1) It is located inside the src folder in the PageRank and PageRankMain classes.

2) The large file was handled first by reading in the compressed and then modifying the PageRank code given so that it would not be O(n^2) this greatly improved run time allowing the large file to be processed faster than it otherwise would have been able to be. In addition, as it relates to counting the terms I gathered the information need to find this data during the code as to not slow down the code and then processed this information using O(n) time to make it fast enough. The main issue that came up in the code was using inefficient methods and algorithms that had to be reworked to improve efficiency.

3) The libraries used for this code are first java.util.HashMap, java.util.Map, java.util.Set, java.util.ArrayList, and java.util.HashSet. All these were used to store and manipulate the data gathered. Next are java.io.BufferedReader, java.io.File, java.io.FileInputStream, java.util.zip.GZIPInputStream, java.io.PrintWriter, java.io.FileWriter, and java.io.InputStreamReader. All of them were used to read and write to files. Finally java.io.FileNotFoundException, java.io.IOException, and java.io.UnsupportedEncodingException. These 3 were used to handle errors.

4) The inlinks and PageRank list are not identical. There are some similarities such as the index and the united states both being 1 and 2 respectively. However, the biggest difference is that years are ranked higher in inlinks than PageRank. In both lists, they are high up however they are higher in the inlinks list. This is most likely due to the fact that anything that happens in or relates to a year will link to it greatly boosting its inlink count compared to PageRank which is based on more than just links. Additionally, another difference is that the only list in the data list_of_countires is high ranked in inlinks but is not present at all in the PageRank list. I believe the likely reason for this is because while many things might link to a list of countries most of the things that link to it will not be high ranked themselves as a list of countries is not important to other high ranked pages like years of technical terms like Population_density.

5) If you set all initial ranks to 0 the PageRank that would take a similar amount of time with no great time difference detectable however the results will have significant differences. As for random values between 0 and 1, it will never converge or not in a reasonable time at least and will therefore not provide and results. If the random surfer is eliminated it will run in a very similar time with no detectable time difference.