

赵 钢,周凌云,周桂良,等. 农产品流通网络拓扑结构统计特性[J]. 江苏农业科学,2013,41(11):417-421.

# 农产品流通网络拓扑结构统计特性

赵 钢<sup>1,2</sup>, 周凌云<sup>1</sup>, 周桂良<sup>1</sup>, 张 浩<sup>1</sup>

(1. 淮阴工学院交通工程学院, 江苏淮安 223003; 2. 南京航空航天大学民航学院, 江苏南京 210016)

**摘要:**从实证角度出发,以度分布为主,研究农产品流通复杂网络的拓扑特征。选取具有代表性的农业区域建立农产品流通复杂网络,考察该网络的拓扑结构特征,分析不同类型节点衔接农户的概率和农产品流通的链路选择特性,研究农产品流通网络节点度分布规律,测算标度指数,并对其进行柯尔莫诺夫-斯米尔诺夫检验。

**关键词:**度分布;农产品流通网络;幂律;标度指数

**中图分类号:**S11;N94 **文献标志码:**A **文章编号:**1002-1302(2013)11-0417-05

近些年来,我国农产品价格频繁剧烈波动,农产品流通体系与经济社会发展的矛盾日益激化。农产品流通系统是社会系统中的一个重要子系统,关系着国计民生与经济社会可持续发展。研究农产品流通系统的拓扑特征,有助于深入理解农产品流通网络的结构特性、动力学特征,有助于揭示农产品流通网络上商品、物资、信息、资金等流动介质的流通效率和传播规律,对于创新农产品流通方式、降低农产品流通成本、保障农产品供给、稳定农产品物价有着重要的意义。系统的内部结构是其功能特性最主要的影响因素。许多现实系统的结构可以用网络的形式表达出来,例如计算机网络、蛋白质结构网络、合作网络等。人们通过多年的研究,认识到自然事物或社会事物聚合形成的一定结构的自然网络、科技网络、社会网络并不是随机联系在一起的。许多现实的网络结构的某些结构特性具有惊人的相似性或一致性。这些规律特性的发现大大推动了复杂网络的研究,使网络科学成为研究复杂系统的主要工具之一。复杂网络的无标度概念提出之后<sup>[1]</sup>,人们对大量的实际网络拓扑特征进行了广泛的实证研究。这些实际网络的研究来源于社会领域网络<sup>[2-5]</sup>、信息领域网络<sup>[6-8]</sup>、技术领域网络<sup>[9-11]</sup>和生物领域网络<sup>[12-14]</sup>。复杂网络度分布特性是复杂网络拓扑结构的重要统计特征,对复杂网络的动力学性质及其功能有着极其深刻的影响。目前,国内外部分学者将复杂网络的研究成果应用到供应链研究中,但这些研究大多是针对制造业产品的供应链进行的。农产品流通网络的结构特点及其动力学特征与制造业商品流通网络存在很大的差异。目前,缺乏应用复杂网络科学的理论研究农产品流通网络结构特性的成果。笔者从实证研究出发,研究农产品流通复杂网络拓扑结构特性,结果显示农产品流通复杂网络符合幂律分布,具有无标度特征。笔者测算了农产品流通复杂网络的标度指数,获取了农户销售农产品的平均销售渠道宽度以及不同类型的流通节点衔接农户的概率和链路选择特性。这些研究揭示出我国农产品流通网络的重要结构特性,

为研究农产品流通网络的发展演化、农产品流通网络的流通效率和动力学特性提供了一定的研究基础。

## 1 农产品流通复杂网络构建

目前,我国已经确立了农产品市场化的流通体制,农业产业化企业、各类农民合作组织、国合商业组织、个体运销户以及经纪人等多种经济成分的市场流通主体共同发展,多形式的农产品流通渠道逐步形成。但是,随着我国经济社会的发展,现有的农产品市场流通体系已不符合经济社会进一步发展的要求,农产品价格频繁剧烈波动,经常出现滞销和买卖贵难现象,严重损害了农产品生产者以及农产品流通主体的利益,严重影响着城镇居民的生活,农产品流通体系的网络结构存在的问题是造成这一现象的主要因素。

