Dear Reviewers and Technical Program Committees Members:

Thank you very much for your careful and valuable comments. We will explain your concerns point by point.

Reviewer#1:

Q1: This paper combines matrix factorization and gnn to solve the community detection problem. The approach is interesting, however, it is not clear why should these two approaches be combined. Any motivation of doing so will make the paper more appealing.

A1: Owing to spatial confinement in the manuscript, the motivation may not be emphasized enough. At first, we want to emphasize the motivation of this paper. Graph convolution neural network is an effective method for community discovery since it can fuse the topology information and the attribute information of nodes by graph convolution and obtain the representation of nodes at a high level. However, unlabeled graph data is very common in real applications. How to learn and discover community structure in a graph by the unsupervised learning method is an important issue. Self-supervised learning is an important unsupervised learning method for a neural network. We propose a self-supervised deep graph convolution neural network (DGCN-NMF) for community discovery. It is different for graph data from traditional structured data (e.g. image, etc.) that the adjacency matrix of a graph is very sparse and high-dimensional ($n \cdot n$). DGNN-NMF introduces residual structure to design a deep graph convolution neural network. From Fig.1 and Eq.(3), it is known that the sparse or high-dimensional matrix brings the following problems: (1). It brings much valueless or redundant information and the valueless or redundant information is propagated into every layer which degrades the performance of DGCN-NMF. (2). That the output matrix of each layer is concatenated with the sparse or high-dimensional matrix also results in more parameters of GNN for DGNN-NMF which increases the training difficulty of GNN. NMF is an effective matrix factorization method and it has many advantages, e.g. simplicity, good interpretability, less storage, etc. Therefore DGCN-NMF introduces NMF to solve the above problems.

In the revised manuscript, we will add the above description to further emphasize the motivation of this manuscript.

Reviewer#2:

Q1: Although the authors explain that Z of DGCN-NMF contains both the high-level and the low-level semantic information of G and A should be restructured from Z, the authors should explain the theoretical validity of the loss definition in eq. (4) in more detail.

Q1: We explain the meaning of Eq.(4) at first. Eq.(4) is the restruction loss of adjacency matrix. Z is the output of MLP with the size $n \cdot k$. Z can be seen as the low dimensional embedding vectors for the nodes of graph G. Therefore $Z \cdot Z^{T}\in \mathbb{R}^{n\times n}$ measures the similarity between any two nodes. A is the adjacency matrix of a graph and $A_{ij}=1$ means there is an edge between the *i*th node and the *j*th node. In community discovery, the similarity of two nodes is generally larger than that of nonadjacent nodes if there is an edge between the nodes. It can be seen that DGCN-NMF uses the embedding vectors of nodes to restruct the graph with minimum loss.

Q2: The authors should briefly explain the data sets.
A2: We do not show the detailed information of the data sets because of the limitation of layout in the previous version. However, we give the websites which list the resources and the descriptions of the data sets. In the revised version, we will add a table into the manuscript to list the main information of the experimental data sets.

Q3: The authors are recommended to explain the loss of the restriction in eq. (5) in more detail. It is difficult to understand the reason why the target matrix of the adjacency matrix and the attribute information is the same ZZ'.
A3: There are two aspects of information for a graph: the topology information and the attribute information of nodes (For example, in a social network, a node represents a person and an edge represents the relationship between two persons. There is some attribute information for each person, e.g. age, height, weight, income, expenses, etc.). In this paper, we propose a deep graph convolution neural network (DGCN-NMF) for community discovery. It is different from the traditional deep neural networks (e.g. CNN, LSTM, VGG-16, ResNet, etc.) that DGCN-NMF can fuse the topology information of a graph and the attribute information of nodes. In other words, the low dimensional embedding Z is the fusion of the topology and attribute information. It is known that there is no label for the nodes of the input graph. DGCN-NMF trains graph convolution neural networks by self-supervised learning. In Q3, we explain the meaning of eq. (4). Eq. (4) aims at restructing the topology information of the input graph with minimum loss. In the same way, we also expect to use the low dimensional embedding Z to restruct the attribute information X with minimum loss. $Z \cdot Z^{T}\in \mathbb{R}^{n\times n}$ measures the similarity of node pairs. However, the similarity between any two nodes can be also measured from the attribute information X. $X \cdot X^{T}\in \mathbb{R}^{n\times n}$ also measures the similarity of node pairs. Eq. (5) aims at restructing the attribute information with minimum loss. Therefore the loss function of DGCN-NMF expects to use the low dimensional embedding Z to restruct both the topology information and the attribute information.

