

# 融合多元信息的多关系社交网络节点重要性研究

罗 浩<sup>1</sup> 闫光辉<sup>1</sup> 张 萌<sup>1</sup> 包峻波<sup>1</sup> 李俊成<sup>1</sup> 刘 婷<sup>1</sup> 杨 波<sup>2</sup> 魏 军<sup>2</sup>

<sup>1</sup>(兰州交通大学电子与信息工程学院 兰州 730070)

<sup>2</sup>(国网甘肃省电力公司信通公司 兰州 730050)

(luoh382@163.com)

## Research on Node Importance Fused Multi-Information for Multi-Relational Social Networks

Luo Hao<sup>1</sup>, Yan Guanghui<sup>1</sup>, Zhang Meng<sup>1</sup>, Bao Junbo<sup>1</sup>, Li Juncheng<sup>1</sup>, Liu Ting<sup>1</sup>, Yang Bo<sup>2</sup>, and Wei Jun<sup>2</sup>

<sup>1</sup>(School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070)

<sup>2</sup>(Information and Telecommunication Company, State Grid Gansu Electric Power Company, Lanzhou 730050)

**Abstract** Identifying critical nodes is one of the principal tasks of social network analysis, and it is essential to understand the structure and dynamic characteristics of the complex networks. However, the analysis framework of node importance mainly focuses on single-relational networks. As a typical model of the real world, the multi-relational network has become one of the hot topics in the field of network science, in which the research on node importance lacks systematic research. Focusing on the study of node importance in multi-relational social networks, we create the directed multiplex network model to describe a multi-relational network and use the representation framework based on tensor algebra to analyze it. Meanwhile, we propose a measure of node importance considered the influence of centrality, prestige, transitivity in multi-relational social networks. Considering the influence of coupling information and the difference of transmission mechanism for node importance on multi-relational networks, in this work we improve the method and propose another more efficient method called IOMEC to evaluate the node importance. Experimental results on four real networks show that the method of information fusion can effectively eliminate the influence on node importance evaluation, which is caused by the coupling information and the transmission mechanism of the multi-relational network. The IOMEC method can measure the importance of nodes more accurately and has lower time complexity. The experimental results demonstrate that centrality and prestige are the main factors to evaluate the node importance and the necessity of considering the transitivity of nodes. In this work we not only provide new ideas and methods for evaluating node importance for multi-relational networks but also expand the application of information fusion technology.

**Key words** multi-relational network; social network; node importance; centrality; prestige; transitivity; information fusion; Dempster-Shafer theory of evidence

收稿日期:2019-06-03;修回日期:2019-11-25

基金项目:国家自然科学基金项目(61662066,61163010);甘肃省青年科技基金计划项目(1606RJYA222)

This work was supported by the National Natural Science Foundation of China (61662066, 61163010) and the Technique Foundation Program for Young Scientists of Gansu Province (1606RJYA222).

通信作者:闫光辉(yanghacademic@163.com)

**摘要** 识别重要节点是社会网络分析领域的重要任务之一,也是理解复杂网络结构和动力学特性的有效方式,迄今发展起来的节点重要性分析框架主要面向单关系网络.多关系网络作为准确刻画现实世界复杂系统的典型建模形式,已成为当前网络科学领域研究的热点,但对于多关系网络的节点重要性研究尚缺乏系统性的研究成果.针对多关系社交网络节点重要性研究问题,通过构建有向多重网络模型和基于张量代数的数学框架对其进行建模和分析,将中心性、声望和传递性作为影响社交网络节点重要性的关键因素,提出了一种面向多关系社交网络的节点重要性度量指标,并针对其存在不足引入 D-S (Dempster-Shafer) 证据理论进行改进,进一步提出了 IOMEC (in-degree out-degree multiplex evidential centrality) 节点重要性度量方法.在 4 个真实网络上的实验结果表明:采取信息融合的方法可以有效消除多关系网络耦合信息和传递机制对节点重要性评测造成的影响,提出的 IOMEC 方法能够更准确地对节点重要性进行度量,并且具有较低的时间复杂度,在论证节点中心性和声望是衡量节点重要程度主要因素的同时,说明了综合考虑节点传递性的必要性.所做工作作为多关系网络节点重要性研究提供新的思路方法的同时,进一步拓展了信息融合技术的应用场景.

**关键词** 多关系网络;社交网络;节点重要性;中心性;声望;传递性;信息融合;D-S 证据理论

**中图法分类号** TP18; TP301

Facebook, Twitter 等大规模社交网络的兴起,为社会网络研究提供了数据支撑,而复杂网络工具的引入,为社会网络分析提供了定量基础.节点重要性研究作为社会网络分析的关键问题之一,迄今发展起来的理论框架主要面向单关系网络结构.作为一种简化的网络科学框架,单关系网络结构不足以准确刻画现实世界中相互关联、彼此依存的复杂系统,即网络的网络(networks of networks, NoN),对 NoN 的研究已逐渐成为复杂网络分析领域富有挑战性的研究热点<sup>[1-2]</sup>.多关系网络(multi-relational networks)作为 NoN 的一种典型存在方式,已成为刻画现实世界各类关系构成复杂系统的常用模型.近年来,国内外学者面向多关系网络,围绕基本度量<sup>[3-6]</sup>、社团结构<sup>[7-10]</sup>、传播扩散<sup>[11]</sup>、谱特征分析<sup>[12]</sup>等基础问题进行了大量研究,也形成了较为丰富的理论成果.但针对多关系网络节点重要性的研究,大多仍是通过构建多个单关系网络,并通过某种规则进行复合的方式进行建模和分析,这种处理方式会丧失各类关系之间的特定结构信息.如何精确刻画多关系网络,探索推广单个网络上已有的节点重要性度量方法到多关系网络中,研究设计符合多关系网络结构特点和动力学特征的节点重要性分析方法,具有重要的理论价值和应用价值.

本文针对多关系社交网络的节点重要性问题,着眼于在线社交平台产生的海量关系数据,通过构建多关系有向网络模型,应用基于张量代数的多层次网络数学框架对其进行表示和分析,融合刻画节点中心性、声望、传递性等影响社交网络节点重要性

因素的度量指标,借鉴单关系网络 ClusterRank, IO-ClusterRank 等节点重要性方法和基于 D-S (Dempster-Shafer) 证据理论的信息融合方法,对多关系有向网络的节点重要性进行了系统研究,提出了面向多关系有向网络的 IOMCR (in-degree out-degree multiplex ClusterRank) 和 IOMEC (in-degree out-degree multiplex evidential centrality) 方法.在 4 个大规模多关系真实网络数据集上的实验结果表明,由于受多关系网络耦合信息和传递机制的影响,直接借鉴 ClusterRank 和 IO-ClusterRank 所提出的 IOMCR 方法,会对节点重要性测量的准确性造成影响,而运用 D-S 证据理论进行多元信息融合的 IOMEC 方法能够有效规避这种影响,相比于其他方法,能够更加准确地测量节点的重要程度.在时间复杂度方面,相对于 IOMCR, IOMEC 方法由于引入了信息融合技术,具有更低的时间复杂度.本文所做工作不仅丰富了多关系网络节点重要性的度量方法,也扩展了多关系网络的基本度量和 D-S 证据理论的实际应用场景.

## 1 相关工作

多关系网络<sup>[13]</sup>的概念起源于社会学领域,社会学家根据行动者(actor)之间的关系属性,对网络中的边进行分类<sup>[14-15]</sup>. Mitchell<sup>[16]</sup>将此类网络称之为拥有“多重(multi-stranded)”关系的网络.被用来研究多关系社会网络的方法包括指数随机图模型、元网络(meta-networks)、元矩阵(meta-matrices),

以及通过块模型 (block modeling) 和关系代数 (relational algebras) 来确定社会角色的其他方法等<sup>[17-18]</sup>。但社会学领域所研究的网络规模相对较小, 定量分析复杂度较低, 且常采用调查的方法对结果进行统计分析, 不适用于大规模网络。

对大规模多关系网络进行研究, 一个基本问题是如何对其进行建模和表示。随着复杂网络分析技术的不断发展, 对于单个网络, 已经建立了较为完备的理论分析框架。鉴于此, 对多关系网络的处理, 起初的处理思路是利用“降维”的思想, 围绕特定领域网络, 针对性地运用社会学、可拓学等其他学科理论, 通过构建加权网络的方式将多关系网络转化为单关系网络<sup>[19-20]</sup>; 或是借助矩阵分解、多子网复合等理论方法对已有网络进行时间划分或快照处理, 对单个网络分别进行研究后再进行复合<sup>[21-22]</sup>。此类简化的处理方式, 会导致原始数据中包含的特定结构信息和不同类型关系之间关联信息的丧失, 不利于后续研究的开展。一些研究围绕“全连通网络”, 运用超图理论开展分析<sup>[20]</sup>, 虽能够对网络进行融合, 保留了信息的完整性, 但“超图”模型通常用于刻画“全连通网络”, 局限性较强。

在计算机科学和计算机线性代数领域, 研究者们运用张量分解 (tensor-decomposition)<sup>[23-24]</sup> 和多路数据 (multiway data) 分析<sup>[25]</sup> 的方法研究了不同类型的多层网络。在物理学领域, 物理学家们围绕多层网络<sup>[26-29]</sup>、网络的网络<sup>[30-31]</sup> 和节点着色网络 (node-colored networks), 已经开展了许多重要的基础性工作。Boccaletti 等人<sup>[32]</sup> 于 2014 年首次提出了多层网络的基本数学定义, 该定义统一了迄今文献上关于多层网络的不同提法。经过近几年的发展, 多层网络模型已逐渐成熟, 形成了一系列符合复杂网络基本特征和动力学机制的理论方法, 为后续研究奠定了基础。对于多关系网络, 可将每一类关系对应于一个网络层, 同时考虑各类关系间的关联信息, 建立多层网络模型开展相关研究。因此, 多层网络模型是表示多关系网络的理想形式, 将多层网络模型运用到多关系网络中, 在应用层面进行拓展, 是值得深入研究的一个基本问题。

在社会网络分析领域, 中心性、声望和传递性是制约社会网络中行动者显著性或重要性的关键因素<sup>[13]</sup>。在复杂网络分析领域, 对于单个网络节点中心性的度量, 已经建立了较完备和丰富的指标和方法<sup>[33]</sup>。根据局部性质划分, 可分为全局中心性、局部中心性、半局部中心性 3 类。全局中心性的经典指标

有介数中心性 (betweenness centrality, BC)<sup>[34]</sup>、接近中心性 (closeness centrality, CC)<sup>[35]</sup>、特征向量中心性 (eigenvector centrality, EC)<sup>[36]</sup> 等。局部中心性的典型指标有度中心性 (degree centrality, DC)<sup>[37]</sup>。为平衡全局中心性的复杂度和局部中心性的因素单一, Chen 等人<sup>[38]</sup> 提出了一种半局部中心性 (semi-local centrality, SLC), 通过考虑节点的 4 阶邻居信息来度量节点重要程度, Gao 等人<sup>[39]</sup> 在其基础上通过引入 D-S 证据理论, 提出了一种证据半局部中心性 (evidential semi-local centrality, ESC)。但上述经典指标都只关注了节点的中心性, 而忽略了传递性在度量重要节点的关键作用。通常用节点的局部聚集系数刻画节点的传递能力, Chen 等人<sup>[40]</sup> 综合考虑了节点的中心性和传递性因素, 同时考虑了节点的二阶邻居对节点重要性造成的影响, 提出了 ClusterRank 指标, 该指标在运行效率和准确度上都要优于前者。声望的概念, 主要通过可以区分行动者发出或者接收“选择”关系的行为进行量化, 因此只能利用有向网络进行研究。在复杂网络分析中, 通常用节点的出度来刻画中心性, 用节点的入度来刻画声望<sup>[13]</sup>。遗憾的是, 声望除用于早期社会网络分析外, 现阶段对有向网络节点重要性的研究中, 研究者们更加关注节点的出度信息, 关注节点的入度信息的研究很少。值得一提的是, Wang 等人<sup>[41]</sup> 研究了有向网络节点出度和入度的分布关系, 针对有向加权的单关系网络提出了 IO-ClusterRank 指标, 其对节点重要性的度量效果要优于传统的 ClusterRank 指标。

