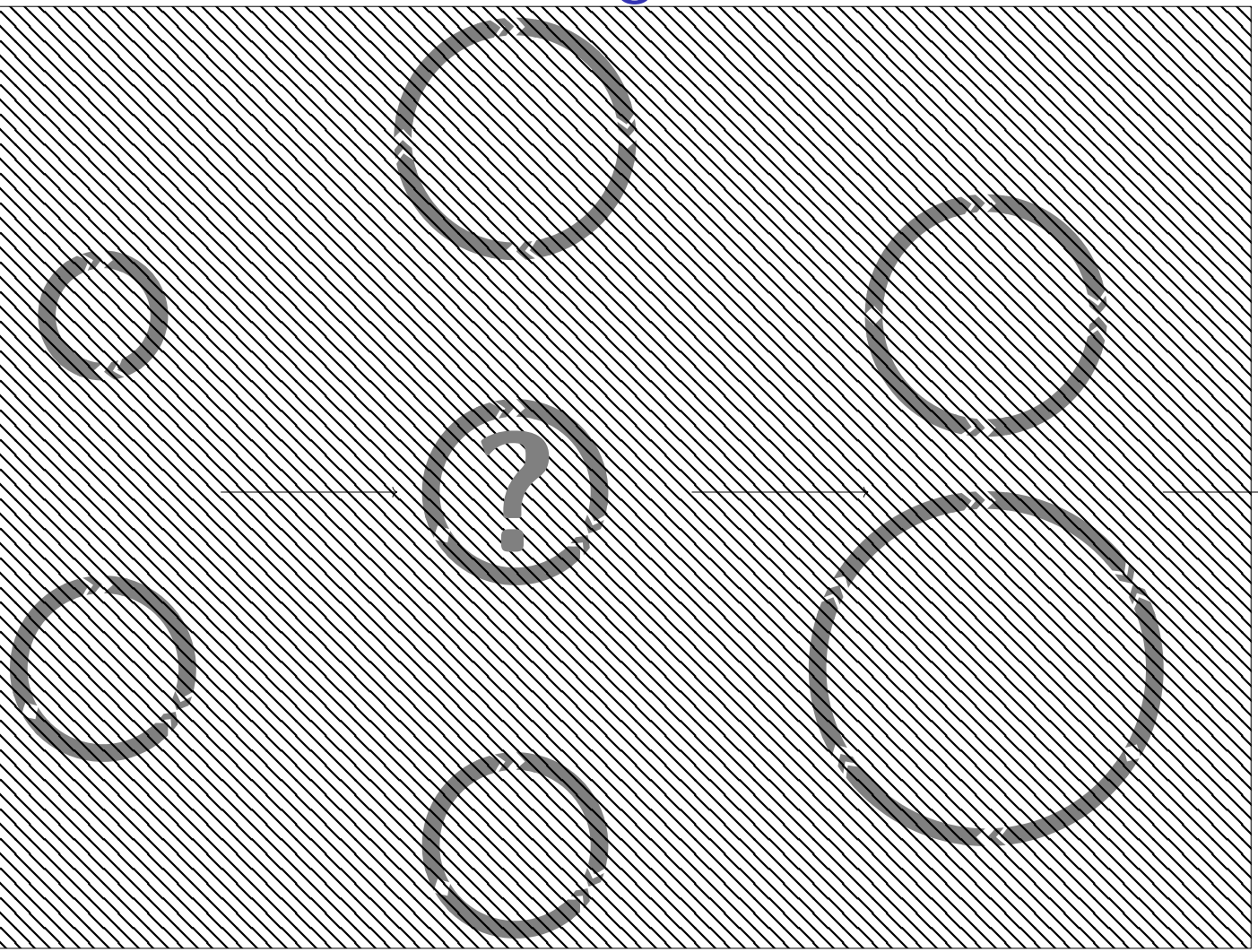
The Genome Halving Problem



The Genome Halving Problem

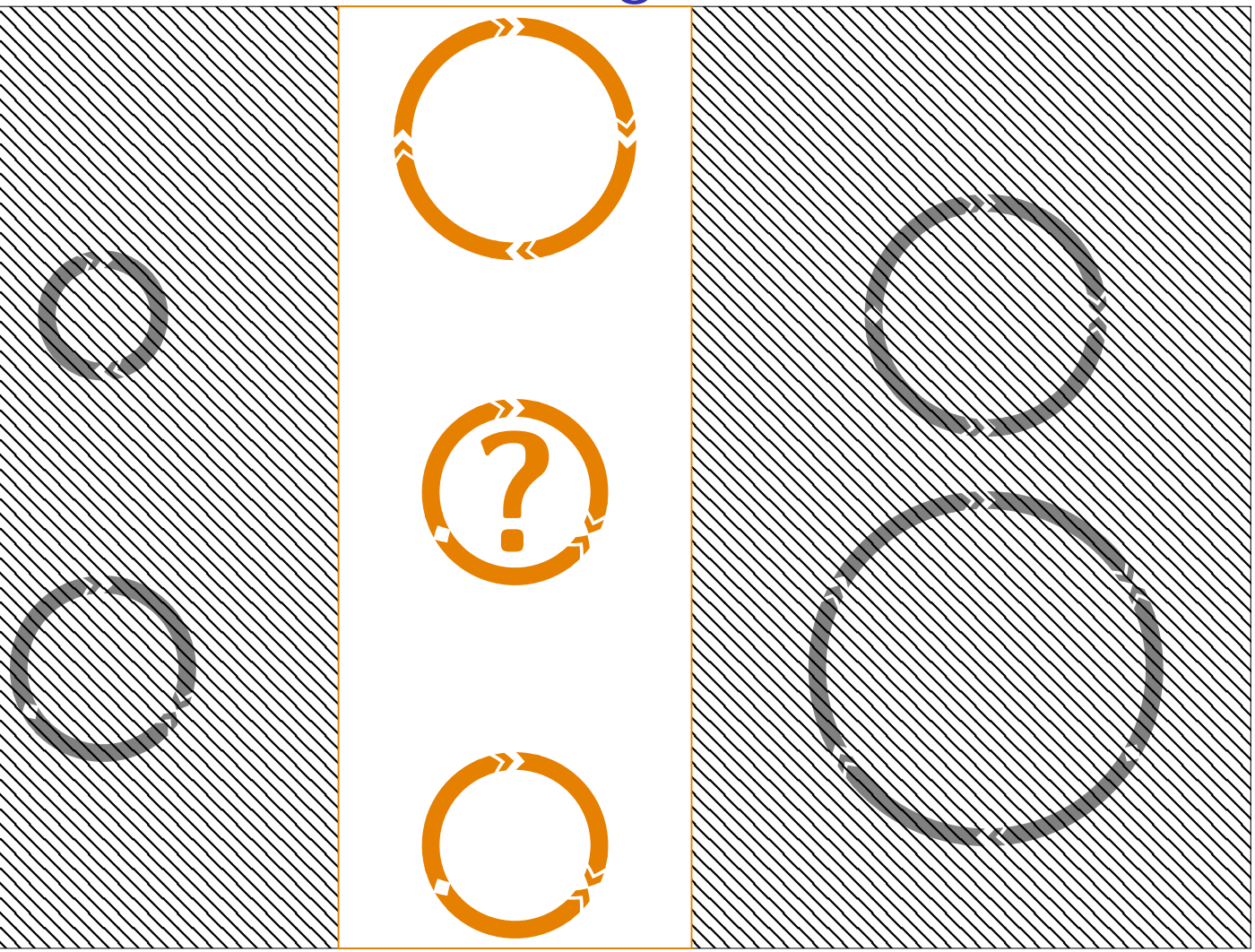


The Genome Halving Problem

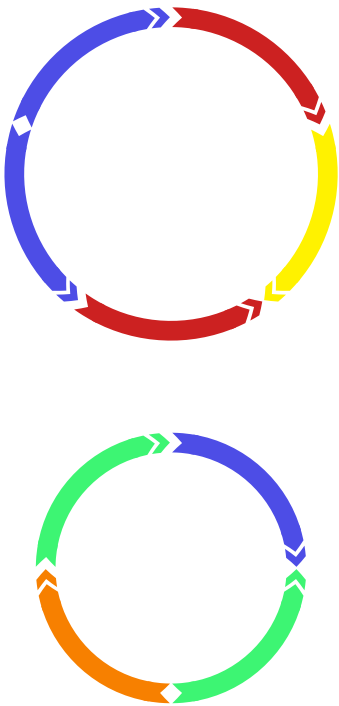


Natural | Structurally Doubled | Even | Natural

The Genome Halving Problem

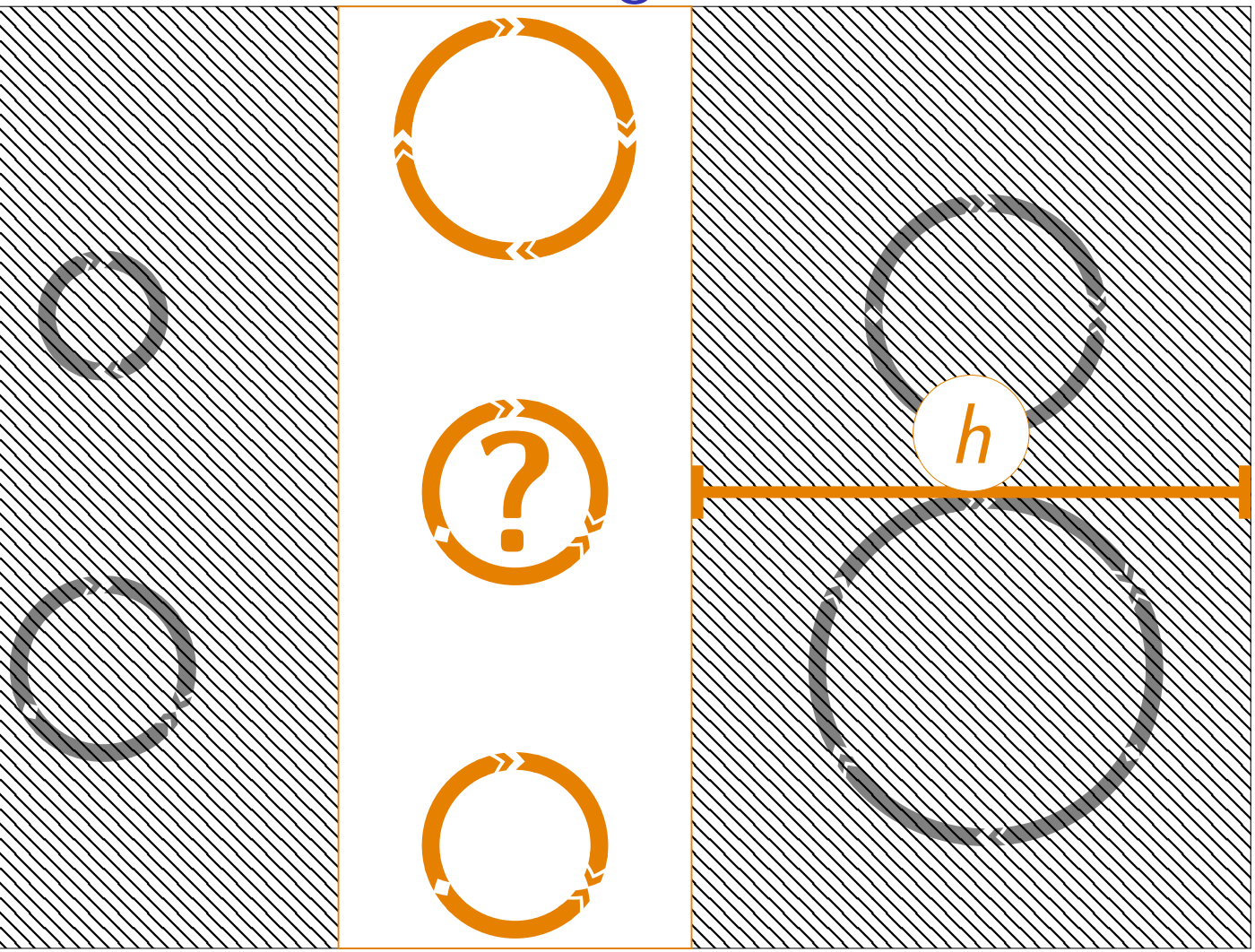


Structurally Doubled

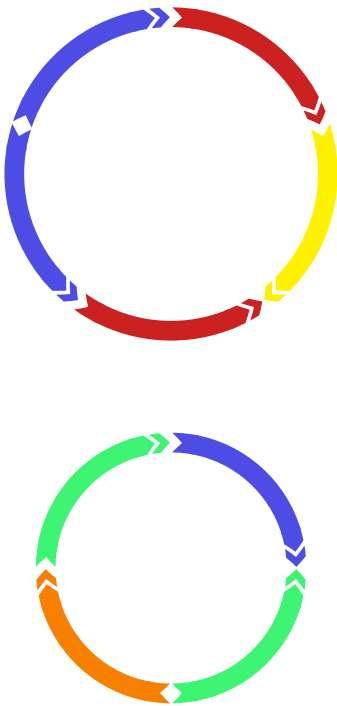


Natural

The Genome Halving Problem

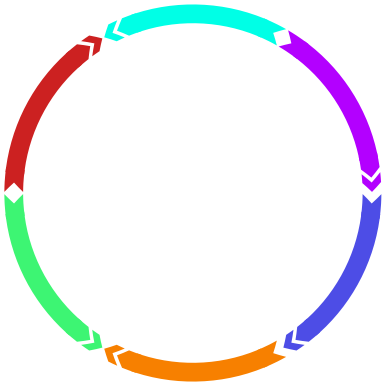


Structurally Doubled

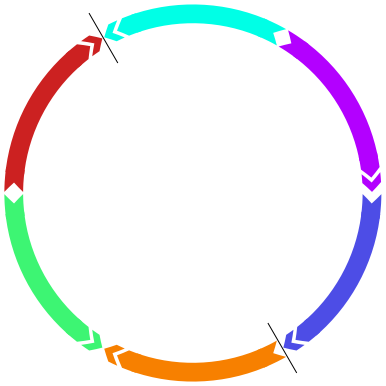


Natural

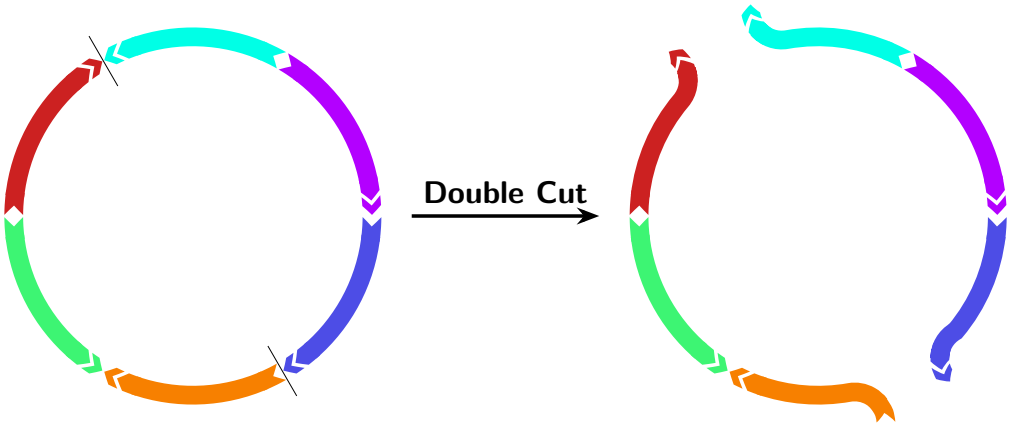
Operations: Double-Cut-And-Join



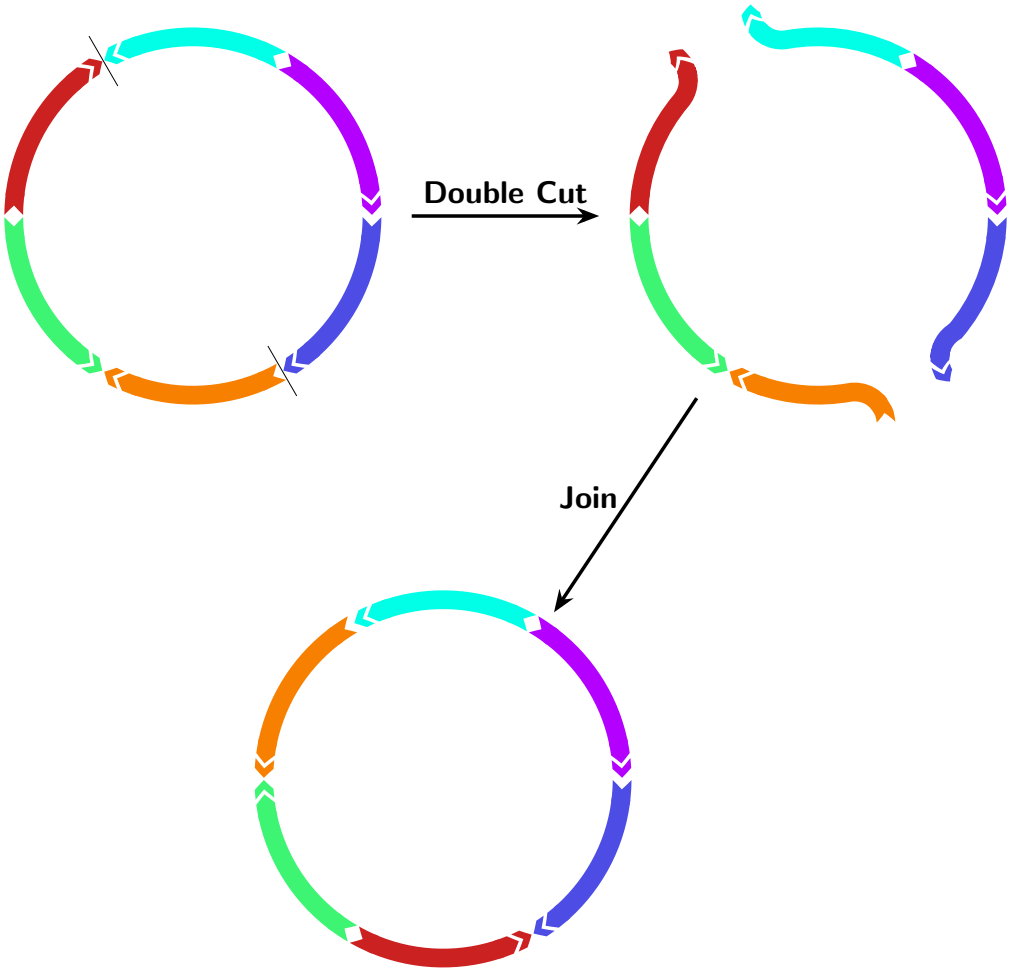
