

Identifying Important Internet Outages

Ryan Bogutz¹, Yuri Pradkin², and John Heidemann²

¹The College of New Jersey; Ewing, NJ

²USC/Information Sciences Institute; Marina del Rey, CA

Abstract—Today, outage detection systems can track outages across the whole IPv4 Internet—millions of networks. However, it becomes difficult to find meaningful, interesting events in this huge dataset, since three months of data can easily include 41 billion observations and millions of outage events. We propose an *outage reporting system* that sifts through this data to find the most *interesting* events. We explore multiple metrics to evaluate “interesting”, reflecting the size and severity of outages. We show that defining interest as the product of size by severity works well, avoiding degenerate cases like complete outages affecting a few people, and apparently large outages that affect only a small fraction of people in an area. We have integrated outage reporting into our existing public website (<https://outage.ant.isi.edu>) with the goal of making near-real-time outage information accessible to the general public. Such data can help answer questions like “what are the most significant outages today?”, “did Florida have major problems in an ongoing hurricane?”, and “are there power outages in Venezuela?”.

I. INTRODUCTION

Network outages, while rare, have tangible consequences in our increasingly connected world. With the advent of the smart home, Internet-of-things (IoT) devices, mobile phones and tablets, as well as voice-over-IP (VoIP) systems, citizens are increasingly reliant on the Internet. As a result, the loss of Internet connectivity can disrupt our daily lives. For instance, a recent routing problem between a customer of Verizon and Cloudflare interfered with millions of users’ access to popular sites like Google, Discord, and Amazon [16]. Outages also occur as the result of political interference, such as in the Egyptian revolution of 2011 [6]. Finally, we have shown that network outages reflect the impact of hurricanes and natural disasters [9], [12]—outages in the Internet can be used to infer the extent of problems in the physical world.

Several different systems today track Internet outages, both globally with active probing [12], globally with passive observations [8]. Some are specifically targeted at weather events with active probing [14]. This data can be of use to the general public, to understand natural disasters or problems with their network; to scientists and policymakers, to improve our Internet; and to network operators, to diagnose problems in networks and plan network improvement. USC makes all their outage data available for research use [1] and operates a Google-maps-style website with a global view and time travel (see <https://outage.ant.isi.edu> [2]).

However, it is hard to make sense of what outages are important, with gigabytes of data collected over multiple years, and an entire world to look at with hundreds of dots indicating potential outages each minute. When browsing the website, it

is easy to miss events lasting a short time, and it is time-consuming to play out days of data. Even with direct access to the data in a database, queries are hard to formulate and it is not clear what to look for.

Our contribution is twofold: we explore *metrics that identify important events* in this voluminous data, finding that the product of event size and severity does a good job of identifying “interesting” events. Second, we use this metric to *provide a daily report of the ten “most interesting” events*. We use this tool to explore a week of outage data. Our tool is available on the public Internet and is integrated into the USC/ISI outage tool at <https://outage.ant.isi.edu> as a sidebar. Our goal is to make outage data accessible to the general public.

II. BACKGROUND AND RELATED WORK

Prior outage studies have focused on methods to detect outages, but relatively little work has considered analysis and visualization. Although in our work we used data specifically from *Trinocular* outage detection system and integrated our reports with *Trinocular* outage maps, we believe our methodology and metrics can be adapted by other outage detection systems that aggregate concurrent outages. Next, we describe some outage detection systems and prior work on visualizing outages.

A. *Trinocular* Outage Detection

Trinocular is an outage detection system that uses Bayesian inference to minimize probing traffic while reliably detecting outages across millions of edge networks [12]. *Trinocular* monitors /24 IPv4 network prefixes (each a set of 256 adjacent IP addresses) or *address blocks*, observing each network once every 11 minutes (a “*Trinocular* round”). Early *Trinocular* processing was batched, with results generated once a quarter. In 2019 we deployed *Near-Real-Time* *Trinocular* providing preliminary results within two hours of an outage. We continue to update the data quarterly with more complete, batch-processed analysis.

