异构信息网络融合方法研究综述*

田鹏伟[1,2] 张娴[1] 胡正银[1] 董坤[1,2] 许海云[1]
[1] 中国科学院成都文献情报中心 成都 610041 [2] 中国科学院大学 北京 100190

摘要: [目的/意义] 异构信息网络融合对于异构信息网络本身及其相关应用意义重大。综述异构信息网络融合方法,并进行客观的分析和评价,以期为进一步研究提供新的思路。 [方法/过程] 在对异构信息网络及其相关概念进行辨析的基础上,对异构信息网络融合方法进行调研、分析与归纳,评述该领域的研究现状,提出未来可能的研究方向。 [结果/结论] 异构信息网络融合方法分为基于元路径提取、多重关系网络及超图/超网络建模等五种类型,具体方法各有优势与局限;当前异构信息网络融合研究尚处起步阶段,研究方法有待丰富;基于元路径提取的融合方法显现不足;基于异构信息网络融合的应用型研究需进一步开拓。

关键词: 异构信息网络 网络融合 元路径 超图 超网络 聚类 分类
分类号: G250
DOI: 10.13266/j.issn.0252-3116.2017.07.019

随着大数据的迅速发展与计算能力的不断提升,各类学科越发期望通过一定的手段对多种数据(如农业数据、医疗数据、专利数据等)展开分析,挖掘这些数据中的有价值部分[1]。当前,面对数据量、数据维度的暴增,要想全面把握研究对象的特征,仅从单一维度对数据进行挖掘,其结论的准确性和全面性已显现不足,需引入多维视角。此外,真实的世界是异构的,不同类型事物间存在复杂的关系,而传统的网络分析方法难以直接迁移到异构信息网络(Heterogeneous Information Network,HIN)研究中,需利用异构信息网络对真实的世界进行表征。目前,就情报研究而言,随着分析工作精细化与准确化程度的不断提高,对多种信息进行融合并开展深入分析的要求也日益增加[2],但就如何对这些多源、异构信息通过网络建模进行融合仍存有诸多不足,这也是当前制约情报分析向前推进的重要因素。因此,本文拟对已有异构信息网络融合方法进行梳理,明确各类方法的优缺点及适用性,并对未来情报研究领域的多维数据分析做出展望。

1 异构信息网络相关概念

传统的网络分析方法[3]实质是基于同构信息网络开展的,是对异构信息网络某一维度的映射,或为研究的方便性而限定某些特征形成的同构信息网络。因此,该类研究往往不能对研究对象及其关系进行全面把握。因而,需利用异构信息网络对所研究问题进行建模。

1.1 异构信息网络

异构信息网络是由 Y. Sun 等[4]于 2009 年提出的,随后,基于异构信息网络的研究层出不穷,在情报学、数据挖掘、信息检索等领域成为热点。现就异构信息网络相关概念梳理如下(类似概念总结见 C. Shi 等的论文[5]):

(1) 异构信息网络。异构信息网络,指的是网络中节点对象类型及关系类型有许多种类,对于节点类型和关系类型而言,至少应存在两种。给定对象类型集合 $A=\{A\}$ 与关系类型集合 $R=\{R\}$,信息网络[6-7]通常定义为:

一个有向图 $G(V,E)$ 且具有一个对象类型映射函数 $\Phi: V\rightarrow A$ 和一个链接类型映射函数 $\varphi: E\rightarrow R$,其中,每个对象 $v\in V$ 属于一个特定的对象类型,记为 $\Phi(v)\in A$;每一个链接 $e\in E$ 属于一个特定的关系类型,

* 本文系中国科学院科技服务网络计划(STS 计划)项目"中国科学院知识产权信息服务"(项目编号: KFJ-SW-STS-190)与广东省省级科技计划项目"再生医学和组织工程知识集成服务技术研发与应用"(项目编号: 2016A040403098)研究成果之一。
作者简介: 田鹏伟(ORCID: 0000-0001-6784-3617),硕士研究生;张娴(ORCID: 0000-0002-6297-1190),副研究员,硕士生导师,通讯作者,E-mail: zhangx@clas.ac.cn;胡正银(0000-0002-5699-9891),副研究员,博士;董坤(ORCID: 0000-0001-8455-9),博士研究生;许海云(ORCID: 0000-0002-7453-3331),副研究员。
收稿日期: 2017-01-06 修回日期: 2017-03-20 本文起止页码: 137-144 本文责任编辑: 王善军

