

**毕业论文**

（本 科 生）

**论文题目（中文） 基于话题组织与社交网络的知识社区兴趣发现**

**论文题目（外文） Topic Hierarchy and Social Network Based Interest**

**Discovery on a Knowledge Sharing Community**

**学 生 姓 名**  **沈碧螺**

**导师姓名、职称**  **张军 讲师**

**学生所属学院**  **管理学院**

**专 业**  **信息管理与信息系统**

**年 级**  **2014级**

兰州大学教务处

**基于话题组织与社交网络的知识社区兴趣发现**

**摘 要**

传统推荐算法如协同过滤基于内容、用户与标签进行推荐，但受限于标签信息存在的诸多问题，对基于标签进行推荐的方法的研究还不够充分。本文首先讨论了知识社区的组成，然后提出可以基于知识社区中的话题组织，作为标签体系的延伸，从而构建用户兴趣图，再基于用户关注的人的影响力，衡量其对用户的影响权重，根据权重进行兴趣图融合，从而发现用户潜在兴趣，进行推荐。此外，本文以知乎社区为例，使用真实数据与实例，详细地展示了方法的每个步骤，讨论了在生产环境中基于该方法构建推荐系统时需要注意的问题。

**关键词：**社交网络，推荐系统，数据挖掘

**TOPIC HIERARCHY AND SOCIAL NETWORK BASED INTEREST DISCOVERY ON A KNOWLEDGE SHARING COMMUNITY**

**Abstract**

Traditional recommendation algorithms like collaborative filtering are based on content, users and labels. Limited to low quality of labels, studies on label-based recommendation methods is not enough. In this paper, we first discuss the components of a knowledge sharing community, then propose that we can construct interest graph of users based on topic hierarchy of the community which can be seen as an extension of labels, merge all the interest graphs of followees of a user according to influence of followees, discover potential interest and do recommendation. In addition, taking Zhihu for example, we perform every steps of our algorithm using real world data and discuss some questions of building a recommendation system in production.

**Key words:** social network, recommendation system, data mining

# 1 引言

随着互联网的不断发展，移动互联网的兴起，微博、朋友圈、贴吧等社交平台占据了人们的生活。许多用户开始感到疲倦，从而产生了社交之外的需求，其中就包括自我提升的需求。传统的知识问答平台如百度知道等，由于缺乏有效的管理，并不能满足新时代用户的需求。而类似知乎、果壳、Quora等专业性更强、问答质量更高的知识分享平台，则开始受到用户的欢迎，可以称之为知识社区。知识社区包含了一套完善的话题组织方式，整体形成一张有向无环图，可以看做以往标签体系的延伸。而如何利用这些额外信息，扩展传统的推荐算法，进行更好的内容推荐，是一个亟待解决的问题。

从经典的协同过滤发展到现代推荐算法，资源(Resource)、用户(User)与标签(Label)一直是核心关注点，由此产生了基于资源（内容）的方法，基于用户（社交关系）的方法与基于标签的方法。Jinyu Hu等总结了这些观点，将其概括为6个角度，包括UBR(user-based algorithm from resource point)，UBT，RBT，RBU，TBR，TBU等[1]。除此之外，相似度计算在协同过滤中发挥了重要作用，主要包括皮尔逊相似度，余弦相似度和斯皮尔曼等级相关等。现在，协同过滤中用户相似度的计算一般从资源角度出发[2]。Frank等提出了一个在社交网络中基于用户信任度的推荐模型，为如何利用用户之间的关系创造了一种全新的视角[3]。在基于标签的方法上，De Gemmis等提出可以利用社会化标签信息来描述用户的兴趣特征[4]。Bischoff等提出了社会化标签存在的许多问题，并通过案例研究发现超过50%的标签都是无用的[5]。在网络结构的研究上，唐晓波等提出了一种基于混合图的随机游走方法，通过建立融合用户、资源及其关系的混合图，采用重启动随机游走算法，进行相似度计算，得到用户推荐列表[6]。Liu Y等总结了近年来图摘要理论与应用的发展[7]。这些为本文基于图与社交网络的内容推荐的理论与应用研究奠定了基础。

综上所述，现代推荐系统从多角度利用了用户、资源与标签信息，但在标签信息的使用上不够彻底，而且遇到了许多如标签质量过低、标签关系混乱等问题。而知识社区中的话题组织，则不存在社会化标签的诸多问题，同时天然形成一张结构化的有向无环图，可以很方便地进行研究。本文以知乎为例，通过研究知识社区的组成，发现与用户的兴趣相关的要素，然后综合考虑这些要素，计算得到用户兴趣图，通过一定的规则，结合用户关注的人的兴趣图，得到用户的潜在兴趣，进行推荐。

# 2 知识社区的组成

## 2.1 概述

知识社区主要由用户、话题、问题、回答与文章等要素组成。这些要素之间存在复杂的显性关系与隐形关系。如用户与用户之间有直接的关注与被关注关系，也存在不明显的点赞回答、收藏文章等关系。知乎的简化实体关系图如图1。下面将具体讨论这些要素和关系，以及它们与用户兴趣之间的关系。

