目 录

[1. An improved gravity model to identify influential nodes in complex networks based on k-shell method 2](#_Toc106300382)

[1.1 基本信息 2](#_Toc106300383)

[1.2 论文内容 2](#_Toc106300384)

[1.3 实证分析 6](#_Toc106300385)

[1.4 结论 15](#_Toc106300386)

## 1. An improved gravity model to identify influential nodes in complex networks based on k-shell method

基于K壳法的改进的重力模型来识别复杂网络中具有影响力的节点

### 1.1 基本信息

期刊：《Knowledge-Based Systems》

发表时间： 2 June 2021

作者： Xuan Yang, Fuyuan Xiao

作者信息：

School of Computer and Information Science, Southwest University, Chongqing, 400715, China

分区：



### 1.2 论文内容

#### 一、摘要

寻找复杂网络中的重要节点是一个基本问题。最近提出了许多方法来解决这个问题，但以前的研究大多有局限性，其中很少考虑网络的局部和全局信息。节点的位置是网络中节点的一个重要属性，但以前在识别节点的重要性时很少被考虑。为了解决这个问题，我们在K壳算法的基础上提出了一种改进的引力中心度，名为KSGC，用于识别复杂网络中具有影响力的节点。我们的方法考虑了节点的位置，与原来的重力中心度相比，更加合理。在现实世界的网络中进行的一些实验表明，我们的方法可以有效地评估复杂网络中节点的重要性。

#### 二、引言

随着网络科学的快速发展，复杂网络在不同领域引起了广泛关注[1]，如社会[2,3]、生物、物理[4-6]、时间序列[7,8]、交通[9,10]、免疫策略[11,12]等等[1317]。世界充满了复杂性和不确定性，而现实世界的系统可以用复杂网络来建模[18,19]。复杂系统中的一些特殊节点可能决定整个网络的许多结构特性[20-22]。因此，研究人员更加关注识别重要的节点，这些节点可以准确地揭示复杂网络中的隐藏属性。

到目前为止，存在许多经典的方法来识别复杂网络中的重要节点，包括DC、BC、CC、EC。然而，这些经典的方法都有自己的缺点。DC只关注周围连接节点的数量，不考虑整个网络的整体结构[27]。BC和CC从全局角度对节点的重要性进行排序，但由于其计算复杂度高，在大规模复杂网络中不能很好地表现出来。至于特征向量中心性，其不能用于加权网络的特性降低了其应用。最近，人们提出了一些新的措施来提高识别节点的性能。例如，Wen和Jiang提出了一种基于模糊局部维度的方法来识别有影响力的节点[36]。Zhao等人同时考虑了一个节点的自我重要性和全局重要性[37]。Zareie等人使用灰狼优化（GWO）技术，这是一种生物启发的优化算法来寻找种子节点[38]。由于数据融合的良好能力，证据理论已经被用来处理这个问题[39]。

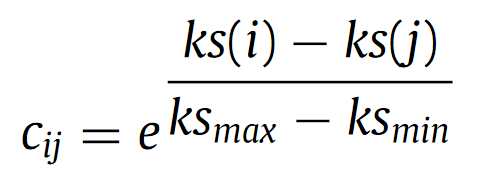
重力定律是计算两个物体之间相互作用的基本物理规则。两个物体之间的相互作用与两个物体的质量成正比，与两个物体之间距离的平方成反比。受引力公式的启发，Li等人[40]提出了引力中心性（GC）模型，该模型利用节点的度作为其质量，利用两个节点之间的最短路径距离作为其距离。而在引力中心性模型中，两个节点之间的互动只与它们的度数和距离有关，这表明两个节点之间的吸引力是一样的。在现实中，每个节点的吸引能力可能是不同的。Liu等人[41]通过考虑网络中节点的权重改进了这一模型，改进后的中心度量被命名为WGC，可以更适用于现实世界的网络。

鉴于上述讨论，在本文中，我们将节点的位置考虑在内。也就是说，位于网络中心的节点比位于网络边缘的节点更有能力吸引其他节点[42]。在k-shell算法的基础上提出了一个改进的引力模型[43]，以识别网络中具有影响力的节点。由k-shell值差异代表的节点之间的位置差异被用作**吸引系数**，该系数可以调整网络中中心节点的吸引力。本文提出的方法结合了本地信息和全球信息。这里进行了几个实验，最后的结果显示我们的方法表现良好。本文的其余部分组织如下。在第2节中简要介绍了一些初步的知识。在第3节中，我们提出了名为KSGC的方法来识别有影响力的节点。在第4节中应用了几个实验来说明所提出的方法的合理性。第5节给出了结论。

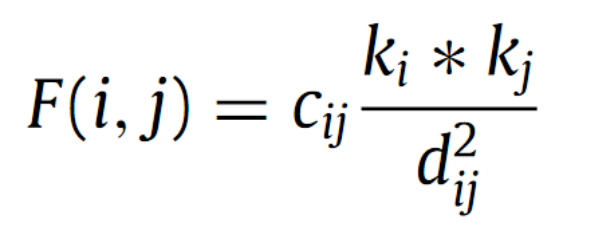
#### 三、模型解析

在一个现实世界的系统中，系统内核的元素比边缘的元素更稳定。一个主要原因是内部元素之间的连接密度比边缘高。我们可以用k-shell值来表示一个节点在网络中的位置。实际上，一个K壳值为n的节点会与网络中的其他n个节点形成一个n+1的集团（clique）。【n-clique】表示一个由n个节点组成的完整子图。它是一个稳定的局部结构，因为它的内部连接密度是最高的。显然，一个k-shell值较高的节点意味着它属于一个更稳定的局部结构，这表示它在遭受外部影响时，会从这个clique中的其他节点获得更多的帮助。换句话说，k-shell值较高的节点可能在一定程度上抵消了外部影响，可能有更多的能力去影响节点。在以前关于基于k-shell的方法的研究中，k-shell值只被用作节点本身的特征。而在本文中，上述分析促使我们考虑，在计算两个节点之间的相互作用时，有必要考虑到由k-shell差异所代表的位置差异。

