arXiv:2301.07711v1 [stat.OT] 18 Jan 2023

用于有向无环图中心性分析的 Matlab 程序

作者： Richard S.J. Tol*

本文介绍了用于网络中心性分析的全新 Matlab 函数。本文采用广义平均距离，而非平均距离。如果优先考虑较近的关系，这种接近中心性度量也适用于非连通图。本文考虑的是到选定节点的距离，而非到所有节点的距离。除了垂直接近中心性度量（内接近中心性和外接近中心性）之外，本文还提出了水平交叉接近度。

关键词：网络中心性；内接近中心性；外接近中心性；交叉接近度；Matlab

I. 引言

有向无环图，也称多树，可用于描绘和分析广义上的祖先关系。Tol (2022*b*) 研究经济学诺贝尔奖得主的学术谱系，Tol (2022*a*) 则研究诺贝尔奖候选人。第一篇论文关注外接近中心性——谁是影响力最大的教授？（卡尔·克尼斯）第二篇则关注内接近中心性——与诺贝尔奖得主关系密切，是否会提高获诺贝尔奖的机会？（会）

Matlab 提供的网络邻近性度量很少，而且没有一种能将邻近性限定在选定节点范围内，而这正是上述研究所需要的。本文扩展了 Matlab 函数的功能范围，使其能够计算多树中的接近中心性。

II. 数据获取

学术谱系数据可在AcademicTree.org上找到。函数getactree.m根据一位学者的个人标识符创建其学术谱系。该函数调用函数 getparents.m，抓取这位学者的姓名，以及其教授的姓名和 ID。教授—学生配对以有向图的形式存储，这是 Matlab 用于有向图的数据结构。 getactree.m还会调用函数 addgraph.m，添加父母节点的父母节点，然后递归调用自身，添加父母节点的父母节点的父母节点，如此类推。

函数 getdesctree.m 会创建一位学者的后代多树，并通过其个人标识符进行识别。

*Tol：苏塞克斯大学经济系，英国法尔默 BN1 9SL，r.tol@sussex.ac.uk；阿姆斯特丹自由大学环境研究所，荷兰；阿姆斯特丹自由大学空间经济学系，荷兰；阿姆斯特丹廷贝亨研究所，荷兰；德国慕尼黑 CESifo；美国科罗拉多州戈尔登科罗拉多矿业学院佩恩公共政策研究所。

此过程可以针对任意学者列表重复执行。函数 MERGEDI-GRAPHS.M 会合并两个图，并在添加相关学者时删除重复条目。

III. 邻近性

图中节点 i 到图中其余部分的距离，可以用 H¨older 平均数来衡量

$$D_i(h) = \left(\frac{1}{|J|} \sum_{j \in J} D_{j,i}^h \right)^{\frac{1}{h}} \tag{1}$$

其中 $D_{j,\,i}$ 是节点 i 到任意节点 j的距离，也就是节点 i 和节点 j之间的边数。集合 J 通常包含所有节点 $j \neq i$，但也可以限定为具有特定特征的节点。在上述引用的应用中，J 仅包含诺贝尔奖得主。

对于 $h = 1$，H¨older 平均数就是算术平均数。可以使用 Matlab 函数 CENTRALITY计算，该函数包含在标准发行版中。请注意，$D_i(1) = \infty$，除非节点 i 是集合 J中所有其他节点的后代。[1] 因此，它不太适用于非连通图。

对于 $h = -1$，H¨older 平均数是调和平均数；如果网络中的某些节点无法到达，该值仍有界。换言之，调和平均数适用于连通子图和非连通子图：对于无法到达的节点$D_{j,\,i} = \infty$ 因此 $1/D_{j,\,i} = 0$。 Marchiori 和 Latora (2000) 将此作为距离度量，Gil-Mendieta 和 Schmidt (1996) 则将其倒数作为接近度量。