Trinocular is generally accurate; it has been found to detect 100% of outages lasting longer than one round [12]. Recent work has used analysis of CDN data to confirm that most blocks are correct, but shown that some blocks report many false outages [13]. We have traced this problem to sparse blocks and shown improvements to address this problem [3].

We have visualized *Trinocular* data in an interactive website since 2017 [2]. For this visualization, each address block is geolocated (with the MaxMind GeoLite City database),

then mapped to a *grid cell* defined by a 0.5-degree latitude/longitude region, or 1 or 2-degree cells when zooming out. The visualization shows the number of networks that are out in each cell by the area of a circle, while the fraction of networks that are out is encoded in the color, ranging from blue to white to red. This website provides the usual interactive map web features (zoom, pan). It supports time travel to any date for which we have data, and can play back the data with an animation to show evolution of outages over time. This visualization thus provides a calibrated, interactive global map of Internet outages, with updates being posted in near-real-time.

This outage detection system and website encompasses an enormous amount of data. Trinocular scans millions of networks (from 3.5M in 2014 to more than 4.2M in 2019), and each quarter generates more than 41B observations. Unfortunately, thousands of small outages happen all the time, and with observation noise, it is easy to lose large events in a background of persistent, small outages.

Our goal is to address this gap, providing reports that identify key outage events in this mass of data, and to integrate these reports into the existing website. The result is to empower users to focus their attention on important events rather than searching for them.

B. Related Outage Detection Systems

A number of other systems also report outage data. Schulman and Spring’s Thunderping [14] was developed concurrently with our system, and focuses on outages of residential networks that occur due to weather events. They have visualized their results with videos, but they do not (to our knowledge) support interactive visualization.

CAIDA has detected outages in passive observations through their network telescope [7]. The algorithms behind their approach have recently been formalized as Chocolate [8]. They visualize the results of this system in IODA, described below.

The website Downtdetector uses crowd-sourced data to provide some information about network and service outages [10]. Unfortunately their exact methodology is proprietary, and the precision of their results is unknown.

C. Outage Visualization

At least three existing systems visualize network outages.

Lin Quan et al. reported early work on non-geographic visualization of Internet outages [11]. While useful for identifying correlated network events, this work targeted network experts and not the general public.

Downtdetector aggregates status reports of varying online services, such as Twitter and Youtube, as well as service providers, such as Comcast and Optimum, and visualizes outages in two ways: a geographic heatmap of reported outages and a histogram of reported outages over time [10]. Downtdetector is useful for describing outages of particular services, but does not provide a broader insight into outage visualization across the whole visible internet.

Internet Outage Detection and Analysis (IODA) is a sophisticated website visualizing several types of outage data collected by CAIDA [5]. It provides three levels of spatial

granularity: country, region, and autonomous system (AS), and it also allows time travel and some types of queries. Their dashboard highlights events based on “Alert Area”, a factor that considers the size of the change by the outage duration and how many detection methods see it. In comparison, our website complements this work by providing a more targeted geolocation (grid cells of latitude/longitude rather than country or region), and uses different metrics to highlight what is considered important. We do not use area alerts because our outages consist of many blocks which often have different start and end times. Future work may evaluate our metrics against their groupings, and compare their alert area over our data.

III. METHOD

A. Problem Statement and System Overview

Our goal is to find important outages. To reach this goal, we need to understand what makes an outage an important one. We seek to define a metric to reflect the importance of each outage, allowing us to prioritize more important ones over less important.

We consider outages important when they affect many people in a noticeable way. We consider three factors as part of interest: size, severity, and change. The *size* of the outage is how many people are affected. The more people are affected, the bigger is the problem. The *severity* is defined as the fraction of networks in the area that have problems. When *everyone* in an area loses Internet access, this may indicate that something significant is going on, such as a complete power outage or devastation by a hurricane, whereas if only some networks are affected, it could mean only one ISP may be experiencing problems. Finally, *rate of change* is important to consider. Outages that last for hours or days (perhaps due to the time required to physically restore downed utility lines) are important to highlight when they occur, but less important a day later. Measurement error or shifts in ISP use sometimes also result in outages that persist for days or weeks. We consider such events of lesser interest.

