



Network science meets respiratory medicine for OSAS phenotyping and severity prediction

Stefan Mihaicuta¹, Mihai Udrescu², Alexandru Topirceanu² and Lucretia Udrescu³

¹ Department of Pulmonology, Victor Babes University of Medicine and Pharmacy, Timisoara, Romania

² Department of Computer and Information Technology, University Politehnica of Timisoara, Timisoara, Romania

³ Faculty of Pharmacy, Victor Babes University of Medicine and Pharmacy, Timisoara, Romania

ABSTRACT

Obstructive sleep apnea syndrome (OSAS) is a common clinical condition. The way that OSAS risk factors associate and converge is not a random process. As such, defining OSAS phenotypes fosters personalized patient management and population screening. In this paper, we present a network-based observational, retrospective study on a cohort of 1,371 consecutive OSAS patients and 611 non-OSAS control patients in order to explore the risk factor associations and their correlation with OSAS comorbidities. To this end, we construct the Apnea Patients Network (APN) using patient compatibility relationships according to six objective parameters: age, gender, body mass index (BMI), blood pressure (BP), neck circumference (NC) and the Epworth sleepiness score (ESS). By running targeted network clustering algorithms, we identify eight patient phenotypes and corroborate them with the co-morbidity types. Also, by employing machine learning on the uncovered phenotypes, we derive a classification tree and introduce a computational framework which render the Sleep Apnea Syndrome Score (SAS_{Score}); our OSAS score is implemented as an easy-to-use, web-based computer program which requires less than one minute for processing one individual. Our evaluation, performed on a distinct validation database with 231 consecutive patients, reveals that OSAS prediction with SAS_{Score} has a significant specificity improvement (an increase of 234%) for only 8.2% sensitivity decrease in comparison with the state-of-the-art score STOP-BANG. The fact that SAS_{Score} has bigger specificity makes it appropriate for OSAS screening and risk prediction in big, general populations.

Submitted 9 October 2016

Accepted 10 April 2017

Published 9 May 2017

Corresponding author

Mihai Udrescu, mudrescu@gmail.com

Academic editor

Massimiliano Zanin

Additional Information and
Declarations can be found on
page 21

DOI 10.7717/peerj.3289

© Copyright

2017 Mihaicuta et al.

Distributed under

Creative Commons CC-BY 4.0

OPEN ACCESS

Subjects Bioinformatics, Computational Biology, Health Policy, Respiratory Medicine, Computational Science

Keywords Network science, Sleep apnea, Phenotypes, Prediction score, Prediction specificity

INTRODUCTION

Obstructive Sleep Apnea Syndrome (OSAS) is a serious clinical disorder caused by abnormal breathing pauses that occur during sleep; this results in sleep fragmentation and excessive daytime somnolence (*Simon & Collop, 2012; Fischer et al., 2012; Lévy et al., 2014*). There are studies reporting the epidemic incidence of OSAS, with worrying increasing rates over the last 20 years (*Young, Peppard & Gottlieb, 2002; Punjabi, 2008;*

Peppard et al., 2013). If not properly diagnosed and treated, OSAS increases the morbidity and perioperative risks (Memtsoudis, Besculides & Mazumdar, 2013; McNicholas, Bonsignore & of EU Cost Action B26, 2007; Rossi, Stradling & Kohler, 2013; Utriainen et al., 2013; Sánchez-de-la Torre, Campos-Rodriguez & Barbé, 2013). Indeed, when it remains undetected, OSAS rapidly creates serious cardiovascular, respiratory and nutritional problems (McNicholas, Bonsignore & of EU Cost Action B26, 2007; Yaggi et al., 2005; Bakker, Montesi & Malhotra, 2013). Therefore, it is essential that OSAS be detected at an early stage, which can only be achieved through preventive actions such as extensive population screening (Pelletier-Fleury et al., 2004).

Apnea severity is indicated by the Apnea-Hypopnea Index, *AHI*; this represents the number of breathing pauses of at least 10 s, recorded over one hour of sleep. As such, any patient can be classified in one of the following *AHI* categories: normal or low-risk apnea (*L*) for $AHI < 5$, mild sleep apnea (*Mi*) for $5 \leq AHI < 15$, moderate sleep apnea (*Mo*) for $15 \leq AHI < 30$, and severe sleep apnea (*Se*) when $AHI \geq 30$.

Usually, polysomnography (PSG) is used as the reference method (i.e., gold standard) of OSAS diagnosis, which is based on measuring *AHI*. However, PSG is expensive, time-consuming, and generally not adequate for population screening (Simon & Collop, 2012; Schlosshan & Elliott, 2004). Since OSAS has a significant prevalence and the PSG-based exhaustive investigation is not feasible when screening a large population, OSAS predictors are preferred for monitoring (Udrescu et al., 2014; Topirceanu et al., 2014).

In current medical practice, there are three major predictive models based on questionnaires, namely Berlin, STOP, and STOP-BANG (Netzer et al., 1999; Silva et al., 2011; Farney et al., 2011; Chung et al., 2008a; Chung et al., 2008b). Published studies indicate STOP-BANG as the best available predictive score, due to its high sensitivity: 83.6% for $AHI > 5$, 92.9% for $AHI > 15$, and 100% for $AHI > 30$. However, STOP-BANG has a low specificity (56.4% for $AHI > 5$, 43% for $AHI > 15$, and 37% for $AHI > 30$) (Chung et al., 2008a; Chung et al., 2008b; Chung et al., 2014) which prevents the usage of this score for population screening. Although there are notable attempts for improving STOP-BANG's specificity (Chung et al., 2014), they are mainly targeting narrow-type cohorts such as perioperative patients.

Consequently, our paper aims at analysing the general case, (i.e., with all patient categories from a given geographical area being taken into account for screening), and not just some specific cohorts. To this end, our research is underpinned by a complex network perspective on uncovering OSAS phenotypes. Indeed, network science is already successfully used in medicine at the disease-level (Barabási, 2011), including respiratory applications (Diez, Agustí & Wheelock, 2014; Faner et al., 2014; Divo et al., 2015). Our network-based approach on OSAS risk factors allows for better, more accurate OSAS phenotype identification, which in turn leads to a new predictive score (SAS_{Score}). In comparison with the state-of-the-art, our OSAS risk prediction score achieves significantly better specificity in predicting actual *AHI* categories, which makes our SAS_{Score} very appropriate for screening big populations as part of preventive medicine programs.

METHODS

Network Medicine has received a lot of attention during the last decade (*Menche et al., 2015*; *Goh et al., 2007*; *Barabási, 2007*; *Vidal, Cusick & Barabasi, 2011*); this trend is fuelled by the fact that complex network science can bring significant advances in various medical fields like genomics (*Sharma et al., 2013*; *Rozenblatt-Rosen et al., 2012*), drug-target interaction (*Yildirim et al., 2007*), or cell metabolism (*Barabasi & Oltvai, 2004*; *Han, 2008*). Consequently, it has been recently suggested that network medicine can be also used for addressing important problems in respiratory medicine (*Faner et al., 2014*; *Divo et al., 2015*).

Description of databases

In order to use complex network tools for OSAS research, we need real-world OSAS patient datasets. Unfortunately, OSAS patients datasets are scarce and not public; such a situation is justified by multiple aspects: big data techniques were only recently considered as tools for respiratory medicine and OSAS, all patients must undergo hospital polysomnography (which entails a complex, expensive and time-consuming process), while coordinated research efforts for gathering data were only recently introduced.

For instance, the biggest such OSAS database, namely European Sleep Apnea DAtabase—ESADA (*Hedner et al., 2011*), is not public and it has gathered data from 15,956 patients in 24 sleep centers from 16 countries since 2007. Also, a recent OSAS study (*Marti-Soler et al., 2016*) where the validation is similar to our approach uses only one (private) validation database, comprising 1,101 patients (*Santos-Silva et al., 2012*).

As a result, in order to perform network investigation on OSAS, we built our own Apnea Patients Database (APD) consisting of consecutive patients with suspicion of sleep breathing disorders, which were evaluated at Victor Babes Regional Hospital in Timisoara (Western Romania) between March 2005 and March 2012 under the supervision of the hospital's Ethics Committee (internal briefing note no. 10/12.10.2013). At the initial visit, the study protocol was clearly explained to obtain the patient's consent and the acceptance of referral physicians. Subsequently, respiratory polygraphy was performed using both Philips Respironics's Stardust polygraph (2005) and MAP's POLY-MESAM IV (1998). PSG was carried out with Philips Respironics' Alice 5 Diagnostic Sleep System, according to the appropriate guidelines (*Rechtschaffen & Kales, 1968*). The polygraphy was performed both at home and at the hospital, whereas PSG measurements were performed at the hospital under medical supervision. To preserve the information accuracy, all collected data were carefully verified; throughout this process, we have ensured complete data confidentiality. Our observational, retrospective study employs only standardized non-invasive procedures that exclude all useless investigations. Moreover, visits did not entail additional effort for the patients or supplemental budget for the clinic.

All 1,371 patients that completed the sleep study protocol and signed informed consent are included in the APD, each with corresponding 108 breathing parameters and anthropometric measurements. The APD distribution of measured *AHI* is presented in [Fig. 1](#).

In order to verify if there is any difference between apnea and non-apnea populations in terms of how risk factors associate and converge, we built a 611 people non-OSAS database

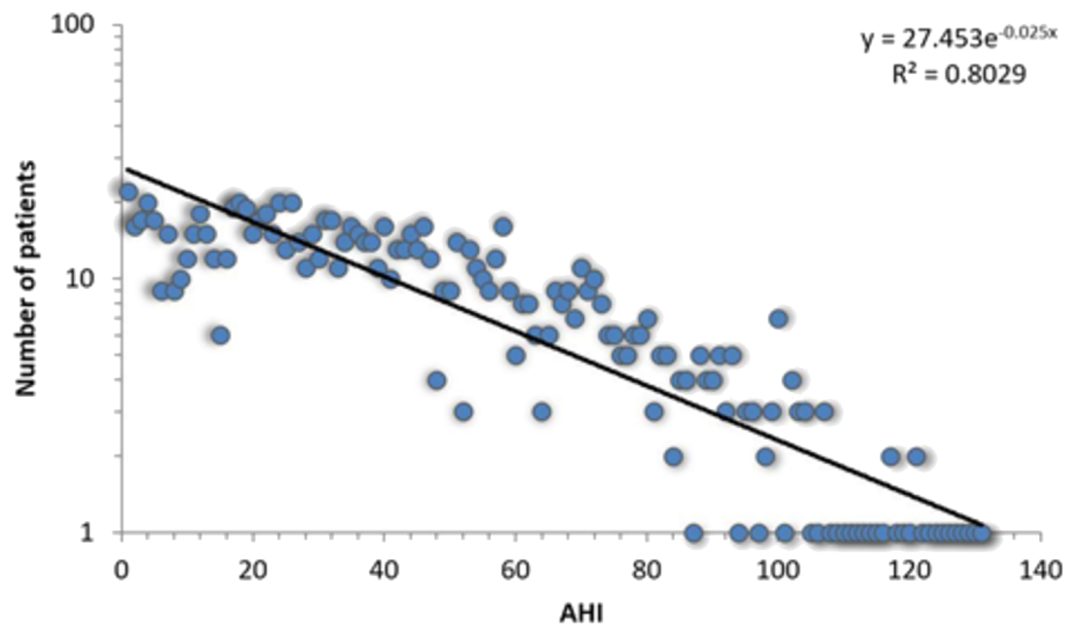


Figure 1 Log-linear histogram representing the number of patients with a certain Apnea-Hypopnea Index (AHI) value, along with the corresponding distribution fit. The best fit is the exponential function $f(x) = 27.453 \cdot e^{-0.025x}$.

NAD (using the same procedure as for the APD). Also, to evaluate the prediction score derived from our study, we gathered a distinct test database TD (over a distinct period of time: the fall of 2013) consisting of 231 patients by following the same procedure. [Figure 2](#) presents the distinct roles of our three databases, as well as the relationship between them.