Operations: Double-Cut-And-Join



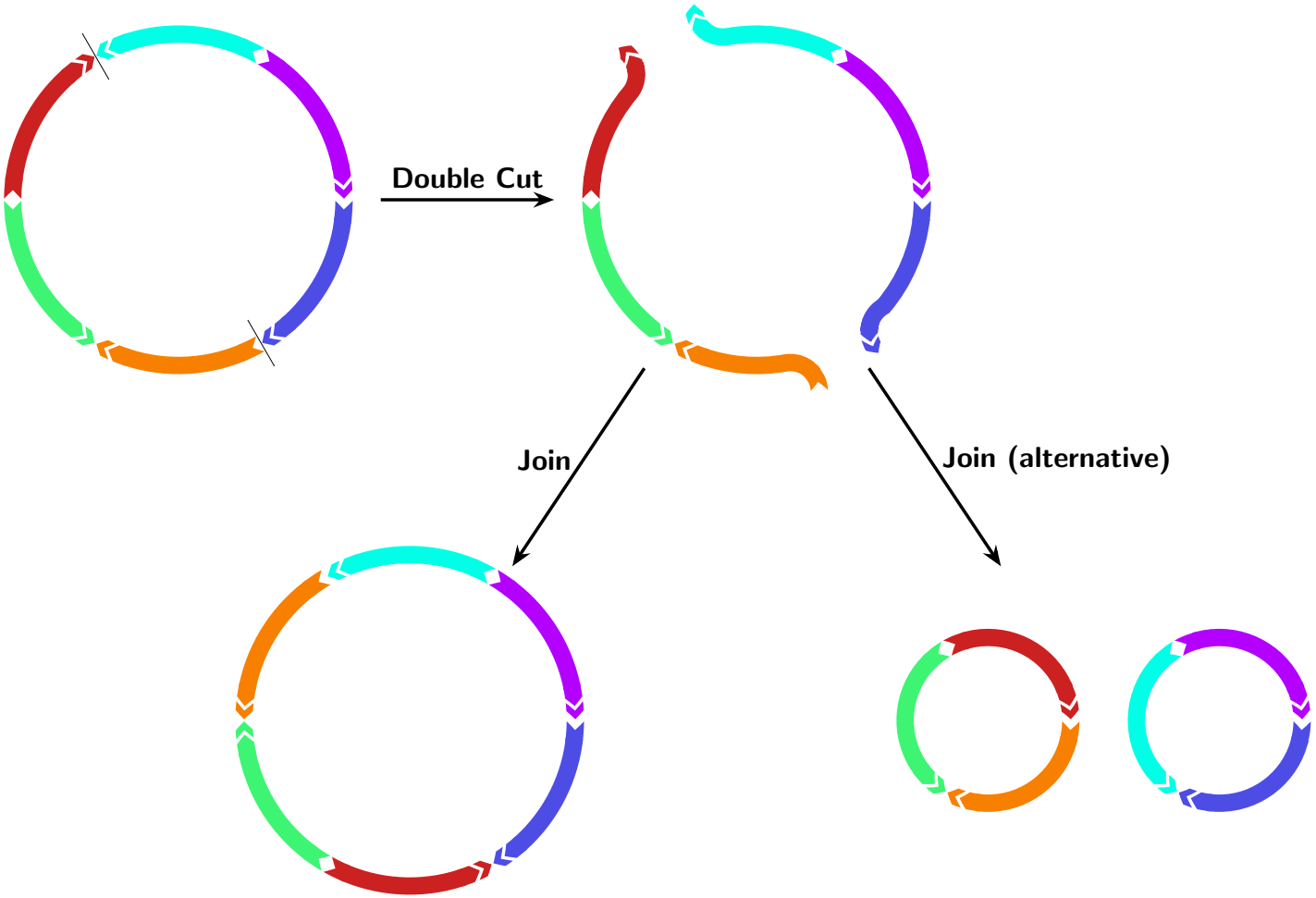
Operations: Double-Cut-And-Join



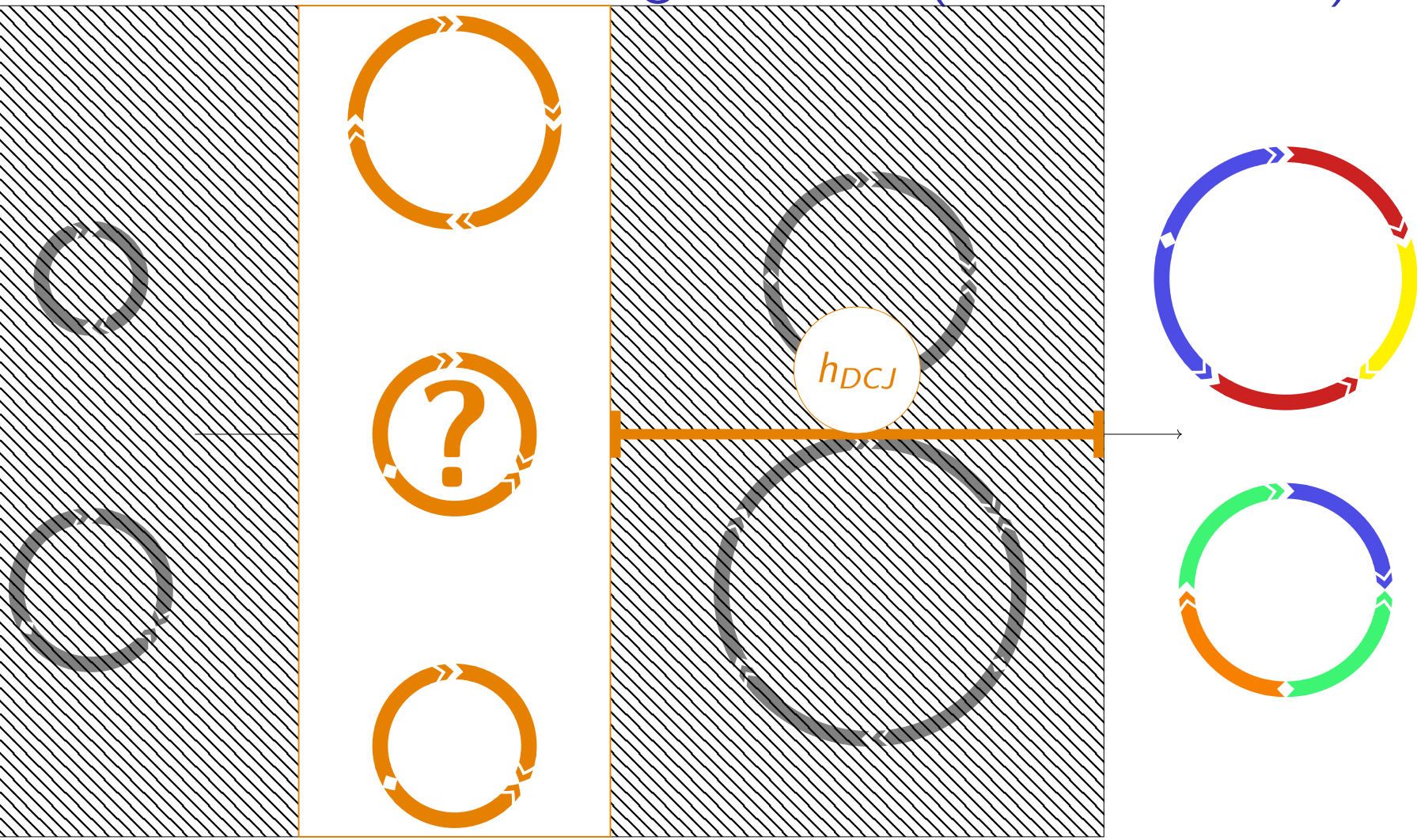
Operations: Double-Cut-And-Join



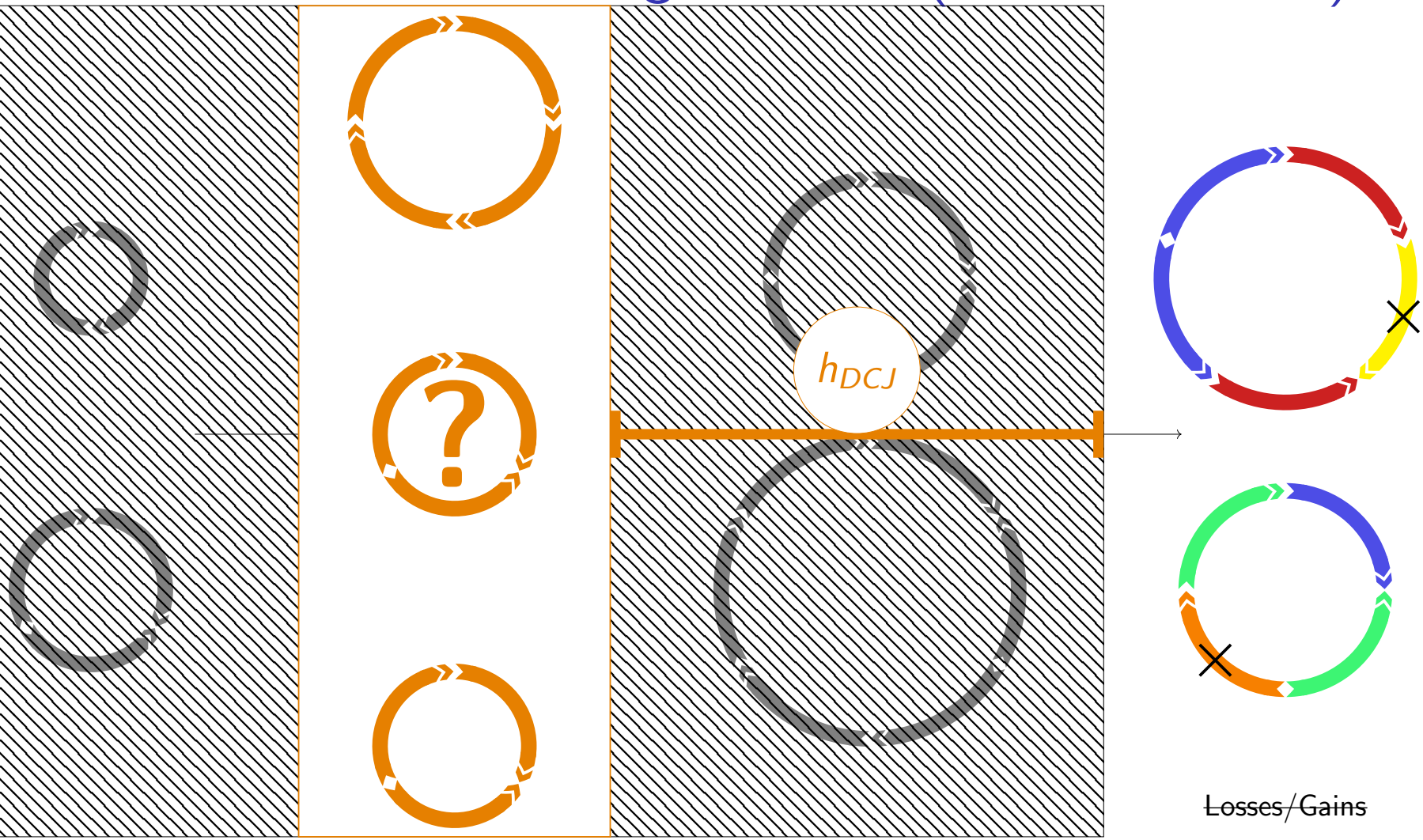
Operations: Double-Cut-And-Join



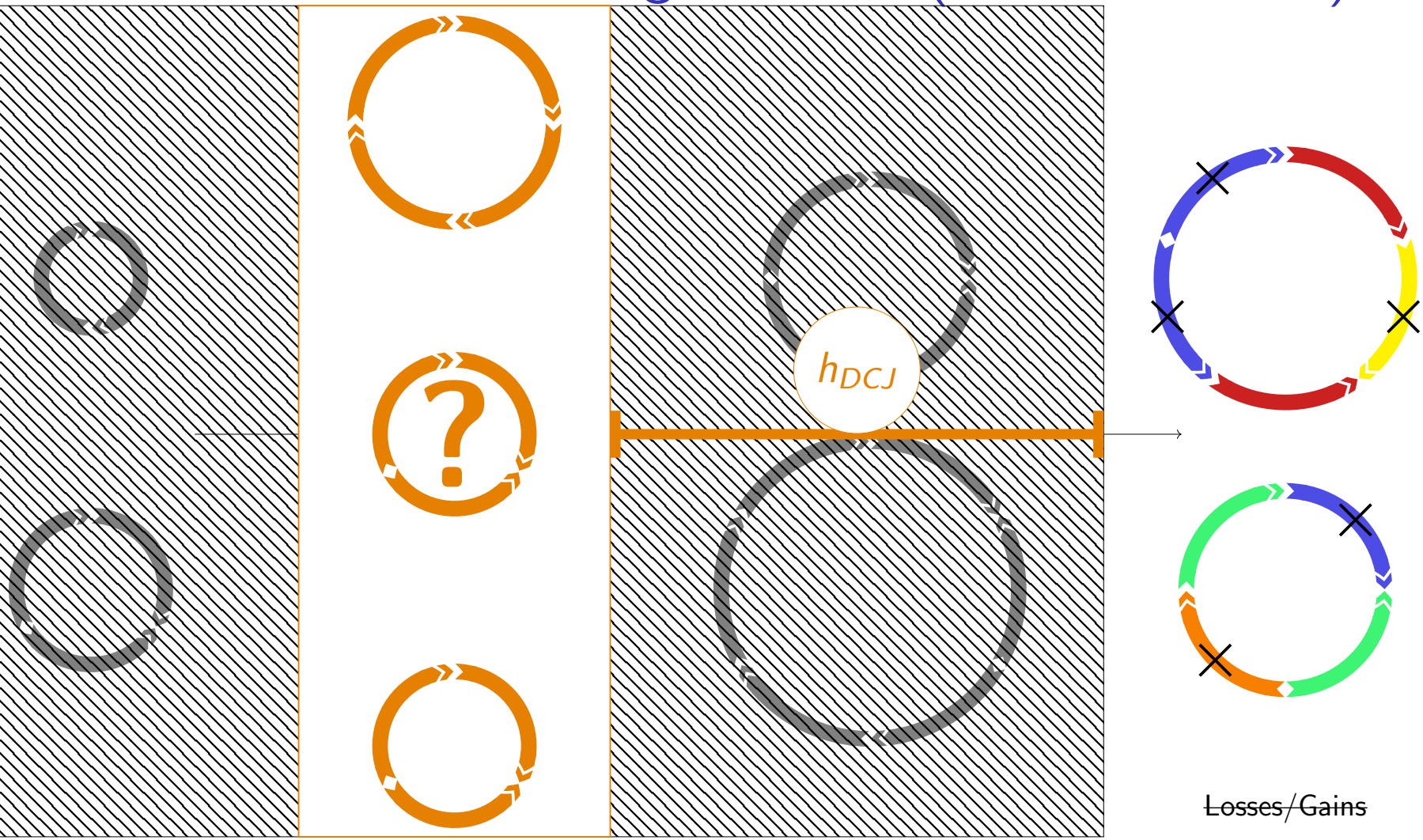
The Classic DCJ Halving Problem (Mixtacki, 2008)



The Classic DCJ Halving Problem (Mixtacki, 2008)



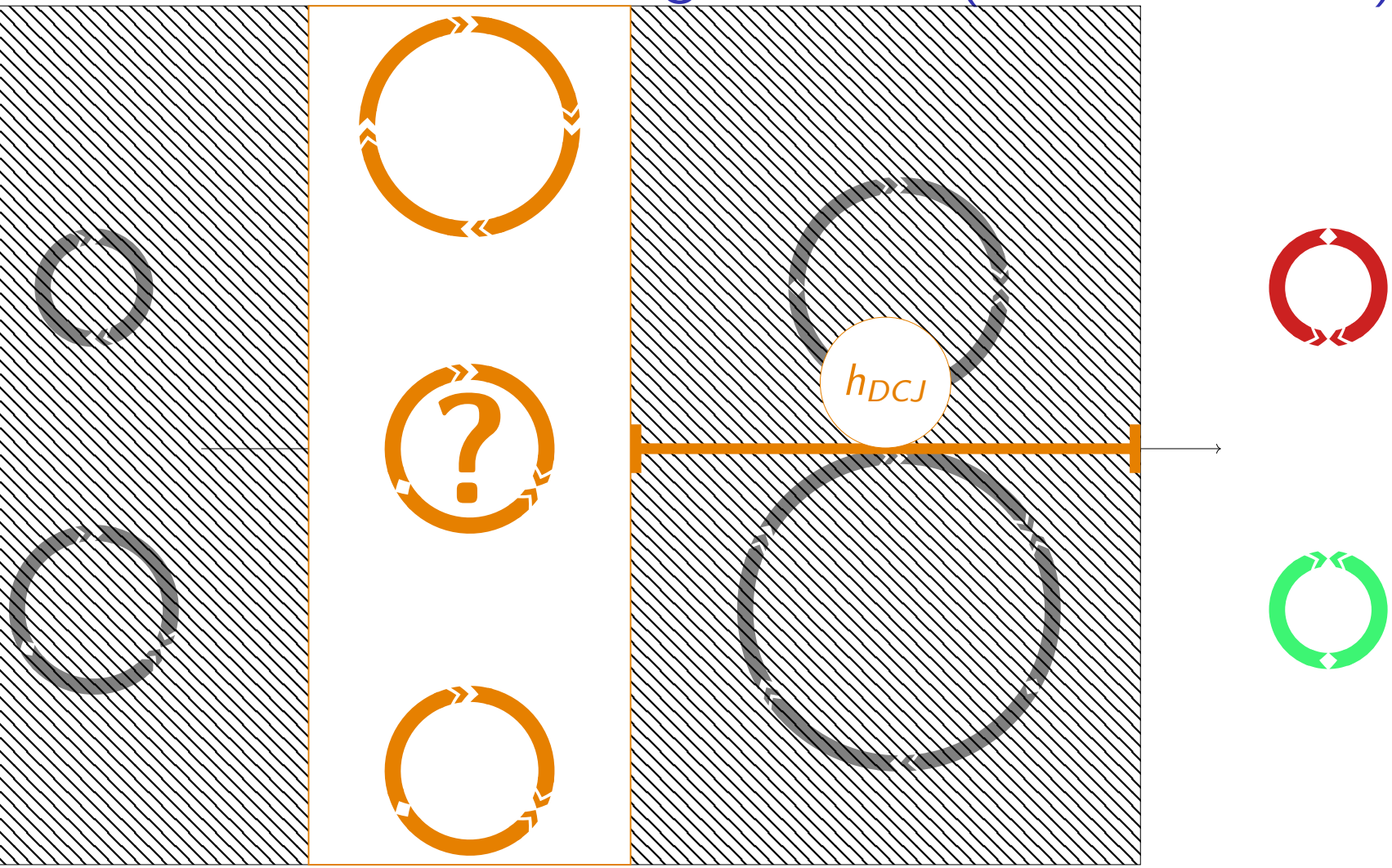
The Classic DCJ Halving Problem (Mixtacki, 2008)



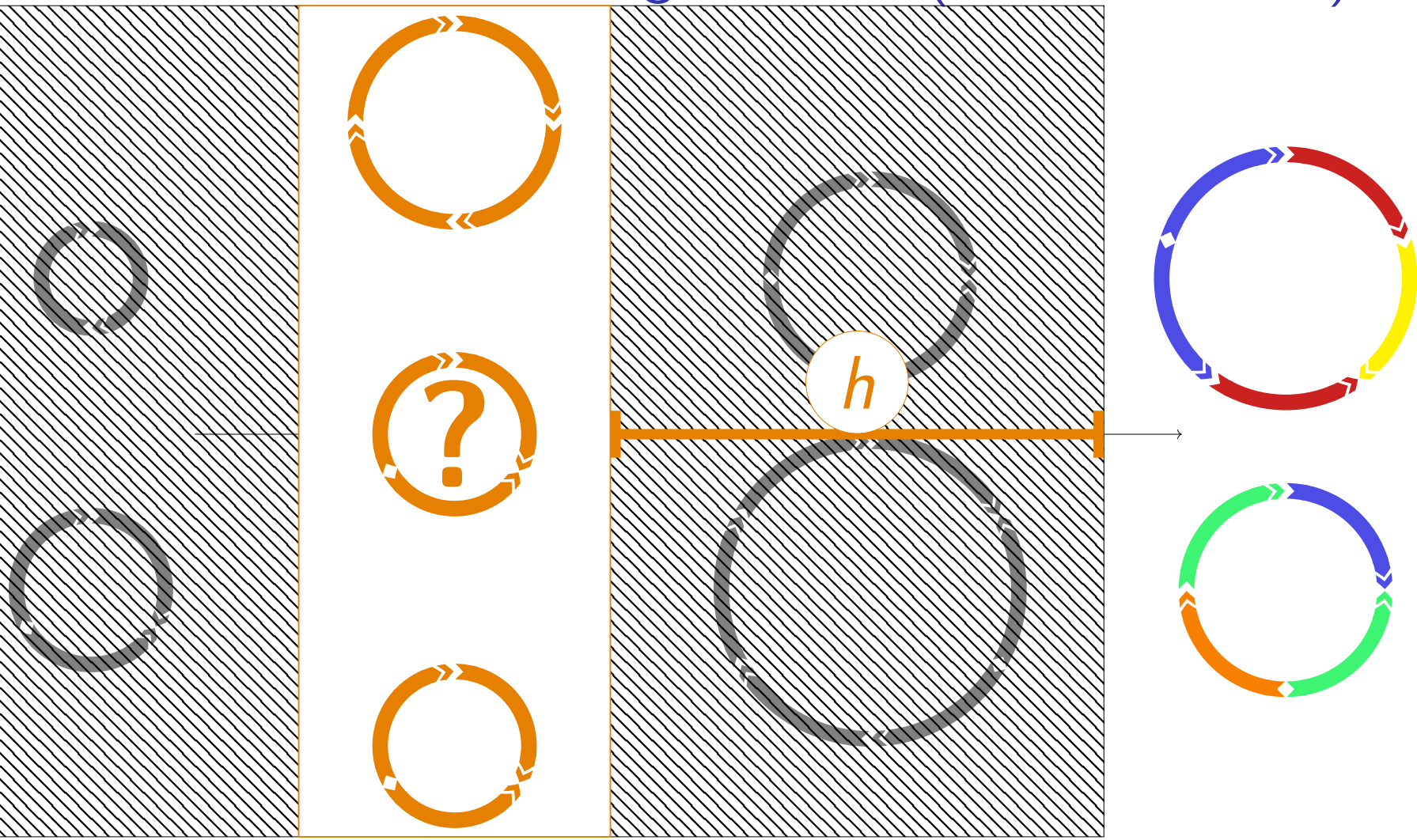
~~Losses/Gains~~

~~Ambiguous Markers~~

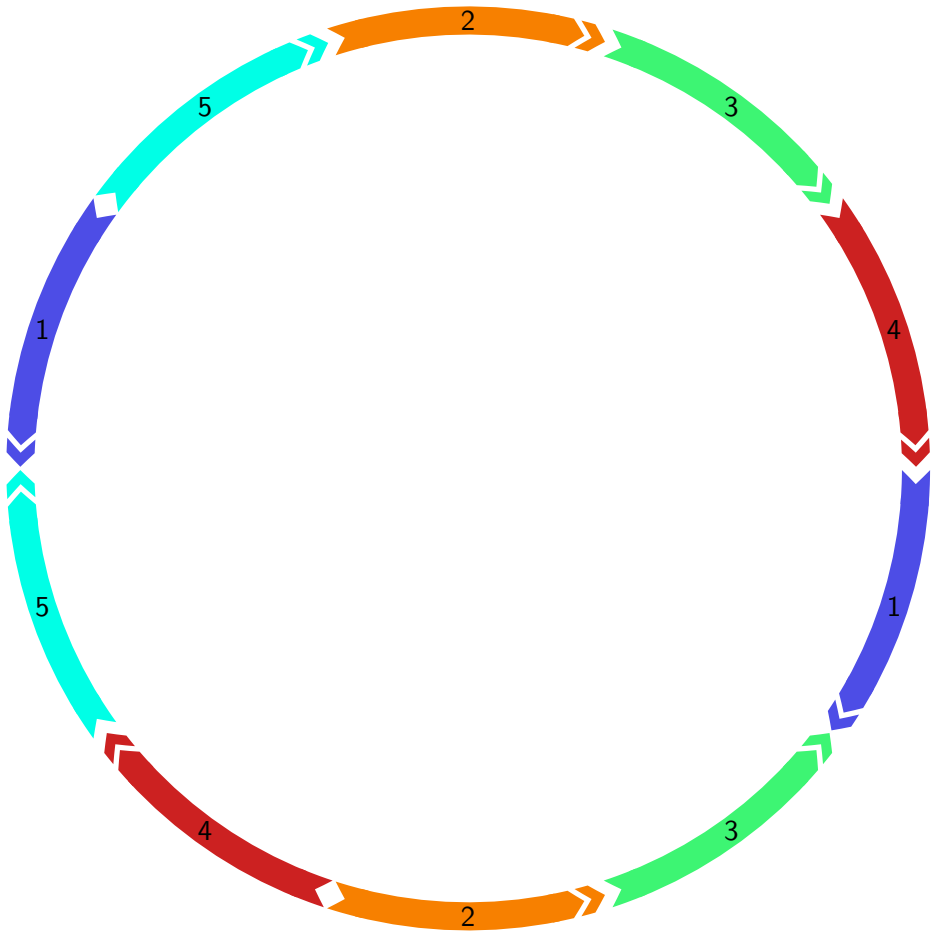
The Classic DCJ Halving Problem (Mixtacki, 2008)



