

deepwalk

先阐述语料库中词项之间的关系和网络结点具有相似的分布，提出random walk 得到每个结点的上下文 Context(w)，将(w,Context(x))作为样本；将nlp中skip-gram模型推广到网络表示方法中，对目标函数不断优化，梯度下降更新结点的表示向量，最后得到网络表示的矩阵。

优化问题：

$$\text{minimize} -\log Pr(\{v_{i-w}, \dots, v_{i-1}, v_{i+1}, \dots, v_{i+w}\} | \Phi(v_i))$$

deepwalk算法如下

```
Input: graph G(V,E), walks per vertex y, window size w, embedding size d
Output: matrix of vertex representations  $\phi \in \mathbb{R}^{|V| \times d}$ 
Initialization: Sample  $\phi$  from  $U^{|V| \times d}$ 
Build a binary tree T from V
for i=0 to y:
    o = Shuffle(V)
    for each  $v_i \in o$ :
        wvi=Randomwalk(G,  $v_i$ , t)
        SkipGram( $\phi$ , wvi, w)
```

SkipGram(ϕ , Wvi, w) 算法如下

```
for each  $v_j \in wvi$ :
    for each  $u_k \in wvi[j-w:j+w]$ 
        J( $\phi$ ) =  $-\log Pr(u_k | \phi(v_j))$ 
         $\phi = \phi - \alpha * \partial J / \partial \phi$ 
```

LINE

在deepwalk基础上提出一阶相似度和二阶相似度

1. **一阶相似度**是两个顶点直接相连，则表示后也应该相似
2. **二阶相似度**是两个顶点通过中介（邻居）相连，公有的邻居越多，表示后越相似

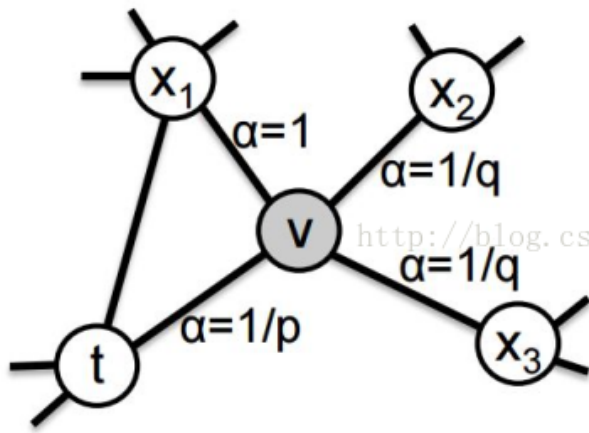
node2vec

对deepwalk改进了随机游走的策略，使用**BFS**和**DFS**两种方式进行游走，定义了两个参数

和q，对**BFS**和**DFS**进行平衡，使得得到的序列既体现局部信息，又体现宏观信息，有很高的适应性

BFS：体现微观局部信息，**DFS**体现宏观信息

p (Return parameter 返回概率参数) 对应**BFS**，p控制调回上一结点的概率，t跳到v后，有1/p的概率跳回到t
q(In Out parameter 离开概率参数) 对应**DFS**，q控制跳到其他结点的概率，t跳到v后，有1/q的概率跳到不与t直接相连的结点



$$\alpha_{pq}(t, x) = \begin{cases} \frac{1}{p} & \text{if } d_{tx} = 0 \\ 1 & \text{if } d_{tx} = 1 \\ \frac{1}{q} & \text{if } d_{tx} = 2 \end{cases}$$

其中, d_{tx} 表示 t 与 x 的距离

WWW

A3embed:Attribute Association Aware Network Embedding

analyze the attributes together with network structure. In social networks many users have similar attributes are not connected to each other. Structural proximity is not sufficient to explain whether nodes in a network are similar or dissimilar. This method can bring us a huge opportunity to capture the nodes' underlying similarity.

通过属性相似度表示学习, 可以找到属性相似, 兴趣相似的用户

GPSP: Graph Partition and Space Projection based Approach for Heterogeneous Network Embedding

对于异质网络, 结点的类型和关系都是多样的。同质网络的方法并不适用因为不能区分不同类型的对象和关系。本文提出了一种基于图分割和空间投影的方法做异构网络的表示。**首先**, 按照边的类型, 将原图分割为很多子网络, 这些子网络有两种类型:

1. 同质网络 (相同类型的结点和关系)
2. 二部图 (相同关系, 但是结点有两种类型)

其次, 使用传统的网络表示对每个同质子网络做表示, **第三步**, 对于每个二部图, 对这个二部图的关系类型, 学习相关类型的结点关系的表示向量; **最后**, 通过二部子图的关系将不同类型的同构子图连接起来, 形成异构网络的最终表示。

MELL:Effective Embedding Method for Multiplex Networks

现有的网络表示方法主要用于单层的网络, 本文的方法用于多路复用网络。多层网络的每层是相同类型的结点, 将每层的结点做嵌入到低维空间, 再通过原图将嵌入层之间进行层连接, 捕获并表示每层之间的连通性, 得到层向量。通过嵌入向量和层向量做链接预测

Network Embedding Based Recommendation Method in Social Networks

首先, 使用node2vec模型进行网络表示, 将每个用户映射到低维空间, 然后讲他们合并到矩阵分解(MF)模型中, 该模型结合潜在和表示后的特征用户推荐。本文的方法可以利用社交网络信息, 还可以利用协同过滤模型进行推荐。

RaRE:Social Rank Regulated Large-scale Network Embedding

根据结点的同趋性，网络中相连的结点的表示应该具有更高的相似度，相似度高的结点意味着他们的属性相近，然而，对于社会等级高的名人，他们往往拥有大量的粉丝，但是他与粉丝之间没有那么强烈的相似度，相反，一些初级群体拥有很高的相似主题却被忽略。我们提出的网络表示模型考虑两种因素的关系：1.属性相似（proximity based factory） 2.流行度高(social rank)，这两种关系是相互依赖的：等级低的结点指向等级高的结点更看重 2.流行，反之，更看重1.属性。我们的方法在构建网络表示过程中平衡这两种影响