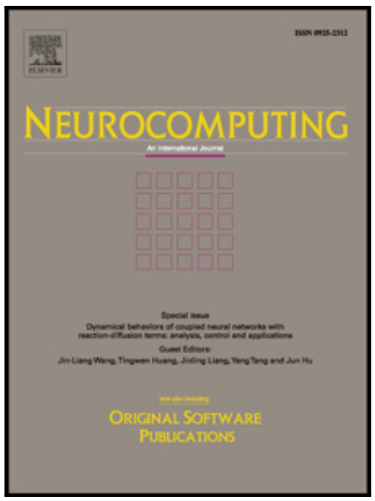

Incorporating network structure with node contents for community detection on large networks using deep learning

Jinxin Cao , Di Jin , Liang Yang , Jianwu Dang

PII: S0925-2312(18)30098-5
DOI: 10.1016/j.neucom.2018.01.065
Reference: NEUCOM 19267

To appear in: *Neurocomputing*

Received date: 3 June 2017
Revised date: 19 December 2017
Accepted date: 28 January 2018

Please cite this article as: Jinxin Cao , Di Jin , Liang Yang , Jianwu Dang , Incorporating network structure with node contents for community detection on large networks using deep learning, *Neurocomputing* (2018), doi: 10.1016/j.neucom.2018.01.065

Incorporating network structure with node contents for community detection on large networks using deep learning

Jinxin Cao[1], Di Jin[1*], Liang Yang[1], and Jianwu Dang[1, 2]

[1]School of Computer Science and Technology, Tianjin University, Tianjin 300350, China

[2]School of Information Science, Japan Advanced Institute of Science and Technology, Ishikawa 923-1292, Japan

{alfred718china, jindi}@tju.edu.cn, yangliang@vip.qq.com, jdang@jaist.ac.jp

Abstract: Community detection is an important task in social network analysis. In community detection, in general, there exist two types of the models that utilize either network topology or node contents. Some studies endeavor to incorporate these two types of models under the framework of spectral clustering for a better community detection. However, it was not successful to obtain a big achievement since they used a simple way for the combination. To reach a better community detection, it requires to realize a seamless combination of these two methods. For this purpose, we re-examine the properties of the modularity maximization and normalized-cut models and fund out a certain approach to realize a seamless combination of these two models. These two models seek for a low-rank embedding to represent of the community structure and reconstruct the network topology and node contents respectively. Meanwhile, we found that autoencoder and spectral clustering have a similar framework in their low- rank matrix reconstruction. Based on this property, we proposed a new approach to seamlessly combine the models of modularity and normalized-cut via the autoencoder. The proposed method also utilized the advantages of the deep structure by means of deep learning. The experiment demonstrated that the proposed method can provide a nonlinearly deep representation for a large-scale network and reached an efficient community detection. The evaluation results showed that our proposed method outperformed the existing leading methods on nine real-world networks.

Keyword: community detection, deep learning, spectral clustering, modularity maximization, normalized-cut, node contents.

1 Introduction

The development of Internet has led to producing more and more variety of data, such as online comments, product reviews and co-author networks, which have affected all aspects of people's lives, and thus the analysis of those data has attracted more and more attention of researchers in various fields. One hot topic in the studies of such social media or online data is to discovery the underlying structure with group effect, which is the so-called community structure. The vertices (users) related to the communities in the network can be divided into groups, in which vertices have more multiple connection but the connections are relatively sparse in the whole network. Those individuals or users belonging to the same community share common profiles or have common interests. The identification of communities consisting of users with similarity is very important, and has been applied in many areas, *e.g.*, sociology, biology and computer science. For example, in biology, some different units

* Corresponding author.