

HELA: Inferring AS Relationships With a Hybrid of Empirical and Learning Algorithms

Xingang Shi[✉], *Member, IEEE*, Zitong Jin[✉], Bin Xiong, Xinyao Huang, Xiaotian Xi[✉],
Dan Li, *Senior Member, IEEE*, Han Zhang[✉], Zhiliang Wang[✉], *Member, IEEE*, and Xia Yin, *Senior Member, IEEE*

Abstract—Knowledge of the business relationships between Autonomous Systems (ASes) is the basis for studying many aspects of the Internet. Despite the significant progress achieved by the latest inference algorithms, their inference results still suffer from errors on many special or critical links, thus hindering many relationship-related applications. We take an in-depth analysis on the challenges inherent in inferring AS relationships, including complex routing policies, limited and biased vantage point (VP) coverage, as well as a lack of accurate validation data. To address these challenges, we introduce HELA, a framework for inferring AS relationships with a hybrid of empirical and machine learning algorithms. HELA incorporates an array of grouping, voting, and machine learning algorithms and allows flexible substitution of each. We systematically evaluate various combinations of them to determine the most effective one for HELA. Furthermore, we describe the collection of varied validation datasets, including BGP community and RPSL records from Internet Routing Registries (IRRs), as well as OneStep community. Our up-to-date dataset corrects errors in previous published validation sets, contains 95% more labelled links, and exhibits a closer alignment to actual link distribution. Using routing data and validation datasets composed for each month during 2021 ~ 2023, we access HELA's superiority in both inference accuracy and stability compared to the state-of-the-art inference algorithms, i.e., AS-Rank, ProBLINK, and HELA's predecessor TopoScope. In particular, HELA achieves up to $2.9\times$ reduction on error rates across overall datasets, up to $2.6\times$ reduction with a $4.5\times$ decrease on standard deviation on incomplete and biased datasets, and up to $1.7\times$ reduction on various sources of validation datasets.

Index Terms—Internet, autonomous system, business relationship, inference algorithm, machine learning.

I. INTRODUCTION

NOWADAYS, the Internet is composed of more than 70,000 autonomous systems (ASes). Based on their business contracts, these ASes propagate prefixes and exchange

routing information with each other to control how traffic will be routed. In a simplified manner, the business contracts are often represented by the AS-to-AS relationships, including peer-to-peer (p2p), customer-to-provider (c2p) and provider-to-customer (p2c). As both the scale and the complexity of the Internet grow at an unprecedented rate, a good knowledge of the relationships between ASes is of great importance for understanding, operating and consolidating the Internet, such as studying the robustness of the Internet [1], [2], detecting route hijacks, route leaks and route bottlenecks [3], [4], [5], [6], designing various routing based attacks and their counter measurements [7], [8], [9], [10], as well as deploying secure routing mechanisms [11], [12], [13], [14].

However, achieving a sound knowledge of AS relationships is not easy, since they are usually kept confidential and have to be inferred from various relevant information. To our knowledge, nearly all inference algorithms [15], [16], [17], [18], [19], [20], [21], [22], [23], [24] use BGP routes collected at a set of vantage points (VPs) [25], [26], [27] as their primary data, despite other related information like IXP peering [28], [29], [30], BGP community [31], [32] and even active probing or route manipulation can be used as supplement. In this paper, we take an in-depth look at the problem of inferring AS relationships using BGP routes collected by measurement platforms at a few locations, because this is the (currently) only viable solution for Internet-scale inference.

We show that inference algorithms are facing several fundamental challenges inherent in this scheme. First, BGP routes are the interaction results of complex routing policies beyond simple AS relationships. For example, although valley-free routing [33] based on p2c and p2p relationships is widely followed in the Internet, exceptions are also pervasive [34], and there exist more complex relationships such as partial transit and hybrid relationships [35]. In addition, routing paths are often affected by non-negligible noises, e.g., temporary oscillations, configuration mistakes, etc. Such complexity will naturally obscure our inference. Second, our actual observation on routing paths is severely disturbed by the ill-distributed vantage points (VPs). The routes reported by a VP is a partial view of the Internet. Moreover, publicly available VPs are typically concentrated around a few geographical locations, mostly in the upper tiers in the Internet hierarchy. An aggregation of their partial views may further aggravate the observation bias. Last, there lacks of an adequate set of real-world relationships that can be used as ground-truth for a fair evaluation of the inference results. The validation data used in the latest works [23], [36] gradually becomes outdated, and thus its accuracy degrades over time. In addition, a recent study [37] shows

Received 16 August 2024; revised 19 January 2025 and 18 March 2025; accepted 14 May 2025; approved by IEEE TRANSACTIONS ON NETWORKING Editor K. Chen. Date of publication 29 May 2025; date of current version 17 October 2025. This work was supported in part by the National Natural Science Foundation of China under Grant 62221003 and in part by the Zhongguancun Laboratory. (*Corresponding author: Zitong Jin.*)

Xingang Shi, Han Zhang, and Zhiliang Wang are with the Institute for Network Sciences and Cyberspace, Tsinghua University, Beijing 100190, China, and also with the Zhongguancun Laboratory, Beijing 100086, China (e-mail: shixg@cernet.edu.cn).

Zitong Jin and Xiaotian Xi are with the Zhongguancun Laboratory, Beijing 100086, China (e-mail: jinztt@mail.zgclab.edu.cn).

Bin Xiong and Xinyao Huang are with the Institute for Network Sciences and Cyberspace, Tsinghua University, Beijing 100190, China.

Dan Li and Xia Yin are with the Department of Computer Science and Technology, Tsinghua University, Beijing 100190, China, and also with the Zhongguancun Laboratory, Beijing 100086, China.

Digital Object Identifier 10.1109/TON.2025.3572148

2998-4157 © 2025 IEEE. All rights reserved, including rights for text and data mining, and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission.

See <https://www.ieee.org/publications/rights/index.html> for more information.

Authorized licensed use limited to: Tsinghua University. Downloaded on October 15, 2026 at 11:48:02 UTC from IEEE Xplore. Restrictions apply.

its distribution can also be biased. We note that, many of these aspects will become more exacerbated when the Internet further grows.

To systematically address these challenges, we develop HELA, a framework that uses a **Hybrid of Empirical and Learning Algorithms** to infer AS relationships in the Internet from fragmentary observations on routing paths. On the one hand, HELA fully utilizes the knowledge of many experts who have rich experiences in operating or analyzing inter-domain routing. The empirical rules summarized by them reflect the roles of (and the relationships between) many ASes determined by the long evolved Internet economics, and can be quite accurate for those ASes. The credibility of these rules is further enhanced through the grouping of BGP routes by prefix propagation and observability, followed by a voting mechanism. On the other hand, HELA uses learning methods to automatically discover features that cannot be easily identified with human experience. Interestingly, with a careful organization of these methods, they can help each other to mitigate their own ineffectiveness due to the aforementioned challenges. Within this framework, we further study dozens of combinations of empirical, grouping and learning algorithms to choose the one HELA finally uses.

Given that previous validation datasets are outdated and publicly available methods for collecting such sets are persistent obsolete, HELA incorporates current AS import/export policies and AS remark identification rules, and collects validation datasets from multiple sources, including Internet Routing Registries (IRRs), BGP communities, and OneStep. These sets, which are updated monthly, cover diverse categories of AS links and are, on average, $\sim 95\%$ larger than the most recent published set [23]. Moreover, our new dataset is more up-to-date, rectifying 236 erroneous links in the previous validation dataset, and exhibits closer to actual link distribution, with a JS-Divergence reduced $3.2\times$ compared to the prior dataset.

Using a series of routing data and validation datasets composed for each month during 2021 $\sim$ 2023, we evaluate HELA against three state-of-the-art inference algorithms, including AS-Rank [22], ProbLink [23], as well as HELA's predecessor TopoScope [36]. We find HELA consistently improves both inference accuracy and stability on various aspects:

1. For overall accuracy, HELA achieves an error rate of up to $1.3 \sim 2.9\times$ smaller, where the rate is up to $1.4 \sim 5.0\times$ smaller for c2p links, and $1.5 \sim 1.8\times$ smaller for p2p links. For hard links that are often misclassified by common algorithms and critical links essential for Internet traffic, HELA can reduce the inference error rate by up to $2.9\times$ and $2.7\times$, respectively.

2. When we intentionally obfuscate the routing data by making it more incomplete or biased, HELA demonstrates even greater superiority. For such datasets, HELA's average error rate is up to $2.6\times$ smaller, and its standard deviation of error rate is up to $4.5\times$ smaller.

3. Across validation datasets sourced from IRR communities, IRR RPSL, and OneStep communities, which exhibit variations in link distribution, HELA consistently outperforms other algorithms, achieving up to a $1.7\times$ reduction in error rates. Notably, the new dataset assembled by HELA is 95%

larger compared to the prior validation dataset [23], and it results in a rate error reduction of $\sim 2.8\times$ for all algorithms.

The paper is organized as follows. Section II introduces existing research efforts on AS relationship inference. Section III discusses the corresponding challenges, as well as the basic ideas of HELA. Section IV details the framework design of HELA, our selection of the exact algorithms, as well as our method to build an up-to-date validation dataset, based on which Section V thoroughly evaluates HELA against three state-of-the-art algorithms. Section VI concludes the paper.¹

II. BACKGROUND AND RELATED WORK

ASes exchange reachability information and control how Internet traffic will be routed using the Border Gateway Protocol (BGP). To select its preferred path toward a destination, each AS utilizes a decision procedure where the most important factor can be roughly represented by its relationships to its neighboring ASes. Two classes of AS relationships, including customer-provider (c2p, or p2c in the reverse direction) and peer-peer (p2p), are the most common [15]. In a p2c relationship, a customer AS pays its provider AS for transiting traffic, while in a p2p relationship, two ASes bilaterally transit traffic to themselves or their customers for free. Such relationships are not mandated, but are formed during the evolvement of Internet economics. A good understanding of the Internet level AS relationships is important and challenging.

Typically, AS relationships are kept confidential, and a common approach to infer them is based on extracting as many BGP AS paths as possible from BGP routing tables or update messages. A few measurement platforms, including Route Views [25], RIPE RIS [26] and Isolario [27], have made great efforts to publish such information. They consist of multiple route collectors (RCs), to which many routers, often called vantage points (VPs), volunteer to report their BGP routes. The imperfection of the BGP routes collected by VPs has well been recognized. For example, 40% of the AS links are believed to be missed, and there is also a high redundancy between different platforms [38], [39], [40]. The former is mainly due to the choice that a VP will not report all its routes, and there are only a limited number of VPs.

Inferring AS relationships with publicly available BGP routing data is a long-studied problem. Although information such as [28], [29], and [30] or BGP community [31], [32] can also help, the number of AS links they affect is typically orders of magnitudes smaller, and thus are only used for assistance. Earlier algorithms devise classification rules based on empirical knowledge about the property of ASes and routing paths. The pioneering work by Gao [15] relies on the valley-free principle in the Gao-Rexford (GR) model [33]. The routing policy means that, due to its natural economic interest, an AS will not transit traffic between its providers or peers, so a p2p or c2p link cannot occur after a p2c or p2p link. Their algorithm infers provider ASes who lie in the middle of AS paths and have large degrees, and peer ASes who have similar degrees. A bunch of algorithms [16], [17], [18], [19], [20], [21] then try to maximize the number of valley-free paths

¹The source code of HELA, our validation datasets and inference results, as well as other supplement materials, will be published on our website.

or other particular links, e.g., p2c links where the provider has a larger degree than the customer. However, these assumptions have oversimplified the problem.

AS-Rank [22] builds on the fact that a few transit-free ASes at the top tier of the Internet form a clique, and proposes empirical rules on more advanced features such as transit degrees and customer cones. It is the first to use real AS relationships as ground truth to validate its accuracy, and has been regarded as a standard tool in this area [41], including inferring the relationships between IPv6-enabled ASes [42]. However, many steps in AS-Rank still introduce non-negligible errors [36]. Based on the fact that the majority of AS links are relatively easy to infer, CoreToLeaf [23] proposes to assume valley-freeness of the paths that pass through those transit-free ASes, and find p2c links with three simple steps. It can only infer part of the AS relationships, but it is much simpler than AS-Rank and can achieve good accuracy in our evaluation.

