文章编号: 1007-5321(2013)01-0001-13

社会搜索研究综述

程时端， 郭 亮， 王文东
(北京邮电大学 网络与交换技术国家重点实验室，北京 100876)

摘要: 社会搜索涉及众多研究领域，首先对比分析了社会搜索与传统搜索模式的关系，阐述了社会搜索的定义; 其次介绍了社会搜索的理论基础，并对社会搜索的研究现状和目前采用的关键技术进行了归纳总结; 分析了社会搜索目前面临的主要问题，并提出了社会搜索未来的发展方向，试图为该研究领域勾画出一个较为全面和清晰的概貌.

关 键 词: 社会搜索; 在线社会网络; 信息检索; 搜索引擎
中图分类号: TP393 文献标志码: A

A Survey on Social Search

CHENG Shi-duan， GUO Liang， WANG Wen-dong
(State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommanications, Beijing 100876, China)

Abstract: Social search is closely related to numerous research fields. First, the relationship between social search and traditional search pattern is analyzed and compared, the definition of social search is elaborated. Then, the theoretical basis of social search is introduced, the current status and key technology of social search are summarized. Finally, the new challenges of social search are analyzed, and the prospective research direction of social search is proposed. A more comprehensive and clear overview of the research area is outlined.

Key words: social search; online social networks; information retrieval; search engine

随着 web2.0 技术的迅猛发展，在线社会网络(OSNs，online social networks)作为一种新型的网络平台如雨后春笋般涌现，吸引了来自学术界和产业界的广泛关注. 人们在互联网上通过各种关系相互联系在一起，形成一个个规模庞大、关系复杂并且内容丰富的在线社会网络，用户在 OSNs 中通过交友、通信、协作、共享和发布内容等方式进行交互. OSNs 是真实社会网络的在线扩展，与社会网络具有相似的特征. 对于社会网络的研究始于 1967 年的小世界实验，哈佛大学的社会心理学家 Milgram[1] 通过一些社会调查对真实社会网络进行了实证研究，给出了著名的“六度分离”推断: 地球上任意 2 个人之间的平均距离是 6.199 8a，Watts 和 Strogatz[2] 在《Nature》杂志上撰文介绍了第 1 个小世界网络模型——WS 小世界模型. 小世界实验一方面揭示了社会网络的小世界特性; 另一方面也验证了社会网络是可搜索的[2-7]，即网络中的节点在只知道局部信息的前提下，自己可以有效地找到较短甚至最短路径.

随后在互联网上大量涌现的 OSNs 站点与社会网络服务都以六度分离理论为基础，帮助用户维护并扩展自己的人脉，提供各种社会化的服务. 互联网的应用模式也开始从传统的“人机对话”模式逐渐转变为“人与人对话”的模式[8-9]. 目前典型的 OSNs 站点有基于社交的 Facebook、MySpace、Orkut，微博

收稿日期: 2012-03-16
基金项目: 中央高校基本科研业务费专项资金项目; 国家自然科学基金项目(61271041); 移动重大专项项目(2012ZX03002008)
作者简介: 程时端(1940—)，女，教授，博士生导师，E-mail: chsd@bupt.edu.cn.

站点 Twitter、新浪微博，还有专用于商务联系的 LinkedIn、XING 以及基于位置的在线社会网络 Foursquare、Brightkite 等.

OSNs 站点经历了快速的增长，导致互联网中不断涌现大量的用户生成内容，在为信息的流动与传播提供便利的同时，也对传统的信息检索方式带来了新的挑战：用户对信息获取的需求逐渐偏向个性化与精准化，传统的信息检索方法已经不能满足人们的信息需求[10]. 传统搜索引擎虽然掌控着庞大的数字帝国，不断地更新索引甚至努力发展实时搜索，然而其找到的信息依然是古老陈旧的. 随着社会化应用越来越普及，涉及的人群也更为广泛，人们浏览网页和对网页的喜好程度比基于网站之间链接关系的算法更为准确. 此外，搜索结果的可信程度不仅与事实有关，还与用户好友的意见和情感有关. 将基于事实的搜索结果与用户好友的经验结合在一起，能够反映出用户所信任的人的意见和网络上的集体智慧. 因此，传统的只分析内容而忽略人的因素的搜索方式已经无法满足用户日益增长的信息需求和高效、方便获取信息资源的要求. 与搜索引擎不同，专注于社交关系的 OSNs 站点拥有海量的真实用户数据，包括用户的身份、用户之间的关系以及用户的在线活动等信息. 搜索引擎可以充分利用这些用户数据推测用户的兴趣喜好，提高用户对搜索结果的满意程度. 在这种背景下，为了适应搜索领域社会化发展的需要，传统搜索引擎纷纷将战略重点从原来的信息检索转向基于用户社会图谱的社会化搜索，分别经历了从第一代的目录式分类搜索、第二代的超链接搜索到第三代社会搜索的发展历程.

第一代的目录式分类搜索以 Yahoo、Infoseek 和 Altavista 为代表，搜索的目标是网站和网址. 这一代搜索反映的是搜索引擎的查全率，搜索结果的好坏往往通过反馈结果的数量来衡量. 第二代搜索引擎是以 Google 和 Baidu 为代表的超链接搜索. 搜索的目标是分散于整个互联网各个角落的信息，其工作原理如图 1 所示. 首先通过爬行器搜集、抓取互联网上的网页信息，同时由索引器对信息进行提取和组织，建立索引数据库，再由查询分析器根据用户输入的查询关键字，在索引数据库中快速检索出文档，进行文档与查询的相关度评价，对将要输出的结果进行排序，最后由搜索引擎接口将结果返回给用户. Google 使用 PageRank 算法，利用链路分析方法对 web 文档进行排序，将重要性的概念引入搜索中来，

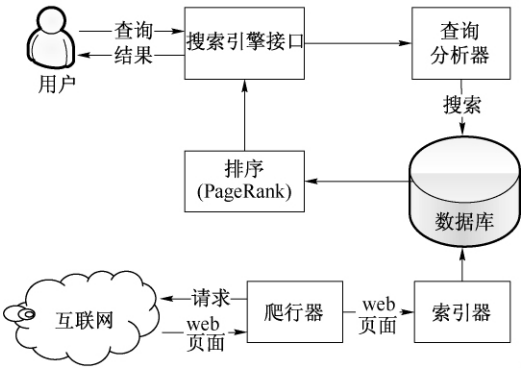

图 1 第二代搜索引擎工作原理

以衡量某个站点的受欢迎程度. 第二代搜索引擎相对于第一代搜索引擎来说，主要特点是提高了搜索的查准率，缺点是搜索与查询无关，即搜索引擎不区分用户，为所有用户提供相同的服务. 无论查询发起者的个性化需求是什么，只要输入的查询关键字相同，都返回同样的搜索结果. 第三代搜索引擎即社会搜索. OSNs 出现之后，社会搜索吸引了越来越多的信息检索研究人员的关注. 社会搜索是对传统搜索模式的一种补充，它将传统 web 搜索与用户的社会图结构结合起来[11]，将传统搜索引擎查找相关信息的功能演变为社会搜索中查找能够提供正确信息的用户. 虽然是基于人的搜索，但社会搜索的最终目的不是查找人，而是通过搜索众多用户的集体智慧来获取和改善搜索结果，帮助用户获得更加准确的信息.

社会搜索是一个跨学科的研究领域，它涉及数据挖掘、人工智能、自然语言处理、信息检索、社会学等多个方面. 近几年来，越来越多的学者致力于社会搜索的研究，取得了许多研究成果，国际上发表的论文数量大幅度增加，但是目前关于社会搜索的综述文章非常少. 因此，笔者综述了社会搜索的研究现状和目前采取的关键技术，并总结了社会搜索面临的问题和未来的发展方向，试图为该研究领域勾画出一个较为全面和清晰的概貌.

