

Songtao PENG, Xincheng SHU, Zhongyuan RUAN, Zegang HUANG, Qi XUAN

# Classifying multiclass relationships between ASes using graph convolutional network

© Higher Education Press 2022

**Abstract** Precisely understanding the business relationships between autonomous systems (ASes) is essential for studying the Internet structure. To date, many inference algorithms, which mainly focus on peer-to-peer (P2P) and provider-to-customer (P2C) binary classification, have been proposed to classify the AS relationships and have achieved excellent results. However, business-based sibling relationships and structure-based exchange relationships have become an increasingly nonnegligible part of the Internet market in recent years. Existing algorithms are often difficult to infer due to the high similarity of these relationships to P2P or P2C relationships. In this study, we focus on multiclassification of AS relationship for the first time. We first summarize the differences between AS relationships under the structural and attribute features, and the reasons why multiclass relationships are difficult to be inferred. We then introduce new features and propose a graph convolutional network (GCN) framework, AS-GCN, to solve this multiclassification problem under complex scenes. The proposed framework considers the global network structure and local link features concurrently. Experiments on real Internet topological data validate the effectiveness of our method, that is, AS-GCN. The proposed method achieves comparable results on the

binary classification task and outperforms a series of baselines on the more difficult multiclassification task, with an overall metrics above 95%.

**Keywords** autonomous system, multiclass relationship, graph convolutional network, classification algorithm, Internet topology

## 1 Introduction

As a typical complex network, the Internet is composed of more than 70000 autonomous systems (ASes). An AS is a network unit that has the freedom to determine its intradomain routing policies. Information is exchanged between ASes through Border Gateway Protocol (BGP) to achieve global reachability and can be used to construct the AS-level topological graph. The business relationships on the edge of the AS-level graph can be classified into three types: 1) customer-to-provider (abbreviated as C2P or P2C depending on the directionality of the relationship), 2) peer-to-peer (P2P), and 3) sibling relationship (S2S). Related studies of AS relationship include topology flattening (Gill et al., 2008; Labovitz et al., 2010), network congestion detection (Sundaresan et al., 2017; Dhamdhere et al., 2018; Smith and Schuchard, 2018), Internet security check (Karlin et al., 2008; Gill et al., 2011; Cohen et al., 2016; Cho et al., 2019), variable routing protocol-based attack designing (Apostolaki et al., 2018), and network characteristic analysis (Tozal, 2018).

Knowledge of the business relationships between ASes is essential to understanding the behaviour of the Internet routing system. However, AS relationships are treated as private information and are unavailable. Since the AS relationship was defined in 2001 (Gao, 2001), many inference algorithms (Dimitropoulos et al., 2007; Gregori et al., 2011; Luckie et al., 2013; Giotsas et al., 2014; Susan Varghese and Ruan, 2015; Jin et al., 2019; 2020; Shapira and Shavitt, 2020) have been proposed, and they

Received December 9, 2021; accepted May 28, 2022

Songtao PENG, Zhongyuan RUAN (✉), Zegang HUANG, Qi XUAN  
Institute of Cyberspace Security, College of Information Engineering,  
Zhejiang University of Technology, Hangzhou 310023, China  
E-mail: zyruan@zjut.edu.cn

Xincheng SHU  
Institute of Cyberspace Security, College of Information Engineering,  
Zhejiang University of Technology, Hangzhou 310023, China; Depart-  
ment of Electrical Engineering, City University of Hong Kong, Hong  
Kong, China

This work was partially supported by the Key R&D Program of Zhejiang (Grant No. 2022C01018), the National Natural Science Foundation of China (Grant Nos. U21B2001 and 61973273), the Zhejiang Provincial Natural Science Foundation of China (Grant Nos. LY21F030017 and LR19F030001), and the Major Key Project of PCL (Grant Nos. PCL2022A03, PCL2021A02, and PCL2021A09).

include the following common characteristics: 1) BGP data are sourced from the platforms that are collected in real-time for network monitoring, troubleshooting, and research; 2) Internet exchange point (IXP, neutral nodes that are used to improve the efficiency of traffic exchange) and S2S relationship (the two connected ASes belong to the same organization) are known information during the classification process; 3) Their focus is shifted from most easily inferred links to a small number of critical links that are difficult to infer. Various advanced algorithms, including AS-Rank (Luckie et al., 2013), ProbLink (Jin et al., 2019), and TopoScope (Jin et al., 2020), have been proposed. These algorithms have a lower prediction error rate (Luckie et al., 2013) and behave surprisingly well in predicting hard links (Jin et al., 2019) or hidden links (Jin et al., 2020). However, the above studies assume that the S2S relationships are known and classify P2P and P2C without considering the other relationships to simplify the network structure.

In recent years, IXP has been an integral part of the Internet's high-level structure. It serves as a centralized exchange platform for establishing connections between different operators, providing reciprocity to the Internet market and contributing to the flattening of the Internet (Castro et al., 2014). In this study, we focus on these relationships that are few but not negligible on the Internet to approach the authentic situation. On the basis of the exchange of structure, we define the relationship between

IXP and AS as X2X. We analyze the homogeneity and heterogeneity of the four relationships (i.e., P2C, P2P, S2S, and X2X, as shown in Fig. 1) using indicators from advanced methods, such as triplet and distance to clique (Luckie et al., 2013), and find that the P2C and S2S relationships have strong similarities that make it difficult to determine whether the ASes belong to the same organization. The features of the X2X relationship are similar to those of P2P. In other words, distinguishing S2S from P2C and X2X from P2P is challenging for the existing classification framework.

To solve this multiclassification problem, we introduce two new features, common neighbor ratio and AS type, and develop a new graph convolution network (GCN) framework called AS-GCN. The convolution process of AS-GCN on the Internet topology is a good simulation of the valley-free path characteristic (Gao, 2001). Therefore, AS-GCN can highly generalize the relevant feature information from the training set to the testing set. In particular, the main contributions of this study are summarized as follows.

(1) We focus on the multiclassification of the AS relationship for the first time and construct an experimental dataset containing four types. We aggregate the results of multiple superior algorithms based on the hard vote to obtain a more comprehensive training set, thereby avoiding the deviation and limitations of the training set of the previous research from a single platform.

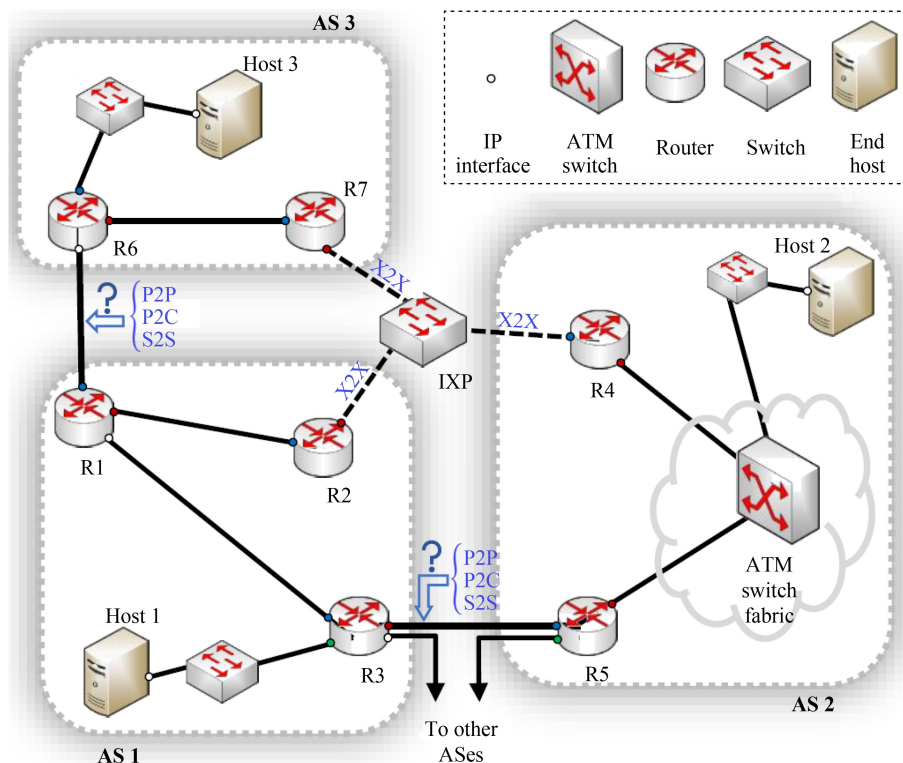


Fig. 1 Brief description of the Internet topology and AS relationships.

(2) We introduce two new features, common neighbor ratio and AS type, and develop a new GCN framework AS-GCN to solve this multiclassification problem under complex scenes. This framework considers global network structures and local link features. To the best of our knowledge, this study is the first to use GCN for addressing this issue.

