

Stitching the Internet’s Blind Spots: 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†

Hasso Plattner Institute and University of Potsdam, Germany

*{firstname.lastname}@student.hpi.uni-potsdam.de

†{firstname.lastname}@hpi.de

Abstract—Outage measurements remain fragmented and difficult to cross-validate. This study addresses these challenges by merging three major datasets: Analysis of Network Traffic Outages (ANT), Cloudflare Radar Outage Center (CROC), and Internet Outage Detection and Analysis (IODA), into a unified 2024 outage view, integrated into the Internet Yellow Pages (IYP) for topology-aware analysis. Our results show that while ANT provides broad coverage and IODA adds supplementary events, CROC’s manual curation yields the most reliable cross-validation. Case studies of the July 2024 Bangladesh shutdown and West Africa submarine cable cut demonstrate that merged data confirms well-known events and reveals discrepancies such as extended outages and additional impacted regions. This work is the first reproducible merger of ANT, CROC, and IODA datasets, extending IYP to incorporate outages and submarine cable information, enhancing transparency and reliability in understanding Internet disruptions.

I. INTRODUCTION

In 2024, Internet shutdowns reached a new high [1], and submarine cables attracted increased attention as potential points of vulnerability [2], [3]. This is alarming since the Internet underpins modern life: it carries essential services (education, health, finance, civic participation) and is a primary channel for public discourse. Interruptions to connectivity, whether accidental, physical, or state-ordered, do more than inconvenience users: they disrupt access to vital information and services and can directly curtail freedom of expression, assembly, and the ability of civil society to monitor and report abuses.

Global monitoring revealed over 80 nationwide outages in 2024 (Fig. 1), underscoring both the importance of coordinated outage detection methods and the scale and temporal variability of events that such systems must recognize and validate. Additional complexity is introduced when submarine cables, the lifelines of global interconnection, are taken into account. A single failure can cascade across regions, since the majority of intercontinental traffic is tunneled through undersea cables. Recent incidents with submarine cables, including cable outages in the Baltic Sea, the Red Sea and West Africa highlight

the importance of robust, multi-source outage measurements [4]. However, the current measurement ecosystem lacks consistency, comparability, and reproducibility, limiting the ability to easily validate and investigate Internet outages, especially in cases of forced Internet cut-offs.

Research Gap and Objectives. Despite the growing importance of outage measurement, the measurement landscape remains fragmented. Internet measurement projects vary in their definitions of “outage”, their temporal and spatial scales, and their publication formats, each introducing geographic and methodological biases. That fragmentation makes it hard to (a) cross-validate events across sources, (b) confidently attribute cause, and (c) provide civil society and policy actors with a single contextualized view that links an outage to affected Autonomous Systems (ASes), countries, and critical physical assets such as submarine cables. Current datasets suffer from four distinct failure modes: false positives that are hard to filter without corroboration; missed events where only one dataset detects a confirmed outage; attribution errors where the source (censorship vs. cable cut) cannot be determined from a single dataset; and temporal inaccuracies where reported start/end times differ from ground truth. ANTs ping-based detection, for instance, yields a large share of single-source events, making false positives difficult to separate from genuine outages without corroboration from other systems. Conversely, confirmed events can be missed entirely: both IODA and CROC failed to capture Senegal during the West Africa cable cut, despite independent confirmation that the country was affected [6]. Other data sets expose different blind spots. OONI’s distributed probes are well suited to detecting censorship signals, but cannot by themselves distinguish a government-ordered shutdown from a physical infrastructure failure without integration with infrastructure-aware sources such as IODA or ANT [7], [8]. Finally, temporal boundaries may be misestimated, as in Bangladesh, where Internet Society Pulse reported the shutdown as ending on July 23, while merged evidence indicates disruptions persisted until July 28–31, later corroborated by [9].

This heterogeneity complicates cross-validation, attribution and integration with topology context such as AS relation-

Authors contributed equally to this paper.

by human intervention. Syamkumar et al. extend the scope with BigBen, a telemetry framework that applies NTP-based passive monitoring for large-scale Internet event detection [14]. Shah et al. propose Disco, which detects disconnections in long-lived TCP connections of RIPE Atlas probes, enabling near-real-time detection without generating measurement traffic [15]. Addressing the scale of results, Bogutz et al. introduce a severity-based reporting system integrated with the ANT project to surface the most significant events for public use [16]. While these approaches are effective in their domains, they operate independently and provide limited opportunity for cross-validation of events across sources.

A parallel line of work addresses censorship and government-directed Internet shutdowns, which are directly relevant because they manifest as infrastructure-level outages detectable by the same probing systems used for physical failures. Aceto and Pescapé survey existing censorship detection techniques and tools, demonstrating the breadth of the field and the lack of a shared detection standard [17]. Raman et al. measure censorship filter deployment at global scale with FilterMap, highlighting the difficulty of blockpage detection across heterogeneous ISP configurations [18]. Tsai et al. automate censorship event detection at scale with CenDTect, applying decision-tree learning to large measurement datasets without manual inspection [19]. Jin et al. present Disguiser, an end-to-end framework for fully automated global censorship detection across DNS, HTTP, and HTTPS [20]. Bischof et al. further bridge the two fields by providing the first longitudinal analysis that merges manually curated shutdown datasets with large-scale outage data to distinguish government-ordered disruptions from spontaneous failures [21]. Master and Garman provide a worldwide view of nation-state censorship across 70 countries, confirming that attribution and formatting inconsistencies remain the primary barriers to integrating censorship data across sources [22]. Two widely used censorship observatories, OONI and Censored Planet, collect large-scale probe measurements and are relevant to this landscape: their exclusion from this study is discussed explicitly below.

A smaller body of work has begun to tackle unification directly. Fontugne et al. compile heterogeneous public Internet measurement data into a knowledge-graph schema - the Internet Yellow Pages (IYP) - harmonizing 46 datasets through an ontology that remains stable as source schemas evolve [11]. While this demonstrates the technical feasibility of dataset integration, a fully automated and standardized outage center that merges and cross-validates multi-source outage events does not yet exist. Our work builds on IYP as both infrastructure and context layer, contributing the missing automated merge and cross-validation step. Open Observatory of Network Interference (OONI) documents Internet disruptions and censorship through volunteer-run probe measurements and cross-validation with external sources, including IODA data [23]. However, OONI was excluded from the merging framework for two reasons. First, its internal use of IODA signals means it is not independent of IODA, which would introduce circular cross-validation. Second, OONI’s measurement method uses

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 [13]	Manually verified AS/country disruption [10]	Corroboration of ping, BGP, telescope signals [32], [33]
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

fixed lists of websites to probe whether they are accessible, a censorship-detection approach, rather than measuring raw IP reachability across address blocks, which is the basis for ANT and IODA’s outage definitions [24]. Censored Planet was also excluded: it uses remote measurement techniques (Augur, Satellite, Hyperquack) designed specifically for censorship detection rather than IP reachability [25], [26], and the 2024 processed dataset was not accessible for this study (access requires a billing account) [27].

However, existing platforms do not yet provide automated merging or cross-validation of multi-source outage data.

