Inferring Multiple Relationships Between ASes using GCN

Songtao Peng, Xincheng Shu, Zhongyuan Ruan, Zegang Huang, and Qi Xuan

Institute of Cyberspace Security, Zhejiang University of Technology

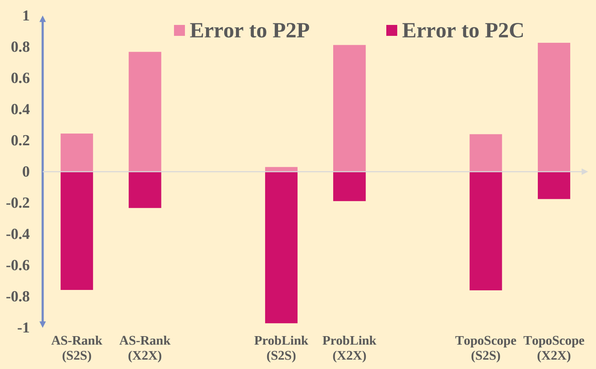
网络空间安全研究院 INSTITUTE OF CYBERSPACE SECURITY

BACKGROUND

- As a typical complex network, the Internet is now composed of more than 70000 Autonomous Systems (ASes).
- Routing information is then exchanged between ASes through Border Gateway Protocol (BGP) to achieve global reachability. Meanwhile, routing information can be effectively used to construct the topological graph of the AS interconnection and thus can implement strategic decisions at the AS-level.
- In AS-level topology, the business relationships between connected ASes are broadly classified into (1) Customer-Provider (C2P), (2) Peer-Peer (P2P), and (3) Sibling relationship (S2S).
- A series of studies on these relationships include topology flattening, Internet security check, network congestion detection, etc.

PURPOSE

Precisely understanding the business relationships between ASes is essential for studying the Internet structure. So far, many inference algorithms have been proposed to classify the AS relationships, which mainly focus on P2P and P2C binary classification and have achieved excellent results. However, there are other types of AS relationships in actual scenarios, i.e., the business-based sibling (S2S) and structure-based exchange (X2X) relationships, which were neglected in the previous studies. We use strong correlation features and advanced methods to ensure the accuracy of binary classification and achieve multiple relational classification.



The default classification of S2S and X2X in binary classification without a priori knowledge.

CONTRIBUTION

In particular, the main contributions of this paper are summarized as follows:

(1) We focus on the multi-classification of the multiple relationships between ASes for the first time. Meanwhile, we aggregate multiple superior algorithms based on the idea of hard voting to obtain the more comprehensive dataset, which well solves the deviation and limitation of the training data of the previous research just from a single platform (**Part. 1**).

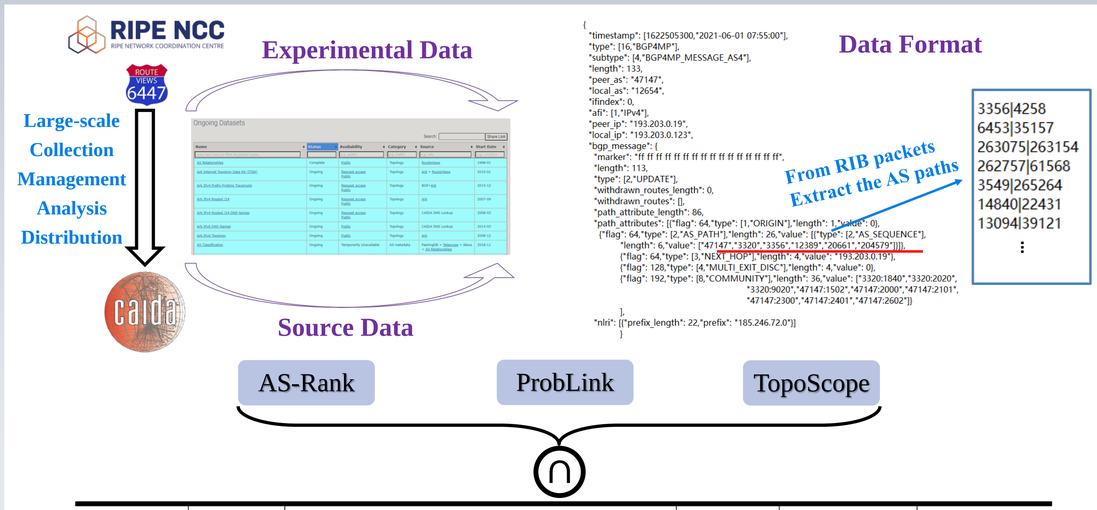
(2) By synthesizing and analyzing the important features, we found the similarities and differences between P2P, P2C, S2S and X2X (**Part. 2**).

(3) We developed a new framework AS-GCN to solve the multiple relationships inference problem under complex scene. As far as we know, this is the first trial to use GCN (**Part. 3**).

(4) Comprehensive experiments show the outstanding performance of our AS-GCN, by comparing with a series of baselines, especially on the more challenging multi-classification task.(**Part. 4**).

