计算物理作业 2：Google 矩阵的本征值

井伟松
SA21004055

更新：*October 6, 2021*

1 Google 矩阵本征值的求解

本次作业的目标是生成维数 $N > 2000$ 的 Google 矩阵，并求解它的本征值。

1.1 Google 矩阵的介绍

Google 矩阵是 Google 公司提出 PageRank 算法的时候需要研究的一种矩阵。这类矩阵可以写成以下形式 (1; 2)：

$$\boldsymbol{G} = \alpha \boldsymbol{S} + \frac{1-\alpha}{N}\boldsymbol{E}$$

其中 $\boldsymbol{G}$ 为 $N \times N$ 的 Google 矩阵，$\boldsymbol{S}$ 为 Markov 矩阵，是由一个图的邻接矩阵 $\boldsymbol{A}$ 生成的。邻接矩阵 $\boldsymbol{A}$ 的定义为若存在一个由节点 j 指向节点 i 的链接，则 $A_{i,j} = 1$；否则有 $A_{i,j} = 0$（是一般意义上的邻接矩阵定义的转置）。$\boldsymbol{S}$ 是由邻接矩阵 $\boldsymbol{A}$ 的每一列进行归一化得到的，也就是当一列的元素都不为 0 时，有 $S_{i,j} = A_{i,j}/\sum_k A_{k,j}$；如果某一列的元素都为 0，则令这一列的元素都为 $\frac{1}{N}$(1; 2)。这个矩阵实际上是模拟一个上网者的行为：如果一个人在某个网页上存在许多条链接，那么他会随机选取一条进行浏览；如果没有链接的话就跳转到另外一个网页上。α 是阻尼系数，在这里的意义是一个人浏览完一个网页之后，选择点击这个网页上超链接的概率。如果他不点击这个超链接，则会选择键入一个新的地址（概率 $1-\alpha$），并假设键入每一个地址的概率是相同的，都是 $\frac{1}{N}$，这就是 Google 矩阵表达式中第二项的实际意义（$\boldsymbol{E}$ 是 $N \times N$ 的全 1 矩阵）。一般情况下可以取 $\alpha = 0.85$。从实际意义上来讲，矩阵 $\boldsymbol{A}$ 的对角线元素需要为 0，但是从数学角度来看，$\boldsymbol{A}$ 对角线元素不为 0 意味着存在自身指向自身的元素，这一点不影响整个图的本质性质，因此在探究实际生成 Google 矩阵的时候不要求 $\boldsymbol{A}$ 的对角线元素必须为 0。

可以理论分析得出 Google 矩阵的最大本征值为 1，第二大本征值模的上界为 $1-\alpha$。最大本征值对应的本征向量就是 Pagerank 向量。这个向量反应了每个网络节点的“受欢迎程度”，可以作为网站排名的依据。

1.2 本征值的求解

为了求解 Google 矩阵的本征值，首先要想办法随机生成 Google 矩阵。在这里为了简便起见，采用完全随机的网络的进行研究。由于一般网页的链接数量比较少，大概在 10 个左右，因此得到的邻接矩阵 $\boldsymbol{A}$ 一般是稀疏矩阵。生成随机网络的方法为首先选定维数 N 和网页链接的平均值 $mean$，然后对邻接矩阵 $\boldsymbol{A}$ 的每个元素以概率 $\frac{mean}{N}$ 赋值 1，以概率 $1-\frac{mean}{N}$ 赋值 0，再按上面的公式算出 Google 矩阵。

下面选取 $N=3000, mean=10$ 的情形对随机稀疏网络的 Google 矩阵的本征值进行求解。通过 Matlab 的`eig`函数可以很快地计算出本征能谱。下面选取不同的 α 数值，研究各个本征值的绝对值大小与本征值谱的分布。取 $\alpha=1$ 的情形，作出的本征能谱如下所示：

