scientific reports

OPEN

The role of the social network in the study of adherence to diabetic retinopathy screening programs

Andreia Penso Pereira[1]✉, Ana Afonso[2,5], Raul M. S. Laureano[3]✉ & Fernando Buarque de Lima Neto[4]

Diabetic retinopathy screenings are a vital strategy to avoid the severe consequences of this disease. However, their success depends on the adherence of the target population. The present work aims to review the adherence to diabetic retinopathy screening, more specifically the influence of the persons with diabetes' social network (contacts between persons with diabetes) on their screening behaviour. The used data set comprises information of 75,921 persons with diabetes, distributed by 20 Primary Health Centre Groups of the Portuguese North Region. Persons with diabetes of the same Group were organized in an N-by-N matrix, resulting in 20 social networks. Network metrics were calculated and its relationship with the adherence to screening was analysed using two perspectives: correlation between global network metrics and adherence rate; cluster analyses based on node level metrics. The results obtained show that: (1) Less connected networks, strongly divided into communities and with a great number of connected components, present the highest adherence rates. (2) The node level metrics allow the identification of groups where the problem of non-adherence is especially high. (3) The non-adherence phenomenon is especially evident in a small group of highly connected individuals. We believe that these results are of utmost relevance as a starting point for future research and as support to the planning of interventions related to diabetic retinopathy screening adherence.

Keywords Health screenings, Diabetic retinopathy screening, Social networks, Screening adherence, Patient-level factors influence in adherence to screening

Nowadays it is consensual that social networks influence health behaviours[1]. Since the early works on the influence of social network characteristics on suicidal behaviour[2–4], there has been a growing interest in this field of research. Several studies proved the influence of the social network structure in disease spreading[5,6], the habit of smoking cigarettes[7,8], physical activity and eating habits[9–12]and risk behaviours[13]. Concerning screenings adherence, the research is much sparser, and the conclusions are not consistent. A first study concluded that social networks have an important influence on cancer-screening behaviour among low-income, older Mexican American women[14], however, in the continuation of the research, the authors found out that the effect is not universal across Hispanic groups[15]. Other researchers, aimed to identifying the main characteristics of successful interventions to promote cancer screening adherence and concluded that effective interventions need to use a variety of strategies, including the structure of the social network[16]. A study focused on the relationship between social network characteristics and breast cancer screening practices among employed women, found statistically significant relationships between network characteristics and screening behaviour, after removing the effects of previous mammography screening adherence[17]. A study that examines the influence of social networks in colorectal cancer screening adherence found that individuals who are socially isolated are less likely to adhere to the screening[18]. Another study regarding the influence of the social network in cancer screening adherence concluded that the screening behaviour of siblings, friends, or co-workers does not have significant influence

[1]Information Sciences, Technologies and Architecture Research Center (ISTAR-IUL), Instituto Universitário de Lisboa (ISCTE-IUL), Av. das Forças Armadas, 1649-026 Lisboa, Portugal. [2]Global Health and Tropical Medicine, GHTM, Associate Laboratory in Translation and Innovation Towards Global Health, LA-REAL, Instituto de Higiene e Medicina Tropical, IHMT, Universidade NOVA de Lisboa, UNL, Rua da Junqueira 100, 1349-008 Lisboa, Portugal. [3]Instituto Universitário de Lisboa (ISCTE-IUL), Business Research Unit (BRU-IUL), Av. das Forças Armadas, 1649-026 Lisboa, Portugal. [4]Universidade de Pernambuco (UPE), Escola Politécnica, Computer Engineering, (POLI/PPG-EC), Rua Benfica, 455 – Bloco "C", Recife 50720-001, Brazil. [5]Escola Superior de Saúde de Santarem, Instituto Politécnico de Santarém, Quinta do Mergulhão, Srª da Guia, 2005-075 Santarém, Portugal. ✉email: andreia_marisa_pereira@iscte-iul.pt; raul.laureano@iscte-iul.pt

on cancer screening behaviours[19]. Table 1 presents more details about the strengths and main outcomes of these research.

The literature focused on identifying the factors that induce the adherence to diabetic retinopathy (DR) screenings, highlights the importance of some sociodemographic variables, health status, knowledge about the disease, and previous health related behaviour[20–23]. Despite de interesting conclusions of these studies, we could not find any research related with the influence of the persons with diabetes' social network on the adherence to DR screening. The present work aims to contribute to fulfil this gap, reviewing adherence to DR screening, more specifically the influence of the persons with diabetes' social network structure (contacts with other persons with diabetes) in the adherence to DR screening. We aim to identify: 1—the global metrics of the persons with diabetes' social network and if they are significantly related with DR screening adherence rate, and 2—specific groups of persons with diabetes, concerning their individual social network features and their screening behaviours.

We believe that the results obtain could be particularly relevant as a starting point for future research and also as a framework to support the planning of interventions related with adherence to DR screenings.

	Refs	Authors	Year	Strengths and main outcomes
The influence of social network characteristics on health-related behaviour	2	E. Durkheim	1897	Durkheim studied the connections between individuals and society, demonstrating the usefulness of sociology as a science. Abnormally low or high levels of social integration can result in increased suicide rates
	3	B. A. Pescosolido e S. Georgianna	1989	This article analyses the characteristics of individuals' social networks to deepen the study of Durkheim's general proposition regarding the protective power of religion, with regard to suicide
	4	P. S. Bearman e J. Moody	2004	The authors concluded that friendship environment affects suicidality and that female adolescents' suicidal thoughts are significantly increased by social isolation
	5	A. S. Klovdahl, E. A. Graviss, A. Yaganehdoost, M. W. Ross, A. Wanger, G. J. Adams e J. M. Musser	2001	The authors use social network methods to reconstruct a tuberculosis outbreak network and to quantify the relative importance o persons and places in that outbreak (betweenness' centrality). This work provides the basis for a new approach to outbreak investigation and disease control
	6	L. A. Meyers, B. Pourbohloul, M. E. Newman, D. M. Skowronski e R. C. Brunham	2005	Traditionally epidemiology assumed that each individual has an equal chance of spreading the disease to everyone else, this study questions this assumption. The authors apply epidemiology methods to the contact network to illustrate that, for a single value of R0, any two outbreaks, even in the same environment, can have very different epidemiological outcomes
	7	S. T. Ennett e K. E. Bauman	1993	Social network theory and analysis were applied to examine the relation between adolescents' social positions and current smoking prevalence. The authors conclude that the chances of being a smoker are significantly higher for isolated adolescents. The relationship was not explained by demographic variables or the number of friends who smoked
	8	N. A. Christakis e J. H. Fowler	2008	The authors used network analytic methods and longitudinal statistical models to determine the extent of the person-to-person spread of smoking and quitting behaviour. The authors concluded that the social network is relevant to these behaviours
	9	J. Zhang, D. Brackbill, S. Yang, J. Becker, N. Herbert e D. Centola	2016	The authors demonstrate that social networks can play an important role in the design of more effective interventions for increasing children's physical activity
	10	J. Zhang, D. Brackbill, S. Yang e D. Centola	2015	
	11	J. Zhang, D. A. Shoham, E. Tesdahl e S. Gesell	2015	
	12	N. A. Christakis e J. . H. Fowler	2007	The authors concluded that the social network is relevant to the behavioural trait of obesity, and obesity appears to spread through social ties
	13	T. W. Valente, S. C. Watkins, M. N. Jato, A. Van Der Straten e L. P. M. Tsitsol	1997	The authors studied the association between social networks and contraceptive use. They concluded that the personal network is associated with contraceptive use. This association is even more significative than the individual characteristics usually considered relevant
The influence of social network characteristics on screening behaviour	14	L. Suarez, L. Lloyd, N. Weiss, T. Rainbolt e L. Pulley	1994	This research aims to determine the extent to which differences in social networks are relevant for adherence to breast and cervical cancer-screening, among low-income Mexican American women. The authors concluded that social networks seem to be a relevant factor for cancer-screening behavior in this group of women
	15	L. Suarez, A. G. Ramirez, R. Villarreal, J. Marti, A. McAlister, G. A. Talavera, E. Trapido e E. J. Perez-Stable	2000	The focus of this research is the influence of social integration on cancer screening participation of Hispanic women. The authors concluded that social networks have the potential to change screening behaviour. However, they also highlight that the influence of the social network was not universal across Hispanic groups and was stronger for Pap smear than for mammography screening behaviour
	16	B. Curbow, J. Bowie, M. A. Garza, K. McDonnell, L. B. Scott, C. A. Coyne e T. Chiappelli	2004	The authors preformed a comprehensive literature review of community-based breast, cervical and colorectal cancer screening interventions, aiming to Identify characteristics of the most successful ones. Their results show that effective interventions combined a variety of strategies, including the use of social networks
	17	J. D. Allen, A. M. Stoddard e G. Sorensen	2008	The authors examined the relationship between social network characteristics and adherence to breast cancer screening. The results obtained indicate that social network characteristics have a modest impact on screening, and that previous adherence is the main predictive factor of future behaviour
	18	J. Ye, S. D. Williams e Z. Xu	2009	The aim of this research was to analyse the relationship between social networks and colorectal cancer screening adherence. The authors concluded that individuals who were socially isolated were less likely to adhere to colorectal cancer screening
	19	N. L. Keating, A. J. O'Malley, J. M. Murabito, K. P. Smith e N. A. Christakis	2011	The aim of this research was to assess if adherence to screening for breast, prostate, or colorectal cancer is influenced by the screening behaviours of friends, coworkers, and close family members. The authors concluded that the screening behaviours of the network contacts had minimal influence on screening behaviours