(3) Comprehensive experiments show the outstanding performance of our framework by comparing it with a series of baselines, especially at the more challenging multiclassification task. Such results indicate that AS-GCN can be a better choice in classifying diverse relationships of ASes.

The rest of this paper is organized as follows. Section 2 reviews the related work and describes the inference algorithms proposed in recent years. Section 3 explains the source and processing of the data. Section 4 presents the challenges of the current research and introduces the basic ideas combined with the GCN model. Section 5 discusses the design and implementation of our framework AS-GCN. Section 6 conducts extensive experiments and hyperparameter analysis. Section 7 provides the conclusion.

## 2 Related work

In this part, we will briefly review the background and the related work on AS relationships, and the inference techniques.

### 2.1 AS relationship

As officially described, an AS is usually defined as a collection of Internet Protocol prefixes under the control of one or more network operators, and the global reachability of information between ASes is achieved through BGP. Route Views and RIPE Network Coordination Centre (NCC) collect BGP routing information through route collectors (vantage points (VPs)) (Orsini et al., 2016) located around the world, which are the main data source used to map the AS-level Internet topology.

The Internet topology at the AS-level is typically modeled by using a graph where each node is an AS and each link represents a business relationship between two ASes. These relationships reflect who pays who when traffic is exchanged between ASes and are key to the proper functioning of Internet systems. AS relationships have typically been categorized into three types: C2P, P2P, and S2S (Motamedi et al., 2015). Customers in a C2P relationship pay the provider to access the Internet, and the other two types, P2P and S2S, are in general settlement-free, indicating that each party of the P2P and S2S relationships does not exchange money and prefers a win-win partnership.

Böttger et al. (2018) found a steady increase in IXP deployments across continents over the last decade. IXPs have facilitated the bypassing of tier-1 Internet service providers, allowing the corresponding organizations to take over the central role of tier-1. The public P2P network ecosystem in Brazil facilitates the formation of a nonprofit business environment for multilateral agreements under the Internet market by maintaining IXPs (Brito et al., 2016). Numerous studies (Castro et al., 2014; Böttger et al., 2018) have demonstrated that IXPs contribute to the flattening of the Internet. In previous years, IXPs were ignored because the Internet was in a developmental stage and was more concerned with the interconnectivity of the network and global reachability. With the competitive nature of the communication companies and the Internet market and the reciprocity offered by IXPs, the importance of IXPs in the Internet network is increasing annually. IXP is no longer a point that can simply be removed.

Understanding of AS relationship is vital to the technical research and economic exchanges of the interdomain structure of the Internet. The relationship between ASes is regarded as private information by various organizations, institutions, and operators and is not published on an open platform. Various excellent AS relationship inference algorithms have been proposed by considering the Internet as a complex network to predict the AS-level structural relationship of the Internet, which is of particular importance for Internet security.

### 2.2 Inference techniques

Gao (2001) first proposed to enhance the representation of the AS graph by defining multiple business relationships and provided an assumption that valid BGP paths are valley-free (Qiu et al., 2007) (i.e.,  $[C2P/S2S]^n[P2P]^{(0,1)}[P2C/S2S]^m$ ,  $n \geq 0$ ,  $m \geq 0$ , a path consists of zero or more C2P or S2S links, followed by zero or one P2P links, followed by zero or more P2C or S2S links, and the shape is composed of an uphill path and a downhill path or one of the two). This assumption plays an important role in the later process of inference algorithm research. A series of methods (Di Battista et al., 2003; Dimitropoulos et al., 2007; Gregori et al., 2011) has been proposed to enhance the inference performance by improving the valley-free feature. However, the subsequent studies prove that only considering degree and valley-free may be insufficient to infer the complex relationships of the Internet.

Unlike previous approaches, AS-Rank (Luckie et al., 2013) does not seek to maximize the number of valley-free paths but relies on three assumptions about Internet interdomain structure: 1) An AS enters into a provider relationship to become globally reachable; 2) A peering clique of ASes exists at the top of the hierarchy; 3) No cycle of P2C links is found. On the basis of these

assumptions, the features of a clique, transit degree, and BGP path triplets that meet the practical importance are proposed to realize the inference. AS-Rank has been used on CAIDA (Center for Applied Internet Data Analysis) until now due to its high accuracy.

Giotsas et al. (2014) rethought the problem of inferring complex AS relationships proposed by Luckie et al. (2013) and presented a new algorithm to infer the two most common types of AS relationships: Hybrid relationships and partial transit relationships. Extending the use of BGP, traceroute, and geolocation data, their inferences achieved 92.9% and 97.0% positive predictive values on hybrid and partial transit relationships, respectively.

ProbLink (Jin et al., 2019) is the first probabilistic AS relationship inference algorithm based on Naive Bayes to reduce the error rate and overcome the challenges of inferring hard links, such as nonvalley-free routing, limited visibility, and nonconventional peering practices. This approach demonstrates its practical importance in detecting route leaks, inferring complex relationships, and predicting the effect of selective advertisements. TopoScope (Jin et al., 2020) uses ensemble learning and Bayesian network to reduce the observation bias and reconstruct the Internet topology by discovering hidden links.

Susan Varghese and Ruan (2015) used the AdaBoost (Friedman et al., 2000) algorithm to train a model that predicts the link types in a given AS graph using two node attributes, degree and minimum distance, to a tier-1 node, but neither the choice of data set nor the setting of the experimental group is sufficiently rigorous. Shapira and Shavitt (2020) proposed a deep learning model BGP2VEC based on natural language processing. In recent years, many methods have mainly focused on the inference of the relationship between P2P and P2C. Giotsas et al. (2014) showed that the interconnections of ASes of a real Internet are more complex and diverse. Therefore, more suitable algorithms for inferring complex relationships should be developed to better understand the increasingly large Internet topology.

---

### 3 Data set

In this section, we describe the sources of the BGP path data and introduce the standard dataset for labeling the S2S and X2X relationships. On the validation dataset, we select the same dataset as many algorithms by focusing on community attributes (Giotsas and Zhou, 2012), and our voting-based training dataset is more representative.

#### 3.1 BGP paths

We collect BGP paths from Route Views and RIPE NCC, which are the most popular projects managing route

information collectors and making their data flow accessible and usable by any researcher. Currently, the two projects manage 29 and 25 collectors, respectively, with more than 1000 VPs in total (this number is growing over time), and continuously collect most of the routing information worldwide. To make the experimental results general and realistic, we downloaded the data flow files and extracted the BGP paths with IPv4 prefixes on the first day of April, August, and December each year from 2014 to 2018.

During the data preprocessing stage, we first remove the paths that have unassigned AS number (ASN) using the table provided by the Internet Assigned Numbers Authority. We sanitize the BGP path (Katz-Bassett et al., 2011) containing AS loops, that is, the same ASN in the path can only be adjacent to each other. We compress the paths that have the same consecutive ASN (i.e., from “A B C C” to “A B C”).

#### 3.2 S2S relationship

We use CAIDA’s AS-to-Organization (AS2Org) mapping dataset, which is derived from WHOIS information provided by Regional and National Internet Registries. The dataset is available from October 2009 onward, and the new dataset is added in each quarter. This dataset contains the ASN and the organization it belongs to, so we can classify the links that its endpoints are managed by the same organization as S2S relationship. In the following binary classification experiment, we remove all edges of the S2S relationship during preprocessing, and we randomly split it into the train, validation, and test set in the multiclassification experiment.

#### 3.3 IXP list

Our IXP list is obtained by accessing the PeeringDB database with data labeled “Routing Server” and extracting its corresponding ASN. IXP provides a neutral shared exchange structure, and clients can exchange traffic with each other after establishing a P2P connection. Previous studies removed the BGP paths that contained IXPs to study the relationships between unneutral ASes. In this study, the connection between IXP and other ASes is defined as the X2X relationship, and we expect to recognize this type of relationship through some empirical analysis. The number of AS contained in the newly extracted list can be considered the lower bound because not all IXPs have the label “Route Server”. A total of 336 IXPs are found in this list on 12/01/2020.

#### 3.4 Experimental dataset

**Training Dataset:** We obtain BGP paths from Route Views and RIPE NCC on the first day of April in

**Table 1** Training dataset established from the intersection of the three methods

Dataset	Total links	Real situation			Intersection	Coincidence rate	Intersection accuracy
		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%

2014–2018 as our source data. The intersection of the result sets obtained by the three inference methods on the same day after unified preprocessing is used as the experimental data for subsequent experiments. The reasons are as follows.

(1) The accuracy of each relationship inference approach is more than 90% in recent years, representing that the existing methods have mastered the general features of the Internet topology structure. However, the use of the inferred results of some publicly available methods as a data set is inaccurate and has obvious biases.

(2) The scale of the Internet is growing, and the number of VPs is increasing. This condition indicates that the data in recent years are more complete in terms of quantity and structure. The amount of data determines the model’s ability to express the real situation to a certain extent.

