

# Multi-View Attribute Graph Convolution Networks for Clustering

Jiafeng Cheng<sup>1</sup>, Qianqian Wang<sup>1\*</sup>, Zhiqiang Tao<sup>2</sup>, Deyan Xie<sup>1</sup> and Quanxue Gao<sup>1,3†</sup>

<sup>1</sup>State Key Laboratory of Integrated Services Networks, Xidian University

<sup>2</sup>Northeastern University

<sup>3</sup>Unmanned System Research Institute, Northwestern Polytechnical University

xd.jiafengcheng@gmail.com, qianqian174@foxmail.com, {zqtaomail, weisong1220}@gmail.com, qxgao@xidian.edu.cn

## 1. 论文背景

图神经网络（GNNs）在处理图结构数据方面取得了相当大的成就。然而：

- (1) 现有的方法不能将可学习的权值分配给邻域内的不同节点，
- (2) 并且由于同时忽略了节点属性和图重构，缺乏鲁棒性。
- (3) 对于不同视图之间的一致性关系，没有明确地考虑相似性距离度量。
- (4) 此外，现有的多视图 GNN 方法主要关注多图的情况，而忽略了同样重要的属性多样性，也就是设计用以解决多视图属性的图结构数据的 GNN 仍未得到充分的探索。

因此，论文提出了一种新的多视图属性图卷积网络（MAGCN）模型中的聚类任务。MAGCN 设计了双路径编码器，用于映射图的嵌入特征和学习视图一致性信息。具体来说，第一种路径开发了多视图属性图注意网络，以减少噪声/冗余性，并学习多视图图数据的图嵌入特征。第二种路径开发一致的嵌入编码器来捕获不同视图之间的几何关系和概率分布的一致性，自适应地为多视图吸引找到一致的聚类嵌入空间。

## 2. 模型框架

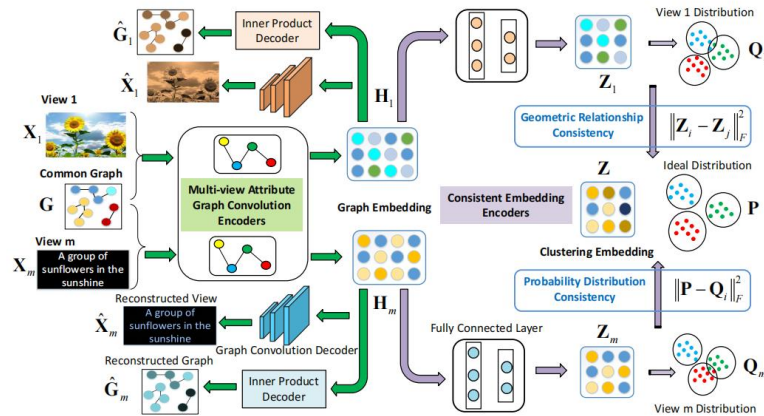


Figure 2: The framework of Multi-View Attribute Graph Convolution Networks for Clustering (MAGCN). It contains two key components: 1) Multi-view attribute graph convolution encoders with attention mechanism: they are used to learn graph embedding from node attribute and graph data. Attribute and graph reconstruction are executed for end-to-end learning. 2) Consistent embedding encoders further obtain a consistent clustering embedding among multiple views through geometric relationship and probability distribution consistency.

### 3. 方法介绍

#### 总体框架

模型包含两个主要模块：多视图属性图卷积编码器和一致的嵌入编码器。首先通过多视图属性图卷积编码器将多视图图数据  $X$  编码为图嵌入  $H$ 。然后将图嵌入  $H$  输入到的一致嵌入编码器，得到一致的聚类嵌入  $Z$ ，最终在  $Z$  计算的理想嵌入内在描述空间上进行聚类。

#### 多视图属性图卷积编码器

对于多视图中的任意一视图，过程可图示化为：

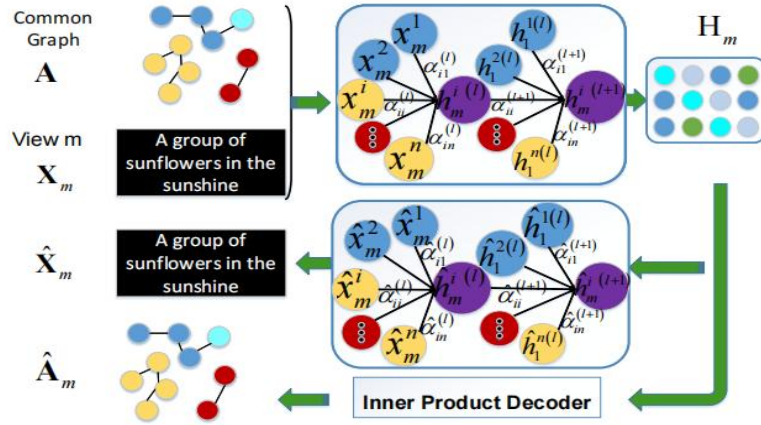


Figure 1: Part of Multi-view Attribute Graph Convolution Encoder for view  $m$ .

