Институт проблем передачи информации

Российской академии наук

Лаборатория «**Математических методов и моделей в биоинформатике**»

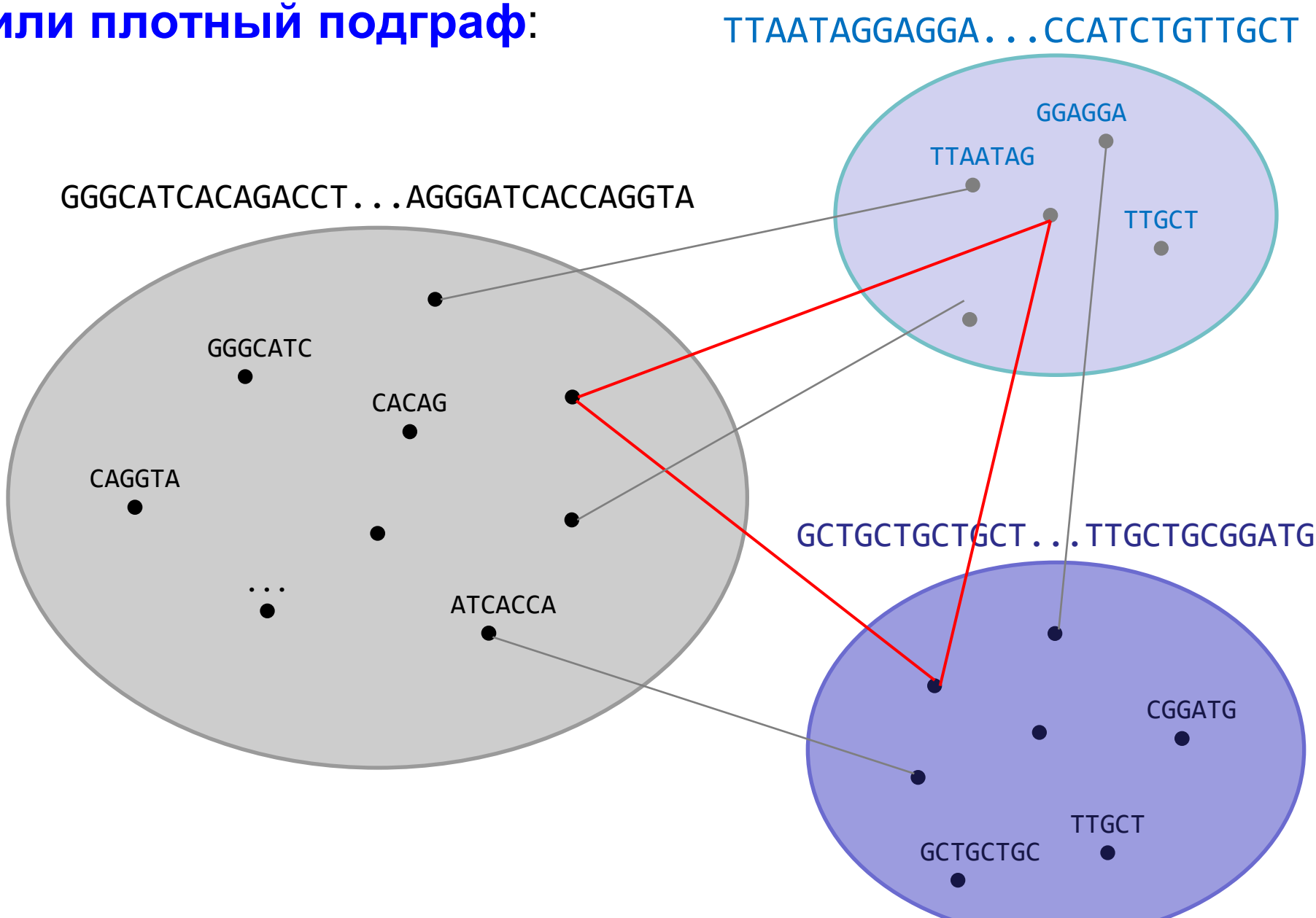
сайт лаборатории: **http://lab6.iitp.ru/**

адрес: **lyubetsk@iitp.ru**

В. Любецкий, К. Горбунов, А. Селиверстов

ОПТИМИЗАЦИЯ НА ГРАФАХ

Типичная задача: дан **многодольный граф**, ребро – близость или иная характеристика пар слов. В нём ищем **клику или плотный подграф**:



Какого размера многодольные графы реально возникают в прикладных задачах? **Число долей 10^5**, длина последовательности **10^9**, в ней берутся короткие слова, т.е. вершин в доле 10^9: всего **вершин и степень каждой 10^{14}**.

Найти слова, которые встречаются с заданной точностью в m=**10^5** последовательностях, с указанием их позиций в ней.

Точность слов, например, 6 редакционных операций при их длине 150. **Построение графа включается в задачу**, так как исходно даны **только все последовательности**.

Трудоёмкость свыше **10^{28}; при этом нужна общая память – более терабайтного объёма**.

Даже при этих скромных оценках *нельзя решать задачу с использованием современных суперкомпьютеров*.

1) Задача кластеризации в многодольном графе

состоит в разбиении графа на кластеры: в кластере как можно **меньше вершин из одной доли**, а как можно **больше долей в кластере**.

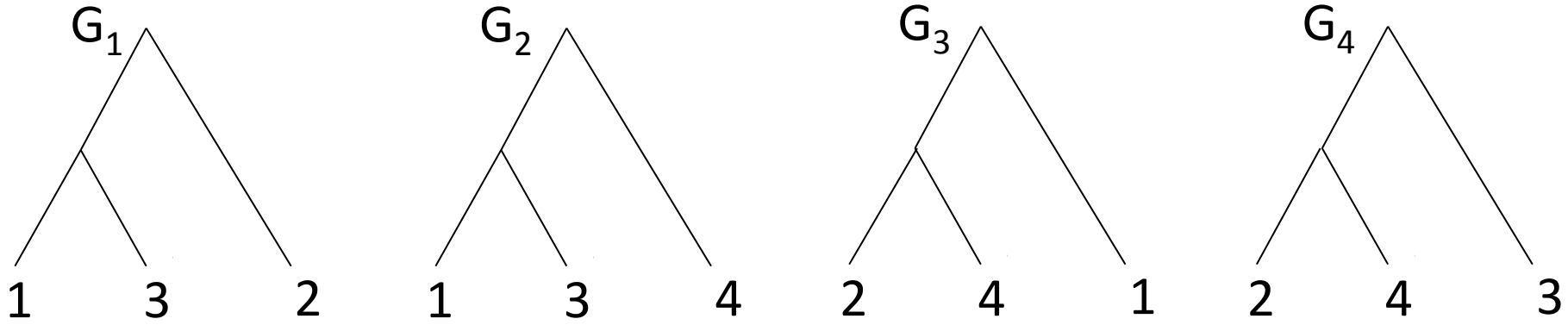
Новая трудность – задача 2х-параметрическая!

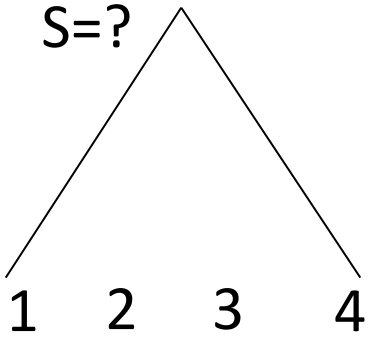
У нас есть алгоритм, довольно эффективно решающий эту задачу; на основе построения остовного дерева (что проблема). Нужно лучше, это – задача.

