Evolution of the Internet k-Dense Structure

Chiara Orsini, Enrico Gregori, *Associate Member, IEEE*, Luciano Lenzini, and Dmitri Krioukov

Abstract—As the Internet autonomous sytem (AS)-level topology grows over time, some of its structural properties remain unchanged. Such time-invariant properties are generally interesting because they tend to reflect some fundamental processes or constraints behind Internet growth. As has been shown before, the time-invariant structural properties of the Internet include some most basic ones, such as the degree distribution or clustering. Here, we add to this time-invariant list a nontrivial property—k-dense decomposition. This property is derived from a recursive form of edge multiplicity, defined as the number of triangles that share a given edge. We show that after proper normalization, the k-dense decomposition of the Internet has remained stable over the last decade, even though the Internet size has approximately doubled, and so has the k-density of its k-densest core. This core consists mostly of content providers peering at Internet eXchange Points, and it only loosely overlaps with the high-degree or high-rank AS core, consisting mostly of tier-1 transit providers. We thus show that high degrees and high k-densities reflect two different Internet-specific properties of ASs (transit versus content providers). As a consequence, even though degrees and k-densities of nodes are correlated, the relative fluctuations are strong, and related to that, random graphs with the same degree distribution or even degree correlations as in the Internet, do not reproduce its k-dense decomposition. Therefore an interesting open question is what Internet topology models or generators can fully explain or at least reproduce the k-dense properties of the Internet.

Index Terms—dK-graphs, k-dense, Internet topology, network evolution.

I. INTRODUCTION

THE DISCOVERY of power laws in the Internet in 1999 [1] came as a big surprise to many, and as a source of major disbelief to some. Even more surprising is that over the last decade, an increasing number of increasingly refined and complete macroscopic Internet topology measurements and data sources (with one exception, WHOIS) show that the power-law distribution of autonomous system (AS) degrees in the Internet has remained exceptionally stable, i.e., time-invariant [2]–[7]. One has to always keep in mind that these macroscopic measurements may miss a significant percentage of links. Indeed, there have been several studies supporting this expectation: Huge percentages of links are reported as missing in [8]–[12], for example. These studies rely on proprietary data collected from only a few ASs. Therefore, questions about statistical significance of the reported results, which are difficult or impossible to reproduce due to the proprietary nature of the data, are well grounded. However these questions may be not so important in view of that all these studies report that a majority of missing links are peering, while customer–provider links appear to be covered well in macroscopic topology measurements. These results are expected, given an ease at which peering links are "set up" at Internet exchange points (IXPs); for example: As soon as an AS connects to an IXP and declares an open peering policy, it can exchange traffic with any other openly peering AS at the same IXP. It is instructive to compare the setup of such "links" to the process behind setting up customer–provider links, often involving complicated business decisions, negotiations, and payment agreements [13], [14]. In other words, peering links are definitely more volatile than customer–provider links. At the same time, due to the specifics of Internet topology measurements, the higher in the customer–provider AS hierarchy are the two ASs connected by a peering link, the higher the probability that topology measurements detect this link. Therefore, a majority of missing peering links are between peripheral ASs, and they might affect the results presented here as discussed in Section III-B.

The time-invariant nature of the power-law distribution of (customer–provider) degrees, as well as of strong clustering, has recently found a new explanation as a consequence of tradeoff optimization between AS popularity and similarity in Internet evolution [15]. This optimization is rather general and applies not only to the Internet, but also to social and biological networks, and even to spacetime in our accelerating universe [16]. More generally, any time-invariant property of an evolving network structure is a candidate to reflect some fundamental forces or constraints behind network evolution, unless this property is a simple statistical consequence of some other property—degree distribution, for example. These forces and constraints can be quite general, applicable to many different networks, as is the case with degree distribution and clustering, or they can be specific to a given network, in which case they likely reflect some specific functions that this network performs.

Here, we show that the Internet k-dense decomposition is time-invariant and explain this invariance via Internet-specific AS data reflecting different functions and business roles of different ASs. The k-dense decomposition [17] of a graph G is a hierarchy of nested subgraphs H_k, $H_k \subset H_{k-1} \subseteq G$, $k = 3, 4, \ldots, k_{\text{MAX}}$, induced by edges belonging to $k - 2$ or more triangles within H_k. When $k = 2$, H_k is equal to G. A more practical definition of k-dense is the following: A link (i, j) belongs to a k-dense subgraph H_k if and only if nodes i and j that

C. Orsini is with the Cooperative Association for Internet Data Analysis, University of California San Diego (CAIDA/UCSD), La Jolla, CA 92093USA, and also with the Institute of Informatics and Telematics, Italian National Research Council (IIT/CNR), Pisa 56124, Italy, and the Department of Information Engineering, University of Pisa (IET/UNIPI), Pisa 56122, Italy (e-mail: chiara@caida.org).

E. Gregori is with the Institute of Informatics and Telematics, Italian National Research Council (IIT/CNR), Pisa 56124, Italy (e-mail: enrico.gregori@iit.cnr.it).

L. Lenzini is with the Department of Information Engineering, University of Pisa (IET/UNIPI), Pisa 56122, Italy (e-mail: l.lenzini@iet.unipi.it).

D. Krioukov is with the Cooperative Association for Internet Data Analysis, University of California San Diego (CAIDA/UCSD), La Jolla, CA 92093USA (e-mail: dima@caida.org).

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.
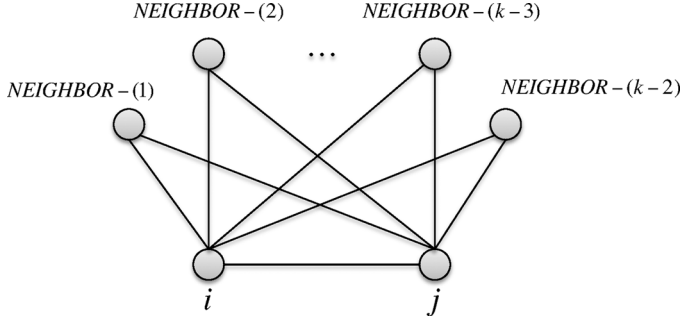
2

IEEE/ACM TRANSACTIONS ON NETWORKING

Fig. 1. Link i, j of multiplicity $m(i, j) = k - 2$. Edge multiplicity is the number of neighbors common to the two nodes that this edge connects [18].

this link connects have at least $k - 2$ common neighbors within H_k; see Fig. 1. If a link belongs to H_k but not to H_{k+1}, it is said to have the k-dense-index equal to k. The k-dense-index of a node is the maximum k-dense-index of its incident edges.

The k-dense-index of an edge is thus a recursive variant of edge multiplicity, defined as the number of triangles in G of which an edge is a member. Edge multiplicity was introduced and studied in [19] and [20], where it was also shown that the edge multiplicity distribution in many real networks is either power-law or fat-tailed, and that many existing network models fail to reproduce this property. In [21], the authors study the k-dense decompositions of different networks to scrutinize the global organization of clustering in them. The authors find that in most networks, triangles overlap similarly to how they overlap in random graphs with the same node degree and clustering distributions, and not as in more structured network models with modular organization. In [18] and [22], the authors perform a thorough analysis of the k-dense decomposition of a single Internet topology snapshot. They find that the k_{MAX}-dense ASs play an important role in Internet connectivity (43% of AS links in the Internet are attached to these ASs), and that there exist some correlations between the presence of dense subgraphs in the Internet AS-level topology and: 1) ASs participating in one or more IXPs, and 2) ASs with a wide geographical scope (i.e., ASs that have points of presence in more than one continent). Compared to this work, the high-level summary of our main results here is as follows.

- We focus on *evolution* of the Internet's k-dense decomposition: We base our study on a series of Internet topology snapshots collected from May 2004 to May 2012 and find that some important k-dense properties are time-invariant over the entire period.
- We show that the Internet's k-dense properties are statistically significant using the dK-series methodology [23].
- We collect and use data related to IXPs, business relationships, and organization of customer–provider hierarchies to understand the drivers behind the formation of AS communities with different k-densities.
- We find some dependencies between peering policies and complexity of the k-densest core in the Internet.

In more detail, our main results are the following.

1) The maximum k-dense-index k_{MAX} exhibits a clear growing trend, increasing from 29 in 2004 to 48 in 2012 (Section IV-A), while the number of ASs in the densest core $H_{k_{\mathrm{MAX}}}$ is small and fluctuating—it is 59 in 2004

and 60 in 2012. The k_{MAX} growth can only partially be attributed to the growing average degree. Other factors are likely related to increasing open peering at IXPs: All the 60 ASs in the 2012 $H_{k_{\mathrm{MAX}}}$ are present at at least one IXP (Section V-A).

2) After proper normalization (Section IV-B) the Internet k-dense decomposition is shown to be time-invariant in Section IV-C. The Internet-specific interpretation of this result appears in Sections V-B and V-C.

3) A significant part of this interpretation is centered around the observation that the degree of an AS and its k-density reflect two different Internet-specific properties of the AS, two different AS business roles. Related results include the following.

   a) While some correlations between AS degree and k-density are present as expected, the relative fluctuations are strong, i.e., nodes with the same degree or k-density have high-variance distributions of their k-densities and degrees, respectively (Section IV-D).

   b) Random graphs having the same degree distribution or even degree correlations as the Internet, do not have the same k-dense decomposition (Section IV-E).

   c) The analysis of PeeringDB data reveals that while high-degree or high-rank ASs tend to be tier-1 transit providers, high-k-density ASs tend to be tier-2 and content providers peering at IXPs (Section V-C).

