Available online at www.sciencedirect.com

ScienceDirect

Energy Procedia 126 (201709) 907–914

Energy Procedia

www.elsevier.com/locate/procedia

72nd Conference of the Italian Thermal Machines Engineering Association, ATI2017, 6-8 September 2017, Lecce, Italy

Internal Combustion Engine sensor network analysis using graph modeling

A. Corsini[a], F. Bonacina[a]*, S. Feudo[a], A. Marchegiani[a], P. Venturini[a]

[a]*Department Mechanical and Aerospace Engineering, Sapienza University of Rome, Via Eudossiana 18, I00184 Rome, Italy**

Abstract

In recent years there has been a rapid development in technologies for smart monitoring applied to many different areas (e.g. building automation, photovoltaic systems, etc.). An intelligent monitoring system employs multiple sensors distributed within a network to extract useful information for decision-making. The management and the analysis of the raw data derived from the sensor network includes a number of specific challenges still unresolved, related to the different communication standards, the heterogeneous structure and the huge volume of data.

In this paper we propose to apply a method based on complex network theory, to evaluate the performance of an Internal Combustion Engine. Data are gathered from the OBD sensor subset and from the emission analyzer. The method provides for the graph modeling of the sensor network, where the nodes are represented by the sensors and the edge are evaluated with non-linear statistical correlation functions applied to the time series pairs.

The resulting functional graph is then analyzed with the topological metrics of the network, to define characteristic proprieties representing useful indicator for the maintenance and diagnosis.

Peer-review under responsibility of the scientific committee of the 72nd Conference of the Italian Thermal Machines Engineering Association

Keywords: Sensor Network, Graph Modeling, Big Data Analytics, Complex Systems.

* Corresponding author. Tel.:+39 0773476521.
E-mail address: fabrizio.bonacina@uniroma1.it

Peer-review under responsibility of the scientific committee of the 72nd Conference of the Italian Thermal Machines Engineering Association
10.1016/j.egypro.2017.08.160

1. Introduction

A sensor network (SN) comprises a group of tiny devices and wireless infrastructure that monitor and record conditions in any number of environments. With the recent diffusion of Micro-Electro-Mechanical Systems technology, the SNs has received a significant attention in the real world scenario and several example of their applications are available in multiple fields, such as power grids, smart buildings, industrial process, transport and logistics, military applications, environmental monitoring, human-centric applications, etc. [1,2].

Looking at internal combustion engines, different on-board-diagnostic (OBD) control unit, defined engine control module ECM or alternatively powertrain control module (PCM) are typically installed. Customary OBD system are designed and calibrated to detect single component fault at the required malfunction criteria rather than every combination of multiple component degradations. The basic concept of OBD systems is to result in malfunction indicator light (MIL) illumination after a fault has been detected on two or three consecutive driving cycles [3]. But it is still difficult to detect a fault when the standard operation conditions of any components have only partially reduced its efficiency, especially under critical working conditions (i.e. at high speed and low loads) [4]. In some recent works, multivariate statistical analysis approach to process the monitored variables (i.e. instantaneous engine speed) has been evaluated for detecting the multiple misfire event in multi-cylinder diesel engine [5]. To be able to identify the cause of the fault, multivariate analysis and the principal component analysis (PCA) was used to investigate the relationship between process parameters, energy variables and emissions [6]. Data Analytics for SNs to perform intelligent analysis on sensor-collected data is considered as a complex area where several issues are still unresolved. In fact, these data seem to have the characteristics of the "4 V" (Volume, Velocity, Variety and Value) typical of the big data, due to their massive amount, the speed required for their collection, processing and use, the heterogeneity (variety of communication standards and data formats), their high redundancy and noise [7].

As reported in Zhou et al. [8], traditional data analysis in statistics, data mining, machine learning, data management and data visualization may result inappropriate in dealing with the sensor network big data.

