《大数据分析与处理》第4次作业

姓名： 李燕琴 学号： 20195633 日期： 2022-05-19

1. 考虑下图中所示的树，使用BFS和DFS算法遍历图，然后列出每种算法中访问节点的顺序（25分）



BFS：v1,v2,v3,v4,v5,v6,v7,v8,v9,v10,v11,v12

DFS：v1,v2,v4,v8,v9,v5,v10,v11,v3,v6,v12,v7

2. 使用Prim算法，计算下图的最小生成树（25分）



1. 随机选择起点v1，设边权和weight sum为ws；
2. {v1,v2},ws=5;
3. {v1,v2,v3},ws=11;
4. {v1,v2,v3,v4},ws=18;
5. {v1,v2,v3,v4,v5},ws=23;
6. {v1,v2,v3,v4,v5,v6},ws=31;
7. {v1,v2,v3,v4,v5,v6,v7},ws=36;
8. {v1,v2,v3,v4,v5,v6,v7,v8},ws=43;
9. {v1,v2,v3,v4,v5,v6,v7,v8,v9},ws=51.

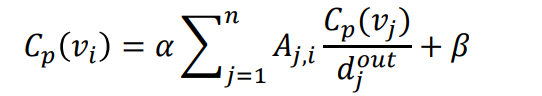
得最小生成树值为51。

3. 计算给定下列条件时下图的PageRank值（25分）



* α=1, β=0
* α=0.85, β=1
* α=0, β=1

根据PageRank算法的代数原理：



可改写为：

徽标

描述已自动生成

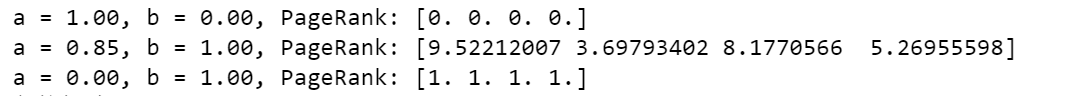
由题意得上图的邻接矩阵为：

A=图片包含 日历

描述已自动生成, D = 日历

描述已自动生成

根据代数计算公式，编写代码，得到结果如下：



|  |
| --- |
| 源代码 |
| A = np.array([      [0,1,1,1],      [1,0,0,1],      [1,0,0,0],      [0,0,1,0]  ])  D = np.diag(A.sum(axis=1))  # 出度矩阵  def pageRankAlgebra(a,b):      I = np.eye(A.shape[0])      tmp = I-a\*np.dot(A.T,np.linalg.inv(D))      C = b\*np.dot(np.linalg.inv(tmp),np.ones(A.shape[0]))      print("a = %.2f, b = %.2f, PageRank:"%(a,b),C)  pageRankAlgebra(1,0)  pageRankAlgebra(0.85,1)  pageRankAlgebra(0,1)  print("邻接矩阵:\n",A)  print("出度矩阵:\n",D) |

4. 简述羊群效应、信息级联、创新扩散和流行病传播的主要联系与区别（25分）

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 羊群效应 | 信息级联 | 创新扩散 | 流行病传播 |
| 网络结构可见性或个体知识域 | 所有人的决策对所有用户可见，接近一个完全图 | 在一组个体中扩散，且个体之间通过网络连接，只能观察到其近邻（好友）的决策行为，可用信息少 | 外在高度可见，但网络传播结构未知；在可以了解到所有用户的决策行为时，用户本身的决策行为更容易受到其近邻的影响。 | 通常假设传播的网络结构未知 |
| 影响途径 | 受到所有用户的影响 | 只受近邻影响 | 受到内部（传播本身特征）、外部（新事物本身特征）因素影响 | 受到全局影响，可以用来分析全局的传播模式 |