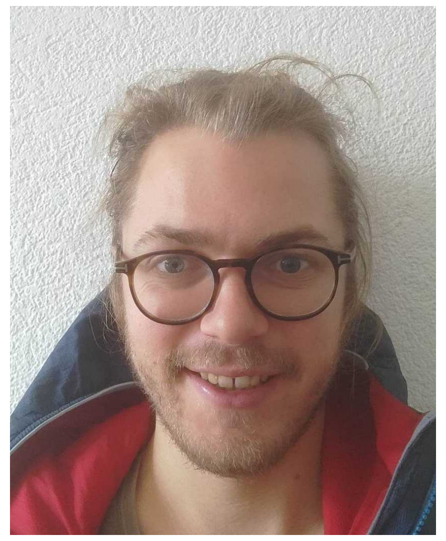
Adversarially Robust Distributed Optimization

A Unified Breakdown Analysis of Byzantine Robust Gossip

Renaud Gaucher

Redeem retreat
September 2025



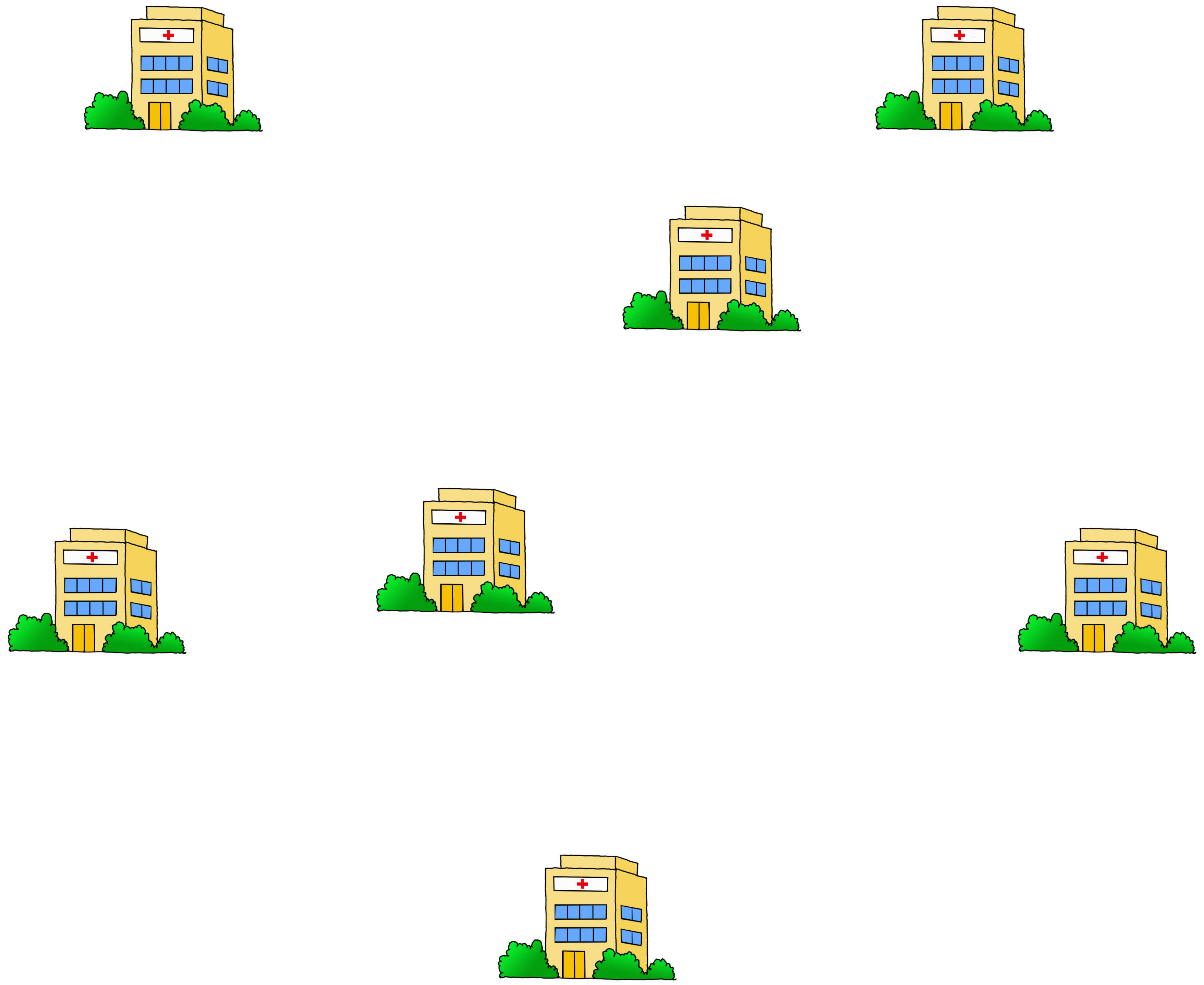
Aymeric Dieuleveut

École polytechnique

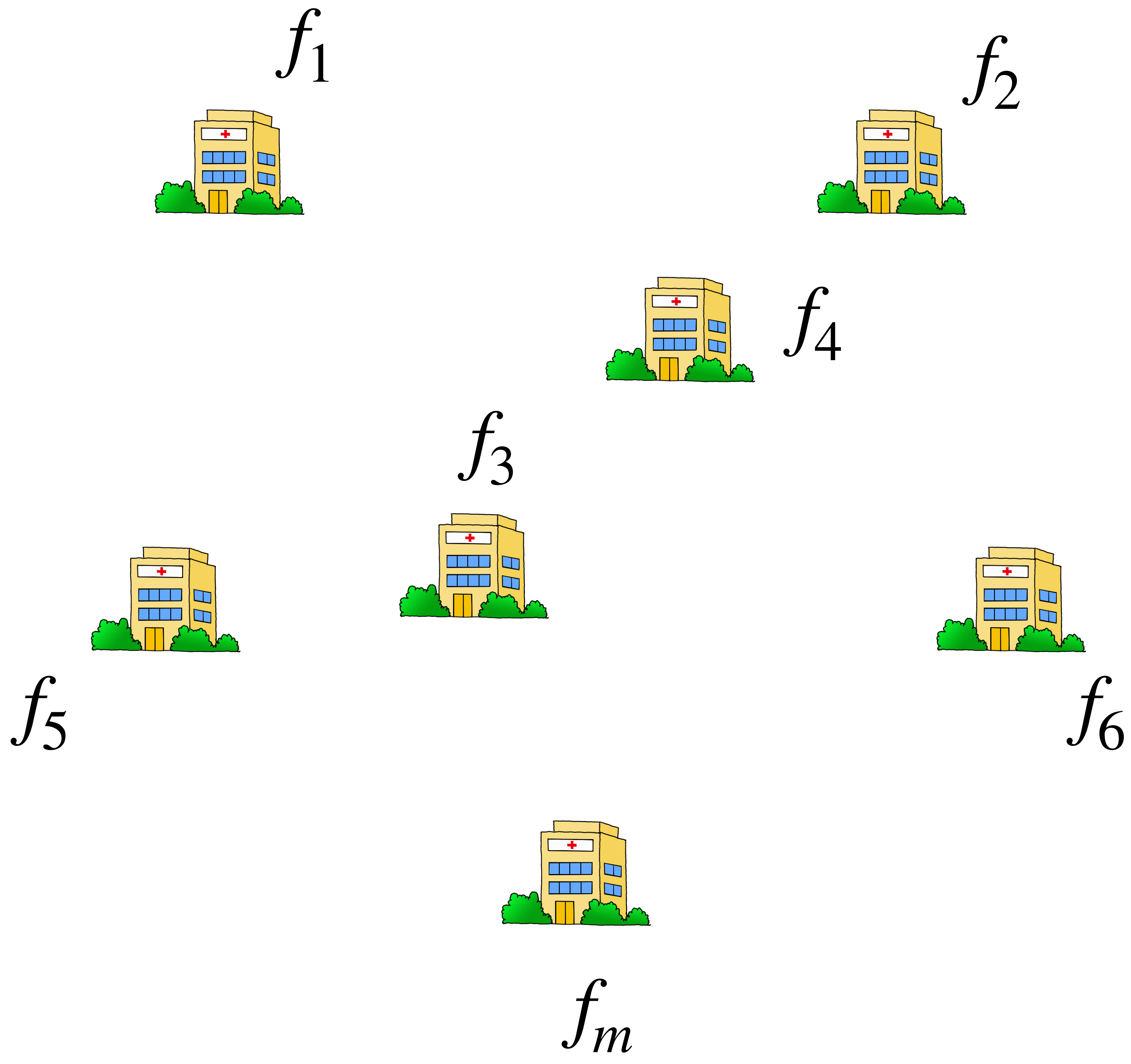
Hadrien Hendrikx

Inria Grenoble

Distributed Optimization in Machine Learning



Distributed Optimization in Machine Learning

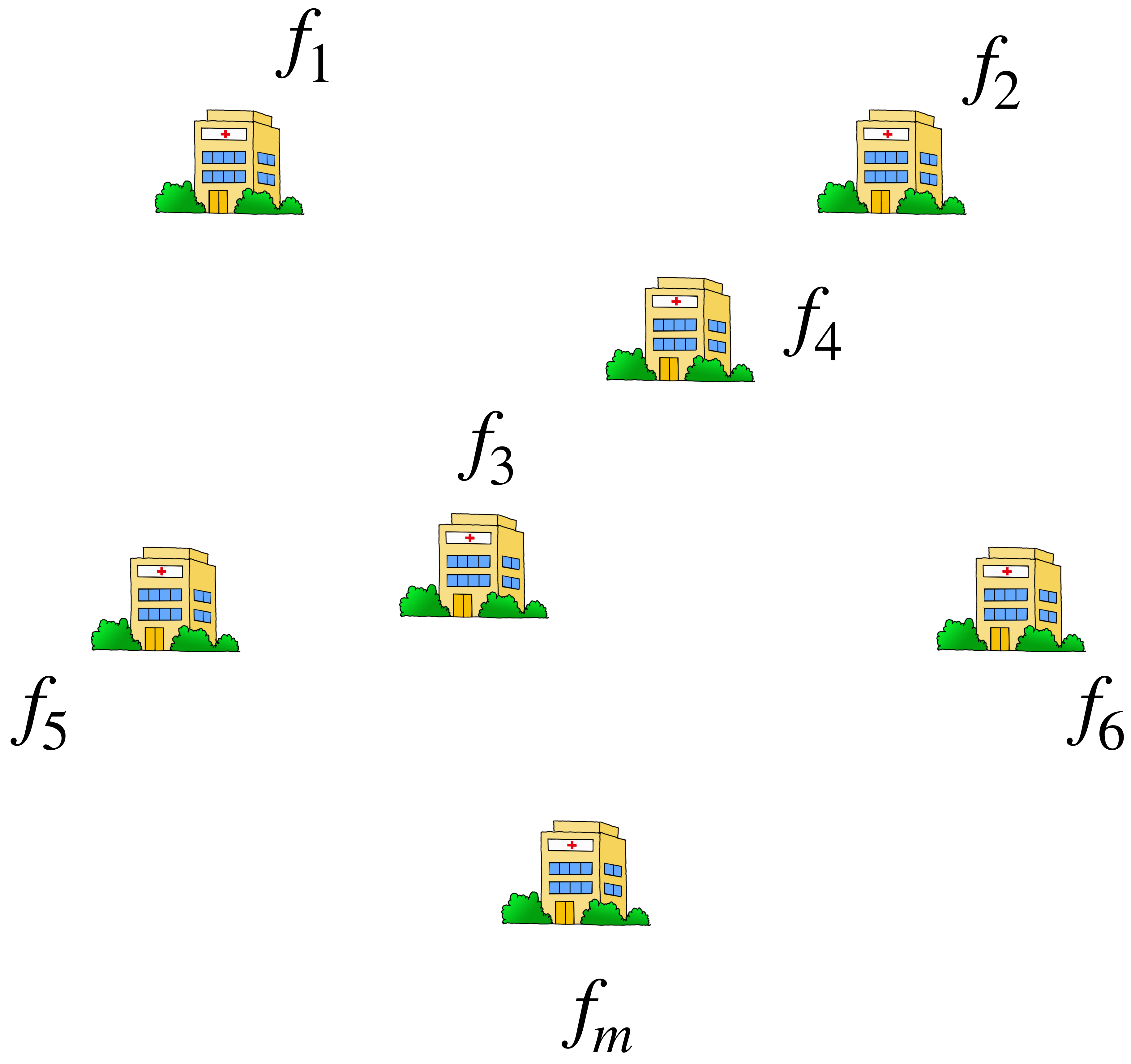