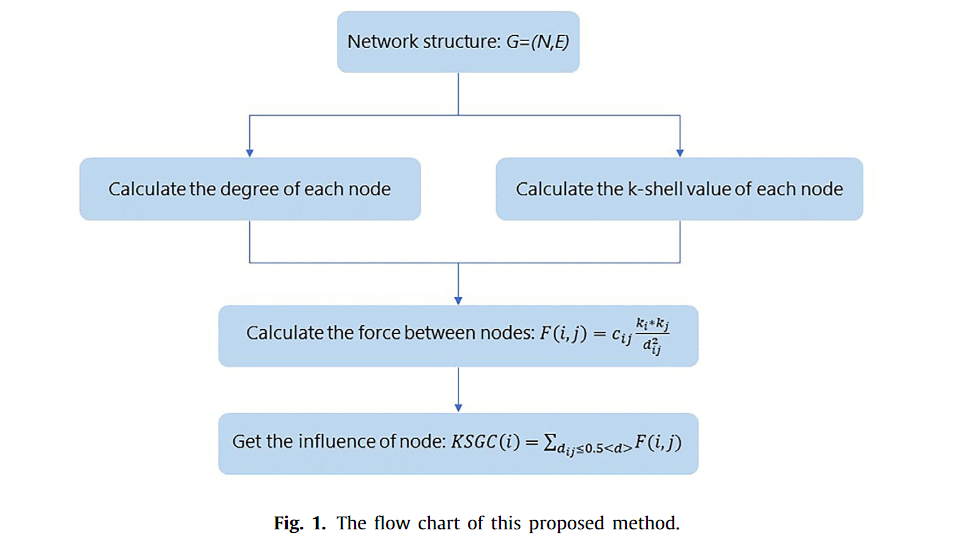
重力模型计算一对节点之间的相互作用来衡量一个节点的影响力。在这个模型中，一个节点的影响力包括两个方面。首先是一个节点的程度。换句话说，如果一个节点有更多的周围邻居，就有更多的节点可以直接受到影响，因此，该节点具有更高的影响力。第二是节点与其他节点之间的距离。一个节点对另一个节点的影响与它们的距离的平方成反比。他们的距离越小，他们的相互作用就越大。在本文中，我们假设一个节点影响另一个节点的能力与它们的位置有关。也就是说，位于网络中心部分的节点可能比位于外围的节点具有更大的吸引力。而我们用K壳值来代表一个节点的位置。节点i对节点j施加的吸引系数是cij：



其中ks(i)和ks(j)分别是节点i和节点j的k-shell值。而ksmax和ksmin是网络中最大和最小的k-shell值。所以节点i对节点j施加的力可以定义为F(i, j):



显然，如果节点i和节点j有不同的K壳值，F（i，j）不等于F（j，i）。而当ks(i)>ks(j)时，节点i对节点j施加的力高于重力模型（Cij必然大于1）；当ks(i)<ks(j)时，情况正好相反（Cij必然小于1），而当ks(i)=ks(j)时，我们的方法就退化为重力模型。而一个节点的重要性是该节点对网络中其他节点施加的力量之和。图1显示了所提方法的流程图。



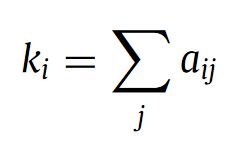
该方法的具体步骤如下：

**第1步：构建网络**

给定一个连接的无定向网络，我们可以用G=(N,E)来表示它。为了表示节点之间的连接关系，这里引入邻接矩阵A。aij=1表示节点i和节点j之间有连接，aij=0则相反。

**第2步：计算每个节点的度和不同节点对之间的最短距离**

如上所述，两个节点之间的力与它们的质量成正比，与它们距离的平方成反比。在我们的模型中，最短路径（dij）被认为是节点i和节点j之间的最短距离。一个节点的质量由其度数表示：



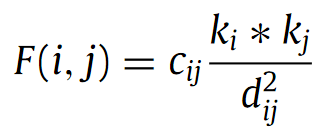
其中，ki是节点i的度数，aij是邻接矩阵中的元素，表示节点i和节点j之间的连接关系。

**第3步：计算每个节点的K壳值**

由于考虑到节点的位置，这里需要得到每个节点的K壳值。具体计算方法已在第2.2节中提到。

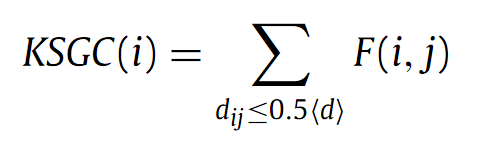
**第4步：计算一个节点对另一个节点施加的力**

两个节点之间的力由三部分组成。有两个节点的度数，两个节点之间的距离和它们的K壳差。公式（8）是计算节点i对节点j施加的力的公式。



**第5步：获得节点的影响**

最后，节点的影响由该节点对其他节点施加的力的总和表示。为了减少计算的复杂性，我们也继承了Li.提出的网络的截断半径。因此，本文提出的基于引力中心度（KSGC）的K壳定义如下：

