



MILDDC算法及其在飞机模块化生产网络分析中的应用

An Algorithm-MILDDC and its Applying in Airplane Modular Production Networks Analyzing

戴爱明^{1,2} 肖灵机^{2/1} 江西财经大学 2南昌航空大学

摘要: 在中心度、节点最近邻和局部社团链接度最小增量有效融合的基础上,提出了一种基于中心度的最小链接度增量社团结构探测算法(Minimal Increment of Link Degree Based on Degree Centrality, MILDDC),并将MILDDC算法应用于某型飞机模块化生产网络并进行社团结构分析,计算结果说明,该算法具有一定的适用性,时间复杂度约为线性时间,并且对包含噪声、孤立点以及异形状的网络社团结构探测具有较强的鲁棒性。

Abstract: Based on effective integration of central degree, node nearest neighbor and minimum local community link degree increments, the thesis proposes Detecting Community Structure Algorithm by Minimal Increment of Link Degree Based on Degree Centrality (MILDDC), applies the algorithm in a type of airplane modular production network and analyzing its community structure, the result of calculation proves that the algorithm is applicable. Time complexity of MILDDC is about linear time, and the algorithm has a strong robustness on community structure mining including noise, isolated point and different shape networks.

关键词: 社团结构; 节点中心度; 链接度增量; 模块化生产网络; 飞机模块化

Keywords: Community Structure; Node Degree Centrality; Increment of Link Degree; Modular Production Networks; Airplane Modular

0 引言

复杂网络的核心研究内容是揭示复杂网络功能和结构之间的内在联系。社团结构探测对理解复杂网络结构特征属性至关重要^[1]。现实世界许多复杂网络全局拓扑结构呈现出由多个分散、自治实体的局部行为通过多种自组织方式涌现而成。鉴于各个节点的局部中心度(local centrality)能推断出隐藏的网络全局拓扑结构, Costa等人提出了一种基于网络的Hub节点探测社团结构的算法(l -

壳(l -shell)算法)^[2]。该算法的基本思想是:从网络中若干个具有最大度的节点出发,同时进行 l -壳的扩展,直到所有的节点都在一个 l -壳之内。 l -壳算法的最大局限性在于需要事先知道网络社团的数目,以确定同时进行扩散的“核心”数目,且每个社团内部都要包含一个Hub节点。此外, l -壳算法还要求网络中的这些社团是等“直径”的,否则,就很容易导致一个社团包含另一个社团的边缘节点。另外,虽然 l -壳算法在寻找网络社团结

构的时候只需要用到节点局部信息,但是它仍然需要知道整个网络的拓扑结构。Bagrow和Boltz在 l -壳算法基础上提出了BB算法^[3],这种算法实际上是一种自学习的算法,即某个节点该如何利用它周围节点的信息来自发地寻找它所在的社团结构。算法的大致思想就是从已知节点开始,通过 l -壳的扩展传播寻找该节点所在的社团结构。从起始节点到它的最近邻,然后到它最近邻的最近邻,以此类推,所有被 l -壳访问的节点,都要计算跟它们相关的两个变量:暴露度(Emerging Degree)和总暴露度(Total Emerging Degree)。这种算法并不完善,

* 基金项目:“十一五”优秀航空科学基金(2006ZG56015)、国家自然科学基金(71263040/G0304)、江西省社科规划项目(11YJ67)及江西省软科学项目(赣财教字[2007]136号)项目资助。



因为 l -壳的扩展很有可能会“溢出”该节点所在的社团。算法的准确与否很大程度上取决于所选的初始节点在社团中所处的位置。如果这个节点本身处于它所在社团的边缘位置,则 l -壳扩展的结果就会同时得到两个或者两个以上的社团。只有当这个节点处在它所在社团的中心位置时,即它到社团内其他节点的距离都是近似相等时,利用这个算法得到的局部社团结构才是准确的。针对局部社团结构的评价Clauset提出了一种衡量指标——局部模块度(Local Modularity),并根据局部模块度增量给出了相应的划分社团结构的算法^[4]。Clauset算法的时间复杂度为 $O(k_2 d)$,其中 k 表示社团的成员数, d 表示网络的平均节点度。Clauset算法和BB算法一样,局部社团探测的准确性很绝程度上取决于所选的初始节点在社团中所处的位置。

