目 录

[1. Identifying influential nodes in complex networks: Effective distance gravity model 2](#_Toc105350924)

[1.1 基本信息 2](#_Toc105350925)

[1.2 论文内容 2](#_Toc105350926)

[1.3 实证分析 10](#_Toc105350927)

[1.4 结论 21](#_Toc105350928)

## 1. Identifying influential nodes in complex networks: Effective distance gravity model

识别复杂网络中具有影响力的节点：有效距离引力模型

### 1.1 基本信息

**期刊**：*《Information Sciences》*

**发表时间**： 17 January 2021

**关键词：**Complex networks, Influential nodes, Gravity model, Effective distance, Susceptible infected model

**作者**： Qiuyan Shang, Yong Deng, Kang Hao Cheong

**作者信息**：

电子科技大学基础与前沿科学研究院

电子科技大学英才学院

陕西师范大学教育学院

新加坡科技与设计大学Science, Mathematics and Technology Cluster

SUTD-马萨诸塞州技术学院国际设计中心

**分区**：



### 1.2 论文内容

**一、摘要**

复杂网络中重要节点的识别通常使用传统的欧氏距离，它只关注节点之间的局部静态地理距离，却忽略了现实世界网络中节点之间的动态互动。为了识别有影响力的节点，静态和动态信息都应该被考虑。为了解决这个问题，本文提出了一个原创的、新颖的、具有有效距离的引力模型，用于识别基于信息融合和多层次处理的有影响力的节点。本文的方法能够综合考虑复杂网络的全局和局部信息，并利用有效距离来整合静态和动态信息。此外，所提出的方法可以帮助本文挖掘现实世界网络中隐藏的拓扑结构，获得更准确的结果。利用易受感染的模型、Kendall相关系数和现有的八种识别方法，对12个不同的真实网络进行了模拟实验。

1. **引言**

近年来，复杂网络的研究已经引起了巨大的关注。许多现实世界的问题可以作为网络科学的一部分进行分析，例如，互联网安全、网络控制系统和社会网络。因此，识别复杂网络中的有影响力的节点在结构和功能方面都起着重要作用，这成为一个重要的研究领域。它可以应用于各个领域来解决现实问题，如疾病和谣言控制、网络系统、生物学、社会系统、影响力最大化、时间序列预测、社区寻找和信息传播。有许多现有的方法来评估节点的影响力。其中一类方法关注节点的局部信息，但忽略了全局信息，如DC、K壳分解、特征向量中心性（EC）和PageRank（PC）。另一种类型表明，节点的影响力主要取决于网络中的连接性，如接近中心（CC）和介数中心（BC）。尽管这些识别方法通常可以给出合理的结果，但它们也有很多局限性。例如，DC只考虑邻居的影响，而忽略了全局影响。PC在有向网络上有很好的表现，但在无向网络上的表现却不是很好。CC和BC对网络结构非常敏感，其复杂度也很高。此外，这些方法只考虑节点的局部信息或全局信息，不够现实。

最近提出了各种新颖的识别方法来解决局部和全局信息的问题。例如，Chen等人利用了路径多样性进行识别。Ren等人提出了一个基于迭代资源分配的模型。Deng等人通过反平方律检测重要节点。受万有引力定律的启发，Li等人提出了一种基于引力模型的算法，称为引力模型（GM）。Jiang等人从分形属性的角度，提出了模糊局部维度来检测有影响力的节点。此外, Wen等人通过局部信息维度来识别重要节点. 由于证据理论对不同来源的数据融合的效率，一些对有影响力的节点进行排序的证据方法已经被研究。还有一些其他的方法用于识别有影响力的节点，如基于随机漫步的方法，基于TOPSIS的方法，熵模型和准拉普拉克中心度。

现有的大多数方法只关注节点的静态信息，而忽略了现实世界网络中节点之间的动态信息，从而导致结果不尽人意。在现实世界中，大多数复杂的网络都有隐藏的动态拓扑结构，包含了节点之间的动态信息。例如，电力网络中节点之间的电力传输，社交网络中用户之间的信息流，蛋白质网络中化学物质的交换，疾病传播网络中病毒的传播。在可观察的几何结构背后，网络中还隐藏着主导动态过程的动态结构，特别是在现实世界的应用中。这种动态拓扑结构包含的信息可以帮助本文更好地识别现实世界中的重要节点，比如**节点之间的动态交互信息**。据评估，这种动态信息与明确的静态信息一样重要，不能被忽视。因此，如何有效地挖掘网络的隐性拓扑结构和节点之间的动态交互信息，已经成为一个备受关注的话题。Brockmann等人提出的**有效距离**可以帮助本文解决这个问题。【有效距离是一种从概率中抽象出来的距离】。它主要关注网络中节点的互动，这是判断的主要依据。【有效距离的核心是通过邻接矩阵计算概率来发现两个节点之间最可能的路径】，它充分考虑了现实世界网络中节点之间的动态信息交互。

这项工作的主要贡献在于本文提出的有效距离引力模型。首先，有效距离引力模型全面考虑了节点的局部信息和网络的全局信息，因此与传统的只关注局部或全局信息的中心性方法相比，具有优势。第二，本文提出的方法可以挖掘出网络中隐藏的动态结构和节点之间的动态交互信息，这些信息在网络的实际运行中起着主导作用。第三，动态信息和显性静态信息的结合可以更好地检测复杂网络中的重要节点。第四，本文模型中的累积中心度得分计算方法减少了由不稳定结构引起的识别误差。

基于本文提出的方法，本文利用SI模型对12个网络进行了多种实验，并与现有的8种方法进行了比较。本文的实验结果表明，与现有的方法相比，本文提出的方法具有稳健性和合理性。

**三、模型解析**

1. **有效距离Effective distance**

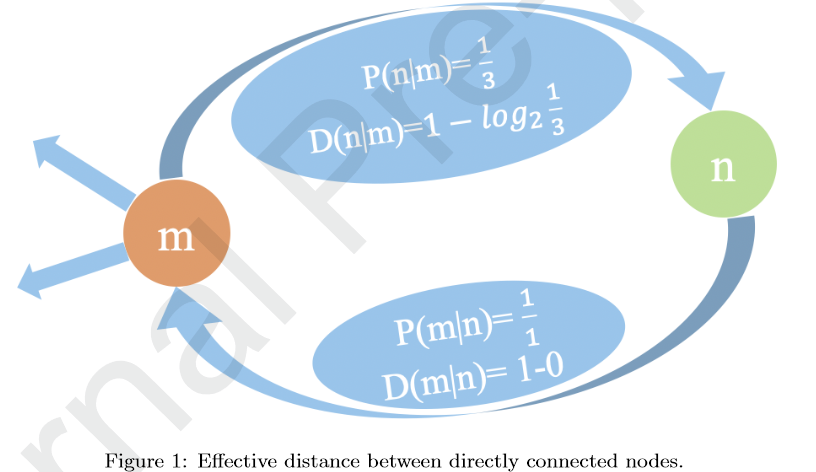
从彼此直接相连的节点m到节点n的有效距离定义为：

*注：“这个有效距离的概念反映了这样一个观点：小部分的交通量n→m实际上相当于一个大的距离，反之亦然。对数是有效长度作加法运算的要求的结果，鉴于沿多步骤路径的概率是乘法的。”——Brockmann*

其中是节点m到节点n的有效距离值（如果它们是直接连接的）。是节点m到节点n的概率。可以通过以下公式得到：

其中，是节点m在有向图中的出度或节点m在无向图中的度。是图G的邻接矩阵中的元素。

Figure 1揭示了直接连接节点之间有效距离的计算。显然，从Figure 1中可以看出， ≠ ， ≠ 。节点i到自身的有效距离为0。与欧氏距离的一个重要区别是，有效距离是不对称的（m到n和n到m的距离并不一定相等）。



对于间接连接的节点，它们之间的有效距离可以通过过境方式获得。例如，从节点C到节点F的有效距离可以通过形如：

= +

的形式得到。如果从节点m到n点有多条路径，本文将使用这两个节点之间最短的路径。根据网络中最短路径的思想，从节点m到n的所有有效距离中选择最短的一条作为最终有效距离：

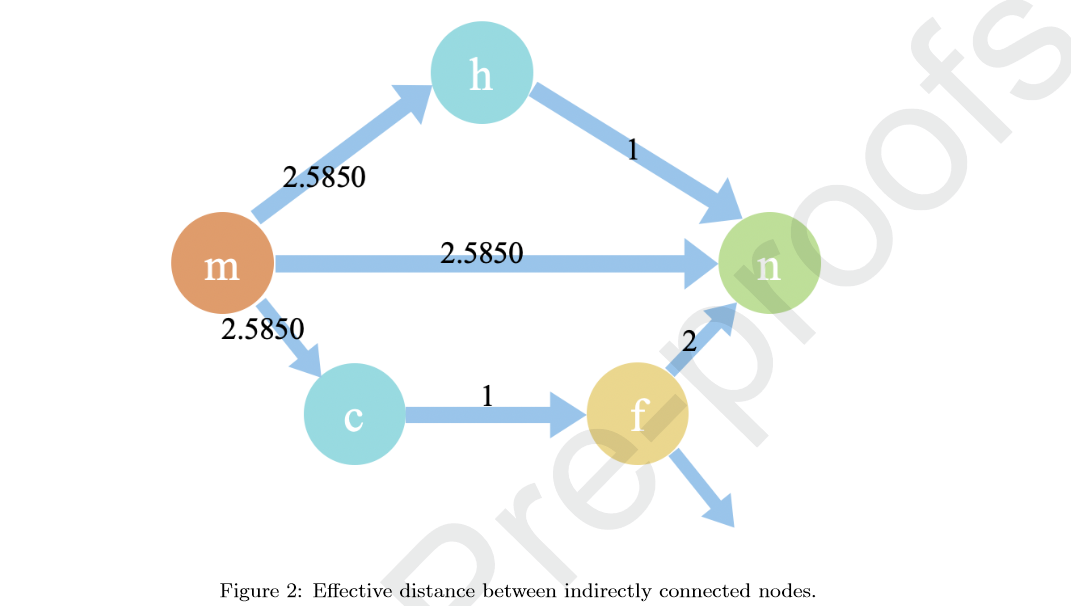
其中代表节点m到节点n的不同有效距离。值得注意的是，和通常不相等。Figure 2是一个计算间接连接节点之间有效距离的简单例子。可以发现，从节点m到节点n有三条可能的路径，其中：

