THEOREM 3.5. *Let $\|\cdot\|_F$ be the matrix Frobenius norm. Then*

$$\left\|\boldsymbol{M} - \hat{\boldsymbol{M}}\right\|_F \le \frac{\mathrm{vol}(G)}{b d_{\min}} \sqrt{\sum_{j=k+1}^{n} \left| \frac{1}{T} \sum_{r=1}^{T} \lambda_j^r \right|^2};$$

$$\left\|\log \boldsymbol{M}' - \log \hat{\boldsymbol{M}}'\right\|_F \le \left\|\boldsymbol{M}' - \hat{\boldsymbol{M}}'\right\|_F \le \left\|\boldsymbol{M} - \hat{\boldsymbol{M}}\right\|_F.$$

PROOF. The first inequality can be seen by applying the definition of Frobenius norm and Eq. 10.

For the second inequality, first to show $\left\|\log \boldsymbol{M}' - \log \hat{\boldsymbol{M}}'\right\|_F \le \left\|\boldsymbol{M}' - \hat{\boldsymbol{M}}'\right\|_F$. According to the definition of Frobenius norm, sufficient to show $\left|\log \boldsymbol{M}'_{i,j} - \log \hat{\boldsymbol{M}}'_{i,j}\right| \le \left|\hat{\boldsymbol{M}}'_{i,j} - \boldsymbol{M}'_{i,j}\right|$ for any i, j. Without loss of generality, assume $\boldsymbol{M}'_{i,j} \le \hat{\boldsymbol{M}}'_{i,j}$.

$$\begin{aligned}\left|\log \boldsymbol{M}'_{i,j} - \log \hat{\boldsymbol{M}}'_{i,j}\right| &= \log \frac{\hat{\boldsymbol{M}}'_{i,j}}{\boldsymbol{M}'_{i,j}} = \log\left(1 + \frac{\hat{\boldsymbol{M}}'_{i,j} - \boldsymbol{M}'_{i,j}}{\boldsymbol{M}'_{i,j}}\right) \\ &\le \frac{\hat{\boldsymbol{M}}'_{i,j} - \boldsymbol{M}'_{i,j}}{\boldsymbol{M}'_{i,j}} \le \hat{\boldsymbol{M}}'_{i,j} - \boldsymbol{M}'_{i,j} = \left|\hat{\boldsymbol{M}}'_{i,j} - \boldsymbol{M}'_{i,j}\right|,\end{aligned}$$

where the first inequality is because $\log(1+x) \le x$ for $x \ge 0$, and the the second inequality is because $\boldsymbol{M}'_{i,j} = \max\left(\boldsymbol{M}_{i,j}, 1\right) \ge 1$. Next to show $\left\|\boldsymbol{M}' - \hat{\boldsymbol{M}}'\right\|_F \le \left\|\boldsymbol{M} - \hat{\boldsymbol{M}}\right\|_F$. Sufficient to show $\left|\boldsymbol{M}'_{i,j} - \hat{\boldsymbol{M}}'_{i,j}\right| \le \left|\boldsymbol{M}_{i,j} - \hat{\boldsymbol{M}}_{i,j}\right|$ for any i, j. Recall the definition of $\boldsymbol{M}'$ and $\hat{\boldsymbol{M}}'$, we get $\left|\boldsymbol{M}'_{i,j} - \hat{\boldsymbol{M}}'_{i,j}\right| = \left|\max(\boldsymbol{M}_{i,j}, 1) - \max(\hat{\boldsymbol{M}}_{i,j}, 1)\right| \le \left|\boldsymbol{M}_{i,j} - \hat{\boldsymbol{M}}_{i,j}\right|$. □