Features and rules can also be automatically learned from observed routing data. Varghese and Ruan [43] is an early attempt on this direction. To mitigate the uncertainties or even conflicts caused by the imperfection of empirical rules, UNARI [24] proposes a probability model (based on the valley-free assumption) that explicitly considers the inference uncertainty. However, it focuses more on reasoning the value of such a model, but less on detailed algorithm design and performance evaluation. Actually, its accuracy is much worse than AS-Rank under our evaluation. ProbLink [23] independently uses a Bayesian probability model to incorporate multiple link features. It discovers that a set of *hard links* are very difficult to classify, and demonstrates its superiority on these hard links.

TopoScope [36] combines AS-Rank's empirical rules and Bayesian Network in a way where each one helps the other. The most important idea is using the former to infer a portion of links with high certainty, but leaving the rest portion to the latter probability model. TopoScope enables more accurate and stable inference on AS relationships, and demonstrates its help on real applications like detecting routing leaks, inferring AS paths and analyzing routing bottlenecks. This paper further extends TopoScope into a framework where different empirical or learning algorithms are substituted and compared, and introduces an updated method for collecting validation datasets to thoroughly evaluate existing inference algorithms.

III. CHALLENGES AND BASIC IDEAS

In this section, we elaborate on the inherent challenges an AS relationship inference algorithm faces, which lead to the solution framework we will explore. The statistics presented in this section are computed based on BGP routes collected in Jan 2023, while we believe the challenges may become more exacerbated when the Internet further evolves. The datasets are retrieved from Route Views and RIPE RIS, and more details about them will be deferred to Section V-A.

Complexity of the Internet routing. It's well known that, p2c and p2p are the two most common relationships between neighboring ASes, and valley-free routing [33] based on such relationships is widely followed in the Internet. However, such a model oversimplifies the Internet routing. First, there also exist more complex relationships, such as hybrid [35], [44],

TABLE I
LIMITED AND BIASED OBSERVATION BY EACH VP

type	#VP	all links	p2c links (average)	p2p links (average)
full VPs	381	482,692	96,509 (66.1%)	3,017 (0.8%)
partial VPs	2,034	176,638	799 (0.5%)	28 (0.007%)
all VPs	2,415	516,723	15,892 (10.9%)	499 (0.1%)

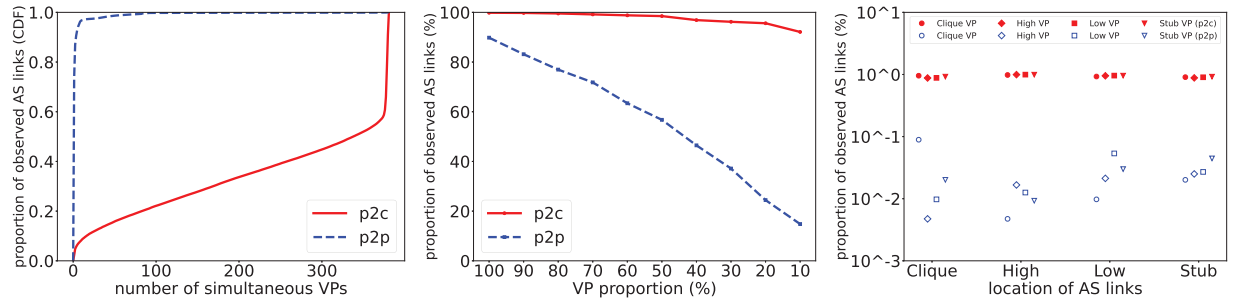
where two ASes inter-connect at multiple locations with different relationships, and partial transit [35], where a customer relationship is restricted to a limited scope. Second, although valley-free routing represents the natural economic interest of an AS, exceptions are also pervasive [34]. In the routing tables we analyze, there are $\sim 9\%$ paths violating the valley-free principle.² Third, current Internet routing is based on prefixes rather than ASes. The routing policies for different prefixes of the same AS may differ at any location, which introduces another level of complexity of the routing paths. For example, in the routing data of Jan 2023, we find multiple optimal paths exist in between approximately 36% source-destination AS pairs. Last, routing paths are often affected by non-negligible noises (e.g., temporary oscillations, configuration mistakes, etc.). Some noisy routes can be removed with certain heuristics, while the remaining ones, together with normal routes, will be treated indiscriminately in an inference algorithm. Consequently, these complexities inevitably obscure our inference on the basic AS relationships.

Although an in-depth analysis of the sources of complexity is beyond the scope of this paper, HELA's integration of empirical rules and learning algorithms ensures accurate and robust inference. The challenge then lies in how to organize them and let them effectively help each other.

Limited and biased observations from each VP. Existing inference algorithms typically use routing data published by Route Views [25] and RIPE RIS [26], collected from around 2,400 VPs. These VPs are divided into two categories, i.e., full VPs, each of which reports a large number of routes to RCs, and partial VPs, each of which reports only a small number of routes because they often deliberately limit their exportation of only a small address space (e.g., only report their routes for some /8 subnets). The routes reported by each VP represent its observations on the Internet, from which an inference algorithm will extract AS links and infer AS relationships. Table I lists the average number and percentage of p2c or p2p links that can be observed by each VP, indicating that a single VP, even a full VP, only achieves a limited and biased view on the global Internet. Specifically, a full VP roughly observes only around 1.12% (3,132) p2p links and 71.87% (109,214) p2c links, w.r.t. the total numbers of p2p and p2c links reported by all the VPs.³ The huge imbalance between p2p and p2c observations is mainly caused by the underpinning AS relationships and the no-valley principle, where routes learned from a peer or a provider neighbor will

²This rate is computed based on the relationships inferred with AS-Rank [22], but using other inference algorithms does not greatly affect the result.

³Since partial VPs contribute much less information, in what follows, we only analyze routes collected by full VPs.



(a) links observed simultaneously by VPs (b) link coverage decreases with less VPs (c) links observed by VPs at different tiers

Fig. 1. Limited and biased observation by a set of VPs (relationships are inferred by AS-Rank on our dataset in Jan 2023).

TABLE II
UNBALANCED DISTRIBUTION OF VPS

location	#ASes	#VPS in RV+RIS
clique: ASes published by CAIDA	19	17 (89.5%)
high-tier: customers of clique ASes	1348	904 (67.1%)
low-tier: other ASes	9602	1398 (14.6%)
sub-tier: ASes without customers	58681	96 (0.2%)

only be advertised to customers. Further analysis also shows that, p2p links observed by VPs are typically within its one (73%) or two hops (20%). Existing inference algorithms often rarely consider the limited and biased VP observations.

Aggregated observation bias due to a limited number of VPs. The observation bias of each VP not only exists, but also differs a lot from each other. This will lead to an aggregated observation bias closely related to the set of VPs an inference algorithm takes as input. To illustrate this, Figure 1a depicts the cumulative distribution (CDF) of how many p2p or p2c links can be observed simultaneously by more than a certain number of VPs. It is interesting to see that, normally a p2c link will be observed by a large number of VPs, while a p2p link will typically be observed by only a very limited number of VPs, which are most probably the VPs close to this link. For example, around 73.8% p2p links can be seen by only one VP, while only 4.3% p2p links can be seen by at least 10 VPs. On the contrary, more than 83.1% p2c links can be observed by at least 50 VPs. This means removing a small number of VPs will not affect the observations of p2c links greatly, but will have a noticeable influence on that of p2p links.

The lack of observations on AS links will trick empirical inference algorithms that depend much on the total degree, the transit degree, or the neighbors of an AS. Figure 1b shows when we gradually remove the observations of more VPs, the number of observed links will also decrease, with the number of p2p observations decreasing much faster.⁴

To overcome the sensitivity of empirical rules, HELA uses a simple but powerful ensemble learning technique, i.e., grouping and voting, to find relationships that can be empirically derived but are not sensitive to the removal of observations. However, care must be taken when we divide groups.

Aggregated observation bias due to unbalanced VP locations. Bias originates not only from the small cardinality of the VP set, but also from how the VPs distribute themselves

TABLE III
DISTANCES FROM OBSERVED P2P LINKS TO A VP

distance	#p2p links	distance	#p2p links
0	950	1	1,050
2	623	3	349
4	292	5	245
6	159	7	119

in the network. In Table II, we list the number of ASes and VPs located in different tiers of the Internet, where we roughly divide the Internet hierarchy into four tiers, i.e., clique, high-tier, low-tier and stub-tier. It is obvious that, existing VPs are more concentrated in the upper tiers. For example, in our dataset, 89.5% clique ASes and 67.1% high-tier AS act as VPs, while less than 1% low-tier or stub ASes act as VPs.

Figure 1c further shows how this imbalance of VP locations affects observations, where we divide both links and VPs into four different tiers, and plot the proportion (in log scale) of links in each tier that can be observed by VPs in each tier. To be specific, each link is double-counted, once for each of its two endpoints. For p2c links, the observation capabilities of VPs in different tiers are very similar. However, for a p2p link, it can be observed always better by VPs located in the same tier as itself, than by VPs in other tiers. The distribution of distances between a VP and its observed p2p links, as shown in Table III, further indicates that VPs are more likely to observe p2p links closer to themselves. Besides, it is more difficult for VPs in upper tiers to observe p2p links in lower tiers, than it is for VPs in lower tiers to observe p2p links in higher tiers. These phenomena are indeed due to the limited capability of observing faraway or low layer p2p links mentioned above. Thus, the concentration of VPs in higher tiers will cause imbalanced observations.

In a probabilistic inference algorithm that relies on a statistical model of link features, the weight of each feature is often related to its occurrence frequency, which is in turn affected by the number of links related to that feature. For example, in ProbLink [23], the non-path feature of a link is closely related to whether another neighboring link also appears in many different paths. The observed occurrences must have been affected by the bias mentioned above, and introduce around 20% errors according to our evaluation (Section V-B).

To solve this problem, HELA tests various learning models, including supervised, unsupervised and semi-supervised ones, and chooses a Bayesian Network model that takes into account

⁴The p2p curve does not start from 100% because some p2p links can be observed only by partial VPs and are not included here.

the inter-dependencies among link features, especially the observation bias in different tiers. In this way, the uneven distribution of observed features originating from the biased VP distributions can be effectively mitigated.

It is also worth noting that, different datasets exhibit different biases. For example, the VPs provided by Isolario [27] are relatively more concentrated in lower layers. In our evaluation (Section V-D), we will intentionally use datasets that are more incomplete or biased to demonstrate the superiority of HELA.

Coverage and bias of validation datasets. A good validation dataset is also critical to fairly verify the performance of any inference algorithm. A common approach to find a reasonable amount of ground-truth on AS relationships involves using BGP community [45] and Routing Policy Specification Language (RPSL) [46]. BGP communities allow ASes to attach values in BGP update messages, to denote certain property (e.g., its peer relationships or geographic locations) or expected action (e.g., traffic blackholing, selective advertisement, etc.). However, the exact semantics of community values may either be published or kept private, limiting the ability to construct a comprehensive community dictionary. For example, Vasileios and Shi [34] built a dictionary of 1286 community values from 224 different ASes in 2012. RPSL enables ASes to record their import/export rules in IRR, which can help us identify p2c relationships (e.g., an AS exporting all its neighbors to another AS). Using community dictionary built by Vasileios and Shi [34], RPSL from RIPE NCC, as well as privately collected data, As-Rank [22] built a validation dataset consisting of $\sim 41\text{K}$ links whose relationships can be annotated, and validated $\sim 36\text{K}$ (29%) of their inferred 126K links. ProbLink [23] used the same community dictionary to build their own validation dataset of $\sim 45\text{K}$ link relationships, covering around 21% of their inferred 213K links. However, the community dictionary and RPSL information have become out-dated over time and thus introduced errors in the validation dataset. A recent study [37] also shows the distribution of the validation dataset (which is determined by the ASes who publish those communities) is biased towards some geographical and topological regions.