Our system generates outage reports (§III-B) that highlight important outages for a given day based on a specific metric to rank all outages on that day (§III-C). We have integrated the reports into our public website, and we also have a dedicated report generation page for experiments (§III-D).

Report generation builds on our existing outage detection system [12] for archival data, and our new, near-real-time implementation for data in the last quarter. In either case, all outage data is loaded into a database.

B. Generating Outage Reports

Report generation draws upon outage data that is stored in a database. This data comes from near-real-time Trinocular (for data in the last quarter), with updates every 15 minutes, or from batch-processed Trinocular [12] for older data. The database contains over 660M data points collected from 4M networks over three years, and records the number of measured and currently out networks in each geographic grid cell. (Grid cells are defined by 0.5, 1, or 2-degree latitude/longitude squares.)

A report is produced by running an SQL query against this database. In the next section we describe the several

TABLE I
IMPORTANCE METRICS TO RATE OUTAGES. WE CONSIDER BOTH ABSOLUTE VALUES, AND THE *change* IN THESE VALUES RELATIVE TO 22 MINUTES AGO.

size	number of networks in each cell
severity	fraction of networks out in the cell
interest-1	size \times severity
interest-2	size \times severity ²

different metrics to evaluate which outages over the day are most important. Each metric value is computed at runtime (using a simple SQL join). We plan to create a caching system to decrease the runtime of a query from several seconds to a fraction of a second. To provide context for outage locations, we add geographic place names for each grid cell. These place names were pre-computed from an on-line database of largest cities [15]; we pre-compute one for each grid cell, so place-name-look up is a database join and need not be done on-line.

In principle, reports can cover any time period and any subset of the globe. Our stand-alone report page supports several standard time periods (12 hours, 1 day, and 1 week) and geographic regions (global, country and continent).

In addition to generated reports, we also support a pre-computed list of major historical outages on many different days. This list is stored separately in the database (for easy update), but events are added manually. We currently include major outages such as hurricanes (Harvey, Irma, Maria from 2017), service outages (such as the 2014 Time Warner outage), and similar large events.

C. Metrics for Outage Prioritization

To identify the most important outages in each report, we rank all outages on the day by some *importance metric*. The report displays only the top ten events by that rank. In addition to ranking metrics, we report only one event for each grid cell per reporting period, so that a long-term problem in one location does not consume all ten spots.

Table I shows the metrics we consider: size, severity, and two versions of “interest”. In addition to considering these values in absolute terms, we also looked at the *change* in each value relative to 22 minutes in the past. For example, we consider both size, and change-in-size (number of networks out now minus those out 22 minutes ago).

Size and severity directly reflect the information in the database, following the number or percent of networks out in each geo-cell. However, we found that both of these metrics often over-emphasize unimportant outages. *Size* emphasizes large outages, but our geolocation database (MaxMind) places all networks for each country at one specific location when it is uncertain about the city-level location of that network. As a result, most countries have many, many networks in one place (in the United States, this is in in Kansas). These artificial “hot spots” may have large numbers of unreachable networks, often for long periods of time, yet they reflect artifacts of geolocation and outage measurement more than actual problems.

Severity is an alternate metric, selecting grid cells where most or all networks are down at some time. Severity over-emphasizes unimportant events for the opposite reason as size:

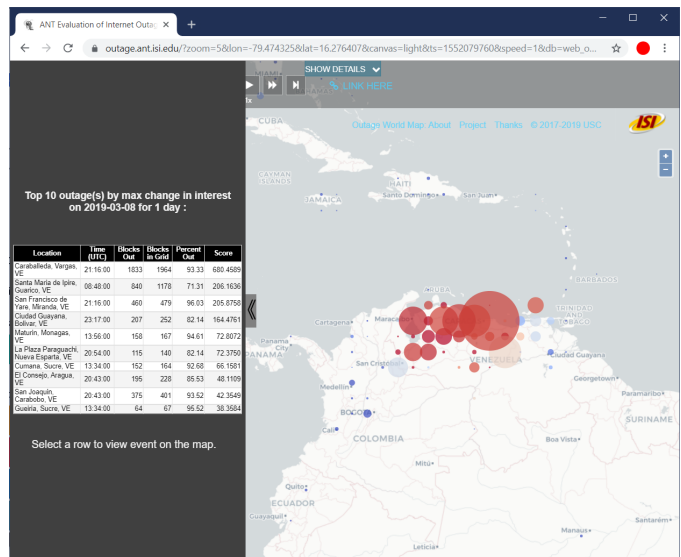


Fig. 1. Our public-facing outage website with the report sidebar expanded, showing an important Venezuelan outage.