Analysis of APD and TD

As patients within TD are used to validate our OSAS prediction with SAS_{Score} , which was obtained by processing patients from APD, we analyse if the distribution of parameters in TD is not too close to the corresponding distributions in APD. Such an investigation is required considering that, although data for the two databases were gathered over distinct periods of time, all measurements were performed in a given geographical region.

To this end, we present the distributions of the most relevant parameters in our research (Age A , Body Mass Index BMI , Neck Circumference NC , High Blood Pressure HBP , and Epworth Sleepiness Score ESS) within the validation population (TD) and the apnea patients database (APD) in [Table 1](#) under the form of measured averages and their corresponding standard deviations, as well as Gini coefficients. We rely on Gini coefficients for a quantitative measure of data dispersion.

We also provide a visual comparison of AHI and relevant risk factor parameters distributions in APD and TD (see [Fig. 3](#)). All these results show that data were randomly gathered, so that the main parameters are normally distributed. However, Gini coefficients (especially for A , BMI and ESS) indicate an important difference between APD and TD distributions. Moreover, [Fig. 3](#) shows a significantly different AHI histogram for TD in comparison with APD. As such, in APD there are many patients with $AHI > 120$, whereas

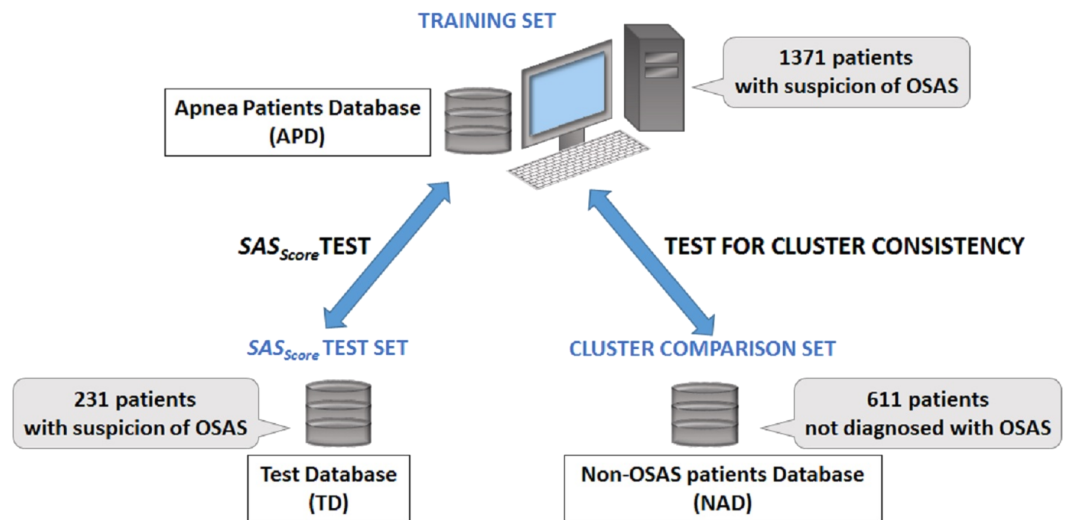


Figure 2 Description of databases used in our research. The main Apnea Patients Database (APD), comprising 1,371 consecutive patients which arrive at the hospital with suspicion of OSAS between 2005 and 2012, is used to build patient phenotypes and to render the SAS_{Score} . The distinct Test Database (TD), comprising 231 consecutive patients which arrive at the hospital with suspicion of OSAS in 2013, is used to verify the sensitivity and specificity of predicting patient's AHI and OSAS categories. The Non-OSAS patients Database (NAD) uses consecutive assessed people whom are not diagnosed with OSAS during the spring 2015—summer 2016 period, in order to test for cluster consistency (i.e., compare how risk factors converge in clusters for OSAS patients in comparison with people without OSAS).

Table 1 AHI and relevant risk factor parameters distribution in the Apnea Patients Database (APD) and Test Database (TD), given as average values plus standard deviation, as well as Gini coefficients. We only considered boolean values for the High Blood Pressure (i.e., if the patient has high blood pressure or not); in this case, we provided the percentages of people with high blood pressure.

Parameter	APD		TD	
	Average	Gini	Average	Gini
AHI	40.04 ± 27.83	0.388	44.39 ± 26.13	0.334
BMI	33.01 ± 7.94	0.576	32.62 ± 7.41	0.468
A (yrs.)	51.73 ± 12.44	0.539	52.01 ± 13.61	0.397
NC (cm.)	42.01 ± 9.94	0.51	42.64 ± 5.12	0.482
HBP	67.47%	N/A	64.93%	N/A
ESS	10.63 ± 5.38	0.362	11.38 ± 5.19	0.422

in TD there is none such patient. Also, in APD the largest number of patients associated to an AHI value correspond to AHI values <20; in contrast, in TD, the largest number of patients with a given AHI value correspond to AHI values around 40.

Building the patient network

An unweighted network is a graph (V, E) , which consists of a set of vertices (or nodes) V and a set of edges (or links) E that represent connections $[v, w] \in E$ between certain pairs of vertices $v, w \in V$. We build the unweighted Apnea Patients Network (APN), by assigning vertices and edges: each node corresponds to a distinct patient in our OSAS

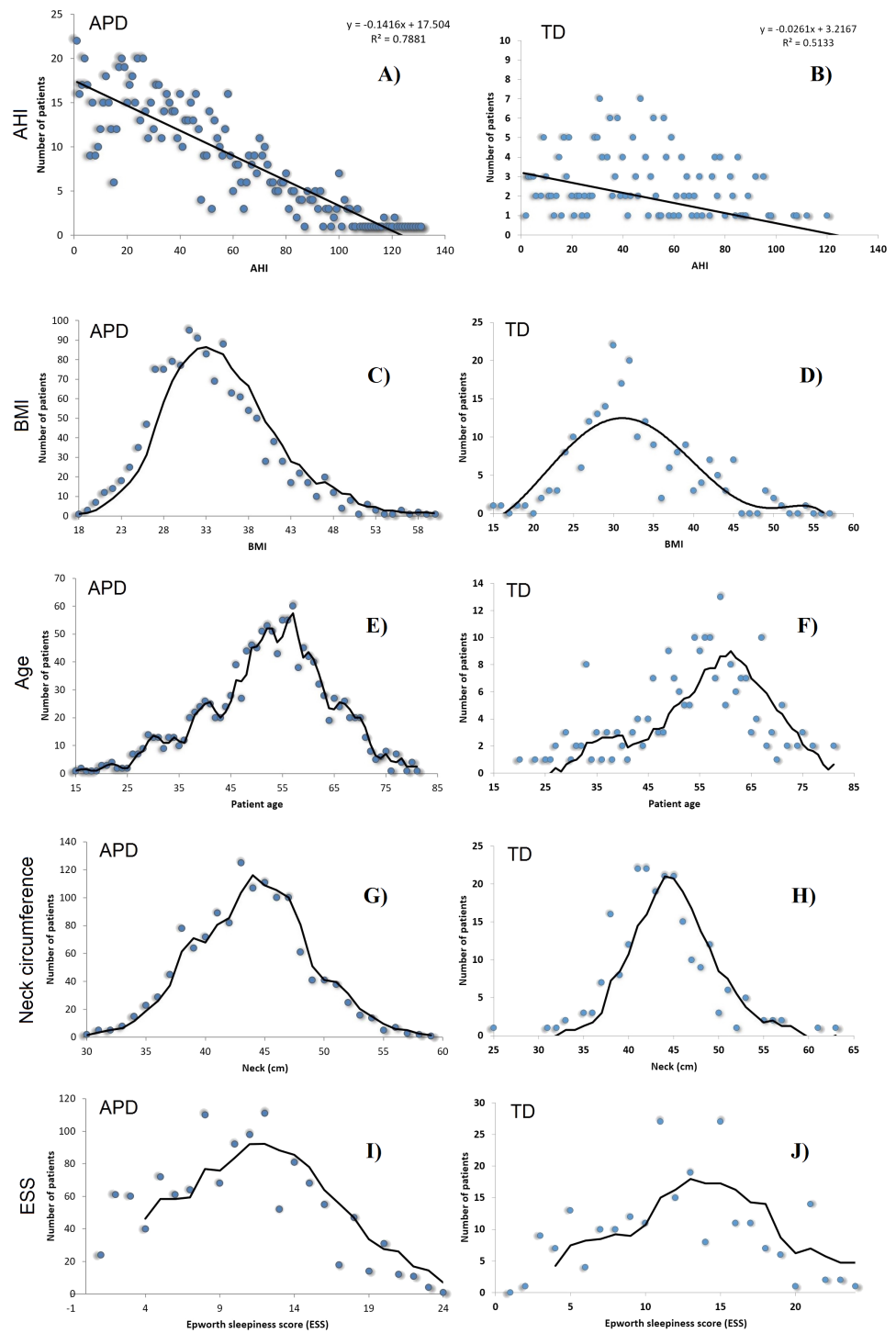


Figure 3 AHI distributions in APD (A, C, E, G, I) and TD (B, D, F, H, J), as well as the normal distributions of BMI, Age A, Neck Circumference NC and ESS, for patients in both APD and TD.

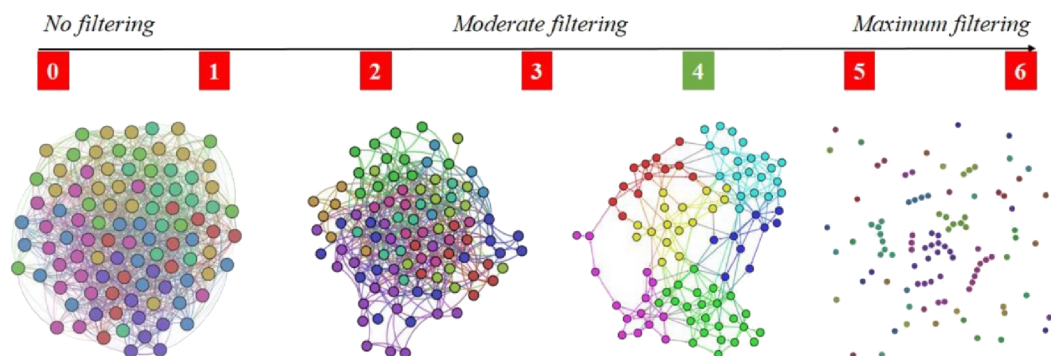


Figure 4 Apnea Patients Network (APN) edge filtering, by considering different definitions for $f_{RFC} = 1$, when we adopt the x -out-of-6 criteria ($x = 1, 2, 3, 4, 5$, and 6). The visual result indicate $x = 4$ as the best solution, because the edge density is convenient for rendering topological clusters with energy model layouts. Other values for x will generate too dense or too sparse networks.

patients database APD, while an edge (link) is created between two vertices if there is a risk factor compatibility between the patients represented by the two vertices (nodes).

The risk factor compatibility is a binary function $f_{RFC} \in \{0, 1\}$ (0 means incompatibility and 1 means compatibility) based on six parameters with high relevance for OSAS: age, gender, BMI, neck circumference, blood pressure (systolic and diastolic), and Epworth Sleepiness Score. We build our APN by considering that $f_{RFC} = 1$ if at least four out of six parameters are identical; otherwise $f_{RFC} = 0$.

The six parameters are selected from the pool of all relevant risk factors (all measured parameters can be found in the [Supplemental Information 1](#) because they can be measured easily and objectively; such objective measurements can be performed anywhere, and are widely accepted in the medical literature (Lévy *et al.*, 2014). In contrast, other scores consider snoring and witnessed apnea episodes as factors, but these are parameters which cannot be observed or measured objectively.

