



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-04733587v2

HAL Id: hal-04733587

<https://hal.science/hal-04733587v2>

Submitted on 27 Jan 2025

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.

Emergence of Nestedness in the Public Internet Peering Ecosystem

Justin Loye¹, Sandrine Mouysset², and Katia Jaffrès-Runser²

¹ Internet Initiative Japan, Tokyo, Japan
jloye@iiij.ad.jp

² IRIT, Université de Toulouse, CNRS, Toulouse INP, UT3, Toulouse, France
sandrine.mouysset@irit.fr, katia.jaffres-runser@irit.fr

Abstract. Nestedness is a property of bipartite complex networks that has been shown to characterize the peculiar structure of biological and economical networks. 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. 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. 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.

Keywords: Peering ecosystem, Nestedness, Link prediction

1 Introduction

Nestedness is a graph property that characterizes the peculiar structure of certain real bipartite complex networks. This property has been observed mainly in two types of biological networks: *i*) the so-called "mutualistic" networks [2], 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 [32]. Only more recently has this structural property been identified in non-biological networks: in the international trade network [11] and in online communication systems [6].

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. 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) [12]. 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”. This ecosystem is not very well known because individual peering relationships between ASes are notoriously difficult to measure [24] [35] and the information is not disclosed by interested parties for competitive and security reasons.

To study the nestedness of the public Internet peering ecosystem, we leverage the PeeringDB dataset to construct two graphs that capture either the relationships leading to a mutualistic network or the ones leading to a biogeographic network. Since ASes and IXPs gain mutual benefit from their association (ASes gain access to novel ASes and IXPs gain in attractiveness), the first graph is the AS [is member of] IXP graph. To capture the relationship leading to a biogeographic network in the second graph, we model the process of selective immigration of an AS to a new geographic area. ASes are actors that may have a geographical strategy in choosing the IXP to connect. For instance, they could select IXPs leading to densely populated areas. 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.

2 Background

We adopt the following notation. A graph $G(E, V)$ is composed of a set of edges or links E , and a set of vertices or nodes V and is described by an adjacency matrix \mathbf{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 \mathbf{B} . We denote by $d(i)$ the degree of node i . A graph is connected if there exists a succession of links connecting every pair of nodes.

2.1 Dataset

The dataset generally used by the scientific community to study the evolution of public peering is PeeringDB [33]. 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 [8] [20], and can be modeled as a bipartite graph representing the AS [is a member of] IXP relationship [27]. PeeringDB is being collected on a daily basis by CAIDA [19], allowing for temporal studies.

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 [8][19]. The data is now mostly uploaded automatically and it is in ASes best interests to give correct information. 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 [7] and CDN deployment [25], PeeringDB does not report on private network interconnects (PNIs).

2.2 Nestedness

Nestedness [22] 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. Figure 1-(b)). As a consequence, it is 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. Figure 1-(c)). Such a structure is representative of a perfectly nested network.

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

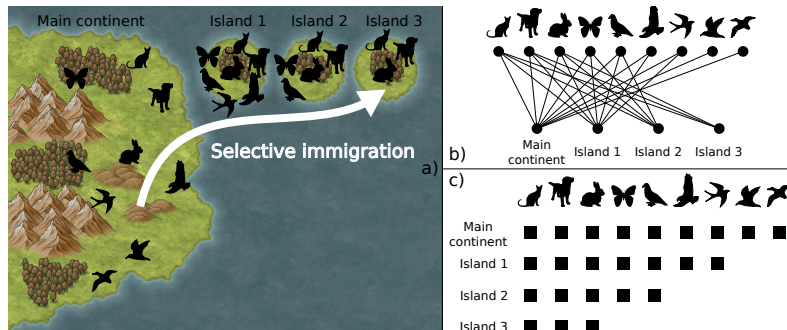


Fig. 1: Illustration of a perfectly nested biogeographic network.

might look for nestedness at the partition-level to identify communities where nodes arrange themselves in a nested manner.

3 Modeling PeeringDB for a nestedness study

After introducing the networks requirements for a nestedness study, we present how the AS [is member of] IXP and AS [is present at] country graphs are built and conditioned to meet these requirements.

3.1 Constraints on the bi-adjacency matrix

The nested biological networks commonly studied in the literature are present in the database *web of life* [17]. Most of them are bipartite, with bi-adjacency matrix presenting the following properties [10]:

- 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). Therefore 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 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 [29][31].

