Official URL:

https://doi.org/10.1109/ICDIM.2017.8244684

To cite this version:

Belghache, Elhadi and Georgé, Jean-Pierre and Gleizes, Marie-Pierre *DREAM: Dynamic data Relation Extraction using Adaptive Multi-agent systems.* (2017) In: Twelfth International Conference on Digital Information Management (ICDIM), 12 September 2017 - 14 September 2017 (Kyushu University, Fukuoka, Japan).

DREAM: Dynamic data Relation Extraction using Adaptive Multi-agent systems

Elhadi Belghache
IRIT, University of Toulouse
Toulouse, France
Elhadi.Belghache@irit.fr

Jean-Pierre Georgé
IRIT, University of Toulouse
Toulouse, France
Jean-Pierre.George@irit.fr

Marie-Pierre Gleizes
IRIT, University of Toulouse
Toulouse, France
Marie-Pierre.Gleizes@irit.fr

***Abstract*—Understanding data is the main purpose of data science and how to achieve it is one of data science challenges, especially when dealing with big data. In order to find meaning and relevant information drowned in the data flood, while overcoming big data challenges, one should rely on an analytic tool able to find relations between data, evaluate them and detect their changes and evolution over time. The aim of this paper is to present the *DREAM*[1] tool for dynamic data relations discovery and dynamic display based on a collective artificial intelligence *Adaptive Multi-Agent System (AMAS)* that uses a new data similarity metric, the *Dynamics Correlation*. It is currently being applied in the *neOCampus* operation, the ambient campus of the University of Toulouse III - Paul Sabatier.**

***Keywords*—Big Data; Adaptive Multi-Agent Systems; Dynamic Analytics; Dynamics Correlation**

I. Introduction

G. Piatetsky-Shapiro defined the *Knowledge Discovery from Data (KDD)* process [1], which has became the standard data analytics pipeline. The most important step of the KDD process is finding hidden patterns through data mining. These patterns are used to build *"Models"*, a more compact or a more useful representation of the raw data.

In response to the rising big data challenges [2] we provide a computer system able to build and display, in real time from huge amounts of data, a new *model* in form of a dynamic graph, that adjusts itself to adapt to changes in data content and structure, wherein a node represents a data source (sensor stream, database attribute, data file column...) and an edge exhibits a better-meaning correlation between the two related data sources to help the users find relevant relations.

To process these big data our system relies on a bio-inspired collective artificial intelligence (Adaptive Multi-Agent Systems) that uses a new analytical tool (Dynamics Correlation), described in the following.

The next sections of the paper will focus on a new similarity metric designed to detect the data dynamics, how it is used by a collective artificial intelligence with several experiments.

II. Analytical tools

Our system relies on a new analytical tool, that can handle dynamic big data, designed from conventional analytical tools as explained in this section.

[1]*DREAM* stands for *D*ynamic data *R*elation *E*xtraction using *A*daptive *M*ulti-agent systems

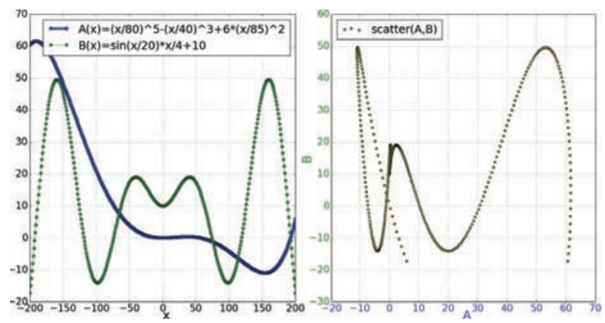

Fig. 1: Non-linearity.

A. Correlation coefficient

When investigating relations between data, one relies first on the most spread analytical tool, the statistical correlation defined whit the Pearson's correlation coefficient r [3] as follows:

$$r(A,B) = \frac{\sum_{i=1}^{n}(a_i b_i) - n\bar{A}\bar{B}}{n\sigma_A\sigma_B} \qquad \text{(II.1)}$$

Where,

- A, B are two variables (data features).
- $\bar{X}$ is the mean of X.
- σ_X is the standard deviation of X.
- n is the number of data points (values).

