Robust Distributed Estimation: Extending Gossip Algorithms to Ranking and Trimmed Means

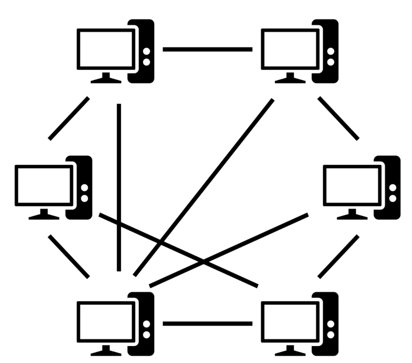
Anna van Elst   Igor Colin   Stephan Clémençon

LTCI, Télécom Paris, Institut Polytechnique de Paris

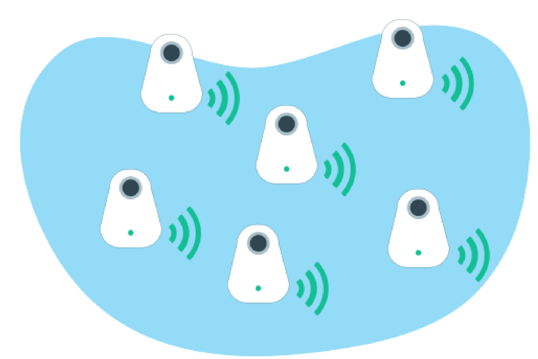
Gossip Algorithms: Where and Why?

- When data is naturally distributed over a **network**.
- If we have **resource** and **communication** constraints.
- If we don't have access to a **central server**.

⇒ For decentralized AI/ML



Peer-to-Peer (P2P) Networks

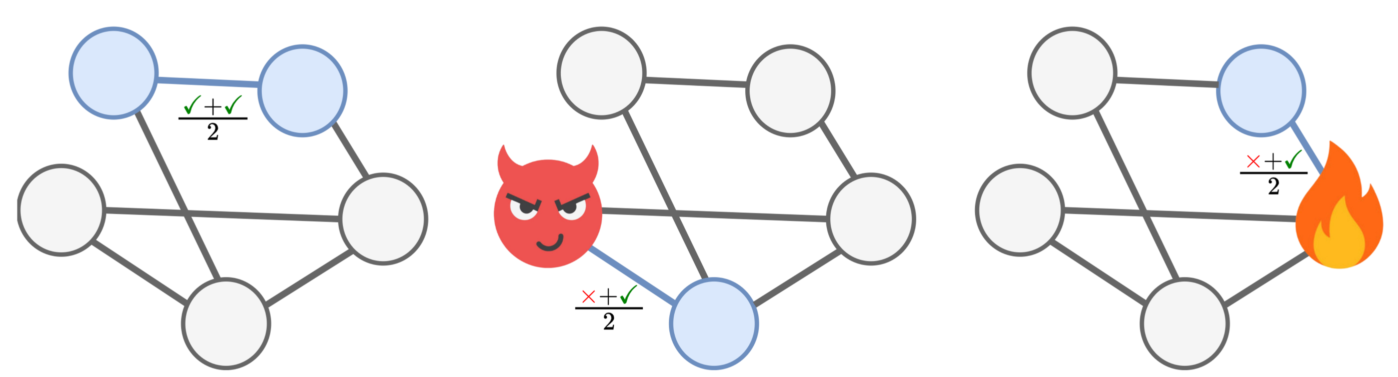


Wireless Sensor Networks (Internet of Things)

Can We Design More Robust Gossip Algorithms?

- Mean-based gossip algorithms are vulnerable to **data contamination.**
- Previous work rely on **local robust** aggregation and fail in most cases.

⇒ We develop a **global robust** aggregation method using **trimmed means**.



Gossip algorithms are not robust to corrupted or malicious nodes.

Problem Formulation and Framework

- **Communication network:** connected undirected graph $G=(V,E)$ with n nodes. Each node k holds observation X_k.
- **Huber's contamination model:** fraction ε of corrupted nodes.
- **Synchronous/asynchronous setting**: global/local iteration counter.

