

一种面向多维复杂网络的节点传播重要性算法

张 昕 王慧慧 严 沛 郭 阳

(辽宁大学信息学院 沈阳 110004)

摘 要 如何度量节点在网络拓扑结构中的重要程度,一直是复杂网络相关领域中的研究热点。现有的研究大多面向单维网络,针对现实网络结构往往是多维共存的问题,提出了维度相似性的定义来度量各维度间的关系。考虑实际信息传播过程中信息衰减对节点重要性的影响,给出传播衰减率的定义,并通过全连接单维网络传播无损假设及对应算法确定衰减系数取值。进一步给出节点重要性的计算方法,在算法中利用复杂网络小世界特性,限定最长传播跳数,使得算法兼顾时间效率与精确度。在真实网络上进行了验证,实验结果表明,与传统的节点度以及节点介数方法相比,该算法在精确度与时间效率方面均具有一定优势。

关键词 多维网络,节点重要性,维度相似性,衰减系数,最长传播跳数

中图分类号 TP301 文献标识码 A

Node Propagation Importance Algorithm for Multi-dimensional Complex Networks

ZHANG Xin WANG Hui-hui YAN Pei GUO Yang

(School of Information, Liaoning University, Shenyang 110004, China)

Abstract How to measure node importance in the network topology has always been a research hotspot in the field of complex networks. Most of the existing researches are oriented to single dimensional networks. Therefore, aiming at the fact that there is often a multi dimensional coexistence in real-world network structure, the definition of dimensional similarity was proposed to measure the relationship between dimensions. Considering the impact of information attenuation on node importance in actual process of information propagation, the definition of propagation attenuation rate is given. The value of attenuation coefficient is determined by propagation non-destructive assumption on a fully connected single dimensional network and corresponding algorithm. And the node importance algorithm is given further. The small network characteristics of the complex network are utilized in the given algorithm to limit the maximum propagation hops, so that the algorithm takes into account both time efficiency and accuracy. The experimental results on the real network show that the proposed algorithm has certain advantages in accuracy and time efficiency compared with traditional node degree and node betweenness methods.

Keywords Multidimensional network, Node importance, Dimensional similarity, Attenuation rate, Maximum propagation hops

1 引言

在科技迅速发展进步的同时,各类复杂网络中信息传播的规模与速度均呈现急剧增长的趋势,例如社交媒体网络的飞速增长与新闻消息的快速扩散,交通网络的持续建设与运输能力的不断提升等。研究各种貌似互不相同的网络之间的共性和处理它们的普适方法^[1],恰是复杂网络的研究重点。现实网络更适合抽象为多维复杂网络^[2-3],例如由铁路、水路、航空 3 种不同的交通方式构成的交通网络就是一种三维网络关系。但现阶段的研究工作大多针对单维复杂网络,针对多维网络的研究成果较少。因此,结合现实网络,探索多维复杂网络中的不同维度间的关系,具有重要的理论及应用价值。

节点重要性是复杂网络研究的热点之一,常用的节点重要性度量方法有节点度^[4]与节点介数^[5]。节点度即节点邻边

的数量^[6],其数值越大表明节点对周边的影响力越大,重要程度越高。该方法的优点是计算效率高,但有存在局部片面性。节点介数表示网络中最短路径经过该节点的数量比例^[7],其数值越大表明节点的枢纽作用越大,重要程度越高。该方法适用于全局网络^[8-9],但计算复杂度较高。

目前,对多维复杂网络中节点重要性的相关研究也有了一定的成果。文献[10]中将度的数量转换为维度数量的累加,即维度数量越多表示度数越大。这种方法忽视了维度之间的关系。杨建祥等提出了多维网络局部介数中心度度量算法^[11],将节点之间的维度数量看作路径的数量。该方法考虑了多个维度之间的关系,将多维信息映射为不同的语义;但该方法在映射的过程中容易丢失部分信息,而且仅仅适用于一些特定的场合。

本文受国家自然科学基金项目(61472169),辽宁省发改委工程实验室项目(2016-294),辽宁省博士科研启动基金项目(20170520323)资助。

张 昕(1979-),男,博士,副教授,CCF 会员,主要研究方向为数据挖掘、复杂网络;王慧慧(1992-),女,硕士生,主要研究方向为复杂网络, E-mail:meng210021@163.com(通信作者)。

本文重点研究多维复杂网络中节点的传播重要性,主要工作包括:

