



(21) 申请号 202110665380.0

(22) 申请日 2021.06.16

(65) 同一申请的已公布的文献号

申请公布号 CN 113704570 A

(43) 申请公布日 2021.11.26

(73) 专利权人 香港理工大学深圳研究院

地址 518057 广东省深圳市南山区粤海街道高新技术产业园南区粤兴一道18号
香港理工大学产学研大楼205室

(72) 发明人 陈家进 刘松柏 林秋镇 李青

(74) 专利代理机构 深圳市君胜知识产权代理事务所(普通合伙) 44268

专利代理师 谢松 徐凯凯

(51) Int. Cl.

G06F 18/2323 (2023.01)

G06F 16/901 (2019.01)

G06N 3/126 (2023.01)

G06N 3/0455 (2023.01)

G06N 3/084 (2023.01)

G06N 3/0895 (2023.01)

(56) 对比文件

CN 102413029 A, 2012.04.11

CN 102708407 A, 2012.10.03

CN 104268629 A, 2015.01.07

CN 106934722 A, 2017.07.07

CN 107276843 A, 2017.10.20

US 2020210864 A1, 2020.07.02

梁世娇. 基于进化多目标优化的复杂网络社区检测算法的研究. 中国优秀硕士学位论文全文数据库(基础科学辑). 2021, (第2期), 正文全文.

王聪; 柴争义. 基于多目标进化的复杂网络社区检测. 计算机技术与发展. 2020, (第06期), 正文全文.

审查员 张倩玉

权利要求书2页 说明书9页 附图4页

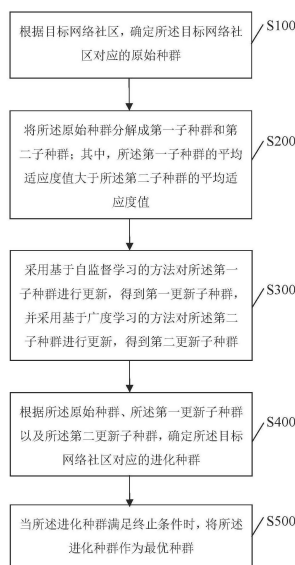
(54) 发明名称

基于自监督学习式进化的大规模复杂网络社区检测方法

(57) 摘要

本发明公开了基于自监督学习式进化的大规模复杂网络社区检测方法, 包括步骤: 根据目标网络社区, 确定目标网络社区对应的原始种群; 将原始种群分解成第一子种群和第二子种群; 采用基于自监督学习的方法对第一子种群进行更新, 得到第一更新子种群, 并采用基于广度学习的方法对第二子种群进行更新, 得到第二更新子种群; 根据原始种群、第一更新子种群以及第二更新子种群, 确定目标网络社区对应的进化种群; 当进化种群满足终止条件时, 将进化种群作为最优种群。本发明将大规模复杂网络社区结构检测问题建模成一个两目标的优化问题, 通过基于自监督学习式进化方法和基于广度学习的方法同时优化这两个目标, 实现对大规模复杂网

络的最优社区结构检测。



1. 一种基于自监督学习式进化的大规模复杂网络社区检测方法,其特征在于,包括步骤:

根据目标网络社区,确定所述目标网络社区对应的原始种群;

将所述原始种群分解成第一子种群和第二子种群;其中,所述第一子种群的平均适应度值大于所述第二子种群的平均适应度值;

采用基于自监督学习的方法对所述第一子种群进行更新,得到第一更新子种群,并采用基于广度学习的方法对所述第二子种群进行更新,得到第二更新子种群;

根据所述原始种群、所述第一更新子种群以及所述第二更新子种群,确定所述目标网络社区对应的进化种群;

当所述进化种群满足终止条件时,将所述进化种群作为最优种群;

所述根据目标网络社区,确定所述目标网络社区对应的原始种群,具体包括:

根据目标网络社区,确定所述目标网络社区对应的目标网络;

根据所述目标网络,确定所述目标网络社区对应的原始种群;

所述根据所述目标网络,确定所述目标网络社区对应的原始种群,具体包括:

确定所述目标网络中各网络节点各自分别对应的度值;其中,所述度值为连接的网络节点的数量;

根据所有各网络节点的度值,确定所述目标网络的候选社区中心节点;

根据所述候选社区中心节点,确定所述目标网络社区对应的原始种群;

所述目标网络社区是指由若干个网络节点与各网络节点之间的关系所形成的社区,网络节点的数量超过预设节点数量,且不同网络节点之间边的数量超过预设边数量,网络节点的数量为十万级或百万级;所述目标网络社区包括:互联网世界中的万维网、交通系统中的路线网、电子领域中的集成电路、电力系统中的大型电力网络、生物系统中的细胞神经网络到蛋白质互相作用网络、社会关系中的社交网络中的一种。

2. 根据权利要求1所述的基于自监督学习式进化的大规模复杂网络社区检测方法,其特征在于,所述检测方法还包括:

当所述进化种群不满足终止条件时,将所述进化种群作为原始种群,并继续执行将所述原始种群分解成第一子种群和第二子种群的步骤,直至所述进化种群满足终止条件时,将所述进化种群作为最优种群。

3. 根据权利要求1所述的基于自监督学习式进化的大规模复杂网络社区检测方法,其特征在于,所述采用基于自监督学习的方法对所述第一子种群进行更新,得到第一更新子种群,具体包括:

对所述目标网络中的所有网络节点进行分组,得到若干个变量组;

确定各变量组各自分别对应的自编码器;其中,所述自编码器包括编码单元和解码单元;

针对所述第一子种群中的每个解个体,根据该解个体和所有编码单元,得到该解个体对应的表示空间;对所述表示空间进行进化操作,得到新的表示空间;根据所述新的表示空间和所有解码单元,确定该解个体对应的更新的解个体,以得到第一更新子种群;其中,所述表示空间的维度小于所述解个体的维度。

4. 根据权利要求3所述的基于自监督学习式进化的大规模复杂网络社区检测方法,其

特征在于,所述采用基于广度学习的方法对所述第二子种群进行更新,得到第二更新子种群,还包括:

确定所述第一子种群的若干个目标解个体;其中,所述目标解个体的数量与所述变量组的数量相同;