Distributed Optimization in Machine Learning

Number of nodes in the network

local loss of node i

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

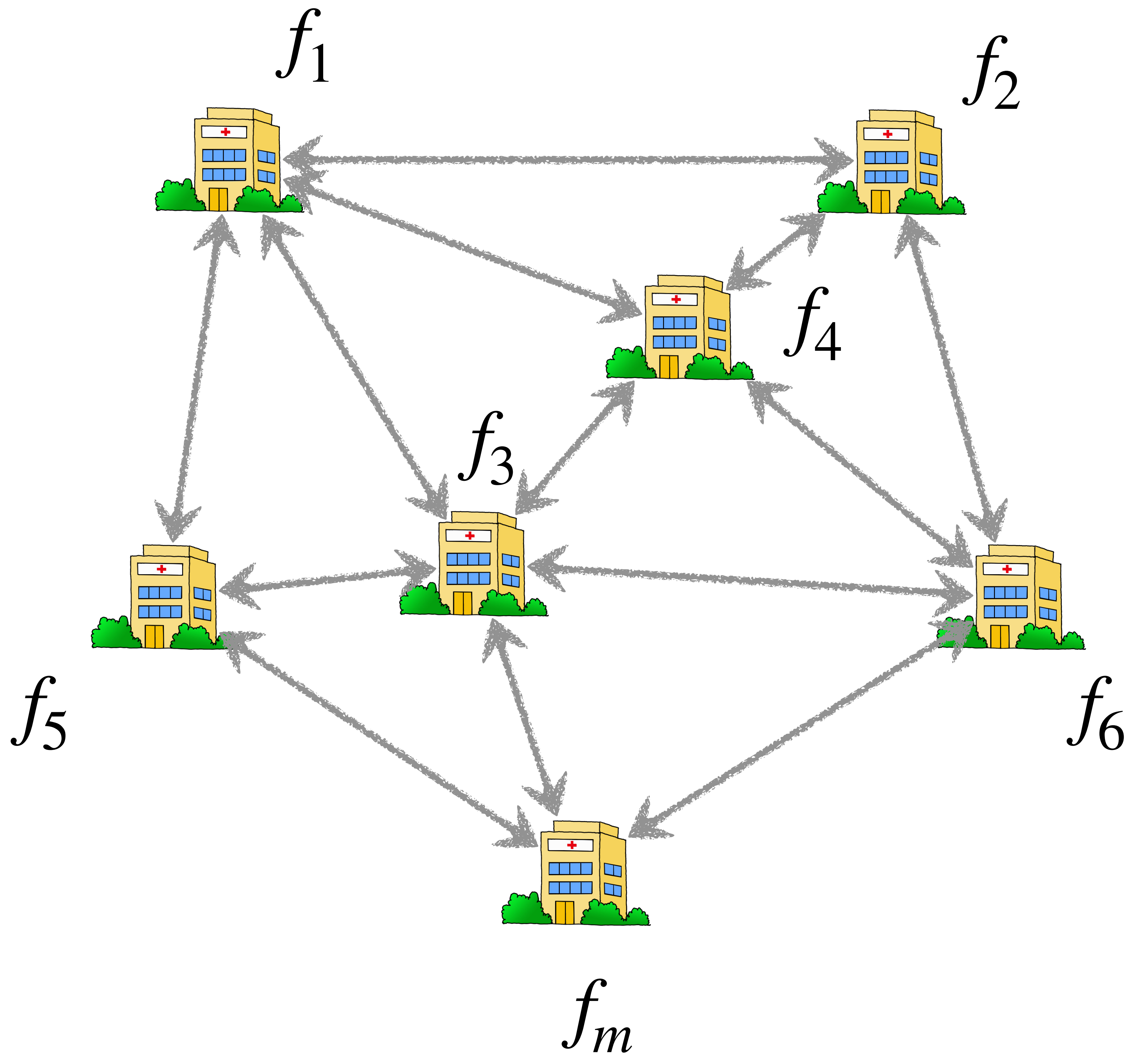


Distributed Optimization in Machine Learning

Number of nodes in the network

local loss of node i

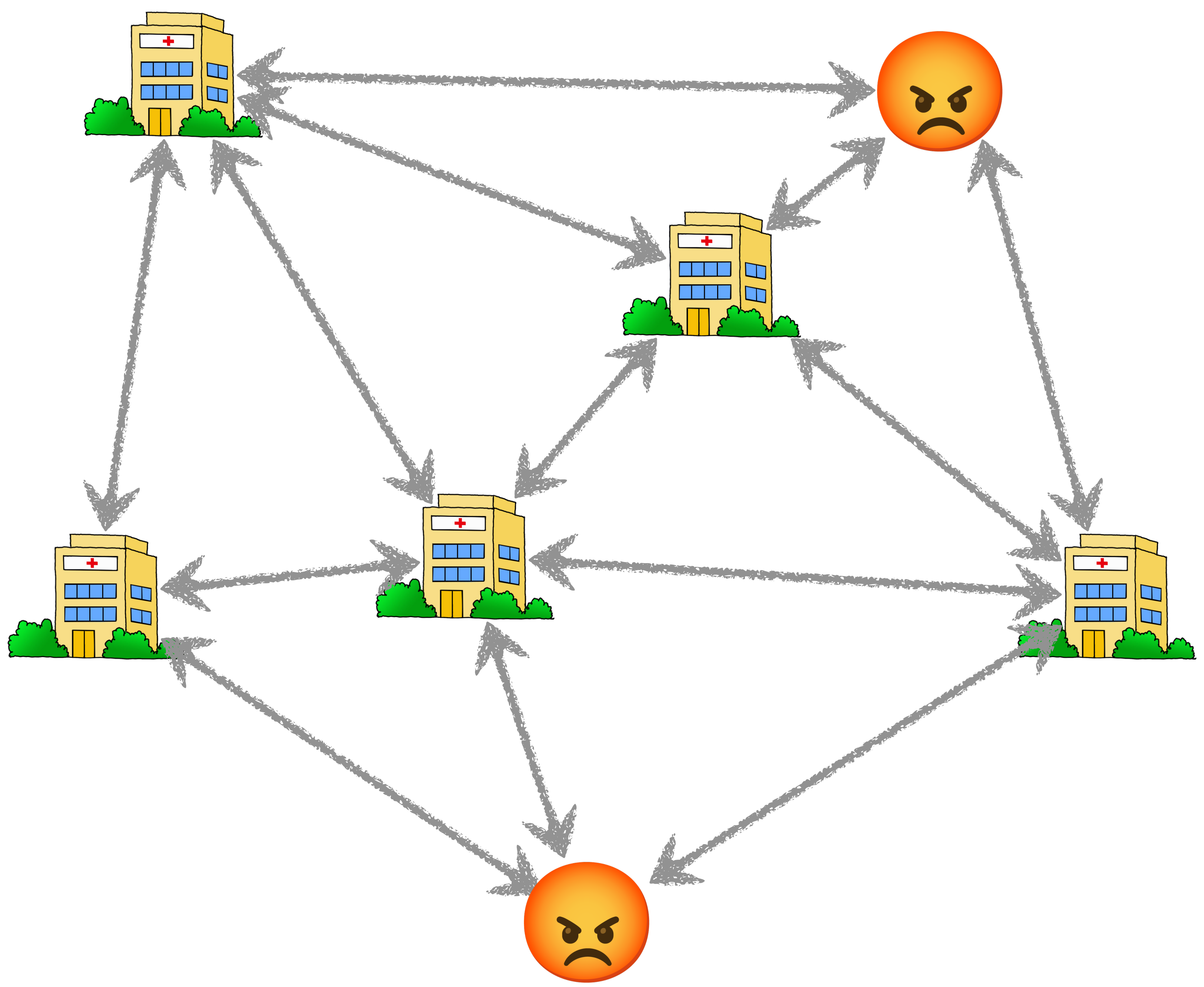
$$\min_{x \in \mathbb{R}^d} f(x) = \frac{1}{m} \sum_{i=1}^{m} f_i(x)$$



- Nodes can access their local loss function only
- Nodes collaborate to minimize the sum
- Synchronous communications