---

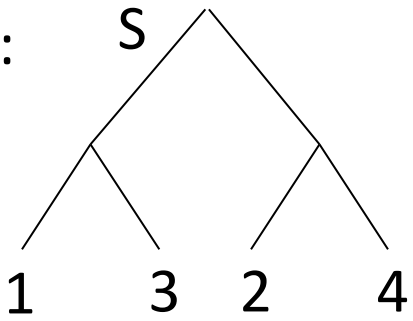
2) Задача согласования большого числа деревьев в единое дерево

«**Среднее**» дерево = **согласование** деревьев. Для хорошо обусловленных данных решение однозначно:

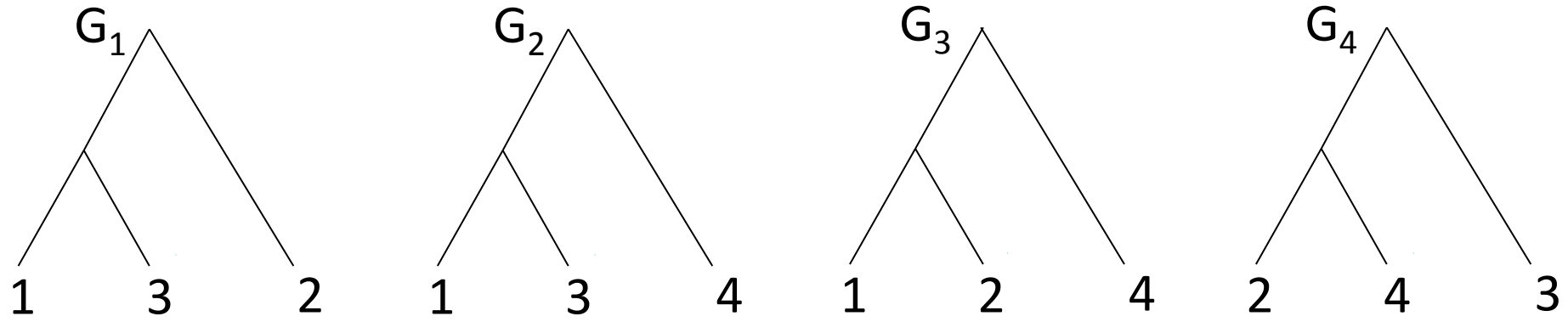


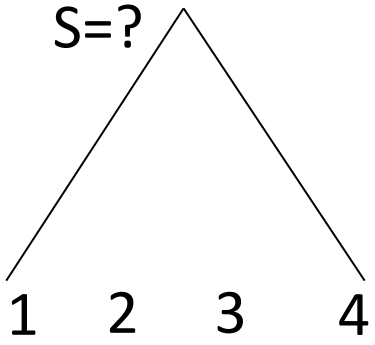


Решение:

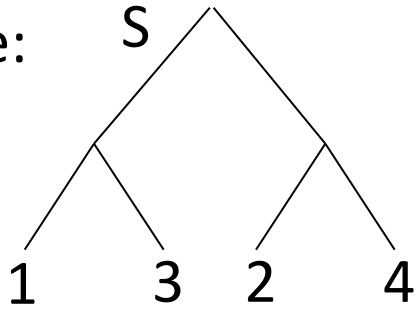


Согласование деревьев. Для плохо обусловленных данных решение неоднозначно:

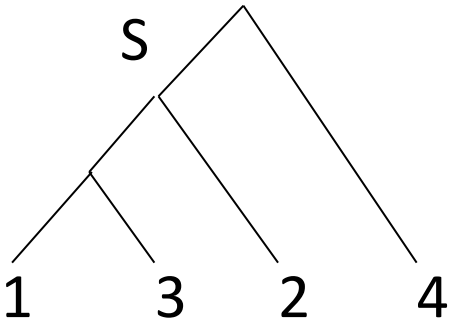




Решение:

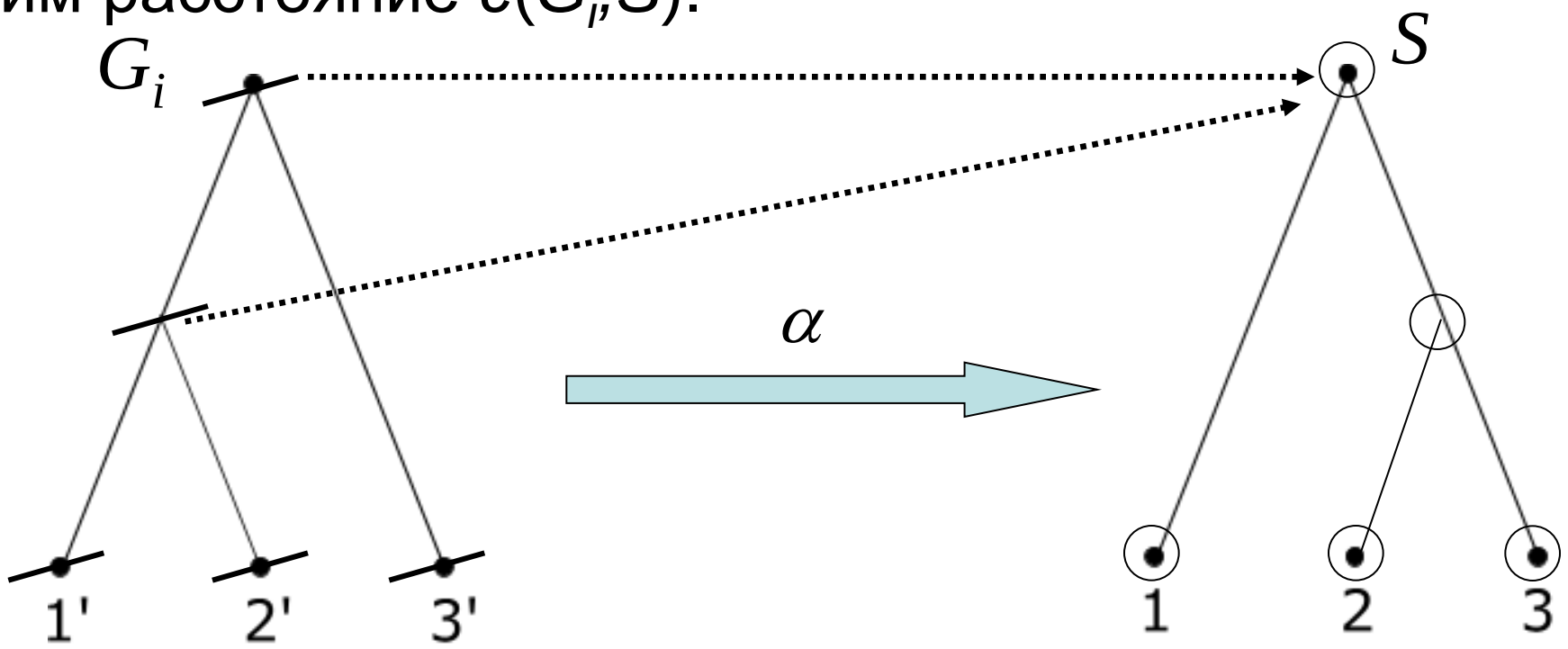


или



Прежде всего, определим **расстояние** от дерева до дерева. Затем **минимизируем** $\sum_i c(G_i, S)$ **сумму расстояний** в пространстве всех деревьев, которое пробегает искомое S .

Итак, определим отображение α и по нему вычислим расстояние $c(G_i,S)$.



$c(G_i,S) = 2$

Нами получен **кубический** алгоритм, который при некоторых условиях **точно!** находит такой минимум.

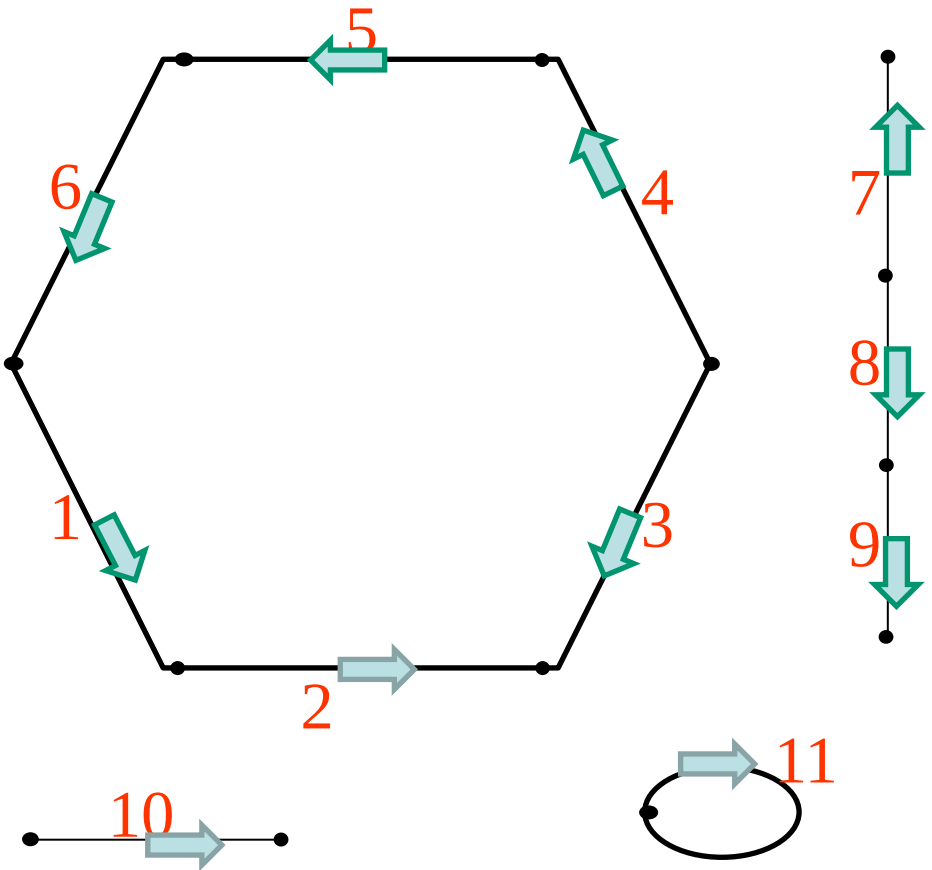
3) Редакционное расстояние между CC-графами.