1)分析了维度之间的关联程度,提出了维度相似性的概念,体现了多个维度之间拓扑结构的相似程度;

2)给出传播衰减率的定义,合理描述了信息在传播过程中必然产生的衰减情况,并通过全连接单维网络传播无损假设及对应算法,推算得出该定义中衰减系数的值;

3)利用复杂网络小世界特性,限定最长传播跳数,提高节点重要性的计算的时间效率;

4)给出节点传播重要性计算方法,并通过实验验证算法的有效性。

2 相关定义

本文研究对象为无向无权图复杂网络,可以抽象为图 $G=(V,E)$,其中 V 表示节点集, E 表示边集。 E 中每条连边都在 V 中有一对顶点与之对应,记为 $e=(u,v)$,其中 $e \in E$ 且 $u,v \in V$ 。节点 u 的邻居节点集合记为 $N_G(u)=\{u_i | u_i \in V, (u,u_i) \in E\}$ 。而多维复杂网络可以看作多个单维复杂网络的复合,令 $G_i=(V,E_i)$ 表示其中第 i 个单维网络(称为第 i 维网络或维度 i),则整个 n 维网络表示为 $G_n=(V,E_n)$,其中 $E_n = \bigcup_{i=1}^n E_i$ 。

多维复杂网络中节点的传播重要性是指节点在信息传播过程中通过不同维度累积的信息量大小。令 $IN(v)$ 表示节点 v 的信息量,其值越大,说明节点 v 接收信息的能力就越强,在多维复杂网络中的拓扑位置越重要。由于节点之间可能存在多种维度,因此在理想状态,即不存在传播衰减的多维复杂网络中,节点信息量应为沿各维度传播结果的合并,即 $IN(v) = \bigcup_{u \in N_G} \bigcup_{i=1}^n IN_i^u(v)$,其中 n 为网络维数。

2.1 维度相似度

节点通过不同维度所接收到的信息量是有差异的。由于多维复杂网络的多个维度之间不仅存在语义差异,而且它们之间也是互相关联的,因此要找出各维度间的相似性对维度间的关系进行量化,据此来度量不同维度间的信息传播差异。

定义 1(维度间共同路径) 对于节点对 $u,v \in V$,若以其为端点的路径 P 同时存在于 G_i 与 G_j ($i \neq j$),则称路径 P 为维度 i 与维度 j 的维度间共同路径。

图 1 所示为一个 2 维网络,分别由实线与虚线表示两个维度的边,则路径 $F-D-A-B-E-C$ 及其所有子路径均为两个维度的维度间共同路径。

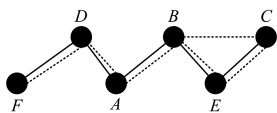


图 1 2 维网络共同路径

定义 2(维度相似度) 令 G_i 为基本维度,其路径数量为 $|P_i|$, G_i 与 G_j 的维度间共同路径数为 $|P_{sim}|$,则 G_i 和 G_j 的相似度为:

$$Sim_{G_i}(G_j) = \frac{|P_{sim}|}{|P_i|} \quad (1)$$

考查维度间相似性时,确立某个维度作为基本维度,以便

量化维度间的关系。显然,相似度值越大,即维度间共同路径的数量越多,则两个维度间的相似度越高,二者的传播过程越接近。由于基本维度的选取对相似度计算存在影响,进而影响信息量的量化结果,因此为了更准确地衡量节点的传播重要性,实际计算时可以每次选取不同的维度作为基本维度,并将计算平均值作为最终结果。

2.2 传播衰减率

在实际的信息传播过程中,必然有一部分信息会丢失,这种由于信息在通过传输介质进行信息传播时产生信息丢失的现象就是传播衰减。

定义 3(传播衰减率) 信息在维度 G_j 中每传播一跳,假设由一个节点 u 传播至其邻居节点,传播过程中遗失的信息量与节点 u 信息量的比例就称为传播衰减率,记为:

$$U_j^i = 1 - (1 - \mu) * Sim_{G_i}(G_j) \quad (2)$$

其中, μ 为衰减系数,表示信息传播时每经过一跳的折损率,是网络的一个固有系数。本文将基本维度的初始信息量取值为 1,用维度相似度近似衡量维度 G_j 的初始信息量,且衰减系数越大,该维度衰减率也就越大。上述定义通过综合考虑衰减系数与维度相似度,体现了不同维度中传播衰减的差异。