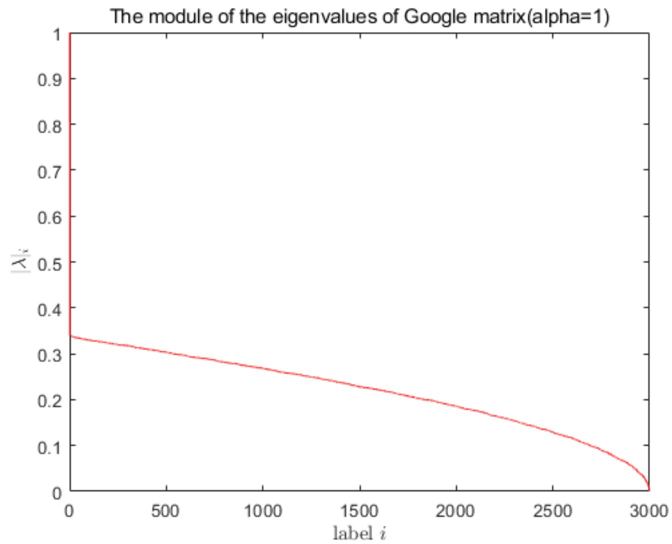

图 1: $\alpha = 0.1$ 时 Google 矩阵的本征值的模与序号之间的关系（降序排列）

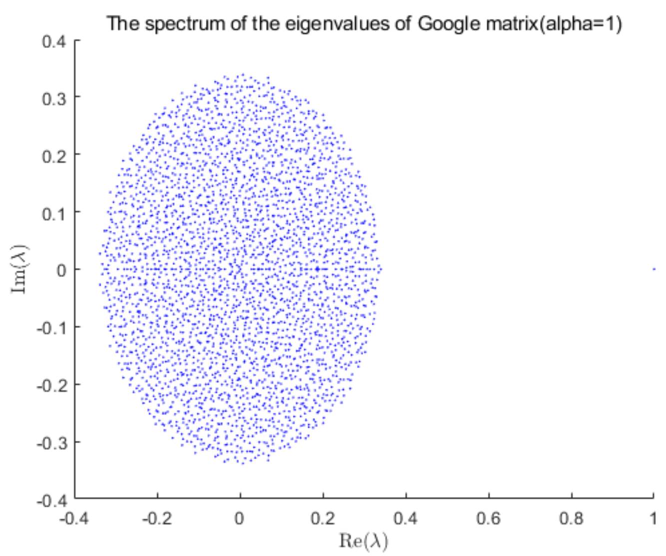

图 2: $\alpha = 0.1$ 时 Google 矩阵的本征值谱

取 $\alpha = 0.85$，本征值谱如下所示：

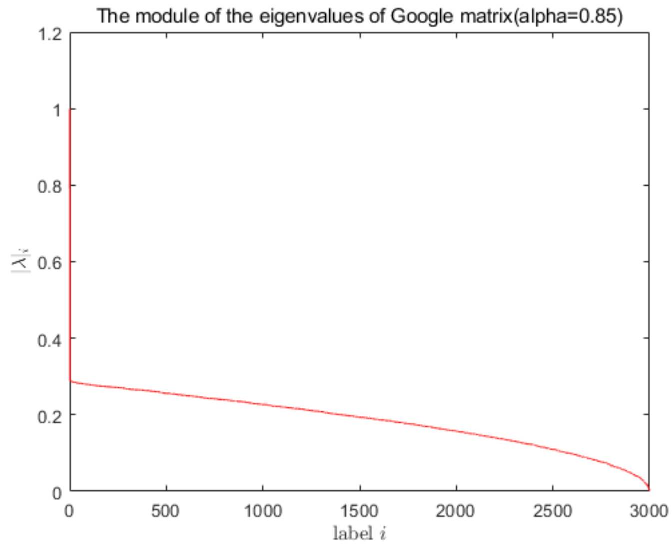

图 3: $\alpha = 0.85$ 时 Google 矩阵的本征值的模与序号之间的关系（降序排列）

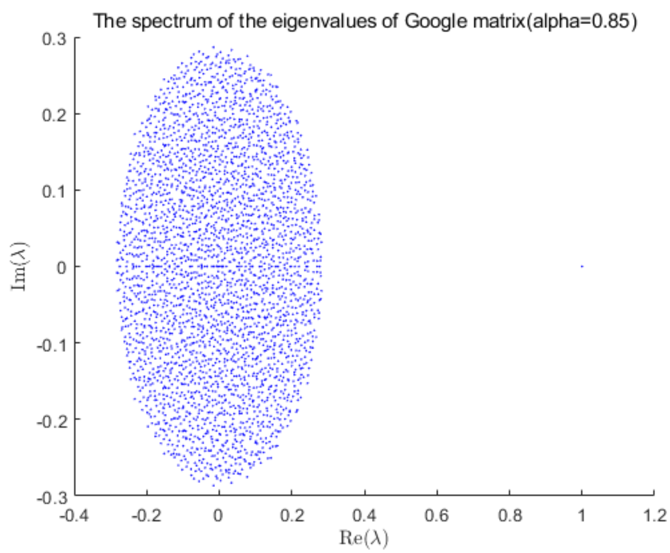

图 4: $\alpha = 0.85$ 时 Google 矩阵的本征值谱

取 $\alpha = 0.5$，本征值谱如下所示：

