目 录

[1. GMM: A Generalized Mechanics Model for Identifying the Importance of Nodes in Complex Networks 2](#_Toc104120221)

[1.1 基本信息 2](#_Toc104120222)

[1.2 论文内容 2](#_Toc104120223)

[1.3 实证分析 7](#_Toc104120224)

[1.4 结论 27](#_Toc104120225)

## 1. GMM: A Generalized Mechanics Model for Identifying the Importance of Nodes in Complex Networks

GMM：用于识别复杂网络中节点重要性的广义力学模型

### 1.1 基本信息

期刊：《Knowledge-Based Systems》

发表时间：29 December 2019

作者：Fan Liu, Zhen Wang, Yong Deng

作者信息：

电子科技大学，基础与前沿科学研究院、英才学院和数学科学学院

西北工业大学机械工程学院，光学图像分析与学习中心

电子科技大学，基础与前沿科学研究院

分区：



### 1.2 论文内容

**一、摘要**

本文提出了一个使用全局信息和局部信息的改进的重力模型。为了验证该方法的有效性，在总共十个真实网络上进行了一些实验。特别是，提出了一个创新的基于网络的实验质量评估，以验证识别节点重要性的方法。

文章结构：

背景概念与定义介绍

WGravity模型构建

节点排序测试

有效性测试（网络质量Q和攻击TOP节点方法）

**二、引言**

Fei等人提出了一种基于反平方律的衡量节点影响力的方法，即如果一个节点的度数越大，离其他节点越近，该节点的影响力就越大。然而，在一些应用中，有必要考虑节点的整体影响力。Gao等人提出了一个非Markov信息传播模型。在网络中添加边也是优化信息传播模型的一个重要途径。此外，一些基于节点相互作用的重力模型对于大型网络来说是很耗时的，Li等人通过引入截断半径解决了这个问题。

总的来说，如何更好地利用局部和全局信息来评估节点的影响仍然是一个问题。例如，Wen和Deng提出了一种基于局部维度的方法来识别重要节点。为了更好地解决上述问题，有必要提出一个更加普遍化的模型。

受万有引力定律的启发，本文提出了用于识别复杂网络中节点重要性的**广义力学模型（GMM）**，该模型考虑了局部和全局信息。如果一个节点的度数越大（局部信息），与其他节点的距离越短，权重越大（全局信息），该节点的影响力就越大。此外，GMM继承了Li等人提出的重力模型的截断半径，以处理大规模网络的耗时问题。

本文的贡献总结如下。

1、 提出了一个用于评估节点影响的广义力学模型。基于这个模型，一个加权重力模型（WGravity）被应用于节点重要性的评估。WGravity模型结合全局和局部信息来评估一个节点的影响。

2、 提出了一种验证节点重要性评估的新方法。基于复杂网络的连接性，提出了复杂网络质量。然后，通过攻击网络中的节点，检验了网络的最终复杂网络的质量。

**三、模型解析**

**（一）初步了解**

1. **重力中心性模型 Gravity centrality**

重力模型方法是将引力的形式引入复杂网络，用来表征两个节点之间的相互作用。对其他所有节点的作用越大，节点就越重要。重力中心性的定义如下：

其中，ki是节点i的度数，dji是节点j和节点i之间的最短距离，CG(j)是节点j的中心性分数。

1. **特征向量中心性（EC）**

EC为网络中的所有节点分配相对分数。EC是一种复杂的中心测量方法。一个节点的连接有很少的节点。如果它的EC值很高，它也很重要。给定一个n×n矩阵A，xj是归一化最大特征向量的第j个元素的值。特征向量中心度（EC）的定义如下：

其中λ是A的最大特征值，aji代表节点j和节点i之间的联系，是节点j的中心性分数。

**（二）WGravity模型**

1. **算法设计**

**步骤1：构建网络**

给定一个真实的网络，构建图G=（V，E）来表示该网络，其中V代表节点，E代表边。

**步骤2：计算节点的度**

在力学模型中，两个物体的相互作用与质量成正比，与距离的平方成反比。在提议的方法中，节点的度被比作现实世界中物体的质量。

其中是节点i的度数，是节点i和节点j之间的连接。

**步骤3：计算最大的归一化特征向量的值**

在现实世界中，每个物体的影响力是不同的。物体的影响力越大，与其他物体的相互作用力也越大。衡量每个物体的影响力也成为一个关键问题。受EC方法启发，采用最大的特征向量的值作为每个节点的权重。

其中和是邻接矩阵A最大的特征值和归一化的特征向量，是特征向量的第i个值。

解释：

在以前的研究中，重力模型没有考虑到每个节点的权重，也就是说，节点间互动能力都一样。实际中，每个节点的能力不同，有必要考虑节点的权重。受EC的启发，从全局角度出发，通过给每个节点分配一个相对分数来区分不同节点的权重。当特征向量中的所有元素都被要求为非负值时，只有最大的特征值才能得到满足。

（Perron-Frobenius定理：设A为非负不可约矩阵，矩阵的谱半径定义为矩阵A所有特征值的绝对值的最大值。则有：

1. 且是矩阵A的一个单特征值
2. A有一个对应于的正特征向量
3. A的每个非负特征向量都对应于特征值）

**步骤4：计算节点的相互作用力**

在现实世界中，两个物体之间的互动往往与质量成正比，与距离的平方成反比。在这个模型中，节点的度被当作质量，节点之间的最短距离被当作节点之间的距离。WGravity中心度的定义如下：

其中是节点i和节点j之间的最短距离，R是每个节点的影响半径（截断半径），影响半径设定为（假设每个节点的影响不超过网络半径的一半），即所有和节点i的最短距离小于等于R的点。

**算法总结：**

综上所述，GMM的整个步骤如图1所示：

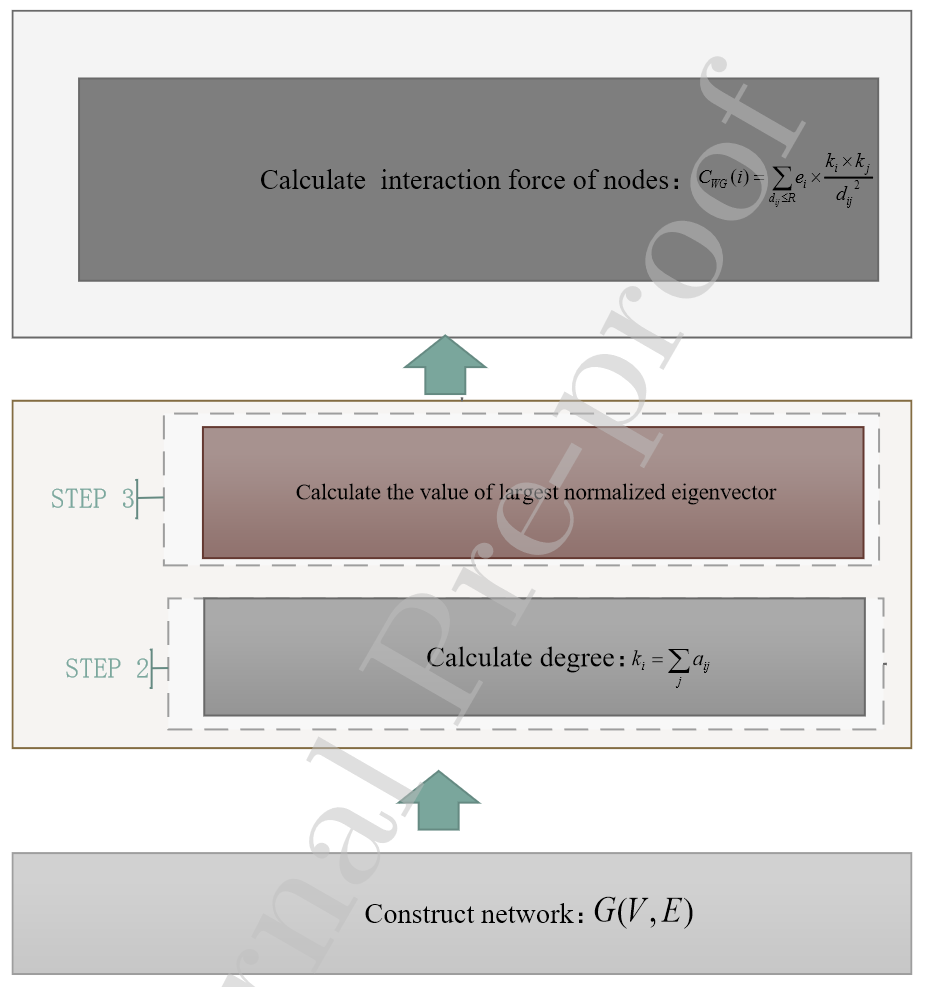


图 1 第一步是建立一个网络。第二步计算每个节点的节点度，作为每个节点互动的质量。第三步，给每个节点分配一个权重。第四步计算每个节点互动的中心性分数。

**2、算法时间复杂度分析**

考虑一个有n个节点和m条边的网络G，邻接矩阵A。本文提出的方法的时间复杂度总共分为三个项：节点的最短路径为、邻接矩阵A对应的最大特征向量e为、计算节点的度为。所以它的总时间复杂度是：

其中是任意节点半径R内的最大节点数。

### 1.3 实证分析

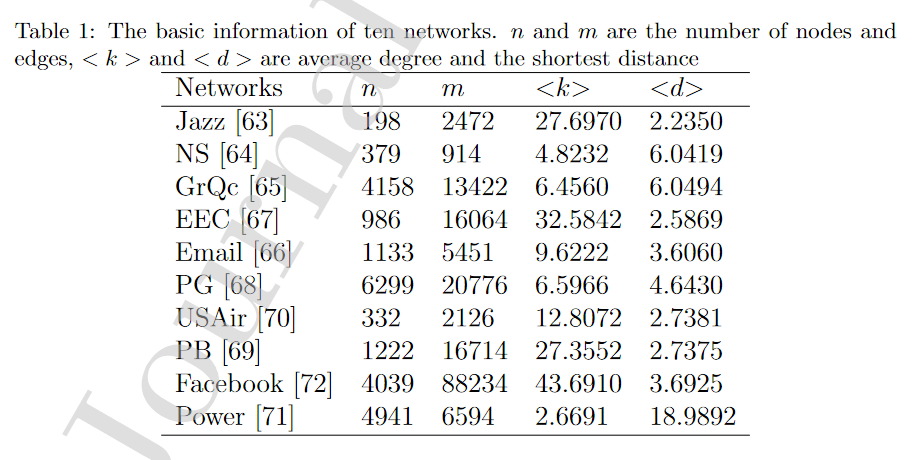
为了测试GMM模型的有效性，在十个真实网络中进行了五个不同的实验，并与六个不同的方法进行了比较。

