Detecting Events in the Dynamics of Ego-centered Measurements of the Internet Topology

Assia Hamzaoui, Matthieu Latapy, Clémence Magnien
LIP6 – CNRS and UPMC
104, avenue du Président Kennedy – 75016 Paris – France
Firstname.Lastname@lip6.fr

***Abstract*—Detecting events such as major routing changes or congestions in the dynamics of the internet topology is an important but challenging task. We explore here a *top-down* approach based on a notion of statistically significant events. It consists in identifying statistics which exhibit a homogeneous distribution with outliers, which correspond to events. We apply this approach to ego-centerd measurements of the internet topology (views obtained from a single monitor) and show that it succeeds in detecting meaningful events. Finally, we give some hints for the interpretation of such events in terms of network events.**

I. Introduction

The study of the internet topology as obtained from measurements attracted recently much attention, see *e.g.* [13], [21], [24]. One of the key challenges in this field is nowadays to describe the time evolution of this topology. However, getting information on this is extremely difficult, as our measurement abilities are limited [20] and object's dynamics are very complex [6], [23], [5], [17], [19], [22].

In [19], the authors propose to focus on a part of the topology, which they call an *ego-centered view*. It consists in what a single machine, called *monitor*, may see of the internet topology. It is basically captured by running `traceroute` measurements from the monitor to a given set of randomly chosen destinations, and iterating this process every few minutes. See Section III and [19] for details.

This approach proved to be successful in capturing interesting information [22], [19]. One of the main perspective is to use it to try to detect *events* in the dynamics of ego-centered views, which has not been explored yet. This would lead to insight on the routing (and routing trees) dynamics, and may have important applications in network security (one may monitor a part of the network, or the network around a given machine).

However, previous work has shown that the dynamics of ego-centered views is intense [22], [6]: new nodes are continuously observed, load-balancing introduces frequent changes in paths, etc. Distinguishing *events*, *i.e.* abnormal dynamics in this data is therefore challenging. It may even be impossible, as many *events* may occur at the same time, and throughout the measurement. Conversely, it is not clear that some events are very different from others, and that ego-centered measurements allow to capture some.

We present here the first method to automatically and rigorously detect events in the dynamics of ego-centered views of the internet topology. Intuitively, the internet topology is subject to a substantive dynamics, which can be described as *regular*, and sometimes undergo deep and unusual changes, that we call *events*. For instance load balancing is a *normal* routing dynamics. In contrast, failures in some parts of the network, creating major links and congestions may be seen as *events*.

The *bottom-up* approach is a natural approach to study such events. It consists in the study of the system in detail in order to characterize what normal dynamics and events are expected, and then to search for traces of events. However this approach raises difficult problems: first it is not known what characterizes the dynamics, either normal or abnormal [8]. In addition, there are many different events [10], [18], that cannot necessarily be identified [28]. Finally intuition on the dynamics and its effects often is misleading, as some experiments confirm [22].

The other complementary approach is the *top-down* one. In this approach, one observes the object from the outside by measuring it, as it is done for a living organism for instance. After calculating various statistics, one may then define an event as a statistically significant anomaly. The advantage of this approach is that knowledge of the system is not essential as a first step (but it is necessary for event interpretation). Furthermore, this approach is general, once established, it may be applied to different case studies. Another strong point of this method is that it is rigorous and objective and can help to discover unexpected features of the dynamics. Finally, the *top-down* approach is very suitable for event detection in *complex systems* in the interest, and our goal in this paper is to explore it in the case of ego-centred measurements of the internet topology.

We describe our methodology in Section II. We present the internet topology data which we use in Section III. We apply the method in Sections IV to VI, considering different statistics. Finally, we present some interpretations in Section VII, which show that the detected events are indeed significant from a networking perspective.

II. Methodology

A first possible approach for event detection is bottom-up: from a knowledge of which events may occur on the internet, one may attempt to define statistics to monitor, and which would indicate such events. This supposes a good knowledge of the actual dynamics of the topology, though, and the impact that events have on it. It also means that one has to guess the impact events will have on measurements, although they are very partial and biased views of the whole internet. In addition, correlating these events to statistics may turn out to be very difficult.

It must be clear that the current situation makes this approach extremely difficult. Current knowledge of internet dynamics in general, and events occurring on it in particular is very limited [19], [7], [22], [12], [25], [26]. It is not known which impact routing changes or local failures have on the internet topology. There is no database giving the list of all events occurring on the network (even the events occurring in a single AS are rarely and poorly documented, see Section VII).

Finally, the bottom-up approach, though appealing, is extremely challenging, and it seems out of reach in our current knowledge of the internet topology and its dynamics.