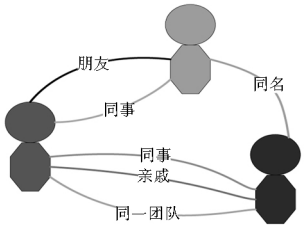
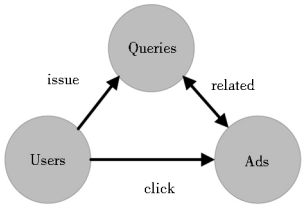
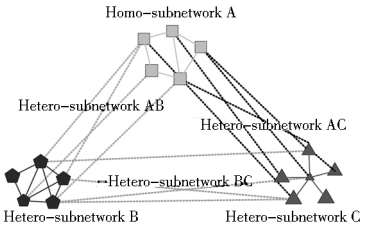
记为 $\varphi(e) \in R$。若两个链接同属一个关系类型，则两个链接具有相同的开始对象类型和结束对象类型。而当对象类型 $\lceil A \rceil > 1$ 或者关系类型 $\lceil R \rceil > 1$ 时，称该网络为异构信息网络，否则称为同构信息网络。

(2) 同质/异质异构信息网络。依据网络节点的同质、异质特性又可分为同质异构与异质异构信息网络，即若网络中节点对象集合 A 都属同一类型，而关系集合 R 属不同类型，且网络拓扑结构中存在多层网络结构体，则该网络称之为同质异构信息网络[8]，否则属于异质异构信息网络，常见于图像处理、通信领域等[9]。其中，各网络层级节点之间的链接无要求。若对象集合 A 和关系集合 R 都只有一种类型，且仅有一层网络结构体，则属同构信息网络。

1.2 异构信息网络及其相关概念辨析

在提出异构信息网络前已存有多种网络概念，如多维关系、多模网络与复杂网络等，如表1所示：

表1 异构信息网络相关概念辨析

名称	概念内容	网络特性	示意图
多维关系网络	网络中节点集合 A 只有一种类型，关系集合 R 存在多种分类	主要体现节点间的多关系集合，如常见社交网络、生物网络等[10]	 图1 多维关系网络[11]
多模网络	由多种类型节点(节点类型 > 3)构成的网络(由 L. Tang 等[12]在二模网络基础上提出)	多模体现在节点类型的不同，研究多模对象间的链接关系[13]	 图2 多模网络[12]
复杂网络	网络拓扑结构具有一般性，节点间链接关系既非规则也非纯粹随机[14]	真实的网络是复杂网络[15]，具有重尾性、高聚集系数及社区化，如万维网、社交、神经网络等	 图3 复杂网络[16]

通过上述分析，笔者认为多维关系网络与多模网络相比，除链接关系都可涵盖多类之外，最大不同在于后者建模的节点类型不同，但二者均是异构信息网络的一种特例，只是异构信息网络建模更强调节点间的交互作用，挖掘其中的语义内容[17]，使分析更具准确性。与复杂网络相比，异构信息网络侧重对微观结构的解读[7]，而复杂网络主要集中对宏观信息的把握[18]。实际应用中因数据分析需求、网络规模及应用场景的不同各有侧重。

2 异构信息网络融合方法研究进展

所谓异构信息网络融合，是指对已存在的若干信息网络进行融合使其拓扑结构合为一体，或对客观世界复杂系统中的若干组实体进行网络结构重新构建。当前，融合方法种类繁多，研究角度不尽相同，难以给出明确的分类体系。本文试图从网络结构角度对现有研究方法进行总结，主要分析网络的内、外部结构两个方面。

利用关键词“异构信息网络、异构网络”+“融合”在 CNKI 中进行检索，及“heterogeneous information network、heterogeneous network”+“fusion”于谷歌学术中进行检索，共获得约 240 篇文献，通过人工删选(重点关注图情与计算机领域)剩余 200 篇。

2.1 网络内部结构方面

从异构信息网络内部结构切入，大体集中于三类

研究：基于2－模或3－模网络建模的融合、基于元路径提取的融合及基于多重关系网络的融合。

2.1.1 基于2－模或3－模网络建模的融合 2－模/3－模网络建模是最简单的异构信息网络融合研究。通常是在网络结构中获取具有代表性的异质节点，通过节点间的链接关系构建网络达成融合。张自立等[19]利用2－模网络建模对文献特征网络共现关系进行融合，揭示某研究领域的重要机构及研究热点。许海云等[20]将2－模网络社区识别方法应用于交叉学科网络，增强了研究主题识别效果。从模型构建的角度来说网络建模的同时也是融合的过程。也有学者将“2－模网络”向“多模网络”推进。L. Leydesdorff[21]利用3－模网络建模将作者－期刊－关键词联系起来，反映真实的合作关系、研究主题等。李长玲等[22]以知识网络为例构建3－mode共现关系网络，挖掘代表性作者及合著团队、经典和热点研究等。上述研究说明利用2－模或3－模网络建模可对异构信息网络进行融合。

2.1.2 基于元路径提取的融合 从网络的内部结构入手提取异构信息网络中的“元路径”是另一种解决思路。有学者提出利用元路径提取对异构信息网络进行融合[23]。所谓元路径是指在异构信息网络中获取不同对象间的关联路径。即给定异构信息网络G(A，R)，A的元路径P定义为：$A_1 \xrightarrow{R_1} A_2 \xrightarrow{R_2} \cdots \xrightarrow{R_L} A_{L+1}$，表示节点$A_1$和$A_{L+1}$间的关系组合。$R = R_1 \circ R_2 \circ \cdots \circ R_{L+1}$，其中，$\circ$表示关系间的组合。如图4所示：

