

# 复杂网络建网算法研究

唐绍明, 沈黎\*, 张睿超, 陈俊宇, 向宇杰

湖南农业大学信息与智能科学技术学院, 湖南 长沙

收稿日期: 2022年5月10日; 录用日期: 2022年6月10日; 发布日期: 2022年6月17日

## 摘要

复杂网络是研究复杂系统的重要手段之一。如何将已有数据高效地映射成为复杂网络即是复杂网络建网算法研究的主要内容之一。在收集整理了大量的文献资料后, 文章详细综述了几种复杂网络的建网算法, 并进一步分析了各自算法的优缺点。

## 关键词

复杂网络, 建网算法

# Research on Establishing Network Algorithm of Complex Networks

Shaoming Tang, Li Shen\*, Ruichao Zhang, Junyu Chen, Yujie Xiang

College of Information and Intelligence, Hunan Agricultural University, Changsha Hunan

Received: May 10<sup>th</sup>, 2022; accepted: Jun. 10<sup>th</sup>, 2022; published: Jun. 17<sup>th</sup>, 2022

## Abstract

Complex network play an important role in the research of complex system. How to map complex network using existing data is the primary research content. This paper makes a detailed introduction to several algorithms for establishing complex network and analyzes the advantage and disadvantage of each algorithm.

## Keywords

Complex Network, Establishing Network Algorithm

\*通讯作者。



## 1. 引言

在自然界和人类社会中，个体总是或多或少被环境所影响，同时也不停地影响着环境。复杂网络是描述个体间相互关系的模型[1]，由于其灵活普适的特点，已经广泛应用在自然科学、社会和科学工程技术上，并取得了许多突破性的进展。复杂网络已经成为国际学术界一个新兴的研究热点，并且取得了飞速的发展。

随着复杂网络分析方法的不断成熟和完善，如何将现实世界映射成复杂网络也已经成为研究的关键问题。2008年，Lacasa等人提出了可视图算法。2009年，Luque水平可视图算法。2019年，田甜等人提出基于频域复杂网络分解的局部特征提取建网新方法[2]。2020年，李姝等人提出基于随机图的复杂网络建网方法[3]。在复杂网络的研究过程中，寻求一种直观、简便而又有效的建网方法已成为众多学者追求的目标。本文总结了复杂网络涉及的一些基本概念，并综述了复杂网络建网的常见算法并分析了各自的优缺点。

## 2. 复杂网络的基本概念

二元组  $G=(V,E)$ ， $n=|V|$ ， $m=|E|$  表示一个无向无权图，其中  $n$  和  $m$  是节点和边的数目， $V$  表示图中节点集合， $E$  表示图中边的集合。将复杂系统中按照某种规则抽象出来的个体映射为节点，复杂系统中个体之间的某种关系映射为边，这样复杂系统映射的二元组被称为复杂网络。复杂网络有如下几个基本概念。

### 2.1. 网络的度和度分布

度和度分布是复杂网络中基本而又重要的概念。节点  $v$  的度是指所有与该节点相连的总边数[4]，在复杂网络中随机选择一个节点该节点的度恰好为  $k$  的概率定义为  $P(k)$ ，则分布函数  $P(k)$  为网络中所有节点的度分布。ER 随机网络的度分布可以近似用泊松分布来描述，所以这种网络也被称为均匀网络。然而，通过对来自生物、信息和社会等不同领域的真实网络的大量研究表明，实际网络的度分布明显不同于泊松分布，而是服从幂律分布的，即

$$P(k) \sim k^{-\gamma} \quad (1)$$

这类网络称为无标度网络。

### 2.2. 网络直径 $D$ 和网络的平均路径长度 $L$

网络直径为网络节点之间距离的最大值，即

$$D = \max \{D_{i,j} \mid i \neq j; i, j = 1, 2, \dots, |V|\} \quad (2)$$

网络的平均路径长度为网络中任意两个节点之间的平均距离，即

$$L = \frac{2}{N(N-1)} \sum_{i \neq j} D_{i,j} \quad (3)$$

这里的  $D_{i,j}$  为无权网络中连接节点  $i$  和  $j$  所经过的最少边的个数， $N$  为网络中节点的总个数。网络的

平均路径长度是对复杂系统中信息传递效率的重要度量指标[5]。

### 2.3. 节点聚集系数

节点  $v$  的度为  $K_v$ ,  $E_v$  为节点  $v$  的  $K_v$  个邻居节点之间连接的边的个数, 那么节点  $v$  的聚集系数为

$$C_v = \frac{2E_v}{K_v(K_v - 1)} \quad (4)$$

节点的聚集系数描述了复杂网络中与同一个节点连接的两节点之间也相互连接的平均概率, 定量地刻画了复杂网络的局域结构性质。

## 3. 时间序列转换为复杂网络

近年来, 研究人员发现可以通过复杂网络来分析非线性时间序列的很多内在特征, 因此将时间序列映射为复杂网络是分析系统的关键环节。下面, 本文将介绍几个经典的时间序列转换复杂网络的建网方法。

