

文章编号: 1003-0077(2018)07-0137-06

面向拓片信息的甲骨字网络构建与分析

焦清局^{1,2,3}, 高峰^{1,2}, 金园园¹, 熊晶^{1,2}, 刘永革^{1,2,3}

- (1. 安阳师范学院 计算机与信息工程学院, 河南 安阳 455000;
2. 河南省甲骨文信息处理重点实验室, 河南 安阳 455000;
3. 汉语海外传播河南省协同创新中心, 河南 安阳 455000)

摘要: 未识甲骨字的考释是甲骨文研究最重要的内容,也是历史学家和计算机学家研究甲骨文遇到的最大瓶颈。甲骨文研究积累的数据已体现出海量化和系统化。因此,该文以甲骨文拓片为基础数据,通过建模定义甲骨字之间的距离,进而构建甲骨字网络。在此网络之上,分析网络的度分布、局部连接比率、聚类系数、模块度等相关特性。结果表明:构建的甲骨字网络不仅能充分反映甲骨文系统的单音节词多和复音节词少的古文字特征,而且能捕捉甲骨文拓片的语义单元,并具有很强的模块特性。该文构建的网络及其特性可为历史学家和网络甲骨学家揭示未知甲骨字的语义提供新的数据和理论基础。

关键词: 甲骨字;拓片;复杂网络;构建;分析

中图分类号: TP391

文献标识码: A

Construction and Analysis of Rubbing-oriented Oracle Character Network

JIAO Qingju^{1,2,3}, GAO Feng^{1,2}, JIN Yuanyuan¹, XIONG Jing^{1,2}, LIU Yongge^{1,2,3}

- (1. School of Computer & Information Engineering, Anyang Normal University, Anyang, Henan 455000, China;
2. Key Laboratory of Oracle Information Processing in Henan Province, Anyang, Henan 455000, China;
3. Collaborative Innovation Center of International Dissemination of Chinese Language, Anyang, Henan 455000, China)

Abstract: Predicting semantics of unknown oracle bone inscriptions is an important topic, and also the bottleneck for histories and computer scientists in oracle bone inscriptions research. Based on the massive accumulated data of oracle, we first define the distance of two oracle character by modeling on big rubbing information, then construct the network of oracle characters in this work, and, finally, analyze the degree distribution, local link rate, clustering coefficient and modularity of the network. Experiments show that the constructed network has a strong module structure, reflects the feature of more monosyllabic words and less polysyllabic words, and captures semantic units of rubbings.

Key words: oracle bone inscriptions; rubbing; complex network; construction; analysis

0 引言

甲骨文是一种距今有 3 500 多年历史的古汉字,其所记载的内容极其丰富,涉及商代社会生活的诸多方面,不仅包括政治、军事、文化、社会习俗等内容,而且涉及天文、历法等科学技术^[1]。对甲骨文进

行深入的研究与探讨,可以对语言文字学、考古学、历史学、社会人类学等学科产生深刻的影响,具有重要的文化价值和传承意义。目前国内外共收藏的甲骨片约有 15 万片^[2](数字会随时间的变化而变化,如新出土甲骨片及甲骨片的缀合),被发现的甲骨文单字约有 4 378 个^[3](2010 年前所见殷墟甲骨字),但其中可释者仅为 1 682 个^[3]。因此,考释字是甲

收稿日期: 2017-07-24 定稿日期: 2017-11-27

基金项目: 国家社会科学基金(16@ZH017A3);国家语委科研规划项目(YB135-50);国家自然科学基金(U1504612);河南省高等学校重点科研项目(17B520001);河南省教育厅人文社会科学研究一般项目(2017-ZZJH-005)

骨文研究的主要任务。一百多年来,经过几代学者的努力,在甲骨文字的考释方面取得了丰硕的成果^[4]。如郭沫若是第一个运用历史唯物主义研究古文字的人,善于将文字考释与史料分析相结合,进而研究中国古代社会。他在甲骨文字考释方面取得了令人瞩目的成绩。历史学家唐兰,他不但考释出很多难以辨识的甲骨文字,还建立了各种较为完整和系统的研究方法,如对照法、推勘法、偏旁分析法等。其他如董作宾、徐中舒也在甲骨文考释方面取得了丰硕的成果。然而,纯人工地研究甲骨文也存在许多缺陷:①专家们对甲骨文的辨识和翻译只能依赖于长期的学术钻研和经验累积,而且培养一名甲骨文专家是无法速成的,往往需要一二十年甚至更长的时间^[1];②甲骨文字之间不是孤立存在的,它们之间相互联系和作用,从而形成一个成熟的文字系统。然而,历史学家研究甲骨文的考释时,只是孤立的研究,并不能从系统的角度揭示甲骨文字的场景和语义。