3 衰减系数的确定

由于在多维复杂网络中,衰减系数 μ 是一个固有系数,其大小只与网络本身有关,而与边所属的维度和信息量的大小无关,因此可以通过对单维网络的分析推算出 μ 的近似值。

考虑一种极限情况,即在全连接的单维网络中,信息由某一节点经由所有可能路径传播至另一节点,其中的可能路径不含环路。图 2 所示为全连接的单维网络中信息传播的可能路径,用实线表示,其中节点 A 与节点 B 分别为传播源点与传播终点。图 2 中子图(a)所示为长度为 1 的路径,即节点 A 与节点 B 的直接连接,子图(b)、子图(c)和子图(d)分别展示长度为 2,3,4 的可能路径之一。

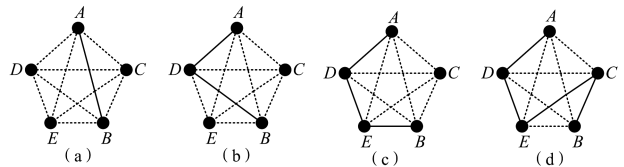


图 2 全连接单维网络

虽然各条路径经过的跳数不同,信息衰减情况也不相同,但可以认为整合至目的节点后的信息量等于初始信息量,即全连接单维网络传播无损。这与现实网络中的情况接近,比如在成员之间互相认识的社交网络中传播某个消息,虽然成员相互之间经口头或邮件等方式传递的信息与事实存在出入,但经过全面的交流后,最终每人都能获得与事实基本一致的信息。

衰减系数确定算法的具体描述如算法 1 所示。

算法 1 衰减系数确定算法

输入:节点集合 V ,衰减系数 μ 初始值

输出:衰减系数 μ

Begin

for(V 中每个节点 v) {

 令 v 为未标记;

```

IN(v)=∅;
}
n=|V|;
在 V 中任选一节点 vsource;
标记 vsource;
IN(vsource)=S;
while(n>1){
  for(V 中每个未标记节点 v'){
    IN(v')=IN(vsource) * (1-μ) ∪ IN(v');
  }
  vsource = V 中任一未标记节点;
  标记 vsource;
  n--;
}
vlast = V 中剩余的唯一未标记节点;
if(IN(vlast)=S)
  减小 μ 的值;
  goto begin;
else if(IN(vlast)<S * 95%)
  增大 μ 的值;
  goto begin;
else
  return μ;
End

```

该算法由任选的传播源点出发,经过不含重复边的路径传播信息至唯一终点。以图 2 为基础举例说明不含重复边路径的含义:设传播源点为节点 A,终点为节点 E,则路径 ABCE 与 ACBE 仅考虑其中一条,即 BC 与 CB 为重复边。通过排除重复边,算法可以避免信息回传对信息量累积的影响。若终点信息量等于源点初始信息量,则说明衰减不足,若终点信息量远小于源点初始信息量,则说明衰减过度,据此调整衰减系数的取值。本文认定终点信息量大于源点初始信息量的 95% 且不等于源点初始信息量即为信息基本无损,以此确定衰减系数的最终取值。

若全连接网络中节点数量较多,则每一节点均可得到来自大量不同路径的信息输入,使得合并后的节点信息量等于或接近初始信息量,无法体现衰减系数的作用。因此,选取节点数为 5~8 的全连接单维网络进行实验,对不同网络均进行多次实验,每次输入不同的初始 μ 值,取多次实验结果的平均值为不同网络对应的 μ 值,如表 1 所列。

表 1 μ 的取值分布

n	5	6	7	8
μ	0.18	0.16	0.20	0.18

由表 1 可知,各网络推算所得的对应 μ 值比较接近,验证了全连接单维网络传播无损假设的正确性,本文取表 1 所列 μ 值的均值,即 0.18 为多维复杂网络衰减系数的值。

4 节点传播重要性评估算法

依据第 2 节中多维复杂网络节点传播重要性的定义,考查信息在网络中传播后节点的累积信息量,并将其作为节点传播重要性的度量指标。信息在多维网络中传播时,受到维度相似性及衰减系数的共同影响,若衰减后到达某节点的信

息越多,则该节点的传播重要性就越大。

4.1 最大传播跳数

小世界^[12-14]特性是复杂网络的基本性质之一,意指虽然复杂网络的规模很大,但是实际网络中节点与节点之间的平均距离与网络中所有节点数量相比只是一个极小的数。因此,在信息从传播源点到其他节点的传播过程中,需要经过的路径跳数往往较小。选择能覆盖绝大多数节点的合适的传播跳数 k,将大幅提高多维复杂网络节点传播重要性的计算效率。本文选取不同规模的经典单维数据集和 multidimensional 数据集进行了计算,考查不同 k 值情况下传播的节点覆盖率,即传播源点所能到达的节点数与网络中所有节点数的比例。

