WILEY Expert Systems

**ARTICLE**

# Ramex-Forum: a tool for displaying and analysing complex sequential patterns of financial products

## Pedro Tiple[1] | Luis Cavique[2] | Nuno Cavalheiro Marques[3]

[1] Research Department, GoBusiness Finance, Lisbon, Portugal

[2] Universidade Aberta, Lisbon, Portugal

[3] NOVALINCS-DI/FCT, Universidade Nova de Lisboa, Caparica, Portugal

**Correspondence**

Nuno Cavalheiro Marques, NOVALINCS-DI/FCT, Universidade Nova de Lisboa, Caparica, Portugal.
Email: nmm@fct.unl.pt

## Abstract

Financial data provides a valuable up-to-date knowledge of the world economy. However, it is presented in extremely large data volumes, in diverse formats, and is constantly being updated at a high speed. The Ramex-Forum algorithm is oriented to guide financial experts in finding new and relevant information. We present a sensitivity analysis and new visualizations using an improved version of the Ramex-Forum algorithm. The proposed algorithm is applied to two case studies – the petroleum production chain and the European financial institutions risk analysis. Different combinations of parameters and new ways to visualize data are used. Results highlight the importance of Ramex-Forum for analysing relevant relationships in price variations in financial markets.

### KEYWORDS

business intelligence, data analysis, European financial data, petroleum production chain, rule learning

## 1 | INTRODUCTION

Sequential pattern mining is a subtopic of data mining, where the values are organized in a sequence, which aims to find relevant patterns in data samples. This includes extracting frequent occurring patterns and comparing sequences for similarities. Sequence mining techniques are typically supported by association rule learning. Most of the sequence pattern discovery techniques present three common handicaps: the need of parameters, the huge number of rules that do not permit a global view, and scalability problems.

Ramex algorithm by Cavique (2007; 2015) provides a comprehensive view of the sequences, providing the user the visualization of the data sequences with a special kind of tree, a poly-tree, which shows all the items, but only the most relevant sequences, thus retrieving a bird-eye view of the dataset structure. Ramex has been implemented in different scenarios: web mining by Cavique (2007), marketing by Cavique & Coelho (2008), Cavique (2015), and financial studies by Marques and Cavique (2013) and Tiple (2014). Ramex-Forum comparison with other sequence mining algorithms is discussed in the study of Cavique (2015). Previous results show that the use of poly-tree methods highlights relevant patterns that are usually hidden, for example, by process mining methods. Here, the Ramex-Forum original proposal of Marques and Cavique (2013) will be improved and tuned for finding sequential patterns using two case studies. The goals of this work are the following: (a) to define products' influence based on time series databases,

(b) to represent products' influence in a way that makes sequences easy to recognize by a non-expert user, and (c) to analyse the influences on financial products.

This work follows EPIA–KDBI 2015 presentation (Tiple et al., 2015) describing a study on how Ramex-Forum can be used to quantify how the price of the crude oil (raw material) can influence the price of manufactured products. Extended examples with improved definitions, additional measures regarding data correlation, and a second data set representing the dependences in major European financial institutions are presented for this special issue of the Expert Systems Journal of Knowledge Engineering.

The paper is structured as follows: Section 2 defines the moving average (MA) strategies and influential products. Section 3 presents Ramex-Forum algorithm. In Section 4, the petroleum production chain problem is presented, and different combinations of parameters are shown. In Section 5, a second case study regarding European financial institutions is presented and discussed. Finally, in Section 6, we draw some conclusions.

## 2 | RELATED WORK

In this section, concepts are presented and definitions are established in order to be reused in the following sections. MA strategies are introduced, and the definition of poly-tree is established.

wileyonlinelibrary.com/journal/exsy

## 2.1 | Moving average strategies

The MA is widely used to create decision rules in financial trading environments (e.g., Gunasekarage & Power, 2001; Fong & Yong, 2005). We assume a crossover strategy to buy and sell financial products (e.g., Chan, 2009). Given the product price index in time $t$, $I(t)$, the time series will be smooth using MA, (e.g., Makridakis & Wheelwright, 1978).

An MA of a set of numbers is obtained by calculating the averages of the corresponding subsets with a predefined length of $n$ days. A short-term MA can be expressed by $n = 5$, and a long-term MA for values of $n = 100$, 200, or larger. The $n = 100$ with length $n$ days, in time $t$ is given by the following:

$$MA(t, n) = \sum_{j=t-n}^{t} I(j)/n.$$

Crossover strategies can be defined by using two MAs, a short-term MA and a long-term MA. When the short-term MA crosses the long-term MA and remains below, it is a selling signal as it indicates a downward shifting trend. This is known as a "sell signal", represented in Figure 1 with a red arrow.

On the other hand, when the short-term MA crosses above the long-term MA, it is a buying signal as it indicates the trend is shifting up. This is known as a "buy signal", represented in Figure 1 with a green arrow.

The buy or sell decision rule for each instant of time $t$ can be defined as follows:

- Buy signal(t), if $MA(t, \text{short - term}) > MA(t, \text{long - term}) \cdot (1 + \varepsilon)$

- Sell signal(t), if $MA(t, \text{short - term}) \leq MA(t, \text{long - term}) \cdot (1 - \varepsilon)$

Where $\varepsilon$ is a small value for a smoothing coefficient aimed at attenuating noise.

In Figure 1, $MA(t, 40)$ and $MA(t, 160)$ were used as short and long-term MAs ($n$ is set in trading days), and we set $\varepsilon = 0$. To avoid the time lag before signal detection and for easier parameter analysis, in this paper, we have set the short-term MA to daily prices (i.e., $I(t)$). Values for parameter $\varepsilon$ are studied in Section 2.



**FIGURE 1** Financial product (normalized DJI index in black) and respective short-term (blue) and long-term moving average (red) and crossover identifying Buy signal (green) and Sell signal (red)

## 2.2 | Graphs, trees and poly-trees concepts

One of the goals of this work is to find new ways of data visualization, where a non-expert user can easily recognize sequences. On the one hand, graphical representation of a cyclic graph does not have a good visualization, and the identification of large sequences is almost impossible. On the other hand, we want to include all the products. Following the study of Rebane and Pearl (1987), the trade-off between visualizing all products and finding large sequences is given by a subgraph with N nodes and N-1 edges: the poly-tree.

Given a connected undirected graph, a spanning tree of the graph is a subgraph that connects all the vertexes. A minimum weight spanning tree is the spanning tree with a weight that is lower than or equal to the weight of every other spanning tree. This problem is easily solved using a greedy algorithm. The algorithms proposed by Kruskal and Prim are well-known examples (e.g., Cormen et al., 2009). In Kruskal's algorithm, the edges are chosen without worrying about the connections to previous edges, but avoiding the cycles. In Prim's algorithm, the tree grows from an arbitrary root.

In oriented or direct graphs (digraphs), the number of arcs entering a node is called in-degree, and the number of arcs leaving a node is called out-degree. An oriented tree is a direct acyclic graph with an in-degree node equal to one, with the exception of the root, where the in-degree is equal to zero. A poly-tree is a relaxed oriented tree, with one (and only one) path between any pair of nodes, where the in-degree of a node can be greater than one. The term poly-tree was coined by Rebane & Pearl (1987), and the poly-tree algorithm is well-known for inference in Bayesian networks.

A tree, with a in-degree node equal to one, a poly-tree, with a node in-degree greater than one, and a direct acyclic graph, with two or more paths between a pair of nodes, are shown in Figure 2.
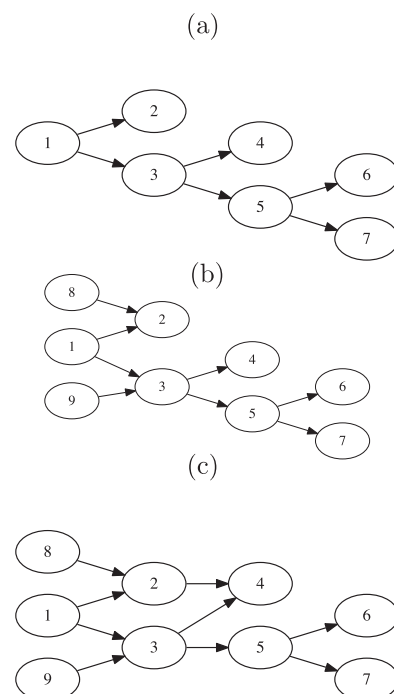


**FIGURE 2** (a) Tree, (b) poly-tree, and (c) direct acyclic graph

The algorithms of Prim and Kruskal are used in undirected graphs. For directed graphs, Edmonds (1967) branching algorithm and Fulkerson (1974) algorithm can be referred. However, both algorithms generate trees, with in-degree zero or one. To find the maximum weighted poly-tree, there are few or nonexistent bibliographic references.

# 3 | RAMEX-FORUM ALGORITHM

Most of the sequence pattern discovery techniques, based on association rules, present three common handicaps:

(i). Parameters: The user must specify a minimum support threshold to find the desired patterns. A useless output can be expected by pruning either too many or too few items. The process must be repeated interactively, which becomes very time consuming for large databases.

(ii). Number of Rules: The association rules' systems that support the item-set and sequence mining usually generate a huge number of rules, and therefore, it is difficult for the user to decide which rules to use.

