

订阅DeepL Pro以编辑此演示文稿。  
访问[www.DeepL.com/pro](https://www.deepl.com/pro?cta=edit-document)，了解更多信息。

[Physica A 433 (2015) 344-355](http://dx.doi.org/10.1016/j.physa.2015.03.042)

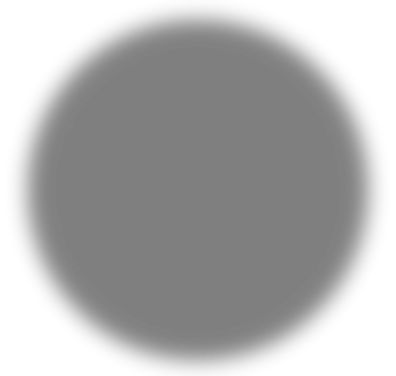


内容列表可在[ScienceDirect上](http://www.elsevier.com/locate/physa)找到

物理学A

杂志主页: [www.elsevier.com/locate/physa](http://www.elsevier.com/locate/physa)

利用全局多样性和局部拓扑结构特征来识别有影响力的网络传播者



傅宇翔 [a](#_bookmark0), 黄中元 [b](#_bookmark1),⇤, 孙传才 [a](#_bookmark0)

a *国立交通大学计算机科学系，新竹市塔学路1001号，300。*

b *长庚大学工学院电子与计算机工程系计算机科学与信息工程专业，桃园市文华一路259号，333。*

兴发娱乐官网手机版市人民政府

* 我们结合全局多样性和局部特征来识别有影响力的节点。
* 提出了一个稳健可靠的两步框架作为节点排名措施。
* 一系列的实验结果表明，我们的方法表现良好且稳定。

我的朋友们，你们好吗？

*文章的历史。*

2014年6月3日收到的

2015年1月30日收到修订版，2015年4月16日上网。

*关键词。*

节点多样性 熵

社会网络分析 *K*-shell分解 SIR流行病模型

A B S T R A C T

识别传播思想、信息或传染病的最有影响力的个人是一个受到网络研究者极大关注的话题，因为这种识别可以帮助或阻碍信息传播、产品曝光和传染病检测。枢纽节点、高间隔度节点、高接近度节点和高*K*壳节点已被确定为良好的初始传播者。然而，很少有人利用网络结构中的节点多样性来衡量传播能力。本文描述的两步框架使用了一个稳健可靠的测量方法，结合全局多样性和局部特征来识别最有影响力的网络节点。一系列易感-感染-恢复（SIR）流行病模拟的结果表明，我们提出的方法在与各种复杂网络数据集相关的单一初始传播者情况下表现良好且稳定。

© 2015年作者。由Elsevier B.V.出版。这是一篇开放性的文章，在CC

BY-NC-ND许可证（[http://creativecommons. org/licenses/by-nc-nd/4.0/](http://creativecommons.org/licenses/by-nc-nd/4.0/)）。

# 简介

网络传播现象是研究的重点，从通过在线社会媒体网站的信息传播，到病毒式营销，到流行病的识别和控制，再到电网和互联网的级联故障，等等[[1-9](#_bookmark12)]。目前正在建立和测试识别关键传播者的策略，以加速信息传播，增加产品曝光，检测传染病爆发，并执行早期干预策略[[10](#_bookmark19)]。拓扑结构是这个网络传播识别过程中的一个核心概念[[1](#_bookmark12),[2](#_bookmark13)]。

在社会网络分析中，用于识别有影响力的网络节点的中心性措施大致分为局部和全局[[3](#_bookmark14),[7](#_bookmark17),[11](#_bookmark20)]。度数中心性，定义为一个焦点节点所连接的节点数量，衡量节点在网络中的参与度。然而，大多数研究者所青睐的测量网络节点影响力的技术没有考虑全局拓扑结构的重要性。两种最广泛使用的全局中心度量方法用于

\* 通讯作者。电话。+886 3 211 8800x3474；传真。+886 3 211 8700.

*电子邮件地址：*[gscott@mail.cgu.edu.tw](mailto:gscott@mail.cgu.edu.tw)（C.-Y. Huang）。

<http://dx.doi.org/10.1016/j.physa.2015.03.042>

0378-4371/© 2015年作者。由Elsevier B.V.出版。这是一篇根据CC BY-NC-ND许可证[（http://creativecommons. org/ licenses/by-nc-nd/4.0/](http://creativecommons.org/licenses/by-nc-nd/4.0/)）公开获取的文章。

克服这些限制的方法是间隔性和紧密性。间性中心性评估的是一个节点位于其他两个节点之间最短路径上的程度，它决定了网络流量。紧密性中心性被定义为从一个焦点节点到所有其他节点的最短距离的倒数之和。影响力与对优势网络位置的占有有关。优势的三个基本来源是高程度、高接近度和高间隔度。在简单的网络结构中，这些优势往往是单独变化的。在复杂的网络中，这些位置特征之间可能存在明显的不一致，这意味着一个传播者的位置可能在某些方面是有利的，而在其他方面是不利的。

除了中心度量之外，*k*-shell分解分析的结果表明，位于核心层的网络节点比位于外围层的节点能够传播到更广泛的区域[[1](#_bookmark12),[2](#_bookmark13)]。虽然每个节点的传播能力不同，但那些具有相似*k*-shell值的节点被认为具有同等重要性。有人提出了一种方法，以相同的*k*壳层中的度中心性对节点的网络传播能力进行排序，以达到调整排名列表的目的[[8](#_bookmark18)]。为了对传播者进行排名，一种被称为混合度分解（MDD）的方法在分解过程中加入了原本被忽略的度数节点[[3](#_bookmark14),[6](#_bookmark16),[12](#_bookmark21)]。然而，研究人员仍然表现出忽视网络拓扑结构和节点多样性的重要性的倾向，尽管它们与社区经济发展等因素存在正相关关系[[13](#_bookmark22)]。此外，用户访问的地点的熵值与这些用户在社会网络中的社会关系的数量呈正相关[[14](#_bookmark23)]。节点度、间隔度和接近度中心性的综合熵值已被用于创建复杂的网络可视化[[15](#_bookmark24),[16](#_bookmark25)]。

受过去对网络拓扑结构和节点多样性研究的启发，我们利用熵的概念开发了一种稳健可靠的方法来测量节点的传播能力，并识别复杂网络中的超级传播者节点。这种测量方法可以用来分析受特定个体节点影响的全局网络拓扑层和局部邻域节点的数量。我们的假设是，*K*壳分解[[1](#_bookmark12),[2](#_bookmark13)]可用于全局分析的目的，具有高程度的全局多样性和局部中心性的节点能够穿透多个全局层并影响复杂网络局部层中的大量邻居。为了测量节点的影响力，我们提出了一个两步走的框架，用于获取复杂网络中的全球和本地节点信息。在第一步中，使用诸如复杂网络的社区检测算法[[5](#_bookmark15),[17](#_bookmark26),[18](#_bookmark27)]或核心/外围网络层的*K*壳分解算法来获得全局节点信息，之后用熵来评估网络节点的全局多样性。第二步，通过使用各种类型的局部中心性（包括度中心性）来获得局部节点信息。最后，全局多样性和局部特征被结合起来以确定节点的影响力。在我们的实验中，传播能力是以一段时间内恢复的节点总数来衡量的。我们使用SIR（易感-感染-恢复）流行病模拟[[2](#_bookmark13),[19](#_bookmark28),[20](#_bookmark29)]与各种社会网络数据集[[21-25](#_bookmark30)]比较了所提措施的传播能力和社会网络的本地/全球中心性。

# 背景介绍

为了表示一个复杂的社会网络，让一个无向图*G*=*（V，E）*，其中*V*是网络的节点集，*E*是网络的边集。让*n* = |*V* |表示网络节点的n um}ber，*m* = |*E*|表示边的数量。网络结构为

表示为一个邻接矩阵*A*=

*a ij*= 0.

