



一种基于离散数据从局部到全局的网络重构算法

徐翔 朱承 朱先强

**Discrete data based local-to-global network reconstruction algorithm**

Xu Xiang Zhu Cheng Zhu Xian-Qiang

引用信息 Citation: *Acta Physica Sinica*, 70, 088901 (2021) DOI: 10.7498/aps.70.20201756

在线阅读 View online: <https://doi.org/10.7498/aps.70.20201756>

当期内容 View table of contents: <http://wulixb.iphy.ac.cn>

---

## 您可能感兴趣的其他文章

### Articles you may be interested in

复杂系统重构

Complex system reconstruction

物理学报. 2020, 69(8): 088906 <https://doi.org/10.7498/aps.69.20200001>

基于复杂网络动力学模型的无向加权网络节点重要性评估

Evaluation methods of node importance in undirected weighted networks based on complex network dynamics models

物理学报. 2018, 67(9): 098901 <https://doi.org/10.7498/aps.67.20172295>

节点中心性对复杂网络传播模式的影响分析

Analysis of the effect of node centrality on diffusion mode in complex networks

物理学报. 2017, 66(12): 120201 <https://doi.org/10.7498/aps.66.120201>

基于区域密度曲线识别网络上的多影响力节点

Identifying multiple influential nodes based on region density curve in complex networks

物理学报. 2018, 67(19): 198901 <https://doi.org/10.7498/aps.67.20181000>

多频正弦混沌细胞神经网络及其复杂动力学特性

Multi-frequency sinusoidal chaotic neural network and its complex dynamics

物理学报. 2020, 69(24): 240501 <https://doi.org/10.7498/aps.69.20200725>

基于领域相似度的复杂网络节点重要度评估算法

Node importance measurement based on neighborhood similarity in complex network

物理学报. 2017, 66(3): 038902 <https://doi.org/10.7498/aps.66.038902>

# 一种基于离散数据从局部到全局的网络重构算法\*

徐翔 朱承<sup>†</sup> 朱先强

(国防科技大学, 信息系统工程重点实验室, 长沙 410073)

(2020年10月22日收到; 2020年11月15日收到修改稿)

网络的结构和功能彼此相互影响, 网络上的功能往往体现为网络上的动力学过程, 网络上的动力学过程通过网络中的行为表象数据进行体现. 因此, 根据网络上可观测的相关数据对网络结构进行重构将成为可能. 本文拟解决如何根据网络上可观测的离散数据还原网络拓扑结构的问题, 提出了在网络局部利用每一条离散数据对应节点的相似程度来推测节点间发生连边的可能性, 通过多条离散数据重构网络各个局部拓扑并将由多条数据得到的局部拓扑进行叠加, 最终重构出整个网络的全局拓扑结构的算法. 为了验证算法的可行性与准确性, 在小世界、无标度和随机网络中进行了网络重构实验, 通过在三种不同类型及不同规模的网络中进行网络重构实验可以看出, 网络重构算法在不同类型网络中的表现也不同, 且网络的平均度值也会影响网络重构算法对数据的要求. 为了验证算法的适用性, 对三个实际网络进行了网络重构实验, 结果显示算法能够适用实际较大规模网络的重构. 该算法具有很好的适用性和准确度, 适合不同类型网络的拓扑结构重构场景.

**关键词:** 网络重构, 复杂网络, 动力学, 离散数据**PACS:** 89.75.Fb, 64.60.aq, 89.75.-k**DOI:** 10.7498/aps.70.20201756

## 1 引言

网络作为复杂系统的抽象已经广泛存在于现实世界, 从生物界的食物网<sup>[1]</sup>到大脑中的脑网络<sup>[2]</sup>、现代社会中的电力网络<sup>[3]</sup>、Internet<sup>[4]</sup>、社交网络<sup>[5]</sup>等等. 网络中的节点代表系统中的实体要素, 网络中各节点间的连边表示系统中各实体之间的相互作用关系. 然而, 人们一般对现实中的复杂系统知之甚少, 不了解系统内部的相关结构, 例如, 生态系统中各个物种之间的相互影响关系以及大脑中各个部分之间的相互作用关系等. 虽然系统中各要素之间的作用关系较难获得, 但随着系统的逐渐演化, 与系统行为相关的演化数据会被保留下来. 例如, 在生态系统演化的过程中, 不同演化时期存在的物种种类和物种数量可以获得; 2019年末到

2020年初爆发的新型冠状病毒在不同城市和国家的感染情况数据<sup>[6]</sup>也可以得到. 通过对系统演化过程中产生的相关数据进行分析和处理, 可以对系统中隐藏的结构和动态过程进行挖掘, 这类研究问题被称为动力学网络重构<sup>[7-12]</sup>. 在现实世界, 很多网络中的数据能够体现网络上的动态过程, 例如: 交通网络中的流量、车速, 社交网络中的点赞数、转发数等. 网络的结构具有自适应性质, 网络结构自适应辨识问题<sup>[13]</sup>对网络结构重构具有一定的帮助. 综上, 如何根据网络上可观测的相关数据对未知结构的网络进行拓扑重构是一个重要且有研究价值的问题.

目前, 对网络拓扑进行重构的工作较为丰富, 包括格兰杰因果关系 (Granger Causality)<sup>[14-17]</sup>方法, 通过因果推断来判断变量之间的关系, 该方法对成对变量具有较好的适用性, 变量数量达到三个

\* 国家自然科学基金 (批准号: 71571186, 61703416) 和湖南省研究生创新基金 (批准号: CX20190041) 资助的课题.

<sup>†</sup> 通信作者. E-mail: zhucheng@nudt.edu.cn

或者以上时,推断的结果可能会出现错误. 压缩感知 (compress sensing)<sup>[10,18,19]</sup> 方法通过将网络上的动力学过程转化成压缩感知方法能够处理的欠定线性系统,利用可观测到的时间序列对网络的拓扑进行重构. 压缩感知被广泛应用于电子工程尤其是信号处理中,用于获取和重构稀疏或可压缩的信号. 该方法的优势是通过获取少量的信号数据重构出原始信号. 除此之外,相关性方法能够根据网络节点之间的相关性进行网络拓扑重构. 文献 [20] 针对网络噪声干扰问题提出了一种结合 QR 分解 (QR decomposition) 和压缩感知的方法对网络进行结构重构. 相关性方法在其他领域也有很多应用<sup>[21,22]</sup>, 该方法的优点是简单快速,适合大规模网络拓扑重构问题,但对数据的数量和质量要求较高.

在面对网络结构重构问题时,一些工作基于信息论<sup>[23]</sup> 进行研究. 与简单的利用相关系数作为节点相关性的依据相比,与信息论相关的指标能更好地反映不同条件下节点之间的相关性程度. 常用的基于信息论的指标有互信息<sup>[24]</sup> (mutual information)、传输熵<sup>[25]</sup> (transfer entropy) 和因果熵<sup>[26,27]</sup> (causation entropy) 等. 文献 [28] 利用传输熵对无线传感网的拓扑进行了推测,但该网络的规模较小. 网络重构的方法还有很多,文献 [29,30] 较为详细地综述了相关的方法.

本文通过借鉴文献 [31] 中利用 SIR 模型产生网络数据的方法进行网络拓扑结构还原,产生数据的具体方法将在第 3 节阐述. 通过利用产生的初始数据,我们的贡献有以下几点: 第一,与传统的利用相关性指标<sup>[32]</sup> 进行节点相关性计算不同,首先统计被相同感染节点同时感染的不同节点数量,然后再统计任意两个节点同时被感染的数量,综合考虑了疾病在节点间的传播过程以及网络中不同节点之间的相互作用,更加全面地刻画了网络节点之间的相关性; 第二,与基于时序数据网络重构<sup>[33,34]</sup> 方法不同,我们的方法只需要离散的数据,即对数据之间的时间相关性没有要求,较大程度降低了获取数据的难度; 第三,使用了从局部到全局的结构重构方法,充分利用了每条数据对网络结构的影响,提高了网络拓扑重构的准确性且计算复杂度较低.

## 2 网络重构问题描述

网络重构问题根据对网络初始结构的了解

程度可以分为两类: 一类为已知部分初始网络结构,对剩余未知部分进行推理或预测,这类问题一般被称为链路预测<sup>[35]</sup> 问题; 另一类为对初始网络结构完全未知,一般需要知道网络上的动力学过程或能够获取网络上的观测数据.