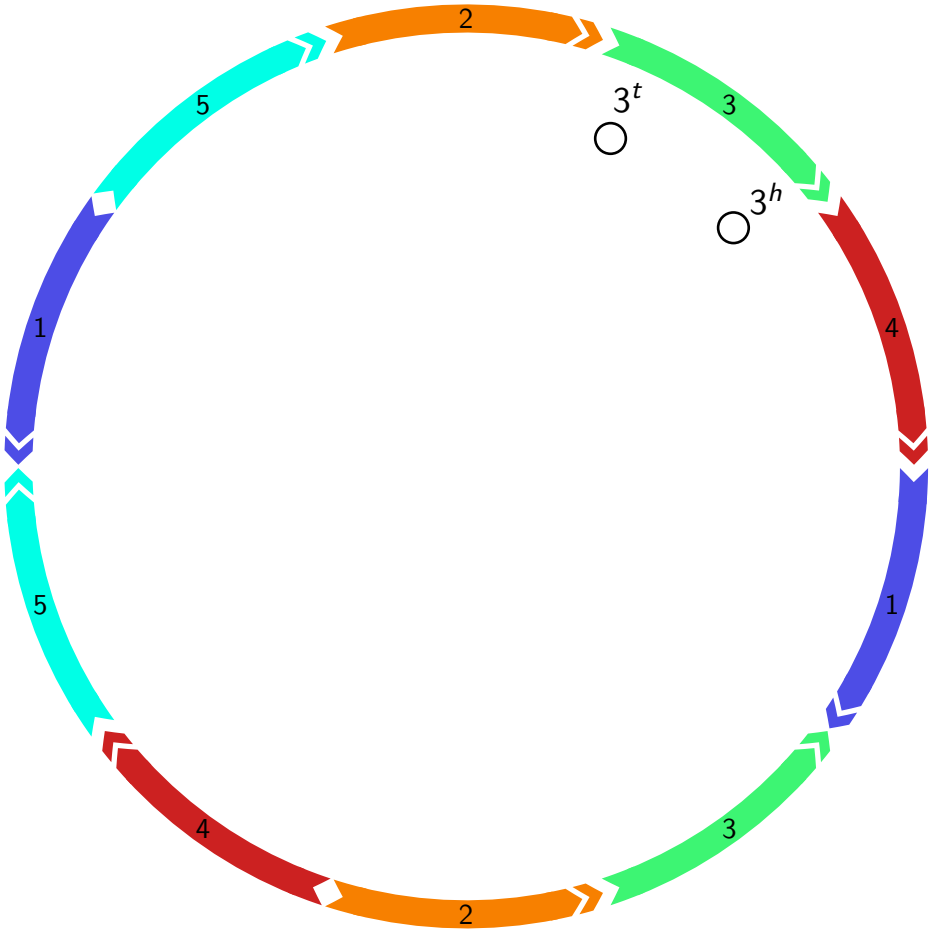
The Classic DCJ Halving Problem (Mixtacki, 2008)



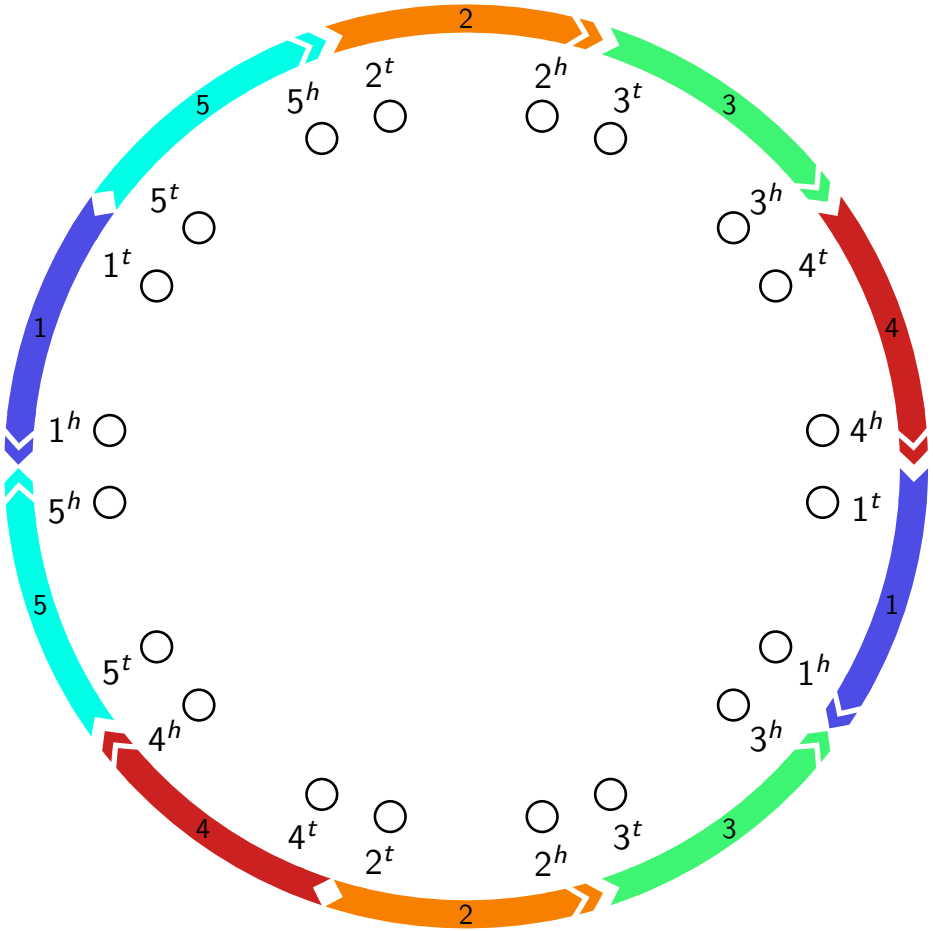
Halving with (~~Losses~~,~~Ambiguous~~): Natural Graph



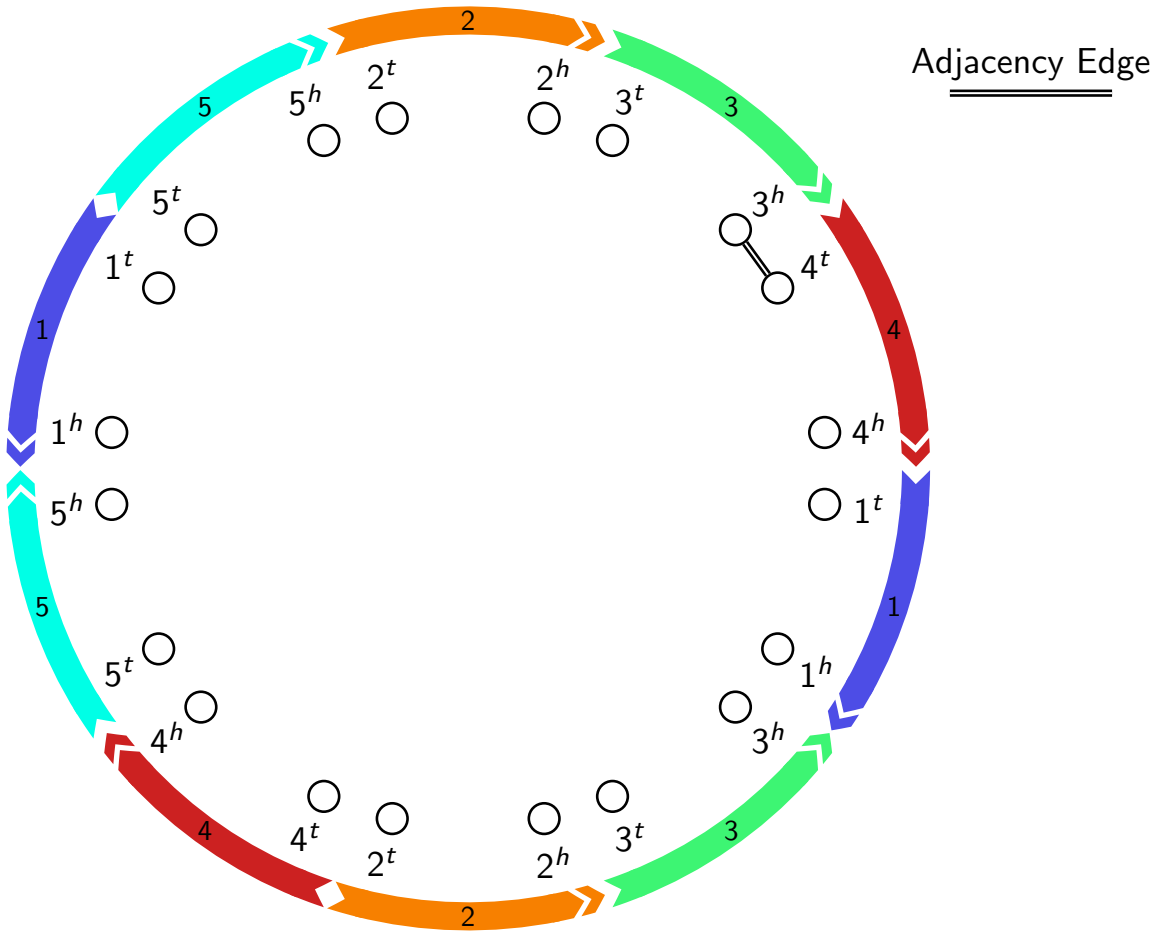
Halving with (~~Losses~~,~~Ambiguous~~): Natural Graph



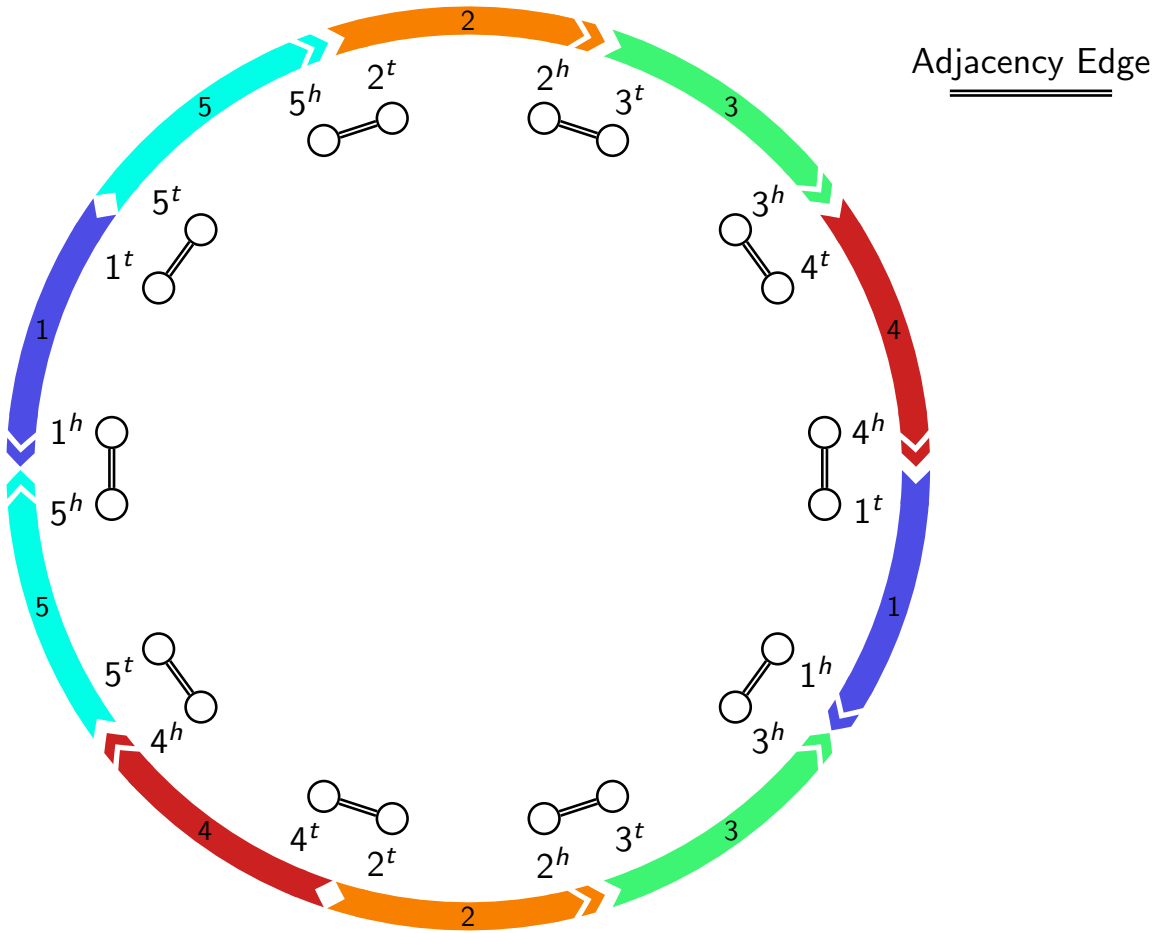
Halving with (~~Losses~~,~~Ambiguous~~): Natural Graph



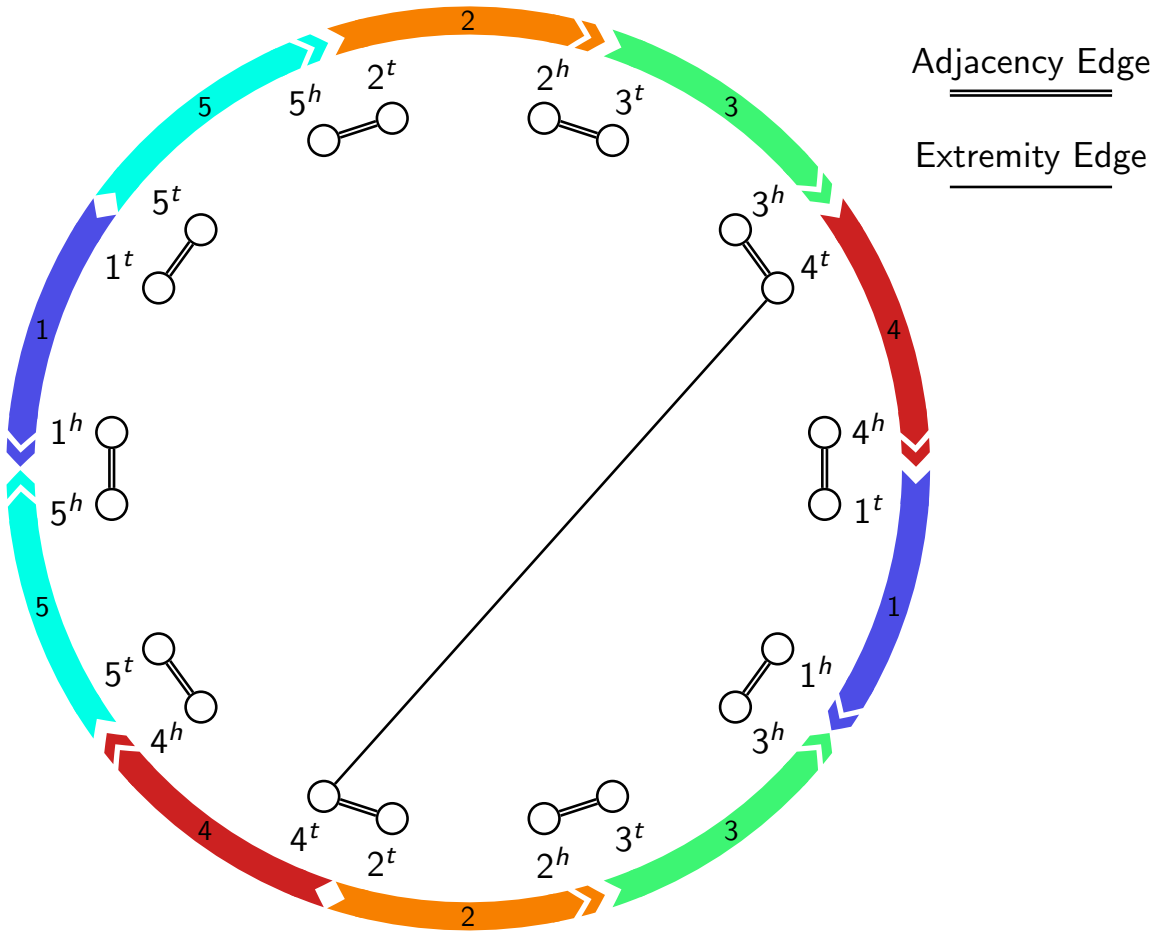
Halving with (~~Losses~~,~~Ambiguous~~): Natural Graph



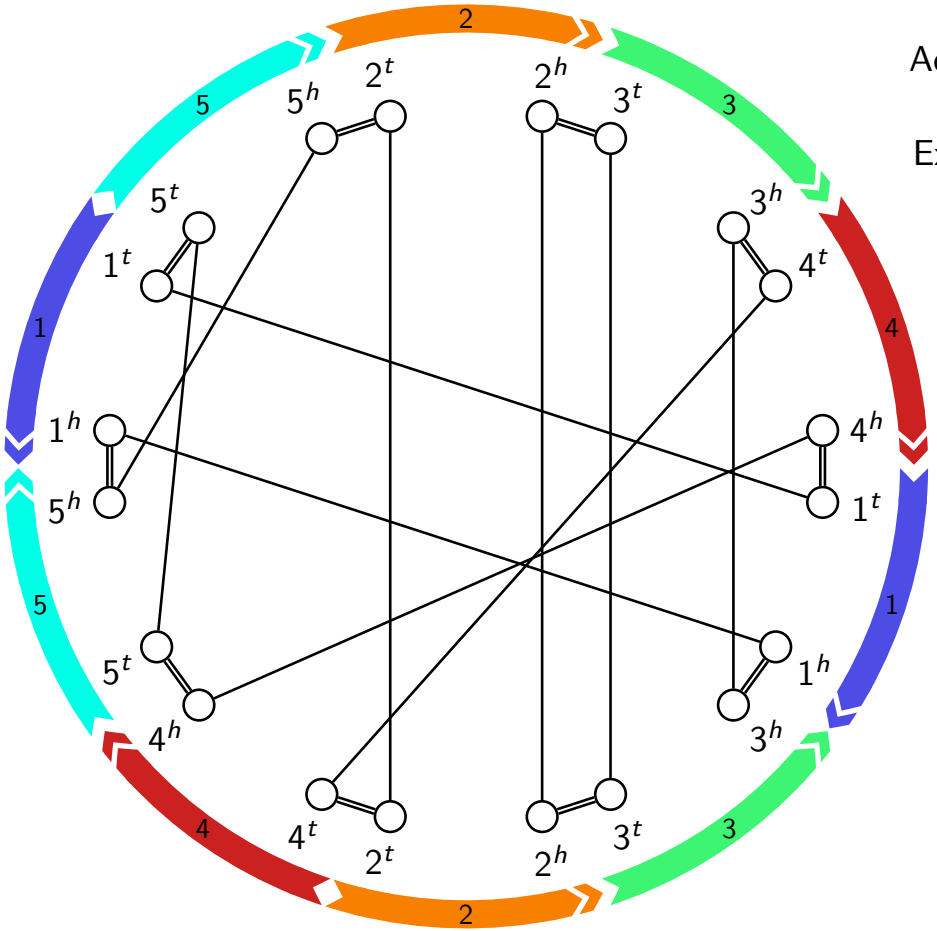
Halving with (~~Losses~~,~~Ambiguous~~): Natural Graph



