

## Supplementary figures

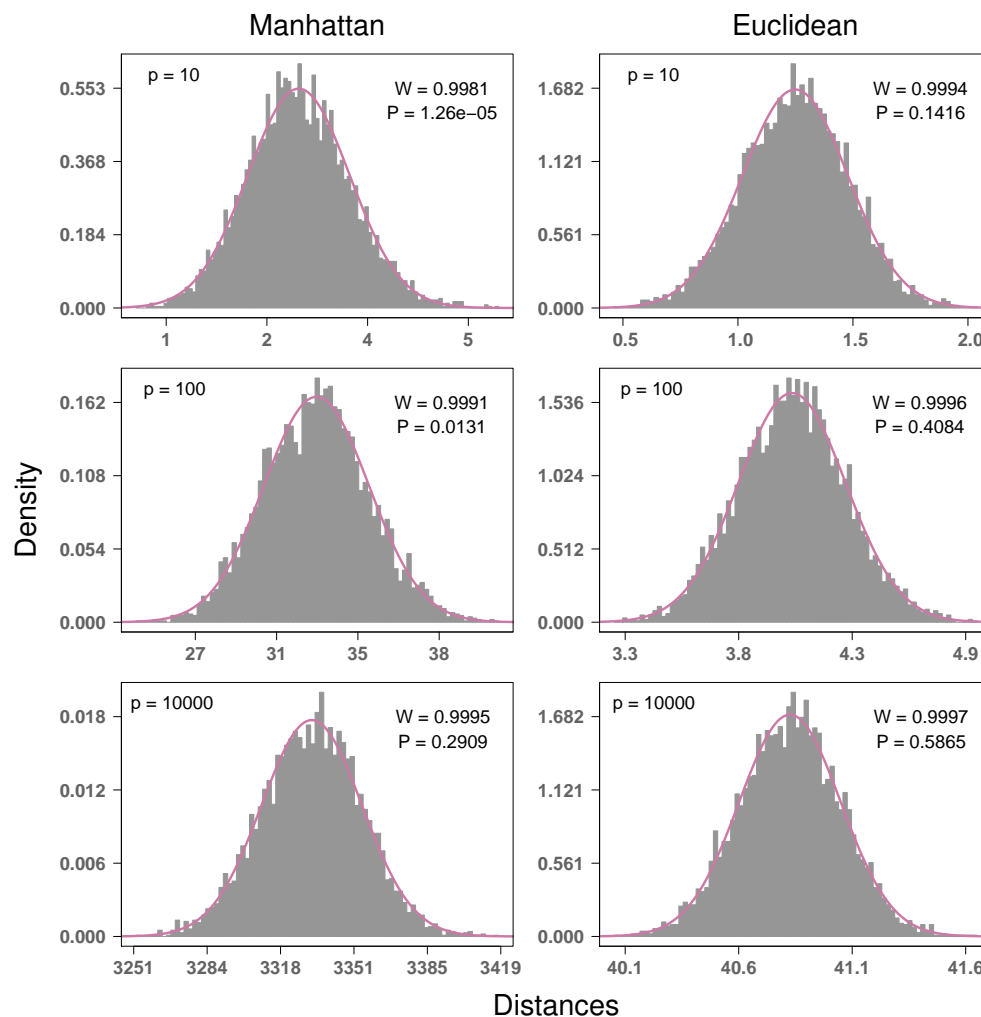
[illegible]





Figure 1 displays four histograms showing the distribution of distances between the two most similar sequences for different sample sizes ( $p$ ). The x-axis represents 'Distances' and the y-axis represents 'Density'. Each plot includes a red curve representing a theoretical distribution.

- Top Left ( $p = 10$ ):** The distribution is highly skewed, with a peak density of approximately 2.369. The parameters are  $W = 0.9551$  and  $P = 2.2e-16$ .
- Top Right ( $p = 100$ ):** The distribution is more spread out, with a peak density of approximately 0.157. The parameters are  $W = 0.9956$  and  $P = 4.3e-11$ .
- Bottom Left ( $p = 1000$ ):** The distribution is more concentrated, with a peak density of approximately 0.030. The parameters are  $W = 0.9994$  and  $P = 0.0761$ .
- Bottom Right ( $p = 10000$ ):** The distribution is very concentrated, with a peak density of approximately 0.008. The parameters are  $W = 0.9997$  and  $P = 0.6533$ .