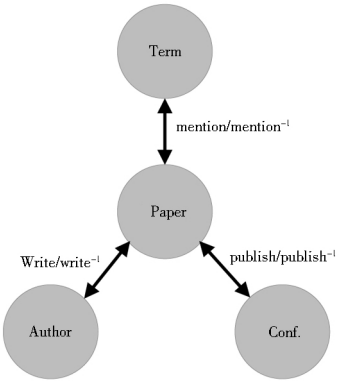


图4 异构信息网络模式[24]

其中，涉及的元路径有“$author \xrightarrow{write} paper \xrightarrow{write^{-1}} author$”“$author \xrightarrow{write} paper \xrightarrow{publish^{-1}} conference \xrightarrow{publish} paper \xrightarrow{write^{-1}} author$”等，可见不同元路径其链接的节点不同，且多条元路径可看作是对不同网络层级的关联。

基于元路径的多种优势，利用元路径提取进行异构信息网络融合研究不断涌现，有学者将元路径提取同其他研究方法相结合达到融合目的。G. Fu等[25]提出一种基于元路径提取的监督型机器学习模型，对生物医学领域的异构信息网络进行有效的融合。Y. Sun等[26]结合元路径提取与用户指导提出PathSelClus算法，利用机器学习思想对大规模网络进行聚类，增强聚类质量。Y. Zhao等[27]提出了一种新的基于元路径提取的非负矩阵分解框架(MPNMF)对异构信息网络进行融合，从而提高提高聚类的性能。也有学者就元路径提取本身存在的不足进行修正。B. Shi等[28]与C. Meng等[29]针对元路径提取的不足，如数据规模小[30]、元路径需手动设定[31]，将研究拓展至更大的数据集(如Yago、Wikipedia)。X. Kong等[32]在原有元路径的基础上提出链接路径依赖关系概念，利用对象间的依赖关系提出集体分类方法。C. Luo等[33]引入关系路径概念，提出一种半监督学习框架SemiRPClus，而链接路径依赖与关系路径其实质仍是元路径提取思想。

2.1.3 基于多重关系网络的融合 多重关系网络的提出是为符合研究者的目的[34]，即从网络结构内部提取出不同关系构建多种同质/异质信息网络，而融合操作主要是对这些子网络进一步处理形成融合。其中，包括线性与非线性两种运算。

(1)线性运算，即对融合子网络赋权后，通过线性加权方式使得融合结果形成一个整合网络。X. He等[35]人利用文本相似度调节文本链接的强度并采用线性相加方式与同被引网络整合为一个加权邻接网络。X. Liu等[36]提出了一种混合聚类框架，即对多种数据源构建的混合网络进行线性加权处理，解决大规模期刊的混合聚类问题。D. R. Liu等[37]提出一种混合专利分类方法，即利用线性加权方式对多重专利网络分类结果进行融合，提高了专利查询和分类的准确性。M. Magnani等[38]融合线上与线下多重社交关系网络，挖掘相似节点并进一步开展链接预测及社团识别研究。

(2)非线性运算，虽然线性融合方法可解决异构信息网络融合，但其并非最优选择。F. Janssens[39]认为加权线性组合存在局限，提出利用逆Fisher卡方方法将词－文档矩阵与引用－文档矩阵融合。随后，有学者利用迭代运算中进行概率赋值和矩阵相乘两种方式展开非线性融合。B. T. Dai等[40]利用多元概率分布对多重关系网络建模，提出广义随机块模型方法进

行网络融合操作。吴蕾等[41]结合多主题模型与马尔科夫逻辑网提出融合概率图模型,结果显示该方法比单纯指定路径权重的分类算法准确性更高。刘彤等[34]利用元矩阵对由专利及其属性构成的关系网络进行表征,引入LDA算法融合多重关系网络,进而开展子群划分研究。张邦佐等[42]结合矩阵相乘与元路径提取提出一种基于矩阵分解的推荐算法,提高推荐的精准性。

2.2 网络整体结构方面

仅从网络内部结构入手有时并不能表征研究对象的整体特征,因而,需通过网络整体结构对网络建模与整合,主要涉及基于聚类/分类的融合与基于超图/超网络建模的融合。

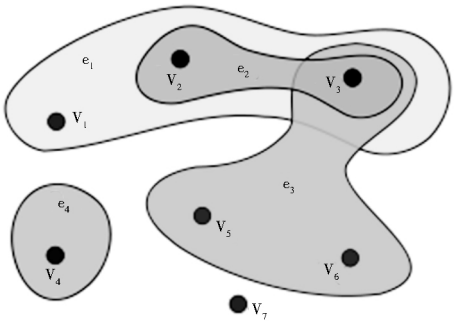
2.2.1 基于聚类/分类的融合 聚类与分类操作贯穿网络分析方法始终,此处笔者重点关注从网络整体结构角度出发考虑的聚类和分类操作。