Using **ranks**, the goal is to compute a **robust** average: denoting $m = \lfloor n\alpha \rfloor$,

$$\underbrace{\bar{x}_\alpha}_{\alpha\text{-trimmed mean}} = \frac{1}{n-2m}\sum_{k=m+1}^{n-m} \underbrace{X_{n(k)}}_{\text{Order statistic}} = \frac{1}{n}\sum_k f\left(\underbrace{r_k}_{\text{Rank}}\right)\cdot \underbrace{X_k}_{\text{Observation}}$$

GoRank – A Gossip Algorithm for Ranking

- Ranks can be computed using **pairwise comparisons**:

$$\underbrace{r_k}_{\text{Rank of node } k} = 1 + n\left(\frac{1}{\underbrace{n}_{\text{\# of nodes}}}\sum_{l=1}^{n} \underbrace{\mathbb{I}_{\{X_k > X_l\}}}_{\text{Pairwise comparison}}\right) = 1 + n r'_k$$

- For each iteration t, select a **random edge** $(i,j) \in E$ with probability $1/|E|$ and the two connected nodes **swap** their observations $X_i \leftrightarrow X_j$.
- Then, the rank estimates are updated using a **running average**:

$$\underbrace{R_k(t)}_{\text{Rank estimate}} = 1 + n\left(\frac{1}{\underbrace{C_k(t)}_{\text{Counter}}}\sum_{s=1}^{C_k(t)} \mathbb{I}_{\{X_k > \underbrace{Y_k(s)}_{\text{Auxiliary observation}}\}}\right) = 1 + nR'_k(t)$$

Theorem 1&2 (Convergence of GoRank Estimates).

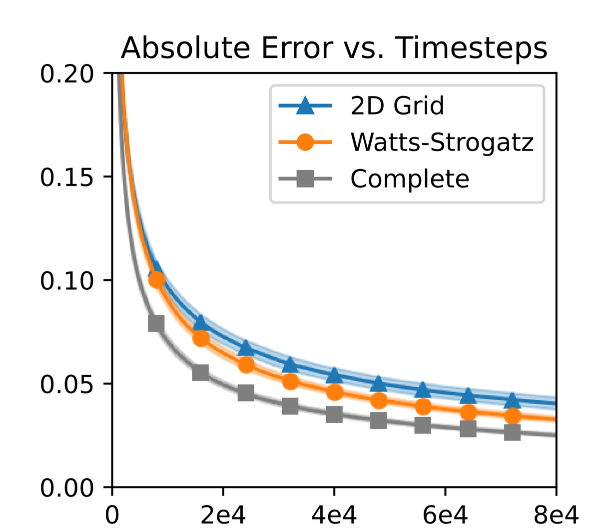
Define $c = \lambda_2/|E|$ the connectivity of the graph with λ_2 denoting the spectral gap of the graph Laplacian. Then, $\forall k \in [n], \forall t \geq 1$,

$$|\mathbb{E}[R_k(t)] - r_k| \leq \frac{\sigma_k}{ct} \quad \& \quad \mathbb{E}|R_k(t) - r_k| \leq \mathcal{O}\left(\frac{\sigma_k}{\sqrt{ct}}\right).$$

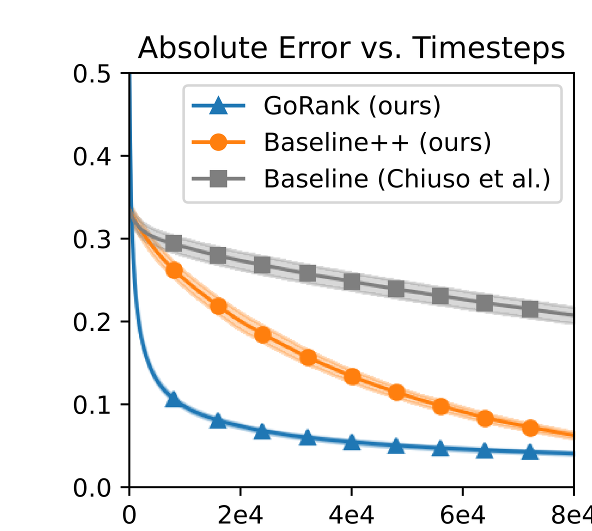
▶ What makes the theoretical analysis work?

