Bridging Disparate Views on the DCJ-indel Model

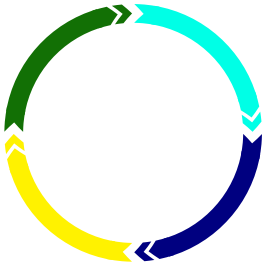
WABI 2023

Leonard Bohnenkämper

Bielefeld University

September 4, 2023

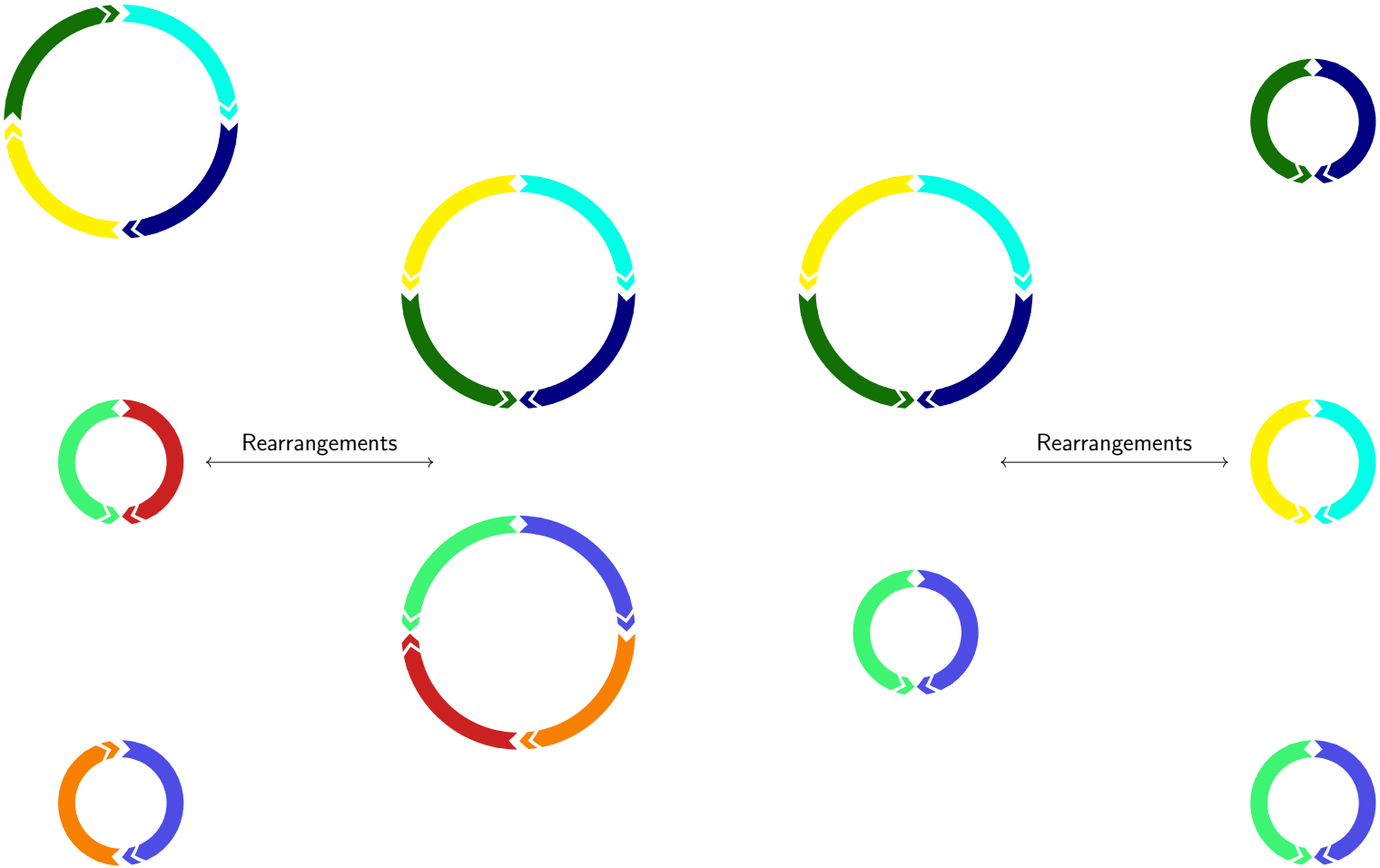
The Genomic Distance Problem



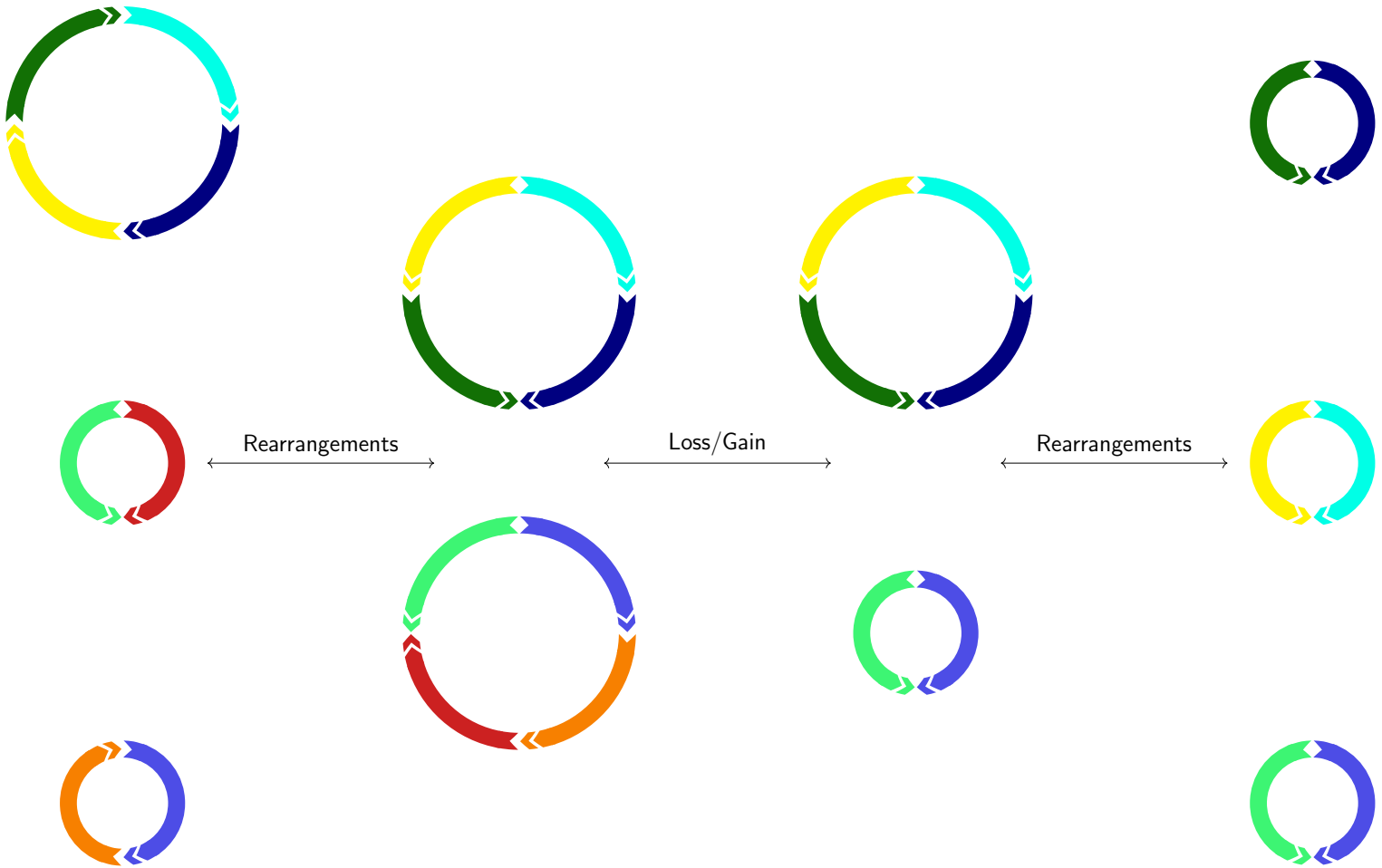
Genome $\mathbb{A}$

Genome $\mathbb{B}$

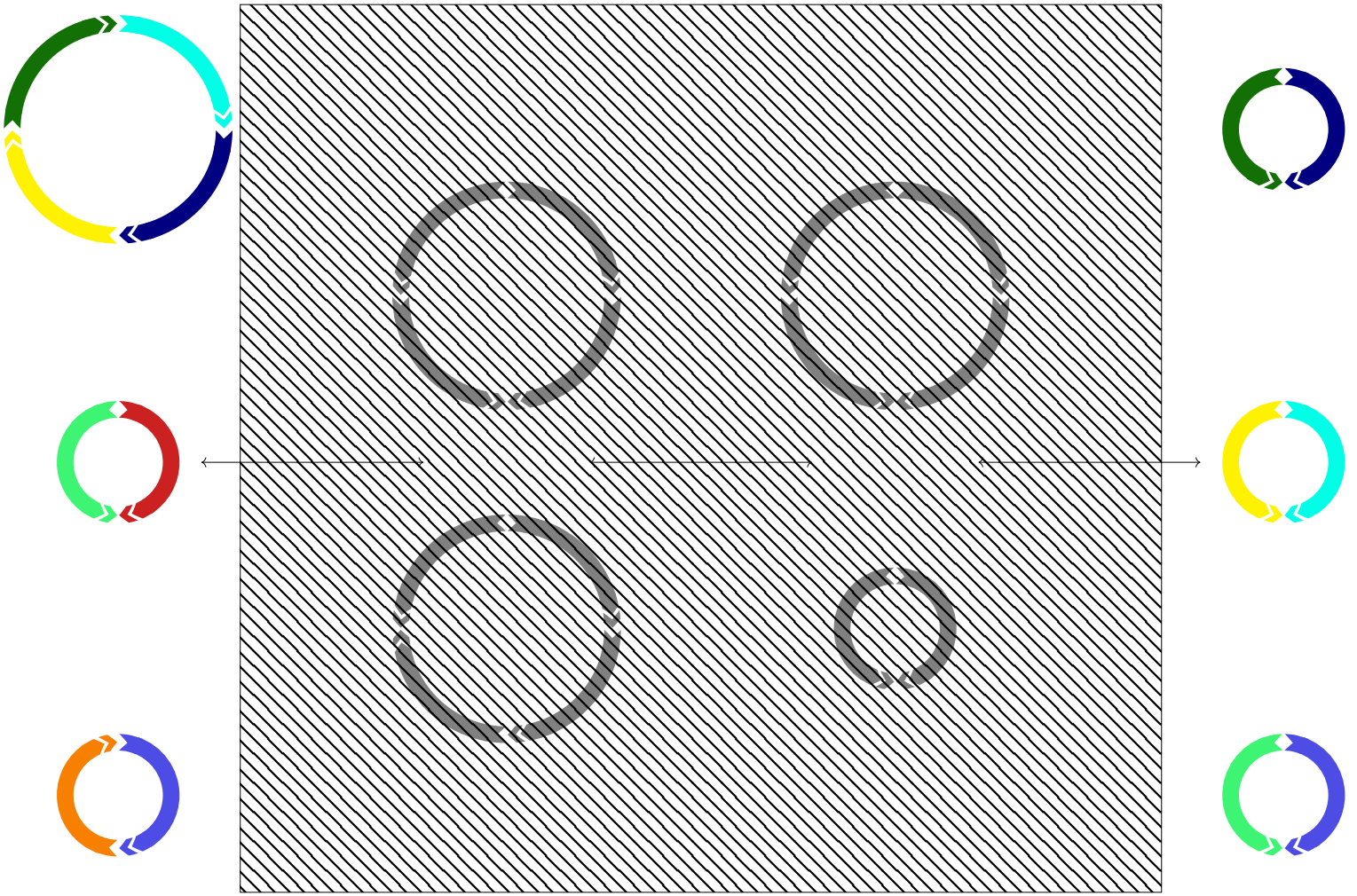
The Genomic Distance Problem



The Genomic Distance Problem



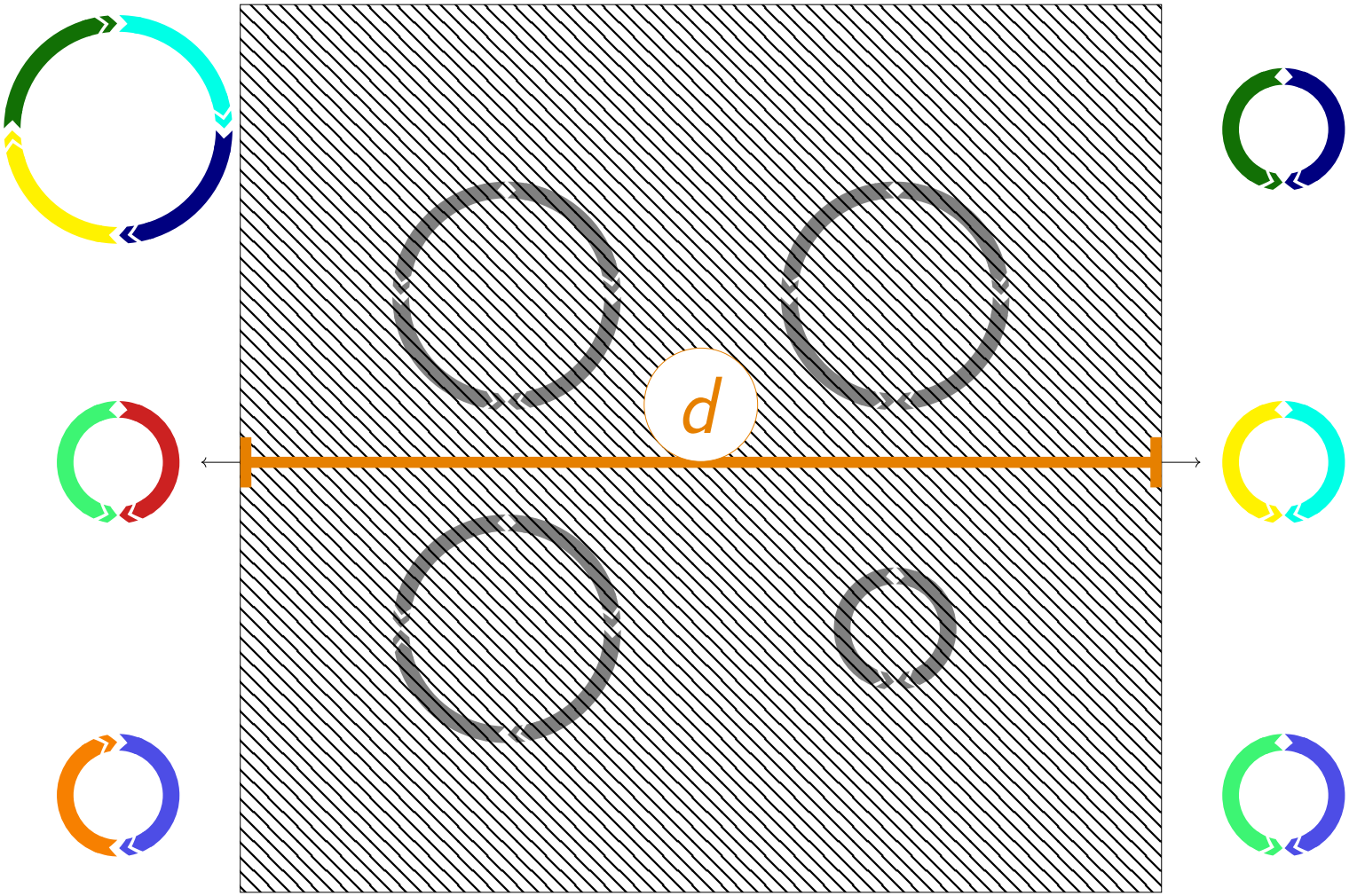
The Genomic Distance Problem



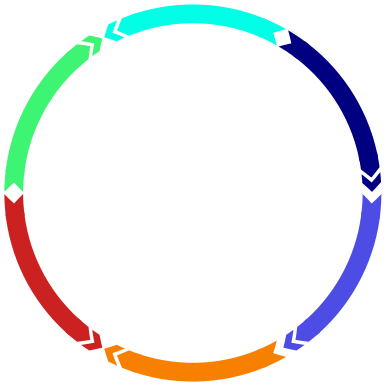
Genome $\mathbb{A}$

Genome $\mathbb{B}$

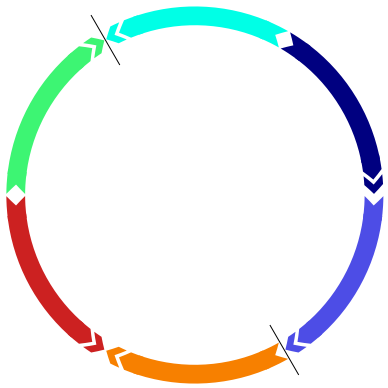
The Genomic Distance Problem



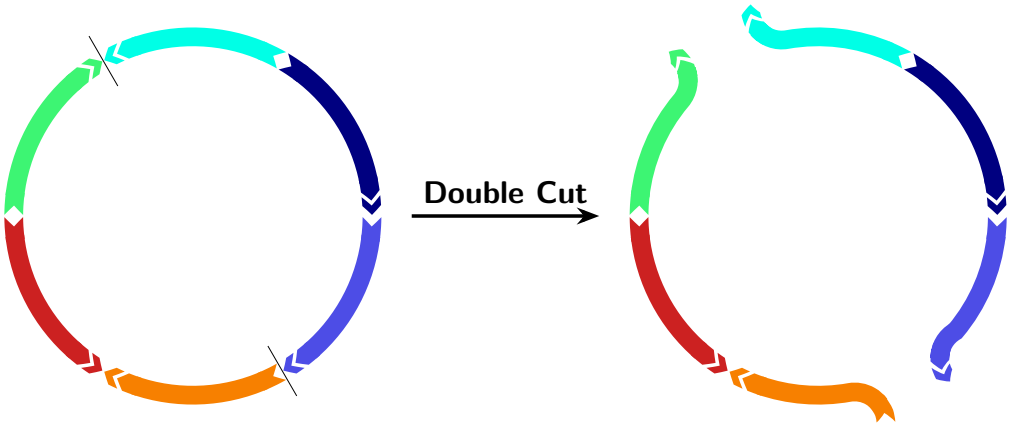
Operations: Double-Cut-and-Join (DCJ)



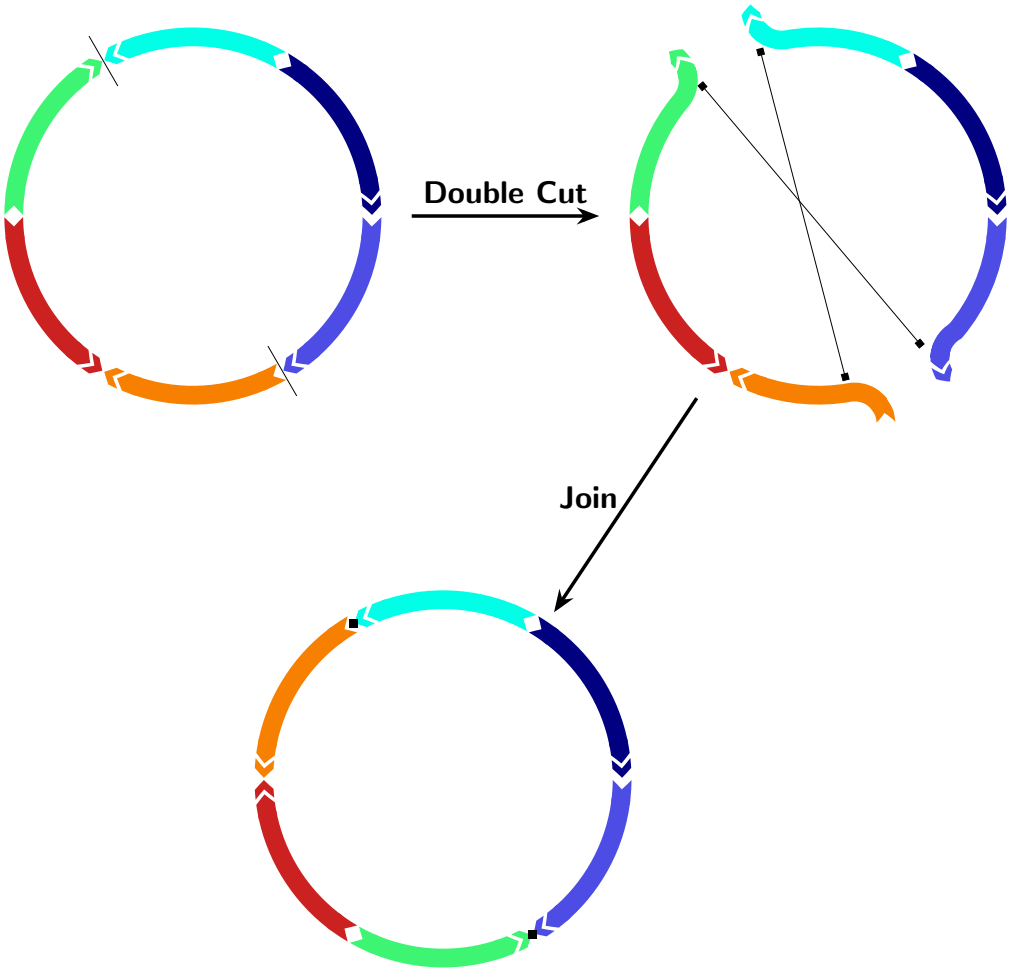
Operations: Double-Cut-and-Join (DCJ)



