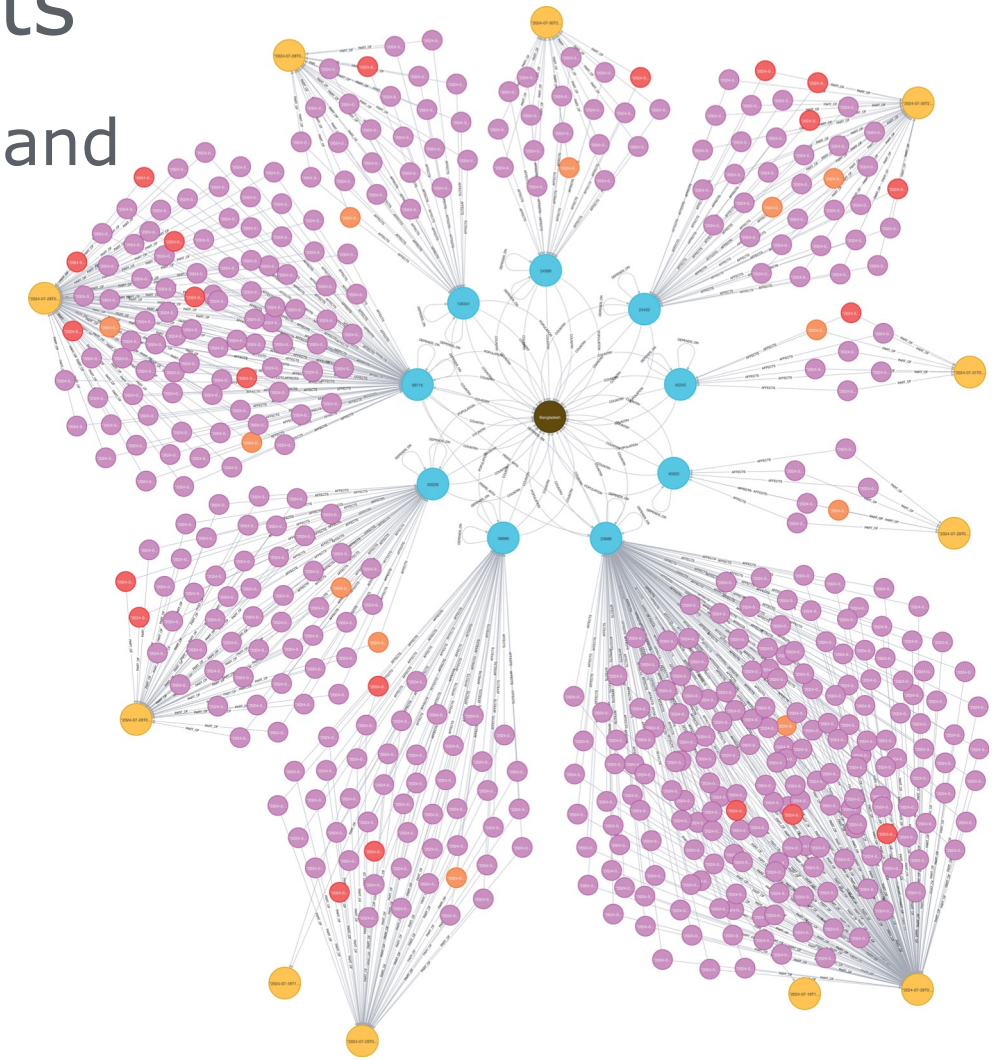
Hasso Plattner Institut
Digital Engineering · Universität Potsdam

Stitching the Internet's Blindspots

Unifying Outage Datasets for Visualization and Cross-Validation

Florian Künzig, Katrin Vollhardt, Maximilian Paß,
Sofya Generalova, Christian Kaiser, Simon Nowack,
Vasilis Ververis, and Vaibhav Bajpai

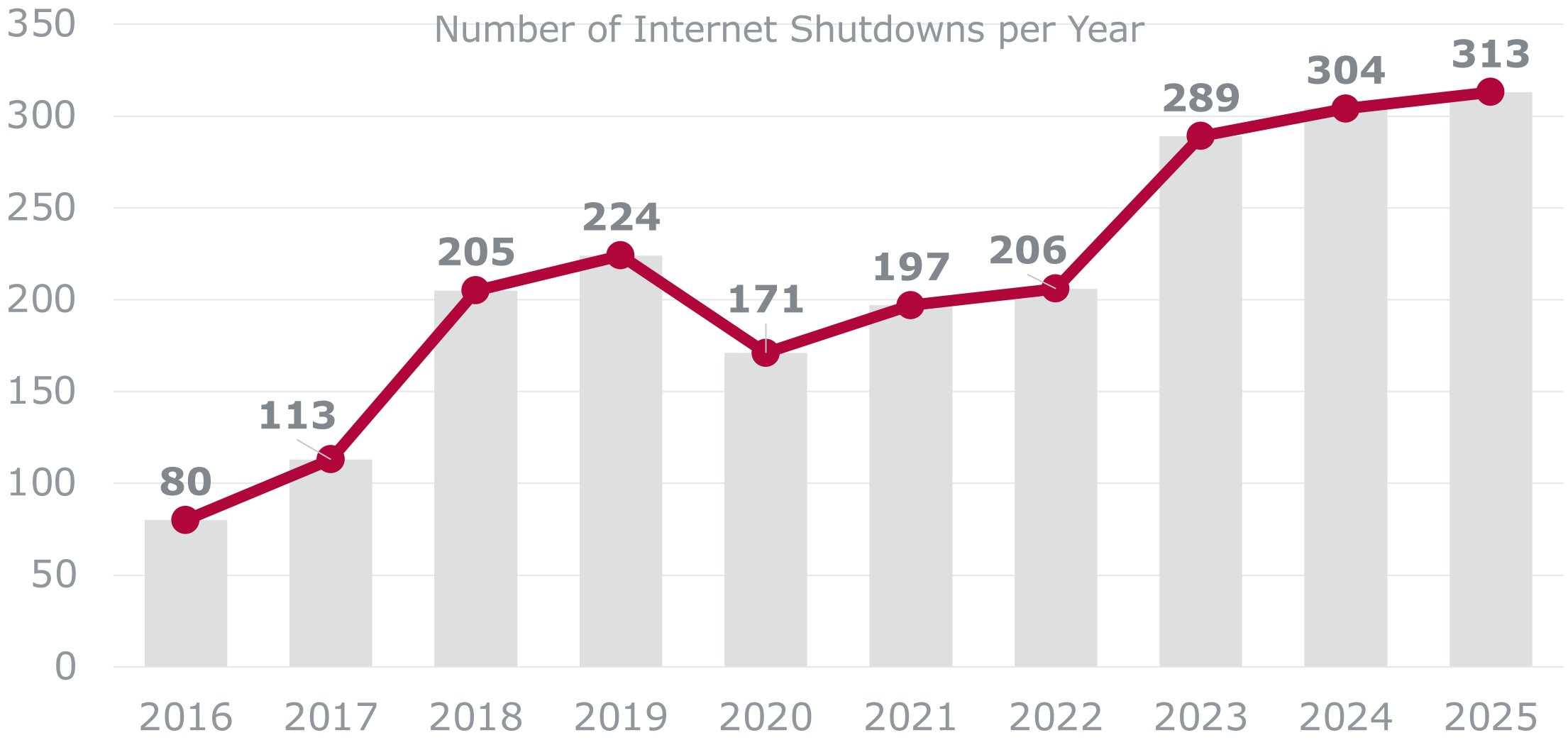
TMA Conference 2026, Delft, Netherlands



Design IT.
Create Knowledge.