In this paper we propose to apply an analytical procedure to analyze sensor network data gathered from an internal combustion engine (ICE), based on the Complex Network analysis techniques. Similar methods have been successfully applied in biology and biomedicine [9], but in current literature there are no application cases for ICE monitoring systems. For example, in neuroscience, complex networks are used to describe the structure of relationships between the various regions of the brain and create diagnostic models based on recorded measures from functional magnetic resonance imaging (fMRI), electroencephalography (EEG) or magnetoencephalography (MEG) [10].

The main purpose of this paper is to test the effectiveness of these analytical methods to identify different operating conditions of an ICE. For these reason, we performed three different laboratory tests in order to reproduce a standard situation (thus reproducing urban condition) and two possible failures: i) disabling the exhaust gas recirculation (EGR) valve and ii) reducing by 50% the section of the intake air duct.

An accurate description of the proposed method for sensor network data analysis is given in Chapter 2, while in Chapter 3 the case study is defined. Chapter 4 shows the results obtained by applying the method on the case study and finally, in Chapter 5, the conclusions are presented.

2. Data analysis model

The proposed data analysis approach is based on the graph modeling of the sensor network and the subsequent study of the characteristic topological metrics, with the aim of extracting useful information on the analyzed system. Specifically, the method can be summarized in the following steps:

a) Extractions of the time series
b) Data cleaning and recovering
c) Data modeling with an unweighted graph
d) Definition of the functional graph
e) Analysis of the topological metrics of the graph

At the initial stages (a,b), the heterogeneous data acquired by the sensors are temporally re-aligned and any outlier or background noises are removed. In step (c) a fully-connected graph is created in which each monitored variable represents a node and all the nodes are connected to each other with non-directed edges. The output of this preliminary phase is an unweighted graph model of the sensor network.

The next step (d) consists in attributing a weight to all the links, by defining a connectivity matrix represented with a functional graph. This process is based on the evaluation of the correlation between the pairs of time series X1 (t) ... XN (t) recorded from the N different sensor nodes, for each pair of nodes i and j. The magnitude of the statistical correlations represents the weighted links wij of the resulting functional graph (see Fig. 1).

Among the number of the methods to compute those functional links [11], we applied a nonlinear multivariate correlation method that returns a non-directional functional graph, the Mutual Information (MI). The MI is a probabilistic function able to quantitatively analyze the information shared between any two random variables X and Y, and reads as:

$$MI(X;Y) = \sum_{x,y} p(x,y) ln \frac{p(x,y)}{p(x)p(y)} \tag{1}$$

where, p(x), p(y), and p(x,y) are respectively the marginal and the joint probability density functions of X and Y.

Unlike other correlation methods, MI has the great advantage to measure any type of relationship between variables and it is not affected from variability in spatial transformations (the logarithm argument in Eq. 1 is dimensionless) [12]. Measure of similarity is widely used for example in signal analysis and image processing applications [13, 14]. However the MI is based on statistical functions that have to be estimated with a finite number of samples. This estimation can be performed by different approaches among which the histogram based approach has been extensively used for the advantages in terms of low computational complexity [15]. In the present study, MI was calculated within a fixed range of 100 data entry window, sliding over the entire duration of the tests. As such for the monitored variables we got n-100 functional graphs.

Finally, in the last step (e), the topological metrics for each functional graph obtained in the previous point are evaluated. Topological metrics can be classified into three main subclasses: distance, connection, and spectral. The distance class includes the metrics, which provides information about the graph size (e.g. the number of nodes a message has to cross to reach its destination). The connection class includes all the metrics providing information related to the structure of the network (e.g. the number of a node's neighbors), together with metrics that help grouping nodes into clusters or hierarchies. Finally, the spectral class includes the eigenvalues and eigenvectors of a graph [16]. Looking at the specific sensor network, the number of nodes of the functional graphs remains constant, while the correlations between the monitored variables (edge weight) and the edge number (in the case of MI = 0, the edge is automatically removed) vary over time. For this reason, we have considered two topological metrics able to vary according to the number of edges and/or to their specific weight, namely the Diameter and the Average Weighted Degree Centrality (AWDC).