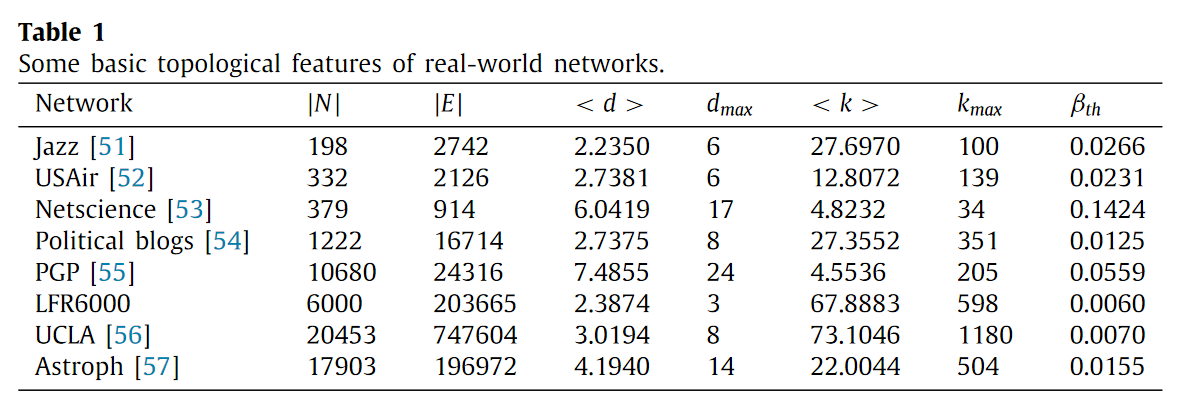


这个公式同时考虑了局部和全局信息。如果一个节点的度数较大，与其他节点的距离较短，此外，其K壳值也相对高于其周围的节点，那么这个节点的重要性就会更大。而随着截断半径的设定，一个节点的影响力不会超过网络直径的一半。总之，这项工作结合了引力模型和kshell算法来识别复杂网络中节点的重要性。本文的亮点是，我们提出了一个吸引系数来修改成对节点之间的相互作用。通过引入这个吸引系数，当网络中存在两个k-shell值不同的节点时，它们之间的吸引就会不同。一个K壳值较高的节点比K壳值较低的节点更有能力吸引其他节点。因此，这是一种更全面的方法，它从全局角度捕捉到了一对节点之间的差异。

### 1.3 实证分析

#### 一、数据来源

在本节中，我们使用了六个真实世界的网络和一个由LFR基准模型[58]生成的合成网络来测试所提模型的性能。而LFR6000网络是由LFR基准模型生成的，参数设置为节点数为6000，平均度数为60，最大度数为500，社群结构的混合参数为0.8。有关数据的详细信息可参见表1。

