Assessing the Internet in the SEE Region

Jim Cowie

Internet History Initiative

Fellow, Berkman Klein Center for Internet & Society at Harvard University

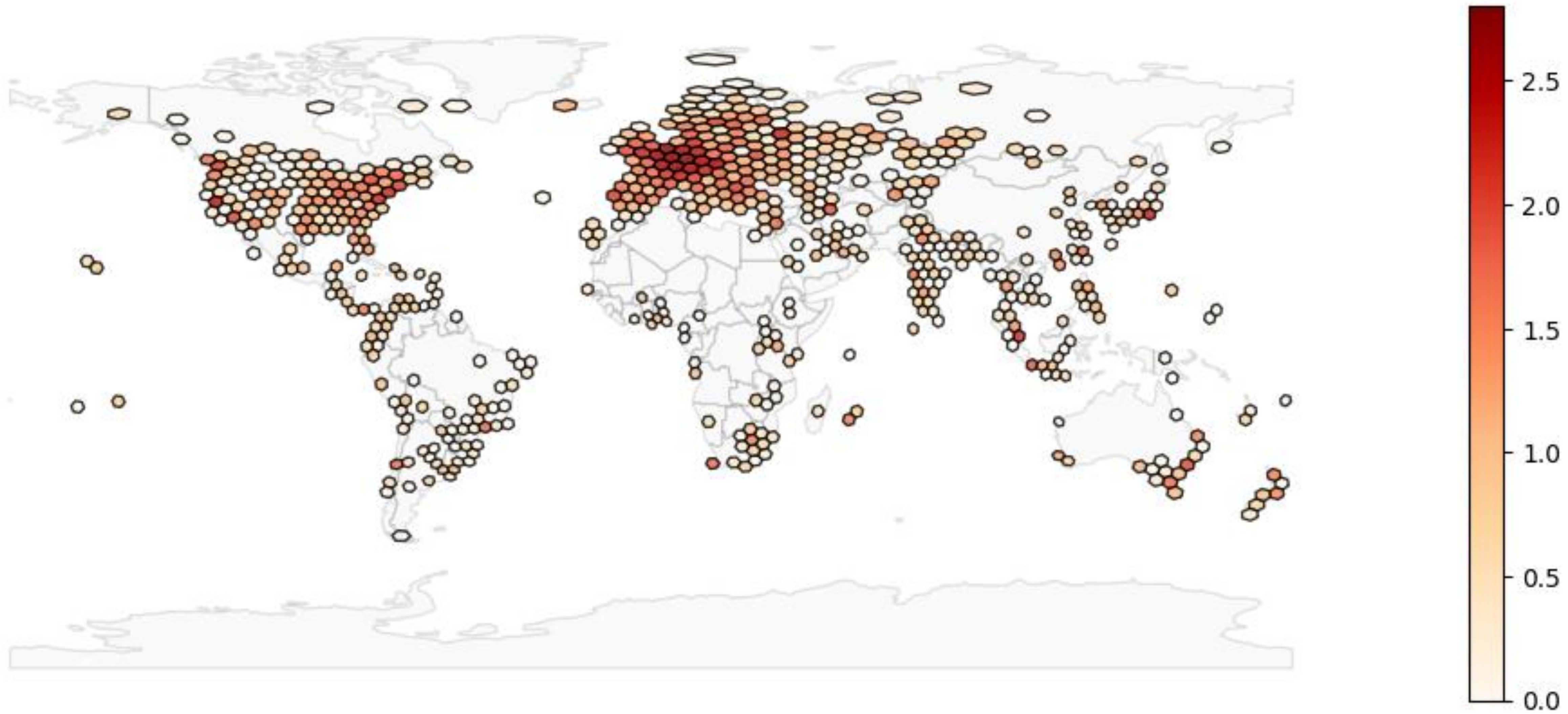
RIPE Southeast European Regional Meeting

Sofia, 7-8 April 2025

Internet History Initiative: Research Goals

- Collect and preserve the network operators' community legacy of Internet measurement datasets
- Extract time series data that reflect key aspects of regional Internet growth and diversification
- Study similarities and differences in Internet development across world regions
- Make these time series available to researchers studying different (potentially non-technical) aspects of international development

RIPE Atlas Global Probe Density (logscale)



Assessing the Regional Internet, Two Ways

Network perspective

- How many routers can we reach in *k* hops from our region?
- How many routers can we reach within *t* milliseconds?

Content perspective

- How do popular sites choose to serve our region?
- Where do large DNS resolvers serve our market from?

Part 1: Network Perspective

In a perfect world, we would always have a comprehensive assessment of the sites our customers are paying us to connect them to – perfectly anticipating their future needs.

We'd purchase enough transit and build enough peering relationships to satisfy our customers with low-latency, high-throughput service to the counterparties they want to talk to, all over the world – even when the Internet is under stress, or damaged, or parts of it are shut down.

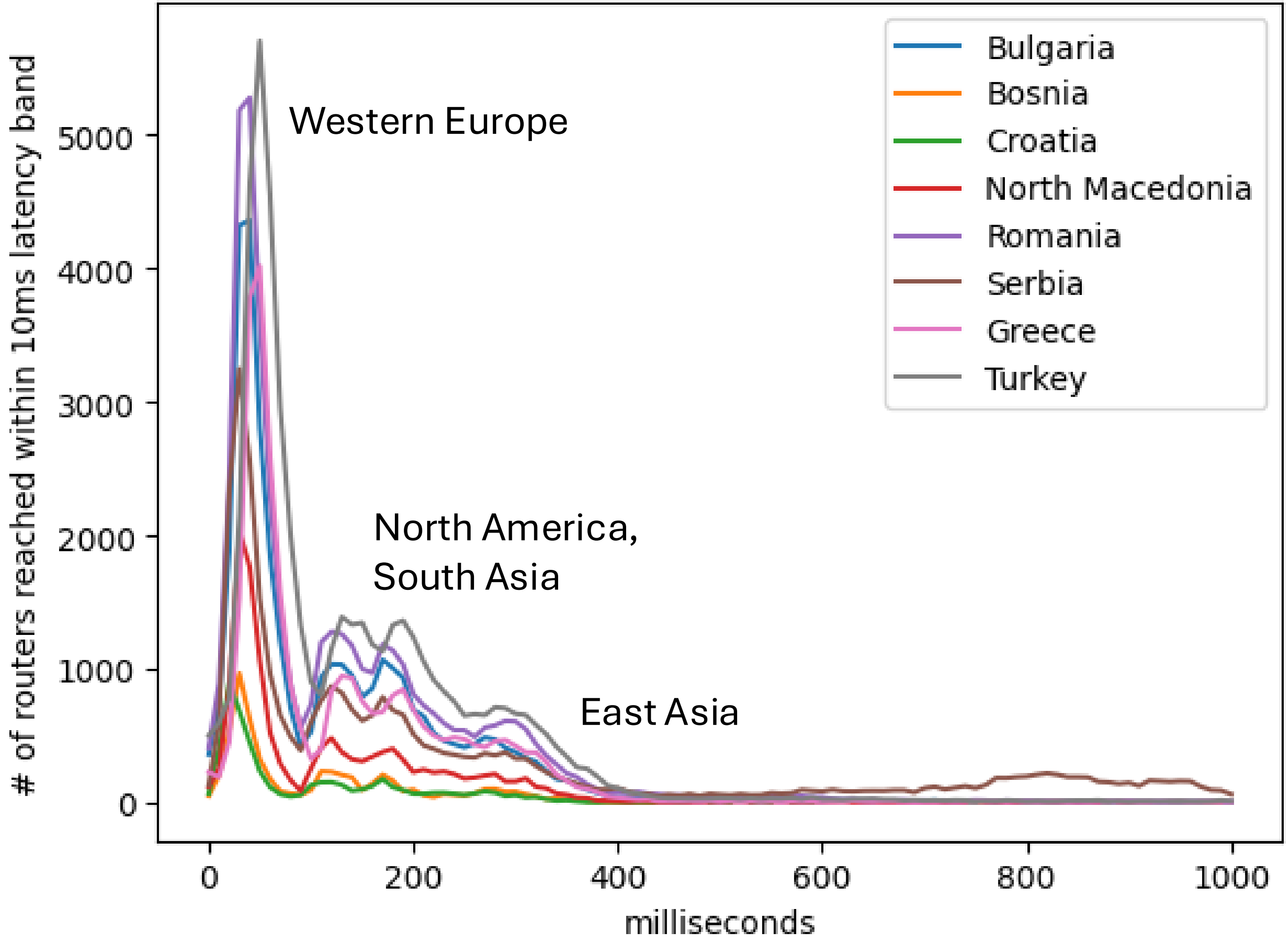
In reality, though…

We don't know these things with certainty. And we certainly can't predict how bad things might get in the future. But we can build a model and compare our region to others.

One simple model lets the set of **RIPE ATLAS Anchors** stand in for our customers' global counterparts. Anchors traceroute to each other continually, and their geolocation is reasonably well recorded.

