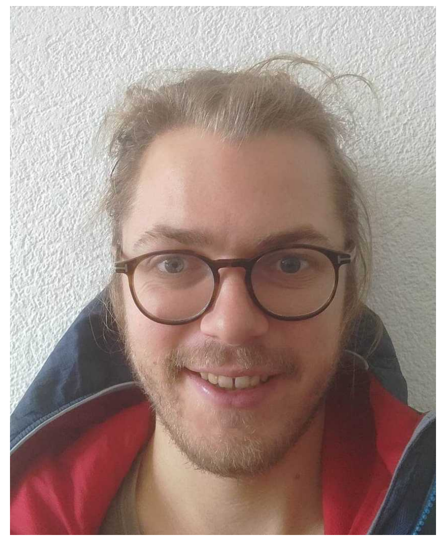
Achieving Optimal Breakdown for Byzantine Robust Gossip

Renaud Gaucher

Thoth Seminar - October 2024

Aymeric Dieuleveut



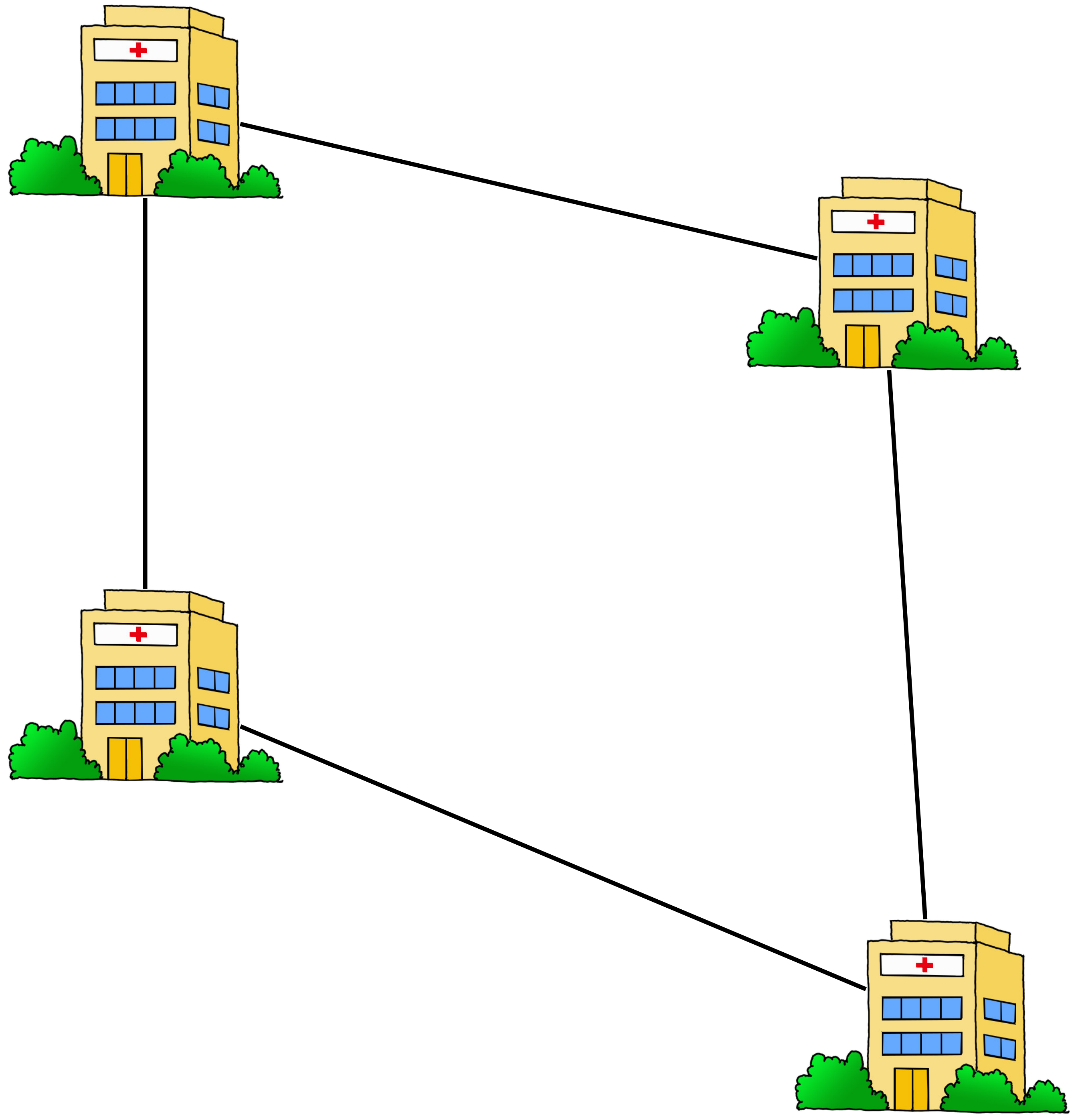
Hadrien Hendrikx

Distributed Optimization in Machine Learning

Number of nodes in the network

$$\min_{x \in \mathbb{R}^d} f(x) = \sum_{i=1}^{m} f_i(x)$$

Local loss of node i



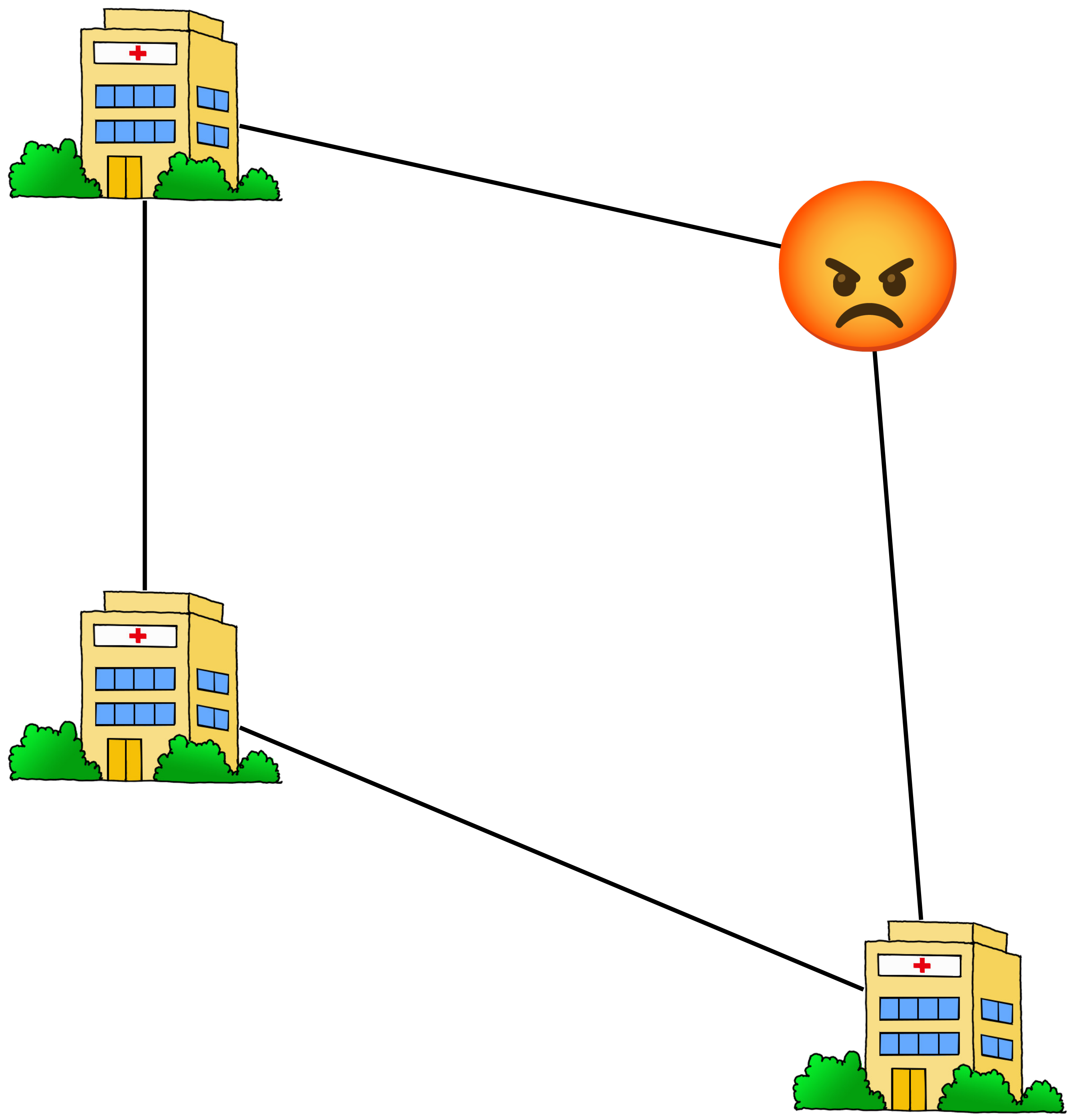
- Each node has only access to a local parameter and his local loss function
- Nodes collaborate to find a global objective