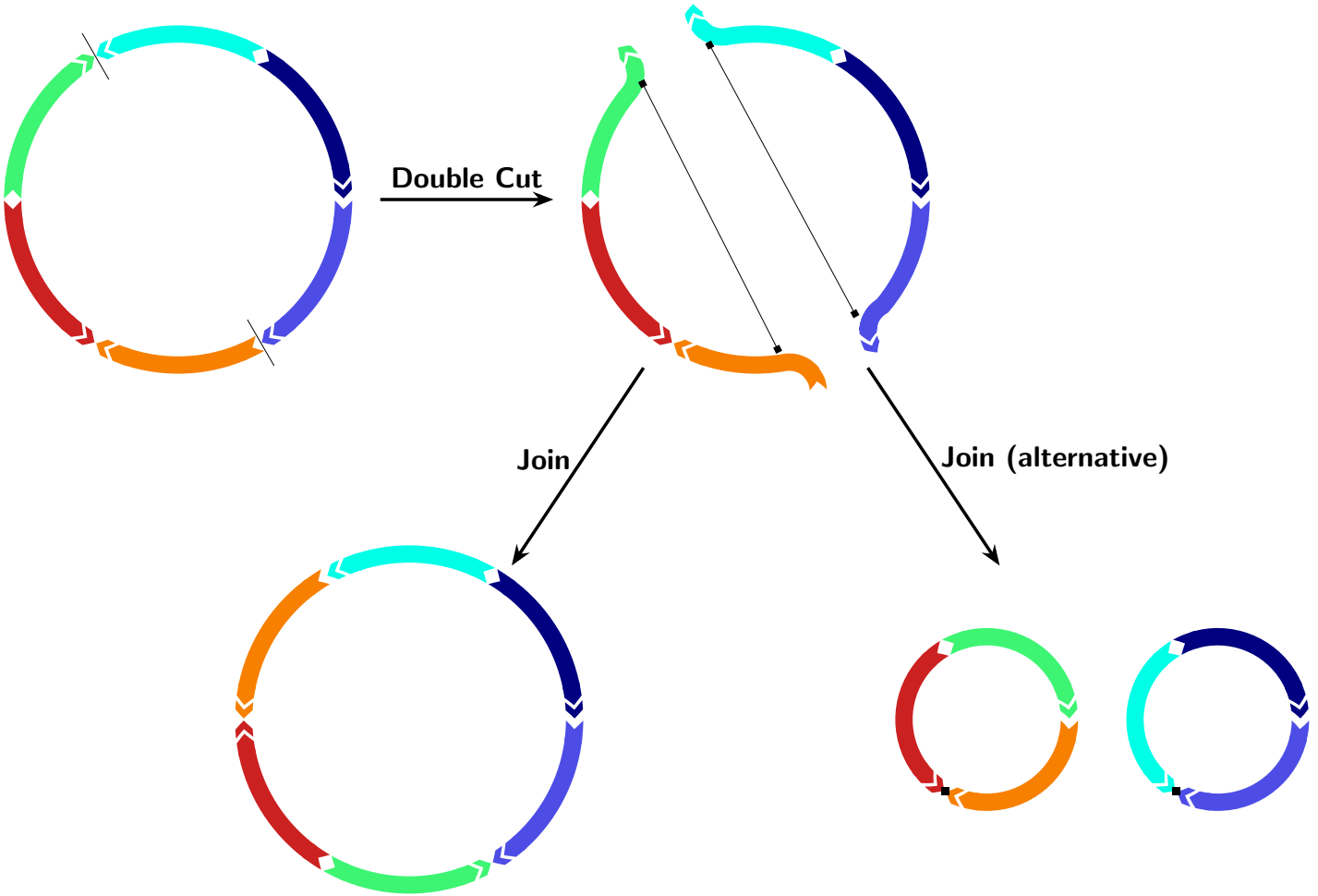
Operations: Double-Cut-and-Join (DCJ)



Operations: Double-Cut-and-Join (DCJ)

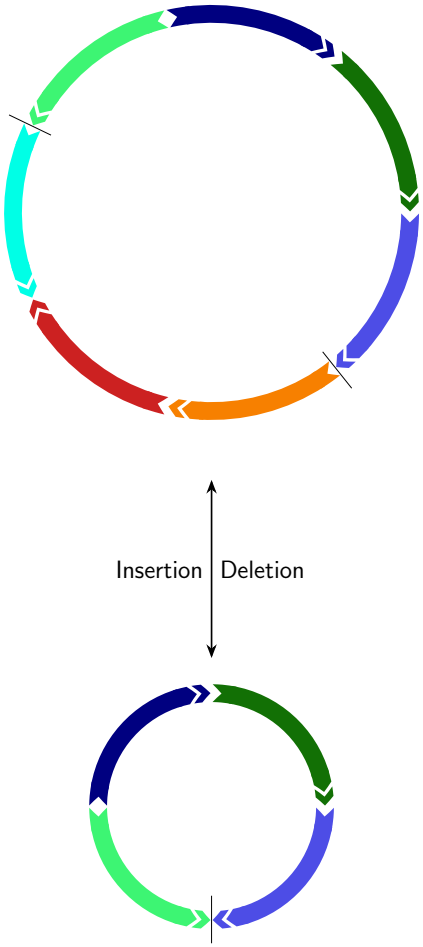


Operations: Double-Cut-and-Join (DCJ)

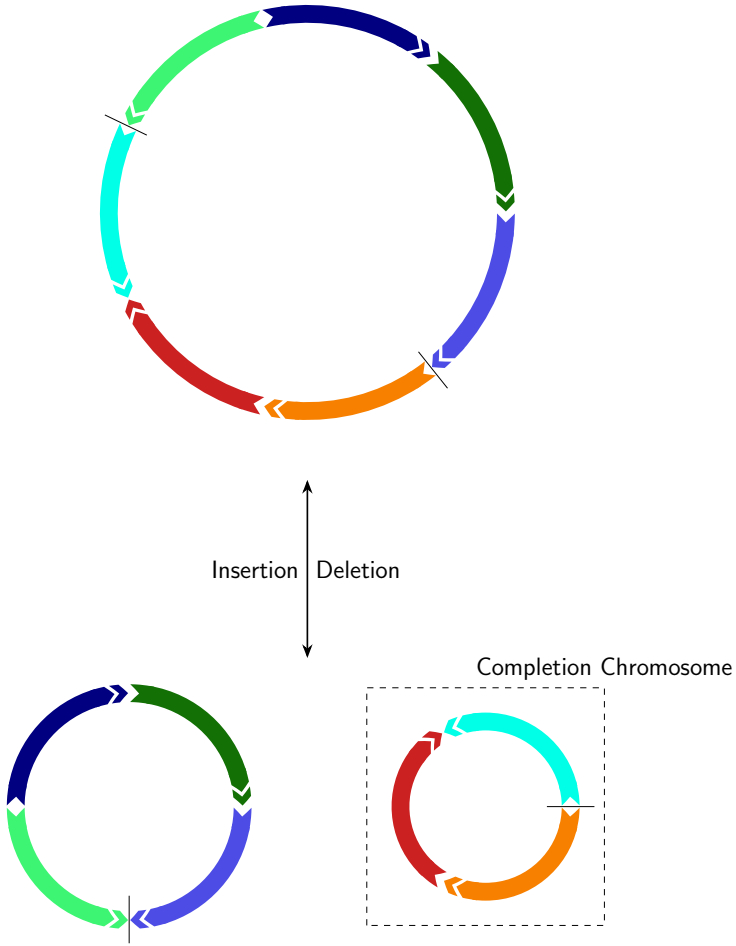


Operations: Indels

Explicitly (BWS, 2011)



Via Completion (Compeau, 2013)



Problems with the BWS- and Compeau-Conceptualizations

- Unwieldy formulas:

Problems with the BWS- and Compeau-Conceptualizations

- Unwieldy formulas:
  - BWS:

$$d_{DCJ}^{id}(A,B) = d_{DCJ}(A,B) + \sum_{C} \lambda(C) - 2P - 3Q - 2T - S - 2M - N$$

Problems with the BWS- and Compeau-Conceptualizations

- Unwieldy formulas:
  - BWS:

$$d_{DCJ}^{id}(A,B) = d_{DCJ}(A,B) + \sum_{C} \lambda(C) - 2P - 3Q - 2T - S - 2M - N$$

	Sources	Reusable Resultants	Unsat. after safe internal ops.					Unsat. after all safe ops. in group					No partner due to
			$A \circ a$	$B \vert a$	$a \vert b$	$A \vert b$	$B \circ b$	$A \circ a$	$B \vert a$	$p_{a \vert b}$	$A \vert b$	$B \circ b$	
P													
1	$AA_{\mathcal{AB}} + BB_{\mathcal{AB}}$		1	1	2	1	1	0	0	0	0	0	
Q													
1	$2AA_{\mathcal{AB}} + BB_{\mathcal{A}} + BB_{\mathcal{B}}$		2	2	2	2	2	0	0	0	0	0	
2	$2BB_{\mathcal{AB}} + AA_{\mathcal{A}} + AA_{\mathcal{B}}$		2	2	2	2	2	0	0	0	0	0	
T													
1	$AA_{\mathcal{AB}} + BB_{\mathcal{A}} + AB_{\mathcal{AB}}$		2	2	2	1	1	0	0	0	0	0	
2	$AA_{\mathcal{AB}} + BB_{\mathcal{B}} + AB_{\mathcal{BA}}$		1	1	2	2	2	0	0	0	0	0	
3	$BB_{\mathcal{AB}} + AA_{\mathcal{A}} + AB_{\mathcal{BA}}$		2	2	2	1	1	0	0	0	0	0	
4	$BB_{\mathcal{AB}} + AA_{\mathcal{B}} + AB_{\mathcal{AB}}$		1	1	2	2	2	0	0	0	0	0	
5	$2AA_{\mathcal{AB}} + BB_{\mathcal{A}}$	$AA_{\mathcal{B}}$	2	2	2	2	0	0	0	0	2	0	$P_{A\vert b} \to T1, Q1, P1;$
6	$2AA_{\mathcal{AB}} + BB_{\mathcal{B}}$	$AA_{\mathcal{A}}$	2	0	2	2	2	2	0	0	0	0	$P_{A \circ a} \to T2, Q1, P1;$
7	$2BB_{\mathcal{AB}} + AA_{\mathcal{A}}$	$BB_{\mathcal{B}}$	2	2	2	0	2	0	0	0	0	2	$P_{B \circ b} \to Q2, P1, T3;$
8	$2BB_{\mathcal{AB}} + AA_{\mathcal{B}}$	$BB_{\mathcal{A}}$	0	2	2	2	2	0	2	0	0	0	$P_{B\vert a} \to Q2, P1, T4;$
S													
1	$AA_{\mathcal{A}} + BB_{\mathcal{A}}$		2	2	0	0	0	0	0	0	0	0	
2	$AA_{\mathcal{B}} + BB_{\mathcal{B}}$		0	0	0	2	2	0	0	0	0	0	
3	$AB_{\mathcal{AB}} + AB_{\mathcal{BA}}$		1	1	2	1	1	0	0	0	0	0	
4	$AA_{\mathcal{AB}} + BB_{\mathcal{A}}$	$AB_{\mathcal{BA}}$	1	2	1	1	0	0	1	1	1	0	$P_{B\vert a} \to S1, T5, T1;$ $\star P_{a\vert b} \to T5, T1, P1;$

Problems with the BWS- and Compeau-Conceptualizations

- Unwieldy formulas:
  - Compeau:

$$d_{DCJ}^{ind}(\Pi,\Gamma) = N - [(c + p^{\pi,\pi} + p^{\gamma,\gamma} + \lfloor\frac{p^{\pi,\gamma}}{2}\rfloor)$$
$$+ \frac{1}{2}(p_{even}^{0} + \min\{p_{odd}^{\pi}, p_{even}^{\pi}\} + \min\{p_{odd}^{\gamma}, p_{even}^{\gamma}\} + \delta)]$$

Problems with the BWS- and Compeau-Conceptualizations

- Unwieldy formulas:
  - Compeau:

$$d_{DCJ}^{ind}(\Pi,\Gamma) = N - [(c + p^{\pi,\pi} + p^{\gamma,\gamma} + \lfloor \frac{p^{\pi,\gamma}}{2} \rfloor) + \frac{1}{2}(p_{even}^{0} + \min\{p_{odd}^{\pi}, p_{even}^{\pi}\} + \min\{p_{odd}^{\gamma}, p_{even}^{\gamma}\} + \delta)]$$

$\delta = 1$ when $p^{\pi,\gamma}$ odd and either $p_{odd}^{\pi} > p_{even}^{\pi}, p_{odd}^{\gamma} > p_{even}^{\gamma}$ or $p_{odd}^{\pi} < p_{even}^{\pi}, p_{odd}^{\gamma} < p_{even}^{\gamma}$; otherwise $\delta = 0$.

Problems with the BWS- and Compeau-Conceptualizations

- Unwieldy formulas:
  - BWS:

$$d^{id}_{DCJ}(A,B) = d_{DCJ}(A,B) + \sum_{C} \lambda(C) - 2P - 3Q - 2T - S - 2M - N$$

  - Compeau:

$$d^{ind}_{DCJ}(\Pi,\Gamma) = N - [(c + p^{\pi,\pi} + p^{\gamma,\gamma} + \lfloor \frac{p^{\pi,\gamma}}{2} \rfloor) + \frac{1}{2}(p^{0}_{even} + \min\{p^{\pi}_{odd}, p^{\pi}_{even}\} + \min\{p^{\gamma}_{odd}, p^{\gamma}_{even}\} + \delta)]$$

Problems with the BWS- and Compeau-Conceptualizations

- Unwieldy formulas:
  - BWS:

$$d^{id}_{DCJ}(A,B) = d_{DCJ}(A,B) + \sum_{C} \lambda(C) - 2P - 3Q - 2T - S - 2M - N$$

  - Compeau:

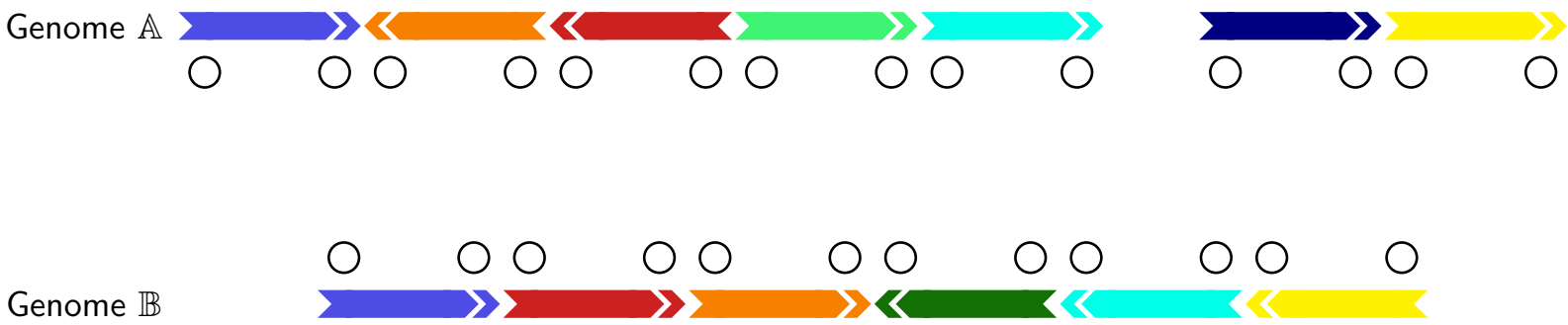
$$d^{ind}_{DCJ}(\Pi,\Gamma) = N - [(c + p^{\pi,\pi} + p^{\gamma,\gamma} + \lfloor \frac{p^{\pi,\gamma}}{2} \rfloor) + \frac{1}{2}(p^{0}_{even} + \min\{p^{\pi}_{odd}, p^{\pi}_{even}\} + \min\{p^{\gamma}_{odd}, p^{\gamma}_{even}\} + \delta)]$$

- Models known to be equivalent, but unclear how formulas relate.

