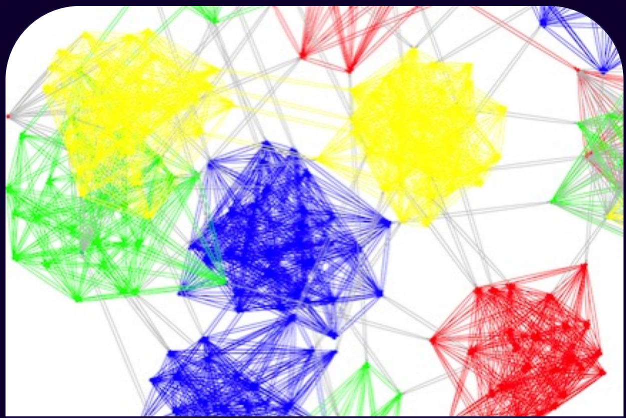
Реберные разбиения графов

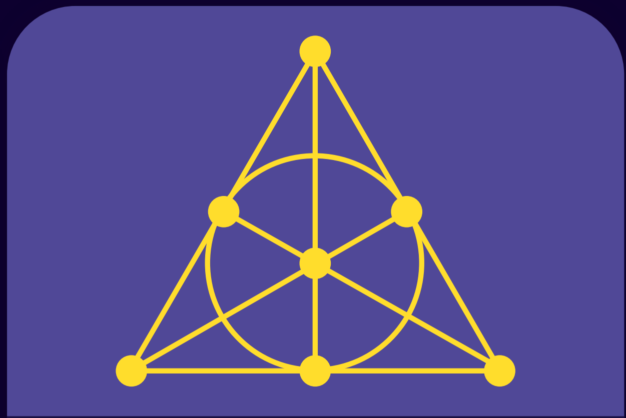
Якунин Александр
R&D центр Т-Банка

План

От распределенной обработки данных



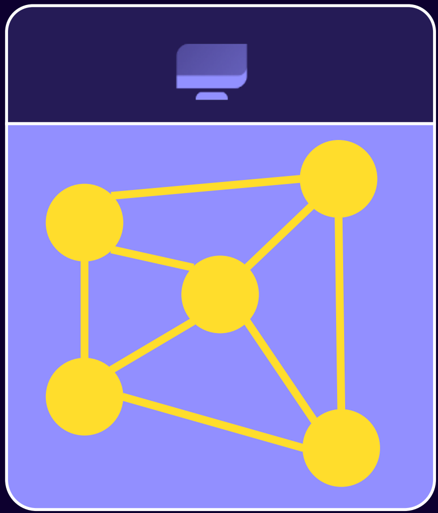
Через дискретную оптимизацию



К геометрии над конечными полями

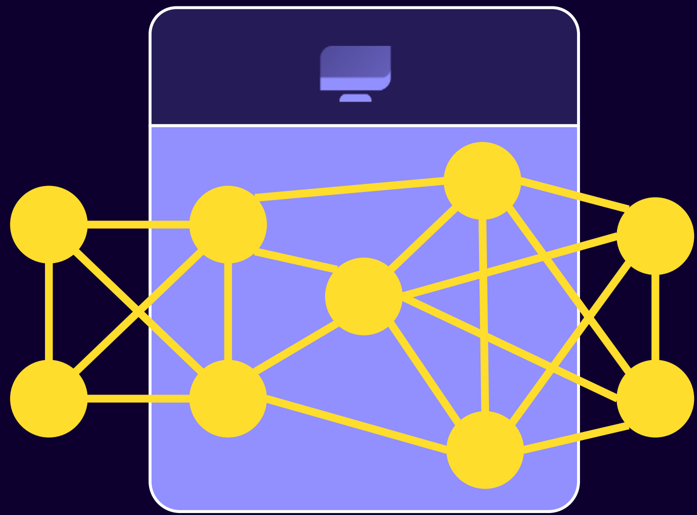
Распределенная обработка данных

- Хотим считать алгоритмы на графах



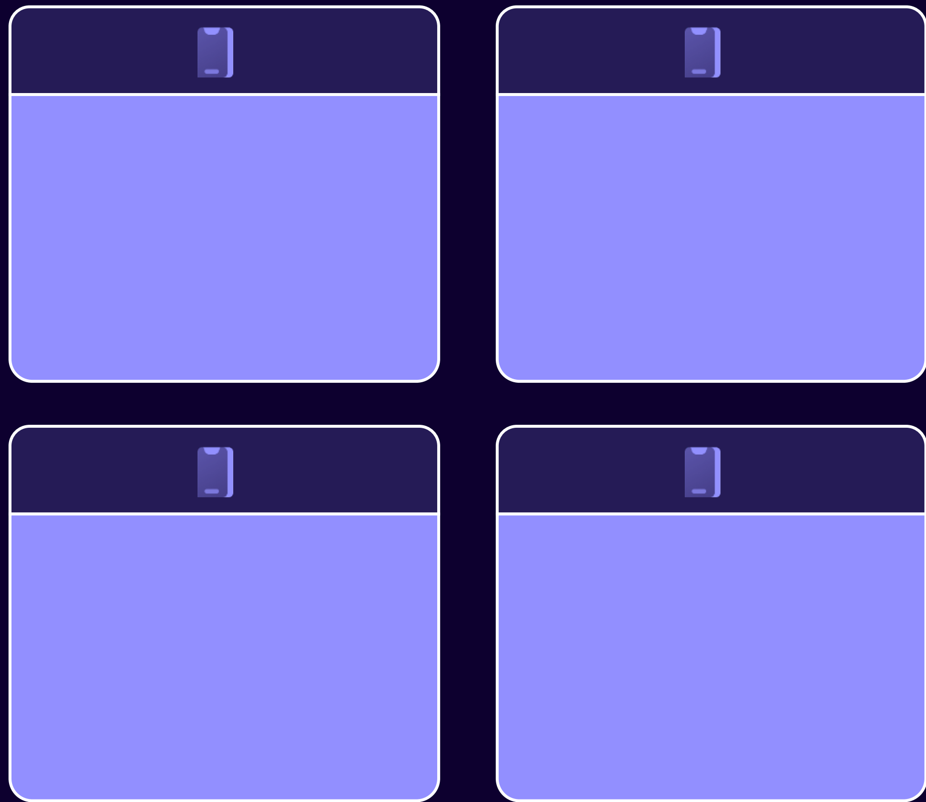
Распределенная обработка данных

- Хотим считать алгоритмы на графах
- А если граф слишком большой?



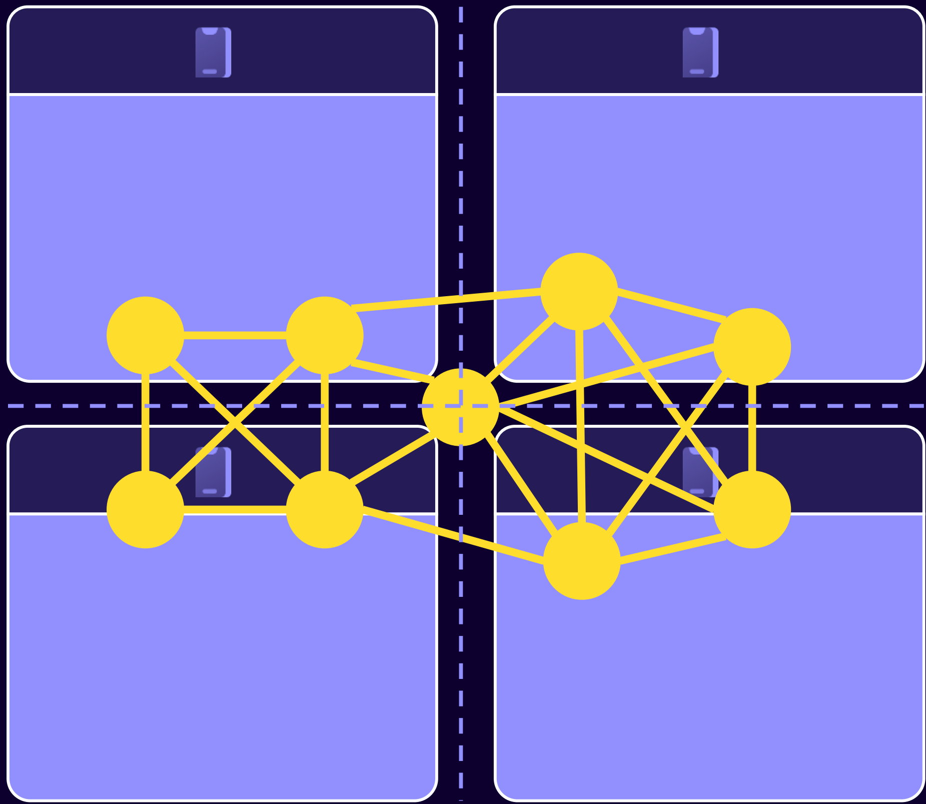
Распределенная обработка данных

- Хотим считать алгоритмы на графах
- А если граф слишком большой?
- Используем кластер из нескольких вычислительных узлов



Распределенная обработка данных

- Хотим считать алгоритмы на графах
- А если граф слишком большой?
- Используем кластер из нескольких вычислительных узлов
- Нужно разбить граф на части