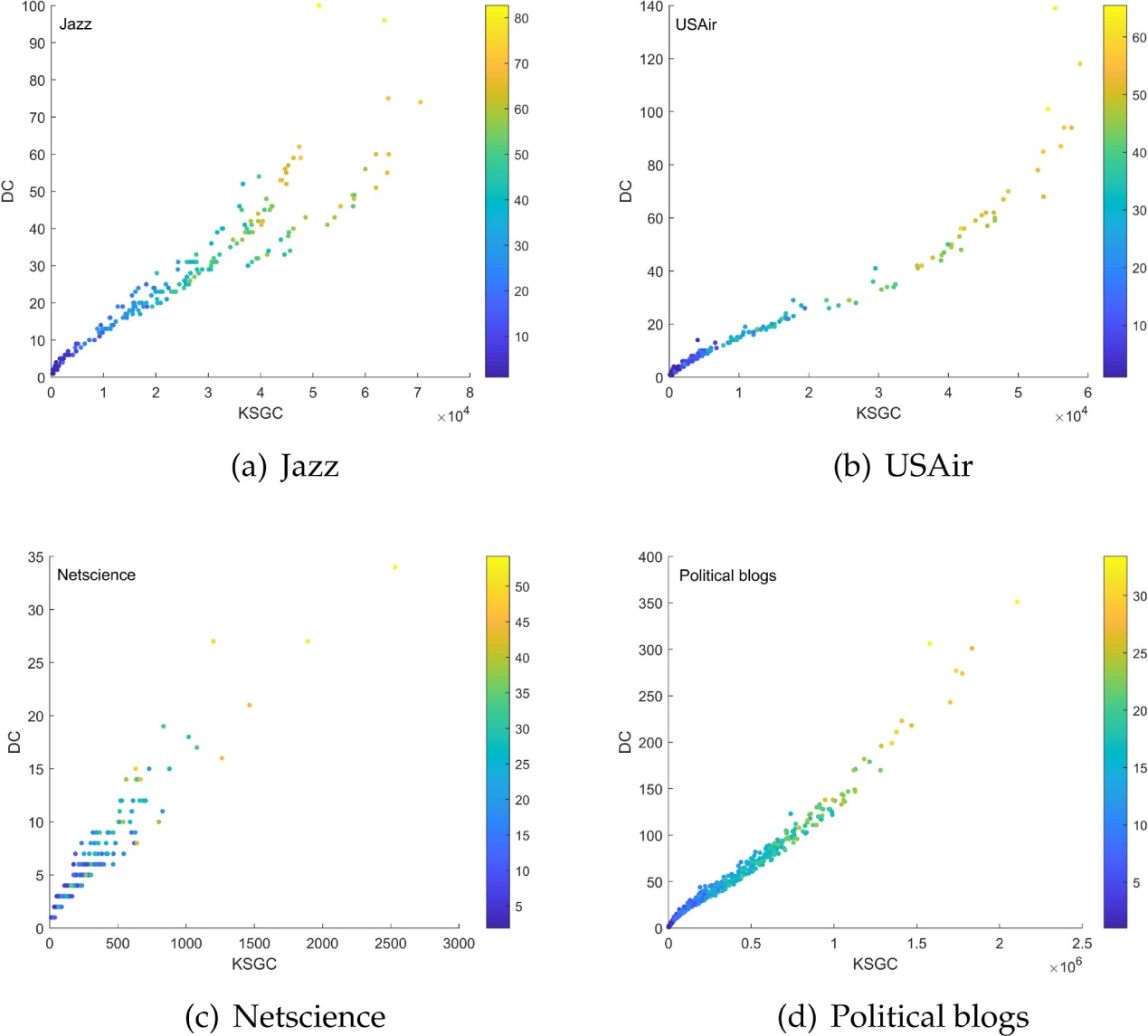


在实验阶段，为了验证我们提出的方法的正确性，我们用DC、BC、CC和EC来探索它们与KSGC的关系。为了测试我们方法的性能，我们使用了一些最先进的基于k-shell的方法，如混合度分解（MDD）、分层k-shell（HKS）、改进的k-shell混合（KSHI）、k-shell迭代因子（KS-IF）和基于潜在边缘权重的k-shell指数（WKS），以及包括GC和WGC的重力模型方法作为比较对象。

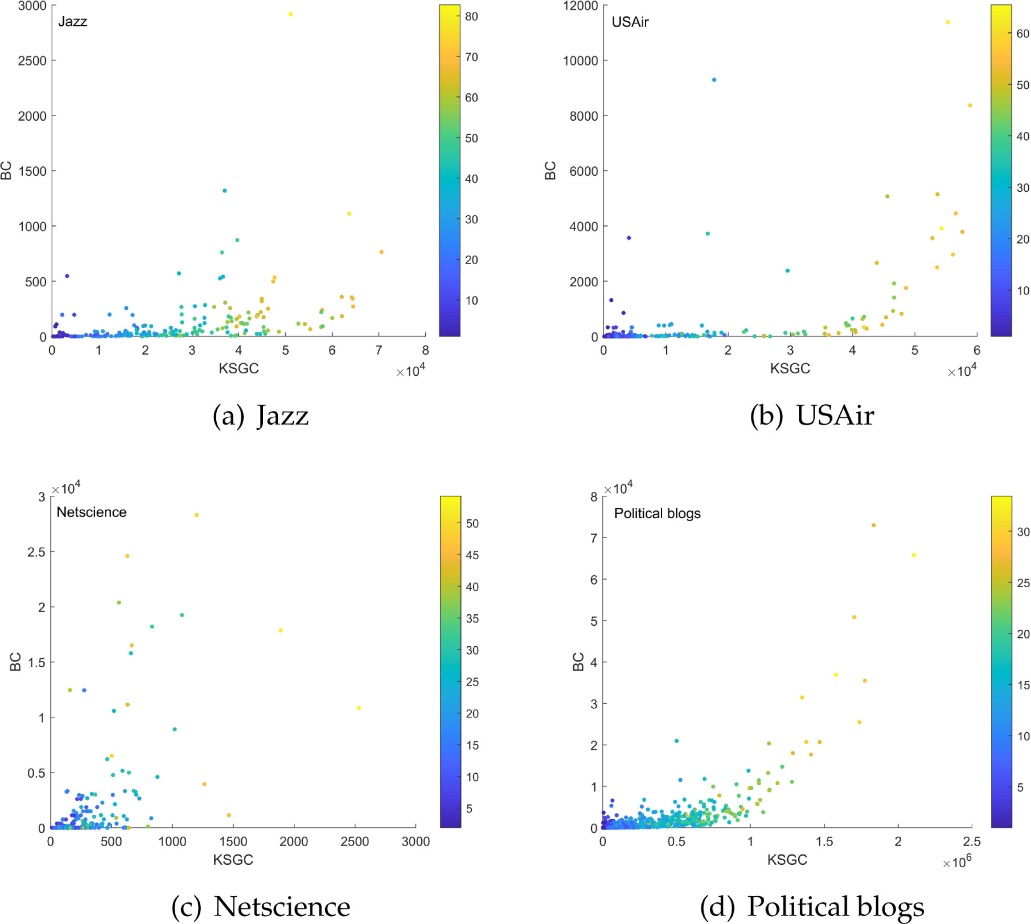
#### 二、KSGC与其他方法的关系

为了找到KSGC和其他方法之间的关系，这里介绍了几个真实世界网络中不同方法之间的关系图。选择DC、BC、CC和EC进行比较。实验结果在图2-5中给出。图中的每个点代表网络中的一个节点。图中各点的坐标表示用不同方法得到的数值，各点的颜色表示当β=βth时，不同网络中使用SI模型的节点的感染能力，由F(10)得到。

