

Trend Mining in Social Networks: A Study Using A Large Cattle Movement Database

Puteri N.E. Nohuddin (1), Rob Christley (2), Frans Coenen (1), Christian Setzkorn (2)

(1) Department of Computer Science, University of Liverpool, UK.

(2) School of Veterinary Science, University of Liverpool and National Centre for Zoonosis Research, Leahurst, Neston, UK

Email: {Puteri.Nohuddin, coenen, robc, c.setzkorn}@liverpool.ac.uk

Abstract. This paper reports on a mechanism to identify temporal spatial trends in social networks. The trends of interest are defined in terms of the occurrence frequency of time stamped patterns across social network data. The paper proposes a technique for identifying such trends founded on the Frequent Pattern Mining paradigm. The challenge of this technique is that, given appropriate conditions, many trends may be produced; and consequently the analysis of the end result is inhibited. To assist in the analysis, a Self Organising Map (SOM) based approach, to visualizing the outcomes, is proposed. The focus for the work is the social network represented by the UK's cattle movement data base. However, the proposed solution is equally applicable to other large social networks.

Keywords: Social Network Analysis, Trend Mining, Trend Visualization.

1 Introduction

A *Social Network* is an interconnected structure that describes communication (of some form) between individuals. Social Network Mining is concerned with the identification of patterns within such networks. Typical applications include: the identification of disease spreading patterns from dynamic human movement [2], monitoring users' topics and roles in email distributions [15], and filtering product ratings from online customer networks for viral marketing strategies [7]. Social network mining approaches tend to be founded on graph mining or network analysis techniques. Typically, social network mining is undertaken in a static context, a "snapshot" is taken of the network which is then analysed. Little work has been done on the dynamic aspects of social network mining. Further, most current work does not take into consideration the relationship between network nodes and their associated geographical location. The work described in this paper seeks to address these two issues.

In the context of this paper, social networks are viewed in terms of a sequence of snapshots taken of the network at discrete time intervals. The patterns of interest are identified using Frequent Pattern Mining (FPM) techniques, which relate groups of data items that frequently appear together. An adaptation of the Total From Partial (TFP) FPM algorithm [5] is used, Trend Mining TFP (TM-TFP), to discover trends in sequences of time stamped social networks. The trends are described in terms of sets of frequency measures associated with specific frequent patterns that occur across the

network. These sequences are conceptualised as time series or “trend lines”. However, using the TM-TFP approach, given appropriate input conditions, a great many such trends may be identified. The trends of interest are very much application dependent. We may be interested in increasing or decreasing trends, or trends that represent seasonal changes. In addition, we may wish to identify “flat” trend lines, or sudden changes in trends. This paper therefore also suggests a visualisation technique, founded on a Self Organising Map (SOM) approach, to cluster similar trends. This allows end users to focus on the application dependent trend lines of interest.

The focus of the work described in this paper is the UK’s cattle movement database. This is a UK government funded initiative, managed by The Department for Environment, Food and Rural Affairs (Defra), introduced in 1998 in response to a cattle disease epidemic. The database records the movement of all cattle between pairs of locations in Great Britain. As such, these locations can be viewed as network nodes, and the movement of cattle as weighted links between node pairs. All data entries are time stamped, and thus snapshots of the network can be obtained. Sequences of patterns, trends, can thus be identified within the network. Similarly, each location (node) is referenced geographically and thus a spatial element can be added into the analysis. The activity evidenced in the databases was grouped at monthly intervals to obtain a sequence of “snapshots”. For any given month the number of network nodes was in the region of 56,300, and the number of links in the region of 73,600, in other words the size of each time stamped network was substantial.

The overall contribution of this paper may thus be summarized as follows. Firstly, a mechanism for identifying spatial-temporal trends in large social networks is described. Secondly, a technique is presented to support the analysis/visualisation of the outcomes by clustering similar trends. Thirdly, a “real-life” application of the techniques is presented and evaluated.

The rest of this paper is organized as follows. Section 2 provides a brief background of FPM, social network mining and the SOM technology used. Section 3 describes the cattle movement database (social network) application. An overview of the proposed spatio-temporal social network trend mining framework is then given in Section 4. A full evaluation of the framework is reported in Section 5; and, finally, Section 6 provides some conclusions.