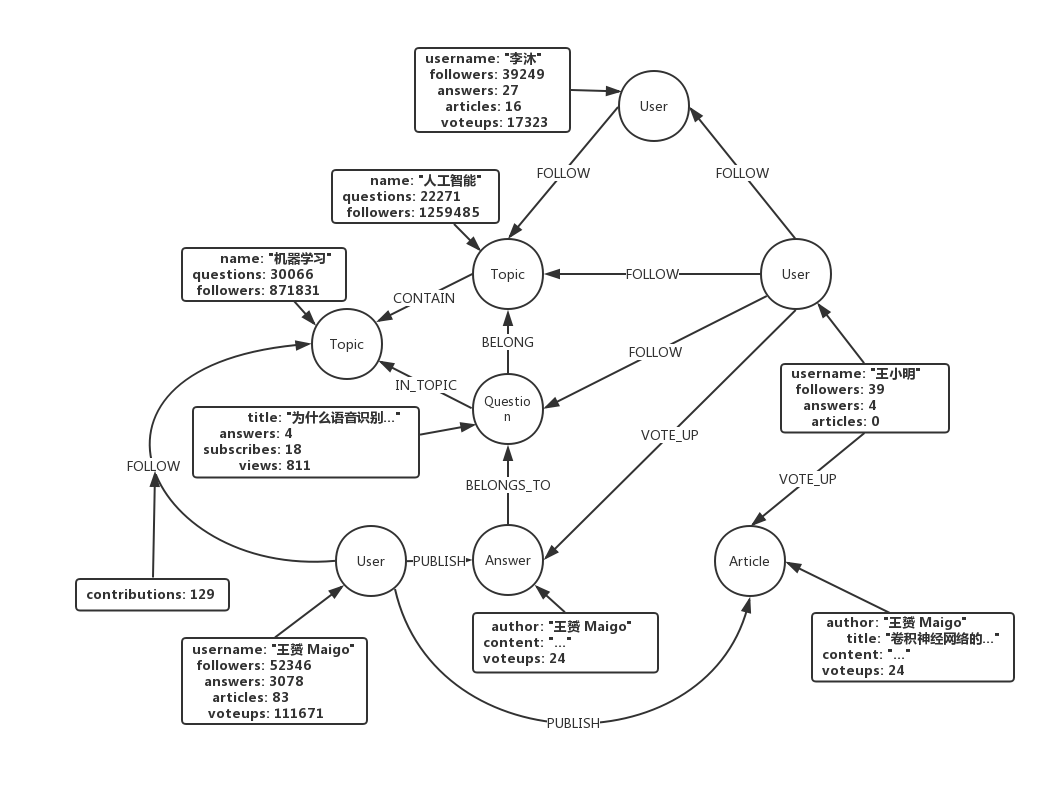


图2.1 知识社区简化关系图

## 2.2 话题

话题即知识社区中问题的分类，任何问题都必须属于一个及以上的话题。在知乎社区，全部话题通过父子关系构成一个有根无循环的有向图。「根话题」即为所有话题的最上层的父话题。不能确定话题的会归到「未归类」话题。节点之间可能存在跨级连接，如知识管理是管理学的子级话题，信息管理也是管理学的子级话题，同时知识管理是信息管理的子级话题。

话题与问题、用户之间的关系，使话题节点包含了许多额外的属性（Attributions），这些属性一定程度上体现了话题的活跃程度。用户可以关注话题，所以话题的属性包括关注数，如知乎的管理学话题有110,772人关注，但同时管理学的子级话题知识管理有132,140人关注。这引出了父子话题的属性继承问题。在关注数这一属性上，不继承是合理。如果父级话题继承子级话题的关注数，那么越宽泛的话题关注数就越多，这显然是不合理的。关注宽泛话题的人，对宽泛的问题感兴趣，关注专业话题的人，则可能只对小领域的专业性强的问题感兴趣，而关注数不继承则更能体现这种倾向。问题必须属于一个及以上的话题，属于话题的属性还包括问题数。如知乎的信息管理话题包含16,162个问题，管理学话题包含267,019个问题，这说明问题数属性是继承属性。而这又引出了父子话题的关系继承问题。在问题属于话题这个关系上，该问题表现为：一个问题属于某话题，那么该问题是否也属于这个话题的父级话题。知乎的默认策略是继承，但同时提出了弱关系的概念。在话题中可以设置与某个父级话题是否是弱关系，默认为否，如果为是，则该话题的问题不会出现在父级话题中。

话题的问题数，关注数在一定程度上体现了话题的活跃程度。而话题的活跃程度与用户对话题的感兴趣程度存在显著的正相关，如微博热搜，百度搜索热点等都由此而生。同时话题的活跃程度越高，说明该话题下有越多的内容可以进行推荐。但如果只考虑话题活跃程度，不考虑用户之间的差异，则效果并不会很好。特别在知识社区，用户对于知识的需求有更强的专业性，泛化的热门话题完全不能满足这种需求。所以在话题的活跃程度之外，还需要考虑其他更加能体现个性化与专业性的要素。可以发现，用户回答某个问题，而这个问题属于某个话题，所以用户在该话题贡献了一个回答。这在一定程度上可以体现用户对该话题的兴趣，同时表明了用户在该话题领域的专业性。所以在计算用户对话题的感兴趣程度时也将考虑这个要素。

