Tournesol: The Game-Theoretical Challenges of Developing a Secure Online Collaborative Platform

Lê Nguyên Hoang, Calicarpa & Tournesol

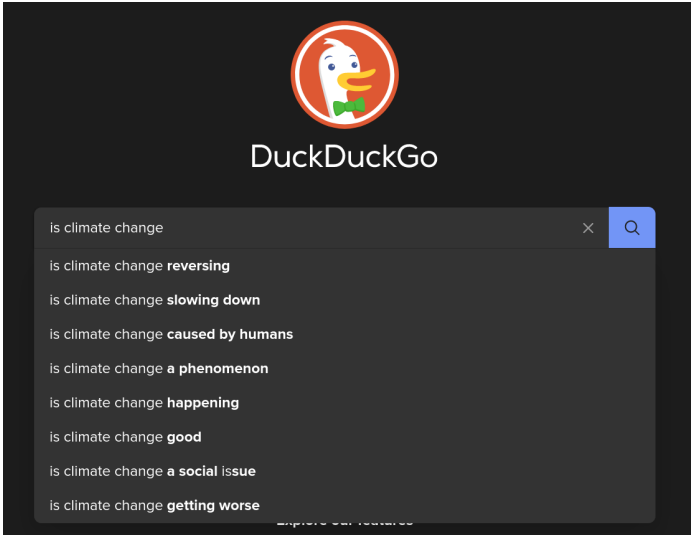
Joint work with Adrien Matissart, Louis Faucon, Aidan Jungo, Romain Beylerian, Julien Fageot, Youssef Allouah, Rachid Guerraoui, Oscar Villemaud, Sadegh Farhadkhani...

Workshop Dynamic Games and Applications, October 2023

Section 1

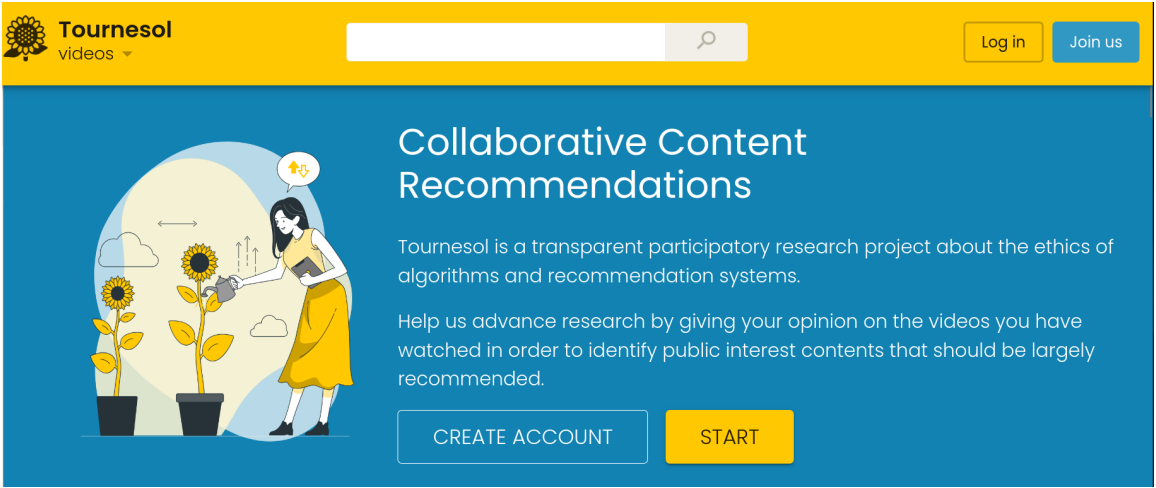
The context

Millions of billions of dilemmas...



... currently solved by an implicit game!

