

Homework 2

Problem 1

Google matrix and its eigenvalues: Google matrix is a particular stochastic matrix that is used by Google's PageRank algorithm. The matrix represents a graph with edges representing links between pages. Please calculate the eigenvalues of the google matrix ($N > 2000$ is required)

谷歌矩阵的定义

谷歌矩阵具有形式

$$G = \alpha S + (1 - \alpha)ee^T/N$$

其中 α 是一个预先给定的数值，代表用户随机浏览的倾向性

$$S = H + ea^T/N$$

而 H 是网页之间关系的邻接矩阵 A 归一化所得矩阵：

$$A_{ij} = \begin{cases} 1, i \rightarrow j \\ 0, i \nrightarrow j \end{cases}$$

同时

$$H_{ij} = \begin{cases} \frac{A_{ij}}{\sum_k A_{ik}}, \sum_k A_{ik} \neq 0 \\ 0, \sum_k A_{ik} = 0 \end{cases}$$

e 是全1向量

$$e = (1, 1, \dots, 1)^T$$

a 是指示悬挂节点的指标向量

$$a_i = \begin{cases} 0, \sum_k A_{ik} \neq 0 \\ 1, \sum_k A_{ik} = 0 \end{cases}$$

矩阵 H 代表用户按照网页之间链接浏览的状态转移，矩阵 S 考虑了悬挂节点（无外链网页）的修正，矩阵 G 考虑了用户在网页内链接以外的跳转行为，得到的就是谷歌矩阵。