(1)聚类是一种无监督学习行为,即将属性相似的数据聚集到一个簇内,相异的数据划分在不同的簇中。Y. Sun等[43]基于bi-typed型异构信息网络,提出一种基于排序(互增强)的聚类方法RankClus,其结果优于NCut[44]、SimRank[43]算法。随后,Y. Sun等[45]针对RankClus不能解决跨多对象的异构信息网络融合问题,将研究对象扩展到"星型"异构信息网络上,提出迭代增强融合方法NetClus生成高质量的网络簇。之后,Y. Sun等[46]在NetClus的基础上又考虑网络属性、语义结构在聚类中重要程度不一,提出一种自动化识别路径的概率模型算法GenClus。M. Zitnik等[47]采用直和方式形成新的矩阵,提出一种惩罚矩阵三因素分解方法,具有较好的聚类效果。随后,利用M. Zitnik等提出的融合算法学者们开展了一系列研究,如病毒预测研究[48]、精准医疗研究[49]等。

(2)分类与聚类操作是截然不同的两种技术,需根据已有的标记信息进行分类。分类研究已在同构信息网络中取得一定成果[50-51],基于异构信息网络的分类研究虽处起步阶段,但已引起学者们的关注。M. Ji等[52]借鉴转换分类思想,提出了一种基于图的正则化分类框架GNetMine,即对不同类型的对象和链接进行统一处理,使用相同的分类标准,实证表明该方法较算法LLGC[53]具有一定优越性。C. Luo等[24]突破GNetMine算法假设,认为网络中不同分类对象具有不同分类标准,提出利用关系路径促进分类的HetPathMine算法,结果显示HetPathMine的分类结果较准确。

2.2.2 基于超图、超网络建模的融合 除利用多重关系网络建模之外,还可站在网络集合之外对其"俯瞰",而超图、超网络是有效的"俯瞰"手段,可描述最具一般的异构关系,克服经典图论在多元关系表征上的局限。

(1)基于超图的融合。超图[54-55]概念由C. Berge于1973年提出,近年随着多源数据、异构信息网络的盛行,学者们逐渐关注超图理论及其应用。目前对于超图的定义比较认可的是:超图是一种形如$H=(V,E)$的节点与边对,其中V称之为节点或顶点的一组元素,E是V的非空子集。通俗来说,超图是图的泛化,其中一条边(E)可以包含任何数量的顶点(V),一个顶点(V)可以隶属多个超边(E),如图5所示:



$X=\{v_1,v_2,v_3,v_4,v_5,v_6,v_7\}$ 和 $E=\{e_1,e_2,e_3,e_4\}=\{\{v_1,v_2,v_3\},\{v_2,v_3\},\{v_3,v_5,v_7\},\{v_4\}\}$

注:X表示由$v_1,v_2,v_3,v_4,v_5,v_6,v_7$节点构成的集合;$E$表示超边集合,其中集合$\{v_1,v_2,v_3\}$的超边为$e_1$

图5 超图示意图[56]

因超图是图的泛化,所以异构信息网络可看作同构信息网络的共现。有学者将超图应用在异构信息网络融合研究中。如在计算机领域,F. Chen等[57]充分考虑多模网络内部与多模态间的依赖性,采用超图对微博信息建模用于情绪预测研究。罗永恩等[58]结合共享熵提出一套模块划分方法,对原始异构信息网络进行聚类划分,该方法具有良好的融合效果。在情报学领域,蔡淑琴等[59]利用超图模型有效的标识知识间的多元关系及层次与非层次结构,达到融合目的。许海云等[20]指出借鉴超图思想将有助于基于学科交叉网络融合进行主题发现研究。

(2)基于超网络的融合。超网络最早由Y. SHEFFI[60]提出,王众托等[61]把高于而又超于现存网络的网络称为超网络(super network),使其含义明确。

超网络是基于网络的网络,利用超网络对异构信息网络进行建模达成融合具有自身优势。Q. Suo等[62]针对由书籍、电影和音乐评论构成的异构信息网

络，提出了一种应用超网络分析用户评级数据的新方法。滕立[63]在超网络基础上提出共现网络分析方法，消除了异构信息网络中的孤立节点。Y. S. Zhao等[64]通过研究用户与项目间的关系，将其映射到超网络中，并利用距离测量方法进行网络融合，挖掘用户偏好特征。

3 异构信息网络融合方法研究述评

异构信息网络融合主要解决多维、异构数据的网络建模、聚类/分类问题，通过对研究对象整体特征的归纳以实现聚类/分类，而融合的关键在于特征提取和融合算法的效果，而具体应用方面显现不足。另外，即便些许融合研究中涉及具体应用型数据集(如专利、DBLP等)也只是针对性研究，难以移植于其他数据集，其应用拓展性存在局限。

