Loqman Salamatian - Research Statement

My research develops rigorous models that piece together the **hidden structure of the Internet** from the fragments that can be directly observed. I focus on where visibility is most limited and where that lack of visibility most constrains our ability to improve the network: **within private infrastructures** (Curvature-Based Analysis, SIGMETRICS'22 [4] / CACM Research Highlight [6]), at the **borders between networks** (MetAScritic, IMC'24 [2]), where **users actually connect** (APNIC, IMC'24 [2]), and during **performance degradations** (HERMES, NSDI'26 under review [1]). The models I develop provide operators, researchers, and policymakers with the insights needed to **diagnose problems more efficiently** and **build a more resilient network.**

Research vision. The Internet is a **critical infrastructure.** Its reliability underpins the reliability of the systems built on top of it, from AI to cloud services. Yet its operation depends on thousands of independently run networks that must coordinate without central control. Because the **Internet's decentralized design** precludes any unified view of its whole, operators, researchers, and policymakers must base their decisions on fragments: routing updates, which show the paths networks choose to advertise; traceroutes, which reveal the routes packets take from a handful of scattered vantage points across the Internet; speed tests, which measure throughput and latency from the user's device but rely on voluntary participation, yielding dense data where users experience issues and little where they do not. With incomplete data, operators often **misdiagnose failures**, **fix symptoms** instead of causes, and implement changes that unintentionally trigger **cascading failures**.

For decades, that limited visibility was manageable: redundancy, human expertise, and loosely coupled systems kept the Internet resilient despite its blind spots. But traffic that once traversed measurable networks now flows through opaque private backbones, via encrypted protocols and national infrastructures shaped by political interests. The result is a widening gap between how much we depend on the Internet and how little we understand its behavior.

To overcome these limits, I build **models and systems that extend traditional measurements with explainable inference,** enabling us to understand the Internet's structure and properties even where direct observation falls short. Each model reports its confidence in each inference, enabling operators to know when to trust an inference and when more measurement is needed. This work presents both **an engineering challenge**—building systems that can process and act on Internet-scale data in real-time—and **a scientific one**—building models that explain Internet behavior from incomplete measurements and make their reasoning explicit and transparent. For example, in my project **HERMES** [1], we built a large-scale measurement system that continuously processes millions of user-initiated speed tests and executes millions of traceroutes each day. On their own, this data is too irregular and noisy to diagnose Internet problems, but **when combined and rigorously modeled**, they reveal **network degradations** and where along the measured path they occur. When the existing signals are insufficient, the system automatically coordinates and launches targeted follow-up measurements, guided by the model, to resolve the ambiguity in real-time. In this way, **theory and measurement work in tandem**: the model interprets what we see, and, when the explanation is underdetermined by existing samples, the model directs the system to issue more measurements. Together, they reveal the origin of the slowdown and guide operators toward the segments that should be avoided to restore normal performance.

The challenge of **working with fragments** is not confined to performance data; it surfaces across many facets of the Internet, including the **geographically distributed physical infrastructure** that carries traffic, the **topology of interconnected networks**, and the **users** whose presence and activity drive demand. My contribution is a **unified approach** that treats missing data as an inherent property of the Internet that must be explicitly modeled. In **Curvature-Based Analysis** [4], I introduced a geometric framework that uses end-to-end latency to reconstruct the shape of otherwise opaque private backbones and pinpoint the narrow corridors where geography and deployment choices make these networks most vulnerable to disruption. In **MetAScritic** [2], I reframed the problem of discovering the Internet's interconnection as an inference task, using the parts of the network we can measure to implicitly learn the economic and geographic patterns that shape these relationships and to fill in the parts we cannot see directly. Across all these projects,

I build **theory-driven systems** that use **existing measurements** to infer the **network's hidden structure** and, when **uncertainty remains**, **guide new measurements** to resolve it.