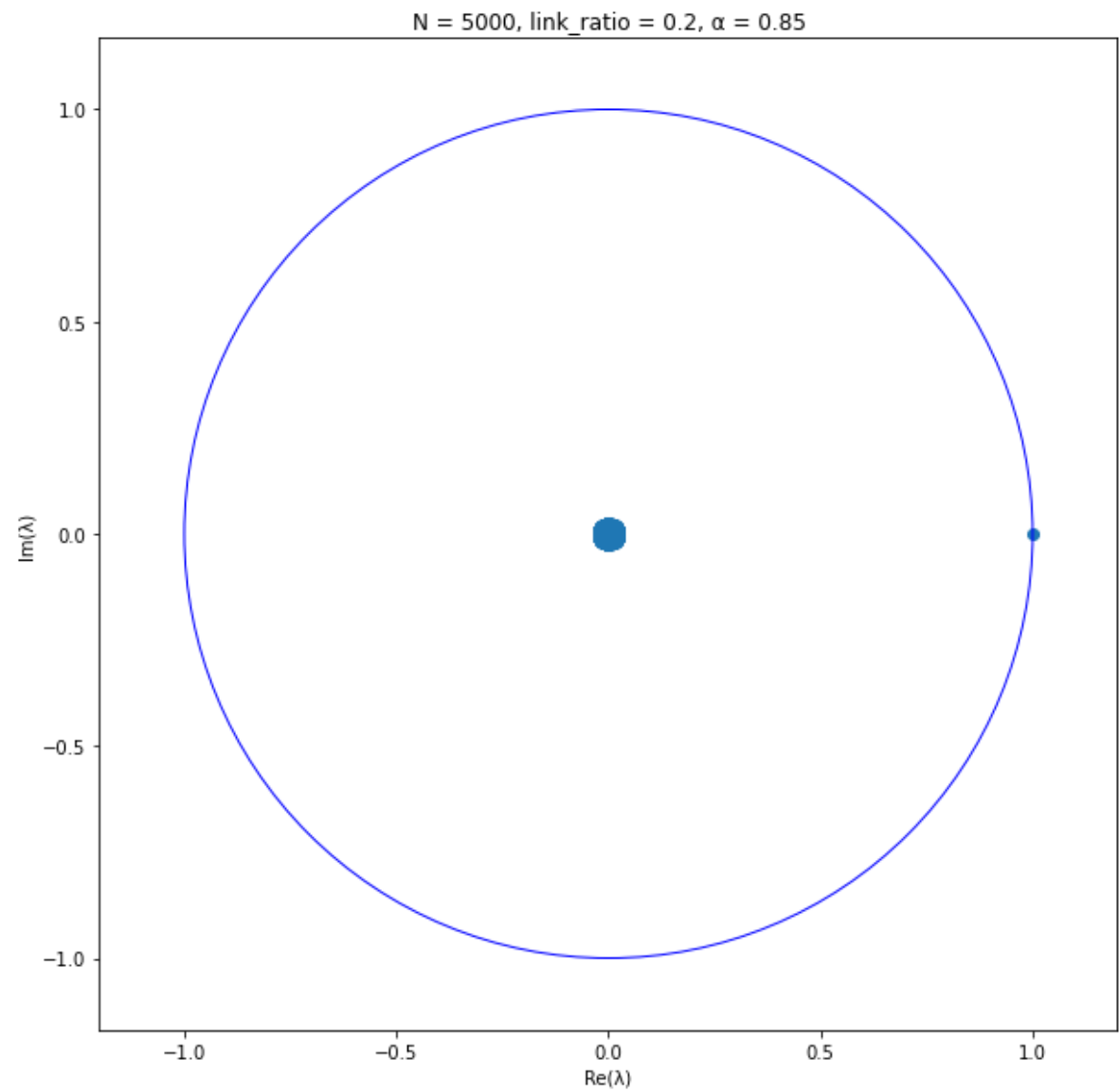
谷歌矩阵的生成和特征值计算

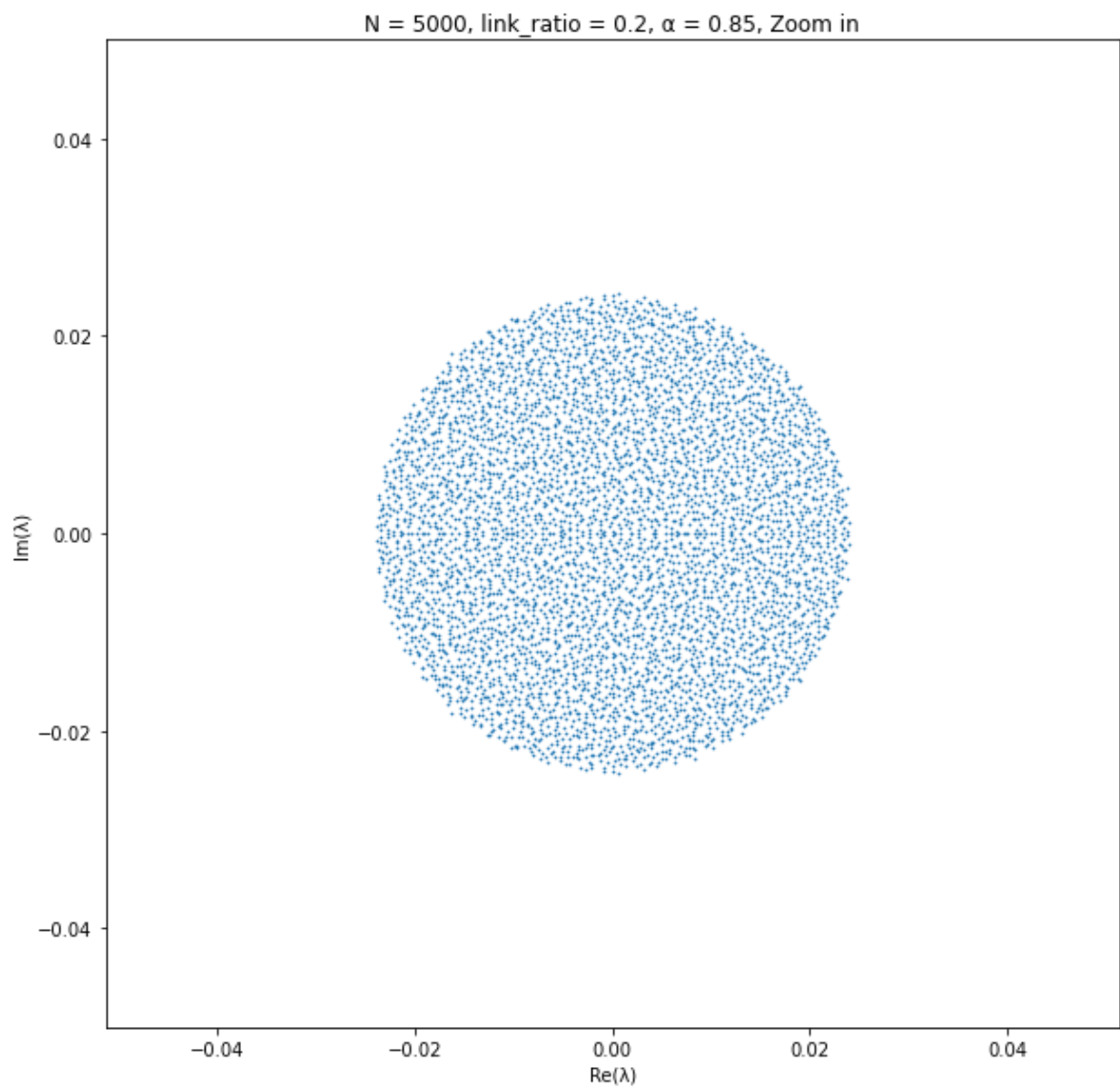
这里采用 Python 的 `numpy` 数值计算包来生成此矩阵。这里只概述生成中思路，具体代码请参考附件中的 `ipynb` 文件。

