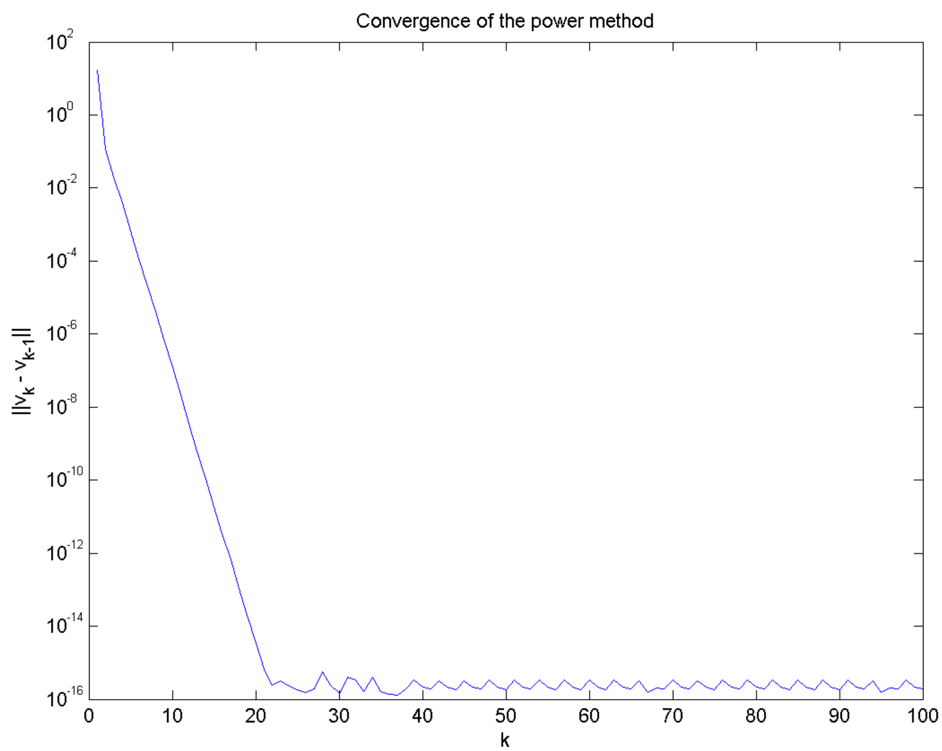
1 PageRank

1.

The indexes of the three webpages with largest indexes are: 873, 50 and 81.



2.

$$
\begin{aligned}
v_k &= \frac{T^T v_{k-1}}{||T^T v_{k-1}||} \\
&= C_1 T^T v_{k-1} \\
&= C_2 (T^T)^2 v_{k-2} \\
&\vdots \\
&= C_k (T^T)^k v_0 \\
&= \frac{(T^T)^k v_0}{||(T^T)^k v_0||} \quad \text{since } ||v_k|| = 1
\end{aligned}
$$

We also have for k large enough two possible approximations:

$$T^k v_0 = a_1\pi + \sum_{i=2}^{n} a_i \lambda_i^k u_i \approx a_1\pi$$

but also

$$\sum_{i=2}^{n} a_i \lambda_i^k u_i \approx a_2 \lambda_2^k u_2$$

So we have:

$$\begin{aligned}
||v_k - v_{k-1}|| &= ||\frac{a_1\pi + \sum_{i=2}^n a_i\lambda_i^k u_i}{||a_1\pi + \sum_{i=2}^n a_i\lambda_i^k u_i||} - \frac{a_1\pi + \sum_{i=2}^n a_i\lambda_i^{k-1} u_i}{||a_1\pi + \sum_{i=2}^n a_i\lambda_i^{k-1} u_i||}|| \\
&\approx ||\frac{a_1\pi + \sum_{i=2}^n a_i\lambda_i^k u_i}{|a_1|} - \frac{a_1\pi + \sum_{i=2}^n a_i\lambda_i^{k-1} u_i}{|a_1|}|| \\
&= \frac{1}{|a_1|}||a_1\pi + \sum_{i=2}^{n} a_i\lambda_i^k u_i - a_1\pi - \sum_{i=2}^{n} a_i\lambda_i^{k-1} u_i|| \\
&= \frac{1}{|a_1|}||\sum_{i=2}^{n} a_i(\lambda_i^k - \lambda_i^{k-1})u_i|| \\
&\approx \frac{1}{|a_1|}||a_2(\lambda_2^k - \lambda_2^{k-1})u_2|| \\
&= \left|\frac{a_2}{a_1}\right| \cdot |\lambda_2^{k-1}| \cdot |\lambda_2 - 1| \cdot ||u_2|| \\
&= \left|\frac{a_2}{a_1}\right| \cdot |\lambda_2^{k-1}| \cdot |\lambda_2 - 1| \quad (||u_2|| = 1) \\
&= \left|\frac{a_2}{a_1}(\lambda_2 - 1)\right| \cdot |\lambda_2|^{k-1}
\end{aligned}$$

It is worth noting at this point that for large values of k the logarithm of the difference computed above decreases linearly.

$$\begin{aligned}
\ln ||v_k - v_{k-1}|| &\approx \ln\left|\frac{a_2}{a_1}(\lambda_2 - 1)\right| + (k-1)\ln|\lambda_2| \\
&= \ln|\lambda_2|k + \ln\left|\frac{a_2}{a_1}(\lambda_2 - 1)\right| - \ln|\lambda_2| \\
&= Ak + B \quad \text{with} \quad A = \ln|\lambda_2| \quad \text{and} \quad B = \ln\left|\frac{a_2}{a_1}(1 - \frac{1}{\lambda_2})\right|
\end{aligned}$$

For a very small value of k, the result obtained above doesn't seem to reflect what is going with the plot. It is likely that our simplifications do not hold for such small values of k. For $5 \leq k \leq 20$ the result seems to be satisfied, but for $k > 20$ we get strange results. This is because the power method has converged to machine precision ($\approx 10^{-16}$) and all that we are looking at are roundoff errors, which are random in nature.

To estimate the value of k, we may look at the part of the plot where the behaviour is linear, and find the slope of that line. Then we may estimate

$$k = e^{slope}$$