The reason for adopting the 4-out-of-6 criterion is that it assures the right amount of link density in the APN, meaning that there are enough links so that the APN is connected, but not too many links so that communities (i.e., clusters) can be rendered with energy model layouts (Noack, 2009). As Fig. 4 shows that the 4-out-of-6 link filtering represents the best alternative, we use this criterion to build the APN. To the best of our knowledge, this link filtering procedure is original and has not been used before in such network-based approaches.

APN clustering

We clustered the APN by using a dual clustering methodology: energy-model layouts plus modularity classes, similar to the approach from (Udrescu *et al.*, 2016; Udrescu *et al.*, 2014). Energy-models are force directed network layout algorithms, namely visual tools that assign certain positions in the Euclidian space to both nodes and edges (Noack, 2009). To this end, we used the Force Atlas 2 algorithm (Jacomy *et al.*, 2014) as the network layout; this new layout is very effective in clustering various types of complex networks, as it is based on previous theoretical foundations of force directed attraction–repulsion algorithms

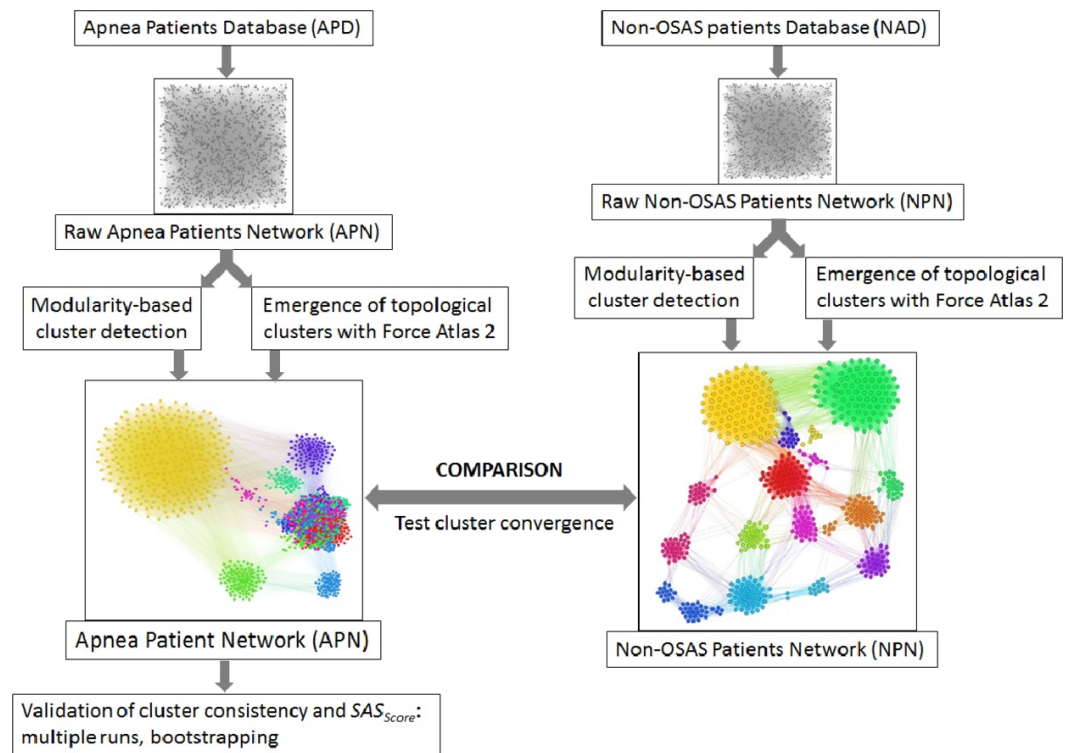


Figure 5 Overview of the proposed dual clustering methodology (modularity classes represented with distinct colors and topological clusters emerged from running the force directed layout Force Atlas 2), along with testing cluster formation against non-OSAS control patients (i.e., showing that risk factors converge differently for OSAS and non-OSAS patients), cluster consistency and SAS_{Score} validation.

(Fruchterman & Reingold, 1991; Noack, 2003). Indeed, Force Atlas 2 is clustering complex networks by producing well-defined topological clusters. The overview of the entire clustering process, including testing cluster convergence with non-OSAS control patients and validation of SAS_{Score} , is presented in Fig. 5.

For a network (V, E) , a layout algorithm running in an Euclidean k -dimensional space R^k places each vertex $v \in V$ to a corresponding position $p_v \in R^k$ and assigns an Euclidean distance $|p_v - p_w|$ to each edge $[v, w] \in E$.

In particular, energy model layout algorithms are developed as attraction–repulsion (A-R) force systems (Noack, 2009). As such, in an A-R system, adjacent vertices attract while all the other pairs of vertices repulse; this is the emerging mechanism which leads to the formation of groups of vertices with dense connections that we interpret as communities or clusters. The A-R force values are proportional to the power (A or R) of the Euclidean distances between the nodes: the attraction between adjacent vertices v and w is $|p_v - p_w|^A \cdot \overrightarrow{p_v, p_w}$, and the repulsion between any 2 vertices $v, w \in V$ is $|p_v - p_w|^R \cdot \overrightarrow{p_v, p_w}$ ($\overrightarrow{p_v, p_w}$ is the unit vector from v to w). To generate topological communities that are consistent with connection densities, thus having the advantage of emphasizing distinct communities and clusters (Jacomy et al., 2014), attraction between two nodes has to decrease with the Euclidean distance between the nodes, while repulsion has to increase

with the Euclidean distance, therefore we have $A \geq 0$ and $R \leq 0$; such representative force-based layouts are the Fruchterman and Reingold model ($A = 2, R = -1$) (Fruchterman & Reingold, 1991), and the LinLog model ($A = 0, R = -1$) (Noack, 2003). For all A-R energy models, the resulted edge positions are determined by a local energy minima situation (Noack, 2009), as described in Eq. (1):

$$\min \left\{ \sum_{[v,w] \in E, v \neq w} \left(\frac{|p_v - p_w|^{A+1}}{A+1} - \frac{|p_v - p_w|^{R+1}}{R+1} \right) \right\}. \quad (1)$$

In addition to the layout algorithm, we used modularity-based network clustering (Girvan & Newman, 2002), a method that was proven to be effective in network medicine (Diez, Agustí & Wheelock, 2014; Faner et al., 2014). Network clustering consists of assigning each vertex $v \in V$ to one of the disjoint vertex subsets (or clusters) C_i , such that $\cup_i C_i = V$. In our APN clustering approach, modularity classes C_i are represented with distinct colors. Because the APN is an unweighted network, the modularity of any clustering is defined in Eq. (2), where $|E_{C_i}|$ is the number of edges in cluster C_i , $|E|$ is the total number of edges in the network, d_{C_i} is the total degree¹ for nodes in cluster C_i , and d is the total degree for all nodes in the network:

$$\sum_{C_i} \left(\frac{|E_{C_i}|}{|E|} - \frac{\frac{1}{2}d_{C_i}^2}{\frac{1}{2}d^2} \right). \quad (2)$$

Noack demonstrated that energy-model layout algorithms produce topological clusters that are equivalent with those rendered by modularity-based clustering (Noack, 2009). However, force-directed layouts provide additional topological information about clusters. As such, for a more accurate analysis, it is recommended that both modularity clustering and force directed layouts are used (Noack, 2009; Jacomy et al., 2014; Udrescu et al., 2016).

RESULTS

APN analysis

The APN representation resulted from our clustering methodology is presented in Fig. 6, where the distinct colors correspond to distinct modularity classes, and the well-defined topological clusters are explained accordingly. In Fig. 6, we interpret the eight topological clusters as distinct phenotypes, and provide the risk factors prevalence as percentages (L, Mi, Mo, Se)% for each such cluster/phenotype.

Non-OSAS Patients Network (NPN) analysis

Using the information from the 611 people non-OSAS database (NAD), we employ the same procedure as for the APN from Fig. 6. The NAD represents the control population, consisting of people that are not diagnosed with OSAS. The result of applying our methodology on NAD patients is presented in Fig. 7, where the colors correspond to distinct modularity classes; at the same time, topological communities rendered with the energy-model layout Force Atlas 2 are indicated and explained.

Upon visual inspection, Fig. 7 suggests that in the non-OSAS control population there are more patterns of risk factors association, which leads to a number of 12 topological

¹The degree of a vertex represents the number of its incident edges.

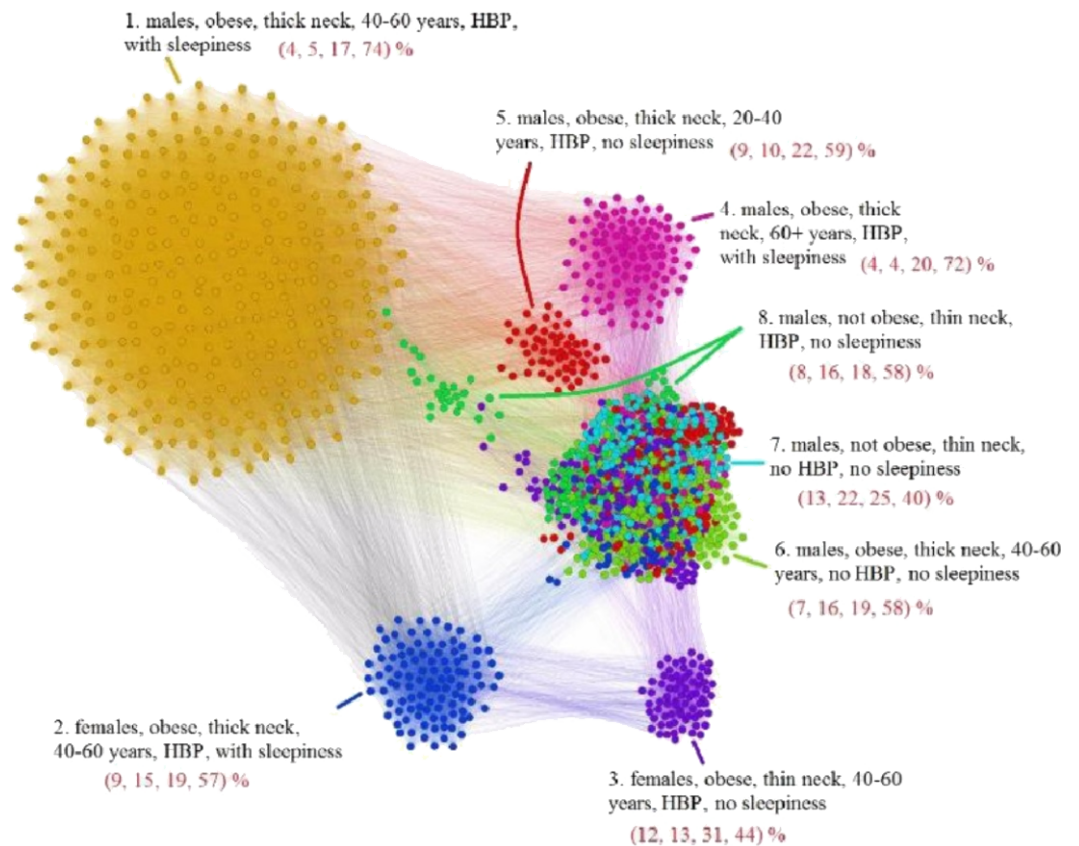


Figure 6 Apnea Patients Network (APN) obtained with data from the Apnea Patients Database (APD), according to the risk factor compatibility relationship, using our dual network clustering methodology (i.e., modularity classes and energy-model layouts). The assigned colors correspond to modularity classes, and the 8 topological clusters are indicated. For each topological cluster, statistics are provided in red (as percentages) for all AHI risk groups: low, mild, moderate, and severe, using the format (*L*, *Mi*, *Mo*, *Se*)% (e.g., in Cluster 2 the patients are distributed on risk groups as follows: 9% *L*, 15% *Mi*, 19% *Mo*, 57% *Se*). The risk group classification is made with AHI values that are obtained by actually performing polysomnography (PSG) and polygraphy.

clusters and modularity classes that are not correlated with OSAS or AHI risk groups. As such, according to our network-based methodology, it occurs that the six considered risk factors consistently converge only for the individuals with OSAS.

