18th International Conference on Information Fusion
Washington, DC - July 6-9, 2015

Hidden Relationships Discovery through High-Level Information Fusion

Claire Laudy
Thales Research & Technology
Palaiseau, France
claire.laudy@thalesgroup.com

***Abstract*—We previously proposed an approach and a framework for high level information fusion that we extended in order to manage uncertain information. The fusion algorithm is graph-based and generic and can be parametrized in order to provide different fusion operations. In this paper, we use uncertain graph-based fusion in order to discover social knowledge from a network of information. We aim at discovering hidden relationships between persons within a social network community from a collection of heterogeneous data sets. We apply our approach on data sets provided by the 2014 VAST Challenge. Data is available regarding the working relationships between persons, but the private relationships that may exist between the actors of the scenario are not explicit. Using semantic information fusion, we show that clues about these private relationships can be discovered, and we give information about the certainty associated to the discovered knowledge.**

I. Introduction

We previously proposed an approach and a framework for high level information fusion [1]. The information fusion algorithm embedded in the framework is a graph-based generic fusion algorithm that can be parametrized so to provide different fusion operations. Furthermore, we extended the approach in order to manage uncertain information [2].

Our aim in this paper is twofold. On the one hand, this works evaluates and validates our theoretical results on uncertain high level information fusion. On the other hand, we propose a method based on the semantic fusion of high level and heterogeneous information that aims at discovering hidden relationships between persons within a social network community. To do so, we use the graph based information fusion operations defined previously, namely *information synthesis*, *information fusion* and *information query*, in order to discover new semantic knowledge in a collection of heterogeneous data sets. The case study is based on data provided by the 2014 VAST Challenge. Data is available regarding the working relationships between persons, but the private relationships that may exist between the actors of the scenario are not explicit. Using semantic information fusion, we show that clues about these private relationships can be discover, and we give information about the certainty associated to the discovered knowledge.

This paper is organized as follows. Section 1 is dedicated to the study of related work. In section 2, we recall our theoretical results on high level information fusion. Section 3 is dedicated to the case study, which aims at discovering relationships between persons from a collection of heterogeneous data sets. The discovery of such relationships enables building a social network containing these persons.

II. Related Works

Social media have become of major importance in people's lives. A lot of activities occur and/or are reported on the internet through Social Media. Following this trend, research on Social Media and Social Networks is more and more advanced and diversified.

One of the trends of the research on social network analysis concerns the search for communities. [3], [4], [5], [6] and [7] for instance, insist on the importance of such a task in domains as varied as criminal forensics, advertisement, scientific collaboration and entertainment.

As the work presented in [8], our work relies on previous results on graph based high level information fusion. The authors present an application that enables users to query a social network for specific patterns of information. The process relies on the TrusT graph matching algorithm previously presented in [9].

Within the area of community mining, the work presented in [10] selects and combines relations coming from different networks with different types of relations. The aim is to discover answers to specific questions and this way, discover hidden communities. The work relies on an optimization approach and a linear regression algorithm. Contrary to this approach, our work makes a big use of semantics that is embedded in the high level information.

[11] focuses on the importance of taking the strength of relationships into account within Social Network Analysis (SNA) and Social Media communities analysis. The authors use an unsupervised learning process to assess the strength of friendship relations in a social network, from persons interactions and user similarities.

[6] aims at finding the strongest relations in a social network. To do so, the authors rely on a comparison of the neighbors of the nodes of the network. They use a link prediction algorithm, also described in [12].

As in the work presented in [6], we attach importance to the strength of the relations within a social network. However, our aim differs as we do not aim at selecting the strongest relations, but discovering all the relations and their associated strength, relying on uncertain testimonies of interaction in real

978-0-9964527-1-7©2015 ISIF

life. The two approaches are complementary, in a aim of SNA for forensics, for instance.

In [3], the authors present a work on mining missing or invisible relations in a social network. They rely on a game theoretic framework to explain the formation of communities and then explore the potential missing relations in a use centric process relying on the loyalty of the users towards the communities.

With the aim of identifying criminals from groups of individuals, [5] presents work on the detection of hidden relationships and activities in social networks. One of the interesting ideas is that the authors take into account the existence of different social networks and different types of relations in the social networks. They discover hidden relations among persons by using a Relation Extraction process and a regression based algorithm that associates objects of the social network.

In [13], the authors describe the use of information contained in different social networks in order to feed information to a main network. Information is gathered from users' daily life. The aim is very similar to ours, except that the author extract user interest rather than relationships. It is however very complementary in an objective of using heterogeneous insights in order to discover hidden relationships.

III. High-level Information Fusion Functions

A. Graph based Information Representation

Graph based representations appear to be naturally well adapted to soft data. Our approach relies on the use of bipartite graphs, more specifically a subset of the conceptual graphs ([14], [15]) to represent soft data and knowledge. The conceptual graphs formalism is a model that encompasses a basic ontology (called *vocabulary*), graph structures and operations on the graphs. The vocabulary defines the different types of concepts and relations that exist in the modeled application domain, while the graphs provide a representation of the observations which are provided by the information sources.