The Diameter is defined as the length $\max_{i,j} d(i,j)$ of the "longest shortest path" between any two graph vertices (i,j), where d(i,j) is a graph distance [17]. The diameter is, then, the maximal distance between all pairs of nodes, when paths which backtrack, detour, or loop are excluded from consideration.

The Weighted Degree Centrality WDC of a node is the sum of weighted values of the edges connecting node i and its neighbors j. It reads as:

$$WDC(i) = \sum_{j}^{N_i} W_{i,j} \tag{2}$$

Where N_i is the set of neighbors of node i and $W_{i,j}$ refers to the weighted value of edge between node i and its neighbor j [18, 19]. As previously defined, the WDC is a specific attribute of the single node, so for a unique characterization of the graph we have considered the Average Weighted Degree Centrality.

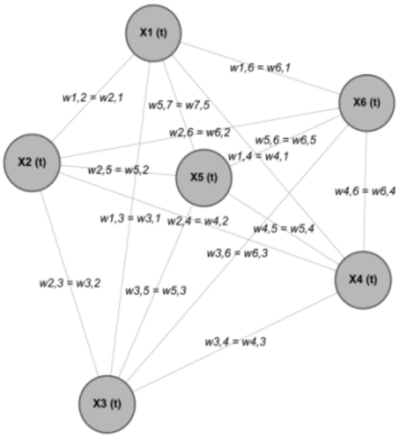

Figure 1. Example of a functional graph structure of a 6-node sensor network at time t.

Table 1. Characteristics of the engine used for tests.

Engine	1.9 JTD
Cycle	Diesel, 4 strokes
Total displacement	1910 cm^3
Number of cylinders	4
Max. Torque	205 Nm @ 1500 rpm
Max. Power	77 kW @ 4000 rpm
Injection type	Common-rail, Bosch Unijet

3. Case Study

3.1 Experimental setup

A four strokes, common-rail, turbocharged Diesel engine with a JTD injection system was used for the tests. The main characteristics of the engine are reported in Table 1.

The engine was installed in the Department of Mechanical and Aerospace Engineering Laboratory, at the Sapienza University of Rome, Italy. In order to install the engine on the test bench some parts were removed or modified. In particular, the air intake was connected to the room ventilation system so that the intake air is at constant temperature (about 25 °C) and pressure (101500 Pa). Moreover, flywheel and gearbox have been removed.

The test bench is equipped with a Schenk hydraulic brake. A Cardan joint connects the engine drive shaft and the hydraulic brake. Despite the modifications reported, the engine is equipped with the original auxiliary and injection systems and its own ECU. Therefore, the engine can be considered in automotive configuration. Further details on the experimental setup can be found in [20, 21].

3.2 Measuring chain

Figure 2 shows a sketch of the measuring chain used in the present study. A specific in-house software (coded in Python) was developed for data acquisition. It couples the data coming from ECU with those measured by the Bosh BEA 350 emissions monitoring unit (EMU). The code is upgraded respect of the one already realized for the acquisition of all the engine parameters handled by the ECU [23]. Since the EMU allows only few recording of the measured data and it cannot be directly connected to a PC, we adopted a video capture and recognize approach. Each time the software acquires data from the ECU, through a standard webcam it also takes a picture of the EMU monitor runtime showing emissions data. At the end of each test, a post-processing tool using OCR (Optic Character Recognition) technique analyzes the photos and converts the information they contain into numbers. In this way, we couple the engine and emissions data. The software was implemented on Raspberry Pi system.

The most relevant engine parameters are thus acquired by connecting a PC to the ECU and (indirectly) to the EMU, while torque and power are measured through the hydraulic brake of the test bench. Table 2 summarizes all the monitored variables.

3.3 Experiments details

In some previous publications (see for instance [22, 23]), the engine behavior was studied performing a series of tests. Torque, power, specific fuel consumption and some other relevant quantities have been measured at different pedal position in order to explore the complete map of the engine.