1. **数据来源**

本文总共在现实世界中的10个数据集上进行了实验。分别是：

1. Jazz，一个爵士乐手的网络。[Jazz musicians (konect.cc)](http://konect.cc/networks/arenas-jazz/)
2. NS，一个由科学家合作的网络。[Network science (konect.cc)](http://konect.cc/networks/dimacs10-netscience/)
3. GrQc，一个发表在预印本的网络。[GrQc Collaboration Networks](https://networkrepository.com/ca-GrQc.php)
4. Email，描述了成员之间相互发送电子邮件的网络。
5. EEC，描述了一个欧洲研究成员交换邮件的网络。
6. PG，描述了一个文件共享的网络。[Pretty Good Privacy (konect.cc)](http://konect.cc/networks/arenas-pgp/)
7. PB，一个博客网络。[Political blogs (konect.cc)](http://konect.cc/networks/dimacs10-polblogs/)
8. USAir，美国运输航空网络。[USA airline](http://vlado.fmf.uni-lj.si/pub/networks/data/mix/USAir97.net)
9. Facebook，社交用户网络[SNAP: Network datasets: Social circles](http://snap.stanford.edu/data/ego-Facebook.html)
10. Power，美国电力网络。[US power grid (konect.cc)](http://konect.cc/networks/opsahl-powergrid/)

有关数据的具体信息如表1所示，n和m是节点和边的数量，<k>和<d>是平均度和最短距离。。



1. **节点的中心性分数计算**

在这个实验中，WGravity模型被用来计算节点的中心性分数。为了比较，重力模型、DC、CC、BC、PC和EC也被应用于同一数据集。实验结果显示在图2至图11中。在热图中，节点的颜色越深，说明该节点的重要性越大。

首先，计算每种方法，得出每个节点的中心性得分。然后，将所有的分数归一化。最后，通过热图对实验结果进行比较。从图2到11可以看出，所有方法得到的节点的相对重要性的分布是一致的。CC方法计算节点重要性的数值过大。WGravity模型和重力模型节点的分布非常接近。重力模型和WGravity模型的表现在判断每个节点的重要性上相对一致，但WGravity模型在全局上有明显的不同。

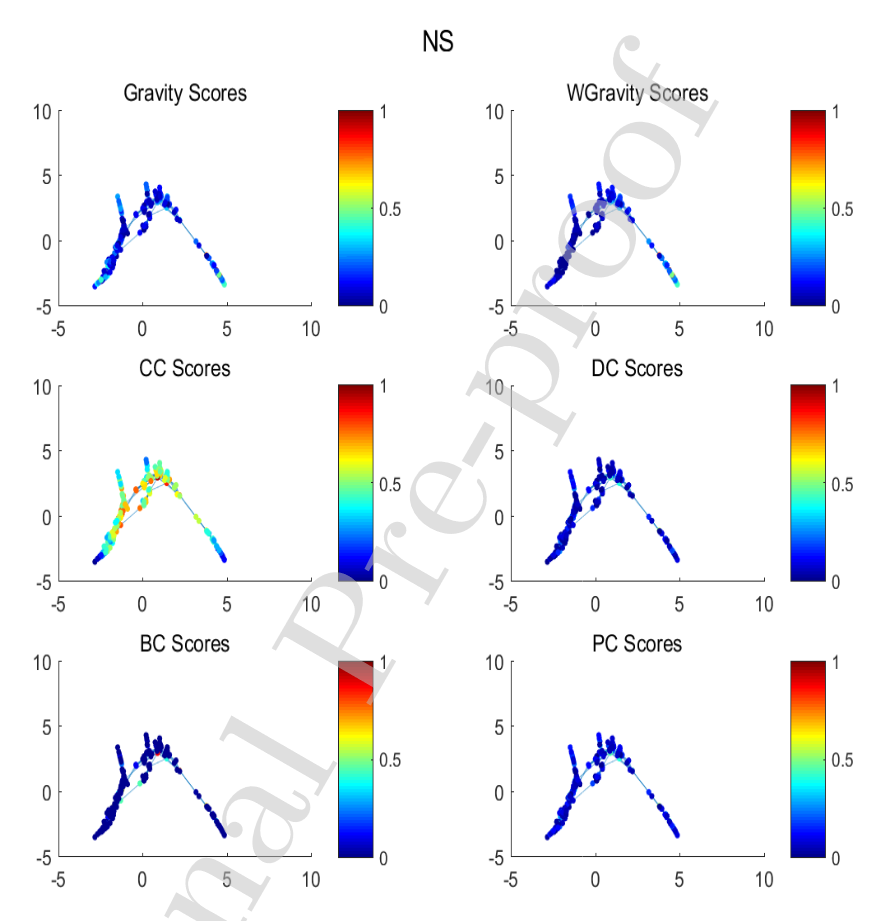
1. **Jazz**

在Jazz网络中，重力模型和GMM模型、DC的表现基本相同，CC中每个节点的数值都很高。根据得出的WGravity分数，GMM模型认为，Jazz网络中右边网络的节点比左边的节点更重要。



