

Visualisation of Trend Pattern Migrations in Social Networks

Puteri N.E. Nohuddin¹(✉), Frans Coenen², Rob Christley³,
and Wataru Sunayama⁴

¹ Institute of Visual Informatics, Universiti Kebangsaan Malaysia, Bangi, Malaysia
`puteri.iv@ukm.edu.my`

² Department of Computer Science, University of Liverpool, Liverpool, UK
`frans@liverpool.ac.uk`

³ School of Veterinary Science, University of Liverpool and National Centre
for Zoonosis Research, Leahurst, Neston, UK
`robc@liverpool.ac.uk`

⁴ Graduate School of Information Sciences, Hiroshima City University,
Hiroshima, Japan
`sunayama@sys.info.hiroshima-cu.ac.jp`

Abstract. In data mining process, visualisations assist the process of exploring data before modeling and exemplify the discovered knowledge into a meaningful representation. Visualisation tools are particularly useful for detecting patterns found in only small areas of the overall data. In this paper, we described a technique for discovering and presenting frequent pattern migrations in temporal social network data. The migrations are identified using the concept of a Migration Matrix and presented using a visualisation tool. The technique has been built into the Pattern Migration Identification and Visualisation (PMIV) framework which is designed to operate using trend clusters which have been extracted from big network data using a Self Organising Map technique. The PMIV is also aimed to detect changes in the characteristics of trend clusters and the existence of communities of trend clusters.

Keywords: Trend analysis · Trend clustering · Visualisation · Self organising maps · Frequent patterns

1 Introduction

Our proposed Pattern Migration Identification and Visualisation (PMIV) framework detects changes in trend clusters in social network data. The changes are in terms of: (i) trend cluster membership, specifically pattern migrations, (ii) the nature of the trend clusters, i.e. size and existence of clusters in a sequence of data, and (iii) communities of trend clusters that are connected with one another in which pattern migrations occurred. The PMIV framework is proposed to ease the users in decision making and strategic planning. The framework consist of two main process: (i) Migration Matrix that identify changes in trend clusters,

and (ii) Visuset, a visualisation software tool to illustrate the result in effective manner.

The rest of the paper is organised as follows: In Sect. 2 the related work of trend cluster analysis and visualisation are described. Section 3 explains the concept of proposed PMIV framework. Then, Sect. 4 provides some description on the social network datasets used for the demonstration. Section 5 describes the Migration Matrices algorithm for identifying pattern migration followed by Sect. 6 the customised visualisation tool is introduced. Section 7 demonstrates the trend cluster analysis using PMIV framework. Finally, the paper is concluded in Sect. 8.

2 Previous Work

The proposed Pattern Migration Identification and Visualisation framework is founded on a number of mechanisms, namely frequent pattern mining and Self Organising Maps (SOMs) clustering, trend cluster analysis and Visuset software. However, the work in this paper focuses only on the trend cluster analysis and visualisation. The previous work on frequent pattern mining and SOM clustering can be found in [5,6]. The section is concluded with a review of some alternative approaches to trend cluster analysis (Subsect. 2.1) and visualisation in data mining (Subsect. 2.2) related to the work described in this paper.

2.1 Trends Cluster Analysis

Trends act as indicators to inform the direction and/or update on events or occurrence of situations which normally involved with temporal data. The identification of trends can be done in many ways. For example, there a number of related work on trends in time stamped in terms of Jumping and Emerging Patterns. Emerging Patterns describe patterns with frequency counts change between time stamps [4]. Whereas, Jumping Patterns describe patterns whose support counts change drastically (jump) from one time stamp to another. The concept of frequent pattern trends defined in terms of sequence of frequency counts has also been adopted in [7] in the context of longitudinal patient datasets. In [7], trends are categorised according to pre-defined prototypes. Likewise, the work described in this paper collected trends from frequency counts of temporal frequent patterns discovered from a sequence social network datasets. Thus, with the trends, further analysis are able to carry out for the decision makings.