We propose here a top-down approach for event detection in the dynamics of ego-centered views of internet topology. We first consider ego-centered measurements and define simple statistics which may capture key dynamics (Sections IV to VI). We then look for *statistically significant* events, *i.e.* situations in which the observed statistics deviate significantly from the ones usually observed [29]. These will correspond to events, and interpreting them in terms of network events needs further examination (Section VII).

The key point of our method is therefore to be able to define statistics which allow event detection, *i.e.* which exhibit a clear notion of *normal* vs *abnormal* dynamics.

When one considers a statistics, three typical situations may occur:

- First, the observed values may be homogeneous, which means that they all have approximately the same value, and that we never observe a significant deviation from this behavior. In such cases, the statistics is of no help in detecting events: there is never a clear deviation from the *normal* value.
- The observed values may be heterogeneous by nature, which means that there is no *normal* value, and therefore no *abnormal* value indicating events.
- Finally, the observed values may be homogeneous but with some outliers, which indicate abnormal events from a statistical point of view.

Therefore, we are interested in statistics which have a homogeneous behavior, but exhibit some outliers. When studying a statistic, we will observe its dynamics and the distribution of its values (*i.e.* for each *x*, the number of times the statistics has value *x*). We will then try to distinguish between the three kinds of situations above.

Studying such distributions and deciding on their nature however is a subtle task. We will use here two complementary approaches: manual inspection of the distributions by plotting them in various scales, and automatic fit to classical distributions which characterize the three behaviors of interest pointed out above.

In order to perform manual inspection we plot each distribution in lin-lin, lin-log and log-log scales. We moreover plot the inverse cumulative distributions (*i.e.* for each *x*, the number of times the statistics has value greater than *x*) which are often easier to read. The use of these three scales makes it possible to highlight exponential and polynomial decreases, which are hallmarks of homogeneity and heterogeneity respectively. This gives a first mean to gain insight on the distribution and the corresponding statistics.

Once a global shape has been identified in this way, one may try to fit the observed distributions to classical model distributions. One must keep in mind that many models may be considered. Moreover automatic fitting is a difficult task, and results may be misleading [30], [9]. This is why we complement it with manual inspection of plots.

We consider here three typical model distributions, which correspond to the three situations described above: the normal distribution (*i.e.* $P(x) = \frac{1}{\nu\sqrt{2\pi}} e^{\frac{-1}{2}(\frac{x-\mu}{\nu})^2}$) is a typical homogeneous distribution with well defined mean μ and standard deviation ν, which has an exponential decrease; the power-law distribution (*i.e.* $P(x) \sim x^{-\alpha}$) is a typical heterogeneous distribution characterized by its exponent α and which has no *normal* value; and we model homogeneous distributions with outliers by first identifying possible outliers with Grubb's test [14], then removing them and fit the remaining distribution to the normal distribution as above.

We will perform all fits using the classical Maximum Likelihood Estimation (MLE) [11]. Notice however that, in our context, the fit is not the outcome of greatest interest: we are interested in *how much* this fit is relevant. To estimate this, we use one of the widely used goodness of fit called Kolmogorov-Smirnov (KS) test [27].

Finally our event detection method consists in the following steps:

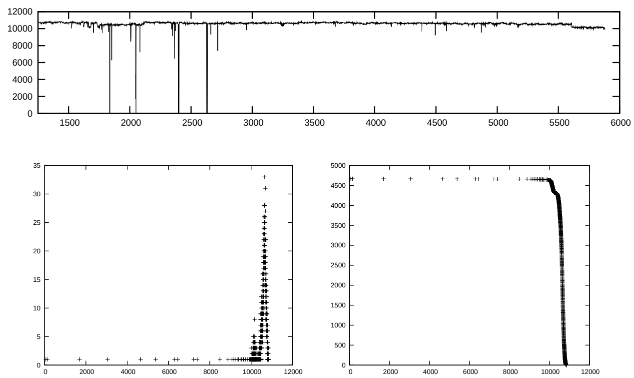
- Define statistics which possibly allow event detection.
- Study the calculated statistics in order to keep those exhibiting homogeneous distributions with outliers. The detected outliers will characterize our statistically meaningful events.
- Explore the detected outliers, with view to interpret them in term of network events.

III. Data

We use here the data described in [19]. It consists in periodic ego-centered measurements of the internet topology: a monitor runs periodic `traceroute` like measurements towards a set of 3000 random targets, waits for 10 minutes, and iterates this (which is called a *round* of measurement). There is approximately four rounds of measurements per hour (approximately 100 per day). Such measurements were performed from more than 100 monitors (mainly PlanetLab machines [2]) and for several months. The obtained datasets are freely available [3]. We computed the statistics on measurements from different monitors, and obtained similar results in all cases. We present in this paper statistics on representative cases.

IV. Number of nodes at each round

The first statistics that one may consider in order to study the dynamics of ego-centered views is naturally the number N_i of nodes observed in each round i of measurement. We display it in Figure 1 (top).