Table 1. State of the art.

Methods

Data

This original research is based on data of the Portuguese North Regional Health Administration (ARSN) concerning persons with diabetes convened for DR population-based screening.

Were selected the subjects that met cumulatively the following inclusion criteria:

- Persons with diabetes registered in the ARSN's primary health care units;
- Persons with diabetes convened for DR screening during the year 2018.

The subjects correspond to 75,921 persons with diabetes invited for DR screening, distributed by twenty ACES, as illustrated in Table 2. Figure 1 illustrates the geographic area covered by each ACES.

The initial data set includes the following variables: age; gender; 7-digit home address postal code; professional status; ACES; Primary Health Care Unit; type of Primary Health Care Unit; family file number in Primary Health Care Unit (encrypted); existence or not of a family physician, number of consultations at the Primary Health Care Unit in the last 12 months, and type of diabetes. Subsequently, data from the National Institute of Statistics (INE)[25]were used to obtain the variables "income (median)" and "educational qualifications (distribution)", by postal code with 7 digits, as these variables are identified in the literature as strongly related to the adherence rate[22].

All methods of data gathering were carried out in accordance with relevant guidelines and regulations. The experimental protocol was approved by ARSN.

The authors did not have any direct contact with the subjects participating in the study.

The data obtained from INE are publicly available and of a general nature, not allowing the identification of the subjects involved. The data provided by ARSN were collected by the institution, in accordance with the legislation applicable in the Portuguese Public Administration, including informed consent from all subjects and/or their legal guardian(s)[26,27].. Moreover, the data provided by ARSN for this research went through a set of mechanisms that guarantee the protection of privacy (for example encryption and anonymization), and all the procedures were duly endorsed by the ARSN ethics committee, in strict compliance with all issues related to access to Public Administration data, and the data protection regime.

Network construction

We consider three types of relationships obtained through the variables provided by ARSN. The first relationship (probability equal to 1) was based exclusively on the existence of a close family relationship, obtained through the family file number registered in the primary health care unit. The second type of connection is based on

ACES	Resident population	Persons with diabetes	Persons with diabetes invited for DR screening
Alto Ave	256,696	22,028	5561
Alto Tâmega e Barroso	94,143	9062	0
Aveiro Norte	113,188	9603	1653
Baixo Tâmega	182,125	14,766	2514
Barcelos/ Esposende	154,645	13,312	4140
Braga	181,494	13,140	2691
Douro Sul	74,095	7316	0
Espinho/ Gaia	183,524	16,365	919
Famalicão	133,832	10,176	3279
Feira/ Arouca	161,671	12,137	3307
Gaia	152,062	12,660	1840
Gerês/ Cabreira	108,913	9052	3040
Gondomar	166,522	15,148	2984
Maia/ Valongo	229,164	17,564	4516
Marão e Douro Norte	105,025	10,030	0
Porto Ocidental	136,369	12,038	174
Porto Oriental	101,222	9743	2857
Póvoa de Varzim/ Vila do Conde	142,941	12,575	3226
Santo Tirso/ Trofa	110,529	10,698	2674
ULS Alto Minho	244,836	22,253	7802
ULS Matosinhos	175,478	15,129	3660
ULS Nordeste	136,252	13,511	4180
Vale Sousa Norte	161,792	13,465	2008
Vale Sousa Sul	175,852	13,724	0

Table 2. Distribution of the persons with diabetes invited for DR Screening, number of persons with diabetes and resident population by ACES.