II. METHODOLOGY

To cross-validate Internet outages, datasets are selected based on their structure, granularity, temporal resolution, and mutual compatibility, and then integrated with Internet topology data. We focus on three representative platforms: ANT, CROC, and IODA, which offer complementary measurement perspectives through active probing, manual curation, and multi-source aggregation (Fig. 2).

A. Dataset Landscape

The Internet measurement landscape comprises a variety of data sources, with differing levels of mutual dependence and diverse data processing techniques. Platforms such as Internet Society Pulse and AccessNow’s KeepItOn [28], [29] curate reports of large-scale disruptions but do not provide measurement data or formal validation procedures. Measurement-driven outage datasets include ANT, CROC, IODA, and OONI. These projects differ in coverage, validation, and collection method as shown in Table I: ANT relies on active IPv4 probing, IODA combines multiple automated signals, and CROC provides manually verified and cause-attributed outages [10], [30]. OONI complements these datasets by documenting Internet disruptions and censorship through probe measurements and cross-validation with external sources [31].

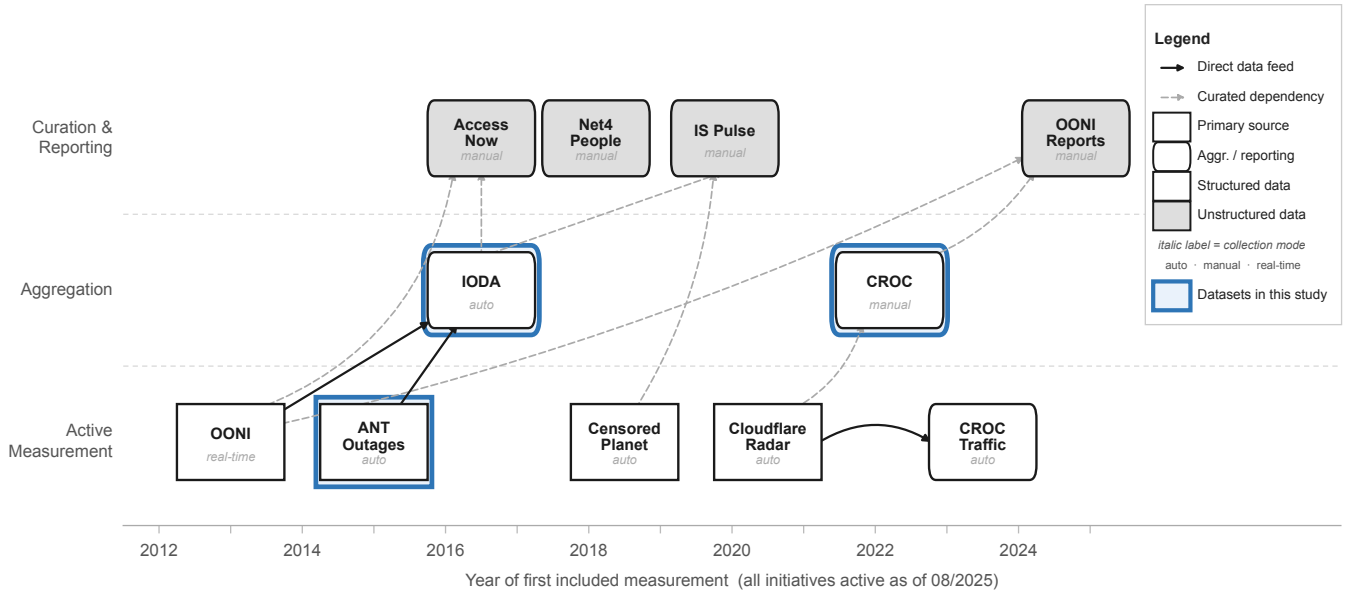


Fig. 2: Overview of the Internet outage measurement landscape and data-flow dependencies among major monitoring initiatives. Axes encode processing level and first measurement year; shapes, fills, and arrows distinguish source type, data structure, and dependency type. Blue nodes mark the three datasets merged in this study: ANT, IODA, and CROC, which provide complementary reach, detection signals, and curated validation.

B. Data Processing, Merging, and Tools

Suitable datasets from the previously mentioned sources were selected based on the following criteria: 1) **Data structure**: Only datasets with structured data were considered to facilitate analysis without extensive pre-processing. 2) **Spatial resolution**: Datasets needed to provide event data at least at the AS level, enabling geographic assignment and identification of related incidents. 3) **Temporal resolution**: Each incident was required to include a start time and duration. 4) **Data independence**: To accurately compare event coverage, datasets had to be independent, avoiding circular data references. Based on these criteria, ANT, CROC, and IODA were deemed compatible for merging. While IODA data contributes to CROC’s verification, it remains independent. Our analysis focuses on 2024, since ANT dataset for 2025 is not yet available.

Data Processing: We standardized timestamps across datasets, deriving end times where necessary. Nationwide CROC events were expanded to all ASes in the country using IYP. To minimize noise and reduce dataset size, only AS-level events observed in at least two datasets were retained. Since outage datasets aim to detect the same real-world events and time-coincident disruptions in the same AS very likely share a root cause, we used the temporal dimension for maximal clustering. Note that the case studies apply a causal query approach that retrieves all outages for rank-1 ASes in affected countries, including single-source events, in order to explicitly surface coverage gaps across datasets.

Merging Algorithm: Outage events from all datasets were

gathered by AS, sorted by start time, and iteratively merged if their time intervals overlapped. Merged outages span from the earliest start to the latest end time of overlapping events, with references from each resulting Outage node to all its constituent OutageEvent nodes maintained for traceability.

Internet Yellow Pages: We integrated merged outages into the IYP knowledge graph to enable topology-aware analysis and to make use of querying the data using the Cypher language. Together with the possibility of visually plotting graph results this supports contextual analysis of outages, revealing impacts on dependent ASes. Furthermore, using TeleGeography’s submarine cable data, a SubmarineCable node was introduced and linked to existing Country nodes by a :LANDS_IN relationship [34], enabling high-level views of affected cables and related outages. In summary, this methodology integrates, merges, and contextualizes heterogeneous outage data to analyze dataset coverage, cross-validation, and topology-aware insights.

Software, Dataset, and Reproducibility: The code and configurations will be publicly released to ensure reproducibility. The CROC and IODA datasets are publicly available; the ANT dataset requires a separate request under a non-disclosure agreement [30].

III. RESULTS AND ANALYSIS

We first compare AS coverage and outage-duration distributions, followed by an analysis of their event overlaps after merging.

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

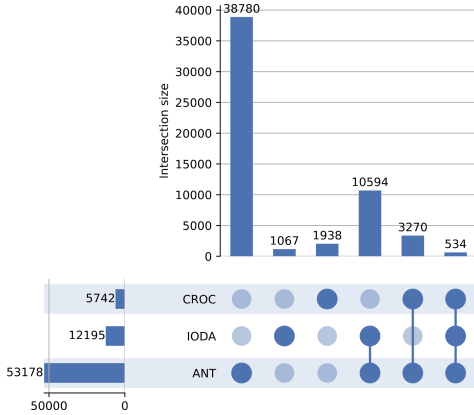


Fig. 3: UpSet plot visualizing AS intersections among ANT, CROC, and IODA datasets. Despite its smaller footprint, CROC shows disproportionately high event-level agreement with ANT (49.3%), reflecting the precision gains of manual curation.