= 2.5850 + 1 = 3.5850

= 2.5850

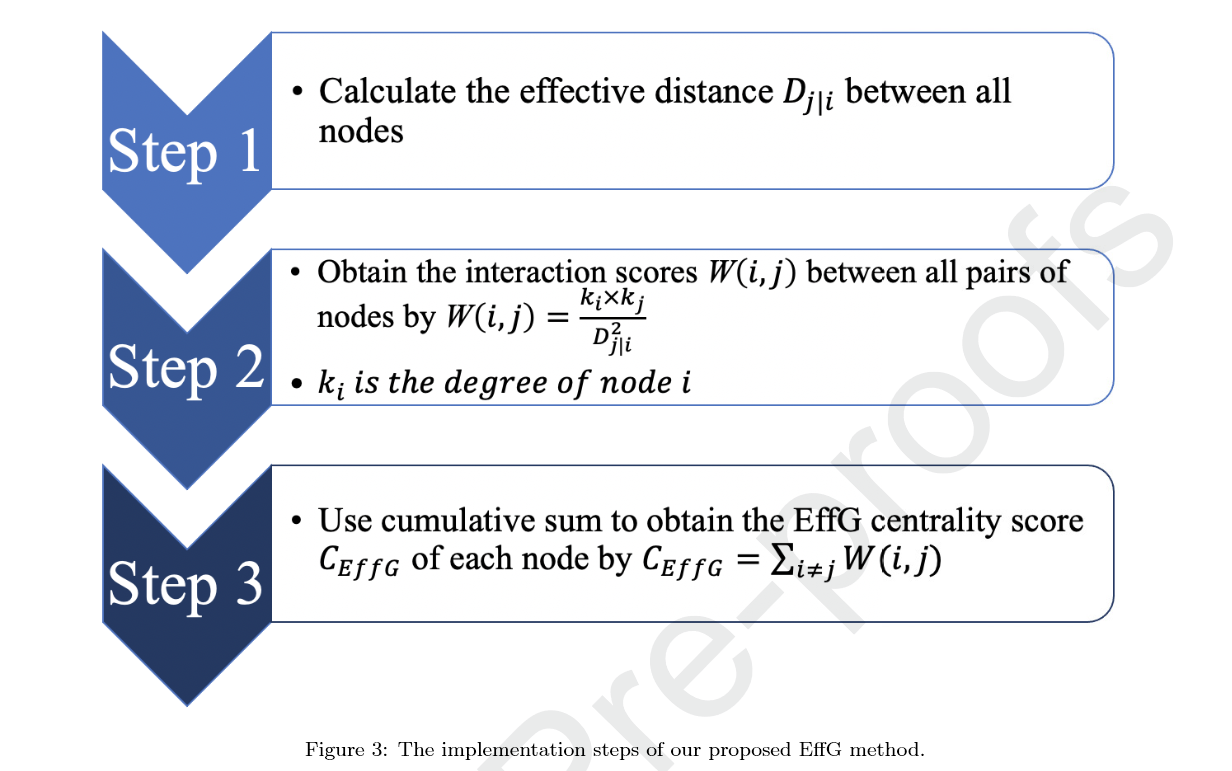
= 2.5850 + 1 + 2 = 5.5850

则最终的有效距离是最短路径，等于2.5850。



1. **有效距离引力模型EffG（本文提出）**
2. **有效距离引力模型EffG**

EffG方法的框架如图3所示，下面将进一步给出计算的细节。



**步骤1：计算所有节点之间的有效距离**

Dirk Brockmann和Dirk Helbing提出了【有效距离】，如果用概率来构建一个新的距离度量，以取代传统的地理距离，那么复杂的时空模式可以被简化为简单和均匀的波传播模式。当给定疾病的参数和网络结构时，从有效距离获得的通量信息可以可靠地预测疾病的到达时间。因此，利用有效距离来有效挖掘网络中隐藏的动态拓扑结构和节点间的动态信息是合理的。节点间的有效距离可以通过前文给出的公式计算：

路径概率：

有效距离：

最终有效距离：

**步骤2：计算所有节点对之间的互动分数**

基于引力公式，节点之间的具体【**交互得分】**定义如下：

其中和分别为节点i和节点j的度数。是节点i到节点j的有效距离。表示节点i和节点j之间的交互得分。

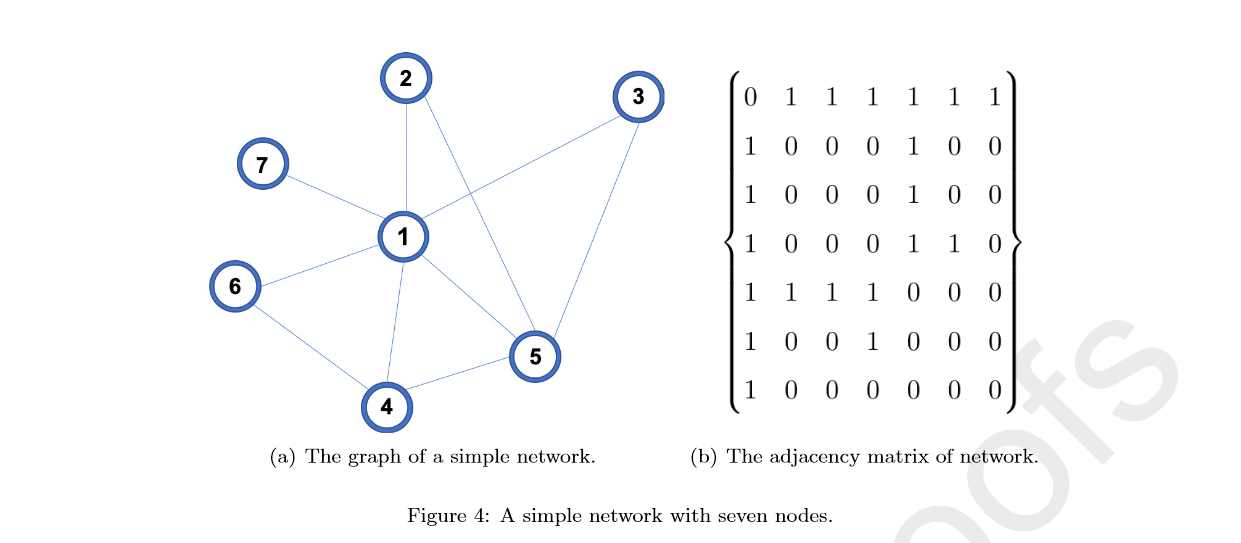
**步骤3：使用累积总分来获得每个节点的EffG中心性得分**

在获得节点之间的交互得分后，每个节点的EffG得分由其与网络中其他节点的交互得分的累积之和得到。因此，【**EffG中心性分数**】可以定义如下：

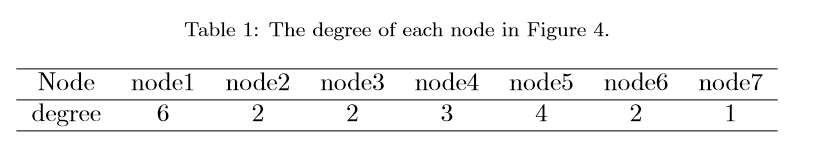
其中，N是网络中所有节点的数量，代表节点i的EffG中心性得分。

1. **数据示例**

为了更好地解释EffG，下面给出一个数字例子，说明EffG如何工作。以节点2为例，计算它的EffG分数。Figure 4(a)是一个网络的图，Figure 4(b)表示邻接矩阵。