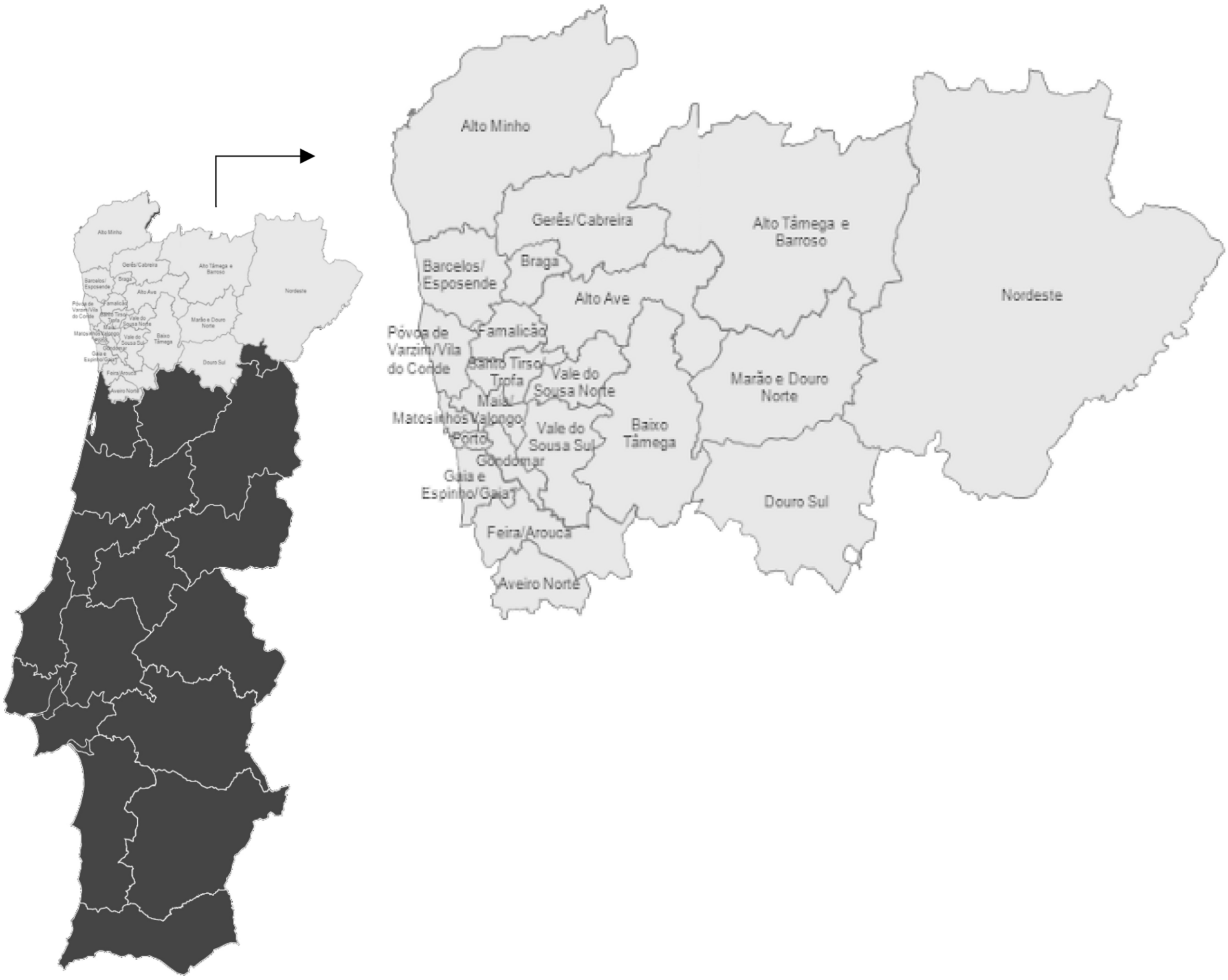

Fig. 1. ARSN's ACES geographical coverage. (Adapted from[24]).

the variables age and postal code with seven digits of the area of residence, namely the possible existence of a relationship was considered when the persons with diabetes live in the same postal code and the age difference is less than 5 years. The third type of relationship is based on the primary care unit where the person with diabetes is registered and the number of times he/she has had an appointment in the last 12 months. We considered the possibility of the existence of a relationship when the persons with diabetes are enrolled in the same health unit and had at least 5 consultations in the last 12 months. As it was not possible to accurately determine the probabilities in the last two types of relationships, it was decided to consider the value of 0.25 in both cases. The literature focused on the study of human interactions in near geographic spaces[28,29] provided some support for the reasonableness of the assumption in the case of geographic proximity to the residence (second type of relationship). As for the third type of relationship, based on diabetes consultations in the same health unit, we were advised by ARSN experts. According to them, it is common practice to schedule these appointments for the same day and at the same hour, for reasons of logistical ease of services. Therefore, it is not uncommon for the time spent in the waiting room to be prolonged and for persons with diabetes to end up establishing some complicit relationships.

After defining the probability of a social relation, persons with diabetes (nodes) of the same ACES were organized in an N-by-N square matrix using SPSS Modeler 18.2 software. The data entries represent a relationship between a pair of nodes (edges). Twenty social networks were built, one for each ACES, which reflect the way the diabetic population relates to each other. Figure 2presents the graphic representation of each of the 20 social networks (one for each ACES) built on Gephi software, using the force Atlas algorithm for network spatialization and to help its interpretation[30]. Visually the networks are made up of dots, which correspond to the persons with diabetes, and edges that represent the existence of a relationship between two persons with diabetes in the network. The edges are thicker the greater the probability of the relationship existing. Different colours were assigned to correspond to different communities. For example, in the case of ACES Espinho/Gaia, made up of health centres in the area covered by two main cities (Espinho and Gaia), we can observe the existence of two main communities (blue and green dots) strongly connected. There are also numerous less connected persons with diabetes (red dots).

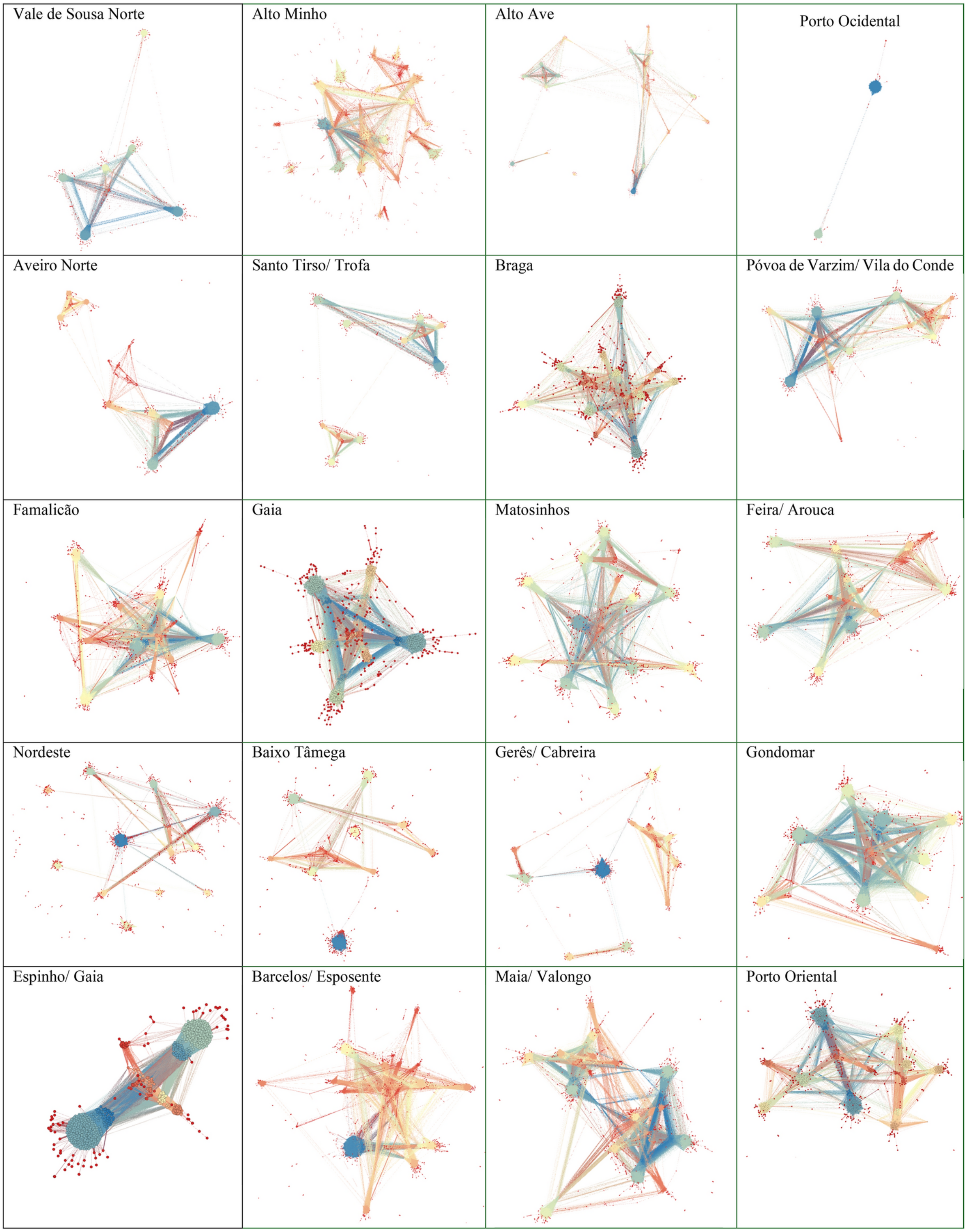

Fig. 2. Network's graphic representation.