A. Comparison of Measurement Datasets

Table II summarizes 2024 outage datasets, showing CROC with the fewest events, IODA nearly two orders of magnitude larger, and ANT with over 8.3e6 outages. Despite its size, ANT’s median outage duration (88 min.) is short compared to CROC’s (840 min.), reflecting its lower degree of event aggregation. The merged dataset’s median (66 min.) aligns closely with IODA’s (60 min.). ANT covers the most ASes, followed by IODA and CROC; the merged dataset spans over 56,000 unique ASes. Overlaps are substantial: 14,000 ASes appear in at least two datasets, and 534 are common to all three, indicating complementarity and redundancy (see Table II). As ANT probes all 5.3M /24 IPv4 blocks every 11 minutes, any brief unreachability is an event, including BGP flaps and noise. The large count reveals ANT’s measurement breadth, not the outage severity [13]. CROC’s much longer median duration (840 min. vs. ANT’s 88 min.) reflects the step of human validation, consulting with IODA and selection for sustained outages [10], [35]. A finer grained view on dataset overlap is provided in Fig. 3. This UpSet plot shows that most ASes are unique to one dataset, with ANT contributing the largest exclusive share (38,780). Notable overlaps exist between ANT–IODA (10,594), ANT–CROC (3,270), and IODA–CROC (1,067), with 534 ASes shared by all three datasets.

Table III quantifies the intersections of measured outage

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

TABLE IV: AS ranking by outage counts in 2024.

AS	Country	Outage Count
24547	China	2526
39392	Czechia	1301
42965	Germany	1234
6877	Ukraine	1215
2501	Japan	1175
27877	Argentina	1072
24554	India	1058
133772	Hong Kong	1056
35154	Russian Federation	1019
27983	Argentina	974

events per dataset using the presented merging algorithm for intersection. ANT outages are 90% unique, with modest confirmation by IODA and minimal by CROC. IODA shows 81% unique outages. CROC exhibits the strongest overlap, with nearly half (49.3%) coinciding with ANT events, reflecting its manual validation process and higher precision. Cross-validation is strongest in CROC. 49.3% of its outages are confirmed by ANT, and 5.4% by all three datasets, reflecting its human-curation process. By contrast, ANT and IODA are largely complementary: only 9.3% and 18.9% of their outages are mutually confirmed. This means that for the majority of individual events, datasets disagree, but for major incidents cross-dataset agreement is achievable, as shown in the case studies. Further analysis focuses on the geographical coverage of datasets and the spatial concentration (AS level) of measured outages: Figures 4a–4b illustrate geographic coverage across datasets. The merged dataset retains the dominant patterns but reflects the underlying AS population distribution which can bias country-level visualizations. This dynamic is illustrated by the case of the United States, which appears prominently in red on the outage map and exhibits a high absolute number of outages, yet none of its ASes appear among those with the highest outage counts (see Table IV).

Table IV ranks ASes by outage count, identifying globally problematic networks (e.g., AS 24547 in China). Unlike country-level views, AS-level rankings avoid bias from nations with many ASes. Taken together, the quantitative comparison, overlap analysis, global mapping, and AS-level ranking demonstrate both the promise and the challenges of integrating outage datasets. Methodological differences among CROC, IODA, and ANT introduce biases in coverage, duration, and

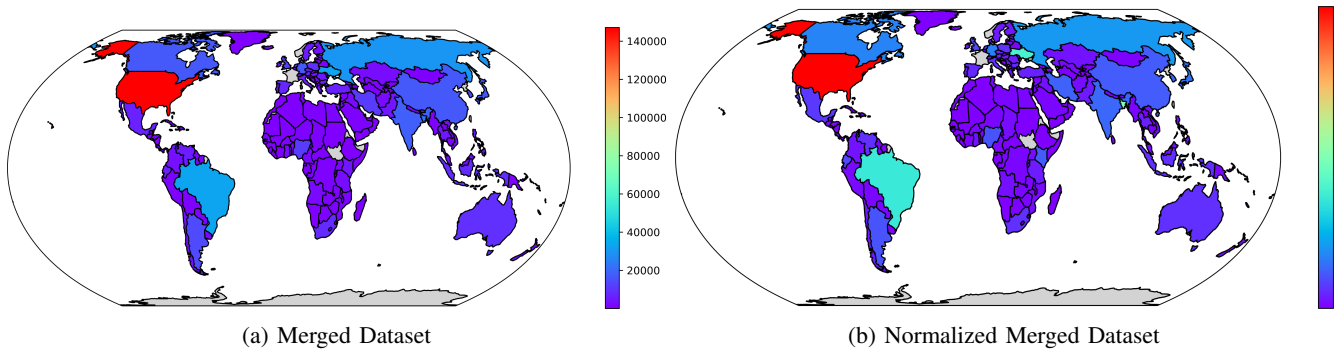


Fig. 4: Equal Earth projection map showing Internet outages by raw magnitude and a normalized merged dataset. Colors reflect each country’s total outage counts. Normalization ensures equal contribution from all sources. Multi-source spikes are most prominent in Bangladesh and Ukraine, confirming those as the strongest cross-validated outage regions in 2024.

validation rates. We establish CROC as highest-precision dataset and anchor while ANT provides broad data but also introduces noise, resulting in a merged dataset that reveals events invisible to any single source, such as the identification of globally problematic networks such as AS 24547 (China).

IV. MEASUREMENT CASE STUDIES

The comparative analysis in Tables II and III shows the diversity and complementarity of available outage datasets. While these aggregate statistics clarify source coverage and overlap, they remain abstract. To demonstrate the practical value of dataset integration and validate that merging yields meaningful, context-rich insights, we conducted case studies. By examining specific events, we illustrate how merged outage data can be leveraged within IYP to analyze both government-directed shutdowns and infrastructure-related disruptions. To represent diverse outage types, we selected one government-directed Internet shutdown and one outage caused by a submarine cable cut due to a natural event, both from 2024. While outages could be identified through data exploration on the consolidated data alone, we conducted case studies on well-documented incidents. This approach allows for cross-verification of our insights with external reports, emphasizing the accuracy and practical value of our methodology.

For nationwide censorship incidents, the Internet Society Pulse database serves as ground truth. Building on this, Fig. 1 summarizes the number of distinct nationwide government-caused censorship events per country and month in 2024. Only full nation shutdowns are shown. To showcase a government-directed shutdown, we selected the July 2024 shutdown in Bangladesh. This event was chosen due to its spontaneous occurrence and the range of Internet disruption observed, from partial to complete, showcasing the full range of nation-state responses. As a regular outage, we examine the West Africa submarine cable outage starting mid-March 2024, also reported by the Internet Society [36]. Together, these case studies offer a comprehensive perspective on outages and censorship, serving as a practical testbed for validating our

```
MATCH (o:Outage)-[:PART_OF]-(oe:OutageEvent)-[:AFFECTS]-(a_j
↳ :AS)-[:COUNTRY]-(c:Country)
WHERE c.name = "Bangladesh" AND oe.startDate >=
↳ datetime("2024-07-18T00:00:00Z")
AND oe.startDate < datetime("2024-07-23T23:59:59Z")
AND a.asn IN [24389,24432,45245,45925,23688,58689,58715, j
↳ 63526,135341]
RETURN a, c, oe, o
```

Listing 1: Cypher query from Bangladesh case study

unified, topology-aware approach to measuring Internet disruptions.

