

lec17.3 SNA Centrality Measures

Social Network Analysis - Measures of centrality and prestige

术语 (Terminology) :

1. Actors:

- **定义**: 感兴趣的对象 (例如, 人、组织、设备等)
- **例子**: 在社交网络中, 用户是 actors

2. Interactions or Relationships:

- **定义**: actors 之间的交互或关系
- **例子**: 在社交网络中, 朋友关系或交流互动是 interactions 或 relationships

3. 图的表示:

- **定义**: 这些演员和关系可以用图 $G = (V, E)$ 来表示, 其中:
 - V 代表节点, 表示 actors
 - E 代表边, 表示演员之间的关系

4. 例子:

- **Facebook**: 用户和朋友关系
- **LinkedIn**: 用户和专业关系
- **Co-authorship network**: 研究人员和他们的共同作者关系

网络分析的应用

1. 研究结构属性 (Study structural properties) :

- 分析网络的整体结构特性, 例如节点的度分布、路径长度、聚类系数等

2. 识别“中心”或“有影响力”的节点 (Identify "central" or "influential" nodes) :

- 通过中心性度量 (如度中心性、接近中心性、中介中心性等) 找出网络中最重要或最具影响力的节点

3. 识别关键节点 (Identify critical nodes) :

- 找出在网络中起关键作用的节点, 这些节点的失效可能会导致网络的断裂或功能失效

4. 检测由一组演员形成的“社区” (Detect "communities" formed by a group of actors) :

- 使用社区检测算法识别网络中的群体或子网络, 这些群体内部的节点联系紧密, 但与外部节点的联系较少

5. 检测异常子结构 (Detect abnormal substructures) :

- 识别网络中的异常模式或结构, 例如网页图中的链接农场 (link farms) 等

6. 新链接预测 (New link prediction) :

- 预测网络中未来可能出现的新连接, 常用于推荐系统和社交网络

7. 社交影响分析 (Social influence analysis) :

- 分析节点之间的影响力和传播路径, 了解信息或行为在网络中的传播过程

8. 其他 (etc.) :

- 图论和网络分析还有许多其他应用, 例如网络优化、流量分析、生物网络分析等

Measures of vertex “importance”

中心性度量 (Measures of Centrality) ——用于无向图

1. 度中心性 (Degree Centrality) :

- **定义**: 度中心性是通过计算节点的度（即连接到该节点的边的数量）来评估节点的重要性
- **意义**: 度数越高的节点通常在网络中越重要，因为它们直接连接的其他节点更多

2. 接近中心性 (Closeness Centrality) :

- **定义**: 接近中心性是通过计算节点与所有其他节点之间的平均最短路径长度来评估节点的重要性
- **意义**: 接近中心性高的节点能够在更短的路径上连接到网络中的其他节点，因此在传播信息或资源时更为高效

3. 中介中心性 (Betweenness Centrality) :

- **定义**: 中介中心性是通过计算节点在所有最短路径中出现的频率来评估节点的重要性
- **意义**: 中介中心性高的节点通常位于网络的关键路径上，充当信息或资源流动的桥梁

威望度量 (Measures of Prestige) ——用于有向图

1. 度威望 (Degree Prestige) :

- **定义**: 度威望通过计算节点的入度（即有多少其他节点指向该节点）来评估节点的重要性
- **意义**: 入度高的节点在网络中被更多的其他节点指向，通常具有更高的威望或影响力

2. 接近威望 (Proximity Prestige) :

- **定义**: 接近威望通过计算节点与指向该节点的其他节点之间的平均最短路径长度来评估节点的重要性
- **意义**: 接近威望高的节点可以更快地被网络中的其他节点访问，表明它们在网络中的中心位置

这些度量方法提供了不同的视角来评估节点在网络中的重要性，适用于各种分析和应用场景。例如：

- **度中心性**可以用于识别网络中的超级连接节点（如社交网络中的名人）
- **接近中心性**有助于识别在网络中具有战略位置的节点
- **中介中心性**适用于识别控制信息或资源流动的关键节点
- **度威望**和**接近威望**则专注于有向图中的节点影响力和可达性

