10-K Documetric Similarity

Dara O Shayda
dara@lossofgenerality.com

Nov 10th 2013

Prelude

This is an investigation into putative relationship between the 10-K filings for US public companies and their corresponding stock prices.

We show that the relationship is tangibly extant and not a conceptual nor a theoretical one.

We start by stablishing the Cosine Similarity between the Section 7 (Management Discussion) of about 1000 US companies, for 2012. We set the threshold at 60% for similarity between the said sub-sections of the filings.

Then we take the corresponding stock prices for 2013 and instead of focusing on a small period of time and wondering about fluctuations, we go straight to the Trend (de-noised) curves of these stocks and compute the distance between the curves (Canberra Metric) to establish a measure for similarity. 2013 was chosen as a lag of time to allow for reaction to the 10-K filing of 2012.

No other knowledge about the market or specific stock is assumed. No further assumptions.

The code below is a simple implementation of the above to verify the concepts and observe the limitations.

As we see below the relationship between the 10-K filings' Section 7 and Trend curves are real and small number of failures are observed.

In specific if the distance between two 10-K filings is small, then the distance between their corresponding Trend curves is also small in more rigorous manner:

$$d(s_1,\ s_2) \ll d(s_i,\ \text{tr}) \implies d'(T(s_1),\ T(s_2)) \ll d'(T(s_i),\ T(\text{tr}))\quad i = 1, 2$$

Where d is Cosine Similarity (or whatever metric used for similarity in 10-K document space), s_1 and s_2 are two stocks with similarity more than 60%, tr is a an outlier much further away from s_1 and s_2 than distance between s_1 and s_2, d' is the metric in Trend curve space of the stock (Canberra Distance) and T is the Trend function. $\ll$ stands for much less than.

If the stocks' Cosine Similarity for their 10-K filings form a connected sub-graph, then we test for max-min:

$$\max\left(d'\left(\left(T\left(s_i\right),\ T\left(s_j\right)\right)\right\}\right) \ll \min(d'\left(\left(T\left(s_i\right),\ T(\text{tr})\right)\right\}) \quad \forall\, i,\ j$$

s_i and s_j are connected in the Cosine Similarity graph of the 10-K filings. This equation basically says:

if the similarity connected stocks in 10-K space with Cosine Distance form a low-radius cluster which is afar from a random-pick outlier (tr), then their Trend curve space for price in Canberra Distance also forms a low-radius cluster which is afar from the Trend curve of the random-pick!

This is indeed a very weak off-shoot definition of isometry. Let's call this relative-isometry with pre-served outlier.

Remark 1: *The above test is not a full-proof test, but good enough to show that there is such a relation-ship as depicted above.*

Remark 2: *The above test fails in some occasions and indeed the statistics of such failures would indicate the good-ness of the said relationship.*

Remark 3: *Canberra Distance might not always be a good metric, therefore variations of such metric for similarity might result in much more accurate Documentric algorithms.*

To investigate further, a set of random-pick outliers were selected to investigate the statistical nature of the above algorithm. In above inequalities the term $d'\left(\left(T\left(s_i\right),\ T(\text{tr})\right)\right.$ was replaced by the mean of all such distances summing over the set of outliers, let's represent that by $\overline{d}'\left(\left(T\left(s_i\right),\ T(\text{tr})\right)\right.$ and therefore:

$$\max\left(d'\left(\left(T\left(s_i\right),\ T\left(s_j\right)\right)\right\}\right) \ll \min\left(\overline{d}'\left(\left(T\left(s_i\right),\ T(\text{tr})\right)\right\}\right) \quad \forall\, i,\ j$$

as the desired inequality to check for Documetric Similarities mapped into Trend curve similarities.

Results are quite similar, strong indication of a relationship between the 10-K filings and stock prince Trend curves.

Disclaimer

We provide 'cloud' for computing the machine learning algorithms and data mining for forecast, classifica tion and clustering applications. By no means we provide any guarantees for the models to be of certain accuracy or performance. It is the responsibility of the user to fine-tune the algorithms to produce the

desired results. For that matter we provide a large variation for a certain algorithm and different versions of the algorithm.

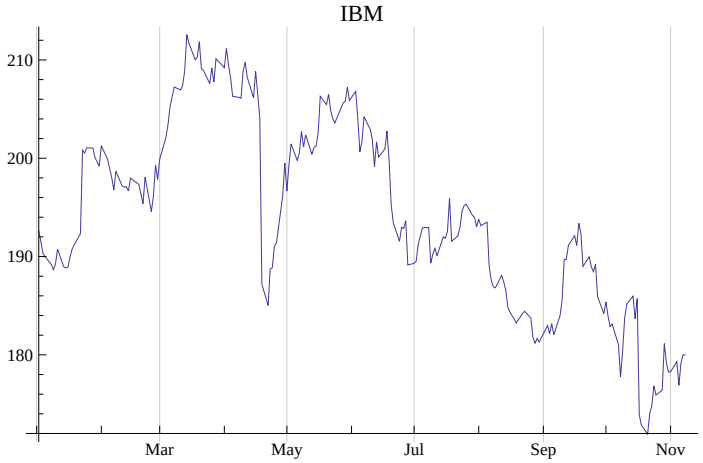
Trend & Wavelets

Choose a stock and download this year's price data:

In[2]:=
```
price = FinancialData["IBM", "Jan. 1, 2013"];
```

In[3]:=
```
DateListPlot[price, Joined → True, Frame → False, PlotLabel → "IBM"]
```

Out[3]=

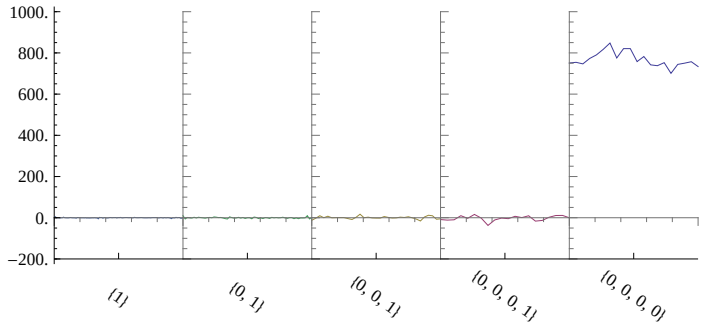


Choose a Wavelet basis to break down the price function above:

In[4]:=
```
dwd = DiscreteWaveletTransform[price[[All, 2]], BiorthogonalSplineWavelet[3, 3], 4];
```

In[5]:=
```
WaveletListPlot[dwd, PlotLayout → "CommonYAxis", Ticks → Full]
```

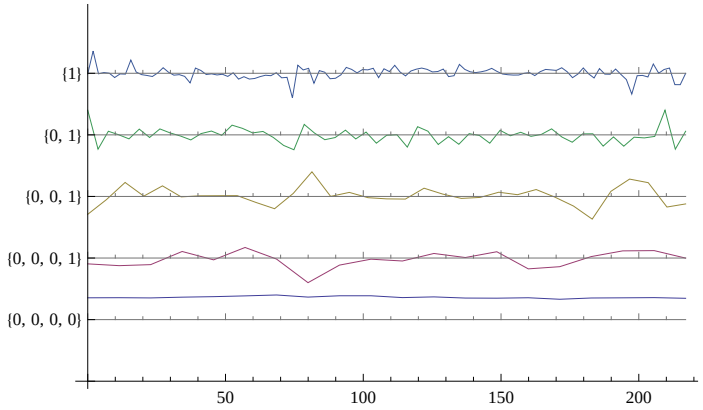
Out[5]=



{0, 0, 0, 0} is the Trend function for the price of the stock and the rest are noise levels:

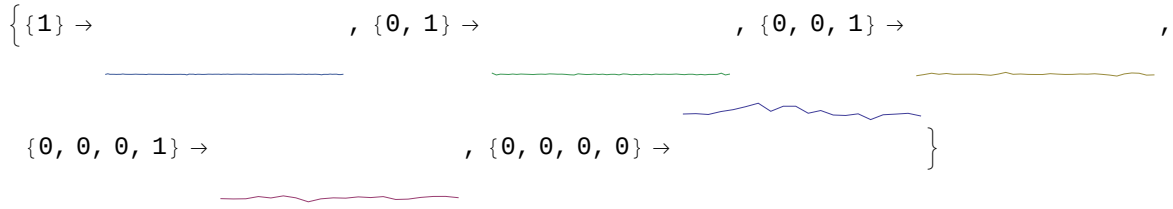
In[6]:=
```
WaveletListPlot[dwd, Ticks → Full]
```

Out[6]=



In[7]:=
```
dwd[Automatic, {"ListPlot"}]
```

Out[7]=



See which Wavelet index contributes most to the price function:

In[8]:=
```
dwd["EnergyFraction"]
```

Out[8]=
```
{{1} → 0.0000128351, {0, 1} → 0.0000475163,
 {0, 0, 1} → 0.000118123, {0, 0, 0, 1} → 0.000256589, {0, 0, 0, 0} → 0.999565}
```

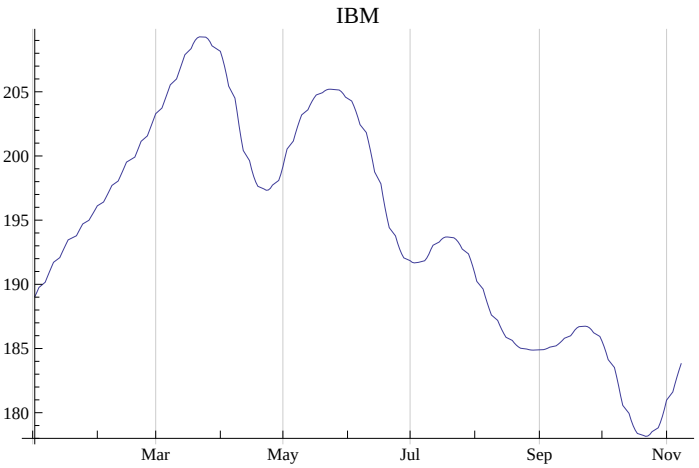
{0, 0, 0, 0} → 0.999565 clearly indicates that this index includes most of the information about the shape of the price function.

Inverse transform the Wavelet attempting to approximating the original price function by removing the noise and only taking into account the most contribution by the Trend:

In[9]:=
```
trIBM = InverseWaveletTransform[WaveletThreshold[dwd, {"Hard", 100}]];
```

In[10]:=
```
DateListPlot[Transpose[{price[[All, 1]], trIBM}],
 Joined → True, Frame → False, PlotLabel → "IBM"]
```

Out[10]=

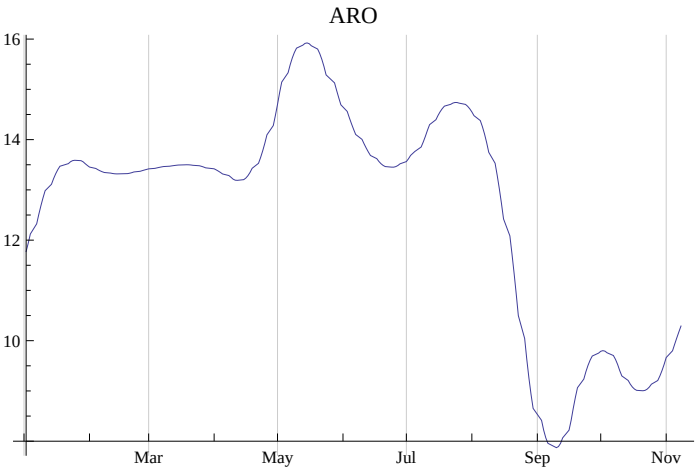


Trend & Similarity

Find the Trend curve for two stocks and by using Canberra metric compute the distance between the two curves and thus obtain a measure for their similarity:

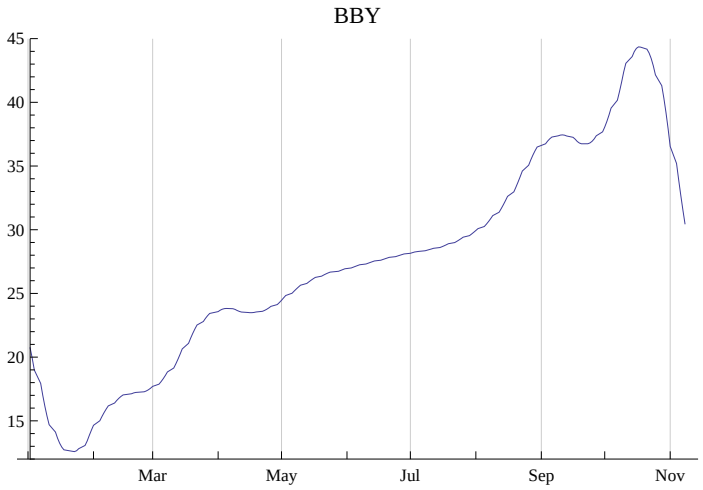
In[11]:=
```
price = FinancialData["ARO", "Jan. 1, 2013"];
dwd =
  DiscreteWaveletTransform[price[[All, 2]], BiorthogonalSplineWavelet[3, 3], 4];
trARO = InverseWaveletTransform[WaveletThreshold[dwd, {"Hard", 100}]];
DateListPlot[Transpose[{price[[All, 1]], trARO}],
 Joined → True, Frame → False, PlotLabel → "ARO"]
```

Out[14]=



In[15]:=
```
price = FinancialData["BBY", "Jan. 1, 2013"];
dwd =
  DiscreteWaveletTransform[price[[All, 2]], BiorthogonalSplineWavelet[3, 3], 4];
trBBY = InverseWaveletTransform[WaveletThreshold[dwd, {"Hard", 100}]];
DateListPlot[Transpose[{price[[All, 1]], trBBY}],
 Joined → True, Frame → False, PlotLabel → "BBY"]
```

Out[18]=



Find their Canberra distance:

In[19]:=
```
CanberraDistance[trARO, trBBY]
```

Out[19]= 73.7717

Random-Pick Outlier

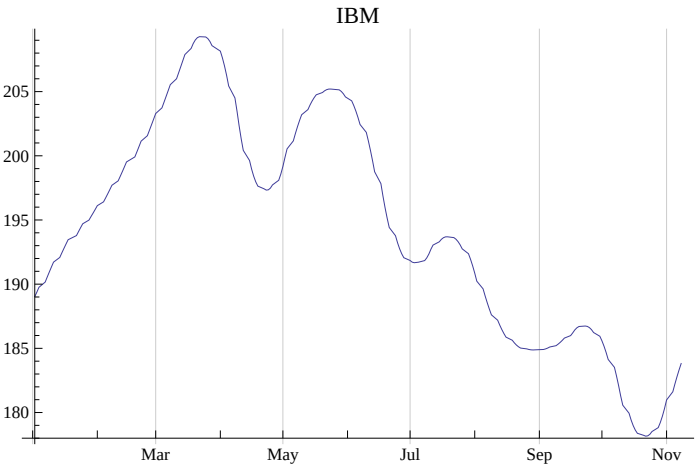
Now lets choose a random stock that has no Documetric Similarity to any other stocks and for that matter use as reference to see if the dis-similarity also persists between the price curves:

In[20]:=
```
tr = {};
price = FinancialData["IBM", "Jan. 1, 2013"];
dwd =
  DiscreteWaveletTransform[price[[All, 2]], BiorthogonalSplineWavelet[3, 3], 4];
AppendTo[tr, InverseWaveletTransform[WaveletThreshold[dwd, {"Hard", 100}]]];
DateListPlot[Transpose[{price[[All, 1]], Last[tr]}],
 Joined → True, Frame → False, PlotLabel → "IBM"]

price = FinancialData["ACO", "Jan. 1, 2013"];
dwd =
  DiscreteWaveletTransform[price[[All, 2]], BiorthogonalSplineWavelet[3, 3], 4];
AppendTo[tr, InverseWaveletTransform[WaveletThreshold[dwd, {"Hard", 100}]]];
DateListPlot[Transpose[{price[[All, 1]], Last[tr]}],
 Joined → True, Frame → False, PlotLabel → "ACO"]
```