Basic **conceptual graphs** are bipartite graphs containing concept and relation nodes. Figure 1 gives an example of a conceptual graph. The rectangular boxes represent concept nodes and the ovals represent relation nodes.

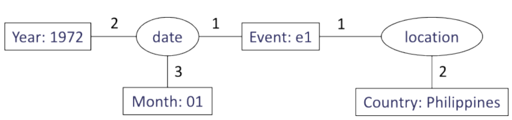

Fig. 1. Example of a conceptual graph

The term **concept** is used to refer to a concept node. The concepts represent the "things" or entities that exist. A concept is labeled with two components: the conceptual type and the individual marker.

The **conceptual type** defines the category to which the entity belongs. For instance, in Figure 1 the concept `[Country:Philippines]` is an instance of the category `Country`, i.e., its conceptual type is `Country`.

The **individual marker** relates a concept to a specific object of the world. The object represented by `[Country:Philippines]` has the name (or value) `Philippines`. The individual markers may also be undefined. An undefined or generic individual marker is either blank or noted with a star $\star$, if the individual object referred to is unknown.

The term **relation** is used to refer to a relation node. The relation nodes of a conceptual graph indicate the relations that hold between the different entities of the situation that is represented. Each relation node is labeled with a relation type that points out the kind of relation that is represented.

The notion of **vocabulary** was defined in [15]. The concept types and the conceptual relation types, which are used to label the concept and relation nodes, are organized in hierarchies.

Formally, we denote the set of concept types as T_C, the set of relation types as T_R and the set of individual markers that are used to labeled the concept nodes as $\mathrm{markers}$, which defines a vocabulary $\mathcal{V} = (T_C, T_R, \mathrm{markers})$. A basic conceptual graph G is then defined by a 4-uple $G = (C_G, R_G, E_G, l_G)$, where

- (C_G, R_G, E_G) is a finite undirected and bipartite multigraph. C_G is the set of concept nodes. R_G is the set of relation nodes, and E_G is the set of edges.
- l_G is a naming function of the nodes and edges of the graph G which satisfies:
  1) A concept node c is labeled with a pair $l_G(c) = (type(c), marker(c))$, where $type(c) \in T_C$ and $marker(c) \in \mathrm{markers} \cup \{\star\}$.
  2) A relation node r is labeled by $l_G(r) \in T_R$. $l_G(r)$ is also called the type of r.

B. Specialization and generalization of graphs

A specialization/generalization relationship is defined on the graphs. These relationships are used for the query function. The aim of the query is indeed to find all the sub graphs of the information graph that are specializations of the query graph. Therefore, the query is expressed as a generic graph.

1) Relationships between conceptual types: Given the hierarchical nature of the vocabulary, a partial order holds among the set of conceptual types T_C, interpreted as a relation of specialization: $t_1 \leq t_2$ means that t_1 is a specialization of t_2, that is to say that any instance of the class denoted by t_1 is also an instance of the class denoted by t_2.

2) Relationships between concepts: Given the order on T_C, we can also partially order the concepts that are defined on $T_C \times \{\mathrm{markers} \cup \{\star\}\}$, by a specialization relation as follows. Let $c_1 = [T_1 : m_1]$ and $c_2 = [T_2 : m_2]$ be two concept nodes, we define:

$$c_1 \leq c_2 \quad \text{iff} \quad \begin{cases} T_1 \leq T_2 \\ m_2 = \star \quad \text{or} \quad sim(m_1, m_2) \geq thres \end{cases} \tag{1}$$

where sim is a similarity function and $thres$ a user-defined threshold. According to the different applications,

$sim(m_1, m_2)$ and $thres$ may be defines empirically, after a statistical study or heuristically. In the present work, they are defined heuristically.

3) Relationships between graphs: We also define a specialization relation between graphs. This relation is denoted by $\sqsubseteq$ (in order to avoid confusion with the specialization relation $\leq$ between concepts). Let A and B be two basic conceptual graphs. $\mathcal{C}_A$ and $\mathcal{R}_A$ denote the set of concepts and relations of the graph A, defined over the vocabulary $\mathcal{V}$. Denoting as P_{AB} the set of graph isomorphisms between A and B, we have:

$$A \sqsubseteq B \Leftrightarrow \exists p \in P_{AB}, \left\{ \begin{array}{rcl} p : \mathcal{C}_A, \mathcal{R}_A & \rightarrow & \mathcal{C}_B, \mathcal{R}_B \\ c_A, r_A & \mapsto & c_B, r_B \\ \forall c_A \in \mathcal{C}_A, & & c_A \leq c_B \\ \forall r_A \in \mathcal{R}_A, & & r_A = r_B \end{array} \right.$$

C. Generic Fusion Algorithm

The InSyTo Synthesis platform encompasses a generic graph based fusion algorithm made of two interrelated components. The first component is a generic sub-graph matching algorithm, which itself relies on the use of fusion strategies.

