I am a theoretical computer scientist working on problems in networking and machine learning, and I primarily use combinatorial methods to solve problems. The main theme in my research is to **understand how information, structure, and dynamics interact in communication and learning systems**. I have a broad interest in networking and ML, and in particular I have spent a considerable amount of time studying the effects of communication network structure on data integrity [13, 11, 15, 14, 18, 17, 19]. I also have studied models for DNA data storage [12], graph based private data retrieval [16], and interactive graph learning [4]. I work to understand how data moves and transforms in stochastic or distributed systems, and how we can develop algorithms for data recovery in such systems. I want to find common principles that can be applied through networking, learning theory, and data reconstruction. To this end, the rest of my research statement is organized around the following concrete questions,

1. How is data quality impacted by the topology of a communication network? My work with Michelen [13] effectively answers this for static networks, and builds on my previous work [11]. It is an ongoing research project with Srivastava and Ulukus [18, 19] to answer this for dynamic networks
2. How many imperfect observations are needed to recover a combinatorial object? I am interested in algorithms and lower bounds for two concrete models of this question: reconstructing DNA from subsequences [12], and reconstructing automata from random walks [10].
3. What are the foundations of self-attention in Large Language Models? I am collaborating with researchers across several universities to analyze LLMs through the lens of dynamical systems and interacting particle systems.

1 Gossip Networks

Imagine a network of self-driving cars that wish to share information about congestion, accidents, and signal disruptions. One of the simplest communication protocols used in practice is known as *gossip*, in which each car randomly exchanges information with nearby vehicles over time. A fundamental question that arises is: how new is the information held by a given vehicle at a particular moment? The freshness of stored data is a critical quantity as it could be the difference between an accident versus a near miss. The freshness of data is also known as the **age of information** (AoI), which was formally introduced as a metric in networking by Kaul, Yates, and Gruteser [9].

The model I consider was originally introduced by Roy Yates in a pair of papers [25, 26] and has garnered significant interest in recent years. In this framework, a network is represented by a graph whose vertices exchange packets continuously in time according to independent Poisson processes. As vertices share data, they only store the most up-to-date data they have received so far. This ensures that the whole network is essentially a sensor for the 'state of the world', where separate nodes have imperfect but correlated tracking of the world. My goal is to understand how the overall network structure impacts the quality of sensing as measured by the AoI metric.

Combinatorial Approaches

The original work of Yates [25, 26] is concerned with the long-term average age of vertices in the complete graph and cycle graph. His proof technique utilized tools from control theory to express the AoI as a certain recursion on subsets of the graph. Using tools from the next section, **I've strengthened this foundational result** [15] with no reference to control, and in a more concise proof. The recursive formula is the starting point for many works in age of gossip. The main technical challenge for finding the AoI is that the recursion tree is exponential in size of the graph. I became interested in AoI when I attended a talk by Srivastava and Ulukus [20] where they bounded the recursion using properties of the 2-D square grid. Their proof contained hints of a connection between AoI and the connectivity of the graph. As a PhD student **I identified the Cheeger constant as a combinatorial measure** for AoI in many classes of graphs [11]. This lead me to obtain tight bounds on the AoI in new graph families. In the same paper, I proved AoI is a monotone graph property (solving an open problem and justifying earlier results [21, 8]) and leveraged this fact to **prove a thresholding phenomena in the Erdős-Reyni random graph**. Additionally during my postdoc appointment I've worked closely with PhD student Arunabh Srivastava to **identify networks where the push and pull protocols differ significantly** in their performance [17]. My work has **inspired other researchers** at Southeastern University [22, 23, 24] and the University of North Carolina at Charlotte [6] to study gossip networks from this combinatorial lens.