(iii). Scalability: Because most of the existing algorithms use a lattice structure in the search space and need to scan the database more than once, they are not compatible with very large databases.

In order to create holistic environments in data mining, micro-patterns and macro-patterns must be differentiated. The micro-patterns correspond to small percentages of data; for instance, in association rules, it is common to have a measure of support that includes support values ≥5%, with high confidence rules being chosen. On the other hand, the macro-patterns involve a large percentage of data, for example, in the regression model, all data elements are used. The micro-patterns are characterized by high confidence, while macro-patterns are characterized by high support. There are other examples of micro-patterns: in sequence mining, a support ≥1% is frequent; in the classification problem, by using decision trees, each branch of the tree corresponds to a small percentage of the data; in the classification problem using the k-nearest neighbor, the comparisons are made using a reduced number of k elements. Finally, regarding macro-patterns in techniques such as regression, hypothesis testing, clustering or reduction of attributes, all data are taken into account.

The proposed algorithm was coined with the Latin name, Ramex, meaning "branch of a tree" and Forum, meaning "market". Ramex-Forum introduces a new vision for classic problems of sequence mining, allowing the search of macro-patterns instead of micro-patterns.

The purpose of Ramex-Forum algorithm is to show a bird's-eye view of complex systems with multiple dependencies.

Given a sequence of measurements of the products collected over time, we define it as a time series of distinct financial products: *Prod(i)*, $i \subseteq 1 \ldots m$. The Ramex algorithm uses a poly-tree sequence model with two phases. The first phase transforms pairs of time series into a network with m nodes, and the second phase searches for the most relevant product sequences.

*Ramex-Forum, Two Phase Algorithm*
**Input:** raw data time series of *Prod(i)*;
**Output:** influence poly-tree
1. Problem transformation creates a network where the nodes are the products and the edges are a measure of influence between products (Section 1).
2. Search for the most influential poly-tree of the products (Section 3.2).

## 3.1 | Ramex-forum algorithm – phase I

In the first phase, the problem transformation takes place, where a dataset is replaced by a network of products.

In the network, each node corresponds to a product, and each transition represents the influence of one product over another product. The weight of each edge is calculated with an influence measure.

In this paper, we define "influenced product" as a financial product influenced by another product if there are many occurrences of the same signal (*buy sell*), in previous periods of time. We introduce the concept of "signal counter" that corresponds to the number of consecutive days with the same buy or sell signal.

$$MA(t, short \text{-} term) = MA(t, long \text{-} term)$$

**Definition 1.** A market signal counter of a specific product P is denoted by signal - counter$(P, t)$ and it is characterized by the parameter $t$ corresponding to the time and a date or a numeric timestamp associated to the event. As presented in Section 2, Buy/Sell signal$(t)$ switches whenever a crossover occurs, i.e., the difference between short and long-term *MA*s is zero. Buy signal$(t)$ is represented by positive integer numbers and Sell signal$(t)$ is represented by negative integer numbers. At the crossover point the signal-counter is set to zero, otherwise it is incremented by one, as follows:

- signal - counter$(P, t) = 0$ if MA$(t, short \text{-} term) = $ MA$(t, long \text{-} term)$
- signal - counter$(P, t) = $ signal - counter$(P, t - 1) + 1$ if Buy signal$(t)$
- signal - counter$(P, t) = $ signal - counter$(P, t - 1) - 1$ if Sell signal$(t)$

Table 1 presents the signal count of five financial products. For buying trend, positive figures are used and for selling trend, negative figures are used in the signal counters.

As stated before, the purpose of this paper is to find products that influence other products, when there is a co-occurrence of two market signals. With an interval between events of, for example, 200 days the

**TABLE 1** Signal-counters of _ve _nancial products

| Time | A | B | C | D | E |
|---|---|---|---|---|---|
| 20150919 | 1 | −8 | 0 | 3 | 4 |
| 20150920 | 2 | −9 | 1 | 4 | 5 |
| 20150921 | 3 | 0 | 2 | 5 | 0 |
| 20150922 | 4 | 1 | 3 | 0 | −1 |
| 20150923 | 5 | 2 | 4 | −1 | 0 |
| 20150924 | 6 | 3 | 0 | −2 | 1 |

**TABLE 2** Influence of products for latency of 4 and Buy signal

| Time t | A→B | A→C | D→A | E→A | C→B | B→E | D→C | E→C | E→D | Others |
|---|---|---|---|---|---|---|---|---|---|---|
| 20150919 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0 |
| 20150920 | 0 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 0 |
| 20150921 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| 20150922 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 20150923 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 20150924 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| $\sum$ | 3 | 4 | 3 | 2 | 2 | 1 | 2 | 1 | 2 | 0 |

influence is not plausible, so a threshold should be imposed. The difference of days of the occurrences is limited to a number of days or latency $\delta$.

**Definition 2.** *Product X influences product Y, that is, $X \rightarrow Y$, with a maximum latency $\delta$ and for a specific signal in time t denoted by Influence$(X, Y, \delta, signal\_, t)$, when the following is true:*

- $0 < (|\,signal - counter(X, t)| - |\,signal - counter(Y, t)|\,) \leq \delta$
- and signal - counter$(X, t) \neq 0$
- and signal - counter$(Y, t) \neq 0$
- and signal$(X, t) =$ signal$(Y, t) =$ signal$\_$
- signal$\_ \in Buy$ , $Sell$

To exemplify the influence products, let $\delta = 4$, a signal type of Buy (all positive values), and Table 1 data. Comparing products A and B, the sum of the function *Influence$(A, B, 4, Buy, t)$* returns the value 3. The counts for every product are presented in Table 1, and the figures for every pair of products *A, B, C,* and *D* are shown in Table 2.

Given the cumulative data in Table 2, an influence graph for buy signal can be represented in Figure 3.

The sum of influence counts is a simple and efficient way to express the relevance of the influence between each pair of products. The influence graph presents cycles, and even with five nodes, it is difficult to establish useful rules. The use and exploration of the whole graph will be developed in the second phase of the Ramex-Forum algorithm.

## 3.2 | Ramex-forum algorithm – phase II

Because the graph is difficult to read, a transformation must be performed. In order to transform the graph into a tree or poly-tree, two heuristics can be used.
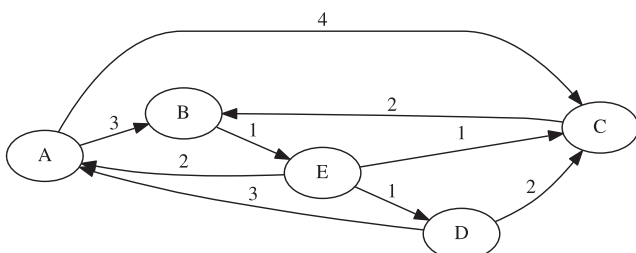


**FIGURE 3** Influence graph

When a starting node is known, the Forward Heuristic, which generates a tree of products, can be applied. When there is no starting node, we apply the Back-and-Forward Heuristic that is able to create a poly-tree with all the financial products.

To find the most influence poly-tree sequence of financial products, we are going to use the Back-and-Forward heuristic, because there is no information about a starting vertex.

**Poly-tree generation – Back-and-Forward heuristic, phase II**
**Input:** graph $G(V, E)$
**Output:** poly-tree $PT$
Initialize $PT$;
*for* $|V|$-1 *times* **do**
    **foreach** *edge* $\subseteq G\backslash PT$ **do**
    Calculate $x = argmax($ forward-node $x$ not visited in $G$ and
    connected with $PT$ or back-node $x$ not visited in $G$ and connected
    with $PT$);
    **end**
    *Update solution PT with node x*;
**end**

The Back-and-Forward algorithm is based on the Prim algorithm. Prim algorithm finds the minimum or maximum spanning-tree in an undirected graphs. Back-and-Forward heuristic finds the maximum weighted poly-tree by applying the same technique to directed graphs.

The algorithms of Prim and Kruskal are used in weighted undirected graphs to solve the minimum (or maximum) spanning tree in polynomial time. For weighted directed graphs, the algorithms of Edmonds (1967) and Fulkerson (1974) also generate optimum branching in polynomial time. However, both algorithms generate trees, with in-degree zero or one. To find the maximum weighted poly-tree in a weighted directed graph, we developed a heuristic, based on the Prim algorithm. The algorithm complexity is also polynomial, and we believe it finds the optimal poly-tree, because it explores all the weighted arcs as the Prim algorithm does in undirected graph.

The time complexity of the Back-and-Forward heuristic is equal to the Prim algorithm, which is $O(E \cdot N)$ with $E$ edges and $N$ nodes.

The result of the influence poly-tree of our running example is given in Figure 4, with a sum of the influence arcs of 12.

At this stage, we can infer that products D and E are the most influencing products for the given set, because their signal-counter is greater than the following ones.

Finally, we summarize the goals, the steps, and the parameters of Ramex-Forum. The Ramex-Forum goal is to create a clear visualization
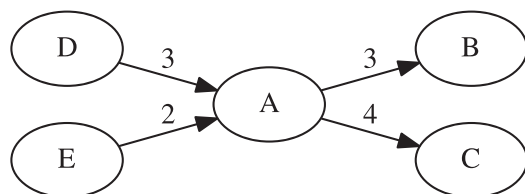
**FIGURE 4** Influence poly-tree

of influence products, for example, products that influence other products. The algorithm is divided into two steps, the graph generation and the transformation into a poly-tree. The algorithm only uses two parameters:

- value $\varepsilon$, a small value to smooth the noise in MA s, presented in Section 2;
- latency $\delta$, to limit the difference of days between the occurrences of two products.