话题组织可以看做标签体系的延伸，解决了标签存在的许多问题，但仍然有一些问题无法避免。社会化标签非常稀疏，存在大量的同义词，无用词，表意不明的词，这些问题通过知识社区的专业用户共同编辑与平台专门的管理得到了基本的解决，但即使没有这些低质量的标签干扰，话题图仍然比较稀疏。图3为以信息管理话题为根，遍历其所有下级节点得到的话题图。信息管理有132,140名关注者，包含10,000多个问题，但有112个下级话题，其中只有15个话题拥有超过10,000个关注者，剩下的话题关注者分布如图3，其中又只有12个话题有超过2,000个问题，剩下的话题包含问题数分布如图4。可以通过图摘要算法，合并某些话题节点，从而一定程度上缓解这个问题。

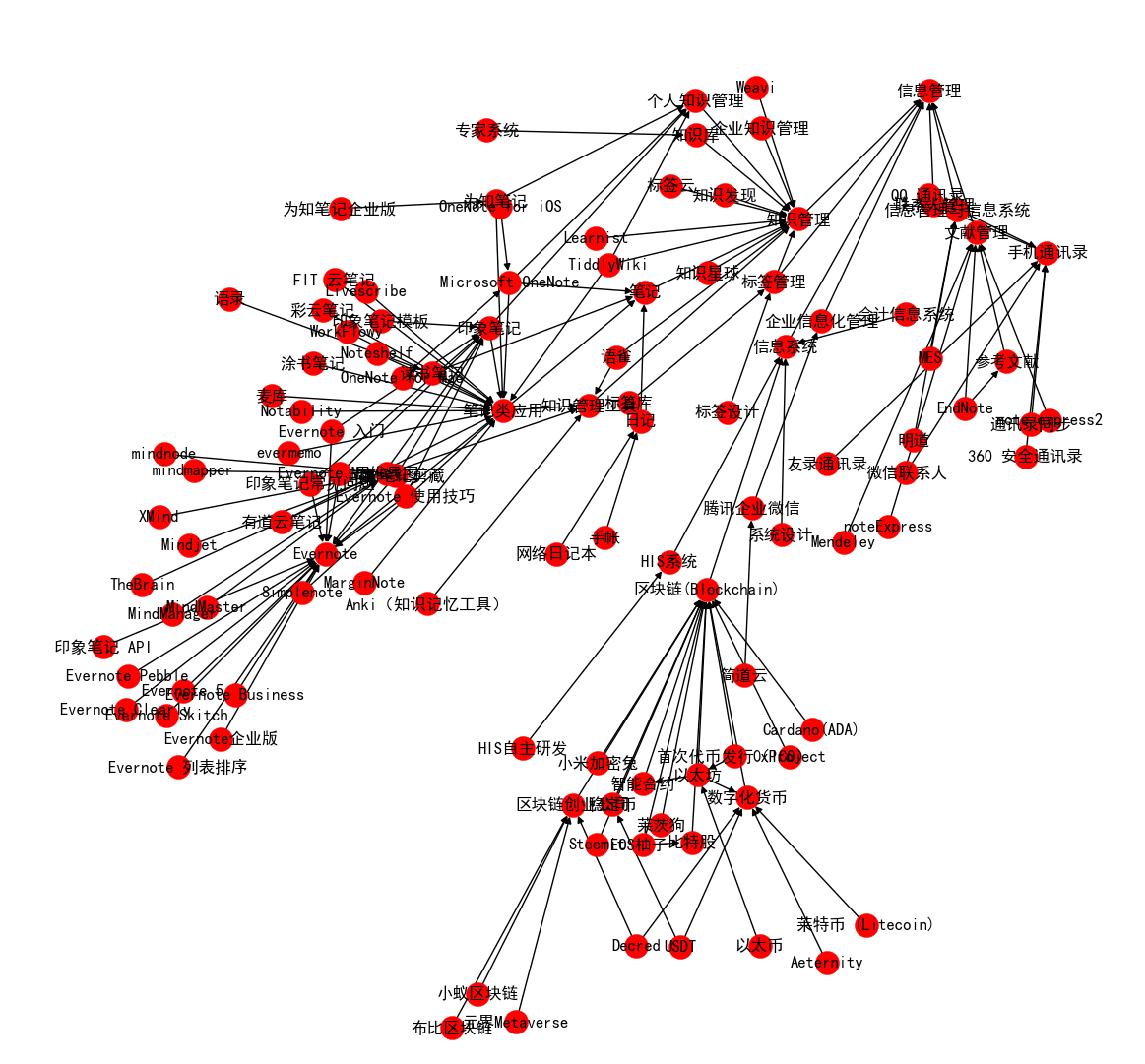


图2.2 信息管理话题组织

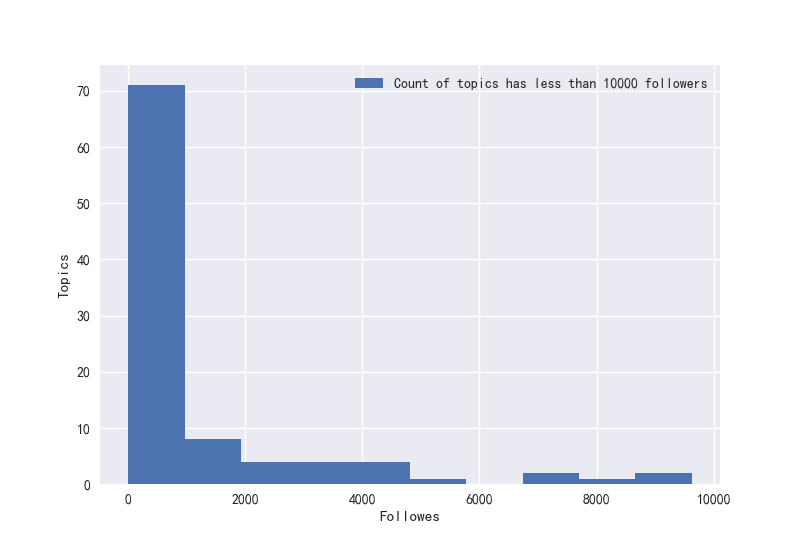


图2.3 信息管理话题中关注者低于10,000的下级话题分布

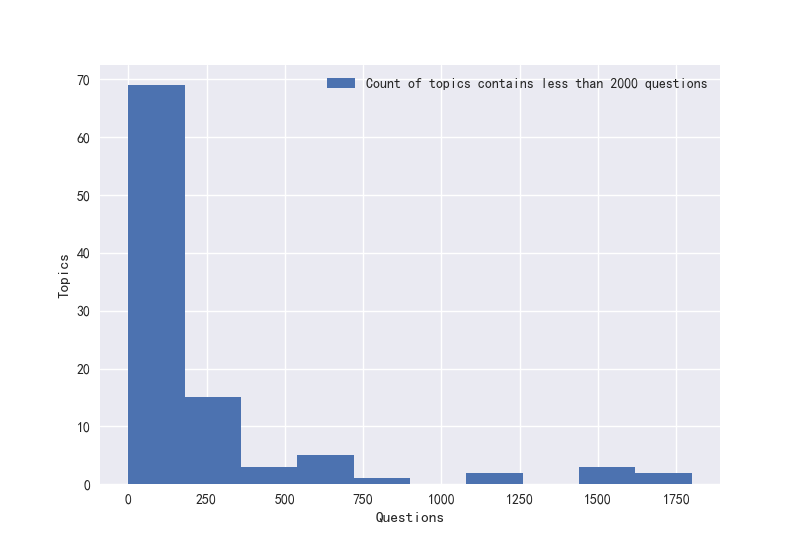


图2.4 信息管理话题中包含问题数低于2,000的下级话题分布

## 2.3 问题与回答

问题与回答是知识社区的核心。问题由用户提出，隶属于一个或多个话题，包含了标题与内容。知乎的问题是开发编辑的，即问题提出后，任何人都可以更改问题隶属的话题，标题及描述。回答是用户对于问题的答案，没有标题，只有内容。其他用户可以赞同、感谢和收藏回答，还可以针对回答发表评论。用户可以关注问题，在关注之后，每当这个问题有新的回答，就将通知用户。通知可以理解为高一级的推荐，对用户的影响更大，但如果用户不感兴趣，造成的负面影响也越大。

