

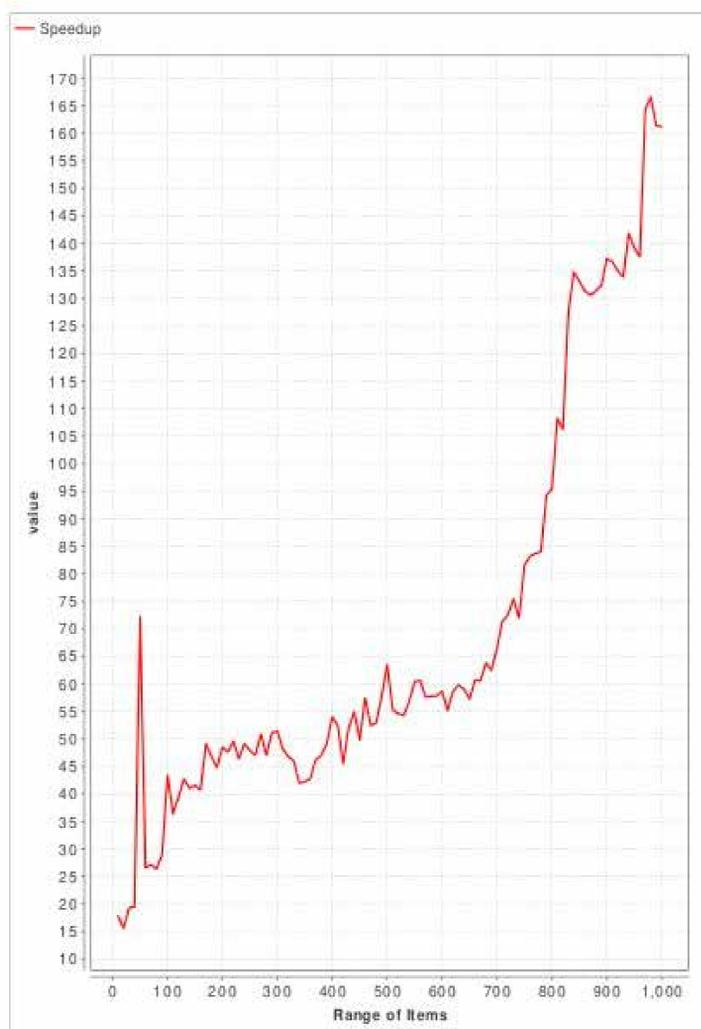
# Counting Coincidences Makes Sense.

## How to be objective

The Associator generates a graded list of Association Rules from binominal data, using a parallel implementation of the A Priori algorithm to generate Frequent Itemsets, and Iterative Proportional Fitting to standardise the contingency tables that the Itemsets contain, and from which rules are extracted.

This standardisation results in all measures of interest ranking the contingency tables in the same order. The implementation extends the work that appears in Tan, Pang-Ning; Kumar, Vipin; and Srivastava, Jaideep; "Selecting the right objective measure for association analysis."

## How to be fast enough



**Frequent Itemset mining:** early GPU based implementations using CUDA+Thrust accelerated a RapidMiner application by more than 100 times using modest onboard graphics cards.

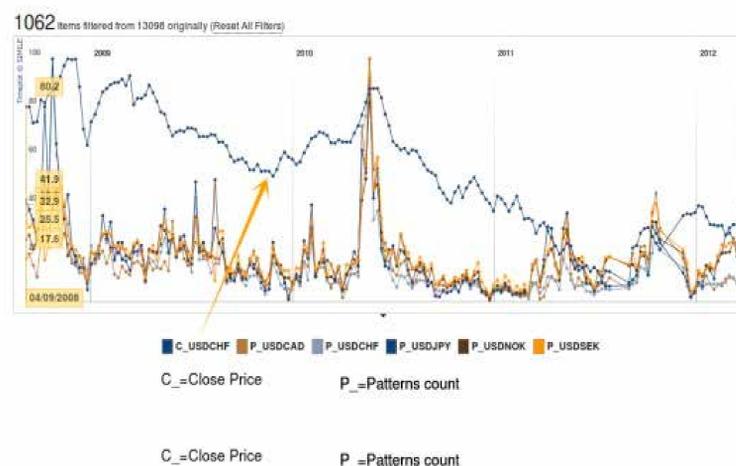
Crucially, the acceleration is proportional to the dimensions of the problem.

## Use case - Intermarket Dynamics

A comprehensive survey of the billions of combinations of 37 major currency pairs, minute to minute, over the post crash years reveals clear spikes in identifiable patterns.