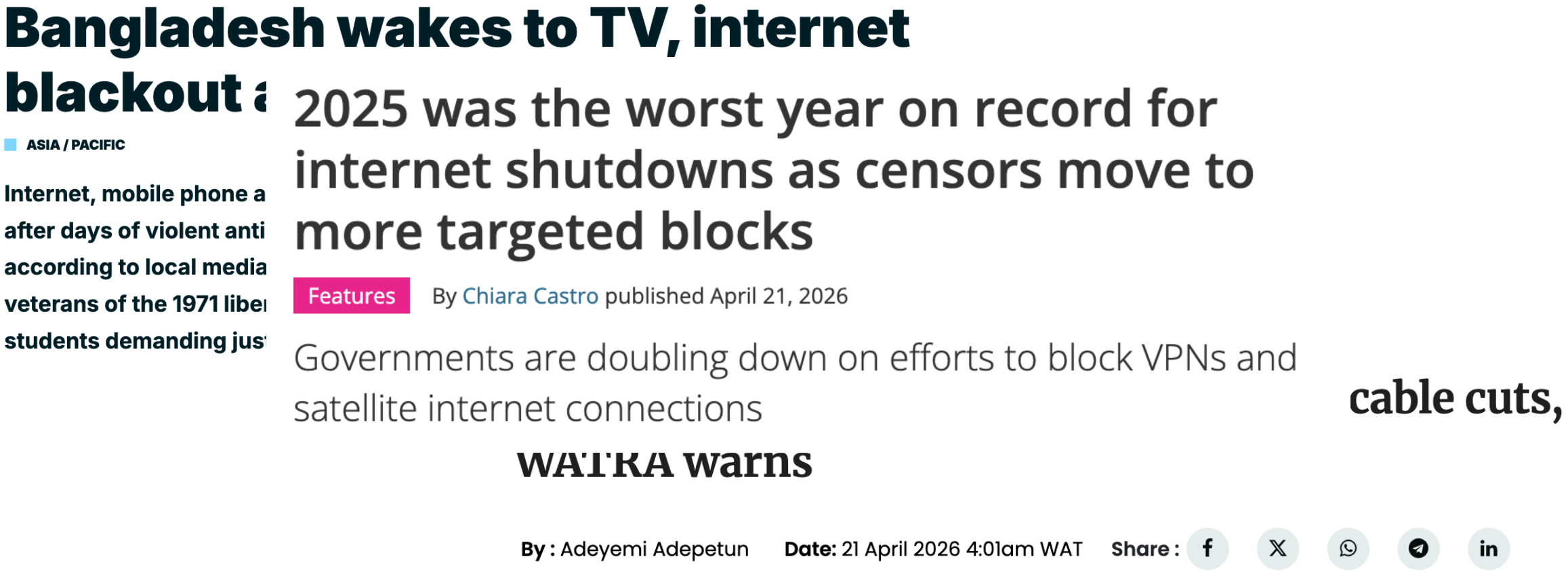
www.hpi.de

Internet Outages Keep Increasing in Numbers



Source: Access Now, „Rising repression meets global resistance: Internet Shutdowns in 2025“, 31st of March 2026.

Internet Outages Keep Increasing in Numbers



Bangladesh wakes to TV, internet blackout a

ASIA / PACIFIC

Internet, mobile phone a
after days of violent anti
according to local media
veterans of the 1971 liber
students demanding just

2025 was the worst year on record for internet shutdowns as censors move to more targeted blocks

Features By Chiara Castro published April 21, 2026

Governments are doubling down on efforts to block VPNs and satellite internet connections

cable cuts,

WATRA warns

By : Adeyemi Adepetun **Date:** 21 April 2026 4:01am WAT **Share :**

Outage Datasets Are Fragmented

- Every dataset defines „Outage“ differently
- Outage data shows incident in isolation

- Four failure modes:
  1. False positives
  2. Missed events
  3. Attribution errors
  4. Temporal inaccuracy

- No unified, reproducible way to consolidate & validate

Outage Datasets Are Fragmented

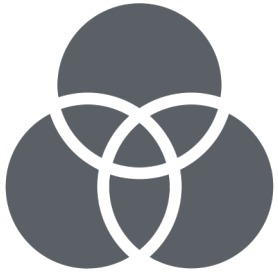
- Every dataset defines „Outage“ differently
- Outage data shows incident in isolation

- Four failure modes:
  1. False positives
  2. Missed events
  3. Attribution errors
  4. Temporal inaccuracy

We define an **Outage** as an **unavailability** in service at:
- A certain **point in time**
- A certain **AS**

- No unified, reproducible way to consolidate & validate

Research Questions



Can heterogeneous outage datasets be merged qualitatively and reproducibly using existing features?

Do the datasets cross-validate one another?

Does embedding merged Outages in a topology knowledge graph (IYP) improve analysis vs. dataset-level alone?

Contributions

Internet Outage Measurement Comparison

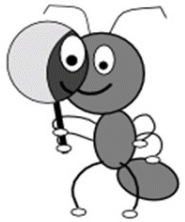
Simple Reproducible Merging Algorithm for multiple Internet Outage datasets

Embedding of merged outages into knowledge graph environment for graphic, causal analysis