计算选取的实际网络包括单维的 Zachary 空手道社团成员关系网络^[15] 与美国大学生橄榄球网络^[16],以及多维的 WikiElec 网络^[17]。其中,Zachary 网络是由 34 个节点组成的小型社员关系网络,节点对间的连边代表成员与成员间是朋友关系;橄榄球网络是对美国大学生橄榄球联赛某个赛季的比赛情况进行统计构成的中型网络,节点对间的连边表示两个球队间曾打过比赛;WikiElec 网络是从维基百科的投票选举产生的数据(其中包括大约 7000 位维基用户与 10 万次历史投票结果)中抽取的数据集,节点代表投票者,节点之间的边代表着用户 A 对用户 B 投票,其中按照投票的态度将边分为 3 个维度。在这 3 个网络上不同 k 值下的覆盖率计算结果如图 3 所示,当 k 取值过小时,3 个网络中覆盖率均较低,直至 k=4 时,覆盖率达到 90% 左右,当 k=6 时,覆盖率达到 95% 以上。

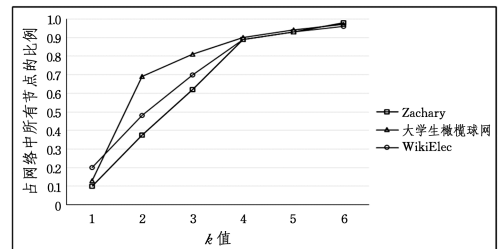


图 3 不同 k 值下的覆盖率

4.2 传播重要性算法

多维复杂网络的节点传播重要性算法的具体描述如算法 2 所示。

算法 2 传播重要性算法

输入: n 维网络 G_n (其中基本维度为 G_b), 衰减系数 μ , 最长传播跳数 k, 传播源点占节点总数的比例 r

输出: 网络 G_n 中全部节点的信息量

Begin

for(其他维度 G_i)

$P = E_b \cap E_i$;

//P 表示基本维度与维度 i 重合边的集合;

// E_b 表示基本维度 B 的边集合;

Count = |P|;

Count_b = | E_b |;

for(k)

$P_i = P$;

$P_{ib} = P_i$ 在基本维度中长度 +1 的路径集合

$P_{id} = P_i$ 在维度 G_i 中长度 +1 的路径集合;

$P = P_{id} \cap P_{ib}$;

Count += |P|;

$$\text{Count}_b += |P_{ib}|;$$

$$\text{Sim}_{Gb}(G_i) = \text{count} / \text{count}_b;$$

依比例 r 选取部分节点为传播源点;

for (所有传播源点 v)

 节点集合 $S = v$;

 节点集合 $Q = v$ 全部维度上邻节点;

 for(k)

 for(Q 中所有节点 v')

$$\text{IN}(v') = \text{IN}(v') \cup \text{IN}(v_s) * (1 - \mu) * \text{Sim}_{Gb}(G_i);$$

 // v_s 为 v' 在 S 中的邻节点

 // i 为 v_s 与 v' 间的任选维度

$S = Q$;

$Q = Q$ 的邻节点集合 $- S$;

End

算法包含 2 个主要部分:第一部分是在指定基本维度的前提下计算其他维度与基本维度的相似性,每一维度均由维度边集之间的重合部分开始计算,即由长度为 1 的共同路径累积搜索至最长传播跳数所限定的长度;第二部分是依据输入的选择比例选取一定数量的传播源点,由每个源点出发传播信息至最长传播跳数覆盖范围,并在传播过程中考虑信息衰减,得到所有节点最终接收的信息量。

4.3 节点传播重要性的排序及分析

图 4 所示为一个交通网络关系示意图,节点代表地点,节点间的连边代表两地直接可达,其中不同线型连边表示不同维度,即不同交通方式。该网络中共有 25 个节点,以及来自 3 个维度的 52 条连边,其中:实线边与其连接的节点构成维度 1,代表两个地点之间可以通过乘坐公交的交通方式到达;虚线边与其连接的节点构成维度 2,代表两个地点间可以通过地铁的交通方式到达;双点线边与其连接的节点构成维度 3,代表两地点间可通过步行方式到达。3 个维度的边集并不完全相同,即有的地点间可以通过多种不同的交通方式直接到达,而有的地点间只有一种方式直接到达,甚至不存在直连而只能中转。

