(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
02 November 2017 (02.11.2017)

WIPO | PCT

(10) International Publication Number
WO 2017/187401 A1

(51) International Patent Classification:
G06F 17/30 (2006.01)

(21) International Application Number:
PCT/IB2017/052469

(22) International Filing Date:
28 April 2017 (28.04.2017)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/329,310 29 April 2016 (29.04.2016) US

(71) Applicant: **ECOLE POLYTECHNIQUE FEDERALE DE LAUSANNE (EPFL)** [CH/CH]; EPFL-TTO, EPFL Innovation Park J, 1015 Lausanne (CH).

(72) Inventors: **VANDERGHEYNST, Pierre**; Ch. de Montaney 50, 1024 Ecublens (CH). **RICAUD, Benjamin**; Rue du Miroir 253, 74500 Maxilly (FR). **BENZI, Kirell**; Ch. de la Gravière 2, 1008 Prilly (CH).

(74) Agent: **ROLAND, André**; c/o ANDRE ROLAND SA, P.O. Box 5107, 1002 Lausanne (CH).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report (Art. 21(3))*

(54) **Title:** SYSTEM, DEVICE, AND METHOD FOR CONTEXTUAL KNOWLEDGE RETRIEVAL AND DISPLAY

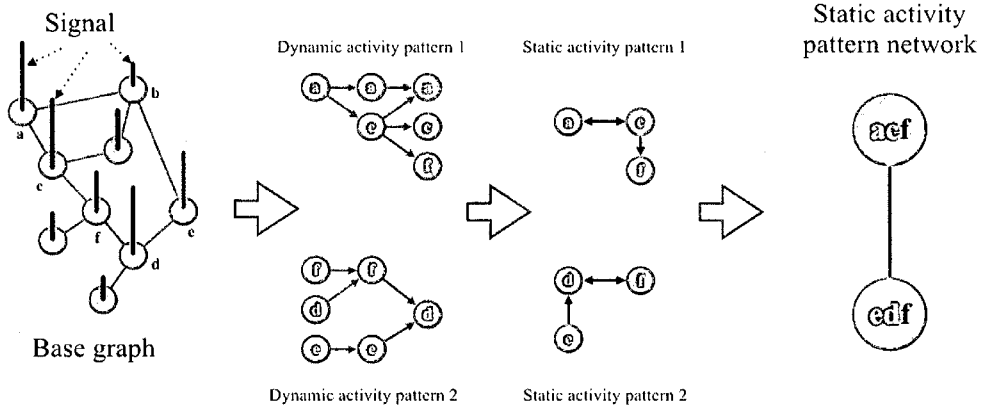

FIG. 9

(57) **Abstract:** A method to extract patterns of activity from time series of data structured as a network of objects, the method comprising the steps of creating a base graph of similar or related items from a given corpus by comparing intrinsic features of the items, combining time varying data on nodes or edges of the base graph, creating dynamic activity patterns by connecting nodes of the base graph according to a measure of activity from the time series of data, and creating static activity patterns by folding the dynamic activity patterns according to identifiers of the nodes from the base graph.

WO 2017/187401 A1

SYSTEM, DEVICE, AND METHOD FOR CONTEXTUAL KNOWLEDGE RETRIEVAL AND DISPLAY

CROSS-REFERENCE TO RELATED APPLICATIONS

(1) The present application claims priority to the United States provisional patent application with the Serial No. 62/329,310 that was filed on April 29, 2016, the entire contents thereof herewith incorporated by reference.

FIELD OF THE INVENTION

(2) The invention relates to a system, a device, and method to extract, rank and display contextual knowledge from time varying data on a network of objects.

BRIEF DISCUSSION OF THE BACKGROUND ART

(3) The number of Internet users grows steadily, reaching for the first time the cap of 3 billion in 2015. Every year, the amount of digital information created doubles and by 2020, around 40 of trillion gigabytes of storage will be required to store these data. To make sense of all those data and create business insights, companies are deploying vast infrastructures to cope with the load. On the software side, new frameworks and algorithms are invented to be able to process data at this scale. In this regard, the creation of a graph (network) of objects allows to bring structure and extract useful analytics for companies. Indeed, the emergence of large-scale graph analytics frameworks such as Google Pregel, Apache GraphX, or Dato GraphLab has proven to be really helpful to analyze similarity between products, connections between pages or user to user interactions in the case of social networks.

(4) In another domain of data analytics, streaming computation systems such as Spark Streaming or Apache Storm are rapidly developing to face the demand of fast analytics on the fly

without having to load all from the disk. It can also reduce significantly the amount of storage needed if raw data are dropped in the process.

(5) While these systems solve some of the problems of static graph analysis and time series analysis, time varying data on the nodes or edges of a graph that could be changing through time is completely left out of the equation. More specifically, it appears that there is a crucial need of a scalable method that could extract, rank, link and display activity patterns on a network by combining a graph of objects and time series on its nodes in order to extract contextual knowledge. The context is given by the specific arrangement of nodes and edges in the network as well as their time values or common dynamic activity. Accordingly, in light of the above deficiencies in the field of the displaying of contextual knowledge, novel and substantially improved solutions are desired.

SUMMARY

(6) According to one aspect of the present invention, a method to extract patterns of activity from time series of data structured as a network of objects is provided. Preferably, the method comprising the steps of creating a base graph of similar or related items from a given corpus by comparing intrinsic features of the items, combining time varying data on nodes or edges of the base graph, creating dynamic activity patterns by connecting nodes of the base graph according to a measure of activity from the time series of data, and creating static activity patterns by folding the dynamic activity patterns according to identifiers of the nodes from the base graph.

(7) According to another aspect of the present invention, a system configured to extract patterns of activity from time series of data structured as a network of objects is provided, the system preferably including a computer device that has access to memory, and a network

interface. Moreover, the computer device is preferably configured to create a base graph of similar or related items from a given corpus of data from the memory by comparing intrinsic features of the items, combine time varying data on nodes or edges of the base graph, create dynamic activity patterns by connecting nodes of the base graph according to a measure of activity from the time series of data, and create static activity patterns by folding the dynamic activity patterns according to identifiers of the nodes from the base graph.

(8) According to yet another aspect of the present invention, a non-transitory computer readable medium is provided, the computer readable medium having computer instructions recorded thereon. The computer instructions are preferably configured to perform a method for extracting patterns of activity from time series of data structured as a network of objects when executed on a computer having memory. Preferably, the method includes the steps of creating a base graph of similar or related items from a given corpus by comparing intrinsic features of the items, combining time varying data on nodes or edges of the base graph, creating dynamic activity patterns by connecting nodes of the base graph according to a measure of activity from the time series of data, and creating static activity patterns by folding the dynamic activity patterns according to identifiers of the nodes from the base graph.

(9) The above and other objects, features and advantages of the present invention and the manner of realizing them will become more apparent, and the invention itself will best be understood from a study of the following description with reference to the attached drawings showing some preferred embodiments of the invention.

BRIEF DESCRPTION OF THE SEVERAL VIEWS OF THE DRAWINGS

(10) The accompanying drawings, which are incorporated herein and constitute part of this specification, illustrate the presently preferred embodiments of the invention, and together

with the general description given above and the detailed description given below, serve to explain features of the invention:

(11) FIG. 1 depicts a schematic representation of a graph generated by the present method representing an event in a context, based on data extracted from Wikipedia, representing the average visitor behavior in a given time period, according to an aspect of the present invention;

(12) FIG. 2 depicts a schematic representation of a timeline of the event of FIG. 1;

(13) FIG. 3 depicts a schematic representation of a method to generate a spatio-temporal graph according to another aspect of the present invention;

(14) FIG. 4 depicts a schematic representation of the generation of spatio-temporal edges from the nodes and their associated binary masks according to yet another aspect of the present invention;

(15) FIG. 5 depicts a schematic representation of static activity and dynamic activity with base graph G and spatio-temporal graph H according to an aspect of the present invention;

(16) FIG. 6 depicts a schematic representation of a static activity pattern according to still another aspect of the present invention;

(17) FIG. 7 depicts a schematic representation of a graphical user interface to show, browse and explore patterns resulting from the method according to an aspect of the present invention;

(18) FIG. 8 shows a schematic perspective view of a system or device for implementing the method according to another aspect of the present invention;

(19) FIG. 9 shows a schematic representation of a flowchart performing the method according to another aspect of the present invention;

(20) FIG. 10 shows a schematic representation of a flowchart performing an exemplary embodiment for user product recommendations;

(21) FIG. 11 shows a schematic representation of a flowchart performing classification of graphs that represent news patterns based on activity patterns, as still another exemplary embodiment; and

(22) FIG. 12 shows a schematic representation of a flowchart performing a financial analysis to track evolution of traded securities based on activity patterns, as yet another exemplary embodiment.