This paper describes trend clusters that has been grouped using an unsupervised clustering method. The clustering method adopted the technology of SOM, a type of artificial neural network, was first introduced by Kohonen [1]. A SOM is an effective visualisation method to translate high dimensional data into a low dimension grid (map), with $x \times y$ nodes.

As noted above, trends are grouped to form clusters. These clusters are therefore referred to as *trend cluster*. The trend cluster analysis, described in this

paper, involves observing and recognizing cluster changes in terms of (say) cluster size or cluster membership. There are several reported studies concerning the detection of cluster changes and cluster membership migration. Denny *et al.* [8] proposed a technique to detect temporal cluster changes using SOMs to visualize emerging, splitting, disappearing, enlarging or shrinking clusters in the context of taxation datasets. Lingras *et al.* [9] proposed the use of Temporal Cluster Migration Matrices (TCMM) for visualizing cluster changes in e-commerce site usage. As will become apparent later in this paper, a related idea founded on the concept of migration matrices, will be proposed.

The proposed trend cluster analysis is also founded on the Newman Hierarchical Clustering algorithm to detect the communities of clusters that interacted as small groups rather than as a big group within a social network. Hierarchical clustering is widely used in cluster analysis tools. Examples include: identification of the similarity and dissimilarity between cancer cells clusters [10], detecting road accident “black spots” using a road traffic cluster analysis [11] and determining the relationship between various industries based on the movements of financial stock prices [12]. As noted in Sect. 6, Hierarchical clustering identifies communities of clusters according to some similarity value [2].

2.2 Visualisation

Visualisation is a tool to assist users explore, understand and analyse data. It also enables researchers and other users to investigate datasets to identify patterns, associations, trends and so on. They should provide an effective representation of the processed data and help to interpret any related concerns and issues. Thus, an effective data visualisation can help users make robust decisions based on the data being presented. Applications in strategic planning, service delivery and performance monitoring have been supported using data visualisation tools immensely. Since data mining usually involves extracting “hidden” information from large datasets, the result interpretation process can get considerably complicated. This is because, in data mining, it extracts information from a database that the user did not already know about. Therefore there are many ways to graphically represent a result model, the visualisations that are used (hopefully) to describe the relationships between data attributes to the users.

In this paper, the work are related to visualisation of changes happened in trend clusters collected within social network data. The aim of visualisation is similar to the application in the general data mining so as to highlight the relationship of changes in the trends data and cluster membership. There is some reported work on data visualisation of trends [3] and cluster change [8]. The work described this paper, describes a visualisation method for: (i) detecting large amounts of frequent patterns migrations from one trend cluster in i epoch¹ to another trend cluster in $i + 1$ epoch, and (ii) identifying communities of trend clusters in social networks. The customised visualisation tool for this framework is called Visuset.

¹ An epoch is defined in terms of a start and an end time stamp.

Visuset is a 2-D visualisation software tool that was developed for chance discovery [15]. It represents node communities, using a 2-D drawing area, based on the Spring Model [13]. It highlights which nodes are connected directly and indirectly with other nodes in detected communities which are depicted as “islands”. Nishikido *et al.* [14] presented Visuset as an animation interface to illustrate change points in keyword relationship networks. This was considered to be a chance discovery tool because it discovered significant candidates (keywords) that benefited the utilization and selection process. Visuset provides a clear animation of communities of cluster to highlight which clusters connect to which clusters. The significance of Visuset is that the research described in this thesis utilizes Visuset to support trend cluster analysis and visualisation of significant dynamic cluster changes in sequences of data.

3 The Pattern Migration Identification and Visualisation (PMIV) Framework

The PMIV framework is directed at finding interesting pattern migrations between trend clusters and trend changes in social network data. In this work, trends are trend line representing frequency counts of binary valued frequent patterns discovered using Trend Mining-Total From Partial (TM-TFP) algorithm [5, 6] in epochs of social network data. A trend can be said to be interesting if its “shape” changes significantly between epochs. Thus to perform further analysis, the trends then are clustered into similar “shape” using SOM.

