arXiv:2304.01737v1 [physics.soc-ph] 4 Apr 2023

有向加权网络和马尔可夫链的高阶同配性

Alberto Arcagni[1] Roy Cerqueti[2,3] Rosanna Grassi[4,*]

[1] 罗马智慧大学MEMOTEF系，Via del Castro Laurenziano, 9 - 00161，意大利罗马
alberto.arcagni@uniroma1.it

[2] 罗马智慧大学经济与社会科学系，Piazzale Aldo Moro, 5 - 00185，意大利罗马
roy.cerqueti@uniroma1.it

[3] GRANEM，昂热大学，SFR CONFLUENCES，F-49000 昂热，法国

[*] *Corresponding author*

[4] 米兰比可卡大学统计与定量方法系，Via Bicocca degli Arcimboldi, 8 - 20126，意大利米兰

rosanna.grassi@unimib.it

摘要

本文提出了一类适用于加权有向网络的新型同配性度量。我们扩展了经典的 Newman 的度-度同配性，将度以外的节点属性纳入考量。此外，我们提出通过长度大于一的有向路径连接节点，从而得到高阶同配性。我们将这些度量应用于贸易网络这一典型案例，并进行了实证分析。重要的是，我们展示了这一全局网络指标与马尔可夫链状态的自相关密切相关。

关键词：网络、高阶同配性、随机游走、马尔可夫链。

1 引言

非对称相互作用是自然界许多复杂系统的特征。有向网络是表示这些复杂情形的合适工具。尽管这类系统在现实世界的各种现象中显著存在，但有向网络的形式化表示仍未得到应用，这可能是因为这类模型在数学处理上存在困难。事实上，一些网络指标在有向框架下并不容易处理；有时，在某些情况下，针对有向网络的网络度量仍缺乏恰当的定义。从这一角度出发，本文聚焦于为有向网络提出同配性度量的广义概念。这个主题具有重要意义，因为同配性是一种全局指标，能够为了解网络结构提供有用见解。在 Newman（2002）的经典定义中，同配性由一种全局度量表示，该度量基于节点度之间的皮尔逊相关。该度量的取值范围为 $[-1, 1]$，在同配网络中取正值，在异配网络中取负值。度-度同配性度量的含义很简单：在无向网络中，同配性越高，高度节点越倾向于连接其他高度节点。相反，在异配网络中，高度节点则倾向于连接低度节点。同配性度量可以沿两个不同方向扩展。一方面，可以考虑节点的其他定量属性，而非度；另一方面，可以不再局限于节点间的邻接关系——这是 Newman 的度-度同配性的基础——转而提出更一般的节点连接方式。本文致力于满足这两项重要要求。事实上，在前一种情况下，基于度的同配性对节点间的一致性提供了有限的视角。将其推广到不同的节点属性，可以克服这一问题，扩展该系数的含义，使其能够描述节点之间除度之外的相似性或差异性（参见 Meghanathan（2016）、Arcagni 等人（2019）、Yuan 等人（2021））。在后一种情况下，值得一提的是 Arcagni 等人（2017）提出的 Newman 同配性指数推广形式。作者给出了一个闭式公式，为无向无权网络的同配性提供了统一方法；具体而言，他们考虑的顶点对不一定彼此相邻，也可以通过……相连

路径、最短路径和随机游走。重要的是，通过长度大于一的路径连接节点，使我们能够洞察彼此相连但并不相邻的节点之间的相似性。事实上，高度节点可能彼此没有直接连接，但仍可通过低度节点保持同等程度的连接。

在 Arcagni 等人（2017）中，基于路径的新型同配性系数概念被称为高阶同配性。对于某些特定情形，高阶同配性能够通过对控制连接的矩阵作适当定义，得出 Newman 指数以及其他基于同配性的度量（Estrada 等人，2008；Mayo 等人，2015；Van Mieghem 等人，2010）。沿着这一思路，Arcagni 等人（2019）将高阶同配性扩展为一类专用于加权无向网络的高阶度量，并在特定情形下得出 Leung 和 Chau（2007）提出的加权同配性指数。有向网络中的同配性问题也已得到研究（Yuan 等人，2021）。具体而言，Foster 等人（2010）和 Piraveenan 等人（2010）提出了对 Newman 公式的一种自然改进，根据节点入度和出度的组合，将同配性度量编码为四种有向度量。这些系数量化了入度较高的节点倾向于与入度较高的节点相连的程度；其他情形也同样适用，即入度与出度、出度与入度，以及出度与出度的组合。然而，目前仍缺少一种适用于加权有向网络的高阶同配性度量。

本文通过填补这一空白，为这场讨论作出了贡献。我们针对加权有向网络的一般情形，引入一种新的 h –与 $h \geq 1$ –阶同配性度量概念。根据既有文献，并在必要时，我们将这一概念简称为高阶同配性度量。本文通过纳入加权有向网络，完善了 Arcagni 等人（2017，2019）提出的定义。此外，所提出的概念建立在节点属性之上，该属性不一定是度或强度。由此，我们在上述两个方向上扩展了 Newman 同配性系数。

值得注意的是，所提出的同配性度量在马尔可夫链领域具有重要的解释意义（有关这类随机过程的综述，参见 Norris 等人，1998年

）。事实上，有向加权网络特别适合用于描述在由节点标识的离散状态集上随机流动的情况。在马尔可夫链的图表示中，连接状态/节点 i 与状态/节点 j 的弧，意味着从 i 跳转到 j 的概率非零。因此，复杂网络中弧的权重（经过适当归一化后）自然对应于在相应马尔可夫链中观察到从一个状态流向另一个状态的概率。在阐述复杂网络与马尔可夫链之间的联系时，我们沿用了文献中一些重要研究所开辟的路径。值得一提的是 G´omez 等人（2010年），他们将复杂网络与马尔可夫链结合起来研究疾病传播模型。在相同的应用背景下，Iannelli 等人（2017年）探讨了源自马尔可夫链理论的一般基于网络的度量所包含的信息。

我们的创新之处在于，将高阶同配性——一种涉及相邻节点以外节点的网络拓扑性质——与齐次离散时间马尔可夫链联系起来。

具体而言，我们考虑相对于给定顶点中心性度量的节点同配性。 h 阶同配性指数的取值范围为 $[-1,1]$；当通过长度为 h 的游走连接的节点在所考虑的中心性度量上具有较高的一致性（分别为不一致性）时，该指数趋近于 1（分别为-1）。这样的取值范围表明，在马尔可夫链的背景下， h 阶同配性指数可以解释为滞后 h 阶的自相关项。这正是我们的情形。事实上，阶数h 就是从一个节点到另一个节点的游走长度。如果我们设想每条连边耗时一个时间单位，那么所考虑的游走就耗时 h 个时间单位。从这个角度看，阶数 h可视为离散时间马尔可夫链中的一个时间变量；该链的状态由网络节点及其中心性度量构成，转移概率则与网络连边的权重适当相关。

对阶数为 h 的同配性度量的这种解读，使我们能够自然地搭建起复杂网络与随机过程之间的桥。由此，我们得以从网络的高阶同配性所包含的信息，转向其底层马尔可夫链的动力学性质。具体而言，网络的时间维度

以及自相关所捕捉到的规律——这些都隐藏在网络结构中——在转向马尔可夫链理论时便清晰显现。

本文结构如下。第 2 节介绍数学符号。第 3 节提出加权有向图中高阶同配性的一种新定义。第 4 节展示所提出指标在贸易网络中的应用。第 5 节说明如何在有限状态的离散时间齐次马尔可夫链背景下解释高阶同配性度量。结论见第 6 节。

