项目进展情况:

1. 2015年2月—2015年3月进展

(1)在未接触giraph源码之前我们先进行了对pregel算法进行了了解(注:giraph作为pregel的替代算法)阅读了论文《大规模图计算模型PreqeI》：粗略了解pregel的大致结构以及运行状态

接下来阅读了论文论文《Pregel: A System for Large-Scale Graph Processing》：加深了我们pergel的运行机制有了比较详细的了解，对学习giraph有了初步的认识。

(2)接下来阅读了《Comparing community structure identification》,《A Genetic Algorithm with Local Search》，《Optimizing Graph Algorithms on Pregel-like Systems》等论文学习了当今对图研究的相关的论文，为本研究最终的目的：将上述文章中所介绍的将这些算法应用到Web链接组成的图或SNS中人与人之间的各种关系图的实际处理系统中，提供了思路。

(3)订购了《hadoop实战》这本书，了解了hadoop的运行原理，初步学习对hadoop集群具体使用方法，以及相关的机制。

②2015年4月—2015年6月进展:

(1)完成了VM虚拟机上面对hadoop的配置，并且成功运行。

创建了1个主节点和3个分支节点，完成了对hadoop分布式结构的部署。

(2)从giraph官网上下载了giraph的源码，放在主节点中，在主节点上完成了构建girph源码的maven程序的部署，并且成功编译了上述giraph形成了可运行的jar包

(3)在主节点上装并调试了eclipse 并且导入了giraph的源码方便分析，调试，和修改

(4)安装了运行giraph源码所需要的zookeeper软件，使在单机调试giraph源码的时候顺利的运行：使从giraph官网上面给出例的代码测试样例运行出来得到了与官网一致的正确输出。

(5)在2步中maven构建的基础之上在HDFS上面成功分布式运行了giraph的代码。

③ 2015年9月—2016年1月(现阶段)：

1. 深入对giraph源码中图处理的例子:最短路径算法，佩奇排名算法以等算法的源代码进行了深入的分析，掌握了giraph源码的运行方式并以矢量图的形式简洁的表示出来。
2. 已经能在基于giraph数据结构的前提下写出简单的算法比如统计计算过程中每个超步中活跃点的个数

项目阶段成果（字数精炼在200字内）:

1. 掌握了giraph源码的运行机制，并能够基于该机制写出算法。
2. 能使上述写出的算法在hadoop分布式平台上成功的运行。
3. 能处理图，并且能通过giraph算出数据(比如最短路径，佩奇排名)
4. 能够对实验中所出的错误进行调试和修改。

存在的问题：

1. linux的eclipse在调试giraph源码的时候无法看到部分变量。

参考解决办法：通过对软件设置的改变，来达到调试的目的。

1. 在分布式的环境下调试程序仍然是我们面临的问题(注因为分布式是多线程并发的所以不能像原来一样的单步调试)

参考解决办法：在相应的地方设置输出来判定数据的输出是否正确。

1. 自己写的算法仍然不够:

解决方法：更多的实验性的编写算法并实现，使自己加深对giraph的熟悉