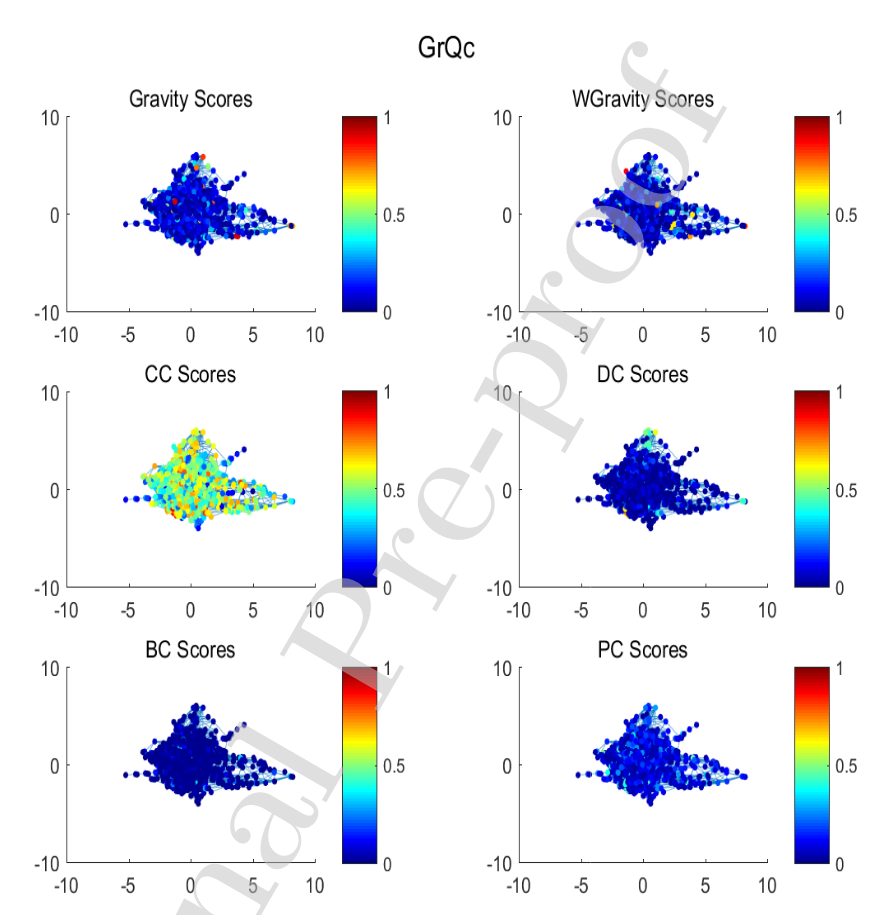
1. **NS**

在NS网络的实验中，PC、BC、WGravity和Gravity很难区分节点重要性的差异，但CC在全球范围内给出了一个显著的差异。



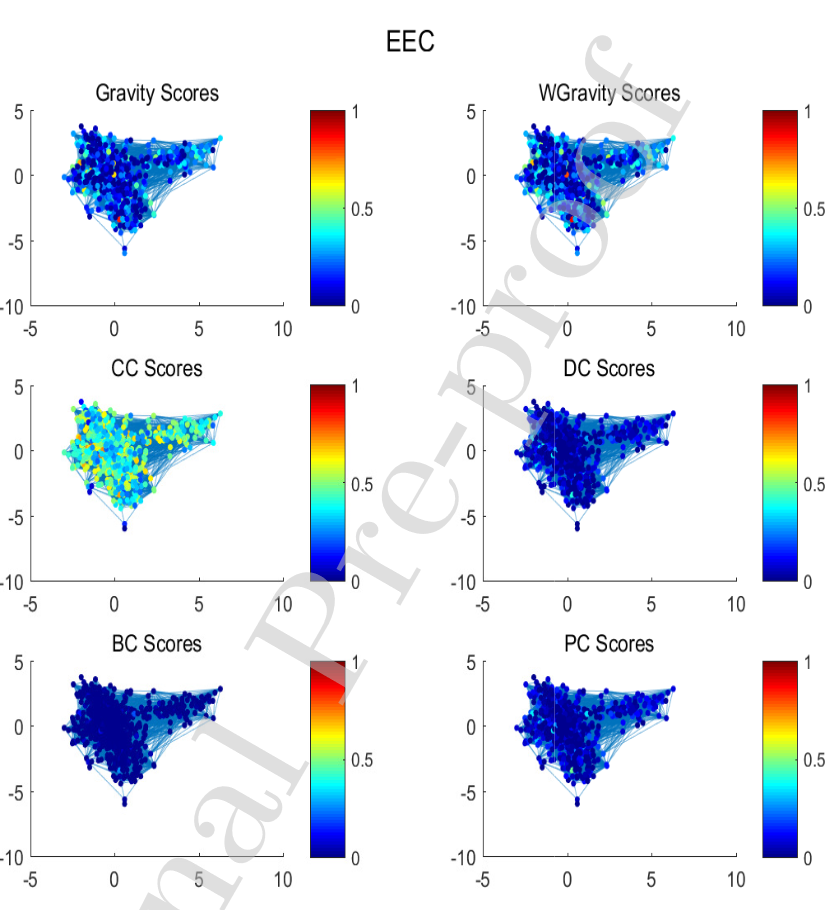
1. **GrQc**

在GrQc网络中，WGravity和Gravity之间有轻微的差别，但每个节点的重要性顺序在全球范围内没有太大差别。



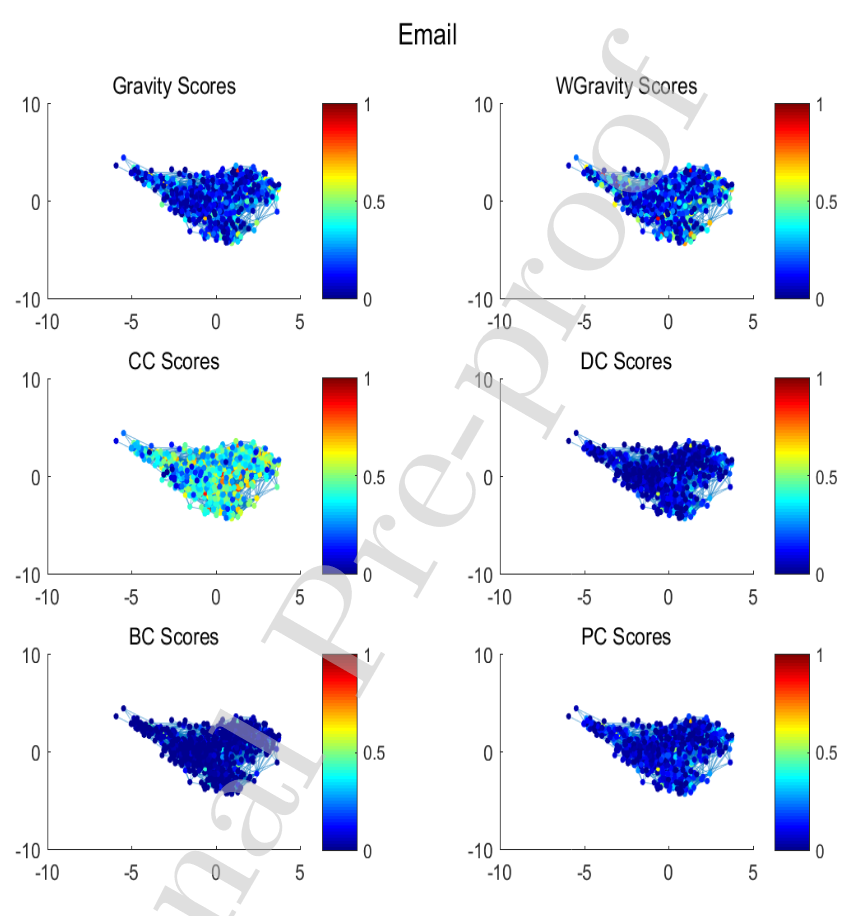
1. **EEC**

在EEC网络中，BC、PC和DC很难区分每个节点的差异，但WGravity和Gravity可以更好地判断节点的重要性。



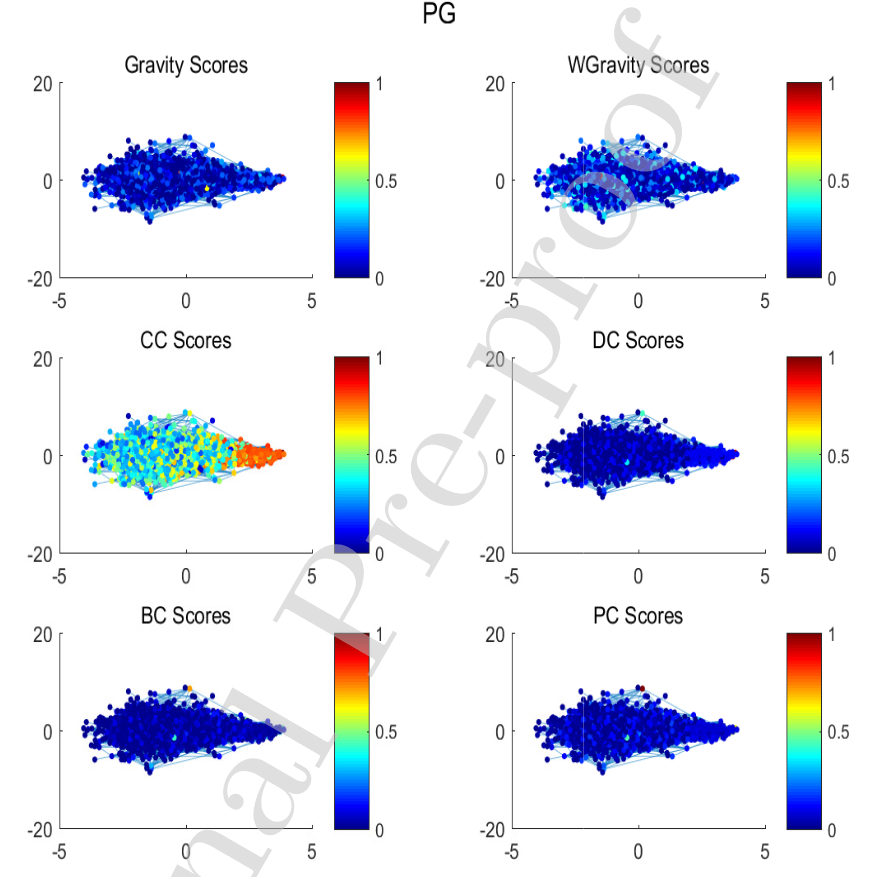
1. **Email**

在Email网络中，CC对节点的重要性赋予了较高的数值，但可以区分每个节点的差异。然而，BC和DC则很难区分节点之间的差异。



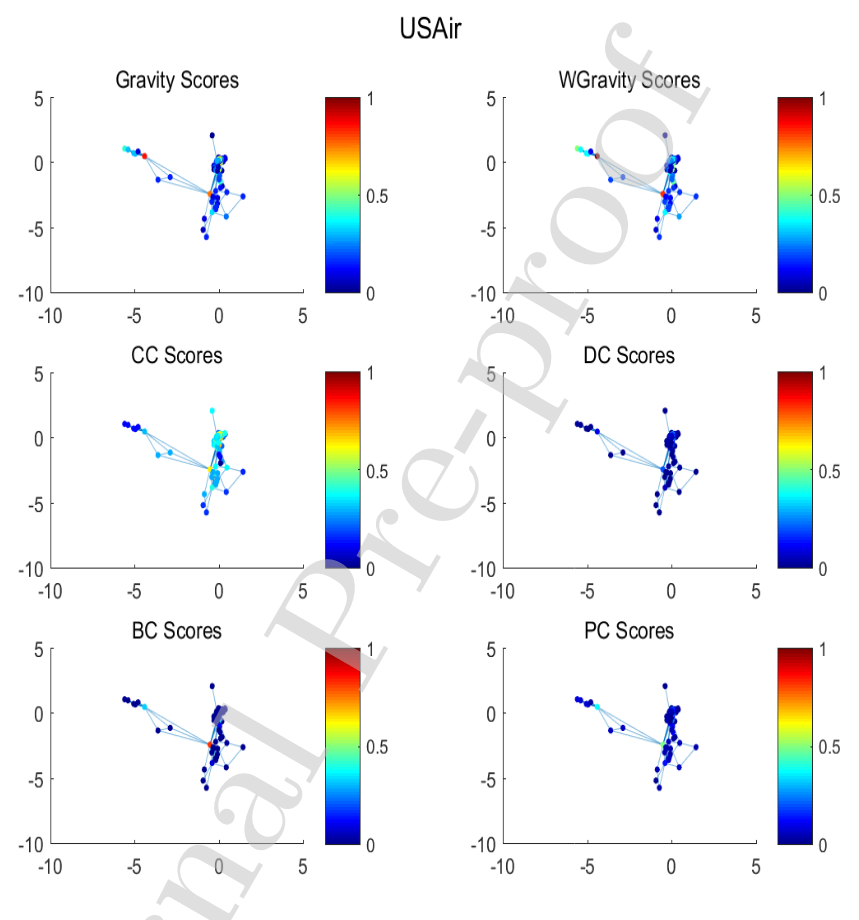
1. **PG**

在PG网络中，CC在判断全局节点的重要性方面有优势。Gravity和WGravity比较一致，尤其是节点重要性的分布方面。



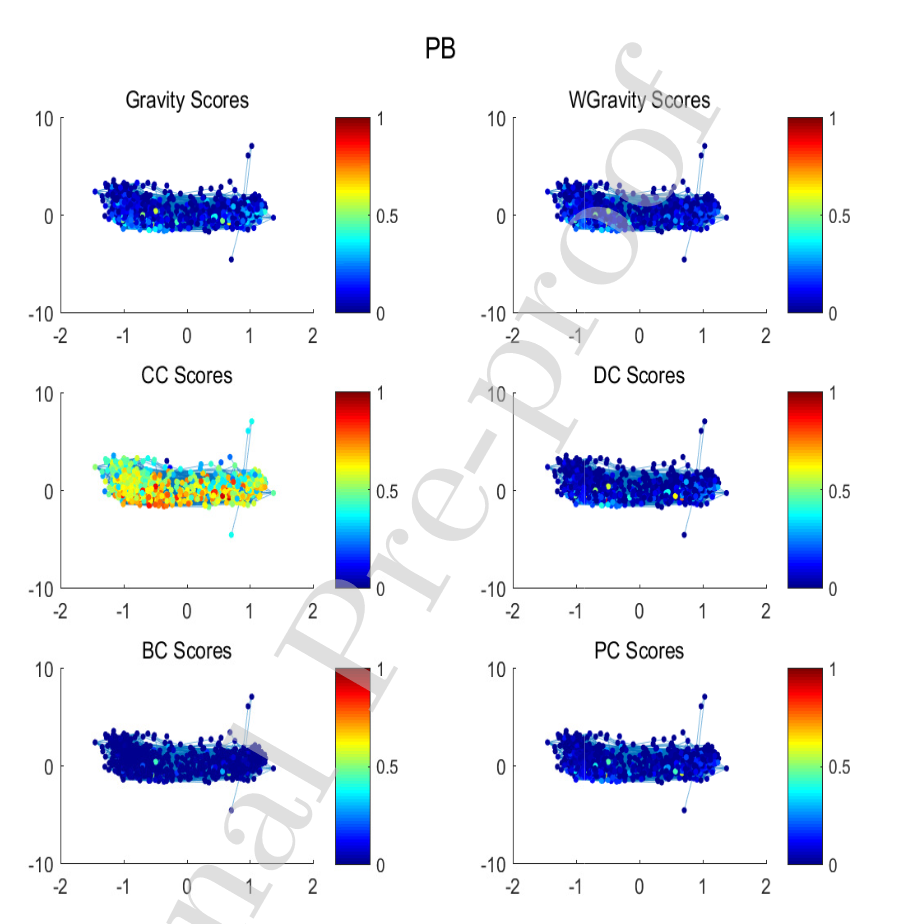
1. **USAir**

在USAir网络中，WGraviy与Gravity略有不同。



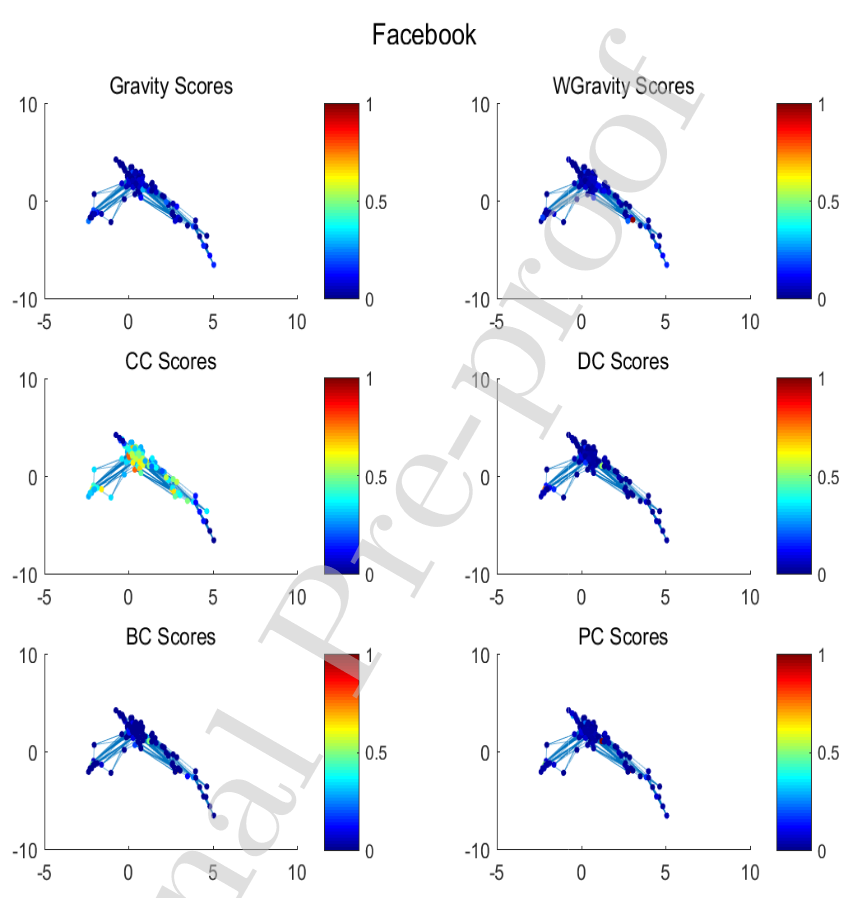
1. **PB**

在PB网络中，CC可以清楚地看到节点的全局重要性的差异。



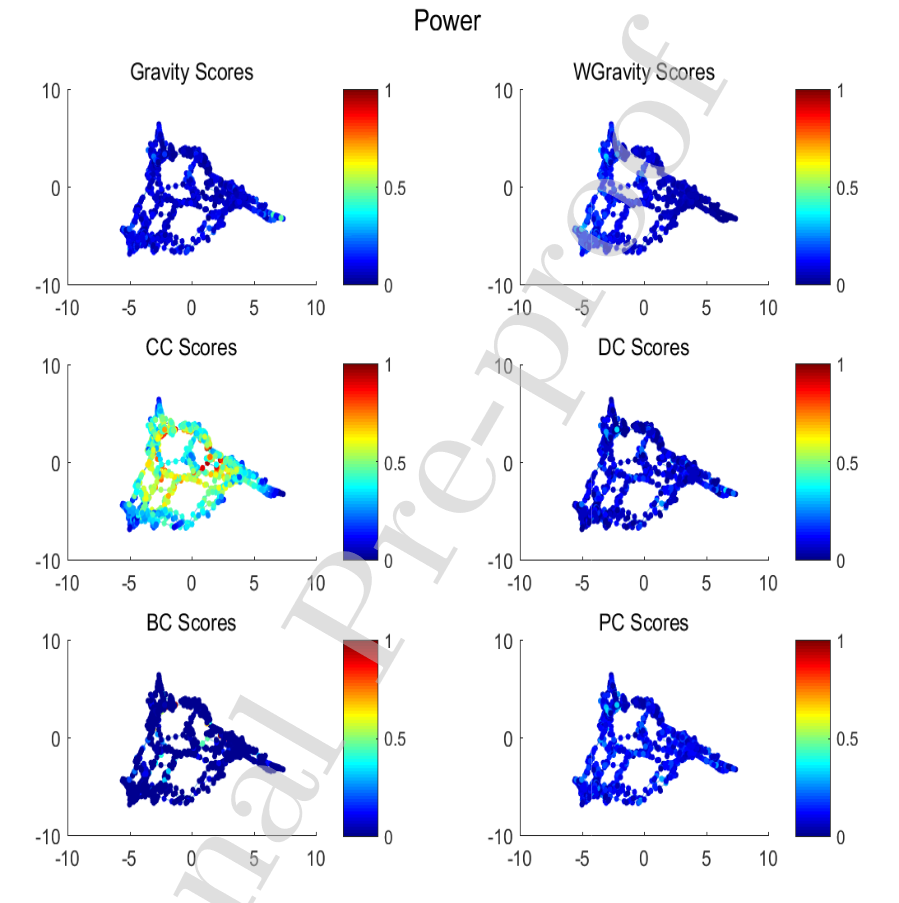
1. **Facebook**

以下是Facebook网络的测试，节点的颜色越深，说明该节点的重要性越大。



1. **Power**

以下是Power网络中的测试结果。

****

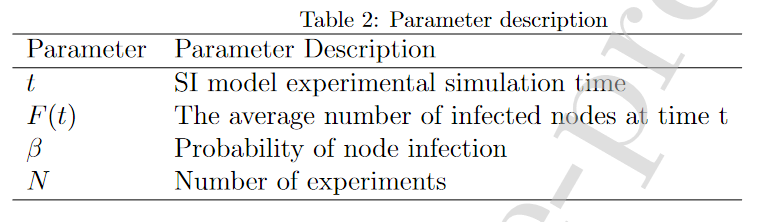
1. **WGravity模型有效性验证**