2 Background

This section provides some necessary background to the work described in the rest of this paper. It commences with a brief review of FPM, then continues with a consideration of current work in trend mining, social network mining and SOM technique.

FPM was first popularized in the context of Association Rule Mining (ARM). The catalyst ARM algorithm is generally acknowledged to be the Apriori algorithm [1]. Many alternative ARM and FPM algorithms have since been proposed. The FPM algorithm used with respect to the work reported in this paper is the TFP (Total From Partial) algorithm [5]. TFP uses a tree structure, the P-tree, in which partial support counts are stored; and a second tree structure, the T-tree (a reverse set enumeration tree

data structure) facilitates fast “look up” ([4]). TFP offers advantages, with respect to many other FPM algorithms, in terms of computational efficiency and storage.

Trend mining is concerned with the identification of patterns that change over time. Trends are typically defined in terms of time series. One example is Google Trends, a public web facility recently introduced by Google to identify trends associated with keyword search volume across various global regions and in various languages [22]. Trend recognition processes can be applied to qualitative and also to quantitative data, such as forecasting financial market trends based on numeric financial data, and usage of text corpora in business news [18]. In the context of this paper, trends are defined in terms of the changing frequency of individual patterns. A similar concept has been used in the context of Jumping Emerging Pattern (JEP) mining. For example in Khan et al. [9], a moving window was used to identify such patterns (patterns whose support changes significantly over time).

There has been a rapid increase in attention, within the data mining community, regarding social network analysis. This is because of the demand to exploit knowledge from the large amount of data that has been collected with respect to the social behavior of users in online environments. A social network depicts the structure of some social entities, and normally comprise actors who are connected through one of more class of links [20]. To analyze this structure, many social network analysis techniques have been proposed which map and measure the relationships and flows between people, organizations, groups, computers or web sites. Social network mining can be applied in a static context, which ignores the temporal aspects of the network; or in a dynamic context, which takes temporal aspects into consideration. In a static context, we typically wish either: (i) to find patterns that exist across the network, or (ii) cluster (group) subsets of the networks, or (iii) build classifiers to categorize nodes and links. In addition, given the dynamic context, we wish to identify trends or change points within social networks. A further point of interest is the geographical relationships represented by nodes. Given the availability of spatio-temporal data, we can determine the relationship between nodes by evaluating the spatio-temporal co-occurrences of events in social networks [14]. Thus, we are able to detect changes or abnormal patterns in communication (movement) behaviour in the network.

Self Organising Maps (SOMs), or Self-Organizing Feature Map (SOFM), were first proposed by Kohonen [11, 10]. Essentially, SOMs are a neural network based technique designed to reduce the number of data dimensions in some input space by projecting it onto a “map”, which plots the similarities of the input data by grouping similar data items together (i.e. clustering the data). The algorithm is based on unsupervised and competitive learning, and typically operates by first initialising a $n \times m$ matrix of nodes where each node is to be associated with a cluster. Currently, there is no scientific method for determining the best value for n , i.e. to identify how many clusters should be represented by a SOM, however the $n \times m$ value does define a maximum number of clusters; in most cases, on completion of the SOM algorithm, some nodes will be empty [6]. Since SOM are based on competitive learning, the output nodes on the map compete among each other to be stimulated to represent the input data. Eventually, some nodes can be empty without any input vectors. The authors implemented a SOM using Matlab toolbox functions. The toolbox provided functions, based on the Kohonen

SOM algorithm, that determine the distance between input data and represented similar input data on the map [19]. A SOM approach was adopted because it could group similar trends, and thus enhance the analysis of the TM-TFP result, without requiring prior input of the number of desired clusters ($n \times m$). It also represented a “tried and tested” approach that had been successfully used in many engineering applications such as patterns recognition and process monitoring [12].

3 The Cattle Movement Data Base

The work described in this paper focuses on the UK Cattle Tracing System (CTS) database. Nevertheless, the proposed technique is equally applicable to other types of data sets with similar properties. The CTS is maintained by the British Cattle Movement Service (BCMS). The CTS database is the core information source for Defra’s RADAR (Rapid Analysis and Detection of Animal-related Risks) database with regard the birth, death and movement of cattle in Great Britain. The required recording is undertaken using a range of mechanisms including post, telephone and a dedicated website [21]. Cattle movements can be one off movements to final destinations, or movements between intermediate locations [16]. In short, the movement types include cattle imports, movements between locations, on movement in terms of births and off movements in terms of death. CTS was introduced in September 1998, and updated in 2001 to support the disease control activities. Currently, the CTS database holds some 155 Gb of data.