Bangladesh Internet Shutdown. In July 2024, the Bangladeshi government cut off the Internet, widely reported as a near-total nationwide outage [37]. The event was independently detected by ANT, CROC and IODA, making it a suitable test case for cross-dataset validation. We analyzed the merged dataset using a local IYP instance with topology data from July 2024 [38]. Since a national Internet shutdown is usually carried out on Internet Service Provider (ISP) level, we focused our analysis on the most relevant ISPs in Bangladesh. This reduced the complexity of investigation from more than 1700 Bangladesh-affiliated ASes, to nine ASes, which have been identified based on AS-population coverage data within IYP. Out of these nine ASes, four belong to major mobile operators and five to large broadband providers (see Table VI). This analysis is supported by external reports indicating that Bangladesh’s 190+ million Internet users primarily access the Internet via cellular networks operated by Grameenphone, Robi Axiata, Banglalink, and Teletalk, whereas the approximately 13 million broadband users are served by numerous smaller ISPs [39], [40]. Therefore, this case study focuses on these nine key ASes.

To further investigate the outage, we queried our locally hosted IYP instance for ASes affected between 18 and 23 July 2024, merging overlapping AS-level events reported across ANT, CROC, and IODA datasets (see Listing 1).

The resulting graph is visualized automatically, as shown in Fig. 5, illustrating the outage affecting the specified

TABLE V: West Africa Cable Cut: affected ASes and operators, outage start and end times, with dataset coverage.

AS	Country	Cable	Operator	Outage Start	Outage End	ANT	CROC	IODA
2860	Portugal	[Africa Coast to Europe (ACE), SAT-3/WASC, West Africa Cable System (WACS), MainOne]	NOS_COMUNICACOES NOS COMUNICACOES, S.A.	01.01. 00:00	01.01. 00:00	x	-	x
3352	Spain	[Africa Coast to Europe (ACE), SAT-3/WASC, West Africa Cable System (WACS)]	Telefonica_de_Espana TELEFONICA DE ESPANA S.A.U.	01.01. 00:00	01.01. 00:00	x	-	x
3215	France	[Africa Coast to Europe (ACE)]	AS3215 Orange S.A.	11.03. 03:21	17.03. 13:15	x	-	-
36912	Cameroon	[SAT-3/WASC, West Africa Cable System (WACS)]	ORANGE	12.03. 16:28	28.03. 00:00	x	x	x
37461	Guinea	[Africa Coast to Europe (ACE)]	ORANGE-	14.03. 05:00	16.03. 15:46	x	x	-
37309	Gambia	[Africa Coast to Europe (ACE)]	QCell	14.03. 05:00	14.03. 05:51	x	x	-
37094	Liberia	[Africa Coast to Europe (ACE)]	OrangeLiberia	14.03. 05:00	17.03. 00:00	x	x	-
29571	Côte d'Ivoire	[Africa Coast to Europe (ACE), SAT-3/WASC, West Africa Cable System (WACS), MainOne]	ORANGE-COTE-IVOIRE	14.03. 07:40	17.03. 20:28	x	x	x
8346	Senegal	[Africa Coast to Europe (ACE), SAT-3/WASC, MainOne]	SONATEL SONATEL-AS Autonomous System	14.03. 07:44	14.03. 08:17	x	-	-
30986	Ghana	[Africa Coast to Europe (ACE), SAT-3/WASC, West Africa Cable System (WACS), MainOne]	SCANCOM	14.03. 07:45	12.04. 07:15	x	x	x
37424	Benin	[Africa Coast to Europe (ACE), SAT-3/WASC]	Spacetel	14.03. 07:45	25.03. 15:12	x	x	-
29465	Nigeria	[Africa Coast to Europe (ACE), SAT-3/WASC, West Africa Cable System (WACS), MainOne]	VCG-AS MTN NIGERIA Communication limited	14.03. 10:30	17.03. 11:00	x	x	x
24691	Togo	[West Africa Cable System (WACS)]	TOGOTEL-AS TogoTelecom, Togo	14.03. 10:30	15.03. 03:30	x	x	-
36996	Namibia	[West Africa Cable System (WACS)]	TELECOM-NAMIBIA	14.03. 10:30	14.03. 12:30	x	-	x
8346	Senegal	[Africa Coast to Europe (ACE), SAT-3/WASC, MainOne]	SONATEL SONATEL-AS Autonomous System	14.03. 13:48	14.03. 14:03	x	-	-