(3) The intersection of the inference results of multiple methods can be regarded as a method of hard voting based on probability, and our vote focuses on the result consistently. On the basis of the result of previous research, we greatly reduce the error of the dataset we constructed.

Table 1 shows the results of BGP paths in the same period under the current three advanced inference models and their number of intersections and proportions.

**Validation Dataset:** The community attribute (Giotsas and Zhou, 2012) is an optional BGP route tag that simplifies the implementation of the routing policy, which includes information about the relationships between some important AS pairs. Our dataset is collected by ProbLink (Jin et al., 2019), which contains the same labeled datasets based on community attributes from 2014 to 2018 on April 1st each year. Similar to the previous series of work (Luckie et al., 2013; Jin et al., 2019; 2020), we treat this validation set as a “best-effort” dataset to reproduce the existing inference models and evaluate our AS-GCN model. Table 2 shows the details of this validation dataset from 2014 to 2018.

## 4 Challenges and basic ideas

In this section, we discuss the major challenge of current inference algorithms and then propose our basic ideas. In particular, we conduct a quantitative analysis using the 04/01/2017 dataset.

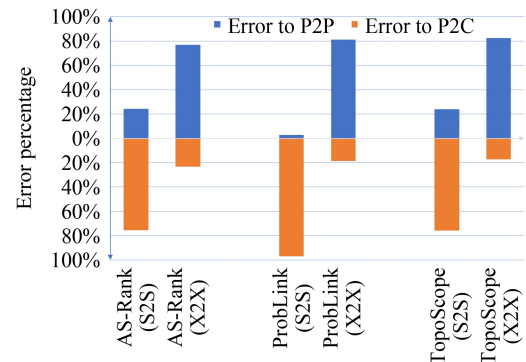
**Table 2** Validation dataset

Dataset	Validation (P2C/P2P)	All links	(S2S/X2X)
04/01/2014	44474 (26210/18264)	146167	1663/1164
04/01/2015	42512 (26528/15984)	169019	1906/1881
04/01/2016	49074 (28214/20860)	221573	2133/3957
04/01/2017	49604 (29362/20242)	219713	2276/2686
04/01/2018	44868 (30284/14584)	223221	2760/4467

### 4.1 AS relationship classification

Traditionally, Internet topological graph is constructed on the basis of the corresponding relationship between different organizations. The most frequently considered business relationships that need to be inferred are P2P (cooperation and mutual benefit) and P2C (service is accompanied by money exchanges). S2S and X2X (connect to IXP for more cost-effective communication) relationships are often considered to be known apriori.

The S2S relationship is not highly distinguishable from other types, and we conduct experiments to evaluate the inference of the existent algorithms for special types. Figure 2 gives a histogram of the misclassification of the two types of relationships of S2S and X2X by different algorithms on the 04/01/2017 dataset. The trends of the three mainstream algorithms are consistent with each other: More than 75% of S2S links are classified as P2C, and more than 75% of the X2X links are classified as P2P. For instance, 2387 S2S links are inferred by the ProbLink method, where 2316 (approximately 97%) are classified as P2C. Such results indicate the strong similarity

**Fig. 2** Fraction of the misclassification of complex relationships by different algorithms.

between S2S and P2C and between X2X and P2P, making it relatively difficult to distinguish them by using the current methods.

We believe that the characteristics of communication mode within the organization and the role of the IXP are all necessary for a better understanding and fine management of the Internet. According to what we have observed, only considering several appropriate features to distinguish the four types of AS relationships for multi-classification problems is difficult. In recent years, graph-based structural features and graph machine learning have shown excellent performance in numerous downstream tasks. As a typical complex network, the Internet is a good attempt to use structural features and graph learning models.

#### 4.2 Aggregate surrounding information

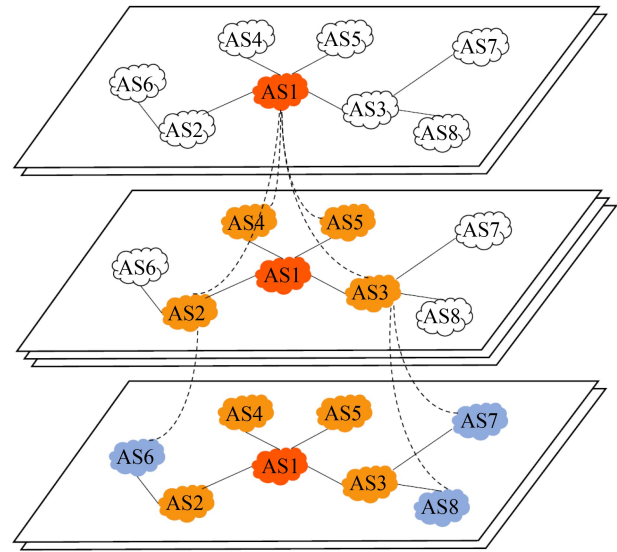
In previous work, the triplet feature (Luckie et al., 2013) is used to simulate the valley-free path in a probabilistic manner, that is, its focus on what the general situation is is similar to the neighboring edges of each edge in a BGP path. A BGP path can be typically decomposed into adjacent link pairs or three consecutive links. For example, “6939|4826|38803|56203” can be decomposed into link pairs: “6939 4826 38803”, “4826 38803 56203”, or into 3 consecutive links: “Null-<6939, 4826>-<4826, 38803>”, “<6939, 4826> - <4826, 38803> - <38803, 56203>”, “<4826, 38803>-<38803, 56203>-Null”.

The triplet features are local properties based on the first-order neighbors of ASes in the Internet graph. In this sense, we can use higher-order structural information beyond first-order neighbors for classification by employing GCNs (Kipf and Welling, 2016). From the row vector perspective of matrix multiplication, the graph-based convolution process is equivalent to the aggregation operation of the feature vectors of the neighbor nodes.

The operation of such aggregating neighbor information is given by the schematic, as shown in Fig. 3. The process of graph convolution operation can realize efficient filtering operation on graph data, and GCN brings powerful fitting capabilities by stacking and modifying multiple GCN layers. The model structure will be elaborated in the algorithm design in the next section.

## 5 Methodology

In this section, we propose AS-GCN, a GCN-based framework, to classify the AS relationship under complex Internet structures automatically and adaptively. The framework runs in two steps. First, we extract eight features of the nodes (ASes) and the edges (links). Second, we input the graph structure and feature information into our model for training and comparing with



**Fig. 3** Schematic of the aggregating surrounding information process of the GCN model.

ground truth to minimize the negative log-likelihood loss and to solve binary and multiclass AS relationship classification problems.

#### 5.1 Design and analysis of features

We calculate the difference of features between the target AS and the source AS to represent the corresponding link before feature analysis.

$$\Delta = |f_{AS1} - f_{AS2}|, \quad (1)$$

where  $f_{AS1}$  and  $f_{AS2}$  represent the feature values of AS1 and AS2, respectively, and  $\Delta$  is the difference. This method is used by default in the feature analysis below.

##### 5.1.1 Degree and transit degree

Degree is one of the most basic indices in graph theory to describe the importance of a node. The degree of a node  $i$  in an undirected network is defined as the number of edges directly connected to it. Transit degree (Luckie et al., 2013) is the number of unique neighbors appearing in the transit paths of an AS by extracting triplets. For example, suppose paths AS1, AS2, AS3, and AS4 are found, then two triplets, (AS1, AS2, AS3) and (AS2, AS3, AS4), are extracted. AS2 and AS3 have two different neighbor nodes from the extracted triplets, so their transit degree is 2. The ASes with a transit degree of zero are stub ASes, which are at the outermost layer of the network (i.e., AS1 and AS4). The transit degree is more suitable for describing the relationship clusters of ASes on the Internet and mainly reflects the scale of the customer service and cooperation of an AS. Therefore, degree and transit degree are used as important graph

structure features for subsequent algorithm design.

### 5.1.2 Distance to clique

A clique is a cluster of ASes existing at the top of the hierarchy, and the internal ASes form a transit-free state among each other (which can be regarded as a P2P relationship). Inferring clique is the first step in extracting the distance to clique features. Similar to the previous work (Luckie et al., 2013), the distance to clique feature is mainly used to capture the distance from the ASes to the network center. This feature is based on such an assumption: High-tier ASes have a large number of customers, so they are easier to be peers to achieve mutual benefit, and the better strategy for the ASes at the low-tier is to rely on the top ASes for achieving global accessibility and forming P2C relationship.

We first construct an undirected graph by extracting the AS links in the BGP paths as edges. We then calculate the average shortest path distance from each AS to the clique ( $\sum_{i=1}^N D_i/N$ ,  $N$  is the number of the clique, and  $D_i$  is the shortest distance from an AS to the  $i$ th AS in the clique) and map it to the corresponding link using Eq. (1). Finally, we use the 04/01/2017 validation dataset to evaluate the four types of AS links, and the distribution of distance to clique is shown in Fig. 4(a). This feature well reflects the considerably difference between the P2P and P2C relationships, as well as the similarity of the two pairs of AS relationships (i.e., P2C and S2S, P2P and X2X). This condition is one of the reasons that the X2X and S2S relationships are difficult to distinguish by using previous inference algorithms.

