Balancing Interests: A Policy-Based Approach to Internet Peer Selection

Martin Thodi
University of Cape Town
South Africa
thdmar002@myuct.ac.za

Amreesh Phokeer
Internet Society
Mauritius
phokeer@isoc.org

Josiah Chavula
University of Cape Town
South Africa
josiah.chavula@uct.ac.za

,

***Abstract*—This paper presents a novel, policy-aware approach to selecting and evaluating potential peering partners among Autonomous Systems in the Internet ecosystem. We develop a quantitative scoring model that combines requirements from peering policy documents with AS profile features from public datasets. Our model computes compatibility scores for potential AS peering pairs based on mutual requirement satisfaction, addressing limitations in existing approaches that often overlook specific policy requirements. We validate the proposed model using machine learning techniques and demonstrate its robust predictive performance across diverse ASes. The results reveal the individualized nature of peering strategies and the varying importance of different factors across ASes. This research contributes to a deeper understanding of Internet interconnection dynamics and provides a valuable tool for network operators to optimize their peering strategies.**

***Index Terms*—peering, autonomous systems, relationships, mutual satisfaction**

I. Introduction

The Internet's global connectivity is fundamentally built upon relationships between Autonomous Systems (ASes), primarily through transit arrangements and settlement-free peering [1]. While offering significant benefits such as reduced transit costs and improved network performance, settlement-free peering often involves a time-consuming establishment process that can span several months [2]. Furthermore, some ASes face challenges in identifying and evaluating optimal peering partners [2].

To address this challenge, we propose a novel peer selection model that carefully considers individual ASes' policies and requirements. Our approach leverages natural language processing (NLP) techniques to extract relevant information from unstructured peering policy documents. This extracted information is then combined with AS profile features from public peering datasets to develop a scoring model. By doing so, we build upon previous work while addressing key limitations of existing approaches, particularly their failure to adequately account for the requirements outlined in AS peering policies.

The proposed framework is designed to function in a copilot capacity, providing explainable automated recommendations while leaving the final decision to human actors. This ensures that nuanced factors and business considerations not captured by our model can be incorporated into the final peering decision, facilitating more efficient and mutually beneficial peering arrangements in the Internet ecosystem. The key contribution of our research is a policy-aware approach that computes compatibility scores for potential AS peering pairs based on mutual requirement satisfaction.

The paper is structured as follows: Section II covers related work, Section III describes our data, Section IV details our selection model, and Section V outlines our evaluation methodology. Results are presented in Section VI, followed by discussion in Section VII. Section VIII concludes the paper.

II. Related Work

Previous research on AS peering partner selection has approached the problem from various angles. Dhamdhere [3] investigated provider and peer selection strategies in an evolving Internet ecosystem, focusing on the economic implications of different selection approaches. While informative, this work did not aim to recommend specific peering partners.

Economic and game-theoretic aspects of peering have been explored in several studies [4], [5], [6], [7], [8], [9], [10], [11]. However, these models either focused on fair pricing or simplified network scenarios, overlooking crucial real-world characteristics such as skewed traffic matrices, geographic colocation constraints, and ratio-based peering policies.

Dey et al. [12] introduced Meta-peering, an automated peer selection framework using three key metrics: Willingness Score, Affinity Score, and Felicity Score. Their model achieved 85% accuracy in identifying existing peering relationships, demonstrating its potential as a tool for network administrators.

Building on this, Ibne Alam et al. [13] incorporated internal routing costs into the peer selection model. They introduced the Peering Stability (PS) metric to assess the stability of ISP peering relationships, estimating a price of stability and showing that stable peering points exist between ISP pairs.

Most recently, Mustafa et al. [14] developed a machine learning model to identify potential peering relationships between ASes. Their approach utilized data from PeeringDB and CAIDA ASRank, determining key factors influencing peering decisions. While they analyzed peering policy documents from US-based ISPs, this information was not directly incorporated into their ML model.

Our work advances these prior studies by explicitly incorporating ASes' preferences, requirements, and restrictions into the decision model. We use natural language processing techniques to extract relevant information from unstructured peering policy documents, enabling us to compute a mutual compatibility score between ASes based on their policies and attributes. This approach results in a more applicable, accurate, and explainable model that provides clear justification for peering decisions and offers 'diagnostic' predictions to guide AS operators in identifying suitable peering partners.