1 社会搜索定义

美国 Judysbook. com 网站在 2005 年对“Social Search”注册了商标，由此社会搜索一词开始在学术界和产业界广泛流行起来. Judysbook. com 网站对社会搜索的解释为：不同于大多数搜索引擎所做的把寻找本地信息的人引向陌生的网站，社会搜索使得搜索者不仅能够得到各种本地信息，还能与他/她所

信赖的社会关系们交流与分享对于各个信息的推介与评价. 维基百科对社会搜索的解释为：社会搜索或社会搜索引擎是将查询发起者的社会图结构考虑在内的 web 搜索类型. 与传统搜索分析文档内容和文档之间的链路结构相比，社会搜索提供的结果更能够体现用户创建或用户相关的内容.

不同的研究团体对于社会搜索也有不同的定义. Horowitz 等[12] 从搜索目标和信任关系 2 个角度对比了传统 web 搜索与社会搜索的不同. 首先，传统搜索引擎的目的是查找能够满足用户需求的文档或资源，而社会搜索是为了查找能够满足用户需求的人；其次，传统搜索中的信任是基于权威度的，而在社会搜索中，信任则是基于用户之间的亲密度关系. Teevan J 等[13] 通过分析查询日志，对比了传统 web 搜索与 Twitter 中社会搜索的主要区别. Twitter 搜索不仅实时性强，且信息多与人相关，主要被用来跟踪内容. Twitter 搜索结果包含很多社会内容和事件信息，而 web 搜索结果更多的是基本的事实和导航内容. 传统搜索引擎建立了信息与信息之间的关系，OSNs 建立了人与人之间的关系，而社会搜索则是要将信息与人关联起来，重建一种人与信息之间的映射.

总体来说，社会搜索是用社会化的形式将某一类型的信息聚拢到一起，将搜索的范围从整个互联网精确到某一范围内具有特定特征的信息上来，再引入特定的“量身定制”的搜索算法进行信息匹配，从而得到更加精确的搜索结果. 其本质是利用用户的在线社会网络，将 web 搜索与用户的 social graph 结合起来，同时强调用户的社会交互在搜索中的作用[14]，以达到提高搜索质量与相关性的目的.

2　理论基础

社会搜索是将传统搜索与 OSNs 结构结合起来的新型搜索模式. 从真实社会网络到在线社会网络，对 Social Networks 的研究已经有 40 多年的历史，一些著名的理论和方法为社会搜索领域的研究提供了支撑.

2.1　六度分离理论和无标度特性

在六度分离的小世界实验中，人们搜索目标对象时采用的是分散式的贪婪算法，即当前信件的持有者基于局部信息，以最有可能到达目标对象的方式来传递信件，实验结果虽然在某种程度上反映了社会网络的小世界特征，实际上却只有少部分的信件最终送到了收信人手中，实验最终的完成率很低，可信度也非常低. 因此，后续的一些研究人员开始在不同的场景下对小世界实验的结论进行验证[2-7,15]，并得出结论：社会网络中任意一对节点之间存在短路径. 这里的短路径指的是任意两点之间的平均距离与网络中节点总数的对数成正比[2].

Kleinberg[4-5] 在没有损耗的情况下证明：小世界中的短路径不仅存在，而且可以找到. Watts 等[6] 构建了一个层次网络模型，采用简单最近原则的下一步路由策略，证明构造出的网络是可搜索的. Dodds 等[3] 也通过实验说明：查询链能够成功完成，除了搜索能力，用户个人的传递动机也非常重要.

除了小世界性质，无标度特性是社会网络的又一重要特征. 无标度网络模型于 1999 年首次被提出，Barabasi 和 Albert 基于网络的增长特性和优先连接特性构造出第一个 BA 无标度网络[16]. 在线社会网络同样具有无标度特性[17]，OSNs 中节点的度分布函数具有幂律分布的形式. 无论是小世界特征还是无标度特性，对于在线社会网络中的资源搜索都是非常有利的，但又引出一系列新的问题：人们在社会网络中如何执行搜索？社会网络的哪些特征使其满足可搜索性？如何利用这些特征有效地搜索社会网络等. 这些问题都是社会搜索领域的研究重点.

2.2　社会网络分析

社会搜索不仅与搜索策略相关，还依赖于 OSNs 的网络结构信息、用户的交互行为方式等. 社会网络分析(SNA，social network analysis) 研究的出发点就是对行动者构成的社会关系结构进行量化分析. 这里的行动者可以表示单个用户，也可以表示群组. 社会网络分析方法将这些行动者看作网络中的节点，将节点之间的连接链路看作社会关系或信息的流动方式，为在线社会网络提供一种可视的、数学化的分析方法.

SNA 涉及复杂网络、数据挖掘、知识管理、数据可视化、统计分析和信息传播等多个领域，主要包括对用户节点的分析、对社会关系的分析、对社会群体的分析以及对社会网络拓扑结构的分析. 在社会搜索中，SNA 方法主要用来建立社会关系模型，研究分析 OSNs 的社会关系结构、用户交互行为模式以及社会关系结构与用户行为模式的相互影响作用等[17-20]. 在社会搜索的信息检索过程中结合 SNA 技

术有助于提高搜索结果的质量[21].

2.3 弱连接理论

弱连接理论由美国社会学家 Granovetter 提出[22-23].他将社会网络中人与人之间的关系按照沟通频率简单地划分为强连接和弱连接,通过寻访麻省牛顿镇的居民如何找工作来探索社会网络,发现人们在找寻工作时,关系紧密的强连接朋友反倒没有那些关系一般甚至只是偶尔见面的弱连接朋友更能发挥作用.

弱连接理论强调了弱连接在社会关系中的重要地位.强连接将人们紧密的联系起来,形成一个个小的圈子,而弱连接将这些小圈子连接起来形成一张更大的网络.虽然不如强连接关系紧密,但弱连接具有很高的传播效率.Dodls[3]进一步证明了成功的社会搜索主要是利用弱连接,而不需要高度连接的 hubs.因此,在社会搜索中,为了提高搜索效率,应该充分利用和发挥弱连接低成本与高效率的优势.

综上所述,理论基础与社会搜索的关系如图2所示.六度分离理论与无标度特性为社会网络的可搜索性提供依据,SNA 通过分析社会网络的结构特征和用户特征,为社会搜索中的信息提供量化的分析方法,弱连接理论则有助于提高社会搜索的效率.

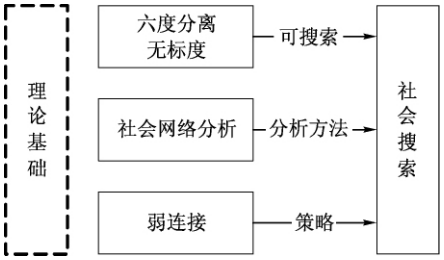

图2 理论基础与社会搜索的关系

3 研究分类

社会搜索借助社会网络的力量来参与搜索引擎结果的筛选和整理过程.从专家智慧和集群智慧角度出发,可以将社会搜索分为两大类.专家智慧主要关注社会网络中的搜索策略和社会搜索模型,而集群智慧旨在通过社会网络的特征来重新整理搜索结果.基于集群智慧的社会搜索又可以分为基于标签的社会搜索、基于社会网络结构的搜索、社会化推荐和个性化搜索4大类.下面分别进行论述.

3.1 搜索策略

搜索策略的核心是在网络中寻找从源到目的地代价最小的路径.在许多实际网络中,单个节点无法充分掌握整个网络的全局结构,甚至不知道目标节点在网络中的位置,这就需要一定的搜索策略,通过制定一定的规则来指导消息的传递过程.在社会搜索中,搜索策略即描述用户查询在 OSNs 中的分散搜索过程.