4) The structure of densest core $H_{k_{\mathrm{MAX}}}$ is statistically determined by its degree distribution alone (Section IV-F). We interpret this result via open peering policies of participating ASs, which do not choose their peers based on their degrees (Section V-D). On the contrary, selective peering policies, present elsewhere in the Internet, likely introduce degree correlations, providing a new interpretation of the main result in [23], where the AS-level topology of the Internet was found to be statistically determined by its degree correlations.

Some other *related work* dealing with Internet evolution includes [5], which evaluates how different structural properties, related to the distributions of node degrees, centralities, path lengths, community structure, etc., change over time, from January 2002 to January 2010. The study relies on the Cramér–von Mises criterion to identify changes between the distributions and finds that most distributions remain unchanged, except for the average path length and clustering coefficient. These changes are interpreted as a consequence of peering policy changes. The different growth dynamics of the IPv4 and IPv6 topologies from 1997 to 2009 are juxtaposed in [24]. The main result is that IPv4 topology growth had a phase transition in 2001, while IPv6 had a different phase transition in 2006. The authors of [25] focus on topology liveness and completeness problems, comparing different Internet topology measurement data sources for the period from January 2004 to December 2006. Two evolution trends are highlighted in this work: 1) customer networks are the major cause of the overall topology growth; 2) transit providers tend to form increasingly denser structures. The monumental study [3] analyzes the evolution of customer–provider connections in the Internet from January 1998 to January 2010. AS links and nodes are labeled by business relationships and roles, and studied separately.
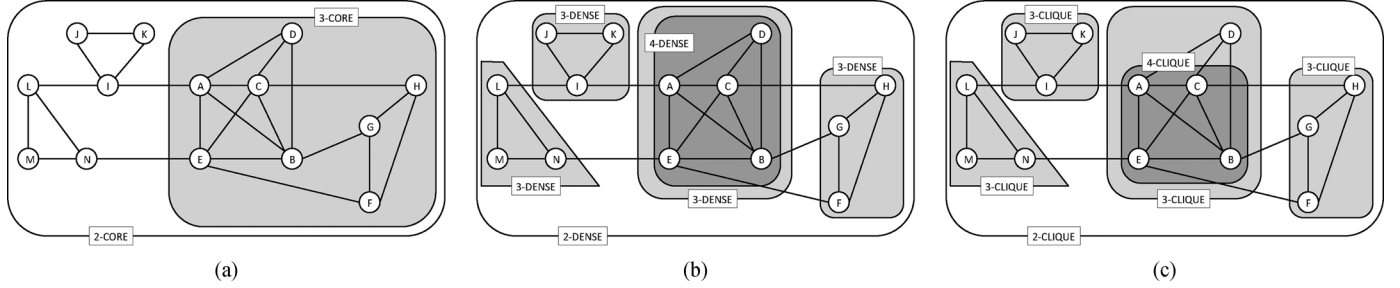
This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

ORSINI *et al.*: EVOLUTION OF INTERNET k-DENSE STRUCTURE

3

Fig. 2. Decompositions of a sample graph by the (a) k-core, (b) k-dense, and (c) k-cliques (CPM) methods.

The authors find that enterprise networks and content/access providers at the periphery are the main contributors to the overall Internet growth. They also study rewiring activity and find that content/access providers appear as most active in that regard. The metrics that are perhaps the most similar to the k-dense decomposition are the k-core decomposition [26]–[29] and the k-clique percolation method for community detection [30]. A k-core of a graph is its subgraph whose node degrees within the subgraph are greater or equal to k, and a k-clique community is a maximal union of adjacent k-cliques, where two k-cliques are adjacent if they share $k - 1$ nodes. The evolution of the k-core decomposition of the Internet from December 2001 to December 2006 is studied in [31]. Similar to the k-dense decomposition that we analyze here, the k-core decomposition also appears time-invariant according to [31]. However, contrary to the maximum k-density, the maximum k-coreness does not exhibit any growing trend. To the best of our knowledge, it remains unclear if these results in general, and AS coreness in particular, can find any Internet-specific interpretations.

II. k-DENSE METHOD

A. Motivations

The k-core and k-dense decompositions, as well as the k-clique percolation method (CPM), all identify densely connected communities that may still have many connections going outside, contrary to the modularity maximization [32], for example, where the number of such outside connections is minimized. We believe that these k-methods suit better the AS Internet reality, where for instance a group of large Internet service providers (ISPs) in a common geographical region is likely to be highly interconnected and thus form a dense community, but at the same time the same ISPs may have a large number of connections directed outside of this community to their dispersed customers [33].

The three k-methods are similar also in that they all are deterministic, recursive, and provide a set of nested communities of increasing density. Yet, they are different in how this density is defined. Any node in a k-core has at least k connections to other nodes in the same k-core; connected nodes in a k-dense community share at least $k - 2$ common neighbors and hence overlapping triangles belonging to the same community; while CPM communities are groups of overlapping maximal k-cliques—see Fig. 2 for a basic comparison of these decompositions of a small sample graph. A more detailed comparison of these k-methods can be found in [33].

We believe the k-dense decomposition is optimal among the three methods. Due to its common-neighbor requirement, the k-dense approach suggests stronger relationships among nodes belonging to the same community compared to a k-core decomposition and yields a more insightful view of how the network is organized. Indeed, if we know that a node is k-dense, then we immediately have richer information about the local structure of the node's neighborhood. We know not only that the node has degree at least $k - 1$ and so do its $k - 1$ neighbors, but also that with each of these neighbors the node shares at least $k - 2$ common neighbors, forming a highly clustered structure of strongly overlapping triangles. As shown in [15], this clustering indicates strong similarity between the involved ASs. Compared to CPM, the k-dense decompositions is less sensitive to noise [33]. From the computational perspective, while the k-dense decomposition is slightly more complex than the k-core decomposition, it is much less complex than CPM.

B. Definitions

The k-dense decomposition is a recursive graph decomposition [17], based on edge multiplicity [19], [20]. The multiplicity $m_G(i, j)$ of edge (i, j) in graph G is the number of triangles in G containing the edge, or equivalently, the number of common neighbors of connected nodes i and j; see Fig. 1. By definition, the k-dense subgraph H_k of graph G is the subgraph induced by all the links with multiplicity larger or equal to $k - 2$ *in the subgraph*

$$m_{H_k}(i, j) \geq k - 2. \tag{1}$$

This subgraph can be obtained from G by iterative pruning all the links with multiplicity smaller than $k - 2$. All the links in H_k have multiplicity equal or greater than $k - 2$ in G as well, $m_{H_k}(i, j) \geq k - 2 \Rightarrow m_G(i, j) \geq k - 2$, but the converse is not generally true. That is, links with $m_G(i, j) \geq k - 2$ are only *candidate* H_k links. The k-dense decomposition of G is a set of nested subgraphs $H_{k_{\text{MAX}}} \subset \cdots \subset H_{k+1} \subset H_k \subset \cdots \subset G$, where $H_{k_{\text{MAX}}}$ is the smallest and densest nonempty subgraph. An example of k-dense decomposition is shown in Fig. 2(b).

A link is said to have the k-*dense-index* equal to k^* if it belongs to H_{k^*} but not to H_{k^*+1}. The set of all the links with the k-dense-index equal to k^* is called the k^*-dense-shell.

A node is said to have the k-*dense-index* equal to k^* if k^* is the maximum k-dense-index of its incident links. The set of all the nodes with k-dense-index equal to k^* is called the k^*-dense-set.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

4

IEEE/ACM TRANSACTIONS ON NETWORKING

TABLE I
NUMBER OF AS NODES, AS LINKS, AND THE VALUES OF k_{MAX}

	2004	2005	2006	2007	2008	2009	2010	2011	2012
Nodes	17,858	20,486	23,044	26,101	29,042	32,379	35,583	38,888	42,419
Links	50,326	59,382	66,781	79,931	90,588	102,362	113,846	127,558	146,271
k_{MAX}	29	33	32	34	40	37	39	42	48

III. DATA

A. Description

We use the AS-level topology data from the UCLA Computer Science Department's Internet Research Lab (IRL) [34], [35]. We collect yearly Internet snapshots dated from May 2004 to May 2012. Specifically, for each year, we download the data corresponding to May 31, and then discard all the links with the *last seen* attribute older than May 1. Table I reports the numbers of AS nodes and links in each snapshot.

We emphasize that BGP and traceroute-based data provide incomplete and biased views of the real Internet AS-level topology. For example, many connections between leaf (low-degree) ASs are hidden from monitors located in hub (high-degree) ASs [12]. Changing numbers of monitors introduce various artifacts and aberrations in the observed topologies. Since only the best paths are typically announced and used for routing, the observed views are highly incomplete. Specific to traceroute measurements, there are many nonresponding ASs, especially leaf ASs comprising a majority of Internet ASs. There are also many issues with mapping IP addresses to AS numbers, including IP addresses for which there are multiple or no mapping ASs, and many other vagaries [36].

Our decision to use the IRL BGP data was motivated primarily by the observation that the number of ASs in this dataset is relatively consistent with the number of the Advertised AS Count growth in the CIDR report [37].

B. Incompleteness

While these Internet topology measurements are known to be incomplete, many studies have demonstrated that a vast majority of missing links are peering connections [8]–[12]. The reason for this is that a topology measurement monitor located in AS X can observe only X's peering connections and the peering connections between ASs that are above X in the customer–provider AS hierarchy [8]. As a consequence, peering connections between large providers tend to be detected by topology measurements with high probability, while peering links between small ASs are mostly missing.

Based on these observations, we believe that a k-dense analysis of the complete topology would differ from the analysis below as follows.
- The k_{MAX} index would probably be higher.
- Some ASs would have higher k-dense indices.
- There would probably be a larger number of small k-dense communities disconnected from the giant k-dense components (H_ks consist of one connected component for most values of k [33]).

