



# 基于地理加权中心节点距离的网络社区发现算法

万 幼<sup>1,2</sup> 刘耀林<sup>1,2,3</sup>

1 武汉大学资源与环境科学学院,湖北 武汉,430079

2 武汉大学地理信息系统教育部重点实验室,湖北 武汉,430079

3 地球空间信息技术协同创新中心,湖北 武汉,430079

**摘要:**提出一种基于地理加权中心节点距离的网络社区发现算法 (geographical weighted central node distance based Louvain method, GND-Louvain)。该算法扩展了传统复杂网络领域的经典社区发现方法 Louvain, 利用地理加权中心节点来度量社区发现过程中的空间距离关系, 并将此距离衰减效应加入到距离模块度模型中, 以此来计算和评估空间网络社区划分结果的质量, 并探究了空间社区发现结果不稳定的原因。通过定义节点计算顺序, 保证了社区发现结果的质量和稳定性。利用中国铁路网线路数据, 设计了 5 种不同空间约束的空间社区发现对比性实验。结果证明, GND-Louvain 算法的准确性最高, 并且算法结果最稳定。

**关键词:**社区发现; 空间网络; 复杂网络; 空间约束; 模块度

中图分类号:P208

文献标志码:A

空间网络是节点位于一个可物理度量空间中的网络<sup>[1]</sup>。通常是在一个二维空间,采用欧氏距离度量。社区结构是复杂网络的一个基本属性<sup>[2]</sup>。复杂网络的社区发现是识别一个节点的集合,使得集合内节点之间的相互作用比它们与集合外节点的相互作用更强<sup>[3]</sup>。而空间网络的社区发现能获得分布在地理空间中的非随机连接所产生的、具有高内聚低耦合的子网络社区结构,从而揭示城市内部和城市之间的空间交互模式<sup>[4]</sup>、道路交通网络的空间分布格局<sup>[5]</sup>、人群活动行为模式<sup>[6]</sup>等,因此具有很好的研究意义。

与传统复杂网络的结构构成不同,空间网络中节点间的连接概率不仅与其自身的度有关,还与空间距离有关。其连接概率呈现随距离衰减的效果<sup>[7-8]</sup>。当涉及到由地理现象和人类活动所构成的空间网络社区发现时,除网络拓扑关系外,常需要考虑其他空间约束的影响。针对空间网络社区发现的研究方法可分为 4 种<sup>[9]</sup>:①忽略空间距离关系,直接采用传统的社区发现方法对网络拓扑结构进行社区挖掘<sup>[10]</sup>;②在传统社区发现方法中加入空间邻近约束条件,获得区域内部空间连通的社区结果<sup>[11]</sup>;③在网络连接的属性中加入

空间距离的反作用影响,重构空间网络,再利用传统社区发现方法挖掘社区结构<sup>[12]</sup>;④改进网络模块度计算的概率模型,建立距离与拓扑关联关系相结合的节点间联系的联合概率函数,作为距离模块度计算的期望值部分,从而挖掘出剔除距离因素之后仍然紧密联系的空间社区<sup>[13-14]</sup>。此类方法识别出的社区结果可以分析除了空间距离影响之外,其他构成社区空间结构的因素,因此可以获得更多有价值的知识<sup>[13-14]</sup>。如 Expert 等<sup>[13]</sup>采用剔除距离影响的模块度模型,从比利时手机通话数据中识别出在空间上存在重叠的两大社区,而两大社区的空间分布揭示了比利时的两大语言体系的所在区域。

现有空间网络社区发现方法存在如下的两个问题:①对社区空间距离的度量方式有多种形式,但缺乏由此造成的社区发现结果的差异性分析研究。如图 1 所示,在度量两个子社区 1 和 2 之间的距离时,可使用所有连接边/存在空间邻近区域的连接边之间的最短/平均/最大距离<sup>[15]</sup>,或子社区的地理几何中心节点距离,或子社区的度中心节点距离。不同的距离度量方式得到的距离值差异很大,这个差异直接影响到它们之间距离衰减

效应的计算,进而影响空间社区发现的结果。到目前为止,还未有针对此影响探究何种度量方式更加准确和有效的研究。②缺乏对空间网络社区发现方法稳定性的分析研究。在空间网络社区发现中最普遍采用的是基于快速展开策略的模块度最大化方法 Louvain<sup>[16]</sup>。此方法的效率和准确性很高,适合于大型网络<sup>[1,17]</sup>,但其社区发现结果存在不稳定性。当网络节点输入顺序改变时,会导致计算结果的显著不同。