We use the correlation magnitude r^2 to measure the strength of the correlation : the greater r^2 is, the stronger is the correlation.

However, the correlation coefficient has a major downside: non-linearly correlated variables are considered as independent ($r \approx 0$) but they are in fact correlated and therefore it causes a loss of relevant information. The non-linear correlation is mainly due to:

- **Non-linearity:** when at least one of the variables is non-linear and random-like (see figure 1).
- **Time shift:** For example in figure 2 $A(x) = (\frac{x}{10})^2$ and $B(x) = (\frac{x}{10} + 5)^2$, both are square functions that take the same argument ($\frac{x}{10}$), with a delay of 5 for the second one (B).
- **Non-linearity and Time shift:** see figure 4.

In order to distinguish between a true independence and a non-linear correlation, one can look for a pattern or a shape

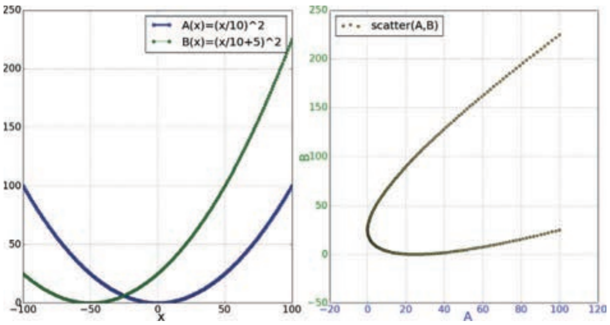

Fig. 2: Time shift.

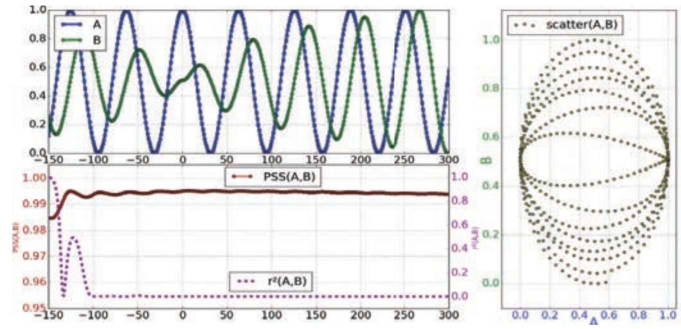

Fig. 4: Phase Space Similarity VS Correlation Coefficient.

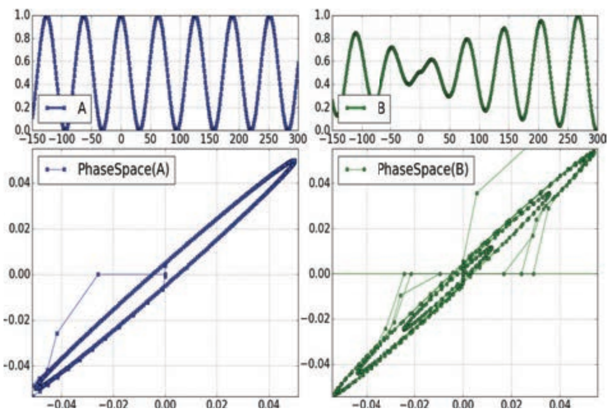

Fig. 3: Phase Space.

in the scatter plot (see figure 2, figure 1 and figure 4), which suggests a non-linear correlation. Conversely, a uniform scatter plot indicates an independence. Though, this process can be quiet difficult and non-systematic. Hence another analytical tool based on a similar process is needed.

B. Phase Space Similarity

The defining feature of a scatter plot is its graphical representation of the relation between two variables over time. This representation may be so complex that it looks like a uniform scatter, which hints at an independence, as a result of the projection of one complex variable over another one.

We exploit another representation of the variables, that avoids the projection, to study the relation between them. This representation, known as the *Phase Space*, came from physics [4] and is built following the behavior of a single variable over time.

1) Phase Space: The Phase Space PS is a collection of points, whose coordinates are the difference of successive data points (values) in a sliding window, defined as follows:

$$(psx_{A_i}, psy_{A_i}) = (A_i - A_{i-1}, A_{i+1} - A_i) \tag{II.2}$$

$$PS_A = \{(psx_{A_i}, psy_{A_i}), \forall i \in [1, n-1]\} \tag{II.3}$$

For example, in figure 3 the Phase Space of a sinusoidal function is represented as an ellipse by reason of the cyclic nature of the sinusoidal function. Moreover, all the time shifted variables have the same Phase Space.

