# 学术博客推荐网络的 h 度实证
## ——以科学网博客为例

谭　旻[1]　许　鑫[2]
[1](浙江大学信息资源管理系　杭州 310027)
[2](华东师范大学商学院信息学系　上海 200241)

**摘要:** 【**目的**】研究 h 度这一新型带权信息网络分析框架在学术博客推荐网络中的特性。【**方法**】以科学网博客 2013 年数据为基础，构建学术博客推荐网络，计算 h 度等相关参数，并辅以信息可视化进行讨论。【**结果**】学术博客推荐网络中，高 h 度节点的产生可由信息源(博主)所持有的学术知识内涵导致，也可因信息源提供话题的兴趣外延引发; h 度($h_A$)与节点带权度($N_A$)之间存在形如 $N_A=b\times h_A^2$ 的近似函数关系；高 h 度节点通常成为网络中心部分的局部子群组织者。【**局限**】h 度并非完美指标，后续研究可试用改进型 h 度进行拓展。【**结论**】h 度可作为学术博客推荐网络分析的测度工具之一，对于此类社群的管理可从高 h 度节点入手。

**关键词:** 社会性媒体　学术博客　信息推荐关系　社会网络分析　信息网络　h 度
**分类号:** G203

## 1　引　言

推荐是学术博客中的常见信息行为。信息推荐关系反映了推荐主体 A 对于被推荐主体 B 所发布的信息内容于某种意义上的认可或赞同，在诸多信息交互行为中，它表征了行为主体双方产生的一类相对直接的信息关联。分析信息关联可用网络分析方法，该方法主要以图论方法作为数理基础，从量化分析的角度研究各类网络的结构及优化问题，在信息研究领域有广泛的应用空间[1-2]。近年来，信息领域还出现了关于信息类网络的特有分析框架，例如 Zhao 等将社会网络分析方法与 h 指数相结合，提出测度权重网络的基本指标——h 度(h-degree)[3]。这一测度关注了网络中的重要节点，对联系数量和联系强度做出平衡，具有符合幂律特性、易于计算等特点，在国际上也有一系列的应用和扩展[4-8]。

测度方法是对学术博客及其相关网络进行研究的基础。h 度作为一种新兴的信息网络测度参数，虽已有许多应用和实证，但大多局限于文献类信息网络，而在互联网类信息网络中的特点与有效性尚停留在假设阶段。在这样的背景下，本文尝试引入 h 度分析学术博客推荐网络。

## 2　研究背景

互联网为信息的传递提供了更快捷的渠道，而信息能在瞬息万变的网络海洋中得到传播与推广，重要的途径之一即为信息推荐。例如，电子商务领域中，可通过分析用户信息与行为，给出恰当的商品信息推荐，从而引导消费[9]。日常生活中，周围人群的推荐往往是个体接受信息的一个重要来源。而借助网络的优势，具有相同或相近兴趣的用户更容易组成群体，在获取所需信息的同时，还存在着强烈地表达观点、与其他用户分享经验以及实现自我价值的需求[10]。在博客平台中，对于博文或博客的推荐、分享及点赞等功能，也可在广义上理解为互联网信息推荐的具体形式。学术博客中的推荐关系既表征了博主之间的人际交流，也抽象地反映了信息实体之间的信息互动，从而可构建

通讯作者：许鑫，ORCID：0000-0001-7020-3135，E-mail：xxu@infor.ecnu.edu.cn。

为网络形式进行研究。

当然，作为一种网络大类，信息网络与其他网络大类(如社会网络和生物网络等)自然存在较根本的差别，从而在整体结构上、局部节点与联系的处理上，移植于社会网络分析的传统方法未必总是最佳选择。在这样的背景下，信息领域学者也开始尝试研究信息网络分析专用工具。2011 年，鉴于信息网络多为联系含有权重信息的带权网络这一性质，Zhao 等结合 h 指数[11]的独有算法，提出 h 度及多种相关测度作为带权信息网络的统计研究方法[3]。带权网络中节点的 h 度指该节点与其他 h 个节点保持不低于 h 的联系强度。这一指标的特点包括侧重测量网络中的联系强度结构、可对网络中的联系数量和联系强度进行平衡、符合网络中的幂律特性、含义简洁且易于计算。此后，h 度及相关参数在国际上形成了系列研究。Schubert 在 h 度的框架基础上提出了测度合作网络的合作能力指数(Partnership Ability Index) [4]，而 Rousseau 则指出合作能力指数实际上是 h 度的一种特例[5]，Zhao 等也将 h 度扩展到有向网络[12]，其他一些扩展和改进也随之展开[6-9]。h 度的测度核心是联系权重，而博客推荐网络中关注的要点也是推荐关系及推荐次数，推荐关系为网络联系，推荐次数即是联系权重。由此可见，h 度符合博客推荐网络中的应用场景，有望成为研究博客中推荐行为的一类观测指标。本文将对此进行探索性实验。