### 3.1. 可视图复杂网络建网方法

2008 年 Lacasa [6] 等人基于可视角度提出一种新颖的时间序列转换复杂网络的快速算法, 即可视图算法。如图 1 所示, 该方法首先将时间序列的每个数据点以直方图的形式描绘到坐标轴上, 直方条的高度表示时间序列的数值大小。若两个直方条顶端处相连且没有被其他直方条阻断, 则时间序列中的节点可映射为复杂网络中有连边的两个节点。通过对时间序列中的所有数据点进行处理, 从而将时间序列转换为复杂网络。

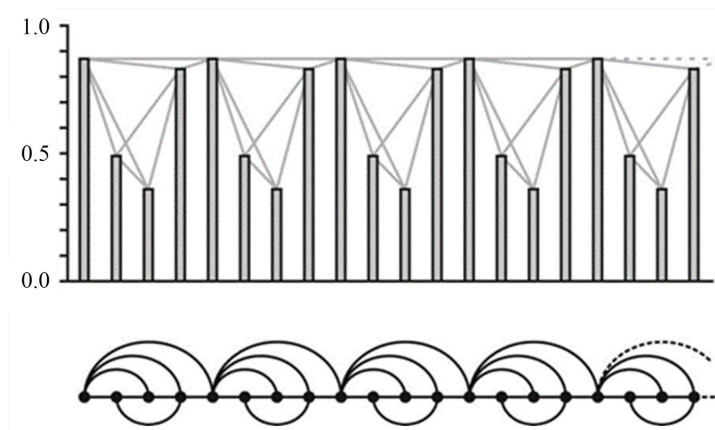


Figure 1. Algorithm theory of visibility graph  
图 1. 可视图算法原理图

可以看出, 可视算法将时间序列和复杂网络有效地连接起来, 并在复杂网络中继承了时间序列的大部分性质, 对于不同类型的信号具有较强的区分能力。

### 3.2. 水平可视图算法

2009 年, Luque [7] 在前人研究的基础上提出了水平可视图算法。如图 2 所示, 该算法使用经典可视图的方法将数据点转换到坐标轴上, 但与经典算法不同的是规定只能从时间点水平地向左右两边看。那么, 一个数据点能够观察到另一个数据点, 当且仅当它们之间没有被障碍物阻挡水平视线。

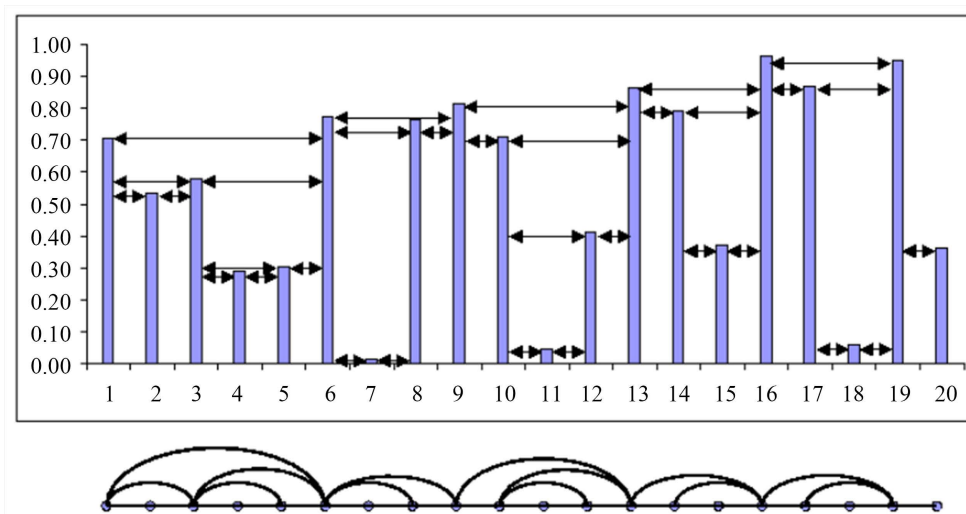


Figure 2. Algorithm theory of horizontal visibility graph  
图 2. 水平可视图算法原理图

可以看出，水平可视图算法相对简单，实现容易。但缺点是可能会造成时间序列中的一些重要信息的损失。

### 3.3. 基于相空间重建网算法

1980 年，Packard 等人总结了基于相空间重构的时间序列转换算法：坐标延迟法和导数法。但是在实际应用中，我们常常使用坐标延迟重构法来对时间序列进行相空间重构。对于确定的时间序列  $\{X_i, i=1, 2, \dots, n\}$ ，该算法选取合适的最佳延迟时间  $\tau$  并嵌入  $m$  维相空间中，则重构的相空间为

$$Y_j = (x_j, x_{j+\tau}, \dots, x_{j+(m-1)\tau}) \quad (5)$$

这里  $Y_j$  表示在  $m$  维相空间中的新的状态向量即为复杂网络中的节点。然后依据向量之间的相似性来确定网络节点之间是否相连，从而将原始时间序列转换为复杂网络。从上面的描述可以看出，嵌入维数和延迟时间的计算是该算法的核心内容。其求解方法的研究成果也层出不穷。其中比较有代表性的算法是利用虚假最近邻法计算嵌入维数和利用平均互信息法计算延迟时间。

