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

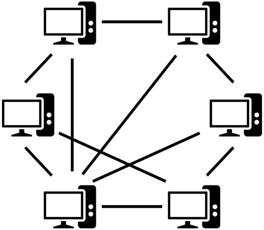
Can we design more robust gossip algorithms?

Anna van Elst, Igor Colin, Stephan Clémençon

LTCI, Télécom Paris, Institut Polytechnique de Paris

Gossip Algorithms: Where and Why?

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

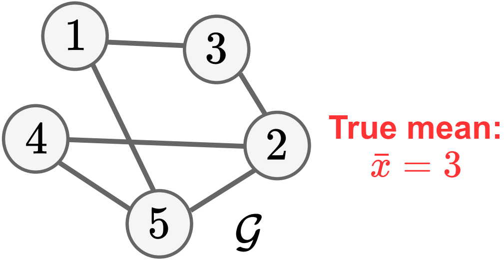


Peer-to-Peer (P2P) Networks

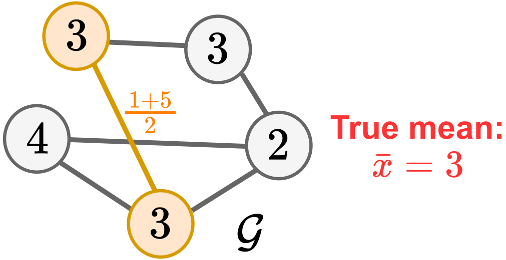
Wireless Sensor Networks

...and many more!

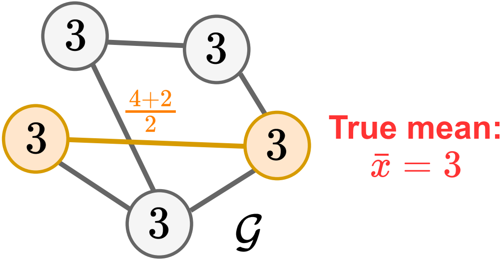
Simple Example: Gossip for Averaging



Select random $(i, j) \in E$, average X_i and X_j



Select random $(i, j) \in E$, average X_i and X_j



Select random $(i, j) \in E$, average X_i and X_j

What Makes the Theoretical Analysis Work?

- Averaging matrix: $t > 0$, for a random edge $(i, j) \in E$,

$$\mathbf{W}_2(t) = \mathbf{I}_n - \frac{1}{2}\left(\mathbf{e}_i - \mathbf{e}_j\right)\left(\mathbf{e}_i - \mathbf{e}_j\right)^\top .$$

- Its expectation with respect to the edge sampling process:

$$\mathbf{W}_2 = \mathbb{E}\left[\mathbf{W}_2(t)\right] = \mathbf{I}_n - \underbrace{\mathbf{L}}_{\text{Graph Laplacian}}/(2|E|).$$

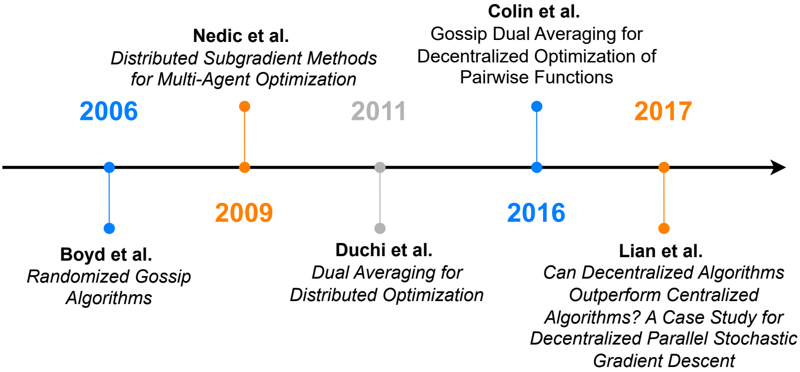
- Properties: symmetric, doubly stochastic and

$$\left\|\mathbf{W}_2 - \mathbf{1}_n\mathbf{1}_n^\top/n\right\|_{\mathrm{op}} \leq 1 - \underbrace{c}_{\text{Graph's connectivity}}/2 < 1.$$

- Convergence in expectation:

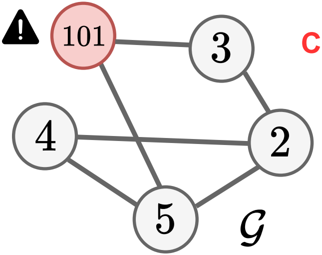
$$\|\mathbb{E}[\mathbf{Y}(t)] - \bar{x}_n\mathbf{1}_n\| \leq \left(1 - \frac{c}{2}\right)^t \|\mathbf{X} - \bar{x}_n\mathbf{1}_n\| .$$

Applications of Gossip for Decentralized Learning



When Gossip Goes Wrong...