(23) Herein, identical reference numerals are used, where possible, to designate identical elements that are common to the figures. Also, the representations in the drawings are simplified for illustration purposes and may not be depicted to scale.

BRIEF DESCRIPTION OF THE SEVERAL EMBODIMENTS

(24) The invention relates to a system, device, and method for contextual knowledge retrieval based on activity patterns. According to aspects of the invention, it is proposed a system, device and method to extract, rank, link and display information preferably includes the following steps:

(25) Creating or indexing a graph or network of related items. This base graph of related items can be composed of all or a part of the set of objects connected by natural relationships or similarity using intrinsic features of objects such as categories, topics, keywords, words, signals, etc. The graph of related items can be un/directed, dynamic, weighted and composed of multiple type of edges (multigraphs). Collecting activity time series such as number of visits, likes, tweets, favorites, activity score or signal through time on the nodes and/or edges of said graph.

(26) Extracting and indexing dynamic activity patterns that are combinations of the network and the time series associated to the nodes. High values in the time series indicate an activity of the associated nodes during a particular time period. The dynamic activity patterns are sub-networks of connected nodes with a common time activity. These dynamic activity patterns are represented in the form of a multilayer graph where each node is associated to an identifier, a time stamp (or layer identifier) and a value. Each node is connected to its neighbors (and/or itself) on the next layer according to the edges of the base graph and thus can be mapped back to the base graph of related items and to the original time series of activity. The extraction of theses activity patterns can either be done on a given historical corpus or in an online manner, outputting dynamic activity patterns on the fly.

(27) Computing static activity patterns from dynamic activity patterns. A static activity pattern is obtained by flattening the multilayer graph: nodes and edges of the dynamic activity pattern a grouped so as to appear once in the new network. The value associated to each node/edge of the static graph of activity represents its total time activity over the time period.

(28) Enriching and indexing the content of static activity patterns with information about the dynamics of the pattern activation such as the time stamp and duration of the event, the user activity, images, audio, video, ontologies, pages, description, categories, topics, geographical origin, languages, social activity and any other features that can be extracted from the content of the graph of related items. This information are properties that can be associated to nodes, edges or to the static activity pattern. At this stage it is possible to use a classification method to assign patterns to a specific set of classes or clusters. Both dynamic and static patterns are kept for further analysis.

(29) Creating a network of static activity patterns where each node is a static activity pattern connected to others if they are similar. The measure of similarity can be defined in several ways including common items, topics, users, etc.

(30) Retrieving and ranking static/dynamic activity patterns by relevance for a given query. The query could contain any content such as words, topics, categories, time-period, and activity related criteria such as the most popular patterns for a given topic. A query can also be composed of another static/dynamic activity pattern or created from a specific network algorithm on the static activity network.

(31) Displaying the results for given query as a combination of possibly hideable panels including: a search bar, a list of related results for a given query, an interactive visualization of the static activity pattern network, an interactive visualization of the static activity pattern and its associated timeline of showing how nodes activates through time and a panel displaying information on the pattern and about each node. This last panel could also display an embedded view of a webpage, image, video, sound, associated to a node.

(32) According to at least one aspect of the present invention, a device, system and method for contextual knowledge retrieval based on activity patterns is provided, that is preferably divided in three principal parts. First, the extraction of patterns is described, then the indexation and finally the visualization, presentation of these patterns.

(33) Regarding the pattern, FIG. 1 shows a schematic representation of an exemplary graph generated by the present method representing an event in a context, based on data extracted from Wikipedia, representing the average visitor behavior in a given time period, according to an aspect of the present invention. In this representation, the web activity related to the German Wings airplane crash is analyzed. In the circles, different related Wikipedia pages

are shown, associated with a number of web visits per hour. The representation could be a display of a graphical user interface.

(34) In the method, the first step consists in creating the base graph G, of related items. Each node represents an object and each edge between two (2) nodes encodes a relationship between these nodes. An edge can be naturally extracted from the dataset e.g. hyperlinks, ontologies or be constructed by using a similarity measure between the nodes of the graph. The measure of similarity can be created by using a distance function of any or a sum of the features attached to a node such as page, category, topic, user, signal, image, etc. or the activity value associated to the node. The distance function can be diverse such as Euclidean, Cosine, Hamming or learned from the data. This base graph can be stored in memory or on the disk of one or several computers either in the form of a graph database, a graph analytics framework or as a list of nodes and edges. The nodes and edges of the base graph can also be streamed from any other medium and eventually dropped after computations. In the case of a graph changing through time, a binary vector, m, indicating whether the edge exists at a given time stamp can be stored alongside the edge of the base graph. An edge is said active at a given time stamp if its binary value is set to true. The same is done for nodes of the base graph.

(35) The second step of the method consists of collecting the signal over the nodes or edges of the base graph. This signal, S, gives a value, or possibly a vector of values, for each node and time stamp e.g. number of tweets per minutes, number of click or visits per hours. This is not restricted to Internet data and could be applied in all sort of domains such as medical data with fMRI or electrodes activity per seconds etc. where electrodes are parts of a network of sensors. This signal can be stored in memory or on the disk of one or several computers in a SQL, NoSQL, NewSql database, a streaming framework or as plain files. This signal can also be

streamed from any other medium and eventually dropped after computations. From the signal, a binary mask M indicating whether the node/edge of the base should be active at a given time stamp is created by filtering the signal. This filter can consist in a fixed or adaptive threshold or a function mapping from the signal values to a binary space. A node is said active at a given time stamp if its binary value is true as shown in FIG. 4.

(36) The next step consists in creating a temporary graph, the spatio-temporal graph H, that holds all the dynamic activity patterns extracted from a combination of the signal and the base graph. To do so, each active node at a given time step it connected to its active neighbors in the next time step according to the active edges of the base graph G, creating spatio-temporal edges. To give more control to the creation of dynamic activity patterns, an additional binary vector p is also stored alongside the node/edge of the base graph. When set to false it will disable the creation of a spatio-temporal edge that would have normally been created. Optionally, an active node can be connected to its active itself on the next layer. Their corresponding "control" vector p should also to be set to true for a given time step. The whole spatio-temporal graph creation process is shown in FIG. 5, given a base graph G of three (3) nodes A, B, C and signal over three time steps, t_1, t_2, t_3. At each time step, only the active nodes and edges are colored or otherwise highlighted on a display. Spatio-temporal edges connect nodes across layers forming two (2) disconnected sets of nodes and edges, that are two distinct dynamic activation patterns. To extract these disconnected subgraphs from the temporary spatio-temporal graph H, a possible implementation uses the connected or weakly connected components algorithms based on disjoint-set data structure, breadth-first search or depth-first search or implementation for fully dynamic graphs.

(37) A possible implementation of the creation of spatio-temporal edges from the nodes and their associated binary masks is shown in FIG. 6. Similarly, to the standard model, for each edge e the destination vertex mask is shifted before performing a logical AND with the source vector. The mask m_e indicating if the activation state of an edge and the mask p_e are all subject to a logical "AND" between the source and shifted destination masks. The same process is used at the node level to account for self spatio-temporal edges.

(38) From dynamic activity patterns the next step of the method computes static activity patterns. As shown in FIG. 5, the nodes of the spatio-temporal graph H correspond to the dynamic activity of their corresponding nodes on the base graph G. Dynamic activity patterns which are sub-graphs of the spatio-temporal graph H can thus be folded in a non-dynamic fashion to get back to the space of the base graph. For instance, considering a dynamic pattern d with three (3) spatio-temporal edges and four (4) spatio-temporal nodes, (A_1, B_2), (B_1, A_2), (A_2, B_3), the resulting static pattern s would only contain two (2) nodes and two (2) edges: (A, B) and (B, A). Additional properties such as the number of nodes/edges folded can be stored on the static activity pattern and used later on for various analysis. The static activity pattern thus represents the total activity of the dynamic activity pattern through its time span. An example of static activity patterns is shown in FIG. 1.