	Goodness of fit
normal distribution	0.82527
normal distribution with outliers	0.91261
power law distribution	0.00217

Fig. 1. Top row: number N_i of nodes observed at each round of measurement, as a function of the number of measurements round i performed. Middle row, left: distribution of N_i; right: inverse cumulative distribution of N_i. Bottom row: goodness of fit of the distribution with the three studied distributions models.

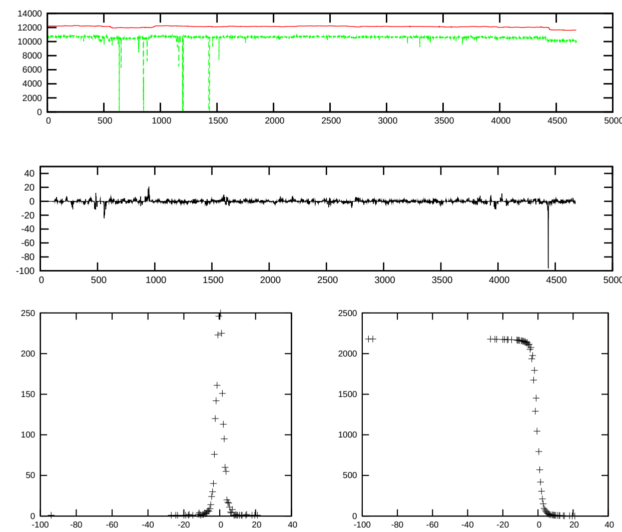
This plot shows that the number of nodes at each round is rather stable with some exceptions. Most of the time, it oscillates close to a mean value slightly above 10 600. However, one may observe that this value changes near rounds 2 100 and 5 600: during some time after these rounds, the number of nodes oscillates close to a different value. In addition to these changes in the average value, the plot also exhibits sharp downward peaks. On the contrary, no upper peak is visible.

These observations are confirmed by the distribution of the value of N_i and the goodness of fit test, see Figure 1. Indeed, the distribution reveals two distinct regimes, with many values around 10 050 and 10 600. Otherwise, the distribution is clearly homogeneous with outliers. The presence of abnormally low values (points on the left) but no abnormally high values corresponds the presence of downward peaks but no upward ones.

It must be clear that downward peaks, although they are very clear statistical outliers, bring little information: they may be caused by local connectivity failures, which have the effect that ego-centered views are (partly) blank during one or a few rounds. This kind of event is trivial.

On the contrary, an upward peak would indicate an interesting event: it would mean that we suddenly observe significantly more nodes at one round. However, there is no upper peak, which is a non-trivial fact: one may easily imagine scenario where such peaks would appear. Figure 1 shows that such scenario do not occur in practice. As a consequence, one cannot detect events by observing abnormally high values of N_i.

Finally, the only notable dynamics in the number N_i of nodes observed at each round are changes in the mean values around which it oscillates. We detect such changes as follows: we associate to each i the median of values N_i to $N_i{+}100$, that we denote by M_i, then we define D_i = M_i - $M_i - 1$. We plot these values in Figure 2. It appears clearly that this method succeed in identifying events, *i.e.* outliers in D_i distribution.



	Goodness of fit
normal distribution	0.76231
normal distribution with outliers	0.98069
power law distribution	0.00241

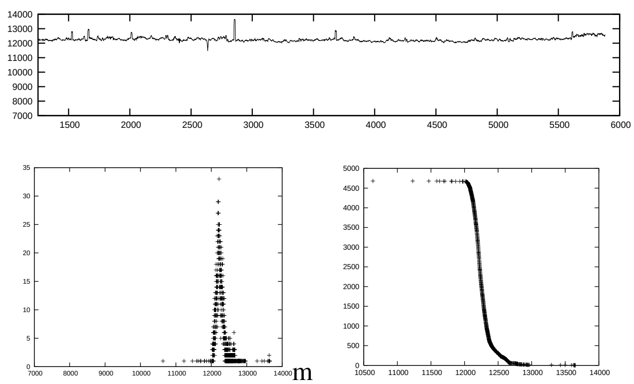
Fig. 2. From top to bottom: N_i and M_i as functions of i (N_i and M_i actually overlap but we shifted up M_i for readability); the variations D_i of M_i, the distribution and the inverse cumulative distribution of D_i; goodness of fit of the distribution with the three studied distributions models.