表2 异构信息网络融合方法对比

融合角度	融合方式	优势	不足
数学运算	线性	模型易构建，权重易调整	人为因素大，不宜直接加权[39]
	非线性	矩阵分解优势明显，概率模型语义表达准确	概率建模假设各对象独立[40]，对矩阵大小的要求较严苛
网络结构(内部)	2－模/3－模建模	模型易构建，便于可视化	异构节点类型较少，研究中未能充分体现语义作用
	元路径提取	融合准确性高，具有语义表征性	路径提取需人工标注，路径不宜过长，不能表针更具细粒度的语义
	多重关系网络	网络间关联紧密，模型易构建，便于计算	子网络是同构性质的，且子网络间通常需具有较强的链接关系
网络结构(整体)	聚类/分类	融合“监督”思维，发挥机器学习优势	聚类簇需人为设定，方法应用对数据集依赖性强
	超图/超网络	将图划分方法拓展到异构信息网络中	现有研究多集中于概念、数学模型构建等，具体方法及应用较少

在尽量减少人为参与的情况下，如何更大程度地进行自动化融合是当前研究的目标，各种方法在融合处理中虽有不足，但基于元路径提取或聚类/分类融合方式，无论在融合特征的数目、网络复杂程度等方面均表现较优。

3.2 研究方法有待在大规模网络中验证

当前，融合方法多基于小型数据集，且多为结构化数据集，如专利数据集、DBLP[65]及IMDb[66]等。其中，2－模/3－模网络建模融合中涉及的数据集显然不足以与DBLP或IMDb相提并论；所谓元路径提取多数集中在三种节点中，路径长度并未超过8；多重关系网络构建多数集中在三或四种网络模型，而更多维异构子网络并未在研究中体现；而基于超图/超网络的融合研究目前集中对原有多维、异构数据重新建模，而非针对大规模异构数据进行处理。因此，面向较大数据集合(如Facebook、Twitter及DBpedia[67]等)验证现有异构信息网络融合方法，其可行性与质量均有待商榷。

3.3 偏重算法研究，应用性研究较少

目前，异构信息网络融合重算法研究，应用性研究较少，主要表现在：基于2－模/3－模网络建模融合相对成熟，存在些许应用性研究；基于元路径提取相关研究仍处起步阶段，针对性应用研究较少；多重关系网络融合及聚类/分类融合研究则注重算法间的比较，追求网络搭建。总体看来，异构信息网络融合已开始由线性向非线性过渡，且建模维度更加丰富，现阶段异构信息网络融合研究已经具备一定的积累，形成了一些基本的分析方法，但也存在些许局限与不足。

3.1 网络种类繁多，研究方法难以统一

由于多维关系、多模网络及异构信息网络等概念的存在，研究者们基于不同的概念提出各自的研究方法，但未对上述概念加以辨析，并对其方法的共性给予分析，导致融合方法存在混淆。此处将对现有融合方法进行对比，揭示其优势与不足，如表2所示：

4 未来研究展望

当前，情报学领域的异构信息网络融合研究相对较少，该类探究尚处研究初期，其中融合方法构建与应用性研究均有待完善。此外，随着数据维度和体量的剧增，异构信息网络结构的划分、语义关联等研究面临挑战。通过分析异构信息网络现有融合方法存在的缺陷与不足，提炼出可用于情报分析的融合方法，并进一步深入拓展该类方法。

4.1 归纳网络概念并探究有效的融合方法

首先，对比并归纳现有信息网络概念；其次，针对特定的数据集构建不同的异构信息网络，并对其进行取舍达成融合要求；最后，研究更深层次、广泛性融合方式，具体阐述如下：

(1)对比并归纳现有网络概念。通过对不同的网络概念进行对比、归类，有利于学者们理清网络建模思路，针对不同数据集提取所需特征并进行特定“异构信息网络”构建，采用特定的融合方法，同时也有助于进一步研究深层次融合体系。

(2) 融合异构信息网络,达成情报分析需求。依据不同的分析需求获取特定的特征构建异构信息网络,或对同质信息网络进行加权融合使得融合结果更符合情报需求。如何进一步提取更具表征性、语义特性的元路径并以自动化方式辅之,更是未来研究的热点。尽可能多地构建异质多重关系网络,同时应充分发挥聚类和分类操作在融合处理中的作用。有效利用超图/超网络的建模思想,促进网络整体和内部结构融合方法的优势互补。

(3) 研究更深层次融合方法。借助矩阵相乘、分解运算是解决由异构特征构建的异构信息网络融合的有效途径,但需注意运算中各分解值所表征的实际含义。利用概率建模捕捉异构信息网络中各链接的语义重要性,或利用信息互增强迭代运算进行网络融合均是未来可以继续深入的研究方向。此外,将矩阵表征与概率建模相结合的融合概率图模型将大大提高异构信息网络融合的效率及其效果[38]。但关于该方面的研究目前较少,且针对具体融合研究的应用涉及不多,是未来可进一步推进的地方。