## 4 | PETROLEUM PRODUCTION CHAIN

Petroleum is one of the most important resources in the developed world, and it is still a major variable influencing the economy and markets. The price of petroleum and its derivatives is not influenced simply by supply and demand; taxes, speculation, wars, costs in refinement, and transportation all contribute in setting prices. Having that in mind, it is already clear that there will not be a 1:1 price ratio between a source material and its derivatives. Another relevant aspect is that because there is a lengthy refinement process, a significant increase in the price of the source material should reflect in the price of its derivatives after a period equal to the time it takes to refine – usually within 3–4 weeks (Borenstein et al., 1996). Moreover, following Suviolahti (2009), due to its high economic importance and cost, the price of crude oil should always reflect on the final price.

When a product is transformed from a source material, its final price will be dependent on several things like transformation process, distribution and marketing costs, taxes, and demand. These costs vary from product to product; however, the price of the source material will always reflect on the final price (Suviolahti, 2009). The price of crude oil has a heavy weight on the prices of its derived products, from the price at the refineries to the retail price of all its products. Also, due to production chain, the price at the refinery should affect the retail price on the surrounding area, while price fluctuations at city level will be faster than state averages.

This section presents a study on a method to quantify how the price of the crude oil (raw material) can influence the price of manufactured products by using Ramex-Forum.

### 4.1 | Data and information gathered

Following the study of Gary and Handwerk (2001), petroleum is refined into a relatively extensive list of categories, with each category having hundreds of subproducts, namely petroleum, can be processed

and refined into several other goods and substances such as gasoline, kerosene, asphalt, paraffin, plastics, and lubricants. Moreover, this division and classification mostly depends on its social usage. A set of historical data is provided by the U.S. Energy Information Administration.[1] This is a very large publicly available repository of historical values for a wide range of petroleum related subjects. The time-frame of available data ranges start from the 1980s to 2006 (different prices began being tracked at different times). Studied data was downloaded on June 2014 and ranges from January 2006 to June 2014. The variations in the prices of these products are also compared with the stock value of the companies that produce them. For this, the stock market value of eleven corporations dedicated to extracting, processing, and selling of crude and crude related products were selected (from an internal database), and its stock value was compared with the previously mentioned products.

The different products that represent data for 55 prices are detailed in Table 3. The prices are separated into retail/bulk price and spot price; this means that for some items the price is taken from retail sellers and for other items it is the security price at that day. The data is separated into four categories: known benchmarks of Hammoudeh et al., (2008) for crude oil (West Texas Intermediate as OklahomaWTI, European Brent, and the OPEC Basket – an average of the oil price in 12 oil-exporting developing nations); refinery price for Gasoline (New York Harbor and Gulf Coast), RBOB Gasoline (Los Angeles), Diesel (New York Harbor, Gulf Coast, and Lost Angeles), Kerosene Jet (Gulf Coast), Propane (Gulf Coast), and Heating Oil (New York Harbor); national (Retail Gas Price as RGP, Retail Diesel Price as RDP), state (StateRGP, StateRDP), and city (CityRGP, CityRDP) averages for regular gasoline and diesel; Corporation stock values.

### 4.2 | Parameter setup

Different parameter sets in Ramex-Forum yield significantly distinct results. This paper studies the influence and best values for parameters $\delta$, $\varepsilon$ thresholds, and MA size. Focus will be put on the *Buy* comparison because in the selected data, the increase/decrease of prices is very asymmetrical with a strong lean towards increases.

For better parameter comparison, it is useful to consider an additional measure. The *average edge weight* consists of summing the weight of all the edges in the output of the first stage of the Ramex-Forum algorithm and dividing that sum by the number of edges in the graph:

$$average\ edge\ weight(V, E) = \sum weight(e)/|E|\,,\ \forall\,e \in E.$$

#### 4.2.1 | Parameter $\delta$

Parameter $\delta$ was analyzed regarding its effect on the average edge weight changes. The result can be seen in Figure 5(a). The chart shows the average edge weight change for each increment in the value of $\delta$. Each line represents the results obtained using different MA sizes of 40, 80, 120, 160, 200, or 240 days. Several big spikes can be seen every 5 days; this is because gas and diesel prices at the pump are only registered on a weekly basis, so for each 5 day increase in $\delta$, the algorithm will pick up another change in value. This makes the analysis

---

[1]The used data was downloaded from http://www.eia.gov/ on June 2014 and ranges from January 2006 to June 2014.

**TABLE 3**    Regions, divided by product type

| Spot | Retail gas price | Retail diesel price | Corporation |
|---|---|---|---|
| European Brent | Boston, California | California | Chevron |
| Gulf Coast Diesel | Chicago, Cleveland | East Coast | Hess |
| Gulf Coast GAS | Colorado, Denver | Gulf Coast | Exxon Mobil |
| Gulf Coast Kerosene Jet | East Coast, Florida | Midwest | Murphy Oil |
| Los Angeles Diesel | Gulf Coast, Houston | Rocky Mountain | Holly Frontier |
| Los Angeles RBOB | Los Angeles, US | West Coast | Vulcan Materials |
| NY Harbor Diesel | Miami, Midwest | | Valero Energy |
| NY Harbor GAS | Minnesota, New York City | | Occidental Petrl. |
| NY Harbor Heating Oil | New York, Texas | | Devon Energy |
| OklahomaWTI | Rocky Mountain, Ohio | | Anadarko Petrl. |
| OPEC Basket | Seattle, San Francisco | | |
| Texas Propane | Massachusetts, Washington | | |
| | West Coast | | |

(a)



(b)
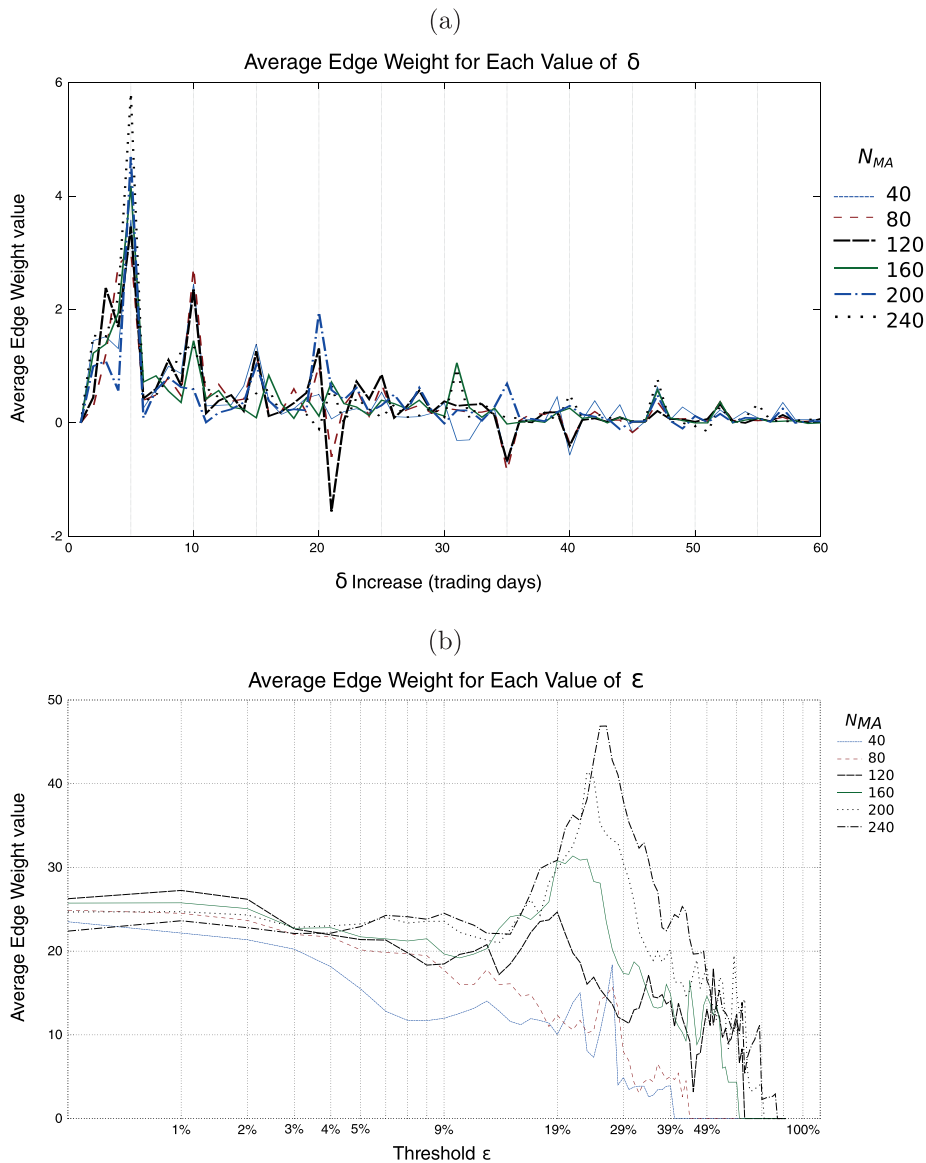


**FIGURE 5**    Graphs showing the change in: (a) average edge weight with each increment of using the Buy comparison; (b) average edge weight and number of nodes with each increment in the threshold interval for using the Buy comparison

somewhat harder, but it is still useful as changes in retail prices are now clearly identified. The first thing that stands out is that in the first week, there is already a noticeable increase in the average edge weight; however, some of it is due to influences between retail prices and not only from refinery to retail prices. Second, after the fourth week, the individual increases in δ barely produce a meaningful increase in value, still the cumulative increases are significant. The parameter δ was fixed at a value of 30 working days (around 6 weeks: 2 weeks more than the expected).