Geography: A Geometric View of Cloud Backbone Connectivity. At its core, the Internet is a **physical network shaped by geography.** Because distance and physical path determine fundamental limits on latency and redundancy, geography is the source of many performance characteristics and failure modes on the Internet. But these geographic imprints are increasingly hidden, as the networks that now carry most of the Internet's traffic—Google, Azure, and AWS—deliberately obfuscate ***traceroute***, the traditional tool for studying connectivity. This lack of transparency is a real structural risk for everyone since any one of these networks experiencing an outage can render most of the Web inaccessible. This raises a fundamental challenge: *how can we learn about infrastructures that resist direct observation?* Our work demonstrates that geometry offers a promising path forward. Even when traceroute fails, the delays that packets experience still carry a structured imprint of geography. By **treating latency as a manifold** laid over physical space, we can recover properties of connectivity that would otherwise remain latent. In our paper [4], we demonstrated this by constructing manifold views of **Google, Azure, and AWS** global networks. The key insight is that curvature—a measure of how paths deviate from straight-line distance—encodes the network's hidden structure. In particular, negative curvature marks the critical bridges that tie regional clusters together, revealing the essential backbone links. The resulting manifold views reveal the physical topology underpinning these modern cloud networks. For example, they show that Microsoft's Azure backbone forms two tightly coupled clusters, North America–Europe and Asia–Pacific, joined by a fragile Tokyo–Seattle bridge, whereas Google's backbone offers richer transpacific diversity but no direct Europe–Asia path. These geometric insights offer **a quantitative perspective** on how the global **backbones** of these cloud providers **differ**, based solely on end-to-end latency, and highlight where **each remains most vulnerable** to cable failures or geopolitical disruptions.

Our follow-up system, **Matisse** [9, 10], extended this approach by incorporating **time as an extra dimension**. Instead of a static manifold, it models Internet latency as a dynamic surface that evolves with daily network conditions, transforming latency time series into smooth surfaces that show how networks "breathe": where congestion predictably recurs, which intercontinental links oscillate between alternate routes, and how delay surfaces warp in response to events. This dynamic perspective enables operators to **track the health of network paths** over time and quickly **pinpoint degraded regions**.

Topology Inference: Recovering Hidden Connectivity with MetAScritic. On top of the Internet's **geographic substrate** lies its **network topology**—the map of which networks connect to which. This view is central to operators as they rely on it to **assess resilience** and **plan interconnection**, while researchers use it to study **routing behavior** and **security risks**. Yet, despite decades of efforts, most inter-network connectivity remains hidden to anyone who is not part of the networks on each side of a given link. The primary public data sources, BGP collectors, rely on a small number of networks that **voluntarily** share some of the routes they observe from their viewpoints. Each contributor exposes only the routes it has learned from its own neighbors, who may themselves export only a subset of what they observe. As a result, **visibility fades quickly** beyond the cooperating networks. The links that matter most—those between large content providers (e.g., Google) and the access networks serving residential users (e.g., Spectrum)—are therefore among the least visible.

To recover these **missing connections**, we must first understand where **networks have the opportunity to interconnect;** that is, we need to know where they could physically meet. Our earlier work, **iGDB** [10], laid this essential foundation by systematically mapping the Internet's physical and logical layers—**identifying the facilities and cities where each network operates**. By establishing where networks are physically colocated, iGDB provides the geographic context needed to reason about potential interconnections. But physical opportunity alone does not guarantee a connection. My system **MetAScritic** [2] tackles the core inference problem: **determining which of these possible links actually exist**. We model the interconnection matrix city by city, with entries indicating whether two networks are connected, and use the observed entries (known links) and network attributes to infer the unobserved ones using **matrix completion**. Our key hypothesis is that **interconnection decisions follow common strategies** shaped by economics, geography, and traffic demand. As a result, the inter-network connectivity matrix is **low-rank**: peering choices can be explained by a small set of latent factors. However, Internet measurement data violate the usual assumptions of matrix