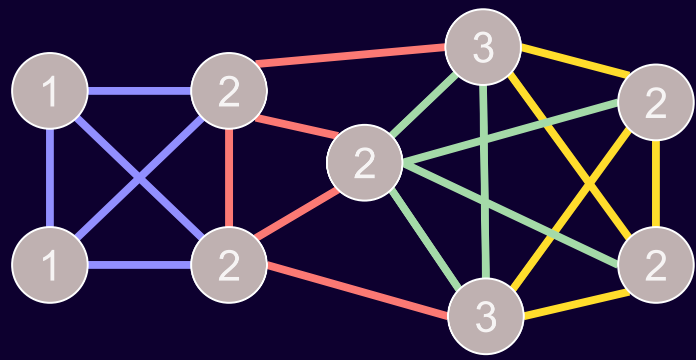


Реберное разбиение графа

Задача

Ищем разбиение ребер графа на K частей

$$E = E_1 \sqcup \cdots \sqcup E_k$$



Пример сбалансированного разбиения

с фактором репликации 2

Реберное разбиение графа

Задача

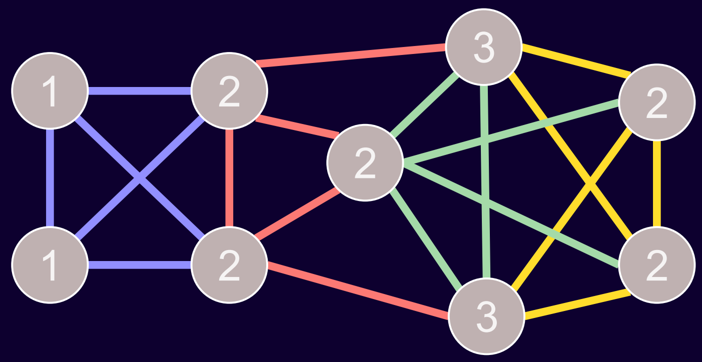
Ищем разбиение ребер графа на K частей

$$E = E_1 \sqcup \cdots \sqcup E_k$$

Баланс

Хотим, чтобы части были примерно одинакового размера

$$\max_i |E_i| \leq \frac{m(1+\varepsilon)}{k}$$



Пример сбалансированного разбиения

с фактором репликации 2

Реберное разбиение графа

Задача

Ищем разбиение ребер графа на К частей

$$E = E_1 \sqcup \dots \sqcup E_k$$

Баланс

Хотим, чтобы части были примерно одинакового размера

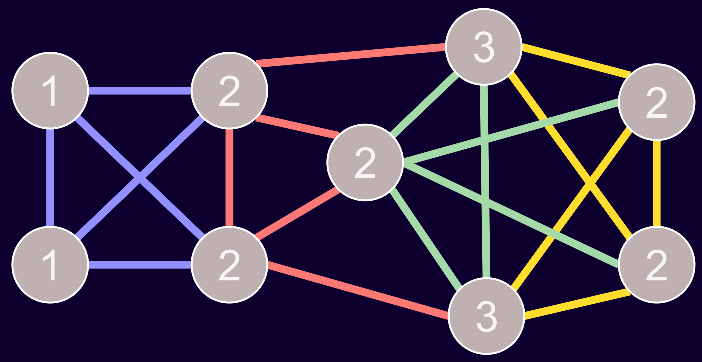
$$\max_i |E_i| \le \frac{m(1+\varepsilon)}{k}$$

Коммуникация

Хотим, чтобы у вершин было как можно меньше разноцветных ребер

$$\frac{1}{n}\sum_{v \in V} rf(v) \to min$$

Где $rf(v) = |\{i: E_i \cap E(v) \neq \emptyset\}|$



Пример сбалансированного разбиения

с фактором репликации 2

Реберное разбиение графа

Задача

Ищем разбиение ребер графа на K частей

$$E = E_1 \sqcup \dots \sqcup E_k$$

Баланс

Хотим, чтобы части были примерно одинакового размера

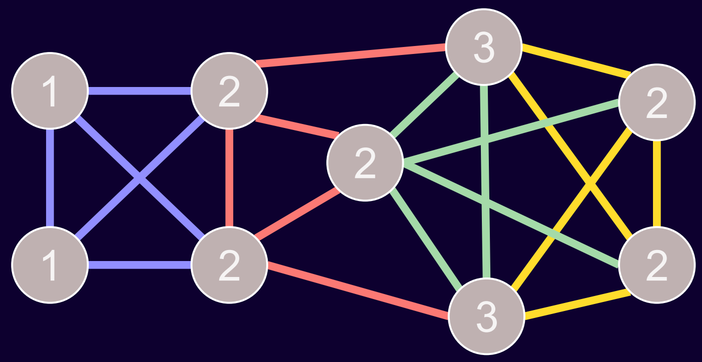
$$\max_i |E_i| \le \frac{m(1+\varepsilon)}{k}$$

Коммуникация

Хотим, чтобы у вершин было как можно меньше разноцветных ребер

$$\frac{1}{n}\sum_{v \in V} rf(v) \to min$$

Где $rf(v) = |\{i : E_i \cap E(v) \neq \emptyset\}|$



Пример сбалансированного разбиения

с фактором репликации 2

Теорема

Эта задача NP-трудная

Случайное разбиение

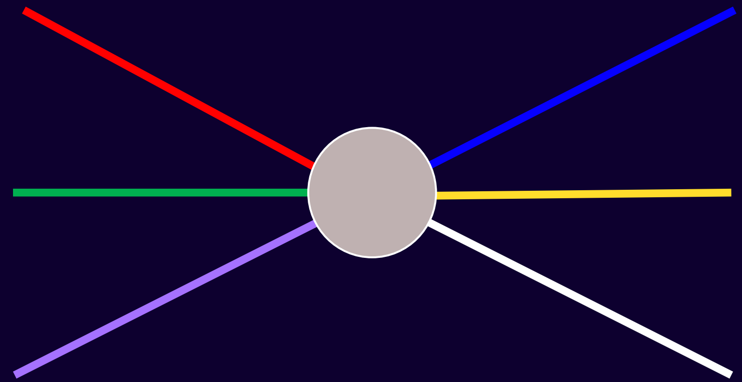
Алгоритм

Случайно красим каждое ребро

Баланс: почти идеальный
По закону больших чисел

Фактор репликации: максимально плохой

$$rf(v) \approx \min(d(v), k)$$



Разбиение «по вершинам»

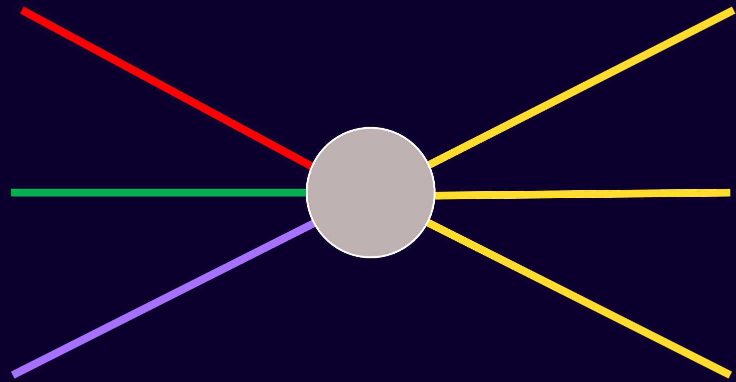
Алгоритм

1. Случайно красим <u>вершины</u>
2. Каждое ребро красим в цвет одной из вершин

Фактор репликации:

$$rf(v) \approx \min\left(\frac{d(v)}{2}, k\right)$$

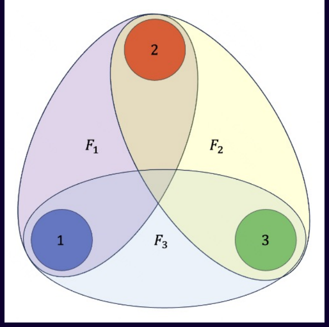
Хорошо работает если k сильно больше средней степени

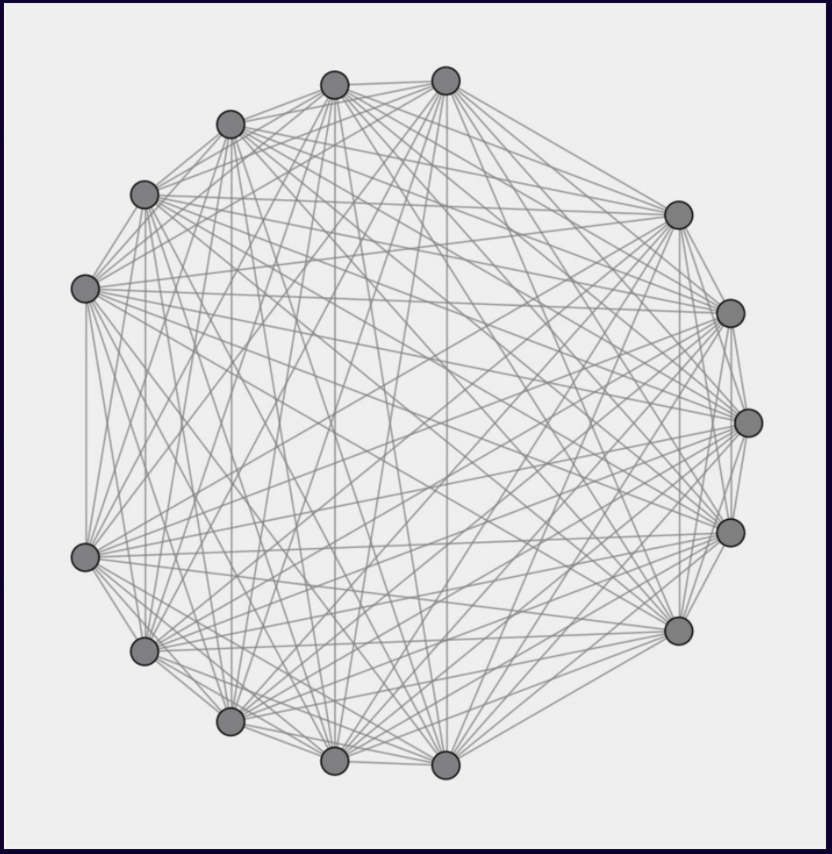


Разбиение «через множества»