2 预备知识

本节概述与有向复杂网络相关的一些基本定义和符号，这些内容将用于介绍我们的方法学方案。

关于更详细的论述，请参阅 Bang-Jensen 和 Gutin（2008）；Newman（2010）；Gross 和 Yellen（2003）。

有向图（或有向图） $N = (V,\ A)$ 是由集合 V 和 A 组成的有序对，其中 V 是有限顶点（或节点）集合，A 是 m 中顶点构成的有序对（弧）集合；若（$i,\ j$）和/或（$j,\ i$）属于 A，则顶点 i 和 j 相邻。有向游走是由顶点和弧组成的序列，从 i 到 j，其中所有弧的方向相同。

可以为每条弧（$i,\ j$）赋予权重 $w_{ij} > 0$，从而得到加权有向图。此外，假设 $w_{ij} = 0$ 当且仅当（$i,\ j$）$\notin A$。权重 w's 收集在一个实 n 阶方阵 $\mathbf{W}$ 中（即加权邻接矩阵）。根据定义，这种矩阵的元素同时描述了 V 中顶点之间的邻接关系以及弧上的权重。在此背景下，加权有向网络是一个有向图 $N =$（$V,\ A$），并配备一个加权非对称矩阵 $\mathbf{W}$。在无权情况下，非零权重只能取单位值，因此矩阵$\mathbf{W}$ 仅提供邻接关系的信息。对于无权网络，我们将$\mathbf{W}$ 记为 $\mathbf{A}$ （即邻接矩阵）。

矩阵 $\mathbf{A}$ 的 $k-$ 次幂的（$i,\ j$）元素，是从 i 到 j 的长度为 k 的有向游走数量。

由于 i n 度、出度、入强度和出强度是最常用的中心性度量——

在同配性的语境下，正如 Newman（2002）的原始定义所示，我们为这些度量引入一种特定记法，并给出其定义。

我们将节点 $i \in V$的入度 d_i^I（相应地，出度 d_i^O）定义为指向（相应地，从） i出发的弧的数量。于是，入度向量和出度向量分别为$\mathbf{d}^I$和$\mathbf{d}^O$。顶点的度向量表示为$\mathbf{d} = \mathbf{d}^I + \mathbf{d}^O$。类似地，文中给出了节点 $i \in V$的入强度和出强度（分别为s_i^I 和s_i^O）以及入强度向量和出强度向量（分别为$\mathbf{s}^I$ 和$\mathbf{s}^O$）的定义。具有 $d_i^I = 0$ 和 $d_i^O > 0$ 的顶点 $i \in V$称为源点，因为它是其出发弧的起点。类似地，具有 $d_i^O = 0$ 和 $d_i^I > 0$ 的顶点 $i \in V$称为汇点 ，因为它是每条入弧的终点。

3 高阶同配性的一种新定义

我们考虑加权有向网络顶点的两种中心性度量，并汇总其取值，以此定义加权有向图的高阶同配性，从而将 Arcagni 等人（2017，2019）提出的相关概念扩展为皮尔逊相关指数在 n维向量$\mathbf{x}$ 和$\mathbf{y}$中的一种特定应用。对于有向图，与$\mathbf{x}$ 和$\mathbf{y}$ 相关、长度为 $h \geq 1$ 的高阶同配性指数可定义如下

$$r_h = r(\mathbf{x}, \mathbf{y}, \mathbf{E}_h) = \frac{\mathbf{x}^\top \left(\mathbf{E}_h - \mathbf{p}_h \mathbf{q}_h^\top\right) \mathbf{y}}{\sqrt{\left[\mathbf{x}^\top \left(\mathbf{D}_{\mathbf{p}_h} - \mathbf{p}_h \mathbf{p}_h^\top\right) \mathbf{x}\right] \left[\mathbf{y}^\top \left(\mathbf{D}_{\mathbf{q}_h} - \mathbf{q}_h \mathbf{q}_h^\top\right) \mathbf{y}\right]}}, \tag{1}$$

其中$\mathbf{E}_h$ 是一个归一化的 n平方矩阵，其第 i行、第 j列的元素记为 $e_{ij}^{(h)}$，其中 $\|\mathbf{E}_h\|_1 = \sum_{i,\ j=1}^{n} |e_{ij}| = 1$ 而$\mathbf{p}_h = \mathbf{E}_h \mathbf{1}$、$\mathbf{q}_h = \mathbf{E}_h^\top \mathbf{1}$，其中$\mathbf{1}$ 是元素均为 1 的相容向量，而$\mathbf{D}_{\mathbf{p}_h}$ 和$\mathbf{D}_{\mathbf{q}_h}$ 分别是以向量$\mathbf{p}_h$ 和$\mathbf{q}_h$为主对角线元素的对角矩阵。

特别地，我们假设$\mathbf{E}_1 = \mathbf{W}/\|\mathbf{W}\|_1$；~~正如~~我们将在第 5.2 节看到的，矩阵$\mathbf{E}_h$也与马尔可夫链 $\forall\, h \geq 1$有关。因此，方程（1）中的同配性度量取决于适当的中心性度量$\mathbf{x}$ 和$\mathbf{y}$，它们分别与一个具有 h 步的随机过程的初始状态和终止状态相关。

根据 Newman（2002）的观点，我们指出，在处理同配性时，节点的入强度和出强度也发挥着重要作用，这一点在

加权有向图中同样如此。事实上，这些中心性度量提供了有关网络拓扑结构的不同信息。因此，我们详细描述这些情况，同时也阐释高阶同配性度量在这一特定背景下的含义。

具体而言，为了全面研究加权有向网络的高阶同配性度量，我们考虑入强度和出强度的所有可能组合，并从公式 (1) 出发，定义加权有向图的四种高阶同配性指数。

1. $r_h^{OI} = r(\mathbf{s}^O, \mathbf{s}^I, \mathbf{E}_h)$（出-入）

   对于通过长度为 h 的有向游走相连的节点，该系数使用公式 (1) 计算出强度向量与入强度向量之间的相关性。具体而言，较高的 r_h^{OI} 值表示这些顶点的出强度与入强度高度一致，而接近 -1 的 r_h^{OI} 值则表示高度不一致。

2. $r_h^{IO} = r(\mathbf{s}^I, \mathbf{s}^O, \mathbf{E}_h)$（入-出）该指数衡量经过长度为 h 的有向游走后，起始顶点的入强度与到达顶点的出强度之间的一致性和不一致性。

3. $r_h^{II} = r(\mathbf{s}^I, \mathbf{s}^I, \mathbf{E}_h)$（入-入）该系数用于评估在存在长度为 h 的有向游走时，起始顶点的入强度与结束顶点的入强度之间的相关性。

4. $r_h^{OO} = r(\mathbf{s}^O, \mathbf{s}^O, \mathbf{E}_h)$（出-出）该高阶同配性指数描述了通过长度为 h的有向游走相连时，起始顶点的出强度与终止顶点的出强度之间的相关性。

注意，给定 h，不一定有 $r_h^{OI} = r_h^{IO}$，因为索引取决于矩阵 $\mathbf{E}_h$ 所描述的有向游走分布，而该矩阵不一定是对称的。