### 5.1.3 Assign VP

VPs (VP can be intuitively understood as the first node of the AS path) are typically distributed in many different geographic locations, especially at the upper tiers of the Internet hierarchy. The number of VPs is extremely

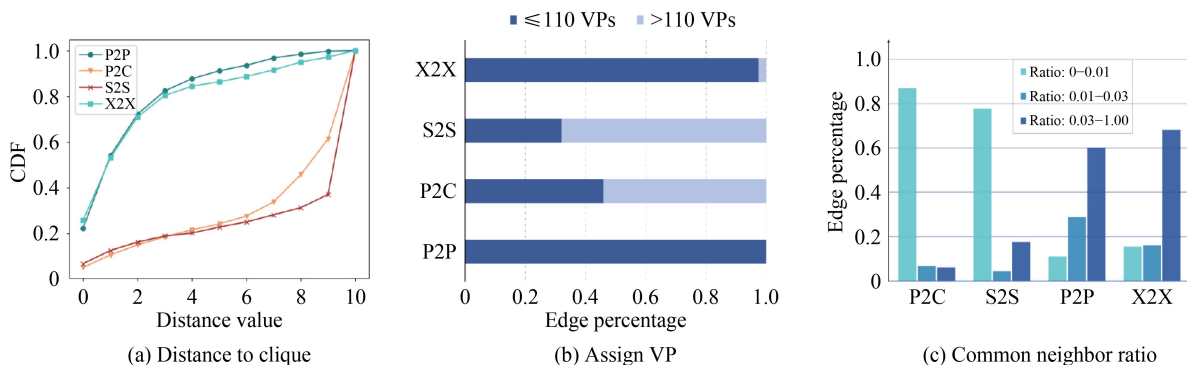
limited compared with the scale of the complete Internet structure. Here, we analyze the case of AS links that can be detected by each VP. Figure 4(b) shows that more than 97% of P2P and X2X links can be detected by less than 110 VPs (Jin et al., 2020), while more than half of P2C and S2S links are captured by more than 110 VPs. Hence, for the single feature “assign VP”, the two pairs of AS relationships (i.e., P2C and S2S, P2P and X2X) tend to be similar. This result once again confirms the challenge of the multiclass relationship classification problem.

### 5.1.4 Common neighbor ratio

This feature is defined as the ratio of common neighbors between ASes to the total number of neighbors, which is a new key point that we combined with the practical importance and the graph structure. Our intuitive idea is as follows: ASes belonging to the same organization (S2S relationship) are unlikely to connect to other types of ASes at the same time due to business. IXPs are used to promote the interconnection and exchange of the backbone of the Internet and connect many important ASes due to their functions. Therefore, IXPs’ common neighbor ratio will be higher than other types of AS. As shown in Fig. 4(c), the proportion of P2P and X2X relationships with large common neighbor ratios exceeds 60%, which validates our intuitive idea. However, the two pairs of AS relationships are indistinguishable up to the common neighbor ratio.

### 5.1.5 Distance to VP

Different from distance to clique, we also pay attention to the distance from each node (AS) to the VP in each BGP path. This feature indicates that we expect to count the distance set from the target AS to VP in all BGP paths to reflect the position of a link in all paths (i.e., front, middle, or end) and reflects the valley-free path characteristic. The distance to VP value of the node will be



**Fig. 4** Analysis of the distance to clique, assign VP, and common neighbor ratio. (a) Cumulative distribution function (CDF) of absolute distance between ASes to clique for different relationships. (b) Distribution of the number of VPs with a threshold of 110 that can be detected on each relationship’s links. (c) Distribution of different common neighbor ratio for different relationships.

expressed as a set of integers because the same node will appear in several paths. In the face of these integer sets, the mean value of the set represents the universality of the node position, and the maximum and minimum values of the set represent the specificity of the node position. As shown in Fig. 5, the use of the average value of distance to VP is more discriminative among the four types. The following feature importance analysis presented in Section 6.5 can fully prove that the importance of the mean value is higher than the maximum and minimum values.

### 5.1.6 Node hierarchy

The Internet obeys a certain hierarchical structure (Carmi et al., 2007). Considering the hierarchical features of ASes, we focus on the distribution of different types of AS links in the Internet topology. Referring to the work of *K*-shell (Carmi et al., 2007), we use transit degree to decompose the Internet into three components.

(1) All ASes in the clique form the core layer, and the number of ASes in this layer is the lowest. The ASes in the clique have a large average degree.

(2) The rest of the ASes with zero transit degree are considered as the shell of the Internet. Simultaneously, cases are observed where some ASes are directly connected to the core. This part constitutes the largest component.

(3) The remaining ASes are all classified into one category.

The hierarchy ratio is shown in Fig. 6(a), and the performance characteristics are the same as above.

### 5.1.7 AS type

The type of organization of an AS determines its function and affects its relationship with its neighbors. We obtain the AS classification dataset from CAIDA (*As\_classification*). The ground truth data are extracted from the self-reported business type of each AS list in PeeringDB. AS type can be summarized into three main categories: 1) Transit/Access. This type of ASes is classified to be either a transit and/or access provider. 2) Content. This type of ASes provides content hosting and distribution systems. 3) Enterprise. This type of ASes includes various organizations, universities, and companies. We also add the fourth type: 4) Unknown. This group contains those ASes that do not have a clear type and the neutral ASes (i.e., IXP) that do not belong to the first three categories. In the experiment, 40% of X2X edges contain ASes of the Unknown type, which is well distinguished from the other three types of ASes, as shown in Fig. 6(b).

Table 3 shows a summary of features. For the misclassification phenomenon prevalent during the above discussion of each feature, our opinion is that P2P and X2X are similar in that they are generally at the upper layers of the Internet and enhance the efficiency of Internet communications, while P2C and S2S are generally at the edge of the Internet and are more dependent on higher layer facilities.

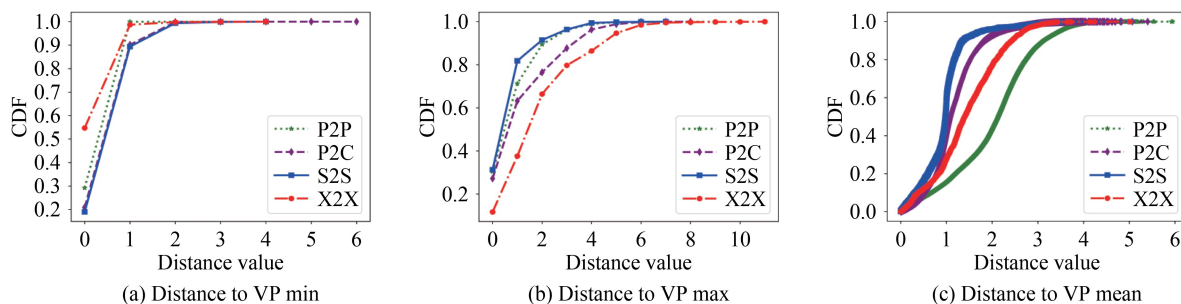


Fig. 5 Cumulative distribution function of the distance to VP using different calculation methods on four types of AS relationships.

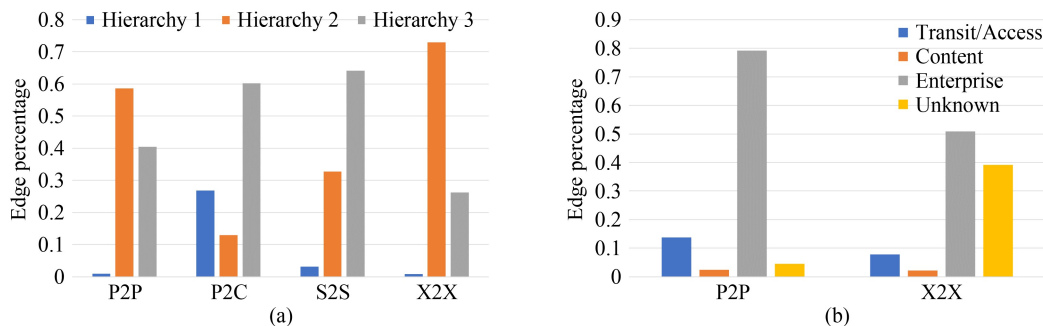


Fig. 6 (a) Hierarchical ratio of the AS-level Internet topology; (b) Distribution of the four AS types between P2P and X2X.