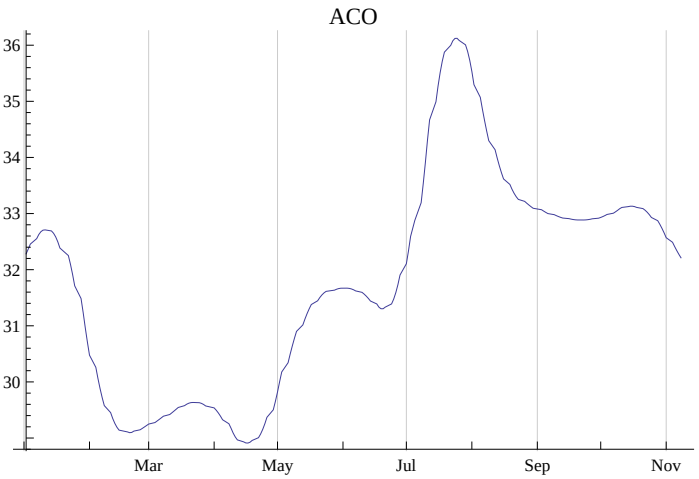
```
price = FinancialData["ALE", "Jan. 1, 2013"];
dwd =
  DiscreteWaveletTransform[price[[All, 2]], BiorthogonalSplineWavelet[3, 3], 4];
AppendTo[tr, InverseWaveletTransform[WaveletThreshold[dwd, {"Hard", 100}]]];
DateListPlot[Transpose[{price[[All, 1]], Last[tr]}],
 Joined → True, Frame → False, PlotLabel → "ALE"]

price = FinancialData["CATM", "Jan. 1, 2013"];
dwd =
  DiscreteWaveletTransform[price[[All, 2]], BiorthogonalSplineWavelet[3, 3], 4];
AppendTo[tr, InverseWaveletTransform[WaveletThreshold[dwd, {"Hard", 100}]]];
DateListPlot[Transpose[{price[[All, 1]], Last[tr]}],
 Joined → True, Frame → False, PlotLabel → "CATM"]

price = FinancialData["ARRY", "Jan. 1, 2013"];
dwd =
  DiscreteWaveletTransform[price[[All, 2]], BiorthogonalSplineWavelet[3, 3], 4];
AppendTo[tr, InverseWaveletTransform[WaveletThreshold[dwd, {"Hard", 100}]]];
DateListPlot[Transpose[{price[[All, 1]], Last[tr]}],
 Joined → True, Frame → False, PlotLabel → "ARRY"]

price = FinancialData["ADBE", "Jan. 1, 2013"];
dwd =
  DiscreteWaveletTransform[price[[All, 2]], BiorthogonalSplineWavelet[3, 3], 4];
AppendTo[tr, InverseWaveletTransform[WaveletThreshold[dwd, {"Hard", 100}]]];
DateListPlot[Transpose[{price[[All, 1]], Last[tr]}],
 Joined → True, Frame → False, PlotLabel → "ADBE"]

price = FinancialData["ACAD", "Jan. 1, 2013"];
dwd =
  DiscreteWaveletTransform[price[[All, 2]], BiorthogonalSplineWavelet[3, 3], 4];
AppendTo[tr, InverseWaveletTransform[WaveletThreshold[dwd, {"Hard", 100}]]];
DateListPlot[Transpose[{price[[All, 1]], Last[tr]}],
 Joined → True, Frame → False, PlotLabel → "ACAD"]

price = FinancialData["ALXA", "Jan. 1, 2013"];
dwd =
  DiscreteWaveletTransform[price[[All, 2]], BiorthogonalSplineWavelet[3, 3], 4];
AppendTo[tr, InverseWaveletTransform[WaveletThreshold[dwd, {"Hard", 100}]]];
DateListPlot[Transpose[{price[[All, 1]], Last[tr]}],
 Joined → True, Frame → False, PlotLabel → "ALXA"]


price = FinancialData["CRIS", "Jan. 1, 2013"];
dwd =
  DiscreteWaveletTransform[price[[All, 2]], BiorthogonalSplineWavelet[3, 3], 4];
AppendTo[tr, InverseWaveletTransform[WaveletThreshold[dwd, {"Hard", 100}]]];
DateListPlot[Transpose[{price[[All, 1]], Last[tr]}],
 Joined → True, Frame → False, PlotLabel → "CRIS"]

price = FinancialData["AVHI", "Jan. 1, 2013"];
dwd =
  DiscreteWaveletTransform[price[[All, 2]], BiorthogonalSplineWavelet[3, 3], 4];
AppendTo[tr, InverseWaveletTransform[WaveletThreshold[dwd, {"Hard", 100}]]];
```

```
DateListPlot[Transpose[{price[[All, 1]], Last[tr]}],
 Joined → True, Frame → False, PlotLabel → "AVHI"]

price = FinancialData["ASIA", "Jan. 1, 2013"];
dwd =
  DiscreteWaveletTransform[price[[All, 2]], BiorthogonalSplineWavelet[3, 3], 4];
AppendTo[tr, InverseWaveletTransform[WaveletThreshold[dwd, {"Hard", 100}]]];
DateListPlot[Transpose[{price[[All, 1]], Last[tr]}],
 Joined → True, Frame → False, PlotLabel → "ASIA"]
```

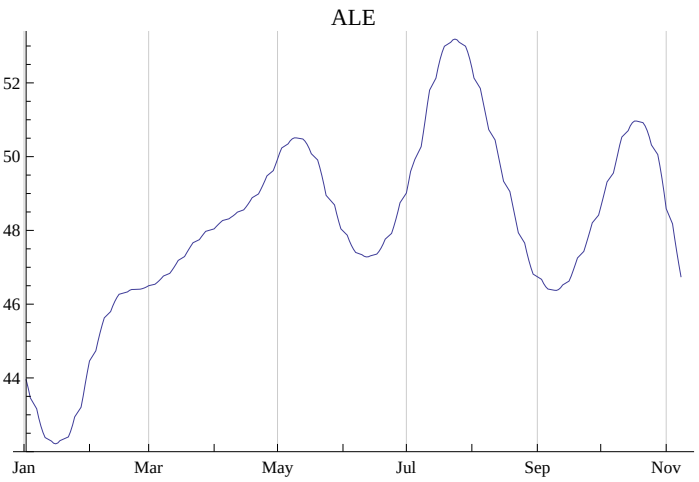
Out[24]=



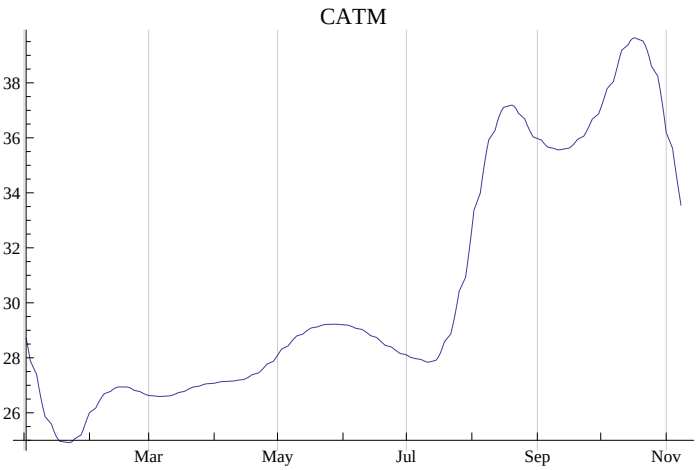
Out[28]=



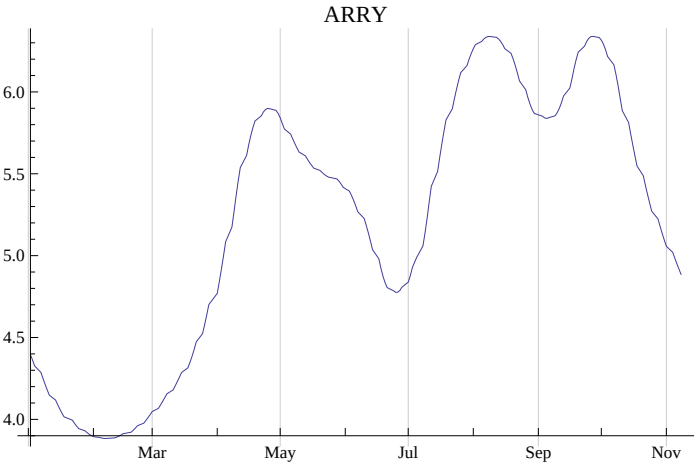
Out[32]=



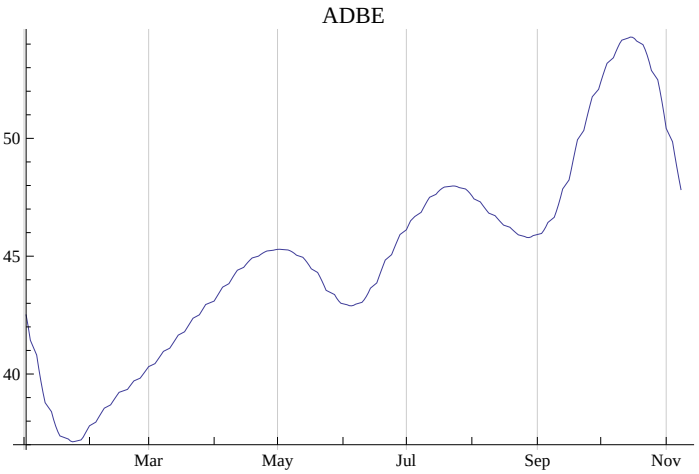
Out[36]=



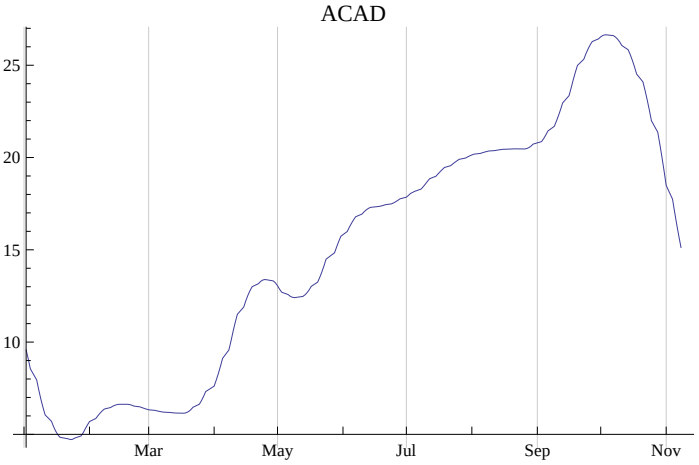
Out[40]=



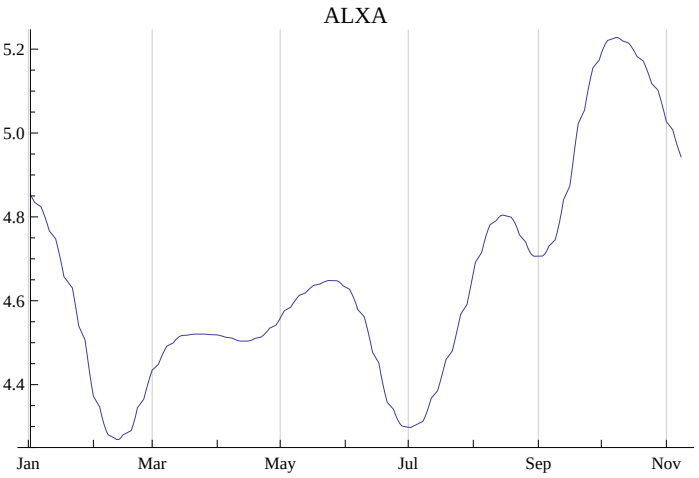
Out[44]=



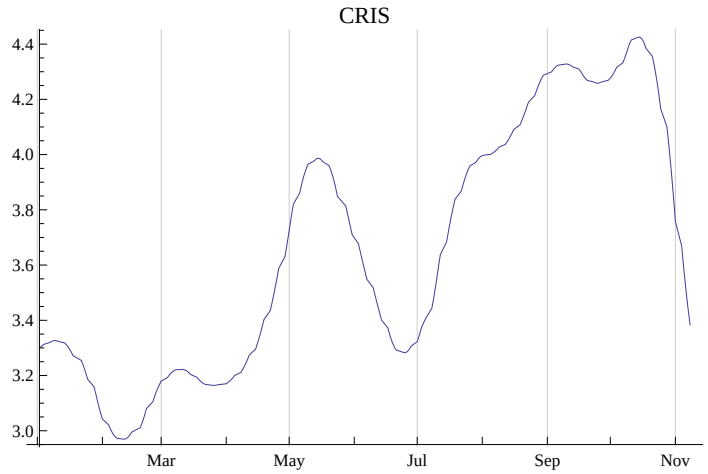
Out[48]=



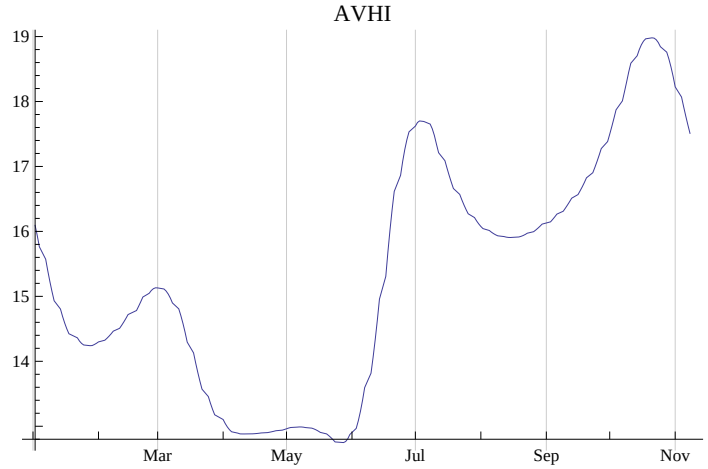
Out[52]=



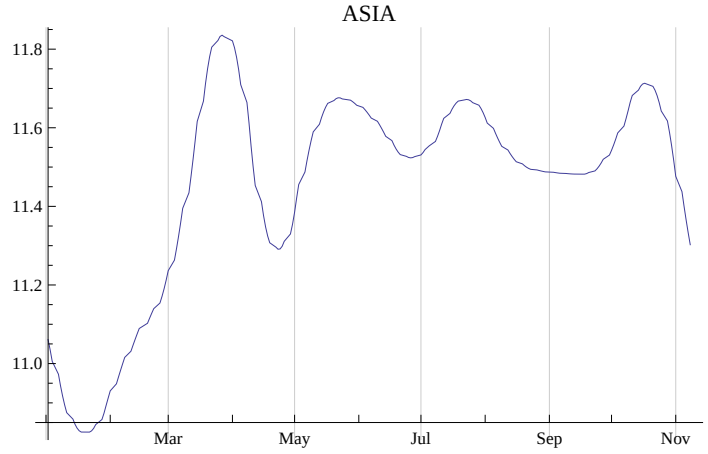
Out[56]=



Out[60]=



Out[64]=



Then compute the distance between each Trend curve and the random pick set, as you can see is much larger than the distance between the two Trend curves i.e. much dis-similar:

In[65]:=
```
data = Table[CanberraDistance[trARO, tr[[i]]], {i, 1, Length[tr]}]
Mean[data]
Sqrt[Variance[data]]
```

Out[65]= {190.637, 94.106, 127.139, 88.9482, 88.2489,
121.56, 57.3494, 97.8092, 117.139, 27.18, 21.5428}

Out[66]= 93.7872

Out[67]= 47.8149

Compute the mean distance to the random-pick outliers:

In[68]:=
```
data = Table[CanberraDistance[trBBY, tr[[i]]], {i, 1, Length[tr]}]
Mean[data]
Sqrt[Variance[data]]
```

Out[68]= {163.529, 29.2695, 63.2106, 18.7545, 144.419,
56.8086, 69.4445, 149.673, 162.93, 56.9357, 82.9226}