4 H世界贸易网络中的高阶同配性 k

为说明我们定义的优势，我们展示了高阶同配性在贸易网络中的一个应用。

数据涉及国家之间交易的商品，贸易额以美元计[1]。我们构建了出口和进口加权网络，其中顶点代表国家，弧上的权重表示2021年222个国家之间进口或出口商品的贸易额。请注意，顶点数量多于公认国家的数量，因为 ComTrade 还纳入了几乎无人居住的地区。因此，下文中的“国家”一词采用广义定义，也包括这些地区。

关于国家之间的连边，预计出口网络与进口网络互为补充，因为从 i 到 j 的进口应等于从 j到 i的出口。但实际情况并非如此，因为数据记录的是报告方国家正式申报的与其贸易伙伴之间的贸易，而部分申报缺失或存在不一致。即使考察2021年以前的数据，这一问题仍然存在。

这里我们采用了与 Antonietti 等人（2022）类似的方法，构建了一个汇总进口和出口数据的对称贸易网络。不过，与 Antonietti 等人（2022）不同，我们保留了网络中的弧方向，因此构建的是一个有向网络。

更具体地说，设 $\mathbf{W}_I$ 和 $\mathbf{W}_E$ 为加权邻接矩阵，用于表示各国之间进出口贸易的规模。因此，贸易网络定义为 $N = (V, \mathbf{W})$，其中矩阵 $\mathbf{W}$ 的元素 w_{ij} 为:

$$w_{ij} = \max(w_{ij}^E, w_{ji}^I),$$

分别为矩阵 $\mathbf{W}_E$ 和 $\mathbf{W}_I^\top$的 w_{ij}^E 和 w_{ji}^I 第 $ij-$ 个元素。这样，非对称矩阵 $\mathbf{W}$ 表示从国家 i 到国家 j的商品流。值得注意的是，在构建 $\mathbf{W}$ 时，我们选择进口和出口中的较大值。实际上，与 Antonietti 等人（2022）的观点一致，我们假设申报值越高，所包含信息的质量就越高。

[1]数据可免费获取并下载，详见联合国商品贸易统计数据库（2022）)

4.1 网络主要特征分析

首先，我们分析一些经典的网络特征。

顶点数为 $n = 222$。令$\mathbf{A}$ 为将$\mathbf{W}$ 中弧的权重忽略后得到的邻接矩阵。网络密度定义为弧的总数与潜在弧的最大数量之比[2] $(n-1)n$。

因此，密度为

$$\frac{\|\mathbf{A}\|_1}{(n-1)n} = \frac{\sum_{i,j=1}^{n} a_{ij}}{(n-1)n} = 0.4996.$$

尽管人们可能会认为贸易网络是完整的，但本例并非如此。这是因为存在一些小国家和地区，它们是边缘节点，可能只报告了部分出口或进口。

中心性度量的分布，尤其是度和强度的分布，提供了其他重要信息。顶点的度和强度分别表示一个国家的贸易往来次数和规模，由此可以了解贸易的多样化程度和强度。图1展示了入度和出度以及入强度和出强度的边际分布。入强度和出强度采用以10为底的对数刻度进行图示。

“入”度量与“出”度量的分布十分相似，但值得强调度和对数强度分布的一些特征：

1. 在 $x-$ 轴上，对应国家数量的度值为数百，而强度值范围为 10^6 至 10^{12};

2. 对数强度分布呈钟形;

3. 度分布呈两极分化：超过半数的国家与少于100个其他国家相连，而十分之一的国家几乎与所有其他国家相连。

正如我们稍后将看到的，这一特征对于寻找中心国家和外围国家的簇具有重要意义。

现在，我们通过四种中心性度量比较各国：入度、出度、入对数强度和出对数强度。

[2] N 是一个没有自环的有向网络

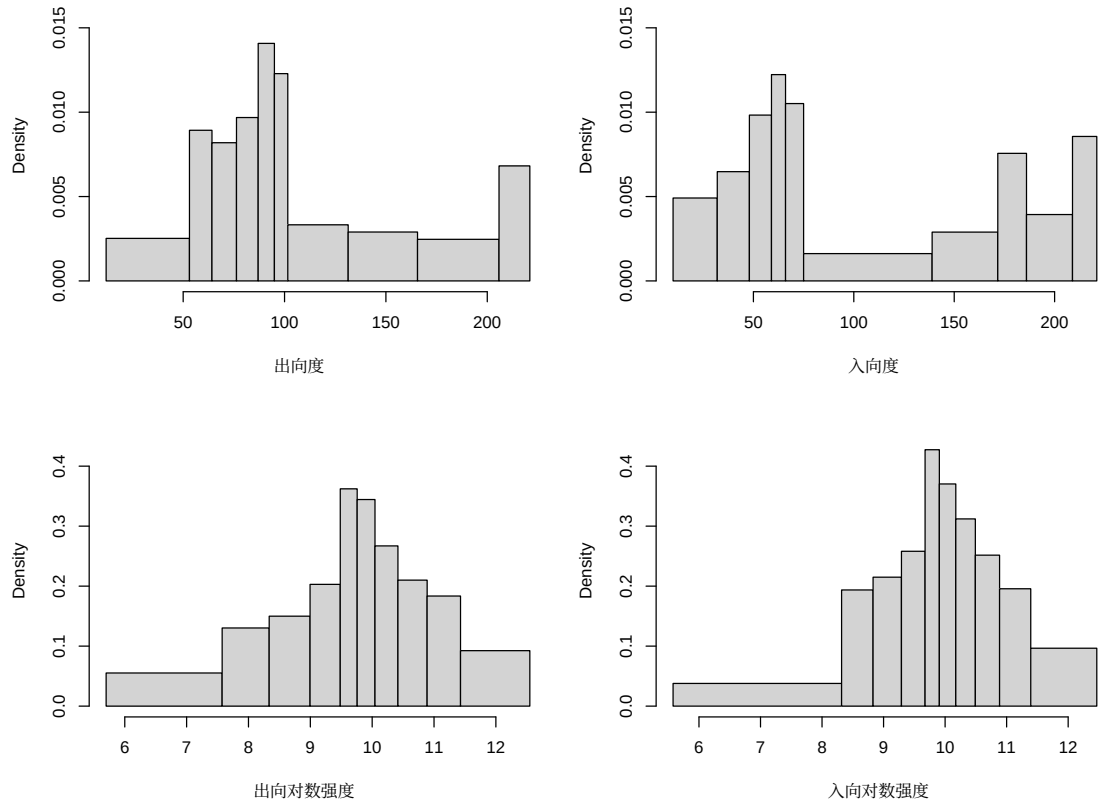

图1：所考察的中心性度量的边际分布。直方图的组别代表十分位数。

表 1 显示了所考察变量之间的强相关性（使用皮尔逊系数计算）。

正如预期，“入向”和“出向”方向高度相关（度和对数强度的相关系数分别为 0.883 和 0.947 ）。不过，度和对数强度中心性之间也存在很强的线性 相关性，数值介于 0.771 和 0.792 之间。尽管相关性并不代表因果关系，但这些较高的数值表明，连接数量的增加会使体量呈指数增长。

现在，我们分析标准化变量的主成分。第一主成分代表了整体变异性的 87%（将前两个主成分合在一起考虑时，变异性上升至 96%），且其载荷全部为正。因此，其得分可用于按照中心性对各国家进行一种综合且粗略的排序。排名前 10 位的中心国家按递增顺序为：加拿大、比利时、日本、意大利、英国，

	入度	出度	入对数强度	出对数强度
入度	1			
出度	0.883	1		
入对数强度	0.792	0.785	1	
出对数强度	0.742	0.771	0.947	1