Используем систему множеств $F_1 \dots F_r \subset \{1 \dots k\}$

В примере $\{1, 2\}, \{1, 3\}, \{2, 3\}$



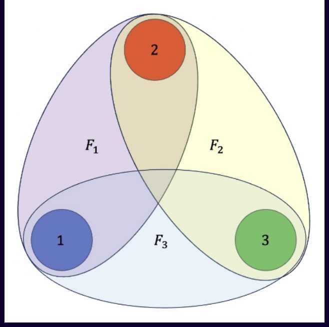


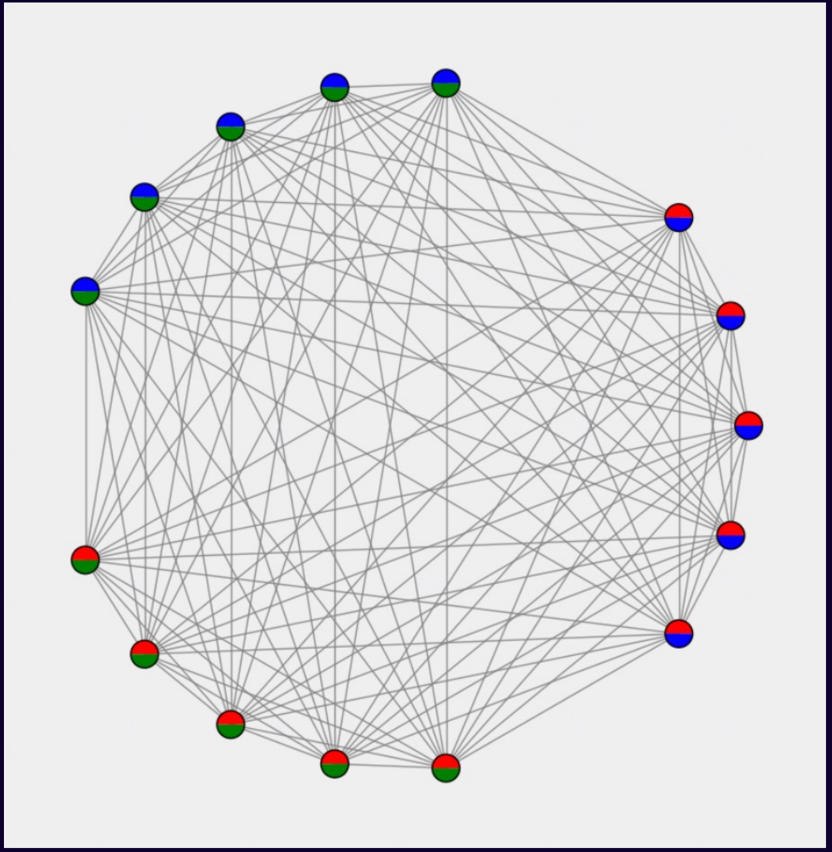
Разбиение «через множества»

Используем систему множеств $F_1 \dots F_r \subset \{1 \dots k\}$

В примере $\{1, 2\}, \{1, 3\}, \{2, 3\}$

1. Для каждой вершины v
   случайно выберем $F_v \in \{F_1 \dots F_r\}$



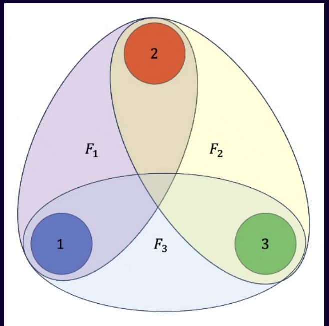


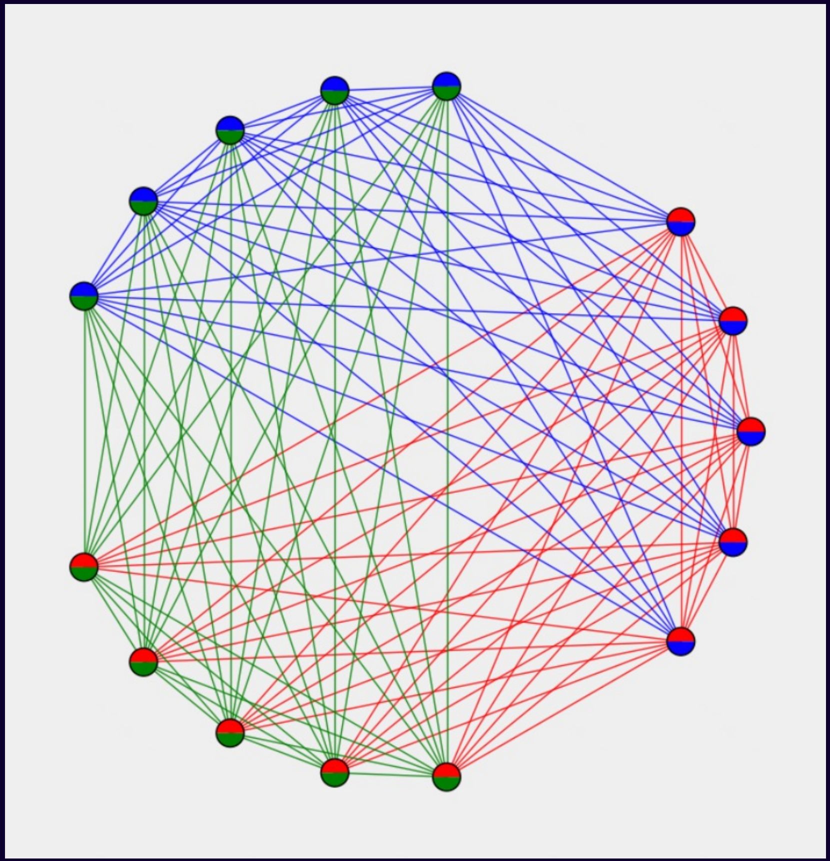
Разбиение «через множества»

Используем систему множеств $F_1 \dots F_r \subset \{1 \dots k\}$

В примере $\{1, 2\}, \{1, 3\}, \{2, 3\}$

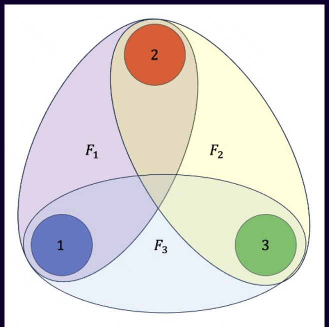
1. Для каждой вершины v
   случайно выберем $F_v \in \{F_1 \dots F_r\}$
2. Каждое ребро $e = vu$
   раскрасим в цвет $\mathrm{i} \in F_v \cap F_u$