Out[69]= 90.7179

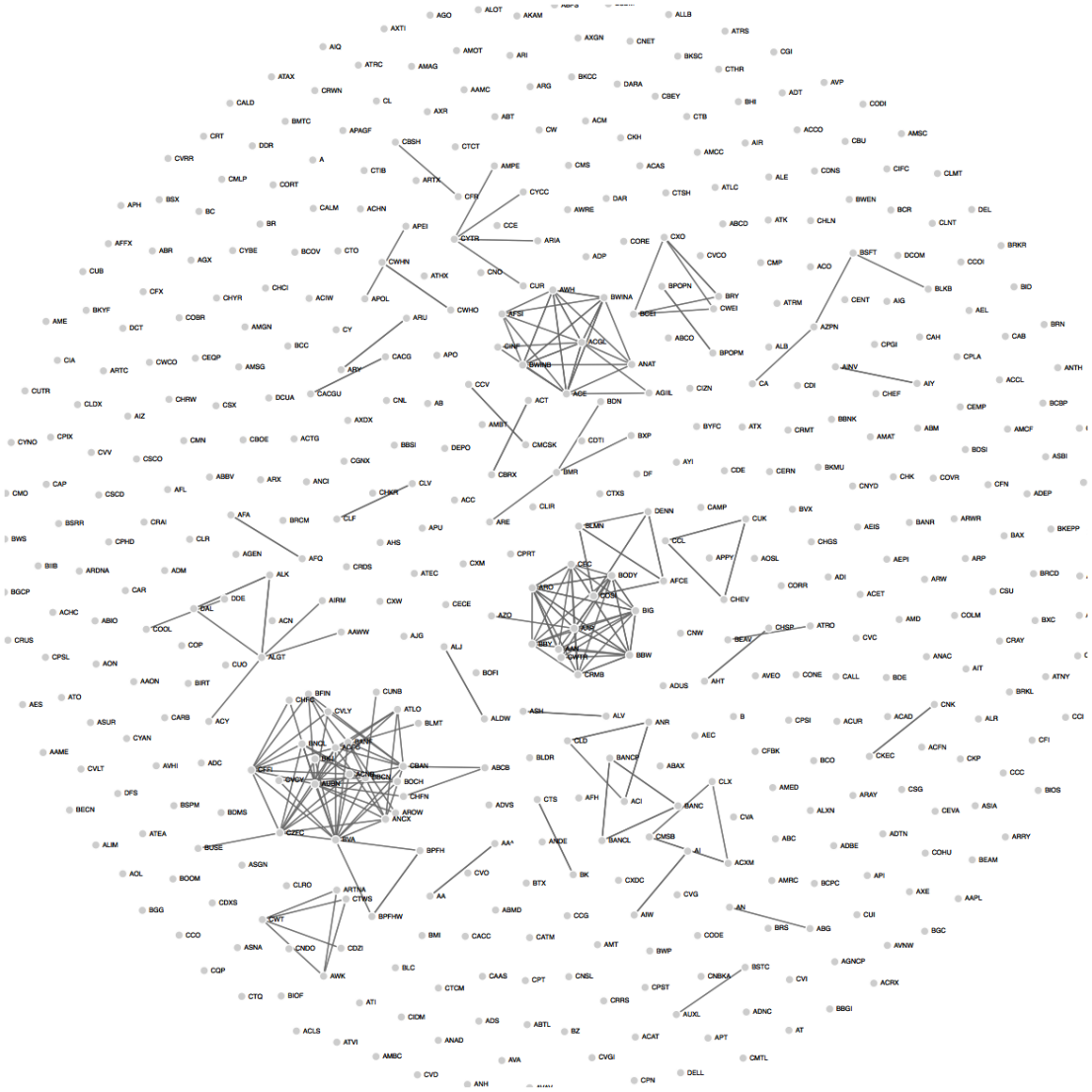
Out[70]= 54.2137

10-K Filings 'Cosine Similarity

Go to URL: http://thorek01.lossofgenerality.com/visualize/SimilarityGraph.html

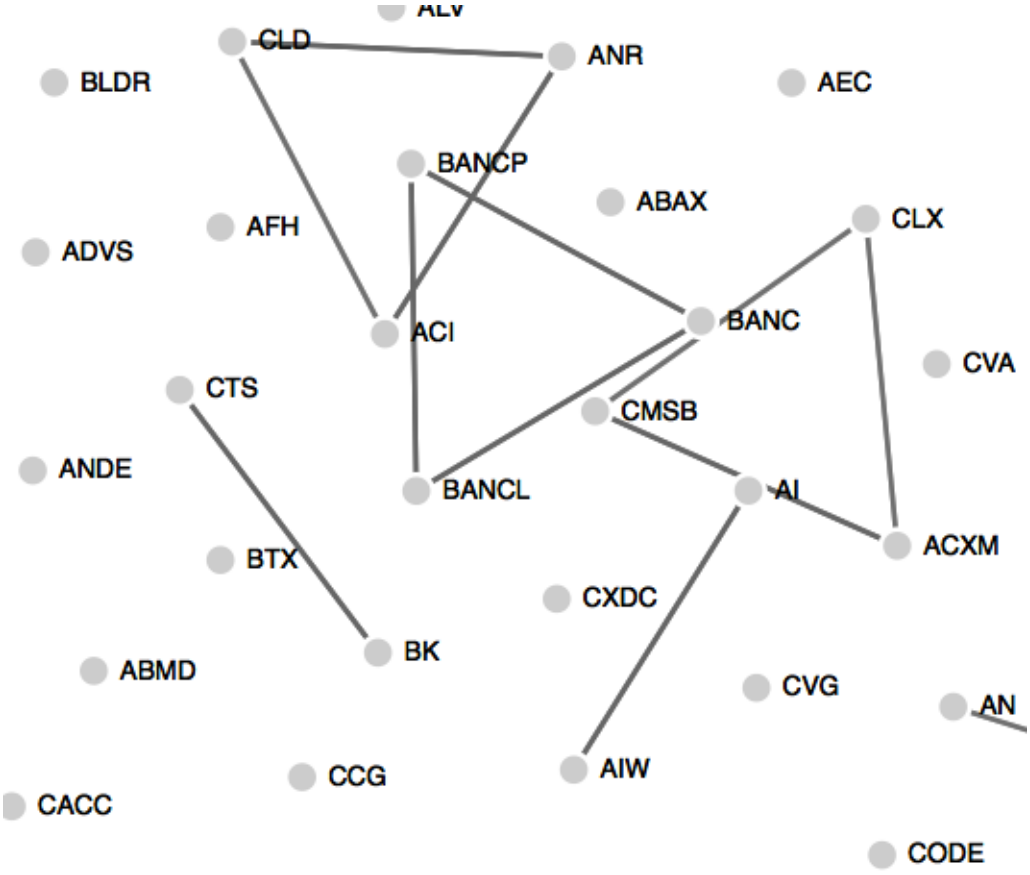
Actual 2012 section 7 (Management Discussion) of about 1000 10-K securities filings in USA. The edges represent 60% or more similarity between the section 7.

As you can see the 60% similarities are established for certain number of companies.



Let’s choose some of these similarities and investigate to see if the documetric similarities of 10-K filings do reflect the price curve similarity.

Similarity: CTS, BK

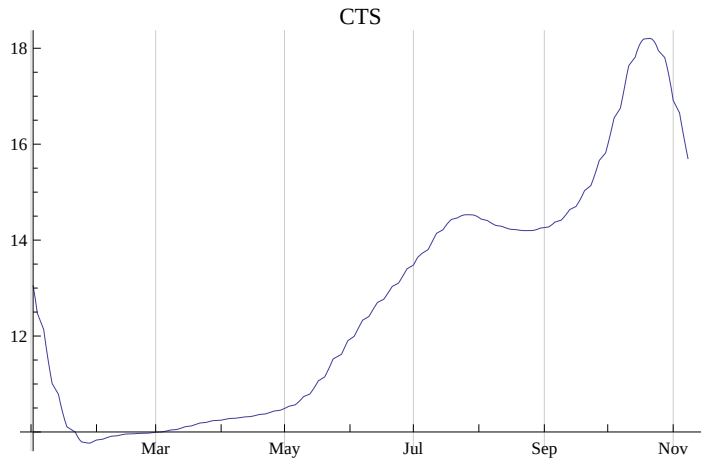


Find the Trend curve for CTS:

In[71]:=

```
price = FinancialData["CTS", "Jan. 1, 2013"];
dwd =
  DiscreteWaveletTransform[price[[All, 2]], BiorthogonalSplineWavelet[3, 3], 4];
trCTS = InverseWaveletTransform[WaveletThreshold[dwd, {"Hard", 100}]];
DateListPlot[Transpose[{price[[All, 1]], trCTS}],
 Joined → True, Frame → False, PlotLabel → "CTS"]
```

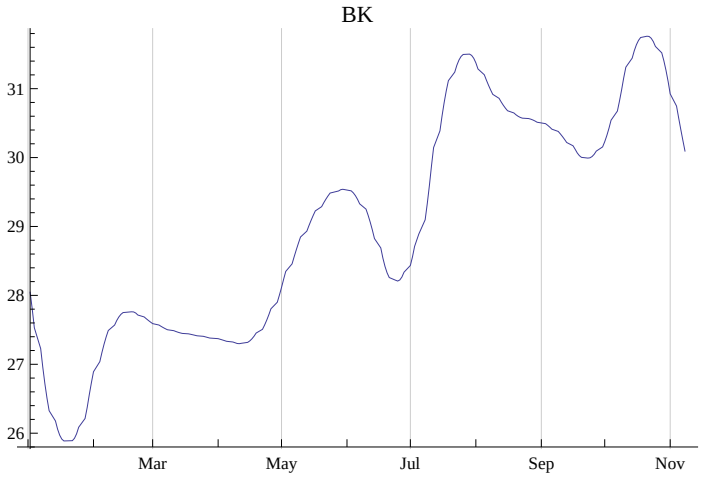
Out[74]=



Find the Trend curve for BK:

In[75]:=
```
price = FinancialData["BK", "Jan. 1, 2013"];
dwd =
  DiscreteWaveletTransform[price[[All, 2]], BiorthogonalSplineWavelet[3, 3], 4];
trBK = InverseWaveletTransform[WaveletThreshold[dwd, {"Hard", 100}]];
DateListPlot[Transpose[{price[[All, 1]], trBK}],
 Joined → True, Frame → False, PlotLabel → "BK"]
```

Out[78]=



Use weighted Manhattan Distance e.g. Canberra Distance to establish a measure of distance or similarity between the Trend Curves:

In[79]:=
```
CanberraDistance[trBK, trCTS]
```

Out[79]= 85.8195

OK!

Measure the Canberra Distance between the Trend curves of CTS and BK to the Trend curve of the random-pick which we show by tr (array) variable below:

In[80]:=
```
data = Table[CanberraDistance[trBK, tr[[i]]], {i, 1, Length[tr]}]
Mean[data]
Sqrt[Variance[data]]
```

Out[80]= {160.451, 9.9231, 53.7317, 7.81697, 151.148,
 46.6443, 76.5341, 156.984, 168.516, 67.0727, 94.0523}

Out[81]= 90.2613

Out[82]= 60.3641

In[83]:=
```
data = Table[CanberraDistance[trCTS, tr[[i]]], {i, 1, Length[tr]}]
Mean[data]
Sqrt[Variance[data]]
```

Out[83]= {190.128, 93.9779, 126.776, 90.0408, 89.8183,
122.129, 37.3165, 98.9199, 118.914, 21.5773, 18.6184}

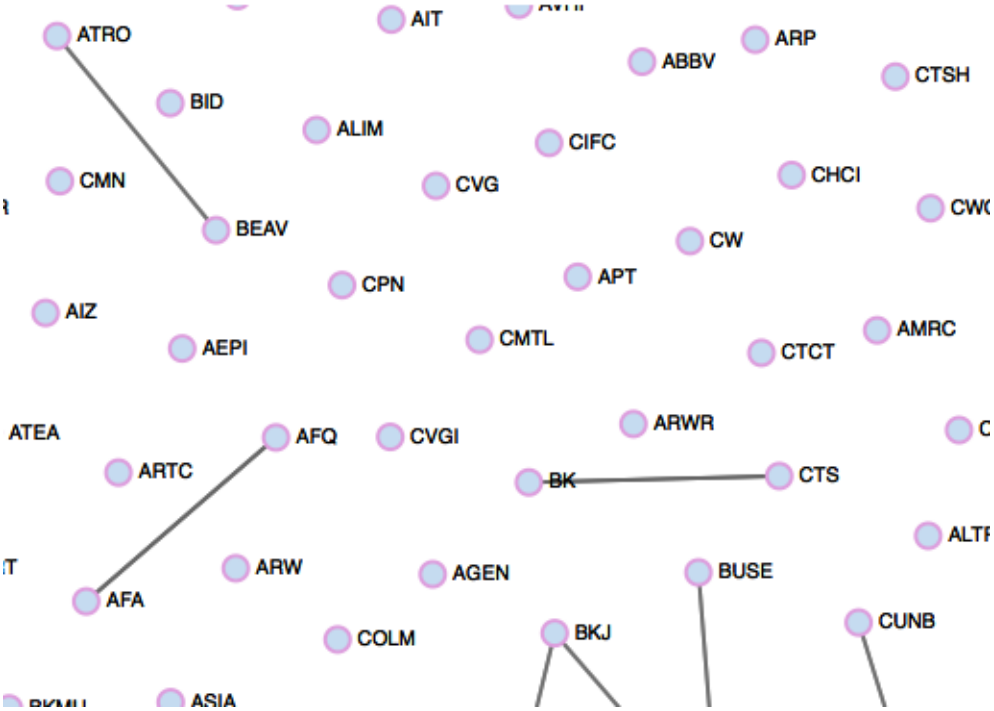
Out[84]= 91.6561

Out[85]= 50.8569

So the computation shows:

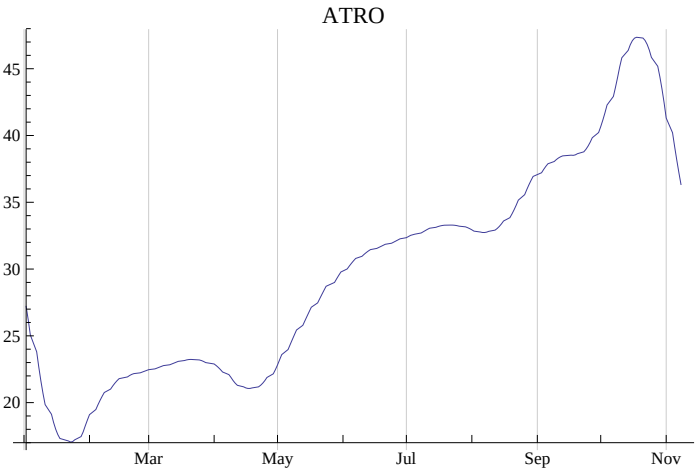
1. The two stocks are close in Trend curve similarity as they are close in Cosine Documetric Similarity
2. The random stock that is afar in terms of Cosine Documetric Similarity is also afar in Canberra distance between the corresponding Trend curves

Similarity: ATRO, BEAV



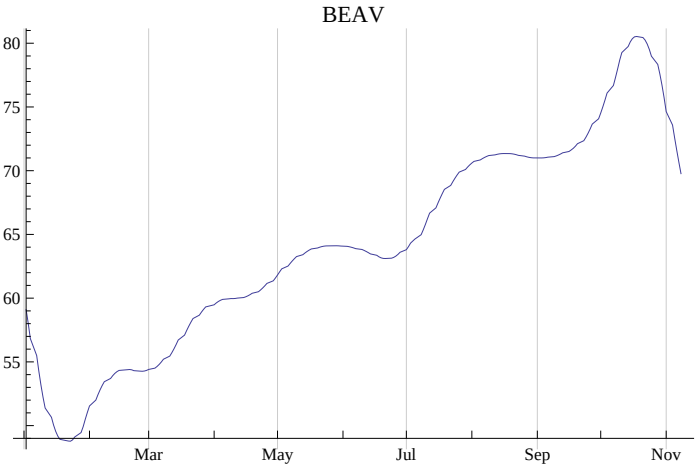
In[86]:=
```
price = FinancialData["ATRO", "Jan. 1, 2013"];
dwd =
  DiscreteWaveletTransform[price[[All, 2]], BiorthogonalSplineWavelet[3, 3], 4];
trATRO = InverseWaveletTransform[WaveletThreshold[dwd, {"Hard", 100}]];
DateListPlot[Transpose[{price[[All, 1]], trATRO}],
 Joined → True, Frame → False, PlotLabel → "ATRO"]
```

Out[89]=



In[90]:=
```
price = FinancialData["BEAV", "Jan. 1, 2013"];
dwd =
  DiscreteWaveletTransform[price[[All, 2]], BiorthogonalSplineWavelet[3, 3], 4];
trBEAV = InverseWaveletTransform[WaveletThreshold[dwd, {"Hard", 100}]];
DateListPlot[Transpose[{price[[All, 1]], trBEAV}],
 Joined → True, Frame → False, PlotLabel → "BEAV"]
```

Out[93]=



Distance between the two Trend price curves:

In[94]:=
```
CanberraDistance[trATRO, trBEAV]
```

Out[94]= 81.6118

OK!