随着对甲骨文的深入研究和其材料的数字化,甲骨文的数据已体现出海量化和系统化^[5],这为使用计算机技术研究甲骨文提供了数据上的保障。甲骨文的研究也渐渐进入了以计算机研究和人工研究相辅相成的时代。为了解决甲骨文资料库的缺陷,2007年,刘永革开发了甲骨文字的编辑软件^[6]。甲骨文字编辑软件对甲骨文字的可视化研究开创了一种资料编辑的新方法,也为国内外的甲骨文学者之间的相互交流提供了一个新的平台。2008年,顾绍通等人对计算机中如何输入甲骨文字也做出重要研究^[7]:首先对甲骨文字形进行深入研究之后发现,甲骨文的字形具有一定的规律。通过对《殷墟甲骨刻辞类纂》中收录的甲骨文字形的拓扑结构进行深入分析之后,整理出了569个甲骨文字的偏旁部首。再把这569个偏旁部首分别配置到标准键盘的26个英文字母上之后,通过拆分取码和现代汉字的拼音方法,使用标准键盘即可输入3673个甲骨文字(包含异体字合文)。甲骨文字输入法的问世对甲骨文字的数字化、展示、有效保护和方便使用,具有非常重要的现实意义。2013年,酆格斐等人通过对原始拓片的预处理,并结合数学形态学方法提出12项指标描述甲骨字特征,这些特征较好地反映了甲骨文字的笔画形态和结构^[8]。2014年,安阳师范学院的高峰对诸多甲骨文语义进行研究后,构建了一个甲骨文领域的语义词典系统^[9]。作者通过不断地对国内外甲骨文语义词典的对比研究后,从其特点出发,对甲骨文语料中的原材料进行了精细的加

工处理,并在甲骨文语义分类的基础上对实词做出了很多属性描述,并建立了计算机的半自动化辅助加工模块,用来服务于甲骨文辅助机器的翻译和考释工作。2014年,熊晶以许多甲骨文学家建立的甲骨文字库为基础,提出一种计算机辅助翻译甲骨文的方法^[10]。2016年,中山大学、北京微软研究院以及瑞士日内瓦大学的研究人员通过图像处理的方式识别甲骨字^[11]。2016年,德国马克斯普朗克研究所和上海大学的研究人员构建了有关动物甲骨字的认知网络^[12]。虽然一些计算机学家在研究甲骨文方面取得了显著的成绩,但是还存在一些问题需要继续深入研究。如现有的算法在研究甲骨文字时,并没有从系统的角度研究,导致使用计算机方法预测未知甲骨文字的语义距离还很远。不仅如此,现有的计算机方法研究甲骨文时,并没有考虑甲骨字之间的联系对场景和语义的影响。

复杂网络是描述复杂系统的一种有效工具,语言系统的网络化抽象为研究语言提供了新的视角和手段^[13]。目前,人们已经构建了汉语的同现网络^[14]、句法网络^[15]、语义场网络^[16]等。语言网络的构建及其特征的分析对研究语言系统背后的形成机制和演化规律具有重要的意义^[13]。在本文中,我们使用甲骨拓片信息构建甲骨文字网络,并对网络的特性进行详细分析。本文的研究结果能为历史学家和网络甲骨学家揭示未知甲骨字的语义提供新的数据和研究思路。

1 甲骨字网络构建

本文以收集的72151片甲骨文拓片为研究对象,进而通过建模构建甲骨字网络。由于甲骨拓片历史久远,拓片的损坏比较严重。因此,在构建网络之前,对其进行相应的处理。第一,如果在一个拓片中,字和字之间有残缺的情况,用省略号代替;第二,除去没有甲骨字的拓片;最后共得到71455片甲骨文拓片、6199个已识和未识甲骨字。