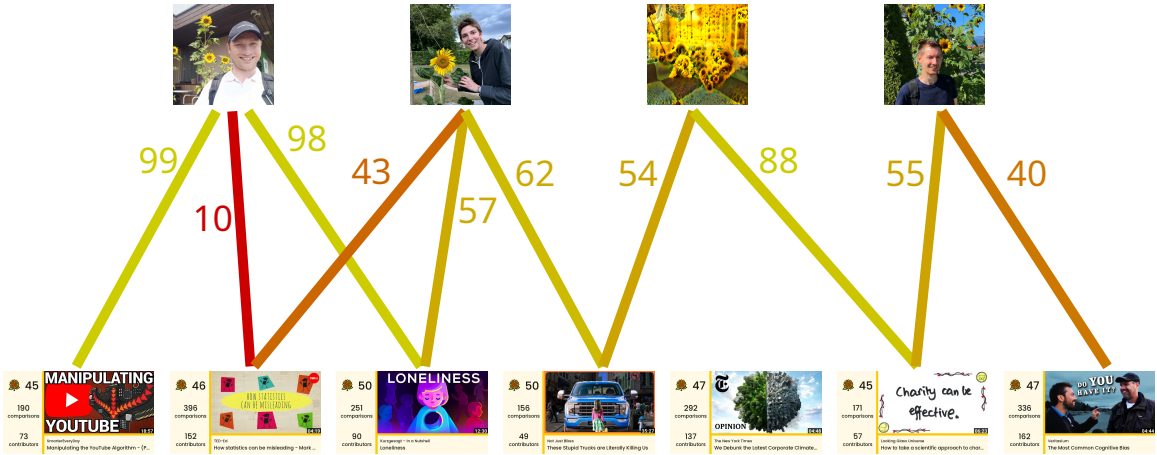
Tournesol


Tournesol
videos
Log in
Join us
Collaborative Content Recommendations
Tournesol is a transparent participatory research project about the ethics of algorithms and recommendation systems.
Help us advance research by giving your opinion on the videos you have watched in order to identify public interest contents that should be largely recommended.
CREATE ACCOUNT
START

Section 2

The robust sparse voting problem

Sparse voting



The French reviewers problems

Note: Some of my best friends are Parisian and Marseillais.

The French reviewers problems

Note: Some of my best friends are Parisian and Marseillais.

The Parisien reviewer problem

Some content may be mostly scored by complain-addict reviewers.

The French reviewers problems

Note: Some of my best friends are Parisian and Marseillais.

The Parisien reviewer problem

Some content may be mostly scored by complain-addict reviewers.

The Marseillais reviewer problem

Some content may be mostly scored by exaggeration-addict reviewers.

Our resilience to diverging expression styles

Von Neumann - Morgenstern utility functions

Cardinal preferences are defined up to any positive affine transformation.

Our resilience to diverging expression styles

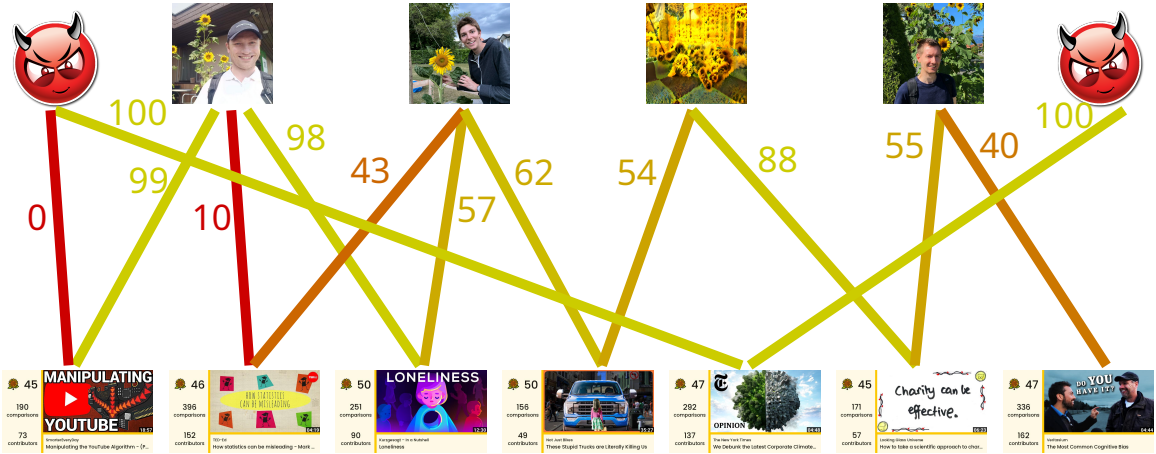
Von Neumann - Morgenstern utility functions

Cardinal preferences are defined up to any positive affine transformation.

Sparse unanimity (informal)

If all voters have the same VNM preference, then we must recover it, even under sparse voting.