可以使用函数 HARMONICCENTRALITY 进行计算。该函数接受任意多树（Matlab 中的有向图）作为参数，并返回其中所有节点的调和平均中心性。函数 HARMONICNOBELITY 还接受一个额外参数，即一个布尔向量，为每个节点指定 0 或 1。该函数会返回所有节点到标记为 1 的节点的调和平均距离。

调和平均距离可用于强调邻近性，而弱化远距离关系。h 越负，近邻关系就越突出。函数 HOLDERCENTRALITY 和 HOLDERNOBELITY 与上述函数相同，但多了一个参数 $h < 0$。

方程（1）是一个外接近中心性度量。在多树中，外接近中心性衡量祖先关系。若在方程（$D_{j,\,i}$ 替换为$D_{i,\,j}$，则1）将得到一个内接近中心性度量，用于衡量后代关系。

上述函数有一个可选参数，可将输出从默认的外接近中心性切换为内接近中心性。

外接近中心性和内接近中心性度量垂直距离、父母节点和子节点。水平距离，即交叉接近中心性，也值得关注——兄弟节点的影响力可能与父母节点一样大

[1]在标准版本中，集合 J 包含图中的 all 个节点。

。在多树中，节点 i 到 j 的水平距离定义为

$$H_{i,j}(n) = \frac{|D_{k,i} = D_{k,j} = n|}{\max(|D_{k,i} = n|, |D_{k,j} = n|)} \tag{2}$$

也就是说，距离等于第 n 代共同祖先的数量除以祖先的最大数量。在生物学中，$H_{i,\ j}(1) = 1$ 适用于兄弟姐妹，$H_{i,\ j}(1)$ =0.5 适用于同父异母或同母异父的兄弟姐妹，而 $H_{i,\ j}(1) = 0$ 适用于其他所有人。 $H(i,j)(2) = 0.5$ 适用于第一代表亲， $H(i,j)(3) = 0.25$ 适用于第二代表亲，以此类推。

函数 CROSSDISTANCE会针对多树中所有选定节点，返回任意阶数 n 对应的矩阵 H 。两个节点之间的距离由函数 HORZDIST计算。[2]

构建出水平距离矩阵 H 后，方程（1）中广义平均数的倒数即定义了交叉接近中心性。函数 CROSS-CLOSENESS返回图中所有节点的这一度量。

IV. 结论

我从三个方面扩展了 Matlab 的中心性度量库。首先，标准接近中心性度量采用距离的算术平均数。我提出以广义平均距离取而代之；此外，若该度量优先考虑邻近关系而非远距离关系，它还适用于非连通图。其次，我将与所有节点的邻近性改为与选定节点的邻近性。第三，除了基于垂直距离的外接近中心性和内接近中心性之外，我还定义了基于水平距离的交叉接近中心性。我提供了收集学术谱系数据的工具，但这些中心性度量适用于任意有向无环图。

五、代码

所有提及的程序均位于GitHub。

参考文献

Gil-Mendieta、Jorge 和 Samuel Schmidt。1996。《墨西哥的政治网络》。*Social Networks*，18(4)：355–381。2

Marchiori、Massimo 和 Vito Latora。2000。《小世界中的和谐》。*Physica A: Statistical Mechanics and its Applications*，285(3)：539–546。2

托尔，理查德·S·J. 2022*a*. “诺贝尔奖得主学生培养出诺贝尔奖得主教授。” arXiv 讨论论文 2207.04441。 1

[2]这些函数用于考察共同祖先关系。很容易对其进行调整，以衡量共同后代关系，但我想不出这有什么应用。

托尔，理查德·S·J. 2022*b*. “克尼斯学派的兴起：经济学诺贝尔奖得主的师生网络。”《欧洲经济思想史杂志》，0(0)：1–24。 1

此图“logit.png”可从以下位置获取“PNG”格式版本：

http://arxiv.org/ps/2301.07711v1