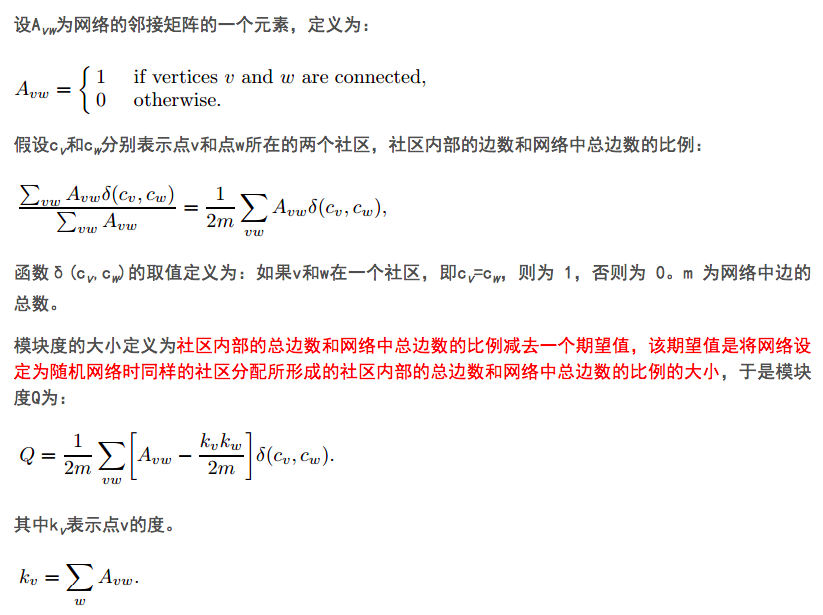
无监督学习的评估方法

**无监督学习的常见分类：**

1. 聚类
2. 降维
3. 异常检测

**聚类常见评估方法：**

* 根据聚类组内距离，以及组间距离，聚类数量等：
  + sum of squared error (SSE) 组内平方误差和 VS 聚类数量
  + **Calinsky criterion：** *SSB/(K-1))/(SSW/(n-K))*
* k是聚类数，N是样本数，SSw是我们之前提到过的组内平方和误差， SSb是组与组之间的平方和误差，SSw越小，SSb越大聚类效果越好，所以Calinsky criterion值一般来说是越大，聚类效果越好
  + 图聚类评估 – Modularity（模块度）：

