Social Life Networks: A Multimedia Problem?

Amarnath Gupta
San Diego Supercomputer Center
Univ. of California San Diego
La Jolla, USA
a1gupta@ucsd.edu

Ramesh Jain
Computer Science
Univ. of California Irvine
Irvine, CA 92617
jain@ics.uci.edu

ABSTRACT

Connecting people to the resources they need is a fundamental task for any society. We present the idea of a technology that can be used by the middle tier of a society so that it uses people's mobile devices and social networks to connect the needy with providers. We conceive of a world observatory called the Social Life Network (SLN) that connects together people and things and monitors for people's needs as their life situations evolve. To enable such a system we need SLN to register and recognize situations by combining people's activities and data streaming from personal devices and environment sensors, and based on the situations make the connections when possible. But is this a multimedia problem? We show that many pattern recognition, machine learning, sensor fusion and information retrieval techniques used in multimedia-related research are deeply connected to the SLN problem. We sketch the functional architecture of such a system and show the place for these techniques.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous; D.2.8 [**Software Engineering**]: Metrics—*complexity measures, performance measures*

General Terms

Multimedia, Situation model, Social Networks

Keywords

Situation Recognition, Event Monitoring

1. INTRODUCTION

Connecting people to the resources they need is a fundamental task for any society. Connecting patients to the appropriate health care facility, farmers to the right people in the right agricultural advice centers, disaster victims to the best shelters, are actions that every society and its societal agents (Government, Military, non-profit organizations) carry out on a regular basis. In developing countries, the needs for security, food, water, health care, education, and basic transportation are still unsolved. [9] argued that one can think of the world population as a pyramid. the large middle section of the pyramid can, with sufficient technology support, be assisted to find the connections they need to improve their lives. Their lack of current, factual information creates many problems, including ineffective resource distribution, wasted aid, and poor living quality. With the right technology infrastructure, the MOP segment of the human society can be connected to the resources they need.

We imagine a new world where all people are connected in a giant social network, where every node is a real-life entity (people, organizations, facilities, equipment, ...) and every edge is a local connection between a node pairs that are spatially, emotionally, or logically close to them. This would include their neighbors, doctors, their local grocery stores and police stations, their homes and workplaces as well as things in their homes and workplaces, their local transportation and places in their villages and cities. This social network, accessible from their simple and smart phones and computers, is always keeping up to date information about the physical entities they represent. The whole network acts as a "world observatory", and yet provides only information to people related to their "circles". We call such a world-status-bearing network a *Social Life Network* (SLN). Although such an idealized, abstract network does not exist today, we postulate that if we unite all social networks and interconnected mobile devices capturing every moment of our lives, web sites tracking our browsing behavior, electronic commerce transactions we make every day, and bring them under a single network, it is not too far-fetched to see that we are already very close to having an SLN.

The SLN is, by definition, a dynamic entity. It is designed to be always current and constitutes a model of the evolving world; it captures events and situations in the life of each person or thing and collective events and situations that affect any segment of the human society represented in the SLN. Therefore, an SLN has a fundamental difference compared to standard social networks. In an SLN, properties (and hence nodes and edges) of the graph are always evolving; therefore paths between two nodes might emerge or disappear as the model evolves. This gives rise to a generic problem that we refer to as the *SLN Problem*: "Can we connect a person node in an SLN to a resource node as the person's situation evolves and needs to be connected to the resource node?" In other words, can we connect a person to a matching resource as and when the person's situation

MM'13, October 21–25, 2013, Barcelona, Spain.

http://dx.doi.org/10.1145/2502081.2502279.

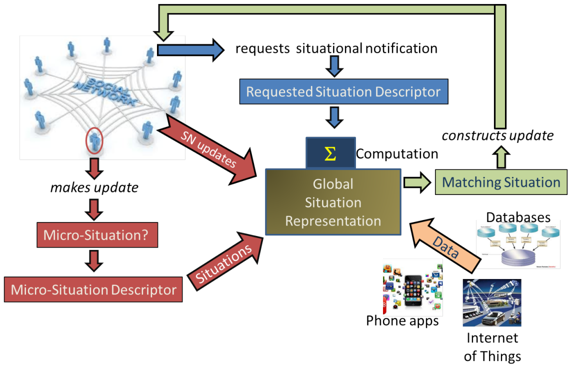

Figure 1: A schematic sketch of a Social Life Network

evolves? To our knowledge, this is a new unsolved problem for the next generation of social networks.

At this point, it would seem that this next-generation social networks problem is completely unrelated to multimedia systems. We however argue that multimedia systems should not be narrowly defined as a discipline that operates on sensory signals like audio, images and video to extract semantics. Rather, it should be viewed as a wealth of representational and computational techniques that model the real world in terms of multidimensional signals and applies some specific classes of operations on them – if any engineering problem can make effective use of these techniques, it falls within the purview of multimedia systems. In the case of the SLN problem we posit that today's social network solutions do not adequately capture the spatiotemporal dependencies of real world situations in a way that multimedia representations can and therefore, we should be able to use multimedia techniques for this problem. Based on this postulation, we intuitively hypothesize that *one can automatically derive a media-representation of the evolving properties of the SLN model, and utilize it for situation detection that enables a solution for the SLN problem by using computing techniques from multimedia systems.*

The Intuition. A Social Life Network can be pictorially represented as shown in Figure 1. It builds on a social network, defined in a broader sense, where not only people, but other objects like mobile applications, databases, and the Internet of Things (e.g., household appliances and cars) also provide information. All of these information sources send their "signals" to a separate Situation Management Module. This module monitors the updates made by the members of the social network as well as other information from the networked world, and determines if any of the updates may qualify as a potential micro-situation, based on some criteria set by the administrators of the system; if an update passes the qualification test, it gathers contextual information about the update and constructs a micro-situation object, which it sends to a global Situation Processor. On the other hand, another member posts a request for situations it would like to be notified on. This request is also converted into a desired Situation Description, and sent to the same processor. The processor matches the situation request with the already-ingested situation updates, and upon finding a successful match, constructs a suitable update message, and dispatches it to the appropriate requesters.

As we have seen in Situation Recognition Systems [25, 26], the computation preformed inside the processor can make use of spatial, temporal and semantic representations used by multimedia techniques. The primary intuition behind a multimedia representation is simple. *If real-world events can be captured using a suitable combination of human and electronic sensors, then interesting situations can be recognized as the occurrence of spatiotemporal patterns in the combined sensor-space. These pattern computations can be performed by first computing application dependent features from these combined sensors, and then applying appropriate matching techniques on these feature sets.* In the next section, we elaborate on this intuitive principle.

