POLITECNICO DI TORINO
Repository ISTITUZIONALE

Poster: Towards Contextual Representations for BGP Analytics

Original
Poster: Towards Contextual Representations for BGP Analytics / Huang, K., Mellia, M., Loye, J., Krenc, T., Fontugne, R.. - (2026), pp. 1207-1208. (2026 ACM Internet Measurement Conference Karlsruhe (DE) October 12 - 16, 2026) [10.1145/3777912.3847590].

Availability:
This version is available at: 11583/3016159 since: 2026-09-30T11:58:47Z

Publisher:
ACM

Published
DOI:10.1145/3777912.3847590

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

18 November 2026

Poster: Towards Contextual Representations for BGP Analytics

Kai Huang, Marco Mellia
Politecnico di Torino

Justin Loye, Thomas Krenc, Romain Fontugne
IIJ Research Laboratory

ACM Reference Format:
Kai Huang, Marco Mellia and Justin Loye, Thomas Krenc, Romain Fontugne. 2026. Poster: Towards Contextual Representations for BGP Analytics. In *Proceedings of the 2026 ACM Internet Measurement Conference (IMC '26), October 12–16, 2026, Karlsruhe, Germany*. ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3777912.3847590

1 Introduction

The networking and security communities have long analyzed Border Gateway Protocol (BGP) data to understand the structure and behaviors of the Internet's inter-domain routing ecosystem for a wide range of applications, from characterizing Autonomous Systems (ASes) and their relationships [2], identifying routing patterns and anomalies [3], to inferring BGP communities [4] semantics, etc. Many of these analyses, however, rely on heuristics or manually engineered features, requiring substantial domain knowledge and human effort. Such approaches are commonly tailored for specific tasks and derived from partial observations, that may unable to fully capture complex and context-dependent patterns of the routing ecosystem and introduce biases, making them hard to generalize.

Recent advances in Artificial Intelligence (AI) have shown that self-supervised pre-training can learn generalizable and transferable representations (i.e. embeddings) from large-scale unlabeled data, enabling strong performance across a wide range of downstream applications. Motivated by this success, we investigate whether self-supervised representation learning with large volumes of BGP routing data can capture meaningful routing characteristics without task-specific ground truth, and provide reusable representations for different downstream applications, offering an automated alternative for BGP data analytics.

We present our preliminary exploration of self-supervised BGP embeddings by pre-training a BERT-like Transformer model and validating its learned representations. Through several preliminary downstream tasks, we find that the learned embeddings capture meaningful characteristics of AS that can support accurate inference with simple classifiers. These initial results motivate us to further explore the capabilities and limitations of the learned representations under more challenging tasks and different scenarios. Ultimately, our goal is to develop a foundation model that can effectively capture the structure and dynamics of inter-domain routing semantics and support a broad range of analytics.

2 Initial Model Design

Following the intuition that an AS path is, by definition, an ordered sequence of ASes, adapting approaches from Natural Language Processing (NLP) provides a natural way to model such routing sequences. Prior work such as BGP2Vec [5] has demonstrated the value of learning AS representations from their co-occurrence in routing paths. We aim to investigate whether a Transformer-based model can further exploit the information in each AS path and learn richer representations that can also support context-dependent analysis of routing paths at different levels.

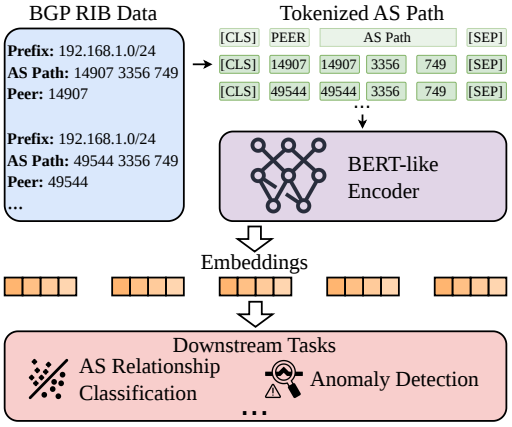


Figure 1: Overview of BGP embedding pipeline.

