

Younes Ben Mazziane

University of Avignon

✉ younesbenmazziane@gmail.com

Education

Postdoc

UNIVERSITY OF AVIGNON

Avignon, France

Nov. 2024 - now

- **Supervisors:** Francesco De Pellegrini and Eitan Altman
- **Topic:** Bidding Strategies in Proportional Allocation Auctions

Ph.D. in Computer Science

CENTRE INRIA D'UNIVERSITÉ CÔTE D'AZUR

Sophia Antipolis, France

Oct. 2020 - Oct. 2024

- **Supervisors:** Sara Alouf and Giovanni Neglia
- **Title:** Probabilistic Analysis for Caching

M.Sc. in Computer Science (Ubiquitous Networking)

POLYTECH NICE SOPHIA UNIVERSITÉ CÔTE D'AZUR

Sophia Antipolis, France

2019 - 2020

- **Scholarship:** Labex UCN@Sophia

M.Sc. in Electrical and Computer Engineering

INSTITUT NATIONAL DES POSTES ET TÉLÉCOMMUNICATIONS

Rabat, Morocco

2017 - 2020

Classes Préparatoires aux Grandes Ecoles (MP)

LYCÉE OMAR IBN LKHATTAB

Meknes, Morocco

2015 - 2017

Research Overview & Publications

My Ph.D. thesis, Probabilistic Analysis for Caching, provides rigorous performance guarantees for probabilistic streaming and caching algorithms. The analysis of these algorithms is complex, as it involves studying analytically intractable stochastic processes. We circumvent this difficulty by introducing novel surrogate processes that capture the real dynamics yet remain tractable. These surrogates are then analyzed with Markov-process techniques, renewal theory, concentration inequalities, or online convex optimization, yielding tight performance bounds.

In my postdoc, I'm working on repeated games arising from the competition among agents in proportional allocation auctions. The objective is to capture the system's evolution based on agents' selfish behavior. To answer these questions, our study integrates concepts from game theory and multi-agent online convex optimization.

STREAMING ALGORITHMS: COUNT-MIN SKETCH WITH CONSERVATIVE UPDATES

Context. Online counting of items' occurrences in a data stream with a high arrival rate is crucial to extract key features of the stream, e.g., detection of top k items. Because the stream may contain billions of distinct items, streaming algorithms are forced to operate with a memory sublinear in the number of distinct items. Under such tight space constraints, streaming tasks like detecting the top- k items inevitably incur some errors. The Count-Min Sketch (CMS) is a widely adopted solution: it keeps a compact array of counters that supports approximate frequency queries for each item on the stream, offering a trade-off between memory and accuracy. A refinement, CMS with Conservative Update (CMS-CU), boosts accuracy but unlike the vanilla CMS whose memory-accuracy trade-off is well understood, theoretical guarantees for CMS-CU are still missing.

Contributions We develop tractable analytic frameworks that bound the frequency estimation error of CMS-CU in two complementary regimes: (i) frequent items under stationary stochastic streams [C1, J1] and (ii) infrequent items under arbitrary streams [C5, J3]. For both settings, the main obstacle is dealing with processes similar to "balls and bins with power of d ," except that, on ties, all minimal bins among selected ones are incremented

instead of a tie-breaking rule where a single bin is incremented. In [C1, J1], we show that the analysis boils down to studying a constrained variant of the aforementioned process: a random hypergraph is drawn once, its vertices are the counters, each hyperedge encodes an admissible d -tuple, and each hyperedge is selected at every step according to a fixed, non-uniform probability distribution modeling the heterogeneity of the items in the stream. On the other hand, in [C5, J3], all hyperedges are admissible and have homogeneous probabilities of being selected.

[J3] Younes Ben Mazziane, Othmane Marfoq.

Universal and Tight Bounds on Counting Errors of Count-Min Sketch with Conservative Updates.
In **Proc. ACM Meas. Anal. Comput. Syst**, Volume 9, 2025.

[C5] Younes Ben Mazziane, Othmane Marfoq.

Universal and Tight Bounds on Counting Errors of Count-Min Sketch with Conservative Updates.
ACM **SIGMETRICS** 2025 - International Conference on Measurement and Modeling of Computer Systems.

[J1] Younes Ben Mazziane, Sara Alouf, Giovanni Neglia.

Analyzing Count Min Sketch with Conservative Updates.
In **Computer Networks**, Volume 217, 2022.

[C1] Younes Ben Mazziane, Sara Alouf, Giovanni Neglia.

A Formal Analysis of the Count-Min Sketch with Conservative Updates.
In **INFOCOM** 2022 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS).