(39) In the case where the static activity patterns are considered to have too many nodes to represent a coherent set of related items, it is possible to run a community detection algorithm to extract smaller and more meaningful communities of nodes. A possible implementation could include the algorithms. If done so, the corresponding dynamic activity pattern should also be partitioned in as many communities as found by the algorithm, using the mapping given by the

algorithm on the static activity pattern. Both dynamic and static patterns are kept for further analysis.

(40) Each static activity pattern is then enriched with all possible information given by the base graph G and external information that can be mapped to the nodes of G such as page, overall activity, image, audio, video, social activity, ontologies, categories, topics, geographical origin, languages, etc. In addition, information extracted from the dynamic activity patterns can also be included in the static activity pattern such as duration of activation from each as well some topological properties of the network such centrality or flow measures. At this stage, it is possible to use a clustering or classification method such as k-means, support vector machine or deep neural networks to assign patterns to a specific set of classes or clusters using a combination of features described previously.

(41) A final step of the patterns extraction may consist in creating a network of static activity patterns where each node is a pattern connected by a measure of similarity. This measure of similarity can be created by a combination of static activity patterns features or by the clustering previously described. For instance, two static activity patterns could be linked if they share a common node from the base graph G, for example at a different point in time, or if they belong in the same topic.

(42) Regarding the pattern retrieval, once static and dynamic patterns of activity are extracted, the proposed device, system, and method consists in indexing and ranking the patterns by relevance upon the user query using all or a part of their features. The query can be formulated as a combination of different words, topics, categories, communities, time-period, and activity related criteria such as the most active/popular/visited pattern for example. Different queries can be combined to filter the results more precisely.

(43) In addition to the previously described method for querying the patterns, it is also possible to use network traversal algorithms to rank and query the data. Using the graph of static activity patterns, it is possible to explore the neighborhood of a node at one or more hops or traverse the graph using shortest-path algorithms. Connected components and community detection algorithms previously described can also be used to rank relevant part of the static activity pattern network for a given query. Finally, a query can also be composed of a part of static activity pattern using subgraph isomorphism algorithms.

(44) Regarding pattern display, the last part of the method describes the user interface to browse and explore patterns extracted by the proposed method, for example a graphical user interface that is shown on a display or screen. An overview of the interface is shown in FIG. 7.

(45) The user interface is made of a different sets of panels that can be hidden or expanded on request, for example in the form of a graphical user interface that can be operated by a touchpad, mouse or other pointing device. It is composed of a search bar to input user queries. Upon click, keyboard press or when typing, a list of elements is displayed. These elements can be the name of the node, the pattern (shown in FIG. 1), the closest matching node, activity timeline of the pattern (shown in FIG. 2), the spatio-temporal graph (shown in FIG. 3), a relevance score or any information that could be extracted from static activity patterns. It is possible to also display a small picture alongside each result that visualize the static activity pattern. Another panel represents the visualization of the results in the static activity pattern network. In this interactive map, each retrieved node could be highlighted. This visualization of the static activity pattern network can be queried by the user by selecting nodes or tracing path between nodes of the network.

(46) Upon user interaction with a result, a panel representing the static activity pattern is displayed in form of a network where each node is a node from the base graph G. The results panel can be folded to a minimal or preview state to free space for the rest of the interface. All or part of the features of the pattern can be displayed on the nodes and edges of the visualized network. In addition, a panel representing the associated timeline of the event showing how nodes of the static activity patterns activates through time can be displayed. The last panel shows a detailed description of the pattern or a detailed description of a node/edge when the user interacts with it. In the case when the node represents a displayable content such as webpage, image, video, sound, etc. the content can be directly displayed in the panel.

(47) FIG. 8 shows an exemplary device and system for implementing the method described above. The system includes a data processing device 20, for example but not limited to a personal computer, Macintosh™ computer, laptop, notebook, netbook. A dataset 12 is schematically shown, that can be located locally in a storage 26 associated with processing device 2, or can be accessed via the network 40, for example the Internet, from various servers 50 and storage 60. Processing device 20 can be equipped with one or several hardware microprocessors and with internal memory. Also, processing device 20 is connected to a data input device, for example a keyboard 24 to provide for user instructions for the method, and a data display device, for example a computer screen 22, to display different stages and final results of the data processing steps of the method. For example, the graphical user interface 25 can be shown, as schematically shown in FIG. 7. Also, the base graph G, spatio-temporal graph H, nodes, edges, static and dynamic activity patterns etc. can also be represented in computer screen 22. Processing device 20 is also connected to a network 40 via a network interface, for example the Internet, to access various cloud-based and network based services, for example but

not limited to cloud or network servers 50, cloud or network data storage devices 60, specific web servers like Wikipedia, etc. The method described above can also be performed on hardware processors of one or more servers 50, and the results sent over the network 40 for rendering and display on computer screen 22 via processing device 20.

(48) Processing device 20 can be equipped with a data input/output port, for example a CDROM drive, Universal Serial Bus (USB), card readers, storage device readers, to read data, for example computer readable and executable instructions, from non-transitory computer-readable media 30, 32. Non-transitory computer-readable media 30, 32 are storage devices, for example but not limited to external hard drives, flash drives, memory cards, USB memory sticks, CDROM, Blu-Ray™ disks, optical storage devices and other types of portable memory devices that are capable of temporarily or permanently storing computer-readable instructions thereon. The computer-readable instructions can be configured to perform the method, as described above, when loaded to processing device 20 and executed on a processing device 20 or a cloud or other type of network server 50, for example the one shown in FIG. 8.

(49) FIG. 9 shows various stages in the method in an exemplary representation. First, on the left side, a base graph G is shown having various related items, represented as nodes or edges of the graph, and depicting signals that are acting on the individual nodes/edges. The connections between the nodes/edges represent a measure of similarity of the nodes/edges. Next, from the base graph G, two exemplary dynamic activity patterns 1, 2 are created by connecting nodes of the base graph according to a measure of activity from the time series of data. In the dynamic activity pattern, nodes are connected to their neighbors on the base graph G over time if they both have a sufficiently important measure of activity in two consecutive time steps. Thereafter, two exemplary static activity patterns are created by folding the dynamic activity

patterns 1, 2 according to identifiers of the nodes from the base graph G. In the last stage of the method as shown, an exemplary static activity pattern network is shown. Here, each node represents a static activity pattern and each edge encodes a measure of similarity, typically by sharing a common node of the base graph G, between static activity patterns. This network can be visualized in an interactive manner to ease the exploring and discovery of related patterns of activity.

(50) Based on the above described method and its implementation as a device or in a system, there are several novel applications, as shown with exemplary embodiments in the following paragraphs. For example, the proposed method, device and system has numerous applications in different fields. It is basically possible to use in a large variety of data and event analysis situations as long as it is possible to create a graph and extract a signal on the nodes/edges of a graph.

(51) As an exemplary embodiment, the present method, device and system can be used to recommend products, as schematically shown in the flowchart of FIG. 10. Here, a graph of similar products is created and dynamically updated according to several factors, for example the availability in stock, geographical location of the product or of similar products, availability at different shops or users as well as the weather for instance. The signal on the graph can be defined as the number of items bought per product per time unit, for example per hour, day or any other relevant time-series. The method, device and system creates activity patterns linking different products together according to what users bought in a particular context. A recommendation scheme to a user could consist in ranking the best activity patterns given some past events, including at least one of historical purchases and items already put in the cart. A more personalized layer of recommendation could be added by using clustering methods and

traversals on the network of static activity patterns to propose a smooth shopping experience, dynamically creating contextual bundles of products frequently bought together, tailored specifically to this user in this context as shown in FIG. 10.

(52) Next, as another exemplary embodiment, FIG. 11 shows a schematic flowchart for a method to detect and remove or otherwise make "fake news" unavailable, or to indicate a level of authority and accuracy of a news article, by examining the veracity of news articles. Considering a graph of web pages linked by hypertext links and a signal on the graph given by the number of visits or searches for a topic, according to another aspect, the method, device and system could output subgraphs of connected articles relevant in the current context as show in FIG. 11. Using label propagation techniques and natural language processing it becomes feasible to blacklist or approve an entire subgraph of related news depending on how trusted the websites issuing the news are. The usage of the novel user interface schematically illustrated in FIG. 7 could also dramatically ease the process of exploring, flagging, and/or approving articles.