Сначала напомним это **расстояние** для **последовательностей**, оно широко используется.

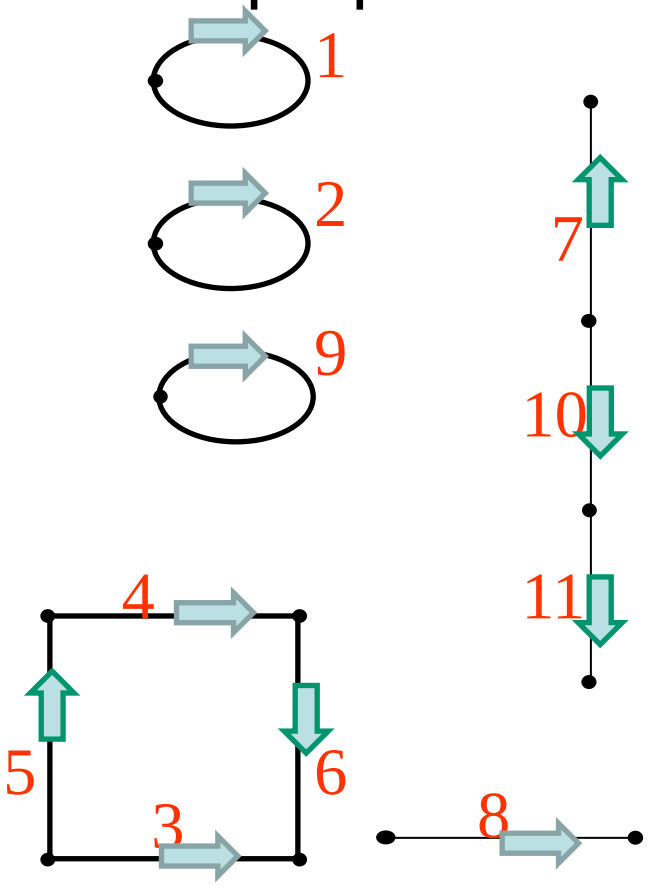
Типичный пример:

Содер-ое понятие расстояния между графами появилось недавно! Фиксирован набор операций, преобразующих один граф в другой. Найти **расстояние – наименьшую длину последовательности операций**, которая преобразует *a* в *b*, **и саму последовательность этих операций**. //**СС-граф** состоит из произвольного числа **цепей и циклов**, каждое ребро направлено и помечено. Важный частный случай графов!

граф *a*



граф *b*



Вариант Задачи: каждой **операции назначена своя цена**, **найти цену самой дешёвой последовательности операций** которые переводят граф *a* в граф *b*, и **саму эту последов-ть**.

Нами найден **линейный алгоритм** (точно! при некотором условии) решающий эту Задачу для стандартного набора операций:

расклейка **двух склеек** в графе и склеивание четырёх освободившихся краёв рёбер по-другому;

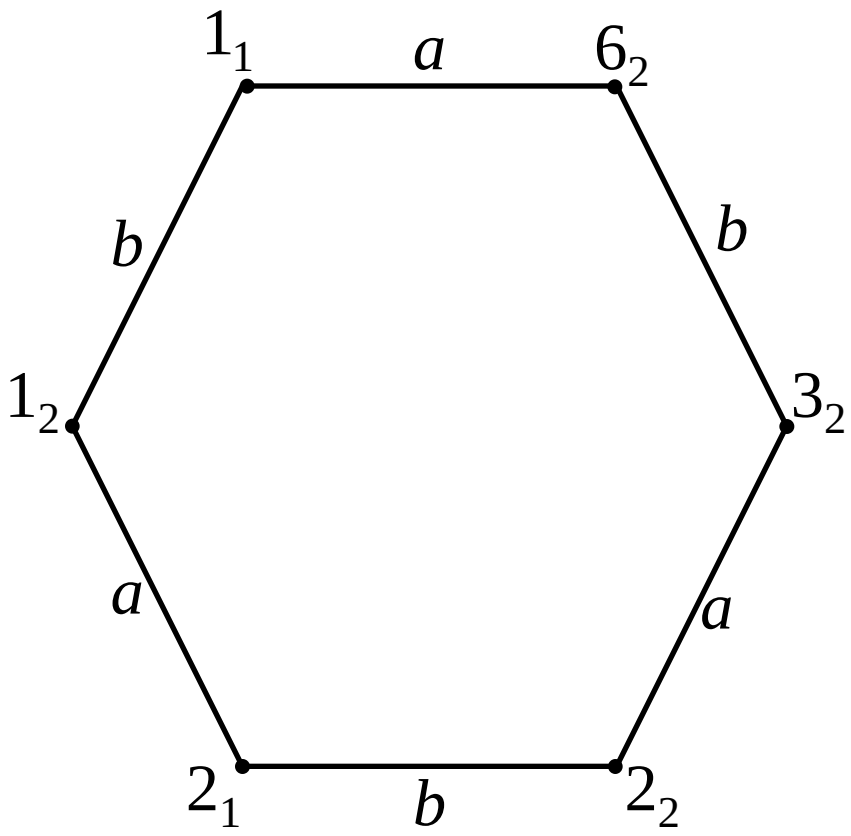
расклейка **одной склейки** и склеивание одного из освободившихся краёв ребра с каким-то свободным краем;

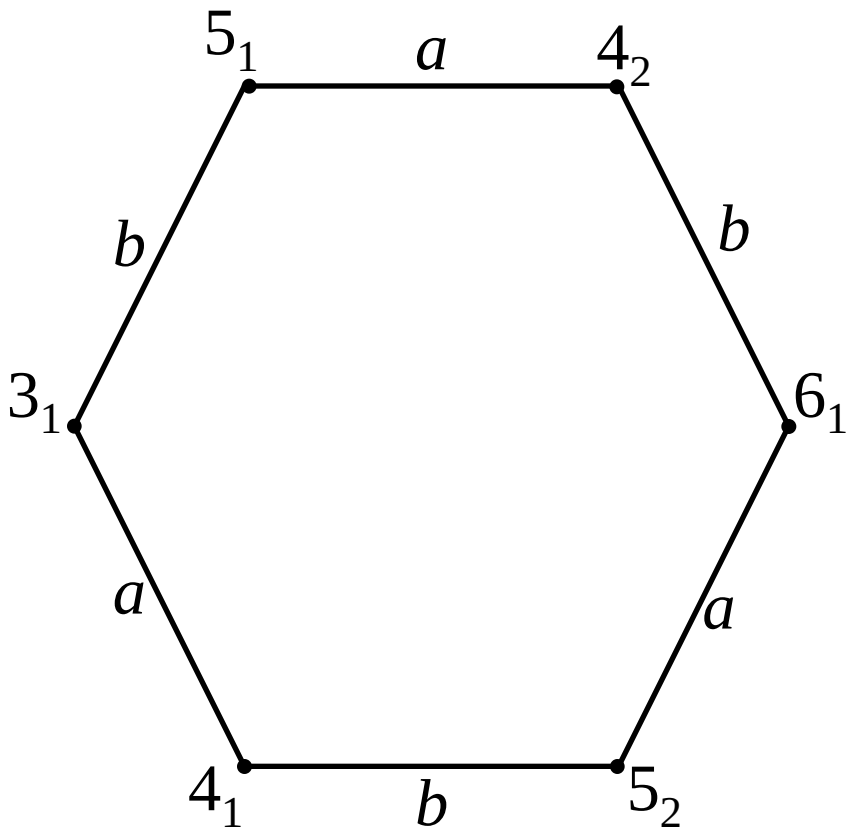
расклейка одной склейки и **склейка** двух свободных краёв;

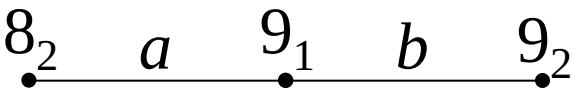
удаление/вставка связного участка рёбер в графе.

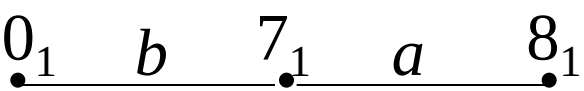
Исследование этой задачи для других наборов операций представляет также большой интерес.