Social networks evaluation metrics

After obtaining the 20 social networks, the Gephi software was also used to calculate two different sets of metrics, differentiated by the level of analysis: metrics at the level of the whole network; and node level metrics. The first set of metrics provides more compact information and allows the assessment of the overall structure of the network, giving insights about important properties of the underlying social phenomena. The second explores individual metrics to understand how the position of a node (individual) within the overall structure of the graph, helps to understand behavioural patterns. Tables 3 and 4present, respectively, the list of global and individual metrics calculated for this research[31]. Modularity, Connected Components, Average Degree, and

Network Metrics	Description
Number of Nodes	Number of individuals composing the network
Number of Edges	Number of relations (interactions) between individuals
Average Degree	The average degree is the mean of the degrees of all nodes in a network
Avg. Weighted Degree	Average sum of weights of the edges of nodes
Network Diameter	The diameter is given by the maximum eccentricity of the set of vertices in the network. Sparser networks have generally greater diameter than full matrices, due to the existence of fewer paths between pairs of nodes. This metric gives an idea about the proximity of pairs of nodes in the network, indicating how far two nodes are, in the worst of cases
Graph Density	Density can explain the general level of connectedness in a network. It is given by the proportion of edges in the network relative to the maximum possible number of edges. It goes from a minimum of 0, when a network has no edges at all, to a maximum of 1, when the network is perfectly connected (also called complete graph or clique)
Modularity	Modularity metrics strength of division of a network into communities (modules, clusters). Metrics takes values from range $< -1, 1 >$. Value close to 1 indicates strong community structure. When $Q = 0$ then the community division is not better than random
Connected Components	Connected components refer to a set of vertices that are connected to each other by direct or indirect paths. In other words, a set of vertices in a graph is a connected component if every node in the graph can be reached from every other node in the graph
Avg. Cluster Coefficient	The local clustering of each node is the fraction of triangles that actually exist over all possible triangles in its neighbourhood. Roughly speaking it tells how well connected the neighbourhood of the node is. If the neighbourhood is fully connected, the clustering coefficient is 1 and a value close to 0 means that there are hardly any connections in the neighbourhood. The average clustering coefficient of a graph is the mean of local clustering
Avg. Path Length	Average path length is a concept in network topology that is defined as the average number of steps along the shortest paths for all possible pairs of network nodes. It is a measure of the efficiency of information or mass transport on a network

Table 3. Description of network level metrics.

Node Level Metrics	Description
Eccentricity	The eccentricity measure captures the distance between a node and the node that is furthest from it
Closeness Centrality	Closeness centrality is a measure that indicates how close a node is to all the other nodes in a network. A high closeness centrality means that there is a large average distance to other nodes in the network
Harmonic Closeness centrality	Harmonic Centrality is a variant of Closeness Centrality, that reverses the sum and reciprocal operations in graphs with unconnected clusters, the harmonic centrality could be a better indicator of centrality than closeness centrality
Betweenness Centrality	Betweenness centrality is a measure based on the number of shortest paths between any two nodes that pass through a particular node. Nodes around the edge of the network would typically have a low betweenness centrality. A high betweenness centrality might suggest that the individual is connecting various parts of the network together
Degree	The degree of a node is the number of relation (edge) it has. It is the sum of edges for a node
Weighted Degree	The weighted degree is based on the number of edges for a node but pondered by the weight of each edge. It is the sum of the weight of the edges
Authority	The authority indicates the value of the information that the node holds. The relevance of an authority is "measured" by the number of inward links (or simply by the number of links in undirected graphs)
Modularity Class	Modularity class identifies nodes that are more densely connected than to the rest of the network. Those nodes have the same modularity class
Component Number	A connected component of an undirected graph is a maximal set of nodes such that a path connects each pair of nodes. The component number identifies a group of nodes that belong to the same components
Clustering	Clustering is the fraction of triangles that do exist over all possible triangles in its neighbourhood. Roughly speaking it tells how well connected the neighbourhood of the node is. If the neighbourhood is fully connected, the clustering coefficient is 1 and a value close to 0 means that there are hardly any connections in the neighbourhood
Triangles	Counts the number of triangles for each node in the graph. A triangle is a set of three nodes where each node has a relationship to the other two. In graph theory terminology, this is sometimes referred to as a 3-clique Triangle counting is used to detect communities and measure the cohesiveness of those communities. It is also used to determine clustering coefficients
Eigen centrality	Eigenvector centrality is a centrality index that calculates the centrality of a node based not only on their connections, but also based on the centrality of that node's connections

Table 4. Description of node level metrics.

Average Path Length are the most relevant global metrics for our research. Degree and Weight Degree are the most relevant individual metrics.

Global and node level network metrics relationship with the adherence to screening was analysed using two different perspectives: correlation between global level network metrics and ACES adherence rate and cluster analyses based on node level network metrics. For the first analysis, Pearson and Spearman correlations[32] between global level network metrics and the ACES adherence rate were calculated, based on the results obtain from each ACES network (Table 3). The clusters analysis using the k-means algorithm was conducted with IBM SPSS Modeler 18.2[33],, being the inputs, the node level metrics described in Table 4, and the "Adherence", that assumes the value 1 when the diabetic adhered to the screening and 0 otherwise.

Statistical analysis

This research aims to analyse the influence of the persons with diabetes' social network structure in the adherence to DR screening, by prosecuting to specific objectives: 1 – identify the correlation of the networks global metrics and the ACSE screening adherence rate. 2 – identify specific groups of persons with diabetes, concerning their individual social network features and their screening behaviours. To achieve the first goal, global network

