

Genome rearrangements

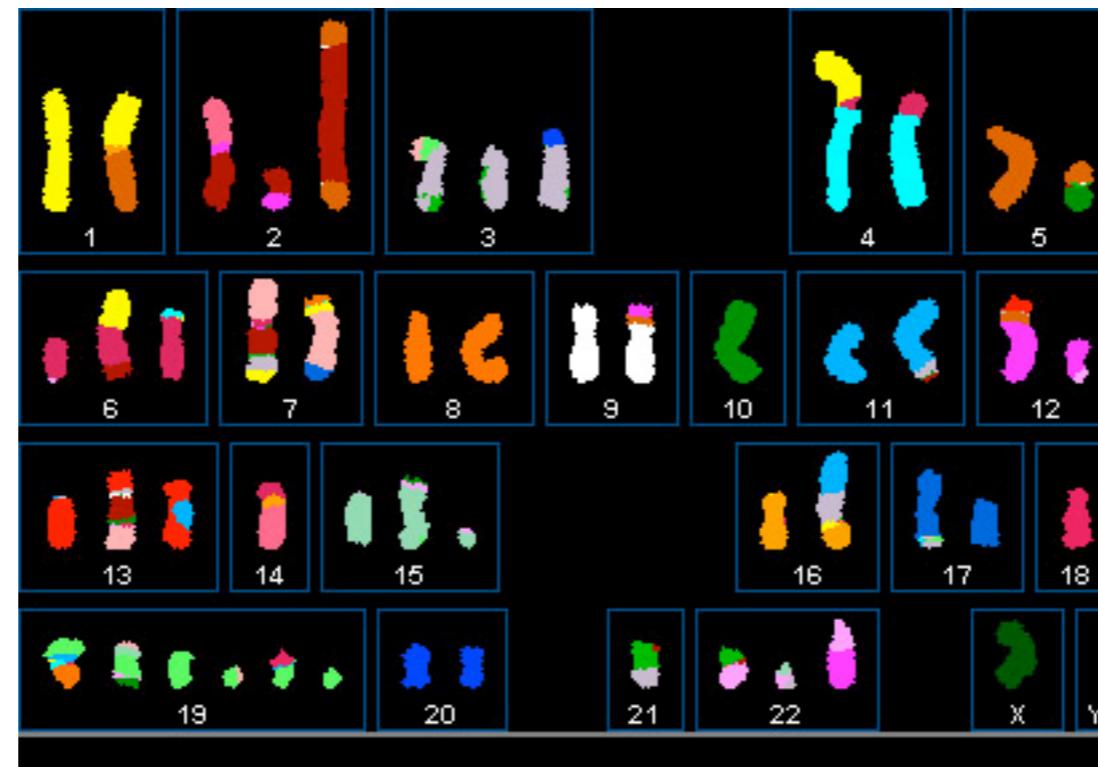
Aganezov Sergey jr. 2011 CS Center

Genome rearrangements

- Эволюция
- Раковые заболевания
- Предсказание будущих эволюционных изменений

Genome rearrangements

- Раковые заболевания



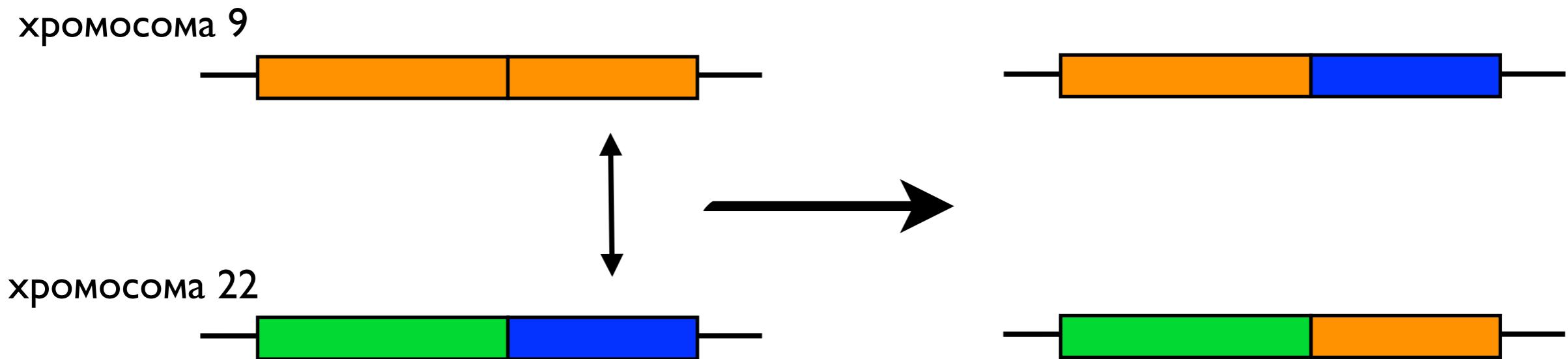
Genome rearrangements

- “Philadelphia” chromosome



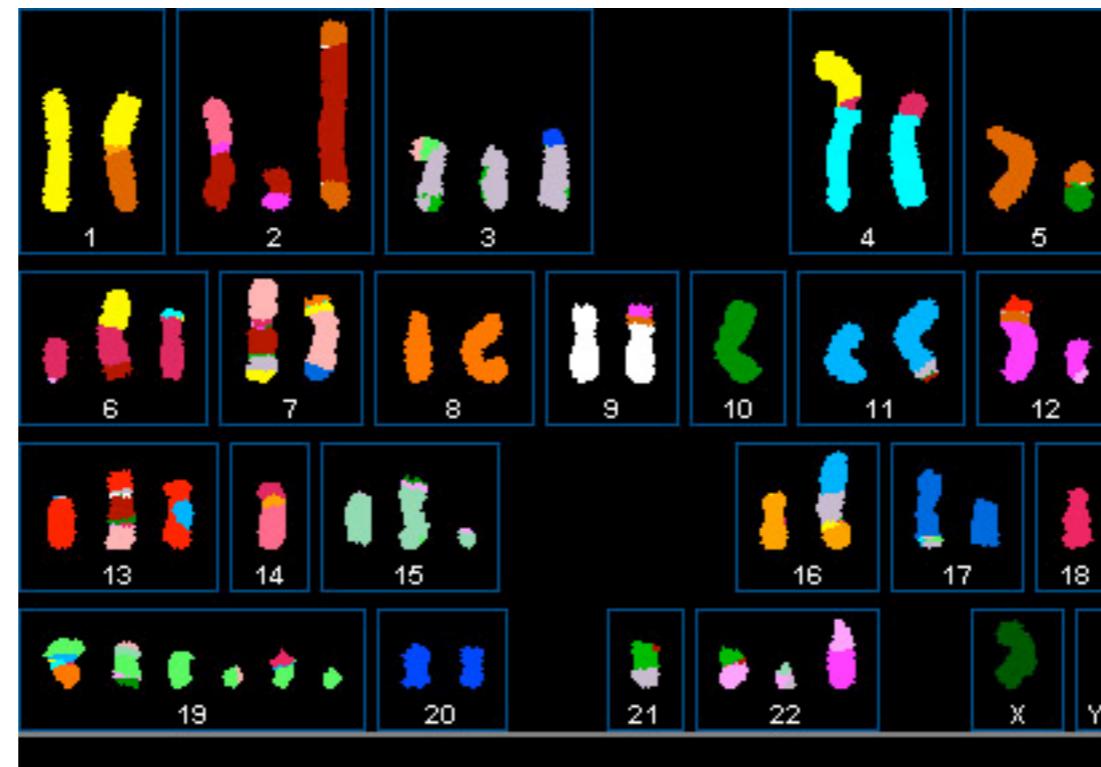
Genome rearrangements

- “Philadelphia” chromosome

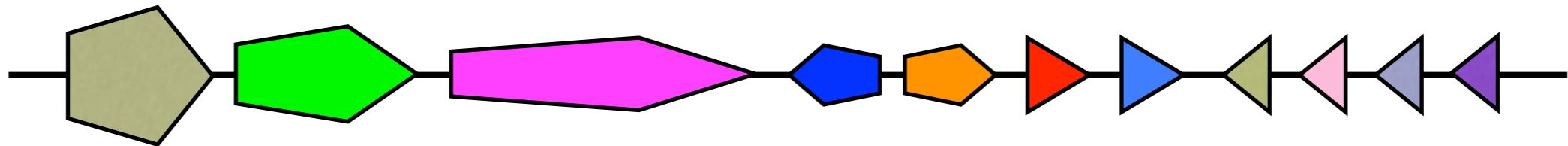


Genome rearrangements

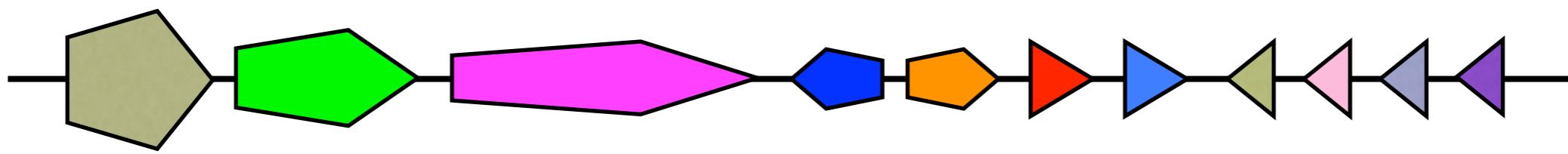
- Раковые заболевания



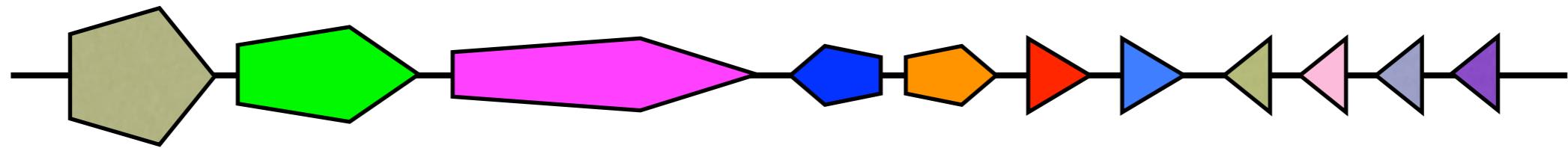
неизвестный предок



неизвестный предок

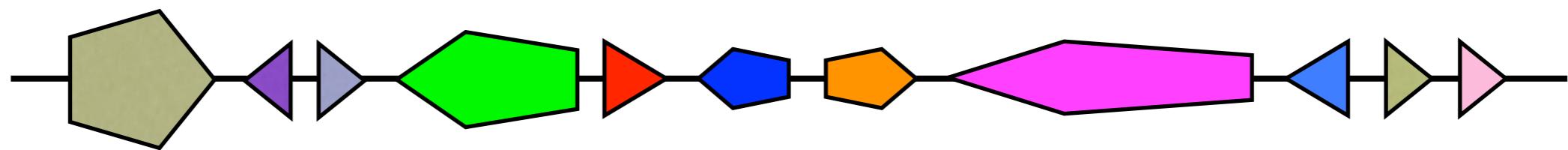


неизвестный предок



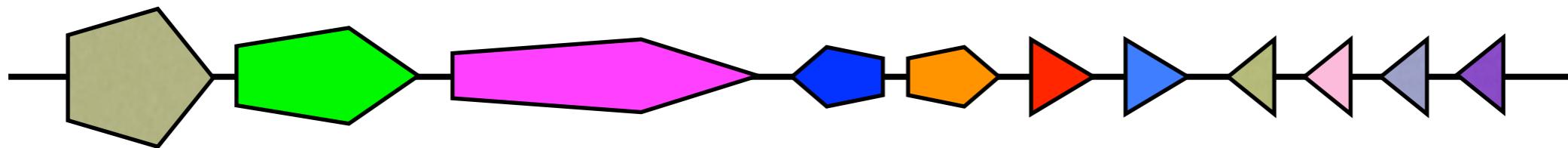
80 000 000 лет

мышь

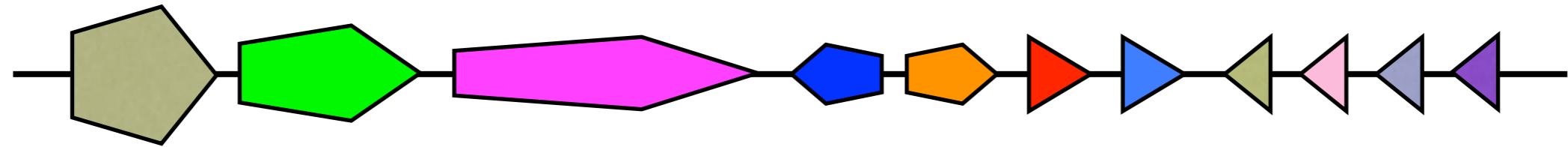


80 000 000 лет ранее...