2) Phase Space Similarity metric: As a similarity metric the *Phase Space Similarity (PSS)* should vary from 0 to 1, meaning full dissimilarity and full similarity respectively. Accordingly, we define the PSS metric for an automatic comparison between two phase spaces, as follows:

$$Max\_ED(A,B) = \begin{cases} \sqrt{2} & \text{if A and B are normalized} \\ \sqrt{\begin{array}{c}\left(\max\limits_{2\le i\le n-1} psx_i - \min\limits_{2\le i\le n-1} psx_i\right)^2 \\ + \\ \left(\max\limits_{2\le i\le n-1} psy_i - \min\limits_{2\le i\le n-1} psy_i\right)^2\end{array}} & \text{else} \end{cases} \tag{II.4}$$

$$PSD(A,B) = \frac{\sum_{i=2}^{n-1} \sqrt{\left(psx_{A_i} - psx_{B_i}\right)^2 + \left(psy_{A_i} - psy_{B_i}\right)^2}}{(n-2)\ Max\_ED(A,B)} \tag{II.5}$$

$$PSS(A,B) = \left(1 - PSD(A,B)^2\right)^2 \tag{II.6}$$

If the data are not normalized their phase spaces will have different scales and the PSS will, probably, be negative even though they have the same dynamics. Thus, the PSD in equation II.5 should be divided by the *Maximal Euclidean Distance* (equation II.4).

As shown in figure 4, the correlation coefficient r^2 starts with high values then quickly decreases and remains very close to 0. While the PSS varies between 0.99 and 1 meaning a high similarity between the phase spaces of the two non-linear and time shifted variables A and B, which indicate a non-linear relation between the variables that the coefficient correlation couldn't expose.

3) Local Phase Space Similarity: The PSS, like the correlation coefficient, uses an arithmetic mean to compute the mean euclidean distance of the phase spaces points from the beginning (equation II.5). Therefore, it gives an overall similarity value and muffles short *Situations of Interest (SI)*, time intervals where data (values) are correlated, contained in data that are mostly not correlated (see figure 5). In other words, relevant *Situations of Interest* may be drowned in the data as a consequence of the PSS memory (arithmetic mean).

Therefore, we define the *Local PSS (LPSS)* as a PSS without memory, meaning it is a PSS given only the last phase space points, as follows:

$$LPSD(A_i, B_i) = \frac{\sqrt{\left(psx_{A_i} - psx_{B_i}\right)^2 + \left(psy_{A_i} - psy_{B_i}\right)^2}}{Max\_ED(A,B)} \tag{II.7}$$

$$LPSS(A_i, B_i, m) = \frac{\sum_{j=i-m+1}^{i} \left(1 - LPSD(A_j, B_j)^2\right)^2}{m}, m \geq 1 \tag{II.8}$$

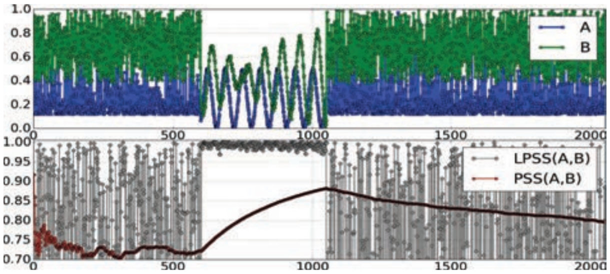

Fig. 5: Local Phase Space Similarity (LPSS) with $m = 1$.

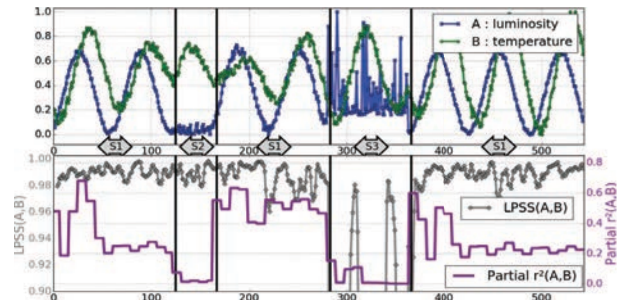

Fig. 6: Dynamics Correlation applied on *neOCampus* sensor data (luminosity and temperature) with artificial noise in $S2$ and $S3$.