We adopt a BERT-like encoder model, which is an encoder-only Transformer designed to produce contextual representations of input sequences for downstream tasks such as classification and regression. We show the overview of the proposed approach in Fig. 1. We first extract AS paths from all available RIB snapshots of RIPE RIS on June 26th, 2026. We filtered out all paths with only one AS, and remove all routes that the origin prefix is larger than /8. We convert each path into a sequence of AS tokens. The tokenized AS paths are then used to self-supervise the BERT-like encoder.

AS Tokenization. We treat each AS path as a sequence in which every AS number (ASN) corresponds to a token. We construct the vocabulary from the ASes observed in the training data, together with special tokens required by the model. All reserved ASes[1] are mapped to a dedicated unknown token. We prepend the peer ASN as a conditioning token, allowing the model to capture the observation context regardless if the peer ASN is absent from the AS path, i.e. when peering with an IXP route server. Its fixed position after [CLS] distinguishes it from regular AS-path tokens.

BERT-like Encoder. Given the fact AS paths are much shorter comparing with natural-language sequences, we start our first exploration with a BERT-like Transformer encoder with 4 Transformer layers ($L = 4$), 4 attention heads ($A = 4$), and a hidden dimension $H = 128$, which is substantially smaller than standard BERT model ($L = 12$, $A = 12$, $H = 768$). As input to the Transformer encoder, each AS token is mapped to a learnable token embedding and combined with a positional embedding. We pre-train the model

[1]We follow the Special-Purpose Autonomous System (AS) Numbers listed in RFC7249

This work is licensed under a Creative Commons Attribution 4.0 International License.
IMC '26, Karlsruhe, Germany
© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2327-8/2026/10
https://doi.org/10.1145/3777912.3847590

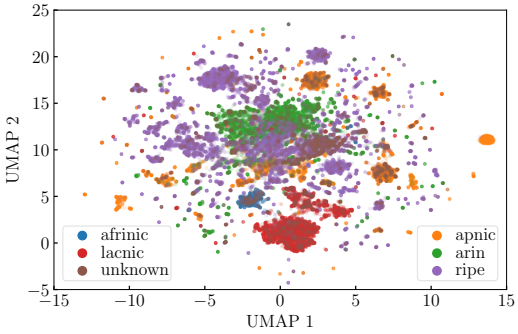


Figure 2: UMAP visualization of embeddings with their RIRs.

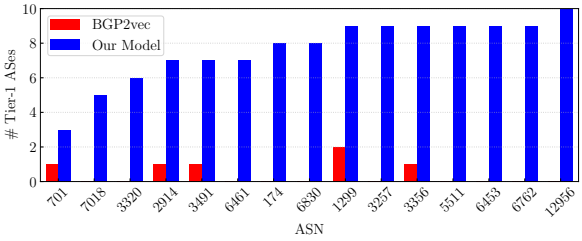


Figure 3: Nearest Neighbors of Tier-1 ASes.

using the Masked Language Modeling (MLM) objective: randomly mask 25% AS tokens in each path and the model learns to recover them from the surrounding route context. Unlike the original BERT, we omit next sentence prediction or segment embeddings, since we consider each AS path independent rather than related with others.

Through pre-training, the model learns a static token embedding for each AS, representing global routing characteristics accumulated across its occurrences in the training data. And given an AS path, the encoder can further produce a contextual embedding for each AS in the path, reflecting its role within that particular path. These embeddings can subsequently be used for different downstream BGP analytics tasks.

3 Preliminary Validation

We test on some downstream simple tasks to validate whether the learned embeddings capture a basic but meaningful AS property that was never explicitly provided during pre-training.