неизвестный предок

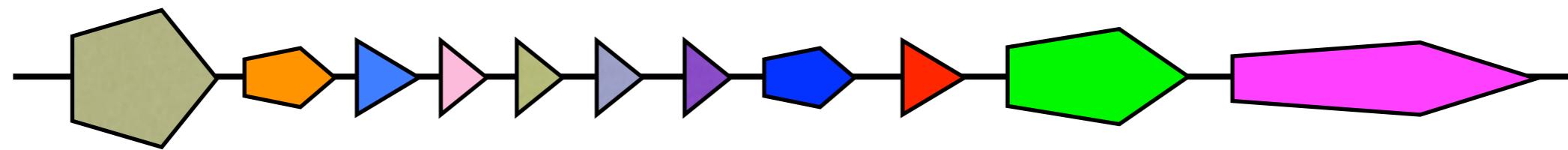


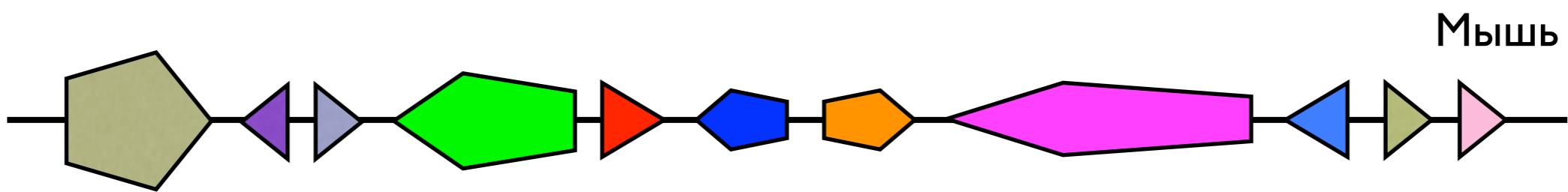
неизвестный предок



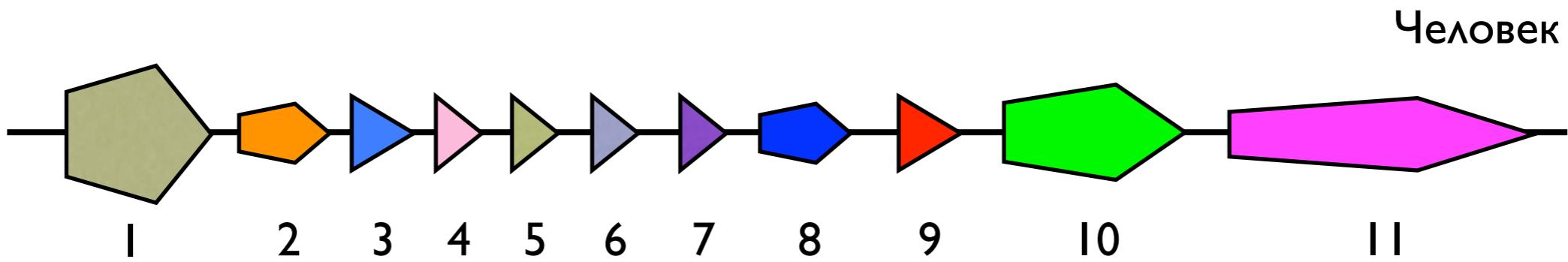
80 000 000 лет

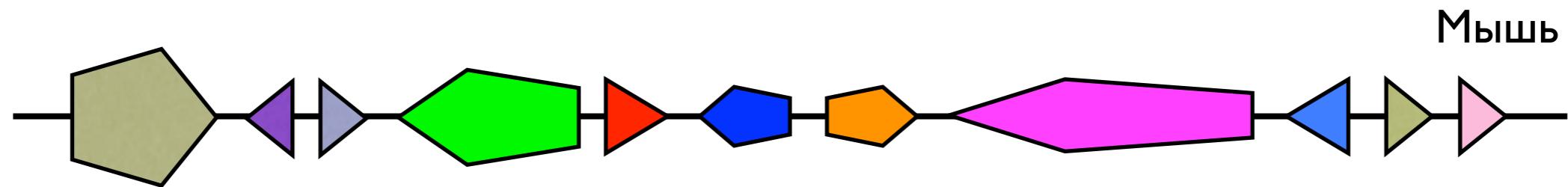
Человек



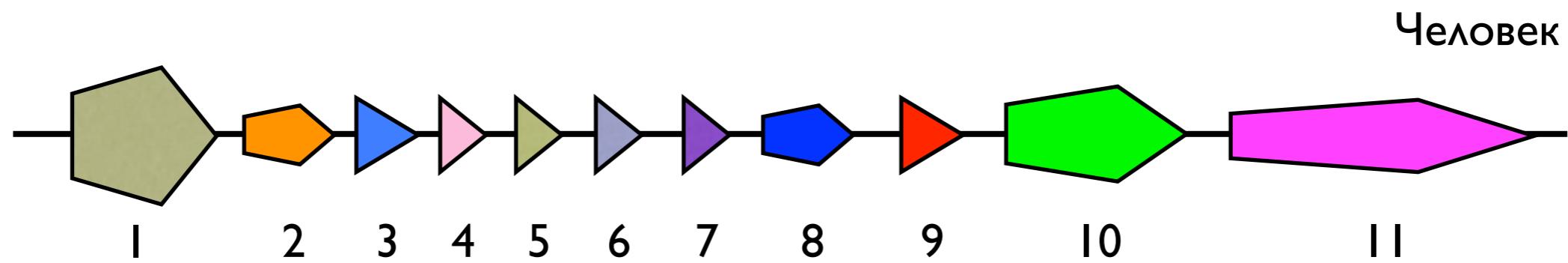


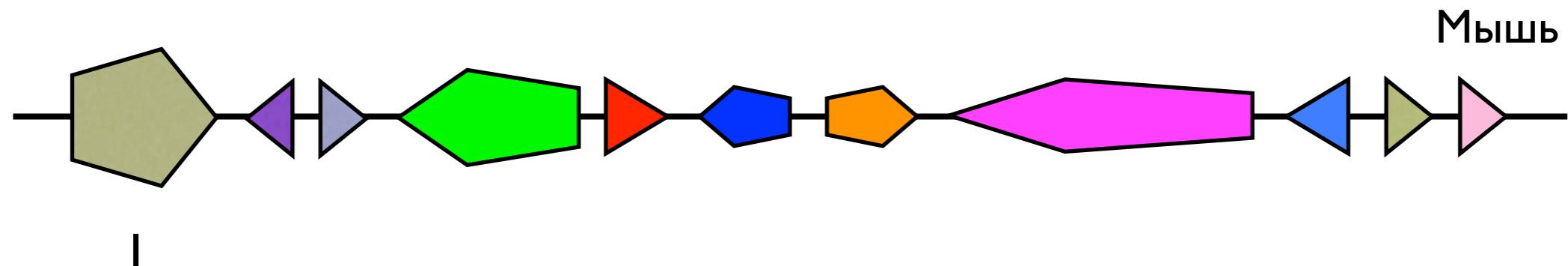
80 000 000 лет



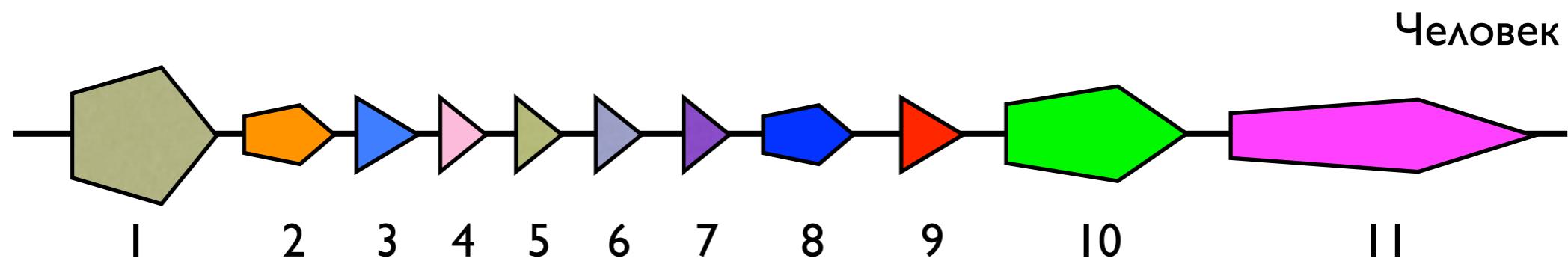


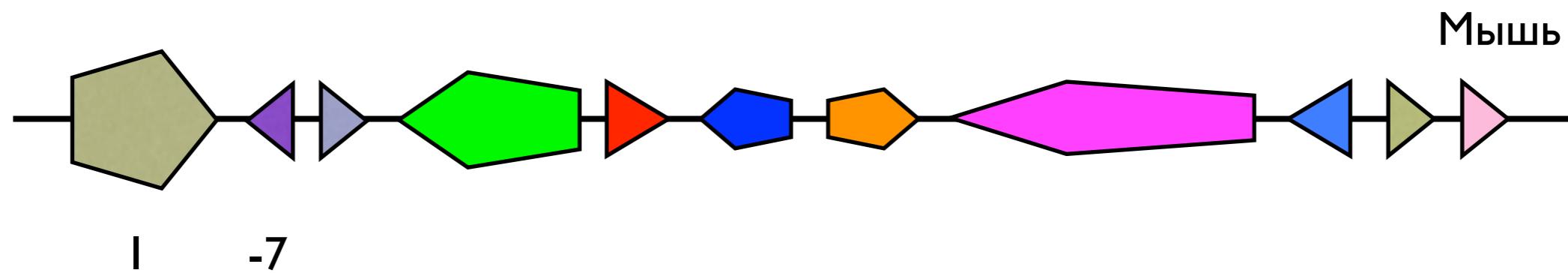
80 000 000 лет



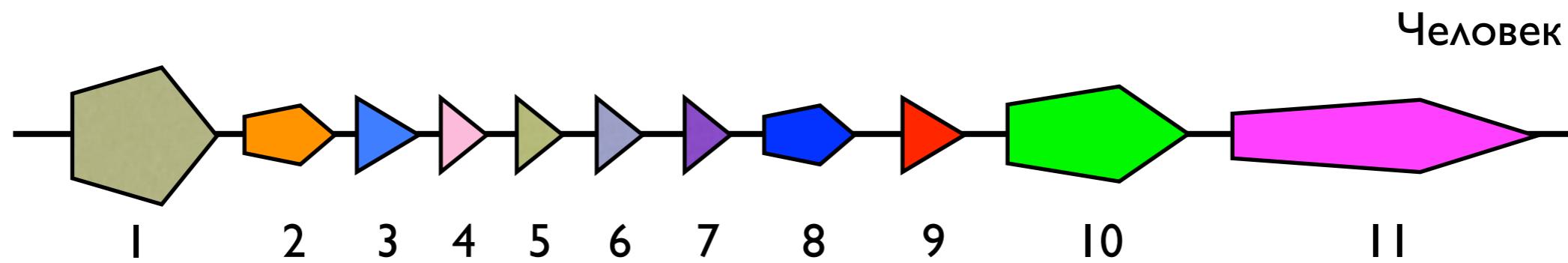


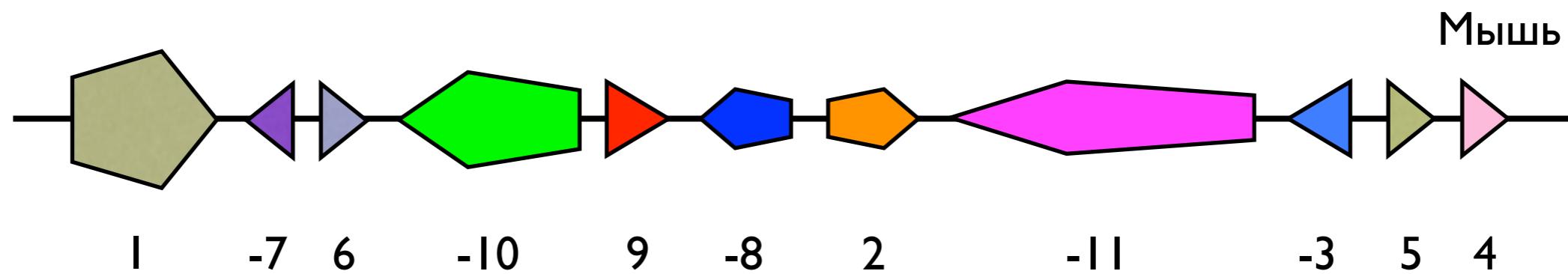
80 000 000 лет



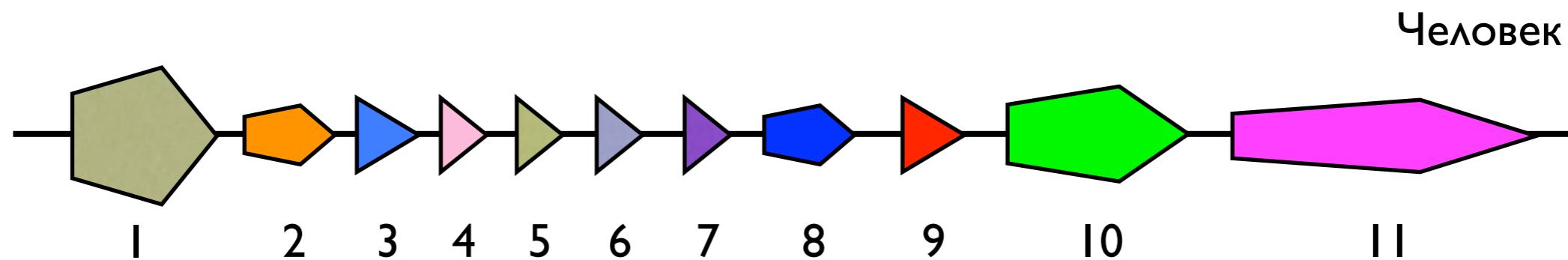