2. SITUATION DETECTION: MULTIMEDIA *VIS À VIS* SLN

Currently, most multimedia situation detection systems (including our own [25, 26]) go through the following generic workflow:

(a) Acquire signals from a few "media" sources like videos, image collections.

(b) Acquire signals from auxiliary sources (e.g., a GPS).

(c) Represent the signals in some novel, interesting way.

(d) Extract novel features that are geared toward object recognition, concept recognition, or action recognition.

(e) Use a classifier to group and categorize the features into some known situation categories.

Since it attempts to solve a larger problem, the SLN setting goes far beyond solving situation recognition problems in several important ways.

(i) We can conceptualize the input to SLN as *messages* – in a sense they form the "raw signals" from which the situations have to be assessed and acted upon. These messages can come from people who actively post them. However, they could also come from sensors (GPS location) that "post" on a user's behalf, from global information sources (rainfall from a weather station, regional drought statistics from a Government agency) or from software applications that transmit highly focused complex data (e.g., prescription refills from a doctor's office; heart rate, activity level from a mobile application).

(ii) The messages would often be partly structured text and semi-structured data from sensors and applications respectively, but they will also contain labeled images, videos and voice information related to messages.

(iii) The situation understanding process must extract, integrate and correlate the relevant information produced by the multiplicity of diverse and complex messages, which will come in asynchronously, following different formats and protocols. This correlation must be computed both at the level of single person's messages (micro-situations as shown in Fig. 1) and at the global situation processor that assimilates micro-situations as well as global situations to determine actions.

(iv) The SLN has to be able to handle multiple representations and sometimes joint representations that facilitates feature extraction and situation recognition. A well-studied analogy is the widespread use of both

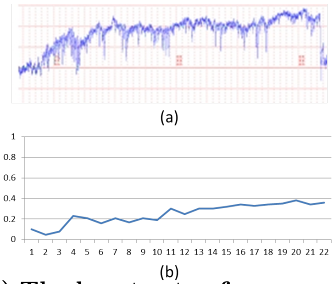

Figure 2: (a) The heart rate of a person as she types a message, (b) the emotionality measurement of the same person

time and frequency domain representations for signal processing and interpretation. These representations are useful for different problems, and hence are equally valuable. An interesting example of joint representation is a spectrograph which is a time-plot of short-term Fourier coefficients computed over small intervals.

(v) Unlike a multimedia information retrieval system, there need not be any explicit query in SLN. The evolving situation of the SLN member *is* a continuous query posed to the rest of the system. However, the processor needs to produce an action only when the combination of the member's micro-situation and the global state of the world "match" in some well-defined way.

(vi) Tasks like "matching" situations and classifying situation-related features will possibly be a combination of automated and human-powered (e.g., through crowdsourcing) because the time-to-response may sometimes be critical for actions like rescuing a flood victim.

Despite these distinctive characteristics of SLN situation processing compared to standard multimedia situation detection, they are deeply connected.

2.1 Signal Formation in SLN

In standard multimedia problems, the signals are a given – one starts with signals from the media source(s), and possibly supplements it with additional data sources. In SLN however, *signals need to be constructed* from asynchronous, multivariate, mixed-modality information coming from possibly uncoordinated source. Often the construction itself is a computational process, and depends on the application in which the situation recognition occurs. In this sense, we blur the distinction between what is called a feature in traditional multimedia system and what we are calling situation-depicting signals in our current setting. As a simple, somewhat unrealistic example, consider the time-series (a) and (b) in Figure 2. The first time-series is measured from the online activity in a network, and depicts the burstiness of activity measured an SLN member on a specific site. The second time-series is measured by a wearable device from the same member, and represents the member's heart rate. As shown in the figure, the member's site activity is frequent, and the heart rate has a moderate elevation. These two indicators, taken individually, say nothing about the member's state or situation. Taken together, they might signify a level of increasing stress in the member as she interacts at the web site, and hence may be considered as the components that can construct a "virtual", semantic *signal* by combining these two measurable signals. The difficult part is making the decision that these signals jointly constitute an event-level signal. Once the decision is made, the problem transfers to the realm of existing or emerging multimedia techniques (e.g., sensor fusion) for the actual construction process. In [21] the authors perform cooperative localization, i.e., accurately detect the actual positions of vehicles based upon the vehicular velocity and yaw-rate reported at 50 Hz, combined with GPS coming through a Global Navigation Satellite System (GNSS) reported at 4Hz. The latter signal however is prone to ambiguity because it gets impacted by the buildings around the vehicle. The technique in [21] uses a Bayes filtering technique to handle the GNSS ambiguity, and then uses raytracing from building models to eliminate undesirable GNSS signals. Coupled with the vehicle motion signals, they construct the "virtual" (fused in their terminology) positional signal. Mapping to our example, combining the message and time-series signals can be achieved if it can be shown that they are generated by the same time-dependent point process (or processes that last for a duration [10]), i.e., the probability of their co-occurrence can be predicted by a single generative model. The details of such combination modeling is beyond the scope of this "ideas" paper.

2.2 An Analysis of SLN Signals

It is evident from the prior discussion that a primary design problem for situation processing in SLN is to identify and characterize the classes of virtual signals that need to be constructed. Here, we define a starting analysis of such signals that we have identified from the current multimedia-related literature.

An SLN signal is by definition a temporal entity that depicts the existence of a complex state, entering into and exiting a state, a state transition with transition parameters (like delays), or a declaratively definable pattern over the states and transitions. The states, transitions, and patterns can be thought of in terms of the following axes.