用户提出了隶属于某个话题的问题，或者回答了这个问题，或者赞同、感谢和收藏了该问题的回答，都表现了用户对于这个问题及问题隶属的话题的兴趣。但其体现的程度是有所区别的。提出问题一般代表了最大程度的兴趣，用户希望对这个问题及其所属的话题有深入了解。这个问题的动态将得到用户的高度关注，相关问题的内容也将得到用户的较大关注。这是基于内容（资源）的推荐方法的一部分，但是本文更加关心的是用户关注某个话题的问题的行为对于用户对话题的感兴趣程度的补充。因为用户的使用习惯相差很大，有一部分用户只会关注很少的话题，但会关注大量的问题。在这种情况下，用户关注的主题就不能刻画用户兴趣，必须使用用户关注的问题的主题加以补充。对于问题的回答的赞同、感谢和收藏同理，所以在必要的时候，我们也会考虑这些要素。

要通过问题与回答刻画用户兴趣，它们各自之间就必须有一定的区分度。一个无人问津的问题很快就会沉没在问题的海洋中，而一个广受关注的问题每天都会有新的回答产生。用户对这两者的关注程度是不一样的。问题的属性包括被浏览数，关注数，回答数，提出时间，隐性的属性还可能有所有回答总赞同数等，可以综合考虑这些要素，从而得出问题的重要程度。

## 2.4 用户

用户是知识社区的主体，是问题的提出者，是答案的贡献者，是话题的编辑者。用户与用户凭借显性的关注与被关注关系形成一个社交网络。对于某个用户来说，他关注的人对用户有很大影响，而关注他的人对用户影响较小。Frank等提出用户信任度的概念，用户对关注的人有较大的信任度。所有，我们认为用户对关注的人关注的话题也有较大的兴趣。这与传统的基于好友和基于用户相似度的方法是截然不同的。