To fairly evaluate HELA, we use communities and RPSL freshly updated in IRRs and OneStep [47] to produce a series of up-to-date relationship validation datasets. Despite variations in the composition and accuracy of different validation datasets, our inference results always perform superior accuracy across any of them. Collectively, these validation datasets provide a more accurate, uniform, and reliable large-scale validation resource.

IV. DESIGN OF THE HELA FRAMEWORK

A. Overview

Based on the above ideas, we design a general framework named HELA (i.e., Hybrid of Empirical and Learning Algorithms) for accurate, robust and extensible inference on AS relationships. The input data HELA needs are BGP paths that can be locally observed by VPs, and its output is the p2p or p2c/c2p relationships for observed links. Inferring relationships that are more complex [35] or for unobserved links [36], are out of the scope of this paper. Unless explicitly

stated otherwise, the statistics presented in this section are computed based on the same dataset as in Section III.

As Figure 2 depicts, HELA works in four steps as follows.

1. Divide the whole set of BGP routes into several groups. In this step, various grouping strategies, e.g., grouping by the VPs who observe those routes, or by the ASes who originate those routes, can be substituted.

2. Apply certain empirical algorithms to each group of routes, so that if a link appears in multiple groups, we get multiple inference results for it. We then use voting to find a set of *consensus links*, for which a majority of groups can reach an agreement on their inferred results. From the *consensus links*, with our empirical rule on the observability of p2p or p2c links, we further identify a few *trusted links* and label them with high confidence. Links other than the *trusted links* are called *fuzzy links*, and we compute a probability for a *fuzzy link* based on the voting results. In this step, various empirical algorithms, e.g., AS-Rank [22], Core2Leaf [23]), can be substituted into the framework to perform a first-stage inference.

3. Feed the data with first-stage inference results into certain machine learning models, and infer the relationship on each *fuzzy link*. Many learning methods, such as supervised, semi-supervised and unsupervised learning, can be substituted into the framework in this step. The model features are extracted from BGP routes, but may differ in each learning method.

4. Identify community dictionary and RPSL information from multiple sources, to extract a set of labeled links from all observed BGP routes. Vary the combination of algorithms that can be substituted in steps 1 $\sim$ 3, and use this set of labeled links to find the best combination as HELA's final choice. We note that, both the validation sources and BGP routes used in this step are specific for Jan 2023, while we will use a serial of datasets from 2021 to 2023 to evaluate HELA against existing inference algorithms in Section V.

Noting that similar to algorithms like AS-Rank and ProbLink, HELA infers AS relationships solely from observed AS path links, and all evaluations in this study are based on these observable links. For the inference of unobservable (i.e., hidden) links in the Internet, refer to TopoScope, as HELA does not incorporate additional mechanisms for this purpose. Furthermore, to ensure fairness, all evaluations involving TopoScope exclude its hidden link inference component.

B. Empirical Algorithms for Relationship Inference

Before explaining the steps in HELA, we first introduce the empirical algorithms we have investigated to study their contributions to HELA, in particular, the trade-off between inference accuracy and coverage. We test three different algorithms, each employing a set of manually designed rules. The first is AS-Rank [22], the de-facto algorithm for AS relationship inference. The second is Core2Leaf, an algorithm proposed by Jin et al. [23] based on their argument that many links are actually easy to infer. It relies on valley-freeness of the paths passing through tier-1 ASes, and achieves a similar level of accuracy on those easy-to-infer links. We further propose Core2Leaf-PC, a refined version of the Core2Leaf algorithm, which employs stricter rules to infer a (smaller) portion of links with higher confidence. In this variant, p2p relationships

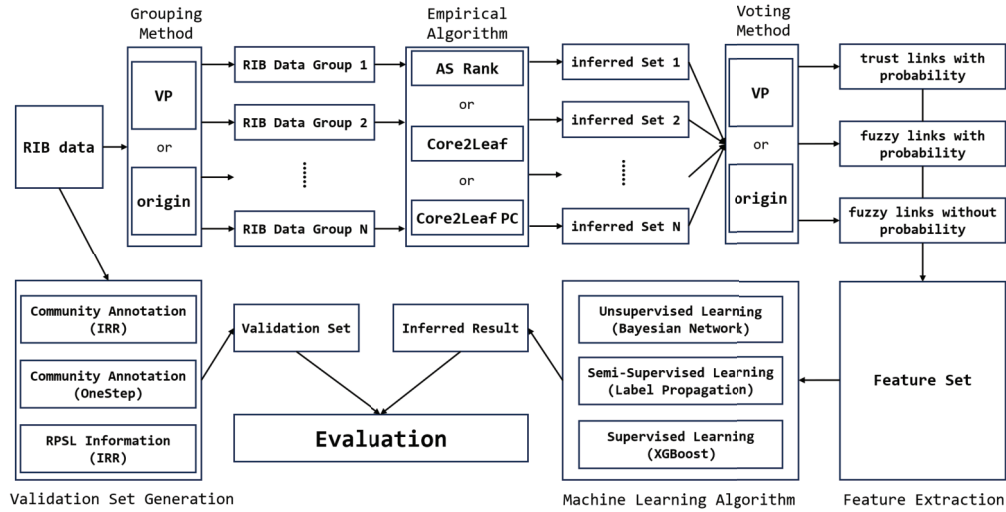


Fig. 2. Structure and Workflow of the HELA framework.

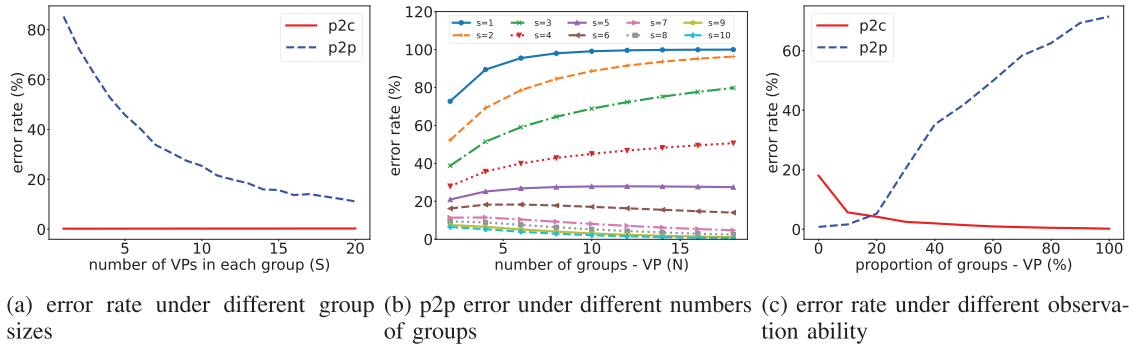


Fig. 3. Inference error rate of AS-Rank with VP based grouping.

will only be inferred when they are between tier-1 ASes (i.e., with high confidence), but it will leave many uninferred links.

Using routing data from 2021 to 2023 as a validation dataset, we evaluated the inference accuracy and coverage of the three empirical algorithms when applied independently of HELA. Core2Leaf-PC demonstrated the lowest error rate, averaging 0.9%, while AS-Rank showed a slightly higher error rate of 1.0%. Core2Leaf exhibited the highest error rate, averaging 2.4%. However, the average percentages of links they can infer are 28.6% (Core2Leaf-PC), 100% (AS-Rank) and 99.7% (Core2Leaf) respectively. Evidently, Core2Leaf-PC prioritizes accuracy at the expense of reduced coverage.

C. Grouping and Voting

We design and evaluate several grouping and voting methods. To deal with the observation biases of VPs, we divide routing paths according to the VPs who observe them. To deal with the complexity in prefix-based routing, we divide routing paths according to the corresponding Origin ASes. We then use voting to make a first-stage inference on different links with different levels of confidence.

1) *Dividing According to VPs*: As mentioned in Section III, inference algorithms are sensitive to the aggregated observation capability. There are two implications. First, removing observations on some links may cause inference errors on other links. Second, the sensitivity related to each link differs.

To reduce such sensitivity, we deliberately induce sensitivity by limiting an inference algorithm to observations from only a small group of VPs, then from consensus achieved by many different groups, to find links which are less sensitive to the algorithm. Essentially, we are using ensemble learning, i.e., voting, to reduce variances caused by algorithm sensitivity. In order to illustrate the characteristics of grouping and voting among VPs in this section, we use AS-Rank, if not otherwise stated, as the empirical algorithm in the HELA framework.

In Figure 3a, we plot the average inference error rate when applying AS-Rank to the observations (i.e., routes in Jan 2023) of a group of S VPs. The error rate is calculated as follows: for p2c relationship, a true positive (tp) happens when a p2c link is inferred correctly, and a false positive (fp) happens when a p2p link is mis-classified as p2c, or the inferred provider-customer direction is wrong. Then the error rate of p2c is $\frac{\#fp}{\#tp + \#fp}$. The same goes for p2p relationship. We can see the error rate of p2c (i.e., the probability that a p2p or a c2p link is mis-classified as p2c) is relatively low and stable. However, the high error rate of p2p drops to around 20% only after the group size S grows over 10, and becomes stable afterward. This clearly demonstrates how the removal of VPs causes aggregated observation bias and inaccurate inference, especially on p2c links that are less observed (i.e., they are mis-classified as p2p and cause false positives for p2p).

Suppose we use N different groups of VPs, and each group independently gives an inference result with an error

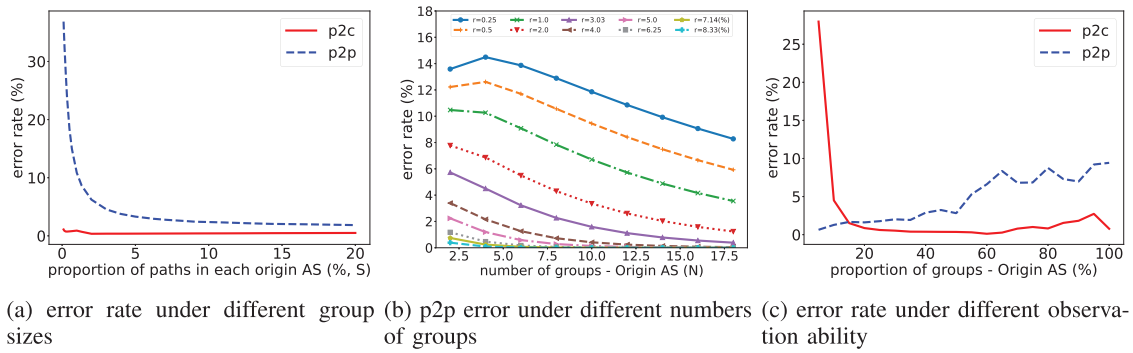


Fig. 4. Inference error rate of AS-Rank with Origin AS based grouping.

probability of p . The probability that at least $\lceil N/2 \rceil$ out of N groups, i.e., the majority of them, give a correct result is

$$P(X \geq \lceil N/2 \rceil) = \sum_{x=\lceil N/2 \rceil}^N \binom{N}{x} p^x (1-p)^{N-x}. \quad (1)$$

According to the probability of a successful majority voting in equation (1), we plot the error rate for p2p links with different group sizes (S) and different numbers of groups (N) in Figure 3b, where p can be found from Figure 3a. It can be seen that, when the group size S is larger than 10 and the group count N is larger than 6, the voting will nearly always give a right answer. On the contrary, if the group size is small such that a single group has a large error rate, it will be harder to get a correct answer. In practice, we first set $S = 10$, and determine N according to the total number of full VPs. If VPs are abundant, we will increase S appropriately when there are enough groups (e.g., $N > 10$). For example, in our evaluations presented in Section V-B and Section V-C, there are around 380 full VPs, and we use $S = 38$ and $N = 10$ in the grouping stage. In addition, we include partial VPs in one group.