At the same time, a considerable percentage of low-degree ASs would unlikely be affected by this incompleteness because most ASs whose business is not related to Internet operations connect only to their providers via *observable* customer–provider links.
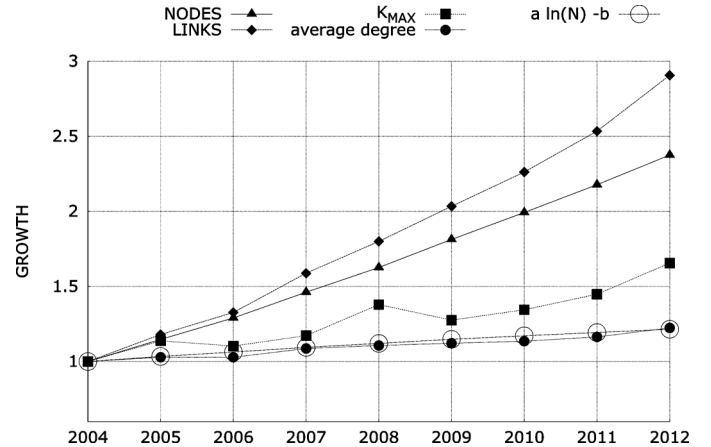

Fig. 3. Internet growth in terms of numbers of nodes, links, k_{MAX}-dense-index, and average degree. The black triangles show $N(t)/N(t_0)$, the number of nodes in the graph at time t divided by the number of nodes at time $t_0 = 2004$, $N(t_0) = 17,858$. The black rhombuses show the same ratio for the number of links, $M(t)/M(t_0)$, $M(t_0) = 50,326$. The black squares are $k_{\mathrm{MAX}}(t)/k_{\mathrm{MAX}}(t_0)$, $k_{\mathrm{MAX}}(t_0) = 29$. The black circles show the average degree of the graph at time t divided by the average degree at time $t_0 = 2004$, $\overline{k}(t_0) = 5.64$. The average degree grows logarithmically with the Internet size, $\overline{k} \approx a \ln N - b$, with $a = 1.3$ and $b = 7.5$. The empty circles are $(a \ln N(t) - b)/(a \ln N(t_0) - b)$, showing the quality of this approximation.

IV. INTERNET'S k-DENSE DECOMPOSITION

Treating the data described in Section III as a sorted set of AS-level graphs ordered by year, we report in this section the results of the statistical analysis of Internet's k-dense decomposition and its evolution over time. The interpretation of these results is deferred to Section V.

A. Basic Trends

Similar to other studies, e.g., [3] and [25], we have not observed and do not report any significant changes in the basic graph properties—degree distribution, degree correlations, clustering, betweenness, and shortest path distributions—even though the graph has grown significantly; see Table I and Fig. 3. We see that the number of links M has been growing faster than the number of nodes N, meaning that the average degree $\overline{k} = 2M/N$ has been increasing. This increase appears to be a logarithmic function of N, $\overline{k} \approx a \ln N - b$, with $a = 1.3$ and $b = 7.5$, in agreement with [15] and [16]. The increasing trend of k_{MAX} growing from 29 in 2004 to 48 in 2012 indicates that more densely connected parts in the Internet have been forming in the course of its growth.

B. k-Dense Normalization

For each snapshot, we first compute the number of nodes and links with a given k-dense-index, i.e., the number of nodes in each k-dense-set and the number of links in each k-dense-shell. Since the graph in the snapshots and its k_{MAX} are growing, to be

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

ORSINI *et al.*: EVOLUTION OF INTERNET k-DENSE STRUCTURE   5
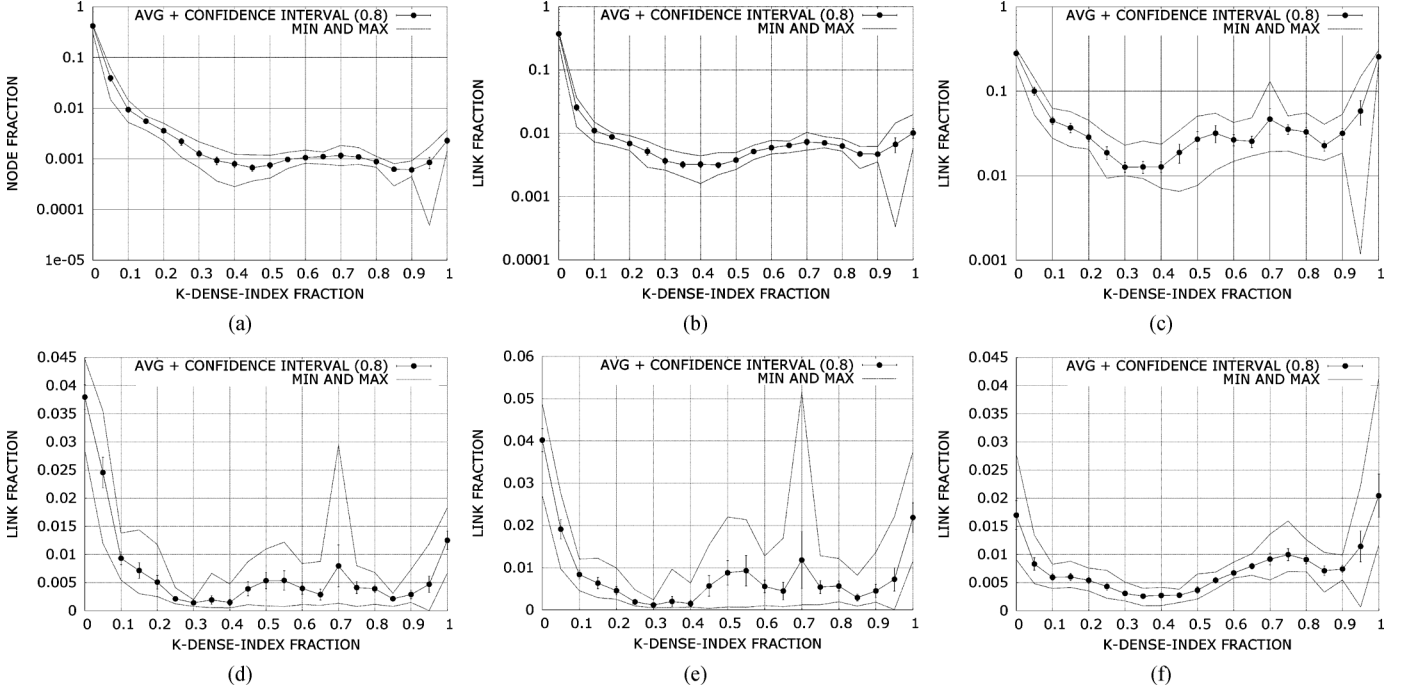

Fig. 4. Normalized k-dense decomposition of the Internet. The x-axes are linearly binned with the bin size equal to 0.05, and each plot shows the average (black dots and solid lines), 80% percentile (vertical bars), and minima and maxima (dashed lines) of the corresponding values computed across the nine historical snapshots of the Internet. (a) Fraction of nodes with a given k-dense-index, i.e., the number of nodes in the k-dense-set divided by the total number of nodes in the corresponding snapshot. (b) Fraction of links with a given k-dense-index, i.e., the number of links in the k-dense-shell divided by the total number of links in the corresponding snapshot. (c) Fraction of links attached to a given k-dense-set, i.e., the number of links whose one or two ends are attached to nodes in a given k-dense-set, divided by the total number of links in the corresponding snapshot. (d) Number of links with one end attached to a 2-dense-set node and with the other end attached to a k-dense-set node, divided by the number of links attached to the 2-dense-set. (e), (f) Corresponding fractions for links attached to 3- and k_{MAX}-dense-sets.

able to properly compare the nine snapshots, we next perform the following normalization, mapping k-dense-indices and corresponding numbers of nodes and links to fractions with values between 0 and 1.

- *x-axis normalization*: Map each index k in each snapshot to what we call the *k-dense-index fraction*

$$x = \frac{k - k_{\mathrm{MIN}}}{k_{\mathrm{MAX}} - k_{\mathrm{MIN}}} \qquad (2)$$

  where k_{MIN} and k_{MAX} are the minimum and maximum values of the k-dense-index in the snapshot.
- *y-axis normalization*: Divide the corresponding number of nodes or links by the total number of nodes or links in the snapshot.

In all the considered snapshots, there are no nodes of degree 0, so that $k_{\mathrm{MIN}} = 2$ for links and nodes in all the snapshots. This means that $x = 0$ corresponds to the k-dense-index equal to 2, while $x = 1$ corresponds to the k-dense-index equal to k_{MAX}, which grows with time. Upon such normalization, it becomes difficult to read off the absolute values of k-dense-indices, nodes, or links from the resulting plots, but this normalization is helpful to see if there are any size- and time-invariant statistical trends.

C. k-Dense Decomposition

In Fig. 4(a) and (b), we report the normalized fractions of nodes and links for each k-dense-index, averaged over the nine

historical snapshots. These average values are highly representative for each snapshot from 2004 to 2012, in the sense that the distributions of the actual values across the snapshots are quite narrow, as indicated by the 80% percentile and min/max deviations from the averages in the figures. In other words, the shown normalized statistics are fairly stable over time, i.e., time- and size-invariant. In particular, a vast majority of nodes and links in all the snapshots have low k-dense-indices, not belonging to any dense communities.

