

文章编号: 1004-4353(2015)01-0074-05

基于加权网络的教科书知识结构分析 ——以韩国高一《科学》教材为例

郭优¹, 卢海洋¹, 曹铷¹, 崔雪梅^{2*}

(1. 延边大学理学院 物理系; 2. 延边大学师范学院: 吉林 延吉 133002)

摘要: 通过构建加权网络和骨架网络的方法对韩国 2009 年课程改革后出版的融合型高一《科学》教科书和第 7 次课程改革后出版的模块型高一《科学》教科书进行比较分析, 结果显示模块型教科书的骨架网络呈多核心特征, 而融合型教科书骨架网络只有一个核心术语, 其他术语以同心圆形式扩散. 此结果表明, 韩国 2009 年课程改革后出版的融合型高一《科学》教科书有效地融合了科学知识, 实现了课程改革的理念.

关键词: 加权网络; 知识结构; 骨架网络

中图分类号: N94

文献标识码: A

Analysis of textbook knowledge structure based on weighted network: taking the textbook of Science in Senior One of Korea for example

GUO You¹, LU Haiyang¹, CAO Ru¹, CUI Xuemei^{2*}

(1. Department of Physics, College of Science, Yanbian University;

2. Normal College, Yanbian University; Yanji 133002, China)

Abstract: The integrative type textbook of Science for Senior One published after curriculum reform of Korea in 2009 and the modular type textbook of Science for Senior One published after the 7th curriculum reform were comparatively analyzed by constructing the weighted network and the skeleton network. The results show that the skeleton network of modular type textbook appears the property of multiple cores but the skeleton network of integrative type textbook has sole core term with other terms spread around it in concentric circles. These findings suggest that the integrative textbook of Science for Senior One published after curriculum reform of Korea in 2009 effectively fused scientific knowledge and realized the idea of curriculum reform.

Key words: weighted network; knowledge structure; skeleton network

利用复杂网络分析方法研究复杂系统的特性在很多领域被使用, 如神经网络的研究^[1]、万维网分布的研究^[2]、好莱坞演员网络研究^[3]、蛋白质相互作用网络研究^[4]、航空客运网络研究^[5]和人类疾病传播网络研究^[6]等等. 初期网络分析的焦点聚焦在连接度(degree)的分布特性上, 随着研究的深入, 网络分析的内容逐渐趋向于层次结构(hierarchical structure)、模块结构(modular struc-

ture)和社团结构(community structure)等方面^[7-9], 这些研究主要是通过分析网络的信息后寻找共同体结构或基于信息建构加权网络得出骨架结构(minimal spanning tree). 目前, 对教科书所呈现的知识网络特性的研究较少, 本文在先行研究^[10]中基于二元网络对英文《普通物理》教科书和《牛津英语词典》进行了分析, 结果显示该教科书的物理知识网络的度分布满足幂律分布, 且具有层次

性模块结构.

韩国在 2009 年课程改革中为了体现科学知识的融合,将高一《科学》课程修订为《融合科学》,课程目标定为使学生从整体的视角理解科学的意义和作用^[11].为了验证韩国 2009 年课程改革的理念是否体现在教科书中,本文利用加权网络的分析方法,对韩国 2009 年课程改革后出版的教科书和之前第 7 次课程改革后出版的教科书进行比较分析.

1 研究资料及方法

1.1 研究对象

研究对象为韩国第 7 次课程改革后出版(2001~2002 年)的两本模块型高一《科学》教科书(天才教育出版社^[12]、金星出版社^[13])和 2009 年课程改革后出版(2011 年)的两本融合型高一《科学》教科书(天才教育出版社^[14]、金星出版社^[15]),其中:两本模块型教科书中将科学知识分为“能量”、“物质”、“生命”、“地球”和“环境”等 5 个领域,对应于物理、化学、生物、地球科学等 4 门课程;两本融合型教科书将科学知识分为“宇宙和生命”、“科学和文明”两个一级主题,一级主题又分为“宇宙的起源与进化”、“太阳系和地球”、“生命的进化”、“信息通信和新材料”、“人类的健康和科学技术”、“能量和环境”等 6 个二级主题,以使学生从整体的视角去理解科学知识和科学探究的价值.