3.2 AS-IXP graph

The AS [is member of] IXP graph, referred to as *AS-IXP graph*, is a bipartite graph composed of 11407 ASes reporting 31889 associations to 780 IXPs. It is obtained with PeeringDB membership data as proposed in [27], for which we weight the links with the port size attribute that represents its capacity in Gbps. We consider only the main connected component, which comprises more than 99% of the total of nodes.

We filter out nodes in order to meet the constraints given earlier. We studied 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 [8] as the sum of the port sizes of an AS’ links). We chose to filter out the nodes with a degree $d \leq 3$ in order to *i*) reduce the eccentricity to 0.5 ; *ii*) make the network small enough for a tractable nestedness study on 2681 nodes ; *iii*) and have the least restrictive filter possible. 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: degree, PageRank (to select major traffic destination) and reverse PageRank (to select major traffic sources). For each metric, we select the top 100 ASes and keep the union of the three sets. The final AS-country graph is composed of 182 ASes and 112 countries.

PageRank and reverse PageRank metrics are computed on a weighted and directed AS-IXP graph model of PeeringDB as given in [20]. Weights stem from the routers' port size reported in PeeringDB and edge directions from the reported traffic balance of ASes (Heavy or Mostly Inbound or Outbound, Balanced).

Next, we solely work with the unweighted version of the two AS-country and AS-IXP graphs because: *i*) nestedness studies are mainly carried out on unweighted graphs, and are therefore better understood ; *ii*) many nestedness tools don't account for weights ; *iii*) to simplify the scope of our analysis (for example, in the link prediction study, predicting the link weight is harder than predicting the appearance 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

Here, 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 [40],[3] and [22].

4.1 Detecting nestedness

Statistical significance of 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) [22]. State-of-the-art implementations of this model are known as "Bascompte's PP" [2] and "corrected PP" [30], the latter being a correction of the widely used Bascompte's PP aiming to further reduce the number of type I errors.

The nestedness statistical significance of the network to be evaluated is assessed with the p -value and the z -score [40]. 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 [3]. A p -value less than 0.05 is considered sufficient to say that the network is more nested than the statistical set [3]. 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 To evaluate the nestedness of a network, it is considered good practice to use several metrics. We consider the popular “Nestedness metric based on Overlap and Decreasing Fill” (noted NODF) [1], a variant of NODF taking into account a null model called “ $\tilde{\eta}$ ” [36], and the “spectral radius metric” (noted ρ)[4][37].

4.2 Nestedness in AS-country graph

The results of the statistical tests of nestedness for the AS-country graph are given in Table 1. 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.

4.3 Nestedness in AS-IXP graph

For the AS-IXP graph, the results presented in Table 1 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.

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

Graph	Nestedness metric			
	Null model	$\tilde{\eta}$	NODF	Spectral radius
AS-country		0.001 (80.59)	0.001 (61.04)	0.001 (29.97)
AS-IXP	PP (Bascompte)	0.001 (103.94)	0.001 (86.61)	0.001 (94.58)
AS-IXP nested comp.		0.001 (144.36)	0.001 (113.21)	0.001 (56.31)
AS-country		0.001 (27.34)	0.001 (20.93)	0.001 (11.42)
AS-IXP	PP (corrected)	0.945 (-1.63)	0.977 (-1.94)	0.001 (39.68)
AS-IXP nested comp.		0.001 (27.97)	0.001 (19.53)	0.001 (16.79)

However, it has been shown in [20] that the IXPs membership network exhibits a community structure that correlates with the geographic proximity of

ASes and IXPs. Thus, we leverage a metric developed in [36] to quantify the presence of nested communities. This metric called In-Block Nestedness (IBN) is an adaptation of the metric $\tilde{\eta}$ also introduced in [36]. It can be optimized in the same way as the classical modularity [16][14] to obtain a network partition. In order to find nested communities, the authors of [36] optimize IBN with the extremal optimization (EO) method [5] [13]. In the present work, we use the implementation of this algorithm made available by the authors of [28] to find nested communities, and compare it to the classical modularity communities. We observe that the modularity tends to partition the network in communities of similar sizes. The 15 hypergiant ASes identified in [8], *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 [8] 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 1 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.

5 Use cases

The triangular structure in the bi-adjacency matrix which is characteristic of nestedness is yet to be revealed with node-level metrics. 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.

5.1 Node-level metrics

The simplest node-level metric we can use 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 [34]. 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.