#### 编码

- 节点属性（同一般 GCN 一样）

$$\mathbf{H}_m^{(l)} = \sigma \left( \mathbf{D}^{-\frac{1}{2}} \mathbf{G}' \mathbf{D}^{-\frac{1}{2}} \mathbf{H}_m^{(l-1)} \mathbf{W}^{(l)} \right)$$

- 图结构

该论文为了更好的为自身节点和邻域节点分配可学习的权重，在节点之间使用了带有共享参数的注意力机制。对于第  $l$  层编码器，可学习的相关性矩阵定义

$$\mathbf{S} = \varphi \left( \mathbf{G} \odot t_s^{(l)} \mathbf{H}_m^{(l)} \mathbf{W}^{(l)} + \mathbf{G} \odot t_r^{(l)} \mathbf{H}_m^{(l)} \mathbf{W}^{(l)} \right)$$

⊙是指具有广播能力的元素级乘法， $\phi$ 表示激活函数，通常设置为s型激活函数

(即  $s(x) = (\frac{1}{1 + \exp^{-x}})$ ) ,  $t_s^{(l)}, t_n^{(l)} \in R^{1 \times d_l}$  分别表示与它们自身节点和邻居节点相关的

的可训练参数。

将  $\mathbf{S}$  归一化得到最终的相关系数  $\mathbf{G}$ ，因此  $G_{ij}$  为

$$\mathbf{G}_{ij} = \frac{\exp(\mathbf{S}_{ij})}{\sum_{k \in \mathbf{N}_i} \exp(\mathbf{S}_{ik})}$$

$\mathbf{N}_i$  是与节点  $i$  相邻的所有节点的集合。

**解码**（属性节点特征的重构和结构信息的重构）

#### • 节点属性矩阵的重构

在解码过程中，使用相同数量的层作为解码器的编码器，每个解码器层都试图反转其相应的编码器层。解码过程就是编码过程的逆过程。

$$\hat{\mathbf{H}}_m^{(l-1)} = \sigma \left( \hat{\mathbf{D}}^{-\frac{1}{2}} \hat{\mathbf{G}}' \hat{\mathbf{D}}^{-\frac{1}{2}} \hat{\mathbf{H}}_m^{(l)} \hat{\mathbf{W}}^{(l)} \right)$$

#### • 图结构的重构

重构的图结构由内积编码器实现

$$\hat{\mathbf{G}}_m^{ij} = \phi(-h_m^i \top h_m^j)$$

$h_m^i, h_m^j$  是图嵌入特征的节点  $i$  和  $j$ ， $\phi(\cdot)$  是内积函数。

面向多视图数据，其重构损失应该对应于每个视图，包含节点属性特征重构和结构信息重构，可以将其计算如下：

$$\mathcal{L}_{re} = \min_{\theta} \sum_{i=1}^M \left\| \mathbf{X}_i - \hat{\mathbf{X}}_i \right\|_F^2 + \lambda_1 \sum_{i=1}^M \left\| \mathbf{G} - \hat{\mathbf{G}}_i \right\|_F^2$$

$\theta$  是是多视图属性图卷积编码器的网络参数。

## 一致的嵌入编码器

对于视图  $m$ ，首先采用非线性特征提取映射进行图嵌入  $H_m$ 。 $H_m$  被映射到低维空间  $Z_m$ 。映射函数为  $g_m(H_m; \eta) \rightarrow Z_m$ ，其中  $\eta$  表示编码器参数。使用一致的聚类层来学习由所有  $Z_m$  自适应集成的公共聚类嵌入特征  $Z = \sum \beta_i Z_i$  作为是低维特征空间上的自适应视图融合。

