



**HAL**  
open science

# Emergence of Nestedness in the Public Internet Peering Ecosystem

Justin Loye, Sandrine Mouysset, Katia Jaffrès-Runser

► **To cite this version:**

Justin Loye, Sandrine Mouysset, Katia Jaffrès-Runser. Emergence of Nestedness in the Public Internet Peering Ecosystem. Proceedings of Complex Networks 2024, Dec 2024, Istanbul, Turkey. hal-04733587

**HAL Id: hal-04733587**

**<https://hal.science/hal-04733587v1>**

Submitted on 14 Oct 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License

# Emergence of Nestedness in the Public Internet Peering Ecosystem

JUSTIN LOYE\*, Internet Initiative Japan Research Laboratory, Japan

SANDRINE MOUYSSSET\*, Institut de Recherche en Informatique de Toulouse, Université Paul Sabatier, France

KATIA JAFFRÈS-RUNSER\*, Institut de Recherche en Informatique de Toulouse, Toulouse INP-ENSEEIH, France

Nestedness is a property of bipartite complex networks that has been shown to characterize the peculiar structure of biological and economical networks. In a nested network, a node of low degree has its neighborhood included in the neighborhood of nodes of higher degree. Emergence of nestedness is commonly due to two different schemes: i) mutualistic behavior of nodes, where nodes of each class have an advantage in associating with each other, such as plant pollination or seed dispersal networks; ii) geographic distribution of species, captured in a so-called biogeographic network where species represent one class and geographical areas the other one. Nestedness has useful applications on real-world networks such as node ranking and link prediction.

Motivated by analogies with biological networks, we study the nestedness property of the public Internet peering ecosystem, an important part of the Internet where autonomous systems (ASes) exchange traffic at Internet eXchange Points (IXPs). We propose two representations of this ecosystem using a bipartite graph derived from PeeringDB data. The first graph captures the AS [is member of] IXP relationship which is reminiscent of the mutualistic networks. The second graph groups IXPs into countries, and we define the AS [is present at] country relationship to mimic a biogeographic network. We statistically confirm the nestedness property of both graphs, which has never been observed before in Internet topology data. From this unique observation, we show that we can use node metrics to extract new key ASes and make efficient prediction of newly created links over a two-year period.

CCS Concepts: • **Computing methodologies** → **Model development and analysis**; • **Networks** → **Public Internet**; **Topology analysis and generation**.

Additional Key Words and Phrases: internet topology, nestedness, public peering

## ACM Reference Format:

Justin Loye, Sandrine Mouysset, and Katia Jaffrès-Runser. 2023. Emergence of Nestedness in the Public Internet Peering Ecosystem. 1, 1 (December 2023), 19 pages. <https://doi.org/10.1145/nnnnnnnn.nnnnnnnn>

## 1 INTRODUCTION

Nestedness is a graph property that characterizes the peculiar structure of certain real complex networks. These complex networks are composed of two classes of nodes, with interactions only observable between nodes of two different classes. They are thus modeled using a bipartite graph. Such a bipartite complex graph is nested when, for any pair of nodes  $(i, j)$  of the same class, if the degree of  $i$  is greater than the degree of  $j$ , then the neighborhood of  $j$  is included in the neighborhood of  $i$ . This property has been observed mainly in two types of biological bipartite networks: *i*) the so-called "mutualistic" networks [4], such as pollination or seed dispersal networks, where the nodes of each class have a mutual benefit to associate with each other and *ii*) the biogeographic networks representing the distribution of species over geographical sites [41]. Only more recently has this structural property been identified in non-biological

\*All authors contributed equally to this research.

Authors' addresses: Justin Loye, [jloye@ijj.ad.jp](mailto:jloye@ijj.ad.jp), Internet Initiative Japan Research Laboratory, Tokyo, Japan; Sandrine Mouysset, [sandrine.mouysset@irit.fr](mailto:sandrine.mouysset@irit.fr), Institut de Recherche en Informatique de Toulouse, Université Paul Sabatier, Toulouse, France; Katia Jaffrès-Runser, Institut de Recherche en Informatique de Toulouse, Toulouse INP-ENSEEIH, Toulouse, France, [katia.jaffres-runser@irit.fr](mailto:katia.jaffres-runser@irit.fr).

networks: in the international trade network [13], where it accounts for the ability of countries to produce products of increasing complexity, and in online communication systems [8] as well.

Here, by making an analogy with mutualistic and biogeographic networks, we question for the first time to our knowledge the concept of nestedness in data describing the topology of the Internet. The nestedness property being the result of a dynamic process of evolution, it can be leveraged to identify key players and predict the evolution of the network structure to some extent.

At the coarsest granularity, the Internet is composed of autonomous systems (ASes) that can be, for example, content providers, Internet service providers, or network service providers. They essentially interconnect with each other in a client-to-provider or peer-to-peer relationship, thanks to infrastructures that can be private or public. A significant amount of Internet traffic between ASes occurs in public infrastructures known as Internet Exchange Points (IXPs) [14]. IXPs are generally non-profit organizations that operate in metropolitan areas with the objective of facilitating the peering interconnections of the ASes that are members, in exchange for a fee from the ASes to cover their operating costs. Together, ASes and IXPs form what is known as the “public peering ecosystem”.

There is no single or complete view of the Internet topology, making it difficult to understand the Internet at the AS level. The Internet is notoriously difficult to measure [32, 44], and information is not disclosed by its constituents for competitiveness and security reasons. Current AS-level topology data [21] is known to miss most peering links [1, 21] due to their invisible nature [36], even though these links are a growing part of the Internet thanks to developments in the public peering ecosystem in recent years [9]. The dataset generally used by the scientific community to study the evolution of public peering is PeeringDB [42]. PeeringDB contains self-reported data about ASes, IXPs, and the membership of ASes at IXPs. This database is expressive enough to identify the major players in the peering ecosystem [10, 28], and can be modeled as a bipartite graph representing the AS [is a member of] IXP relationship [35]. PeeringDB is being collected on a daily basis by CAIDA [27], allowing for temporal studies.

In this work, we study the nestedness of the public Internet peering ecosystem by defining two graphs from the PeeringDB dataset. First, we know that ASes and IXPs gain mutual benefit from their association, with the former ones gaining access to novel ASes and the latter ones gaining attractiveness. Thus, we construct the AS [is member of] IXP graph to capture the essence of a mutualistic biological network. Second, we construct another graph on the basis of an analogy with the biogeographic networks. Suppose an AS expands its business to a new geographic area. Then, it is more likely that the AS will establish itself in a certain order over time: starting with the most densely populated or strategically important places and then expanding to less populated or strategic places. This results in a process of selective immigration similar to the one causing nestedness in biogeographic networks (cf. Fig. 1). We mimic such graph by grouping IXPs into countries, and defining an AS [is present at] country graph. If the analogies are relevant, we expect to find nested networks that, through the applications of the nestedness concept such as node ranking and link prediction, improve our understanding of the public peering ecosystem.

Our study makes the following contributions:

- We reveal the nested structure of the topology of two networks representing the public peering ecosystem using PeeringDB;
- We identify new key peering actors from metrics related to nestedness. These actors are for instance high degree ASes able to access hard to reach areas of the world;
- Following a 27-month temporal study, we show that nestedness can be leveraged to effectively predict new link creation.

Depending on the acceptance of the paper, we provide the code to reproduce the results presented hereafter and we make the tools implemented easy to use for the community.

This paper is organized as follows. In [section 2](#), we present background on the public peering ecosystem and nestedness. In [section 3](#), we extract two graphs from PeeringDB and detail the necessary preconditioning for a nestedness study. The nestedness of both graphs is then assessed in [section 4](#). Finally, we study in [section 5](#) two applications of nestedness: new key ASes identification and link prediction.

## 2 BACKGROUND

### 2.1 Mathematical notations

Next, we consider that a graph  $G(E, V)$  is composed of a set of edges or links  $E$ , and a set of vertices or nodes  $V$ . It is described by an adjacency matrix  $A$ . A graph is said to be bipartite if all links are shared between two disjoint subsets of nodes, and if the graph is not directed, it can be fully described by its bi-adjacency matrix  $B$ . We denote by  $d(i)$  the degree of a node  $i$ , i.e. its number of links. A graph is said to be connected if there exists a succession of links connecting every pair of nodes.

### 2.2 PeeringDB

PeeringDB is the authoritative source of information about the public peering ecosystem. It is a non-profit and user-maintained database made to facilitate interconnections of ASes at IXPs. It consists in self-reported information about ASes (ASN, business type, traffic levels, ratio of inbound or outbound traffic, ...), IXPs (facilities, country, ...), and the membership of ASes at IXPs. As of March 1st, 2021, 869 IXPs and 21388 ASes are registered.