在本节中，首先介绍SI模型和肯德尔相关系数Kendall’s tau。然后，将比较WGravity与其他方法之间的关系。

**（一）前置定义**

1. **SI模型**

SI模型可用于评估网络中每个节点的传输能力并反映每个节点的影响。在SI模型中，网络的节点有两种状态：易感状态和感染状态。感染状态的节点以一定的概率感染周围的易感节点。表2描述了在SI模型中使用的参数：



其中，t为SI模型的实验仿真时间，F(t)为t时刻感染节点的平均数，β为节点感染的概率，N为实验次数。

1. **肯德尔相关系数Kendall’s tau**

给定两组序列和，xi和yi分别是和的第i个位置的值。让(xi, yi)为一组序列对。如果xi>xj且yi>yj ，或者xi<yi且yi<yj ，那么（xi，yi）被记录为一个正的序列对。否则，它被记录为一个负的序列对。Kendall’s tau的定义如下：

其中n+和n-是正数序列对和负数序列对的数量，n是序列对的总数量。

Kendall’s tau反映了两个序列的相关性。Kendall’s tau的值越大，两个序列之间的相似度就越高。

**（二）模型有效性对比**

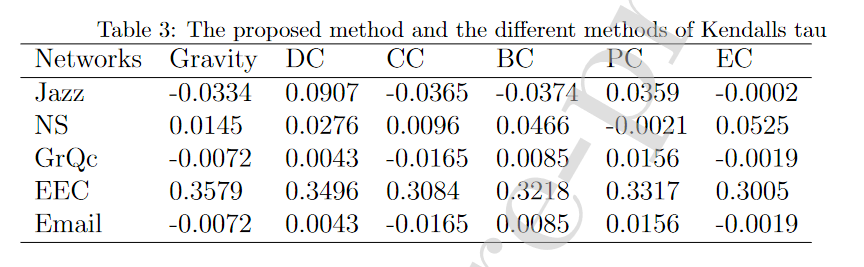
在本实验中，采用了 种方法对模型进行了对比，分别是：1、基于SI模型将其他方法与WGravity进行相似性考察；2、在SI模型下观察TOP100个节点的平均受感染节点数；3、SI模型标准中心性序列为标准考察相似度；4、通过攻击重要节点来比较网络的质量

1. **基于SI模型将其他方法与WGravity进行相似性考察**

在这个实验中，通过比较WGravity和其他方法之间的关系来证明该方法的有效性。所有方法都是在数据集Jazz、NS、GrQc、EEC、Email上进行的。在实验的SI模型中，感染概率β被设定为0.1，时间t被设定为45，所有实验被设定为100次。

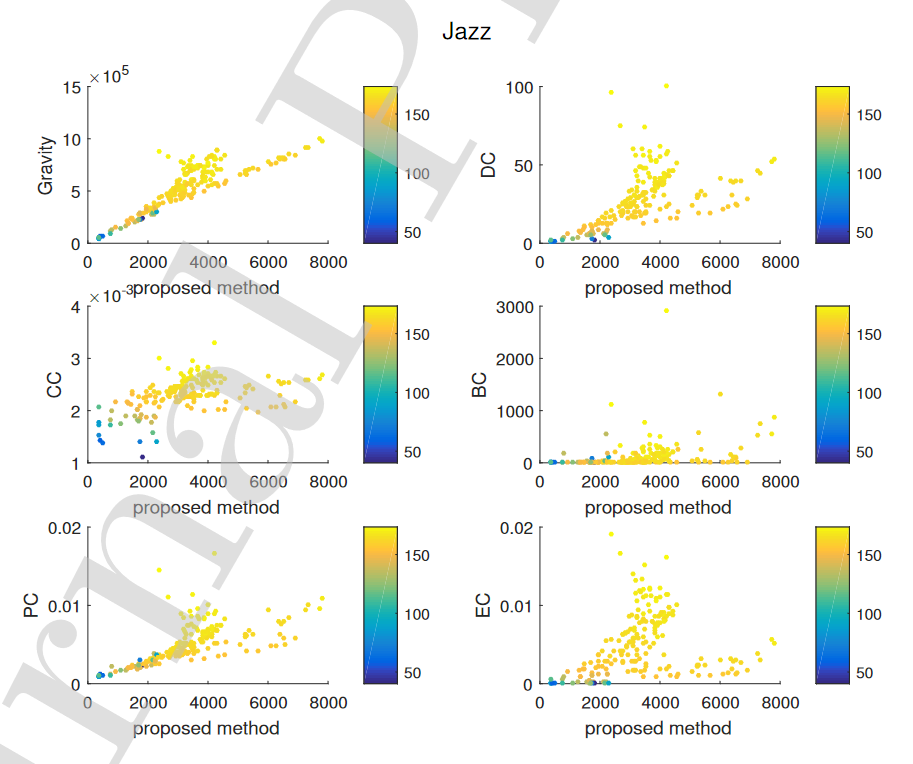
方法：首先，通过前面的节点重要性测试，将每个节点按降序排序。如果节点的中心性分数越大，说明节点的重要性就越大；然后，通过SI模型对每个节点的重要性进行评估。采用了Kendall’s tau系数与分布图进行验证。