Similar to the joint degree distribution, showing how nodes of different degree interconnect [23], we next focus on how different k-dense-sets interconnect. In Fig. 4(c), we report the normalized fraction of links attached to nodes in a given k-dense-set. Similarly to Fig. 4(a) and (b), we observe that these statistics are also stable over time. We also observe that there are two classes of k-dense-sets to which a vast majority of all links are attached—those with the smallest and largest k-dense-index fractions close to 0 and 1. Upon examination of the original data, we find that the left peak in Fig. 4(c) is formed by the links attached to the 2- and 3-dense sets, while the right peak is due to the links attached to the k_{MAX}-dense set. This observation motivates us to further focus on the links attached to these three sets and show where their other ends go in Fig. 4(d)–(f). We see that 2- and 3-dense-attached links go mostly to other low-dense-index nodes and to the nodes in the k_{MAX}-dense-set, and that k_{MAX}-dense nodes direct a considerable percentage of their connections not only to nodes in low-dense-sets, but also to nodes in densely-connected parts of the network. We also notice
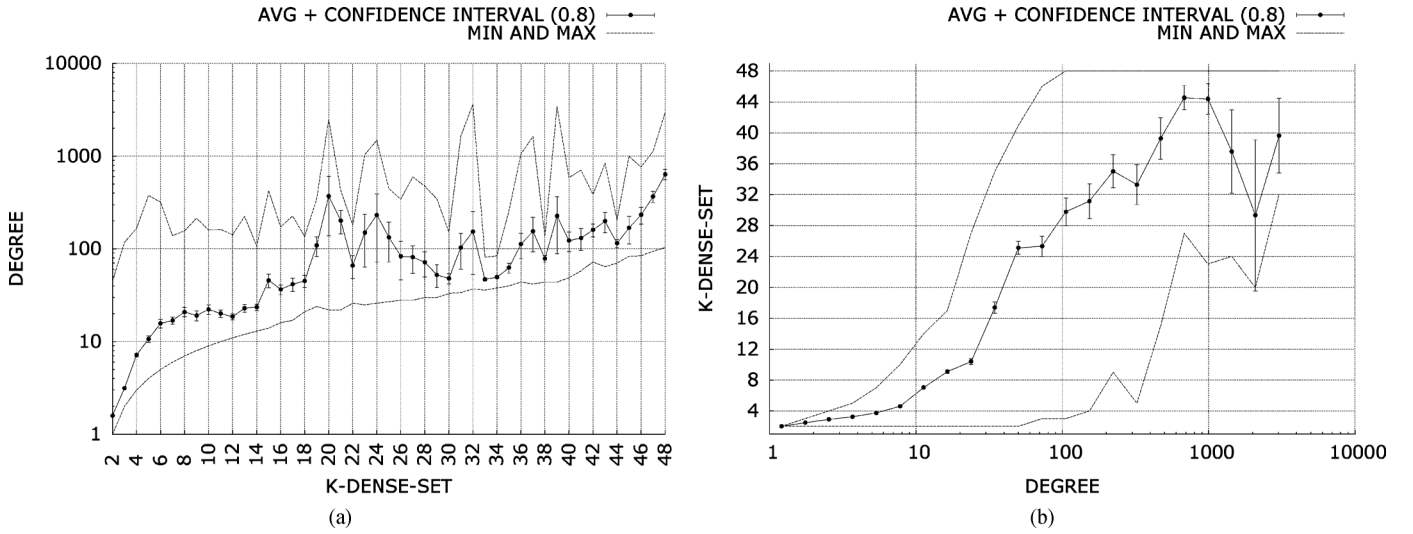
This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

6

IEEE/ACM TRANSACTIONS ON NETWORKING

Fig. 5. Degrees of nodes versus their k-dense-indices in the 2012 Internet snapshot. The figures show the averages, 80% percentiles, and min/max values for each distribution: (a) distributions of node degrees in k-dense-sets and (b) distributions of k-dense-indices of nodes with degrees falling within a given range obtained by logbinning the degree values.

TABLE II
AVERAGE PROPERTIES OF NODES IN THE 2-, 3-, AND k_{MAX}-DENSE-SETS:
AVERAGE DEGREE $\overline{k}$, AVERAGE CLUSTERING COEFFICIENT $\overline{c}$, AND
AVERAGE BETWEENNESS CENTRALITY $\overline{b}$

	$\overline{k}$	$\overline{c}$	$\overline{b}$
2-dense-set	1.643	0	$2.1 \cdot 10^3$
3-dense-set	3.161	0.746	$9.3 \cdot 10^3$
k_{MAX}-dense-set	403.8	0.194	$3.0 \cdot 10^6$

that all the considered normalized statistics appear to be quite stable over time, as indicated by the narrow 80% percentiles.

Given that a vast majority of links are attached to nodes in 2-, 3-, or k_{MAX}-dense sets, we further characterize these sets in Table II. The significant percentages of links attached to 2- and 3-dense sets are not surprising in view of that a majority of nodes belong to these sets; see Fig. 4(a). We see in Table II that these peripheral sets are populated with nodes of low average degree and betweenness centrality of the order of the number of nodes N in the graph. The k_{MAX}-dense-set, on the other hand, consists of a small number of nodes with high average degree and betweenness of the order of N^2. That is, these nodes have a central position in the network, and they are a key element shaping the overall connectivity, with many links attached to them [Fig. 4(c)], which motivates us to further analyze the structure of this k_{MAX}-dense subgraph. However, before doing so in Section IV-F, we first answer the following two natural questions.

1) Does the degree of a node fully define its k-dense-index, making this index a repetitive statistics providing no new information compared to node degree?
2) Do random graphs having the same degree distribution or degree correlations as the Internet fully reproduce all its k-dense properties, making them statistically insignificant, that is, casting them as statistical consequences of that this graph has the observed (joint) degree distribution?

D. *k-Dense-Index Versus Node Degree*

By definition, a node with k-dense-index equal to k has degree greater or equal to $k - 1$, while a node with degree k has

a k-dense-index smaller or equal to $k + 1$. These constraints do not preclude strong relative fluctuations between these two properties within the allowed bounds. For example, a node with a high degree can be the center of a hub-and-spoke configuration and have a lower k-dense-index than a node with a lower degree belonging to a small clique. In Fig. 5, we show the relation between the k-dense-indices of nodes and their degrees in the 2012 snapshot. As expected, the average degree of nodes in a given k-index-set as a function of k exhibits a growing trend, however the fluctuations are high. In particular, the degrees of nodes in the 2-dense-set vary from 2 to 45, while the degree of nodes in the k_{MAX}-dense-set ($k_{\mathrm{MAX}} = 48$) vary from 103 to 2964. The other way around, nodes of degree 100, for example, have k-indices ranging from 12 to 45, and the node with maximum degree 174 (Cogent Communications inc., $k = 3655$) has the k-dense-index equal to $32 < k_{\mathrm{MAX}}$. We conclude that the k-index is not fully determined by the node degree, an observation that gains further support and Internet-specific explanation in Section V.

E. *Statistical Significance of the k-Dense Properties*

Even though the node degree does not fully determine the node k-index, it may still be the case that random graphs having the same degree distribution as the Internet also have exactly the same k-dense properties, meaning that these properties are nothing but statistical artifacts, as it turned out to be the case with the rich club connectivity in the Internet [38]. To check if a similar story applies to the k-dense properties, we construct dK-random graphs [23] for $d = 0, 1, 2$ as described in Section IV-F. We generate 10 realizations for each d. These graphs are random graphs with the same average degree, degree distribution, or joint degree distribution as in the Internet 2012 snapshot. In Table III, we report the distributions of the k_{MAX} values in these random graphs for each d, while in Fig. 6 we show the k-dense link decompositions of these graphs and juxtapose them against the Internet's. We observe that neither degree distribution nor joint degree distribution fully reproduce
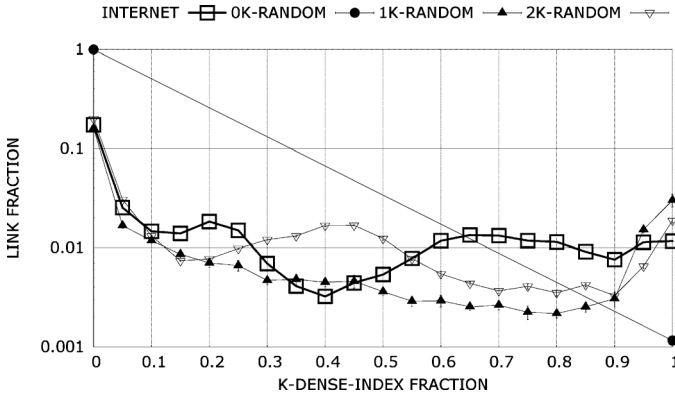
Fig. 6. Normalized k-dense decomposition of links in Internet 2012 and its dK-randomizations for $d = 0, 1, 2$. The figure is similar to Fig. 4(b) and shows the averages (data points and solid lines) and 80% percentiles (vertical bars) of the distributions of k-dense link fractions in dK-random graph instances for $d = 0, 1, 2$.

TABLE III
k_{MAX}-DENSE-INDEX IN INTERNET 2012 AND ITS dK-RANDOMIZATIONS FOR $d = 0, 1, 2$. THE TABLE SHOWS THE AVERAGES $\bar{k}_{\mathrm{MAX}}$ AND STANDARD DEVIATIONS σ OF THE DISTRIBUTIONS OF THE k_{MAX}-INDEX VALUES IN dK-RANDOM GRAPH INSTANCES FOR $d = 0, 1, 2$

	$\bar{k}_{MAX}$	σ
Internet	48	-
$0K$-random	3	0
$1K$-random	68	4.05
$2K$-random	44.3	0.46

TABLE IV
BASIC PROPERTIES OF k_{MAX}-DENSE CORES IN THE NINE HISTORICAL SNAPSHOTS: N THE NUMBER OF NODES, M THE NUMBER OF LINKS, $D = 2M/[N(N-1)] \approx \bar{k}/N$ THE GRAPH DENSITY

Year	k_{MAX}	N	M	D
2004	29	59	1,381	0.807
2005	33	55	1,318	0.888
2006	32	44	872	0.922
2007	34	97	3,159	0.678
2008	40	63	1,763	0.903
2009	37	55	1,353	0.911
2010	39	60	1,595	0.901
2011	42	81	2,745	0.847
2012	48	60	1,703	0.962

the k-dense properties of the Internet, meaning that these properties have their own statistical significance, which is another observation that finds a natural Internet-specific explanation in Section V. Yet another property that finds an explanation rooted in Internet specifics in that section is an unexpectedly simple structure of the k_{MAX}-index core that we analyze next.