由于甲骨文系统是中国最早的文字系统,语言特性还处于萌芽的状态。因此,它和现有的成熟文字系统有很大的区别:第一,在甲骨文系统中,同一拓片的甲骨字描述了同一个场景(或称语义单元),如,战争、天象、婚娶等,但也有可能不同拓片中的甲骨字描述不同场景。第二,在甲骨文系统中,单音节词较多,而复音节词较少。这也是古文字系统特有的属性。

为了构建甲骨字网络,需要定义甲骨字和甲骨字之间的相似性距离。由于甲骨文系统的同一场景

或语义单元是以拓片为单位,所以,如果在一个拓片中,两个甲骨字之间在 n 阶 Markov 链的条件下同时存在,则认为这两个甲骨字之间应存在一条边。与刘知远构建汉语网络不同^[14],本文中,在两个甲骨字之间定义了相应的权重。对于同一拓片上的两个甲骨字(这两个甲骨字可以是已识或未识),它们分别用 i 和 j 表示,那么这两个字之间的距离为 w_{ij} (见公式 1)。不仅如此,在 n 阶 Markov 链中 n 在现代汉语中经常取值为 2^[14],因为现在的文字系统有大量的词语。而在甲骨文系统,很少有词组的出现。因此,在构建网络时,对于不同拓片, n 值选择为拓片上甲骨字的个数。

$$w_{ij} = 10^{\frac{1}{\text{interval}}} \quad (1)$$

$$\text{interval} = \begin{cases} l_j - l_i & i \text{ 和 } j \text{ 之间无省略号} \\ \frac{\max(\text{length})}{2} + (l_j - l_i) & i \text{ 和 } j \text{ 之间有省略号} \end{cases} \quad (2)$$

其中式(1)中的参数 interval 由式(2)定义。在式(2)中, l_j 和 l_i 分别表示甲骨字 i 和 j 在拓片中的位置且字 j 在字 i 的后面。如果甲骨字 i 和 j 之间有省略号,即它们之间有残缺的甲骨字,那么参数 interval 由两部分组成:一部分是 $\frac{\max(\text{length})}{2}$,参数 length 是 71 455 拓片中甲骨字的个数, $\max(\text{length})$ 表示含有最多甲骨字拓片的长度。

对于所有的 71 455 片甲骨片,我们以 6 199 个甲骨字为基础,构建 6 199 × 6 199 大小的相似性矩阵 \mathbf{M} (即权重网络)。首先,如果两个甲骨字 i 和 j 在同一甲骨片上出现,使用式(1)~(2)进行计算,

并把 w_{ij} 赋予 \mathbf{M}_{ij} 处;其次,如果两个甲骨字在不同的拓片上出现,那么把这两个字在不同拓片上计算的相似性距离在同一个 \mathbf{M} 位置上叠加。图 1 给出了计算相似性矩阵 \mathbf{M} 的一个简单示意图。图中有两个拓片,共有 9 个甲骨字。以这 9 个甲骨字为基础,构建大小为 9 × 9 的相似性矩阵。例如,利用式(1)~(2)计算甲骨字 2 和甲骨字 4 之间的距离,然后把 $w_{2,4}$ 的值放在相似矩阵 $\mathbf{M}_{2,4}$ 的位置上;其次,如果两个甲骨字在不同拓片上同时出现,需要分别计算这两个甲骨字在不同拓片上的相似性距离,然后相加放在相似性矩阵对应的位置上。如图中的甲骨字 5 和 6,分别计算甲骨字 5 和 6 在拓片 1 和 2 上的距离 $w_{5,6}^1$ 和 $w_{5,6}^2$,然后把 $w_{5,6}^1$ 和 $w_{5,6}^2$ 相加放在相似性矩阵 $\mathbf{M}_{5,6}$ 的位置上。最后,依据 71 455 片拓片信息,得到 6 199 个甲骨字之间的相似矩阵,这个矩阵共包含了 160 964 条有权重边。为了保证边信息能真实反映甲骨字之间的拓片信息,我们只保留了权重大于 5 的边,而删去权重小于 5 的边。其原因如下:一,如果边的权重小于 5,从式(1)~(2)的定义可知,此条边连接的两个甲骨字不直接相邻,即不会构成复音节词;二,由于文中构建甲骨字权重网络采用的是叠加方法,即两个甲骨字之间的权重是由这两个甲骨字在 71 455 片拓片中计算的距离之和。如果边的权重小于 5,说明此条边连接的两个甲骨字在 71 455 片拓片中计算的距离之和小于 5,因此这两个甲骨字用来描述同一个场景或语义单元的可能性很小。经过处理,文中构建的甲骨字网络包含 5 474 个已识和未识甲骨字和 75 611 条边。