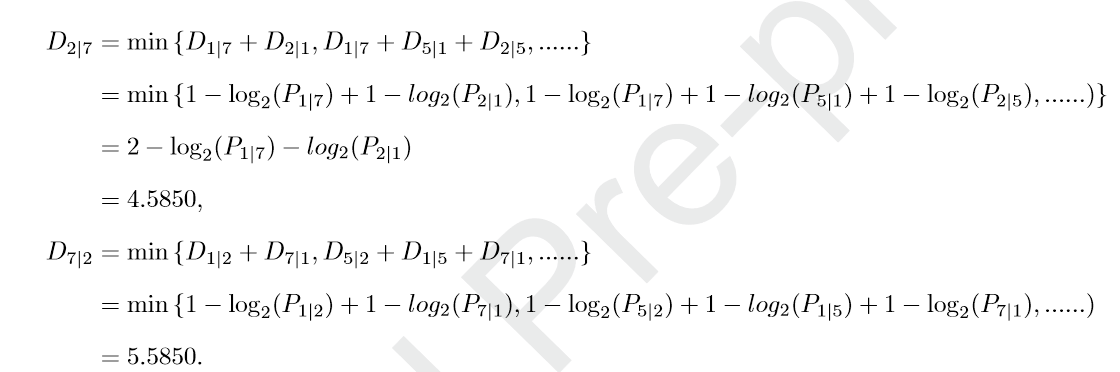


图中每个节点的度数如表1所示：

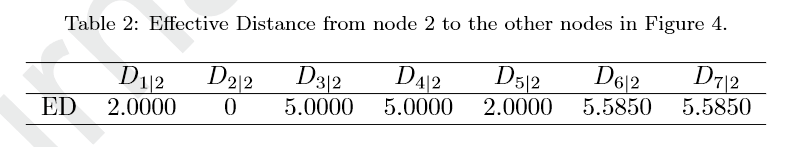


* **第一步：计算所有节点的有效距离**

首先，通过有效距离公式计算节点之间的有效距离。在正常情况下，。则节点2和节点7之间的有效距离可以计算如下：

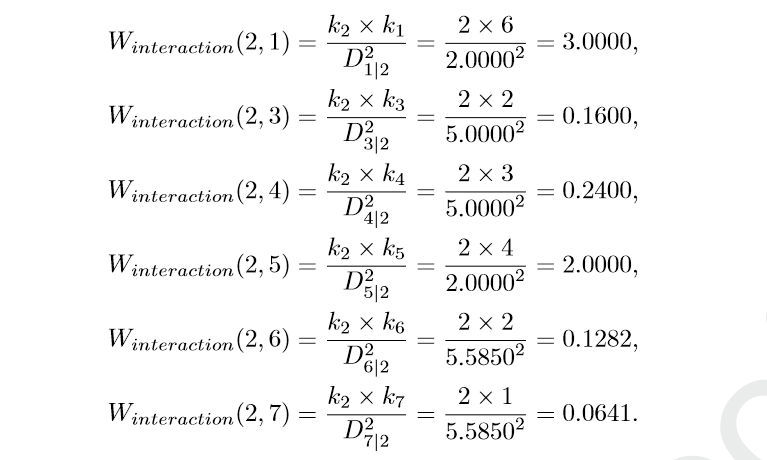


正如所观察到的，和是不同的，正如本文上面所讨论的。使用同样的程序，也可以计算出从节点2到其他节点的有效距离。结果显示在表2中：



* **第二步：计算所有节点对之间的相互作用分数**

随后，节点2和其他节点之间的互动分数 可以通过互动分数公式计算如下：

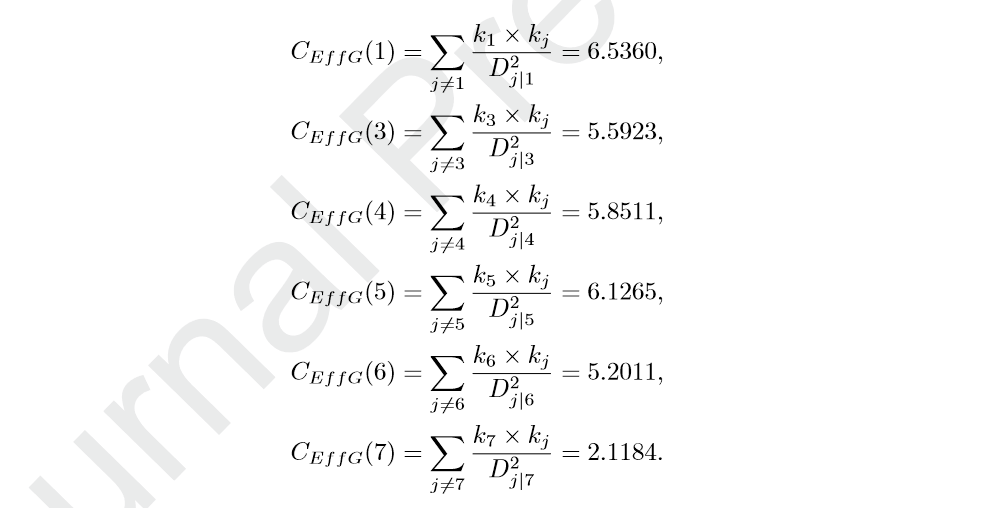


* **第三步：使用累积总分来获得节点2的EffG中心性得分**

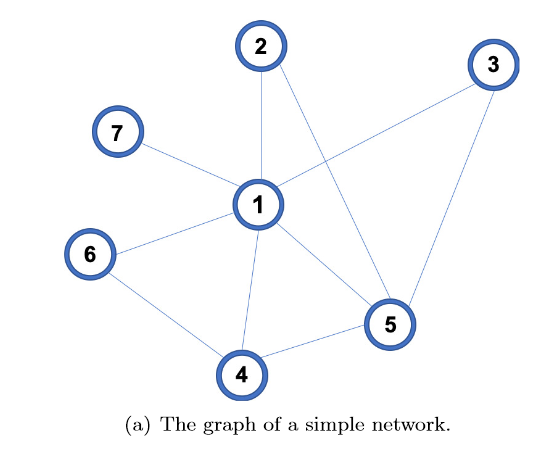
节点2的EffG得分可以通过公式（11）得到，如下所示：



其他节点的EffG得分可以通过同样的程序计算，如下所示：



计算结果结合图4(a)可以表明，本文提出的EffG方法是实用和客观的。

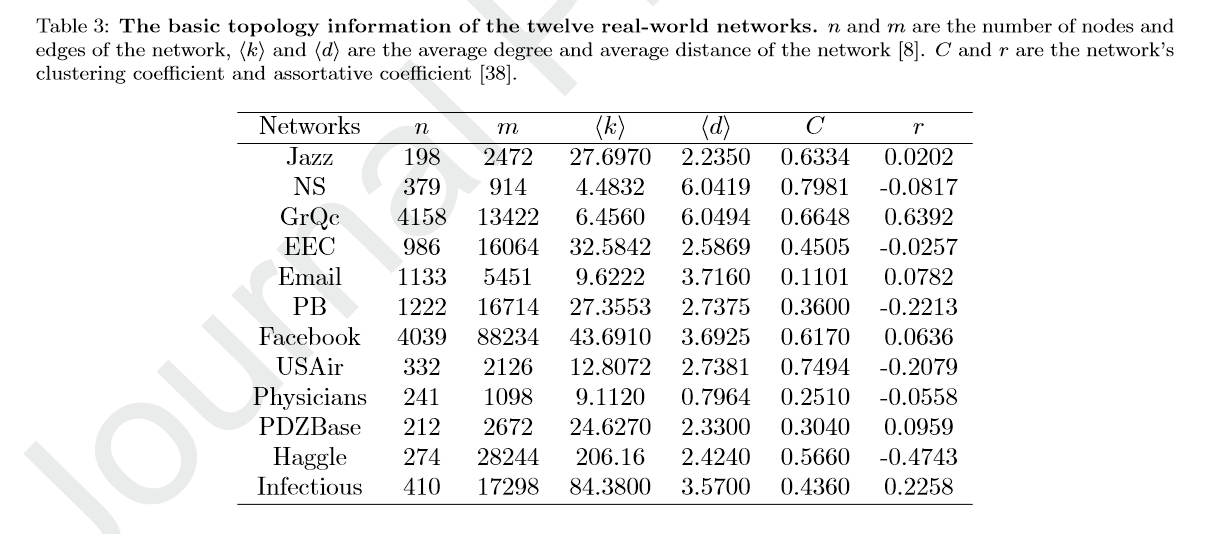


### 1.3 实证分析

为了验证方法可行性和有效性，在12个网络上进行了5次实验，与现有8种方法进行了比较，包括DC、CC、EC、PC、BC、Gravity、准拉普拉斯中心度（QL）和基于资源分配的迭代方法（Ira）。

1. **数据来源**

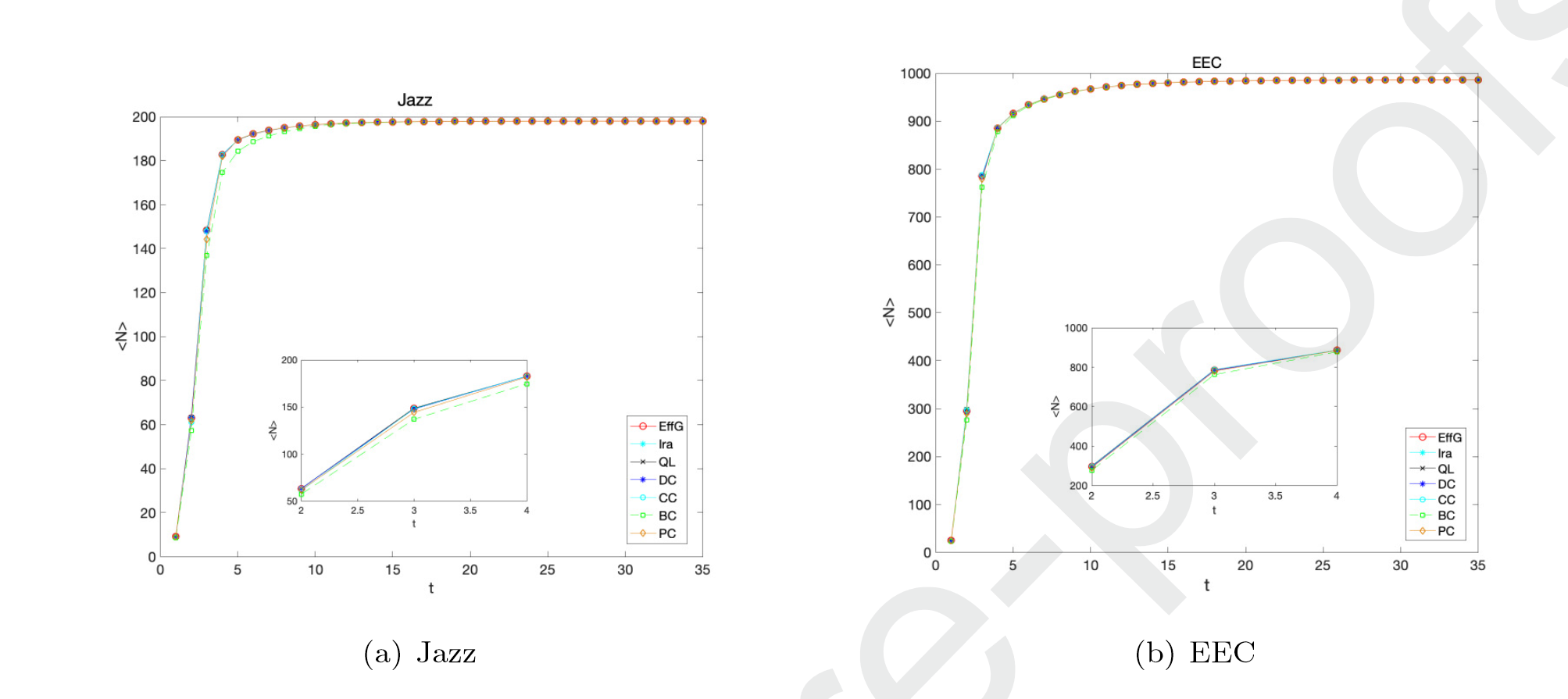
12个网络上分别是Jazz、NS、GrQc、Email、EEC、Facebook、PB、PDZBase、Haggle、Infectious、Physicians和USAir，包括两个通信网络（Email、EEC），一个交通网络（USAir），两个社交网络（Facebook、PB），三个合作网络（Jazz、NS、GrQc、Physicians），一个蛋白质-蛋白质交互网络（PDZBase）和一个无线移动设备网络（Haggle）。网络相关信息显示在表3中：

