

Matlab 在客户—供应商网络中的应用研究

Research on Application of Matlab in Customer-Supplier Networks

郎香香¹ 彭哲² 张婷婷³

Lang Xiangxiang Peng Zhe Zhang Tingting

(1. 厦门大学管理学院, 福建 厦门 361005 2 厦门大学王亚南经济研究院, 福建 厦门 361005 ;

3. 中国农业银行北京分行, 北京 100022)

(1.School of Management, Xiamen University, Fujian Xiamen 361005 ;

2.Wang Yanan Institute for Studies in Economics, Xiamen University, Fujian Xiamen 361005 ;

3.Beijing Branch, Agricultural Bank of China, Beijing 100022)

摘要 构建并度量客户—供应商网络对管理学和产业组织学的研究具有重要意义。由于缺乏相关技术的支持, 管理学一些领域的研究尚停留在理论阶段, 缺乏实证支持。本文通过引入图论和社会网络的相关技术, 以中国国家统计局提供的投入产出表数据为基础, 介绍了构建行业层面的客户—供应商网络的方法。鉴于复杂网络的相关指标难以进行手工度量, 本文介绍了采用 Matlab 软件计算网络集群系数和平均最短路径的程序和方法, 为后续实证研究奠定了基础。

关键词 客户—供应商网络, 投入产出表, 集群系数, 最短路径

中图分类号: O212 ; TP393 文献标识码: A 文章编号: 1671-4792(2014)04-0006-08

Abstract Constructing and characterizing a customer-supplier network has a great significance for the researches on management science and industrial organization. However, due to the lack of relative technical support, the research of related fields in management remains a theoretical target and lack of empirical support. By introducing graph theory and relative technology of social network and based on the data of input-output table provided by National Bureau of Statistics, this paper introduces a way to build the customer-supplier network at the industrial level. As it is difficult to manually measure the relative index of complex network, this paper describes a procedure and ways to calculate the computer network clustering coefficient and the average shortest path length by using Matlab, laying a basis for future studies.

Keywords Customer-Supplier Network, Input-Output Tables, Clustering Coefficient, Shortest Path

0 引言

客户—供应商网络的研究, 在管理学和产业组织学等领域有着重要的意义。例如, 管理学中的供应链理论认为, 供应链是由原材料零部件供应商、生产商、批发经销商、零售商、运输商等一系列企业组成的, 每家企业都是“链”中的一个节点。原材料零部件

依次通过“链”中的每个企业, 逐步变成产品, 产品则通过一系列流通配送环节, 最后交到最终用户手中 (刘丽文, 2013)。随着市场竞争的日趋激烈, 越来越多的企业意识到, 单靠自身的努力难以取胜, 有必要建立多个企业的“联盟”, 共同对抗其他“联盟”。而“盟友”的首选, 是与本企业业务相关联的上游或下

游企业,即供应链上的其他成员。如果“链”上的各企业未能相互合作,没有增强整个“链”的竞争力,这条“链”就会处在一种不合理的状态,导致各企业之间的交易成本居高不下,甚至使企业内部的管理绩效互相抵消。

此外,产业组织学中的垄断竞争理论也与供应商—客户关系密切相关。根据产业组织学中的买方抗衡势力理论,企业行业整合会形成一定的市场垄断,垄断企业会提升产品价格,压缩下游企业的利润空间。而下游的客户企业也会进行相应的行业整合,提高自身的购买力,以对抗供应商市场势力的上升。该理论虽然得到了很多学者的支持和认可,但是,对该理论的实证检验并不多。原因是,企业的客户—供应商网络错综复杂,给行业整合对上下游企业影响的研究造成了一定的阻碍。

缺乏供应商网络、客户网络的分析技术,一定程度限制了对相关理论的实证检验。为此,本文介绍了图论和社会网络的相关技术,详细说明了如何使用 Matlab 软件来衡量网络的相关指标,为后续的实证检验奠定了基础。本文的具体操作如下:首先根据中国国家统计局提供的《投入产出基本流量表》分别计算行业投入系数表和行业产出系数表;其次根据系数表计算客户行业与供应商行业的相邻矩阵;最后,根据相邻矩阵计算行业集中性程度、集群系数和平均最短路径的指标。