Description of phenotypes

In order to have a clear characterization of our rendered phenotypes, we are tracking the OSAS comorbidities (as recorded in the APD) within the APN. To this end, we consider the comorbidity types: *cardiovascular* (e.g., hypertension or stroke), *nutritional* (e.g., obesity or diabetes), and *respiratory*-related (e.g., COPD or asthma). Figure 8 presents the highlighted comorbidities within the APN by using distinct colors for comorbidity types that appear individually, as well as for comorbidity type overlaps (cardiovascular + nutritional, cardiovascular + respiratory, nutritional + respiratory, cardiovascular + nutritional + respiratory), and patients without known comorbidities.

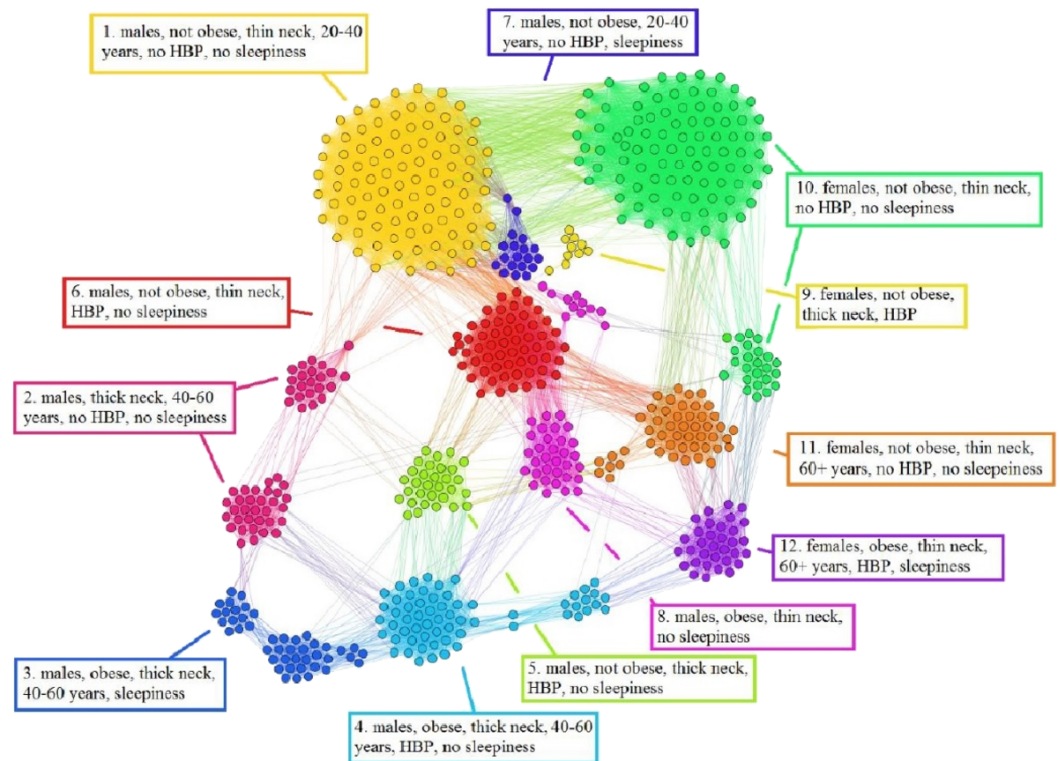


Figure 7 Non-OSAS Patients Network (NPN), obtained with data from the non-OSAS database (NAD) control population database. The colors correspond to 12 modularity classes. Also, the energy-model layout Force Atlas 2 generates 12 topological clusters, which are described in terms of risk factors and their correspondence with the modularity classes.

In light of comorbidity and AHI risk groups statistics provided in [Table 2](#) for each cluster resulted from our network analysis (as illustrated in [Figs. 6](#) and [8](#)), we characterize the phenotypes as follows:

- **Phenotype 1:** Mostly patients within the *Se* AHI risk group, which are generally obese males with thick neck, high blood pressure, sleepiness, and age between 40 and 60 years. For a large majority of these patients, all comorbidity types overlap.
- **Phenotype 2:** The large majority of these patients have *Mo* and *Se* apnea forms; they are obese females with thick neck, high blood pressure, sleepiness, age between 40 and 60 years. In this phenotype there are no patients with only respiratory comorbidities and only few of them have single nutritional comorbidities.
- **Phenotype 3:** The patients have mostly *Mo* and *Se* apnea, but there are less *Se* forms in comparison with other phenotypes; they are obese females with thin neck, high blood pressure, no sleepiness, and age between 40 and 60 years. This phenotype does not contain patients with only respiratory comorbidities.
- **Phenotype 4:** Mostly *Se* patients; however, there is a significant number of *Mo* individuals, which are generally obese males with thick neck, high blood pressure,

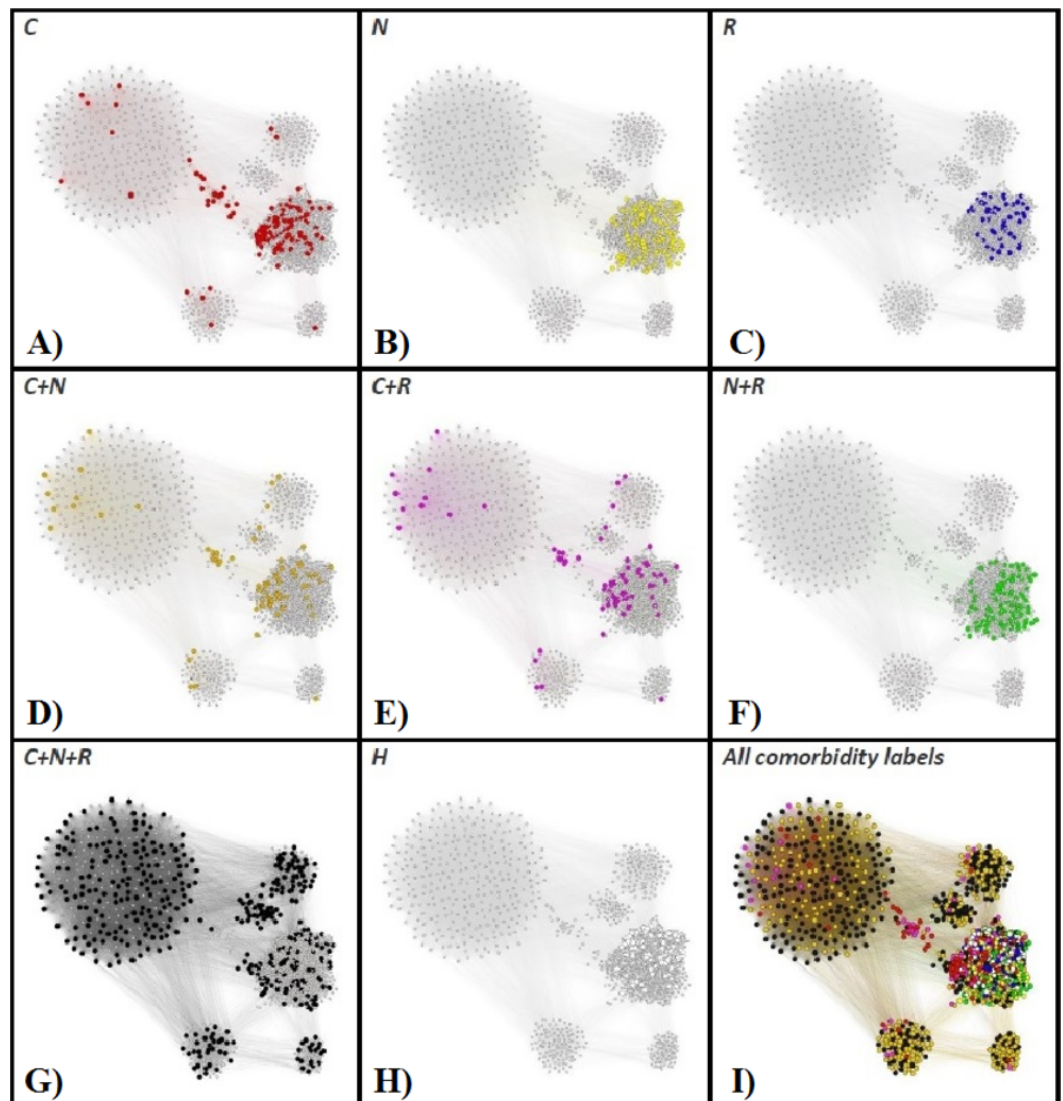


Figure 8 The Apnea Patients Network (APN), with highlighted individual comorbidities and associations of comorbidities. The red nodes correspond to patients that have only cardiovascular comorbidities (C in A), the yellow nodes to only nutrition-related (N in B), and blue nodes represent OSAS patients with only respiratory comorbidities (R in C). Patients with overlapping comorbidity types are represented within the APN as follows: orange nodes correspond to cardiovascular plus nutritional comorbidities ($C + N$ in D), purple nodes to cardiovascular plus respiratory comorbidities ($C + R$ in E), green nodes to nutritional plus respiratory comorbidities ($N + R$ in F), and black nodes to the superposition of cardiovascular, nutritional and respiratory comorbidities ($C + N + R$, G). The OSAS patients without known comorbidities are highlighted in the APN as white nodes (H in H). We also provide the APN where all nodes are labeled according to their comorbidity or comorbidities overlapping in I.

sleepiness, and over 60 years. In this phenotype only a few patients have the single respiratory comorbidities type.

- **Phenotype 5:** Mostly *Se*, *Mo*, and *Mi* patients, which are obese young males with thick neck, high blood pressure, no sleepiness, and age between 20 and 40. In this phenotype,

Table 2 Description of the eight relevant Apnea Patients Network (APN) phenotypes. The phenotypes (Ph) are listed with a short description in terms of most or least predominant AHI risk groups (low risk- *L*, mild- *Mi*, moderate- *Mo*, and severe- *Se*), significant combinations of the 6 objective parameters, and most/least predominant comorbidity types (cardiovascular- *C*, nutritional- *N*, respiratory- *R*, and without comorbidities- *H*) or comorbidity types overlaps (*C + N*, cardiovascular + nutritional; *C + R*, cardiovascular + respiratory, *N + R*, nutritional + respiratory; *C + N + R*, cardiovascular + nutritional + respiratory). For each phenotype, we provide the corresponding percentages for comorbidity types and comorbidity type associations, as well as the percentage of patients pertaining to one of the AHI risk groups; the boldface entries correspond to representative values, in terms of simple majority. In phenotype descriptions, HBP stands for high blood pressure.

Ph.	Description	Comorbidity types and associations (%)								AHI risk groups (%)			
		<i>C</i>	<i>N</i>	<i>R</i>	<i>C + N</i>	<i>C + R</i>	<i>N + R</i>	<i>C + N + R</i>	<i>H</i>	<i>L</i>	<i>Mi</i>	<i>Mo</i>	<i>Se</i>
1	Mostly <i>Se</i> with <i>C + N + R</i> comorbidities	2.38	–	–	32.65	3.4	–	61.57	–	4	5	17	74
2	Mostly <i>Mo</i> and <i>Se</i> , thick neck females, no <i>R</i> , few <i>N</i>	6.49	0.65	–	61.04	3.9	0.65	27.27	–	9	15	19	57
3	Mostly <i>Mo</i> , <i>Se</i> , thin neck females, no <i>R</i>	13.66	10.24	–	48.29	1.95	2.93	22.93	–	12	13	31	44
4	Mostly <i>Mo</i> and <i>Se</i> , elderly males with thick neck, few <i>R</i>	6.52	3.62	–	36.96	5.8	3.62	44.48	–	4	4	20	72
5	Mostly <i>Mi</i> , <i>Mo</i> , <i>Se</i> , obese young males, mostly <i>N</i>	2.47	18.52	1.23	19.75	3.09	20.99	32.72	1.23	9	10	22	59
6	Mostly <i>Mo</i> , <i>Se</i> , 40–60 yrs. obese males, no HBP, mostly <i>N</i>	2.58	33.55	5.81	10.97	1.29	31.61	7.74	6.45	7	16	19	58
7	Less <i>Se</i> , non-obese, thin neck, no HBP males, mostly <i>H</i> , some <i>R</i>	8.84	0.68	27.21	0.68	4.76	–	–	57.83	13	22	25	40
8	Less <i>Se</i> , non-obese, thin neck, HBP males, mostly <i>C</i>	38.79	2.59	–	9.48	28.45	2.59	18.1	–	8	16	18	58

almost all patients have nutritional comorbidities or comorbidity overlaps that include the nutritional type.