本论文使用了两种一致性约束方式：

### • 几何关系一致性

假设  $Z_m, Z_b$  是由一致的嵌入编码器得到的视图  $m$  和  $b$  的低维空间特征矩阵。用它们来计算一个几何关系的相似度分数，如  $si(Z_m, Z_b)$ ，其中  $si(\cdot)$  是一个相似度函数。取最简单的差分相似度函数，则取几何关系的一致性  $L_{geo}$ ：

$$L_{geo} = \min_{\eta} \sum_{i \neq j}^M \|Z_i - Z_j\|_F^2$$

$\eta$  是编码器的网络参数

### • 概率分布一致性

使用学生的  $t$ -分布，作为一个核来度量集成节点表示  $z_i$  和质心  $\mu_j$  之间的相似性。因此， $Z$  的原始概率分布  $Q$  为

$$q_{ij} = \frac{(1 + \|z_i - \mu_j\|^2 / \alpha)^{-\frac{\alpha+1}{2}}}{\sum_{j'} (1 + \|z_i - \mu_{j'}\|^2 / \alpha)^{-\frac{\alpha+1}{2}}}$$

$\{\mu_j\}_{j=1}^k$  是  $k$  个初始聚类质心， $\alpha$  是学生  $t$ -分布的自由度， $q_{ij}$  是将节点  $i$  分配给聚类  $j$  的概率。将  $q$  提高到二次幂，并与每簇的频率归一化，得到如下

$$p_{ij} = \frac{q_{ij}^2/f_i}{\sum_{j'} q_{ij'}^2/f_{j'}}$$

$f_j = \sum_i q_{ij}$  是软簇频率。据此，将目标定义为概率分布一致性损失在  $Z_m$  的软分配

$Q_m$  和  $Z$  的辅助分布  $P$  如下：

$$\mathcal{L}_{pro} = \min_{\eta} \sum_{m=1}^M \rho_m \|\mathbf{Q}_m - \mathbf{P}\|_F^2$$

总损失（总优化目标）

$$\mathcal{L} = \min_{g, c, \mathbf{P}} \mathcal{L}_{re} + \lambda_2 \mathcal{L}_{geo} + \lambda_3 \mathcal{L}_{pro}$$

优化总体损失  $\mathcal{L}$ ，从聚类嵌入特征  $Z$  中学习辅助分布  $\mathbf{P}$ ，然后从辅助分布  $\mathbf{P}$  中预测每个节点的聚类。对于节点  $i$ ，其聚类可以用  $p_i$  计算，其中概率值最高的指数是  $i$  的聚类。因此，可以得到节点  $i$  的聚类标签为：

$$y_i = \arg \max_k (p_{ik})$$

## 4. 实验分析

### 数据集选取

| Database | Attribute1 | Attribute2 | Classes | Nodes  | Edges  |
|----------|------------|------------|---------|--------|--------|
| Cora     | 1,433      | 2,708      | 7       | 2,708  | 5,429  |
| Citeseer | 3,703      | 3,327      | 6       | 3,327  | 4,732  |
| Pubmed   | 500        | 19,717     | 3       | 19,717 | 44,438 |