There is a wide diversity among the observed situation-depicting variables that define SLN signals. Roughly, we can think of them as measurements, derived measurements, references, reference-based measurements, aggregates, relationships and propagations. A measurement is a direct observation (e.g., GPS coordinates) that is used to compute states (of entities, or phenomena or locations), and is often subject to stability and accuracy issues. To construct a situation signal from a measurement one often goes through a discretization or robust estimation process before it is actually used. In [17], McGovern *et al* use a SAX-based [24] discretization technique for detecting situations in two different scenarios. In the first an individual's state of motion (going up the stairs vs. down the stairs) is computed from discretized measurements of joint trajectories; in the second, the states of supercell thunderstorms are recognized using wind speed, barometric pressure and so forth. A derived measurement is computed from one or more measurements. Frequency of a single signal, the joint probability of a multivalued signal are classical derived measurements used for situation detection. A reference is the mention of one entity (or event) in another entity. Attaching a picture to a message, citing an event in a tweet are examples of references. The entity (or event) in question can either be the concept of the entity (or event), or a concrete instance thereof. For example, a bag of words based feature description associated with a concept name is a reference to a concept; on the other hand, referring to a specific labeled event (e.g., a much discussed violent crime) is an example of the latter. From a multimedia viewpoint, this mention and its temporal behavior can be viewed as a situation-depicting signal. As a simple example, just counting the number of occurrences of "Boston Marathon" (an instance of a `marathon race` concept) in messages and pictures over time starting April 2013 will determine that an evolving situation has occurred about this event. In a more complex case, [23] have developed a traffic situation detection method, where a vehicle specifies its own state (e.g., stopped at an intersection) by referencing the other entities and events in its vicinity (e.g., entity:car-in-front c, event:c "suddenly-stopped"). Reference-based measurements are variables derived from references. Continuing the traffic metaphor, a time-plot of the number of cars in a highway segment (a contextual locational variable) that entered a `stopped` state because the `car-in-front` had a `suddenly-stopped` event designates the onset of a traffic jam, and perhaps an accident situation. The last two examples also show the use of aggregates in the construction of a situation-depicting signal. But we would like to specially mention aggregates because they are always parameterized over a collection mechanism. In the Eventshop system [25, 26] the aggregate structure comes in the form of a spatial grid that stores the aggregate of messages based on a theme at every time-point. The situation detection occurs by specifying a tree of algebraic operations that compute patterns on the spatial grid. A relationship conforms to the notion of a "relational feature" in multimedia systems; with this category of "signals", a situation emerges because of the interaction between multiple-parties, where a party can be any combination of entities, locations and events. The traffic situation example above can be viewed as a rudimentary relationship. A more suitable case has been developed in [2] where multi-party situations are analyzed by classifying the state of each person with an SVM, computing the motion of a person through sensors, and the distances between people as a proxy for interaction (they do not consider non-physical interaction like conversation or electronic communication). These virtual signals are fused to detect situations like "introduction" or "game playing". Finally, propagation refers to signals that arise due a *flow* over a structure. Two well-researched examples are optical flow [19], and network flow that are widely used in literature as models of an evolving situation. As a concrete example, a road segment network is a network that is laid over a spatial grid, and traffic flow, measured from vehicle location and velocity as well as from inductive loops can be measured over segments of the network. The situation-depicting flow signal is computed using an origin-destination flow model in this case by combining road-segment data (i.e. data from the edges of the network) with spatiotemporally acquired data through Emage operations [25]. The propagation is then measured by observing the variation of the flow over the structure of the network, i.e., as a time series at every node of the road network.

In the beginning of this section, we noted the importance of combining multiple "raw signals" to create situation-depicting signals. Most of the efforts described in the foregoing paragraphs and systems like them make use of one or more combination technique. We describe some of them.

The EventShop system [5] achieves this integration in two stages. In the first stage a *stel*, i.e., a spatiotemporal element that aggregates the count of some physical variable (collected from one or more data sources) is created on a single semantic theme at a single time point. More generally, however, it can use any other aggregation function that constructs the stel-value. A spatial grid (an array) of stels constructs what is called an *Emage*. A sequence of spatial grids (i.e., a time-series of arrays with application-defined spatial and temporal resolutions) constitute an Emage-sequence for a single variable. In stage 2, when multiple variables need to be combined, the system lets the user define a tree of algebraic operations that create a new Emage sequence from existing Emage sequences. Emutem, a healthcare application described in [18] aims to combine signals from microphones, physiological parameter sensors (e.g., pulse rate) and infrared occupation sensors to detect situations like fall and heart attack. Since all signals are continuous, they use a Gaussian mixture model to specify each type of signal and then use a fuzzy information combination model with Gaussian membership functions. At a more symbolic level, [19] creates an activity recognition system for healthcare – it uses an SVM to classify the individual activity signals and then apply a rule-based information combination method on the results of classification. In a smart home application, [3] uses an ontology-driven rule engine to real-time, continuous recognition of household activities like making coffee and watching TV. Finally, [8] constructs "logical sensors" by combining the outputs of hardware sensors using subjective logic, which is a modification of the well-known Dempster-Shafer model for belief propagation.

We note that the foregoing discussion is not about how the SLN should operate, but to establish the intuition that many of the current representation and analysis machinery used by the multimedia, pattern recognition and pervasive computing communities can be viewed through the glasses of situation processing. Next, we turn our attention to the actual information processing in an SLN as we see it.

3. SITUATION PROCESSING IN ACTION

3.1 Three Use Cases

We present three uses cases with distinct foci to develop the mechanics of situation processing within the SLN system. In the first two cases, assume that the SLN works atop a social network S which has individual, group and system type members. For every member, the SLN provides a special "friend" called the *Situation Box*, which can be treated just as another member of the network – when desired, the user will post messages to the Situation Box, tag it in pictures, other entities and events, send voice messages to it, and respond to any information that is posted by the Situation Box. The Situation Box, which is run by the SLN, can also provide messages to the human member. If the member so desires, she can direct her mobile device applications to post information to this situation box on her behalf. An application like Moves (`http://www.moves-app.com/`) could, for instance, auto-post the location and activity on behalf of the member if desired. Whenever the Situation Box receives any user input, it is registered by the SLN as a user event. For non-human members, however, every posting made is considered to be a contributing event. For the third case, assume that the SLN works on the Internet of Things and vehicular ad-hoc networks (VANET).

Elderly Person in a Disaster. Sue, an elderly member of the SLN, lives by herself and has difficulty walking, a complaint she has often made to the Situation Box. Sue has environmental allergies from multiple allergens including dust and mold, and often suffers from allergy-induced asthma. Whenever she suffers an allergy attack, usually after cleaning her house, she complains to the Situation Box, but usually feels better in a day or two. At the time of our interest, Sue is aware of an impending hurricane predicted to sweep over her town. In her repeated and increasingly anxious complaints to the Situation Box, she says she is concerned about her house flooding. She also expresses deep concern that she will have severe mold allergy, like she had five years ago, if her house gets flooded. The SLN should assess her situation and connect her with family, or an evacuation team or emergency medication services to help her.