最简单的搜索策略是广度优先搜索,搜索的每一步都向所有邻居节点传递查询消息,直到找到目标节点的任一个邻居为止.该策略的优点是搜索速度快,通常只需要短短几步就能将查询遍布整个网络.然而随着网络规模的扩大,该策略产生的大量的查询消息流量会导致网络拥塞.因此,在实际应用中,就需要同时考虑效率和开销的搜索策略,下面分别介绍基于节点度的搜索、随机游走策略、基于相似性的搜索、分布式搜索模型、基于社会行为的路由.

1) 基于节点度的搜索

节点的度定义为与该节点连接的其他节点的数目,即邻居节点的数量,在有向网络中又分为出度和入度.出度表示从该节点指向其他节点的边的数目,入度则表示从其他节点指向该节点的边的数目.节点的度在某种程度上能够反映该节点的重要性.Mislove 等[17]通过对4个流行的 OSNs 站点进行大规模测量分析,证明了 OSNs 中节点的度符合幂律分布特性,OSNs 是无标度网络.

Adamic 等[24]认为,为达到理想的搜索效果,算法必须符合“按度序列搜索”的设想.因此,2001年首次提出最大度搜索策略.在假设每个节点都知道自己所有邻居度的前提下,依据邻居节点的度在网络中执行搜索:算法的每一步均选择当前节点度最大的邻居作为下一步节点,重复这个过程直到找到目标节点的任一邻居.Adamic 在 Gnutella 网络上验证了最大度搜索策略的有效性.然而,该策略只适用于幂律指数介于2.0~2.3之间的无标度网络,对于幂律指数大于2.3的无标度网络搜索效率极低.同时,最大度搜索没有利用目标节点的信息.

2) 随机游走策略

与最大度搜索类似,随机游走策略(RWS,random walk search)同样不考虑目标节点的信息,每个节点随机选择一个邻居作为下一步节点,重复这个过程直到寻找到目标节点的任一个邻居.根据算法对访问节点的不同限制条件,可以将随机游走策略分为3种类型,即无限制的随机游走、不返回上一步节点的随机游走和不重复访问节点的随机游走.

随机游走是一个基础的动态过程,它的优点是

能够快速访问远程连接，在规则网络、随机网络、小世界网络和无标度网络中，都有关于随机游走策略的研究工作.

3) 基于相似性的搜索

与最大度搜索和随机游走策略不同，基于相似性的搜索要将目标节点的信息考虑在内，搜索的每一步是寻找与目标节点最相似的邻居节点作为下一步节点. 这里的相似性有多种度量方式，包括地理位置临近度、兴趣相似度、职业相似度等. 相似性搜索利用社会网络的趋同特性，即拥有相似特征的人更倾向于建立连接，这一特性能够帮助寻找最短路径.

针对相似性搜索最早的研究工作是 Kleinberg 的网格模型[4-5]. Kleinberg 将网络中的 N 个节点分布在一个二维网格上，节点之间的距离定义为节点之间的网格步数，每个节点通过短程或长程连接与网格中的其他节点相连. 采用基于局部拓扑信息(网格结构、目标节点在网格上的位置和路由路径上节点的位置及其长程连接)的贪婪路由策略执行搜索，即节点将信息传给离目标节点最近的短程或长程连接的邻居节点.

随后，Watts 等[6]提出了一种基于社会距离的层次网络模型. 在该模型中，节点根据职业、地理位置、兴趣等特性聚集成一些比较小的群，这些群又根据他们的共同特性聚集成规模更大的群，从下往上一层一层汇聚，最终形成一个层次网络. 节点间的社会距离定义为 $y_{ij} = \min_h x_{ij}^h, h = 1, 2, \cdots, H$. 其中 x_{ij} 定义为节点 i 和 j 最低的共同上级所在的层数，同一个群中两节点间的社会距离为 1，H 为分层标准重数. 同样是采用贪婪路由策略，节点选择距离目标节点社会距离最近的邻居作为下一步节点.

Adamic 等[25]在真实的 HP 实验室 Email 联系人和学生社交网站两种场合下，根据网络中节点的 3 种不同属性(节点的度、节点在实验室组织层次图中的位置、节点的物理位置)分别采用以下 3 种不同的策略进行搜索，包括最大度搜索、基于目标节点组织层次的搜索、基于目标节点地理位置的搜索. 实验结果表明，利用目标节点信息的两种搜索策略的效率胜过仅利用邻居节点度的最大度搜索策略. 而利用目标节点职业位置的搜索策略则优于利用目标节点地理位置的搜索策略.

还有一些研究工作将网络的趋同性与度分布特性结合起来设计消息的路由策略. Simsek[26]提出了 EVN(expected-value navigation)算法，同时结合网络的趋同性和 degree disparity，在只知道网络局部拓扑的情况下寻找目标节点，该算法无论是在小世界网络还是在无标度网络中都能够有效地执行分布式搜索. Imsek[27]也提出一种相似性与度混合的路由规则 EVN. 节点将信息传递给具有最优化度量值的邻居节点，依次下去，直到将信息传递到目标节点. 度量函数定义为节点之间相似性度量参数 q_{ij} 与节点度 k_i 的混合函数. 若网络中不存在相似性，则等同于最大度搜索策略；若所有节点的度均相等，则类似于其他相似性搜索策略. 但是在实际网络中，这 2 种情况都很少见.

4) 分布式搜索模型

分布式搜索模型是在网络上传递用户查询的分布式搜索系统，其关键任务是在不利用中心索引的情况下高效地将查询请求转发到能够正确回复的节点. SQM(social query model)就是一个典型的分布式搜索的概率模型，其路由策略的有效性取决于网络中任一节点发起查询之后，其他节点予以正确回复的概率. 传统的马尔可夫决策过程和 PageRank 模型都可以看作是 SQM 的特例[28].

SQM 模型基于节点的专业程度、响应速度和路由策略等实际因素进行建模，为网络中的所有节点同时提供最优的查询路由策略. 其基本思想如下：

网络中的节点集为 $S = \{x_1, x_2, \cdots, x_n\}$. N_i 表示节点 x_i 的邻居节点集合. 任一节点 x_i 收到查询时，有 3 种可能的操作，即丢弃该查询、响应该查询、将查询转发给其他节点.

查询模型由以下 4 部分组成.

① Expertise：专业级别 e_i 表示节点 x_i 响应查询的概率. x_i 将查询转发给 N_i 中邻居的概率为 $(1 - e_i)$. 用 $n \times 1$ 的向量 $\boldsymbol{e}$ 来表示所有节点的专业级别.

② Correctness：当节点 x_i 以概率 e_i 响应查询时，令 w_i 表示答案正确的概率. 用 $n \times 1$ 的向量 w 来表示正确度.

③ Response Rate：节点 $x_j \in N_i$ 收到来自 x_i 的查询请求，以概率 r_{ij} 接受该查询，以 $(1 - r_{ij})$ 的概率丢弃该查询. 用 $n \times n$ 的矩阵 $\boldsymbol{R}$ 来表示响应率.

④ Policy：节点 x_i 将查询转发给邻居节点时要依据一定的策略 Π^i. Π^i 是 N_i 的概率分布. 节点 x_i 以概率 $(1 - e_i)$ 将查询转发，转发给邻居节点 x_j 的概率为 Π_j^i. 节点 x_j 可以概率 r_{ij} 接受，或者以概率 $(1 - r_{ij})$ 忽略该查询. 网络中所有节点的整体策略用一个 $n \times n$ 的矩阵 $\boldsymbol{\Pi}$ 来表示，其中第 i 行为 Π^i.

查询路由的核心问题是：节点 x_i 得到由它发起查询正确答案的概率 P_i. 给定一个策略 Π，节点在2种情况下可以得到正确答案.