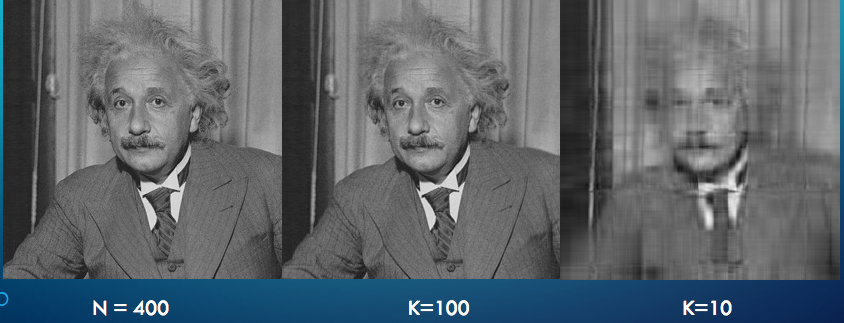


**降维常见评估方案：**

* 目标：压缩率和还原度：
  + 例：深度学习 AutoEncoder 降维，优化目标：压缩率和还原度

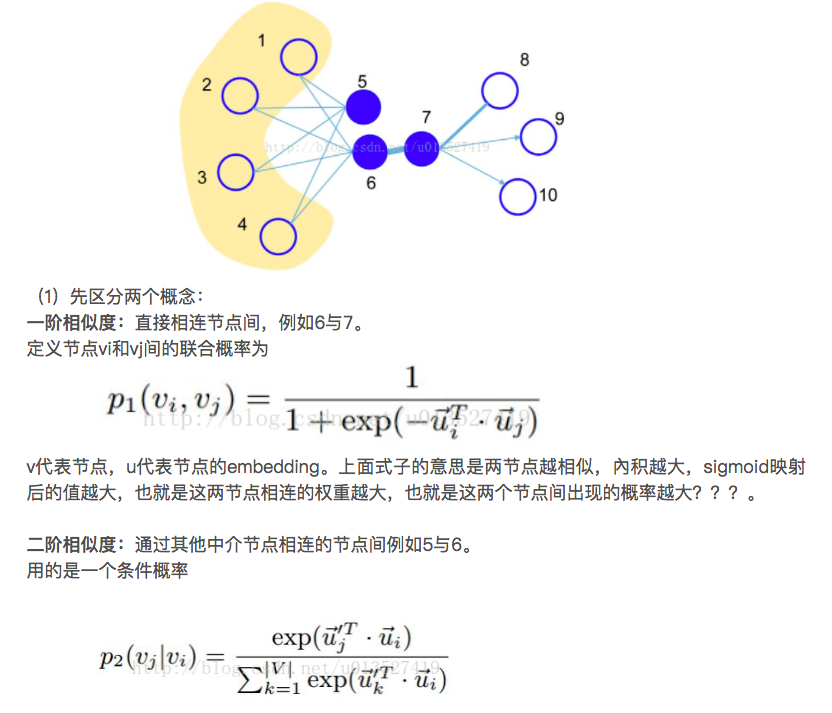
****

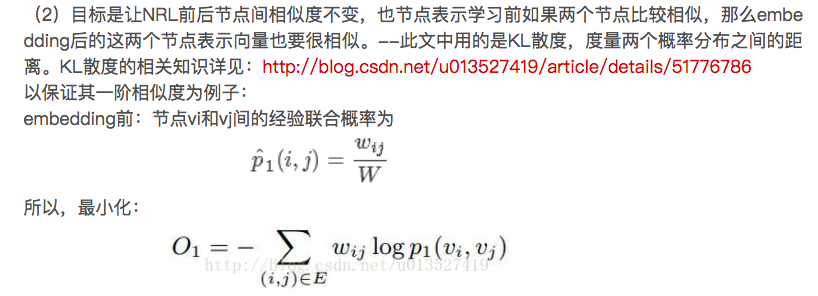
* + 例：SVD奇异值分解，优化目标：压缩率和还原度

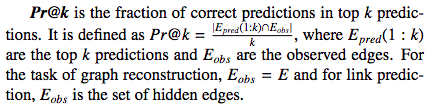
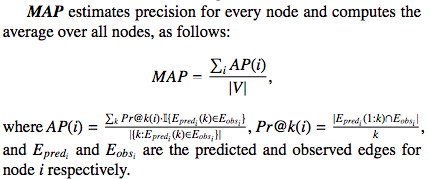
****

* + **word Embedding :**
  + **network Representation （GRAPH EMBEDDING）**

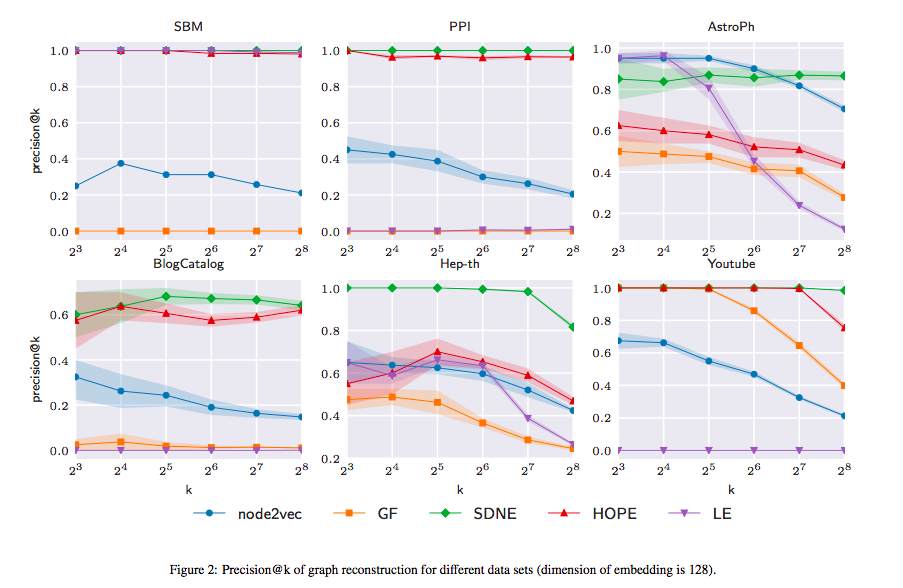
1. **尽量保留原网络结构：相似度定义如下：**

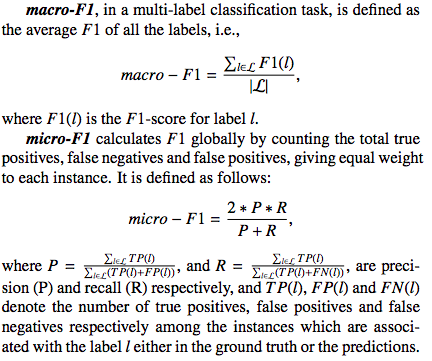


****

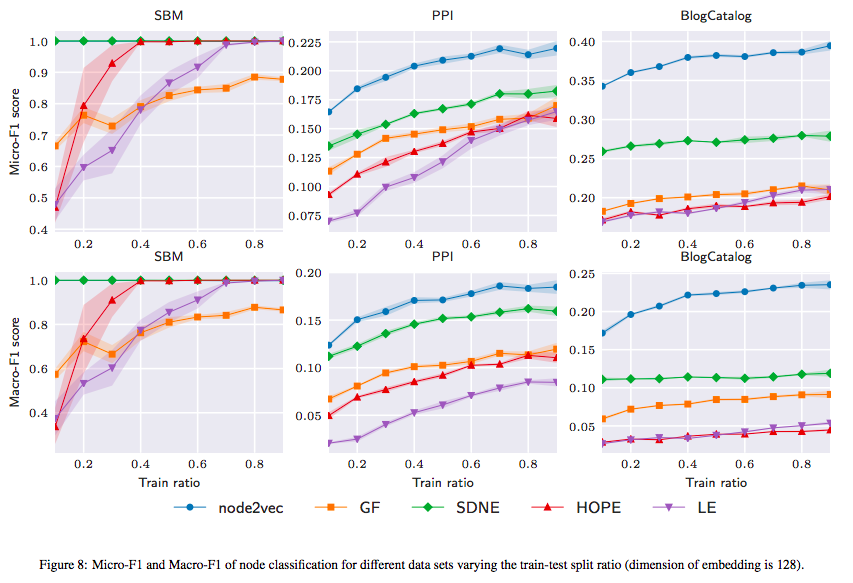
1. **尽量大的压缩率。**
2. **结合Link Prediction（有监督学习）问题再进行评估。评估指标有: Pr@k 和 MAP**
   1. ****
   2. ****

**常见的Graph Embedding算法应用在Link Prediction效果如下：**

****

1. **结合Node Classification问题进行评估：评估指标有macro-F1和 micro-F1:**
   1. ****

**常见的Graph Embedding算法应用在Node Classification效果如下：**

****