The CTS database comprises a number of tables, the most significant of which are the *animal*, *location* and *movement* tables. The animal table gives information about each individual animal, referenced by an ID number, such as the breed of the animal (185 different breeds are recorded) and animal date of birth. The location table gives details about individual locations, again referenced by a unique ID number, such as its grid coordinates (Easting and Northing) and location type. The most common location types are Agricultural Holding, Landless Keeper, Market, Common Land and slaughterhouses. A total of thirteen different categories of location are recognised within the database. The movement table, in the context of the work described here, is the most significant. Each record in the movement table describes the movement of one animal from a *sender* location to a *receiver* location. The table also includes the date of the movement.

For the work described in this paper, information was extracted from these tables into a single data warehouse that comprised sets of time stamped data “episodes”. The temporal granularity used through out was one month. The number of CTS records represented in each data episode was about 400,000, each record representing a cattle movement instance. Each record in the warehouse comprised: (i) a time stamp, (ii) the number of cattle moved, (iii) the breed, (iv) the senders location in terms of Eastings and Northings, (v) the “type” of the sender’s location, (vi) the receivers location in terms of Eastings and Northings, and (vii) the “type” of the receiver’s location. If two different breeds of cattle were moved at the same time from the same sender location to the same receiver location this would generate two records in the warehouse. The maximum number of cattle moved between a pair of locations for a single time stamp could be more than 40 animals.

In summary, the cattle movement warehouse can be interpreted as a social network where nodes described farms or holding areas and links described cattle movements between farms and/or holding areas. The number of animals moved, of a given breed type, between any pair of nodes, was regarded as a link weighting. The spatial magnitude of movement between nodes can be derived from the location grid values.

By applying the proposed spatio-temporal trend mining technique to the CTS social network, trends describing cattle movements, across time and space, can be identified. Within these trends we can also identify sub-trends, i.e. trends contained within trends. As noted above, a trend describes the fluctuations of the frequency of a frequent pattern over time. Frequent patterns often contain sub-patterns (sub-sets) which also have trend lines associated with them. Trend-lines belonging to such patterns are identified as sub-trends of the trends associated with the parent (super-set) patterns. Generally, sub-trends display similar trends to their parent trends, but not necessarily so.

Some previous studies of the CTS database have been conducted. Green and Kao [8], who conducted an analysis of the CTS database, confirmed that the number of movement record decreased as the distance between farm location increase. Another significant example is the study undertaken by Robinson and Christley [17], who identified a number of trends in the database. These trends demonstrated that the UK cattle population was in constant flux, and that both seasonal and long term patterns might be identified to support model predictions and surveillance strategies. However, Robinson and Christley's study was very "high level" and considered temporal trends across the entire CTS database without acknowledging spatial factors. The approach describes in this paper serves to identify trends that exhibit both spatial and temporal aspects.

4 The Social Network Trend Mining Framework

This section provides an overview of the proposed social network trend mining framework. The framework comprises two principal components: (i) the trend mining unit, and (ii) the visualization unit. A block diagram giving a high level view of the framework is presented in Figure 1.

The Trend Mining unit (represented by the elements in the top half of Figure 1) is responsible for the identification of trend lines, one per identified pattern, across the input social network. An established FPM algorithm, TFP [5] (introduced in Section 2) was extended for this purpose. The resulting software, Trend Mining TFP (TM-TFP), took as input a sequence of time stamped binary valued data tables, and a minimum support threshold (S); and produced a sequence of trend lines. It should be noted that if a particular pattern, at a particular time stamp, fell below the support threshold it was deemed to be "not relevant" and an occurrence value of 0 was recorded in the trend line, however the pattern was not "thrown away". TM-TFP utilizes the T-tree and P-tree data structures used in TFP. These were essentially set enumeration tree structures designed to enhance the efficiency of the TFP algorithm. In addition, TM-TFP made use of a further tree data structure, the TM tree, that combined the identified frequent item sets into trend lines.