对于所述第二子种群中的每个解个体,根据所有目标解个体以及所有变量组对该解个体进行进化操作,确定该解个体对应的更新的解个体,以得到第二更新子种群。

5.根据权利要求1所述的基于自监督学习式进化的大规模复杂网络社区检测方法,其特征在于,所述将所述原始种群分解成第一子种群和第二子种群,具体包括:

确定所述原始种群对应的目标空间;所述目标空间包括:各解个体各自分别对应的目标函数值,所述目标函数值包括KKM目标函数值和RC目标函数值;

将所述目标空间分解成若干个子空间;其中,每个子空间包括至少两个解个体分别对应的目标函数值;

确定各子空间中各解个体各自分别对应的适应度值;

将各子空间的解个体分为第一类解个体和第二类解个体;其中,所述第一类解个体中解个体的适应度值大于所述第二类解个体中解个体的适应度值;

根据所有子空间中第一类解个体,确定所述第一子种群;

根据所有子空间中第二类解个体,确定所述第二子种群。

6.根据权利要求1所述的基于自监督学习式进化的大规模复杂网络社区检测方法,其特征在于,所述根据所述原始种群、所述第一更新子种群以及所述第二更新子种群,确定所述目标网络社区对应的进化种群,具体包括:

根据所述原始种群中各解个体各自分别对应的目标函数值、所述第一更新子种群中各解个体各自分别对应的目标函数值以及所述第二更新子种群中各解个体各自分别对应的目标函数值,确定所述目标网络社区对应的进化种群。

7.一种计算机设备,包括存储器和处理器,所述存储器存储有计算机程序,其特征在于,所述处理器执行所述计算机程序时实现权利要求1至6中任一项所述方法的步骤。

8.一种计算机可读存储介质,其上存储有计算机程序,其特征在于,所述计算机程序被处理器执行时实现权利要求1至6中任一项所述的方法的步骤。

基于自监督学习式进化的大规模复杂网络社区检测方法

技术领域

[0001] 本发明涉及计算机技术领域,尤其涉及的是一种基于自监督学习式进化的大规模复杂网络社区检测方法。

背景技术

[0002] 进入21世纪的大数据时代,以因特网为代表的信息技术的迅猛发展使人类社会大步迈入了大规模复杂网络时代。在现实世界中许多系统以网络的形式存在,从互联网世界中的万维网到交通系统中的路线网,从电子领域中的超大规模集成电路到电力系统中的大型电力网络,从生物系统中的细胞神经网络到蛋白质互相作用网络,从社会关系中的社交网络到科学家之间的合作关系网,大规模复杂网络无处不在。社区结构检测是大规模复杂网络分析中的一项关键任务,它有助于理解以大规模复杂网络为代表的系统的特性,且对实际应用具有重要的意义。大规模复杂的网络结构由几十万甚至上百万对节点与节点之间的关系组成。大规模复杂网络研究的主要问题之一是对社区结构的高效快速检测。现实世界的网络通常由功能单元组成,这些功能单元以网络模块或社区的形式表现出来,同一社区内部的节点在连接程度非常紧密,不同社区之间的连接相对稀疏。社区结构作为复杂网络的重要特性,在一定程度上可以反映网络的重要特征。因此,识别社区结构对于描述网络组织结构和理解复杂系统至关重要。但是,随着网络规模的不断扩大使得网络结构变得非常复杂,很多传统的方法在处理大规模复杂网络的社区结构检测时存在一些缺陷。现有技术中存在的以下问题:(1) 现有基于聚类的社区检测方法大多需要一定的先验知识来预先指明目标网络中社区的数量,而对于大规模复杂网络社区检测,网络中社区的数量是很难预先确定的;(2) 现有的基于进化算法的社区检测方法只能处理小规模的网络社区检测问题,对于大规模复杂网络社区的检测,传统进化算法的搜索效率低且收敛速度很慢,因此很难实现的对大规模复杂网络社区的快速高效检测。

[0003] 因此,现有技术还有待于改进和发展。

发明内容

[0004] 本发明要解决的技术问题在于,针对现有技术的上述缺陷,提供一种基于自监督学习式进化的大规模复杂网络社区检测方法,旨在解决现有技术中很难实现对大规模复杂网络社区的快速高效检测的问题。

[0005] 本发明解决技术问题所采用的技术方案如下:

[0006] 一种基于自监督学习式进化的大规模复杂网络社区检测方法,其中,包括步骤:

[0007] 根据目标网络社区,确定所述目标网络社区对应的原始种群;

[0008] 将所述原始种群分解成第一子种群和第二子种群;其中,所述第一子种群的平均适应度值大于所述第二子种群的平均适应度值;

[0009] 采用基于自监督学习的方法对所述第一子种群进行更新,得到第一更新子种群,并采用基于广度学习的方法对所述第二子种群进行更新,得到第二更新子种群;

[0010] 根据所述原始种群、所述第一更新子种群以及所述第二更新子种群,确定所述目标网络社区对应的进化种群;

[0011] 当所述进化种群满足终止条件时,将所述进化种群作为最优种群。

[0012] 所述的基于自监督学习式进化的大规模复杂网络社区检测方法,其中,所述检测方法还包括:

[0013] 当所述进化种群不满足终止条件时,将所述进化种群作为原始种群,并继续执行将所述原始种群分解成第一子种群和第二子种群的步骤,直至所述进化种群满足终止条件时,将所述进化种群作为最优种群。

[0014] 所述的基于自监督学习式进化的大规模复杂网络社区检测方法,其中,所述采用基于自监督学习的方法对所述第一子种群进行更新,得到第一更新子种群,具体包括:

[0015] 对所述目标网络中的所有网络节点进行分组,得到若干个变量组;

[0016] 确定各变量组各自分别对应的自编码器;其中,所述自编码器包括编码单元和解码单元;

[0017] 针对所述第一子种群中的每个解个体,根据该解个体和所有编码单元,得到该解个体对应的表示空间;对所述表示空间进行进化操作,得到新的表示空间;根据所述新的表示空间和所有解码单元,确定该解个体对应的更新的解个体,以得到第一更新子种群;其中,所述表示空间的维度小于所述解个体的维度。

[0018] 所述的基于自监督学习式进化的大规模复杂网络社区检测方法,其中,所述采用基于广度学习的方法对所述第二子种群进行更新,得到第二更新子种群,还包括:

[0019] 确定所述第一子种群的若干个目标解个体;其中,所述目标解个体的数量与所述变量组的数量相同;

[0020] 对于所述第二子种群中的每个解个体,根据所有目标解个体以及所有变量组对该解个体进行进化操作,确定该解个体对应的更新的解个体,以得到第二更新子种群。

[0021] 所述的基于自监督学习式进化的大规模复杂网络社区检测方法,其中,所述根据目标网络社区,确定所述目标网络社区对应的原始种群,具体包括:

[0022] 根据目标网络社区,确定所述目标网络社区对应的目标网络;

[0023] 根据所述目标网络,确定所述目标网络社区对应的原始种群。

[0024] 所述的基于自监督学习式进化的大规模复杂网络社区检测方法,其中,所述目标网络包括若干个网络节点;

[0025] 所述根据所述目标网络,确定所述目标网络社区对应的原始种群,具体包括:

[0026] 确定所述目标网络中各网络节点各自分别对应的度值;其中,所述度值为连接的网络节点的数量;

[0027] 根据所有各网络节点的度值,确定所述目标网络的候选社区中心节点;

[0028] 根据所述候选社区中心节点,确定所述目标网络社区对应的原始种群。

[0029] 所述的基于自监督学习式进化的大规模复杂网络社区检测方法,其中,所述将所述原始种群分解成第一子种群和第二子种群,具体包括:

[0030] 确定所述原始种群对应的目标空间;所述目标空间包括:各解个体各自分别对应的目标函数值,所述目标函数值包括KKM目标函数值和RC目标函数值;

[0031] 将所述目标空间分解成若干个子空间;其中,每个子空间包括至少两个解个体分

别对应的目标函数值；

[0032] 确定各子空间中各解个体各自分别对应的适应度值；

[0033] 将各子空间的解个体分为第一类解个体和第二类解个体；其中，所述第一类解个体中解个体的适应度值大于所述第二类解个体中解个体的适应度值；

[0034] 根据所有子空间中第一类解个体，确定所述第一子种群；

[0035] 根据所有子空间中第二类解个体，确定所述第二子种群。

[0036] 所述的基于自监督学习式进化的大规模复杂网络社区检测方法，其中，所述根据所述原始种群、所述第一更新子种群以及所述第二更新子种群，确定所述目标网络社区对应的进化种群，具体包括：

[0037] 根据所述原始种群中各解个体各自分别对应的目标函数值、所述第一更新子种群中各解个体各自分别对应的目标函数值以及所述第二更新子种群中各解个体各自分别对应的目标函数值，确定所述目标网络社区对应的进化种群。

[0038] 一种计算机设备，包括存储器和处理器，所述存储器存储有计算机程序，其中，所述处理器执行所述计算机程序时实现上述任一项所述方法的步骤。

[0039] 一种计算机可读存储介质，其上存储有计算机程序，其中，所述计算机程序被处理器执行时实现上述任一项所述的方法的步骤。

[0040] 有益效果：本发明将大规模复杂网络社区结构检测问题建模成目标的优化问题，通过基于自监督学习式进化方法和基于广度学习的方法优化目标，实现对大规模复杂网络的最优社区结构检测。

附图说明

[0041] 图1是本发明中基于自监督学习式进化的大规模复杂网络社区检测方法的第一流程图。

[0042] 图2是本发明中基于参考方向引导的聚类方法分解得到子种群的示意图。

[0043] 图3是本发明中随机分解变量组的示意图。

[0044] 图4是本发明中基于自监督学习的方法更新子种群的示意图。

[0045] 图5是本发明中本发明中基于自监督学习式进化的大规模复杂网络社区检测方法的第二流程图。

具体实施方式

[0046] 为使本发明的目的、技术方案及优点更加清楚、明确，以下参照附图并举实施例对本发明进一步详细说明。应当理解，此处所描述的具体实施例仅仅用以解释本发明，并不用于限定本发明。

[0047] 请同时参阅图1-图5，本发明提供了一种基于自监督学习式进化的大规模复杂网络社区检测方法的一些实施例。

[0048] 如图5所示，本发明实施例中基于自监督学习式进化的大规模复杂网络社区检测方法，包括以下步骤：

[0049] 步骤S100、根据目标网络社区，确定所述目标网络社区对应的原始种群。

[0050] 具体地，目标网络社区是指由大量节点与各节点之间的关系所形成的社区。原始

种群是指根据目标网络社区初始化得到的种群。目标网络社区具体为大规模复杂网络社区,本发明将大规模复杂网络社区结构检测问题建模成一个两目标的优化问题。先根据目标网络社区,得到目标网络社区对应的原始种群。由于原始种群是根据目标网络社区初始化得到的种群,没有进行两目标优化,因此,原始种群并不是最优种群。

[0051] 步骤S100具体包括:

[0052] 步骤S110、根据目标网络社区,确定所述目标网络社区对应的目标网络。

[0053] 具体地,这里可以将一个大规模复杂网络社区抽象成一个目标网络,目标网络可以用无向图表示,即 $G(V, E)$,其中 $V = \{v_1, v_2, \dots, v_n\}$ 表示目标网络中所有网络节点的集合, $E = \{(v_i, v_j) | v_i \in V, v_j \in V, \text{and } i \neq j\}$ 表示目标网络中不同网络节点之间边的集合, $n = |V|$ 表示目标网络中网络节点的数目。

[0054] $G(V, E)$ 可以进一步用它的邻接矩阵 $A = (A_{i,j})^{n \times n}$ 表示,其中 $A_{i,j} = A_{j,i}$ 且 $A_{i,j} \in \{0, 1\}$ 。这里一方面 $A_{i,j} = 0$ 表示 $G(V, E)$ 中的第 i 个网络节点和第 j 个网络节点之间没有边相连,另一方面 $A_{i,j} = 1$ 表示 $G(V, E)$ 中的第 i 个网络节点和第 j 个网络节点之间有边相连。

[0055] 网络社区的网络节点的数量超过预设节点数量,且不同网络节点之间边的数量超过预设边数量时,则可以将该网络社区作为大规模复杂网络社区。本发明中所述目标网络包括若干个网络节点,网络节点的数量超过预设节点数量,不同网络节点之间边的数量超过预设边数量。