Probabilistic Approaches and Dynamic Networks

Towards the end of my PhD I began collaborating with Marcus Michelen on other approaches to study AoI beyond the long-term average paradigm introduced in [25]. Our feeling was the Cheeger constant was not the tightest measure for AoI. We quickly realized that the correct measure had to do with *local structure* of the graph, as opposed to the global structure the Cheeger constant captures. In our work published in the Transactions of Information Theory [13], we proved a connection between AoI and the statistical physics model 'first passage percolation' (FPP), which allowed us to **develop a novel combinatorial parameter tightly capturing the AoI in arbitrary graphs**. We leveraged this new connection to **solve numerous open problems** in the area. Notably we solved the following questions from the survey [7]: the AoI scaling on square lattices, random geometric graphs, a cycle + random matching, and found a family of graphs with arbitrary polynomial AoI scaling. In a followup work I used the connection to FPP to recover the recursive identity in [25] with a very short proof and no control theoretic machinery [15]. In particular **my derivation holds in a stronger setting** of infinite, locally finite graphs and the time limit can be replaced with a suitably large stochastic time. In some recent unpublished work I've also found an explicit recursive formula for the moment generating function for FPP on arbitrary graphs, building on methods developed in the AoI community.

My work with Michelen provided a new analytic framework which resolved many open problems, so when I started my postdoc I was interested in moving beyond static graphs. Dynamic graphs are an extremely rich area, and there is ample room for creativity in formulating models and problems. It was pointed out to me by Srivastava that the recursive/ combinatorial methods I was quite familiar with break down in even simple dynamic models. Fortunately with the connection between gossip and FPP we had another tool to tackle problems. I began a collaboration and mentorship with Srivastava, and we started by studying a simple 2-state dynamic model [19] which already required some nontrivial analysis. We showed if the switching between states was fast enough, the network performance does not degrade. We followed this up with a study on arbitrary transitions rates and states [18], and identified regimes of slow, fast, and intermediate scaling. We also identified counterexamples to intuitive conjectures for dynamic models. We are close to submitting a journal paper which **unifies and generalizes** many of these results and methods for dynamic models on a finite state space. Finally, we have a paper **going beyond AoI** investigating misinformation spread and information degradation in networks [14].

2 Reconstruction Problems

In addition to networking question, I am generally interested in problems of the form 'how many noisy observations of a combinatorial object are necessary/ sufficient to recover the original object?' The following two sections highlight some of my work in this direction.

String and Tree Reconstruction

The string reconstruction problem was introduced over 20 years ago as a simplified model for DNA data storage [1]. An arbitrary binary string S is passed through a deletion channel where each bit is removed with probability p — we call a (random) subsequence from this process a *trace* of S. The string reconstruction problem asks 'how many traces are needed to recover an arbitrary string S with success probability $1 - \delta$?' A similar problem was asked in the case of trees [2], and I have proven reductions from tree reconstruction to string reconstruction. My first contribution was to show that **tree reconstruction is no easier than string reconstruction** [12]. This is somewhat surprising, since trees have richer structure we may have hoped to use their structure to obtain better sample complexity bounds. The idea of my reduction was to encode strings as a special type of tree, and show that the tree deletion channel acts essentially as the string channel on these special trees. As a converse, I showed there is a labeling of the vertices of any tree which allow it to be recovered via a coupling to the string reconstruction problem. Therefore, the problem of recovering labelled trees from traces is equivalent to the string reconstruction problem, up to some insignificant polynomial factors. Finally, I proved **numerous identities for the deletion channel** itself, including a connection between bit probabilities and generating functions induced by strings.

Learning DFA from Random Walks

Learning DFA has a rich history tracing back to the early days of learning theory as a model for autonomous robot navigation. Early hardness results for passive learning were complemented by polynomial sample complexity bounds for active and query learning models. I'm interested in a generalization of a learning model proposed by Freund et al. [3]. In the model I consider, an agent performs a random walk on a DFA with random transitions, and random 0-1 labels on each state. The goal of the learning procedure is to eventually make no mistakes when predicting the state labels of the DFA. In the case of an unbiased random walk Freund et al. proved there is an algorithm which stops making mistakes after $O(n^5)$ steps in the random walk. It turns out that their algorithm does not generalize to biased random walks. A significant portion of my PhD thesis [10] is dedicated to constructing a new algorithm which learns using these extremely biased random walks. The main challenge in this setting is that very rare events contribute a significant number of errors in any algorithm. I have **constructed an algorithm that learns the core of the DFA** with polynomially many mistakes via an agnostic learning procedure. There are uncommon, but not very rare events which give the learner useful information that can be leveraged to build a representation of the most visited sites in the DFA. Some of my future work is to extend this core-building algorithm to learn the whole DFA.