80 000 000 лет



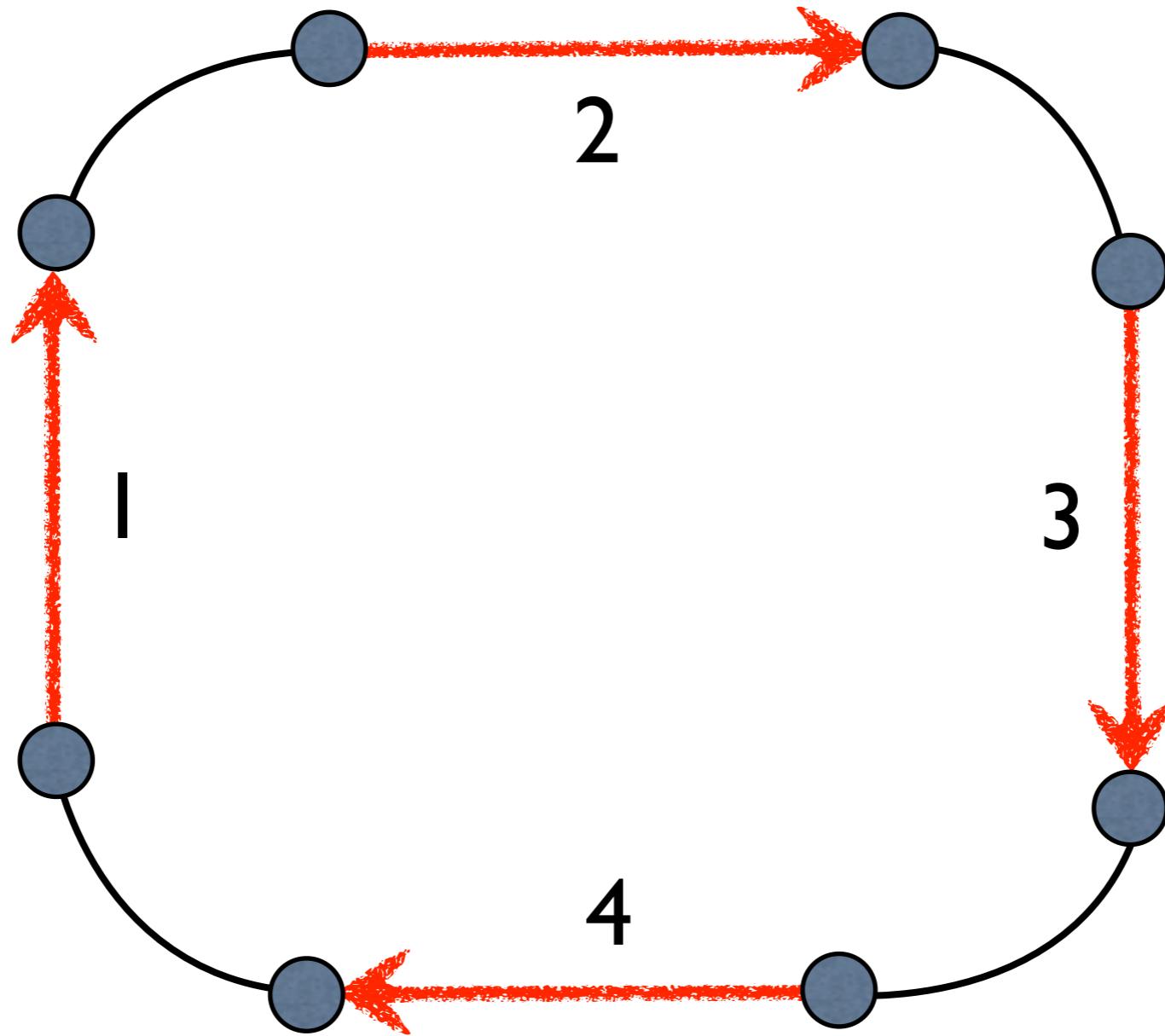


80 000 000 лет



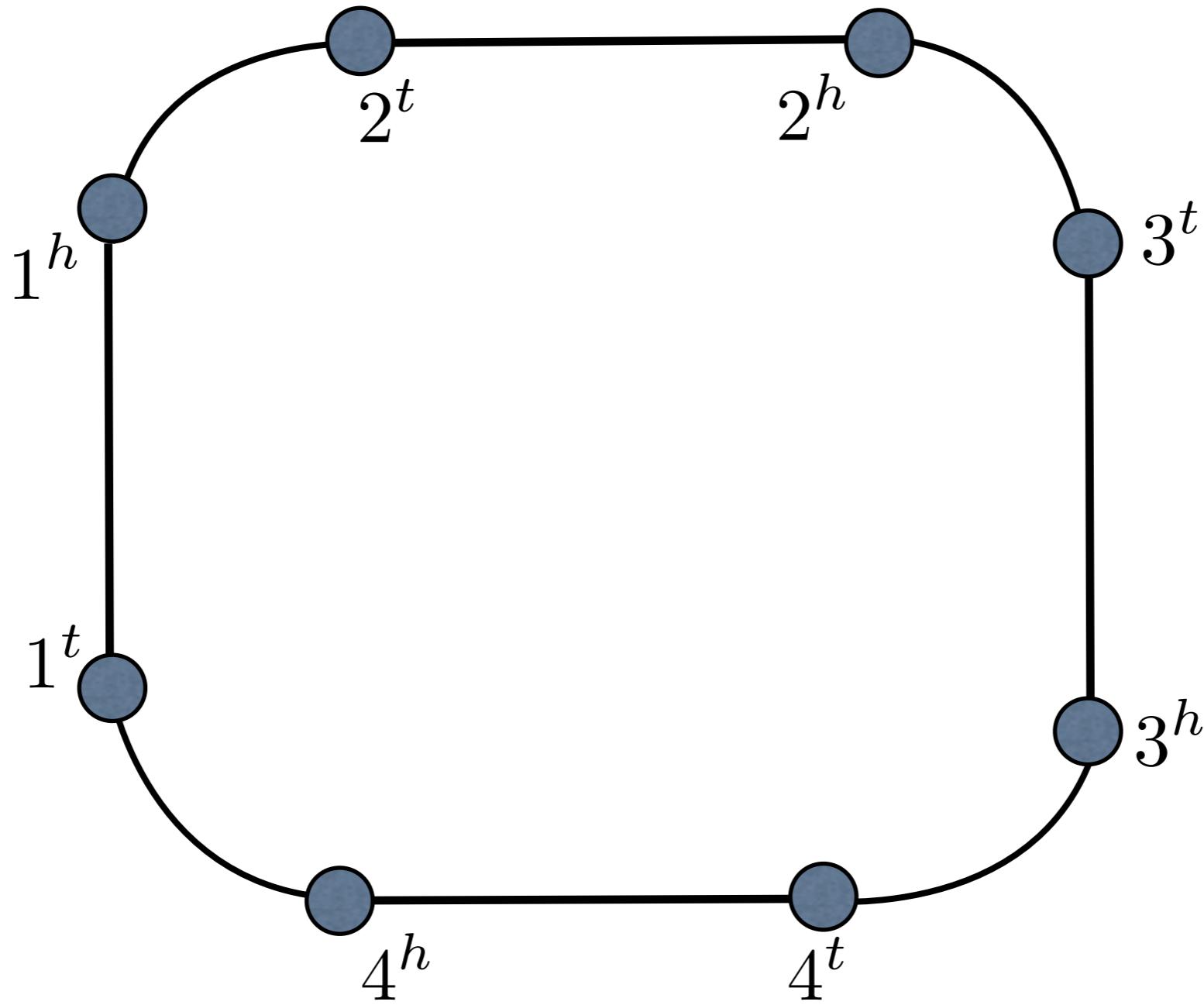
Циклический геном

$$A = (1, 2, 3, 4)$$



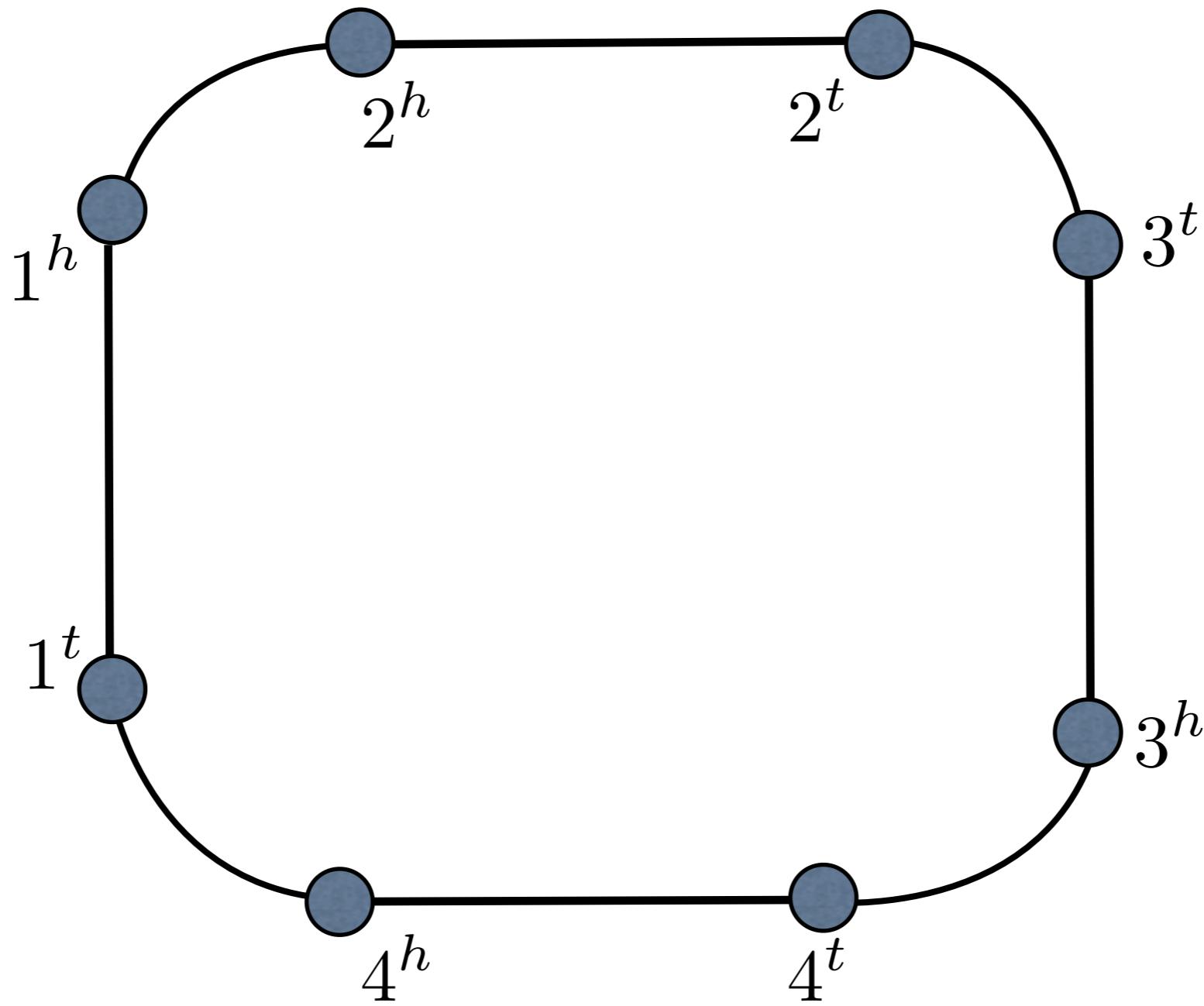
Циклический геном

$$A = (1, 2, 3, 4)$$



Циклический геном

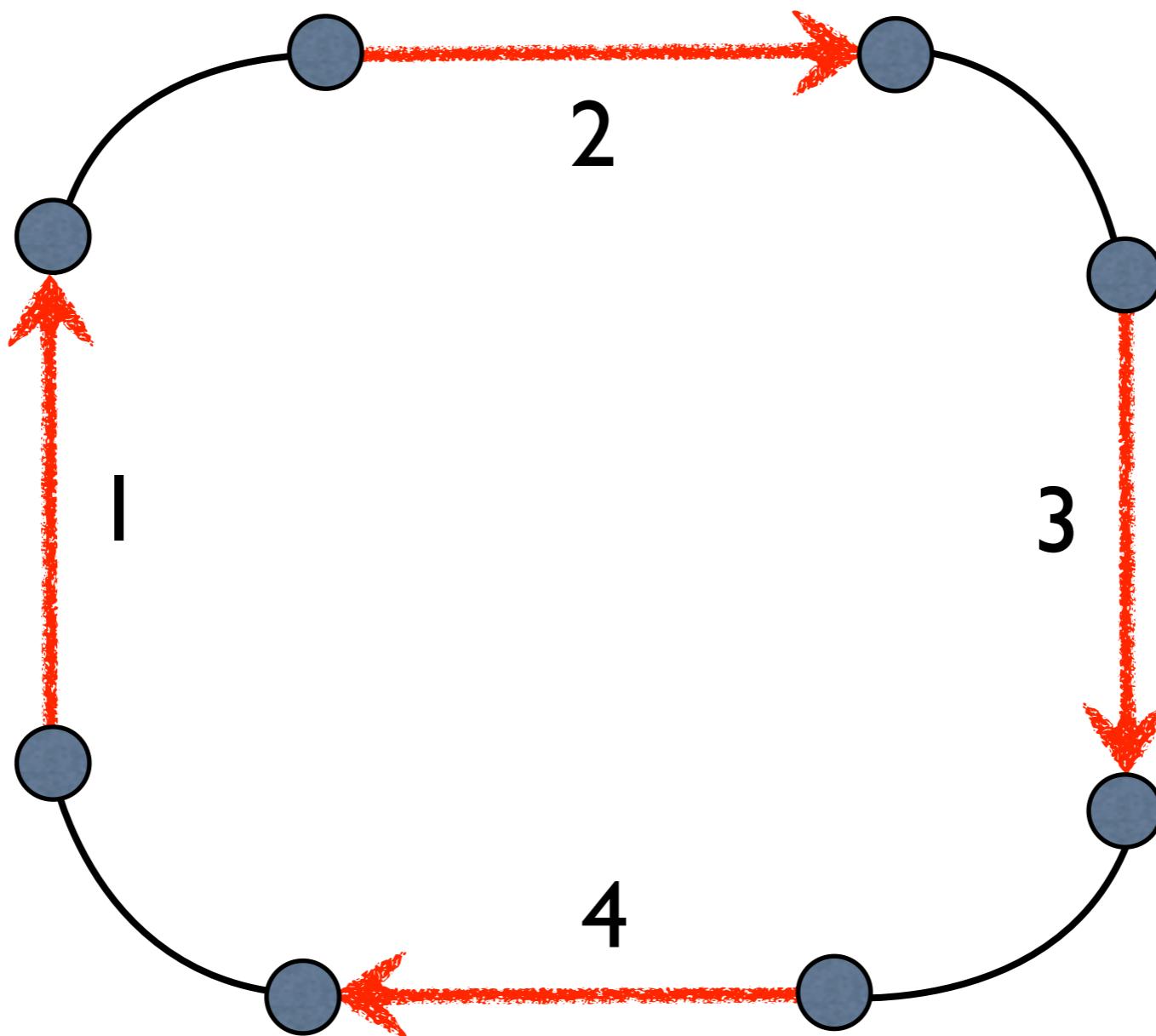
$$A = (1, -2, 3, 4)$$



2-break операции

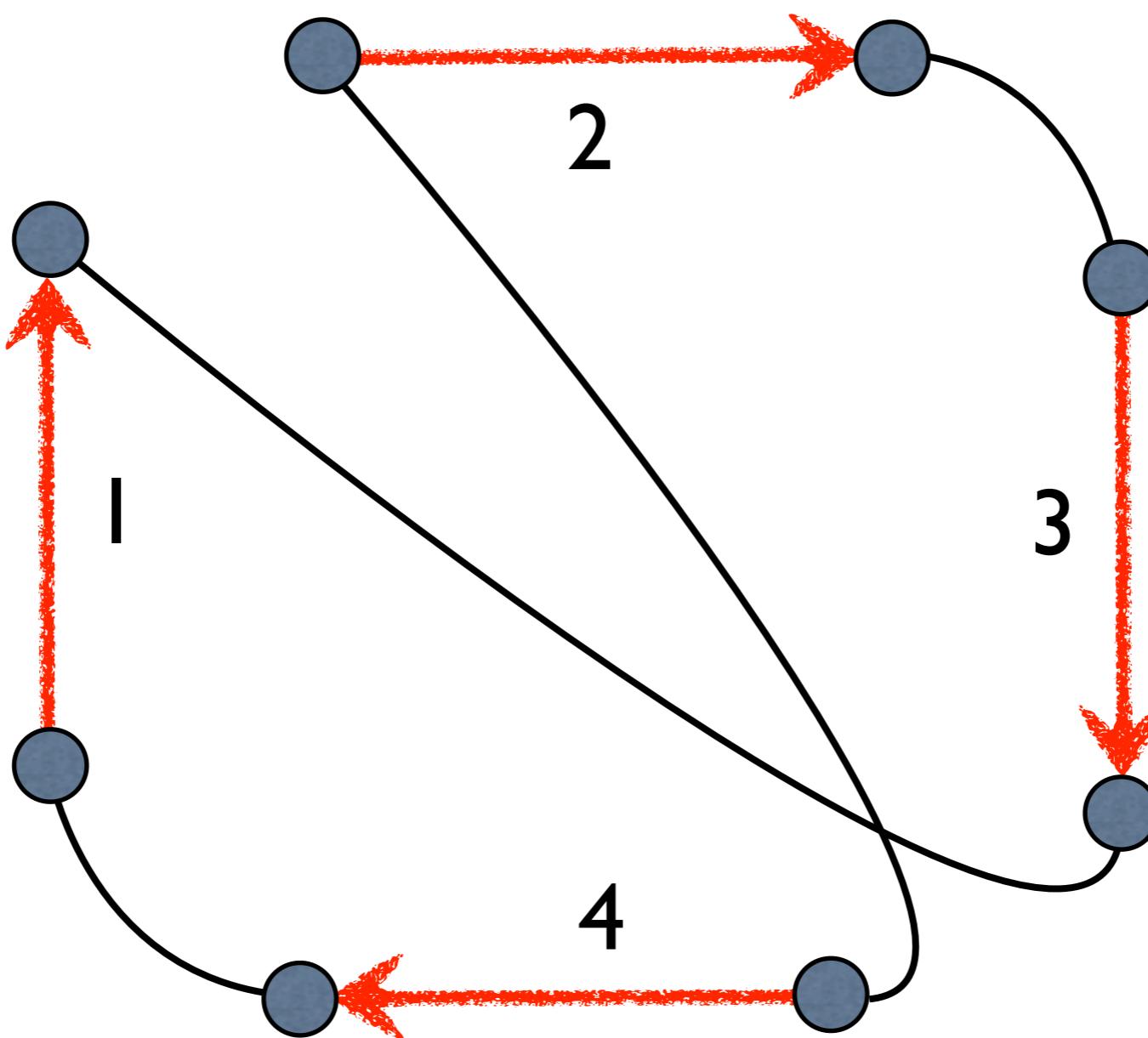
Reversal

$$A = (1, 2, 3, 4)$$



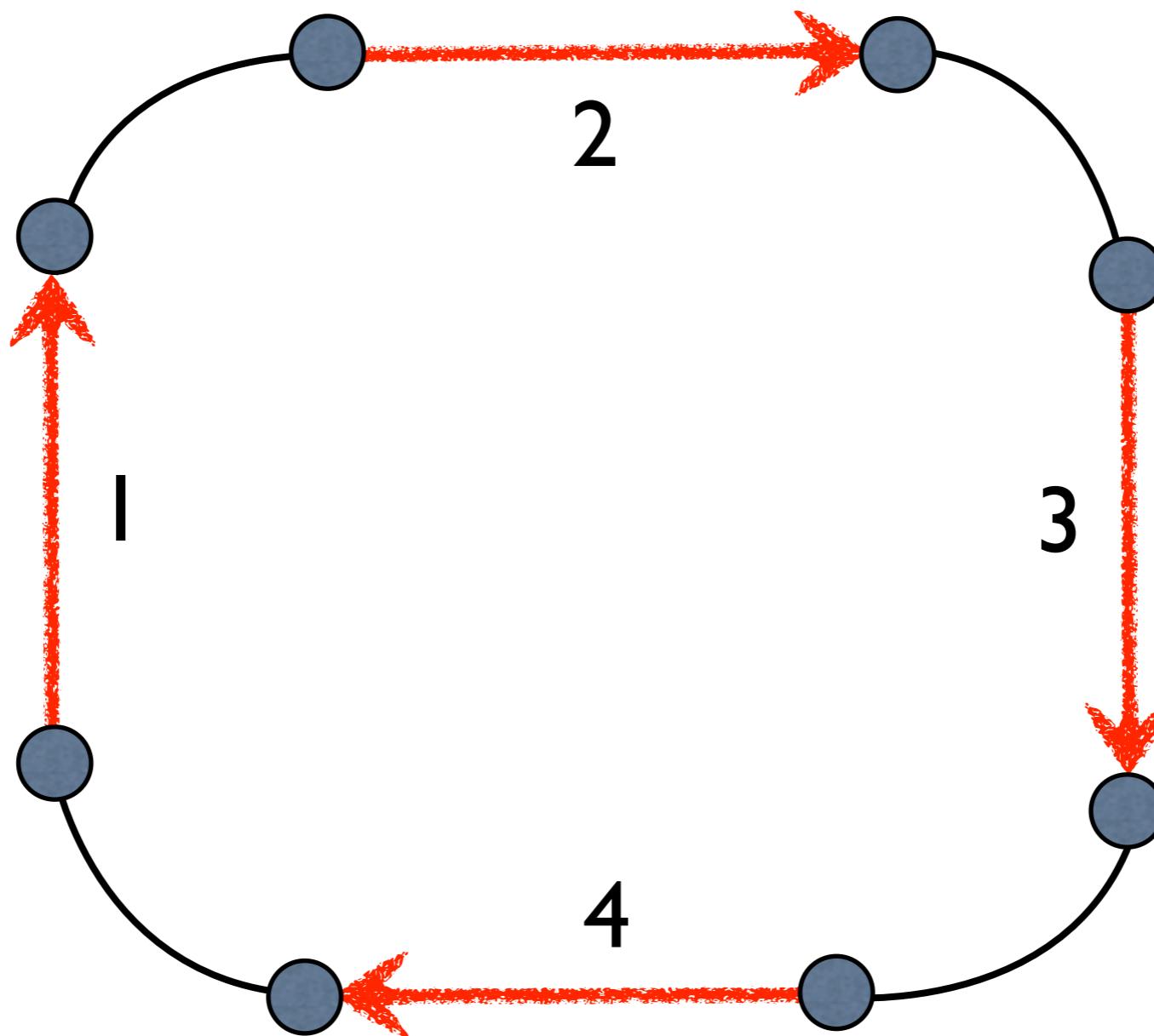
Reversal

$$A = (1, -3, -2, 4)$$



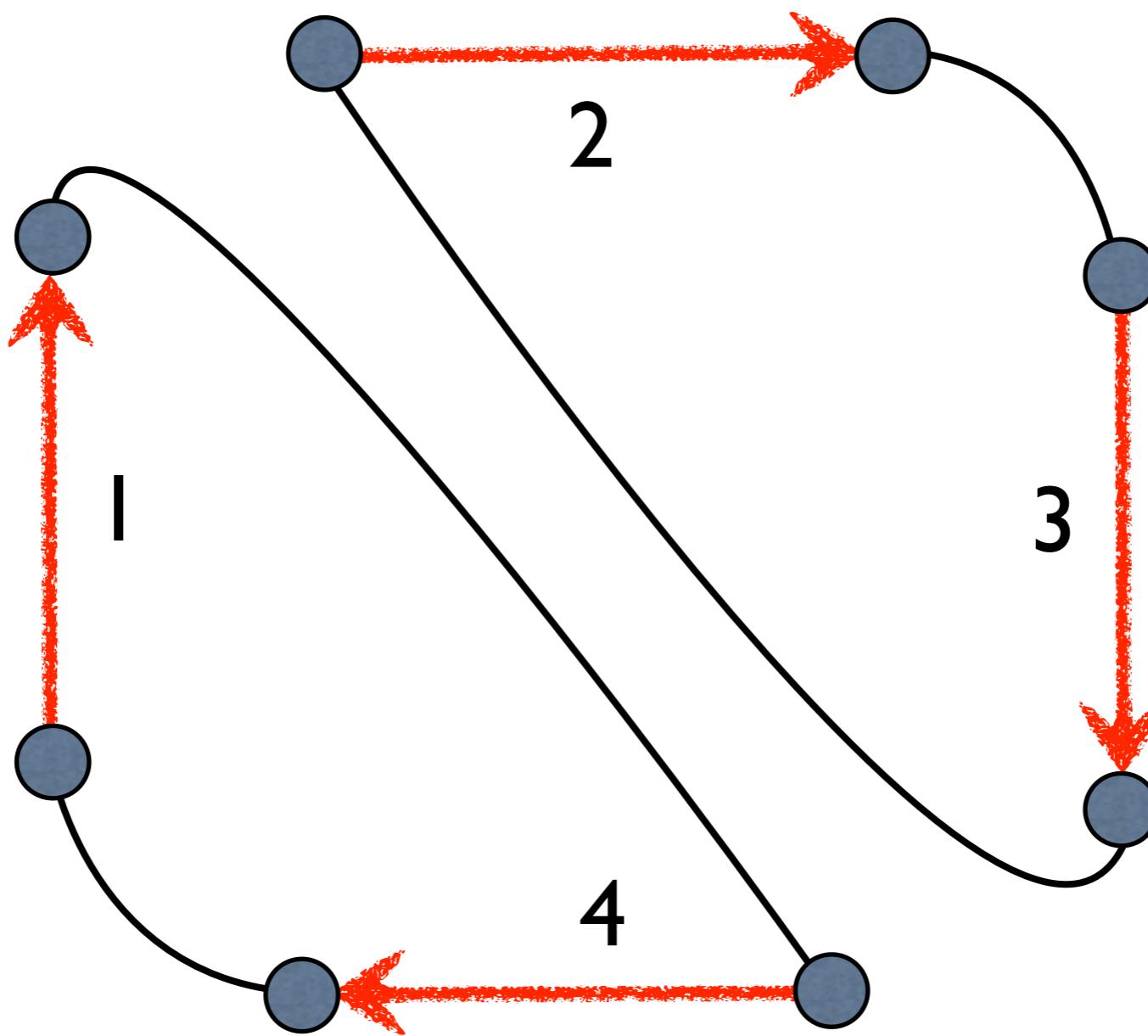
Fissions

$$A = (1, 2, 3, 4)$$