[0056] 将对目标网络 $G(V, E)$ 进行社区结构检测建模成一个两目标的优化问题,目的是通过寻找一组最优中心节点来实现最优的社区结构检测,主要包括以下步骤:

[0057] 目标网络 $G(V, E)$ 中的每一个网络节点都可以潜在地看作是某一个社区的中心节点,中心节点的个数用 k 表示,这也就意味着目标网络 $G(V, E)$ 被检测为以各自的中心节点为聚类中心的 k 个社区。

[0058] 确定社区中心节点的过程中,每一个解个体可以定义为一个 n 维的二进制向量 $b = (b_1, b_2, \dots, b_i, \dots, b_n) \in \{0, 1\}^n$,其中 n 表示整个网络中节点的个数, b_i 表示第 i 个网络节点 v_i 是否为中心节点,如果 v_i 是中心节点那么 $b_i = 1$,否则 $b_i = 0$ 。

[0059] 确定社区中心节点后,整个目标网络中的所有网络节点可以划分为两个类别,即中心节点的集合 CN ,以及非中心节点的集合 NC 。

[0060] 通过计算 CN 中的中心节点以及 NC 中的非中心节点之间的距离来定义它们之间的关系矩阵 $U = (U_{i,j})^{(n-k) \times k}$,其中 $U_{i,j}$ 表示 NC 中的第 i 个非中心节点 NC_i 和 CN 中的第 j 个中心节点 CN_j 之间的隶属度, $U_{i,j}$ 的具体计算如以下公式所示

$$U_{i,j} = \frac{1}{\sum_{l=1}^k \left(\frac{\text{dis}(NC_i, CN_l)^2}{\text{dis}(NC_i, CN_l)} \right)}$$

[0061]

[0062] 其中 $\text{dis}(NC_i, CN_j)$ 表示节点 NC_i 和节点 CN_j 之间的基于谱聚类的欧式距离,1表示社区的序号, k 表示社区的数量。

[0063] 把 CN 中的第 j 个中心节点 CN_j 当成是第 j 个社区 C_j 的中心点,然后根据非中心节点与中心节点之间的关系矩阵 U ,如果 $U_{i,j} = \max_{l=1, \dots, k} U_{i,l}$,就可以把 NC 中的第 i 个非中心节点 NC_i 划分到第 j 个社区 C_j 中。

[0064] 可以把目标网络 $G(V, E)$ 检测为具有 k 个社区的结构 $C = \{C_1, C_2, \dots, C_k\}$,那么目标网

络 $G(V, E)$ 中社区结构检测的两个优化目标就可以定义为如下:

$$[0065] \quad \min \begin{cases} \text{KKM} = 2(n-k) - \sum_{i=1}^k \left(\frac{L(C_i, C_j)}{|C_i|} \right) \\ \text{RC} = \sum_{i=1}^k \left(\frac{L(C_i, V \setminus C_j)}{|C_i|} \right) \end{cases}$$

[0066] 其中 $L(C_i, C_j) = \sum_{u \in C_i, w \in C_j} A_{u,w}$, $A_{u,w}$ 是目标网络 $G(V, E)$ 的邻接矩阵。这里的第一个目标函数KKM表示所有社区的内部链接密度,第二个目标RC表示不同社区之间的外部链接密度。通过最小化这两个目标函数可以确保最终的检测结果中每一个社区内网络节点之间的连接是密切的而不同社区之间的联系是稀疏的。

[0067] 步骤S120、根据所述目标网络,确定所述目标网络社区对应的原始种群。

[0068] 具体地,确定目标网络后,可以根据目标网络确定目标网络社区对应的原始种群。

[0069] 步骤S120具体包括:

[0070] 步骤S121、确定所述目标网络中各网络节点各自分别对应的度值;其中,所述度值为连接的网络节点的数量。

[0071] 步骤S122、根据所有各网络节点的度值,确定所述目标网络的候选社区中心节点。

[0072] 步骤S123、根据所述候选社区中心节点,确定所述目标网络社区对应的原始种群。

[0073] 目标网络中每一网络节点的度值定义为与之相连的网络节点的个数,这样就可以计算出目标网络中每个网络节点的度值。例如,目标网络中某一网络节点与7个其它网络节点连接,则该网络节点的度值为7;再如,目标网络中某一网络节点与10个其它网络节点连接,则该网络节点的度值为10。

[0074] 将度值较大的网络节点以较大的概率初始化为候选中心点,因为位于社区中心的网络节点很可能与其他网络节点有更多的连接。因此,初始的解个体中度值越大的网络节点越有可能成为中心节点,即对应的二进制变量大概率 $b_i = 1$ 。例如,度值最大的网络节点初始化为中心节点的概率为95%,而度值最小的网络节点初始化为中心节点的概率为5%。

[0075] 根据以上的策略初始化一个拥有N个解个体的原始种群P,其中,每一个解个体是n维的二进制向量b,因此可形成一个n维的变量空间,当n的值非常大时(例如n的值达到几十万甚至几百万),也即网络中的网络节点非常多时,称之为大规模复杂网络的变量空间或者决策空间。

[0076] 步骤S200、将所述原始种群分解成第一子种群和第二子种群;其中,所述第一子种群的平均适应度值大于所述第二子种群的平均适应度值。

[0077] 具体地,平均适应度值是指原始种群中各解个体的适应度值的平均值,通过平均适应度值将原始种群分解成第一子种群和第二子种群。第一子种群的平均适应度值大于第二子种群的平均适应度值。

[0078] 步骤S200具体包括:

[0079] 步骤S210、确定所述原始种群对应的目标空间;所述目标空间包括:各解个体各自分别对应的目标函数值,所述目标函数值包括KKM目标函数值和RC目标函数值。