表 1：中心性度量之间的相关矩阵。

荷兰、法国、德国、中国和美国。

我们使用四个主成分而非相关性较强的原始变量进行了层次聚类分析。我们计算了国家之间的欧几里得距离，并使用完全法定义组间距离。

层次方法建议分为两组。其中一组包含 116 个小型领土和国家，它们在体量方面几乎处于边缘地位；另一组由其余 106 个国家组成，控制着大部分体量。我们将后一组称为“主网络”。

图 2　　直观展示了贸易网络，并呈现出

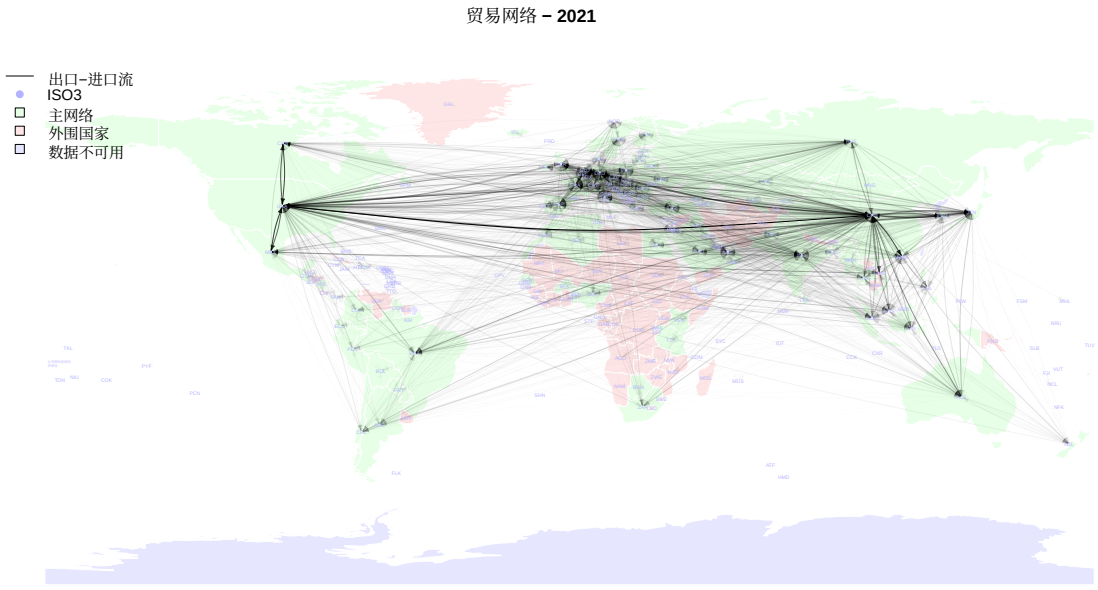

图 2：贸易网络的世界地图。各国根据所属聚类分别以绿色和粉色标示。

聚类分析得出的群组，即主网络和外围国家。然而，这些群组之间的关系仍有待探究。正如我们将在下一小节看到的，同配性可以从中心性相关的角度给出答案。

4.2 高阶同配性

显然，计算高阶同配性系数时，可以采用中心性度量与邻接矩阵的多种组合。本节仅关注在世界贸易[3]背景下能够对结果作出较直观解释的情形。

需要注意的是，在有向网络中，对于公式 (1) 定义的同配性系数，向量 $\mathbf{x}$ 与$\mathbf{y}$并不对称。事实上，在本应用中，$\mathbf{x}$ 和$\mathbf{y}$ 是中心性度量；正如第 3 节所指出的，所得系数取决于我们分别用$\mathbf{x}$ 和$\mathbf{y}$表示的出中心性和入中心性。我们对所用中心性度量作如下经济学解释:

- 出向中心性度量供给，即顶点 i i 所代表的国家 i 对商品的出度和出强度。
- 入向中心性度量需求，即顶点i 所代表的国家 i对商品的入度和入强度。
- 边的方向表示出口，即，从 i 指向 j 的箭头表示从国家 i 到国家 j的出口。

鉴于上述中心性度量的经济含义，本应用中最合适的有向同配性是出入同配性 r_h^{OI}，以便有效追踪国家间商品交换的路径。

图3展示了随着 h 增加而计算得出的高阶直接同配性，分别基于无权网络和加权网络。在这两种情况下，该网络在阶数 $h = 1$上呈异配；这意味着连接数较多的出口商，以及

[3]我们还计算了所有其余情形下的高阶同配性度量，结果可按需提供。

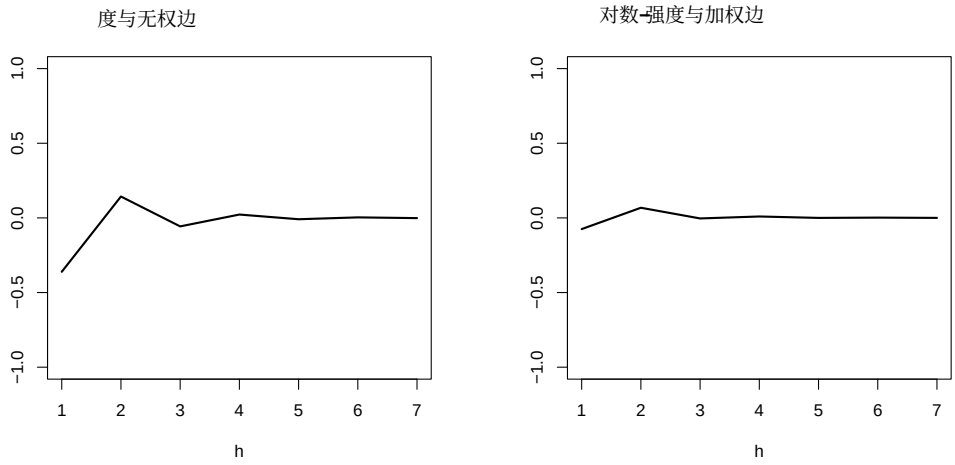

图3：无权（左）和加权（右）网络的进出口高阶同配性。

交易量较大的出口商往往与规模较小的进口商进行贸易。请注意，同配性的绝对值很快就低于 0。5，这是因为大型贸易商也需要彼此进行贸易（例如，美国和中国），或许小型贸易商之间也会如此。

在阶数为 $h = 2$ 时，同配性变为正值。无向网络也观察到了类似行为（见 Arcagni 等人，2017、2019），这可能是阶数为 $h = 1$时异配行为的结果。事实上，如果出口商和进口商在第一步倾向于属于不同群体，那么在第二步中，他们就倾向于属于同一群体。这种异配行为与同配行为之间的交替，分别会在奇数阶和偶数阶持续出现。

考虑权重时，由于表 1 所报告的中心性度量之间存在相关性，高阶同配性的行为也类似。不过，考虑权重后，我们观察到变化幅度显著减小。事实上，弧上的权重将通过聚类分析识别出的主网络与其余部分区分开来，而贸易主网络几乎是完全网络，这正如预期。

最后，我们观察到，随着阶数增加，网络中商品预期流量的随机性也随之增加，因此同配性趋于 0。

5 基于马尔可夫链的高阶同配性解释

本节从具有有限状态的离散时间齐次马尔可夫链的转移概率出发，对高阶同配性度量所包含的信息进行形式化表述。作为前提，我们先介绍马尔可夫链的一些基本定义和记号。有关完整论述，读者可参阅 Norris 等人（1998），例如。

5.1 马尔可夫链：基本定义与符号

具有有限状态的齐次离散时间马尔可夫链是一个随机过程 $\mathcal{X} = (X(t) : t \in \mathbb{N})$，满足以下条件。

$(P1)$ 存在一个含有 n 个元素的集合 V，使得对每个 $t \in \mathbb{N}$，$X(t) \in V$。V 的元素是 $\mathcal{X}$ 的状态，而集合 V 是该马尔可夫链的状态空间。

$(P2)$ 马尔可夫性质成立，即：

$$P(X(t+1) = j | X(t) = i, X(t-1) = i_{t-1}, \ldots, X(0) = i_0) = P(X(t+1) = j | X(t) = i),$$

for each $t \in \mathbb{N}$ and $i, j, i_0, \ldots, i_{t-1} \in V$.