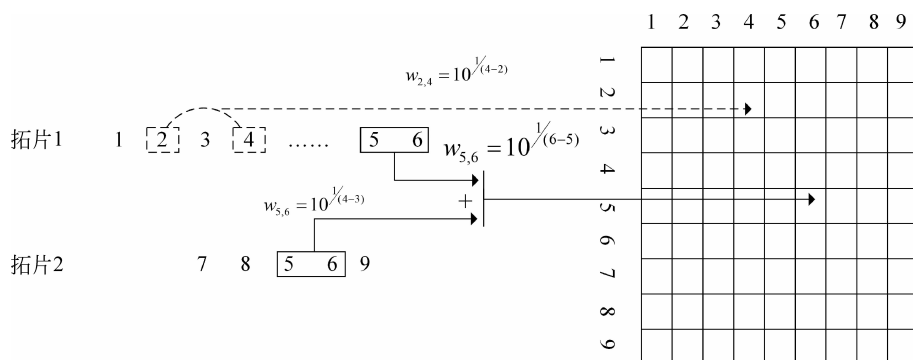


图 1 甲骨字之间距离计算图

本文使用的构建甲骨字网络方法具有三个创新点:一是在构建网络的过程中,充分利用了拓片在甲骨文系统中作为语义单元的信息,即拓片中的甲骨字不论是已识或未识,根据式(1)~(2)都可以构

建它们之间的距离。因此,未识和已识的甲骨字出现在同一个网络中,这种现象为我们依据已识的语义信息破译未识甲骨字提供可能;二是构建网络的方法体现了甲骨文系统中复音节词较少的古文字特

征;三是在构建网络的过程中赋予甲骨字之间相应的权重,利于分析甲骨字之间的同现信息。

2 甲骨字网络特性分析

为了验证甲骨字网络是否具有真实网络的特性,本文对构建网络的度分布、局部连接比率、聚类系数、模块度等特性进行分析。

2.1 度分布

一个节点的度是此节点的邻接节点的个数或者是节点连接边的个数。如果我们把节点度为 k 的数目占网络节点总数目的比例记为 p_k ,那么网络中不同度的统计分布即为度分布^[17]。为了方便推断一个甲骨字在 71 455 片拓片中同时出现的信息,即一个甲骨字和其邻接甲骨字共同描述同一个场景或语义单元信息,我们把构建的权重网络简化为无权重的网络,然后计算度分布。在本文构建的网络中,甲骨字的度表现为甲骨字之间的权重值(或连接边的个数),而甲骨字之间权重分布表现为网络的度分布。图 2 给出了未识和已识甲骨字网络的度分布图,从图中我们可以看到甲骨字网络的度分布符合无标度分布^[18](scale-free distribution),无标度分布意味着网络中大部分节点度的取值较小,但是会有少数节点度的取值非常大。在甲骨字网络中的度分布说明:一方面大部分甲骨字的度值比较小,比如,度值小于 10 的甲骨字(即此甲骨字有 10 个相邻甲骨字)占总甲骨字的比例为 76.6%,而度值小于 17、50 的甲骨字占总甲骨字的比例分别为 82.1%、91.1%。在甲骨文字系统中,较小的度值代表描述同一个场景或语义单元所需的甲骨字也较少。另一方面,有少数的甲骨字有很大的度值,如甲骨字“卜”和“贞”字之间的权重高达 203 756,如果假设这两个甲骨字直接相连,那么“卜”和“贞”在 71 455 个拓片中至少出现 20 375 次。不仅如此,“卜”和其他甲骨字的度值也较大。通过相关的文献我们得知^[1]:在甲骨字系统中,单音节名词占大多数;而动词占少