The visualization unit was developed using the SOM Toolbox in Matlab [19]. The toolbox made use of functions and learning processes based on the Kohonen SOM al-

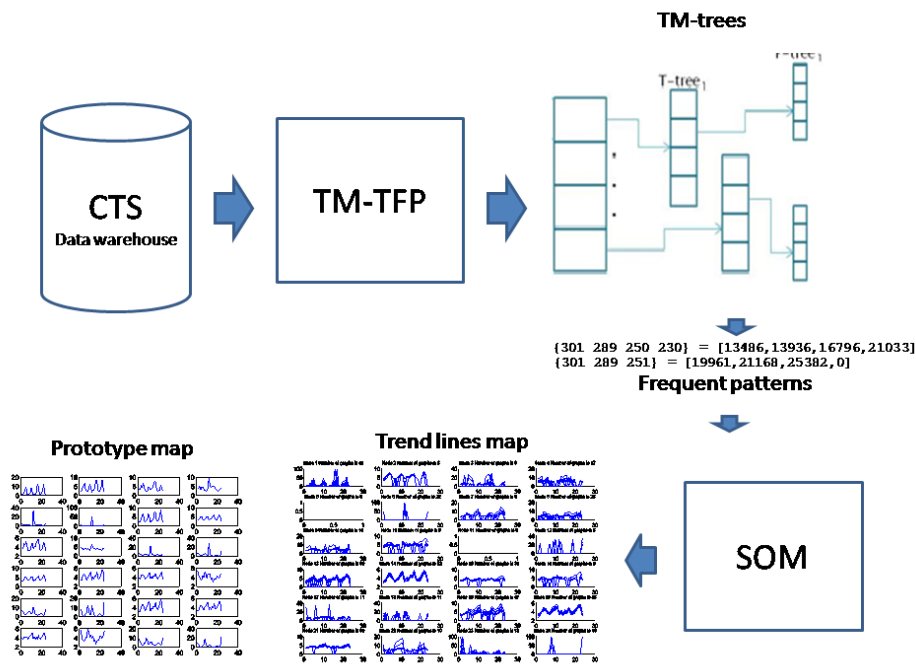


Fig. 1. Block diagram outlining social network trend mining framework

gorithm. The authors extended the software to produce prototype and trend line maps. As will be illustrated in the following section, the prototype map displays the characteristics of each identified trend line cluster, while the trend lines map gives information regarding the number of trends that exist in each cluster.

5 Evaluation

To evaluate the proposed framework, numerous experiments were conducted, the most significant of which are reported in this section. All the reported experiments were directed at the cattle movement database introduced in Section 3. This section is divided into three Sub-sections as follows. Sub-section 5.1 describes the necessary preprocessing conducted on the input data. Sub-section 5.2 reports on the evaluation of the TM-TFP algorithm. Sub-section 5.3 considers the SOM visualisation/clustering mechanism adapted to support the analysis of the trend lines identified using TM-TFP.

5.1 Data Preprocessing, Normalisation and Discretisation

Several subsets of the CTS database were used for the evaluation. These ranged from between six month’s of data to two year’s of data time stamped at monthly intervals. There were twenty four monthly data sets collected from within the time period of January 2005 to December 2006 inclusive.

The FPM algorithm used, TFP [5], in common with most other FPM algorithms, operates using binary valued data. The data had therefore to be normalised and discretized. The Easting and Northing coordinate values were divided into k kilometer sub-ranges, experiments using both $k = 50$ and $k = 100$ were conducted. The effect of this ranging was to sub-divide the geographic area represented by the CTS into a $k \times k$ grids, allowing for the inclusion of trends that express both spatial and temporal relationships. The number of cattle moved value (m) was discretised into five sub ranges: $m \leq 10$, $11 \leq m \leq 20$, $21 \leq m \leq 30$, $31 \leq m \leq 40$ and $m > 40$. The non-linear distribution was used so that a roughly equal proportion of records was included in each sub-range. The end result of the normalisation/discretisation exercise was a table schema comprising 265 attributes where $k = 100$, and 305 attributes where $k = 50$.

5.2 Evaluation of TM-TFP

This Sub-section presents an analysis of the proposed TM-TFP algorithm. Experiments were conducted using a sequence of support thresholds: $S = 2\%$, $S = 5\%$ and $S = 8\%$. Four different *temporal windows* were used, 6 months, 12 months, 18 months and 24 months, corresponding to 6, 12, 18 and 24 data episodes respectively. In this analysis, the Easting and Northing coordinate of datasets were normalised according to 100 km grid squares (i.e. $k = 100$).