- **Phenotype 6:** Consists of mostly *Mo* and *Se* apnea; the patients with this phenotype are generally obese males with thick neck, no high blood pressure, no sleepiness, and middle aged (40–60 years old). Their comorbidities are mostly nutritional-related (either single nutritional comorbidity or an association of comorbidities that contains the nutritional type).
- **Phenotype 7:** Patients with mostly *Mo* and *Se* apnea, but with less *Se* forms in comparison with other phenotypes; this phenotype's patients are generally males of all ages with thin neck, no high blood pressure, and no sleepiness. The majority of these patients have no comorbidities; however, those who have a comorbidity tend to have respiratory-related problems.
- **Phenotype 8:** Patients mostly within *Se*, *Mo*, and *Mi* AHI risk groups; they are males from all age groups with thin neck, no sleepiness, but with high blood pressure. These patients tend to have a single cardiovascular comorbidity type or an association of comorbidities that include the cardiovascular type.

OSAS risk prediction with SAS_{Score}

Classifying any new patient in one of the phenotypes can be performed by adding the new patient to the APN and then running the modularity class and force-directed layout algorithms in Gephi one more time. However, during the screening process, physicians are frequently unable to perform these rather complex and time consuming steps (i.e., manipulation of databases and managing Gephi plugins), because of the obvious constraints.

In order to deal with this problem, we propose a simplified solution for classifying *de novo* patients using a computer algorithm that is implemented as a web-based application. As such, we employ supervised machine learning in order to classify any new person in one of the eight validated phenotypes, based on the six relevant parameters. To this end, we choose decision tree learning, because decision trees are easy to use, quick, and intuitive for medical personnel. We use R and available recursive tree partitioning libraries from R Studio in order to perform recursive tree partitioning, thus generating a phenotype classification tree ([Therneau et al., 2010](#); [Hothorn, Hornik & Zeileis, 2006](#)).

Given the eight phenotypes determined with our network procedure, and the features extracted from them, we label each patient from the APD with the cluster/phenotype to which it pertains. From this point, we employ supervised learning methods and then test them in order to choose the tree which provides the highest reliability while maintaining a reasonable complexity.

The tested mining algorithms are: recursive partition tree, conditional inference tree, evolutionary tree, oblique tree, maptree, naive Bayes, random forest and linear regression ([Therneau et al., 2010](#); [Hothorn, Hornik & Zeileis, 2006](#); [Grubinger, Zeileis & Pfeiffer, 2011](#)). The algorithm test procedure relies on applying the resulted decision trees on *de novo* patients from the Test Database (TD). We take all patients in TD and assign each of them to a phenotype with the classification tree. In parallel, we add the TD patients to the APD,

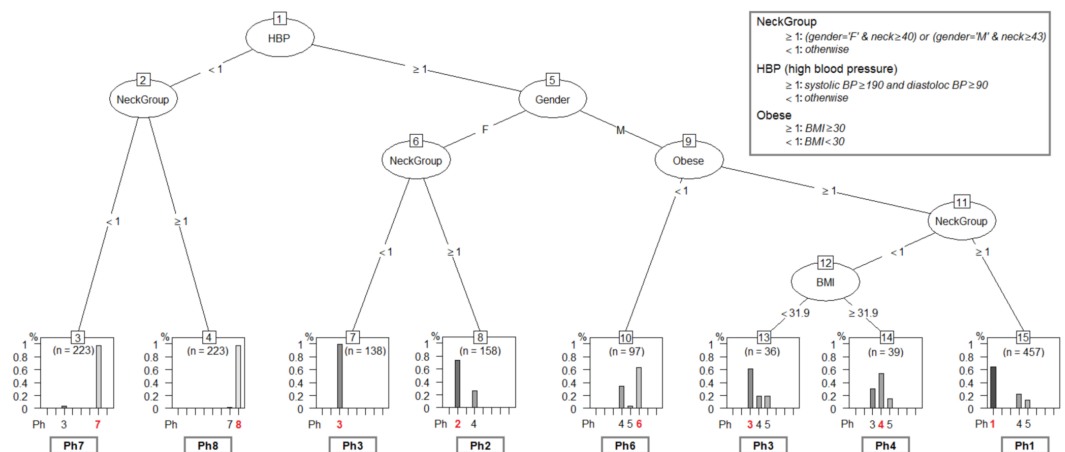


Figure 9 The phenotype classifying tree, obtained by running *evtree* on patients from APD. Each new person is assigned to one of the eight phenotypes, denoted as *Ph1–Ph8* (dominant phenotypes are emphasized). Decisions are made according to some of the six relevant parameters (gender-*G*, age- *A*, body mass index-*BMI*, systolic blood pressure-*SBP*, diastolic blood pressure-*DBP*, Epworth sleepiness score-*ESS*), but also according to variables computed from the six parameters (High Blood Pressure-*HBP*, Neck Group-*NG*, Obesity-*Ob*), as described in the figure legend.

and rerun the entire layout clustering procedure. Indeed, we obtain the same eight patient phenotypes, and we consider the phenotype labels that are assigned to patients in TD by our network clustering procedure as being the reference (i.e., correct phenotype assignment). Therefore, we quantify the efficiency of each decision tree by comparing with the reference represented by the network-based classification. The test results indicate the evolutionary tree (*evtree*) [Grubinger, Zeileis & Pfeiffer \(2011\)](#), which is given in [Fig. 9](#), as the best method for our classification problem.

The classification tree in [Fig. 9](#) uses the six relevant parameters (gender- *G*, age- *A*, body mass index- *BMI*, systolic blood pressure- *SBP*, diastolic blood pressure- *DBP*, Epworth sleepiness score- *ESS*), but also variables that are computed from the six parameters (High Blood Pressure- *HBP*, Neck Group- *NG*, Obesity- *Ob*) according to [Eqs. \(3\)– \(5\)](#):

$$HBP = \begin{cases} 1 & \text{if } SBP \geq 140 \text{ and } DBP \geq 90 \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

$$NG = \begin{cases} 1 & \text{if } (G = \text{female and } NC \geq 40) \text{ or } (G = \text{male and } NC \geq 43) \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

$$Ob = \begin{cases} 1 & \text{if } BMI \geq 30 \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

When we test the classification tree in [Fig. 9](#) on patients from TD and compare the results against the reference classification (i.e., network-based), the resulted classification accuracy is 69.30%; the detailed prediction accuracy results for TD are given in [Table 3](#).

Table 3 Prediction table for each phenotype and associated classification error, as obtained by employing emtree on TD. Each line in the table shows (in bold characters) how many patients in TD are correctly classified in the corresponding phenotype.

		Patients actually belonging to phenotype								Error
		Ph 1	Ph 2	Ph 3	Ph 4	Ph 5	Ph 6	Ph 7	Ph 8	
Patients predicted in	Ph 1	294	0	0	103	60	0	0	0	35.67%
	Ph 2	0	116	0	42	0	0	0	0	26.58%
	Ph 3	0	0	163	11	12	1	0	0	12.36%
	Ph 4	0	0	8	17	1	0	0	0	34.61%
	Ph 5	0	0	0	0	0	0	0	0	100%
	Ph 6	0	0	0	33	3	61	0	0	37.11%
	Ph 7	0	0	7	0	0	0	216	0	3.14%
	Ph 8	0	0	0	2	0	0	3	218	2.24%

Table 4 Average values for the relevant parameters (body mass index- *BMI*, neck circumference- *NC*, systolic blood pressure- *SBP*, diastolic blood pressure- *DBP*, Epworth sleepiness score- *ESS*), which are computed for each of the eight clusters in Fig. 6.

	Cluster/Phenotype							
	1	2	3	4	5	6	7	8
$BMI_a^{Cluster}$	36.83	37.54	31.88	33.75	34.99	32.93	22.74	28.67
$NC_a^{Cluster}$	47.66	43.9	35.91	46.79	43.17	43.66	33.39	37.36
$SBP_a^{Cluster}$	143.96	145.32	139.48	140.56	133.74	125.46	119.66	134.55
$DBP_a^{Cluster}$	89.46	87.61	85.80	84.28	84.43	80.75	73.16	84.41
$ESS_a^{Cluster}$	11.49	10.77	7.74	10.10	8.58	9.50	6.28	8.90

In order to generate our SAS_{Score} from the classification tree in Fig. 9, we follow the next sequence of steps:

1. Perform anthropometric measurements on each new patient.
2. Classify each patient in one of the eight phenotypes (using the classification from Fig. 9).
3. Refer to cluster normalization in order to get phenotype parameter averages from Table 4.
4. Compute SAS_{Score} using Eq. (6).

Therefore, having the six recorded parameters for any new patient that has to be evaluated, and the average values $BMI_a^{Cluster}$, $NC_a^{Cluster}$, $SBP_a^{Cluster}$, $DBP_a^{Cluster}$, $ESS_a^{Cluster}$ with $Cluster \in \{1, 2, \dots, 8\}$ as presented in Table 4, the SAS_{Score} is automatically computed by our computer application according to Eq. (6), then correspondingly displayed by the web-based interface.

$$SAS_{Score} = \frac{BMI}{BMI_a^{Cluster}} + \frac{NC}{NC_a^{Cluster}} + \frac{1}{2} \left(\frac{SBP}{SBP_a^{Cluster}} + \frac{DBP}{DBP_a^{Cluster}} \right) + \frac{ESS}{ESS_a^{Cluster}}. \quad (6)$$

SAS_{Score} values are ≥ 1 and generally < 7 ; this range of values can further serve to classify patients as being at risk of developing OSAS or not. A patient with $SAS_{Score} > threshold$ is considered at risk, whereas $SAS_{Score} \leq threshold$ means that there is no OSAS risk. In

Table 5 Average anthropometric measurements with standard deviations (expressed as percentages (%)) for the anthropometric parameters (Age *A*, *BMI*, Neck Circumference *NC*, High Blood Pressure *HBP*, Epworth Sleepiness Score *ESS*, Apnea-Hypopnea Index *AHI*) in the APN clusters/phenotypes after 100 runs (for instance, the average age in cluster 2 is 54.11 years with a standard deviation of 3.51%).

Cluster	<i>A</i>	<i>BMI</i>	<i>NC</i>	<i>HBP</i>	<i>ESS</i>	<i>AHI</i>
1	51.89 ± 0	36.83 ± 0	47.66 ± 0	100% ± 0	11.49 ± 0	50.75 ± 0
2	54.11 ± 3.51	37.73 ± 0.86	44.13 ± 0.47	92% ± 5.22	10.59 ± 0.99	38.3 ± 0.49
3	56.14 ± 1.87	31.88 ± 1.27	35.87 ± 1.13	87% ± 5.77	7.8 ± 2.45	30.92 ± 3.09
4	66.47 ± 0.28	33.84 ± 1.55	44.15 ± 2.46	90% ± 6.78	10.02 ± 1.64	42.99 ± 3.3
5	34.54 ± 1.34	35.09 ± 1.53	44.52 ± 2.39	51% ± 5.27	8.94 ± 3.26	49.66 ± 4.94
6	51.23 ± 2.45	32.94 ± 1.66	45.23 ± 3.08	0% ± 0	9.85 ± 2.82	44 ± 4
7	43.59 ± 2.62	24.6 ± 5.77	33.01 ± 1.35	7% ± 10.57	6.46 ± 2.51	25.93 ± 2.9
8	55.92 ± 4.47	27 ± 3.55	39.96 ± 4.51	96% ± 10.42	8.75 ± 3.64	34.29 ± 2.66

order to attain the main objective of our paper, namely to make population-wide OSAS monitoring and screening effective, we consider the higher specificity as being more important, so we choose $threshold = 3.9$, which determines a sensitivity of 0.8025 and a specificity of 0.4189. Taken together, these results obtained with the TD, consisting of *de novo* patients only, suggest that our SAS_{Score} significantly outperforms STOP-BANG in terms of specificity (i.e., it is 2.34 times better), while remaining only slightly worse than STOP-BANG in terms of sensitivity (8.2% decrease). We consider these results as being particularly relevant, because the distribution of *AHI* in the TD is notably different from the distribution of *AHI* in APD (see Fig. 3 panels A and B).

