Representation learning for social networks using Homophily based Latent Space Model

Pranav Nerurkar**
panerurkar_p16@ce.vjti.ac.in
Veermata Jijabai Technological Institute
Mumbai, Maharashtra

Madhav Chandane†
mmchandane@ce.vjti.ac.in
Veermata Jijabai Technological Institute
Mumbai, Maharashtra

Sunil Bhirud‡
sgbhirud@ce.vjti.ac.in
Veermata Jijabai Technological Institute
Mumbai, Maharashtra

ABSTRACT

Representing data in the form of a graph (network) is becoming an increasingly common approach for modelling complex systems. Graph representation models have thus emerged as a unified vocabulary across scientific domains. The advantage of graphs for visualization of systems is that graph theoretic literature can be used for their analysis. Therefore, systems from online social networking, economics, biology, internet, citation and e-commerce are being modelled as graphs (entities as nodes and relationships as edges). Inspite of many advantages in analysis and inferencing, there are several disadvantages associated with network models. To overcome these disadvantages, Network Representation Learning (NRL) has emerged as a popular solution.

Representative learning is utilized as conventional graph representation of a social network creates inter-dependency between the nodes of the system. Because of such inter-dependency it is not conceivable to straightforwardly apply machine learning strategies for their examination. To conquer such a disadvantage, network embedding frameworks represent a graph in a latent space where the coupling between nodes no longer exists. The present inquiry proposes an ERGM based model for making low dimensional graph representation of social network dependent on homophily (HLS). Such a model has favorable circumstances over other representative learning procedures explored in the literature as it very well may be utilized even on social networks with side information. This favorable position is exhibited on four data-sets that have categorical and real attributes. HLS model likewise has a second preferred standpoint that it is probabilistic model over current models which are heuristic inclined.

CCS CONCEPTS

• **Computer systems organization** → **Embedded systems**; *Redundancy*; Robotics; • **Networks** → Network reliability.

*All authors contributed equally to this research.

COINS, May 5–7, 2019, Crete, Greece

ACM ISBN 978-1-4503-6640-3/19/05…$15.00
https://doi.org/10.1145/3312614.3312627

KEYWORDS

representation learning, latent space models, social networks

ACM Reference Format:
Pranav Nerurkar, Madhav Chandane, and Sunil Bhirud. 2019. Representation learning for social networks using Homophily based Latent Space Model. In *INTERNATIONAL CONFERENCE ON OMNI-LAYER INTELLIGENT SYSTEMS (COINS), May 5–7, 2019, Crete, Greece.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3312614.3312627

1 INTRODUCTION

Representing genuine or synthetic frameworks as networks (graphs) has difficulties as strategies to investigate them depend on iterative, combinatorial advances or require computationally costly stochastic simulations [1, 2]. An answer for these issues could be the use of distributed or parallel models to enhance the handling and speed of calculation. Be that as it may, the social network graph representation leads to challenges in parallelization of calculations due to linkages between nodes (vertices). Probabilistic frameworks expect independent and identically distributed ($i.i.d$) observations in a vector space. Therefore machine learning too can not be applied to social network information as the information points (vertices) in them are combined with one another and henceforth do not satisfy the $i.i.d$ criteria [3]. Moreover, Graphs are complex topological structures with no spatial locality like grids, there is no fixed ordering or reference point in them and dynamic or multi-modal features may also be present in them. All these factors have to be tackled to make social network data amenable for analysis [4].

An answer for these issues is Network Representation Learning (NRL). Nework embeddings are portrayals of the networks in a low-dimensional vector space where the inborn coupling between the vertices of the system is no longer present. Once the coupling is removed the nodes in the graph resemble $i.i.d$ observations becoming suitable for machine learning. Hence, downstream network applications, for example, link prediction, clustering, node classification, network visualization perform better on vector embeddings over original social network data [5–8].

Before Network representation learning methods were proposed, social network analysis (SNA) was performed by means of extensive feature engineering. This required use of kernel functions or graph statistics which made the entire exercise data dependent. NRL frameworks eliminated the need for feature engineering and made SNA data independent. Network embedding methods are broadly classified as adjacency based [9–13], multi-hop based [14–20] and random walk based methods [21–26].