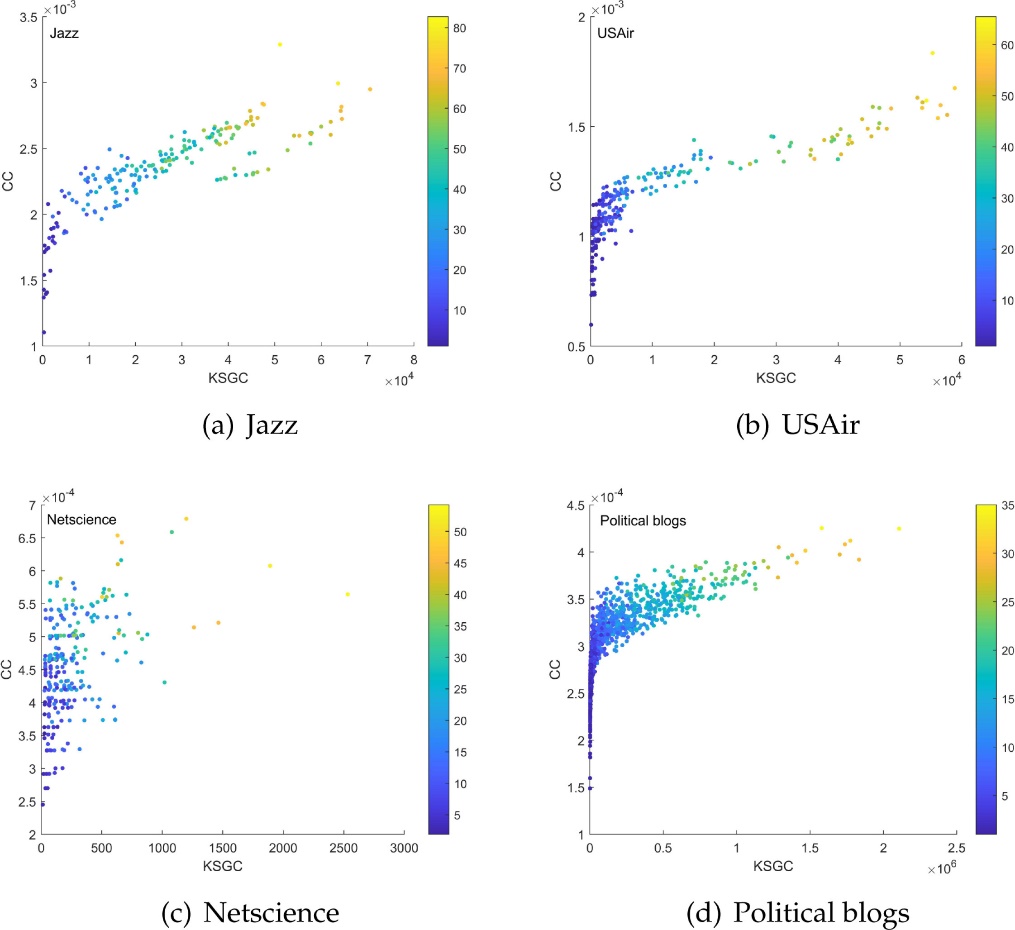
从图2可以看出，DC和KSGC有很强的正相关关系，这意味着DC大的节点的KSGC值也会很大。