F. Structure of the k_{MAX}-Dense Core

Given the importance of $H_{k_{\mathrm{MAX}}}$, the densest and innermost subgraph core (Section IV-C), we next characterize its structure in more detail. In Table IV, we show the basic properties of $H_{k_{\mathrm{MAX}}}$ subgraphs in all the nine snapshots. We see that these subgraphs are quite small and dense (link density $D = 1$ in complete graphs).

Similar to Section IV-E, we next perform the standard dK-statistical analysis [23] of these subgraphs. The procedure is as follows. First, we focus on the latest 2012 snapshot, extract the $H_{k_{\mathrm{MAX}}}$ subgraph from it, and treat this subgraph

as a separate graph. In particular, node degrees are computed within $H_{k_{\mathrm{MAX}}}$. Then, we construct 20 random graphs having exactly the same average degree as this graph, using the standard $\mathcal{G}_{N,M}$ (Erdős–Rényi [39]) construction procedure of throwing M edges onto $N(N-1)/2$ node pairs uniformly at random. These graphs are called $0K$-random graphs. Then, we construct 20 random graphs having exactly the same degree sequence as $H_{k_{\mathrm{MAX}}}$ using the fast generalized Havel–Hakimi algorithm from [40]. This algorithm is guaranteed to always quickly succeed as soon as the degree sequence is graphical, i.e., realizable. The resulting graphs are called $1K$-random graphs. Then, we compute the basic structural graph properties that in neither $0K$- nor $1K$-random graphs are guaranteed to be the same as in the original $H_{k_{\mathrm{MAX}}}$ graph. The results are in Fig. 7.

The average neighbor degree is not guaranteed to be captured by $1K$-random graphs. Only $2K$-random graphs, having exactly the same joint degree distribution as the original graph, guarantee to reproduce this degree correlation statistics. Yet, we find that $1K$-random graphs have the same average neighbor degree as $H_{k_{\mathrm{MAX}}}$. Clustering is guaranteed to be fully captured only by $3K$-random graphs, reproducing the frequencies of triangular subgraphs. Yet, we find that $1K$-random graphs have the same clustering as $H_{k_{\mathrm{MAX}}}$. The frequencies of motifs of size 3 or 4 are guaranteed to be fully captured only by $3K$- or $4K$-random graphs, the latter reproducing the frequencies of subgraphs of size 4 by definition, but we find that there are no statistically significant deviations of the frequencies of these motifs in $1K$-random graphs from the corresponding counts in $H_{k_{\mathrm{MAX}}}$. Finally, the global betweenness or shortest path length distributions are not guaranteed to be captured by dK-random graphs with any $d < N$, but again we see that these global statistics are well reproduced in $1K$-random graphs, although not as well as the local statistics. At the same time, none of the considered structural properties is closely approximated by $0K$-random graphs (except for the shortest path length distribution).

Collectively, these observations imply that even though the $H_{k_{\mathrm{MAX}}}$ core is so dense (link density $D = 0.962$) and close to being a complete graph that one could expect it to be well approximated by $0K$-random graphs, this expectation is actually wrong, but $1K$-random graphs capture closely the basic structural properties of this core. We applied the same dK-analysis to the $H_{k_{\mathrm{MAX}}}$ cores in all the other snapshots and found them all to be $1K$-random as well. In other words, $1K$-randomness of the $H_{k_{\mathrm{MAX}}}$ core is another time-invariant property.

V. INTERPRETATION OF INTERNET'S k-DENSE PROPERTIES

In this section, we interpret and explain the statistical results from Section IV using specifics of the Internet, where nodes are ASs and links are business relationships between ASs. That is, a link between a couple of ASs is a business agreement between the two organizations that enables these networks to exchange traffic. Recall that business relationships can be very roughly split into two classes: *provider–customer* and *peering*. Providers announce all the destinations to their customers and thus forward all the traffic that their customers forward to them. Peers mutually announce a limited set of destinations, typically just their own destinations and their customer networks. Providers
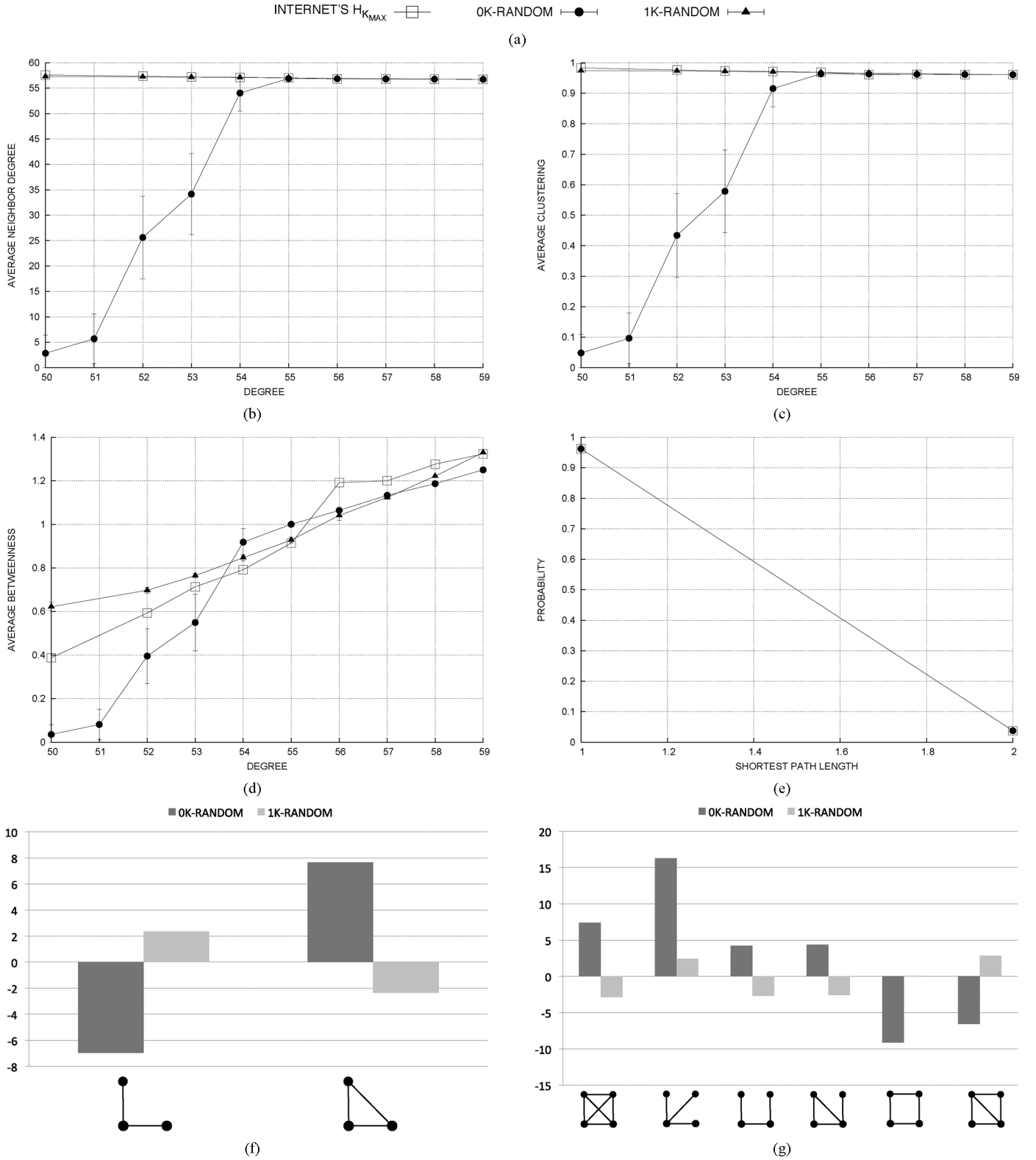
This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

8

IEEE/ACM TRANSACTIONS ON NETWORKING

Fig. 7. Basic structural properties of the $H_{k_{\mathrm{MAX}}}$ core in Internet 2012 versus its dK-randomizations for $d = 0, 1$. The figure shows the averages (points and solid lines) and 80% percentiles computed across different dK-random graph instances. Plots (b), (c), and (d) show the average neighbor degree, clustering, and betweenness of nodes of a given degree, while (e) shows the shortest path length distribution. Plots (f) and (g) show the z-scores z for motifs of size 3 and 4, $z = (x - \mu)/\sigma$, where x is a number of occurrences of a given subgraph in $H_{k_{\mathrm{MAX}}}$, while μ and σ are the mean and standard deviation of the distribution of the corresponding occurrences computed across different dK-random graph instances.