3 Foundations of Large Language Models

In my postdoc I've become interested in the mathematics of large language models (LLMs). The key feature that separates LLMs from other neural networks is the *self-attention layer*, which allows the LLM to extract non-Markovian relationships in language. It was recently shown that the self-attention layer behaves as a non-linear ordinary differential equation (ODE) [5]. Current techniques can quantitatively describe the behavior of the self-attention layer for a limited range of parameters, which is not realistic for practitioners. I am currently working with researchers from Rice University, University of Wisconsin Madison, Oxford University, and University of Warsaw to extend the set of analytic techniques available to describe these dynamics. Additionally, I have posted code on my GitHub which allows us to run experiments and visualizations of the dynamics. We are particularly interested in **finding an attracting family of solutions** to the ODE as this would provide insights into how the depth of an LLM impacts the expressivity in outputs. Our techniques mirror some classical work in the Kuramoto model for synchronized oscillators. Our first project is well underway, and we anticipate releasing our study soon.

Future Work

I will continue working on problems in communication, learning, and their intersection. The work I've highlighted above leaves many open problems with varying degrees of difficulty. Some of the easier problems arise naturally from different modeling assumptions in gossip networks, such as considering hypergraphs or different communication protocols. I am more interested in questions that go beyond simple model tweaks. Some examples of problems I'm excited to work on are below.

1. Can we apply combinatorial insights from gossip to solve problems in statistical physics? For instance, there are open problems in the computational complexity of exactly or approximately finding the first passage time between vertices, which may have answers based on tools I've developed. These questions have precedent in the literature, being answered for many other models such as Ising and monomer-dimer systems. The study of the first-passage time on infinite lattices with exponential weights, known as the *Eden Model*, have natural analogues in gossip which we understand very well. I've highlighted some connections in a recent pre-print [15], and will continue exploring this direction moving forward.

2. What is the natural measure of complexity for reconstruction problems? Machine learning theorists have devised many notions of dimension for classes of learning problems (VC dimension for PAC learning, Littlestone dimension for online learning etc.) which capture some notion of hardness for learning. So far no useful measure exists for reconstruction problems (total varaition distance is a measure but is unwieldy for analysis). Already in my PhD thesis [10] I attempted to construct a measure for the string reconstruction problem using discrete Fourier techniques, and I plan to resume this direction, with a broader focus on a general definition of reconstruction problems.

3. The LLM attention mechanism appears to compress an arbitrary initial distribution through its subsequent applications. What is the nature of this (lossy) compression, and why is it so effective as an intermediate stage in generating output text? Our work to find an attracting family of solutions implies bounds on the amount the initial data is compressed. I would like to explore information theoretic explanations as to why this lossy compression still leads to reasonable model performance.

References

[1] Tugkan Batu, Sampath Kannan, Sanjeev Khanna, and Andrew McGregor. Reconstructing strings from random traces. In *SODA*, pages 910–918. SIAM, 2004.

[2] Sami Davies, Miklós Z. Rácz, and Cyrus Rashtchian. Reconstructing trees from traces. In Alina Beygelzimer and Daniel Hsu, editors, *COLT*, volume 99, pages 961–978. PMLR, 2019.

[3] Yoav Freund, Michael Keams, Dana Ron, Ronitt Rubinfeld, Robert E. Schapire, and Linda Sellie. Efficient learning of typical finite automata from random walks. pages 315–324, June 1993. 25th Annual ACM Symposium on Theory of Computing, STOC 1993 ; Conference date: 16-05-1993 Through 18-05-1993.

[4] Xing Gao, Thomas Maranzatto, and Lev Reyzin. A unified analysis of dynamic interactive learning. In *2023 59th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 1–8, 2023.