Although it is a self-reported database, PeeringDB is considered to be up-to-date and correct. The data is now mostly uploaded automatically and it is in ASes best interests to give correct information. As stated in [10], “[PeeringDB] has a very good standing in the network operators community, which naturally has very big interest in having reliable peering information available”. The authors give examples of important organisations that rely on PeeringDB like Google, Netflix and Cloudflare. Comparing three years of PeeringDB snapshot with BGP data, authors in [27] show that “PeeringDB membership is reasonably representative of the Internet’s transit, content, and access providers in terms of business types and geography of participants”. Thus, in line with these previous works, we consider PeeringDB as ground truth.

However, it’s important to note that PeeringDB only tells us about the public peering ecosystem, which is only a sub-part of the inter-domain Internet. Although IXPs are an important part of today’s Internet, enabling local traffic exchange, tier 1 ASes bypassing [9] and CDN deployment [33], PeeringDB does not report on private network interconnects (PNIs). PNIs are a preferred means of interconnection for ASes with the means to do so. For example, Tier-1s use PNIs because, for them, public peering means losing a customer. This is also the case for CDNs: Facebook (Meta) favors connections via PNI rather than IXP for reasons of monitoring and congestion control [45].

### 2.3 Nestedness

Nestedness is a graph property commonly found in real world data such as biological and economical networks. Formally, a graph is perfectly nested when, for any pair of nodes  $(i, j)$ , if the degree of  $i$  is greater than the degree of  $j$ , then the neighborhood of  $j$  is included in the neighborhood of  $i$ . If the graph is bipartite, as is often the case in nestedness studies, then the same condition holds for any pair of nodes within the same class (cf. [Fig. 1-\(b\)](#)). As a consequence, it is

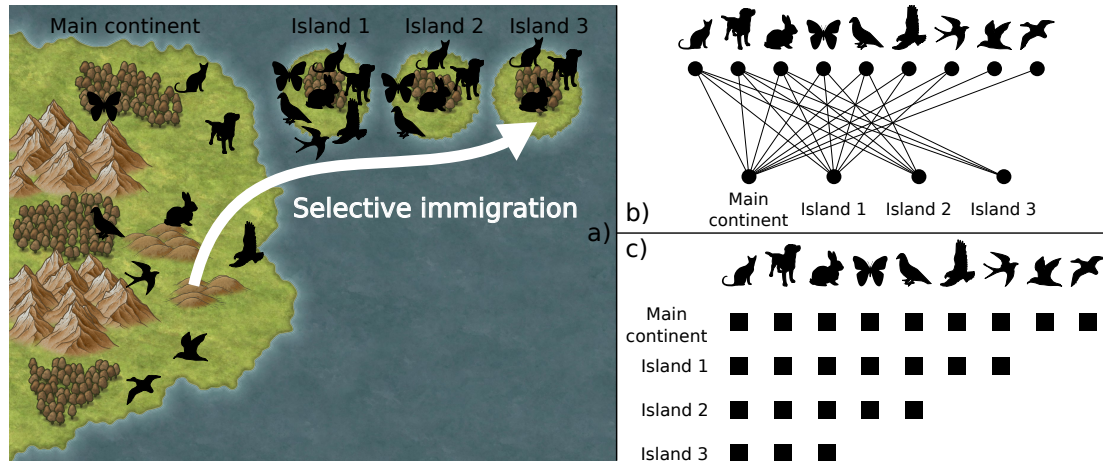


Fig. 1. Illustration of a perfectly nested biogeographic network. a) The selective immigration process causing the nested structure b) Bipartite biogeographic network c) A possible permutation of the rows and columns of the bi-adjacency matrix (here ordered by decreasing degree) showing the characteristic triangular structure.

possible to find an arrangement of rows and columns such as its bi-adjacency matrix presents a triangular shape on the top left corner (cf. Fig. 1-(c)). Such a structure is representative of a perfectly nested network. For a recent nestedness literature overview, we refer the reader to [30].

Nestedness is a property characterized at different levels. At the graph-level, it consists in quantifying how much the graph is close to a perfectly nested graph. After that, one might be interested in node-level nestedness that consists in finding the arrangement of rows and columns allowing to reveal the triangular structure in the adjacency matrix. This ordering represents a new ranking of the nodes in terms of their individual contribution to nestedness. And finally, one might look for nestedness at the partition-level to identify communities where nodes arrange themselves in a nested manner.

Nestedness was first conceptualized to explain the geographic distribution of animal species resulting from a selective immigration process. Such an immigration process is illustrated in Fig. 1-(a): a set of species living on a main continent immigrates to islands by crossing only small distances of water. As a result, the species farthest from the main continent are necessarily a subset of the nearest species, inducing the nested structure. This structure has been first observed in a biogeographic network representing the distribution of mammals in archipelagos [41]. This distribution is best represented by an unweighted bipartite graph, with nodes in classes “mammalians” or “archipelagos” and a link signifying the presence of species at geographical sites. It was then observed in other biological bipartite networks where the nodes of each class mutually benefit from their association. These so-called “mutualistic” networks are for example plant pollination networks or seed dispersal networks. In these two examples, the nodes of the first class are plants whose reproduction is facilitated, while the nodes of the second class are animals finding food.

Since then, nestedness has drawn the attention of network scientists for its useful applications. In a nested network, the node ordering w.r.t. to their individual contribution to nestedness is the best strategy of attack to disconnect the network [16]. Moreover, new links are more likely to appear in the characteristic upper triangular structure, therefore it can be leveraged for link prediction [13].

### 3 MODELING PEERINGDB FOR A NESTEDNESS STUDY

Motivated by analogies with biological nested networks, we study the nestedness of two networks derived from PeeringDB data. Since ASes and IXPs derive mutual benefit from their association, the AS [is member of] IXP graph is reminiscent of a mutualistic biological network. The other analogy stems from an intuition on the way ASes expand their reach in the network to new geographic areas. They may start with densely populated areas or strategically important places to maximize revenue, and later develop to less populated or less strategic places. This results in a process of selective immigration similar to the biogeographic network illustrated in Fig. 1). We mimic such graphs by grouping IXPs into countries, and defining an AS [is present at] country graph.

In order to analyze the nestedness property of both graphs, it is necessary to ensure that the graphs are well conditioned, i.e. that their adjacency matrix respects the same properties as the ones manipulated in previous works on nestedness.

#### 3.1 Constraints on the bi-adjacency matrix

The nested biological networks commonly studied in the literature are present in the database *web of life* [25]. Most of them are bipartite, with bi-adjacency matrix presenting the following properties [12]:

- The ratio between the number of rows  $N_r$  and columns  $N_c$  is balanced (there is an equilibrium between the number of nodes in both classes). As a consequence, the eccentricity  $e = |N_r - N_c| / (N_r + N_c)$  should be between 0 and 0.5.
- The matrix has a small size  $N = N_r + N_c$  (at most 1000 nodes in both classes). As a consequence, numerous tools related to the evaluation of nestedness do not scale up.
- The matrix is neither too sparse nor too dense. Since nestedness is based on the comparison of overlapping neighborhoods, graphs with many low-degree nodes are penalized when detecting nestedness, even if these low-degree nodes do not play an important role in the overall structure of the graph. The scarce studies involving large networks therefore filter out the nodes with the fewest connections [38, 40]. We stress that filtering nodes is also a widespread practice in other fields. For example the bipartite user [has rated] movie graph from the movielens100k dataset [23], widely used to test recommendation algorithms, contains only users with a degree superior to 20.

In the rest of this section, we present how the AS [is member of] IXP and AS [is present at] country graphs are built and conditioned to meet these requirements.

#### 3.2 AS-IXP graph

The AS [is member of] IXP graph, referred to as *AS-IXP graph* in the remainder of this paper, is a bipartite graph composed of 11407 ASes reporting 31889 associations to 780 IXPs. It is obtained with PeeringDB membership data as proposed in [35], for which we weight the links with the port size attribute. We consider only the main connected component, which comprises more than 99% of the total nodes.