评价指标：ACC, NMI, ARI



实验结果

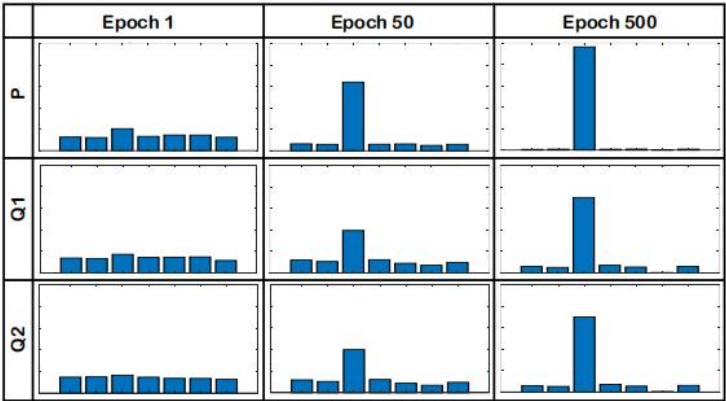
• 使用比较算法的评价指标

| Database                                  | Info.  | Cora  |       |       | Citeseer |       |       | Pubmed |       |       |
|---|--|-------|-------|-------|----------|-------|-------|--------|-------|-------|
| Metric                                    |  | ACC   | NMI   | ARI   | ACC      | NMI   | ARI   | ACC    | NMI   | ARI   |
| K-means [MacQueen, 1967]                  | $\mathbf{X}_1$                               | 0.500 | 0.317 | 0.239 | 0.544    | 0.312 | 0.285 | 0.580  | 0.278 | 0.246 |
| Graph Encoder [Tian <i>et al.</i> , 2014] | $\mathbf{G}$                                 | 0.301 | 0.059 | 0.046 | 0.293    | 0.057 | 0.043 | 0.531  | 0.210 | 0.184 |
| Deep Walk [Perozzi <i>et al.</i> , 2014]  | $\mathbf{G}$                                 | 0.529 | 0.384 | 0.291 | 0.390    | 0.131 | 0.137 | 0.647  | 0.238 | 0.255 |
| DNGR [Cao <i>et al.</i> , 2016]           | $\mathbf{G}$                                 | 0.419 | 0.318 | 0.142 | 0.326    | 0.180 | 0.043 | 0.468  | 0.153 | 0.059 |
| M-NMF [Wang <i>et al.</i> , 2017b]        | $\mathbf{G}$                                 | 0.423 | 0.256 | 0.161 | 0.336    | 0.099 | 0.070 | 0.470  | 0.084 | 0.058 |
| DCCA [Andrew <i>et al.</i> , 2013]        | $\mathbf{X}_1 \& \mathbf{X}_2$               | 0.436 | 0.214 | 0.160 | 0.450    | 0.221 | 0.204 | 0.443  | 0.048 | 0.040 |
| DCCA [Andrew <i>et al.</i> , 2013]        | $\mathbf{G} \& \mathbf{X}_1 \& \mathbf{X}_2$ | 0.583 | 0.416 | 0.342 | 0.513    | 0.283 | 0.238 | 0.508  | 0.097 | 0.097 |
| DCCAE [Wang <i>et al.</i> , 2015]         | $\mathbf{X}_1 \& \mathbf{X}_2$               | 0.472 | 0.289 | 0.221 | 0.503    | 0.240 | 0.211 | 0.537  | 0.122 | 0.092 |
| DCCAE [Wang <i>et al.</i> , 2015]         | $\mathbf{G} \& \mathbf{X}_1 \& \mathbf{X}_2$ | 0.564 | 0.311 | 0.241 | 0.534    | 0.310 | 0.230 | 0.581  | 0.239 | 0.186 |
| GAE [Kipf and Welling, 2016]              | $\mathbf{G} \& \text{best} \mathbf{X}$       | 0.530 | 0.397 | 0.293 | 0.380    | 0.174 | 0.141 | 0.632  | 0.249 | 0.246 |
| VGAE [Kipf and Welling, 2016]             | $\mathbf{G} \& \text{best} \mathbf{X}$       | 0.592 | 0.408 | 0.347 | 0.392    | 0.163 | 0.101 | 0.619  | 0.216 | 0.201 |
| MGAE [Wang <i>et al.</i> , 2017a]         | $\mathbf{G} \& \text{best} \mathbf{X}$       | 0.684 | 0.511 | 0.448 | 0.661    | 0.412 | 0.414 | 0.593  | 0.282 | 0.248 |
| ARGAE [Pan <i>et al.</i> , 2018]          | $\mathbf{G} \& \text{best} \mathbf{X}$       | 0.640 | 0.449 | 0.352 | 0.573    | 0.350 | 0.341 | 0.681  | 0.276 | 0.291 |
| ARVGAE [Pan <i>et al.</i> , 2018]         | $\mathbf{G} \& \text{best} \mathbf{X}$       | 0.638 | 0.450 | 0.374 | 0.544    | 0.261 | 0.245 | 0.513  | 0.117 | 0.078 |
| DAEGC [Wang <i>et al.</i> , 2019]         | $\mathbf{G} \& \text{best} \mathbf{X}$       | 0.704 | 0.528 | 0.496 | 0.672    | 0.397 | 0.410 | 0.671  | 0.266 | 0.278 |
| GATE [Salehi and Davulcu, 2019]           | $\mathbf{G} \& \text{best} \mathbf{X}$       | 0.658 | 0.527 | 0.451 | 0.616    | 0.401 | 0.381 | 0.673  | 0.322 | 0.299 |
| MAGCN-view 1                              | $\mathbf{G} \& \mathbf{X}_1$                 | 0.710 | 0.553 | 0.476 | 0.698    | 0.418 | 0.403 | 0.683  | 0.321 | 0.310 |
| MAGCN-view 2                              | $\mathbf{G} \& \mathbf{X}_2$                 | 0.594 | 0.409 | 0.327 | 0.621    | 0.363 | 0.366 | 0.539  | 0.261 | 0.227 |
| MAGCN                                     | $\mathbf{G} \& \mathbf{X}_1 \& \mathbf{X}_2$ | 0.751 | 0.598 | 0.532 | 0.711    | 0.458 | 0.462 | 0.691  | 0.331 | 0.321 |

Table 2: Clustering results of various methods on three databases. Best results are highlighted in red and the suboptimal results are marked in blue. *Info.* means the input of different methods:  $\mathbf{G}$  donates the graph structure,  $\mathbf{X}_1$  and  $\mathbf{X}_2$  represent the node feature of view 1 and view 2.