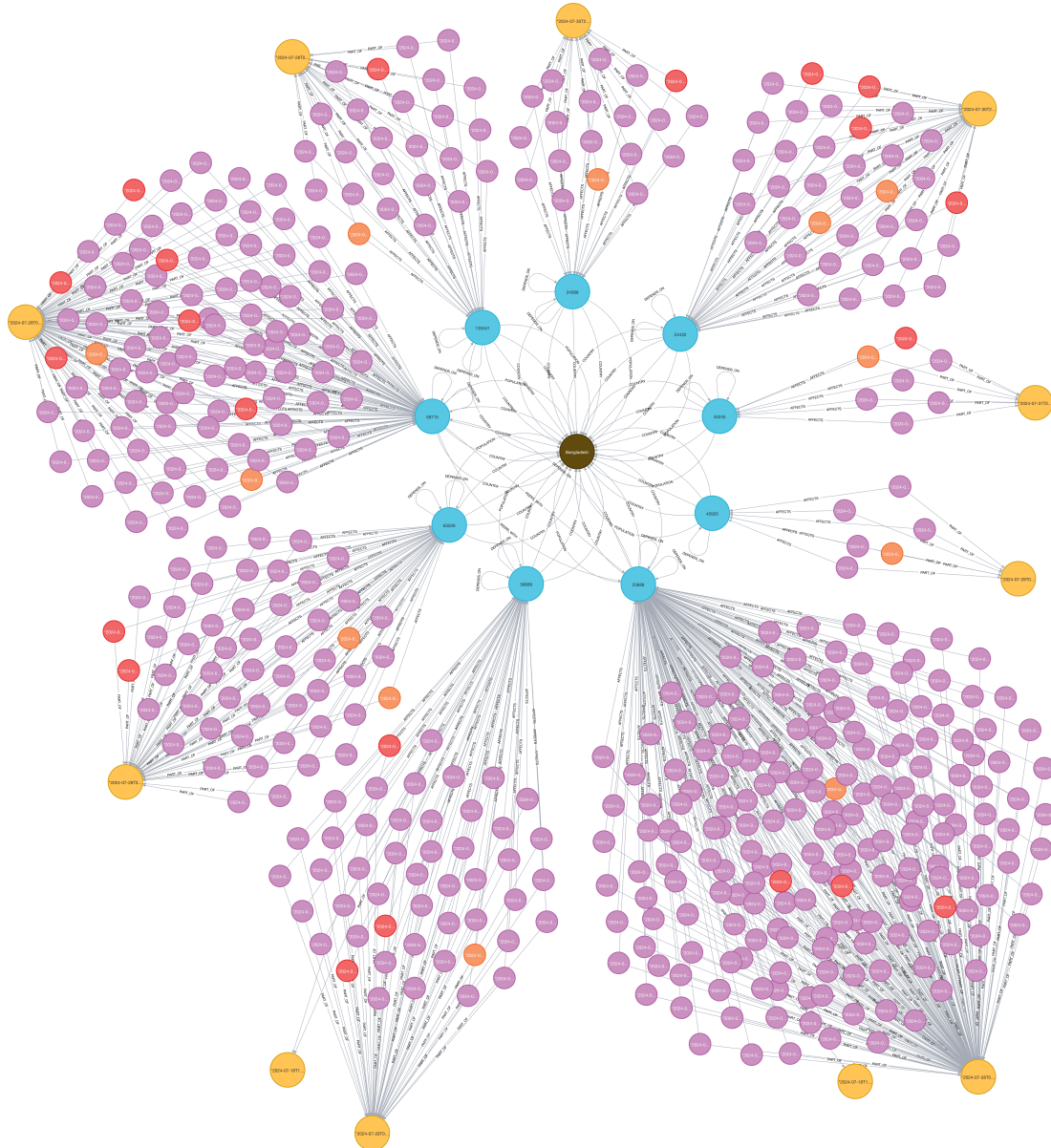


Fig. 5: Topological overview of the Bangladeshi Internet shutdown, showing ASes (cyan) of the four mobile network operators and the five largest ISPs affected by Outages (yellow). Each outage results from merged AS-level Outage Events (red, orange, purple). All ASes are arranged around the Bangladesh country node (brown), ordered as in Table VI (starting top center, clockwise).

TABLE VI: **Bangladesh Internet Shutdown (July 2024): affected ASes, operators, services, outage intervals, and dataset coverage.**

AS	ISP	Network	Outage Start	Outage End	ANT	CROC	IODA
24389	GrameenPhone	Cellular	17.07. 19:30	30.07. 22:00	x	x	x
24432	Robi-Axiata	Cellular	18.07. 00:30	30.07. 22:00	x	x	x
45245	Banglalink	Cellular	17.07. 20:15	31.07. 00:00	x	x	x
45925	Teletalk	Cellular	18.07. 14:45	28.07. 09:00	x	x	-
23688	Link3	Broadband	18.07. 11:44	18.07. 11:55	x	-	-
			18.07. 12:45	28.07. 09:00	x	x	x
58689	ICC	Broadband	18.07. 11:35	18.07. 11:50	x	-	x
			18.07. 13:40	29.07. 09:31	x	x	x
58715	Earth	Broadband	18.07. 12:00	28.07. 09:00	x	x	x
63526	SSD-Tech	Broadband	18.07. 14:20	28.07. 09:00	x	x	x
135341	Orange Comm.	Broadband	18.07. 11:20	28.07. 09:00	x	x	x

ASes: The country of Bangladesh (represented by the brown node) appears as a common link connecting all nine ASes (shown in cyan). Each AS has `OutageEvent` nodes linked to it, originating from different data sources. The color of the `OutageEvent` nodes indicates their origin. ANT nodes are indicated in purple, CROC nodes in orange, IODA nodes in red. Finally, the merging algorithm merged each AS’s `OutageEvent` nodes into one or more yellow `Outage` nodes, depending on their temporal overlap. Table VI provides an overview of the merged `Outages`, including their time window, AS affiliation, and the source of their `OutageEvents`. For all nine ASes (in Table VI), the merged view confirms a multi-day outage affecting both mobile and broadband operators. The merged timeline reveals an earlier onset of disruptions for some mobile networks and a longer duration of disruption than initially reported publicly. This finding is consistent with independent OONI measurements published following our analysis [9]. The case study also illustrates a limitation: merging can inflate outage durations, when single, long-tail measurements or partial network failures are included. Moreover, events from unrelated causes such as a BGP misconfiguration and a cable cut that happen to coincide in time for the same AS will be merged into a single outage node, potentially obscuring distinct incidents. The merged dataset consistently confirmed outages across at least two datasets for each AS (see Fig. 5), visualizing cross-dataset confirmation for every `Outage` node in the graph, and extending the known shutdown duration beyond any single source.

On March 14, 2024, multiple submarine cables suffered failures off the coast of Côte d’Ivoire (Africa Coast to Europe (ACE), Submarine Atlantic 3/West Africa Submarine Cable (SAT-3), West Africa Cable System (WACS), and MainOne), causing disruptions in 13 countries [41]. The incident is visible in ANT, CROC and IODA, making it a representative case for infrastructure outages. We queried IYP for outages on March 14 2024 affecting ASes connected to impacted cables and traced resulting events through AS-and `Country-SubmarineCable` relationships. Though 13 countries were initially reported [6], our submarine cable view found 18 affected countries, including France, Portugal, and Spain, linked

by cable landings. This demonstrates the wider impact of the cut and the effectiveness of querying `SubmarineCable` nodes to identify outages.

West Africa Submarine Cable Cut. To focus on major networks, we restricted analysis to rank-1 ASes per country (by population coverage). Table V shows that in most affected countries outages are confirmed by at least two datasets, and in several cases by all three, consistent with external reports [36]. A small number of anomalies were observed, such as earlier apparent onset in Cameroon and transient outages in Burkina Faso and Senegal that were not confirmed across all datasets. These cases highlight sensitivity to false positives, timing noise in individual measurements, and demonstrate that merging can strengthen validation and expose inconsistencies. A detailed analysis of the West Africa Cable Cut confirmed that all reported outages were reliably detected in the measurement datasets, with nearly every outage cross-validated by at least two independent sources. Fig. 6 illustrates the IYP deduction chain for this case study, tracing submarine cable infrastructure through to confirmed outage events.

V. DISCUSSION AND CONCLUSION

This section reviews how the results address the three initial research questions and highlights contributions toward a centralized platform for tracking Internet censorship and outages.

RQ1: *Can heterogeneous Internet outage datasets be merged qualitatively and reproducibly using their existing features, and what challenges arise?* We present a straightforward merging method aligning outages by time and AS, requiring minimal preprocessing besides spatial harmonization of nationwide events in CROC (see Section II-B). While assuming nationwide outages affect all ASes in a country may overestimate impact, this heuristic enabled qualitative, reproducible merging. The temporal dimension introduced challenges regarding inflation of outage durations, potentially obscuring distinct events, as seen in Section IV. All datasets except for ANT are publicly accessible; the ANT dataset requires a separate request under a non-disclosure agreement.