### 4.2.2 | Parameter ε

Parameter ε was studied trying to find the best combination of parameters. The algorithm was run several times, and the average edge weight value was recorded for each run. The best values for threshold interval and the MA size are represented in Figure 5(b). The graph shows the progression of the average edge weight, in relation to increase in threshold size. The parameters that lead to the highest increase in average weight can be clearly identified as the MA size of 240 days with a threshold of around 26% of the MA. However, things change when the influence event count is also considered: Increasing the threshold rapidly decreases the number of detected events (a threshold of 26% will reduce the number of events by about 80%). In this case, the starting average is around 130 events and falls to 30, in the 3-year period analysed: a very low average number of events. For this case study, the choice was made to maximize the event count so that a broader spectrum of influences can be detected instead of restricting the analysis to situations where the prices rise or fall sharply (detected with higher threshold values). Small increases in the ε threshold will raise the average weight while only lowering the event count by small amounts. Nevertheless, random fluctuations do not advise going for a threshold of 0%, so this trade-off seems to favor the usage of smaller values for ε.

### 4.2.3 | Moving average size, $N_{MA}$

The choices for available MA sizes were based on the study of Marques and Cavique (2013), and the graphs show that maximizing this parameter yields the best results. However, the user still needs to take into account what it means to increase the MA size: the bigger the MA is, the smoother the curve will be and, thus, the more the MA will behave like a noise filter (i.e., by becoming less and less sensitive to small changes in the behavior of the product). The values overlap for small ε values, and it is hard to read the effects of the first increments in a linear scale, so Figure 5(b) uses a logarithmic scale for representing ε values, showing that the 240 and 120 MA sizes have a very similar behavior. The average weight for an MA of 120 days has a higher starting value than the 240 days one. This means that for a *buy signal*, best thresholds are the following: $ε = 1\% \wedge δ = 30 \wedge N_{MA} = 120$.

Figure 6 shows that it is possible to find more than just sequential patterns with these parameters. Graph colors were added to each node according to their product type. These colors show a clear grouping of product types, with same color nodes mostly close to each other. This was expected for gas-to-gas and diesel-to-diesel influences. However, even the stock, refinery, and reference benchmark prices tend to

group together at least in pairs. Furthermore, refineries are almost exclusively related to the same type of product, gas producing refineries are connected to retail gas prices, and diesel producing refineries are connected to diesel retail prices. The Gulf Coast GAS refinery node does not exactly meet the previous observation as it is shown influencing some diesel products, even so, this might be a positive thing as it will alert an attentive analyst to the weight behind the Gulf Coast refinery gas prices. After further analysis, Gulf Coast GAS is identified as the most influential node as it has at least one detected event for all other products, and its average edge weight is the highest by a margin of 5%. Probably due to huge oil production in this area, it is mostly the start of oil production chain. In the study of Tiple (2014), it was also observed that specificities of gas usage in the Rocky Mountain retail gas price could trigger third level dependences. Also, the most glaring aspect of the graph is how influential specific products are. The tree is not just an assorted web of relations but groups of products aggregating around very influential/influenced products. There are some expected trend setters like the OPEC Basket that is used as a benchmark for oil price, the Gulf Coast refineries, and then some unexpected like the Minnesota retail gas price. For other tested data and parameters, equal graphs were observed. Indeed, color coding also showed very similar results with strong groupings of colors and some few selected products influencing groups of others.

## 5 | EUROPEAN FINANCIAL INSTITUTIONS

Financial institutions are heavily dependent upon each other, often having capital invested or borrowing money between themselves. This leads to an interesting case study because Ramex-Forum might shed new light on how these institutions relate to each other based on their market behavior.

This time there is no clear production chain like in the petroleum case study, but new interesting sequences of relations are expected.

### 5.1 | Dataset

For this case study, the company stock price and the 5-year credit default swap (CDS) were used for a 10-year period (2005–2015). The presented Ramex-Forum poly-tree was generated with up-to-date data on November 6, 2015. This case study focuses on all west European banking institutions that are marked as active in GoBusiness Bloomberg Finance database on October 7, 2015 (the date where the process was started).

A total of 141 products were collected, and of those, 29 had CDS related to their available products. These products are distributed around 19 different countries with the distribution being very heavily tilted to Great Britain. The top five countries by number of products (stock price or CDS) in the dataset are the following:

1. Great Britain (34)

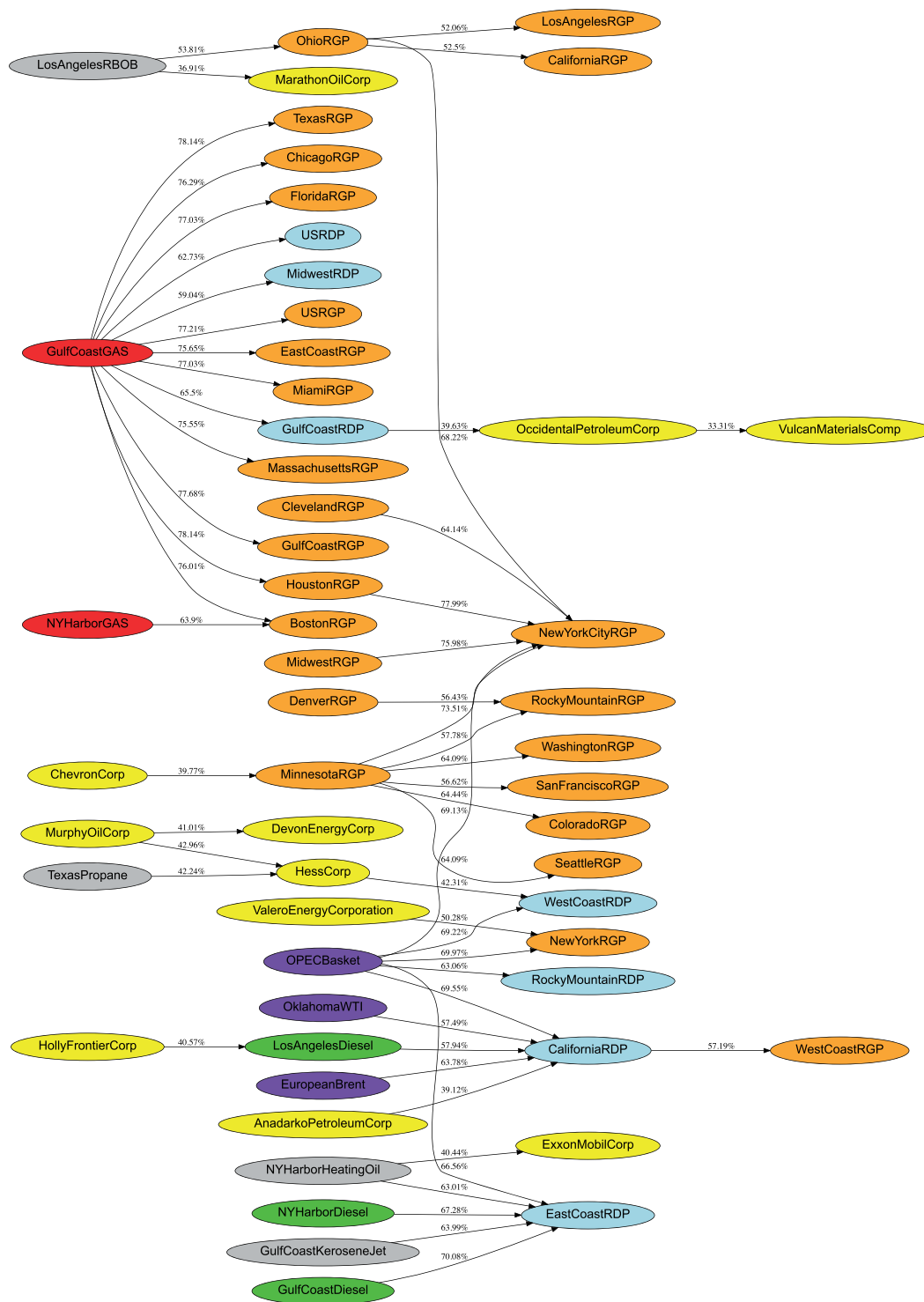2. Italy (19)

3. Germany (18)

4. Spain (17)

5. France (13)

**FIGURE 6**  Graph showing the resulting Buy tree after applying Ramex-Forum on the data with the selected parameters

Remaining countries with counts per country are the following: Sweden (12), Switzerland (10), Netherlands (8), Austria (7), Belgium (6), Greece, Ireland, and Portugal (5 products each), Russia (4), Norway (3), Turkey (2), and Czech Republic and Denmark (1 product each).

The Global Industry Classification Standard classification was used to identify the industry group of each of the banking institutions collected. They fall in three groups: Banks (84), Diversified Financials (41), and Insurance (5).