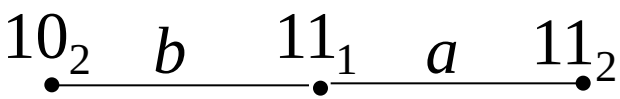
Задачу **удалось решить через показ на новом графе**, какие края рёбер склеены в *a+b.* Т.е. теперь единый граф показывает склеенность краёв сразу в двух исходных графах!



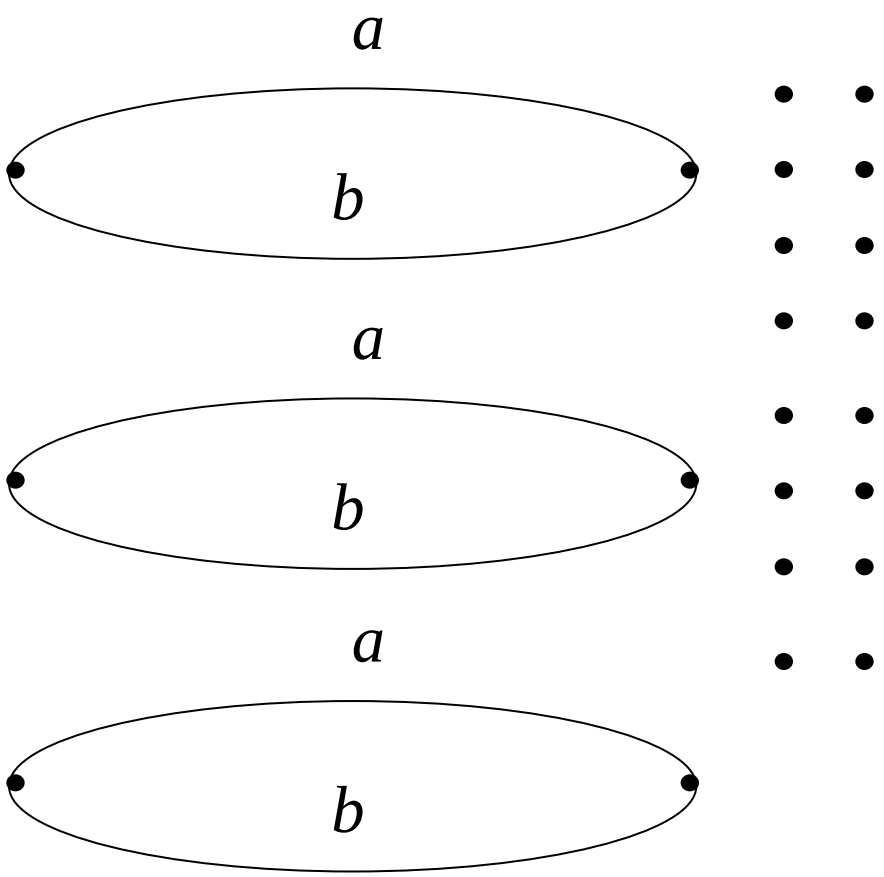






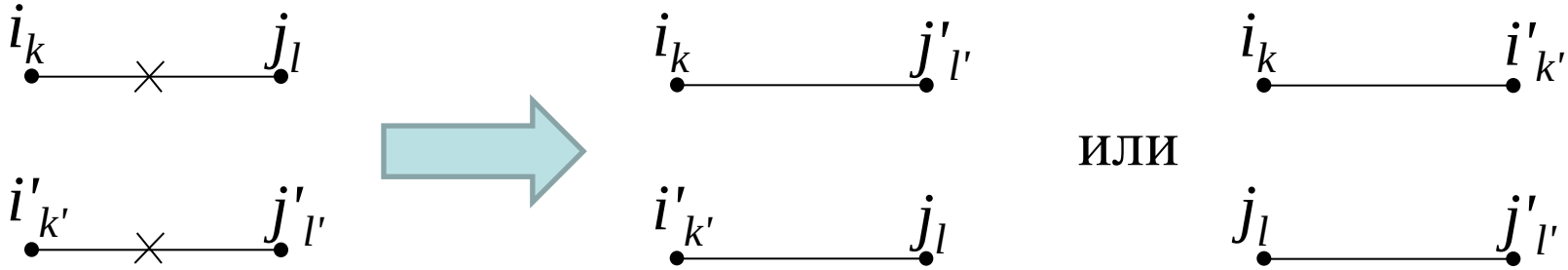


Исходная задача эквивалентна **приведению этими операциями графа краёв к следующему виду:**

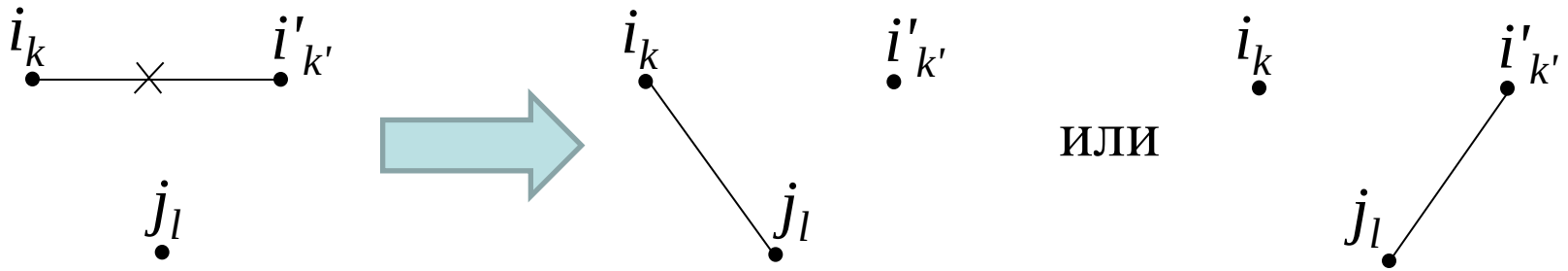


Соот-ие операции над «графом краёв» выглядят так:

1)

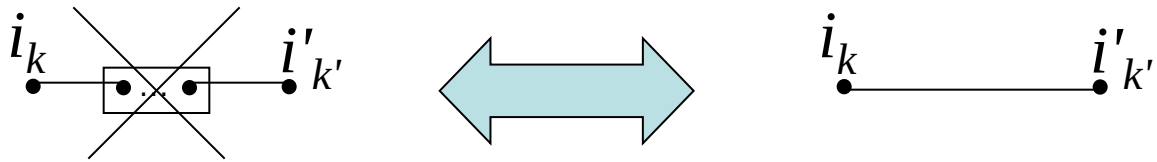


2)



3)

4)



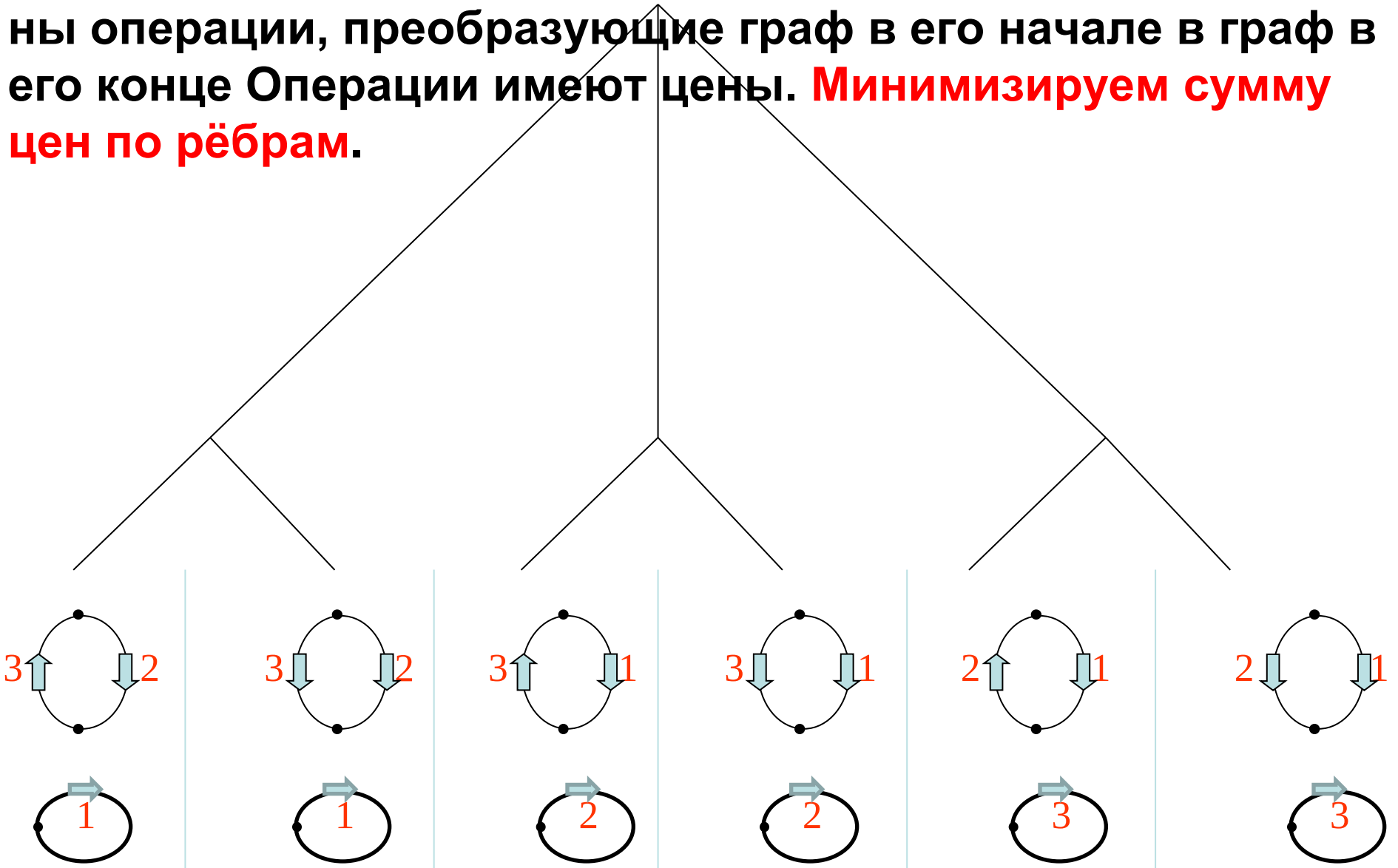
В случае цен приходится **налагать условия**. Например, такие: сначала обозначим цены разреза c_1, склейки c_1', полуторной переклейки $c_{1.5}$, двойной переклейки c_2.
Годятся соотношения цен: $c_2 \le c_1 \le c_1' \le c_{1.5}$ («циклическое») и $c_1 \le c_1' \le c_{1.5} \le c_2$ («линейное»).