- **Swapping matrix**: $t > 0$, for $(i,j) \in E$, $\mathbf{W}_1(t) = \mathbf{I}_n - (\mathbf{e}_i - \mathbf{e}_j)(\mathbf{e}_i - \mathbf{e}_j)^\top$.
- Its **expectation** is: $\mathbf{W}_1 = \mathbb{E}[\mathbf{W}_1(t)] = \mathbf{I}_n - \underbrace{\mathbf{L}}_{\text{Graph Laplacian}}/|E|$.
- **Properties:** symmetric, doubly stochastic and

$$\left\|\mathbf{W}_1 - \mathbf{1}_n\mathbf{1}_n^\top/n\right\|_{\text{op}} \leq 1 - \underbrace{c}_{\text{Graph's connectivity}} < 1.$$



(a) Graph topology (*i.e.*, connectivity) has an impact on convergence.



(b) Comparison with previous work: GoRank vs. Chiuso et al.

GoRank shows great performance on large networks and outperforms previous work.

GoTrim – A Gossip Algorithm for Trimmed Means

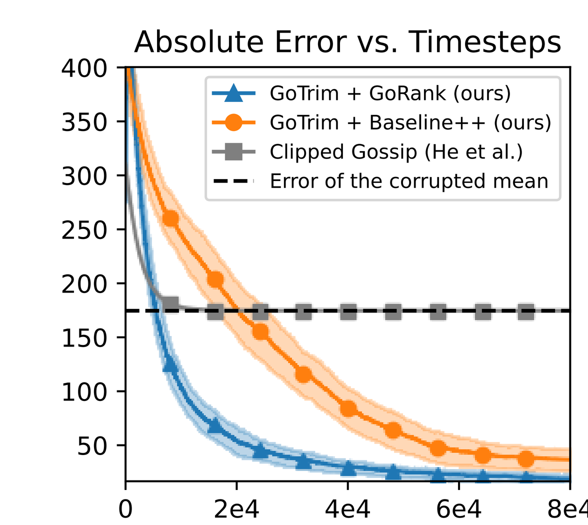
- Computes a **weighted average using current ranks**: $\frac{1}{n}\sum_k f(r_k)\cdot X_k$.
- **Idea 1**: Perform standard (weighted) **gossip-based averaging** using current ranks with weight: $w_k = f(r_k)$.
- **Idea 2**: Since ranks vary, **correct past errors** by injecting: $\left(w_k^{\text{new}} - w_k^{\text{old}}\right)\cdot X_k$.

Theorem 3&4 (Convergence and Robustness of GoTrim Estimates).

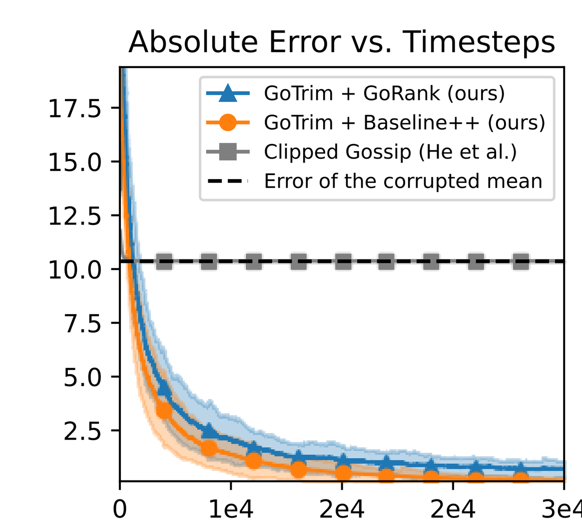
Define $c = \lambda_2/|E|$ the connectivity of the graph. Then, $\forall t \geq T^*$,

$$\|\mathbb{E}[\boldsymbol{Z}(t)] - \bar{x}_\alpha \mathbf{1}_n\| \leq \mathcal{O}\left(\frac{1}{c^2 t}\right).$$

Moreover, the τ-breakdown point at iteration t is $\approx \alpha - \frac{1}{\sqrt{t-T}}$.



(a) Simulated Dataset



(b) Basel Luftklima Dataset

GoTrim quickly improves over the naive corrupted mean.

Key Takeaways & Future Work

✓GoRank and GoTrim show **strong empirical performance**

✓Our Algorithms come with **solid theoretical guarantees**.

✓ We plan to address:

- Extension to multivariate data
- Decentralized robust optimization
- Byzantine Robustness

References

[1] Chiuso et al. Gossip algorithms for distributed ranking. In ACC. IEEE, 2011.

[2] He et al. Byzantine-robust decentralized learning via clippedgossip. *arXiv preprint*, 2022.

[3] Huber et al. *Robust statistics*. John Wiley & Sons, 2011.