



网络表示学习

Graph Representation Learning

汇报人：薛荧荧 汇报时间：2018年7月



目录

CONTENTS

01

研究背景与意义

02

模型简介

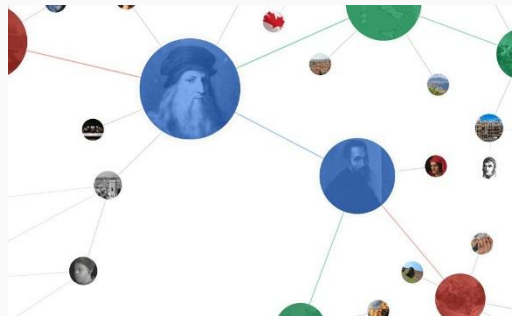
03

效果、应用、发展

04

存在问题

Graph Representation Learning (GRL, 网络表示学习/图特征表示学习)



motivation

网络结构在日常生活中非常**普遍**：社交网络、学术引用网络、知识图谱。

随着互联网的发展，网络结构日益**复杂、庞大**。

如何高效地在大规模网络上进行网络分析任务，例如节点分类、聚类、知识图谱中实体相关性计算、链接预测、社区发现等？

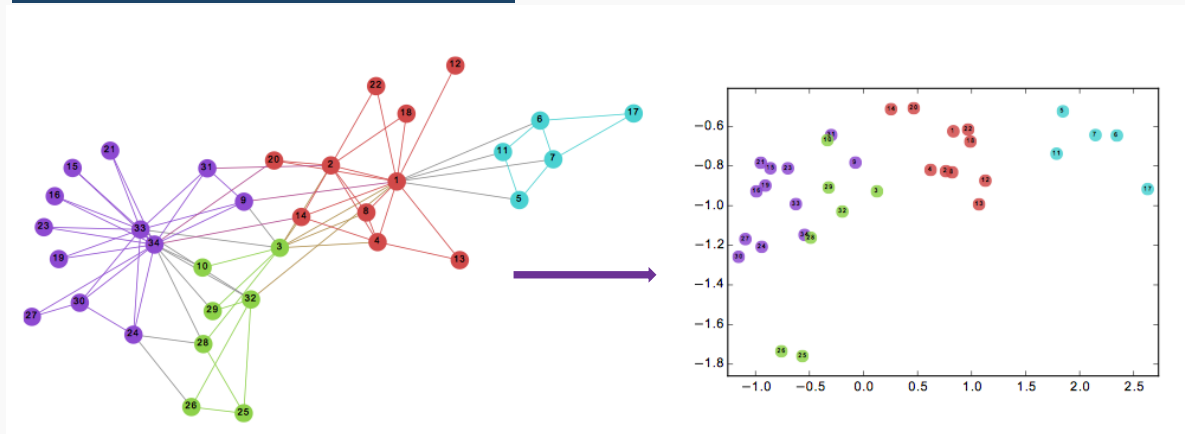
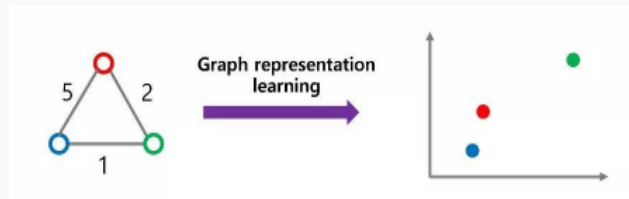
为了进行相应的网络分析任务，最重要的问题就是如何利用网络中的信息，对网络中的节点进行有效的特征表示。

网络表示的质量，对于进行后续的网络分析任务至关重要。

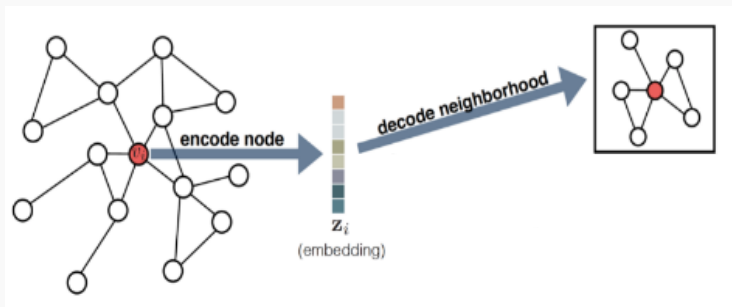
Graph Representation Learning (GRL, 网络表示学习/图特征表示学习)

GRL也被称作 Network Embedding、Graph Embedding。

将图中每一个节点都映射到一个低维向量空间，变成一个多维向量（向量化（Embedding）技术），并且在此空间内保持原有图的结构信息或距离信息。



Graph Representation Learning (GRL, 网络表示学习/图特征表示学习)