*aij*

和*a ij*2 *Rn* ，其中*a ij*= 1，如果节点*i*和*j*之间存在链接，否则

的

度（或局部）中心性是衡量复杂网络中节点影响力的一种简单而有效的方法。让*Cd（i）表示节点i*的度中心性。高度中心性表示一个节点与其邻居之间有大量的联系。*NBh（i）表示节点i*在*h*跳距离上的邻居集合。因此，节点*i*的度中心性被定义为

*n*

X

*Cd(i)* = |*NB h(i)*| = *a ij* (1)

*j*=1

其中，*NBh（i）*是节点*i*在*h*-hop距离上的邻居数量；在大多数情况下，h1[[7](#_bookmark17)]。

|| =

间性中心度或依赖性衡量的是复杂网络中经过某个节点的最短路径的比例。让*C(bi)表示节点i*的间性中心度。间性值高，表明复杂网络节点位于重要的通信路径上。因此，节点*i*的介于度中心性定义为

*C b(i)* = X

*v2V*

s6=*t*=6

*Q st(i)*

*ǞǞǞ*

(2)

其中，*Qst（i）*是指从节点*s*到节点*t*经过节点*i的*最短路径数，*Qst*是指从节点*s*到节点*t*的最短路径总数[[3](#_bookmark14),[7](#_bookmark17),[11](#_bookmark20)]。

紧密性（也称为全局）中心性衡量一个节点到其他节点的最短路径的平均长度。让*C(li)*表示节点*i*的紧密性中心度。高的紧密性中心度值表示一个节点位于



**图1.** SIR（易感者-感染者-恢复者）流行病模拟模型。

复杂网络的中心，而且该节点到其他节点的平均距离比亲近度中心度低的节点要短。节点*i*的紧密性中心性定义为

*n*

X

*C l(i)* = 1 *, l i*

*i*

1

= *n* -

*d ij* (3)

*j*=1

其中*li*是节点*i*到其他节点的最短路径的平均长度，*dij*是节点*i*到节点的距离

*j* [[11](#_bookmark20)].

*k*-shell分解法[[1](#_bookmark12),[2](#_bookmark13)]为复杂网络中的每个节点迭代分配一个*k*-shell层值。在第一步中，让*k* 1并移除所有*C d(n) k 1的节点*。在移除之后，剩下的一些网络节点的度数可能是*k* 1。节点被不断地从网络中剪除，直到没有*k* 1的节点。所有被移除的节点都被分配一个*ks 1*的*k*壳值。下一步需要一个类似的过程：让*k 2*，修剪节点，并给所有被移除的节点分配一个*k*-shell值为2。这个过程重复进行，直到所有的网络节点都被删除并分配了*k*-shell索引。该方法揭示了复杂网络的重要特征--例如，所有的互联网节点都可以被划分为核心、对等连接的组件或孤立的组件[[1](#_bookmark12)]。

= =

= =

== =

*K*-shell指数和*K*-核心层的概念有细微差别。*k*-shell指数是一个全局指标，代表一个节点所处的网络核心层。较高的*k*-shell指数代表内部核心层节点比外围层节点更重要。*k*-核心层是一个由具有*ks k*-shell指数超过给定值*k的*节点组成的子网络[[1](#_bookmark12),[2](#_bookmark13)]。高*k*壳指数的节点比具有相同程度中心性值的节点能够感染更多的邻居[[2](#_bookmark13),[6](#_bookmark16),[8](#_bookmark18)]。

根据上述的全局衡量标准，节点网络位置是通过分析整个网络结构和相邻节点网络位置的相对关系来确定。节点网络位置信息是通过计算所有对的最短路径的间性中心度和所有最短路径的平均长度的接近性中心度来获得。从外围层到核心层的*K*壳网络结构分解表明涉及整个网络结构的全局属性；相反，''局部''指标涉及节点信息，如连接数。

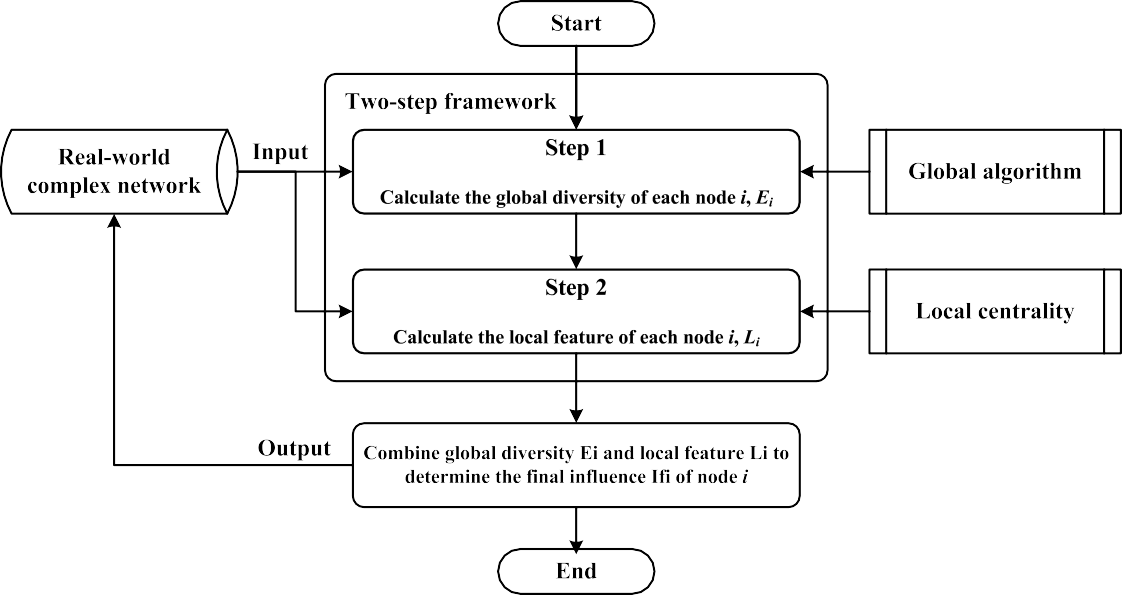
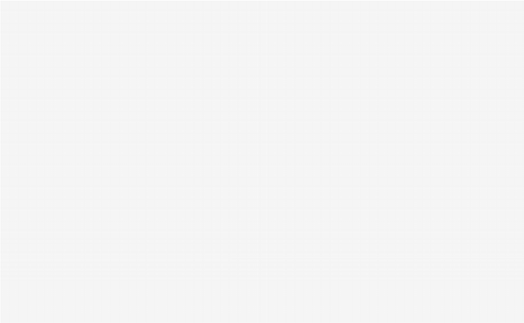
[图1](#_bookmark3)所示的SIR流行病模型[[2](#_bookmark13),[19](#_bookmark28),[20](#_bookmark29)]已被广泛用于多个领域，研究信息、谣言、生物疾病和其他现象在人群中的理论传播过程。感染性''概念是上述传播现象的一般属性。用流行病学的行话来说，思想和谣言可以表示为疾病病原体，导致受感染的个体在更大的人群中传播疾病。个人在康复或接受疫苗后可能有病原体抗体，因此变得无法再感染。由于这一特点，SIR流行病模型被广泛用作描述疾病状态和研究传染性传播过程的通用模型。关于人口的社会网络结构，SIR模型可以应用于网络传播动力学，疾病传播是一般网络传播现象的例子。SIR模型已经被修改和扩展（如SEIR[[26](#_bookmark32)]和SIHR[[27](#_bookmark33)]），以研究不同疾病和谣言在网络中的传播动态[[28](#_bookmark34)]。

SIR模型由三种状态组成：易感（*S*）、感染（*I*）和恢复（*R*）。*S*集的节点容易受到信息或疾病的影响，*I*集的节点能够感染邻居，而*R*集的节点具有免疫力，不能被重新感染。最初，几乎所有的网络节点都在*S*集，少数被感染的节点（有时是一个人）充当传播者。在每个时间步骤中，*I*节点以预先确定的感染率*β*感染它们的邻居，之后它们以恢复率成为恢复节点。 让*S(t)*表示时间*t*的易感节点的数量，*I(t)*表示