The graph matching component takes care of the overall structures of the initial and fused observations. It is in charge of the structural consistency of the fused information, regarding the structures of the initial observations, within the fusion process.

The fusion strategy part is made of similarity, compatibility and functions over elements of the graphs to be fused (see equation 1 for instance). They enable the customization of the generic fusion algorithm according to the context in which it is used.

D. Different fusion functions

According to the fusion strategies that are used, InSyTo graph fusion algorithm provides three different operations. These operations are depicted on figure 2 and described hereafter. This paper illustrates the use of these different operations for a global objective of knowledge discovery.

1) Information Synthesis: **Information synthesis** enables one to collect and organize information about a specific subject. Through information synthesis, all the gathered information items are organized into a network. The redundant part of the information items are detected and eliminated.

The fusion strategies are used within information synthesis, in order to enable the fusion of information items that are slightly different but describe the same situation of the real life. These discrepancies may appear when different sources of information with potentially different levels of precision for instance, are used to draw a picture of an on-going situation (see [1]).

2) Information Query: All the instances of information corresponding to a specified graph pattern may be found within a network of information, through the **information query** function.

Within the query function, the strategies that are used in the algorithm follow the specialization/generalization relationships defined before (III-B). We use a *subsomption* strategy and a whole-structure conservation constraint.

The subsumption strategy is used within the nodes to nodes comparison between the query and the data graphs. That is to say that the concepts of the query graph must be more general than the ones of the data graph in order to be fused.

The specialization relationship between the query and the data graphs also imply that the structure of the query graph must be entirely found in the data graph. The query function relies on the search for injective homomorphism between the query graph and the data graph.

3) Information Fusion: When a model of a situation of interest (e.g. an activity involving a specific person at a specific date) is available, one may want to monitor the situation and trigger further processes if an instance of such a situation is happening. Therefore, different observations, potentially coming from different sources, are filtered out in order to keep observations of interest only. They are then assembled through **information fusion** in order to provide a representation of the ongoing situation of interest, as precise as possible.

The model of situation is, within information fusion, more generic than the observation graphs. Further more, fusion strategies may be used, as for the Information Synthesis function. The use of the model constraints the structure of the fused observation.

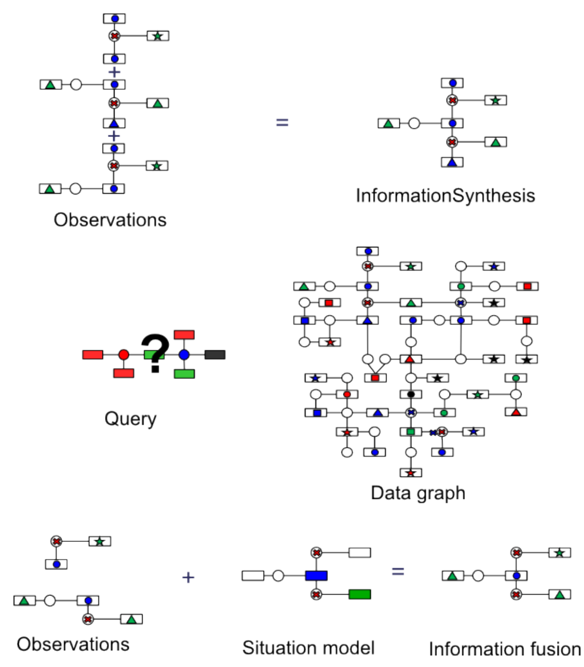

Fig. 2. InSyTo three functions

E. Uncertain Information Fusion

In [2], we introduced the management of uncertainty in the information fusion function. The approach is inspired by [16] and based on the extension of the belief function theory [17],

[18]. The extension enables using the theory within high level information fusion.

In this section, $\otimes$ denotes a fusion operator for the fusion of reliable observations about an event, as defined in the preceding section.

Let $\mathcal{M}$ be the knowledge model associated to a particular event of interest. Suppose we receive an observation $A \sqsubseteq \mathcal{M}$ about this event. Similarly as in ([19], [16]), reliability in the present work means the following: if this observation can be assumed to be reliable, then our knowledge about the event becomes A, and if this observation is assumed to be not reliable, then it is not useful and must be discarded, which amounts to knowing nothing about the event. Note that in our approach, knowing nothing and knowing $\mathcal{M}$ about the event of interest are considered equivalent.

This classical view of the notion of reliability can be extended to the situation where we receive two observations A and B, as follows. There are four elementary cases to consider with respect to the reliability of these observations:

1) If they are both not reliable, then we discard both of them and we only know $\mathcal{M}$ about the event;
2) If observation A is reliable and observation B is not reliable, then we discard observation B and our knowledge about the event is A;
3) If observation A is not reliable and observation B is reliable, then we discard observation A and our knowledge about the event is B;
4) If they are both reliable, then we know $A \otimes B$ about the event.