ACES	Nodes	Edges	Average Degree	Avg. Weighted Degree	Network Diameter	Graph Density	Modularity	Connected Components	Avg. Cluster Coefficient	Avg. Path Length	Adherence rate
Vale do Sousa Norte	2008	197,154	196.369	4933.765	9	0.098	0.801	32	0.943	3.123	75.94%
ULS Alto Minho	7802	846,604	217.022	5480.531	14	0.028		103	0.933	5.302	73.36%
Alto Ave	5561	467,894	168.277	4235.794	14	0.03	0.907	124	0.942	4.567	72.49%
Aveiro Norte	1653	121,016	146.42	3691.349	10	0.089	0.753	25	0.935	3.628	70.68%
Santo Tirso/ Trofa	2674	240,044	179.539	4518.811	11	0.067	0.85	49	0.943	4.104	70.05%
Braga	2691	137,013	101.831	2559.922	10	0.038	0.864	72	0.915	3.207	67.61%
Famalicão	3279	272,923	166.467	4195.669	10	0.051	0.806	94	0.91	3.098	66.95%
Gaia	1840	155,317	168.823	4250.951	11	0.092	0.751	49	0.907	2.727	65.71%
UlS Matosinhos	3660	329,296	179.943	4520.082	10	0.049	0.881	96	0.942	3.221	65.40%
ULS Nordeste	4180	299,093	143.107	3608.026	15	0.034	0.892	134	0.928	5.105	64.67%
Baixo Tâmega	2514	205,256	163.29	4108.055	14	0.065	0.825	47	0.958	4.414	64.60%
Gerês/ Cabreira	3040	290,352	191.021	4810.411	13	0.063	0.832	52	0.946	4.697	63.63%
Espinho/ Gaia	919	105,076	228.675	5827.258	7	0.249	0.514	16	0.922	2.168	63.06%
Barcelos/ Esposende	4140	488,785	236.128	5957.391	10	0.057	0.779	52	0.939	3.423	62.37%
Maia/ Valongo	4516	477,864	211.632	5343.08	13	0.047	0.847	65	0.913	3.806	60.66%
Porto Oriental	2857	305,344	213.751	5380.942	9	0.075	0.813	45	0.937	2.901	57.38%
Gondomar	2984	356,444	238.903	6007.976	11	0.08	0.831	49	0.943	2.938	56.70%
Feira/ Arouca	3307	335,054	202.633	5106.033	11	0.061	0.848	32	0.916	3.3	54.07%
Póvoa Varzim/ Vila do Conde	3226	310,813	192.692	4844.529	10	0.06	0.83	49	0.929	3.177	52.37%
Porto Ocidental	174	5397	62.034	1561.207	7	0.359	0.452	4	0.984	2.232	44.36%

Table 5. Network level metrics. As showed, the adherence rate varies between 44.36% in ACES Porto Ocidental, and 75.94% in ACES Vale do Sousa Norte.

Network Measure	Correlation with Adherence Rate
Nodes	0.371
Edges	0.260
Average Degree	0.137
Avg. Weighted Degree	0.136
Network Diameter	0.363
Graph Density	−0.513**
Modularity	0.435*
Connected Components	0.440*
Avg. Cluster Coefficient	−0.261
Avg. Path Length	0.481**

Table 6. Pearson linear correlation coefficients. Notes: * p-value < = 0.1; ** p-value < = 0.05.

metrics were calculated, and Pearson and Spearman correlation coefficients were determined to assess a possible linear or ordinal association between each of the network metrics and the adherence rate[32]. To accomplish the second goal, was performed a node level cluster analyses based on diabetic (node) level network metrics, and his screening behaviour. Chi-square tests were performed to access the underline interference of socio-economic, health related and previous DR screening behaviour in the cluster's formation.

Results

Networks metrics and screening adherence

For each one of the 20 social networks were calculated 10 global metrics. Table 5 presents the values obtained and Table 6 presents the Pearson linear correlation coefficients between each measure and the adherence rate.

Modularity, Connected Components, and the Average Path Length present significant positive and moderate linear correlations (values between 0.4 and 0.5), showing that when these metrics increase the adherence rate also tends to increase. Graph Density is negatively and moderate correlated with the adherence rate, presenting a Pearson coefficient of −0.513, showing that when this metric increases the adherence rate tends to decrease.

Metrics for individual nodes and clusters analysis

The quality of the clusters analysis is classified as "fair" (Silhouette measure of cohesion and separation of 0.4. The centroids (average of the input variables) of each one of the five clusters obtained is presented in Table 7. Moreover, we perform the parametric test analysis of variance (one-way ANOVA) to determine whether

		Cluster—1	Cluster—2	Cluster—3	Cluster – 4	Cluster – 5	Total	ANOVA F test	η	η2
Inputs	Adherence	Yes (52.20%)	Yes (100.00%)	Yes (65.90%)	Yes (74.50%)	No (100.00%)	Yes (65,0%)	18,106.51 **	0.697	0.486
	Authority	0	0.05	0	0	0.03	0.01	43,952.45 **	0.890	0.792
	Closeness centrality	0.26	0.34	0.97	0.31	0.35	0.33	41,541.57 **	0.884	0.782
	Clustering	0.68	0.96	0.19	0.96	0.94	0.84	6658.12 **	0.604	0.365
	Degree	45.22	380.38	16.13	223.52	363.74	186.63	39,608.48 **	0.880	0.774
	Eccentricity	8.35	7.22	1.26	7.78	7.07	7.55	9438.939 **	0.670	0.449
	Eigen centrality	0.02	0.91	0	0.16	0.78	0.22	64,359.98 **	0.921	0.848
	Harmonic closeness centrality	0.29	0.42	0.98	0.36	0.43	0.38	38,785.89 **	0.878	0.770
	Triangles	2280.07	73,573.81	1284.42	25,137.59	64,469.60	24,519.94	18,497.75 **	0.784	0.615
	Weighted Degree	1141.34	9592.14	422.17	5621.78	9188.08	4700.93	38,687.49 **	0.877	0.770
	Betweenness centrality	1455.51	4203.59	2.48	5621.60	5280.26	4144.13	42.25 **	0.060	0.004
Size		28.66% (13,272)	9.29% (4,300)	4.55% (2,105)	50.71% (23,484)	6.80% (3,150)	100.00% (46,311)	100.00% (46,311)	100.00% (46,311)	100.00% (46,311)

Table 7. Clusters' centroids, ANOVA F test, eta (η) measure of association and effect size (η2). Notes * p_ value < = 0.1; ** p_value < = 0.05.

there are any statistically significant differences between the means of the input variables in the five clusters (independent groups) and we assess the effect size using the eta squared ratio (η^2). The results show significant differences between the means of all the input variables in the five clusters (all $p < 0.001$) and for all the input variables the effect size is considered large, except for Betweenness Centrality ($\eta^2 = 0.004$) where it is considered negligible Thus, these results reinforce the quality of the clusters and the differences between the five clusters[34].

The observation of each cluster characteristics leaded to the following analysis:

Cluster 1 – Poorly connected with low adherence: subjects with few connections to other elements of the diabetic community (Degree = 45.22; Weighted Degree = 1141.34), with an adherence below average, although most of the members adhere to the screening program (adherence rate = 52.2%).

Cluster 2 –Very connected, adherents: subjects very connected with other members of the diabetic population (Degree = 380.38; Weighted Degree = 9592.14), who adhered to screening (adherence rate = 100%).

Cluster 3 – Isolates, with average adherence: have few or no links to other persons with diabetes in the network (Degree = 16.13; Weighted Degree = 422.17). This group adherence is close to average.

Cluster 4 – Reasonably connected, with high adherence: subjects with a reasonable number of connections (Degree = 223.52; Weighted Degree = 5621.78) and an adherence above average.

Cluster 5 –Very connected, non-adherent: subjects very connected with other members of the diabetic population (Degree = 363.74; Weighted Degree = 9188.08), who did not adhere to screening.