面向多关系网络, 如何借用已有的物理量去描述其基本拓扑结构和性质尚面临巨大挑战, 将单关系网络中的成熟工具和方法扩展到多关系网络中已成为当前复杂网络研究的一大热门方向。最初的做法是将不同关系的节点和连边以不同权重的方式进行合并, 从而将多关系网络转化成为单关系网络, 再利用传统物理量去分析其各种结构属性。将多关系网络合并成单关系网络, 会导致一些关系间重要信息的丢失, 同时还存在如何设定各种权重的问题<sup>[42]</sup>。因此, 发现一套适合多关系网络的网络理论框架的必要性开始凸显。多层网络模型的提出为多关系网络的研究奠定了理论基础, 国内外学者针对多层网络度量指标的研究做了大量研究工作, 也取得了一些成果, 多层网络的中心性、路径长度、聚集系数等概念被相继提出, 为后续研究奠定了基础。但多层网络的度量指标, 大多数是在特定条件和应用场景下提出的, 尚未形成类似于单关系网络丰富经

典的度量指标体系,其可扩展性需进一步增强.值得一提的是,Battiston 等人<sup>[43]</sup>在向量的基础上提出了多层网络节点度(node degree)、节点强度(node strength)等概念,Cozzo 等人<sup>[44]</sup>从路径和环的概念出发,基于多层网络的传递性提出了一种多层网络聚集系数(clustering coefficient)的概念,二者都是从多层网络的基本表征和动力学特性出发的,具有较好的可扩展性,成为目前比较经典的多层网络度量指标,为后续研究提供了基本遵循.特别是在传递性研究方面,通过在实际数据上进行实验和分析,表明在社会网络和交通网络中,层内聚集性和层间聚集性的差异是不同的,这实际上反映了在这些网络中节点传递性的起源机制大不一样.他们提供的聚集系数定义表达式为不同类型的网络传递性研究提供了新的思路.

综上,现阶段对于多关系网络的研究比较理想的处理方式是构建多层网络模型并运用相关指标进行分析,而对于多层网络 2 种较为经典的度量指标,主要是围绕多层网络的拓扑结构和传递机制进行研究,如何从多层网络的基本模型和表示框架出发,运用这些指标对多关系网络的节点重要性开展研究还存在很多问题,也是本文工作的出发点之一.

## 2 基础理论

本节主要对文中用到的基础理论和符号记法进行介绍和说明.

### 2.1 爱因斯坦求和约定和符号记法

在张量计算体系中,为使计算公式简洁,通常将求和符号 $\sum$ 省略,即采用爱因斯坦求和约定.满足 2 个规则<sup>[45]</sup>:

#### 1) 哑指标规则

① 在同一项中,以一个上标与一个下标成对地出现,表示遍历其取值范围求和.如:

$$A_a^a = \sum_{a=1}^N A_a, \quad (1)$$

$$A^a B_a = \sum_{a=1}^N A_a B_a, \quad (2)$$

$$A_\beta^a B_\alpha^\beta = \sum_{\alpha=1}^N \sum_{\beta=1}^N A_{\alpha\beta} B_{\alpha\beta}. \quad (3)$$

② 每一对哑指标的字母可以用相同取值范围的另一对字母任意替换,其意义不变(必须指代明确).如:

$$A^a B_a = A^\beta B_\beta, \quad (4)$$

$$A_\beta^a B_\gamma^\beta C_\alpha^\gamma \neq A_\beta^a B_\beta^\beta C_\alpha^\beta. \quad (5)$$

在式(5)中,由于已经存在 $\beta$ ,若将 $\gamma$ 替换为 $\beta$ ,则会造成指代不明.

③ 涉及张量积之间比率的方程式,应将爱因斯坦求和约定分别应用到分子与分母.如:

$$\frac{A_\beta^a B_\gamma^\beta}{C_\beta^a A_\alpha^\beta} = \frac{\sum_\beta A_{\alpha\beta} B_{\beta\gamma}}{\sum_\alpha \sum_\beta C_{\alpha\beta} A_{\beta\alpha}}. \quad (6)$$

#### 2) 自由指标规则

若一个指标在表达式的各项中都在同一水平上出现并且只出现一次,或者全为上标,或者全为下标,表示该表达式在该自由指标的 $N$  维取值范围内都成立,即代表了 $N$  个表达式:

$$A_\beta^a B_\gamma^\beta = \sum_{\beta=1}^N A_\beta^a B_\gamma^\beta, \quad (7)$$

其中, $\beta$  为哑指标, $\alpha$  和 $\gamma$  均为自由指标.

同理,一个表达式中的某个自由指标可以全体地换用相同取值范围的其他字母,意义不变.

在欧几里德空间中,张量的协变向量通常用列向量 $a \in \mathbb{R}^N$  来表示,其分量记为 $a^\alpha$ ,( $\alpha = 1, 2, \dots, N$ ),对偶向量(逆变向量)可用对应的转置向量(行向量)表示,其分量记为 $a_\alpha$ .

为避免混淆,本文对张量代数框架下的符号记法做 2 个约定:

1) 用拉丁字母(如 $i, j, k \dots$ )来表示张量;用希腊字母(如 $\alpha, \beta, \gamma \dots$ )来表示张量分量.例如 $a_i^\alpha$  表示第 $i$  个张量的分量.

2) 用不带波浪线的字母(如 $i, \alpha \dots$ )指代网络中的节点;用带波浪线的字母( $\tilde{i}, \tilde{\alpha} \dots$ )指代网络中的关系类别.

### 2.2 ClusterRank 与 IO-ClusterRank 指标

节点的局部聚集系数与信息传播的范围及网络进化中节点的度增长能力均呈负相关.鉴于此,Chen 等人<sup>[40]</sup>提出的 ClusterRank 指标是一种针对有向网络节点重要性的半局部中心性指标,该指标不仅将邻居节点数量作为衡量中心性的一个指标,同时考虑了传递性对节点重要程度的影响,即聚集系数越大节点的传播能力越弱.ClusterRank 算法定义为

$$Cr_i = f(c_i) \sum_{j \in \Gamma_i} (k_j^{\text{out}} + 1), \quad (8)$$

$$c_i = \frac{N_{jk}}{k_i^{\text{out}} (k_i^{\text{out}} - 1)}, \quad (9)$$

其中, $k_j^{\text{out}}$  为节点 $v_j$  的出度, $\Gamma_i$  为节点 $v_i$  的邻居集合, $f(c_i)$  为 $v_i$  的局部聚集系数 $c_i$  的线性负相关函

数,  $c_i$  越大,  $f(c_i)$  越小. 式(9)中,  $N_{jk}$  为  $\Gamma_i$  中节点  $v_j$  和  $v_k$  存在连边的数量. 通过实验对比, ClusterRank 算法优于 PageRank 和 LeaderRank 算法, 并且计算复杂度与后二者比较相对较低.

但是, ClusterRank 指标只考虑了节点的出度, 并没有考虑节点的入度信息, Wang 等人<sup>[41]</sup>通过对有向网络出入度的分析, 基于 ClusterRank 提出了一种针对加权有向网络的节点重要性度量指标 IO-ClusterRank, 定义为

$$r_i = \frac{1}{k_i} (k_i^{\text{out}} r_i^{\text{out}} + k_i^{\text{in}} r_i^{\text{in}}), \quad (10)$$

$$r_i^{\text{out}} = f(c_i^{\text{out}}) \sum_{j \in \Gamma_i^{\text{out}}} (s_j^{\text{out}} + w_{ij}), \quad (11)$$

$$r_i^{\text{in}} = f(c_i^{\text{in}}) \sum_{j \in \Gamma_i^{\text{in}}} (s_j^{\text{in}} + w_{ij}), \quad (12)$$

其中,  $k_i^{\text{out}}$  和  $k_i^{\text{in}}$  分别为节点  $v_i$  的出度和入度,  $c_i^{\text{out}}$  和  $c_i^{\text{in}}$  分别为节点  $v_i$  出聚集系数和入聚集系数,  $w_{ij}$  为节点  $v_i$  和  $v_j$  的权重,  $s_j^{\text{out}}$  和  $s_j^{\text{in}}$  分别为节点  $v_j$  的出强度和入强度. 实验结果表明: IO-ClusterRank 指标在出入度差异较大的网络中具有明显优势, 从而说明在节点重要性研究中考虑入度信息的必要性.

### 2.3 D-S 证据理论

D-S 证据理论是由 Dempster 和 Shafer<sup>[46]</sup> 建立的一套数学理论, 是对概率论的进一步扩充, 在表示和处理有限个不确定问题时具有明显优势, 适用于专家系统、人工智能、模式识别和系统决策等领域的实际问题. 本节主要介绍证据理论处理流程涉及到的辨识框架(frame of discernment)、基本概率指派(basic probability assignment, BPA) 以及 Dempster-Shafer 证据组合规则等基本概念和计算方法.

**定义 1.** 辨识框架. 在证据理论中, 通常将所考察判断对象的全体用一个两两互斥的有限完备集合来表示, 该集合称为辨识框架, 记为

$$\Psi = \{\psi_1, \psi_2, \dots, \psi_N\}, \quad (13)$$

$\Psi$  的幂集  $2^\Psi$  记为

$$2^\Psi = \{\emptyset, \psi_1, \dots, \psi_N, \psi_1 \cup \psi_2, \dots, \psi_1 \cup \psi_2 \cup \psi_3, \dots, \Psi\}. \quad (14)$$

**定义 2.** 基本概率指派. 设  $\Psi$  是辨识框架,  $2^\Psi$  构成命题集合,  $\forall A \subseteq \Psi$ , 若函数  $m: 2^\Psi \rightarrow [0, 1]$  满足 2 个条件:

$$\begin{cases} m(\emptyset) = 0, \\ \sum_{A \subseteq \Psi} m(A) = 1, \end{cases} \quad (15)$$

则称  $m$  为基本概率指派,  $m(A)$  为命题  $A$  的基本概率数, 被视为准确分配给  $A$  的信度.

**定义 3.** Dempster-Shafer 组合规则. 设  $m_1$  和  $m_2$  为 2 组基本概率指派, 对应的对象分别为  $A_1, A_2, \dots, A_k$  和  $B_1, B_2, \dots, B_l$ , 用  $m$  表示  $m_1$  和  $m_2$  组合后的新证据. Dempster-Shafer 组合规则定义:

$$\begin{cases} m(\emptyset) = 0, \\ m(A) = \frac{1}{1-k} \sum_{A_i \cap B_i = A} m_1(A_i)m_2(B_i), \end{cases} \quad (16)$$

其中,  $k = \sum_{A_i \cap B_i = \emptyset} m_1(A_i)m_2(B_i)$  称为冲突系数, 用于衡量证据对象间的冲突程度,  $k$  越大, 则冲突越大. 当  $k=1$  时, 组合规则无法使用, 即 BPA 存在矛盾.

## 3 多关系网络的张量表示和基本度量

为研究多关系网络, 首先需构建一种精确的数学表示方法, 以及与之配套的分析工具. 本文采用基于张量代数的多层网络表示框架对多关系网络进行建模和分析. 该框架的优势在于可扩展性强, 多重网络、时序网络、相互依赖网络等均可用多层网络模型进行构建. 本节主要介绍多关系网络的基本模型和张量表示方法, 以及本文中所用到的一些基本度量指标的计算方法.

### 3.1 多关系网络的基本模型

参考 Boccaletti 等人<sup>[32]</sup> 提出的多层网络数学定义, 给出多关系网络的图定义:

**定义 4.** 多关系网络. 一般地, 多关系网络的图结构可用一个二元组表示, 记为  $G_R = (V, E_R)$ . 其中,  $V = \{v_1, v_2, \dots, v_N\}$  表示顶点集,  $N$  为网络中顶点个数, 边集  $E_R = \bigcup_{\alpha} E_R^{[\alpha]}$ ,  $\alpha = 1, 2, \dots, M$ , 其中  $E_R^{[\alpha]}$  表示第  $\alpha$  类关系的边集,  $M$  表示关系类别数量.