often charge their customers using the 95th percentile measurement schema [14], i.e., the cost of the service depends on the amount of traffic exchanged. Peering is usually free of charge (unless maintenance costs are considered). The setup of public peering is greatly simplified by IXPs, facilities where each participant AS can create a single peering connection to any other
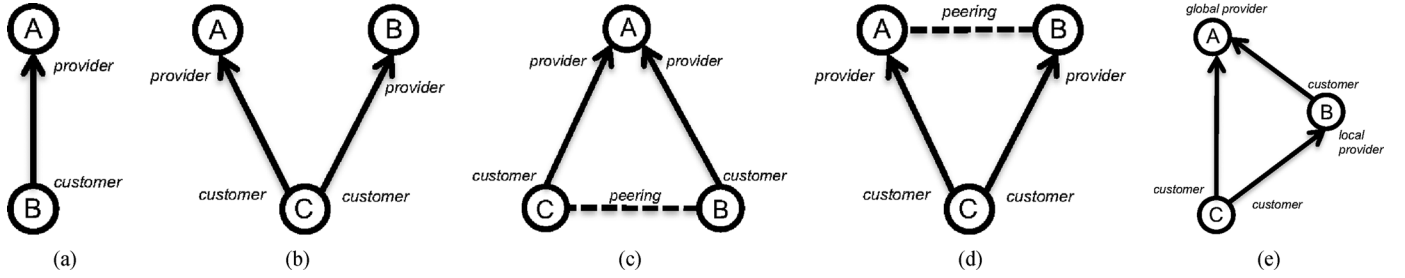
This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

ORSINI *et al.*: EVOLUTION OF INTERNET k-DENSE STRUCTURE

9

Fig. 8. Connectivity patterns of ASs with k-dense-indices equal to 2 and 3. (a) Single provider. (b) Multihoming. (c) Peering customers. (d) Peering providers. (e) Multihoming with local and global providers.

participants that accept to peer (*open peering* policy, or to a specific subset of those (*selective/restrictive peering* policy). In what follows, we explain the main observations from Section IV in view of these Internet-specific realities.

A. Growth of the k_{MAX}-Dense Index

Even though the size of the $H_{k_{\mathrm{MAX}}}$ core fluctuates over time (Table IV), the k_{MAX}-dense index steadily grows (Fig. 3). In this section, we provide some evidence that this growth is correlated with (if not due to) the proliferation of IXPs.

The first piece of evidence is the structural change of the Internet topology in the last decade due to different growth rates of peering and provider–customer relationships. The original economic driver behind peering is to reduce costs that customers pay to their providers. However, as the price of transit has been steadily decreasing, this driver becomes less prominent, unless large volumes of traffic are exchanged. As shown in [13], peering grows fast, contributing much more to middle AS tiers, compared to the tier-1 ASs. Large content providers (CPs) and content delivery networks (CDNs), which generate a significant percentage of the total Internet traffic [13], [41], are primary drivers behind this process because of the following reasons.

1) A shorter path between these networks and subscribers provides better performances.
2) Although the traffic exchanged is highly asymmetric, for most ISPs, the connection to the content providers is vital.

Furthermore, [41] asserts that a dominant percentage of AS-level traffic flows directly between large CPs, CDNs, data centers, and consumer networks, and that 150 ASs originate more than 50% of the Internet inter-AS traffic. Reference [11] also supports our claim by describing the ground truth behind a large European IXP: The authors shows that the amount of peering links established in this facility is unexpectedly huge—about 50 000 peering links among approximately 350 AS members. Yet another piece of evidence can be found in [18]: The percentage of ASs that are members of at least one IXP within a given k-dense set is a rapidly growing function of k. In addition, all the k_{MAX}-dense ASs are members of at least one IXP.

To further support our claim here, we perform analysis similar to [18]. We collect the May 2012 information about the 60 k_{MAX}-dense ASs in our 2012 snapshot from PeeringDB [42], a project maintaining a database aimed at facilitating the exchange of information related to peering: *what networks are peering, where they are peering, and if they are likely to peer*

with you. We find that 58 out of the 60 ASs have a peering record in the database, and that *all* these 58 ASs are members of the Deutscher Commercial Internet Exchange (DE-CIX) [43], one of the largest Internet Exchange Points worldwide, located in Frankfurt, Germany, with the current membership count of more than 450 ASs. We investigate the profile of the remaining two ASs and find that they both are also DE-CIX members: One is a telecommunication company with an unknown peering policy; the other is a hosting service provider with an open peering policy.

B. 2- and 3-Dense-Sets

Nodes with the k-dense-index equal to 2 or 3 are a vast majority of all ASs in the Internet. These peripheral ASs contribute most to the overall network growth: They have small degrees, but a large percentage of all connections in the Internet are attached to these ASs [3], [25]. All ASs of degree 1 and all ASs with zero clustering belong to the *2-dense-set*. These customer ASs connect to the Internet, but their business is not related to Internet operations: They set up their connections to obtain Internet connectivity, thus all they need is a transit provider. Sometimes, for backup, higher availability, or other purposes, they set up agreements with multiple provider, i.e., they purchase transit from more than one provider (multihoming). If the two providers of a multihomed AS happen to be connected, then a triangle is formed, so that the multihomed AS belongs to the *3-dense-set* (Fig. 8).

C. k_{MAX}-Dense-Set Versus High-Rank ASs

The k_{MAX}-dense ASs form the densest-connected community by definition. The easiest way to establish such dense connectivity in practice is by connecting to a large IXP and declaring an open peering policy. Given that CPs and CDNs benefit from peering with any willing-to-peer ASs [45], it is quite plausible that CPs and CDNs are main players behind the formation of this densely interconnected substructure. Surprisingly, the adoption of an open peering policy is an emerging phenomenon among network service providers (NSPs) [46], i.e., tier-2 ASs provided with an own backbone network that purchase transit from an upstream provider and resell it to other ASs. Although these ASs usually adopt a selective peering policy, as they do not want to peer with potential customers, such peering connections help tier-2 ASs to provide a better end-user experience to their customers [47].

The PeeringDB data shown in Table V(a) and (d) confirm this statement. Indeed, large percentages of the k_{MAX}-dense ASs:

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

10   IEEE/ACM TRANSACTIONS ON NETWORKING

TABLE V

AS PROPERTIES FROM PEERINGDB. THE k_{MAX}-*dense* COLUMN REFERS TO THE 60 k_{MAX}-DENSE ASs IN THE 2012 SNAPSHOT, WHILE THE *degree-rank* AND *AS-rank* COLUMNS REFER TO THE 60 HIGHEST-DEGREE AND HIGHEST-AS-RANK ASs [44]. THE DATA FOR THE HIGHEST-ADDRESS- AND PREFIX-RANK ASs ARE SIMILAR TO THE HIGHEST-AS-RANK ASs. (a) BUSINESS TYPE. (b) GEOGRAPHIC TYPE. (c) TRAFFIC VOLUME. (d) TRAFFIC RATIO. (e) PEERING POLICY

	k_{MAX}-dense	degree-rank	AS-rank
Cable/DSL/ISP	12	9	5
Content	4	1	0
Educational/Research	1	3	1
Network Service Provider	42	42	43
Non-Profit	1	0	0

(a)

	k_{MAX}-dense	degree-rank	AS-rank
Asia Pacific	0	0	2
Europe	28	20	8
North America	0	2	5
Regional	17	8	4
Global	15	25	30

(b)

	k_{MAX}-dense	degree-rank	AS-rank
100 Mbps - 10 Gbps	10	6	0
10 Gbps - 1 Tbps	41	36	29
1Tbps+	2	7	12
Unknown	7	11	19

(c)

	k_{MAX}-dense	degree-rank	AS-rank
Balanced	30	42	37
Mostly Inbound	7	6	9
Mostly Outbound	21	7	3
Unknown	2	5	11

(d)

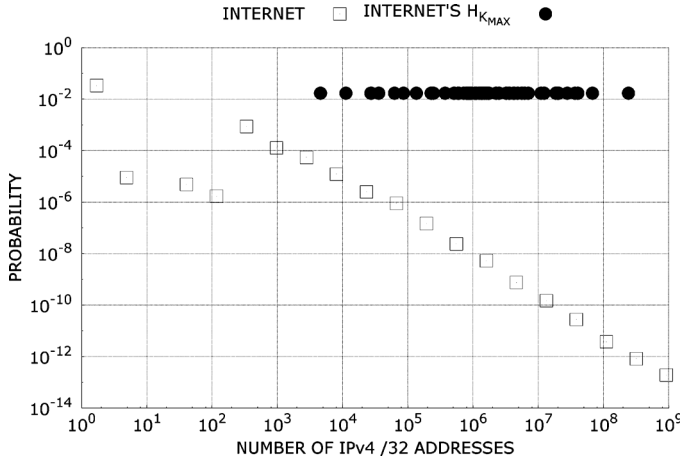|	k_{MAX}-dense	degree-rank	AS-rank
Open	32	15	7
Restrictive	1	11	15
Selective	26	29	27
Unknown	1	5	11

(e)

Fig. 9. Distributions of customer cone sizes for the k_{MAX}-dense ASs (black circles) and all ASs in the Internet (empty squares).

- can be considered as content or network service providers;
- direct most of their traffic outbound;
- have an open peering policy.

The percentage of ASs with selective peering policies (almost all NSPs) is also significant, but all these ASs are good candidates for selective peering as well, explaining their high k-density. Indeed, one commonly considered aspect in the peer selection process is the symmetry of the exchanged traffic [48]. We do not have access to traffic statistics, but we can use the number of IP addresses in the customer cone [49] as a proxy. The customer cone of an AS is the set of ASs that can be reached from the AS following only provider-to-customer links. In other words, it is the set of destinations that can be reached *for free* upon peering with the AS. The distributions of the customer cone sizes (the numbers of /32 IP addresses in the customer cone, to be precise) shown in Fig. 9 indicate that the customer cone sizes of k_{MAX}-dense ASs are large, but their distribution is narrower that in the rest of the Internet, thus making k_{MAX}-dense ASs potential peer candidates even in the selective peering case.