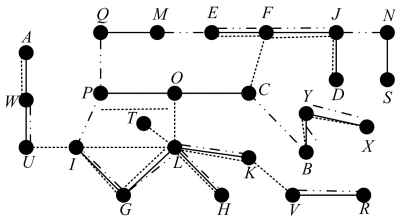


图 4 交通网络关系图

依据第 2 节中维度相似性的定义,以及衰减系数的推算结果,计算该交通网络在选择不同基本维度时各维度的传播衰减率,并取平均值为各维度传播衰减率的最终取值,如表 2 所列。

表 2 交通网络中各维度传播的衰减率

基本维度	计算维度		
	维度 1	维度 2	维度 3
维度 1	0.18	0.32	0.38
维度 2	0.50	0.18	0.55
维度 3	0.51	0.51	0.18
平均	0.40	0.34	0.37

分别将该交通网络中的每个节点作为传播源点进行广播,其携带的初始信息量相同。经过信息传播后得到节点重

要性程度的排序结果,将重要性前 8 的节点与按节点度以及节点介数方法得到的节点重要性排序结果进行比较,如表 3 所列。

表 3 计算节点重要性的不同方法的结果对比

次序	度数	节点介数	3 个源节点	5 个源节点	全部源节点
1	L	L	L	L	L
2	F	K	G	G	G
3	G	F	K	F	F
4	I	E	F	K	K
5	J	I	I	I	I
6	Y	J	J	J	J
7	O	G	E	E	E
8	P	Y	O	O	O

在该交通网络中,如果算法可以按照各个地点的便利程度来对不同的地点进行排序,则认为算法是准确有效的。由表 3 给出的排序结果可知,节点介数体现的重要性排序与本文算法选取全部节点作为传播源点所得结果大致相同,但部分节点存在差异。介数法排序中认为节点 K 比节点 G 更重要,但从实际的多维网络角度分析,虽然两个节点都与节点 L 直接相连,但 G 节点的度数比 K 节点的度数大,也就是说,实际生活中可以通过更多的交通方式从地点 G 到达其他地方,所以地点 G 更便利。介数排序认为节点 E 比节点 I 重要,实际上节点 I 的度大于节点 E 的度,而且节点 I 与网络中度最大的节点 L 直接相连,所以节点 I 比节点 E 更重要。用度排序的方法认为节点 F 比节点 G 更重要,实际上节点 G 与网络中最重要节点 L 直接相连,也就是说节点 G 到达最重要的地点需要经过的最短路径更少,所以节点 G 比节点 F 更重要。根据上述示例计算的对比结果可以看出,相对于传统方法,采用本文计算节点传播重要性的方法能够更为准确地刻画节点的重要性。

在实际的计算过程当中,选取像节点 A 和节点 R 这种重要程度较低的节点作为传播源点的意义不大,且选择较少数量的传播源点进行广播比较节点重要性,能够大大提高计算效率。因此,进一步选取该交通网络中度数最大的前 3 个节点 L, F, G 以及前 5 个节点 L, F, G, J, I 作为传播源点进行广播,得到节点重要性前 8 的节点排序结果亦在表 3 中给出。从对比结果来看,该交通网络选取部分节点作为传播源点得到的结果与选取所有节点作为传播源点得到的结果相差较小,保持与网络实际情况基本一致,相对于介数方法以及度方法仍具有一定优势。

5 实验结果与分析

本文选取 WikiElec 网络与 Alex Arenas 网络进行实验对比分析。Alex Arenas 网络^[18]是由 Rovirai Virgili(Tarragona)大学的某个研究小组的 E-mail 网络数据组成,用于分析研究小组内的社会个体及社会关系。由于这两个网络数据中并不包含节点重要性,因此本文在其上模拟用户日常行为,进行信息的随机扩散,以较长时间后节点所获信息量排序为基准数据,具体过程为:

1) 网络中每个节点初始持有 1 个单位的信息,且各节点信息均不相同。令每个节点以随机的时间间隔(100~10 000 ms)等概率选取随机的邻居节点以及随机的维度传递信息,经过足够长的时间后,按节点所持有不同信息的总数排序,即为节