Byzantine Distributed Optimization

Some *unknown* units are Byzantine - malicious and omniscient adversaries

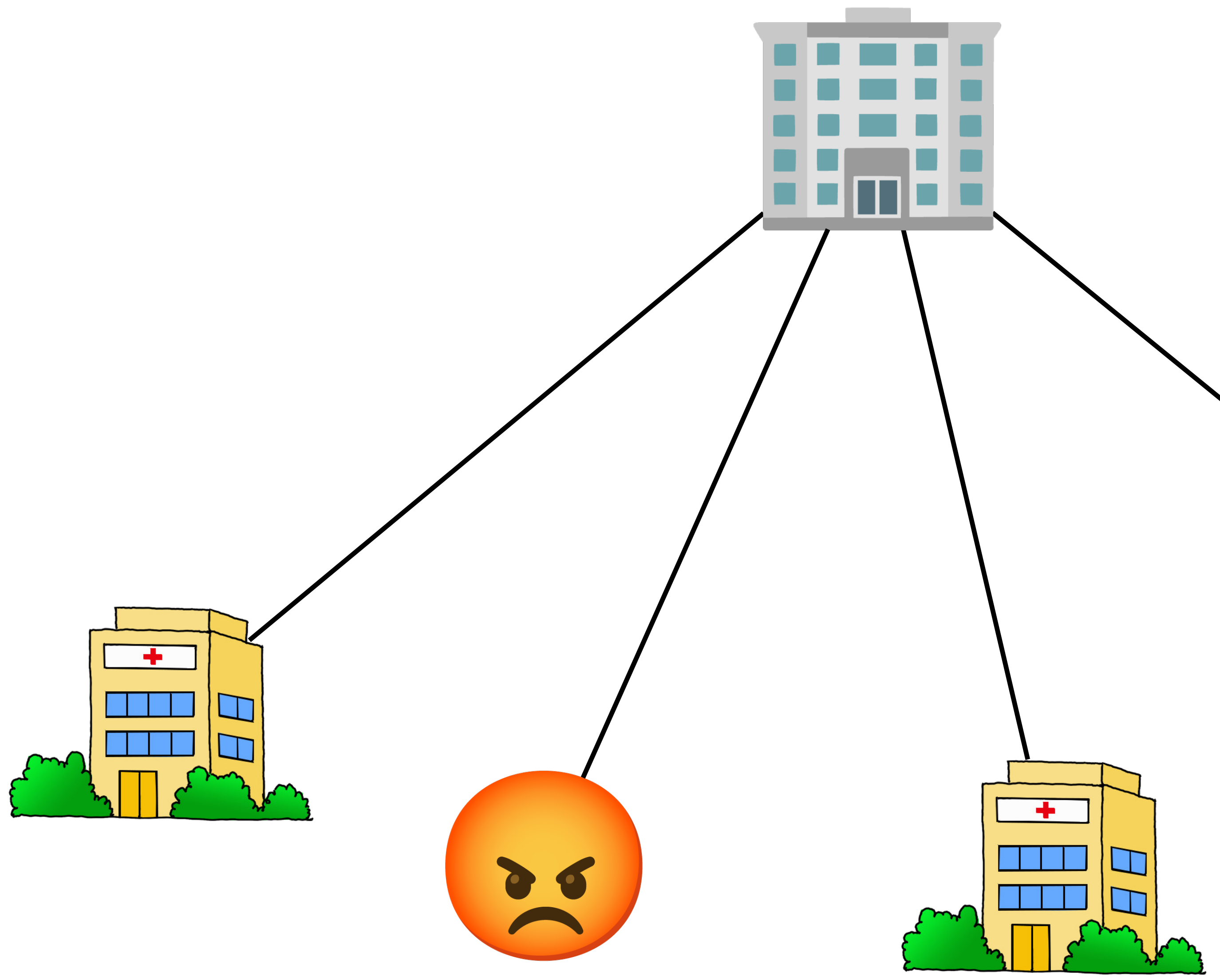
$$\min_{x \in \mathbb{R}^d} f_h(x) = \sum_{i\ honest} f_i(x)$$