针对网络结构部分已知的情况,可以利用链路预测相关方法进行网络结构的重构,典型的链路预测方法包括最大似然估计<sup>[36]</sup> 和概率模型<sup>[37]</sup> 等. 链路预测的思想是对给定的一个网络,为网络中没有连边的节点对  $(x, y)$  赋予一个得分值  $S_{xy}$ , 对网络中所有没有连边的节点对的得分值按照从大到小排序,排在最前面的节点对形成的连边概率最大. 链路预测的思想可以用于网络拓扑重构,需要做的是尽可能保证每条链路预测准确性以及预测链路的完整性. 利用链路预测中的相似性<sup>[38]</sup> 概念,通过对相似节点进行判断从而获得相似节点间的连边关系.

对网络进行重构有以下几点困难: 第一,网络结构的复杂性,实际中的很多网络具有节点数量多、连接关系复杂的结构特点; 第二,网络各节点之间的交互关系一般为非线性; 第三,关于网络动态过程的数据较难获得,包括数据的数量和质量. 同时,在获取网络数据过程中还会存在一定的干扰数据,即噪声会对网络重构的精度产生影响; 第四,实际存在的网络大多数为时序动态网络,且存在双层甚至多层的结构,因此如何对时序网络和多层网络结构进行重构<sup>[39]</sup> 也是一个重要的问题.

## 3 基于离散数据的网络重构算法

### 3.1 网络数据

网络上的动力学过程有很多,我们采取了经典的 SIR 疾病传播过程<sup>[40]</sup> 来产生网络数据,具体过程如下.

在未知网络中任选一个节点作为感染节点,设置疾病传播概率  $\beta = 0.2$ , 节点恢复概率  $\mu = 1$ , 一定时间之后,统计网络中最终稳定状态下被感染节点的数量,这一过程产生了网络的一条数据,以此类推,重复上述操作,可以获得关于网络的多条数据信息. 文献 [41] 使用了和本文相似的数据形式,不同的是该论文产生数据使用的动力学过程与本文不同,且网络中节点的状态会以一定的概率相互转移,即使用的是非终态数据.

$$\begin{matrix}
 & \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 & \dots & N \end{matrix} \\
 \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ \vdots \\ M \end{matrix} & \left[ \begin{array}{cccccccc}
 0 & 0 & 1 & 0 & 1 & 0 & \dots & 1 \\
 1 & 1 & 1 & 0 & 1 & 0 & \dots & 1 \\
 0 & 0 & 1 & 1 & 0 & 0 & \dots & 1 \\
 1 & 0 & 1 & 1 & 0 & 0 & \dots & 1 \\
 1 & 0 & 1 & 0 & 0 & 1 & \dots & 0 \\
 0 & 0 & 1 & 1 & 0 & 1 & \dots & 0 \\
 \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1
 \end{array} \right]
 \end{matrix}$$

图 1 网络初始二值数据矩阵  
Fig. 1. Initial binary data matrix of the network.

为了更加直观地表示网络数据并便于后续的相关计算, 将网络中被感染的节点状态设置为“1”, 未被感染的节点设置为“0”, 则可以得到网络初始二值数据矩阵, 数据格式如图 1 所示. 其中每一行代表不同的数据, 即不同的感染节点, 每一列表示

网络中不同的节点. 从该数据矩阵可以看出, 不同数据之间相互独立, 不存在时间上的相关性, 即数据是离散的.

### 3.2 相关定义

为了方便地介绍我们提出的基于离散数据的网络重构算法, 给出以下相关定义.

#### 定义 1 二值数据矩阵

给定一个图  $G = (V, E, S)$ ,  $V$  表示图中的节点集合,  $E$  表示图中的连边集合,  $S$  表示图中各节点的状态集合, 当节点  $j$  被第  $i$  次选取的感染源节点感染时,  $S_{ij} = 1$ , 反之  $S_{ij} = 0$ . 二值数据矩阵  $\mathbf{S}_{M \times N} = (S_{ij})_{M \times N}$ , 其中  $M$  表示网络中数据的数量,  $N$  为网络节点数量.

例如, 一个拥有 16 条数据 8 个节点的网络的二值数据矩阵可表示如下:

$$\mathbf{S}_{16 \times 8} = \begin{bmatrix}
 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \\
 1 & 1 & 1 & 0 & 1 & 0 & 0 & 1 \\
 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 \\
 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\
 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\
 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 \\
 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \\
 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 \\
 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\
 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 \\
 1 & 1 & 1 & 0 & 0 & 1 & 0 & 1 \\
 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\
 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\
 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\
 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\
 1 & 1 & 0 & 0 & 1 & 0 & 1 & 1
 \end{bmatrix}$$

#### 定义 2 相同感染源数量

给定一个图数据矩阵  $\mathbf{S}_{M \times N}$ , 定义网络中任意两节点的相同感染源数量为  $n_{kj} = \sum_{i=1}^M S_{ik} S_{ij}$ ,  $k \neq j$  (当  $k = j$  时规定  $n_{kj} = 0$ ), 其中  $S_{ik}$  表示节点  $k$  在第  $i$  次选取感染源节点时的状态,  $S_{ij}$  表示节点  $j$  在第  $i$  次选取感染源节点时的状态,  $M$  表示数据数量. 两节点的相同感染源数量越大, 则两节点间的相似性越高, 两节点间存在连边的概率越大, 反之亦然. 例如, 图 2 中  $n_{12} = 4$ .

#### 定义 3 相同感染源矩阵

给定一个二值数据图  $G = (V, S)$  和图数据矩阵

$\mathbf{S}_{M \times N}$ , 称  $\mathbf{A}^G = (n_{kj})_{N \times N}$ ,  $k \neq j$  (当  $k = j$  时规定  $n_{kj} = 0$ ) 为相同感染源矩阵,  $n_{kj}$  为节点  $k$  和节点  $j$  的相同感染源数量,  $N$  为二值数据图节点数量.

例如, 图 2 所示数据矩阵对应的相同感染源矩阵为

$$\mathbf{A}^G = \begin{bmatrix}
 0 & 4 & 5 & 4 & 3 & 4 & 4 & 7 \\
 4 & 0 & 3 & 0 & 2 & 1 & 2 & 4 \\
 5 & 3 & 0 & 4 & 2 & 7 & 5 & 9 \\
 4 & 0 & 4 & 0 & 1 & 5 & 5 & 4 \\
 3 & 2 & 2 & 1 & 0 & 0 & 2 & 4 \\
 4 & 1 & 7 & 5 & 0 & 0 & 3 & 5 \\
 4 & 2 & 5 & 5 & 2 & 3 & 0 & 6 \\
 7 & 4 & 9 & 4 & 4 & 5 & 6 & 0
 \end{bmatrix}.$$

$$S_{16 \times 8} = \begin{bmatrix} 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ \boxed{1} & \boxed{1} & 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ \boxed{1} & \boxed{1} & 1 & 0 & 0 & 0 & 1 & 1 \\ \boxed{1} & \boxed{1} & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ \boxed{1} & \boxed{1} & 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}$$

图 2 节点 1 和节点 2 的相同感染源数量

Fig. 2. The same number of infection sources in node 1 and node 2.

**定义 4** 二值数据子图

给定一个二值数据图  $G = (V, S)$  和图数据矩阵  $S_{M \times N}$ , 针对任意数据  $i$ , 称节点集合  $V^i = \{v_j | S_{ij} = 1\}$  中的节点构成的图为二值数据子图  $G^i$ .

例如, 由  $S_{16 \times 8}$  可以得到数据 1 对应的子图节点集合为

$$G^1 = (V^1, S^1) = (\{v_3, v_5, v_7, v_8\}, \{1, 1, 1, 1\}).$$

**定义 5** 子图相同感染源矩阵

给定一个二值数据子图  $G^i$ , 称  $A^i = (n_{kj})_{N_i \times N_i}$ ,  $k \neq j$  (当  $k = j$  时规定  $n_{kj} = 0$ ) 为二值相同感染源矩阵, 其中  $k, j \in V^i, N_i$  为第  $i$  次选取感染源节点时对应的二值数据子图节点数量.

