Northeastern University, Khoury College of Computer Science

CS 6220 Data Mining | Assignment 3

Due: February 15, 2023(100 points)

Sibo Wu

Github Repo

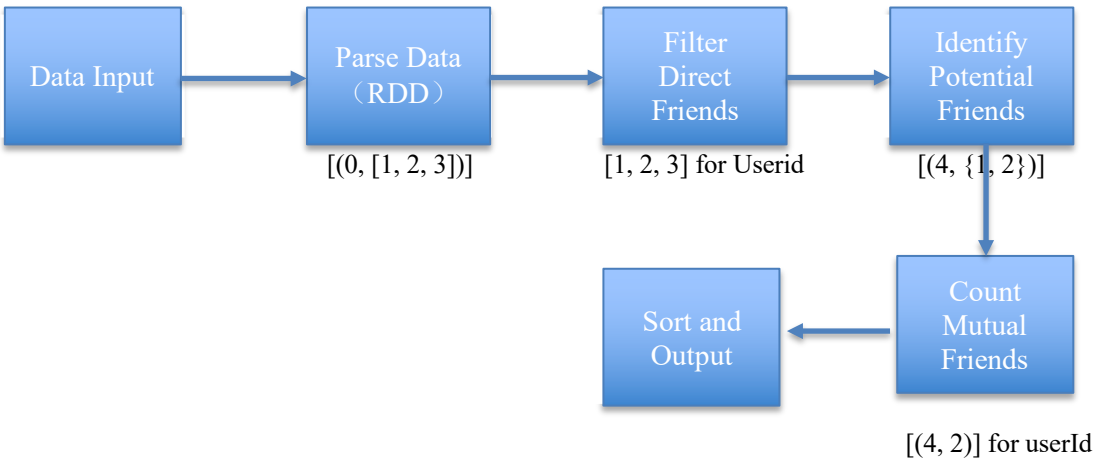
People You Might Know

Pipeline Sketch

1. Parse the social network data to create an RDD of users and

their direct friends.

2. For each user, identifie potential friends by excluding direct connections and calculating the number of mutual friends.
3. Rank these potential connections based on the count of mutual friends, and for each user, we output the top N recommendations.



Algorithm

```
def generate_recommendations(user_id, user_friends, N=10):
    """
    Generate recommendations for a given user based on the number of
mutual friends.

    Parameters:
    - user_id (int): The ID of the user for whom to generate
recommendations.
    - user_friends (RDD): An RDD of (user_id, [friends_list]) tuples.
    - N (int): The number of recommendations to generate.

    Returns:
    - A list of user IDs representing the algorithm's recommendation of
people that the user might know,
      ordered by decreasing number of mutual friends.
    """

    # Extract the list of direct friends for the given user
    direct_friends = user_friends.filter(lambda x: x[0] ==
user_id).flatMap(lambda x: x[1]).collect()

    # Generate potential recommendations
    potential_recs = user_friends \
        .filter(lambda x: x[0] != user_id) \
        .flatMap(lambda x: [(friend, {x[0]}) for friend in x[1] if
friend not in direct_friends]) \
        .reduceByKey(lambda a, b: a | b) \
        .map(lambda x: (x[0], len(set(direct_friends) & x[1]))) \
        .filter(lambda x: x[1] > 0 and x[0] != user_id)  # Exclude the
user's own ID from recommendations

    # Sort by number of mutual friends (descending) and then by user ID
(ascending)
    top_recs = potential_recs.sortBy(lambda x: (-x[1],
x[0])).map(lambda x: x[0]).take(N)

    return top_recs
```

Output

User ID 924: Recommendations: 439,2409,6995,11860,15416,43748,45881
User ID 8941: Recommendations: 8943,8944,8940
User ID 8942: Recommendations: 8939,8940,8943,8944
User ID 9019: Recommendations: 9022,317,9023
User ID 9020: Recommendations: 9021,9016,9017,9022,317,9023
User ID 9021: Recommendations: 9020,9016,9017,9022,317,9023
User ID 9022: Recommendations: 9019,9020,9021,317,9016,9017,9023
User ID 9990: Recommendations: 13134,13478,13877,34299,34485,34642,37941
User ID 9992: Recommendations: 9987,9989,35667,9991
User ID 9993: Recommendations: 9991,13134,13478,13877,34299,34485,34642,37941