**Table 3** Features table

No.	Features	Key points	Newly proposed
1	Degree	Structure-based	
2	Transit degree	Structure-based	
3	Distance to clique	Internet-based	
4	Assign VP	Internet-based	
5	Common neighbor ratio	Structure-based	√
6	Distance to VP min	Internet-based	
7	Distance to VP max	Internet-based	
8	Distance to VP mean	Internet-based	
9	Node hierarchy	Internet-based	
10	AS type	Internet-based	√

### 5.2 Model framework

As shown in Fig. 7, AS-GCN mainly consists of five layers.

**Input Layer:** For the binary classification problem (i.e., P2P and P2C), the other two types of links (i.e., S2S and X2X) will be removed from the dataset in advance. Given an undirected and unweighted graph  $G = (V, E)$ , which represents a static Internet AS-level network, where  $V$  denotes the set of ASes, and  $E \in V \times V$  denotes the set of business relationships between ASes.  $v_i \in V$  denotes an AS, and  $e_{ij} = (v_i, v_j)$  denotes an AS link pointing from  $v_i$  to  $v_j$ . The adjacency matrix  $A$  is an  $n \times n$  matrix with  $A_{ij} = 1$  if  $e_{ij} \in E$  and  $A_{ij} = 0$  if  $e_{ij} \notin E$ . The graph has node attributes  $X$ , where  $X \in \mathbb{R}^{n \times d}$  denotes the feature matrix, and  $x_v \in \mathbb{R}^d$  represents the feature vector of node  $v$ .

**Feature Layer:** The feature layer is mainly to construct its feature vector for each node and generate a corresponding weight value for each edge. The feature vector is composed of the seven important features mentioned above (i.e., degree, transit degree, distance to clique, distance to VP, assign VP, node hierarchy, and AS type) and the related features after normalization, and

common neighbor ratio constitutes the weight of the edges.

**GCN Layer:** In the previous section, we introduced the basic idea of using graph convolution operation. GCN is a semisupervised learning algorithm for graph-structured data and utilizes Laplace transform to make the node aggregate the features of higher-order neighbors. In particular, GCN instantiates  $A$  to be a static matrix closely related to the normalized graph Laplacian for obtaining the following expression

$$\hat{A} = \tilde{D}^{-1/2} \tilde{A} \tilde{D}^{-1/2}, \quad (2)$$

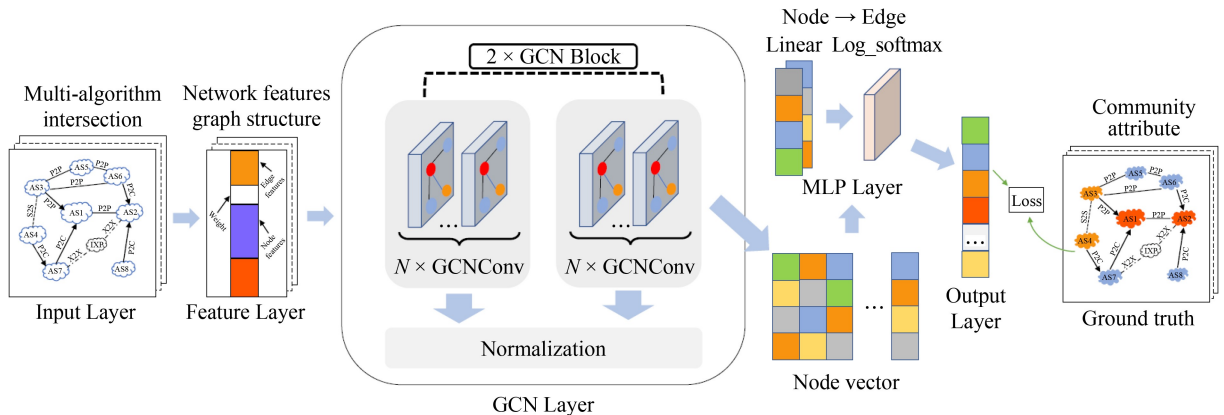
where  $\tilde{A} = A + I$ ,  $\tilde{D}_{ii} = \sum_j \tilde{A}_{ij}$ , with  $A$  being the adjacency matrix and  $I$  being the identity matrix.  $\hat{A}$  can be regarded as a graph displacement operator.

GCN performs convolution operation in the spectral domain, and each operation can aggregate an additional layer of features. The spectral convolution function is formulated as

$$H^{(l+1)} = \sigma(\hat{A}H^{(l)}W^{(l)}), \quad (3)$$

where  $W^{(l)}$  denotes the layer-specific trainable weight matrix, and  $\sigma(\cdot)$  is a nonlinear activation function (i.e., sigmoid function).  $H^{(l)} \in \mathbb{R}^{n \times k}$  represents the matrix of activation at layer  $l$ , where  $n$  and  $k$  denote the number of nodes and output dimensions of layer  $l$ , respectively. In particular, we set  $H^{(0)} = X$  as initialization input. Each convolution operation will capture the neighbor's features of an additional layer. For example,  $A$  and  $X$  perform matrix multiplication in the first round of convolution, and they are equivalent to nodes aggregating the features of their first-order neighbors. The more such multiplications, the more layers of information are abstractly merged.

We design a model that contains two identical blocks, each with two-layer GCN, to achieve semisupervised node classification on a graph with a symmetric adjacency matrix  $A$ . Our forward model then takes the block's



**Fig. 7** Model framework of AS-GCN.

simple form

$$f(\mathbf{X}, \mathbf{A}) = \text{Norm}(\text{ReLU}(\hat{\mathbf{A}}\text{ReLU}(\hat{\mathbf{A}}\mathbf{X}\mathbf{W}^{(0)}))\mathbf{W}^{(1)}), \quad (4)$$

where  $\mathbf{W}^{(0)}$  is an input-to-hidden weight matrix, and  $\mathbf{W}^{(1)}$  is a hidden-to-output weight matrix. Rectified linear unit (ReLU) is the activation function used to normalize the result. The neural network weights  $\mathbf{W}^{(0)}$  and  $\mathbf{W}^{(1)}$  are trained by using gradient descent.

**Multilayer perceptron (MLP) Layer:** In this layer, the information aggregated by the node is transformed into the corresponding edge. Let  $\mathbf{Z}$  denotes the output of GCN layer, we can obtain the following expression

$$\mathbf{Y}_{ij} = \text{Log\_softmax}(\text{Linear}(\mathbf{Z}_i, \mathbf{Z}_j)), \quad (5)$$

where  $\mathbf{Z}_i$  and  $\mathbf{Z}_j$  denote the output vector of the two end nodes corresponding to the edge  $e_{ij}$  in the graph through the GCN layer, respectively.  $\mathbf{Y}_{ij}$  is a probability vector that is used to classify the label of the corresponding edge. Compared with the method of the dot product to map edges, MLP generates a vector representation for each edge and is more suitable for classification tasks.

For semisupervised multiclassification, we apply the cross-entropy loss as the objective function, which is formulated as

$$\mathcal{L}(\theta) = -\sum_{\theta_{ij} \in E} y_{ij} \log(\mathbf{Y}_{ij}), \quad (6)$$

where  $y_{ij}$  denotes the label of edge  $e_{ij}$ , and  $\theta$  denotes all parameters needed to be learned in the model.

**Output Layer:** The model output is compared with ground truth to minimize the negative log-likelihood loss. In the course of the experiment, the training set is used to train the model for determining its weight and bias, the validation set makes the model in the best state by adjusting the hyperparameters, and the test set is used only once during the entire experiment to evaluate the performance of our model. We save the model with the highest experimental result.

## 6 Evaluation

In this section, we evaluate our GCN-based classification algorithm, AS-GCN, on the experimental dataset. We compared the performance of AS-GCN with the stable and accurate algorithms proposed before, such as AS-Rank, ProbLink, and TopoScope, and machine learning methods only considering proposed features, such as support vector machine (SVM), 1-nearest-neighbor (1NN), random forest (RF), Xgboost, LightGBM, and MLP. We proved the superiority of AS-GCN in the following three aspects.

(1) Under snapshots at different times, the accuracy of AS-GCN on binary classification can be maintained above 97%, which is higher than that of many superior

algorithms, such as AS-Rank, ProbLink, and TopoScope.

(2) The overall accuracy of multiclassification tasks evaluated on equal proportions of datasets can reach higher than 95%. Precision and recall show a general advantage.

(3) The performance of the model can be explained to a certain extent through the parameter optimization and feature importance analysis of the model.

### 6.1 Baseline methods

To evaluate the effectiveness of the model, our model is compared with the nine different methods falling into two categories: 1) relationship inference methods, including AS-Rank, ProbLink, and TopoScope; and 2) feature-based methods, including SVM, 1NN, RF, Xgboost, LightGBM, and MLP. These baselines are described as follows.

(1) **AS-Rank** (Luckie et al., 2013). The state-of-the-art AS relationship inference technique called the ‘‘AS-Rank’’ algorithm takes 11 intricate steps to label each link as C2P or P2P.