## 3 博客推荐网络中的 h 度

博客推荐网络是由博客社区中的信息推荐行为抽象出的分析框架，可描述为：以博主(或博客)为网络节点、以博主(或博客)之间的推荐行为关系为网络联系的节点和联系集合，此集合即构成博客推荐网络这一信息网络的现实类型。该类网络中，博主(或博客)为网络内的信息源实体，而推荐行为构成表征信息在不同实体之间流动或传播的指示性标志。从而，不同信息实体之间的推荐次数代表了联系的权重与信息的流量，成为关键性分析要素，而整个博客推荐网络，事实上呈现为典型的带权信息网络。

h 度已在多种文献类带权信息网络中得到测试和应用。作为分析带权信息网络可能的有益工具，本文将其推广到博客推荐网络的定量研究中。为了使概念明晰，将分析对象确定为博客社区中的博主。博客推荐网络中的 h 度，即为博主在推荐行为中汇集的受关注度指标，根据 h 度的原始定义[3]，博主 h 度定义如下：

博客推荐网络中，博主 A 的 h 度 $h_A$ 指有 $h_A$ 位其他博主各自在一定统计期内推荐了博主 A 所发布的信息内容不少于 $h_A$ 次。

由此定义可见，理论上，博客推荐网络中 h 度测度的是博主(测评对象节点)被其他博主(邻近节点)所推荐的强度与频度的综合，同时与是否广泛地被社群中的其他博主所推荐也有直接关系。一位博主的 h 度高，要求其不仅要有忠实的“粉丝”给予长期密切的关注和频繁的推荐，同时也要在信息议题上有较多的受众。从网络结构的角度，高 h 度的博主有成为博客社群中意见领袖的明星潜质。实践中，一位博主长期保持较高的 h 度，说明其博客的信息内容对于较多的其他博主有较强的持续影响，故无论是博客社区的运营者、网络舆情的挖掘者还是该博客社区的新用户，都可将其作为重点关注的对象，从而更有效地管理、研究和融入。就学术博客社区而言，高 h 度的博主还说明其学术议题和思想能受到更多不同学科学者的青睐，并引发更多不同或交叉领域的思考和流传。

显然，与文献类信息网络相比，h 度在互联网类信息网络中的表征意义略有不同。在文献的共引网络中，h 度主要反映核心文献之间的关联及在研究主题中的价值。在作者合作网络中，h 度侧重挖掘相关主题研究中的引领性作者。在学科引文网络中，h 度重点体现学科在科学知识交流体系中的贡献与位置。而在博客推荐网络中，h 度测度对象是信息源集合——博主，计算时也涉及到信息流动的方向和流量——推荐关系与频次，故而 h 度在此情景中主要表征博主在由推荐行为所交织而成的信息互动中的影响力。广受推荐的博主，不仅影响了众多社区成员，对于各成员的影响深度也较强，从而对网络具有一定的组织与控制能力。

## 4 实例研究

### 4.1 数据源与处理

以科学网(ScienceNet.cn)的博客社区作为实例[12-14]。采用网络爬虫程序于 2013 年 1 月–2014 年 2 月期间连续采集科学网博客的整体数据。其中博主信息和推荐信息作为本文核心数据，用于构建学术博客推荐网

络。数据处理流程如下:

(1) 从下载数据中提取博主的信息和博主之间的推荐行为;

(2) 将博主名作为节点名，并列为推荐矩阵的行列名;

(3) 统计节点之间的推荐行为和推荐次数，作为矩阵的对应元素值，构建出完整的推荐矩阵;

(4) 计算相关网络和博主节点的 h 度等指标。其中 h 度的计算过程如表 1 所示:

表 1　学术博客推荐网络中的博主 A 的 h 度计算

| 序号 | 博主 A 的相连节点 | 推荐博主 A 的次数 | 判定 |
|---|---|---|---|
| 1 | 博主 B | 18 | 1 < 18 |
| 2 | 博主 C | 12 | 2 < 12 |
| 3 | 博主 D | 9 | 3 < 9 |
| 4 | 博主 E | 8 | 4 < 8 |
| 5 | 博主 F | 7 | 5 < 7 |
| 6 | 博主 G | 5 | 6 > 5 |
| … | … | … | … |
| n | … | k | n > k |

(注: 先按推荐 A 的推荐次数降序排列, $h_A$=5。)

由于科学网博客不能自我推荐，故结果的推荐矩阵中对角线元素全部为 0。因为其他的推荐与自我推荐在表征意义上存在区别，今后的推荐网络研究中也建议将推荐矩阵对角线置为 0。为避免网络规模过大并掩盖核心信息，只保留被推荐次数不低于 5 次的博主及相连节点作为网络节点，最终矩阵含 595 个博主节点及相互之间 159 567 条推荐联系，网络中联系总强度为 354 024。

### 4.2　结果分析与讨论

(1) 网络主要 h 度结果分析

根据以上数据可算得，2013 年间案例学术博客推荐网络中每条推荐联系的平均强度为 2.22 次。即平均而言，有推荐关系的博主之间，每半年才推荐一次，频度不高。由此可见，在学术博客这类以学者为主要构成的网络社群中，推荐行为并不随意。而学者在推荐信息时，也较为慎重。当然，核心节点的出现几乎是人与信息所形成群落中的必然现象。例如，博主曹聪在一年中被推荐了 7 853 次，平均每天被推荐 21.52 次，网络中平均每位博主推荐了 13.20 次。在科学网博客这一学术性社区中，可谓备受瞩目。曹聪的 h 度也是网络中最高的，如表 2 所示:

表 2　科学网学术博客推荐网络中 h 度不低于 20 的博主(节点)

| 序号 | 博主名 | h 度 | 序号 | 博主名 | h 度 |
|---|---|---|---|---|---|
| 1 | 曹　聪 | 43 | 17 | 戴德昌 | 23 |
| 2 | 李学宽 | 40 | 18 | 李土荣 | 23 |
| 3 | 杨正瓴 | 40 | 19 | 张玉秀 | 23 |
| 4 | 陆俊茜 | 38 | 20 | 苏德辰 | 22 |
| 5 | 武夷山 | 37 | 21 | 陈楷翰 | 22 |
| 6 | 赵美娣 | 32 | 22 | 刘　洋 | 22 |
| 7 | 徐　晓 | 31 | 23 | 刘全慧 | 21 |
| 8 | 张忆文 | 30 | 24 | 刘艳红 | 21 |
| 9 | 陈小润 | 30 | 25 | 郑小康 | 21 |
| 10 | 吕　喆 | 29 | 26 | 李天成 | 21 |
| 11 | 李伟钢 | 27 | 27 | 许培扬 | 20 |
| 12 | 钟　炳 | 26 | 28 | 王春艳 | 20 |
| 13 | 庄世宇 | 26 | 29 | 刘　立 | 20 |
| 14 | 朱晓刚 | 25 | 30 | 陈冬生 | 20 |
| 15 | 蔡庆华 | 25 | 31 | 陈　安 | 20 |
| 16 | 陈湘明 | 24 | 32 | 鲍海飞 | 20 |

由表 2 可见，博客推荐网络中 h 度较高的博主多为科学网博客中的知名博主。本研究中的 h 度从信息推荐行为的角度，较好地体现了网络中受到广泛关注且被推荐强度较高的重要学术信息源。如曹聪的 h 度为 43，说明 2013 年中有 43 位其他博主推荐了他的博文至少 43 次。实际上，案例网络中有 335 位博主都推荐了不少于 5 次，占所有博主节点的 56%，体现出他在网络中的广泛影响力。曹聪于美国哥伦比亚大学博士毕业后，在世界著名大学和研究机构长期从事科学文化与科技政策研究，近期在 *Science* 等著名学术期刊发表了多篇关于中国科技政策与体制的研究论文，并与国内学者有很多交流与合作。故其博客发布的博文信息常与国内科学文化现状和科技体制改革等学术界热点议题有关，能引起不同学科博主的广泛讨论，并产生共鸣，从而获得较多的认同意义上的推荐。虽然科学网是以科学信息讨论为主的学术博客，但博客这一社会性媒体的载体性质，决定了这一类信息载体在内容上与期刊、图书等传统纸质信息载体存在重要区别。特别是在具体信息内容上，自然地出现了多样性。如表 2 中 h 度排名第 2 的中国科学院的李学宽作为化学领域的研究者，在讨论化学之余，他也在博客中发