笔者所在的课题组选取苏北区域农产品流通结构进行深入研究,因为苏北地区是我国农产品的重要生产区域,选择这一区域进行研究具有一定的代表意义。本研究以江苏省淮安市淮阴区袁集镇、新渡乡,清浦区城南乡、黄码乡以及江苏省盐城市建湖县建阳镇、蒋营镇等乡镇的调研结果为有效样本,绘制出农产品流通复杂网络图(图1)。

在农产品流通网络的实证研究中研究全局网络是不现实的,也是不可能的,往往只能考察实际网络的局部,为了区别实际全局网络而把所考察的网络称之为数据网络。在前述的农产品数据网络中不可避免地存在一些孤立的节点,但在实际的全局网络中是不可能存在孤立节点的。经过大量调研发现,在实际的农产品流通网络中,在一定时间长度内,度值为1的节点也是几乎不存在的。因此,为了消除数据内在噪声,笔者在进行数据分析时先后剥去网络0-壳和1-壳。

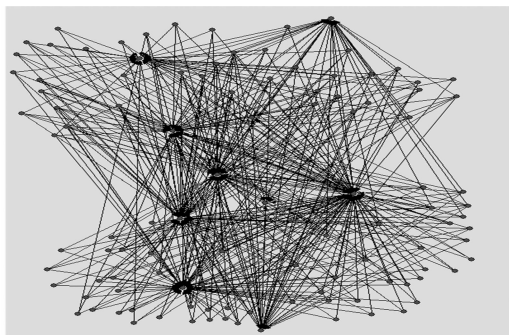
## 2 农产品流通网络的平均度特性

节点的度是刻画网络节点属性的最重要的参数,网络中所有节点的度的平均值是网络的平均度,记作 $\langle k \rangle$ 。节点平均度可以视作网络的0阶度分布特性<sup>[15]</sup>。给定网络G的邻接矩阵 $A = (a_{ij})_{N \times N}$ ,则有 $k_i = \sum_{j=1}^n a_{ij} = \sum_{j=1}^n a_{ji}$ ,  $\langle k \rangle = \sum_{i,j=1}^n a_{ij} = \frac{1}{n} k_i = \frac{2M}{n}$ ,其中M是网络的连边数。有向网络的节点的度包括出度和入度。节点入度 $k_i^{in} = \sum_{j=1}^n a_{ji}$ ,节点出度 $k_i^{out} = \sum_{j=1}^n a_{ij}$ 。

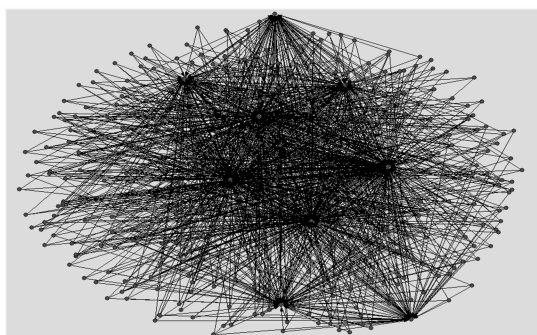
收稿日期:2013-04-21

基金项目:江苏省高校哲学社会科学基金(编号:2012SJB630006);江苏省社会科学基金(编号:10EYD031,10EYD029)。

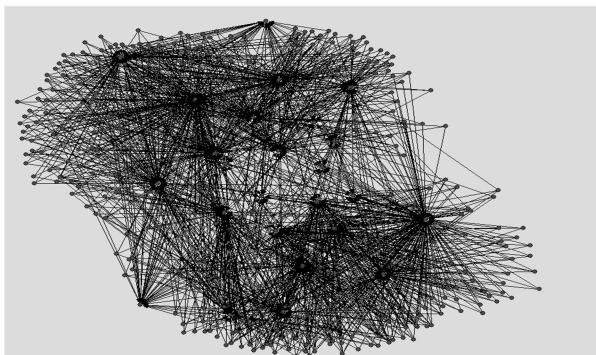
作者简介:赵 钢(1977—),男,山西忻州人,博士研究生,讲师,研究方向为供应链技术与系统管理。E-mail:zgzhaoang@163.com。



