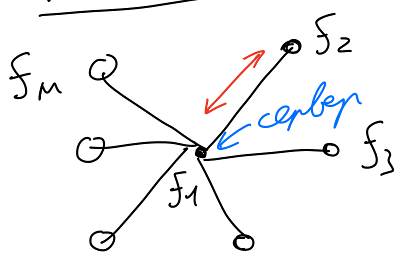
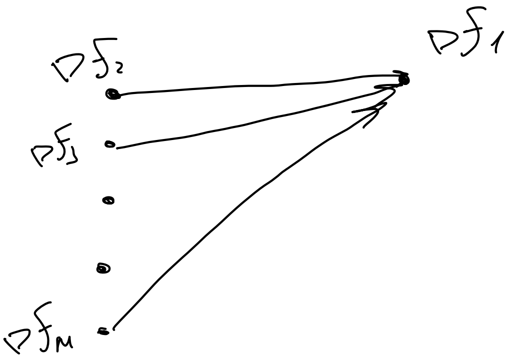
• До этого

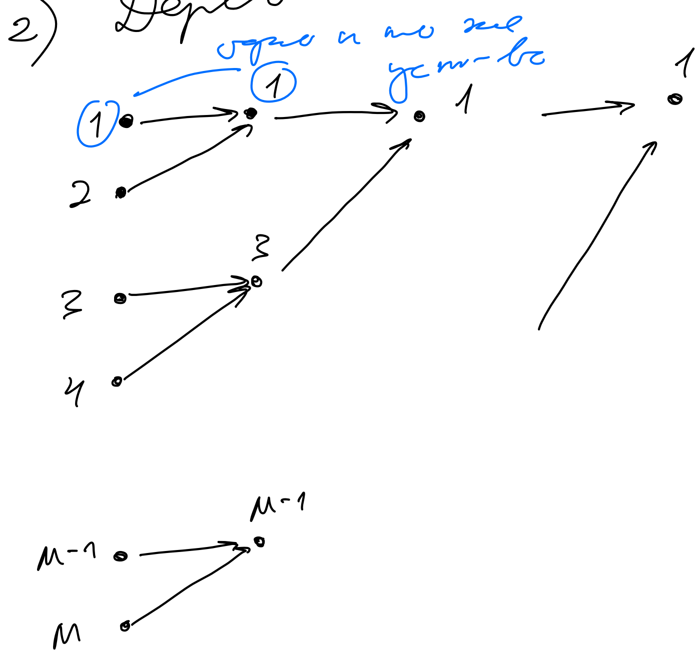


1) Базовый – „сервер"



нагрузка ~ M

2) Дерево



нагрузка на 1 уст-во

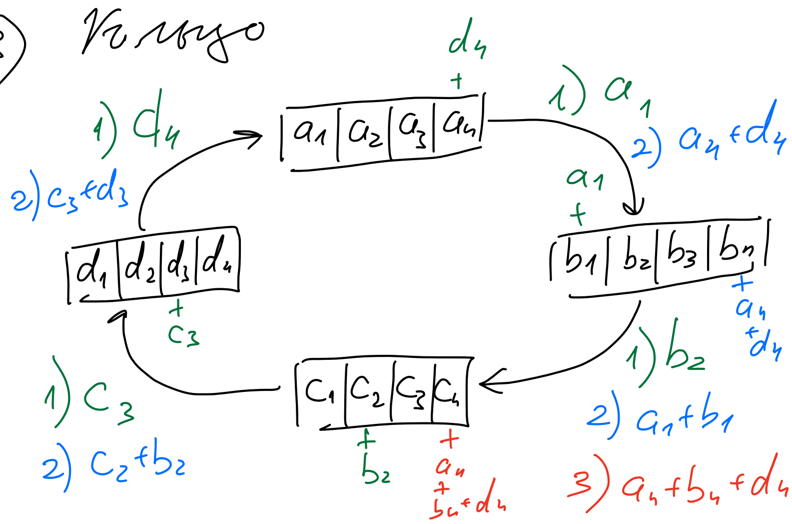
$\log_2 M$

проблема:

⊖ неравномерность использ. устройств

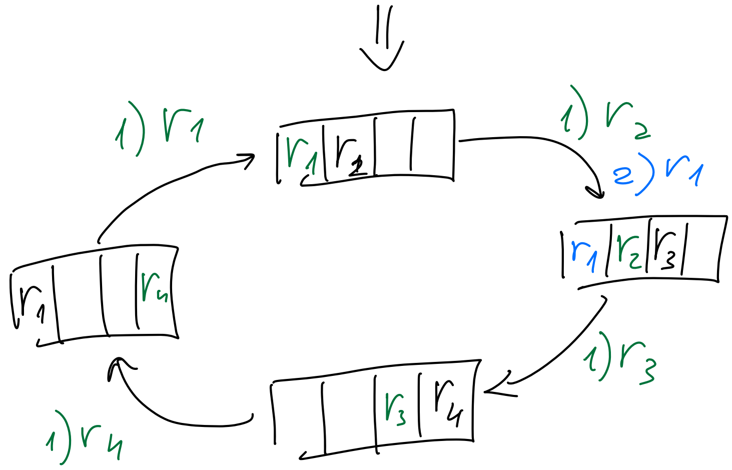
⊖ зависимость от M

3) Кольцо



нагрузка

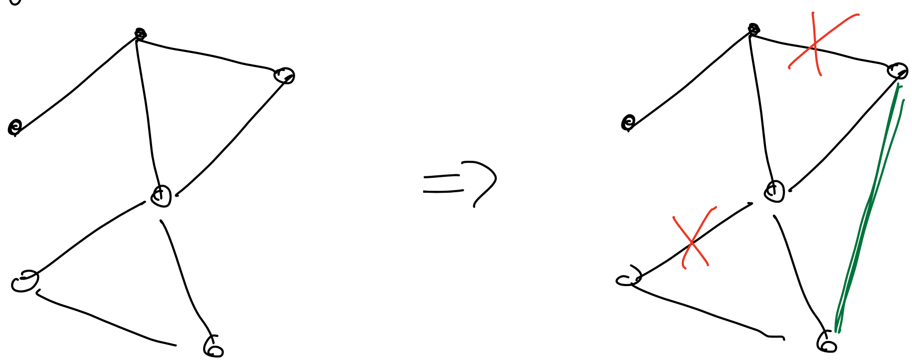
$O(1)$ для всех устройств



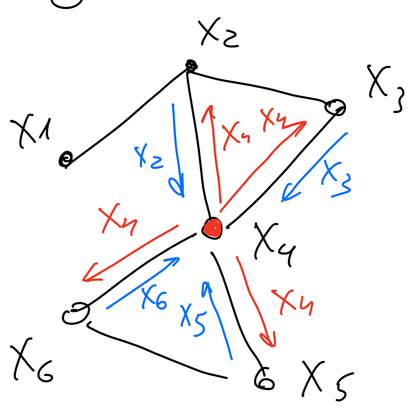
подгрузка

$O(1)$ для всех устройств

Что делать, если ребра в графе связи неустойчивы



gossip / диффузия



$$x_4^k \to x_4^{k+1} = \frac{1}{5}\left(x_3^k + x_5^k + x_6^k + x_2^k + x_4^k\right)$$

$$x_4^k \to x_4^{k+1} = \alpha x_4^k + \beta_1 x_3^k + \beta_2 x_5^k + \beta_3 x_6^k + \beta_4 x_2^k$$

$$\alpha + \beta_1 + \beta_2 + \beta_3 + \beta_4 = 1 \ ?$$

стохастичность

веса в матрицу

M – mixing матрица

$x^{k+1} = Mx^k$ ← итерации gossip

удобная запись

Пример

$$M_{ij} = \begin{cases} 0 & i \not\leftrightarrow j \\ \dfrac{1}{N(i)+1} & i \leftrightarrow j \end{cases}$$

Что потребовать для сходимости к среднему?