例如, 图 2 数据矩阵中数据 1 对应的子图相同感染源矩阵为

$$A^1 = \begin{bmatrix} 0 & 2 & 5 & 9 \\ 2 & 0 & 2 & 4 \\ 5 & 2 & 0 & 6 \\ 9 & 4 & 6 & 0 \end{bmatrix}.$$

**3.3 子图重构**

给定任意数据  $i$  对应的子图相同感染源矩阵  $A^i$ , 对矩阵中每一行进行最大共同数据数搜索, 对最大数据数处两节点进行连边, 得到重构子图  $G_i = (V^i, E^i)$ ,  $E^i = \bigcup_{p, q \in V^i} \{(p, q) | n_{pq} = \max(n_{pq})\}$ , 其中  $V^i$  为子图节点集合,  $E^i$  为子图连边集合,  $n_{pq}$  为节点  $p$  和节点  $q$  的相同感染源数量; 当出现多个相同的最大共同数据数时, 依次选取其中的每个最大值对应的两节点进行连边, 对得到的不同子图进行度方差计算, 并将度方差值小的子网络作为最终子网的结构, 度方差计算公式为

$$\sigma_i^2 = \frac{1}{N_i} \sum_{j \in V^i} (k_j - \bar{k}_i)^2, \quad (1)$$

其中  $\sigma_i^2$  表示第  $i$  条数据对应子图的度方差值;  $k_j$  表示子图中节点的度值;  $\bar{k}_i$  表示子图的平均度值,  $N_i$  为子图节点数量. 例如, 数据 1 对应的子图重构过程如图 3 所示.

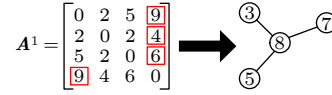


图 3 子图重构过程

Fig. 3. Subgraph reconstruction process.

**3.4 子图叠加**

对所有数据得到的重构子图  $G_i$  进行叠加, 即对子图  $G_i$  中的相同节点进行重叠, 最终得到图  $G$  的拓扑, 即  $G = (V, E)$ ,  $V = \bigcup_{i=1}^M V^i$ ,  $E = \bigcup_{i=1}^M E^i$ , 其中  $V$  表示图  $G$  的节点集合,  $E$  表示图  $G$  的连边集合,  $M$  表示数据数量,  $E^i$  为第  $i$  条数据对应重构子图的连边集合,  $V^i$  为第  $i$  条数据对应重构子图的节点集合, 子图叠加过程如图 4 所示. 对所有数据得到的子图进行叠加得到网络全局拓扑, 即得到网络的邻接矩阵  $A$ .

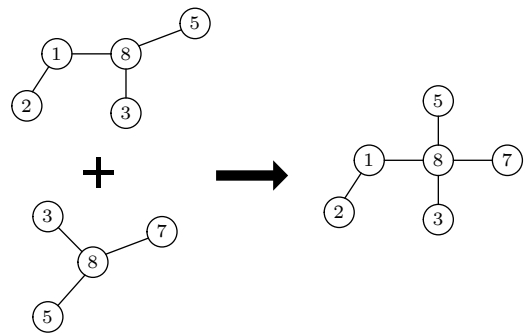


图 4 子图叠加过程

Fig. 4. Subgraph superposition process.

子图叠加数学计算过程如下:

$$\begin{aligned} G^{12} &= ((V^1 \cup V^2), (E^1 \cup E^2)) \\ &= (\{3, 5, 7, 8\} \cup \{1, 2, 3, 5, 8\}, \{(1, 2), \\ &\quad (1, 8), (3, 8), (5, 8)\} \cup \{(3, 8), (5, 8), (7, 8)\}) \\ &= (\{1, 2, 3, 5, 7, 8\}, \{(1, 2), (1, 8), \\ &\quad (3, 8), (5, 8), (7, 8)\}), \end{aligned}$$

网络重构算法流程如下所示:

**Algorithm** Network Topology Reconstruction

**Input:** Binary data matrix  $S_{M \times N}$

- 1:  $n_{kj} = \sum_{i=1}^M S_{ik} S_{ij}$
- 2:  $A^G = (n_{kj})_{N \times N}$
- 3: for  $i = 1$  to  $M$  do
- 4:      $V^i = \{v_j | S_{ij} = 1\}$
- 5:      $G^i \leftarrow V^i$
- 6:     for  $v_i$  in  $G^i$  do
- 7:          $A^i = (n_{kj})_{N_i \times N_i}$
- 8:         if number of  $\max(n_{pq})=1$
- 9:              $E^i = \bigcup_{p,q \in V^i} \{(p, q) | n_{pq} = \max(n_{pq})\}$
- 10:            else
- 11:              $E^i = \bigcup_{p,q \in V^i} \{(p, q) | \min(\sigma_i^2)\}$
- 12:         end for
- 13:      $G_i = (V^i, E^i)$
- 14: end for
- 15:  $V = \bigcup_{i=1}^M V^i$
- 16:  $E = \bigcup_{i=1}^M E^i$
- 17:  $G = (V, E)$

**Output:** Network topology  $G$

## 4 实验结果分析

### 4.1 网络重构指标

采用真正例率 (true positive rate, TPR) 和假正例率 (false positive rate, FPR) 分别表示网络重构的准确率和误差, TPR 指标越高, FPR 指标越小则说明网络重构的效果越好<sup>[42–43]</sup>. TPR 和 FPR 指标计算公式如下:

$$\text{TPR} = \text{TP}/(\text{TP} + \text{FN}), \quad (2)$$

$$\text{FPR} = \text{FP}/(\text{TN} + \text{FP}), \quad (3)$$

其中 TP(true positive), FP(false positive), TN (true negative) 和 FN(false negative) 分别表示真正例数、假正例数、真反例数和假反例数.

### 4.2 三种网络重构效果分析

为了验证本文算法的适用性, 针对不同规模的 WS 小世界网络、BA 无标度网络<sup>[44]</sup> 和 ER 随机网络<sup>[45]</sup> 进行了网络重构实验, 网络的相关拓扑属性如表 1 所列, 其中  $N$  表示网络节点数量,  $E$  表示网络连边数量,  $\langle k \rangle$  表示网络的平均度,  $C$  表示网络的集聚系数,  $\langle l \rangle$  表示网络的平均路径.

表 1 三类网络的基本拓扑特征

Table 1. Basic topological features of the three types of networks.

WS networks	$N$	$E$	$\langle k \rangle$	$C$	$\langle l \rangle$
WS100	100	200	4	0.099	3.61
WS200	200	400	4	0.078	4.17
WS300	300	600	4	0.057	4.51
WS400	400	800	4	0.052	4.74
WS500	500	1000	4	0.082	4.96
WS600	600	1200	4	0.062	5.12
WS700	700	1400	4	0.071	5.25
WS800	800	1600	4	0.079	5.43
WS900	900	1800	4	0.068	5.48
WS1000	1000	2000	4	0.068	5.58
BA networks	$N$	$E$	$\langle k \rangle$	$C$	$\langle l \rangle$
BA100	100	196	3.92	0.155	3.11
BA200	200	396	3.96	0.075	3.41
BA300	300	596	3.97	0.073	3.53
BA400	400	796	3.98	0.068	3.62
BA500	500	996	3.98	0.037	3.81
BA600	600	1196	3.98	0.044	3.77
BA700	700	1396	3.98	0.034	3.95
BA800	800	1596	3.99	0.023	3.93
BA900	900	1796	3.99	0.020	4.06
BA1000	1000	1996	3.99	0.027	4.14
ER networks	$N$	$E$	$\langle k \rangle$	$C$	$\langle l \rangle$
ER100	100	487	9.74	0.117	2.249
ER200	200	2056	20.56	0.103	2.004
ER300	300	4579	30.65	0.103	1.936
ER400	400	7936	39.68	0.099	1.917
ER500	500	12284	49.14	0.098	1.909

#### 4.2.1 WS 小世界网络实验

图 5 为不同规模的 WS 小世界网络重构实验效果. 由图 5 可以发现, 随着网络数据的增加, 不同节点规模的 WS 小世界网络重构效果也越来越好, 且最终都能够完全重构出网络的拓扑. 还可以发现, 随着网络规模的增加, 需要的网络数据量也随之增加, 但从最终达到平衡的数据数量来看, 需要的信息数量与网络节点呈线性变化, 即对网络数据数量的需求与网络节点数量是同一个数量级. 从对 WS 小世界网络的重构实验结果可以看出, 本文算法对网络拓扑还原具有较高的准确性, 能够适应不同规模的网络, 且对网络数据数量的要求不高.