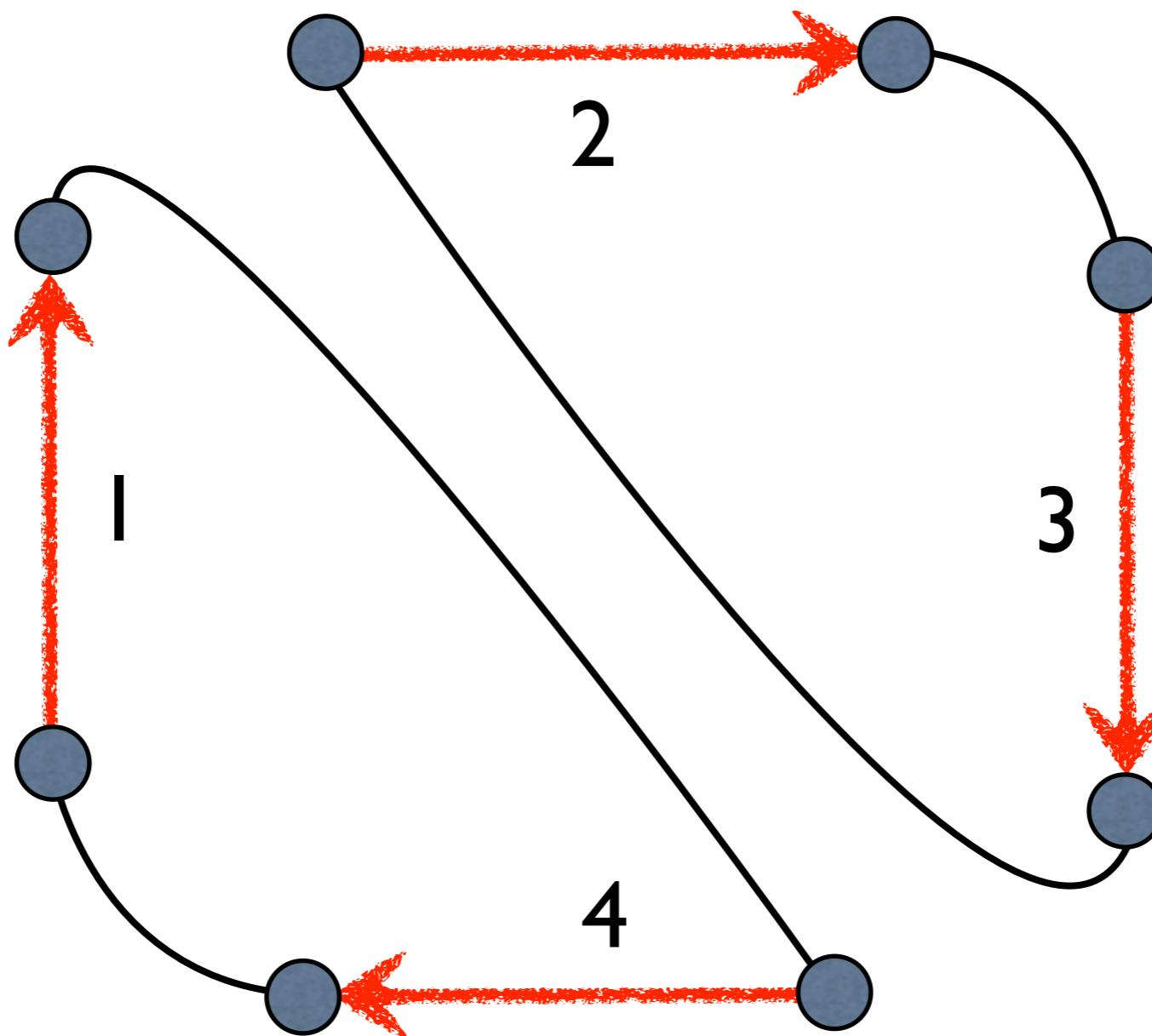
Fissions

$$A = (1, 2, 3, 4) \longrightarrow A = (1, 4) (2, 3)$$



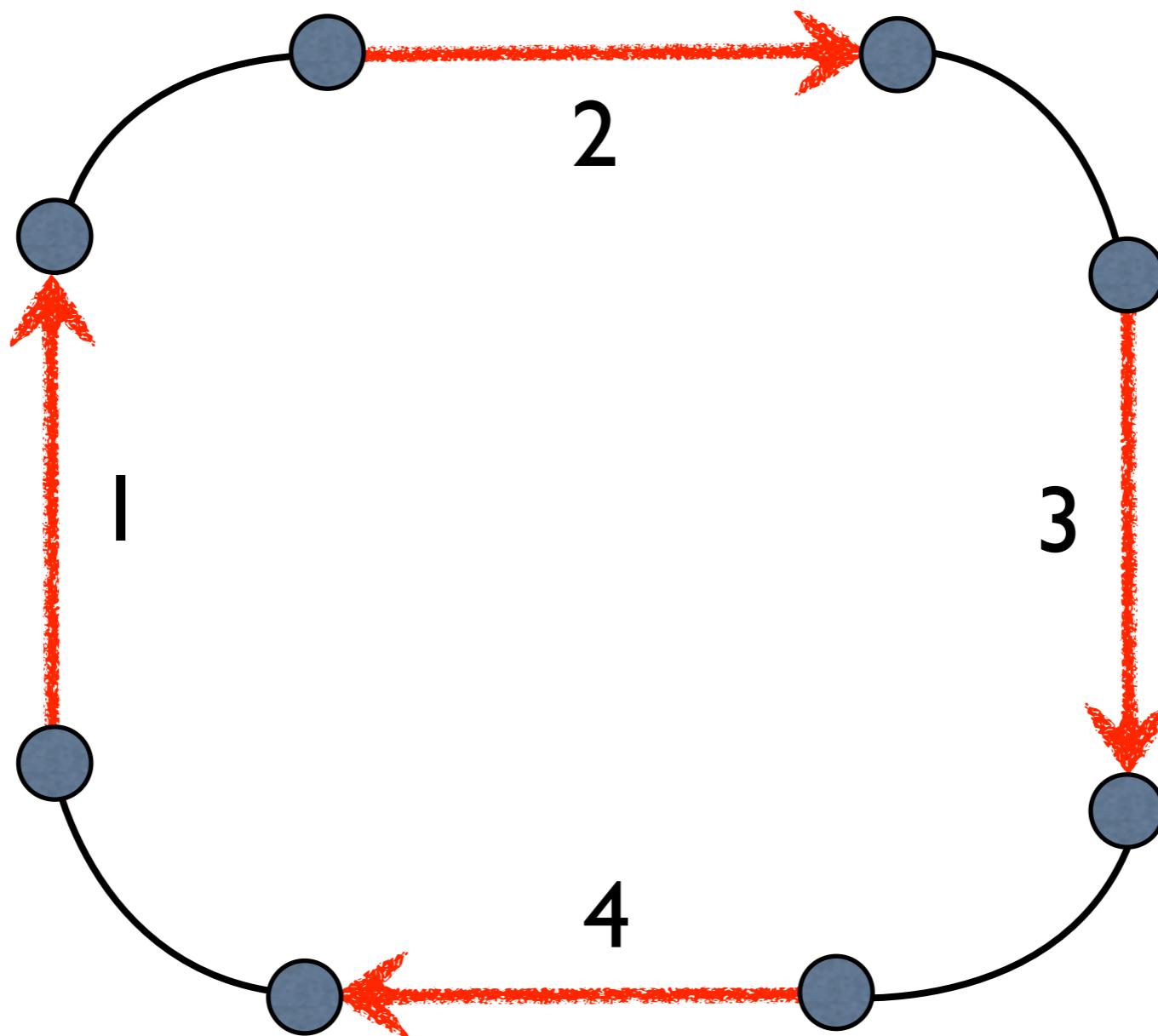
Fusion

$$A = (1,4) (2,3)$$

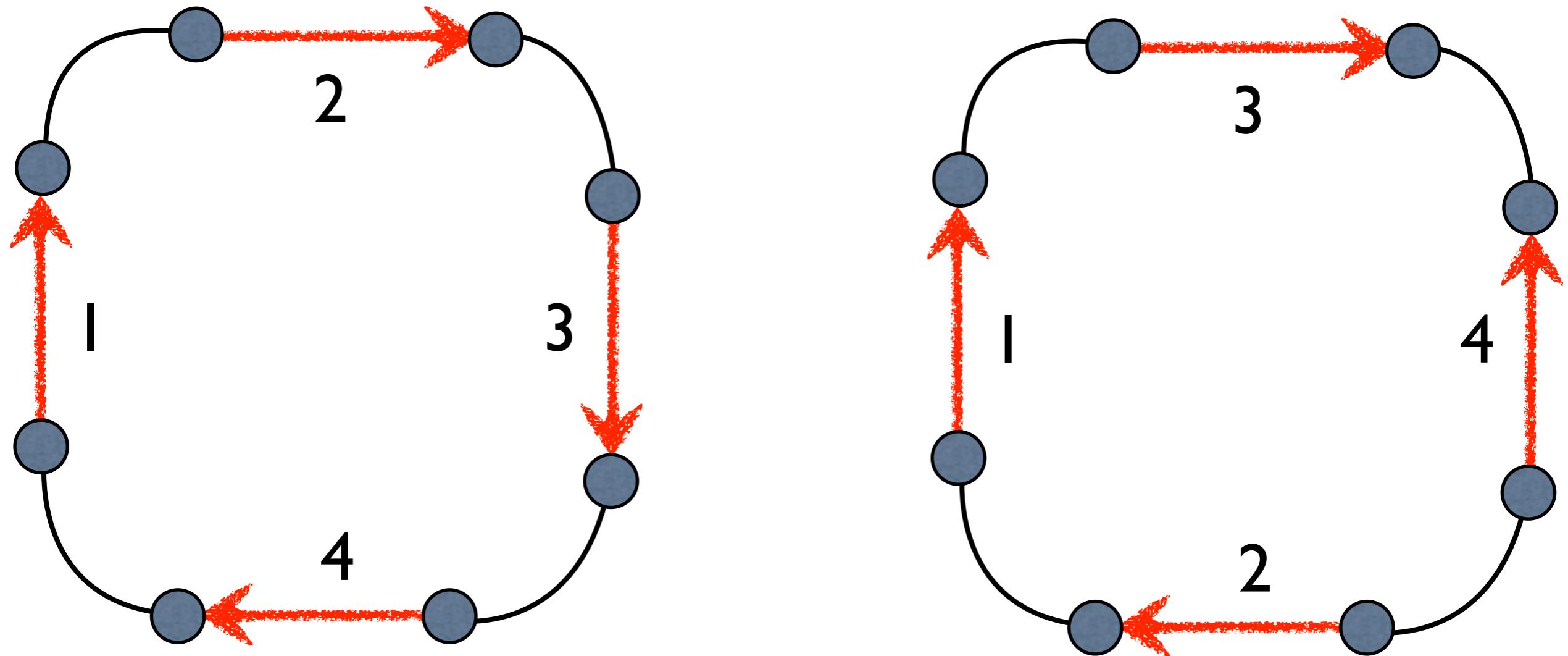


Fusion

$$A = (1,4) \ (2,3) \longrightarrow A = (1,2,3,4)$$

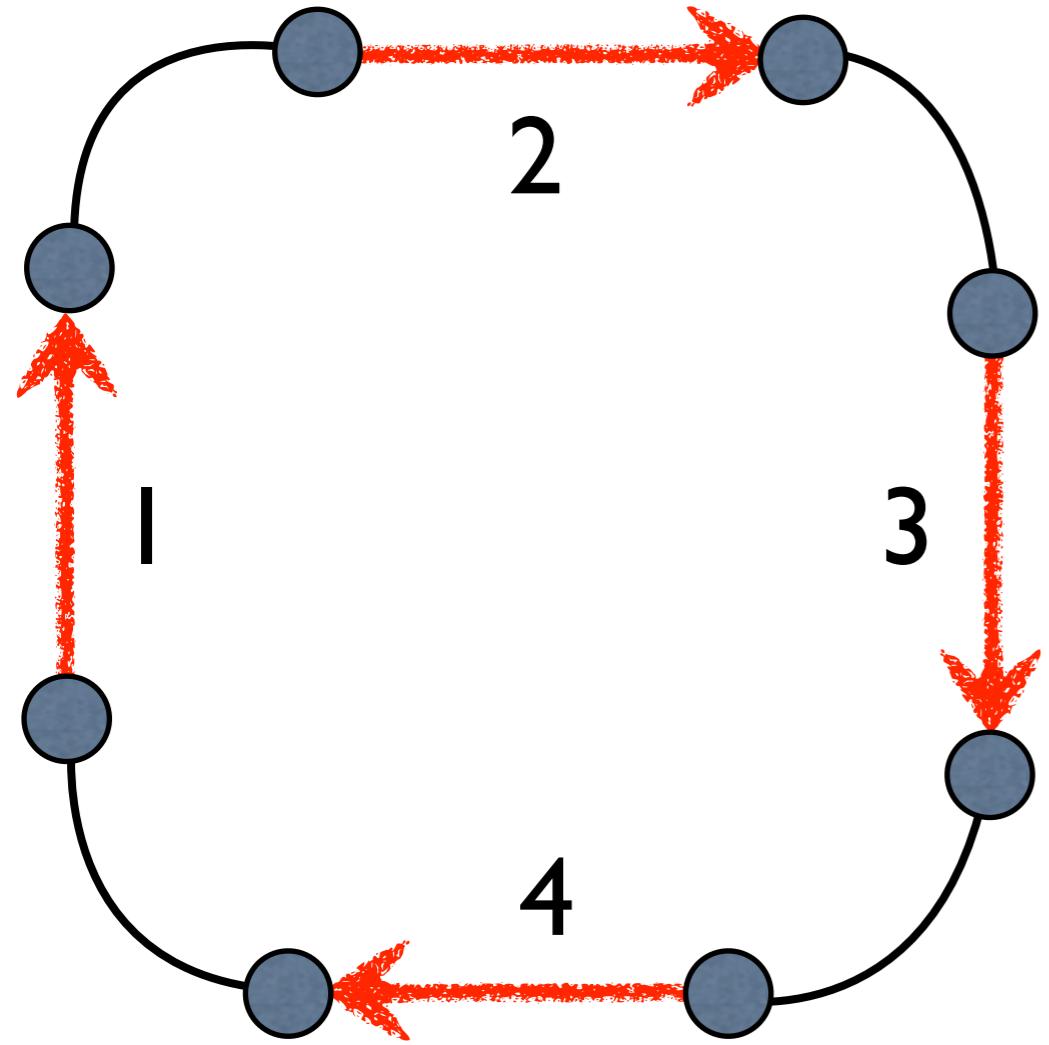


Breakpoint graph

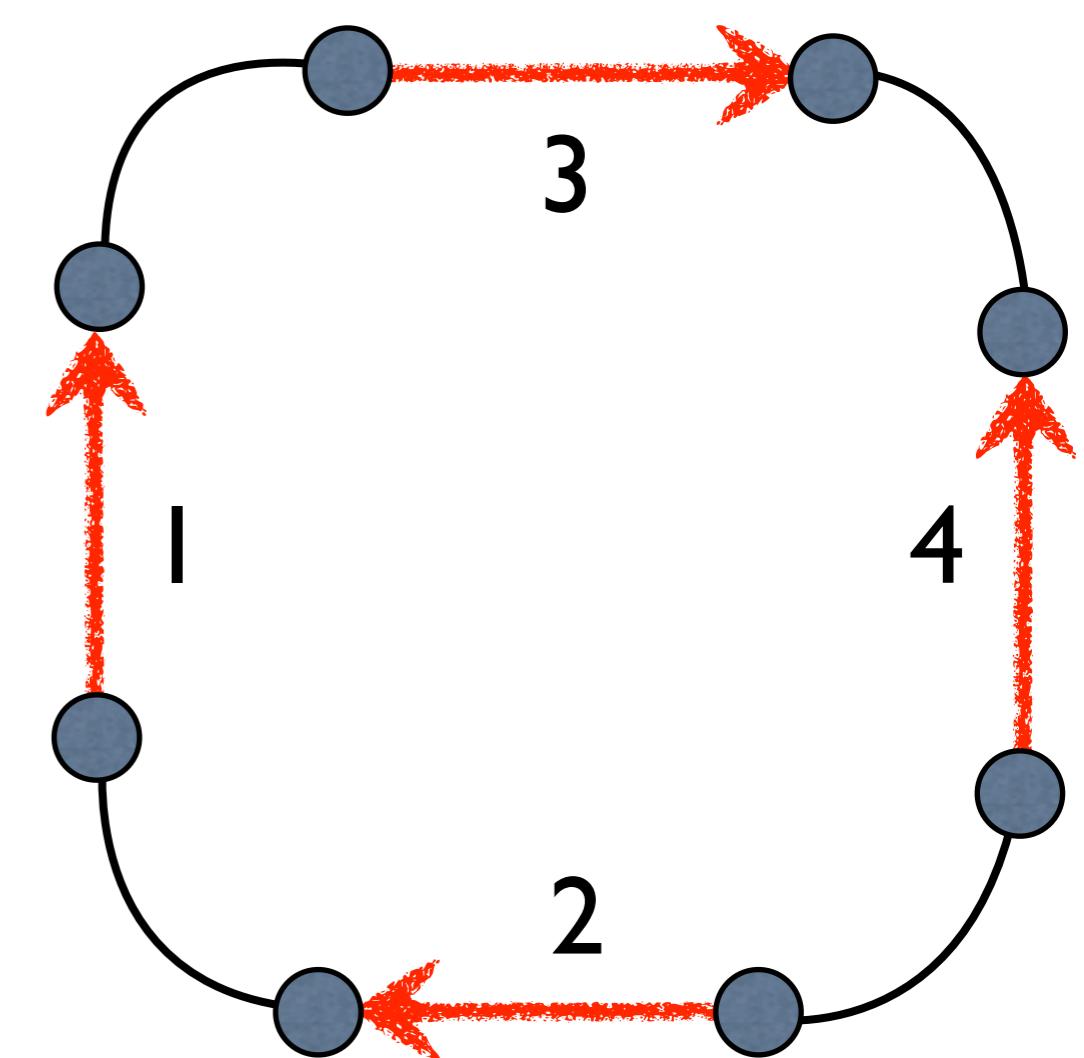


Breakpoint graph

$A = (1, 2, 3, 4)$

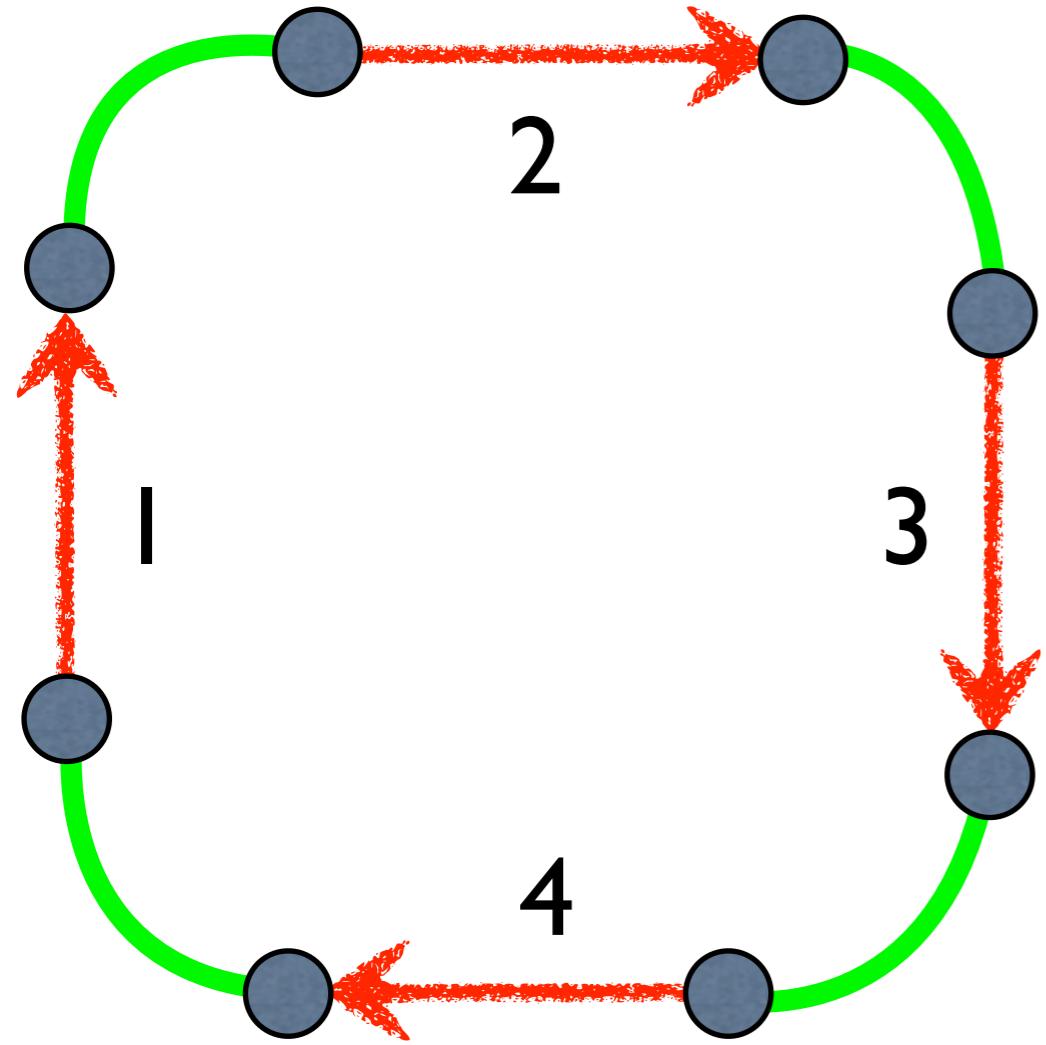


$B = (1, 3, -4, 2)$

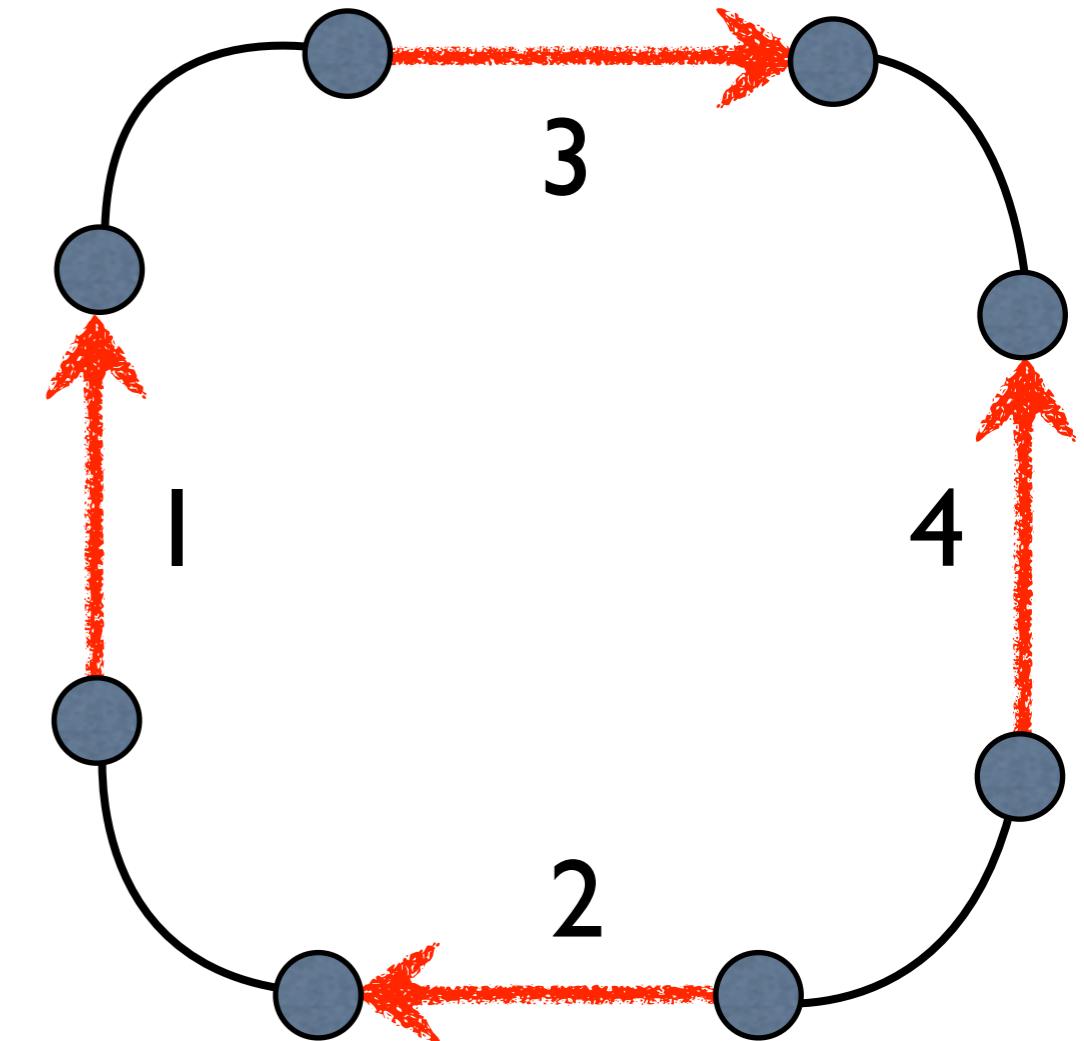


Breakpoint graph

$A = (1, 2, 3, 4)$