关注的人之间也要有区分度，不同的人对用户的影响程度是不同的。影响力较大的名人，大V对用户的影响较大，反之则影响较小。与用户影响力相关的属性包括关注数，回答问题数，获得的赞同数，感谢数，回答被收藏数，以及是否是某个话题的优秀回答者，回答被官方推荐次数等。例如李开复博士是知乎的用户之一，他有1,049,377个关注者，回答过109个问题，总共获得118,906次赞同，26,640次感谢，回答被收藏81,057次，10个回答被官方推荐。毫无疑问，李开复博士是知乎上最有影响力的人物之一，关注了他的用户肯定对他的一言一行非常感兴趣，同时也有很大的可能对他关注的话题与问题感兴趣。Vczh是知乎社区的大V，主要活跃于计算机科学话题，他有712,531个关注者，回答过19,129个问题，总共获得2,111,672次赞同，235,419次感谢与315,863次收藏，4个回答被官方推荐。可以看出，赞同、感谢与收藏的数量与回答数有很大关系，所以在计算影响力的时候，需要综合考虑这两个因素。如果一个用户同时关注了李开复与vczh，有理由相信李开复对他的影响会更大，即使他原来主要关注计算机科学话题。

# 3 推荐模型与算法

## 3.1 图数据库

在传统的推荐模型中，数据一般存储于关系数据库中。在准备数据时，程序员根据业务逻辑，编写并执行SQL语句，从数据库中提取数据。但是从图1的简化关系图中可以发现，实体之间的关系非常复杂，而且经常需要根据多个关系才能得到某些需要的特征。要完成这些需求，需要编写复杂的SQL语句，维护这些SQL也是一个巨大的负担。造成这种现象的主要原因是，我们想要表达的是图模型，实体与关系，或者说节点与边，都非常重要。所以，我们转向一种新的、更加适合现在的场景的数据库：图数据库

图数据库是一个基于图数据模型的，带有增删查改（CRUD）功能的实时数据库管理系统。不同于其他数据库，关系在图数据库中是第一位的。这意味着应用程序不必使用外键或如MapReduce之类的外部处理去推断数据之间的关系。图数据库的数据模型同时也比其他关系型数据库或者NOSQL数据库更简单，更具表现力。[8]

Cypher是一种类似于SQL的声明式图查询语言，它使我们可以更加高效地查询与更新图数据。Cypher本来由Neo4j公司为neo4j图数据库创造，在2015年通过openCypher开源，现在也被其他图数据库如SAP HAHA和AgensGraph支持。

下面将使用neo4j为数据存储，使用Cypher表达数据查询逻辑，使用伪代码表达计算逻辑，来具体讨论推荐模型与算法的实现。

## 3.2 兴趣图构建

全部话题通过父子关系构成一张有根的有向无环图。用户关注的话题同样可以用图来表示，如果这些话题之间存在父子关系，则话题节点之间会有边，由子话题指向父话题。在图模型中，所有的实体都是节点。所以可以把用户看做节点，用户与话题之间存在多重边。例如用户A关注了话题T，则存在一条边从用户节点A指向话题节点T，边上的属性可能包括关注时间。如果用户A回答过问题Q，而问题Q属于话题T，则可以认为用户节点A指向话题节点T的边上有另一个属性：贡献答案数。显然，这个属性是可以由用户节点、问题节点与话题节点之间的关系得到，属于冗余属性。但一方面我们认为这个属性的计算比较昂贵，另一方面我们认为它非常重要，所以进行可以提前计算，存在图数据库中。

依据前面对各要素的讨论，与用户兴趣相关的主要包括关注的话题的相对影响力和用户对话题的贡献度。话题相对影响力主要依赖于包含问题数与关注数，计算公式如下，

(1)

其中表示话题T的相对影响力，表示话题T的包含问题数，表示话题T的关注数，表示相对系数。用户对话题的贡献度体现在贡献答案数。而这个属性并不能简单地平均。大部分用户在关注的大部分话题领域没有贡献任何回答，少部分贡献了1个回答，贡献多余10个回答的非常罕见。所以对其进行一定的变换，关键点转换表如表3.1。当贡献数大于20时，使用线性变换。

表3.1 贡献回答数转换表

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 回答数 | 0 | 1 | 5 | 10 | 20 | 30 |
| 转换值 | 1 | 5 | 15 | 25 | 40 | 50 |

通过多项式回归拟合这些关键点，可以得到变换函数g，如图3.1所示。

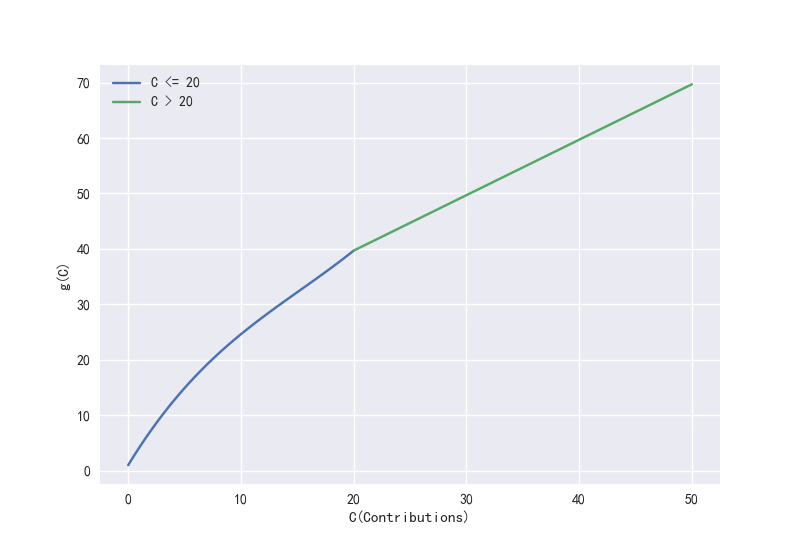


图3.1 贡献回答数变换函数g

变换之后，再对这些值进行归一化（Normalize）得到，

(2)