度中心性 (Degree Centrality)

定义

度中心性 $C_D(i)$ 的定义：

- 无向网络中一个节点 i 的度中心性 $C_D(i)$ 等于该节点的度（即与该节点直接相连的边的数量），除以节点可能的最大度数
- 公式表示为：

$$C_D(i) = \frac{\deg(i)}{n-1}$$

其中， $\deg(i)$ 是节点 i 的度数， n 是图中节点的总数

动机 (Motivation)

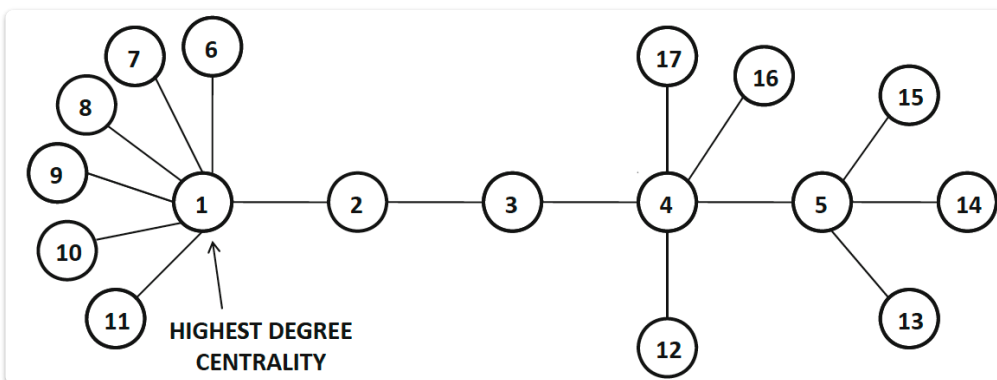
- 度数较高的节点通常是网络中的枢纽节点，它们在网络中更为中心，并且能够将网络中远离的部分连接在一起
- 高度中心性节点通过其大量的连接，能在网络中迅速传播信息或资源

应用

- **识别关键节点**：在社交网络中，高度中心性节点可能是社交活跃的用户，能够迅速影响其他用户
- **网络优化**：在通信网络中，度中心性高的节点可以作为关键路由节点，以提高网络的传输效率
- **疫情防控**：在传染病传播模型中，识别高中心性的节点可以帮助制定更有效的隔离和防控策略

度中心性的局限性

- **局限性**：
 - **局部信息**：度中心性主要使用局部信息，只考虑给定节点 i 的直接邻居节点，而不考虑超出直接邻居的其他节点
 - **忽略整体结构**：这种方法在某种程度上忽略了网络的整体结构
- **例子**：
 - 图片中的示例展示了节点1，它拥有最高的度中心性，因为它有最多的直接邻居。然而，尽管节点1在局部网络中是一个枢纽节点，它并不能被视为整个网络的中心节点
 - 从图中可以看到，节点1的直接邻居（节点6到11）都只与节点1相连，并没有其他连接。这意味着，虽然节点1有高的度中心性，但它的影响力和重要性仅限于它的局部区域



接近中心性（Closeness Centrality）

定义

接近中心性用于衡量一个节点在网络中相对于其他节点的“接近”程度。具体定义如下：

- 适用于无向和连通图
- **平均最短路径距离 Average shortest path distance, $AvDist(i)$** ：从节点 i 开始到网络中所有其他节点的最短路径的平均值
- 公式表示为：