Honest nodes only



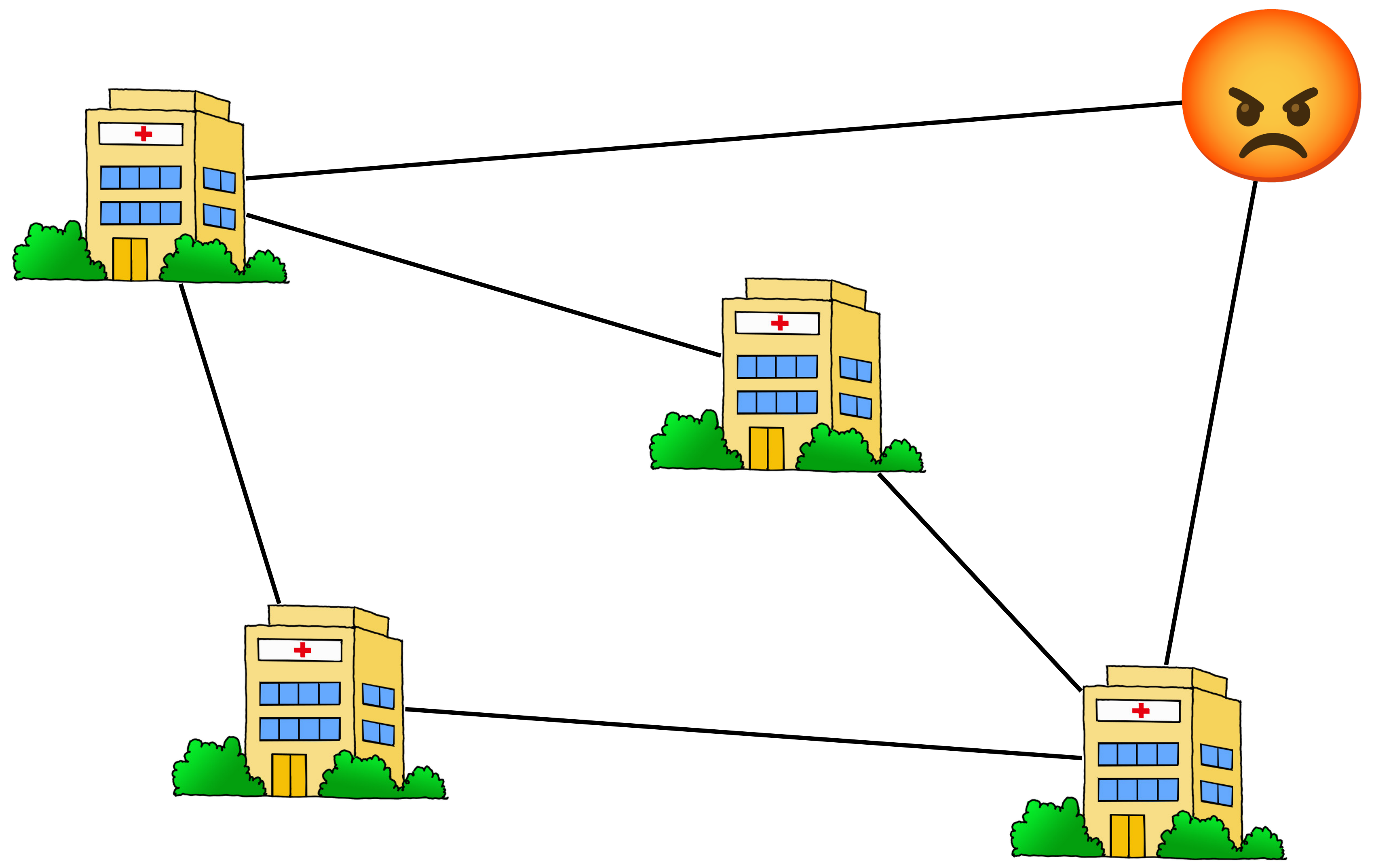
Communication model

Federated



- All nodes connected to a trusted central server

Decentralized



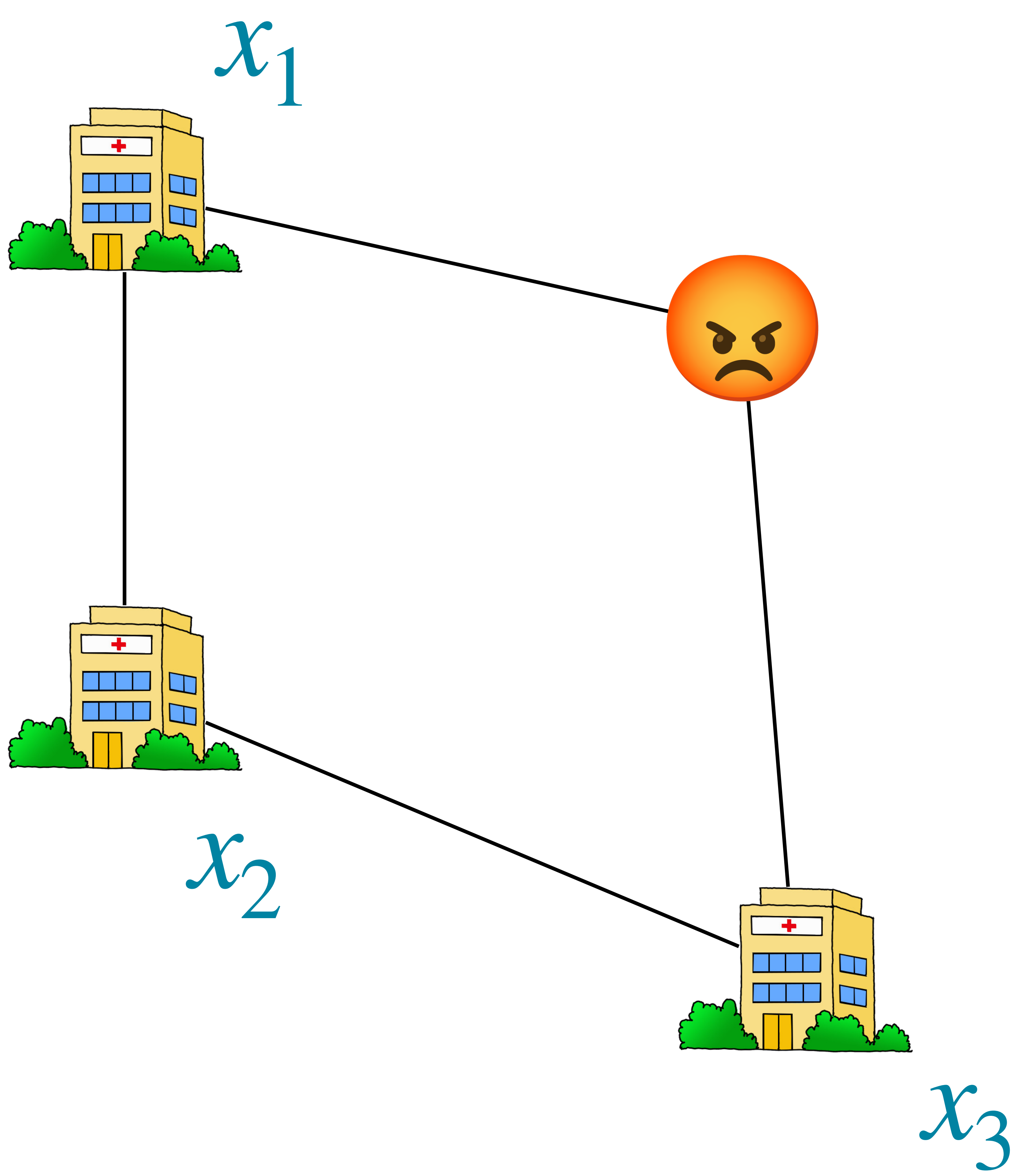
- Nodes linked by a communication graph

Sub-problem: Robust distributed averaging

Honest nodes holds an initial parameter x_i

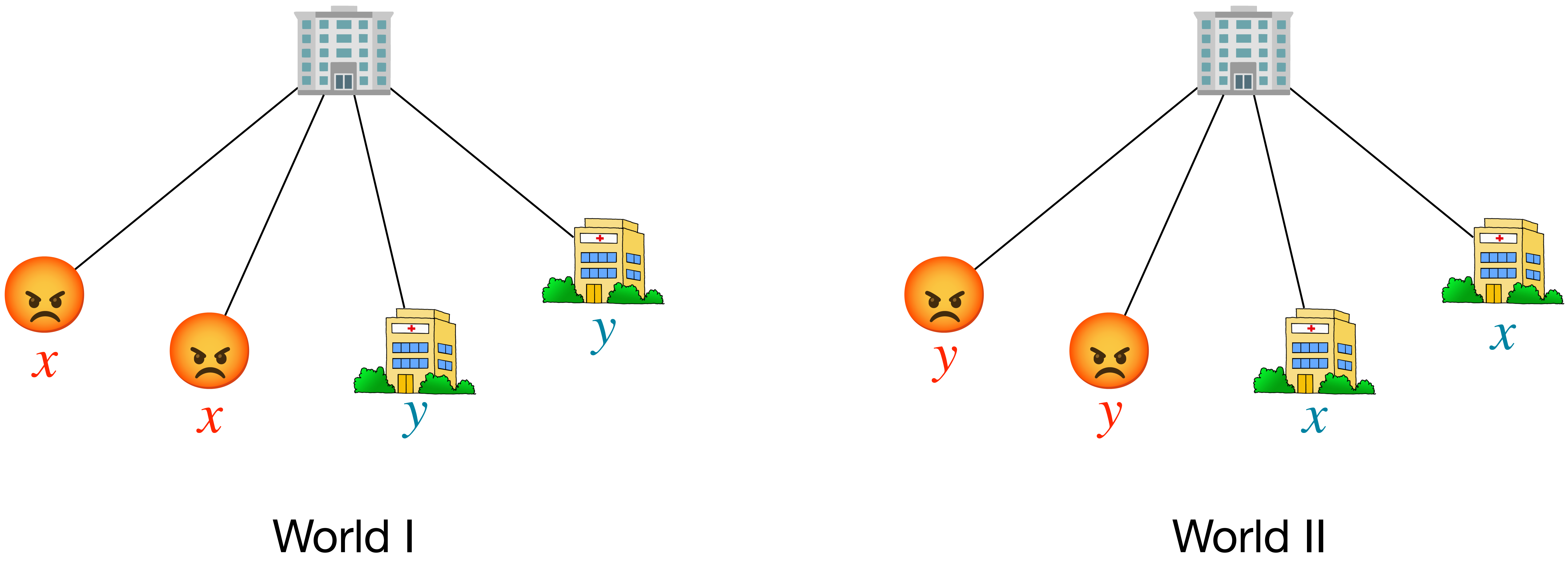
Finding the average $\frac{1}{h}\sum_{i\ honest} x_i$ boils down to solving