severity always prefers cells where all (or almost all) networks are out. There are many grid cells which happen to have 20 or fewer /24 networks. With so few networks, loss of a few networks greatly changes severity. And in these sparsely networked areas there is often only one provider, so these networks may all fail together. Thus use of the severity metric tends to select networks in sparsely populated areas (with few networks) that happen to have an unlucky event.

Our final metrics we collectively call *interest*, and they reflect the product of size and severity (interest-1) or size and severity-squared (interest-2). Our goal in ranking by the product is to allow *both* factors (size and severity) to play a role, so we may rank some large outages that effect only a few people the region, or smaller events that effect most people. Interest-2 metric puts more emphasis on severity than interest-1.

We consider the *change* version of these metrics to emphasize *shifts* in outages as more interesting than static outages. *Change-size* is the difference in size now compared to 22 minutes ago, and *change-severity* is similar. *Change-interest-1* and *change-interest-2* look compute the base metrics and then calculate the difference between now and the metric 22 minutes ago.

We prefer change-interest-1, since it finding important outage per day while avoiding the degenerate cases of size and severity (alone). We found change-interest-2 to perform similarly in our evaluation, but we prefer -1 out of concern that squaring the severity may favor 80% outages in small regions over 20% outages in other regions that affected many more people. We evaluate these metrics in detail in §IV-A.

D. Website Integration

We have integrated reports into the existing outage website, and we also provide a stand-alone reports web page to support our evaluation.

Figure 1 shows our public-facing website with the report sidebar expanded. Website visitors select the date using date

Rank	Location	Coordinates	Time (UTC)	Blocks Out	Blocks in Grid	Percent Out	Score
1	Caraballeda, Vargas, VE	(10.75, -66.75)	2019-03-07 21:26:00	1263	1583	79.79	881.3550
2	Duaca, Lara, VE	(10.25, -69.25)	2019-03-07 21:26:00	407	464	87.72	270.3072
3	Bejuma, Carabobo, VE	(10.25, -68.25)	2019-03-07 21:48:00	532	600	88.67	260.5044
4	San Francisco de Yare, Miranda, VE	(10.25, -66.75)	2019-03-09 16:31:00	448	476	94.12	253.1740
5	Santa Maria de Ipire, Guanico, VE	(8.25, -65.75)	2019-03-08 08:48:00	840	1178	71.31	206.1636
6	Concepcion, Zulia, VE	(10.75, -71.75)	2019-03-07 13:33:00	297	388	76.55	181.8535
7	Ciudad Guayana, Bolivar, VE	(8.25, -62.75)	2019-03-08 23:17:00	207	252	82.14	164.4761
8	Cave Springs, Arkansas, US	(36.25, -94.25)	2019-03-07 18:52:00	323	668	48.35	140.8372
9	San Joaquin, Carabobo, VE	(10.25, -67.75)	2019-03-07 21:37:00	275	364	75.55	140.5539
10	Fechain, Nord-Pas-de-Calais, FR	(50.25, 3.25)	2019-03-06 23:48:00	290	662	43.81	127.0115

Fig. 2. The report-specific webpage with report options shown at the top.

selector control; expanding the sidebar (by clicking on the chevron-marked tab) generates the report. Clicking on any line of that report causes the map to geographically recenter on that outage and shift time to the its start. The public website supports only our best metric (change-interest-1), it ranks outages only for the currently selected day and for the whole world. In the future we plan to add an option for historical queries and for narrowing searches to a fixed list of regions.

To support our evaluation of different metrics, time periods, and regions, we also provide a web page that is dedicated to reports and includes additional controls to select each different aspect of the report. Figure 2 shows one such report. Here, clicking on a row opens up a new tab showing the event.