III. Description of Data

A. PeeringDB

PeeringDB [15] is a free, open-source database where ASes list their profiles, brief peering requirements, and contact information. We used a CAIDA-maintained snapshot of PeeringDB from May 30, 2024 [16]. Our proposed scores use data from three tables: *net* for AS information, *ix* for internet exchange point details, and *netixlan* for information about AS presence at IXPs.

B. Peering Policy Documents

ASes provide detailed peering policy documents outlining their requirements, preferences, and restrictions. These policies typically cover several key areas including operational, traffic volume, geographical coverage requirements and routing policies. General and operational requirements set foundational expectations for potential peers, including maintaining a 24/7 Network Operations Center (NOC), updating PeeringDB records, implementing Mutually Agreed Norms for Security (MANRS), and utilizing Internet Routing Registry (IRR) for routing management.

Traffic considerations are crucial, with ASes specifying minimum peak traffic volumes for different peering arrangements to justify costs. Some also require specific ratios of outbound to inbound traffic for fair relationships. Geographic coverage needs are often addressed, with ASes requiring peers to connect at multiple locations to distribute traffic loads evenly.

Routing policies detail traffic management specifications, including accepted route types and preferred routing strategies. ASes with selective or restrictive policies may include specific limitations, such as restrictions on peering with certain customers or minimum transit customer requirements.

We used NLP techniques to extract key information from these policy documents, including traffic volume requirements for IXP and PNI peering, minimum required peering locations, and traffic ratios for a balanced exchange.

IV. Peer Selection Model

We conceptualize the peering decision as an interactive process optimized from the deciding autonomous system's perspective. This model considers both the requirements of the deciding AS and the mutual compatibility between potential peering partners. We propose that when two ASes mutually meet each other's stated requirements, it likely results in stable peering relationships between them.

To model this concept, we identify seven key criteria from peering policy documents and PeeringDB and calculate three scores: two measuring individual AS satisfaction and one quantifying mutual satisfaction.

We use the following general approach to compute these scores. For a given peering requirement i, an AS K specifies whether this requirement must be fulfilled and the minimum fulfilment amount. K's satisfaction score, a value between 0 and 1, is computed as a relative measure of what it wants compared to what the potential peer P is offering.

Let $k_{sat}(i)$ be K's satisfaction score for requirement i, $p_{offer}(i)$ be the value P can offer based on its policy or attributes, and $k_{wants}(i)$ be the minimum value K requires. Then:

$$k_{sat}(i) = \begin{cases} \min\left(\frac{p_{offer}(i)}{k_{wants}(i)}, 1\right) & \text{if } i \text{ is required} \\ 1 & \text{otherwise} \end{cases} \quad (1)$$

This equation captures how satisfied AS K is with what P offers. Much like a percentage: if P offers more than or exactly what K wants, K is 100% satisfied (score = 1). If P offers less than K wants, K's satisfaction is proportional to how much P offers compared to what K wants. If AS *K* does not require *i* to be fulfilled, then it will be satisfied with whatever P is offering. This formulation ensures that more open ASes - which have fewer requirements, will also accounted for. The satisfaction score for P, $p_{sat}(i)$, is computed similarly.

The mutual satisfaction score for requirement i is:

$$mut_{sat}(i) = k_{sat}(i) \times p_{sat}(i) \quad (2)$$

This equation measures how well both ASes' requirements are met simultaneously. By multiplying their individual satisfaction scores, we ensure that if either AS is unsatisfied (score close to 0), the mutual satisfaction will also be low. Only when both ASes are highly satisfied will the mutual satisfaction be high. For instance, if K is 90% satisfied and P is 80% satisfied, their mutual satisfaction would be 0.72 or 72%.

We also compute a satisfaction disparity score:

$$disp(i) = \frac{|k_{sat}(i) - p_{sat}(i)|}{k_{sat}(i) + p_{sat}(i)} \quad (3)$$

Equation 3 measures how unequal the satisfaction levels are between the two ASes. The numerator captures the absolute difference in satisfaction, while the denominator normalizes the score.

These scores quantify the compatibility between potential peering partners for each requirement. While ASes share similarities, we recognize that each makes peering decisions based on its unique strategy and business goals. To account for these differences, we employ a machine learning approach to learn the weights each AS assigns to different scores. This allows us to predict individualised peering decisions and provide explanations for those decisions.