Here we focus our attention on the engine behavior under some simple failure situations. In particular, we wanted to reproduce somehow an urban situation. In this conditions, the gas pedal position (gpp) rarely goes over 35-40%, thus we decided to test the engine at an average gpp, namely 20%. Then, the engine speed was set to 3000 rpm which is the value at which the engine exerts about its maximum power for the given gpp. Assuming these operative points, we performed three different experiments in order to reproduce a standard situation and two possible failures. In the standard situation, the engine works without any modification; a first failure was reproduced disabling the exhaust gas recirculation (EGR) valve. The last test is performed reducing by 50% the section of the intake air duct, thus reproducing a clogged air filter. The engine was brought to the chosen operating points and left at this condition for 30 min. During the experiments, the data acquisition frequency was about 0.125 Hz. Table 3 summarizes the operative conditions of the performed tests.

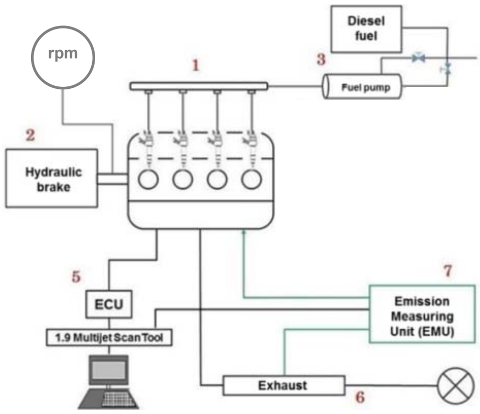

Figure 2. Layout of the experimental setup and measurement chain.

Table 2. List of the monitored variables with the specific meter used for the measurements.

1 = Engine speed [rpm]	8 = Inj. FuelQuantity-Main [mm3/str]	15 = Boost Pressure [mbar]	22 = CO2 [% vol.]
2 = Engine Temp.[°C]	9 = Inj. Time (Pre) [ms]	16 = EGR Valve Opening [%]	23 = O2 [% vol.]
3 = Gas Pedal Position [%]	10 = Inj. Time (Main) [ms]	17 = Fuel Pressure [bar]	24 = NO [ppm]
4 = Fuel Temperature [°C]	11 = Inj. Advance (Pre) [deg]	18 =Oil Temperature [°C]	25 = Brake Power [kW]
5 = Fuel Consumption [l/h]	12 = Inj. Advance (Main) [deg]	19 = CO [% vol.]	
6 = Total FuelQuantity [mm3/str]	13 = Intake Air Quantity [mg/str]	20 = Lambda [-]	
7 = Inj. FuelQuantity-Pre [mm3/str]	14 = Air Temp. Boost Manifold [°C]	21 = HC [ppm]	

Meters used: Variables 1-18 OBD Sensors; Variables 19-24 Emission Analyzer; Variable 25 Hydraulic Brake.

4. Results

Figure 3 shows torque (Nm), power (kW) and specific fuel consumption (g/kWh) curves at full load. Maximum torque equals about 260 Nm at 2250 rpm; maximum power equals 73 kW at about 3000 rpm; minimum fuel consumption is achieved at about 2250 rpm and equals 170 g/kWh.

Figure 4 shows the functional graphs obtained by applying the SN model to the data recorded during the tests at the investigated three operating conditions. The graphs at the top of Fig. 4 are obtained by applying the Force Atlas Algorithm [24], which visually emphasizes the complementarity between nodes (the most influential nodes are grouped at the center of the graph, while the least influential ones are placed at greater gradual distances).

The lower part of Fig. 4 shows the graphs obtained for the same three different operating conditions by applying the Circular Layout [24], which has the advantage of clearly ranking all the existing links between the pairs of nodes and emphasizing the most important ones.

The different colors of the nodes are related to the specific monitoring systems adopted (OBD sensors, emission analyzer or hydraulic brake). The size and color intensity of the nodes are based on the Weighted Degree Centrality value, while the thickness and the color intensity of the edges are based on the value of the weight calculated with the mutual information.