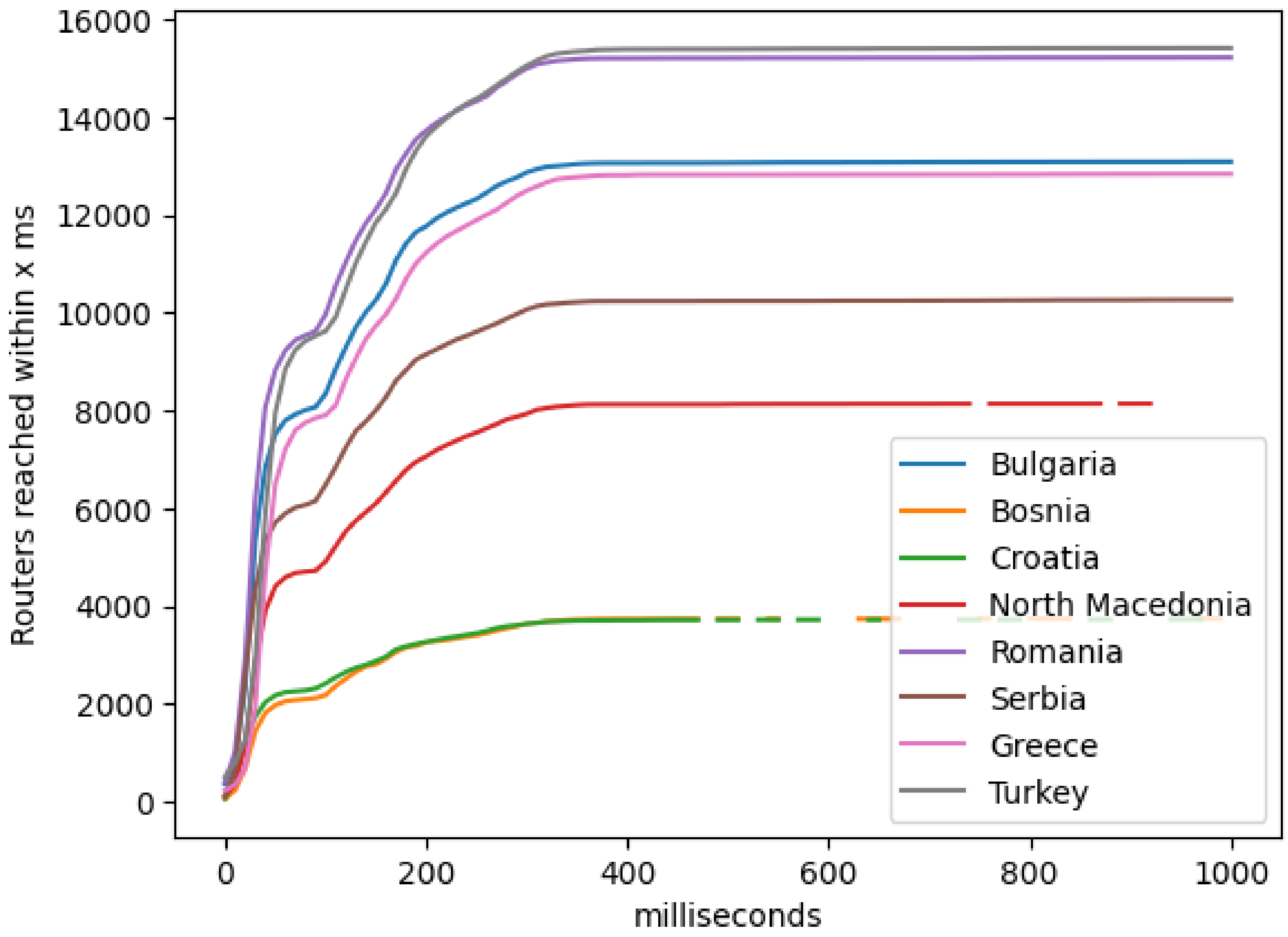
Let's examine traces from **anchors in our region**, to **all the other anchors.**

- How many routers can we reach within *t* milliseconds?
- How many routers can we reach in *k* hops from our region?



How many unique routers are seen in traceroutes within each 10ms latency band, moving out from anchors in the given country?

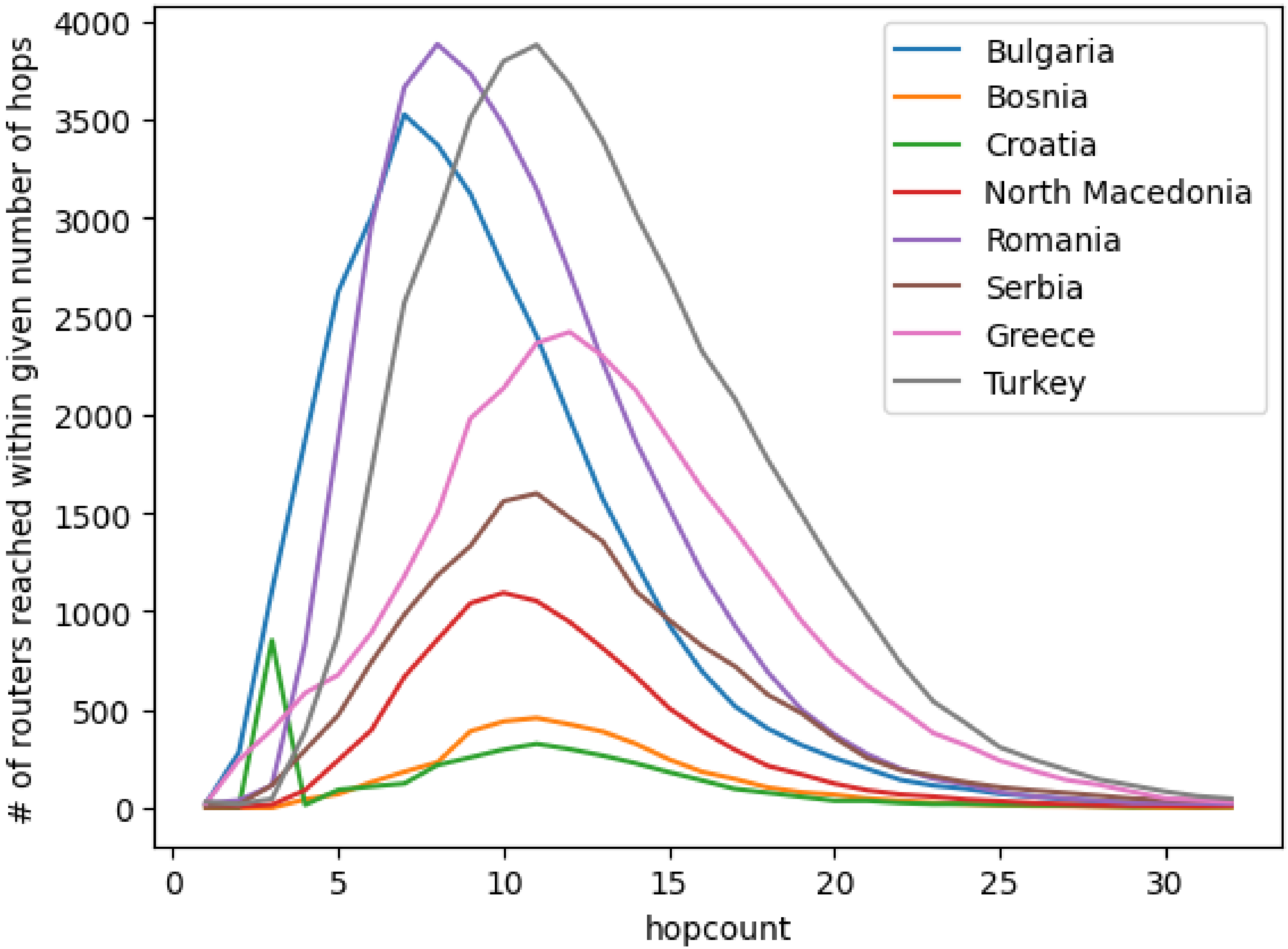
This reflects the geographic density within the anchor set (western EU bias).



How many routers are encountered within no more than **X ms**, moving out from anchors in the given country?

"Higher" for larger anchor sets with diverse routing

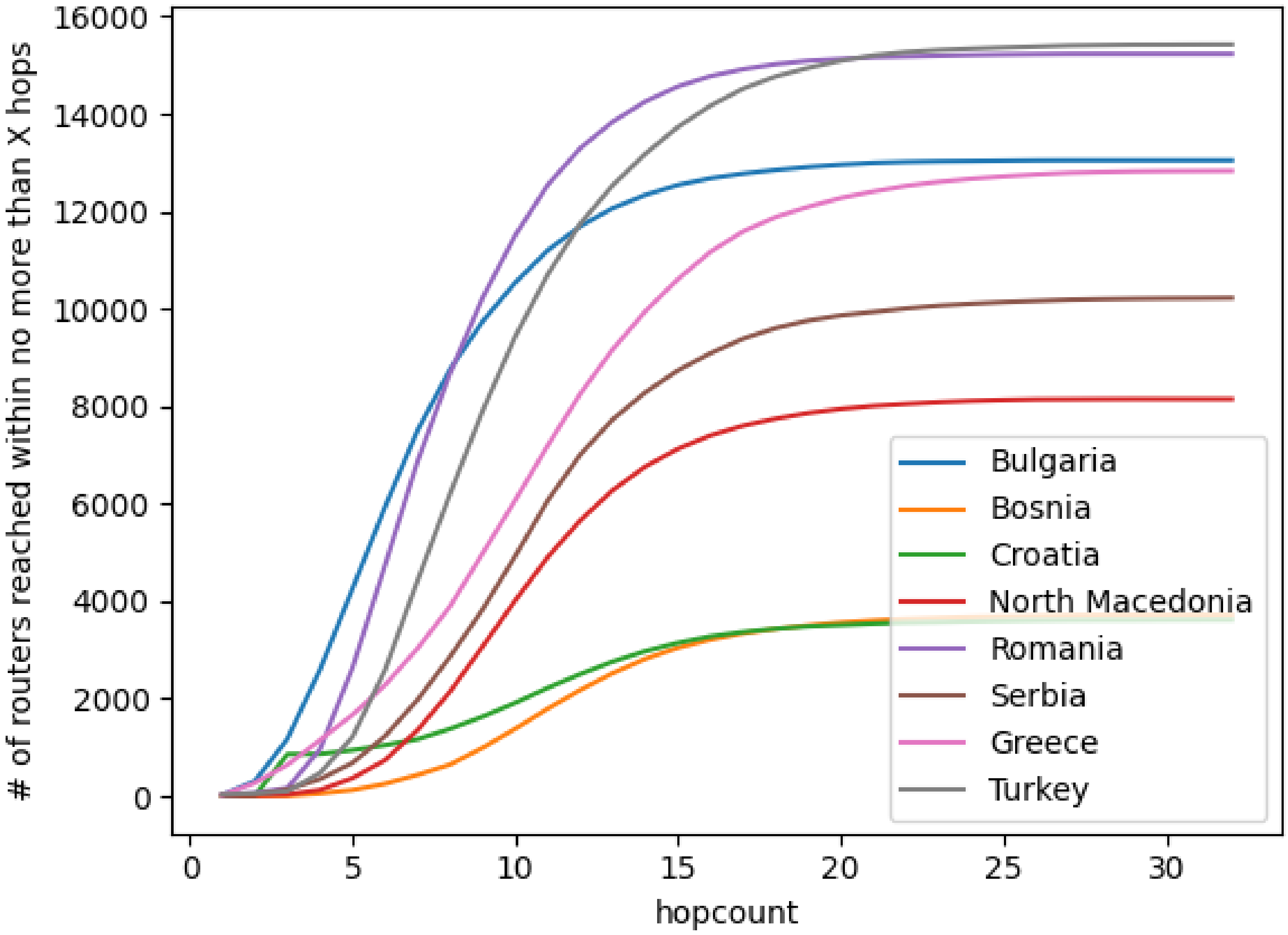
"Steeper" reaches more of the Internet faster (shorter paths)



How many routers are encountered in exactly **X hops**, moving out from anchors in the given country?