The subsequent subsections provide detailed descriptions of the proposed scores.

TABLE I: Traffic Profiles and Ratios

Profile	Ratio (Out:In)	Traffic Skew (Out/In)
Heavy Inbound	1:3	0.3
Mostly Inbound	1:2	0.5
Balanced	1:1	1.0
Mostly Outbound	2:1	2.0
Heavy Outbound	3:1	3.0

A. Geographical Coverage Score

This score quantifies how well two ASes meet each other's multi-location peering requirements. Each AS specifies whether peering at multiple locations is required and the percentage of locations a peering partner should cover.

We compute this score using peering locations from PeeringDB, assigning numerical identifiers to all possible peering locations. Each AS's presence is represented by a multi-hot binary vector. The co-location of two ASes is computed as the dot product of their presence vectors.

Formally, for an AS K with a presence vector X_p and a geographical coverage requirement K_{geo} (as a percentage), and a potential peer P with a presence vector P_p and a geographical coverage requirement P_{geo}, we compute P's coverage of K's locations as:

$$P_{cover\_K} = \frac{X_p \cdot P_p}{X_p \cdot X_p} \tag{4}$$

Where the numerator represents shared locations and the denominator represents K's total locations.

The geographic coverage satisfaction score for K is then computed as:

$$k\_sat\_geo = \begin{cases} \min(P_{cover\_K}/K_{geo}, 1) & \text{if required} \\ 1 & \text{otherwise} \end{cases} \tag{5}$$

P's satisfaction score is computed similarly. We calculate the mutual satisfaction and disparity scores using equations 2 and 3 respectively.

B. Traffic Ratio Score

Some ASes require balanced traffic exchange, specifying ratios to indicate acceptable imbalances. For example, a 2:1 ratio means an AS will accept up to twice as much traffic as it sends, while a 1:1 ratio indicates equal exchange. ASes not considering traffic ratios state this in their policy and omit ratio specifications.

To calculate the traffic ratio, we analyse each AS's traffic profile and ratio requirements. Table I shows the encoding of self-reported traffic profiles from PeeringDB.

The Traffic Skew Metric represents the relative outbound traffic for each profile.

For an AS K with traffic skew K_s, traffic ratio requirement *Out:In*, and potential peer P with traffic skew P_s, the maximum imbalance K allows is:

$$K_{rmax} = K_s \times Out \tag{6}$$

The satisfaction score of the traffic ratio requirement r for K is computed as:

$$k\_sat\_ratio = \begin{cases} 1 & \text{if not } r \text{ or } P_s \leq K_s \\ 1 - \frac{P_s - K_s}{K_{rmax}} & \text{if } K_s < P_s \leq K_{rmax} \\ -\min\left(\frac{P_s - K_{rmax}}{K_{rmax}}, 1\right) & \text{if } P_s > K_{rmax} \end{cases} \tag{7}$$

This equation yields a satisfaction score of 1 when K has no ratio requirement or when P's traffic skew is below K's own skew ($P_s \leq K_s$). For cases where P's skew exceeds K's but remains within the maximum acceptable ratio ($K_s < P_s \leq K_{rmax}$), the score decreases linearly from 1 to 0. When P's skew exceeds K's maximum acceptable ratio ($P_s > K_{rmax}$), the score becomes negative, with penalties increasing up to -1 for larger exceedances.

The satisfaction score for P is computed similarly. The mutual satisfaction and disparity scores are then calculated using equations 2 and 3, respectively.

C. Traffic Volume Score

ASes typically specify minimum peak volume requirements to justify peering costs. This metric quantifies how well two ASes meet each other's traffic volume requirements. While ASes use precise traffic measurements for decision-making, researchers must estimate these volumes using publicly available data.

We use stipulated minimum requirements from policy documents to compute satisfaction scores. For ASes without specified volume requirements, we interpolate values based on ASes of the same type and traffic levels. Traffic levels are self-reported in PeeringDB using 18 interval categories (e.g., 0-20 Mbps, 20-100 Mbps, ..., 100+ Tbps).