$(P3)$ 一步转移概率不随时间变化，即：

$$P(X(t+1) = j | X(t) = i) = P(X(1) = j | X(0) = i),$$

对于每个 $t \in \mathbb{N}$ 和 $i, j \in V$。我们将所有一步转移概率收集在一个n阶方阵 $\mathbf{P}$ 中，其一般元素为 $p_{ij} = P(X(1) = j | X(0) = i)$。$\mathbf{P}$ 是马尔可夫链 $\mathcal{X}$ 的转移概率矩阵。

正如我们将在下文看到的，网络的节点集合 V 与马尔可夫链的状态空间相同。

性质（$P1$）–（$P3$）使我们能够通过三个要素来识别一个马尔可夫链：状态空间 V、转移概率矩阵 $\mathbf{P}$ 和初始概率分布

$\pi_0 = (\pi_0(1) \ldots, \pi_0(n))$, 其中 $\pi_0(i) = P(X(0) = i)$, 对每个 $i \in V$ 都成立。特别地, 条件（$P3$）表明马尔可夫链 $\mathcal{X}$ 是齐次的。

此外, 性质（$P3$）可以推广到对每个 $h \in \mathbb{N}$而言不随时间变化的 h步转移概率情形。直接计算可得

$$P(X(t+h) = j|X(t) = i) = P(X(h) = j|X(0) = i), \tag{2}$$

对每个 t, $h \in \mathbb{N}$和 i, $j \in V$均成立。我们将 h步转移概率汇集在n阶方阵$\mathbf{P}_h$中。该矩阵第 i行、第 j列的元素为$p_{ij}^{(h)} = P(X(h) = j|X(0) = i)$。

由于状态空间 V 是有限的，矩阵$\mathbf{P}_h$ 是$\mathbf{P}^h$，即 h步转移概率矩阵是一步矩阵$\mathbf{P}$的 h 次幂。

5.2 马尔可夫链背景下的高阶同配性

给定 $h \geq 1$，我们回顾一下：$\|\mathbf{E}_h\|_1 = 1$；因此，参照马尔可夫链理论，通用元素 $e_{ij}^{(h)}$ 可表示联合概率，即 i 和 j 分别是在时间 t 和 $t+h$时 $\mathcal{X}$的位置状态，对每个 t均成立:

$$e_{ij}^{(h)} = P(X(h) = j, X(0) = i). \tag{3}$$

考虑满足（3）的马尔可夫链 $\mathcal{X}$，并根据条件概率的定义，可以写出

$$\mathbf{E}_h = \mathbf{D}_{\pi_0}\mathbf{P}^h \tag{4}$$

其中 $\mathbf{D}_{\pi_0}$ 是一个对角矩阵，阶数为 n 的平方，其对角线元素为初始分布 π_0 的各个成分。由于矩阵 $\mathbf{P}$，且 $\mathbf{P}^h$ 在 $h > 1$ 下按行随机，$\mathbf{P}^h\mathbf{1} = \mathbf{1}$，因此:

$$\mathbf{E}_h\mathbf{1} = (\mathbf{D}_{\pi_0}\mathbf{P}^h)\mathbf{1} = \mathbf{D}_{\pi_0}(\mathbf{P}^h\mathbf{1}) = \mathbf{D}_{\pi_0}\mathbf{1} = \pi_0 \tag{5}$$

于是 $\pi_0 = \mathbf{p}^h$, $h \geq 1$，其中 $\mathbf{p}^h$ 是公式 (1) 中出现的向量。事实上，由于它们是 $e_{ij}^{(h)}$ 联合概率，根据方程 (3):

$$\sum_{j=1}^{n} e_{ij}^{(h)} = \sum_{j=1}^{n} P(X(h) = j, X(0) = i) = P(X(0) = i) = \pi_0(i) \, \forall i = 1, ..., n$$

也就是说，$\mathbf{p}_h$ 是初始概率分布。

显然，一般而言，并不存在唯一的、满足式 (4) 的状态空间为 V 的马尔可夫链 $\mathcal{X}$。然而，根据式 (5)，设定 $\mathbf{E}_1 = \mathbf{w}\,\|\mathbf{w}\|_1$ 后，向量 π_0 是唯一的。因此，一旦选定网络 $N = (V, A)$，马尔可夫链 $\mathcal{X} = (X(t) : t \in \mathbb{N})$ 就是唯一的，其中 V,$\mathbf{P}$ 和 π_0 均为固定值。

请注意，对于网络中作为汇点的每个节点 $i \in V$，矩阵 $\mathbf{P}$ 必须是随机矩阵；因此，在主对角线上添加 1，以便在马尔可夫链中为该汇点关联一个吸收态，并且 $\pi_0(i)$ 必须为 0，以免改变网络的拓扑结构，并确保对 $h = 1$ 满足式 (4)。相反，若 i 不是汇点，则 $\pi_0(i) > 0 \forall i \in V$，因为无论$\mathbf{E}_1$ 中的 $\mathbf{P}$取何值，式 (4) 都会返回一行零值。

与公式 (5) 类似，

$$\mathbf{E}_h^\top \mathbf{1} = (\mathbf{D}_{\pi_0}\mathbf{P}^h)^\top \mathbf{1} = (\mathbf{P}^h)^\top \mathbf{D}_{\pi_0}\mathbf{1} = (\mathbf{P}^h)^\top \pi_0 = \pi_h \tag{6}$$

于是 $\pi_h = \mathbf{q}_h$, $h \geq 1$，其中 $\mathbf{q}_h$ 是公式 (1) 中出现的向量。同样，根据公式 (3):

$$\sum_{i=1}^{n} e_{ij}^{(h)} = \sum_{i=1}^{n} P(X(h) = j, X(0) = i) = P(X(h) = j) = \pi_h(j) \,\forall j = 1, ..., n$$

也就是说，$\mathbf{q}_h$ 是经过 h 步后的概率分布。

考虑到这一点，我们可以在马尔可夫链的语境下重写高阶同配性指数（定义见公式 1）：

$$r_h = r(\mathbf{x}, \mathbf{y}, \mathbf{D}_{\pi_0}\mathbf{P}^h) = \frac{\mathbf{x}^\top \left(\mathbf{D}_{\pi_0}\mathbf{P}^h - \pi_0\pi_h^\top\right)\mathbf{y}}{\sqrt{\left[\mathbf{x}^\top \left(\mathbf{D}_{\pi_0} - \pi_0\pi_0^\top\right)\mathbf{x}\right]\left[\mathbf{y}^\top \left(\mathbf{D}_{\pi_h} - \pi_h\pi_h^\top\right)\mathbf{y}\right]}}, \tag{7}$$