Where m is a smoothing factor of the metric output used to lessen occasional disturbances. In our experiments, we set m to 5 (see figure 6), which results in 15 data points (values) because a single phase space point needs three data points. With $m = 5$, the $LPSS$ fits the data dynamics well enough while filtering outlier values.

The data shown in figure 4 have a high PSS when analyzed as it is. When these correlated data are inserted in random data (see figure 5) they are considered as a *situation of interest (SI)*. On one hand, when the SI starts, the PSS increases slowly although it doesn't reach the same values as in figure 4 as a consequence of the random data before the SI that lowers the overall PSS. Then, after the ending of the SI, the PSS keeps decreasing. On the other hand, the $LPSS$ varies randomly between 0 and 1 for the data before and after the SI. Furthermore, the $LPSS$ points out very well to the correlated data by varying between 0.96 and 1 during the SI.

C. Dynamics correlation

When one variable is constant-like and the other one is highly dynamic, like respectively B and A in $S3$ in figure 6), the phase space of the former overwhelms the phase space of the latter leading phase space based similarity metrics to be falsely high.

To fix this issue of false positives, we define the *Dynamics Correlation* metric, a combination of the $LPSS$ with the coefficient correlation computed for the data in situations of interest ($Partial\ r^2$), as follows:

- If $LPSS \geq 0.95$ and $Partial\ r^2 \geq 0.01$ it's a true situation of interest ($S1$).
- If $LPSS \geq 0.95$ and $Partial\ r^2 < 0.01$ it's a false situation of interest ($S2$).
- If $LPSS < 0.95$ it isn't a situation of interest (3).

To see the 3 possible outputs of the *Dynamics Correlation*, we use luminosity [5] data and temperature [6] data with some artificial noise (figure 6), generated during the 9 first days of July 2017 by ambient sensors of one room of the *neOCampus* operation [7].

To take full advantage of the *Dynamics Correlation*, we associate it (see section IV) with a collective artificial technology described in the next section.

III. Adaptive Multi-Agent Systems theory

To keep up with the *Data Flood*, conventional data analytic techniques require more storage capabilities and computing power due to their centralized processing pipeline.

In a centralized processing pipeline, even though distributed, there is ultimately one unit, or very few, that gather the processing sub-results to compute the final result, which induces a bottleneck equivalent to a cognitive overload in the brain when the number of the inputs or perceptions reaches the limit of the system. Natural systems like bird flocking, animal herding and bacteria, are decentralized and thrive in harmful and highly dynamic environments. which have led to the family of bio-inspired collective artificial intelligence.

This section presents the *Adaptive Multi-Agent Systems (AMAS)*, a collective artificial intelligence used to design our big data analytic tool.

A. Multi-Agent Systems

A Multi-Agent System (MAS) [8] is defined as a macro-system composed of autonomous agents which pursue individual objectives and which interact in a common environment to solve a common task. The autonomy of an agent is a fundamental characteristic: an agent is capable of reacting to its environment and act from its own decision, relying only on a limited and localized knowledge of the environment.

B. Self-Adaptive Multi-Agent Systems

A self-Adaptive Multi-Agent Systems (AMAS) is a MAS able to adjust itself, organize itself, heal itself, etc. [9] to remain in a well-functioning state after a perturbation.

A designer following this approach focuses on giving the agents a local view (and partial knowledge) of their environment, means to detect problematic situations and guidelines to act in a cooperative way, meaning that the agents will try to achieve their goals while respecting and helping the other agents around them as best as they can [10]. The fact that the agents do not follow a global directive towards the solving of the problem but collectively build this solving, produces an emergent [11] problem solving process that explores the search space of the problem in original ways.

The difficulty here is to give the agents the right local behavior in order to get the right global function and a good adaptation capability since there is no formal process, which translates the behavior of the components and their interactions into a well-defined global function.