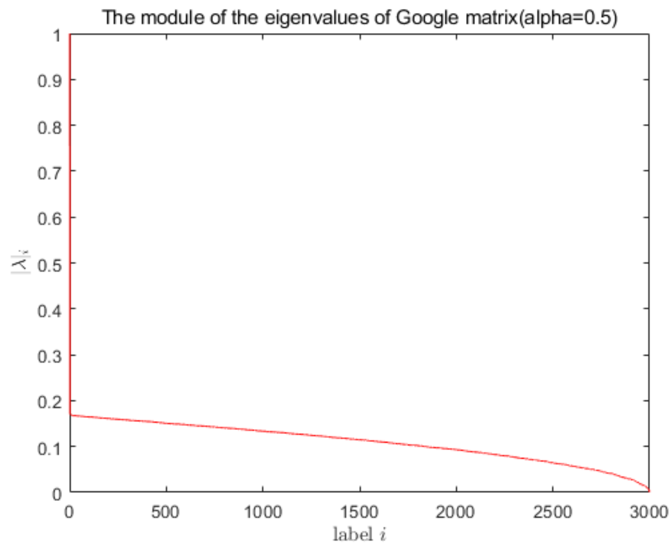

图 5: $\alpha = 0.5$ 时 Google 矩阵的本征值的模与序号之间的关系（降序排列）

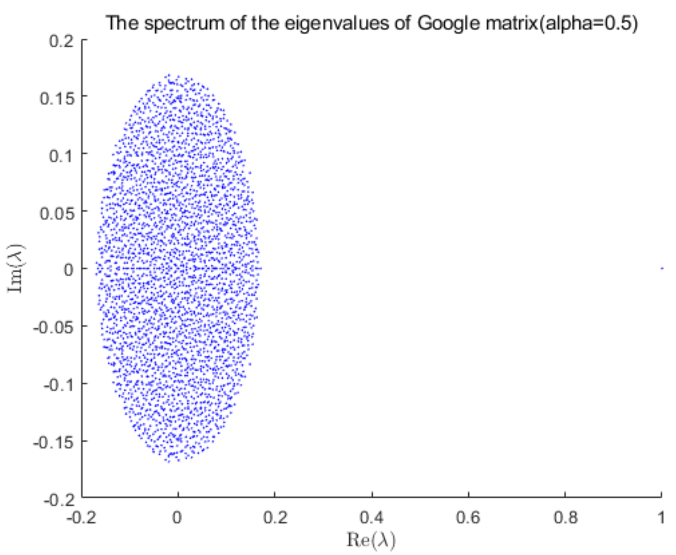

图 6: $\alpha = 0.5$ 时 Google 矩阵的本征值谱

取 $\alpha = 0.1$，本征值谱如下所示：

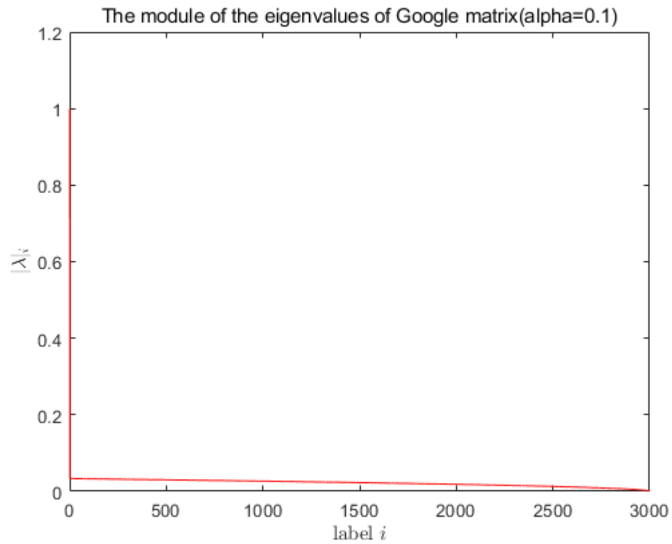

图 7: $\alpha = 0.1$ 时 Google 矩阵的本征值的模与序号之间的关系（降序排列）

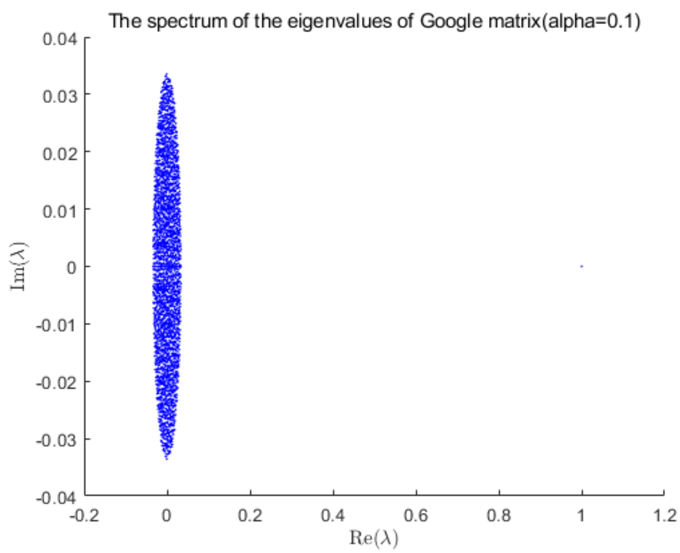

图 8: $\alpha = 0.1$ 时 Google 矩阵的本征值谱

1.3 一些讨论

从得到的本征值谱来看，选取的几种 α 数值得到的本征值谱的特征都是类似的：最大的特征值都是 1，且只有一个，其他特征值密布在一个圆形范围内（直线看起来很像椭圆，但是是由于 x, y 轴的比例尺不一样）。且 α 越大，圆的半径越小。$\alpha = 1, 0.85, 0.5, 0.1$ 的情况第二大本征值的模分别在 $0.35, 0.3, 1.75, 0.035$ 左右，表明第二大本征值的模与 α 基本是成正比的。这个结论与文献 (2) 中采用随机 AB 模型（具体生成方法参见文献 (3)）以及随机化的社区模型生成的网络得到的本征值谱较为类似。但是对于实际网络的分析却展示出了不同的结果：在 $\alpha \neq 0$ 时第二大本征值很接近 $1 - \alpha$；$\alpha = 0$ 时存在很多个简并本征值 1，且本征值在 0-1 之间分布都较为稠密。实际网络的本征值谱如下所示：