(53) As another exemplary embodiment, the method, device and system can be used to extract trends on a corpus of documents, for example news articles or advertisement, for example as a data mining tool. The proposed method gives much richer and cleaner results than methods based on correlation as it leverages the structure of a graph of related documents to remove noise. Thanks to its scalable implementation, the method, device and system allows to analyze large corpus of documents such as a Wikipedia, part of the Web and social networks such as Facebook or Twitter. With a much finer analysis of how monitored keywords and users relate to each other through time, it possible to optimize ads and track precisely user behavior. The method can also be used as a drop-in replacement for Google Trends with a richer a better

precision on worldwide news coverage using Wikipedia as base corpus to detect trends as shown in the FIG. 1.

(54) According to yet another exemplary embodiment, the method, device and system can be used to ease user interaction with complex datasets such as a patent database or scientific publication archives. The graph of documents is created by similarity according to text features as well as metadata such as authors, jurisdiction, editor, domain, publisher, and so forth. The signal on the graph is defined as the number of clicks / visits on each document, for example via different access databases, in the patent context for example but not limited to patentscope.com, espacenet.com, patents.google.com, uspto.gov, lexisnexis, westlaw. The method, device and system outputs activity patterns encoding relevant groups of documents according to user searches. Moreover, related patents could be clustered using human intelligence instead of only relying on text analysis. In addition, these patterns could then be explored interactively on a user interface and ultimately help users finding relevant documents, as exemplarily shown in FIG 7.

(55) According to still another exemplary embodiment, the method, device and system can be used to track user cohorts on a specific website, to perform website analytics or A/B testing of websites. By collecting user clicks across sessions, it is possible to construct a spatio-temporal graph of user activity on the domain. Next, the method, device and system can then be used to cluster users and thus extract insightful analytics on user behavior. For example, the method, device and system could be used to optimizing the layout of a website according to the user in a specific context. For instance, an activity pattern can be extracted when a group of users respond to a newsletter by visiting a website and skimming through the pages in a particular fashion. For instance, an activity pattern could consist in a series of articles around a political leader followed by the results on a given election. This contextual pattern can be used to

tune the user interface e.g. a user with an activity on the website resembling this pattern will changed dynamically to improve retention rate as well as ads.

(56) The method, device and system can be used as a financial indicator to track the evolution of traded securities, such as stocks, shares, commodities or other financial products. Given a graph of products created according to their likelihood to be exported by a country using the Standard International Trade Classification (SITC) categories or another type of securities classification scheme, and product activity extracted using text-analysis and human intelligence from financial news sites, such as but not limited to Bloomberg, Reuters, Financial Times, activity patterns can be extracted. By encoding the causality between products in time in a given context from historical data, it is now possible to predict whether a shortage of a given raw material will influence the price of derived products and in which time frame as schematically shown in FIG. 12.

(57) While the invention has been disclosed with reference to certain preferred embodiments, numerous modifications, alterations, and changes to the described embodiments, and equivalents thereof, are possible without departing from the sphere and scope of the invention. Accordingly, it is intended that the invention not be limited to the described embodiments, and be given the broadest reasonable interpretation in accordance with the language of the appended claims.

CLAIMS:

1. A method to extract patterns of activity from time series of data structured as a network of objects, the method comprising the steps of:

creating a base graph of similar or related items from a given corpus by comparing intrinsic features of the items;

combining time varying data on nodes or edges of the base graph;

creating dynamic activity patterns by connecting nodes of the base graph according to a measure of activity from the time series of data; and

creating static activity patterns by folding the dynamic activity patterns according to identifiers of the nodes from the base graph.

2. The method of claim 1, further comprising the step of:

creating a static activity graph where each node is a static activity pattern linked by a measure of similarity using intrinsic features of the static activity patterns.

3. The method of claim 1, further comprising the steps of:

indexing the static activity patterns using intrinsic features of the static activity patterns;

ranking a list of matching static activity patterns by relevance upon user query composed of at least some features of the static activity patterns;

creating user queries from network traversals of the static activity patterns; and

creating user queries from results of clustering or community detection algorithms on the static activity patterns.

4. The method of claim 1, further comprising the steps of:

displaying a panel representing the static activity pattern in a display in form the base graph, upon user interaction with a result of a query.

5. A system configured to extract patterns of activity from time series of data structured as a network of objects, the system including a computer device that has access to memory, and a network interface, the computer device configured to:

create a base graph of similar or related items from a given corpus of data from the memory by comparing intrinsic features of the items;

combine time varying data on nodes or edges of the base graph;

create dynamic activity patterns by connecting nodes of the base graph according to a measure of activity from the time series of data; and

create static activity patterns by folding the dynamic activity patterns according to identifiers of the nodes from the base graph.

6. The system of claim 5, wherein the computer device is further configured to:

create a static activity graph where each node is a static activity pattern linked by a measure of similarity using intrinsic features of the static activity patterns.

7. The system of claim 5, wherein the computer device is configured to:

index the static activity patterns using intrinsic features of the static activity patterns;

rank a list of matching static activity patterns by relevance upon user query composed of at least some features of the static activity patterns;

create user queries from network traversals of the static activity patterns; and

create user queries from results of clustering or community detection algorithms on the static activity patterns.

8. The system of claim 5, wherein the computer device is configured to:

instruct a display device to display a panel representing the static activity pattern in form the base graph, upon user interaction with a result of a query.

9. A non-transitory computer readable medium, the computer readable medium having computer instructions recorded thereon, the computer instructions configured to perform a method for extracting patterns of activity from time series of data structured as a network of objects when executed on a computer having memory, the method comprising the steps of:

creating a base graph of similar or related items from a given corpus by comparing intrinsic features of the items;

combining time varying data on nodes or edges of the base graph;

creating dynamic activity patterns by connecting nodes of the base graph according to a measure of activity from the time series of data; and

creating static activity patterns by folding the dynamic activity patterns according to identifiers of the nodes from the base graph.

10. The non-transitory computer readable medium of claim 9, further comprising the step of:

creating a static activity graph where each node is a static activity pattern linked by a measure of similarity using intrinsic features of the static activity patterns.

11. The non-transitory computer readable medium of claim 9, further comprising the steps of:

indexing the static activity patterns using intrinsic features of the static activity patterns;

ranking a list of matching static activity patterns by relevance upon user query composed of at least some features of the static activity patterns;

creating user queries from network traversals of the static activity patterns; and

creating user queries from results of clustering or community detection algorithms on the static activity patterns.

12. The non-transitory computer readable medium of claim 9, further comprising the steps of:

displaying a panel representing the static activity pattern in a display in form the base graph, upon user interaction with a result of a query.

