

实验 4 图的三角形计数

1. 实验要求

实验背景

图的三角形计数问题是一个基本的图计算问题，是很多复杂网络分析（比如社交网络分析）的基础。目前图的三角形计数问题已经成为了 Spark 系统中 GraphX 图计算库所提供的一个算法级 API。本次实验任务就是要在 Hadoop 系统上实现 Twitter 社交网络图的三角形计数任务。

实验任务

一个社交网络可以看做是一张图（离散数学中的图）。社交网络中的人对应于图的顶点；社交网络中的人际关系对应于图中的边。在本次实验任务中，我们只考虑一种关系——用户之间的关注关系。假设“王五”在 Twitter/微博中关注了“李四”，则在社交网络图中，有一条对应的从“王五”指向“李四”的有向边。图 1 中展示了一个简单的社交网络图，人之间的关注关系通过图中的有向边标识了出来。。

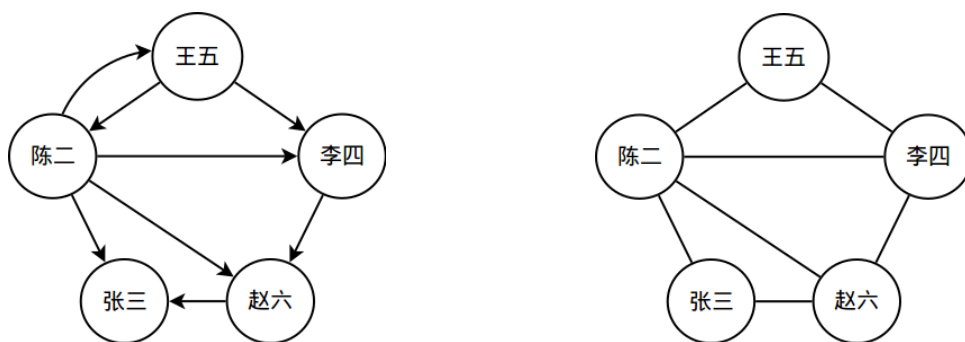


图 1 一个简单的社交网络示例。左侧的是一个社交网络图；右侧的图是将左侧图中的有向边转换为无向边后的无向图。

本次的实验任务就是在给定的社交网络图中，统计图中所有三角形的数量。

在统计前，需要先进行有向边到无向边的转换，依据如下逻辑转换：

IF ($A \rightarrow B$) OR ($B \rightarrow A$) THEN A-B

“ $A \rightarrow B$ ”表示从顶点 A 到顶点 B 有一条有向边。 $A-B$ 表示顶点 A 和顶点 B 之间有一条无向边。一个示例见图 1，图 1 右侧的图就是左侧的图去除边方向后对应的无向图。

请在无向图上统计三角形的个数。在图 1 的例子中，一共有 3 个三角形。

4.输出结果文件的截图，输出结果文件在 HDFS 上的路径（某些情况下助教会检查 HDFS 上的输出文件）

- 1.本次实验的程序运行会比较耗时，请尽早开始实验，以免最后非常仓促。
- 2.本次实验会占有较大的比重。