模块型教科书中的物理术语由韩国物理学会发布的物理术语组成,化学、生物、地球科学、环境等领域的术语由教科书索引中的术语加上常用术语组成,一共为 1 048 个科学术语.融合型教科书的术语由教科书索引中的术语加上常用科学术语组成,一共为 1 299 个科学术语.模块型和融合型教科书中 有 337 个相同术语,将两种教科书的科学术语组合后最终得到 2 010 个科学术语.4 本教科书中出现的科学术语个数如表 1 所示.表 1 中 N 表示术语个数, $\langle K \rangle$ 表示平均度.

表 1 4 本教科书中出现的科学术语个数及平均度

课程改革时期	出版单位	N	$\langle K \rangle$
第 7 次	天才教育出版社	1 095	10.24
	金星出版社	936	8.82
2009 年	天才教育出版社	992	20.10
	金星出版社	1 033	23.02

1.2 加权网络分析方法

在本文建构的加权网络中,节点是科学术语,连边的权值为两个术语在整本教科书的同一个句子中同时出现的次数,即两个术语的配对出现的频数 s (两个术语在同一个句子中同时出现一次以上均视为一次).两个术语在同一个句子中出现,表明两个术语间有联系,即两个术语在同一个句子中出现次数能够表明两个术语关系的密切程度.

利用不同加权网络连边权值代表不同重要度的特点,建构没有环形结构的骨架网络或树状结构网络,将有 N 个节点的网络用 $N - 1$ 个连线连接^[16].将网络的全部连边由权值最大的连边开始按照权值逐渐变小的顺序进行连接,连接途中如果某个连边形成环形结构,那么不连接该连边.用这种方法,可以利用 $N - 1$ 个连边连接所有节点,即可以建构无环的骨架网络,虽然该方法在骨架网络中会丢失一部分数据,但由于没有环形结构,可将复杂的网络结构特征呈现在二维平面上.

骨架网络中需要分析的第 1 个特性是从网络中心节点到其他各节点的距离 d_i^m ,网络的中心节点为到其他各节点的距离的平均值最小的节点;第 2 个特性为节点连接的同质性,它代表在相连于节点 i 的 k_i 个节点中,与节点 i 属于相同主题的节点所占比例.连接的同质性系数 h 定义为

$$h_i = k_i^s / k_i, \tag{1}$$

其中 k_i 为节点 i 的度, k_i^s 为与节点 i 相连的 k_i 个节点中与节点 i 属于相同主题的节点个数.

2 研究结果

2.1 《科学》教科书知识网络的统计特性

模块型和融合型教科书中科学术语出现的频数 f 和配对出现频数 s 的分布如图 1 所示.图中两本模块型教科书分别用 CJ1(天才教育出版社)和 KS1(金星出版社)来标注,两本融合型教科书分

别用 CJ2(天才教育出版社)和 KS2(金星出版社)来标注。

从图 1(a) 的术语出现频数 f 分布可以看出, 融合型教科书的最大 f 值比模块型教科书的最大 f 值大。将 f 分布图用幂律指数截止函数 $P(f) \sim f^{-\alpha} e^{-f/\tau}$ 拟合, 该函数由幂函数和指数函数构成, 其中 α 代表幂函数减少的快慢, τ 为呈现指数函数特性的时间常数, τ 越大呈现指数函数的区域越小, 反之亦然。当 τ 足够大时, 可以认为函数的特性接近于幂函数。在 f 分布图中: 模块型教科书的 $\alpha=2.0, \tau=53 \sim 85$; 融合型教科书的 $\alpha=1.4, \tau=96 \sim 133$ 。由此可知, 融合型教科书的 f 在更大区域内呈现幂函数特性, 并且减少得相对缓慢, 其分布偏重于 f 值大的术语。这是因为, 在模块型教科

书中为了分别组成 5 个领域的知识, 术语出现频数高于融合型教科书, 并且术语出现的次数相差不大; 而在融合型教科书中, 重要的术语在整个教科书中被频繁使用, 因此这类术语的出现频数相对较大。