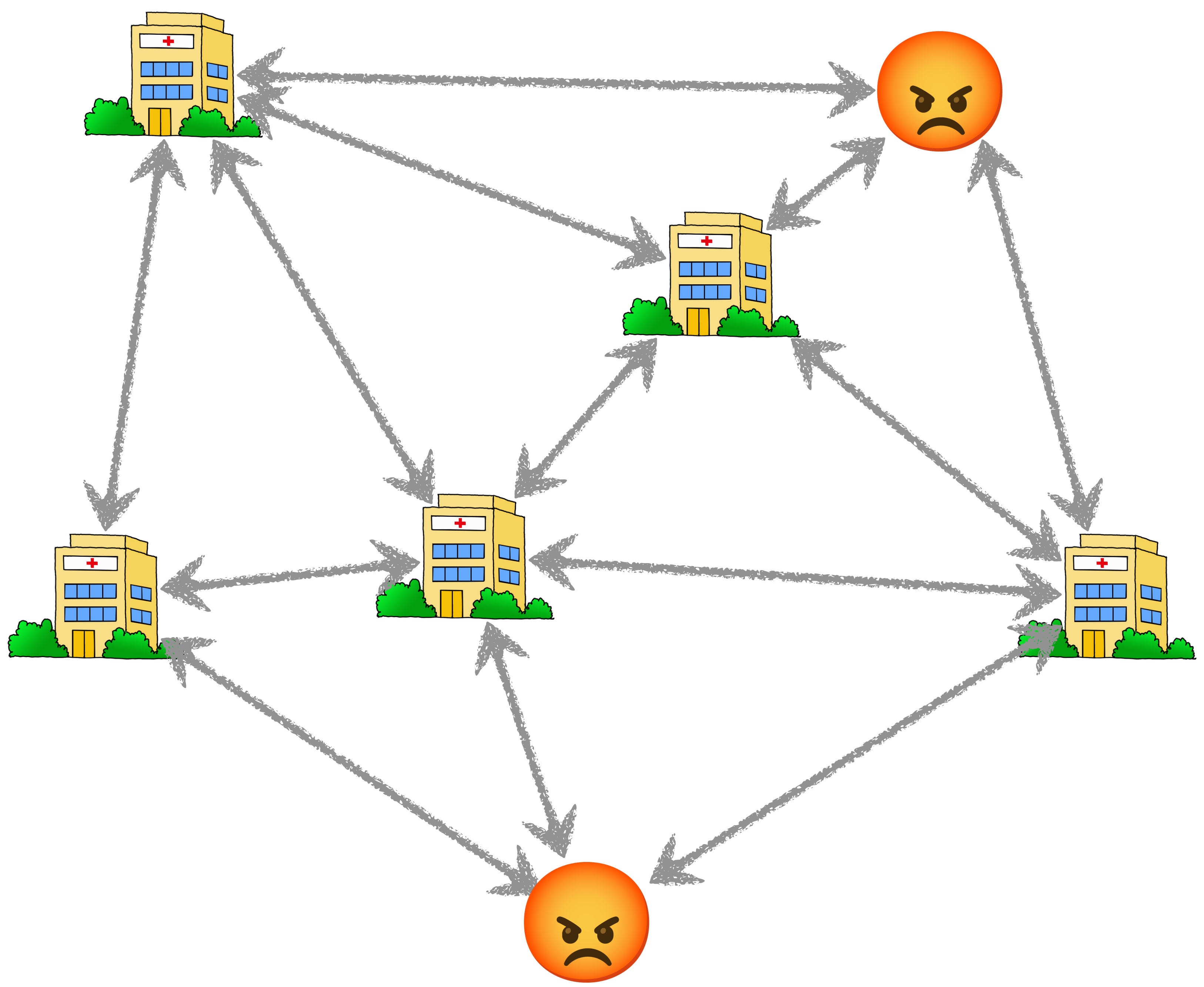
Distributed Optimization with Adversaries (Byzantines)

Goal: $$\min_{x \in \mathbb{R}^d} \frac{1}{|\text{honest}|} \sum_{i \in \text{honest}} f_i(x)$$



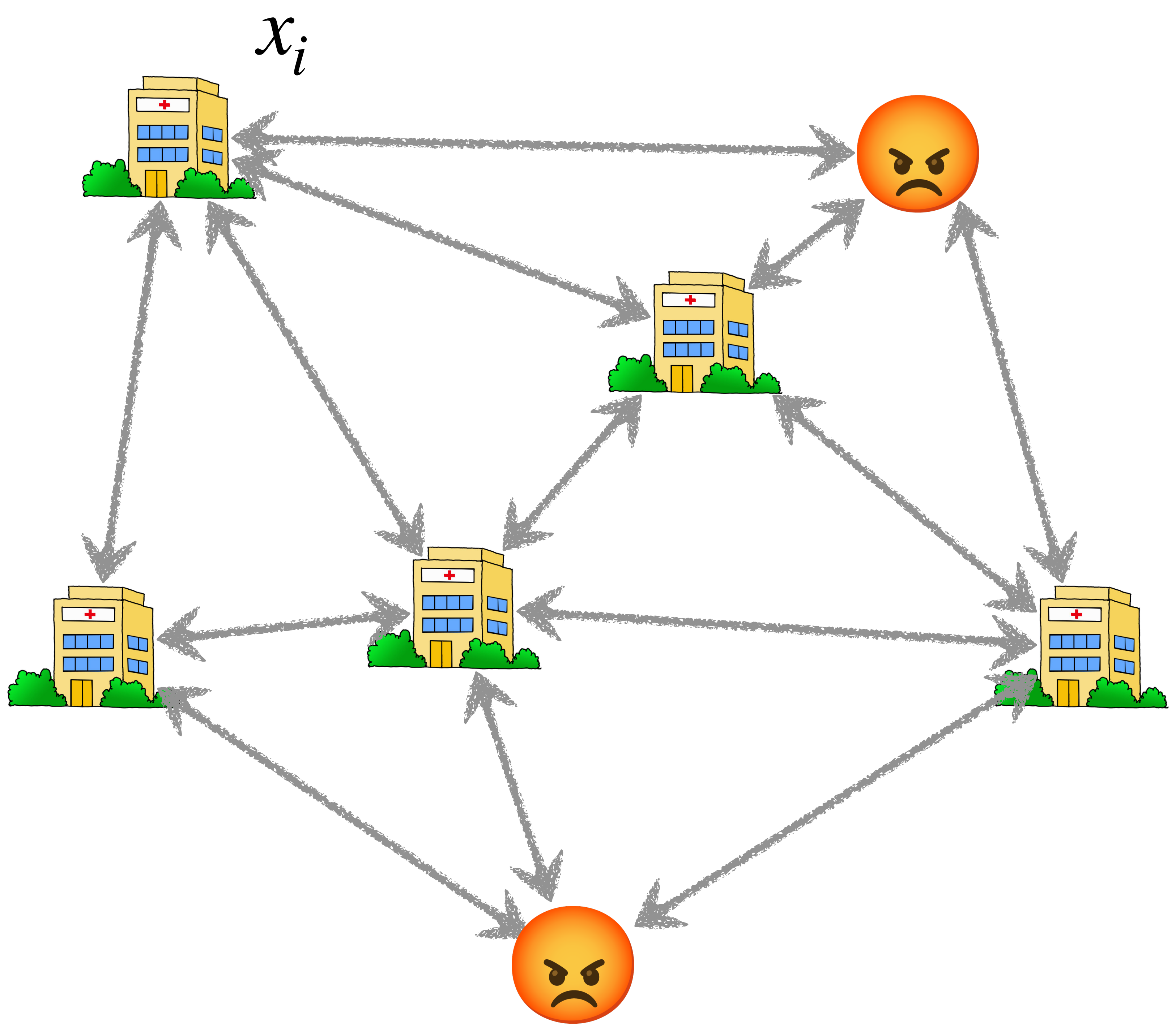
Distributed Optimization with Adversaries (Byzantines)

Goal: $$\min_{x \in \mathbb{R}^d} \frac{1}{|\text{honest}|} \sum_{i \in \text{honest}} f_i(x)$$



Distributed Optimization with Adversaries (Byzantines)

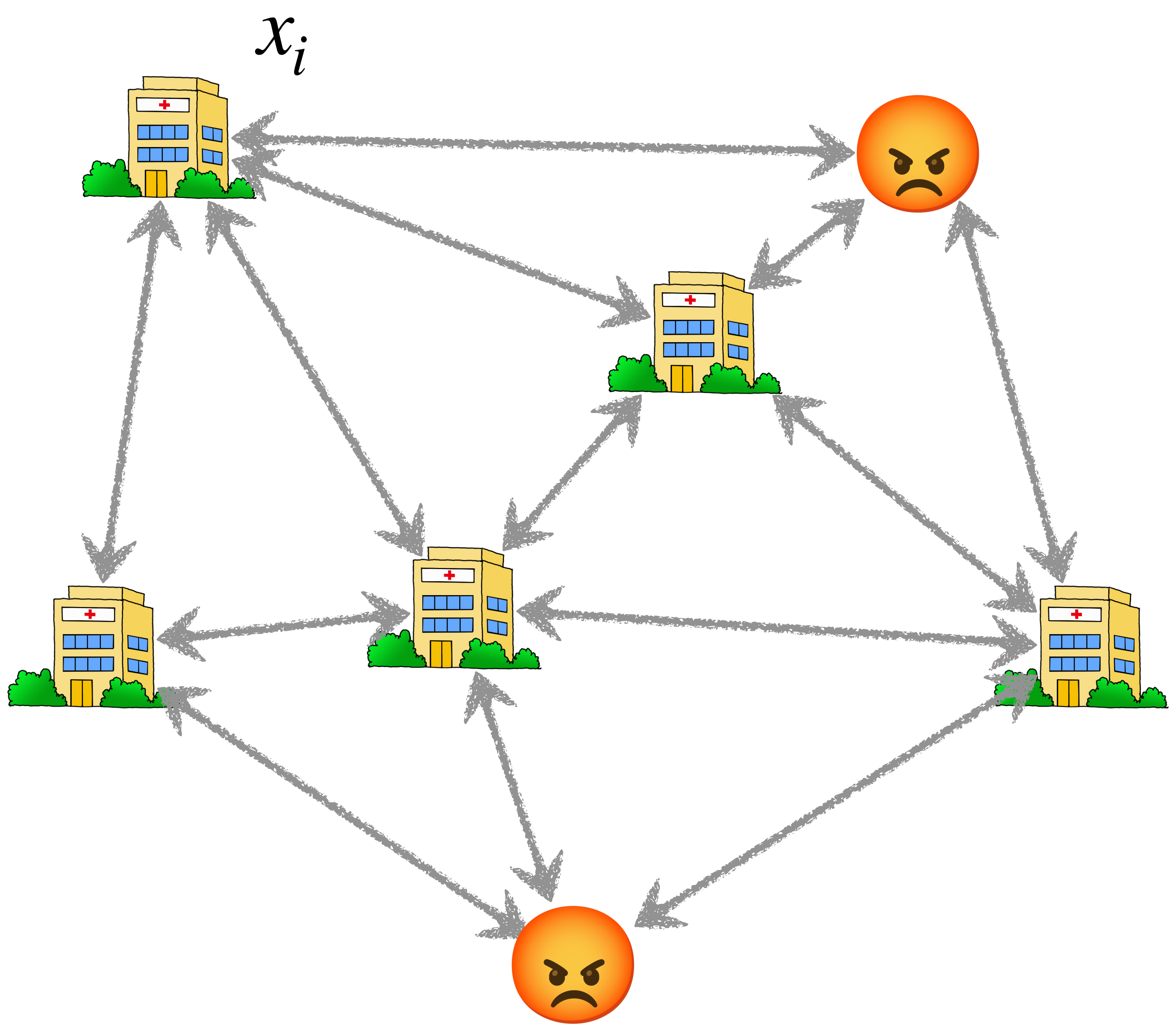
Goal: $$\bar{x}_h^0 = \frac{1}{|\text{honest}|} \sum_{i \in \text{honest}} x_i^0$$