Halving with (~~Losses~~,~~Ambiguous~~): Natural Graph



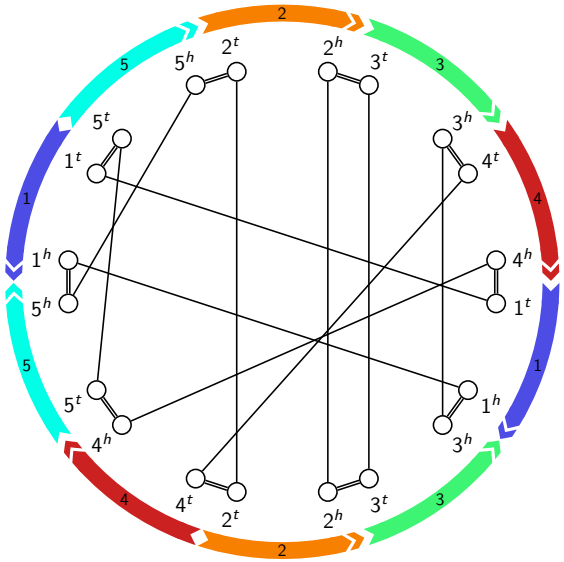
Halving with (~~Losses~~,~~Ambiguous~~): Natural Graph

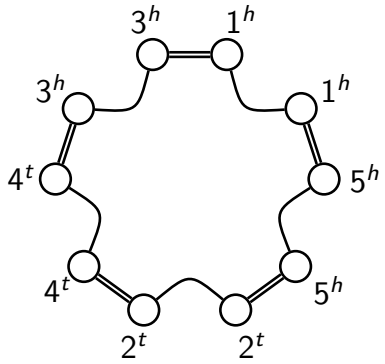


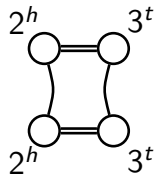
Adjacency Edge

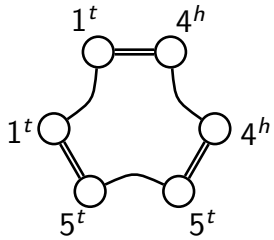
Extremity Edge

Halving with (~~Losses~~,~~Ambiguous~~): Distance

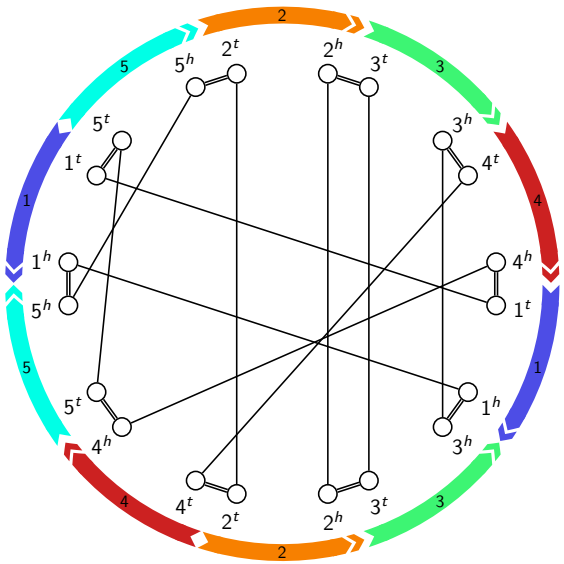


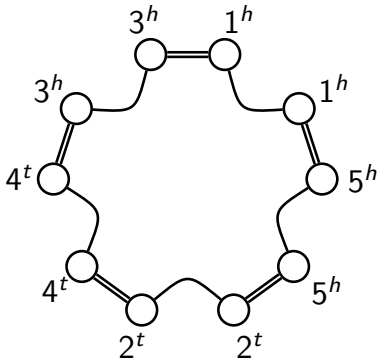


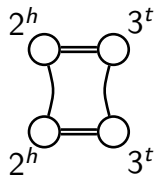


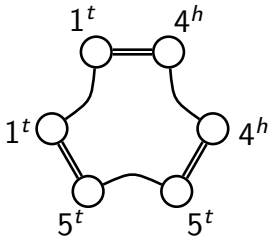


Halving with (~~Losses~~,~~Ambiguous~~): Distance





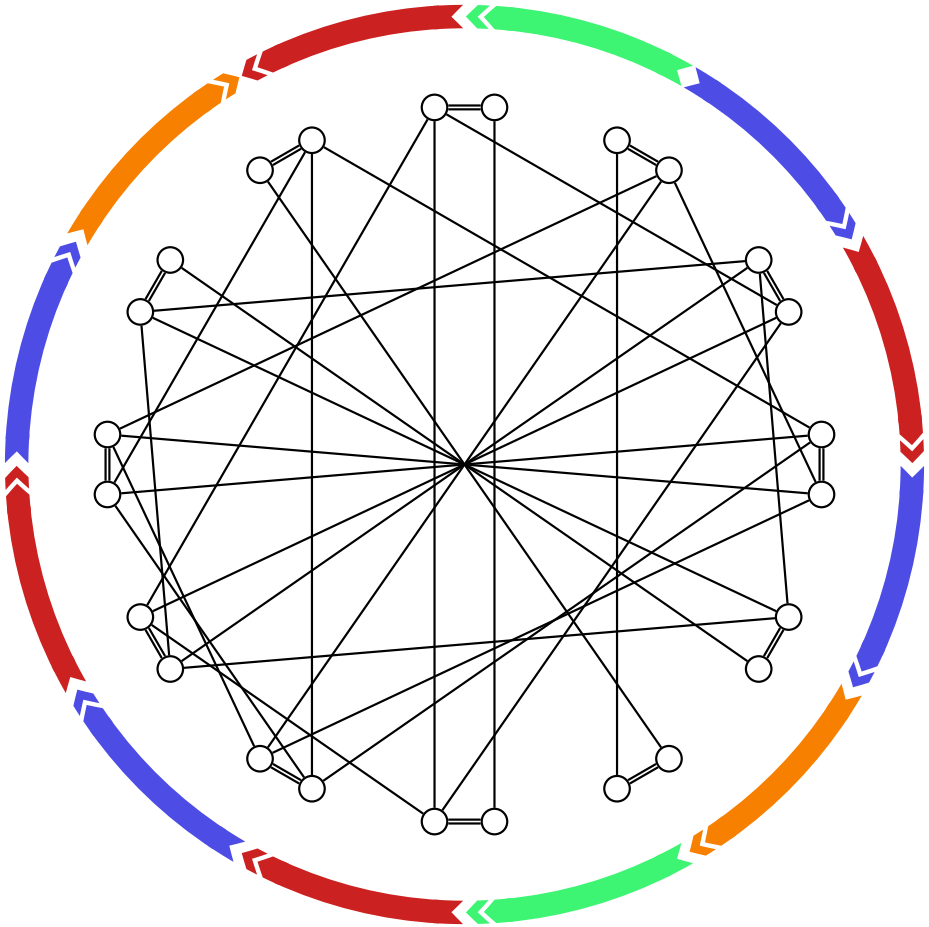




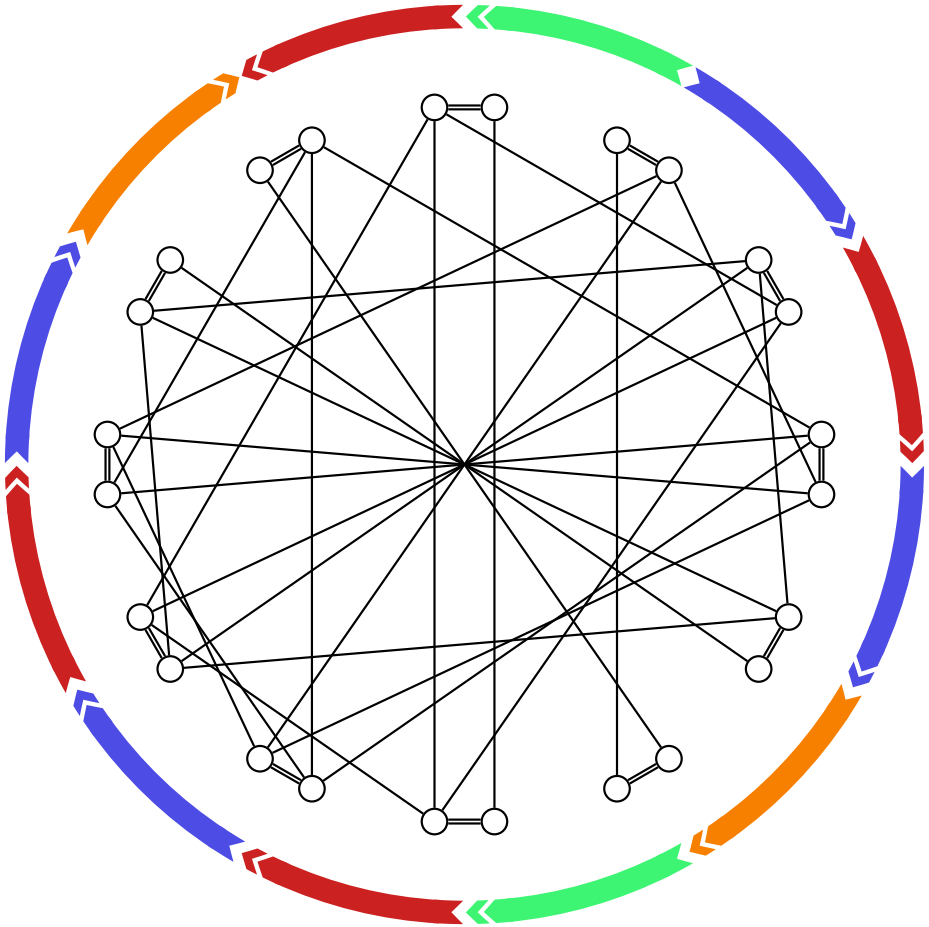
For (circular) duplicated genomes:

$$h_{DCJ} = n - c_e$$

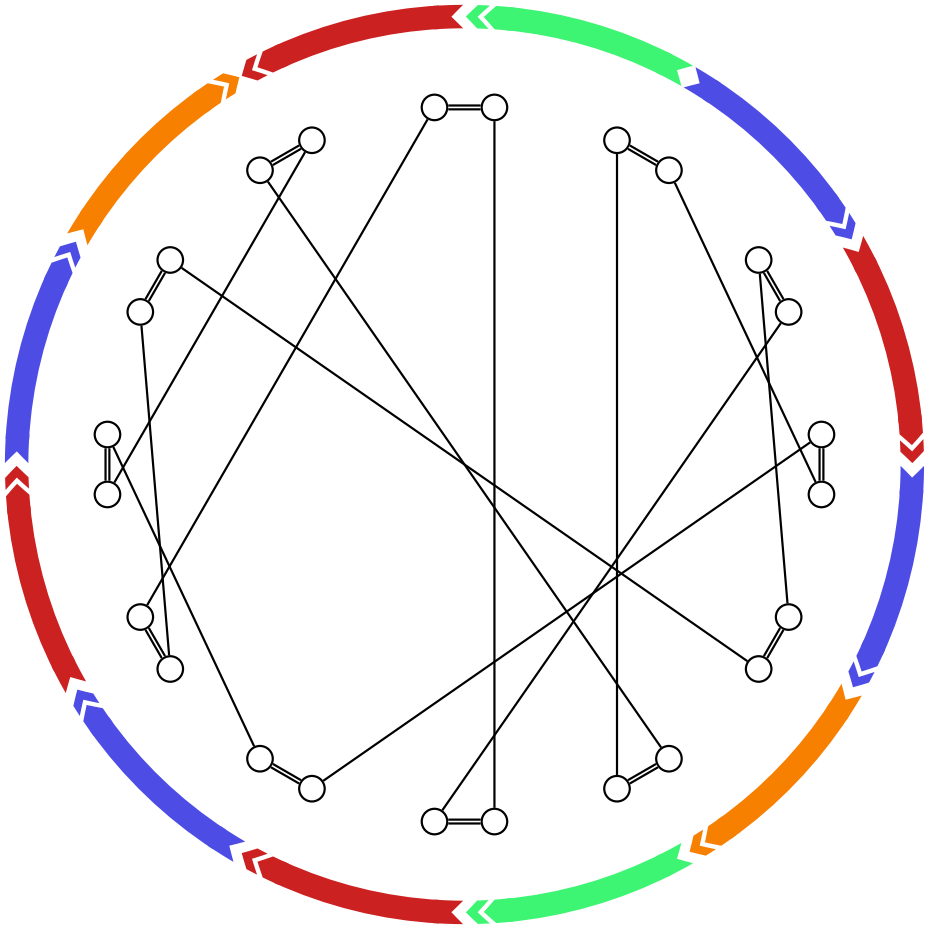
Halving with (~~Losses~~,Ambiguous): Supernatural Graph



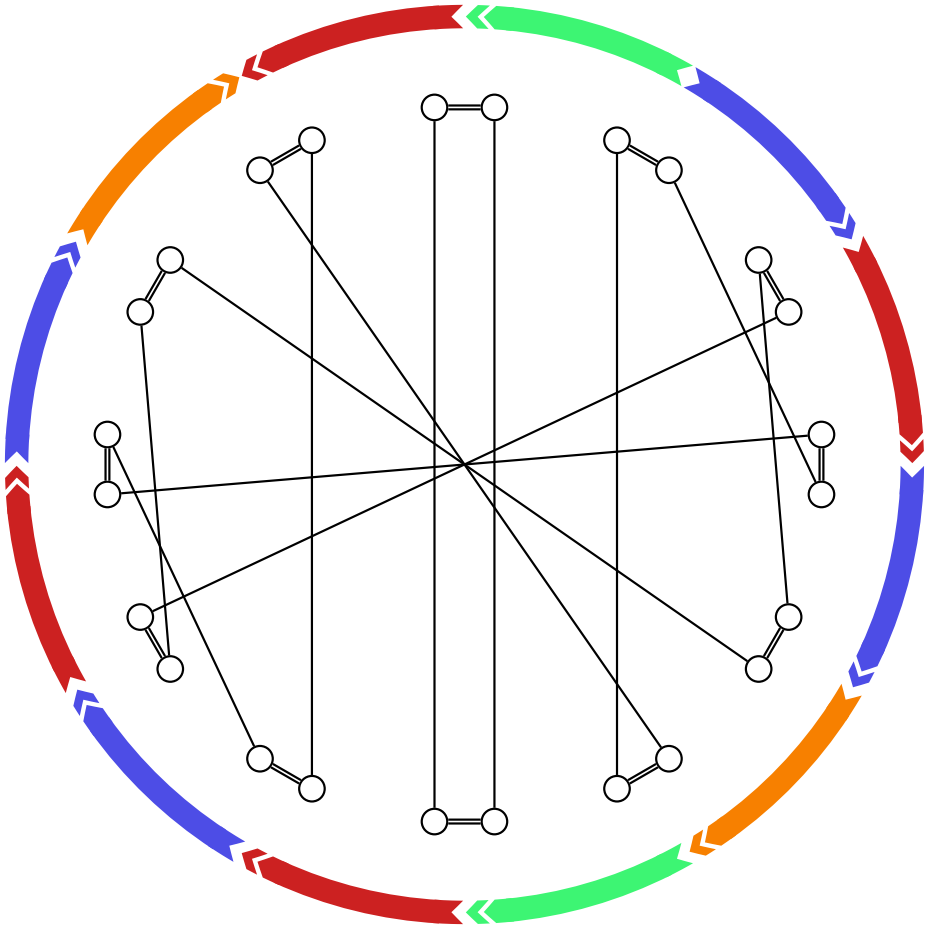
Halving with (~~Losses~~,Ambiguous): Distance



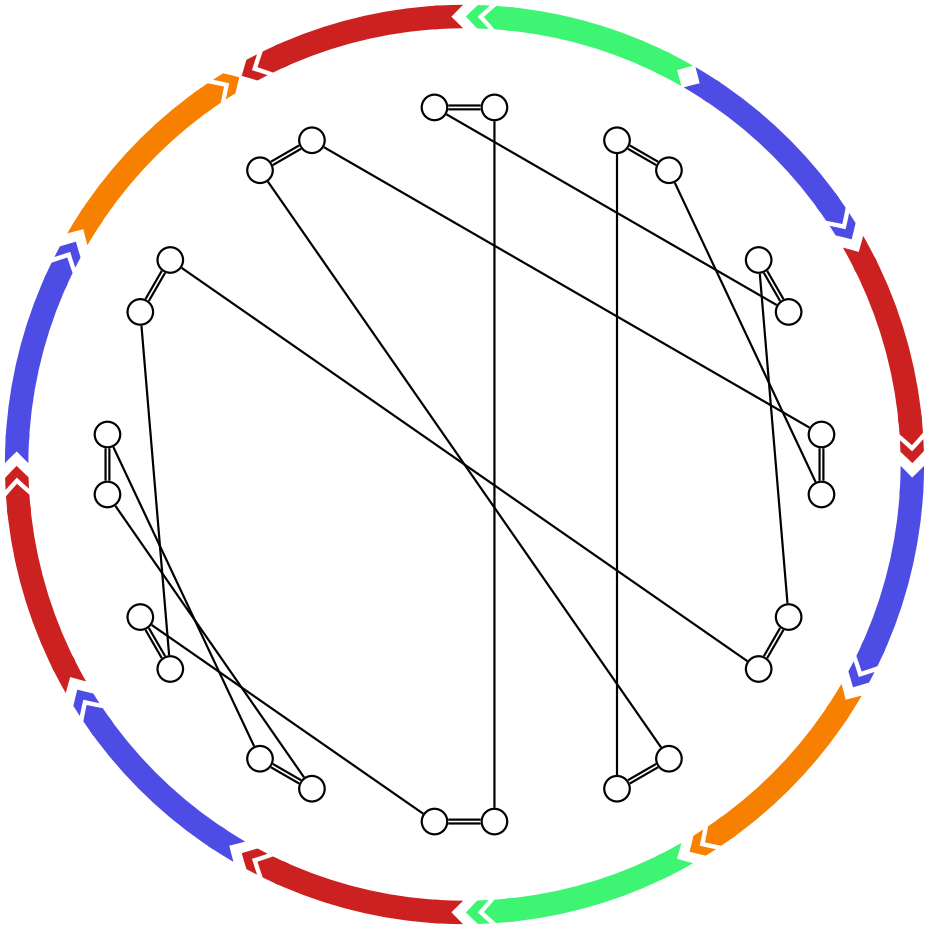
Halving with (~~Losses~~,Ambiguous): Distance