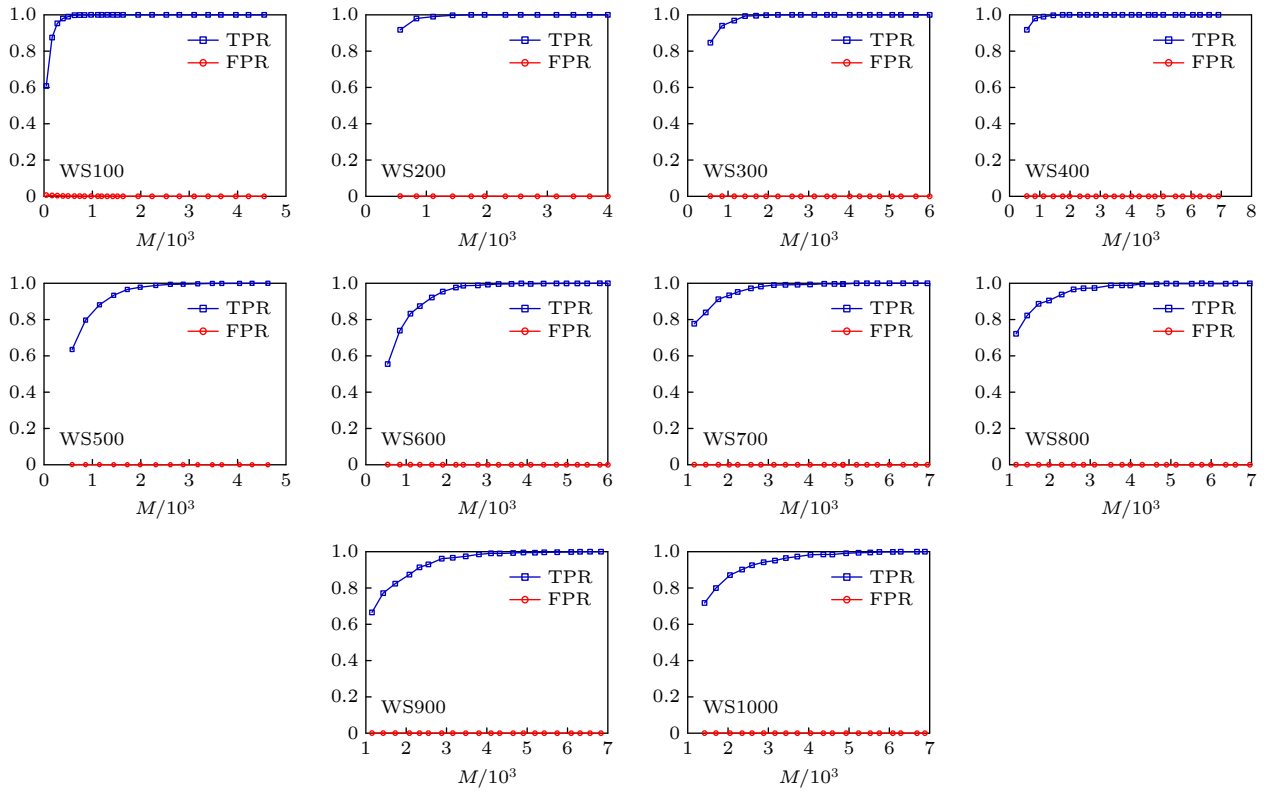


图 5 不同规模的 WS 小世界网络重构实验效果

Fig. 5. Experimental results of WS small world network reconstruction with different scales.

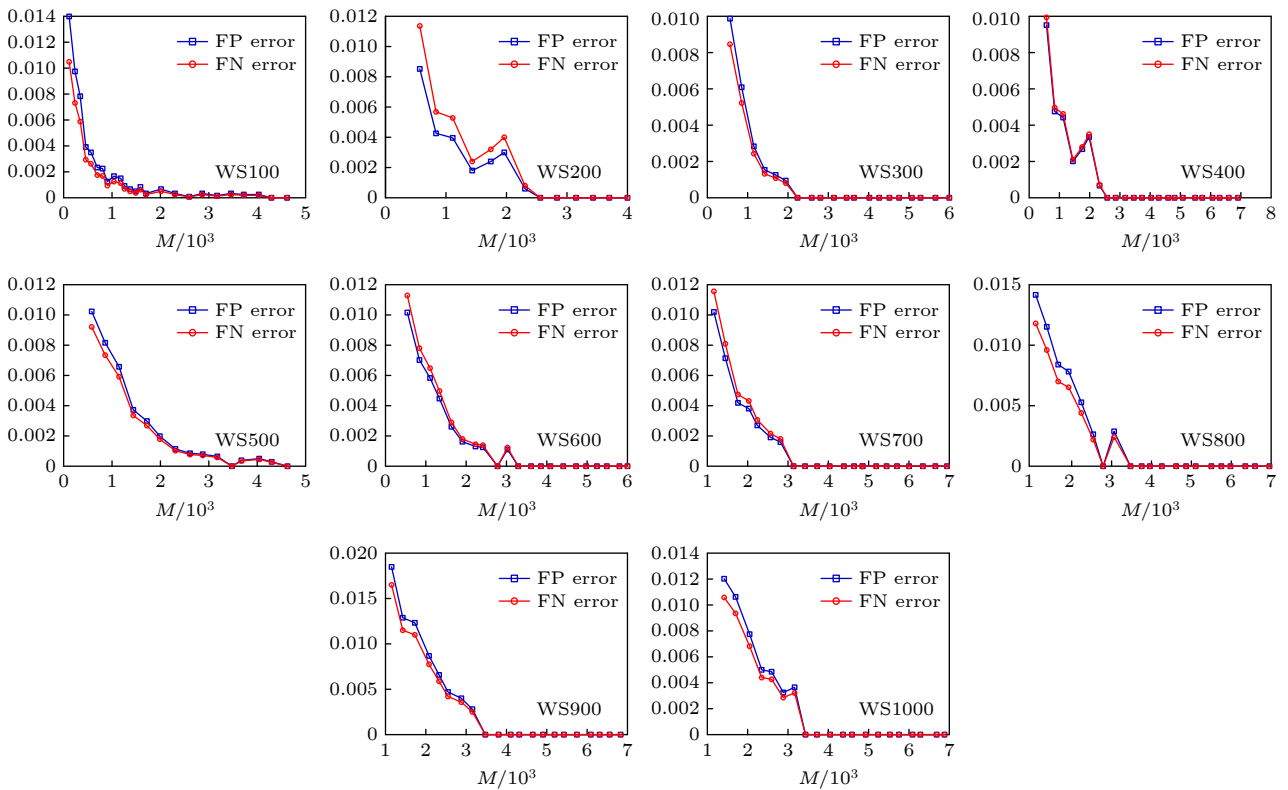


图 6 不同规模的 WS 小世界网络重构误差分析

Fig. 6. Error analysis of WS small world network reconstruction with different scales.

为更直观地反映算法的重构效果, 定义了多边  
 重构误差  $e_{FP}$  和少边重构误差  $e_{FN}$  指标, 计算公式如下:

$$e_{FP} = FP/TP, \quad (4)$$

$$e_{FN} = FN/TP. \quad (5)$$

如图 6 所示, 在不同节点规模的 WS 小世界网络

重构实验过程中, 随着实验数据的增加, 网络重构  
 实验的多边重构误差  $e_{FP}$  和少边重构误差  $e_{FN}$  逐渐  
 减小, 最终趋近于 0, 该实验误差分析进一步说明  
 了本文算法的准确性.

为了研究网络平均度值对网络重构效果的影  
 响, 对 WS 小世界网络, 分别对网络平均度值为 4,

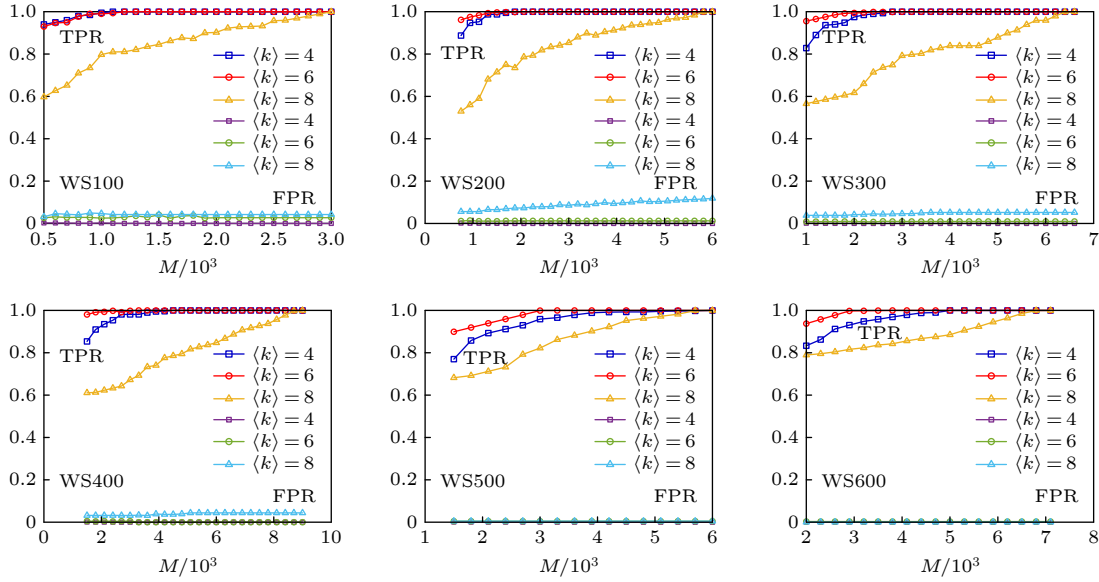


图 7 WS 小世界网络不同平均度值对网络重构实验效果的影响

Fig. 7. Influence of different average degrees of WS small world network on network reconstruction experiment.