所有的实验结果都显示在图12至16中。表3显示了WGravity和其他方法的Kendalls tau值的比较。可以看出，WGravity与重力模型、PC、CC、DC得到的结果有正相关的关系。其中，本文的方法与重力模型有高度的相似性。



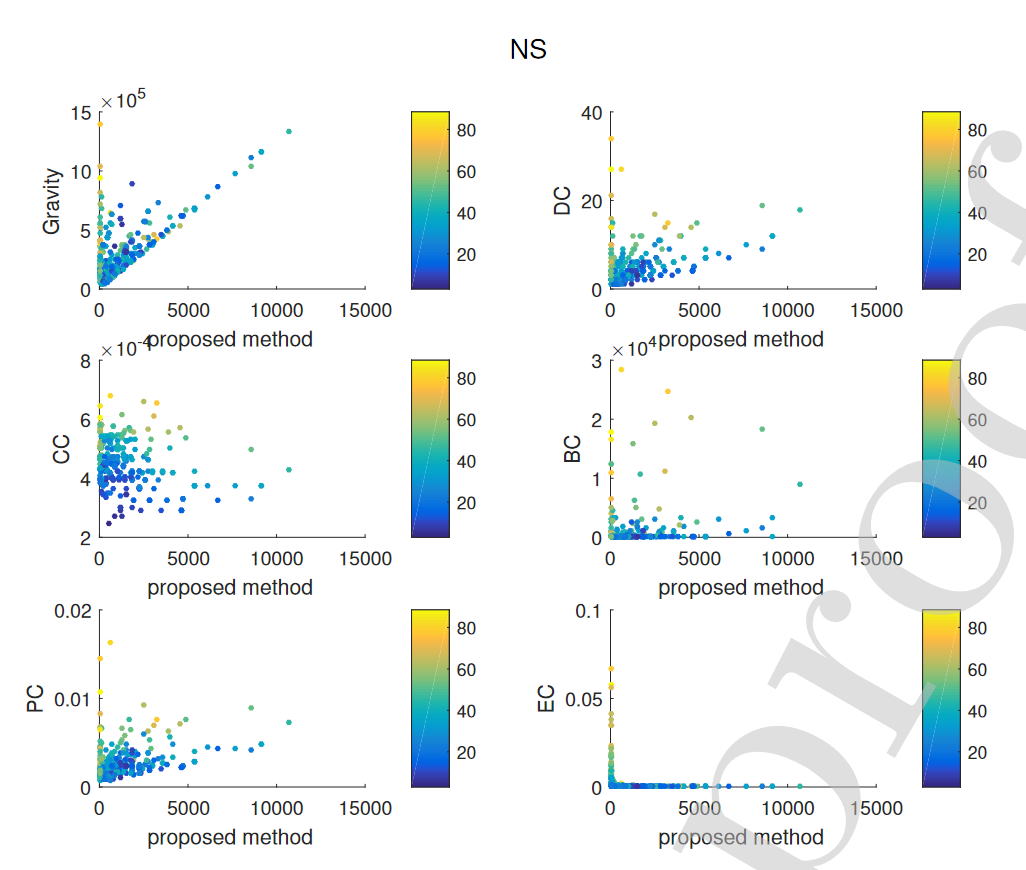
1. Jazz

图中描述了在Jazz网络中，WGravity与其他方法之间的关系。横轴和纵轴都是节点的中心性分数。WGravity与PC和重力模型显示出良好的线性关系。颜色越深，节点在SI模型中的影响就越大。如果散点图的斜率趋势表明WGravity与比较方法之间的相关性。如果斜率是正的，说明是正相关的。