进一步, 针对多关系有向网络, 可将其分层表示为层内网络(intra-layer network)和层间网络(inter-layer network)两部分, 每一层对应于一种关系. 其中, 层内网络用  $G_{RI} = (V, E_{RI})$  表示, 层内边集可记为

$$E_{RI} = \bigcup_{\alpha} \{ \langle v_i^{[\alpha]}, v_j^{[\alpha]} \rangle \mid \langle v_i^{[\alpha]}, v_j^{[\alpha]} \rangle \in E_{RI}^{[\alpha]} \}, \quad (17)$$

其中,  $\langle v_i^{[\alpha]}, v_j^{[\alpha]} \rangle$  表示层  $\alpha$  内, 节点  $v_i$  指向节点  $v_j$  的有向边的集合. 层间网络用  $G_{RC} = (V, E_{RC})$  表示, 层间边集可记为

$$E_{RC} = \bigcup_{\alpha, \beta} \{ \langle v_i^{[\alpha]}, v_i^{[\beta]} \rangle \mid (v_i^{[\alpha]}, v_i^{[\beta]}) \in E_{RC}^{[\alpha, \beta]} \}. \quad (18)$$

不同于层内网络, 由于多关系网络各层间均存在彼此的交互关系, 故层间边集中各层对应节点间为一种双向关系(也可视为无向关系).

为避免与节点标号混淆, 在多关系网络的抽象

定义中,我们用带有 $[\cdot]$ 的希腊字母 $([\alpha], [\beta] \text{ 等})$ 的上标来指示多层网络的层标号。 $E^{[\alpha]}$ 组成元素称为多层网络层内连边,区别于每个 $E^{[\alpha, \beta]}$ 的表示层间连边的组成元素.

由定义 4 可得,用于构建多关系网络的多层网络模型是一种“层间节点对齐”型的多层网络,称之为多重网络(multiplex network).图 1 展示了一个具有 3 层的无向多层网络和多重网络的示意图,可以看出,多重网络属于多层网络的特殊类型,其唯一的层间连接模式是来自不同层之间对应节点的连接.

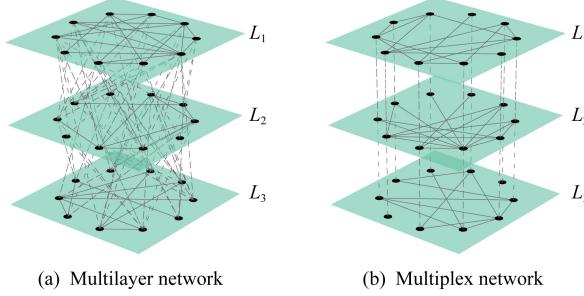


Fig. 1 Schematics of multilayer and multiplex networks

图 1 多层网络与多重网络原理图

### 3.2 多关系网络的张量表示框架

由 3.1 节可知,可用多重网络模型对多关系网络进行建模,本节主要介绍基于张量代数的多关系网络框架,本文中关于张量运算的公式符号记法采用 2.1 节介绍的爱因斯坦求和约定.下面分别介绍单层网络和多层网络的张量表示方法<sup>[47]</sup>:

#### 1) 单层网络的张量表示

有别于通常情况下的单关系网络,本文中所讨论的单层网络(monoplex networks)属于多层网络的特殊情况,需满足与时间无关和同构 2 个条件.

记单层网络为  $G_{\text{monoplex}} = \{V, E\}$ , 节点集记为  $V = \{v_1, v_2, \dots, v_N\}$ , 将节点集中每一个节点与向量空间  $\mathbb{R}^N$  中状态向量相对应,例如,节点  $v_i$  所对应的状态向量分量为:  $\mathbf{e}_i^a = (0, \dots, 1, 0, \dots, 0)^\top$ , 此式中第  $i$  个元素为 1, 其余元素全为 0, 它在向量空间  $\mathbb{R}^N$  中所对应的逆变(对偶)向量分量记为:  $\mathbf{e}_i^j = (0, \dots, 1, 0, \dots, 0)$ . 通过克罗内克积来构建描述节点间关系的张量空间,即:  $\mathbb{R}^N \otimes \mathbb{R}^N = \mathbb{R}^{N \times N}$ . 由此,二阶标准基张量可定义为

$$\mathbf{E}_{\beta i}^{aj} = \mathbf{e}_i^a \otimes \mathbf{e}_\beta^j, i, j = 1, 2, \dots, N. \quad (19)$$

为描述简洁,通常省略符号  $\otimes$ , 式(19)可记为  $\mathbf{E}_{\beta i}^{aj} = \mathbf{e}_i^a \mathbf{e}_\beta^j$ , 本文后续部分将采用省略  $\otimes$  的通用记法.

进一步,用  $a_j^i$  来表示节点  $v_i$  与  $v_j$  间的连接情况(连接为 1, 反之为 0), 则对应的邻接张量(在矩阵理论框架下称之为邻接矩阵)可由二阶张量来表示:

$$\mathbf{A}_\beta^a = a_j^i \mathbf{E}_{\beta i}^{aj}, \mathbf{A}_\beta^a \in \mathbb{R}^{N \times N}. \quad (20)$$

同理,对于加权网络,可用  $w_{\beta i}^i$  来表示节点  $v_i$  与  $v_j$  间的连接权重,则所对应的加权邻接张量为

$$\mathbf{W}_\beta^a = w_{\beta i}^i \mathbf{E}_{\beta i}^{aj}, \mathbf{W}_\beta^a \in \mathbb{R}^{N \times N}. \quad (21)$$

#### 2) 多层网络的张量表示

对于多关系网络,可用多层框架对其进行描述,其中每一类关系用单独一层  $\tilde{k}$  来表示( $\tilde{k} = 1, 2, \dots, M$ ). 用二阶张量  $\mathbf{A}_{\beta \tilde{k}}^a$  表示第  $\tilde{k}$  层层内邻接张量(即同一层内节点间的关系).

对于多层网络,来自层  $\tilde{h}$  的节点  $v_i$  可以连接到任意一层  $\tilde{k}$  的任意节点  $v_j$ , 为了表示包含多个层关系的信息,引入二阶层间邻接张量,记为  $\mathbf{C}_{\beta \tilde{k}}^{a \tilde{h}}$ , 特别地,对应于一对表示相同层  $\tilde{k}$  的层间邻接张量等于该层的层内邻接张量,即:  $\mathbf{C}_{\beta \tilde{k}}^{a \tilde{k}} = \mathbf{A}_{\beta \tilde{k}}^a$ .

类似于单层网络,在空间  $\mathbb{R}^M$  中,引入标准基向量,其分量记为  $\mathbf{e}_k^{\tilde{\gamma}}$ , 其中  $\tilde{k}, \tilde{\gamma} = 1, 2, \dots, M$ . 对于空间  $\mathbb{R}^{M \times M}$ , 可构造二阶标准基张量,记为:  $\mathbf{E}_{\beta \tilde{h}}^{\tilde{\gamma} \tilde{k}} = \mathbf{e}_h^{\tilde{\gamma}} \mathbf{e}_{\beta \tilde{h}}^{\tilde{k}}$ .

根据上述定义,多层邻接张量可用层间邻接张量  $\mathbf{C}_{\beta \tilde{k}}^{a \tilde{h}}$  和二阶标准基张量来表示,即:

$$\mathbf{A}_{\beta \tilde{\beta}}^{\tilde{\gamma} \tilde{\gamma}} = \mathbf{C}_{\beta \tilde{k}}^{a \tilde{h}} \mathbf{E}_{\beta \tilde{h}}^{\tilde{\gamma} \tilde{k}} = a_{jk}^{ih} \mathbf{E}_{\beta i}^{aj} \mathbf{E}_{\beta h}^{\tilde{\gamma} \tilde{k}} = a_{jk}^{ih} \mathbf{e}_{\beta \tilde{h}}^{a \tilde{\gamma} \tilde{k}}, \quad (22)$$

其中,  $a_{jk}^{ih}$  为一实数, 表示  $\tilde{h}$  层节点  $v_i$  与  $\tilde{k}$  层节点  $v_j$  之间的连接情况(可能不对称);  $\mathbf{e}_{\beta \tilde{h}}^{a \tilde{\gamma} \tilde{k}} \equiv \mathbf{e}_i^a \mathbf{e}_\beta^j \mathbf{e}_h^{\tilde{\gamma}} \mathbf{e}_{\tilde{h}}^{\tilde{k}}$  为空间  $\mathbb{R}^{N \times N \times M \times M}$  中 4 阶标准基张量(即秩 4 张量).

由上述讨论可得,基于张量代数的表征框架适用于有向/无向、有权/无权各类网络.

为在某些情况下便于计算,在空间  $\mathbb{R}^{N \times N \times M \times M}$  中,可将多层邻接张量  $\mathbf{A}_{\beta \tilde{\beta}}^{\tilde{\gamma} \tilde{\gamma}}$  通过矩阵化(或称为展开运算)表示为一个  $NM \times NM$  的矩阵,这对于计算算法的实现是非常有用的<sup>[23]</sup>. 在已有研究中,关于多层网络的社区检测、扩散、聚集系数等研究都使用了矩阵表示法进行计算和分析<sup>[43-44, 48-50]</sup>.

事实上,根据定义 4 和克罗内克积的运算规则,可以很方便地得到该矩阵,所得结果与通过动力学扩散方程得到的矩阵是一致的<sup>[49]</sup>. 下面以构造多关系网络的邻接矩阵为例,对其构造方法予以说明:

根据定义 4,可分别构建层内网络和层间网络所对应的邻接矩阵.

**定义 5.** 层内邻接矩阵. 多关系网络第  $\alpha$  类关系所对应的层内连接可由一个邻接矩阵来表示,记为  $\mathbf{A}^{[\alpha]} \in \mathbb{R}^{N \times N}$ , 则多关系网络的层内邻接矩阵定义为

$$\mathbf{A}_{\text{RI}} = \bigoplus_{k=1}^M \mathbf{A}^{[a]} \in \mathbb{R}^{MN \times MN}. \quad (23)$$

**定义 6.** 层间邻接矩阵, 多关系网络所对应的层间连接可由一个加权邻接矩阵来表示, 记为  $\mathbf{W}_c \in \mathbb{R}^{M \times M}$ , 取单位矩阵  $\mathbf{E} \in \mathbb{R}^{N \times N}$ , 则多关系网络的层间邻接矩阵定义为

$$\mathbf{W}_{\text{RC}} = \mathbf{W}_c \otimes \mathbf{E} \in \mathbb{R}^{MN \times MN}. \quad (24)$$

由定义 5 和定义 6 可得到多关系网络的邻接矩阵, 即:

$$\mathbf{A}_{\text{R}} = \mathbf{A}_{\text{RI}} + \mathbf{W}_{\text{RC}} \in \mathbb{R}^{MN \times MN}. \quad (25)$$

### 3.3 多关系网络的基本度量

#### 1) 多关系网络的度中心性

在张量代数框架下首先给出单层网络度中心性的表示方法<sup>[47]</sup>, 并在多层网络上进行推广:

##### ① 单层网络的度中心性

首先定义“1-向量”, 即分量元素全为 1 的向量, 记为

$$\mathbf{u}^a = (1, 1, \dots, 1)^T \in \mathbb{R}^N.$$

同理,  $\mathbf{U}_a^{\beta} = \mathbf{u}_a \mathbf{u}^{\beta}$  为元素全为 1 的二阶张量, 称之为“1-张量”.