Dataset Selection

Censored Planet

#KeepItOn



Mozilla

Internet Society Pulse

CLOUDFLARE

IODA
INTERNET OUTAGE DETECTION AND ANALYSIS

Tor Metrics

Dataset Selection Criteria

1. Data Structure
   - Structured data such as CSV instead of unstructured data (e.g. reports, forum posts)

2. Spatial resolution
   - Geographic assignment at least until the AS level

3. Temporal resolution
   - Each event must include start time and duration

4. Data independence
   - Avoidance of circular references

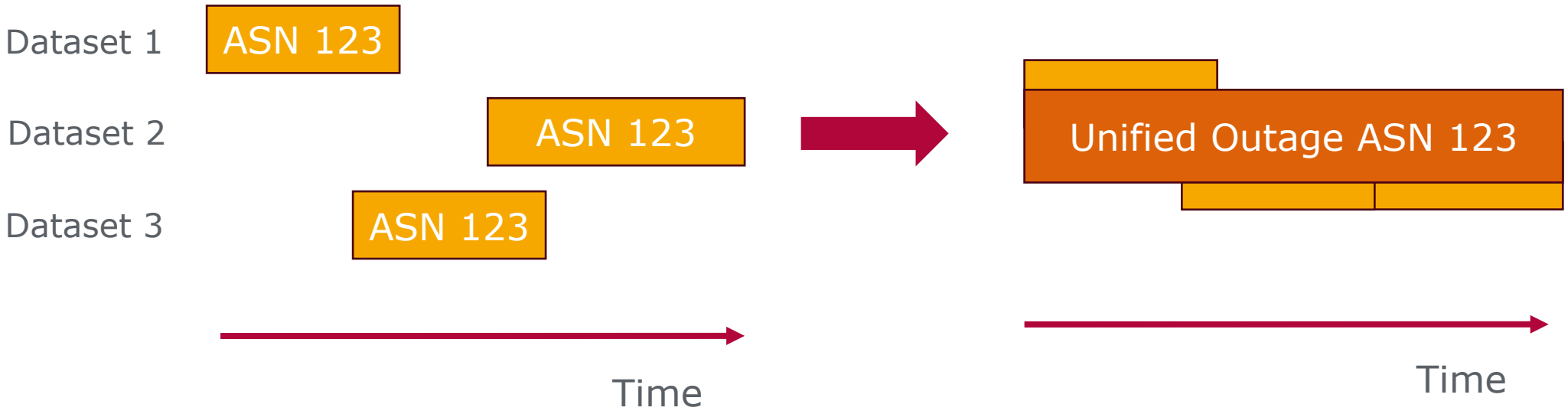
Selected Datasets

	 ANT	CROC	IODA
Outage definition	IPv4 /24 unreachability	Manually verified AS/country disruption	Ping, BGP and telescope signals
Spatial resolution	/24 block → AS	AS and country	AS or country
Temporal threshold	>= 11min probing round	No fixed threshold	Near real-time
Validation	Automated	Human-curated	Automated
	Broad reach, High FPR	**Small dataset, Precisely Curated**	**Diverse Signals, Near real-time**

Combining Multiple Internet Outage Datasets

AS: Which Autonomous System (AS) was affected

Time: When the outage occurred, start and end in UTC



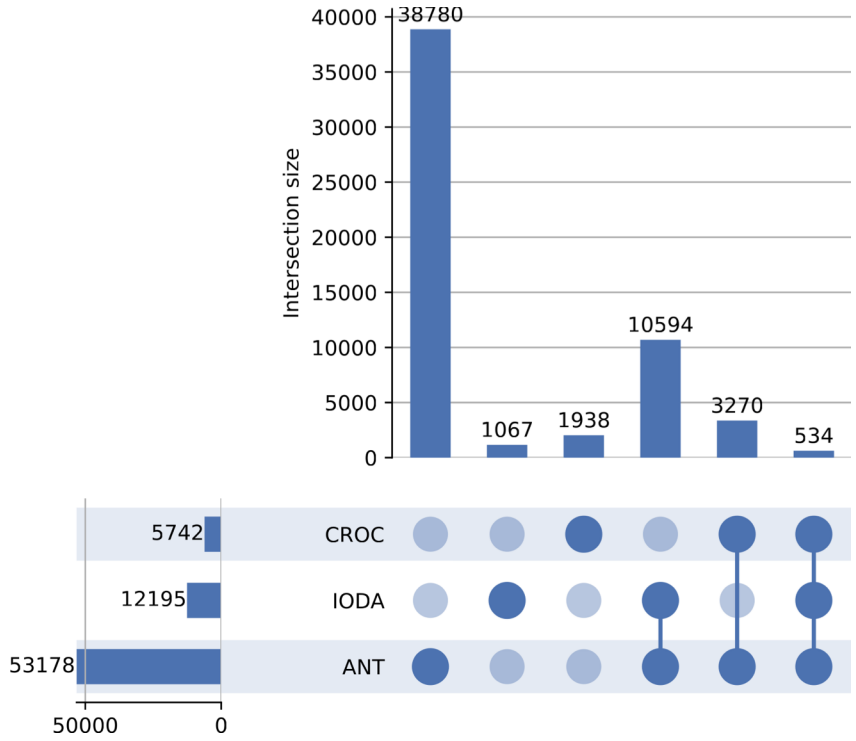
Results: Do the Datasets Cross-Validate?

- Most ASes are unique to one dataset
- CROC serves as anchor:

  49,3% of its ASNs are also included in ANT,

  5,4% by all three
- ANT & IODA are largely complementary only

  (only 9,3% of ANT AS are in IODA)



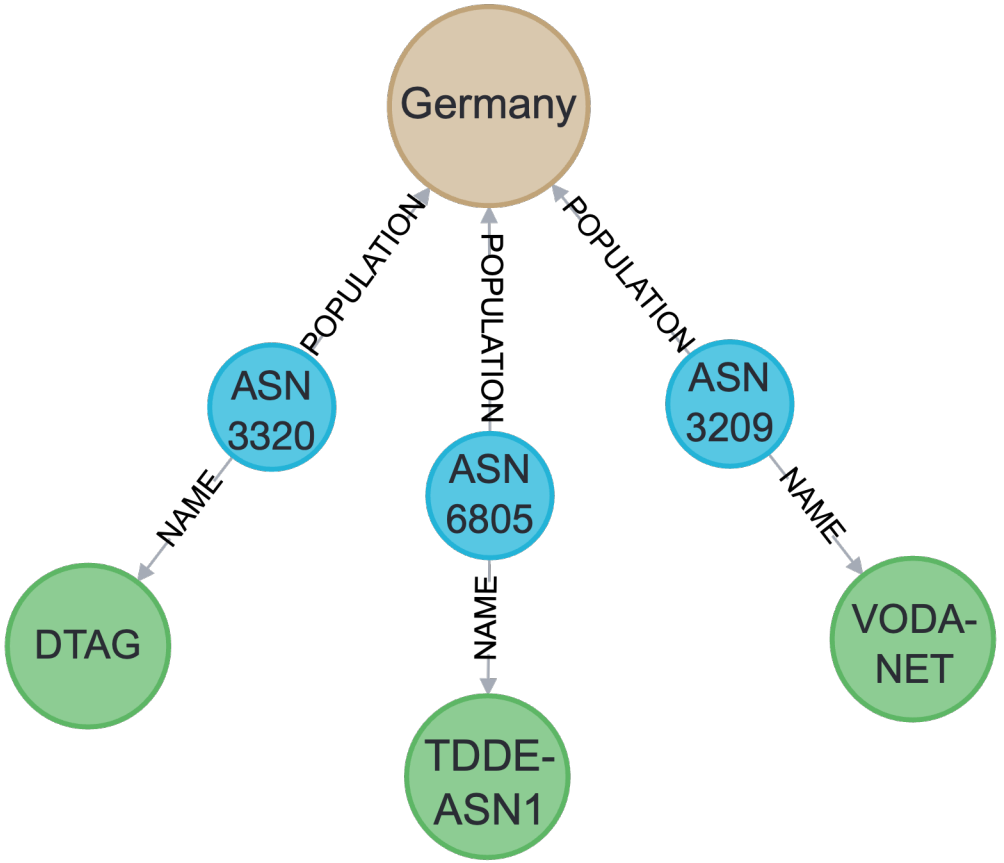
Results: Scale & Coverage

TABLE II: Summary statistics of 2024 outage datasets: event counts, median durations

	CROC	IODA	ANT	Combined
# Outage Events	9028	283,434	8,300,515	1,258,560
Median Duration (min)	840	60	88	66