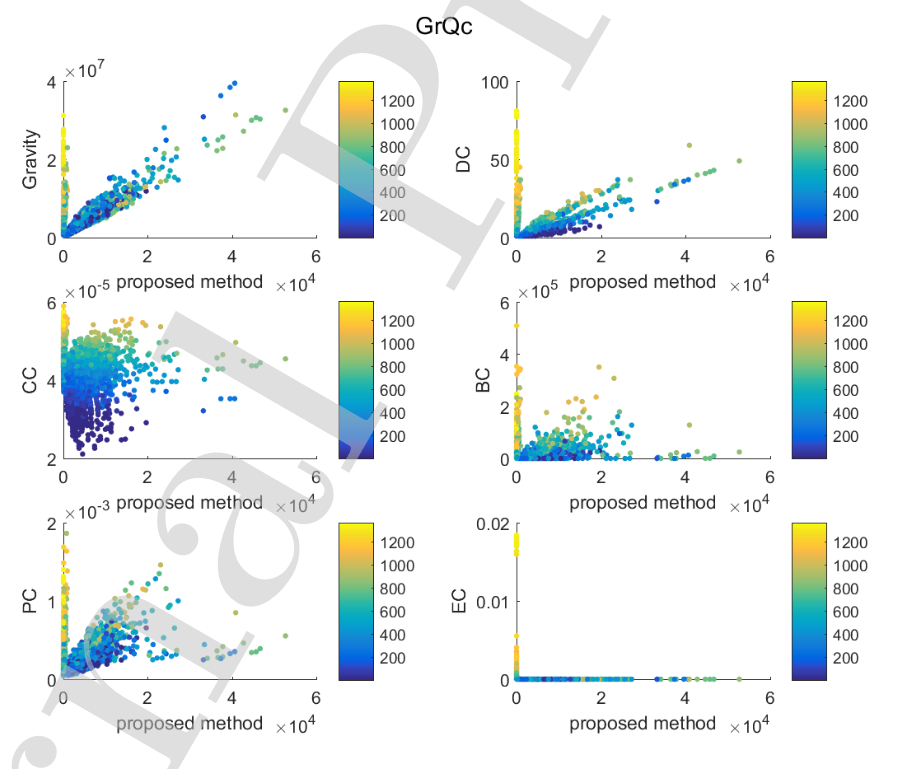
1. NS

该图描述了在NS网络中所提出的方法与其他方法之间的关系。本文的方法显示出与重力模型的强烈相关性。



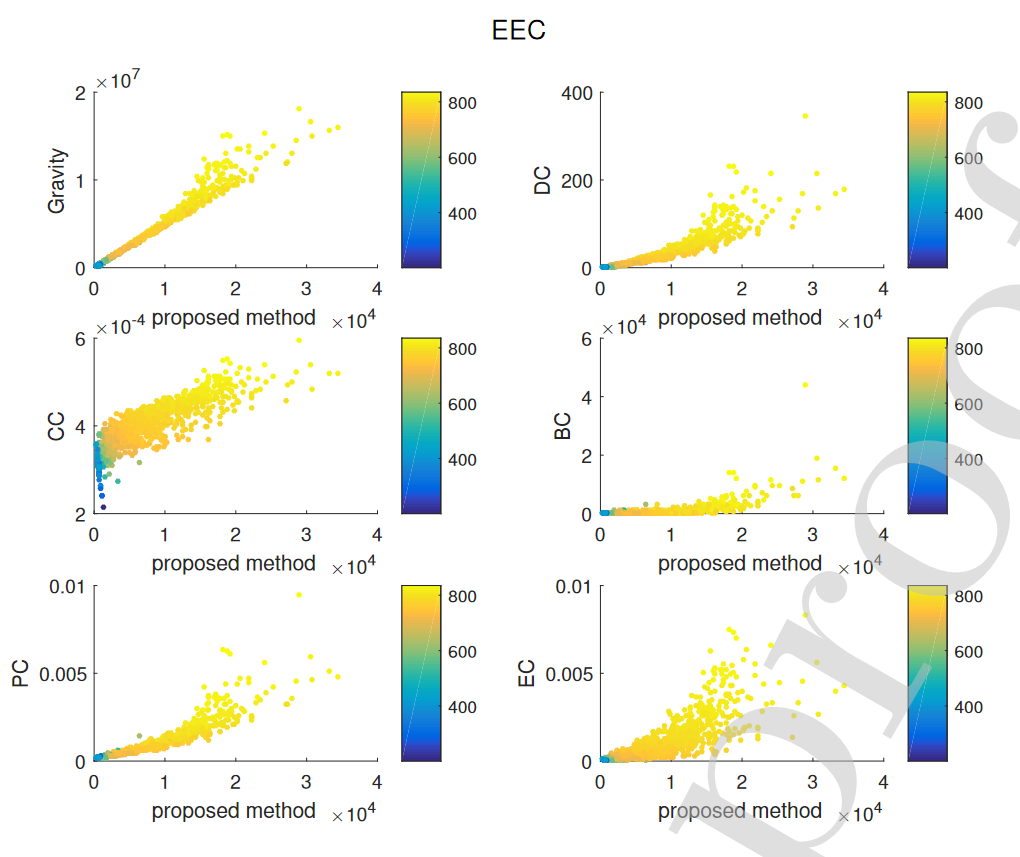
1. GrQc

图中描述了GrQc中所提方法与其他方法之间的关系。本文的方法与PC显示出良好的相关性，与CC和BC的差异不同。



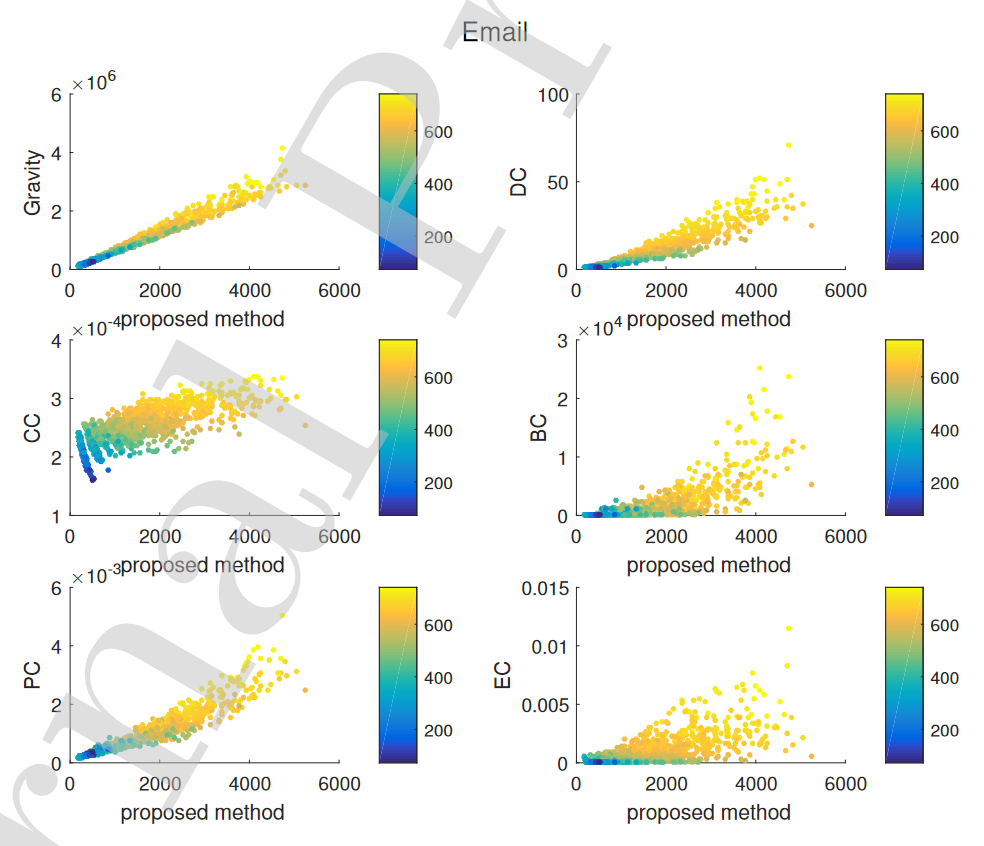
1. EEC

该图描述了EEC中所提出的方法与其他方法之间的关系。本文的方法与其他六种方法都显示出强烈的相关性。



1. Email

该图描述了Email网络中所提出的方法与其他方法之间的关系。WGravity与重力模型、DC和PC显示出强烈的相关性。

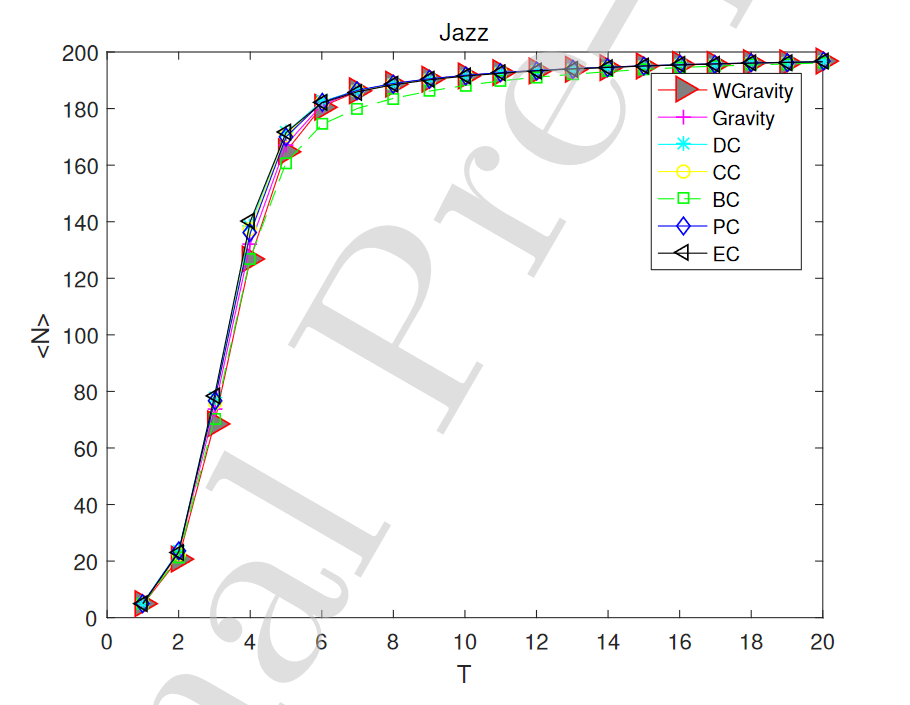


1. **基于SI模型观察TOP节点的平均受感染节点数**

在网络中，如果一个节点比较重要，它的传染能力就比较高。在这个实验中，比较了所提方法和其他方法所选节点的平均感染力。首先，从上一部分的两种方法中各选取前100个节点。然后，将这前100个节点作为SI模型中的感染源。最后，计算每种方法的平均感染节点数。实验结果如图17至21所示。如果节点越重要，它能传播的节点就越多。因此，曲线上升得越快，选择作为种子的序列集就越重要。从实验结果可以看出，WGravity在Email、EEC和Jazz中与其他方法具有相同的效果。

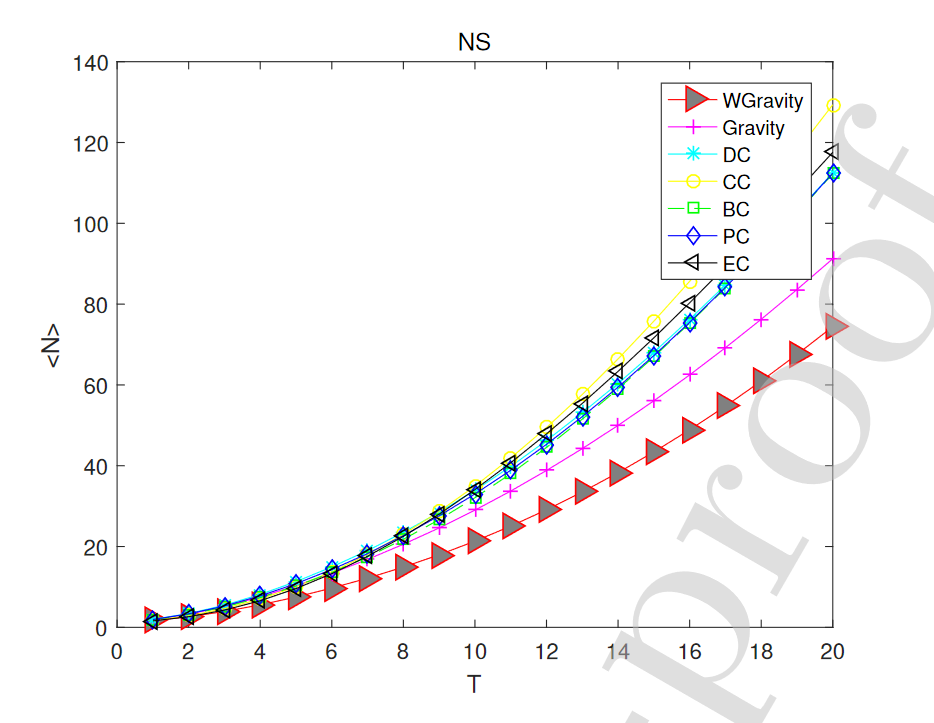
1. Jazz

该图比较了不同方法的情况下，在Jazz网络中选择的前100个节点的平均受感染节点的数量。如果曲线上升较快，说明所选的前100个节点对SI模型的影响较大。F(t)表示时间t时节点感染的平均数，所有方法的上升趋势基本相同，但时间T=4后BC的平均影响数最少。



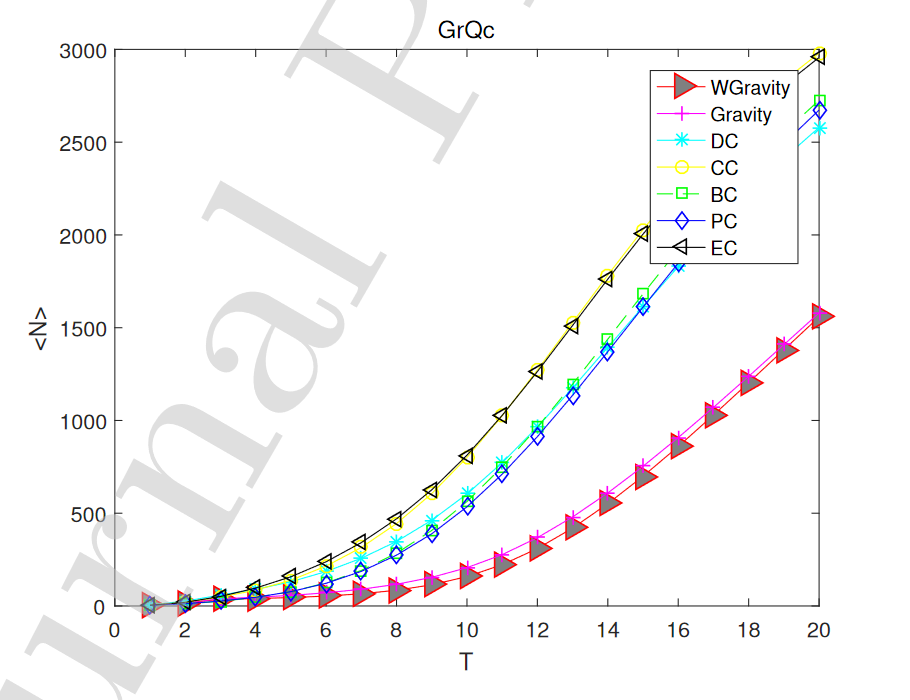
1. NS

该图比较了不同方法的情况下，在NS网络中选择的前100个节点的平均受感染节点的数量。所有方法在NS中的表现都明显不同。CC的前100个节点的影响最大，曲线上升较快。WGravity的前100个节点的影响数最少。



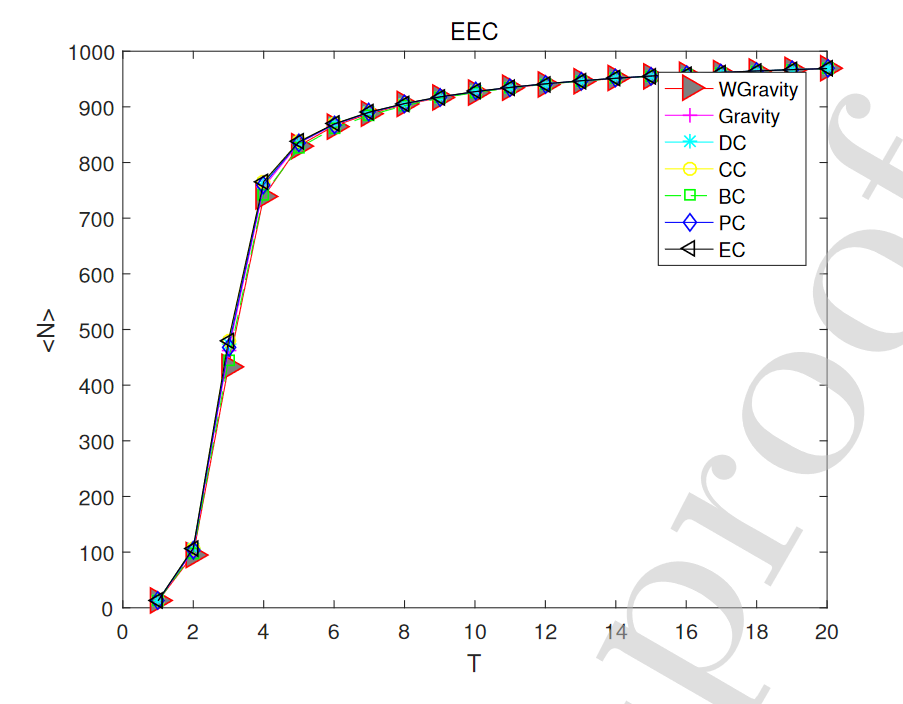
1. GrQc

该图比较了不同方法的情况下，在GrQc网络中选择的前100个节点的平均受感染节点数。CC和EC的前100个节点的影响数趋势一致，而且曲线上升较快。WGravity的前100个节点的影响数最少。