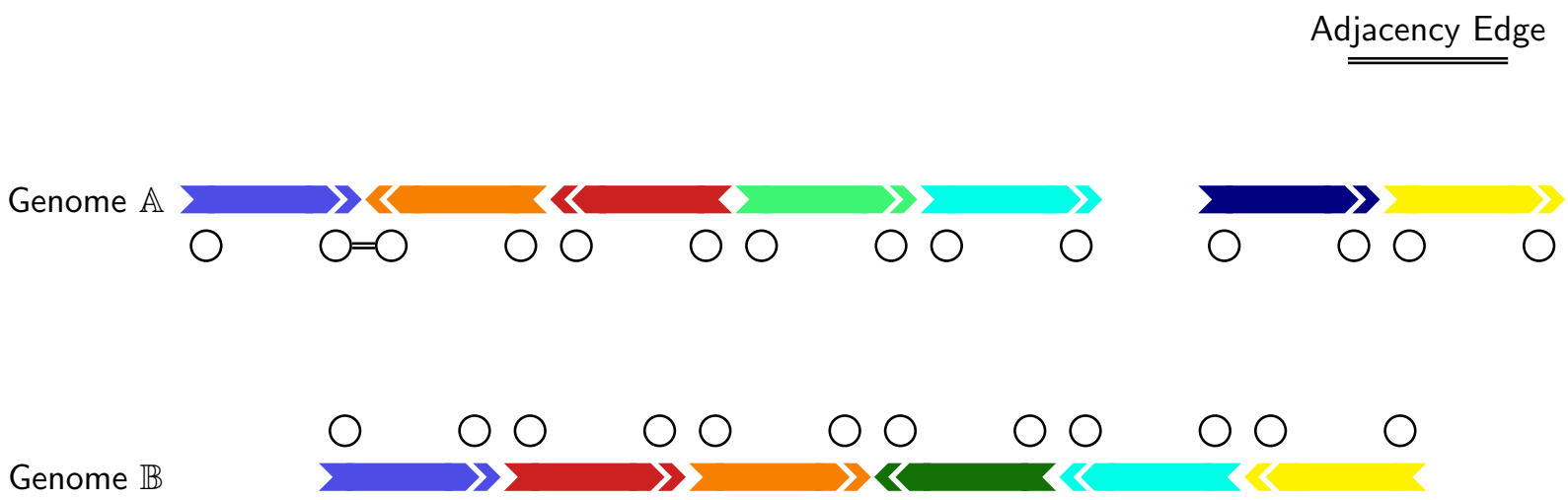
The Relational Diagram

The Relational Diagram

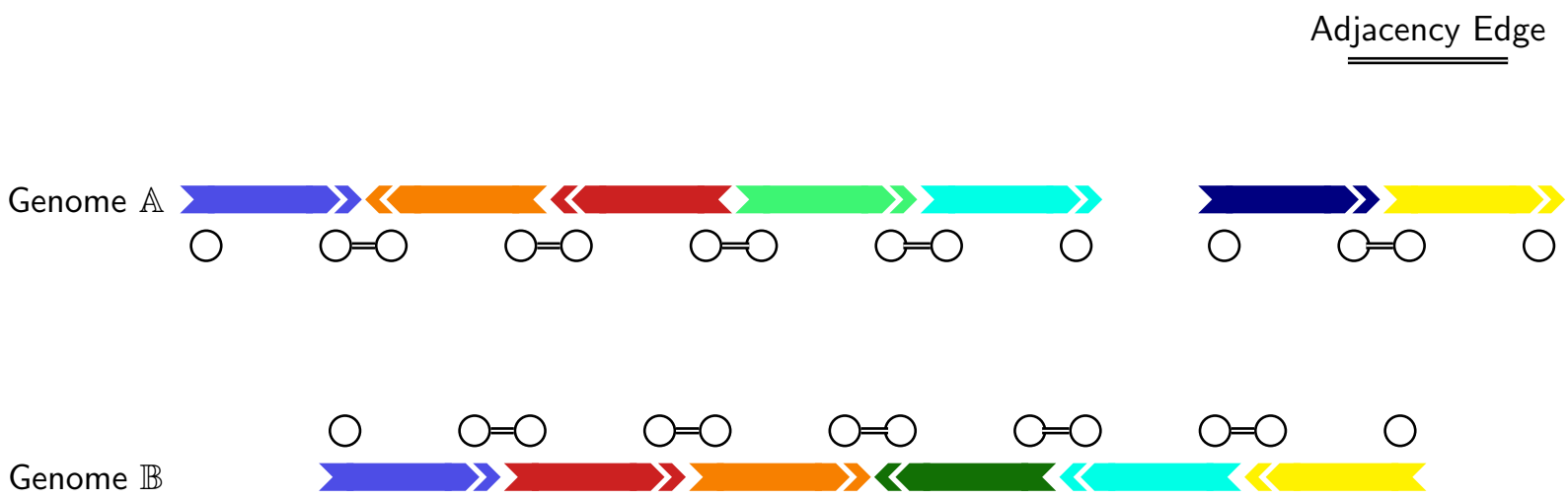
The Relational Diagram



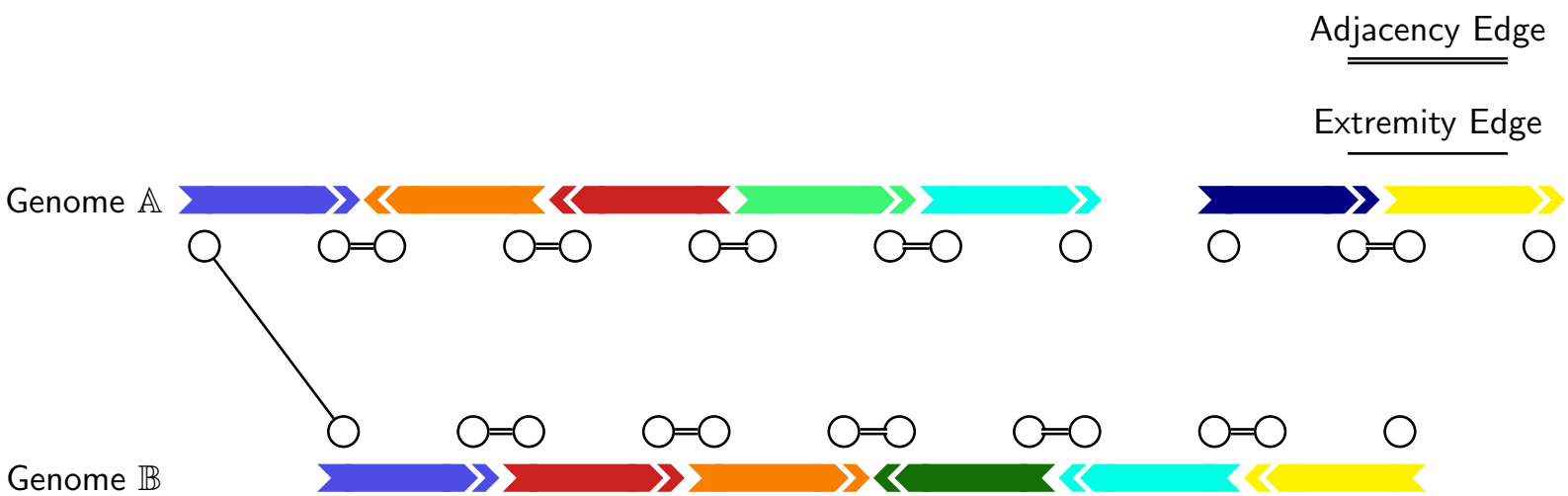
The Relational Diagram



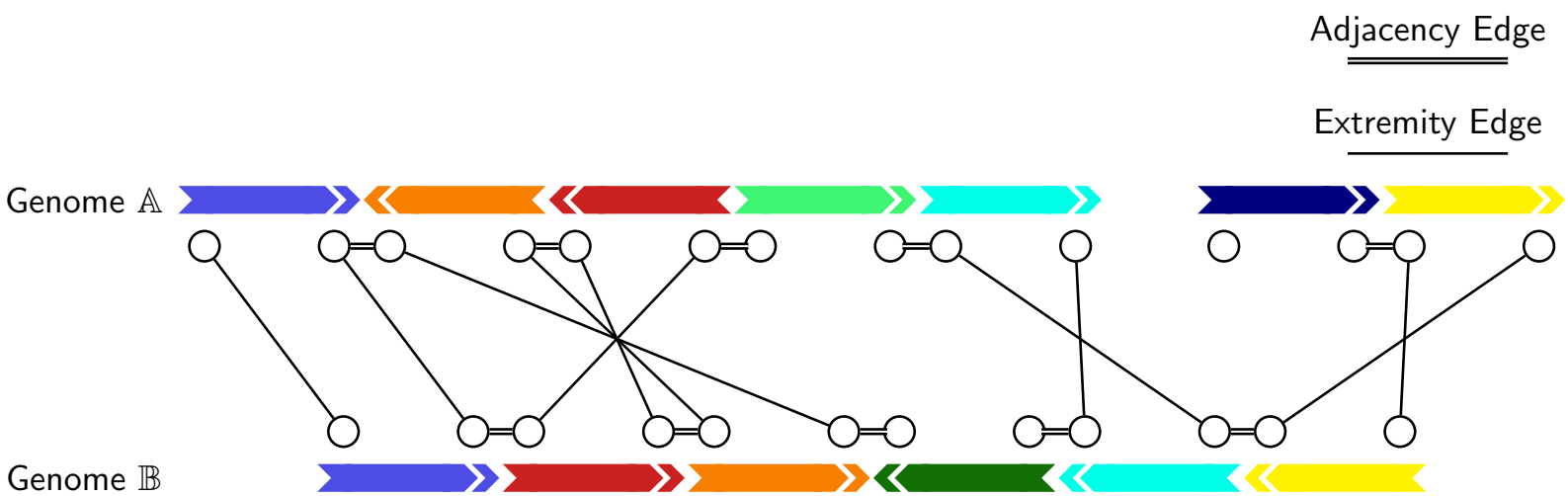
The Relational Diagram



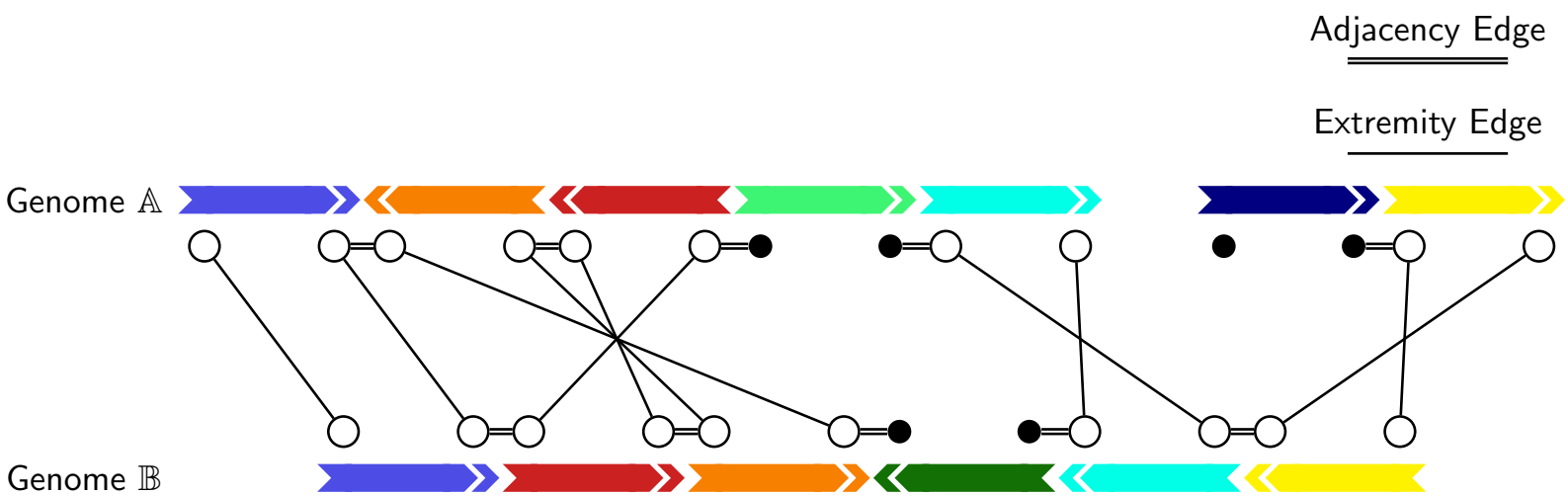
The Relational Diagram



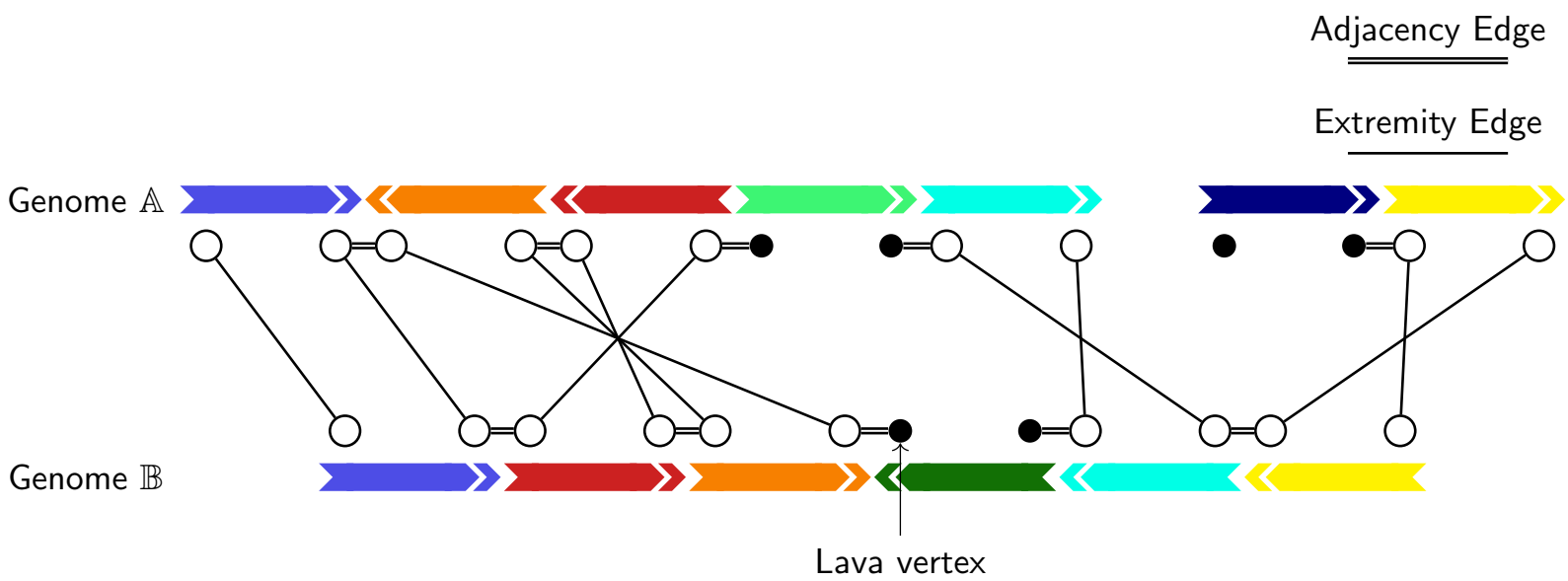
The Relational Diagram



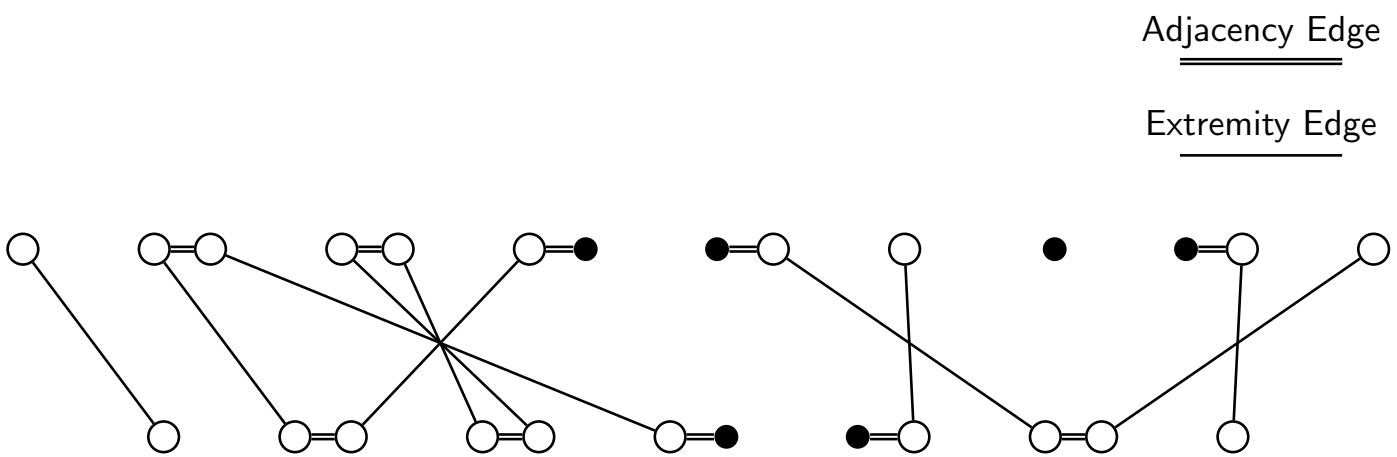
The Relational Diagram



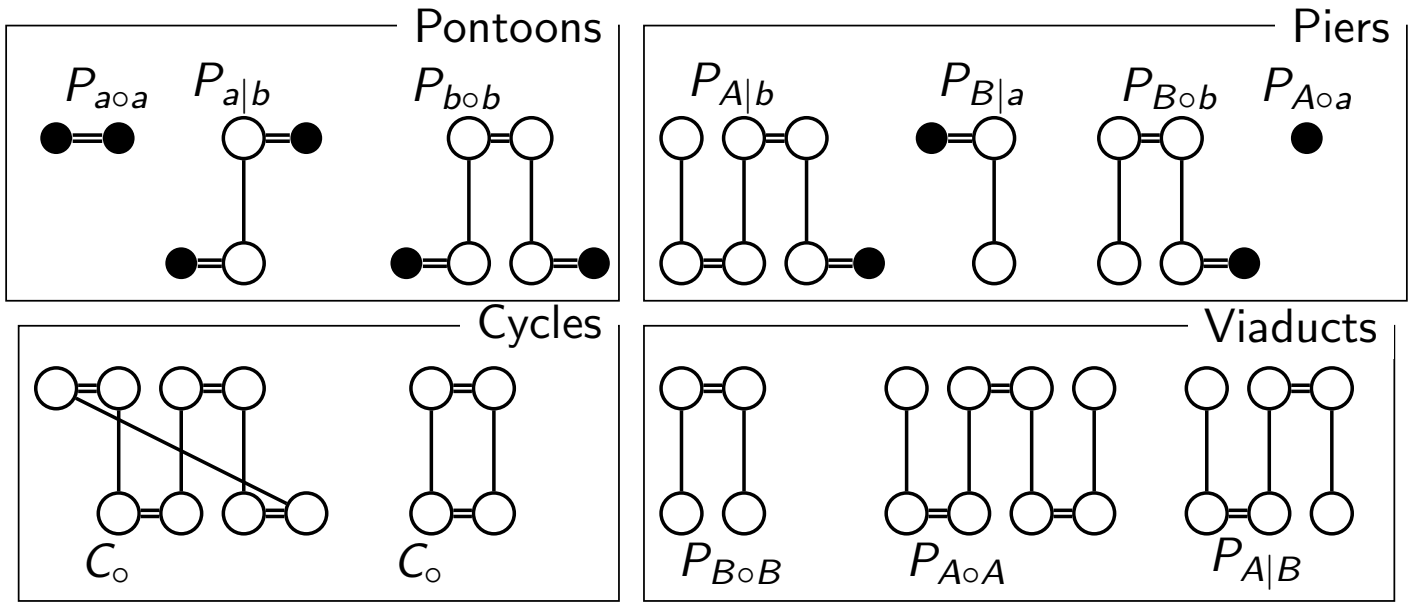
The Relational Diagram



The Relational Diagram

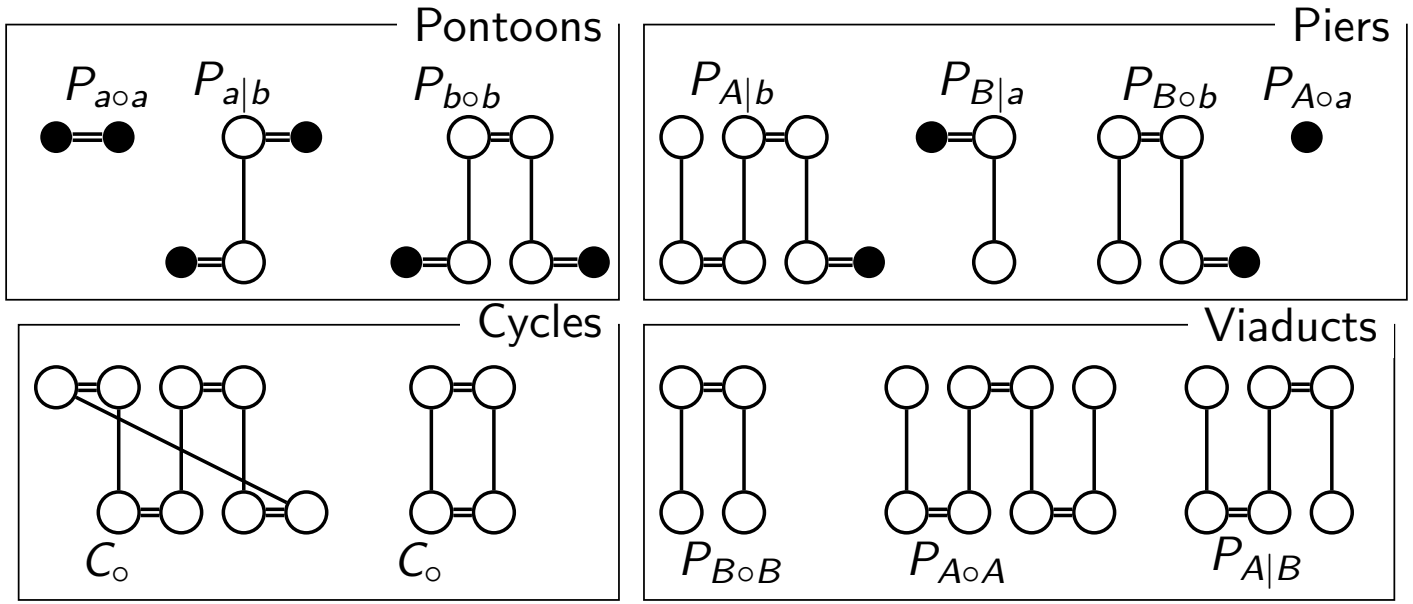


Types of Components



Earlier work, Recomb-CG 2023.

A New Formula



Earlier work, Recomb-CG 2023.

$$d_{DCJ}^{id}(\mathbb{A},\mathbb{B}) = n - c_\circ + \left\lceil \frac{p_{a|b} + \max(p_{A\circ a}, p_{B|a}) + \max(p_{A|b}, p_{B\circ b}) - p_{A|B}}{2} \right\rceil$$

Safe Operations

$$d_{DCJ}^{id}(\mathbb{A}, \mathbb{B}) = n - c_\circ + \left\lceil \frac{p_{a|b} + \max(p_{A\circ a}, p_{B|a}) + \max(p_{A|b}, p_{B\circ b}) - p_{A|B}}{2} \right\rceil$$

Safe Operations

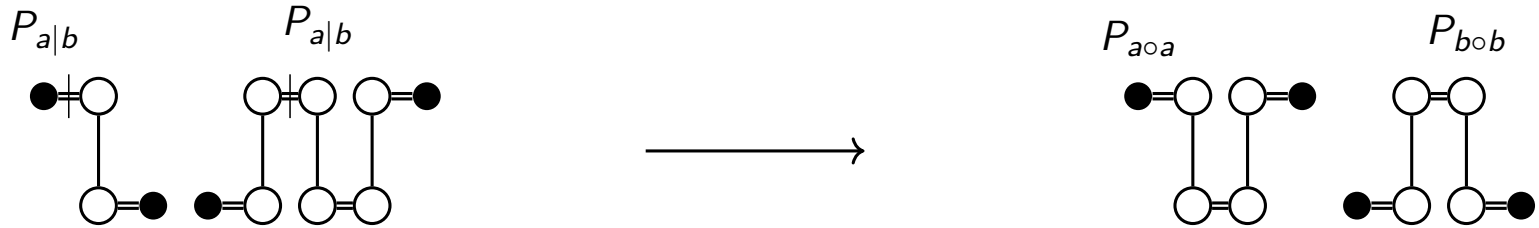
$$d_{DCJ}^{id}(\mathbb{A},\mathbb{B}) = n - c_\circ + \left\lceil \frac{p_{a|b} + \max(p_{A\circ a}, p_{B|a}) + \max(p_{A|b}, p_{B\circ b}) - p_{A|B}}{2} \right\rceil$$

Safe operations reduce the formula by 1, no matter the rest of the graph.

Safe Operations

$$d_{DCJ}^{id}(\mathbb{A},\mathbb{B}) = n - c_\circ + \left\lceil \frac{p_{a|b} + \max(p_{A\circ a}, p_{B|a}) + \max(p_{A|b}, p_{B\circ b}) - p_{A|B}}{2} \right\rceil$$

Safe operations reduce the formula by 1, no matter the rest of the graph.

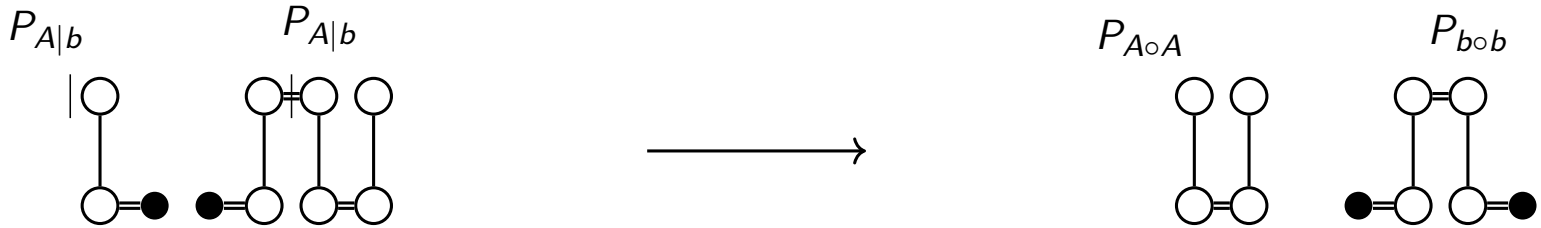


Safe operation

Safe Operations

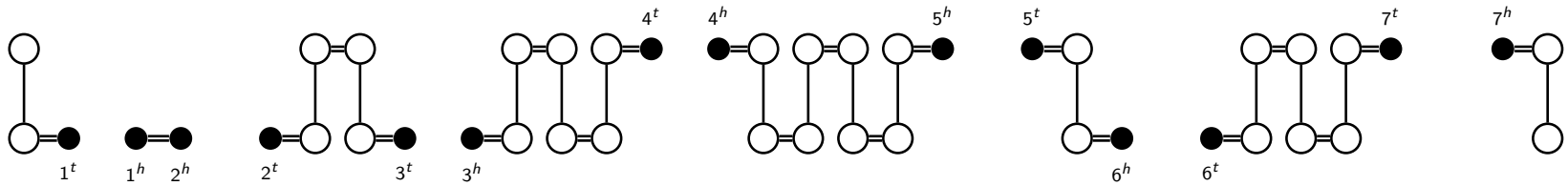
$$d_{DCJ}^{id}(\mathbb{A}, \mathbb{B}) = n - c_\circ + \left\lceil \frac{p_{a|b} + \max(p_{A\circ a}, p_{B|a}) + \max(p_{A|b}, p_{B\circ b}) - p_{A|B}}{2} \right\rceil$$

Safe operations reduce the formula by 1, no matter the rest of the graph.