IV. RESULTS

We next evaluate our report generation system, comparing different metrics of importance (§IV-A). To make these results concrete we examine one week of outages, from 3 March 2019 to 10 March 2019. We examine many important events that we find (§IV-B) and determine who is affected by them (§IV-D).

A. How do ranking metrics compare?

We first want to compare the different metrics we proposed to rank outage importance (see Table I).

We would like to know how the ranking metrics compare in ranking the same outage. We believe our preferred interest metric will rank outages of high size and severity greater than the other metrics. We will consider three different outages from the same day with the following characteristics: high size and high severity, small size and high severity, high size and low severity.

Table II, Table III, Table IV and Table V show the “change” version of the metrics for our day of interest (3 March 2019). The non-change metrics are omitted here due to space limitations, but they are available in our technical report [4].

We see that all show the same first outage in India at 9:05Z. However, the second-place outage varies: *size* focuses on a large outage of 2000 networks in Brazil, but affecting only 4% of the networks there. This event is degenerate—that location is the default Brazil location (Duque de Caxias), so there were a large number of outages there because there are many networks for which geolocation is poor. By contrast, the

second *severity* finds is a much smaller place in Brazil (Rio Brilhante) where half the networks failed. This event is the degenerate case for severity—the problem is so severe because that region has only 31 measurable networks in it.

Finally, both *change-interest-1* and *change-interest-2* identify a Russian outage as their second choice, where 52 of 117 networks fail, a 44% outage affecting more than 50 networks. This example shows the interest metric’s ability to find a balance of size and severity. The third choice for interest-1 and -2 shows their difference: change-interest-2 selects a smaller outage in Tunisia (Bin Qirdan, 68 of 107 networks fail for 63%, while change-interest-1 selects a larger but less severe outage in Pakistan (Malir Cantonment with 153 of 1439 networks failing, 11%). (Note that change-interest-2 appears to invert its second and third choices, even though the product of the Russian outage is lower, its *change* is greater.)

These examples show our preference for change-interest-1 at finding many people who are affected while avoiding the degenerate case of the default location in each country.

B. Do the ranking metrics find important outages?

We next confirm that our metrics find interesting outages. From the prior section, we expect the *interest-1* metric will find more interesting outages in its top-ten list each day compared to other rankings that will be distracted by degenerate cases. To evaluate metric success, we next look at the top-ten list for each metric, determine how many events are interesting (high in size and severity, dynamically changing). For the purpose of this experiment, we will define high in size as over 500 blocks, high in severity as over 50%, and dynamically changing as changing in the last 22 minutes on our outage map.

Table VI shows the the number of interesting outages found by each metric on the week of 3 March 2019. We find that the non-interest metrics do not perform well—severity locks on to tiny outages and does not recognize larger ones, and size tends to identify large but static outages. Size and change-interest-2 perform similarly, finding 3 outages.

We also see that the change- metrics consistently do better. Non-change metrics tend to highlight static locations rather than actual changes.

We see that the change-interest-1 and change-interest-2 metrics do better than change-size and change-severity, because they find a balance of size and severity.

Surprisingly, we find change-interest-2 does slightly better than change-interest-1, with 29 over 24 events over the week. We should perhaps reconsider our preference for change-interest-1, although manual examination of specific events causes us to favor it.

C. Can We Confirm Outages?

Although the focus of this paper is not on the accuracy of the outages that are reported, our reports can help verify outages. We next look briefly at the outages we find to confirm how they look in the underlying data. That is, we want to make sure that the outages reported by the tool are correct in terms of the time of occurrence, location, and size. We expect the reporting tool will match the raw data within a threshold.