时间*t*的感染节点数，*R（t）*是时间*t*的恢复节点数，*⇢（t）*=*R（t）/N*是免疫节点的比例。SIR模型中的节点总数为*S（t）*+*I（t）*+*R（t）*=*n*。

# 拟议的措施

有影响力的社会网络传播者应该满足两个网络拓扑条件：高全局多样性和高局部特征。首先，节点的全局多样性是由邻居的网络位置决定的。邻居位置信息的更大差异增加了节点的多样性--也就是说，邻居在所有网络层（社区）的分布是平等的，否则他们的网络位置是相似的。有一个预期，高全局多样性的节点在传播过程的早期阶段会非常迅速地传播信息、想法或谣言。第二，节点的本地特征是用邻居连接的总和来衡量的。这里的期望是，高局部特征的节点将在大量的候选节点中引发早期和快速积累的传染性传播。



**图2.** 拟议的计算网络节点影响力的两步框架。

[图2](#_bookmark4)所示的两步框架是我们提出的在复杂网络中获得全局和局部节点信息的方法。在步骤1中，全局算法（如社区检测、图聚类、*K*壳分解）被用来分析复杂网络中节点的全局特征。结果被用来计算节点的全球多样性。在第2步中，度中心性被用来测量局部节点特征。最后，全局多样性和局部特征被结合起来，以确定复杂网络节点的最终影响力。

在步骤 1 中，以 *k*-shell 分解法为例，获取复杂网络中节点的全局信息。获得节点的*k*-shell值后，用香农熵[[29](#_bookmark35)]计算全局多样性，然后用它来描述一个节点影响多少网络层。根据公式[（4）](#_bookmark5)的定义，最大熵表示一个节点能够与复杂网络的所有层平等连接的情况，最小熵为0表示一个节点的所有连接都在复杂网络的同一层。[如图3](#_bookmark6)所示，节点*i*的*k*-shell熵，确保其邻居的*k*-shell值更加多样化，定义为

ksmax

X

*E (iXi* *)* = - *pi* *xj* - log2*p* *ix j* (4)

*j*=1

*p ix j xj*

ksmax

=

jP=1 *xj*

*Ei (Xi)* - Emin

*E*ˆ *X* =*( )*

iilog2 ksmax - Emin

(5)

(6)

where *Xi* = {1*,* 2*, . . . , ks*max} denotes the *k*-shell values of the neighbors of node *i, pi xj* the probability of the *xj*-core layer of neighbors, *xj* the number of nodes in the *xj*-core layer of the complex network, and *E*ˆ*i (Xi)* the normalized *k*-core entropy required for the case under consideration.

在第二步中，节点的度中心性被用来分析复杂网络中的局部特征值；邻居的度中心性也被考虑。高的影响值表明一个节点和它的邻居具有高的度中心性，反过来表明该节点能够达到最广泛的局部范围。节点*i*的局部特征被定义为

*L i*= 对数 2

X

*jeNBh*=1 *(i)*

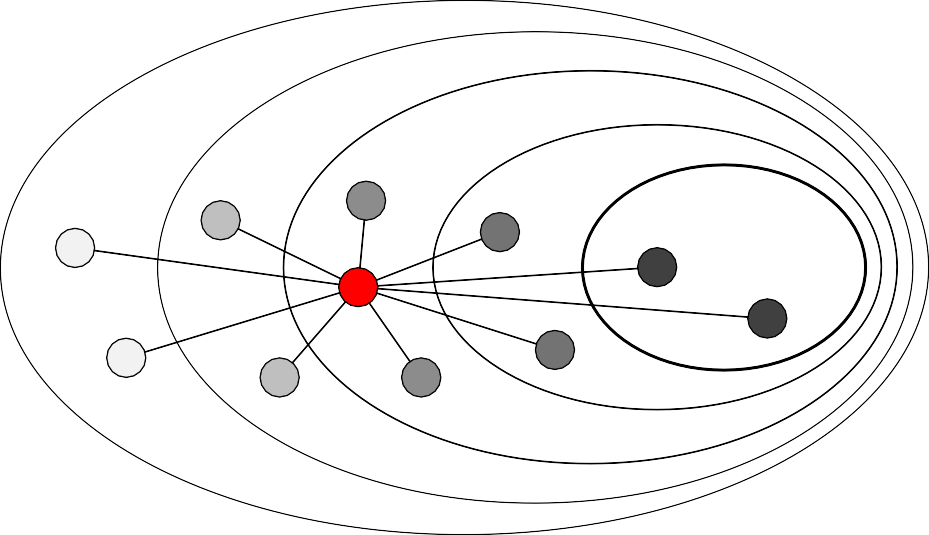
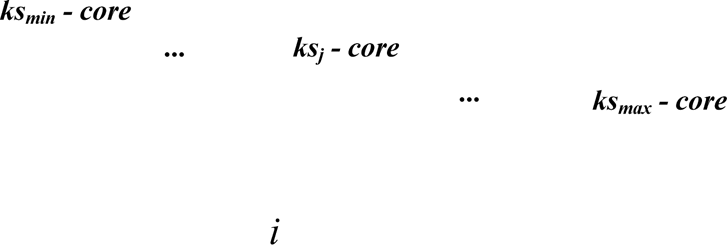
*L i*- *L* min

*L*ˆ =

iLmax - Lmin

*C d(j)*! (7)

(8)



**图3.** 用*K*壳熵来描述邻居在网络层中的分布情况的图示。

其中，*Cd（j）*是邻居*j*的度数中心性，*NBh*（1*i）*是节点*i*在*h*跳距离的邻居集合。*L (ii)*可以扩展成为''邻居的邻居''版本，即考虑节点*i*的所有2跳距离的邻居，*L*ˆ*i (i)*是所考虑情况下所需的归一化局部特征。

=

根据有影响力的传播者的定义，*Ei*和*Li*被同时考虑，以使复杂网络中的节点*i*的传播能力最大化。被识别的节点有望连接到不同网络层的枢纽节点。最后，全局多样性*Ei*和局部特征*Li*相结合，表示*IFi*，即节点*i*的最终影响力，定义为

*IF* = *iE i*- *L i。* (9)

时间和空间的复杂性对某些应用来说是很重要的；所提出的两步框架（包括合并步骤）的时间和空间复杂性在[图2](#_bookmark4)中显示。在第一步中，与计算全局多样性值*E*相关的时间复杂度为*O ( k n)*，即每个节点访问其所有的邻居以获取*k*壳值，其中*k*为平均节点度中心度，空间复杂度表示为存储节点的*E*值的*O (n)*。在第二步中，计算局部特征值*L*的时间复杂度为*O k n*2，即每个节点积累其邻居的度中心性值之和，其中空间复杂度表示为存储节点的*L*值的*O（n）*。在组合步骤中，时间复杂度为*O（n）*，意味着最终的影响*IF*值是*E*乘以*L*（或一个加法操作，如果

h i -

h i - h i

introducing the logarithm to equation [[11](#_bookmark20)]), and space complexity is denoted as the *O (n)* of storing the *IF* values of nodes. Total time complexity is expressed as *O* h*k*i2 · *n* + h*k*i · *n* + *n* = *O* h*k*i2 + h*k*i + 1 · *n* and total space complexity as