① 响应该查询并且知道正确答案，概率为 $w_i e_i$；

② 以概率 $(1-e_i)$ 转发该查询，将查询发送给邻居 x_j，概率为 Π_j^i. 随后，节点 x_j 以概率 $(1-r_{ij})$ 忽略该查询，或者以 r_{ij} 的概率接受该查询，并尝试给出答案. 假设 x_j 能够给出正确答案的概率为 P_j，则对于任意的 P_i，有如下递归公式：

$$P_i = w_i e_i + (1-e_i)\sum_{j\in N_i}^{n} \Pi_j^i\left((1-r_{ij})\times 0 + r_{ij}P_j\right) = w_i e_i + (1-e_i)\sum_{j\in N_i}^{n} \Pi_j^i r_{ij} P_j$$

另外一个分布式搜索模型是 iLink [29]. 同样是在社会网络中实现搜索和消息路由的模型，它与SQM的主要区别在于对消息的路由转发由 Supernode 负责，即查询消息由超级节点根据专业程度和响应速度等进行路由.

分布式搜索模型最典型的例子是 Aardvark [12]. Aardvark 是最早的社会化搜索引擎之一，负责提供问答式的社会网络服务. Aardvark 通过特定的问题，将用户和最有可能拥有答案的用户连接在一起. 用户之间采用对称的亲密度度量，亲密度通过加权特征的向量夹角余弦相似度计算得出，包括词汇匹配、profile 相似性以及真实生活中的社会连接等.

5) 基于社会行为的路由

基于社会行为的路由旨在利用社会网络中用户的行为特征和方法来指导网络中的消息传递过程. 在日常生活中，人们会通过朋友或朋友的朋友来传递消息，Daly 等 [30] 基于这种朋友圈的思想提出了一种 SimBet 路由算法，基于 SNA 技术在 DTN 网络中实现路由. Costa 等 [31] 提出一种基于社会意识的路由算法，通过选择具有最佳功效值的消息载体来实现消息的路由. 功效值代表了节点作为转发节点的能力，与节点的位置信息和可能与目标节点分享同一兴趣的概率有关. 与目标节点分享同一兴趣的概率可通过卡尔曼滤波器进行预测.

3.2 基于标签的社会搜索

Tagging 即标签，是用户对互联网上的内容以 keywords 或 tags 的形式进行标记和分类，以便于进行内容的组织、过滤和搜索的过程. 标签将用户和资源联系起来，在描述资源特征的同时也体现了用户的兴趣偏好. 而社会化书签是采用标签的方式对资源进行内容聚合，通过使用不同的标签把不同主题的资源归类，从而实现知识管理与信息共享的系统. 社会化书签的主要功能包括存储、分享、发掘链接信息，其最大的特点是集合众人之力，以用户自定义标签的方式对网站进行分类. 最具代表性的社会化书签系统是 Del. icio. us，中文名为“美味书签”.

标签和书签信息都是由用户自行添加，为内容组织和资源共享提供了新的方式，可以通过它们挖掘用户的喜好并预测用户行为，从而进一步提高社会搜索的效率，为用户提供个性化的服务.

Xu 等 [32] 将 tags 简单分为5类.

① Content-based tags：描述对象的内容或一个对象所属的分类，如“电脑”、“lucene”.

② Context-based tags：描述上下文信息. 如描述位置和时间信息的“北京”、“2012-01-01”.

③ Attribute tags：描述对象隐含的内在属性，如“Nikon”(相机的类型)、“homepage”(网页的类型).

④ Subjective tags：主观描述对象的标签，如“funny”、“cool”.

⑤ Organizational tags：标识个人信息，或是对某项任务的描述，如“my paper”、“todo”.

Cattuto 等 [33] 从路径长度的角度说明了基于标签的信息发现确实是有效的. 在标签关系图中，即使图中的节点数增加，标签之间的距离也保持的非常小. 作者通过实验证明，资源之间的平均路径长度为3.5，即从任意一个资源出发，平均经过3.5步就可以发现另外一个资源. 这样短的路径长度说明，在搜索过程中，可以利用标签对资源进行组织分类的特点来改进搜索.

标签信息不仅可以很好地概括相应的 web 页面，其数量在一定程度上也可以反映一个 web 页面的流行程度. 除了利用标签进行信息发现、共享以及社团的排序，标签也可以帮助社会搜索完成信息的组织、搜索和提取工作. Bao 等 [34] 定义了3类与 social search 相关的用户，包括 web 页面创建者、页面标注者和搜索引擎用户. web 页面创建者负责创建页面并将页面通过链路连接起来，页面标注者使用 annotations 组织并记录共享页面的用户，搜索引擎用户则通过使用搜索引擎获取信息. 这3类用户相互重叠，同一个用户可以充当多个角色. Bao 等 [34] 提出2个算法说明了社会注解，即标签信息在搜索过程中发挥的作用. 首先利用 SocialSimRank 算法计算标注信息与 web 查询之间的相似度；再通过 So-

cialPageRank 算法度量 web 页面的流行程度，从用户的角度来描述页面的质量.

典型的社会化书签系统通常包含 3 种数据单元.

① Triple：〈$user_i$，tag_j，url_k〉三元组，表示用户 i 为 URL k 标记了 tag j.

② Post：用户标记过的 URL 或其他相关的元数据.

③ Label：一个 label 是一个〈tag_i，url_k〉对，表示系统中至少存在一个包含 tag i 和 URL k 的三元组.

由于用户标签信息的相关度很高，采用书签系统作为搜索的数据源，在搜索结果中得到的 URLs 通常是最近更新，也是最权威的. 然而与大规模的 web 数据量相比，del. icio. us 等书签站点只提供了少量的数据. 用来注解 URLs 的标签虽然相关性很强，但是大都能通过上下文来确定. 例如有将近 1/6 的标签出现在页面的标题中，有 1/2 的标签出现在页面正文中. 除了页面内容，许多标签可以直接通过 URL 的域名来确定[35]. 这说明目前由社会化书签系统提供的 URLs 为搜索引擎提供的数据仍然不够充分.

社会化书签系统的核心数据结构是 folksonomy，分众分类. Folksonomy 可以看作是由不同用户为资源添加的 tag 批注的轻量级分类结构，由用户、标签、资源以及他们之间的三重关系组成. 利用 folksonomy 结构不仅能够提高社会搜索的检索精度，也能够提高检索的查全率. 文献[36-37]中通过建立用户—标签关系模型和标签—资源关系模型，无缝地将标签映射到依赖于某个特定用户查询的资源上，从而帮助用户寻找到最相关的社会媒体信息. 有些研究工作也关注 folksomony 的模型和排序方法. Hotho 等[38]提出了 FolkRank 搜索算法，用来挖掘 folksonomy 的结构. 该算法可以在 folksonomy 里发现社团，也可以用来结构化搜索结果. Krause 等[39]通过对比 folksomony 排序与搜索引擎排序发现，与传统搜索引擎用户相比，社会化书签系统的用户只关注少量的话题. Del. icio. us 用户大多关注 IT 领域，系统中前 20 个 top tags 有 11 个是与计算机相关的术语. 但是 2 个系统 URLs 的重叠度很高，这说明社会化书签系统的用户更倾向于对传统搜索引擎排列较高的 web 页面进行标记. 还有一些研究注重分析社会化书签系统中标签的流行程度和分布规律[40]，通过对标签结构和分布规律的学习有助于实现知识共享与发现.

除了利用标签描述内容（资源是什么），也可利用标签信息来描述意图（资源可以用来干什么）. Purpose tags[41]方法将给定资源所有可能提供的服务组合起来，每个资源能提供的服务不只限于它的内容，也与跟他交互的代理有关. Purpose tagging 重点捕捉意图而非内容，通过分析资源服务的不同作用，purpose tags 能够协调用户的意图与社会化应用系统中内容之间的关系.

3.3　基于社会网络结构的搜索

以 PageRank 为代表的基于链路分析方法的搜索排序技术已经有 10 多年的发展历史，而社会搜索将用户的个性化信息和人际关系叠加到链路分析之上，通过利用社会网络的结构特征来实现搜索和排序[42-43].