Validation of clustering consistency

In this subsection, we verify that rendering the clusters in Fig. 6 is not mere serendipity, and it is not induced by some fortunate heterogeneity of patients. To this end, we perform random shuffling and bootstrapping test investigations.

Because our clustering methodology starts with a random state, namely it starts with the raw network where nodes are randomly placed and the links have corresponding lengths (see Fig. 5), our first shuffling test consists of running the procedure in Fig. 5 many times in order to see if we get statistically consistent results. Therefore, we run our dual clustering procedure 100 times on the same APD, and then measure the distribution of anthropometric values for each phenotype. The result of our random shuffling is given in Table 5 which presents average anthropometric measurements with standard deviations (expressed as percentages) for the each APN cluster/phenotype after 100 runs. Indeed, the deviations from the average values are very small, emphasizing the consistency of our clustering procedure in Fig. 5.

Our second test approach entails generating test APDs from the original patient dataset, in order to perform bootstrapping. To do so, we generate 10 new APD datasets with the same number of patients as the original APD by randomly selecting patients from the original APD database. Therefore, in the test APDs, some of the original patients may be missing, while others may be present two or more times. Next, we apply the same clustering methodology from Fig. 5 and find that the same phenotypes emerge.

Table 6 Defining phenotype characteristics given as majority values with corresponding percentages, as measured in the original APN communities.

Cluster	G	Ob	TN	AG	HBP	SLP
1	M (100%)	1 (100%)	1 (100%)	3 (100%)	1 (100%)	1 (61%)
2	F (100%)	1 (96%)	1 (100%)	3 (92%)	1 (100%)	1 (56%)
3	F (80%)	1 (81%)	0 (100%)	3 (69%)	1 (96%)	0 (70%)
4	M (100%)	1 (82%)	1 (77%)	4 (100%)	1 (99%)	1 (52%)
5	M (100%)	1 (96%)	1 (83%)	2 (100%)	1 (100%)	0 (64%)
6	M (88%)	1 (87%)	1 (100%)	3 (57%)	0 (100%)	0 (57%)
7	M (67%)	0 (61%)	0 (99%)	3 (50%)	0 (100%)	0 (77%)
8	M (98%)	0 (100%)	0 (55%)	3 (84%)	1 (100%)	0 (58%)

Table 7 Characteristic features of phenotypes averaged over 10 randomized APNs. Majority average percentages for the dominant values are rounded to the nearest integer, while the colours represent how close these averages are to the measurements on the original APN from Table 6: bold entries correspond to very close matches (percentage difference $\leq 6\%$), normal character entries correspond to a good match ($6\% < \text{percentage difference} \leq 15\%$), while grey-italics table entries correspond to significant differences ($> 15\%$).

Cluster	G	Ob	TN	AG	HBP	SLP
1	M (100%)	1 (100%)	1 (100%)	3 (100%)	1 (100%)	1 (60%)
2	F (100%)	1 (96%)	1 (100%)	<i>3 (74%)</i>	1 (100%)	1 (60%)
3	F (90%)	1 (81%)	0 (100%)	3 (63%)	1 (87%)	0 (69%)
4	M (95%)	1 (86%)	1 (88%)	4 (100%)	1 (97%)	1 (53%)
5	M (97%)	1 (97%)	1 (86%)	2 (100%)	1 (79%)	0 (62%)
6	M (95%)	1 (87%)	1 (96%)	<i>3 (73%)</i>	0 (100%)	0 (58%)
7	<i>M (81%)</i>	0 (64%)	0 (89%)	<i>3 (75%)</i>	0 (99%)	0 (70%)
8	<i>M (82%)</i>	0 (84%)	<i>0 (80%)</i>	3 (68%)	1 (100%)	0 (71%)

The characteristics of the original APN phenotypes from Fig. 6 are provided in Table 6; Table 7 shows the averaged characteristics, for each cluster, over the randomized 10 APNs obtained by bootstrapping. In Table 7 distinct character types suggest how close (bold and normal characters) or how far (grey italics) are the phenotype characteristics resulted from bootstrapping from the original APN values.

The values from Table 7 indicate that the dominant characteristics of each cluster in the test APNs are very similar to the cluster characteristics from the original APN. As a result, even if we create new APDs with corresponding APNs by shuffling the patients from our dataset, the bootstrapping procedure yields the same phenotypes. The bootstrapping procedure reveals that some phenotypes emerge as more stable when randomized (e.g., clusters 1, 3, 4, 5), with a small offset compared to the original APN. The lack of convergence over multiple randomizations could be interpreted as a lack of representativeness as an OSAS phenotype. As such, clusters 2 and 6 are slightly less representative, and clusters 7 and 8 are notably variable, suggesting that the last two phenotypes may not be so well characterized, due to the reduced amount of patient data.

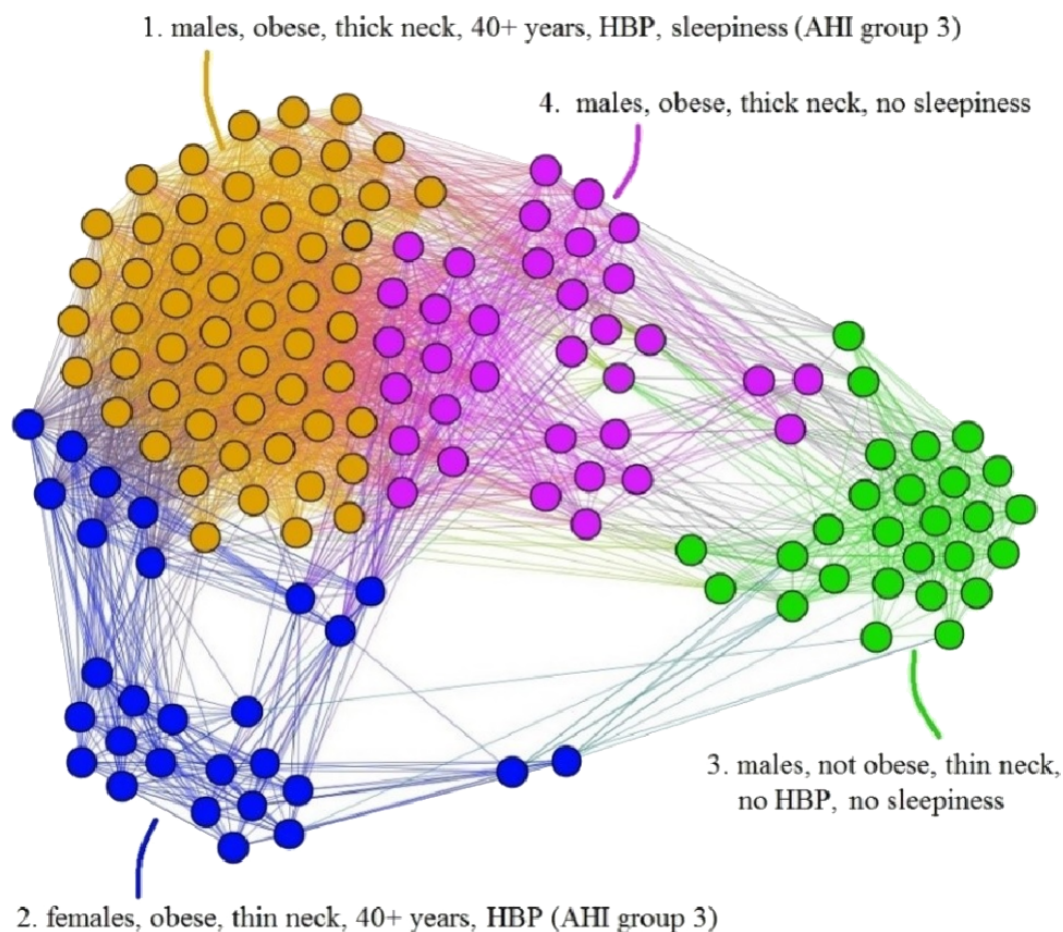


Figure 10 Representation of Test Database-TD patients, according to the risk factor compatibility, using the proposed network-based methodology. Although the number of patients is significantly smaller in comparison with the primary Apnea Patients Network-APN, the visual inspection reveals that the emerged clusters are similar to clusters from Fig. 6 (APN).

We also create a test patient network (TPN), similar to building the APN, and then apply our dual clustering methodology. The result is presented in Fig. 10; upon visual inspection it can be noticed that the clusters emerged in TPN are similar to the clusters from Fig. 6's APN, even if the number of patients is significantly smaller in TD; this result suggests that the association and convergence of risk factors in OSAS patients is indeed a non-random, consistent process.

DISCUSSION

The proposed method is not the first to cluster apnea patients (*Joosten et al., 2012; Vavougiou et al., 2016; Ye et al., 2014*), but to the best of our knowledge it is the first network-based approach used for clustering apnea patients. Another important feature is that our network-based methodology employs only easy-to-measure, objective clustering parameters (*AHI* is used only for phenotype evaluation). This way, our clustering methodology emphasizes the

high complexity of OSAS phenotypes, from typical (cluster 1—yellow) to the less obvious ones (clusters 6–8).

When defining the eight apnea phenotypes, besides the force-directed layout, we also use modularity class clustering. In Fig. 6, the phenotypes based on modularity classes are generally consistent with the topological clusters resulted from applying the Force Atlas 2 layout. However, the visual inspection of Fig. 6 reveals that phenotypes 5, 6, 7, and 8 tend to spatially overlap; this tendency is much stronger for phenotypes 6, 7, and 8. Such a tendency for overlapping phenotypes that characterize patients with generally mild and moderate OSAS is also suggested by Joosten *et al.* (2012). This observation may indicate that these phenotypes are interrelated and generally hard to distinguish even in clinical practice. Still, some of the nodes in these clusters (e.g., cluster 8) have a clear tendency towards separation from the overlapping; this indicates that we probably need more patients/nodes, in order to completely segregate Cluster 8. As our APN will grow over time, the less convergent clusters 7 and 8 might become more representative. Indeed, the fact that even in the original APN from Fig. 6 clusters 7 and 8 present significant overlapping and that their topological segregation from other clusters is somehow fuzzy confirms the conclusion of our bootstrapping investigation. Nonetheless, we preferred to use the distinct modularity classes in conjunction with the topological clusters because they bring more information, i.e., more detail which can be useful for medical analysis.

From a medical standpoint, we note that our dual clustering method renders distinct male and female clusters; this observation is consistent with the state of the art medical literature which holds gender as a very important predictor of OSAS. For instance, in a 2009–2013 study on 272,705 patients from North America, referred for home sleep apnea testing, clinical OSAS features are found to be more common in males than females (Cairns, Poulos & Bogan, 2016). Other studies performed on 23,806 (Gabbay & Lavie, 2012), and 1,010 (Vagiakis *et al.*, 2006) patients respectively, show clear differentiation between the two genders in terms of AHI distribution and severity.