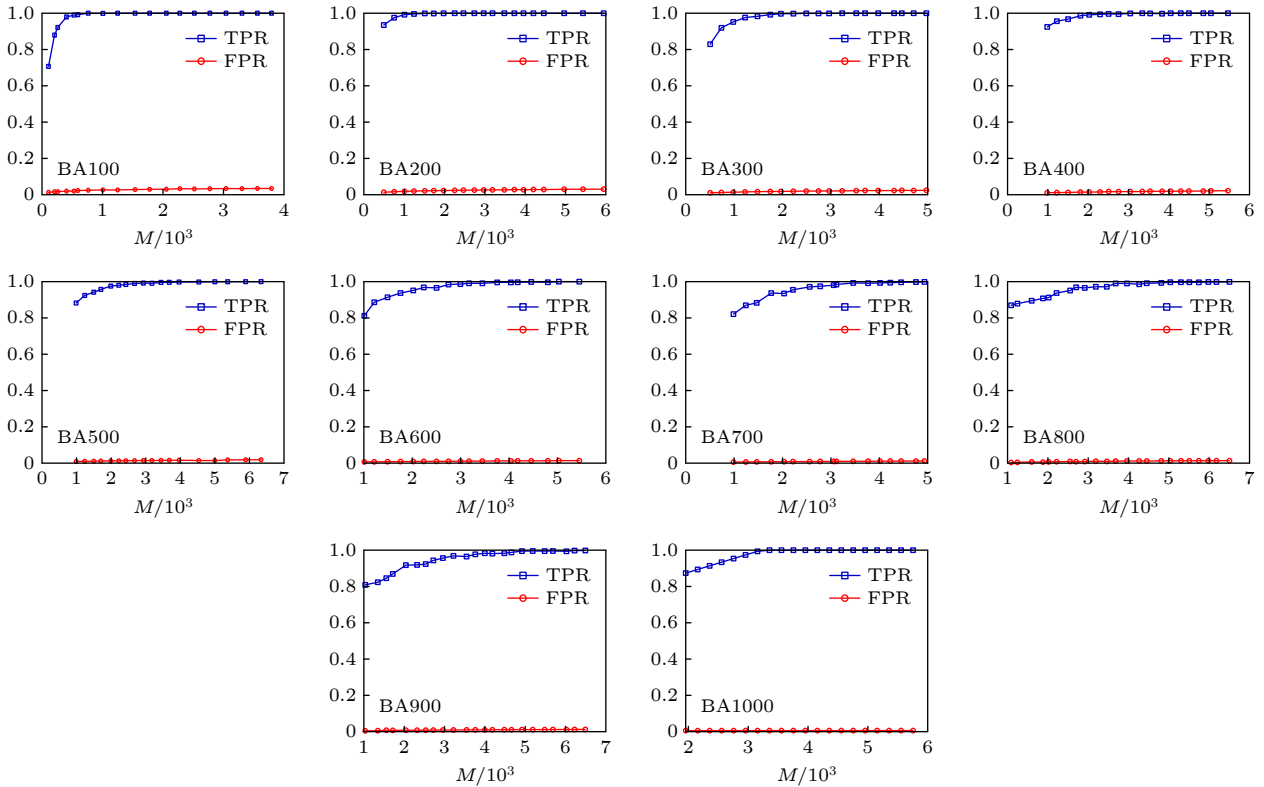


图 8 不同规模的 BA 无标度网络重构实验效果

Fig. 8. Experimental results of BA scale-free network reconstruction with different scales.



6 和 8(即  $\langle k \rangle = 4$ ,  $\langle k \rangle = 6$  和  $\langle k \rangle = 8$ ) 的网络进行了网络重构实验, 实验结果如图 7 所示. 从图 7 可以看到, 随着网络平均度值的增加, 要达到相同的网络重构效果, 需要的网络数据量更多, 原因是网络的平均度值越大网络中各个节点的连边数量越

多, 则网络整体的连边数量也随之增加, 所以需要更多的网络数据来重构网络. 从网络节点数量为 200—600 (即 WS200—WS600) 的重构效果可以发现, 网络平均度值为  $\langle k \rangle = 6$  的重构效果在相同网络数据情况下的重构效果比网络平均度值为

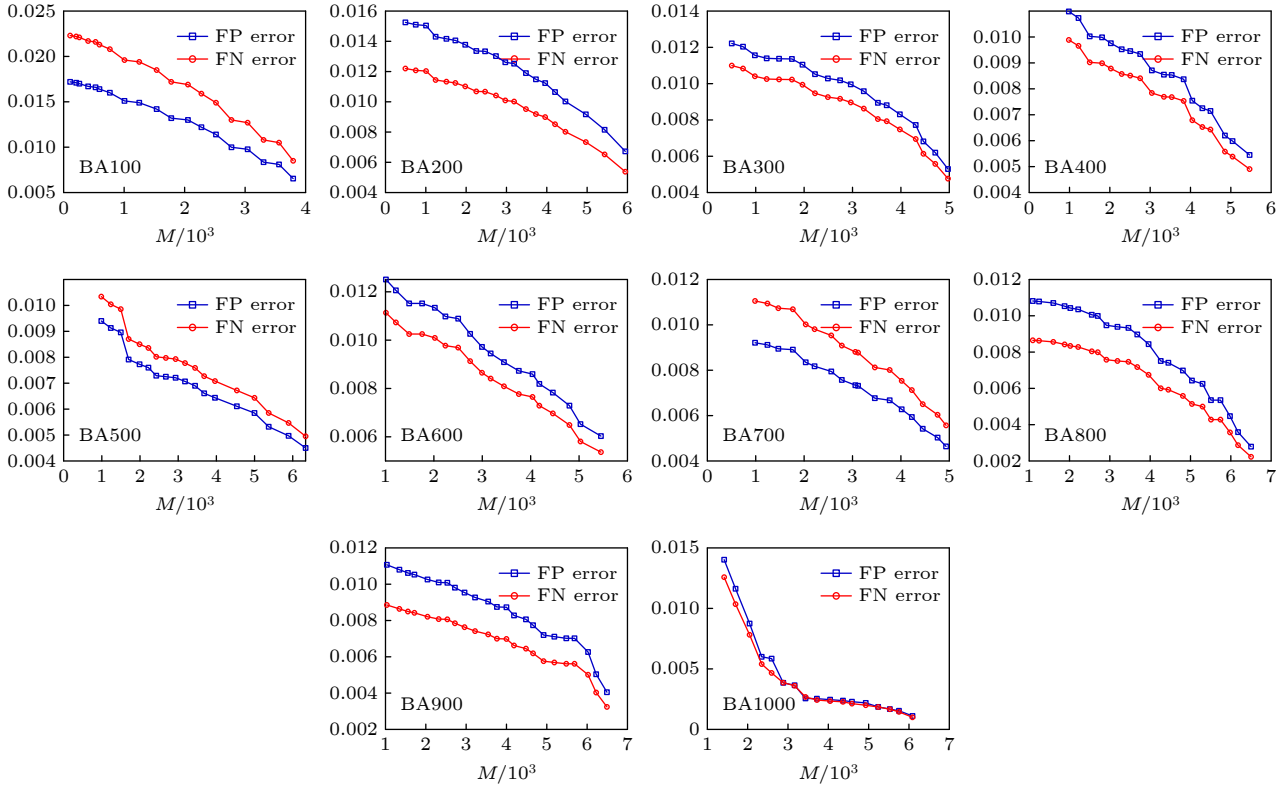


图 9 不同规模的 BA 小世界网络重构误差分析

Fig. 9. Error analysis of BA scale-free network reconstruction with different scales.