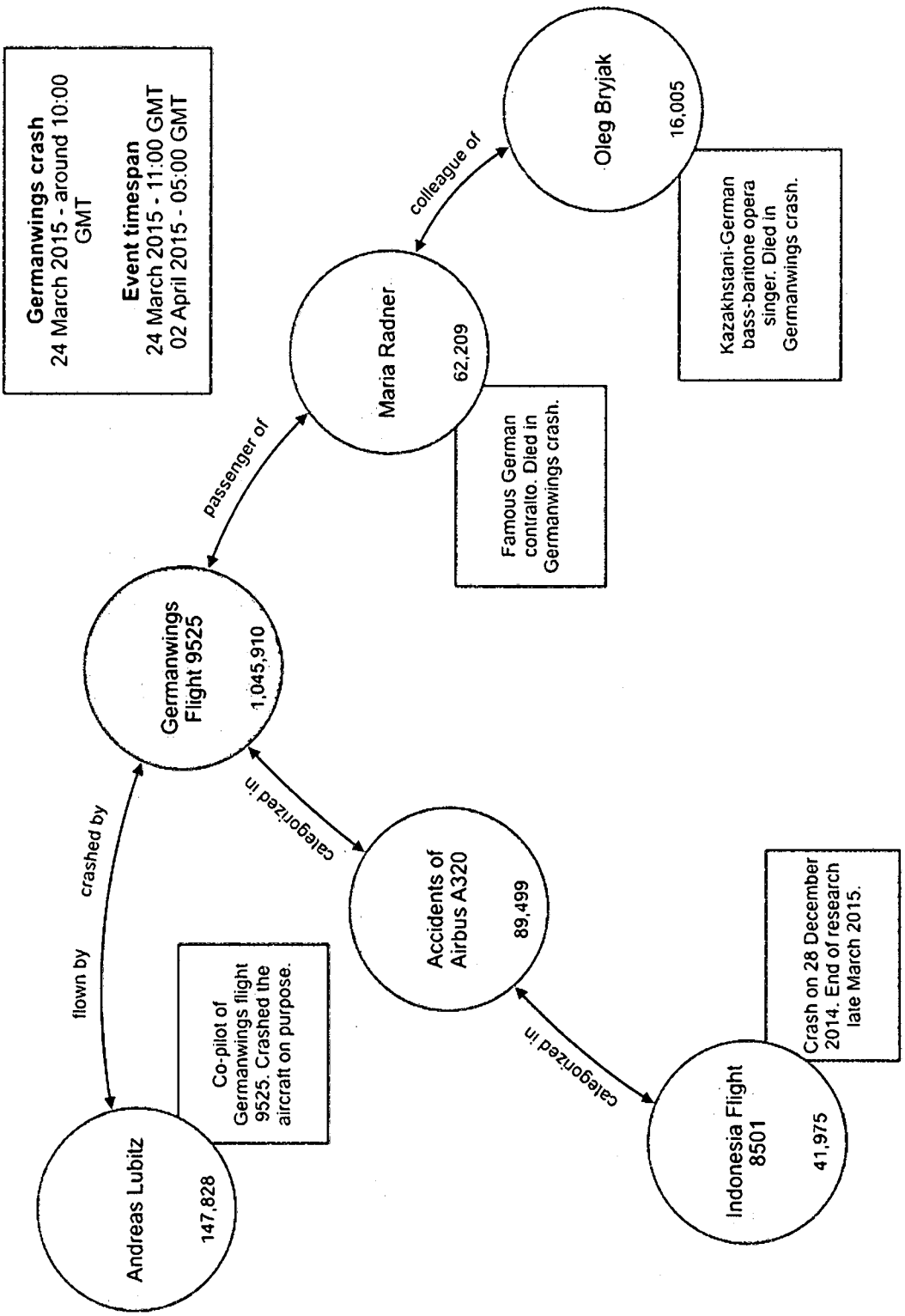
Germanwings crash
24 March 2015 - around 10:00 GMT
Event timespan
24 March 2015 - 11:00 GMT
02 April 2015 - 05:00 GMT
Oleg Bryjak
16,005
Kazakhstani-German bass-baritone opera singer. Died in Germanwings crash.
colleague of
Maria Radner
62,209
Famous German contralto. Died in Germanwings crash.
passenger of
Germanwings Flight 9525
1,045,910
categorized in
Accidents of Airbus A320
89,499
categorized in
Indonesia Flight 8501
41,975
Crash on 28 December 2014. End of research late March 2015.
crashed by
flown by
Andreas Lubitz
147,828
Co-pilot of Germanwings flight 9525. Crashed the aircraft on purpose.

FIG. 1

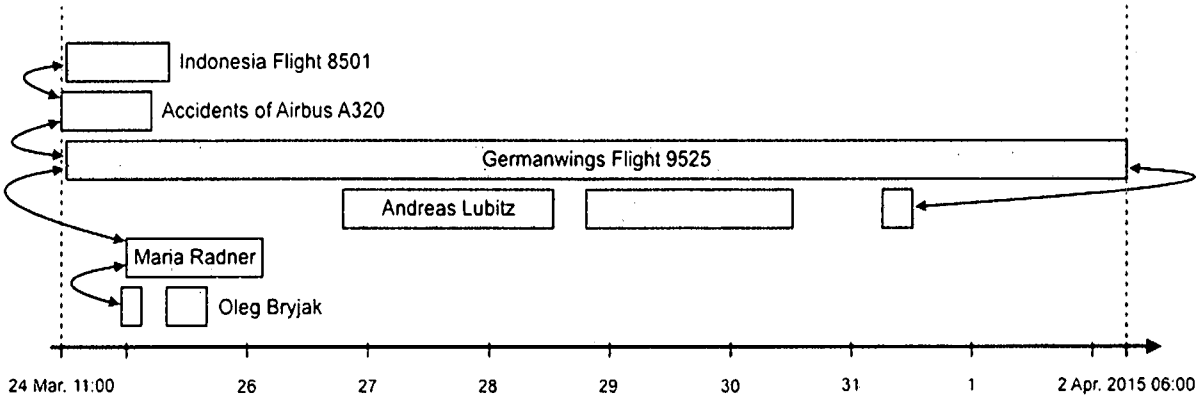
Indonesia Flight 8501
Accidents of Airbus A320
Germanwings Flight 9525
Andreas Lubitz
Maria Radner
Oleg Bryjak
24 Mar. 11:00
26
27
28
29
30
31
1
2 Apr. 2015 06:00

FIG. 2

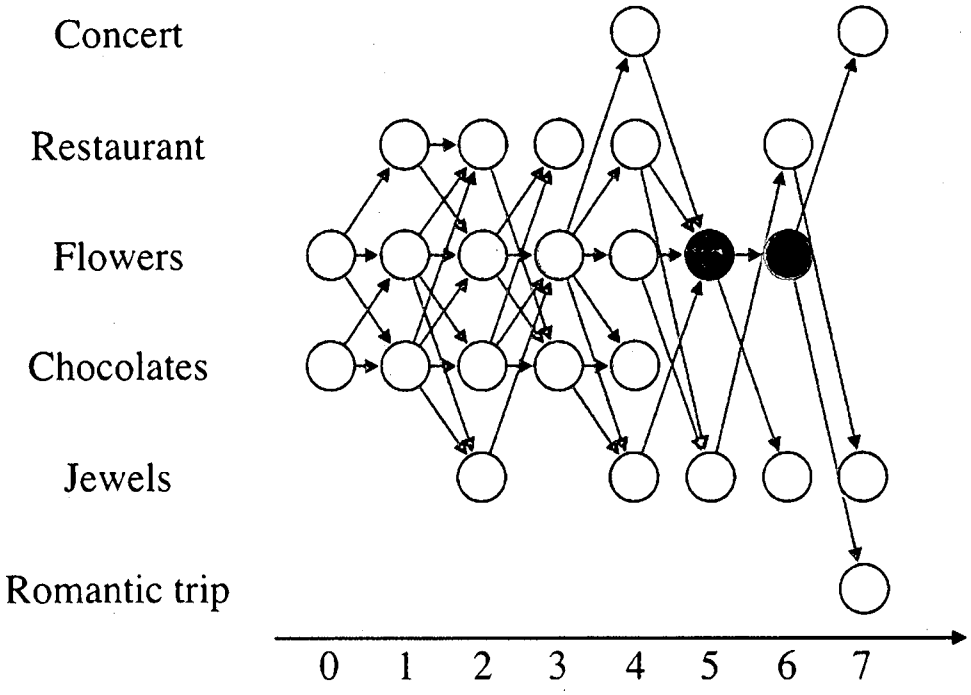
Concert
Restaurant
Flowers
Chocolates
Jewels
Romantic trip
0 1 2 3 4 5 6 7

FIG. 3

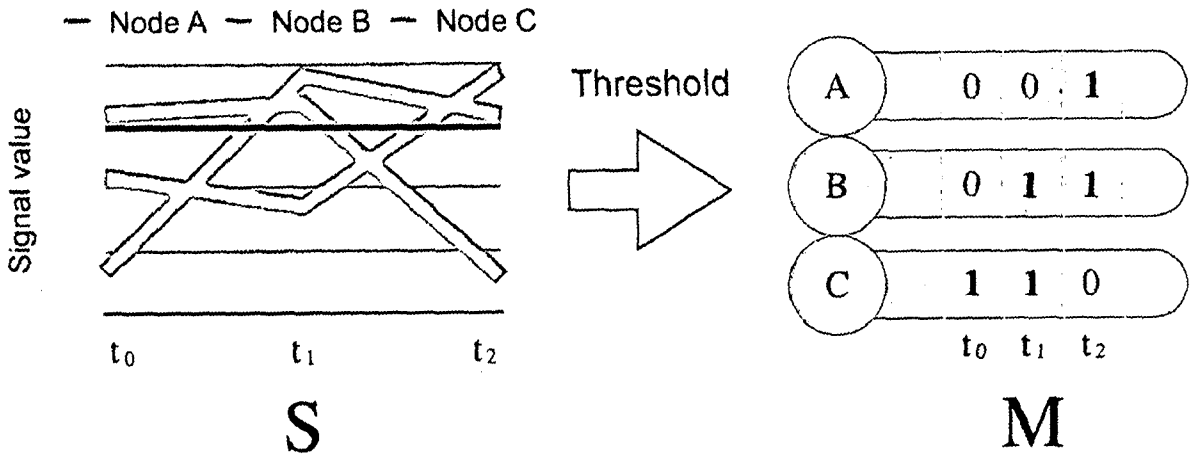
Node A
Node B
Node C
Signal value
Threshold
A 0 0 1
B 0 1 1
C 1 1 0
t0 t1 t2
t0 t1 t2
S
M

