平均距离计算

**数据集来源：**

小数据集：http://konect.uni-koblenz.de/networks/douban

大数据集：<http://konect.uni-koblenz.de/networks/livejournal-groupmemberships>

**数据集基本统计：**

Vertexs num = 154908

Edges num = 327162

avgDist = 5.10

Diameter = 9

**数据集HyperAnf结果：**

avgDist = 4.88

**数据集抽样方法结果：**

抽样k个样本，为抽样方差的无偏估计

1. 计算k个样本的均值当作整体均值的估计，即
2. 计算k个样本的方差的无偏估计当作整体的方差估计，即
3. 根据中心极限定理，无论原始数据满足何种分布，随机抽样k个样本，连续抽样n次，这n次抽样的均值都会满足正态分布，该正态分布为
4. 计算正态分布为为2个标准差的置信区间，根据正态分布的68-95-99.7的经验法则，即计算正态分布为的置信度为95%的置信区间（）
5. 不停的增大k值，观察抽样均值正态分布的统计数据

随着抽样样本个数k增大，均值很快就收敛，抽样均值的正态分布趋势越集中，2个标准差的置信区间在逐渐减小

The 1 round , Cost Time = 2.435067 min.

SampleSize = 10, Mean = 5.151680, Sigma = 0.652229.

Confident that : total mean 95 percent in [4.739175,5.564186]

The 2 round , Cost Time = 2.384917 min.

SampleSize = 20, Mean = 5.138057, Sigma = 0.507150.

Confident that : total mean 95 percent in [4.911252,5.364861]

The 3 round , Cost Time = 3.321983 min.

SampleSize = 30, Mean = 5.117406, Sigma = 0.469201.

Confident that : total mean 95 percent in [4.946077,5.288734]

The 4 round , Cost Time = 2.499300 min.

SampleSize = 40, Mean = 5.120099, Sigma = 0.506294.

Confident that : total mean 95 percent in [4.959994,5.280203]

The 5 round , Cost Time = 3.362150 min.

SampleSize = 50, Mean = 5.090705, Sigma = 0.535899.

Confident that : total mean 95 percent in [4.939130,5.242280]

The 6 round , Cost Time = 2.400317 min.

SampleSize = 60, Mean = 5.088548, Sigma = 0.525080.

Confident that : total mean 95 percent in [4.952973,5.224123]

The 7 round , Cost Time = 2.801850 min.

SampleSize = 70, Mean = 5.083020, Sigma = 0.498011.

Confident that : total mean 95 percent in [4.963972,5.202067]

The 8 round , Cost Time = 2.942117 min.

SampleSize = 80, Mean = 5.082336, Sigma = 0.507507.

Confident that : total mean 95 percent in [4.968854,5.195818]

The 9 round , Cost Time = 5.613250 min.

SampleSize = 90, Mean = 5.099874, Sigma = 0.505322.

Confident that : total mean 95 percent in [4.993342,5.206405]

The 10 round , Cost Time = 2.923817 min.

SampleSize = 100, Mean = 5.117239, Sigma = 0.488681.

Confident that : total mean 95 percent in [5.019503,5.214975]

The 11 round , Cost Time = 2.353917 min.

SampleSize = 110, Mean = 5.125184, Sigma = 0.497958.

Confident that : total mean 95 percent in [5.030227,5.220141]

The 12 round , Cost Time = 2.653667 min.

SampleSize = 120, Mean = 5.108776, Sigma = 0.486976.

Confident that : total mean 95 percent in [5.019867,5.197685]

**数据集基本统计：**

Vertexs num = 13,891,479

Edges num = 112,307,385

avgDist = 4.73

Diameter = 28

**数据集HyperAnf结果：**

avgDist = 6.57 与真实平均最短路径相差比较大，小数据集上的差距会稍微小一些

**数据集抽样方法结果：这几天跑不动**