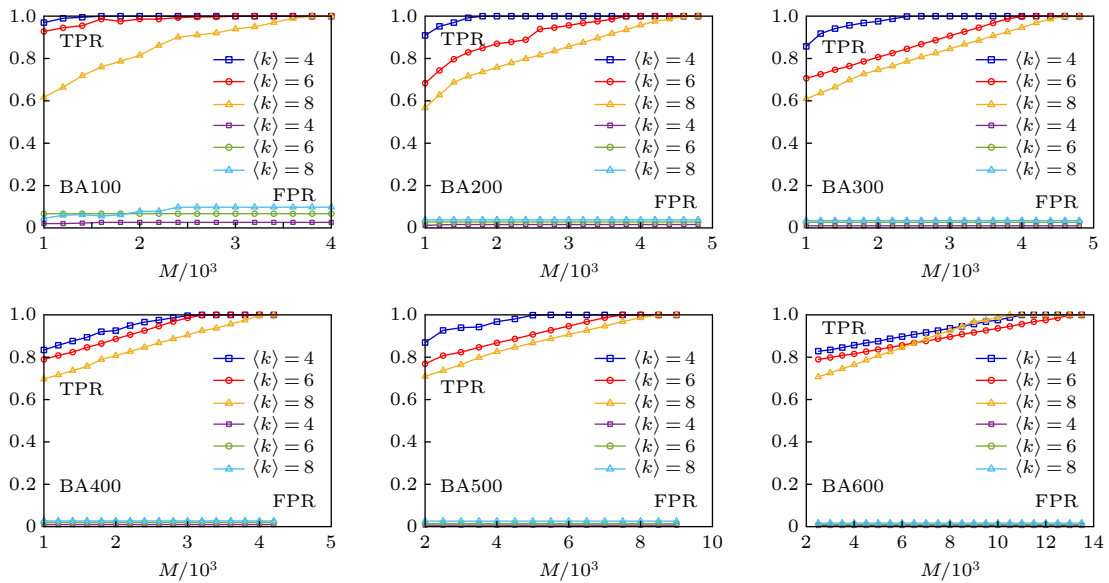


图 10 BA 无标度网络不同平均度值对网络重构实验效果的影响

Fig. 10. The influence of different average degree values of BA scale-free network on network reconstruction experiment.

$\langle k \rangle = 4$  的网络重构效果好, 原因可能是 WS 小世界网络具有较高的集聚性, 因此网络平均度值在一定范围内对网络数据的需求量较少, 即相同数量的网络数据能够很好地被节点及其邻居节点“利用”。

#### 4.2.2 BA 无标度网络实验

从图 8 可以看出, 与 WS 小世界网络类似, 随着网络数据的增加网络重构效果也越来越好. 图 9 展示了实验误差曲线, 总体上来说网络重构误差随着实验数据的增加逐渐减小. 从图 10 可以看出,

在相同网络数据的情况下, 网络平均度值越大, 网络的重构效果越差, 且平均度值越大网络重构需要的数据越大.

#### 4.2.3 ER 随机网络实验

图 11 展示了不同规模的 ER 随机网络重构效果, 相比同等规模的 WS 小世界和 BA 无标度网络, ER 随机网络需要更多的网络数据. 图 12 展示了两种重构误差的变化情况, 可以发现, 两种重构误差变化的趋势基本一致, 误差随实验数据的增加

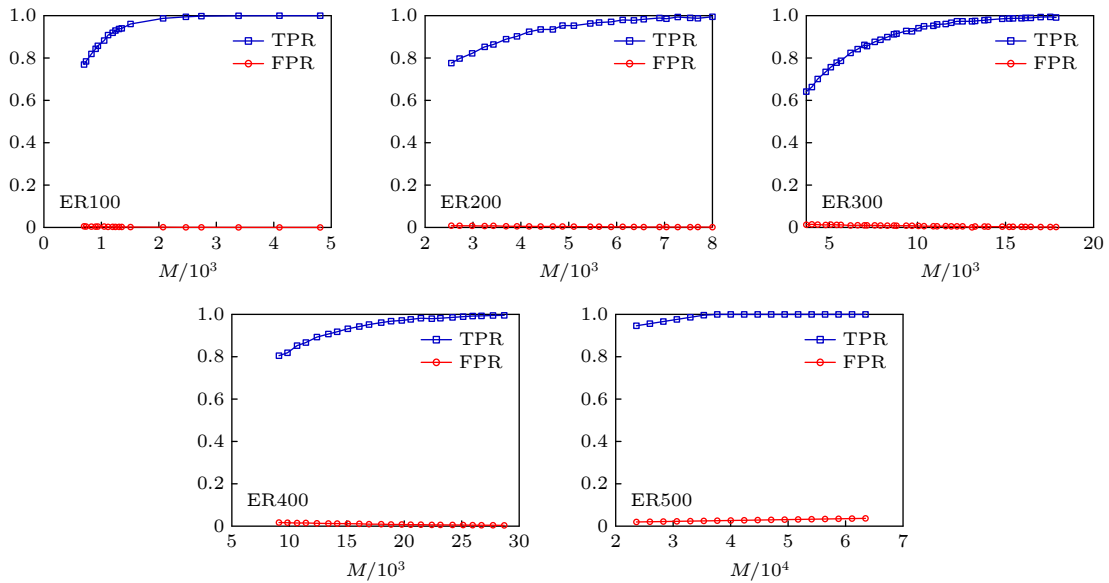


图 11 不同规模的 ER 随机网络重构实验效果

Fig. 11. Experimental results of ER random network reconstruction with different scales.

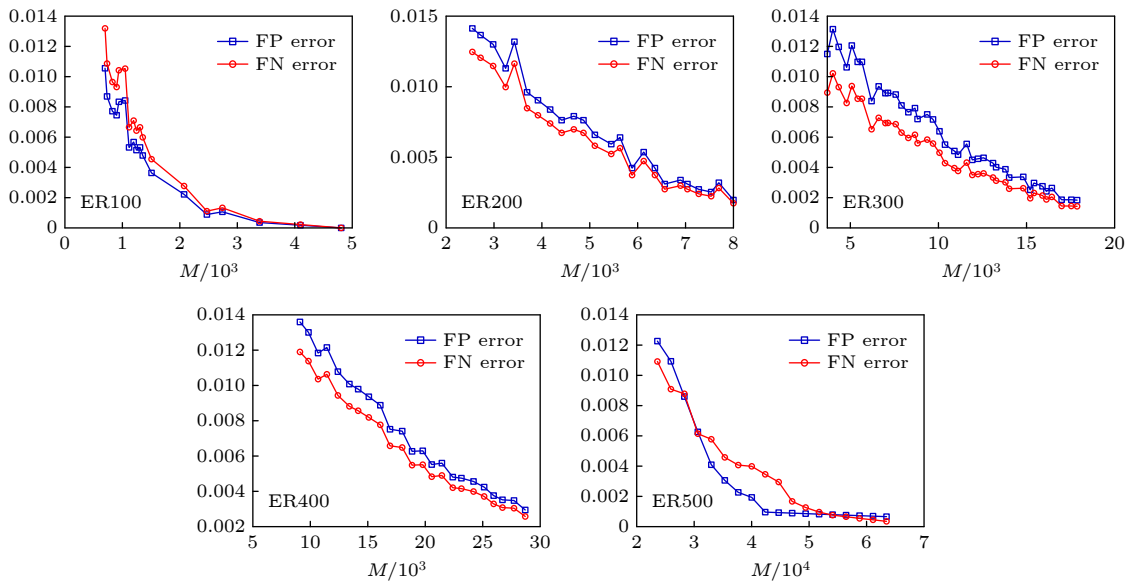


图 12 不同规模的 ER 小世界网络重构误差分析

Fig. 12. Error analysis of ER random network reconstruction with different scales.

逐渐减小. 除此之外, 对具有不同平均度值的 ER 随机网络进行网络重构实验, 发现网络重构的效果与网络的平均度值基本没有关系, 从图 13 可以发现三条曲线基本重合.

#### 4.2.4 三种网络对比实验

为了更直观地比较不同网络重构效果, 同时对 WS, BA 和 ER 网络进行网络重构实验, 实验结果见图 14. 从图 14 可以看出, 在相同网络数据下可以发现 WS 和 BA 网络的重构效果类似, ER 网络

则需要更多的网络数据.