This classification will be useful to further understand why some relations between products happen.

## 5.2 | Credit default swap

An additional type of financial product is used in this case study, the CDS. By not only looking at the stock price of a company, new layers of complexity can also be introduced into the dataset resulting in richer results.

*"A credit default swap is a particular type of swap designed to transfer the credit exposure of fixed income products between two or more parties. In a credit default swap, the buyer of the swap makes payments to the swap's seller up until the maturity date of a contract. In return, the seller agrees that, in the event that the debt issuer defaults or experiences another credit event, the seller will pay the buyer the security's premium as well all interest payments that would have been paid between that time and the security's maturity date."*[2]

Because it behaves like an insurance, the more risk associated to a certain product, the higher the insurance price will be, or vice versa for lower risk. The perceived risk of a certain product can be assessed by reports, news, or any other indicator the CDS issuer deems relevant; however, because stock prices react to the same indicators, it is expected that the price fluctuations of a product's CDS, although inverse to stock prices, still express dependencies among financial products.

These credit default swaps will behave inversely to the market, putting them in their own isolated bubble of behavior. In this way, no direct relation should be detected with stock market *Buy/Sell* events.

## 5.3 | Correlation between products

While the precision measure gives an idea on how two products behaved sequentially in the observed period, just from that, it is hard to know how trustworthy that observation is. For instance, in low frequency events, a question arises: Does the observed behavior result from an actual relationship or was it just a random chance that produced those results? In this case, correlation coefficients are used as an additional measure to complement Ramex-Forum extracted relations. A correlation coefficient gives an indication of how similar the observed and predicted binary classifications are. They usually assign a value between [−1; +1], where −1 means totally uncorrelated, 0 is the same as a random prediction, and 1 means total correlation.

The most commonly used correlation coefficient is Pearson's; however, due to the characteristics of the Ramex-Forum algorithm and its output, the Matthews Correlation Coefficient (a rescaled form of Pearson's coefficient) was chosen. It is simply a different approach of calculating the Pearson's coefficient, by looking at the confusion matrix of all considered product pairs and using that to calculate de correlation coefficient. The Mathews Correlation Coefficient (MCC) is given by the equation:

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

This equation takes the following into account:

- True positives (TP) – when the positive prediction matches the outcome.
- True negatives (TN) – when the negative prediction matches the outcome.
- False positives (FP) – when the positive prediction does not match the outcome.
- False negatives (FN) – when the negative prediction does not match the outcome.

A positive prediction happens when for a given relation among two products, the influencing product crosses its MA, and the influenced product also crosses its MA in the same direction (both products have either a buy or sell signal simultaneously).

A negative prediction happens when both the influencing and influenced products are not producing signals of the same type or at the same time.

With this new measure a filter for finding macropatterns can be applied to prune relations that are not correlated enough, producing more meaningful results with less noise. One way to classify the scale of available correlation values is as follows:

- 0 – No relationship
- ±0.30 – A weak relationship
- ±0.50 – A moderate relationship
- ±0.70 – A strong relationship
- ±1 – A perfect relationship

In general, we will be looking into relations with at least a moderate positive relationship; as such, the graph was pruned of edges with a value of *MCC* < 0.5. Still, due to the higher complexity of the poly-tree and for more concise and illustrative results in this paper, we choose to focus more only on very relevant connections, and so, the minimum threshold was slightly raised to 0.6. This value was perceived as a good compromise for a qualitative discussion because it was simultaneously filtering less strong connections, without removing interesting macropatterns that result from the second phase Ramex-Forum. When the final graph is produced for analysis, the MCC values will also be displayed next to each edge so that the reader is aware of the strength of the correlation in that relation.

In order to generate a single connected poly-tree, the inclusion of weak relations was a direct consequence of only having the signal-counters as the criteria for relating products. The use of MCC filtering also has the advantage of allowing additional objective criteria, so the algorithm can now relax this restriction. The consequences of removing this restriction for low MCC values are also analysed on this study.

## 5.4 | Results

Ramex-Forum was set with the parameters established the previous section (i.e., $\delta = 60$, $\varepsilon = 0.1$, $N_{MA} = 120$). Algorithm performance was almost instantaneous with present dataset. However, in order to

---

[2]http://www.investopedia.com/terms/c/creditdefaultswap.asp, retrieved November 13, 2015.
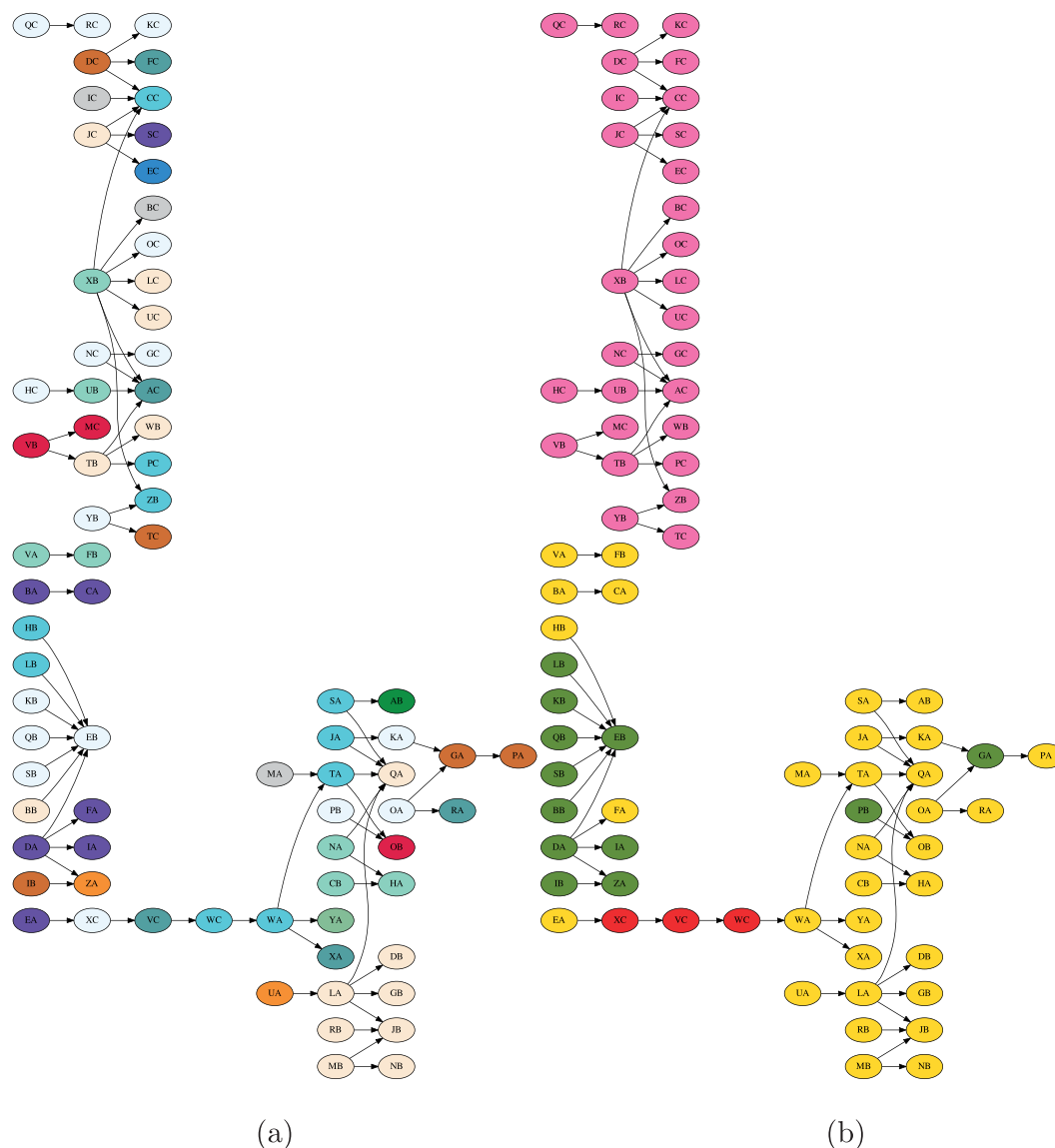
**FIGURE 7** Graphs showing a generic view of main dependences on major European financial institutions colored by country (graph a, on the left) and by sector (graph b, on the right). Detailed output is presented in appendix

better evaluate the "real world" performance of the Ramex-Forum algorithm, two additional datasets of different dimensions for a 5 year period were then used for additional performance measures. The tests were done using a MS-Windows PC Intel *i5–6600* at 3.5 GHz with 8Gb memory. The time it took for each run of the algorithm and the memory used by the JAVA virtual machine (VM) was recorded. According to the measured results, for 100 products, Ramex-Forum took 45 ms in phase I and 15 ms in phase II. The JAVA VM used 54 and 37 Mb (respectively for phase I and phase II). A more demanding test used 1000 products. In this second test, 235 Mb was used for phase I, and 255 Mb of memory was used for phase II. Overall running time was 7.9 s.

Figure 7 shows a generic view of the poly-tree for financial products when *MCC* > 0.6. Appendix A presents a table with codes and available data, not used by the algorithm but used to assign colors in image. Appendix B shows a detailed per-node view of this poly-tree.