TM-TFP identified frequent patterns within each time stamped data episode, with their individual support values. These patterns were then related across the data to identify trend lines. The total number of trend lines identified, using support thresholds of 2%, 5% and 8% are presented in Table 1. For example, for the six month data input, for $S = 2\%$, 1993 trend lines were identified; while with $S = 5\%$, 523 trend lines were identified, and with $S = 8\%$ 222 trend lines. From Table 1, it can be seen that a great many trend lines are discovered. As the temporal window is increased there is a corresponding (although slight) increase in the number of identified trend lines. As would be expected, increasing the support value had the effect of decreasing the number of identified patterns (trends), but at the risk of missing potentially significant trends.

Duration (months)	Support Threshold		
	2	5	8
6	1993	523	222
12	2136	563	242
18	2175	570	248
24	2204	580	257

Table 1. Number of trend lines identified using TM-TFP algorithm when $k = 100$

Duration (months)	Support Threshold		
	2	5	8
6	39.52	29.09	28.11
12	78.58	58.91	55.43
18	114.85	88.12	83.02
24	156.96	118.61	115.48

Table 2. Run time values (seconds) using the TM-TFP algorithm

For completeness, Table 2 presents a sequence of “run time” values for the reported sequence of experiments so as to give an indication of the time complexity of the TM-TFP trend mining algorithm. Increases in the size of the temporal window (number of

data episodes) and the support thresholds gave rise to corresponding linear increases in TM-TFP run time.

5.3 Evaluation of SOM Visualisation

From the above, it can be seen that a great many trends can be identified, the number of trends increases as the temporal window size is increased, and the support value is decreased. This presents a challenge concerning the analysis and interpretation of the identified trend lines. More generally, we can observe that powerful mining algorithms, such as TM-TFP, may overwhelm the end user with too many patterns. Further, the trend lines produced by algorithms such as TM-TFP can occur in many shapes.

The proposed solution is to use a SOM approach to cluster the trends, and consequently ease the interpretation. As noted in Section 2, a SOM consists of several nodes with associated prototypes. A prototype is a vector of the same dimensionality as the original data. The training algorithm adjusts the prototypes, enabling the SOM to spatially order high-dimensional data in much lower dimensional space whilst maintaining complicated underlying relationships and distributions. Time series can be presented to the trained SOM whereby the node with the most similar prototype is declared the winner. Similarity may be determined in a number of ways but the most straight forward is the Euclidean distance measure. Due to the training process, we can expect similar time series to be won by spatially close nodes. Hence, if nodes are arranged as a two dimensional grids, we may expect nodes in the top left corner to win time series of very different shapes in comparison to those in the lower right corner. This is a very useful property of SOMs. Furthermore, large numbers of time series can be reduced to a few meaningful coordinates whilst maintaining their underlying relationships. Domain experts can combine neighbouring nodes to form clusters, reducing the number of patterns even further. This might lead to a handful of clusters containing (say) steady, increasing, decreasing and constantly changing time series.

For the analysis described in this Sub-section, the SOM was seeded with 6×4 nodes, each node representing a cluster. The authors have tested several sizes of SOM to generate clusters of trend lines. However, there is no specific method to determine the most appropriate size of a SOM map, and a 6×4 was found to be the most effective. The bigger the $n \times m$ size of a SOM map the greater the possible number of clusters that may be identified. For example, in medical images clustering, it has been shown that a higher size SOM gives a better image clustering results [3]. For discussion purposes, in this Sub-section, nodes are sequentially numbered by row, reading from top-left to bottom-right. Recall that each trend line represents a sequence of support values. Thus, the authors applied a distance function¹ and a neighbourhood function² to determine similarity. The visualization results are presented in Figures 2, 3 and 4 for a sequence of support thresholds ($S = 2\%$, $S = 5\%$ and $S = 8\%$) using the twenty four month data set, the largest of the four data sets experimented with. Each figure comprises a pair of “maps”, the top map presents the prototypes for the identified clusters and the below map the trends contained in each cluster. As might be expected more clusters

¹ Euclidean function is the distance function used to calculate shortest distance.

² Gaussian function is used to determine the neighbourhood size on the map.