$$AvDist(i) = \frac{\sum_{j=1}^n dist(i, j)}{n - 1}$$

其中， $dist(i, j)$ 表示节点 i 到节点 j 的最短路径距离， n 是网络中的节点总数

接近中心性 Closeness Centrality, $C_C(i)$

- 节点 i 的接近中心性是平均距离 $AvDist(i)$ 的倒数。

其中：

- $f_{jk}(i) = \frac{q_{jk}(i)}{q_{jk}}$ 是通过节点 i 的最短路径的比例，直观上 $f_{jk}(i)$ 表示节点 i 在节点 j 和节点 k 之间调控信息流的控制程度
 - q_{jk} 是节点 j 和节点 k 之间最短路径数量
 - $q_{jk}(i)$ 是节点 j 和节点 k 之间经过节点 i 的最短路径数量
- $\binom{n}{2} = \frac{n(n-1)}{2}$ 表示节点对的数量
- **计算步骤：**
 1. 找出所有节点对 (j, k) 之间的最短路径
 2. 计算每个节点对 (j, k) 的最短路径数量 q_{jk}
 3. 计算通过节点 i 的最短路径数量 $q_{jk}(i)$
 4. 计算每个节点对 (j, k) 上的 $f_{jk}(i)$
 5. 对所有节点对 (j, k) 的 $f_{jk}(i)$ 求和并取平均值

边介数中心性 (Edge Betweenness Centrality)

- **定义：**介数中心性不仅可以应用于节点，还可以应用于边。边介数中心性通过计算通过一条边的最短路径数量来衡量该边的重要性
 - **计算方法：**通过一条边的最短路径数量与总的最短路径数量的比值来计算
 - **高介数边：**具有高介数中心性的边通常连接图中不同簇（社区）的节点

与接近中心性的比较

- **接近中心性 (Closeness Centrality)：**
 - **基于距离：**接近中心性是基于节点到其他所有节点的最短路径距离来衡量节点的中心性的
 - **不考虑节点的重要性：**接近中心性不考虑节点在网络中作为中介的重要性，即通过该节点的最短路径的数量
- **介数中心性 (Betweenness Centrality)：**
 - **关键性 (Criticality)：**介数中心性衡量的是节点作为其他节点之间的桥梁的重要性，即通过该节点的最短路径数量
 - **信息流控制：**介数中心性的重要性在于确定那些在社交网络中对信息流控制最大的节点

重要性

- **信息流控制：**
 - 介数中心性高的节点在社交网络中具有重要的中介作用，它们是信息在网络中流动的重要路径
 - 这些节点的失效可能会导致网络中断裂或信息流的阻塞，因此它们在网络结构和功能中起着关键作用

介数中心性的关键点

1. **介数中心性的范围：**
 - 介数中心性的值介于0和1之间
 - 这个范围确保了介数中心性值可以标准化，并在不同规模的网络中进行比较
2. **高值对应于更高的介数：**
 - 介数中心性值越高，表示该节点在网络中的中介作用越显著
 - 高介数中心性的节点通常控制着大量的信息流动和路径，因此在网络中具有重要的桥梁作用
3. **适用于非连通网络：**

- 介数中心性可以在非连通的网络中定义
- 即使网络不是完全连通的，介数中心性仍然可以用来衡量各个子图中的节点的重要性

例子

- **社区检测 (Community Detection)**：节点和边介数中心性的概念广泛应用于社区检测算法中，如 Girvan-Newman 算法
 - **Girvan-Newman 算法**：一种基于边介数中心性的社区检测算法，通过逐步移除高介数边来识别网络中的社区结构