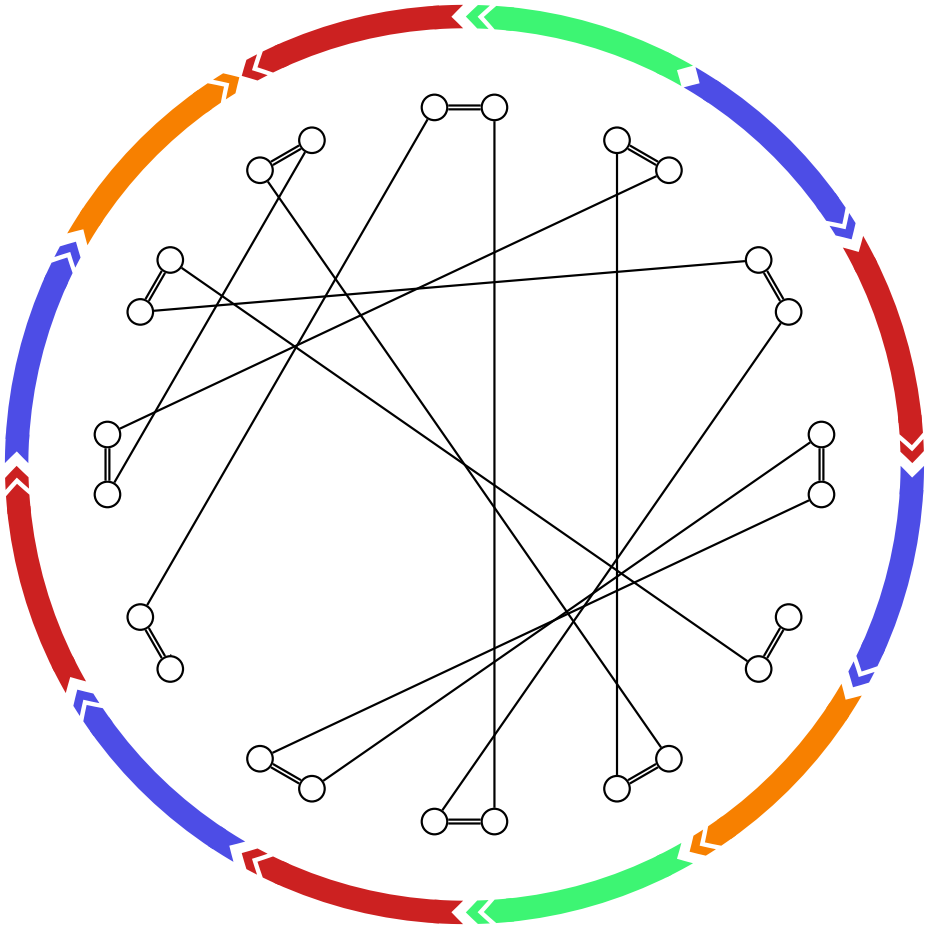
Halving with (~~Losses~~,Ambiguous): Distance



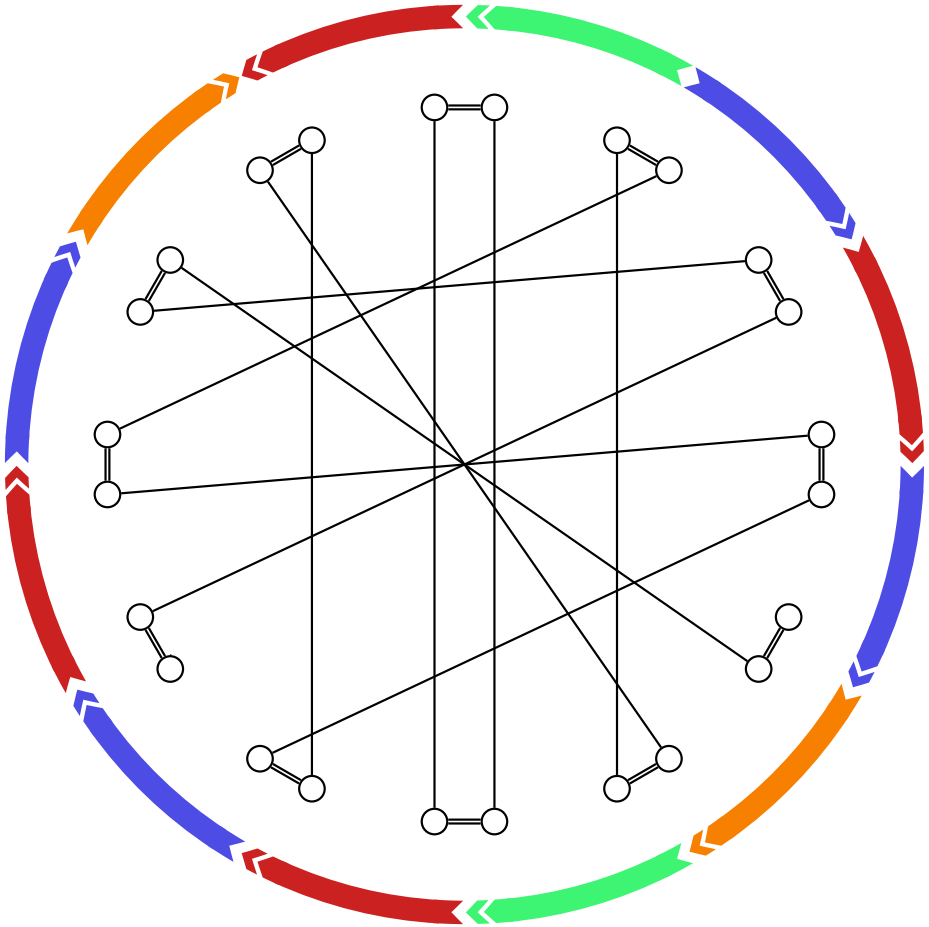
Halving with (~~Losses~~,Ambiguous): Distance



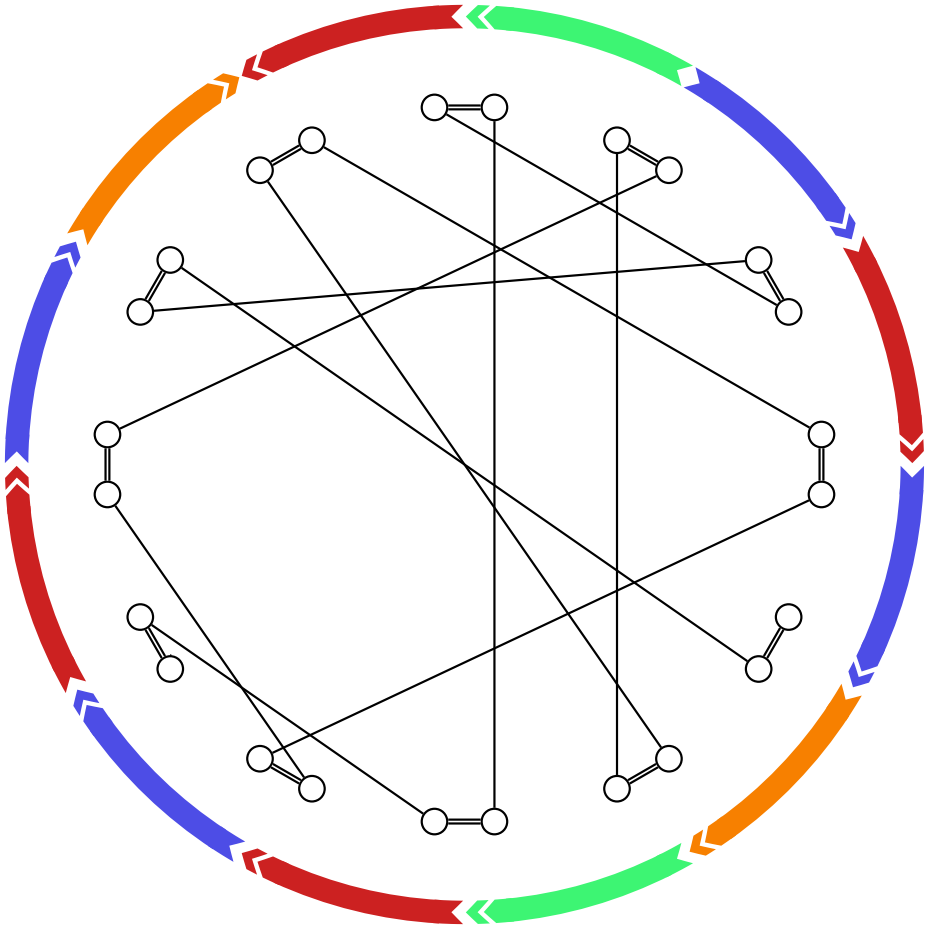
Halving with (~~Losses~~,Ambiguous): Distance



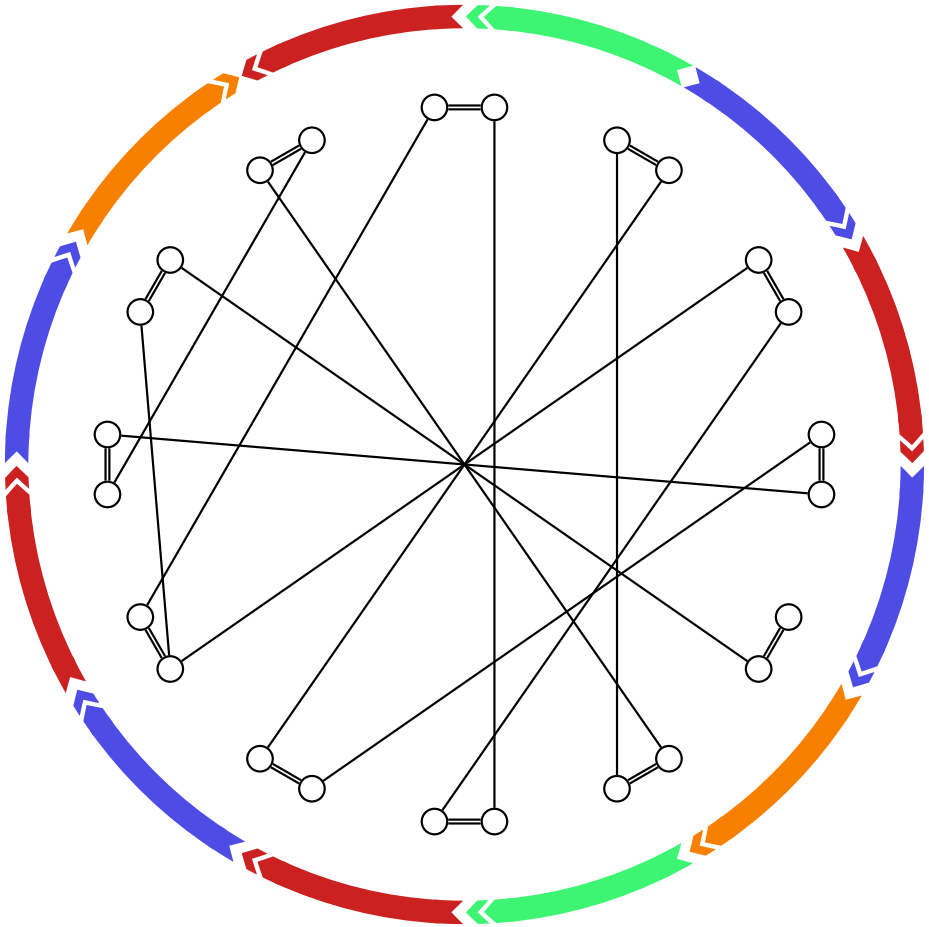
Halving with (~~Losses~~,Ambiguous): Distance



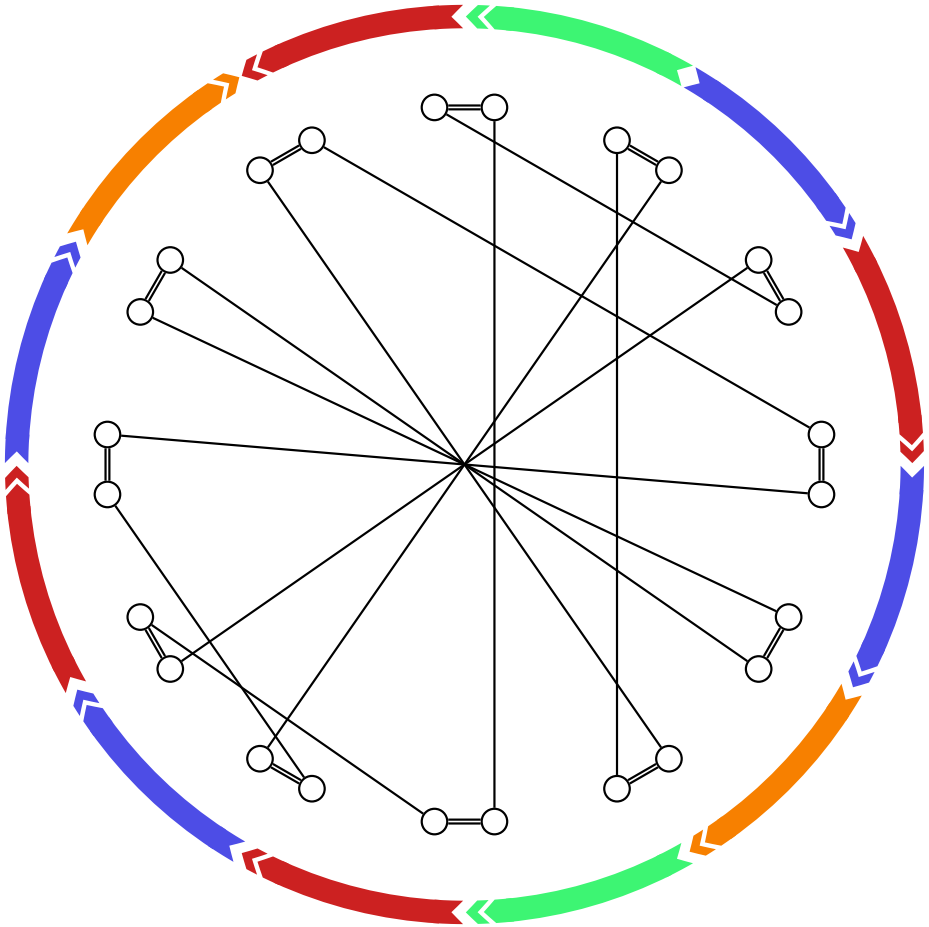
Halving with (~~Losses~~,Ambiguous): Distance



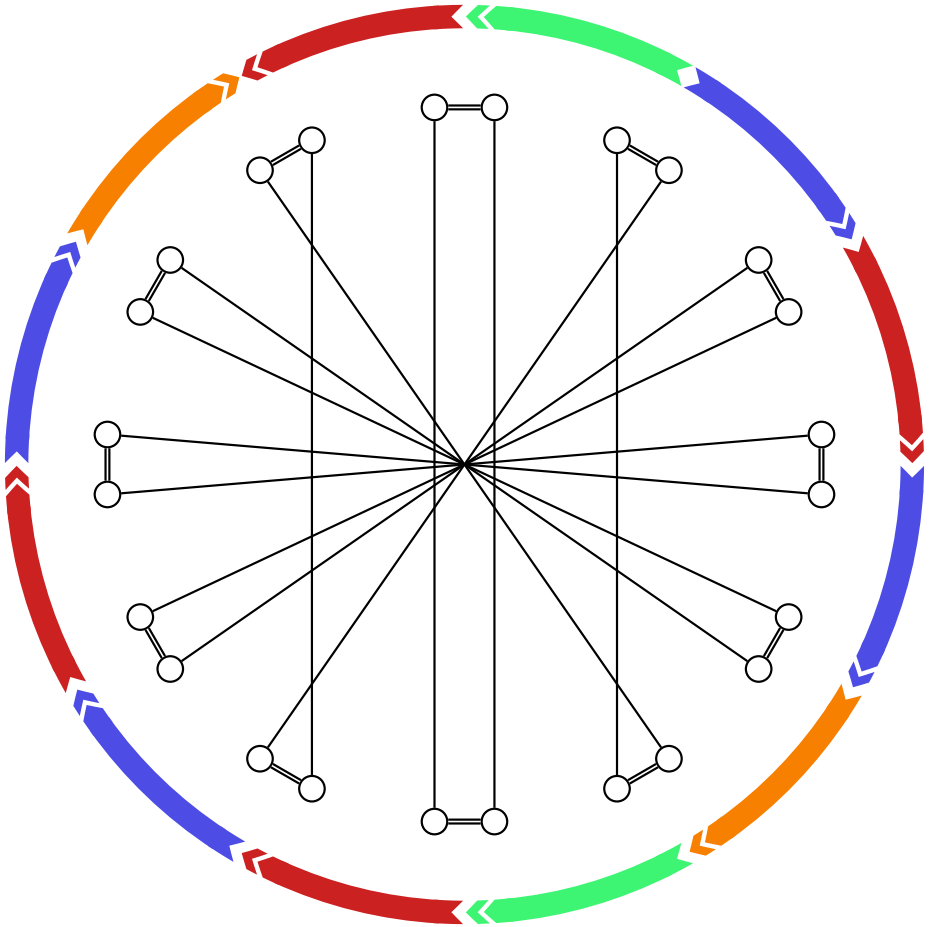
Halving with (~~Losses~~,Ambiguous): Distance



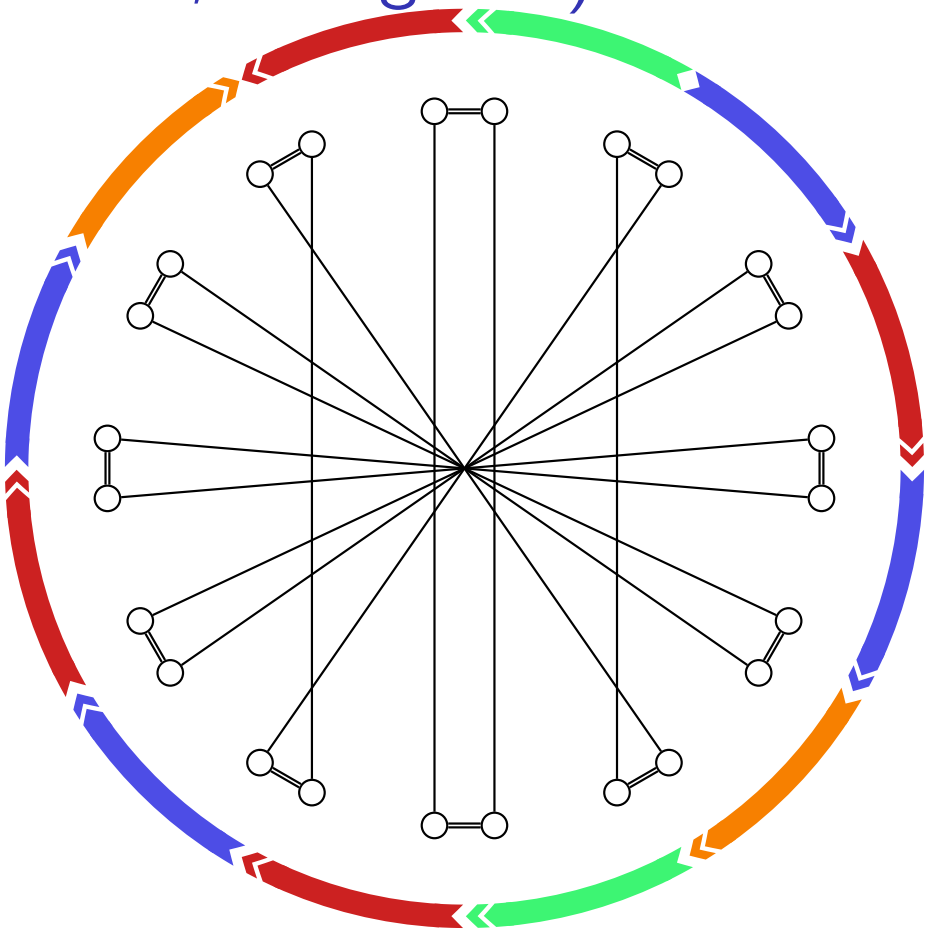
Halving with (~~Losses~~,Ambiguous): Distance



Halving with (~~Losses~~,Ambiguous): Distance



Halving with (~~Losses~~,Ambiguous): Distance



$$h_{DCJ} = \min_{\substack{\text{Consistent} \\ \text{Decomposition } D}} (n - c_e(D))$$

Halving with (~~Losses~~,Ambiguous): Solution

- NP-hard

Halving with (~~Losses~~,Ambiguous): Solution

- NP-hard
- Solvable via ILP:

Minimize

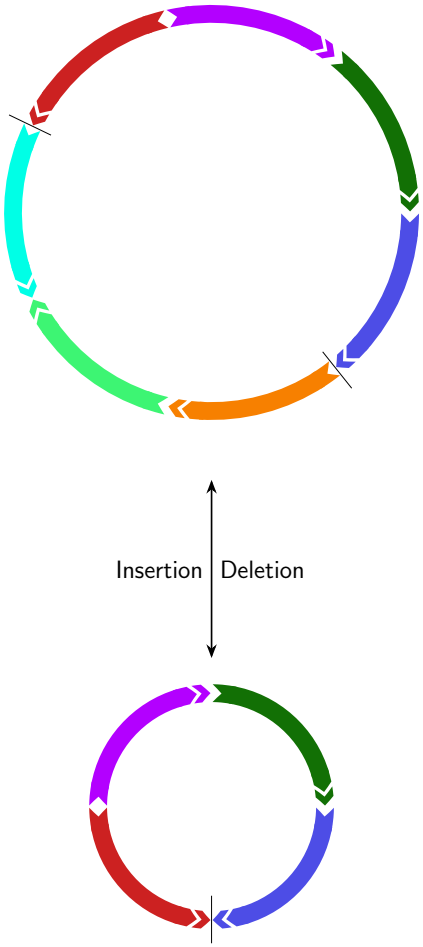
$$\frac{1}{2}\sum_{e\in E^{\star}_{\equiv}} x_e - \sum_{v\in V^{\star}} c_v + \sum_{e\in E^{\star}_{\Gamma}} b_e + \sum_{\substack{C \text{ circ.}\\ \text{chromosome}}} s_C$$