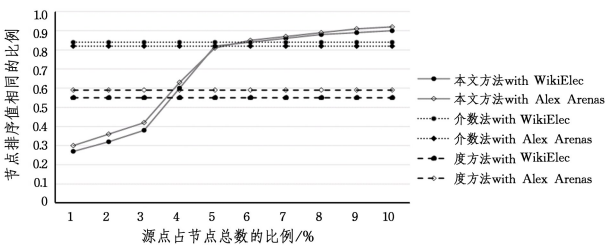
点重要性基准排序。

2)网络中每个节点初始持有同样大小的信息集,且各信息集内容均不相同,与方法1)类似,令每个节点随机传递信息,但每跳传递均有信息衰减,经过足够长的时间后,按节点所持有信息集的大小排序,即为节点重要性基准排序。

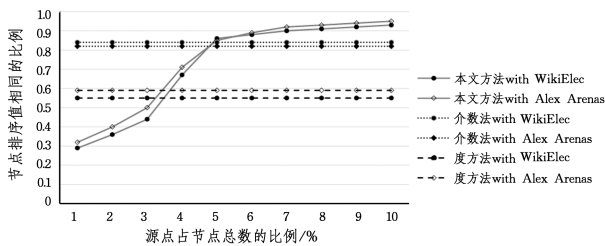
将方法1)与方法2)分别在 WikiElec 网络与 Alex Arenas 网络上运行足够长时间(本节实验中运行时间为 72h),即可得到本文实验对比所用的 4 组基准数据。在不同最长传播跳数以及不同传播源点数目情况下运行本文算法,分别取不同维度为基本维度进行实验,并以各次实验结果的均值为最终结果。通过在精确度与时间效率 2 个方面与节点介数及节点度方法进行比较,论证本文方法的优越性。

5.1 传播源点

考虑到实验数据集规模较大,且由 4.3 节分析可知,选取部分节点作为传播源点来计算节点重要性,其误差在可接受范围之内。因此,本节实验中选取度值最大的 2%~10% 的节点为传播源点,结合 4.1 节分析结果,令传播跳数 $k=6$ 。图 5 给出不同传播源点比例下本文方法的计算结果与节点介数以及节点度方法的精确度对比,其中图(a)为基准数据与本文算法的运行均不考虑传播过程中信息衰减的情况,图(b)为考虑信息衰减且衰减系数按第 3 节所得设为 0.18 的情况。节点介数与节点度方法不需要选取部分源点,所以其准确程度及运行时间效率均为固定值,于图中显示为平行于横轴的直线。



(a) 不考虑信息衰减情况



(b) 考虑信息衰减情况

图 5 选取不同的传播源点计算节点重要性的准确程度

由图 5 可以看出,当选取的传播源点占节点总数的比例较低时,无论是否考虑信息衰减,节点重要性的计算精确度均较低。如图 5(a)所示,在不考虑信息衰减的情况下,当选取的传播源点占比达到 5% 时,在 2 个实验数据集上的计算精确度均超过 80%,接近节点介数方法的精确度,且优于节点度方法。而当选取的传播源点占比为 10% 时,在 WikiElec 数据集和 Alex Arenas 数据集上的计算精确度分别达到 90% 和 92%。在图 5(b)所示的考虑信息衰减的情况下,当选取的传播源点占比达到 5% 时,在 2 个实验数据集上的计算精确度均接近 85%,优于节点介数及节点度方法。而当选取的传播源点占比为 10% 时,在 WikiElec 数据集和 Alex Arenas 数据

集上的计算精确度分别达到 93% 和 95%。可以看出,本文方法在计算节点重要性的精确度方面具有一定优势,特别是在考虑信息衰减即更为符合实际网络的情况下优势更为明显。

图 6 给出了不同传播源点比例下本文方法与节点介数方法的时间效率对比,考虑到节点度方法的时间效率极佳,实验中不与其进行对比。另外,由于是否考虑信息衰减对算法的时间效率几乎没有影响,因此仅在考虑信息衰减的情况下进行对比。

由图 6 可以看出,当选取的传播源点占节点总数的比例较低时,节点重要性的计算时间效率均较高,且 WikiElec 数据集优于 Alex Arenas 数据集。当选取的传播源点占比达到 5% 时,在 2 个实验数据集上的运行时间增长均接近一倍,而当选取的传播源点占比为 10% 时,在 2 个实验数据集上的运行时间均接近节点介数方法。显然,若继续增加传播源点占比,本文方法的时间效率将低于节点介数方法。