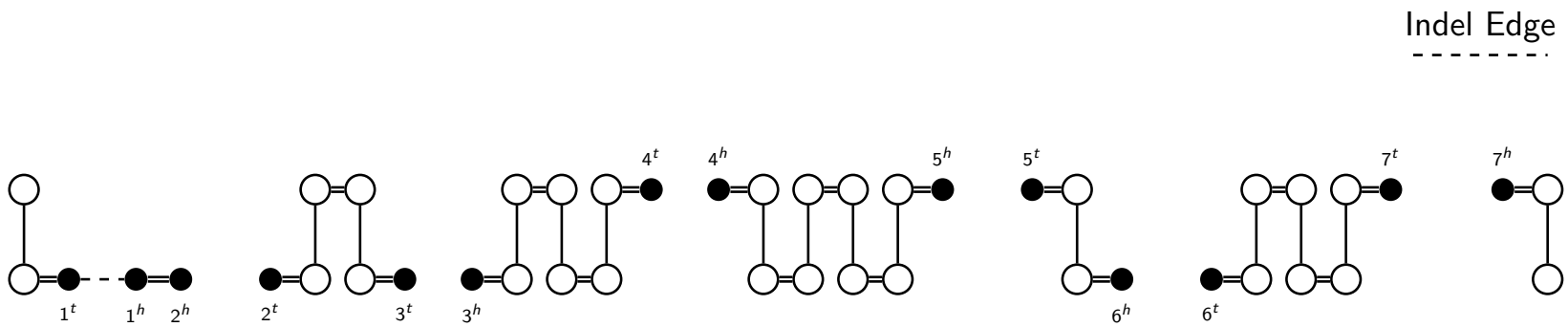


Unsafe operation

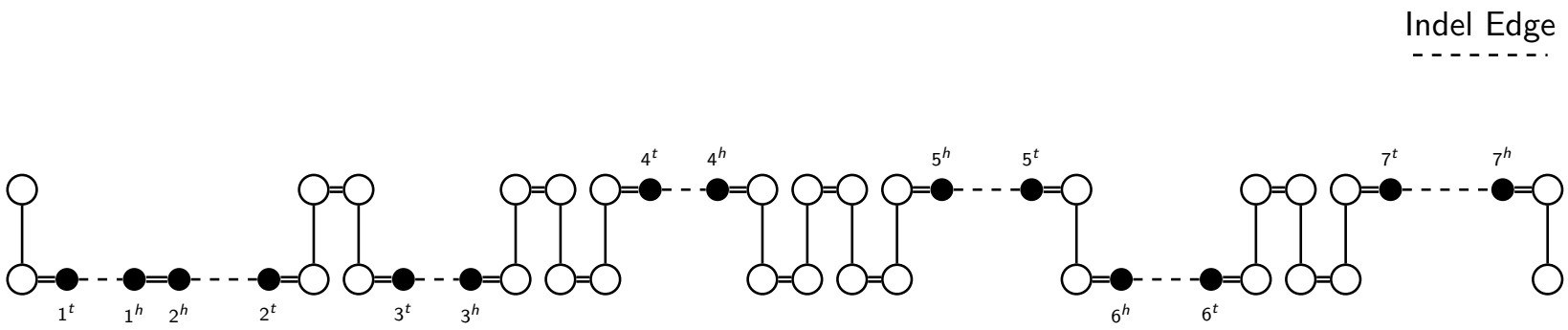
BWS-Conceptualization: Basics



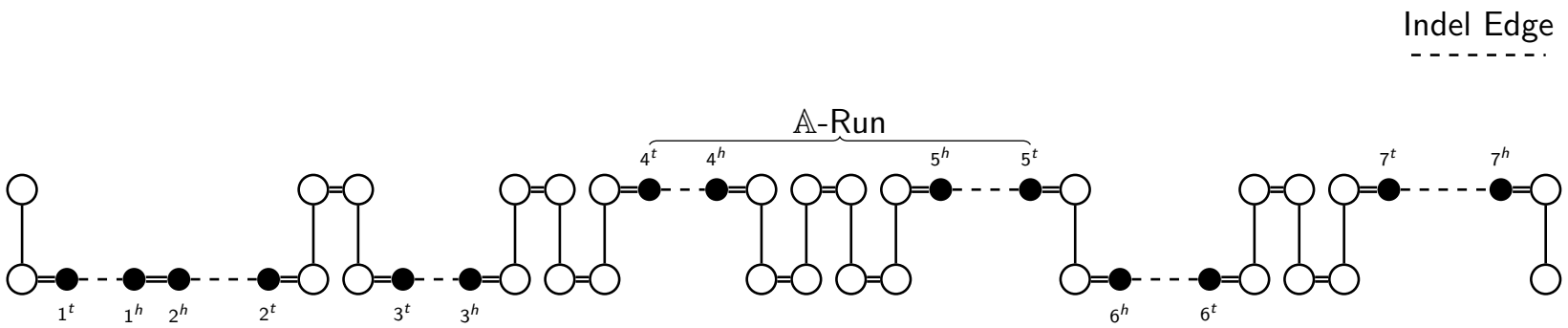
BWS-Conceptualization: Basics



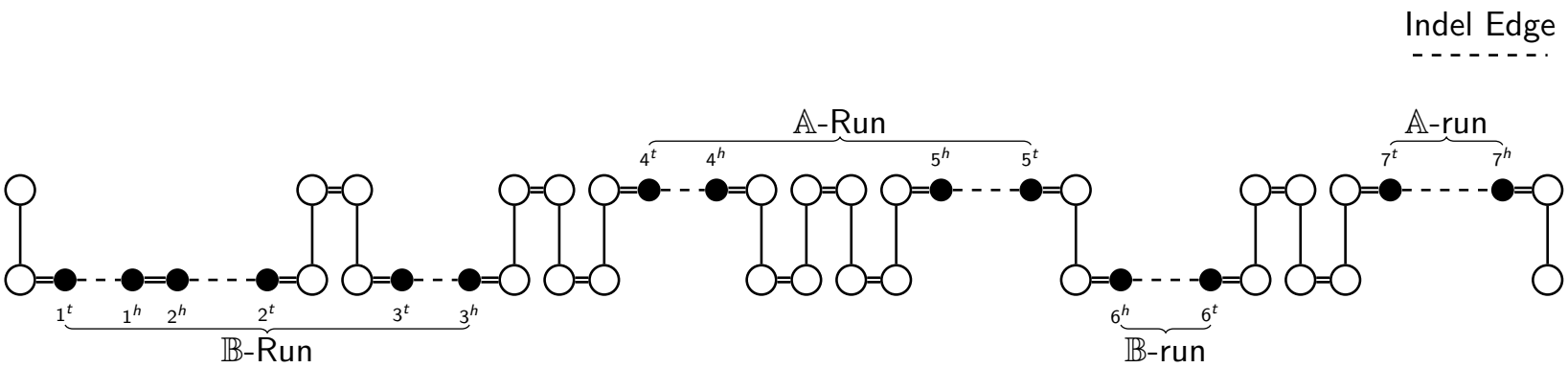
BWS-Conceptualization: Basics



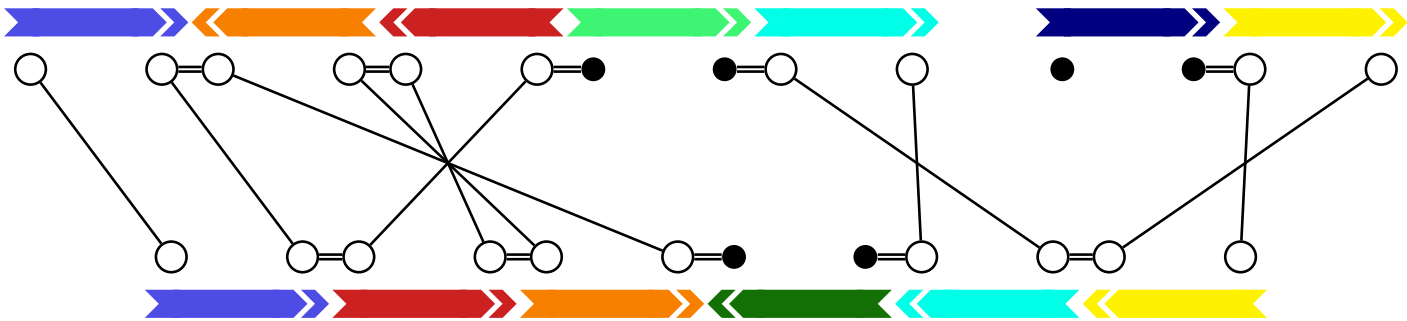
BWS-Conceptualization: Basics



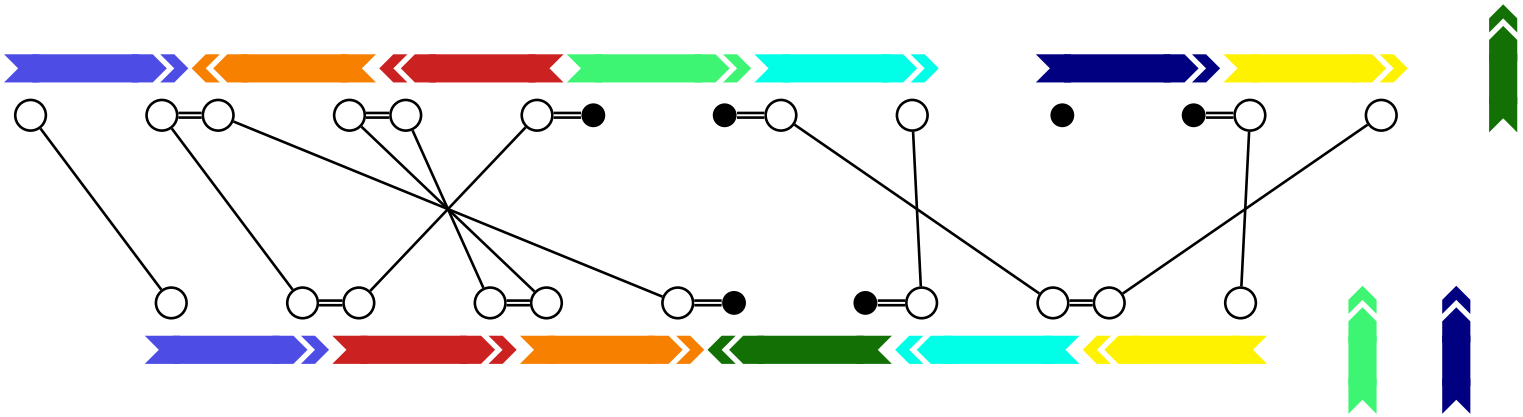
BWS-Conceptualization: Basics



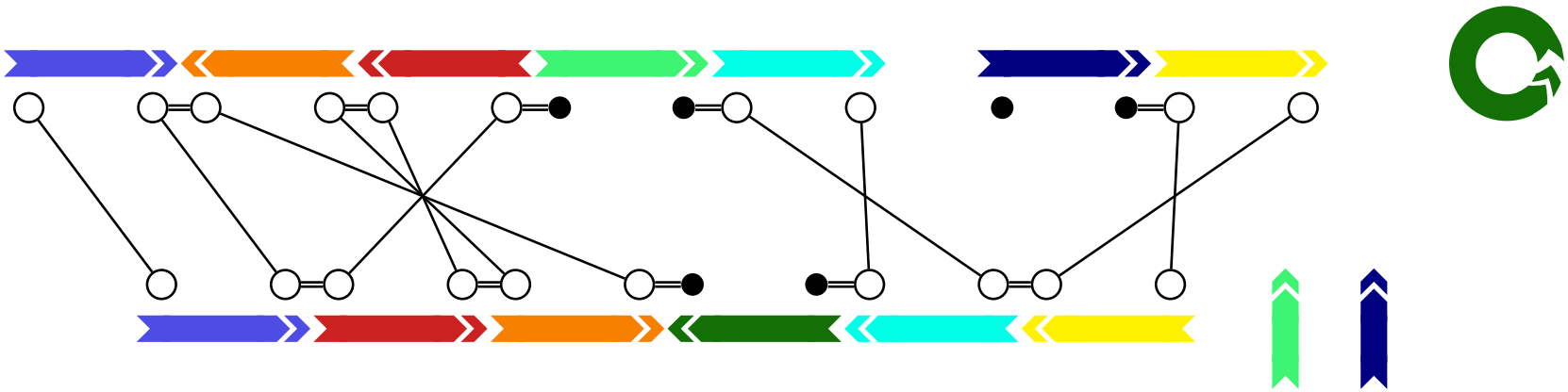
Compeau-Conceptualization: Basics



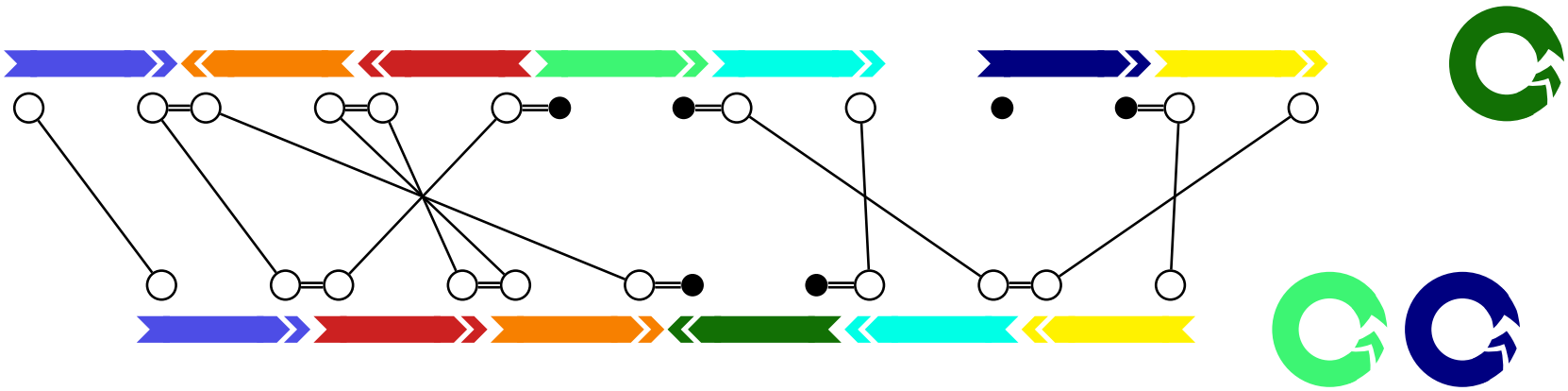
Compeau-Conceptualization: Basics



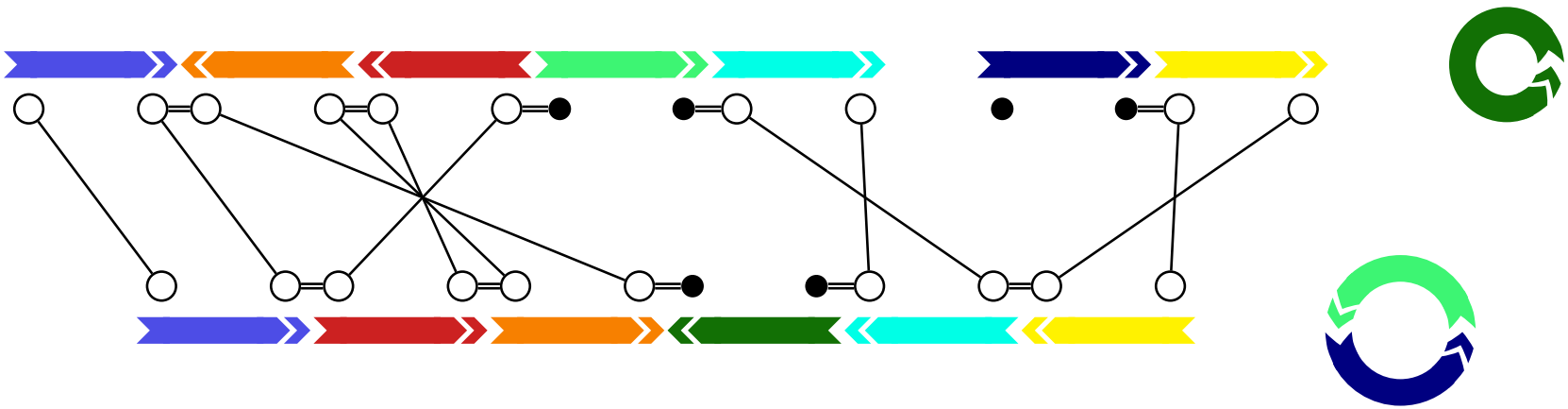
Compeau-Conceptualization: Basics



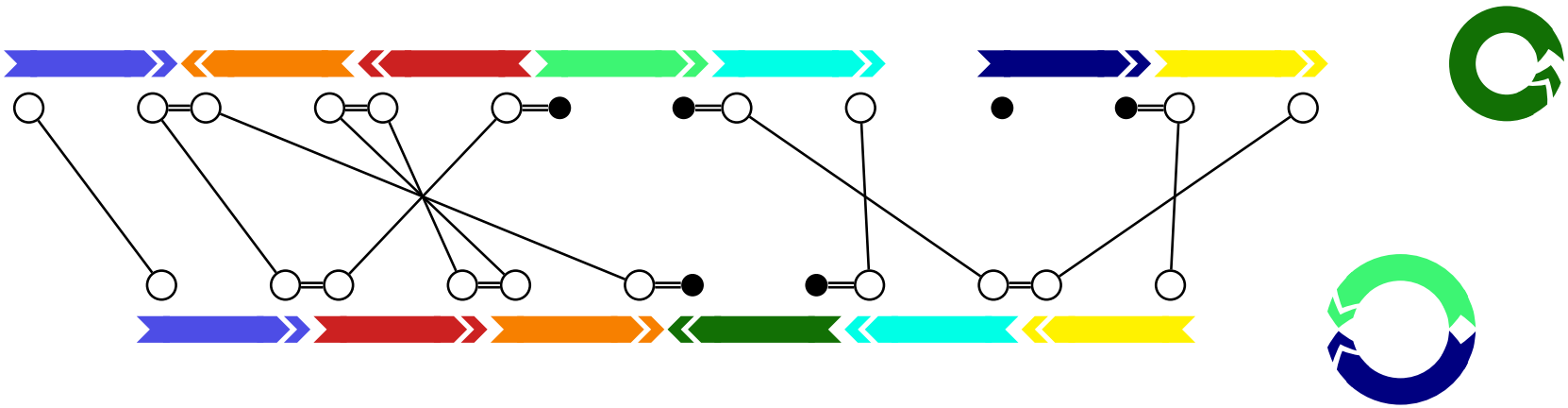
Compeau-Conceptualization: Basics



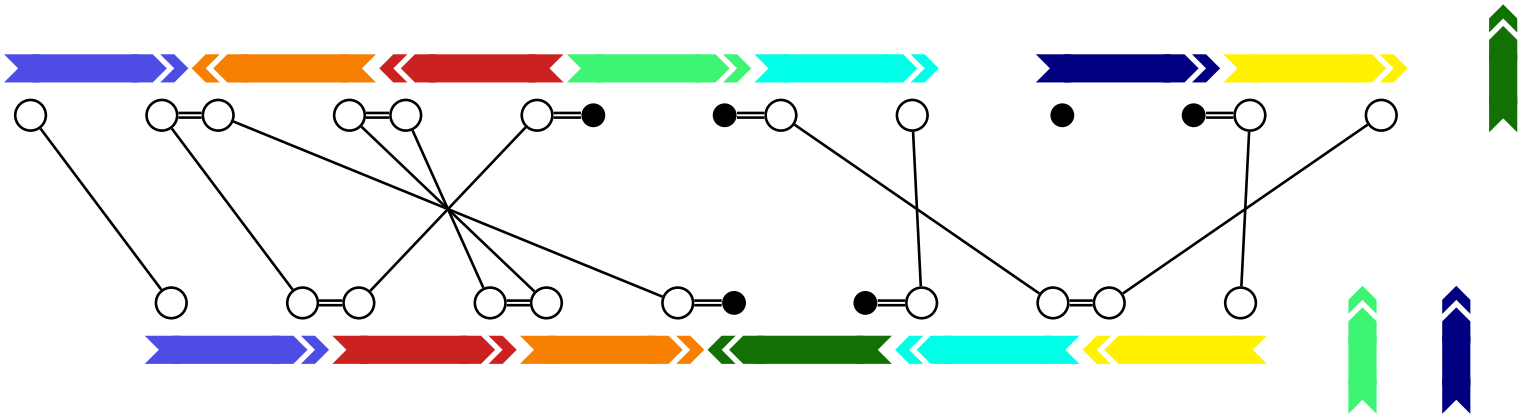
Compeau-Conceptualization: Basics



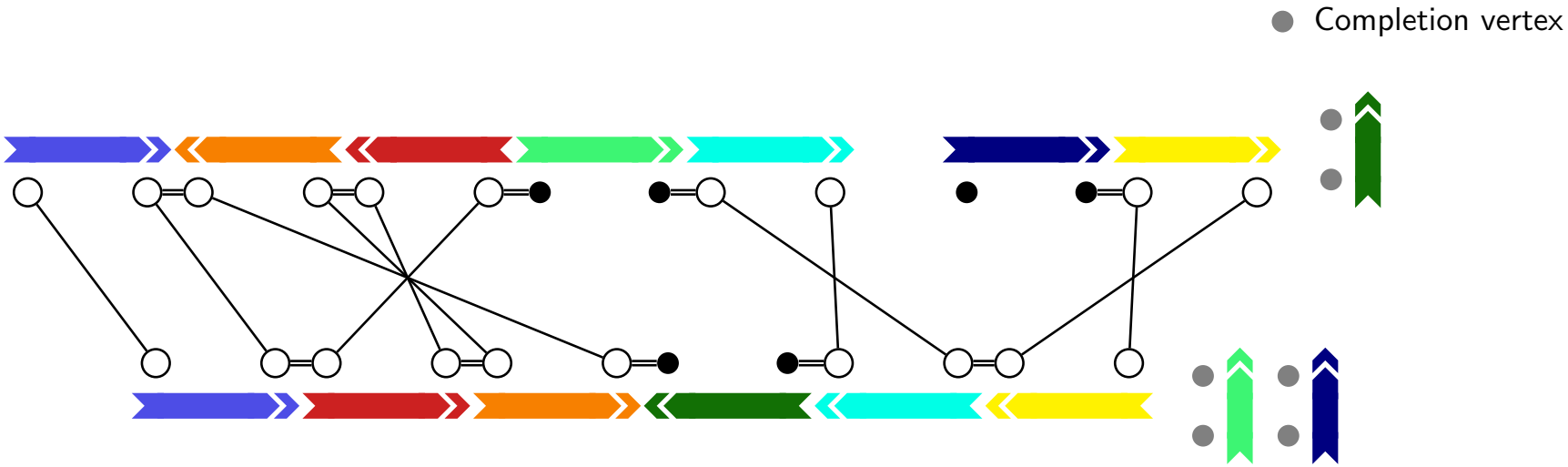
Compeau-Conceptualization: Basics



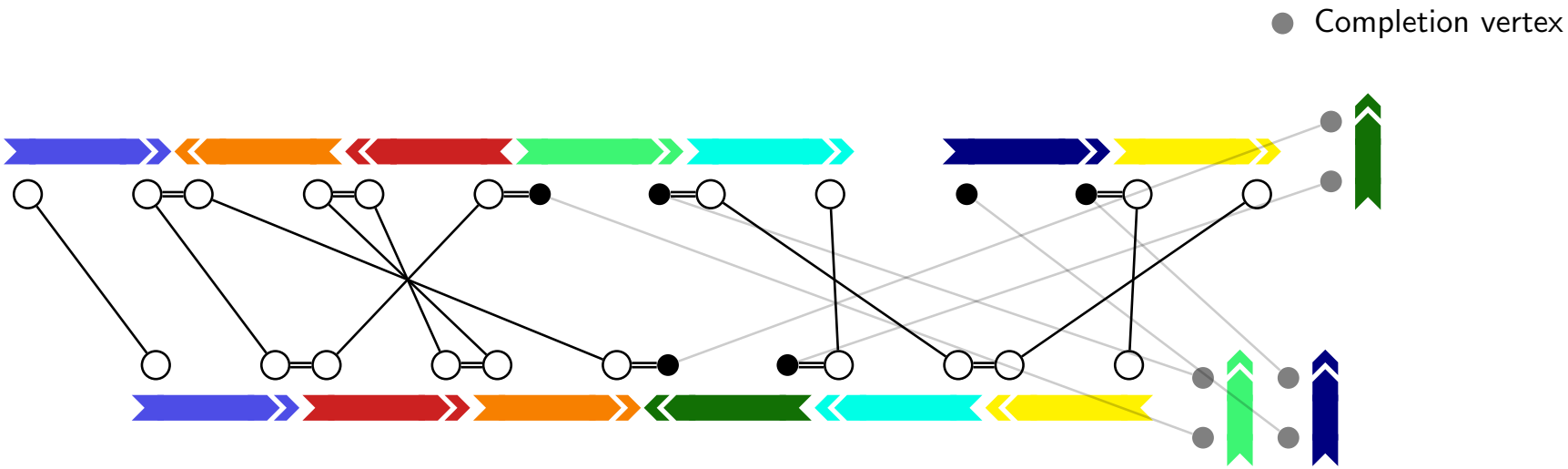
Compeau-Conceptualization: Basics



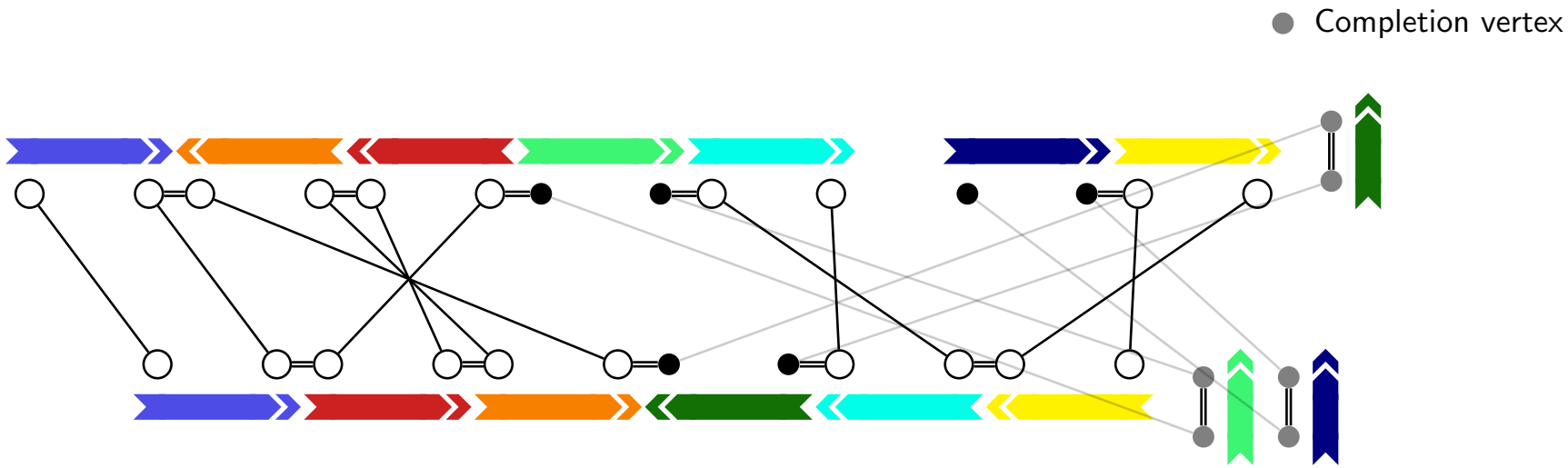
Compeau-Conceptualization: Basics



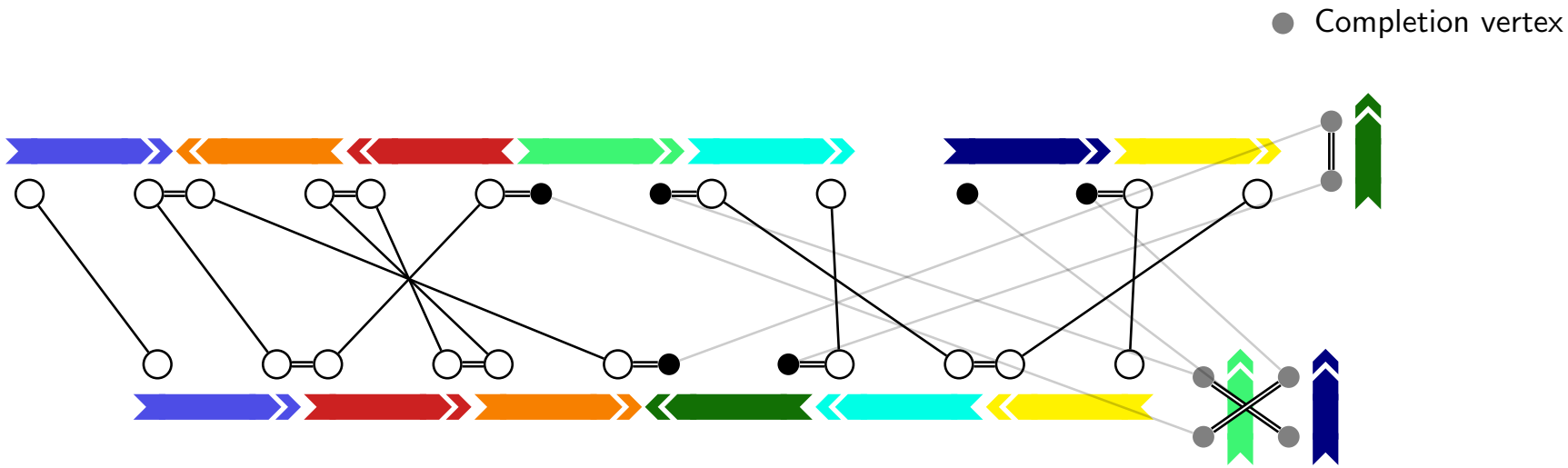