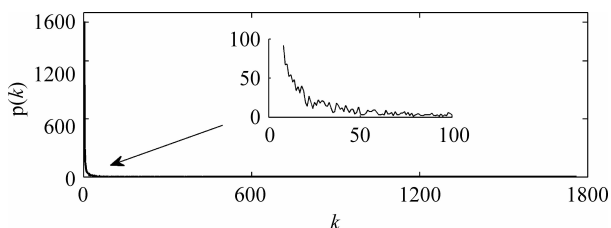


图 2 甲骨字网络的度分布

数,并且在动词中,祭祀动词占多数。“卜”字是常用的动词,经常和其他名词相连使用,因此,“卜”字具有较大的度值。以上分析也说明我们构建的甲骨字网络能充分反映甲骨文系统的语言信息。

2.2 局部连接比率

局部连接比率(local-links-rate,LLR)是一种衡量网络局部特性的指标^[19],如式(3)所示。由于网络中的边信息比节点信息更能反映网络的各种特性。因此,LLR 的定义是基于网络的连接信息,而不是节点信息。对网络中任意一条边 e ,被它连接的两个节点为 n_1 和 n_2 。首先计算这两个节点的共同邻接节点(common node,CN),然后统计共同邻接节点之间存在的边数(local-links)。最后,计算局部连接比率 LLR。图 3 给出了计算 LLR 的实例图。对于图 3 中的一条实线边,连接它的两个节点(中空结点)共享四个节点(实心节点),这四个节点之间存在三条边(点形边)。因此,实线边的 LLR 值为 3/4。如果一个网络的平均 LLR 大于 2,那么这个网络有较强的局部特性^[19]。通过计算,甲骨字网络的 LLR 值高达 26.678 7,说明甲骨字网络具有很强的局部特性,即描述同一个场景(或语义单元)的甲骨字在甲骨字网络中相互之间连接的边较为稠密。

$$LLR = |\text{local-links}| / CN \quad (3)$$

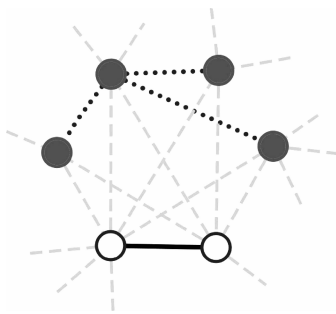


图 3 局部连接比率计算示意图^[19]

2.3 聚类系数

对于无权重网络,聚类系数用来描述一个节点的直接相邻节点之间的边的连接情况^[20],它的定义如式(4)所示。而对于有权重的网络,计算聚类系数比较复杂。在本文中使用无权重网络下的聚类系数,主要因为两种类型的聚类系数的定义对于衡量网络的特性不会带来本质的影响^[21]。图 4 给出计算节点聚类系数的示意图。对于图中的中空节点,它有五个邻接节点(实心节点),这五个节点之间共

有七条边(点形边),那么中空节点的聚类系数为:

$$\frac{7}{5 \times (5-1)/2} = 0.7。$$

一个网络的聚类系数是网络中所有节点聚类系数的平均值。通过计算,甲骨字网络的聚类系数为 0.594 4。较高的聚类系数意味着节点的邻接节点之间存在更程度的交互关系,即这个节点和其邻接节点更稳固地聚集成模块结构^[22]。在甲骨字网络中,较高的聚类系数意味着一个甲骨字和其邻接的甲骨字参与描述同一场景或语义单元的概率较高。

$$CC_i = \frac{e_i}{U_i \times (U_i - 1)/2} \quad (4)$$

U_i 表示节点 i 的邻接节点数, e_i 表示 U_i 个邻接节点之间存在的边数。

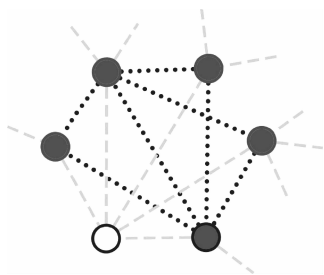


图4 聚类系数计算示意图

2.4 模块度

模块(module,或称社团)结构是复杂网络的一个基本特性,也是复杂网络研究的重点内容。模块是网络的一个子集,它要求模块中节点之间的边连接紧密,而不同模块之间节点的边连接稀疏。图5是一个含有12个节点和三个模块的网络示意图^[23]。模块内的节点具有相似的属性,依据这一特点,模块结构已在很多领域取得了成功的应用。如在蛋白质相互作用网络中,功能相似的蛋白质在网络中往往以模块的形式存在。因此,通过挖掘模块结构可以预测未知蛋白质的功能;在人类社会中,人以类聚是模块结构在社会网络中的真实反映。社会学家可以利用模块结构研究人们的心理行为、兴趣爱好等。通过构建包含已知和未知语义的甲骨字网络,在此基础上分析此网络是否具有模块度特性,进而利用模块内结点的属性,可以预测同一模块内未知甲骨字的语义信息。如在图5中下方的一个含有五个节点的模块中,假如我们已知其中四个甲骨字描述了某种场景信息(如婚娶),那么根据模块结构中节点具有相同属性的特性,可以推测剩余一个未知语义的甲骨字也用来描述婚娶信息。