Distributed Optimization with Adversaries (Byzantines)

Goal: $$\bar{x}_h^0 = \frac{1}{|\text{honest}|} \sum_{i \in \text{honest}} x_i^0$$

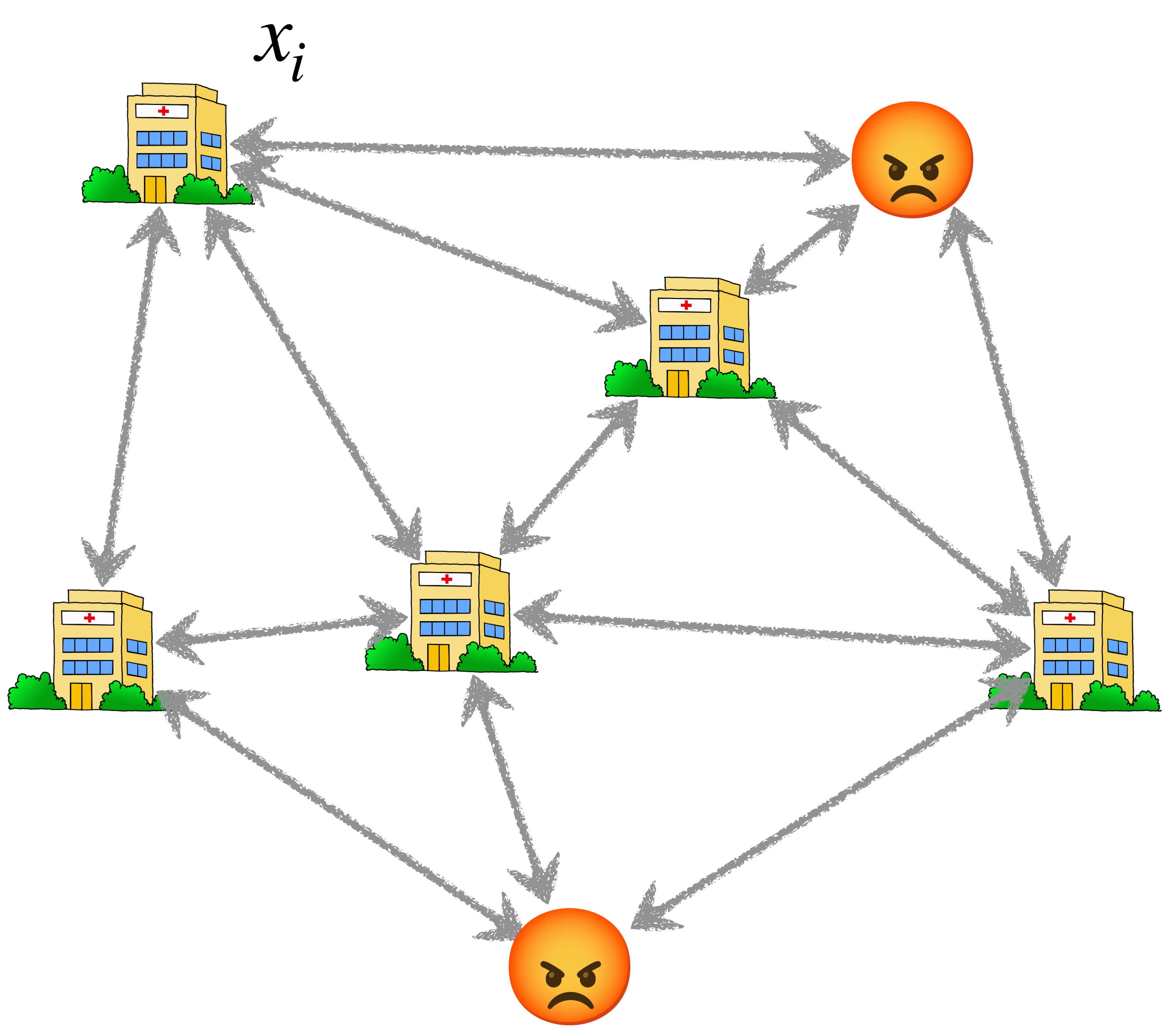
Each honest node has at most b Byzantine neighbors



Distributed Optimization with Adversaries (Byzantines)

Goal: $$\bar{x}_h^0 = \frac{1}{|\text{honest}|} \sum_{i \in \text{honest}} x_i^0$$

Each honest node has at most b Byzantine neighbors

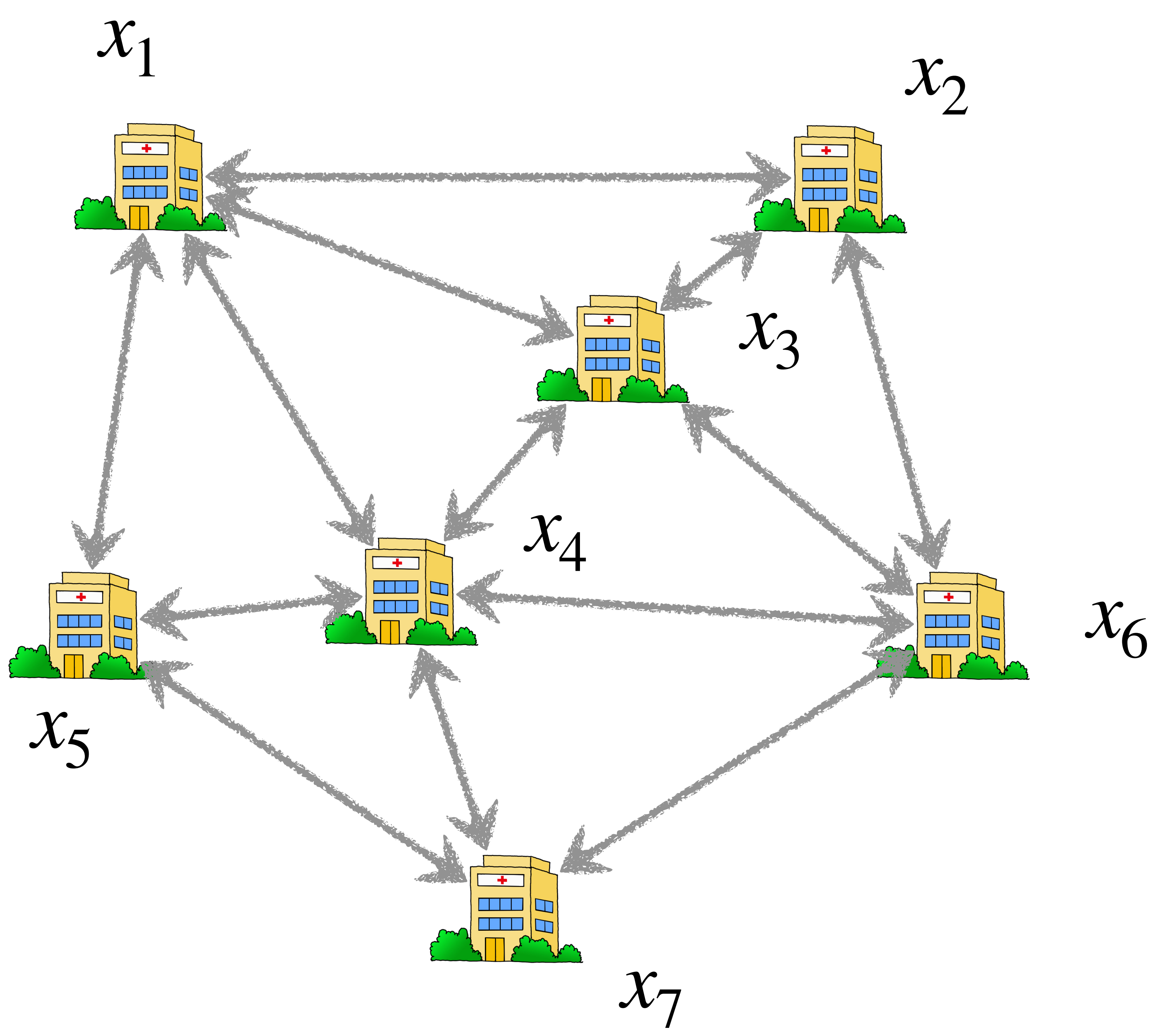


Definition: r - robustness

$$\frac{1}{|\text{honest}|} \sum_{i \in \text{honest}} \left\| x_i^t - \bar{x}_h^0 \right\|^2 \leq r \frac{1}{|\text{honest}|} \sum_{i \in \text{honest}} \left\| x_i^0 - \bar{x}_h^0 \right\|^2$$