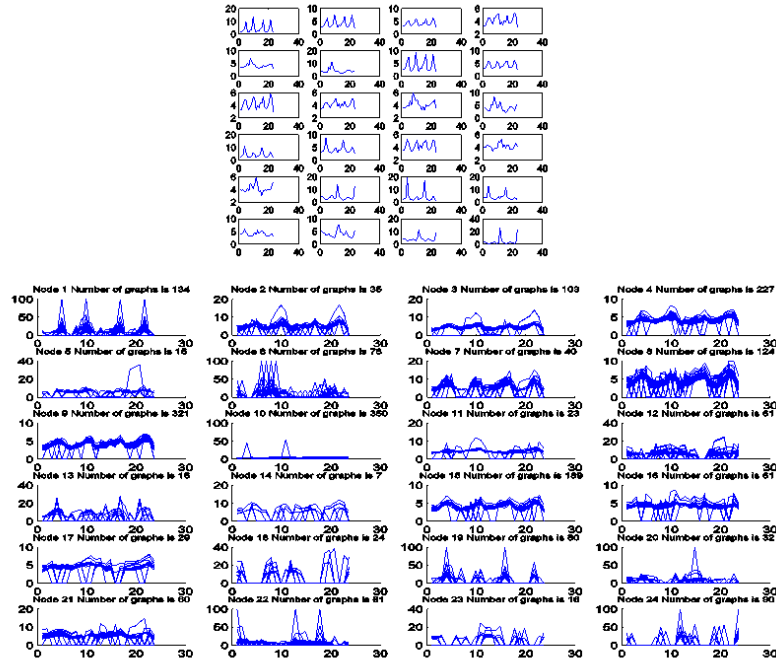


Fig. 2. SOM frequent patterns($S = 2\%$)

were found with $S = 2\%$ than with $S = 5\%$ and $S = 8\%$. Further analysis verified that subsets of the frequent patterns tended to be included in the same cluster as their supersets. Thus, in each cluster, only certain attribute labels/columns were combined; therefore providing for more effective, focused and understandable result interpretation by the end user.

Some specific trend examples (taken from Figures 2, 3 and 4) are: (i) the pattern (Sender holding locations in area B and movement 25-30 animals of Breed British Friesian and receiver holding locations in area D) is in constant flux across the 24 months (area B and D are identifiers for specific grid squares in the geographic area under consideration), and (ii) the pattern (Movement 20-25 animals and Breed Highland and gender female) is static across the time period. The largest number of trends contained in a single cluster was 360 (node 10 in Figure 2). Some nodes, for example numbers 2, 3, 4, 7, 8, 9 and 15 in Figure 2, show “clear-cut” types of trend lines in the respective clusters. Regardless of the “outlier” trend lines shown in each node, the prototype SOM portrays the prototype trends in each node.

The maps presented in Figures 3 and 4 (for $S = 5\%$ and $S = 8\%$) show similar clusters. Further analysis established that this was because use of a lower support threshold resulted mostly in the identification of subsets of those identified using $S = 8\%$ (although for $S = 2\%$ additional patterns were also discovered that were not found using $S = 8\%$).

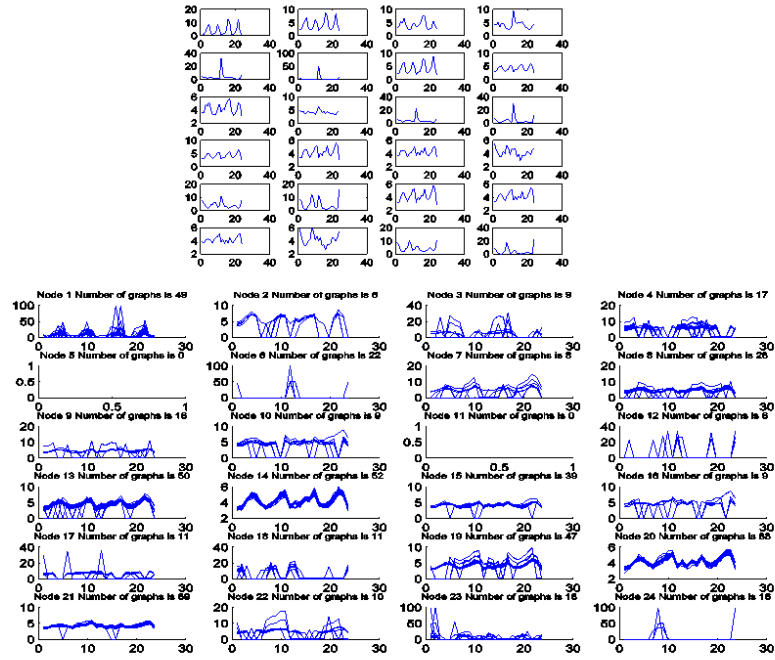


Fig. 3. SOM frequent patterns ($S = 5\%$)