4.2 推动融合研究面向大数据集情报分析与应用性研究

(1) 推动融合研究面向大型异构数据集合。随着情报分析数据维度的不断增加和体量上涨,异构信息网络融合研究已不仅局限于解决中小规模异构数据集的聚类/分类问题,而应在结合高效计算方式(如人工智能、深度学习等)的前提下,构建维度更多的异构信息网络,全面探究研究对象的特征,其关键在于如何进行数据特征的提取及网络构建。

(2) 推动融合研究面向情报分析应用性研究。异构信息网络融合研究可用于解决多源、异构数据整合,尤其针对情报研究中的多维数据分析问题,通过提取多维数据特征、构建异构信息网络进而利用上述融合方法进行综合性、全面性分析实现对研究对象的聚类/分类。如针对多学科交叉进行主题发现研究,开展多源、异构的文献检索研究,进行各类在线文献系统推荐研究,以及基于异构信息网络融合的科研人才评价体系建设等。这些需求的满足都迫切的需要异构信息网络融合方法的支撑。

参考文献:

[1] CHEN C L P, ZHANG C Y. Data-intensive applications, challenges, techniques and technologies: a survey on big data [J]. Information sciences, 2014, 275(11): 314 - 347.

[2] 孟祥保. 图书情报学交叉融合与发展——基于国外35种核心期刊的引文分析[J]. 图书情报知识, 2012(5): 50 - 58.

[3] PARK H W. Hyperlink network analysis: a new method for the study of social structure on the web [J]. Connections, 2003, 25(1): 49 - 61.

[4] SUN Y, HAN J, ZHAO P, et al. Rankclus: integrating clustering with ranking for heterogeneous information network analysis [C]//Proceedings of the 12th international conference on extending database technology: advances in database technology. New York,: ACM, 2009: 565 - 576.

[5] SHI C, LI Y, ZHANG J, et al. A survey of heterogeneous information network analysis [J]. Computer science, 2015, 134(12): 87 - 99.

[6] JI M, HAN J, DANILEVSKY M. Ranking-based classification of heterogeneous information networks [C]//Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining. New York: ACM, 2011: 1298 - 1306.

[7] SUN Y, HAN J. Mining heterogeneous information networks: principles and methodologies [J]. ACM SIGKDD explorations newsletter, 2012, 14(2): 439 - 473.

[8] FEDERICA. Innovating in computing network architectures [EB/OL]. [2016 - 11 - 03]. http://www.fp7-federica.eu/infrastructure/architech.php.

[9] 王海涛, 付鹰. 异构网络融合——研究发展现状及存在的问题[J]. 数据通信, 2012(2): 18 - 21.

[10] BOUANAN Y, RIBAULT J, FORESTIER M, et al. Modeling and simulation of human reaction in a multidimensional social network [J]. IFAC-PapersOnLine, 2015, 48(3): 592 - 597.

[11] BERLINGERIO M, COSCIA M, GIANNOTTI F, et al. Foundations of multidimensional network analysis [C]//International conference on advances in social networks analysis and mining. Piscataway: IEEE, 2011: 485 - 489.

[12] TANG L, LIU H, ZHANG J, et al. Community evolution in dynamic multi-mode networks [C]//Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining. New York,: ACM, 2008: 677 - 685.

[13] 王娜, 李霞, 徐红英. 社会网络分析之社区发现研究[J]. 深圳大学学报(理工版), 2014, 31(1): 35 - 42.

[14] BOCCALETTI S, LATORA V, MORENO Y, et al. Complex networks: structure and dynamics [J]. Physics reports, 2006, 424(4/5): 175 - 308.

[15] STROGATZ S H. Exploring complex networks [J]. Nature, 2001, 410(6825): 268 - 76.

[16] HWANG T H, KUANG R. A heterogeneous label propagation algorithm for disease gene discovery [C]//Proceedings of the 2010 SIAM International conference on data mining. Philadelphia: SIAM, 2010: 583 - 594.

[17] SUN Y, TANG J, HAN J, et al. Community evolution detection in dynamic heterogeneous information networks [C]//Proceedings of the eighth workshop on mining and learning with graphs. New

York：ACM，2010：137－146.

[18] RAVASZ E，BARABASI A，OLTVAI Z. Hierarchical organization of complex networks [EB/OL]. [2016－10－11]. https://repository.library.northeastern.edu/files/neu:331069/fulltext.pdf.

[19] 张自立，张紫琼，李向阳. 基于2－模网络的科研单位和关键词共现分析方法 [J]. 情报学报，2011，30(12)：1249－1260.

[20] 许海云，郭婷，岳增慧，等. 基于TI指标系列的情报学学科交叉主题研究 [J]. 情报学报，2015，34(10)：1067－1078.

[21] LEYDESDORFF L. What can heterogeneity add to the scientometric map? Steps towards algorithmic historiography [EB/OL]. [2016－12－20]. http://leydesdorff.net/mcallon/mcallon.pdf

[22] 李长玲，刘非凡，魏绪秋. 基于3－mode网络的领域主题演化规律分析——以知识网络研究领域为例 [J]. 情报理论与实践，2014，37(12)：104－110.