*O (n* + *n* + *n)* = *O (3n)*，通过将*E*、*L*和*IF*值存储在同一内存空间并忽略，可以减少为*O ( n)*

的合并步骤，将时间复杂度降低到*O*

# 实验结果和讨论

hki2 + hki

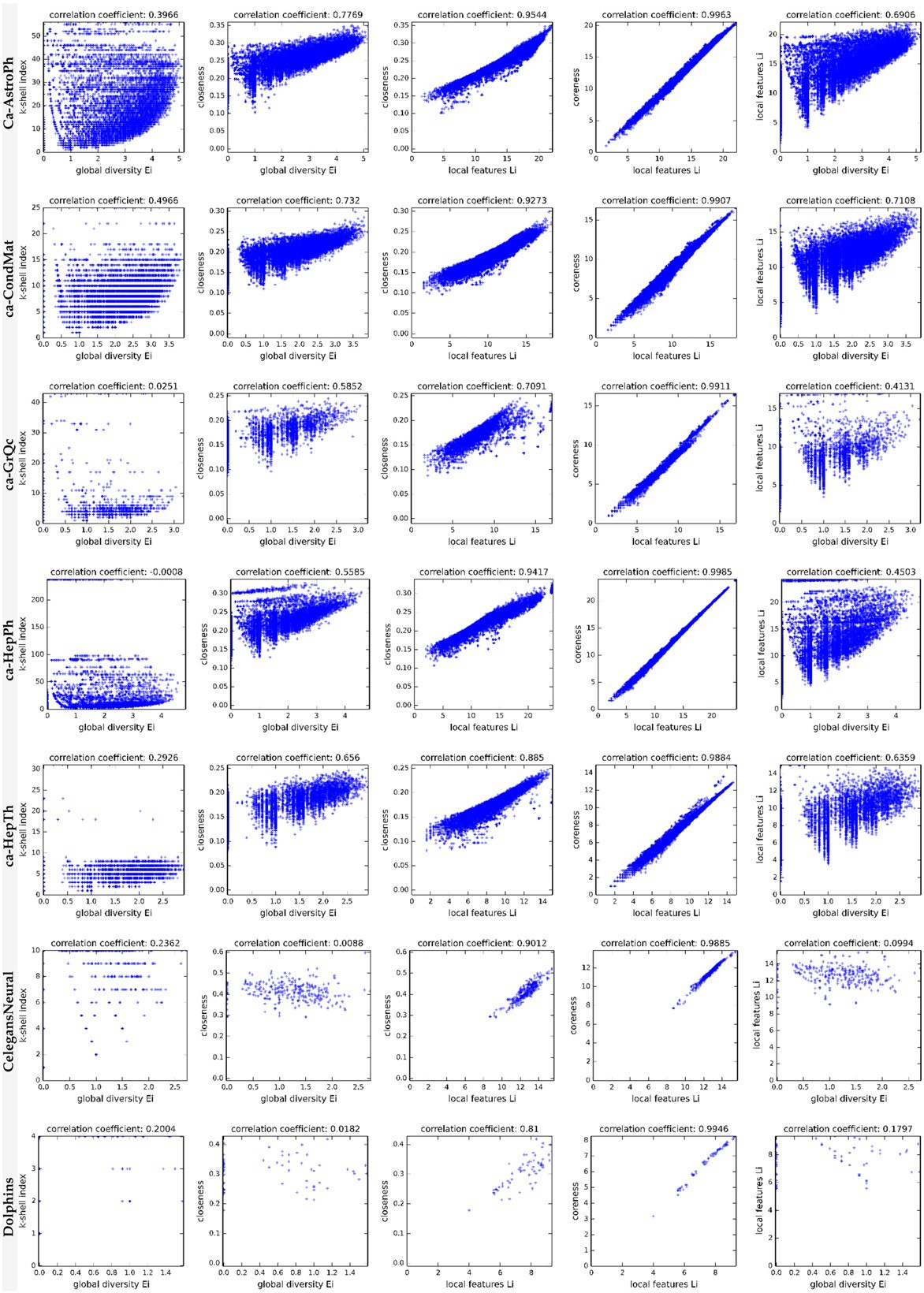
*- n* .

[表1](#_bookmark9)和[图4](#_bookmark8)分别显示了基本的复杂网络属性和巨型连接部件（GCC）网络结构的分析结果。[图4](#_bookmark8)中的相关系数*r*被分为高（*r*=0*.*7）、中（0*.*4 5 *r<*0*.*7）或低（*r<*0*.*4）[[30](#_bookmark36)]。基于这些分类，[图4](#_bookmark8)（a）显示了*E*和*K*壳指数之间的相关性。被确定为具有高相关性的通信网络类型包括Email-Contacts（*r* 0*.*82）、Email-Enron（*r 0.*78）和PolBlogs（*r 0.*74）--也就是说，在这些网络中，位于内部核心层的高全球多样性节点与其他网络核心层进行通信。在[图4](#_bookmark8)(b)中，关于*E*和接近度之间的相关性，大多数网络数据集（13个中的11个）至少有中等程度的相关性，这意味着高全球多样性节点倾向于网络中心，比其他节点有更低的通信成本。在[图4](#_bookmark8)(c)中，关于*L*和接近度之间的相关性，13个数据集中有12个具有高的相关性，这说明高的局部特征节点也倾向于网络中心，可以用来近似接近度中心性和减少计算成本。在[图4](#_bookmark8)(d)中，关于*L*和核心度[[31](#_bookmark37)]之间的相关性，所有的网络数据集都有很高的相关性（*r* = 0*.*9），这意味着*L*也可以用来近似核心度--即达到最广泛的局部范围的能力。在[图4](#_bookmark8)(e)中，关于*E*和*L*值之间的相关性，大多数数据集至少有中等程度的相关性，这意味着较高的相关性表明网络中有大量的核心层和邻居受到影响。

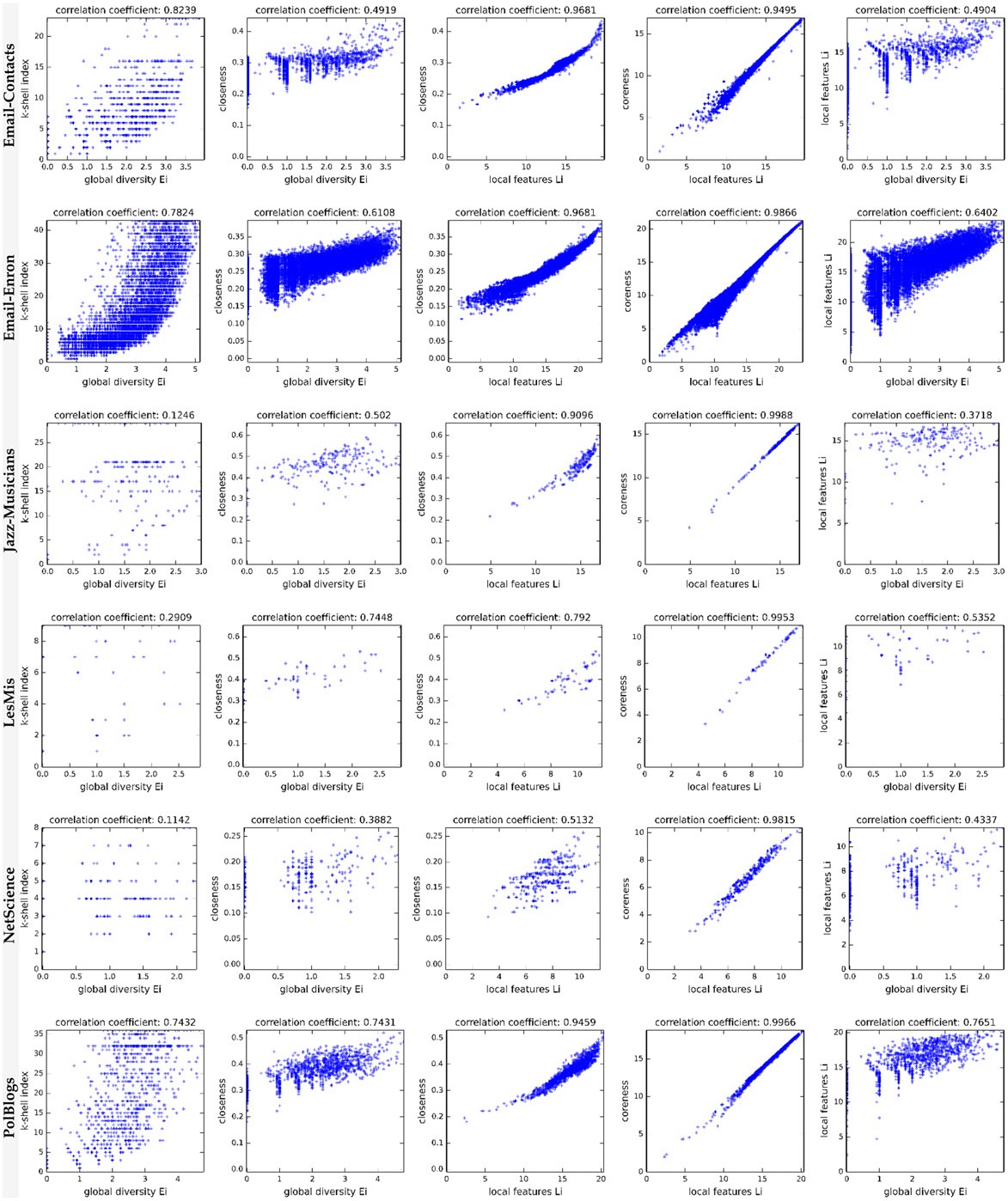
= =

=

总之，在PolBlogs数据集的所有结果中都发现了高度的关联性。原始PolBlogs数据集中的网络是一个关于美国政治的网络日志之间的超链接的定向网络[[22](#_bookmark31)]，这意味着频繁的互动和