(2) **ProbLink** (Jin et al., 2019). ProbLink is the first explainable probability model used to infer AS relationships. It uses the general Naive Bayes framework to incorporate multiple link features for inferring.

(3) **TopoScope** (Jin et al., 2020). TopoScope is a framework used to accurately recover AS relationships from such fragmentary observations. TopoScope uses ensemble learning and Bayesian network to mitigate the observation bias originating from a single VP and from the uneven distribution of available VPs.

(4) **SVM** (Cortes and Vapnik, 1995). SVM classifier is a well-known supervised learning model in classification tasks.

(5) **1NN** (Peterson, 2009). 1NN method uses Euclid distance to calculate the similarities and differences.

(6) **RF** (Breiman, 2001). RF contains multiple decision tree classifiers. The classification result of the test data is determined by the score formed by the number of votes of the classification tree.

(7) **Xgboost** (Chen and Guestrin, 2016). Xgboost is a boosting algorithm, which is a boosted tree model and integrates many tree models to form a strong classifier. The tree model used is the Classification and Regression Tree (CART) model.

(8) **LightGBM** (Ke et al., 2017). LightGBM is a new gradient boosting decision tree implementation with two novel techniques: Gradient-based One-Side Sampling and Exclusive Feature Bundling.

(9) **MLP** (Arber et al., 1997). MLP is a three-layer feedforward model using backpropagation.

### 6.2 Evaluation metrics

To evaluate the performance of the proposed model, the following metrics are used: Recall, Precision, and

Accuracy. These parameters are derived by using the following equations

$$Recall = \frac{TP}{TP + FN}, \quad (7)$$

$$Precision = \frac{TP}{TP + FP}, \quad (8)$$

$$Accuracy = \frac{TP + FN}{TP + TN + FP + FN}, \quad (9)$$

where  $TP$ ,  $TN$ ,  $FP$ , and  $FN$  refer to true positive, true negative, false positive, and false negative, respectively.

Our model is implemented on PyTorch (Paszke et al., 2017). The parameters are updated by using the Adam algorithm (Kingma and Ba, 2014). Each experiment runs 200 epochs in total. In the binary classification, all results are obtained by training the AS-GCN using Adam with weight decay of  $5 \times 10^{-4}$  and an initial learning rate of 0.1. We use two blocks, where each block has two standard GCN layers (i.e., AS-GCN setting can be represented as  $2 \times 2$ ), to learn the graph structure. In the multiclassification, all results are obtained by training the AS-GCN using Adam with weight decay of 0 and an initial learning rate of 0.05. We use the  $2 \times 1$  setting of AS-GCN to learn the graph structure.

### 6.3 Binary classification

As a binary classification task, inferring the relationship between P2P and P2C is the main focus of many existing algorithms. We validate our AS-GCN on the same data set. Our experimental snapshots are taken from the BGP path on the first day of April each year from 2014 to 2018, and our validation set is the “best-effort” validation data set based on community attributes.

Table 4 shows the experimental results on three snapshots selected at the same time each year from 2016 to 2018. We use the same validation sets for different methods. As shown in Table 4, our AS-GCN outperforms the traditional inference methods with Accuracy higher than 97%, showing its excellence and stability although the three traditional inference algorithms show good performance and generality in terms of Accuracy, above 93% and even up to 97%. And, the Accuracy gap between these feature-based methods and AS-GCN is more

obvious. The results of RF, Xgboost, and LightGBM show that ensemble learning, which integrates the weak classifiers into strong classifiers, is better than the single classifier. The neural network-based MLP, which shows poor results on this task, exhibits the significant enhancement of graph structure features for the AS relationship classification task.

### 6.4 Multiclassification

Classification of multiple relationships for the Internet is more difficult than binary classification based on the analysis in Section 4. Considering more types of links can help describe the Internet topology more precisely to better understand its evolution mechanism.

Given that S2S and X2X relationships only occupy a small part of the entire Internet, we randomly sample P2P and P2C relationships to make the four types balanced and to prevent the poor training effect caused by the severe imbalance of samples in different types. We make  $P2P : P2C : S2S : X2X \approx 1 : 1 : 1 : 1$  and divide the four types of AS relationships into separate training, validation, and test sets, with a ratio of 6 : 2 : 2 for multiclassification experiments. Reflecting the performance of the model is difficult if Accuracy is used as the only evaluation metric. Thus, we consider Precision and Recall. Accuracy measures the global sample prediction, and each class needs to calculate its corresponding value separately for Precision and Recall. After downsampling, we can view each class equally, and we use a macroaveraging method that directly averages the evaluation metrics of the different classes by adding them together and giving the same weight to all classes.

The multiclassification results are shown in Table 5, where AS-GCN behaves surprisingly well in classifying the four types of relationships, with Accuracy, Precision, and Recall all exceeding 95%. The ensemble-based learning approach has comparable performance. However, the outperformance of our model against all baselines suggests that improving the performance by simply increasing the number and quality of features is insufficient, and that a more appropriate model is needed. Methods, such as AS-Rank and ProbLink, are not listed here because they are specifically designed for binary classification. Such results validate again that our AS-GCN is more general and can be used to effectively recognize different types of relationships between ASes.

**Table 4** Accuracy of AS link classification methods for binary classification experiment

Dataset	Traditional inference methods			Feature-based methods					AS-GCN	
	AS-Rank	ProbLink	TopoScope	SVM	1NN	RF	Xgboost	LightGBM		MLP
04/01/2016	96.75	96.80	96.44	61.74	81.50	93.14	94.92	91.74	77.30	<b>97.51</b>
04/01/2017	94.19	95.08	93.77	62.70	81.10	82.04	91.10	88.90	78.00	<b>97.51</b>
04/01/2018	<b>97.37</b>	96.55	96.21	73.31	74.10	79.46	92.65	86.18	69.20	97.20

**Table 5** Accuracy, Precision, and Recall of the multiclassification experiment

Dataset	Links	Methods	Accuracy (all)	Precision				Macro Precision	Recall				Macro Recall
				P2P	P2C	S2S	X2X		P2P	P2C	S2S	X2X	
04/01/2016	11093	SVM	73.78	56.00	99.00	84.00	81.00	80.00	96.00	83.00	89.00	30.00	74.50
		INN	81.40	78.00	75.00	87.00	88.00	82.00	73.00	82.00	88.00	85.00	82.00
		RF	94.71	94.37	99.79	88.99	95.19	94.59	97.20	97.20	94.29	90.49	94.80
		Xgboost	94.66	94.87	98.99	89.34	94.86	94.52	96.20	97.80	93.81	91.04	94.71
		LightGBM	94.26	94.43	98.58	91.63	92.24	94.22	95.00	97.00	91.19	90.49	93.42
		MLP	66.24	55.00	77.00	76.00	60.00	67.00	65.00	64.00	94.00	49.00	68.00
		AS-GCN	<b>95.32</b>	94.37	99.60	92.36	94.60	<b>95.23</b>	97.20	98.80	92.14	92.87	<b>95.25</b>
04/01/2017	10962	SVM	73.60	56.00	99.00	82.00	77.00	78.50	94.00	81.00	89.00	30.00	73.50
		INN	79.57	76.00	74.00	87.00	85.00	80.50	71.00	79.00	89.00	82.00	80.25
		RF	92.62	90.17	99.17	86.68	94.66	92.67	97.20	95.20	92.76	85.85	92.75
		Xgboost	93.28	92.69	98.37	87.21	94.72	93.25	96.40	96.40	94.12	86.78	93.43
		LightGBM	93.33	92.34	98.39	87.83	94.41	93.24	96.40	97.60	91.40	88.08	93.37
		MLP	64.91	51.00	71.00	77.00	71.00	67.50	73.00	62.00	95.00	34.00	66.00
		AS-GCN	<b>95.54</b>	95.85	99.40	93.76	93.28	<b>95.57</b>	97.20	99.20	91.03	94.42	<b>95.46</b>
04/01/2018	13227	SVM	76.60	55.00	99.00	86.00	89.00	82.25	95.00	79.00	89.00	55.00	79.50
		INN	82.77	75.00	77.00	91.00	87.00	82.50	72.00	83.00	87.00	87.00	82.25
		RF	92.98	91.80	93.96	91.23	94.08	92.77	94.00	96.40	88.30	93.26	92.99
		Xgboost	93.96	93.69	95.45	91.35	94.77	93.82	95.00	96.40	89.62	94.57	93.90
		LightGBM	92.69	91.49	96.78	91.99	91.57	92.96	94.60	96.20	84.53	94.46	92.45
		MLP	71.83	56.00	84.00	89.00	70.00	74.75	70.00	47.00	89.00	79.00	71.25
		AS-GCN	<b>95.38</b>	95.42	99.00	93.26	94.60	<b>95.57</b>	95.80	99.40	92.09	94.94	<b>95.56</b>

### 6.5 Parameter and feature analysis

In this subsection, we first conduct some sensitivity analysis of hyperparameters on the 04/01/2017 dataset. We analyze how different choices of the hyperparameters may affect the performance of our AS-GCN. Second, we analyze the importance of seven features in detail by controlling other variables. Finally, we make a visual analysis for the multiclassification results of AS links.