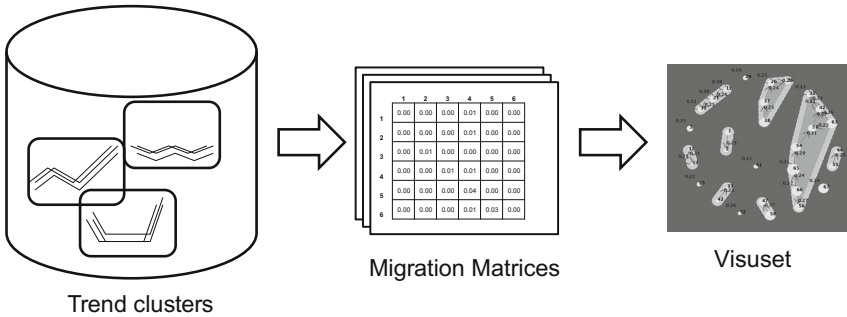


Fig. 1. Schematic illustrating the operation of the proposed framework

Figure 1 gives a schematic of the PMIV framework. The input to the PMIV framework is a set of $TC = \{\tau_1, \tau_2, \dots, \tau_n\}$ partitioned according to m SOMs (note that n should typically be determined by the number of nodes in SOM that contains groups of similar trends). The set of trends associated with a SOM E_j where the complete set of trends, in a sequence of SOM E , is then given by $E = \{E_1, E_2, \dots, E_m\}$.

The next stage is to detect pattern migrations in the sets of trend clusters TC between k pairs of SOMs E_j and E_{j+1} , where $k = m - 1$. This is achieved by generating a sequence of Migration Matrices for each pair. The Migration Matrices provide the information on members of trend clusters in SOM E_j (map node number) moved to trend clusters in SOM E_{j+1} . This information is used to determine values the communities of trend clusters and illustrate the animation of pattern migrations using Visuset (Sect. 6).

The final stage is to illustrate the pattern migration visualisation using Visuset. The objective is then to identify interesting of pattern migrations of individual trend clusters that exist across the set E . Note that, some patterns may remain in the trend cluster for the entire sequence of m maps. Some other patterns may fluctuate between clusters. In addition, the changes in size of temporal trend clusters can also be observed.

4 Social Network Dataset

The work described in this paper is directed at three specific social networks. The second was extracted from the Cattle Tracing System (CTS) database in operation in Great Britain (GB) and is referred to as the CTS Network. The first is a customer network extracted from an insurance company's database referred to as the Deeside insurance quote network. The third is the logistic item cargo distribution network from Malaysian Armed Forces (MAF). In all mentioned networks comprises of sets of n time stamped datasets $D = \{d_1, d_2, \dots, d_n\}$ partitioned into m epochs.

Each data set d_i comprises a set of records such that each record describes a social network node pairing, the description consists of some subset of a global set of attributes A that describes the network. There are $2^{|A|} - 1$ patterns that may exist in any given dataset. The *support* (s) for a pattern I in a dataset d_i is the number of occurrences above σ , a support threshold, of the pattern in d_i expressed as a percentage of the number of records in d_i . As mentioned above, the pattern and its support are discovered in [5, 6]. The n support counts for each patterns is set to represent a pattern's trend. Then, the large number of trends are group into trend clusters for further analysis using PMIV.