1 行业投入系数和产出系数的核算

投入产出表记录了每个行业的产出对其他行业产出的贡献,可以反映出行业之间的客户—供应商

关系。中国国家统计局每五年统计一次各行业的投入产出额,具体包括三张表格:《投入产出基本流量表(中间使用部分)》、《投入产出直接消耗系数表》和《投入产出完全消耗系数表》。本文主要采用的是《投入产出基本流量表(中间使用部分)》。需要注意的是,历次发布的投入产出表,所涉及行业略有不同。例如,2002年的投入产出表涉及122个行业,而2007年则涉及135个行业。

对于“投入产出基本流量表”,记第*i*行第*j*列所列内容为 A_{ij} 。其中*i*是行(供应商行业),*j*是列(客户行业)。 A_{ij} 表示客户行业*j*的产出所需的行业*i*的直接投入值(不含行业增加值)。

投入系数 I_{ij} 是用 A_{ij} 除以行业*j*的总投入(即总产出) A_{0j} (基本流量表最后一行的第*j*个元素)后所得的系数^①。 I_{ij} 代表的是客户行业*j*每一元的产出所需要的供应商行业*i*产出的金额,它显示了供应商行业*i*的投入对于客户行业*j*产出的相对重要性。

相应地可以计算产出系数 $O_{ij}=A_{ij}/A_{0i}$,即用 A_{ij} 除以供应商行业*i*的总产出 A_{0i} (基本流量表最后一列的第*i*个元素)后所得的系数。 O_{ij} 表示供应商行业*i*每一元的产出销售给行业*j*的金额,它显示了客户行业*j*对供应商行业*i*的销售的相对重要性。

为了直观地说明矩阵 $I=(I_{ij})$ 和 $O=(O_{ij})$ 的计算方法,这里以2007年的投入产出表为基础,截取了“石油和天然气开采业”及其相关行业的数据进行解释。表一列出了7个关联行业的投入产出基本

^①若不考虑行业增加值,则总投入与总产出的金额相等。

流量。

表一 投入产出基本流量表(A_{ij})

		客户行业 j							总产出
		1	2	3	4	5	6	7	
供 应 商 行 业 i	1	491,212	122,463	118,771	65,585	130,241	15,390	20,636	11,082,895
	2	127,991	10,999,316	3,595,764	8,208,650	19,025,709	2,043,369	661,108	179,150,123
	3	5,542,419	117,753,997	1,282,450	1,874	3,858,428	3,712,378	107,590	95,348,874
	4	6,344	22,631	154,535	1,435,930	26,812	16,125	101,323	27,283,871
	5	625	48,551	24,727	1,569	6,545,242	6,942,307	34,089,651	79,097,903
	6	1,998	38,650	505	2,266	661,304	9,013,366	642,400	43,092,022
	7	3,896	236,350	63,135	2,791	1,359,964	937,454	28,768,250	121,864,868
总投入		11,082,895	179,150,123	95,348,874	27,283,871	79,097,903	43,092,022	121,864,868	

注 :1、燃气生产和供应业 2、石油及核燃料加工业 3、石油和天然气开采业 4、航空运输业 5、合成材料制造业 6、化纤制造业 7、塑料制品业

数据来源：中国国家统计局 2007 年 135 个部门的投入产出表

根据表一可以计算投入系数表和产出系数表。用 A_{ij} 除以行业 j 的总产出(最后一行)就可以得到投入系数表(见表二) 除以行业 i 的总产出(最后一列)就可以得到产出系数表(见表三)。

表二 投入系数表(I_{ij})

	燃气生产和供应业	石油及核燃料加工业	石油和天然气开采业	航空运输业	合成材料制造业	化纤制造业	塑料制品业
燃气生产和供应业	4	0	0	0	0	0	0
石油及核燃料加工业	1	6	4	30	24	5	1
石油和天然气开采业	50	66	1	0	5	9	0
航空运输业	0	0	0	5	0	0	0
合成材料制造业	0	0	0	0	8	16	28
化纤制造业	0	0	0	0	1	21	1
塑料制品业	0	0	0	0	2	2	24