[5] Borjan Geshkovski, Cyril Letrouit, Yury Polyanskiy, and Philippe Rigollet. A mathematical perspective on transformers, 2023.

[6] Irtiza Hasan and Ahmed Arafa. Version age of information with contact mobility in gossip networks, 2025.

[7] Priyanka Kaswan, Purbesh Mitra, Arunabh Srivastava, and Sennur Ulukus. Age of information in gossip networks: A friendly introduction and literature survey, 2023. Available online at arXiv:2312.16163.

[8] Priyanka Kaswan and Sennur Ulukus. Age of gossip in ring networks in the presence of jamming attacks, 2022.

[9] Sanjit Kaul, Roy Yates, and Marco Gruteser. Real-time status: How often should one update? In *2012 Proceedings IEEE INFOCOM*, pages 2731–2735, 2012.

[10] Thomas Maranzatto. *Combinatorial Methods for Learning and Information Theory*. PhD thesis, University of Illinois Chicago, https://tmaranzatto.github.io/files/Thesis.pdf, 2024.

[11] Thomas Jacob Maranzatto. Age of gossip in random and bipartite networks. In *2024 IEEE International Symposium on Information Theory (ISIT)*, pages 1173–1178, 2024.

[12] Thomas Jacob Maranzatto. Tree trace reconstruction - reductions to string trace reconstruction. In *2024 IEEE International Symposium on Information Theory (ISIT)*, pages 885–890, 2024.

[13] Thomas Jacob Maranzatto and Marcus Michelen. Age of gossip from connective properties via first passage percolation. *IEEE Transactions on Information Theory*, 71(11):9019–9027, 2025.

[14] Thomas Jacob Maranzatto, Arunabh Srivastava, and Sennur Ulukus. Information degradation and misinformation in gossip networks. *2025 IEEE International Symposium on Information Theory (ISIT)*, pages 1–6, 2025.

[15] Thomas Jacob Maranzatto and Sennur Ulukus. Deriving moments in the age of gossip process from percolation, 2025.

[16] Shreya Meel, Xiangliang Kong, Thomas Jacob Maranzatto, Itzhak Tamo, and Sennur Ulukus. Private information retrieval on multigraph-based replicated storage. In *2025 IEEE International Symposium on Information Theory (ISIT)*, pages 1–6, 2025.

[17] Arunabh Srivastava, Thomas Jacob Maranzatto, and Sennur Ulukus. Age of gossip with the push-pull protocol, 2024.

[18] Arunabh Srivastava, Thomas Jacob Maranzatto, and Sennur Ulukus. Age of gossip with time-varying topologies. In *2025 IEEE International Symposium on Information Theory (ISIT)*, pages 1–6, 2025.

[19] Arunabh Srivastava, Thomas Jacob Maranzatto, and Sennur Ulukus. Information freshness in dynamic gossip networks, 2025.

[20] Arunabh Srivastava and Sennur Ulukus. Age of gossip on a grid. In *2023 59th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, September 2023.

[21] Arunabh Srivastava and Sennur Ulukus. Network connectivity–information freshness tradeoff in information dissemination over networks, 2024. Available online at arXiv:2405.19310.

[22] Han Xu, Jiayu Pan, Yinfei Xu, Shuo Shao, and Tiecheng Song. Timely gossip on lines: Hybrid ageing. In *2025 IEEE International Symposium on Information Theory (ISIT)*, pages 1–6, 2025.

[23] Han Xu, Yinfei Xu, and Tiecheng Song. Timely gossip with age-dependent networks. In *2024 IEEE International Symposium on Information Theory (ISIT)*, pages 1167–1172, 2024.

[24] Han Xu, Yinfei Xu, and Tiecheng Song. Moment analysis of age-dependent gossip networks. *IEEE Transactions on Information Theory*, 71(1):1–22, 2025.

[25] Roy D. Yates. The age of gossip in networks. In *2021 IEEE International Symposium on Information Theory (ISIT)*. IEEE, July 2021.

[26] Roy D. Yates. Timely gossip. In *2021 IEEE 22nd International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pages 331–335, 2021.