可以看出，基于相空间重构算法有着坚实的数学基础，但在嵌入维数和延迟时间估算过程中存在不确定因素，而且无法精准给出连边关系的判断阈值，导致该算法的鲁棒性较差。

## 4. 改进的无标度网络演化算法

无标度网络的经典演化模型，例如：Price 模型和 BA 模型，虽然能模拟现实中的部分复杂网络，但是由于建网过程中，只是在单位时间内增加固定的节点和连边，这是不符合实际网络中非固定时间间隔增长的特征的。下面将介绍改进的网络演化算法。

### 4.1. 基于均匀更新计数的建网算法

在实际网络中，网络节点的增长时间间隔是不固定的。针对这一特点，该算法设定网络节点增加的时间间隔和每个新增节点的连接边数服从均匀分布。具体算法如下：

1) 初始化：生成  $N$  个节点，在任意一对节点间以概率  $p$  添加一条边。任何两两节点之间不能有重边和自环。

2) 增长: 按时间间隔  $t$  增加节点, 每个时间间隔只增加一个节点, 每个新增节点与网络中  $m$  个已知节点相连, 即新增  $m$  条边, 这里  $t$  和  $m$  服从均匀分布。

3) 优先连接: 每个新增的节点与已存在节点之间连接概率为  $\pi$ 。

4) 结束: 当经过  $N$  个时间间隔后, 输出生成的网络模型。

可以看出, 基于均匀更新计数的建网算法考虑了网络非匀速增加的特征, 更符合实际网络的特点。

## 4.2. 基于高斯更新计数的建网算法

考虑到实际中很多复杂网络的增长时间间隔是在不断变化的。在不同的时期, 增长间隔是不相同的, 那么我们可以将这种时间间隔的增长方式看成随机过程中的高斯分布。由此提出基于高斯更新计数的建网算法, 具体算法如下:

1) 初始化: 生成  $N$  个节点, 在任意一对节点间以概率  $p$  添加一条边。任何两两节点之间不能有重边和自环。

2) 增长: 按时间间隔  $t$  增加节点, 每个时间间隔只增加一个节点, 每个新增节点与网络中  $m$  个已知节点相连, 即新增  $m$  条边, 这里  $t$  服从高斯分布,  $m$  服从均匀分布。

3) 优先连接: 每个新增的节点与已存在节点之间连接概率为  $\pi$ 。

4) 结束: 当经过  $N$  个时间间隔后, 输出生成的网络模型。

可以看出, 基于高斯更新计数的建网算法中的增长时间间隔更具普遍性, 但其缺点是计算量较大, 实现较复杂。

## 5. 结语

21 世纪是复杂性和网络化的世纪。国内外已经掀起了复杂网络理论及其应用研究热潮, 如何将现实世界映射成为复杂网络则是研究的一个重要基础。在本文中, 我们介绍了几种常见的转换算法, 虽然算法实现原理各不相同, 但都能较高效地实现复杂网络的建网, 从而模拟现实世界。当然, 由于算法的不同特点, 建网过程中难免会损失掉一些重要信息, 如何利用新颖的算法来避免复杂网络建网过程中信息的丢失, 这是今后研究的方向。

## 基金项目

湖南农业大学大学生创新训练项目(XCX2021006)。

## 参考文献

- [1] 郭世泽, 陆哲明. 复杂网络基础理论[M]. 北京: 科学出版社, 2018.
- [2] 田甜, 温广瑞. 一种新的复杂网络建模和特征提取方法[J]. 振动、测试与诊断, 2019, 6(2): 17-19.
- [3] 李姝, 邵志刚. 基于随机图的复杂网络建模方法研究[J]. 小型微型计算机系统, 2020, 41(9): 1925-1938.
- [4] 汪小帆, 李翔, 陈关荣. 复杂网络理论及其应用[M]. 北京: 清华大学出版社, 2006.
- [5] 刘金龙. 复杂网络的重分形分析算法研究及其应用[D]: [博士学位论文]. 湘潭: 湘潭大学, 2017.
- [6] Lacasa, L., Lucas, B., Ballesteros, F. and Nuno, J.C. (2007) From Time Series to Complex Networks: The Visibility Graph. *Proceedings of the National Academy of Science of the United States of America*, **105**, 4972-4975. <https://doi.org/10.1073/pnas.0709247105>
- [7] Luque, B., Lacasa, L., Ballesteros, F. and Luque, J. (2009) Horizontal Visibility Graphs: Exact Result for Random Time Series. *Physics Review E*, **80**, Article No. 046103. <https://doi.org/10.1103/PhysRevE.80.046103>