Table 3. Testing conditions.

Name	Fuel	Gpp (%)	Engine Speed (rpm)	EGR	Air intake section
Standard	gasoil	20	3000	ON	100 %
50% clogged air filter	gasoil	20	3000	ON	50 %
EGR valve OFF	gasoil	20	3000	OFF	100 %

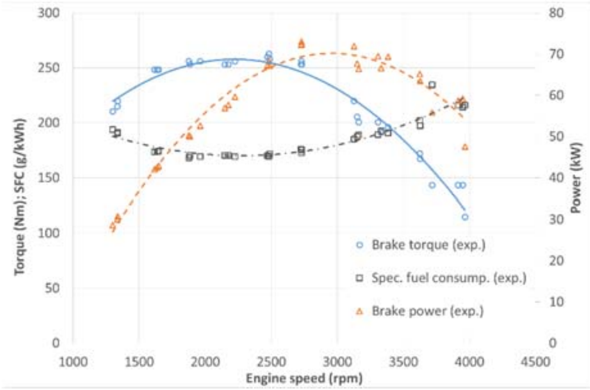

Figure 3. Torque, power and specific fuel consumption curves of the tested engine at full load.

As visible in the Force Atlas, Figure 4.a, the main influencers in standard operating conditions are the brake power (node 1), the fuel temperature (node 4) and the EGR valve opening (node 16). When moving to the Circular Layout, then, it is possible to infer also a strong correlation between the engine speed (node 1), the brake power (node 25) and the set-point of the EGR valve.

When the EGR valve is switched off (Figure 4.b), in addition to the expected decrease of the influence of the specific variable (node 16), the failure of the valve (which also regulates the emissions) is triggered by the lambda probe (node 20) which emphasizes its importance on the graph. The difference in operation of the lambda probe with respect to the standard case implies a different regulation of the input air and the amount of fuel in the main injection, both from the ignition advance (node 12) and the injection fuel quantity (node 8).

When clogging the air filter (Figure 4.c), the system recognizes the insufficient amount of air according to the rotation regime, so the weight of node 1 and node 25 decreases respect to the standard conditions.

In order to provide some hints, Figure 6 shows the time plot of the two topological metrics, respectively the Diameter (Figure 5.a) and the Average Weighted Degree (Figure 5.b), calculated for all the graphs considered for each operating condition.

Notably, both Diameter and Average Weighted Degree temporal plot remains clearly distinct within specific ranges corresponding to the three tested operating conditions in the monitoring interval. The observation of the averaged values, summarized in Table 3, it is possible to note that each operating condition is uniquely identified by specific topological values.

5. Conclusions

In this paper an analytical model was proposed for heterogeneous data gathered from sensor networks installed in ICE. The approach derives from the analysis of signals typically adopted in neuroscience and is based on the graph modeling of the network where the edge weights are defined by the evaluation of the MI between all the pairs of nodes. Finally, the topological metrics of the graph are extrapolated, with the aim of identifying the synthetic indicators characteristic of a given condition of the monitored system. Specifically, the method was tested for the identification of two abnormal operating conditions in the ICE: 50% clogged air filter and EGR valve OFF.

For the laboratory tests, various monitoring systems (OBD sensors, emission analyzer and hydraulic brake) have been installed, each with a specific data acquisition and transmission protocol, for a total duration of 30 minutes and a frequency of acquisition of 0.125 Hz. The results show that the method is able to provide identification patterns of each of the analyzed conditions represented by a specific graph characterized by well-defined correlations between the pairs of monitored variables (see Fig. 4). Furthermore, through the determination of the topological metrics of the graphs (Diameter and Average Weighted Degree) it is possible to define synthetic indicators that uniquely identify the condition investigated (see Fig. 5 and Table 4). The analysis of these indicators can be used as basis to develop both a sensor failure check strategy and a malfunction prediction. The diameter of the graph and the average weighted degree, together with other information provided by functional graphs of several operating conditions, can provide a combination of indicators that are unique or typical of few specific operating conditions. On these bases, one can develop a tool to predict possible malfunction of the system.