Robust sparse voting



Our voting security guarantee

Lipschitz resilience (informal)

The vote is L-Lipschitz resilient if the votes of a contributor only affect the output scores by at most L.

Our voting security guarantee

Lipschitz resilience (informal)

The vote is L-Lipschitz resilient if the votes of a contributor only affect the output scores by at most L.

Generalization to the case of continuous voting rights

L-Lipschitz resilience is equivalent to demanding that the vote be L-Lipschitz continuous in voting rights (considering ℓ_1 norm for voting rights vector, and ℓ_∞ norm for output scores).

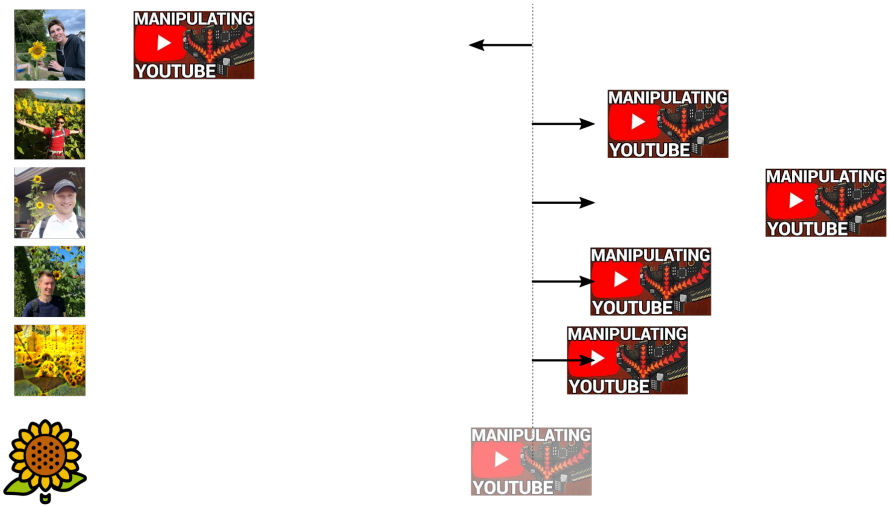
The big question

Is there a sparsely unanimous and
Lipschitz resilient voting algorithm?