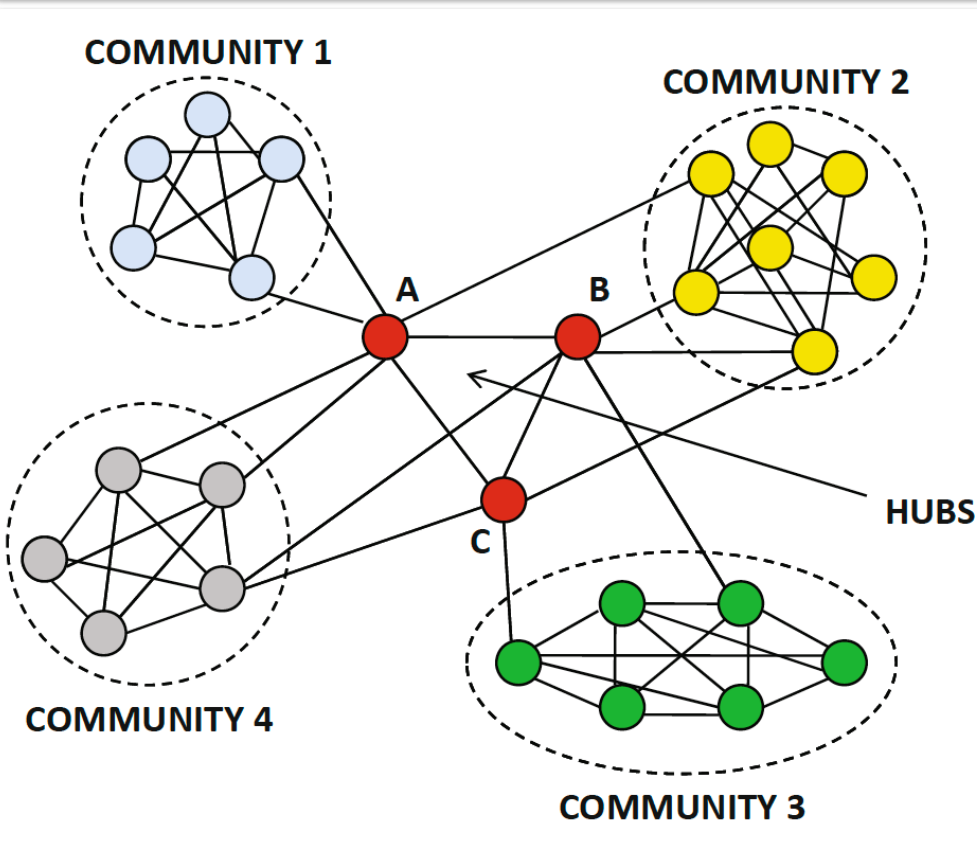
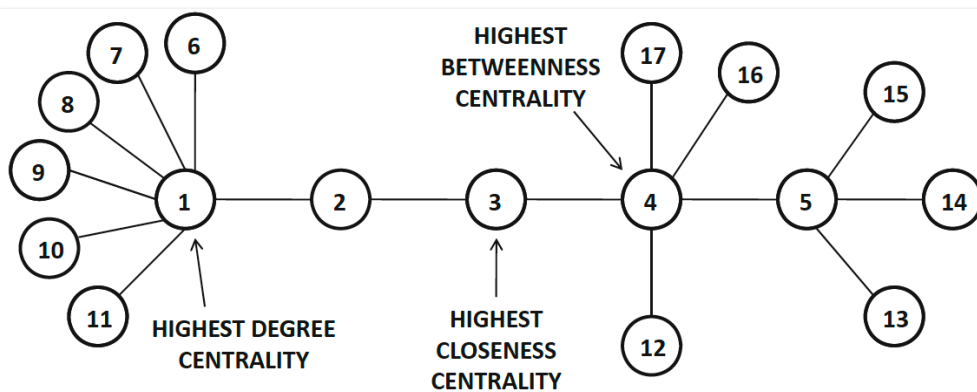
例子

假设有 4 个节点 A, B, C, D 构成的网络，其中：

- $q_{AB} = 2$ 表示节点 A 和节点 B 之间有两条最短路径。
- $q_{AB}(C) = 1$ 表示这两条路径中有一条经过节点 C 。

那么， $f_{AB}(C) = \frac{q_{AB}(C)}{q_{AB}} = \frac{1}{2} = 0.5$ 。

计算 C 的介数中心性 $C_B(C)$ 时，需要对所有可能的节点对重复上述步骤，并将结果平均。



度威望 (Degree Prestige)