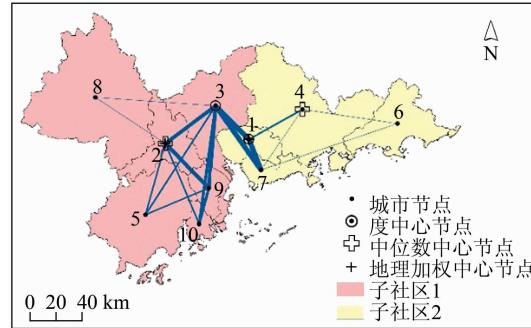


图 1 两个网络子社区之间的距离度量方式  
Fig. 1 Different Distance Measures Between Two Meta-Communities

针对以上两个问题,本文提出一种新的空间网络社区发现算法(geographical weighted central node distance based Louvain method, GND-Louvain)。算法采用地理加权中心节点来度量子社区间的距离,并在模块度公式中加入距离衰减效应影响,利用距离模块度,发现隐含的有意义的网络社区结构。然后通过限定空间节点的输入和计算次序,在保证社区发现结果质量的同时提高了方法的稳定性。

## 1 GND-Louvain 算法理论方法

本节首先介绍空间网络中地理加权中心节点的定义和计算方法,然后给出基于地理加权中心节点的空间网络社区发现方法。

### 1.1 空间网络的地理加权中心节点

复杂网络领域针对网络节点中心性的测量主要有3个指标:点度中心度、接近中心度和介数中心度<sup>[18]</sup>。其中,节点的接近中心度是该节点到其他所有节点的最短路径之和的倒数,可反映节点与网络中其他节点的接近程度。地理加权中心节点的定义借鉴接近中心度指标,并融合了空间点模式分析中加权中位数中心节点的概念<sup>[19-20]</sup>。它以网络节点间最短路径衡量节点间的距离,最终选取到其他所有网络节点的距离之和最小的(即接近中心度最大)的那个节点。图1中节点2和

1分别是子社区1和2的地理加权中心节点。

地理加权中心节点综合考虑了空间网络所具有的空间距离和网络连接特征,被广泛应用于交通网络中各种资源的合理配置中,以最小化资源获取的代价。

### 1.2 GND-Louvain 社区发现算法

空间网络的社区发现以社区内网络节点间的关联最大化为目标进行网络节点的分割或合并操作。模块度最大化是广为应用的一种可量化的社区发现方法<sup>[2-3]</sup>。特别是空间网络的社区发现,基本都是基于模块度最大化算法的扩展或改进<sup>[1, 12-13]</sup>。

**定义1** 传统模块度(NG-modularity)<sup>[21]</sup>。对于网络一个给定的划分  $C = \{c_1, c_2 \dots c_q\}$ , 传统模块度的计算公式为:

$$M(C) = \frac{1}{2m} \sum_{c \in C} \sum_{i, j \in c} W_{ij} - P_{ij} \quad (1)$$

式中,  $m$  是网络总边数;  $i, j$  是同一个社区的节点对, 节点  $i, j$  之间连接的权重  $W_{ij}$  用连接数来度量;  $P_{ij}$  是随机情况下节点  $i, j$  之间连接的概率,

$$P_{ij} = \frac{k_i k_j}{2m}, \text{ 其中 } k_i \text{ 和 } k_j \text{ 分别是节点 } i \text{ 和 } j \text{ 的度}.$$

模块度的计算公式度量了真实网络与随机情况下,社区内部的边的联系紧密程度和社区之间的边的联系的紧密程度。网络划分的模块度可以用来衡量社区划分的质量。模块度取值在[-1, 1]之间,通常当模块度大于0.3时,认为网络中存在有非随机模式产生的社区结构。且模块度取值越大,社区内部的连接越紧密。

**定义2** 距离模块度(D-modularity)<sup>[14]</sup>。对于网络一个给定的划分  $C = \{c_1, c_2 \dots c_q\}$ , 距离模块度的计算公式为:

$$M_{dist}(C) = \frac{1}{2m} \sum_{c \in C} \sum_{i, j \in c} W_{ij} - P_{ij} \quad (2)$$

与传统模块度的随机连接概率不同,距离模块度的随机连接概率为:

$$P_{ij} = \frac{\hat{P}_{ij} + \hat{P}_{ji}}{2m}, \hat{P}_{ij} = \frac{k_i k_j f(d(v_i, v_j))}{\sum_{v_q \in V} k_q f(d(v_q, v_i))} \quad (3)$$

此处增加了一个距离衰减函数  $f$  来度量节点间随距离  $d$  衰减的连接概率。本文中使用指数型的距离衰减函数<sup>[7-8]</sup>,其定义如下:

$$f(x) = e^{-(x/\sigma)^2} \quad (4)$$

式中,  $x$  是距离值;  $e$  是自然对数底;  $\sigma$  是控制距离衰减幅度的变量。 $\sigma$  可被设定为一个数据驱动的自适应数值(例如,当前空间网络所有点对距离的平

