Homework Assignment

Yaghoub Shahmari - 98100883

4.1:

$$k_{max} = k_{min} N^{1/\gamma - 1}$$

Network	Internet	Science Collaboration	Actor Network	Protein Interactions
$k_m ax$	153	72	166019	56

4.2:

a.

The normalization condition is:

$$\int_{k_{\min}}^{k_{\max}} q_k \, dk = \int_{k_{\min}}^{k_{\max}} p_k \, dk = 1$$

Plugging in $p_k = Ck^{-\gamma}$ and solving, we get:

$$C = \frac{\gamma - 1}{k_{\max}^{1-\gamma} - k_{\min}^{1-\gamma}} \Rightarrow A = \frac{\gamma - 2}{\gamma - 1} \frac{k_{\max}^{1-\gamma} - k_{\min}^{1-\gamma}}{k_{\max}^{2-\gamma} - k_{\min}^{2-\gamma}}$$

b.

$$\langle k \rangle_q = \int_{k_{\min}}^{k_{\max}} k q_k \, dk = AC \int_{k_{\min}}^{k_{\max}} k^{1-\gamma+1} \, dk = \frac{\gamma - 2}{3 - \gamma} \frac{k_{\max}^{3-\gamma} - k_{\min}^{3-\gamma}}{k_{\max}^{2-\gamma} - k_{\min}^{2-\gamma}}$$

c.

$$\langle k \rangle_p = C(\gamma - 2)(k_{\max}^{2-\gamma} - k_{\min}^{2-\gamma})$$

Plugging in the given values $\gamma = 2.1$, $k_{\min} = 1$, $k_{\max} = \sqrt{N} \approx 10^4$, we get:
$\langle k \rangle_p = 0.446$ and $\langle k \rangle_q = 61.2$
So $\langle k \rangle_q > \langle k \rangle_p$ showing the friendship paradox.

d.

This "friendship paradox" occurs because high degree nodes are more likely to be connected to any given node. So when choosing a neighbor at random, we are more likely to select these high degree nodes compared to if we selected nodes randomly. So the neighbors tend to have higher degree on average than the nodes themselves.

Summary of the reference:

This article empirically analyzes the distribution of citations for scientific publications using two large datasets - over 780,000 papers published in 1981 catalogued by the Institute for Scientific Information (ISI) and over 24,000 papers published in Physical Review D journals from 1975-1994. By constructing a Zipf plot that ranks papers by number of citations, the author finds evidence that the asymptotic tail of the citation distribution follows a power law, with the number of papers cited x times decaying as $x^{-\alpha}$, where α is approximately 3. This suggests that a small fraction of papers are cited very heavily, while most papers receive few citations. The continuing evolution of citation statistics over time is also highlighted.