模块度(modularity)^[24]不仅是一种用来挖掘网

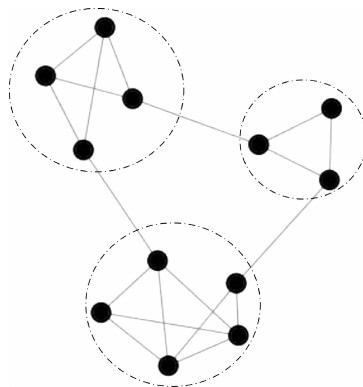


图5 网络模块结构示意图^[23]

络中模块结构的方法,而且是一种用来衡量网络是否具有模块度的标准。虽然基于模块度的方法具有“分辨率限制”(Resolution limit)的问题^[25],但它仍然被广泛用于判断一个网络是否具有模块结构的评价标准。对于有权重的网络,模块度(Q)的定义如式(5)所示。

$$Q = \sum_{v=1}^{n_c} \left[\frac{W_v}{W} - \left(\frac{S_v}{2W} \right)^2 \right] \quad (5)$$

其中, n_c 是网络划分的模块个数, W 是网络中所有边的权值之和, W_v 是模块 v 内部所包含的边的权重和, S_v 是所有与模块 v 内部的点相关联的边的权重和。利用模块度方法对甲骨字网络进行分析,得到的模块度的值为 0.292 1。根据文献^[26]我们得知,如果一个网络的模块度大于等于 0.3,说明这个网络具有很强的模块特性。另外,从局部连接比率和聚类系数可以说明我们构建的甲骨字网络具有较强的局部特性。综上所述,甲骨字网络具有良好的模块结构属性,这种属性为我们通过识别模块结构进而破译未知甲骨字的语义提供了直接数据和理论上的依据。

3 讨论

作为一个新的研究方向,语言网络正在悄然兴起,并取得了一些有意义的结果^[27]。本文第一次使用大规模的拓片信息创新性地构建了甲骨字网络,其主要表现在以下三个方面:一是构建的网络充分捕捉了甲骨文系统的语义单元信息;二是构建网络的方法保留了甲骨文系统单音节词多、复音节词少的特性;三是构建的网络中边权重反映了甲骨字之间的同现关系。

在甲骨字网络之上,深入研究了网络的度分布、局部连接比率、聚类系数和模块结构特性。本文构建的甲骨字网络可为网络甲骨学家和历史学家预测未知甲骨字的场景和语义提供数据支持和直接的研究

理论思路。但是,本文还存在一些不足之处需要我們进一步研究:第一,式(2)中参数 length 选取具有不合理性: length 表示的是同一拓片上两个甲骨字之间残缺字的个数,因此, length 的值不能设置过大。如何利用甲骨文字系统的意义选取 length 的值是我们需要进一步研究的问题。第二,由于破译未知甲骨字的语义是甲骨学研究的重要内容,因此另外一个不足之处是本文既没有利用构建的网络破译已识甲骨字的一些偏旁部首的可能含义,也没有依据已识甲骨字推断未识甲骨字或其偏旁部首的可能含义。如何设计高效算法推断未识甲骨字或其偏旁部首的可能含义是我们下一步重点研究的内容。