2) *Dividing According to Origin ASes*: When we divide routes according to VPs, each group fed into the empirical algorithms represents a partial view on the global Internet topology. To take into account the complexity of prefix-based routing policies, we consider another dimension of division, i.e., let one group capture a partial view on the global routing behavior. More specifically, for one group, we select $r\%$ of the routes originated by each AS, so that each group contains $r\%$ of the paths towards every AS. Note that this is different from using paths from $r\%$ Origin ASes, since the latter will be very much like using paths observed by some VPs. In Figure 4a, we plot the average inference error rate with respect to r (i.e., the group size), where AS-Rank is applied on each group. We can see the curves for p2p and p2c exhibit a similar trend to those in Figure 3a, but the error rate of p2p links drops much faster when r increases. Using the same way as Figure 3b, we plot the error rate for p2p links with different group sizes (r) and numbers of groups (N) in Figure 4b according to the probability of a successful majority voting in equation (1). Consequently, when a group contains more than 5% of the paths towards every AS (e.g., $r > 5\%$), the inference can be considered reliable.

3) *Voting for First-Stage Inference*: Voting depends on the assumption that the error probability p in each group is small,

but this is not always true, especially for p2p links. To deal with this, we first increase the voting threshold. In particular, we construct *consensus links* where at least 80% of groups agree on their classification. Then, we filter these *consensus links* to find *trusted links* based on the following findings.

We conduct multiple voting scenarios with different group sizes S from 1 to 20, assessing the average inference error rate for p2p and p2c links observed by a specific proportion of groups. For instance, Figure 3c shows that p2p links seen by 30% of groups have an average error rate of about 20%. This data reveals that p2p links are more accurately classified when observed by fewer groups, while p2c links are more reliably identified when seen by more groups. The rationale is that links observed by a greater number of VPs are more likely to be p2c links, as illustrated in Figure 1a. So some of the inferred p2p links are actually mis-classified p2c links, particularly when observed by more groups. Based on this, we use a new rule: from *consensus links*, we only select p2p links seen by at most 20% of groups and p2c links seen by at least 80% of groups as *trusted links*.

For grouping by Origin ASes, its error rate is smaller than grouping by VPs under the same number of groups. Since the error probability p in equation (1) is smaller, we set the threshold for selecting *consensus links* to at least 50% groups. Likewise, based on Figure 4c, we select p2p consensus links that are observed by at most 50% groups, and p2c consensus links observed by at least 50% groups, as *trusted links*.

For the trade-off between quantity and accuracy, we construct two sets of *trusted links* based on the VP-based and Origin AS-based grouping. Next, we discard conflicting ones between them, and compute their union or intersection. For the other two empirical algorithms, the same way is used to tune voting parameters and three sets of *trusted links* are constructed. Now the relationships of *trusted links* are already determined, and all the other links are called *fuzzy links*, which have been assigned different types (or even no type if an empirical algorithm says so) by different numbers of groups.

D. Learning Algorithms for Fuzzy Link Inference

The observed characteristics of a link often deviate from what strict empirical rules have expected, so we try to use machine learning algorithms to further classify *fuzzy links*. We consider three categories of algorithms, i.e., unsupervised, semi-supervised and supervised learning algorithms.

TABLE IV
FEATURES USED IN BAYESIAN NETWORK

features	descriptions (with respect to a link ℓ)
relation proportion (N1*)	for ℓ 's two endpoint ASes, the proportion of their provider/peer/customer neighbors
triplet (N2)	the proportion of different relationships on the adjacent links of ℓ
distance to clique (N3)	the average distance from ℓ to clique ASes
#observing VPs (V1)	the number of VPs that can observe ℓ
VP position (V2*)	the hierarchical distribution of VPs that can observe ℓ
non-path-before (V3)	the number of adjacent p2p/p2c links (which never appear before ℓ on any path) ℓ has
#origin ASes (O1*)	the number of ASes who originate paths that pass through ℓ
origin position (O2*)	the hierarchical distribution of the ASes who originate paths that pass through ℓ
non-path-after (O3*)	the number of adjacent p2p/p2c links (which never appear after ℓ on any AS path) ℓ has
neighbor ratio (L1*)	for the two endpoint ASes of ℓ , the ratio between the numbers of their neighbors
co-located IXP (L2)	the number of IXPs that the two endpoint ASes of ℓ are co-located in
co-located facility (L3)	the number of facilities that the two endpoint ASes of ℓ are co-located in

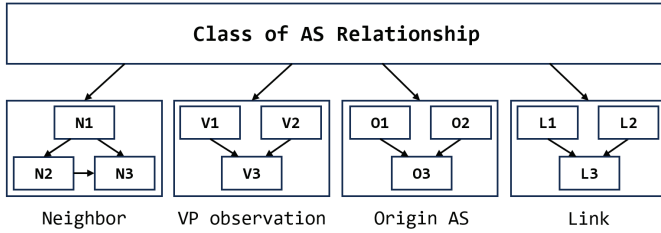


Fig. 5. Bayesian Network structure.

The specific models we use are Bayesian Network (BN), Label Propagation (LP) and XGBoost (XGB) respectively. Despite all three learning methods having more complex models, they often take a long time to train. For the relatively simple scenario of inferring AS relationships, these three models are deemed adequate.

1) *Bayesian Network*: ProbLink [23] employs a naive Bayesian model to estimate the probability of a link having a specific relationship. However, this approach oversimplifies the problem, as AS link features are interdependent, and combining too many features can lead to issues such as overfitting. To address this, HELA utilizes a Bayesian Network (BN) as its unsupervised learning model, which accounts for feature interdependencies and improves accuracy. The BN's operational principles and parameter determination follow those outlined in TopoScope and are omitted here for brevity.

HELA's advancements over TopoScope primarily lie in BN feature selection and dependency structure. Table IV lists the features used by BN in HELA, with asterisks marking those introduced by HELA, while others are inherited from ProbLink. These features are categorized into four groups. The first is mostly related to the surrounding environment of links, the second is related to VP observation, the third is related to origin ASes, and the fourth is related to the relation between the two endpoint ASes. Figure 5 illustrates the dependency

structure among these features, with arrows from the class of AS relationship to each category indicating that all features depend on the AS relationship classification.

2) *Label Propagation*: For semi-supervised learning, we utilize the Label Propagation (LP) algorithm [48], whose basic idea is to identify analogous features for a *fuzzy link* within a set of *trusted links*, thereby determining the relationship of this *fuzzy link*. This methodology entails two principal phases: similarity matrix construction and label propagation.

In the first step, a graph is constructed, whose nodes represent links, including trusted links and fuzzy links (i.e., labeled links and unlabeled links). The edge between node i and j denotes the similarity w_{ij} between them, defined as:

$$w_{ij} = \exp \left(-\frac{\|F_i - F_j\|^2}{\alpha^2} \right) \quad (2)$$

where F_i and F_j denote the eigenvectors of node i and j respectively, and α is the hyperparameter.

For the second step, labels are passed through the edges, and the greater the weight of an edge, the easier it is for a label to be propagated over. Assume that there are N edges and an $N \times N$ probabilistic transition, matrix P is defined as:

$$P_{ij} = P(i \rightarrow j) = \frac{w_{ij}}{\sum_{m=1}^k w_{im}} \quad (3)$$

where P_{ij} denotes the probability of label propagation from node i to j .

Suppose the number of trusted links is L , we define an $L \times 3$ label matrix Y_L whose i th row represents the label vector corresponding to the i th trusted link. Specifically, the vector is $(1, 0, 0)$ if the relationship of the link is p2p (similarly, $(0, 1, 0)$ for p2c and $(0, 0, 1)$ for c2p). Likewise, assume that there are U fuzzy links and a $U \times 3$ labels matrix Y_U . Combining the two matrices results in a label matrix $F = [Y_L; Y_U]$ with dimension $N \times 3$, in which a random assignment is applied to Y_U . Note that the vectors in the label matrix are probability labels, representing the probability of a relationship of a link.

The label propagation algorithm is as follows:

- 1) Propagate $F = PF$.
- 2) Reset the label vector corresponding to the trusted links $F_L = Y_L$.
- 3) Repeat from step (1) until F converges.

Since we treat the class of trusted links as their final result, step (2) deviates from the original algorithm, which row-normalizes F . In the LP method, the sole parameter to be determined is α . To manage the large number of links, we apply the KNN algorithm [49] to preserve the similarity among the k nearest neighbors for each node. After model training and parameter optimization, $\alpha = 0.2$ is the optimal setting.

3) *XGBoost*: The XGBoost [50] is a machine learning algorithm based on gradient boosting, which is less prone to overfitting and provides better generalization than traditional decision trees. HELA utilizes the XGBoost package [51] in Python to train the model, with the trusted links as the training set. After completing training, input the feature vectors of the fuzzy links and the model will output their relationships.

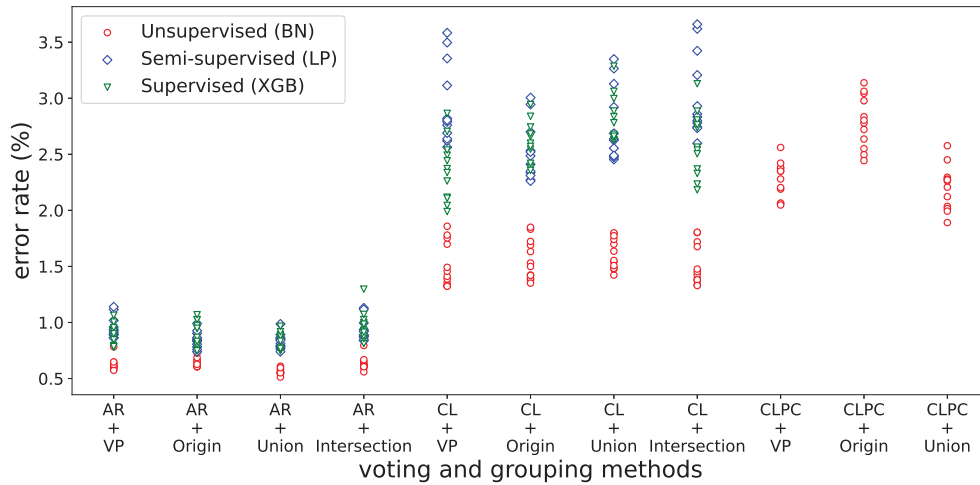


Fig. 6. Overall inference error rate for different combinations.

E. Algorithm Combination

For inferring the relationship of trusted links, there are 12 algorithm combinations in HELA, which are 3 heuristic algorithms (AS-Rank, Core2Leaf, Core2Leaf-PC, abbreviated as AR, CL and CLPC respectively) $\times$ 4 grouping and voting methods (VP, Origin AS, Union and Intersection). For inferring the fuzzy links, there are three machine learning algorithms, namely Bayesian Network (BN, unsupervised), Label Propagation (LP, semi-supervised), and XGBoost (XGB, supervised). As a result, HELA contains a total of $12 \times 3 = 36$ algorithm combinations. However, since the number of fuzzy links obtained from Core2Leaf PC is relatively large, the semi-supervised and supervised learning algorithms perform poorly. We exclude these experiments, leaving $36 - 8 = 28$ combinations finally. Figure 6 depicts the overall inference error rate.

In the decision of machine learning algorithm, Bayesian Network outperforms both Label Propagation and XGBoost. AS-Rank beats other heuristic algorithms. With AS-Rank as the heuristic algorithm and Bayesian Network as the machine learning algorithm, choosing Union as the grouping and voting method provides the lowest error rate and the highest stability. This outcome can be attributed to three main factors. First, Bayesian Networks effectively capture feature dependencies, reducing overfitting, which is common in other methods impacted by a large number of features. Second, AS-Rank, while maintaining accuracy, significantly increases the number of inferred links compared to other methods, providing a reliable baseline for learning inference. Third, the Union method accounts imbalance for both source ASes and VPs, ensuring accurate inference of both well-observed links around VPs and under-observed links near small source ASes. The result confirms our suspicion that links obtained by grouping and voting among VP and AS are precise enough. HELA will set them as trusted links to prevent machine learning algorithms from inferring them incorrectly.