首先，要生成代表网站链接关系的邻接矩阵 A 。这里采用 `numpy` 的随机数生成函数 `np.random.rand` 实现，该函数可以生成任意大小的、由 $0 \sim 1$ 之间随机数组成的矩阵。为了控制生成结果中的链接强度，我们为添加一个参数 r ，如果 `np.random.rand` 生成矩阵的元素小于这个值，那么该处 A 的值为1否则取\$0\$。关于矩阵操作和运算，这里使用 `einops` 保证代码可读性。最终的求值过程利用 `numpy.linalg.eigen` 完成，这个函数会输出所给矩阵的特征向量和特征值。

结果和讨论

首先, 针对 $N = 5000, r = 0.2, \alpha = 0.85$ 的情况, 计算进行了 83.29 秒, 所得结果如下:

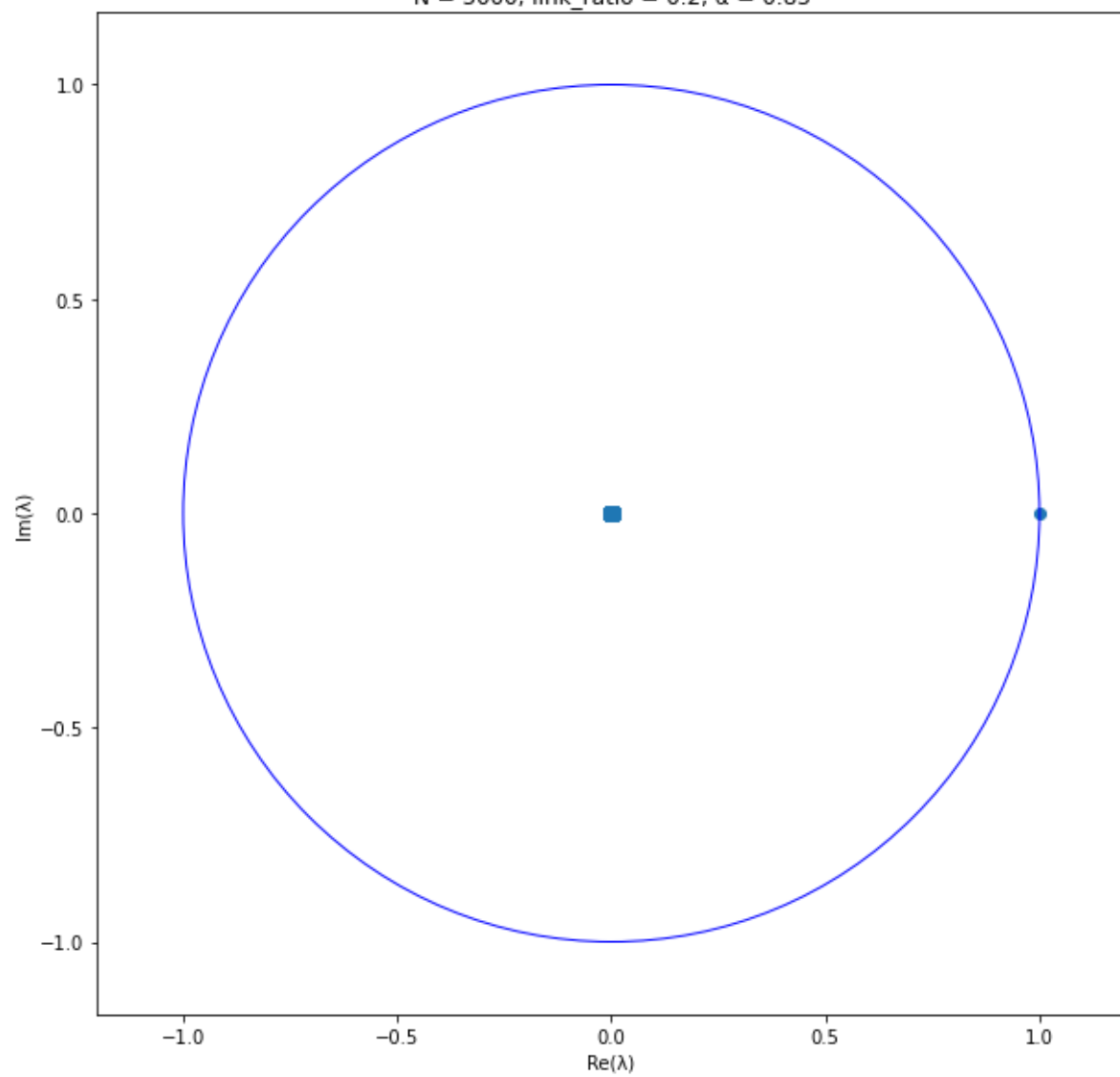


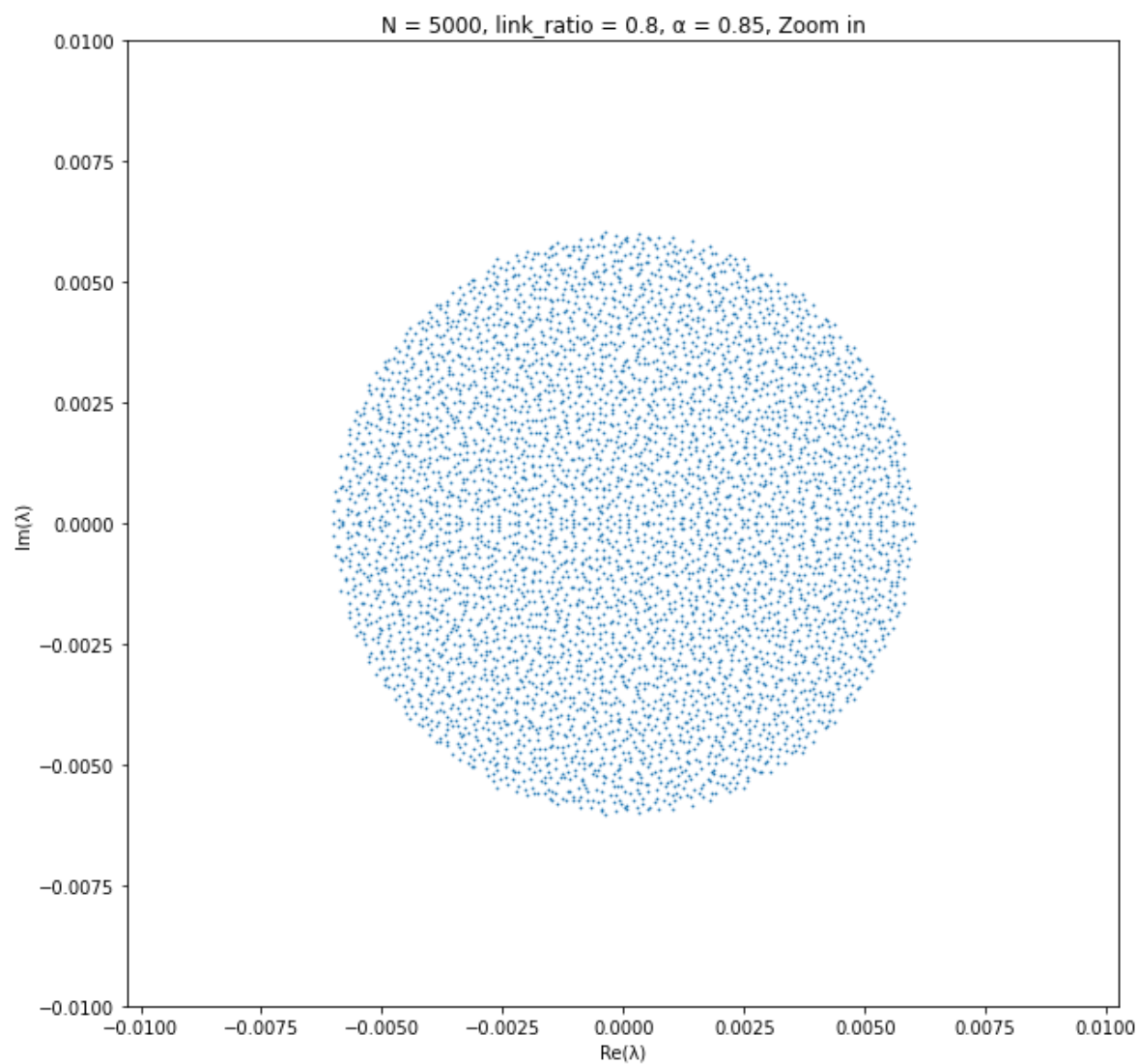


除了一个特征值1之外，其他所有的特征值都处在中心的很小区域内.在上图中我们给出了特征值在复平面上原点附近的分布，可以发现这些根的分布近乎均匀，并且位于一个小圆盘上.另一方面，这些根关于实轴是对称的.