从图 1(b) 可以看出, 术语配对出现频数 s 分布的特性与术语出现频数 f 分布的特性相似。用函数 $P(s) \sim s^{-\beta} e^{-s/\tau}$ 拟合的结果如图 1 所示, 其中模块型教科书的 $\beta=3.7 \sim 3.8, \tau=19 \sim 33$; 融合型教科书的 $\beta=2.3, \tau=19 \sim 26$ 。虽然模块型教科书和融合型教科书呈现幂函数的区域相似, 但融合型教科书的 s 分布减少得相对缓慢, 这表明融合型教科书中 s 值大的术语对更多。

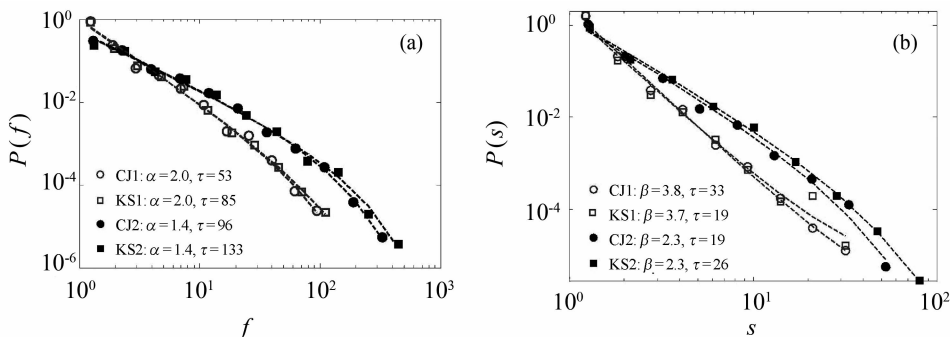


图 1 模块型和融合型教科书的术语出现频数 f 和术语配对出现频数 s 的分布 (双对数坐标)

2.2 《科学》教科书知识网络的特性

为了更加准确地分析《科学》教科书知识结构的特性, 将 s 作为连边的权值建构科学知识的加权网络。为了在二维平面上了解知识结构的特点, 基于加权网络建构骨架网络, 并借助 Pajek^[17] 软件画出骨架网络, 如图 2 和图 3 所示。图 2 和图 3 中节点的颜色深浅代表节点所属的主题。一个术语在哪个主题中出现的次数最多, 那么此术语就属于该主题。

从图 2 和图 3 可看出, 模块型和融合型教科书的骨架网络有结构上的区别, 其中模块型教科书骨架网络的核心骨架结构沿着中心轴线性扩散, 而融合型教科书骨架网络有单一的中心, 整体结构围绕这个中心以同心圆的形式扩散。为了定量描述这种特性, 找出网络的中心后求出网络中心到各节点距离 d 的分布, 分布情况如图 4 所

示。从图 4 可知, 融合型教科书的 d 值整体上比模块型教科书的小。各教科书所有术语的 d_i^m 平均值 $\langle d \rangle$ 分别为 7.02(CJ1)、6.49(KS1)、5.12(CJ2)、4.95(KS2), 模块型教科书的 $\langle d \rangle$ 大于融合型教科书。另外, 在模块型教科书的距离分布图(图 4)中呈现相当于各领域中心节点的多个峰值, 但在融合型教科书中只有一个峰值。这是由于融合型教科书的骨架网络呈现以同心圆形式扩散的特征, 而模块型教科书分为 5 个领域构成模块化结构, 其线性结构特征来自加权网络的模块性特征。

在图 2 和图 3 所示的骨架网络中, 连接相同主题节点的连边数 L_s 占总连边数 ($L=N-1$) 的比例 (L_s/L) 分别为 0.86(CJ1)、0.85(KS1)、0.69(CJ2)、0.67(KS2), 这说明在模块型教科书的骨架网络中, 连接相同主题节点的连边数比融合型教科书多。