"Higher" for larger anchor sets with diverse routing

"Left-leaning" reaches more of the Internet faster (shorter paths, fewer routers traversed)

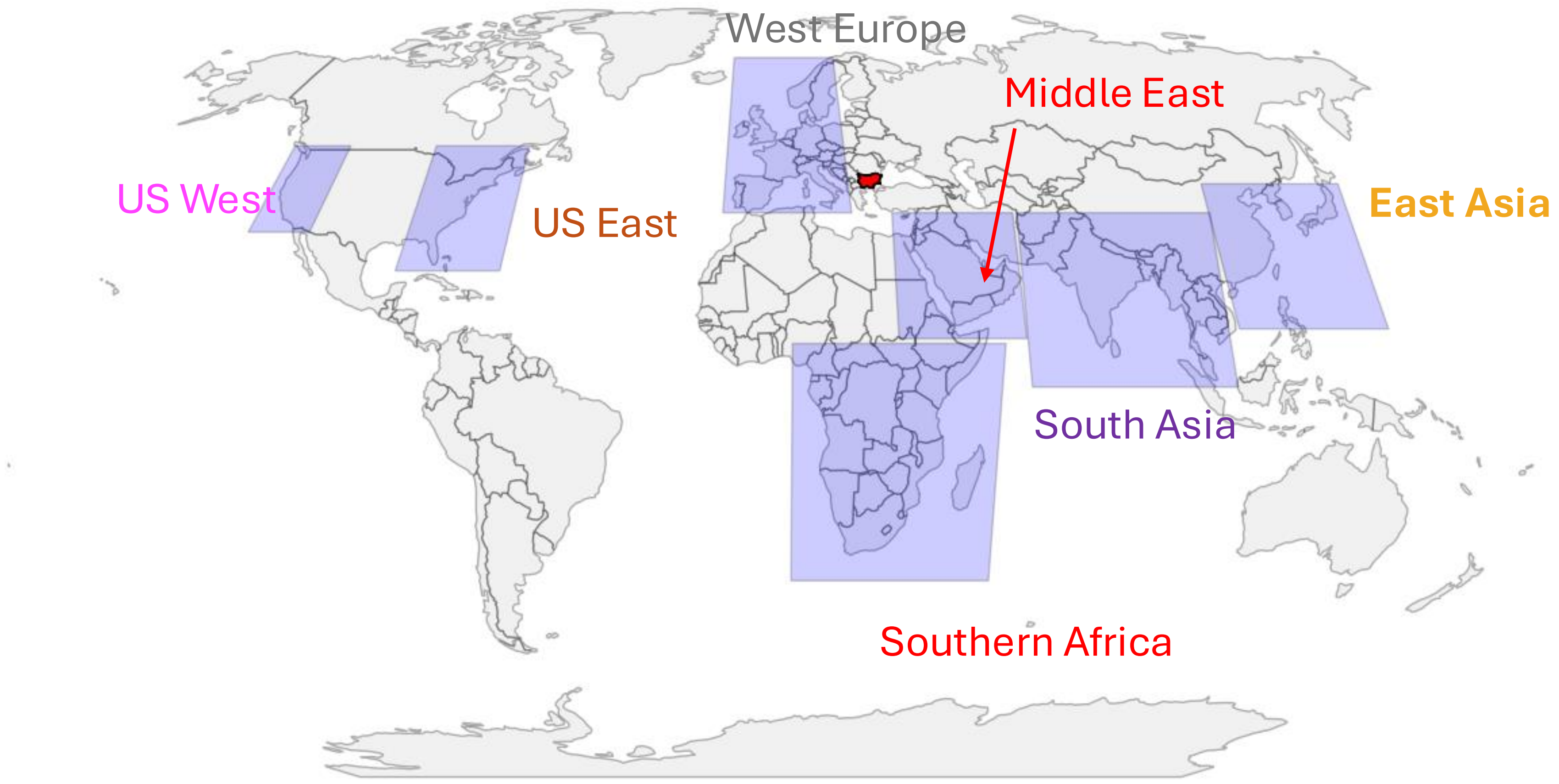


How many routers are encountered within **no more than X hops**, moving out from anchors in the given country?

"Higher" for larger anchor sets with diverse routing

"Steeper" reaches more of the Internet faster (shorter paths, fewer routers traversed)

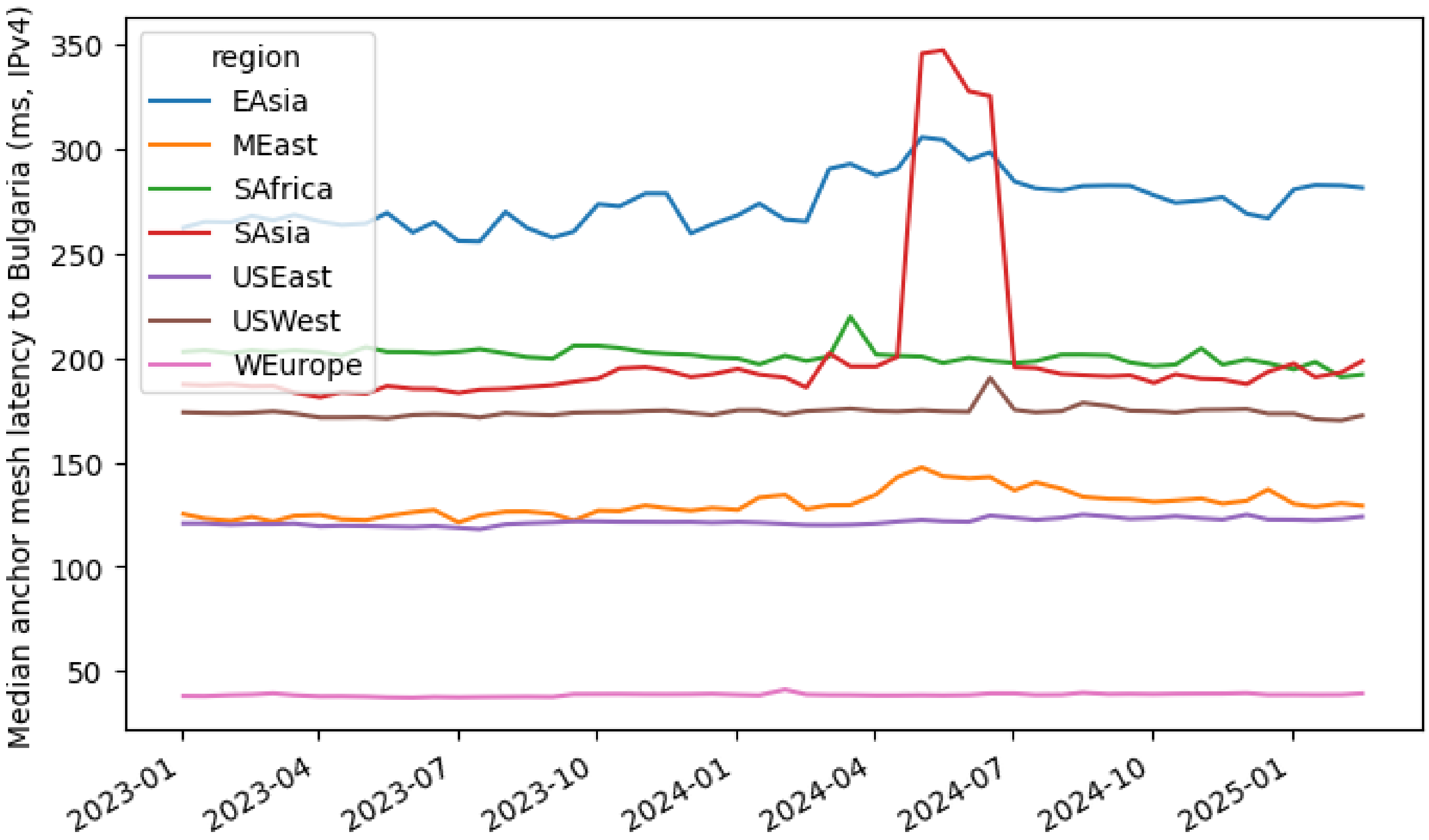
How else can we tell a story about regional trends?



Let's identify seven regions to study.

We'll examine median historical latencies between anchors in these regions, and the anchors hosted within Bulgaria.

Bulgarian IPv4 Latencies 2023-2025: Stable.



East Asia (the long way around, ~275ms)

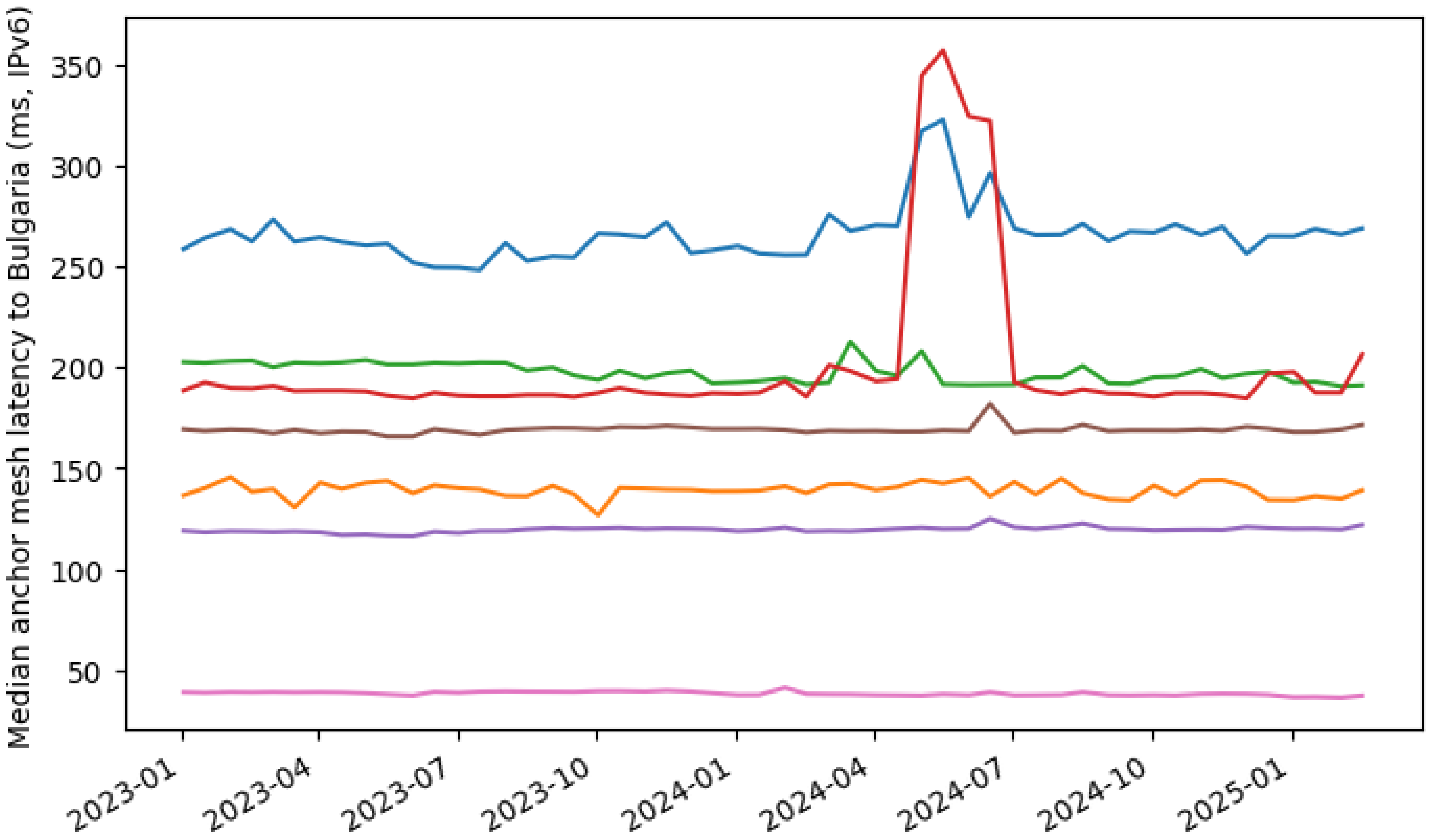
South Africa (~200)
South Asia (~200 or ~325 when SMW5 breaks)
US West (~175 via Atlantic and across USA)