布了大量关于摄影方面的信息，从而吸引了大批对摄影有兴趣的学者参与交流，引发了较多基于兴趣意义上的推荐，并获得较高的 h 度。这两个例子说明，学术博客中，节点的高 h 度不仅可以由传统的学术观点认同所导致，也可由其他兴趣热点所引发。

缘于对信息的敏感和分享信息的专业训练，图书情报领域学者常是信息交流新工具的先行体验者。表 2 中的高 h 度学者也有多位属于图书情报领域。例如，中国科学技术信息研究所的武夷山，除了在博客中讨论图书情报学和文献计量学相关知识，还经常发布关于社会、文化、管理和语言学等众多领域的学术讨论，吸引了大量各学科学者参与。浙江大学的赵美娣常及时跟进重要的社会热点，并能以学者的视角进行思辨，引发热烈的讨论。中国医学科学院的许培扬则充分体现了信息学者的信息素养，分享了大量关于图书情报和医学等领域的信息，成为网络中突出的信息源。这些案例再次说明，在学术博客的推荐网络中，无论是进行学术讨论，还是进行社会热点与兴趣分享，都可能产生较高的 h 度。h 度也可一定程度作为学术影响力之外的多样化社会影响力和关注度参量。

(2) h 度在案例网络中的模型实证

h 度的成功不仅因为其含义明晰、计算简洁，另一主要原因是其与传统方法之间常能找到有趣的相依关联。Hirsch 曾给出 h 度与总被引次数 C 之间的近似函数关系[11]:

$$C = a \times h^2 \quad (1)$$

其中，a 为文献信息中引征行为下的系数，Hirsch 认为 a 在 3 到 5 之间[11]。与此类似，若假设在网络情境下也有此规律，则设 $N_A$ 为节点 A 的带权度 (Weighted Degree), $h_A$ 为节点 A 的 h 度，则应有:

$$N_A = b \times h_A^2 \quad (2)$$

其中, b 为互联网信息中推荐行为下的系数。根据本文案例网络数据，可得到如图 1 所示的结果。由图 1 所见，在案例推荐网络中，h 度也较好地符合了公式(2)，可与 Hirsch 提出的理论模型相互印证。实证结果不仅拟合优度高，估计出的系数 b=3.62 也与 Hirsch 推测的理论系数值接近。这一结果说明, h 度在学术博客推荐网络中的运用可得到理论模型的支撑，其基础规律与 h 指数在文献信息的引证行为中的理论性质有相通之处。图 1 中的模型还揭示了在学术博客推荐网络中，节点 h 度数值在总体上也能体现节点带权度的总值，h 度越高，节点带权度通常也越高。当然，社会领域的模型几乎都是近似规律，由图 1 中散点分布可见，大部分散点只是在拟合函数曲线周围分布，而非刚好落于曲线上。原则上, h 度与带权度只能相互粗略估计，而不能精确互推。学术博客推荐网络中，节点带权度等于被推荐次数，公式(2)显示了在此类网络中，博主的 h 度与其总计的被推荐次数之间具有幂指数为 2 的近似幂律函数关系。