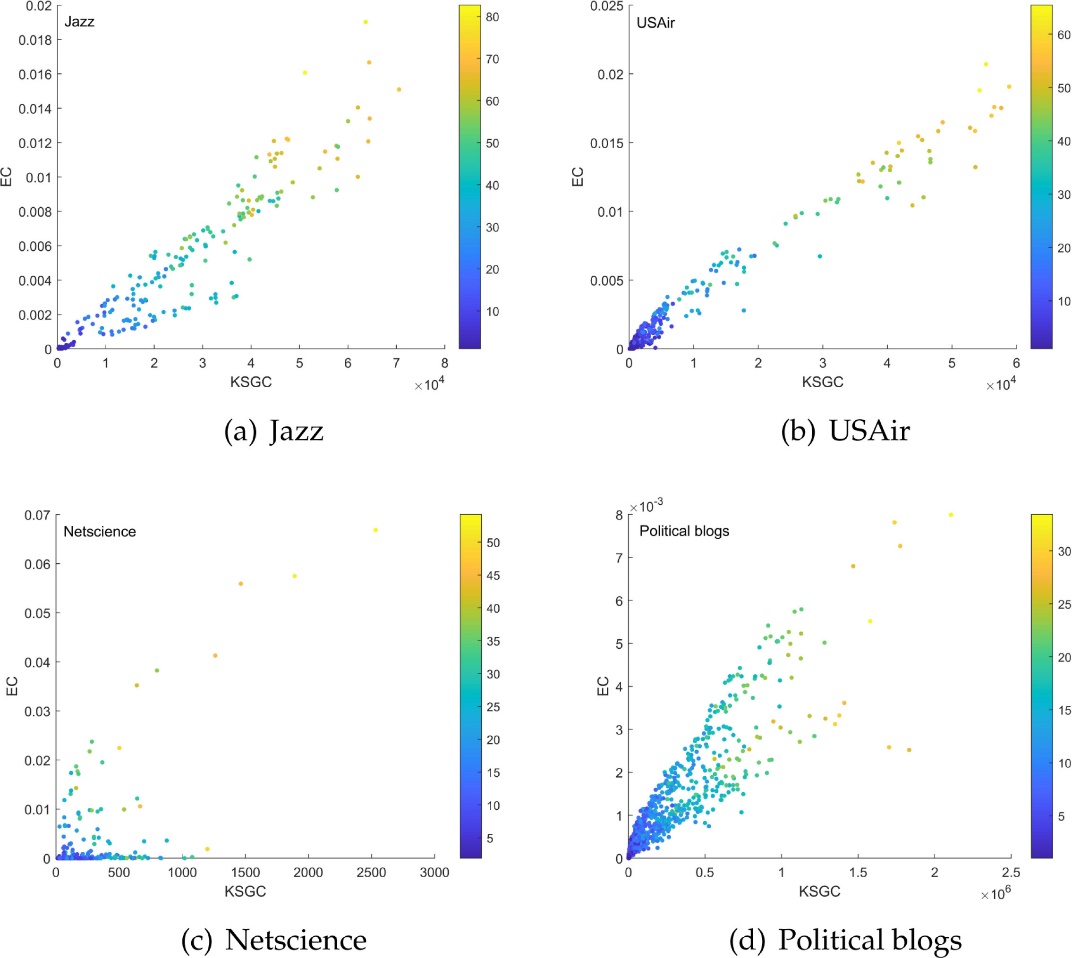
然而，正如图3中的结果，KSGC与BC几乎没有关系。



从图4中可以观察到KSGC和CC之间的关联性。可以发现，除了Netscience网络，KSGC和CC之间的关系是正的。但相关程度没有DC那么强。



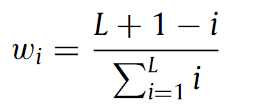
在图5中，和CC一样，KSGC和EC在Netscience网络中也没有什么关联性。而在另外三个网络中，EC与KSGC的正相关比CC更明显。



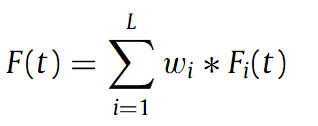
总的来说，除了BC之外，所提出的KSGC与一些经典的中心性方法有着不同程度的相关性。另一个值得注意的关键点是，KSGC值高的节点有更高的感染能力，这将证明KSGC的合理性。

#### 三、传播能力的对比

重要性越大的节点在网络中的传播能力就越强。为了比较不同方法得到的高排名节点的传播能力，采用SI模型，将不同方法中的高排名节点作为初始感染节点。我们用F（t）来表示网络中特定时间t的感染节点数量。F（t）随着感染过程t的增加而增加，最终在网络中大部分节点被感染时达到一个稳定值。我们在每种方法中选择前5%的节点作为初始感染节点。此外，我们给不同排名的节点分配不同的权重。排名指数为i的节点的权重可以定义为。