Insect Infestation in India. Dr. Thomas Reinhardt, an agricultural entomologist in Berlin studies south Asian insects, especially insects that infest agricultural plants. His maintains two circles on this SLN, personal and professional. The professional circle consists of his colleagues, and a special entity called "Problems", where he expects to gather interesting real-world problems related to his area of interest. He links his "Problems" node to the Situation Box, which gathers situations as requested by people who post "Desired Situations" on the board, with the expectation that the SLN will find *matching situations* and respond to the posts. Far away, in a few states of Northwest India, farmers are experiencing an unexplained problem with the crop they are growing – the leaves are discoloring and the plants are dying. They have started using their phones to send updates to their Situation boxes. These updates include pictures and steps they have already taken. The SLN is expected to have Dr. Reinhardt be interested in the case, and perhaps contact the Government to offer his expertise.

Situation-aware Vehicular Network. An non-human SLN is built on top of a VANET, where the members are cars, traffic sensors on freeways, including inductive loops, periodic reports from emergency roadside services, and mobile traffic applications like Waze (`www.waze.com`). For this SLN, information updaters are smart cars whose positions on a road segment, velocities, accelerations are known at every k seconds, and dynamic relationships are formed between a car and other cars within a d distance. Let us assume every car wants a notification if it is heading toward an impending traffic jam or accident situation (i.e., this "query" is made by every car) based on status updates by its neighbors and then wants to connect to a car at the nearest exit to enquire about the local traffic situation. The SLN also has access to traffic volume and speed estimates, historical average and peak traffic for any time of the day, and weather reports. The expectation from the SLN is that it will update potentially affected cars within a reasonable time so that they can advise the drivers of a different route.

3.2 Overall Functionality of the SLN

Let us recapitulate the primary situation processing tasks in a social life network through our use cases.

(a) For every registered member of the SLN, it must *record and monitor the events generated by the member* based on (i) information offered by the member and (ii) the output on any devices and applications that monitor the member's state. These user events must be transformed to situation-depicting signals and must be interpreted in context to derive what we call a *micro-situation* in Figure 1. In the elderly person's case, the SLN must derive a progressive sequence of micro-situations. In the case of the Indian farmers, the micro-situations sent by every farmer can be gathered into a *macro-situation*, a data structure to hold an aggregate view of the total crisis. As we will see, the techniques of situation-depicting signal generation in Sec. 2.2 will be used here.

(b) The SLN also records, maintains and monitors the *interest and capability* declared by any member that can play the role of a resource provider. The SLN converts this declaration to a **match target**, which is notionally like a continuous query plan in a data stream management system, but with an important distinction. Provider capabilities and interests are not static, define-once structures. They evolve over time, sometimes based on dynamically changing situations, and sometimes just with time. For the elderly lady, the nearest rescue center may have to change its plans as it gets filled up as more people seek shelter; but the shelter five miles away has just expanded its capacity.

(c) It assimilates, computes and monitors the *state of the world* (i.e., the world within the SLN's purview), by collecting contextual information sent by physical, social, Internet etc. from a large body of external information resources (e.g., weather, news, police advisories). However, the information from these resources can show wide variability in their timeliness, frequency of update, accuracy, coverage and granularity. If the Vehicular Network gets the information that the police has just closed an exit 2 miles down the freeway, it must adjust its computation of traffic jam likelihood. Most of the information combination techniques from Section 2.2 apply here.

(d) It must *match the micro and macro situations to the declared interests and capabilities* of the resource providers. Since both these factors evolve with time, the matching process must be time aware. Techniques for situation matching will use algorithms from database, information retrieval and multimedia communities. Further, since the quality of match strongly depends on the contextual data like location of the situation, conditions under which the situations occur and any other correlated data will play a significant role in the matching process.

(e) Finally, once the match has occurred, *the result of the match must be delivered back* to the members whose situations can be addressed – it is desirable that this delivery of result is always accompanied by a concrete action. In the case on Indian Farmers, the SLN will have to construct semi-structured multimedia summary packet of the evolving state of the ailing crops (and their environment) that is comprehensive enough for Dr. Reinhardt.

These five components form the basic building blocks of the SLN architecture. The next section discusses the operations of these components in more detail.

4. FUNCTIONAL ARCHITECTURE OF SITUATION PROCESSING IN SLN

The members in an SLN can have three roles – they can be solution seekers (SS), solution providers (SP) or information providers (IP). A single member can serve in any of the three roles, but may have only one role at a time. Any transaction with the SLN (e.g., through the Situation Box) must be identified as one of these roles. In the following we refer to member actions with respect to their roles.

4.1 Formation of Micro and Macro-Situations

To create a micro-situation from an SS member's member events, the SLN uses a basic situation constructor and a number of specialized situation constructor modules that work cooperatively. These constructors operate on the messages and related information, and creates/updates one or more *situation description structures* (SDS), data structures that are shared by all situation handling modules.

The Analyzer. The activity analyzer takes as input a member activity and analyzes it for augmentation. The most common activity is sending messages. A text message `M` has an optional `[object, [tag]]` portion that depicts a list of objects (e.g., pictures) that can be posted and the list of tags that can be associated with each object. The tag may be assigned by the user or maybe assigned by an automatic or human algorithm. In the Indian farmers case, people can attach the tag "severe discoloration" to the picture of the wheat plan leaves. `M` is always analyzed for content and references. Every content analysis function assigns one or more *scores* to the message, which are then attached to the message. For example, entity and emotion detection algorithms can be applied to `M`, producing a referred entity list, and a list of {`emotion, strength`} pairs. The entity (e.g., "flood water") will be recognized using a relevant ontology and the emotions (e.g., `[(fear, 0.6), (sadness, 0.2)]`) will be recognized using techniques described in [14, 1]. Specifically, techniques like [14] relates the emotional (e.g., fear) to the causing entity (e.g., mold-allergy). An activity like adding a description to a picture is treated like a new message with the picture as the reference. An entity (e.g., a voice recording) with no textual tag will only be algorithmically analyzed.