To meet the constraints outlined earlier, we study in Fig. 2 the evolution of adjacency matrix properties (eccentricity and network size  $N$ ) according to node filters based on degree and weighted degree (we use for the later the terminology “port capacity” introduced in [10] as the sum of the port sizes of an AS’ links). As highlighted in red in Fig. 2, we conclude to filter out the nodes with a degree  $d \leq 3$ . The main reason for this is to reduce the eccentricity to 0.5 while having the least restrictive filter possible. The chosen filter has the advantage as well of making the network small

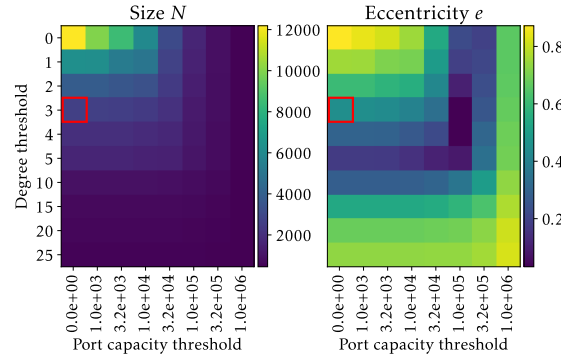


Fig. 2. Impact of filtering on properties of the AS-IXP graph. The chosen filter  $d \leq 3$  is highlighted in red.

enough for a nestedness study, with 2681 nodes. It reduces the number of nodes with low degree and therefore penalizes the nestedness detection, which is particularly necessary since half of the ASes in PeeringDB have a single connection. We do not apply a filter on the port capacity because its effect is less pronounced than the one of the filter on degrees, and could result in the removal of high degree ASes with low associated port sizes. The resulting graph presents 1942 ASes connected to 739 IXPs with 17689 links.

### 3.3 AS-country graph

The AS [is present at] country is referred to as the *AS-country graph* in the following. We construct the AS-country graph by merging, in the unfiltered AS-IXP graph, the IXPs by their reported countries. We obtain the presence of 11407 ASes in 112 countries, inducing a strong imbalance between both classes. To have a balanced ratio, we select the most central ASes in the graph according to three metrics. First, we select the top 100 ASes in decreasing order of degree in the AS-IXP graph. Second, we select the top 100 ASes in decreasing order of PageRank and reverse PageRank in a weighted and directed graph model of the AS-IXP graph as proposed in [28]. In this paper, authors leverage the router’s port size and the reported traffic balance of ASes (Heavy or Mostly Inbound or Outbound, Balanced) of PeeringDB to build a weighted and directed bipartite graph that captures the communication capacity of each link. PageRank (resp. reverse PageRank) metric applied to this graph can be used to identify nodes concentrating (resp. emitting) traffic. The final set of ASes is the union of the three top 100 sets. The resulting set is composed of 182 ASes.

In the following analysis, we only work with the unweighted version of the two AS-country and AS-IXP graphs. We do this for reasons of convenience: i) nestedness studies are mainly carried out on unweighted graphs, and are therefore more mature ii) many nestedness tools used don’t take weights into account iii) to simplify the scope of our analysis (for example, in the link prediction study, predicting the link weight is more difficult than predicting the presence of a new link). From an application point of view, this restriction amounts to working only on the absence and presence of ASes, which we argue is still valuable information for stakeholders.

## 4 NESTEDNESS ASSESSMENT

In this section, we characterize for the first time to our knowledge the presence of nestedness in Internet topology data. We first introduce the methods used to assess nestedness that are based on shared best practices outlined in [50],[5] and [30].

#### 4.1 Detecting nestedness

*Statistical significance of metrics.* To evaluate the nestedness of a network, it is considered good practice to use several metrics. Most metrics do not include a null model and return a value that, in absolute terms, has no clear interpretation. It is therefore necessary to compare this value to a null ensemble consisting of a large number of nestedness measures on random networks sharing properties with the network to be evaluated. We generate the null ensembles with the null model Proportional-Proportional (PP) that is commonly used since it offers a balance between errors of type I (identifying non-nested networks as nested) and type II (rejecting nested networks) [30]. This model is also used in community detection in the popular modularity measures of Newman [22] and its bipartite counterpart [3]. State-of-the-art implementations of this model are known as “Bascompte’s PP” [4] and “corrected PP” [39], the latter being a correction of the widely used Bascompte’s PP aiming to further reduce the number of type I errors. Random networks generated by those models have nodes that share on average the same degree as the nodes of the network to test.

Once a null ensemble is generated, the nestedness statistical significance of the network to be evaluated is assessed with the  $p$ -value and the  $z$ -score [50]. The  $p$ -value is the fraction of the null ensemble with higher nestedness than the test network. If there is no network in the null ensemble that is more nested than the network to be tested, then we conservatively set the value of  $p$  to be the inverse of the ensemble size, as proposed in [5]. A  $p$ -value less than 0.05 is considered sufficient to say that the network is more nested than the statistical set [5]. On the other hand,  $p > 0.95$  indicates an anti-nested pattern. The  $z$ -score is the difference between the test network measure and the mean of the null ensemble measures, normalized by its standard deviation.

*Graph-level metrics.* Graph-level nestedness metrics quantify how much a network deviates from a perfectly nested structure. They leverage the structure of the  $N_r \times N_c$  bi-adjacency matrix, where  $N_r$  and  $N_c$  are respectively the number of nodes in both classes. By convention, we will refer to nodes in one class as the “row nodes”, the others being the “column nodes”. In the following, we report the three main metrics whose descriptions are given in the reference [30].

The first metric we consider is the popular metric called *Nestedness metric based on Overlap and Decreasing Fill*, noted NODF [2]. Given a pair of row nodes  $(i, j)$  with degrees such that  $\mathbf{d}(i) > \mathbf{d}(j)$ , we expect that their number of common neighbors  $O_{i,j} = \sum_{\alpha} A_{i,\alpha} A_{j,\alpha}$  is equal to  $\mathbf{d}(j)$  for a perfectly nested network, or smaller than  $\mathbf{d}(j)$  otherwise. The total NODF value  $\eta$  is calculated from the NODF value of each class. The NODF value per row  $\eta^R$  is defined by:

$$\eta^R = \sum_{i,j} \frac{O_{i,j}}{\mathbf{d}(j)} \Theta(\mathbf{d}(i) - \mathbf{d}(j)), \quad (1)$$

where  $\Theta$  is the Heaviside function, defined by  $\Theta(x) = 1$  if  $x > 0$ ,  $\Theta(x) = 0$  if  $x \leq 0$ . This function allows to count only the pairs of  $(i, j)$  row nodes such that  $\mathbf{d}(i) > \mathbf{d}(j)$ . The maximum contribution of a pair is 1. Thus, the maximum value of  $\eta^R$  is equal to the number of pairs of row nodes  $N_r(N_r - 1)/2$ , with  $N_r$  the number of row nodes. In the same way, the NODF metric per column  $\eta^C$  is given by :

$$\eta^C = \sum_{\alpha,\beta} \frac{O_{\alpha,\beta}}{\mathbf{d}(\beta)} \Theta(\mathbf{d}(\alpha) - \mathbf{d}(\beta)), \quad (2)$$

where  $O_{\alpha,\beta} = \sum_i A_{i,\alpha} A_{i,\beta}$ . The maximum attainable value is  $N_c(N_c - 1)/2$ , with  $N_c$  the number of columns. Finally, the nestedness value of a network is given by the total NODF which is the ratio between the NODF value computed on



the graph  $\eta^R + \eta^C$  and the maximum attainable value:

$$\eta = \frac{\eta^R + \eta^C}{\frac{N_r(N_r-1)}{2} + \frac{N_c(N_c-1)}{2}}. \quad (3)$$

The second metric we consider is a variant of NODF taking into account a null model proposed in [46] :

$$\tilde{\eta} = \frac{2}{N_r + N_c} \left\{ \sum_{i,j} \frac{O_{i,j} - \langle O_{i,j} \rangle}{(N_r - 1)d(j)} \Theta(d(i) - d(j)) + \sum_{\alpha,\beta} \frac{O_{\alpha,\beta} - \langle O_{\alpha,\beta} \rangle}{(N_c - 1)d(\beta)} \Theta(d(\alpha) - d(\beta)) \right\}, \quad (4)$$

with the expectation of the number of common neighbors  $\langle O_{i,j} \rangle$  given by considering that the neighbors of each node are drawn randomly, respecting their degree. In this case, the probability that the two nodes have a common link is  $\langle O_{i,j} \rangle = d(i)d(j)/N_c^2$ . The same reasoning holds for any pair of  $\alpha, \beta$  nodes of the other class. If the network is not more nested than the null model, then  $\tilde{\eta} \leq 0$ , and the closer  $\tilde{\eta}$  is to its maximum value of 1, the more nested the network is.

Lastly, we consider the metric known as spectral radius. It has been shown in [6] that the spectral radius  $\rho$  of a matrix, i.e. the eigenvalue with the highest modulus, is maximal when this matrix is perfectly nested. Staniczenko et al. [47] statistically study the behavior of  $\rho$  on matrices with imperfect nestedness, to conclude that  $\rho$  can be used to quantify the nestedness of real complex networks.

#### 4.2 Nestedness in AS-country graph

The results of the statistical tests of nestedness for the AS-country graph are given in Table 1a. The  $p$ -value, associated with a high  $z$ -score, shows that no null ensemble provided a more nested network than the AS-IXP graph. We can conclude that the AS-country network is highly nested.