那么，长度为 h 的高阶同配性指数表示马尔可夫链在时间距离 h 处的自相关，其计算对象不是状态 $v_i \in V$，而是它们经过变换后的结果$\mathbf{x}$ 和 $\mathbf{y}$。事实上，在网络语境下，中心性度量可以是定义在节点集 V 上的实函数的像。

5.3 一些示例

我们给出一些玩具示例，以说明新引入的同配性度量在马尔可夫链中的作用。为此，我们考察一些加权网络实例，其相关的马尔可夫链具有有意义的状态分类。下面详细介绍。

我们考察四个网络 $N_k = (V, \mathbf{W}_k)$，其中 $k = 1, \ldots, 4$，共享同一组节点 $V = \{a, b, c, d, e\}$，其加权邻接矩阵 $\mathbf{W}_k$ 定义如下：

$$\mathbf{W}_1 = \begin{pmatrix} 0 & 2 & \mathbf{2} & 2 & 0 \\ 396 & 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 & 396 \\ 0 & 2 & \mathbf{2} & 2 & 0 \end{pmatrix}, \quad \mathbf{W}_2 = \begin{pmatrix} 0 & 2 & \mathbf{196} & 2 & 0 \\ 396 & 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 & 396 \\ 0 & 2 & \mathbf{196} & 2 & 0 \end{pmatrix}$$

$$\mathbf{W}_3 = \begin{pmatrix} 0 & \mathbf{2} & 198 & 0 & 0 \\ 0 & 0 & 0 & 100 & 0 \\ 150 & 0 & 0 & 0 & 150 \\ 0 & 200 & 0 & 0 & 0 \\ 200 & 0 & 200 & 0 & 0 \end{pmatrix}, \quad \mathbf{W}_4 = \begin{pmatrix} 0 & \mathbf{198} & 198 & 0 & 0 \\ 0 & 0 & 0 & 100 & 0 \\ 150 & 0 & 0 & 0 & 150 \\ 0 & 200 & 0 & 0 & 0 \\ 200 & 0 & 200 & 0 & 0 \end{pmatrix}$$

网络如图4所示。通过观察该图，值得注意的是，N_1 和 N_2 具有相同的拓扑结构，仅有两个权重不同。具体而言，从 c没有流出；也就是说，c 是一个汇点。然而，在前一种情况下，c 的流入具有显著较低的权重，而在后一种情况下，c 的流入则特别高。就权重的绝对值而言（即考虑矩阵$\mathbf{W}_1$和 $\mathbf{W}_2$），这一证据成立；从相对角度来看也是如此（即考虑下方所示的转移概率矩阵$\mathbf{P}_1$ 和$\mathbf{P}_2$ ）。

此外，N_3 和 N_4 具有相同的拓扑结构，并且节点类 $\{b, d\} \subset V$没有流出。在这两种情况下，都可能通过从 a到 b的弧，使来自集合 $\{a, c, e\}$的流入进入该节点类。该弧在网络 N_3中的权重较低，而在 N_4中的权重则显著较高。

现在我们来考察中心性度量。在表 2 中，我们报告了每个网络的入强度$\mathbf{s}_k^I$ 和出强度$\mathbf{s}_k^O$分布, $k = 1, ...,4$以及平均值。请注意，由于求和的结合律，入强度和出强度的平均值相同。

我们根据相应中心性度量的平均值，将节点分为大节点或小节点。从表 2 可以看出，在网络 N_1中，节点 a 和 e 是大的接收者，但属于小的传播者。相反，b 和 d 是小

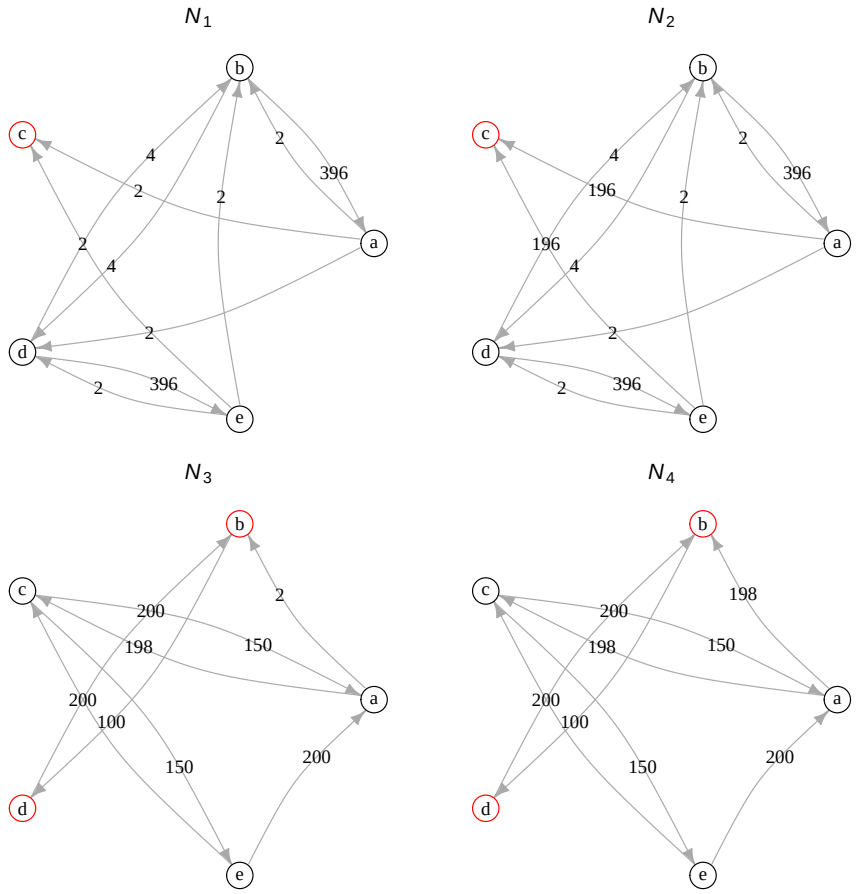

图4：四个示例的图形表示。

V	$\mathbf{s}_1^I$	$\mathbf{s}_1^O$	$\mathbf{s}_2^I$	$\mathbf{s}_2^O$	$\mathbf{s}_3^I$	$\mathbf{s}_3^O$	$\mathbf{s}_4^I$	$\mathbf{s}_4^O$
a	396	6	396	200	350	200	350	396
b	8	400	8	400	202	100	398	100
c	4	0	392	0	398	300	398	300
d	8	400	8	400	100	200	100	200
e	396	6	396	200	150	400	150	400
mean		162.4		240		240		279.2

表 2：每个示例的入强度$\mathbf{s}_k^I$和出强度$\mathbf{s}_k^O$ $k = 1, \ldots, 4$。

接收者、大规模传播者和节点 c 是边缘节点。在网络 N_2中，节点中心性与网络 N_1中的相同，只有节点 c 例外，它变成了一个大规模

接收者。在网络 N_3中，吸收类 $\{b,d\}$ 中的节点是边缘节点，而节点a 是一个大型接收者，但只是小规模传播者。节点 e 是一个小型接收者，但却是大规模传播者，而 c 是中心节点。网络 N_4 中的中心性只有节点 a 不同，该节点也变成了大规模传播者，因此成为中心节点。因此， b 受到这种变化的影响，成为一个大型接收者。