Middle East (~125 via Europe and Med)
US East (~125 via Atlantic)

West Europe (sub-50ms direct routes)

RIPE Atlas latency measurements to Bulgarian anchors, IPv4 RTT, daily median, 1st and 15th of the month, Jan 2023-Feb 2025

Bulgarian IPv6 Latencies 2023-2025: Stable.



East Asia (the long way around, ~275ms)

South Africa (~200)
South Asia (~200 or ~350 when SMW5 breaks)
US West (~160 via Atlantic and across USA)

Middle East (~150 via Europe and Med)
US East (~120 via Atlantic)

West Europe (sub-50ms direct routes)

RIPE Atlas latency measurements to Bulgarian anchors, IPv6 RTT, daily median, 1st and 15th of each month, Jan 2023-Feb 2025

Part 2: Content Perspective

Using the entire ATLAS Anchor set as a model for what we care about can only take us so far.

Users care a lot about fast access to specific, highly distributed content.

- How do popular sites choose to serve our region?
- Where do large DNS resolvers serve our market from?

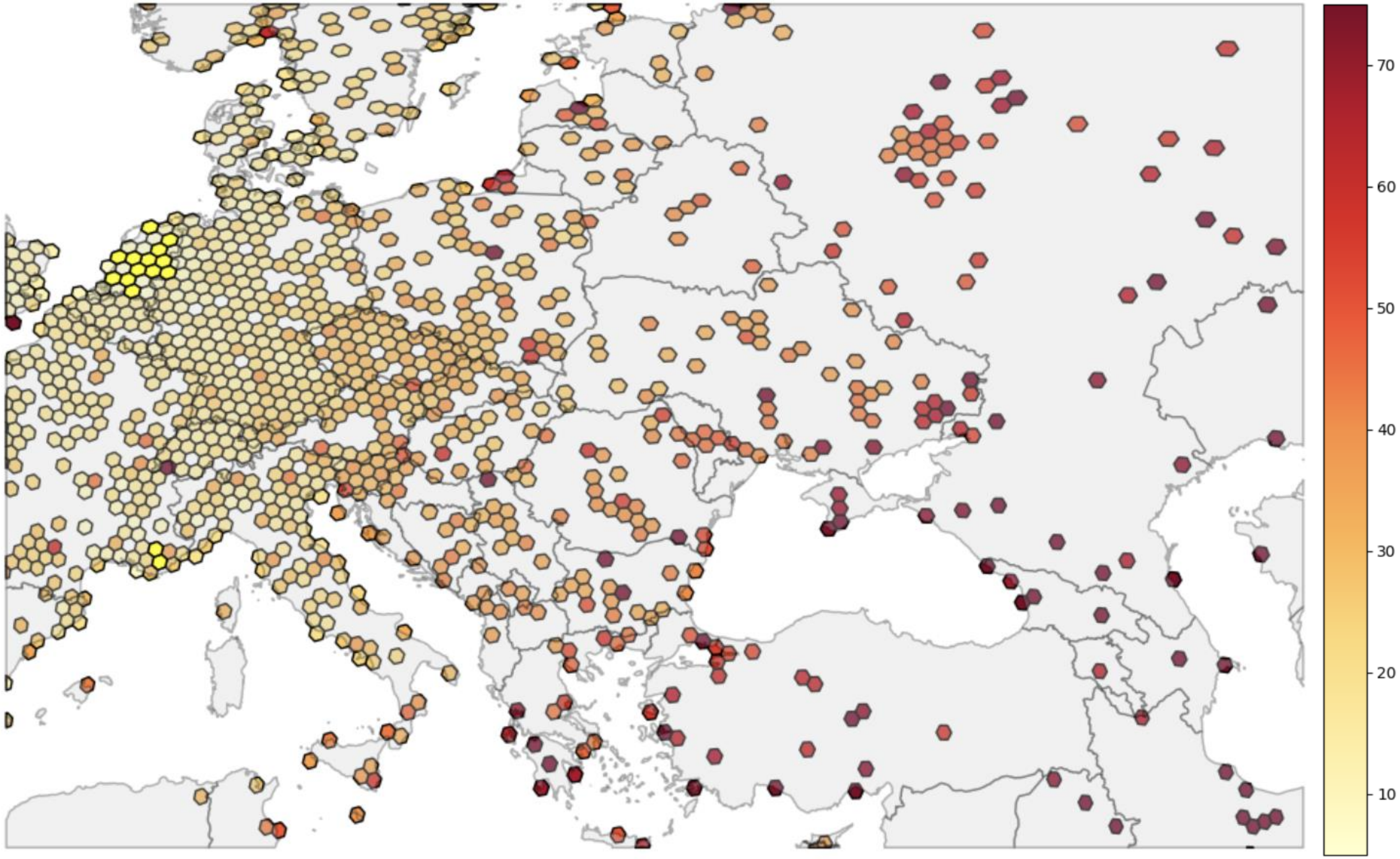
Popular Site Latencies

- The Atlas probes perform periodic pings to various sites
- Google, Facebook, Wikipedia .. as resolved on-probe
- Bonus: measurements in both IPv4 and IPv6 when available!

These days, this tells us a lot about the centralization or edge distribution of popular sites.

Let’s look at latencies and host mappings seen 1 April 2025!

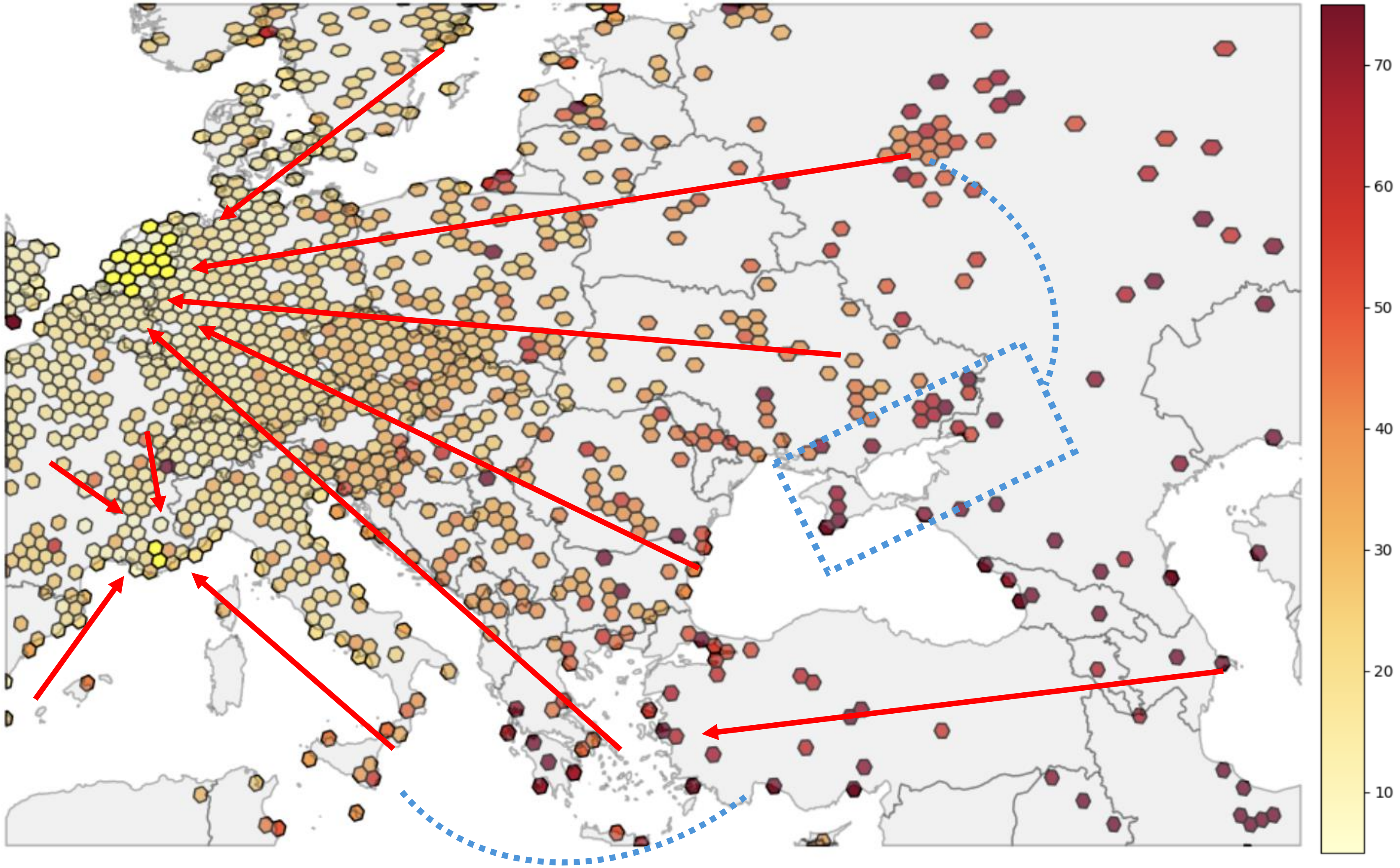
Wikipedia: Classic Centralized Hosting



Yellow hexes: median ping under 5ms to Wikipedia (IPv4)

