Community Embeddings with Bayesian Gaussian Mixture Model and Variational Inference

Anton I. N. Begehr
Graduate School of Business
National Research University Higher School of Economics
Moscow, Russia
a.begehr@fu-berlin.de

Prof. Dr. Petr Panfilov
Graduate School of Business
National Research University Higher School of Economics
Moscow, Russia
ppanfilov@hse.ru

***Abstract*—Graphs, such as social networks, emerge naturally from various real-world situations. Recently, graph embedding methods have gained traction in data science research. The graph and community embedding algorithm ComE aims to preserve first-, second- and higher-order proximity. ComE requires prior knowledge of the number of communities K. In this paper, ComE is extended to utilize a Bayesian Gaussian mixture model with variational inference for learning community embeddings (ComE BGMM+VI), similar to ComE+. ComE BGMM+VI takes K as the maximum number of communities and drops components through the trade-off hyperparameter weight concentration prior. The advantage of ComE BGMM+VI over the non-Bayesian ComE for an unknown number of communities K is shown for the small Karate club dataset and explored for the larger DBLP dataset.**

***Index Terms*—graph, embedding, community embedding, ComE, Bayesian, variational inference, Gaussian mixture, expectation maximization**

I. Introduction

Graphs, such as social networks, knowledge graphs, content-rating graphs, and communication networks, emerge naturally from various real-world situations. Analyzing these graphs leads to findings and understanding of the underlying structures, coherences, and dependencies. Recently, methods for embedding graph's nodes into lower-dimensional Euclidean spaces, called graph embeddings, have gained traction in multiple areas of data science research.[**?**]

Community Embeddings, in addition to embedding a graph's nodes through first- and second-order proximity, also preserve higher-order proximity by embedding clusters present in the graph data. The graph and community embedding algorithm ComE aims to preserve first-, second- and higher-order proximity by embedding a graph's nodes and communities.[**?**] ComE requires prior knowledge of the number of communities K. In this paper, ComE is extended to utilizing a Bayesian Gaussian mixture model with variational inference for learning community embeddings (ComE BGMM+VI), similar to ComE+ published by **?** in **?** .[**?**] ComE BGMM+VI takes K as the maximum number of communities and drops components through a trade-off hyperparameter.

The reported study was partially supported by RFBR grant №20-07-00958. The paper was prepared within the framework of the HSE University Project Group Competition 2020-2022.

The recent **?** graph embeddings algorithm ComE is extended similarily to the **?** ComE+ by taking a Bayesian approach. The open-source code for the Bayesian approach to ComE's community embedding is published on GitHub[1] and serves as a contribution to community embedding research.[**?**]

The original ComE paper ComE by **?** and the ComE+ paper ComE+ by **?** in combination with the ComE source code have provided the basis and architecture for the community embeddings utilized in this work.[**? ? ?**]

Multiple surveys and articles on graph embeddings were consulted to build a full picture of the current state of graph embedding research. Especially the **?** survey $\text{Goyal}_2018by?andthe?paperrossi20tkdd - rolesby?haveproventobepri$

On part of comparing the Bayesian Gaussian mixture model with variational inference to the Gaussian mixture model with expectation-maximization, the **?** book Bishop06 by **?** includes essential statistics and information science knowledge and explanations.[**?**]

II. Community Embedding

[1] at ComE_BGMM_GH