均值)。根据指数函数性质, 高于此平均值的距离  $x$  的连接, 将获得较大的距离衰减影响, 而低于此平均值的连接, 会获得较小的距离衰减影响。并且当  $\sigma$  接近无穷大时, 距离模块度变为传统模块度。

GND-Louvain 算法扩展了经典的基于模块度的社区发现方法 Louvain。它采用距离模块度作为度量社区内部连接紧密性的标准。并且, 在社区发现的过程中, 不断地重新计算并更新子社区的地理加权中心节点, 以此度量子社区之间的距离衰减效应。具体的, GND-Louvain 算法是一个类似凝聚层次聚类的迭代分层的过程, 主要有 3 个步骤:

- 1) 通过迭代方式遍历节点集合, 找出使每个节点模块度变化最大的连接节点与之合并, 直到当前节点集合的任意节点移动都不会再使模块度增加时, 结束迭代, 并形成一层社区分割结果。

- 2) 将获得的社区分割结果的每个子社区视为新的节点, 子社区之间的连接视为新的边, 以此构成一个新的空间网络, 再代入步骤 1) 进行迭代遍历, 形成更高一层的社区分割结果。

- 3) 当步骤 2) 的重复迭代过程不能使模块度再增加时, 输出最终的社区分割结果。

## 2 实验分析

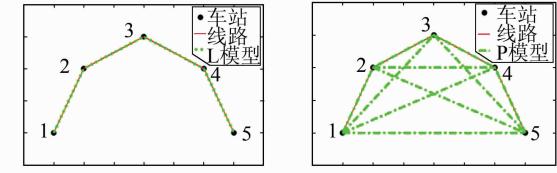
### 2.1 空间网络的数据建模

空间网络数据的来源是全国铁路交通网络数据。利用网络爬虫技术, 从中国铁路客运服务中心网站获取了 2017 年 4 月间所有在运行的铁路客运线路数据, 包含 7 种列车类型(城铁、动车、高铁、普快、特快、临客和卧铺)和 5 470 条运行线路, 然后对数据进行了预处理和网络建模。

数据预处理包括: 将所有火车站对应到所在地级市; 以城市名替代原铁路线路中的站点名; 以城市的内几何中心计算空间距离矩阵; 以各城市地理边界计算空间邻接矩阵。海南省因与大陆城市无邻接关系, 其省内的线路数据被删除。

铁路网络的建模方式主要有两种: L 空间模型和 P 空间模型<sup>[22]</sup>。L 空间模型中, 在一条经过多个点的线路上, 只有相邻的两个节点才被认为是直接相连的。P 空间模型中, 由一条线路连接的所有节点都被认为是两两相连的。图 2 描述了一条铁路线路基于两种不同空间模型所形成的网络。比较发现, P 空间模型忽略了两个节点之间的拓扑邻近约束, 而注重于是否有关联, 因此更适于做社区发现分析<sup>[23]</sup>。利用 P 空间模型构建的

铁路网络可用一个加权连接矩阵来表达, 矩阵元素就是所有途经两个城市的列车总数。



(a) L 空间模型 (b) P 空间模型

Fig. 2 The Space L Model and Space P Model for Train Nets

在所有 7 种火车类型中, 由于城际铁路的运行线路少, 范围分散, 并且与高铁列车在速度上匹配, 路程上互补, 因此将其合并到高铁线路中。最终获取了 315 座城市所组成的 6 个不同类型的空间网络, 共包含 4 147 条铁路运行线路数据。6 个空间网络所含城市和连接的个数等属性信息如表 1 所示。

表 1 6 个不同类型空间网络的节点和边个数

Tab. 1 The Number of Nodes and Edges of Six Networks

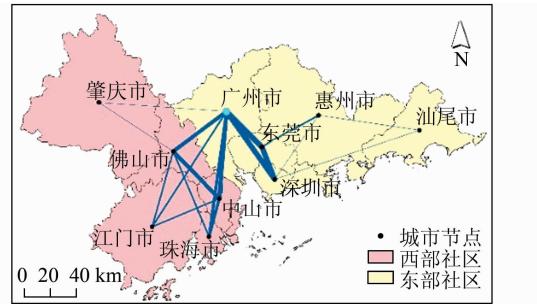
空间 网络	铁路 类型	城市 数	连接数	所有连边的平 均长度/km	所有节点间平均 距离( $\sigma$ 值)/km
CG-net	高铁	167	5 044	740	997
D-net	动车	175	1 824	501	1 174
K-net	普快	294	13 920	895	1 326
T-net	特快	212	3 812	913	1 291
X-net	临客	203	1 370	449	1 379
Z-net	卧铺	177	3 337	1 004	1 203