with $r < 1$

Gossip communication



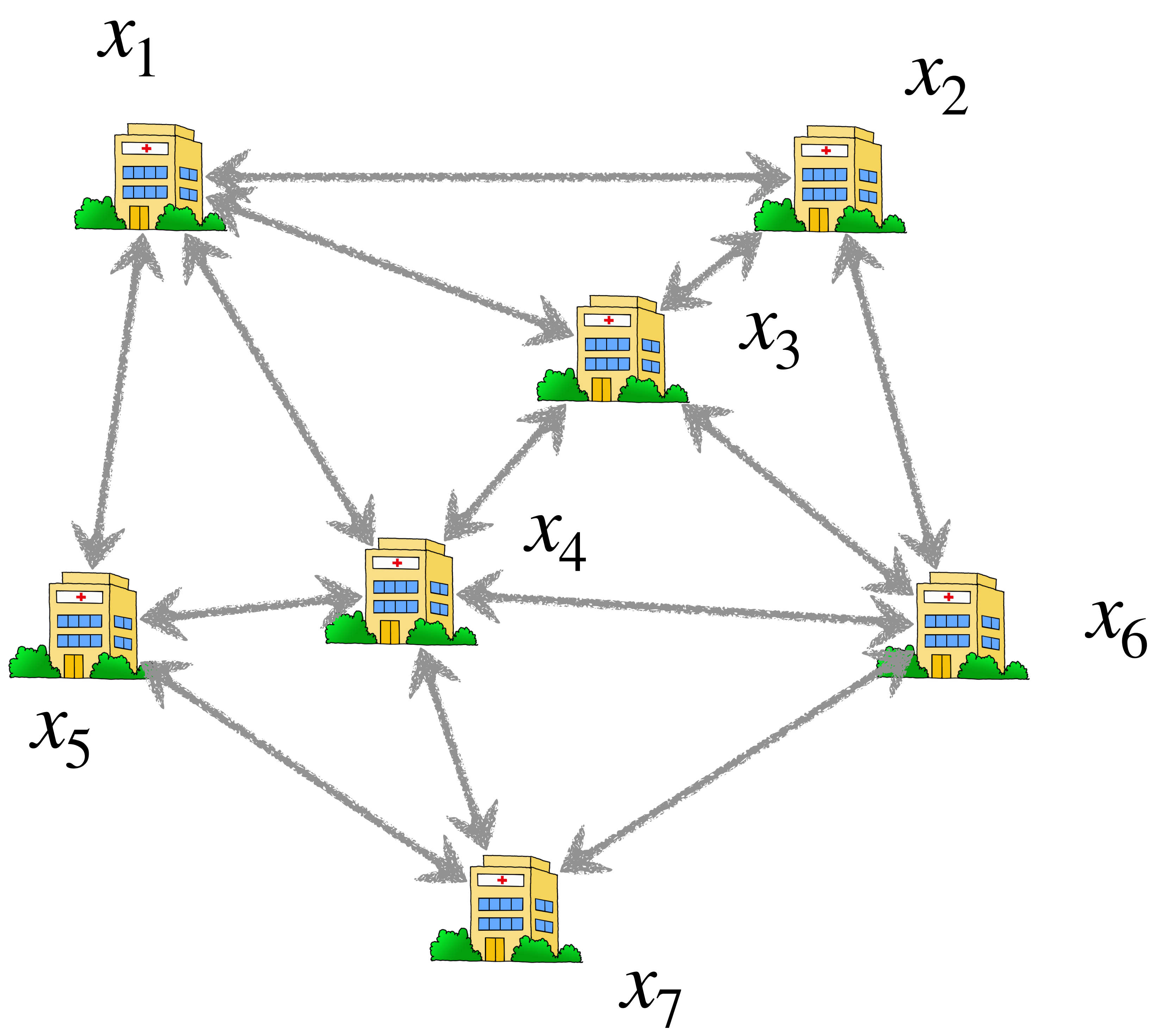
Goal

$$\bar{x} = \frac{1}{m} \sum_{i=1}^{m} x_i$$

Gossip communication

Update of node i

$$x_i^{t+1} = x_i^t - \eta \sum_{j \in \text{neighbors(i)}} \left(x_i^t - x_j^t \right)$$



Goal

$$\bar{x} = \frac{1}{m} \sum_{i=1}^{m} x_i$$

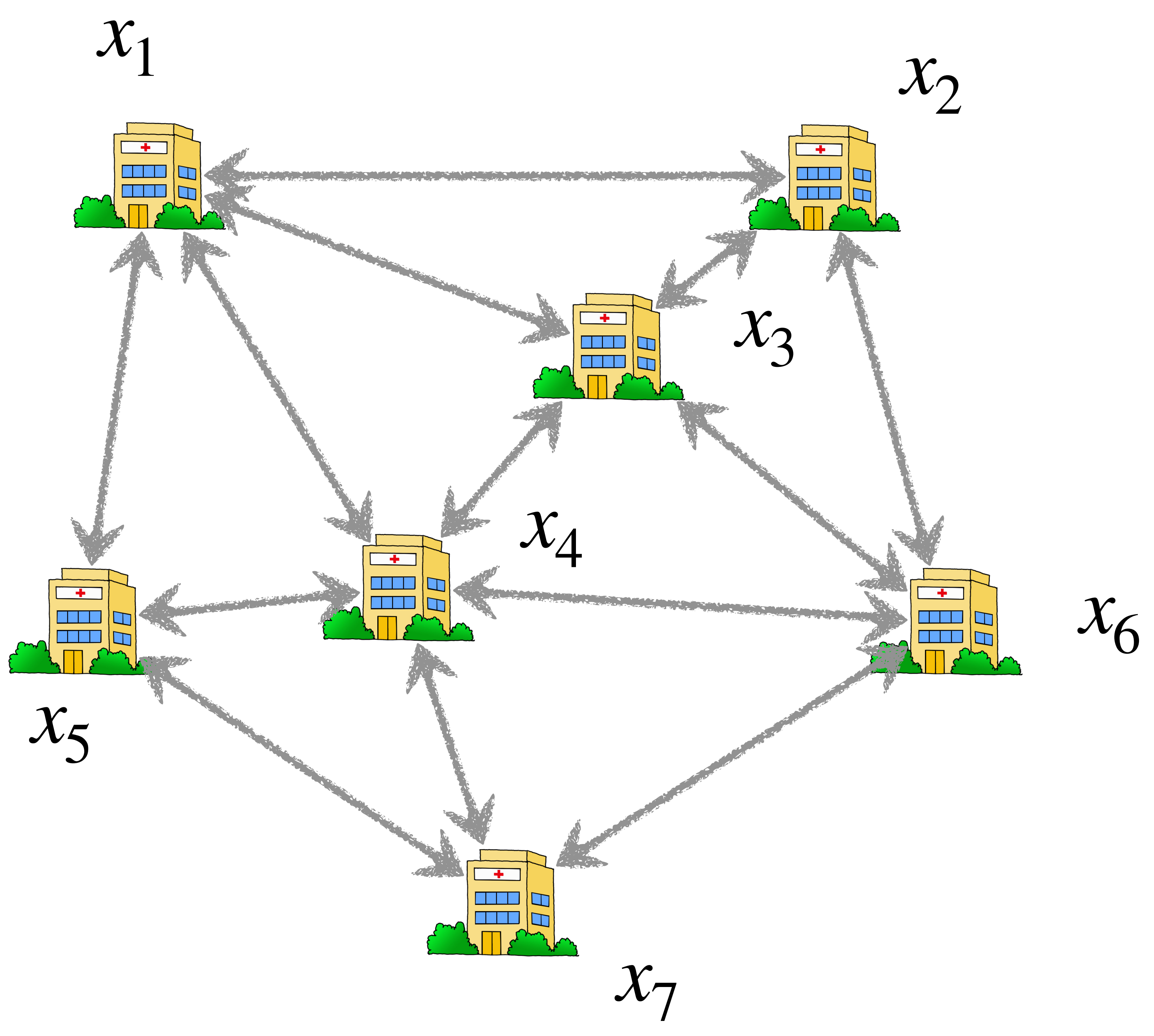
Gossip communication

Update of node i

$$x_i^{t+1} = x_i^t - \eta \sum_{j \in \text{neighbors(i)}} \left(x_i^t - x_j^t \right)$$

Using $L = \text{Diag(degrees)} - \text{Adjacency}$ and $X^t = \begin{pmatrix} x_1^t \\ \vdots \\ x_h^t \end{pmatrix}$

$$X^{t+1} = (I - \eta L) X^t$$



Goal

$$\bar{x} = \frac{1}{m} \sum_{i=1}^{m} x_i$$

Gossip communication

Update of node i

$$x_i^{t+1} = x_i^t - \eta \sum_{j \in \text{neighbors(i)}} \left(x_i^t - x_j^t \right)$$

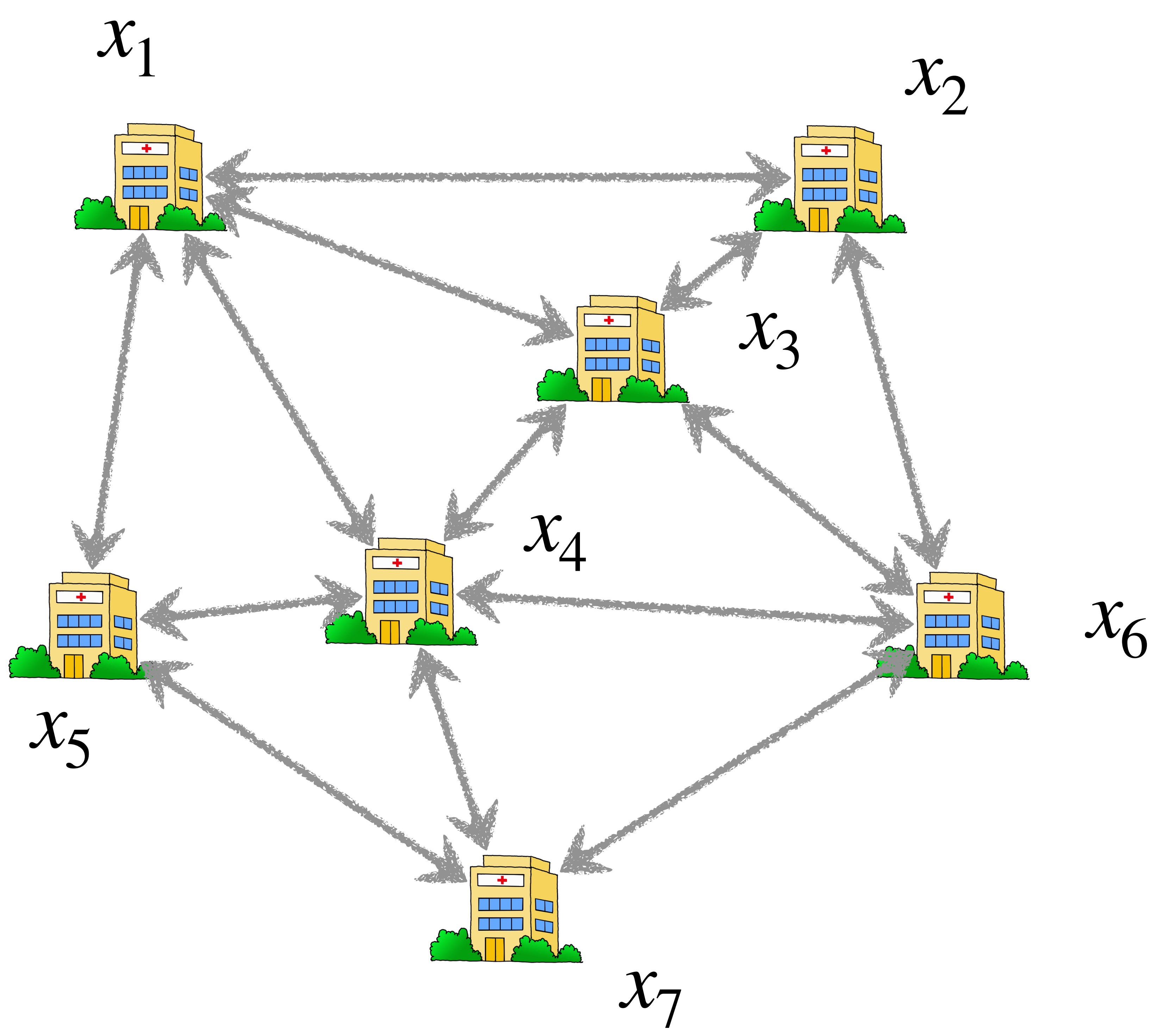
Using $L = \text{Diag(degrees)} - \text{Adjacency}$ and $X^t = \begin{pmatrix} x_1^t \\ \vdots \\ x_h^t \end{pmatrix}$

$$X^{t+1} = (I - \eta L) X^t$$

Theorem (folklore)