Figure 7 (b) is particularly interesting. Indeed, just by looking at price variations, the Ramex-Forum algorithm was able to identify main

price types that are being considered. It is particularly interesting to note the *island* composed by all CDS prices. Also, on a macropattern-based analysis, it is very interesting to notice the *SchrodersPLC* (ED code) node. Schroders is one of Europe's largest asset management companies, and it is based in Great Britain. Based on Ramex-Forum result, a simple Google search has shown a recent Italian based Exor SPA (code BB) press release [3] that presents that Exor is now owner of part of *The Economist Group* jointly with Schroders PLC.[4] It is out of the scope of this article to make a deep analysis of present relations between these companies, and previous relations may be due to pure chance, so let us analyse two more *trivial* relations that are easy to see and understand in Ramex-Forum poly-tree. *Banco Santander SA CDS*

---

[3]https://www.exor.com/home/Media/01_Comunicati-stampa/2015/38_CS_12082015.html, retrieved 13th November, 2015.

[4]*The "A" special shares are held by individual shareholders including the Cadbury, Rothschild, Schroder and other family interests as well as a number of staff and former staff*, according to http://www.economistgroup.com/results_and_governance/ownership.html, retrieved November 13, 2015.

(code *XB*), based in Spain, is one of the major European banks. So it is expectable that other financial institutions in Europe do depend on its behavior, namely its UK-based company *Santander UK PLC CDS* (code *OC*). A little less expected is its high relation (and correlation) of the price of their CDS with other major financial banks, namely *BNP Paribas CDS* (code *ZB*), *Credit Agricole SA CDS* (code *CC*), and even *Commerzbank* (code *AC*)

Following a more systematic approach, we also check the Ramex poly-tree's longest path that starts in *Swedbank AB* (code *EA*). It is interesting to note its path to the Prudential PLC insurance company (code *XC*), that by itself is in relation with *Allianz* (code *VC*) and then, indirectly, to *AXA* (code *WC*), and still indirectly to *Natixis* French Bank (code *WA*), which in turn influences directly the stock price of *Commerzbank* (code *TA*), *Bank of Ireland* (code *AY*), and *Credit Agricole* (code *XA*). Again, it is outside the scope of this paper to make an in-depth analysis of the European financial sector, and we leave that in-depth analysis to further studies.

Finally, we should stress three important aspects on this poly-tree: the already discussed per-country and per-sector strong correlation (please recall that this information is not provided in any way to the algorithm), and the fact that relations among institutions on a long path in Ramex are probably noncausal. Indeed, from our longest path in Figure 7, if *EA* is connected to *XC* with some probability of 0.33, *XC* is connected to *VC*, and then *VC* with *WC* both with probability 0.5, so the indirect relation from *EA* to *WC* would be of just $0.3 \times 0.5 \times 0.5 = 0.07$. Nevertheless, a further microanalysis of these type of proposals may help to highlight hidden dependencies in data.

## 6 | CONCLUSIONS

The presented work focus on two case studies, using real world data and a deep analysis. Our study aims to provide illustrative and useful examples of Ramex-Forum. The signal-to-noise ratio on the petroleum production chain analysis shows that sequential patterns of prices can provide a much deeper description of product dependencies based on events. The per-sector and per-country clear dependence on financial products show how useful Ramex-Forum can be for helping to gather financial intelligence. Moreover, the $\delta$ and $\varepsilon$ parameters seem both consistent, intuitive, and an adaptable alternative for measuring long-term dependencies that are not directly possible with more instantaneous methods.

The introduction of MCC empirically proved itself to be quite relevant on studied data, because it provided an easy way for the analyst to focus first on more strong and systematic relations in poly-trees. Also, by relaxing the restriction that limits the solution to a single poly-tree, the independence of certain node groups is immediately visible. This makes the final result both easier to read and more informative.

Regarding current performance of the algorithm, with studied datasets, the output is almost instantaneous, so, results are compatible with present project needs. Despite this, even after tests of simultaneous analysis with 1000 products (a value that is clearly beyond the scope of our future project needs), we have noticed that the main overhead was due to the JAVA VM garbage collector as used in the current implementation (almost 500 Mb were used for JAVA VM in both phases of Ramex-Forum). Nevertheless, even so, the results were achieved in less than 8 s, a value that is fast enough for analysing the available data-stream.

From our study, the main difficulty on using Ramex-Forum was due to the high sensibility of the algorithm to slight changes in data. Indeed, relations that appear on one run, might disappear on the next run, with only a small variation in parametrization. This happens because the second step of the algorithm looks for the heaviest edges. So, a change in parametrization that changes edges by just 0.01% may cause one edge to be heavier than the other, resulting in that an almost equally strong (but slightly weakest) edge will be replaced. Moreover, that would have repercussions on the whole poly-tree because the new relation could mean that another previously existent relation will not be taken into account because the poly-tree cannot have cycles. In future work, it might be interesting to extend the poly-tree definition by allowing the locking of relations. For example, by providing a starting tree and adding to that relevant and already analyzed relations. This way the user can safely change parameters and see how the result changes around a specific set of interesting nodes.

Future work in this area should now follow two complementary paths. On one hand, even more stability and aggregative power must be given to the Ramex-Forum user. On the other hand, a tool capable of using these results for further information retrieval using different information sources, such as financial news, should be pursued. Regarding aggregative power, some relevant nodes in a generalized poly-tree should be able to express and aggregate different financial and economic variables that need to be taken into account. So far, only the connections themselves have been considered. If the influence weights are also taken into account, the analysis becomes more complex. Also, relations among hyper-graphs are being studied. In the current version, only pairs of nodes are being analyzed, creating graphs. In future work, the authors plan to use more than two node comparisons, creating hyper-graphs. Moreover, some related studies (Matos et al., 2014) show that less classical hybrid approaches can also be used to complement the crossover event detection approach and are good candidates for future experiments. In that respect, previous work in expert systems will be highly useful for providing the user with distinct visualizations and multi-criteria comparisons among products, for example, helping to build correlation based on financial simulators. A simple simulator based only on Ramex correlations was already described in the study of Tiple (2014).

Regarding information retrieval, graph analysis can select relations among companies on a given time frame. This way, a news browser can be built, namely for consulting financial and economic news relevant in the context of events aggregated by poly-tree edges. It would be of particular interest to include a better characterization of the algorithm behavior during a global market crisis, namely by helping to quantify the consequences and the causes behind the recent crisis in oil prices or the still ongoing economic financial crisis.

academics assure the independence of achieved results. Namely, all mentioned financial companies were selected due to the correlations of their observed data-patterns, directly depicted from data extracted for this study from Blomberg available data. The authors had no previous relation with any of the mentioned companies and were not aware of any causal relation among those companies.

## REFERENCES

Borenstein, S., Shepard, A., and of Economic Research, N. B. (1996). Sticky prices, inventories, and market power in wholesale gasoline markets. Number no. 5468 in NBER working paper series. National Bureau of Economic Research.

Cavique, L. (2007). A network algorithm to discover sequential patterns. In *In progress in artificial intelligence*, EPIA07, LNAI 4874, (pp. 406–414). Berlin: Heidelberg. Springer-Verlag.

Cavique, L. (2015). Ramex: a sequence mining algorithm using poly-trees. In *In Proceedings of the pervasive information systems chapter*. World CIST, Berlin: Heidelberg. Springer-Verlag.

Cavique, L., & Coelho, J. (2008). Descoberta de padrões sequenciais utilizando árvores orientadas. *Revista de Ciências da Computação*, 3(3).

Chan, E. (2009). *Quantitative trading: How to build your own algorithmic trading business*. Wiley Trading: Wiley.

Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2009). *Introduction to algorithms, third edition* (3rd ed.). Cambridge, Massachusetts: The MIT Press.

Edmonds, J. (1967). Optimum branchings. *Journal of Research of the National Bureau of Standards*, 71B, 233–240.

Fong, W. M., & Yong, L. H. M. (2005). Chasing trends: recursive moving average trading rules and internet stocks. *Journal of Empirical Finance*, 12(1), 43–76.

Fulkerson, D. (1974). Packing rooted directed cuts in a weighted directed graph. *Mathematical Programming*, 6(1), 1–13.

Gary, J., & Handwerk, G. (2001). *Petroleum refining* Taylor & Francis: Institut francais du petrole publications.

Gunasekarage, A., & Power, D. M. (2001). The profitability of moving average trading rules in south asian stock markets. *Emerging Markets Review*, 2(1), 17–33.

Hammoudeh, S., Ewing, B. T., & Thompson, M. A. (2008). Threshold cointegration analysis of crude oil benchmarks. *The Energy Journal*, 29(4), 79–96.

Makridakis, S., & Wheelwright, S. (1978). *Forecasting: methods and applications*. Wiley/Hamilton series in management and administration: Wiley.

Marques, N. C., & Cavique, L. (2013). Sequential pattern mining of price interactions. In Advances in Artificial Intelligence - Proceedings of the Workshop Knowledge Discovery and Business Intelligence, EPIA-KDBI, Portuguese Conference on Artificial Intelligence, pp.314–325.

Matos, D., Marques, N., & Cardoso, M. (2014). Stock market series analysis using self-organizing maps. *Revista de Ciências da Computação*, 9(9), 79–90.

Rebane, G., & Pearl, J. (1987). The recovery of causal polytrees from statistical data. In *Third Conference on Uncertainty in Artificial Intelligence*, pp.222–228.

