Poster:Inferring Multiple Relationships between ASes using GCN

I. ABSTRACT

Precisely understanding the business relationships between Autonomous Systems (ASes) is essential for studying the Internet structure [1]. So far, many inference algorithms [2], [3], [4], [5], [6] have been proposed to classify the AS relationships, which mainly focus on Peer-Peer (P2P) and Provider-Customer (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 and structure-based exchange relationships, which were neglected in the previous studies. These relationships are usually difficult to be inferred by existing algorithms because there is no discrimination on the designed features compared to the P2P or P2C relationships (Fig. 1).

In our work, we construct a set of experimental datasets containing three types of business relationships (P2C, P2P, S2S) and one structure relationship (X2X), and focus on the multiple relationships between ASes for the first time. First, we summarize the differences between AS relationships under the structural and attribute features, and the reasons why multiple relationships are difficult to be inferred. We then introduce new features and propose a Graph Convolutional Network (GCN) framework, AS-GCN (Fig. 3), 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 [2] characteristic, and the framework also takes into account the global network structure and local link features simultaneously. The experiments on real Internet topological data validate the effectiveness of our method, i.e., AS-GCN achieves excellent results on the easy binary classification task (Table. I), the accuracy rate is generally higher than 97 %, and outperforms a series of baselines on the more difficult multi-classification task (Table. II), with the overall accuracy above 95 %. Finally, we analyze the importance of each feature in the multi-classification problem (Fig. 2), and apply t-SNE dimensionality reduction technology to visualize the results (Fig. 4).

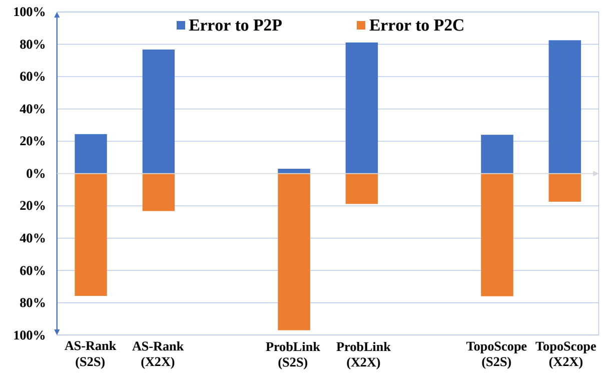

Fig. 1: The fraction of misclassifying S2S and X2X into P2P and P2C of complex relationships by different algorithms.

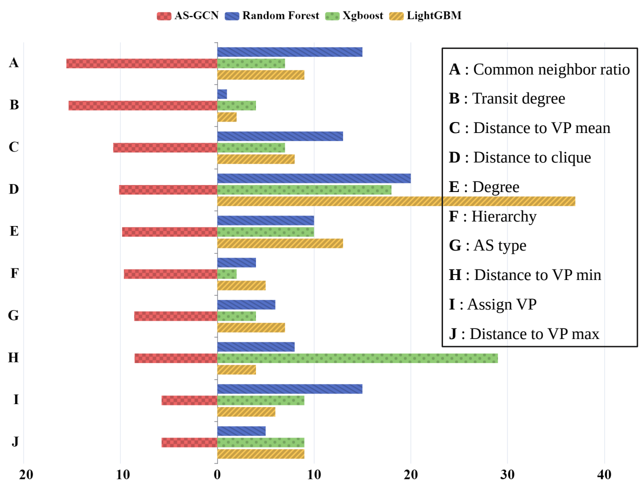

Fig. 2: Feature importance analysis on AS-GCN and its comparison methods (Random Forest, Xgboost, LightGBM). Common neighbor ratio shows its relatively important position.

Referencias

[1] Reza Motamedi, Reza Rejaie, and Walter Willinger, "A survey of techniques for internet topology discovery," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 2, pp. 1044–1065, 2014.

[2] Lixin Gao, "On inferring autonomous system relationships in the internet," *IEEE/ACM Transactions on networking*, vol. 9, no. 6, pp. 733–745, 2001.

[3] Matthew Luckie, Bradley Huffaker, Amogh Dhamdhere, Vasileios Giotsas, and KC Claffy, "As relationships, customer cones, and validation," in *Proceedings of the 2013 conference on Internet measurement conference*, 2013, pp. 243–256.