Figure 8 gives the parameter optimization process of AS-GCN on the 04/01/2017 dataset for the multiclassification. The main parameters considered here are GCN structure, learning rate, and weight decay. We obtain the optimal parameter in each round of experiments and then use it as a fixed value before evaluating the next parameter. In Fig. 8, the optimized parameters are obtained as follows: The setting of the AS-GCN structure is  $2 \times 1$ . The learning rate is 0.05. The model does not have an obvious overfitting phenomenon, so the weight decay parameter is set to 0. Simultaneously, the final output of each GCN block needs to be added to the normalization operation after the activation function ReLU. We share the same set of parameters on different experimental datasets.

Figure 9 and Table 6 show the importance of the

features, which is a method of scoring and sorting the input features of a classification model, and reveals the relative importance of each feature when making a classification. We evaluate each feature of the AS-GCN model using Accuracy by removing only one specified feature from the all at once to form the left side of the histogram in Fig. 9. The feature importance evaluation results under the other three comparison algorithms are shown in the right side of the figure. For AS-GCN, the importance score (denoted as  $S$ ) can be formulated as follows

$$S_i = \frac{|Acc_{org} - Acc_i|}{\sum_{j=1}^d |Acc_{org} - Acc_j|} \times 100\%, \quad (10)$$

where  $Acc_{org}$  represents the Accuracy of the original model,  $Acc_i$  represents the Accuracy by removing the  $i$ th dimension feature, and  $d$  is the number of features involved in importance analysis. From Fig. 9 and Table 6, the proposed common neighbor ratio is the most important feature in AS-GCN, and distance to clique and distance to VP are relatively important features in feature-based methods. We also average the ranking of importance scores obtained by the four methods. The average ranking is listed in Table 6. Distance to clique and common neighbor ratio are manifested as the strong important

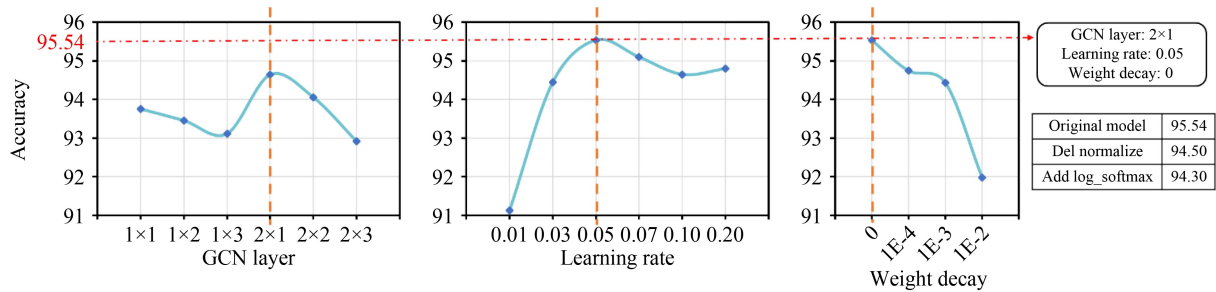


Fig. 8 Parameter sensitivity analysis for AS-GCN model.

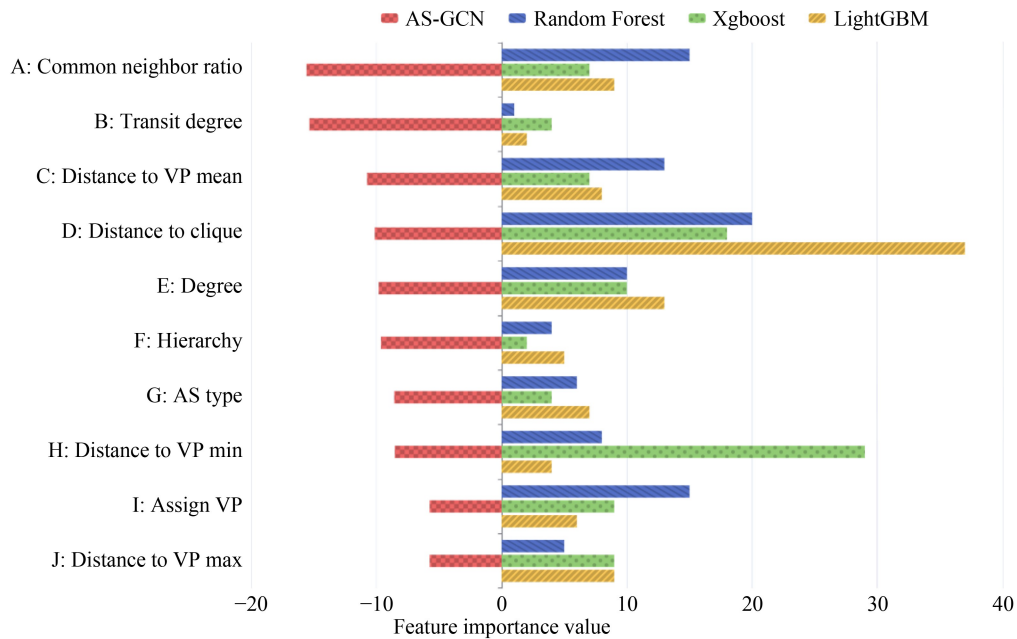


Fig. 9 Feature importance analysis on AS-GCN and its comparison methods.

Table 6 Feature importance analysis

Features removed	Accuracy	Importance score	Average ranking
Common neighbor ratio	91.45	15.60	3.25
Transit degree	91.51	15.37	7.75
Distance to VP mean	92.72	10.76	4.75
Distance to clique	92.88	10.14	2
Degree	92.96	9.84	3.75
Hierarchy	93.01	9.65	8.25
AS type	93.29	8.58	7
Distance to VP min	93.30	8.54	6
Assign VP	94.03	5.76	5.75
Distance to VP max	94.03	5.76	6.50

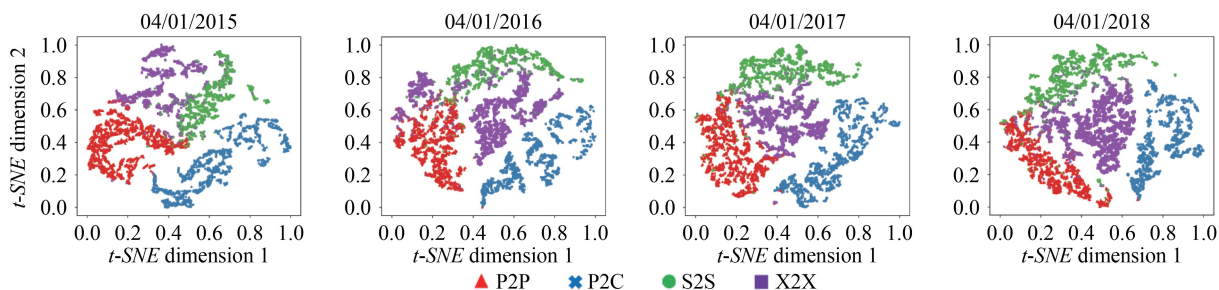
features in each algorithm. On the contrary, transit degree as the exclusive feature of AS-level network topology does not show excellent performance.

We use a dimensionality reduction algorithm *t-SNE* (van der Maaten and Hinton, 2008) to project the high-dimensional vector output by AS-GCN into the 2D space

for intuitive visualization. Figure 10 shows the degree of aggregation and discrimination between the four types of AS relationships on different datasets, proving the effectiveness of our proposed AS-GCN framework once again.

## 7 Conclusions and future work

In this study, we add the Internet facility IXP to the analysis of Internet structure for the first time and construct a reliable dataset for multiple classification tasks. Facing the advantages and disadvantages of traditional inference methods, we investigate the AS relationship inference problem from two perspectives: More features (common neighbor ratio and AS type) and a more fitting model (AS-GCN). The advantage of our model is that it combines AS relational features, graph structure features, and graph learning methods. Our experimental study shows that this new framework effectively achieves excellent performance on binary classification tasks and exceeds a series



**Fig. 10** Visualization of multiclassification obtained by AS-GCN using *t-SNE* algorithm.

of baselines on multiclassification tasks with evaluation metrics above 95%. The efficiency and stability exhibited by AS-GCN provide some support for research in anomaly detection and location, complex relationship inference, and network service recommendation.

Several challenges can be explored in future studies. Further study on the classification of complex relationships (i.e., hybrid and partial transit relationships) with machine learning methods is needed. The dataset problems have been a particularly critical part of machine learning methods. The labels of our datasets come from inference algorithms, so we cannot guarantee that they are 100% correct. We are also eager for a standard dataset to evaluate and improve our inference methods better. Although we have proposed several new features, such as type and common neighbor ratio, the lack of features with strong distinction remains a problem in the multiclassification task. Therefore, the representative features of each type of relationship must be determined in future studies of the various types of business relationships.