Although this is outside the scope of our work, we have also evaluated the nestedness of the weighted version of this graph, which takes into account the sum of port sizes that an AS has in a country. Our goal is to assess that the nested structure is not an artefact caused by making the graph unweighted, i. e. by the binarization of the adjacency matrix. It is indeed possible for a binary bi-adjacency matrix to be nested, but to have a weight distribution on the triangular structure following a non-nested or even an anti-nested pattern (for an illustration we refer the reader to Fig. 2 in [47]). To do this, we use the methodology introduced in [47], which consists in choosing a metric and a null ensemble suitable for weighted graphs. We therefore use the spectral radius metric on the weighted adjacency matrix, and also measure it on an ensemble of 10000 random graphs generated by a null model keeping the source and destination of the original links fixed but randomly permuting the weights between links. We find that the graph is strongly nested, as no random graph is more nested than it ( $p$ -value of 0.0001) and the  $z$ -score is high at 11.16. We can therefore conclude that the observed nestedness is not an artefact caused by making the graph unweighted.

#### 4.3 Nestedness in AS-IXP graph

For the AS-IXP graph, the results presented in Table 1b are quite mixed. Among the six combinations of null ensembles and metrics, four indicate a strongly nested network with  $p < 0.05$  while two identify the network as strongly anti-nested with  $p$  close to 0.95. Thus, we cannot conclude on the nested structure of the AS-IXP graph.

However, it has been shown in [28] that the IXPs membership network exhibits a community structure that correlates with the geographic proximity of ASes and IXPs. Thus, one might wonder if nestedness should not rather be sought in the network's communities, i.e. at the level of the partition. To do so, we leverage a metric recently developed in [46] to quantify the presence of nested communities, also called "blocks" due to the resulting structure in the adjacency matrix.

Table 1. Statistical test results, presented as the  $p$  value and  $z$  score in parentheses, of the graph-level nestedness.

(a) AS-Country				(b) AS-IXP			
Nestedness metric	$\tilde{\eta}$	NODF	spectral radius	Nestedness metric	$\tilde{\eta}$	NODF	spectral radius
Null model				Null model			
PP (Bascompte)	0.001 (80.59)	0.001 (61.04)	0.001 (29.97)	PP (Bascompte)	0.001 (103.94)	0.001 (86.61)	0.001 (94.58)
PP (corrected)	0.001 (27.34)	0.001 (20.93)	0.001 (11.42)	PP (corrected)	0.945 (-1.63)	0.977 (-1.94)	0.001 (39.68)

#### 4.4 In-Block Nestedness

This metric called In-Block Nestedness (IBN) is an adaptation of the metric  $\tilde{\eta}$  (Equation 4) also introduced in [46]. It can be optimized in the same way as the classical modularity [18, 22] to obtain a network partition.

We first define the number of nodes in the same class and community of a row node  $i$  by  $C_i = \sum_j \delta(g(i), g(j))$ , where  $g(i)$  refers to the community of a node  $i$  and  $\delta$  is the Kronecker delta, i.e.  $\delta(i, j) = 1$  if  $i = j$ ,  $\delta = 0$  otherwise. The same equation holds for a column node  $\alpha$ :  $C_\alpha = \sum_\beta \delta(g(\alpha), g(\beta))$ . The number of common neighbors now includes community membership:

$$O_{i,j} = \sum_{\alpha=1}^{N_r} A_{i,\alpha} A_{j,\alpha} \delta(g(i), g(\alpha)), \quad O_{\alpha,\beta} = \sum_{i=1}^{N_c} A_{i,\alpha} A_{i,\beta} \delta(g(i), g(\beta)). \quad (5)$$

The value of IBN  $\tilde{I}$  is then

$$\tilde{I} = \frac{2}{N_r + N_c} \left\{ \sum_{i,j}^{N_r} \left[ \frac{O_{i,j} - \langle O_{i,j} \rangle}{\mathbf{d}(j)(C_i - 1)} \Theta(\mathbf{d}(i) - \mathbf{d}(j)) \delta(g(i), g(j)) \right] + \sum_{\alpha,\beta}^{N_c} \left[ \frac{O_{\alpha,\beta} - \langle O_{\alpha,\beta} \rangle}{\mathbf{d}(\beta)(C_\alpha - 1)} \Theta(\mathbf{d}(\alpha) - \mathbf{d}(\beta)) \delta(g(\alpha), g(\beta)) \right] \right\}. \quad (6)$$

In order to find nested communities, the authors of [46] optimize  $\tilde{I}$  with the extremal optimization (EO) method [7, 17]. In the present work, we use the implementation of this algorithm made available by the authors of [37] in [15]. We will compare this nested community partition to the classical notion of strongly connected community partition, the latter being obtained by optimizing Barber's modularity [3] with the same implementation of the EO algorithm [15, 37].

The composition of the largest communities obtained by modularity and IBN optimization is given in Table 2 and 3 respectively. We can observe that the modularity tends to partition the network in communities of similar sizes. The 15 hypergiant ASes identified in [10]<sup>1</sup>, i.e. ASes leveraging IXPs to have a global reach, are split between communities 0 (40%) and 7 (60%), which does not capture their actual global footprint on the Internet. For the IBN on the contrary, a large portion of the nodes and all of the hypergiants are grouped in a single community. If we consider the port capacity metric, introduced in [10] as the sum of all port sizes reported in PeeringDB for each AS, this community is even more predominant, containing ASes that account for 77% of total port capacity. The connected subgraph induced by this community is highly nested, with the results presented in Table 4 showing the lowest possible  $p$ -value and a high  $z$ -score. Motivated by these observations, we will study in the following this community which we call the main nested component of the AS-IXP graph.

<sup>1</sup>namely Apple Inc, Amazon.com, Facebook, Google Inc. Akamai Technologies, Yahoo!, Netflix, Hurricane Electric, OVH, Limelight Networks Global, Microsoft, Twitter Inc., Twitch, Cloudflare and Verizon Digital Media Services.

Table 2. Community composition of the modularity optimization by the extremal optimization algorithm.

Community	ASes proportion (%)	IXPs proportion (%)	Port capacity proportion (%)	Unique countries	Countries distribution
6	19.0	8.9	11.2	21	FR: 14   GB: 9   IT: 8   CZ: 5   BG: 4
0	15.1	16.0	36.4	25	US: 68   HK: 5   TZ: 5   CA: 4   PH: 4
10	11.7	7.3	4.0	3	ID: 25   DE: 15   UA: 14
4	11.3	11.0	1.9	27	AU: 15   DE: 12   CN: 5   NL: 5   RO: 5
1	11.0	11.9	9.3	12	BR: 36   US: 29   CA: 8   EC: 3   AO: 2
7	9.6	21.0	25.4	71	IN: 18   PL: 11   JP: 10   US: 9   ES: 6
16	7.0	7.7	4.3	9	RU: 34   US: 7   LT: 4   EE: 3   LV: 3
12	6.5	5.7	2.5	9	AU: 24   NZ: 6   CL: 3   BO: 2   CN: 2

Table 3. Same as Table 2 for communities obtained by the optimization of In-Block Nestedness  $\tilde{I}$ .

Community	ASes proportion (%)	IXPs proportion (%)	Port capacity proportion (%)	Unique countries	Countries distribution
1	44.4	45.6	77.3	100	US: 86   DE: 14   CA: 12   IN: 11   GB: 9
98	7.0	3.1	3.0	15	ID: 3   US: 3   BO: 2   DE: 2   GB: 2
31	5.6	1.9	2.3	3	AU: 12   AR: 1   US: 1
38	5.3	4.6	4.0	2	BR: 33   PT: 1
19	5.1	3.0	1.1	3	ID: 20   AR: 1   HK: 1
14	5.1	4.1	2.3	4	RU: 26   UA: 2   PH: 1   UZ: 1
53	5.0	2.2	0.3	11	DE: 4   CH: 2   US: 2   CA: 1   CN: 1

Table 4. Same as Table 1 for the community 0 of the AS-IXP graph.

Null model	Nestedness metric	NODF	spectral radius
	$\tilde{\eta}$		
PP (Bascompte)	0.001 (144.36)	0.001 (113.21)	0.001 (56.31)
PP (corrected)	0.001 (27.97)	0.001 (19.53)	0.001 (16.79)

## 5 USE CASES

Nestedness has been exhibited for the AS-country and AS-IXP networks, but the triangular structure in the bi-adjacency matrix which is characteristic of nestedness is yet to be revealed. Such a structure is obtained by a relevant ordering of the rows and columns of the bi-adjacency matrix, and has several applications:

- The derivation of a new and original ranking of nodes that offers a novel interpretation of their importance in the network;
- The prediction of link appearance in the network based on the idea that a new link is more likely to appear in the triangular structure of the matrix.