对于无向无权网络而言, 其对应的邻接张量  $\mathbf{A}_{\beta}^a$  可以解释为应用于“1-向量”的 1 个算子, 由此可定义无权无向网络的度中心性向量(度向量)为

$$\mathbf{k}_{\beta} = \mathbf{A}_{\beta}^a \mathbf{u}_a \in \mathbb{R}^N. \quad (26)$$

通过将度向量投影到对应于节点  $v_i$  的标准基向量上, 可得到节点  $v_i$  的度中心性, 即:

$$k_i = \mathbf{k}_{\beta} e_i^{\beta}. \quad (27)$$

##### ② 多关系网络的度中心性

与单层网络相似, 通过使用适当阶的“1-张量”对单层网络情况进行相同的投影, 可以得到多关系网络的度中心性张量(multiplex degree centrality, MDC), 即

$$\mathbf{K}^a = \mathbf{A}_{\beta\delta}^{\alpha\bar{\gamma}} \mathbf{U}_{\gamma}^{\bar{\delta}} \mathbf{u}^{\beta} = \mathbf{k}_{\bar{h}\bar{k}}, \quad (28)$$

其中,  $\mathbf{k}_{\bar{h}\bar{k}}$  为对应于层  $\bar{h}$  和  $\bar{k}$  之间连接的中心性向量. 类似地, 可在多关系网络上得到相应的节点度、平均度、度的二阶矩、度的方差的定义.

节点度(node degree):

$$K_i = \mathbf{K}_{\beta} e_i^{\beta}. \quad (29)$$

平均度(mean degree):

$$\langle K \rangle = \frac{\mathbf{K}_{\beta} \mathbf{u}^{\beta}}{\mathbf{U}_{\rho}^{\rho}}. \quad (30)$$

度的二阶矩(the second moment of the degree):

$$\langle K^2 \rangle = \frac{\mathbf{K}_{\beta} \mathbf{K}^{\beta}}{\mathbf{U}_{\rho}^{\rho}}. \quad (31)$$

度的方差(the variance of the degree):

$$\text{var}(K) = \frac{\mathbf{K}_{\beta} \mathbf{K}^{\beta}}{\mathbf{U}_{\rho}^{\rho}} - \frac{\mathbf{K}_{\alpha} \mathbf{K}_{\beta} \mathbf{U}^{\alpha\beta}}{(\mathbf{U}_{\rho}^{\rho})^2}. \quad (32)$$

对于有向网络, 使用不同的克罗内克积, 可分别得到出度中心性和入度中心性的张量表示:

$$\mathbf{K}^a = \mathbf{A}_{\beta\delta}^{\alpha\bar{\gamma}} \mathbf{U}_{\gamma}^{\bar{\delta}} \mathbf{u}^{\beta}, \quad (33)$$

$$\mathbf{K}_{\beta} = \mathbf{A}_{\beta\delta}^{\alpha\bar{\gamma}} \mathbf{U}_{\gamma}^{\bar{\delta}} \mathbf{u}_{\alpha}. \quad (34)$$

由式(33)和式(34)可得, 使用张量代数的网络表征框架有助于理解有向网络中的方向性, 因为入度与出度的克罗内克积的运算规则是不同的.

#### 2) 多关系网络的局部聚集系数

Cozzo 等人<sup>[44]</sup>从路径和环的概念出发, 提出了多层网络局部聚集系数和全局聚集系数的概念. 在单层网络中, 可用邻接矩阵来计算网络中路径和环的数量<sup>[51-52]</sup>. 类似地, 将多层网络中的行走称之为“超行走(supra-walk)”, 并通过层内邻接矩阵和层间邻接矩阵来计算多层网络的局部聚集系数. 按照 Cozzo 等人提出的多层网络聚集系数的计算方法, 本文给出一种多关系网络局部聚集系数(multiplex local clustering coefficient, MLCC)的简化计算方法. 由 3.1 节可知, 本文运用多重网络模型来表示多关系网络, 则在多关系网络中, 通过“超行走”得到的三元闭环结构有且仅有图 2 所示的 5 种基本形式, 其中较大节点(红色)代表闭环的起始节点, 层内路径用实线表示, 层间路径用虚线表示, 较粗实线(橙色)代表层内第 2 步行走路径. 通过“超行走”所构成的三元闭环可通过将各网络层聚合到一层得到.

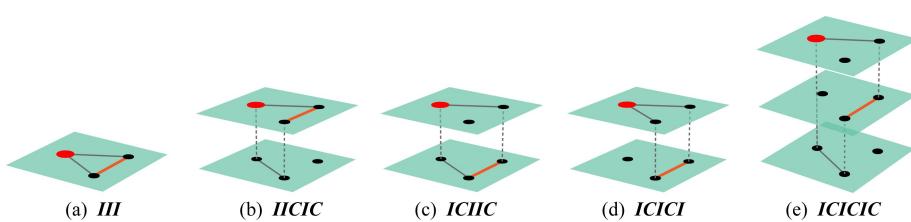


Fig. 2 Sketch of elementary cycles

图 2 基本闭环结构

根据 3.2 节多关系网络邻接矩阵的构造方法,可得到多关系网络节点聚集系数的定义。

**定义 7.** 多关系网络的局部聚集系数。令  $M, I, C$  分别表示多关系网络的邻接矩阵、层内邻接矩阵、层间邻接矩阵,

$$\Omega = \{\text{III, IICIC, ICIIC, ICICI, ICICIC}\}$$

为多关系网络中基本三元闭环结构集合,  $\xi$  为集合  $\Omega$  中的元素,

$$\hat{\Omega} = \{\text{IFI, IFCIC, ICFIC, ICFCI, ICFCIC}\}$$

在构成基本三元闭环结构的“超行走”过程中,层内第 2 步发生在一个完全图上的闭环结构集合。其中,  $F = \bigoplus_a F^{[a]}$ ,  $F^{[a]} = J^{[a]} - E^{[a]}$  为层  $a$  中一全连接网络的邻接矩阵,  $\hat{\xi}$  表示  $\hat{\Omega}$  中的元素, 则多关系网络节点  $v_i$  的局部聚集系数可定义为

$$C_{M,i} = \frac{\sum_{\xi \in \Omega} \omega_{\xi} (\xi)_{ii}}{\sum_{\hat{\xi} \in \hat{\Omega}} \omega_{\hat{\xi}} (\hat{\xi})_{ii}}, \quad (35)$$

其中,  $\omega_{\xi}$  和  $\omega_{\hat{\xi}}$  代表“超行走”分别跨越 1 层、2 层和 3 层所占的贡献, 在无特定背景情况下, 一般设  $\omega_{\xi} = \omega_{\hat{\xi}} = 1/3$ 。

## 4 融合多元信息的节点中心性度量方法

现实世界的社交数据中存在许多有向关系, 比如微博网络中的转发、关注关系, 引文网络中的引用和被引用关系, 都可以用有向边予以刻画。因此, 本节基于第 2 节和第 3 节介绍的基础理论, 用有向多重网络的张量表示框架刻画多关系社交网络, 综合考虑社交网络中节点中心性、声望、传递性对节点重要性造成的影响, 分别用多关系网络中节点的出度中心性、入度中心性、局部聚集系数进行量化, 运用信息融合的思想, 对多关系网络中的节点重要性度量方法进行了研究。

### 4.1 基于 IO-ClusterRank 的节点重要性排序方法

借鉴 2.2 节介绍的 ClusterRank 和 IO-ClusterRank 的思想方法, 并利用多关系网络的度量指标进行推广, 提出了一种基于 ClusterRank 的多关系社交网络节点的半局部中心性度量方法 IOMCR。IOMCR 在考虑节点本身及其邻居节点中心性和声望的同时, 还考虑了节点的传递能力。IOMCR 的定义:

$$v_{\text{IOMCR},i} = \frac{f(C_{M,i})}{K_i} \sum_{a=1}^M R_i^{[a]}. \quad (36)$$

节点的局部聚集系数与其传递能力一般呈负相关关系, 即其局部聚集系数越小, 说明该节点传递能

力越强。式(36)中,  $f(C_{M,i})$  为节点  $v_i$  的局部聚集系数  $C_{M,i}$  的非线性负相关函数, 本文取  $f(C_{M,i}) = 10^{C_{M,i}} \cdot K_i$  为节点  $v_i$  的度中心性值,  $R_i^{[a]}$  的计算:

$$R_i^{[a]} = K_{[a]} R_i^{[a]} + K_i^{[a]} R_i^{[a]}, \quad (37)$$

$$R_i^{[a]} = \sum_{j \in \Gamma_i} (K_j^{[a]} + 1), \quad (38)$$

$$R_i^{[a]} = \sum_{j \in \Gamma_i} (K_j^{[a]} + 1), \quad (39)$$

其中,  $K_{[a]}$  和  $K_i^{[a]}$  分别代表节点  $v_i$  的出度和入度, 用以刻画该节点的中心性和声望。 $R_i^{[a]}$  和  $R_i^{[a]}$  分别刻画了节点  $v_i$  邻居的中心性和声望强度,  $K_j^{[a]}$  和  $K_j^{[a]}$  分别为  $v_i$  的邻居节点集合  $\Gamma_i$  中节点  $v_j$  的出度中心性和入度中心性。

### 4.2 融合多元信息的节点重要性排序方法

4.1 节提出的 IOMCR 指标采用了一种正相关性融合方法, 即认为节点的中心性和声望越显著, 其传递性也就越强。但在真实网络中, 很多节点往往不具备这种正相关关系, 例如在微博网络中, 一些用户虽然中心性和声望都很显著(即关注和被关注的用户都很多), 其可以接收到很多消息, 但却不喜欢转发(即传递性很弱), 还有一些用户热衷于转发自己的消息(即传递能力较强), 但关注和被关注的用户却很少(即中心性和声望不显著)。综上, 中心性、声望、传递性之间往往呈现出的是一种不确定性关系, 而 IOMCR 方法往往需要在中心性、声望、传递性三者之间具有相对明确的正相关性条件下, 才能表现出较好性能。而在多关系网络中, 由于传递机制具有差异性, 往往三者之间不具备明确的正相关关系。故 IOMCR 指标具有一定的局限性, 其在度量节点重要性程度时, 节点中心性、声望、传递性三者信息之间有可能会相互抵消, 从而对度量节点重要程度造成影响。

考虑到 D-S 证据理论<sup>[39,46,53]</sup>是处理不确定性问题的理想办法, 本文引入 D-S 证据理论, 对 IOMCR 进行了改进, 分别将节点的中心性和声望信息与节点的传递性信息进行融合, 进一步提出了 IOMEC 方法:

$$v_{\text{IOMEC},i} = \frac{v_{\text{OME},i} + v_{\text{IMEC},i}}{2}. \quad (40)$$

其中,  $v_{\text{OME},i}$  为节点中心性和传递性信息融合后的度量值,  $v_{\text{IMEC},i}$  为声望信息和传递性信息融合后的度量值, 取二者的平均值作为最终的度量值。以计算  $v_{\text{OME},i}$  为例, 介绍 D-S 证据理论融合节点中心性和传递性信息的方法步骤:

### 1) 计算节点中心性和传递性量化指标

分别计算节点的出度中心性向量  $\mathbf{K}^a$  和节点的聚集系数向量  $\mathbf{C}_M$ , 记  $K^i$  和  $C_{M,i}$  分别为  $\mathbf{K}^a$  和  $\mathbf{C}_M$  中对应第  $i$  个位置的元素值, 分别计算  $f(C_{M,i}) = 10^{-C_{M,i}}$ , 所得结果向量记为  $\mathbf{F}$ . 统计  $\mathbf{K}^a$  和  $\mathbf{F}$  中元素的最大值和最小值, 分别记为  $K^M, K^m, F_M, F_m$ . 一般情况下,  $K^M \neq K^m, F_M \neq F_m$ .

### 2) 确定节点重要程度的辨识框架

按照 2.3 节中 D-S 证据理论的处理流程, 首先应确定考察对象的辨识框架  $\Psi$ . 本文所考察的是节点的重要程度, 简单起见, 可将网络中节点的重要性分为节点重要和节点不重要 2 类, 用 I 表示节点重要, NI 表示节点不重要, 则记  $\Psi = \{I, NI\}$ . 显然,  $\Psi$  中的元素两两互斥且为有限完备集合.