### 2.2 样本网络实验分析

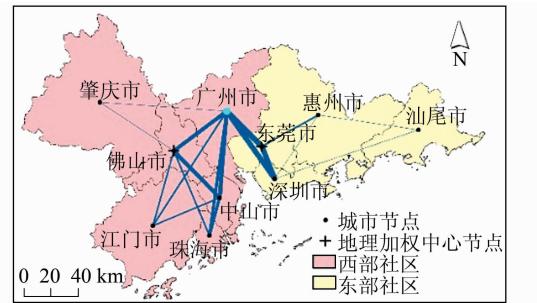
本节选用广东省的城铁网络数据作为样本数据, 进行 GND-Louvain 算法的准确性和有效性实验分析。并将其结果与基于传统模块度的 Louvain 方法(简称 NG-Louvain)进行比较。两个算法的社区分割结果都是将城铁网络划分为了两个子社区, 结果的可视化效果如下图 3 所示。

图 3 显示出在两种不同算法下, 广东省省会广州市的计算结果存在差异。在 NG-Louvain 算法中, 使用传统模块度只考虑了空间网络的拓扑关系, 此时由于广州与东南部两座距离相近的东莞和深圳市之间具有高的连接数, 因此被分割到东部社区中。而在 GND-Louvain 算法中, 综合考虑了空间网络的拓扑和距离因素, 广州被分割到西部社区中。虽然广州与东莞和深圳之间的实际连接数更高, 但由于距离邻近, 其期望连接数也很高, 导致距离模块度的值不如其与西南部城市的高。GND-Louvain 算法的结果显示出, 去掉距离

衰减的因素影响后,西南部城市实际上获得了更多额外的与广州市联系的隐含因素。相对而言,东部城市与广州市的联系则主要依赖于其与广州市的距离更近。



(a) NG-Louvain 社区分割结果



(b) GND-Louvain 社区分割结果

图3 两种不同模块度的社区发现结果比较  
Fig. 3 Community Detection Result of Two Modularities on a Sample Network

表2采用两种模块度公式计算了两个算法社区分割结果的模块度数值。NG-Louvain 算法获得了略高于 GND-Louvain 算法的传统模块度值,而 GND-Louvain 算法则获得了高出 NG-Louvain 算法很多的距离模块度值。由于图3显示基于距离模块度的 GND-Louvain 算法能获得更有价值的隐含信息,本文接下来的实验也将采用距离模块度进行社区发现的质量评价。

表2 两个社区发现算法结果的模块度数值对比

Tab. 2 Modularities of Two Partition Results

社区分割结果	传统模块度	距离模块度
NG-Louvain 社区分割结果	0.190	0.079
GND-Louvain 社区分割结果	0.181	0.118

### 2.3 铁路网络实验分析

本节针对全国铁路线路数据构建的6个真实的地理空间网络,设计了5个基于不同社区距离度量方式的算法,来对比验证 GND-Louvain 算法的准确性和稳定性。除上节不考虑距离衰减因素的 NG-Louvain 算法外,另外3个都是基于距离模块度的,分别是:①基于地理几何中心节点距离的 D-Louvain 算法<sup>[14]</sup>;②基于网络度中心节点距离的 DD-Louvain 算法;③基于空间邻近约束的

D-REDCAP ( regionalization with dynamically constrained agglomerative clustering and partitioning) 算法。其中 D-Louvain 算法和 DD-Louvain 算法的实现与 GND-Louvain 算法类似,仅在距离计算时采用了不同的中心点。此3个算法和 GND-Louvain 都采用了 Louvain 来实现模块度的最大化,以消除各方法在社区迭代过程中可能存在的其他差异。而 D-REDCAP 算法是在经典的区域分割算法 REDCAP<sup>[15]</sup> 中用距离模块度矩阵替代了原有的模块度矩阵,并且模块度最大化采用的是层次聚类方法。该方法依据最初的节点间距离矩阵获得节点合并的距离模块度增益矩阵,并不断合并存在空间邻近关系且模块度增益最大的两个节点或子社区。在社区迭代合并过程中,需要不断更新计算的是模块度增益矩阵。参考 Guo<sup>[15]</sup> 的建议,子社区合并后,新社区与其他子社区之间的模块度增益采用的是二者间所有模块度增益的平均值。而最终社区发现结果质量的衡量采用的是距离模块度,其数值越高,则说明社区发现的结果越好。其中,NG-Louvain 算法结果也使用了式(2)重新计算获得其距离模块度值,以方便进行统一比较。

实验中发现,基于 Louvain 快速展开策略实现距离模块度最大化,虽然效率快,准确性高,但稳定性不足。不同的节点输入次序会得到差异性很大的结果。因此,对4个基于 Louvain 的算法进行了重复百次以上的运算(重复次数初始化为网络节点总个数),每次运算均随机产生节点输入次序。然后对获取的距离模块度值进行范围和频次统计。图4是针对6种不同空间网络,采用5种算法的距离模块度取值结果。蓝柱是重复运算的最小模块度取值,代表算法的最差结果,其取值越大说明算法越好;红柱代表重复运算的模块度波动范围,代表了算法的稳定性,其取值越小说明算法越好。二者之和是重复运算得到的最大模块度值,代表算法的最好结果,其取值也是越大越好。

