01 Facebook friends PySpark Assignment

Assignment:

Question 1: Apache Spark RDD Operations - Computing Average Number of Friends by Age

You are working with a dataset called "fakefriends.csv" that contains information about people and their social connections. The CSV file has the following structure:

- Column 0: Person ID
- Column 1: Name
- Column 2: Age
- Column 3: Number of Friends

Task:
Write a Spark application using RDDs to analyze this dataset and compute the average number of friends for each age group. Your solution should:

1. Parse each line of the CSV file to extract age and number of friends
2. Group the data by age
3. Calculate the average number of friends for each age
4. Display the results showing each age and the corresponding average number of friends

Expected Output Format:

```
(age, average_number_of_friends)
```

Requirements:

- Use Spark RDDs (not DataFrames)
- Implement the parsing function to extract relevant fields
- Use appropriate RDD transformations like `map`, `mapValues`, and `reduceByKey`
- Handle the aggregation logic to compute averages correctly
- Display the final results

Bonus: Explain how the `mapValues` and `reduceByKey` operations work together to compute the average in your solution.