然后汇总与，得到话题兴趣度，

(3)

在生产环境中，因为用户对于话题的兴趣在短时间不会发生很大变化，所以建议通过离线批处理得到话题兴趣度，然后存储在图数据库的话题节点与用户节点的边上。而在兴趣图的构建过程中，需要的是一个图结构，所以抽取出一张只包含用户关注的话题节点的子图，将话题兴趣度作为相应话题节点的属性。如果用话题节点大小表示兴趣度，以用户Scott为例，其兴趣图可视化如图3.2。

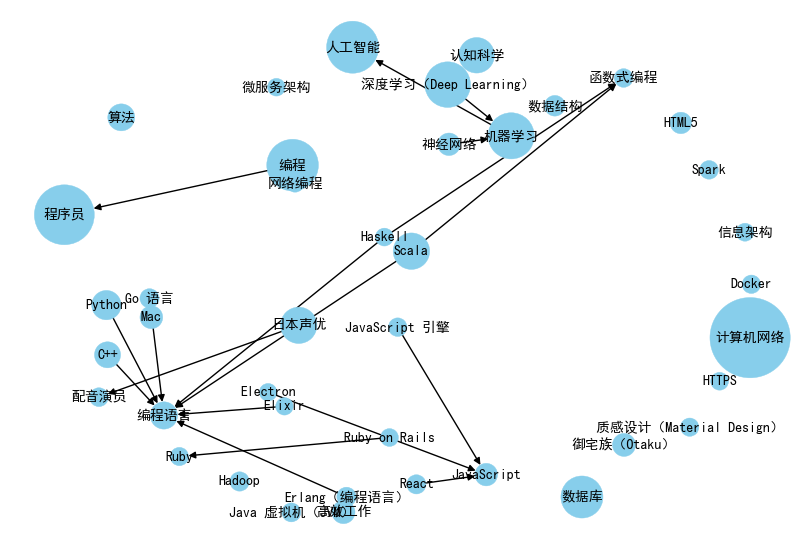


图3.2 用户Scott兴趣图可视化结果

通过兴趣图，可以很直截了当地发现该用户的职业是程序员，对多种编程语言都很感兴趣，同时对人工智能方向很感兴趣，另外业余还对日本声优与配音演员感兴趣。

另外考虑一种上文提到过的极端情况，即用户没有关注任何话题，但是关注了许多问题。这种现象可能是由用户特殊的使用习惯导致的，它确实是存在的。上面兴趣图的构建方法对于这类用户完全不适用，必须考虑其他方法。可以取用户关注的所有问题的所属话题，进行聚合，统计频次，然后归一化，以此作为话题兴趣度。



图3.3 用户关注问题所属话题聚合

以用户xigig为例，该用户没有关注任何话题，但是关注了120个问题。使用图3.3的Cypher语句，可以得到这些话题及对应的出现频次。结果显示，120个问题涉及到238个话题，其中频次大于1的话题仅有42个，大于5的话题仅有8个，去除频次为1的话题后可视化结果如图3.4。

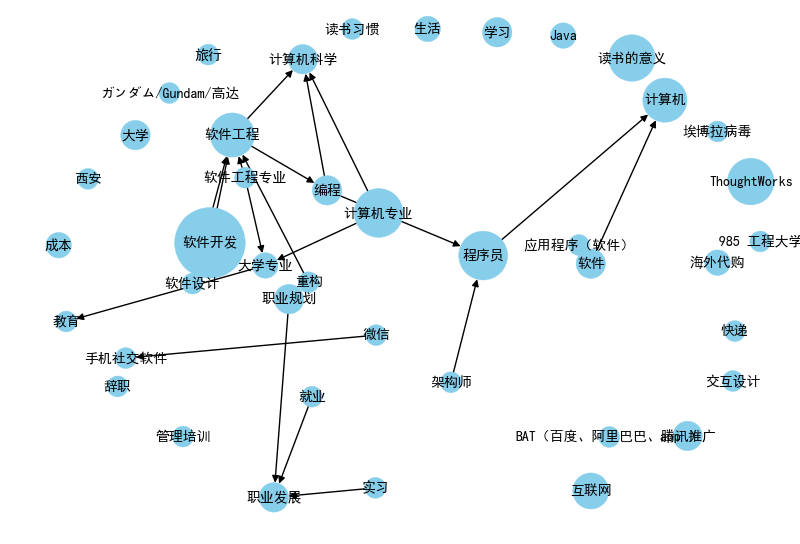


图3.4 用户xigig兴趣图可视化结果

## 3.3 兴趣图融合

通过上述方法，可以构建出用户及其关注的人的兴趣图，然后需要将用户关注的人的兴趣图进行融合，公式如下，

(3)

上式中表示融合图，n表示用户关注的人的数量，表示第i个关注的人的影响权重，表示第i个关注者的兴趣图。下面具体讨论的计算方法。

根据上面的讨论，关注的人U的影响力与U的关注数，回答问题数，获得的赞同数，感谢数，回答被收藏数，以及是否是某个话题的优秀回答者，回答被官方推荐次数等要素相关。其中U的关注数之外的要素都与回答直接相关，可以将其表示为，其公式如下，