除了小世界现象和无标度特性，拥有明显的社团结构是 OSNs 的又一个重要特征. 社团同标签一样，是一种有效组织信息的方式. 所谓社团结构，是指整个社会网络由若干个“group”或“cluster”构成. 每个社团内部节点之间的连接非常紧密，而各个社团之间的连接则相对来说比较稀疏. 社团结构是对社会网络信息的中观度量，体现了社会网络传播信息的能力. 如何把搜索与社团结构两者有效地整合起来，形成人与人可以交流的、基于群体特征的搜索，是社会搜索领域需要研究的问题.

基于社团结构的搜索通常分为两步：社团结构的识别与发现；社团用户参与搜索的过程.

对于网络社团结构的研究已经有很长的历史. 早期的社团发现方法大多借鉴图形分割的理论，如著名的 Kernighan-Lin 算法和基于 Laplace 图特征值的谱平分法. 2002 年，社团发现由 Girvan 和 Newman 提出[44]，基本思想是采用分裂式的层次聚类方法，不断地从网络中移除边介数最大的边. 边介数定义为网络所有最短路径中经过该边的路径数目占最短路径总数的比例，反映的是边在整个网络中的桥接作用. 随后，Newman 等[45]又相继提出了基于分裂式层次聚类和凝聚式层次聚类[46-47]的社团发现方法. 按照不同的实现途径，又可以将社团发现技术分为基于 HITS 的技术、基于有向二分图的技术和基于网络流量的技术.

除了社会网络、神经网络等复杂网络，在 web 页面和博客中也普遍存在社团结构的概念. Li 等[48]通过分析 web 页面和 blogs，利用句子、文档中单词和

实体的并发情况来发现社团结构. 单词和实体的并发出现通常隐含了它们之间的连接关系. 笔者将社团发现转换为图聚类问题,提出一个分层的聚类算法. 首先生成一个加权的名字实体图,再将名字实体图聚类为社团. 给定一个 web 文档集合,发现包含同一个实体社团结构的工作流程如下:

1) 将 web 文档转换为自由文本;

2) 将每个文档切割为句子,用一个命名实体解析器提取出包含多个名字实体的句子;

3) 将句子转换为无向图,名字实体为顶点,包含超过一个名字实体的句子为边,边的权重为实体名字的并发频率;

4) 使用凝聚聚类算法将加权图聚类为社团.

除了辨别可能的社团结构,定义和解释已有的社团也有利于社会搜索的信息发现过程. Du 等[49]在提出社团检测算法 ComTector 的同时,将网络拓扑信息和节点的自然属性结合起来,提出一种描述已发现社团的通用命名方法.

对于社会搜索还有一种解释是社团用户参与搜索的过程. 在检测到社团结构之后,社会搜索的任务是如何利用这些社团属性来改进搜索结果.

Peter 等[50]提出一种分布式 CWS(collaborative web search)模型,以协作的方式,利用社团搜索者形成的对搜索结果的推荐,来补充搜索引擎返回的结果. 该模型将搜索行为存储为一个社团矩阵(H_c,hit-matrix),将查询与结果关联起来,如 $H_c(i,j)$ 表示社团 C 中的成员为查询 Q_i 选择页面 P_j 的次数. 协作方式采用一种动态的信任模型,用户之间的连接强度由信任分值来决定. 用户 u_t 和 u_s 之间的信任分值,定义为 u_t 选择过的 u_s 的推荐的比例. 随着网络的演进,该模型的推荐功能和搜索的性能会逐步提高.

Agichtein 等[51]通过挖掘大型社区问答平台 Yahoo! Answers 中的内容和用户交互信息,利用社团反馈自动地识别高质量的内容. 也有研究在传统的搜索方法中引入对社会距离的度量来改善搜索结果的排序效果[52]. 首先修改聚类算法(RNM,recursive neighborhood mean)[53],在 modified-RNM 的基础上采用 k-means 方法识别拥有社团属性的子群,分析子群中的成员对搜索结果的预测能力. 实验结果表明,搜索用户更倾向于在自己所在的子群内部选择搜索结果.

社团具有重叠特性和分层结构,而全球分层的节点不能捕捉重叠群组之间的关系. 因此,Ahn 等[54]人重新考虑用链路群组代替节点来定义社团,这种非传统的方法可以让 overlap 与 hierarchy 不冲突,既可以表示分层结构,又能够表现出社团的重叠特性. 基于链路的社团结构揭露了网络中的重叠和分层组织是相同现象的 2 个方面,社会搜索可以利用社团的这种重叠特性和分层结构来指导消息的路由与专家定位过程.

3.4 社会化推荐

社会搜索的本质与传统搜索是一致的,都是帮助用户找到他们想要的信息. 在使用传统搜索引擎时,要得到相关度高的搜索结果,需要用户键入合适的关键词进行查询. 然而,随着互联网上信息的爆炸式增长,用户的需求很难通过几个简短的关键字来准确地描述. 面对搜索引擎返回的海量结果,用户也无法做出快速而有效地选择. 因此,在社会搜索中,需要一种能够推测用户兴趣和偏好的技术,主动地向用户做信息推荐. 社会化推荐就是在社会媒体应用中引入推荐技术,主动为用户推荐可能感兴趣的信息或服务. 推荐与搜索技术互补,利用数据挖掘和知识发现技术,根据用户个人的习惯和偏好等向用户推荐信息. 推荐的内容通常包括商品、服务等多种类型,这里将被推荐的信息类型统称为项目. 根据推荐算法的不同,可以将推荐技术分为基于内容的推荐、协同过滤推荐和混合推荐.

基于内容的推荐不需要依赖用户对项目的评价意见,而是根据用户已经选择过的项目,推荐其他具有类似属性的项目给用户. 基于内容推荐算法的根本在于信息获取和信息过滤. 首先通过特征提取方法得到项目内容特征向量 $\boldsymbol{C}(s)$,推荐系统则根据用户所评价项目的特征,预测用户的兴趣偏好,从而为用户进行推荐. 假设对用户 c 的描述信息向量用 $U(c)$ 表示,在基于内容的推荐系统中,推荐函数通常被定义为[55] $\mu(c,s) = \mathrm{score}(\boldsymbol{U}(c),\boldsymbol{C}(s))$. 其中,score 值可以利用向量的余弦夹角等方法计算得出,最后根据 $\mu(c,s)$ 值对项目进行排序,将 $\mu(c,s)$ 值最大的若干个项目作为结果推荐给用户 c. 除了传统的基于信息获取的推荐方法之外,还有一些研究利用机器学习的方法,如贝叶斯分类、聚类分析等,通过分析已有的数据得出模型,再进行基于模型的推荐. 基于内容推荐算法的研究还包括自适应过滤和阈值设定等.

总体来说,基于内容的推荐算法可以提供精确

的推荐结果[56]. 然而该类算法只适用于被推荐的项目具有丰富的内容信息，并且能够自动对内容进行提取，如对书籍、文章和书签的推荐，对视频、图片等地推荐则不适合.

协同过滤(CF，collaborative filtering)是应用最广泛、最成功的推荐技术，它的基本思想是具有相似兴趣爱好的用户会对同一个项目表现出相似的偏好. 协同过滤首先需要利用用户的历史信息计算用户之间的相似性，然后利用相似用户对项目的评价来预测目标用户对特定项目的喜好程度，系统再根据这一喜好程度对目标用户进行推荐. CF 推荐算法是从用户角度进行推荐，在用户之间形成一种自助、协同式的社会推荐模式，对推荐对象没有特殊的要求，能够处理视频、音乐、图片等难以进行文本结构化表示的项目. 可以将 CF 推荐算法分为两类[55]：基于记忆的算法和基于模型的算法. 基于记忆的算法首先计算用户之间的相似度[57]，通过聚合分析所有相似用户对项目的评分，来预测目标用户的喜好. 基于模型的算法利用用户对大量项目的评分数据来学习并推断用户行为模型，然后基于该模型采用概率统计的方法对新项目进行预测评分[58-60].