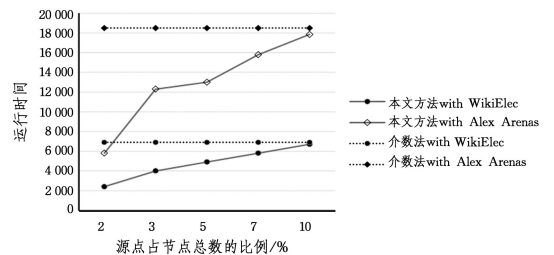
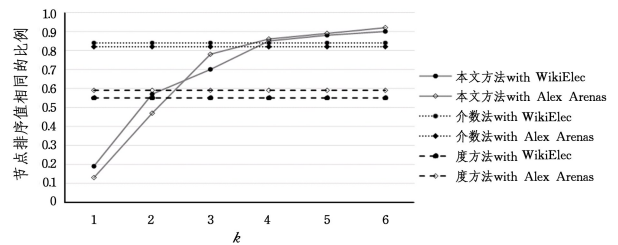


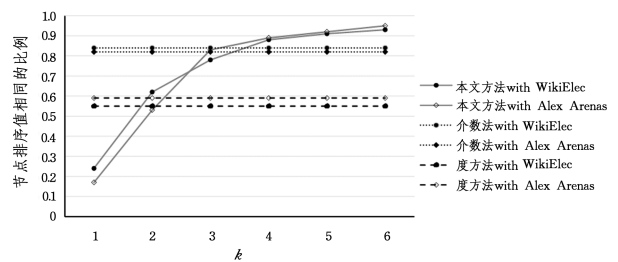
图 6 选取不同的传播源点计算节点重要性的效率

5.2 最大传播跳数

本节实验中令最大传播跳数 k 取值 1~6,结合 5.1 节实验结果,选取度值最大的前 10% 的节点为传播源点。图 7 给出不同最大传播跳数下本文方法的计算结果与节点介数以及节点度方法的精确度对比,其中图 7(a)为基准数据与本文算法的运行均不考虑传播过程中信息衰减的情况,图 7(b)为考虑信息衰减的情况。与 5.1 节类似,节点介数与节点度方法不涉及最大传播跳数,所以其准确程度及运行时间效率均为固定值,于图中显示为平行于横轴的直线。



(a) 不考虑信息衰减情况



(b) 考虑信息衰减情况

图 7 k 取不同值时准确度的变化

由图 7 可以看出,当 k 取值过小时,精确度也非常低。在图 7(a)中,当 $k=4$ 时,在 2 个实验数据集上的计算精确度均

接近 85%,略优于节点介数方法,大幅优于节点度方法。在图 7(b)所示的考虑信息衰减的情况下,当 $k=4$ 时,在 2 个实验数据集上的计算精确度均接近 90%,优势更加明显。而无论是否考虑信息衰减,当 $k=6$ 时均与 5.1 节实验结果相同,不再赘述。

图 8 给出了最大传播跳数 k 取不同值时本文方法与节点介数方法的时间效率对比,与 5.1 节相同,仅在考虑信息衰减的情况下进行对比。

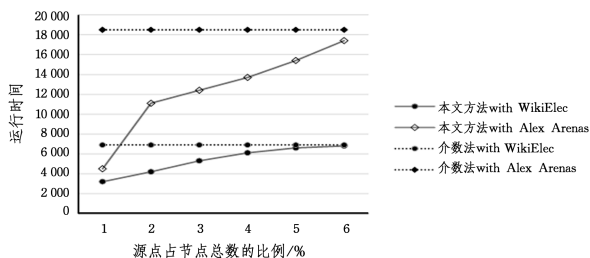


图 8 k 值不同时效率的变化

由图 8 可以看出,当 k 值较小时,节点重要性的计算时间效率均较高,且 WikiElec 数据集优于 Alex Arenas 数据集。当 $k=4$ 时,在 WikiElec 数据集上的运行时间增长约一倍,在 Alex Arenas 数据集上运行时间增长约 2 倍。而当 $k=6$ 时,在 2 个实验数据集上的运行时间均接近节点介数方法。因此可以得出与 5.1 节类似的结论,即若 k 值继续增加,则本文方法的时间效率将低于节点介数方法。

综合看来,本文方法选取的传播源点数越多,最大传播跳数 k 取值越大,则算法的准确程度越高,但运行时间也随之增加。当传播源点占比与最大传播跳数位于适当的区间时,本文方法在计算精确度与时间效率方面均占据明显优势;且相对于信息传播无衰减的理想情况,本文方法在考虑信息衰减的情况下优势更为突出,因此更加适用于实际网络中节点重要性的计算。