For an AS K with traffic volume requirements K_{vol} and a potential peer P with requirements P_{vol}, K's satisfaction score is computed as:

$$k\_sat\_vol = \min(P_{vol}/K_{vol}, 1) \tag{8}$$

P's satisfaction score is computed similarly. The mutual satisfaction and disparity scores are calculated using equations 2 and 3 respectively.

We acknowledge that this method does not use true measured traffic volumes. However, we assume that ASes typically initiate peering requests after observing exchanged traffic volumes close to their policy requirements. The potential partner then evaluates this request against its own minimum traffic requirements. Thus, our estimate serves as a reasonable proxy for the decision-making process.

D. General Requirements Score

This metric assesses how well two ASes meet each other's general requirements. We evaluate four typical requirements: updated PeeringDB record, 24x7 NOC, use of IRR for routing rules, and implementation of MANRS.

TABLE II: Summary of Sampled Autonomous Systems

Type	Count	Scope	Policies	Multi-Geo	Ratios
access	5	Africa (2), Europe (1), South America (2)	open (2), selective (2), restrictive (1)	No (4), Yes (1)	No (5)
content	7	global (5), Asia Pacific (1), Africa (1)	selective (4), open (3)	No (4), Yes (3)	No (7)
transit	8	global (6), Europe (1), Africa (1)	restrictive (4), selective (3), open (1)	Yes (8)	No (6), Yes (2)

We represent these requirements as a multihot binary vector, where 1 indicates a requirement and 0 indicates its absence. Similarly, an AS's features are represented by a multihot vector F_X, where 1 denotes the presence of a feature for AS X.

For an AS K considering a potential partner P, K's general requirements satisfaction is computed as:

$$k\_gen\_sat = \frac{R_K \cdot F_P}{R_K \cdot R_K} \tag{9}$$

Where R_K is K's requirements vector and F_P is P's features vector. The satisfaction score for P, *p_gen_sat* is computed in the same way. The disparity and mutual satisfaction scores are computed using equations 2 and 3 respectively.

E. Maximum Advertised IP Prefixes Scores

ASes specify the maximum IPv4 and IPv6 prefixes to be advertised after peering in PeeringDB. These prefixes represent networks under an AS, and their maximum allowed exchange provides insight into the AS's size and capacity. This score captures the difference in size and capacity, hypothesizing that similarly sized ASes make better peers. Additionally, ASes wanting to advertise as many prefixes as possible will be better paired with those having similar policies.

For an AS K with maximum allowed prefixes K_{pmax} and a potential peering partner P with P_{pmax}, we compute the satisfaction scores for IPv4 and IPv6 maximum prefixes requirement as:

$$k\_sat\_ipvX = \min(P_{pmax}/K_{pmax}, 1) \tag{10}$$

Where X is the IP version number (4 or 6). We compute P's satisfaction score similarly, along with their mutual and disparity scores.

F. Port Capacity Score

This score compares the port capacity of potential peering partners at shared locations. PeeringDB records each AS's port capacity at every facility where it has a presence. ASes prefer congestion-free peering links for efficient traffic exchange, making them better suited to partners with similar port sizes to prevent bottlenecks. We use the reported port sizes to compute individual, mutual, and disparity scores for an AS and its potential peering partner.

V. Methodology

We assess the proposed scoring model's predictive capability using a machine learning approach, with CAIDA's AS relationships dataset [17] as ground truth. This dataset, derived from BGP table snapshots, contains AS peering relationships inferred by a statistical algorithm.

A. Hypothesis and Model Class Rationale

We hypothesize that our developed scores capture AS peering decision processes and can predict peering relationships. To test this, we employ a binary classification approach using a Random Forest classifier. This method allows us to learn AS selection function weights, assess score predictive power, and capture non-linear relationships while resisting overfitting. It also provides feature-importance insights, helping evaluate the relative impact of different factors on AS peering decisions.

B. Sample Size and Data Preparation

Our study focused on a diverse set of 20 ASes, purposefully selected to represent various types, sizes, operational scopes, and regions of registration. For each of these target ASes, we identified potential peering partners using PeeringDB data. We considered an AS a potential peer if it shared at least one peering location with the target AS. We then computed mutual satisfaction scores for each potential peering pair using the method described in the previous section. These mutual scores served as direct input features for our classification task, without additional processing.