In this section, we investigate both applications for the two PeeringDB networks. Thus, we first present the node-level nestedness metrics derived by unveiling the triangular graph structure following recent literature methods [30]. It allows us to identify new important nodes of the peering ecosystem. Next, we show how leveraging this triangular structure allows the effective prediction of newly created links.

### 5.1 Node-level metrics

The simplest node-level metric we can leverage to re-order rows and columns of the bi-adjacency matrix to reveal nestedness is the node degree. It is a classical metric fulfilling the decreasing degree condition of nestedness, and has a simple interpretation. However, its definition does not take into account the second and only other property of nestedness: the neighborhoods' overlap.

Another possible ordering of nodes is given by the popular state-of-the-art BINMATNEST algorithm [43]. The authors propose a graph-level nestedness metric sensitive to the nodes ordering in the matrix. A genetic algorithm is deployed to maximize this graph-level metric by swapping the rows and columns. The ranking of nodes is given by the row or the column indices observed in the optimal isomorphism. This metric clearly aims at accounting for the overlap between neighborhoods property.

Lastly, we introduce the fitness-complexity metric that also produces state-of-the-art ordering [26], but which has the advantage over BINMATNEST of having a clear interpretation and a simple and efficient implementation. It is in the context of economic complexity, first introduced by economists to analyze the country [export] product bipartite network, that the fitness-complexity metric came into being [49]. This approach views the export of a product as the result of a production process that requires a country to have all the capabilities necessary to manufacture the product. In this view, countries with more capacity are more competitive in world trade, as they have the ability to produce and export more products than countries with less capacity. In addition, a notion of complexity is added to products to show that they can only be produced by competitive countries. This notion is itself taken into account for competitiveness: a competitive country exports complex products. This approach is modeled by coupled, non-linear and recursive fitness and complexity equations. The fitness equation measures the economic health of a country based on the quality of its exported products and the complexity equation measures the quality of the products based on the economic health of the countries that export them and the number of countries that export them. For a network described by the adjacency matrix  $\mathbf{A}$ , the **fitness**  $f(i)$  of a country  $i$  is defined as the sum of the complexity scores  $q(\alpha)$  of the products  $\alpha$  exported by this country:

$$f(i) = \sum_{\alpha} A_{i,\alpha} q(\alpha), \quad q(\alpha) = \frac{1}{\sum_i A_{i,\alpha} / f(i)}. \quad (7)$$

The scores of the products depend on the number of countries that export them but also on the economic health (fitness) of the countries that produce them. A competitive country generally exports complex products, a less developed country exports products that are easier to produce. The complexity measure is based on these assumptions. To solve these recursive and coupled sequences, we compute iteratively :

$$f_n(i) = \sum_{\alpha} A_{i,\alpha} q_{n-1}(\alpha), \quad q_n(\alpha) = \frac{1}{\sum_i A_{i,\alpha} / f_{n-1}(i)}, \quad (8)$$

beginning with uniform non-null vectors  $f_0$  and  $q_0$ . At each iteration, these vectors are divided by their mean. For a better convergence, we use a slightly modified version of the complexity in Equation 8 as proposed in [48]:

$$q_n(\alpha) = \frac{1}{\sum_i A_{i,\alpha} (1 - f_{n-1}(i))}. \quad (9)$$

Ordering countries and products by these rankings reveals the triangular structure inherent to nestedness. Since its publication, this algorithm has been used outside the context of economics, notably in biology, to measure the nestedness of nodes and to rank them from most to least nested.

## 5.2 New key players

We have seen that there are several nestedness metrics at the node level. In addition to the traditional degree metric which satisfies the first nestedness condition, namely a degree decrease, there are two specific nestedness metrics known as BINMATNEST and fitness-complexity. After verifying that all metrics reveal the characteristic triangular structure of nestedness, we ask here whether BINMATNEST and fitness-complexity provide an original view compared to the degree by identifying new key ASes, and what this teaches us about the public peering ecosystem.

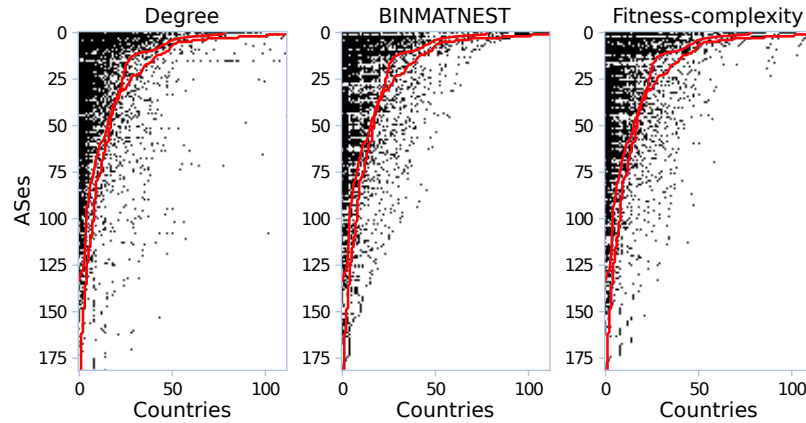


Fig. 3. Ordering of the bi-adjacency matrix of the AS-country graph for different node-level nestedness metrics.

We show the AS-country bi-adjacency matrix ordered by different node-level nestedness metrics in Fig. 3. All metrics reveal a triangular structure characteristic of nestedness. The links are concentrated in a stylized upper left triangle, with a convex hypotenuse caused by the filling rate of the matrix. The fill rate is the ratio between the number of existing links and the number of possible links, i.e.  $\phi = \sum_{i,j} B_{i,j} / (N_r \times N_c)$ . Depending on its value, the hypotenuse of a perfectly nested biadjacency matrix is either a straight line ( $\phi = 0.5$ ), convex ( $0 \leq \phi < 0.5$ ) or concave ( $0.5 < \phi \leq 1$ ). To visualize how marked the AS-IXP triangular structure is, we compare it to a perfectly nested network with the same number of links. The hypotenuse of a perfectly nested network is given by the diversity-ubiquity curves [13]. The diversity and ubiquity curves are respectively the curves giving the degree of the AS and the countries. They are represented in red on the figure. We can see that the density of the links is indeed higher on the left of these curves than on the right, thus giving a visual indication of nestedness for all of the metrics considered.

We find a difference between the degree metric on one side, and the BINMATNEST and fitness-complexity metrics on the other side. Degree ordering has a dense triangle of links but some of them are far to the right of the diagonal, while the other two have no links to the right of the diagonal at the cost of a less dense triangle. This means that the fitness-complexity and BINMATNEST rankings, specific to nestedness, provide a new ranking that is original compared to the traditional degree one. The disagreements in the rankings are reported in Table 5a that lists the top 30 ASes for all three rankings. New players characterized by a large gap between both their nestedness ranking and their degree ranking appear. Facebook’s secondary ASN (now Meta), that is used to distribute its most popular content [31] efficiently through offnet deployment as close as possible to end-users [20], ranks significantly better according to BINMATNEST and fitness, with +13 and +14 increase respectively over the degree. It enters the top 4 AS, and ahead of Facebook’s main AS, which may seem peculiar. Other AS have a significant increase in ranking, such as network service providers ROSTELECOM (+80 and +91) and NETIX communication (+16 and +32). According to the fitness-complexity metric, these new players are interpreted as being able to access hard-to-reach areas of the world, which can not be captured in degree ranking.

Although hard to validate, we believe most of these results are consistent from an operational point of view. Facebook’s secondary AS is used to distribute its most popular content in cached form, which we assume is essential in remote areas where bandwidth is limited. NETIX communication offers global peering solution and advertise its “global reach”

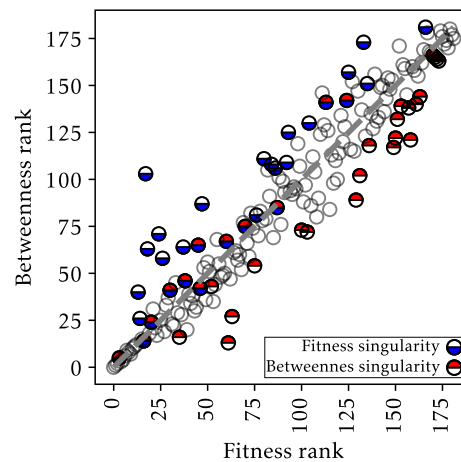


Fig. 4. Correlation between the fitness and betweenness rankings of nodes in the AS-country graph. Singularities, defined as a ranking deviation of at least 10 with respect to the degree, are highlighted.

[34], so they might be specialized in hard-to-reach regions. In the case of ROSTELECOM, we attribute the large increase to a low degree coupled with its presence in both Europe and Asia.