1) связность графа (можно ослабить на период. связность)
2) $M^T = M$ (можно отказаться)
3) стохастичность $\qquad M\underline{1} = \underline{1} \qquad \underline{1}^T M^T = \underline{1}^T$

$\underline{1}$ — ед. вектор $\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$

Важно $\quad x^k \to \bar{x} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \frac{1}{M} \sum_{m=1}^{M} x_m^0 \quad$ (среднее к чему стремимся)

$$\bar{x} = \frac{1}{M} \underline{1}^T x^0 \cdot \underline{1}$$

Хотим $\quad \bar{x}^k = \bar{x}^{k+1}$ (среднее сохраняется)

$\bar{x}^k = \bar{x}^0$ (и равняться начальному)

$$\bar{x}^{k+1} \overset{\text{среднее}}{=} \frac{1}{M} \underline{1}^T x^{k+1} \underline{1} \overset{\substack{\text{gossip} \\ x^{k+1} = Mx^k}}{=} \frac{1}{M} \underline{1}^T M x^k \underline{1}$$

$$\overset{M^T = M}{=} \frac{1}{M} \underline{1}^T M^T x^k \underline{1} \overset{\text{стох.}}{=} \frac{1}{M} \underline{1}^T x^k \cdot \underline{1} = \bar{x}^k \quad \blacksquare$$

$\bar{x}^k$ — среднее сохраняется

• Сходимость к среднему?

$$\underbrace{\|x^{(k+1)} - \bar{x}^0\|_2}_{\text{расстояние до среднего вектора}} = \|Mx^k - \bar{x}^0\|_2 = \|Mx^k - M\bar{x}^0\|_2$$

$M\bar{x}^0 = \bar{x}^0$

$$= \|M(x^k - \bar{x}^0)\|_2 \overset{?}{<} \|x^k - \bar{x}^0\|_2$$

расстояние до среднего уменьшится за 1 шаг.

зависит от матрицы M

$$\le \|M\|_2 \|x^k - \bar{x}^0\|_2$$

спектральная

зависимость от сингулярных чисел (от с.з. M)

$$\le \lambda_{max}(M) \|x^k - \bar{x}^0\|_2$$

с.з. $M \in [-1; 1]$

$\lambda_{max}(M) = 1$

$$= \|x^k - \bar{x}^0\|_2$$

сходимости нет :-(

Но это не проблема

$\lambda(M) = 1$ соответствует собственный вектор $= 1$

- $(x^k - \bar{x}^0) \parallel \mathbb{1} \Rightarrow x^k \parallel \mathbb{1}$

  $\parallel \mathbb{1}$ $\qquad\qquad$ $\Downarrow$

  x^k уже идеально усреднен

- $(x^k - \bar{x}^0) \nparallel \mathbb{1}$

  $(x^k - \bar{x}^0) \in (\text{span}(\mathbb{1}))^\perp$ мы знаем

  на пространстве $\lambda_{max}(M) < 1$

$$\|x^{k+1} - \bar{x}^0\|_2 \le \underbrace{\lambda_{max}(M)}_{<1} \|x^k - \bar{x}^0\|_2$$

поэтому $\bar{x}^k \to \bar{x}^0$

Как сделать с оптимизацией?

1) $x_m^{k+1/2} = x_m^k - \gamma \nabla f_m(x^k)$ ← локальный шаг

   $X^{k+1} = M^T X^k$ ← gossip за большое число Т найдем ≈ среднее

2) $x_m^{k+1/2} = x_m^k - \gamma \nabla f_m(x^k)$

   $X^{k+1} = M X^k$ ← одна итерация

проблема → не сходится из-за $\|\nabla f_m(x^*)\|_2^2$

3) Gradient Tracking

$$\nabla f_m(x^k) \to \nabla f_m(x^*)$$

$$\nabla f_m(x^k) \to y_m^k$$

алгоритм:

$$x_m^{k+1/2} = x_m^k - \gamma y_m^k$$

$$x^{k+1} = M x^{k+1/2}$$

$$y_m^{k+1/2} = y_m^k + \nabla f_m(x_m^{k+1}) - \nabla f_m(x_m^k)$$

$$y^{k+1} = M y^{k+1/2}$$

$$x_m^{k+1} = M\left(x_m^k - \gamma M y_m^{k-1} - \gamma M \nabla f_m(x_m^k) + \gamma M \nabla f_m(x_m^k)\right)$$

$$x^k + M(x^k - x^{k-1}) - \gamma \nabla f(x^k) + \gamma \nabla f(x^{k-1})$$