a. 网络 I (包括 100 个 A 类节点, 2 个 B 类节点, 1 个 D 类节点, 3 个 E 类节点, 2 个 F 类节点。由节点的实际地理空间方位结合 Kamada-Kawa 算法产生。)



b. 网络 II (包括 300 个 A 类节点, 3 个 B 类节点, 21 个 C 类节点, 2 个 D 类节点, 3 个 E 类节点, 2 个 F 类节点。由节点的实际地理空间方位结合 Kamada-Kawa 算法产生。)

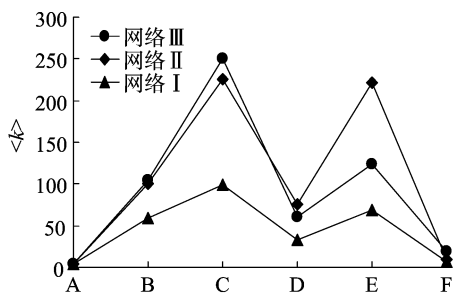


c. 网络 III (包含 500 个 A 类节点, 5 个 B 类节点, 2 个 C 类节点, 2 个 D 类节点, 10 个 E 类节点, 6 个 F 类节点。由节点的实际地理空间方位结合 Frucherman Reingolds 算法产生。)

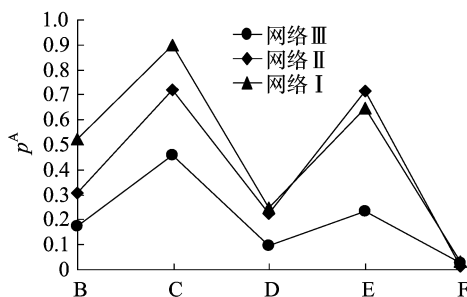
A 表示农户; B 表示收购商或农业产业企业及合作社; C 表示本地 (产地) 批发市场; D 表示销地 (外地) 批发市场; E 表示农贸市场及城市菜市场; F 表示超市及餐饮服务企业

图1 农产品流通复杂网络

如果将图 1 中的网络 I、网络 II 以及网络 III 看作无向网络, 网络 I 平均度  $\langle k^I \rangle = 8.65$ , 网络 II 平均度  $\langle k^{II} \rangle = 9.06$ , 网络 III 平均度, 因网络 III 中的 C 类节点和 B 类节点均比网络 II 与网络 I 多, 所以网络 III 的网络平均度略大于网络 II 和网络 I。农户的平均度对于农产品流通复杂网络研究具有特定的意义, 从它可以看出该农产品流通网络中农户平均销售渠道的宽度。网络 III 的 A 类节点即农户平均度  $\langle k^{A3} \rangle = 4.51$ , 网络 II 的 A 类节点平均度  $\langle k^{A2} \rangle = 4.26$ , 网络 I 的 A 类节点平均度  $\langle k^{A1} \rangle = 4.45$ , 3 个网络农户的度值相差不大, 由此可以推测在图 1 的数据网络中农产品生产者的平均度约为 4。因为在农产品流通复杂网络中, 作为农产品生产者的农户没有入边, 只有出边, 所以图 1 的数据网络中农户的农产品销售渠道平均宽度约为 4, 其中一个渠道是农产品批发市场, 一个渠道是收购商或农业产业企业, 平均有 2 个渠道是不同的农贸市场或城镇菜市场。如果将 3 个数据网络作有向网络处理, 可以看到网络 III 的平均入度和平均出度是相等的, 这不是数据子网的偶然结果, 而是有向网络的必然结果。在有向网络中尽管每个节点的入度和出度不尽相同, 但是网络的平均出度和平均入度是相等的, 这个平凡的事实揭示了复杂网络的一个简单而又普遍的涌现。农产品流通复杂网络中不同类节点的平均度以及邻接农户概率分析结果如图 2 所示。