(4)

上式中，，，，分别代表获得的赞同数，感谢数，回答被收藏数，回答问题数以及回答被官方推荐次数，，表示感谢数的相对系数和被收藏数的相对系数。因为用户赞同一个回答是比较随意的，而感谢是相对慎重的，收藏回答则说明这个回答对用户具有长远价值，所以在公式中加入相对系数。准确的相对系数取值需要对所有用户的赞同数，感谢数，收藏数进行统计分析得到，这里暂且先取为10，为5。表示补偿常数，为了防止出现在极端情况下用户只回答过1个问题而偶然收获高赞导致的过高的情况，这里暂且取为5。表示增加比率，这里取为0.1，因为回答被官方推荐可以充分表明该用户的影响力。对于关注数，我们不做任何变换。最终得到以下公式，

(5)

上式中表示用户U的关注数，可以简单地取为0.5，其他不予赘述。

下面讨论一种极端情况，用户A只关注了李开复与其他几个影响力相对小很多的用户，李开复的w超过了所有其他人的总和。这将导致最终的几乎退化为李开复的兴趣图，显然这是不太合适的。这时需要对关注数与F进行裁剪，使其归一化之后的比例不能大于一定的阈值。

从生产的角度看，上面的影响权重在短时间内不会发生很大变化，即不需要通过实时流计算得到，完全可以通过离线批处理预先计算，然后存储在图数据库的用户节点中。注意到，归一化之前的值仅与用户A本身相关，所以这些影响权重不需存储在用户与关注的人的边上。在需要时，将这些值取出，重新进行归一化即可。

再以用户Scott为例，他总共关注了18个人。在计算这些人的影响权重与兴趣图之后，进行融合，得到融合图。融合图有627个节点，其节点权重分布如图3.5。可以看到，虽然话题节点很多，但权重是相对集中的。权重大于1的话题节点有178个，权重总和占89.9%，权重大于5的话题节点有58个，权重总和占63.3%。若只包含这58个节点，可视化结果如图3.6。

对比图3.5与图3.6，可以发现新增的话题节点，即用户Scott以前没有关注的话题，主要包括健康、数学、英语口语，互联网等。我们认为向程序员推荐健康话题与互联网是比较合适的。而对于一个关注人工智能的程序员，数学与英语口语也应该是他比较感兴趣的。

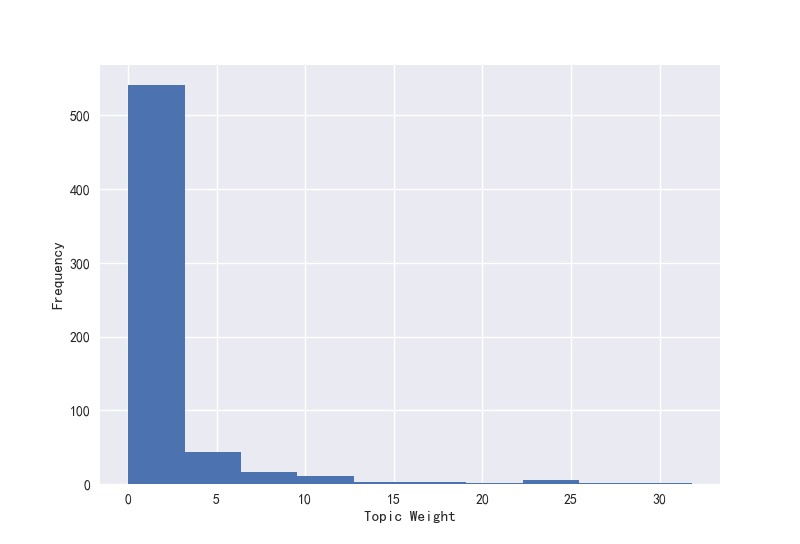


图3.5 融合图话题节点权重分布

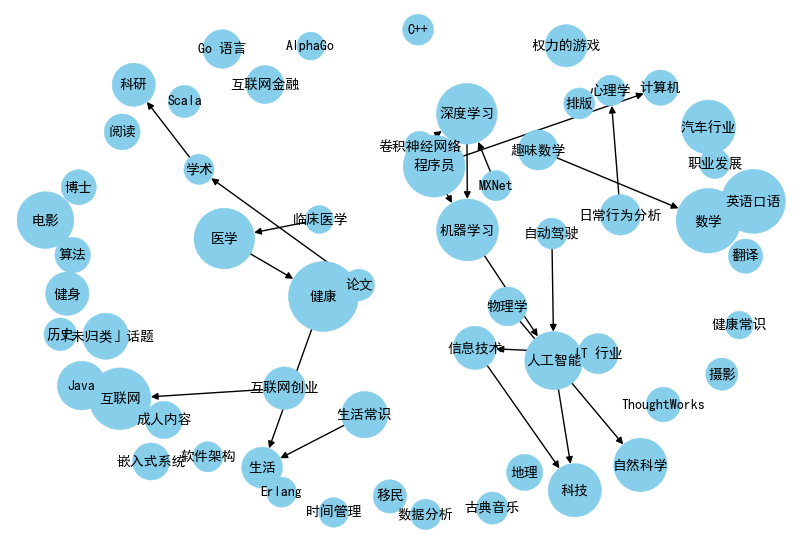


图3.6 融合图话题节点权重分布

从生产的角度看，因为融合图依赖的兴趣图和权重都可以通过离线批处理得到，那么显然融合图也可以通过这种方式生成，而无需实时计算。所以融合图的更新也依赖于兴趣图与权重的更新。我们可以对每个用户维护一个更新数，每当用户关注的话题发生变化，就对更新数进行自增。当更新数增加到一个阈值时，触发用户兴趣图更新，同时触发关注他的人的与权重相关的更新数的自增。这种方式一般称为响应式策略。