本文在以上文献基础上,提出一种新的基于中心度的最小链接度增量社团结构探测算法(MILDDC)。MILDDC算法首先通过节点中心度及节点之间最短路径两个参数确定网络社团结构中心节点集,然后对社团中心节点集进行最近邻聚类产生若干不完整社团,再对社团共享最近邻节点集及其余孤立节点集进行链接度增量计算,将节点划入链接度增量最小的社团,完成全局社团结构的探测。

1 相关定义及算法描述

1.1 相关定义

本文研究的是无方向、无权重的网络图。给定图 $G(V, E)$, $=\{v_1, v_2, \dots, v_n\}$ 表示图中节点集合, $E \subseteq \{(v_i, v_j) : v_i, v_j \in V\}$ 表示节点连接集合, 令无序偶对 (v_i, v_j) 表示节点 $v_i \in V$ 与 $v_j \in V$ 之间的边, (v_i, v_j) 可以用邻接矩阵描述。

定义1.1邻接矩阵(Adjacency Matrix), 其描述图中各节点两两之间的关系。邻接矩阵 $A=(a_{ij})$ 的元素 a_{ij} 定义为:

$$a_{ij} = \begin{cases} 1, & \text{其中, 如果 } (v_i, v_j) \in E, \text{ 元素} \\ 0, & \end{cases}$$

$a_{ij}=1$, 否则, $a_{ij}=0$ 。

定义1.2节点中心度(Degree Centrality), 其基本思想是重要节点是那些拥有与其它节点有较多连接边数的节点。一个网络图中节点的重要性可依据它们度分布的大小进行排序。相应地节点 v_i 的中心度定义为^[5]:

$$C_{v_i}^D = \frac{D_{v_i}}{n-1} = \frac{\sum_{v_j \in V} a_{ij}}{n-1} \quad (1)$$

其中 D_{v_i} 表示节点 v_i 的度, n 表示网络节点数。

定义1.3节点邻居, 其定义是与该节点相连的节点, 表示为:

$$N(v) = \{w \in V | (v, w) \in E\} \quad (2)$$

定义1.4局部社团外部链接度(Outer Link Degree of Local Community), 某个社团内的节点直接与社团外节点相连的边数称为局部社团外部链接度, 记为 L , 其公式如下^[6]:

$$L = \sum_{ij} C_{ij} \delta(i, j) \quad (3)$$

其中: C 表示某已知社团集合,

$$\delta(i, j) = \begin{cases} 1, & v_i \in C, \text{ 且 } v_j \in \bar{C}; \text{ 或} \\ & v_j \in C, \text{ 且 } v_i \in \bar{C} \\ 0, & \text{其他} \end{cases}$$

局部社团外部链接度是一个从局部角度出发度量社团结构的显著性程度的指标。局部社团外部链接度越小, 说明局部社团结构特征越显著。

定义1.5局部社团外部链接度增量(Minimal Increment of Out Link Degree of Local Community), 根据定义1.4可知, 当新的节点 v_i 合并到社团 C 后, 局部社团外部链接度可能会发生一定的变化。我们把节点 v_i 合并前后局部社团外部链接度的变化量称为节点 v_i 的局部社团外