We also report the appearance of new actors in the AS-IXP graph with Table 5b<sup>2</sup>. Among them, we find operators related to domain resolution (DNS) such as RIPE NCC K-Root Operations (improvements over the degree of +30), DNS-OARC-112 (+29), Netnod (+11). The RIPE NCC K-Root Operations and Netnod ASes are associated with the 13 root name servers and thus play a structuring role in the global Internet, which is clearly revealed in this new ranking. These AS are well ranked because they extend their presence to isolated IXPs.

To assert the originality of the fitness metric, we compare it with the betweenness metric proposed in [19], which is fundamental in network science while being original relative to the degree metric. The betweenness centrality quantifies the proportion of shortest paths between all pairs of nodes in a network that pass through a particular node, indicating its importance as a bridge between sub-structures of the network [24]. We plot against each other in Fig. 4 the AS ranks based on the betweenness and the fitness metrics. First of all, we see that the two rankings are correlated, with points distributed along the diagonal. However, when we highlight the ASes with a deviation of at least 10 between one of their rankings and the ranking by degree, which we call singularity, we notice that these singularities are generally located on either side of the diagonal. So, while some AS are singular according to both metrics, the majority are singular according to only the betweenness or the fitness. This shows that fitness is original to both degree and betweenness.

We point out that all the AS rankings presented differ greatly from a CAIDA reference ranking based on ASes customer cone [29]. This is mainly due to the fact that ASes with many customers, such as backbone ASes with no provider known as "tier 1 networks", are less interested in public peering. For them, settlement-free peering means losing a potential customer. Thus, we see the limits and specificity of the PeeringDB dataset. Now that we have shown that nestedness metrics reveal the triangular structure of the PeeringDB networks and give an original view of the public peering ecosystem, we will investigate if the presence of a triangular structure enables efficient link prediction.

<sup>2</sup>We do not show results from BINMATNEST due to performance issues.

Table 5. Comparison of node-level nestedness rankings to the traditional degree ranking. The ranking difference with respect to degree is shown in square brackets.

(a) AS-Country			(b) AS-IXP		
Degree	Binmatnest	Fitness	Degree	Fitness	
1	Packet Clearing House	Packet Clearing House [0]	1	Hurricane Electric	Hurricane Electric [0]
2	Packet Clearing House AS42	Packet Clearing House AS42 [0]	2	Cloudflare	Packet Clearing House [1]
3	Cloudflare	Facebook Inc AS63293 [14]	3	Packet Clearing House	Packet Clearing House AS42 [1]
4	Akamai Technologies	Cloudflare [-1]	4	Packet Clearing House AS42	Cloudflare [-2]
5	Hurricane Electric	Akamai Technologies [-1]	5	Google LLC	Akamai Technologies [2]
6	VeriSign Glob. Reg. Serv.	VeriSign Glob. Reg. Serv. [0]	6	Microsoft	Google LLC [-1]
7	Microsoft	Hurricane Electric [-2]	7	Akamai Technologies	Microsoft [-1]
8	Facebook Inc	Google LLC [1]	8	Facebook Inc	Facebook Inc [0]
9	Google LLC	Netnod [2]	9	Amazon.com	Amazon.com [0]
10	Amazon.com	Facebook Inc [-2]	10	Netflix	VeriSign Global Registry Services [5]
11	Netnod	RIPE NCC K-Root Ops [2]	11	Yahoo!	EdgeCast Networks (VDMS) [1]
12	Subspace	Microsoft [-5]	12	EdgeCast Networks (VDMS)	Netflix [-2]
13	RIPE NCC K-Root Ops	Amazon.com [-3]	13	Limelight Networks Global	Limelight Networks Global [0]
14	Imperva	Subspace [-2]	14	Fastly, Inc.	Fastly, Inc. [0]
15	Netflix	Netskope [12]	15	VeriSign Global Registry Services	Yahoo! [-4]
16	Anexia	Imperva [-2]	16	G-Core Labs	<b>Netnod</b> [11]
17	Facebook Inc AS63293	G-Core Labs [2]	17	Apple Inc.	Equinix Internet Exchange - MLPE [25]
18	Riot Games	Quantil Networks [10]	18	Riot Games	i3D.net [1]
19	G-Core Labs	Telstra (International) [34]	19	i3D.net	<b>RIPE NCC K-Root Operations</b> [30]
20	Yahoo!	Swisscom [5]	20	Blizzard Entertainment	Salesforce.com [14]
21	Valve Corporation	Anexia [-5]	21	Subspace	Netskope [2]
22	SG.GS	Riot Games [-4]	22	SG.GS	Subspace [-1]
23	Zenlayer Inc	Netflix [-8]	23	Netskope	<b>DNS-OARC-112</b> [29]
24	Limelight Networks Global	Wnet Telecom USA [41]	24	Imperva	Riot Games [-6]
25	Swisscom	Zenlayer Inc [-2]	25	Twitch	Apple Inc. [-8]
26	M247	Valve Corporation [-5]	26	Valve Corporation	G-Core Labs [-10]
27	Netskope	Yahoo! [-7]	27	Netnod	Twitch [-2]
28	Quantil Networks	M247 [-2]	28	Swisscom	Imperva [-4]
29	EdgeCast Networks (VDMS)	<b>Rostelecom</b> [80]	29	SoftLayer Technologies, Inc.	Quantil Networks [8]
30	Equinix IX - MLPE	<b>NetIX Com. Ltd.</b> [16]	30	StackPath (Highwinds)	Blizzard Entertainment [-10]

### 5.3 Link prediction

As a more concrete application, we ask whether the nestedness observed in PeeringDB graphs can be used to predict the appearance or disappearance of new links, which is motivated by the fact that nestedness captures the attractiveness of IXPs and countries for ASes looking for novel interconnection opportunities. To assess the quality of the proposed link prediction scheme, the temporal evolution of the network has to be extracted as a ground truth. Thus, we first explain how temporal snapshots of the two graphs are extracted from PeeringDB. Second, we present the prediction method, and describe how the predictions are validated for both graphs. Results are given for both graphs numerically, together with an illustration for the AS-country graph.

*Networks preparation.* The temporal study is made possible by CAIDA [27], which provides the community with daily PeeringDB snapshots from 2013 to present time. Our study window is the period between 2019-01-01 and 2021-03-01, and we choose a time resolution of a month, leading to 27 snapshots of the two graphs.

Since the graphs evolve over time, we study here the largest connected component present over the time window of interest. Thus, we select the nodes of both graphs by considering only the sub-graphs induced by the nodes present in all snapshots. This selection leads to only a few nodes being disconnected from the main connected component, that are removed as well. For the AS-country graph, this results in a network of 181 ASes and 91 countries. For the AS-IXP graph of which we previously studied only the main nested component, we perform a new largest nested community search as described in subsection 4.3, this time at the beginning of the time period. To be comprehensive, our goal is to start from a community and study its evolution, without the knowledge of the future community structure. Once again,

we find a main community strongly nested that is similar to the one presented in Table 3: it includes 51.8% of ASes, 39.7% of IXPs, 80% of all port capacity, 55 countries and all 15 hypergiants. We therefore consider only the sub-graph induced by this community.

For the AS-country and AS-IXP graphs, respectively, the number of links varies from 1863 to 2064 and 6329 to 6893. The links persist over time: for both graphs, a mean of 98% of links of a snapshot are still present in a snapshot taken 6 months later.

*Link prediction models.* We adopt the method of link prediction proposed in [13]. This method is used to predict both the appearance and disappearance of links in nested networks. We only present how to predict appearances for the sake of simplicity, the extension to the prediction of disappearances being straightforward and given in [13]. It consists in i) ordering all unformed links at the beginning of the time period according to their likelihood of appearance ii) predicting the links sequentially according to this order.

The first point is achieved by taking advantage of the triangular structure of the bi-adjacency matrix. Let us consider a probit model of parameters  $\alpha, \beta, \gamma$  that is fit from the nested arrangement of links at the beginning of the time period:

$$A_{i,j} = \alpha d(i) + \beta d(j) + \gamma(d(i) \times d(j)) + \epsilon_{i,j} \quad (10)$$

The probit residual term  $\epsilon$  represents the deviation from the nested arrangement of links and the model. A low negative value corresponds to an unformed link that was expected to be present in the nested structure. It is therefore a link that is likely to be created. On the other hand, high positive value corresponds to a formed link not expected to be present, and thus being likely to disappear in later snapshots.

*Validation.* We assess the quality of the predictor by doing the following checks. Based on the first snapshot and its matrix, we define the order of link appearance (resp. disappearance) following the decreasing (resp. increasing) order of the probit model of (10). For each prediction, we check in the subsequent snapshots if the link has appeared (resp. disappeared) or not.