$$\min_{x\in\mathbb{R}^d} f_h(x) = \sum_{i\ honest} \|x - x_i\|^2$$



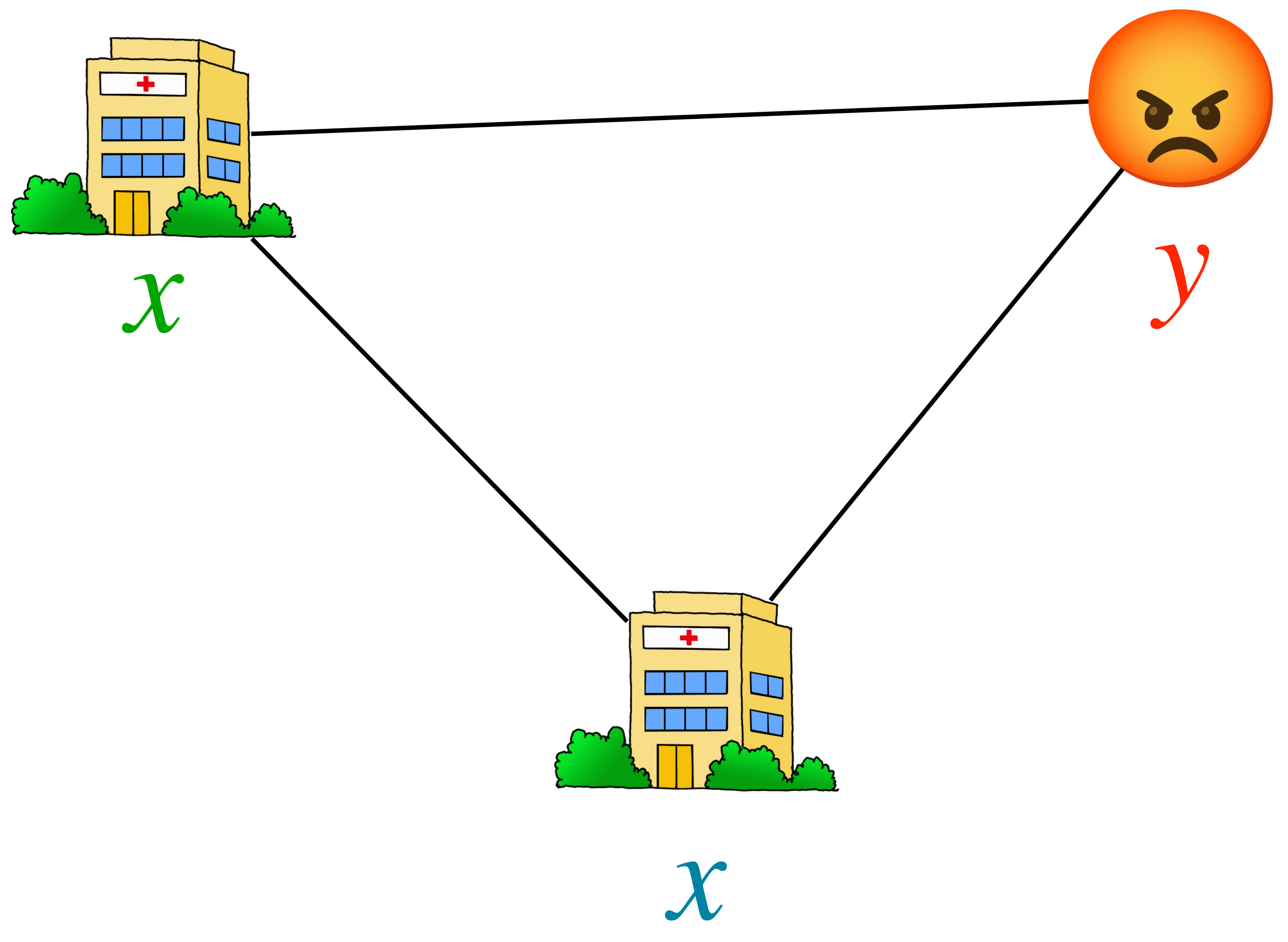
Maximal proportion of Byzantine

Federated: no robustness possible if more than 1/2 of Byzantines



Maximal proportion of Byzantine

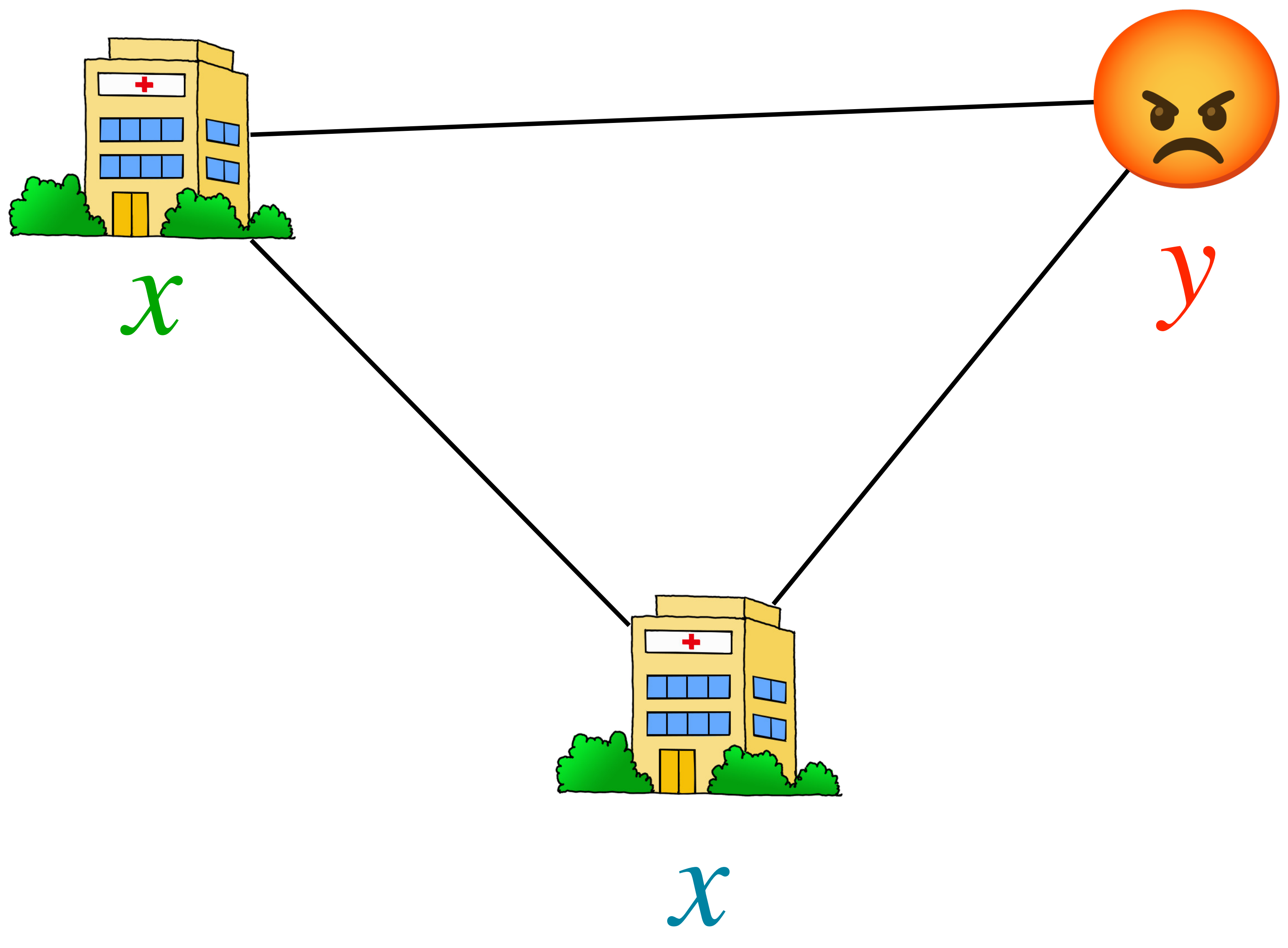
Decentralized: no robustness possible if more than 1/3 of Byzantines