RQ2: *Do the datasets cross-validate each other?* Major disruptions show strong cross-validation across merged outage nodes, with `OutageEvents` from different datasets overlapping in time and AS. However, isolated single-source events mainly from the ping based ANT dataset occurred, especially at outage boundaries, possibly reflecting partial disruptions or false positives due to elevated error rates. Table III shows ANT and IODA are largely complementary, with CROC having the highest overlap proportion due to manual curation. Overall, overlap among all three remains low but sufficient for insightful threefold validation. CROC is the most robust source, with nearly half its events confirmed by others. Low overlap between ANT and IODA indicates either high false positives or differing measurement sensitivities. Future merging strategies and dataset integration must improve coverage and validation. Nonetheless, the naive merging method reveals new insights for major incidents. A systematic cross-validation assessing

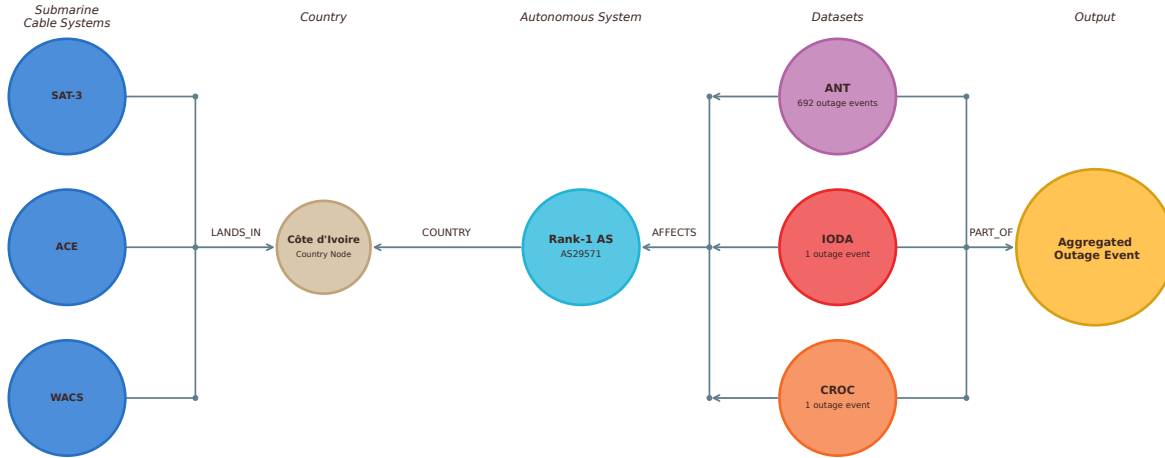


Fig. 6: **Deduction chain from submarine cable infrastructure to confirmed outage, using Côte d’Ivoire’s AS as an example. Following LANDS_IN → COUNTRY → AFFECTS → PART_OF, three cables (SAT-3, ACE, WACS) can be traced to 692 ANT (purple), 1 IODA (red), and 1 CROC (orange) outage events, all merged into a single Outage node (yellow), demonstrating IYP’s ability to link physical infrastructure failures to network-level disruptions.**

the overlap between the outage events captured by our data sources, ANT, CROC, and IODA, and the broader set of Internet shutdown events documented in Internet Society Pulse has not been conducted in this work. Such an analysis could quantify how representative our combined detection pipeline is with respect to a wider corpus of known events, and would complement the per-source validation presented here. We leave this broader cross-validation as a direction for future work.

RQ3: Does a topology-based approach aid Internet outage analysis? Leveraging IYP topology data demonstrated value in assessing large-scale disruptions involving numerous networks, infrastructure, and nations. This approach reduces data volume by focusing on key networks and links entity relations for a holistic outage view. IYP’s queryable database and visualization tools simplify and accelerate complex data interpretation, opening new research avenues for visual-topological outage exploration. Challenges include the country-to-AS mapping’s coarse granularity, which may misattribute ASes geographically. We approximate relevant ASes by population coverage. Submarine cable data was simplified to country-level associations due to lacking precise landing point info, complicated by dynamic routing [42]. IYP tooling limitations include lack of longitudinal mode and high overhead importing snapshot topologies, slowing multi-date analyses. These issues should be addressed in future work. Despite this, the tooling effectively enables intuitive topological data exploration, providing a strong foundation. We conclude heterogeneous outage datasets can be merged with minimal preprocessing. Cross-validation is strongest for large incidents, and topology-based integration reduces complexity while revealing structural biases. This is the first study merging ANT, CROC, and IODA, contextualized with submarine cable infrastructure. We position this work as a first step toward a unified Internet outage

reporting framework. Our approach can process outage data spanning over a year, is queryable with the Cypher language, enriches outage data with IYP’s topology and automatically visualizes results. This allows researchers to conduct their own studies on the same, other, or even yet unknown events.

VI. ETHICAL CONSIDERATIONS

This study builds on publicly available or requestable datasets (ANT, CROC, IODA) and collects no personally identifiable or user-level data. The analysis is limited to infrastructure-level events. Heuristics, such as mapping nationwide outages to all ASes of a country, may lead to overestimation or misattribution, however these limitations are documented to ensure transparency. Given the political sensitivity of some outage measurements, such as government-ordered shutdowns, we have taken care to avoid attributing responsibility to individuals or organizations and to frame results in a strictly technical context. This work follows established community norms for Internet measurement research and does not introduce additional risks to individuals or operators beyond those inherent in the source datasets.

REFERENCES

- [1] A. Now, Z. Rosson, F. Anthonio, and C. Tackett, “KeepItOn-2024-Internet-Shutdowns-Annual-Report,” Feb. 2025. [Online]. Available: <https://www.accessnow.org/wp-content/uploads/2025/02/KeepItOn-2024-Internet-Shutdowns-Annual-Report.pdf>
- [2] Insikt Group, “The Escalating Global Risk Environment for Submarine Cables,” Recorded Future, Inc., Tech. Rep., Jun. 2023. [Online]. Available: <https://go.recordedfuture.com/hubfs/reports/ta-2023-0627.pdf>
- [3] N. Potter, “NATO Plans an Orbital Backup Internet Using Satellite Broadband - IEEE Spectrum,” Dec. 2024. [Online]. Available: <https://spectrum.ieee.org/undersea-internet-cables-nato>
- [4] M. Anderson, “Protecting Undersea Internet Cables: A Tech Challenge - IEEE Spectrum,” May 2024. [Online]. Available: <https://spectrum.ieee.org/undersea-internet-cables-protection-tech>