**Acknowledgements** The authors would like to thank all the members of the IVSN Research Group, Zhejiang University of Technology for the valuable discussions about the ideas and technical details presented in this paper.

## References

- Apostolaki M, Marti G, Müller J, Vanbever L (2018). SABRE: Protecting Bitcoin against routing attacks. arXiv preprint. arXiv:1808.06254
- Arber S, Hunter J J, Ross Jr J, Hongo M, Sansig G, Borg J, Perriard J C, Chien K R, Caroni P (1997). MLP-deficient mice exhibit a disruption of cardiac cytoarchitectural organization, dilated cardiomyopathy, and heart failure. *Cell*, 88(3): 393–403
- Böttger T, Antichi G, Fernandes E L, di Lallo R, Bruyere M, Uhlig S, Castro I (2018). The elusive Internet flattening: 10 years of IXP growth. arXiv preprint. arXiv:1810.10963
- Breiman L (2001). Random forests. *Machine Learning*, 45(1): 5–32
- Brito S H B, Santos M A S, dos Reis Fontes R, Perez D A L, da Silva H D L, Rothenberg C R E (2016). An analysis of the largest national ecosystem of public Internet exchange points: The case of Brazil. *Journal of Communication and Information Systems*, 31(1): 256–271
- Carmi S, Havlin S, Kirkpatrick S, Shavitt Y, Shir E (2007). A model of Internet topology using *k*-shell decomposition. *Proceedings of the National Academy of Sciences of the United States of America*, 104(27): 11150–11154
- Castro I, Cardona J C, Gorinsky S, Francois P (2014). Remote peering: More peering without Internet flattening. In: *Proceedings of the 10th ACM International on Conference on Emerging Networking Experiments and Technologies*. Sydney: Association for Computing Machinery, 185–198
- Chen T, Guestrin C (2016). XGBoost: A scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. San Francisco, CA: Association for Computing Machinery, 785–794
- Cho S, Fontugne R, Cho K, Dainotti A, Gill P (2019). BGP hijacking classification. In: *Network Traffic Measurement and Analysis Conference (TMA)*. Paris: IEEE, 25–32
- Cohen A, Gilad Y, Herzberg A, Schapira M (2016). Jumpstarting BGP security with path-end validation. In: *Proceedings of the ACM SIGCOMM Conference*. Florianopolis: Association for Computing Machinery, 342–355
- Cortes C, Vapnik V (1995). Support-vector networks. *Machine Learning*, 20(3): 273–297
- Dhamdhere A, Clark D D, Gamero-Garrido A, Luckie M, Mok R K P, Akiwate G, Gogia K, Bajpai V, Snoeren A C, Claffy K (2018). Inferring persistent interdomain congestion. In: *Proceedings of the Conference of the ACM Special Interest Group on Data Communication*. Budapest: Association for Computing Machinery, 1–15
- Di Battista G, Patrignani M, Pizzonia M (2003). Computing the types of the relationships between autonomous systems. In: *IEEE INFOCOM. 22nd Annual Joint Conference of the IEEE Computer and Communications Societies*. San Francisco, CA: IEEE, 156–165
- Dimitropoulos X, Krioukov D, Fomenkov M, Huffaker B, Hyun Y, Claffy K C, Riley G (2007). AS relationships: Inference and validation. *Computer Communication Review*, 37(1): 29–40
- Friedman J, Hastie T, Tibshirani R (2000). Additive logistic regression: A statistical view of boosting (with discussion and a rejoinder by the authors). *Annals of Statistics*, 28(2): 337–407
- Gao L (2001). On inferring autonomous system relationships in the Internet. *IEEE/ACM Transactions on Networking*, 9(6): 733–745
- Gill P, Arlitt M, Li Z, Mahanti A (2008). The flattening Internet topology: Natural evolution, unsightly barnacles or contrived collapse? In: *International Conference on Passive and Active Network Measurement*. Cleveland, OH: Springer, 1–10
- Gill P, Schapira M, Goldberg S (2011). Let the market drive deployment: A strategy for transitioning to BGP security. *Computer*

- Communication Review, 41(4): 14–25
- Giotsas V, Luckie M, Huffaker B, Claffy K (2014). Inferring complex AS relationships. In: Proceedings of the Conference on Internet Measurement Conference. Vancouver, BC: Association for Computing Machinery, 23–30
- Giotsas V, Zhou S (2012). Valley-free violation in Internet routing: Analysis based on BGP community data. In: IEEE International Conference on Communications (ICC). Ottawa, ON: IEEE, 1193–1197
- Gregori E, Improta A, Lenzini L, Rossi L, Sani L (2011). BGP and inter-AS economic relationships. In: International Conference on Research in Networking. Valencia: Springer, 54–67
- Jin Y, Scott C, Dhamdhere A, Giotsas V, Krishnamurthy A, Shenker S (2019). Stable and practical AS relationship inference with Prob-Link. In: Proceedings of the 16th USENIX Conference on Networked Systems Design and Implementation. Boston, MA: USENIX Association, 581–597
- Jin Z, Shi X, Yang Y, Yin X, Wang Z, Wu J (2020). TopoScope: Recover AS relationships from fragmentary observations. In: Proceedings of the ACM Internet Measurement Conference. New York, NY: Association for Computing Machinery, 266–280
- Karlin J, Forrest S, Rexford J (2008). Autonomous security for autonomous systems. *Computer Networks*, 52(15): 2908–2923
- Katz-Bassett E, Choffnes D R, Cunha Í, Scott C, Anderson T, Krishnamurthy A (2011). Machiavellian routing: Improving Internet availability with BGP poisoning. In: Proceedings of the 10th ACM Workshop on Hot Topics in Networks. Cambridge, MA: Association for Computing Machinery, 1–6
- Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Q, Liu T Y (2017). LightGBM: A highly efficient gradient boosting decision tree. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, CA: Curran Associates Inc., 3149–3157
- Kingma D P, Ba J (2014). Adam: A method for stochastic optimization. arXiv preprint. arXiv:1412.6980
- Kipf T N, Welling M (2016). Semi-supervised classification with graph convolutional networks. arXiv preprint. arXiv:1609.02907
- Labovitz C, Iekel-Johnson S, McPherson D, Oberheide J, Jahanian F (2010). Internet inter-domain traffic. *Computer Communication Review*, 40(4): 75–86
- Luckie M, Huffaker B, Dhamdhere A, Giotsas V, Claffy K (2013). AS relationships, customer cones, and validation. In: Proceedings of the Conference on Internet Measurement Conference. Barcelona: Association for Computing Machinery, 243–256
- Motamedi R, Rejaie R, Willinger W (2015). A survey of techniques for Internet topology discovery. *IEEE Communications Surveys and Tutorials*, 17(2): 1044–1065
- Orsini C, King A, Giordano D, Giotsas V, Dainotti A (2016). BGPStream: A software framework for live and historical BGP data analysis. In: Proceedings of the Internet Measurement Conference. Santa Monica, CA: Association for Computing Machinery, 429–444
- Paszke A, Gross S, Chintala S, Chanan G, Yang E, DeVito Z, Lin Z, Desmaison A, Antiga L, Lerer A (2017). Automatic differentiation in PyTorch. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, CA: Curran Associates Inc., 1–4
- Peterson L E (2009). *K*-nearest neighbor. *Scholarpedia*, 4(2): 1883
- Qiu S Y, McDaniel P D, Monrose F (2007). Toward valley-free inter-domain routing. In: IEEE International Conference on Communications. Glasgow: IEEE, 2009–2016
- Shapira T, Shavitt Y (2020). Unveiling the type of relationship between autonomous systems using deep learning. In: IEEE/IFIP Network Operations and Management Symposium. Budapest: IEEE, 1–6
- Smith J M, Schuchard M (2018). Routing around congestion: Defeating DDoS attacks and adverse network conditions via reactive BGP routing. In: IEEE Symposium on Security and Privacy (SP). San Francisco, CA: IEEE, 599–617
- Susan Varghese J, Ruan L (2015). A machine learning approach to edge type prediction in Internet AS graphs. Online Paper
- Sundaresan S, Deng X, Feng Y, Lee D, Dhamdhere A (2017). Challenges in inferring Internet congestion using throughput measurements. In: Proceedings of the Internet Measurement Conference. London: Association for Computing Machinery, 43–56
- Tozal M E (2018). Policy-preferred paths in AS-level Internet topology graphs. *Theory and Applications of Graphs*, 5(1): 3
- van der Maaten L, Hinton G (2008). Visualizing data using *t*-SNE. *Journal of Machine Learning Research*, 9(86): 2579–2605