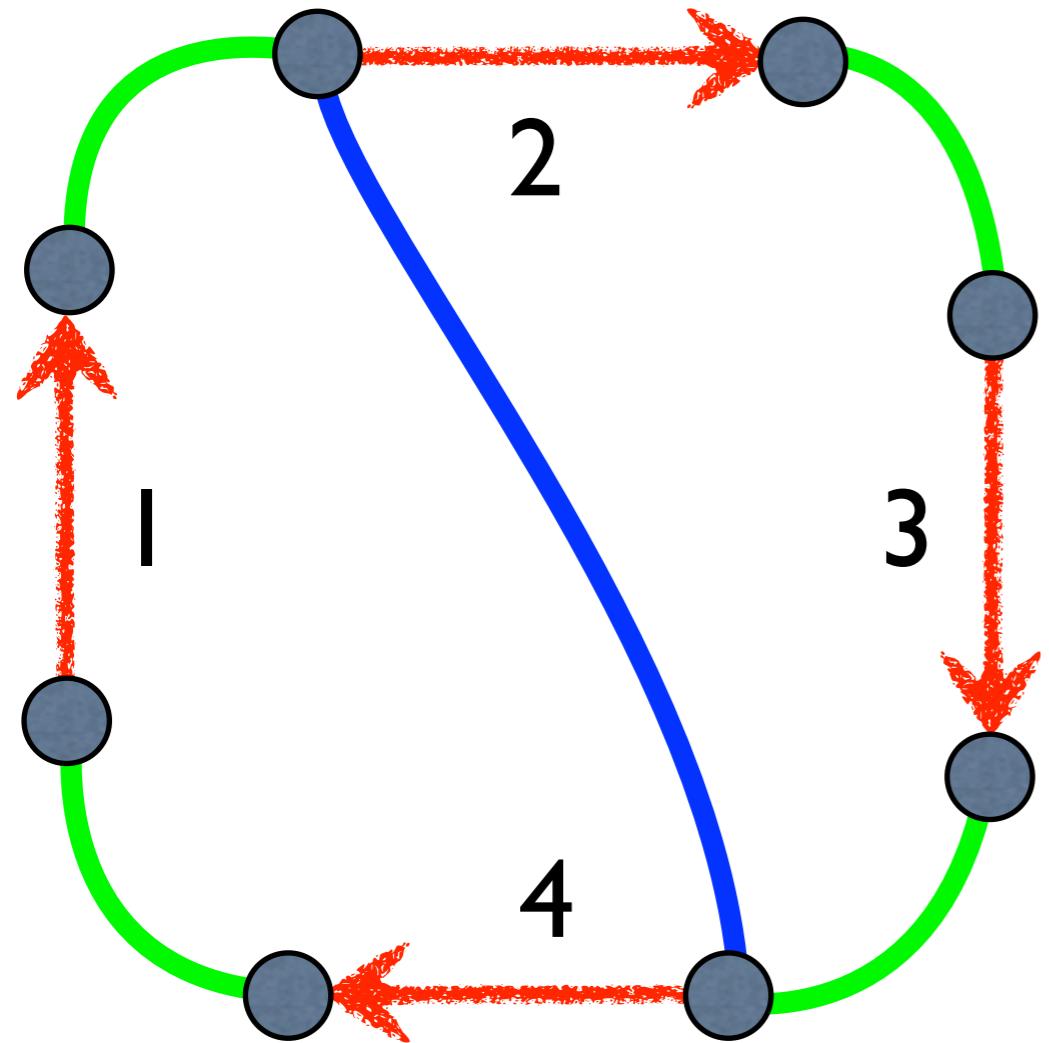


$B = (1, 3, -4, 2)$

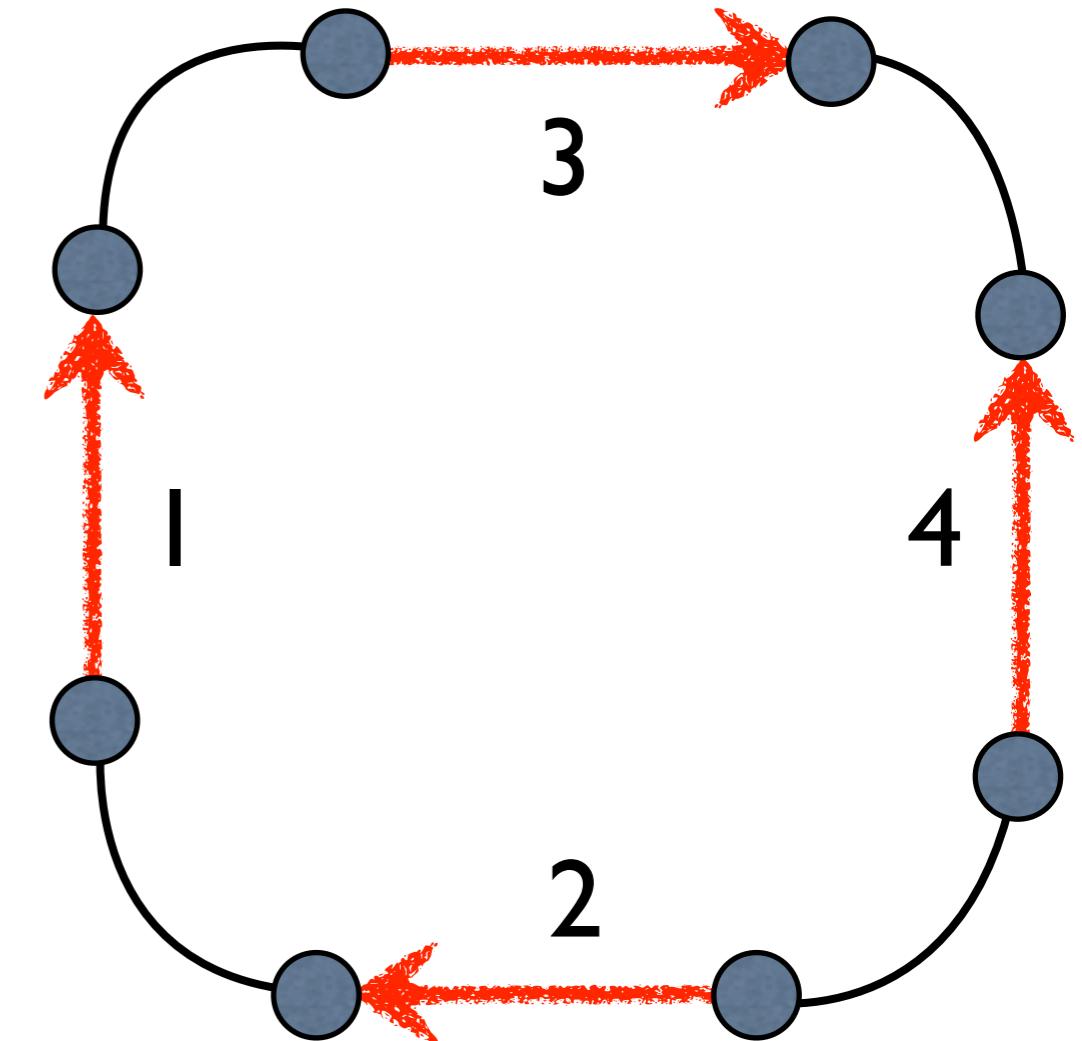


Breakpoint graph

$A = (1, 2, 3, 4)$

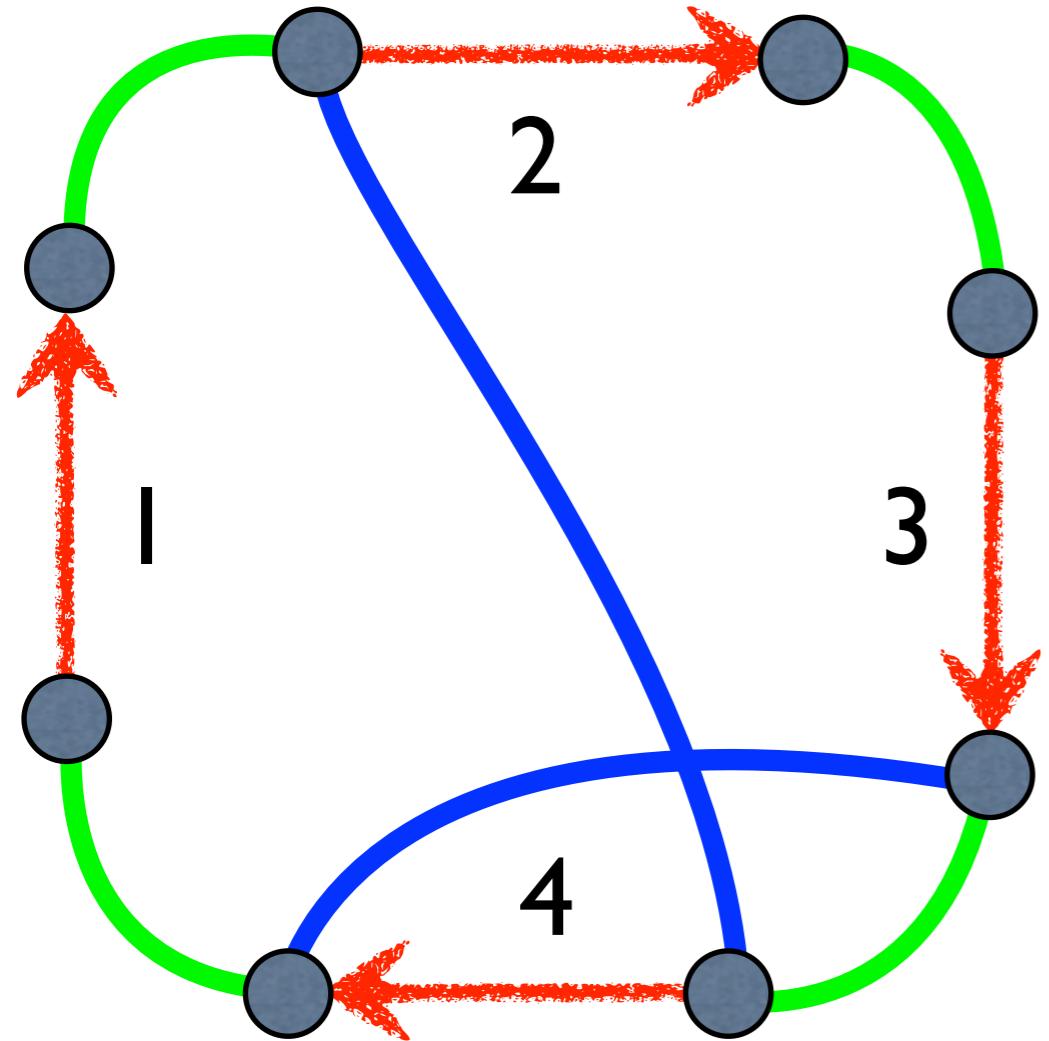


$B = (1, 3, -4, 2)$

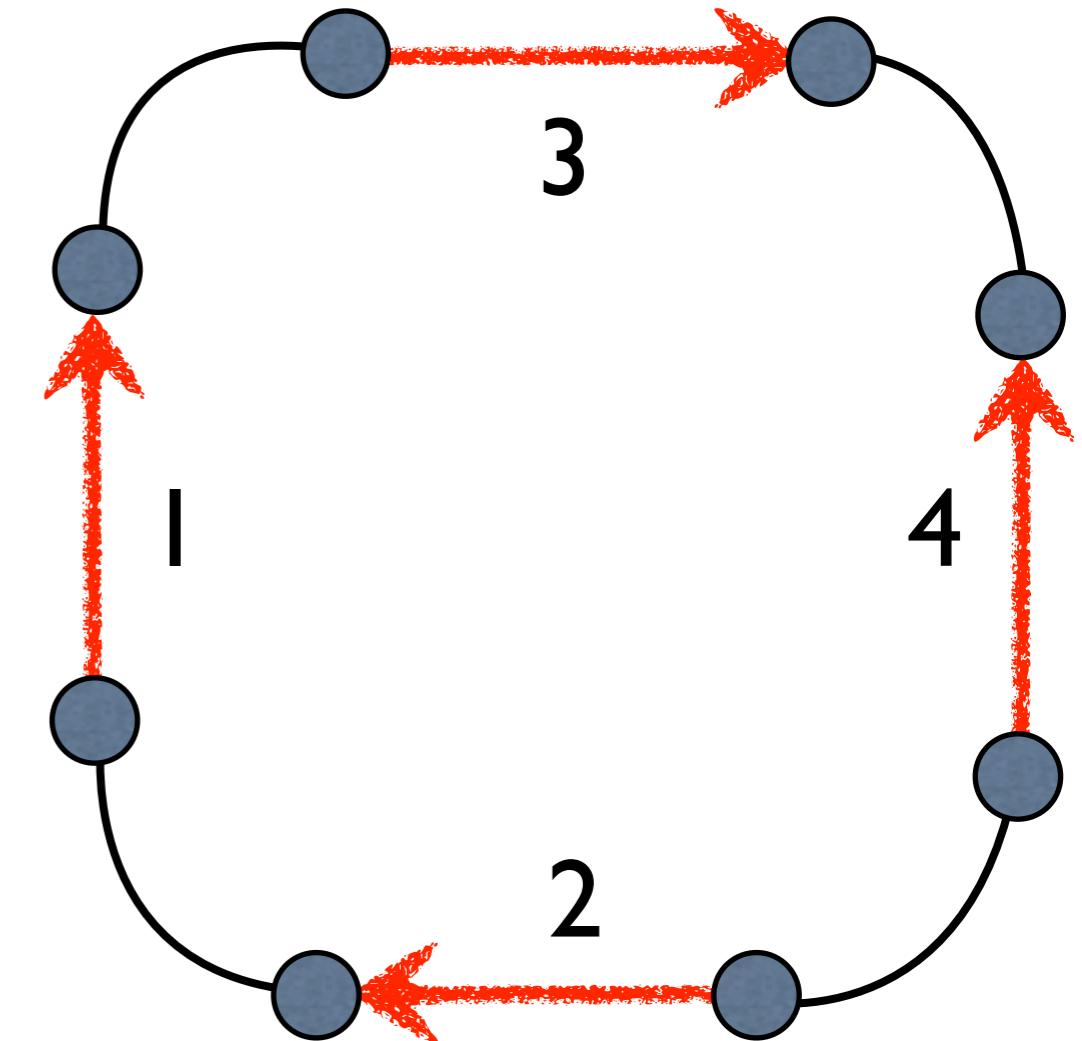


Breakpoint graph

$A = (1, 2, 3, 4)$

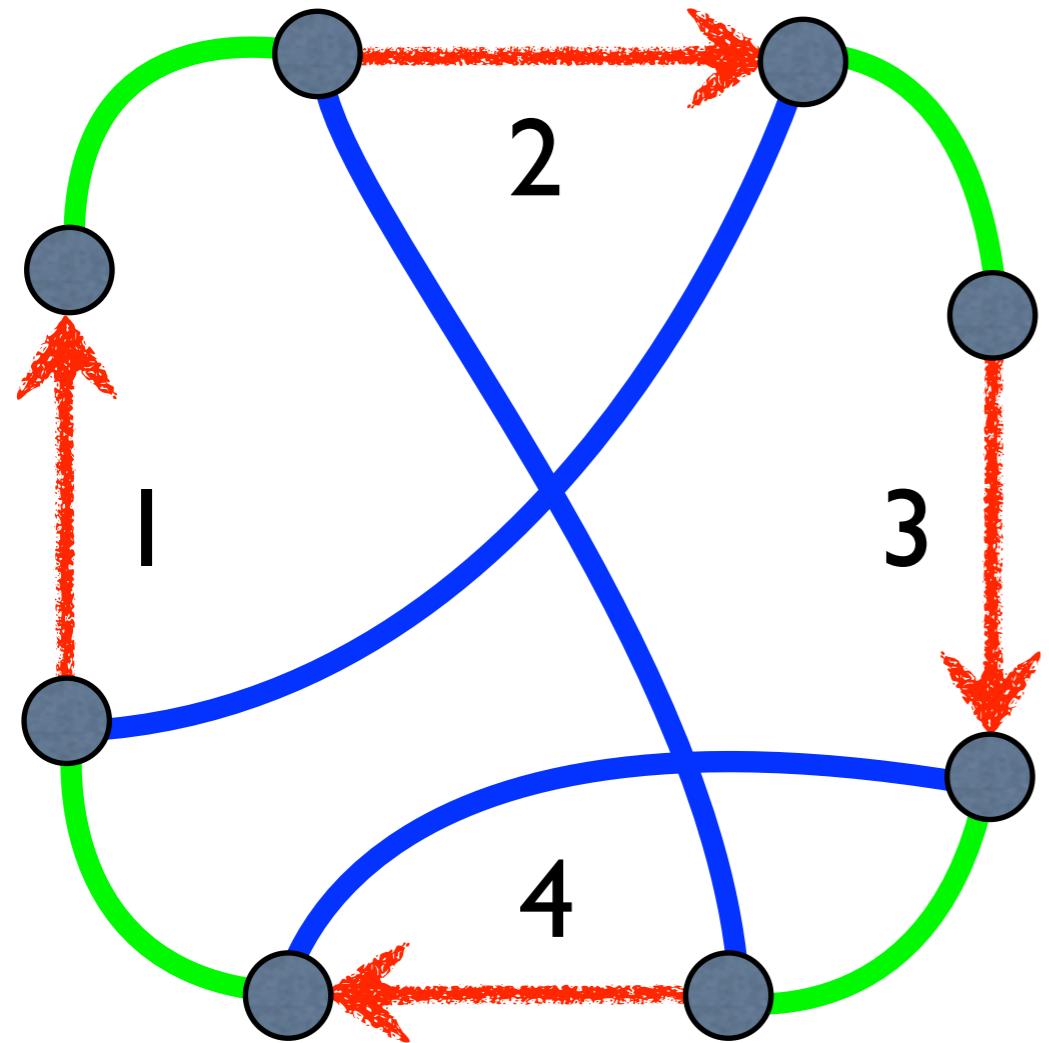


$B = (1, 3, -4, 2)$

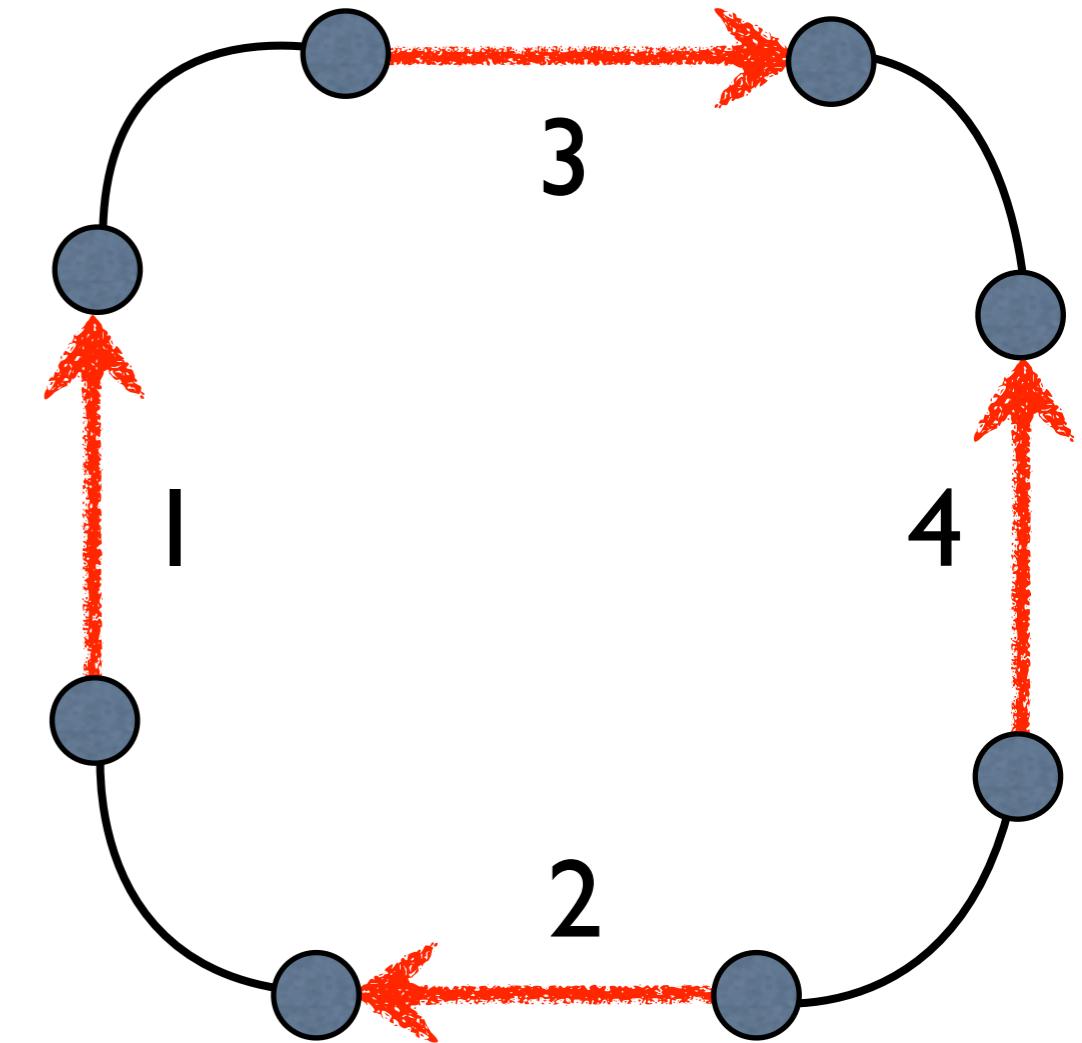


Breakpoint graph

$A = (1, 2, 3, 4)$

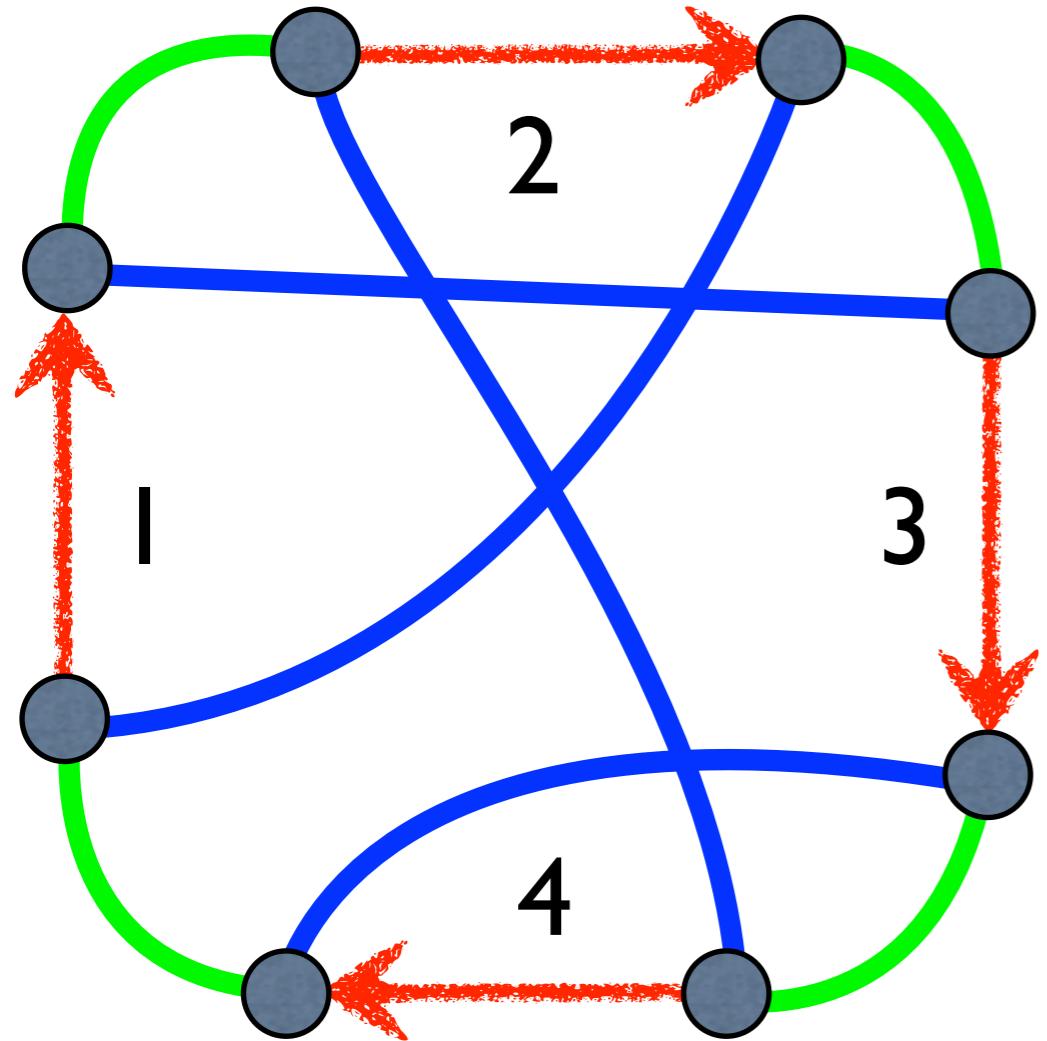


$B = (1, 3, -4, 2)$

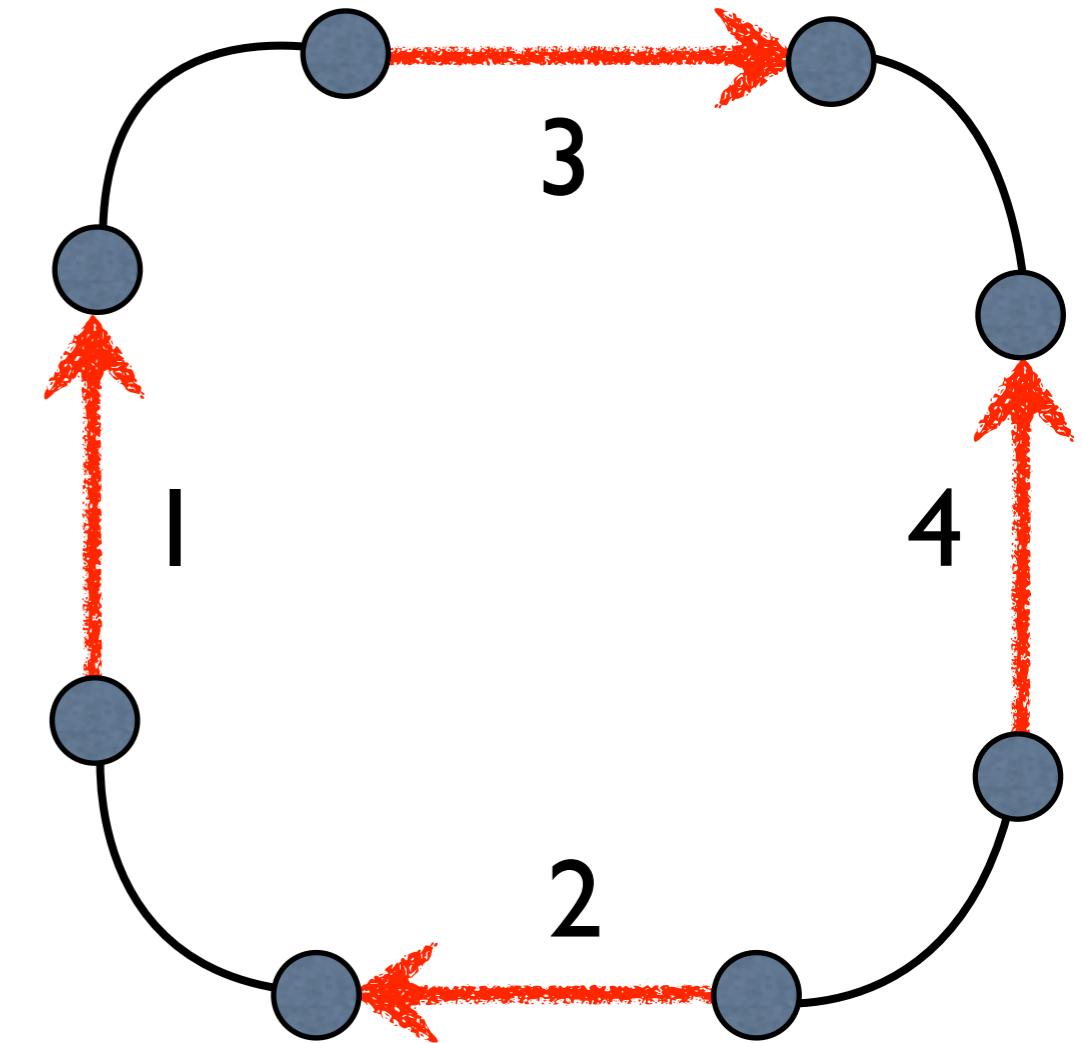


Breakpoint graph

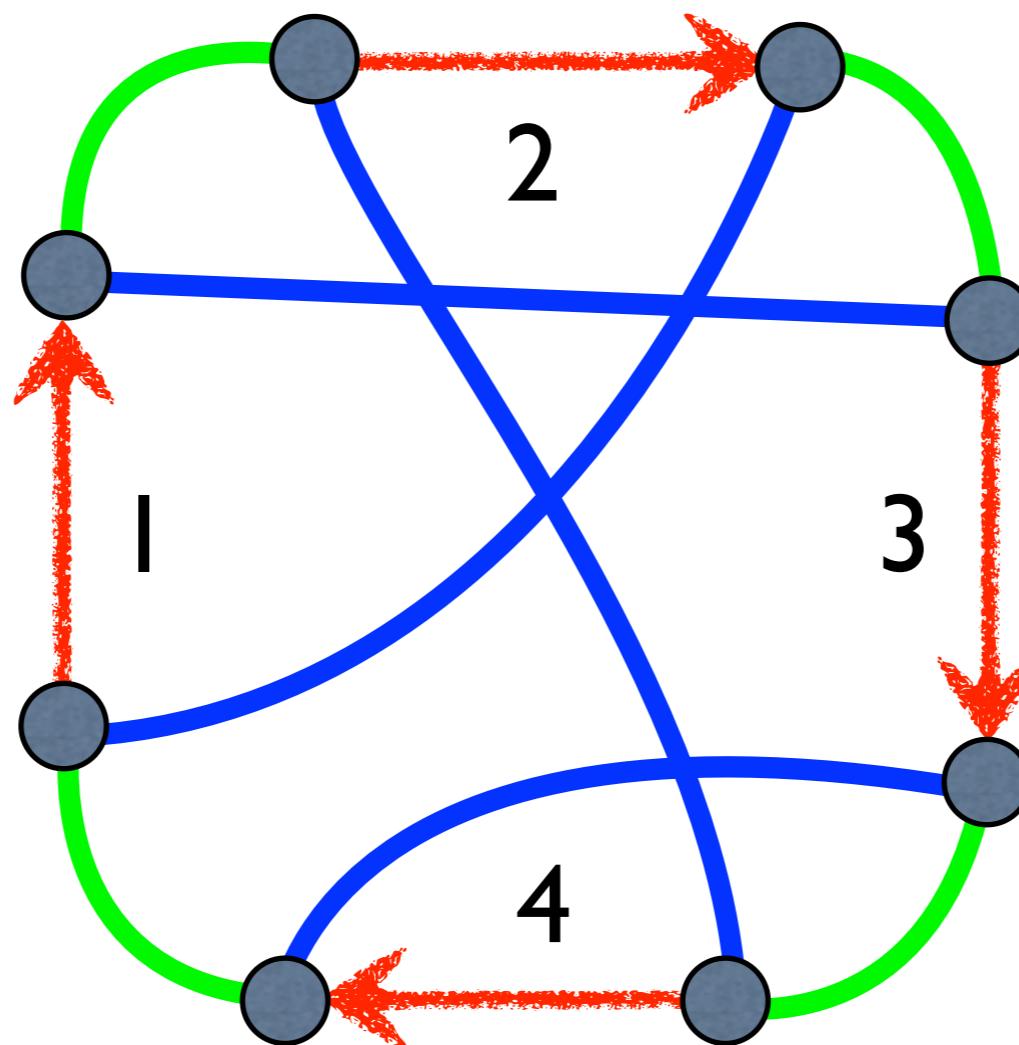