混合推荐通过组合不同的推荐技术，来弥补各推荐技术单独使用时的缺陷. 目前最常见的混合推荐方法是基于内容推荐方法和基于 CF 推荐方法的组合，如推荐列表线性组合的方式，或将不同推荐算法整合到一个统一的框架模型下计算推荐函数等，不同的组合方式适用于不同的应用场景.

除了以上介绍的 3 种推荐技术外，还有周涛等[61-62]提出的基于网络结构的推荐算法、基于知识的推荐方法、基于规则的推荐方法以及不同推荐算法组合的新的混合推荐技术.

虽然这些主流的推荐技术在一定程度上考虑了用户信息以及用户之间的相似性度量，但仍然没有充分利用 OSNs 中用户的社会属性. ReferralWeb[63]是第 1 个结合社会网络与协同过滤技术的系统. 在定位特定专家的同时，能够将用户和专家的社会网络进行可视化，以提供一条推荐关系路径. 一些大规模实证研究结果显示，信息推荐中社会关系往往比推荐内容和用户喜好的匹配程度更加重要. 事实上，用户更喜欢来自朋友而非来自系统的推荐. Schenkel 等[64]提出一种利用社会关联来进行搜索和推荐的框架，在设计评分模型时综合考虑用户的社会关联和项目与标签之间的语义关系. 在该模型中，评分值与用户相关，并取决于查询发起者的社会上下文. 用户查询由一系列标签表示，在评分过程中首先利用简化的 BM25 算法计算与查询用户 u 相关的拥有标签 t 的文档 d 的得分 $s_u(d,t)$. 为了对标签进行语义扩展，引入对标签相似性的度量值 $\mathrm{tsim}(t_1,t_2)$，将目标文档 d 的最终分值定义为 $s_u^*(d,t) = \max_{t' \in T} \mathrm{tsim}(t,t') s_u(d,t')$. 在计算标签相似性时，除了语义关联，也加入了对用户社会关系的考虑，即如果 2 个标签被用户的同一个好友标记过，则相似度更高：

$$\mathrm{tsim}_{so}(u,t,t') = \sum_{u' \in U} F_u(u') \frac{\mathrm{d}f_{u'}(t')}{\mathrm{d}f_{u'}(t \cap t')}$$

其中 $F_u(u')$ 表示用户 u 与 u' 的好友关系强度. 最后，通过对目标文档中与查询相关的所有标签得分计算求和，得出目标文档的最后评分值进行排序和推荐.

McNally 等[65]在协作搜索系统 HeyStaks① 的基础上提出一种信誉机制，从结果推荐的角度来建模单个搜索者的信誉值. 这里的信誉是对查询者可信度的度量. 如果一个查询者对搜索知识做了贡献，并且这个知识在之后的搜索过程中频繁地被其他用户选择，则该查询者的信誉值就高. HeyStaks 系统提供的主要功能是过滤、排序和结果推荐. 通过用户的信誉模型能够识别出一个社团的关键人物，也可以利用用户的信誉值来改进结果的推荐. McNally 等[66]扩展了 HeyStaks 社会搜索模型，利用 HeyStaks 用户的搜索知识、信誉等信息来改善推荐结果.

Guy 等[67]基于这样的假设：与用户好友相关的 items 更能够吸引该用户，考虑利用用户的社会网络来改进推荐系统的效果. 通过收集社会网络信息，基于用户的社会关系研究社会化媒体的推荐. 除了社会关系，与用户标签相关的 items 同样能够提高推荐质量. 因此，Guy 等[68]在推荐系统中同时加入用户的社会关系和标签信息，并且通过实验证明将两者结合起来设计推荐系统能够得到更好的效果.

总体来说，从社会搜索的角度，社会化推荐主要关注推荐技术在社会媒体信息发现中的应用. 而从推荐系统的角度，社会化推荐的目的是利用社会特性来解决推荐系统中的难题，如冷启动等[69].

3.5　个性化

由于知识背景和环境的差异，用户的搜索需求不同，每个人对于相同关键词所期待的搜索结果是

① www.heystaks.com

千差万别的. 因此,需要搜索引擎从传统的面向内容的方式,转变为面向用户的方式,为用户提供个性化的搜索结果. 所谓搜索的个性化,是指利用用户协作,根据用户兴趣偏好以及地理位置等个性化特征提供个性化搜索服务,以满足不同用户的不同需求[70-71]. 社会搜索中融入了大量社会化元素,是实现个性化搜索的天然平台.

有研究工作利用用户过去的隐式活动建立用户profile,根据这个profile来设计关联性反馈框架,利用与用户兴趣相关的隐式信息对web搜索结果进行重新排序[72],以实现个性化搜索. 结果表明,在改善搜索的个性化算法中,用户和用户资料集的表示对个性化是十分重要的,可以利用这些特征来设计有效的个性化搜索算法.

Carmel等[73]利用社会搜索系统SaND(social networks and discovery)获取用户数据,基于用户的社会关联研究个性化社会搜索. SaND系统提供的社会聚合业务包括社会搜索、个性化item推荐、个性化人物推荐、寻找人与人之间的社会路径以及其他的社会网络服务. 将与用户相关的实体分为document、persons、tag、group 4种. SaND负责在社会数据中获取$N(u)$和$T(u)$,$N(u)$表示与用户u相关的用户分级列表,$T(u)$表示与用户u相关的terms分级列表. 用户的profile表示为$P(u)=(N(u),T(u))$,对搜索结果的排序计算方式为

$$S_p(q,e|P(u))=\alpha S_{np}(q,e)+(1-\alpha)\left[\beta\sum_{v\in N(u)}w(u,v)w(v,e)+(1-\beta)\sum_{t\in T(u)}w(u,t)w(t,e)\right]$$

其中:$S_p(q,e|P(u))$为给定用户u的profile和实体e对查询q的个性化排序值;$S_{np}(q,e)$为e对q的非个性化排序值;$w(u,v)$和$w(u,t)$分别为用户v和term t对用户u的关系强度;$w(v,e)$和$w(t,e)$分别为用户v和term t对实体e的关系强度. 一个实体首先由SaND根据非个性化的方式计算出排序值$S_{np}(q,e)$;再通过用户profile中的用户和terms的关系强度对该值进行个性化的修正.

Wang等[74]通过挖掘用户在OSNs上的公开活动,如博客、社会书签等来个性化搜索服务. 个性化搜索方案分为3步.

1) 提取OSNs中用户的兴趣,为每个用户建立并维持兴趣档案.

2) 系统收到来自用户的查询请求,按照如下步骤进行处理:

①将查询转发给当前的搜索引擎,搜索引擎返回一系列与查询相关的web页面,每个页面对应1个相关性分值;

②系统从用户的兴趣档案中获得一个兴趣向量,向量中的每个元素包含1个词和1个表示兴趣度的分值,在为用户建立兴趣向量时,对不同的社会信息资源赋予不同的权重;

③对搜索引擎返回的top n个页面,系统基于该页面与用户兴趣向量的匹配程度为其计算兴趣值;

④对每个页面,综合其相关性分值和兴趣值计算出最终分值,通过个性化参数来调整兴趣值对最终分值的影响程度;

⑤系统根据web页面的最终分值选出要返回给用户的结果列表.

3) 系统根据用户对搜索结果的反馈作调整,调整的目的是找到一系列适合特定用户的参数集.

笔者通过搜集博客、社会书签和标签3类数据资源进行了验证. 结果表明,综合多个社会系统的信息得到的个性化搜索结果比只依靠单个社会系统的信息效果更好.