Darkest red: median ping over 75ms

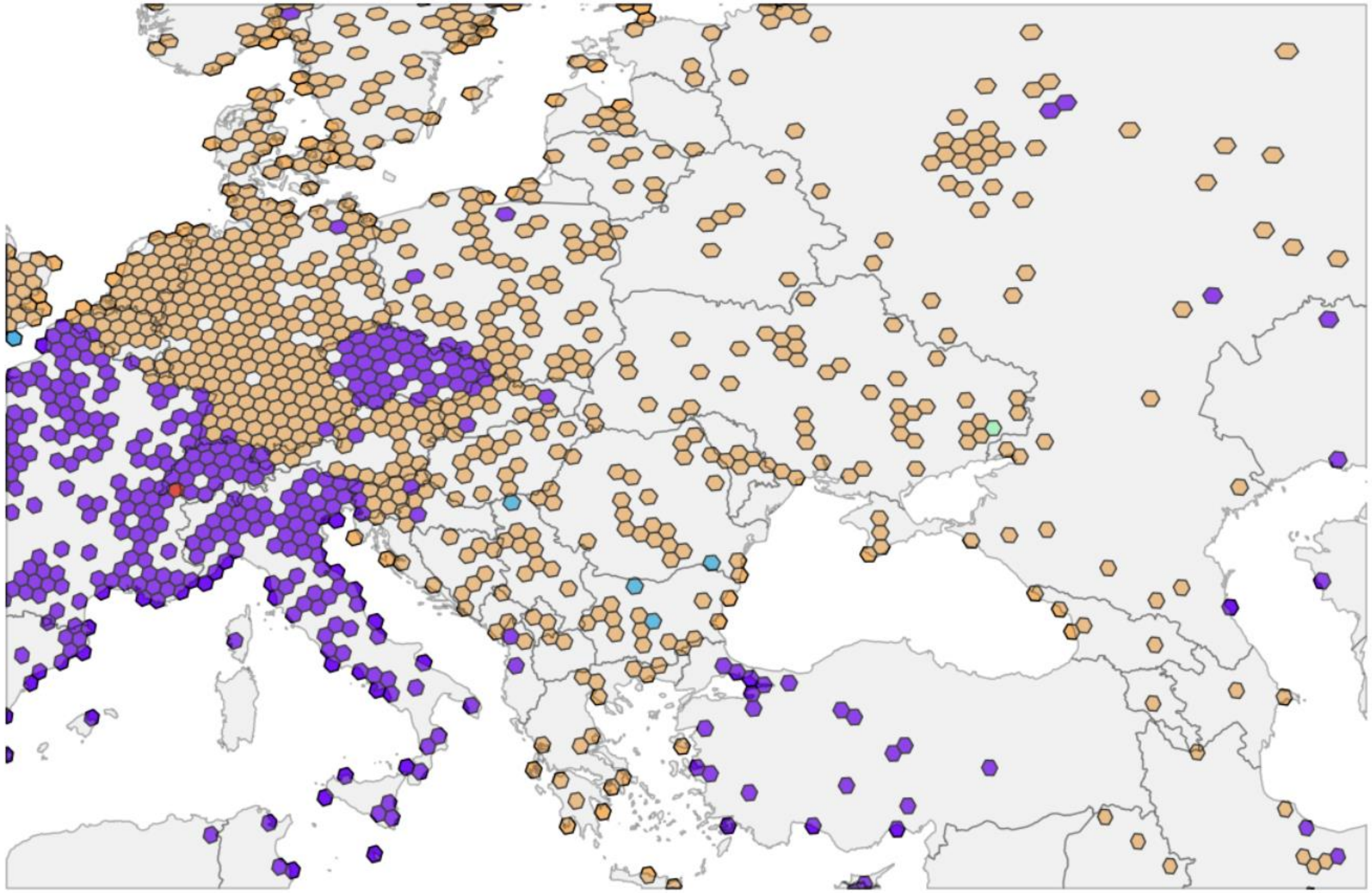
Wikipedia: Classic Centralized Hosting



Speed of light color gradient based on fiber route miles from Amsterdam or Marseilles

We see some higher-latency detours; e.g., occupied Ukraine, via Russia

'Resolve on probe' reveals Wikiwatersheds



Wikipedia.org **IPv4**

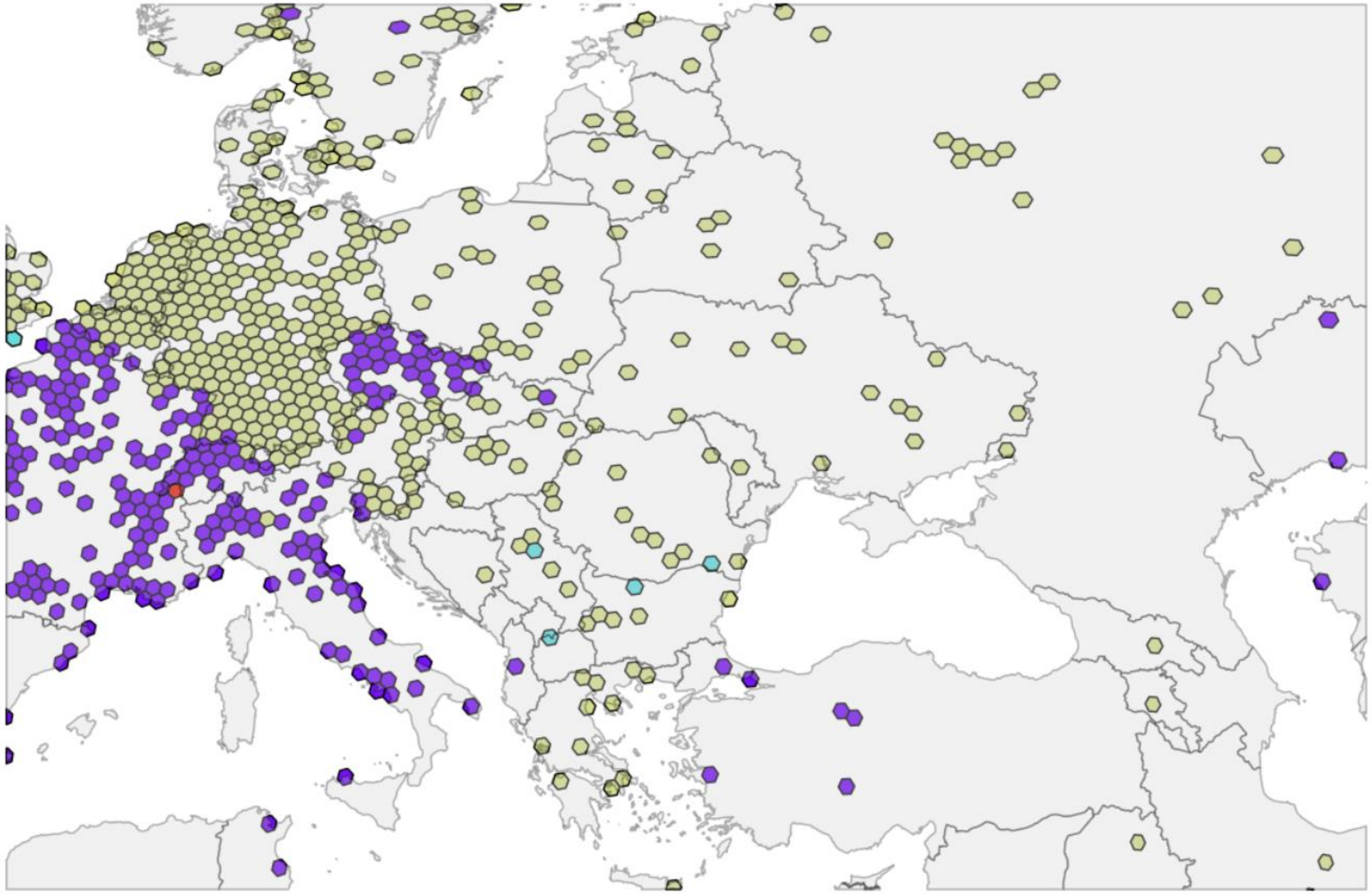
Approximate country-level load balancing:

Purple to Marseilles

Gold to Amsterdam

Blue to Virginia (VPN?)