- What if a node's value is corrupted? All estimates will be corrupted!
- The average is not robust to outliers. We need more robust statistics, e.g., trimmed mean (see Huber's robustness).



Corrupted mean:
$\bar{x} = 23$

True mean:
$\bar{x} = 3$

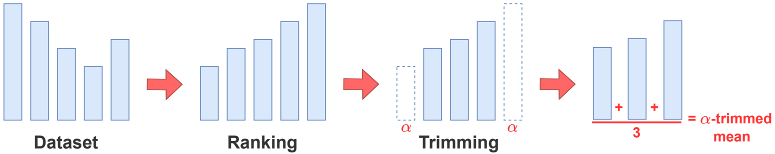
Robust Statistics: Trimmed Means

- Robust average: denoting $m = \lfloor n\alpha \rfloor$,

Order statistic

$$\underbrace{\bar{x}_\alpha}_{\alpha\text{-trimmed mean}} = \frac{1}{n-2m} \sum_{k=m+1}^{n-m} X_{n(k)}$$

- Breakdown point is equal to α.



Problem Formulation and Framework

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

Using ranks, we can compute a robust average:

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

GoRank – A Gossip Algorithm for Ranking (Asynchronous)

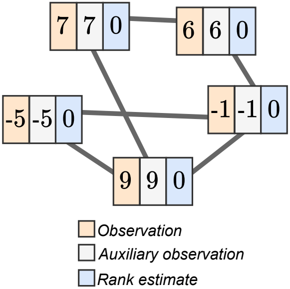
$$r_k = 1 + n\left(\frac{1}{n}\sum_{l=1}^{n}\mathbb{I}_{\{X_k > X_l\}}\right) = 1 + nr'_k$$

Rank of node k

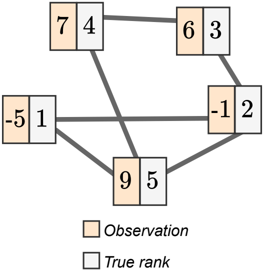
# of nodes

Pairwise comparison

Estimates:

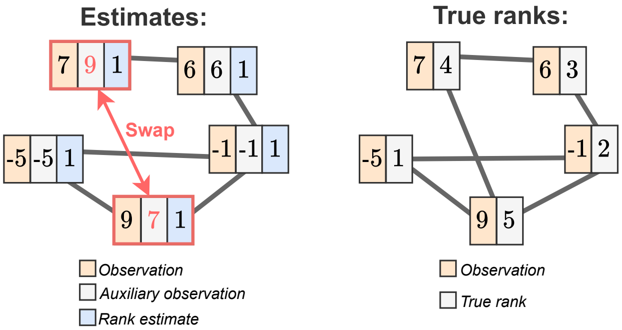


True ranks:



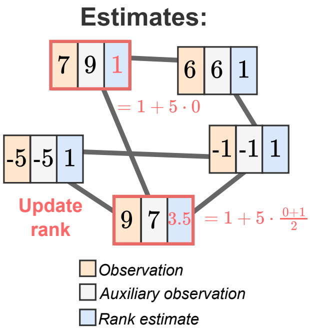
GoRank – A Gossip Algorithm for Ranking (Asynchronous)

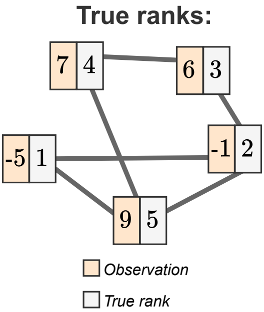
$$\underbrace{R_k(t)}_{\text{Rank estimate}} = 1 + n\left(\frac{1}{\underbrace{C_k(t)}_{\text{Counter}}}\sum_{s=1}^{C_k(t)} \mathbb{I}_{\{X_k > \underbrace{Y_k(s)}_{\text{Auxiliary observation}}\}}\right) = 1 + nR'_k(t)$$