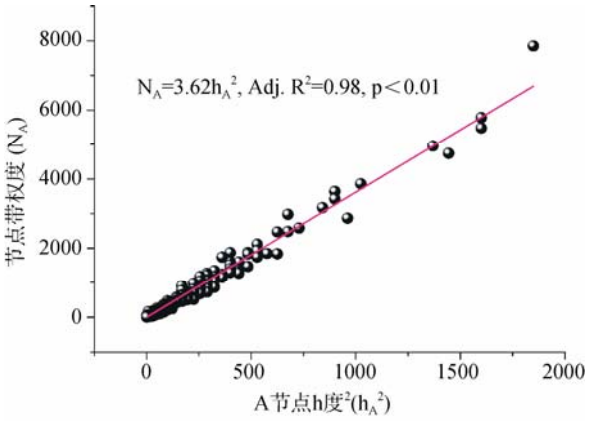


图 1 案例推荐网络中 h 度与节点带权度的近似拟合关系

(3) 高 h 度值节点在网络中的位置讨论

此外，还可注意到图 1 中 h 度值较高的节点数少，而低值较多，这与“二八律”和信息的幂律分布[15]一脉相承，必然导致网络在整体上呈现出核心节点聚于内而边缘节点散于外的结构模式，如图 2(a)的网络整体可视化图形所示。由图 2(b)–图 2(d)可见，三位 h 度最高的博主分别连接和控制了网络中心部分的一部分重要子群。在网络中心部位，曹聪与中部及偏左子群联系密切，李学宽与右下部节点关系较强，而杨正瓴与右部节点群体关联较多。由此可见，在结构上，学术博客推荐网络中的 h 度，主要体现了节点对局域子群的信息影响。当 h 度较高时，其影响的也常是网络的核心部分节点群。由此，当需要对网络中的信息观点进行干预和引导或需促进科学知识更广泛传播时，同时从多个不同的高 h 度的核心节点入手将是一种高效的方式。

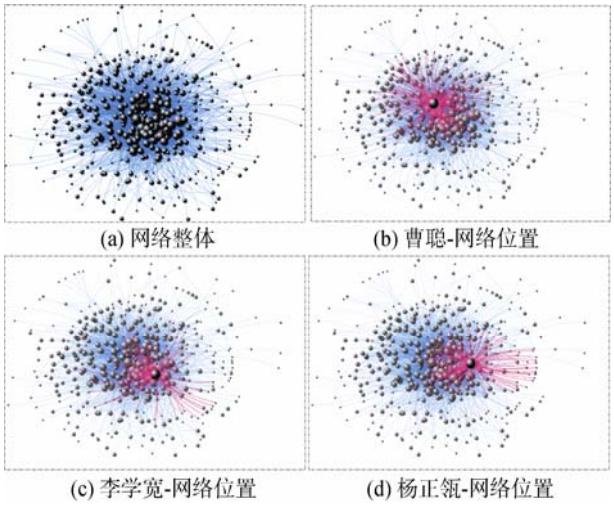


图 2　案例网络整体结果可视化及 h 度排名前 3 的博主(节点)在网络中的位置

## 5　结　语

本文将 h 度这一带权网络基础分析参量引入到非文献类信息网络研究中，并重点关注 h 度在学术博客推荐网络中的特性规律。将科学网博客这一典型的学术博客作为数据源，以信息推荐关系为网络联系类型，构建了学术博客的推荐网络，并以此讨论了 h 度在该类网络中的实证特性。结果发现，学术博客推荐网络中的高 h 度节点的产生既可由信息源(博主)所持有的学术知识内涵导致，也可因信息源提供话题的兴趣外延引发。h 度在推荐行为中符合 Hirsch 提出的基于引征行为的理论模型[11]。h 度($h_A$)与节点带权度($N_A$)之间存在形如 $N_A = b \times h_A^2$ 的近似函数关系，其中系数 b 的值也落入 Hirsch 的估值区间[11]。网络整体可视化图形则显示，高 h 度节点通常成为网络中心部分的局部子群组织者，从而为网络整体知识交流模式或管理提供了启示。

本文将 h 度的应用范围从文献类信息网络扩展到非文献类信息网络，并发现 h 度在学术博客推荐网络中的部分性质，从而为 h 度的后续研究提供参考，也为学术博客研究提供了可选择的工具。然而，作为单一指数，h 度显然并非完美。例如，在学术博客推荐中，只考察相连节点的直接推荐关系，没有考虑间接关系；对于推荐频度和活跃度等时间因素，缺乏明确的度量；而博主的推荐行为模式如何影响 h 度，将是今后值得深入探究的问题。h 度的算法本身存在诸多局限，如适用于整数处理、有时同值较多缺乏区分度等。未来研究可利用 h 度的各类变体，构建更丰富且目标不同的测量参数；也可在更多不同的互联网类信息网络中验证和应用 h 度，并比较其特性规律与本文的结果有何不同。

## 参考文献：

[1] Otte E, Rousseau R. Social Network Analysis: A Powerful Strategy, also for the Information Sciences [J]. Journal of Information Science, 2002, 28(6): 441-453.

[2] Borner K, Sanyal S, Vespignani A. Network Science [J]. Annual Review of Information Science and Technology, 2007, 41: 537-607.

[3] Zhao S X, Rousseau R, Ye F Y. H-Degree as a Basic Measure in Weighted Networks [J]. Journal of Informetrics, 2011, 5(4): 668-677.