We next juxtapose this k_{MAX}-dense-set against different sets of high-rank ASs, in particular high-degree ASs, shedding more light on the different Internet-specific meanings of the k-dense-index and node degree.

Tier-1 is defined as a set of ASs that do not need to pay any transit providers to reach any destination. To accomplish

TABLE VI

60 k_{MAX}-DENSE ASs IN THE 2012 SNAPSHOT VERSUS TOP-60 HIGH-RANK ASs. THE TABLE SHOWS THE OVERLAPS BETWEEN THE SET OF ASs IN $H_{k_{\mathrm{MAX}}}$ AND *address-rank*, *prefix-rank*, AND *AS-rank*: THE SETS OF 60 ASs WITH THE LARGEST NUMBERS OF IPV4 /32 ADDRESSES, IPV4 ROUTING PREFIXES, AND ASs IN THEIR CUSTOMER CONES; AND *degree-rank*: THE SET OF 60 HIGHEST-DEGREE ASs. ALL THE DATA ARE FROM [44]

k-max dense	address-rank	prefix-rank	AS-rank	degree-rank
60	4	6	7	31

this, all the tier-1 ASs connect to each other forming a clique. This structure ensures that all these ASs have high, although not necessarily *highest* k-dense-indices. However, because they provide routes to all the destinations in the Internet without paying any upstream providers, they have *largest* customer cones [8], [49]. Since the main business role of tier-1 ASs (or more generally, of any transit provider) is to sell transit, they have a lot of connections to enterprise customers [Fig. 4(d) and (f)], but the open peering policy would not increase their revenue. Therefore, the tier-1 ASs are more likely to adopt restrictive peering policies. These considerations suggest that these ASs, contrary to a common belief, do not largely belong to the *densest* community, i.e., to $H_{k_{\mathrm{MAX}}}$. In particular, the ASs with the largest customer cones are not k_{MAX}-dense (Fig. 9).

In addition, we compute the overlaps between the set of the 60 ASs in the $H_{k_{\mathrm{MAX}}}$ core of the 2012 snapshot and the top-60 ASs ranked by their customer cone size. The results in Table VI show that these overlaps are not substantial. In the same table, we also show the overlap between the k_{MAX}-max dense ASs and the top-60 highest-degree ASs. From Fig. 5, we know that a high degree does not necessarily mean a high k-dense-index. In Table VI, we find that the overlap between these two AS sets is about 50%, and if we look back at Table V, we observe that these highest-degree or highest-rank ASs have quite different properties, compared to the k_{MAX}-dense ASs. Specifically, the former AS sets have the following:
- a higher percentage of global ASs [Table V(b)];
- a higher percentage of 10Gbps+ ASs [Table V(c)];
- a lower percentage of mostly outbound ASs [Table V(d)];
- a higher percentage of ASs having a restrictive peering policy [Table V(e)];
- a lower percentage of ASs having an open peering policy [Table V(e)].

Simply put, high-degree or high-rank ASs tend to be very large transit providers, while k_{MAX}-dense ASs tend to be either content providers or tier-2s.

D. 1k-Randomness of the $H_{k_{\mathrm{MAX}}}$ Core

If an AS has an *open peering policy*, it simply sets up a certain number of connections (equal to its degree), without *choosing* its neighbors in any way, e.g., based on their degrees or any other properties. On the contrary, an AS with a *selective/restrictive peering policy* always *chooses* its peers and one of the topological properties emerging from this selection process is the degree of the peering candidate, correlated with its customer cone size. In the latter case, we thus cannot expect that the degree distribution alone is sufficient to fully describe the graph. Some nontrivial degree correlations must be present in it, and indeed the Internet AS-level graph as a whole was found to be not $1K$- but $2K$-random in [23]. The large percentage of open-peering k_{MAX}-dense ASs, and the similarity between their customer cone sizes are thus the main factors explaining the $1K$-randomness of the $H_{k_{\mathrm{MAX}}}$ subgraph.

Let us assume for a moment that a large fraction of k_{MAX}-dense links is present at a single large IXP. We do not have access to the peering matrix data of any large IXPs, but the data reported in [11] suggest that this assumption may very well be correct, with DE-CIX being one of possible large IXP candidates. If we also assume that the peering policies declared in PeeringDB reflect the peering policies that each AS follows at each IXP where it has presence, then our explanation of $H_{k_{\mathrm{MAX}}}$'s $1K$-randomness is supported by the data in Fig. 9 and Table V(e).

VI. Conclusion

In summary, as the Internet grows over time, the maximum k-dense index grows as well. That is, the densest and innermost Internet core $H_{k_{\mathrm{MAX}}}$ becomes increasingly denser. Yet, this form of densification can only partially be attributed to the growing average degree. Even though the node degrees and k-dense-indices are correlated, the relative fluctuations are strong, and the two statistics reflect two different Internet-specific properties of AS nodes. High-degree and high-k-dense-index ASs tend to be transit and content providers, respectively. The latter form the densest community in the Internet, which only loosely overlaps with the high-rank ASs forming the tier-1 core. The structure of the $H_{k_{\mathrm{MAX}}}$ core is relatively simple. Statistically, it is almost fully determined by its degree distribution, a property that can be explained by open peering policies that many content providers and NSPs tend to follow at IXPs. Most importantly, after proper normalization, all the considered k-dense properties of the Internet appear time-invariant. In particular, a vast majority of all AS links in the Internet are attached either to ASs with the k-dense-index equal to 2 or 3, or to the k_{MAX}-dense ASs.

Speaking more generally, the Internet's k-dense properties, derived from a recursive variant of edge multiplicity measuring the frequency and density of triangle overlaps, appear to be a statistically significant and time-invariant structural properties that cannot be fully captured by either degree distribution ($1K$-distribution) or degree correlations ($2K$-distribution). The

$3K$-distribution, the distribution of subgraphs of size 3, may very well capture the Internet's k-dense properties. Even though this conjecture is quite plausible, it is not guaranteed to be correct because reproducing the frequency of degree-labeled triangles does not automatically imply that the whole k-dense decomposition hierarchy of nested subgraphs H_k is correctly reproduced as well. Therefore, it is interesting to investigate what existing or new Internet topology models and generators are capable of explaining or at least reproducing the k-dense properties of the Internet. In that respect, the recent results in [21] are interesting as they indicate that some basic k-dense properties of the Internet and other networks are fairly similar to those in random graphs that have the same degree and *clustering* distributions. A more detailed analysis is needed to see how accurately such graphs reproducing clustering capture the full spectrum of the k-dense properties considered here. Yet, it would not be too surprising if such an analysis does find that clustering indeed defines fairly accurately all the k-dense properties. Such results would not be surprising because clustering and k-dense properties are closely related, while the specifics of clustering are consequences of similarity forces driving network evolution, and taking these forces into account predicts remarkably well not only a long list of structural properties of different networks including the Internet, but also many details of the dynamics of their growth [15].

References

[1] M. Faloutsos, P. Faloutsos, and C. Faloutsos, "On power-law relationships of the Internet topology," *Comput. Commun. Rev.*, vol. 29, no. 4, pp. 251–262, Aug. 1999.

[2] G. Siganos, M. Faloutsos, P. Faloutsos, and C. Faloutsos, "Power laws and the AS-level Internet topology," *IEEE/ACM Trans. Netw.*, vol. 11, no. 4, pp. 514–524, Aug. 2003.

[3] A. Dhamdhere and C. Dovrolis, "Twelve years in the evolution of the Internet ecosystem," *IEEE/ACM Trans. Netw.*, vol. 19, no. 5, pp. 1420–1433, Oct. 2011.

[4] R. Pastor-Satorras and A. Vespignani, *Evolution and Structure of the Internet: A Statistical Physics Approach*. New York, NY, USA: Cambridge Univ. Press, 2004.

[5] B. Edwards, S. A. Hofmeyr, G. Stelle, and S. Forrest, "Internet topology over time," *CoRR* 2012, abs/1202.3993 [Online]. Available: http://arxiv.org/abs/1202.3993

[6] B. Huffaker, M. Fomenkov, and K. Claffy, "Internet topology data comparison," CAIDA, University of California, San Diego, CA, USA, Tech. Rep., 2012.

[7] B. Zhang, R. Liu, D. Massey, and L. Zhang, "Collecting the Internet AS-level topology," *Comput. Commun. Rev.*, vol. 35, no. 1, pp. 53–61, Jan. 2005.

[8] X. Dimitropoulos, D. Krioukov, M. Fomenkov, B. Huffaker, Y. Hyun, K. Claffy, and G. Riley, "AS relationships: Inference and validation," *Comput. Commun. Rev.*, vol. 37, no. 1, pp. 29–40, Jan. 2007.

[9] Y. He, G. Siganos, M. Faloutsos, and S. V. Krishnamurthy, "Lord of the links: A framework for discovering missing links in the Internet topology," *IEEE/ACM Trans. Netw.*, vol. 17, no. 2, pp. 391–404, Apr. 2009.

[10] R. Oliveira, D. Pei, W. Willinger, B. Zhang, and L. Zhang, "The (in)completeness of the observed Internet AS-level structure," *IEEE/ACM Trans. Netw.*, vol. 18, no. 1, pp. 109–122, Feb. 2010.

[11] B. Ager, N. Chatzis, A. Feldmann, N. Sarrar, S. Uhlig, and W. Willinger, "Anatomy of a large European IXP," in *Proc. ACM SIGCOMM*, New York, NY, USA, 2012, pp. 163–174.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

12   IEEE/ACM TRANSACTIONS ON NETWORKING

[12] E. Gregori, A. Improta, L. Lenzini, L. Rossi, and L. Sani, "On the in-completeness of the AS-level graph: A novel methodology for BGP route collector placement," in *Proc. IMC* , 2012, pp. 253–264.