**图4.**我们的传播实验中使用的复杂网络的统计结果。散点图的子图从左到右按以下顺序排列。(a)全球多样性和*K*壳指数，(b)全球多样性和接近度，(c)局部特征和接近度，(d)局部特征和核心度，以及(e)全球多样性和局部特征。*x-y*轴属性之间的相关系数在每个子图的顶部。



**图4.** (*续*)

网络内部发生的交流。在这种类型的网络中，网络传播者可能是疾病、信息或谣言传播中最具影响力的初始节点。然而，根据[表1](#_bookmark9)和[图4](#_bookmark8)所示的数据，我们不能提供明确的证据表明基本的网络属性或相关系数可以用来对网络结构进行分类，以描述复杂的网络传播动态。

在传播实验中，我们使用了三种网络数据集分类：科学合作、传统社会和其他。衡量标准是度、间性和紧密性中心度；*K*壳分解；核心度[[31](#_bookmark37)]；PageRank[[34](#_bookmark40)]；以及我们提出的方法。传播实验和SIR流行病模型参数如下。对每个网络数据集进行5000次模拟，每次模拟包括50个时间步长，每个措施的前1个节点作为初始传播者。我们实验中使用的SIR流行病模型的*β*感染率为

**表1**

本项目中使用的真实世界网络的属性。当原始网络被断开时，我们只考虑最大的连接网络组件。

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 网络类型 | 网络 | 描述 | *N* | *E* | (*c*) | kmax | (*k*) | ksmax | (*ks*) | *H* | *r* | *βthd* | *β* |
| 协作 | ca-AstroPh  ca-CondMat | arxiv.org的astro-ph中的共同作者。合著在 | 17 903  21 363 | 196 972  91 286 | 0.63  0.64 | 504  279 | 22.00  8.55 | 56  25 | 13.11  5.12 | 2.99  2.63 | 0.20  0.13 | 0.02  0.04 | 0.03  0.05 |
|  | ca-GrQc | Cond-mat类别。  Gr-qc中的合著 | 4 158 | 13 422 | 0.56 | 81 | 6.46 | 43 | 4.58 | 2.79 | 0.64 | 0.06 | 0.15 |
|  | ca-HepPh | 类别。  共同著作在 | 11 204 | 117 619 | 0.62 | 491 | 21.00 | 238 | 15.93 | 6.23 | 0.63 | 0.01 | 0.05 |
|  | ǞǞǞ | hep-ph类别。  共同著作在 | 8 638 | 24 806 | 0.48 | 65 | 5.74 | 31 | 3.41 | 2.26 | 0.24 | 0.08 | 0.12 |
|  |  | 第七类。 |  |  |  |  |  |  |  |  |  |  |  |
| 社会 | 爵士乐-音乐人 电子邮件-联系方式 | 1920年代爵士乐手之间的合作。计算机科学系的电子邮件联系人 | 198  12 625 | 2 742  20 362 | 0.62  0.11 | 100  576 | 27.70  3.23 | 29  23 | 17.27  1.65 | 1.40  34.25 | 0.02  -0.39 | 0.03  0.01 | 0.04  0.05 |
|  | 电子邮件-安隆 | 大学学院。  伦敦。  安然公司的电子邮件数据集。 | 33 696 | 180 811 | 0.51 | 1383 | 10.73 | 43 | 5.73 | 13.27 | -0.12 | 0.01 | 0.05 |
|  | 雅各布氏症候群神经系统 | 的神经网络。  C. elegans 线虫。 | 297 | 2 148 | 0.29 | 134 | 14.46 | 10 | 7.98 | 1.80 | -0.16 | 0.04 | 0.06 |
| 其他 | 海豚 | 62只海豚中的频繁关联。 | 62 | 159 | 0.26 | 12 | 5.13 | 4 | 3.16 | 1.33 | -0.04 | 0.15 | 0.15 |
|  | 吕思勉 | 悲惨世界》网络。 | 77 | 254 | 0.57 | 36 | 6.60 | 9 | 4.73 | 1.83 | -0.17 | 0.08 | 0.08 |
|  | 网易科学 | 网络科学合作。 | 379 | 914 | 0.74 | 34 | 4.82 | 8 | 3.47 | 1.66 | -0.08 | 0.12 | 0.20 |
|  | PolBlogs | 政治博客。 | 1 222 | 16 714 | 0.32 | 351 | 27.36 | 36 | 14.82 | 2.97 | -0.22 | 0.01 | 0.02 |

[如表1](#_bookmark9)所示。根据以前的至少一项研究，大的感染率在传播措施方面没有区别[[2](#_bookmark13)]。为了给每个网络数据集分配一个合适的感染率，感染率是通过比较理论上的流行病阈值*βthd*和参考研究中使用的数字来确定的[[31](#_bookmark37)]。恢复率总是被设定为1[[2](#_bookmark13)]，这意味着被感染集*I*中的每个节点在感染其邻居后立即进入恢复集*R*。基于网络的模拟步骤如下。

*H* = (*k*)*/*(*k*)2，程度异质性[[32](#_bookmark38)]。

*βthd*=(*k*)*/*(*k*)，理论上的流行病阈值[[33](#_bookmark39)]。

=

第一步。在初始化过程中，所有节点都处于*S*状态，除了(top-1)初始撒布器处于*I*状态。

第2步。在每个时间步骤*t中*，处于*I*状态的每个节点根据感染率*β*随机地感染其邻居，然后进入*R*状态（即1）。*⇢(t)的*累积传染率是在每个时间步骤结束时计算的恢复节点的数量。

=

第3步。重复步骤2，直到满足最大时间步长的要求--或者，如果有必要，当*I*状态集为空时[[2](#_bookmark13)]。

实验结果和有关传播动力学的细节（即*⇢（*50*）的*结果）[见图5](#_bookmark10)[和表2](#_bookmark11)。我们发现，领先组可以定义为大于最大结果减去1%的不准确系数的措施的传播结果，表示为