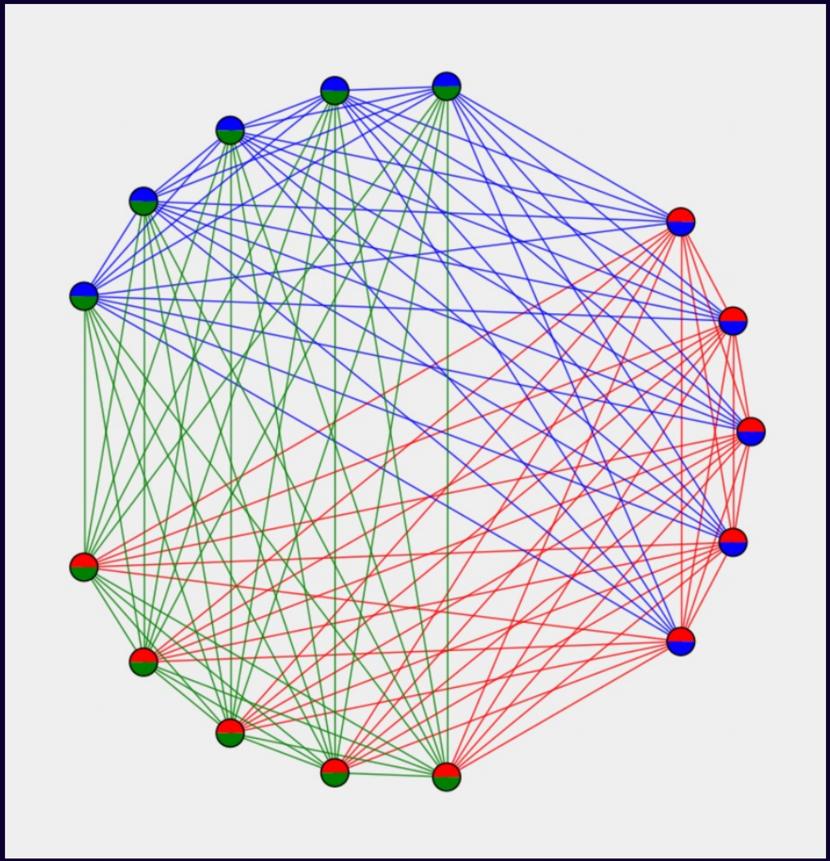




Какими должны быть множества

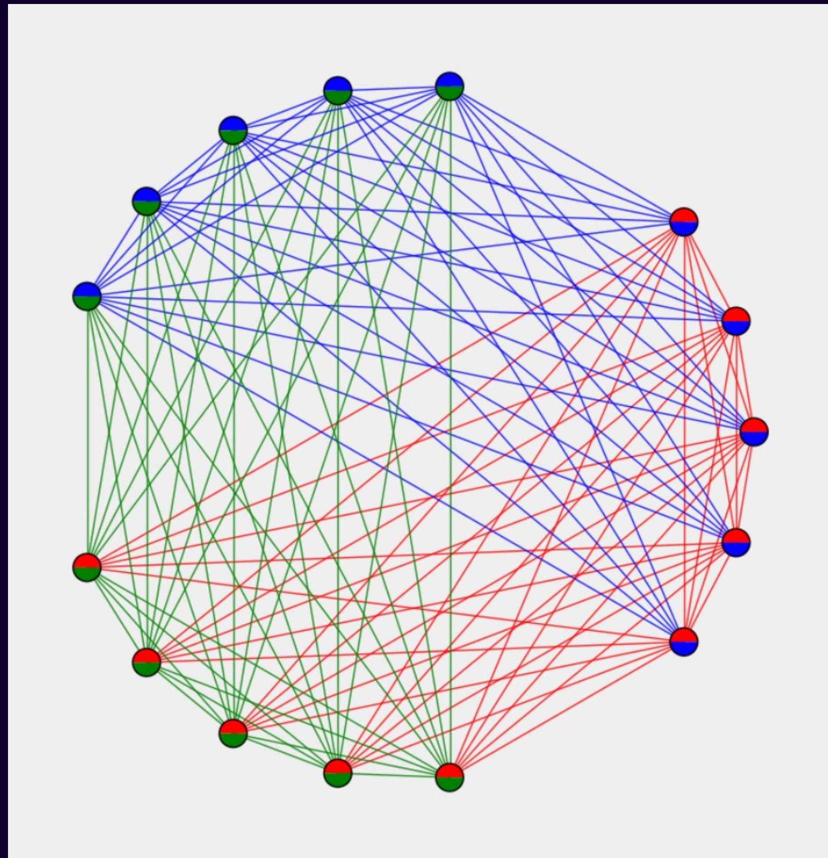
1. Множества попарно пересекаются
   Для корректности алгоритма
2. Система симметричная
   Для сбалансированности разбиения
3. Размер множеств – как можно меньше
   Для уменьшения фактора репликации





Какими должны быть множества

1. Множества попарно пересекаются
   Для корректности алгоритма
2. Система симметричная
   Для сбалансированности разбиения
3. Размер множеств – как можно меньше
   Для уменьшения фактора репликации



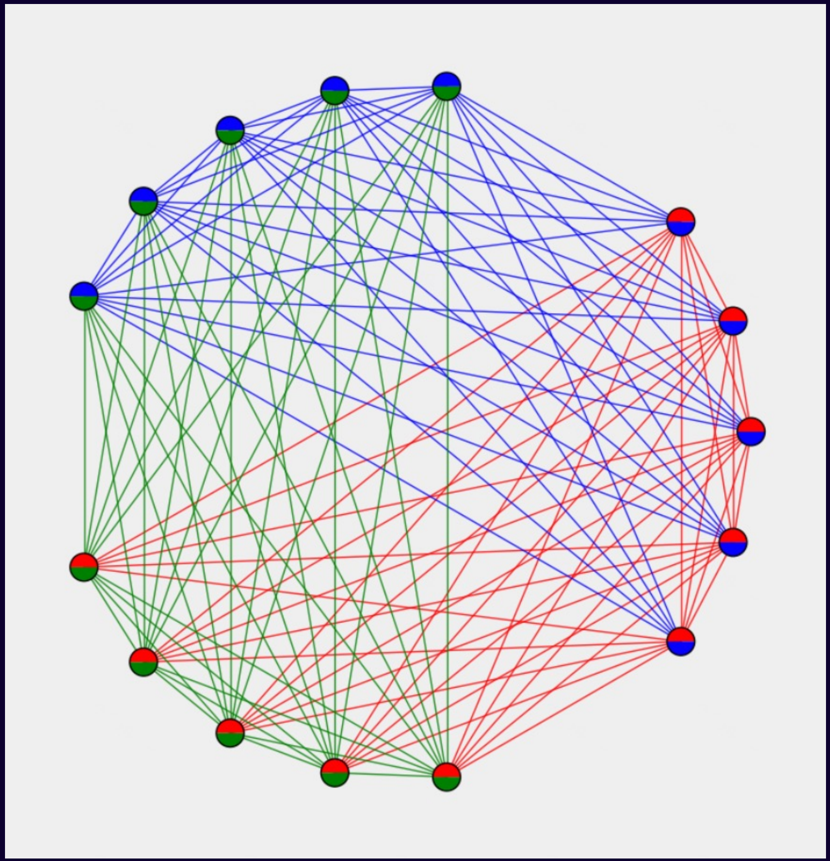
Теорема

В симметричной системе попарно пересекающихся множеств $F_1 \dots F_r$

$$\frac{|\mathrm{F}_1| + \cdots + |\mathrm{F}_r|}{r} > \sqrt{k}$$

Какими должны быть множества

1. Множества попарно пересекаются
   Для корректности алгоритма
2. Система симметричная
   Для сбалансированности разбиения
3. Размер множеств – как можно меньше
   Для уменьшения фактора репликации



Теорема
В симметричной системе попарно пересекающихся множеств $F_1 \ldots F_r$

$$\frac{|\mathrm{F}_1| + \cdots + |\mathrm{F}_r|}{r} > \sqrt{k}$$

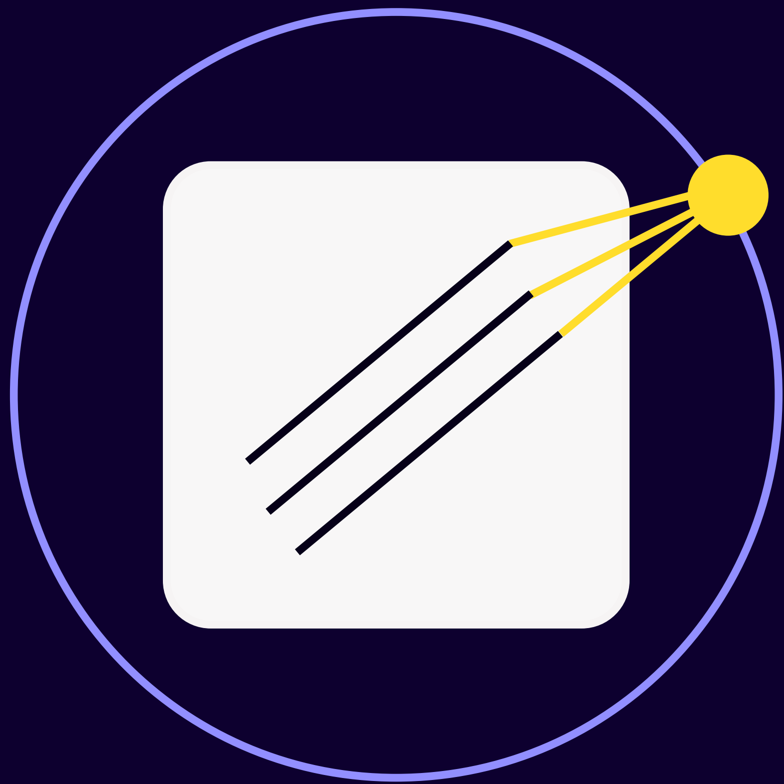
Как находить такие оптимальные системы?

Проективная плоскость

Набор прямых и точек, такой что:

- Любые 2 прямые пересекаются ровно в 1 точке.
- Через любые 2 точки проходит ровно 1 прямая.

Строится поверх плоскости $\mathbb{R}^2$.



Проективная плоскость

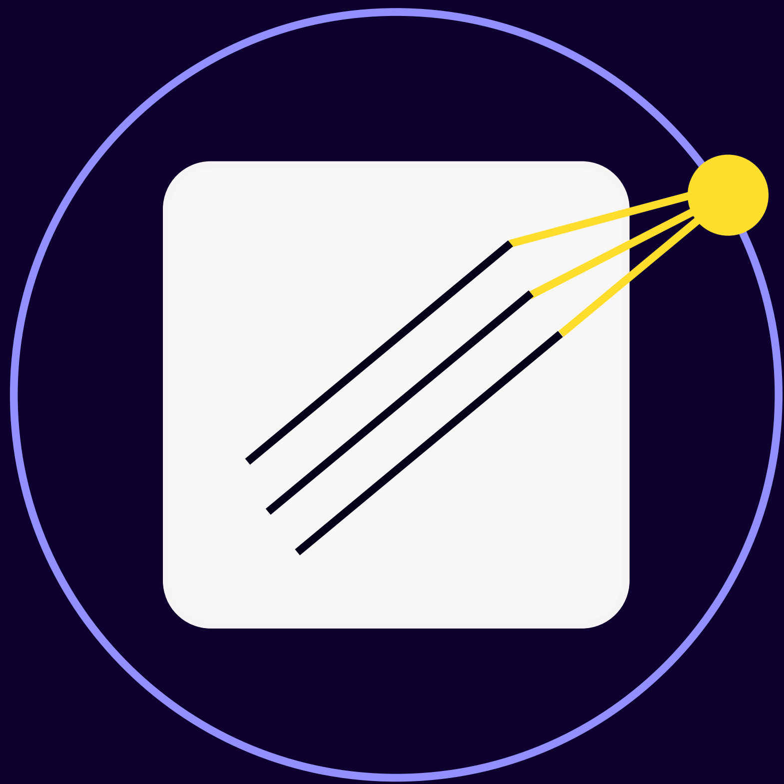
Набор прямых и точек, такой что:

- Любые 2 прямые пересекаются ровно в 1 точке.
- Через любые 2 точки проходит ровно 1 прямая.