In all cases, the maps identified similar prototypes indicating that if we are only interested in identifying prototypes a higher support threshold, which offers the advantage of greater computational efficiency, is sufficient. By applying the SOM technique, the authors were able to provide a lower dimension of prototypes to represent a substantial number of generated trend lines. In each node, trend lines exhibit the frequent itemsets which include all the possible combination of attributes that carry spatial and temporal features. Thus, the changes or fluctuation can be perceived easily for further analysis.

6 Conclusions

In this paper, the authors have described a framework for identifying and visualising spatio-temporal trends in social networks. The trends were defined in terms of trend lines (in effect time-series) representing the frequency of occurrence of individual patterns with time. Firstly, the trends were identified (the first element of the two part framework) using an extension of the TFP algorithm, TM-TFP. The challenge of TM-TFP was that, given a realistically sized database, a great many trends could be identified making them difficult to analyse. Secondly, to address the analysis issue, the framework used a SOM based clustering mechanism which allowed for similar trends to be grouped together. The advantage offered was that this would allow decision makers, and other end users, to focus on relevant trends.

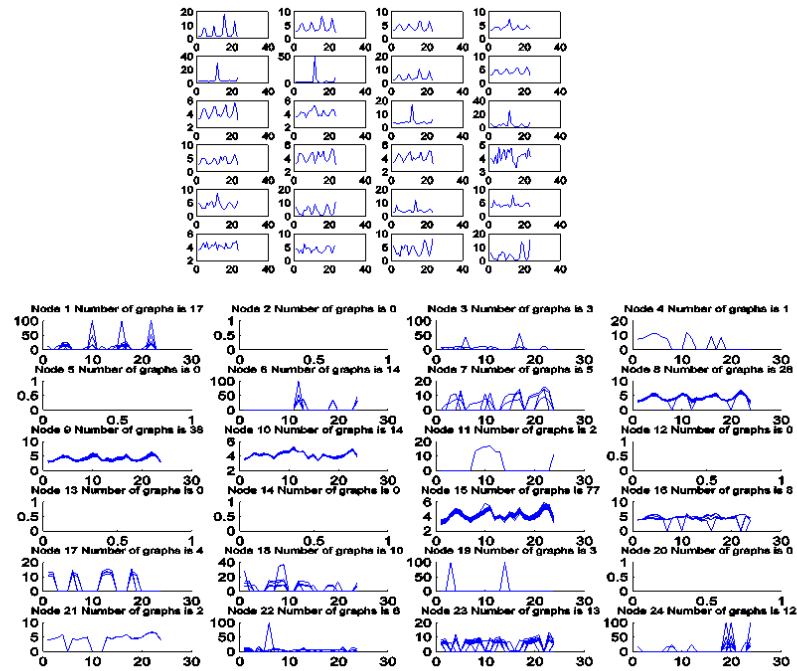


Fig. 4. SOM frequent patterns ($S = 8\%$)

The framework was evaluated using the UK's Cattle Tracking System (CTS) data base. More specifically, it was applied to sequences of snapshots of the database covering from six months to two years of data. The main findings may be summarized as follows:

1. TM-TFP can successfully identify trends in large social networks with reasonable computational efficiency.
2. The SOM clustering/visualisation technique provides a useful mechanism for grouping similar trends.
3. From the SOM, to ease decision makers to spot trend lines in each cluster for further investigation to be taken.

The research team have been greatly encouraged by the results produced to date and are currently investigating further mechanisms where by understanding and advanced analysis of the identified trends can be facilitated.