由表二可知 ,石油和天然气开采业是燃气生产和供应业以及石油及核燃料加工业的主要供应商行业 ,在后两者的总投入中所占的比例分别是 50%和 66%。石油及核燃料加工业是航空运输业和合成材料制造业的主要供应商行业 ,在这两个行业各自的总投入中所占的比例分别是 30%和 24%。合成材料制造业是化纤制造业和塑料制品业的主要供应商行业 ,在这两个行业中的投入比例分别是 16%和

28%。由投入系数表 ,我们可以找出“石油和天然气开采业→石油及核燃料加工业→合成材料制造业→化纤制造业”的供应商链条。

表三 产出系数表(O_{ij})

	燃气生产和供应业	石油及核燃料加工业	石油和天然气开采业	航空运输业	合成材料制造业	化纤制造业	塑料制品业
燃气生产和供应业	4	1	1	1	1	0	0
石油及核燃料加工业	0	6	2	5	11	1	0
石油和天然气开采业	6	123	1	0	4	4	0
航空运输业	0	0	1	5	0	0	0
合成材料制造业	0	0	0	0	8	9	43
化纤制造业	0	0	0	0	2	21	1
塑料制品业	0	0	0	0	1	1	24

由表三可知 ,在石油和天然气开采业的产品销售中 ,石油及核燃料加工业的采购比例是 123% ,这表明石油及核燃料加工业的需求量超过石油和天然气开采业的供应量 ,石油及核燃料加工业是石油和天然气开采业的主要客户。石油及核燃料加工业的总产出中 ,有 6%用于自身生产 ,5%由航空运输业采购 ,11%由合成材料制造业采购。因此 ,航空运输业和合成材料制造业是石油及核燃料加工业的主要客户。对于合成材料制造业 ,最大的客户行业是塑料制品业(采购比例是 43%) ,其次是化纤制造业(采购比例为 9%)。因此 ,我们可以根据产出系数表找出“石油和天然气开采业→石油及核燃料加工业→合成材料制造业→塑料制品业”的客户关系链。虽然这与供应商链条相似 ,但是两个关系链中上下游行业间关系的强度存在较大的差异。

相应的 Matlab 脚本为 :

```
% 导入数据 数据文件为 paper_data.xlsx
A = xlsread('paper_data.xlsx');
N_all = size(A,1);
```

```
% 选取相关行业的产出额
ind_id = [93; 37; 7; 100; 43; 47; 49; N_all];
% 相关行业的投入产出额(表一)
Amat = A(ind_id,ind_id);
N = size(Amat,1)-1;
% 计算投入系数和产出系数(表二和表三)
lmat = A (1:end-1,1:end-1)./repmat (A (end,1:
end-1),N,1);
Omat = A (1:end-1,1:end-1)./repmat (A (1:end-1,
end),1,N);
```

2 根据行业投入产出系数核算相邻矩阵

根据投入系数表和产出系数表构建投入产出的相邻矩阵,为后面网络指标的求解奠定基础。下面以投入系数表(如表二所示)为例,得到相邻矩阵:

第一步,用 0 替换对角元素。

第二步,若 $l = (l_{ij})$ 中任一非对角元素 l_{ij} 大于 1%, 则用 1 替换之^②。

第三步,将 $l=(l_{ij})$ 对称化,如若 $l_{ij}=1$, 则令 $l_{ji}=1$ 。这样,就得到如下相邻矩阵:

$$\begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 \end{pmatrix}$$

^②Fan & Lang(2000)、Fan & Goyal(2006)和 Ahern(2012)等在研究中认为,投入产出系数在 1%的行业存在纵向相关关系,因此,在构造投入产出网络中,本文只加入了投入产出系数大于 1%的行业进行配对,剔除了投入产出关系小于 1%的行业配对。