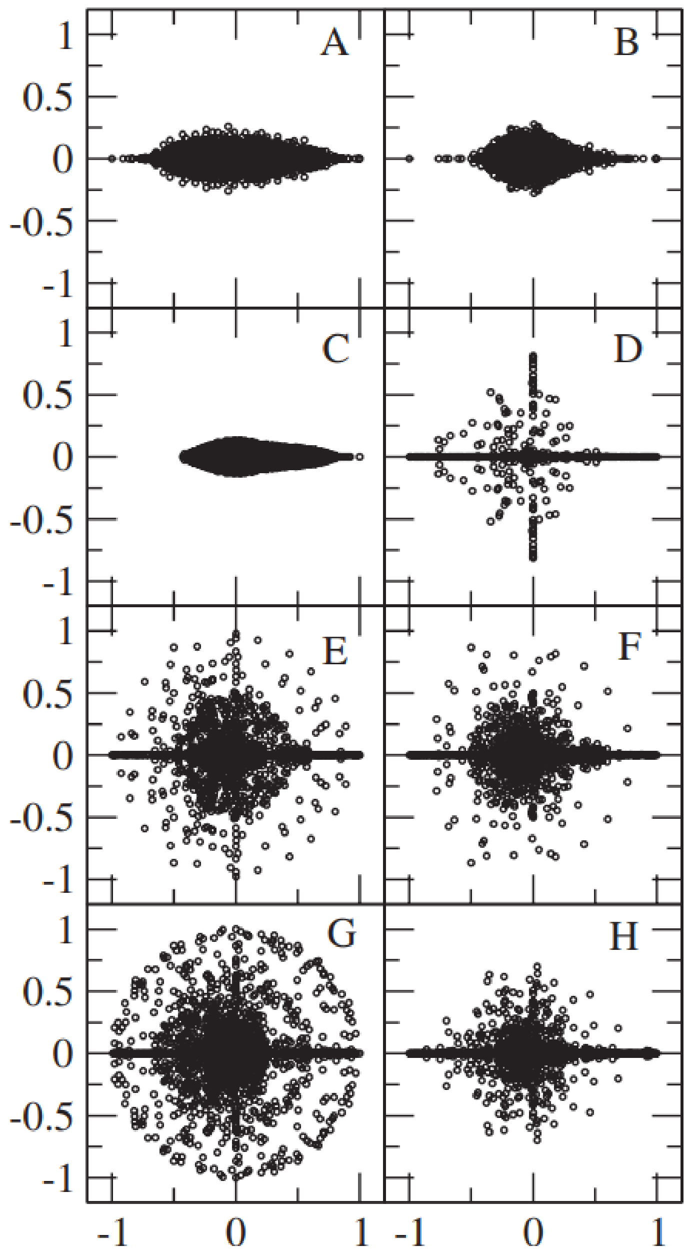

图 9: 采用实际网络得到的 Google 矩阵的本征值谱。图中 A,B,C 采用字典网络作为研究对象；D,E,F,G,H 采用社区万维网网络进行研究。图片来源 (1)

从采用实际网络得到的 Google 矩阵的本征值谱来看，可以看出本征值谱与网络生成的方法有很大关系。字典网络 (A,B,C) 和社区网络 (D,E,F,G,H) 得到的本征值谱具有很明显的差异，且都与随机网络的结果有很大的不同。字典网络的本征值谱基本在靠近实轴的区域有较为稠密的分布，而社区网络分布却相对比较分散，在圆周上也有很多的本征值分布。另外社区网络中也有很多本征值分布在实轴附近，这实际上是网络结构较为对称化的一种体现（若 A 到 B 有链接，则 B 到 A 也更倾向于产生链接）(1)。

随机网络的结果之所以和实际网络差距较大，原因在于实际的网络并不是完全随机的，网络的拓扑结构总是有一些内在的规律。例如网络的连接具有明显的社区性质，也就是节点优先在社区内部生成，并且具有对称化的特点。因此，文献 (1) 中提到了一种染色法来生成随机网络，得到的 Google 矩阵的本征值更接近于实际情形。另外对于本征值谱的问题也存在动力学上的解释：如果一个动力系统的最大本征值和第二本征值的差距越大，表明信息越容易从不同节点之间相互传递。由于实际网络的社区分布性质，导致任意两个节点之间（不同社区的）的信息传递速度可能会很慢，因此最大本征值和第二本征值之间的差距就会变得很小。另外，α 的数值越小，表明用户越倾向于随机输入链接的方法浏览新的网页，也表明说明了系统的随机性就越强，因此导致最大本征值和第二本征值的差距减小 (1)。因此 Google 本征值谱的分布可以揭示出网络本身的一些性质。

由于篇幅和时间的限制，不可能对各种情形一一进行讨论。但是对 Google 矩阵的研究不失为一个很有研究价值的课题。

2 代码

```
% 本程序随机生成Google矩阵并求对应的本征值。
clear;clc;
N=3000; %维数
alpha=0.1; %阻尼系数
mean=10; %平均每个网页的连接数
M_matrix=zeros(N,N); %Markov矩阵
for i=1:N
    for j=1:N
        if(rand<mean/N)
            M_matrix(i,j)=1;
        end
    end
end
for i=1:N
    if(sum(M_matrix(i,:))~=0)
        M_matrix(i,:)=M_matrix(i,:)/sum(M_matrix(i,:));
    else
        M_matrix(i,:)=1/N*ones(1,N);
    end
```

```
end
G_matrix=(alpha*M_matrix+(1-alpha)/N*ones(N,N))';
[vector,value]=eig(G_matrix);
[spectrum,index]=sort(diag(value),'descend');
vector_sort=vector(:,index);
probability=abs(vector(:,1));
figure(1)
plot(1:N,abs(spectrum),'r');
title('The module of the eigenvalues of Google matrix(alpha=0.1)');
xlabel('label $i$','interpreter','latex');
ylabel('$|\lambda|_i$','interpreter','latex');
figure(2)
scatter(real(spectrum),imag(spectrum),2,'b','filled');
title('The spectrum of the eigenvalues of Google matrix(alpha=0.1)');
xlabel('Re($\lambda$)','interpreter','latex');
ylabel('Im($\lambda$)','interpreter','latex');
page_rank=abs(sort(vector_sort(:,1),'descend'));
```

参考文献

[1] Georgeot B, Giraud O, Shepelyansky D L. Spectral properties of the Google matrix of the World Wide Web and other directed networks[J]. Physical Review E, 2010, 81(5): 056109.

[2] Giraud O, Georgeot B, Shepelyansky D L. Delocalization transition for the Google matrix[J]. Physical Review E, 2009, 80(2): 026107.

[3] Albert R, Barabási A L. Topology of evolving networks: local events and universality[J]. Physical review letters, 2000, 85(24): 5234.