On an Intel Xeon Gold 5118 with 500GB of memory, ASRank, ProbLink, and TopoScope require an average of 2 hours, 1 hour, and 2.5 hours, respectively, to complete inference. HELA's inference time varies between

2.5 and 3 hours, depending on the selected algorithm. Consequently, the integration of empirical and learning algorithms does not substantially increase computational complexity. However, HELA's design significantly enhances scalability. As demonstrated in Section V, the rules formulated by AS-Rank 11 years ago no longer reflect current conditions, and ProbLink's simplistic probabilistic inference design results in poor performance on top-tier links. In contrast, HELA's modular structure, which supports both empirical and learning-based algorithms, ensures adaptability to advanced techniques, leveraging the strengths of different algorithms to optimize its performance. We also evaluated the effects of various grouping strategies on HELA. The findings indicate that the average random allocation illustrated in Figure 6 yields the best performance. In contrast, extreme allocation methods, such as segregating observations of top- and bottom-tier VPs or grouping large and small source ASes separately, result in a slight reduction in the number of trusted links (approximately 0.7%~1.8%) and a marginal increase in error rates (0-0.2%). This is primarily due to AS-Rank's higher error rates when inferring under-observed groups, making consensus on certain links challenging. Nonetheless, the learning algorithm identifies reliable features within these links, effectively mitigating the errors.

F. Validation Dataset Construction

In 2012, Vasileios and Shi [34] compiled BGP community dictionaries, which were later integrated into prominent AS relationship inference algorithms such as AS-Rank, ProbLink, and TopoScope for constructing their validation datasets. However, due to the absence of disclosed construction criteria, updated dictionaries were unavailable over an extended period, leading to the gradual obsolescence of these validation datasets. Furthermore, Prehn and Feldmann [37] identified significant bias issues in the validation dataset, potentially compromising the accuracy of analytical results. Consequently, building on prior research, we develop a novel method for constructing BGP community dictionary and parsing RPSL information, aiming to generate a more balanced and comprehensive validation dataset. This approach is achieved by

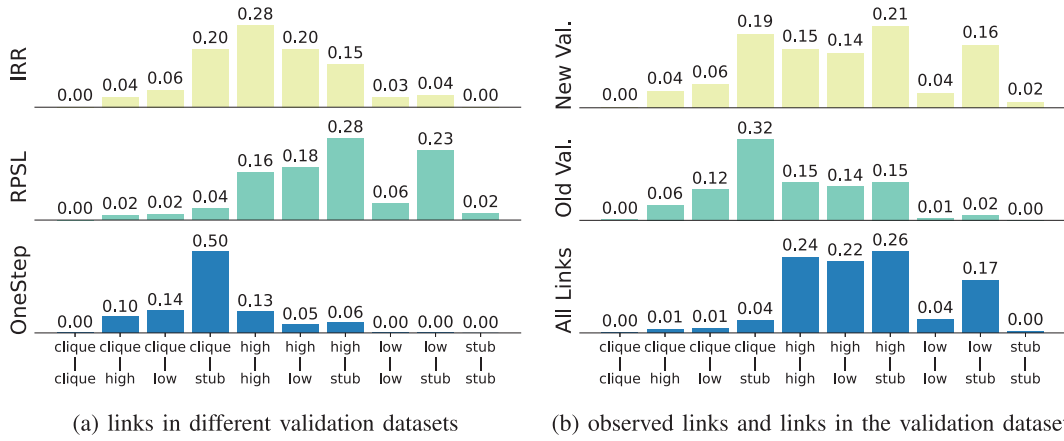


Fig. 7. Distribution of observed links and links in the validation dataset.

TABLE V
COMMUNITY KEYWORDS TO MARK AS RELATIONSHIPS

category to mark	keywords
provider	transit uplink(s)? upstream(s)? provider(s)? carrier(s)?
customer	downlink(s)? downstream(s)? customer(s)? client(s)?
peer	peer(ing)?(s)? ix(p)? (from at via through on by ixp_name)

TABLE VI
STATISTICS OF VALIDATION LINKS

data sources	#all links (p2c/p2p)	#links in the other two validation datasets
IRR community	53,248 (17,957/35291)	8,439
IRR RPSL	34,970 (26,793/8,177)	1,711
OneStep community	16,740 (12,301/4,439)	6,832

gathering data from major Regional Internet Registries (RIRs, including APNIC, ARIN, RIPE NCC, LACNIC, AFRINIC and RADB), as well as OneStep. RIRs facilitate voluntary policy reporting by ASes, enabling the extraction of community dictionary and RPSL data, while OneStep aggregates community records from select upper-tier AS websites, offering insights from a few crucial ASes.

After downloading the RIR file, we first extract key information by fields, such as AS number, AS-set name and corresponding members, and update time. Next, we remove records updated over a year ago and map AS-set annotations to AS links. We then use regular expressions to extract keywords from the remark field, as listed in Table V. Given that ASes often employ complex and non-standard natural language in policy statements, the combination of these rules with specific phrases (e.g., from, at, via, through, on, by, etc.) is also considered to indicate AS relationships. To identify the most commonly used keywords among ASes, we cross-check the AS relationships inferred from different keywords with those from the AS inference algorithm, considering only keywords with over 90% match accuracy as final candidates. These keywords typically have clear meanings, and given the high accuracy of current inference algorithms, we regard this cross-check as reliable. Upon obtaining AS annotations for the community, we collect BGP routes and identify the AS links

associated with these annotations for the validation set. For RPSL records, we use AS-set mappings to identify AS X and AS Y that formulate the import/export policy. If 1) X exports ANY to Y , 2) Y imports ANY from X , and 3) Y does not export ANY to X , the X - Y link is classified as a p2c relationship. Conversely, if neither X nor Y imports or exports ANY to each other, but only exchanges certain neighbors, then X - Y is labeled as a p2p link. These rules also achieve a cross-validation accuracy exceeding 90% with the existing inference algorithm. After applying these rules to the IRR and OneStep datasets, as well as with BGP routes in Jan 2023, the results in Table VI are obtained. It shows that OneStep community and IRR RPSL identify more p2c links, while IRR community finds the most labeled links. Additionally, OneStep and IRR community share about 6k overlapping links, and the total number of links from all sources is 87,909, which is 95% more than the ProbLink validation dataset [23].

To analyze the distributional characteristics of validation datasets from various sources, we classify the links in the validation dataset according to the criteria in Table II and present the findings in Figure 7a. The results indicate that IRR community predominantly involves high AS links, IRR RPSL favors links towards high or low ASes with stub ASes, and OneStep community links are biased between clique and stub ASes. Section V-E will show that HELA maintains accuracy and stability across validation datasets with varying distributions. Furthermore, we compare the distribution of our validation dataset (New Val.), ProbLink's validation dataset (Old Val.), and all AS links in Figure 7b, where our validation dataset exhibits a higher concentration in the L-S group and a reduced presence in the C-S group. The Jensen-Shannon Divergence (JS-Divergence) is a metric commonly employed to quantify the proximity of two distributions, with lower values indicating greater similarity. The JS-Divergence between the new validation dataset and all AS links is 0.05, as opposed to a value of 0.16 for the JS-Divergence between the old validation dataset and all links. It indicates that our new validation dataset exhibits a stronger alignment with the overall link distribution. Additionally, since there are 31 updates in AS community annotation, New Val. and Old Val. differ on 236 links, the impact of which will be evaluated in Section V-E.

V. EVALUATION

In this section, we evaluate HELA against TopoScope, AS-Rank and ProbLink. HELA adopts AS-Rank, Union voting and Bayesian Network as the final algorithm combination for evaluation, and the source codes of TopoScope, AS-Rank and ProbLink are published by their authors. Comparisons with other algorithms, such as UNARI [24] and GR-Model [33], were excluded due to their high error rates ($>20\%$). We demonstrate that HELA can achieve high accuracy in the following aspects:

Overall observations (Section V-B). Compared to the other three algorithms, HELA always performs higher accuracy and stability on datasets from 2021 $\sim$ 2023, with error rates between 0.5% and 0.7%. HELA reduces the average error rate of AS-Rank by $1.7\times$ for p2c links and $1.3\times$ for p2p links, while reducing that of ProbLink by $3.0\times$ for p2c links and $1.5\times$ for p2p links, and reducing that of TopoScope by $\sim 1.3\times$ for p2p links. Voting and Bayesian Networks significantly enhance inference accuracy on trusted and fuzzy links, respectively.

Hard links and critical links (Section V-C). HELA performs up to $2.7\times$ better for a set of *hard links*, and up to $2.9\times$ better for another set of *critical links*, both of which greatly affect the overall inference accuracy.

Worse observations (Section V-D). HELA achieves more accurate and stable results when the VP observations are more incomplete or biased. This situation will be very possible when the Internet grows faster than our observation capability.

Different validation datasets (Section V-E). Despite variations in accuracy and distribution among validation datasets from diverse sources, HELA always exhibits superior precision and stability across all datasets, reducing the error rate by up to $1.7\times$ and the standard deviation by up to $4.0\times$. Besides, the error rates of all algorithms in the old validation dataset are $\sim 2.8\times$ of that in the new validation dataset, highlighting the enhanced accuracy of the new validation dataset.

This section concludes with the replication of experiments from TopoScope to validate the effectiveness of HELA in two practical applications: route leakage and AS path inference.

Comprehensive analysis reveals that AS-Rank heavily suffers on incomplete observations, resulting in low inference accuracy for hard links, typically those in the lower tiers. ProbLink is constrained by the imbalance in VP observations, leading to misclassification of some links with either sufficient or insufficient observations, thereby reducing its accuracy for critical link inference. While TopoScope and HELA address some of these issues, HELA further accounts for the imbalance in observations around source ASes. Consequently, HELA demonstrates superior inference accuracy for links involving smaller source ASes (e.g., “moderately observed,” “clique-stub,” “tier-1 to small ASes,” and “between unequal ASes” links as discussed in Section V-C) compared to TopoScope.

A. Datasets

Our BGP routing data are primarily derived from Route Views [25], RIPE RIS [26], and Isolario [27].⁵ Monthly

⁵Since Isolario has stopped running, we only have data from it for 2021. However, this does not impact our analysis in this section.

TABLE VII

STATISTICS OF ALL LINKS AND LABELED LINKS

year	#all links (p2c/p2p)	#labeled links (p2c/p2p)
2021	490,518 (146,586/343,932)	87,909 (46,825/41,084)
2022	503,063 (152,769/350,294)	88,722 (46,931/41,791)
2023	516,723 (157,125/359,598)	90,876 (48,549/42,327)

snapshots of BGP routing tables, captured on the first day of each month from January 2021 to December 2023, are obtained from ~ 380 VPs by over 40 route collectors, resulting in a total of 36 snapshots. Each snapshot, following preprocessing as described later, is employed as the input dataset to evaluate the inference algorithms. These datasets contain $\sim 500K$ links as shown in Table VII, where the numbers of p2p or p2c links are listed in brackets. Additionally, we intentionally select diverse subsets from each dataset to examine the efficacy of the algorithms under limited and biased Internet observation.

Preprocessing. We preprocess the dataset by filtering mistakes or noises that may impact the inference in four steps:

- 1) Removing paths that contain unassigned ASNs.
- 2) Removing consecutive duplicate ASNs caused by prepending.
- 3) Filtering paths with loops.
- 4) Fetching the ASNs corresponding to the Route Server from PeeringDB and IXPDB, and removing these ASNs from the paths, since the Route Server does not have a commercial relationship with a normal AS.

Validation Dataset. On the first day of each month from 2021 $\sim$ 2023, we employ IRR to build a community dictionary and extract RPSL information, and utilize OneStep for an additional community dictionary. Subsequently, the rules outlined in Section IV-F are applied to determine labeled AS relationships from RPSL and relationships denoted by the community dictionaries within monthly BGP routes. Conflicting relationships within the three validation datasets are then excised, yielding 36 monthly sets. The performance of various algorithms on these validation datasets is assessed in Section V-E, while for other sections, the sets are aggregated to form a composite validation dataset.