Suviolahti, H. (2009). The influence of volatile raw material prices on inventory valuation and product costing. Master Thesis, Department of Business Technology, Helsinki School of Economics.

Tiple, P. (2014). Tool for discovering sequential patterns in financial markets. Master Thesis in Engenharia Informática, Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa.

Tiple, P., Cavique, L., & Marques, N. (2015). Ramex-forum: Sequential patterns of prices in the petroleum production chain. In F. Pereira, P. Machado, E. Costa, & A. Cardoso (Eds.), *Progress in artificial intelligence*, volume 9273 of *lecture notes in computer science*. (pp. 584–589). Springer International Publishing.

## AUTHOR BIOGRAPHIES

Pedro Tiple is a collaborator in the Research Department of GoBusiness Finance and has been participating in projects related to computational finance. He is working mostly with data mining and visualization. He received the degree in Computer Science Engineering from the New University of Lisbon (FCT-UNL) in 2011, the MSc degree in Computer Science Engineering from the New University of Lisbon (FCT-UNL) in 2014 and has been working with GoBusiness Finance as a software engineer and research collaborator since 2013.

Luis Cavique is a Professor of the Computer Science Section (SIFT) in the Department of Sciences and Technology (DCeT) at Universidade Aberta and Research member in LabMAg. He worked in the Polytechnic Education System from 1991 to 2008, namely as Adjunct Professor in the Setubal and in the Lisbon Polytechnic Institute. He received the degree in Computer Science Engineering from the New University of Lisbon (FCT-UNL) in 1988, the MSc degree in Operational Research and Systems Engineering from the Technical Lisbon University (IST-UTL) in 1994, and the PhD degree in Systems Engineering from the Technical Lisbon University (IST-UTL) in 2002. His research areas are in the intersection of Computer Science and Systems Engineering, namely the Heuristic Optimization and the Data Mining. He is author of numerous national and international publications subject to peer-review. He participated in several commissions of national and international conference programs, he is a reviewer in some journals and he is a chief-editor of the Revista de Ciências da Computação da Universidade Aberta.

Nuno Marques is a tenured Assistant Professor at the Computer Science Department of Faculdade de Ciências e Tecnologia from Universidade Nova de Lisboa (FCT/UNL). He is working as a researcher since 1991 and as a lecturer since 1997. His PhD degree was achieved in January 2000 and focused efficient models for part of speech tagging and verbal subcategorization acquisition from large text corpora. Nuno Marques' PhD work has pioneered Text Mining for the Portuguese language. Current work conjoins Machine Learning in Artificial Neural Networks and its applications to real world problems, namely in Text and Data Mining. Since 2009, he is working on applying Data Mining to Financial Data Streams, namely by applying text and data mining in this area. His current research interests include artificial neural networks and deep learning used for knowledge extraction from text and data. Since 2009, his research projects involve both academia and industry.

# APPENDIX A. List of EuroFin Institutions

Bloomberg data, ordered by type of price, code in Figure 7, country code (ISO 3166–1 two alpha) and Bloomberg name. In column type CDS denotes a credit default swap type price and Bank, Insurance or Diversified Finance (Div) denotes the financial sector for stock prices.

**TABLE A1**

| Type | Code | Country | Name |
| --- | --- | --- | --- |
| Bank | AB | AT | Erste Group Bank AG |
| Bank | BA | SE | Nordea Bank AB |
| Bank | CA | SE | Svenska Handelsbanken-A SHS |
| Bank | CB | SP | Banco de Sabadell SA |
| Bank | DB | IT | Mediobanca SPA |
| Bank | EA | SE | Swedbank AB - A Shares |
| Bank | FA | SE | Skandinaviska Enskilda BAN-A |
| Bank | FB | SP | Banco Popular Espanol |
| Bank | GB | IT | UBI Banca SCPA |
| Bank | HA | SP | Banco Santander SA |
| Bank | HB | FR | Wendel |
| Bank | JA | FR | BNP Paribas |
| Bank | JB | IT | Banco Popolare SC |
| Bank | KA | GB | Lloyds Banking Group PLC |
| Bank | LA | IT | Intesa Sanpaolo |
| Bank | MA | NL | ING Groep NV-CVA |
| Bank | MB | IT | Banca Popolare di Milano |
| Bank | NA | SP | Banco Bilbao Vizcaya Argenta |
| Bank | NB | IT | Banca Popol Emilia Romagna |
| Bank | OA | GB | Barclays PLC |
| Bank | OB | PT | Banco Comercial Portugues-R |
| Bank | PA | CH | Credit Suisse Group AG-REG |
| Bank | QA | IT | UNICREDIT SPA |
| Bank | RA | DE | Deutsche BanK AG-Registered |
| Bank | RB | IT | Banca Popolare di Sondrio |
| Bank | SA | FR | Societe Generale SA |
| Bank | TA | FR | Credit Agricole SA |
| Bank | UA | BE | KBC Groep NV |
| Bank | VA | SP | Caixabank S.A |
| Bank | WA | FR | Natixis |
| Bank | XA | DE | Commerzbank AG |
| Bank | YA | IE | Bank of Ireland |
| Insurance | VC | DE | Allianz SE-REG |
| Insurance | WC | FR | AXA SA |
| Insurance | XC | GB | Prudential PLC |

| Type | Country | Code | Name |
| --- | --- | --- | --- |
| CDS | AC | DE | Commerzbank AG CDS |
| CDS | BC | NL | Cooperatieve Centrale Raiffe CDS |
| CDS | CC | FR | Credit Agricole SA CDS |
| CDS | DC | CH | Credit Suisse Group AG CDS |
| CDS | EC | DK | Danske Bank A.S |
| CDS | FC | DE | Deutsche Bank AG CDS |
| CDS | GC | GB | HBOS PLC CDS |

**TABLE A1** (Continued)

| Type | Country | Code | Name |
| --- | --- | --- | --- |
| CDS | HC | GB | HSBC Bank PLC CDS |
| CDS | IC | NL | ING Bank NV CDS |
| CDS | JC | IT | Intesa Sanpaolo SpA CDS |
| CDS | KC | GB | Lloyds Bank PLC CDS |
| CDS | LC | IT | Mediobanca SpA CDS |
| CDS | MC | PT | Novo Banco SA CDS |
| CDS | NC | GB | Royal Bank of Scotland PLC CDS EUR |
| CDS | OC | GB | Santander UK PLC CDS EUR |
| CDS | PC | FR | Societe Generale SA CDS EUR |
| CDS | QC | GB | Standard Chartered BANK CDS EUR |
| CDS | RC | GB | Standard Chartered PLC CDS EUR |
| CDS | SC | SE | Svenska Handelsbanken AB CDS EUR |
| CDS | TB | IT | Banca Monte dei Paschi di Si CDS |
| CDS | TC | CH | UBS CDS EUR |
| CDS | UB | SP | Banco Bilbao Vizcaya Argenta CDS |
| CDS | UC | IT | UniCredit CDS EUR |
| CDS | VB | PT | Banco Comercial Portugues SA CDS |
| CDS | WB | IT | Banco Popolare SC CDS |
| CDS | XB | SP | Banco Santander SA CDS |
| CDS | YB | GB | Barclays Bank PLC CDS |
| CDS | ZB | FR | BNP Paribas SA CDS |
| Div | BB | IT | Exor SPA |
| Div | DA | SE | Investor AB-B SHS |
| Div | EB | GB | Schroders PLC |
| Div | GA | CH | UBS Group AG-REG |
| Div | IA | SE | Industrivarden AB-A SHS |
| Div | IB | CH | Pargesa Holding SA-BR |
| Div | KB | GB | 3I Group PLC |
| Div | LB | FR | Eurazeo |
| Div | PB | GB | Henderson Group PLC |
| Div | QB | GB | Close Brothers Group PLC |
| Div | SB | GB | Intermediate Capital Group |
| Div | ZA | BE | Groupe Bruxelles Lambert SA |

(*Continues*)