For this experiment, we look across the largest interesting outages in the week beginning on 3 March 2019. We then

TABLE II
TOP 10 CHANGE-SIZE FOR MARCH 3

Rank	Location	Lat/Lon	Time	Blocks Out	All Blocks	Percent Out	Metric Score
1.	Sriramnagar, Telangana, IN	(17.25, 78.25)	08:54:00	202	980	20.61	195
2.	Duque de Caxias, Rio de Janeiro, BR	(-22.75, -43.25)	07:15:00	2028	48441	4.19	110
3.	Kampung Bukit Tinggi, Pahang, MY	(3.25, 101.75)	16:47:00	116	4802	2.42	100
4.	Malir Cantonment, Sindh, PK	(24.75, 67.25)	06:31:00	153	1439	10.63	67
5.	Amouguer, Meknes-Tafilalet, MA	(32.25, -4.75)	06:31:00	369	21757	1.70	61
6.	Chandler, Arizona, US	(33.25, -111.75)	09:05:00	54	4343	1.24	52
7.	Shahr-e Qods, Tehran, IR	(35.75, 51.25)	09:27:00	131	16675	0.79	51
8.	Wan Chai, Wanchai, HK	(22.25, 114.25)	18:26:00	234	25859	0.90	51
9.	Meiling, Jiangxi Sheng, CN	(28.75, 115.75)	07:59:00	182	4220	4.31	47
10.	Meiling, Jiangxi Sheng, CN	(28.75, 115.75)	07:48:00	170	4210	4.04	47

TABLE III
TOP 10 CHANGE-SEVERITY FOR MARCH 3

Rank	Location	Lat/Lon	Time	Blocks Out	All Blocks	Percent Out	Metric Score
1.	Sriramnagar, Telangana, IN	(17.25, 78.25)	08:54:00	202	980	20.61	0.1995
2.	Rio Brilhante, Mato Grosso do Sul, BR	(-21.75, -54.25)	12:34:00	15	31	48.39	0.1613
3.	Mahbubnagar, Telangana, IN	(16.75, 78.25)	09:05:00	16	54	29.63	0.1407
4.	Raglan, Waikato, NZ	(-37.75, 174.75)	06:09:00	7	8	87.50	0.1250
5.	Raglan, Waikato, NZ	(-37.75, 174.75)	05:58:00	7	8	87.50	0.1250
6.	Jauharabad, Punjab, PK	(32.25, 72.25)	14:24:00	15	22	68.18	0.1166
7.	Nash, Texas, US	(32.25, -96.75)	07:04:00	37	278	13.31	0.1006
8.	Jalalpur, Punjab, PK	(32.75, 74.25)	12:45:00	8	21	38.10	0.0953
9.	Jalalpur, Punjab, PK	(32.75, 74.25)	12:34:00	8	21	38.10	0.0953
10.	Paita, South Province, NC	(-22.25, 166.25)	18:37:00	42	240	17.50	0.0917

TABLE IV
TOP 10 CHANGE-INTEREST-1 FOR MARCH 3

Rank	Location	Lat/Lon	Time	Blocks Out	All Blocks	Percent Out	Metric Score
1.	Sriramnagar, Telangana, IN	(17.25, 78.25)	09:05:00	275	1051	26.17	56.4977
2.	Achinsk, Krasnoyarskiy, RU	(56.25, 90.25)	06:09:00	52	117	44.44	22.2646
3.	Malir Cantonment, Sindh, PK	(24.75, 67.25)	06:31:00	153	1439	10.63	11.1211
4.	Ghatkesar, Telangana, IN	(17.25, 78.75)	09:05:00	41	131	31.30	10.1810
5.	Qalyub, Muhafazat al Qalyubiyah, EG	(30.25, 31.25)	00:50:00	947	8176	11.58	9.4791
6.	Duque de Caxias, Rio de Janeiro, BR	(-22.75, -43.25)	07:15:00	2028	48441	4.19	9.0204
7.	Tepusteca, Yoro, HN	(15.25, -86.25)	03:02:00	70	475	14.74	7.3732
8.	Teupasenti, El Paraiso, HN	(14.25, -86.75)	03:02:00	8	10	80.00	6.0667
9.	Behbahan, Khuzestan, IR	(30.75, 50.25)	01:23:00	13	25	52.00	5.9980
10.	Bin Qirdan, Madanin, TN	(33.25, 11.25)	14:24:00	68	107	63.55	5.7731

