S&P 500 write-up

We use the list of S&P 500 constituents as of Jun 28, 2017 (505 companies in total) to download the full range of historical daily adjusted close price data (dating back to 1970) for these companies. Then to find companies that exist throughout the whole time period from 1985 to now, we pick the companies that have data through the first 6 trading days in the year 1985, in total 149 companies.

We construct minimum spanning trees based on correlation matrices on daily return in 100-day windows. In case of dirty data, we only consider stocks with more than 95 valid data points in each 100-day window, and we fill the non-existent data points with the last available value, and if the non-existent data points are in the front of the window, we fill them with the next available values.

We then divide the stocks into clusters for each window based on the minimum spanning tree. Then in each cluster we pick the single stock with the highest betweeness measure. With this as the universe on that date, we construct an optimized portfolio using Markowitz’s method. The portfolio is re-balanced every two weeks (10 trading days).

SP500NewmanBetweennessWeighted.pdf