Compeau-Conceptualization: Basics



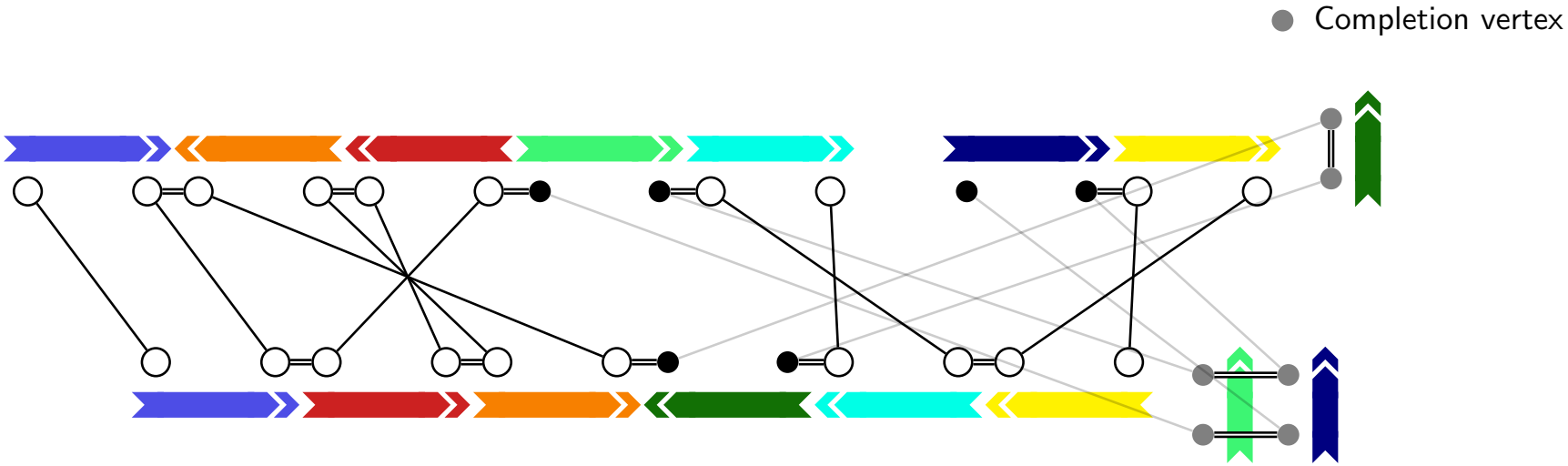
Compeau-Conceptualization: Basics



Compeau-Conceptualization: Basics

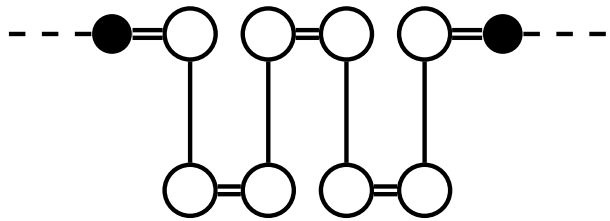


Compeau-Conceptualization: Basics



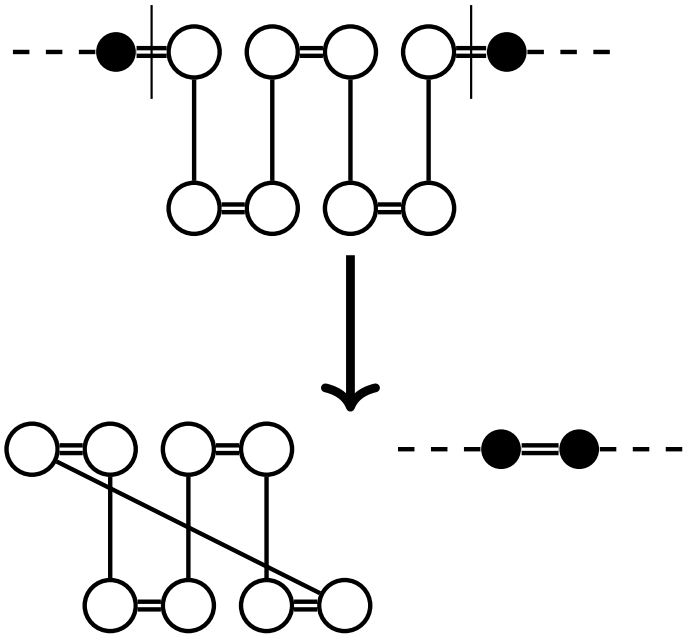
Bridging BWS- and Compeau-Conceptualizations

BWS: Runs can be accumulated.



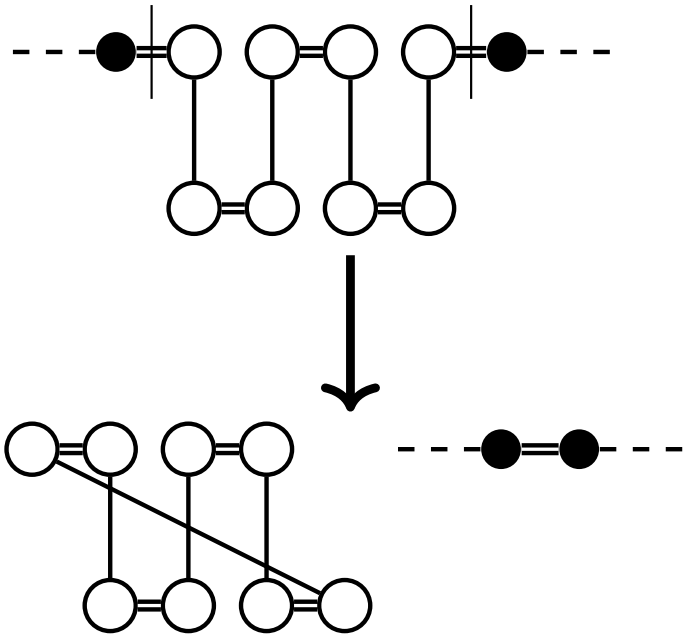
Bridging BWS- and Compeau-Conceptualizations

BWS: Runs can be accumulated.

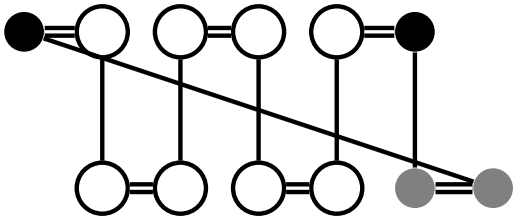


Bridging BWS- and Compeau-Conceptualizations

BWS: Runs can be accumulated.

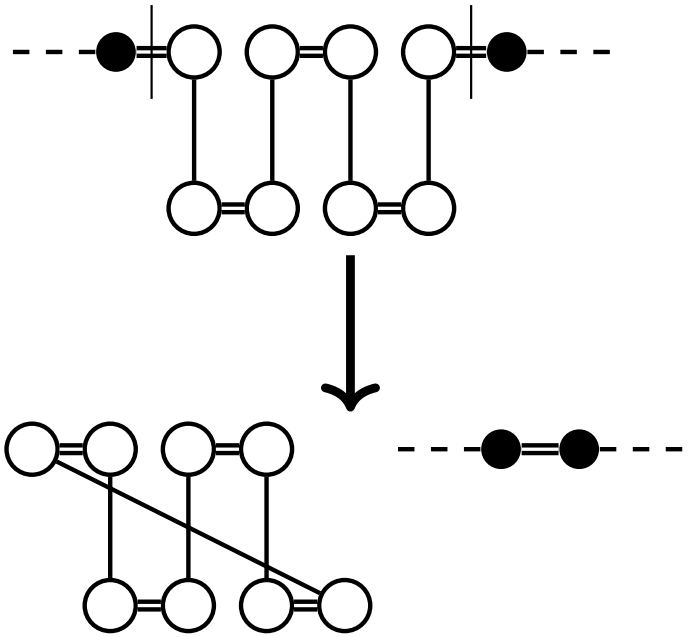


Compeau: Optimal 1-bracelet

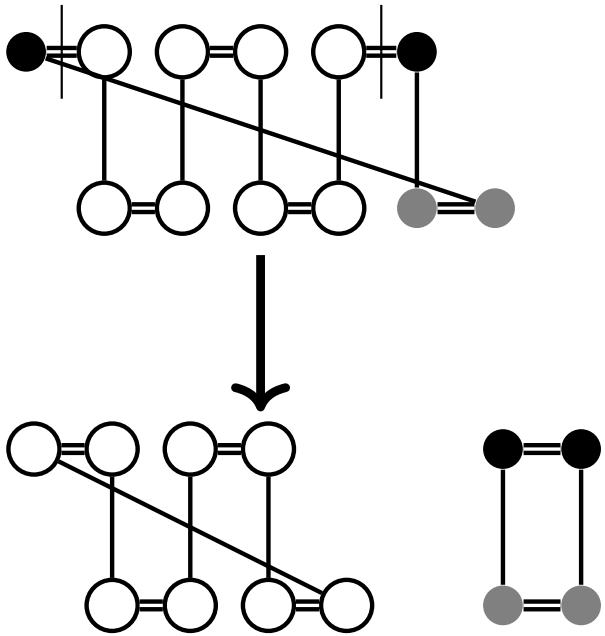


Bridging BWS- and Compeau-Conceptualizations

BWS: Runs can be accumulated.

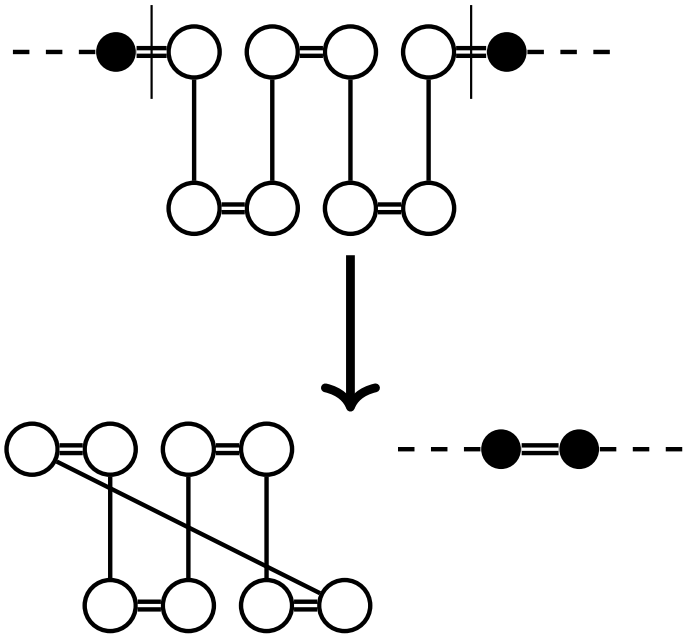


Compeau: Optimal 1-bracelet

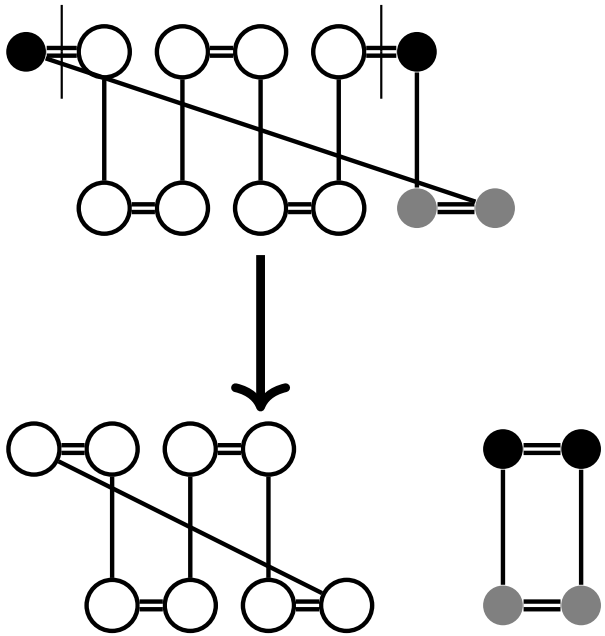


Bridging BWS- and Compeau-Conceptualizations

BWS: Runs can be accumulated.



Compeau: Optimal 1-bracelet



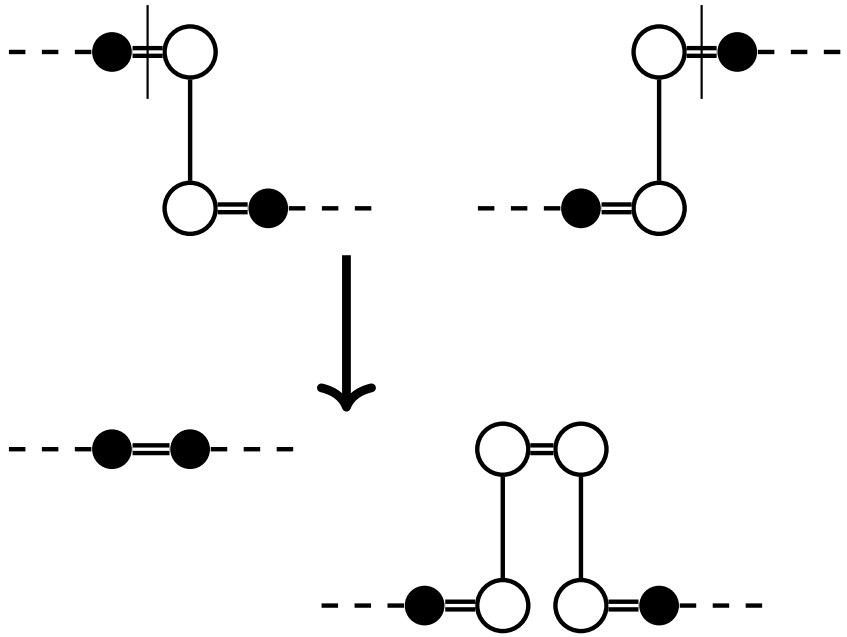
Common Ground: Safe operation!