B. Overall Accuracy

First, we present the inference error on all links. The error rate is defined as in Section IV-C. A link is counted as a false positive (#fp) if its inferred relationship differs from the actual one (e.g., a p2c link misclassified as p2p or c2p, or a p2p link misclassified as p2c or c2p). Conversely, correctly inferred links are counted as true positives (#tp). The error rate is then given by $\frac{\#fp}{\#tp + \#fp}$. Figure 8a depicts the overall error distribution for HELA, TopoScope, AS-Rank and ProbLink, respectively. We can see that, compared with the other three, HELA is nearly always more accurate and stable, with error rates between 0.5% and 0.7%. AS-Rank is very unstable, with an error rate varying between 0.6% and 1.6%, mainly because its empirical rules are not flexible enough to adapt to the variations in the data. Although ProbLink has a smaller variation due to its probabilistic principle, its average error is

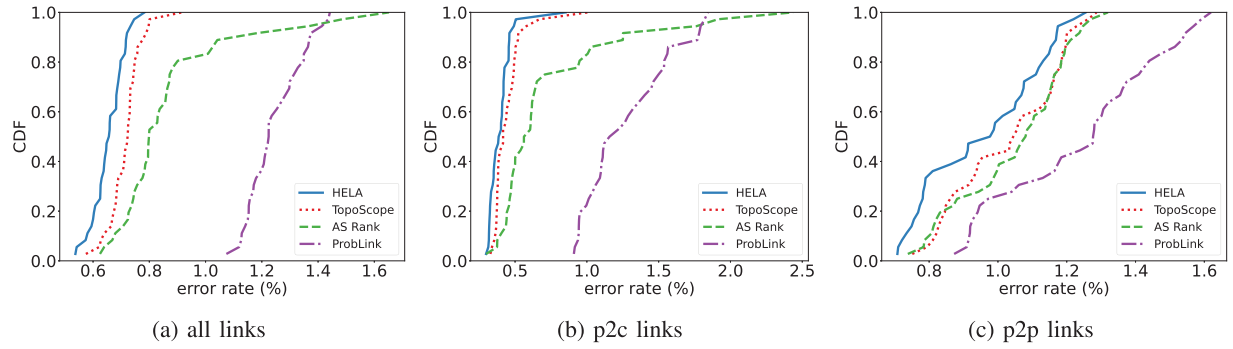


Fig. 8. Inference error rate for different links from 2021 ~ 2023.

TABLE VIII
ERROR REDUCTION RATIO FROM 2021 ~ 2023

	reduction ratio against AS-Rank				
	all	p2c	p2p	critical	hard
max	$2.89 \times$	$4.98 \times$	$1.71 \times$	$2.89 \times$	$1.79 \times$
min	$1.10 \times$	$0.93 \times$	$1.12 \times$	$1.10 \times$	$1.14 \times$
average	$1.45 \times$	$1.67 \times$	$1.30 \times$	$1.45 \times$	$1.36 \times$
	reduction ratio against ProbLink				
	all	p2c	p2p	critical	hard
max	$2.47 \times$	$3.87 \times$	$1.82 \times$	$2.47 \times$	$2.69 \times$
min	$1.76 \times$	$2.3 \times$	$1.34 \times$	$1.76 \times$	$1.48 \times$
average	$2.07 \times$	$2.97 \times$	$1.53 \times$	$2.07 \times$	$1.97 \times$
	reduction ratio against TopoScope				
	all	p2c	p2p	critical	hard
max	$1.31 \times$	$1.37 \times$	$1.49 \times$	$1.31 \times$	$1.43 \times$
min	$1.07 \times$	$0.87 \times$	$1.08 \times$	$1.07 \times$	$1.11 \times$
average	$1.19 \times$	$1.04 \times$	$1.28 \times$	$1.19 \times$	$1.29 \times$

always the largest, roughly double that of HELA. The overall distribution trend of TopoScope is similar to that of HELA, but HELA achieves a higher precision. This is because HELA adopts the Union voting method to obtain more accurate trusted links from the Origin AS grouping and voting, which prevents them from being inferred incorrectly in the second stage.

We further plot the errors for p2c or p2p links in Figure 8b and Figure 8c. For the p2c links, HELA and TopoScope have the lowest error rates, while AS Rank and ProbLink perform relatively poorly. For the p2p links, the distribution trends of all algorithms are consistent, but HELA performs better than TopoScope and AS-Rank, while ProbLink performs the worst.

In Table VIII, we list the relative error reduction ratio HELA achieves for different links. For example, compared with AS-Rank, ProbLink and TopoScope, HELA respectively reduces the overall error rate by 1.45 or 2.07 or 1.19 $\times$ on average, and by 2.89 or 2.47 or 1.31 $\times$ in the best case. The results therein show that AS-Rank performs better than ProbLink on average, especially for p2c links,⁶ but is less stable.

Figure 9 further reveals the individual contribution of each step in the HELA framework, by depicting how well voting and machine learning algorithms, as well as the other three algorithms, can classify the relationships for p2p or p2c trusted links and fuzzy links. It can be seen that, HELA achieves high precision and performs better than

AS-Rank for each type of trusted links. However, HELA-Bayes and ProbLink perform poorly for overall trusted links. This indicates that using voting to find the trusted links can effectively mitigate against observation uncertainty, protecting them from erroneous empirical rules or biased probabilistic models. Additionally, HELA performs better than TopoScope for overall trusted links. In particular, HELA and TopoScope perform almost equally perfectly in p2c trusted link inference (Figure 9b), but TopoScope performs worse than HELA and AS-Rank for p2p trusted links (Figure 9c), since the trusted links in HELA contain links inferred in the second stage of TopoScope (derived from grouping and voting by Origin AS in HELA), which are more suitable for empirical inference.

On the other hand, Figure 9d illustrates the greater challenge in accurately inferring fuzzy links compared to trusted links, with significant variations in error patterns and algorithm performance rankings. Specifically, voting and AS-Rank do not work well for p2p fuzzy links (Figure 9f), suggesting the superiority of probabilistic models over empirical rules for these cases. In contrast, for p2c fuzzy links (Figure 9e), ProbLink is much worse than the others, while the Bayesian Network, which considers the dependency of different features, especially the concentration bias of VPs, achieves a more accurate and stable result. Finally, HELA performs better than TopoScope for each type of fuzzy links, due to the addition of features based on Origin AS to the HELA.

C. Hard Links and Critical Links

ProbLink argues that a set of *hard links* (i.e., links of small degree ASes or observed by 50 ~ 100 VPs, non-VP and non-Tier1 links, clique to stub links, and conflict links in CoreToLeaf) are the main cause of inference errors, so we are interested in how well the inference algorithms can deal with them. In Figure 10, we plot how well different algorithms can infer the relationships for the five groups. The error distribution curves show HELA nearly always performs the best in all groups, surpassing even TopoScope.

Hard links introduce inference errors, but not necessarily errors that have great impacts on applications that depend on the (real or inferred) relationships. In this regard, we believe some AS links play a more important role than others. For example, tier-1 transit links, or bottleneck links [6], [52] are likely to have a greater impact on the performance of various

⁶The average reduction ratio achieved by HELA for p2c links is 1.67 (v.s. AS-Rank), 2.97 (v.s. ProbLink) and 1.04 (v.s. TopoScope), while that for p2p link is 1.3 (v.s. AS-Rank), 1.53 (v.s. ProbLink) and 1.28 (v.s. TopoScope).

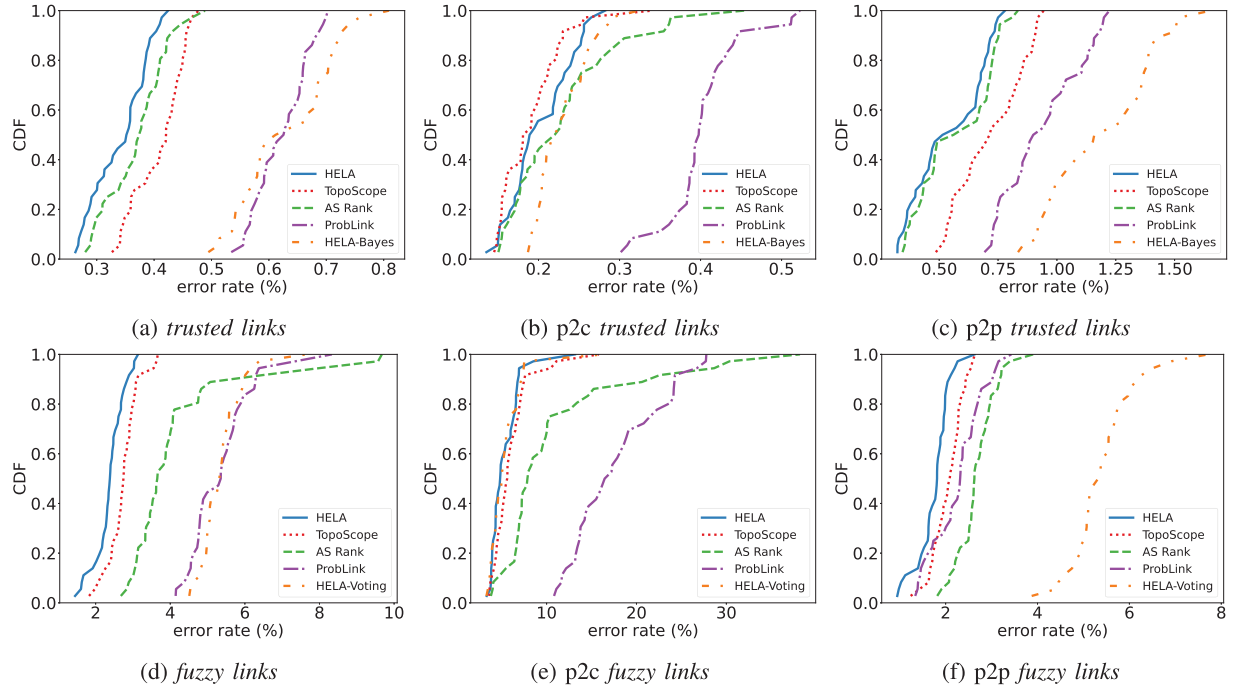


Fig. 9. Inference error rate for *trusted links* and *fuzzy links* from 2021 ~ 2023.

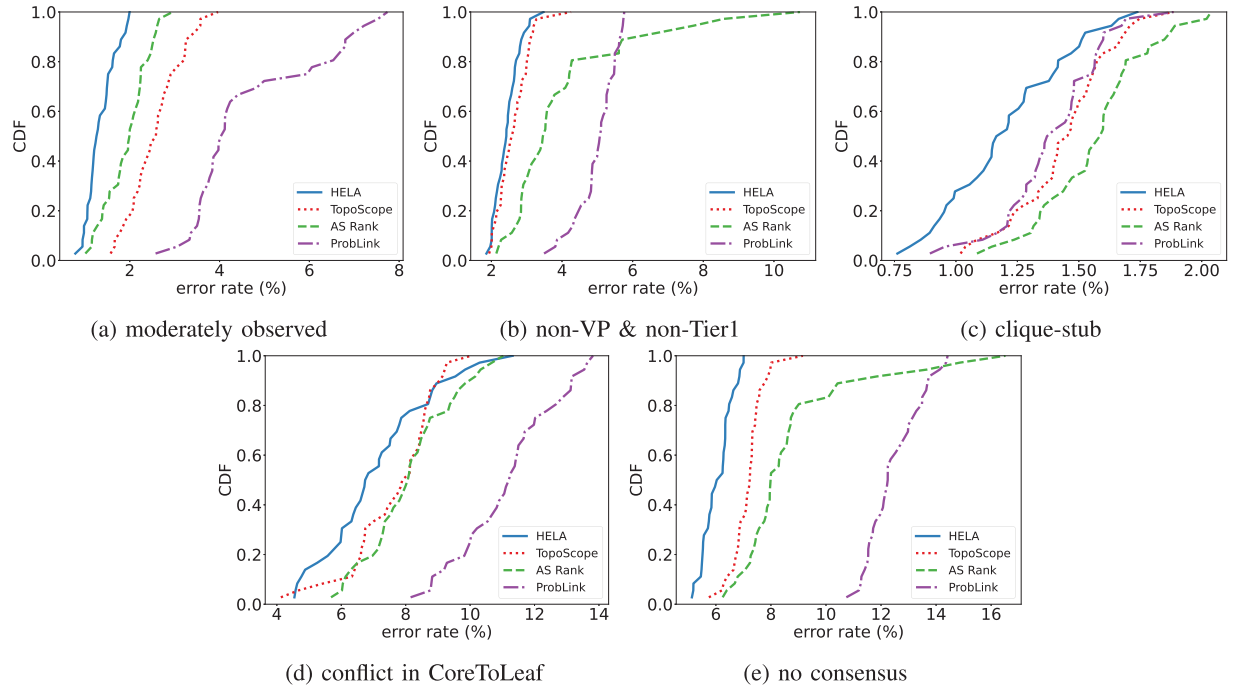


Fig. 10. Inference error rate for *hard links* from 2021 ~ 2023.

applications, and the inference accuracy for their relationships deserves our attention.⁷ For this reason, we select a few links which TopoScope call *critical links*, and compare how well different algorithms can infer their relationships. TopoScope defined a set of *critical links* as the following five categories.