$A = (1, 2, 3, 4)$



$B = (1, 3, -4, 2)$

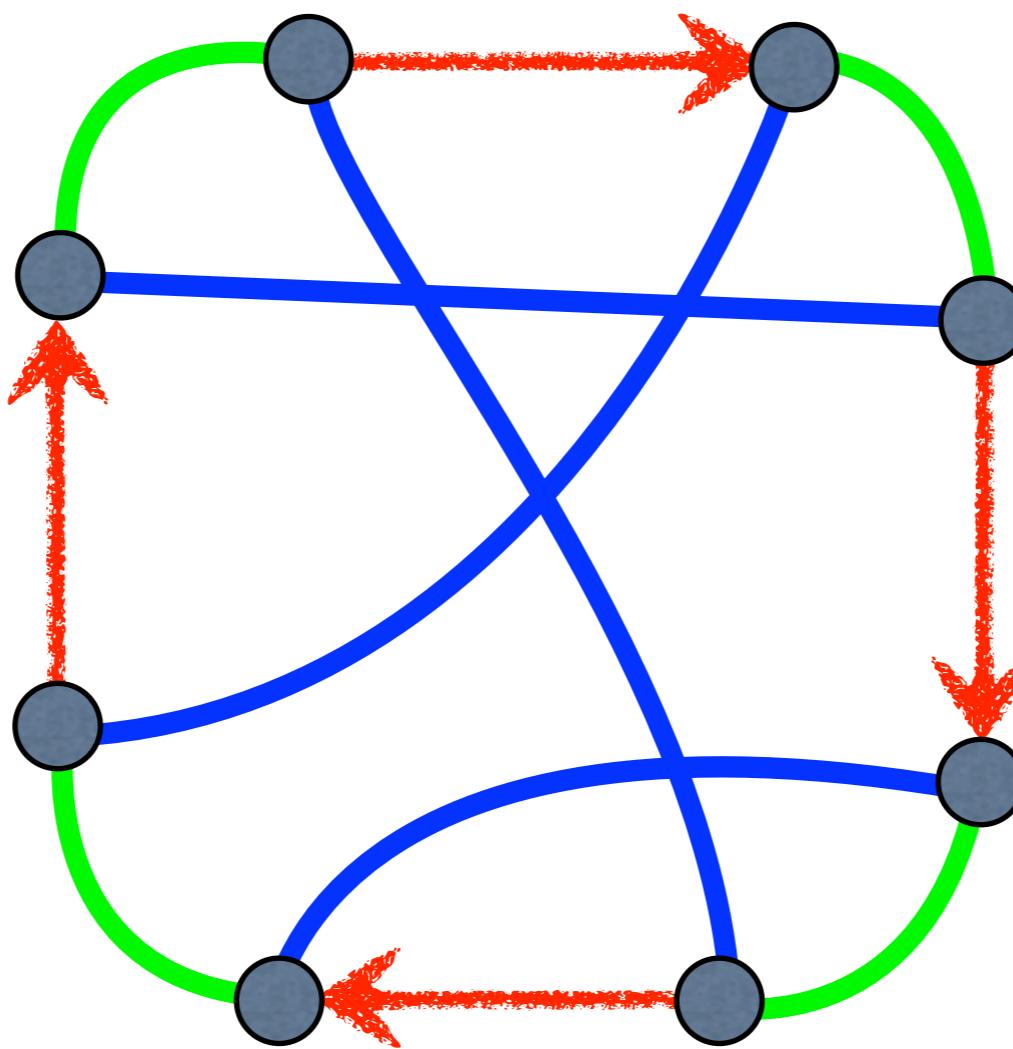


Breakpoint graph



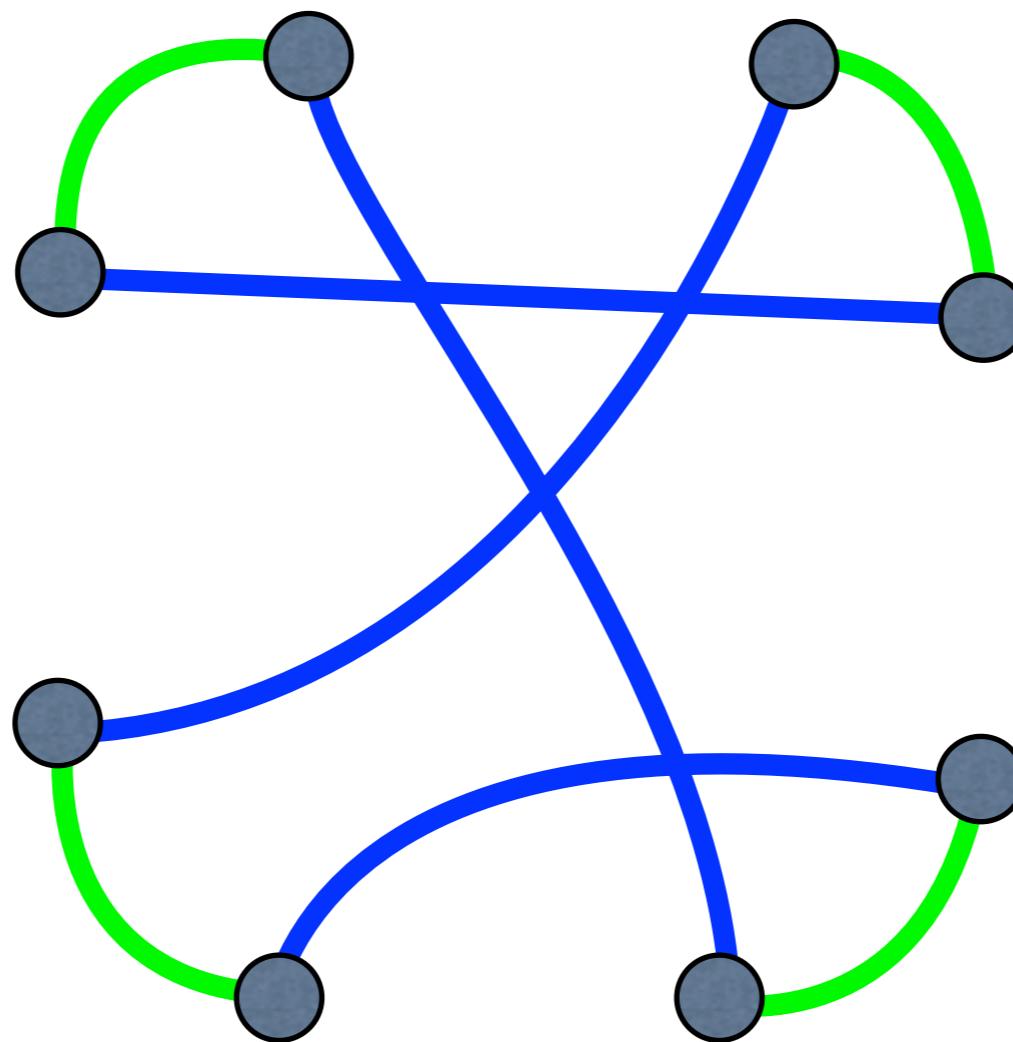
A
B

Breakpoint graph



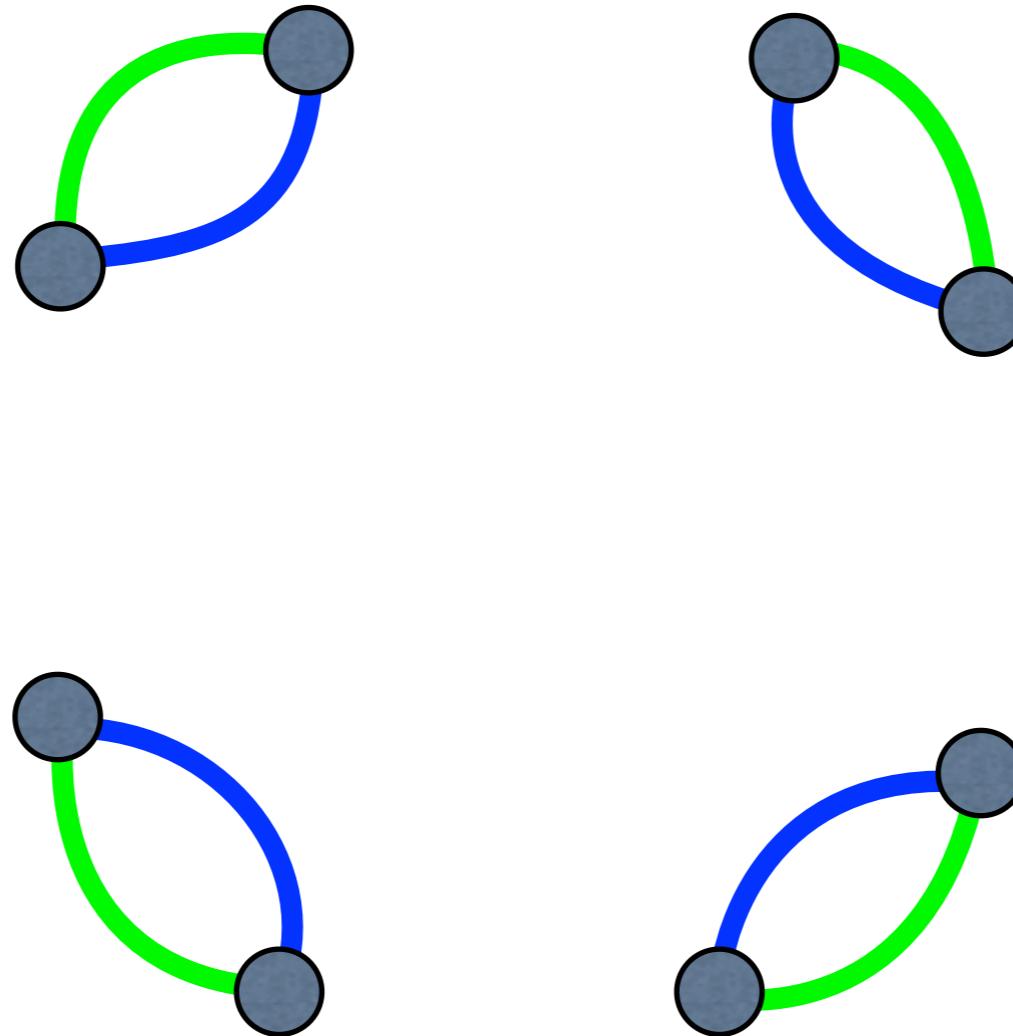
A
B

Breakpoint graph



A
B

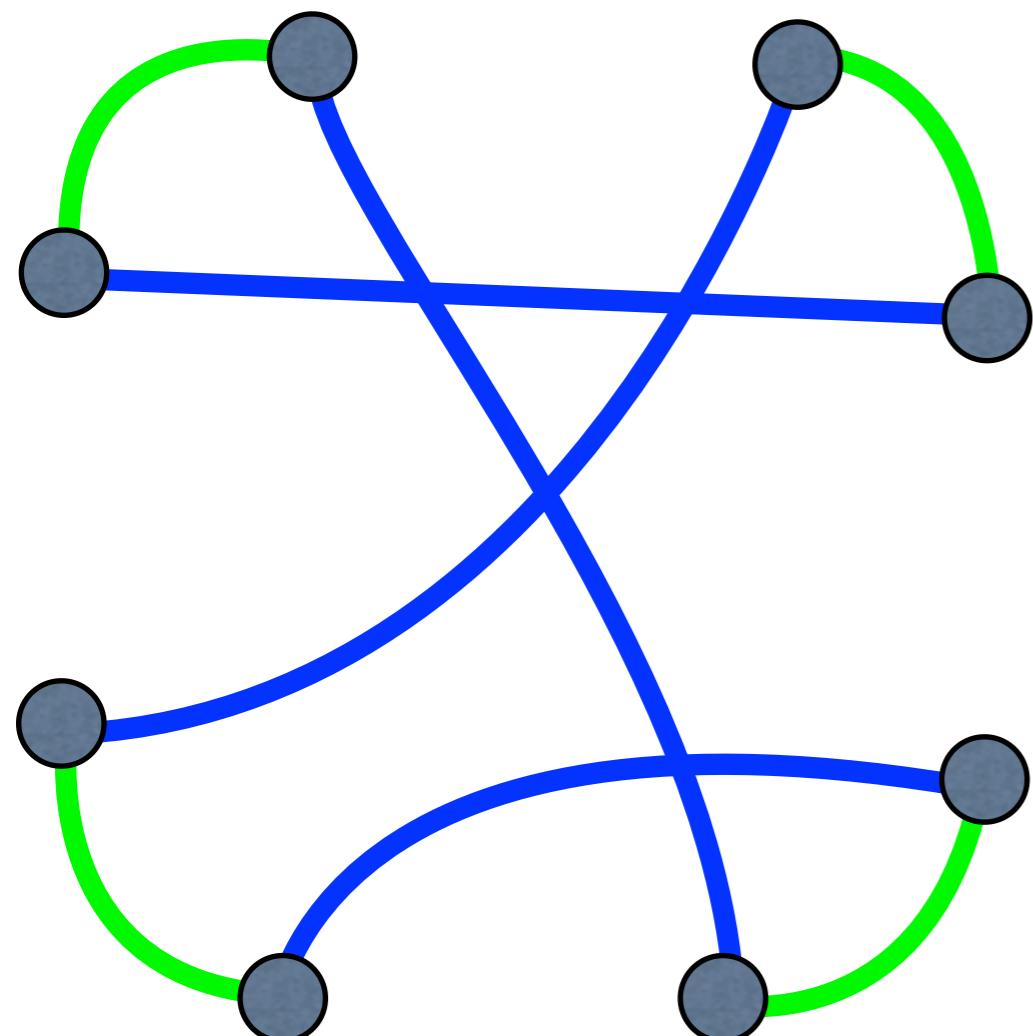
Breakpoint graph идентичных геномов



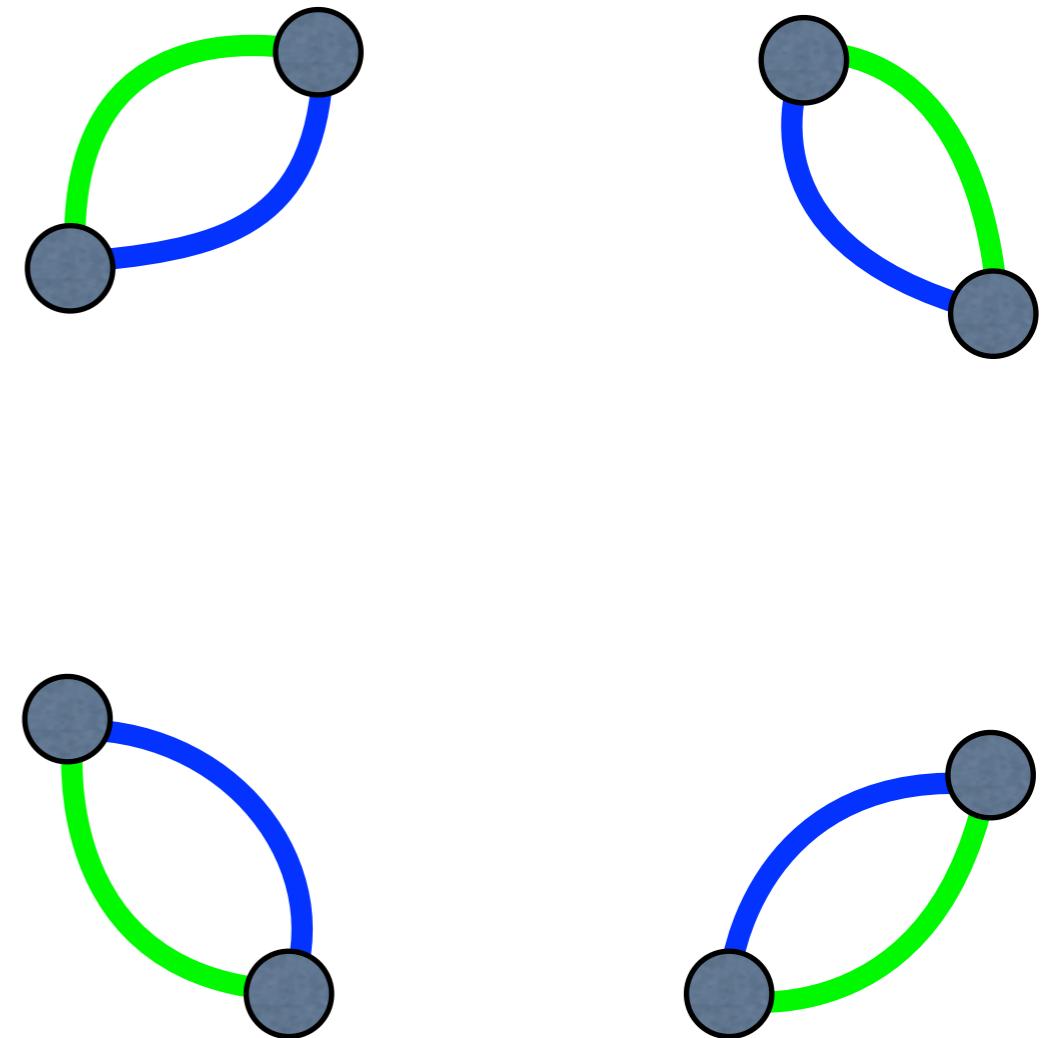
A —————
B —————

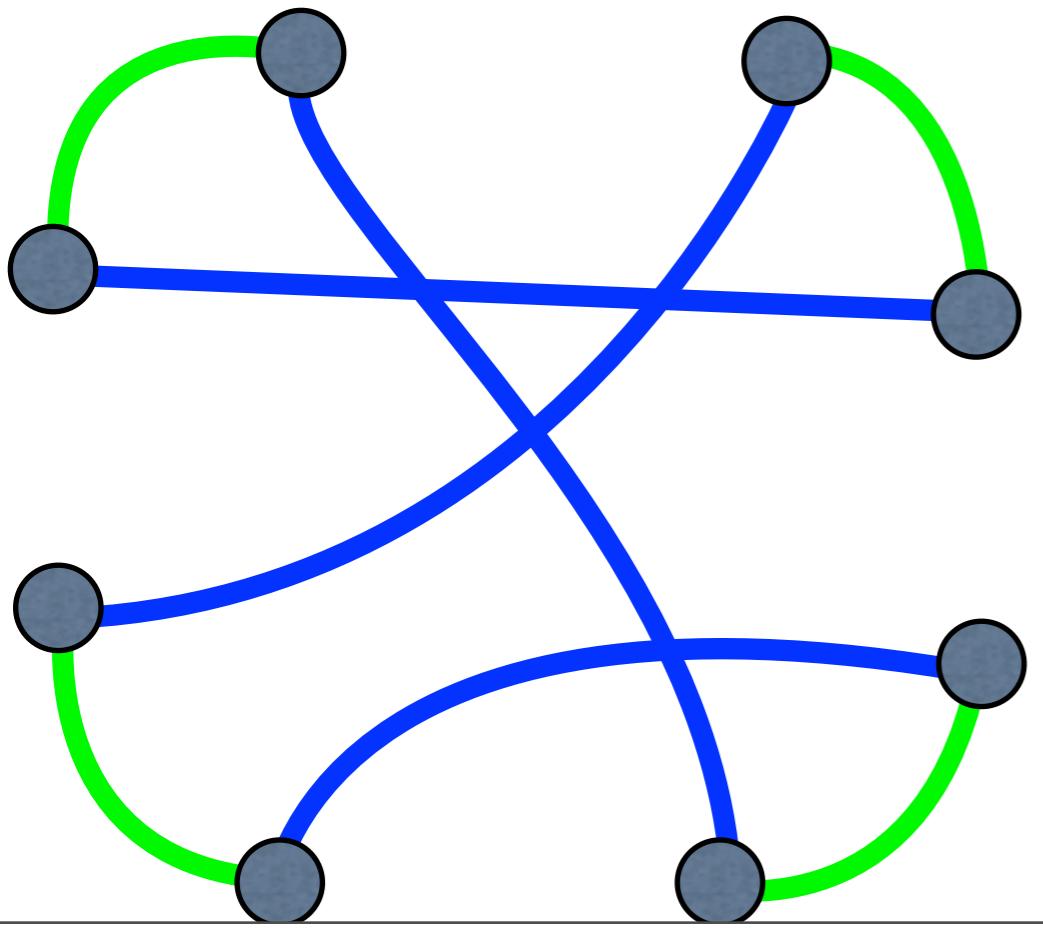
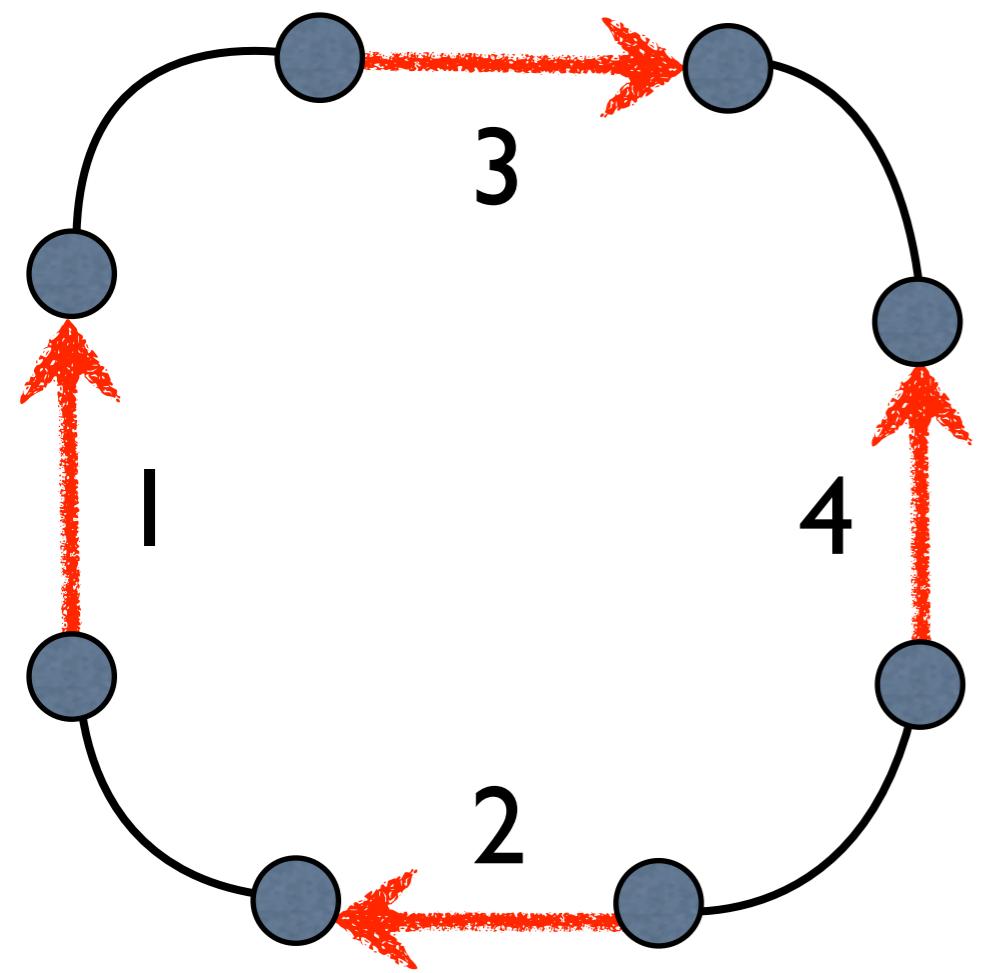
Breakpoint расстояние

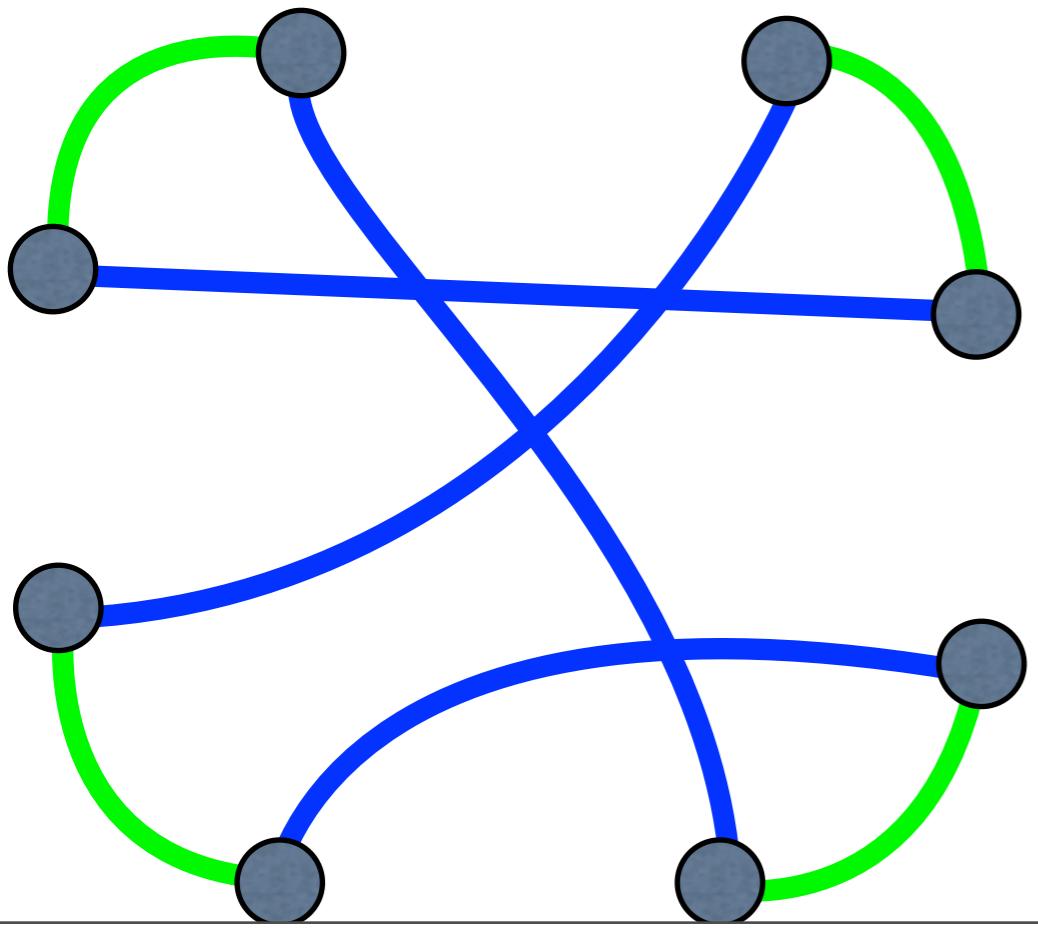
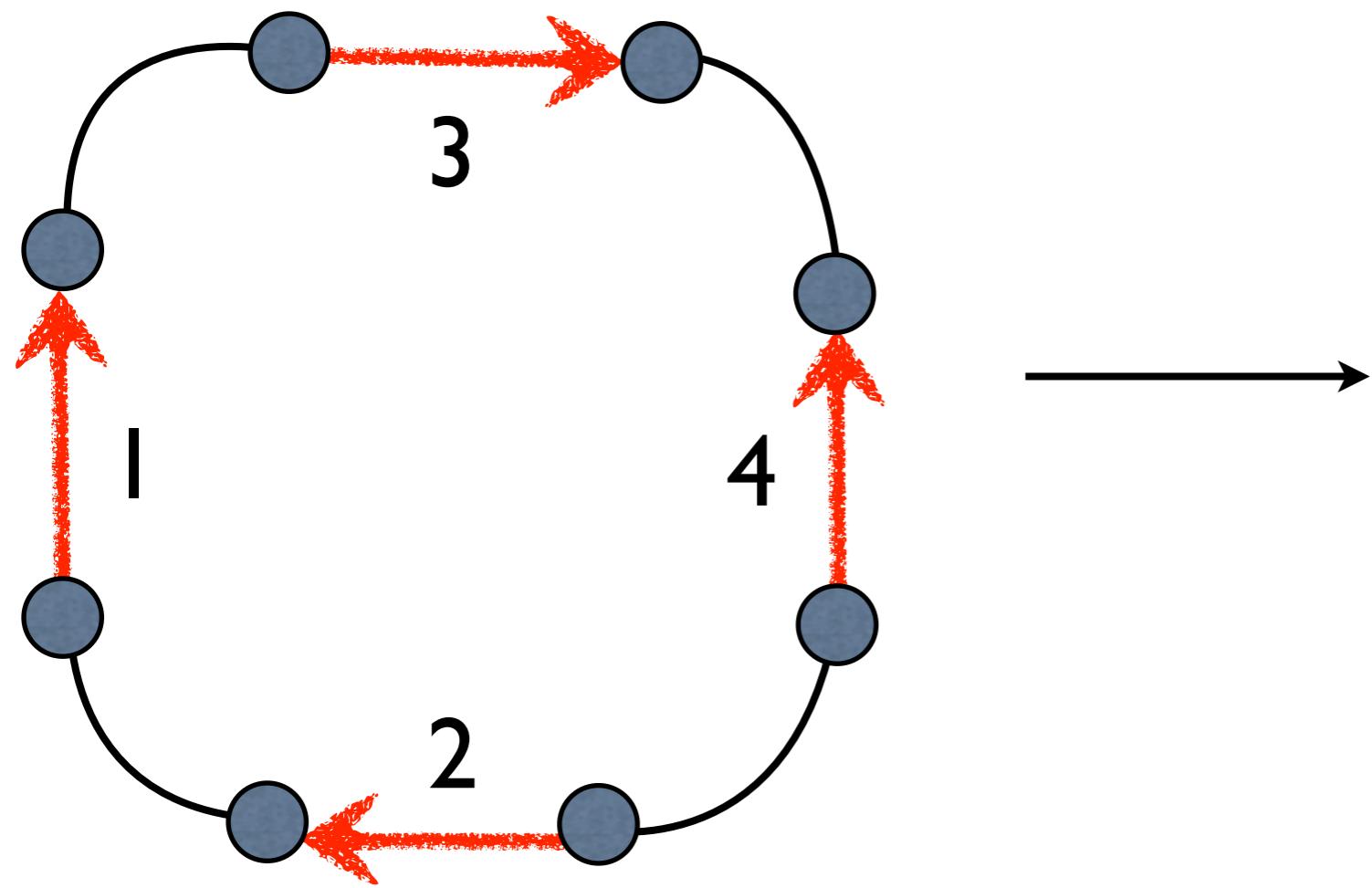
что есть