Contrary to the conclusions of the previous studies, that state that individuals who are socially isolated are less likely to adhere to the colorectal cancer screening[18], the results obtained in this research revealed that, in DR screening, the group of "isolated" persons with diabetes is not the most problematic regarding adherence (adherence rate of 65.9%, slightly higher than the global adherence rate of 65.0%). In fact, the non-adherence phenomenon is especially evident in cluster 5, a group of highly connected individuals with 100% of non-adherence, which represents 6.7% of the target population. The second cluster with the lower adherence rate (52.2%) is cluster 1, a group of individuals with few connections with other persons with diabetes, but higher connected than the "Isolates" group, with corresponds to 28.8% of the target population.

Regarding a more general characterization of the persons with diabetes of each cluster, Table 8 presents Chi-squared test of independence and Cramer's V, between the five clusters, socio-demographic, health status, and health services utilization variables.

As we can see there is a significant relationship between the cluster and most of the variables analysed ($p < 0.001$). However, Cramer's V show that generally those variables have a low effect, except for the Number of primary health care consultations in the last 12 months (Cramer's V = 0.31). Indeed, the persons with diabetes of clusters 1 e 3 tend to have fewer consultations in the past 12 months. There is also a slight tendency for users of UCSP (health units dedicated to users without a family doctor) and/or USF model A (transition model for model B health units) to be more prevalent in clusters 1 and 3 (Cramer's V = 0.17).

Discussion

The present work aims to analyse the influence of the persons with diabetes' social network, more specifically their contacts with other members of the target population, in the individual decision of adhere or not to the screening. More specifically, it is intended to: analyse the influence of the characteristics of social networks in different regions on the adherence rates; categorize the persons with diabetes based on their social network and screening behaviour.

To conduct this research, global and node level network metrics were calculated and its relationship with the adherence to screening was studied using two different perspectives: correlation between global level network metrics and the Primary Health Centre Group adherence rate; cluster analyses based on node level network metrics.

Type	Variable	Cluster 1 - Poorly connected, low adherence	2 - Very connected, adherents	3 - Isolates, average adherence	4 - Reasonably connected, high adherence	5 - Very connected, non-adherent	Total	Chi-squared test Cramer's V
Size		28.7% (13,272)	9.3% (4,300)	4.6% (2,105)	50.7% (23,484)	6.8% (3,150)	100.0% (46,311)	
Sociodemographic	Age (years)							
	[18;39]	0.6%	0.6%	1.1%	1.0%	1.6%	0.9%	X2=469.45** Cramer's V=0.05
	[39;54]	6.9%	6.8%	10.9%	8.1%	8.4%	7.8%	
	[54;64]	22.5%	22.0%	28.9%	21.9%	16.3%	22.0%	
	[64;74]	33.9%	35.5%	35.0%	33.2%	25.7%	33.2%	
	≥74	36.1%	35.0%	23.8%	35.8%	48.0%	36.1%	
	Gender							
	Masculine	49.2%	52.0%	46.1%	53.1%	57.3%	51.8%	X2=117.59** Cramer's V=0.05
	Feminine	50.8%	48.0%	53.9%	46.9%	42.7%	48.2%	
	Degree of urbanization f the area of residence							
	0	10.0%	10.6%	7.2%	8.7%	6.5%	9.0%	X2=1378.51** Cramer's V=0.10
	1	20.4%	14.6%	12.8%	13.5%	7.5%	15.1%	
	2	27.3%	19.8%	36.1%	26.5%	13.0%	25.6%	
	5	42.3%	55.0%	43.9%	51.3%	73.0%	50.2%	
	Professional status							
	Active	42.7%	38.8%	53.1%	37.4%	32.9%	39.5%	X2=404.74** Cramer's V=0.05
	Student	0.2%	0.1%	0.7%	0.3%	0.5%	0.3%	
	Not active	11.5%	13.5%	11.0%	13.4%	12.6%	12.7%	
	Retired	45.1%	47.3%	34.6%	48.5%	53.7%	47.1%	
	Unknown	0.5%	0.3%	0.6%	0.4%	0.3%	0.4%	
	Income (median)							
	Unknown	2.2%	3.6%	3.2%	3.9%	3.8%	3.3%	X2=689.13** Cramer's V=0.06
	≤ 8,511	8.6%	8.3%	13.4%	10.2%	9.0%	9.6%	
	[8,511;9,811]	19.0%	16.6%	17.2%	20.6%	16.5%	19.3%	
	[9,811;11,167]	20.7%	18.9%	18.3%	19.3%	14.8%	19.3%	
	[11,167;12,649]	22.1%	26.5%	21.1%	16.0%	21.0%	19.3%	
	[12,649;17,400]	19.3%	18.6%	18.7%	19.3%	21.5%	19.4%	
	≥17,400	8.1%	7.5%	8.0%	10.6%	13.4%	9.7%	
	Education							
	Less than elementary school	2.2%	3.6%	3.2%	3.9%	3.8%	3.3%	X2=394.88** Cramer's V=0.06
	Less than middle school	20.3%	16.8%	21.3%	21.4%	17.3%	20.4%	
	Less than high school	55.3%	56.7%	55.6%	50.6%	47.3%	52.5%	
	High school	12.5%	14.4%	10.3%	11.8%	16.2%	12.5%	
	College degree	9.6%	8.6%	9.6%	12.2%	15.4%	11.2%	
Relationship with health services	Type of Health Unit							
	UCSP	24.0%	22.2%	17.6%	12.7%	19.9%	17.5%	X2=2686.02** Cramer's V=0.17
	USF A	25.7%	0.3%	28.9%	24.9%	7.6%	21.9%	
	USF B	50.3%	77.5%	53.5%	62.4%	72.5%	60.6%	
	Followed by family physician							
	No	13.8%	8.5%	15.9%	9.2%	16.3%	11.2%	X2=114.38** Cramer's V=0.04
	Yes	86.2%	91.5%	84.1%	90.8%	83.7%	88.8%	
	Number of consultations at the Primary Care Unit in the last 12 months							
	0	3.1%	0.0%	5.7%	0.0%	0.0%	1.1%	X2=17972.36** Cramer's V=0.31
	1	2.8%	0.0%	4.3%	0.0%	0.0%	1.0%	
	2–3	35.1%	0.0%	51.5%	0.3%	0.0%	12.6%	
	4–6	36.2%	33.6%	32.4%	35.3%	31.7%	35.0%	
	≥ 6	22.8%	66.4%	6.2%	49.1%	33.9%	41.6%	