参考文献

- [1] 江铭虎. 自然语言处理[M]. 北京: 高等教育出版社, 2006: 388-389.
- [2] 胡厚宣. 八十五年来甲骨文材料之再统计[J]. 史学月刊, 1984(5): 15-22.
- [3] 李宗琨. 甲骨文字编[M]. 北京: 中华书局, 2012.
- [4] 夏大兆. 甲骨文字用研究[D]. 合肥: 安徽大学博士学位论文, 2014.
- [5] 熊晶, 高峰, 吴琴霞. 甲骨文大规模基础数据的语义挖掘研究[J]. 现代图书情报技术, 2015, 2: 7-14.
- [6] 刘永革, 栗青生. 可视化甲骨文字编辑器设计与实现[J]. 安阳师范学院学报, 2007, 5: 35-38.
- [7] 顾绍通, 马小虎, 杨亦鸣. 基于字形拓扑结构的甲骨文输入编码研究[J]. 中文信息学报, 2008, 22(4): 123-128.
- [8] 酆格斐, 顾绍通, 杨亦鸣. 基于数学形态字的甲骨拓片字形特征提取方法[J]. 中文信息学报, 2013, 27(2): 79-85.
- [9] 高峰, 田喜平, 刘永革. 甲骨文领域语义词典的构建研究[J]. 安阳师范学院学报, 2014, 5: 43-47.
- [10] 熊晶, 高峰, 吴琴霞. 甲骨文计算机辅助翻译技术研究[J]. 科学技术与工程, 2014, 14(2): 1671-1815.
- [11] Guo J, Wang C, Roman-Rangel E, et al. Building hierarchical representations for oracle character and sketch recognition[J]. IEEE Transactions on Image Processing, 2016, 25(1): 104-118.
- [12] Dress A, Grünewald S, Zeng Z. A cognitive network for oracle bone characters related to animals[J]. International Journal of Modern Physics, 2016, 30(4): 1630001.
- [13] 赵恂怡, 刘海涛. 语言网络研究的数学模型[J]. 中文信息学报, 2015, 29(6): 46-53.
- [14] 刘知远, 孙茂松. 汉语同现网络的小世界效应和无标度特性[J]. 中文信息学报, 2007, 21(6): 52-58.
- [15] 赵恂怡, 刘海涛. 语言同现网、句法网、语义网的构建与比较[J]. 中文信息学报, 2014, 28(5): 24-31, 65.
- [16] 杨华, 姬东鸿, 萧国政. 汉语语义场网络中的无标度分布现象[J]. 中文信息学报, 2015, 29(3): 34-43.
- [17] Barrat A, Weigt M. On the properties of small-world network models[J]. The European Physical Journal B-Condensed Matter and Complex Systems, 2000, 13(3): 547-560.
- [18] Barabási A, Albert R. Emergence of scaling in random networks[J]. Science, 1999, 286(5439): 509-512.
- [19] Jiao Q, Huang Y, Shen H. Community mining with new node similarity by incorporating both global and local topological knowledge in a constrained random walk[J]. Physica A: Statistical Mechanics and its Applications, 2015, 424(2015): 363-371.
- [20] Newman M. The structure and function of complex networks[J]. Siam Review, 2003, 45(2): 167-256.
- [21] 汪小帆, 李翔, 陈关荣. 网络科学导论[M]. 北京: 高等教育出版社, 2012: 96-103.
- [22] 林枫, 刘云, 江钟立. 汉字网络的历时性模式探析[J]. 复杂系统和复杂性科学, 2012, 9(3): 50-61.
- [23] Fortunato S. Community detection in graphs[J]. Physics Reports, 2010, 486(3): 75-174.
- [24] Newman M, Girvan M. Finding and evaluating community structure in networks[J]. Physical Review E, 2004, 69(2): 026113.
- [25] Fortunato S, Barthelemy M. Resolution limit in community detection[J]. Proceedings of the National Academy of Sciences, 2007, 104(1): 36-41.
- [26] Clauset A, Newman M, Moore C. Finding community structure in very large networks[J]. Physical Review E, 2004, 70(6): 066111.
- [27] 韩普, 王东波, 路高飞, 等. 语言网络研究进展[J]. 中文信息学报, 2014, 28(1): 9-18.



焦清局(1983—), 通信作者, 博士, 讲师, 主要研究领域为复杂网络理论及其应用、甲骨文信息处理、模式识别等。

E-mail: qjjiao@aynu.edu.cn



金园园(1982—), 硕士, 助教, 主要研究领域为甲骨文信息处理。

E-mail: jinyuan@aynu.edu.cn



高峰(1977—), 硕士, 讲师, 主要研究领域为自然语言与机器学习、甲骨文信息处理。

E-mail: gaof97@163.com