The cattle movement network was extracted from the Cattle Tracing System (CTS) database in operation in Great Brittain (GB). The CTS database was introduced in September 1998 and updated in 2001 as a result of a number of outbreaks of bovine diseases. The database is maintained by DEFRA, the Department for Environment, Food and Rural Affairs, a UK government department. The database records all cattle movements in GB, each record describes the movement of a single animal (cattle), identified by a unique ID number, between two *holding locations* (farms, markets, slaughter houses, etc.). However, the CTS database can be interpreted as a social network, where each node represents a geographical location and the links the number of cattle moved between locations. The links describe specific types of cattle movement, thus there could be more than one link between pairs of nodes. An example of a pattern of cattle movement that might be attached to a link is:

$$\{NumberOfAnimalMoved = \{50\}, BeastType = Liung, AnimalAge = \{1 : 3\}, PTI = 4 \text{ and } Senderarea = 13\}$$

where the attribute label *PTI* is the Parish Testing Interval which describes the frequency of disease detection testing for each node; the value is between 1 and 4 years. The number of cattle attached to the link is 50. Each node describes a location defined in terms of 100 Km grid squares. Four years worth of data, from 2003 to 2006, were divided into four epochs of 12 months each. After discretisation and normalisation, the average number of nodes within a single network was 150,000 and the average number of links was 300,000 with 445 attributes.

The Deeside Insurance Quote network was extracted from a sample of records taken from the customer database operated by Deeside Insurance Ltd. (collaborators on the work described in this paper). Two years of data, from 2008 to 2009, were obtained comprising, on average, 400 records per month. In total, the data set comprised 250 records with (after discretisation and normalisation) 314 attributes. The data was divided into two epochs comprising 12 months each. The Deeside can also be viewed as a network, the nodes comprised postal areas (characterised by the first few digits of UK post/zip codes), and the links are the number of requests for specific types of insurance quotes received for the given time stamp. The Deeside office is viewed as the “super node” that can have many links to the outlying nodes (customers’ postcodes). An example of a pattern of Deeside network, attached to a link, might be:

$$\{CarType = Vauxhall, EngineSize = \{1500 : 1999\}, OffenceCode = SP, Fine = \{200 : 500\} \text{ and } Gender = Male\}$$

where the value $\{1500:1999\}$ states the *EngineSize* is within the range 1500 and 1999, the value SP is an *OffenceCode* indicating exceeding the speed limit, and the value $\{200:300\}$ indicates a *Fine* of between 200 and 500. The average number of records in a single (one month) time stamps was about 800, and the average number of links 200.

The MAF Logistic Cargo Distribution dataset described the shipment of logistics items for Malaysian Army, Air Force and Navy. The example of logistic items were vehicle, medicines, military uniforms, ammunition and repair parts. The dataset was extracted from the 2008 to 2009 records to form 2 episodes with 12 time stamps each. In the MAF network, logistic items were sent from a number of division logistic headquarters to brigades and then to specific battalions in West and East Malaysia. The location of headquarters, brigades and battalions are the nodes of the MAF network. These offices were viewed as being sender and receiver nodes (in a similar manner as described for the CTS) and the shipments as links connecting nodes in the network. Each month comprised of some 100 records. An example of a pattern of MAF network, attached to a link, might be:

$$\{Item = 1tonnetruck, Sender = 4Armor, Receiver = 1Armor, ShipmentCost = \{200000 : 500000\}\}$$

where the Item described the logistic items that were sent from Sender to Receiver as mentioned in the example. The ShipmentCost described the estimated total cost between MYR200000 and MYR500000 of the cargo distribution from sender's location to receiver's location.

5 Migration Matrices

As mentioned above, the trend clusters are generated from the clustering of large numbers of trends into the epochs of SOMs E_m . In the Migration Matrices (MMs) algorithm, the changes in trends associated with patterns can be measured by interpreting the SOMs in terms of a rectangular (2-D plane) where each point in the plane represents a SOM node. Thus, given a sequence of trend line SOM maps comparisons can be made to see how trends associated with individual frequent patterns change by analyzing the nodes in which they appear. The MMs algorithm is described in Algorithm 1.1.