与这些示例中的网络相关的齐次马尔可夫链具有相同的状态空间 $V=\{a,b,c,d,e\}$。我们将与网络N_k关联的马尔可夫链记为 $\mathcal{X}_k$，其中 $k=1$、.....、$\mathbf{4}$。 $\mathcal{X}_k$ 的转移矩阵和初始概率分布可由邻接矩阵 $\mathbf{W}_k$ 得到，并分别记为$\mathbf{P}_k$ 和 $\pi_{0、k}$。下面列出这些转移矩阵。

$$\mathbf{P}_1=\begin{pmatrix} 0 & 0.\bar{3} & 0.\bar{3} & 0.\bar{3} & 0 \\ 0.99 & 0 & 0 & 0.01 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0.01 & 0 & 0 & 0.99 \\ 0 & 0.\bar{3} & 0.\bar{3} & 0.\bar{3} & 0 \end{pmatrix}, \quad \mathbf{P}_2=\begin{pmatrix} 0 & 0.01 & 0.98 & 0.01 & 0 \\ 0.99 & 0 & 0 & 0.01 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0.01 & 0 & 0 & 0.99 \\ 0 & 0.01 & 0.98 & 0.01 & 0 \end{pmatrix}$$

$$\mathbf{P}_3=\begin{pmatrix} 0 & 0.01 & 0.99 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0.5 & 0 & 0 & 0 & 0.5 \\ 0 & 1 & 0 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 & 0 \end{pmatrix}, \quad \mathbf{P}_4=\begin{pmatrix} 0 & 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0.5 & 0 & 0 & 0 & 0.5 \\ 0 & 1 & 0 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 & 0 \end{pmatrix}$$

由于等式（5），初始概率分布与出强度成正比。它们可写为:

$$\begin{aligned} \pi_{0,1}^{\top} &= [0.008, 0.493, 0, 0.493, 0.007] \\ \pi_{0,2}^{\top} &= [0.1\bar{6}, 0.\bar{3}, 0, 0.\bar{3}, 0.1\bar{6}] \\ \pi_{0,3}^{\top} &= [0.1\bar{6}, 0.08\bar{3}, 0.25, 0.1\bar{6}, 0.\bar{3}] \\ \pi_{0,4}^{\top} &= [0.284, 0.072, 0.215, 0.143, 0.287]. \end{aligned}$$

马尔可夫链的状态 $\mathcal{X}$ 可以轻松分类。 c 是$\mathcal{X}_1$ 和 $\mathcal{X}_2$ 的吸收态，状态集合 $\{b,d\}$ 是马尔可夫链 $\mathcal{X}_3$ 的吸收类,

$\mathcal{X}_4$。根据相关网络，被 c 和$\{b,d\}$ 吸收的概率分别代表区分 $\mathcal{X}_1$ 与 $\mathcal{X}_2$ 以及 $\mathcal{X}_3$ 与 $\mathcal{X}_4$的元素。

现在，我们将考察同配性，即上述马尔可夫链各状态对应的出强度与入强度之间的相关性。

h	N_1	N_2	N_3	N_4
1	0.7718	0.0828	0.5404	0.5503
2	−0.7664	−0.0739	0.5204	−0.0309
3	0.6031	0.0507	0.3515	0.2663
4	−0.6031	−0.0470	0.5067	−0.0130
5	0.5086	0.0361	0.4034	0.1688
6	−0.5550	−0.0339	0.4551	−0.0236
...	...	...	...	...
9995	0	0	0.0098	0.0718
9996	0	0	−0.0098	−0.0718
9997	0	0	0.0098	0.0718
9998	0	0	−0.0098	−0.0718
9999	0	0	0.0098	0.0718
10000	0	0	−0.0098	−0.0718

表 3：四个示例中 r_h^{OI} 的首次和末次评估值。

图 5 展示了网络 N_1 和 N_2的高阶同配性出-入指标 r_h^{OI}。在 N_1 中，进入吸收态 c 的概率很小；因此，我们可以认为其影响微乎其微。节点 a 和 e 的出强度较小、入强度较大，与之相反，节点 b 和 d则具有相反的特征。我们将节点对 $\{a,e\}$ 确定为第 1 组，将节点对 $\{b,d\}$ 确定为第 2 组。在第一步 $h=1$，同配性主要源于第 1 组的出强度与第 2 组的入强度之间的相关性，反之亦然。可以观察到两者之间存在一致性，因此呈现正同配性。经过两步后 $h=2$，如果马尔可夫链从第 1 组开始，则很可能跳转到第 2 组，再返回第 1 组，从而产生异配行为。若马尔可夫链从第 2 组开始，情况也相同。因此，随着 r_h^{OI}的值增大，这

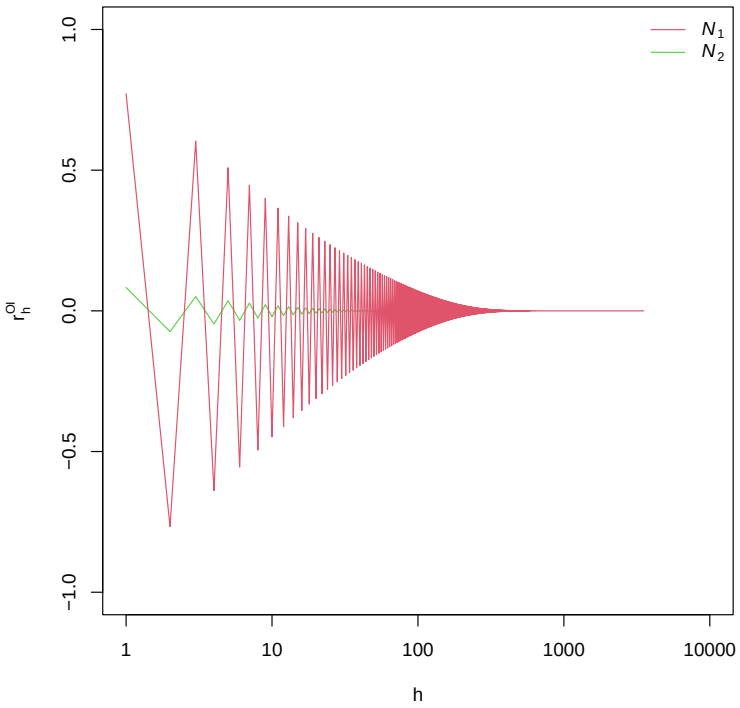

图 5： r_h^{OI}随着网络 N_1 和 N_2的 h 值增大。

会导致奇数和偶数 h值分别对应正、负同配性之间交替变化，这一点已由表 3 证实。渐近的零同配行为是由随机性和吸收态 c的存在共同造成的。在网络 N_2中，可以观察到相同的行为，但振荡幅度较小。这是因为第 1 组节点的出度增加，而且进入吸收态 c 的概率更高；在该状态下，中心性没有变异性。

请注意，即使汇点中的强度度量没有变异性，方程 (1) 中也不会出现不定式，因为到达汇点的概率小于 1。此外，在网络 N_1中，到达汇点的概率很高，因此很快就达到渐近值。相反，在网络 N_2中，经过同配行为与异配行为之间长时间的交替后，才达到渐近值。