GoRank – A Gossip Algorithm for Ranking (Asynchronous)

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





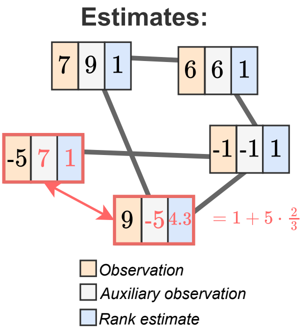
GoRank – A Gossip Algorithm for Ranking (Asynchronous)

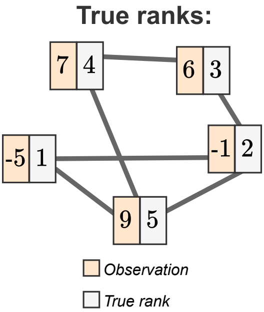
$$\boxed{R_k(t)} = 1 + n\left(\frac{1}{\boxed{C_k(t)}}\sum_{s=1}^{C_k(t)} \mathbb{I}_{\{X_k > \boxed{Y_k(s)}\}}\right) = 1 + nR'_k(t)$$

Rank estimate

Counter

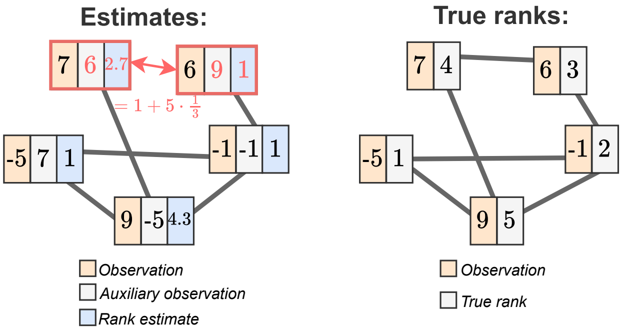
Auxiliary observation



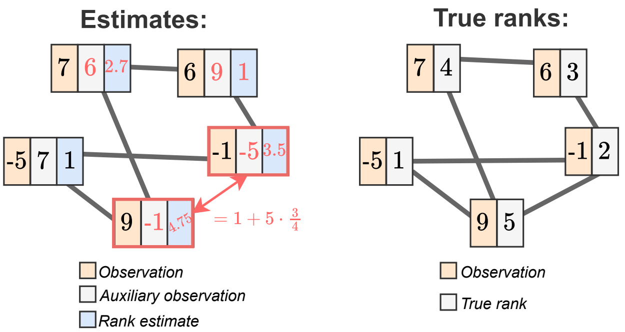


GoRank – A Gossip Algorithm for Ranking (Asynchronous)

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



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



GoRank Shows Great Empirical Performance

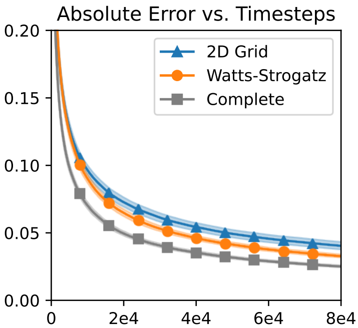


Figure 1: Graph topology has an impact on convergence.

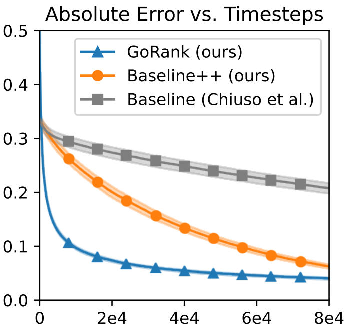


Figure 2: GoRank outperforms previous work in large graphs.

GoRank: Theoretical Analysis

	GoRank (synchronous)
Rank Estimator of Node k	$1 + n \cdot \frac{1}{t} \sum_{s=1}^{t} \mathbb{I}_{\{X_k > Y_k(s)\}}$
Convergence in Expectation (Theorem 1)	$\mathcal{O}\left(\frac{\sigma_k}{ct}\right)$
Expected Abs. Error (Theorem 2)	$\mathcal{O}\left(\frac{\sigma_k}{\sqrt{ct}}\right)$

Table 1: Convergence Rate Bounds.