V. Number of nodes in consecutive rounds

The fact that the number N_i of nodes observed at each round is very stable does *not* mean that the observed nodes are always the same: consecutive rounds may see different ones. Such changes may be evidenced by observing the number N_i^p of distinct nodes in p consecutive rounds, for a given integer p. We display in Figure 3 the case $p = 5$ (top row). This plot shows that, like N_i, N_i^5 is very stable and oscillates around a mean value [1]. As expected, this value is

[1]It also experiences changes of regime, like N_i, for instance around round 5 600 in Figure 3. Notice that, in this case, the new average value for N_i^5 is larger than before, while it was lower for N_i. This means that, although we see less nodes in each round, the nodes we see vary more from one round to another. This gives some hints on further understanding the event which occurred, but deepening this is out of the scope of this paper.

larger than the one for N_i, but it is far from 5 times larger. This shows that many nodes appear in several consecutive rounds. Moreover, upper peaks appear on this plot, which make it very different from the one of N_i in Figure 1. The distribution and the goodness of fit test are presented in this same figure (middle and bottom rows): it confirms the presence of a clear mean value, but also points out clear statistical outliers, both abnormally low (as before) and abnormally high (which is new). This observation is important for event detection: there are specific times (pointed out by the peaks in Figure 3) at which an abnormal number of new nodes appear in a series of consecutive rounds. This gives a new way to detect statistically significant events.

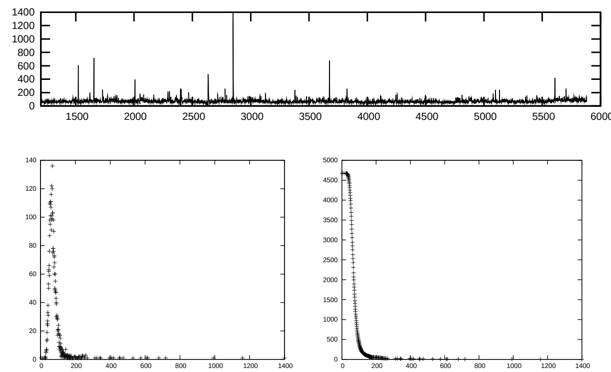


	Goodness of fit
normal distribution	0.61312
normal distribution with outliers	0.92411
power law distribution	0.00408

Fig. 3. Top row: number N_i^5 of distinct nodes observed during five consecutive previous rounds of measurements, as a function of the number of measurements rounds performed. Middle row, left: the distribution of N_i^5; right: the cumulative distribution of N_i^5. Bottom row: goodness of fit of the distribution with the three studied distributions models.

However, automatic detection using this approach is not trivial: as the observed mean value may change during time, and as upper peaks (which we want to detect) may be smaller than these variations of the mean, we may miss some events, and making the difference between a statistically sound event and normal dynamics may be difficult. In order to solve this problem, we observe the number of *appearing nodes*, *i.e.* the number of nodes observed in a series of rounds but not observed in the previous rounds. To do so, we consider two integers p and c (for *previous* and *current*, respectively) and compute for all i the number of distinct nodes which we observe in rounds i to $i+c-1$, but not in rounds $i-p$ to $i-1$, which we denote by a_i. Notice that observing the number d_i of disappearing nodes is also natural. We observed similar results for appearing and disappearing nodes, and so we focus here on appearing nodes. We considered wide ranges of possible values for c and p, and observed little difference, if any, as long as they were greater than 1 or 2 and lower than 100. We illustrate here the obtained results with $p = 10$ and $c = 2$, see Figure 4.

The obtained plot exhibits clear upper peaks, independent of the current mean value of N_i^c, which is confirmed by



	Goodness of fit
normal distribution	0.86490
normal distribution with outliers	0.97263
power law distribution	0.00703

Fig. 4. Top: number a_i of appearing nodes, for all round index i and values $p = 10$ and $c = 2$, as a function of the number of measurements rounds performed. Middle, left: the distribution of a_i; right: inverse cumulative distribution of a_i. Bottom row: goodness of fit of the distribution with the three studied distributions models.

the distributions and the goodness of fit test. We obtain this way a method for automatic detection of statistically meaningful events, by considering outliers of the appearing nodes distribution as event.