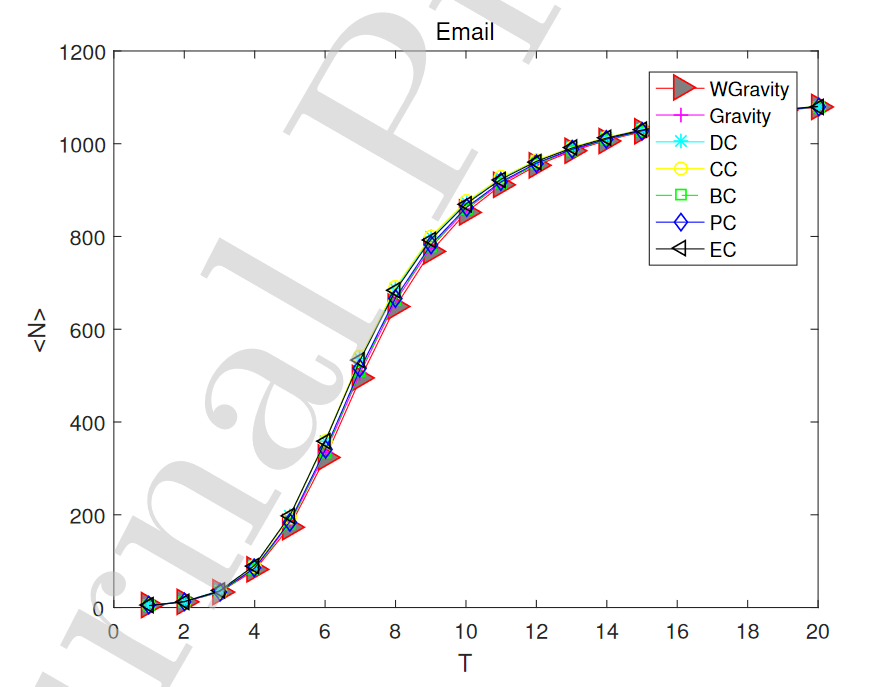
1. EEC

该图比较了不同方法的情况下，在EEC网络中选择的前100个节点的平均感染节点数。所有方法在EEC网络中的差异并不明显，上升的趋势是一致的。在前两步中，所有的曲线都迅速上升，在第八步中逐渐收敛。然而，EC明显上升的趋势比其他方法略高。



1. Email

该图比较了不同方法的情况下，在Email网络中选取的前100个节点的平均受感染节点数。所有方法在Email网络中的差异并不明显，上升的趋势是一致的。然而，可以看出，受Wgravity方法选出的节点影响的节点数量略小。

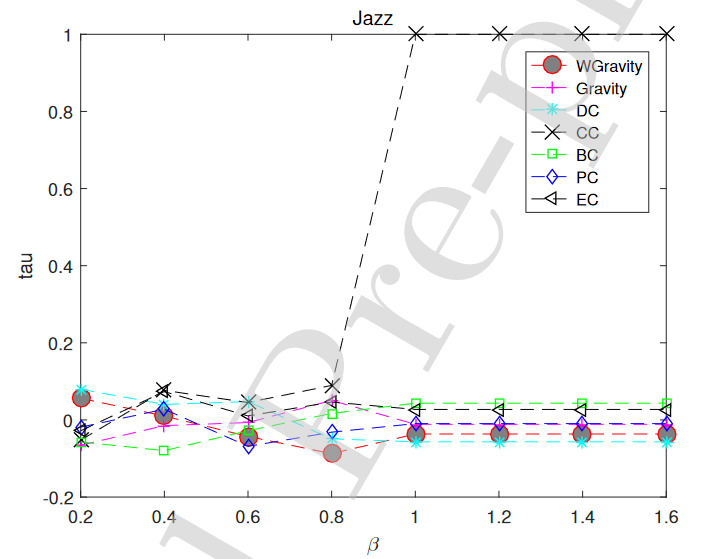


1. **SI模型标准中心性序列为标准考察相似度**

在所有的数据集中，在一个SI模型中给出一个传播概率β，从而得到一个标准的中心性序列，然后我们将其他方法与这个序列进行比较。在所有实验中，每个SI模型独立进行100次，β从0.2到1.6变化。实验结果如图22至25所示。可以看出，随着β的不断增加，CC和SI的序列不断接近。本文提出的方法与其他方法在网络EEC上得出的序列很接近。

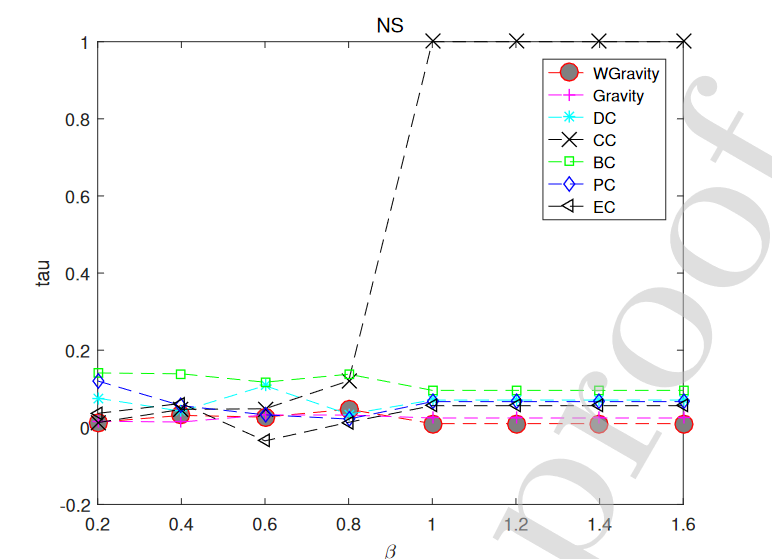
1. Jazz

当β变化时，该图描述了Jazz网络中SI生成的标准序列与其他方法生成的序列之间的差异。当β增加时，CC的tau值逐渐接近1，意味着逐渐与SI相同，而DC和SI的差异最大。



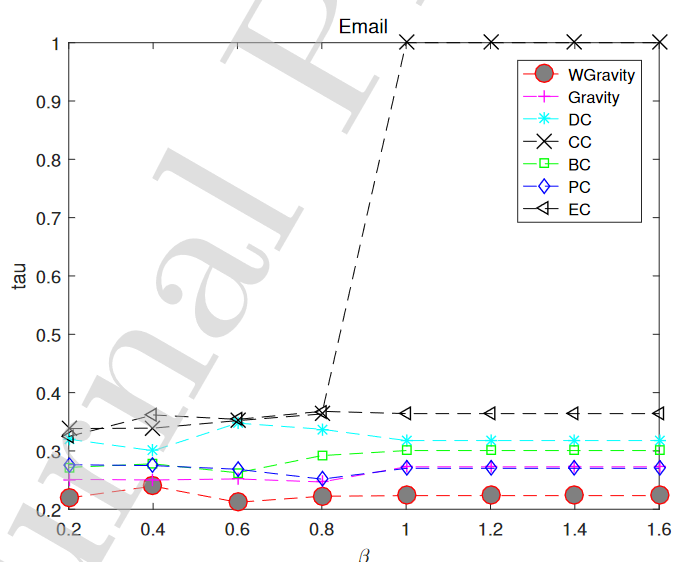
1. NS

当β发生变化时，该图描述了NS网络中SI生成的标准序列与其他方法之间的差异。当β增加时，CC的tau逐渐接近于1，意味着逐渐与SI相同。EC有所波动，当β=0.6时，WGravity和SI序列的tau值最低。



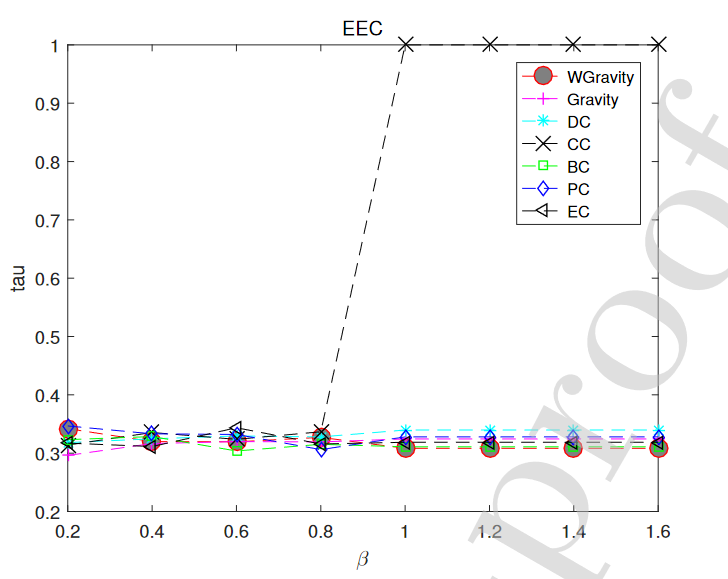
1. Email

当β发生变化时，该图描述了在Email网络中由SI产生的标准序列和其他方法之间的差异。WGravity和SI序列的tau是最低的。EC和SI序列的tau值是第二高的。



1. EEC

当β发生变化时，该图描述了在EEC网络中由SI生成的标准序列与其他方法之间的差异。除了CC，其他方法与SI序列的tau值是最低的。



1. **通过攻击重要节点来比较网络的质量**

网络的连接性可以作为网络质量的一个指标。**如果网络的连通性比较好，那么网络的质量就比较高**。在这个实验中，我们通过攻击每个方法来测试网络的质量，以给出一个顶级节点的序列。**网络的质量**定义如下：