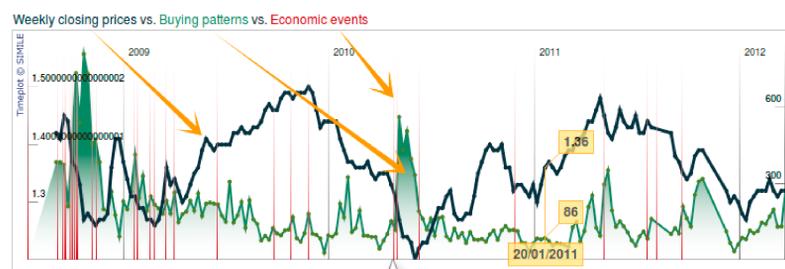
The analysis matches rises for a particular currency pair in one minute against rises in **any** pairs in the **previous** minute; any match that recurs sufficiently frequently is then counted as a pattern.

*Example: Pattern spikes that signal change in the outlook for the Dollar against the Swiss Franc*



*Example: pattern spikes that signal change on EUR-USD, and tie in with real-world events.*

### Cues for a next minute rise for EURUSD



Sources: Tradestation for prices, ParaCue for Economic commentary.

Move your mouse over the plots to

**Greek bailout, EFSF created**

On May 2, the EU and IMF announce a \$146 billion financial rescue package for Greece to address its sovereign debt crisis in exchange for the country enacting strict austerity measures. Less than two weeks later, the EU and IMF agree to create a temporary eurozone stability mechanism—the European Financial Stability Facility—worth \$1 trillion. The move comes in conjunction with a decision by the European Central Bank to buy eurozone government bonds on the open market in an effort to provide an added safety net for the euro

Comparing the effect of those two events on different currency pairs also provides an insight into the perceived distance between that market and the epicenter of the shock.

For example, the Lehmans collapse hit GBPUSD much harder than the first Greek bailout, whereas exactly the reverse is true for EURGBP.

**More generally**

Data-mining fast moving markets needs a gear change, not more statistical assumptions!

Parallel counting enables empirical analysis previously deemed beyond timely reach, and is already revealing minute to minute patterns of crowd re-action in Foreign Exchange markets that are familiar to traders, but very difficult to reconcile with statistical models.

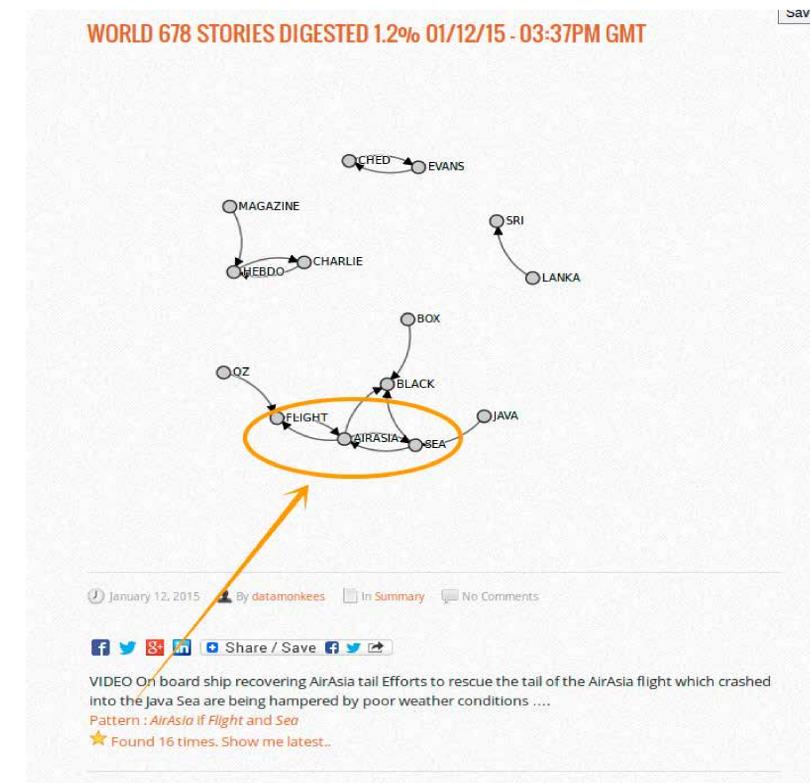
## Use case – Key Theme Identification

Putting key words into a search engine will return a collection of relevant documents.

But what if you already have a collection of documents, and need to identify the key phrases that they contain?

The News Digest at meme-machines.com looks for exactly such keyword patterns 24/7

*Example: Hundreds of news stories, four big themes, fourteen words.*



Association Rules can be combined and accumulated into usable rulebases. So the meme-machines.com News Desk also records the 'ifs' and 'thens' for later research.

*Example: What words connect to 'Ebola', and how?*

```
?- lookup('Ebola',If-Then).
If = ['Africa', 'Chimerix', 'City', 'Contact', 'Doctor', 'Health', 'Hospital', 'Liberia', 'Nurse', 'Outbreak', 'Patient', 'Quarantine', 'US', 'Virus', 'West', 'York'],
Then = ['Africa', 'City', 'Leone', 'Nurse', 'Quarantine', 'Sierra', 'West', 'York'].
```