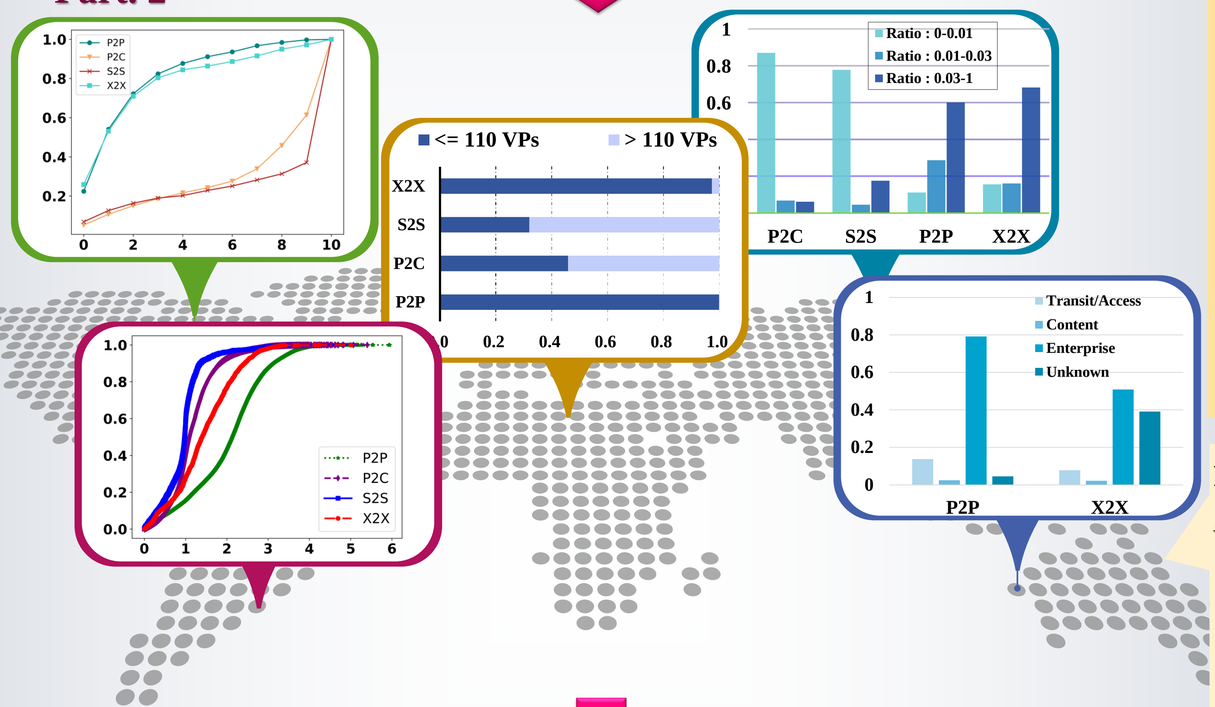
MATERIALS AND METHODS

Part. 1

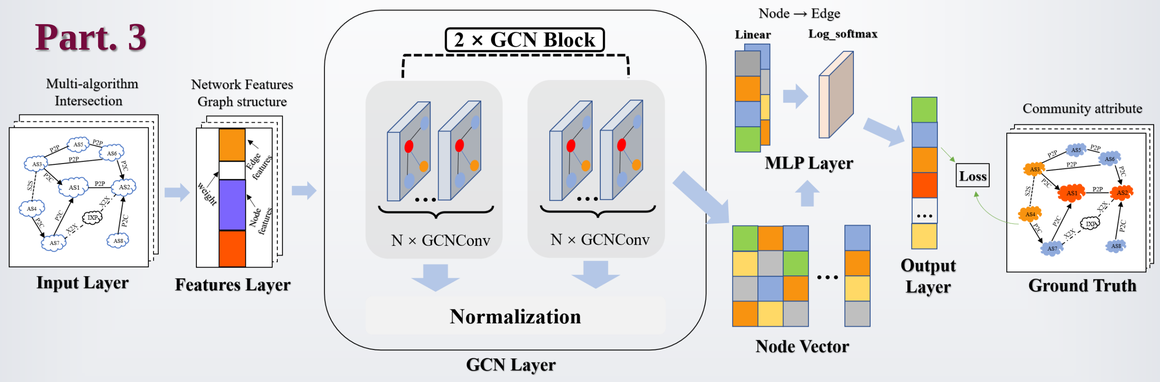


Date	Total links	Real Situation			Intersection	Coincidence rate	Intersection Acc
		AS-Rank	ProbLink	TopoScope			
2016	265633	254297	254466	258538	221272	83.30%	98.13%
2017	255694	246117	249752	249755	214465	83.88%	97.06%
2018	263219	254554	254700	255352	214281	81.97%	99.26%

Part. 2



Part. 3



Part. 4

Table 1: Experimental *Accuracy* of Binary Classification

Date	Traditional inference methods			Feature-based methods			AS-GCN
	AS-Rank	ProbLink	TopoScope	RF	Xgboost	LightGBM	
2016	96.75	96.80	96.44	87.58	94.85	91.18	**97.51**
2017	94.19	95.08	93.77	95.82	95.49	96.44	**97.51**
2018	**97.37**	96.55	96.21	84.71	93.37	83.25	97.20