[4] Vasileios Giotsas, Matthew Luckie, Bradley Huffaker, and KC Claffy, "Inferring complex as relationships," in *Proceedings of the 2014 Conference on Internet Measurement Conference*, 2014, pp. 23–30.

[5] Yuchen Jin, Colin Scott, Amogh Dhamdhere, Vasileios Giotsas, Arvind Krishnamurthy, and Scott Shenker, "Stable and practical {AS} relationship inference with problink," in *16th {USENIX} Symposium on Networked Systems Design and Implementation ({NSDI} 19)*, 2019, pp. 581–598.

[6] Zitong Jin, Xingang Shi, Yan Yang, Xia Yin, Zhiliang Wang, and Jianping Wu, "Toposcope: Recover as relationships from fragmentary observations," in *Proceedings of the ACM Internet Measurement Conference*, 2020, pp. 266–280.

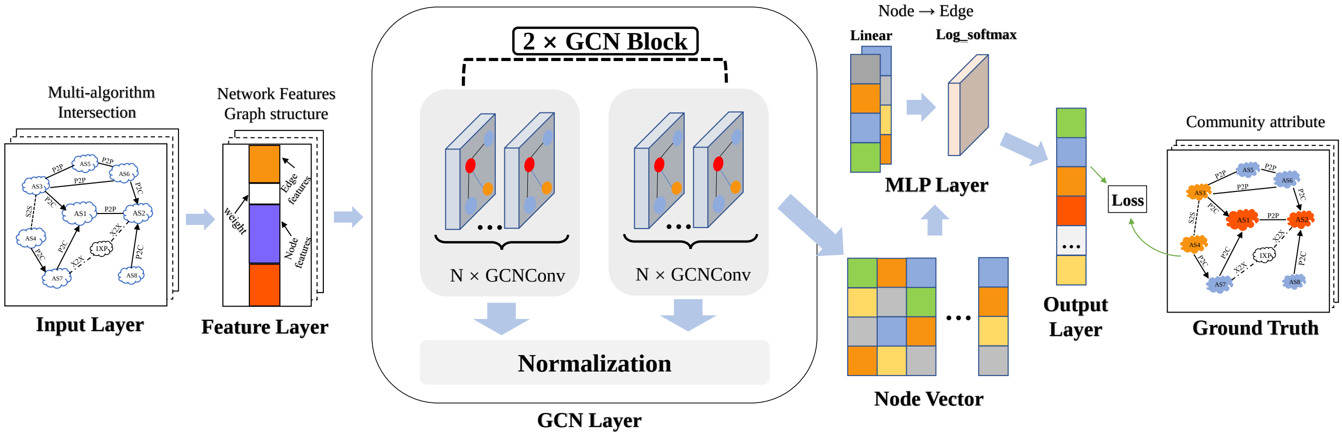

Fig. 3: Model Framework of AS-GCN. AS-GCN mainly conists of five types of layers, i.e., Input layer, Feature layer, GCN layer, MLP layer, and Output layer.

Table I: The *Accuracy* of AS link inference methods for binary classification experiment.

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 II: The *Accuracy*, *Precision* and *Recall* of the three feature-based methods and our AS-GCN model for multi-classification experiment.

Date	Links	Methods	Acc.	Precision				Recall			
			(all)	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**

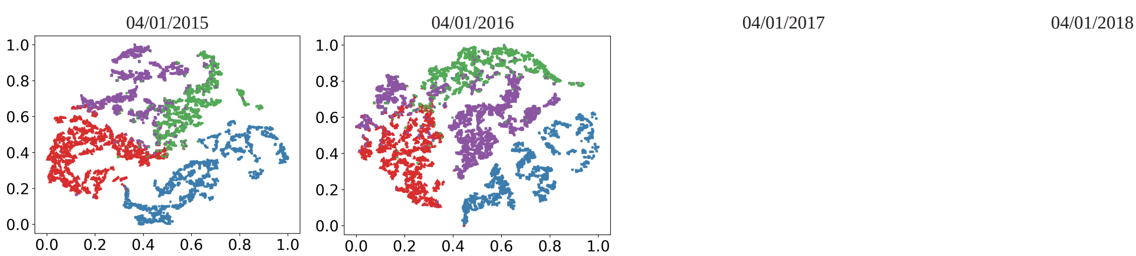

Fig. 4: Visualization of multi-classification obtained by AS-GCN using t-SNE algorithm.