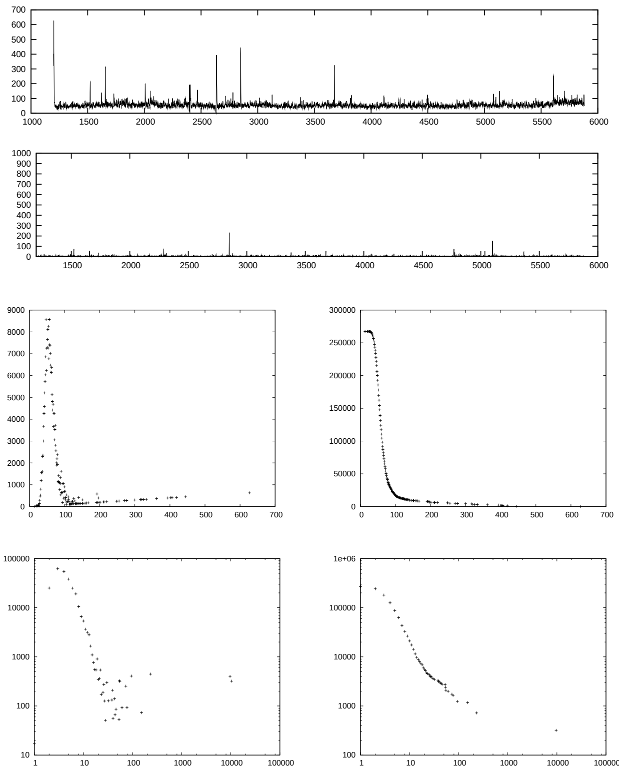
VI. Connected components

We have seen in the previous section that, at some particular moments, an abnormal number of nodes appear in our ego-centered views of the Internet topology. However, we said nothing on their *structure*: are they scattered in the observed topology? are they grouped? or do they belong to several small groups?... Intuitively for instance, an important routing change may lead to the discovery of a new part of the network, which would be revealed by the appearance of nodes forming a connected component in our ego-centered views.

In order to investigate this, we study the connected components of newly appearing nodes. More precisely, for all i, we select the appearing nodes as defined above, and consider links observed between these nodes. We then compute the connected components of this graph, which we call *connected components of appearing nodes*. As in the previous section, we use $p = 10$ and $c = 2$, which give results representative of what we observed on wide ranges of these values.

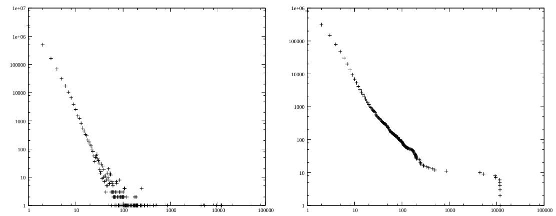
We show in Figure 5 the number of connected components observed for all i, as well as the size of the largest one for all i, together with the distributions of these values and their goodness of fit test. The statistically abnormal events detected with the number of connected components statistics are the same as the ones detected in the previous section. This shows that events detected using the number of appearing nodes are events in which many connected components appear, among which at least a large one.

Observing connected components makes it possible to go further. Indeed, it has the advantage that, at each round, several values are observed: for all i several connected components



	Goodness of fit
normal distribution	0.84923
normal distribution with outliers	0.98033
power law distribution	0.31121
	Goodness of fit
normal distribution	0.61298
normal distribution with outliers	0.58903
power law distribution	0.42118

Fig. 5. From top to bottom: number of connected components of appearing nodes; size of the largest one; distributions of the number of connected components of appearing nodes; inverse cumulative distributions of these last values; distributions of size of the largest connected components of appearing nodes; inverse cumulative distributions of these last values; goodness of fit of the distributions with the three studied distributions models. Here we considered $p = 10$ and $c = 2$



	Goodness of fit
normal distribution	0.21142
normal distribution with outliers	0.18645
power law distribution	0.85318

Fig. 6. Top row, left: the distribution of the size of *all* connected components of appearing nodes observed during the measurements; right: the inverse cumulative distribution of the values on the left. Bottom row: the goodness of fit of the distribution with the three studied distributions models. Here we considered $p = 10$ and $c = 2$

may appear, and we may consider their size. This leads to the distribution of the size of *all* appearing connected components, whichever round they appear in, presented in Figure 6.

This distribution does not exhibit a clear difference between *normal* values and *abnormal* ones, though: the distribution is well fitted by a power-law, as the goodness of fit test on the figure 6 (bottom row) shows. As a consequence, we cannot use it to detect events that would be revealed by the appearance of an abnormally large connected component.

Notice that one may go further by computing various properties of connected components (their density, average degree, or clustering coefficient, for instance), and then observing their distribution. This may lead to the identification of statistically meaningful events. However this is out of the scope of this paper and left for future work.

VII. Towards event interpretation

In previous sections we have presented a methodology and some statistics which make it possible to detect statistically significant events. More precisely, we are able to point out moments in time at which events occur, and to identify nodes and links involved in these events. The ultimate goal of this procedure is to further study detected events and in particular to *interpret* them in term of network events (such as node or link failures, or congestions). This is crucial for a true understanding of internet dynamics and for network monitoring.

Event interpretation is challenging, though, because the current knowledge of the internet dynamics is limited, but also because of the size of the data, its ego-centered (thus biased) nature, and its lack of clear structure. Ideally, one may use a database of events occurring in the internet and match such events to the ones we detect (and conversely). This is not feasible in general, though, as no complete such database exists. Only partial information is available for some specific AS, which we explore in Section VII-A below.