References

1. Aggrawal, C., Yu, P.: A Condensation Approach to Privacy Preserving Data Mining. In EDBT. LNCS, vol. 2992, pp. 183-199. Springer, Berlin (2004)
2. Read, J.M., Eames, K.T.D., Edmunds, W.J.: Dynamic Social Networks and the implications for the spread of infectious disease. J. R. Soc. Interface, vol. 5, pp. 1001-1007 (2008)

3. Chalabi, Z., Berrached, N., Kharchouche, N., Ghellemallah, Y., Mansour, M., Mouhadjer, H.: Classification of the Medical Images by the Kohonen Network SOM and LVQ. *Journal of Applied Sciences*, vol. 8(7), pp. 1149-1158 (2008)
4. Coenen, F., Leng, P., Ahmed, S.: Data Structures for association Rule Mining: T-trees and P-trees. *IEEE Transactions on Data and Knowledge Engineering*, vol. 16(6), pp. 774-778 (2004)
5. Coenen, F.P., Goulbourne, G., Leng, P.: *Computing Association Rules Using Partial Totals. Principles of Data Mining and Knowledge Discovery. LNCS*, vol. 2168, pp. 54-66. Springer Berlin / Heidelberg (2001)
6. Cottrell, M., Rousset, P.: A powerful Tool for Analyzing and Representing Multidimensional Quantitative and Qualitative Data. In *Proceedings of IWANN 97. LNCS*, vol. 1240, pp. 861-871. Springer Berlin / Heidelberg (2006)
7. Domingos, P.: Mining Social Networks for Viral Marketing. *IEEE Intelligent Systems*, vol. 20(1), pp. 80-82 (2005)
8. Green, D.M., Kao, R.R.: Data quality of the Cattle Tracing System in Great Britain. *Veterinary Record*, vol. 161(13), pp. 439-443 (2007)
9. Khan, M.S., Coenen, F., Reid, D., Tawfik, H., Patel, R., Lawson, A.: A Sliding Windows based Dual Support Framework for Discovering Emerging Trends from Temporal Data. *KBS Journal*, to be appeared in 2010
10. Kohonen, T.: *The Self Organizing Maps. Neurocomputing Elsevier Science*, vol. 21, pp. 1-6 (1998)
11. Kohonen, T.: *The Self Organizing Maps. Series in Information Sciences*, vol. 30. Springer, Heidelberg (1995)
12. Kohonen, T., Oja, E., Simula, O., Visa, A., Kangas, J.: Engineering applications of the Self-Organizing Map. *Proceedings of the IEEE*, vol. 84(10), pp. 1358-1384 (1996)
13. Krause G., Blackmore C., Wiersma S., Lesneski C., Woods CW., Rosenstein N.E., Hopkins R.S.: Marijuana use and Social Networks in a Community Outbreak of Meningococcal Disease. *South Medical Journal*, vol. 94(5), pp. 482-485 (2001)
14. Lauw, H., Lim, E., Pang, H., Tan T.: *Social Network Discovery by Mining Spatio-Temporal Events. Computational Mathematical Organization Theory*, vol 11(2), pp. 97-118. Springer Netherlands (2005)
15. McCallum, A., Corrada-Emmanuel, A., Wang, X.: Topic and Role Discovery in Social Networks with Experiments on Enron and Academic Email. *Journal of Artificial Intelligence Research*, vol. 30, pp. 249-272 (2007)
16. Mitchell A., Bourn D., Mawdsley J., Wint W., Clifton-Hadley R., Gilbert M.: Characteristics of cattle movements in Britain : An analysis of records from the Cattle Tracing System. *Animal Science*, vol. 80, pp. 265-273 (2005)
17. Robinson, S. And Christley, R. M.: Identifying temporal variation in reported births, deaths and movements of cattle in Britain. *BMC Veterinary Research*, pp. 2-11 (2006)
18. Streibel, O.: Trend Mining with Semantic-Based Learning. *Proceedings of CAiSE-DC (2008)*
19. Vesanto, J., Himberg, J., Alhoniemi E., Parhankangas, J.: Self-Organizing Map in Matlab: the SOM Toolbox. In *Proceedings of the Matlab DSP Conference (200)*
20. Wasserman, S., Faust, K.: *Social Network Analysis: Methods and Applications. Cambridge University Press (2006)*
21. Defra. Livestock movements, identification and tracing: Cattle Tracing System. <http://www.defra.gov.uk/foodfarm/farmanimal/movements/cattle/cts.htm>
22. Google Trends. <http://www.google.com/intl/en/trends/about.html>