[illegible]

Figure 2 displays four histograms showing the distribution of distances for different sample sizes ( $p$ ), with a red curve representing a normal distribution fit. The x-axis is labeled "Distances" and the y-axis is labeled "Density".

- Top Left ( $p = 10$ ):** The distribution is highly skewed and sparse. The y-axis ranges from 0.000 to 4.502. The x-axis ranges from 0 to 4. The fit parameters are  $W = 0.9637$  and  $P = 2.2e-16$ .
- Top Right ( $p = 100$ ):** The distribution is more concentrated and bell-shaped. The y-axis ranges from 0.000 to 0.315. The x-axis ranges from 19 to 33. The fit parameters are  $W = 0.9971$  and  $P = 4.0e-11$ .
- Bottom Left ( $p = 1000$ ):** The distribution is very smooth and bell-shaped. The y-axis ranges from 0.000 to 0.051. The x-axis ranges from 213 to 279. The fit parameters are  $W = 0.9995$  and  $P = 0.2096$ .
- Bottom Right ( $p = 10000$ ):** The distribution is extremely smooth and bell-shaped. The y-axis ranges from 0.000 to 0.014. The x-axis ranges from 2365 to 2565. The fit parameters are  $W = 0.9998$  and  $P = 0.8876$ .

[illegible]

Figure 2 displays four histograms showing the distribution of distances between the two most similar sequences for different sample sizes ( $p$ ). The x-axis represents Distances, and the y-axis represents Density. Each plot includes a pink normal distribution curve and statistical values  $W$  and  $P$ .

- Top Left ( $p = 10$ ):** The distribution is highly skewed to the right.  $W = 0.9790$ ,  $P = 2.2e-16$ .
- Top Right ( $p = 100$ ):** The distribution is more symmetric but still shows some skewness.  $W = 0.9986$ ,  $P = 0.0002$ .
- Bottom Left ( $p = 1000$ ):** The distribution is nearly symmetric and bell-shaped.  $W = 0.9996$ ,  $P = 0.3184$ .
- Bottom Right ( $p = 10000$ ):** The distribution is very symmetric and bell-shaped.  $W = 0.9997$ ,  $P = 0.6069$ .

[illegible]

Figure 1 displays four histograms showing the distribution of distances between pairs of nodes for different values of  $p$  (10, 50, 150, and 300). The x-axis represents Distances, and the y-axis represents Density. Each plot includes a normal distribution fit curve (pink line) and statistical values  $W$  and  $P$ .

- Top Left ( $p = 10$ ):** The distribution is centered around a distance of approximately 105. The density ranges from 0.000 to 0.048. The fit curve is a normal distribution.  $W = 0.9985$ ,  $P = 0.0001$ .
- Top Right ( $p = 50$ ):** The distribution is centered around a distance of approximately 2784. The density ranges from 0.000 to 0.008. The fit curve is a normal distribution.  $W = 0.9996$ ,  $P = 0.4214$ .
- Bottom Left ( $p = 150$ ):** The distribution is centered around a distance of approximately 25304. The density ranges from 0.000 to 0.003. The fit curve is a normal distribution.  $W = 0.9996$ ,  $P = 0.3407$ .
- Bottom Right ( $p = 300$ ):** The distribution is centered around a distance of approximately 101150. The density ranges from 0.000 to 0.001. The fit curve is a normal distribution.  $W = 0.9997$ ,  $P = 0.5533$ .

October 14, 2019

Figure 1 displays four histograms showing the distribution of distances for different values of  $p$  (10, 50, 150, and 300). Each plot includes a grey histogram, a pink normal distribution curve, and statistical values  $W$  and  $P$ .

- Top Left ( $p = 10$ ):** The x-axis ranges from 11 to 19. The y-axis (Density) ranges from 0.000 to 0.287.  $W = 0.9985$ ,  $P = 0.0001$ .
- Top Right ( $p = 50$ ):** The x-axis ranges from 350 to 395. The y-axis (Density) ranges from 0.000 to 0.061.  $W = 0.9996$ ,  $P = 0.5309$ .
- Bottom Left ( $p = 150$ ):** The x-axis ranges from 3113 to 3231. The y-axis (Density) ranges from 0.000 to 0.024.  $W = 0.9996$ ,  $P = 0.4321$ .
- Bottom Right ( $p = 300$ ):** The x-axis ranges from 12067 to 12305. The y-axis (Density) ranges from 0.000 to 0.012.  $W = 0.9997$ ,  $P = 0.4653$ .

The x-axis for all plots is labeled "Distances" and the y-axis is labeled "Density".

October 14, 2019