To capture the unique peering preferences of each AS, we trained a separate classifier for each target AS. This approach allows us to represent an expressive peer selection function tailored to each target AS, accounting for potential variations in peering strategies across different ASes. Table II summarizes the characteristics of the 20 selected ASes.

C. Classification Model Configuration and Evaluation

Our evaluation strategy is designed to provide a robust assessment of the model's performance.

Train-Test Split: We use a 50-50 split to create the training and testing datasets. This conservative split ensures that we have a large portion of unseen data for final evaluation, providing a stringent test of our model's generalisation capabilities. **Performance Metrics**: We evaluate the model using accuracy, precision, recall, F1-score and ROC-AUC metrics. Accuracy measures the overall correctness of predictions, while precision and recall help us understand the balance between false positives and false negatives. The F1-score provides a single metric balancing precision and recall, and ROC-AUC assesses the model's ability to distinguish between classes. We also conduct a feature importance analysis to understand which factors influence peering relationship predictions.

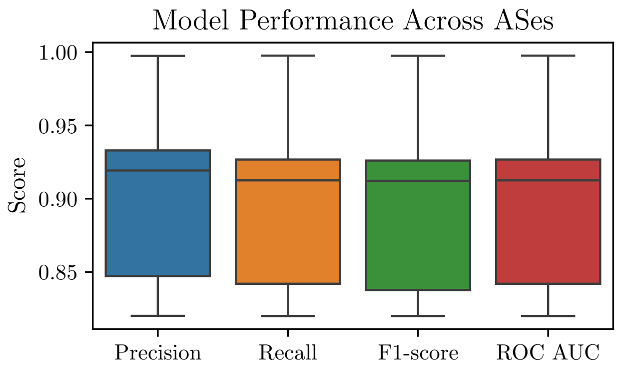

Fig. 1: Model Performance

VI. Empirical Results

A. Model Performance

Figure 1 summarises model performance across the 20 ASes. The models show strong performance with median values above 0.90 for all metrics. Precision has the highest median and smallest spread, while recall shows slightly more variability. F1-score and ROC AUC remain high, reflecting overall effectiveness and discrimination ability. The lower whiskers are above 0.80 for all ASes, indicating robust performance even in challenging cases. This consistent high performance suggests the proposed approach effectively captures key factors influencing peering decisions across diverse network operators.

B. Feature Importance Analysis

Figure 2 shows the important features for eight diverse ASes. The feature importance analysis reveals significant variability in peering decision factors, highlighting the individualized nature of these relationships. Technical factors, particularly traffic volume, port size, and maximum prefix exchange, consistently emerge as the most important features across most ASes. Traffic volume stands out as a top factor for six out of eight ASes, underscoring its critical role in peering decisions. Geographic and general requirements generally show lower importance, though their significance varies among ASes. Traffic ratio is typically the least important feature, often having minimal impact even for ASes such as GTT Communications and PCCW Global that require it. Content providers like Netflix and Meta tend to prioritize port size and traffic volume more consistently than transit providers such as GTT Communications, which display more varied patterns.

VII. Discussion

Our primary objective was to develop and validate quantitative metrics that could effectively capture the multifaceted nature of peering relationships based on the mutual satisfaction of policies. The results strongly support the efficacy of this approach. The model demonstrated robust predictive performance across diverse ASes, indicating that our proposed scores effectively capture key peering decision factors.

The peer selection model offers immediate practical value for peering coordinators in both reactive and proactive scenarios. When evaluating incoming peering requests, coordinators can leverage our quantitative metrics to rapidly assess compatibility between their AS and potential partners using concrete policy requirements and network attributes. Additionally, the model enables coordinators to proactively identify promising potential partners by analysing compatibility scores before initiating contact, increasing the likelihood of successful peering negotiations while avoiding partnerships unlikely to succeed. Our ongoing research focuses on extending this work in two critical directions: developing a comprehensive peering decision optimization framework and designing a peer selection copilot prototype for real-world implementation. These developments aim to bridge the gap between theoretical modelling and operational deployment, providing network operators with practical tools to enhance their peering strategies while maintaining the transparency and efficiency offered by our explainable metrics.