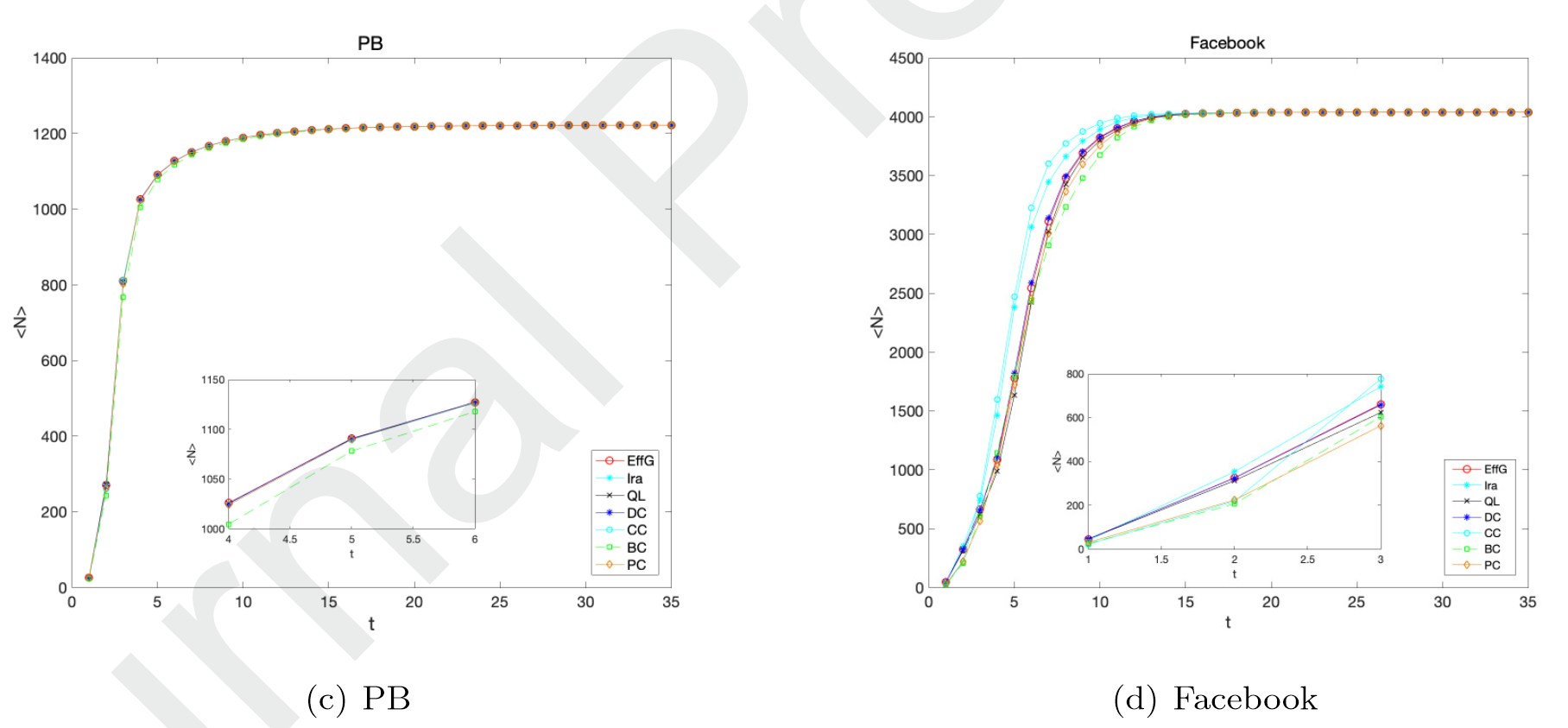


1. **用SI模型进行评估**

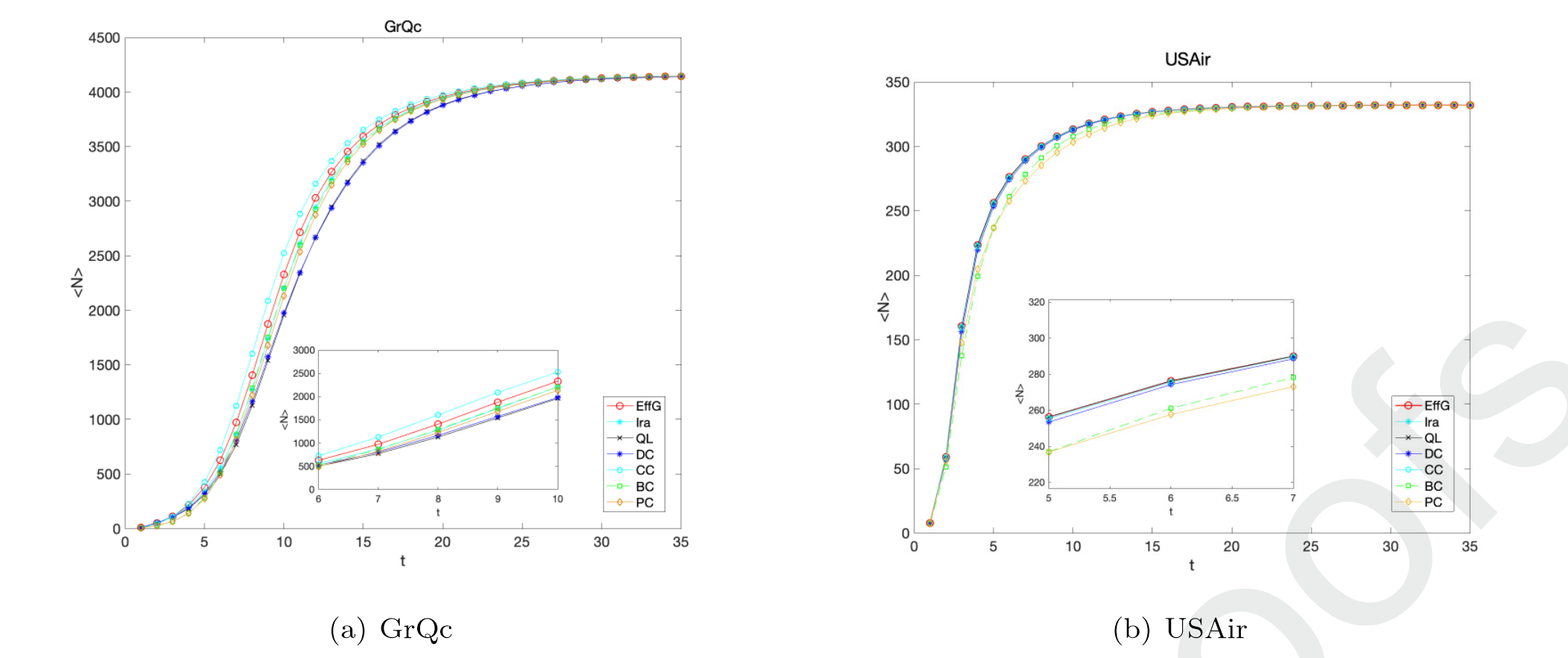
实验选择用不同方法排列的前100个节点，分别作为SI模型的初始感染节点。最后，分别计算每种方法的平均感染节点数F（t）。在本文的实验中传播概率β被设定为0.2。

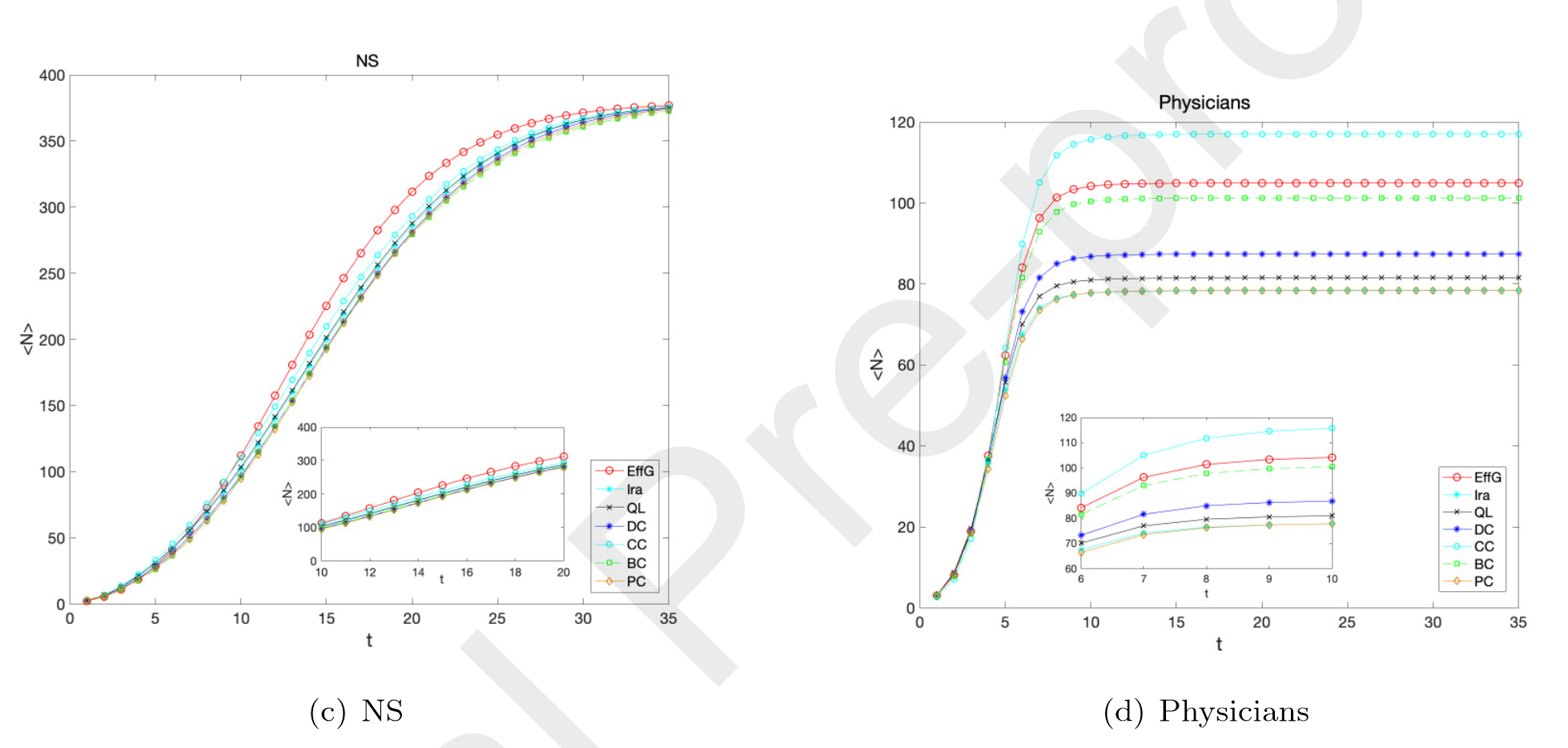
在图5（a）-5（c）中，这些方法之间的差异并不明显，这意味着它们基本上是一致的。BC的曲线总是最低的，而QL、Ira和DC对应的曲线通常在中间。