4 面临的问题和发展趋势

针对社会搜索,笔者概述了其理论基础和主要研究方向,分别阐述了每个研究方向的基本内容、常用方法和研究现状. 作为一个相对新兴的、跨学科的研究领域,社会搜索的优势满足了用户的需求,各种社会化搜索引擎得到了迅速的普及. 然而目前社会搜索研究中还存在以下许多亟待解决的问题.

1) 缺乏对用户需求的语义理解. 由于用户需求中大量的不确定因素,单纯考虑用户需求中的关键字与网络资源的匹配程度是不合理的,需要了解用户需求的具体语义,进一步去匹配过滤.

2) 不能实时地获取信息. 现有的工作是定期从OSNs站点爬取数据资源,而OSNs中用户生成的内容是随时间变化的,传统方法获取的信息具有一定的滞后性.

3) 缺乏统一的用户模型. 用户上下文知识构建是社会搜索模型中必需的组成成分,目前缺乏统一的模型来建模用户上下文知识.

4) 缺乏统一的评价标准. 由于社会搜索更强调个性化结果,传统搜索引擎的评价标准已不再适用. 这就需要一个新的,能够反映用户对搜索结果主观

态度的统一的评价标准.

针对社会搜索的研究现状，未来研究中值得关注的方向有以下 2 个方面.

1) 移动化. 随着科技的高速发展，手机已经成为信息传递的主要设备之一，利用手机上网已经成为人们获取信息资源的主流方式. 近年来随着社会化媒体的大量涌现，基于手机 GPS 的商业搜索应用得到了快速的发展. 许多移动终端业务支持用户创建、维护并加强社会关联，这些业务能够通过共享信息或资源来帮助用户建立有用的社交网络[75]. 因此，将社会搜索与移动应用融合起来，一方面，在社会搜索中加入移动上下文，在搜索过程中挖掘手机的不同应用，将这些应用与现有的手机功能结合起来，设计基于移动场景的社会化搜索；另一方面，在移动搜索中融入社会化元素[76]. 由于手机设备的处理能力较弱，对带宽和屏幕显示也有一定的限制，使得移动搜索对精准性和个性化的要求更高. 因此，在利用移动搜索普适计算特性的同时，分析学习用户在其社会网络中的特征与行为，能够为用户提供更加实时、精确的个性化结果[77].

2) 语义化. 用户习惯于使用自然语言模式进行信息检索，网络语言使用灵活，不遵循严格的用词和语法规则，自动分析的难度很大[78]. 因此，如何有效地分析查询语义，找到用户所需的信息是社会搜索中的关键问题. 语义信息检索是解决这一问题非常有潜力的方法. 在社会搜索中采用语义分析和推理技术，对用户各种形式的查询输入进行分析、理解与再组织，再利用语义关联分析技术发现用户感兴趣的资源之间的关联，能够提高社会搜索引擎的查准率与查全率. 因此，将语义检索技术和社会网络信息结合起来，为用户提供个性化的语义搜索服务，成为社会搜索未来的发展趋势之一.

参考文献：

[1] Milgram S. The small world problem [J]. Psychology Today, 1967, 1(1): 60-67.

[2] Watts D J, Strogatz S H. Collective dynamics of small-world networks [J]. Nature, 1998, 393(6684): 440-442.

[3] Dodds P, Muhamad R, Watts D J. An experimental study of search in global social networks [J]. Science, 2003, 301(5634): 827-829.

[4] Kleinberg J. Navigation in a small world [J]. Nature, 2000, 406(6798): 845-845.

[5] Kleinberg J. The small-world phenomenon: an algorithmic perspective [C]//STOC 2000. Portland: ACM, 2000: 163-170.

[6] Watts D J, Dodds P S, Newman M E J. Identity and search in social networks [J]. Science, 2002, 296: 1302-1305.

[7] Leskovec J, Horvitz E. Planetary-scale views on a large instant-messaging network [C]//WWW 2008. Beijing: ACM, 2008: 915-924.

[8] Kleinberg J. The convergence of social and technological networks [J]. Commun ACM, 2008, 51(11): 66-72.

[9] Shadbolt N, Berners-Lee T. Web science emerges [J]. Scientific American, 2008, 299(4): 76-81.

[10] Dhyani D, Ng W K, Bhowmick S S. A survey of web metrics [J]. ACM Comput Surv, 2002, 34(4): 469-503.

[11] Mao Yuqing, Shen Haifeng, Sun Chengzheng. Google + facebook: a social-network-optimized web search approach [C]//iUBICOM 2011. Newcastle: British Computer Society, 2011: 1-8.

[12] Horowitz D, Kamvar S. The anatomy of a large-scale social search engine [C]//WWW 2010. Raleigh: ACM, 2010: 431-440.

[13] Teevan J, Ramage D, Morris M R. Twitter Search: a comparison of microblog search and web search [C]//WSDM 2011. Hong Kong: ACM, 2011: 35-44.

[14] Evans B M, Chi E H. Towards a model of understanding social search [C]//CSCW 2008. San Diego: ACM, 2008: 485-494.

[15] Goel S, Muhamad R, Watts D. Social search in "small-world" experiments [C]//WWW 2009. Madrid: ACM, 2009: 701-710.

[16] Barabasi A L, Albert R. Emergence of scaling in random networks [J]. Science, 1999, 286(5439): 509-512.

[17] Mislove A, Marcon M, Gummadi P K, et al. Measurement and analysis of online social networks [C]//IMC 2007. San Diego: ACM, 2007: 29-42.

[18] Kumar R, Novak J, Tomkins A. Structure and evolution of online social networks [C]//KDD 2006. Philadelphia: ACM, 2006: 611-617.

[19] Ahn Y Y, Han S, Kwak H, et al. Analysis of topological characteristics of huge online social networking services [C]//WWW 2007. Banff: ACM, 2007: 835-844.

[20] Backstrom L, Huttenlocher D P, Kleinberg J, et al. Group formation in large social networks: membership, growth, and evolution [C]// KDD 2006. Philadelphia:

ACM, 2006: 44-54.

[21] Magnani M, Montesi D, Rossi L. Conversation retrieval for microblogging sites [J]. Information Retrieval, 2012, 15(3-4): 354-372.

[22] Granovetter M. The strength of weak ties [J]. The American Journal of Sociology, 1973, 78(6): 1360-1380.

[23] Granovetter M. Getting a job [M]. Cambridge: Harvard University Press, 1974.

[24] Adamic L, Lukose R, Puniyani A, et al. Search in power-law networks [J]. Physical Review E, 2001, 64(046135): 1-8.

[25] Adamic L, Adar E. How to search a social network [J]. Social Networks, 2005, 27(3): 187-203.

[26] Simsek O, Jensen D. Decentralized search in networks using homophily and degree disparity [C]//IJCAI 2005. Edinburgh: Morgan Kaufmann Publishers Inc, 2005: 304-310.

[27] Imsek O J D. Navigating networks by using homophily and degree [J]. PNAS, 2007, 105(35): 12758-12762.

[28] Banerjee A, Basu S. A social query model for decentralized search [C]//SNA-KDD 2008. Las Vegas: ACM, 2008: 63-72.

[29] Davitz J, Yu J, Basu S, et al. ILink: search and routing in social networks [C]//SIGKDD 2007. San Jose: ACM, 2007: 931-940.

[30] Daly E, Haahr M. Social network analysis for routing in disconnected delay-tolerant MANETs [C]//MobiHoc 2007. Montreal: ACM, 2007: 32-40.

[31] Costa P, Mascolo C, Musolesi M. Socially-aware routing for publish-subscribe in delay-tolerant mobile Ad hoc networks [J]. IEEE Journal on Selected Areas in Communications, 2008, 26(5): 748-760.

[32] Xu Zhichen, Fu Yun, Mao Jianchang, et al. Towards the semantic web: collaborative tag suggestions [C]// WWW 2006. Edinburgh: ACM, 2006: 36-43.

