the same time[4–7, 11]. However, we find we cannot directly adopt them into our problem setting with two main reasons. First, we need a unified framework that can handle both process-level and network-level events together. Second, according to the scale of our data, it becomes extremely difficult to run on a matrix/graph structure with millions of columns/nodes.

7 CONCLUSION

In this paper, we propose a unified optimization framework that can tackle two problems in the domain of enterprise information networks — host community detection and host anomaly assessment. Our perspectives in both tasks are based on host behaviors. This particular domain comes up with unique data characteristics, new community formulation, and great challenges of community detection and anomaly assessment. We propose an embedding-based model to investigate intricate behavioral patterns of each host purely based on their historical events. Empirical studies on real enterprise information networks show our proposed model can effectively identify host communities and assess host behavioral anomaly status, and outperform other popular community detection methods. An interesting direction for further exploration would be applying the proposed framework to other applications (such as social networks) and tasks (such as root cause analysis).

REFERENCES

[1] Ting Chen, Lu-An Tang, Yizhou Sun, Zhengzhang Chen, Haifeng Chen, and Guofei Jiang. 2016. Integrating community and role detection in information networks. In *Proceedings of the 2016 SIAM International Conference on Data Mining*. SIAM, 72–80.

[2] Ting Chen, Lu An Tang, Yizhou Sun, Zhengzhang Chen, and Kai Zhang. 2016. Entity embedding-based anomaly detection for heterogeneous categorical events. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, IJCAI 2016, New York, NY, USA, 9-15 July 2016*. 1396–1403.

[3] Zhengzhang Chen. 2012. *Discovery of informative and predictive patterns in dynamic networks of complex systems*. Ph.D. Dissertation. North Carolina State University.

[4] Zhengzhang Chen, William Hendrix, Hang Guan, Isaac K. Tetteh, Alok N. Choudhary, Fredrick H. M. Semazzi, and Nagiza F. Samatova. 2013. Discovery of extreme events-related communities in contrasting groups of physical system networks. *Data Mining and Knowledge Discovery* 27, 2 (2013), 225–258.

[5] Zhengzhang Chen, William Hendrix, and Nagiza F Samatova. 2012. Community-based anomaly detection in evolutionary networks. *Journal of Intelligent Information Systems* 39, 1 (2012), 59–85.

[6] Zhengzhang Chen, Kanchana Padmanabhan, Andrea M Rocha, Yekaterina Shpanskaya, James R Mihelcic, Kathleen Scott, and Nagiza F Samatova. 2012. SPICE: Discovery of phenotype-determining component interplays. *BMC Systems Biology* 6, 1 (2012), 40.

[7] Zhengzhang Chen, Kevin A. Wilson, Ye Jin, William Hendrix, and Nagiza F. Samatova. 2010. Detecting and tracking community dynamics in evolutionary networks. In *Proceedings of the 2010 IEEE International Conference on Data Mining Workshops (ICDMW'10)*. 318–327.

[8] Aaron Clauset, Mark EJ Newman, and Cristopher Moore. 2004. Finding community structure in very large networks. *Physical Review E* 70, 6 (2004), 066111.

[9] Boxiang Dong, Zhengzhang Chen, Hui (Wendy) Wang, Lu-An Tang, Kai Zhang, Ying Lin, Zhichun Li, and Haifeng Chen. 2017. Efficient discovery of abnormal event sequences in enterprise security systems. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management (CIKM'17)*. 707–715.

[10] Santo Fortunato. 2010. Community detection in graphs. *Physics Reports* 486, 3 (2010), 75–174.

[11] Jing Gao, Feng Liang, Wei Fan, Chi Wang, Yizhou Sun, and Jiawei Han. 2010. On community outliers and their efficient detection in information networks. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 813–822.

[12] Michael Gutmann and Aapo Hyvärinen. 2010. Noise-contrastive estimation: A new estimation principle for unnormalized statistical models.. In *AISTATS*, Vol. 1. 6.

[13] Greg Hamerly and Charles Elkan. 2004. Learning the k in k-means. In *Advances in neural information processing systems*. 281–288.

[14] Steven A Hofmeyr, Stephanie Forrest, and Anil Somayaji. 1998. Intrusion detection using sequences of system calls. *Journal of Computer Security* 6, 3 (1998), 151–180.