Строится поверх плоскости $\mathbb{R}^2$.

А что если заменить $\mathbb{R}$ на другое *поле?*

Например, на $\mathbb{Z}_p$, где p – простое число?



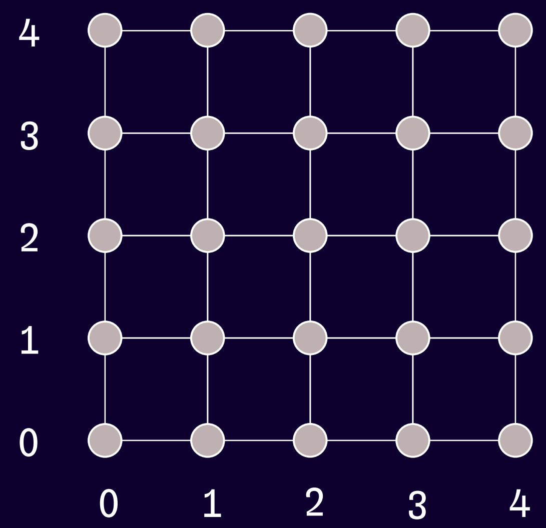
КПП: конструкция

Точки:

- Точки сетки $\mathbb{Z}_p^2$

Прямые:

- Обычные $\{(x, ax+b): x \in \mathbb{Z}_p\}$



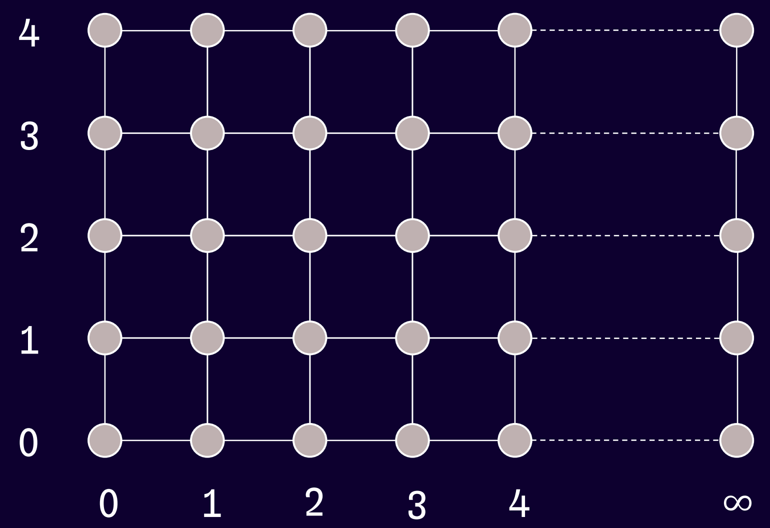
КПП: конструкция

Точки:

- Точки сетки $\mathbb{Z}_p^2$
- Бесконечно удаленные точки $(\infty, 0) \dots (\infty, p-1)$

Прямые:

- Обычные $\{(x, ax+b) : x \in \mathbb{Z}_p\} \cup \{(\infty, a)\}$



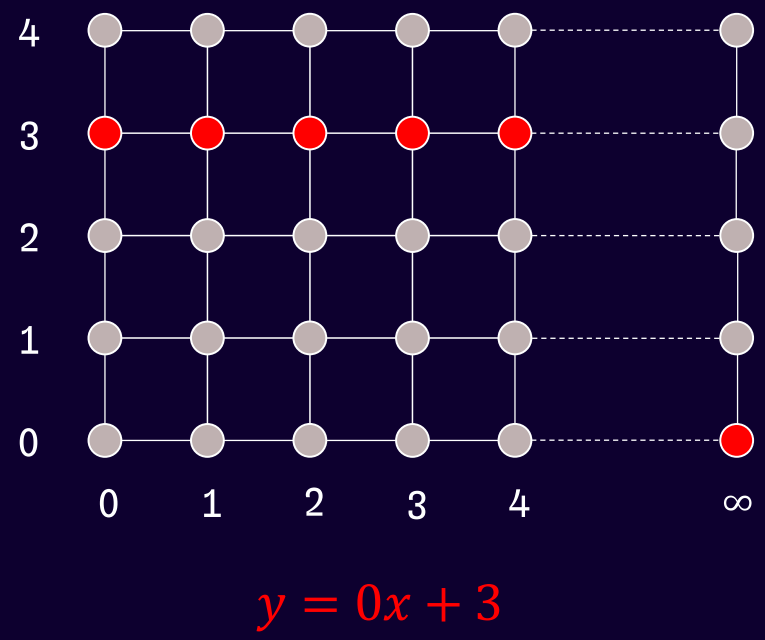
КПП: конструкция

Точки:

- Точки сетки $\mathbb{Z}_p^2$
- Бесконечно удаленные точки $(\infty, 0) \dots (\infty, p-1)$

Прямые:

- Обычные $\{(x, ax+b): x \in \mathbb{Z}_p\} \cup \{(\infty, a)\}$



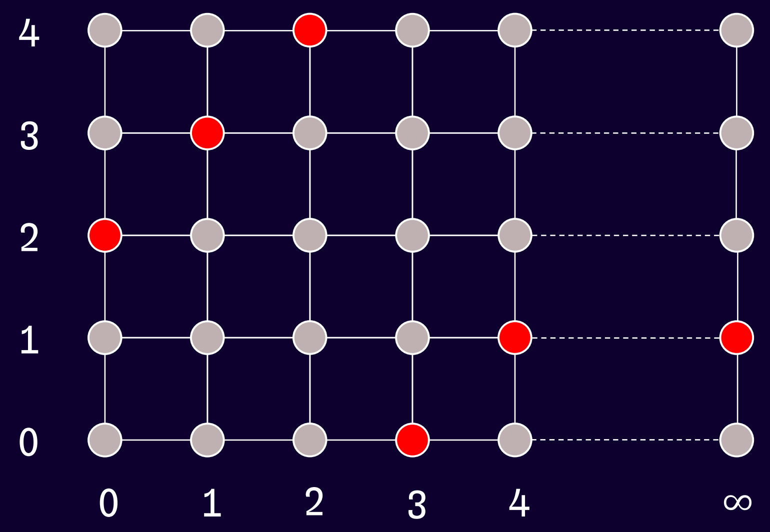
КПП: конструкция

Точки:

- Точки сетки $\mathbb{Z}_p^2$
- Бесконечно удаленные точки $(\infty, 0) \dots (\infty, p-1)$

Прямые:

- Обычные $\{(x, ax+b): x \in \mathbb{Z}_p\} \cup \{(\infty, a)\}$



$y = 1x + 2$

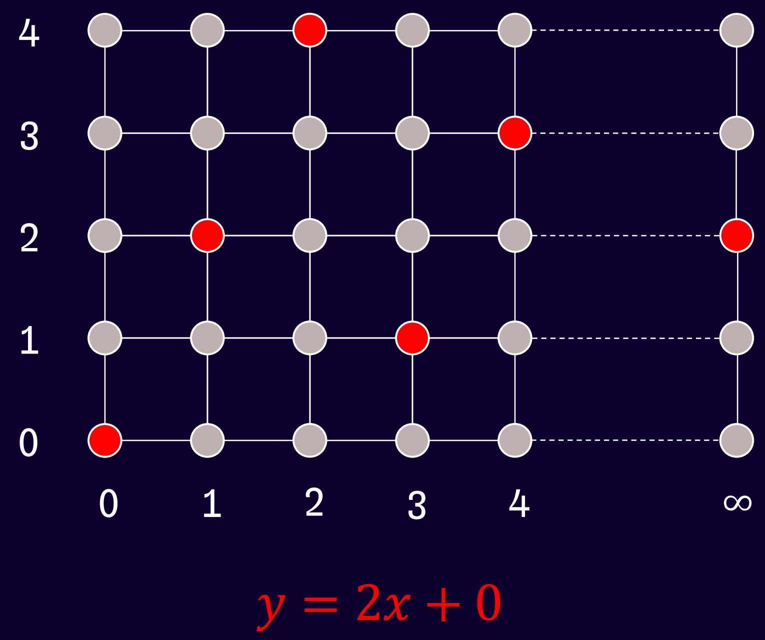
КПП: конструкция

Точки:

- Точки сетки $\mathbb{Z}_p^2$
- Бесконечно удаленные точки $(\infty, 0) \dots (\infty, p-1)$

Прямые:

- Обычные $\{(x, ax+b) : x \in \mathbb{Z}_p\} \cup \{(\infty, a)\}$



КПП: конструкция

Точки:

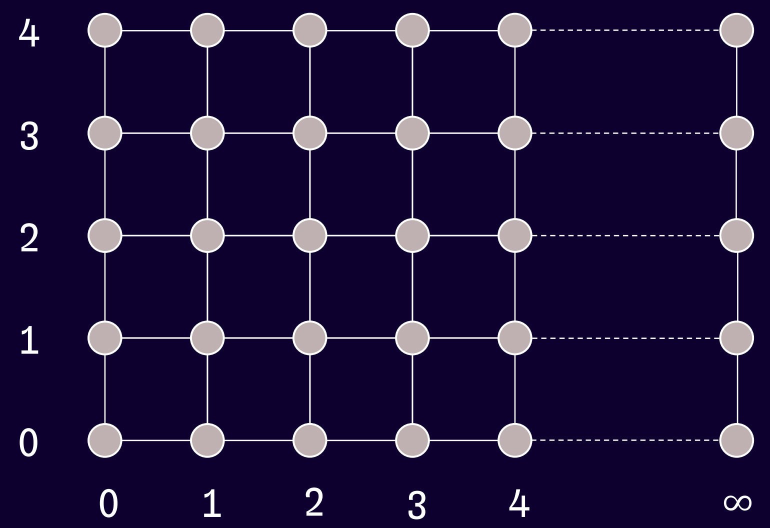
- Точки сетки $\mathbb{Z}_p^2$
- Бесконечно удаленные точки $(\infty, 0) \dots (\infty, p-1)$

Прямые:

- Обычные $\left\{(x, ax+b)\colon x \in \mathbb{Z}_p\right\} \cup \{(\infty, a)\}$

Утверждение:

Любые 2 прямые пересекаются ровно в 1 точке



КПП: конструкция

Точки:

- Точки сетки $\mathbb{Z}_p^2$
- Бесконечно удаленные точки $(\infty, 0) \ldots (\infty, p-1)$

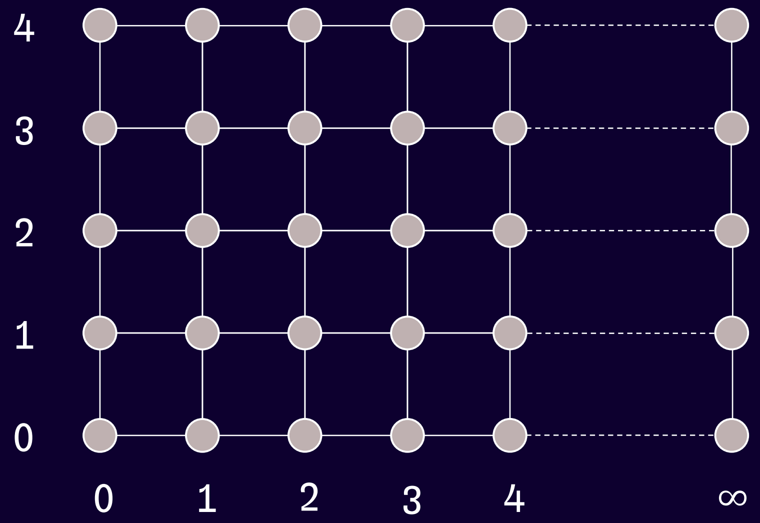
Прямые:

- Обычные $\{(x, ax+b) : x \in \mathbb{Z}_p\} \cup \{(\infty, a)\}$

Утверждение:

Любые 2 прямые пересекаются ровно в 1 точке

$$\begin{cases} a_1 x + b_1 = y \\ a_2 x + b_2 = y \end{cases}$$



КПП: конструкция

Точки:

- Точки сетки $\mathbb{Z}_p^2$
- Бесконечно удаленные точки $(\infty, 0) \ldots (\infty, p-1)$

Прямые:

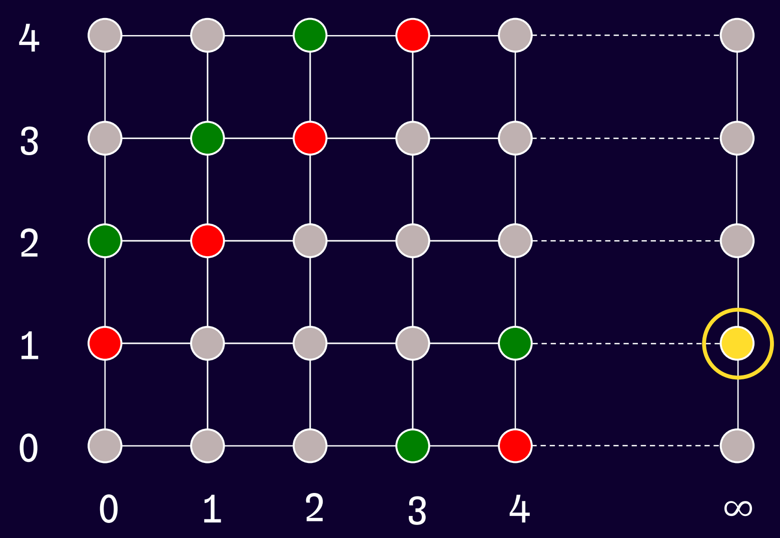
- Обычные $\{(x, ax+b) : x \in \mathbb{Z}_p\} \cup \{(\infty, a)\}$

Утверждение:

Любые 2 прямые пересекаются ровно в 1 точке

$$\begin{cases} a_1 x + b_1 = y \\ a_2 x + b_2 = y \end{cases}$$

Если $a_1 = a_2$: пересечение по (∞, a_1)



$y = 1x + 1$

$y = 1x + 2$

КПП: конструкция

Точки:

- Точки сетки $\mathbb{Z}_p^2$
- Бесконечно удаленные точки $(\infty, 0)$ … $(\infty, p-1)$

Прямые:

- Обычные $\{(x, ax+b) : x \in \mathbb{Z}_p\} \cup \{(\infty, a)\}$

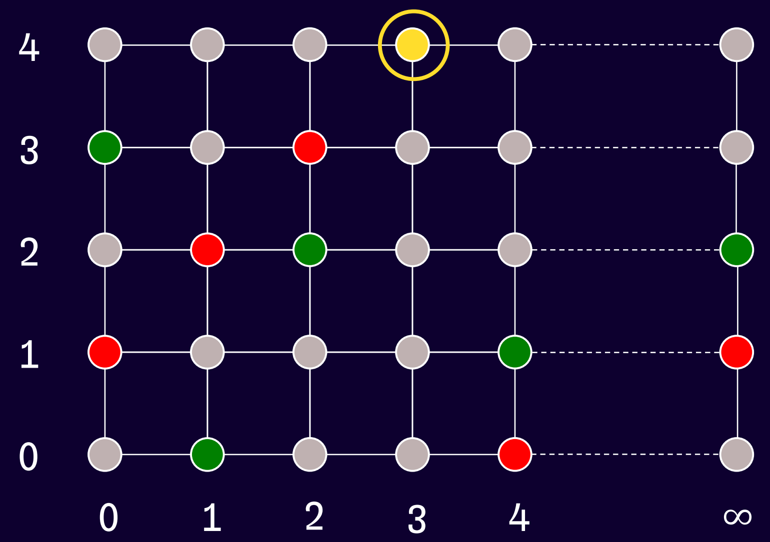
Утверждение:

Любые 2 прямые пересекаются ровно в 1 точке

$$\begin{cases} a_1 x + b_1 = y \\ a_2 x + b_2 = y \end{cases}$$

Если $a_1 \neq a_2$:

$$x = \frac{b_2 - b_1}{a_1 - a_2}, \qquad y = a_1 x + b_1$$



$y = 1x + 1$
$y = 2x + 3$

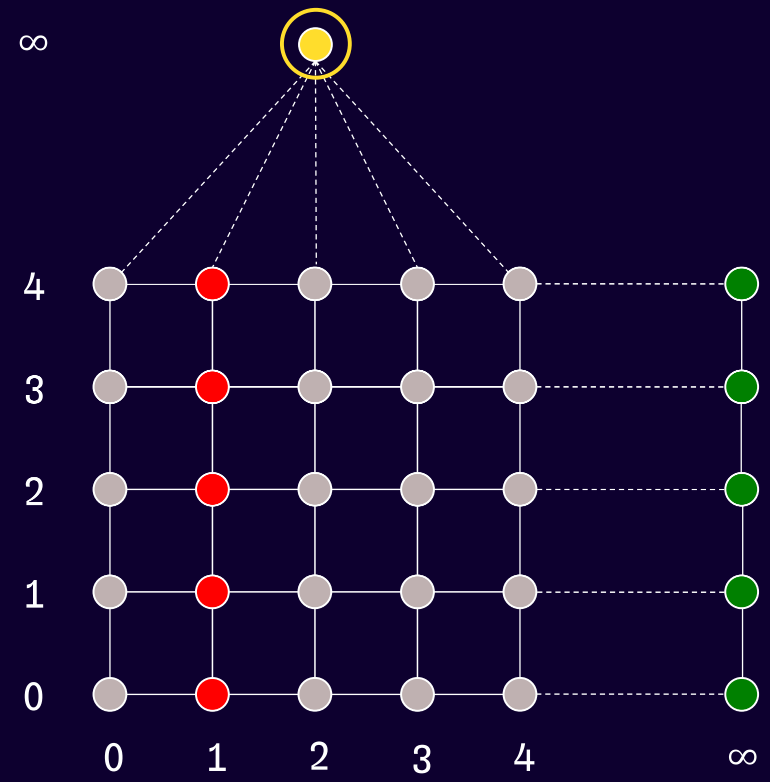
КПП: конструкция

Точки:

- Точки сетки $\mathbb{Z}_p^2$
- Бесконечно удаленные точки $(\infty, 0) \dots (\infty, p-1)$
- Вертикальная бесконечно удаленная точка $(1, \infty)$

Прямые:

- Обычные $\{(x, ax+b) \colon x \in \mathbb{Z}_p\} \cup \{(\infty, a)\}$
- Вертикальные $\{(a, y) \colon y \in \mathbb{Z}_p\} \cup \{(1, \infty)\}$
- Бесконечно удаленная $\{(\infty\,, y) \colon y \in \mathbb{Z}_p\} \cup \{(1, \infty)\}$



$x = 2$

БУ прямая

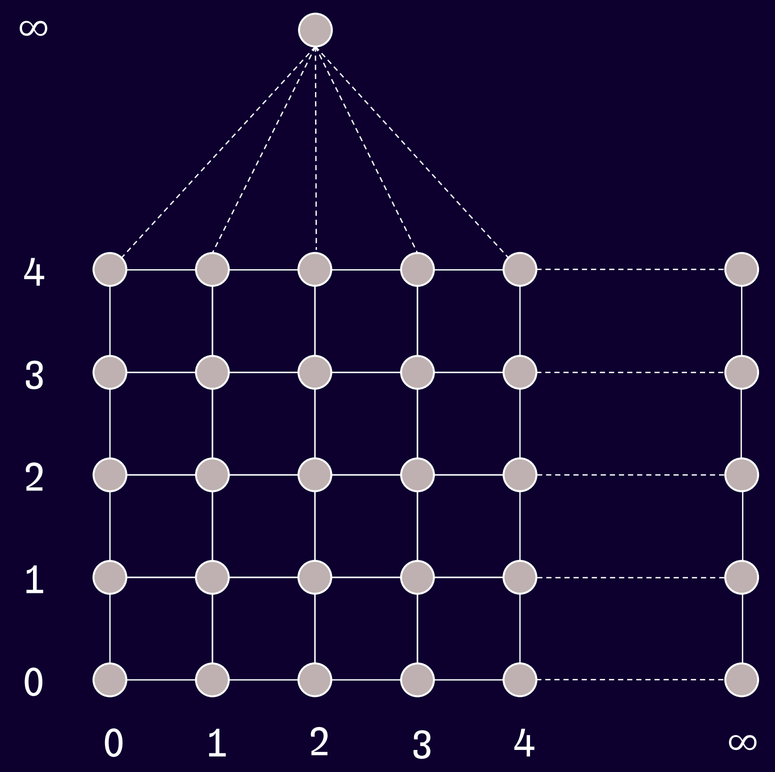
КПП: свойства

- Количество точек: $p^2 + p + 1$
- Количество прямых: $p^2 + p + 1$

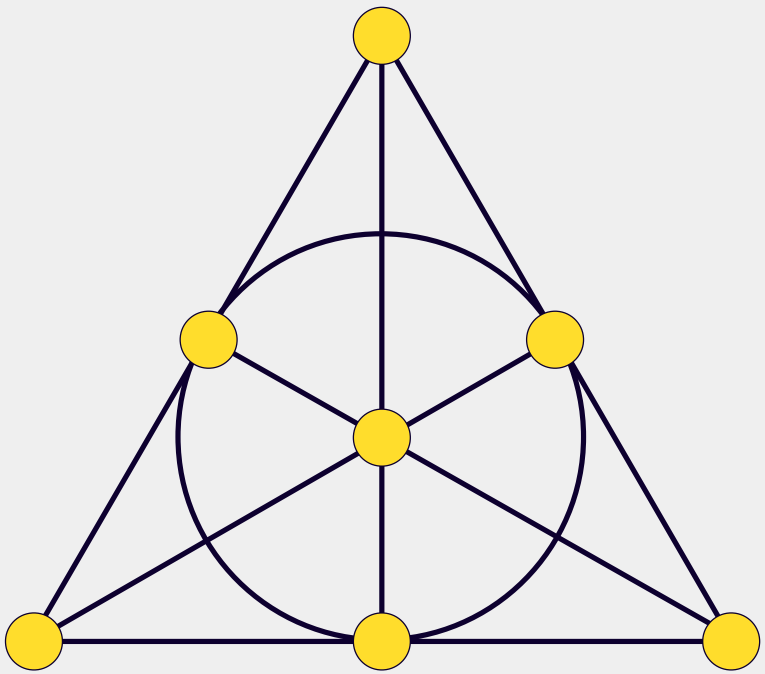
- В каждой прямой $p + 1$ точка
- Каждая точка содержится в $p + 1$ прямой

- Любые 2 прямые пересекаются ровно в 1 точке
- Через любые 2 точки проходит ровно 1 прямая