A. Study Limitations & Future Work

Our study has two main limitations. First, we validated our model using CAIDA's AS relations data, which is generated by an inference algorithm. While this algorithm is highly accurate, only 34% of its outputs were validated by network operators. Despite this drawback, CAIDA's dataset remains the most comprehensive publicly available resource and is widely used in research. Second, our model only considers IXP peering, not PNI, as we rely on PeeringDB for location data, which lacks most private PoP information. Future work could explore alternative evaluation methods and incorporate router-level AS topology to include PNI. We are working to enhance our model by including additional metrics that capture economic, security, and qualitative factors influencing peering decisions, which will increase the model's richness and expressiveness.

VIII. Conclusion

This study introduces a quantitative, data-driven method for identifying and assessing potential peering partners in the Internet ecosystem. We have successfully demonstrated the effectiveness of the approach in predicting peering relationships using machine learning techniques.

Ethical Consideration

This work does not raise any ethical issues.

Acknowledgment

This work was financially supported by the Internet Society (ISOC) and AFRINIC, through the Computer Science Department at the University of Cape Town.

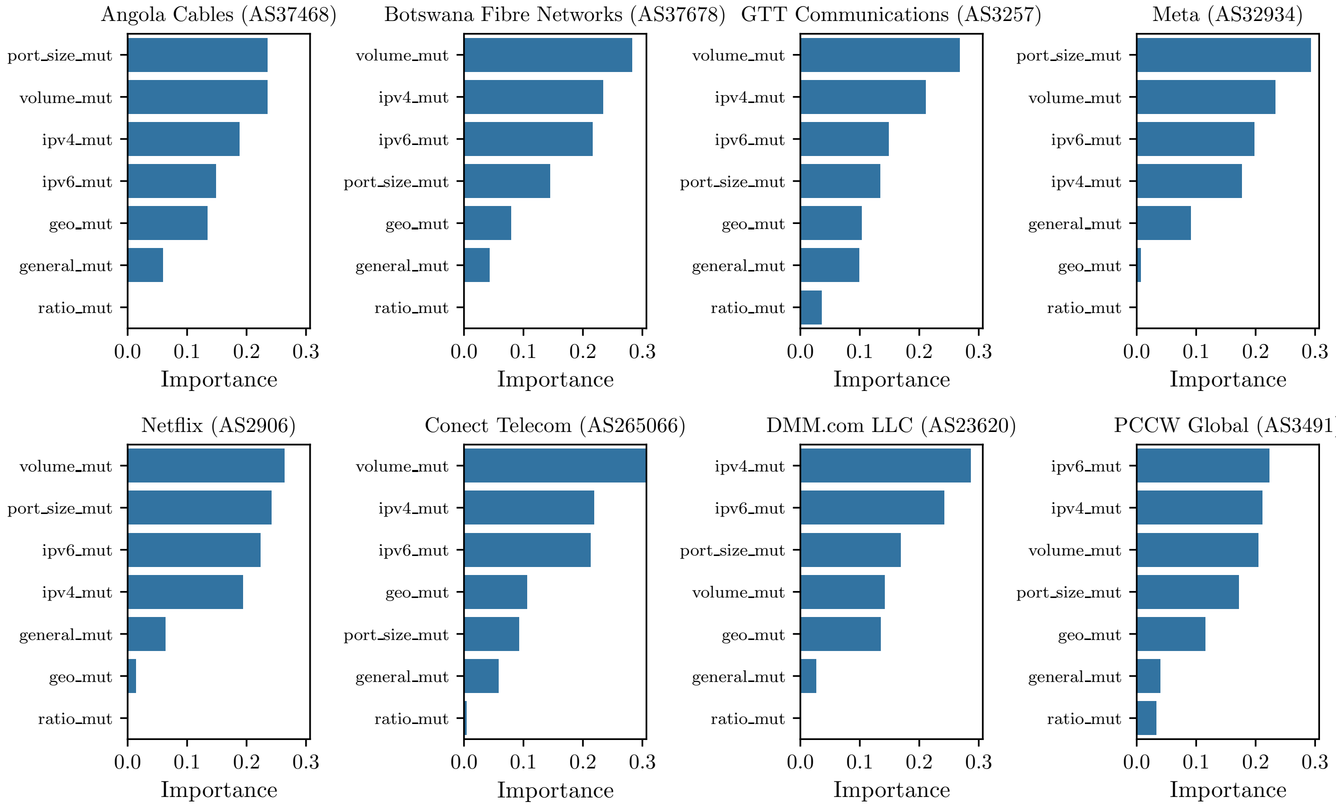

Fig. 2: Feature Importance

References

[1] William Norton. *The Internet Peering Playbook: Connecting to the Core of the Internet*. DrPeering Press, 2014 edition edition, 2014.