[23] SUN Y，HAN J，YAN X，et al. PathSim：meta path-based top-k similarity search in heterogeneous information networks [J]. Proceedings of the VLDB endowment，2011，4(11)：992－1003.

[24] LUO C，GUAN R，WANG Z，et al. HetPathMine：A novel transductive classification algorithm on heterogeneous information networks [M]//Advances in Information Retrieval. 2014：210－221.

[25] FU G，DING Y，SEAL A，et al. Predicting drug target interactions using meta-path-based semantic network analysis [J]. BMC bioinformatics，2016，17(1)：1－10.

[26] SUN Y，NORICK B，HAM J，et al. PathSelClus：integrating meta-path selection with user-guided object clustering in heterogeneous information networks [J]. ACM transactions on knowledge discovery from data，2012，7(3)：723－724.

[27] ZHAO Y，SUN Z，XU C，et al. Meta-path based nonnegative matrix factorization for clustering on multi-type relational data [C]//International joint conference on neural networks (IJCNN). Piscataway：IEEE，2015：1－8.

[28] SHI B，WENINGER T. Mining interesting meta-paths from complex heterogeneous information networks [C]//IEEE International conference on data mining workshop. Piscataway：IEEE，2014：488－495.

[29] MENG C，CHENG R，MANIU S，et al. Discovering meta-paths in large heterogeneous information networks [C]//Proceedings of the 24th international conference on World Wide Web. New York：ACM，2015：754－764.

[30] SUN Y，BARBER R，GUPTA M，et al. Co-author relationship prediction in heterogeneous bibliographic networks [C]//International conference on advances in social networks analysis and mining. Piscataway：IEEE，2011：121－128.

[31] SHI C，KONG X，HUANG Y，et al. HeteSim：a general framework for relevance measure in heterogeneous networks [J]. IEEE transactions on knowledge and data engineering，2014，26(10)：2479－2492.

[32] KONG X，YU P S，DING Y，et al. Meta path-based collective classification in heterogeneous information networks [C]//Proceedings of the 21st ACM international conference on information and knowledge management. New York：ACM，2012：1567－1571.

[33] LUO C，PANG W，WANG Z. Semi-supervised clustering on heterogeneous information networks [C]//Pacific-Asia conference on knowledge discovery and data mining. Berlin，German：Springer，2014：548－559.

[34] 刘彤，杨冠灿，侯元元. 基于多重关系整合的专利网络分析方法研究与应用 [J]. 情报理论与实践，2016，39(2)：59－63.

[35] HE X，ZHA H，DING C H Q，et al. Web document clustering using hyperlink structures [J]. Computational statistics & data analysis，2002，41(1)：19－45.

[36] LIU X，SHI Y，FRIZO J，et al. Weighted hybrid clustering by combining text mining and bibliometrics on a large-scale journal database [J]. Journal of the American society for information science & technology，2010，61(6)：1105－1119.

[37] LIU D R，SHIH M J. Hybrid-patent classification based on patent-network analysis [J]. Journal of the American society for information science & technology，2011，62(2)：246－256.

[38] MAGNANI M，MICENKOVA B，Rossi L. Combinatorial analysis of multiple networks [EB/OL]. [2016－10－20]. http://ia601009.us.archive.org/19/items/arxiv－1303.4986/1303.4986.pdf.

[39] JANSSENS F，GLANZEL W，MOOR B D. A hybrid mapping of information science [J]. Scientometrics，2008，75(3)：607－631.

[40] DAI B T，CHUA F C T，LIM E P. Structural analysis in multi-relational social networks [EB/OL]. [2016－10－12]. http://epubs.siam.org/doi/pdf/10.1137/1.9781611972825.39.

[41] 吴蕾，张文生，王珏. 异构信息网络数据上的融合概率图模型 [J]. 计算机科学与探索，2014，8(6)：712－718.

[42] 张邦佐，桂欣，何涛，等. 一种融合异构信息网络和评分矩阵的推荐新算法 [J]. 计算机研究与发展，2014(s2)：69－75.

[43] JEH G，WIDOM J. SimRank：a measure of structural-context similarity [C]//Proceedings of the eighth ACM SIGKDD international conference on knowledge discovery and data mining. New York：ACM，2002：538－543.

[44] SHI J，MALIK J. Normalized cuts and image segmentation [J]. IEEE transactions on pattern analysis and machine intelligence，2000，22(8)：888－905.

[45] SUN Y，YU Y，HAN J. Ranking-based clustering of heterogeneous information networks with star network schema [C]//Proceedings of the 15th ACM SIGKDD international conference on knowledge discovery and data mining. New York：ACM，2009：797－806.

[46] SUN Y，AGGARWAL C C，HAN J. Relation strength-aware clustering of heterogeneous information networks with incomplete attributes [J]. Computer science，2012，5(5)：394－405.

[47] ZITNIK M，ZUPAN B. Data fusion by matrix factorization [J]. IEEE transactions on pattern analysis & machine intelligence，2015，37(1)：41－53.