SDS Objects. For every member the SLN Analyzer creates an $ECPS$(*event context parameter space*) instance, a multidimensional structure that is updated for every member activity. Based upon prior research that analyzes temporal statements about states and events [11] we can view the event e (e.g., posting a high-fear message, attaching pictures of one's body with the label "rash") in terms of the ECPS:

- $TIM(e)$, the mean time-of-occurrence of e
- $LOC(e)$, the location of e, if available
- $DUR(e)$, the duration of e
- $CAT(e)$, the category(-ies) of e based on some ontology
- $ENT(e)$, the set of entities related to e
- $EVT(e)$, events occurring within $\pm\Delta T$ of e
- $AFR(e)$, the average frequency if e occurs repeatedly
- $AGE(e)$, the length of time since the first event like e

The parameter $CAT(e)$ is assigned by a classification algorithm (e.g., [13, 29, 15, 16]) that analyzes `M` and labels it with a known term from an ontology. The labeling is crucial because depending on the event category, it can be sent to a special handler that performs deeper situational analysis of the event. For example, medical-events (e.g., allergy and asthma in our example) are dispatched to a medical micro-situation generator (discussed next). In general $CAT(e)$ is multi-valued because more than one category can be assigned to e. The emotional classification of messages, for instance, can be a second category associated with e. Parameters $TIM(e), LOC(e), DUR(e), AGE(e)$ should be viewed at multiple levels of granularity. The frequency of occurrence should be represented by the minute, hour, day, the time of the day (e.g., evening), and so forth because the event pattern for constructing the reference model can be found at any level of granularity. Parameters $ENT(e)$ and $EVT(e)$ are computed only from the content of the augmented message `M` and its history for the current member, and not on any global situations.

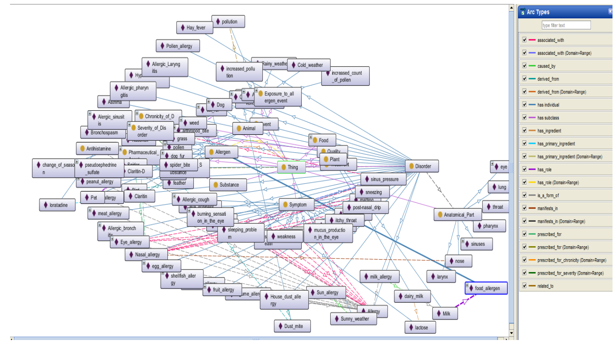

Figure 3: A portion of the Allergy Asthma Ontology

When the categories are more specific, the micro-situation detector can produce a much richer situation description using the relevant ontologies. Figure 3 shows a portion of the allergy-asthma ontology. This ontology contains the knowledge diseases and their classifications, symptoms, drugs for different forms of allergies, allergens that are responsible for different symptoms and so forth. When Sue's messages are passed to this disease aware situation handler, it now navigate the ontology to elicit additional information by posting messages back to Sue's Situation Box – it can ask whether Sue has the requisite medication to last long enough, extract the response and place it in the micro-situation it constructs for Sue.

A secondary SDS object used by the SLN is an *evolving subspace cluster tracker (ESCT)*. Based on the $ECPS$, we compute subspace clusters (assuming a finite set of event and entity categories) over the above parameters for a finite duration of time, followed by a cluster transition tracking technique. Subspace clustering techniques have been heavily used in signal processing [28], multimedia research [4, 22]. Cluster transition analysis has been used in multimedia applications ([7] for audio signal segmentation), and data mining [27, 12], where cluster transitions are categorized into external transitions (new cluster, cluster survival, cluster split, cluster absorption, cluster death) and internal transitions (no change, size change, compactness change, location change). If Sue posts a high-strung emotional message most Thursday mornings complaining about an allergy attack with a general reference to the housecleaning event, and reports a relief message by Friday afternoons, we will soon see a steady cluster in the (allergy-ThursdayMorning-housecleaning) subspace, and another steady cluster (relief-FridayPostMorning-AllergyMedication) subspace. The steady occurrence of these two time-delayed clusters will form a *ref-*

erence pattern for some of Sue's posting behavior. However, during the flood situation we will witness high-strung fearful postings that are more random in time, but more clustered in entities `flood, water, mold` – this shift in the clustering structure will be reflected through a set of *interestingness metrics* that compare clusters. Cluster survival is measured by testing if the two cluster means in the same region but at two different time points differ by less than half standard deviation. To detect compactness transitions, we use the difference of the standard deviations of the clusters X, Y. Cluster absorption is measured by absorption ratio, the portion of clusters of time t_i absorbed by clusters of time t_{i+1}. Cluster location shift is measured by a *skewness* metric [20]:

$$\gamma(X) = \frac{\frac{1}{|X|}.\Sigma_{x \in X}(x - \mu(X))^3}{\frac{1}{|X|}.(\Sigma_{x \in X}(x - \mu(X))^2)^{3/2}}$$

Micro-situation constructor. The concrete construction process of a micro-situation is performed as follows. For every SDS touched by a user activity (more generally, a set of user activities), there is a scoring mechanism that evaluates the *importance score* of the current activity. Different SDS objects and activity type combinations compute the score in different ways. For the cluster tracker SDS, the importance score can be computed by creating a union of conjunctive rules over the different measures of interestingness described above. Thus one can state that the importance score grows linearly if either the external or the internal transition score exceeds the threshold of 0.5, and is the higher of the two scores beyond 0.5. When the importance score is beyond a second threshold, the SDS objects are passed on to the Situation Matching module.

Macro-situation constructor. A macro-situation is processed the same way as a micro-situation except that in this case the system tracks SDS objects like ECPS and ECST for multiple members instead of individual members. But macro-situations often have a spatial variability that is captured by location-aware SDS object. We use Emage, the grid-like structure used in Eventshop [5, 25] as explained in Section 2.2 as a spatial SDS object. Each Emage is associated with the collection of messages from which it was derived, and is therefore connected to other SDS objects that are related to the same set of member-events.

4.2 Construction of a Match Target

Let us now look at Situation Processing from the SP-members' standpoint. In the simplest case, an SP-member uses a simple user interface to specify the nature of service or interest she can provide. Depending on the specified service, the SLN uses an ontology to acquire more information about the service. If a disaster shelter provides an accommodation service to 50 people, the SLN will simply ask for the location of the shelter, and create a variable to count the remaining capacity of the shelter at any point in time. The location information will be used to match potential victims using a distance measure between a victim and the shelter.