The reasoning described in the previous paragraph can be formalized as follows. Let $\mathcal{H}_A = \{h_A, \neg h_A\}$ be the assumption space on the reliability of observation A, where h_A (respectively $\neg h_A$) denotes that observation A is reliable (respectively unreliable). Similarly, let $\mathcal{H}_B = \{h_B, \neg h_B\}$ be the assumption space on the reliability of observation B. The set of possible elementary assumptions on the reliability of these two observations is denoted by $\mathcal{H}_{A\times B}$ and defined by $\mathcal{H}_{A\times B} = \mathcal{H}_A \times \mathcal{H}_B = \{(h_A, h_B), (h_A, \neg h_B), (\neg h_A, h_B), (\neg h_A, \neg h_B)\}$. We can define a mapping $\Gamma_{A,B}$ from $\mathcal{H}_{A\times B}$ to $\Pi^*(\mathcal{M})$, which assigns to each elementary hypothesis $h \in \mathcal{H}_{A\times B}$, the result of the fusion of the two observations A and B. $\Gamma_{A,B}(h)$ indicates how to interpret these observations in each of their configuration $h \in \mathcal{H}_{A\times B}$. We have:

$$\begin{array}{rcl}\Gamma_{A,B}(h_A, h_B) & = & A \otimes B;\\ \Gamma_{A,B}(h_A, \neg h_B) & = & A;\\ \Gamma_{A,B}(\neg h_A, h_B) & = & B;\\ \Gamma_{A,B}(\neg h_A, \neg h_B) & = & \mathcal{M}.\end{array}$$

The difficulty is that in general we have uncertain knowledge about the reliability of the observations. We consider in this paper that this uncertainty is represented by a probability distribution $prob^{\mathcal{H}_{A\times B}}$ defined on space $\mathcal{H}_{A\times B}$. Following [16], this uncertainty is transferred through $\Gamma_{A,B}$ onto space $\Pi^*(\mathcal{M})$ in the form of a probability distribution $prob^{\Pi^*(\mathcal{M})}$ defined on $\Pi^*(\mathcal{M})$ by:

$$prob^{\Pi^*(\mathcal{M})}(C) = \sum_{h:\Gamma_{A,B}(h)=C} prob^{\mathcal{H}_{A\times B}}(h), \quad \forall C \in \Pi^*(\mathcal{M}).$$

In this context, $prob^{\Pi^*(\mathcal{M})}(C)$ is the probability that our knowledge about the event of interest be in the form of conceptual graph $C \sqsubseteq \mathcal{M}$. In short, it is the probability of knowing C. For instance, we may assume that observations A and B have independent probabilities q_A and q_B, respectively, of being reliable, in which case we obtain:

$$\begin{array}{rcl}prob^{\Pi^*(\mathcal{M})}(A \otimes B) & = & q_A \cdot q_B;\\ prob^{\Pi^*(\mathcal{M})}(A) & = & q_A \cdot (1 - q_B);\\ prob^{\Pi^*(\mathcal{M})}(B) & = & (1 - q_A) \cdot q_B;\\ prob^{\Pi^*(\mathcal{M})}(\mathcal{M}) & = & (1 - q_A) \cdot (1 - q_B).\end{array}$$

The extension of this approach to the case of more than two partially reliable observations does not raise any theoretical issue. One should be only be aware that the order in which observations are handled may matter depending on whether $\otimes$ is associative.

The belief function theory is extended as follows. Let $prob^{\Pi^*(\mathcal{M})}$ represent our uncertain knowledge about an event of interest. It may for instance be the result of the merging of several partially reliable observations according to the scheme above.

It is insightful to remark that $prob^{\Pi^*(\mathcal{M})}$ is quite close formally to a mass function [17], [18], since a mass function on a finite set Ω is formally a probability distribution on the power set of Ω. Indeed, this comparison can be used to define some concepts inspired from belief function theory, in the present knowledge representation framework, which deals with uncertain and soft knowledge. In particular, we may define the degree of support (or degree of certainty) $Sup(A)$ of a conceptual graph $A \sqsubseteq \mathcal{M}$ as:

$$Sup(A) = \sum_{B \sqsubseteq A, B \neq \diamond} prob^{\Pi^*(\mathcal{M})}(B).$$

This definition is directly inspired from the definition of the belief function associated to a mass function [17], [18]. Its introduction is motivated by the fact that in some problems, such as the one of finding clues of the veracity of an hypothesis studied in Section IV, we may only be interested in a given graph $A \sqsubseteq \mathcal{M}$ and in particular by how much the knowledge derived from the available observations supports this graph.

IV. Uncertain Meeting Discovery

A. The VAST Challenge

The goal of the Visual Analytics Challenges is to advance Visual Analytics Evaluation through Competitions [20]. The Visual Analytics Benchmarks Repository contains resources to improve the evaluation of technology. Benchmarks contains data sets and tasks, as well as materials describing the uses of those benchmarks (the results of analysis, contest entries, controlled experiment materials etc.) Most benchmarks contain ground truth described in a solution provided with the benchmark, allowing accuracy metrics to be computed.

