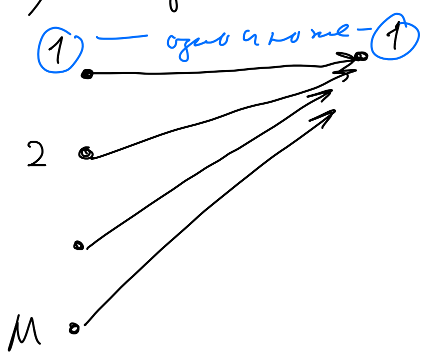
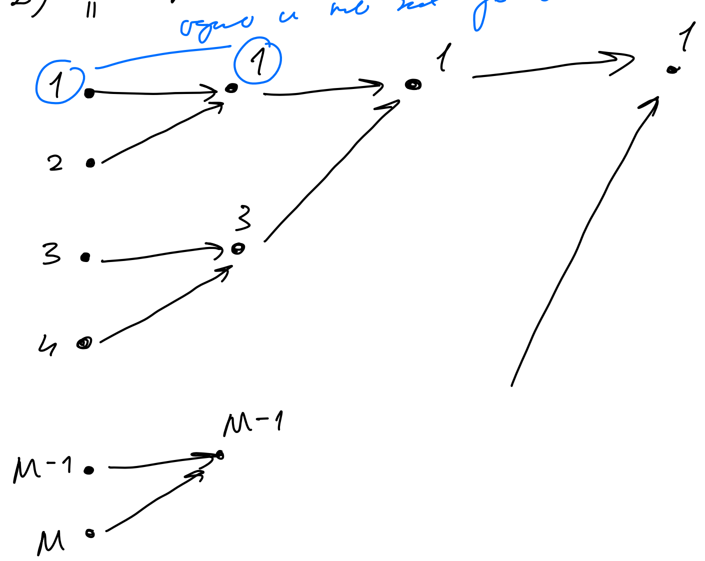
• Способы организации общения

1) Базовый – „сервер"



нагрузка ~ M

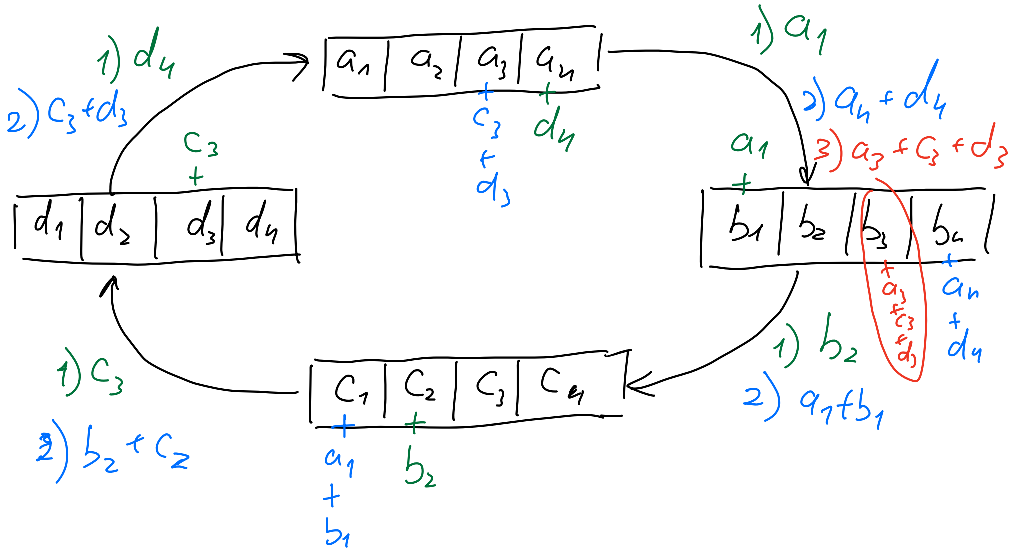
2) „Дерево"



нагрузка на 1ое ус-во
$\log_2 M$

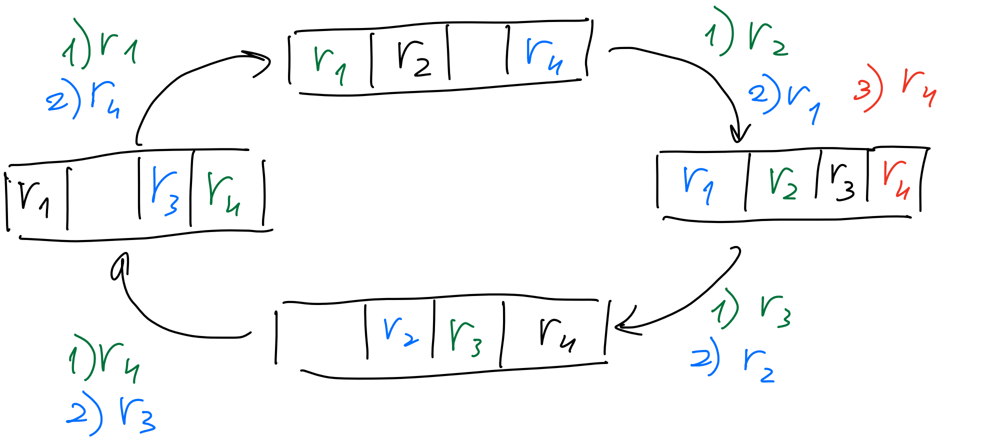
проблемы:
⊖ зависимость от M
⊖ неравномерность исп. устройств

3) „Кольцо"