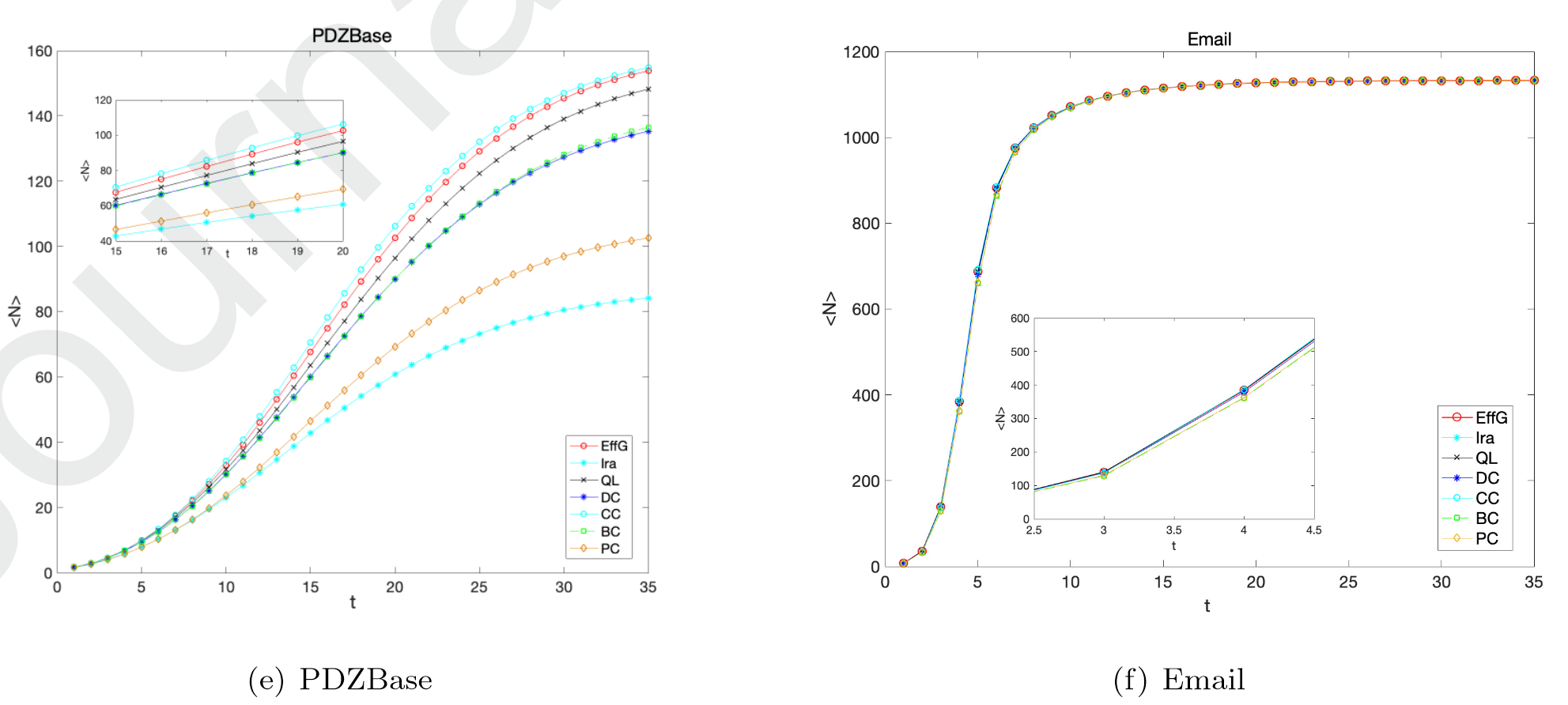




在图6（a）-6（b）中，CC的曲线总是最高，而EffG的曲线是第二高。在图6（c）-6（e）中，所有方法的表现都明显不同。与EffG和CC相对应的曲线较高，上升速度也比其他的快。这意味着CC和EffG选择的前100个节点的影响力更大。此外，如图6（c）所示，NS网络中EffG的曲线最高，上升速度也最快。综上所述，可以看出，本文提出的方法EffG和CC对应的曲线总是处于最高或第二高的位置。此外，在上述所有网络中，它们所对应的曲线的斜率也非常大，这意味着二者所选择的初始节点集具有更强的感染能力。换句话说，**CC和本文提出的方法EffG可以更准确地选择有影响力的节点**。然而，在上述大多数网络中，PC和BC所对应的曲线上升得比其他的要慢。



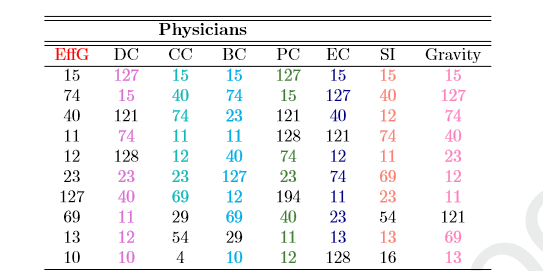




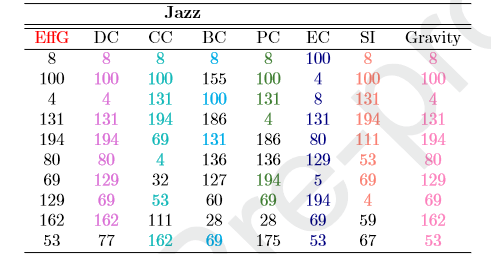
1. **比较排名结果相似性**

表4列出了用不同方法，包括EffG、DC、BC、CC、PC、EC、Gravity和β=0、t=20的SI模型，对JAZZ等网络中前10个重要节点的排名。

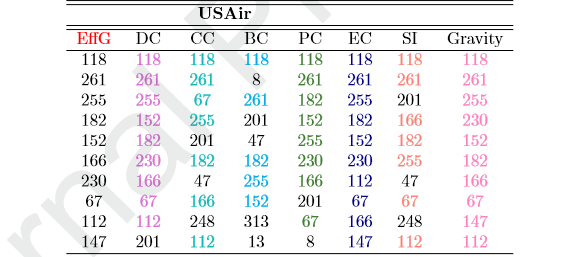
在Physicians网络中，与EffG最相似的列表是Gravity和BC，因为它们之间一致的节点数量达到了9个。其他方法和EffG之间一致的节点数为7-8，这也是相对较高的，但低于BC和Gravity。



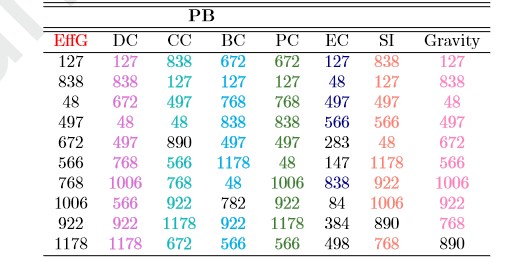
在 JAZZ网络中，与EffG最相似的列表是Gravity和EC，因为它们选择的前10个节点是一样的。然而，BC的节点与EffG最不一致，这与Physicians网络的情况不同。



在USAir中，与EffG最一致的列表仍然是EC和Gravity，最不一致仍然是BC，类似Jazz网络。在CC和SI模型中，与EffG一致的节点数量减少，与PC的节点数量增加了。此外，可以看出，所有的方法都成功地将118号节点确定为USAir中最具影响力的节点。



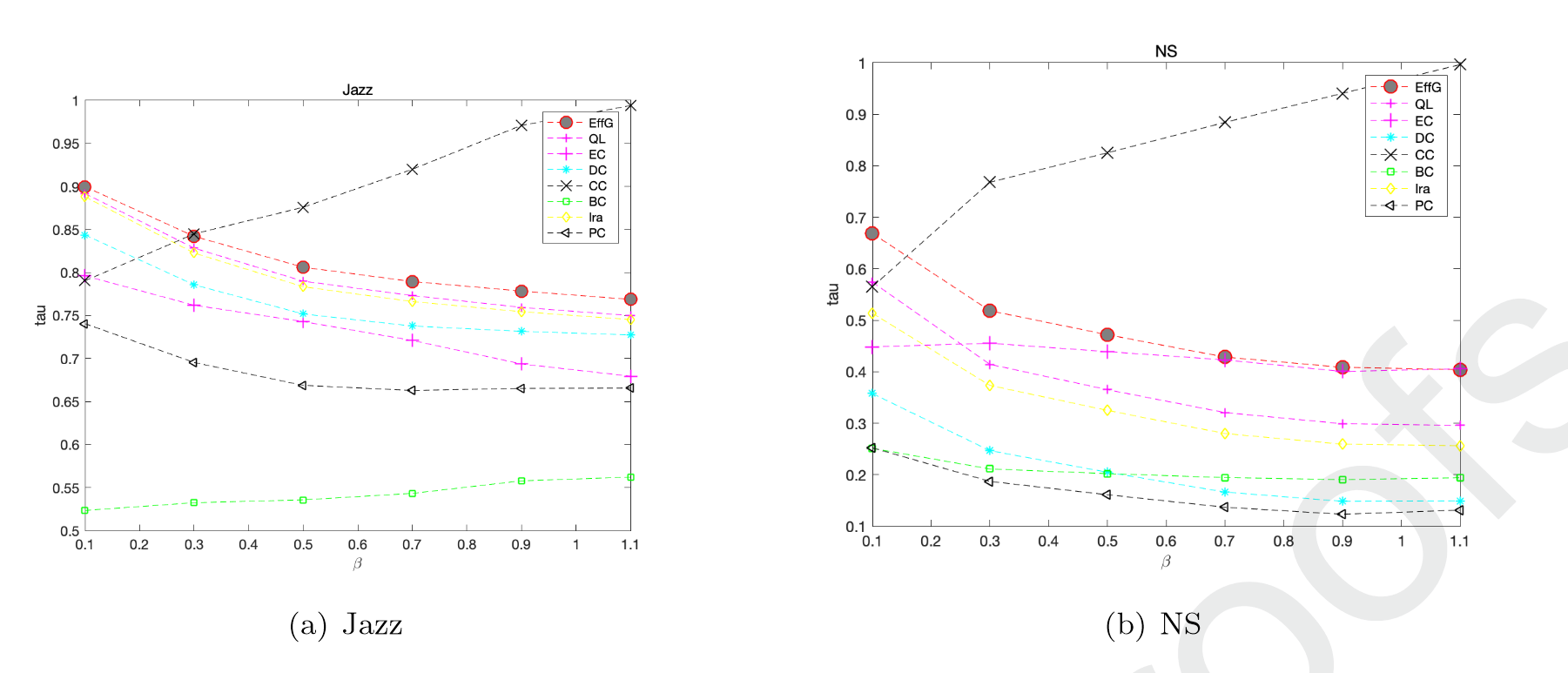
观察PB网络的结果，与EffG最一致的列表是DC和PC。其次CC、BC、SI模型和Gravity都有9个与EffG相同的节点，而EC只有5个。总之，在不同的网络中，EffG在不同网络中与不同方法有接近的表现。这一现象反映了EffG既具有全局和局部属性，也具有静态和动态信息，因为它与不同类型的方法在不同的网络上是一致的。这一现象也表明，EffG能更好地适应不同的网络。此外，在表4所示的4个真实世界网络中，EffG和其他方法的一致节点数量很高。与其他方法的高吻合度证实了本文提出的方法的合理性。