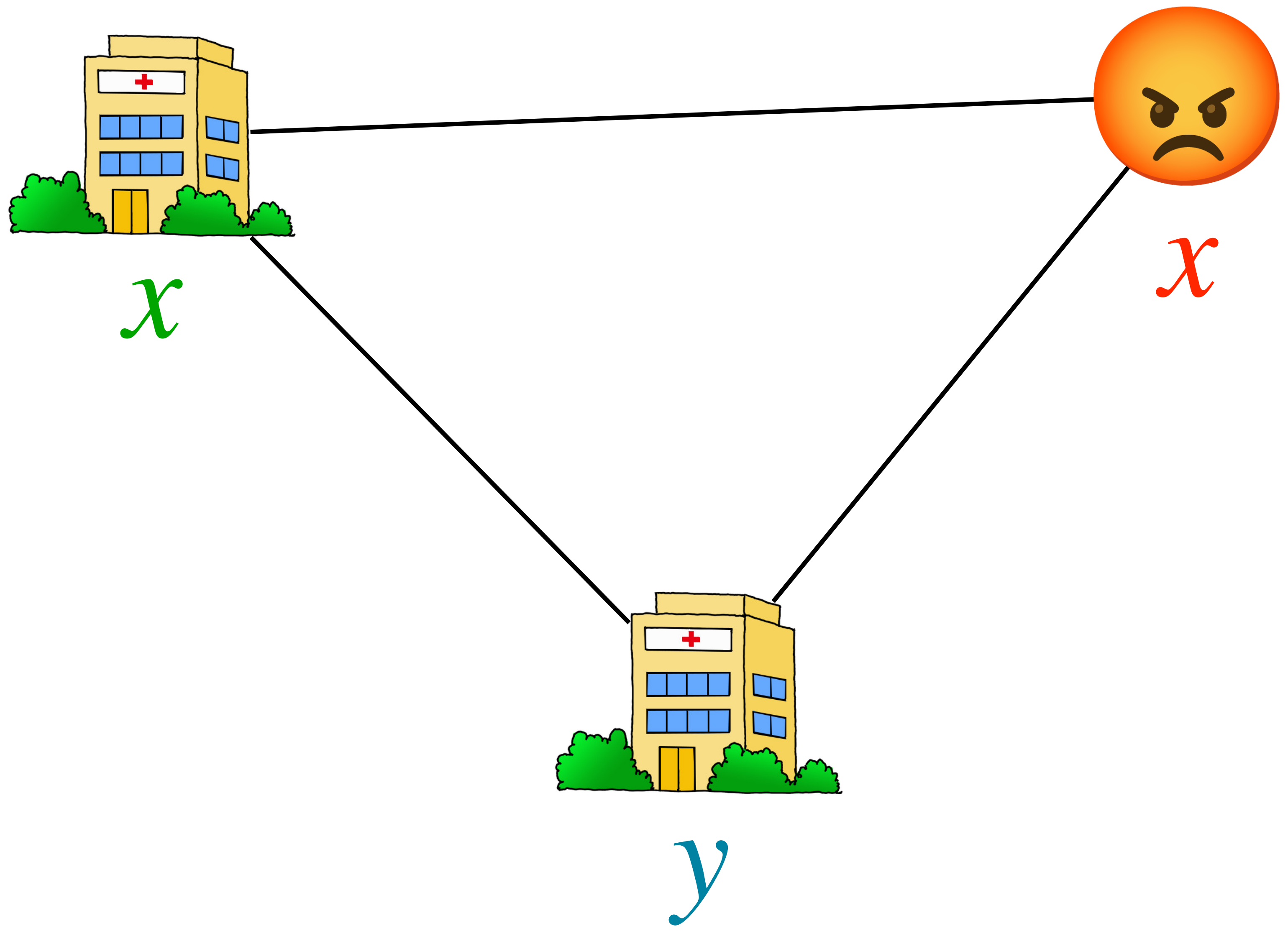
World I

Maximal proportion of Byzantine

Decentralized: no robustness possible if more than 1/3 of Byzantines



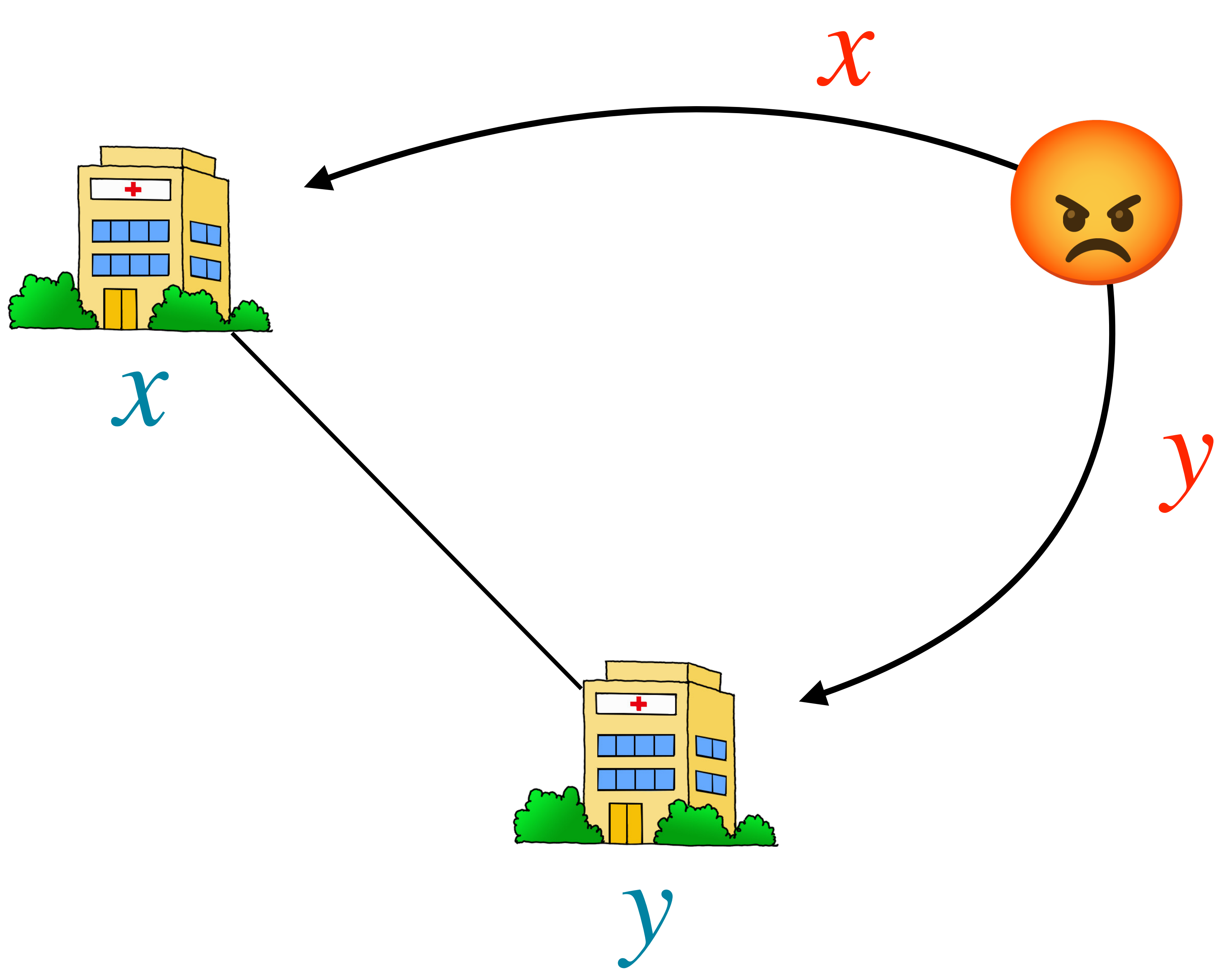
World I



World II

Maximal proportion of Byzantine