从图4可以得到如下结论:

1) 4个基于 Louvain 的算法总是获得较 D-REDCAP 更大的距离模块度。原因除了采用凝聚层次聚类的固有缺陷外,D-REDCAP 在节点合并时采用严格的空间邻域约束,此约束对模块度最大化策略影响很大。

2) 4个基于 Louvain 的算法在重复百次以上运算后,都能获得相近的最大模块度取值,说明了 Louvain 的快速展开策略的准确性较高。而对比模块度取值范围情况,GND-Louvain 算法在6个

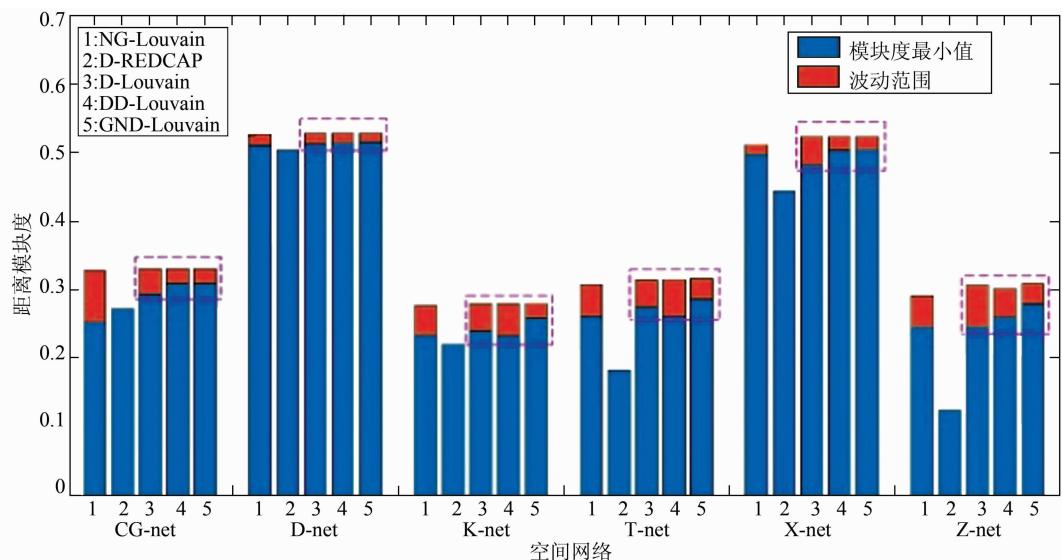


图 4 6 个空间网络社区发现结果的模块度大小比较

Fig. 4 Modularities Comparison of Six Train Nets

网络的社区发现中的波动范围总是最小的,也说明了算法是相对稳定的。

3) 相关性分析结果发现,距离模块度取值范围的波动与空间网络的连接边平均长度显著的正相关。

如图 5 所示,3 种距离模块度 Louvain 算法的稳定性受到空间网络本身的影响。但与 D-Louvain 和 DD-Louvain 相比,地理加权中心节点距离的 GND-Louvain 算法受空间网络本身的影响总是最小的,且变化的幅度也总是最小的。因此可认为,GND-Louvain 算法在准确性和稳定性上,优于其他几种距离和邻域约束的社区发现算法。

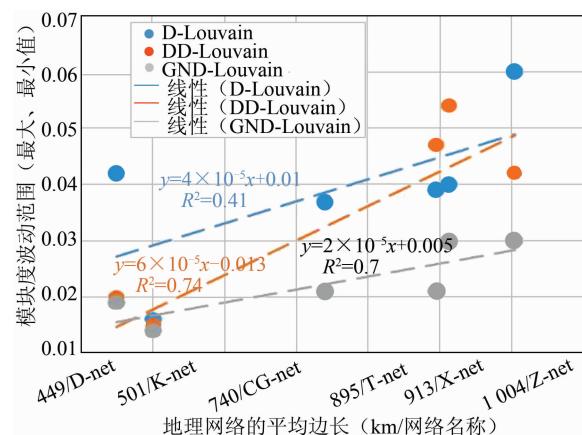


图 5 3 个距离模块度算法的模块度波动范围与网络平均边长度之间的线性回归分析

Fig. 5 Linear Regression Analysis Between Range of Modularity and Average Distance of Connections

此外,对比各算法重复实验后模块度取值的频率分布,GND-Louvain 算法在其中 4 个网络(CG-net、D-net、K-net、X-net)的实验中都是在最