[0080] 步骤S220、将所述目标空间分解成若干个子空间;其中,每个子空间包括至少两个

解个体分别对应的目标函数值。

[0081] 步骤S230、确定各子空间中各解个体各自分别对应的适应度值。

[0082] 步骤S240、将各子空间的解个体分为第一类解个体和第二类解个体；其中，所述第一类解个体中解个体的适应度值大于所述第二类解个体中解个体的适应度值。

[0083] 步骤S250、根据所有子空间中第一类解个体，确定所述第一子种群。

[0084] 步骤S260、根据所有子空间中第二类解个体，确定所述第二子种群。

[0085] 目标网络G(V,E)中社区结构检测的两个优化目标为目标函数KKM和RC，计算出原始种群P中的每一个解个体的目标函数值，从而构成一个二维的目标空间，目标空间中的目标函数值有KKM目标函数值和RC目标函数值，因此，该目标空间为二维的目标空间。

[0086] 在二维的目标空间中采用一种基于参考方向引导的聚类方法将原始种群P分解为两个子种群：第一类子种群 P_1 和第二类子种群 P_2 ，具体如图2所示。

[0087] 首先，在基于KKM和RC的二维目标空间中均匀地采样N个参考向量 r^1, r^2, \dots, r^N ，这样，整个二维目标空间被划分为N个子空间。

[0088] 根据参考向量引导的权值和来计算每个子空间中解个体的适应度值。

[0089] 将适应度值较大的解个体聚为第一类解个体，然后将适应度值较小的解个体聚为第二类解个体。例如，以子空间中解个体的适应度值的中位值作为参考，适应度值大于或等于该中位值的解个体归为第一类解个体，适应度值小于该中位值的解个体归为第二类解个体。当然，还可以以子空间中解个体的适应度值的平均值作为参考，适应度值大于或等于该平均值的解个体归为第一类解个体，适应度值小于该平均值的解个体归为第二类解个体。

[0090] 将所有子空间的根据所有子空间中第一类解个体聚为所述第一子种群。根据所有子空间中第二类解个体聚为所述第二子种群，从而得到第一类子种群 P_1 和第二类子种群 P_2 。当然，由于第一类解个体的适应度值大于第二类解个体的适应度值，则第一类子种群 P_1 的平均适应度值大于第二类子种群 P_2 的平均适应度值。

[0091] 相比于 P_2 中的解个体的质量， P_1 中的解个体对于优化KKM和RC这两个目标整体表现出更好的收敛性和多样性。适应度值是指解个体的权值和，具体为解个体中目标的值乘以相应的参考向量中的权值然后再相加。

[0092] 步骤S300、采用基于自监督学习的方法对所述第一子种群进行更新，得到第一更新子种群，并采用基于广度学习的方法对所述第二子种群进行更新，得到第二更新子种群。

[0093] 具体地，在得到两个子种群后，采用不同方法分别对第一子种群和第二子种群进行更新。采用基于自监督学习的方法对所述第一子种群进行更新，且采用基于广度学习的方法对所述第二子种群进行更新。

[0094] 步骤S300具体包括：

[0095] 步骤S310、对所述目标网络中的所有网络节点进行分组，得到若干个变量组。

[0096] 步骤S320、确定各变量组各自分别对应的自编码器；其中，所述自编码器包括编码单元和解码单元。

[0097] 步骤S330、针对所述第一子种群中的每个解个体，根据该解个体和所有编码单元，得到该解个体对应的表示空间；对所述表示空间进行进化操作，得到新的表示空间；根据所述新的表示空间和所有解码单元，确定该解个体对应的更新的解个体，以得到第一更新子种群；其中，所述表示空间的维度小于所述解个体的维度。

[0098] 步骤S340、确定所述第一子种群的若干个目标解个体；其中，所述目标解个体的数量与所述变量组的数量相同。

[0099] 步骤S350、对于所述第二子种群中的每个解个体，根据所有目标解个体以及所有变量组对该解个体进行进化操作，确定该解个体对应的更新的解个体，以得到第二更新子种群。

[0100] 将整个大规模复杂网络的n维变量空间随机分解为M个变量组： $g_1, g_2, \dots, g_i, \dots, g_j, \dots, g_M$ ，如图3所示。

[0101] 每一个变量组 g_i 是一组随机网络节点的集合 $g_i = \{v_{i,1}, v_{i,2}, \dots, v_{i,m}\}$ ，且 $g_i \cap g_j = \emptyset \quad \forall i \neq j$ ， i, j 表示变量组序号， M 表示变量组的数量， m 表示变量组中网络节点的数量。

[0102] 基于变量分组， P_1 和 P_2 中的每一个解个体的变量空间可以相应的分解为M个子变量空间，目的是通过对大规模的变量空间分化为多个低维的子空间，然后分别在子空间中进行进化搜索，即分而治之的策略来提高进化搜索的有效性。

[0103] 对第一子种群 P_1 中的每个解个体按照基于自监督学习的方法进行更新。

[0104] 根据M个变量分组 g_1, g_2, \dots, g_M ，用 P_1 中的解个体训练出M个自编码器，如下图4所示。在训练自编码器时，可以采用基于梯度下降的反向传播方法进行训练，损失函数可以是最小平方误差。

[0105] 每个变量分组对应一个自编码器，其中每一个自编码器是一个3层的神经网络，包括输入层、隐藏层和输出层，输入层和输出层中的神经元个数都是与其对应的变量组中的变量个数，而隐藏层神经元的个数为1个。

[0106] 根据以上训练好的自编码器， P_1 中的每个解个体可以自监督式的得到它的一个M维的表示 (w_1, w_2, \dots, w_M) ，其中 w_i 代表它通过与第i个变量组 g_i 相对应的自编码器而得到的隐藏层上神经元的编码单元。

[0107] P_1 中解个体在通过自监督构成的M维表示空间中进行进化操作，包括交叉和变异，从而进化学习到它们新的M维空间的表示。

[0108] 将 P_1 中解个体在M维表示空间中新学习到降维解依次通过相对应的M个自建编码器的解码单元解码回原来的n维变量空间，从而有效的更新 P_1 中解个体，得到第一更新子种群。

[0109] 通过在降维的M维表示空间根据自监督式进化学习来更新 P_1 中的解个体，可以有效的提升进化搜索的效率，并且能加快整个种群的收敛速度。

[0110] 对子种群 P_2 中的每个解个体按照基于广泛学习的方法进行更新。

[0111] 因为 P_1 中的解个体对于优化KKM和RC这两个目标整体表现出更好的收敛性和多样性，所以，对于 P_2 中的每个解个体，从子种群 P_1 中随机挑选M个解个体进行进化操作。