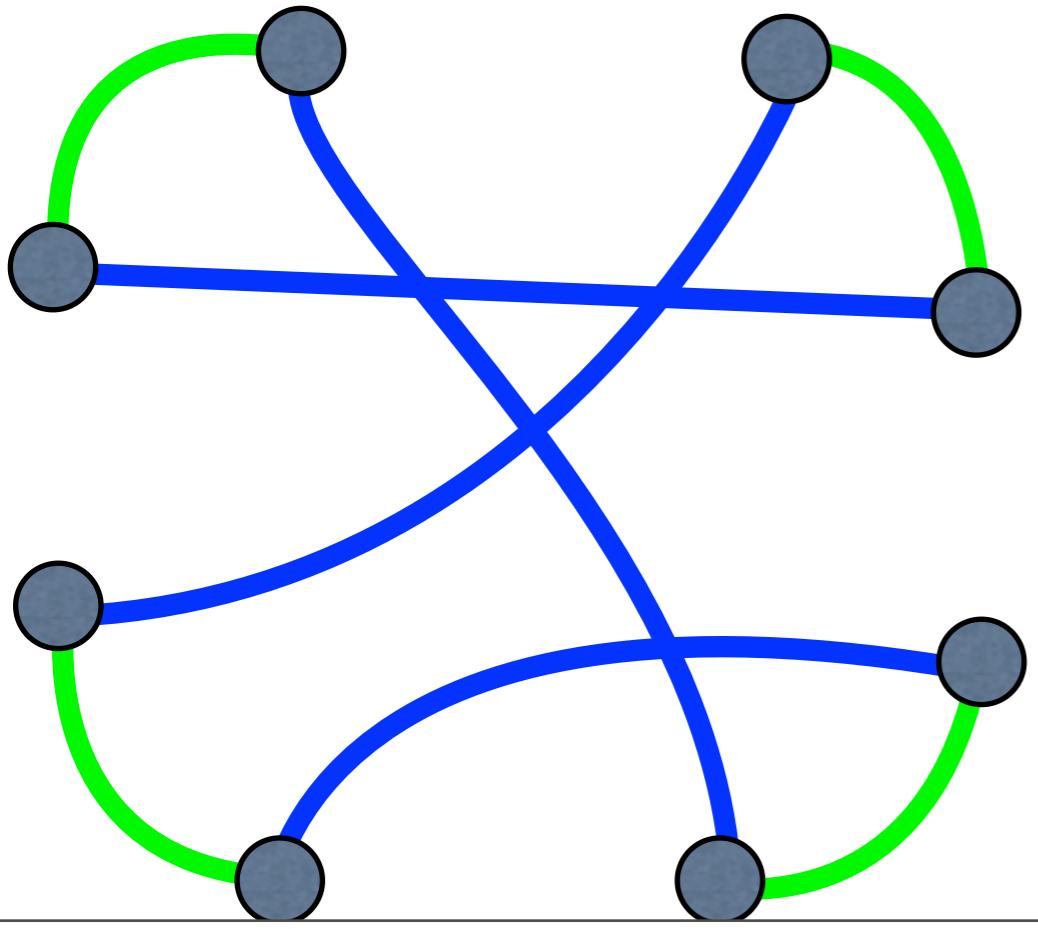
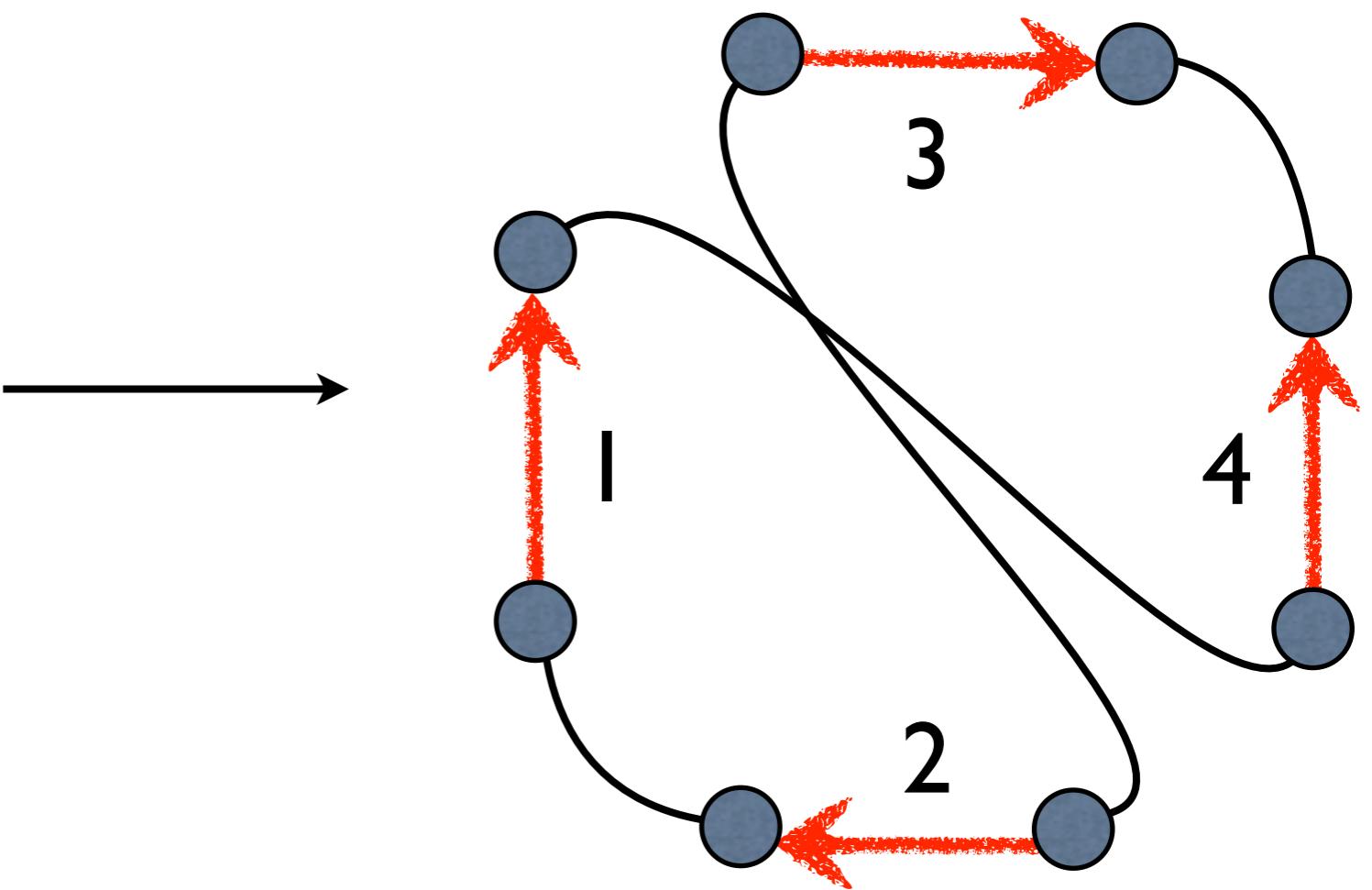
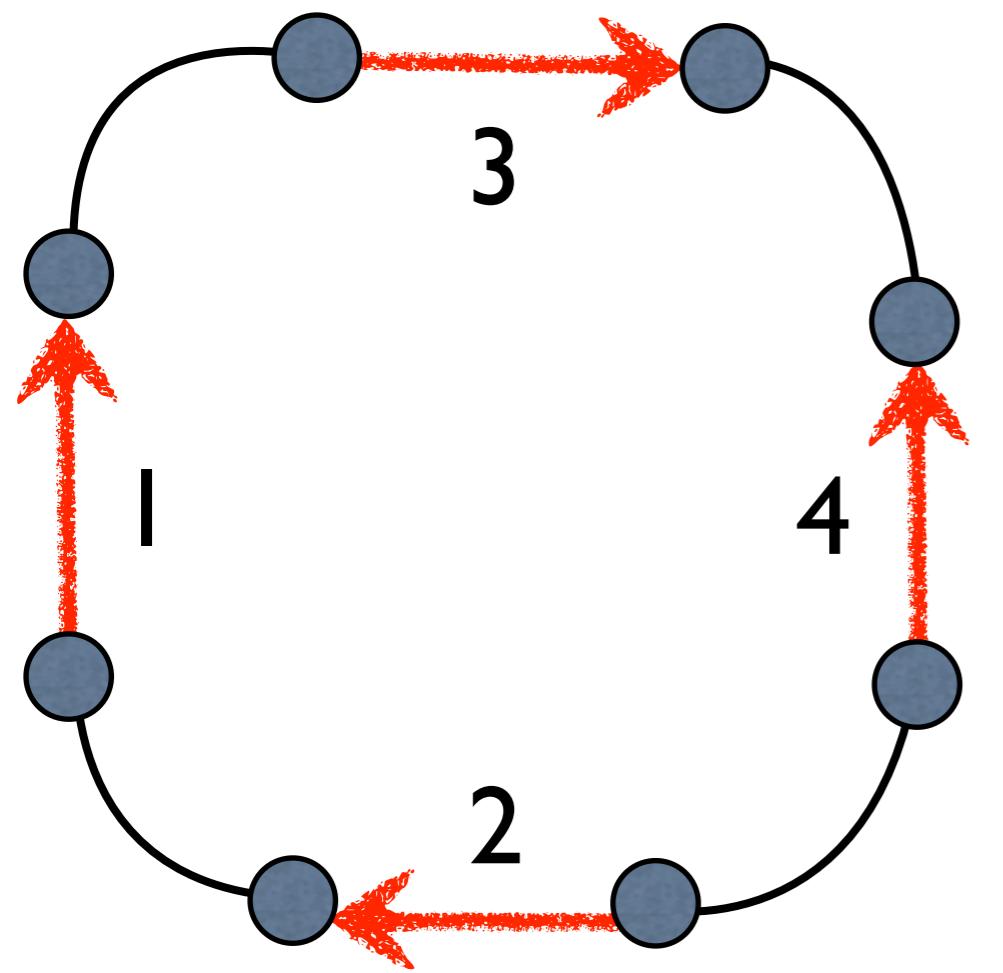


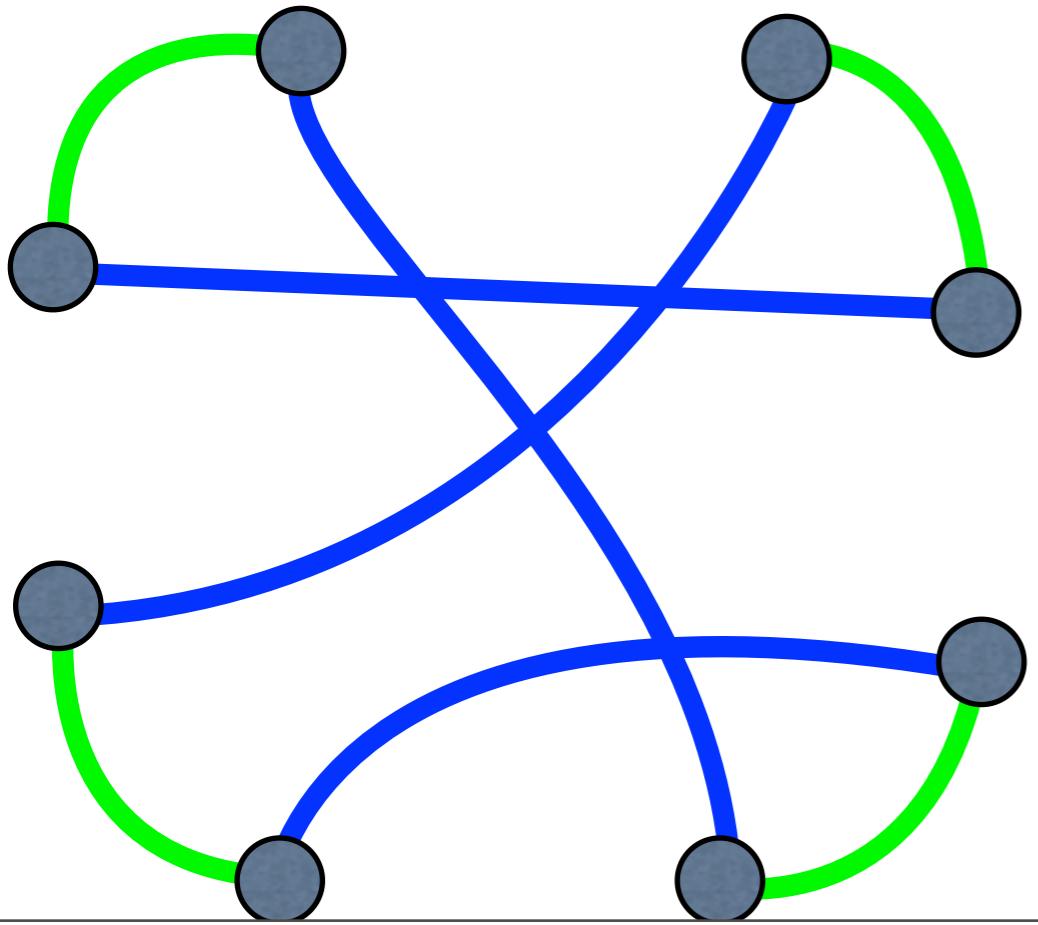
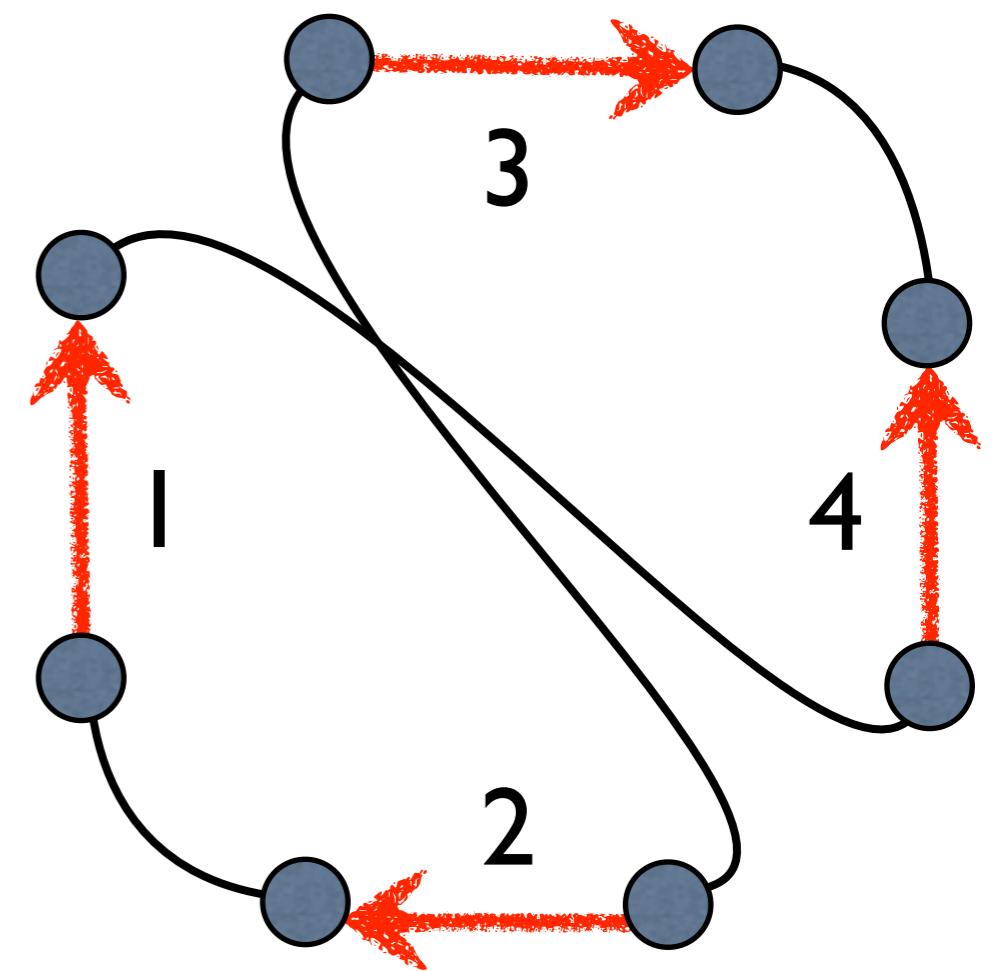
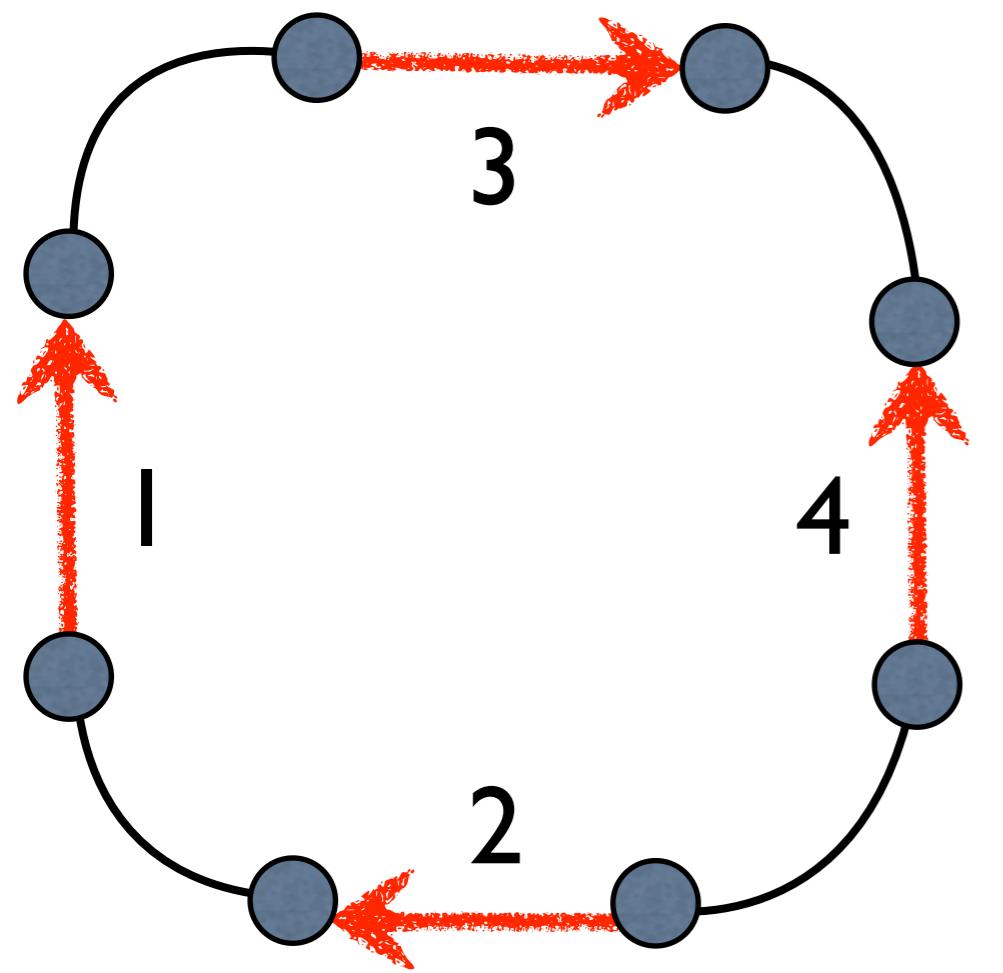
что хочется

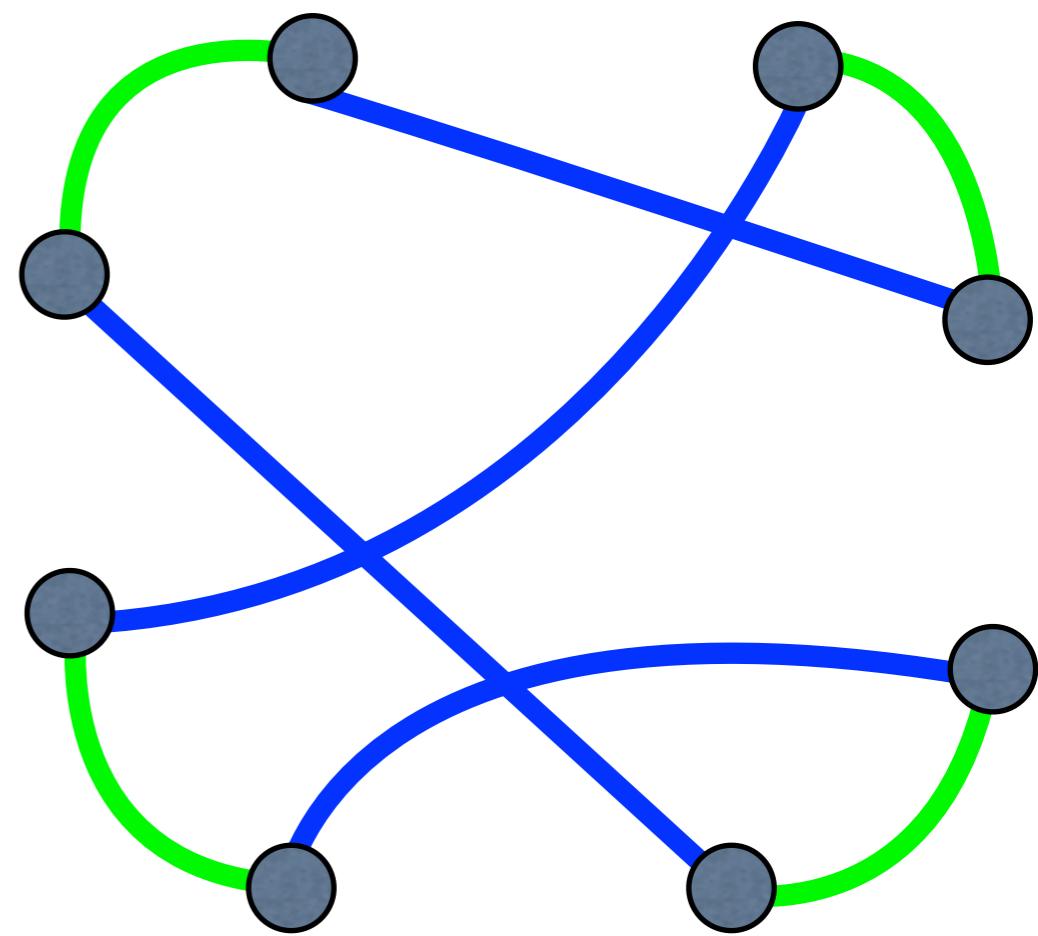
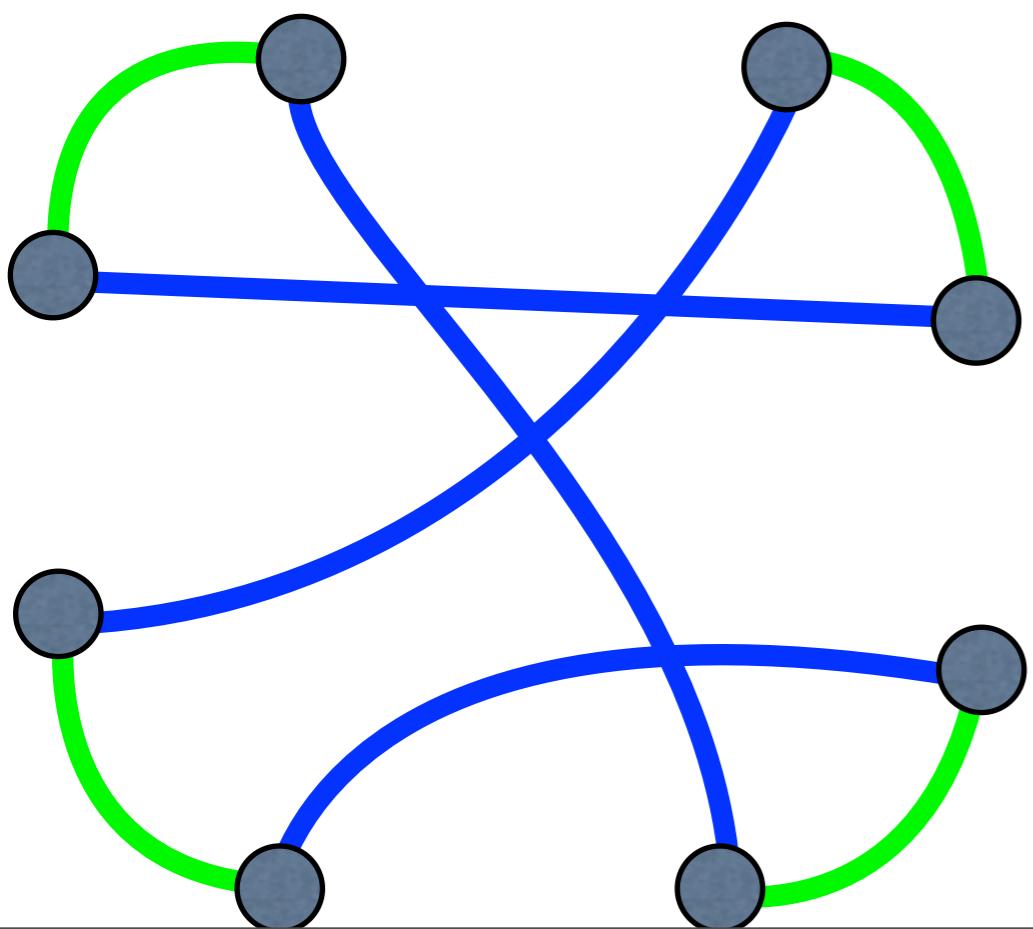
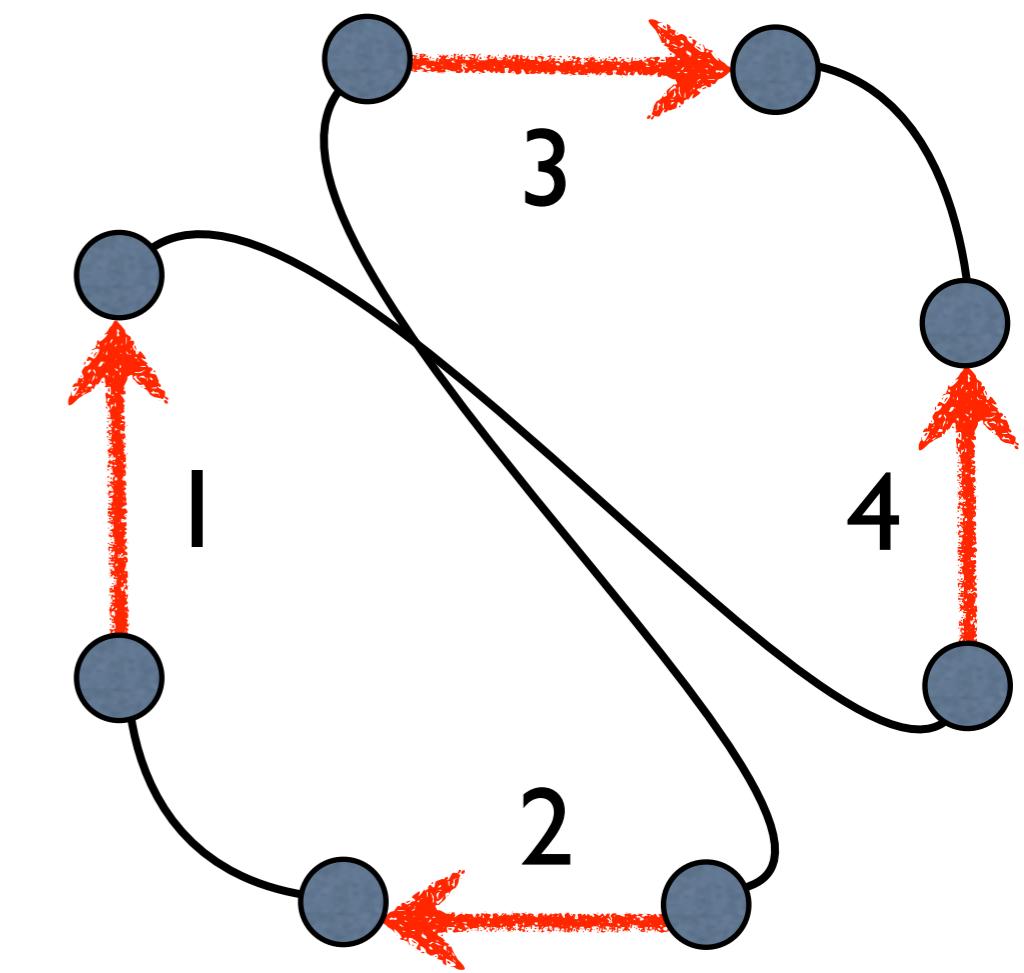
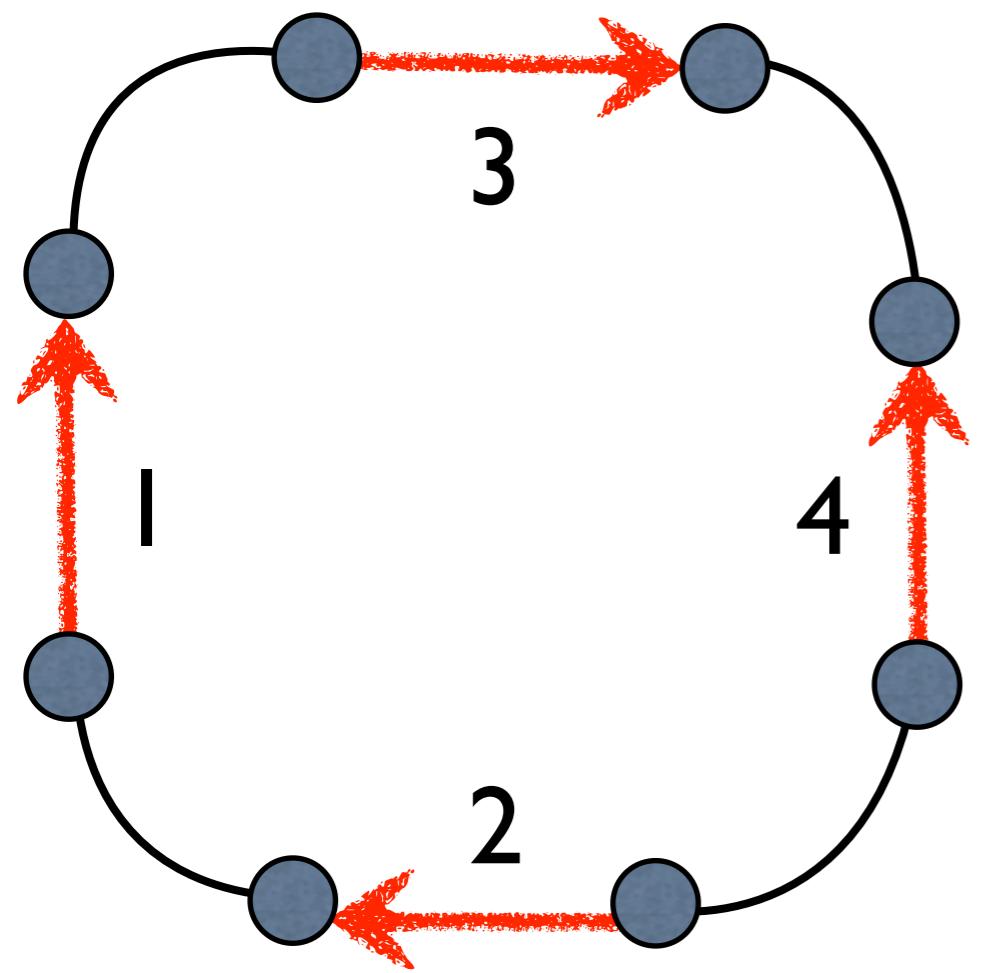