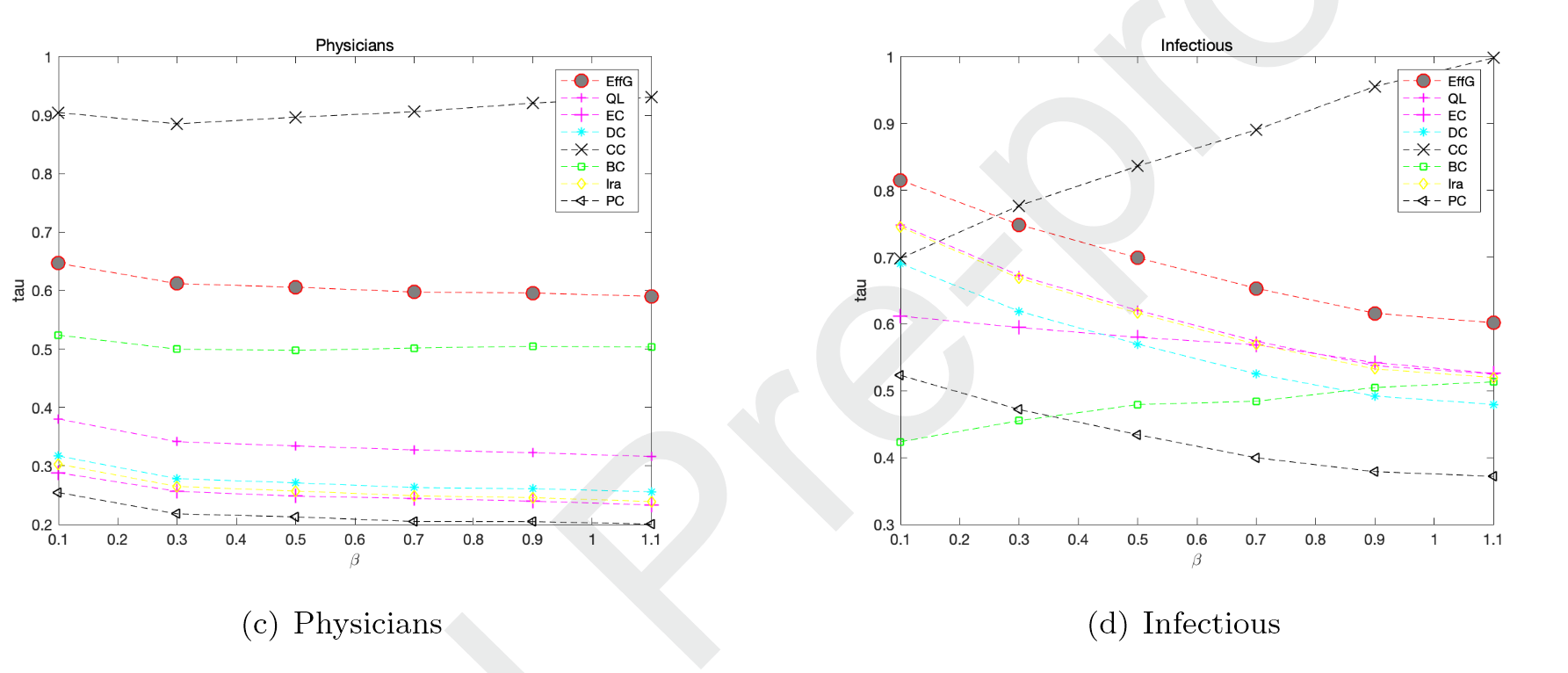
1. **本文提出方法与SI模型的关系**

两个序列X和Y的Kendall相关系数定义如下：

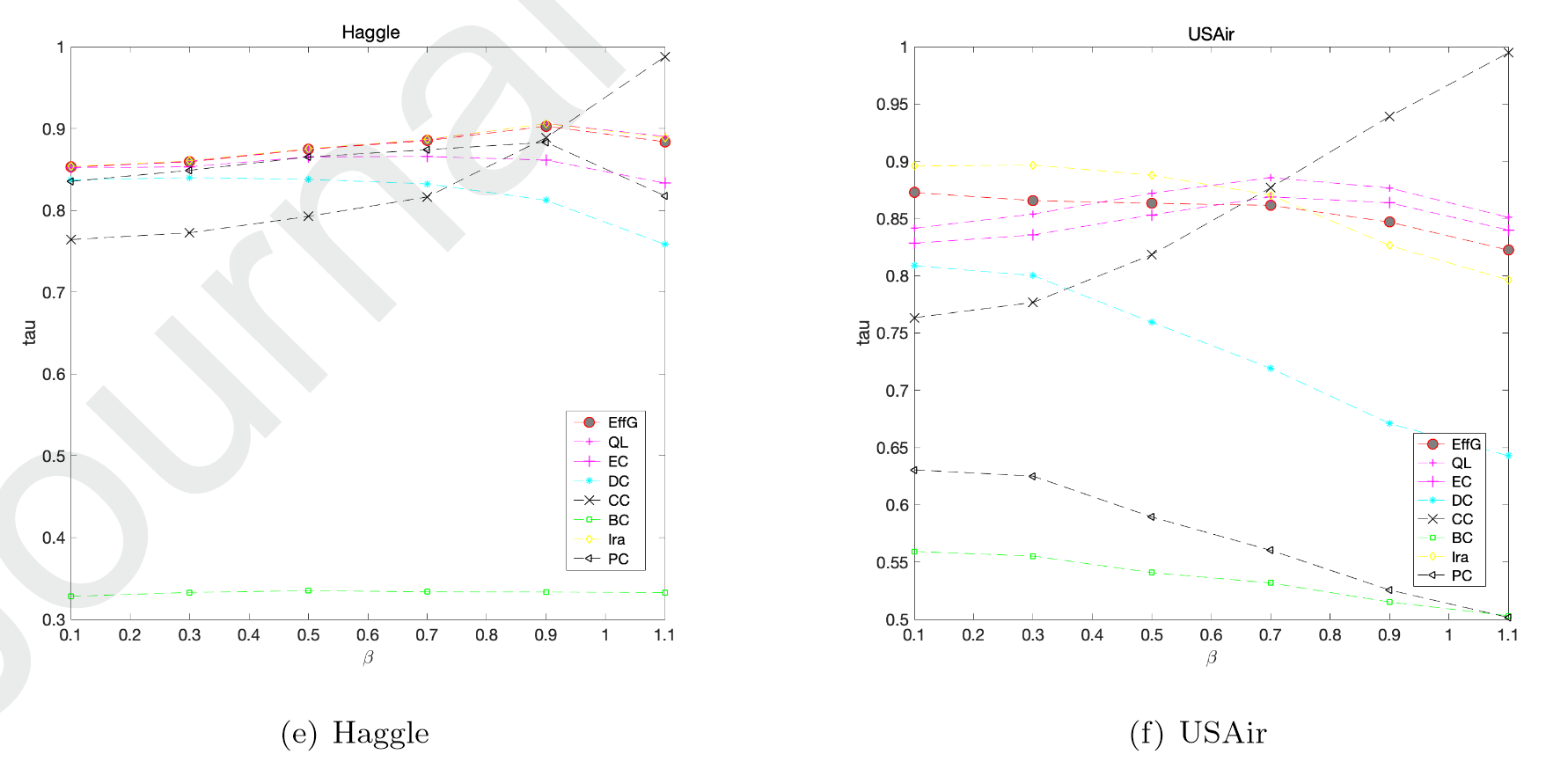
在本实验中，对该方法有效性的评价是基于与SI模型的相关性。在所有数据集中，分别赋予SI模型以不同感染概率β，获得一个标准中心性序列。然后计算SI模型序列和其他方法得出序列的Kendall相关系数。在实验中，感染概率β从0.1变化到1.1，SI模型独立运行120次，感染时间t=20，取不同网络的平均值。实验结果如图7所示，其中tau代表Kendall相关系数的值。tau值越大，表明中心性方法和SI模型之间的正相关越强。