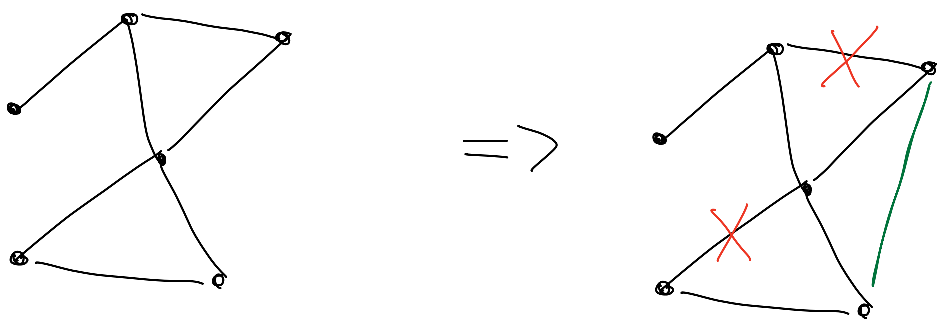
нагрузка
$O(1)$
на все
устройства

⇓

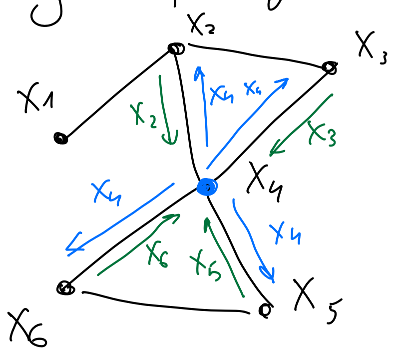


пересылка $O(1)$ на устройства

Что делать, если связь (ребра в графе) между устройствами неустойчивы



gossip / диффузии



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

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

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

в виде матрицы

удобно $\begin{cases} M \text{ — mixing матрица} \\ x^{k+1} = Mx^k \end{cases}$ ← итерация gossip в матричном виде

- <u>Сходимость gossip процедуры</u>

Предположения на граф и M:

1) связность графа (можно ослабить на периодич. связность)

2) $M^T = M$ (можно ослабить)

3) стохастичность M $\qquad M\underline{1} = \underline{1}$ $\qquad \underline{1}^T M^T = \underline{1}^T$

вектор из 1

Хотим усреднить: $x^k \to \bar{x} = \underbrace{\left(\frac{1}{M}\sum_{m=1}^{M} x_m^0\right)\cdot \underline{1}}_{\frac{1}{M}\,\underline{1}^T x^0 \cdot \underline{1}}$

В процессе мы хотим $\bar{x}^k = \bar{x}^{k+1}$ (среднее сохранялось)

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

$$\underset{\text{среднее}}{\bar{x}^{k+1} =} \frac{1}{M}\,\underline{1}^T x^{k+1}\cdot \underline{1} \underset{\text{gossip}}{=} \frac{1}{M}\,\underline{1}^T M x^k \cdot \underline{1}$$

$$\underset{M^T = M}{=} \frac{1}{M}\,\underline{1}^T M^T x^k\, \underline{1} \underset{\text{стох. } M}{=} \frac{1}{M}\,\underline{1}^T x^k \cdot \underline{1} = \underset{\text{среднее сохранилось}}{\bar{x}^k}$$

- Сходимость к среднему?

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

$M\bar{x}^0 = \bar{x}^0$ (стох.)

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

уменьшение расстояния

зависит от матрицы

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

спектральная

хотим $\|M\|_2 < 1$ (зависит от сингуляр. числа, в нашем случае от $\lambda_{max}(M)$)

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

$\lambda_{max}(M) = 1$ для этих матриц

$\lambda(M) \in (-1;1)$

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

сходимости нету :-(

Но это не проблема:

$\lambda(M) = 1$ соответствует с.в. $= \underline{1}$

$x^k - \bar{x}^0$? ← на это действует оператор

- $(x^k - \bar{x}^0) \parallel 1 \Rightarrow x^k \parallel 1$ $\Downarrow$ x^k уже усреднен

  ($\bar{x}^0$ ↑ $\parallel 1$)

  gossip не нужен — все усреднено

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

  $(x^k - \bar{x}^0) \in (\mathrm{span}(1))^\perp$

  ↑ вот здесь $|\lambda_{max}(M)| < 1$

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

поэтому $x^k \to \bar{x}^0$

⊕ линейная сходимость

⊖ процедура неточная

⊙ зависимость от свойств графа и M

- Как связать с оптимизацией?

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

$$(x_1^{k+1} \ldots x_m^{k+1}) = \underset{\text{с помощью gossip}}{\text{усредняя}} (x_1^{k+1/2} \ldots x_m^{k+1/2}, T)$$

T — число итераций (больше, тем лучше точность усредн.)

⊕ как будто бы параллел. метод

⊖ накопление ошибки усреднения

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

$$(x_1^{k+1} \ldots x_m^{k+1}) = \underset{\text{с помощью gossip}}{\text{усредняя}} (x_1^{k+1/2} \ldots x_m^{k+1/2}, 1)$$

⊕ лучше с точки зрения коммуникаций

⊖ метод не сходится до точного решения из-за $\|\nabla f_m(x^*)\|_2$

3)
$$\nabla f_m(x^k) \to \nabla f_m(x^*) \to = 0$$ только в „усредн." формате

усреднять не только x, но ∇f

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

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

x_m^{k+1} — с помощью gossip 1 итерации

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

y_m^{k+1} — с помощью gossip 1 итерации

$x_m^k \to x^*$

$$y_m^{k+1/2} = y_m^k + \underbrace{\nabla f_m(x_m^{k+1})}_{\to \nabla f_m(x^*)} - \underbrace{\nabla f_m(x_m^k)}_{\to \nabla f_m(x^*)}$$

y_m^{k+1} — замешивается к среднему

Gradient Tracking

(+) сходится до точного решения

(−) общение по 2 векторам

Все, что обсуждалось: $M \to M(k)$

на меняющуюся M