### 3) 构建基本概率指派函数

令  $m_{K,I}(i)$  和  $m_{K,NI}(i)$  分别为在  $K^i$  下节点  $v_i$  重要和不重要的概率,  $m_{F,I}(i)$  和  $m_{F,NI}(i)$  分别代表在  $f(C_{M,i})$  下节点  $v_i$  重要和不重要的概率. 在无特定条件情况下, 通常约定该概率指派服从均匀分布, 根据均匀分布的概率密函数, 可定义上述 4 个指标的 BPA 函数:

$$m_{K,I}(i) = \frac{|K^i - K^m|}{K^M - K^m + 2\mu}, \quad (41)$$

$$m_{K,NI}(i) = \frac{|K^i - K^M|}{K^M - K^m + 2\mu}, \quad (42)$$

$$m_{F,I}(i) = \frac{|F_i - F_m|}{F_M - F_m + 2\lambda}, \quad (43)$$

$$m_{F,NI}(i) = \frac{|F_i - F_M|}{F_M - F_m + 2\lambda}. \quad (44)$$

$\mu$  和  $\lambda$  为修正参数, 取值对节点排序没有影响, 通常取  $\mu, \lambda \in (0, 1)$ .

### 4) 进行 D-S 证据组合

首先定义考察对象的集合, 记为

$$M(i) = \{m_I(i), m_{NI}(i), m_\Psi(i)\}, \quad (45)$$

其中,  $m_I(i)$  为表示节点  $v_i$  重要的概率,  $m_{NI}(i)$  表示节点  $v_i$  不重要的概率,  $m_\Psi(i)$  表示节点  $v_i$  或者重要或者不重要的概率. 根据式(16), 可分别计算  $m_I(i), m_{NI}(i), m_\Psi(i)$ .

### 5) 计算 OMEC.

根据贝叶斯理论, 在无其他先验知识的情况下, 将  $m_\Psi(i)$  等可能分配给  $m_I(i)$  和  $m_{NI}(i)$ , 进一步修正节点  $v_i$  重要的概率  $M_I(i)$  和不重要的概率  $M_{NI}(i)$ , 即:

$$M_I(i) = m_I(i) + \frac{m_\Psi(i)}{2}, \quad (46)$$

$$M_{NI}(i) = m_{NI}(i) + \frac{m_\Psi(i)}{2}. \quad (47)$$

由式(46)和式(47)可得, 当  $M_I(i)$  越大、 $M_{NI}(i)$  越小时, 说明节点  $v_i$  越重要. 鉴于此, 对  $M_I(i)$  和  $M_{NI}(i)$  进行组合, 可得到  $v_{OMEC,i}$  的度量值:

$$v_{OMEC,i} = M_I(i) - M_{NI}(i) = m_I(i) - m_{NI}(i). \quad (48)$$

为保证概率的非负性和归一性, 将式(48)进行标准化处理, 进一步修正为

$$v_{OMEC,i}^* = \frac{|v_{OMEC,min}| + v_{OMEC,i}}{\sum_{i=1}^N (|v_{OMEC,min}| + v_{OMEC,i})}. \quad (49)$$

类似地, 将  $v_{OMEC,i}$  计算过程中的出度中心性向量  $\mathbf{K}^a$  改为计算入度中心性向量  $\mathbf{K}_a$ , 即可得到融合声望和传递性的  $v_{IMEC,i}$  度量指标.

## 4.3 时空复杂度分析

多关系网络中, 记节点数为  $n$ , 关系数为  $m$ .

### 1) 时间复杂度分析

计算多关系网络出度中心性和入度中心性的时间复杂度均为  $O(mn)$ , 计算节点的局部聚集系数时间复杂度为  $O(m^3n^3)$ . 由于 IOMCR 方法和 IOMEC 方法的计算均考虑了衡量节点中心性、声望和传递性的度量指标, 故二者的计算复杂度均依赖于计算节点出度中心性、入度中心性和局部聚集系数的时间复杂度. 在已知节点出度中心性、入度中心性和局部聚集系数的条件下, IOMCR 方法由于考虑了节点的一阶邻居信息, 其时间复杂度为  $O[mn(\langle K^a \rangle + \langle K_a \rangle)]$ , 而 IOMEC 方法将层间信息通过 D-S 证据理论进行了融合, 仅关注了节点本身, 故其时间复杂度为  $O(n)$ . 在单独计算二者的时间复杂度方面, IOMEC 方法要优于 IOMCR 方法, 但综合考虑节点的出度中心性、入度中心性和聚集系数的时间复杂度后, 二者的时间复杂度均为  $O(m^3n^3)$ .

### 2) 空间复杂度分析

鉴于通过构建 4 阶张量对多关系网络进行表征, 故其空间复杂度为  $O(m^2n^2)$ . 在代码实现过程中, 由于考虑到邻接张量的稀疏性, 采用稀疏张量对其进行压缩处理, 可以有效降低多关系网络的空间复杂度.

## 5 实验与讨论

通过在 4 个真实社交网络数据集上进行相关实验, 从运行时间和度量节点重要程度准确性 2 个方面对所提出方法进行了评测, 并选取多关系网络的度中心性和特征向量中心性<sup>[47]</sup> 2 种经典指标进行对比来说明提出方法的有效性, 为避免与单关系网

络中的度量指标进行混淆,将多关系网络中的度中心性度中心性记为 MDC、特征向量中心性记为 MEVC.实验平台为 Intel Core i9-9900K 3.6 GHz, 64 GB RAM, 操作系统为 CentOS 7, 代码由 Python 实现.

选取了 4 个规模不同的真实社交网络数据集<sup>[54]</sup> MoscowAthletics2013, NYClimateMarch2014, MLKing2013, Cannes2013, 这 4 个数据集均是围绕特定事件从 Twitter 平台获得的, 包含转发、提及、回复 3 种有向关系. 4 个数据集的基本统计信息如表 1 所示:

Table 1 The Basic Information of Experimental Data

表 1 实验数据集的基本信息

Data Set	$N_r$	$N_u$	$N_l$	$\langle K \rangle$	$\langle K^2 \rangle$
MoscowAthletics2013	3	88804	210250	10.69	3286.23
NYClimateMarch2014	3	102439	353495	12.85	6785.34
MLKing2013	3	327707	396671	8.40	1559.62
Cannes2013	3	438537	991854	10.46	3160.37

## 5.1 评价标准

在节点重要性研究中, 通常利用传播动力学模型对节点重要性的排序结果进行评价, 本文采用 SIR 模型<sup>[11,32,55-57]</sup>. SIR 模型包括 3 个状态:

1) 易感(susceptible)状态  $S(t)$ , 表示易感染但尚未感染疾病的个体数量.

2) 感染(infected)状态  $I(t)$ , 表示已经感染, 并且能够将疾病传播给易感个体的个体数量.

3) 恢复(recovered)状态  $R(t)$ , 表示感染个体已经恢复, 并且不会被再次感染的个体数量.

在 SIR 模型中, 节点重要性的评价标准一般有 2 项: 一种为节点的平均传播范围, 即达到稳态时感染态  $I(t)$  和恢复态  $R(t)$  的数量之和, 数量越大, 说明该节点越重要. 另一种为传播速率, 可从传播初始阶段感染态  $I(t)$  和恢复态  $R(t)$  的数量和达到稳态时所用时间 2 方面来评判, 数量越大、所用时间越小, 说明该节点越重要; 为方便说明, 将第  $t$  个时间迭代步时的感染态和恢复态的数量之和记为  $F(t)$ , 感染态的平均数量记为  $F_i(t)$ , 达到稳态时的数量记为  $F(t_c)$ .

由于网络规模不同, SIR 模型的传染概率和恢复概率依赖于各层层内网络的拓扑结构. 通常, 传染概率的设置依赖于各层内网络的渐近传播阈值  $\beta_c^{[a]}$ . 渐进传播阈值为传播过程中应取的最小传播概率<sup>[33]</sup>, 定义为

$$\beta_c^{[a]} = \frac{\langle K^{[a]} \rangle}{\langle (K^{[a]})^2 \rangle - \langle K^{[a]} \rangle}. \quad (50)$$

根据贝叶斯理论和条件概率, 各层传染概率设置为

$$p_i^{[a]} = \frac{\lambda_i \beta_c^{[a]}}{N_r}, \quad (51)$$

其中,  $\lambda_i$  为放大系数,  $N_r$  为多关系网络的关系数. 恢复概率通常依赖于各层的平均度, 类似地, 各层恢复概率设置为

$$p_r^{[a]} = \frac{\lambda_r}{N_r \langle K^{[a]} \rangle}. \quad (52)$$

由于传播过程是依概率进行感染传播, 即使 2 组实验条件完全相同, 最终得到的结果也会存在一定差异. 为此, 为此选取 100 次独立实验所得结果的期望值作为评价结果.

## 5.2 运行时间分析

通过统计 4 个真实网络数据集上各算法的实际运行时间来比较提出方法的运行速率, 并对 4.3 节中的时间复杂度分析结果进行验证. 为方便说明, 记  $t_K, t_L, t_I, t_{IE}, t_E$  分别为多关系网络中计算多重度中心性、多重局部聚集系数、IOMCR 方法、IOMECE 方法以及多重特征向量中心性的运行时间,  $t_I^*$  和  $t_{IE}^*$  分别为单独计算 IOMCR 方法和 IOMECE 方法的运行时间. 根据 4.3 节对各算法的时间复杂度分析可知,  $t_I$  和  $t_{IE}$  依赖于  $t_K$  和  $t_L$ , 即:

$$t_I = t_K + t_L + t_I^*, \quad (53)$$

$$t_{IE} = t_K + t_L + t_{IE}^*. \quad (54)$$

由于 4 个网络规模不同, 我们通过统计各方法运行时间占总运行时间的比重来说明算法运行时间的长短, 即:

$$P(t) = \frac{t_{\text{method}}}{t_{\text{total}}}. \quad (55)$$

其中,  $t_{\text{method}}$  表示计算指定方法时所消耗的时间,  $t_{\text{total}}$  表示计算各方法所消耗的总时间.

如图 3 所示, 在单独计算 IOMCR 方法和 IOMECE 方法时,  $t_{IE}^*$  要明显少于  $t_I^*$ . 图 4 描述了 5 种指标在 4 个网络上的运行时间, 可以看出,  $t_I, t_{IE}, t_K$  三者之间差别不大, 说明在计算 IOMCR 方法和 IOMECE 方法的过程中, 主要时间消耗在计算多重局部系数的环节上, 而多重度中心性和多重特征向量中心性由于都是线性时间复杂度, 故二者运行时间很短, 在整个计算过程中基本可以忽略不计. 综上, 实验结果与 4.3 节中时间复杂度的分析是一致的, 在运行时间上, IOMECE 方法要优于 IOMCR 方法.