2014 VAST Challenge presents three inter-related mini-challenges and an overall Grand Challenge to test challengers skills. The scenario is the following. In the roughly twenty years that Tethys-based GAStech has been operating a natural gas production site in the island country of Kronos, it has produced remarkable profits and developed strong relationships with the government of Kronos. However, GAStech has not been as successful in demonstrating environmental stewardship.

In January, 2014, the leaders of GAStech are celebrating their new-found fortune as a result of the initial public offering of their very successful company. In the midst of this celebration, several employees of GAStech go missing. An organization known as the Protectors of Kronos (POK) is suspected in the disappearance, but things may not be what they seem.

Among the large amount of data available within the VAST 2014 challenge, this paper focuses on the use of a subset of it. Within the scenario, law enforcement has been given access to the personal and business credit and debit card transactions for the local GAStech employees for the two weeks preceding the kidnapping. Many of the GAStech employees also use loyalty cards to gain discounts or extra benefits at the businesses they patronize, and law enforcement has been given access to two weeks of this loyalty card data as well.

B. Overall Relationship discovery process

Within the global scenario described in the VAST Challenge, we focus on the discovery of extra-professional relationships between persons working for GasTech.

The Kronos scenario doesn't include data that can be directly used as testimony of extra-professional relationships between persons. Given the places in which we know that the employees have activities (thanks to the use of their cards data), we try to discover the ones that may know each other, given the clue that they may have shared activities. We process the credit and loyalty card data sets so to provide two sources of information.

The sources will provide testimonies of potential private meetings (i.e. outside from workplace) between persons. Once a network of private meetings is obtain, the degree of the personnel relationship between two persons is computed thanks to the uncertain fusion operation. Figure 6 depicts the overall process and we detail it hereafter.

In this work, the similarity function $sim(m1, m2)$ is only used for Period concepts. The similarity of two periods is processed according to the normalized duration of their intersection. For two period markers

- $p1$ and $p2$, $sim(p1, p2) = 1$ if $p1$ and $p2$ are equals.
- $sim(p1, p2) = 0.5 * norm(intersect(p1, p2)/norm(p1)$ otherwise.

Other concepts have to have equal markers in order to be compatible regarding fusion, i.e. $sim(m1, m2) = equals(m1, m2)$.

C. Transaction Network construction

The `Transaction` Network contains a description of all the transactions that were recorded through the two cards data sets. It is built in a three steps process.

- Step 1. Each data set is parsed to transform the recorded transactions into a set of graphs of the form depicted on Figure 3.

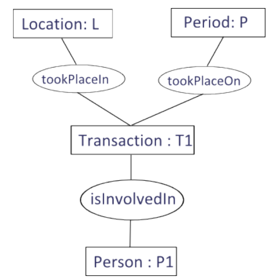

Fig. 3. Model graph of a Transaction.

- Step 2. The two sets of graphs are queried so to find records describing the same transaction. The corresponding record are fused into a single one.
  For each transaction of the first data set, the second one is queried with a query graph of the form as in Figure 3, where the `Period` and `Location` are specified. The fusion strategy used enables getting answers with compatible periods (i.e. overlapping periods)
- Step 3. A new transaction network is built by aggregating all the remaining transactions through the *information synthesis* operation. No fusion strategy is used in this operation, so that all the transaction are kept unfused in the resulting transaction network.

D. Potential Private Meeting Discovery

Once the transaction network has been constructed, the system search for potential private meeting between persons. These potential meetings occur when two persons are at the same place at the same moment.

In other words, we will get clues that a `PrivateMeeting` has potentially occurred when :

- A `Person` P1 realizes a transaction T1 in a `PlaceOfInterest` L at a `Period` D1, and
- a `Person` P2 $\neq$ P1 realizes a transaction T2 $\neq$ T1 in a `PlaceOfInterest` L at a `Period` D2 that is compatible with D1.

This process is depicted in 6 and the use of the different fusion functions is detailed hereafter.

- Step1: Query for activities. All the activities known for each person are listed. To do so, the *query* operation is used with a graph depicting a transaction (Figure 3) where the `Person` is specified.

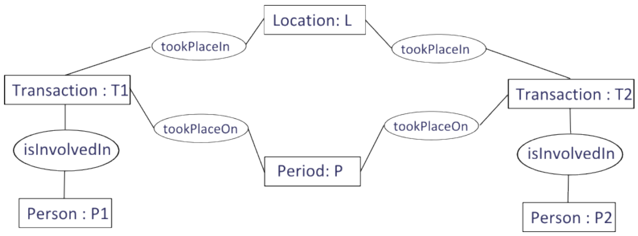

Fig. 4. Query for transactions from two persons at the same place and period.

- Strep 2: Query for transactions that occurred at the same place and at compatible periods. This query is built using the results of the previously obtained activities. The *query* operation is used with a graph query of the form depicted in Figure 4, where a part of the graph is specialized (according to information contained in the activities obtained before) and the other is generic, so that all compatible transactions are retrieved.
- Step 3: Generation of a potential private meetings between the two persons involved in the answers to the previous queries. For each couple of transactions performed by different persons, at the same place and date, we generate a graph describing a potential private meeting between these persons. This graph has the form depicted on Figure 5. All the private meetings graphs are aggregated into a private meeting network, thanks to the *information synthesis* operation.