Q4: The 6th and 9th lines in section 2.2: the loss the restruction -> the loss of the restruction.
A4: Thank you for your carefulness. We have revised the errors and avoid any grammatical and spelling errors in the revised manuscript.

Reviewer#3:
Q1: It is hard to follow eq-2 and eq-3. The H matrix is factored into U and V. Is V discarded afterwards? eq-3 seems problematic, is the H_l on the right-hand-side redundant?
A1: For a graph, U is a matrix with the size n*d and V is a matrix with the size d*n. U can be seen as the low dimensional embedding vectors of H and V can be seen as the context embedding vectors. However, the context is not the main concern of this paper. Therefore V is discarded.
We want to emphasize that there is no redundancy for Eq(3). Eq(3) is a key equation to update the output of the lth layer. DGCN-NMF introduces ResNet structure to design graph convolution neural network. After the lth layer outputs $H_{l}$, $H_{l}$ is updated by concatenating the outputs of the current and previous layers. The new matrix which concatenates $H_{0}, \cdots,$

H_{l}$ is copied to $H_{l}$ and it is as the input of the (l+1)th layer. Therefore the H_l on the right-hand-side is not redundant and Eq.(3) is correct.

Q2: (relatively minor) The motivation for eq-4 is unclear: since the learned Z is high-level representation of nodes, asking the ZZ^T to reconstruct the adjacent matrix seems odd, as adjacent matrix mainly captures low-level information (first-order connectivity of graph).
A2: This comment concerns the core of this paper. From Eq.(4), it restructs the adjacent matrix and seems that DGCN-NMF only captures low-level information. However, it is not true. DGCN-NMF restructs the adjacent matrix and does not mean it mainly captures low-level information. It is known that GCN is equal to a low-pass filter and can fuse the topology information of the graph and the attribute information of nodes. A deep GCN with $l$ layers can obtain the graph information with $l$ hops (Reference: arXiv:2009.14332). It means Z captures multi-hop information of the graph although Z is used to restruct the adjacent matrix. Therefore GCN does not mainly capture low-level information (first-order connectivity of graph) and it can also capture high-level information.

Q3: The writing of the paper should be polished. There are multiple incorrect or unnatural uses. For instance, in Introduction, “However, due to the non-structure, the high dimension...”non-sturctute should be replaced with a noun. “For \forall Hi (i=0, 1, 2...)”, “DGCN-NMF is to...” should probably be replaced as “DGCN-NMF aims at...”
A3: Thank you for your valuable suggestions. We have revised the irregular expressions in the revised version.

Q4: The motivation to use NMF should be discussed, why not SVD?
A4: Thank you for your valuable suggestion. The motivation refers to Q1 to reviewer#1. The reason why SVD is not introduced is that DGCN-NMF aims at preserving the real semantics of nodes in community discovery. There are two aspects of information for a graph: the topology information and the attribute information of nodes (For example, in a social network, a node represents a person and an edge represents the relationship between two persons. There is some attribute information for a person, e.g. age, height, weight, income, expenses, etc.). It is obvious that each element of the adjacency matrix is nonnegative and the attribute information of nodes is also always nonnegative. The factorization of NMF can preserve nonnegative, but SVD cannot ensure nonnegative factorization. Therefore we introduce NMF not SVD into DGCN-NMF.

Q5: The authors should give a concise description about the dataset used in the paper. A web link is not sufficient: the website can become inaccessible, or the reader may not always have internet access.
A5: The response to the descriptions of the datasets refers to A2 to reviewer#2. In this paper, we give three links and the inaccessible link is updated.
Code: https://github.com/hmliangliang/DGCN-NMF
Data sets: http://memetracker.org/data/index.html
and http://www-personal.umich.edu/~mejn/netdata/
All websites are accessible in the revised version.