a. 不同类节点的平均度



b. 不同类节点邻接农户概率

A、B、C、D、E、F 同图 1。

图2 流通网络中不同类节点的平均度和邻接农户概率

从图 2-a 可见,节点平均度最高的是 C 类节点,除农户类节点外,节点度值加和最大的是 E 类节点;从图 2-b 可见,平均邻接农户概率最高的是 C 类节点,次高的是 E 类节点,即 A 类节点倾向与 C 类和 E 类节点直接连接。由此可见,农户倾向于将农产品直接向本地批发市场和农贸市场销售的概率最大,C 类节点和 E 类节点对于所调查网络的农产品销售具有非常重要的作用。邻接 A 类节点概率最小的是 F 类节点,其次是 D 类节点;A 类节点到达 D 类节点大多经过 2 跳,或经过 B 类节点或经过 C 类节点。由此可见,农户的农产品经超市出售的概率很小,即农超对接比例非常小,农户的农产品直接向外市场销售的概率比较小,向外市场销售的农

产品大多经过收购商、企业或农民合作社及本地批发市场。

### 3 农产品流通网络度分布特性

#### 3.1 度分布情况

从网络中任选一个节点,它的度正好为  $k$  的概率称之为该网络的度分布。网络的度分布可以视为网络的 1 阶度分布特性,它刻画了网络中不同度的节点在网络中出现的概率。在分析农产品流通复杂网络的数据网络时是基于 1-核网络进行的,为了可以较准确地判断出农产品流通复杂网络的度分布规律,绘制了前述农产品供应网络度分布的线性、半对数、双对数坐标图(图 3)。

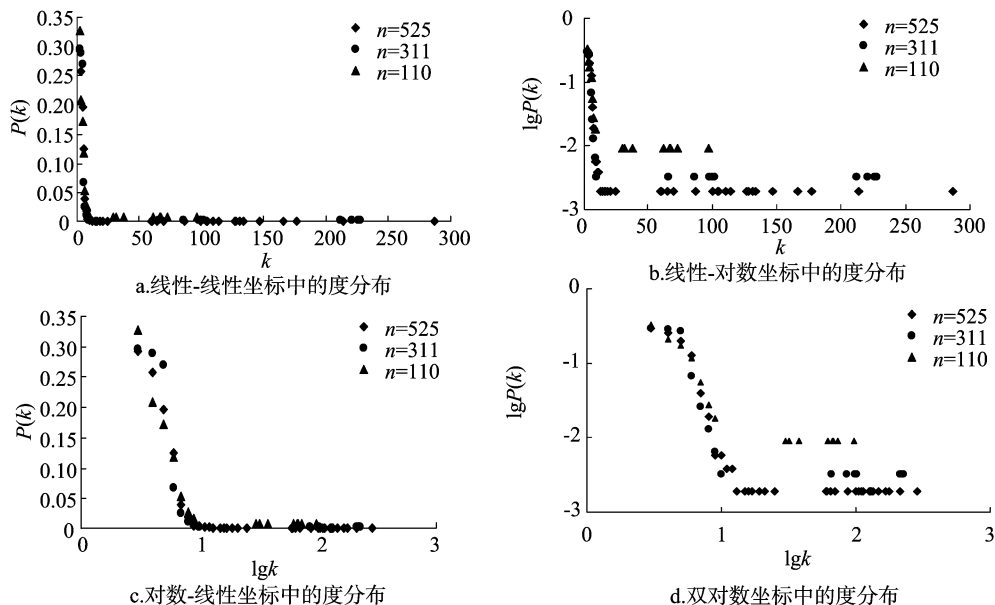


图3 农产品流通复杂网络的度分布

依据图 3 排除了指数分布、Gamma 分布、偏斜高斯分布、广延指数分布的可能性<sup>[16-17]</sup>。右下方的这些出现概率最小的数据点呈现很大的统计涨落。在左上方这些数据点中,因其出现频率大,所以每个数据都包含着大量的实证数据,具有最高的统计可信度,因此应该对拟合结果有关键作用。而右下方的这些出现概率小的数据点,每个包含极少的实证数据,大多仅包含 1 个实证数据,因此应该在拟合过程中少加考虑。对数据网络的度分布的拟合见图 4。