One may also try to interpret detected events by visualizing the data. To do so, graph drawing is appealing, but current methods are unable to handle large graphs and/or produce drawings which are easy to interpret. Some insight may however be obtained this way, and we explore this in Section VII-B.

In the following, we select one of the most interesting statistics for detecting events, the number of appearing nodes in several consecutive rounds (Section V). We apply it to a typical measurement and select events detected this way.

Moreover, we will use a data reduction technique which will be of great help. It consists in focusing on the part of the data involved in the event under concern. To do so we first identify the set E of nodes involved in the event *i.e.* nodes significant appearance's for the chosen statistics. We then select the destinations such that a path from the monitor to the destination contains at least a node in E. Finally we keep only the part of the measurement obtained with these destinations, which is equivalent to measurements conduced with this reduced set of destinations. In this measurement the dynamics of nodes in E is sill visible and most nodes not involved in the event are removed.

A. Correlation with known events

In order to help maintenance and provide better services, some ISP record *events* occurring in their network and document them. This information is partial, poorly structured, and needs manual inspection [16], [15], but this is of great of interest here as it makes it possible to match statistically significant events which we detect to known network events reported in these databases.

Abilene [1] is one the main ISP to provide rich information on events occurring in their network [4]. A database of tickets describing such events is freely available on line; we display typical instances in Figures 8 and 9.

To match a detected event to such a ticket, we proceed as follows. First we select a statistical meaningful event with our method as explained above, and then we localize the the timestamps at which it happens (pointed out with peaks on the corresponding plot). Correlating this event with abnormal event consists in finding in the Abilene database a set of tickets such that the timestamps of these tickets overlap the timestamps of our event, and *affected* field cites elements whose addresses appear in the set E of involved nodes. We therefore have to collect the IP addresses of the elements cited in the ticket in order to check their presence in E.

An example of result is displayed in Figure 7. Among the detected events, two of them are correlated with tickets, and have some particularity. The first statistical event that we pointed out is followed by a significant decrease in the number of appearing nodes. This event is correlated with the trouble ticket of Abilene shown in Figure 8. A second event comes after on the plot, followed by an equivalent significant increase in the number of appearing nodes. Inspecting this second event leads to its correlation with the ticket in Figure 9, which actually is the ticket declaring that the problem cited in the first ticket ended. In this case, thus, there is a perfect fit between the two statistical events under concern and the one depicted in Figures 8 and 9.

```
SUBJECT:       Internet2 IP Network Peer SINET (CHIC) Outage
AFFECTED:      Peer SINET (CHIC)
STATUS:        Unavailable
START TIME:    Thursday, May 17, 2007, 11:47 AM (1147) UTC
END TIME:      Pending
DESCRIPTION:   Peer SINET's connection the Internet2 IP
               Community is unavailable.  SINET Engineers
               have been contacted, however, no cause of
               outage has been provided yet. SINET is multi-homed.
TICKET NO.:    10201:45
TIMESTAMP:     07-05-18 00:40:43 UTC
```

Fig. 8. An example of Abilene trouble ticket which corresponds to the first event pointed out in the Figure 7. It describes a technical intervention under the ticket number 10201:45. The involved network elements are cited in the field AFFECTED. The begin and the end timestamps are given, and details are provided in field DESCRIPTION.

B. Graph drawing

One may also examine a detected event by manipulating a drawing of the underlying graph. Although many drawing methods exist, with different advantages and limitations, in most cases the size of our data is prohibitive. To this regard, being able to identify a moment in time at which an event occurs and focusing on the nodes involved in the event as described above both are crucial: this data reduction leads to graphs of a few thousand nodes, which several software are able to manipulate (and draw).

```
SUBJECT:       Internet2 IP Network Peer SINET (CHIC) Resolved
AFFECTED:      Peer SINET (CHIC)
STATUS:        Available
START TIME:    Thursday, May 17, 2007, 11:47 AM (1147) UTC
END TIME:      Friday, May 18, 2007, 3:51 AM (0351) UTC
DESCRIPTION:   Peer SINET was unavailable to the the Internet2 IP
               Network Community.  SINET Engineers reported the
               reason for outage was due to a fiber cut in New York.
               SINET is multi-homed.
TICKET NO.:    10201:45
TIMESTAMP:     07-05-18 07:39:16 UTC
```

Fig. 9. The trouble ticket corresponding to the second event pointed out in the Figure 7.

One may then draw in different colors the appearing, disappearing and stable nodes and/or links. Figure 10 displays a typical example. Such manual examination of events using graph manipulation software opens the way to a more detailed understanding of detected events, and to their interpretation in terms of network events.