Subject to:

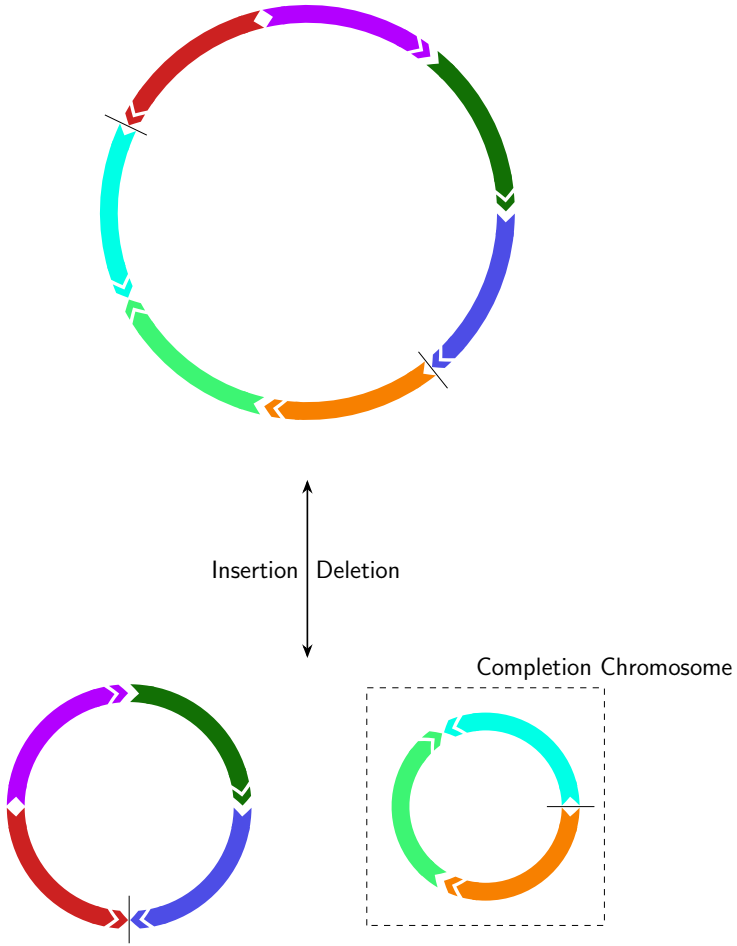
Core		
cc.01:	$x_e = 1$	$\forall e \in E_{\Gamma}^{\perp}$
cc.02:	$x_e = x_d$	e, d sibling edges
cc.03A1:	$\sum_{uv\in E^{\star}_{\Gamma}} x_{uv} = 1$	$\forall u \in V^{\star}$
cc.03A2:	$\sum_{uv\in E^{\star}_{\equiv}} x_{uv} = 1$	$\forall u \in V^{\star}$
cc.04:	$l_u - l_v \leq i(u)(1 - x_{uv})$	$\forall uv \in E^{\star}$
cc.05:	$i(u)c_u \leq l_u$	$\forall u \in V$

Dealing with Losses/Gains: Indels

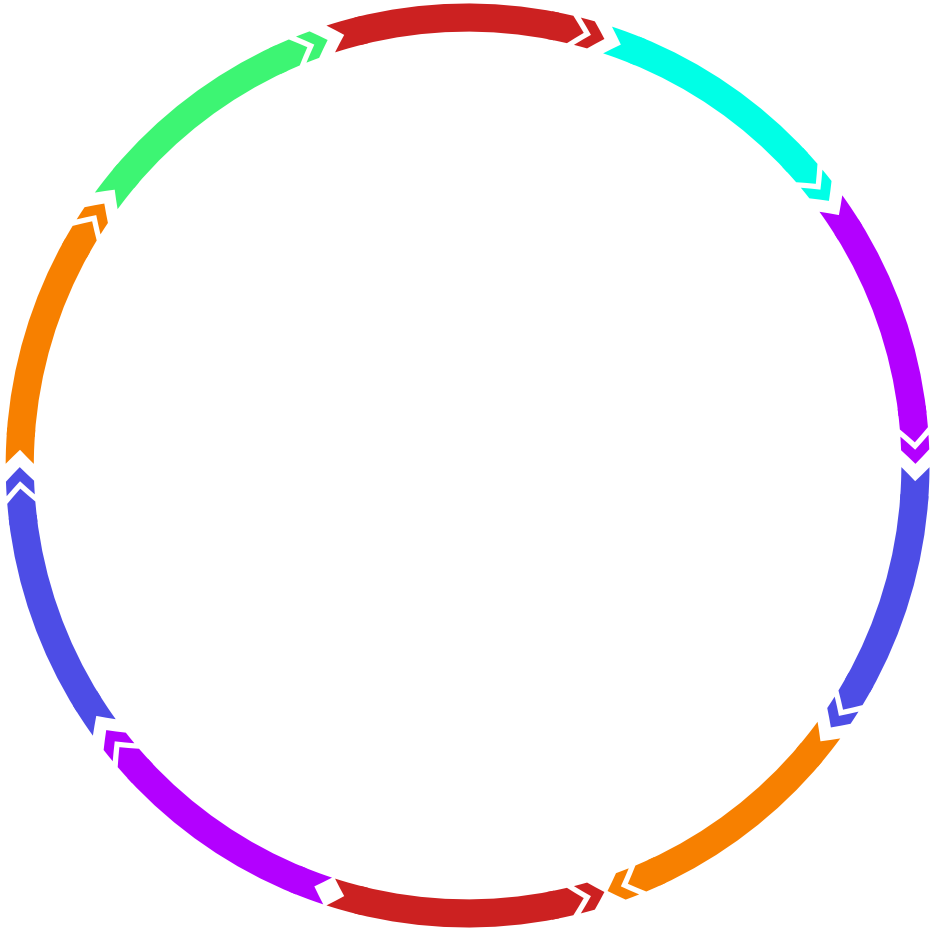
Explicitly (BWS, 2011)

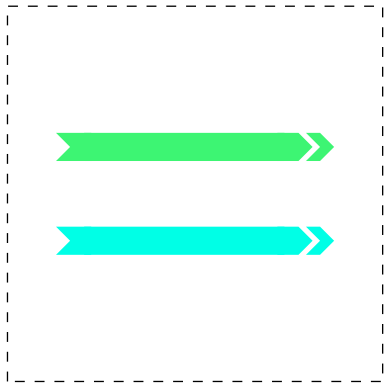


Via Completion (Compeau, 2013)

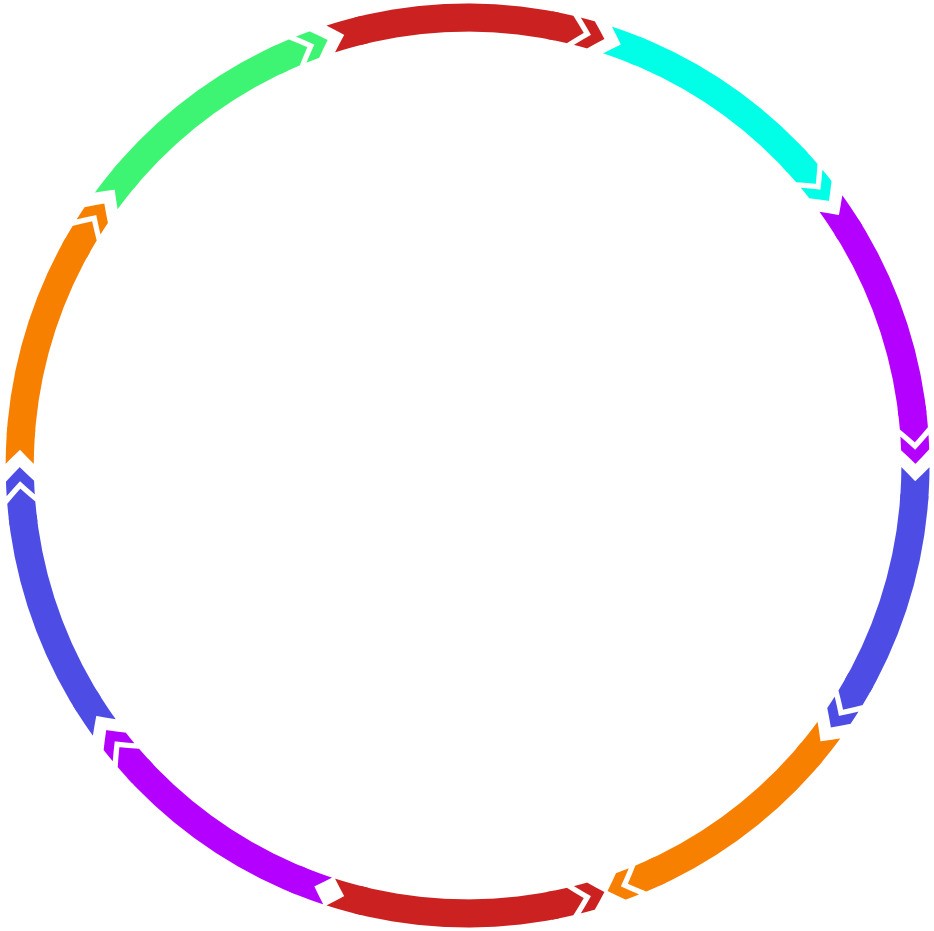


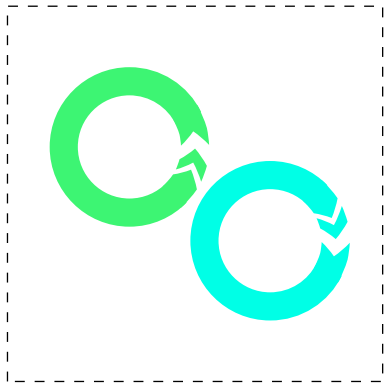
Compeau-style Halving with (Losses,~~Ambiguous~~)



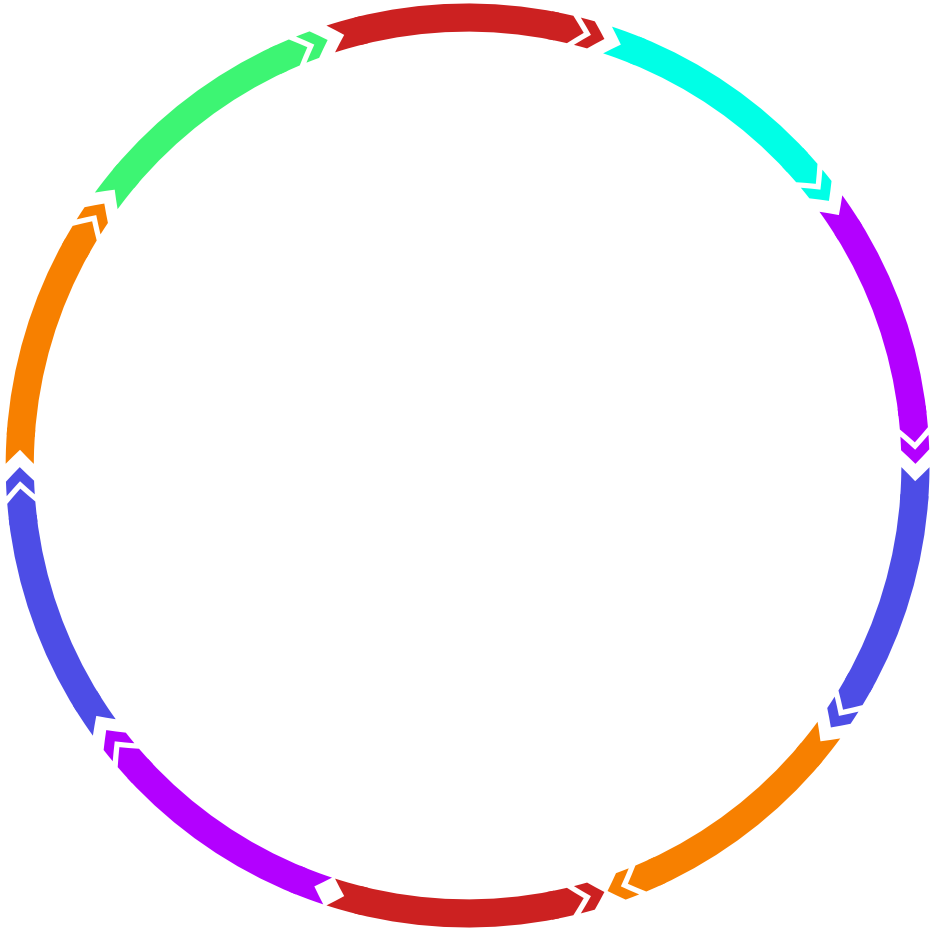


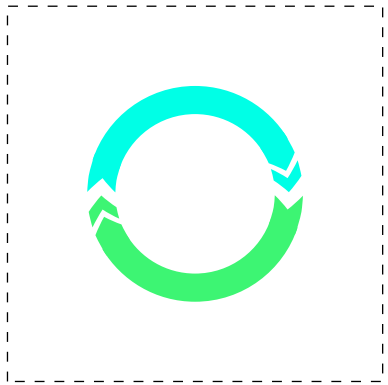
Compeau-style Halving with (Losses,~~Ambiguous~~)



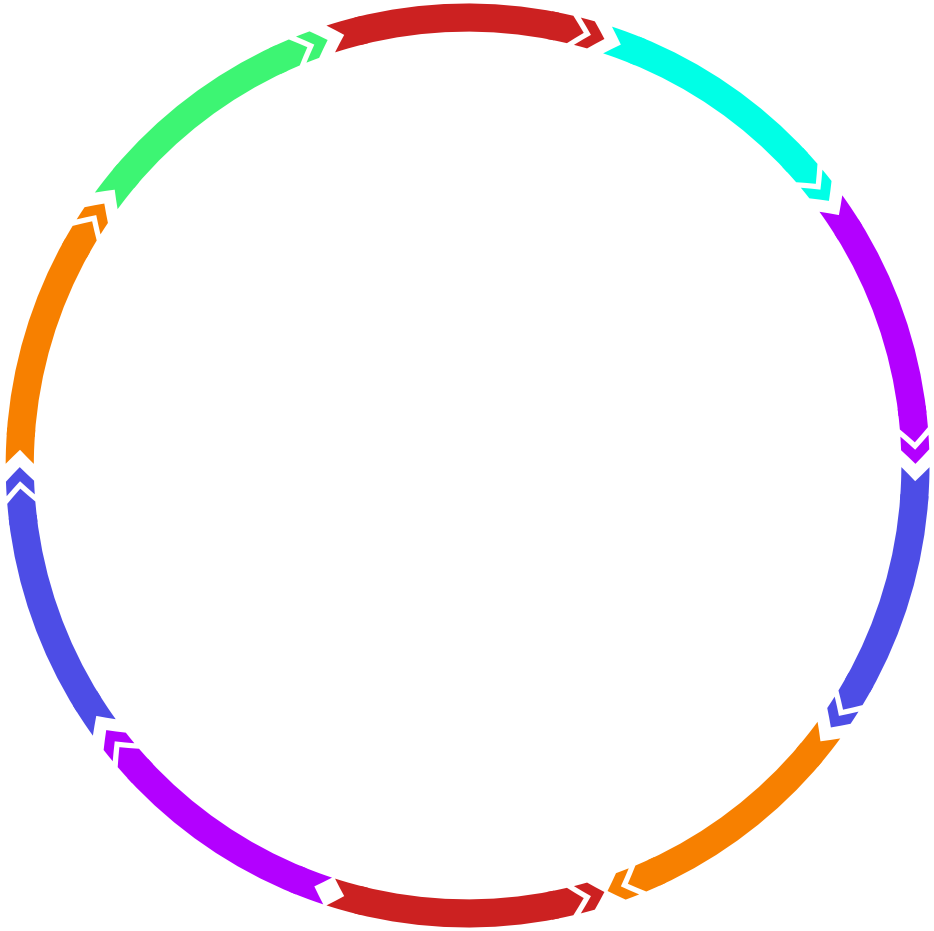


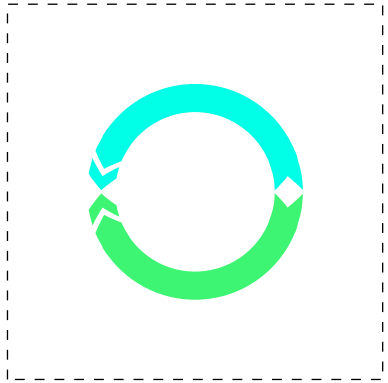
Compeau-style Halving with (Losses,~~Ambiguous~~)



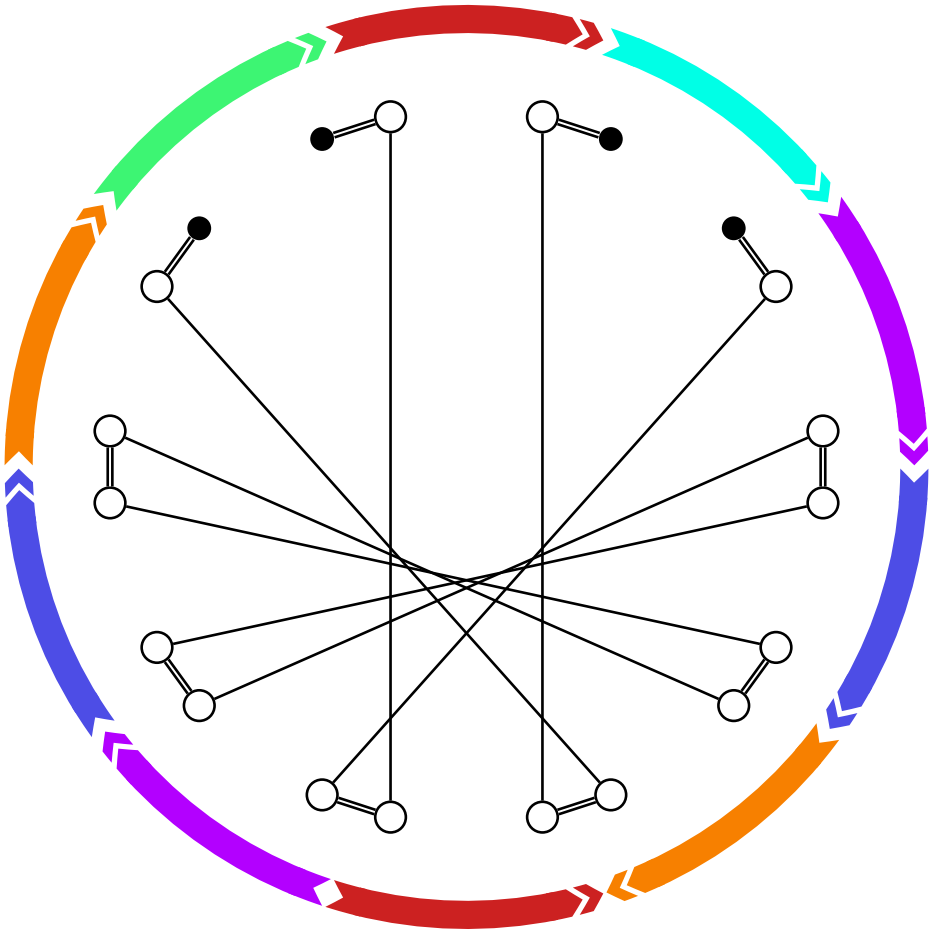


Compeau-style Halving with (Losses,~~Ambiguous~~)