#### 4.3 三种实际网络重构效果分析

为了更好地说明本文算法的适用性, 选取了三个实际网络进行网络重构实验, 三个网络的具体属性数据如表 2 所列. 其中, Euroroad 和 Minnesota 为公路网数据集, 相关数据可以在 <http://networkrepository.com/road.php> 上获取; Power Grid 数据集由 Duncan Watts 和 Steven Strogatz 编制, 数据可在 <http://cdg.columbia.edu/cdg/datasets> 上获取.

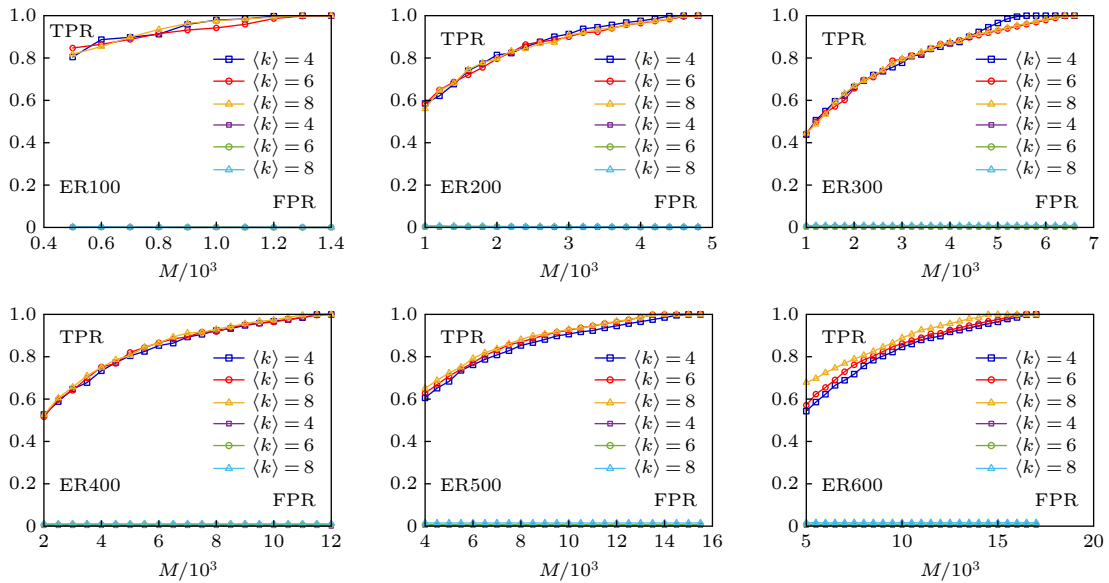


图 13 ER 随机网络不同平均度值对网络重构实验效果的影响

Fig. 13. Influence of different average degree of ER random network on network reconstruction experiment.

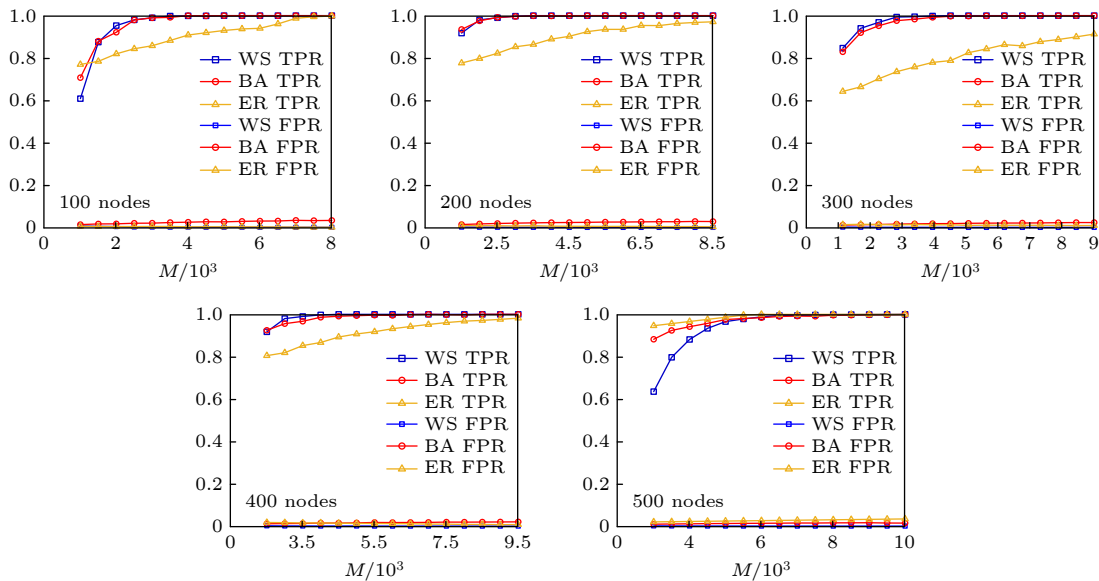


图 14 三种不同网络在相同数据下的重构效果对比

Fig. 14. Comparison of reconstruction effects of three different networks under the same data.

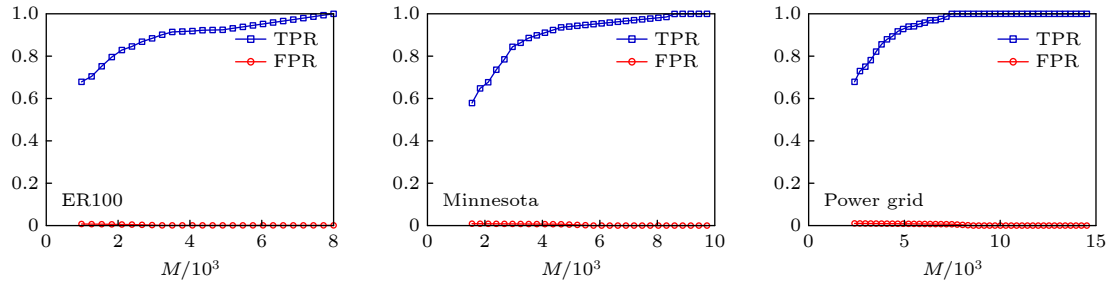


图 15 三个实际网络重构实验效果

Fig. 15. Experimental results of three practical network reconstruction.

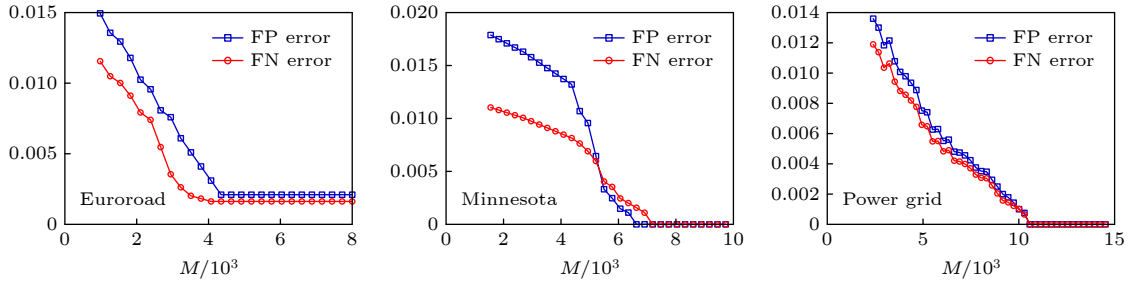


图 16 三个实际网络重构误差分析

Fig. 16. Error analysis of three practical network reconstruction.

表 2 三个实际网络的基本拓扑特征

Table 2. Basic topological characteristics of three practical networks.

实际网络	$N$	$E$	$\langle k \rangle$	$C$	$\langle l \rangle$
Euroroad	1174	1417	2.414	0.020	18.371
Minnesota	2642	3303	2.500	0.017	35.349
Power Grid	4941	6594	2.669	0.107	18.989

图 15 展示了三个实际网络的重构效果, 可以发现网络边数 (节点) 越多, 重构网络需要的数据越多, 随着使用数据的增加, 网络的重构效果也逐步提高. 图 16 展示了三个实际网络对应的重构误差变化曲线, 可以看出, 三个实际网络的重构误差随着数据量的增加都呈现下降趋势, 最终都趋于 0.