VIII. Conclusion

In this paper we propose and implement a method to automatically and rigorously detect events in the dynamics of ego-centered views of the internet topology. It relies on a notion of statistically significant events. We define simple statistics to do so interestingly, all kinds of distributions are obtained: homogeneous and heterogeneous, which do not lead to the detection of events, and homogeneous with outliers, which do. We also provide approaches to interpret the detected events by drawing them and correlating them to known networking events

Our main perspectives for this work are of course to explore more subtle statistics and to improve event detection. For instance, one may compute the distance between extremities of newly appearing links (before they appear). One may also conduct more case studies to gain insight on the events we observe. In order to help such interpretation, one may simulate ego-centered measurements on a graph with simulated dynamics (random removals/additions of nodes/links, for instance). This would shed light on relation between what we observe with such measurements and actual events, which is crucial in our context. Going further, one may observe events using measurements conduced from several monitors. Some events may be invisible from single monitors, giving complementary views.

References

[1] Abilene backbone network. At http://www.internet2.edu/.
[2] The planetlab platform. At http://www.planet-lab.org/.
[3] Radar datasets. At http://data.complexnetworks.fr/Radar/.
[4] Technical discussion for the internet2 network. At https://listserv.indiana.edu/archives/internet2-ops-l.html.
[5] Réka Albert and Albert-László Barabási. Topology of evolving networks: Local events and universality. *Physical Review Letters*, 85(24):5234+, December 2000.

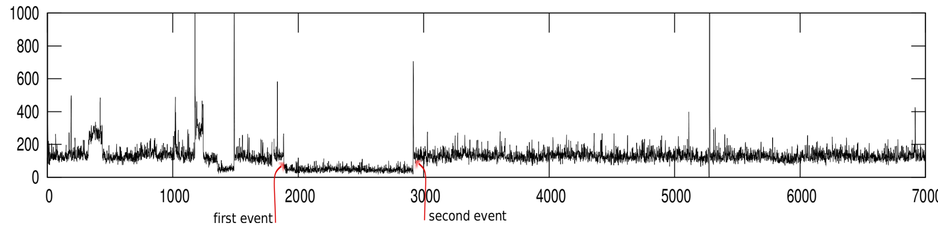


Fig. 7. Number a_i of appearing nodes, for all i and values $p = 10$ and $c = 2$. The two arrows denote two statistical events that were correlated with two known events tracked by the Abilene trouble tickets of Figure 8 and 9

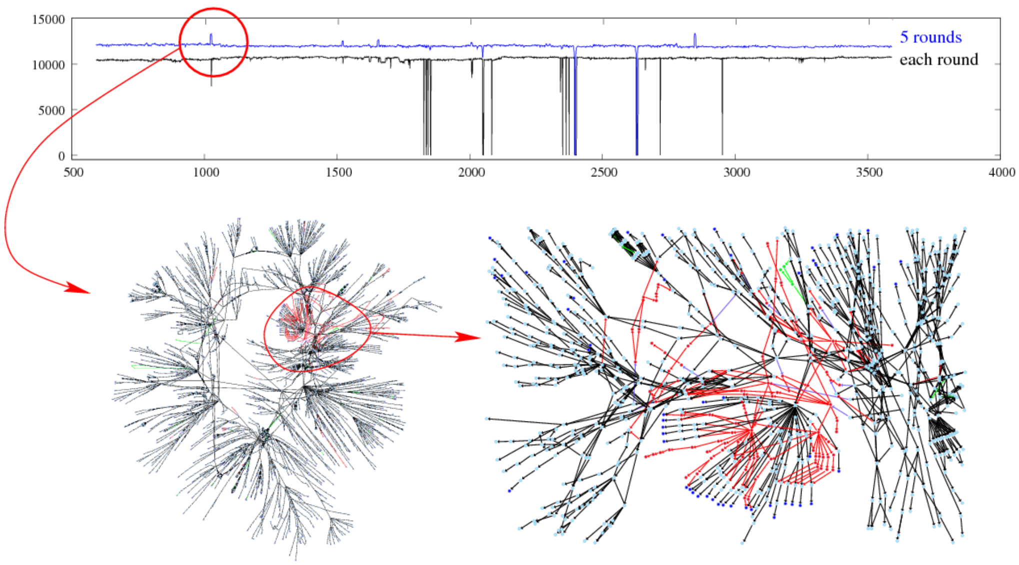


Fig. 10. Top row: the number N_i^5 of distinct nodes observed during five consecutive rounds of measurements, and the number of nodes observed at each round of measurement, *i.e.* N_i. Bottom row: the graph drawing of topological changes observed during the detection of an event with a zoom on the corresponding area on the ego-centered view. Appearing nodes and links are in red.

[6] Augustin, Brice, Cuvellier, Xavier, Orgogozo, Benjamin, Vigerand Fabien, Friedman, Timur, Latapy Matthieu, Magnien Clémence, and Teixeira Renata. Avoiding traceroute anomalies with paris traceroute. In *IMC '06: Proceedings of the 6th ACM SIGCOMM conference on Internet measurement*, pages 153–158, New York, NY, USA, 2006. ACM.

