Find the Mutual Friends

Summary:

This assignment was quite challenging but rewarding in the end. The mapper reads a file containing a userID and a friends list on each line. The userID and each friendID in the friends list is then paired together and emitted as a key with the original friends list as the value. Next, the reducer accepts the mappers results and groups all like keys and their associated friends lists. After the grouping, the intersection of the friends lists is produced and the reducer emits the userID and friendID pair along with the intersecting friends lists (mutual friends)

To run this program do the following:

1. Put the sample_friends.txt, friend_mapper.py, and friend_reducer.py files in the same directory (or your Hadoop directory if using streaming)
2. Run the following command
    - From the command line type the following,
        - Windows – more sample_friends.txt | py friend_mapper.py | py friend_reducer.py
        - Linux - cat sample_friends.txt | python friend_mapper.py | python friend_reducer.py
        - Note: “python” or “py” is required before the mapper and reducer files if the python env variable is not specified
    - From the Hadoop cluster
        - Use “hdfs dfs – put <file.ext><directory>” to put the sample data, mapper, and reducer files above in a directory such as “/input”
            - Example: hdfs dfs -put sample_friends.txt /input
        - hdfs dfs -cat /input/sample_friends.txt | python friend_mapper.py | python friend_reducer.py

Testing Results:

```
[redacted]>more sample_friends.txt
101,102 103 104 105 106
102,101 103 104
103,101 102 104 105
104,101 102 103
105,101 103
106,101
```

```
[redacted]>more sample_friends.txt | py friend_mapper.py
('101', '102')  ['102', '103', '104', '105', '106']
('101', '103')  ['102', '103', '104', '105', '106']
('101', '104')  ['102', '103', '104', '105', '106']
('101', '105')  ['102', '103', '104', '105', '106']
('101', '106')  ['102', '103', '104', '105', '106']
('101', '102')  ['101', '103', '104']
('102', '103')  ['101', '103', '104']
('102', '104')  ['101', '103', '104']
('101', '103')  ['101', '102', '104', '105']
('102', '103')  ['101', '102', '104', '105']
('103', '104')  ['101', '102', '104', '105']
('103', '105')  ['101', '102', '104', '105']
('101', '104')  ['101', '102', '103']
('102', '104')  ['101', '102', '103']
('103', '104')  ['101', '102', '103']
('101', '105')  ['101', '103']
('103', '105')  ['101', '103']
('101', '106')  ['101']
```

```
[redacted]> more sample_friends.txt | py friend_mapper.py | py friend_reducer.py
101, 102 ['103', '104']
101, 103 ['102', '104', '105']
101, 104 ['102', '103']
101, 105 ['103']
101, 106 []
102, 103 ['101', '104']
102, 104 ['101', '103']
103, 104 ['101', '102']
103, 105 ['101']
```