IV. AMAS FOR REAL-TIME DYNAMICS CORRELATION

For a system of n inputs (variables), it takes $\frac{n(n-1)}{2}$ calls of the dynamics correlation analytical tool to examine all the possible relations, which corresponds to a temporal complexity of $O(n^2)$ if computed sequentially or a spatial complexity if computed in the same time, in either ways, this high complexity prevents the system to scale up.

Thence, for the sake of designing a computer system that produces a dynamic graph model of the data relations in real time, we incorporate the dynamics correlation into a self-adaptive multi-agent system to focus only on the probably correlated data and thus reduce the computing power to find all the data relations.

A. The agents: system architecture

The system is composed of two types of agents: "Percept" and "Correlation".

- **Percept:** a Percept agent represents a unique data stream (a sensor, an attribute in a database table, a column in a CSV file, etc.). The Percept receives the data, normalizes them and sends them to its associated Correlation agents. It also links itself cooperatively to other Percepts, as described in the next subsection, by creating common Correlation agents in order to study the dynamics correlation on the fly. Furthermore, the percept helps other percepts to find dynamics correlations between them by linking them when it is relevant.
- **Correlation:** a Correlation agent is associated with two Percept agents, A and B. When it receives new data values from one of its Percepts, says A, the agent applies the dynamics correlation analytical tool following this procedure:
  1- Put the new data value of the percept A in a small data buffer until the correlation agent receives data from the other percept agent (B).
  2- When there is a new data value B_i coming from the second percept agent (B), get A_i the data value saved in the buffer if there is only one, or the mean value of the data saved in the buffer as a consequence of different data acquisition frequencies of the two percepts.
  For example, in figure 6 the luminosity sensor data [5] are produced each 20s and the temperature sensor [6] generates new data each minute. Thus, each temperature data point corresponds to the mean of the mean of the 3 last luminosity data points.
  3- If $i \geq 3$, compute for the new data couple (A_i, B_i) their phase space points, $(psx_{A_{i-1}}, psy_{A_{i-1}})$ and $(psx_{B_{i-1}}, psy_{B_i - 1})$ respectively, using equation II.2. Else, go back to 1.
  4- Compute the $LPSS$ with equation II.8. If $LPSS \geq 0.95$, it is the beginning of a new situation of interest (SI), then starts computing the correlation coefficient r^2 for SI ($Partial\ r^2$) incrementally, using equation II.1 where the mean and the standard deviation are updated [12] as follow:
  $$\bar{A}_i = \frac{S_i}{i} \qquad with \quad S_i = S_{i-1} + A_i, \quad S_0 = 0$$
  $$\sigma_{A_i} = \frac{Q_i}{i} - \bar{A}_i^2 \ \ with\ \ Q_i = Q_{i-1} + A_i^2,\ \ Q_0 = 0$$
  Else, it's not a situation of interest (S3, figure 6).
  5- When SI ends, because $LPSS < 0.95$ and $Partial\ r^2 \geq 0.01$, the data of SI are dynamically correlated (S3, figure 6). Moreover, if $Partial\ r^2 \geq 0.95$, the data are variating at the same time, else they are time-shifted.
  6- Otherwise, SI ends when $Partial\ r^2 < 0.01$ (see S2 in figure 6), it still is a situation of interest wherein one of the percept has a well-defined dynamic and needs the cooperation of the other agents to find a percept which is correlated with. Then go back to 1.

B. Cooperative behavior & interactions

Cooperation is the engine of the self-organization processes taking place in the system and the heart of our bottom-up method. Cooperation is classically defined by the fact that two agents work together if they need to share resources or competences. We describe the cooperation mechanism of our AMAS as follows:

1- Initially, when the system starts, each data stream is agentified, in other words, a dedicated Percept agent is created to represent and handle the stream.
2- A new Percept agent first builds a random neighborhood, which means it links itself to random Percepts agents by creating common Correlation agents.
3- As soon as a Correlation agent finds a situation of interest, the agent sends it back to its Percepts agents.
4- Then these percepts agents update their mutual correlation and spread it through their neighbors if the situation of interest represents a dynamics correlation.
5- Otherwise, the Percept agent with the well-defined dynamic tries to find a correlation with another neighbor for this data segment (active search) and the other Percept agent puts in contact the former with Percepts agents that have a well-defined dynamic for the same segment as well (passive search).
6- If after a long time the Correlation agent doesn't find any situation of interest, the agent becomes useless and signals it to its Percepts agents in order to launch an inquiry into a potential anomaly (sensors malfunction). Then the agent destroys itself.
7- Likewise, when a Percept agent doesn't receive new situation of interest or doesn't help other (5-passive search) anymore, it expands its neigh1borhood randomly to find new correlations. If this doesn't work the Percept agent raises an anomaly alert of uselessness.
8- Also, according to the openness property of the AMAS theory, when a new Percept agent is created, it will build a small random neighborhood and each of its neighbors suggests to it other interesting percepts.
9- Finally, when a Percept agent is not computing (it has free time) or all of its associated correlation agents are

destroyed, it expands its neighborhood by selecting the neighbors of its neighbors that have similar situations of interest.

C. Graph model building

The data relations model is represented by a dynamic graph, wherein a node expresses a data stream and an edge exhibits a dynamics correlation between two data sources, built and updated by the agents with these additional cooperative behaviors:

1) **The nodes:** are created and destroyed by the percept agents.
   a) When a percept agent starts, it creates a new node labeled with the corresponding data source name.
   b) When a percept agent destroys itself,
      i) it sends a message to warn, its associated correlation agent, of its destruction.
      ii) it removes its node from the graph model.
2) **The edges:** are managed by the correlation agents
   a) After detecting situations of dynamics correlation (s) if $\sum_{s \in S} |s| \geq 50$, where S is the set of all the situations s and $|s|$ is the number of data values contained in S (length of S), then:
      i) if there isn't already an edge, the correlation agent inserts a new edge that expresses a relation between the two nodes corresponding to its percepts.
      ii) update the relation intensity ($RI(A, B)$) with:

$$R\_Len(A,B) = \frac{\sum_{s \in S} |s|}{n}, \text{ } n \text{ is the number of all data values} \quad \text{(IV.1)}$$

$$R\_LPSS(A,B) = \frac{\sum_{s \in S} LPSS(s)}{|S|} \quad \text{(IV.2)}$$

$$R\_Corr(A,B) = \frac{\sum_{s \in S} r^2(s)}{|S|} \quad \text{(IV.3)}$$

$$RI(A,B) = \frac{R\_Len(A,B) + R\_LPSS(A,B) + R\_Corr(A,B)}{3} \quad \text{(IV.4)}$$

   b) When it receives a warning message from one of its percepts agents (1(b)i), the correlation agent destroys itself.
   c) When destroyed, the agent removes its edge

Some examples of such dynamic graph model are illustrated in the next section.

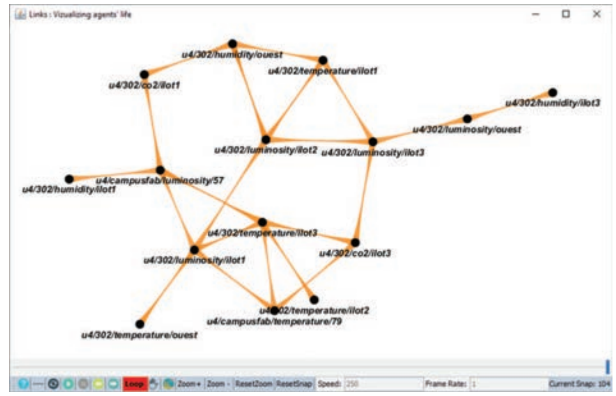

Fig. 7: Graph model after 2H analysis of *neOCampus* data.

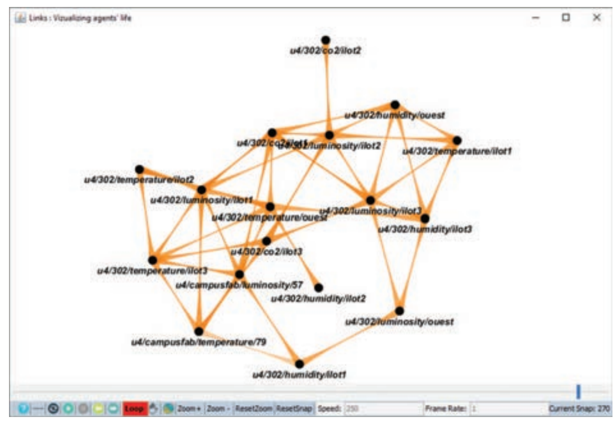

Fig. 8: Graph model after 5H analysis of *neOCampus* data.