We also use the fitness-complexity metric [39] [38] that also produces state-of-the-art ordering [18], but which has the advantage over BINMATNEST of having simple and efficient implementation and a clear interpretation. This metric has been introduced in world trade networks to rank countries and the products they export. It is modeled by coupled, non-linear and recursive fitness and complexity equations. For a network described by the adjacency matrix \mathbf{A} , the **fitness** $\mathbf{f}(i)$ of a country i is defined as the sum of the complexity scores $\mathbf{q}(\alpha)$ of the products α exported by this country:

$$\mathbf{f}(i) = \sum_{\alpha} \mathbf{A}_{i,\alpha} \mathbf{q}(\alpha), \quad \mathbf{q}(\alpha) = \frac{1}{\sum_i \mathbf{A}_{i,\alpha} (1 - \mathbf{f}(i))}. \quad (1)$$

These equations are solved by computing iteratively:

$$\mathbf{f}_n(i) = \sum_{\alpha} \mathbf{A}_{i,\alpha} \mathbf{q}_{n-1}(\alpha), \quad \mathbf{q}_n(\alpha) = \frac{1}{\sum_i \mathbf{A}_{i,\alpha} (1 - \mathbf{f}_{n-1}(i))}, \quad (2)$$

beginning with uniform non-null vectors \mathbf{f}_0 and \mathbf{q}_0 . 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. In our case, the interpretation for the AS-IXP graph (respectively AS-country graph) is as follows: an AS has a high fitness score if it is able to reach many IXPs (countries), some of which are difficult to access. An IXP (country) has a high complexity score if it is associated with many ASes, some of which are capable of reaching hard-to-reach areas.

5.2 New key players

We show the AS-country bi-adjacency matrix ordered by different node-level nestedness metrics in Figure 2. All metrics reveal a triangular structure charac-

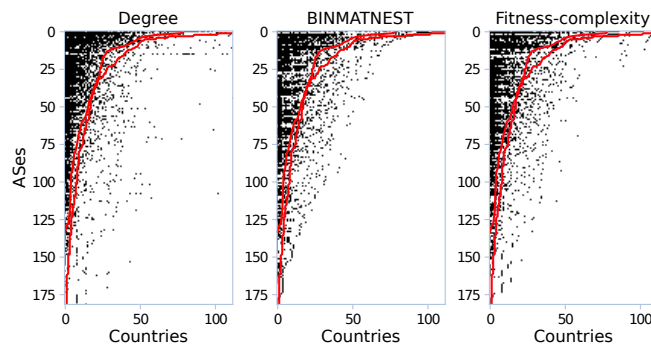


Fig. 2: Ordering of the bi-adjacency matrix of the AS-country graph for different node-level nestedness metrics.

teristic of nestedness. The links are concentrated in a stylized upper left triangle, with a convex hypotenuse caused by the filling rate of the matrix. 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 [11]. 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. We investigated the differences between 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 [23] efficiently through offnet deployment as close as possible to end-users [15], 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” [26], 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 observed the appearance of new actors in the AS-IXP graph with by comparing the degree and fitness-complexity rankings³. 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.

We point out that all the AS rankings presented differ greatly from a CAIDA reference ranking based on ASes customer cone [21]. This is mainly due to the

³We do not show results from BINMATNEST due to performance issues.

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.

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 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. We find a similar main community strongly nested comprising 51.8% of ASes, 39.7% of IXPs, 80% of all port capacity 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 [11]. 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 [11]. 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 α, β, γ that is fit from the nested arrangement of links at the beginning of the time period:

$$\mathbf{A}_{i,j} = \alpha \mathbf{d}(i) + \beta \mathbf{d}(j) + \gamma(\mathbf{d}(i) \times \mathbf{d}(j)) + \epsilon_{i,j} \quad (3)$$

The probit residual term ϵ 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 (3). 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 [9]. 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) 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 Figure 3. 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 only slightly better than a random one, with an AUC of the ROC curve shown in d) of 0.621.

We find the same behavior for the prediction quality of the AS-IXP graph, with an AUC of 0.88 for created links and 0.55 for deleted links. 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.