Algorithm 1.1. Migration Matrices

```

input : FP = Set of all frequent patterns in episodes  $\{e_1, e_2, \dots, e_e\}$ 
output: Sequence of  $(e - 1)$  Migration Matrices
1 Define Table measuring  $|FP| \times e$ ;
2 for  $i \leftarrow 1$  to  $e$  (step through episodes) do
3   | for  $j \leftarrow 1$  to  $|FP|$  (step through the set of FP) do
4   |   | Table[i][j] = SOM node ID for  $e_i$ ;
5   | end
6 end
7 Define  $(e - 1)$  Migration Matrices (MMs), each measuring  $(x \times x)$  where  $x$ 
   is the number of SOM nodes;
8 for  $i \leftarrow 1$  to  $(e - 1)$  do
9   | for  $k \leftarrow 1$  to  $|FP|$  do
10  |   | Increment count at  $MM_i[Table[K][i]][Table[K][i+1]]$ ;
11  | end
12 end
    
```

The algorithm commences by defining a $|FP| \times e$ table. The table is populated with the SOM node IDs, the trend cluster, and for each trend line in the trend cluster has an associated frequent pattern, for SOM maps $E = \{E_1, E_2, \dots, E_m\}$ (line 4). Then in line 7, the algorithm defines a sequence of $m - 1$ MMs for each pair of SOM E_j and E_{j+1} , each measuring $x \times x$. x is the number of SOM nodes that are the number of trend clusters. The process continues by comparing the node numbers of the frequent pattern and counting the pattern migrations for each node ID (trend cluster) between SOM E_j and E_{j+1} (lines 8 and 10).

Subsequently, the algorithm also produces a trend cluster analysis of the pattern migrations between trend clusters. The analysis comprises a comparison of pattern migrations for each pair of SOM E_j and E_{j+1} . The number of patterns migrating from node _{i} in E_j to node _{j} in E_{j+1} are recorded. It is also possible to

determine how the sizes of trend clusters in a given pair of SOMs, E_j and E_{j+1} , change. This analysis thus provides for identification of patterns that move, or do not move, between successive SOM nodes, which may be of interest given particular applications.

The result of MMs are also supported with a visualisation tool. Further description is in the following Sect. 6.

6 Visualisation of Pattern Migrations

The Visualisation of Pattern Migrations produced “pattern migration maps” using Visuset that uses the concept of *Spring Model* [13]. The spring model for drawing graphs in 2-D space is designed to locate nodes in the space in a manner that is both aesthetically pleasing and limits the number of edges that cross over one another. The graph to be depicted is conceptualised in terms of a physical system where the edges represent springs and the nodes objects connected by the springs. Nodes connected by “strong springs” therefore attract one another while nodes connected by “weak springs” repulse one another. The graphs are drawn following an iterative process. Nodes are initially located within the 2D space using a set of (random) default locations (defined in terms of an x and y coordinate system) and, as the process proceeds, pairs of nodes connected by strong springs are “pulled” together. In the context of PMIV The network nodes are represented by the trend clusters in SOM E and the spring value was defined in terms of a *correlation coefficient* (C):

$$C_{ij} = \frac{X}{\sqrt{(|E_{ki}| \times |E_{k+1j}|)}} \quad (1)$$

where C_{ij} is the correlation coefficient between a node (trend cluster) i in SOM E_k and a node j in SOM E_{k+1} (note that i and j can represent the same node but in two different maps), X is the number of trend lines that have moved from node i to j and $|E_{ki}|$ ($|E_{k+1j}|$) is the number trends at node i (j) in SOM E_k (E_{k+1j}). A migration is considered “interesting”, and thus highlighted by Visuset, if C is above a specified minimum relationship threshold (Min-Rel). With respect to all network we have discovered that a threshold of 0.2 is a good working Min-Rel value. The Min-Rel value is also used to prune links and nodes; any link whose C value is below the Min-Rel value is not depicted.