Decentralized: no robustness possible if more than 1/3 of Byzantines



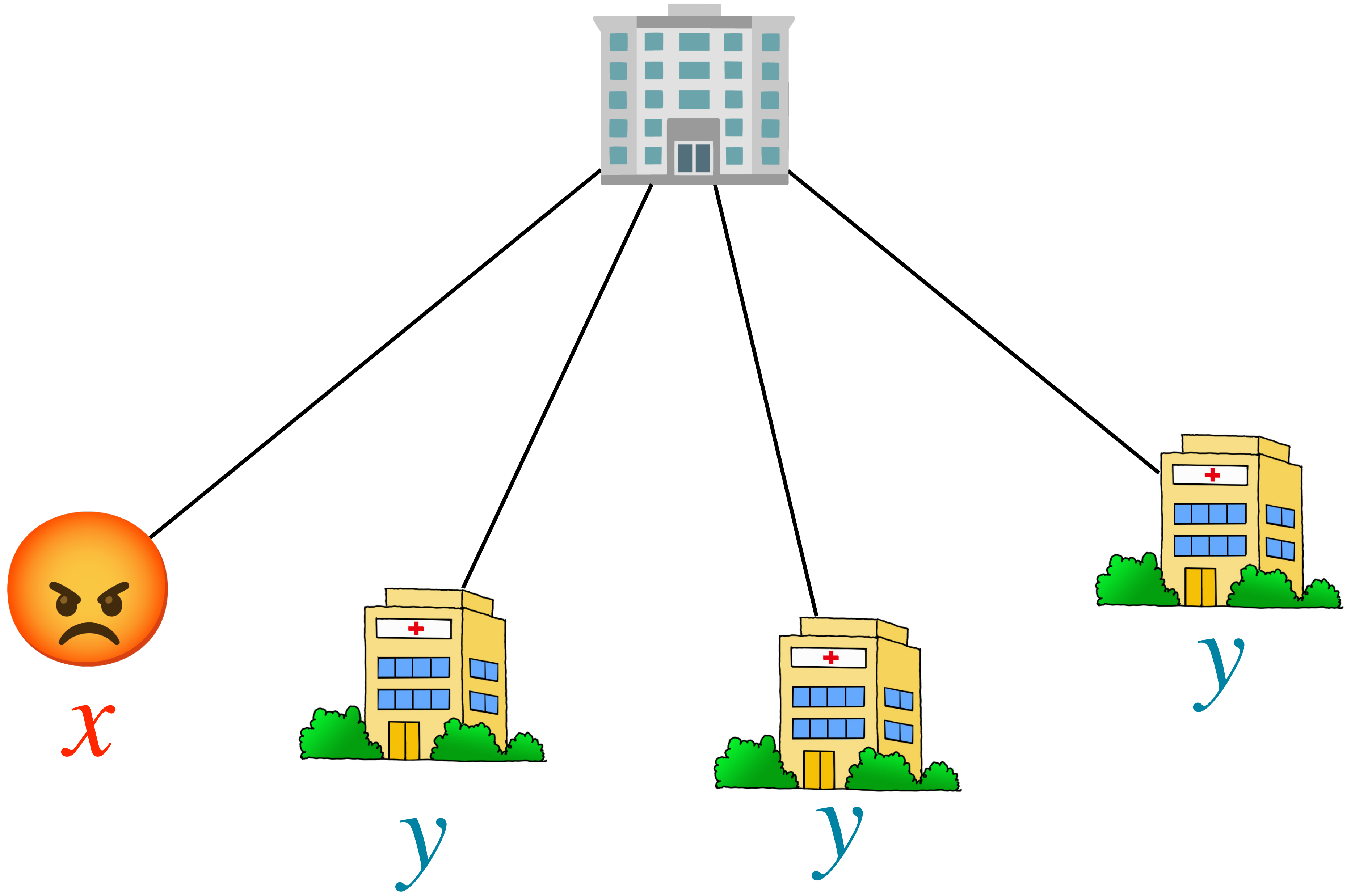
Breakdown points

- Federated communication: 1/2 of Byzantine units
- Decentralized communication: 1/3 of Byzantine units *for fully connected graphs*