[15] Zhiting Hu, Poyao Huang, Yuntian Deng, Yingkai Gao, and Eric Xing. 2015. Entity hierarchy embedding. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, Vol. 1. 1292–1300.

[16] Tsuyoshi Idé and Hisashi Kashima. 2004. Eigenspace-based anomaly detection in computer systems. In *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 440–449.

[17] David J. Ketchen and Christopher L. Shook. 1996. The application of cluster analysis in strategic management research: an analysis and critique. *Strategic Management Journal* 17, 6 (1996), 441–458.

[18] Yan Liu, Alexandru Niculescu-Mizil, and Wojciech Gryc. 2009. Topic-link LDA: joint models of topic and author community. In *Proceedings of the 26th International Conference on Machine Learning*. ACM, 665–672.

[19] Chen Luo, Zhengzhang Chen, Lu-An Tang, Anshumali Shrivastava, Zhichun Li, Haifeng Chen, and Jieping Ye. 2018. TINET: Learning invariant networks via knowledge transfer. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD'18)*. 1890–1899.

[20] Christopher D Manning, Prabhakar Raghavan, Hinrich Schütze, et al. 2008. *Introduction to information retrieval*. Vol. 1. Cambridge university press Cambridge.

[21] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems*. 3111–3119.

[22] Darren Mutz, Fredrik Valeur, Giovanni Vigna, and Christopher Kruegel. 2006. Anomalous system call detection. *ACM Transactions on Information and System Security (TISSEC)* 9, 1 (2006), 61–93.

[23] Mark EJ Newman. 2004. Fast algorithm for detecting community structure in networks. *Physical review E* 69, 6 (2004), 066133.

[24] Mark EJ Newman. 2006. Finding community structure in networks using the eigenvectors of matrices. *Physical Review E* 74, 3 (2006), 036104.

[25] Caleb C Noble and Diane J Cook. 2003. Graph-based anomaly detection. In *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 631–636.

[26] Pascal Pons and Matthieu Latapy. 2005. Computing communities in large networks using random walks. In *International Symposium on Computer and Information Sciences*. Springer, 284–293.

[27] Peter J Rousseeuw. 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics* 20 (1987), 53–65.

[28] Yiye Ruan, David Fuhry, and Srinivasan Parthasarathy. 2013. Efficient community detection in large networks using content and links. In *Proceedings of the 22nd International Conference on World Wide Web*. ACM, 1089–1098.

[29] Jian Tang, Meng Qu, Mingzhe Wang, Ming Zhang, Jun Yan, and Qiaozhu Mei. 2015. Line: Large-scale information network embedding. In *Proceedings of the 24th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 1067–1077.

[30] Robert Tibshirani, Guenther Walther, and Trevor Hastie. 2001. Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63, 2 (2001), 411–423.

[31] Nguyen Xuan Vinh, Julien Epps, and James Bailey. 2010. Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance. *Journal of Machine Learning Research* 11, Oct (2010), 2837–2854.

[32] Joyce Jiyoung Whang, David F Gleich, and Inderjit S Dhillon. 2013. Overlapping community detection using seed set expansion. In *Proceedings of the 22nd ACM International Conference on Information & Knowledge Management*. ACM, 2099–2108.

[33] Zhiqiang Xu, Yiping Ke, Yi Wang, Hong Cheng, and James Cheng. 2012. A model-based approach to attributed graph clustering. In *Proceedings of the ACM SIGMOD International Conference on Management of Data*. ACM, 505–516.

[34] Jaewon Yang, Julian McAuley, and Jure Leskovec. 2013. Community detection in networks with node attributes. In *Proceedings of the IEEE International Conference on Data Mining*. IEEE, 1151–1156.

[35] Tianbao Yang, Rong Jin, Yun Chi, and Shenghuo Zhu. 2009. Combining link and content for community detection: a discriminative approach. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 927–936.

[36] Xi Zhang, Jian Cheng, Ting Yuan, Biao Niu, and Hanqing Lu. 2013. TopRec: domain-specific recommendation through community topic mining in social network. In *Proceedings of the 22nd International Conference on World Wide Web*. ACM, 1501–1510.

[37] Yang Zhou, Hong Cheng, and Jeffrey Xu Yu. 2009. Graph clustering based on structural/attribute similarities. *Proceedings of the 35th International Conference on Very Large Data Bases* 2, 1 (2009), 718–729.