相应的 MATLAB 脚本为:

```
% 将投入系数矩阵转换为二元相邻矩阵 lmat2
%(1)第一步 将对角元素替换为 0
lmat2 = lmat;
lmat2(eye(N)==1)=0;
%(2)第二步 若元素大于 0.01 ,用 1 替换 ,否则 ,
定为 0
lmat2=(lmat2>0.01);
%(3)第三步 对称化、数值化
lmat2 = (lmat2 | lmat2')+0;
```

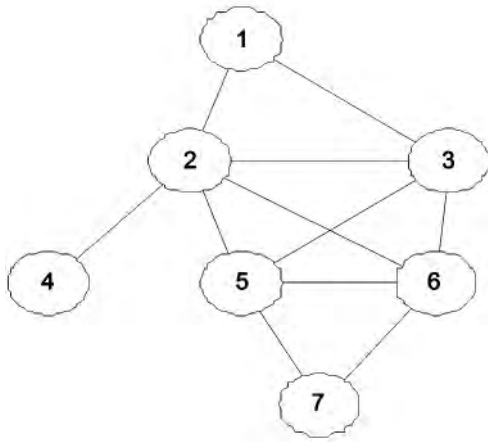
以上是根据投入系数表得出的相邻矩阵,用产出系数表计算相邻矩阵的方法与之类似,只需用 Omat、Omat2 分别替换上述语句中的 lmat 和 lmat2 即可。

3 客户—供应商网络指标的核算

3.1 客户—供应商网络结构图

得出相邻矩阵后,就可以根据相邻矩阵得到 7 个行业之间的供应网络图,如图一所示^③。由图一可知,仅仅是石油和天然气相关的 7 个行业,关联网络就已经较为复杂。若考虑全部的 135 个行业,得到的网络会更错综复杂。这时,用图形来表示行业之间的供应商关联,显然是项艰巨的任务。为此,接下来本文将引入图论和社会网络的相关技术进行分析。

^③这里绘制的“客户—供应商关联图”使用了 Ali Taylan Cemgil 等人编写的“图形版式生成包”(Graph Layout Generation Package); 相关程序的下载地址为 <http://www.cmpe.boun.edu.tr/~cemgil/matlab/graphlayout.zip>,具体说明请参阅: <http://www.cmpe.boun.edu.tr/~cemgil/matlab/layout.html>。在绘制图形时,可能需要调整行业之间的顺序,以便得到简明的关联图。



图一 石油相关行业的供应商—客户关联图

注 :1、燃气生产和供应业 2、石油及核燃料加工业 3、石油和天然气开采业 4、航空运输业 5、合成材料制造业 6、化学纤维制造业 7、塑料制品业

所用的 Matlab 程序为 :

```
% 将 GLGP 包加入路径
addpath .\GraphLayout
% 绘制关联图
draw_layout(Imat2)
```

3.2 客户—供应商网络指标核算

对供应链、产出网络的研究主要涉及客户—供应商网络的密集程度,企业与其客户、供应商关系的密切程度,以及企业在网络中的位置等。因此,本文主要介绍三个指标,以此核算网络密集度和客户—供应商关系的强度。这三个指标是:集群系数、平均最短路径和集中性程度。

(1)集群系数。按照 Watts & Strogatz(1998)的方法,对于行业而言,集群系数等于行业的关联行业

之间的实际关联数与理论最大关联数之比。由于实际关联数至多等于理论最大关联数,因此,集群系数的最大值是 1。

在“客户—供应商网络”中,行业的集群系数越高,行业的客户—供应商网络越紧密,行业中的企业进行供应链整合的重要性更高,难度也更大。从产业组织学的角度来看,处于网络核心位置行业进行行业整合的影响面更广,也更可能引发进一步的行业整合。

下面以图一为例介绍 Watts & Strogatz(1998)对集群系数的核算方法。在图一中,行业 1(燃气生产和供应业)与行业 2(石油及核燃料加工业)、行业 3(石油和天然气开采业)直接相关,对于 2~3 组成的子群,共有 1 个链条(或称“边”),而理论最大的链条数为 $C_2^2=1$,因此,行业 1 的集群系数为 1。