V. Experiments & Applications

Our system is applied mainly to analyze real-time ambient sensor data and time series data sets. The intensity of a relation (equation IV.4), for the experiments presented in this section, are expressed by the opacity of the corresponding edge such the greater is RI the higher is the edge opacity.

A. Real-time data

For a real time experiment, we rely on the *neOCampus* operation [7], the ambient campus of the University of Toulouse III, that is being iteratively equipped with pervasive ambient sensors and effectors. Student activity will be one of the main generator of data. The real time data used for this experiment are generated by sensors of two rooms inside the U4 building of the university: the room *campusfab* located on the ground floor and the room *302* on the third floor.

When the system starts (7) there is few relations, because of the small size of the first neighborhoods of the percepts. Then with their cooperative behavior (section IV) the agents explore more efficiently the sensors network, by focusing the probable relation with some random linking to avoid stagnation, leading to the appearance of some clusters (figure 8)

In the scope of ambient systems, Internet of Things and Smart Cities, our system can provides extra features:

- **Anomaly detection:** when all the relations of one data stream disappear, it may be a normal change of the data stream dynamics or corruption caused by a failure or disturbances. The system points out this anomaly an then the user should investigate if there is an issue with the data stream and fixes it.
- **Data generator:** if a data stream is corrupted, the user can use the dynamics model to producing phantom limb data using the data related data streams.
- **Eco-feedback:** real time discovery of better-meaning correlations between users action and energy consumption to present meaningful feedback to help the users find where and how they can save energy [13].

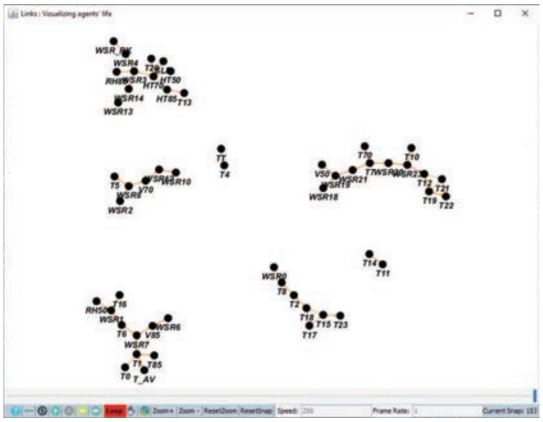

Fig. 9: Graph model after 100 lines of Ozone data set [14].

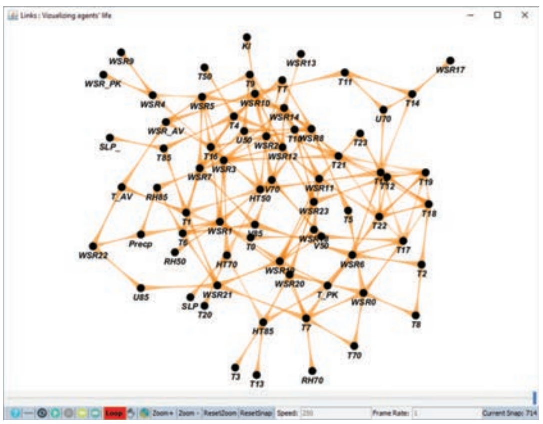

Fig. 10: Graph model after 300 lines of Ozone data set [14].

B. Time series data sets

The UC Irvine Machine Learning Repository (UCI) provides relevant data sets with high number of variables, that highlight the advantage of using our the *DREAM* system to analyze huge data sets. For the next experiments, we use the Ozone Level Detection Data Set [14], which contains 72 variables and 2536 lines (days) with some missing values replaced by the mean. The data set is evolving over time from 1998 to 2004. Our dynamic model can give a first hint for environmental scientists to explain what these variables are and how they actually interact in the formation of ozone [15].