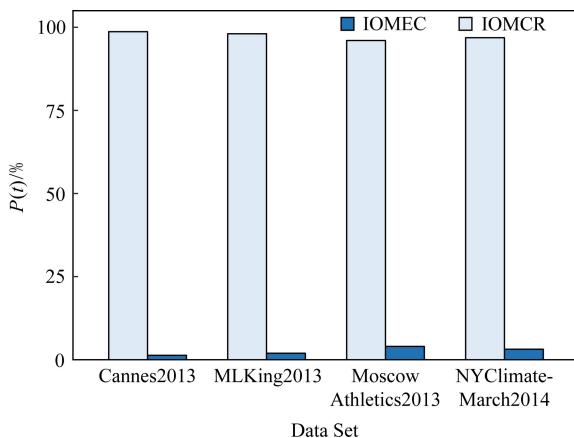


Fig. 3 The run time of IOMCR and IOMECE

图3 IOMCR方法和IOMECE方法运行时间统计图

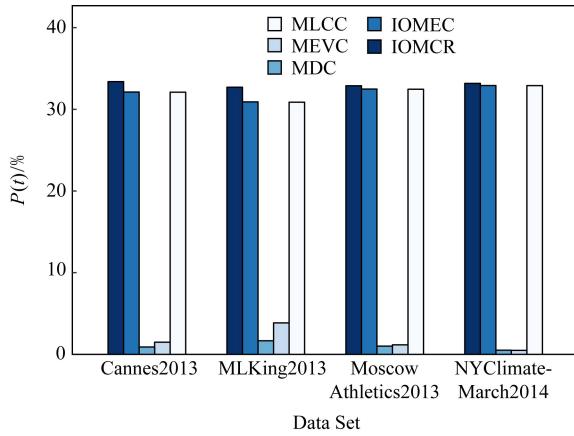


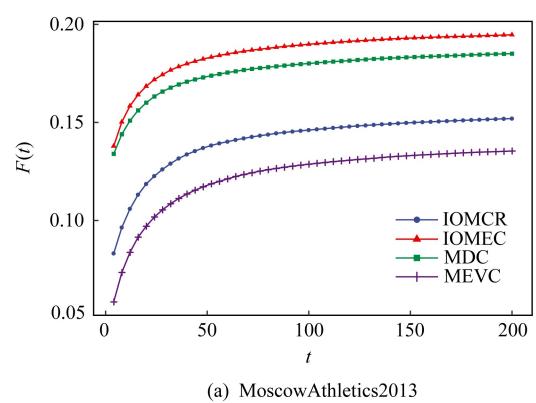
Fig. 4 The run time of measures

图4 各度量方法运行时间统计图

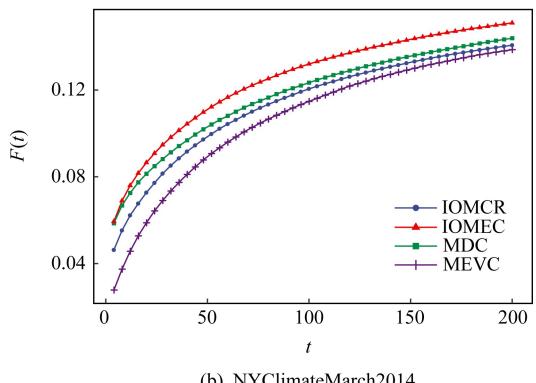
### 5.3 度量准确性分析

本节在4个数据集上进行了4组不同的实验，并通过SIR进行评测，分别从传播范围和传播速率2方面对各度量方法进行评价来验证提出算法在度量节点重要程度的准确性。

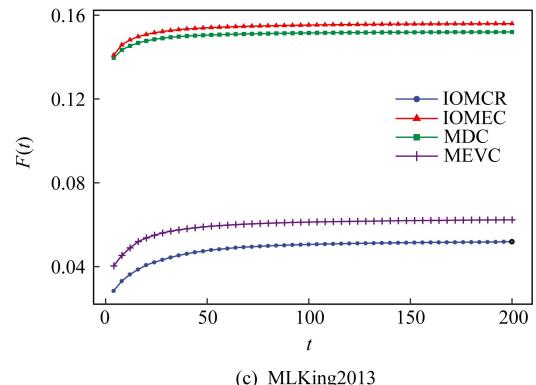
**实验1.**通过观测SIR传播曲线的传播范围和传播速率，对各方法度量节点重要性的准确性进行评测。在4个网络中，按照各方法对节点重要程度的度量值进行降序排序，分别选取前1%的节点作为初始传染源节点，统计给定时间迭代步的 $F(t)$ ，并绘制SIR传播曲线，其中横轴代表时间迭代步数，在前3个网络中，设定最大迭代步数为200，而由于Cannes2013规模较大，设置最大迭代步为500。纵轴代表对应的 $F(t)$ ，通过观测曲线达到稳态时的时间和对应的 $F(t_e)$ 来说明传染源节点的传播范围，通过观测曲线的斜率变化来说明传染源节点的传播速率。图5所示的实验结果显示：



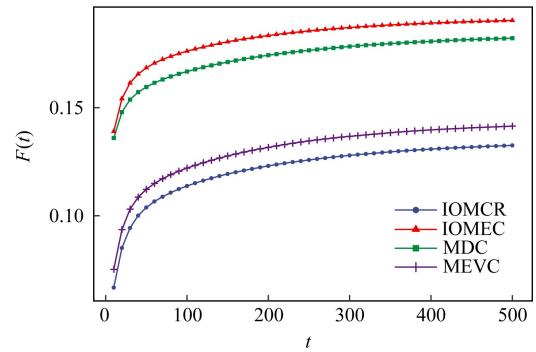
(a) MoscowAthletics2013



(b) NYClimateMarch2014



(c) MLKing2013



(d) Cannes2013

Fig. 5 SIR spreading curves of the top 1% nodes obtained by each method as the initial infectious nodes  
图5 各度量方法所得前1%的节点作为初始传染源节点时的SIR传播曲线

1) 在 4 个网络上,无论是传播时间还是传播范围, IOMEC 方法相比其他方法,都能表现出较好结果,特别是随着网络规模的增加,该优势愈发明显,而多重特征向量中心性表现最差;

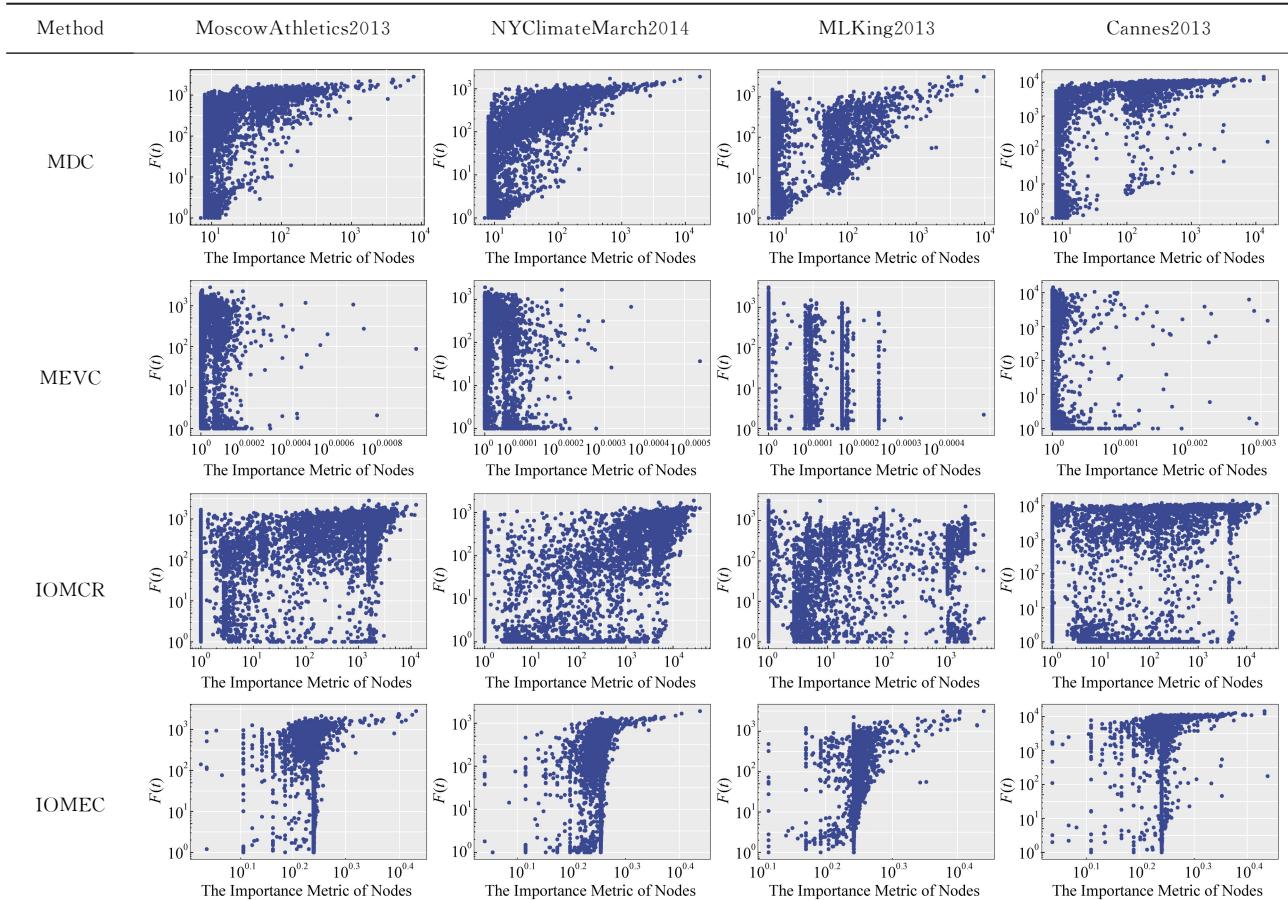
2) IOMCR 方法的评测结果要劣于 IOMEC 方法和多重度中心性,特别是在 MLKing2013 数据集上表现最为突出, IOMEC 方法和多重度中心性得到的重要节点的传播范围,约是 IOMCR 方法得到重要节点传播范围的 3 倍;

3) 多重度中心性总体表现良好,但要劣于提出的 IOMEC 方法,并且随着网络规模的增大,二者差距愈发明显.

**实验 2.** 由于在传播过程中,传染源节点在早期的传播中起到主要作用,即对于节点传播能力的评测,早期的传播更为重要. 本实验主要通过观测各节点传播初期的  $F(t)$  与各方法度量值之间的相关关系,来说明节点的传播能力. 本实验设置  $t = 10$ . 在理想状态下,  $F(t)$  与对应的测量值应呈现正相关关系.

4 个数据集上的实验结果如表 2 所示. 表 2 中,每列子图分别对应同一数据集下 4 种不同方法,坐标轴采用双对数坐标,横轴表示每种方法下各节点的度量值,纵轴表示各节点作为传染源节点,所对应的  $F(t)$  ( $t = 10$ ). 由于网络规模较大,故选取各方法得到前 1000 个重要节点的并集作为实验节点集合.

Table 2 The Correlation Between  $F(t)$  ( $t = 10$ ) and Its Measure  
表 2  $t = 10$  时,  $F(t)$  与各方法度量值间的关系



实验结果表明:多重度中心性和 IOMEC 方法在 4 个网络中都呈现出一定的正相关性,而另 2 种方法均未呈现出明显的正相关性. 在多重度中心性方法中,一些度量值较低的节点反而  $F(t)$  更大且分布较为稠密,而在 IOMEC 方法中这类节点分布稀疏,故 IOMEC 方法总体上要优于多重度中心性方法.

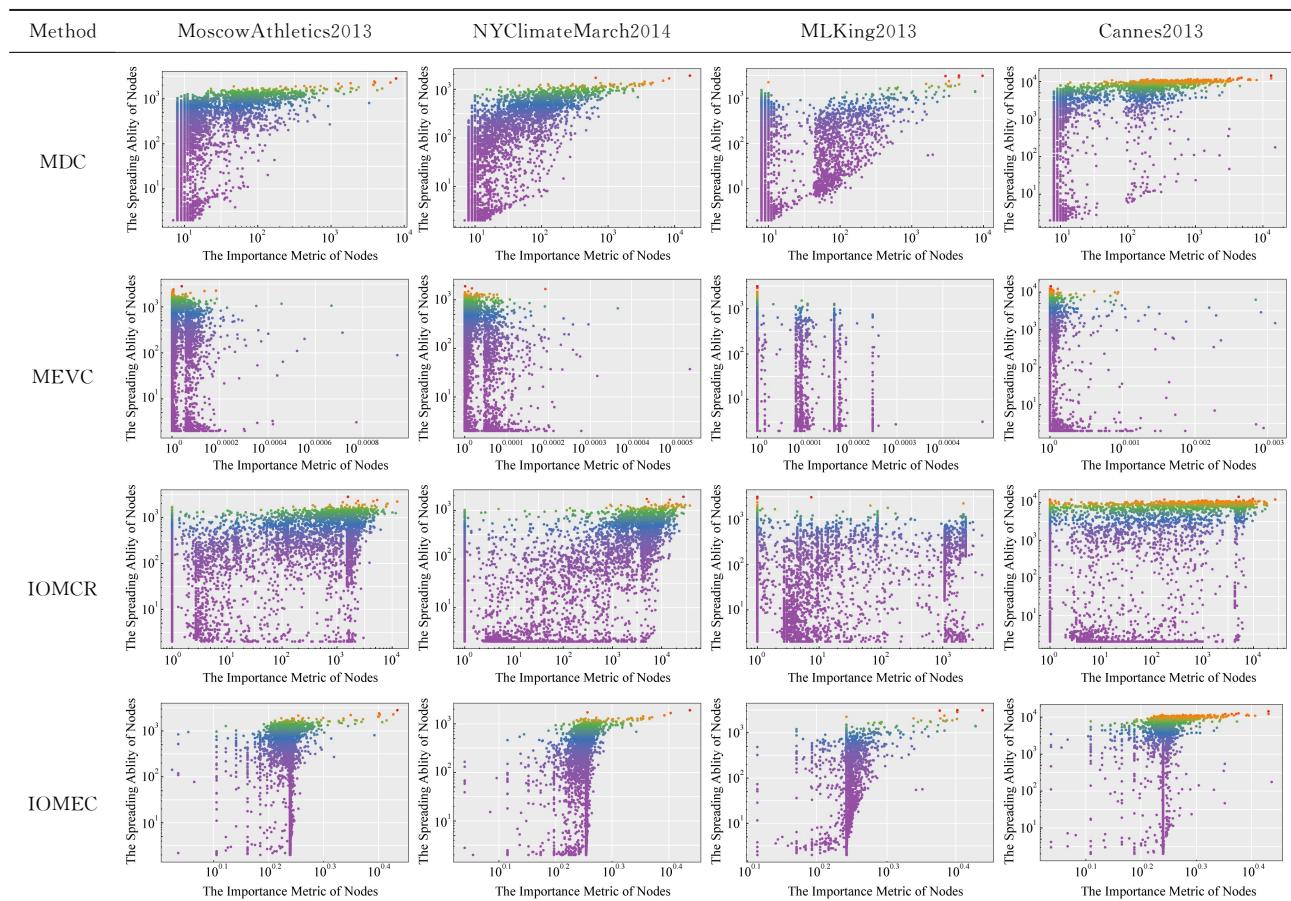
**实验 3.** 为进一步讨论各方法节点的测量值与其传播能力之间的相关关系,并评测各方法对节点重要性度量的准确性. 本实验对各节点在 SIR 模型中得到的稳态值  $F(t_c)$  进行排序,并以此为基准,比较其与各方法对节点重要性度量值之间的相关关系. 表 3 描述了在 4 个网络中,各节点在 SIR 模型中