TABLE V
TOP 10 CHANGE-INTEREST-2 FOR MARCH 3

Rank	Location	Lat/Lon	Time	Blocks Out	All Blocks	Percent Out	Metric Score
1.	Sriramnagar, Telangana, IN	(17.25, 78.25)	09:05:00	275	1051	26.17	16.8058
2.	Achinsk, Krasnoyarskiy, RU	(56.25, 90.25)	06:09:00	52	117	44.44	10.1904
3.	Bin Qirdan, Madanin, TN	(33.25, 11.25)	14:24:00	68	107	63.55	5.2114
4.	Teupasenti, El Paraiso, HN	(14.25, -86.75)	03:02:00	8	10	80.00	5.0089
5.	Trenel, La Pampa, AR	(-35.75, -64.25)	21:11:00	15	17	88.24	4.9294
6.	Manthani, Telangana, IN	(18.75, 79.75)	09:05:00	14	16	87.50	4.6937
7.	Ghatkesar, Telangana, IN	(17.25, 78.75)	09:05:00	41	131	31.30	3.6030
8.	Datian, Guangdong, CN	(22.25, 112.25)	08:21:00	4	4	100.00	3.5000
9.	Madhugiri, Karnataka, IN	(13.75, 77.25)	07:59:00	4	4	100.00	3.5000
10.	Datian, Guangdong, CN	(22.25, 112.25)	08:10:00	4	4	100.00	3.5000

TABLE VI
NUMBER OF INTERESTING OUTAGES PER METRIC PER DAY

day in March:	3	4	5	6	7	8	9	<i>total</i>
size	0	0	0	0	0	0	1	1
change-size	0	0	1	0	5	3	7	16
severity	0	1	0	0	2	0	0	3
change-severity	0	1	1	0	7	4	6	19
interest-1	0	1	3	0	3	1	0	8
change-interest-1	0	0	2	0	8	5	9	24
interest-2	0	1	3	0	3	1	0	8
change-interest-2	3	0	2	0	10	5	9	29

TABLE VII
LIST OF AFFECTED NETWORK PROVIDERS FOR THE CARABALLEDA,
VENEZUELA OUTAGE ON 8 MARCH 2019

<i>Number of Blocks</i>	<i>Network Provider</i>
3210	CANTV Servicios, Venezuela
172	Corporación Telemic C.A.
68	Supercable
40	Universidad Simon Bolivar
37	Net Uno, C.A.
24	Fundación Centro Nacional de Innovación Tecnológica (CENIT)
7	GBLX Global Crossing Ltd.
4	Omnivision C.A.
2	Universidad Pedagógica Experimental Libertador
1	Universidad Católica Andrés Bello

compare the outages that appear in reports with all the outages in the underlying data.

We find that the largest outage for the week was in Brazil, but it is a static outage, which is of little interest to us. The next, more dynamic, largest outage is in Venezuela, affecting 1954 blocks and 98.5% of the grid cell.

Through analysis of raw data, we find that the number of affected blocks in Venezuela listed on the tool (1991) is within 5% of the number of blocks found in the raw data (2065). We also confirm that the raw data recorded an outage at the same location and time. This analysis suggests that the reporting tool reports outages accurately, highlighting events that appear in our raw data.

D. Which network providers are affected?

Finally, outage reports prompt us to look into the actual network operators that encountered problems. Although we do not do that for all outages, we next do that for one case to show how one would use our tool to understand the impact of the Internet outages to some locality.

From our reports, we selected the outage in Caraballeda, Venezuela to examine on 2019-03-08T11:00Z. We selected this outage since it is the largest outage from the week of 3 March that also meets our interest standards. We then extracted the specific networks that went down during this event, within a half hour before and after the given start time, and joined that data with a mapping of IP addresses to Internet Service Providers from public WHOIS databases from the Regional Internet Registries (ARIN, LACNIC, etc.).

Table VII shows the AS most affected for this event. We find that a variety AS are affected by the event, from

telecommunications companies (CANTV Servicios, Corporación Telemic C.A.) to Universities (Universidad Pedagógica Experimental Libertador, Universidad Católica Andrés Bello). We find that CANTV Servicios is affected the most, with 3210 blocks down. Corporación Telemic C.A., the second most affected provider, has a significantly lower number of blocks affected, 172. This suggests the outage event was highly skewed towards one specific AS. Additionally, localized outage data will help understanding the impact of any given outage, and support further examination.

