



西安电子科技大学  
XIDIAN UNIVERSITY

# 复杂网络与群体智能

吴建设

Email: [jshwu@mail.xidian.edu.cn](mailto:jshwu@mail.xidian.edu.cn)

西安电子科技大学人工智能学院

课件下载地址:

<https://web.xidian.edu.cn/jshwu/teach.html>



- (1) 节点相似性
- (2) 基于局部结构相似性指标
- (3) 全局路径特征



- 节点相似性指标用于评价节点之间的相似程度;
- 用途：可作为**链路预测**、**节点分类**的依据。
- 依据：节点相似性度量的依据是**局部结构特征**或者**全局路径特征**。



## ➤ 基于局部结构相似性指标

- 用途：可作为**链路预测**、**节点分类**的依据。
- 依据：节点相似性度量的依据是**局部结构特征**或者**全局路径特征**。



## ➤ 基于局部结构相似性指标

### 1 公共邻居

- 顾名思义，共同邻居指标认为两个节点之间共同的邻居越多，这两个节点的相似性越高。用 $\Gamma(i)$ 表示节点 $v_i$ 邻居集合，则 $\Gamma(i) \cap \Gamma(j)$ 表示两个节点邻居集合的交集。用 $|\cdot|$ 表示集合中节点的数量，则共同邻居相似性指标 $S_{ij}$ 如式（3-1）所示：

$$S_{ij} = |\Gamma(x) \cap \Gamma(y)| \quad (3-1)$$

- 该方法的优点是简单，但是考虑的因素较少。在共同邻居的基础上，考虑其它因素，比如节点度，可以得到其它几种指标。



## ➤ 基于局部结构相似性指标

### 2 Salton相似性指标

- 也称为余弦相似性指标，是由Salton等人提出的，定义如下：

$$S_{ij} = \frac{|\Gamma(i) \cap \Gamma(j)|}{\sqrt{k_i \times k_j}} \quad (3-2)$$

其中， $k_i$ 表示节点 $v_i$ 的度。

- 把邻接矩阵的第 $i$ 行和第 $j$ 行视为两个向量 $a_i$ 和 $a_j$ （元素为0和1），并使用两者夹角的余弦值作为相似度的计算。



## ➤ 基于局部结构相似性指标

### 3 Jaccard相似性指标

这个指标是由Jaccard提出的，定义如下：

$$S_{ij} = \frac{|\Gamma(i) \cap \Gamma(j)|}{|\Gamma(i) \cup \Gamma(j)|} \quad (3-3)$$

可见，Jaccard相似性指标是两个节点邻居集合的交集与并集的比值。



## ➤ 基于局部结构相似性指标

### • 4 Sorenson指标

是由Sorenson在处理植物生态数据时提出的，定义如下

$$S_{ij} = \frac{2|\Gamma(i) \cap \Gamma(j)|}{k_i + k_j} \quad (3-4)$$

### • 5 大度节点有利指标 (HPI)

$$S_{ij} = \frac{|\Gamma(i) \cap \Gamma(j)|}{\min\{k_i, k_j\}} \quad (3-5)$$



## ➤ 基于局部结构相似性指标

### • 6 大度节点不利指标 (HDI)

$$S_{ij} = \frac{|\Gamma(i) \cap \Gamma(j)|}{\max\{k_i, k_j\}} \quad (3-6)$$

### • 7 优先连接指标 (PA)

$$S_{ij} = k_i \times k_j \quad (3-7)$$



## ➤ 基于局部结构相似性指标

### • 8 AA指标

$$s_{ij} = \sum_{z \in \Gamma(i) \cap \Gamma(j)} \frac{1}{\log k_z}$$

该指标主要考虑了两节点**共同邻居的度**的信息，主要考虑到共同邻居中**度小的节点的影响**。这一点很容易理解，例如在新浪微博上甲乙两个人都关注了一个明星，丙丁两个人共同关注了一个普通的人，则丙和丁两个人认识的可能性大于甲和乙两个人认识的可能性。



## ► 基于局部结构相似性指标

### • 9 RA指标

$$s_{ij} = \sum_{z \in \Gamma(i) \cap \Gamma(j)} \frac{1}{k_z}$$

周涛等提出的，RA指标主要基于网络上资源传递的思想，例如在网络中不存在连边的两个节点 $v_i$ 和 $v_j$ ，要想从 $v_i$ 传递一些资源到 $v_j$ ，就需要传递媒介，在网络中他们的共同邻居则充当着这种传递的媒介。从节点 $v_i$ 传递一个单位的资源到 $v_j$ ，首先先平均分配传递到 $v_i$ 的邻居，每个邻居也说平均分配给它的邻居，则 $v_j$ 得到的资源数则为这两个节点的相似度。



## ➤ 基于局部结构相似性指标

- **10** Leicht-Holme-Newman(LHN1)。

不是与可能的最大值进行比较，而是与这样的邻居的预期数目相比较。其定义是

$$s_{ij}^{\text{LHN1}} = \frac{2|\Gamma(i) \cap \Gamma(j)|}{k_i \times k_j}$$

其中的分母 $k_x \times k_y$ 是节点 $x$ 和节点 $y$ 在配置模型中共同邻居的期望。



## ➤ 基于全局路径的相似性指标

### • 1 Katz指标

该指标主要根据考虑了网络的所有路径信息，其定义如下：

$$s_{ij} = \sum_{l=1}^{\infty} \alpha^l \cdot |\text{paths}_{i,j}^{<l>}| = \alpha A_{ij} + \alpha^2 (A^2)_{ij} + \alpha^3 (A^3)_{ij} + \dots$$

式中 $\text{paths}_{i,j}^{<l>}$ 代表路径长度为 $l$ 的路径的个数。 $A_{ij}$ 代表邻接矩阵 $A$ 的第 $i$ 行第 $j$ 列元素。

这里要注意2阶路径、3阶路径的计算方式。

参数 $\alpha$ 小于邻接矩阵最大特征值的倒数，级数收敛，则定义还可以表示为：

$$S = (I - \alpha \cdot A)^{-1} - I$$



## ► 基于全局路径的相似性指标

- 2 Leicht Holme Newman Index (LHN2)指标

此相似度指标基于这样一个想法：如果两个节点它们的邻居节点们是相似的，那这两个节点也是相似的。可以获得自我一致的矩阵公式

$$S = \phi AS + \psi I = \psi(I - \phi A)^{-1} = \psi(I + \phi A + \phi^2 A^2 + \dots)$$

其中 $\phi$ 和 $\psi$ 都是自由元素用于控制相似度指标两个部分的平衡。设置 $\psi = 1$ ，此相似度指标就与Katz相似度指标非常相似。



西安电子科技大学  
XIDIAN UNIVERSITY

谢谢