Нетривиальный эффект этой задачи – **определённые группы цепей (в графе краёв) нужно обязательно преобразовывать вместе**, чтобы достичь минимума!

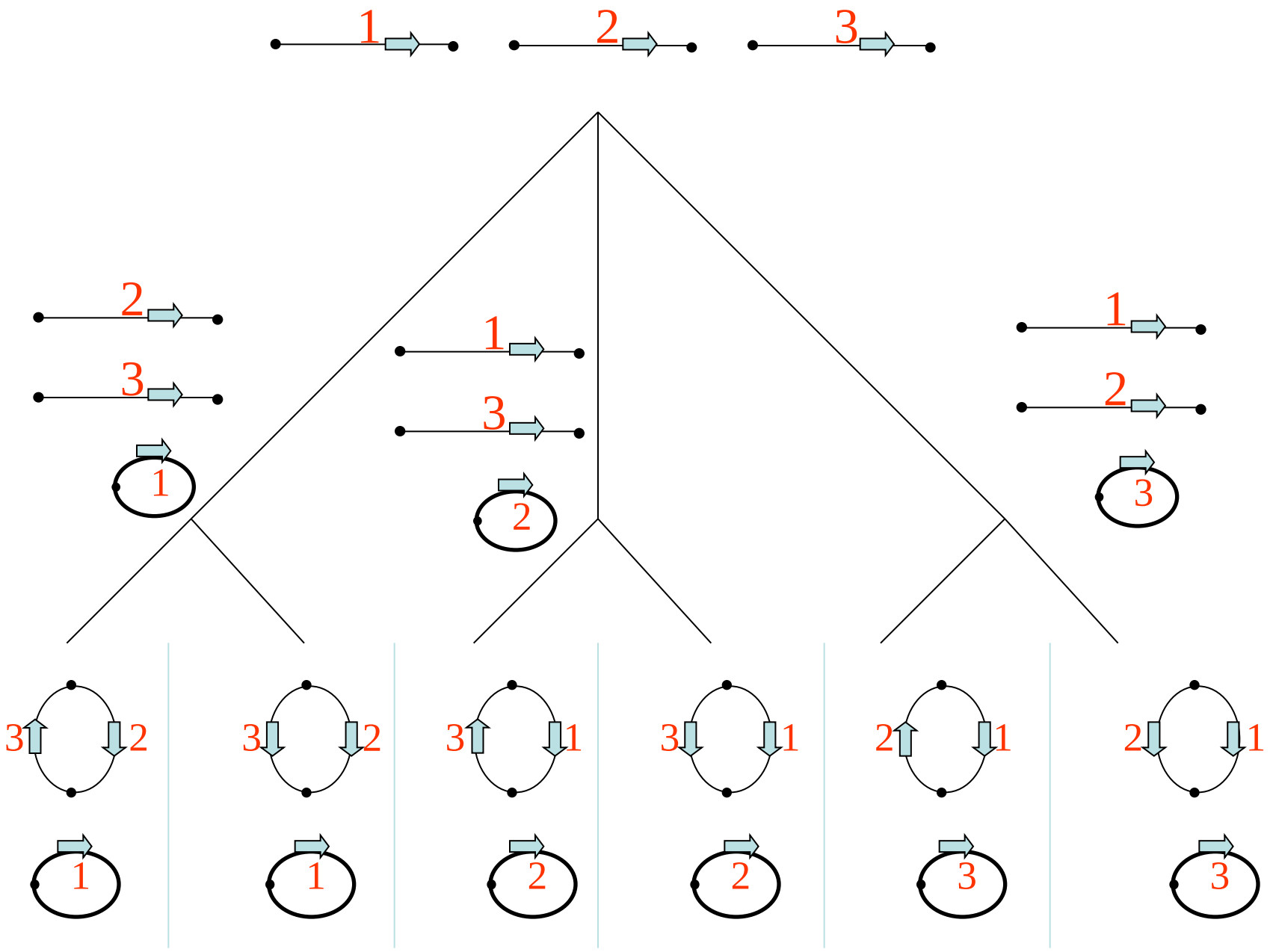
Также важна **задача условной оптимизации**: найти **самую дешёвую последовательность среди самых коротких** последовательностей операций.

Не известно, существует ли полиномиальный по времени алгоритм решения безусловной задачи даже для одного из этих вариантов цен.

4) Оптимальное продолжение на всё дерево СС-графов, заданных в листьях дерева. Расстановка – каждой внутренней вершине приписан свой граф. На рёбрах разрешены операции, преобразующие граф в его начале в граф в его конце Операции имеют цены. Минимизируем сумму цен по рёбрам.



Решение – графы во внутренних вершинах:

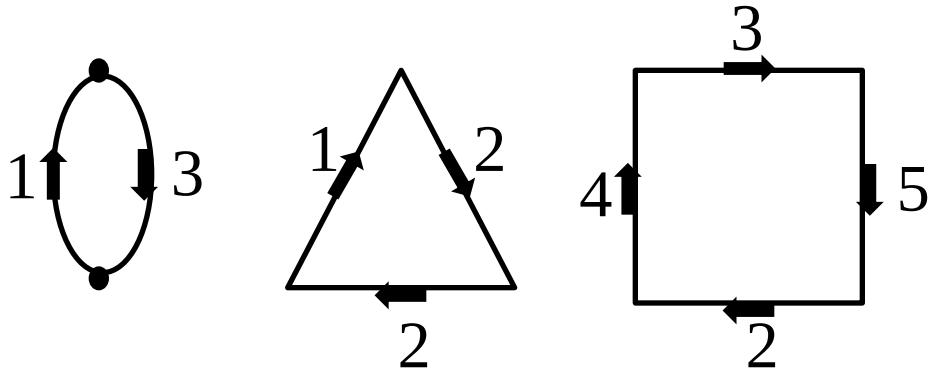


Нами найден кубический от исходных данных алгоритм, точный при некотором условии, который минимизирует цену расстановки.