The resulting graph model can be used for:

- **Dimensionality reduction:** the variables that have a high dynamics correlation can be considered as a single variable.
- **Initialization and tuning of Machine Learning algorithms used for prediction:** uses the relations of the class attribute to select the relevant factor for it prediction. For example, in a neural network set the initial weight of a variable (neuron) with it relation intensity (RI)

VI. Conclusion and Future Work

The speed at which new data is generated, and the need to manage changes in data both for content and structure lead to new rising challenges in what can be called Dynamic Big Data Analytics. This is especially true in the context of complex systems with strong dynamics, as in for instance large scale ambient systems. One existing technology that has been shown as particularly relevant for modeling, simulating and solving problems in complex systems are Multi-Agent Systems. We described and discussed in this article how such a technology can be applied to big data in the form of an *Adaptive Multi-Agent System* where local analytics agents interact in a self-organised way.

This technology is currently being applied to several problems that will show its genericity (i.e. it does not require domain-specific expertise from the engineer that applies it) and validate its interests. The first is the *neOCampus* operation. The second is the *3Pegase* project in which we work with Orange and hospitals in Toulouse among others. The aim is an end-to-end predictive platform for elderly people staying at their own pervasively equipped homes. The third will be the performance and quality validation in well known big data on-line competitions.

Our future work will focus on improving our system for fast real-time correlation detection in dynamic environments with limited computing power by fixing the number of the active agents in the same time and adding to the system a dynamic relation characterization ability using a new reasoning mechanism inspired from logic (Inference to the Best Explanation) and epidemiology (Hill's criteria of causation).

References

[1] G. Piatetsky-Shapiro, "Knowledge discovery in real databases: A report on the ijcai-89 workshop," *AI magazine*, vol. 11, no. 4, p. 68, 1990.

[2] E. Belghache, J.-P. Georgé, and M.-P. Gleizes, "Towards an adaptive multi-agent system for dynamic big data analytics," in *Cloud and Big Data Computing (CBDCom)*. IEEE, 2016, pp. 753–758.

[3] B. E. Cooper, "Correlation and Function Fitting," in *Statistics for Experimentalists*. Elsevier, May 2014, pp. 206–212.

[4] D. D. Nolte, "The tangled tale of phase space," *Physics Today*, vol. 63, no. 4, pp. 33–38, Apr. 2010.

[5] "neocampus-u4-302-ilot1-lux," Jul. 2017. [Online]. Available: https://neocampus.univ-tlse3.fr/domoticz/

[6] "neocampus-u4-302-ilot1-temp," Jul. 2017. [Online]. Available: https://neocampus.univ-tlse3.fr/domoticz/

[7] "The neOCampus projey of the University of Toulouse III - Paul Sabatier," 2017. [Online]. Available: https://neocampus.univ-tlse3.fr

[8] G. Weiß, *Multiagent Systems, A modern Approach to Distributed Artificial Systems*. MIT Press, 1999.

[9] G. Di Marzo Serugendo, M.-P. Gleizes, and A. Karageorgos, "Self-organising software: From natural to artificial adaptation," 2011.

[10] J. P. Georgé, B. Edmonds, and P. Glize, *Making Self-organizing adaptive multi-agent systems work*. Kluwer Publishing, 2004, pp. 321–340.

[11] P. A. Corning, "The re-emergence of emergence, and the causal role of synergy in emergent evolution," *Synthese*, vol. 185, no. 2, pp. 295–317, Mar. 2012, wOS:000303530600009. [Online]. Available: https://link.springer.com/article/10.1007/s11229-010-9726-2

[12] T. Finch, "Incremental calculation of weighted mean and variance," *University of Cambridge*, vol. 4, pp. 11–5, 2009.

[13] P. Piché, E. Belghache, N. Verstaevel, and B. Lartigue, "Impact of eco-feedback on the behavior of campus users," *Energy*, p. 8, 2014.

[14] "UCI Machine Learning Repository: Ozone Level Detection Data Set." [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Ozone+Level+Detection

[15] K. Zhang, W. Fan, X. Yuan, I. Davidson, and X. Li, "Forecasting skewed biased stochastic ozone days: Analyses and solutions," in *Proceedings of the Sixth International Conference on Data Mining*, ser. ICDM '06. Washington, DC, USA: IEEE Computer Society, 2006, pp. 753–764.