In a more complex case, the SLN offers a formal notification language to specify more complex conditions the situations of interest must satisfy. The full language is out of the scope of this paper, but as a simple SLN specification, Dr. Reinhardt's interest on recent south Asian insects infesting agricultural plants can be declared formally as follows.

```
WITH myInterest m                                (1)
WHERE m.topic IN (instance*(O,insect-category-
1))                                              (2)
NOTIFY                                           (3)
WHEN                                             (4)
OCCURS SITUATION s AND                           (5)
s.ISABOUT m AND                                  (6)
S.occursDuring([Now-6,*],month) AND              (7)
S.severity > 0.6                                 (8)
```

Lines 1-2 specify a variable `m` of built-in type `myInterest` whose `topic` attribute can be any transitively derived instance of a specific insect category in an ontology `O`. The `NOTIFY` clause states that the provider should receive the notification. The notification condition uses the built-in expression `OCCURS SITUATION` and assigns the situation a variable `s`. The rest of the declaration specifies properties of the desired situation. Line 6 uses the special property `ISABOUT` to depict the approximate theme of the situation. The difference between `topic` and theme is that the former is a narrow subject associated with an SLN node, while the latter is associated with the aggregated topics of a collection of nodes, and hence has a broader scope. Line 7 uses the system-defined property `occursDuring` to specify the timespan of situation `s` and states that it should be at most 6 months old. Finally, line 8 uses a property `severity` between 0 and 1 and is set to be at least 0.6 in this declaration. The `severity` condition, as we will see, will get converted to the importance score discussed in Section 4.1. The declaration does not use any spatial condition about the situation because it can come from any place.

To convert this declaration into a representation that the situation processor may handle, it needs to be transformed into an automata structure that sits on top of SDS objects and monitors their changing states. We sketch the basic rationale of transformation process without giving the complete algorithm. 1) The ontological condition in the `WITH` clause gets unfolded, and the possible values of `m.topic` gets enumerated; in other words, `m.topic` is replaced by a list of k insects. 2) Since "topic" is known to be an entity reference, and entities are handled by the ESCT objects, the SLN creates a cluster tracking object for each insect so that it can track how cluster properties change with time. Further, it notes that the severity (in fact, an importance score) threshold should be OR-ed over the k clusters, i.e., the condition is satisfied if *any* of the k cluster severity values exceeds 0.6. Since this condition is only on a single entity, it should only use the intrinsic cluster transitions. Specifically, a set of rules will be used to identify that cluster size transition is the only interestingness metric that can be used here. Thus the conditions `s.ISABOUT m` and `s.severity > 0.6` transforms into `importanceScore(insect1.cluster.sizeTransition(deltaT)) > 0.6 OR ... (`k` times)`, where the `deltaT` parameter specifies the time period over which the size transition is observed, and can be nominally calculated as a fraction of the total period of observation (6 months in our case). Thus, if the fraction is 10%, the transition should be observed over some 18 days. A simple implementation of the `importanceScore` function can be to interpret it as % growth, thus a 60% growth over 18 days will satisfy the condition. 3) All user activities contributing to the qualify cluster elements are chosen (i.e., it gets the activity IDs of every user activity that satisfied condition in the previous step). 4) For these activities, get all other SDS objects (e.g., the Emage) (recall that a situation is physically represented as several different SDS objects). 5) Determine which of these SDS objects have a high importanceScore. 6) Create a SituationObject with information collected from the previous step.

The steps above represent a recipe for detecting the *candidate situation(s)* that Dr. Reinhardt maybe interested in. They still need to be evaluated and scored for their recency and overall severity as described in Section 4.4.

The above discussion should not give the impression that setting up the match target is not dependent on the global situation processor. Many times the interest specification of a provider would be dependent an evolving situation that is unrelated to member activities. If the disaster shelter loses power or accessibility, the interest specification might become useless. This information cannot be determined if only user activities are explored. In general, the match target should also inspect the *state of the world*.

4.3 Assimilation of World-State

The "world" in SLN refers to the physical reality represented through a collection of potentially real-time information sources that relevant for an application domain. These sources act as IP-members or are created virtually, by aggregating information from SS-members. Different weather stations can be IP-members that stream a vector of data every minute. On the other hand, for the Vehicular Network, one can construct an "average traffic speed" source that is computed from all cars for each road segment. Since they are not individual specific, they are modeled along the fundamental axes of any situation, namely time, location, thematic metadata, and optionally, any intrinsic observation structure over which the application domain can be interpreted. In many applications, the intrinsic observation structure may be space itself as in the case of [25]. But in our Vehicular Network use case, the road network is a more logical choice to interpret some of the situation dynamics. To allow for this, the SLN framework must admit multiple representations to serve different purposes within the same application, so long as they are in synchrony with one another. The framework also allows multi-resolution information in space and time so that different IP-members can provide information at different resolutions. It is the SLN system's task to keep them aligned with each other, and provide roll-up aggregation, and interpolation operations on the data so that any downstream operator (like convolution) can view all data at the same resolution.

Since there is large variety of information in an SLN network, it must allow multiple models of information assimilation depending on the data. In the case of the Vehicular Network, the likelihood of congestion will depend on external factors like the weather, and it makes sense to develop an estimate of the weather impact $w(r)$ on an arbitrary road segment r. If the weather information is provided by multiple weather stations, then $w(r)$ can be computed as an estimation problem using techniques like Kalman filtering combined with a spatial process like inverse distance weighting or Thessian polygon averaging. On the other hand, if each car also had an environment sensor and could measure rainfall, then $w(r)$ could be computed by a simpler aggregation model. Now consider police reports as an world event source that provides events like accidents, road blocks, police giving a ticket on the roadside and so on. To estimate its impact factor $p(r)$ on road segment r, we have to model it using a network propagation model that will predict (perhaps using machine learning) the spread of congestion probabilities for each event type and then aggregate all such probabilities using a suitable fusion mechanism like a Bayesian network. The takeaway point is that the assimilation model in SLN will never have a single solution, but must be crafted for each application using the same library of techniques predominantly used by the multimedia community. In our own research, we have designed Eventshop [5], where we have adopted a "design studio" approach to developing world assimilation models. Here, the designer will chooses representation models from a model library, data manipulation operators from an operator library, spatial and temporal process models from a process library and so forth, and compose them into an assimilation model to test it on sample data we collect.

Once the appropriate assimilation models are constructed, the SLN needs to combine it with both the micro (and macro) situations, as well as the interest and capability models that demonstrate situation dependency. For the Elderly Person use case, this means that the model of hurricane's effect, including estimates for accessibility, water damage, likelihood of mold damage, and so forth must be computed for Sue's house. If other specific parameters are needed for these computations, they must be obtained, often by direct communication with Sue through her Situation Box. For the shelter provider in the same hurricane situation, the situation model will be used to compute a number of *viability measures* for the provider's offer. If a long power outage occurs, the viability measure can use prior data from the current and similar providers to determine how long they have been operational without power, and then predict a time-based estimate of how long the shelter would be operational. In general, a vector of such situation-influence measures will be added to the interest profile of the predictor, and will be exercised upon a situation match.