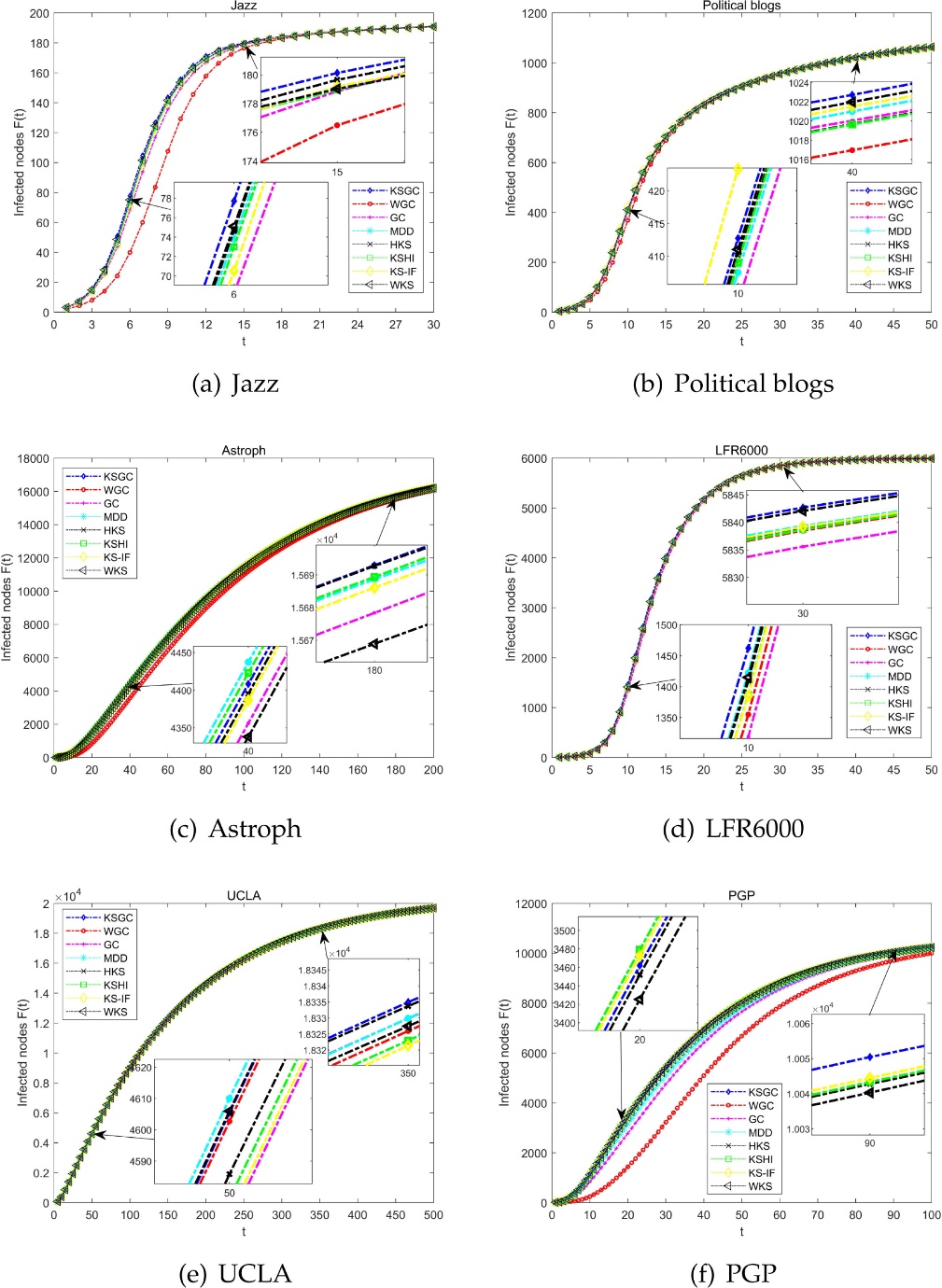


其中L是我们选择的初始感染节点的数量。可以很容易地发现，一个排名较高的节点（排名指数小）会有较高的权重。而F(t)可以得到：



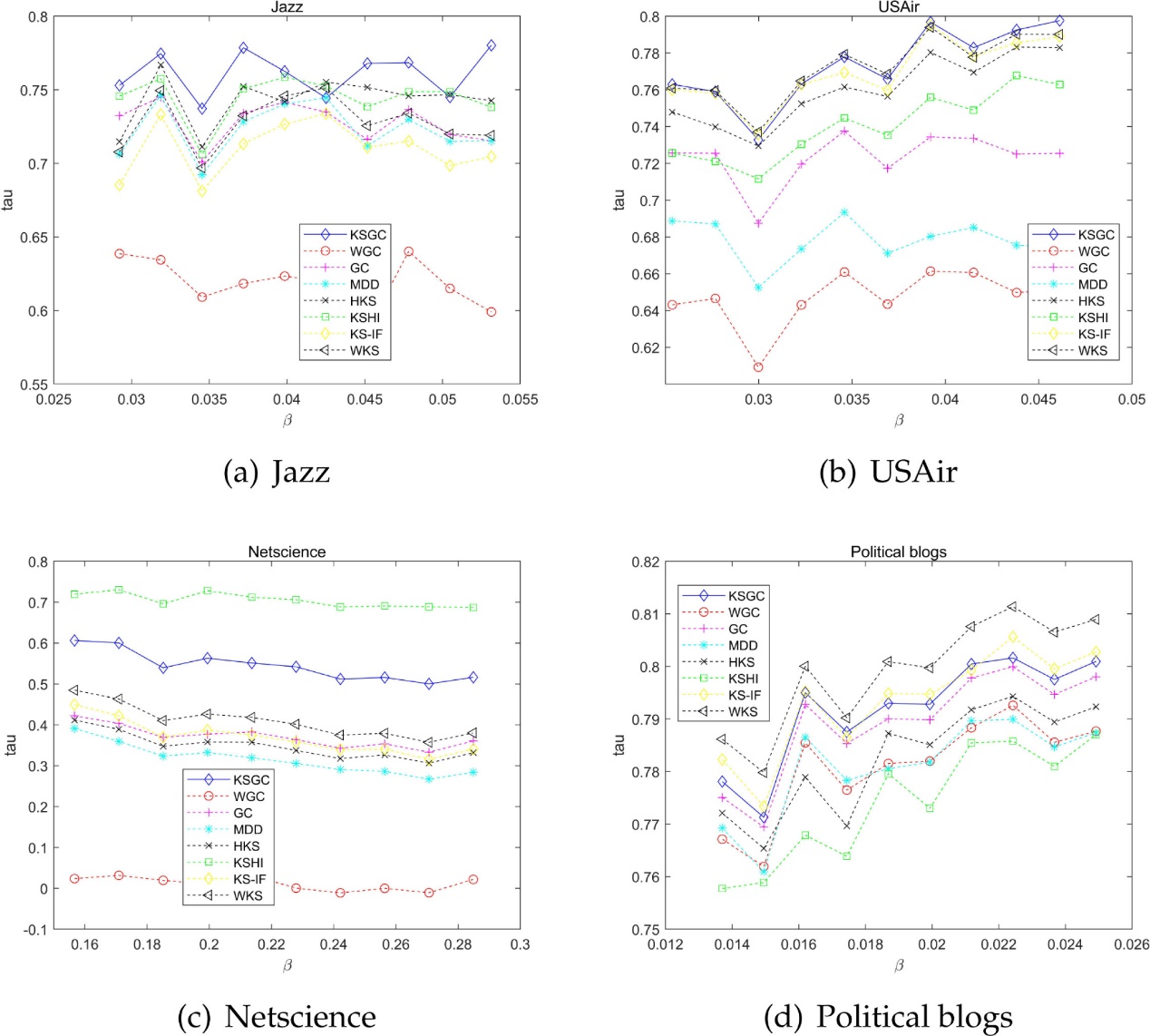
其中Fi(t)是由排名指数为i的初始感染节点在时间t引起的感染节点数量。每个实验在相应的网络中进行了100次，β=βth。最后的F（t）是表示感染能力的平均结果。

