**研究工作总结**

1. **研究内容和研究方法**

研究内容：本项目将研究基于Pregel模型的高性能图处理系统Apache Giraph，研究常用的图处理算法在单平台或者多平台上面的实现，并将这些算法应用到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。

研究方法：本组通过阅读giraph代码学习并理解其运作机制，进而实现某些经典算法，来对比分布式运行算法对大数据处理的优越性。

2. 要说明主要的科学发现、特色和创新之处，并列出具体的内容和必要的数据

具体内容：本组通过编写算法成功处理由2200个点和若干边组成的图，通过比较单机运行和分布运行时耗时，发现基于giraph机制的多机处理的效率远比单机效率高(其中单机运行大概需要2.5而分布式只需要不到20s)从上述数据可以得出基于giraph机制下的图处理程序的优越性。

特色：本项目的特色在于可以基于giraph源码的机制改进绝大多数对大图运算的算法，进而提高该算法的效率。

3. **成果的科学意义和应用前景**  ①成果科学意义

本组所研究的程序能改进所有有关图的算法，通过giraph的机制，使其在分布式环境下更高效的处理与图相关的大数据。

应用前景:

②Giraph是一个高扩展性的交互图形处理系统，本组通过对Giraph进行了改进和提升，实现对大图的分析。Giraph的整体计算模型可以被运用在本组服务器内并行巨大的网络图计算。并将这些算法应用到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

4. **计划（如有变动，须说明）的完成情况、存在的问题、建议及其他需要说明的情况（包括项目研究工作中的难点和经验，本项目研究是否达到预期目标，如未达到，请分析原因和可能的解决途径。今后进一步研究的建议和设想。）)**

①完成情况:

成功在**Giraph**系统平台上成功运行并行算法，且用实例来检测其实用性，比较了其与串行算法的计算效率(并行远远优于串行)

1. 存在问题

不足之处在于在编写算法时某些思路不清晰导致的计算结果与期望值有所差距，本组还在耐心的调试之中。

1. 项目中的难点:

本项目研究的难点在于对giraph源码的分析，通过查阅网上的资料学习，掌握了giraph的相关机制。达到了预期的目标

1. 进一步的研究和设想:

通过giraph改进更多有关图的算法，比较其效率，并推广。

项目取得的主要成果与结果:

成果:

1. 理解并掌握了giraph的算法原理，并且能够基于giraph的机制编程
2. 通过编写算法改进了介数中心性算法，最短路径算法，弱连通分量算法(wcc)使其在分布式下的效率大大的优于在单机模式下的效率。并且保证了正确率
3. 成功运行的并行算法来实现Web链接组成的图或SNS中人与人之间的各种关系图

结果:

完成了既定目标