[48] ZITNIK M，ZUPAN B. Matrix factorization-based data fusion for

drug-induced liver injury prediction [J]. Systems biomedicine, 2014, 2(1): 16-22.

[49] REGENBOGEN S, WILKINS A D, LICHTARGE O. Computing therapy for precision medicine: collaborative filtering integrates and predicts multi-entity interactions [C]. Pacific symposium on biocomputing. Pacific symposium on biocomputing. NIH public access, 2016, 21: 21.

[50] MACSKASSY S A, PROVOST F. Classification in networked data: a toolkit and a univariate case study [J]. Journal of machine learning research, 2007, 8(May): 935-983.

[51] YIN Z, LI R, MEI Q, et al. Exploring social tagging graph for web object classification [C]//Proceedings of the 15th ACM SIGKDD international conference on knowledge discovery and data mining. ACM, 2009: 957-966.

[52] JI M, SUN Y, DANILEVSKY M, et al. Graph regularized transductive classification on heterogeneous information networks [C]// European conference on machine learning and knowledge discovery in databases. Berlin, Heidelberg: Springer, 2010: 570-586.

[53] ZHOU D, BOUSQUET O, LAL T N, et al. Learning with local and global consistency [J]. Advances in neural information processing systems, 2004, 16(16): 321-328.

[54] BERGE C. Graphs and hypergraphs [M]. New York: Elsevier, 1973.

[55] BERGE C. Hypergraphs: Combinatorics of finite sets [M]//North Holland: Elsevier, 1989: 521-552.

[56] Hypergraph [EB/OL]. [2016-12-08]. https://en.wikipedia.org/wiki/Hypergraph#Theorems.

[57] CHEN F, GAO Y, CAO D, et al. Multimodal hypergraph learning for microblog sentiment prediction [C]//Multimedia and Expo (ICME), 2015 IEEE International Conference on. Piscataway: IEEE, 2015: 1-6.

[58] 罗永恩，胡继承，徐茜. 基于超图的多模态关联特征处理方法[J]. 计算机工程，2017，43(1)：226-230.

[59] 蔡淑琴，肖泉，吴颖敏. 基于超图的知识表示及检索相似性度量研究[J]. 图书情报工作，2009，53(8)：102-105.

[60] SHEFFI Y. Urban transportation networks: equilibrium analysis with mathematical programming methods [J]. Transportation science, 1985, 19(4): 463-466.

[61] 王众托，王志平. 超网络初探[J]. 管理学报，2008，5(1)：1-8.

[62] SUO Q, SUN S, HAJLI N, et al. User ratings analysis in social networks through a hypernetwork method [J]. Expert systems with applications, 2015, 42(21): 7317-7325.

[63] 滕立. 基于超网络的作者-机构-国家混合共现网络研究[J]. 情报学报，2015(1)：28-36.

[64] ZHAO Y S, SHEN B. Empirical study of user preferences based on rating data of movies [EB/OL]. [2016-11-12]. http://europepmc.org/backend/ptpmcrender.fcgi?accid=PMC4703247&blobtype=pdf.

[65] DBLP. DBLP XML database [EB/OL]. [2016-12-01]. http://dblp.org/xml/dblp.dtd.

[66] IMDb. IMDb database [EB/OL]. [2016-12-01]. http://www.imdb.com/interfaces.

[67] DBpedia. DBpedia database [EB/OL]. [2016-12-01]. http://wiki.dbpedia.org/downloads-2016-04.

作者贡献说明：

田鹏伟：负责研究设计、文献调研及论文撰写；

张娴：负责框架设计、论文修改及撰写指导；

胡正银：负责框架设计；

董坤：负责观点提炼、论文修改；

许海云：负责观点提炼、论文修改。

Review of Studies on Heterogeneous Information Network Fusion Methods

Tian Pengwei[1,2] Zhang Xian[1] Hu Zhengyin[1] Dong Kun[1,2] Xu Haiyun[1]

[1] Chengdu Library and Information Center, Chinese Academy of Sciences, Chengdu 610041

[2] University of Chinese Academy of Sciences, Beijing 100190

Abstract: **[Purpose/significance]** The study of heterogeneous information networks (HINs) fusion is of great significance for heterogeneous information network analysis and their applications. In this paper, some typical methods are reviewed in order to provide new ideas for further research. **[Method/process]** Based on the comparison of the concepts between HINs and other networks, this paper investigated, analyzed and summarized the research methods of HINs fusion, reviewed the current research in this field, and put forward the future research directions possibly. **[Result/conclusion]** The methods of HINs fusion are divided into five types: meta-path extraction, multi-relational network and hyper-graph/super-network modeling, and so on. and these specific methods have their own advantages and limitations. At present, HINs fusion is still at the initial stage. The research methods and applications based on HINs fusion needs further exploration.

Keywords: heterogeneous information network network fusion meta-path hyper-graph super-network clustering classification