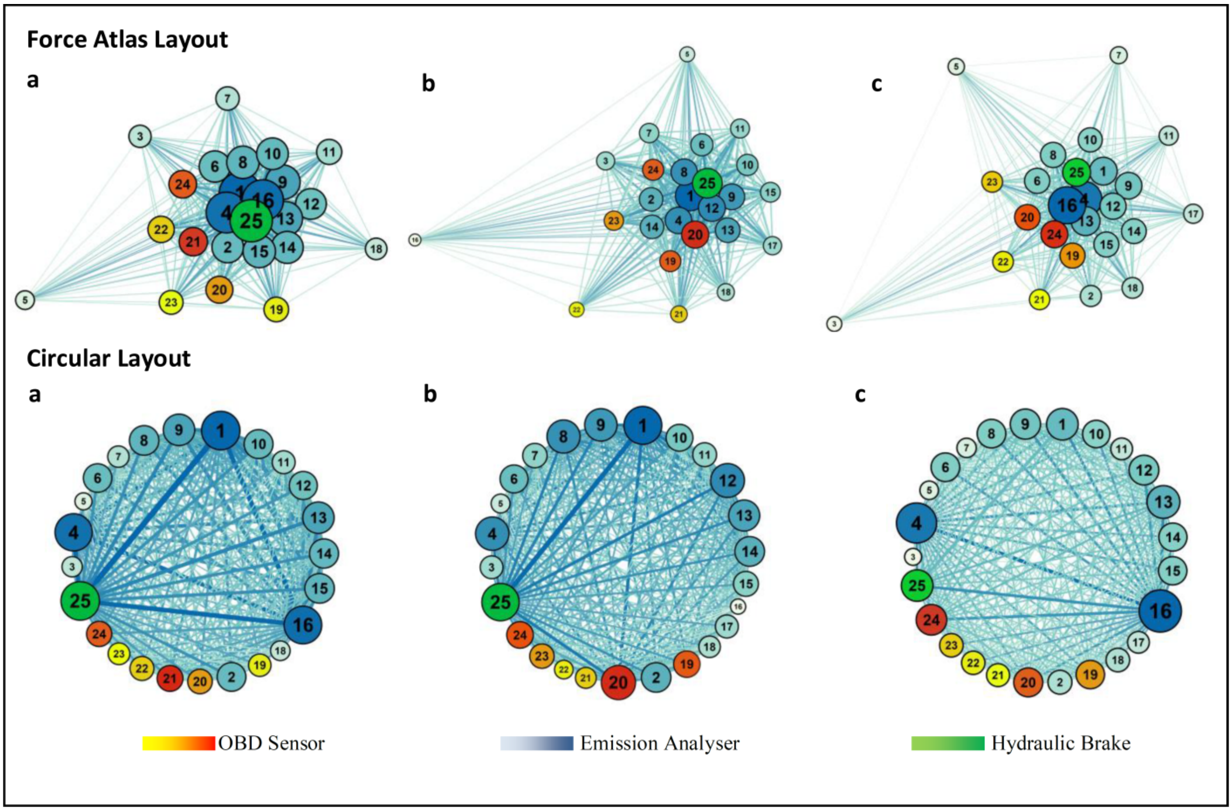

Figure 4. Representation of the functional graph by Force Atlas Layout and Circular layout for 3 different operating conditions: (a) Standard; (b) EGR valve opening; (c) 50% clogger air filter. The size and color intensity of the nodes are based on the weighted degree, while the thickness and the color intensity of the edges are based on the weight measured by the MI.