图 4 中的数据点包含了所考察网络的 97% 的节点,这些包含大量实证数据网络节点、具有最高统计可信度的数据点在双对数坐标中呈一条直线,而且排除了其他分布的可能性,这表明农产品流通复杂网络的绝大部分节点的度分布符合幂律分布。将原始数据按度大小排序并将秩次图绘制在双对数坐标中,结果显示,绝大部分数据点服从幂律分布,和前面的结论吻合。在实证研究中,幂律度分布会出现截断,实际网络中度最大的节点很有限,从而使得在度很高时,幂律分布也不成立。利用度分布图,运用最小二乘法求出标度指数,但是这种方法的缺点是度分布尾部摆动范围较大,标度指数不容易准确估计。

#### 3.2 度的补分布函数

如果随机变量服从幂律分布,会在双对数坐标下表现为

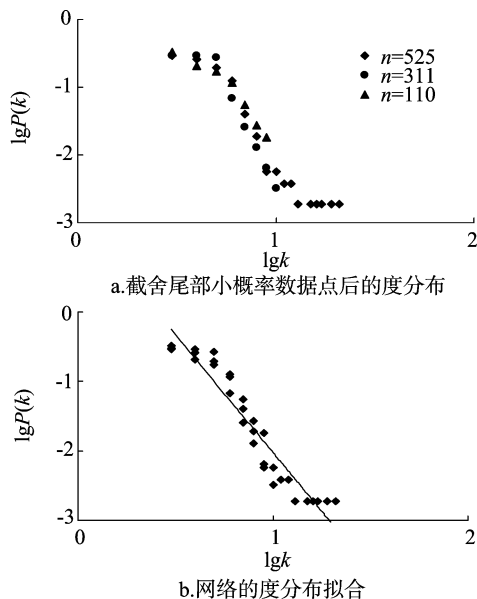


图4 农产品流通复杂网络的度分布拟合

一条直线,然而直线的尾部可能由于统计误差而产生波动,而解决尾部波动的一个方法是去掉尾部数据只拟合前半段,但这样做会丢失一部分信息。考察网络度分布的另一种办法是

对数据点进行装箱处理<sup>[18-19]</sup>,但不论是线性装箱还是对数装箱都属于粗粒化方法,仍然会产生统计噪声。利用补分布是一种较好的办法。使用补分布不需要进行装箱操作,避免了确定装箱宽度的问题,并且对所有数据有较好的使用效果。

对于离散变量有  $F(k) = \sum_{K \geq k} p(k)$ , 幂律分布的补分布  $F(k) = p(K \geq k) = \lambda \sum_{K \geq k} K^{-\gamma}$ ;  $\lambda \int_k^{\infty} d^{-\gamma} dk = \frac{\lambda}{\gamma-1} k^{-(\gamma-1)}$ 。由此可以看出,如果度的概率分布为幂律函数,那么补分布函数也是幂律函数,并且它们之间幂指数相差 1。采用补分布的方法不但可以避免分布描述中的平庸性,避免度分布中出现的扰动,减少的数据量不大,而且累积后的每个数据一般都包含了许多原来的数据,因此会在最大程度上减少统计涨落。

对前述的农产品流通复杂网络的度作补分布,结果如图 5 所示。补分布的一个重要缺陷是在尾部往往会偏离幂律,因此在拟合之前截掉尾部的数据,因为现实中的幂律分布很难在整个区间内都很好地服从幂律。利用图 5 的补分布数据序列拟合得到的补分布幂指数等于 1.78,所以利用补分布的方法得到的网络度分布的标度因子即标度指数等于 2.78。