V. CONCLUSIONS

We have evaluated several different metrics to rank Internet outages. We showed that our change-interest metrics find a good balance between outage size and severity. We have used the metrics and the reporting tool to find important outages and successfully analyze a week of data. Finally, we demonstrated that we find interesting outages and can use this tool to target further investigation. Our reports are available on our public website at <https://outage.ant.isi.edu>, and our data is available at no charge to interested researchers.

Acknowledgments: Ryan Bogutz carried out this work at USC/ISI over summer 2019 as part of the ISI Summer Research Experience for Undergraduates program (NSF award #1659886, PI: Jelena Mirkovic). Yuri Pradkin and John Heidemann’s research is in part sponsored by the Department of Homeland Security (DHS) Science and Technology Directorate, Cyber Security Division (DHS S&T/CSD) via contract number 70RSAT18CB0000014, and by the Air Force Research Laboratory under agreement number FA8750-18-2-0280. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon.

REFERENCES

- [1] ANT Project. ANT project outage datasets. <https://ant.isi.edu/datasets/outage/>, Apr. 2013. Datasets updated quarterly since Nov. 2013.
- [2] ANT Project. Ant internet outages interactive map. <https://outage.ant.isi.edu/> and <https://ant.isi.edu/blog/?p=1141>, Dec. 2017.
- [3] G. Baltra and J. Heidemann. Improving the optics of active outage detection (extended). TR ISI-TR-733, USC/ISI, May 2019.
- [4] R. Bogutz, Y. Pradkin, and J. Heidemann. Identifying important internet outages (extended). TR ISI-TR-735, USC/ISI, Oct. 2019.
- [5] CAIDA. IODA. web page <https://ioda.caida.org/ioda>.
- [6] J. Cowie. Egypt leaves the Internet. Renesys Blog <http://www.renesys.com/blog/2011/01/egypt-leaves-the-internet.shtml>, Jan. 2011.
- [7] A. Dainotti, C. Squarcella, E. Aben, M. Chiesa, K. C. Claffy, M. Russo, and A. Pescapé. Analysis of country-wide Internet outages caused by censorship. In *Proc. ACM IMC*, pages 1–18. ACM, Nov. 2011.
- [8] A. Guillot, R. Fontugne, P. Winter, P. Merindol, A. King, A. Dainotti, and C. Pelsser. Chocolate: Outage detection for internet background radiation. In *Proc. IFIP TMA*, Paris, France, June 2019. IFIP.
- [9] J. Heidemann, L. Quan, and Y. Pradkin. A preliminary analysis of network outages during Hurricane Sandy. TR ISI-TR-2008-685b, USC/Information Sciences Institute, Nov. 2012. (correction Feb. 2013).
- [10] Ookla, LLC. Downtetector. web page <https://downtetector.com/>.
- [11] L. Quan, J. Heidemann, and Y. Pradkin. Visualizing sparse Internet events: Network outages and route changes. In *Proc. First ACM Workshop on Internet Visualization*. Springer, Nov. 2012.
- [12] L. Quan, J. Heidemann, and Y. Pradkin. Trinocular: Understanding Internet reliability through adaptive probing. In *Proc. ACM SIGCOMM*, pages 255–266, Hong Kong, China, Aug. 2013. ACM.
- [13] P. Richter, R. Padmanabhan, N. Spring, A. Berger, and D. Clark. Advancing the art of Internet edge outage detection. In *Proc. ACM IMC*. ACM, Oct. 2018.
- [14] A. Schulman and N. Spring. Pingin’ in the rain. In *Proc. ACM IMC*, pages 19–25, Berlin, Germany, Nov. 2011. ACM.
- [15] A. Thampi. reverse_geocoder. web page https://pypi.org/project/reverse_geocoder/.
- [16] L. Voytko. Major outage brings down discord, reddit, amazon and more. *Forbes*.