Internet Yellow Pages

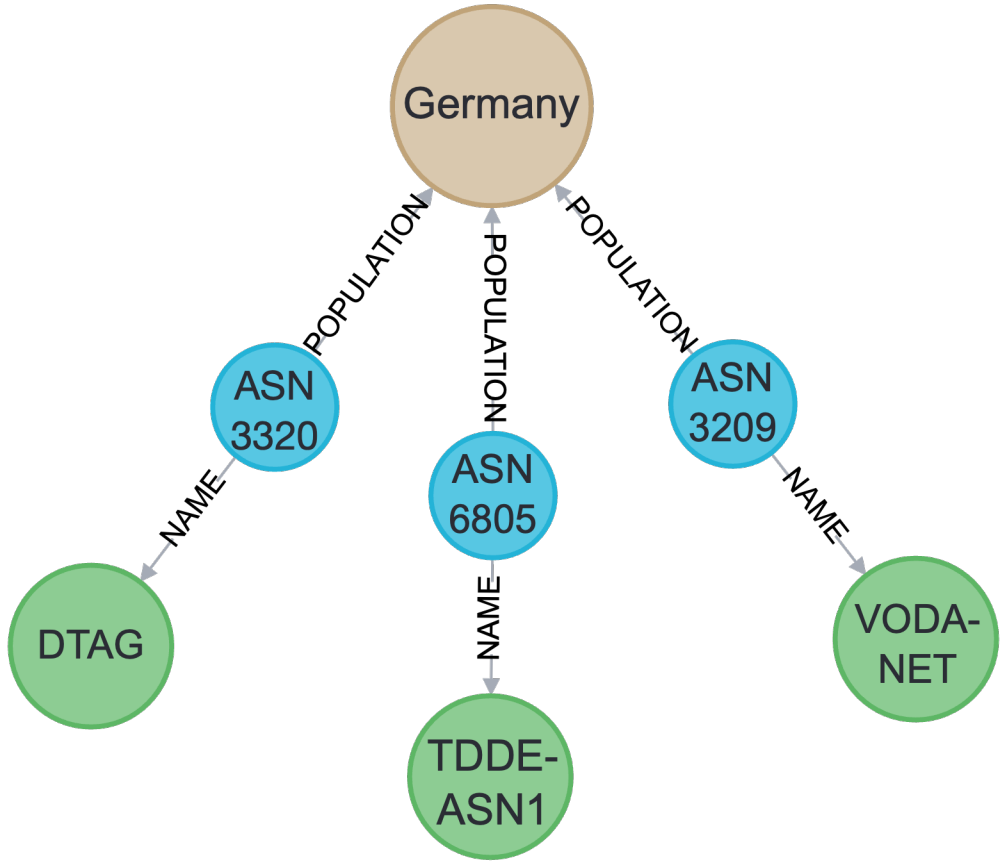
- Knowledge graph of Internet topology
- Contains
  - ASNs
  - IXPs
  - IP-Ranges
  - Statistics: Population and Traffic Ranks
- Integrates several, complementary data sources
- Shows relations visually explorable using Neo4j



Internet Yellow Pages

- Knowledge graph of Internet topology

- Provides:
  - Understanding of the Internet‘s structure
  - In-built visualization
  - Explorability
  - Exposes possible collateral impacts of Outage Events to Internet structure

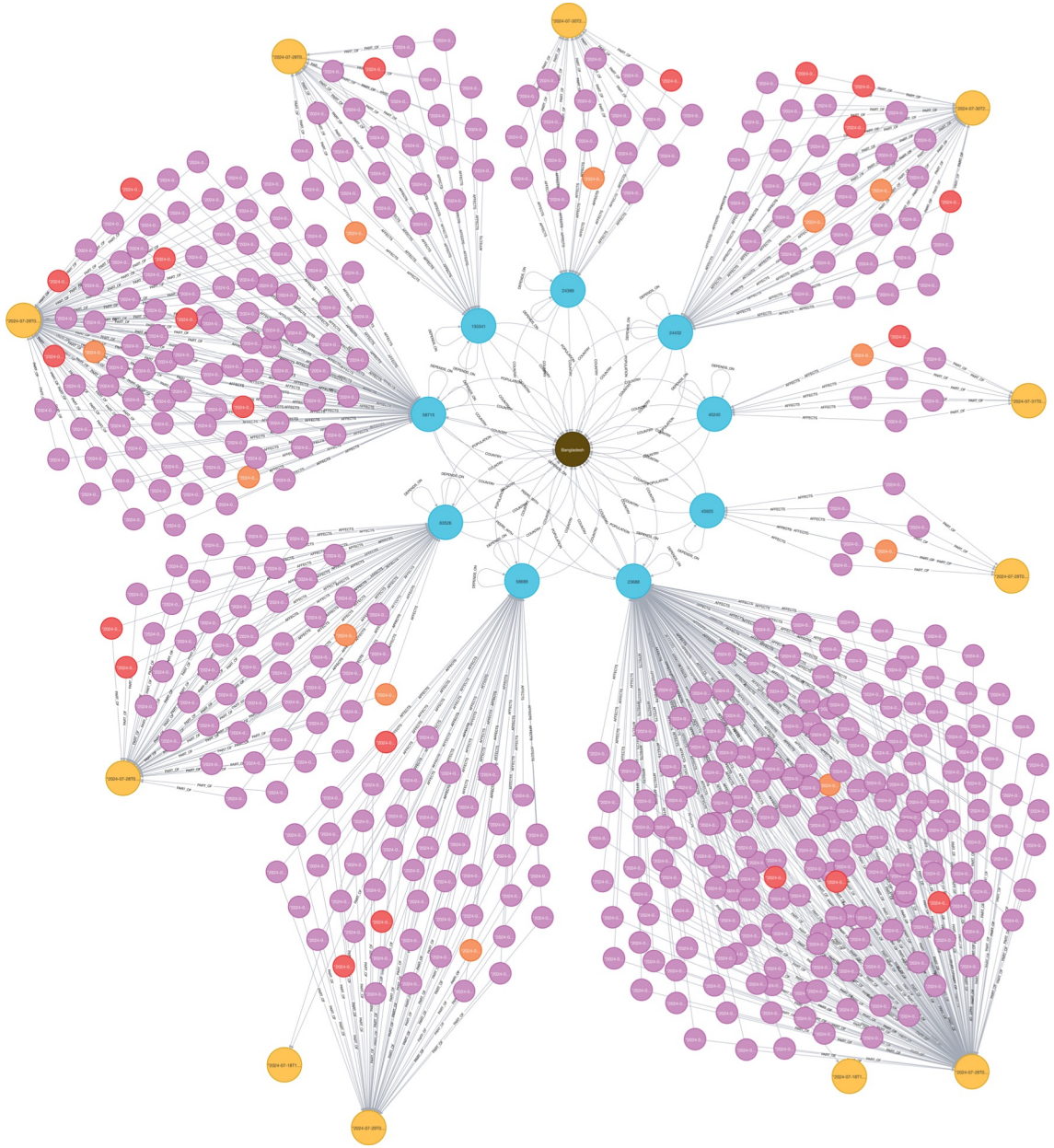


Case Study Bangladesh

- Initial report[1]:

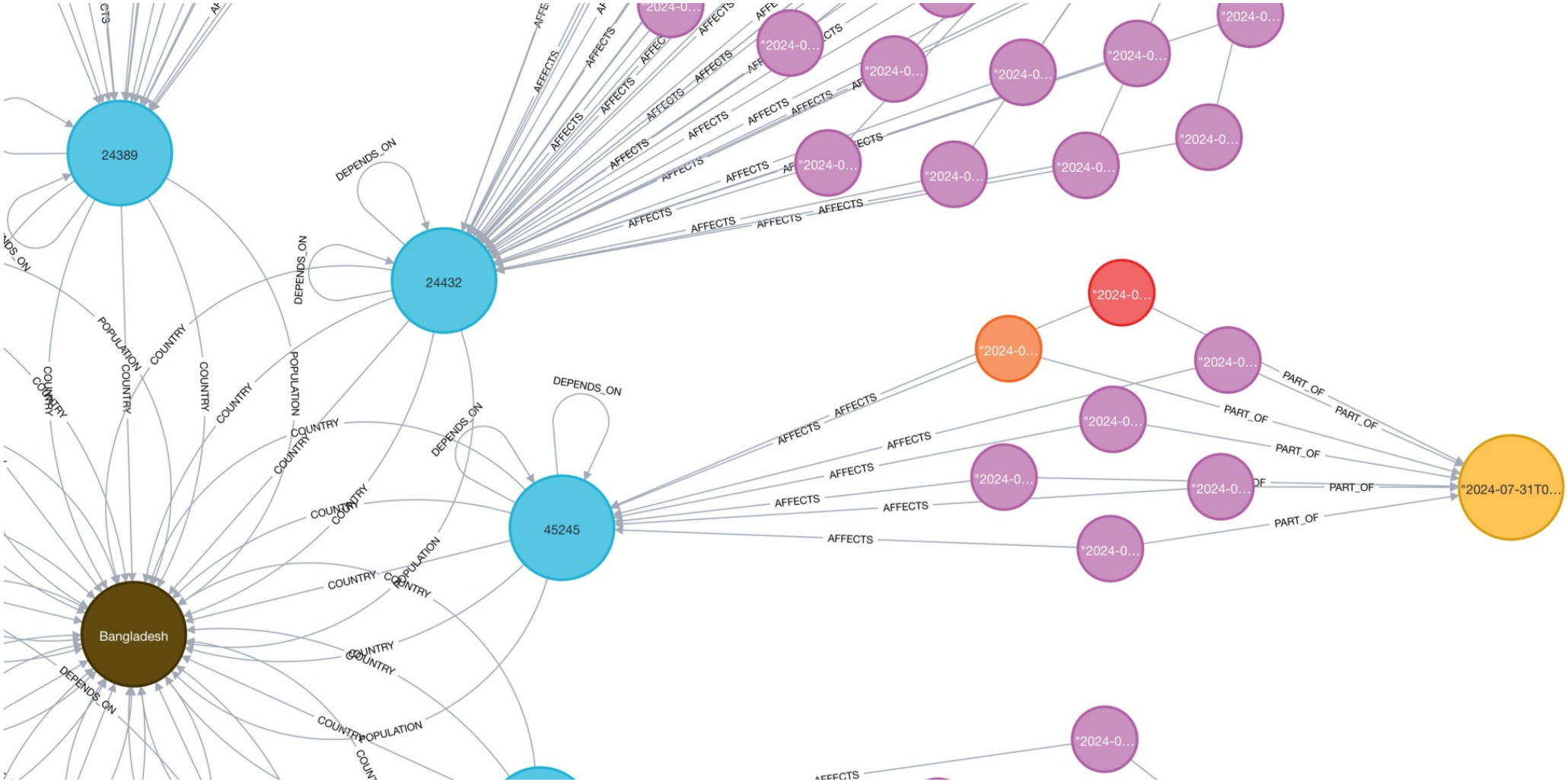
  Disruptions from July 18-23

- Reduced analysis to Top 9 AS (Telcos)



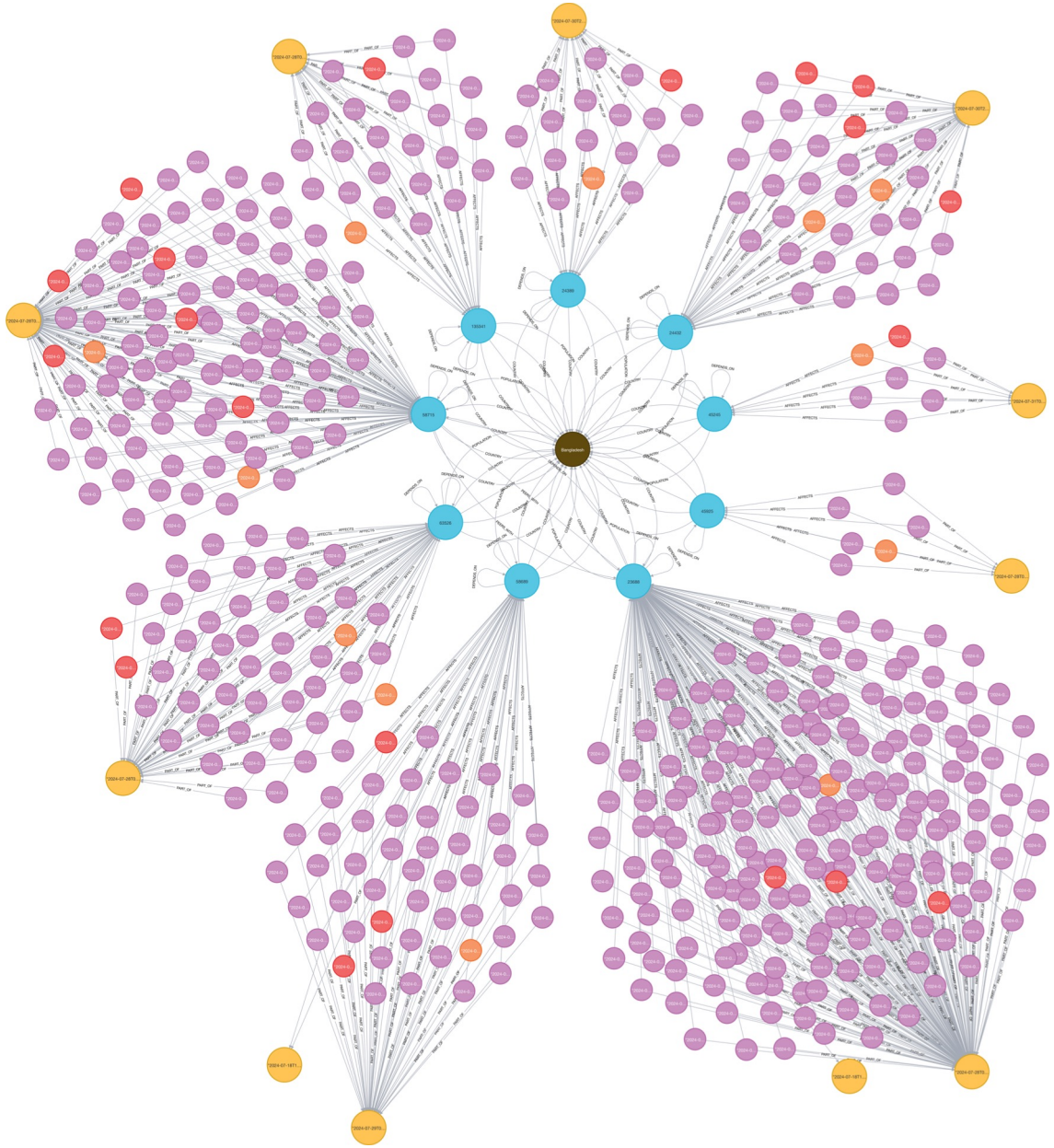
[1] https://pulse.internetsociety.org/en/shutdowns/mobile-internet-shut-down-in-bangladesh-amidst-protests/

Case Study Bangladesh



Case Study Bangladesh

- Initial report[1]:
  Disruptions from July 18-23
- Reduced analysis to Top 9 AS (Telcos)
- Merged timeline shows **earlier onset** and **longer duration** July 17-31
- This timeline was independently confirmed **a year after the shutdown** by OONI[2]