'Resolve on probe' reveals Wikiwatersheds



Wikipedia.org **IPv6**

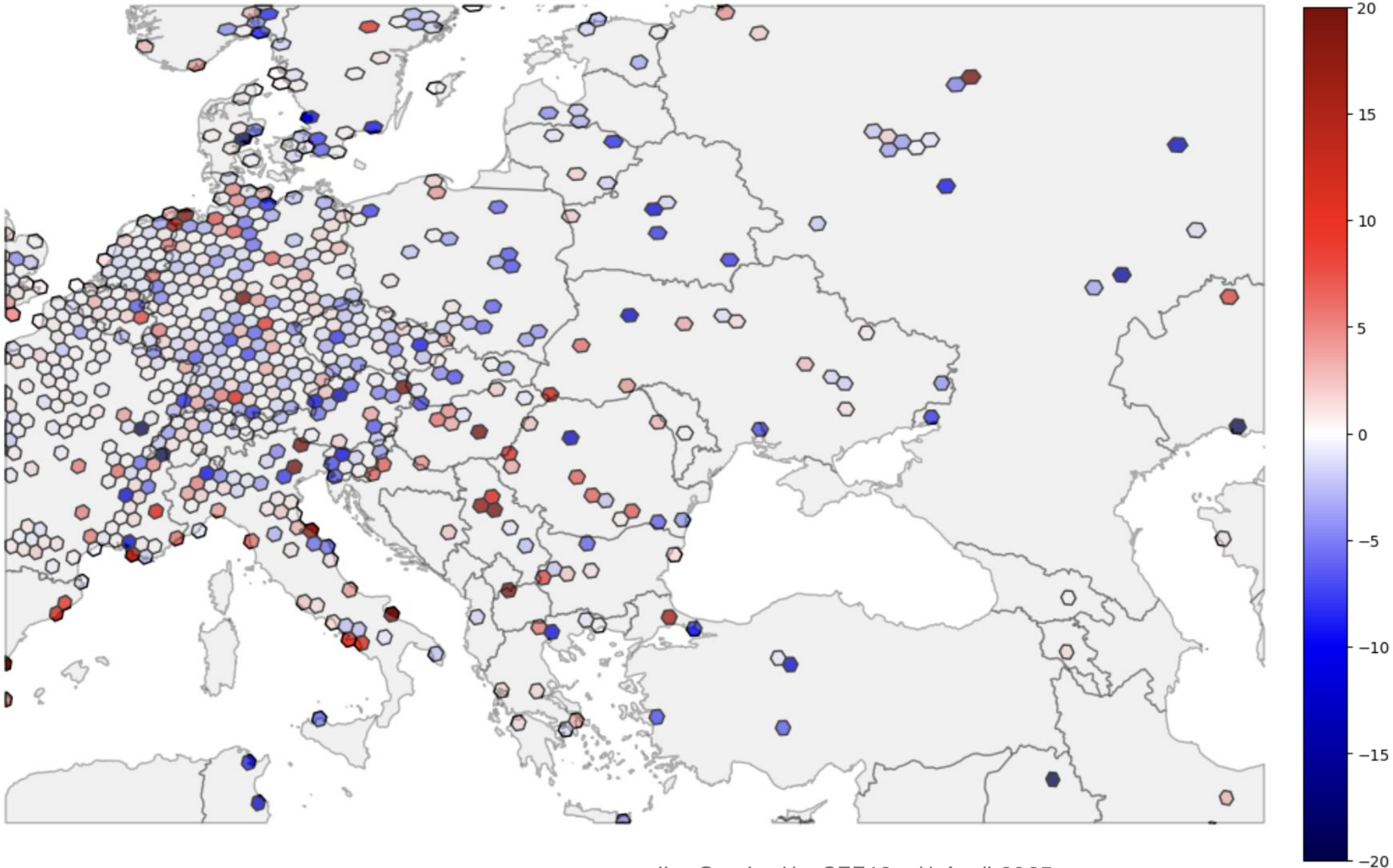
Same approximate country-level load balancing:

Purple to Marseilles

Green to Amsterdam

Blue to Virginia (VPN?)

What’s faster, IPv4 or IPv6? It’s tricky.

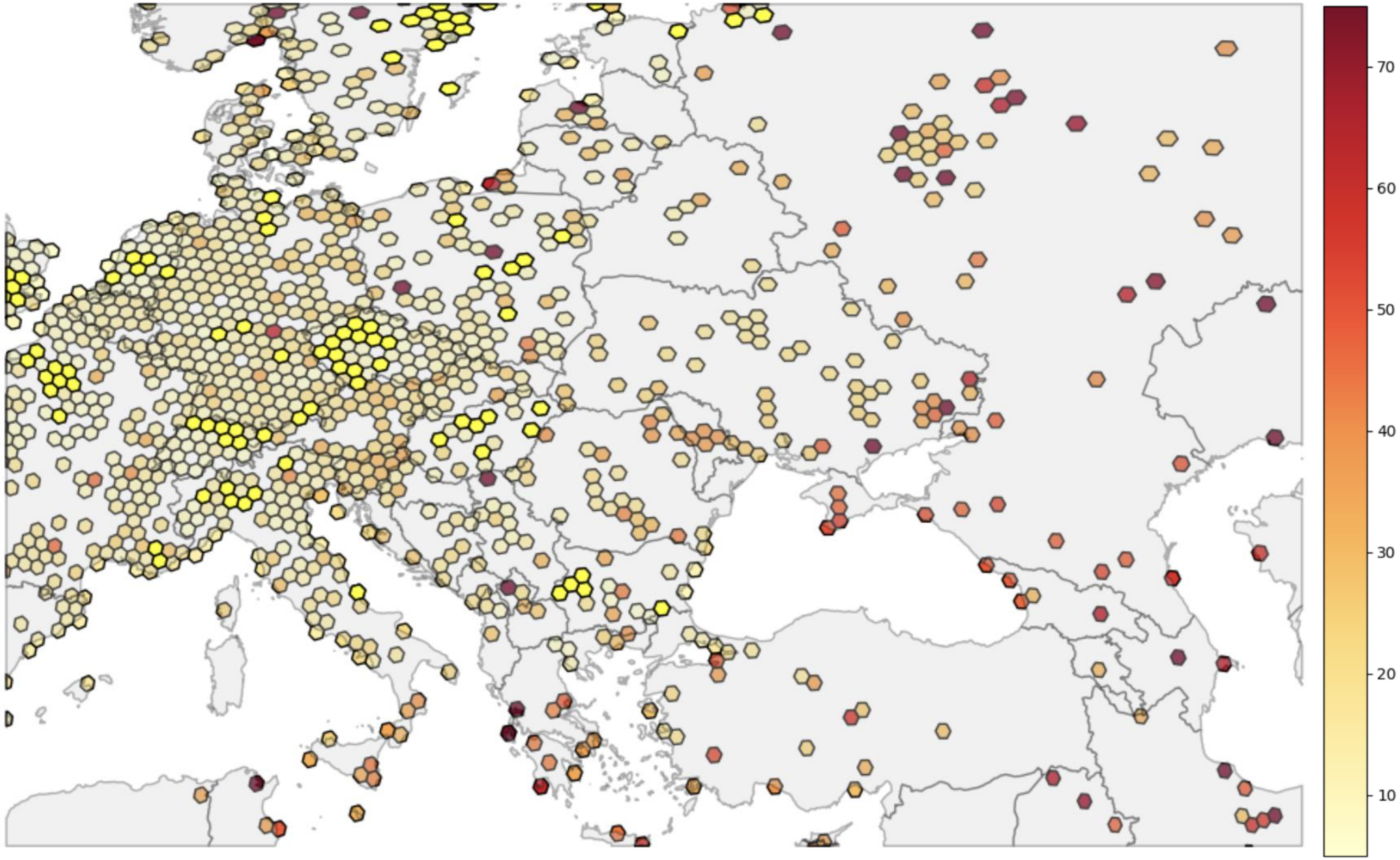


Red: IPv4 is faster ping to Wikipedia.

Blue: IPv6 is faster ping to Wikipedia.

Pretty random, no strong geographic pattern!

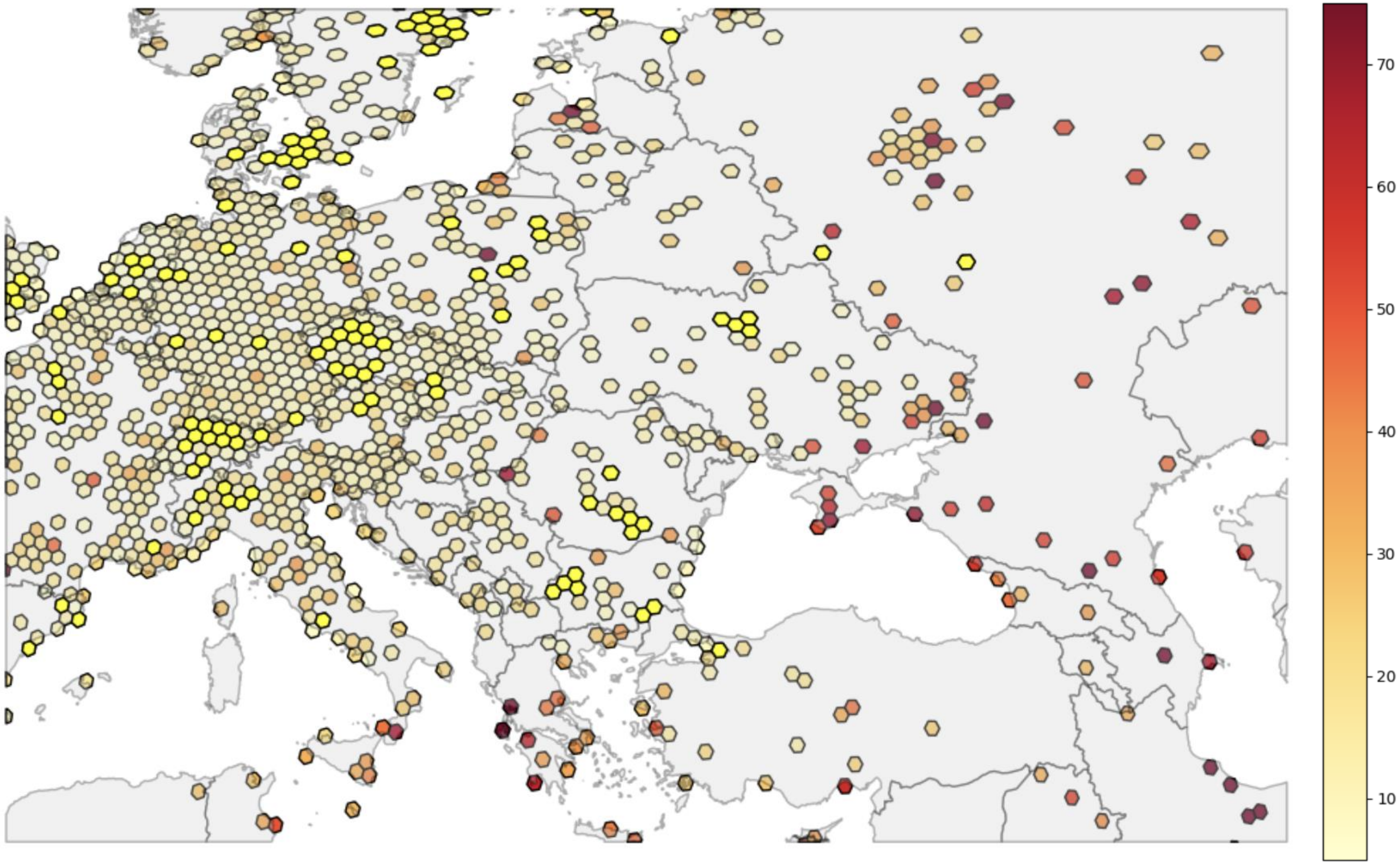
Google: Moderately Distributed Hosting



Yellow hexes: median ping under 5ms to Google (IPv4)

Few such in SEE (Bulgaria the exception)

Facebook: Highly Distributed Hosting



Yellow hexes: median ping under 5ms to Facebook (IPv4)