[13] B. Krogfoss, M. Weldon, and L. Sofman, "Internet architecture evolution and the complex economies of content peering," *Bell Lab. Tech. J.*, vol. 17, no. 1, pp. 163–184, Jun. 2012.

[14] W. B. Norton, "Internet service providers and peering v3.0," 2012 [Online]. Available: http://drpeering.net/white-papers/Internet-Service-Providers-And-Peering.html

[15] F. Papadopoulos, M. Kitsak, M. Serrano, M. Boguñá, and D. Krioukov, "Popularity versus similarity in growing networks," *Nature*, vol. 489, pp. 537–540, Sep. 2012.

[16] D. Krioukov, M. Kitsak, R. Sinkovits, D. Rideout, D. Meyer, and M. Boguñá, "Network cosmology," *Nature Sci. Rep.*, vol. 2, no. 793, Nov. 2012.

[17] K. Saito, T. Yamada, and K. Kazama, "Extracting communities from complex networks by the k-dense method," *IEICE Trans. Fund. Electron., Commun. Comput. Sci.*, vol. E91-A, no. 11, pp. 3304–3311, 2008.

[18] E. Gregori, L. Lenzini, and C. Orsini, "k-dense communities in the Internet AS-level topology graph," *Comput. Netw.*, vol. 57, no. 1, pp. 213–227, 2013.

[19] M. A. Serrano and M. Boguñá, "Clustering in complex networks. I. General formalism," *Phys. Rev. E*, vol. 74, p. 056114, 2006.

[20] V. Zlatić, D. Garlaschelli, and G. Caldarelli, "Networks with arbitrary edge multiplicities," *Europhys. Lett.*, vol. 97, no. 2, pp. 28 005–p1–p5, 2012.

[21] P. Colomer-de Simon, A. Serrano, M. G. Beiro, I. Alvarez-Hamelin, and M. Boguna, "Deciphering the global organization of clustering in real complex networks," 2013 [Online]. Available: http://arxiv.org/abs/1306.0112

[22] E. Gregori, L. Lenzini, and C. Orsini, "k-dense communities in the Internet AS-level topology," in *Proc. 3rd COMSNETS*, 2011, pp. 1–10.

[23] P. Mahadevan, D. Krioukov, K. Fall, and A. Vahdat, "Systematic topology analysis and generation using degree correlations," *Comput. Commun. Rev.*, vol. 36, no. 4, pp. 135–146, Aug. 2006.

[24] G. Zhang, B. Quoitin, and S. Zhou, "Phase changes in the evolution of the IPv4 and IPv6 AS-level Internet topologies," *Comput. Commun.*, vol. 34, no. 5, pp. 649–657, Apr. 2011.

[25] R. V. Oliveira, B. Zhang, and L. Zhang, "Observing the evolution of Internet AS topology," in *Proc. SIGCOMM*, New York, NY, USA, 2007, SIGCOMM '07, pp. 313–324.

[26] S. B. Seidman, "Network structure and minimum degree," *Social Netw.*, vol. 5, pp. 269–287, 1983.

[27] G. Siganos, S. Tauro, and M. Faloutsos, "Jellyfish: A conceptual model for the AS Internet topology," *J. Commun. Netw.*, vol. 8, no. 3, pp. 339–350, Sep. 2006.

[28] M. Gonen, D. Ron, U. Weinsberg, and A. Wool, "Finding a dense-core in Jellyfish graphs," *Comput. Netw.*, vol. 52, no. 15, pp. 2831–2841, Oct. 2008.

[29] J. I. Alvarez-Hamelin, L. Dall'Asta, A. Barrat, and A. Vespignani, "K-core decomposition of Internet graphs: Hierarchies, self-similarity and measurement biases," *Netw. Heterogeneous Media*, vol. 3, no. 2, pp. 371–293, 2008.

[30] G. Palla, I. Derenyi, I. Farkas, and T. Vicsek, "Uncovering the overlapping community structure of complex networks in nature and society," *Nature*, vol. 435, no. 7043, pp. 814–818, 2005.

[31] G.-Q. Zhang, G.-Q. Zhang, Q.-F. Yang, S.-Q. Cheng, and T. Zhou, "Evolution of the Internet and its cores," *New J. Phys.*, vol. 10, no. 12, p. 123027, 2008.

[32] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks," *Proc. Nat. Acad. Sci.*, vol. 99, no. 12, pp. 7821–7826, Jun. 2002.

[33] C. Orsini, "A structural analysis of the Internet AS-level topology," Ph.D. dissertation, Univ. Pisa, Pisa, Italy, 2013.

[34] R. Oliveira, D. Pei, W. Willinger, B. Zhang, and L. Zhang, "Quantifying the completeness of the observed Internet AS-level structure," UCLA, Los Angeles, CA, USA, Tech. Rep. 080026, 2008.

[35] UCLA Computer Science Department's Internet Research Lab, Los Angeles, CA, USA, "Internet topology collection," 2012 [Online]. Available: http://irl.cs.ucla.edu/topology/

[36] K. Keys, Y. Hyun, M. Luckie, and K. Claffy, "Internet-scale IPv4 alias resolution with MIDAR," *IEEE/ACM Trans. Netw.*, vol. 21, no. 2, pp. 383–399, Apr. 2013.

[37] "CIDR report," 2012 [Online]. Available: http://www.cidr-report.org/as2.0/

[38] V. Colizza, A. Flammini, M. A. Serrano, and A. Vespignani, "Detecting rich-club ordering in complex networks," *Nature Phys.*, vol. 2, pp. 110–115, 2006.

[39] P. Erdos and A. Renyi, "On the evolution of random graphs," *Publ. Math. Inst. Hung. Acad. Sci.*, vol. 5, pp. 17–61, 1960.

[40] H. Kim, C. I. D. Genio, K. E. Bassler, and Z. Toroczkai, "Constructing and sampling directed graphs with given degree sequences," *New J. Phys.*, vol. 14, no. 2, p. 023012, 2012.

[41] C. Labovitz, S. Iekel-Johnson, D. McPherson, J. Oberheide, and F. Jahanian, "Internet inter-domain traffic," *Comput. Commun. Rev.*, vol. 41, no. 4, pp. 75–86, Aug. 2010.

[42] "PeeringDB," 2012 [Online]. Available: http://www.peeringdb.com/

[43] DE-CIX, Frankfurt, Germany, "DE-CIX, the Deutscher Commercial Internet Exchange," 2012 [Online]. Available: http://www.de-cix.net/

[44] CAIDA, the Cooperative Association for the Internet Data Analysis, La Jolla, CA, USA, "The CAIDA AS relationships dataset," Jun. 1, 2012 [Online]. Available: http://www.caida.org/data/active/as-relationships/

[45] W. B. Norton, "Content providers (aka content publishers)," 2011 [Online]. Available: http://drpeering.net/white-papers/Ecosystems/Content-Providers.html

[46] A. Lodhi, A. Dhamdhere, and C. Dovrolis, "Peering strategy adoption by transit providers in the Internet: A game theoretic approach?," *Perform. Eval. Rev.*, vol. 40, no. 2, pp. 38–41, Oct. 2012.

[47] W. B. Norton, "The top five motivations to peer," in *The Internet Peering Playbook: Connecting to the Core of the Internet*. Palo Alto, CA, USA: DrPeering Press, 2011.

[48] W. B. Norton, "The folly of peering ratios (as a peering candidate discriminator)," 2012 [Online]. Available: http://drpeering.net/white-papers/The-Folly-Of-Peering-Ratios.html

[49] CAIDA, the Cooperative Association for the Internet Data Analysis, La Jolla, CA, USA, "AS rank," 2012 [Online]. Available: http://as-rank.caida.org/

Chiara Orsini received the Master's and Ph.D. degrees in computer engineering from the University of Pisa, Pisa, Italy, in 2009 and in 2013, respectively.

She is a Post-Doctoral Researcher with the Cooperative Association for Internet Data Analysis (CAIDA), La Jolla, CA, USA. Her research interests focus primarily on studying the structural characteristics of the Internet topology at the autonomous-system level of abstraction.

Enrico Gregori (A'90) received the Laurea degree in electronic engineering from the University of Pisa, Pisa, Italy, in 1980.

In 1981, he joined the Italian National Research Council (CNR), Pisa, Italy, where he is currently a CNR Research Director. He has contributed to several national and international projects on computer networking. He has authored more than 100 papers in the area of computer networks and has published in international journals and conference proceedings. His current research interests include ad hoc networks, sensor networks, wireless LANs, quality of service in packet-switching networks, and evolution of TCP/IP protocols.

Luciano Lenzini (M'07) received the Laurea degree in physics from the University of Pisa, Pisa, Italy, in 1969.

He is a Full Professor in computer networking with the Faculty of Engineering, University of Pisa. He is the author and coauthor of a high number of papers published in journals and conference proceedings. He is the author of numerous industrial patents. He has directed several national an international projects in the area of computer networking. His current research interests include the design and performance evaluation of MAC protocols for wireless networks and the quality of service provision in integrated and differentiated services networks.

Dmitri Krioukov received the Diploma in Physics from the St. Petersburg State University, St. Petersburg, Russia, in 1993, and the Ph.D. degree in physics from Old Dominion University, Norfolk, VA, in 1998.

He started working in the networking industry with Dimension Enterprises, Herndon, VA, USA. Upon its acquisition in 2002 by Nortel Networks, he accepted a Research Scientist position with Nortel, Herndon, VA, USA. In 2004, he moved back to academia, and since then he has been a Senior Research Scientist with the Cooperative Association for Internet Data Analysis (CAIDA), University of California, San Diego (UCSD), La Jolla, CA, USA.