completion: observations are sparse and skewed toward a few networks. To make this formulation viable under these constraints and to produce outputs that operators can interpret and act on, we introduce three specific extensions. First, in addition to known (measured) links, we incorporate **negative evidence**: when a vantage point consistently probes both networks but never observes a link, that absence is treated as evidence that the link does not exist. Second, we developed a theoretical framework for **measurement bias**, showing that public data is skewed toward networks with many vantage points or cooperative operators, while less open networks remain largely unmeasured. To correct for this bias, MetAScritic issues targeted measurements to under-observed networks, actively probing the most uncertain regions of the matrix, thereby expanding visibility in its existing blind spots. Third, we made the system **explainable**: each inferred link comes with a confidence score and a rationale based on concrete features (e.g., network size and type), so operators can both understand why the link was inferred and choose the false-positive/false-negative trade-off that best fits their use case. Before **MetAScritic**, the only available operating points were the public view of the topology (high FN, low FP) and the full-mesh colocation view (low FN, high FP). **MetAScritic** is the first technique to operate between these extremes. Even while operating in the low false positive region, **MetAScritic** produces the **most complete public Internet** map to date, inferring **34x more connections** than existing public datasets reveal; at that operating point, its inferences achieve about **86% accuracy** when validated against independent sources, including operator reports, border-router configurations, and interconnection facilities' internal data.

Performance Tomography: Detecting and Localizing Degradations. Most Internet performance monitors are designed to catch large**, binary failures**—links going down or outages across regions. But for users, the Internet often "fails" in subtler ways: a Zoom call freezing, a video buffering, or a game lagging. Detecting these degradations is difficult. The only public source of performance monitoring data is **user-initiated speed tests**. Tests are run at **irregular times**, usually when users already suspect a problem, and are **prone to noise or problems that are specific to a user's residence** (e.g., poor WiFi configuration). And even when performance drops are visible, it is hard to pinpoint **where along the path** they occur. Without knowing where the problem lies, operators cannot determine whether to **reroute traffic**, **contact another network**, or **fix their own network**. Diagnosing the source also requires visibility into **both directional paths**, since packets may follow entirely different routes each way, and the real cause of a performance problem may lie on the unseen return path.

HERMES [1] addresses these challenges by converting **irregular and biased measurements** into **reliable population-level signals**. A central difficulty is **irregular sampling**: users run tests at unpredictable times, resulting in some hours and locations having dense data, while others have almost none. HERMES corrects for this imbalance by aggregating tests for a given network within a metro area, building time-varying baselines—essentially rolling estimates of "normal" performance—and weighting evidence so that bursts of activity from a single user do not distort population-level trends. We also designed hypothesis tests tailored to crowd-sourced data: rather than assuming independent, uniformly sampled measurements, HERMES combines multiple statistical tests, introducing a novel Wasserstein-distance test, to flag only statistically significant degradations at the group level. To localize events, it reconstructs paths in both directions and looks for common segments shared across affected users, distinguishing between forward-path and return-path issues. And when several explanations remain possible (e.g., multiple metros or candidate networks), it uses a small, carefully chosen set of follow-up measurement probes to disambiguate the real source of the event. In practice, HERMES has identified **hundreds of bottlenecks** at ISPs, **tens of thousands of slowdowns** that have affected users' connections, and rerouting events that have left users on suboptimal paths since its deployment in September 2024. Despite relying solely on user-initiated speed tests, HERMES recovers a view of performance degradations that closely aligns with multiple independent sources of ground truth (including Google's extensive telemetry). The agreement with these sources indicates that opportunistic user measurements, when processed appropriately, are sufficient for monitoring performance at the Internet scale. Moreover, HERMES is the **first system to detect performance drops caused by the reverse path**, an entire class of degradations currently invisible from the perspective of networks delivering content to users.

Future Research. In the next stage of my work, I aim to **make Internet measurement both explanatory and actionable,** developing **models** that clarify why the network behaves as it does, and **systems** that help operators and policymakers act on those explanations.