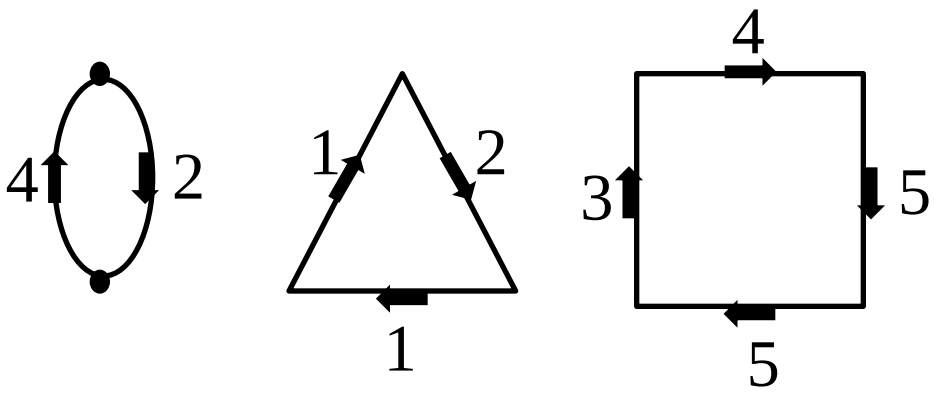
5) Сведение задачи вычисления расстояния между СС-графами в случае ПАРАЛОГОВ к задаче ЦЛП.

Даны:

Граф *a*



Граф *b*

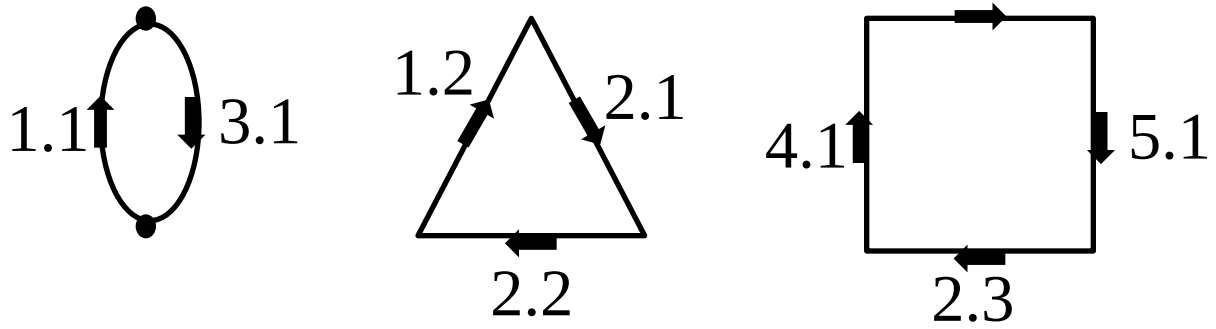


$$F = 0.5\sum x + \sum y - \sum p$$

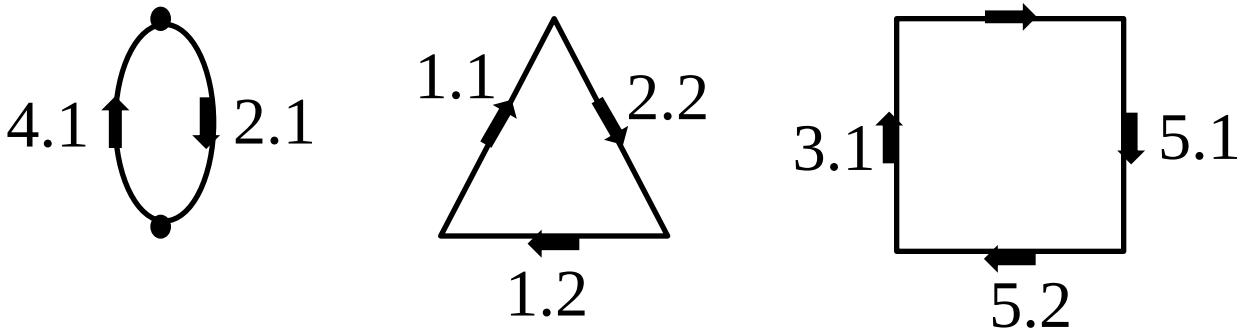
Нужно: нумеровать одноимённые в *a* и в *b* рёбра (называемые *общими*), чтобы минимизировать число операций, преобразующих *a* в *b*. Решить – *минимизировать ц.ф. F* ! Для краткости изложения покажем графы без цепей.

Начнём с нумерации рёбер, чтобы оперировать с ними (первая цифра показывает группу рёбер-паралогов):

Граф *a*

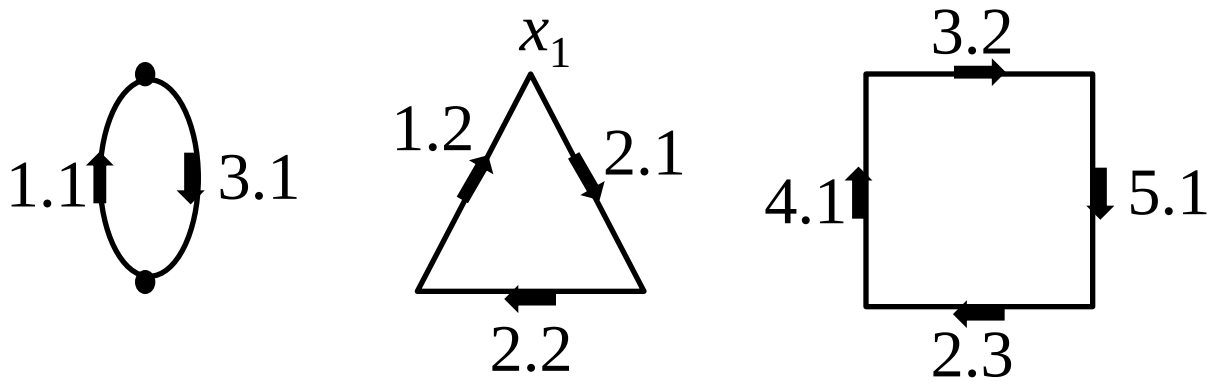


Граф *b*



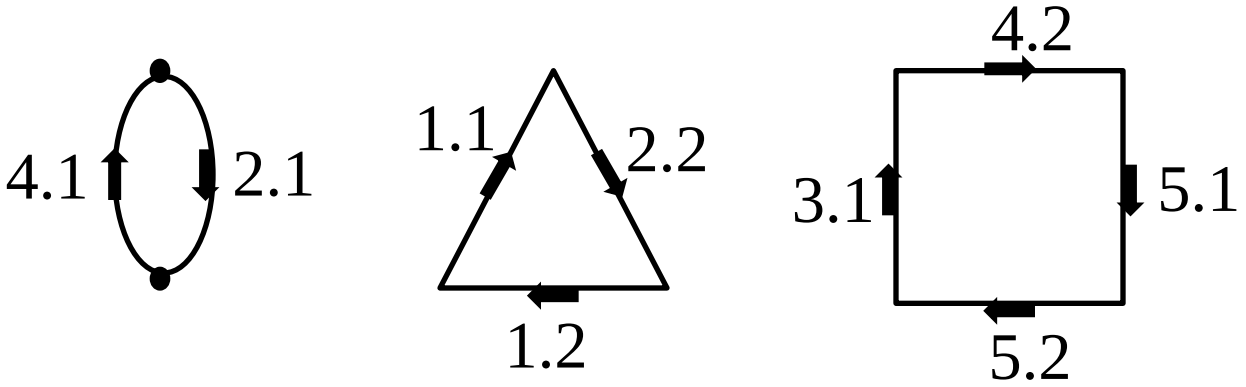
Паре общих рёбер сопоставляется булева переменная *z*, задающая их соответствие. Например, $z_{231}=1$ означает: ребро 2.3 в *a* соответствует ребру 2.1 в *b.* Ограничение: ребру в одном графе соответствует не более одного ребра в другом.

Граф *a*



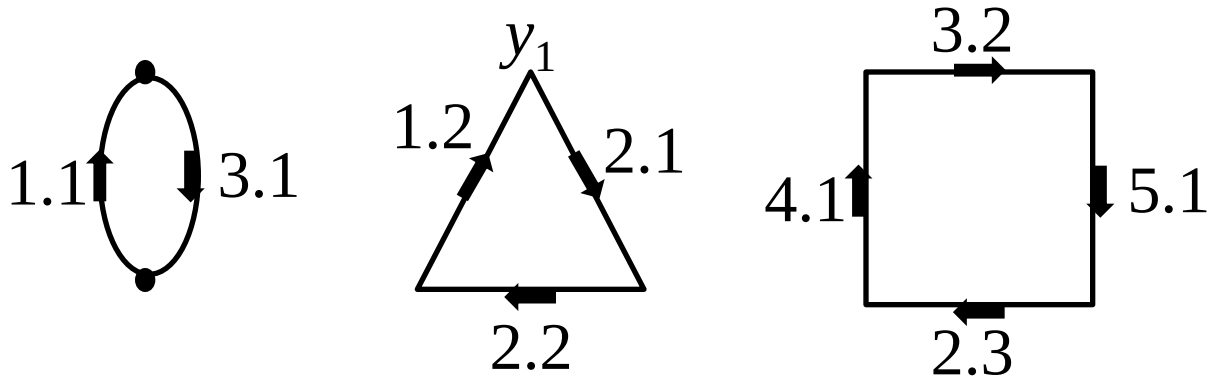
$$x_1 \geq (z_{121} + z_{122}) - (z_{211} + z_{212})$$