Compute the mean distance to the random-pick outliers:

In[95]:=

```
data = Table[CanberraDistance[trATRO, tr[[i]]], {i, 1, Length[tr]}];
Mean[data]
```

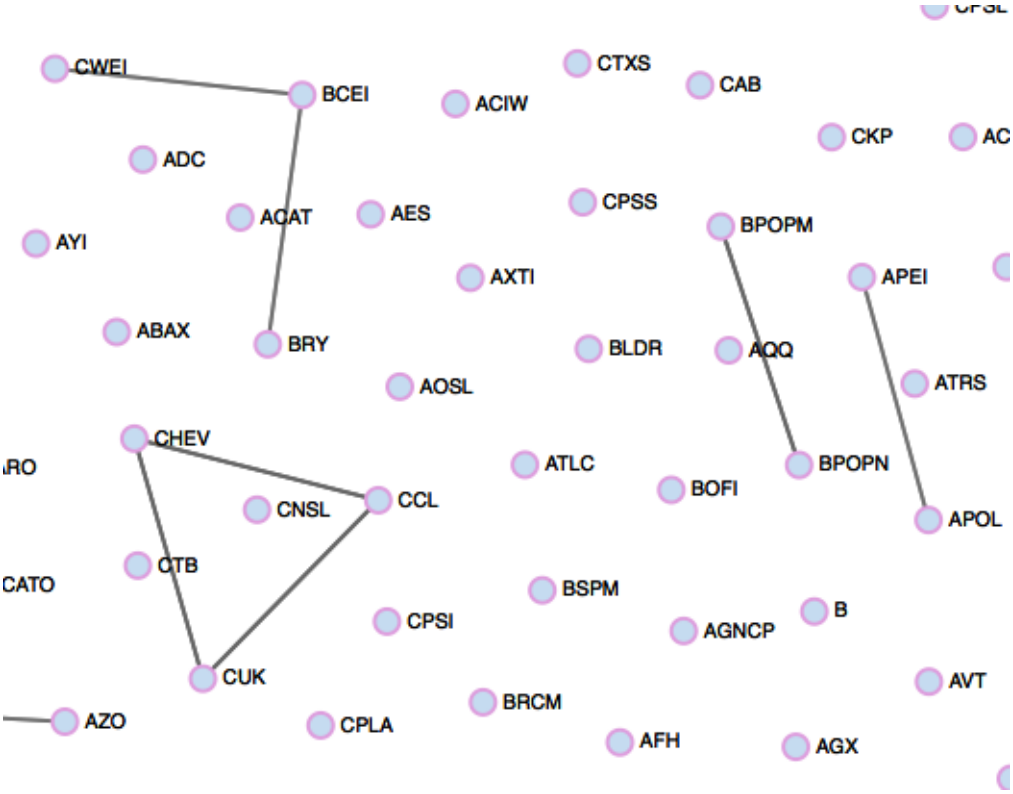
Out[96]= 91.6539

In[97]:=

```
data = Table[CanberraDistance[trBEAV, tr[[i]]], {i, 1, Length[tr]}];
Mean[data]
```

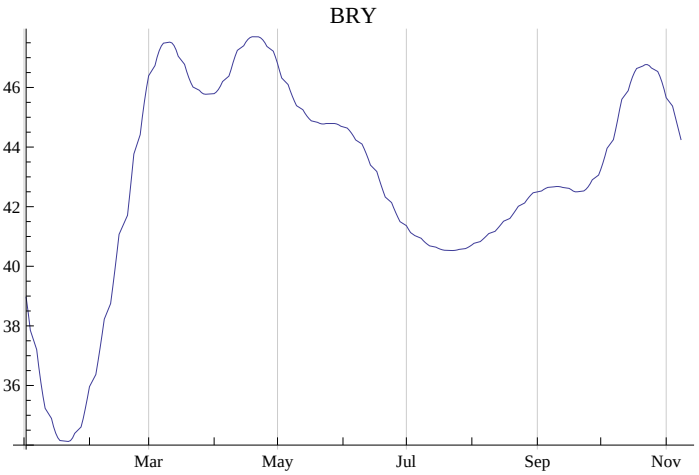
Out[98]= 119.512

Similarity: BRY, BCEI, CWEI



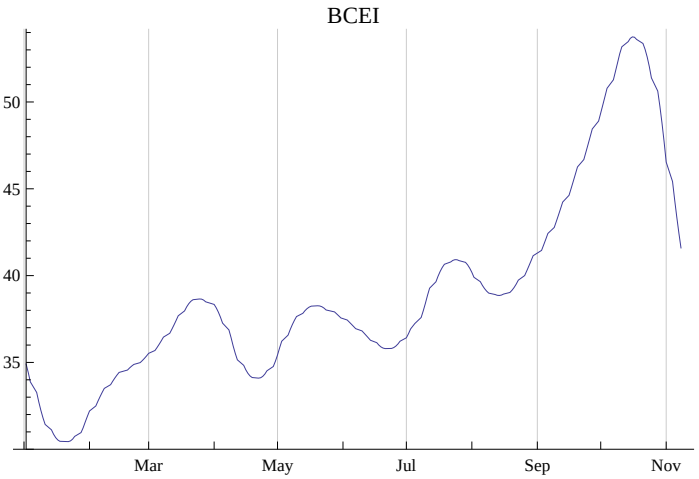
In[99]:=
```
price = FinancialData["BRY", "Jan. 1, 2013"];
dwd =
  DiscreteWaveletTransform[price[[All, 2]], BiorthogonalSplineWavelet[3, 3], 4];
trBRY = InverseWaveletTransform[WaveletThreshold[dwd, {"Hard", 100}]];
DateListPlot[Transpose[{price[[All, 1]], trBRY}],
 Joined → True, Frame → False, PlotLabel → "BRY"]
```

Out[102]=



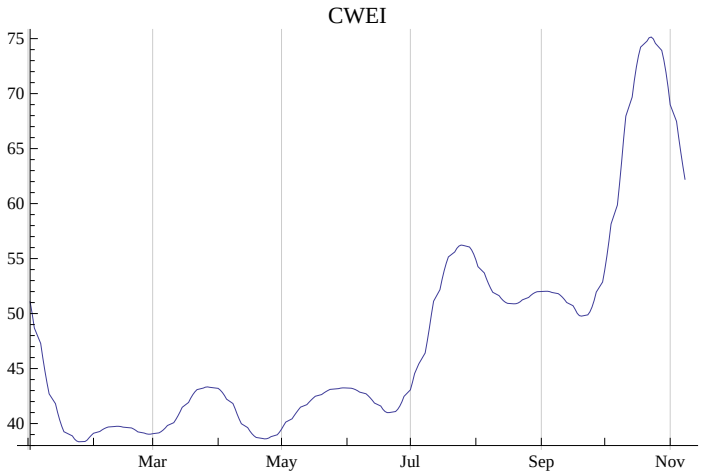
In[103]:=
```
price = FinancialData["BCEI", "Jan. 1, 2013"];
dwd =
  DiscreteWaveletTransform[price[[All, 2]], BiorthogonalSplineWavelet[3, 3], 4];
trBCEI = InverseWaveletTransform[WaveletThreshold[dwd, {"Hard", 100}]];
DateListPlot[Transpose[{price[[All, 1]], trBCEI}],
 Joined → True, Frame → False, PlotLabel → "BCEI"]
```

Out[106]=



In[107]:=
```
price = FinancialData["CWEI", "Jan. 1, 2013"];
dwd =
  DiscreteWaveletTransform[price[[All, 2]], BiorthogonalSplineWavelet[3, 3], 4];
trCWEI = InverseWaveletTransform[WaveletThreshold[dwd, {"Hard", 100}]];
DateListPlot[Transpose[{price[[All, 1]], trCWEI}],
 Joined → True, Frame → False, PlotLabel → "CWEI"]
```

Out[110]=



Compute the pair-wise distance between the Trend price curves:

In[111]:=
```
CanberraDistance[trBRY, trCWEI]
CanberraDistance[trBCEI, trBRY]
CanberraDistance[trBCEI, trCWEI]
```

Out[111]= 18.9564

Out[112]= 14.9414

Out[113]= 20.7979

OK!

Compute the mean distance to the random-pick outliers:

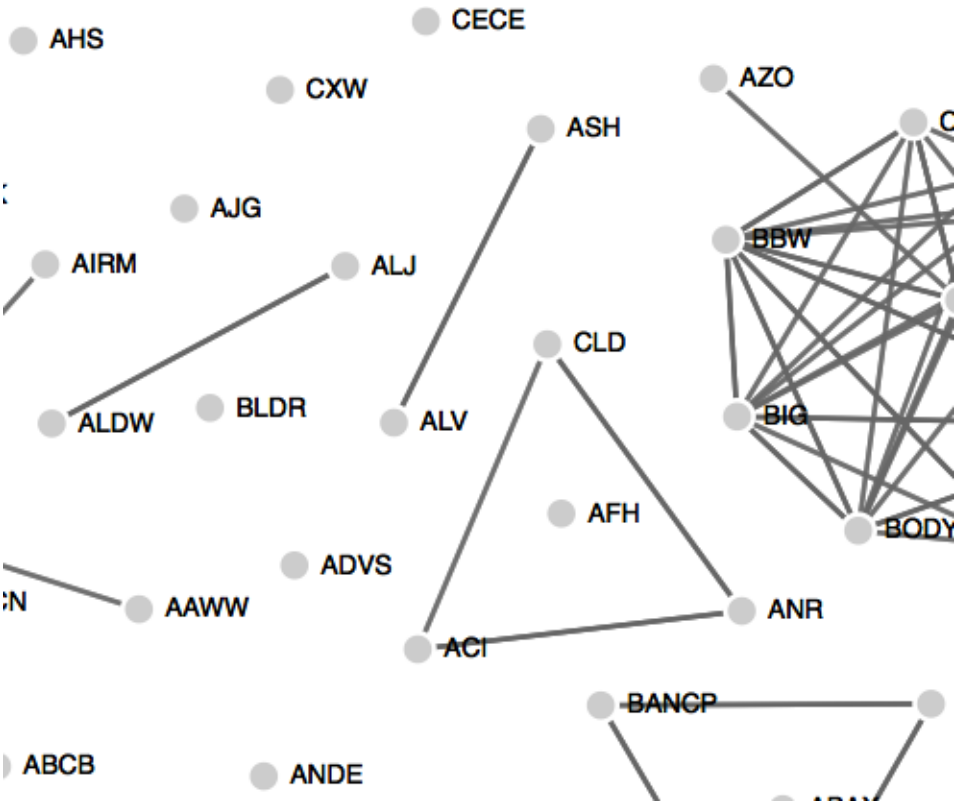
In[114]:=
```
data = Table[CanberraDistance[trBRY, tr[[i]]], {i, 1, Length[tr]}];
Mean[data]
data = Table[CanberraDistance[trBCEI, tr[[i]]], {i, 1, Length[tr]}];
Mean[data]
data = Table[CanberraDistance[trCWEI, tr[[i]]], {i, 1, Length[tr]}];
Mean[data]
```

Out[115]= 99.4717

Out[117]= 96.5732

Out[119]= 103.72

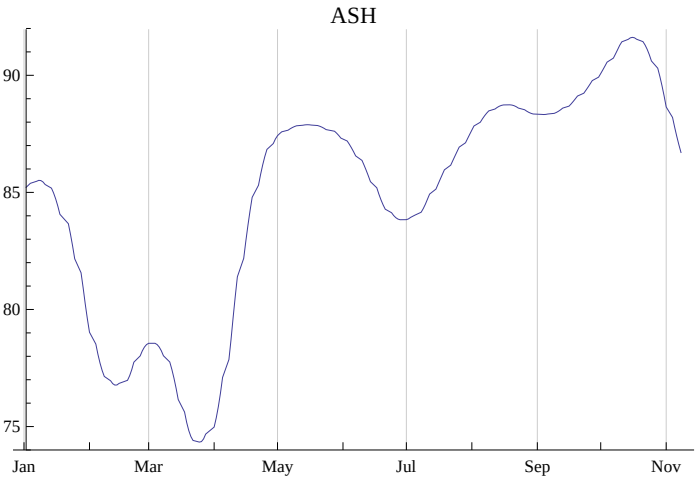
Similarity: ASH, ALV, AXP



In[120]:=

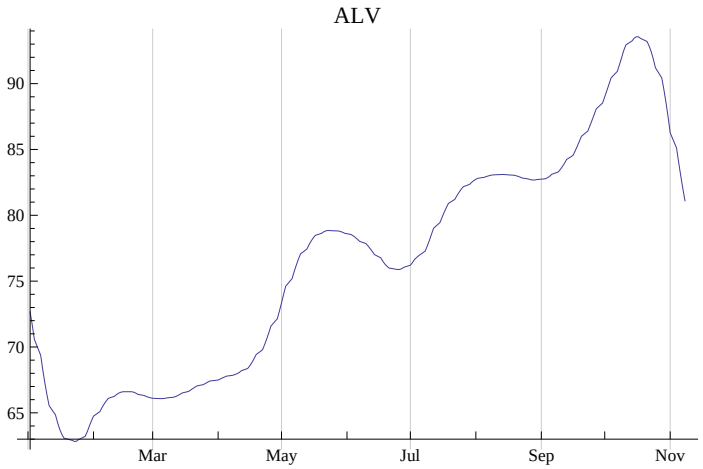
```
price = FinancialData["ASH", "Jan. 1, 2013"];
dwd =
  DiscreteWaveletTransform[price[[All, 2]], BiorthogonalSplineWavelet[3, 3], 4];
trASH = InverseWaveletTransform[WaveletThreshold[dwd, {"Hard", 100}]];
DateListPlot[Transpose[{price[[All, 1]], trASH}],
 Joined → True, Frame → False, PlotLabel → "ASH"]
```

Out[123]=



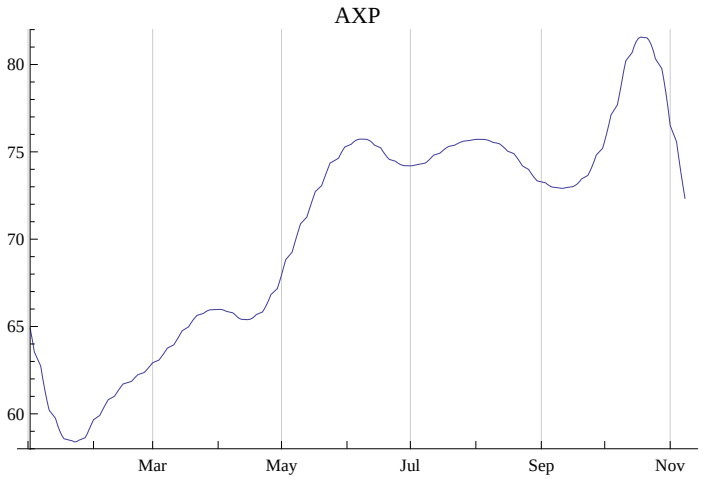
In[124]:=
```
price = FinancialData["ALV", "Jan. 1, 2013"];
dwd =
  DiscreteWaveletTransform[price[[All, 2]], BiorthogonalSplineWavelet[3, 3], 4];
trALV = InverseWaveletTransform[WaveletThreshold[dwd, {"Hard", 100}]];
DateListPlot[Transpose[{price[[All, 1]], trALV}],
 Joined → True, Frame → False, PlotLabel → "ALV"]
```

Out[127]=



In[132]:=
```
price = FinancialData["AXP", "Jan. 1, 2013"];
dwd =
  DiscreteWaveletTransform[price[[All, 2]], BiorthogonalSplineWavelet[3, 3], 4];
trAXP = InverseWaveletTransform[WaveletThreshold[dwd, {"Hard", 100}]];
DateListPlot[Transpose[{price[[All, 1]], trAXP}],
 Joined → True, Frame → False, PlotLabel → "AXP"]
```

Out[135]=



Compute pair-wise distance between the Trend price curves:

In[136]:=
```
CanberraDistance[trALV, trASH]
CanberraDistance[trAXP, trASH]
CanberraDistance[trALV, trAXP]
```

Out[136]= 11.9169

Out[137]= 20.3276

Out[138]= 8.67416

OK!

Compute the mean distance to the random-pick outliers:

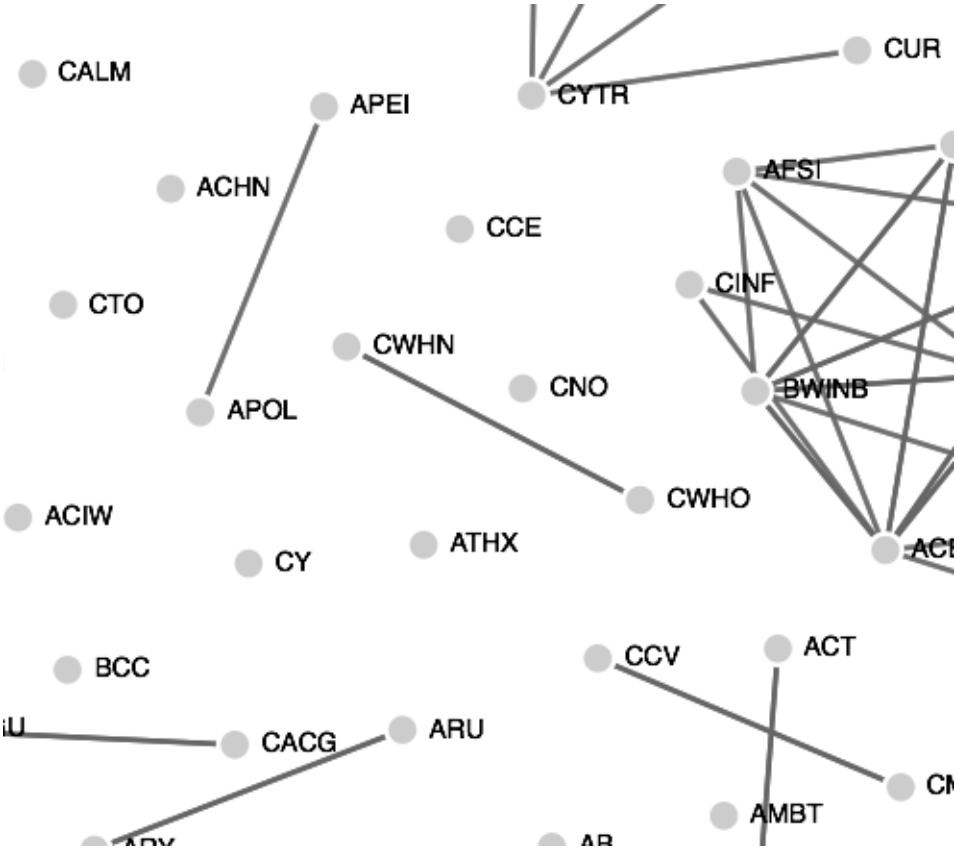
In[139]:=
```
data = Table[CanberraDistance[trALV, tr[[i]]], {i, 1, Length[tr]}];
Mean[data]
data = Table[CanberraDistance[trASH, tr[[i]]], {i, 1, Length[tr]}];
Mean[data]
data = Table[CanberraDistance[trAXP, tr[[i]]], {i, 1, Length[tr]}];
Mean[data]
```