Bridging BWS- and Compeau-Conceptualizations

BWS: Runs can be recombined.

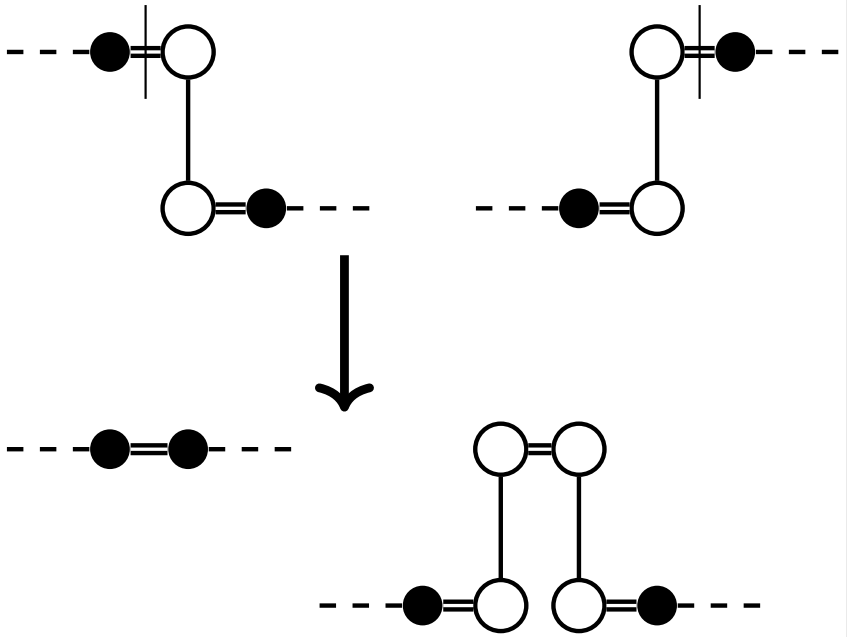
Bridging BWS- and Compeau-Conceptualizations

BWS: Runs can be recombined.

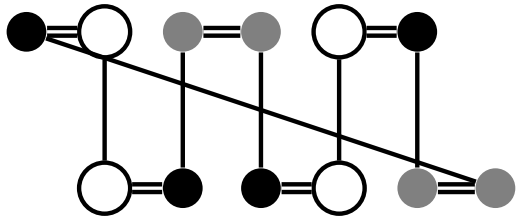


Bridging BWS- and Compeau-Conceptualizations

BWS: Runs can be recombined.

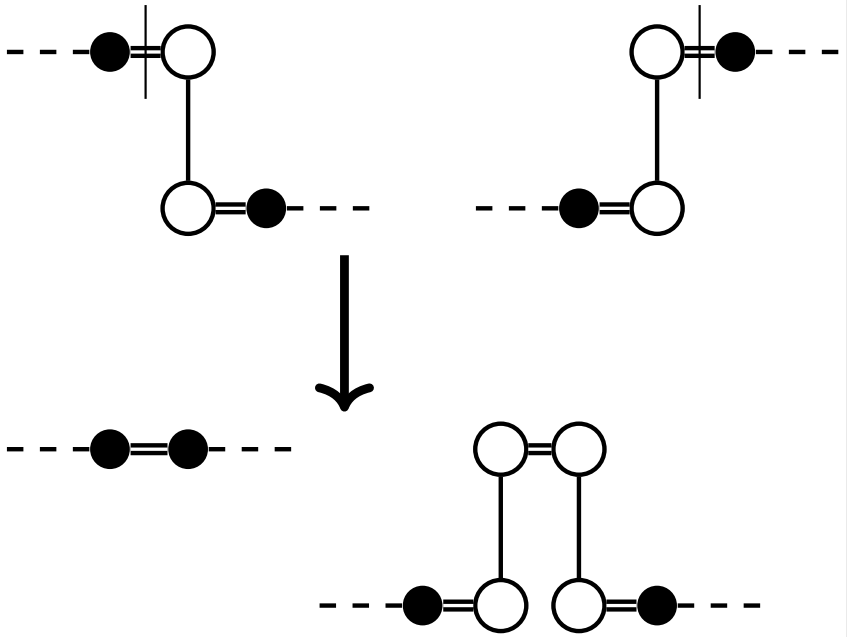


Compeau: Optimal 2-bracelet.

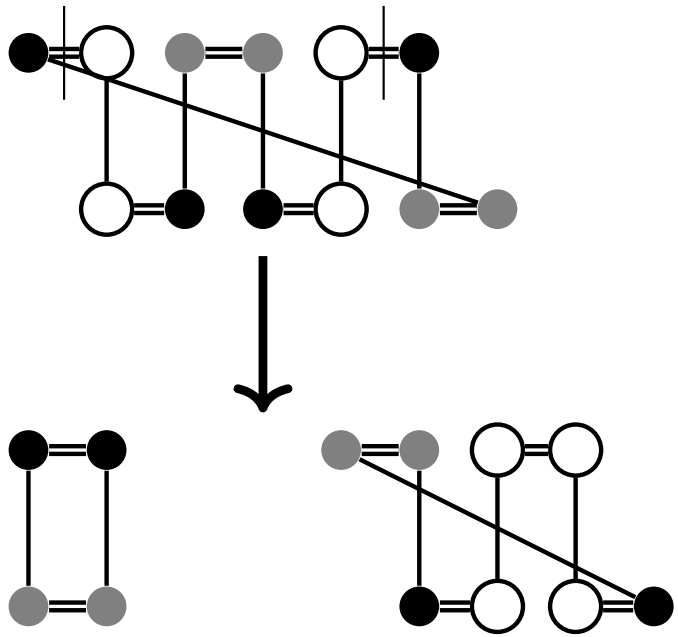


Bridging BWS- and Compeau-Conceptualizations

BWS: Runs can be recombined.

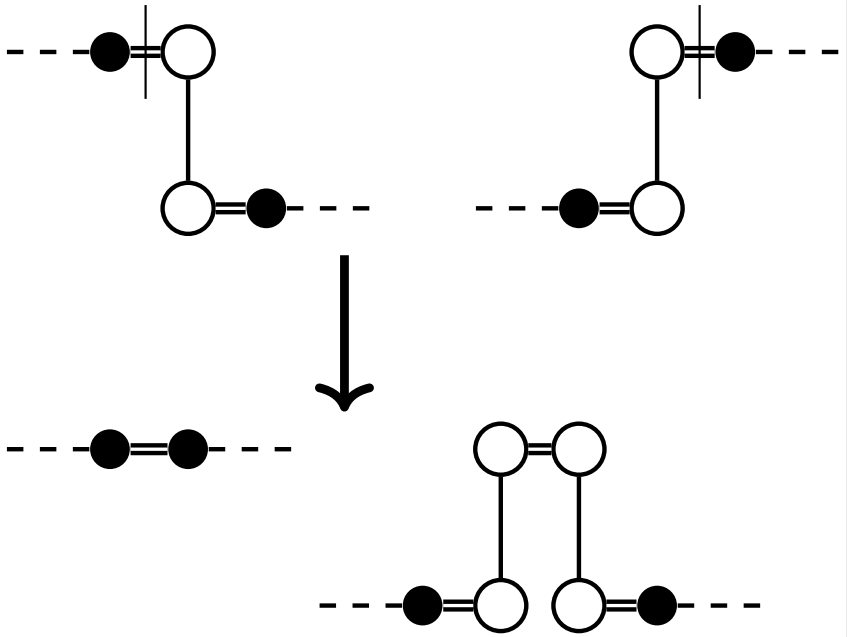


Compeau: Optimal 2-bracelet.

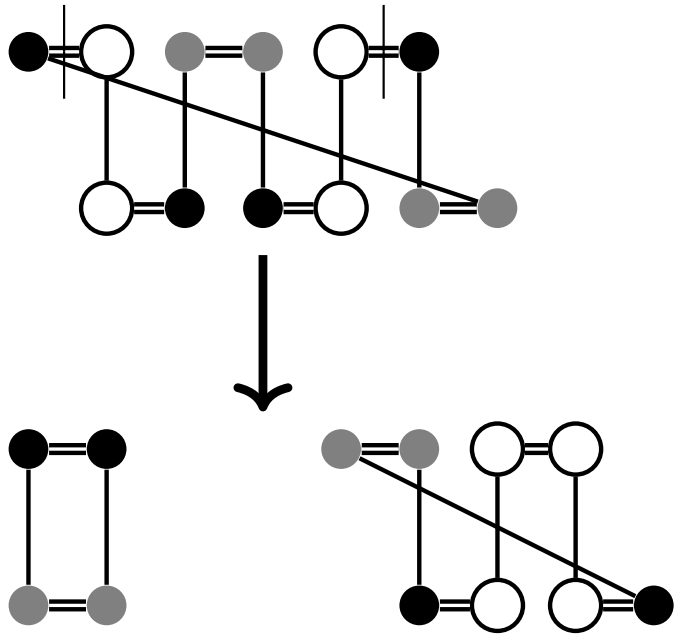


Bridging BWS- and Compeau-Conceptualizations

BWS: Runs can be recombined.



Compeau: Optimal 2-bracelet.

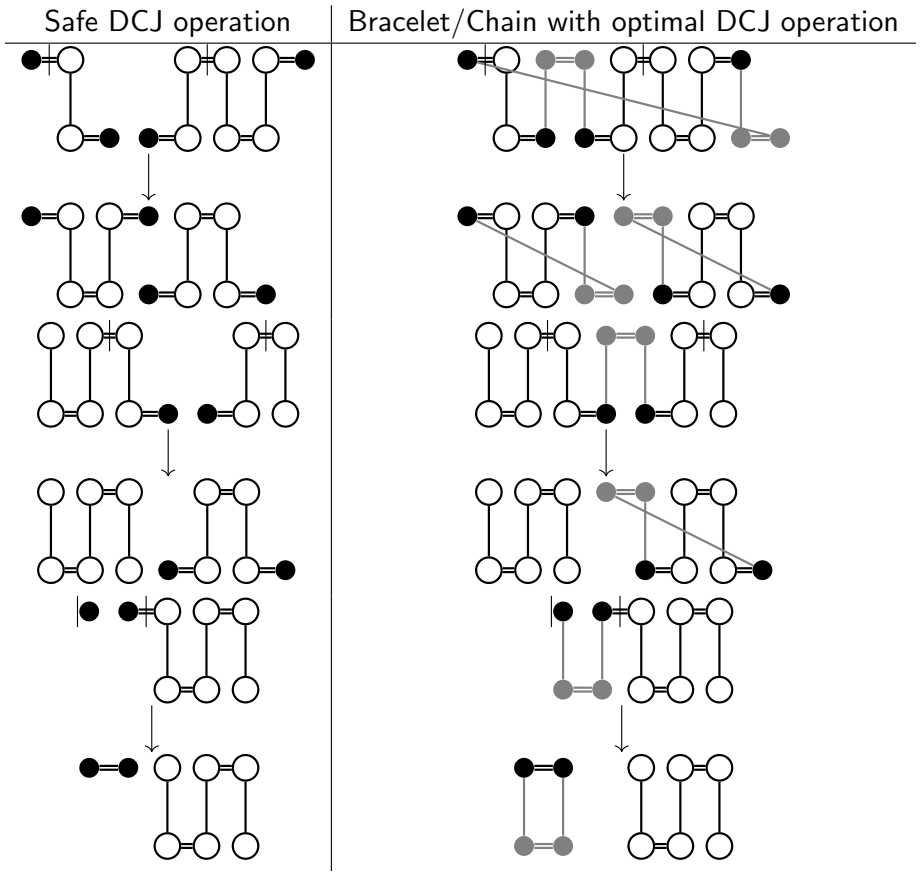


Common Ground: Safe operation!

Safe Operations and the BWS-Conceptualization

	Sources	Reusable Resultants	Unsat. after safe internal ops.					Unsat. after all safe ops. in group					No partner due to
			$A \circ a$	$B \mid a$	$a \mid b$	$A \mid b$	$B \circ b$	$A \circ a$	$B \mid a$	$p_{a \mid b}$	$A \mid b$	$B \circ b$	
P													
1	$AA_{\mathcal{AB}} + BB_{\mathcal{AB}}$		1	1	2	1	1	0	0	0	0	0	
Q													
1	$2AA_{\mathcal{AB}} + BB_{\mathcal{A}} + BB_{\mathcal{B}}$		2	2	2	2	2	0	0	0	0	0	
2	$2BB_{\mathcal{AB}} + AA_{\mathcal{A}} + AA_{\mathcal{B}}$		2	2	2	2	2	0	0	0	0	0	
T													
1	$AA_{\mathcal{AB}} + BB_{\mathcal{A}} + AB_{\mathcal{AB}}$		2	2	2	1	1	0	0	0	0	0	
2	$AA_{\mathcal{AB}} + BB_{\mathcal{B}} + AB_{\mathcal{BA}}$		1	1	2	2	2	0	0	0	0	0	
3	$BB_{\mathcal{AB}} + AA_{\mathcal{A}} + AB_{\mathcal{BA}}$		2	2	2	1	1	0	0	0	0	0	
4	$BB_{\mathcal{AB}} + AA_{\mathcal{B}} + AB_{\mathcal{AB}}$		1	1	2	2	2	0	0	0	0	0	
5	$2AA_{\mathcal{AB}} + BB_{\mathcal{A}}$	$AA_{\mathcal{B}}$	2	2	2	2	0	0	0	0	2	0	$P_{A \mid b} \to T1, Q1, P1;$
6	$2AA_{\mathcal{AB}} + BB_{\mathcal{B}}$	$AA_{\mathcal{A}}$	2	0	2	2	2	2	0	0	0	0	$P_{A \circ a} \to T2, Q1, P1;$
7	$2BB_{\mathcal{AB}} + AA_{\mathcal{A}}$	$BB_{\mathcal{B}}$	2	2	2	0	2	0	0	0	0	2	$P_{B \circ b} \to Q2, P1, T3;$
8	$2BB_{\mathcal{AB}} + AA_{\mathcal{B}}$	$BB_{\mathcal{A}}$	0	2	2	2	2	0	2	0	0	0	$P_{B \mid a} \to Q2, P1, T4;$
S													
1	$AA_{\mathcal{A}} + BB_{\mathcal{A}}$		2	2	0	0	0	0	0	0	0	0	
2	$AA_{\mathcal{B}} + BB_{\mathcal{B}}$		0	0	0	2	2	0	0	0	0	0	
3	$AB_{\mathcal{AB}} + AB_{\mathcal{BA}}$		1	1	2	1	1	0	0	0	0	0	
4	$AA_{\mathcal{AB}} + BB_{\mathcal{A}}$	$AB_{\mathcal{BA}}$	1	2	1	1	0	0	1	1	1	0	$P_{B \mid a} \to S1, T5, T1;$ $\star P_{a \mid b} \to T5, T1, P1;$

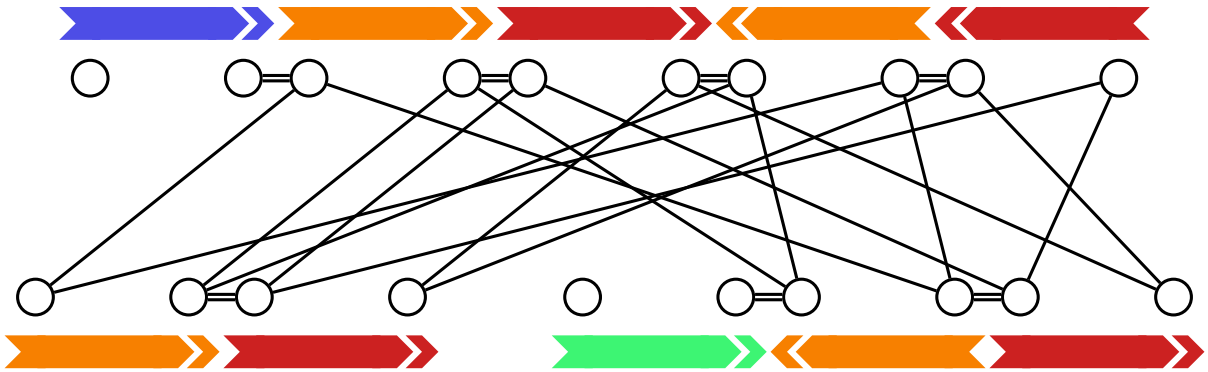
Safe Operations and the Compeau-Conceptualization



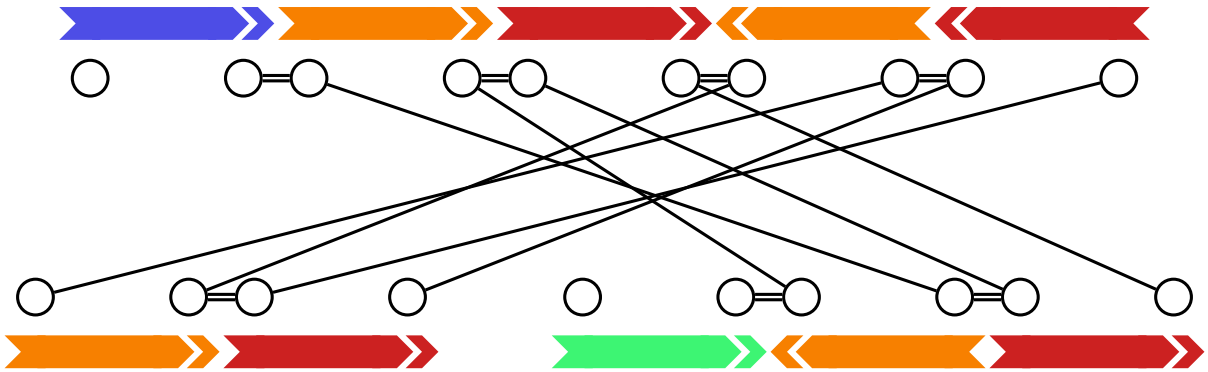
What's the use?

(Besides didactics)

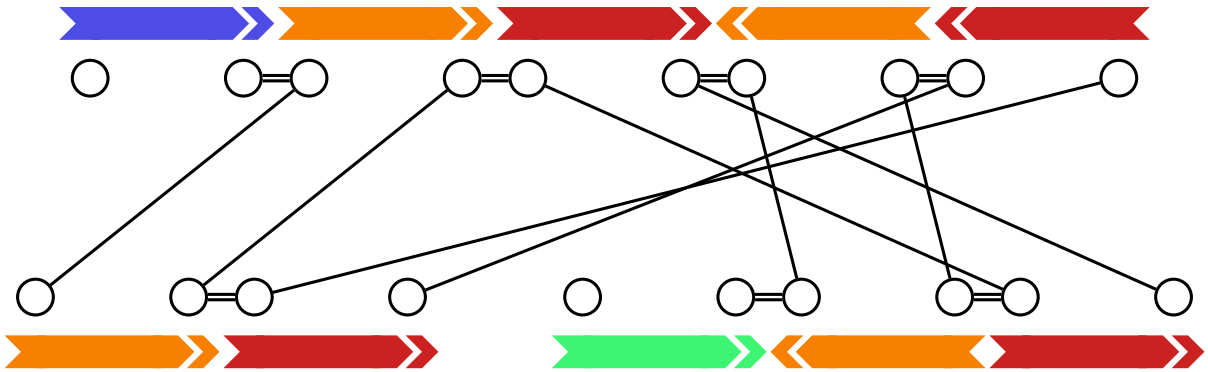
Duplicate Markers make the Problem NP-hard



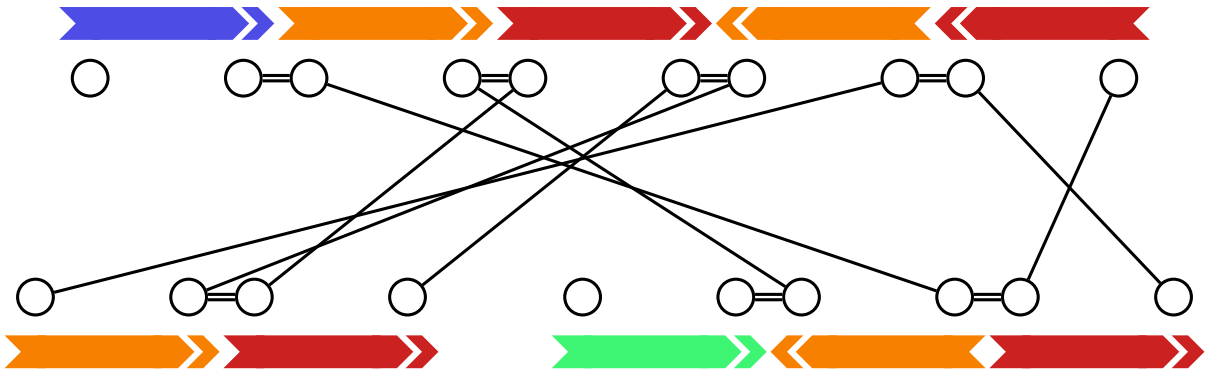
Duplicate Markers make the Problem NP-hard