[7] Brice Augustin, Timur Friedman, and Renata Teixeira. Measuring load-balanced paths in the internet. In *IMC '07: Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*, pages 149–160, New York, NY, USA, 2007. ACM.

[8] Jake D. Brutlag. Aberrant behavior detection in time series for network monitoring. In *LISA '00: Proceedings of the 14th USENIX conference on System administration*, pages 139–146, Berkeley, CA, USA, 2000. USENIX Association.

[9] Qian Chen, Hyunseok Chang, Ramesh Govindan, Sugih Jamin, Scott J. Shenker, and Walter Willinger. The origin of power laws in internet topologies revisited, 2002.

[10] Dorothy E. Denning, Teresa F. Lunt, Roger R. Schell, Mark Heckman, and William R. Shockley. A multilevel relational data model. In *IEEE Symposium on Security and Privacy*, pages 220–234, 1987.

[11] Scott R Eliason and Michael S Lewis Beck. *Maximum Likelihood Estimation: Logic and Practice*. Sage Publications (CA), 1993.

[12] R. Govindan and A. Reddy. An analysis of internet inter-domain topology and route stability. In *Proc. IEEE INFOCOM*, 1997.

[13] Jean-Loup Guillaume, Matthieu Latapy, and Damien Magoni. Relevance of massively distributed explorations of the internet topology: qualitative results. *Comput. Netw.*, 50(16):3197–3224, 2006.

[14] D. Hawkins. *Testing for Normality*. Springer, 1980.

[15] Yiyi Huang, Nick Feamster, and Renata Teixeira. Practical issues with using network tomography for fault diagnosis. *Computer Communication Review 38(5)*, pages 53–58, 2008.

[16] Amelie Medem Kuatse, Renata Teixeira, and Mickael Meulle. Characterizing network events and their impact on routing. *CoNEXT*, page 59, 2007.

[17] Mohit Lad, Ricardo V. Oliveira, Daniel Massey, and Lixia Zhang. Inferring the origin of routing changes using link weights. *ICNP*, pages 93–102, 2007.

[18] Anukool Lakhina, Mark Crovella, and Christophe Diot. Mining anomalies using traffic feature distributions. *SIGCOMM Comput. Commun. Rev.*, 35(4):217–228, 2005.

[19] Matthieu Latapy, Clémence Magnien, and Frédéric Ouédraogo. A radar for the internet. In *ADN 08: 1st International Workshop on Analysis of Dynamic Networks in conjunction with IEEE ICDM 2008*, pages 901–908, 2009.

[20] Lun Li, David Alderson, John Doyle, and Walter Willinger. Towards a theory of scale-free graphs: Definition, properties,and implications. *Internet Mathematics*, 2(4), 2005.

[21] Lun Li, David Alderson, Walter Willinger, and John Doyle. A first-principles approach to understanding the internet's router-level topology. In *SIGCOMM '04: Proceedings of the 2004 conference on Applications, technologies, architectures, and protocols for computer communications*, pages 3–14, New York, NY, USA, 2004. ACM.

[22] Clémence Magnien, Frédéric Ouédraogo, Guillaume Valadon, and Matthieu Latapy. Fast dynamics in internet topology: preliminary observations and explanations. *Fourth International Conference on Internet Monitoring and Protection (ICIMP 2009)*, 2009.

[23] Ricardo V. Oliveira, Beichuan Zhang, and Lixia Zhang. Observing the evolution of internet as topology. *SIGCOMM Comput. Commun. Rev.*, 37(4):313–324, 2007.

[24] Damien Magoniand Jean Jacques Pansiot. Analysis of the autonomous

system network topology. *SIGCOMM Comput. Commun. Rev.*, 31(3):26–37, 2001.

[25] J.-J. Pansiot. Local and dynamic analysis of internet multicast router topology. *Annales des télécommunications*, 62:408–425, 2007.

[26] S.-T. Park, D. M. Pennock, and C. L. Giles. Comparing static and dynamic measurements and models of the internet's AS topology. In *Proc. IEEE Infocom*, 2004.

[27] William H. Press, Saul A. Teukolsky, William T. Vetterling, and B. P. Flannery. *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press, Cambridge, England, 1992.

[28] Balachander Krishnamurthy Subhabrata, Er Krishnamurthy, Subhabrata Sen, Yin Zhang, and Yan Chen. Sketch-based change detection: Methods, evaluation, and applications. In *In Internet Measurement Conference*, pages 234–247, 2003.

[29] Henry C. Jr. Thode. *Identification of Outliers*. CRC Press, 2002.

[30] Larry Wasserman. *All of Statistics: A Concise Course in Statistical Inference*. Springer, 2005.