The quality of the predictor is then assessed by the receiver operating characteristic (ROC) curve [11]. The true positive rate (TPR, predicted links actually created or deleted afterwards) and the false positive rate (FPR, predicted links not created or not deleted afterwards) is evaluated at each prediction. The evolution of the TPR as a function of the FPR gives the ROC curve, and the area under this curve (AUC) contained in the unit square quantifies the quality of the predictor: 1 for the best predictor, 0.5 for a random predictor, and 0 for the worst predictor.

*Results.* The results of the link prediction for the AS-country graphs are shown in Fig. 5. We see in a) that links with negative probit residuals and therefore likely to appear are located inside the triangular structure. These links are proposed first by the predictor, leading to a high predictor quality with an AUC of the ROC curve shown in b) of 0.875, far outperforming a random prediction. The positive probit residuals shown in b) are located outside of the triangle. This time, the prediction based on the nested structure is not significantly better than a random one, with an AUC of the ROC curve shown in d) of 0.621. The same predictor behavior, good for created links and average for deleted links, is found for the AS-IXP graph. The two ROC curves for creation and deletion are shown in Fig. 6. The first has an AUC of 0.88 and the second 0.55. Overall, these results show that the nestedness observed in the PeeringDB ecosystem can be efficiently leveraged to predict the creation of new links. This could serve as an IXP recommender system for network operators.



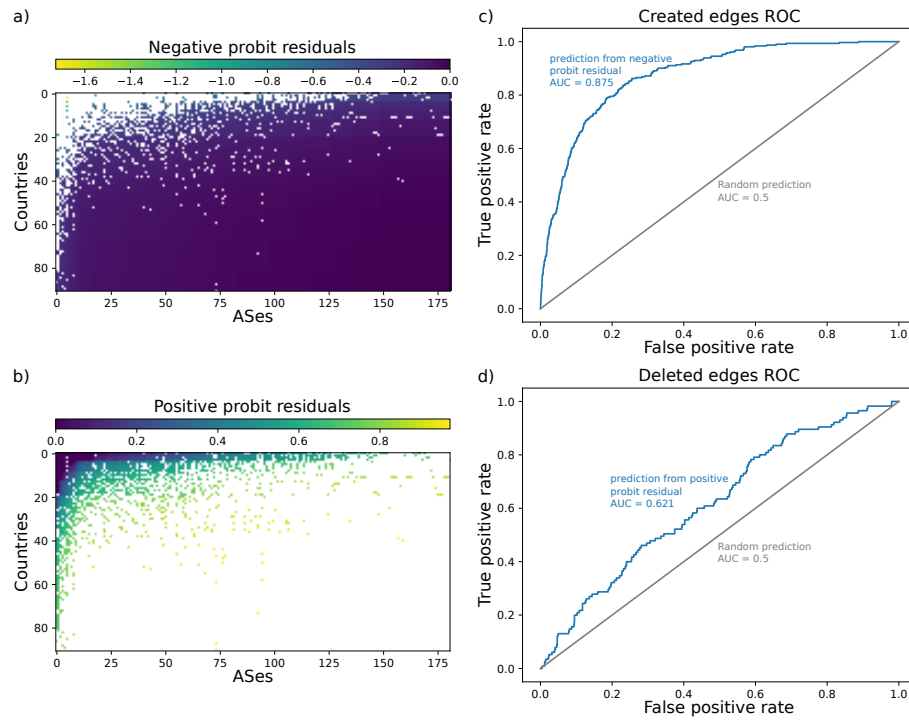


Fig. 5. Link prediction for the AS-country graph. a) Negative probit residuals, i.e. the likelihood for links to appear (links already present and therefore excluded of the prediction are marked in white). b) Positive probit residuals, i.e. the likelihood for links to disappear (links already missing and therefore excluded of the prediction are marked in white). c) ROC of the prediction of newly created links. d) ROC of the prediction of newly deleted links.

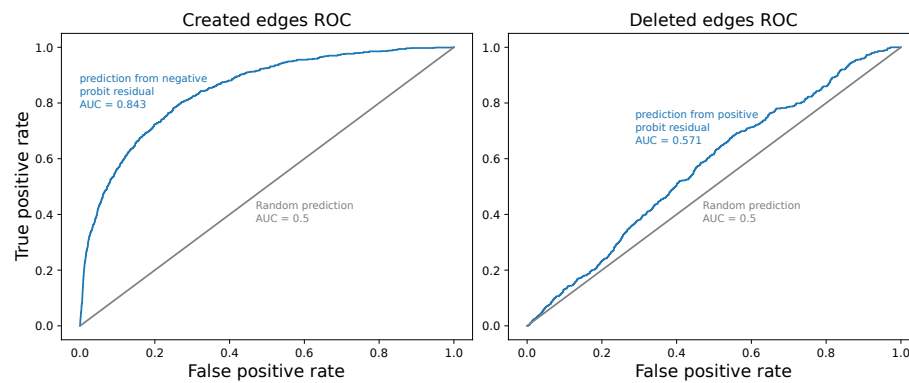


Fig. 6. Link prediction for the AS-IXP graph. Left panel: ROC of the prediction of newly created links. Right panel: ROC of the prediction of newly deleted links.  
Manuscript submitted to ACM

## 6 CONCLUSION

In this work, we have drawn a parallel with biological ecosystems to reveal the nestedness of graph representations of the public peering ecosystem. To the best of our knowledge, this is the first time that nestedness has been observed in Internet topology data. We have shown that nestedness has concrete applications to better understand the public peering ecosystem, an ecosystem that is difficult to grasp because of the lack of topology data due to its confidential nature and measurement limitations. The node-level nestedness metric known as fitness-complexity provides a new interpretable ranking of ASes. This metric highlights the ability of ASes to reach isolated regions of the world and IXPs, which is not captured with the traditional degree and betweenness metrics. With a 27-month time study, we also show that nestedness effectively predicts new associations between ASes and countries in the world, as well as new membership of ASes to IXPs.

We believe these applications will be useful to stakeholders in the public peering ecosystem. For example, link prediction can be seen as an IXP recommender system for network operators. Future work will explore the use of other analysis tools based on deep learning such as graph neural networks to better understand the structure of the Internet.

## ACKNOWLEDGMENTS

We would like to thank Dima Shepelyansky, Marc Bruyère and anonymous reviewers for their helpful comments during the development of this work.