Spectral gap

$$\| X^t - \overline{X}^0 \| \leq \left(1 - \frac{\mu_2(L)}{\mu_{max}(L)} \right)^t \| X^0 - \overline{X}^0 \|$$



Goal

$$\bar{x} = \frac{1}{m} \sum_{i=1}^{m} x_i$$

The Robust Gossip framework

Non-robust update of node i

$$x_i^{t+1} = x_i^t - \eta \sum_{j \in \text{neighbors}(i)} \left(x_i^t - x_j^t \right)$$

The Robust Gossip framework

Robust gossip update of node i

$$x_i^{t+1} = x_i^t - \eta F\Big(\big(x_i^t - x_j^t\big)_{j \in \text{neighbors(i)}} \Big)$$

The Robust Gossip framework

Robust gossip update of node i

$$x_i^{t+1} = x_i^t - \eta\, F\Big(\big(x_i^t - x_j^t\big)_{j \in \text{neighbors(i)}}\Big)$$

Definition: Robust aggregation function

$$\Big\| F(z_1, \ldots, z_n) - \sum_{i \in \text{honest}} z_i \Big\|^2 \leq \rho\, b \sum_{i \in \text{honest}} \| z_i \|^2$$

quality / robustness of F

number of *byzantine* vectors in $z_1, \ldots, z_n$

Instances of robust aggregations

1. Sort $||z_1|| \leq \ldots \leq ||z_n||$

2.a) Remove vectors larger than $||z_{n-b}||$

$$F(z_1, \ldots, z_n) = \sum_{i=1}^{n-b} z_i$$

$\rho = 4$

Instances of robust aggregations

1. Sort $\|z_1\| \leq \ldots \leq \|z_n\|$

2.a) Remove vectors larger than $\|z_{n-b}\|$

$$F(z_1, \ldots, z_n) = \sum_{i=1}^{n-b} z_i$$

$\rho = 4$

2.b) Clip vectors larger at $\|z_{n-2b}\|$

$$F(z_1, \ldots, z_n) = \sum_{i=1}^{n} \frac{z_i}{\|z_i\|} \min \left(\|z_i\|, \|z_{n-2b}\| \right)$$

$\rho = 2$

F-Robust Gossip is r-robust

Theorem

$$\frac{1}{|\text{honest}|} \sum_{i \in \text{honest}} \left\| x_i^1 - \bar{x}_h^0 \right\|^2 \leq r \frac{1}{|\text{honest}|} \sum_{i \in \text{honest}} \left\| x_i^0 - \bar{x}_h^0 \right\|^2$$

with $r = 1 - \dfrac{\mu_2(L) - 2\rho b}{\mu_{max}(L)}$

Algebraic connectivity

In fully connected graphs $\mu_2(L) = |\text{honest}|$

↪ r-robust until a proportion of $1/(2\rho+1)$ aversaries

Tightness of the breakdown point

Theorem

There are arbitrarily sparse graphs and initial values $\{x_i^0\}$ on which, if $2b \geq \mu_2(L)$, no decentralized algorithm is r-robust with r <1

Tightness of the breakdown point

Theorem

There are arbitrarily sparse graphs and initial values $\{x_i^0\}$ on which, if $2b \geq \mu_2(L)$, no decentralized algorithm is r-robust with r <1

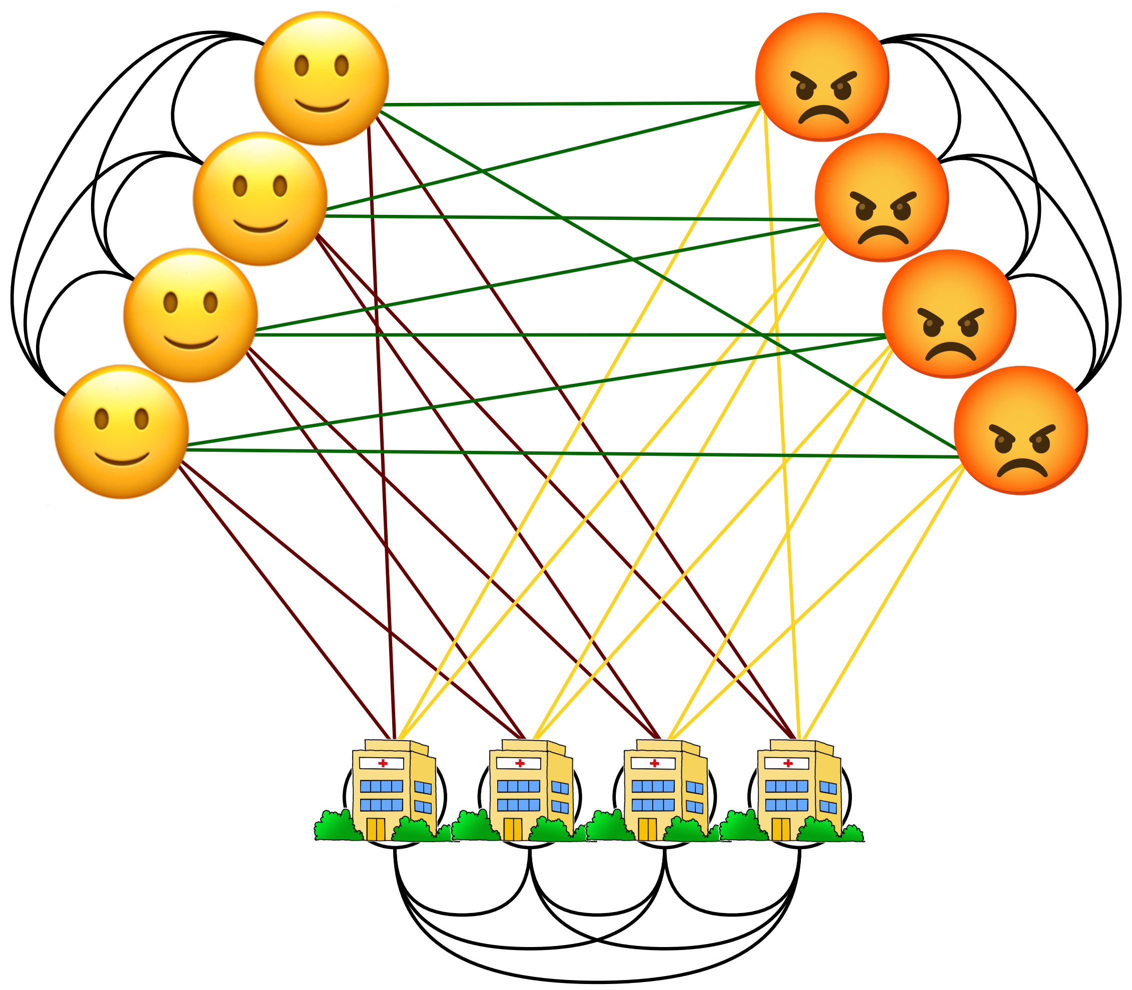
↪ $\rho = 1$ is the best we can have !

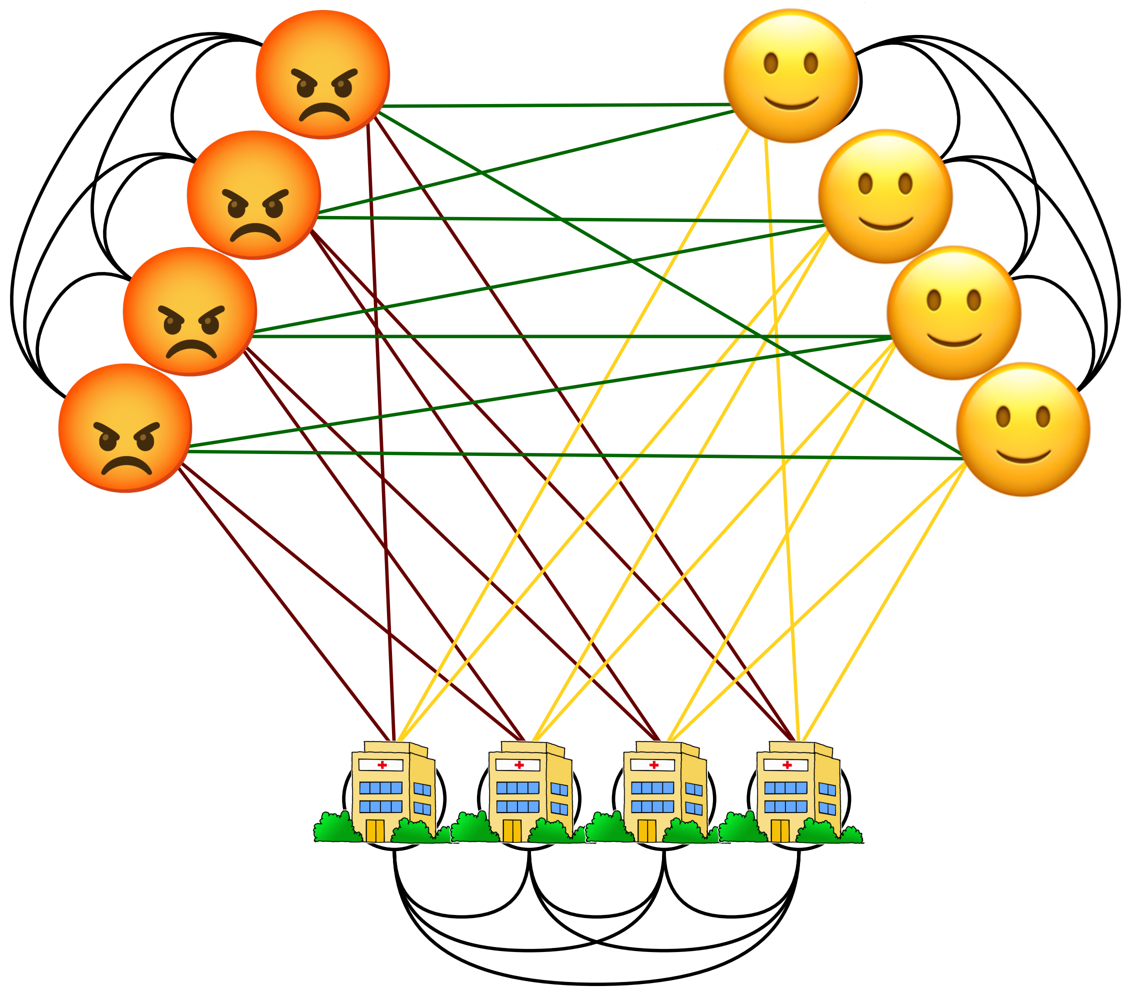
↪ *At most* $1/3$ adversaries in fully-connected graphs

Tightness of the breakdown point

Theorem

There are arbitrarily sparse graphs and initial values $\{x_i^0\}$ on which, if $2b \geq \mu_2(L)$, no decentralized algorithm is r-robust with r <1