大模块度处获得最高频率,而剩余两个网络(T-net 和 Z-net)中是在最大模块度处获得第二高频率。由此也说明,GND-Louvain 算法较另外 4 种算法能更容易地获得最大模块度结果,具有更高的稳定性。

## 2.4 算法稳定性改进

本节实验的目标是在不影响算法准确性的前提下,提升 GND-Louvain 算法的稳定性和效率。分析距离模块度公式,在节点合并时有 3 个因素会影响到模块度的取值:节点度、节点间连接数和节点间距离。节点度越小,节点间连接数和距离值越大,将会获得越大的模块度增益。然而,相对于一个节点,其与其他节点的连接数和距离通常是一对多的关系,无法拿来作为节点合并次序的参考。因此,选择节点度作为影响因子,以实验方法分析节点度升序和降序排列时,GND-Louvain 算法的模块度取值情况。

如图 6 所示,当节点序列固定时,GND-Louvain 算法针对各空间网络都能获得稳定的模块度取值。两类节点度排序情况下,GND-Louvain 算法的模块度实验结果的绝对差异很小。接下来采用最大最小值标准化方法,将此模块度取值与对应空间网络重复多次运算时的模块度取值范围进行标准化,计算相对差异。标准化公式如下:

$$x_{\text{normalization}} = \frac{x - \min}{\max - \min} \quad (5)$$

模块度的标准化转换用了原 GND-Louvain 算法和 D-Louvain 算法重复实验结果,获得的数据体现了节点排序后模块度取值在原算法和 D-Louvain 算法中的排名情况。标准化数值越高,

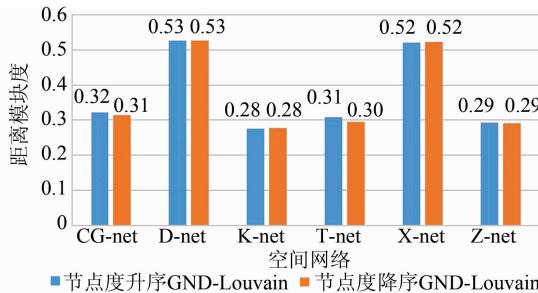


图 6 两类节点度排序情况下 GND-Louvain 算法的  
距离模块度

Fig. 6 Modularities of Two Degree Ordered GND-Louvain  
说明模块度排序值越高,即节点排序的效果越好。  
两类节点度排序情况下,GND-Louvain 算法结果  
的标准化结果如图 7 所示。

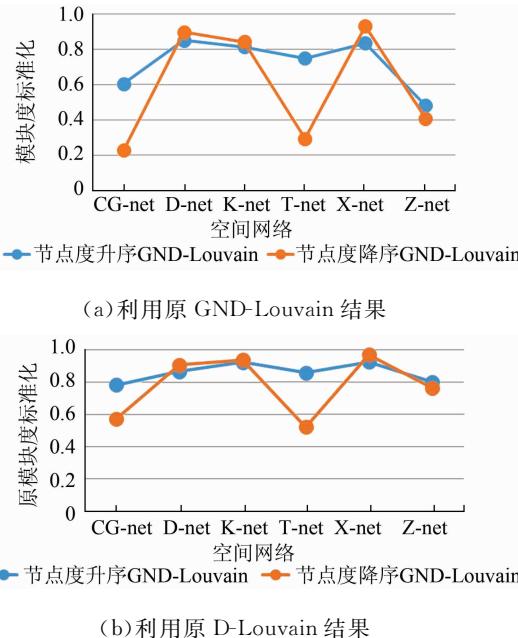


图 7 两类节点度排序下 GND-Louvain 算法模块度  
标准化值

Fig. 7 Normalized Modularities of Two Degree  
Ordered GND-Louvain

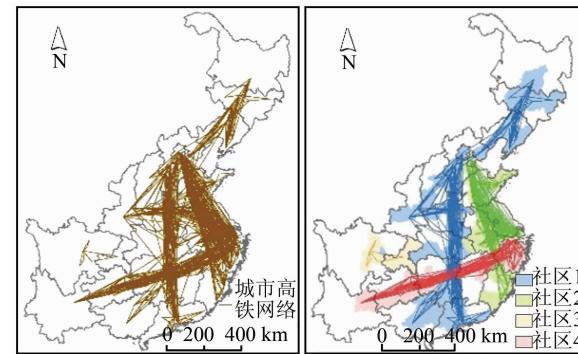
从图 7 的实验结果可以得到如下结论:

1) 节点度升序排列的 GND-Louvain 算法获得了较节点度降序排列更稳定的模块度结果。虽然节点度降序排列的 GND-Louvain 在 4 个网络(D-net、K-net、X-net 和 Z-net)中都获得了较节点度升序排列略好和相近的排序值,但其在另两个网络(CG-net 和 T-net)中却获得了较节点度升序排列低很多的排序值。分析原因是节点度降序排列时,先进行社区合并运算的节点会对模块度的变化影响很大,这导致了最终模块度取值不稳定。

