1. First note that $Z = \frac{1-\theta^N}{1-\theta} \simeq \frac{1}{1-\theta}$ for large $N$.

a) The weights defined in class are given in this case by $w_i = \frac{\pi_i}{\psi_i} = \frac{N}{Z} \theta^{i-1}$, so that for $j \neq i$:

$$a_{ij} = \min \left( 1, \frac{w_j}{w_i} \right) = \min \left( 1, \theta^{j-i} \right) = \begin{cases} 1 & \text{if } j < i \\ \theta^{j-i} & \text{if } j > i \end{cases}$$

which leads to

$$p_{ij} = \begin{cases} \frac{1}{N} \theta^{j-i} & \text{if } j < i \\ \frac{1}{N} + \frac{1}{N} \sum_{k=i+1}^{N} (1 - \theta^{k-i}) & \text{if } j = i \end{cases}$$

b) From the course, we know that

$$\|P^n_i - \pi\|_{TV} \leq \frac{\lambda_1^n}{2\sqrt{\pi_i}}$$

where

$$\lambda_1 = 1 - \frac{1}{w_*} \quad \text{and} \quad w_* = \max_{i \in S} w_i = w_1 = \frac{N}{Z}$$

We conclude therefore that

$$\|P^n_i - \pi\|_{TV} \leq \frac{\sqrt{Z}}{2 \sqrt{\theta^{i-1}}} \left( 1 - \frac{Z}{N} \right)^n$$

For $i = 1$ and large $N$, this bound leads to:

$$\|P^n_1 - \pi\|_{TV} \leq \frac{1}{2\sqrt{1-\theta}} \exp \left( -\frac{n}{N(1-\theta)} \right)$$

while for $i = N$ and large $N$, this bound leads to:

$$\|P^n_N - \pi\|_{TV} \leq \frac{1}{2\sqrt{(1-\theta) \theta^{N-1}}} \exp \left( -\frac{n}{N(1-\theta)} \right) = \frac{1}{2\sqrt{1-\theta}} \exp \left( \frac{N-1}{2} \log(1/\theta) - \frac{n}{N(1-\theta)} \right)$$

c) Because of the last estimate, in order for $\max_{i \in S} \|P^n_i - \pi\|_{TV}$ to be smaller than $\epsilon$, we need that $n \gg N^2$, which gives the desired upper bound on the mixing time. What can actually be shown in this case (but this was not asked) is the following: using the more precise estimate

$$\|P^n_i - \pi\|_{TV} \leq \frac{1}{2} \sqrt{\sum_{k=1}^{N-1} \lambda_k^{2n} \left( \phi^{(k)}_1 \right)^2}$$

we find that this quantity is small (uniformly in $i$) for $n \gg N$ already.
2. a) The transition probabilities are given by
\[ p_{01} = \psi_{01} \min \left( 1, \frac{\psi_{01}}{\psi_{01}} \right) = e^{-2 \beta} \quad p_{21} = \psi_{21} \min \left( 1, \frac{\psi_{21}}{\psi_{21}} \right) = e^{-\beta} \]
\[ p_{10} = \psi_{10} \min \left( 1, \frac{\psi_{10}}{\psi_{10}} \right) = \frac{1}{2} \quad p_{12} = \psi_{12} \min \left( 1, \frac{\psi_{12}}{\psi_{12}} \right) = \frac{1}{2} \]
\[ p_{02} = p_{20} = p_{11} = 0 \quad p_{00} = 1 - e^{-2 \beta} = \frac{1}{2} \quad p_{22} = 1 - e^{-\beta} \] (1)

b) Let us now check that the detailed balance equation is satisfied:
\[ p_{01} \pi_0 = \frac{1}{2} e^{-2 \beta} = p_{10} \pi_1 \]
\[ p_{02} \pi_0 = 0 = p_{20} \pi_2 \]
\[ p_{12} \pi_1 = \frac{1}{2} e^{-\beta} = p_{21} \pi_2. \]

c) As usual, there are several methods to compute the eigenvalues. For example, one can find the three solutions \( \lambda_0, \lambda_1, \) and \( \lambda_2 \) to the equation
\[ \det(P - \lambda I) = 0, \] (2)
where \( I \) is the \( 3 \times 3 \) identity matrix and \( P \) the matrix of the transition probabilities computed in (1).

Another (perhaps even simpler method) method is to solve the following system of equations:
\[ \begin{cases} 
\lambda_0 &= 1 \\
\lambda_0 + \lambda_1 + \lambda_2 &= \text{tr}(P) , \\
\lambda_0 \cdot \lambda_1 \cdot \lambda_2 &= \det(P) 
\end{cases} \]
as we know that the largest eigenvalue is 1, the sum of the eigenvalues equals the trace of \( P \), and their product equals the determinant of \( P \).

Consequently, we obtain
\[ \begin{cases} 
\lambda_0 &= 1 \\
\lambda_1 &= -\frac{e^{-2 \beta}}{4} - \frac{e^{-\beta}}{4} + \frac{1}{2} + \frac{1}{4} \sqrt{e^{-4 \beta} - 2e^{-3 \beta} + e^{-2 \beta} + 4} , \\
\lambda_2 &= -\frac{e^{-2 \beta}}{4} - \frac{e^{-\beta}}{4} + \frac{1}{2} - \frac{1}{4} \sqrt{e^{-4 \beta} - 2e^{-3 \beta} + e^{-2 \beta} + 4} 
\end{cases} \]

d) The spectral gap is given by
\[ \gamma = 1 - \lambda_1 = \frac{1}{2} + \frac{e^{-2 \beta}}{4} + \frac{e^{-\beta}}{4} - \frac{1}{4} \sqrt{e^{-4 \beta} - 2e^{-3 \beta} + e^{-2 \beta} + 4}. \] (3)

Therefore, when \( \beta \) is large, we have
\[ \gamma \approx \frac{1}{4} e^{-\beta}. \] (4)

Remark. The value of \( \beta \) has to be tuned carefully and there is an inherent trade-off in its choice. If we pick \( \beta \) too large, then the spectral gap is small and the convergence to the global minimum occurs very slowly, which reflects the fact that we might get stuck in the local minimum (=state 2). On the other hand, if we pick \( \beta \) too small, then convergence is fast, but the stationary distribution is close to uniform in this case, so there is no guarantee that we land in the global minimum either!