Parsing tweets as spikes

Luíño F. Seoane

June 13, 2013

Using the scripts uploaded in the website we can parse and perform some basic statistics about Inter-Tweet Intervals (ITIs). We can compute the ITI, its average, standard deviation, and Fano factor; definitely showing that the distribution of tweets does not follow a Poisson distribution. The average results for the three data sets available are outlined in the following table:

	Mean ITI	Std ITI	Fano
Data1	363.44	1016.09	2840.76
Data2	191.10	900.10	4239.52
Data3	584.15	1464.61	3672.15

Dismissing any discussion about uncertainty, all results were just rounded to the second decimal place.

We feature further results in figures 1 and 2. In the former, data of ITI are plotted in doubled log scale and fitted to a power law. We do the same for the inverse cumulative distribution ($P(x > X)$, where x is the stochastic variable denoting Inter-Tweet time) in figure 2. In this figure it is made clear that the distribution is not a power-law either, being the next natural suggestion to go for a log-normal distribution.

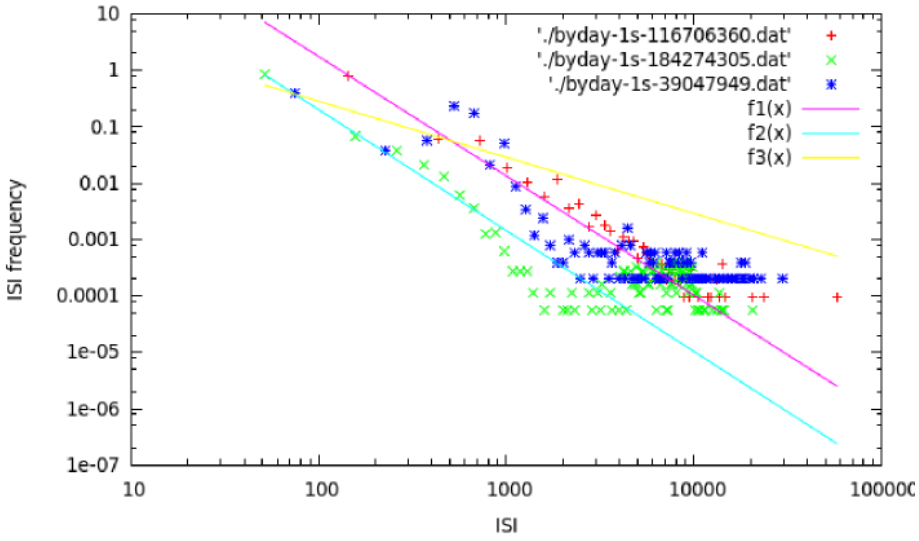

Figure 1: **Log-log plot of the ITI distribution for the three real data-sets available.** Fitting to power-laws could seem acceptable except for the third data set. The outcome of the fits (made with gnuplot) are largely biased by the first data in the series, explaining the ugly fit for this last data set.

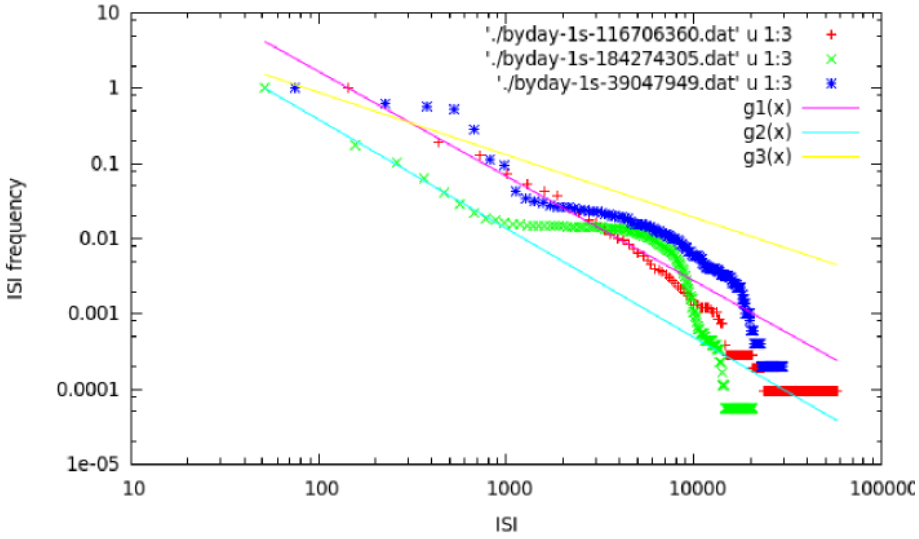

Figure 2: **Log-log plot of the complementary cumulative ITI distribution for the three real data-sets available.** If the ITI were indeed following a power-law, this plot should be scale free in an even more evident way. Instead, we find no scaling region. This suggests that the ITI might follow a log-normal distribution, which is also quite interesting.