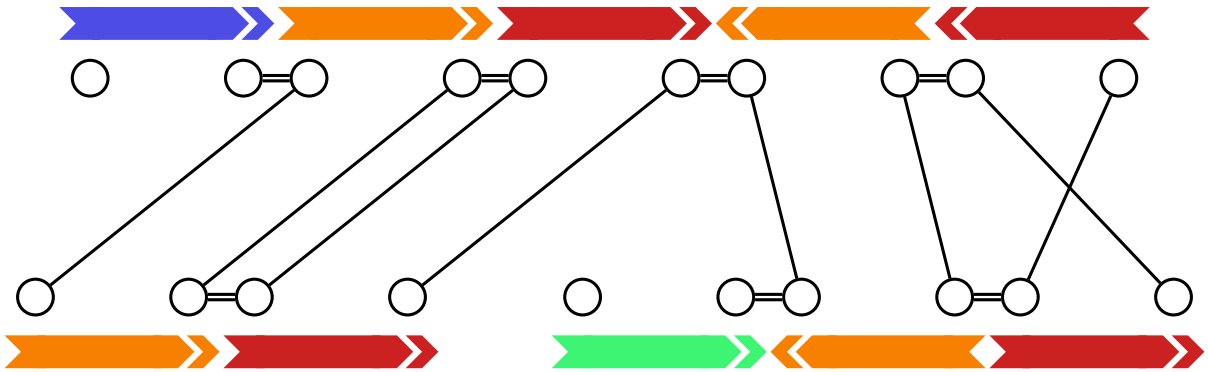
Duplicate Markers make the Problem NP-hard



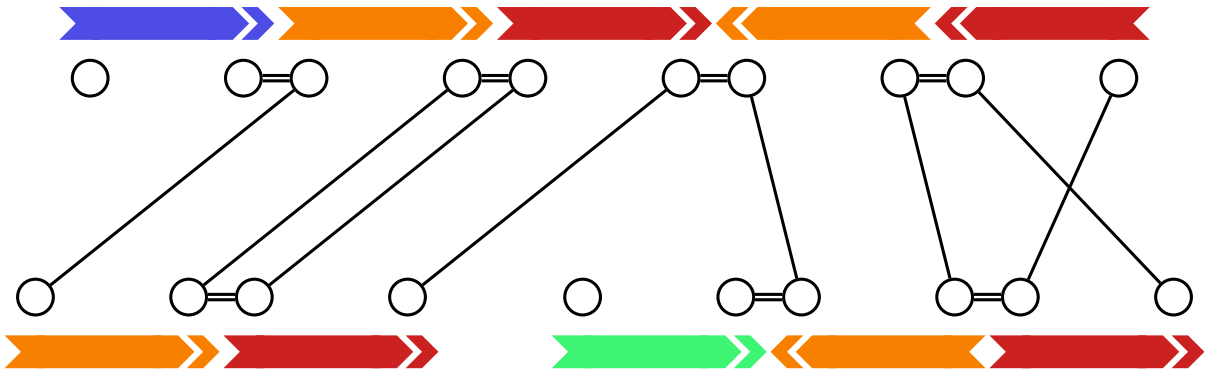
Duplicate Markers make the Problem NP-hard



Duplicate Markers make the Problem NP-hard



Duplicate Markers make the Problem NP-hard

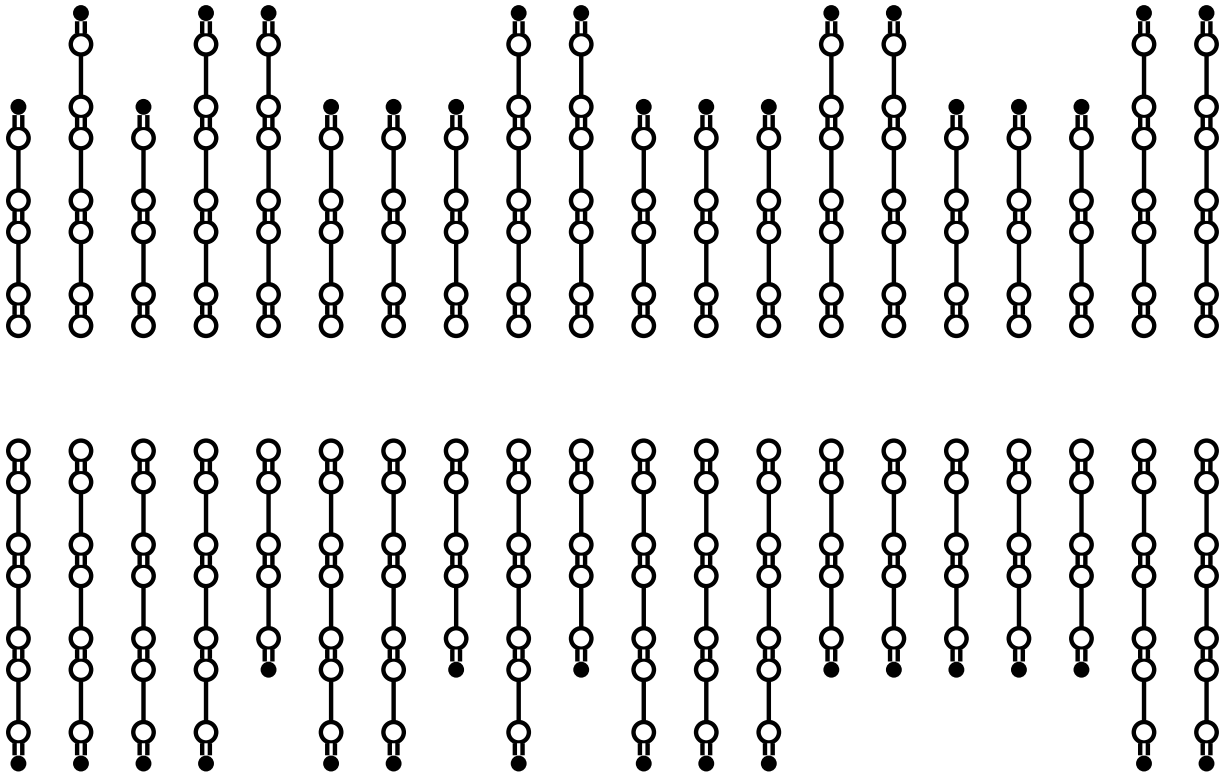


→ Solved by `ding`, ILP based on BWS-model (BBDS,2023).

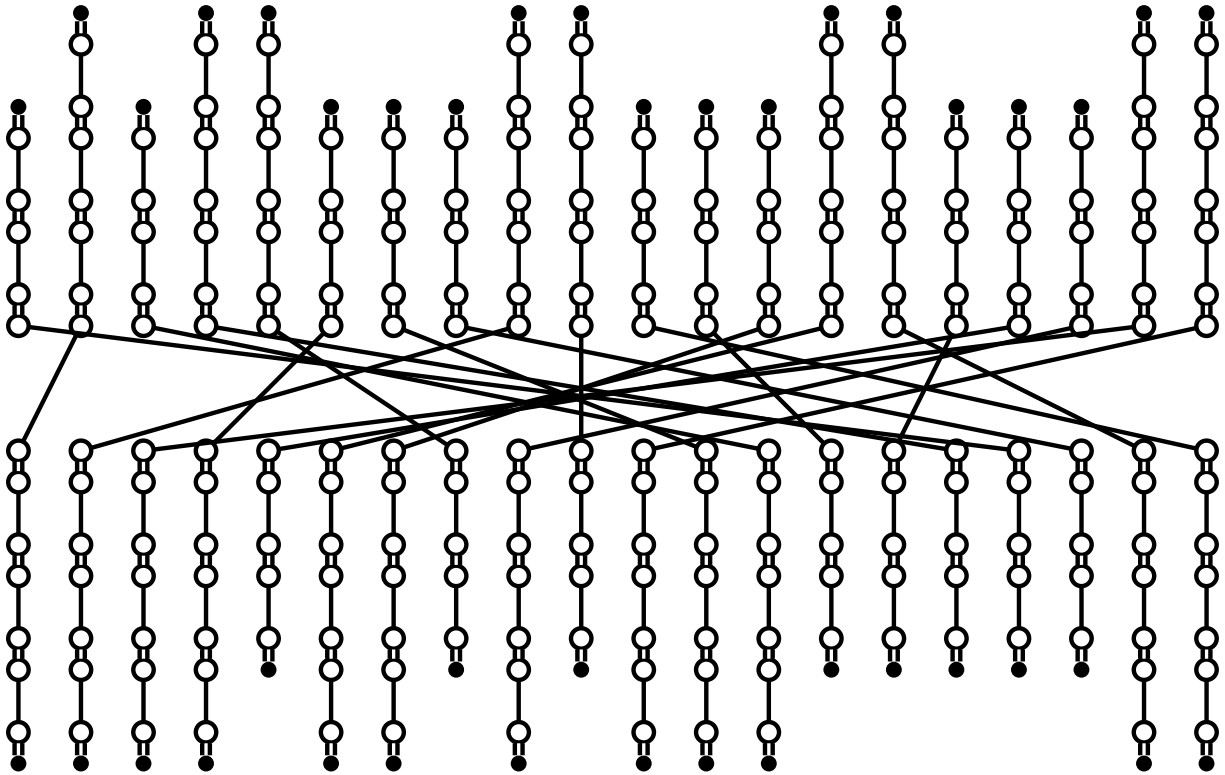
Path-Recombinations in BWS-model are too Complicated for an ILP.

	Sources	Reusable Resultants	Unsat. after safe internal ops.					Unsat. after all safe ops. in group					No partner due to
			$A \circ a$	$B \vert a$	$a \vert b$	$A \vert b$	$B \circ b$	$A \circ a$	$B \vert a$	$p_{a \vert b}$	$A \vert b$	$B \circ b$	
P													
1	$AA_{\mathcal{AB}} + BB_{\mathcal{AB}}$		1	1	2	1	1	0	0	0	0	0	
Q													
1	$2AA_{\mathcal{AB}} + BB_{\mathcal{A}} + BB_{\mathcal{B}}$		2	2	2	2	2	0	0	0	0	0	
2	$2BB_{\mathcal{AB}} + AA_{\mathcal{A}} + AA_{\mathcal{B}}$		2	2	2	2	2	0	0	0	0	0	
T													
1	$AA_{\mathcal{AB}} + BB_{\mathcal{A}} + AB_{\mathcal{AB}}$		2	2	2	1	1	0	0	0	0	0	
2	$AA_{\mathcal{AB}} + BB_{\mathcal{B}} + AB_{\mathcal{BA}}$		1	1	2	2	2	0	0	0	0	0	
3	$BB_{\mathcal{AB}} + AA_{\mathcal{A}} + AB_{\mathcal{BA}}$		2	2	2	1	1	0	0	0	0	0	
4	$BB_{\mathcal{AB}} + AA_{\mathcal{B}} + AB_{\mathcal{AB}}$		1	1	2	2	2	0	0	0	0	0	
5	$2AA_{\mathcal{AB}} + BB_{\mathcal{A}}$	$AA_{\mathcal{B}}$	2	2	2	2	0	0	0	0	2	0	$P_{A\vert b} \to T1, Q1, P1;$
6	$2AA_{\mathcal{AB}} + BB_{\mathcal{B}}$	$AA_{\mathcal{A}}$	2	0	2	2	2	2	0	0	0	0	$P_{A\circ a} \to T2, Q1, P1;$
7	$2BB_{\mathcal{AB}} + AA_{\mathcal{A}}$	$BB_{\mathcal{B}}$	2	2	2	0	2	0	0	0	0	2	$P_{B\circ b} \to Q2, P1, T3;$
8	$2BB_{\mathcal{AB}} + AA_{\mathcal{B}}$	$BB_{\mathcal{A}}$	0	2	2	2	2	0	2	0	0	0	$P_{B\vert a} \to Q2, P1, T4;$
S													
1	$AA_{\mathcal{A}} + BB_{\mathcal{A}}$		2	2	0	0	0	0	0	0	0	0	
2	$AA_{\mathcal{B}} + BB_{\mathcal{B}}$		0	0	0	2	2	0	0	0	0	0	
3	$AB_{\mathcal{AB}} + AB_{\mathcal{BA}}$		1	1	2	1	1	0	0	0	0	0	
4	$AA_{\mathcal{AB}} + BB_{\mathcal{A}}$	$AB_{\mathcal{BA}}$	1	2	1	1	0	0	1	1	1	0	$P_{B\vert a} \to S1, T5, T1;$ $\star P_{a\vert b} \to T5, T1, P1;$

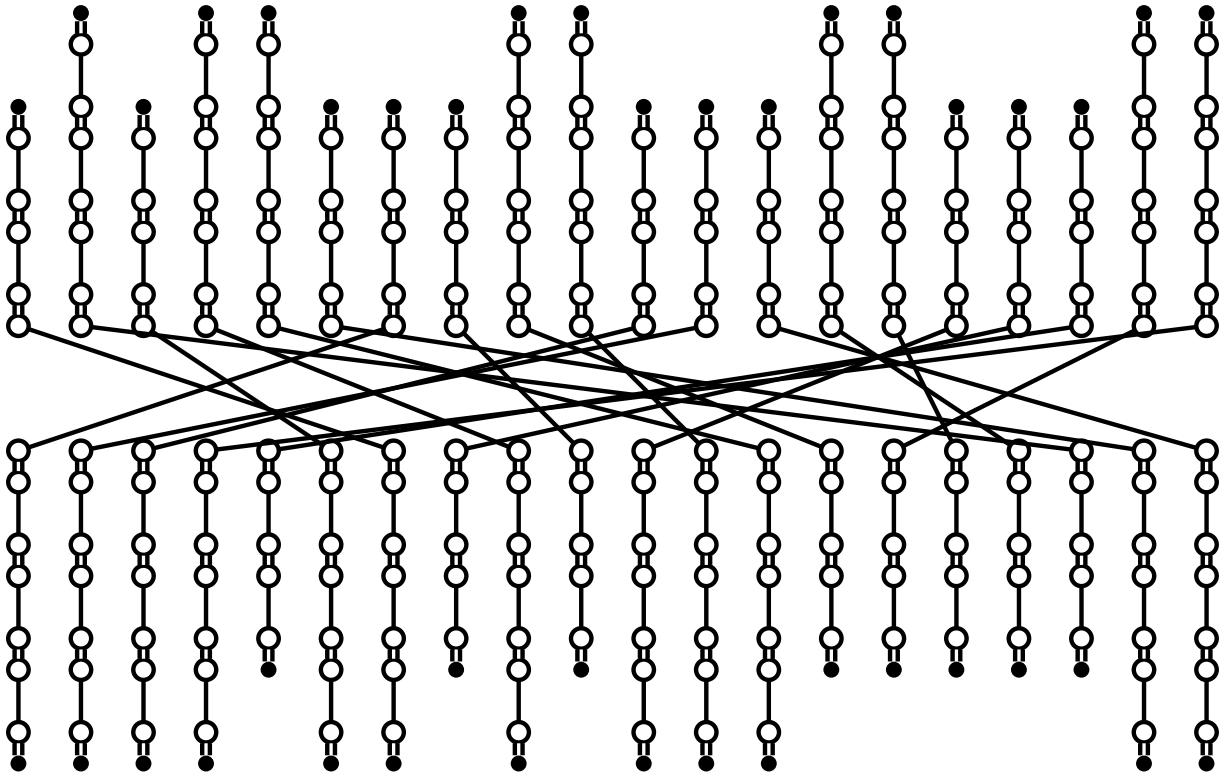
Transforming Paths into Cycles - Capping



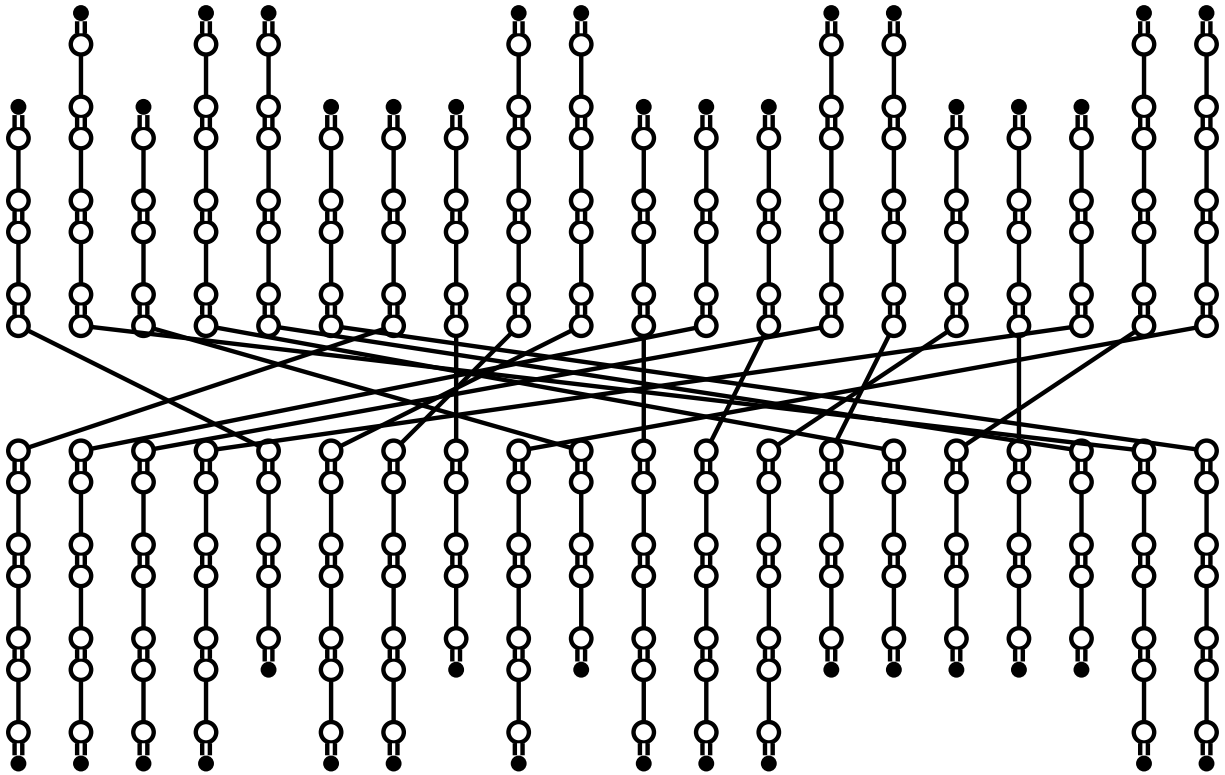
Transforming Paths into Cycles - Capping



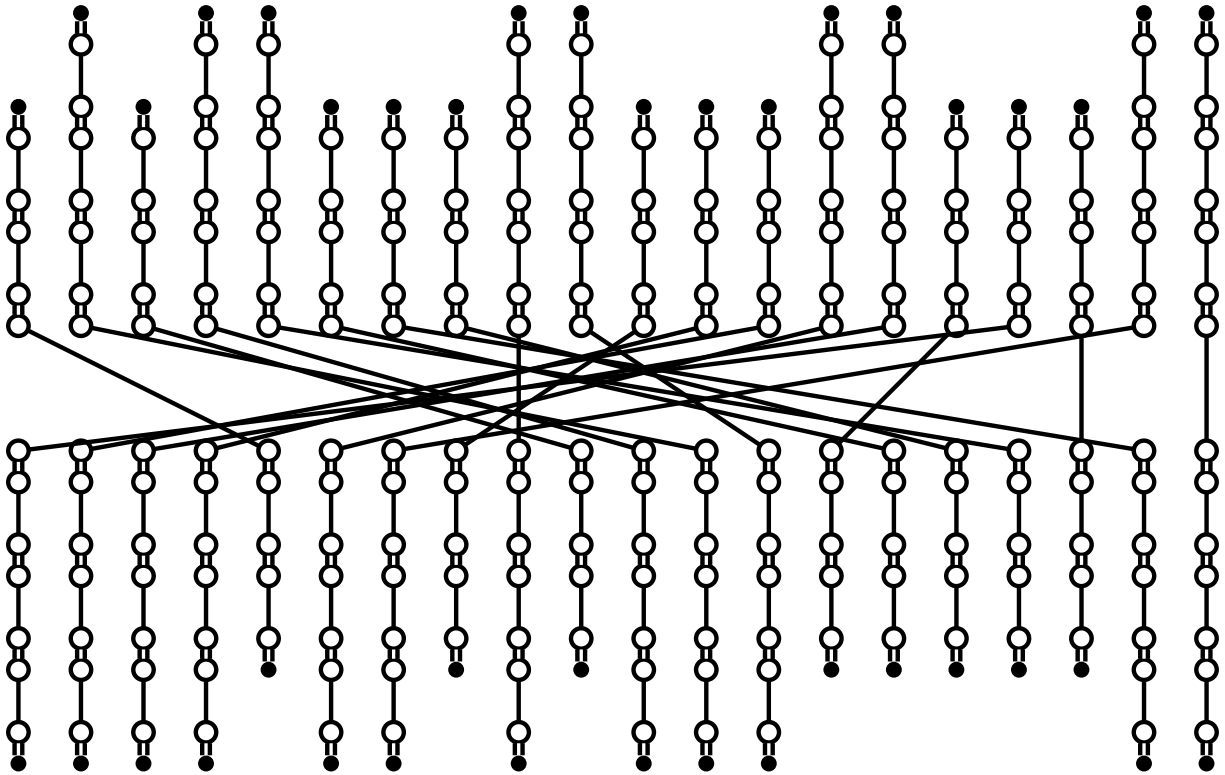
Transforming Paths into Cycles - Capping



