

VilLain: Unsupervised Node Embedding on Hypergraphs via Virtual Label Propagation (Online Appendix)

Anonymous Author(s)

ABSTRACT

Groupwise interactions arise in various scenarios in real-world systems: collaborations of researchers, co-purchases of products, and discussions in online Q&A sites, to name a few. Such higher-order relations are naturally modeled as hypergraphs, which consist of hyperedges (i.e., any-sized subsets of nodes). A popular and promising approach for analyzing such a hypergraph is to represent the nodes in it as structure-preserving embeddings (i.e., low-dimensional vectors), which can be readily used for various downstream tasks, such as node classification and hyperedge prediction. Importantly, it is desirable that such embeddings can be obtained even with minimal requirements, precisely, without relying on labels, features, augmentations, and negative samples. However, previous methods resort to additional information or manipulation of the data, which can be problematic in realistic setups.

In this work, we propose VilLain, a novel hypergraph embedding method based on the propagation of virtual labels (v-labels). Aiming to reproduce higher-order homogeneity, which we observe in real-world hypergraphs, VilLain first learns, for each node, the probability distribution over v-labels. Then, VilLain obtains embeddings by propagating the distributions over the hypergraph. With additional schemes for automatic combination of multiple embeddings while reducing redundancy, VilLain is: **(a) Requirement-free:** not

requiring extra information (e.g., labels and features) or data manipulation (e.g., augmentation and negative sampling), **(b) Versatile:** giving embeddings that are not specialized to specific tasks but generalizable to diverse downstream tasks, **(c) Accurate:** being up to 64.8% and 27.5% more accurate than its competitors for node classification and hyperedge prediction tasks, respectively.

1 INTRODUCTION

2 RELATED WORK

3 PROBLEM STATEMENT

4 PROPOSED METHOD

5 EXPERIMENTAL RESULTS

6 CONCLUSIONS

REFERENCES

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CIKM '22, October 17–22, 2022, Atlanta, Georgia, USA

© 2022 Association for Computing Machinery.

ACM ISBN 978-1-4503-XXXX-X/18/06...\$15.00

<https://doi.org/10.1145/1122445.1122456>