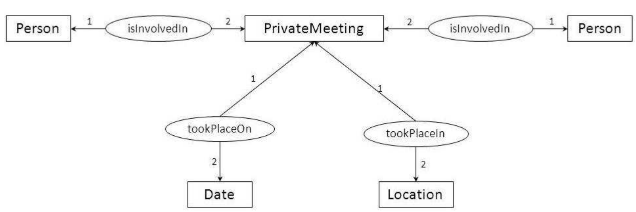

Fig. 5. Private meeting graph

The fact that two persons use their credit cards at the same place almost simultaneously depicts the fact that they might have met. However, this assertion is not sure. Therefore, we grant the potential private meeting graph with a level of reliability.

E. Private Relationships discovery

As said before, our aim is to answer the question "With which degree of certainty do we know that person X knows personally person Y ?". To do so, we apply the uncertain fusion process described above in order to find information items that confirm or contradict this query.

Two persons are known to have a private relationship if they are involved in private meetings. The degree of their relationship is processed according to the degree of support of the fusion of all their potential private meetings. Therefore, the uncertain fusion process is applied with a strategy that enables fusing any `PrivateMeeting` pair of concepts, provided that the persons involved in the meeting are strictly the same (i.e. an "identity" fusion strategy is used for `Person` nodes.)

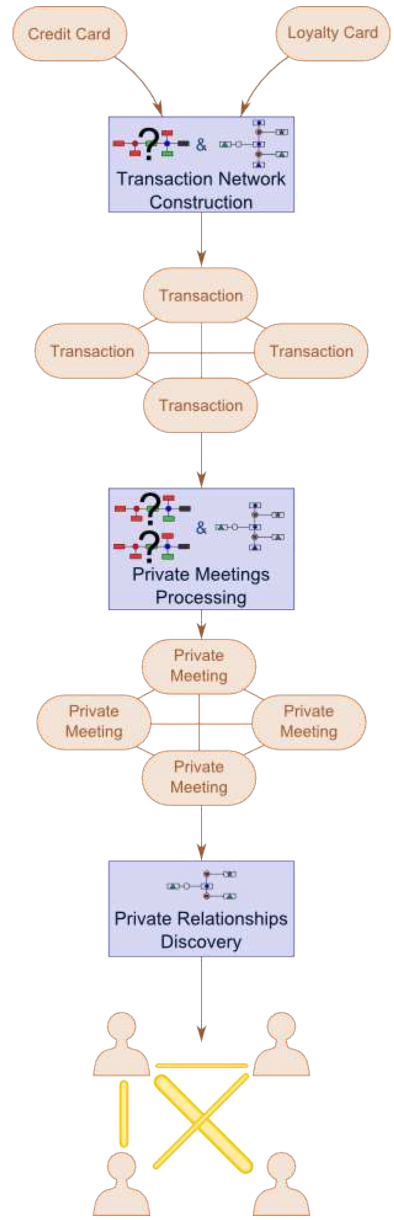

Fig. 6. Private relationships discovery process

F. Results

We applied our Relationship Discovery process to the credit card and loyalty card data of the VAST Challenge. Within the 15 days of data, we discovered more than 1200 potential private meetings between couples of persons. Figure 7 depicts the visualization of the potential meetings a person (Fusil Stenig) had with the other persons of the scenario during one day. The bubbles in the center of the picture depict the locations of the meetings. Their size is proportional to the number of meetings that potentially occurred in this place. The color of the bubbles indicate the certainty that we have that the person has been in the different places. The color of the persons cell depicts the certainty that a person eventually met the displayed actor.

The strength of the relationships between couples of people is depicted on Figure 8 as a network of the persons involved in the scenario. Each bubble corresponds to a person and its size is proportional to the number of links that the person has with the others. When focusing on one single person (left hand side of Figure8), only the linked persons appear. The color of each bubble depicts the certainty of the existence of a private relationship.

The VAST Challenge provided its answers. The kidnapper happened to be the five first listed in the Figure 9, people helping the kidnappers are the 3 following and kidnapped persons are the five following ones. The cells of the table provide the private relationships support that we processed using only a small subset of the data. Several communities have been highlighted by our process, among them the kidnappers one.

We also highlighted a private relationship between Isande Borrasca and Brand Tempestad, which appear to be dating, according the results provided by VAST.

V. Conclusion and Future Work

In previous work ([1], [2], [21]) we described a generic graph based information fusion algorithm. This algorithm is embedded in a High level information fusion platform that enables using it in different contexts. Thanks to the different parametrization, several operations are provided : information synthesis, information fusion, information query and uncertain information fusion. This paper shows the use of these different functions in order to discover private relationships and build a social network, out of the data set provided by the VAST 2014 Challenge.