## 5 总结与展望

针对网络结构完全未知, 网络上的动力学过程已知的网络结构重构问题, 提出了一种基于离散数据从局部到全局的网络重构算法. 通过在网络上模拟 SIR 疾病传播过程来产生网络数据, 利用产生的数据从局部还原到全局叠加, 最终重构出整个网络的拓扑. 我们提出的算法具有快速, 简单的优势,

且适用于不同网络类型. 为了验证算法的准确性和适用性, 在具有不同节点数量的 WS, BA 和 ER 网络上进行了仿真实验, 实验结果表明我们的方法能够准确地还原出不同规模大小的网络拓扑. 为了验证算法的适用范围, 还对三个实际网络进行了重构实验, 由实验结果可以发现, 本文提出的算法同样可行. 目前我们研究的对象属于单层静态网络, 以后的工作可能会考虑如何对动态和多层网络进行拓扑重构.

## 参考文献

- [1] Martinez N D 1991 *Ecol. Monogr.* **61** 367
- [2] Oh S W, Harris J A, Ng L 2014 *Nature* **508** 207
- [3] Watts D J, Strogatz S H 1998 *Nature* **393** 440
- [4] Faloutsos M, Faloutsos P, Faloutsos C 1999 *Sigcomm. Comput. Commun. Rev.* **29** 251
- [5] Scott J 1994 *Sociology* **22** 109
- [6] Ding Xiang Yuan COVID-19 data <https://ncov.dxy.cn/ncovh5/view/pneumonia?from=timeline&isappintalled=0>. [2020-11-16] (in Chinese) [丁香园新型冠状病毒肺炎数据 <https://ncov.dxy.cn/ncovh5/view/pneumonia?from=timeline&isappintalled=0> [2020-11-16]]
- [7] Bongard J, Lipson H 2007 *Proc. Natl. Acad. Sci. U.S.A.* **104** 9943
- [8] Sugihara G, May R, Ye H 2012 *Science* **338** 496
- [9] Marbach D, Costello J C, Küffner R 2012 *Nat. Methods* **9** 796
- [10] Wang W X, Lai Y C, Grebogi C 2016 *Phys. Rep.* **644** 1
- [11] Casadiego J, Nitzan M, Hallerberg S, Timme M 2017 *Nat. Commun.* **8** 1

- [12] Nitzan M, Casadiego J, Timme M 2017 *Sci. Adv.* **3** e1600396
- [13] Zhao H, Li L, Peng H, Xiao J, Yang Y, Zheng M 2017 *Neurocomputing* **219** 5
- [14] Granger C W 1969 *Econometrica* **37** 424
- [15] Runge J, Bathiany S, Bollt E 2019 *Nat. Commun.* **10** 1
- [16] Maziarz M 2015 *J. Phil. Econ.* **8** 86
- [17] Zhou D, Xiao Y, Zhang Y, Xu Z, Cai D 2013 *Phys. Rev. Lett.* **111** 054102
- [18] Wang W X, Yang R, Lai Y C 2011 *EPL* **94** 48006
- [19] Shen Z, Wang W X, Fan Y, Di Z, Lai Y C 2014 *Nat. Commun.* **5** 1
- [20] Li L, Xu D, Peng H, Kurths J, Yang Y 2017 *Sci. Rep.* **7** 1
- [21] Zhang Z, Chen Y, Mi Y, Hu G 2019 *Phys. Rev. E* **99** 042311
- [22] Hempel S, Koseska A, Kurths J, Nikoloski Z 2011 *Phys. Rev. Lett.* **107** 3214
- [23] Akaike H 1973 *Information Theory and an Extension of the Maximum Likelihood Principle* (New York: Springer) pp199–213
- [24] Donges J F, Zou Y, Marwan N, Kurths J 2009 *EPL* **87** 48007
- [25] Stetter O, Battaglia D, Soriano J, Geisel T 2012 *PLoS Comput. Biol.* **8** e1002653
- [26] Sun J, Bollt E M 2014 *Physica D* **267** 49
- [27] Sun J, Taylor D, Bollt E M 2015 *SIAM J. Appl. Dyn. Syst.* **14** 73
- [28] Sharma P, Bucci D J, Brahma S K, Varshney P K 2019 *IEEE T. Netw. Sci. Eng.* **7** 562
- [29] Zhang H F, Wang W X 2020 *Acta Phys. Sin.* **69** 088906 (in Chinese) [张海峰, 王文旭 2020 物理学报 **69** 088906]
- [30] Zhang Z Y, Chen Y, Mi Y Y, Hu G 2020 *Sci. Sin.: Phys. Mech. Astron.* **1** 3 (in Chinese) [张朝阳, 陈阳, 弭元元, 胡岗 2020 中国科学: 物理学 力学 天文学 **1** 3]
- [31] Ma C, Zhang H F, Lai Y C 2017 *Phys. Rev. E* **96** 022320
- [32] Wagner A 2000 *Nat. Genet.* **24** 355
- [33] Levnajić Z 2012 *arXiv*: 1209.0219
- [34] Eagle N, Pentland A S, Lazer D 2009 *Proc. Natl. Acad. Sci. U.S.A.* **106** 15274
- [35] Lü L, Jin C H, Zhou T 2009 *Phys. Rev. E* **80** 046122
- [36] Clauset A, Moore C, Newman M E 2008 *Nature* **453** 98
- [37] Taskar B, Wong M F, Abbeel P, Koller D 2004 *NIPS* **16** 659
- [38] Liu W, Lü L 2010 *EPL* **89** 58007
- [39] Ma C, Chen H S, Li X, Lai Y C, Zhang H F 2020 *SIAM J. Appl. Dyn. Syst.* **19** 1
- [40] Pastor-Satorras R, Castellano C, Van Mieghem P, Vespignani A 2015 *Rev. Mod. Phys.* **87** 120
- [41] Zhang H F, Xu F, Bao Z K, Ma C 2018 *IEEE T. Circuits-I* **66** 4
- [42] Hempel S, Koseska A, Kurths J, Nikoloski Z 2011 *Phys. Rev. Lett.* **107** 054101
- [43] Han X, Shen Z, Wang W X, Di Z 2015 *Phys. Rev. Lett.* **114** 028701
- [44] Barabási A L, Albert R 1999 *Science* **286** 509
- [45] Erdős P, Rényi A 1960 *Publ. Math. Inst. Hung. Acad. Sci.* **5** 17

# Discrete data based local-to-global network reconstruction algorithm<sup>\*</sup>

Xu Xiang    Zhu Cheng<sup>†</sup>    Zhu Xian-Qiang

(Key Laboratory of Information Systems Engineering, National University of Defense Technology, Changsha 410073, China)

( Received 22 October 2020; revised manuscript received 15 November 2020 )

## Abstract

The structure and the function of network interact with each other. The function of network is often reflected as the dynamic process on the network. The dynamic process on the network is reflected by the behavior data in the network. Therefore, it is possible to reconstruct the network structure according to the observed data. This paper aims to solve the problem of how to restore the network topology according to the observable discrete data on the network. In this paper, an algorithm to infer the possibility of edge connection between nodes is proposed by using the similarity degree of each node corresponding to each discrete datum, and by reconstructing each local topology of the network through multiple discrete data, and by superposing the local topology obtained from multiple data, the global topology of the whole network is reconstructed finally. The data in the network are generated by SIR (Susceptible Infective Removed) model with infection probability of 0.2 and recovery probability of 1. Each time, a single node is selected as the infected node, and the final infection state of the network is counted as a network datum. In order to verify the feasibility and accuracy of the algorithm, the network reconfiguration experiments are carried out in small world, scale-free and random networks. Through the network reconstruction experiments in the networks of three different types and different scales, we can see that the performances of network reconstruction algorithm in different types of networks are different, and the average degree of network will affect the requirements for data of the network reconstruction algorithm. In order to verify the applicability of the algorithm, network reconstruction experiments are carried out on three practical networks. The results show that the algorithm can be applied to the reconstruction of large-scale networks. In order to show the accuracy of the algorithm more intuitively, we analyze the network reconstruction error after each network reconstruction experiment. The experiment shows that with the gradual increase of network data, the network reconstruction error gradually decreases and finally approaches to 0. In a nutshell, the algorithm we proposed in this work has good applicability and accuracy, and is suitable for different types of network topology reconstructions.

**Keywords:** network reconstruction, complex networks, dynamics, discrete data

**PACS:** 89.75.Fb, 64.60.aq, 89.75.-k

**DOI:** [10.7498/aps.70.20201756](https://doi.org/10.7498/aps.70.20201756)

<sup>\*</sup> Project supported by the National Natural Science Foundation of China (Grant Nos. 71571186, 61703416) and the Postgraduate Innovation Project of Hunan Province, China (Grant No. CX20190041).

<sup>†</sup> Corresponding author. E-mail: [zhucheng@nudt.edu.cn](mailto:zhucheng@nudt.edu.cn)