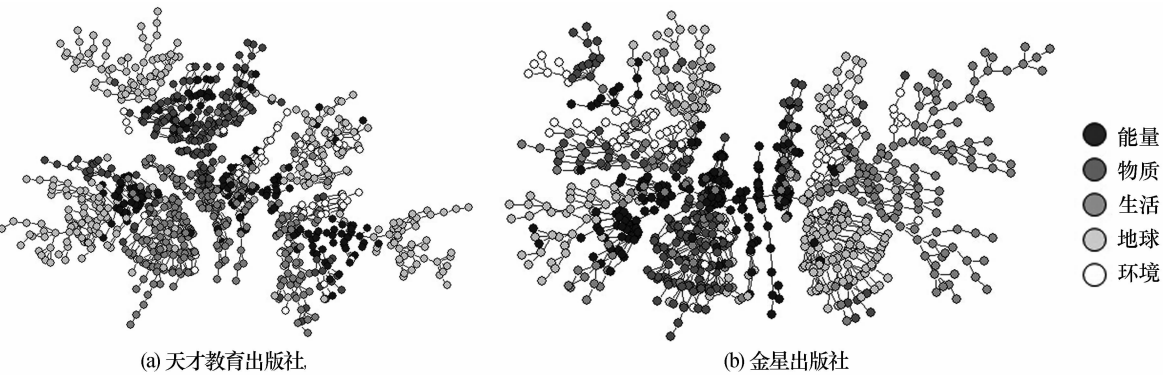


图 2 模块型教科书的骨架网络

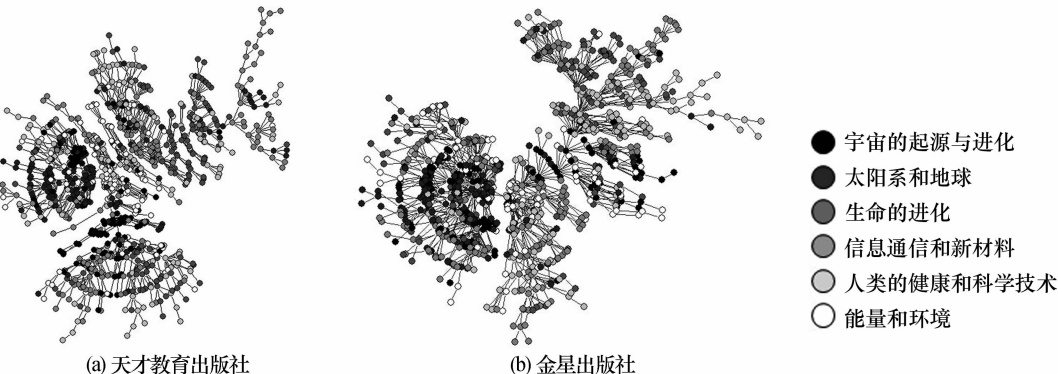


图 3 融合型教科书的骨架网络

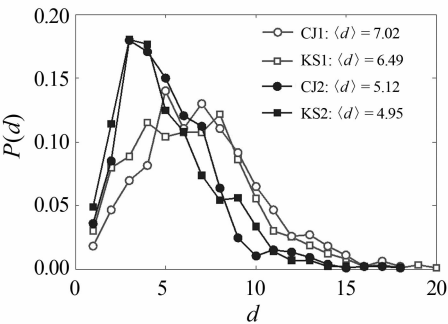


图 4 网络中心到各节点距离的分布图

图 5 为骨架网络连接的同质性系数 h 对节点连接度 k 的分布图. 在 $k < 10$ 的区域内, 模块型教科书的 $h(k)$ 值比融合型教科书的大, 度小的节点相当于树梢的术语, 在模块型教科书中树梢术语属于相同主题的较多, 但在融合型教科书中并未表现出这种情况. 在 $k > 10$ 的区域内, 模块型和融合型教科书的 $h(k)$ 值无明显差异, 表明相当于树干的节点是核心节点, 不仅连接的节点较多, 而且与不同主题的节点相连, 导致 $h(k)$ 值变小. 各教科书所有术语的 h 值的平均值 $\langle h \rangle$ 分别为

0.81(CJ1)、0.78(KS1)、0.65(CJ2)、0.63(KS2).

从图 5 中还可知, 融合型教科书中节点度的最大值比模块型的大, 表 1 中给出的节点的平均度定量地说明了这一点. 这是由于在融合型教科书中术语是在更大的范围内与其他术语相联系, 而在模块型教科书中术语只在模块内相联系.