- [5] Internet Society, “Shutdowns,” Jul. 2025. [Online]. Available: <https://pulse.internetsociety.org/en/shutdowns/>
- [6] J. Tomé and D. Belson, “Undersea cable failures cause Internet disruptions for multiple African countries,” Mar. 2024. [Online]. Available: <https://blog.cloudflare.com/undersea-cable-failures-cause-internet-disruptions-across-africa-march-14-2024/>
- [7] OONI, “Internet Outage Detection & Analysis (IODA),” 2025. [Online]. Available: <https://ooni.org/partners/ioda/>
- [8] A. Master, “Modeling and Characterization of Internet Censorship Technologies,” PhD Thesis, Purdue University, 2023.
- [9] T. I. Raso, S. Afrin, M. A. Chowdhury, M. Xynou, and E. Yachmeneva, “The Longest Silence: Internet Shutdowns During Bangladesh’s 2024 Uprising,” Jul. 2025. [Online]. Available: <https://ooni.org/post/2025-bangladesh-report/>
- [10] C. Azevedo, “Gone offline: how Cloudflare Radar detects Internet outages,” Sep. 2023. [Online]. Available: <https://blog.cloudflare.com/detecting-internet-outages/>
- [11] R. Fontugne, M. Tashiro, R. Sommese, M. Jonker, Z. S. Bischof, and E. Aben, “The Wisdom of the Measurement Crowd: Building the Internet Yellow Pages a Knowledge Graph for the Internet,” in *Proceedings of the 2024 ACM on Internet Measurement Conference*. Madrid Spain: ACM, Nov. 2024, pp. 183–198. [Online]. Available: <https://dl.acm.org/doi/10.1145/3646547.3688444>
- [12] G. Aceto, A. Botta, P. Marchetta, V. Persico, and A. Pescapé, “A comprehensive survey on internet outages,” *Journal of Network and Computer Applications*, vol. 113, pp. 36–63, Jul. 2018, publisher: Elsevier BV. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S1084804518301139>
- [13] L. Quan, J. Heidemann, and Y. Pradkin, “Trinocular: understanding internet reliability through adaptive probing,” in *Proceedings of the ACM SIGCOMM 2013 conference on SIGCOMM*. Hong Kong China: ACM, Aug. 2013, pp. 255–266. [Online]. Available: <https://dl.acm.org/doi/10.1145/2486001.2486017>
- [14] M. Syamkumar, Y. Gullapalli, W. Tang, P. Barford, and J. Sommers, “BigBen: Telemetry Processing for Internet-Wide Event Monitoring,” *IEEE Transactions on Network and Service Management*, vol. 19, no. 3, pp. 2625–2638, Sep. 2022. [Online]. Available: <https://ieeexplore.ieee.org/document/9800915>
- [15] A. Shah, R. Fontugne, E. Aben, C. Pelsner, and R. Bush, “Disco: Fast, good, and cheap outage detection,” in *2017 Network Traffic Measurement and Analysis Conference (TMA)*. Dublin, Ireland: IEEE, Jun. 2017, pp. 1–9. [Online]. Available: <https://ieeexplore.ieee.org/document/8002902>
- [16] R. Bogutz, Y. Pradkin, and J. Heidemann, “Identifying Important Internet Outages,” in *2019 IEEE International Conference on Big Data (Big Data)*. Los Angeles, CA, USA: IEEE, Dec. 2019, pp. 3002–3007. [Online]. Available: <https://ieeexplore.ieee.org/document/9006537/authors>
- [17] G. Aceto and A. Pescapé, “Internet Censorship detection: A survey,” *Computer Networks*, vol. 83, pp. 381–421, Jun. 2015. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1389128615000948>
- [18] R. S. Raman, A. Stoll, J. Dalek, R. Ramesh, W. Scott, and R. Ensafi, “Measuring the Deployment of Network Censorship Filters at Global Scale,” in *NDSS*. San Diego, CA, USA: NDSS, 2020. [Online]. Available: <https://www.ndss-symposium.org/wp-content/uploads/2020/02/23099.pdf>
- [19] E. Tsai, R. S. Raman, A. Prakash, and R. Ensafi, “Modeling and Detecting Internet Censorship Events,” in *Proceedings of the The Internet Society Network and Distributed System Security Symposium (NDSS)*. San Diego, CA, USA: NDSS, 2024. [Online]. Available: <https://www.ndss-symposium.org/wp-content/uploads/2024-409-paper.pdf>
- [20] L. Jin, S. Hao, H. Wang, and C. Cotton, “Understanding the Practices of Global Censorship through Accurate, End-to-End Measurements,” *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, vol. 5, no. 3, pp. 1–25, Dec. 2021. [Online]. Available: <https://dl.acm.org/doi/10.1145/3491055>
- [21] Z. S. Bischof, K. Pitcher, E. Carisimo, A. Meng, R. Bezerra Nunes, R. Padmanabhan, M. E. Roberts, A. C. Snoeren, and A. Dainotti, “Destination Unreachable: Characterizing Internet Outages and Shutdowns,” in *Proceedings of the ACM SIGCOMM 2023 Conference*. New York NY USA: ACM, Sep. 2023, pp. 608–621. [Online]. Available: <https://dl.acm.org/doi/10.1145/3603269.3604883>
- [22] A. Master and C. Garman, “A Worldwide View of Nation-state Internet Censorship,” *Free and Open Communications on the Internet*, no. Free and Open Communications on the Internet 2023(1), pp. 1–21, 2023. [Online]. Available: <https://petsymposium.org/foci/2023/foci-2023-0008.php>
- [23] OONI, “OONI Findings,” 2025. [Online]. Available: <https://explorer.ooni.org/findings>
- [24] —, “OONI: Open Observatory of Network Interference,” 2025. [Online]. Available: <https://ooni.org/>
- [25] R. Sundara Raman, P. Shenoy, K. Kohls, and R. Ensafi, “Censored Planet: An Internet-wide, Longitudinal Censorship Observatory,” in *Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security*. Virtual Event USA: ACM, Oct. 2020, pp. 49–66. [Online]. Available: <https://dl.acm.org/doi/10.1145/3372297.3417883>
- [26] Censored Planet, “Welcome to Censored Planet Observatory’s documentation! — Censored Planet Observatory 2.0 documentation,” 2021. [Online]. Available: <https://docs.censoredplanet.org/>
- [27] —, “Censored Planet,” 2025. [Online]. Available: <https://censoredplanet.org/research/censorship-detection>
- [28] Internet Society, “Internet Society Pulse,” 2025. [Online]. Available: <https://pulse.internetsociety.org/>
- [29] Access Now, “#KeepItOn: fighting internet shutdowns around the world,” 2025. [Online]. Available: <https://www.accessnow.org/campaign/keepiton/>
- [30] ANT, “ANT Internet Outage Datasets,” 2025. [Online]. Available: <https://ant.isi.edu/datasets/outage/index.html>
- [31] OONI, “OONI Explorer - Open Data on Internet Censorship Worldwide,” 2025. [Online]. Available: <https://explorer.ooni.org/>
- [32] CAIDA, “Internet Outage Detection and Analysis (IODA),” Aug. 2016. [Online]. Available: <https://www.caida.org/projects/ioda/>
- [33] IODA, “IODA,” 2025. [Online]. Available: <https://ioda.inetintel.cc.gatech.edu/>
- [34] TeleGeography, “Submarine Cable Map,” Jul. 2025. [Online]. Available: <https://www.submarinecablemap.com/>
- [35] D. Belson and S. Chiang, “The status page the Internet needs: Cloudflare Radar Outage Center,” Sep. 2022. [Online]. Available: <https://blog.cloudflare.com/announcing-cloudflare-radar-outage-center/>
- [36] I. Society, “2024 West Africa Submarine Cable Outage Report,” 2024. [Online]. Available: <https://www.internetsociety.org/resources/doc/2024/2024-west-africa-submarine-cable-outage-report/>
- [37] D. Welle, “Bangladesh: Protesters set light to state broadcaster,” Jul. 2024. [Online]. Available: <https://www.dw.com/en/bangladesh-protesters-set-light-to-state-broadcaster/a-69700479>
- [38] T. I. Project, “TYP Dump 2024/07/15,” Jul. 2024. [Online]. Available: <https://archive.ihrlive.ihrtyp/2024/07/15/>
- [39] M. Frackiewicz, “Wi-Fi, Wires & the Sky: The Full Picture of Internet Access in Bangladesh,” May 2025. [Online]. Available: <https://ts2.tech/en/wi-fi-wires-the-sky-the-full-picture-of-internet-access-in-bangladesh/>
- [40] W. I. Conference, “Bangladesh internet users in 2023 top 131 mln,” Feb. 2024. [Online]. Available: https://www.wicinternet.org/2024-02/18/c_990134.htm?utm_source=ts2.tech
- [41] K. Chege, “What Can We Learn From Africa’s Multiple Submarine Cable Outages?” Aug. 2024. [Online]. Available: <https://pulse.internetsociety.org/blog/what-can-we-learn-from-africas-multiple-submarine-cable-outages>
- [42] M. Weaver, P. Barford, F. E. Bustamante, E. Carisimo, W. Wu, and L. Stokes, “Unsteady Underwater - On the Constancy of Submarine Path Properties,” in *Proceedings of IFIP Networking Conference*. Limassol, Cyprus: IFIP, May 2025, p. 1. [Online]. Available: <https://pages.cs.wisc.edu/~pb/unsteady25.pdf>