## REFERENCES

- [1] Bernhard Ager, Nikolaos Chatzis, Anja Feldmann, Nadi Sarrar, Steve Uhlig, and Walter Willinger. 2012. Anatomy of a Large European IXP. In *Proceedings of the ACM SIGCOMM 2012 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication* (Helsinki, Finland) (*SIGCOMM '12*). Association for Computing Machinery, New York, NY, USA, 163–174. <https://doi.org/10.1145/2342356.2342393>
- [2] Mário Almeida-Neto, Paulo Guimaraes, Paulo R Guimaraes Jr, Rafael D Loyola, and Werner Ulrich. 2008. A consistent metric for nestedness analysis in ecological systems: reconciling concept and measurement. *Oikos* 117, 8 (2008), 1227–1239.
- [3] Michael J Barber. 2007. Modularity and community detection in bipartite networks. *Physical Review E* 76, 6 (2007), 066102.
- [4] Jordi Bascompte, Pedro Jordano, Carlos J Melián, and Jens M Olesen. 2003. The nested assembly of plant–animal mutualistic networks. *Proceedings of the National Academy of Sciences* 100, 16 (2003), 9383–9387.
- [5] Stephen J Beckett, Chris A Boulton, and Hywel TP Williams. 2014. FALCON: a software package for analysis of nestedness in bipartite networks. *F1000Research* 3 (2014).
- [6] Francis K Bell, Dragoš Cvetković, Peter Rowlinson, and Slobodan K Simić. 2008. Graphs for which the least eigenvalue is minimal, I. *Linear Algebra Appl.* 429, 1 (2008), 234–241.
- [7] Stefan Boettcher and Allon G Percus. 2002. Optimization with extremal dynamics. *complexity* 8, 2 (2002), 57–62.
- [8] Javier Borge-Holthoefer, Raquel A Baños, Carlos Gracia-Lázaro, and Yamir Moreno. 2017. Emergence of consensus as a modular-to-nested transition in communication dynamics. *Scientific reports* 7, 1 (2017), 1–9.
- [9] Timm Böttger, Gianni Antichi, Eder L Fernandes, Roberto di Lallo, Marc Bruyere, Steve Uhlig, and Ignacio Castro. 2018. The elusive internet flattening: 10 years of IXP growth. *arXiv e-prints* (2018).
- [10] Timm Böttger, Felix Cuadrado, and Steve Uhlig. 2018. Looking for Hypergiants in PeeringDB. *SIGCOMM Comput. Commun. Rev.* 48, 3 (Sept. 2018), 13–19. <https://doi.org/10.1145/3276799.3276801>
- [11] Andrew P Bradley. 1997. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern recognition* 30, 7 (1997), 1145–1159.
- [12] Matteo Bruno, Fabio Saracco, Diego Garlaschelli, Claudio J Tessone, and Guido Caldarelli. 2020. The ambiguity of nestedness under soft and hard constraints. *Scientific reports* 10, 1 (2020), 1–13.
- [13] Sebastián Bustos, Charles Gomez, Ricardo Hausmann, and César A Hidalgo. 2012. The dynamics of nestedness predicts the evolution of industrial ecosystems. *PloS one* 7, 11 (2012), e49393.
- [14] Nikolaos Chatzis, Georgios Smaragdakis, Jan Böttger, Thomas Krenc, and Anja Feldmann. 2013. On the benefits of using a large IXP as an Internet vantage point. In *Proceedings of the 2013 conference on Internet measurement conference*. 333–346.
- [15] COSIN3-UOC and María Palazzi. 2021. *COSIN3-UOC/nestedness\_modularity\_in\_block\_nestedness\_analysis: nestedness\_modularity\_in\_block\_nestedness\_analysis*. <https://doi.org/10.5281/zenodo.4557009>
- [16] Virginia Dominguez-Garcia and Miguel A Munoz. 2015. Ranking species in mutualistic networks. *Scientific reports* 5, 1 (2015), 1–7.
- [17] Jordi Duch and Alex Arenas. 2005. Community detection in complex networks using extremal optimization. *Physical review E* 72, 2 (2005), 027104.
- [18] Santo Fortunato. 2010. Community detection in graphs. *Physics reports* 486, 3-5 (2010), 75–174.
- [19] Linton C Freeman. 1977. A set of measures of centrality based on betweenness. *Sociometry* (1977), 35–41.
- [20] Petros Gigis, Matt Calder, Lefteris Manassakis, George Nomikos, Vasileios Kotronis, Xenofontas Dimitropoulos, Ethan Katz-Bassett, and Georgios Smaragdakis. 2021. Seven years in the life of Hypergiants’ off-nets. In *Proceedings of the 2021 ACM SIGCOMM 2021 Conference*. 516–533.
- [21] Vasileios Giotsas, Matthew Luckie, Bradley Huffaker, and KC Claffy. 2014. Inferring complex AS relationships. In *Proceedings of the 2014 Conference on Internet Measurement Conference*. 23–30.
- [22] Michelle Girvan and Mark EJ Newman. 2002. Community structure in social and biological networks. *Proceedings of the national academy of sciences* 99, 12 (2002), 7821–7826.
- [23] F Maxwell Harper and Joseph A Konstan. 2015. The movielens datasets: History and context. *Acm transactions on interactive intelligent systems (tiis)* 5, 4 (2015), 1–19.
- [24] Petter Holme, Beom Jun Kim, Chang No Yoon, and Seung Kee Han. 2002. Attack vulnerability of complex networks. *Physical review E* 65, 5 (2002), 056109.
- [25] Bascompte Lab. 2023. *Web of life: ecological networks database*. <https://www.web-of-life.es>
- [26] Jian-Hong Lin, Claudio Juan Tessone, and Manuel Sebastian Mariani. 2018. Nestedness maximization in complex networks through the fitness-complexity algorithm. *Entropy* 20, 10 (2018), 768.
- [27] Aemen Lodhi, Natalie Larson, Amogh Dhamdhere, Constantine Dovrolis, and Kc Claffy. 2014. Using peeringDB to understand the peering ecosystem. *ACM SIGCOMM Computer Communication Review* 44, 2 (2014), 20–27.
- [28] Justin Loye, Sandrine Mouisset, Marc Bruyère, and Katia Jaffrès-Runser. 2022. Global Internet public peering capacity of interconnection: a complex network analysis. (2022). <https://doi.org/10.48550/ARXIV.2206.05146>
- [29] Matthew Luckie, Bradley Huffaker, Amogh Dhamdhere, Vasileios Giotsas, and KC Claffy. 2013. AS relationships, customer cones, and validation. In *Proceedings of the 2013 conference on Internet measurement conference*. 243–256.
- [30] Manuel Sebastian Mariani, Zhuo-Ming Ren, Jordi Bascompte, and Claudio Juan Tessone. 2019. Nestedness in complex networks: observation, emergence, and implications. *Physics Reports* 813 (2019), 1–90.

- [31] Meta. 2023. *Peering policies*. <https://www.facebook.com/peering>
- [32] Reza Motamedi, Reza Rejaie, and Walter Willinger. 2014. A survey of techniques for internet topology discovery. *IEEE Communications Surveys & Tutorials* 17, 2 (2014), 1044–1065.
- [33] Netflix. 2023. *Netflix Open Connect*. <https://openconnect.netflix.com/en/>
- [34] NetIX. 2023. *Company website*. <https://www.netix.net/>
- [35] George Nomikos, Pavlos Sermpezis, and Xenofontas Dimitropoulos. 2017. Re-mapping the internet: Bring the IXPs into play: www. inspire. edu. gr/ixp-map. In *2017 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*. IEEE, 910–915.
- [36] Ricardo V Oliveira, Dan Pei, Walter Willinger, Beichuan Zhang, and Lixia Zhang. 2008. In search of the elusive ground truth: the Internet’s AS-level connectivity structure. *ACM SIGMETRICS Performance Evaluation Review* 36, 1 (2008), 217–228.
- [37] Maria J Palazzi, Javier Borge-Holthoefer, Claudio J Tessone, and Albert Solé-Ribalta. 2019. Macro-and mesoscale pattern interdependencies in complex networks. *Journal of the Royal Society Interface* 16, 159 (2019), 20190553.
- [38] Maria J Palazzi, Jordi Cabot, Javier Luis Canovas Izquierdo, Albert Solé-Ribalta, and Javier Borge-Holthoefer. 2019. Online division of labour: emergent structures in open source software. *Scientific reports* 9, 1 (2019), 13890.
- [39] Maria J Palazzi, Aniello Lampo, Albert Solé-Ribalta, and Javier Borge-Holthoefer. 2021. A comprehensive benchmark for structural assessment in bipartite networks. *bioRxiv* (2021).
- [40] Maria J Palazzi, Albert Solé-Ribalta, Violeta Calleja-Solanas, Sandro Meloni, Carlos A Plata, Samir Suweis, and Javier Borge-Holthoefer. 2021. An ecological approach to structural flexibility in online communication systems. *Nature communications* 12, 1 (2021), 1–11.
- [41] Bruce D Patterson and Wirt Atmar. 1986. Nested subsets and the structure of insular mammalian faunas and archipelagos. *Biological journal of the Linnean society* 28, 1-2 (1986), 65–82.
- [42] PeeringDB. 2023. *Homepage*. <https://www.peeringdb.com/>
- [43] Miguel A Rodriguez-Gironés and Luis Santamaria. 2006. A new algorithm to calculate the nestedness temperature of presence–absence matrices. *Journal of biogeography* 33, 5 (2006), 924–935.
- [44] Matthew Roughan, Walter Willinger, Olaf Maennel, Debbie Perouli, and Randy Bush. 2011. 10 lessons from 10 years of measuring and modeling the internet’s autonomous systems. *IEEE Journal on Selected Areas in Communications* 29, 9 (2011), 1810–1821.
- [45] Brandon Schlinker, Italo Cunha, Yi-Ching Chiu, Srikanth Sundaresan, and Ethan Katz-Bassett. 2019. Internet performance from facebook’s edge. In *Proceedings of the Internet Measurement Conference*. 179–194.
- [46] Albert Solé-Ribalta, Claudio J Tessone, Manuel S Mariani, and Javier Borge-Holthoefer. 2018. Revealing in-block nestedness: detection and benchmarking. *Physical review E* 97, 6 (2018), 062302.
- [47] Phillip Staniczenko, Jason C Kopp, and Stefano Allesina. 2013. The ghost of nestedness in ecological networks. *Nature communications* 4, 1 (2013), 1–6.
- [48] Viktor Stojkoski, Zoran Utkovski, and Ljupco Kocarev. 2016. The impact of services on economic complexity: Service sophistication as route for economic growth. *PloS one* 11, 8 (2016), e0161633.
- [49] Andrea Tacchella, Matthieu Cristelli, Guido Caldarelli, Andrea Gabrielli, and Luciano Pietronero. 2012. A new metrics for countries’ fitness and products’ complexity. *Scientific reports* 2, 1 (2012), 1–7.
- [50] Werner Ulrich, Mário Almeida-Neto, and Nicholas J Gotelli. 2009. A consumer’s guide to nestedness analysis. *Oikos* 118, 1 (2009), 3–17.