结束语 本文研究多维网络中节点重要性的计算方法,重点针对无权无向网络中的信息传播重要性进行研究。主要工作包括:

1) 定义了维度间的共同路径,并进一步定义维度相似度,以此来度量不同维度间的信息传播差异;

2) 为了更为贴近实际网络中信息传播的真实情况,考虑传播过程中信息的衰减,引入传播衰减率等相关概念,并依据全连接单维网络传播无损假设,确定了衰减系数的具体取值;

3) 考虑实际网络普遍具有的小世界特性,提出限定最长传播跳数,从而可以在基本保证计算精确度的同时大幅提高计算的时间效率;

4) 给出节点传播重要性的计算方法,并通过在实际网络上的对比实验,验证本文方法在计算精确度与时间效率方面都优于传统方法,特别是在考虑信息衰减的情况下,优势更为明显。

本文考查了多维网络中维度间的相似性,并基于维度间的共同路径对其进行了量化。考虑到实际网络的复杂性,可以进一步综合更多因素,使维度间相似性的度量更为精确,并在算法效率上对最短路径的提取可以做出进一步优化。另

外,本文工作仅针对无权无向网络,而实际网络中的连接大多具有权重,且部分为有向网络,因此针对有权有向网络的研究具有重要意义。

参考文献

- [1] 汪小帆,李翔,陈关荣. 复杂网络理论及其应用[M]. 北京:清华大学出版社有限公司,2006.
- [2] COSTA L F, OLIVEIRA JR O N, TRAVIESO G, et al. Analyzing and modeling real-world phenomena with complex networks: a survey of applications[J]. *Advances in Physics*, 2011, 60(3): 329-412.
- [3] PEROZZI B, AL-RFOU R, SKIENA S. Deepwalk: Online learning of social representations[C]// *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2014: 701-710.
- [4] 周涛,柏文洁,汪秉宏,等. 复杂网络研究概述[J]. *物理*, 2005, 34(1).
- [5] FREEMAN L C. A set of measures of centrality based on betweenness[J]. *Sociometry*, 1977: 35-41.
- [6] BOCCALETTI S, BIANCONI G, CRIADO R, et al. The structure and dynamics of multilayer networks[J]. *Physics Reports*, 2014, 544(1): 1-122.
- [7] YU H, LIU Z, LI Y J. Key nodes in complex networks identified by multi-attribute decision-making method[J]. 2013.
- [8] MENICHETTI G, REMONDINI D, PANZARASA P, et al. Weighted multiplex networks[J]. *PloS one*, 2014, 9(6): e97857.
- [9] SHCHUROV A A. A multilayer model of computer networks [J]. arXiv:1509.00721, 2015.
- [10] BERLINGERIO M, COSCIA M, GIANNOTTI F, et al. Foundations of multidimensional network analysis[C]// *2011 International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. IEEE, 2011: 485-489.
- [11] 杨建祥,王朝坤,王萌,等. 全动态多维网络局部介数中心度算法[J]. *计算机学报*, 2015(9): 254-4164
- [12] BERLINGERIO M, COSCIA M, GIANNOTTI F, et al. Multidimensional networks: foundations of structural analysis [J]. *World Wide Web*, 2013, 16(5/6): 567-593.
- [13] WATTS D J, STROGATZ S H. Collective dynamics of 'small-world' networks[J]. *Nature*, 1998, 393(6684): 440.
- [14] HENDRY J B. *Rural Vietnam: The Small World of Khanh Hau* [M]. Routledge, 2017.
- [15] BO S, GUO-PING J, YU-RONG S, et al. Rapid identifying high-influence nodes in complex networks[J]. *Chinese Physics B*, 2015, 24(10): 100101.
- [16] XU B, QI J, ZHOU C, et al. Hybrid self-adaptive algorithm for community detection in complex networks [J]. *Mathematical Problems in Engineering*, 2015, 2015.
- [17] SUH B, CONVERTINO G, CHI E H, et al. The singularity is not near: slowing growth of Wikipedia[C]// *International Symposium on Wikis and Open Collaboration*. DBLP, 2009: 1-10.
- [18] LOZANO S, ARENAS A, SÁNCHEZ A. Mesoscopic structure conditions the emergence of cooperation on social networks[J]. *PloS One*, 2008, 3(4): e1892.