部链接度增量, 可以由下式来表示^[6]:

$$\Delta L_i = D_{v_i}^{out} - D_{v_i}^{in} \quad (4)$$

其中, $D_{v_i}^{out}$ 表示节点 v_i 与社团 C 外部节点之间的链接边数, $D_{v_i}^{in}$ 表示节点 v_i 与社团 C 内部节点之间的链接边数。如果节点 v_i 与社团 C 外部节点之间没有链接变数, 这 ΔL_i 计算演变为:

$$\Delta L_i = N - D_{v_i}^{in} \quad (5)$$

其中, $N = \frac{n(n-1)}{2}$, n 表示网络节点数。

局部社团外部链接度增量 ΔL_i 是用来衡量节点 v_i 合并到社团后局部社团结构的显著性程度变化的指标。把社团的某个邻居节点加入该社团, 使局部社团外部链接度增长最慢(或下降最快), 那么就意味着该节点是所有邻居节点中加入到社团后形成的局部社团结构最显著的一个节点。因此, 在对网络节点进行划分时, 可以根据节点的局部社团外部链接度增量 ΔL_i 的值进行选择。 ΔL_i 的计算与社团内外部的链接情况无关, 只与新加入节点的链接情况有关。

1.2 算法描述

算法分为三个步骤, 首先根据节点中心度及节点间最短路径求出社团中心节点集, 然后根据中心节点集最短路径半径将社团中心节点的邻居节点进行聚类, 最后采取最小局部社团外部链接度增量算法将共享最近邻及孤立点进行社团聚类。

1) 社团中心节点搜索算法(ASCCN)

局部中心度(Local Centrality)能推断出隐藏的网络全局拓扑结构^[1], 而节点中心度很好描述了节点在网络中的处于中心的程度。搜索社团中心节点是社团划分正确率的关键控制点。本文提出的社团中心节点搜索算法(Algorithm of Searching Community Center Node, ASSCCN)是首先根据节点中心度确定第一个社团中心节点, 然后采取分裂法,



考虑其余节点本身的节点中心度以及该节点到已知中心节点之间的最短路径。这样可以避免两个距离相近的节点因为中心度高而同时成为不同社团的中心节点。社团中心节点最优评价方法可以由下式计算:

$$M_{\max} = \sum_{i=1}^n ((C_{v_i}^D)^{\alpha} * (d_G(v_i, v_j))^{\beta}) \quad (6)$$

其中 $\alpha + \beta = 1$, α, β 分别表示距离中心度和节点间最短路径的权重系数,它们之间权重分配可以根据情况进行调整。本文经过多次实验,得到较为合理的权重分配是 $\alpha = (1/3)$, $\beta = (2/3)$ 。 $d_G(v_i, v_j)$ 表示节点 v_i 到 v_j 的最短路径长度。考虑 $d_G(v_i, v_j)$ 计量单位的不同,在实际计算过程中需做标准化处理。

社团中心节点搜索算法(ASCCN)通过如下步骤实现:

输入: 社团数目 g , 权重阈值 α, β ;

输出: 社团中心节点集合 $C(g$ 个节点);

Step1: 找出网络中心度最大节点作为初始节点 v_0 加入社团中心节点集 C ;

Step2: 计算网络其余节点 M 值;

Step3: 找出 $M_{\max}$ 并将对应节点合并到初始社团中心集合 C 中;

Step4: 重复Step2~Step3, 直到找到个社团中心节点。

2) 基于中心度的最小链接度增量社团结构探测算法(MILDDC)

考虑复杂网络的小世界特征, 借鉴 l -shell算法和BB算法思想, 以各社团中心节点为核心, 以社团中心节点两两之间最短路径为直径, 将邻居节点进行聚类, 形成 g 个不完整社团, 然后对共享最近邻节点集和孤立点集并采用社团最小链接度增量算法, 完成社团结构的探测。

基于中心度的最小链接度增量社团结构探测算法步骤如下:

输入: G, g, α, β

输出: G 社团划分

Step1: 调用社团中心节点集搜索算法(ASCCN)找出中心节点集 C ;

Step2: 计算中心节点集合 C 最短路径 d ;

Step3: 将中心节点集 C 以 $\left[\frac{d}{2}\right]$ 为半径将邻居节点分别聚类到各自社团集合(共享最近邻除外);