[4] Schubert A. A Hirsch-type Index of Co-author Partnership Ability [J]. Scientometrics, 2012, 91(1): 303-308.

[5] Rousseau R. Comments on “A Hirsch-type Index of Co-author Partnership Ability” [J]. Scientometrics, 2012, 91(1): 309-310.

[6] Yan X B, Zhai L, Fan W. C-index: A Weighted Network Node Centrality Measure for Collaboration Competence [J]. Journal of Informetrics, 2013, 7(1): 223-239.

[7] Abbasi A. H-Type Hybrid Centrality Measures for Weighted Networks [J]. Scientometrics, 2013, 96(2): 633-640.

[8] Zhai L, Yan X B, Zhang G. A Centrality Measure for Communication Ability in Weighted Network [J]. Physica A-Statistical Mechanics and Its Applications, 2013, 392(23): 6107-6117.

[9] 刘玮. 电子商务系统中的信息推荐方法研究[J]. 情报科学, 2006, 24(2): 300-303. (Liu Wei. Research on Information Recommendation in E-commerce Systems [J]. Information Science, 2006, 24(2): 300-303.)

[10] 胡昌平，胡吉明，邓胜利. 基于社会化群体作用的信息聚合服务[J]. 中国图书馆学报，2010, 36(3): 51-56. (Hu Changping, Hu Jiming, Deng Shengli. Information Aggregation Service Based on the Role of Socialization Groups [J]. Journal of Library Science in China, 2010, 36(3): 51-56.)

[11] Hirsch J E. An Index to Quantify an Individual’s Scientific Research Output [J]. Proceedings of the National Academy of Sciences of the United States of America, 2005, 102(46): 16569-16572.

[12] Zhao S X, Ye F Y. Exploring the Directed H-degree in Directed Weighted Networks [J]. Journal of Informetrics, 2012, 6(4): 619-630.

[13] 徐孝娟，赵宇翔，朱庆华．民族志决策树方法在学术博客用户行为中的研究——以科学网博客为例[J]．现代图书情报技术，2014(1): 79-86. (Xu Xiaojuan, Zhao Yuxiang, Zhu Qinghua. Explore User's Behavior of Academic Blog Based on EDTM: Take Blog. Sciencenet as an Example [J]. New Technology of Library and Information Service, 2014(1): 79-86.)

[14] 周春雷，朱向林．科学网图情博客发展现状研究[J]．图书情报知识，2013(5): 98-105. (Zhou Chunlei, Zhu Xianglin. Study on LIS Blogs in Science Net [J]. Document, Informaiton & Knowledge, 2013(5): 98-105.)

[15] Egghe L. Power Laws in the Information Production Process: Lotkaian Informetrics [M]. Elsevier Academic Press, 2005.

**作者贡献声明：**

谭旻：设计并实施实证方案，论文撰写及修订；

许鑫：提出研究问题，部分数据处理，论文最终版本修订。

收稿日期: 2014-11-06

收修改稿日期: 2014-12-29

# The Empirical Study of h-Degree in Recommendation Network of Academic Blogs
## ——Taking ScienceNet.cn Blogs as an Example

Tan Min[1] Xu Xin[2]

[1](Department of Information Resource Management, Zhejiang University, Hangzhou 310027, China)

[2](Department of Information Science, Business School, East China Normal University, Shanghai 200241, China)

**Abstract:** [**Objective**] This paper studies the features of h-degree in recommendation network of academic blogs. [**Methods**] Based on the datasets of blogs in ScienceNet.cn in 2013, construct the recommendation network of academic blogs, calculate the h-degree and related measures, and enter discussion by information visualization. [**Results**] In recommendation network of academic blogs, the generation of nodes with high h-degree is not only caused by academic knowledge connotations which are held by the information source (bloggers), but also because of the interest from topic the information source provided. This paper explores an approximate functional relationship ($N_A=b\times h_A^2$) between h-degree ($h_A$) and node weighted degree ($N_A$). Nodes with high h-degree typically become the organizer of subgroup in the center of a network. [**Limitations**] H-degree is not a perfect indicator, and the future studies will expand the improved h-degree. [**Conclusions**] H-degree can be one of the measurements for recommendation network analysis of academic blogs, and h-degree is also important for community management of this kind community.

**Keywords:** Social media Academic blogs Information recommendation relationship Social network analysis Information network h-degree