VII. CYPHER QUERIES FOR DATA ANALYSIS

```
MATCH (o:Outage) <-[:PART_OF]-(:OutageEvent)-[:AFFECTS]->(a:AS),
      (a)-[:COUNTRY]->(c:Country)
WITH DISTINCT c.country_code as country, count(DISTINCT o) as outages
RETURN country, outages;
```

Listing 2: Cypher query calculating the number of outages per country (Table IV)

```

MATCH (o:Outage)-[:PART_OF]-(oe:OutageEvent)-[:AFFECTS]->(a:AS)
OPTIONAL MATCH (a)-[:COUNTRY]-(c:Country)
RETURN a.asn, c.name, count(distinct(o)) AS outage_count
ORDER BY outage_count DESC
LIMIT 10

```

Listing 3: Cypher query calculating the number of outages per AS, serving as basis for Table IV

```

MATCH (o:Outage)-[:PART_OF]-(oe:OutageEvent)-[:AFFECTS]->(a:AS),
(a)-[:COUNTRY]-(c:Country)
WITH c.country_code AS country, oe.source AS source, count(DISTINCT o) AS
outages_per_source
WITH country, source, 10000 * (outages_per_source - 1) / CASE source
WHEN "CROC" THEN 3605
WHEN "IODA" THEN 89101
WHEN "ANT" THEN 66800
END AS normalized_values
RETURN country, avg(normalized_values) as normalized_outages

```

Listing 4: Cypher query calculating the number of outages per country and normalizing them among the source datasets, serving as basis for Fig. 4

Cypher Queries - Nationwide Internet Shutdown Bangladesh July 2024

```

MATCH (a:AS)--(c:Country)
WHERE c.name = "Bangladesh"
RETURN count(DISTINCT a) AS number_of_unique_AS

```

Listing 5: Cypher query evaluating the number of AS affiliated with the country of Bangladesh

```

MATCH (c:Country {name: "Bangladesh"})
MATCH (a:AS)-[:POPULATION]-(c)
WHERE p.rank <= 15
MATCH (a)-[:NAME]-(n:Name)
WITH a, collect(DISTINCT n.name) AS names, min(p.rank) AS rank
RETURN a, rank, names
ORDER BY rank

```

Listing 6: Cypher query evaluating the top 15 AS by population coverage rank in Bangladesh

Cypher Queries - West Africa Submarine Cable Cut March 2024

```

MATCH (o: Outage)-[:PART_OF]-(oe:OutageEvent)-[:AFFECTS]-(as:AS)-[:COUNTRY]-(c:Coun
try)-[:LANDS_IN]-(sc:SubmarineCable)
WHERE sc.slug IN ["africa-coast-to-europe-ace", "sat-3wasc",
"west-africa-cable-system-wacs", "mainone"]
AND oe.startDate >= datetime("2024-03-14T00:00:00Z")
AND oe.startDate < datetime("2024-03-15T00:00:00Z")
RETURN o, oe, as, c, sc

```

Listing 7: Cypher query identifying 14th March 2024 outages related to the West Africa cable cut, affecting AS in countries connected via ACE, SAT-3/WASC, WACS, and MainOne

```

MATCH (o: Outage)-[:PART_OF]-(oe:OutageEvent)-[:AFFECTS]-(as:AS)-[:COUNTRY]-(c:Coun
try)-[:LANDS_IN]-(sc:SubmarineCable)
WHERE sc.slug IN ["africa-coast-to-europe-ace", "sat-3wasc",
"west-africa-cable-system-wacs", "mainone"]
AND oe.startDate >= datetime("2024-03-14T00:00:00Z")
AND oe.startDate < datetime("2024-03-15T00:00:00Z")
RETURN oe.source, count(DISTINCT oe)

```

Listing 8: Cypher query counting distinct outage events by reporting source for the 14 March 2024 West Africa cable cut affecting ACE, SAT-3/WASC, WACS, and MainOne

```

MATCH (o:Outage)-[:PART_OF]-(oe:OutageEvent)-[:AFFECTS]-(as:AS)-[:COUNTRY]-(c:Coun
try)-[:LANDS_IN]-(sc:SubmarineCable)
WHERE oe.startDate >= datetime("2024-03-14T00:00:00Z")
AND oe.startDate < datetime("2024-03-15T00:00:00Z")
AND c.name IN ["Benin", "Burkina Faso", "Cameroon", "Côte d'Ivoire", "Gambia",
"Ghana", "Guinea", "Liberia", "Namibia", "Niger", "Nigeria", "South Africa",
"Togo", "Senegal"]
RETURN o, oe, as, c, sc

```

Listing 9: Cypher query identifying 14 March 2024 outages related to the West Africa Cable Cut, filtered by countries that have been reported to be affected

```

MATCH (o:Outage)-[:PART_OF]-(oe:OutageEvent)-[:AFFECTS]-(as:AS)-[:COUNTRY]-(c:Coun
try)-[:LANDS_IN]-(sc:SubmarineCable)
WHERE oe.startDate >= datetime("2024-03-14T00:00:00Z")
AND oe.startDate < datetime("2024-03-15T00:00:00Z")
AND c.name IN ["Benin", "Burkina Faso", "Cameroon", "Côte d'Ivoire", "Gambia",
"Ghana", "Guinea", "Liberia", "Namibia", "Niger", "Nigeria", "South Africa",
"Togo", "Senegal"]
RETURN oe.source, count(DISTINCT oe)

```

Listing 10: Cypher query counting distinct outage events by dataset source on 14 March 2024 for reported countries

VIII. GLOSSARY

- ACE Africa Coast to Europe
- ANT Analysis of Network Traffic Outages
- AS Autonomous System
- ASN Autonomous System Number
- Banglalink Banglalink Digital Communications Ltd.
- CROC Cloudflare Radar Outage Center
- CVE Common Vulnerabilities and Exposures
- Earth Earth Telecommunication Pvt. Ltd.
- GrameenPhone GrameenPhone Ltd.
- ICC ICC Communication Ltd.
- IODA Internet Outage Detection and Analysis
- IRI Internet Resilience Index
- ISP Internet Service Provider
- IXP Internet Exchange Point
- IYP Internet Yellow Pages
- Link3 Link3 Technologies Ltd.
- OONI Open Observatory of Network Interference
- Orange Comm. Orange Communication
- Robi-Axiata Robi Axiata PLC
- RQ Research Question
- SAT-3 Submarine Atlantic 3/West Africa Submarine Cable
- SSD-Tech SSD Technologies Ltd. - Mother Company of Carnival Internet
- Teletalk Bangladesh Teletalk Public Ltd.
- WACS West Africa Cable System