针对 $N = 5000, r = 0.8, \alpha = 0.85$ 的情况，计算进行了 73.42 秒，所得结果如下：

$N = 5000$, $\text{link_ratio} = 0.2$, $\alpha = 0.85$

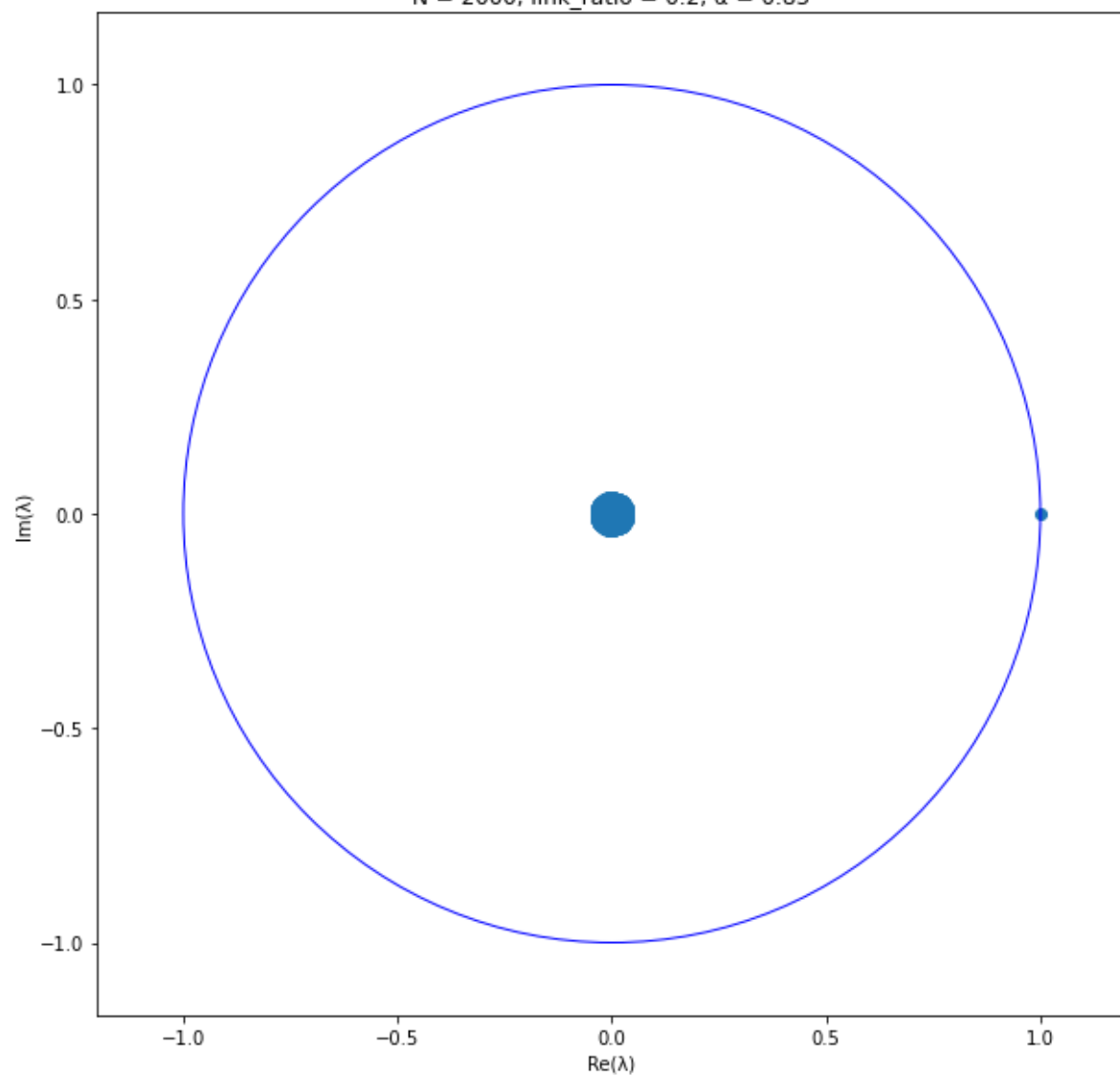


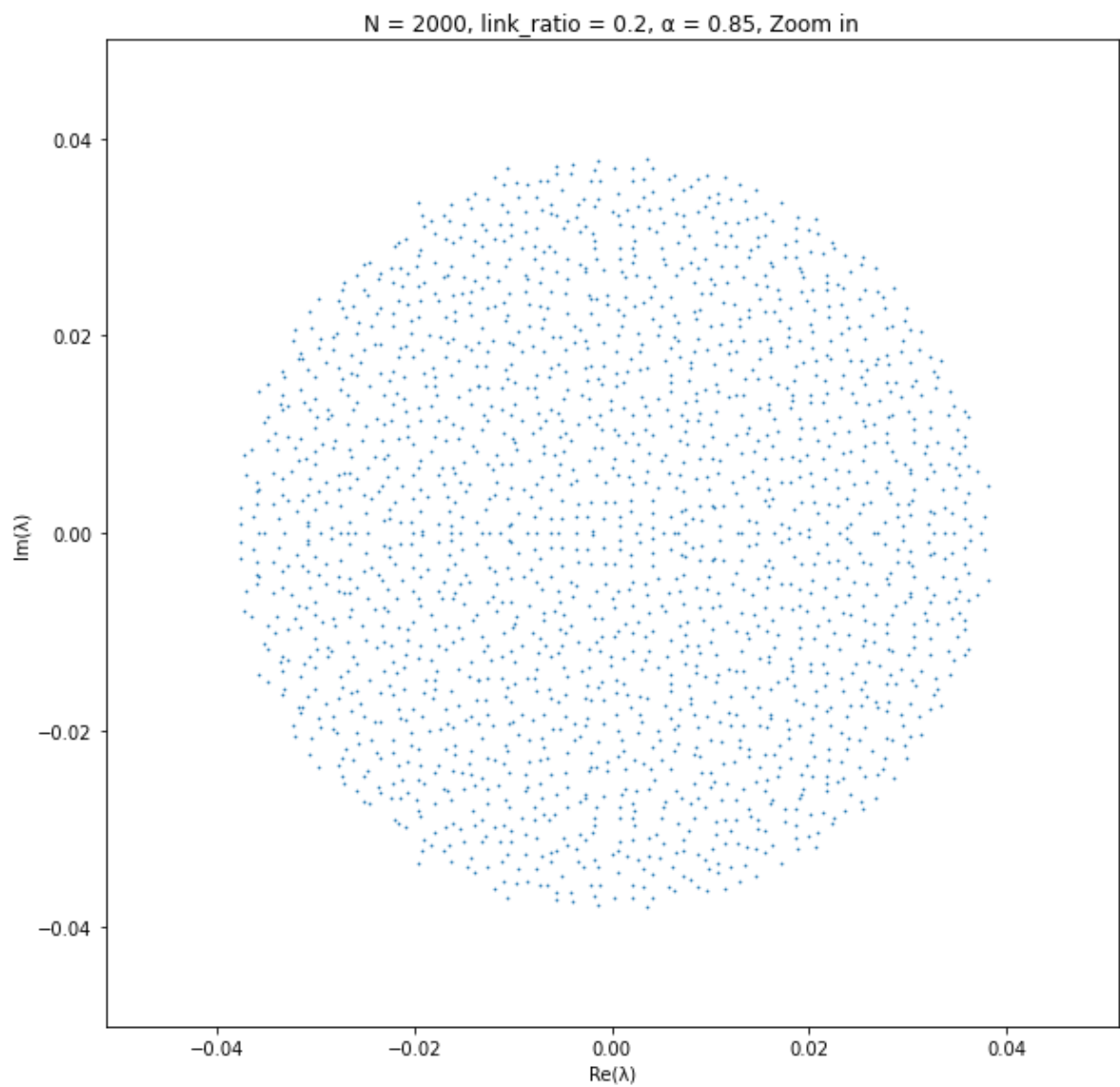