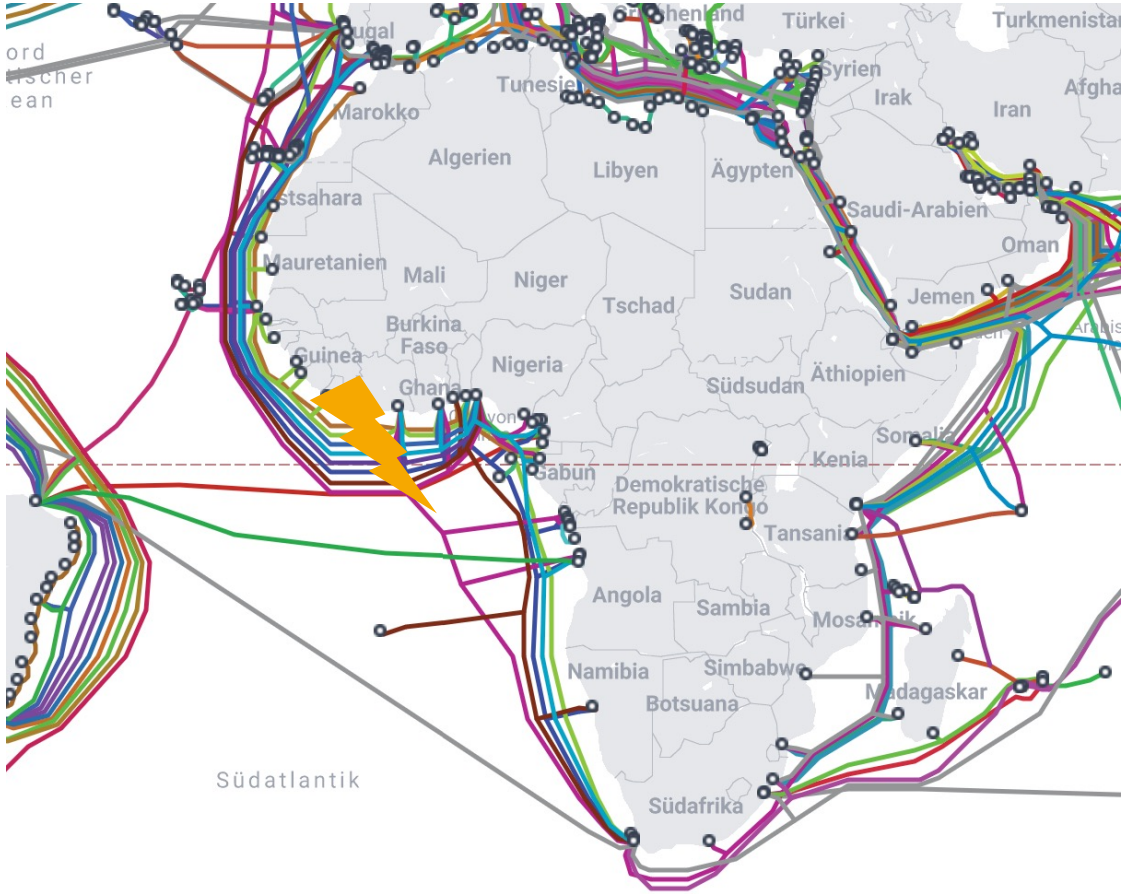


[1] https://pulse.internetsociety.org/en/shutdowns/mobile-internet-shut-down-in-bangladesh-amidst-protests/

[2] https://ooni.org/post/2025-bangladesh-report/

Case Study West Africa

- Initial report[3]: Submarine cable failure
  - 13 African countries
  - March 14th

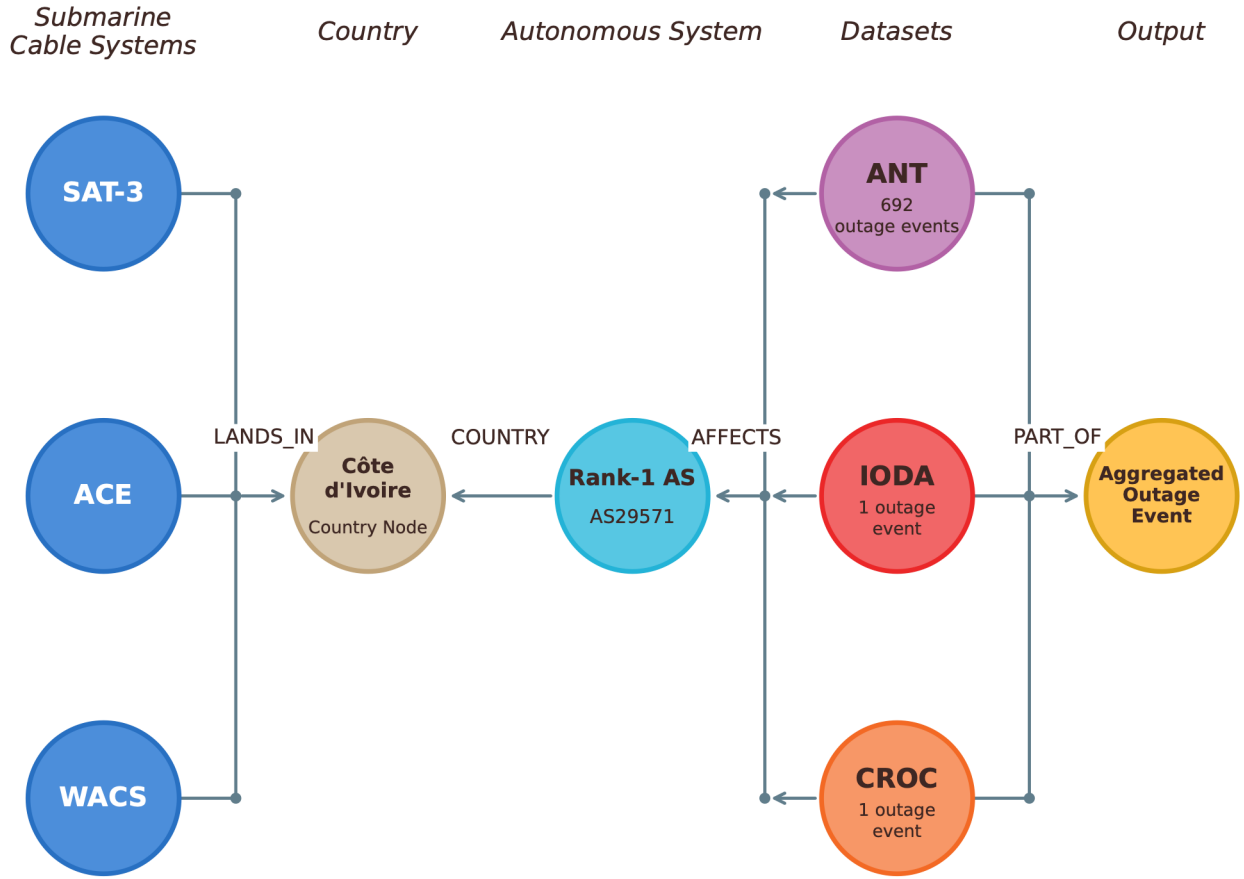


[3]: www.internetsociety.org/resources/doc/2024/2024-west-africa-submarine-cable-outage-report/

Case Study West Africa

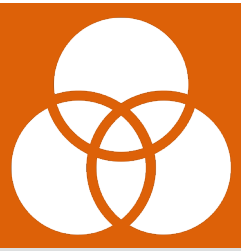
- Initial report[3]: Submarine cable failure
  - 13 African countries
  - March 14th
- ASes reported in 2 datasets
- Anomalies found in the ANT datasets (outage started 2 days early)
- Effects seen in countries not connected to cable, such as Burkina Faso / Senegal

Simplified schematic of Neo4j output



[3]: www.internetsociety.org/resources/doc/2024/2024-west-africa-submarine-cable-outage-report/

Summary



- Yes, using a simple time + AS merging algorithm and minimal preprocessing.
- Challenges remain with overestimation of duration of outages.