Out[140]= 128.488

Out[142]= 133.448

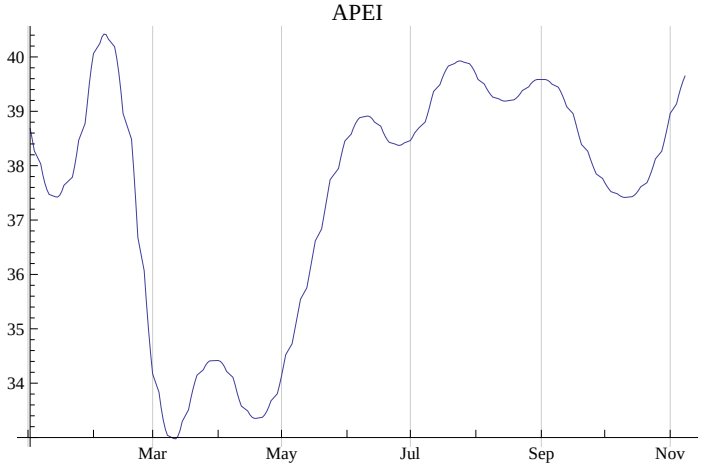
Out[144]= 124.521

Similarity: APEI, APOL



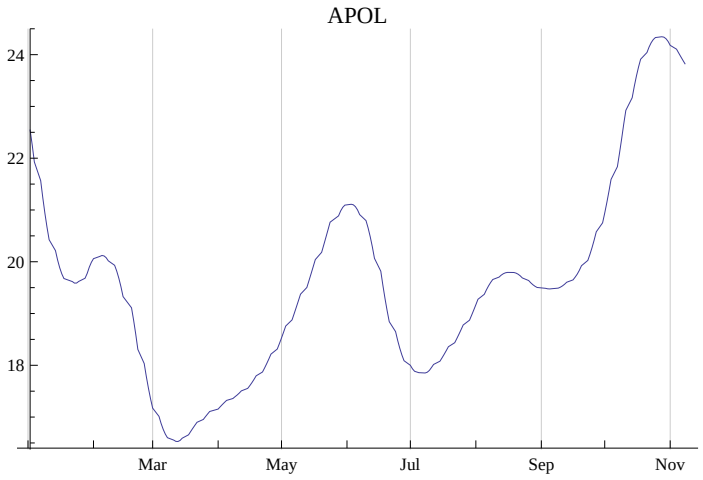
In[145]:=
```
price = FinancialData["APEI", "Jan. 1, 2013"];
dwd =
  DiscreteWaveletTransform[price[[All, 2]], BiorthogonalSplineWavelet[3, 3], 4];
trAPEI = InverseWaveletTransform[WaveletThreshold[dwd, {"Hard", 100}]];
DateListPlot[Transpose[{price[[All, 1]], trAPEI}],
 Joined → True, Frame → False, PlotLabel → "APEI"]
```

Out[148]=



In[149]:=
```
price = FinancialData["APOL", "Jan. 1, 2013"];
dwd =
  DiscreteWaveletTransform[price[[All, 2]], BiorthogonalSplineWavelet[3, 3], 4];
trAPOL = InverseWaveletTransform[WaveletThreshold[dwd, {"Hard", 100}]];
DateListPlot[Transpose[{price[[All, 1]], trAPOL}],
 Joined → True, Frame → False, PlotLabel → "APOL"]
```

Out[152]=



Distance between the Trend curves:

In[153]:=
```
CanberraDistance[trAPEI, trAPOL]
```

Out[153]= 68.0799

OK!

Distance between the APEI and APOL Trend curves and random-pick:

In[154]:=
```
data = Table[CanberraDistance[trAPEI, tr[[i]]], {i, 1, Length[tr]}]
Mean[data]
Sqrt[Variance[data]]
```

Out[154]=
```
{146.666, 17.7465, 26.9815, 23.3424, 163.98,
 20.3085, 98.2274, 168.971, 178.37, 91.0125, 115.228}
```

Out[155]= 95.5303

Out[156]= 64.4907

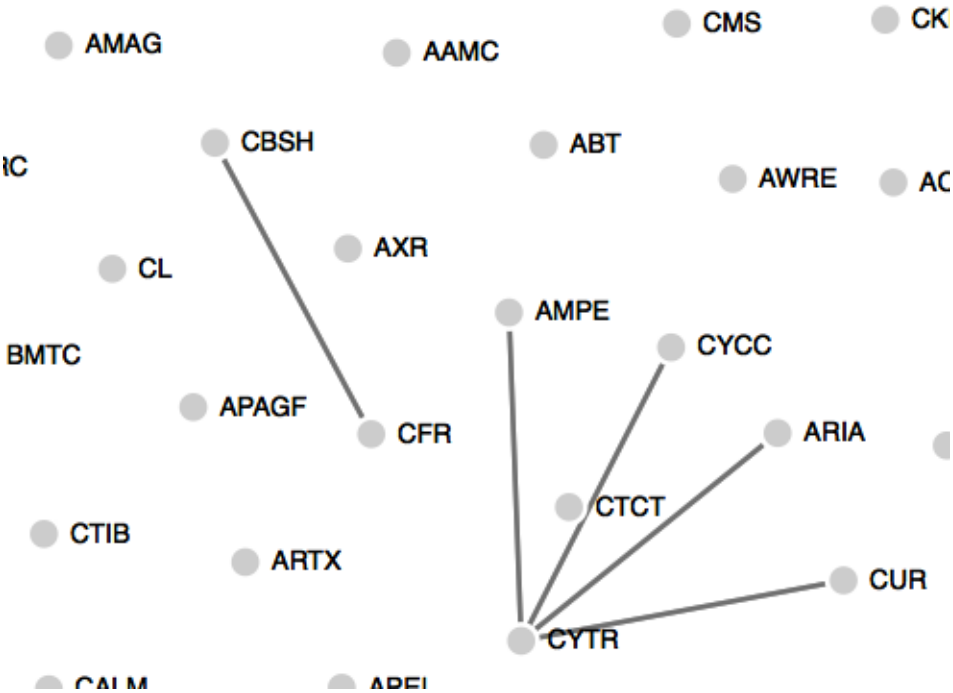
In[157]:=
```
data = Table[CanberraDistance[trAPOL, tr[[i]]], {i, 1, Length[tr]}]
Mean[data]
Sqrt[Variance[data]]
```

Out[157]= {177.061, 51.6571, 91.3897, 46.7502, 125.73, 85.2014, 45.8017, 133.488, 148.63, 26.4448, 56.4549}

Out[158]= 89.8736

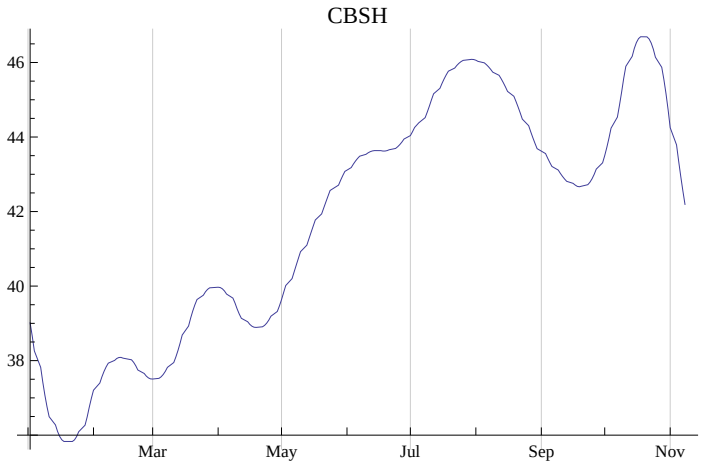
Out[159]= 49.6675

Similarity: CBSH, CFR



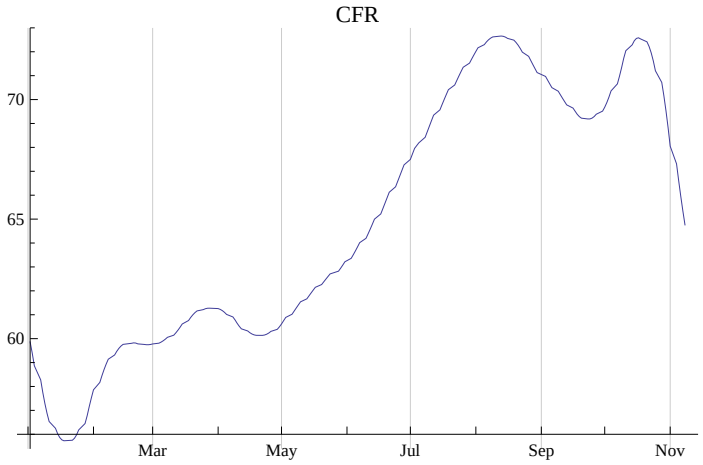
In[160]:=
```
price = FinancialData["CBSH", "Jan. 1, 2013"];
dwd =
  DiscreteWaveletTransform[price[[All, 2]], BiorthogonalSplineWavelet[3, 3], 4];
trCBSH = InverseWaveletTransform[WaveletThreshold[dwd, {"Hard", 100}]];
DateListPlot[Transpose[{price[[All, 1]], trCBSH}],
 Joined → True, Frame → False, PlotLabel → "CBSH"]
```

Out[163]=



In[164]:=
```
price = FinancialData["CFR", "Jan. 1, 2013"];
dwd =
  DiscreteWaveletTransform[price[[All, 2]], BiorthogonalSplineWavelet[3, 3], 4];
trCFR = InverseWaveletTransform[WaveletThreshold[dwd, {"Hard", 100}]];
DateListPlot[Transpose[{price[[All, 1]], trCFR}],
 Joined → True, Frame → False, PlotLabel → "CFR"]
```

Out[167]=



Distance between the CBSH Trend curve and CFR's:

In[168]:=
```
CanberraDistance[trCBSH, trCFR]
```

Out[168]= 47.0029

OK!

Distance between the CBSH and CFR Trend curves and random-pick:

In[169]:=
```
data = Table[CanberraDistance[trCBSH, tr[[i]]], {i, 1, Length[tr]}]
Mean[data]
Sqrt[Variance[data]]
```

Out[169]= {140.046, 29.2517, 15.4746, 34.2169, 168.984,
8.10979, 107.823, 173.373, 182.077, 100.395, 123.359}

Out[170]= 98.4645

Out[171]= 66.321

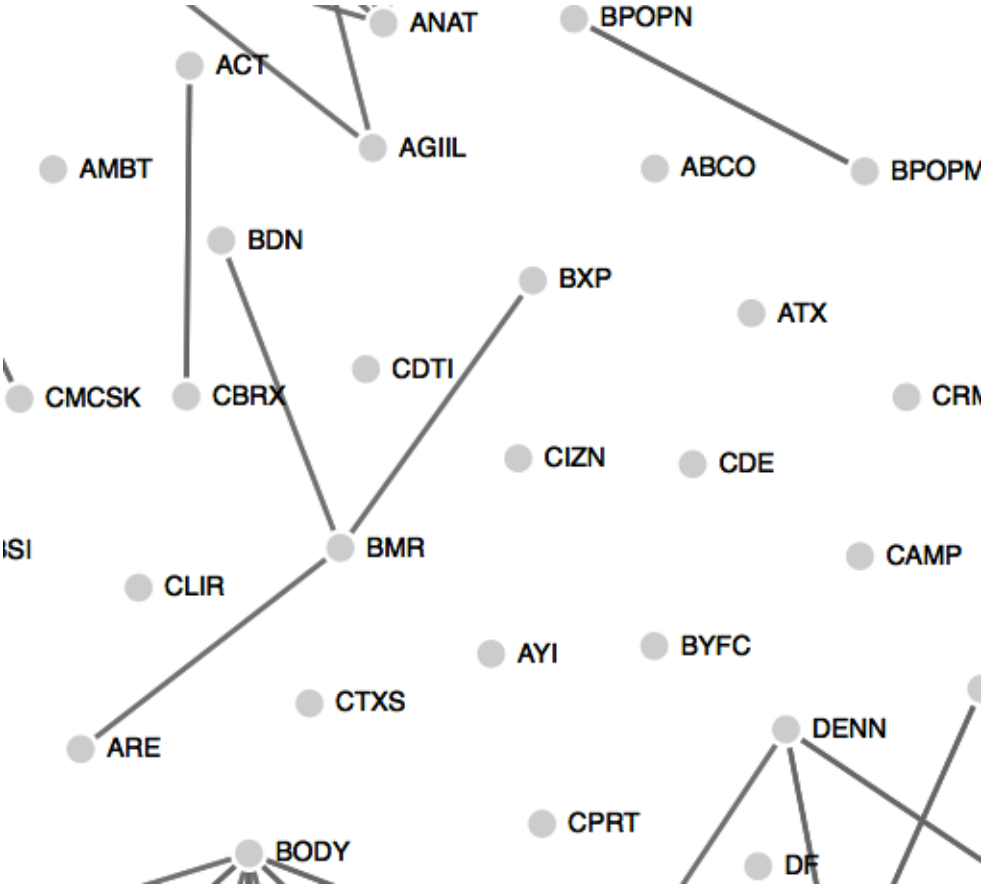
In[172]:=
```
data = Table[CanberraDistance[trCFR, tr[[i]]], {i, 1, Length[tr]}]
Mean[data]
Sqrt[Variance[data]]
```

Out[172]= {108.19, 74.0669, 31.9901, 78.4965, 184.808,
39.2687, 139.022, 187.863, 193.858, 133.973, 151.656}

Out[173]= 120.29

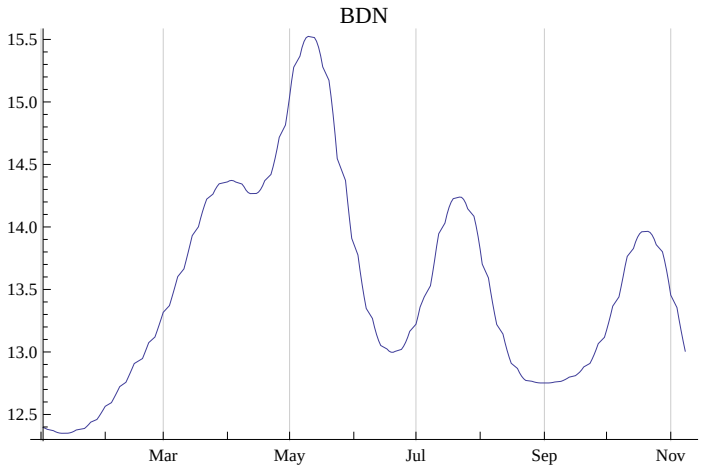
Out[174]= 58.3558

Similarity: BDN, BXP, BMR, ARE



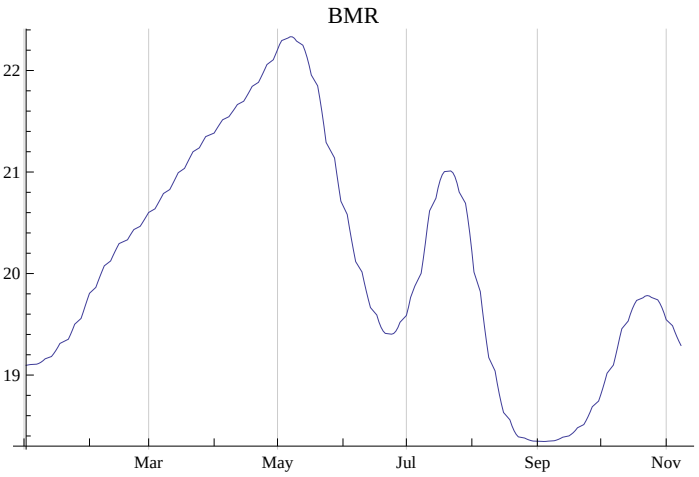
In[175]:=
```
price = FinancialData["BDN", "Jan. 1, 2013"];
dwd =
  DiscreteWaveletTransform[price[[All, 2]], BiorthogonalSplineWavelet[3, 3], 4];
trBDN = InverseWaveletTransform[WaveletThreshold[dwd, {"Hard", 100}]];
DateListPlot[Transpose[{price[[All, 1]], trBDN}],
 Joined → True, Frame → False, PlotLabel → "BDN"]
```