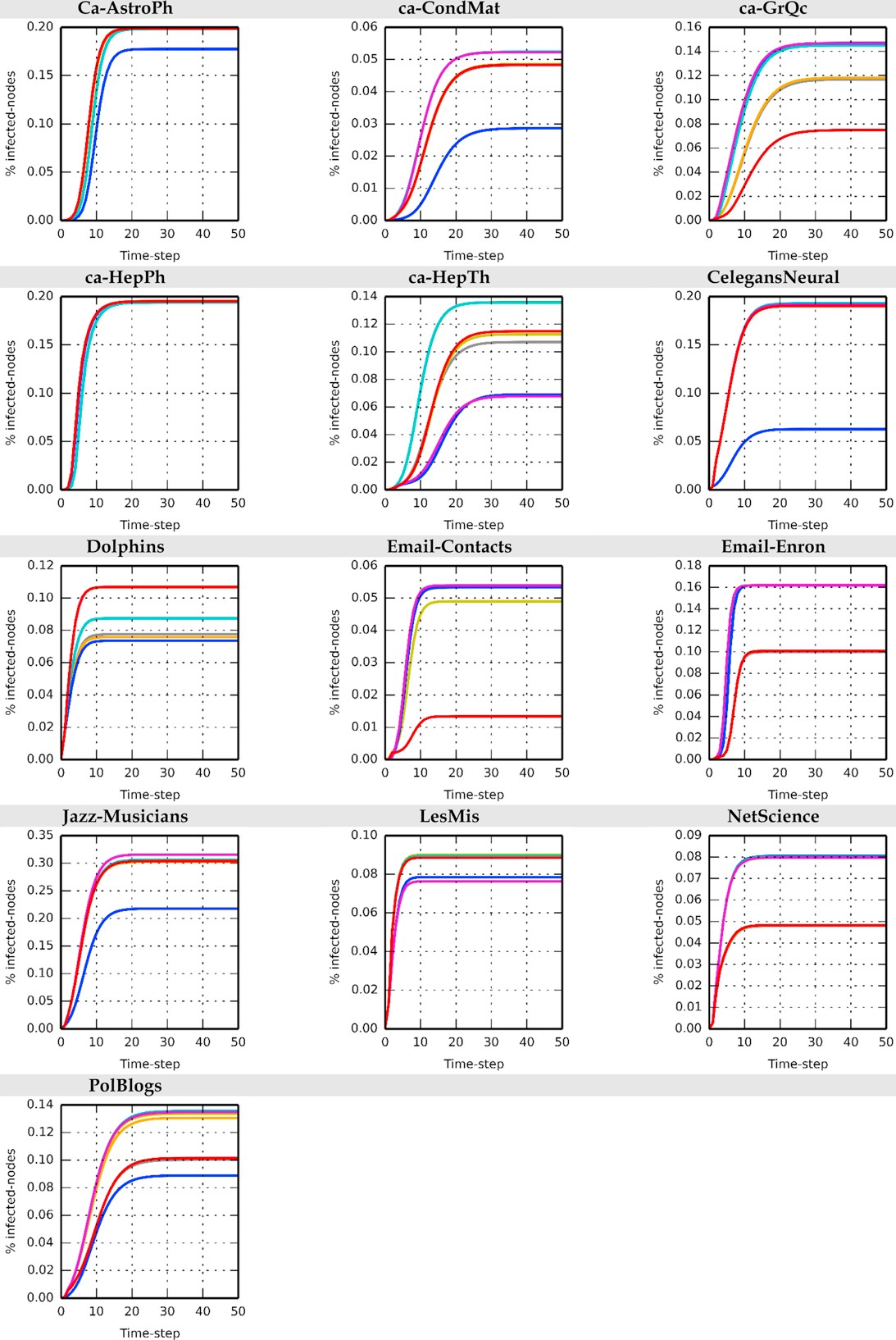
*LG* = {*m* | *pm* (*t)* ≥ *(p*max *(t)* - *err* \* *p*max *(t)) , m* e *M* and *err* e [0*,* 1] } (10)

其中，*M*是实验中使用的措施集，*p*max*（t）*是时间*t*的最大结果，*err*是不准确率（0.01），时间步骤*t* 50。

=

恢复的节点数*⇢（t）*被用来衡量和排列各种措施的传播能力。领先组可以帮助确定一个措施的稳定性，以识别不同网络中的节点的影响。领先组内的措施具有大致相同的传播能力。[表2](#_bookmark11)中的平均等级被用来解释不同网络中的预期等级：平均等级较低的措施被视为在识别好的传播者方面具有更好的辨别力。

根据内部领先组数（衡量稳定性的指标），我们提出的方法在识别不同网络中最有影响力的节点方面表现良好。根据我们的实验结果，所提出的方法能够识别出在复杂网络中作为良好传播者的节点，具有全局的多样性。除了在领先组内，该方法与该组内的其他措施相比，也有更好的排名。被识别的影响力传播者能够通过其多样化的全球连接到达大量的网络节点，能够影响网络层，并且能够对网络层产生最大的传播效果。一个节点的度中心性



**图5.**不同网络的传播动力学结果。图中显示的测量值为：interness（灰色）、closeeness（橙色）、degree（黄色）、*k*-shell（蓝色）、提议的方法（青色）、coreness（品红色）和PageRank（红色）。

**表2**

在使用[表1](#_bookmark9)所示真实世界网络的实验中，不同措施（包括我们提出的方法）的模拟结果比较。网络 *⇢(t* = 50*)*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 巨型连接部件（GCC） | 程度 *BetweennessClosenessk* -coreCorenessPageRankProposed | | | | | | |
| ca-AstroPh | **0***.***1988** 1 | **0***.***1983** 6 | **0***.***1986** 4 | 0*.*1774 7 | **0***.***1987** 2 | **0***.***1987** 2 | **0***.***1984** 5 |
| ca-CondMat | **0***.***0486** 3 | **0***.***0482** 6 | **0***.***0485** 4 | 0*.*0286 7 | **0***.***0523** 2 | **0***.***0484** 5 | **0***.***0523** 1 |
| ca-GrQc | **0***.***1462** 3 | 0*.*1168 6 | 0*.*1182 5 | **0***.***1469** 1 | **0***.***1467** 2 | 0*.*0748 7 | **0***.***1450** 4 |
| ca-HepPh | **0***.***1952** 1 | **0***.***1942** 6 | **0***.***1952** 1 | **0***.***1951** 4 | **0***.***1950** 5 | **0***.***1952** 1 | **0***.***1941** 7 |
| ǞǞǞ | 0*.*1126 4 | 0*.*1070 5 | **0***.***1359** 1 | 0*.*0691 6 | 0*.*0678 7 | 0*.*1148 3 | **0***.***1357** 2 |
| 爵士乐-音乐人 | 0*.*3034 5 | 0*.*3036 4 | 0*.*3024 6 | 0*.*2177 7 | **0***.***3156** 1 | 0*.*3040 3 | **0***.***3060** 2 |
| 电子邮件-联系人 | **0***.***0490** 6 | **0***.***0533** 5 | **0***.***0539** 1 | **0***.***0534** 4 | **0***.***0539** 1 | 0*.*0134 7 | **0***.***0539** 1 |
| 电子邮件-安隆 | 0*.*1008 5 | 0*.*0997 7 | **0***.***1619** 1 | **0***.***1618** 4 | **0***.***1619** 1 | 0*.*1007 6 | **0***.***1619** 1 |
| 雅各布氏症候群神经系统 | **0***.***1909** 5 | **0***.***1922** 2 | **0***.***1916** 4 | 0*.*0627 7 | **0***.***1918** 3 | **0***.***1902** 6 | **0***.***1928** 1 |
| 海豚 | **0***.***1067** 3 | 0*.*0775 5 | 0*.*0757 6 | 0*.*0734 7 | **0***.***1069** 1 | **0***.***1069** 1 | 0*.*0874 4 |
| 吕思勉 | **0***.***0901** 1 | **0***.***0894** 2 | **0***.***0893** 4 | 0*.*0785 6 | 0*.*0764 7 | **0***.***0886** 5 | **0***.***0894** 2 |
| 网易科学 | **0***.***0797** 4 | 0*.*0479 7 | 0*.*0482 6 | **0***.***0805** 1 | **0***.***0799** 3 | 0*.*0483 5 | **0***.***0802** 2 |
| PolBlogs | **0***.***1336** 3 | 0*.*1003 6 | **0***.***1306** 4 | 0*.*0888 7 | **0***.***1347** 2 | 0*.*1014 5 | **0***.***1354** 1 |
| 内部领先的组号。 | 10 | 6 | 9 | 5 | 11 | 6 | 12 |
| 平均等级。 | 3.3846 | 5.1538 | 3.6154 | 5.2308 | 2.8462 | 4.3077 | 2.5385 |