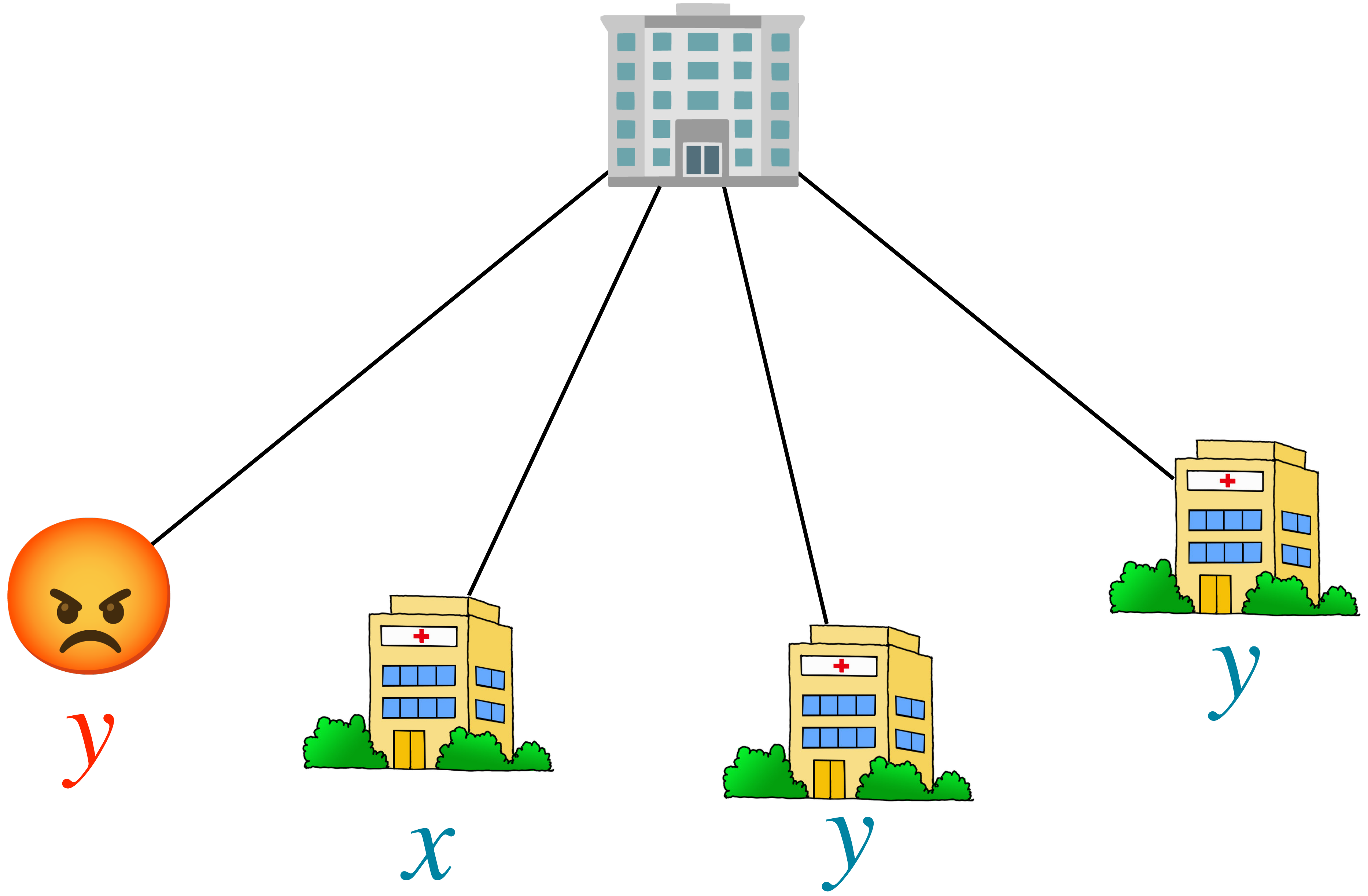
→ No (satisfying) link between the connectivity of the graph and the breakdown point !

Only approximate solutions are reachable

World I



World II



Only approximate solutions are reachable

In the federated setting, under

$$\frac{1}{h}\sum_{i\ honest} \|\nabla f_i(x) - \nabla f_h(x)\|^2 \leq V^2 + H^2\|\nabla f_h(x)\|^2$$

the minimal achievable error is

$$\Omega\left(\frac{b}{h - (1 + H^2)b} \cdot V^2\right)$$

→ Dependence w.r.t graph quantities are still unclear

Breakdown points for arbitrary graphs

Gossip communication without Byzantine

Each node approximately average his neighbours parameters

$$x_i^{k+1} = x_i^k + \eta \sum_{j \sim i} \left(x_j^k - x_i^k \right)$$

Gossip communication without Byzantine

Each node approximately average his neighbours parameters

$$x_i^{k+1} = x_i^k + \eta \sum_{j \sim i} \left(x_j^k - x_i^k \right)$$

Using the graph's Laplacian matrix $W = D - A$

$$X^{k+1} = X^k - \eta W X^k$$

where

$$X^k = \begin{pmatrix} x_1^k \\ \vdots \\ x_h^k \end{pmatrix}$$

Gossip communication without Byzantine

Each node approximately average his neighbours parameters

$$x_i^{k+1} = x_i^k + \eta \sum_{j \sim i} \left(x_j^k - x_i^k \right)$$

Using the graph's Laplacian matrix $W = D - A$

$$X^{k+1} = X^k - \eta W X^k$$

where

$$X^k = \begin{pmatrix} x_1^k \\ \vdots \\ x_h^k \end{pmatrix}$$

Spectral properties give rates of convergence to the average: under $\eta = 1/\mu_h$

$$\|X^k - \overline{X}^0\|^2 \leq \left(1 - \frac{\mu_2}{\mu_h}\right)^k \|X^0 - \overline{X}^0\|^2$$

