Language models

The starting point of the work in the graph is based previous work in distributional semantic.

- The skip-gram model of Mikolov *et al.* [1] and some further improvements (subsampling of frequent words, phrasal words, negative sampling) [2]. A simplified explanation is given in a blog post
- GLOVE by Pennington *et al.* [3]
- PMI factorization of Levy *et al.* [4]
- Theoretical justifications of Arora *et al.* [5]
- Collobert? (actually his model are not exactly in the same but maybe I can extend Hellinger PCA[6] for node embeddings)
- is there a survey on that? (section 5 of `http://u.cs.biu.ac.il/~yogo/nnlp.pdf`, `http://www.grappa.univ-lille3.fr/~gilleron/WordToVector.html`)

Graph models

- deepwalk [7]
- HNE [8] (maybe that one does not use language model? at least not only, because it jointly handles multimedia contents)
- LINE [9]
- subgraph2vec??[10]
- node2vec[11]
- planetoid[12]
- [13]
- iclr17[14]

I should probably draw a table to summarize those methods. Among the various characteristics: using side info, able to deal with out-of-sample problem (ie new nodes added after learning), what kind of walk, use of weights, and so on. (for instance Table 1 in [10] and Table 1 in [12]).

Applications

- Mining location checkins[15] (they use external knowledge to weight the edges of the graph they built, in order for the random walk to be biased towards their prediction task. "This way of edge weight definition drives our random walks to be biased towards ρ biased temporal-locations, and users frequently visiting these ρ-biased locations, which in the end results in each user's neighbors being able to reflect his demographic information, i.e., the intuition of task-specific random walk.")
- communities discovery[16]

- named entity resolution https://arxiv.org/abs/1702.02287

Obviously here the application would be to customized walks when label have binary labels in order to learn a node representation that allow training classic methods to predict remaining signs.

Done already :(http://www.public.asu.edu/~swang187/publications/SiNE.pdf Seventeenth SIAM International Conference on Data Mining (SDM-17)

twice actually https://arxiv.org/abs/1702.06819

5 Conclusion

summarize the contributions and present further directions

References

[1] T. Mikolov *et al.*, "Efficient estimation of word representations in vector space", in *ICLR workshop*, 2013. [Online]. Available: http://arxiv.org/abs/1301.3781.

[2] T. Mikolov *et al.*, "Distributed representations of words and phrases and their compositionality", in *Advances in Neural Information Processing Systems 26*, 2013, pp. 3111–3119. [Online]. Available: http://papers.nips.cc/paper/5021-distributed-representations-of-words-and-phrases-and-their-compositionality.pdf.

[3] J. Pennington *et al.*, "Glove: Global vectors for word representation.", in *EMNLP*, vol. 14, 2014, pp. 1532–43. [Online]. Available: http://www.emnlp2014.org/papers/pdf/EMNLP2014162.pdf.

[4] O. Levy *et al.*, "Neural word embedding as implicit matrix factorization", in *Advances in Neural Information Processing Systems 27*, 2014, pp. 2177–2185. [Online]. Available: http://papers.nips.cc/paper/5477-neural-word-embedding-as-implicit-matrix-factorization.

[5] S. Arora *et al.*, "A latent variable model approach to pmi-based word embeddings", *Transactions of the Association for Computational Linguistics*, vol. 4, pp. 385–399, 2016. [Online]. Available: https://www.transacl.org/ojs/index.php/tacl/article/view/742.

[6] R. Lebret *et al.*, "Word embeddings through hellinger pca", in *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, 2014, pp. 482–490. [Online]. Available: http://ronan.collobert.com/pub/matos/2014_hellinger_eacl.pdf.

[7] B. Perozzi *et al.*, "Deepwalk: Online learning of social representations", in *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2014, pp. 701–710. DOI: 10.1145/2623330.2623732. [Online]. Available: http://doi.acm.org/10.1145/2623330.2623732.

[8] S. Chang *et al.*, "Heterogeneous network embedding via deep architectures", in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2015, pp. 119–128. DOI: 10.1145/2783258.2783296. [Online]. Available: http://doi.acm.org/10.1145/2783258.2783296.

[9] J. Tang *et al.*, "Line: Large-scale information network embedding", in *Proceedings of the 24th International Conference on World Wide Web*, 2015, pp. 1067–1077. DOI: 10.1145/2736277.2741093. [Online]. Available: https://doi.org/10.1145/2736277.2741093.

[10] A. Narayanan *et al.*, *Subgraph2vec: Learning distributed representations of rooted sub-graphs from large graphs*, 2016.

[11] A. Grover *et al.*, “Node2vec: Scalable feature learning for networks”, in *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 855–864. DOI: 10.1145/2939672.2939754. [Online]. Available: http://doi.acm.org/10.1145/2939672.2939754.

[12] Z. Yang *et al.*, “Revisiting semi-supervised learning with graph embeddings”, in *Proceedings of the 33rd International Conference on Machine Learning*, 2016. [Online]. Available: http://jmlr.org/proceedings/papers/v48/yanga16.pdf.

[13] S. Cao *et al.*, “Deep neural networks for learning graph representations”, in *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, 2016, pp. 1145–1152. [Online]. Available: http://www.aaai.org/ocs/index.php/AAAI/AAAI16/paper/viewFile/12423/11715.

[14] T. N. Kipf *et al.*, *Semi-supervised classification with graph convolutional networks*, 2016.

[15] J. Pang *et al.*, *Deepcity: A feature learning framework for mining location check-ins*, 2016.

[16] W. Ding *et al.*, *Node embedding via word embedding for network community discovery*, 2016.