又例如,行业 2 与行业 1、3、4、5、6 都相关。而后 5 个行业之间有 4 个链条(1 与 3,3 与 5,3 与 6,5 与 6),因此,行业 2 的集群系数为 $4/C_2^5=0.2$ 。对其他行业也可采用类似的计算,通过计算,行业 1~7 的集群系数分别为 1、0.40、0.67、0、0.67、0.67 和 1。

这种手工计算只适用于节点较少的小网络,而对于复杂的网络,可以用 Matlab 软件来实现,算法即标准的聚类程序。

相应的 Matlab 脚本为 :

```
Mat=Imat;
C_coef=zeros(N,1);
```

```

For i=1:N
% 寻找子群的交点
joint=find(Mat(i,:)==1);
% 计算交点数
k=length(joint);
% 若交点数为 1, 说明子群只包含一个 i 行业
以外的行业 集群系数为 0
if k<=1
C_coef(i)=0;
% 若交点数大于 1, 说明子群内部有多条关联
% 此时 提取子群内相互关联的行业
else
Mat_sub=Imat(joint,joint);
% 集群系数 = 实际关联数 / 理论关联数
% 实际关联数 = 子群中为 1 的节点个数 / 2, 因
为 2 个点对应一条关联
% 理论关联数 =k*(k-1)/2
C_coef(i)=sum(sum(Mat_sub==1))/(k*(k-1));
End
End

```

(2)平均最短路径。计算各个行业到其他各行业的平均最短路径,可采用 Dijkstra(1959)的两步法:第一步,计算单个行业到网络中其他行业的最短路径;第二步,计算每个行业所有路径的均值。

行业的平均最短路径涉及的是行业在客户—供应商网络中的位置,行业与其他所有行业的最短路径的均值越小,行业越接近网络的核心位置,否则,则越接近网络边缘位置。两个行业之间的最短

路径越小,表明这两个行业的投入产出关系越密切。最短路径的最小值 0,代表的是两个行业是同一行业;其次是 1,表明两个行业之间存在直接的投入产出关系,即这两个行业的投入产出系数大于 1%。

行业之间的最短路径对于管理学的研究有着重要意义。例如,在供应链管理中,处理好与直接客户和供应商的关系可能对企业供应链整合业绩的影响更大。因为直接客户和供应商对于企业生产经营的重要性比间接客户和供应商的影响更为直接、迅速。在企业并购活动中,并购双方处于近距离行业的并购可能比远距离并购的并购绩效更好。因为,并购双方所属行业的投入产出关系在近距离行业并购中更为密切,这有利于并购后的资源整合。

根据以上的分析可知,行业到自身的距离为 0,相邻行业的最短路径记为 1,不相邻行业的距离可定为无穷大。以图一为例,行业 1 到行业 2 的路径有多条:一是直接到行业 2,也可经由关联行业 3 两步达到,也可经行业 3 到行业 5 再到行业 2,等等。前者是 1 步,后者需 2 步,其他路径则需要两步以上。因此,行业 1 到行业 2 的最短路径是 2。

平均最短路径度量的是行业 A 与其他行业的紧密程度,行业 A 到其他行业的平均距离越短,它与其他行业就越紧密。以行业 1 为例,它到行业 1~7 的最短路径分别为 0、1、1、2、2 和 3 步,因此该行业的平均最短路径为 $(0+1+1+2+2+2+3)/7=11/7=1.57$ 。

最后算得的 7 个行业的平均最短路径分别为: 1.57、1.14、1.71、1.14、1.14 和 1.71。

相应的 matlab 脚本为：

% 计算每个行业到其他所有行业的最短距离
矩阵

```
D = graphallshortestpaths(sparse(lmat2));  
D_avg=zeros(N,1);  
D_max=max(D(D~=Inf));  
for i=1:N  
D_i_all = D(i,:);  
% 用 D 的最大元素替换 D 第 i 行为 Inf 的元素  
D_i_all(D(i,:)==Inf)=D_max ;  
% 求各行业最短路径的平均值  
D_avg(i) = mean(D_i_all);  
end
```