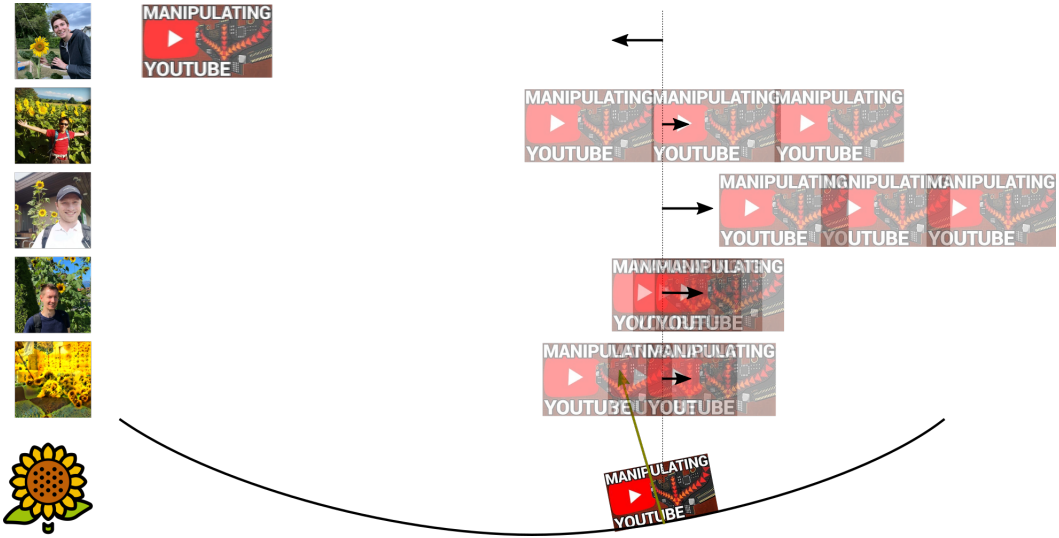
Section 3

Our results

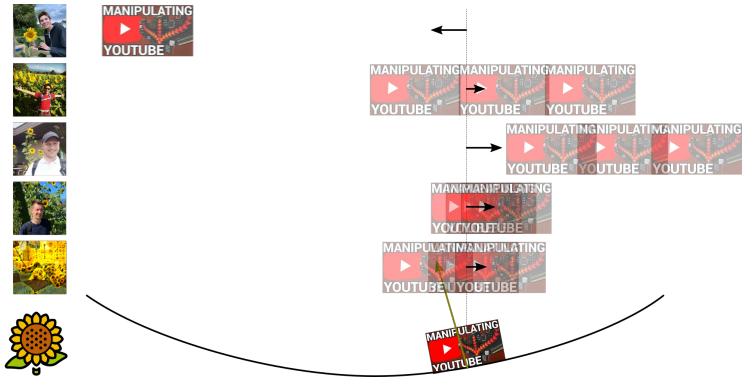
Our key primitive: The Quadratically-Regularized Median



Our key primitive: The Quadratically-Regularized Median



Our key primitive: The Quadratically-Regularized Median



Theorem

$\mathrm{QrMed}_L(\mathbf{x}) \triangleq \arg\min_{z\in\mathbb{R}} \left\{ \frac{1}{2L} z^2 + \sum_{i=1}^{n} |x_i - z| \right\}$ is L-Lipschitz resilient.

The Lipschitz-resilient toolbox (`solidago.py`)

Theorem

$LrMean_L(\mathbf{x}) \triangleq ClipMean\left(\mathbf{x} | QrMed_{L/4}(\mathbf{x}), \frac{LN}{4}\right)$ is L-Lipschitz resilient, and outputs the mean of $\mathbf{x}$ when N is large enough.

The Lipschitz-resilient toolbox (`solidago.py`)

Theorem

$LrMean_L(\mathbf{x}) \triangleq ClipMean\left(\mathbf{x} | QrMed_{L/4}(\mathbf{x}), \frac{LN}{4}\right)$ is L-Lipschitz resilient, and outputs the mean of $\mathbf{x}$ when N is large enough.

Our open-source library `solidago.py` contains other Lipschitz-resilient estimators for quantiles, standard deviation, polarization...

Collaborative scaling is critical

Theorem (informal)

No vote based on individual scaling can guarantee sparse unanimity and Lipschitz resilience.

Our solution: Mehestan

Definition (informal)

1. Min-max normalize each voter's score vector.
2. For each pair of voter (i, j), compare their scaling on alternatives they both scored.
3. For each voter i, aggregate relative scaling compared to j's, using *LrMean*, yielding voter i's score rescaling.
4. Aggregate rescaled scores with *QrMed*.

Our solution: Mehestan

Definition (informal)

1. Min-max normalize each voter's score vector.
2. For each pair of voter (i, j), compare their scaling on alternatives they both scored.
3. For each voter i, aggregate relative scaling compared to j's, using *LrMean*, yielding voter i's score rescaling.
4. Aggregate rescaled scores with *QrMed*.

Theorem (Allouah, Guerraoui, H, Villemaud (2021))

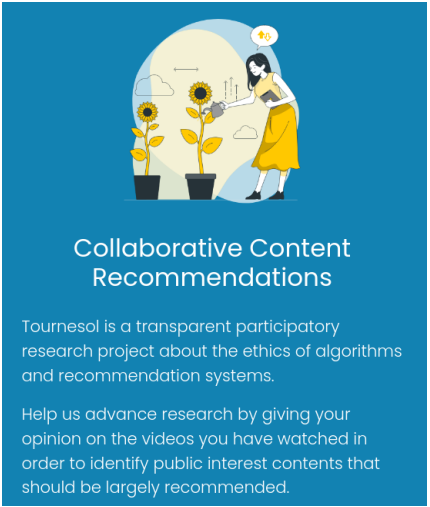
Mehestan is sparsely unanimous and Lipschitz resilience.

Section 4

Conclusion

Tournesol's ambitious goal: Make the web democratic

A flood of exciting governance research questions!



- Proof of Personhood.
- Liquid democracy.
- Web of Trust.
- Lipschitz collaborative filtering.
- Lipschitz Bayesian voting.
- Active learning.
- Diverse and fair recommendations.
- Human-computer interface.
- Cognitive impacts on end user.
- Volition learning.
- Presumption of non-recommendability.
- `tournesol.app/#research`