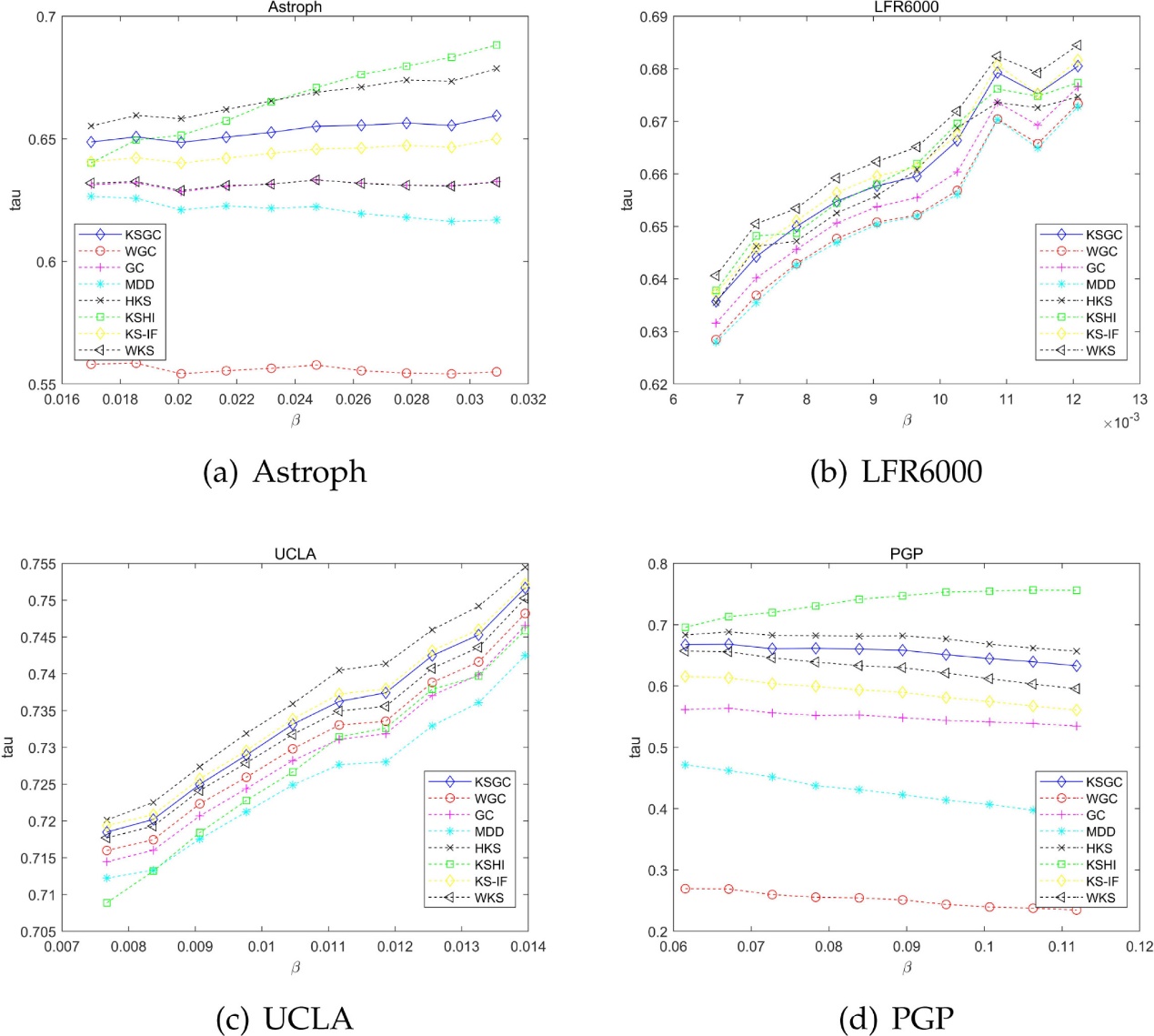
图6描述了不同方法的传播能力的比较。KSGC和其他方法得到的F（t）的上升趋势在所有网络中基本相同。在Jazz和LFR6000网络中，KSGC几乎在整个过程中比其他方法表现得更好。在一些网络中，如政治博客、UCLA和PGP，KSGC在早期感染过程中不是最好的，但其增长速度比其他方法快。而在感染过程的后期，KSGC的F（t）是最高的。在Astroph网络中，KSGC和WKS的增长几乎同时达到稳定状态，并且比其他方法表现更好。综上所述，图6中的大部分结果表明，本发明的方法具有相对优越的性能，这可以证明所提出的KSGC选择的排名靠前的节点在网络中具有更强的传播能力。



#### 四、所提方法的排名准确性

在本节中，我们研究了不同算法的排名准确性。我们使用SI模拟，根据受感染节点的数量在10个步骤中产生排名。而传播率β的值从βth到2 ∗ βth变化，每一步增加10%。最后的排名结果将由100个独立实验的平均值得到。然后，我们用Kendall's tau系数来衡量不同算法的排名表和SI模拟生成的排名表之间的相关性。图7和图8显示了β的变化对8个网络中不同排名算法的Kendall's tau系数（τ ）的影响结果。可以看出，与其他方法相比，这个提议的方法在总体上表现良好。例如，在爵士和美国航空网络中，KSGC的Kendall's tau系数最高。而在其他网络中，即使KSGC不是最好的，但其数值仍然处于相对较高的位置。我们可以发现，KSHI在Netscience和PGP网络中具有最高的Kendall's tau系数，但在政治博客和UCLA网络中表现很差。另一点需要注意的是，在所有的网络中，建议的KSGC比WGC和GC表现得更好，这可以证明我们的改进是有效的。





为了更直观地展示结果，表2中列出了平均Kendall's tau系数值。我们可以得到与上述相同的结论。提出的KSGC和KSHI比其他方法表现更好。



### 1.4 结论

本文提出了一种名为KSGC的新方法来识别网络中具有影响力的节点。这种新方法通过考虑节点的位置来改进现有的引力中心性。它结合了全局信息，如位置、与其他节点的距离和局部信息，如度信息。位于网络中心的节点具有较高的度，与其他节点的距离较短，对其他节点的吸引力较大。最后，节点的重要性可以通过一个节点对网络中其他节点的吸引力之和来衡量。这里对真实世界和合成网络进行了一些实验，以证明所提出的方法的合理性和有效性。

位置是一个节点在识别其重要性方面的一个基本属性。由于经典的k-shell分解方法扩展到了加权网络[65,66]，未来的工作是如何将该方法应用于加权网络。