图 6 展示了网络 N_3 和 N_4的相同结果。这类网络的特征是——

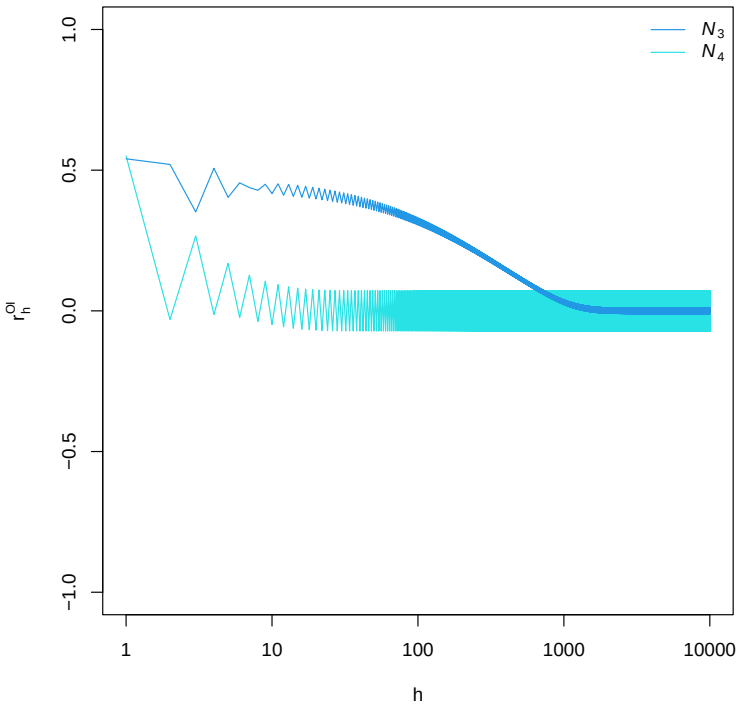

图 6： r_h^{OI}随着网络的 h 值增加， N_3 和 N_4。

由吸收类 $\{b, d\}$ 和互补类 $\{a, c, e\}$ 组成，两者通过从 a 到 b的边所表示的桥相连。两个类都具有正同配性。在 N_3中，桥的权重较低，因此互同配行为占主导。在网络 N_4中，桥的权重非常高，因此较大的振荡可归因于这些类之间的差异。

6 结论

本文通过为加权有向网络引入高阶同配性度量这一概念，填补了复杂网络文献中的空白，其中所考察的节点属性是度和强度中心性的有向版本。这类度量在与直接和加权网络密切相关的应用领域中具有清晰的解释，例如本文所研究的国际贸易领域。此外，我们还强调了所引入的同配性

度量与适当定义的马尔可夫链的自相关之间的紧密联系。

该方法提案具有广泛的适用性，因此能够描述一系列模型中的优先连接现象。一些具有代表性的实证案例可能包括分析迁移流的根源（Cerqueti 等，2019），或研究银行间系统的结构和系统性风险特征（Bo 和 Capponi，2015；Castellano 等，2021；Castiglionesi 和 Eboli，2018；Cerqueti 等，2022，2021；Bargigli 等，2015）。此外，在马尔可夫链的背景下，所提出的同配性度量具有良好的可读性，可有效用于分析具备马尔可夫性质的某些动力随机系统类别的性质。在这方面，值得注意的是，自相关的衰减能够解释底层随机过程的长期记忆特性（参见 Hurst，1951）。这些富有挑战性的主题已列入我们的研究议程。

参考文献

Antonietti, R., Falbo, P., Fontini, F., Grassi, R., 和 Rizzini, G.（2022）。世界贸易网络：国家中心性与COVID-19疫情。《应用网络科学》，7(1)：18。

Arcagni, A., Grassi, R., Stefani, S., 和 Torriero, A.（2017）。复杂网络中的高阶同配性。《欧洲运筹学杂志》，262(2)：708–719。

Arcagni, A., Grassi, R., Stefani, S., 和 Torriero, A.（2019）。扩展同配性：在加权社交网络中的应用。《商业研究杂志》。

Bang-Jensen, J. 和 Gutin, G. Z.（2008）。有向图：理论、算法与应用。Springer Science & Business Media。

Bargigli, L., Di Iasio, G., Infante, L., Lillo, F., 和 Pierobon, F.（2015）。银行间网络的多层结构。《定量金融》，15(4)：673–691。

Bo, L. 和 Capponi, A.（2015）。银行间网络中的系统性风险。《*SIAM*金融数学杂志》，6(1):386–424。

Castellano, R.、Cerqueti, R.、Clemente, G. P. 和 Grassi, R.（2021年）。一种最小化系统性风险的优化模型。数学与金融经济学，15:103– 129。

Castiglionesi, F. 和 Eboli, M.（2018）。银行间网络中的流动性流。金融评论，22(4):1291–1334。

Cerqueti, R.、Cinelli, M.、Ferraro, G. 和 Iovanella, A.（2022）。冲击传播下金融银行间网络的韧性。运筹学年鉴，https://doi.org/10.1007/s10479-022-04567-w。

Cerqueti, R.、Clemente, G. P. 和 Grassi, R.（2019）。迁移流社会经济根源的基于网络的度量。社会指标研究，146：187–204。

Cerqueti, R.、Clemente, G. P. 和 Grassi, R.（2021）。通过高阶聚类系数进行系统性风险评估。运筹学年鉴，299：1165–1187。

Estrada, E.、Hatano, N. 和 Gutierrez, A.（2008）。复杂网络中的“团簇性”混合。统计力学杂志：理论与实验，页码 P03008。

Foster, J. G., Foster, D. V., Grassberger, P. 和 Paczuski, M.（2010）。边的方向与网络结构。美国国家科学院院刊，107(24):10815–10820。

G´omez, S., Arenas, A., Borge-Holthoefer, J., Meloni, S. 和 Moreno, Y.（2010）。基于离散时间马尔可夫链的方法研究基于接触的疾病在复杂网络中的传播。*EPL*（欧洲物理学快报），89(3):38009。

Gross, J. L. 和 Yellen, J.（2003）。图论手册。CRC出版社。

Hurst, H. E. (1951). 水库的长期蓄存能力。美国土木工程师协会汇刊, 116(1):770–799. Iannelli, F., Koher, A., Brockmann, D., H¨ovel, P., and Sokolov, I. M. (2017). 复杂网络中流行病传播的有效距离。物理评论*E*, 95(1):012313. Leung, C. and Chau, H. (2007). 加权同配与异配网络模型。物理学*A*辑：统计力学及其应用, 378(2):591–602.

Mayo, M., Abdelzaher, A., and Ghosh, P. (2015). 复杂网络中的长程度相关性。计算社会网络, 2(1):1–13. Meghanathan, N. (2016). 基于中心性度量的现实世界网络图同配性分析。计算机与信息科学, 9(3):7–25. Newman, M. (2010). 网络：入门。牛津大学出版社。

Newman, M. E. J.（2002）。网络中的同配混合。《物理评论快报》，89(20):208701。

Norris, J. R.、Norris, J. R. 和 Norris, J. R.（1998年）。马尔可夫链。剑桥统计与概率数学系列第2卷。剑桥大学出版社。

Piraveenan, M.、Prokopenko, M. 和 Zomaya, A.（2010）。有向生物网络中的同配混合。《*IEEE/ACM*计算生物学与生物信息学汇刊》，9(1):66–78。

联合国商品贸易统计数据库（2022）。国际贸易统计数据库。数据取自国际贸易统计数据库，https://comtrade.un.org/。

Van Mieghem, P.、Wang, H.、Ge, X.、Tang, S. 和 Kuipers, F. A.（2010）。同配性和度保持重连对网络谱的影响。《欧洲物理学杂志*B*辑》，76(4):643–652。

Yuan, Y., Yan, J., 和 Zhang, P.（2021年）。加权有向网络的同配性度量。《复杂网络杂志》，9(2):cnab017。