2) 节点度升序排列的 GND-Louvain 算法在用原 D-Louvain 结果的模块度标准化中,总是取得较高的排名。因此利用节点度升序策略改进的

GND-Louvain 算法能提升基于几何中心节点距离和距离模度的 D-Louvain 算法的准确性和稳定性。考虑到社区合并过程中的子社区也是一个空间网络,采用空间和网络属性相结合的地理加权中心节点度量空间社区间距离的效果更好。

图 8 以城铁和高铁网络数据实验为例,展示了节点度升序排列的 GND-Louvain 算法的社区分割结果。算法共识别出 4 个社区,展示出由城铁和高铁网络构成的区域空间的交互关系和聚集模式。



(a) 城铁和高铁网络      (b) 网络社区分割结果  
图 8 城铁和高铁网络数据及 GND-Louvain 社区  
分割结果

Fig. 8 High-Speed Train Net and It's Community  
Detection Results by Using GND-Louvain

### 3 结语

与传统复杂网络的社区发现相比,空间网络的社区发现常需要将空间关系作为一个重要的约束条件,因此难度更大。本文通过实际数据的实验,对比了采用距离模块度和传统模块度进行社区发现的不同结果,以说明距离模块度在空间网络社区发现中的重要作用。首次探究了空间网络社区发现方法中采用的不同的距离度量方式对社区发现结果造成的影响程度。通过多组真实铁路网络数据的实验结果对比,验证本文所采用的基于地理加权距离的 GND-Louvain 算法的社区发现结果的准确性和稳定性最好。另一方面,分析了经典 Louvain 的快速展开模块度最大化计算策略存在不稳定性的原因和因素,并用实验证实采用节点度升序排列的 GND-Louvain 算法能在保证计算结果准确性的情况下,提高算法的稳定性。

由于实验选用的铁路网络数据仅代表了一种类型的空间网络,因此 GND-Louvain 方法在其他空间网络上进行社区发现的适用性和有效性还需进一步的验证。后续将针对不同空间网络距离衰

减函数的选取与建模、网络中心节点的选取和计算等问题开展深入的分析和实验研究。

## 参 考 文 献

- [1] Barthélémy M. Spatial Networks[J]. *Physics Reports*, 2011, 499(1-3): 1-101
- [2] Fortunato S. Community Detection in Graphs[J]. *Physics Reports*, 2010, 486(3): 75-174
- [3] Tang Lei, Liu Huan. Community Detection and Mining in Social Media[M]. Wen Yimin, Bi Yingzhou. Beijing: Morgan & Claypool 2013(唐磊, 刘欢. 社会计算: 社区发现和社会媒体挖掘[M]. 刘益民, 闭应洲. 北京: 机械工业出版社, 2013)
- [4] Gao S, Liu Y, Wang Y L, et al. Discovering Spatial Interaction Communities from Mobile Phone Data[J]. *Transactions in GIS*, 2013, 17(3): 463-481
- [5] Yang Jing, Cheng Changxiu, Li Xiaolan, et al. A Similarity Evaluation Method on Spatial Patterns of Network Structures: A Case Study About Bejing Traffic-Network Backbones from 1938 to 2014 [J]. *Geomatics and Information Science of Wuhan University*, 2016, 41(12): 1 593-1 598(杨静, 程昌秀, 李晓岚, 等. 网络结构空间格局相似度分析——以1938—2014年北京市骨干交通网络为例 [J]. 武汉大学学报·信息科学版, 2016, 41(12): 1 593-1 598)
- [6] Liu X, Gong L, Gong Y X, et al. Revealing Travel Patterns and City Structure with Taxi Trip Data [J]. *Journal of Transport Geography*, 2015, 43: 78-90
- [7] Liu Yu, Gong Li, Tong Qingxi. Quantifying the Distance Effect in Spatial Interactions [J]. *Acta Scientiarum Naturalium Universitatis Pekinensis*, 2014, 50(3): 526-534(刘瑜, 龚俐, 童庆禧. 空间交互作用中的距离影响及定量分析 [J]. 北京大学学报(自然科学版), 2014, 50(3): 526-534)
- [8] Chen Y G. The Distance-Decay Function of Geographical Gravity Model: Power Law or Exponential Law [J]. *Chaos, Solitons & Fractals*, 2015, 77: 174-189
- [9] Xu Jun, Chen Yu, Xu Minzheng. Data Mining and Applications in Spatial Networks [J]. *Communications of the CCF*, 2015, 11(11): 40-49(许珺, 陈娱, 徐敏政. 空间网络的数据挖掘和应用 [J]. 中国计算机学会通讯, 2015, 11(11): 40-49)
- [10] Guimera R, Mossa S, Turtschi A, et al. The Worldwide Air Transportation Network: Anomalous Centrality, Community Structure, and Cities' Global Roles[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2005, 102(22): 7 794-7 799
- [11] Guo D S. Flow Mapping and Multivariate Visualization of Large Spatial Interaction Data[J]. *IEEE Transactions on Visualization and Computer Graphics*, 2009, 15(6): 1 041-1 048
- [12] Chen Y, Xu J, Xu M Z. Finding Community Structure in Spatially Constrained Complex Networks [J]. *International Journal of Geographical Information Science*, 2015, 29(6): 889-911
- [13] Expert P, Evans T S, Blondel V D, et al. Uncovering Space-Independent Communities in Spatial Networks[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2011, 108(19): 7 663-7 668
- [14] Liu X, Murata T, Wakita K. Detecting Network Communities Beyond Assortativity-Related Attributes [J]. *Physical Review E*, 2014, 90(1): 012806
- [15] Guo D S. Regionalization with Dynamically Constrained Agglomerative Clustering and Partitioning (REDCAP) [J]. *International Journal of Geographical Information Science*, 2008, 22(7): 801-823
- [16] Blondel V D, Guillaume J L, Lambiotte R, et al. Fast Unfolding of Communities in Large Networks [J]. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10): 10008
- [17] Sobolevsky S, Campari R, Belyi A, et al. General Optimization Technique for High-quality Community Detection in Complex Networks [J]. *Physical Review E*, 2014, 90(1): 012811
- [18] Liu Jun. Lectures on Whole Network Application—A Practical Guide to UCINET[M]. 2nd ed. Shanghai: Gezhi Press, 2014(刘军. 整体网分析——UCINET软件使用指南[M]. 2 版. 上海: 格致出版社, 2014)
- [19] Burt J E, Barber G M, Rigby D L. Elementary Statistics for Geographers[M]. 3rd ed. New York: Guilford Press, 2009
- [20] Lee J, Wong D W. Statistical Analysis with ArcView GIS[M]. New York: John Wiley & Sons, 2001
- [21] Newman M E. Modularity and Community Structure in Networks[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2006, 103(23): 8 577-8 582
- [22] Seaton K A, Hackett L M. Stations, Trains and Small-World Networks[J]. *Physica A: Statistical Mechanics and Its Applications*, 2004, 339 (3): 635-644
- [23] Zhong Ke, Xiao Yu, Xu Jun, et al. Measuring City Centralities Based on the Train Network of China