$$d_2(P, Q) = |P| - c(P, Q)$$

$|P|$ — количество блоков в геноме

$c(P, Q)$ — количество циклов в breakpoint graph

Идентичные гены и их breakpoint graph



$$d_2(P, Q) = |P| - c(P, Q)$$

$|P|$ — количество блоков в геноме

$c(P, Q)$ — количество циклов в breakpoint graph

Proof:

$$d_2(P, Q) = |P| - c(P, Q)$$

$|P|$ — количество блоков в геноме

$c(P, Q)$ — количество циклов в breakpoint graph

Proof:

При каждой 2-break исчезает минимум один цикл и может появится максимум 2 новых

$$d_2(P, Q) = |P| - c(P, Q)$$

$|P|$ — количество блоков в геноме

$c(P, Q)$ — количество циклов в breakpoint graph

Proof:

При каждой 2-break исчезает минимум один цикл и может появится максимум 2 новых

$$\implies \Delta cycles \leq 1$$

$$d_2(P, Q) = |P| - c(P, Q)$$

$|P|$ — количество блоков в геноме

$c(P, Q)$ — количество циклов в breakpoint graph

Proof:

При каждой 2-break исчезает минимум один цикл и может появится максимум 2 новых

$$\implies \Delta cycles \leq 1$$

$c(P, Q)$ — сколько есть циклов $|P|$ — сколько хочется циклов

$$d_2(P, Q) = |P| - c(P, Q)$$

$|P|$ — количество блоков в геноме

$c(P, Q)$ — количество циклов в breakpoint graph

Proof:

При каждой 2-break исчезает минимум один цикл и может появится максимум 2 новых

$$\implies \Delta cycles \leq 1$$

$c(P, Q)$ — сколько есть циклов $|P|$ — сколько хочется циклов

т. к. на при каждой операции число циклов может возрастать не более чем на 1, то наименьшее число операций равно разности исходного количества циклов и необходимого.

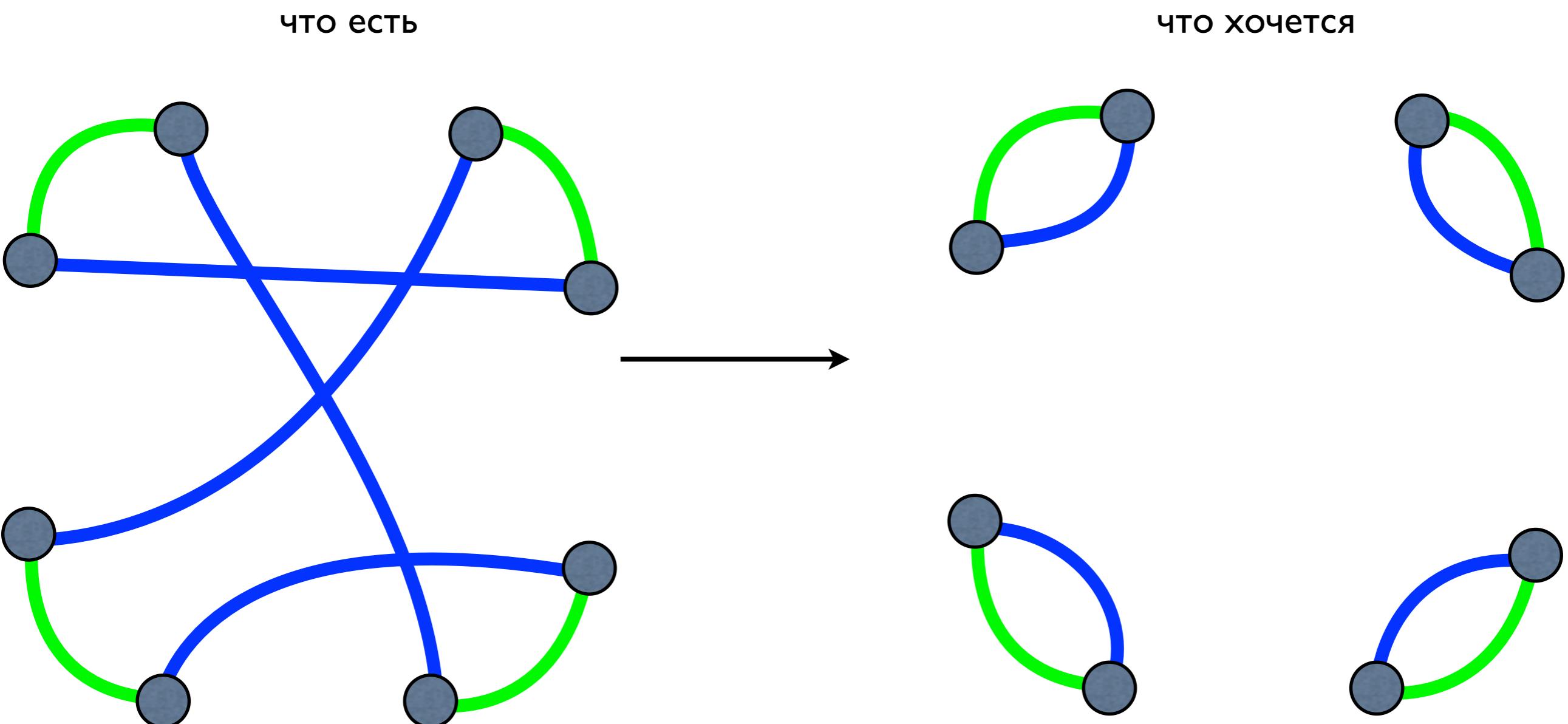


$$d_3(P, Q) = \frac{|P| - c^{odd}(P, Q)}{2}$$

$c^{odd}(P, Q)$ — количество циклов, содержащих нечетное число ребер одного цвета

$|P|$ — количество тривиальных циклов в breakpoint graph (P,P)

Breakpoint graph



$$d_3(P, Q) = \frac{|P| - c^{odd}(P, Q)}{2}$$

$c^{odd}(P, Q)$ — количество циклов, содержащих нечетное число ребер одного цвета

$|P|$ — количество тривиальных циклов в breakpoint graph(P, P)

Proof:

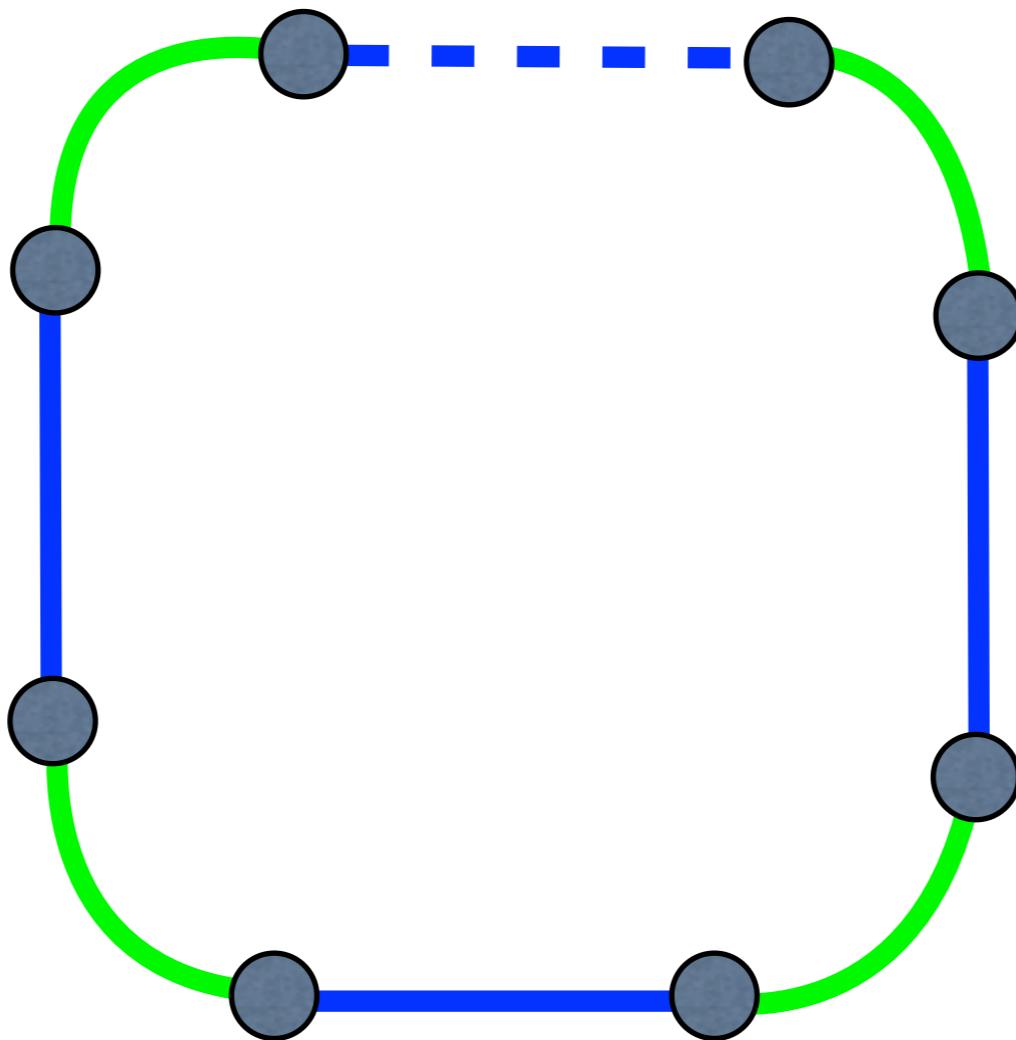
$$d_3(P, Q) = \frac{|P| - c^{odd}(P, Q)}{2}$$

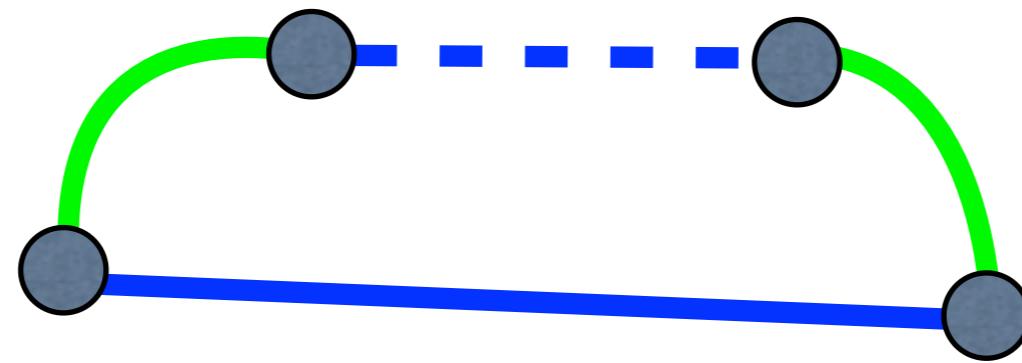
$c^{odd}(P, Q)$ – количество циклов, содержащих нечетное число ребер одного цвета

$|P|$ – количество тривиальных циклов в breakpoint graph(P, P)

Proof:

Если есть нетривиальный нечетный цикл, то он может быть разделен максимум на 3 нечетных цикла с помощью 3-break





$$d_3(P, Q) = \frac{|P| - c^{odd}(P, Q)}{2}$$

$c^{odd}(P, Q)$ — количество циклов, содержащих нечетное число ребер одного цвета

$|P|$ — количество тривиальных циклов в breakpoint graph(P, P)

Proof:

$$d_3(P, Q) = \frac{|P| - c^{odd}(P, Q)}{2}$$

$c^{odd}(P, Q)$ – количество циклов, содержащих нечетное число ребер одного цвета

$|P|$ – количество тривиальных циклов в breakpoint graph(P, P)

Proof:

Если есть нетривиальный нечетный цикл, то он может быть разделен максимум на 3 нечетных цикла с помощью 3-break

$$d_3(P, Q) = \frac{|P| - c^{odd}(P, Q)}{2}$$

$c^{odd}(P, Q)$ – количество циклов, содержащих нечетное число ребер одного цвета

$|P|$ – количество тривиальных циклов в breakpoint graph(P, P)

Proof:

Если есть нетривиальный нечетный цикл, то он может быть разделен максимум на 3 нечетных цикла с помощью 3-break

$$\Delta cycles^{odd} \leq 2$$

$$d_3(P, Q) = \frac{|P| - c^{odd}(P, Q)}{2}$$

$c^{odd}(P, Q)$ – количество циклов, содержащих нечетное число ребер одного цвета

$|P|$ – количество тривиальных циклов в breakpoint graph(P, P)

Proof:

Если есть нетривиальный нечетный цикл, то он может быть разделен максимум на 3 нечетных цикла с помощью 3-break

$$\Delta cycles^{odd} \leq 2$$

Если есть нетривиальный четный цикл, то он может быть разделен максимум на 2 нечетных и один четный циклы с помощью 3-break

$$d_3(P, Q) = \frac{|P| - c^{odd}(P, Q)}{2}$$

$c^{odd}(P, Q)$ – количество циклов, содержащих нечетное число ребер одного цвета

$|P|$ – количество тривиальных циклов в breakpoint graph(P, P)

Proof:

Если есть нетривиальный нечетный цикл, то он может быть разделен максимум на 3 нечетных цикла с помощью 3-break

$$\Delta cycles^{odd} \leq 2$$

Если есть нетривиальный четный цикл, то он может быть разделен максимум на 2 нечетных и один четный циклы с помощью 3-break

$$\Delta cycles^{odd} \leq 2$$

$$d_3(P, Q) = \frac{|P| - c^{odd}(P, Q)}{2}$$

$c^{odd}(P, Q)$ – количество циклов, содержащих нечетное число ребер одного цвета

$|P|$ – количество тривиальных циклов в breakpoint graph(P, P)

Proof:

Если есть нетривиальный нечетный цикл, то он может быть разделен максимум на 3 нечетных цикла с помощью 3-break

$$\Delta cycles^{odd} \leq 2$$

Если есть нетривиальный четный цикл, то он может быть разделен максимум на 2 нечетных и один четный циклы с помощью 3-break

$$\Delta cycles^{odd} \leq 2$$

Таким образом за каждый шаг число нечетных циклов может быть увеличено максимум на 2.



Другие расстояния

$$d_4(P, Q) = \left\lceil \frac{|P| - c_4^1(P, Q) - \lfloor c_4^2(P, Q)/2 \rfloor}{3} \right\rceil$$

$$d_5(P, Q) = \left\lceil \frac{|P| - s_5(P, Q)}{4} \right\rceil$$

где

$$s_5(P, Q) = c_5^1(P, Q) + \min \{c_5^2(P, Q), c_5^3(P, Q)\} + \left\lfloor \frac{\max\{0, c_5^3(P, Q) - c_5^2(P, Q)\}}{3} \right\rfloor$$

Другие расстояния

$$d_k(P, Q) = \left\lceil \frac{|P| - s_k(P, Q)}{k - 1} \right\rceil$$

для $k = 10$

$s_{10}(P, Q)$ содержит более 1000 слагаемых

Сложность

M. Alekseyev, P. Pevzner
“Multi-break rearrangements and chromosomal evolution”
Theoretical Computer Science 395(2-3) (2008), pp. 193-202.

Сложность

$$O(n^{k-2}) + O(n)$$

M. Alekseyev, P. Pevzner
“Multi-break rearrangements and chromosomal evolution”
Theoretical Computer Science 395(2-3) (2008), pp. 193-202.

Сложность

$$O(n^{k-2}) + O(n)$$



$$O(n^{\lfloor k/2 \rfloor - 2}) + O(n)$$

M. Alekseyev, P. Pevzner

“Multi-break rearrangements and chromosomal evolution”
Theoretical Computer Science 395(2-3) (2008), pp. 193-202.

Сложность

$$O(n^{k-2}) + O(n)$$



$$O(n^{\lfloor k/2 \rfloor - 2}) + O(n)$$



$$O(n)^*$$

M. Alekseyev, P. Pevzner

“Multi-break rearrangements and chromosomal evolution”
Theoretical Computer Science 395(2-3) (2008), pp. 193-202.

Сложность

$$O(n^{k-2}) + O(n)$$



$$O(n^{\lfloor k/2 \rfloor - 2}) + O(n)$$



$$O(n)^*$$

* — процесс работает за время
экспоненциальное от k

M. Alekseyev, P. Pevzner

“Multi-break rearrangements and chromosomal evolution”
Theoretical Computer Science 395(2-3) (2008), pp. 193-202.

Breakpoint re-use

$$d_3(Human, Mouse) = 139$$

Если нет переиспользования, тогда будет всего $139 \times 3 = 417$ разрывов. 417 соединяющих связей будут разорваны. Причем каждый раз эти связи будут различны (в свете предположения)

Всего связей для разрывов 281, следовательно какие-то из них точно переиспользуются.

Breakpoint re-use

$$d_3(Human, Mouse) = 139$$

Если нет переиспользования, тогда будет всего $139 \times 3 = 417$ разрывов. 417 соединяющих связей будут разорваны. Причем каждый раз эти связи будут различны (в свете предположения)

Всего связей для разрывов 281, следовательно какие-то из них точно переиспользуются.



Breakpoint re-use

$$d_3(Human, Mouse) = 139$$

Если нет переиспользования, тогда будет всего $139 \times 3 = 417$ разрывов. 417 соединяющих связей будут разорваны. Причем каждый раз эти связи будут различны (в свете предположения)

Всего связей для разрывов 281, следовательно какие-то из них точно переиспользуются.

Но!!!

2-break частный случай 3-break

Breakpoint re-use

Рассмотрим такую серию k -break, что геном P переводится в геном Q с минимальным количеством разрывов

Breakpoint re-use

Рассмотрим такую серию k -break, что геном P переводится в геном Q с минимальным количеством разрывов

Теорема:

любая серия k -break переводящая P в Q делает по меньшей мере $d_k(P, Q) + d_2(P, Q)$ разрывов

Max A. Alekseyev and Pavel A. Pevzner “Are there rearrangements hotspots in the Human genome?”. *PLoS Computational Biology* 3(11) (2007): e209

Breakpoint re-use

Рассмотрим такую серию k -break, что геном P переводится в геном Q с минимальным количеством разрывов

Теорема:

любая серия k -break переводящая P в Q делает по меньшей мере $d_k(P, Q) + d_2(P, Q)$ разрывов

Max A. Alekseyev and Pavel A. Pevzner “Are there rearrangements hotspots in the Human genome?”. *PLoS Computational Biology* 3(11) (2007): e209

Следствие:

существует такая серия 3-break, которая переводит геном P в геном Q за $d_3(P, Q) + d_2(P, Q)$

Breakpoint re-use

Breakpoint re-use

$$d_2(Human, Mouse) = 246$$

Breakpoint re-use

$$d_2(\textit{Human}, \textit{Mouse}) = 246$$

$$d_3(\textit{Human}, \textit{Mouse}) = 139$$

Breakpoint re-use

$$d_2(\text{Human}, \text{Mouse}) = 246$$

$$d_3(\text{Human}, \text{Mouse}) = 139$$

минимальное количество разрывов, необходимо для перевода генома Человека в геном Мыши

Breakpoint re-use

$$d_2(\text{Human}, \text{Mouse}) = 246$$

$$d_3(\text{Human}, \text{Mouse}) = 139$$

минимальное количество разрывов, необходимо для перевода генома Человека в геном Мыши

$$d_2(\text{Human}, \text{Mouse}) + d_3(\text{Human}, \text{Mouse}) = 246 + 139 = 385$$

Breakpoint re-use

$$d_2(\text{Human}, \text{Mouse}) = 246$$

$$d_3(\text{Human}, \text{Mouse}) = 139$$

минимальное количество разрывов, необходимо для перевода генома Человека в геном Мыши

$$d_2(\text{Human}, \text{Mouse}) + d_3(\text{Human}, \text{Mouse}) = 246 + 139 = 385$$

$$385 >> 281$$

Спасибо за внимание!