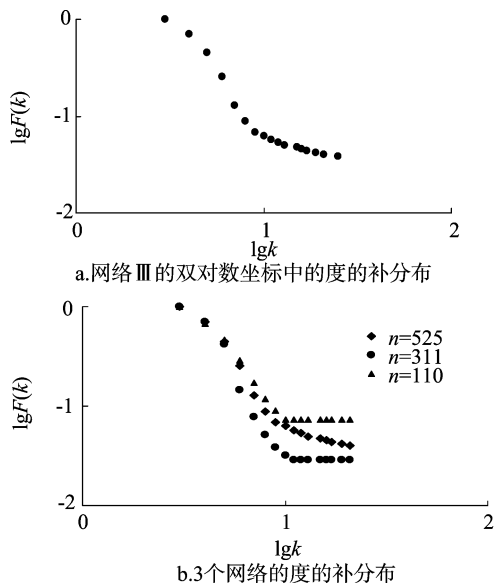


图5 农产品流通复杂网络的度补分布

### 3.3 标度参数的测算

通过对农产品流通复杂网络的度分布和度补分布的分析可知,农产品流通复杂网络的度分布在包含大量实证数据、具有最高统计可信度的数据点中符合幂律分布的特性。通过考察度分布和度的补分布,然后经过最小二乘法拟合得到标度指数估计值。但是,使用 OLS 时隐含 1 个条件,即默认实际数据的对数与理论值的对数之间的偏差服从正态分布,这一假设并不合理,所以 OLS 法产生的估计是有偏估计。所以,用前面 2 种方法估算标度指数并不是最好的方法,两者都是有偏估计。因此,可以在判断幂律分布的前提下,用极大似然法估计标度指数。极大似然法的明显优点在于只要样本量足够大,极大似然估计的参数和未知参数的真值可相差任意小,而且在一定意义上没有比极大似然估计更好的估计。

大部分实际网络的度分布在开始部分并不完全遵循幂律,都是在中部和一部分尾部遵循幂律,实际上幂律分布是当

才成立(当度值为 0 时概率无穷大),其中  $k_{\min}$  为  $k$  符合幂律部分的最小值。因此,对于概率函数  $P(k) = \lambda k^{-\gamma}$ ,做积分近似:  $\lambda: \frac{1}{\int_{k_{\min}}^{\infty} k^{-\gamma} dk} = (\gamma-1)(k_{\min})^{\gamma-1}$ , 则  $P(k): (\gamma-1)$

$$(k_{\min})^{\gamma-1} k^{-\gamma} = \frac{\gamma-1}{k_{\min}} \left(\frac{k}{k_{\min}}\right)^{-\gamma}。$$

如果可以考察到网络中若干节点的度值  $k_1, k_2, \dots, k_n$ , 并且是相互独立的,则似然函数为  $L(k_1, k_2, \dots, k_n; \gamma) = \prod_i \frac{\gamma-1}{k_{\min}} \left(\frac{k_i}{k_{\min}}\right)^{-\gamma-1}$ , 因其对数函数在相同的  $k$  值点达到最大值,所以

$$\ln L = \ln \prod_i \frac{\gamma-1}{k_{\min}} \left(\frac{k_i}{k_{\min}}\right)^{-\gamma-1} = n \ln(\gamma-1) - n \ln k_{\min} - \gamma \sum_i \ln \frac{k_i}{k_{\min}}。$$

$$\text{令 } \frac{\partial \ln L}{\partial \gamma} = 0, \text{ 求解 } \gamma, \text{ 得到极大似然估计的标度参数: } \hat{\gamma} = 1 + \left[ \sum_i \ln \left( \frac{k_i}{k_{\min}} \right) \right]^{-1}。$$

如果网络中满足幂律节点的最小度值和网络节点的平均度比较大,或者是度序列数据量巨大且度值序列大多是连续的情况下,采用上面的公式估计的标度指数和实际数值相差不大。但是,在很多情况下所考察数据网络的满足幂律分布的最小数据都不大,平均度也不大,数据量不是非常多且度序列不连续,在这样的情况下,采用公式  $\hat{\gamma} = 1 + \left[ \sum_i \ln \left( \frac{k_i}{k_{\min}} \right) \right]^{-1}$  计算的标度指数就会比实际值偏大,这是因为用积分对数据序列作了连续化的近似。