FIG. 4

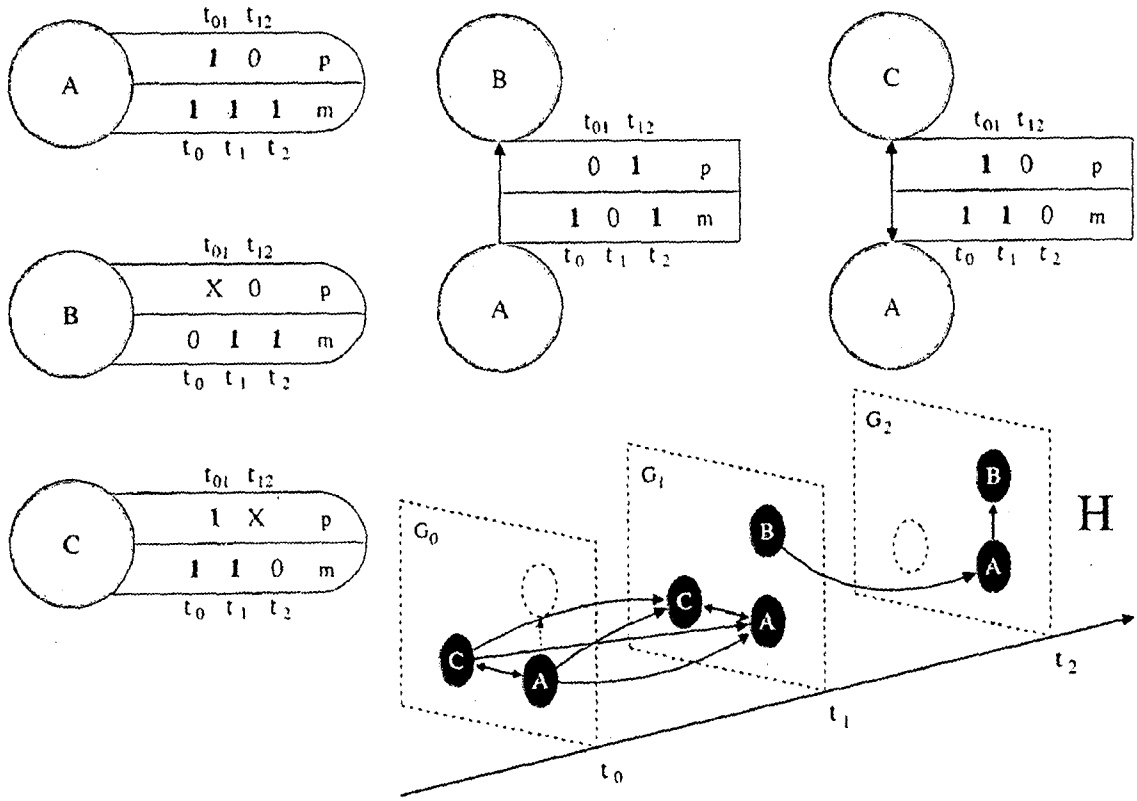
A
B
C
H
G0
G1
G2
t0
t1
t2

FIG. 5

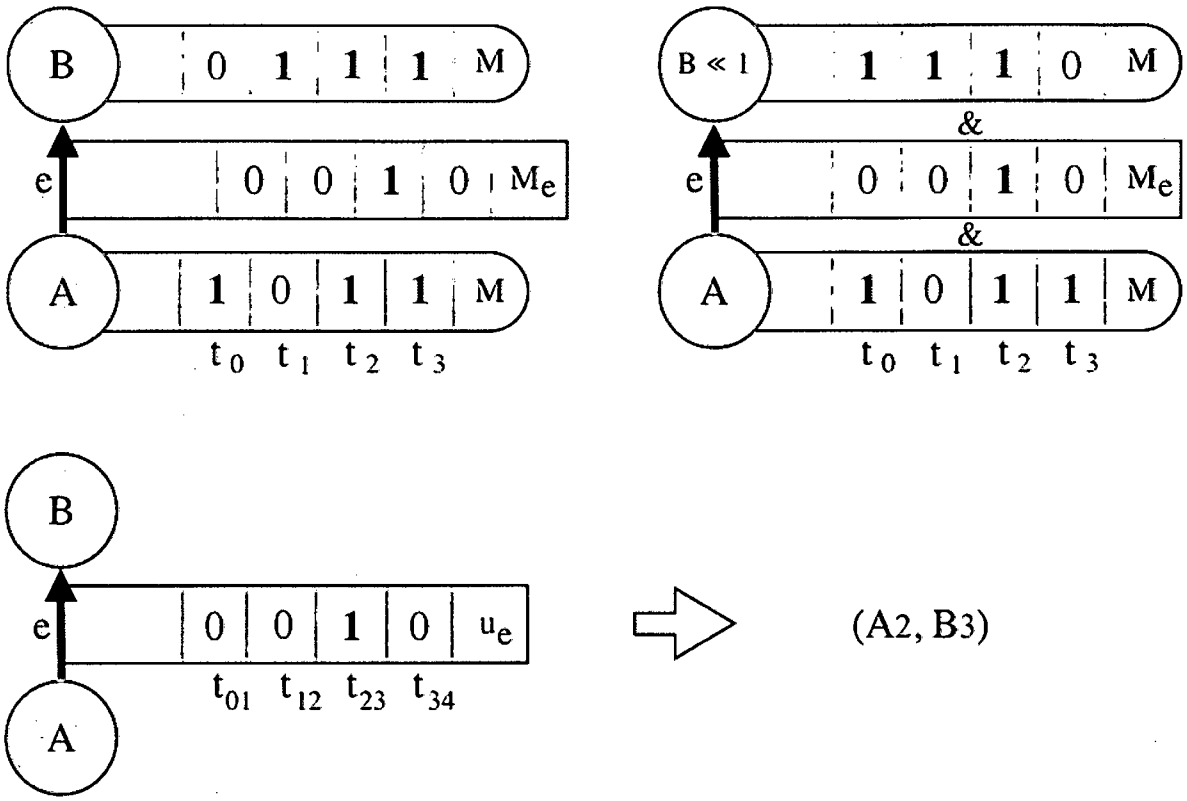
B
0 1 1 1 M
e
0 0 1 0 Me
A
1 0 1 1 M
t0 t1 t2 t3
B ≪ 1
1 1 1 0 M
&
e
0 0 1 0 Me
&
A
1 0 1 1 M
t0 t1 t2 t3
B
e
0 0 1 0 ue
t01 t12 t23 t34
A
(A2, B3)