01 优点

能将复杂的网络信息变成结构化的多维特征，从而实现更方便的算法应用。

传统图算法时间复杂度高且不利于并行计算（节点相似性计算、子图相似性计算问题）

02 难点

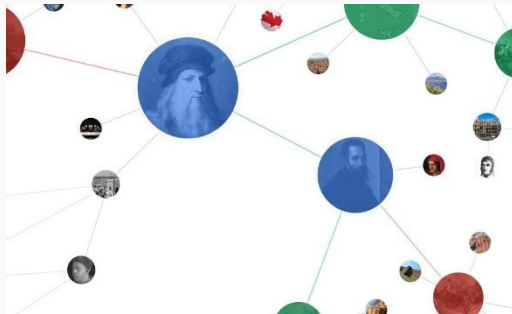
学习属性的选择、提取：不同的向量化表示方法本质上都是对网络信息的一种摘要（节点的近邻关系、角色）。

规模化：现实应用中很多网络包含了大量的节点和边，高效的向量化方法，能够在短时间内处理超大规模的网络，才比较有实际应用的可能性。

向量维度：如何确定合适的向量表示维度，是一个很难的问题，并且也是和具体场景相关的。

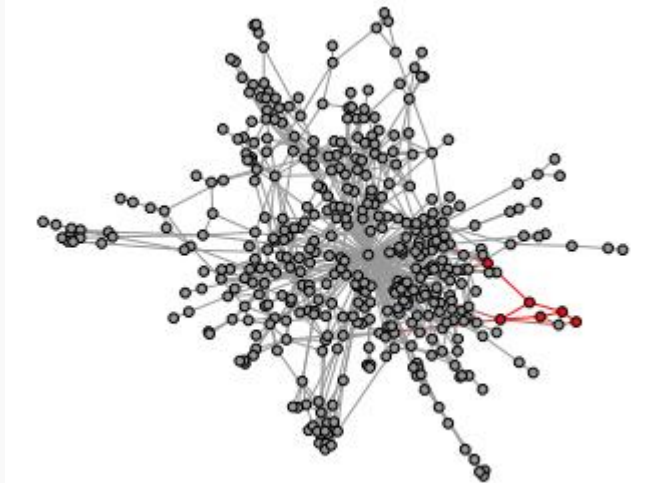
Random walk based Graph Representation Learning(网络表示学习代表模型)

1. 同构网络（文章引用网络）
2. 异构网络（社交网络、知识图谱）



基于随机游走的网络表示学习模型：**deepwalk**

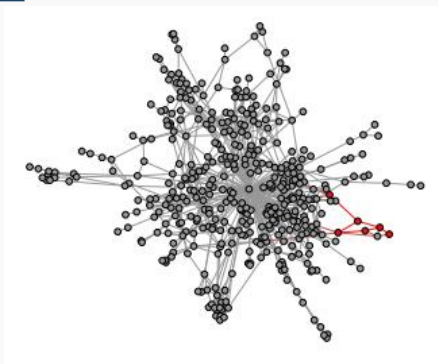
DeepWalk（基于随机游走的网络表示学习代表模型）



思想 对网络局部信息进行特征提取压缩，利用所提取的局部信息完成节点向量化表示

- 步骤**
1. 在网络上每个节点进行随机游走，生成由**网络中节点**构成的**随机游走序列**。
 2. 将上一步游走得到的**节点序列**看成**句子**，**节点**看作**单词**，利用自然语言处理领域广泛使用的训练词向量的神经网络模型——Skip-Gram，来训练网络节点的向量表示。

DeepWalk



Algorithm 1 DEEPWALK(G, w, d, γ, t)

Input: graph $G(V, E)$

window size w

embedding size d

walks per vertex γ

walk length t

Output: matrix of vertex representations $\Phi \in \mathbb{R}^{|V| \times d}$

1: Initialization: Sample Φ from $\mathcal{U}^{|V| \times d}$

2: Build a binary Tree T from V

3: **for** $i = 0$ to γ **do**

4: $\mathcal{O} = \text{Shuffle}(V)$

5: **for each** $v_i \in \mathcal{O}$ **do**

6: $\mathcal{W}_{v_i} = \text{RandomWalk}(G, v_i, t)$

7: SkipGram($\Phi, \mathcal{W}_{v_i}, w$)

8: **end for**

9: **end for**

01

随机游走 (Random Walk)

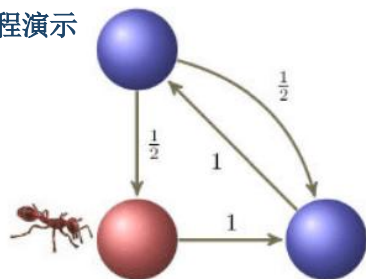
从当前节点均匀、随机地选择下一个节点，然后再从下个节点均匀、随机地选择下一个节点。