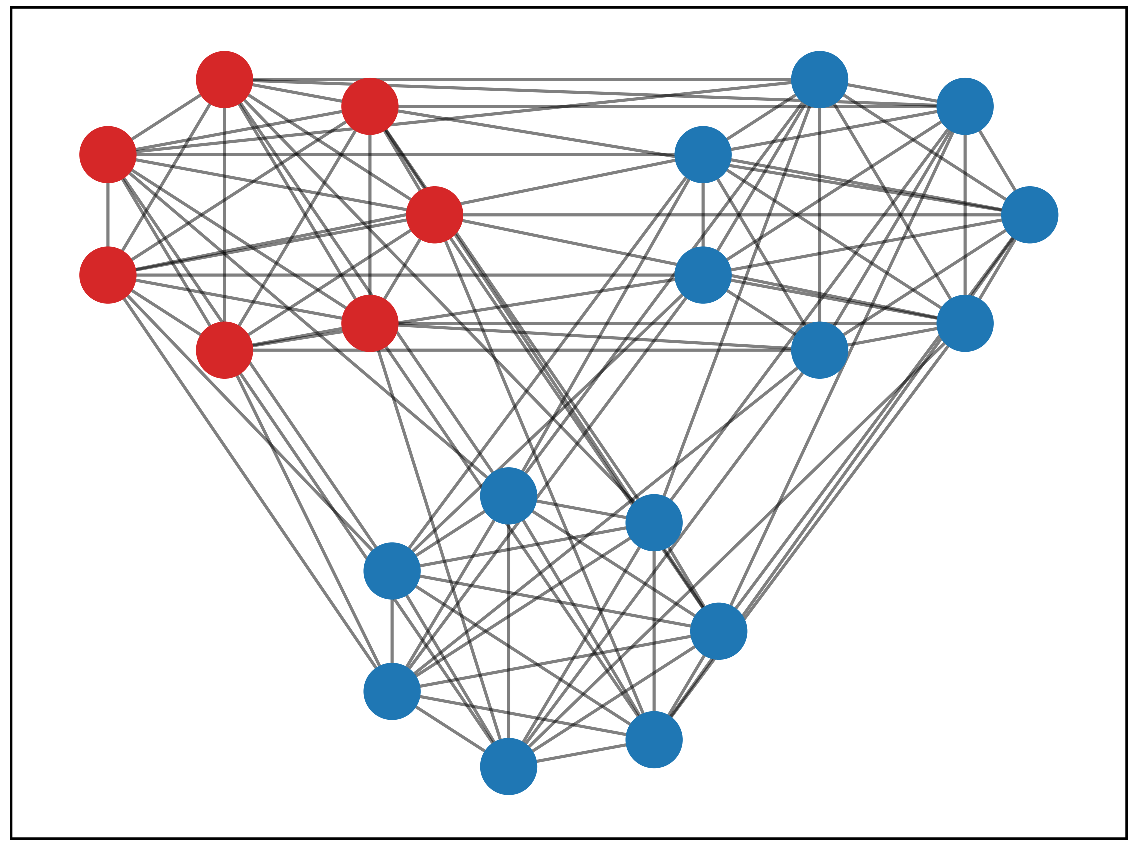
where

$$\overline{X}^0 = \begin{pmatrix} \bar{x}_h \\ \vdots \\ \bar{x}_h \end{pmatrix}$$

Breakdown point in arbitrary graphs

Theorem

For any $b \geq 0$, assume that honest nodes can have up to b Byzantine neighbours. Then for any $h \in 2\mathbb{N}$, $h \geq 2b$, there exists a graph with h honest nodes, and *algebraic connectivity* $\mu_2 = 2b$ on which no communication algorithm can be robust.

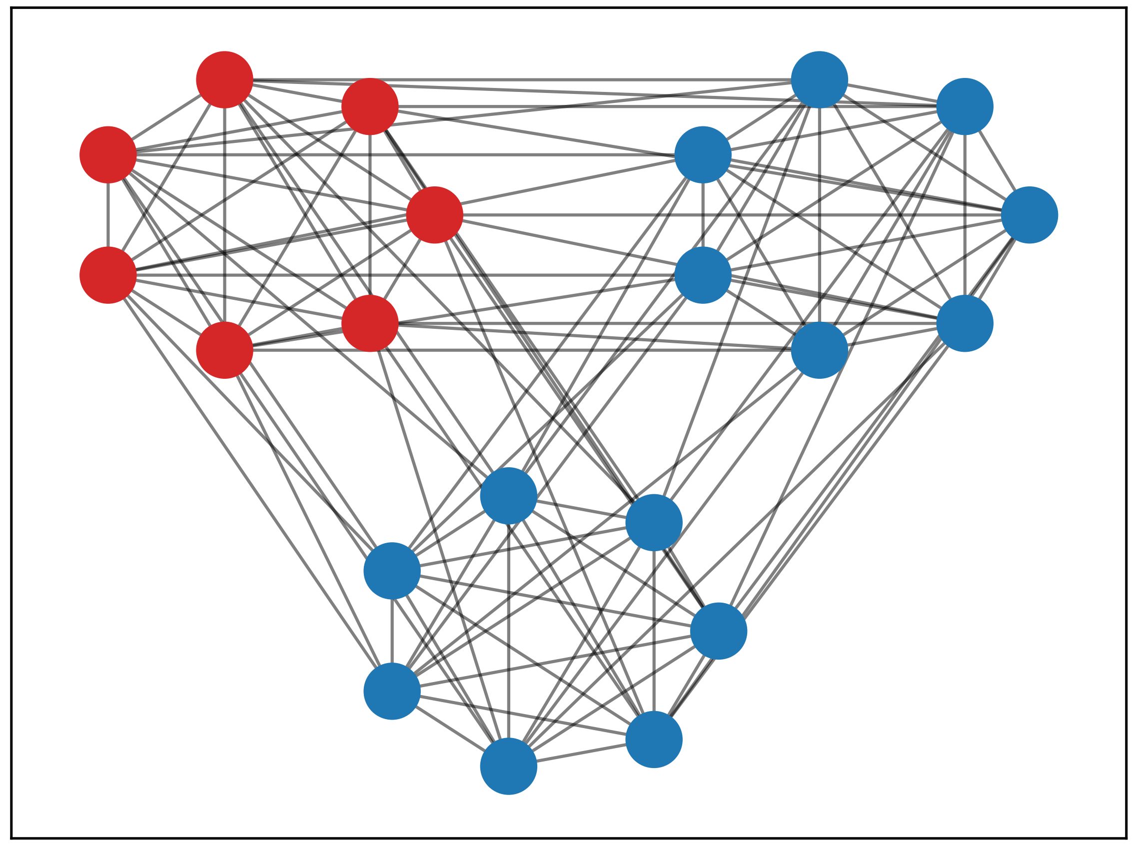


Robust algorithms on arbitrary graphs requires $2b \leq \mu_2$

Breakdown point in arbitrary graphs

Theorem

For any $b \geq 0$, assume that honest nodes can have up to b Byzantine neighbours. Then for any $h \in 2\mathbb{N}, h \geq 2b$, there exists a graph with h honest nodes, and *algebraic connectivity* $\mu_2 = 2b$ on which no communication algorithm can be robust.



Robust algorithms on arbitrary graphs requires $2b \leq \mu_2$

There exists a robust gossip-like algorithm robust when $2(b+1) \leq \mu_2$

Open questions

- Link between graph’s connectivity and achievable error ?
- Influence of Byzantines agents on the convergence rates ?

Thank you for your attention