In current practice, the commonly used score for predicting sleep apnea is STOP-BANG. In comparison with STOP-BANG, our SAS_{Score} significantly improves the prediction specificity (2.34 times better than STOP-BANG), while sensitivity is only slightly degraded. STOP-BANG has high sensitivity because it is a simple heuristic that was especially designed for perioperative patients, where it is essential to identify all potential risks associated with anaesthesia (including OSAS). To further emphasize the higher specificity of SAS_{Score} , we mention that by using our score, only 34% people from NAD are found as at risk of developing OSAS. Moreover, as opposed to STOP-BANG (a fixed questionnaire that cannot be adjusted to specific patients characteristics), SAS_{Score} represents an adaptive methodology. Therefore, as the database grows, better sensitivity and specificity are expected. The classifying tree which leads to rendering SAS_{Score} , as described in section OSAS risk prediction with SAS_{Score} represents a simplified application of our patient clustering/phenotyping method, but this method has the advantage of being applicable in offline conditions, which makes it amenable to clinical practice and population screening. All these considerations indicate SAS_{Score} as an appropriate tool for OSAS screening in large general populations.

Our network-based method represents an application on patients from a given geographical area, therefore we consider that it should be tested for other targeted populations. As such, the network analysis will render new, specific cluster average values (such as BMI_a^{Cluster} , NC_a^{Cluster} , SBP_a^{Cluster} , DBP_a^{Cluster} , ESS_a^{Cluster}). Subsequently, SAS_{Score} values that are specific to the targeted population can be rendered with the SAS_{Score} equation. However, we do not expect that the phenotypes or SAS_{Score} will be significantly different for other populations, since the available medical studies, performed over diverse geographical areas (including a wide array of anthropometric characteristics), show that specific population traits are not particularly relevant for OSAS (*Ralls & Grigg-Damberger, 2012; Lee et al., 2008; Villaneuva et al., 2005*).

Eventually, due to its higher specificity, the SAS_{Score} can be integrated into a large area apnea screening and monitoring procedure, which aims at specifically discovering typical severe cases (easy to investigate with portable devices), without overcrowding sleep laboratories with false positive cases. This way, efficient personalized patient processing can be achieved by making use of prioritization according to the predicted severity level. For instance, this method can be a useful tool for sleep apnea screening in large population categories, such as professional drivers since, at the European level, the new 2014/85/EU directive regarding professional drivers is recommended from January 2016.² In this context, our website <http://sasscore.appspot.com> is a good example of a large-area, accessible OSAS risk prediction tool. Indeed, SAS_{Score} can be conveniently computed in both clinical and population-monitoring practices, due to the fact that it is implemented as easy-to-use smartphone and web-based applications (<https://play.google.com/store/apps/details?id=aerscore.topindustries.aerscore&hl=en> and www.pneumoresearch.ro). To this end, processing data and obtaining the prediction score requires less than 1 min per individual.

CONCLUSION

This paper proposes a new OSAS patients clustering method based on complex network analysis, which leads to identifying OSAS phenotypes. This innovative network medicine approach is extended in order to compute SAS_{Score} , a predictive score for OSAS based on 6 easy-to-measure, objective parameters. The proposed method uses big data, complex networks analysis in order to achieve better specificity in OSAS prediction. As such, our SAS_{Score} can conveniently be used in conjunction with the existing questionnaires for better OSAS prevention through population screening and monitoring, thus paving the way for a personalized patient management process.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This work was supported by Linde Healthcare's Real Fund 2014, through research grant "Morpheus: A Screening and Monitoring System for Sleep Apnea Syndrome", 11289/14.10.2014. M.U. and A.T. were also supported by a grant of the Romanian National

²Commission Directive COMMISSION DIRECTIVE 2014/85/EU of 1 July 2014 amending Directive 2006/126/EC—European Parliament and the Council on driving licences.

Authority for Scientific Research and Innovation, UEFISCDI, project number 31 PED/2017: INCEPTION. There was no additional external funding received for this study. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:

Linde Healthcare's Real Fund 2014.

Morpheus: A Screening and Monitoring System for Sleep Apnea Syndrome": 11289/14.10.2014.

Romanian National Authority for Scientific Research and Innovation, UEFISCDI, project number 31 PED/2017: INCEPTION: Internet of Things Meets Complex Networks for Early Prediction and Management of Chronic Obstructive Pulmonary Disease.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Stefan Mihaicuta conceived and designed the experiments, performed the experiments, analyzed the data, wrote the paper, reviewed drafts of the paper.
- Mihai Udrescu conceived and designed the experiments, analyzed the data, contributed reagents/materials/analysis tools, wrote the paper, prepared figures and/or tables.
- Alexandru Topirceanu analyzed the data, contributed reagents/materials/analysis tools, prepared figures and/or tables, reviewed drafts of the paper.
- Lucretia Udrescu conceived and designed the experiments, performed the experiments, analyzed the data, reviewed drafts of the paper.

Human Ethics

The following information was supplied relating to ethical approvals (i.e., approving body and any reference numbers):

The Ethical Committee of Victor Babes Hospital, Timisoara, Romania approved this study (approval no. 10/12.10.2013).

Data Availability

The following information was supplied regarding data availability:

The raw data has been supplied as [Supplementary Files](#).

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.3289#supplemental-information>.

REFERENCES

- Bakker JP, Montesi SB, Malhotra A. 2013.** Obstructive sleep apnoea: new associations and approaches. *The Lancet. Respiratory Medicine* **1(1)**:e15
[DOI 10.1016/S2213-2600\(12\)70059-0](https://doi.org/10.1016/S2213-2600(12)70059-0).

- Barabási A-L. 2007.** Network medicine—from obesity to the “diseasome”. *New England Journal of Medicine* **357**(4):404–407 DOI [10.1056/NEJMe078114](https://doi.org/10.1056/NEJMe078114).
- Barabási A-L. 2011.** The network takeover. *Nature Physics* **8**(1):14–16 DOI [10.1038/nphys2188](https://doi.org/10.1038/nphys2188).
- Barabasi A-L, Oltvai ZN. 2004.** Network biology: understanding the cell’s functional organization. *Nature Reviews Genetics* **5**(2):101–113 DOI [10.1038/nrg1272](https://doi.org/10.1038/nrg1272).
- Cairns A, Poulos G, Bogan R. 2016.** Sex differences in sleep apnea predictors and outcomes from home sleep apnea testing. *Nature and Science of Sleep* **8**:197–205 DOI [10.2147/NSS.S101186](https://doi.org/10.2147/NSS.S101186).
- Chung F, Yang Y, Brown R, Liao P. 2014.** Alternative scoring models of STOP-BANG questionnaire improve specificity to detect undiagnosed obstructive sleep apnea. *Journal of Clinical Sleep Medicine* **10**(9):951–958 DOI [10.5664/jcsm.4022](https://doi.org/10.5664/jcsm.4022).
- Chung F, Yegneswaran B, Liao P, Chung SA, Vairavanathan S, Islam S, Khajehdehi A, Shapiro CM. 2008a.** STOP questionnaire. A tool to screen patients for obstructive sleep apnea. *The Journal of the American Society of Anesthesiologists* **108**(5):812–821 DOI [10.1097/ALN.0b013e31816d83e4](https://doi.org/10.1097/ALN.0b013e31816d83e4).
- Chung F, Yegneswaran B, Liao P, Chung SA, Vairavanathan S, Islam S, Khajehdehi A, Shapiro CM. 2008b.** Validation of the Berlin questionnaire and American society of anesthesiologists checklist as screening tools for obstructive sleep apnea in surgical patients. *The Journal of the American Society of Anesthesiologists* **108**(5):822–830 DOI [10.1097/ALN.0b013e31816d91b5](https://doi.org/10.1097/ALN.0b013e31816d91b5).
- Diez D, Agustí A, Wheelock CE. 2014.** Network analysis in the investigation of chronic respiratory diseases. From basics to application. *American Journal of Respiratory and Critical Care Medicine* **190**(9):981–988 DOI [10.1164/rccm.201403-0421PP](https://doi.org/10.1164/rccm.201403-0421PP).
- Divo MJ, Casanova C, Marin JM, Pinto-Plata VM, De Torres JP, Zulueta JJ, Cabrera C, Zagaceta J, Sanchez-Salcedo P, Berto J, Baz Davila R, Alcaide AB, Cote C, Celli BR. 2015.** Chronic obstructive pulmonary disease comorbidities network. *European Respiratory Journal* **46**(3):640–650 DOI [10.1183/09031936.00171614](https://doi.org/10.1183/09031936.00171614).
- Faner R, Cruz T, López-Giraldo A, Agustí A. 2014.** Network medicine, multimorbidity and the lung in the elderly. *European Respiratory Journal* **44**(3):775–788 DOI [10.1183/09031936.00078714](https://doi.org/10.1183/09031936.00078714).
- Farney RJ, Walker BS, Farney RM, Snow GL, Walker JM. 2011.** The STOP-Bang equivalent model and prediction of severity of obstructive sleep apnea: relation to polysomnographic measurements of the apnea/hypopnea index. *Journal of Clinical Sleep Medicine* **7**(5):459–465 DOI [10.5664/JCSM.1306](https://doi.org/10.5664/JCSM.1306).
- Fischer J, Dogas Z, Bassetti CL, Berg S, Grote L, Jennum P, Levy P, Mihaicuta S, Nobili L, Riemann D, Puertas-Cuesta FJ, Raschke F, Skene DJ, Stanley N, Pevernagie D. 2012.** Standard procedures for adults in accredited sleep medicine centres in Europe. *Journal of Sleep Research* **21**(4):357–368 DOI [10.1111/j.1365-2869.2011.00987.x](https://doi.org/10.1111/j.1365-2869.2011.00987.x).
- Fruchterman TM, Reingold EM. 1991.** Graph drawing by force-directed placement. *Software: Practice and Experience* **21**(11):1129–1164 DOI [10.1002/spe.4380211102](https://doi.org/10.1002/spe.4380211102).
- Gabbay IE, Lavie P. 2012.** Age-and gender-related characteristics of obstructive sleep apnea. *Sleep and Breathing* **16**(2):453–460 DOI [10.1007/s11325-011-0523-z](https://doi.org/10.1007/s11325-011-0523-z).