定义

- **度威望**:
 - **适用范围**: 度威望只适用于有向网络
 - **计算方法**: 使用节点的入度 (in-degree) 而不是度数 (degree) 来计算
 - **公式**:

$$P_D(i) = \frac{\deg_+(i)}{n - 1}$$

其中, $\deg_+(i)$ 表示节点 i 的入度, n 是网络中节点的总数

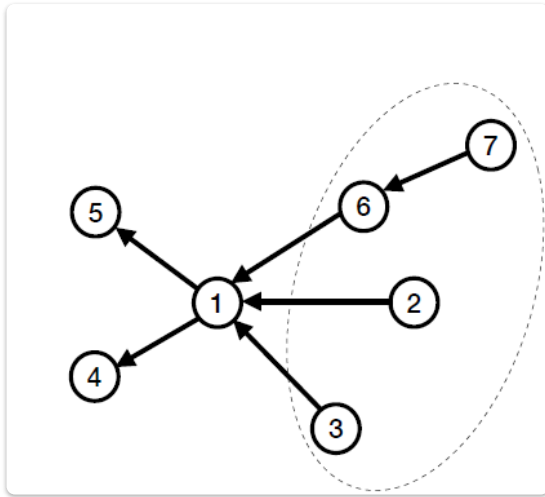
动机

- **高入度的影响**:
 - 只有高入度 (in-degree) 才能提升节点的威望, 因为节点的入度可以被视为对该节点受欢迎程度的投票
 - 一个节点的入度越高, 表示有更多的其他节点指向它, 表明它在网络中具有更高的威望或影响力

接近威望 (Proximity Prestige)

定义

- **接近威望** 仅适用于有向图
- ****影响集 Influence(i)**: 可以通过直接路径到达节点 i 的节点集
 - 图中节点 1 的影响集包括节点 2、3、4、5、6、7



- ****平均最短路径距离 AvDist(i)**: 到达节点 i 的平均最短路径距离
 - 计算公式:

$$\text{AvDist}(i) = \frac{\sum_{j \in \text{Influence}(i)} \text{dist}(j, i)}{|\text{Influence}(i)|}$$

其中, $\text{dist}(j, i)$ 是从节点 j 到节点 i 的最短路径距离, $|\text{Influence}(i)|$ 是影响集的大小

动机

- 仅使用平均距离的倒数 (如接近中心性) 来衡量威望是不公平的, 因为它忽略了影响集的大小

- 例如：
 - 对于节点 6, $\text{AvDist}(6) = 1$, 它的影响集中只有一个节点
 - 对于节点 1, $\text{AvDist}(1) = \frac{5}{4}$, 它的影响集中有四个节点

自然地, 节点 1 应有更高的威望, 因为它的影响集更大

计算方法

- **影响集比例 Influence Fraction**: 影响集大小相对于最大可能影响集的比例

$$\text{InfluenceFraction}(i) = \frac{|\text{Influence}(i)|}{n - 1}$$

其中, n 是网络中的节点总数

- **接近威望 Proximity Prestige, $P_P(i)$** :

$$P_P(i) = \frac{\text{InfluenceFraction}(i)}{\text{AvDist}(i)}$$

接近威望的值在 0 和 1 之间, 高值表示更到的威望

例子

例子

假设一个有向图, 节点1有四个节点 (2, 3, 4, 5) 可以直接到达, 而节点6只有一个节点可以到达:

- 对于节点1:
 - $\text{Influence}(1) = \{2, 3, 4, 5\}$
 - $\text{AvDist}(1) = 5/4$
 - $\text{InfluenceFraction}(1) = 4/6 = 2/3$
 - $P_P(1) = \frac{2/3}{5/4} = \frac{2/3}{1.25} = 0.53$
- 对于节点6:
 - $\text{Influence}(6) = \{7\}$
 - $\text{AvDist}(6) = 1$
 - $\text{InfluenceFraction}(6) = 1/6 \approx 0.167$
 - $P_P(6) = \frac{0.167}{1} = 0.167$

通过这种方式, 节点1的接近威望高于节点6, 反映了它在网络中更高的影响力。