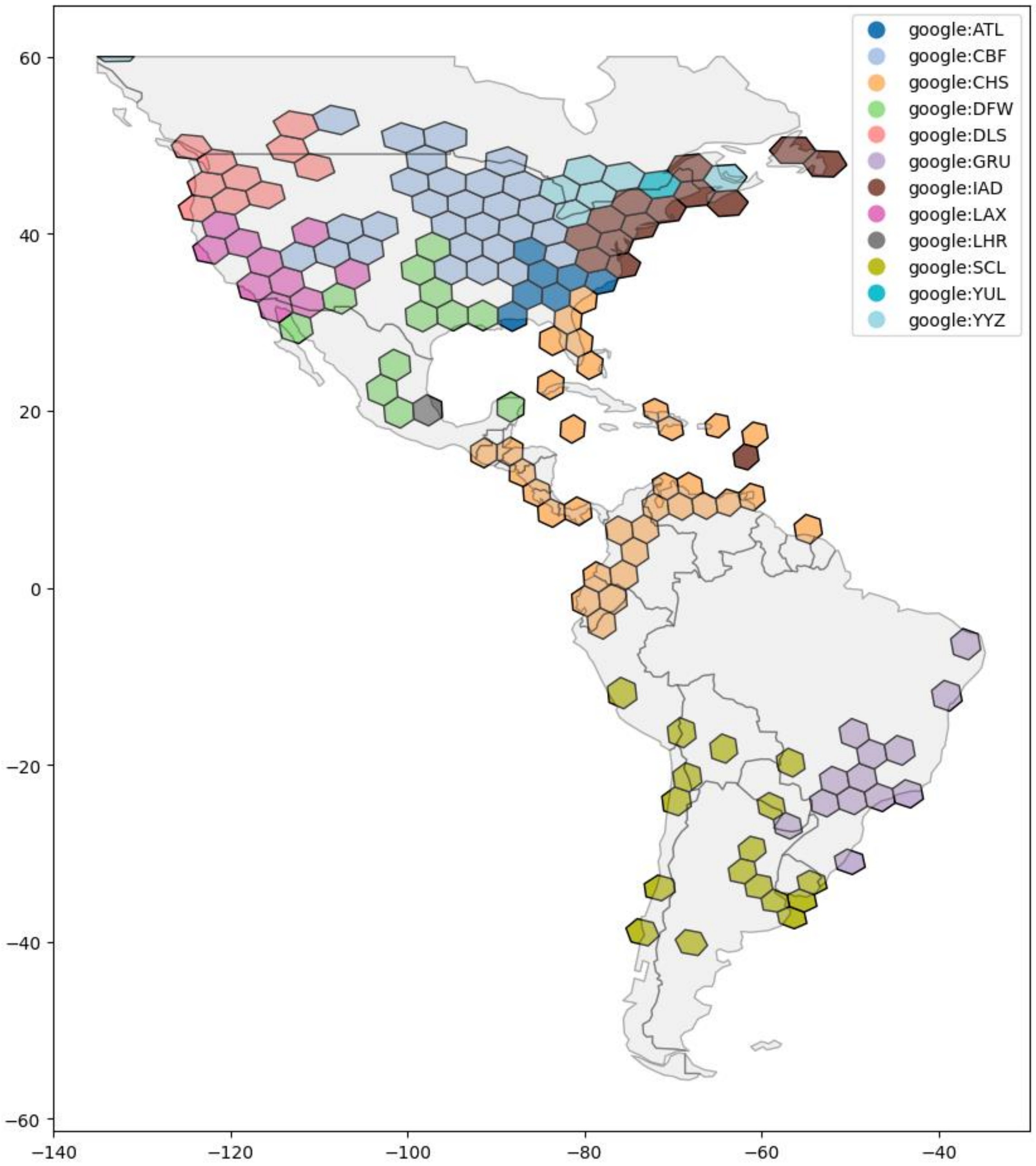
Kiev, Sofia, Bucharest, Istanbul, ...

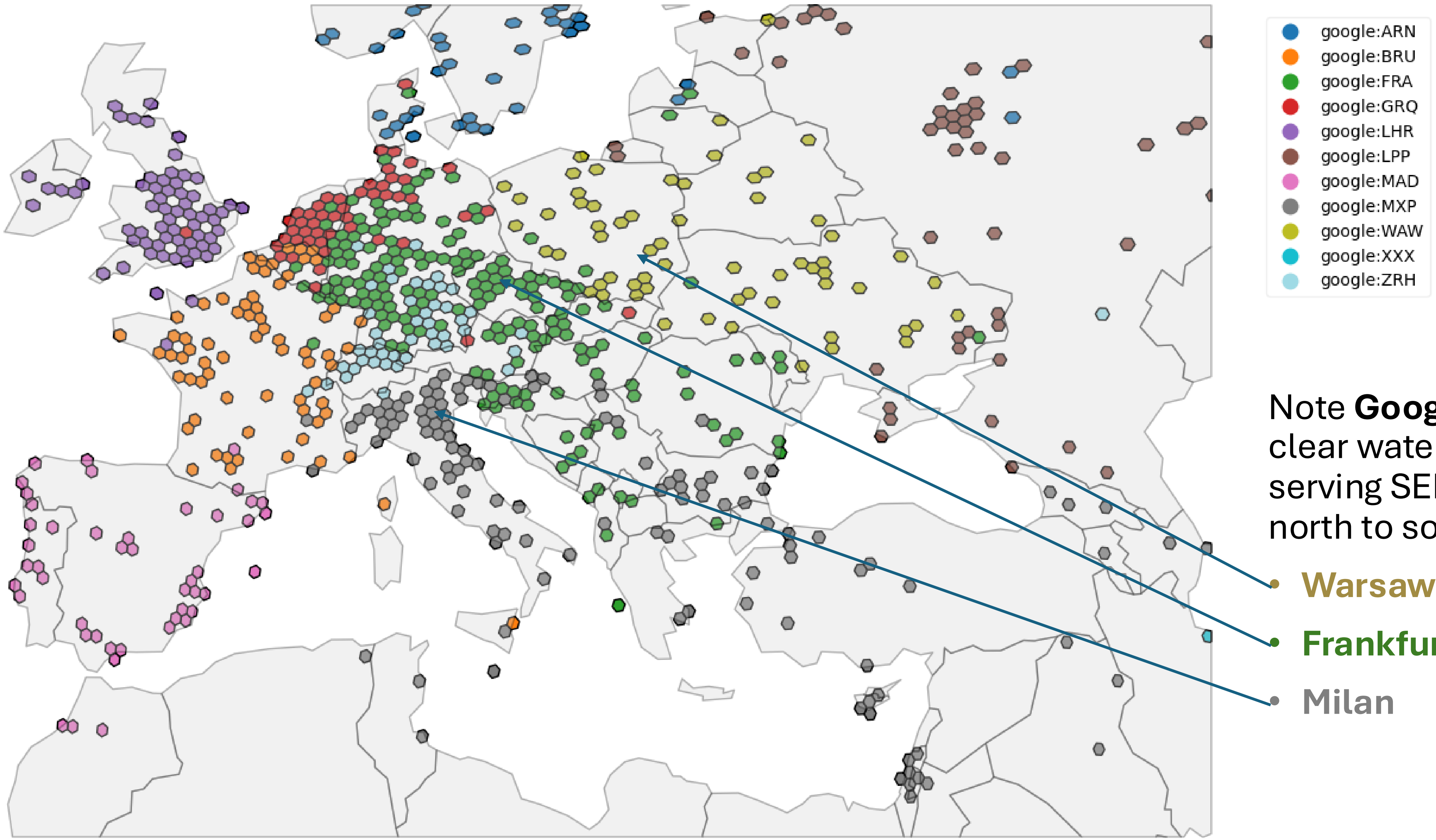
DNS Recursive Resolver Selection

- Two more long-running daily ATLAS experiments allow us to see what recursive resolver makes queries to authoritative resolvers on behalf of an ATLAS probe
- This IP address can be classified as local (often same ASN) or anycast global (e.g., Google 8.8.8.8, Cloudflare 1.1.1.1, Quad9 9.9.9.9)
- Using our knowledge of these DNS services' unicast footprint, we can further determine which specific datacenter hosts the unicast address of the ultimate recursive resolver
- This may be different from the local anycast instance

Example: Google creates ‘watersheds’ for 8.8.8.8 service

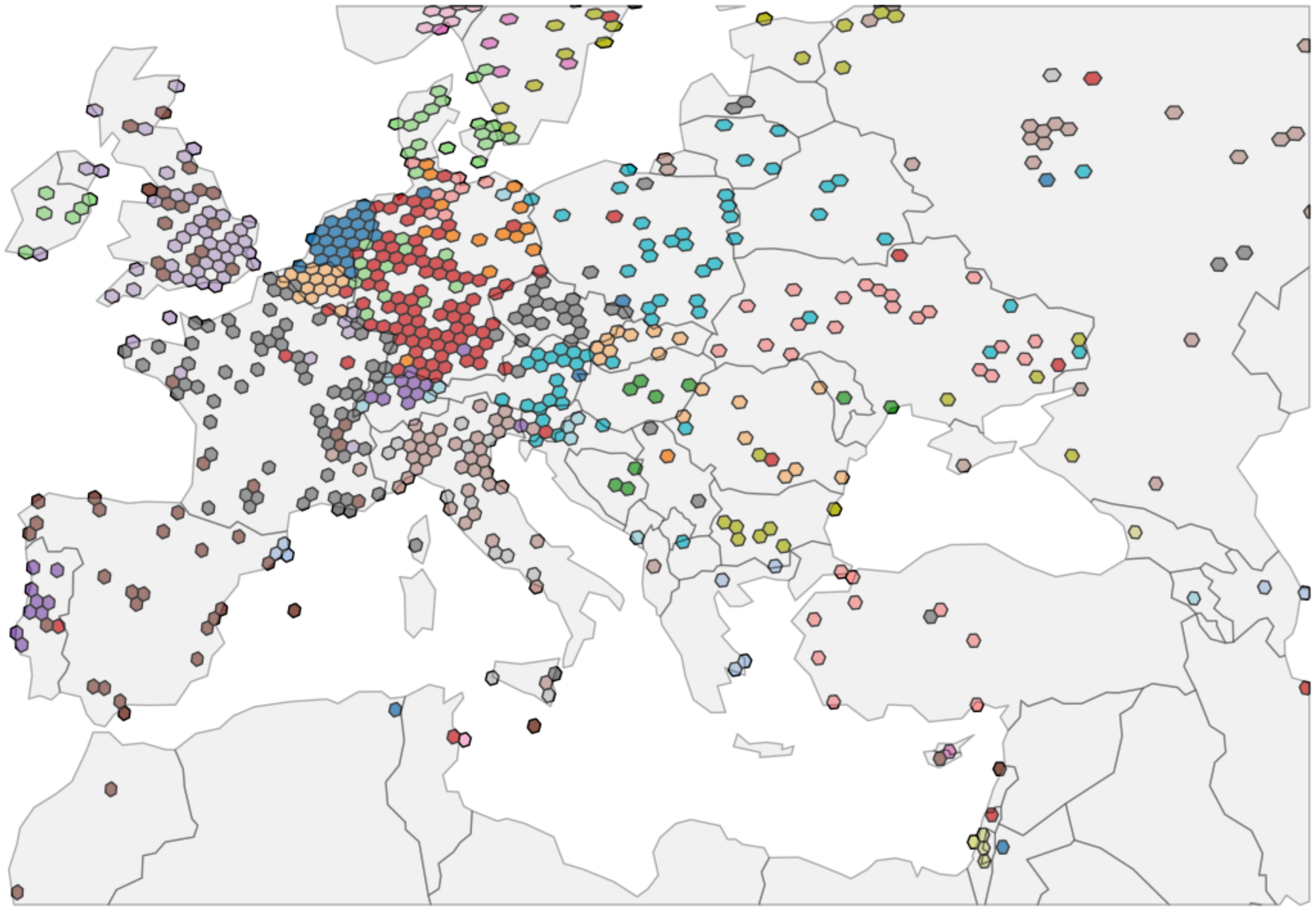
- Each hexagon is colored according to the most common Google datacenter hosting the ultimate unicast resolver address that queries authoritative servers when Atlas probes in that hex make a DNS query
- Most clients here are within 30ms of the ultimate resolver