1. Explanatory Foundations. My research already begins to answer **why the Internet behaves as it does**: **Matisse** recovers hidden infrastructure from the geographic logic of latency, **MetAScritic** explains interconnections in terms of the incentives that drive them, and **HERMES** traces degradations back to their source. The next step is to make this explanatory goal systematic. Today's measurements show **symptoms**, **not causes.** We observe latency spikes, packet loss, and outages, but not whether they stem from overloaded links, routing shifts, or other mechanisms. In my recent HotNets'25 paper [5], I outline how **tools from causal inference** can be adapted to rigorously answer causal questions about the Internet—for example, did the billions spent on rural broadband through the BEAD program actually narrow the performance gap?

Causal inference provides the **language for establishing why relationships exist**. Yet it offers little intuition about how these relationships reshape the network's spatial structure or propagate through its topology. My second goal is therefore to **reconcile topology and geography into a single object** of study by developing a collection of latency manifolds stitched together at their points of interconnection. In this abstraction, each network forms its own manifold, which is internally optimized to minimize path length and delay. At their borders, however, interconnection points act as **"gravitational distortions"** that bend paths under the influence of economic and political forces. This framework embeds **topology and performance in a single geometry of space and time**.

Unifying these causal and geometric perspectives opens a **path toward an interventional Internet science** that can model *counterfactual manifolds*, or how the network's structure and behavior would have changed under alternative decisions or failures. However, the ability to explore these counterfactuals depends on having a principled notion of what a healthy manifold looks like, including quantitative measures of resilience, efficiency, and fairness, as well as a metric to assess how far one network state lies from another. Drawing inspiration from **biology**, where researchers compare cellular manifolds across healthy and diseased trajectories, **I aim to develop a geometric–causal framework for the Internet that detects when the network is drifting away from its desirable state, explains why, and provides operators and policymakers with the tools to steer it back toward health**.

2. Actionable Systems. Users often experience failures that are **missed by any public observatory**. When evaluating **HERMES**, I found numerous posts on public forums where people complained about buffering video or unexplained packet loss. For these users, the only recourse was to call their ISP or ask online if others were affected—a process that frustrates customers and triggers unnecessary technician visits when issues could have been diagnosed remotely. I plan to address this problem with **agentic debugging systems**: fleets of lightweight measurement agents that collaborate across the network to localize faults in real time. When a user experiences a problem, these agents will **coordinate diagnostics** under the guidance of a **centralized controller**, distinguishing between home Wi-Fi issues and bottlenecks along the route. This controller would orchestrate agents, using optimization schemes similar to those in **MetAScritic** and **HERMES**. The long-term goal is a **pervasive debugging service for the Internet** that is always available to users, reducing wasted visits, accelerating diagnosis, and producing actionable explanations that identify the likely root cause, quantify its impact, and indicate who can fix it (e.g., the user, the ISP, or a downstream provider).

This vision connects to a broader research agenda: developing a **theory of measurement-driven optimization**. Networking research has long oscillated between theory and empiricism, between models that idealize the Internet and measurements that reveal its actual behavior. Too often, we have tried to make measurements conform to models rather than letting them reshape our understanding. **Measurement**, however, can serve as a form of **model auditing**: when observations systematically diverge from predictions, they expose the limits of our abstractions and guide the next refinement. The Internet offers a unique opportunity for such **closed-loop reasoning**: unlike most infrastructures, we can safely and continuously experiment with it (e.g., by injecting routing updates, running traceroutes, and observing immediate effects). In this setting, **measurement becomes an active component of optimization**: each probe updates our understanding of the network, and that evolving model guides the next measurement. Realizing this vision requires a new foundation that unifies measurement**,** inference**,** and control. **I aim to build that foundation by developing the algorithms and systems needed for self-directing measurement models that learn, adapt, and optimize as they observe the Internet.**

References

[1] **Loqman Salamatian**, Kevin Vermeulen, Dave Choffnes, Ethan Katz-Bassett, and Phillipa Gill. "HERMES: Repurposing User-Driven Speed Tests to Monitor the Internet." Submitted to the Proceedings of the USENIX Symposium on Networked Systems Design and Implementation (NSDI), 2026 (under review).