REFERENCES

[1] Sanjeev Arora, Yuanzhi Li, Yingyu Liang, Tengyu Ma, and Andrej Risteski. 2016. A latent variable model approach to pmi-based word embeddings. *TACL* 4 (2016), 385–399.
[2] Yoshua Bengio, Aaron Courville, and Pierre Vincent. 2013. Representation learning: A review and new perspectives. *IEEE TPAMI* 35, 8 (2013), 1798–1828.
[3] Austin R Benson, David F Gleich, and Jure Leskovec. 2015. Tensor spectral clustering for partitioning higher-order network structures. In *SDM*. SIAM, 118–126.
[4] Austin R Benson, David F Gleich, and Lek-Heng Lim. 2017. The Spacey Random Walk: A stochastic Process for Higher-Order Data. *SIAM Rev.* 59, 2 (2017), 321–345.
[5] Matthew Brand and Kun Huang. 2003. A unifying theorem for spectral embedding and clustering.. In *AISTATS*.
[6] Shaosheng Cao, Wei Lu, and Qiongkai Xu. 2015. GraRep: Learning graph representations with global structural information. In *CIKM*. ACM, 891–900.
[7] Shaosheng Cao, Wei Lu, and Qiongkai Xu. 2016. Deep Neural Networks for Learning Graph Representations.. In *AAAI*. 1145–1152.
[8] 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.
[9] Dehua Cheng, Yu Cheng, Yan Liu, Richard Peng, and Shang-Hua Teng. 2015. Efficient sampling for Gaussian graphical models via spectral sparsification. In *COLT*. 364–390.
[10] Kewei Cheng, Jundong Li, and Huan Liu. 2017. Unsupervised Feature Selection in Signed Social Networks. In *KDD*. ACM.
[11] Fan RK Chung. 1997. *Spectral graph theory*. Number 92. American Mathematical Soc.
[12] Yuxiao Dong, Nitesh V Chawla, and Ananthram Swami. 2017. metapath2vec: Scalable Representation Learning for Heterogeneous Networks. In *KDD*.
[13] David Easley and Jon Kleinberg. 2010. *Networks, Crowds, and Markets: Reasoning about a Highly Connected World*. Cambridge University Press.
[14] Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin. 2008. LIBLINEAR: A library for large linear classification. *JMLR* 9, Aug (2008), 1871–1874.
[15] David F Gleich, Lek-Heng Lim, and Yongyang Yu. 2015. Multilinear pagerank. *SIAM J. Matrix Anal. Appl.* 36, 4 (2015), 1507–1541.
[16] Aditya Grover and Jure Leskovec. 2016. node2vec: Scalable feature learning for networks. In *KDD*. ACM, 855–864.
[17] William L. Hamilton, Rex Ying, and Jure Leskovec. 2017. Inductive Representation Learning on Large Graphs. In *NIPS*.
[18] Tatsunori B Hashimoto, David Alvarez-Melis, and Tommi S Jaakkola. 2016. Word embeddings as metric recovery in semantic spaces. *TACL* 4 (2016), 273–286.
[19] Roger A. Horn and Charles R. Johnson. 1991. *Topics in Matrix Analysis*. Cambridge University Press. https://doi.org/10.1017/CBO9780511840371
[20] Yann Jacob, Ludovic Denoyer, and Patrick Gallinari. 2014. Learning latent representations of nodes for classifying in heterogeneous social networks. In *WSDM*. ACM, 373–382.
[21] Thomas N Kipf and Max Welling. 2016. Semi-Supervised Classification with Graph Convolutional Networks. *arXiv preprint arXiv:1609.02907* (2016).
[22] Richard B Lehoucq, Danny C Sorensen, and Chao Yang. 1998. *ARPACK users' guide: solution of large-scale eigenvalue problems with implicitly restarted Arnoldi methods*. SIAM.
[23] Jure Leskovec, Kevin J Lang, and Michael Mahoney. 2010. Empirical comparison of algorithms for network community detection. In *WWW*. ACM, 631–640.
[24] Omer Levy and Yoav Goldberg. 2014. Neural Word Embedding as Implicit Matrix Factorization. In *NIPS*. 2177–2185.