We show that, even if the VAST data sets were developed in order to experiment visual analytics tools, high level information fusion is of use for these tools. Information fusion enables preparing the data so that the "visual fusion" and visual analysis is supported and more efficient.

Regarding the VAST data, on going work includes the use of other data sets. Regarding the InSyTo platform, future work will be dedicated to the extension of the uncertain fusion approach to the other operations. We will also adapt the fusion algorithm so that it can be used on data flows rather than on static data sets. This will enable using the high level information fusion within social media and micro-blog monitoring for instance.

regarding future work, the issue of integrating new sources of information is raised. [22] focuses on the detection of relations between persons contributing to discussion through social networking services. This work may be used in order to provide an other source of information for our process. It can, for example, be applied to the e-mail data set provided by the VAST Challenge. The use of such work as systems providing inputs to our system out of the context of the VAST data sets is envisioned. Similarly, work such as [13] could also be used.

References

[1] C. Laudy, *Semantic Knowledge Representations for Soft Data Fusion — Efficient Decision Support Systems - Practice and Challenges From Current to Future*. Chiang Jao Publisher, 2011.

[2] S. Fossier, C. Laudy, and F. Pichon, "Managing uncertainty in conceptual graph-based soft information fusion," in *Proceedings of the 16th International Conference on Information Fusion, FUSION 2013, Istanbul, Turkey, July 9-12*, 2013.

[3] H. Ma, Z. Lu, L. Fan, W. Wu, D. Li, and Y. Zhu, "A nash equilibrium based algorithm for mining hidden links in social networks," *Lecture Notes in Computer Science*, vol. 8287, 2013.

[4] B. Yan and S. Gregory, "Finding missing edges in networks based on their community structure," *Phys. Rev. E*, vol. 85, p. 056112, May 2012.

[5] D. Prakash and S. Surendran, "Article: Detection and analysis of hidden activities in social networks," *International Journal of Computer Applications*, vol. 77, no. 16, pp. 34–38, September 2013, full text available.

[6] D. Cai and T. Feng, "Detecting hidden structure in social networks," Technical Report, Harvard University, Dec 2012., 2012.

[7] C. Laudy, É. Deparis, G. Lortal, and J. Mattioli, "Multi-granular fusion for social data analysis for a decision and intelligence application," in *Proceedings of the 16th International Conference on Information Fusion, FUSION 2013, Istanbul, Turkey, July 9-12*, 2013.

[8] K. Ogaard, H. Roy, S. Kase, R. Nagi, K. Sambhoos, and M. Sudit, "Discovering patterns in social networks with graph matching algorithms," in *Proceedings of the 6th International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction*, ser. SBP'13. Berlin, Heidelberg: Springer-Verlag, 2013, pp. 341–349.

[9] K. Sambhoos, J. Llinas, and E. Little, "Graphical methods for real-time fusion and estimation with soft message data," in *Information Fusion, 11th International Conference on*, 2008.

[10] D. Cai, Z. Shao, X. He, X. Yan, and J. Han, "Mining hidden community in heterogeneous social networks," in *Proceedings of the 3rd International Workshop on Link Discovery*, ser. LinkKDD '05. New York, NY, USA: ACM, 2005, pp. 58–65.

[11] R. Xiang, J. Neville, and M. Rogati, "Modeling relationship strength in online social networks," in *Proceedings of the 19th International Conference on World Wide Web*, ser. WWW '10. New York, NY, USA: ACM, 2010, pp. 981–990.

[12] D. Liben-Nowell and J. Kleinberg, "The link prediction problem for social networks," in *Proceedings of the Twelfth International Conference on Information and Knowledge Management*, ser. CIKM '03. New York, NY, USA: ACM, 2003, pp. 556–559.

[13] F. Su, Y. Liu, L. Zhou, and M. Ye, "Information fusion of crossing network communities based on analyzing interests," in *Bio-Inspired Computing: Theories and Applications (BIC-TA), 2010 IEEE Fifth International Conference on*, 2010.

[14] J.-F. Sowa, *Conceptual Structures. Information Processing in Mind and Machine*, M. Addison-Wesley, Reading, Ed., 1984.

[15] M. Chein and M.-L. Mugnier, *Graph-based Knowledge Representation: Computational Foundations of Conceptual Graphs*. Springer, 2008.

[16] F. Pichon, D. Dubois, and T. Denoeux, "Relevance and truthfulness in information correction and fusion," *International Journal of Approximate Reasoning*, vol. 53, no. 2, 2012.

[17] G. Shafer, *A mathematical theory of evidence*. Princeton University Press, Princeton, N.J., 1976.

[18] P. Smets and R. Kennes, "The transferable belief model," *Artificial Intelligence*, vol. 66, no. 2, 1994.

[19] P. Smets, "Belief functions: The disjunctive rule of combination and the generalized bayesian theorem," *International Journal of Approximate Reasoning*, vol. 9, no. 1, 1993.

[20] V. 2014, http://vacommunity.org/VAST+Challenge+2014.