所提出的 MAGCN 明显优于所有被比较的方法。具体来说，MAGCN 在 Cora 和 Citeseer 上比单个视图的 GCNs 提高了 5%以上，在 Pubmed 上提高了 1%以上，这表明通过几何关系和概率分布的一致性来整合不同的视图是有效的。

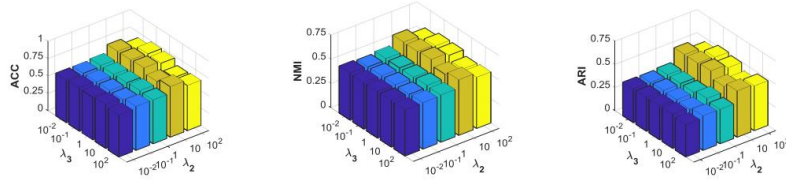
• 概率分布一致性分析



从第三类样本中随机抽取一个样本，计算每个视图的低维表示的原始概率分布  $Q_m$ ，并计算共同特征表示  $\mathbf{Z}$  的目标概率分布  $\mathbf{P}$ 。目标是得到理想的多视图描

述空间。在低维特征  $Z_m$  方面，使用概率分布的一致性作为约束，以减少不同视图之间的差异。如图所示，在初始迭代中，随机初始化使每个类的概率基本相似，不可能找出样本属于哪个类。但第三类的概率随着迭代次数的增加而增加，且在  $Z, Z_1$  和  $Z_2$  上的概率分布趋于一致，这表明理想的多视图描述特征  $Z$  是逐渐学习到的。

#### • 参数的影响（本论文作图错误，将 $\lambda_1, \lambda_2$ 位置放反）



模型中有三个正则化参数： $\lambda_1, \lambda_2, \lambda_3$ ，使用控制变量的方法来分析它们的影响。保持了模型中重构损失的正则参数  $\lambda_1$  不变，并改变了模型中几何关系一致性和概率分布一致性的正则参数  $\lambda_2, \lambda_3$ 。对于概率分布的正则项  $\lambda_3$ ，一致性，当  $\lambda_3$  设置为 10 时，多视图聚类的性能最好。当  $\lambda_2$  值过大时，模型聚类性能不好。因此，在实验中将概率分布一致性的  $\lambda_3$  设置在 10 左右。

#### • 分析不同的视图 2

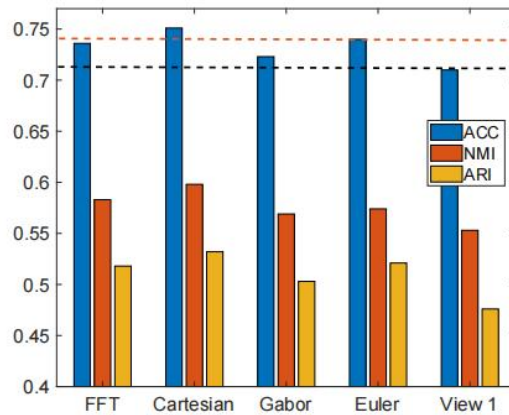


Figure 5: Metrics vs. different view 2 on Cora database.

分析 Cora 在数据库上以不同方式（FFT、笛卡尔、Garbor 和欧拉）构建时模型的性能。为了便于比较，使用视图 1 作为基线。为了保证所有参数的一致性，保证了实验的公平性。如图所示，可以看到，对于所有类型的视图 2，使用两个视图的聚类结果优于单个视图 1（用黑线标记）。此外，笛卡尔乘积方法比视图 2（用红线标记）的其他构造方法效果更好。这就是为什么在所有其他实验中使用笛卡尔积来构造第二个视图。

## 5. 总结

在本文中，提出了一种新的用于聚类的多视图属性图卷积网络（MAGCN），这是一种通用的多视图图神经网络的方法。MAGCN 采用双编码器设计，对提取的特征进行高维重构，并整合低维一致性信息。多视图属性图自动编码器和一致的嵌入编码器网络依次降低了噪声和不同视图之间的差异，最终得到了理想的聚类多视图属性图描述空间。在多视图图结构数据库上的实验结果证明了该方法的有效性，并优于几种最先进的算法。