[25] Hang Li, Haozheng Wang, Zhenglu Yang, and Masato Odagaki. 2017. Variation Autoencoder Based Network Representation Learning for Classification. In *ACL*. 56.
[26] László Lovász. 1993. Random walks on graphs. *Combinatorics, Paul erdos is eighty* 2 (1993), 1–46.
[27] Qing Lu and Lise Getoor. 2003. Link-based Classification. In *ICML*.
[28] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781* (2013).
[29] 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.
[30] Mingdong Ou, Peng Cui, Jian Pei, Ziwei Zhang, and Wenwu Zhu. 2016. Asymmetric Transitivity Preserving Graph Embedding.. In *KDD*. 1105–1114.
[31] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. 2014. DeepWalk: Online learning of social representations. In *KDD*. ACM, 701–710.
[32] S Yu Philip, Jiawei Han, and Christos Faloutsos. 2010. *Link mining: Models, algorithms, and applications*. Springer.
[33] Jianbo Shi and Jitendra Malik. 2000. Normalized cuts and image segmentation. *IEEE PAMI* 22, 8 (2000), 888–905.
[34] A.N. Shiryaev and A. Lyasoff. 2012. *Problems in Probability*. Springer New York.
[35] Chris Stark, Bobby-Joe Breitkreutz, Andrew Chatr-Aryamontri, Lorrie Boucher, Rose Oughtred, Michael S Livstone, Julie Nixon, Kimberly Van Auken, Xiaodong Wang, Xiaoqi Shi, et al. 2010. The BioGRID interaction database: 2011 update. *Nucleic acids research* 39, suppl_1 (2010), D698–D704.
[36] Jian Tang, Meng Qu, and Qiaozhu Mei. 2015. PTE: Predictive text embedding through large-scale heterogeneous text networks. In *KDD*. ACM, 1165–1174.
[37] Jian Tang, Meng Qu, Mingzhe Wang, Ming Zhang, Jun Yan, and Qiaozhu Mei. 2015. LINE: Large-scale information network embedding. In *WWW*. 1067–1077.
[38] Lei Tang and Huan Liu. 2009. Relational learning via latent social dimensions. In *KDD*. ACM, 817–826.
[39] Lei Tang, Suju Rajan, and Vijay K Narayanan. 2009. Large scale multi-label classification via metalabeler. In *WWW*. ACM, 211–220.
[40] Kristina Toutanova, Dan Klein, Christopher D Manning, and Yoram Singer. 2003. Feature-rich part-of-speech tagging with a cyclic dependency network. In *NAACL*. Association for Computational Linguistics, 173–180.
[41] Lloyd N Trefethen and David Bau III. 1997. *Numerical linear algebra*. Vol. 50. Siam.
[42] Grigorios Tsoumakas, Ioannis Katakis, and Ioannis Vlahavas. 2009. Mining multi-label data. In *Data mining and knowledge discovery handbook*. Springer, 667–685.
[43] Cunchao Tu, Han Liu, Zhiyuan Liu, and Maosong Sun. 2017. CANE: Context-aware network embedding for relation modeling. In *ACL*.
[44] Cunchao Tu, Weicheng Zhang, Zhiyuan Liu, and Maosong Sun. 2016. Max-Margin DeepWalk: Discriminative Learning of Network Representation.. In *IJCAI*. 3889–3895.
[45] Ulrike Von Luxburg. 2007. A tutorial on spectral clustering. *Statistics and computing* 17, 4 (2007), 395–416.
[46] Daixin Wang, Peng Cui, and Wenwu Zhu. 2016. Structural deep network embedding. In *KDD*. ACM, 1225–1234.
[47] Cheng Yang, Zhiyuan Liu, Deli Zhao, Maosong Sun, and Edward Y Chang. 2015. Network Representation Learning with Rich Text Information. In *IJCAI*. 2111–2117.
[48] Zhilin Yang, William W. Cohen, and Ruslan Salakhutdinov. 2016. Revisiting Semi-Supervised Learning with Graph Embeddings. In *ICML*. 40–48.
[49] Zhilin Yang, Jie Tang, and William W Cohen. 2016. Multi-Modal Bayesian Embeddings for Learning Social Knowledge Graphs.. In *IJCAI*. 2287–2293.