FIG. 6

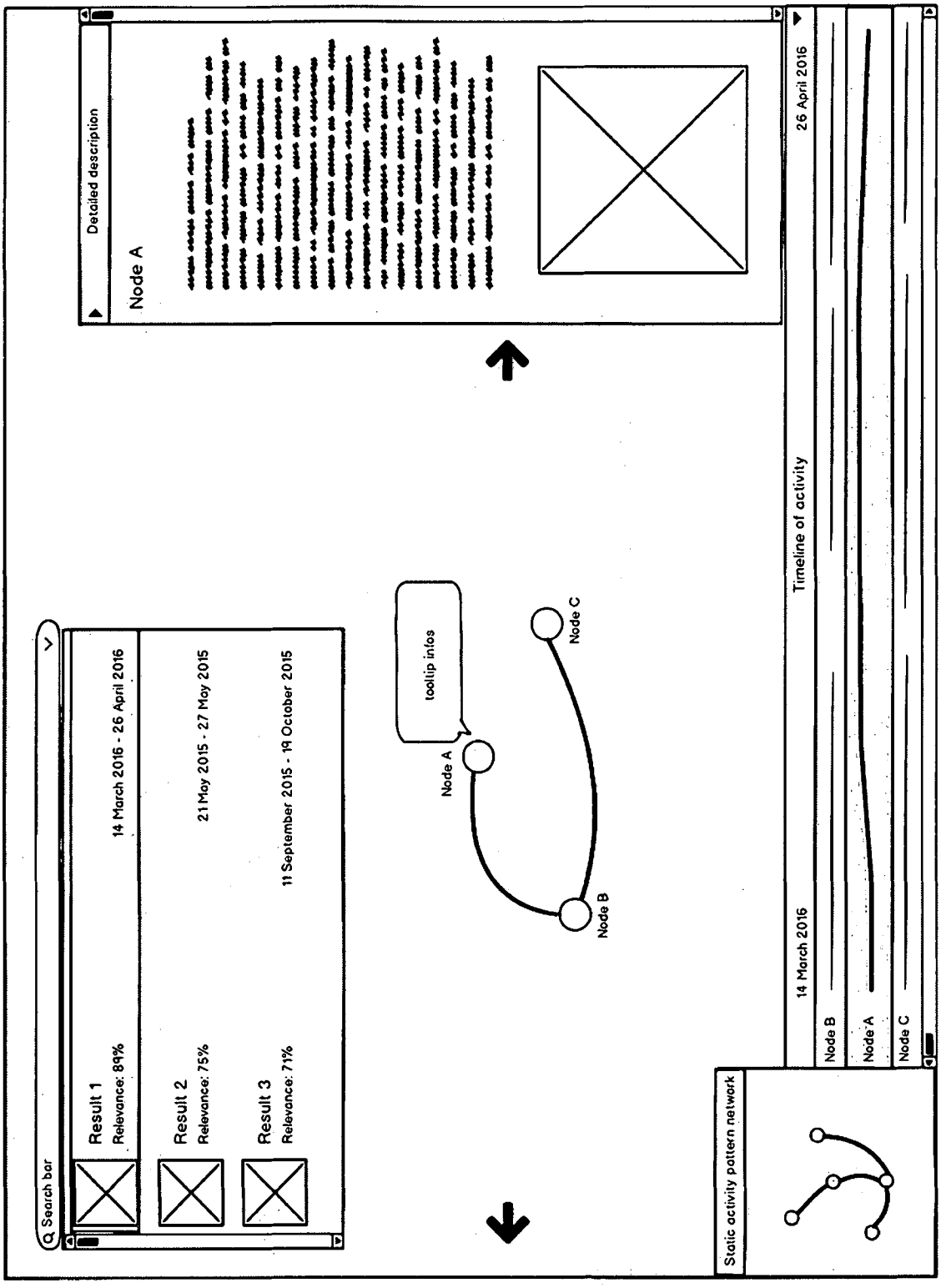
Search bar
Result 1
Relevance: 89%
14 March 2016 - 26 April 2016
Result 2
Relevance: 75%
21 May 2015 - 27 May 2015
Result 3
Relevance: 71%
11 September 2015 - 19 October 2015
Detailed description
Node A
tooltip infos
Node A
Node B
Node C
Timeline of activity
14 March 2016
26 April 2016
Node B
Node A
Node C
Static activity pattern network

FIG. 7

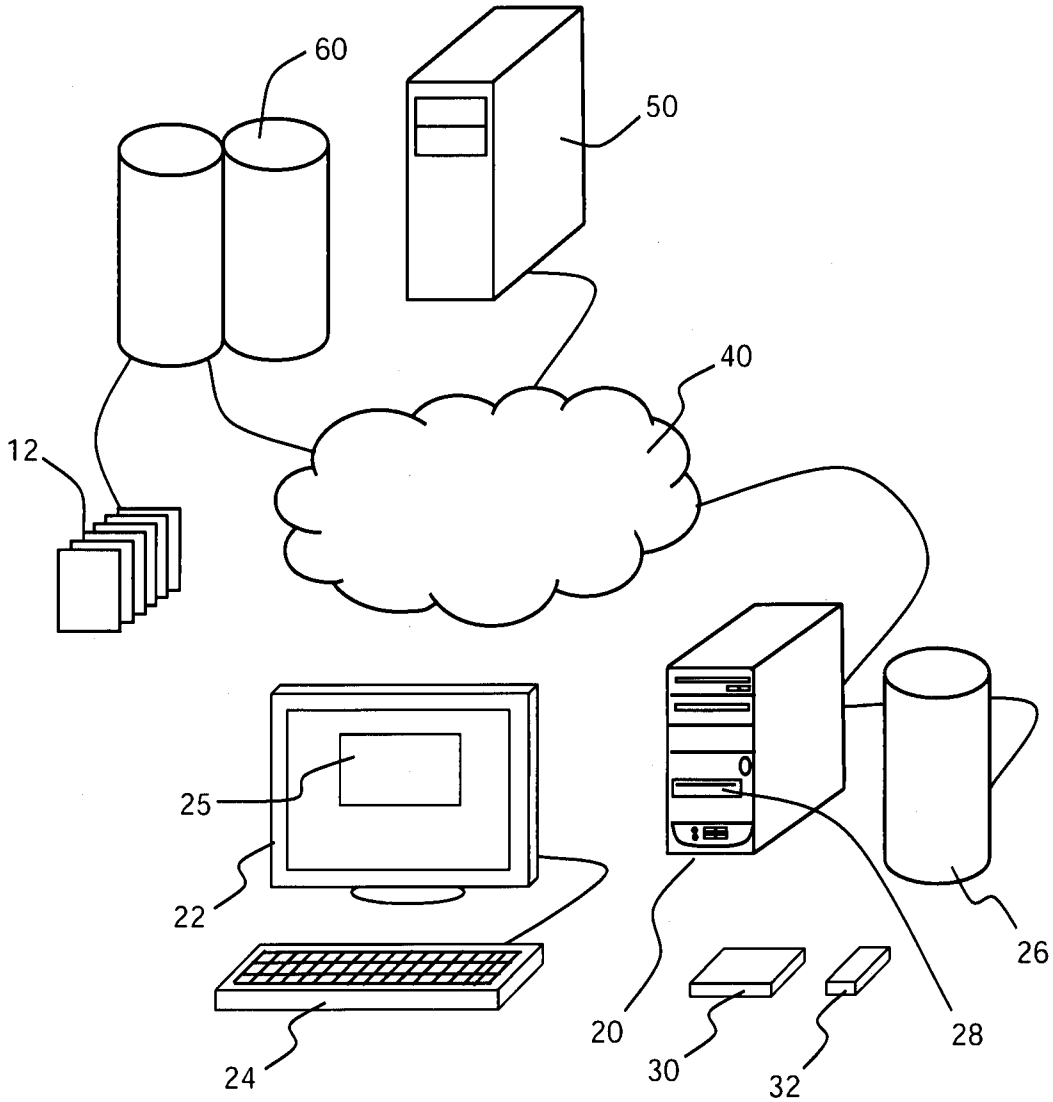

FIG. 8

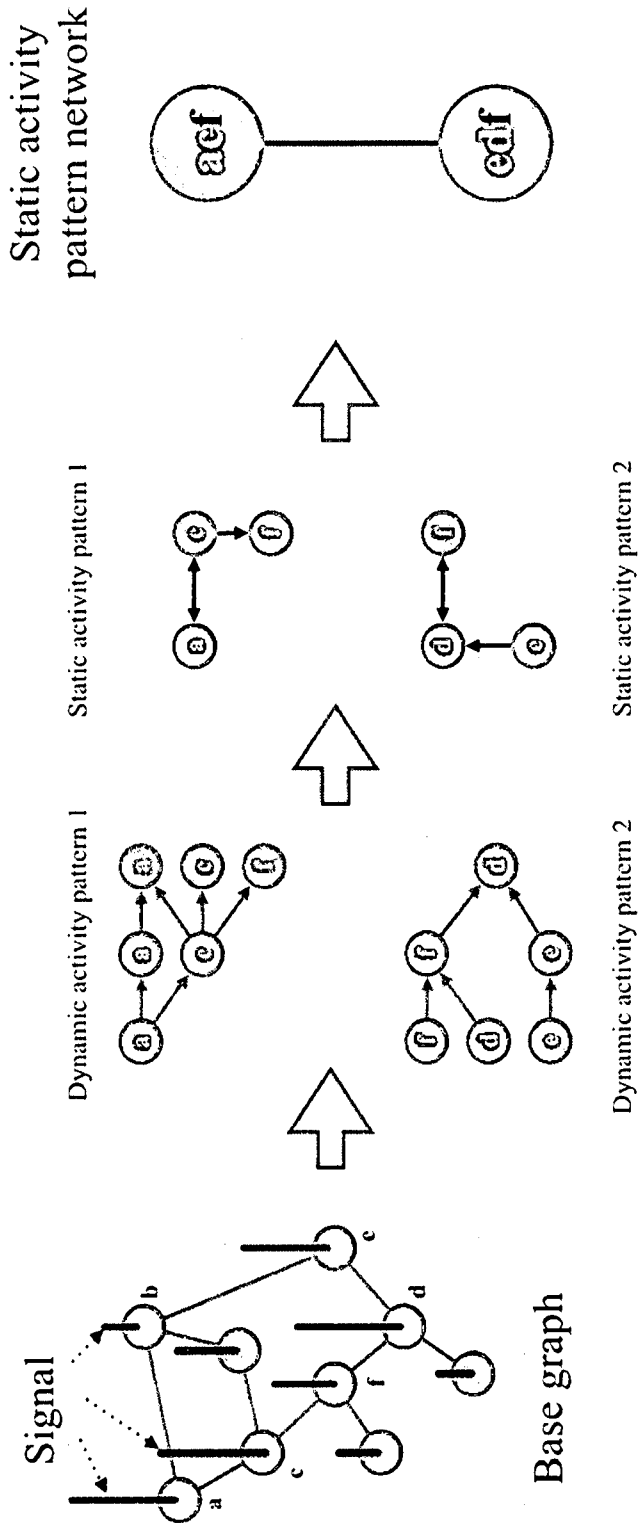
Static activity pattern network
acef
cdf
Static activity pattern 1
Static activity pattern 2
Dynamic activity pattern 1
Dynamic activity pattern 2
Signal
Base graph