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

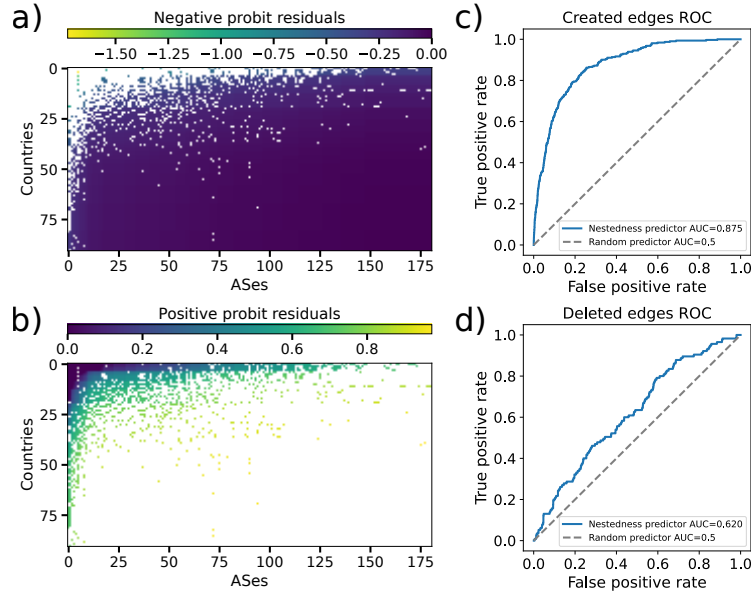


Fig. 3: Link prediction for the AS-country graph.

to better understand the public peering ecosystem, an ecosystem 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.

References

1. Almeida-Neto, M., Guimaraes, P., Guimaraes Jr, P.R., Loyola, R.D., Ulrich, W.: A consistent metric for nestedness analysis in ecological systems: reconciling concept and measurement. *Oikos* **117**(8), 1227–1239 (2008)
2. Bascompte, J., Jordano, P., Melián, C.J., Olesen, J.M.: The nested assembly of plant–animal mutualistic networks. *Proceedings of the National Academy of Sciences* **100**(16), 9383–9387 (2003)
3. Beckett, S.J., Boulton, C.A., Williams, H.T.: Falcon: a software package for analysis of nestedness in bipartite networks. *F1000Research* **3** (2014)

4. Bell, F.K., Cvetković, D., Rowlinson, P., Simić, S.K.: Graphs for which the least eigenvalue is minimal, i. *Linear Algebra and its Applications* **429**(1), 234–241 (2008)
5. Boettcher, S., Percus, A.G.: Optimization with extremal dynamics. *complexity* **8**(2), 57–62 (2002)
6. Borge-Holthoefer, J., Baños, R.A., Gracia-Lázaro, C., Moreno, Y.: Emergence of consensus as a modular-to-nested transition in communication dynamics. *Scientific reports* **7**(1), 1–9 (2017)
7. Böttger, T., Antichi, G., Fernandes, E.L., di Lallo, R., Bruyere, M., Uhlig, S., Castro, I.: The elusive internet flattening: 10 years of ixp growth. *arXiv e-prints* (2018)
8. Böttger, T., Cuadrado, F., Uhlig, S.: Looking for hypergiants in peeringdb. *SIGCOMM Comput. Commun. Rev.* **48**(3), 13–19 (2018). DOI 10.1145/3276799.3276801. URL <https://doi.org/10.1145/3276799.3276801>
9. Bradley, A.P.: The use of the area under the roc curve in the evaluation of machine learning algorithms. *Pattern recognition* **30**(7), 1145–1159 (1997)
10. Bruno, M., Saracco, F., Garlaschelli, D., Tessone, C.J., Caldarelli, G.: The ambiguity of nestedness under soft and hard constraints. *Scientific reports* **10**(1), 1–13 (2020)
11. Bustos, S., Gomez, C., Hausmann, R., Hidalgo, C.A.: The dynamics of nestedness predicts the evolution of industrial ecosystems. *PloS one* **7**(11), e49,393 (2012)
12. Chatzis, N., Smaragdakis, G., Böttger, J., Krenc, T., Feldmann, A.: On the benefits of using a large ixp as an internet vantage point. In: *Proceedings of the 2013 conference on Internet measurement conference*, pp. 333–346 (2013)
13. Duch, J., Arenas, A.: Community detection in complex networks using extremal optimization. *Physical review E* **72**(2), 027,104 (2005)
14. Fortunato, S.: Community detection in graphs. *Physics reports* **486**(3-5), 75–174 (2010)
15. Gígis, P., Calder, M., Manassakis, L., Nomikos, G., Kotronis, V., Dimitropoulos, X., Katz-Bassett, E., Smaragdakis, G.: Seven years in the life of hypergiants’ off-nets. In: *Proceedings of the 2021 ACM SIGCOMM 2021 Conference*, pp. 516–533 (2021)
16. Girvan, M., Newman, M.E.: Community structure in social and biological networks. *Proceedings of the national academy of sciences* **99**(12), 7821–7826 (2002)
17. Lab, B.: *Web of life: ecological networks database* (2023). URL <https://www.web-of-life.es>
18. Lin, J.H., Tessone, C.J., Mariani, M.S.: Nestedness maximization in complex networks through the fitness-complexity algorithm. *Entropy* **20**(10), 768 (2018)
19. Lodhi, A., Larson, N., Dhamdhere, A., Dovrolis, C., Claffy, K.: Using peeringdb to understand the peering ecosystem. *ACM SIGCOMM Computer Communication Review* **44**(2), 20–27 (2014)
20. Loye, J., Mouysset, S., Bruyère, M., Jaffrès-Runser, K.: Global internet public peering capacity of interconnection: a complex network analysis (2022). URL <https://arxiv.org/abs/2206.05146>
21. Luckie, M., Huffaker, B., Dhamdhere, A., Giotsas, V., Claffy, K.: As relationships, customer cones, and validation. In: *Proceedings of the 2013 conference on Internet measurement conference*, pp. 243–256 (2013)
22. Mariani, M.S., Ren, Z.M., Bascompte, J., Tessone, C.J.: Nestedness in complex networks: observation, emergence, and implications. *Physics Reports* **813**, 1–90 (2019)