σ_k = standard deviation

c = spectral gap of the Laplacian

"The absolute error converges in $\mathcal{O}\left(\frac{1}{\sqrt{t}}\right)$."

GoRank: What Makes the Theoretical Analysis Work?

- Swapping matrix: $t > 0$, for a random edge $(i, j) \in E$,

$$\boxed{\mathbf{W}_1(t) = \mathbf{I}_n - (\mathbf{e}_i - \mathbf{e}_j)(\mathbf{e}_i - \mathbf{e}_j)^\top .}$$

- Its expectation with respect to the edge sampling process:

$$\mathbf{W}_1 = \mathbb{E}[\mathbf{W}_1(t)] = \mathbf{I}_n - \boxed{\mathbf{L}}/|E|.$$

Graph Laplacian

- Properties: symmetric, doubly stochastic and

$$\left\|\mathbf{W}_1 - \mathbf{1}_n\mathbf{1}_n^\top/n\right\|_{\mathrm{op}} \leq 1 - \boxed{c} < 1.$$

Graph's connectivity

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:

$$\text{weight} = f(r_k).$$

- **Idea 2**: Since ranks vary, correct past errors by injecting:

$$(\text{weight}^{\text{new}} - \text{weight}^{\text{prev}}) \cdot X_k\,.$$

GoTrim Shows Great Empirical Performance

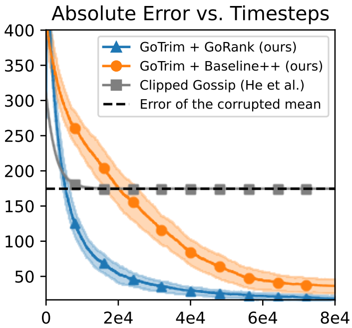


Figure 3: Simulated Dataset

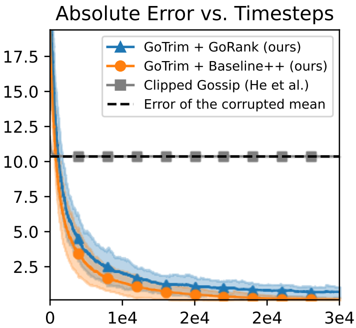


Figure 4: Basel Luftklima Dataset

"GoTrim quickly improves over the corrupted mean."

GoTrim: Theoretical Analysis

	GoTrim (synchronous)
Trimmed Mean Estimator of Node k	$\frac{1}{n}\sum_{k=1}^{n} f(R_k) \cdot X_k$
Convergence in Expectation (Theorem 3)	$\mathcal{O}\left(\frac{1}{c^2 t}\right)$
τ-Breakdown Point (Theorem 4)	$\approx \alpha - \frac{1}{\sqrt{t-T}}$

Table 2: Convergence Rate Bounds.

c = spectral gap of the Laplacian

"The bias converges in $\mathcal{O}\left(\frac{1}{t}\right)$."

Conclusion and Future Work

- We design novel gossip algorithms for ranking (GoRank) and trimmed means estimation (GoTrim).
- We show strong empirical performance and have solid theoretical guarantees.
- **Future work**:
  - Extension to multivariate data
  - Decentralized robust optimization
  - Byzantine Robustness

Our paper:

References

[1] Peter J Huber and Elvezio M Ronchetti. **Robust statistics.** John Wiley & Sons, 2011.

[2] Stephen Boyd et al. **"Randomized gossip algorithms".** In: *IEEE transactions on information theory* 52.6 (2006), pp. 2508–2530.

[3] Igor Colin et al. **"Extending gossip algorithms to distributed estimation of U-statistics".** In: *NeurIPS* 28 (2015).

[4] Alessandro Chiuso et al. **"Gossip algorithms for distributed ranking".** In: *Proceedings of the 2011 American Control Conference*. IEEE. 2011, pp. 5468–5473.

[5] Lie He, Sai Praneeth Karimireddy, and Martin Jaggi. **"Byzantine-robust decentralized learning via clippedgossip".** In: *arXiv preprint arXiv:2202.01545* (2022).