(3)集中性程度。集中性程度显示了网络中某个行业在所有行业的投入产出流中的重要性。这种重要性取决于：与该行业有投入产出关联的行业数目，以及投入产出系数的大小。

用投入系数算得的集中性程度数值较大，表明该行业的产出对于其他行业的产出较为重要，即该行业是其他行业的重要的供应商行业。用产出系数算得的集中性程度数值较高，表明该行业的采购对于其他行业的销售重要性较高，即该行业是其他行业的重要客户行业。

集中性程度的计算方法较为简单，其定义为：某个行业的集中性程度等于该行业对于其他全部行业（包含本行业）的投入（或产出）系数之和。按照行是供应商，列是客户的矩阵（外向）集中性程度是行的

和。

3.3 企业层面的客户—供应商关联核算

根据《中国 2007 年投入产出表部门分类解释》和企业的主营产品资料，划分企业的投入产出行业分类，就可以核算企业层面的客户—供应商网络指标。例如，根据《中国 2007 年投入产出表部门分类解释》合成材料制造业的范围是：(1)初级形态的塑料及合成树脂制造也称初级塑料或原状塑料生产，包括通用塑料、工程塑料、功能高分子塑料的制造。(2)合成橡胶制造，指人造橡胶或合成橡胶及高分子弹性体的生产。(3)合成纤维单(聚合)体的制造，指合成纤维单体和合成纤维聚合物的生产。(4)其他合成材料制造。如果企业的主营产品是以上四类中的一类，就将该企业归为合成材料制造业。此外，还可以通过问卷调查，获取企业的实际客户—供应商数据，然后再进行网络分析。对此，本文介绍的网络指标和 Matlab 程序仍然适用。

当然，用行业层面的数据得到的客户—供应商网络不一定与企业实际的客户—供应商网络相符。但是，使用行业数据的优势在于：它不仅考虑了企业的实际供应商和客户，也考虑企业的潜在供应商和客户，从而有利于对相关理论作实证检验。例如，在研究并购收益在并购双方的分配时，考虑潜在供应商和客户，能更好地反映并购双方的“可替代性”（供应商—客户越多，可替代性越强）和竞争力。一般而言，可替代性强的企业在并购谈判中具有较低的议价能力，因而分配到的并购收益比例也较低。

4 结束语

传统的管理学和产业组织学研究,通常将客户—供应商关系视为简单的线性链条。但实际上,客户—供应商构成了复杂的关联网络。因此,要全面分析企业的客户—供应商网络对企业某些行为(如并购)的影响,就需要度量企业在产业网络中的位置。

本文提供了一些分析客户—供应商网络的方法和技术,介绍了这些度量方法在 Matlab 软件上的实现。这些分析方法为分析客户—供应商网络的密度和强度提供了帮助,为企业经营发展和战略决策影响的分析提供了帮助。

参考文献

[1]刘丽文.供应链管理思想及其理论和方法的发展过程[J].管理科学学报,2003,6(02):81-88.

[2]Ahern K R,Harford J.The importance of industry links in merger waves [J].The Journal of Finance,2014,forthcoming.

[3]Ahern K R. Bargaining power and industry dependence in mergers [J]. Journal of Financial Economics, 2012, 103(03):530-550.

[4]Dijkstra E W.A note on two problems in connexion with graphs [J].Numerische mathematik, 1959,1(01):269-271.

[5]Fan J P H,Lang L H P.The measurement of relatedness: An application to corporate diversification [J].Journal of Business,2000,73(04):629-660.

[6]Fan J P H,Goyal V K.On the Patterns and Wealth Effects of Vertical Mergers [J].The Journal of Business,2006,79(02):877-902.

[7]Watts D J,Strogatz S H.Collective dynamics of 'small-world' networks [J].Nature,1998,393(6684):440-442.

作者简介

朗香香(1985—),女,汉族,山东潍坊人,博士研究生,主要研究方向:公司财务、公司治理、兼并与收购;

彭哲(1984—),女,汉族,湖北荆门人,博士研究生,主要研究方向:应用计量经济学、中国经济;

张婷婷(1982—),女,汉族,湖北黄冈人,博士,主要研究方向:宏观经济、技术经济。