在离散情况下有  $\sum_{k=k_{\min}}^{\infty} k^{-\gamma} = \zeta(\gamma, k_{\min})$ , 其中  $\zeta(\gamma, k_{\min})$  是广义黎曼  $\zeta$  函数。因  $\lambda = 1/\zeta(\gamma, k_{\min})$ , 所以  $p(k) = \frac{k^{-\gamma}}{\zeta(\gamma, k_{\min})}$ ,

$$\text{似然函数为 } L(k_1, k_2, \dots, k_n; \gamma) = \prod_i \frac{j_i^{-\gamma}}{\zeta(\gamma, k_{\min})}。 \text{ 令 } \frac{\partial \ln L}{\partial \gamma} = - \sum \ln k_i - \frac{n}{\zeta(\gamma, k_{\min})} \frac{\partial \zeta(\gamma, k_{\min})}{\partial \gamma} = - \sum \ln k_i - \frac{n \zeta'(\gamma, k_{\min})}{\zeta(\gamma, k_{\min})} = 0,$$

其中  $\zeta'(\gamma, k_{\min})$  是广义黎曼 zeta 函数的偏导。所以,极大似然估计的标度指数  $\hat{\gamma}$  是方程  $-\sum \ln k_i - n \frac{\zeta'(\hat{\gamma}, k_{\min})}{\zeta(\hat{\gamma}, k_{\min})} = 0$  的解。

对方程数值进行求解的过程可以通过 Matlab 实现:给定网络的邻接矩阵,通过调用 Matlab 函数 Sum 可以获得每个节点的度数,然后调用 Matlab 函数 fsolve 进行求解,因为前面利用补分布的方法得到的网络度分布的标度指数等于 2.78,所以将解的初值设定在  $[2, 3]$  区间内,如果求解过程终止或者得到的最终解小于 0,则重新设定初值的估算区间后进行计算。

对于幂律分布,因为当变量趋于 0 时会偏离幂律,所以需要确定变量符合幂律区间的最小值。通过 Kolmogorov - Smirnov 估计,将补分布序列中符合幂律分布的最小值  $k_{\min}$  确定为 3。分别先后采用网络 III 的度补分布和 3 个网络综合数据点的补分布,计算得出所考察农产品流通复杂网络的度补分布的幂指数为  $\hat{\gamma} = 1.75$ , 然后利用补分布幂指数和概率函数的标度指数的关系,得到所考察的农产品流通复杂网络的标度指数为 2.75。由此可以看到,利用极大似然估计法得到的标度指数比利用度的补分布拟合得到的标度指数略小一

些,采用网络Ⅲ的补分布序列进行  $K-S$  检验。按照公式  $D = \max_{k \geq k_{\min}} |S(k) - P(k)|$  [ $S(k)$  表示补分布的实际值,  $P(k)$  表示补分布的理论值] 计算实际值和理论值的累积分布之间的最大距离。在数据网络的补分布中,采用补分布序列大于 20 组的数据点,计算得出  $D = 0.1012$ , 小于  $n = 20$  对应的分位数 0.1257, 也小于  $n = 30$  对应的分位数 0.1048。相应的分位数表可以参见文献[20]。因此可以认定该分布服从幂律分布,且网络的标度指数为 2.75。

#### 4 结论

本研究揭示出农产品流通复杂网络的度分布符合幂律分布,具有无标度特征,标度指数,表明农产品流通复杂网络的度具有有限均值,方差随着农产品流通网络规模的扩大而趋于发散。当地批发市场、销地批发市场、农民专业合作社以及农贸市场或农业产业企业这些度值高的节点会由于方差的发散而增加,但增加的速度因为均值的有限而极缓慢。A 类节点倾向与 C 类和 E 类节点建立链路,表明就目前市场情况下,农户倾向于将农产品直接向本地批发市场和农贸市场销售的概率最大,大多数农产品生产者选择批发市场、收购商和农贸市场进行销售的数量比运往外地市场和超市销售的数量大很多,农户的农产品经超市出售的概率很小,即农超对接比例非常小,农户的农产品直接向外地市场销售的概率比较小,向外地销售的农产品大多经过收购商、企业或农民合作社以及本地批发市场,C 类节点和 E 类节点对于农产品销售具有非常重要的作用。在建立现代高效农产品流通体系的过程中,农产品流通网络中的各类节点和网络结构的进化行为,以及标度指数为 2.75 的幂律特性对农产品流通复杂网络的动力学特征和农产品的流通效率的影响将是下一步的研究重点。