**粗体字**，测量结果是在领先的网络组内。下标，网络在测量中的排名。

和它的邻居，可以用来保持复杂网络的局部层的接触节点的数量。然而，人们注意到各措施之间存在重要的差异。例如，在ca-HepTh、ca-HepTh、Email-Contacts和Email-Enron网络（[表2](#_bookmark11)）的前1位中，亲近度措施表现良好，但在Jazz-Musicians、Dolphins或NetScience网络中则不然。由于该措施想要捕捉的特征在这些网络中可能还不够强大，所以无法确定最有影响力的传播者。

全局多样性*k*和局部特征*k的*平均值被用来确定某些因素组合的预期效果，将网络节点分为四类，并解释为什么要满足两个定义条件。对于节点类型（a），*Ei* *k*和*Li* *k*--即那些被确定为有影响力的传播者的节点能够利用全局多样性和局部特征的优势来影响大多数网络层，并在传播过程的早期阶段感染大量的邻居（如枢纽）。因此，这种节点类型被称为具有最大的传播能力。对于节点类型（b），*E ik*和*Li<k*--即节点影响了许多不同的网络层，但只能够感染少量的邻居节点，也就是说，传播范围取决于邻居是位于内部核心层还是整个高*K*核心层；相应地，这种节点类型的传播活动潜力很大。对于节点类型（c），*Ei<k*和*L* *ik*--即节点只能影响一个或少量的网络层，但它们感染了大量的邻居。这里的传播范围取决于邻居是否位于高*k*核层而不是集群中；这种类型的节点的传播活动潜力与类型（b）相同。对于节点类型（d），*Ei<k*和*Li<k*--即节点只能影响一个或少数网络层，感染少数邻居，表明传播范围被限制在外围层，表明传播能力最小。

( )()

( )()

≥ ( )( )

( ) ≥ ( )

≥ ( )≥ ( )

尽管所提出的方法强调了识别不同网络中最具影响力的节点的稳健性和稳定性，但其局限性可能取决于所涉及的节点类型。例如，对于节点类型（c），在复杂网络中没有全局多样性的情况下，网络的最大*k*壳值会更低，网络规模也会大大缩小。[如表2](#_bookmark11)所示，多尔芬网络中具有高全局多样性的节点无法被识别。在这种情况下，我们提出的方法所识别的节点的传播能力下降到度中心性（忽略第一个项），节点的影响仅限于局部网络层。在没有全局多样性的情况下，公式.[(9)](#_bookmark7)变成

*IF Lii* ，这有利于本地网络层（即程度中心化）。此外，*E*ˆ*i (X* *i)*的归一化全球多样性值

⇡

我们提出的方法产生的参与系数与Teitelbaum等人报告的参与系数相似[[35](#_bookmark41)]，我们观察到的节点的高全局多样性与连接器中心和无亲属中心的参与系数相似，两者都有不同的参与系数。

加权方案通常被添加到方程中以调整不同项的权重。然而，我们想避免与不同网络之间的参数优化有关的问题，因此我们做了一个有目的的决定，保持我们的方程简单，在这个特定的研究中不添加加权方案。这个决定的另一个好处是，不同方程项的权重可以根据网络拓扑结构来确定。例如，假设全局多样性*E*和局部特征*L*的权重被引入*IF*方程中。在Dolphins网络中，全局多样性*E*是不存在的，因此可以忽略，使*IF*为*L*，并允许使用局部特征*L*来确定节点的重要性。这类似于将*E*和*L*的假设权重设置为*↵0.*1和*（*1↵*）0.9*。然而，存在涉及权重的详细调整和参数优化的缺点，不能适用于不同网络数据集的指定目的。

=- =

⇡

越来越多的复杂网络和社会网络研究者正在使用数据驱动的方法来研究在线社会媒体和网站的传播现象。例如，Pei等人[[36](#_bookmark42),[37](#_bookmark43)]跟踪了信息流

涉及使用Facebook、LiveJournal、Twitter和其他在线社交网络数据集的实际扩散过程。他们使用最近邻居的度数之和来近似测量（即*k*-core分解），并发现近似测量的表现几乎与*k*-shell指数一样好，而且优于度数中心和PageRank。假设可以完全获得网络拓扑结构和实际信息流，我们提出的两步框架可以应用于数据驱动的方法，具体如下。

第一步：部署搜索算法（即BFS）来提取用户的实际扩散范围，并使用*K*壳指数或估计值来计算其全局多样性*E*。在数据驱动的方法中，全球多样性*E*被称为基于实际的多样性。在本文中，全球多样性被称为基于拓扑的多样性。

第2步：利用本地特征计算出在传播过程中可能有多少邻居被感染。

第三步：结合基于实际的全局多样性*Eactual*和局部特征*L*，获得网络的节点影响力*IF actual*。

相对于使用理论上的SIR或SIS模型来分析网络中节点的传播能力，数据驱动的方法在直接分析实际的用户扩散过程数据和获得节点影响值方面有很大的潜力。然而，在涉及网络拓扑结构和部分用户数据的情况下，基于拓扑结构的方法和理论上的流行病模型在分析和识别复杂社会网络中的有影响力的传播者方面可能仍然有用。

# 总结

在本文中，我们描述了我们关于计算网络节点影响力的两步框架的建议。在第1步中，使用全局算法来分析全局节点信息，用信息论中的熵概念来协助测量节点的全局多样性。受影响的全球网络层可以用*k*-shell熵来识别。在第二步中，同时考虑节点及其邻居的度中心性，以保持复杂网络局部层中受影响邻居的数量。在最后一步，全局多样性和局部特征被结合起来以确定网络中节点的影响。我们的实验结果表明，所提出的方法表现良好，在不同网络数据集的领先组中保持稳定。换句话说，所提出的方法能够识别最有影响力的节点，作为在不同网络中传播信息、思想或疾病的初始传播者。

我们的计划是为我们的分析增加相当多的细节，引入评估传播能力的复杂方法，并阐明所提出的方法如何受到网络结构的影响。例如，可以使用社区检测算法等全局算法来分析和获得社区网络结构的全局信息，并确定位置和节点作用等因素[[35](#_bookmark41)]如何影响传播者在整个复杂网络中传播信息或疾病的程度。我们还计划研究与网络中多个初始传播者有关的策略。由于选定的传播者的重叠感染区必须最小化[[2](#_bookmark13)]，多个初始传播者的情况可能加速或阻碍复杂网络中的传播。

# 鸣谢

该工作得到了中华民国国家科学委员会的部分资助（MOST-103-2221-E-182-052）。该工作得到了台湾长庚大学高速智能通信（HSIC）研究中心的部分支持。

# 参考文献

1. S.[Carmi,S.Havlin,S.Kirkpatrick,Y.Shavitt,E.Shir,AmodelofInternettopologyusing*k*-shelldecomposition,Proc.Natl.AcadSci.104(27)(2007)](http://refhub.elsevier.com/S0378-4371(15)00304-0/sbref1) 11150-11154.
2. M.[Kitsak,L.K.Gallos,S.Havlin,F.Liljeros,L.Muchnik,H.E.Stanley,H.A.Makse,Identificationofinfluentialspreadersincomplexnetworks,Nat.复杂网络中具有影响力的传播者的识别，Nat.Phys.](http://refhub.elsevier.com/S0378-4371(15)00304-0/sbref2)

6 (11) [(2010)888-893.](http://refhub.elsevier.com/S0378-4371(15)00304-0/sbref2)