To aid the further analysis of the identified pattern migrations it was also considered desirable to identify “communities” within networks (SOM E), i.e. clusters of nodes which were “strongly” connected (feature significant migration). This would indicate significant groupings of patterns whose associated trend lines were changing between SOM E_k and E_{k+1} . An agglomerative hierarchical clustering mechanism, founded on the Newman method [16] for identifying clusters in network data, was therefore adopted. Newman proceeds in the standard iterative manner on which agglomerative hierarchical clustering algorithms are founded. The process starts with a number of clusters equivalent to the number of nodes. The two trend clusters (nodes) with the greatest “similarity” are

then combined to form a merged cluster. The process continues until a “best” cluster configuration is arrived at or all nodes are merged into a single cluster. The overall process is typically conceptualised in the form of a dendrogram. Best similarity is defined in terms of the *Q-value*, this is a “modularity” value which is calculated as follows:

$$Q_i = \sum_{i=1}^{i=n} (c_{ii} - a_i^2) \quad (2)$$

where Q_i is the Q-value associated with the *current* node i , n is the total number of nodes in the network, c_{ii} is the fraction of intra-cluster (within cluster) links (trend lines) in cluster i over the total number of links in the network, and a_i^2 is the fraction of links that end in the nodes in cluster i if the edges were attached at random. The value a_i is calculated as follows:

$$a_i = \sum_{j=1}^{j=n} c_{ji} \quad (3)$$

where c_{ij} is the fraction of inter-cluster links, between the current cluster i and the cluster j , over the total number of links in the network.

In this paper, the implementation of Visuset, trend clusters are depicted as single nodes that might have *self-links* of pattern migrations within the same trend clusters themselves, node pairs linked by an edge, chains of nodes linked by sequence of edges or “islands” of nodes that represent as communities of nodes. This will be demonstrated in Sect. 7.

7 Analysis of PMIV Using Social Network Trend Clusters

Each of the three network is considered in turn in this section. Table 1 provides number of trends and trend clusters for CTS, Deeside Insurance and MAF Cargo Distribution networks. The support threshold (σ) used to mine frequent patterns and trends for CTS network is 0.5 %, and Deeside Insurance and MAF Logistic Cargo networks is 5 %. The trend clusters are discovered, for all networks, using SOM. The number of nodes that consist of trends in SOMs determines the number of trend clusters. Therefore, CTS network has 100 trend clusters, Deeside Insurance and MAF Cargo Distribution networks have 49 trend clusters. CTS network had the biggest SOM node configuration as it is the largest dataset compared to the other two networks.

Then, the trend clusters are processed using MMs algorithm to identify the number of pattern migrations in each pair of SOM E_j and E_{j+1} , a generic example is shown in Table 2. The table shows a MM that provides the numbers of CTS patterns that have migrated from E_{2003} to E_{2004} , $n_{1,1} = 71$, the number of patterns that have stayed (*self-links* in cluster c_1 in both SOM maps, $n_{1,2} = 13$, the number of CTS patterns that have migrated from c_1 in E_{2003} to c_2 in E_{2004} . The Q-values required for the hierarchical clustering are calculated using these numbers of pattern migrations. The numbers of patterns migrated

Table 1. Number of trend clusters for CTS, Deeside Insurance and MAF cargo distribution networks

Networks	Support Threshold σ (%)	Trends in E_j	Trends in E_{j+1}	Number of Trend cluster
CTS	0.5	63117	66870	100
Deeside insurance	5	55241	49983	49
MAF	5	3491	2609	49

Table 2. Migration matrix for CTS network pattern migrations from E_{2003} to E_{2004}

	n_1	n_2	n_3	n_4	n_5	\dots	n_{100}
n_1	71	13	11	0	0	\dots	0
n_2	10	6	13	0	1	\dots	1
n_3	8	0	21	2	0	\dots	0
n_4	0	0	0	3	0	\dots	0
n_5	0	0	0	0	14	\dots	0
\dots	\dots	\dots	\dots	\dots	\dots	\dots	0
n_{100}	0	0	0	0	0	\dots	7

