Counting the number of triangles with Spark

In this application, we are given a file, representing a sample of a the LiveJournal social network that you find at https://snap.stanford.edu/data/soc-LiveJournal1.html. The network is undirected and is described by a `tab`-separated text file with the following format:

```
7	0,31993,40218,40433,1357,21843
```

The first number is the id of a node of the network. It is follow by a comma-separated list of its neighbours. The original dataset was used and is described in the following papers:

- L. Backstrom, D. Huttenlocher, J. Kleinberg, X. Lan. Group Formation in Large Social Networks: Membership, Growth, and Evolution. KDD, 2006.
- J. Leskovec, K. Lang, A. Dasgupta, M. Mahoney. Community Structure in Large Networks: Natural Cluster Sizes and the Absence of Large Well-Defined Clusters. Internet Mathematics 6(1) 29--123, 2009.

Your assignment

Your are asked to compute the number of triangles in the graph, i.e., the total number of *unordered* triples (a,b,c) of vertices that form 3-cliques. In general, solving the problem can be computationally expensive, in the order of $\Theta(n^3)$, with n the number of vertices in the graph. On the other hand, the number of triangles in a (undirected) network is a key indicator of the degree of cohesiveness and social structure. In particular, the number of triangles is necessary to compute the *global* clustering coefficient. In particular, the global clustering coefficient C of a network G is defined as:

$$C = \frac{< Number\ of\ triangles\ in\ G >}{\binom{n}{3}}$$

Solving using Spark

As usual, you find a solution that is not optimized. Still, the solution you find here essentially follows the basic MapReduce implementation of the NodeIterator algorithm presented in

Suri, Siddharth, and Sergei Vassilvitskii. "Counting triangles and the curse of the last reducer." In Proceedings of the 20th international conference on World wide web, pp. 607-614. ACM, 2011.