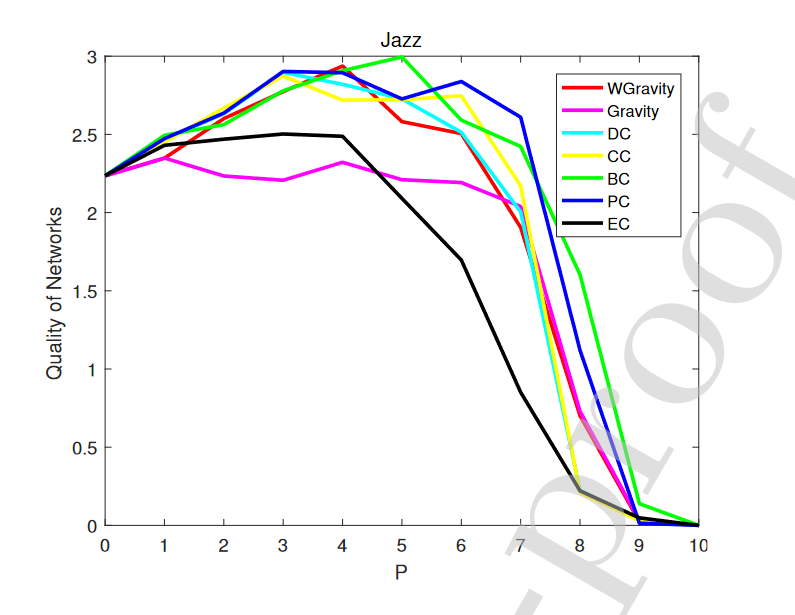
其中是节点i和节点j之间的最短距离，是节点的数量。

在这个实验中，我们按照节点重要性的顺序对P×10%的节点进行攻击。首先，根据每种方法，得出节点重要性的顺序。然后，攻击每个节点并使节点无效。最后，计算出网络的质量。实验结果如图26至30所示。

可以看出，WGravity在数据集Jazz、GrQc中表现出良好的性能。随着攻击的重要节点的百分比继续增加，网络的质量迅速下降。

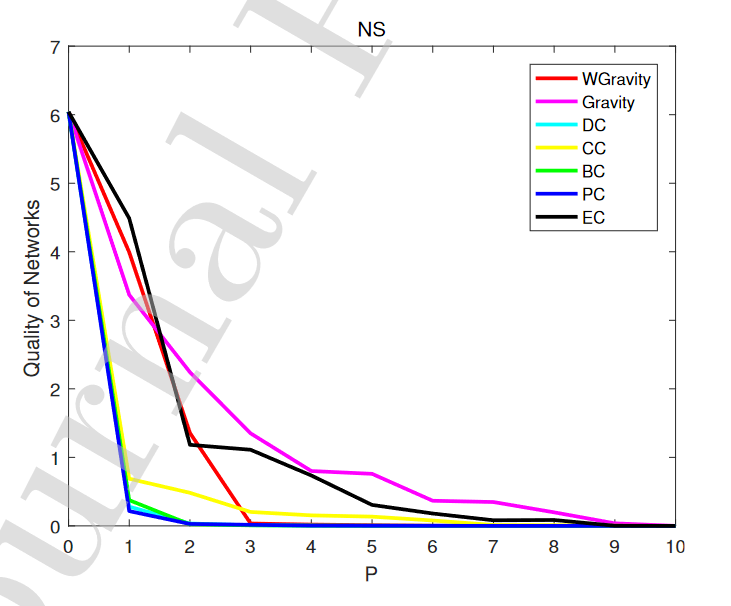
1. Jazz

该图描述了不同方法产生的节点序列受到攻击时，Jazz网络质量的变化。如果网络质量下降较快，说明该方法选择的节点在网络路径中更重要。起初，重力法的网络质量下降最快，然后EC迅速下降。所有方法中BC的下降速度都相对较慢。



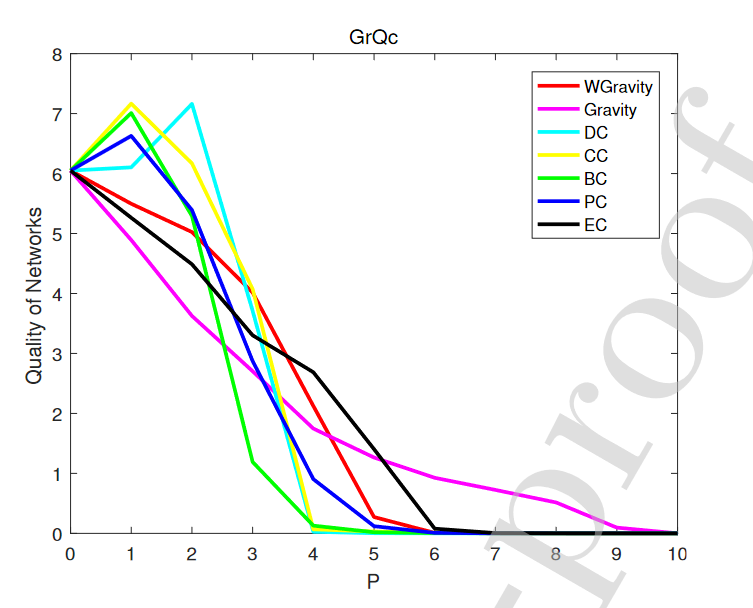
1. NS

图中描述了不同方法产生的节点序列受到攻击时NS网络质量的变化。在所有的方法中，PC下降最快，其次是CC，重力模型下降最慢。



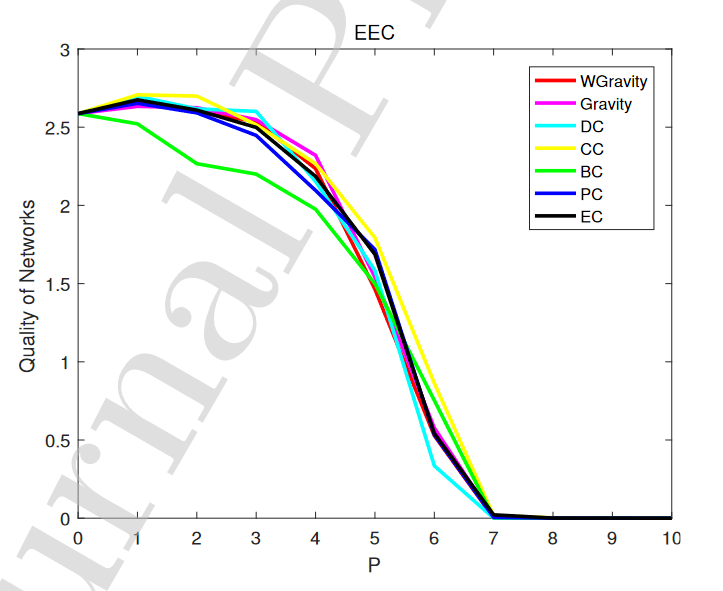
1. GrQc

该图描述了不同方法产生的节点序列受到攻击时GrQc网络质量的变化。起初，Gravity下降得最快，然后PC迅速下降。



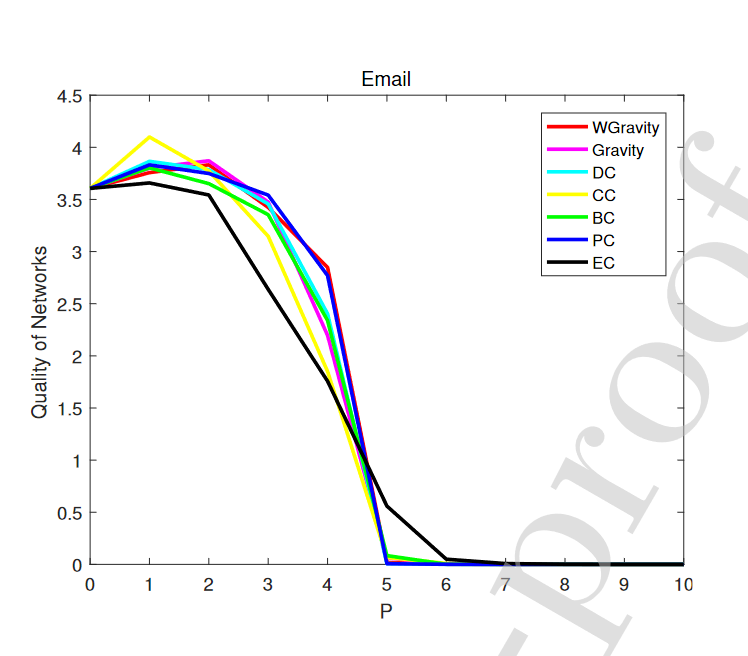
1. EEC

图中描述了不同方法产生的节点序列受到攻击时，EEC网络质量的变化。在所有方法中，BC下降最快，其他方法差别不大。



1. Email

该图描述了不同方法产生的节点序列受到攻击时，电子邮件网络质量的变化。EC掉得最快，其次是CC。其他方法的差别不大。



### 1.4 结论

一、论文结论

本文提出了一个广义力学模型WGravity来评估节点的重要性。为了验证所提方法的有效性，我们将所提方法与五个不同的方法进行了比较。实验结果显示了所提方法的有效性。

此外，还提出了一个创新的实验来验证网络节点的重要性：首先，定义了一个网络质量Q；然后，通过攻击重要节点，计算出网络的质量。

总之，GMM同时考虑了局部和全局信息。**如果一个节点的度数越大，与其他节点的距离越短，权重越大，该节点的重要性就越大**。此外，GMM继承了重力模型的**截断半径**，以处理大规模网络中的耗时问题。重力模型和GMM在判断每个节点的重要性方面的表现相对一致，但GMM在全局上有明显的不同。本文认为，当引入全局信息（节点的相对权重）后，可以从全局的角度来判断节点的重要性差异，这样会更加精细。

二、收获与总结

1. 行文严谨，几乎所有的重要语句都标注了可靠的出处，几乎所有的基础理论和背景都能做到有迹可循。
2. 实证分析和有效性测试相当充分，进行了大量的实验，除直接测试外还通过三种方法进行了有效性测试。绘制的图表数量极多、质量很高、容易理解。
3. 研究的延续性良好，虽然不是特别大的创新，但对前人的研究成果做到了比较好的继承，如EC中心性和截断半径概念。
4. 本文将【节点的度】类比为引力公式中的质量，将【节点间最短距离】类比为引力公式中的距离。或许可以在节点“质量”的概念定义上进行进一步学习和研究，节点间相互作用的“距离”如何更进一步完善也比较值得探讨，后续的阅读中会注意这类问题。
5. 文中提出的网络质量指标Q的定义和【依次攻击一定比例top节点】从而验证节点重要性排序结果合理性的方法值得记录。
6. 将物理学中的引力场转化为复杂网络中的引力场是一个比较有意思的思路，或许在未来可以尝试探索传统学科中还有哪些理论可以恰当地转化为复杂网络中的方法。