ABSTRACT

StackOverflow作为全球最大的技术问答网站之一，已经成为了程序员们学习计算机知识的主要来源。因为它涵盖了广泛的与编程和计算机相关主题。StackOverflow通过构建一个主题标签-主题标签的关系图，分析了在此网站上用户的行为活动和对计算机科学兴趣之间的关系。本文所进行的分析旨在强调所构建的动态网络与静态方法相比，网络随时间的演变。并将文本挖掘技术与传统的预测方法相结合。

1 简介

StackOverflow (SO)是最大的CQA社区之一。

(Community Question Answering）社区，为程序员提供的服务。超过300万用户。它已成为程序员学习计算机知识的主要来源，此网站涵盖了广泛的与编程相关的主题。这个论坛可以用于许多目的，包括获得关于编程语言或的一般信息。编程语言或解决编程问题。许多基于社区的信息网站

基于社区的信息网站，如SO，依靠元数据，称为标签，这 标签，有助于对内容进行索引、分类和搜索，是 内容的分类和搜索，并且是在以下情况下明确要求的 在发布查询时，用户。在SO中，对提交的问题进行标记是强制性的 (每个问题至少有一个标签），每个问题最多可以包含五个标签。只有拥有高级

信誉状态的用户才能生成新的标签，否则将使用现有的标签。本研究的目的是对StackOverflow的标签进行分析，以便通过将标签和它们的发生率表示为一个图来获得关于用户社区的信息。图的节点由标签本身表示，如果两个节点A和B至少有一个查询中同时出现，那么这两个节点之间就有一个公共弧。所得网络的弧线根据弧线所连接的两个标签在同一问题中出现的次数进行加权。在研究过程中，这个标签-标签网络的所有特征都得到了探讨，并与一些标准模型进行了比较：**Barabási-Albert。**

**Erdös-Rényi, Watts-Strogatz, Configuration Model**。特别是，网络的一个特点是，它不是由其用户有意建立的，目的是为了 形成联系的目的（它不是一个侧重于社会关系的网络），但反映了共同的 它不是一个专注于社会关系的网络），而是反映了社区成员的共同利益。该研究的目的之一实际上是为了

事实上，该研究的目标之一是确定用户提出问题最多的IT领域，从而利用服务的学习性。本报告的组织方式如下：第[2]节简要介绍了所使用的数据采集方法、一些汇总的统计数据以及网络的构建，随后在第[3]节中对网络特征进行了初步分析。随后，第[4-6-7]节中讨论了社区发现、链接预测和传播方法。第[5]节介绍了网络的动态分析和 在一年中的几个月里，从它那里获得的社区，从一个时间性的 从整个数据集的时间分析来看。最后，在第 [8]我们提出了对以前的链接预测任务的深化，通过它我们要验证是否 将各自的意义纳入到节点中，将导致 到一个更有效的网络链接预测系统。