Out[178]=



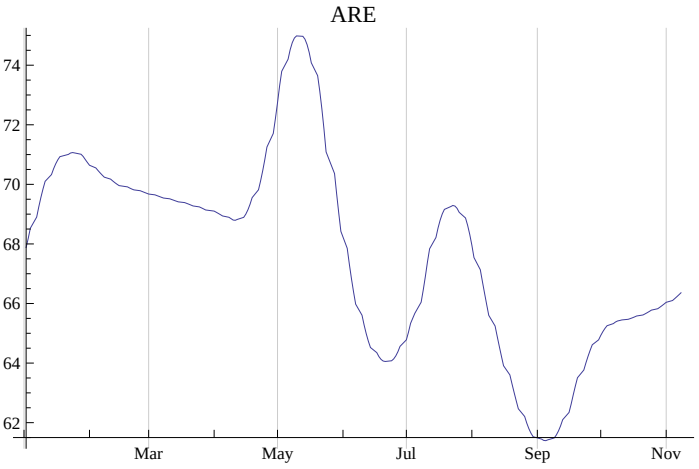
In[179]:=
```
price = FinancialData["BMR", "Jan. 1, 2013"];
dwd =
  DiscreteWaveletTransform[price[[All, 2]], BiorthogonalSplineWavelet[3, 3], 4];
trBMR = InverseWaveletTransform[WaveletThreshold[dwd, {"Hard", 100}]];
DateListPlot[Transpose[{price[[All, 1]], trBMR}],
 Joined → True, Frame → False, PlotLabel → "BMR"]
```

Out[182]=



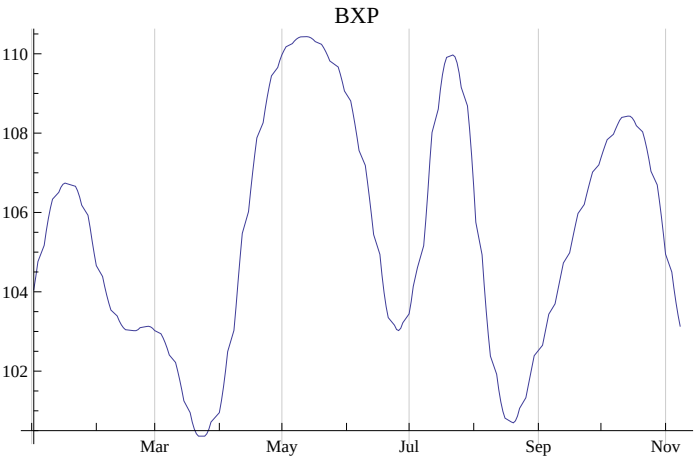
In[183]:=
```
price = FinancialData["ARE", "Jan. 1, 2013"];
dwd =
  DiscreteWaveletTransform[price[[All, 2]], BiorthogonalSplineWavelet[3, 3], 4];
trARE = InverseWaveletTransform[WaveletThreshold[dwd, {"Hard", 100}]];
DateListPlot[Transpose[{price[[All, 1]], trARE}],
 Joined → True, Frame → False, PlotLabel → "ARE"]
```

Out[186]=



In[187]:=
```
price = FinancialData["BXP", "Jan. 1, 2013"];
dwd =
  DiscreteWaveletTransform[price[[All, 2]], BiorthogonalSplineWavelet[3, 3], 4];
trBXP = InverseWaveletTransform[WaveletThreshold[dwd, {"Hard", 100}]];
DateListPlot[Transpose[{price[[All, 1]], trBXP}],
 Joined → True, Frame → False, PlotLabel → "BXP"]
```

Out[190]=



Fail:Partial

In this case the pair-wise Canberra distances of the stock Trend curves are smaller than their individual distances to the random-pick:

In[191]:=
```
CanberraDistance[trBDN, trBMR]
CanberraDistance[trARE, trBMR]
CanberraDistance[trBMR, trARE]
CanberraDistance[trBDN, trBXP]
CanberraDistance[trARE, trBXP]
CanberraDistance[trBMR, trBXP]
```

Out[191]= 42.3255

Out[192]= 117.644

Out[193]= 117.644

Out[194]= 167.632

Out[195]= 47.3833

Out[196]= 147.534

Compute the mean distance to the random-pick outliers:

In[197]:=
```
data = Table[CanberraDistance[trBDN, tr[[i]]], {i, 1, Length[tr]}];
Mean[data]
data = Table[CanberraDistance[trARE, tr[[i]]], {i, 1, Length[tr]}];
Mean[data]
data = Table[CanberraDistance[trBMR, tr[[i]]], {i, 1, Length[tr]}];
Mean[data]
data = Table[CanberraDistance[trBXP, tr[[i]]], {i, 1, Length[tr]}];
Mean[data]
```

Out[198]= 91.8741

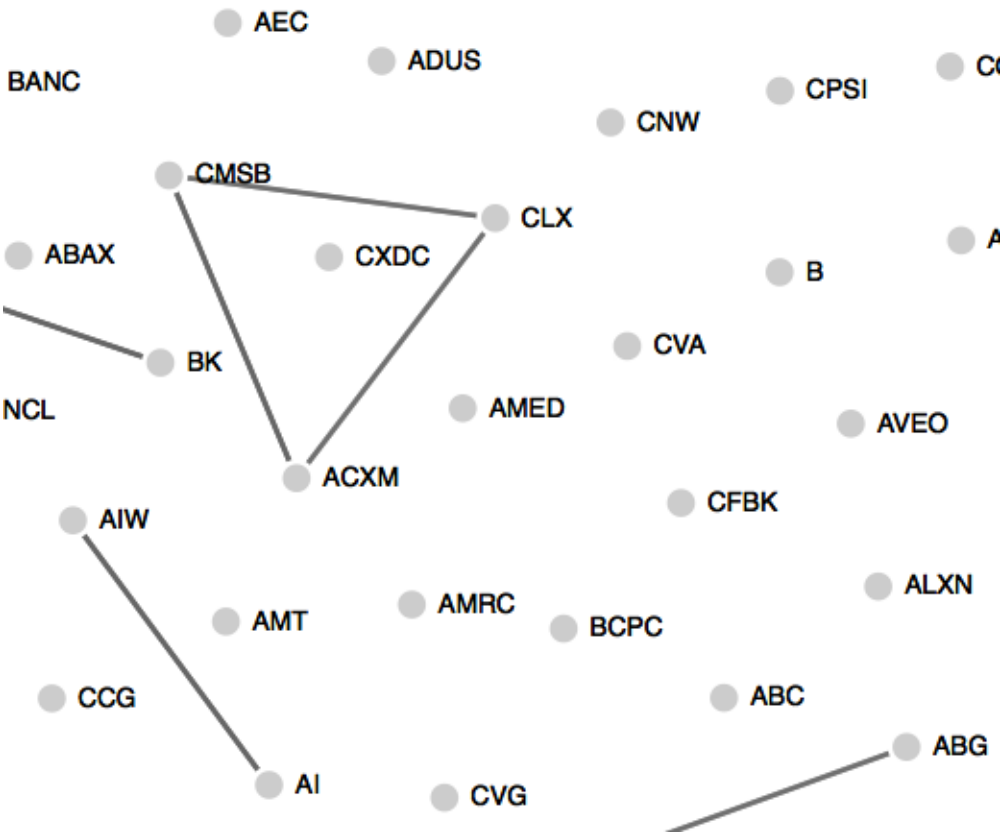
Out[200]= 122.321

Out[202]= 90.3861

Out[204]= 142.441

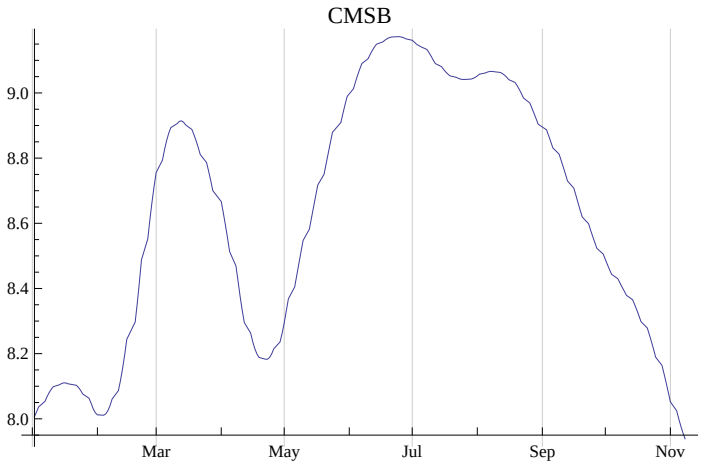
BXP and ARE are causing the problem here, perhaps more investigation needed, or modification of algorithm.

Similarity: CMSB, CLX, ACXM



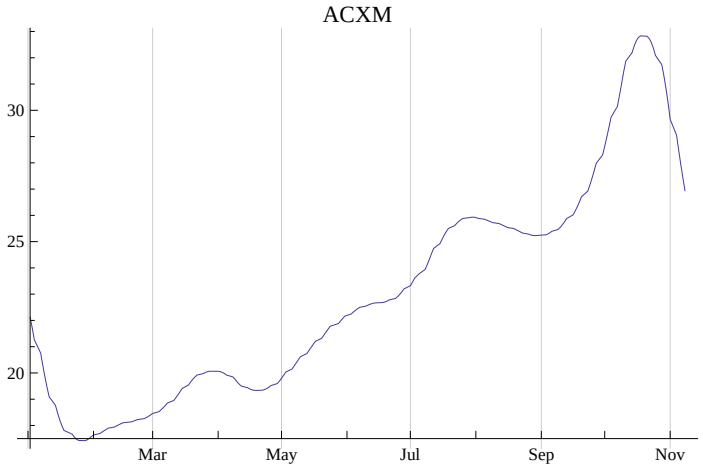
In[205]:=
```
price = FinancialData["CMSB", "Jan. 1, 2013"];
dwd =
  DiscreteWaveletTransform[price[[All, 2]], BiorthogonalSplineWavelet[3, 3], 4];
trCMSB = InverseWaveletTransform[WaveletThreshold[dwd, {"Hard", 100}]];
DateListPlot[Transpose[{price[[All, 1]], trCMSB}],
 Joined → True, Frame → False, PlotLabel → "CMSB"]
```

Out[208]=



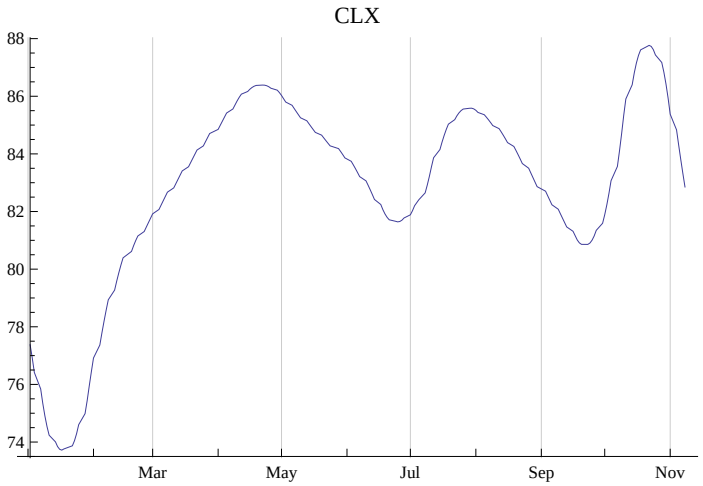
In[209]:=
```
price = FinancialData["ACXM", "Jan. 1, 2013"];
dwd =
  DiscreteWaveletTransform[price[[All, 2]], BiorthogonalSplineWavelet[3, 3], 4];
trACXM = InverseWaveletTransform[WaveletThreshold[dwd, {"Hard", 100}]];
DateListPlot[Transpose[{price[[All, 1]], trACXM}],
 Joined → True, Frame → False, PlotLabel → "ACXM"]
```

Out[212]=



In[213]:=
```
price = FinancialData["CLX", "Jan. 1, 2013"];
dwd =
  DiscreteWaveletTransform[price[[All, 2]], BiorthogonalSplineWavelet[3, 3], 4];
trCLX = InverseWaveletTransform[WaveletThreshold[dwd, {"Hard", 100}]];
DateListPlot[Transpose[{price[[All, 1]], trCLX}],
 Joined → True, Frame → False, PlotLabel → "CLX"]
```

Out[216]=



Let's compute the Canberra distance between the stocks Trend curves:

In[217]:=
```
CanberraDistance[trCMSB, trACXM]
CanberraDistance[trCMSB, trCLX]
CanberraDistance[trCLX, trACXM]
```

Out[217]= 96.9579

Out[218]= 176.063

Out[219]= 123.124

FAIL: Partial

One of the distances i.e. 87.1352 is smaller than the distances above.

In[220]:=
```
data = Table[CanberraDistance[trCMSB, tr[[i]]], {i, 1, Length[tr]}];
Mean[data]
data = Table[CanberraDistance[trCLX, tr[[i]]], {i, 1, Length[tr]}];
Mean[data]
data = Table[CanberraDistance[trACXM, tr[[i]]], {i, 1, Length[tr]}];
Mean[data]
```

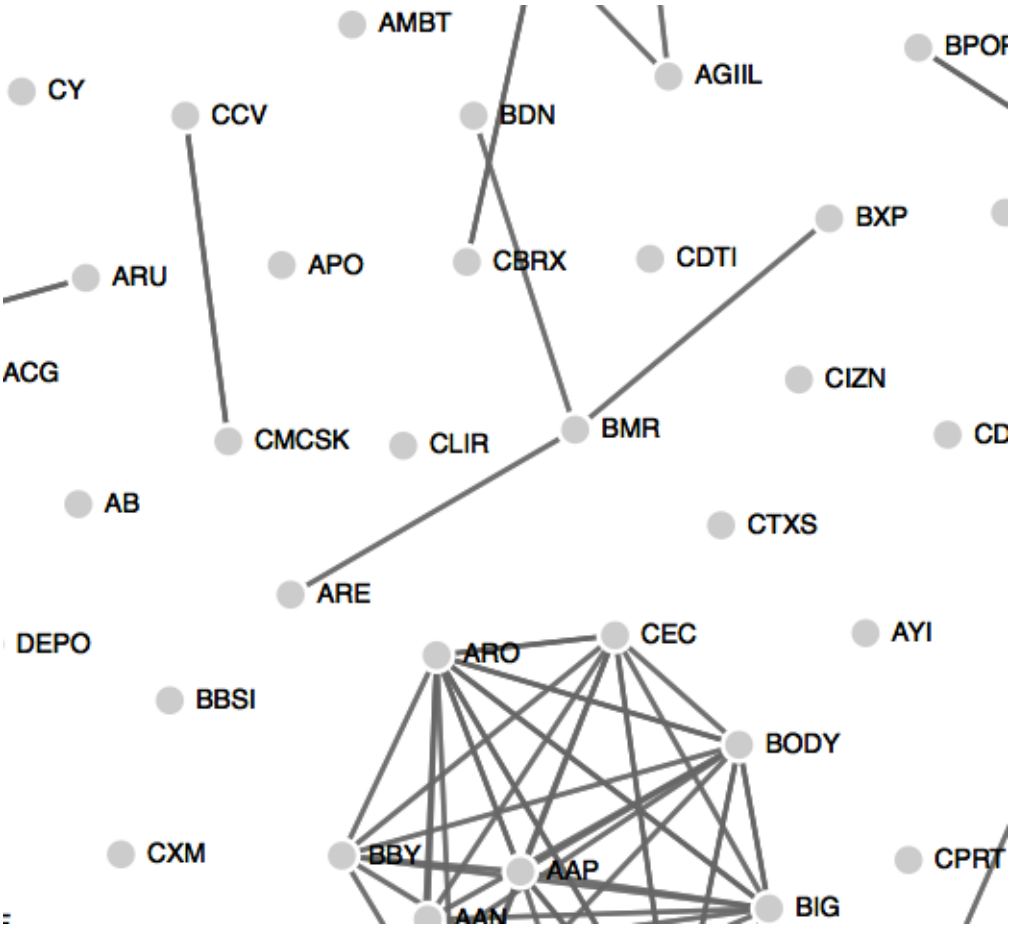
Out[221]= 100.13

Out[223]= 132.339

Out[225]= 90.0301

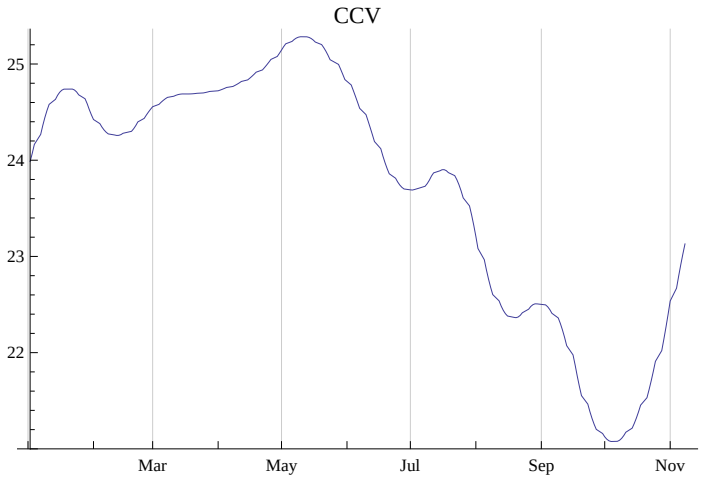
CLX is causing the problem, somehow it should be weeded out.