[2] Aemen Lodhi, Nikolaos Laoutaris, Amogh Dhamdhere, and Constantine Dovrolis. Complexities in Internet Peering: Understanding the "Black" in the "Black Art". In *IEEE INFOCOM (2015) 26 1778-1786*, pages 1778–1786, 2015.

[3] Amogh Dhamdhere. *Provider and peer selection in the evolving Internet ecosystem*. PhD thesis, Georgia Institute of Technology, 2009.

[4] Richard T.B. Ma, Dah Ming Chiu, John C.S. Lui, Vishal Misra, and Dan Rubenstein. Interconnecting eyeballs to content: A shapley value perspective on isp peering and settlement. In *SIGCOMM 2008 Conference and the Co-located Workshops - Proceedings of the 3rd International Workshop on Economics of Networked Systems, NetEcon'08*, pages 61–66, 2008.

[5] Eli A. Meirom, Shie Mannor, and Ariel Orda. Network formation games with heterogeneous players and the internet structure. In *EC 2014 - Proceedings of the 15th ACM Conference on Economics and Computation*, pages 735–752, 2014.

[6] Fehmina Malik, Manjesh K. Hanawal, and Yezekael Hayel. Profit sharing contracts between content and service providers for enhanced network quality. *IEEE Transactions on Network and Service Management*, 19(3):3533–3546, 2022.

[7] Elliot Anshelevich, F. B. Shepherd, and Gordon Wilfong. Strategic network formation through peering and service agreements. *Games and Economic Behavior*, 73(1):17–38, 2011.

[8] Doron Zarchy, Amogh Dhamdhere, Constantine Dovrolis, and Michael Schapira. Nash-peering: A new techno-economic framework for internet interconnections. In *INFOCOM 2018 - IEEE Conference on Computer Communications Workshops*, pages 403–408. Institute of Electrical and Electronics Engineers Inc., jul 2018.

[9] Narine Badasyan and Subhadip Chakrabarti. A simple game-theoretic analysis of peering and transit contracting among Internet service providers. *Telecommunications Policy*, 32(1):4–18, feb 2008.

[10] Xin Wang and Richard T.B. Ma. On Private Peering Agreements between Content and Access Providers: A Contractual Equilibrium Analysis. In *Performance Evaluation Review*, volume 49, pages 57–58. Association for Computing Machinery, jun 2021.

[11] Amogh Dhamdhere, Constantine Dovrolis, and Pierre Francois. A Value-based Framework for Internet Peering Agreements. In *22nd International Teletraffic Congress (ITC 22)*, pages 1–8, Amsterdam, 2010. IEEE.

[12] Prasun Kanti Dey, Shahzeb Mustafa, and Murat Yuksel. Meta-peering: Towards automated ISP peer selection. In *ANRW 2021 - Proceedings of the 2021 Applied Networking Research Workshop*, pages 8–14, New York, NY, USA, 2021.

[13] Md Ibrahim Ibne Alam, Shahzeb Mustafa, Koushik Kar, and Murat Yuksel. Modeling and Automating ISP Peering Decision Process: Willingness and Stability. In *IEEE International Conference on Communications*, volume 2022-May, pages 371–376, 2022.

[14] Shahzeb Mustafa, Prasun Kanti Dey, and Murat Yuksel. Peer Me Maybe?: A Data-Centric Approach to ISP Peer Selection. *Proceedings of the IEEE/IFIP Network Operations and Management Symposium 2022: Network and Service Management in the Era of Cloudification, Softwarization and Artificial Intelligence, NOMS 2022*, 2022.

[15] PeeringDB. Peeringdb. https://peeringdb.com/. (Accessed on 05/13/2024).

[16] CAIDA. The caida ucsd peeringdb dataset, 2024. Accessed on: May 30, 2024.

[17] CAIDA. Caida as relationships dataset. https://www.caida.org/catalog/datasets/as-relationships/, 2024. Accessed on: August 28, 2024.