⁷Unlike the “basic links” in AS-Rank and ProbLink, whose inferred relationships greatly impact the subsequent inference procedure, the accuracy of “critical links” may not necessarily affect the accuracy of other links greatly.

(1) Tier1 AS links. They appear in most AS paths and will affect most applications. (2) Upper tier links. We select all high tier ASes whose degree is larger than 1000. They are large ASes, and similar to tier-1 AS links, their interconnections or connections with tier-1 ASes are also important. (3) Frequently occurring links. We select links that appear more than 10,000 times in the observed AS paths. (4) Links between Tier1 and small ASes (i.e. stub ASes and low tier ASes with small degrees). Some of them may act as important content delivery

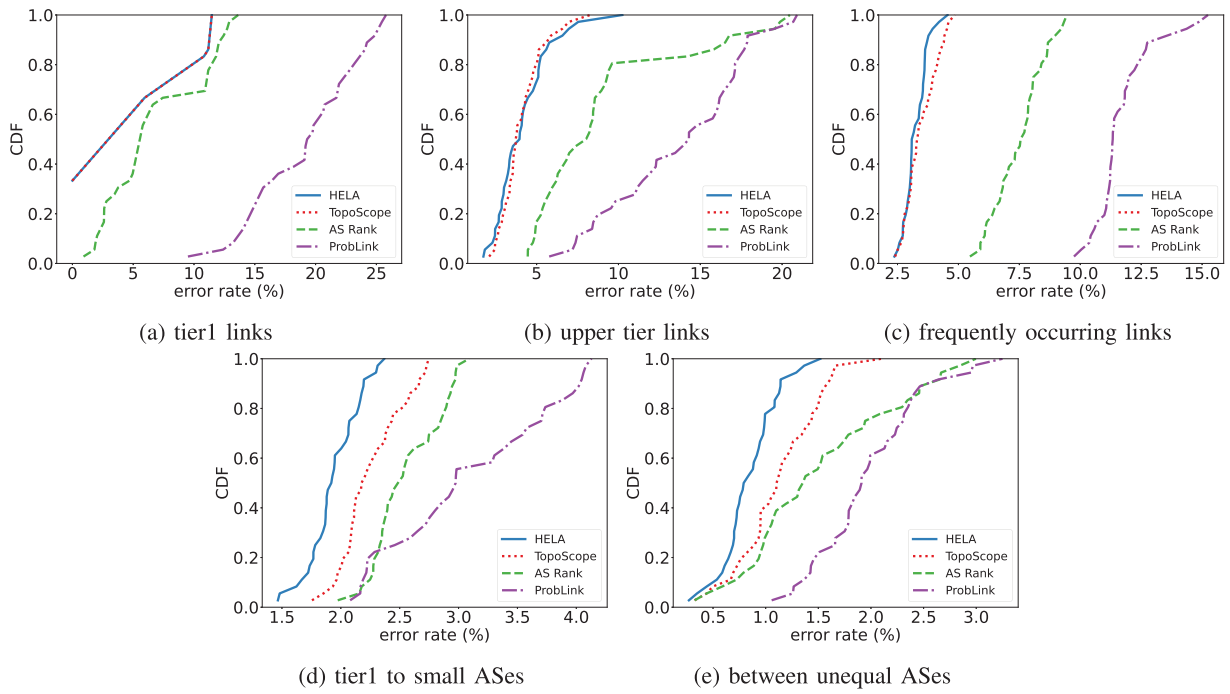


Fig. 11. Inference error rate for *critical links* from 2021 ~ 2023.

channels [53]. (5) Links between highly unequal ASes. They are similar to the last case, where the two endpoint ASes of a link differ greatly in their degrees.

Figure 11 plots the inference accuracy of different algorithms on critical links, where HELA again performs nearly always the best. It is interesting to see that, ProbLink always performs worse than the other algorithms, where its probabilistic model is likely to be tricked by the very high occurrence frequency of those upper layer links. On the other hand, AS-Rank performs worse in the last two groups, illustrating that a lack of observation on small ASes may skew its inference results. Table VIII details the relative improvement of HELA against AS-Rank, ProbLink and TopoScope for these links. On average, TopoScope reduces errors by 1.5 and $1.4\times$ against AS-Rank, by 1.2 and $1.3\times$ against TopoScope, and by 2.1 and $2.0\times$ against ProbLink, for critical and hard links respectively, with maximum ratios reaching 2.9 and $2.7\times$.

D. More Incomplete or Biased Dataset

Both the scale and complexity of the Internet keep growing, whereas the number of VPs may not be able to keep up with the pace. Therefore, we conduct a stress test, where the observation dataset will be more incomplete or biased, to forecast how the algorithms will perform in the future. Specifically, we randomly choose the observations from 20%, 40%, 60% and 80% of all available Route Views and RIPE RIS VPs, from VPs concentrated in the high or low tier of the Internet (as defined in Table II), as well as from VPs from Isolario [27]. For comparison, we also use the full RV+RIS dataset, so there are totally 8 categories of incompleteness or bias datasets. Using routes collected in each month from 2021 ~ 2023, we compose 36 different datasets for each category, so we run 288×8 tests against each algorithm in total.

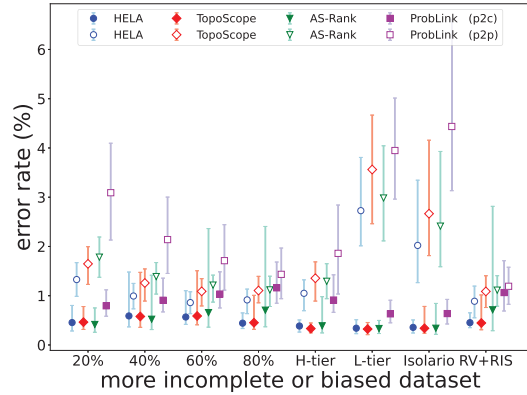


Fig. 12. Inference error rate (min, average, max) in more incomplete and biased datasets (H-tier and L-tier consist of around 75% and 20% of all observations respectively, and the size of the Isolario dataset is around one half of RV+RIS).

Figure 12 plots the error rate of each algorithm in different categories, where a point represents the average error rate, and an error bar represents the corresponding minimum and maximum error rates. HELA performs consistently better on these more incomplete or biased datasets, and the overall distribution trend of TopoScope is similar to that of HELA. For p2c links, ProbLink performs the worst, and HELA, TopoScope, and AS Rank are relatively similar, with HELA exhibiting the lowest fluctuation range and the most stable performance. For p2p links, ProbLink still performs the worst, with TopoScope and AS Rank showing similar performance. Overall, HELA has the lowest fluctuation range and the lowest error rate in the four algorithms.

For p2p and p2c links, HELA can reduce the average error by up to 1.4 and $1.6\times$ compared with AS-Rank, and by up to 2.3 and $2.6\times$ compared with ProbLink. Although HELA performs similarly to TopoScope on p2c links, HELA is more

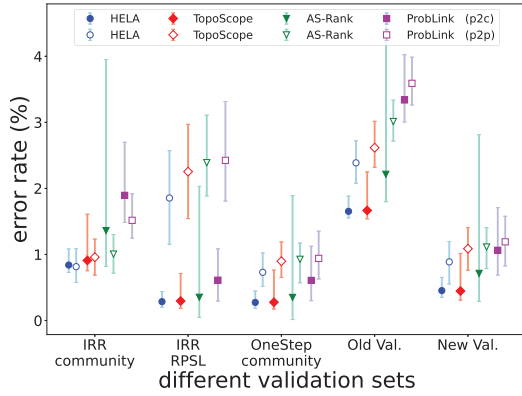


Fig. 13. Inference error rate in different validation datasets.

accurate on p2p links, reducing the average error by up to $1.3\times$. HELA can reduce the standard deviation by as much as 4.5 , 3.3 and $1.6\times$ compared with AS-Rank, ProbLink and TopoScope, indicating that HELA is the most stable algorithm.

E. Different Validation Datasets

In Section IV-F, we delineated the composition of our validation dataset (i.e., IRR community, IRR RPSL, and OneStep community), and show their distinctive hierarchical distributions and link components in Figure 7 and Table VI. This section assesses the performance of various algorithms across these three validation datasets, along with the overall validation dataset and the historical set provided by ProbLink [23], yielding a total of 5 validation datasets. In line with Section V-D, we employed BGP routes collected monthly from 2021 ~ 2023, resulting in 36 distinct datasets for each validation dataset. Consequently, we conducted a total of 36×5 experiments.

Figure 13 presents the inference error rates of each algorithm across different validation datasets, with each point indicating the average error rate and the error bars representing the minimum and maximum error rates. Analysis reveals that, across different validation datasets, the overall performance of the OneStep community is slightly better than that of the new validation dataset, likely due to the predominant distribution of its links in the upper tier of the Internet, which are effectively observed and inferred. The p2c error rate for the IRR community is marginally higher, whereas for p2p links, it is marginally lower. IRR RPSL demonstrates a lower p2c error rate but a relatively higher p2p error rate, which reflects the conservatism of ASes in recording their own import/export policies in RPSL, implying that IRR RPSL alone may not accurately capture p2p relationships. In contrast, the old validation dataset experiences a significant drop in accuracy due to community label updates, with 236 (0.54%) labels being incorrect (it's important to note that this does not encompass all community updates but only those where the new and old validation datasets differ). For the various algorithms, the performance trends across different validation datasets are similar. AS-Rank exhibits substantial fluctuations, and ProbLink shows high error rates. HELA, however, consistently demonstrates more accurate and stable performance, underscoring its robustness.

The average error rate of the old validation dataset is $3.2\times$ higher for p2c links, $2.6\times$ higher for p2p links and $2.8\times$ higher for all links compared to the new validation dataset. Given the observed updates in communities, the new validation dataset is deemed more representative of actual Internet AS relationships. For the IRR community, IRR RPSL, and OneStep community, the respective error rates for p2c and p2p links are $(1.9, 0.9)$, $(0.6, 2.0)$, and $(0.5, 0.8)\times$ higher than those of the new validation dataset. Across different validation datasets, the average error rates of AS-Rank, ProbLink, and TopoScope are at most 1.5 , 2.1 , and $1.2\times$ higher than that of HELA, with the maximum standard deviations being 4.0 , 2.1 , and $1.6\times$ higher, respectively. We believe it reflects the inference accuracy of HELA under various validation scenarios.