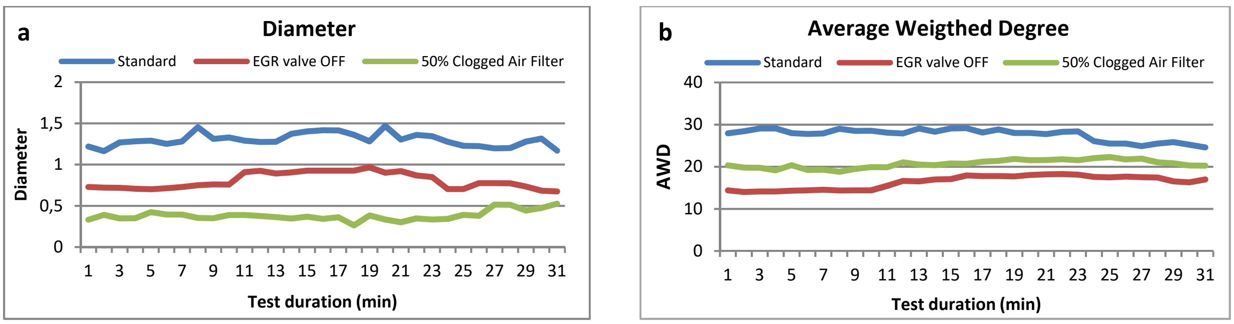

Figure 5. a. Diameter of the graph; b. Average Weighted Degree. Trend over time for the 3 different operating conditions.

For future developments, it would be interesting to test the method by simulating other types of typical failure of ICE, or increase the data acquisition frequency in order to characterize the engine behavior in unsteady state and transitional conditions.

Table 4: Summary of topological metrics (Diameter and AWD) evaluated for each investigated operating condition.

	Standard	50% Clogged air Filter	EGR Valve OFF
AverageWeightedDegree	27,49	20,39	15,45
Diameter	1,30	0,38	0,81

References

[1] J. Yick, B. Mukherjee,D. Ghosal. (2008) "Wireless sensor network survey." *Computer Networks*, vol. 52: 2292–2330.

[2] T. Arampatzis, J. Lygeros, S. Manesis (2005) "A Survey of Applications of Wireless Sensors and Wireless Sensor Networks." *Proceedings of the 13th Mediterranean Conference on Control and Automation Limassol*, Cyprus, June 27-29.

[3] J. Mohammadpour, M. Franchek, and K.Grigoriadis (2011) "A Survey on Diagnostics Methods for Automotive Engines." *American Control Conference on O'Farrell Street*, San Francisco, CA, USAJune 29 - July 01.

[4] G.Rizzoni, S.Onori, M.Rubagottiand (2009) "Diagnosis and Prognosis of AutomotiveSystems: motivations, history and someresults." *Proceedings of the 7th IFAC Symposium on Fault Detection*, Supervision and Safety of Technical ProcessesBarcelona, Spain, June 2009;

[5] C. Hun, A. Li, X. Zhao (2010) "Multivariate statistical analysis strategy for multiplemisfire detectionininternalcombustionengines." *Journal of Mechanical Systems and Signal Processing* , Elsevier vol. 25: 694-703;

[6] A.Corsini, F. Bonacina, S. Feudo, F. Lucchetta, A. Marchegiani (2016) "Multivariate KPI for Energy Management of Cooling Systems in Food Industry." *Energy Procedia*, Proceedings of the 71st Conference of the Italian Thermal Machines Engineering Association, ATI2016, 14-16 September, Turin, Italy, Vol. 101: 606 – 613.

[7] D. Rajeshwari (2015) "State of the Art of Big Data Analytics: A Survey." *International Journal of Computer Applications*, vol. 120: 0975–8887.

[8] K. Zhou, C. Fu, S. Yang (2016) "Big data driven smart energy management: From big data to big insights." *Renewable and Sustainable Energy Reviews* ,vol. 56: 215–225.

[9] F. De Vico Fallani, J. Richiardi, M. Chavez, S. Achard (2014) "Graph analysis of functional brain networks: practical issues in translational neuroscience." *Phil. Trans. R. Soc.* B 369: 20130521.