- Girvan M, Newman ME. 2002.** Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America* **99**(12):7821–7826 DOI [10.1073/pnas.122653799](https://doi.org/10.1073/pnas.122653799).
- Goh K-I, Cusick ME, Valle D, Childs B, Vidal M, Barabási A-L. 2007.** The human disease network. *Proceedings of the National Academy of Sciences of the United States of America* **104**(21):8685–8690 DOI [10.1073/pnas.0701361104](https://doi.org/10.1073/pnas.0701361104).
- Grubinger T, Zeileis A, Pfeiffer K-P. 2011.** emtree: evolutionary learning of globally optimal classification and regression trees in R. Technical report. Working Papers in Economics and Statistics.
- Han J-DJ. 2008.** Understanding biological functions through molecular networks. *Cell Research* **18**(2):224–237 DOI [10.1038/cr.2008.16](https://doi.org/10.1038/cr.2008.16).
- Hedner J, Grote L, Bonsignore M, McNicholas W, Lavie P, Parati G, Sliwinski P, Barbé F, De Backer W, Escourrou P, Fietze I, Kvamme JA, Lombardi C, Marrone O, Masa JF, Montserrat JM, Penzel T, Pretl M, Riha R, Rodenstein D, Saaresranta T, Schulz R, Tkacova R, Varoneckas G, Vitols A, Vrints H, Zielinski J. 2011.** The European sleep apnoea database (ESADA): report from 22 European sleep laboratories. *European Respiratory Journal* **38**(3):635–642 DOI [10.1183/09031936.00046710](https://doi.org/10.1183/09031936.00046710).
- Hothorn T, Hornik K, Zeileis A. 2006.** Unbiased recursive partitioning: a conditional inference framework. *Journal of Computational and Graphical Statistics* **15**(3):651–674 DOI [10.1198/106186006X133933](https://doi.org/10.1198/106186006X133933).
- Jacomy M, Venturini T, Heymann S, Bastian M. 2014.** ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PLOS ONE* **9**(6):e98679 DOI [10.1371/journal.pone.0098679](https://doi.org/10.1371/journal.pone.0098679).
- Joosten SA, Hamza K, Sands S, Turton A, Berger P, Hamilton G. 2012.** Phenotypes of patients with mild to moderate obstructive sleep apnoea as confirmed by cluster analysis. *Respirology* **17**(1):99–107 DOI [10.1111/j.1440-1843.2011.02037.x](https://doi.org/10.1111/j.1440-1843.2011.02037.x).
- Lee W, Nagubadi S, Kryger MH, Mokhlesi B. 2008.** Epidemiology of obstructive sleep apnea: a population-based perspective. *Expert Review of Respiratory Medicine* **2**(3):349–364 DOI [10.1586/17476348.2.3.349](https://doi.org/10.1586/17476348.2.3.349).
- Lévy P, Kohler M, McNicholas WT, Barbé F, McEvoy RD, Somers VK, Lavie L, Pépin J-L. 2014.** Obstructive sleep apnoea syndrome. *Nature Reviews. Disease Primers* **1**:15015–15015.
- Marti-Soler H, Hirotsu C, Marques-Vidal P, Vollenweider P, Waeber G, Preisig M, Tafti M, Tufik SB, Bittencourt L, Tufik S, Haba-Rubio J, Heinzer R. 2016.** The NoSAS score for screening of sleep-disordered breathing: a derivation and validation study. *The Lancet Respiratory Medicine* **4**(9):742–748 DOI [10.1016/S2213-2600\(16\)30075-3](https://doi.org/10.1016/S2213-2600(16)30075-3).
- McNicholas WT, Bonsignore MR, Management Committee of EU Cost Action B26. 2007.** Sleep apnoea as an independent risk factor for cardiovascular disease: current evidence, basic mechanisms and research priorities. *European Respiratory Journal* **29**(1):156–178.

- Memtsoudis SG, Besculides MC, Mazumdar M. 2013.** A rude awakening—the perioperative sleep apnea epidemic. *New England Journal of Medicine* **368**(25):2352–2353 DOI [10.1056/NEJMp1302941](https://doi.org/10.1056/NEJMp1302941).
- Menche J, Sharma A, Kitsak M, Ghiassian SD, Vidal M, Loscalzo J, Barabási A-L. 2015.** Uncovering disease-disease relationships through the incomplete interactome. *Science* **347**(6224):1257601 DOI [10.1126/science.1257601](https://doi.org/10.1126/science.1257601).
- Netzer NC, Stoohs RA, Netzer CM, Clark K, Strohl KP. 1999.** Using the Berlin Questionnaire to identify patients at risk for the sleep apnea syndrome. *Annals of Internal Medicine* **131**(7):485–491.
- Noack A. 2003.** An energy model for visual graph clustering. In: *International symposium on graph drawing*. Springer, 425–436.
- Noack A. 2009.** Modularity clustering is force-directed layout. *Physical Review E* **79**(2):026102.
- Pelletier-Fleury N, Meslier N, Gagnadoux F, Person C, Rakotonanahary D, Oukel H, Fleury B, Racineux J. 2004.** Economic arguments for the immediate management of moderate-to-severe obstructive sleep apnoea syndrome. *European Respiratory Journal* **23**(1):53–60 DOI [10.1183/09031936.03.00066903](https://doi.org/10.1183/09031936.03.00066903).
- Peppard PE, Young T, Barnet JH, Palta M, Hagen EW, Hla KM. 2013.** Increased prevalence of sleep-disordered breathing in adults. *American Journal of Epidemiology* **177**(9):1006–1014 DOI [10.1093/aje/kws342](https://doi.org/10.1093/aje/kws342).
- Punjabi NM. 2008.** The epidemiology of adult obstructive sleep apnea. *Proceedings of the American Thoracic Society* **5**(2):136–143 DOI [10.1513/pats.200709-155MG](https://doi.org/10.1513/pats.200709-155MG).
- Ralls FM, Grigg-Damberger M. 2012.** Roles of gender, age, race/ethnicity, and residential socioeconomic in obstructive sleep apnea syndromes. *Current Opinion in Pulmonary Medicine* **18**(6):568–573 DOI [10.1097/MCP.0b013e328358be05](https://doi.org/10.1097/MCP.0b013e328358be05).
- Rechtschaffen A, Kales A. 1968.** A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects.
- Rossi VA, Stradling JR, Kohler M. 2013.** Effects of obstructive sleep apnoea on heart rhythm. *European Respiratory Journal* **41**(6):1439–1451 DOI [10.1183/09031936.00128412](https://doi.org/10.1183/09031936.00128412).
- Rozenblatt-Rosen O, Deo RC, Padi M, Adelmant G, Calderwood MA, Rolland T, Grace M, Dricot A, Askenazi M, Tavares M, Pevzner SJ, Abderazzaq F, Byrdsong D, Carvunis AR, Chen AA, Cheng J, Correll M, Duarte M, Fan C, Feltkamp MC, Ficarro SB, Franchi R, Garg BK, Gulbahce N, Hao T, Holthaus AM, James R, Korkhin A, Litovchick L, Mar JC, Pak TR, Rabello S, Rubio R, Shen Y, Singh S, Spangle JM, Tasan M, Wanamaker S, Webber JT, Roecklein-Canfield J, Johannsen E, Barabási AL, Beroukhi R, Kieff E, Cusick ME, Hill DE, Münger K, Marto JA, Quackenbush J, Roth FP, DeCaprio JA, Vidal M. 2012.** Interpreting cancer genomes using systematic host network perturbations by tumour virus proteins. *Nature* **487**(7408):491–495 DOI [10.1038/nature11288](https://doi.org/10.1038/nature11288).
- Santos-Silva R, Castro LS, Taddei JA, Tufik S, Bittencourt LRA. 2012.** Sleep disorders and demand for medical services: evidence from a population-based longitudinal study. *PLOS ONE* **7**(2):e30085 DOI [10.1371/journal.pone.0030085](https://doi.org/10.1371/journal.pone.0030085).

- Sánchez-de-la Torre M, Campos-Rodriguez F, Barbé F. 2013.** Obstructive sleep apnoea and cardiovascular disease. *The Lancet Respiratory Medicine* **1**(1):61–72 DOI [10.1016/S2213-2600\(12\)70051-6](https://doi.org/10.1016/S2213-2600(12)70051-6).
- Schlosshan D, Elliott M. 2004.** Sleep 3: clinical presentation and diagnosis of the obstructive sleep apnoea hypopnoea syndrome. *Thorax* **59**(4):347 DOI [10.1136/thx.2003.007179](https://doi.org/10.1136/thx.2003.007179).
- Sharma A, Gulbahce N, Pevzner SJ, Menche J, Ladenvall C, Folkersen L, Eriksson P, Orho-Melander M, Barabási A-L. 2013.** Network-based analysis of genome wide association data provides novel candidate genes for lipid and lipoprotein traits. *Molecular & Cellular Proteomics* **12**(11):3398–3408 DOI [10.1074/mcp.M112.024851](https://doi.org/10.1074/mcp.M112.024851).
- Silva GE, Vana KD, Goodwin JL, Sherrill DL, Quan SF. 2011.** Identification of patients with sleep disordered breathing: comparing the four-variable screening tool, STOP, STOP-Bang, and Epworth Sleepiness Scales. *Journal of Clinical Sleep Medicine* **7**(5):467–472 DOI [10.5664/JCSM.1308](https://doi.org/10.5664/JCSM.1308).
- Simon S, Collop N. 2012.** Latest advances in sleep medicine: obstructive sleep apnea. *CHEST Journal* **142**(6):1645–1651 DOI [10.1378/chest.12-2391](https://doi.org/10.1378/chest.12-2391).
- Therneau TM, Atkinson B, Ripley B. 2010.** rpart: recursive partitioning. R package version, 3. 1–46.
- Topirceanu A, Udrescu M, Muresan A, Mihaicuta S. 2014.** Network-based assessment of sleep apnea syndrome treatment response. *European Respiratory Journal* **44**(Suppl 58):P2215.
- Udrescu L, Sbârcea L, Topirceanu A, Iovanovici A, Kurunczi L, Bogdan P, Udrescu M. 2016.** Clustering drug-drug interaction networks with energy model layouts: community analysis and drug repurposing. *Scientific Reports* **6**:32745 DOI [10.1038/srep32745](https://doi.org/10.1038/srep32745).
- Udrescu M, Topirceanu A, Avram R, Mihaicuta S. 2014.** AER score: a social-network-inspired predictor for sleep apnea syndrome. *Chest* **145**(3):609A-1 DOI [10.1378/chest.1824353](https://doi.org/10.1378/chest.1824353).
- Utriainen KT, Airaksinen JK, Polo O, Raitakari OT, Pietilä MJ, Scheinin H, Helenius HY, Leino KA, Kentala ES, Jalonen JR, Hakovirta H, Salo TM, Laitio TT. 2013.** Unrecognised obstructive sleep apnoea is common in severe peripheral arterial disease. *European Respiratory Journal* **41**(3):616–620 DOI [10.1183/09031936.00227611](https://doi.org/10.1183/09031936.00227611).
- Vagiakis E, Kapsimalis F, Lagogianni I, Perraki H, Minaritzoglou A, Alexandropoulou K, Roussos C, Kryger M. 2006.** Gender differences on polysomnographic findings in Greek subjects with obstructive sleep apnea syndrome. *Sleep Medicine* **7**(5):424–430 DOI [10.1016/j.sleep.2005.12.014](https://doi.org/10.1016/j.sleep.2005.12.014).
- Vavougiou GD, Natsios G, Pastaka C, Zarogiannis SG, Gourgoulianis KI. 2016.** Phenotypes of comorbidity in OSAS patients: combining categorical principal component analysis with cluster analysis. *Journal of Sleep Research* **25**(1):31–38 DOI [10.1111/jsr.12344](https://doi.org/10.1111/jsr.12344).
- Vidal M, Cusick ME, Barabasi A-L. 2011.** Interactome networks and human disease. *Cell* **144**(6):986–998 DOI [10.1016/j.cell.2011.02.016](https://doi.org/10.1016/j.cell.2011.02.016).

- Villaneuva ATC, Buchanan PR, Yee BJ, Grunstein RR. 2005.** Ethnicity and obstructive sleep apnoea. *Sleep Medicine Reviews* **9(6)**:419–436 DOI [10.1016/j.smrv.2005.04.005](https://doi.org/10.1016/j.smrv.2005.04.005).
- Yaggi HK, Concato J, Kernan WN, Lichtman JH, Brass LM, Mohsenin V. 2005.** Obstructive sleep apnea as a risk factor for stroke and death. *New England Journal of Medicine* **353(19)**:2034–2041 DOI [10.1056/NEJMoa043104](https://doi.org/10.1056/NEJMoa043104).
- Ye L, Pien GW, Ratcliffe SJ, Björnsdóttir E, Arnardóttir ES, Pack AI, Benediktsdóttir B, Gislason T. 2014.** The different clinical faces of obstructive sleep apnoea: a cluster analysis. *European Respiratory Journal* **44(6)**:1600–1607 DOI [10.1183/09031936.00032314](https://doi.org/10.1183/09031936.00032314).
- Yıldırım MA, Goh K-I, Cusick ME, Barabási A-L, Vidal M. 2007.** Drug—target network. *Nature Biotechnology* **25(10)**:1119–1126 DOI [10.1038/nbt1338](https://doi.org/10.1038/nbt1338).
- Young T, Peppard PE, Gottlieb DJ. 2002.** Epidemiology of obstructive sleep apnea: a population health perspective. *American Journal of Respiratory and Critical Care Medicine* **165(9)**:1217–1239 DOI [10.1164/rccm.2109080](https://doi.org/10.1164/rccm.2109080).