## APPENDIX B. Full Ramex-Forum Poly-Tree for EuroFin Institutions

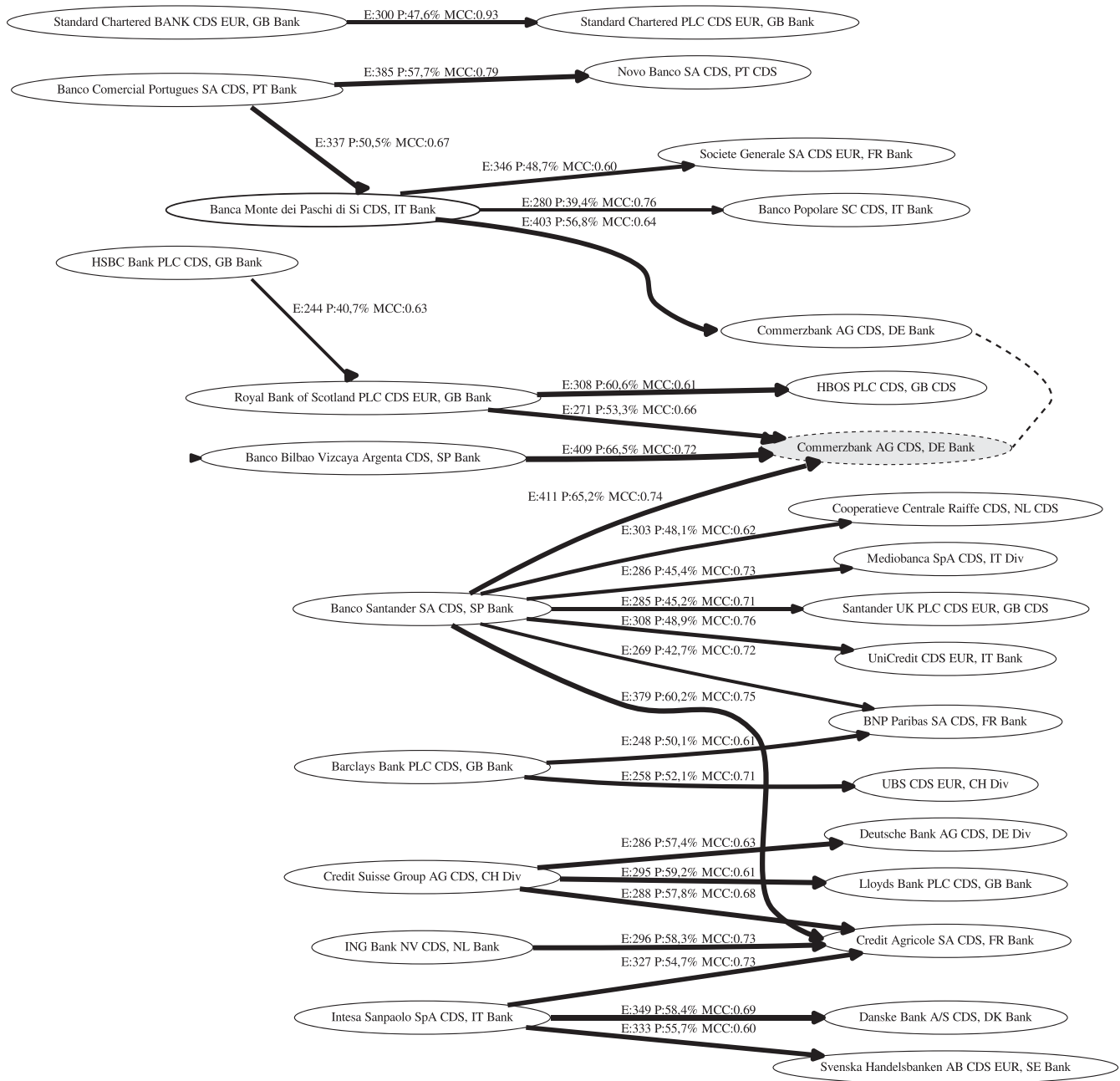Complete EuroFin Poly-Tree, divided by distinct areas. Gray nodes are duplicated for easier presentation.



**FIGURE B1** Graphs showing a generic view of main dependences on major European financial institutions colored by country (graph a, on the left) and by sector (graph b, on the right). Detailed output is presented in appendix
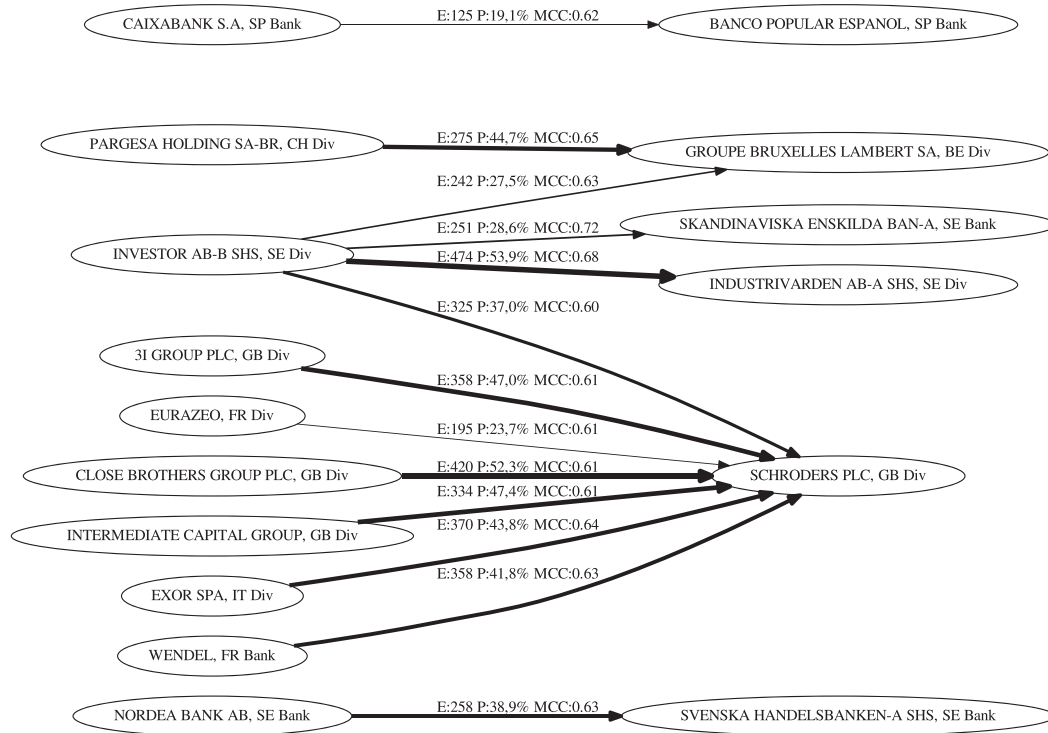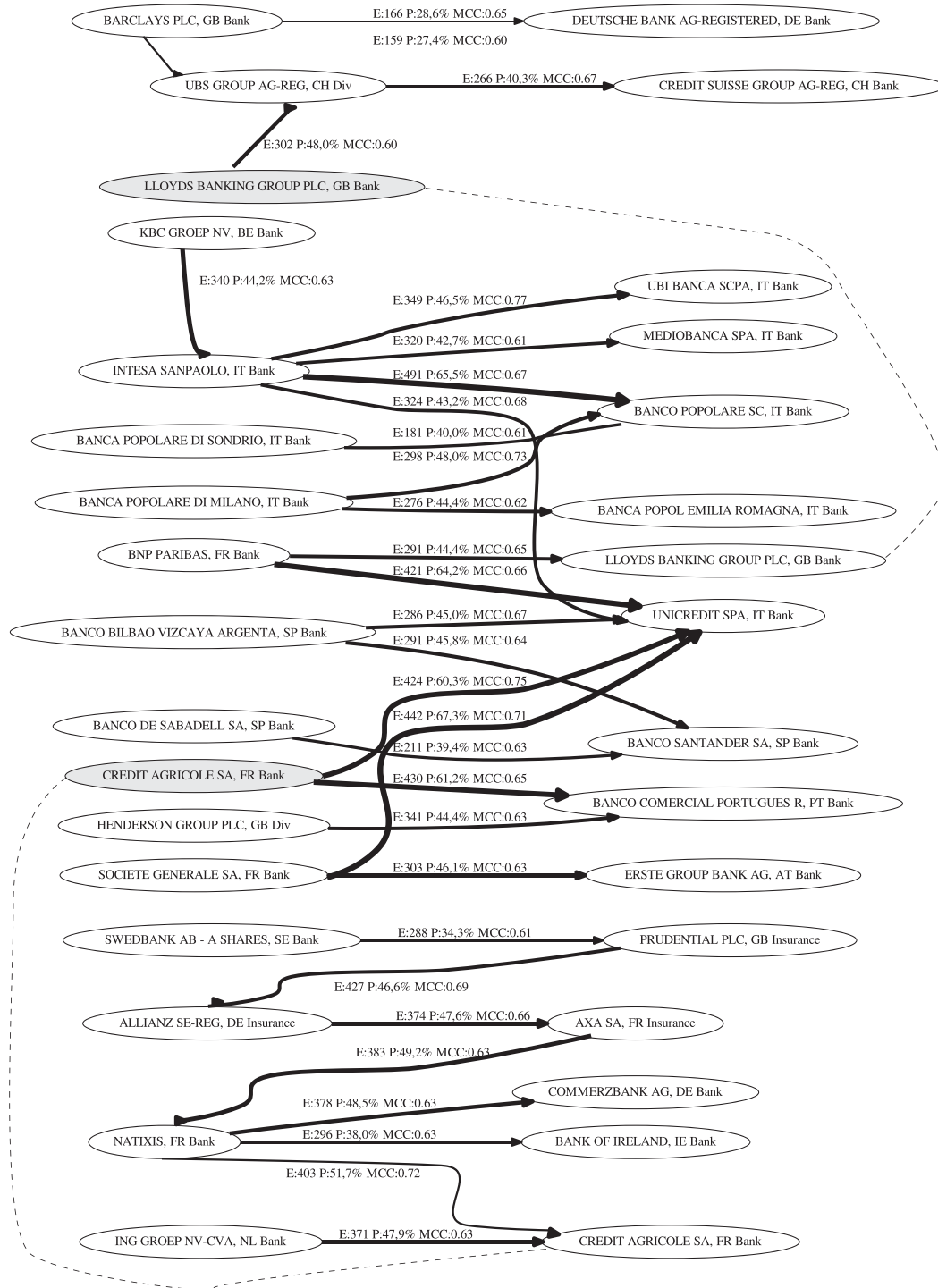
**FIGURE B1** Continued

**FIGURE B1** Continued