## 3.4 推荐策略

首先需要明确的是，基于话题图的推荐是一种补充方法，不能作为主要的推荐方法。典型的使用场景比如用户连续多次刷新尝试获取新的内容，而传统方法的推荐内容已全部推送完，这时可以使用基于融合图的推荐方法，推荐用户未关注但是很大概率感兴趣的话题的相关内容。

最直接的推荐策略就是根据融合图上话题节点的权重排序，然后依次推送这些话题下的活跃内容或精华内容。这种方法最简单，也应该能收到较好的效果。其次，我们可以根据话题节点的权重分布，从中采样，然后推荐。即权重大的话题有更高概率被推荐，权重小的话题也有小概率被推荐。这避免了前一种方法中小概率话题永远不会被推荐的问题。但权重小的话题一般是专业性较强或偏冷门的话题，推荐这种话题有一定的风险性。

最后简单讨论一种可能的方法。在Scott的例子中，权重小于5的话题节点的权重总和达到了30%，而以上两种策略都不能很好地使用这30%。所以可以对融合图使用图摘要方法，将权重小的节点合并，例如合并到其父话题，然后再根据以上策略进行推荐。在合并权重小的节点后，我们有很大概率得到新的、较为泛化的话题，此时风险性也随之降低。

# 4 结语

本文通过研究知识社区的组成，分析了知识社区中各要素与用户兴趣的相关性，讨论了话题组织作为标签体系的延伸的主要优点与存在的问题，然后根据这些理论基础，推导出用户兴趣图的构建方法，图融合方法，最后讨论了基于融合图的推荐策略。Scott例子中的融合图可以表明，基于融合图的方法确实能够发现用户的潜在兴趣，退一步说，也是对用户关注话题的补充。

然而本文提出的方法仍然存在许多问题，需要进一步研究。本文提出的诸多公式都包含了参数，而这些参数该如何取值，是否存在最优解，都是尚不明确的。包括这些公式的形式是否合适，都未经大规模实验或应用检验。本文许多针对生产环境的意见，都是基于主观的工程经验，也需要真正经过生产环境的检验，才能明确其效果。

# 参考文献

[1] Hu J, Gao Z, Pan W. Multiangle Social Network Recommendation Algorithms and Similarity Network Evaluation[J]. Journal of Applied Mathematics,2013,(2013-7-21), 2013, 2013(2013):3600-3611.

[2] Cechinel C, Sicilia M, Ngel, et al. Evaluating collaborative filtering recommendations inside large learning object repositories[J]. Information Processing & Management, 2013, 49(1):34-50.

[3] Walter F E, Battiston S, Schweitzer F. A model of a trust-based recommendation system on a social network[J]. Autonomous Agents and Multi-Agent Systems, 2006, 16(1):57-74.

[4] De Gemmis M, Lops P, Semerangro G, et al. Integrating tags in a semantic content-based recommender[J]. 2008:163-170.

[5] Bischoff K, Firan C S, Nejdl W, et al. Can all tags be used for search?[J]. 2008:193-202.

[6] 唐晓波, 张昭. 基于混合图的在线社交网络个性化推荐系统研究[J]. 情报理论与实践, 2013, 36(2):91-95.

[7] Liu Y, Safavi T, Dighe A, et al. Graph Summarization Methods and Applications: A Survey[J]. 2016

[8] Neo4j, Inc. Why Graph Databases[EB/OL]. https://neo4j.com/why-graph-databases.

[9] Pham M, Cao Y, Klamma R, et al. A Clustering Approach for Collaborative Filtering Recommendation Using Social Network Analysis[J]. Journal of Universal Computerence, 2010, 17(4):583-604.

[10] Jamali M, Ester M. A matrix factorization technique with trust propagation for recommendation in social networks[C]// ACM Conference on Recommender Systems. ACM, 2010:135-142.

[11] Debnath S, Ganguly N, Mitra P. Feature weighting in content based recommendation system using social network analysis[C]// 国际万维网大会. 2008:1041-1042.

[12] Mathur A. Social network based recommendation method and system: US, US8489515[P]. 2013.

[13] 章讯, 龙华, 周芝民. 基于标签改进社交网络好友推荐算法研究与应用[J]. 信息技术, 2017(6):15-18.

[14] 陈克寒, 韩盼盼, 吴健. 基于用户聚类的异构社交网络推荐算法[J]. 计算机学报, 2013, 36(2):349-359.

[15] 景楠, 王建霞, 许皓,等. 基于用户社会关系的社交网络好友推荐算法研究[J]. 中国管理科学, 2017, 25(3):137-146.