Step4: 将共享最近邻节点划入集合 S ;

Step5: 将孤立点集节点划入集合 I ;

Step6: 计算集合 S 中节点与各社团链接度增量, 完成集合 S 中节点社团划分;

Step7: 计算集合 I 中节点与各社团链接度增量, 完成集合 I 中节点社团划分。

基于中心度的最小链接度增量社团结构探测算法可以弥补最近邻聚类的等半径扩散, 同时对包含噪声、孤立点以及任意形状网络结构具有较强的鲁棒性。

1.3 算法时间复杂度分析

给定无向连通图 $G=(V, E)$, 用 n, m 和 g 分别表示 G 中的节点、连边和社团划分数量。计算网络各节点中心度的时间复杂度为 $O(m)$, 生成 g 个社团中心节点集的时间复杂度为 $O(m)$ 。计算社团中心节点集合 C 最短路径时间复杂度为 $O(g^2)$, 计算社团中心节点最近邻时间复杂度变为 $O(gn)$ 。社团新增加一个节点的计算复杂度为 $O(d)$, d 为网络图的平均节点度。对集合 C 和集合 I 寻找社团时间复杂度为 $O(xd)$, x 为集合 C 和 I 中节点数。其中 d, g, x 相对 m 和 n 来说是个很小的数字。本算法的时间复杂度几乎接近于线性。

2 算法测试

为了验证本算法的性能和计算的准确性, 本文把MILDDC算法应用于航空复杂产品模块化生产网络中进行社团结构探测。用Python2.6编写算法程序,

在IBM笔记本电脑((CPU1.80GHz, 内存1.0GB, 硬盘160GB, Windows XP操作系统)上进行了实现。

2.1 飞机模块化结构及模块化生产网络模式分析

飞机制造装配过程是一个复杂的系统工程, 其制造工艺路线长, 涉及加工设备多, 加工处理的信息量大, 产品结构非常复杂, 一般来说, 大型民用飞机可分为机体、发动机和机载设备3个部分。所谓机体指大型飞机中主要用以装载旅客和货物、控制飞机姿态以及在地面支撑飞机的部分, 它又可具体分为机翼、机头、前机身、中机身、后机身、垂尾、方向舵、挂架、平尾、升降舵、雷达罩、起落架。而发动机按照机型配置的不同可以分为双发、三发和四发。机载设备包括航空电子系统、电源系统、空气管理系统、飞行控制系统、液压能源系统、防火系统、供水/水处理系统、氧气系统、雨刷系统以及客舱设备、装饰件等。

飞机的系统复杂性和模块可分性决定了模块化战略成为航空制造企业经营战略的必然选择。目前航空工业逐渐呈现出以大型航空企业为核心, 主系统承包商与分系统承包商以及零部件供应商紧密合作的产业格局。产业市场结构的特点是: 整机制造企业具有高度集中度, 是寡头垄断市场; 发动机等直接为整机配套的系统供应商处于相对集中状态, 是垄断竞争市场; 零部件配套企业相对分散。总装厂商在价值链上处于核心位置, 上、下游零部件供应或经销商依赖于总装厂商。随着市场的开放, 考虑市场的高度动态特性, 考虑模块化生产网络的分工合作以及价值链整体竞争优势, 总装厂商需要对业务流程重新进行设计, “剥离”部分零部件和总成件等非核心业务, 以便投入更多的资源专注于自身的核心竞争力的提升,

