Table 6: The classification performance of MARU over different sizes of embedding dimensions d with 50% of training labeled nodes in Yelp.

Metric	$d = 16$	$d = 32$	$d = 64$	$d = 128$	$d = 256$
Macro-F1	0.6931	0.6933	0.6982	**0.6999**	0.6974
Micro-F1	0.7112	0.7114	0.7154	**0.7186**	0.7156

6 CONCLUSIONS

In this paper, we propose MARU, a novel approach for heterogeneous network embedding by exploiting meta-contexts in random walks. To address the bias caused by conventional random walks, the algorithm of bidirectional extended random walks is proposed to efficiently and fairly capture the comprehensive structural information in the networks. The meta-context aware node embeddings are then designed and optimized to represent properties of the nodes for different local heterogeneous contexts, thereby inferring the node representations based on aggregations over the meta-context distributions. Extensive experiments demonstrate that our proposed approach significantly outperforms state-of-the-art heterogeneous network embedding methods across three general network mining tasks, including multi-label node classification, node clustering, and link prediction. The reasons and insights can be concluded as follows: (1) the algorithm of bidirectional extended random walks effectively alleviates the bias for tail nodes with a theoretical guarantee; (2) the effectiveness of meta-contexts and meta-context aware node embeddings implies that a node can have distinct properties with different local heterogeneous contexts, which benefit the network representation learning; (3) the nature of heterogeneous networks can be much different from the traits of homogeneous networks, so it is crucial to tackle the problems of heterogeneous networks with specific and appropriate technologies.

APPENDIX

A THE PROOF OF COROLLARY 1

Proof. For the one-directional random walk, the expected number of visited tail nodes is $E_o = 1 + \sum_{i=0}^{2n} i \cdot p^i \cdot (1-p)$. For the bidirectional random walk, the expected number of visited tail nodes is $E_b = 1 + 2 \cdot \sum_{i=0}^{n} i \cdot p^i \cdot (1-p)$. Therefore, we have

$$\begin{aligned}\lim_{n\to\infty} E_b - E_o &= \lim_{n\to\infty} (1-p) \cdot \left(\sum_{i=1}^{n} i \cdot p^i - (i+n) \cdot p^{i+n} \right) \\ &= \lim_{n\to\infty} \frac{p \cdot (1-p^n) \cdot (2 \cdot n \cdot p^{n+1} - (2n+1) \cdot p^n + 1)}{1-p} \\ &= \frac{p}{1-p} > 0\end{aligned}$$

□

ACKNOWLEDGEMENT

We would like to thank the anonymous reviewers for their helpful comments. The work was partially supported by NSF DGE-1829071 and NSF IIS-2031187.