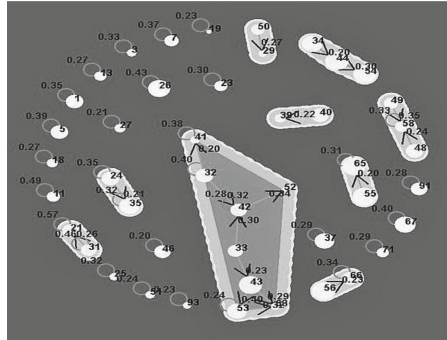


Fig. 2. CTS network pattern migrations for E_{2003} and E_{2004}

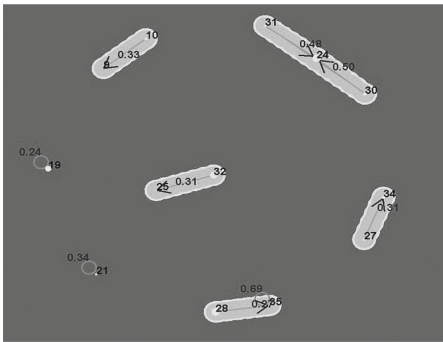


Fig. 3. Deeside insurance network pattern migrations for E_{2003} and E_{2004}

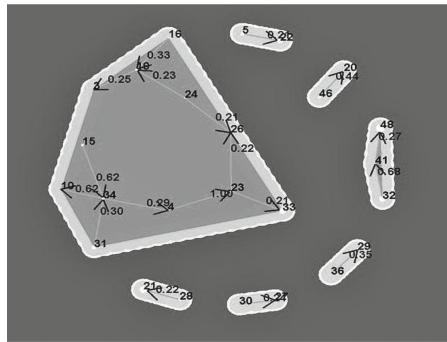


Fig. 4. MAF logistic network pattern migrations for E_{2003} and E_{2004}

in the Table 2, are also used to determine the Q and C-values for visualising the pattern migrations. These Q-values are used to cluster the nodes (trend clusters) so as to detect communities of nodes with pattern migrations. As mentioned, the C-values were used to identify the positions and relationships of trend cluster nodes in the network to support the animation of pattern migrations. Similar MMs were also generated for Deeside Insurance and MAF Cargo Distribution networks

In the second process, Visuset took the generated MMs to illustrate the pattern migration maps. The maps for all three networks are shown in Figs. 2, 3 and 4. Note that, in the maps, trend clusters are represented as nodes and migrations of patterns are shown as links. The C-values threshold for pattern migrations between nodes is above 0.2.

Figure 2 shows that the CTS network pattern migration map between E_{2003} and E_{2004} with 45 nodes out of a total of 100 that had a C-value greater than 0.2. Several islands are displayed, determined using the Newman method described above, including a large island comprising eight nodes. The islands indicate communities of pattern migrations. The nodes are labeled with an identifier (the trend cluster number in E_j), the links with their C-value numbers and link directions show the migration of patterns to new trend cluster in E_{j+1} . From the map we can identify that there are a 30 nodes, of self-links. However, we can deduce that (for example) patterns are migrating from node 34 to node 44, and from node 44 to 54 (thus indicating a trend change). The size of the nodes also indicates how many patterns in the cluster, bigger node has large patterns in it.

Figure 3 depicts the Deeside network pattern migration map between E_{2008} and E_{2009} . There were 13 nodes that have C-value greater than 0.2. Note that the node 28 has C-value of 0.69, the highest pattern migrations to the node 36 in E_{2009} . There were also self-links occurred for only two nodes, 19 and 21. We can also notice that the size of nodes 19, 24, 28 and 32 were among the largest which means the number of cluster members was high compared to other clusters. There were five islands of nodes in which nodes 24, 30 and 31 formed the largest island in the map.

Finally, the pattern migrations map for MAF Cargo Distribution network is shown in Fig. 4. There 24 nodes out of 49 were shown in the map. Unlike the other maps, the network did not have any self-link pattern migrations. The map also had seven islands of nodes, the largest island has about 12 nodes that connected to one another for pattern migrations. There were also some nodes received new cluster members in E_{2009} such as node 10, 26 and 34.