FIG. 9

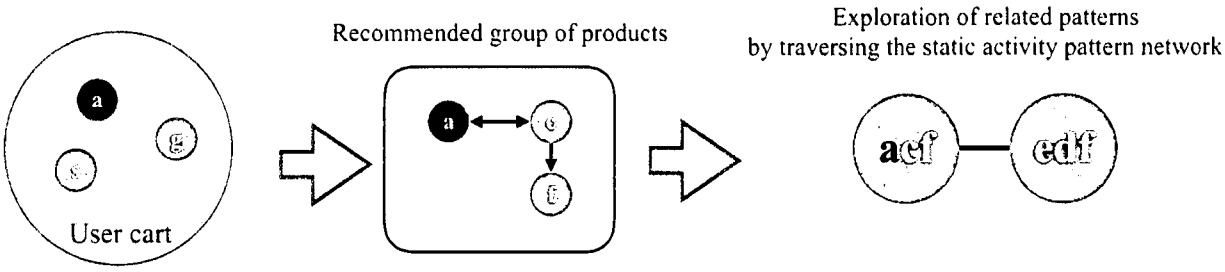
Recommended group of products
Exploration of related patterns
by traversing the static activity pattern network
User cart
acf
edf

FIG. 10

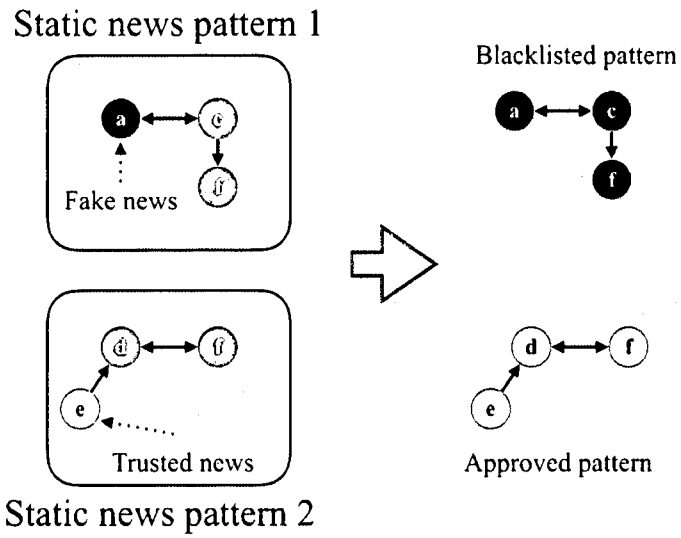
Static news pattern 1
Fake news
Blacklisted pattern
Trusted news
Approved pattern
Static news pattern 2

FIG. 11

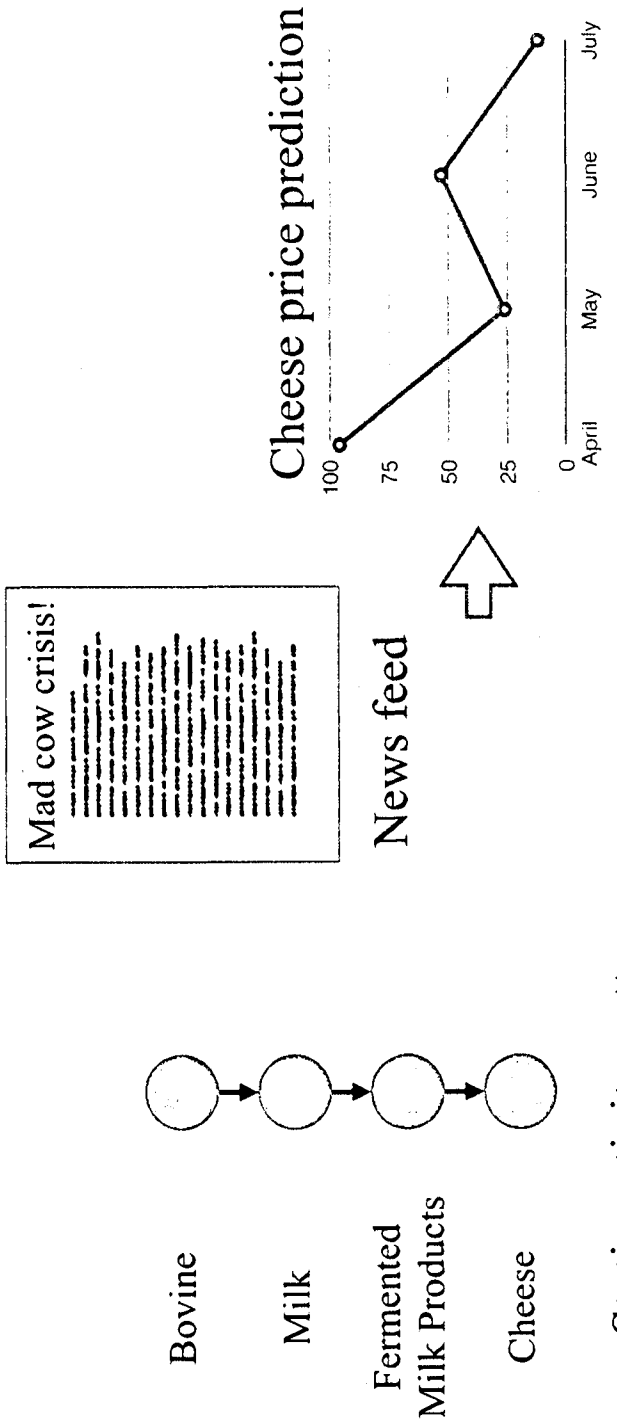
Cheese price prediction
100
75
50
25
0
April
May
June
July
Mad cow crisis!
News feed
Bovine
Milk
Fermented Milk Products
Cheese
Static activity pattern

FIG. 12

INTERNATIONAL SEARCH REPORT

International application No
PCT/IB2017/052469

A. CLASSIFICATION OF SUBJECT MATTER
INV. G06F17/30
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, COMPENDEX, INSPEC, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Kirell Benzi ET AL: "Principal Patterns on Graphs: Discovering Coherent Structures in Datasets", , 1 February 2016 (2016-02-01), pages 1-19, XP055392968, Retrieved from the Internet: URL:https://arxiv.org/pdf/1504.08153v4.pdf [retrieved on 2017-07-21] * section II; figures 1-3 * * section III; figure 4 * -----	1-12
X	US 2011/078189 A1 (BONCHI FRANCESCO [ES] ET AL) 31 March 2011 (2011-03-31) paragraph [0022] - paragraph [0025] -----	1-12

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
21 July 2017	28/07/2017

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Polzer, Andreas

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/IB2017/052469

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2011078189 A1	31-03-2011	US 2011078189 A1	31-03-2011
		US 2012246196 A1	27-09-2012
		US 2014115155 A1	24-04-2014
		US 2017053009 A1	23-02-2017