在图7（a）-7（d）中，EffG对应的曲线总是在第二高的位置，CC的曲线是最高的。然而，BC和PC的曲线比其他方法低。



在图7（d）中，随着β的增加，除了CC以外，所有的曲线都呈现出下降的趋势。

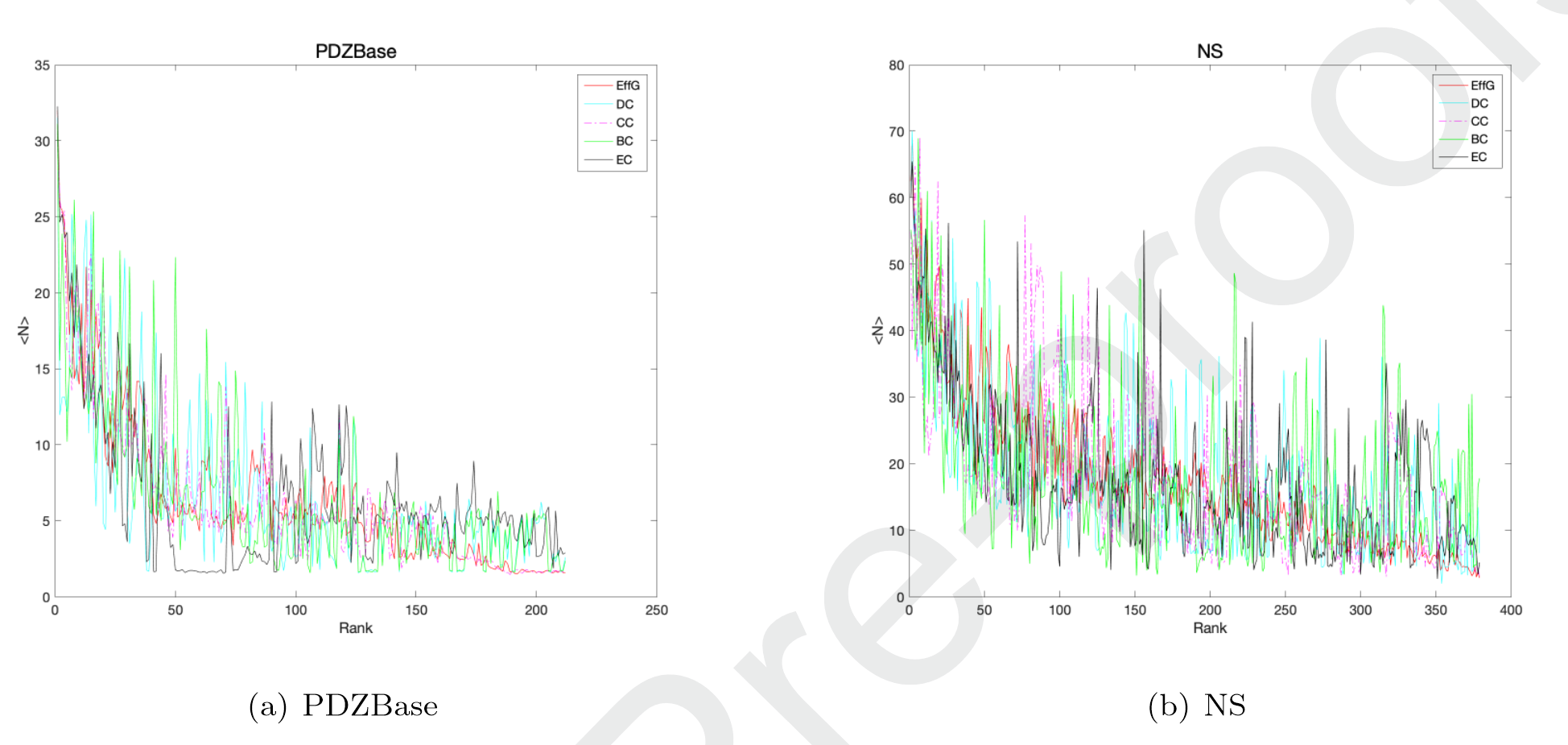


从图7（e）中可以看出，在Haggle网络中，当0.1 < β < 0.9时，EffG的曲线最高，而BC的曲线最低。此外，其他方法的Kendall相关系数也非常接近。如图7（f）所示，当0.7 < β < 1.1时，USAir的Kendall相关系数最小，而CC的相关系数最大。总之，在大多数情况下，CC和SI序列之间的关联性是最强的。SI序列和EffG之间的相关性至少总是第二强的，甚至在某些情况下是最强的。这意味着EffG与SI模型的相关性比大多数现有的方法更强，这证明了其有效性和优越性。

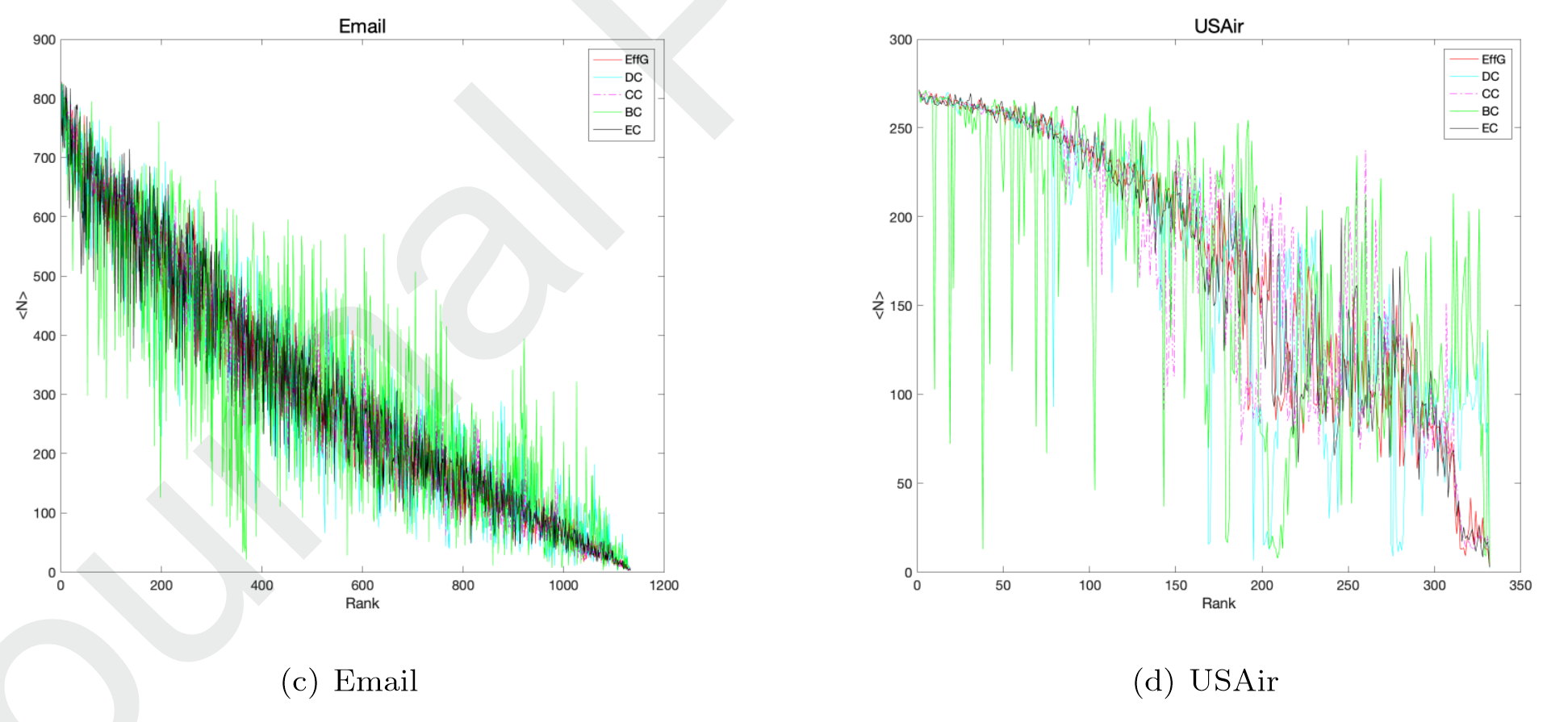
1. **比较所提出的方法和SI模型之间的相似性**

在这个实验中，使用PDZBase、NS、Email和USAir网络。首先，通过DC、BC、CC、EC和EffG得出每个网络节点排名。每个节点将被用作SI模型中的初始感染节点，最终的感染节点数量将在t=15的条件下计算。最后，建立节点排名与它们在t时刻的平均感染节点数之间的相关性，结果如图8所示。

Rank代表节点的排名。排名越高的节点具有更强的感染其他节点的能力，意味着该节点的影响力更大。从图8（a）可以看出，BC和DC的曲线有明显的波动，而它们的下降趋势不是很明显。EffG对应的曲线波动不大，保持持续下降的态势。如图8(b)所示，除EffG外，所有方法的波动都非常明显，而NS的下降趋势也不明显。

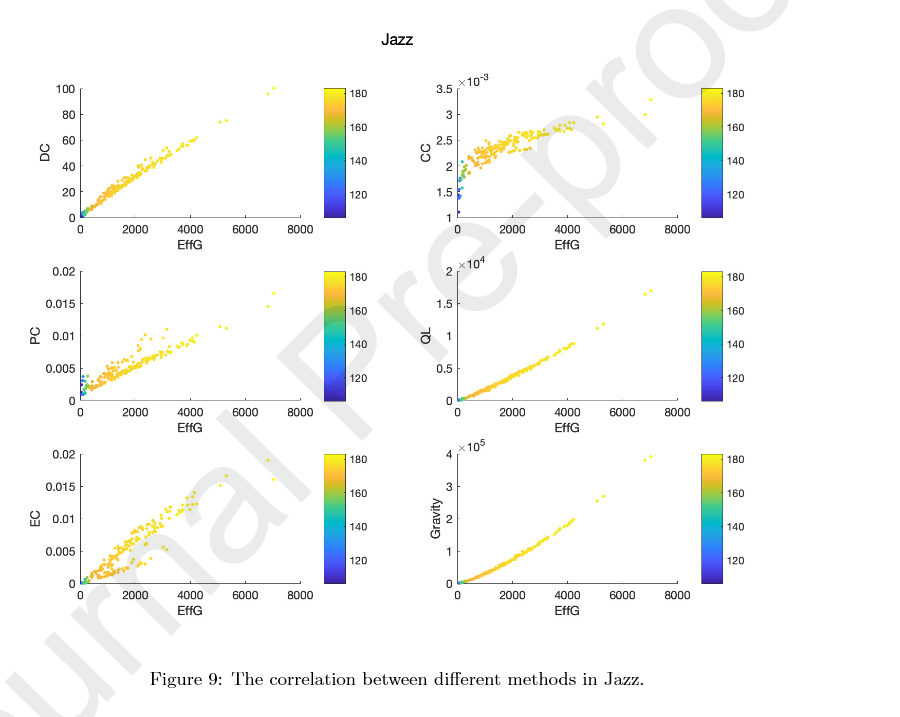


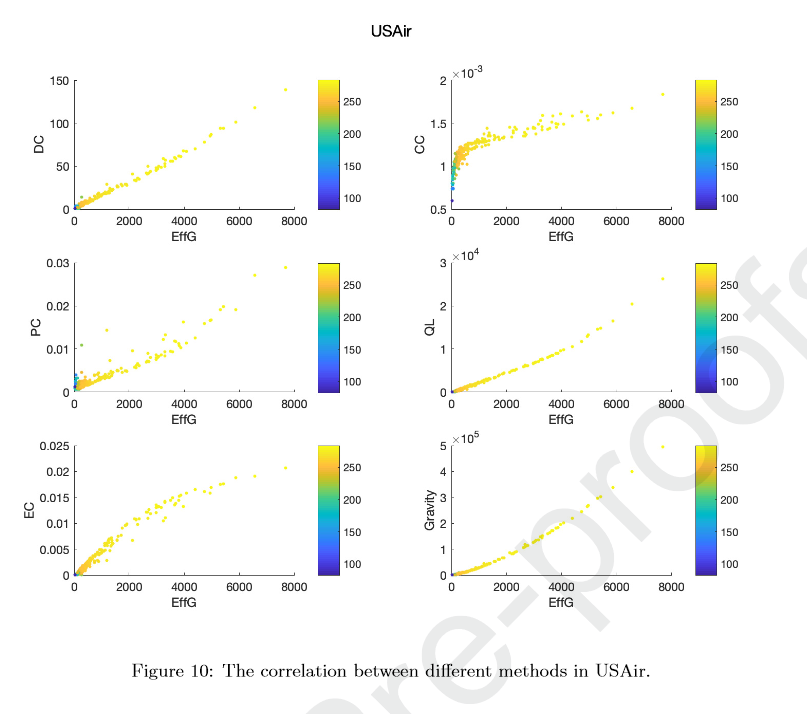
在Email网络中，所有方法的曲线都呈现明显的下降趋势。如图8(d)所示，BC对应的曲线几乎没有下降趋势，波动很大。其他方法则显示出明显的持续下降趋势。综上所述，本文提出的方法EffG对应的曲线是持续下降的，与其他方法相比波动很小。此外，可以很容易地验证，对应于BC的曲线在下降过程中波动很大，没有明显的下降趋势。因此，可以推断，本文提出的EffG方法与其他方法相比，在一定程度上是有效和合理的。