Граф *b*



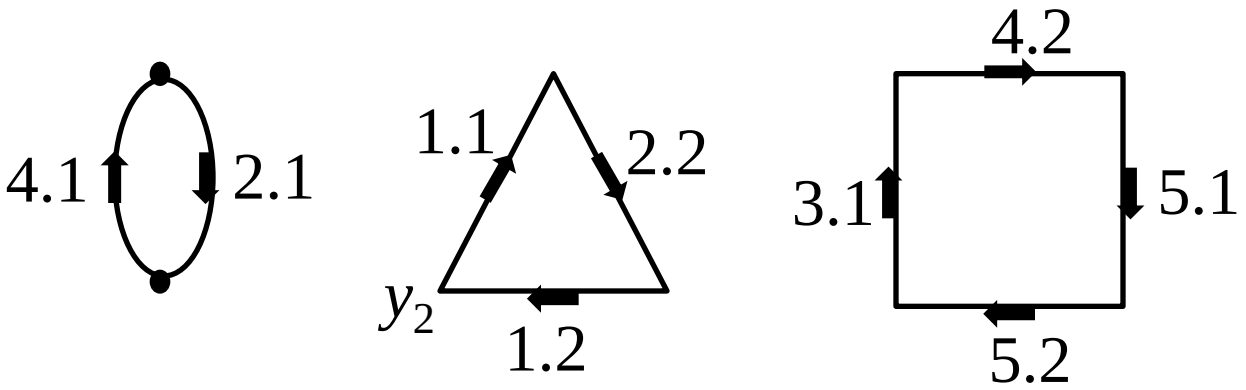
Каждой вершине сопоставляется булева переменная *x*, они указывают, какие вершины в графах соединяют общее ребро с *особым* ребром в соответствии со значениями переменных *z*. Приведённое неравенство выражает: если по *z*, в *a* ребро 1.2 общее, а 2.1 особое, то x_1=1; и т.д.

Граф *a*



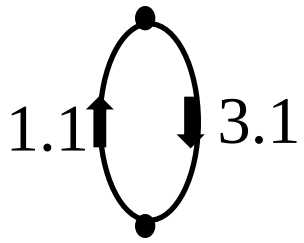
$$y_1 + y_2 \geq z_{122} + (z_{211} + z_{212}) + (z_{111} + z_{121}) - 2$$

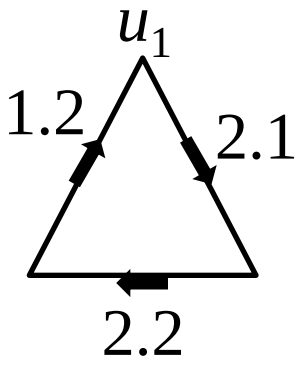
Граф *b*

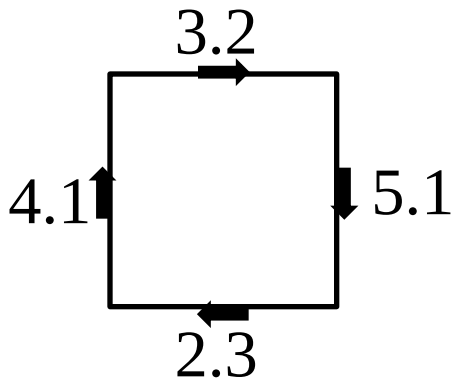


Каждой вершине сопоставляется булева переменная *y*, равная 0, если вершина принадлежит особому ребру; на пути из склеек общих рёбер *y* принимает чередующиеся значения, начиная с 0. Например, приведённое ограничение выражает: если рёбра 1.2 в *a* и в *b* соответствуют друг другу, ребро 2.1 в *a* и ребро 1.1 в *b* общие и y_1=0, то y_2=1.

Граф *a*

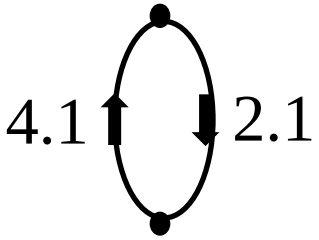


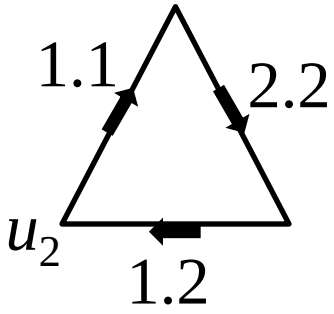


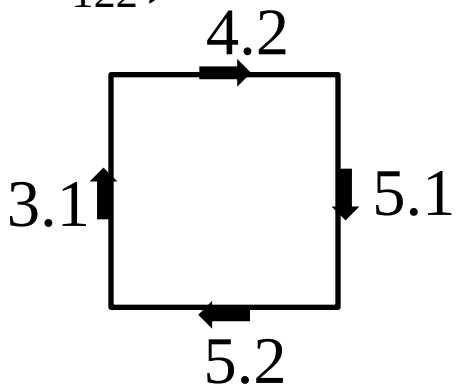


$$u_2 \le u_1 + m_2(1 - z_{122}), \quad u_1 \le u_2 + m_1(1 - z_{122})$$

Граф *b*



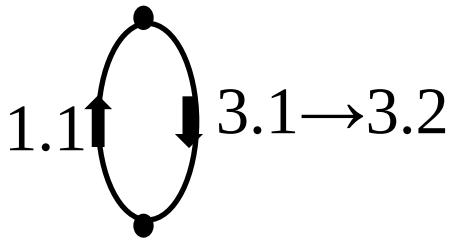


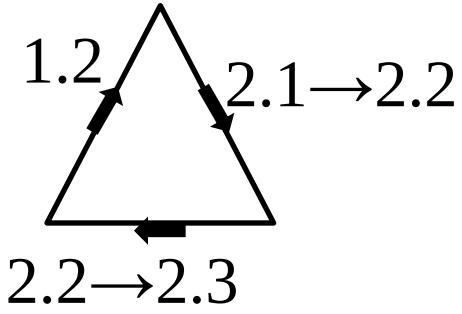


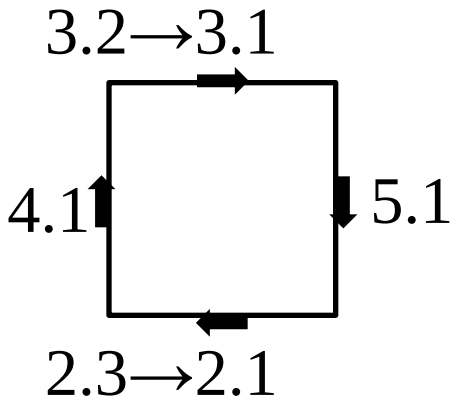
Каждой вершине *k* сопоставляется целочисленная переменная u_k, ограниченная сверху положительным числом m_k, и булева переменная p_k, равная 1, если $u_k=m_k$. Приведённые ограничения выражают: если рёбра 1.2 в *a* и в *b* соответствуют друг другу, то $u_1=u_2$.

Решение для исходной пары графов: z_{112}, z_{121}, z_{212}, z_{231}, z_{321}, z_{411}, z_{511} равны 1, остальные значения 0. Нумерация паралогов в исходном графе такова:

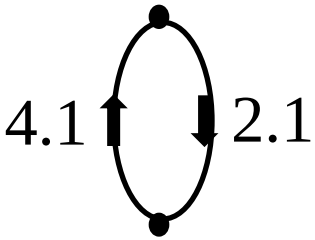
Граф *a*

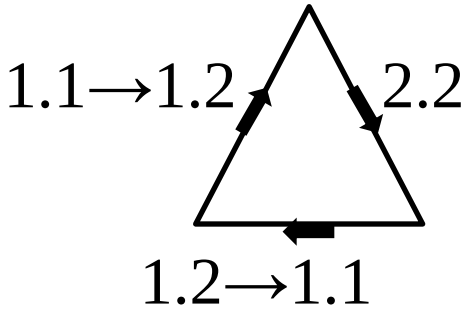


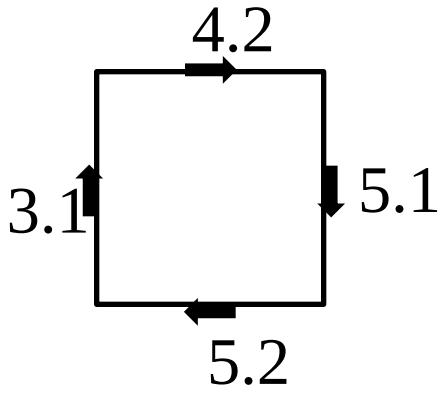




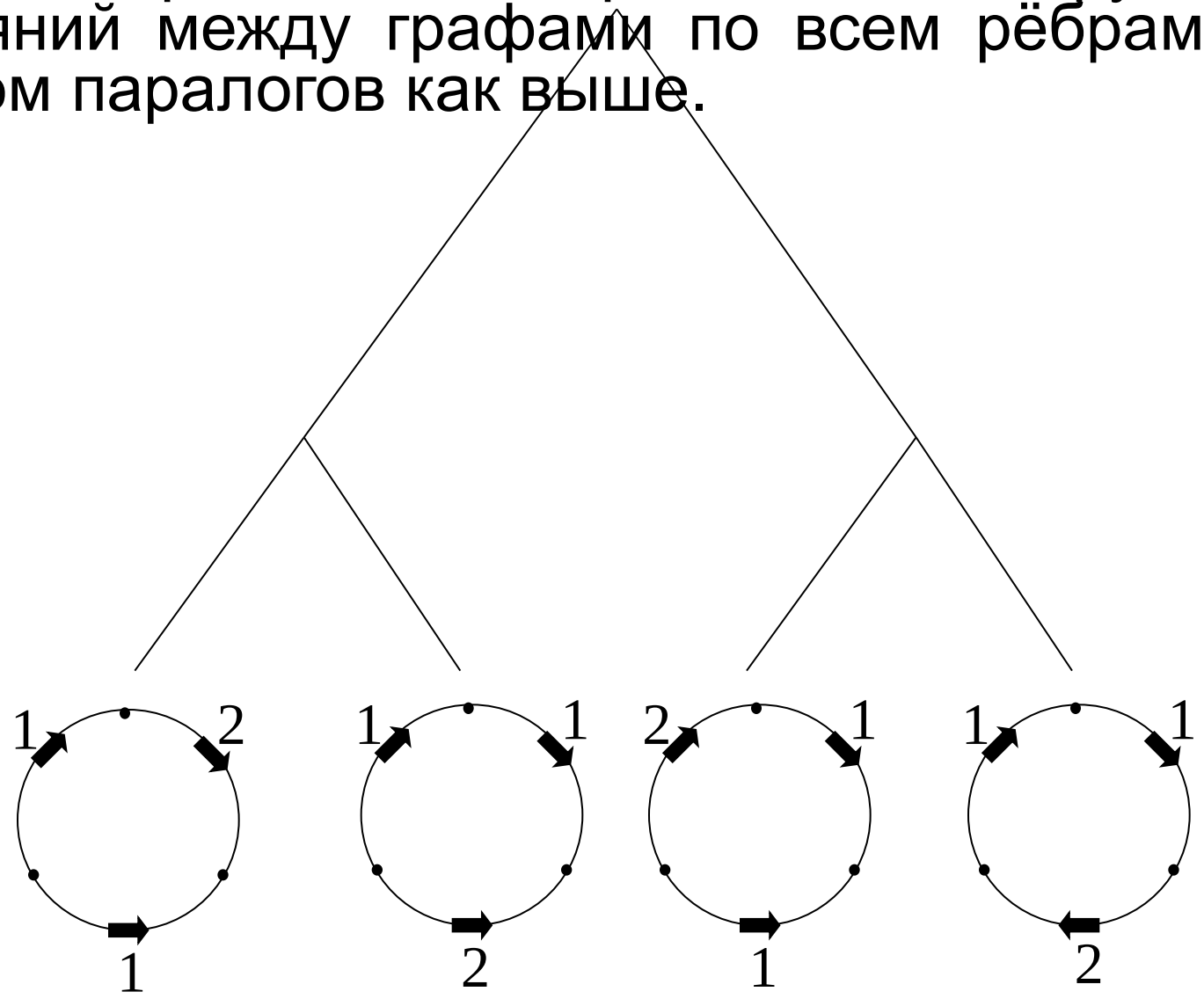
Граф *b*



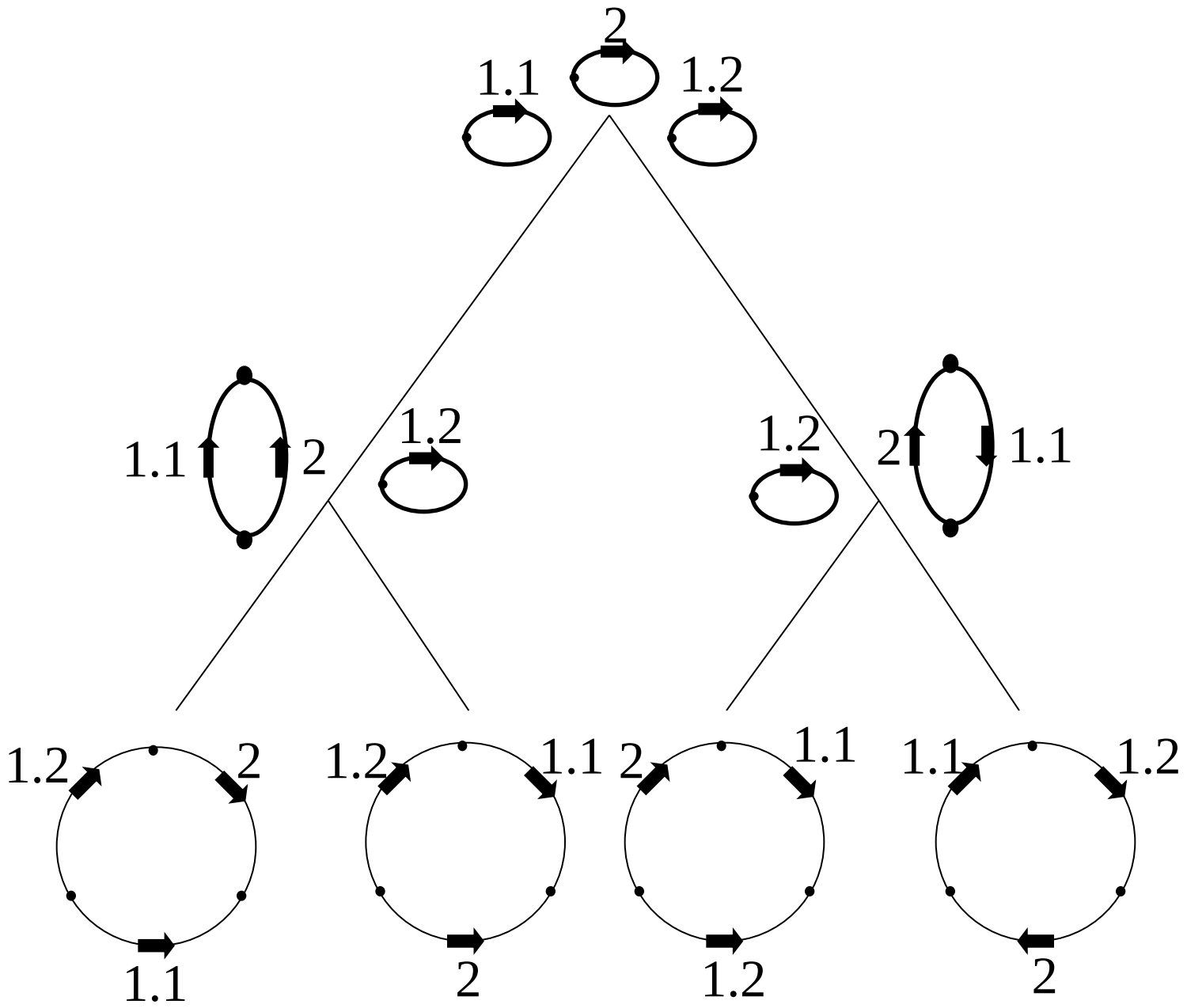




Аналогично к задаче ЦЛП сводится и задача реконструкции графов на данном дереве. Например, в листьях указаны графы. Найти расстановку графов во внутренних вершинах, которая минимизирует сумму расстояний между графами по всем рёбрам дерева с учётом паралогов как выше.



Показано **решение**:



Наши недавние публикации,
эти и другие на странице http://lab6.iitp.ru/ :

V.A. Lyubetsky,
R.A. Gershgorin, A.V. Seliverstov, K.Yu. Gorbunov,
Algorithms for Reconstruction of Chromosomal Structures,
BMC Bioinformatics, 2016, 17:40.

O.A. Zverkov, A.V. Seliverstov, V.A. Lyubetsky,
Regulation of Expression and Evolution of Genes in Plastids of Rhodophytic Branch,
Life, Jan 29 2016, 6:7.

СПАСИБО за внимание