RIR Visualization. For a qualitative view, we first project the static AS embeddings into two dimensions using UMAP and color each AS according to its Regional Internet Registry (RIR). Fig. 2 shows that embeddings exhibit clear separation and clustering among different RIRs, despite the model being trained without RIR labels. This suggests that the embeddings learn meaningful geographic and operational characteristics of ASes implicitly reflected in BGP routing patterns.

Tier-1 Characterization. We next test whether the embeddings capture the hierarchical roles of ASes. For each Tier-1 AS, we look its 10 nearest neighbors in the embedding space and count how many are also Tier-1 ASes. As Fig. 3 shows, our model places Tier-1 ASes substantially closer to one another than BGP2Vec, with most Tier-1 ASes having 7–10 other Tier-1 networks among their 10 nearest neighbors. This suggests that the learned representations capture similarities in the routing roles of top-level transit networks.

Table 1: F1 scores for AS relationship inference.

Test Set	Model	P2C	C2P	P2P	Macro
Cross	BGP2Vec	0.89	0.88	0.97	0.92
	Ours	0.93	0.92	0.98	0.94
Unseen	BGP2Vec	0.85	0.81	0.96	0.89
	Ours	0.90	0.89	0.97	0.92

AS Relationship Inference. We further validate the learned representations on a practical downstream task of inferring relationship between ASes. We concatenate the embeddings of two ASes and feed them into a simple Multilayer Perceptron Classification (MLP) head with the CAIDA AS relationship label [1]. We partition ASes into two disjoint sets and use AS pairs drawn from one set to train the MLP head. We then consider two test settings: **Cross** set, where one AS in each pair is in the training set, and **Unseen** set, where neither AS was seen during MLP training. As Table 1 shows, our model achieves strong performance in both settings and consistently outperforms BGP2Vec, which suggests that the model captures generalizable relationship characteristics that not observed during MLP training.

4 Future Works

We plan to extend the evaluation from AS embeddings to AS-path representations, examining whether the model captures path-level routing behavior, such as distinguishing valid from anomalous routes. We will also explore alternative or complementary self-supervised pre-training strategies from two perspectives. From the *data perspective*, rather than directly training on large volumes of redundant BGP observations, we aim to explore the possibility of constructing more representative routing samples that emphasize meaningful path structures, such as the shortest paths between AS pairs or diverse paths toward a given origin. From the *learning perspective*, we will move beyond MLM and investigate other self-supervised objectives that better capture inter-domain routing properties and path-level semantics.

We will further evaluate and improve the model through more challenging real-world applications, including BGP hijack and route-leak detection, and BGP community semantic inference, etc. More broadly, we aim to investigate the model's robustness and generalizability, particularly against biases introduced by the partial and vantage-point-dependent view of BGP observations, as well as its ability to capture and adapt to the temporal dynamics of the continuously evolving routing ecosystem. Ultimately, these efforts move toward a general and effective representation for BGP analytics.

References

[1] [n. d.]. *The CAIDA AS Relationships Dataset.* Retrieved July, 2026 from https://www.caida.org/catalog/datasets/as-relationships/

[2] Vasileios Giotsas, Matthew Luckie, Bradley Huffaker, and KC Claffy. 2014. Inferring complex AS relationships. In *Proceedings of the 2014 Conference on Internet Measurement Conference.*

[3] Thomas Holterbach, Thomas Alfroy, Amreesh Phokeer, Alberto Dainotti, and Cristel Pelsser. 2024. A system to detect {Forged-Origin}{BGP} hijacks. In *21st USENIX Symposium on Networked Systems Design and Implementation (NSDI 24).*

[4] Thomas Krenc, Matthew Luckie, Alexander Marder, and KC Claffy. 2023. Coarse-grained inference of BGP community intent. In *Proceedings of the 2023 ACM on Internet Measurement Conference.*

[5] Tal Shapira and Yuval Shavitt. 2022. BGP2Vec: Unveiling the latent characteristics of autonomous systems. *IEEE Transactions on Network and Service Management* 19, 4 (2022).