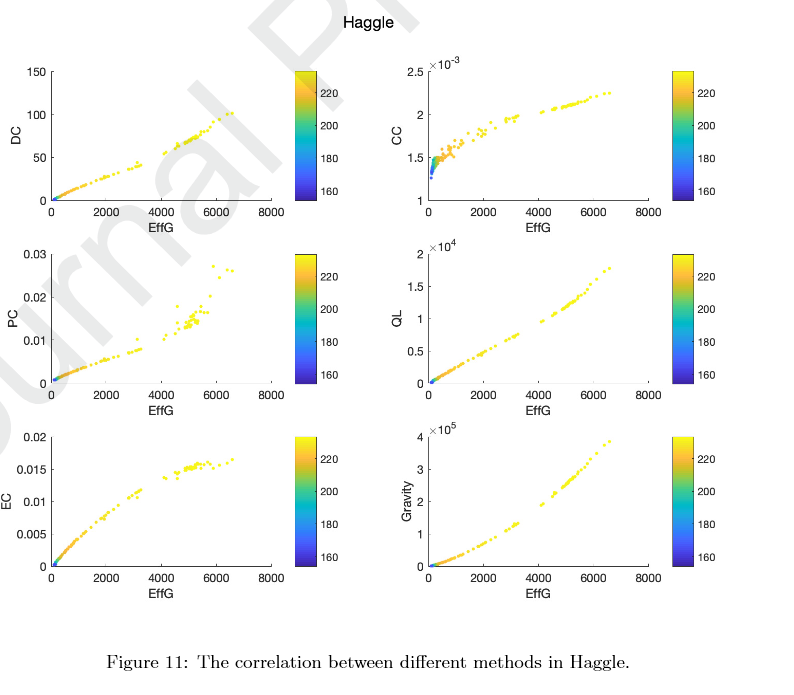


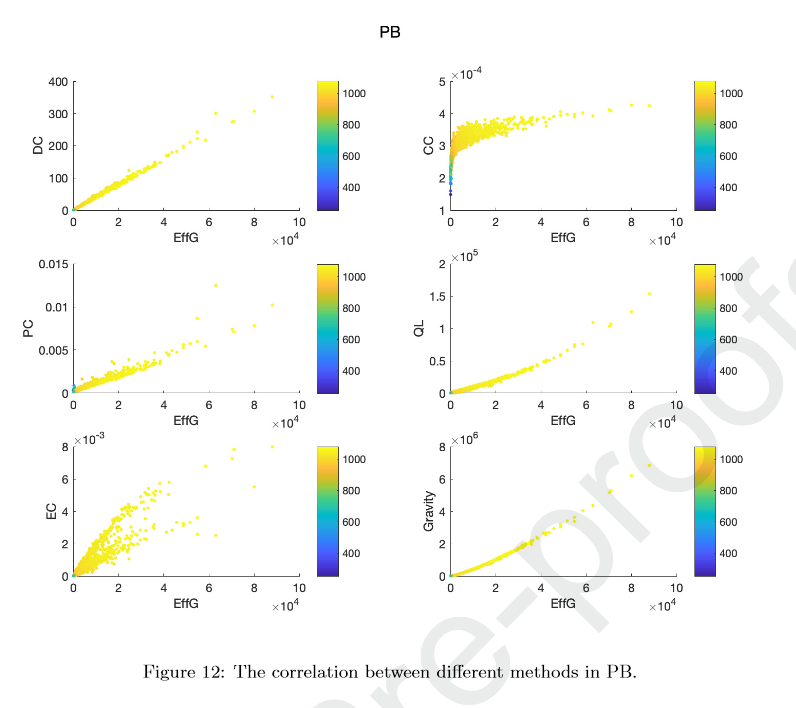
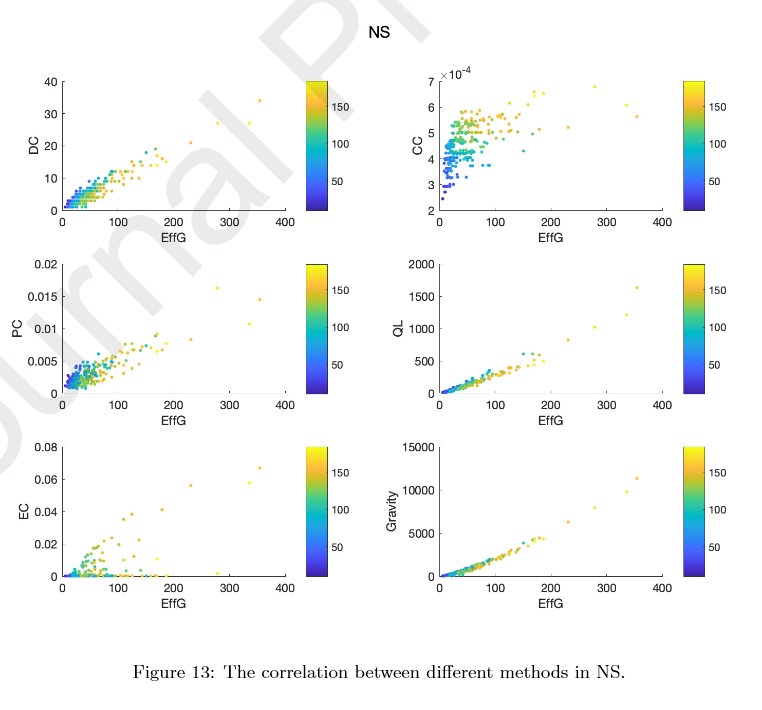
1. **不同方法之间的关系**

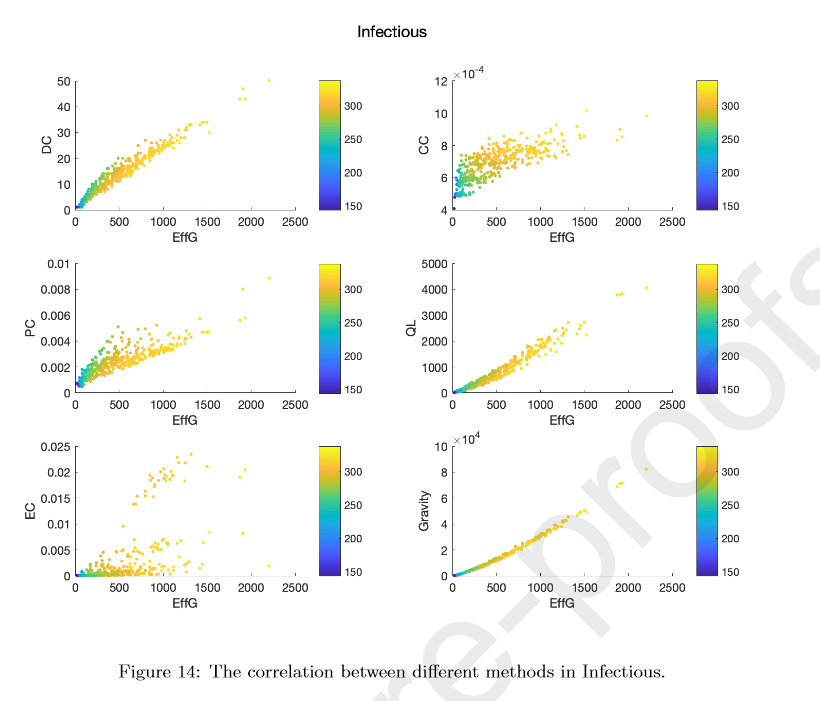
在本实验中，通过比较不同方法得到的中心性分数，得到不同方法之间的关系。图中的坐标轴代表不同的中心性分数。在本实验中使用的SI模型中，感染时间t=20，实验次数K=120，感染率β=0.2。图9-14显示了EffG和其他方法之间的关系。在图9-14中，EffG和其他中心性方法显示出很强的正相关关系。如图13-14所示，与NS网络和感染性网络中的EC相比，本文的方法可以很好地避免许多节点由于中心性得分相同而被捆绑的情况，这对实际应用非常有利。综上所述，所提出的EffG能够有效地识别复杂网络中的重要节点。此外，与其他一些中心性方法相比，它可以用现有的一些方法得到的接近值来识别节点的重要性。











### 1.4 结论

本文提出了一种基于有效距离引力模型的识别有影响力节点的原创性的新方法EffG。在单维因素的基础上，本文提出的EffG模型还综合考虑了节点的局部信息（节点的度）和基于多源信息融合思想的网络的全局信息。一个重要的贡献是，**EffG模型使用有效距离的概念来取代传统的静态欧氏距离**。EffG能够充分利用网络中节点之间的动态信息交流。此外，EffG能够帮助解开网络中隐藏的拓扑结构，这种拓扑结构推动了许多动态信息传播过程。很重要的一点是，EffG对有影响力的节点的识别是与现实世界的条件相一致的。

为了验证该方法的有效性和可行性，本文在12个真实世界的网络上进行了各种实验，并将EffG与现有的8种知名方法进行了比较。实验结果表明，本文的方法在动态信息传播下和多个测试实例中表现良好，从而证明了其在网络科学、生物和社会系统、时间序列和信息传播中的潜在应用。尽管EffG在有影响力的节点识别方面取得了巨大的成功，但作者团队未来工作的动力是EffG是否能被改进以降低其时间复杂度。如此，该方法可以适应大规模网络的使用。