REFERENCES

[1] David Arthur and Sergei Vassilvitskii. 2007. k-means++: The advantages of careful seeding. In *SODA*. 1027–1035.
[2] Hongyun Cai, Vincent W Zheng, and Kevin Chang. 2018. A comprehensive survey of graph embedding: problems, techniques and applications. *TKDE* (2018).
[3] Yelp Dataset Challenge. 2013. Yelp dataset challenge. (2013).
[4] Shiyu Chang, Wei Han, Jiliang Tang, Guo-Jun Qi, Charu C Aggarwal, and Thomas S Huang. 2015. Heterogeneous network embedding via deep architectures. In *KDD*. ACM, 119–128.
[5] Haochen Chen, Bryan Perozzi, Yifan Hu, and Steven Skiena. 2018. Harp: Hierarchical representation learning for networks. In *AAAI*.
[6] Ting Chen and Yizhou Sun. 2017. Task-guided and path-augmented heterogeneous network embedding for author identification. In *WSDM*. ACM, 295–304.
[7] Thomas H Cormen, Charles E Leiserson, Ronald L Rivest, and Clifford Stein. 2009. *Introduction to algorithms*. MIT press.
[8] Hongbo Deng, Jiawei Han, Bo Zhao, Yintao Yu, and Cindy Xide Lin. 2011. Probabilistic topic models with biased propagation on heterogeneous information networks. In *KDD*. ACM, 1271–1279.
[9] Yuxiao Dong, Nitesh V Chawla, and Ananthram Swami. 2017. metapath2vec: Scalable representation learning for heterogeneous networks. In *KDD*. ACM, 135–144.
[10] Federico Errica, Marco Podda, Davide Bacciu, and Alessio Micheli. 2019. A fair comparison of graph neural networks for graph classification. *arXiv preprint arXiv:1912.09893* (2019).
[11] Tao-yang Fu, Wang-Chien Lee, and Zhen Lei. 2017. Hin2vec: Explore meta-paths in heterogeneous information networks for representation learning. In *CIKM*. ACM, 1797–1806.
[12] Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Yoshua Bengio. 2016. *Deep learning*. Vol. 1. MIT press Cambridge.
[13] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. In *NIPS*. 2672–2680.
[14] Aditya Grover and Jure Leskovec. 2016. node2vec: Scalable feature learning for networks. In *KDD*. ACM, 855–864.
[15] Will Hamilton, Zhitao Ying, and Jure Leskovec. 2017. Inductive representation learning on large graphs. In *NIPS*. 1024–1034.
[16] F Maxwell Harper and Joseph A Konstan. 2016. The movielens datasets: History and context. *TIIS* 5, 4 (2016), 19.
[17] Binbin Hu, Yuan Fang, and Chuan Shi. 2019. Adversarial Learning on Heterogeneous Information Networks. In *KDD*. ACM.
[18] Rana Hussein, Dingqi Yang, and Philippe Cudré-Mauroux. 2018. Are Meta-Paths Necessary?: Revisiting Heterogeneous Graph Embeddings. In *CIKM*. ACM, 437–446.
[19] Thomas N Kipf and Max Welling. 2016. Variational graph auto-encoders. *arXiv preprint arXiv:1611.07308* (2016).
[20] David Liben-Nowell and Jon Kleinberg. 2007. The link-prediction problem for social networks. *JASIST* 58, 7 (2007), 1019–1031.
[21] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *NIPS*. 3111–3119.
[22] Tore Opsahl and Pietro Panzarasa. 2009. Clustering in weighted networks. *Social networks* 31, 2 (2009), 155–163.
[23] Mingdong Ou, Peng Cui, Jian Pei, Ziwei Zhang, and Wenwu Zhu. 2016. Asymmetric transitivity preserving graph embedding. In *KDD*. ACM, 1105–1114.
[24] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. 2014. Deepwalk: Online learning of social representations. In *KDD*. ACM, 701–710.
[25] David Martin Powers. 2011. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. (2011).
[26] Benjamin Recht, Christopher Re, Stephen Wright, and Feng Niu. 2011. Hogwild: A lock-free approach to parallelizing stochastic gradient descent. In *NIPS*. 693–701.
[27] Chuan Shi, Binbin Hu, Xin Zhao, and Philip Yu. 2018. Heterogeneous Information Network Embedding for Recommendation. *TKDE* (2018).
[28] Yizhou Sun and Jiawei Han. 2012. Mining heterogeneous information networks: principles and methodologies. *Synthesis Lectures on Data Mining and Knowledge Discovery* 3, 2 (2012), 1–159.
[29] Yizhou Sun, Jiawei Han, Xifeng Yan, Philip S Yu, and Tianyi Wu. 2011. Pathsim: Meta path-based top-k similarity search in heterogeneous information networks. *VLDB* 4, 11 (2011), 992–1003.
[30] Yizhou Sun, Brandon Norick, Jiawei Han, Xifeng Yan, Philip S Yu, and Xiao Yu. 2013. Pathselclus: Integrating meta-path selection with user-guided object clustering in heterogeneous information networks. *TKDD* 7, 3 (2013), 11.
[31] Jian Tang, Meng Qu, and Qiaozhu Mei. 2015. Pte: Predictive text embedding through large-scale heterogeneous text networks. In *KDD*. ACM, 1165–1174.
[32] Jian Tang, Meng Qu, Mingzhe Wang, Ming Zhang, Jun Yan, and Qiaozhu Mei. 2015. Line: Large-scale information network embedding. In *WWW*. 1067–1077.
[33] Lei Tang and Huan Liu. 2009. Relational learning via latent social dimensions. In *KDD*. ACM, 817–826.
[34] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *NIPS*. 5998–6008.
[35] Nguyen Xuan Vinh, Julien Epps, and James Bailey. 2010. Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance. *JMLR* 11, Oct (2010), 2837–2854.
[36] Daixin Wang, Peng Cui, and Wenwu Zhu. 2016. Structural deep network embedding. In *KDD*. ACM, 1225–1234.
[37] Suhang Wang, Jiliang Tang, Charu Aggarwal, Yi Chang, and Huan Liu. 2017. Signed network embedding in social media. In *SDM*. SIAM, 327–335.
[38] Yu Zhao, Zhiyuan Liu, and Maosong Sun. 2015. Representation learning for measuring entity relatedness with rich information. In *IJCAI*.