的传播能力与 4 种方法对节点度量值间的相关关系,表 3 中每一列子图对应同一数据集的不同方法,坐标轴采用双对数坐标,横轴表示各节点在对应方法下的测量值,纵轴表示各节点在 SIR 模型中得到

的稳态值  $F(t_c)$ ,每一个数据点代表网络中的一个节点,节点的颜色表示在传播过程中,  $t=10$  时的传播能力,即  $F(10)$ .类似地,分别选取各方法得到前 1 000 个重要节点的并集作为实验节点集合.

Table 3 The Correlation Between Each Measure and Spreading Ability of SIR

表 3 节点度量值与 SIR 传播能力间的相关关系



实验结果表明:表 3 所示的各节点相关性分布与表 2 所示的节点分布基本一致,说明各节点传播能力在传播过程初期起主要作用.从结果上分析, IOMEC 方法和多重度中心性仍然能够呈现出正相关态势,IOMEC 方法所得到的节点相关性分布要优于多重度中心性;IOMCR 方法和多重证据中心性所得到的结果较差,没有呈现出明显的相关性态势.

**实验 4.**为了进一步验证各方法的性能,本实验研究了按照某种度量方法对节点的重要程度进行排序后,节点排序与在 SIR 传播过程中所对应的平均感染数量  $\langle F(t) \rangle$  ( $t=10$ ) 之间的关系.从理论上分析,性能较好的度量指标,其图像应该呈现递减关系,即随着节点重要程度的降低,对应的受感染节点的平均数会减少.为了对比,本实验特意增加了随机排序(random ranking)这种极端情况.实验结果如

图 6 所示,其中横轴代表各节点排序,纵轴代表对应的受感染节点平均数量.类似地,分别选取各方法得到前 1 000 个重要节点的并集作为实验节点集合.4 个网络上的实验结果表明:

1) 随机排序和多重证据中心性情况最差.二者除未呈现出较好的递减态势外,平均感染数量也非常少,说明二者所得到的节点的传播能力很弱.

2) IOMEC 方法和多重度中心性在 4 个网络上均能够呈现出良好的递减态势,并且在平均感染数量上,IOMEC 方法所得到的结果要优于多重证据中心性,进一步验证了本文提出方法对节点重要程度度量的准确性.

3) 在 MoscowAthletics2013 和 NYClimateMarch 2014 上,IOMCR 方法所得到的结果能够呈现出较好的递减态势,并且筛选出部分节点的传播能力要

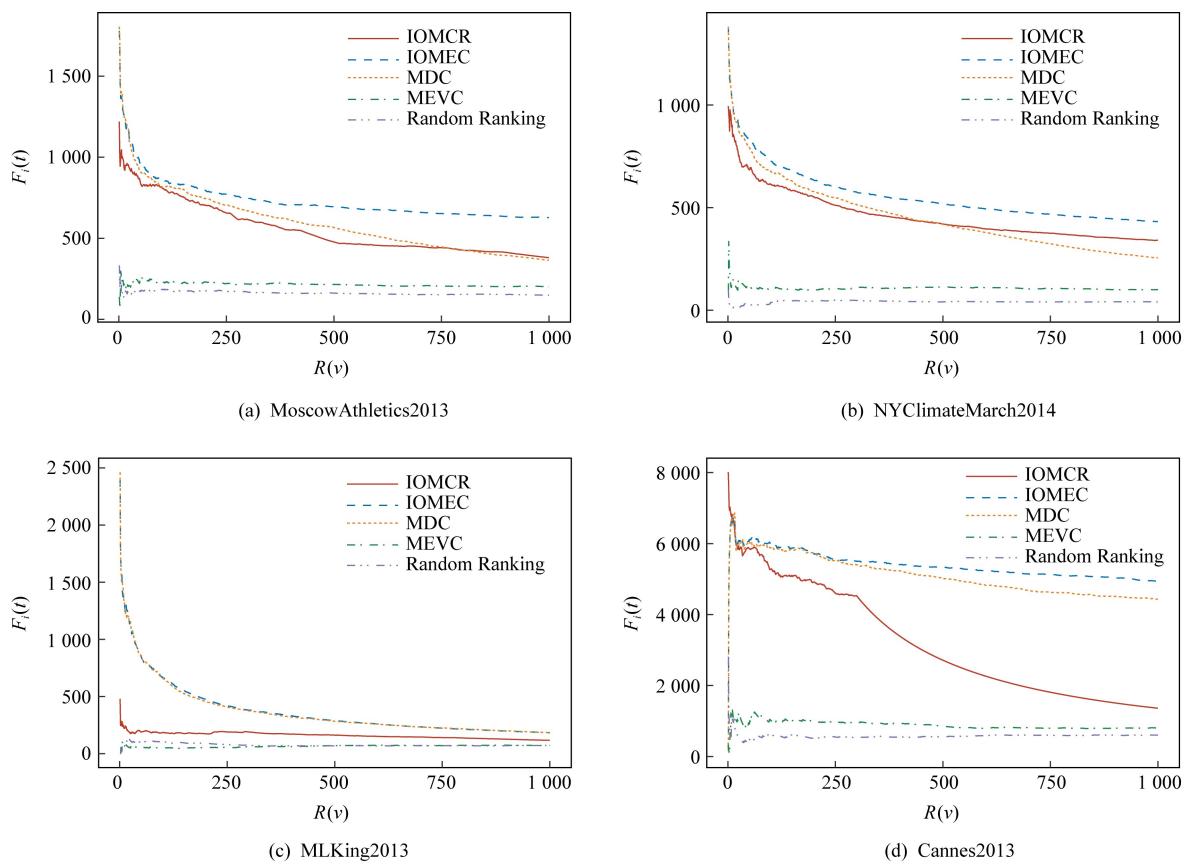


Fig. 6 The relation between the node ordering given by five measures and average infected number in SIR( $t=10$ )

图 6 通过 5 种方法得到的节点排序与 SIR 中对应  $t=10$  时的平均感染数间的关系

优于多重度中心性所得到的结果,但在其他 2 个网络上, IOMCR 方法所得到的结果表现较差,筛选出的节点传播能力与 IOMECE 方法和多重度中心性得到的结果差距明显,尤其在 MLKing2013 中,未呈现出明显的递减趋势,并且平均感染数量很低,进一步说明,随着网络规模的增大,节点传递性的因素愈发重要,并且节点的中心性、声望、传递性之间不具备绝对的正相关性,传统的 ClusterRank 思想存在一定的局限性。

## 6 总 结

本文针对多关系社交网络的节点重要性问题进行了研究,应用有向多重网络模型和基于张量代数的数学框架对其进行建模和分析,综合分析中心性、声望和传递性对节点重要性的影响,提出了 IOMCR 节点重要性度量方法,并针对其存在问题,引入 D-S 证据理论信息融合方法对其进行改进,进一步提出了面向多关系网络的 IOMECE 节点重要性度量方法。在 4 个真实网络上的实验表明 4 个主要结论:

1) 在多关系社交网络中,由于受耦合信息和传递机制差异性的影响,节点的中心性、声望和传递性之间不具有绝对的相关关系,对三者进行相关性融合的 IOMCR 度量方法会对计算结果造成误差,降低了度量节点重要性的准确度,随着网络规模的增大,该问题会更加突出,说明传统的 ClusterRank 思想具有局限性。

2) 通过引入 D-S 证据理论,对节点中心性、声望、传递性进行依概率融合所提出的 IOMECE 度量方法,可以有效消除多关系网络耦合信息和传递机制的差异性对节点重要信息造成的影响,相比于只关注中心性和声望的多重度中心性和只关注节点传递能力的多重特征向量中心性,能够更加准确地挖掘网络中的重要节点,特别是在大规模网络中优势更为明显。

3) 通过 IOMECE 方法、多重度中心性和多重特征向量中心性 3 种方法所得结果的横向对比分析,说明在多关系网络的节点重要性研究中,基于网络拓扑结构的中心性和声望因素占主导地位,而只关注节点传递特性,直接扩展单个网络节点特征向量

中心性的思路方法到多关系网络上,并未得到理想效果.为更加准确的度量节点的重要性程度,在关注节点中心性和声望的同时,也应该关注节点的传递特性,特别是在大规模网络中,节点的传递性作用更加突出.

4) 由于考虑了节点的传递性指标,故提出方法总体具有较高的时间复杂度,相比于 IOMCR 方法,改进的 IOMEC 方法具有更低的时间复杂度.故本文提出方法适用于需要精准挖掘网络中重要节点的应用场景,对于准确度不高的应用场景,可以考虑只关注节点中心性和声望的多重度中心性.

综上,本文所做工作不仅对多关系社交网络的节点重要性研究提供了新的思路方法,也进一步拓展了 D-S 证据理论的应用场景.然而,本文的工作也并不完美,受实验环境的限制,尚不适用于超大规模的多关系网络.近年来,复杂网络呈现出大规模、高维度和动态性的特点,如何建立合理的模型对网络进行表示,设计有效的度量指标或算法识别大规模高维动态复杂网络的关键节点是当前复杂网络节点重要性研究面临的一项挑战,也是我们今后研究工作的重点.