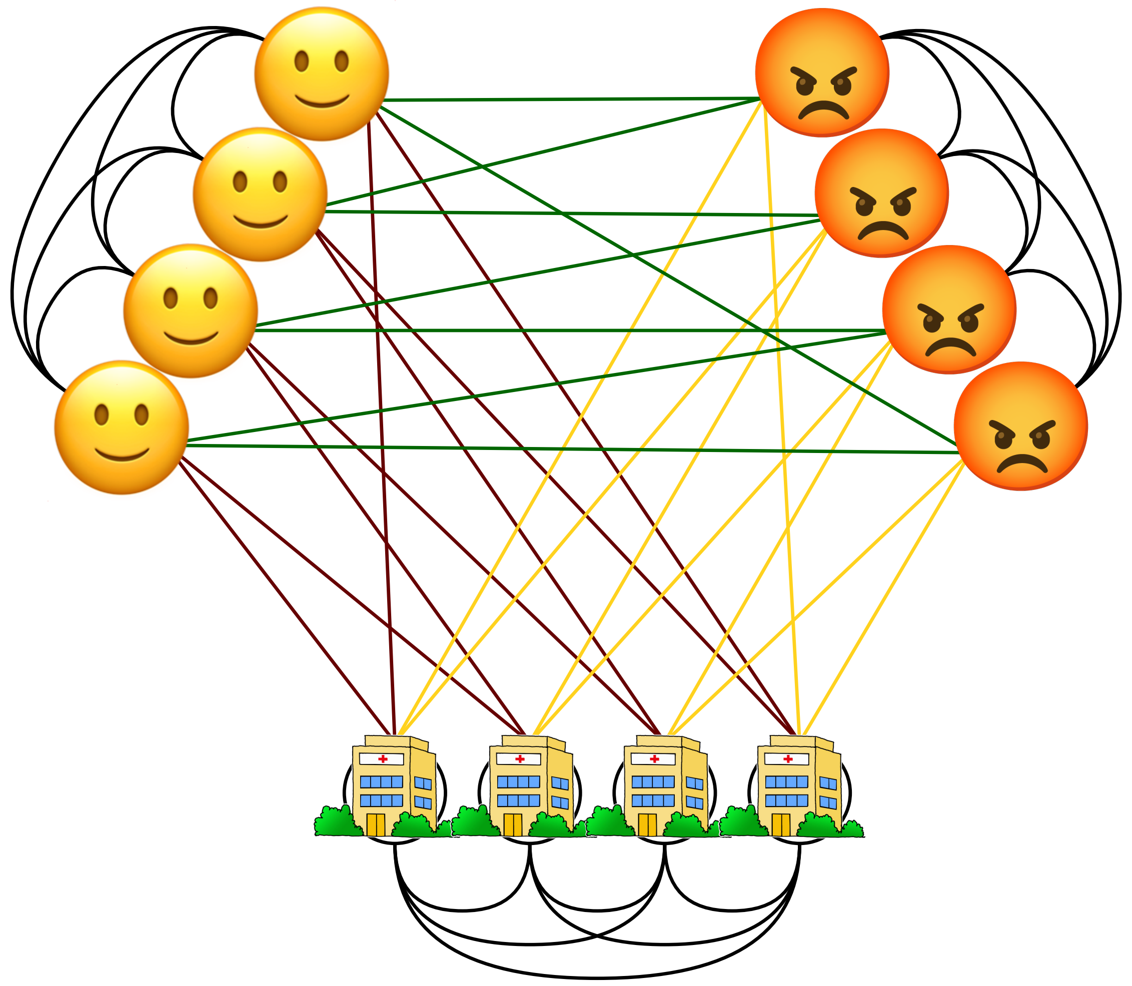




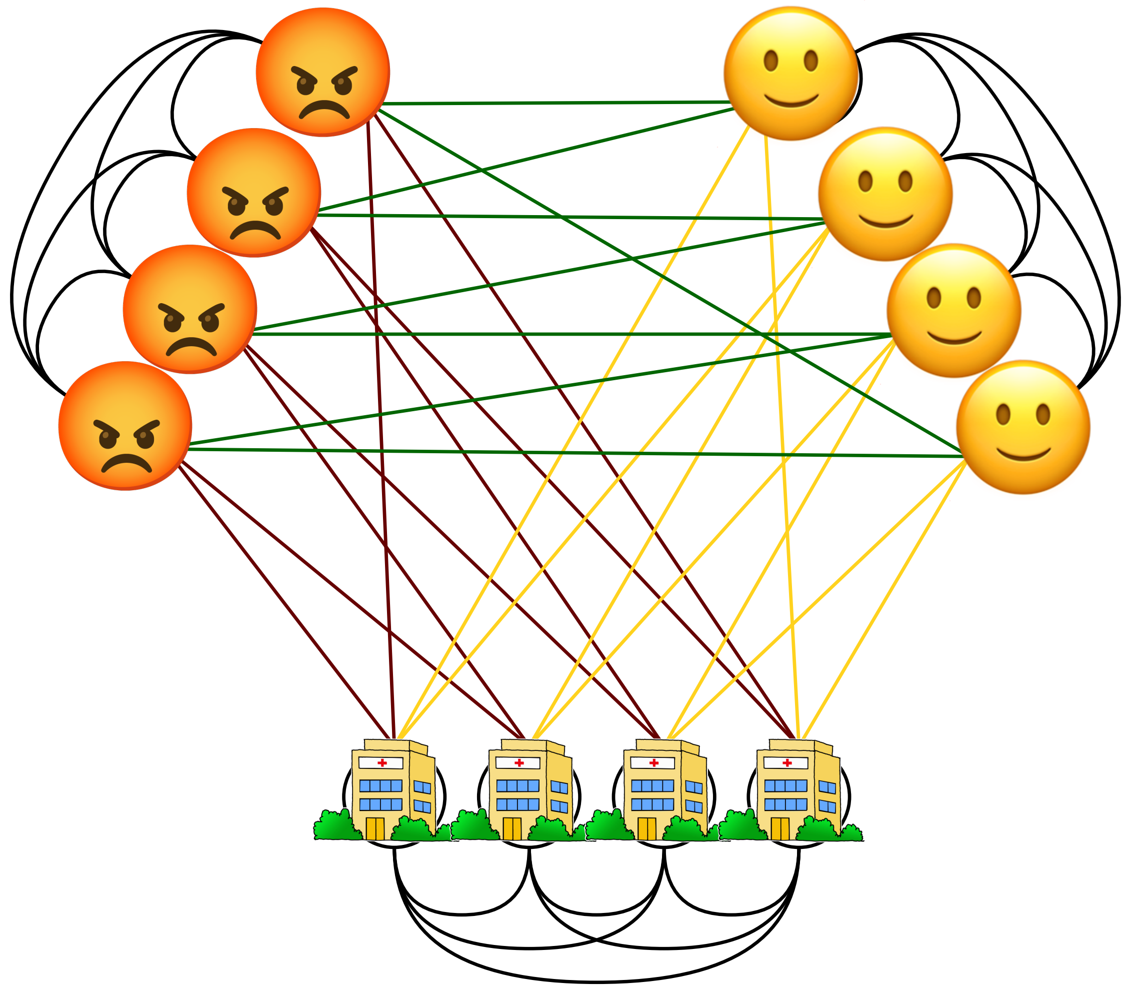
Tightness of the breakdown point

Theorem

There are arbitrarily sparse graphs and initial values $\{x_i^0\}$ on which, if $2b \geq \mu_2(L)$, no decentralized algorithm is r-robust with r <1



????



Asymptotic consensus

« Breakdown ratio » $\delta = 2\rho b/\mu_2(L)$

Spectral gap of the graph $\gamma = \mu_2(L)/\mu_{max}(L)$

Asymptotic consensus

« Breakdown ratio » $\delta = 2\rho b / \mu_2(L)$ Spectral gap of the graph $\gamma = \mu_2(L)/\mu_{max}(L)$

Corrollary: After T iterations of F-RG

$$\frac{1}{|\text{honest}|} \sum_{i \in \text{honest}} \left\| x_i^T - \bar{x}_h^T \right\|^2 \leq \big(1-\gamma(1-\delta)\big)^T \frac{1}{|\text{honest}|} \sum_{i \in \text{honest}} \left\| x_i^0 - \bar{x}_h^0 \right\|^2$$

Asymptotic consensus

« Breakdown ratio » $\delta = 2\rho b / \mu_2(L)$ Spectral gap of the graph $\gamma = \mu_2(L)/\mu_{max}(L)$

Corrollary: After T iterations of F-RG

$$\frac{1}{|\text{honest}|} \sum_{i \in \text{honest}} \left\| x_i^T - \bar{x}_h^T \right\|^2 \leq \left(1 - \gamma(1-\delta)\right)^T \frac{1}{|\text{honest}|} \sum_{i \in \text{honest}} \left\| x_i^0 - \bar{x}_h^0 \right\|^2$$

$$\left\| \bar{x}_h^T - \bar{x}_h^0 \right\|^2 \leq \frac{4\delta}{\gamma(1-\delta)^2} \frac{1}{|\text{honest}|} \sum_{i \in \text{honest}} \left\| x_i^0 - \bar{x}_h^0 \right\|^2$$