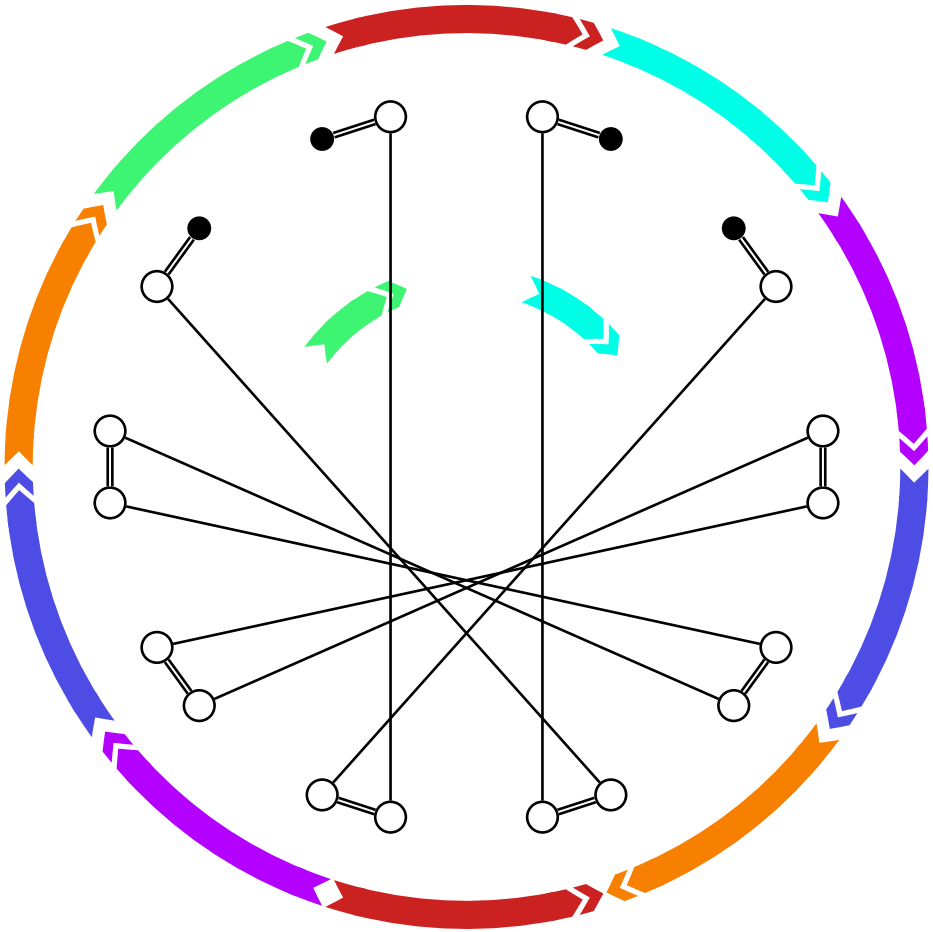




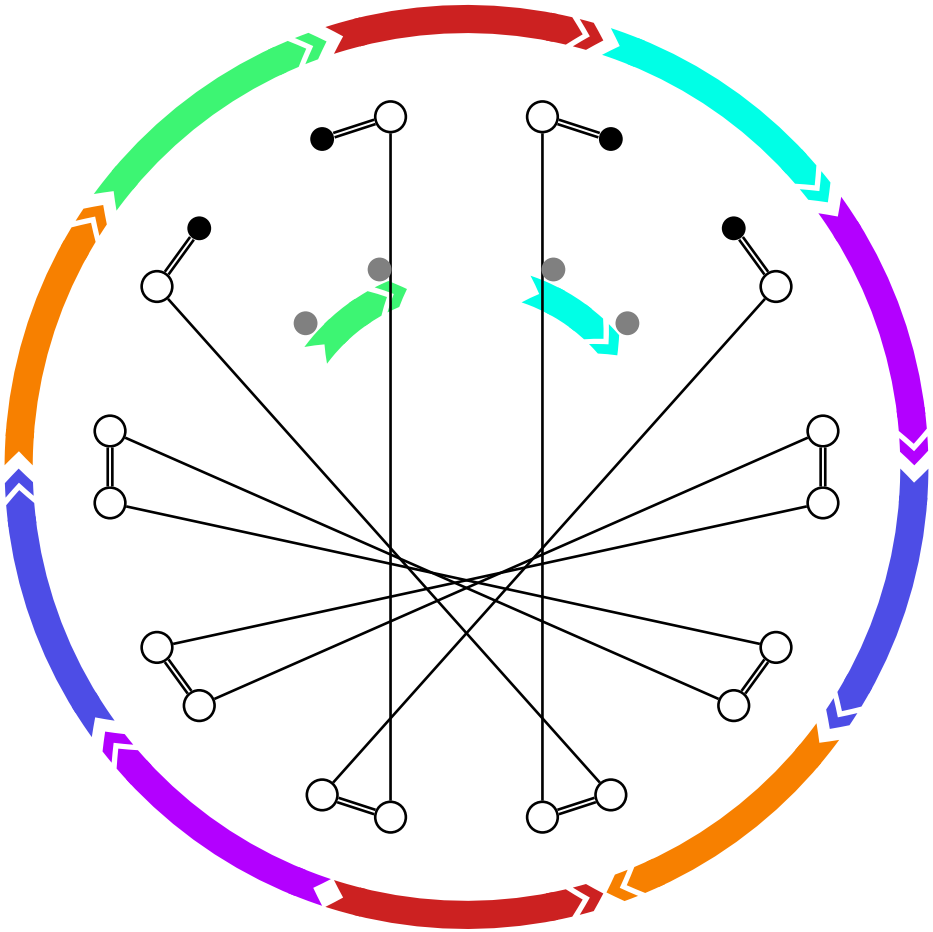
Compeau-style Halving with (Losses,~~Ambiguous~~): Natural Graph



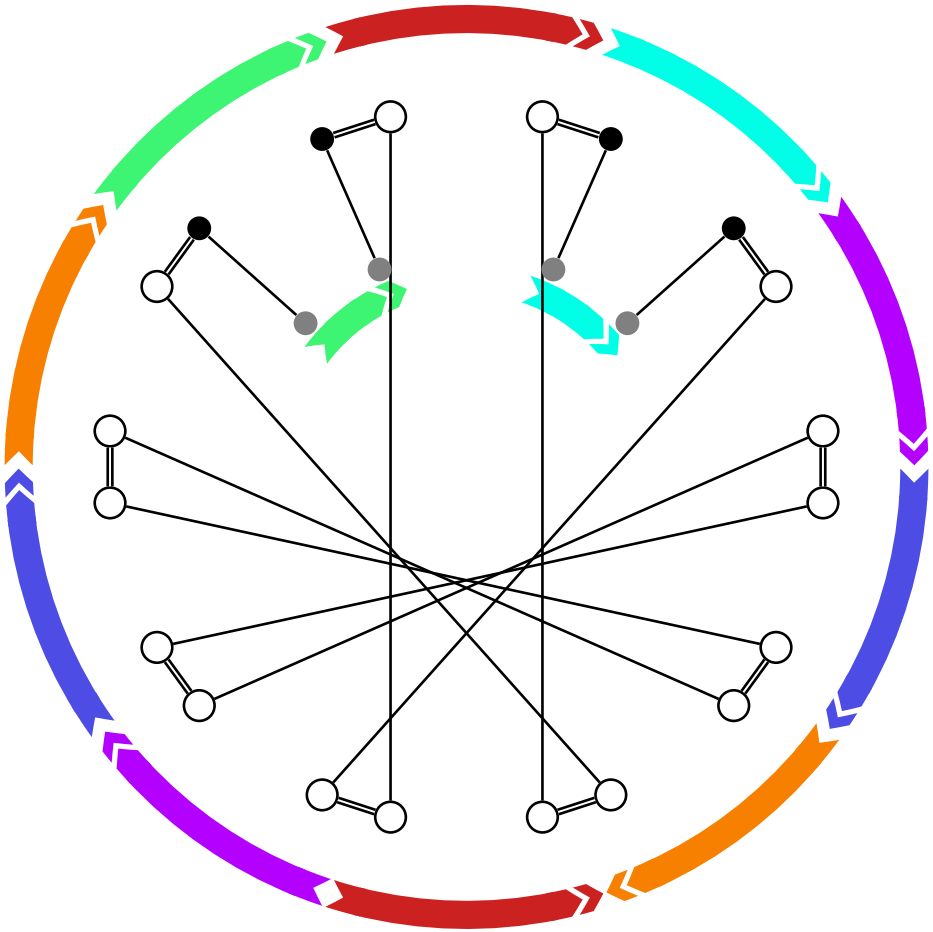
Compeau-style Halving with (Losses,~~Ambiguous~~): Natural Graph



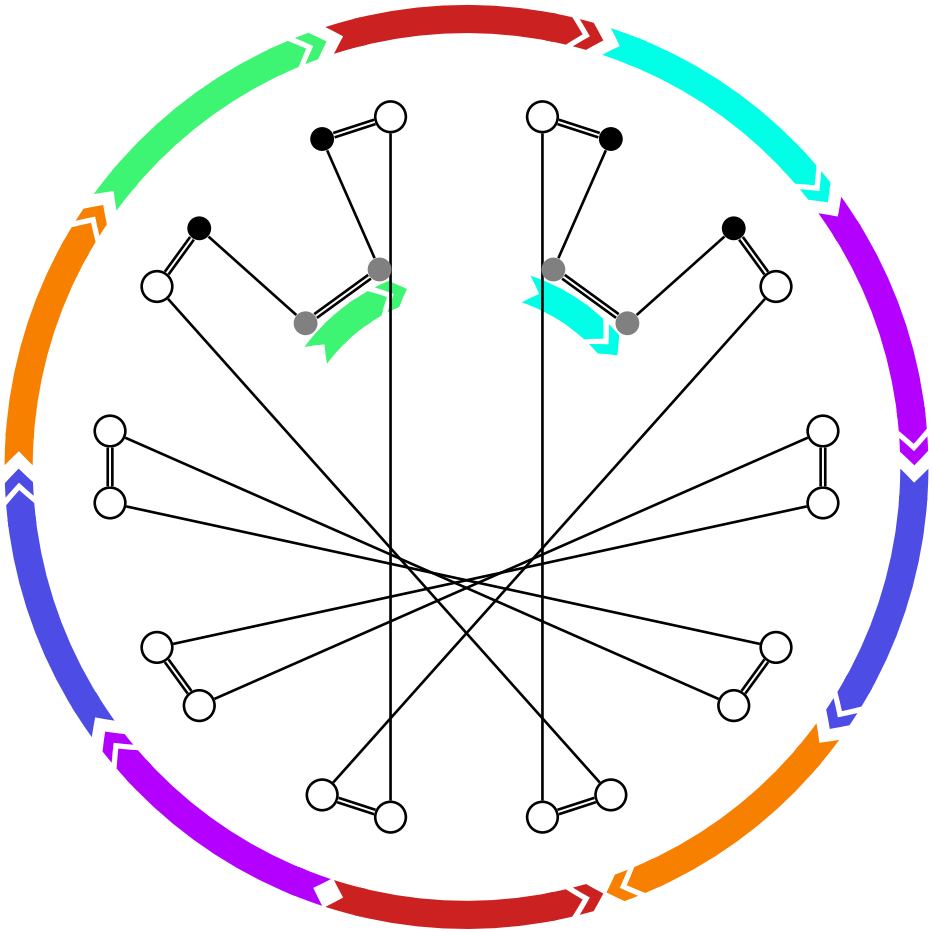
Compeau-style Halving with (Losses,~~Ambiguous~~): Natural Graph



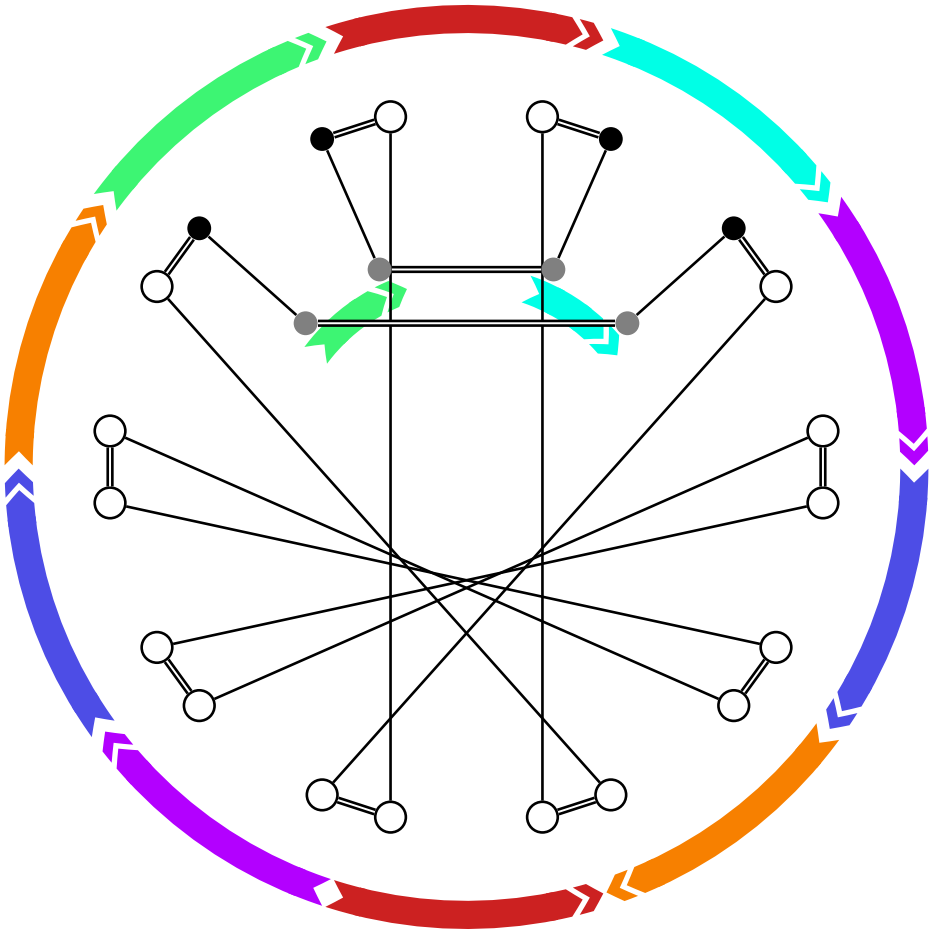
Compeau-style Halving with (Losses,~~Ambiguous~~): Natural Graph



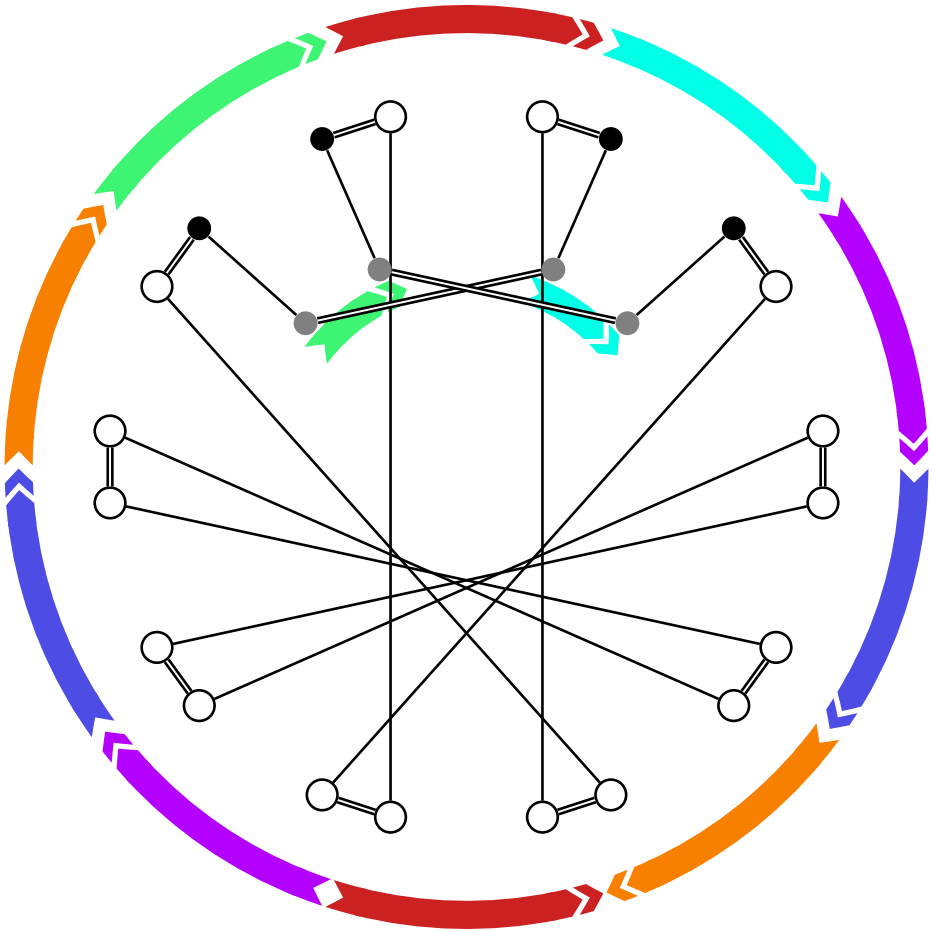
Compeau-style Halving with (Losses,~~Ambiguous~~): Natural Graph



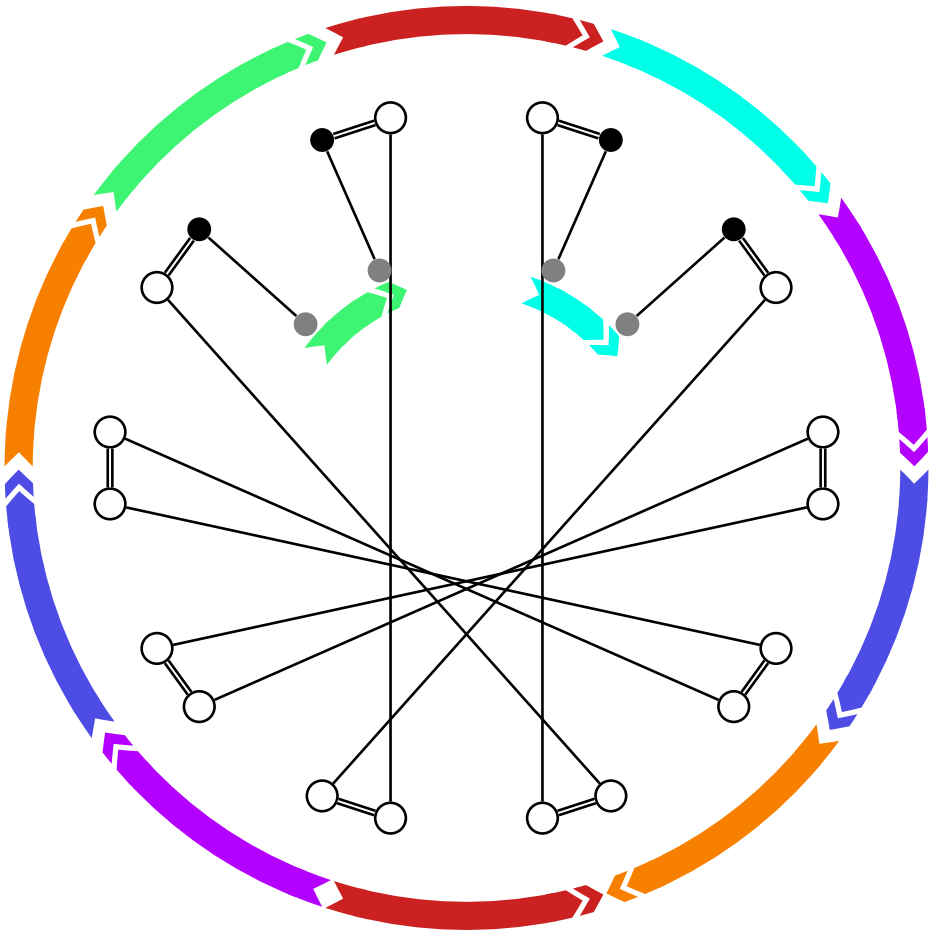
Compeau-style Halving with (Losses,~~Ambiguous~~): Natural Graph



Compeau-style Halving with (Losses,~~Ambiguous~~): Natural Graph

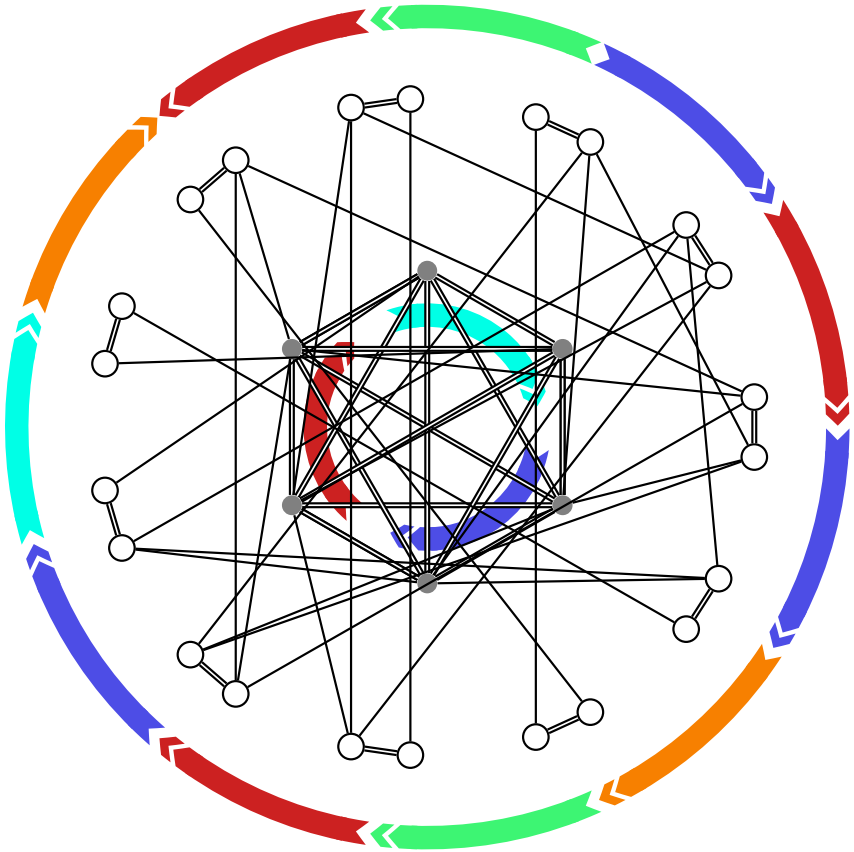


Compeau-style Halving with (Losses,~~Ambiguous~~): Natural Graph



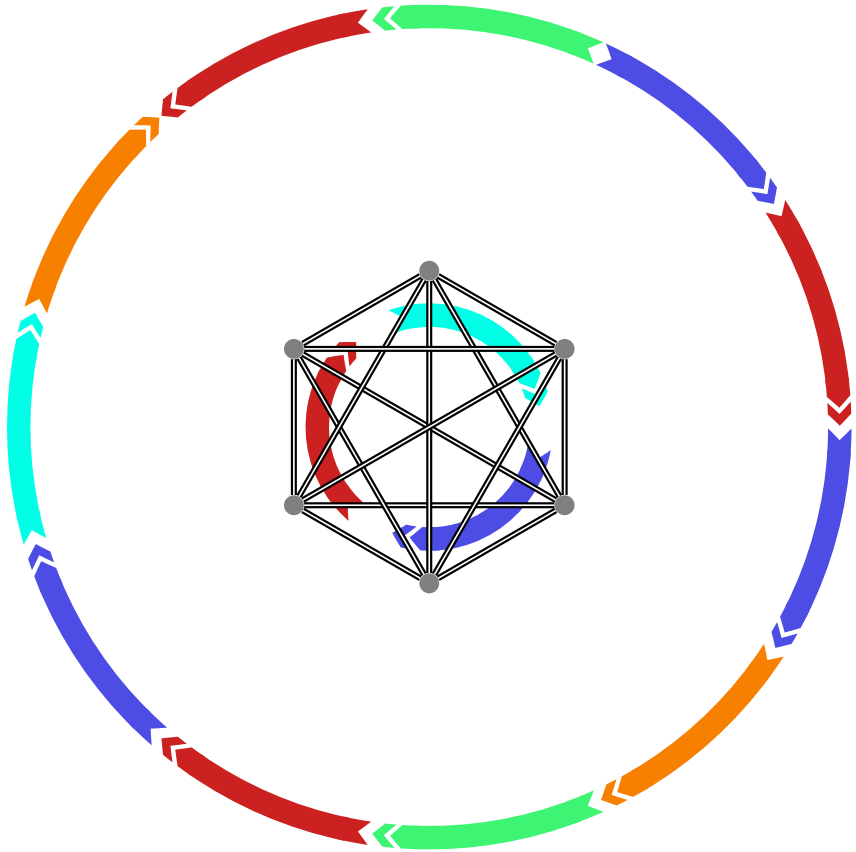
Use the formula for duplicated genomes to determine the optimal completion!