[2] **Loqman Salamatian**, Kevin Vermeulen, Italo Cunha, Vasilis Giotsas, and Ethan Katz-Bassett. "MetAScritic: Reframing AS-Level Topology Discovery as a Recommendation System." In Proceedings of the ACM Internet Measurement Conference (IMC), 2024.

[3] **Loqman Salamatian**, Todd Arnold, Vasilis Giotsas, Calvin Ardi, and Matt Calder. "What's in the Dataset? Unboxing the APNIC User Populations." In Proceedings of the ACM Internet Measurement Conference (IMC), 2024.

[4] **Loqman Salamatian**, Scott Anderson, Joshua Matthews, Paul Barford, Mark Crovella, and Walter Willinger. "Curvature-Based Analysis of Network Connectivity in Private Backbone Infrastructures." In Proceedings of the ACM SIGMETRICS, 2022 (selected as a Communications of the ACM Research Highlight, 2023).

[5] **Loqman Salamatian**. "The Internet as Sisyphus: Repeating Measurements, Missing Causes." In Proceedings of the ACM Workshop on Hot Topics in Networks (HotNets), 2025.

[6] **Loqman Salamatian**, Scott Anderson, Joshua Matthews, Paul Barford, Mark Crovella, and Walter Willinger. "A Manifold View of Connectivity in the Private Backbone Networks of Hyperscalers." Communications of the ACM, 2023.

[7] **Loqman Salamatian**, Frédérick Douzet, Kavé Salamatian, and Kévin Limonier. "The Geopolitics Behind the Routes Data Travel: A Case Study of Iran." Journal of Cybersecurity, 2019.

[8] **Loqman Salamatian**, Todd Arnold, Ítalo Cunha, Jiangchen Zhu, Yunfan Zhang, Ethan Katz-Bassett, and Matt Calder. "Who Squats IPv4 Addresses?" In ACM SIGCOMM Computer Communication Review (CCR), 2023. (Awarded Best of CCR)

[9] Stephen Jasina, **Loqman Salamatian**, Scott Anderson, Paul Barford, Mark Crovella, and Walter Willinger. "Matisse: Visualizing Measured Internet Latencies as Manifolds." Submitted to Passive and Active Measurement (PAM), 2026 (under review).

[10] Stephen Jasina, **Loqman Salamatian**, Paul Barford, Mark Crovella, and Walter Willinger. "A Breath of Fresh Air: Visualizing How Networks 'Breathe.'" In Proceedings of the ACM Workshop on Hot Topics in Networks (HotNets), 2025.

[11] Scott Anderson, **Loqman Salamatian**, Zachary Bischof, Alberto Dainotti, and Paul Barford. "iGDB: Connecting the Physical and Logical Layers of the Internet." In Proceedings of the ACM Internet Measurement Conference (IMC), 2022.

[12] Frédérick Douzet, Louis Pétiniaud, **Loqman Salamatian**, Kévin Limonier, and Kavé Salamatian. "Measuring the Fragmentation of the Internet: The Case of the Border Gateway Protocol (BGP) during the Ukrainian Crisis." In Proceedings of the IEEE International Conference on Cyber Conflict (CyCon), 2020.

[13] Kahlil Dozier, **Loqman Salamatian**, and Dan Rubenstein. "Modeling Average False Positive Rates of Recycling Bloom Filters." In Proceedings of the IEEE Conference on Computer Communications (INFOCOM), 2024.

[14] Kahlil Dozier, **Loqman Salamatian**, and Dan Rubenstein. "Analysis of False Negative Rates for Recycling Bloom Filters (Yes, They Happen!)." In Proceedings of the ACM SIGMETRICS, 2024.