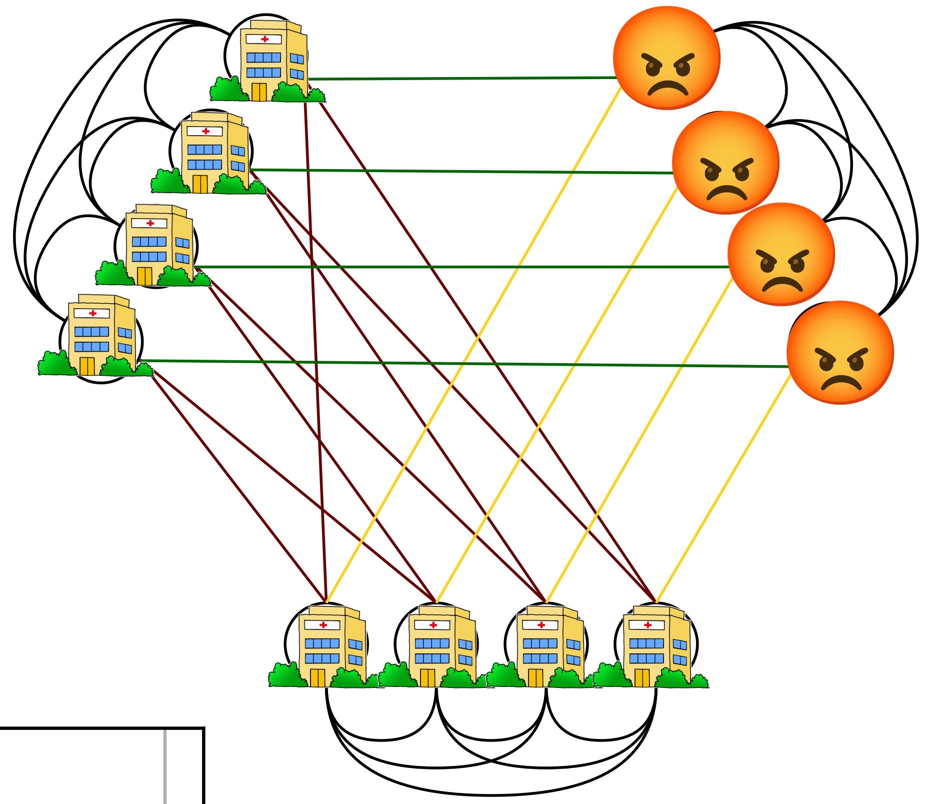
F-RG recovers existing algorithms

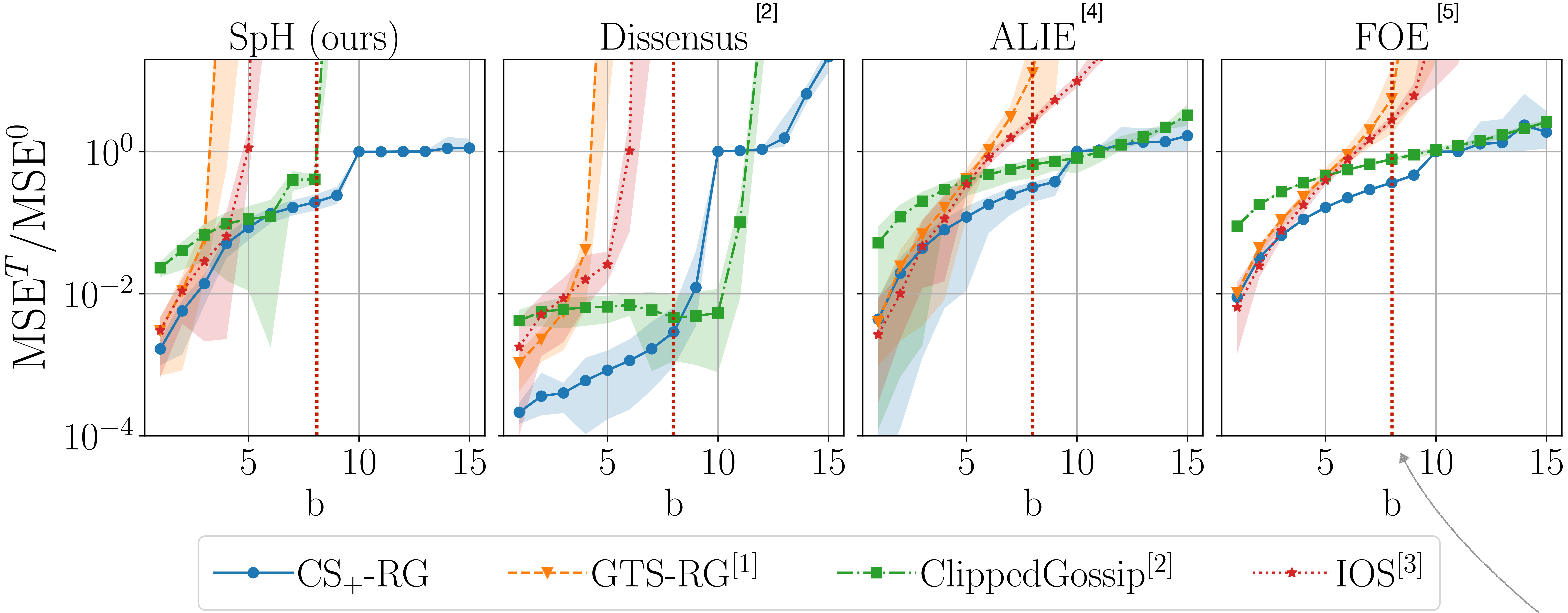
- Trimming + F-RG corresponds, in fully connected graphs, to *Nearest Neighbor Averaging* [1]
- Clipping + F-RG with another *oracle* clipping threshold recovers *ClippedGossip* [2] (w. $\rho = 4$)

[1] Robust collaborative learning with linear gradient overhead, Farhadkhani et al., ICML 2023

[2] Byzantine-Robust Decentralized Learning via ClippedGossip, He et. al. arxiv 2022

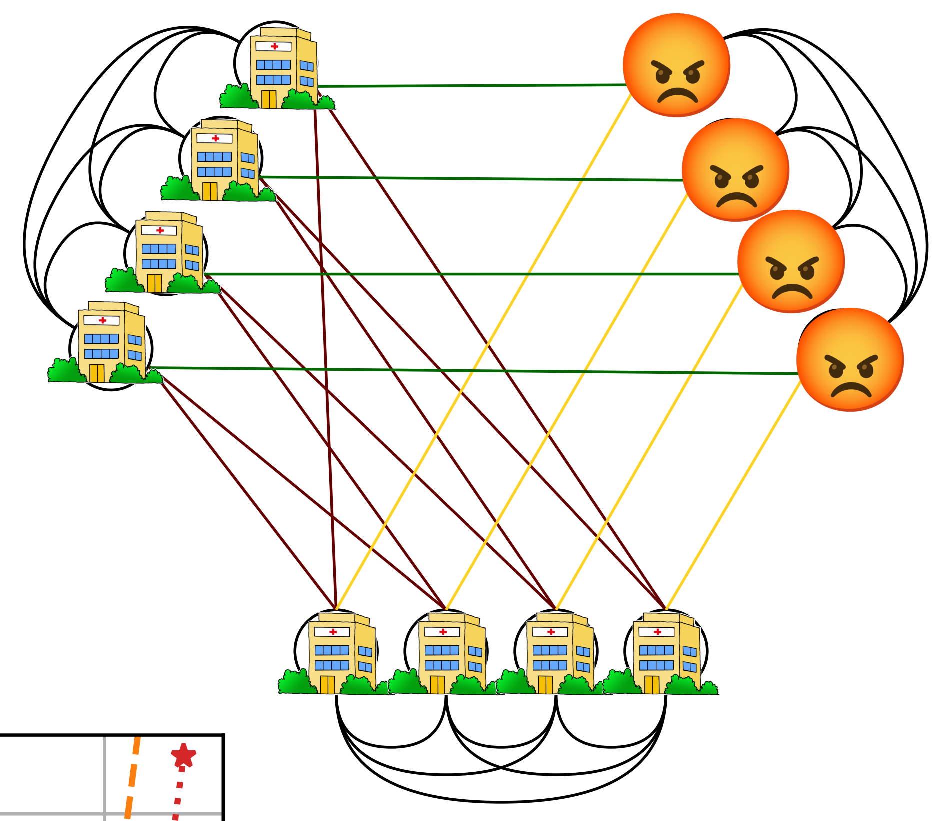
Experiments - communication only

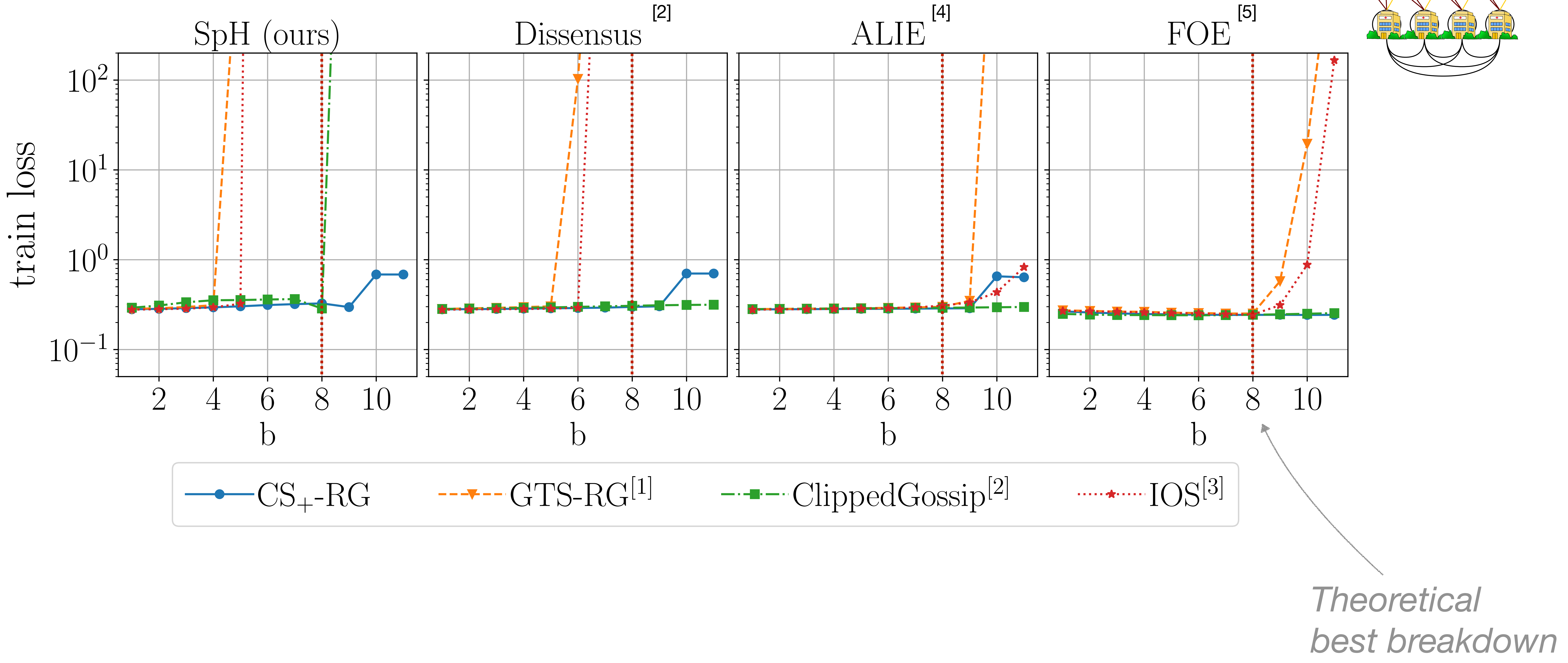




Theoretical best breakdown

Experiments - CNN on MNIST





More in the paper

- ☑ Convergence for local SGD steps + communication with F-RG
- ☑ A new attack that builds on the spectral properties of the graph
- ☑ Experiments

[1] Robust collaborative learning with linear gradient overhead, Farhadkhani et al., ICML 2023

[2] Byzantine-Robust Decentralized Learning via ClippedGossip, He et. al. arxiv 2022

[3] Byzantine-resilient decentralized stochastic optimization with robust aggregation rules, Wu et. al. IEEE tsp 2023

[4] A little is enough: Circumventing defenses for distributed learning, Baruch et. al. NeurIPS 2019

[5] Fall of empires: Breaking byzantine tolerant SGD by inner product manipulation, Xie et. al., UAI, 2020

Experiments - communication w. Erdos Renyi