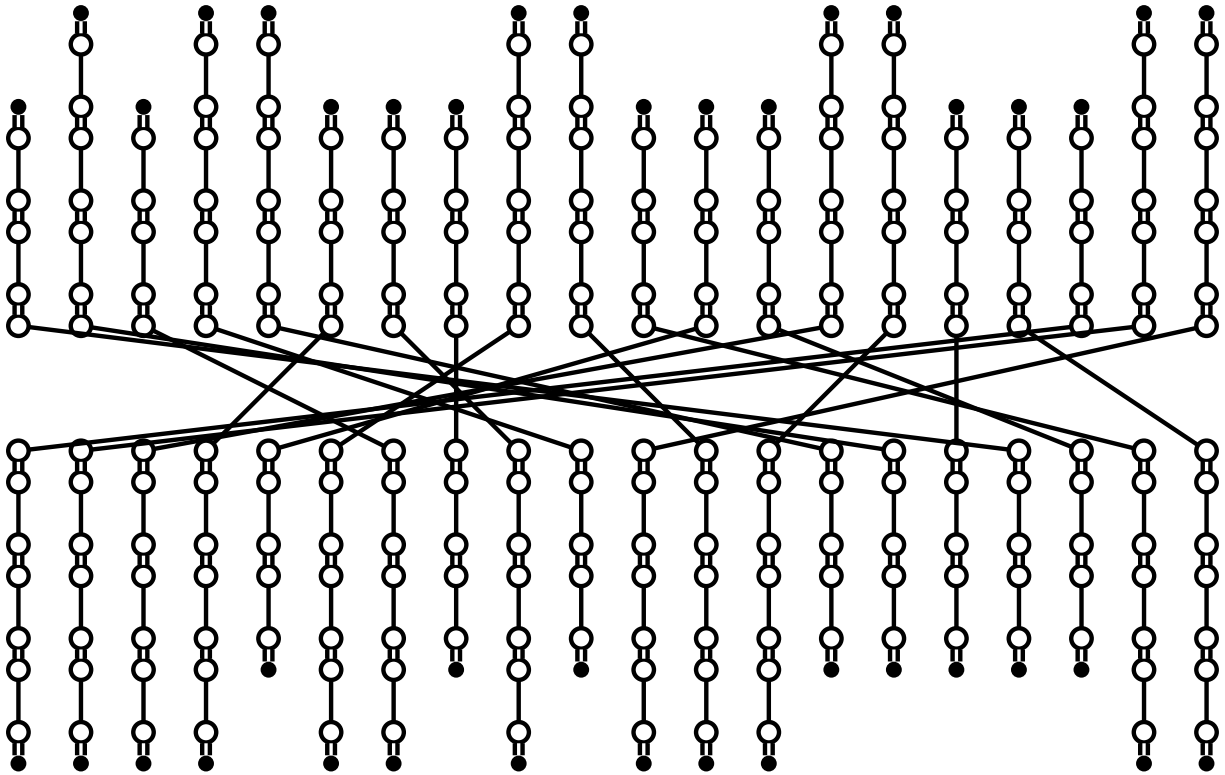
Transforming Paths into Cycles - Capping



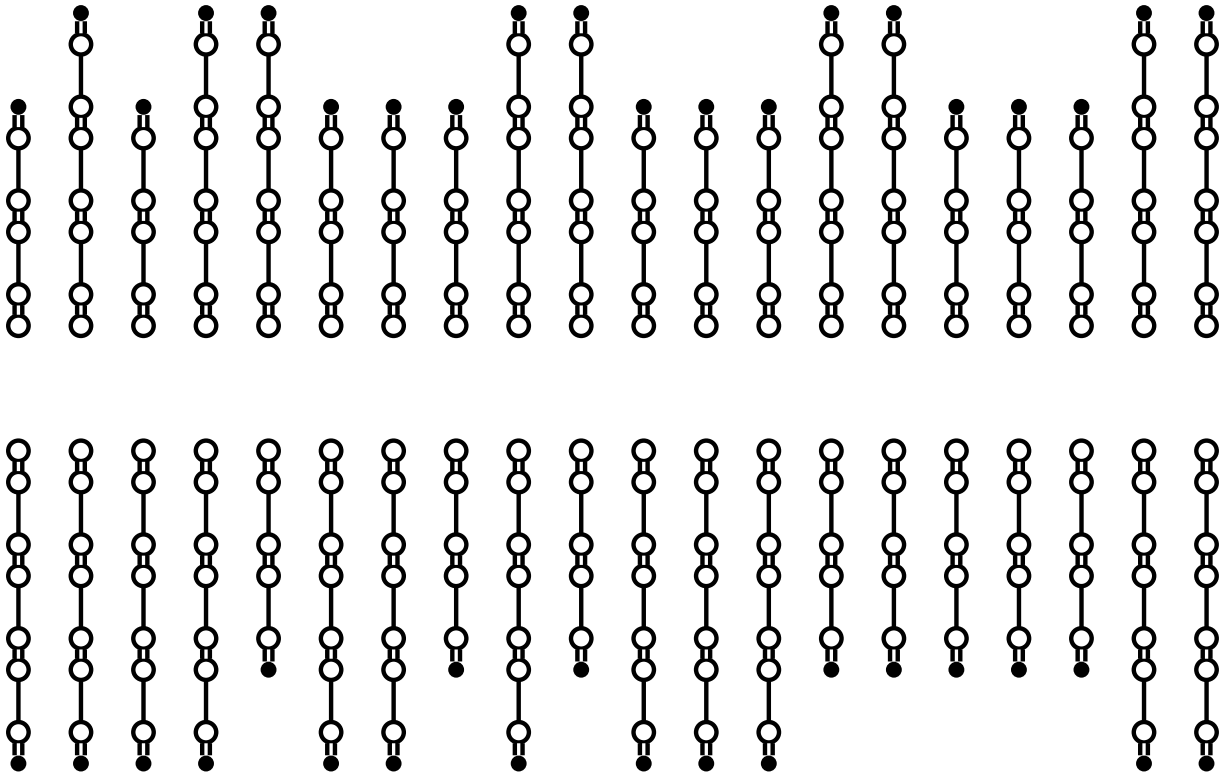
Transforming Paths into Cycles - Capping



Transforming Paths into Cycles - Capping



Transforming Paths into Cycles - Capping

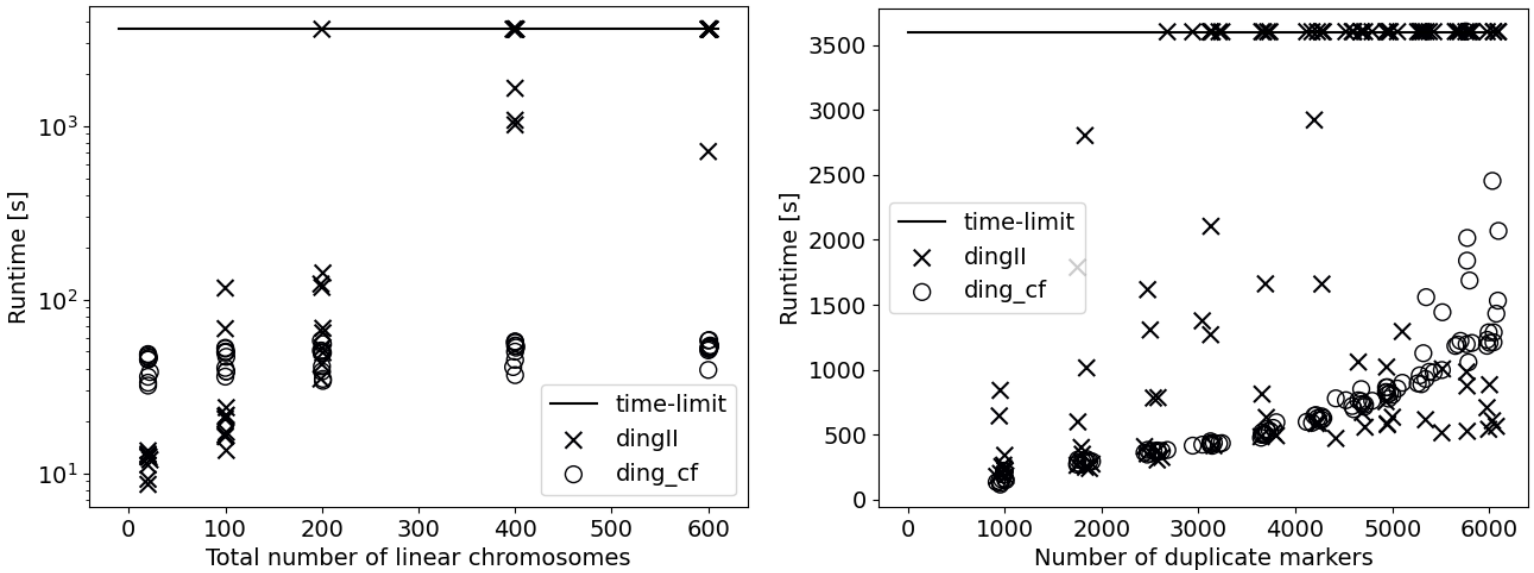


- Superexponential increase of Solution Space. (Rubert & Braga, 2022)

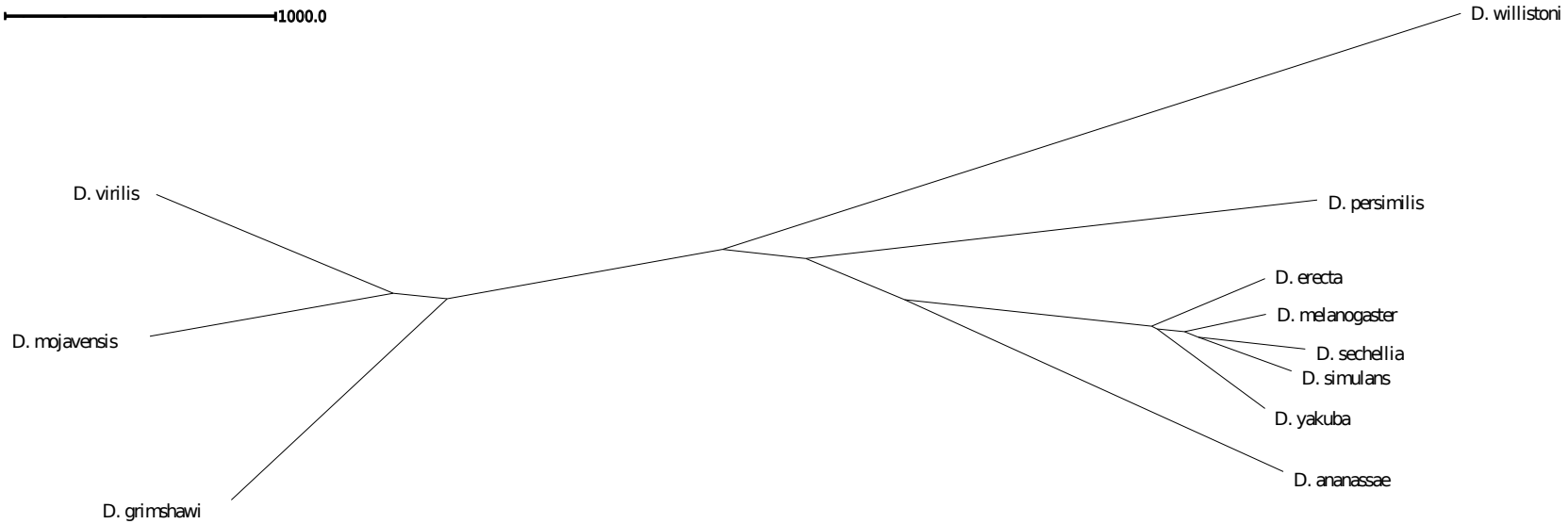
ILP Based on New Formula Avoids Capping!

$$d_{DCJ}^{id}(\mathbb{A},\mathbb{B}) = n - c_\circ + \left\lceil \frac{p_{a|b} + \max(p_{A\circ a}, p_{B|a}) + \max(p_{A|b}, p_{B\circ b}) - p_{A|B}}{2} \right\rceil$$

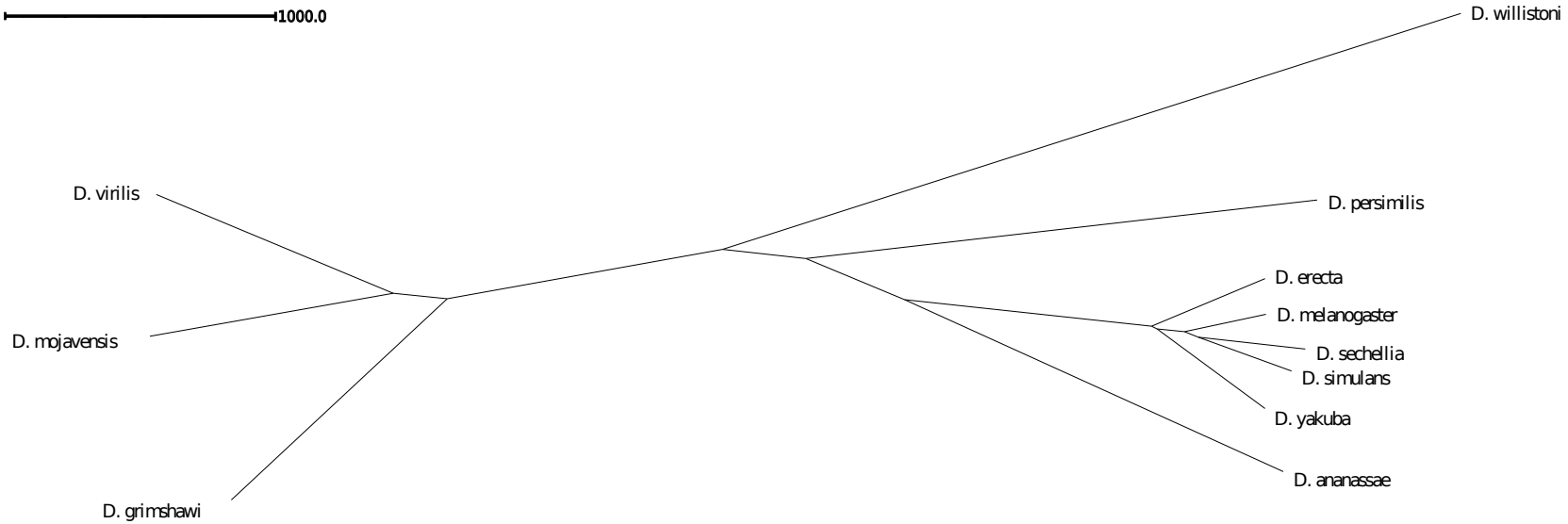
Considerable Performance Advantage when the Number of Linear Chromosomes is High



Computing Rearrangement Phylogenies on Contig-level resolved Genomes

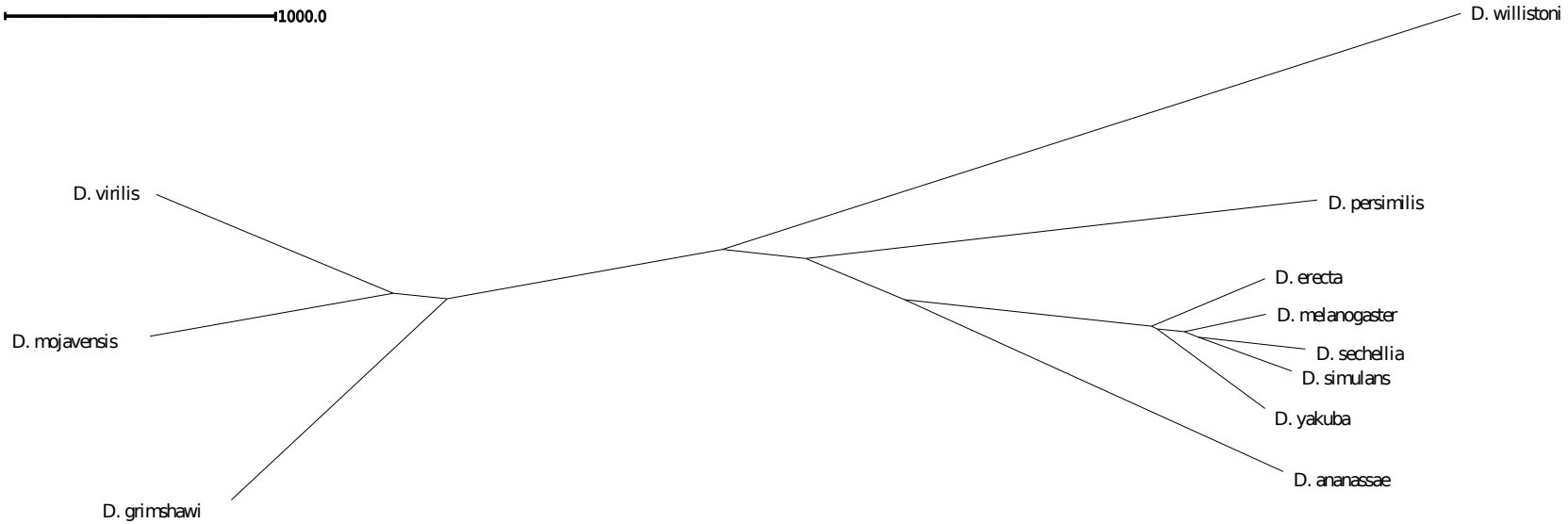


Computing Rearrangement Phylogenies on Contig-level resolved Genomes



- High quality

Computing Rearrangement Phylogenies on Contig-level resolved Genomes



- High quality
- Robust: On average only 0.53% deviation from computed pairwise distances.

Summarized Results

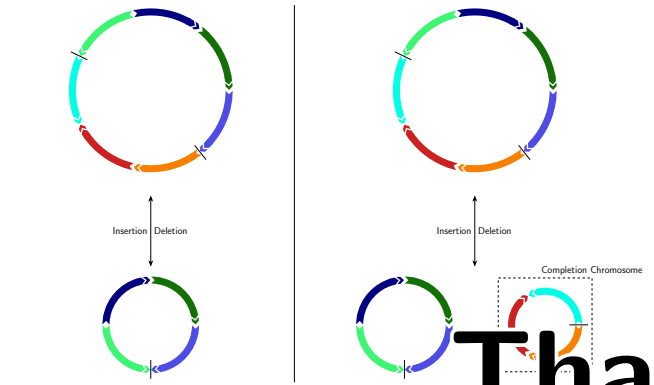
- More compact and simple DCJ-indel formula

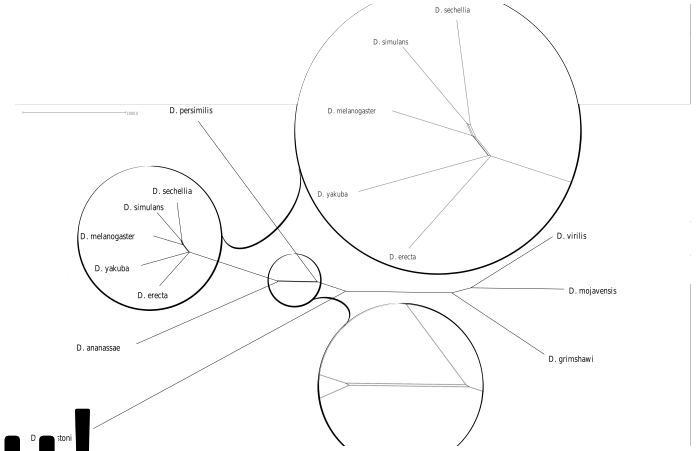
Summarized Results

- More compact and simple DCJ-indel formula
  - Link between BWS and Compeau-conceptualizations

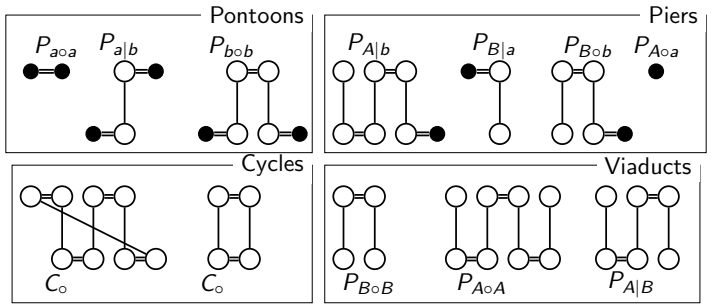
Summarized Results

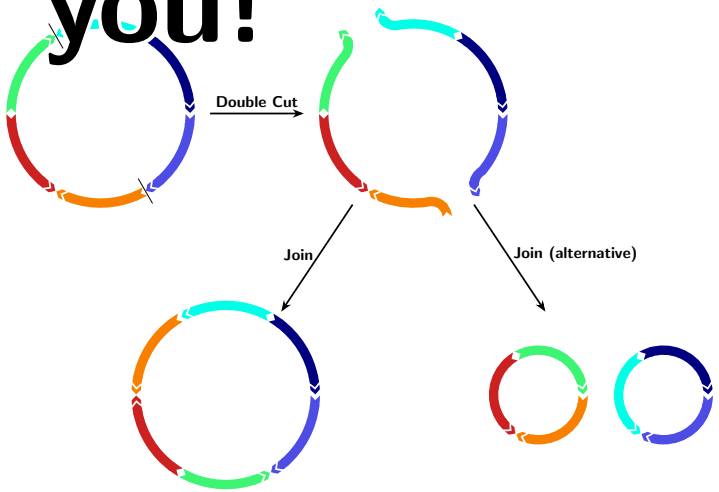
- More compact and simple DCJ-indel formula
  - Link between BWS and Compeau-conceptualizations
- Significant performance improvement of `ding` for high numbers of linear chromosomes/contigs


Insertion
Deletion
Insertion
Deletion
Completion
Chromosome


D. sechellia
D. simulans
D. persimilis
D. melanogaster
D. yakuba
D. virilis
D. erecta
D. mojavensis
D. ananassae
D. grimshawi

Thank you!


Pontoons
$P_{a\circ a}$
$P_{a|b}$
$P_{b\circ b}$
Piers
$P_{A|b}$
$P_{B|a}$
$P_{B\circ b}$
$P_{A\circ a}$
Cycles
$C_\circ$
$C_\circ$
Viaducts
$P_{B\circ B}$
$P_{A\circ A}$
$P_{A|B}$


Double Cut
Join
Join (alternative)