4.4 Situation Matching

The fundamental task in Situation Matching is some combination of the following: deciding which providers' offerings and interests is of maximal use to a specific member (Elderly Person case), deciding which member-presented situations should be packaged for a provider to give him maximal situation information (Indian Farmer case) deciding what situational summary is most useful to a specific member (Vehicular Network case). All these tasks need some form of matching, which is a central capability of many different kinds of systems. In classical information retrieval (including multimedia), matching corresponds to a similarity computation between a query object and a feature object based on multiple criteria, and then a rank computation that aggregates the similarity scores into a total score. In many recommendation systems a match corresponds to a combination of two rank computations – content-based ranking based on item similarity, and collaborative filtering that ranks items used by users with similar profiles. Some modern search engines treat matching as a multi-stage decision process instead of a straightforward similarity based ranking computation. Situation matching in an SLN is an open problem where lessons can be learnt from several research communities. In this paper, we offer some general principles for the problem.

Severity Criteria. Contrary to retrieval and recommendation systems, a unique characteristic of the situation matching is that the member situations that need to find match targets may come with one or more severity criteria that imposes an urgency (i.e., a priority or a time-bound) condition on the match finding process. The severity measures need to be computed based on component-level measures. In Section 4.2, the provider's query uses the `importanceScore` function to compute one measure of severity based on the rate of growing references to the insects affecting the crop. In general, the estimated level of distress expressed in the micro-situations, the spatial extent of the macro-situations, the prior statistics of comparable infestations from the world-state assimilation models, will each compute a severity score. If the situation matching process is implemented as a decision process, it may compute the composite severity score by an evidence accumulation model.

Multi-Match. It is important to recognize that a single situation may satisfy multiple providers. Sometimes these providers will be in competition (i.e., only one shelter is needed for member), and sometimes they can be independently satisfied. For the Indian Farmers, a bank can provide a low-interest loan to the infestation victims independently of Dr. Reinhardt's possible suggestion of the correct plant and soil treatment. This introduces the issue of multi-query optimization because both "queries" can share several situation processing steps to gain overall efficiency. For example, both the bank and the entomologist will utilize the severity metric of estimated distress level and the spatial coverage estimate of the infestation problem. Therefore all computations leading up to these two estimates can be reused. Such reuse is vital to maintain the performance of SLN if it must be applied on an Internet scale. Although the general multi-query optimization is a known NP-complete problem, the SLN system is likely to provide enough contextual constraints on the problem to make it more tractable.

Hybrid Matching. It is now established that for many hard and time-critical information management tasks, completely automated solutions are less practical and less optimal than mixed mode operations that involve more user interaction and crowdsourcing to people. We expect that many situation matching problems need to balance automated matching with human-in-the-loop methods. The interesting questions are when the SLN should decide to consult a human, whom should it consult, and how it would use human-provided input to complete the matching task. In the Indian Farmer case, there is an image interpretation step that determines that discolored and disfigured portions of the leaf images need tagging (to identify potential insect names). So, it is reasonable to have automated techniques to classify images that need to be human tagged, and crowdsource the images or highlighted image regions for actual tagging. Since there might be uncertainty in the user responses, this crowdsourcing task would need to send the correct number of sample images to the correct number of people and aggregate results using mixture of experts model.

4.5 Result Formation and Delivery

Let us revisit the expected result when member situations and providers' offers and interests match in our use cases. For the Elderly Person, the provider(s) must get a member status report and the member needs to get a confirmation that the provider will contact her. For the Indian Farmers, the providers get a comprehensive summary of the situation without necessarily getting the details of any member's status. For the Vehicular Network case, every car gets a "personalized" report that contains a summary of the nearby congestion situation sorted by the its current distance from the congestion point. The common task here is to construct a *situation summary*. As in any summarization task, the objective is to preserve the overall thread of events, and provide more detail on only the salient fragments of the situations ordered in time. Further, whenever an important subevent of the situation occurs, it must be instantiate with sample "cases" which is a composition of "raw" content like messages, images, and voice recordings, but also their situational interpretations defined in terms of the important changes observed in the SDS objects. In the Indian Farmers case, the sudden jump in the photograph submission along with the most dominant increase in the spatial expansion of the infestation situation will be considered a salient situational feature – thus, plots of cluster sizes and before/after snapshots of spatial coverage at these salient time intervals should be part of the summary. Recently [30] have developed an integer programming based approach for video summarization that might be generalized for situation summarization. Comparable approaches have been used for other summarization tasks for news, scholastic text, and other forms of multimedia. We believe this is a particularly suitable subproblem to benefit from multimedia techniques.

5. CONCLUSION

The primary goal of this "ideas paper" is to present a complex, socially relevant problem to the multimedia community rather than offer a definite solution. We show different aspects of the problem and establish that it requires an amalgamation of techniques rather than a fixed approach; further, we identify and address several 'multimedia' related problems required to build emerging systems like the SLN. We have tried to highlight that the information representation methods, feature extraction and comparison algorithms, object and event recognition techniques, learning principles, retrieval strategies, semantic approaches and so forth that are currently developed by the community, can be effectively used for many parts of this problem. However, they must be extended and re-evaluated to meet the timeliness and appropriateness criteria for specific SLN applications.

Testing the idea. We have started creating an experimental testbed for the situation specification and matching problem. In this testbed, microsituations are generated using mobile apps, and world situations are monitored using Eventshop. However, the system is not yet connected to a messaging platform such as the SMSim [6]. This, alongwith scalability testing, is part of our future work.

The Morality Question. An obvious concern with the proposed system is its "big brother" flavor. There are legitimate security and privacy concerns in sending members' personal situations to a central matching machinery, and the potential threat that same situation matching machinery can be used for illegal profiling and immoral exploitation. We recognize the importance this issue and the need to engage in discussions on it, but consciously leave the topic outside the scope of this introductory paper. However, this should not limit other more restricted applications of the same idea. In a personal finance application, tracking the current and expected spending pattern of a family (playing the role of a society) the system may be able to connect a family member to a more reasonable store for buying a certain product. Here, while the state of the world (store prices) is assimilated and used, the information remains "local" to the family. Similarly, in an academic application, a new researcher (defined through a research and interest profile) can be helped by connecting her with the right professionals when attending a conference.

Finally, we thank the reviewers for their insightful comments, which have improved the paper significantly.