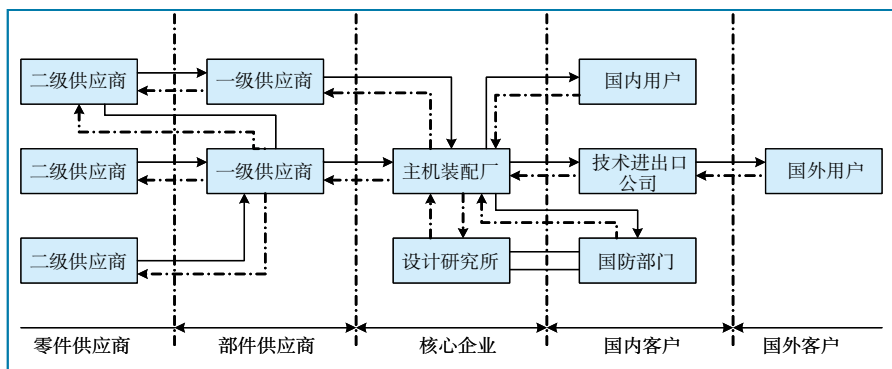


图1 飞机模块化生产运作模型



图2 某型飞机全球模块化生产网络结构图



图3 某型飞机模块化生产网络社团探测结果

打破了传统生产链流程,形成了模块化生产网络供应模式。具体表现为原来由总装厂商生产或装配的零部件转为由模块集成商以模块的形式直接进入总装配线。航空工业模块化生产网络由飞机总装厂商、零部件及原材料供应商、制造服务企业、生产性服务企业以及相关行业管理部门等构成。飞机模块化生产的运作模型如图1所示。

2.2 某型飞机模块化生产网络结构社团结构探测

通过实际调研,某型飞机拥有一级供应商25家,二级供应商98家,三、四级等供应商共197家。根据网络企业间零部件及模块供应等关系绘制该机型全球模块化生产网络结构图(图2)。网络拓扑结构包括312个节点,208条连线,网络统计特征呈现典型的无标度网络特性。

对该生产网络采用本章提出的MILDDC算法进行社团结构探测。社团结构探测结果如图3所示。从图3可以看出,整个模块化生产网络被分为6个社团。分析以机体结构供应商等为中心的飞机模块供应产业集群,该聚类结果符合航空工业产业布局特征。航空产业集群模式在我国已经取得了较好的成绩,例如,西安作为航空老工业基地,是国家的大中型飞机的研制生产基地,在人才、技术、环境配套方面的优势明显,且有陕西西安阎良国家航空高技术产业基地。江西南昌也规划建设航空城,希望建成以中航工业洪都集团和昌河飞机工业集团为核心的航空产业集群。

3 结论

本文在 l -壳算法和BB算法思想

基础上,结合Clusset算法的局部模块度量原则,提出了一种基于中心度的最小链接度增量社团结构探测算法(MILDDC)。MILDDC算法的时间复杂度约为线性时间,并且对包含噪声、孤立点以及任意形状网络结构具有较强的鲁棒性。将该算法应用于某型飞机模块化制造生产网络社团结构的探测,实验结果验证了算法具有一定的正确性和适用性。

AST

参考文献

- [1] Scott J, Social Network Analysis: A Handbook [M]. London: Sage Publications, 2nd Ed., 2002:7-10.
- [2] Costa L F. Hub-Based Community Finding[EB/OL].(2004-03-20), <http://arxiv.org/abs/cond-mat/0405022>.
- [3] Bagrow J P, Bollt E M. A Local Method for Detecting Communities [EB/OL]. (2007-06-25). <http://arxiv.org/abs/cond-mat/0412482>.
- [4] Clauset A. Finding Local Community Structure in Networks[J]. Physical Review E, 2005,72(2). <http://arxiv.org/pdf/physics/0605087.pdf>.
- [5] Wasserman S, Faust K. Social networks analysis: methods and applications [M]. Cambridge: Cambridge University Press, 1994:177-178.
- [6] 王立敏,高学东,武森. 基于最小社团链接度增量的社团结构挖掘算法[J]. 北京科技大学学报, 2009, (01):112-117.

作者简介

戴爱明,博士,副教授,主要研究方向为复杂网络、管理科学与工程;

肖灵机,教授,主要研究方向为企业管、技术创新。