[10] E. Eqlimi, N. RiyahiAlam, M.A. Sahraian, A. Eshaghi, S. RiyahiAlam, H. Ghanaati, K. Firouznia, E.Karami (2014) "Resting State Functional Connectivity Analysis of Multiple Sclerosis and Neuromyelitis Optica Using Graph Theory." *IFMBE proceedings*, vol. 41: 206-209.

[11] M. Zanina, D. Papoc, P.A. Sousa , E. Menasalvas , A. Nicchi, E. Kubike, S. Boccaletti (2016) "Combining complex networks and data mining: Why and how." *Physics Reports*, vol. 635: 1–44.

[12] P. A. Estévez, M. Tesmer, C. A. Perez, J. M. Zurada (2009) "Normalized Mutual Information Feature Selection." *IEEE Transaction On Neural Network*, vol. 20: 189-201.

[13] T. Drugman, , M. Gurban, J. P.Thiran (2007) "Relevant Feature Selection for Audio-visual Speech recognition." *Proc. of the International Workshop on Multimedia Signal Processing*, Crete, Greece: 179 - 182.

[14] J. P. W. Pluim, J. B. A.Maintz, M. A. Viergever (2003) "Mutual Information Based Registration of Medical Images: A Survey." *IEEE Transactions on Medical Imaging*, vol. 22, no. 8, August: 986 - 1004.

[15] P. A. Legg,,P. L. Rosin, D. Marshall, J.E. Morgan (2007) "Proc Improving Accuracy and Efficiency of Registration by Mutual Information using Sturges Histogram Rule." *Proc. Medical Image Understanding and Analysis*, Aberystwyth: 26-30.

[16] J. M. Hernández, P. Van Mieghem "Classification of graph metrics. Hernández." *Delft University of Technology*, Tech. Rep.

[17] D. B. Westn (2001) "Introduction to Graph Theory, 2nd ed." *Englewood Cliffs, NJ: Prentice-Hall.*

[18] T. Opsahl, F. Agneessens, J. Skvoretz (2010) "Node centrality in weighted networks: Generalizing degree and shortest paths." *Social Networks 32*. 245–251.

[19] X. Tang, J. Wang, J. Zhong, Y. Pan (2015) "Predicting Essential Proteins Based on Weighted Degree Centrality." *IEEE/ACM Transacions on Computational Biology and Bioinformatics*, Vol. 11: 407-418.

[20] A. Corsini, G. Fanfarillo, F. Rispoli, P. Venturini (2015) "Pollutant emissions in common-rail diesel engines in extraurban cycle: rapeseed oils vs diesel fuel." *Energy Procedia, 8*: 141-148.

[21] A. Corsini, A. Marchegiani, F. Rispoli, F. Sciulli, P. Venturini (2015) "Vegetable oils as fuels in Diesel engine. Engine performance and emissions." *Energy Procedia, Proceedings of the 69th Conference of the Italian Thermal Machines Engineering Association*, Vol. 81: 942-949.

[22] A. Corsini, R. Di Antonio, G. Di Nucci, A. Marchegiani, F. Rispoli, P. Venturini (2016) "Diesel engine fuelled with different blends of waste cooking oil and gasoil."*Energy Procedia, Proceedings of the 71st Conference of the Italian Thermal Machines Engineering Association*, ATI2016, 14-16 September, Turin, Italy, Vol. 101: 606 – 613.

[23] A. Corsini, V. Giovannoni, S. Nardecchia, F. Rispoli, F. Sciulli and P. Venturini (2012) "Performance of a common-rail Diesel engine fuelled with rapeseed and waste cooking oils." *PROCEEDINGS OF ECOS 2012 - The 25th International Conference on Efficiency, Cost, Optimization, Simulation and Environmental Impact of Energy Systems*, June 26-29, 2012, Perugia, Italy.

[24] Bastian M., Heymann S., Jacomy M. (2009). "Gephi: an open source software for exploring and manipulating networks." *International AAAI Conference on Weblogs and Social Media.*