[33] Cattuto C, Schmitz C, Baldassarri A, et al. Network properties of folksonomies [J]. AI Communications, 2007, 20(4): 245-262.

[34] Bao Shenghua, Xue Guirong, Wu Xiaoyuan, et al. Optimizing web search using social annotations [C]// WWW 2007. Banff: ACM, 2007: 501-510.

[35] Heymann P, Koutrika G, Molina H G. Can social bookmarking improve web search [C]//WSDM 2008. Palo Alto: ACM, 2008: 195-206.

[36] Rawashdeh M, Kim H N, Saddik A E. Folksonomy-boosted social media search and ranking [C]//ICMR 2011. Trento: ACM, 2011: 1-8.

[37] Kim H N, Rawashdeh M, Alghamdi A, et al. Folksonomy-based personalized search and ranking in social media services [J]. Information Systems, 2012, 37(1): 61-76.

[38] Hotho A, Jaschke R, Schmitz C, et al. Information retrieval in folksonomies: search and ranking [J]. The Semantic Web: Research and Applications, 2006, 4011(2006): 411-426.

[39] Krause B, Hotho A, Stumme G. A comparison of social bookmarking with traditional search [C]// Proceedings of the IR research. Glasgow: Springer-Verlag, 2008: 101-113.

[40] Golder S, Huberman B. The structure of collaborative tagging systems [J]. Journal of Information Science, 2006, 32(2): 198-208.

[41] Markus S. Purpose tagging: capturing user intent to assist goal-oriented social search [C]//SSM 2008. Napa Valley: ACM, 2008: 35-42.

[42] Wang Yufeng, Nakao Akihiro, Ma Jianhua. Socially-inspired search and ranking in mobile social networking: concepts and challenges [J]. Journal of Frontiers of Computer Science (FCS), 2009, 3(4): 435-444.

[43] Wang Yufeng, Nakao Akihiro, Vasilakos A V. Doubleface: robust reputation ranking based on link analysis in P2P networks [J]. Cybernetics and Systems, 2010, 41(2): 167-189.

[44] Girvan M, Newman M. Community structure in social and biological networks [J]. PNAS, 2002, 99(12): 7821-7826.

[45] Girvan M, Newman M. Finding and evaluating community structure in networks [J]. Physical Review E, 2004, 69(2): 026113-026128.

[46] Newman M. Fast algorithm for detecting community structure in networks [J]. Physical Review E, 2004, 69(6): 066133-066138.

[47] Clauset A, Newman M, Moore C. Finding community structure in very large networks [J]. Physical Review E, 2004, 70(066111): 1-6.

[48] Li Xin, Liu Bing, Yu P S. Mining community structure of named entities from web pages and blogs [C]//AAAI 2006. Boston: AAAI Press, 2006: 108-114.

[49] Du Nan, Wu Bin, Pei Xin, et al. Community detection in large-scale social networks [C]//SNA-KDD 2007. San Jose: ACM, 2007: 16-25.

[50] Peter B, Barry S. Trusted search communities [C]//

IUI 2007. Honolulu: ACM Press, 2007: 337-340.

[51] Agichtein E, Castillo C, Donato D, et al. Finding high-quality content in social media [C] //WSDM 2008. New York: ACM, 2008: 183-194.

[52] Haynes J, Perisic I. Mapping search relevance to social networks [C] //SNA-KDD 2009. Paris: ACM, 2009: 1-7.

[53] Moody J. Peer influence groups: identifying dense clusters in large networks [J]. Social Networks, 2001, 23(4): 261-283.

[54] Ahn Y, Bagrow J P, Lehmann S. Link communities reveal multiscale complexity in networks [J]. Nature, 2010, 466(7307): 761-764.

[55] Domavicius A, Tuzhilin G. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions [J]. IEEE Transaction on Knowledge and Data Engineering, 2005, 17(6): 734-749.

[56] Pazzani M J, Billsus D. Content-based recommendation systems [J]. The Adaptive Web, 2007, 4321(2007): 325-341.

[57] Sarwar B, Karypis G, Konstan J, et al. Item-based collaborative filtering recommendation algorithms [C] //WWW 2001. Hong Kong: ACM, 2001: 285-295.

[58] Hofmann T. Collaborative filtering via gaussian probabilistic latent semantic analysis [C] //SIGIR 2003. Toronto: ACM, 2003: 259-266.

[59] Marlin B. Modeling user rating profiles for collaborative filtering [C] //NIPS 2003. Vancouver: MIT Press, 2004: 627-634.

[60] Shani G, Brafman R, Heckerman D. An MDP-based recommender system [J]. Machine Learning Research, 2005, 6: 1265-1295.

[61] Zhou Tao, Ren Jie, Medo M, et al. Bipartite network projection and personal recommendation [J]. Phys Rev E, 2007, 76(046115): 1-7.

[62] Zhou Tao, Jiang Luoluo, Su Riqi, et al. Effect of initial configuration on network-based recommendation [J]. Europhys Lett, 2008, 81(58004): 1-4.

[63] Kautz H, Selman B, Shah M. Referralweb: combining social networks and collaborative filtering [J]. Communications of the ACM, 1997, 40(3): 63-65.

[64] Schenkel R, Crecelius T, Kacimi M, et al. Social wisdom for search and recommendation [J]. IEEE Data Eng Bull, 2008, 31(2): 40-49.

[65] McNally K, O'Mahony M P, Smyth B, et al. Towards a reputation-based model of social web search [C] //IUI 2010. Hong Kong: ACM, 2010: 179-188.

[66] McNally K, O'Mahony M P, Coyle M, et al. A case study of collaboration and reputation in social web search [J]. TIST, 2011, 3(1): 100-103.

[67] Guy I, Zwerdling N, Carmel D, et al. Personalized recommendation of social software items based on social relations [C] //RecSys 2009. New York: ACM, 2009: 53-60.

[68] Guy I, Zwerdling N, Ronen I, et al. Social media recommendation based on people and tags [C] //SIGIR 2010. Geneva: ACM, 2010: 194-201.

[69] Guy I, Carmel D. Social recommender systems [C] //WWW 2011. Hyderabad: ACM, 2011: 283-284.

[70] Nagpal A, Hangal S, Joyee R R, et al. Friends, romans, countrymen: lend me your URLs using social chatter to personalize web search [C] //CSCW 2012. Seattle: ACM, 2012: 461-470.

[71] Zhou D, Lawless S, Wade V. Improving search via personalized query expansion using social media [J]. Information Retrieval, 2012, 15(3-4): 218-242.

[72] Teevan J, Dumais S T, Horvitz E. Personalizing search via automated analysis of interests and activities [C] //SIGIR 2005. Salvador: ACM Press, 2005: 449-456.

[73] Carmel D, Zwerdling N, Guy I, et al. Personalized social search based on the user's social network [C] //CIKM 2009. Hong Kong: ACM, 2009: 1227-1236.

[74] Wang Q H, Jin H X. Exploring online social activities for adaptive search personalization [C] //CIKM 2010. Toronto: ACM, 2010: 999-1008.

[75] Ellison N, Steinfield C, Lampe C. The benefits of facebook "friends:" exploring the relationship between college students' use of online social networks and social capital [J]. Journal of Computer-Mediated Communication, 2007, 12(3): 33-36.

[76] Church K, Neumann J, Cherubini M, et al. Social Search Browser: a novel mobile search and information discovery tool [C] //IUI 2010. Hong Kong: ACM, 2010: 101-110.

[78] Ricci F. Mobile recommender systems [J]. Information Technology and Tourism, 2011, 12(3): 205-231.

[79] Pang B, Lee L. Opinion mining and sentiment analysis [J]. Foundations and Trends in Information Retrieval, 2008, 2(1-2): 1-135.