Continued

Type	Variable	Cluster						
		1 - Poorly connected, low adherence	2 - Very connected, adherents	3 - Isolates, average adherence	4 - Reasonably connected, high adherence	5 - Very connected, non-adherent	Total	Chi-squared test Cramer's V
Health Status	Type of diabetes (I or II)							
	Type I	7.2%	7.4%	5.7%	8.0%	11.4%	7.8%	X2=78.68** Cramer's V=0.04
	Type II	92.8%	92.6%	94.3%	92.0%	88.6%	92.2%	
	Body Mass Index (BMI)							
	NA	67.50%	65.20%	64.70%	61.70%	75.70%	64.8%	X2=335.10** Cramer's V=0.04
	≤18.5	0.10%	0.10%	0.00%	0.10%	0.10%	0.1%	
	[18.5;24.9]	5.00%	5.10%	5.20%	5.70%	4.80%	5.4%	
	[25;30]	14.30%	15.70%	16.20%	15.90%	10.50%	15.1%	
	≥30	13.10%	13.90%	14.00%	16.50%	8.80%	14.6%	
	Blood Glucose Leves (HBA1C)							
	NA	67.50%	65.40%	64.90%	61.90%	75.80%	64.9%	X2=338.77** Cramer's V=0.06
	<8	28.70%	29.30%	31.10%	32.80%	19.30%	30.3%	
	≥8	3.80%	5.30%	4.00%	5.30%	4.80%	4.8%	
DR Screening	Days elapsed between calls							
	NA	16.80%	16.10%	16.20%	13.20%	13.70%	14.7%	X2=522.65** Cramer's V=0.05
	<365	21.00%	16.10%	20.50%	22.20%	13.80%	20.6%	
	[365;455]	25.60%	26.10%	31.10%	29.20%	27.00%	27.8%	
	[455;545]	15.30%	17.70%	13.40%	15.80%	15.00%	15.7%	
	[545;635]	6.70%	7.80%	6.80%	5.20%	8.50%	6.2%	
	≥635	14.60%	16.10%	12.10%	14.30%	22.00%	15.0%	
	Number of times the diabetic was convened							
	1	17.10%	16.10%	16.30%	13.50%	13.70%	14.9%	X2=643.42** Cramer's V=0.06
	2	23.90%	20.90%	22.40%	24.20%	24.10%	23.7%	
	3	41.40%	48.80%	36.70%	38.30%	48.00%	40.8%	
	4	15.60%	12.00%	20.70%	20.30%	13.00%	17.7%	
	5	2.00%	2.20%	3.90%	3.70%	1.20%	2.9%	

Table 8. Distribution of the diabetic population variables by cluster, Chi-squared independence test and Cramer's V. Notes: * p_value < = 0.1; ** p_value < = 0.05.

The results revealed that Modularity, Connected Components and the Average Path Length present significant positive Pearson linear correlations and that Graph Density is negatively correlated with the adherence rate.

The second perspective of analysis showed that node level metrics, associated with each diabetic position on the social network, allows the identification of groups where the problem of non-adherence is especially high. The analysis led to the identification of five different clusters:

Cluster 1 – Poorly connected: subjects with few connections to other elements of the diabetic community.

Cluster 2 – Adherents, very connected: subjects very connected with other members of the diabetic population, who adhere to screening.

Cluster 3 – Isolates: have few or no links to other persons with diabetes in the network.

Cluster 4 – Reasonably connected.

Cluster 5 – Non-Adherent, very connected: subjects very connected with other members of the diabetic population, who did not adhere to screening.

Contrary to the conclusions of previous studies, that individuals who are socially isolated are less likely to adhere to the colorectal cancer screening[18], the results obtained in this research demonstrate that, in DR screening, the group of "isolated" persons with diabetes is not the most problematic regarding adherence (adherence rate of 65.9%, slightly higher than the global adherence rate of 65.0%). The non-adherence phenomena is especially evident in cluster 5, a group of highly connected individuals with 100% of non-adherence, which represents 6.7% of the target population. In Portugal, there is a coexistence of a National Health Service, which tends to be free, and Private Health Care Providers where the user bears the costs. In this context, in meetings with ARSN experts, a hypothesis was put forward to explain a small percentage of non-adherence. They believed in the existence of a group of persons with diabetes (although not very significant) who do not adhere to screening because they are being monitored in the private sector. Theoretically, this group would be characterized by higher incomes, residence in urban areas (where most private institutions are located) and higher levels of education. E.g., Cluster 5 seems to bring together these characteristics, however more research will be needed to verify the validity of this hypothesis.

The second cluster with the lower adherence rate (52.2%) is cluster 1, a group of individuals with few connections with other persons with diabetes, but higher connected than the "Isolates" group, representing

28.8% of the target population. The persons with diabetes in this cluster, in general, received fewer previous calls for DR screening (what could indicate that the diabetes is more recent), and have more controlled HBA1C levels (lower risk of DR), which could be part of the explanation to the low screening adherence rate.

Among the strengths of this research, we highlight the used of real data and the large dimension of the data set (75.921 persons with diabetes distributed by 20 Primary Health Centre Groups of the Portuguese North Region Health Administration, invited for screening in 2018), in addition to sound techniques of Statistics and Network Science.

This research presents some limitations, namely, the social networks only comprise ties between members of the target population, neglecting other possible subjects that could influence the persons with diabetes decision of adhere or not to the screening; the links between persons with diabetes (edges) result of plausible, but not factual relations, except for the family relationship. However, even with the assumption of a 0.25 probability for type 2 and 3 relationships, the obtained very different networks that allowed to draw significant conclusions for the problem being studied. The former is next to be investigated in our research as well as with a robust sensitivity analysis to the probabilities here assumed.

Finally, more research is needed to better understand this phenomenon. The influence of the social network in DR screening could be studied considering different groups with different social and demographic features, like in the studies of Suarez et al., concerning cancer screening behaviour[14,15]. Would be important to assess the effectiveness of interventions that take into a count the structure of the social network, aiming to promote adherence to DR screening, and the influence of a broader social network, including members outside the diabetic community, should also be analysed. Some research was done in these areas focusing on cancer screenings[16,17], but not in DR screenings. In future work we intend to focus on some of these topics and test the predictive value of the persons with diabetes' social network features to their DR screening behaviour.

Conclusions

The results obtained allowed us to conclude that the structure of the social network and the position occupied by the diabetic in this network influence the behaviour of adherence to DR screening. Our research showed that less connected networks (where the average number of steps along the shortest path between two nodes is higher), strongly divided into communities and with a great number of connected components present the highest adherence rates. Node level metrics allowed the identification of groups where the problem of non-adherence is especially high. In our research, the non-adherence phenomenon is especially evident in a small group of highly connected individuals, which is contrary to the findings in the literature concerning oncologic screenings.

We think that these results are of utmost relevance as a starting point for future research and as a framework to support decision-making and planning of interventions related with adherence to DR screenings.

Data availability

The data obtained from INE are publicly available [25]. The data from the Portuguese North Region Health Administration, were used under license for the current study, and so are not publicly available. However, data are available from the authors upon reasonable request, addressed to the corresponding author, and with permission of the Portuguese North Region Health Administration.

Received: 18 September 2024; Accepted: 22 November 2024
Published online: 26 November 2024