Compeau-style Halving Natural Genomes (Losses,Ambiguous): Supernatural Graph



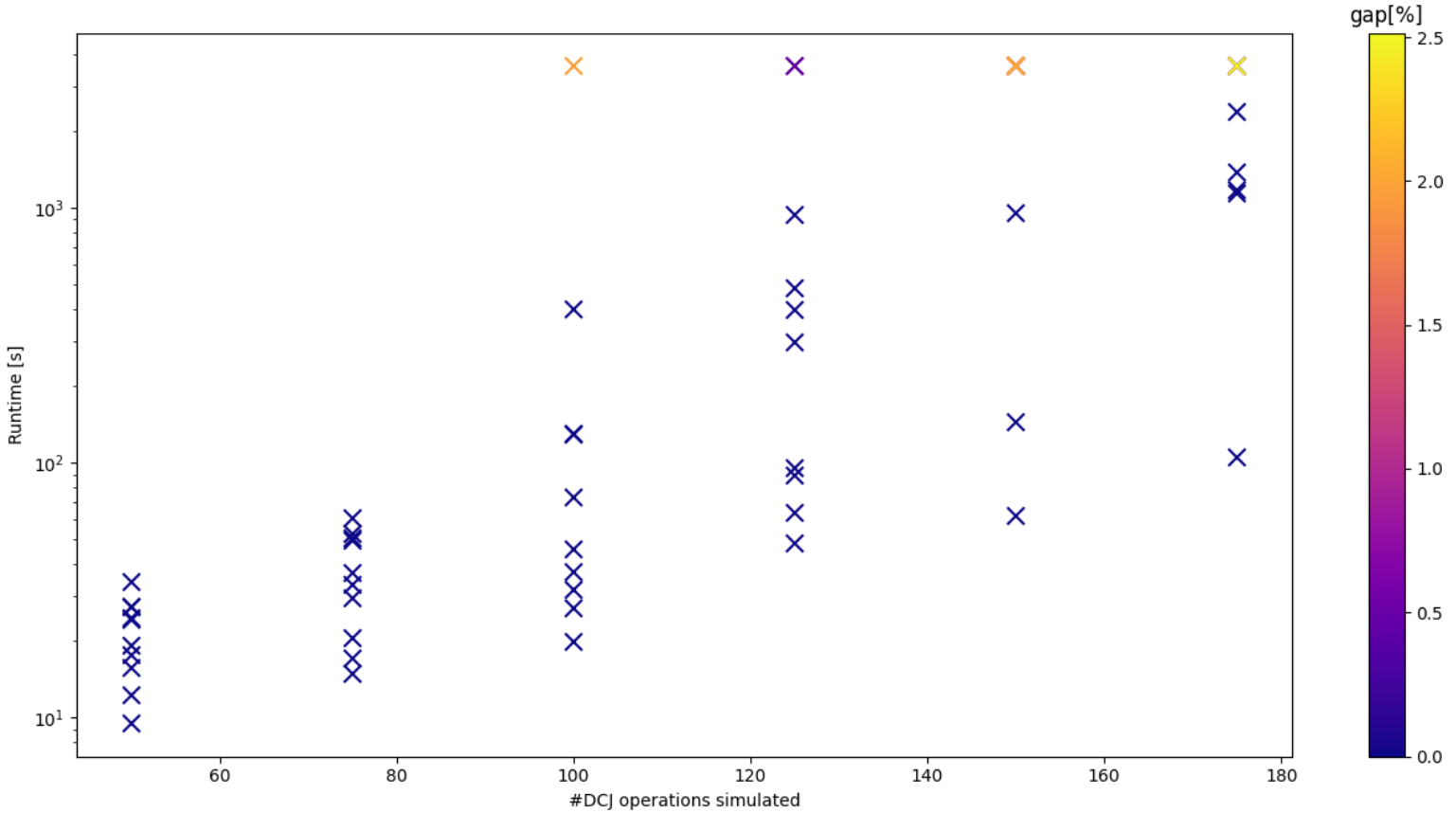
$$h^{id}_{DCJ} = \min_{\substack{\text{Consistent} \\ \text{Decomposition } D}} (n - c_e(D))$$

Compeau-style Halving Natural Genomes (Losses,Ambiguous): Supernatural Graph

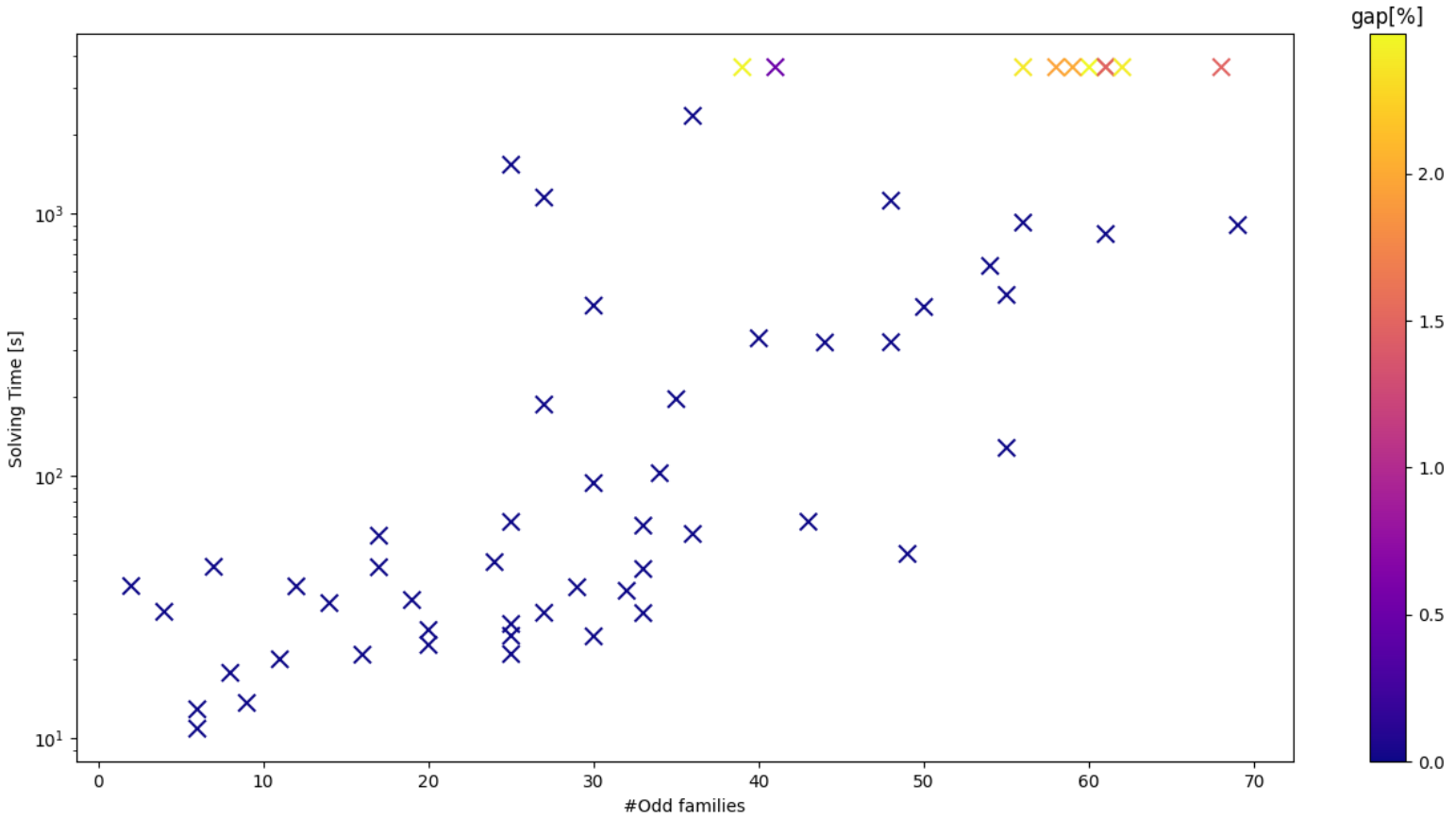


$$h^{id}_{DCJ} = \min_{\substack{\text{Consistent} \\ \text{Decomposition } D}} (n - c_e(D))$$

Compeau-style Halving Natural Genomes: Performance



Compeau-style Halving Natural Genomes: Performance



The BWS Formulation

- Identification of runs

The BWS Formulation

- Identification of runs
- Recombination of runs

The BWS Formulation

BWS

- Identification of runs
- Recombination of runs

Compeau

The BWS Formulation

BWS

- Identification of runs
- Recombination of runs

Compeau

- Closing each even path

The BWS Formulation

BWS

- Identification of runs
- Recombination of runs

Compeau

- Closing each even path
- Closing pairs of odd paths

The BWS Formulation

BWS		Compeau
• Identification of runs	$\equiv$	• Closing each even path
• Recombination of runs		• Closing pairs of odd paths

The BWS Formulation

BWS		Compeau
• Identification of runs	$\equiv$	• Closing each even path
• Recombination of runs	$\equiv$	• Closing pairs of odd paths

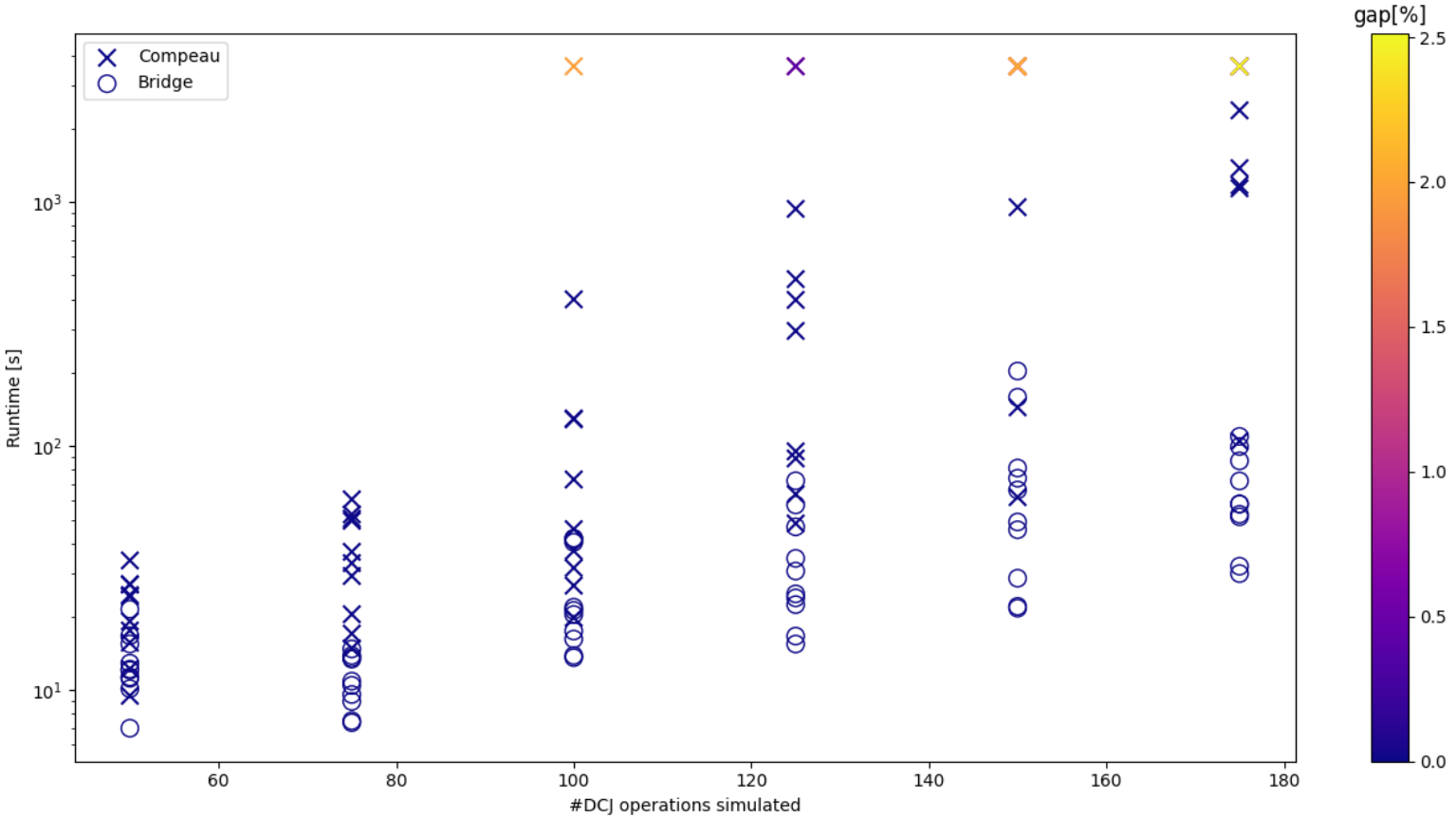
The BWS Formulation

BWS		Compeau
• Identification of runs	$\equiv$	• Closing each even path
• Recombination of runs	$\equiv$	• Closing pairs of odd paths

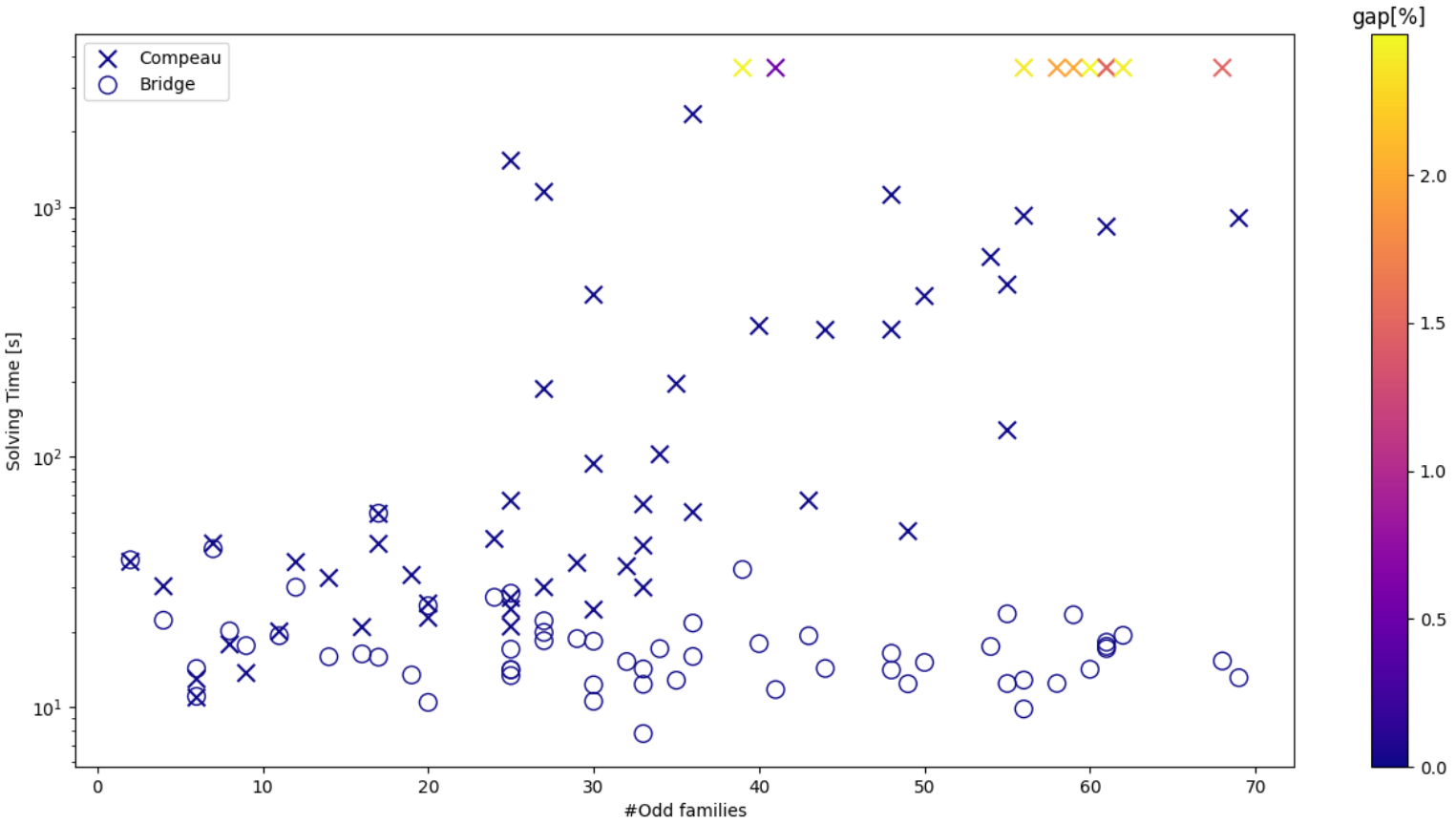
"Bridge" model

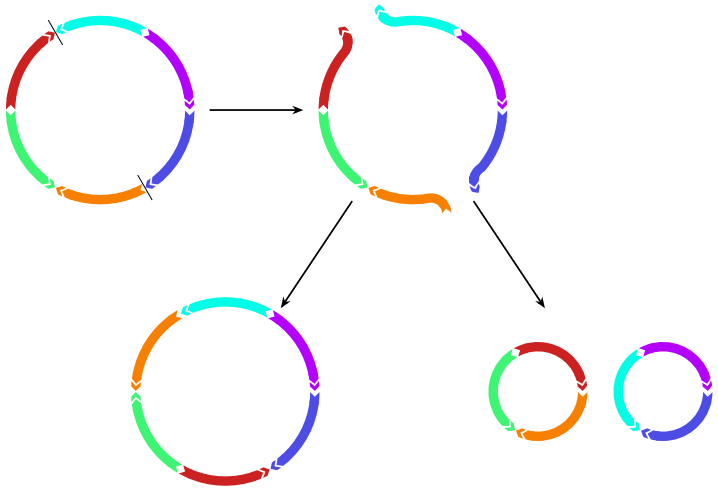
$$h^{id}_{DCJ} = n - c_e + \left\lceil \frac{\tau}{2} \right\rceil$$

“Bridge” Halving Natural Genomes: Performance



“Bridge” Halving Natural Genomes: Performance





Thank you!