## 参 考 文 献

- [1] Fang Jinqing. From a single network to “network of networks” development process: Some discussions on the exploration of multilayer supernetwork models and challenges [J]. *Complex Systems and Complexity Science*, 2016, 13(1): 40-47 (in Chinese)  
(方锦清. 从单一网络向《网络的网络》的转变进程——略论多层次超网络模型的探索与挑战[J]. 复杂系统与复杂性科学, 2016, 13(1): 40-47)
- [2] Battiston F, Nicosia V, Latora V. The new challenges of multiplex networks: Measures and models [J]. *The European Physical Journal Special Topics*, 2017, 226(3): 401-416
- [3] Tu Xiao, Jiang Guoping, Song Yurong, et al. Novel multiplex PageRank in multilayer networks [J]. *IEEE Access*, 2018, 6: 12530-12538
- [4] Ding Cangfeng, Li Kan. Centrality ranking in multiplex networks using topologically biased random walks [J]. *Neurocomputing*, 2018, 312: 263-275
- [5] De Arruda G F, Cozzo E, Moreno Y, et al. On degree-degree correlations in multilayer networks [J]. *Physica D: Nonlinear Phenomena*, 2016, 323/324: 5-11
- [6] Sole-Ribalta A, De Domenico M, Gómez S, et al. Random walk centrality in interconnected multilayer networks [J]. *Physica D: Nonlinear Phenomena*, 2016, 323/324: 73-79
- [7] Liu Yangyang, Zhao Chengli, Zhang Xue, et al. Core structure: The coupling failure procedure in multiplex networks [J]. *Physica A: Statistical Mechanics and Its Applications*, 2018, 507: 1-11
- [8] Ma Lijia, Gong Maoguo, Yan Jianan, et al. Detecting composite communities in multiplex networks: A multilevel memetic algorithm [J]. *Swarm and Evolutionary Computation*, 2018, 39: 177-191
- [9] Ding Rui, Ujang N, Hamid H B, et al. Detecting the urban traffic network structure dynamics through the growth and analysis of multi-layer networks [J]. *Physica A: Statistical Mechanics and Its Applications*, 2018, 503: 800-817
- [10] Pan Lulu, Shao Haibin, Mesbahi M, et al. Structural balance of multiplex signed networks: A distributed data-driven approach [J]. *Physica A: Statistical Mechanics and Its Applications*, 2018, 508: 748-756
- [11] Zheng Chunyuan, Xia Chengyi, Guo Quantong, et al. Interplay between SIR-based disease spreading and awareness diffusion on multiplex networks [J]. *Journal of Parallel and Distributed Computing*, 2018, 115: 20-28
- [12] Pedroche F, Garcia E, Romance M, et al. On the spectrum of two-layer approach and multiplex PageRank [J]. *Journal of Computational and Applied Mathematics*, 2018, 344: 161-172
- [13] Faust K, Wasserman S. *Social Network Analysis: Methods and Applications* [M]. Cambridge, MA: Cambridge University Press, 1994
- [14] Gluckman M. *The Judicial Process Among the Barotse of Northern Rhodesia* [M]. Manchester, England: Manchester University Press, 1955
- [15] Verbrugge L M. Multiplexity in adult friendships [J]. *Social Forces*, 1979, 57(4): 1286-1309
- [16] Mitchell J C. *Social Networks in Urban Situations: Analyses of Personal Relationships in Central African Towns* [M]. Manchester, England: Manchester University Press, 1969
- [17] Pattison P, Wasserman S. Logit models and logistic regressions for social networks: II. Multivariate relations [J]. *British Journal of Mathematical and Statistical Psychology*, 1999, 52(2): 169-193
- [18] Lusher D, Koskinen J, Robins G. Exponential random graph models for social networks [J]. *Social Networks*, 2009, 31(1): 12-25
- [19] Wang Nana, Gao Hong, Liu Wei. Research on a heterogeneous edge multi-graph network model [J]. *CAAI Transactions on Intelligent Systems*, 2017, 12(4): 475-481 (in Chinese)  
(王娜娜, 高红, 刘巍. 异质边多重图网络模型研究[J]. 智能系统学报, 2017, 12(4): 475-481)

- [20] Wang Nana, Gao Hong, Li Shanshan, et al. A research on hypergraph of heterogeneous edge [J]. *Journal of Guangdong University of Technology*, 2017, 34(1): 6–10 (in Chinese)  
(王娜娜, 高红, 李珊珊, 等. 基于异质超边的超图[J]. 广东工业大学学报, 2017, 34(1): 6–10)
- [21] Huang Ruiyang, Wu Qi, Zhu Yuhang. Community detection in dynamic heterogeneous network with joint nonnegative matrix factorization [J]. *Application Research of Computers*, 2017, 34(10): 2989–2992 (in Chinese)  
(黄瑞阳, 吴奇, 朱宇航. 基于联合矩阵分解的动态异质网络社区发现方法[J]. 计算机应用研究, 2017, 34(10): 2989–2992)
- [22] Bing Sheng, Sun Gengxin. Important node detection algorithm for multiple relationships on line social network based on multi-subnet composited complex network model [J]. *Journal of Nanjing University: Natural Sciences*, 2017, 53(2): 378–385 (in Chinese)  
(宾晟, 孙更新. 基于多子网复合复杂网络模型的多关系社交网络重要节点发现算法[J]. 南京大学学报: 自然科学, 2017, 53(2): 378–385)
- [23] Kolda T G, Bader B W. Tensor decompositions and applications [J]. *SIAM Review*, 2009, 51(3): 455–500
- [24] Jo H, Moon H, Ki Baek S. Immunization dynamics on a 2-layer network model [J]. *Physica A: Statistical Mechanics and Its Applications*, 2006, 361(2): 534–542
- [25] Acar E, Yener B. Unsupervised multiway data analysis: A literature survey [J]. *IEEE Transactions on Knowledge and Data Engineering*, 2009, 21(1): 6–20
- [26] Kurant M, Thiran P, Hagmann P. Error and attack tolerance of layered complex networks [J]. *Physical Review E Statistical Nonlinear & Soft Matter Physics*, 2006, 76(2): 26103
- [27] Kurant M, Thiran P. Layered complex networks [J]. *Physical Review Letters*, 2006, 96(13): 138701
- [28] Wei Xiang, Chen Shihua, Wu Xiaojun, et al. A unified framework of interplay between two spreading processes in multiplex networks [J]. *EPL: A Letters Journal Exploring the Frontiers of Physics*, 2016, 114(2): 26006
- [29] Mei Guofeng, Wu Xiaoqun, Wang Yingfei, et al. Compressive-sensing-based structure identification for multilayer networks [J]. *IEEE Transactions on Cybernetics*, 2018, 48(2): 754–764
- [30] Zhou Changsong, Zemanová L, Zamora-López G, et al. Structure-function relationship in complex brain networks expressed by hierarchical synchronization [J]. *New Journal of Physics*, 2007, 9(6): 178
- [31] Zamora-López G, Zhou Changsong, Kurths J. Structural characterization of networks using the cat cortex as an example [G] //Lectures in Supercomputational Neurosciences: Dynamics in Complex Brain Networks. Berlin: Springer, 2008: 77–106
- [32] Boccaletti S, Bianconi G, Criado R, et al. The structure and dynamics of multilayer networks [J]. *Physics Reports*, 2014, 544(1): 1–122
- [33] Lü Lingyuan, Chen Duanbing, Ren Xiaolong, et al. Vital nodes identification in complex networks [J]. *Physics Reports*, 2016, 650: 1–63
- [34] Freeman L C. A set of measures of centrality based on betweenness [J]. *Sociometry*, 1977, 40(1): 35–41
- [35] Sabidussi G. The centrality index of a graph [J]. *Psychometrika*, 1966, 31(4): 581–603
- [36] Bonacich P. Factoring and weighting approaches to status scores and clique identification [J]. *The Journal of Mathematical Sociology*, 1972, 2(1): 113–120
- [37] Albert R, Jeong H, Barabasi A. Diameter of the World-Wide Web [J]. *Nature*, 1999, 401(6749): 130–131
- [38] Chen Duanbing, Lü Lingyuan, Shang Mingsheng, et al. Identifying influential nodes in complex networks [J]. *Physica A: Statistical Mechanics and Its Applications*, 2012, 391(4): 1777–1787
- [39] Gao Cai, Wei Daijun, Hu Yong, et al. A modified evidential methodology of identifying influential nodes in weighted networks [J]. *Physica A: Statistical Mechanics and Its Applications*, 2013, 392(21): 5490–5500
- [40] Chen Duanbing, Gao Hui, Lü Linyuan, et al. Identifying influential nodes in large-scale directed networks: The role of clustering [J]. *PLoS ONE*, 2013, 8(10): e77455
- [41] Wang Yafei, Yan Guanghui, Ma Qingqing, et al. Identifying influential spreaders on weighted networks based on ClusterRank [C] //Proc of Int Symp on Computational Intelligence and Design. Piscataway, NJ: IEEE, 2017: 476–479
- [42] Zhang Xin. Multilayer networks science: Concepts, theories and data [J]. *Complex Systems and Complexity Science*, 2015, 12(2): 103–107 (in Chinese)  
(张欣. 多层复杂网络理论研究进展: 概念、理论和数据[J]. 复杂系统与复杂性科学, 2015, 12(2): 103–107)
- [43] Battiston F, Nicosia V, Latora V. Structural measures for multiplex networks [J]. *Physical Review E*, 2014, 89(3): 32804
- [44] Cozzo E, Kivelä M, Domenico M D, et al. Structure of triadic relations in multiplex networks [J]. *New Journal of Physics*, 2015, 17(7): 73029
- [45] Huang Kezhi, Xue Mingde, Lu Mingwan. Tensor Analysis [M]. 2nd ed. Beijing: Tsinghua University Press, 2003 (in Chinese)  
(黄克智, 薛明德, 陆明万. 张量分析[M]. 2 版. 北京: 清华大学出版社, 2003)
- [46] Han Chongzhao, Zhu Hongyan, Duan Zhansheng. Multi-source Information Fusion [M]. 2nd ed. Beijing: Tsinghua University Press, 2010 (in Chinese)

- (韩崇昭, 朱洪艳, 段战胜. 多源信息融合[M]. 2 版. 北京: 清华大学出版社, 2010)
- [47] De Domenico M, Solé-Ribalta A, Cozzo E, et al. Mathematical formulation of multilayer networks [J]. Physical Review X, 2013, 3(4): 41022
- [48] Gómez S, Díaz-Guilera A, Gómez-Gardeñes J, et al. Diffusion dynamics on multiplex networks [J]. Physical Review Letters, 2013, 110(2): 028701
- [49] Solé-Ribalta A, De Domenico M, Kouvaris N E, et al. Spectral properties of the Laplacian of multiplex networks [J]. Physical Review E, 2013, 88(3): 032807
- [50] Estrada E, Gómez-Gardeñes J. Communicability reveals a transition to coordinated behavior in multiplex networks [J]. Physical Review E, 2014, 89(4): 042819
- [51] Newman M E. Properties of highly clustered networks [J]. Physical Review E: Statistical, Nonlinear, and Soft Matter Physics, 2003, 68(2): 26-121
- [52] Newman M E. Networks: An Introduction [M]. Oxford, England: Oxford University Press, 2010
- [53] Wei Daijun, Deng Xinyang, Zhang Xiaoge, et al. Identifying influential nodes in weighted networks based on evidence theory [J]. Physica A: Statistical Mechanics and Its Applications, 2013, 392(10): 2564-2575
- [54] Omodei E, De Domenico M D, Arenas A. Characterizing interactions in online social networks during exceptional events [J]. Frontiers in Physics, 2015, 3(15): 59
- [55] Reng Xiaolong, Lü Linyuan. Review of ranking nodes in complex networks [J]. Chinese Science Bulletin, 2014, 59(13): 1175-1197 (in Chinese)  
(任晓龙, 吕琳媛. 网络重要节点排序方法综述[J]. 科学通报, 2014, 59(13): 1175-1197)
- [56] Yagan O, Qian Dajun, Zhang Junshan, et al. Conjoining speeds up information diffusion in overlaying social-physical networks [J]. IEEE Journal on Selected Areas in Communications, 2013, 31(6): 1038-1048
- [57] Salehi M, Sharma R, Marzolla M, et al. Spreading processes in multilayer networks [J]. IEEE Transactions on Network Science and Engineering, 2015, 2(2): 65-83



**Luo Hao**, born in 1988. Master candidate. Member of CCF. His main research interest is vital nodes identification in multiplex networks.



**Yan Guanghui**, born in 1970. PhD, professor, PhD supervisor. Senior member of CCF. His main research interests include database theory and system, Internet of things engineering and application, data mining, complex network analysis, etc.



**Zhang Meng**, born in 1995. Master candidate. Her main research interest is high-order structure in complex networks.



**Bao Junbo**, born in 1994. Master candidate. His main research interest is information dissemination in complex networks.



**Li Juncheng**, born in 1994. Master candidate. His main research interest is representation learning in complex networks.



**Liu Ting**, born in 1993. Master candidate. Her main research interest is location privacy protection in mobile internet.



**Yang Bo**, born in 1979. Master, senior engineer. His main research interest is electric power information technology.



**Wei Jun**, born in 1987. Senior engineer. Her main research interest is electric power information technology.