可以发现在原点附近的特征值分布在一个更小的圆盘上，同时仍然保有对称的特点。

针对 $N = 2000, r = 0.2, \alpha = 0.85$ 的情况，计算进行了 9.10 秒，所得结果如下：

$N = 2000$, $\text{link_ratio} = 0.2$, $\alpha = 0.85$

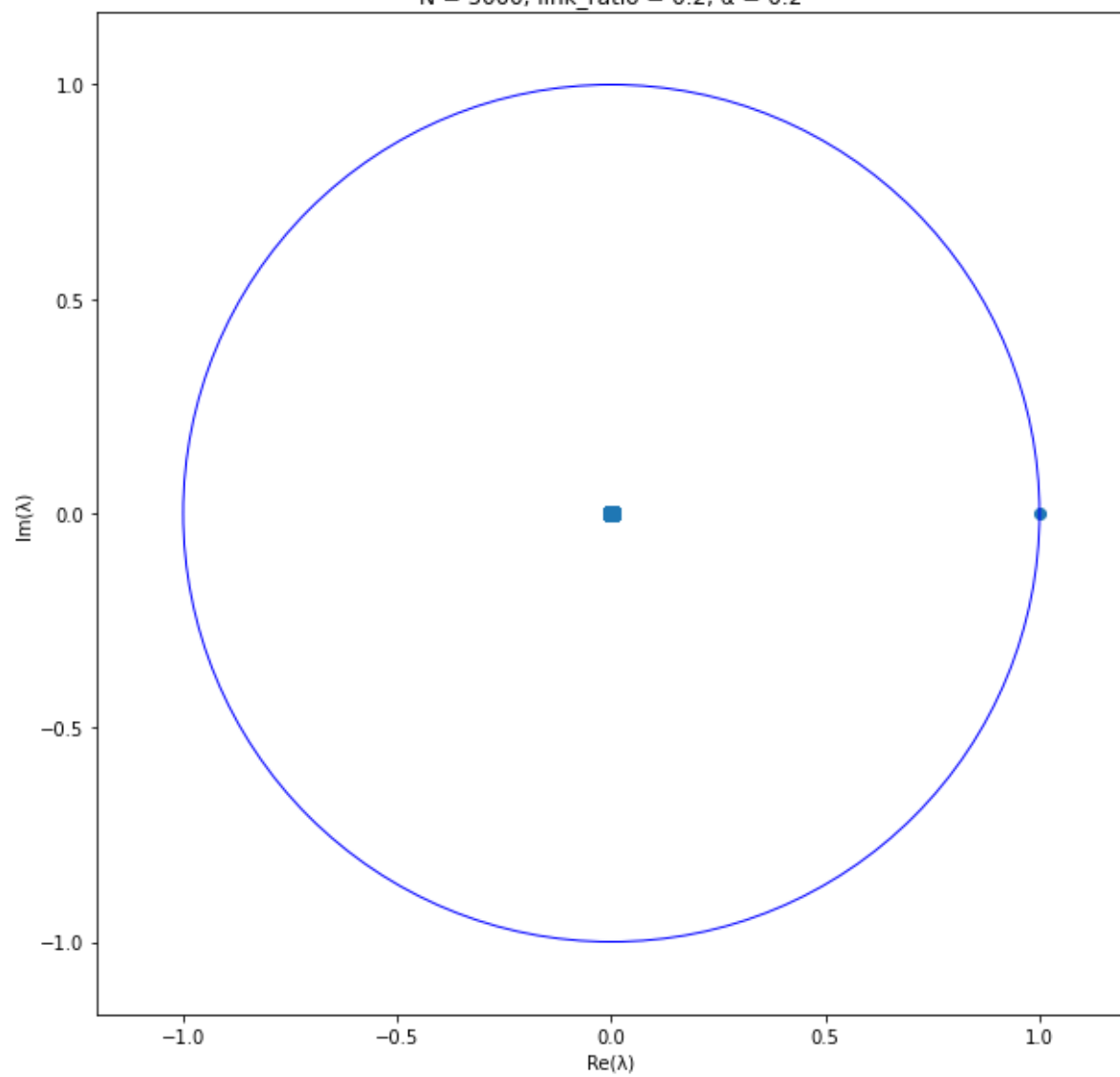


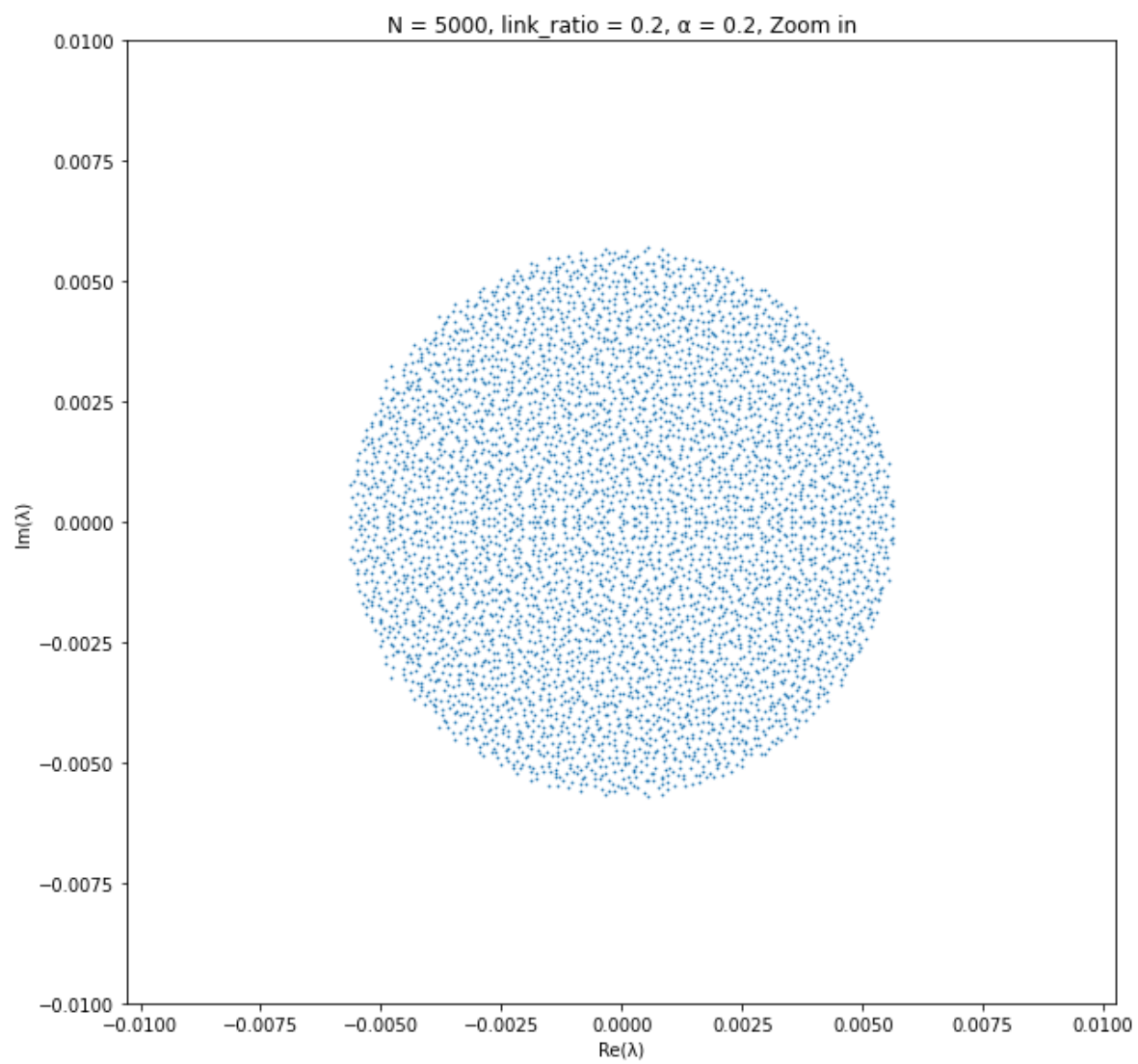


可以发现原点附近特征值的分布与 $N = 5000$ 的原始情况大致相同，但是由于顶点数减少，看起来分布比较稀疏。

针对 $N = 5000$, $r = 0.2$, $\alpha = 0.2$ 的情况，计算进行了 82.14 秒，所得结果如下：

$N = 5000$, $\text{link_ratio} = 0.2$, $\alpha = 0.2$





这时所得的圆盘比原始参数设置中小.