8 Conclusions

In this paper, the authors have described the PMIV framework for detecting changes in trend clusters within social network data. The trend clusters consist of trends are defined in terms of sequences of support counts associated with individual patterns across a sequence of time stamps associated with an *epoch*. The trend clusters are analysed for identifying pattern migrations between pairs of trend clusters found in SOM E_j and E_{j+1} . The pattern migrations are detected using the Migration Matrices algorithm and Visuset, a visualisation software tool. Visuset in PMIV provides useful information: (i) pattern migrations of trend cluster (membership), (ii) changes in trend clusters, and (iii) communities of trend clusters of pattern migrations.

References

1. Kohonen, T.: The self organizing maps. *Neurocomput.* Elsevier Sci. **21**, 1–6 (1998)
2. Han, J., Kamber, M., Pei, J.: *Data Mining: Concepts and Techniques*, 3rd edn. Morgan Kaufmann, San Francisco (2011)
3. Adomavicius, G., Bockstedt, J.: C-TREND: temporal cluster graphs for identifying and visualizing trends in multiattribute transactional data. *J. IEEE Trans. Knowl. Data Eng.* **20**, 721–735 (2008)
4. Dong, G., Li, J.: Efficient mining of emerging patterns: discovering trends and differences. In: *Proceeding of Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 43–52 (1999)
5. Nohuddin, P.N.E., Christley, R., Coenen, F., Setzkorn, C.: Trend mining in social networks: a study using a large cattle movement database. In: Perner, P. (ed.) *ICDM 2010. LNCS*, vol. 6171, pp. 464–475. Springer, Heidelberg (2010)
6. Nohuddin, P.N.E., Coenen, F., Christley, R., Setzkorn, C.: Detecting temporal pattern and cluster changes in social networks: a study focusing UK cattle movement database. In: Shi, Z., Vadera, S., Aamodt, A., Leake, D. (eds.) *IIP 2010. IFIP AICT*, vol. 340, pp. 163–172. Springer, Heidelberg (2010)
7. Somaraki, V., Broadbent, D., Coenen, F., Harding, S.: Finding temporal patterns in noisy longitudinal data: a study in diabetic retinopathy. In: Perner, P. (ed.) *ICDM 2010. LNCS*, vol. 6171, pp. 418–431. Springer, Heidelberg (2010)
8. Denny, Williams, G.J., Christen, P.: Visualizing temporal cluster changes using relative density self-organizing maps. *J. Knowl. Inf. Syst.* **25**, 281–302 (2010)
9. Lingras, P., Hogo, M., Snorek, M.: Temporal cluster migration matrices for web usage mining. In: *Proceedings of the 2004 IEEE/WIC/ACM International Conference on Web Intelligence*, pp. 441–444 (2004)
10. Ng, R., Jorg, S., Sleumer, M.: Hierarchical cluster analysis of SAGE data for cancer profiling. In: *Proceedings of BIOKDD 2001*, pp. 65–72 (2001)
11. Vijayakumar, M., Parvathi, R.M.S.: Concept mining of high volume data streams in network traffic using hierarchical clustering. *J. Eur. J. Sci. Res.* **39**(2), 234–242 (2010)
12. Wittman, T.: *Time-series clustering and association analysis of financial data*. Project thesis (2002)
13. Sugiyama, K., Misue, K.: Graph drawing by the magnetic spring model. *J. Vis. Lang. Comput.* **6**(3), 217–231 (1995)
14. Nishikido, T., Sunayama, W., Nishihara, Y.: Valuable change detection in keyword map animation. In: Gao, Y., Japkowicz, N. (eds.) *AI 2009. LNCS*, vol. 5549, pp. 233–236. Springer, Heidelberg (2009)
15. Ohsawa, Y.: Modeling the process of chance discovery. In: Ohsawa, Y., Abe, A. (eds.) *Advances in Chance Discovery. SCI*, vol. 423, pp. 2–15. Springer, Heidelberg (2013)
16. Newman, M.E.J.: Fast algorithms for detecting community structure in networks. *J. Phys. Rev. E* **69**, 1–5 (2004)