F. Practical Applications

To demonstrate the impact of HELA's accurate inference of AS links on Internet applications, we replicated the route leakage and AS path inference experiments within the TopoScope framework [36], using 2023 data and maintaining all other experimental parameters consistent with TopoScope (details omitted for brevity). The results indicate that for route leakage evaluation, HELA achieves an average inference accuracy of 81.3%, outperforming TopoScope by $1.1\times$ (73.4%), AS-Rank by $1.9\times$ (42.6%), and ProbLink by $2.6\times$ (31.6%). In AS path inference evaluation, HELA's average accuracy reaches 42.1%, exceeding TopoScope by $1.1\times$ (39.7%), AS-Rank by $1.4\times$ (29.6%), and ProbLink by $2.2\times$ (19.1%). These results highlight the substantial advantages of HELA's precise inference of critical Internet links for practical applications.

VI. CONCLUSION

With an in-depth analysis on the challenges in AS relationship inference, we propose HELA, a framework for inferring and evaluating AS relationships with a hybrid of empirical and machine learning algorithms, which makes them complement each other, and allows different combinations. We introduce the internal structure of the HELA and evaluate the combination of various empirical algorithms, grouping and voting methods, and machine learning algorithms. With the optimal ensemble of AS-Rank + Union + Bayesian Network, HELA achieves its optimal performance. Furthermore, we develop a novel BGP community dictionary and RPSL information extraction method, and assemble a comprehensive and well-balanced validation dataset from multiple sources. Through a thorough evaluation across various validation datasets, we confirm HELA's superiority in both accuracy and stability compared to other algorithms, thereby enhancing its applicability to inferring larger and more complex Internet topology.

REFERENCES

- [1] Y. Tu, "How robust is the Internet?," *Nature*, vol. 406, no. 6794, pp. 353–354, Jul. 2000.
- [2] J. C. Doyle et al., "The 'robust yet fragile' nature of the Internet," *Proc. Nat. Acad. Sci. USA*, vol. 102, no. 41, pp. 14497–14502, 2005.
- [3] X. Shi, Y. Xiang, Z. Wang, X. Yin, and J. Wu, "Detecting prefix hijackings in the Internet with Argus," in *Proc. Internet Meas. Conf.*, Boston, MA, USA, 2012, pp. 15–28.

- [4] A. Dhamdhere et al., "Inferring persistent interdomain congestion," in *Proc. Conf. ACM Special Interest Group Data Commun.*, Aug. 2018, pp. 1–15.
- [5] K. Sriram, D. Montgomery, B. Dickson, K. Patel, and A. Robachevsky, "Methods for detection and mitigation of BGP route leaks," Tech. Rep., 2019.
- [6] Y. Yang et al., "Inter-domain routing bottlenecks and their aggravation," *Comput. Netw.*, vol. 162, Oct. 2019, Art. no. 106839.
- [7] J. M. Smith and M. Schuchard, "Routing around congestion: Defeating DDoS attacks and adverse network conditions via reactive BGP routing," in *Proc. IEEE Symp. Secur. Privacy (SP)*, San Francisco, CA, USA, May 2018, pp. 599–617.
- [8] Y. Sun, A. Edmundson, N. Feamster, M. Chiang, and P. Mittal, "Counter-RAPTOR: Safeguarding Tor against active routing attacks," in *Proc. IEEE Symp. Security Privacy*, San Jose, CA, USA, May 2017, pp. 977–992.
- [9] M. Apostolaki, G. Marti, J. Müller, and L. Vanbever, "SABRE: Protecting Bitcoin against routing attacks," in *Proc. Netw. Distrib. Syst. Secur. Symp.*, 2019.
- [10] M. Apostolaki, A. Zohar, and L. Vanbever, "Hijacking Bitcoin: Routing attacks on cryptocurrencies," in *Proc. IEEE Symp. Secur. Privacy (SP)*, May 2017, pp. 375–392.
- [11] A. Cohen, Y. Gilad, A. Herzberg, and M. Schapira, "Jumpstarting BGP security with path-end validation," in *Proc. ACM SIGCOMM Conf.*, 2016, pp. 342–355.
- [12] P. Gill, M. Schapira, and S. Goldberg, "Let the market drive deployment: A strategy for transitioning to bgp security," in *Proc. ACM SIGCOMM Comput. Commun. Rev.*, 2011, vol. 41, no. 4, pp. 14–25.
- [13] S. Goldberg, M. Schapira, P. Hummon, and J. Rexford, "How secure are secure interdomain routing protocols," in *Proc. ACM SIGCOMM Conf.*, Aug. 2010, pp. 87–98.
- [14] R. Lychev, S. Goldberg, and M. Schapira, "BGP security in partial deployment: Is the juice worth the squeeze?," in *Proc. ACM SIGCOMM Conf. Comput. Commun. Rev.*, 2013, pp. 171–182.
- [15] L. Gao, "On inferring autonomous system relationships in the Internet," *IEEE/ACM Trans. Netw.*, vol. 9, no. 6, pp. 733–745, Aug. 2001.
- [16] L. Subramanian, S. Agarwal, J. Rexford, and R. H. Katz, "Characterizing the Internet hierarchy from multiple vantage points," in *Proc. 21st Annu. Joint Conf. IEEE Comput. Commun. Societies*, vol. 2, Jun. 2002, pp. 618–627.
- [17] G. Di Battista, M. Patrignani, and M. Pizzonia, "Computing the types of the relationships between autonomous systems," in *Proc. 32nd Annu. Joint Conf. IEEE Comput. Commun. Societies*, Mar. 2003, pp. 156–165.
- [18] B. Zhang, R. Liu, D. Massey, and L. Zhang, "Collecting the Internet AS-level topology," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 35, no. 1, pp. 53–61, Jan. 2005.
- [19] X. Dimitropoulos et al., "AS relationships: Inference and validation," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 37, no. 1, pp. 29–40, 2007.
- [20] E. Gregori, A. Improta, L. Lenzini, L. Rossi, and L. Sani, "BGP and inter-AS economic relationships," in *Proc. Int. Conf. Res. Netw.*, Jan. 2011, pp. 54–67.
- [21] W. Willinger and M. Roughan, "Internet Topology research redux," in *Proc. ACM SIGCOMM*, Jan. 2013, pp. 1–12.
- [22] M. Luckie, B. Huffaker, A. Dhamdhere, V. Giotas, and K. Claffy, "AS relationships, customer cones, and validation," in *Proc. Conf. Internet Meas.*, Oct. 2013, pp. 243–256.
- [23] Y. Jin, C. Scott, A. Dhamdhere, V. Giotas, A. Krishnamurthy, and S. Shenker, "Stable and practical as relationship inference with ProBLink," in *Proc. NSDI*, 2019, pp. 581–598.
- [24] G. Feng, S. Seshan, and P. Steenkiste, "UNARI: An uncertainty-aware approach to AS relationships inference," in *Proc. 15th Int. Conf. Emerg. Netw. Exp. Technol.*, Dec. 2019, pp. 272–284.
- [25] *University of Oregon Route Views Project*. [Online]. Available: <http://www.routeviews.org/routeviews/>
- [26] *Ripe(RIS)*. [Online]. Available: <https://www.ripe.net/ris/>
- [27] *Isolario*. [Online]. Available: <https://www.isolario.it/>
- [28] E. Gregori, A. Improta, L. Lenzini, and C. Orsini, "The impact of IXPs on the AS-level topology structure of the Internet," *Comput. Commun.*, vol. 34, no. 1, pp. 68–82, Jan. 2011.
- [29] V. Giotas and S. Zhou, "Improving the discovery of IXP peering links through passive BGP measurements," in *Proc. IEEE Conf. Comput. Commun. Workshops (INFOCOM WKSHPS)*, Apr. 2013, pp. 121–126.
- [30] G. Nomikos et al., "O peer, where art thou? Uncovering remote peering interconnections at IXPs," in *Proc. Internet Meas. Conf.*, New York, NY, USA, Oct. 2018, pp. 265–278.
- [31] J. Xia and L. Gao, "On the evaluation of as relationship inferences [Internet reachability/traffic flow applications]," in *Proc. IEEE Global Telecommun. Conf.*, vol. 3, 2004, pp. 1373–1377.
- [32] V. Giotas and S. Zhou, "Inferring as relationships from bgp attributes," Tech. Rep., 2011.
- [33] L. Gao and J. Rexford, "Stable internet routing without global coordination," *IEEE/ACM Trans. Netw.*, vol. 9, no. 6, pp. 681–692, Dec. 2001.
- [34] G. Vasilios and Z. Shi, "Valley-free violation in Internet routing," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2012, pp. 1193–1197.
- [35] V. Giotas, M. Luckie, B. Huffaker, and K. Claffy, "Inferring complex AS relationships," in *Proc. Conf. Internet Meas.*, Nov. 2014, pp. 23–30.
- [36] Z. Jin, X. Shi, Y. Yang, X. Yin, Z. Wang, and J. Wu, "Toposcope: Recover as relationships from fragmentary observations," in *Proc. ACM Internet Meas. Conf.*, 2020, pp. 266–280.
- [37] L. Prehn and A. Feldmann, "How biased is our validation (data) for AS relationships?," in *Proc. 21st ACM Internet Meas. Conf.*, Nov. 2021, pp. 612–620.
- [38] R. Oliveira, D. Pei, W. Willinger, B. Zhang, and L. Zhang, "The (In)Completeness of the observed Internet AS-level structure," *IEEE/ACM Trans. Netw.*, vol. 18, no. 1, pp. 109–122, Feb. 2010.
- [39] H. Chang, R. Govindan, S. Jamin, S. J. Shenker, and W. Willinger, "Towards capturing representative AS-level Internet topologies," *Comput. Netw.*, vol. 44, no. 6, pp. 737–755, Apr. 2004.
- [40] K. Chen, C. Hu, W. Zhang, Y. Chen, and B. Liu, "On the eyeshots of BGP vantage points," in *Proc. IEEE Global Telecommun. Conf.*, Nov. 2009, pp. 1–6.
- [41] *Caida*. [Online]. Available: <http://www.caida.org>
- [42] V. Giotas, M. Luckie, B. Huffaker, and k. claffy, "IPv6 AS relationships, cliques, and congruence," in *Proc. Int. Conf. Passive Act. Netw. Meas.*, Jan. 2015, pp. 111–122.
- [43] J. S. Varghese and L. Ruan, "A machine learning approach to edge type inference in Internet AS graphs," in *Proc. IEEE INFOCOM*, Apr. 2016, pp. 71–76.
- [44] E. Gregori, A. Improta, L. Lenzini, L. Rossi, and L. Sani, "Discovering the geographic properties of the Internet AS-level topology," *Netw. Sci.*, vol. 3, nos. 1–4, pp. 34–42, Dec. 2013.
- [45] R. Chandra, P. Traina, and T. Li, "Bgp communities attribute," Tech. Rep., 1996.
- [46] C. Alaettinoglu et al., "Routing policy specification language (RPSL)," Tech. Rep. RFC 2622, 1999.
- [47] (2024). *Onestep*. [Online]. Available: <https://onestep.net>
- [48] X. Zhu and Z. Ghahramani, "Learning from labeled and unlabeled data with label propagation," Citeseer, Tech. Rep., 2002.
- [49] L. Peterson, "K-nearest neighbor," *Scholarpedia*, vol. 4, no. 2, p. 1883, 2009.
- [50] T. Chen and T. He, "XGBoost: eXtreme Gradient Boosting," *R Package Version 0.4-2*, vol. 1, no. 4, pp. 1–4, 2015.
- [51] dmlc.(2022). *Xgboost Extreme Gradient Boosting*. [Online]. Available: <https://github.com/dmlc/xgboost>
- [52] M. S. Kang and V. D. Gligor, "Routing bottlenecks in the Internet: Causes, exploits, and countermeasures," in *Proc. ACM SIGSAC Conf. Comput. Commun. Security*, 2014, pp. 321–333.
- [53] P. Faratin, D. D. Clark, S. Bauer, and W. Lehr, "Complexity of Internet interconnections: Technology, incentives and implications for policy," Tech. Rep., 2007.