Part-2 : Analyzing Social Networks using GraphX

Dataset used is "**wiki-Vote**". It is taken from SNAP repository.

Following are the results of our analysis:

a. Out-degree

id	outDegree
2565	893
766	773
11	743
457	732
2688	618

b. In-degree

id	inDegree
4037	457
15	361
2398	340
2625	331
1297	309

c. Page Ranks

id	pagerank
4037	32.76139
15	26.253
6634	26.16452
2625	23.51152
2398	18.72839

d. Connected Components

component	count
0	7066
532575944741	3
592705486870	3
936302870556	3
438086664219	2

e. Triangle Counts

id	count
2565	30940
1549	22003
766	18204
1166	17361
2688	14220

Since there were errors in AWS, the code was initially run on databricks. Later, the code was run in IntelliJ and run on AWS. Instructions are in the Readme file.
Link to the databricks file:
https://databricks-prod-cloudfront.cloud.databricks.com/public/4027ec902e239c93eaaa8714f173bcfc/3089703568886817/1834795912323072/2338724615062927/latest.html

References:
https://docs.aws.amazon.com/emr/latest/ReleaseGuide/emr-spark-application.html
https://spark.apache.org/docs/latest/graphx-programming-guide.html
https://snap.stanford.edu/data/