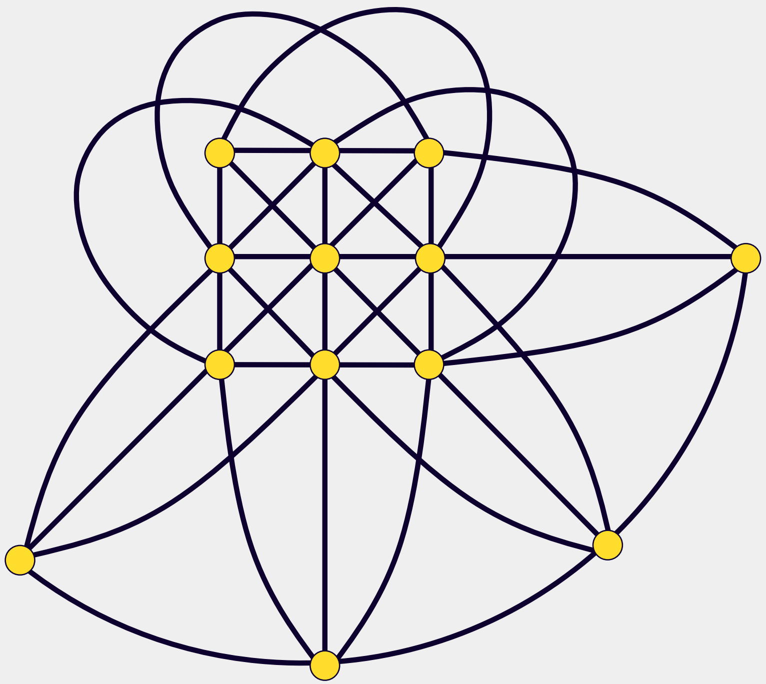
- Является симметричным семейством



КПП: примеры

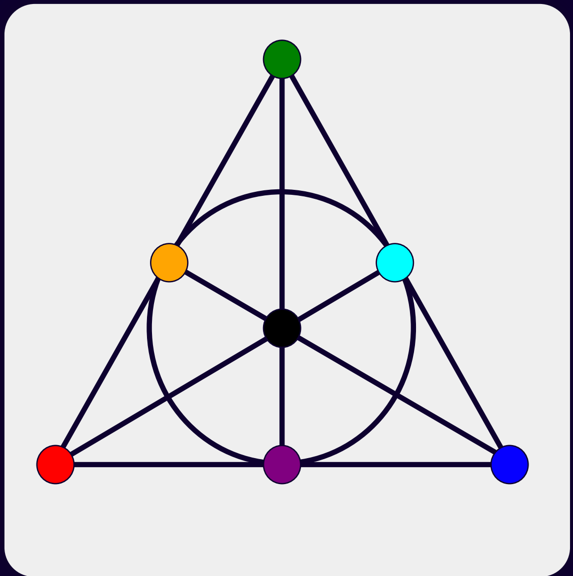


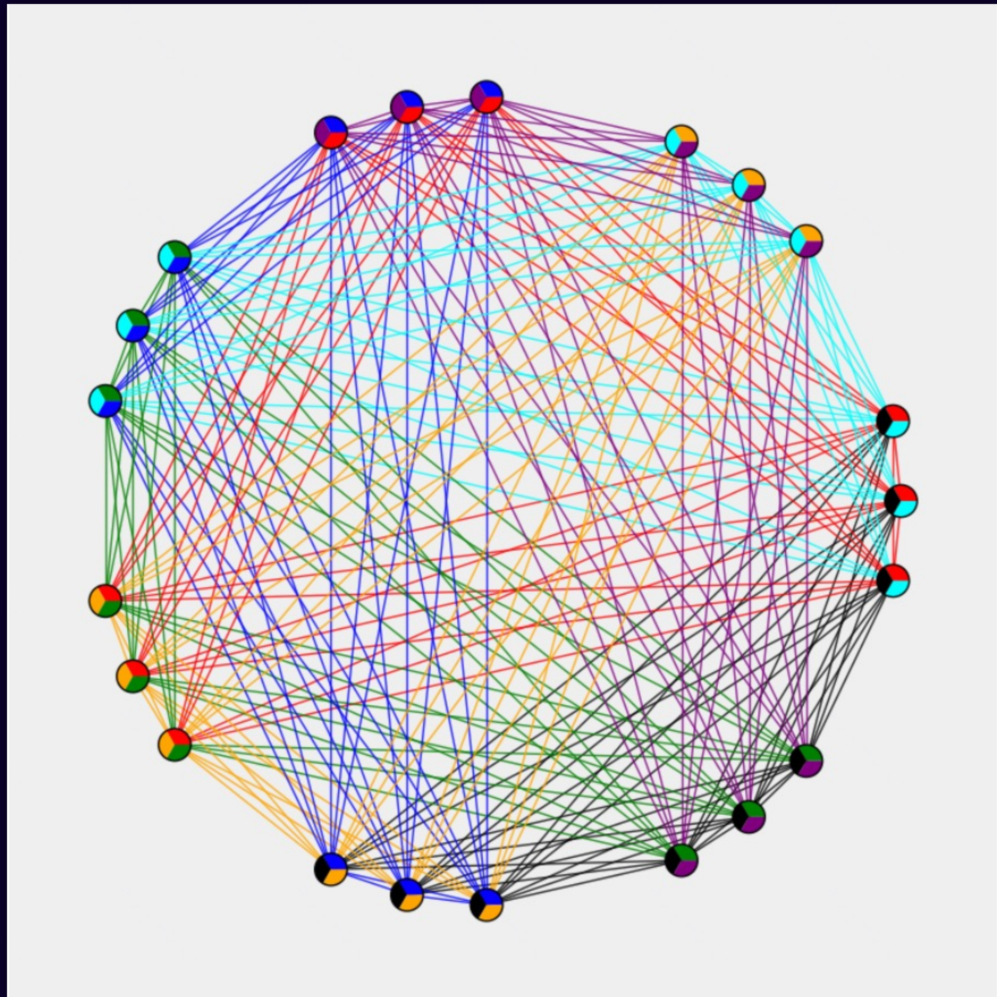
p = 2



p = 3

Разбиение через КПП





Сбалансированные пересекающиеся системы

Вводим более слабую версию симметричных пересекающихся семейств

- Задаем произвольные вероятности множеств и правила выбора пересечений
- Ослабляем условие на симметричность

Получаем оптимальное в худшем случае (для клик) разбиение

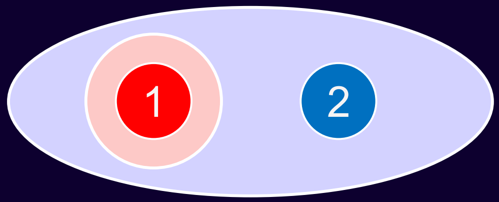
Теорема
Для любого k существует сбалансированная система попарно пересекающихся множеств $F_1 \dots F_r$ т. ч.

$$\frac{|\mathrm{F}_1| + \cdots + |\mathrm{F}_r|}{r} \leq \sqrt{k}(1 + o(1))$$

Следствие
Существует алгоритм реберного разбиения с соответствующими гарантиями.

Теорема
Для любого k сбалансированные пересекающиеся системы дают оптимальное разбиение **полного графа.**

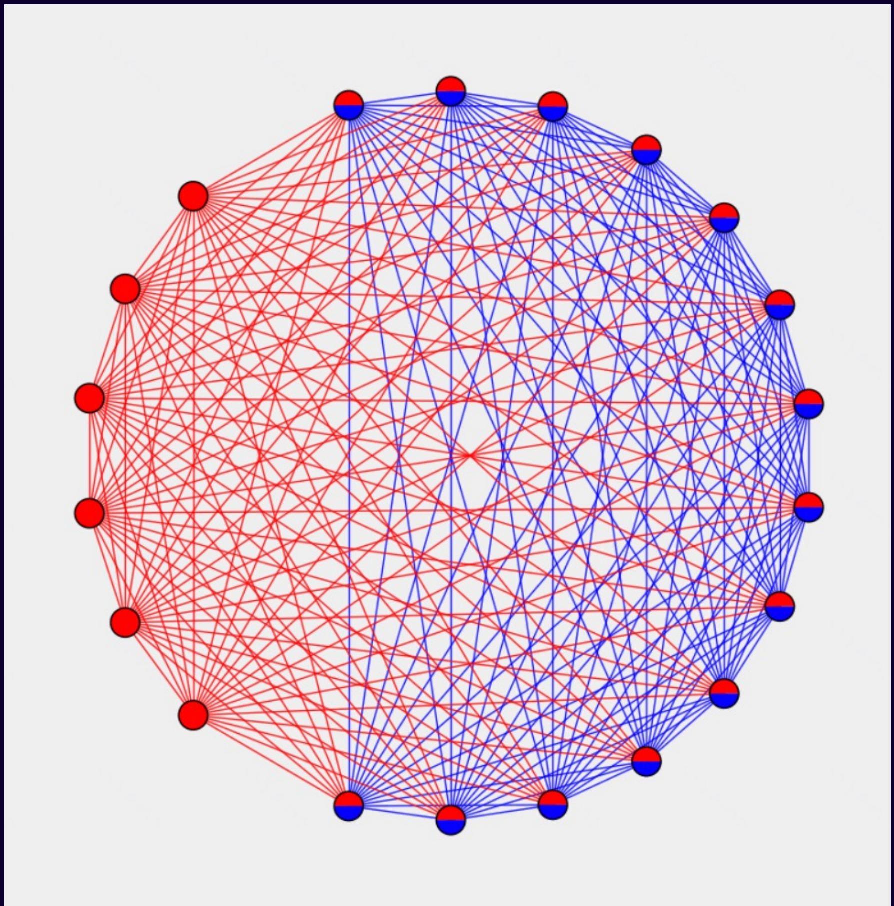
Пример: k=2



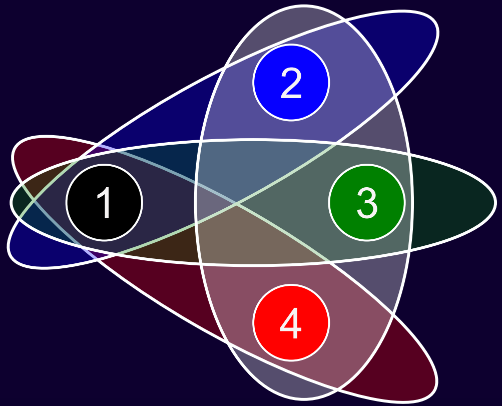
- $\mathcal{F} = \{\{1\}, \{1, 2\}\}$
- $\{1, 2\}$ берем в вероятностью $1/\sqrt{2}$
- Берем 2 всегда когда можно

Получаем BIS со средним размером множеств

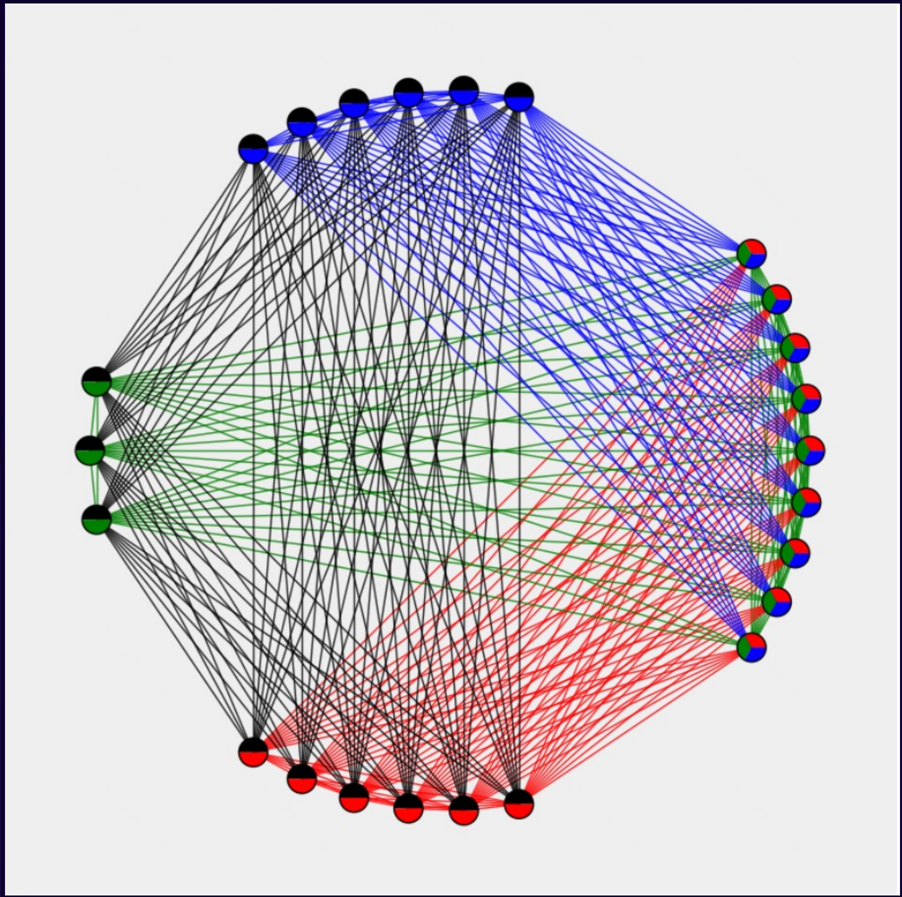
$$1 + 1/\sqrt{2}$$



Пример: k=4



Вероятности множеств	**Самопересечения**
12 – 1/4	12 – 2
13 – 1/8	13 – 3
14 – 1/4	14 – 4
234 – 3/8	234 – 3



Спасибо!