NO-REGRET CACHING AND SIMILARITY CACHING

Context (No-Regret Caching). Caches are small memories that speed up data retrieval. Caching policies strategically select cache content based on previous requests to maximize a utility function, such as the number of requests answered directly by the cache. Worst-case analysis of caching algorithms evaluates their performance against an optimal policy with knowledge of future requests. One way to do that is to consider the regret metric, which is the difference between the utility of a static optimum and that of the algorithm. Numerous caching policies that leverage online convex optimization and exhibit vanishing regret have been proposed. Nevertheless, these algorithms are computationally expensive and necessitate complete knowledge of prior requests, which may not be feasible in practical scenarios like caching at a cellular base station.

Contributions (No-Regret Caching). In [C3,J4], we are the first to show that a simple adaptation of Follow-The-Perturbed Leader algorithm achieves near-optimal regret bounds in the more restrictive setting of partial observability of the requests and in $O(1)$ time for large request sequences.

Context (Similarity Caching). In this problem, items are represented as points within a metric space, enabling the quantification of similarity between any two items through their distance. Similarity caching extends the traditional caching problem by allowing requests for items to be approximately satisfied with similar cached items. The objective is to optimize a utility function that considers how similar the requested item is to the cached item used for the response.

Contributions (Similarity Caching). In [C2,J2], we proposed an analytical framework for computing the hit ratio, i.e., the fraction of requests satisfied directly by the cache, of an adaptation of the well-known Least-Recently-Used (LRU) policy to similarity caching under stationary requests. This computation is challenging as it involves examining a Markov process with exponentially growing state space. To make the problem tractable, we introduced a surrogate similarity caching policy with independent caching decisions and identified conditions where this fictitious policy closely approximates the LRU-based similarity caching variant.

[J4] **Younes Ben Maziane**, Francescomaria Faticanti, Sara Alouf, Giovanni Neglia.
Efficient and Optimal No-Regret Caching under Partial Observation.
Submitted to **IEEE Transactions on Networking** (Major revision).

[C3] **Younes Ben Maziane**, Francescomaria Faticanti, Giovanni Neglia, Sara Alouf.
No-Regret Caching with Noisy Request Estimates.
In **VCC 2023** - IEEE Virtual Conference on Communications.

[J2] **Younes Ben Maziane**, Sara Alouf, Giovanni Neglia, Daniel S. Menasche.
TTL model for an LRU-based similarity caching policy.
In **Computer Networks**, Volume 241, 2024.

[C2] **Younes Ben Maziane**, Sara Alouf, Giovanni Neglia, Daniel S. Menasche.
Computing the Hit Rate of Similarity Caching.
In **GLOBECOM 2022** - IEEE Global Communications Conference.

REPEATED PROPORTIONAL ALLOCATION AUCTIONS GAMES

Context. Decentralized resource allocation in large-scale systems is a fundamental problem that has been extensively studied in network economics. The Kelly, also called proportional-auction, is a simple and efficient mechanism in which each agent receives a share of the resource proportional to its bid. The resulting game has been shown to achieve near-optimal social welfare at its Nash equilibrium. These guarantees, however, assume that agents can actually play this equilibrium, which in turn requires every player to know the utilities of all others. A more realistic setting assumes that each agent knows only its own utility, yet can adapt its bid over repeated, synchronous rounds using limited feedback, such as the aggregate bid of all players.

Contributions. In [C4], we study the repeated game induced by the Kelly mechanism. We prove that when all agents update their bids via a dual-averaging no-regret algorithm and their utilities grow logarithmically with the resource they receive, the system converges to the Nash equilibrium of the one-shot game.

[C4] **Younes Ben Maziane**, Cleque-Marlain Mboulou-Moutoubi, Francesco De Pellegrini, Eitan Altman.
Learning to Bid in Proportional Allocation Auctions with Budget Constraints.
In **WiOpt 2025** - IEEE International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks.

Oral Presentations

SIGMETRICS 2025. *Universal and Tight Bounds on Counting Errors of Count-Min Sketch with Conservative Updates.*

University of Avignon (LIA) Seminars 2024. *Probabilistic Analysis for Caching and Approximate Counting.*

VCC 2023. *No-Regret Caching with Noisy Request Estimates.*

GLOBECOM 2022. *Computing the Hit Rate of Similarity Caching.*

INFOCOM WORKSHOPS 2022. *A Formal Analysis of the Count-Min Sketch with Conservative Updates.*

Teaching

2022 **Introduction à l'informatique par le Web**, (36h TP)

2023 **Base de Programmation en Python**, (26h TD)

2023 **Programmation Avancée en Python**, (48h)

Université
Côte d'Azur,
France.
Polytech Nice
Sophia,
France
Polytech Nice
Sophia,
France