References

1. Zhang, J. & Centola, D. Social Networks and Health: new developments in diffusion, online and offline. *Ann. Rev. Sociol.* **45**, 91–109 (2019).
2. Durkheim, E. *Le Suicide: Etude de sociologie* (Press Universitaires de France, 1897).
3. B. A. Pescosolido and S. Georgianna, "Durkheim, suicide, and religion: Toward a network theory of suicide," *American sociological review,* pp. 33–48, 1989.
4. Bearman, P. S. & Moody, J. Suicide and friendships among American adolescents. *Am. J. Public Health* **94**(1), 89–95 (2004).
5. Klovdahl, A. S. et al. Networks and tuberculosis: an undetected community outbreak involving public places. *Soc. Sci. Med.* **52**(5), 681–694 (2001).
6. Meyers, L. A., Pourbohloul, B., Newman, M. E., Skowronski, D. M. & Brunham, R. C. Network theory and SARS: predicting outbreak diversity. *J. Theor. Biol.* **232**(1), 71–81 (2005).
7. Ennett, S. T. & Bauman, K. E. Peer group structure and adolescent cigarette smoking: a social network analysis. *J. Health Soc. Behav.* **34**, 226–236 (1993).
8. Christakis, N. A. & Fowler, J. H. The collective dynamics of smoking in a large social network. *N. Engl. J. Med.* **358**(21), 2249–2258 (2008).
9. Zhang, J. et al. Support or competition? How online social networks increase physical activity: A randomized controlled trial. *Preventive medicine reports* **4**, 453–458 (2016).
10. Zhang, J., Brackbill, D., Yang, S. & Centola, D. Efficacy and causal mechanism of an online social media intervention to increase physical activity: Results of a randomized controlled trial. *Preventive medicine reports* **2**, 651–657 (2015).
11. Zhang, J., Shoham, D. A., Tesdahl, E. & Gesell, S. Network interventions on physical activity in an afterschool program: An agent-based social network study. *Am. J. Public Health* **105**(S2), S236–S243 (2015).
12. Christakis, N. A. & Fowler, J. H. The spread of obesity in a large social network over 32 years. *N. Engl. J. Med.* **357**(4), 370–379 (2007).
13. Valente, T. W., Watkins, S. C., Jato, M. N., Van Der Straten, A. & Tsitsol, L. P. M. Social network associations with contraceptive use among Cameroonian women in voluntary associations. *Soc. Sci. Med.* **45**(5), 677–687 (1997).
14. L. Suarez, L. Lloyd, N. Weiss, T. Rainbolt and L. Pulley, "Effect of social networks on cancer-screening behavior of older Mexican-American women," *JNCI: Journal of the National Cancer Institute,* vol. 86, no. 10, pp. 775–779, 1994.
15. Suarez, L. et al. Social networks and cancer screening in four US Hispanic groups. *Am. J. Prev. Med.* **19**(1), 47–52 (2000).
16. Curbow, B. et al. Community-based cancer screening programs in older populations: making progress but can we do better?. *Preventive Medicine* **38**(6), 676–693 (2004).

17. Allen, J. D., Stoddard, A. M. & Sorensen, G. Do social network characteristics predict mammography screening practices?. *Health Education & Behavior* **35**(6), 763–776 (2008).
18. Ye, J., Williams, S. D. & Xu, Z. The association between social networks and colorectal cancer screening in American males and females: data from the 2005 Health Information National Trends Survey. *Cancer Causes & Control* **20**(7), 1227–1233 (2009).
19. Keating, N. L., O'Malley, A. J., Murabito, J. M., Smith, K. P. & Christakis, N. A. Minimal social network effects evident in cancer screening behavior. *Cancer* **117**(13), 3045–3052 (2011).
20. Dervan, E., Lillis, D., Flynn, L., Staines, A. & O'shea, D. Factors that influence the patient uptake of diabetic retinopathy screening. *Irish journal of medical science* **177**(4), 303–308 (2008).
21. Mukamel, B. D., Bresnick, H. G., Wang, Q. & Dickey, F. C. Barriers to compliance with screening guidelines for diabetic retinopathy. *Ophthalmic Epidemiology* **6**(1), 61–72 (1999).
22. van Eijk, K., Blom, J. W., Gussekloo, J., Polak, B. C. P. & Groeneveld, Y. K.N.D. van Eijk, J.W. Blom, J. Gussekloo, B.C.P. Polak, Y. Groeneveld, Diabetic retinopathy screening in patients with diabetes mellitus in primary care: Incentives and barriers to screening attendance. *Diabetes Research and Clinical Practice* **96**(1), 10–16 (2012).
23. Newcomb, P. A. & Klein, R. Factors associated with compliance following diabetic eye screening. *Journal of Diabetic Complications* **4**(1), 8–14 (1990).
24. Administração Regional de Saúde do Norte, "Observatório Regional de Saúde," [Online]. Available: http://www.arsnorte.min-saude.pt/observatorio-regional-de-saude/perfis-de-saude/. [Accessed 29 09 2021].
25. INE, Statistics Portugal, [Online]. Available: https://www.ine.pt/xportal/xmain?xpgid=ine_main&xpid=INE. [Accessed Jan. 2023].
26. Lei n.º 58/2019, de 8 de agosto, Diário da República n.º 151/2019, Série I de 2019–08–08, páginas 3 – 40, https://diariodarepublica.pt/dr/detalhe/lei/58-2019-123815982
27. Despacho n.º 1668/2023, de 2 de fevereiro, Diário da República n.º 24/2023, Série II de 2023–02–02, páginas 63 – 68, https://diariodarepublica.pt/dr/detalhe/despacho/1668-2023-206847177
28. Arentze, T., van den Berg, P. & Timmermans, H. Modeling social networks in gographic space: approach and empirical application. *Environment and planning A: economy and space* **44**(5), 1101–1120 (2012).
29. Schram, M., Assendelft, W., van Tilburg, T. & Dukers-Muijrers, N. Social networks and type 2 diabetes: a narrative review. *Diabetologia* **64**(9), 1905–1916 (2021).
30. Jacomy, M., Venturini, T., Heymann, S. & Bastian, M. Force Atlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PloS one* **9**, 98679 (2014).
31. Marsden, P. V. Network data and measurement. *Annual Review of Sociology* **16**(1), 435–463 (1990).
32. Mukaka, M. Statistics corner: A guide to appropriate use of correlation coefficient in medical research. *Malawi Med J.* **24**(3), 69–71 (2012).
33. IBM SPSS Modeler, "SPSS Modeler 18.2 documentation," 18 12 2018. [Online]. Available: https://www.ibm.com/support/pages/spss-modeler-182-documentation. [Accessed 2021 09 17].
34. Sullivan, G. M. & Feinn, R. Using Effect Size - or why the P value is not enough. *J Grad Med Educ.* **3**, 279–282 (2012).

Acknowledgements
We are grateful to the doctors of the Northern Region Health Administration who provided data, insight and expertise that greatly assisted this research.

Author contributions
AP: Conceptualization, Methodology, Data preparation, Network construction, Data analyses. Writing original draft. AA: Conceptualization, Methodology, Expert advice, Writing- Reviewing and Editing. RL and FN: Conceptualization, Methodology, Supervision, Writing- Reviewing and Editing.

Funding
This work is partially funded by national funds through FCT—Fundação para a Ciência e Tecnologia, I.P., under the projects UID/GES/00315/2020 and UIDB/04466/2020, and by the ISCTE-IUL under the project HOPE.

Declarations

Competing interests
The authors declare no competing interests.

Ethics approval
All methods of data gathering were carried out in accordance with relevant guidelines and regulations. The authors did not have any direct contact with the subjects participating in the study. The data obtained from INE are publicly available and of a general nature, not allowing the identification of the subjects involved[25]. The data provided by ARSN were collected by the institution, in accordance with the legislation applicable in the Portuguese Public Administration, including informed consent from all subjects and/or their legal guardian(s)[26,27]. Moreover, the data provided by ARSN for this research went through a set of mechanisms that guarantee the protection of privacy (for example encryption and anonymization), and all the procedures were duly endorsed by the ARSN ethics committee, in strict compliance with all issues related to access to Public Administration data, and the data protection regime. The present research does not include the use of experimental protocols.

Additional information
Correspondence and requests for materials should be addressed to A.P.P. or R.M.S.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.