[J]. *Journal of Geo-Information Science*, 2012,  
14(1): 85-93(钟柯, 肖昱, 许珺, 等. 基于列车运行

网络的中国城市中心性分析 [J]. 地球信息科学学  
报, 2012, 14(1): 85-93)

## Community Detection Algorithm Based on Geographical Weighted Central Node Distance

WAN You<sup>1,2</sup> LIU Yaolin<sup>1,2,3</sup>

1 School of Resource and Environmental Sciences, Wuhan University, Wuhan 430079, China

2 Key Laboratory of Geographic Information System, Ministry of Education, Wuhan University, Wuhan 430079, China

3 Collaborative Innovation Center of Geospatial Technology, Wuhan 430079, China

**Abstract:** Community detection is an important topic in spatial network researches. It can discover interesting spatial patterns, which help understand the spatial structures hidden in the networks. However, community detection in spatial networks is more difficult than traditional networks. Since it has to consider some other spatial correlations, e.g., spatial contiguity, geographic distance. This paper proposes a new spatial community detection algorithm—geographical weighted central node distance based Louvain method (GND-Louvain). It uses a geographical and network dual-constrain to measure the distance and to calculate the distances decay effects between spatial nodes and meta-communities. It also extends the famous fast unfolding community detection algorithm—Louvain, by using the distance-modularity. In addition, a merge order is defined during the GND-Louvain's optimization process to get high quality and stable community results. Comparative experiments are designed on five different community detection methods based on five different spatial constraints, which are ① no distance constraint, ② spatial contiguity edge constraint, ③ geometry median center distance, ④ degree center distance, and ⑤ degree weighted geometry median center distance. The experimental data comes from the Chinese railway lines. And results prove that the newly defined GND-Louvain algorithm can produce more accurate spatial communities than others. And the merge orders also ensure the high quality and stable results.

**Key words:** community detection; spatial network; complex network; spatial constraint; modularity

**First author:** WAN You, PhD, lecturer, specializes in the theories and methods of geo-big data and spatial data mining. E-mail: wanyou@whu.edu.cn

**Corresponding author:** LIU Yaolin, PhD, professor. E-mail: yaolin610@163.com

**Foundation support:** The National Key Research and Development Program of China, No. 2017YFB0503601; the National Natural Science Foundation of China, No. 41471327.