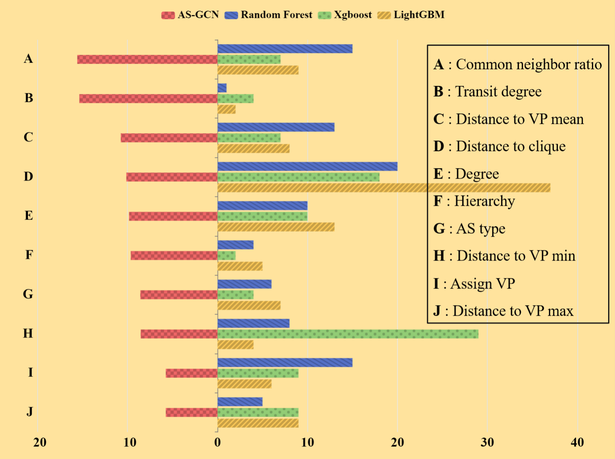
Table 2: Experimental *Accuracy, Precision, Recall* of Multi-Classification

Date	Links	Methods	Acc. (all)	Precision				Recall			
				P2P	P2C	S2S	X2X	P2P	P2C	S2S	X2X
2016	11093	RF	94.71	94.37	**99.79**	88.99	**95.19**	97.20	97.20	**94.29**	90.49
		Xgboost	94.66	**94.87**	98.99	89.34	94.86	96.20	97.80	93.81	91.04
		LightGBM	94.26	94.43	98.58	91.63	92.24	95.00	97.00	91.19	90.49
		AS-GCN	**95.32**	94.37	99.60	**92.36**	94.60	**97.20**	**98.80**	92.14	**92.87**
2017	10962	RF	92.62	90.17	99.17	86.68	94.66	97.20	95.20	92.76	85.85
		Xgboost	93.28	92.69	98.37	87.21	**94.72**	96.40	96.40	**94.12**	86.78
		LightGBM	93.33	92.34	98.39	87.83	94.41	96.40	97.60	91.40	88.08
		AS-GCN	**95.54**	**95.85**	**99.40**	**93.76**	93.28	**97.20**	**99.20**	91.03	**94.42**
2018	13227	RF	92.98	91.80	93.96	91.23	94.08	94.00	96.40	88.30	93.26
		Xgboost	93.96	93.69	95.45	91.35	**94.77**	95.00	96.40	89.62	94.57
		LightGBM	92.69	91.49	96.78	91.99	91.57	94.60	96.20	84.53	94.46
		AS-GCN	**95.38**	**95.42**	**99.00**	**93.26**	94.60	**95.80**	**99.40**	**92.09**	**94.94**

RESULTS

We have proved the superiority of AS-GCN in the following three aspects:

- Under snapshots at different times, the accuracy of AS-GCN on binary classification can be maintained above **97%,** which is comparable with many superior algorithms, such as AS-Rank, ProbLink and TopoScope;
- The overall accuracy of multi-classification tasks evaluated on balanced datasets can even reach higher than **95%.** Precision and recall also show a general advantage.
- Through the parameter feature importance analysis of the model, the performance of the model can be explained to a certain extent.



Feature importance analysis.

Features: *Distance to Clique, Assign VP, Common Neighbor Ratio, Distance to VP mean, AS Type*.

In a word

S2S is similar to P2C

X2X is similar to P2P

P2C, S2S, and P2P, X2X are significantly defferent

CONCLUSIONS

In response to the challenges faced by multi-classification, we analyzed many features and proposed a GCN framework, AS-GCN, to solve the binary classification and multi-classification problem. The convolution process of AS-GCN on the Internet topology is a good simulation of the valley-free path characteristic, and the framework also takes into account the global network structure and local link features simultaneously. Finally, the neural network achieves good performance on the problem by virtue of its important and comprehensive features and its excellent ability to learn inductively.

REFERENCES

1. Motamedi, R., Rejaie, R., & Willinger, W. (2014). A survey of techniques for internet topology discovery.
2. Gao, L. (2001). On inferring autonomous system relationships in the Internet.
3. Luckie, M., Huffaker, B., Dhamdhere, A., Giotsas, V., & Claffy, K. C. (2013). AS relationships, customer cones, and validation.
4. Giotsas, V., Luckie, M., Huffaker, B., & Claffy, K. C. (2014). Inferring complex AS relationships.
5. Jin, Y., Scott, C., Dhamdhere, A., Giotsas, V., Krishnamurthy, A., & Shenker, S. (2019). Stable and Practical {AS} Relationship Inference with ProbLink.
6. Jin, Z., Shi, X., Yang, Y., Yin, X., Wang, Z., & Wu, J. (2020, October). TopoScope: Recover AS Relationships From Fragmentary Observations.

This research supported by the National Natural Science Foundation of China under Grant 61973273 and by the Zhejiang Provincial Natural Science Foundation of China under Grants LY21F030017 and LR19F030001.