- For major incidents: Yes.
- CROC is most robust due to manual curation.
- ANT and IODA are complementary.

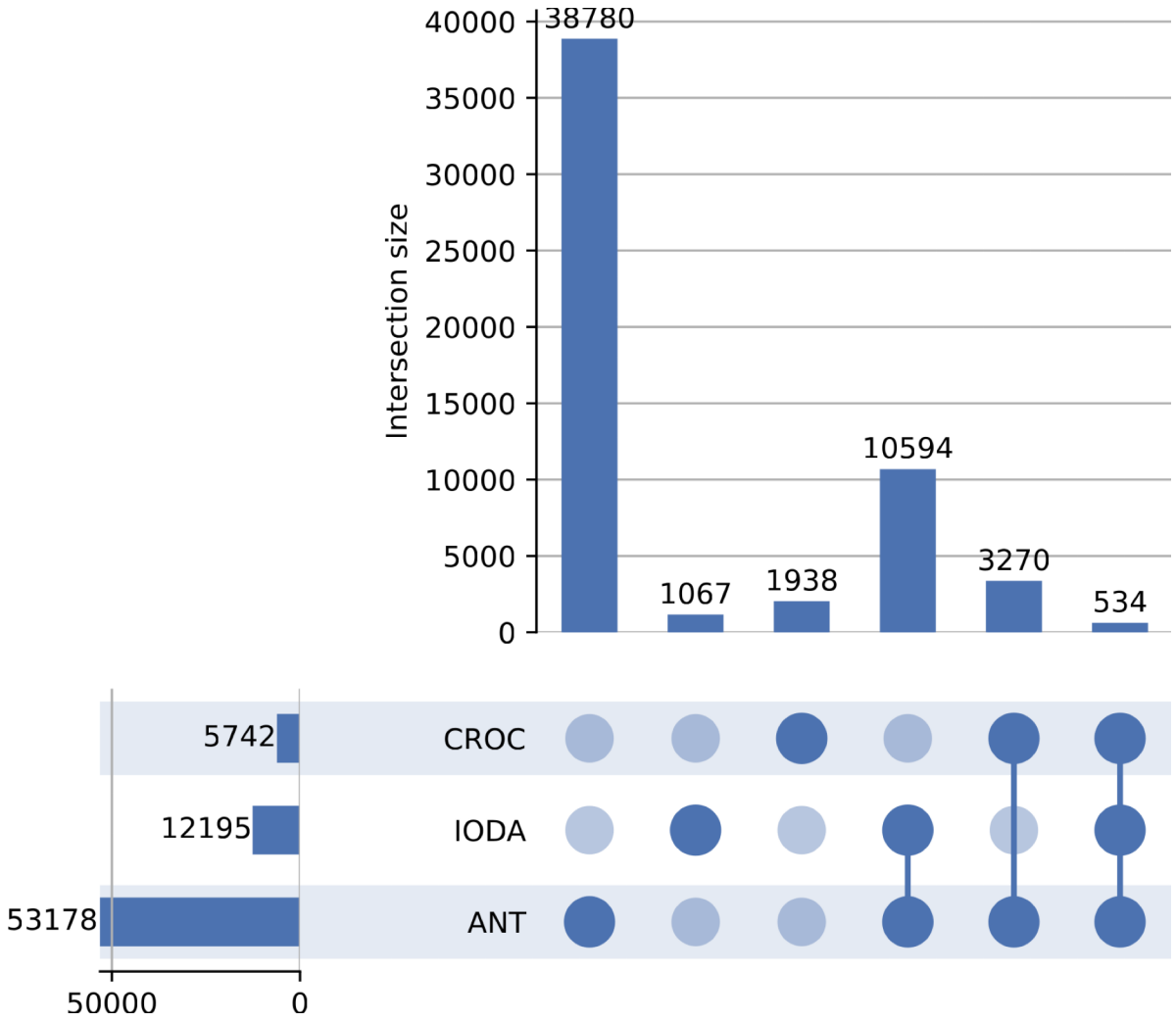
- IYP cuts complexity by visualizing deduction chains
- IYP links to physical infrastructure
- IYP shows collateral damage

Future Work

- Expansion of Outage Analysis to further case studies
- Agentic Automation of Outage Analysis
- Integration of real-time data with the constraint of IYP update intervals
- Release of reproducible Artifact soon