可以将随机游走看作是图结构的局部结构信息，以此来表示图结构中节点的隐含信息（度、重要程度、距离其他节点的距离远近）

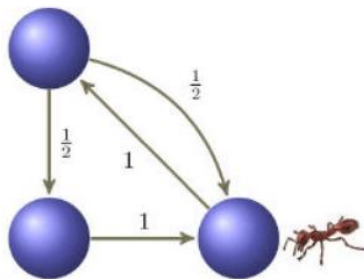
考虑节点 v 的前 k 个节点和后 k 个节点，很好地将网络邻居结构存入向量中。

DeepWalk

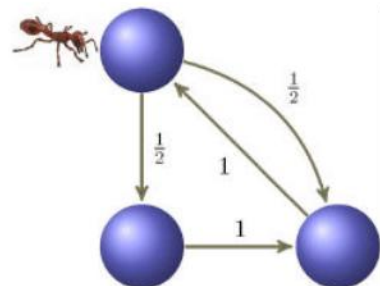
随机游走过程演示



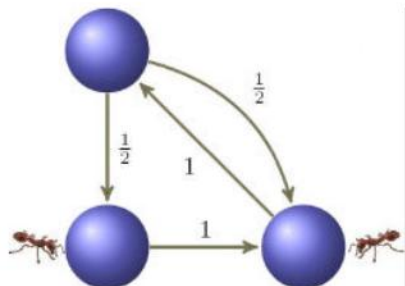
(a) 从起始点（红色）出发



(b) 到达第二个顶点后选择下一个目标



(c) 到达第三个顶点后有两个选择



(d) 依概率到达下一个目标

DeepWalk

02

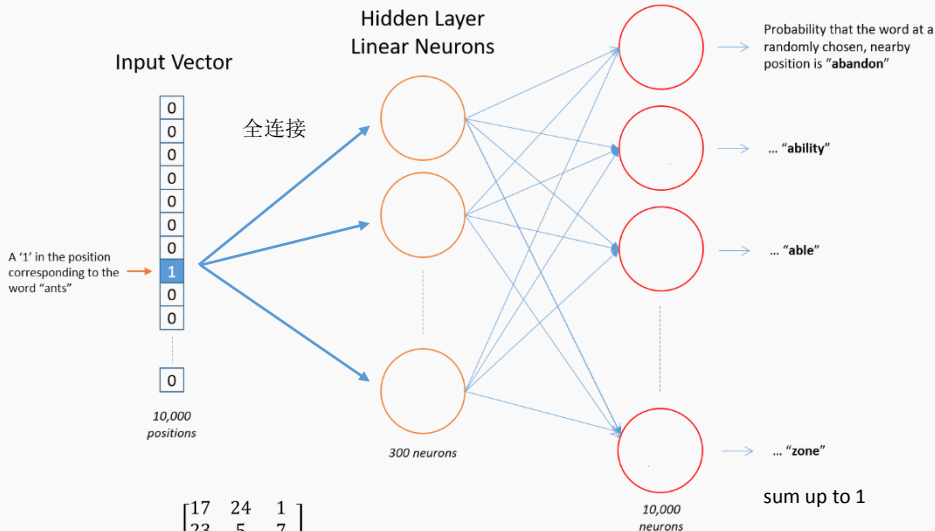
word2Vec之Skim-gram模型

利用上一步游走得到的序列，来训练网络节点的向量表示：
输入中心单词，
输出其周围词出现的可能性（概率）

... an efficient method for learning high quality distributed vector ...

Context
focus word
Context

Output Layer
Softmax Classifier




$$\begin{bmatrix} 0 & 0 & 0 & 1 & 0 \end{bmatrix} \times \begin{bmatrix} 17 & 24 & 1 \\ 23 & 5 & 7 \\ 4 & 6 & 13 \\ 10 & 12 & 19 \\ 11 & 18 & 25 \end{bmatrix} = \begin{bmatrix} 10 & 12 & 19 \end{bmatrix}$$

训练目标：在上下文中出现的单词概率最高



Graph Representation Learning（网络表示学习/图特征学习）

存在问题

1. 目前大部分向量化方法，仅利用网络结构信息，并没有利用网络节点本身的属性（比如文本、图像等多媒体信息）以及边所代表的关系信息，使得结果向量对网络信息的存储很有限。
 2. 大部分模型是对静态网络结构的直推学习，并没有考虑网络时间演化过程中新节点的生成和旧节点的湮灭，而网络的动态特性对理解其性质也至关重要。这个弱点甚至会影响向量化在动态网络上的效果。
- 

The background features a light gray, abstract network pattern. It consists of numerous small dots, some solid black and some light gray, connected by thin, dark gray lines. These lines form a complex web of triangles and polygons, creating a sense of connectivity and structure. The pattern is symmetrical, with a dense cluster of lines and dots on the left and right sides, and a more open, sparser area in the center where the text is located.

THANKS