6. REFERENCES

[1] A. Balahur, J. M. Hermida, and A. Montoyo. Building and exploiting emotinet, a knowledge base for emotion detection based on the appraisal theory model. *IEEE Trans. on Affective Computing*, 3(1):88–101, 2012.

[2] O. Brdiczka, J. L. Crowley, and P. Reignier. Learning situation models in a smart home. *IEEE Trans. Sys. Man Cyber. Part B*, 39(1):56–63, Feb. 2009.

[3] L. Chen, C. D. Nugent, and H. Wang. A knowledge-driven approach to activity recognition in smart homes. *IEEE Trans. Knowl. Data Eng.*, 24(6):961–974, 2012.

[4] E. Elhamifar and R. Vidal. Sparse subspace clustering. In *Proc. of the IEEE Conf. on Comp. Vision and Pat. Recog.(CVPR) 2009.*, pages 2790–2797, 2009.

[5] M. Gao, V. K. Singh, and R. Jain. Eventshop: from heterogeneous web streams to personalized situation detection and control. In *Proc. of the 3rd Ann. ACM Web Science Conf.*, pages 105–108, 2012.

[6] M. Gatti, A. P. Appel, C. Pinhanez, C. dos Santos, D. Gribel, P. Cavalin, and S. B. Neto. Large-scale multi-agent-based modeling and simulation of microblogging-based online social network. In *Proc. of Int. Work. on Multi-Agent-based Simul.*, May 2013.

[7] M. M. Goodwin and J. Laroche. Audio segmentation by feature-space clustering using linear discriminant analysis and dynamic programming. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pages 131–134, 2003.

[8] G. Groh, C. Fuchs, and A. Lehmann. Combining evidence for social situation detection. In *Proc. of 3rd IEEE Conf. on Social Comput.*, pages 742–747, 2011.

[9] R. Jain and D. Sonnen. Social life networks. *IT Professional*, 13(5):8–11, 2011.

[10] M. Karsai, K. Kaski, A.-L. Barabási, and J. Kertész. Universal features of correlated bursty behaviour. *Scientific reports*, 2, 2012.

[11] Z. Kozareva and E. H. Hovy. Learning temporal information for states and events. In *Proc. of the 5th IEEE Int. Conf. on Semantic Computing (ICSC)*, pages 424–429, 2011.

[12] C. Lauschke and E. Ntoutsi. Monitoring user evolution in twitter. In *IEEE/ACM Int. Conf. on Advances in Social Networks Analysis and Mining*, pages 972–977, 2012.

[13] K. Lee, D. Palsetia, R. Narayanan, M. M. A. Patwary, A. Agrawal, and A. N. Choudhary. Twitter trending topic classification. In *IEEE Data Mining Workshops at ICDM*, pages 251–258, 2011.

[14] S. Y. M. Lee, Y. Chen, and C.-R. Huang. A text-driven rule-based system for emotion cause detection. In *Proc. of the NAACL HLT Workshop on Comput. Approaches to Analysis and Generation of Emotion in Text*, pages 45–53, 2010.

[15] C. Li, A. Sun, and A. Datta. Twevent: segment-based event detection from tweets. In *Proc. of the 21st ACM Int. Conf. on Information and Knowledge Management, CIKM*, pages 155–164, 2012.

[16] H. Liu, B. Luo, and D. Lee. Location type classification using tweet content. In *11th Int. Conf. on Machine Learning and Applications (ICMLA)*, pages 232–237, 2012.

[17] A. Mcgovern, D. H. Rosendahl, R. A. Brown, and K. K. Droegemeier. Identifying predictive multi-dimensional time series motifs: an application to severe weather prediction. *Data Min. Knowl. Discov.*, 22(1-2):232–258, Jan. 2011.

[18] H. Medjahed. *Distress Situation Identification by Multimodal Data Fusion for Home Healthcare Telemonitoring*. PhD thesis, Institute National des Telecommunications, Jan 2010.

[19] Y. Nam, S. Rho, and C. Lee. Physical activity recognition using multiple sensors embedded in a wearable device. *ACM Trans. Embed. Comput. Syst.*, 12(2):26:1–26:14, Feb. 2013.

[20] I. Ntoutsi, M. Spiliopoulou, and Y. Theodoridis. Tracing cluster transitions for different cluster types. *Control and Cybernetics*, 38(1):239–259, 2009.

[21] M. Obst, S. Bauer, P. Reisdorf, and G. Wanielik. Multipath detection with 3d digital maps for robust multi-constellation gnss/ins vehicle localization in urban areas. In *IEEE Intelligent Vehicles Symposium (IV)*, pages 184–190, 2012.

[22] G. Petkos, S. Papadopoulos, and Y. Kompatsiaris. Social event detection using multimodal clustering and integrating supervisory signals. In *Proc. of 2nd ACM Int. Conf. on Multimed. Retriev.*, page 23, 2012.

[23] M. Platho, H.-M. Groß, and J. Eggert. Traffic situation assessment by recognizing interrelated road users. In *Proc. of the 15th Int. IEEE Conf. on Intell. Transp. Systems*, pages 1339–1344, 2012.

[24] J. Shieh and E. Keogh. iSAX: indexing and mining terabyte sized time series. In *Proc. of the 14th ACM SIGKDD Int. Conf. on Knowl. Discovery and Data Mining*, pages 623–631, 2008.

[25] V. K. Singh, M. Gao, and R. Jain. Social pixels: genesis and evaluation. In *Proc. of the ACM Int. Conf. on Multimedia*, pages 481–490, 2010.

[26] V. K. Singh, M. Gao, and R. Jain. Situation recognition: an evolving problem for heterogeneous dynamic big multimedia data. In *Proc. of the 20th ACM Multimedia Conf.*, pages 1209–1218, Nov. 2012.

[27] M. Spiliopoulou, I. Ntoutsi, Y. Theodoridis, and R. Schult. Monic: modeling and monitoring cluster transitions. In *Proc. of the 12th ACM Int. Conf. on Knowledge discovery and data mining (SIGKDDD)*, pages 706–711, 2006.

[28] R. Vidal. Subspace clustering. *IEEE Signal Processing Magazine*, 28(2):52–68, 2011.

[29] S. R. Yerva, Z. Miklós, and K. Aberer. Entity-based classification of twitter messages. *IJCSA*, 9(1):88–115, 2012.

[30] Z. Yuan, T. Lu, D. Wu, Y. Huang, and H. Yu. Video summarization with semantic concept preservation. In *Proc. of the 10th Int. Conf. on Mobile and Ubiquitous Multimedia*, pages 109–112, 2011.