1. B.Hou, [Y.Yao,D.Liao,Identifyingall-aroundnodesforspreadingdynamicsincomplexnetworks,PhysicaA391(15)(2012)4012-4017.](http://refhub.elsevier.com/S0378-4371(15)00304-0/sbref3)
2. D.[Chen,L. Lü,M. -S.Shang,Y.C.Zhang,T.Zhou,Identifyinginfluentialnodesincomplexnetworks,PhysicaA391(4)(2012)1777-1787.](http://refhub.elsevier.com/S0378-4371(15)00304-0/sbref4)
3. X.[Zhang,J.Zhu,Q.Wang,H.Zhao,Identifyinginfluentialnodesincomplexnetworkswithcommunitystructure,Knowl.-BasedSyst.42(2013)74-84.](http://refhub.elsevier.com/S0378-4371(15)00304-0/sbref5)
4. A.[Zeng,C.J.Zhang,Rankingspreadersbydecomposingcomplexnetworks,.Lett.A377(14)(2013)1031-1035.](http://refhub.elsevier.com/S0378-4371(15)00304-0/sbref6)
5. S.Gao, [J.Ma,Z.Chen,G.Wang,C.Xing,Ranking the spreadingabilityofnodesincomplexnetworksbasedonlocalstructure,PhysicaA403(2014)](http://refhub.elsevier.com/S0378-4371(15)00304-0/sbref7) 130-147.
6. J.G. [Liu,Z.-M.Ren,Q.Guo, 复杂网络中的扩散影响排名,PhysicaA392(18)(2013)4154-4159.](http://refhub.elsevier.com/S0378-4371(15)00304-0/sbref8)
7. B.[Doerr,M.Fouz,T.Friedrich,Whyrumorsspreadsoquicklyinsocial networks,Commun.acm55(6)(2012)70-75.](http://refhub.elsevier.com/S0378-4371(15)00304-0/sbref9)
8. N.A. [Christakis,J.H.Fowler,Socialnetworksensorsforearlydetectionof contagoutbreaks,PLoSOne5(9)(2010)e12948.](http://refhub.elsevier.com/S0378-4371(15)00304-0/sbref10)
9. M.[纽曼，《网络》。AnIntroduction,OxfordUniversityPress,2010.](http://refhub.elsevier.com/S0378-4371(15)00304-0/sbref11)
10. M.G. [Kendall,Anewmeasureofrankcorrelation,Biometrika30(1-2)(1938)81-93.](http://refhub.elsevier.com/S0378-4371(15)00304-0/sbref12)
11. N.[Eagle,M.Macy,R.Claxton,Networkdiversityandeconomicdevelopment,Science328(5981)(2010)1029-1031.](http://refhub.elsevier.com/S0378-4371(15)00304-0/sbref13)
12. J.Cranshaw, E. Toch, J. Hong, A. Kittur, N. Sadeh, 缩小物理位置和在线社交网络之间的差距, in:第12届ACM泛在计算国际会议论文集，2010年，第119-128页。
13. E.Serin, S. Balcisoy, 基于熵的社会网络敏感性分析和可视化, in:2012年IEEE/ACM社会网络分析与挖掘进展国际会议, ASONAM, 2012, 1099-1104页。
14. S.J. [Russell,P.Norvig,ArtificialIntelligence:AModernApproach,PrenticeHall,2010.](http://refhub.elsevier.com/S0378-4371(15)00304-0/sbref16)
15. M.[Rosvall,C.T.Bergstrom,Aninformation-theoreticframeworkforresolvingcommunitystructureincomplexnetworks,Proc.Natl.AcadSci.104](http://refhub.elsevier.com/S0378-4371(15)00304-0/sbref17) (18) [(2007)7327-7331.](http://refhub.elsevier.com/S0378-4371(15)00304-0/sbref17)
16. M.[Girvan,M.E.J.Newman,社会和生物网络的社区结构,Proc.Natl.Acad99(12)(2002)7821-7826.](http://refhub.elsevier.com/S0378-4371(15)00304-0/sbref18)
17. R.[Pastor-Satorras,A.Vespignani,Epidemicdynamicsandendemicstatesincomplexnetworks,Phys.Rev.E63(6)(2001)066117.](http://refhub.elsevier.com/S0378-4371(15)00304-0/sbref19)
18. C.Y. [Huang,C.L.Lee,T.H.Wen,C.T.Sun,Acomputervirusspreadingmodelbasedonresourcelimitationsandinteractioncosts,J.Syst. Softw.Softw.86(3)](http://refhub.elsevier.com/S0378-4371(15)00304-0/sbref20) (2012) [801-808.](http://refhub.elsevier.com/S0378-4371(15)00304-0/sbref20)
19. 斯坦福大学大型网络数据集收集。[http://snap.stanford.edu](http://snap.stanford.edu/)。
20. 由Mark Newman收集的网络数据集[。http://www-personal.umich.edu/~mejn/netdata/](http://www-personal.umich.edu/%7Emejn/netdata/)。
21. Hernán Alejandro Makse收集的网络数据集[。http://lisgi1.engr.ccny.cuny.edu/~makse/soft\_data.html](http://lisgi1.engr.ccny.cuny.edu/%7Emakse/soft_data.html)。
22. 爵士音乐家网络数据集。<http://www.infochimps.com/datasets/jazz-musicians-network>。
23. P.[Basaras,D.Katsaros,L.Tassiulas,Detectinginfluentialspreadersincomplex,dynamicnetworks,Computer46(4)(2013)24-29.](http://refhub.elsevier.com/S0378-4371(15)00304-0/sbref25)
24. C.Y. [Huang,T.H.Wen,Y.S.Tsai,FLUed:用于模拟流行病动态和评估干预政策新型四层模型,J.Appl.Math](http://refhub.elsevier.com/S0378-4371(15)00304-0/sbref26)2013[(2013)e325816.](http://refhub.elsevier.com/S0378-4371(15)00304-0/sbref26)
25. L.Zhao, [J.Wang,Y.Chen,Q.Wang,J.Cheng,H.Cui,社会网络中的SIHR谣言传播模型,PhysicaA391(7)(2012)2444-2453.](http://refhub.elsevier.com/S0378-4371(15)00304-0/sbref27)
26. C.Y. [Huang,C.T.Sun,H.C.Lin,Influenceoflocalinformationonsocialsimulationsinsmall-worldnetworkmodels,J.Artif..Soc.Simul.8(4)(2005).](http://refhub.elsevier.com/S0378-4371(15)00304-0/sbref28)
27. C.E. [Shannon,Amathematicaltheoryofcommunication,BellSyst.Tech.J.27(3)(1948)379-423.](http://refhub.elsevier.com/S0378-4371(15)00304-0/sbref29)
28. C.P. [Dancey,J.Reidy,StatisticsWithoutMathsforPsychology,fifthed.,PrenticeHall,HarlowEngland,NewYork,2011.](http://refhub.elsevier.com/S0378-4371(15)00304-0/sbref30)
29. J.Bae, [S.Kim,Identifyingandrankinginfluentialspreadersincomplexnetworksbyneighborhoodcoreness,PhysicaA395(2014)549-559.](http://refhub.elsevier.com/S0378-4371(15)00304-0/sbref31)
30. H.B. [Hu,X.F.Wang,量化复杂网络异质性的统一指标,PhysicaA387(14)(2008)3769-3780.](http://refhub.elsevier.com/S0378-4371(15)00304-0/sbref32)
31. C.[Castellano,R.Pastor-Satorras,Thresholdsforepidemicspreadinginnetworks,Phys.Rev.Lett.105(21)(2010)218701.](http://refhub.elsevier.com/S0378-4371(15)00304-0/sbref33)
32. L.Page, [S.Brin,R.Motwani,T.Winograd,ThePageRankcitationranking:给网络带来秩序，技术报告，斯坦福信息实验室，1999。](http://refhub.elsevier.com/S0378-4371(15)00304-0/sbref34)
33. T.[Teitelbaum,P.Balenzuela,P.Cano,J.M.Buldú,Communitystructuresandroledetectioninmusicnetworks,Chaos18(4)(2008)043105.](http://refhub.elsevier.com/S0378-4371(15)00304-0/sbref35)
34. S.Pei, [H.A.Makse,Spreadingdynamicsincomplexnetworks,J.StatMech.2013(12)(2013)P12002.](http://refhub.elsevier.com/S0378-4371(15)00304-0/sbref36)
35. S.Pei, [L.Muchnik,J.J.S.Andrade,Z.Zheng,H.A.Makse,Searchingforsuperspreadersofinformationinreal-worldsocialmedia,Sci.Rep.4（2014）。](http://refhub.elsevier.com/S0378-4371(15)00304-0/sbref37)