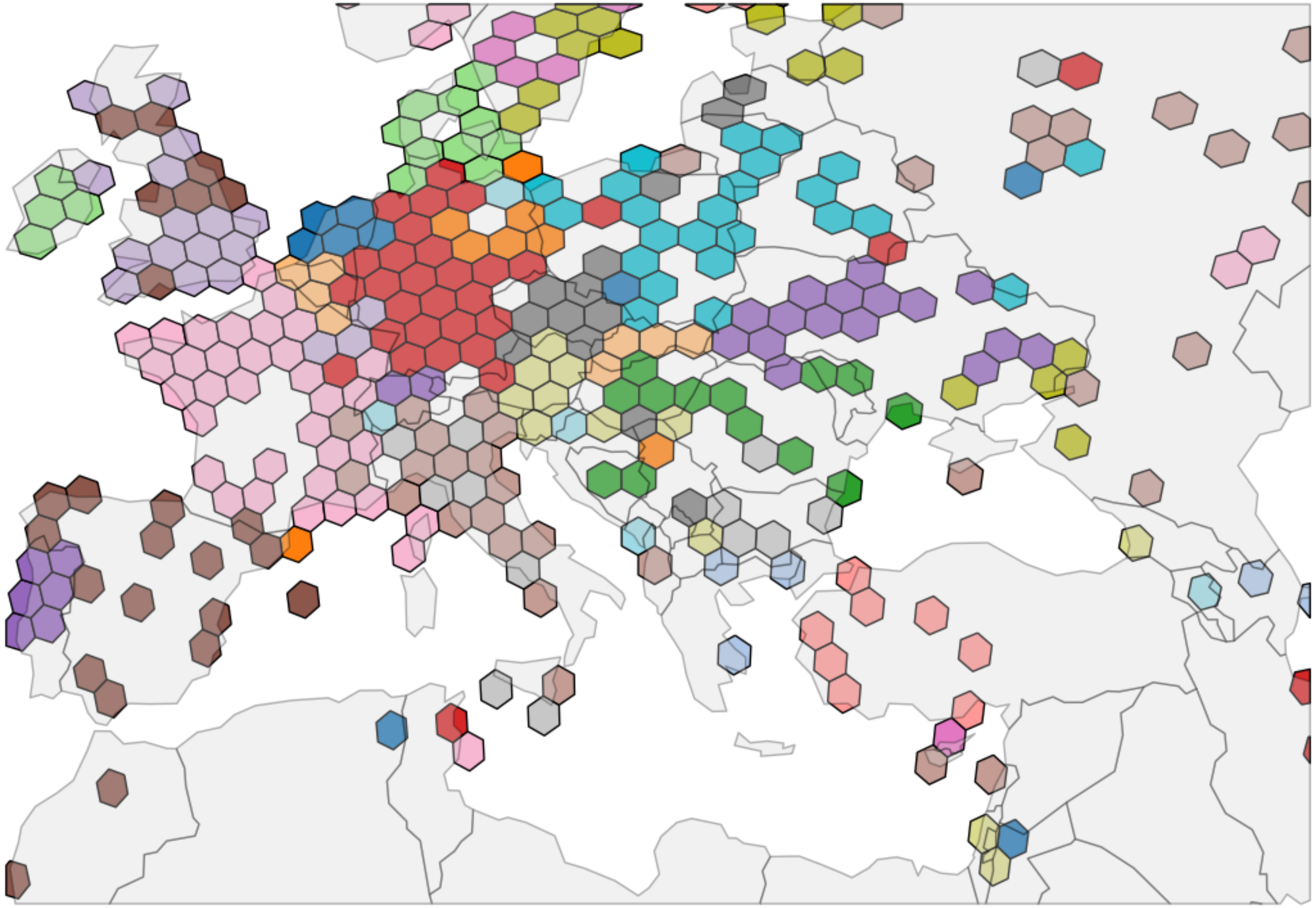


Note **Google's** clear watersheds serving SEE, from north to south:

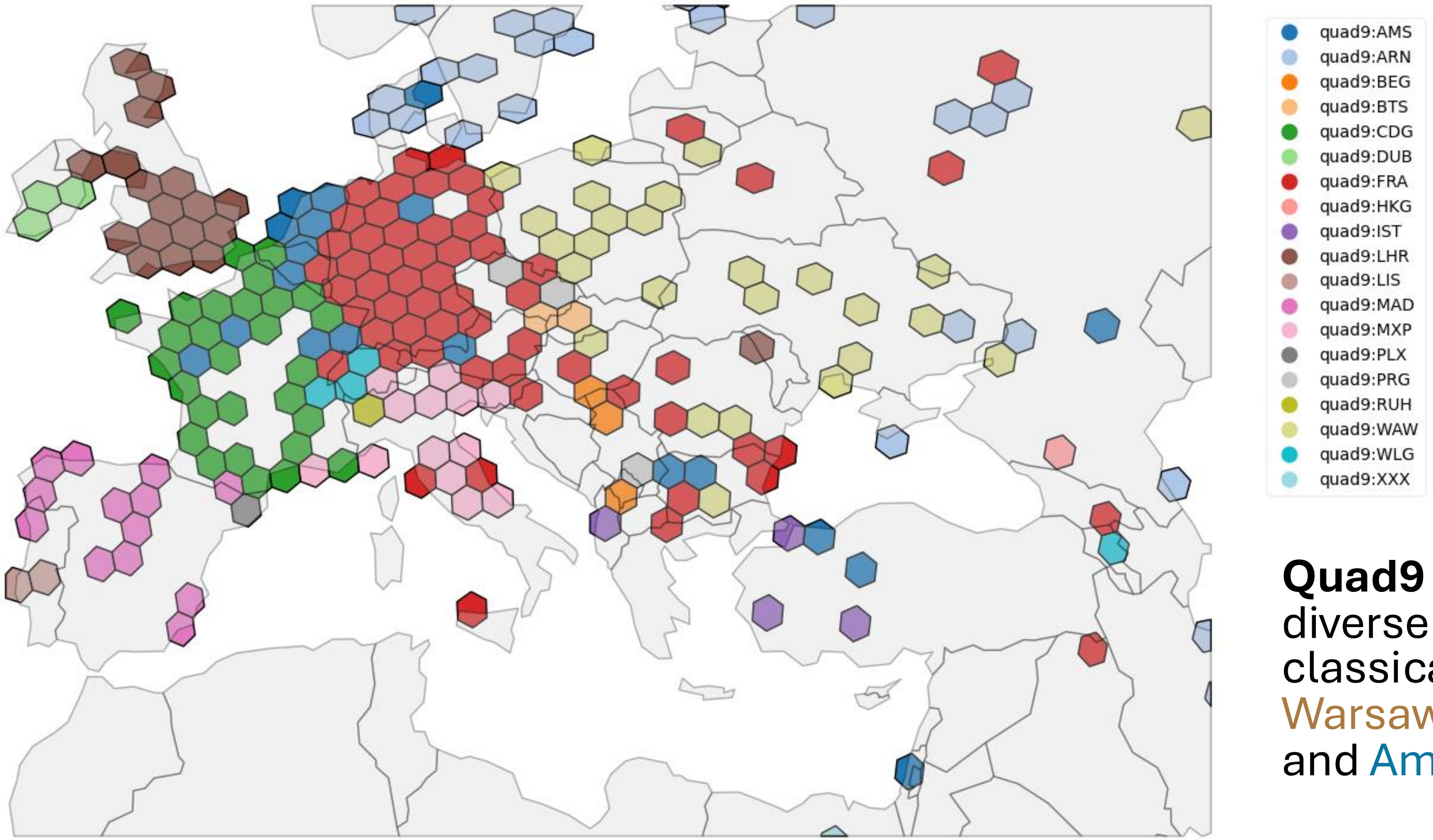
- **Warsaw**
- **Frankfurt**
- **Milan**



Cloudflare has much finer-grained watersheds, including local service in Sofia, Istanbul, Zagreb,Bucharest, Bratislava, Chisinau, ...



Coarsening the hexgrid makes it easier to spot the patterns in **Cloudflare's** deployment.



Quad9 serves SEE from diverse locations: classically Frankfurt and Warsaw, but also Istanbul and Amsterdam.

Conclusions

- RIPE Atlas is a rich source of periodic observations that help us understand how our region is connected, and how large content providers choose to serve our region.
- The history of these measurements will help us tell the story of how Internet in this region evolved.
- Best of all: these datasets are free and open to interpretation!
- I'm always glad to talk to students and other researchers who have ideas for potential data studies.

Thanks!

https://internethistoryinitiative.org

Mastodon: @IHI@cooperate.social

jacowie@cyber.harvard.edu