[0112] 对于 P_2 中的每个解个体 $x \in P_2$ ，分别挑选得到的M个解个体 $\{y_1, y_2, \dots, y_M\}$ ，在变量分组 g_1, g_2, \dots, g_M ，进行交叉和变异。具体过程是： x 与 y_1 进行进化操作时只限制改变 x 中属于 g_1 的变量， x 与 y_2 进行进化操作时只限制改变 x 中属于 g_2 的变量，以此类推。这样 x 的一次更新中分别与来自 P_1 中的M个解个体进行了进化学习，即广度学习，从而得到第二更新子种群。

[0113] 步骤S400、根据所述原始种群、所述第一更新子种群以及所述第二更新子种群，确定所述目标网络社区对应的进化种群。

[0114] 具体地,得到第一更新子种群和第二更新子种群后,根据所述原始种群、所述第一更新子种群以及所述第二更新子种群,确定所述目标网络社区对应的进化种群。

[0115] 步骤S400具体包括

[0116] 步骤S410、根据所述原始种群中各解个体各自分别对应的目标函数值、所述第一更新子种群中各解个体各自分别对应的目标函数值以及所述第二更新子种群中各解个体各自分别对应的目标函数值,确定所述目标网络社区对应的进化种群。

[0117] 具体地,合并原始种群P以及新的两个子种群 P_1, P_2 中的所有解个体,然后进行环境选择,环境选择是在根据合并种群中的解个体对于优化KKM和RC这两个目标整体表现情况,择优选取其中多样性和收敛性表现好的N个解作为下一代的进化种群P'。

[0118] 例如,采取基于参考方向引导的聚类方法,先将以上合并种群分成两个子种群,然后保留其中表现较好(适应度值高)的种群作为下一代的进化种群P'。

[0119] 步骤S500、当所述进化种群满足终止条件时,将所述进化种群作为最优种群。

[0120] 具体地,终止条件可以是进化种群的迭代次数超过预设迭代次数,也可以是进化种群的平均适应度值大于预设适应度阈值。当进化种群达到终止条件时,则将进化种群作为最优种群,并输出最终进化种群P'中的帕累托最优解集,确定最优的社区结构检测。

[0121] 步骤S600、当所述进化种群不满足终止条件时,将所述进化种群作为原始种群,并继续执行将所述原始种群分解成第一子种群和第二子种群的步骤,直至所述进化种群满足终止条件时,将所述进化种群作为最优种群。

[0122] 具体地,当进化种群不满足终止条件时,则要将进化种群作为原始种群进行迭代,具体返回到步骤S200,直至进化种群满足终止条件,将所述进化种群作为最优种群。

[0123] 综上,本发明公开了一种基于自监督学习式进化的大规模复杂网络社区检测方法,主要解决现有技术中存在的以下问题:(1)现有基于聚类的社区检测方法大多需要一定的先验知识来预先指明目标网络中社区的数量,而对于大规模复杂网络社区检测,网络中社区的数量是很难预先确定的;(2)现有的基于进化算法的社区检测方法只能处理小规模的网络社区检测问题,对于大规模复杂网络社区的检测,传统进化算法的搜索效率低且收敛速度很慢,因此很难实现的对大规模复杂网络社区的快速高效检测。本发明的实现步骤为:(1)将目标大规模复杂网络抽象成一个无向图,初始化系统参数,设置迭代停止条件;(2)将大规模复杂网络社区结构检测问题建模成一个两目标的优化问题;(3)基于网络图中度值较大的节点作为候选社区中心节点的策略初始化一个原始种群P;(4)在目标空间采用一种基于参考方向引导的聚类方法将原始种群P分解为两个子种群: P_1, P_2 ;(5)将整个大规模复杂网络的变量空间随机均匀地分解为M个组;(6)对子种群 P_1 中的每个解个体按照基于自监督学习的方法进行进化更新;(7)对子种群 P_2 中的每个解个体按照基于广度学习的方法进行进化更新;(8)合并原始种群P和两个更新子种群 P_1, P_2 中的解个体,通过环境选择更新下一代进化种群;(9)判断终止条件是否满足;(10)输出最终的进化种群P中的帕累托最优解,并确定最优的社区结构检测。

[0124] 本发明将大规模复杂网络社区结构检测问题建模成一个两目标的优化问题,其中第一个目标是最大化同一社区内节点之间的链接密度,第二个目标是使得不同社区之间的链接密度最稀疏,然后通过基于自监督学习式进化方法同时最优化这两个目标,以实现目标大规模复杂网络的最优社区结构检测。

[0125] 本发明具有如下效果：

[0126] 1. 在基于分而治之的多目标进化搜索模式中结合自监督学习策略来提升对大规模复杂网络社区结构进行检测的性能。

[0127] 2. 通过自监督学习策略将原大规模复杂网络社区结构进行检测的问题转换为低维简单网络中的社区结构检测问题,从而可以提升计算效率以及搜索的收敛速度。

[0128] 3. 通过引入基于广度学习的策略,保证进化搜索过程中种群的多样性,从而提升检测结果的准确性。

[0129] 基于上述任意实施例所述的基于自监督学习式进化的大规模复杂网络社区检测方法,本发明还提供了一种计算机设备的较佳实施例:

[0130] 本实施例的计算机设备包括存储器和处理器,所述存储器存储有计算机程序,所述处理器执行所述计算机程序时实现如下步骤:

[0131] 根据目标网络社区,确定所述目标网络社区对应的原始种群;

[0132] 将所述原始种群分解成第一子种群和第二子种群;其中,所述第一子种群的平均适应度值大于所述第二子种群的平均适应度值;

[0133] 采用基于自监督学习的方法对所述第一子种群进行更新,得到第一更新子种群,并采用基于广度学习的方法对所述第二子种群进行更新,得到第二更新子种群;

[0134] 根据所述原始种群、所述第一更新子种群以及所述第二更新子种群,确定所述目标网络社区对应的进化种群;

[0135] 当所述进化种群满足终止条件时,将所述进化种群作为最优种群。

[0136] 基于上述任意实施例所述的基于自监督学习式进化的大规模复杂网络社区检测方法,本发明还提供了一种计算机可读存储介质的较佳实施例:

[0137] 本实施例的计算机可读存储介质,其上存储有计算机程序,所述计算机程序被处理器执行时实现如下步骤:

[0138] 根据目标网络社区,确定所述目标网络社区对应的原始种群;

[0139] 将所述原始种群分解成第一子种群和第二子种群;其中,所述第一子种群的平均适应度值大于所述第二子种群的平均适应度值;

[0140] 采用基于自监督学习的方法对所述第一子种群进行更新,得到第一更新子种群,并采用基于广度学习的方法对所述第二子种群进行更新,得到第二更新子种群;

[0141] 根据所述原始种群、所述第一更新子种群以及所述第二更新子种群,确定所述目标网络社区对应的进化种群;

[0142] 当所述进化种群满足终止条件时,将所述进化种群作为最优种群。

[0143] 应当理解的是,本发明的应用不限于上述的举例,对本领域普通技术人员来说,可以根据上述说明加以改进或变换,所有这些改进和变换都应属于本发明所附权利要求的保护范围。

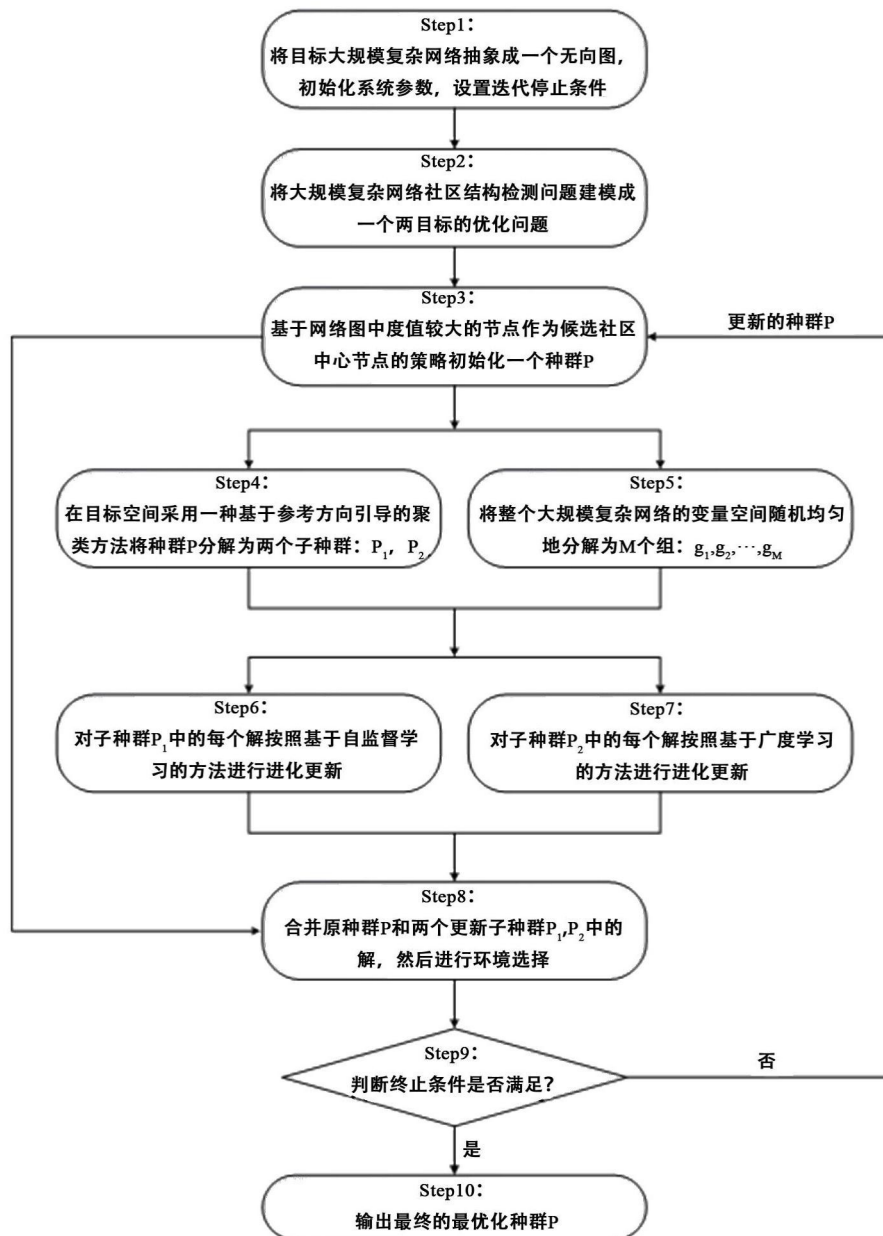


图1

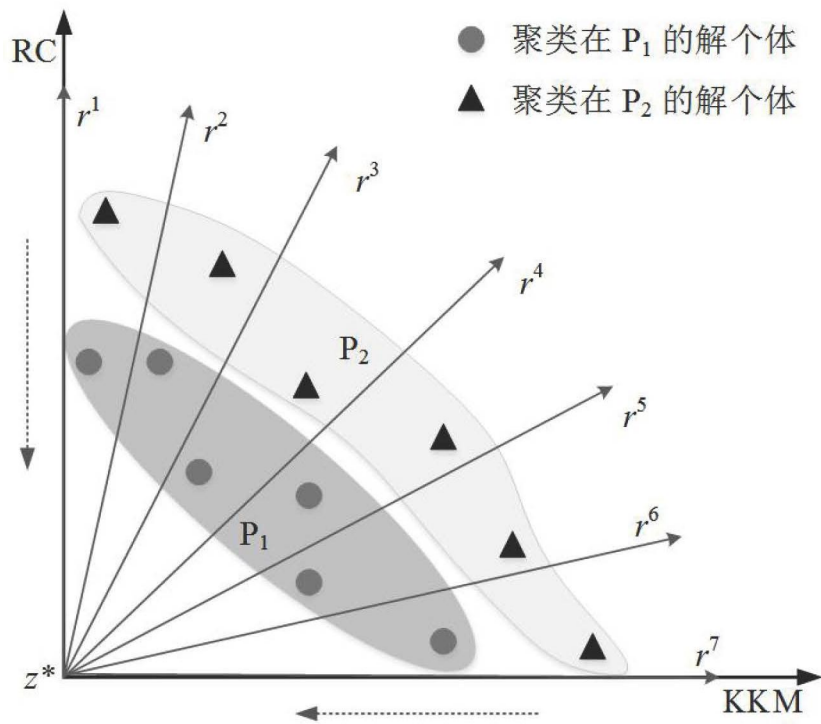


图2

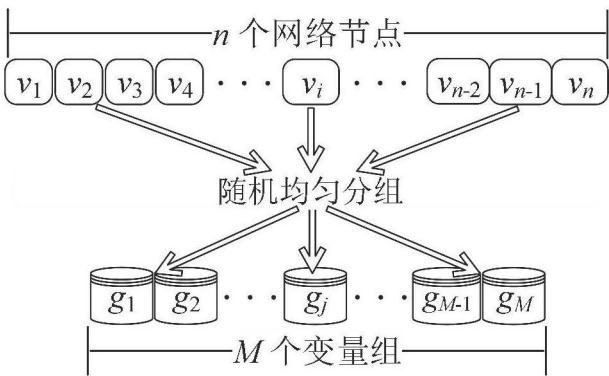


图3

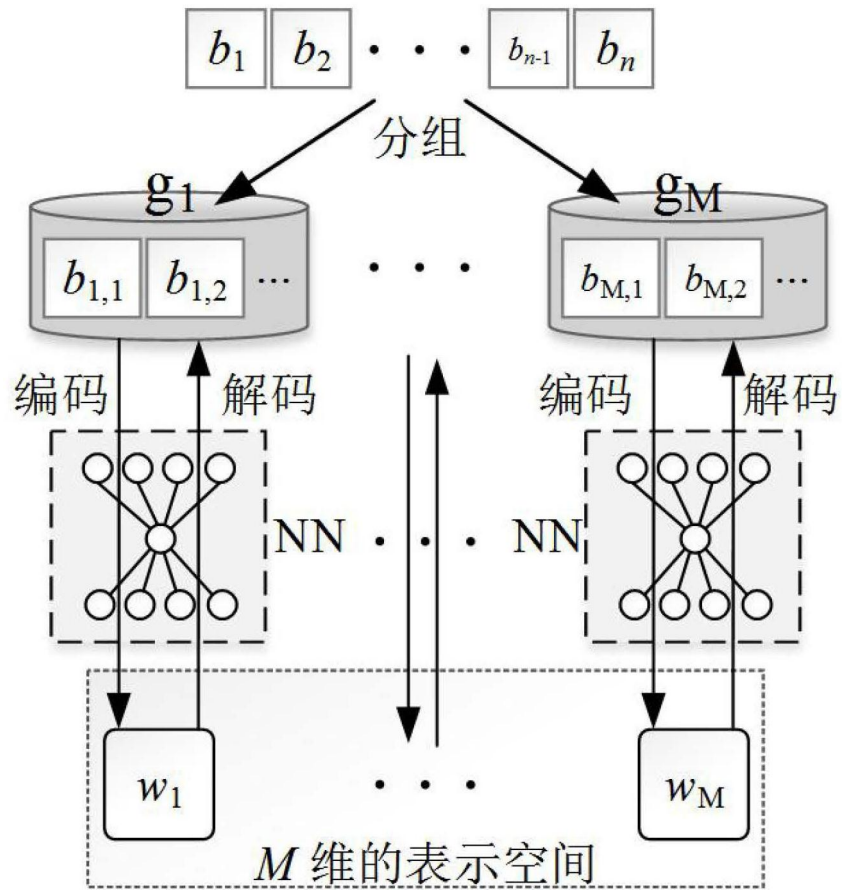


图4

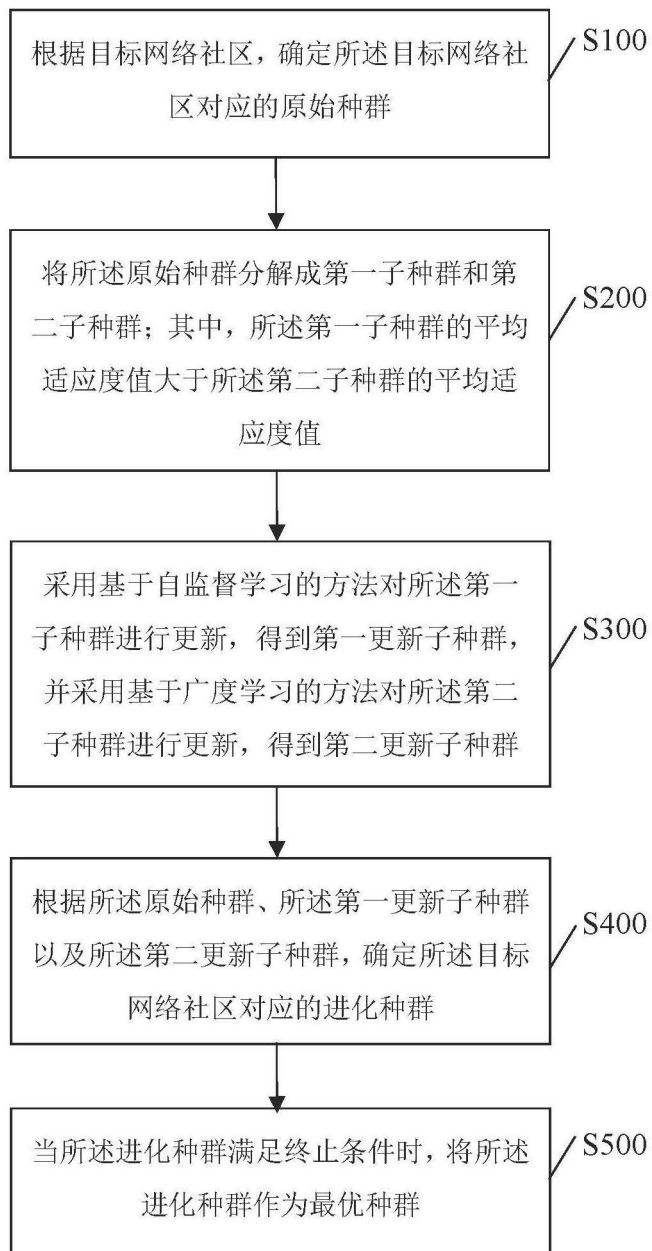


图5