Backup Slides

Upset Plot ASN Intersection

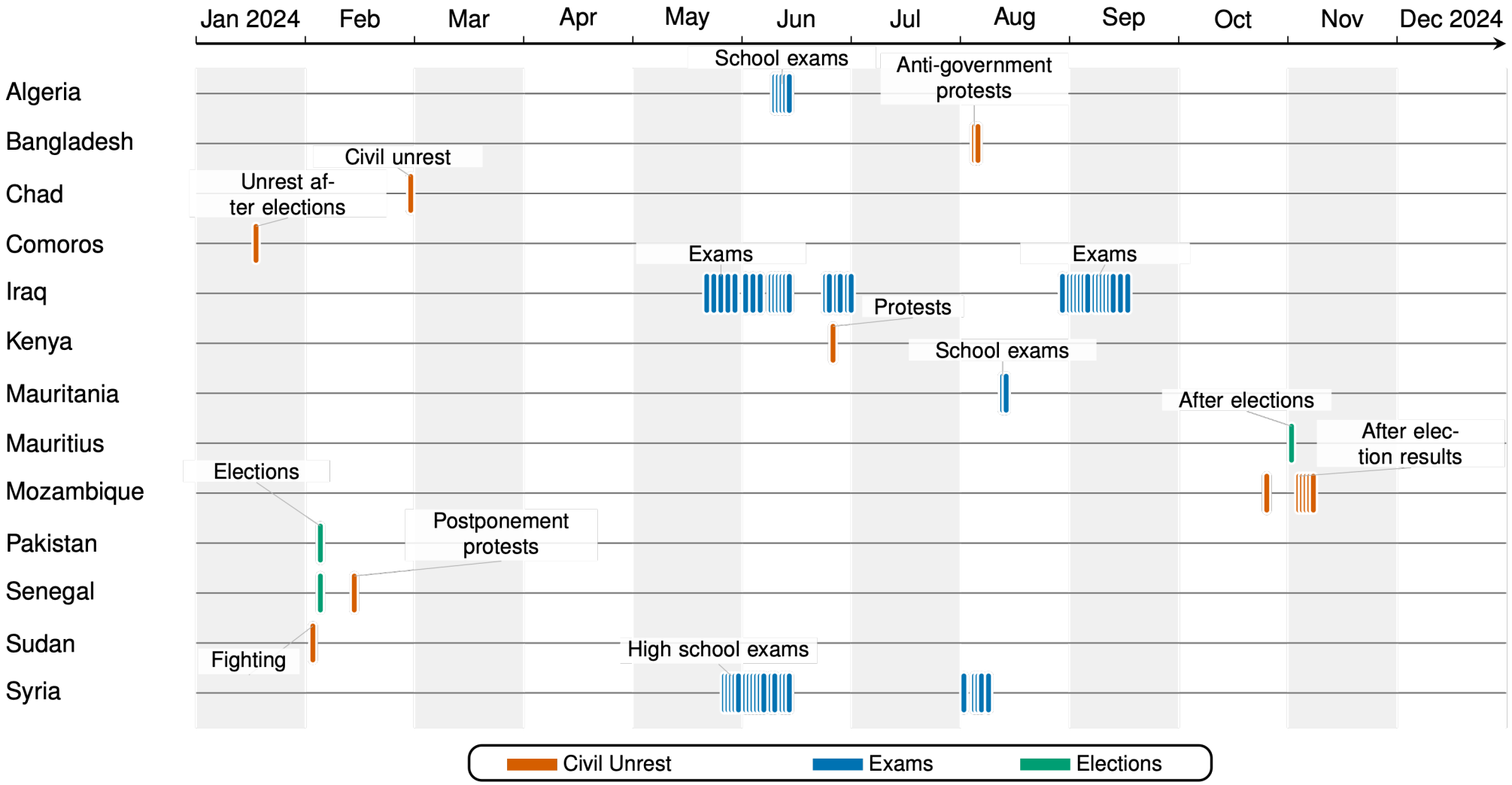


Results: Unified Dataset Characteristics

TABLE II: Summary statistics of 2024 outage datasets: event counts, median durations, and AS coverage.

	CROC	IODA	ANT	Combined
# Outage Events	9028	283,434	8,300,515	1,258,560
Median Duration (min)	840	60	88	66
AS Coverage	5753	12,206	53,178	56,183
AS in ≥ 2 datasets	—	—	—	14,398
AS in all 3 datasets	—	—	—	534

In 2024, Internet Outages Were At An All Time High



Source: Internet Society Pulse, „Who Shuts Down the Internet the Most?"

Dataset Dependencies

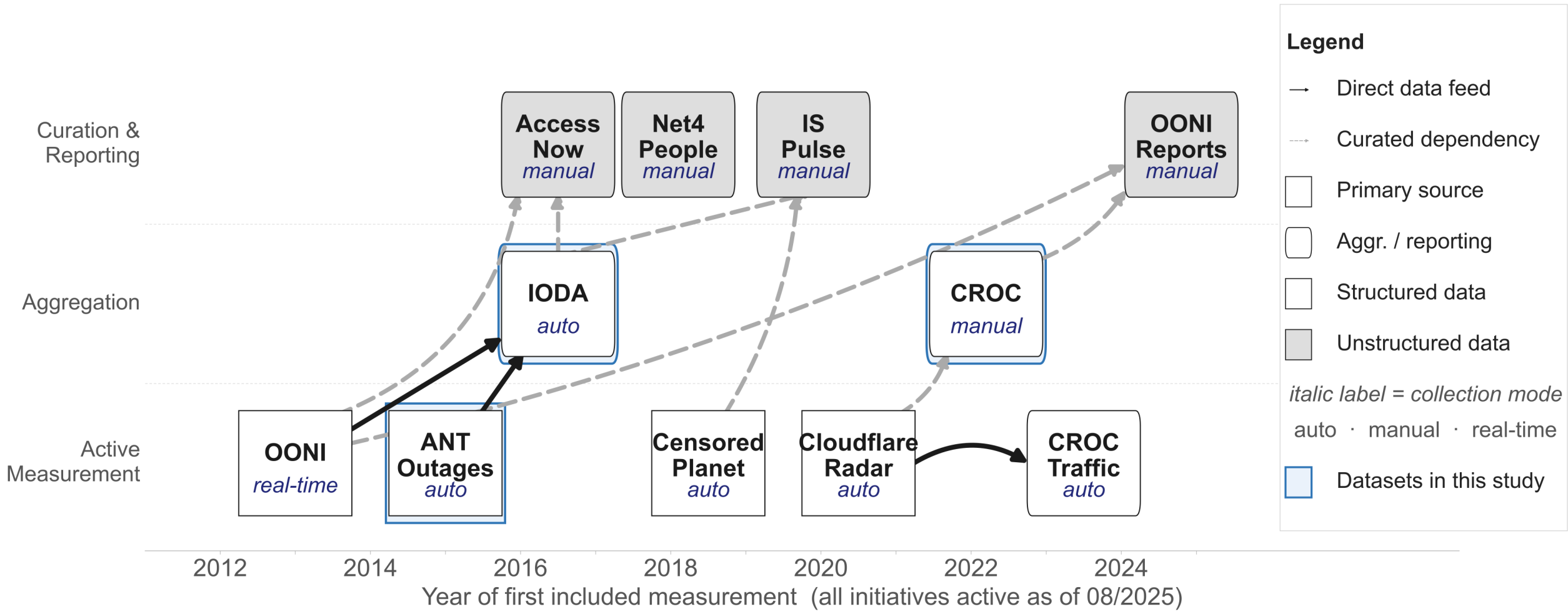
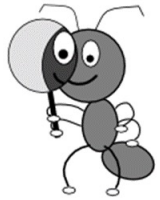


TABLE I: Comparison of datasets. Merging these datasets combines their complementary strengths, improving outage coverage across scale, validation, and signal diversity.

Property	ANT	CROC	IODA
Outage definition	IPv4 /24 unreachability	Manually verified AS/country disruption	Corroboration of ping, BGP, telescope signals
Spatial resolution	/24 block → AS	AS + country	AS or country
Temporal threshold	≥11 min probing round	No fixed threshold	Near-realtime
Collection mode	Automated probing	Manual curation	Automated multi-signal
Validation	Automated	Human-curated	Automated
Key limitation	High false-positive rate	Small coverage footprint	BGP-dependent detection

Selected Datasets



- Active Probing (CAIDA) (Trinocular instance)
- Google Transparency Report
- Traffic data from Merit Network Telescope

- Data from cloudflare routers
- Build up CROC as the one platform for outages

- Internet Outage Data by Active Probing (Trinocular)

TABLE III: Unique and overlapping outages by dataset. Percentages are row-normalized by the reference dataset and are not global proportions.

Intersection	CROC (%)	IODA (%)	ANT (%)
Only ANT	–	–	90.0
Only IODA	–	81.0	–
Only CROC	44.8	–	–
ANT & IODA	–	18.9	9.3
ANT & CROC	49.3	–	0.6
IODA & CROC	0.5	0.0	–
ANT, CROC & IODA	5.4	0.1	0.1

Integrating Outages into Internet Yellow Pages (IYP)

- Topology Knowledge Graph enables causal questions
- Added node types
  - `OutageEvent`: Individual events associated to each `Outage` to retain traceability
  - `Outage`: Cluster of `OutageEvents` for an AS
  - `SubmarineCable`: Associated Submarine Cables per Country

- Added relationship types
  - `Outage - :AFFECTS → AS`
  - `OutageEvent - :PART_OF → Outage`
  - `SubmarineCable - :LANDS_IN → Country`