













**Figure S4: Probability density functions, rank-based distributions (Zipf's plots) and Heaps' plots for all the 35 data sets shown in Table 1.** The indices in the plots are exactly the same as the ones in Table 1. Some data sets are small, and thus the fittings can only cover the blue areas. Especially the H1N1 data set (No. 34) is very small and thus only the Zipf's plot can provide some statistics, while the probability density function cannot be well fitted.