#### 参考文献:

- [1] Barabási A L, Albert R. Emergence of scaling in random networks [J]. Science, 1999, 286(5439): 509–512.
- [2] Davis G F, Greve H R. Corporate elite networks and governance changes in the 1980s [J]. American Journal of Sociology, 1997, 103(1): 1–37.
- [3] Newman M E J, Strogatz S H, Watts D J. Random graphs with arbitrary degree distributions and their applications [J]. Physical Review E, 2001, 64(2): 1–17.
- [4] Davis G F, Yoo M, Baker W E. The small world of the American corporate elite, 1982—2001 [J]. Strategic Organization, 2003, 1(3): 301–326.
- [5] Barabási A H, Ravasz E. Evolution of the social network of scientific collaborations [J]. Physica A, 2002, 311: 590–614.
- [6] Albert R H, Barabási A L. Diameter of the World Wide Web [J]. Nature, 1999, 401: 130–131.
- [7] Barabási A L, Albert B, Jeong H. Scale-free characteristics of random networks: The topology of the World Wide Web [J]. Physica A, 2000, 281(1/2/3/4): 69–77.
- [8] Jaffe A B, Trajtenberg M, Romer P M. Patents, citations and innovations: a window on the knowledge economy [M]. Cambridge: the MIT Press, 2002.
- [9] Latora V, Marchiori M. Is the Boston subway a small-world network? [J]. Physica A, 2002, 314(1/2/3/4): 109–113.
- [10] Cancho R F, Janssen C, Solé R V. Topology of technology graphs: Small world patterns in electronic circuits [J]. Phys Rev E, 2001, 64: 1–5.
- [11] Jordano P, Bascompte J, Olesen J M. Invariant properties in coevolutionary networks of plant-animal interactions [J]. Ecology Letters, 2003, 6: 69–81.
- [12] Jeong H, Tombor B, Albert R, et al. The large-scale organization of metabolic networks [J]. Nature, 2000, 407: 651–654.
- [13] Guelzim N, Bottani S, Bourguin P, et al. Topological and causal structure of the yeast transcriptional regulatory network [J]. Nature Genetics, 2002, 31(1): 60–63.
- [14] 汪小帆, 李翔, 陈关荣. 网络科学导论 [M]. 北京: 高等教育出版社, 2012.
- [15] Shang M S, Lu L Y, Zhang Y C, et al. Empirical analysis of web-based user-object bipartite networks [J]. EPL, 2010, 90(4): 1–5.
- [16] Zhou T, Wang B H. Modelling collaboration networks based on non-linear preferential attachment [J]. International Journal of Modern Physics C, 2007, 18(2): 297–314.
- [17] White E J, Jessica L. On estimating the exponent of power-law frequency distributions [J]. Ecology, 2008, 89(4): 905–912.
- [18] Goldstein M L, Morris S A, Yen G G. Problems with fitting to the power-law distribution [J]. European Physical Journal B, 2004, 41(2): 255–258.

**更正:**《江苏农业科学》2013 年第 41 卷第 10 期 401–404 页所刊论文《土地资源约束对经济增长的影响及对策研究——以苏州市为例》,漏登了基金项目:教育部人文社会科学青年基金(编号 12YJC790019);江苏省高校哲学社会科学研究基金(编号:2012SJB790044);江苏省苏州市哲学社会科学研究课题(编号:12-C-22)。特此更正,并向作者和读者致歉。

(《江苏农业科学》编辑部)