[21] C. Laudy, C. Gouguenheim, and J. Dréo, "Applying mapreduce principle to high level information fusion," in *Information Fusion, 17th International Conference on*, 2014.

[22] M. Song, W. Lee, and J. Kim, "Extraction and visualization of implicit social relations on social networking services," in *Extraction and Visualization of Implicit Social Relations on Social Networking Services*, 2010.

Fig. 7. Private meetings visualization

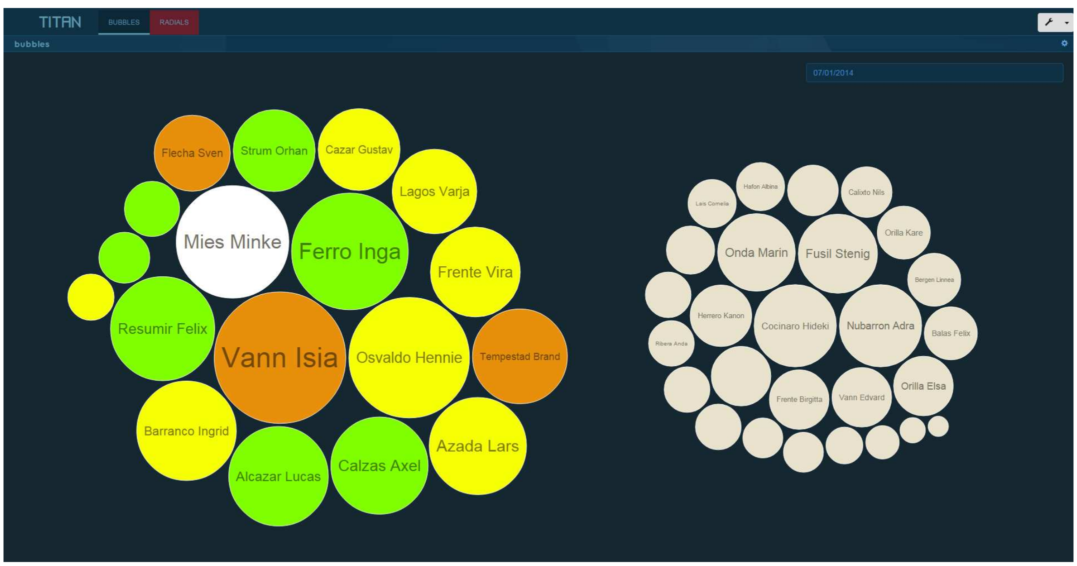

Fig. 8. Relationships visualization

	Osvaldo Hennie	Vann Isia	Ferro Inga	Bodrogi Loreto	Mies Minke	Forluniau Carla	Mies Haber Ruscella	Lais Cornelia	Pantanal Rachel	Barranco Ingrid	Campo-Corrente Ada	Strum Orhan	Vasco-Pais Willem	Borrasca Isande	Tempestad Brand
Osvaldo Hennie	-	0.91	0.88	0.53	0.68	0.28	0.0	0.58	0.0	0.77	0.0	0.0	0.0	0.0	1.0
Vann Isia	0.91	-	0.87	1.0	0.50	0.52	0.50	0.50	0.0	0.72	0.93	0.0	0.0	0.0	1.0
Ferro Inga	0.88	0.87	-	0.72	0.77	0.0	0.57	0.57	0.0	0.0	0.0	0.0	0.0	0.0	0.73
Bodrogi Loreto	0.53	1.0	0.72	-	0.86	0.0	0.59	0.0	0.0	0.0	0.93	0.0	0.0	0.0	0.0
Mies Minke	0.68	0.50	0.77	0.86	-	0.0	0.32	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.56
Forluniau Carla	0.28	0.52	0.0	0.0	0.0	-	0.0	0.0	0.0	0.55	0.0	0.0	0.0	0.0	0.0
Mies Haber Ruscella	0.0	0.50	0.57	0.59	0.32	0.0	-	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Lais Cornelia	0.58	0.50	0.57	0.0	0.0	0.0	0.0	-	0.0	0.0	0.0	0.0	0.0	0.0	0.32
Pantanal Rachel	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-	0.0	0.0	0.0	0.0	0.0	0.0
Barranco Ingrid	0.77	0.72	0.0	0.0	0.0	0.55	0.0	0.0	0.0	-	0.51	0.83	0.0	0.62	0.0
Campo-Corrente Ada	0.0	0.93	0.0	0.93	0.0	0.0	0.0	0.0	0.0	0.51	-	0.78	0.0	0.84	0.0
Strum Orhan	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.83	0.78	-	1.0	0.0	0.0
Vasco-Pais Willem	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	-	0.0	0.0
Borrasca Isande	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.62	0.84	0.0	0.0	-	1.0
Tempestad Brand	1.0	1.0	0.73	0.0	0.56	0.0	0.0	0.32	0.0	0.0	0.0	0.0	0.0	1.0	-

Fig. 9. Relationships between kidnappers and kidnapped persons