23. Meta: Peering policies (2023). URL <https://www.facebook.com/peering>
24. Motamedi, R., Rejaie, R., Willinger, W.: A survey of techniques for internet topology discovery. *IEEE Communications Surveys & Tutorials* **17**(2), 1044–1065 (2014)
25. Netflix: Netflix open connect (2023). URL <https://openconnect.netflix.com/en/>
26. NetIX: Company website (2023). URL <https://www.netix.net/>
27. Nomikos, G., Sermpezis, P., Dimitropoulos, X.: 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), pp. 910–915. IEEE (2017)
28. Palazzi, M.J., Borge-Holthoefer, J., Tessone, C.J., Solé-Ribalta, A.: Macro-and mesoscale pattern interdependencies in complex networks. *Journal of the Royal Society Interface* **16**(159), 20190,553 (2019)
29. Palazzi, M.J., Cabot, J., Canovas Izquierdo, J.L., Solé-Ribalta, A., Borge-Holthoefer, J.: Online division of labour: emergent structures in open source software. *Scientific reports* **9**(1), 13,890 (2019)
30. Palazzi, M.J., Lampo, A., Solé-Ribalta, A., Borge-Holthoefer, J.: A comprehensive benchmark for structural assessment in bipartite networks. *bioRxiv* (2021)
31. Palazzi, M.J., Solé-Ribalta, A., Calleja-Solanas, V., Meloni, S., Plata, C.A., Suweis, S., Borge-Holthoefer, J.: An ecological approach to structural flexibility in online communication systems. *Nature communications* **12**(1), 1–11 (2021)
32. Patterson, B.D., Atmar, W.: Nested subsets and the structure of insular mammalian faunas and archipelagos. *Biological journal of the Linnean society* **28**(1-2), 65–82 (1986)
33. PeeringDB: Homepage (2023). URL <https://www.peeringdb.com/>
34. Rodriguez-Gironés, M.A., Santamaria, L.: A new algorithm to calculate the nestedness temperature of presence–absence matrices. *Journal of biogeography* **33**(5), 924–935 (2006)
35. Roughan, M., Willinger, W., Maennel, O., Perouli, D., Bush, R.: 10 lessons from 10 years of measuring and modeling the internet’s autonomous systems. *IEEE Journal on Selected Areas in Communications* **29**(9), 1810–1821 (2011)
36. Solé-Ribalta, A., Tessone, C.J., Mariani, M.S., Borge-Holthoefer, J.: Revealing in-block nestedness: detection and benchmarking. *Physical review E* **97**(6), 062,302 (2018)
37. Staniczenko, P., Kopp, J.C., Allesina, S.: The ghost of nestedness in ecological networks. *Nature communications* **4**(1), 1–6 (2013)
38. Stojkoski, V., Utkovski, Z., Kocarev, L.: The impact of services on economic complexity: Service sophistication as route for economic growth. *PloS one* **11**(8), e0161,633 (2016)
39. Tacchella, A., Cristelli, M., Caldarelli, G., Gabrielli, A., Pietronero, L.: A new metrics for countries’ fitness and products’ complexity. *Scientific reports* **2**(1), 1–7 (2012)
40. Ulrich, W., Almeida-Neto, M., Gotelli, N.J.: A consumer’s guide to nestedness analysis. *Oikos* **118**(1), 3–17 (2009)