Similarity: CCV, CMCSK



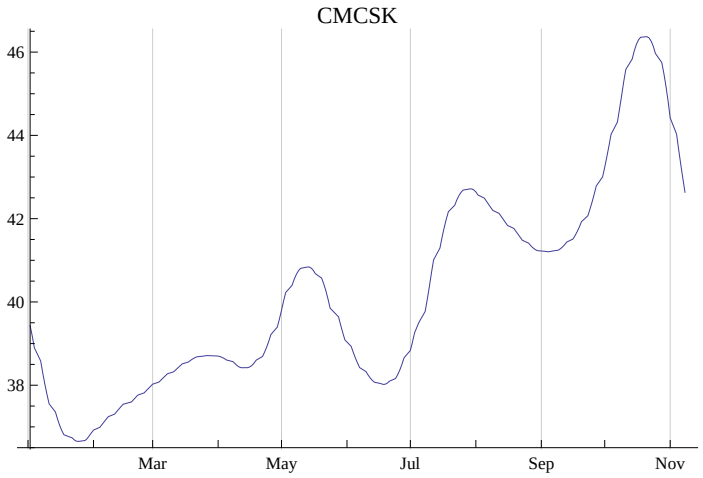
In[226]:=
```
price = FinancialData["CCV", "Jan. 1, 2013"];
dwd =
  DiscreteWaveletTransform[price[[All, 2]], BiorthogonalSplineWavelet[3, 3], 4];
trCCV = InverseWaveletTransform[WaveletThreshold[dwd, {"Hard", 100}]];
DateListPlot[Transpose[{price[[All, 1]], trCCV}],
 Joined → True, Frame → False, PlotLabel → "CCV"]
```

Out[229]=



In[230]:=
```
price = FinancialData["CMCSK", "Jan. 1, 2013"];
dwd =
  DiscreteWaveletTransform[price[[All, 2]], BiorthogonalSplineWavelet[3, 3], 4];
trCMCSK = InverseWaveletTransform[WaveletThreshold[dwd, {"Hard", 100}]];
DateListPlot[Transpose[{price[[All, 1]], trCMCSK}],
 Joined → True, Frame → False, PlotLabel → "CMCSK"]
```

Out[233]=



Distance between the Trend curves for CMCSK and CCV:

In[234]:=
```
CanberraDistance[trCMCSK, trCCV]
```

Out[234]= 56.0144

OK!

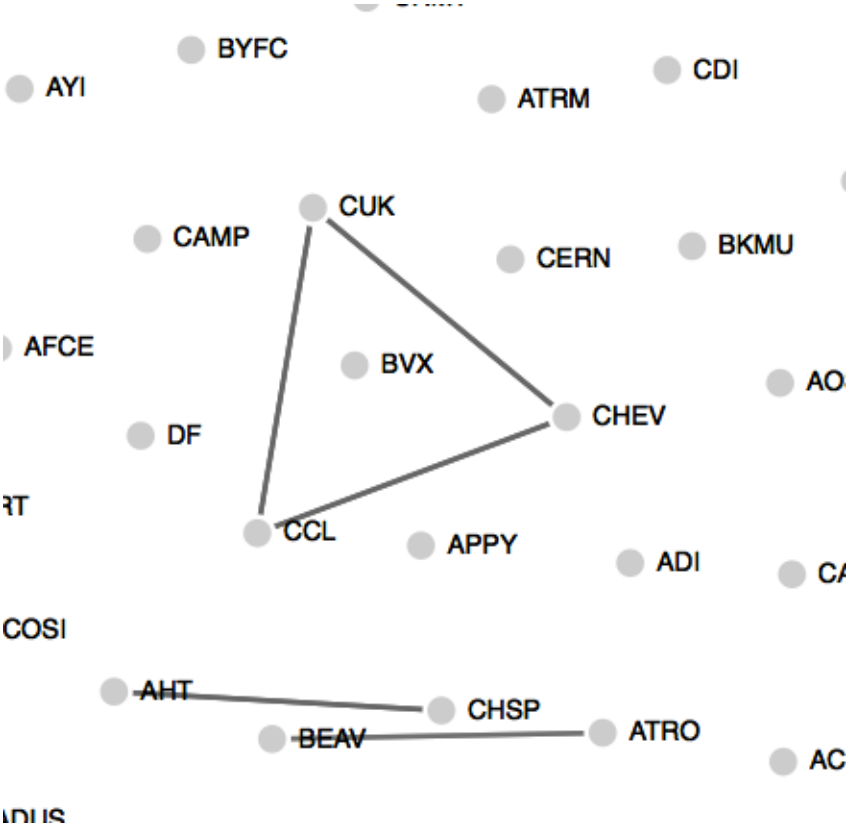
Much smaller than the corresponding distance to the random-pick:

In[235]:=
```
data = Table[CanberraDistance[trCMCSK, tr[[i]]], {i, 1, Length[tr]}];
Mean[data]
data = Table[CanberraDistance[trCCV, tr[[i]]], {i, 1, Length[tr]}];
Mean[data]
```

Out[236]= 97.4875

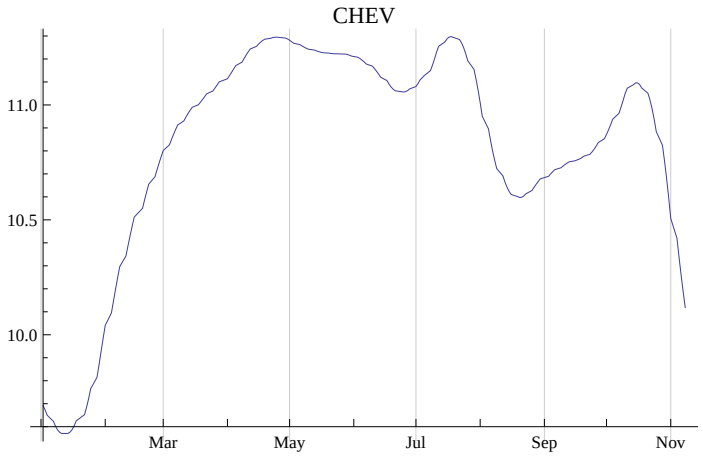
Out[238]= 90.3896

Similarity: CUK, CHEV, CCL



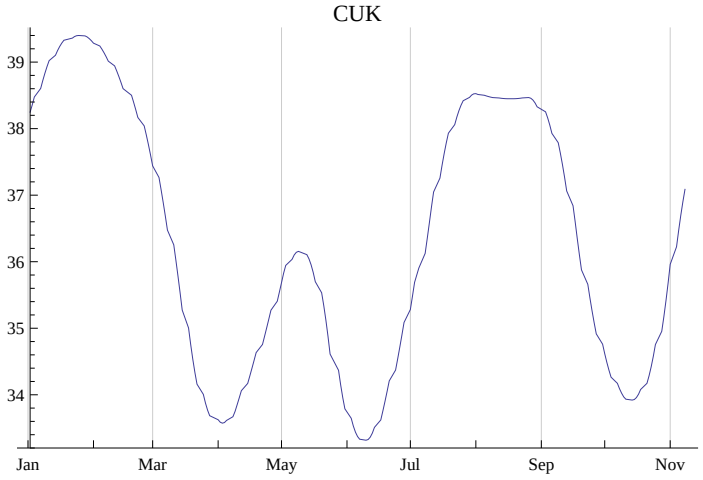
In[239]:=
```
price = FinancialData["CHEV", "Jan. 1, 2013"];
dwd =
  DiscreteWaveletTransform[price[[All, 2]], BiorthogonalSplineWavelet[3, 3], 4];
trCHEV = InverseWaveletTransform[WaveletThreshold[dwd, {"Hard", 100}]];
DateListPlot[Transpose[{price[[All, 1]], trCHEV}],
 Joined → True, Frame → False, PlotLabel → "CHEV"]
```

Out[242]=



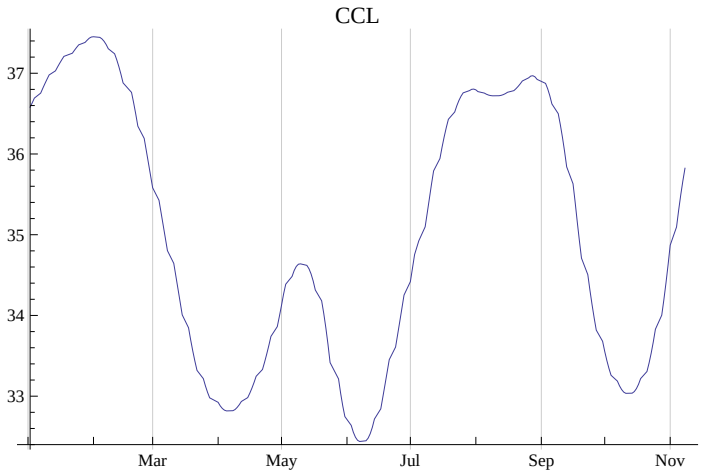
In[243]:=
```
price = FinancialData["CUK", "Jan. 1, 2013"];
dwd =
  DiscreteWaveletTransform[price[[All, 2]], BiorthogonalSplineWavelet[3, 3], 4];
trCUK = InverseWaveletTransform[WaveletThreshold[dwd, {"Hard", 100}]];
DateListPlot[Transpose[{price[[All, 1]], trCUK}],
 Joined → True, Frame → False, PlotLabel → "CUK"]
```

Out[246]=



In[247]:=
```
price = FinancialData["CCL", "Jan. 1, 2013"];
dwd =
  DiscreteWaveletTransform[price[[All, 2]], BiorthogonalSplineWavelet[3, 3], 4];
trCCL = InverseWaveletTransform[WaveletThreshold[dwd, {"Hard", 100}]];
DateListPlot[Transpose[{price[[All, 1]], trCCL}],
 Joined → True, Frame → False, PlotLabel → "CCL"]
```

Out[250]=



Distances between the Trend curves are smaller than the distances to the random-pick:

In[251]:=
```
CanberraDistance[trCHEV, trCCL]
CanberraDistance[trCUK, trCCL]
CanberraDistance[trCUK, trCHEV]
```

Out[251]= 114.295

Out[252]= 4.05922

Out[253]= 117.171

FAIL

Compute the mean distance to the random-pick outliers:

```
data = Table[CanberraDistance[trCHEV, tr[[i]]], {i, 1, Length[tr]}];
Mean[data]
data = Table[CanberraDistance[trCCL, tr[[i]]], {i, 1, Length[tr]}];
Mean[data]
data = Table[CanberraDistance[trCUK, tr[[i]]], {i, 1, Length[tr]}];
Mean[data]
```

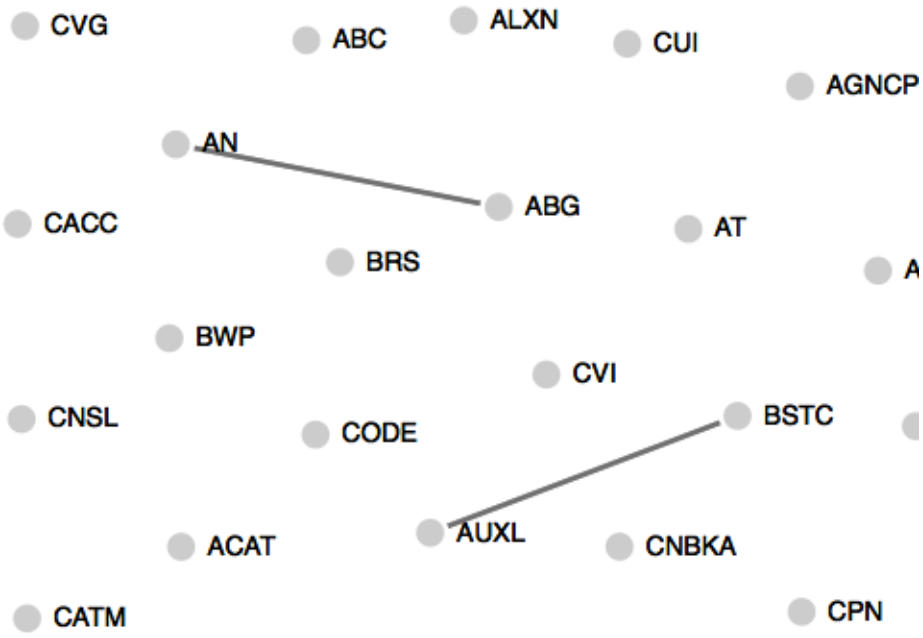
Out[255]= 94.5014

Out[257]= 93.6576

Out[259]= 94.7333

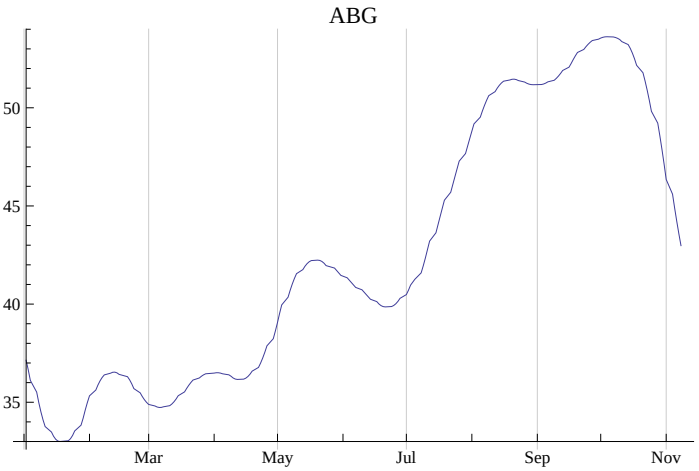
Out[261]= 94.7333

Similarity: AN, ABG



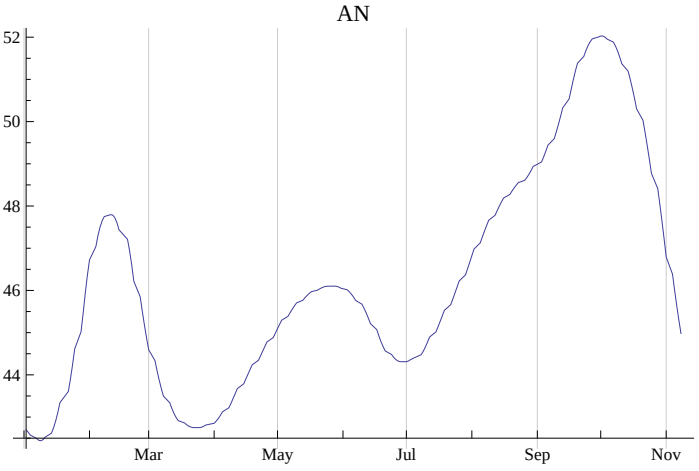
In[262]:=
```
price = FinancialData["ABG", "Jan. 1, 2013"];
dwd =
  DiscreteWaveletTransform[price[[All, 2]], BiorthogonalSplineWavelet[3, 3], 4];
trABG = InverseWaveletTransform[WaveletThreshold[dwd, {"Hard", 100}]];
DateListPlot[Transpose[{price[[All, 1]], trABG}],
 Joined → True, Frame → False, PlotLabel → "ABG"]
```

Out[265]=



In[266]:=
```
price = FinancialData["AN", "Jan. 1, 2013"];
dwd =
  DiscreteWaveletTransform[price[[All, 2]], BiorthogonalSplineWavelet[3, 3], 4];
trAN = InverseWaveletTransform[WaveletThreshold[dwd, {"Hard", 100}]];
DateListPlot[Transpose[{price[[All, 1]], trAN}],
 Joined → True, Frame → False, PlotLabel → "AN"]
```

Out[269]=



Canberra distance for the two stocks' Trend curve:

In[270]:=
```
CanberraDistance[trABG, trAN]
```

Out[270]= 12.9153

OK!