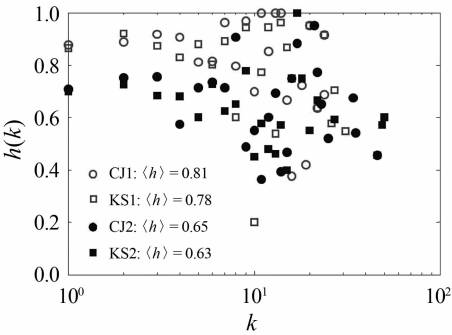


图 5 连接的同质性系数的分布图

3 结论

本文研究结果表明, 两种教科书的骨架网络表现出不同的特性: ①在模块型教科书的骨架网

络中,各领域核心术语呈线性排列在骨架结构的轴上,表现出多核心的特点;在融合型教科书骨架网络中,只有一个核心术语,其他术语以同心圆形式扩散,这种结构上的差异导致融合型教科书的骨架网络中节点间距离的平均值比模块型教科书的小。②在模块型教科书中,属于相同主题的术语间连边较多,而融合型教科书中较少,这是由于模块型教科书的骨架网络中连线的同质性系数的平均值大于融合型教科书。以上结果说明,韩国 2009 年课程改革的理念在高一《科学》教科书中得到了很好的体现,利用本文方法可针对我国中学《综合科学》教科书的知识整合的情况进行分析,为教材编写提供依据。

参考文献:

- [1] Watts D J, Strogatz S H. Collective dynamics of 'small-world' networks[J]. *Nature*, 1998, 393: 440-442.
- [2] Barabási A L, Albert R, Jeong H, et al. Power-law distribution of the world wide web[J]. *Science*, 2000, 287(5461): 2115.
- [3] Barabási A L, Albert R. Emergence of scaling in random networks[J]. *Science*, 1999, 286: 509-512.
- [4] Jeong H, Mason S P, Barabási A L, et al. Lethality and centrality in protein networks[J]. *Nature*, 2001, 411: 41-42.
- [5] Amaral L A N, Scala A, Barthélemy M, et al. Classes of small-world networks[J]. *Proc Natl Acad Sci USA*, 2000, 97: 11149-11152.
- [6] Goh K I, Cusick M E, Valle D, et al. The human disease network[J]. *Proc Natl Acad Sci USA*, 2007, 104: 8685-8690.
- [7] Newman M E J, Girvan M. Finding and evaluating community structure in networks[J]. *Phys Rev E*, 2004, 69: 026113.
- [8] Newman M E J. Fast algorithm for detecting community structure in networks[J]. *Phys Rev E*, 2004, 69: 066133.
- [9] Kim D H, Noh J D, Jeong H W. Scale-free trees: the skeletons of complex networks[J]. *Phys Rev E*, 2004, 70: 046126.
- [10] 崔雪梅, 李凤月, Han SeungKee. 物理知识网络的特性分析[J]. *复杂系统与复杂性科学*, 2013, 10(2): 30-36.
- [11] Korea Foundation for the Advancement of Science & Creativity. Development of a National Science Curriculum(2010-14)[M]. Seoul: Korea Foundation for the Advancement of Science & Creativity, 2010.
- [12] Cha D W, Kim H S, Lee M S, et al. High School Science[M]. Seoul: Chunjae Education, 2002.
- [13] Lee M W, Jeon S W, Choi B S, et al. High School Science[M]. Seoul: Kumsung Publishing, 2002.
- [14] Cho H S, Kang D H, Kang T W, et al. High School Science[M]. Seoul: Chunjae Education, 2011.
- [15] Ahn T I, Ahn J S, Hahn I S, et al. High School Science[M]. Seoul: Kumsung Publishing, 2011.
- [16] Kruskal J B. On the shortest spanning subtree and the traveling salesman problem[J]. *Proc Am Math Soc*, 1956, 7: 48-50.
- [17] Batagelj V, Mrvar A. Pajek: program for large networks analysis [DB/OL]. [2013-11-20]. <http://vlado.fmf.uni-lj.si/pub/networks/pajek/>.