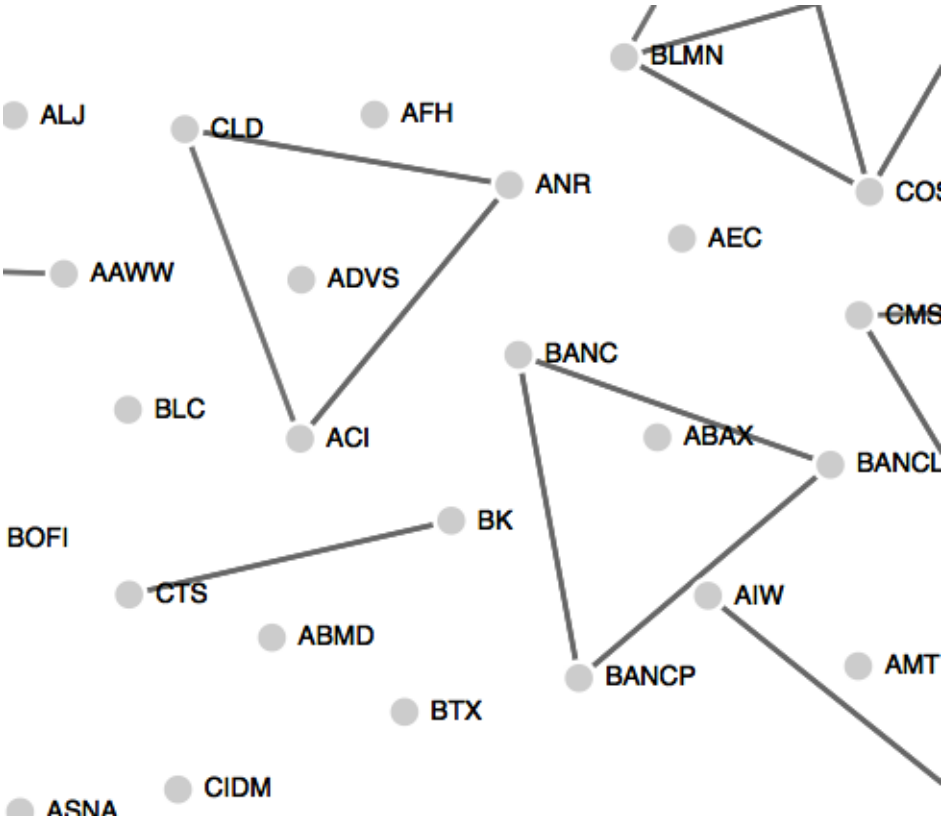
How far they are from the random-pick:

In[271]:=
```
data = Table[CanberraDistance[trABG, tr[[i]]], {i, 1, Length[tr]}];
Mean[data]
data = Table[CanberraDistance[trAN, tr[[i]]], {i, 1, Length[tr]}];
Mean[data]
```

Out[272]= 99.3463

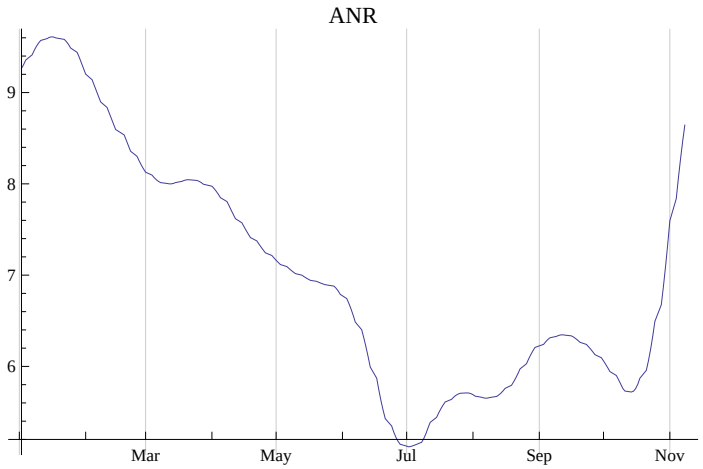
Out[274]= 102.071

Similarity: ANR, ACI, CLD



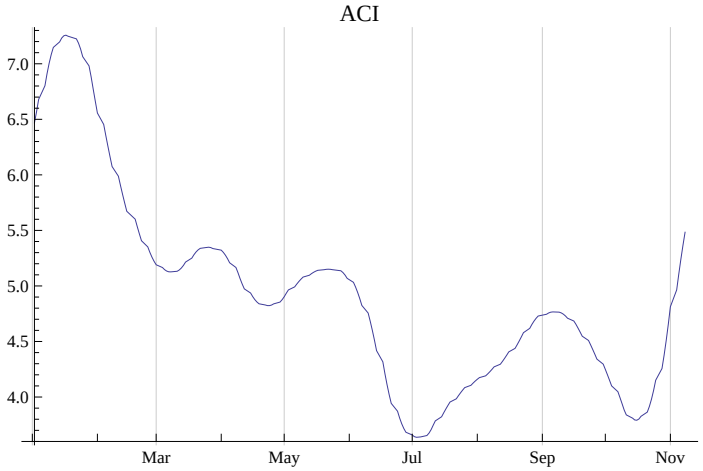
In[275]:=
```
price = FinancialData["ANR", "Jan. 1, 2013"];
dwd =
  DiscreteWaveletTransform[price[[All, 2]], BiorthogonalSplineWavelet[3, 3], 4];
trANR = InverseWaveletTransform[WaveletThreshold[dwd, {"Hard", 100}]];
DateListPlot[Transpose[{price[[All, 1]], trANR}],
 Joined → True, Frame → False, PlotLabel → "ANR"]
```

Out[278]=



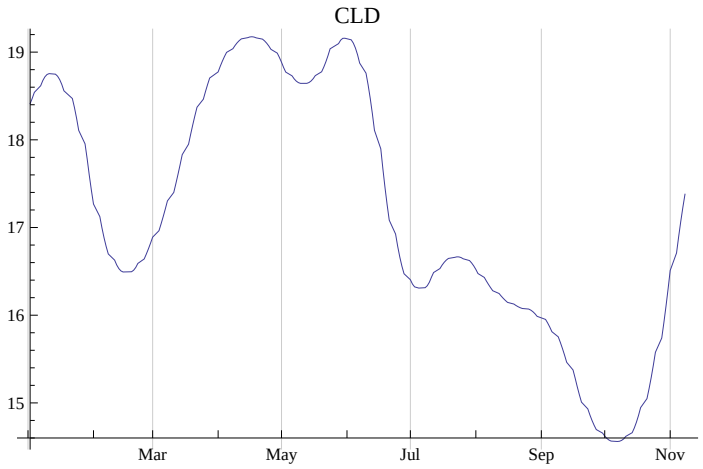
In[279]:=
```
price = FinancialData["ACI", "Jan. 1, 2013"];
dwd =
  DiscreteWaveletTransform[price[[All, 2]], BiorthogonalSplineWavelet[3, 3], 4];
trACI = InverseWaveletTransform[WaveletThreshold[dwd, {"Hard", 100}]];
DateListPlot[Transpose[{price[[All, 1]], trACI}],
 Joined → True, Frame → False, PlotLabel → "ACI"]
```

Out[282]=



In[283]:=
```
price = FinancialData["CLD", "Jan. 1, 2013"];
dwd =
  DiscreteWaveletTransform[price[[All, 2]], BiorthogonalSplineWavelet[3, 3], 4];
trCLD = InverseWaveletTransform[WaveletThreshold[dwd, {"Hard", 100}]];
DateListPlot[Transpose[{price[[All, 1]], trCLD}],
 Joined → True, Frame → False, PlotLabel → "CLD"]
```

Out[286]=



Pair-wise Canberra distances less than the distance to Trend curve of the random-pick:

In[287]:=
```
CanberraDistance[trANR, trACI]
CanberraDistance[trCLD, trACI]
CanberraDistance[trANR, trCLD]
```

Out[287]= 38.0626

Out[288]= 120.682

Out[289]= 91.5921

Fail: Partial

Random-pick Canberra distances:

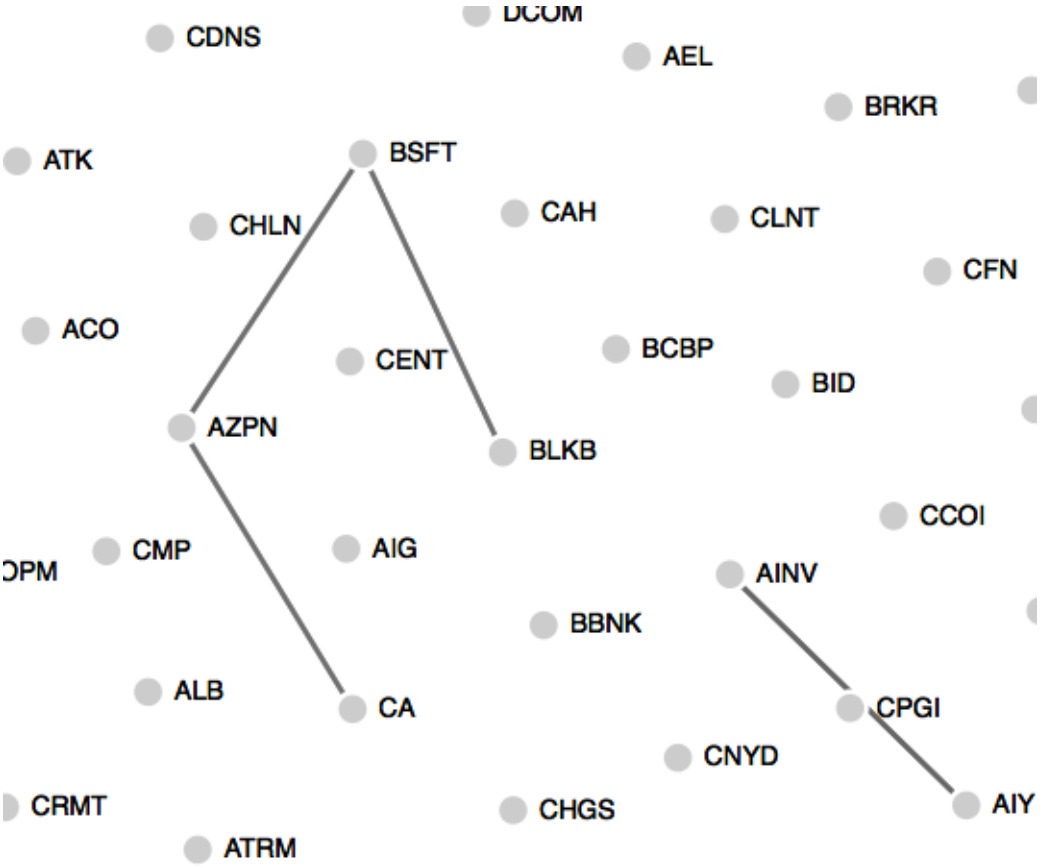
In[290]:=
```
data = Table[CanberraDistance[trACI, tr[[i]]], {i, 1, Length[tr]}];
Mean[data]
data = Table[CanberraDistance[trCLD, tr[[i]]], {i, 1, Length[tr]}];
Mean[data]
data = Table[CanberraDistance[trANR, tr[[i]]], {i, 1, Length[tr]}];
Mean[data]
```

Out[291]= 113.244

Out[293]= 91.0136

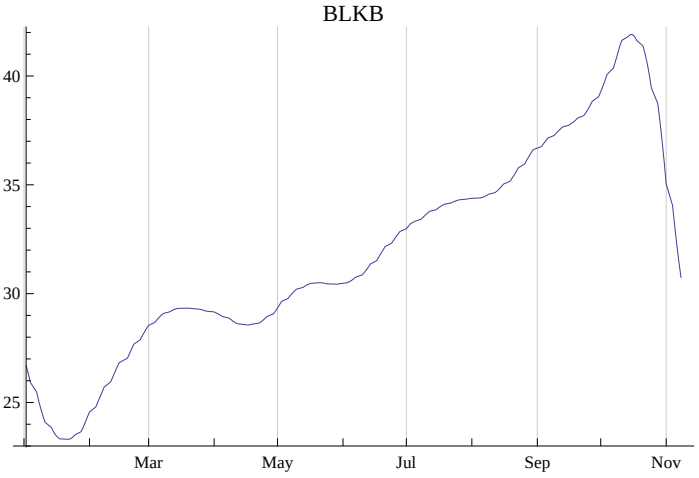
Out[295]= 104.745

Similarity: BLKB, BSFT, AZPN, CA



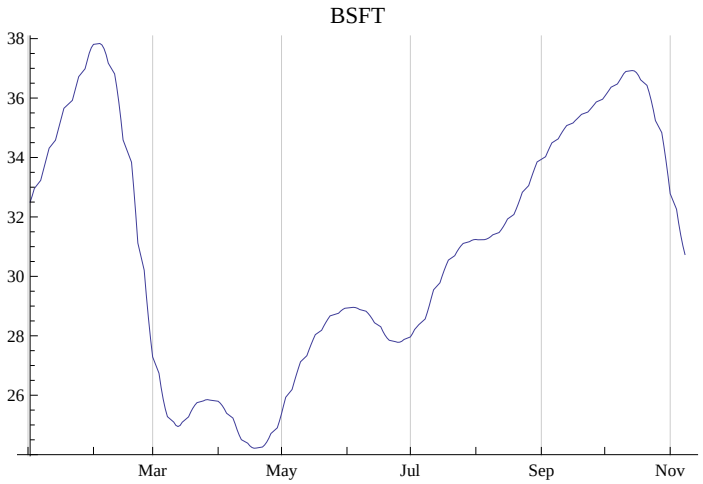
In[296]:=
```
price = FinancialData["BLKB", "Jan. 1, 2013"];
dwd =
  DiscreteWaveletTransform[price[[All, 2]], BiorthogonalSplineWavelet[3, 3], 4];
trBLKB = InverseWaveletTransform[WaveletThreshold[dwd, {"Hard", 100}]];
DateListPlot[Transpose[{price[[All, 1]], trBLKB}],
 Joined → True, Frame → False, PlotLabel → "BLKB"]
```

Out[299]=



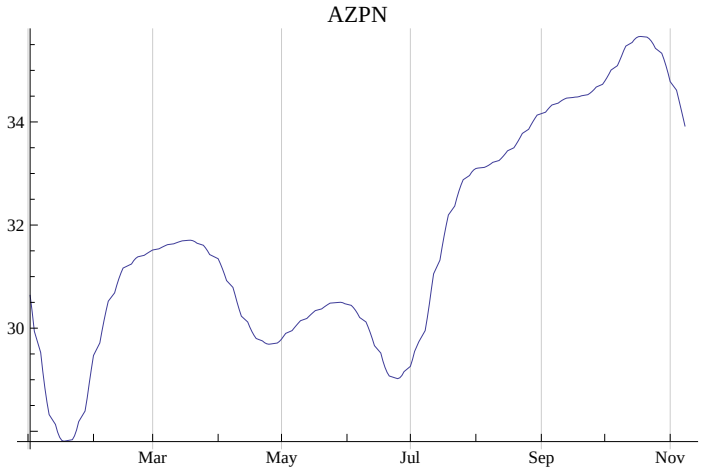
In[300]:=
```
price = FinancialData["BSFT", "Jan. 1, 2013"];
dwd =
  DiscreteWaveletTransform[price[[All, 2]], BiorthogonalSplineWavelet[3, 3], 4];
trBSFT = InverseWaveletTransform[WaveletThreshold[dwd, {"Hard", 100}]];
DateListPlot[Transpose[{price[[All, 1]], trBSFT}],
 Joined → True, Frame → False, PlotLabel → "BSFT"]
```

Out[303]=



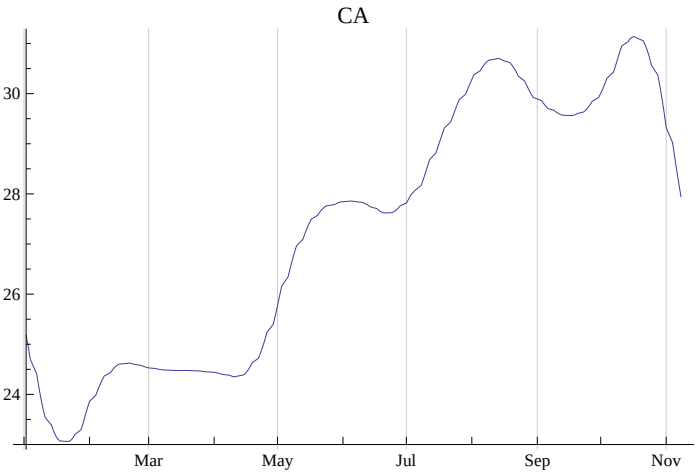
In[304]:=
```
price = FinancialData["AZPN", "Jan. 1, 2013"];
dwd =
  DiscreteWaveletTransform[price[[All, 2]], BiorthogonalSplineWavelet[3, 3], 4];
trAZPN = InverseWaveletTransform[WaveletThreshold[dwd, {"Hard", 100}]];
DateListPlot[Transpose[{price[[All, 1]], trAZPN}],
 Joined → True, Frame → False, PlotLabel → "AZPN"]
```

Out[307]=



In[308]:=
```
price = FinancialData["CA", "Jan. 1, 2013"];
dwd =
  DiscreteWaveletTransform[price[[All, 2]], BiorthogonalSplineWavelet[3, 3], 4];
trCA = InverseWaveletTransform[WaveletThreshold[dwd, {"Hard", 100}]];
DateListPlot[Transpose[{price[[All, 1]], trCA}],
 Joined → True, Frame → False, PlotLabel → "CA"]
```

Out[311]=



Pair-wise Canberra distances less than the distance to Trend curve of the random-pick:

In[312]:=
```
CanberraDistance[trBLKB, trBSFT]
CanberraDistance[trBSFT, trAZPN]
CanberraDistance[trAZPN, trCA]
```

Out[312]= 15.6775

Out[313]= 10.9233

Out[314]= 16.467

OK!

Random-pick Canberra distances:

In[315]:=
```
data = Table[CanberraDistance[trBLKB, tr[[i]]], {i, 1, Length[tr]}];
Mean[data]
data = Table[CanberraDistance[trBSFT, tr[[i]]], {i, 1, Length[tr]}];
Mean[data]
data = Table[CanberraDistance[trAZPN, tr[[i]]], {i, 1, Length[tr]}];
Mean[data]
data = Table[CanberraDistance[trCA, tr[[i]]], {i, 1, Length[tr]}];
Mean[data]
```

Out[316]= 91.4096

Out[318]= 91.2411

Out[320]= 91.3239

Out[322]= 90.2452