UNISINOS

Programa de Pós-Graduação em

Computação Aplicada

Doutorado Acadêmico

Henrique Damasceno Vianna

Pompilos: A Social Aware Model for Preventive Care of Non Communicable Diseases

São Leopoldo, 2019

Henrique Damasceno Vianna

Pompilos: A Social Aware Model for Preventive Care of Non Communicable Diseases

Thesis presented as a partial requirement to obtain the Doctor's degree from the Postgraduate Program in Applied Computation of the University of Vale do Rio dos Sinos — UNISINOS

Advisor:
Prof. Dr. Jorge Luis Victória Barbosa

São Leopoldo
2019

V617p Vianna, Henrique Damasceno.
Pompilos : a social aware model for preventive care of non communicable diseases / Henrique Damasceno Vianna. – 2019.
164 f. : il. ; 30 cm.

Tese (doutorado) – Universidade do Vale do Rio dos Sinos, Programa de Pós-Graduação em Computação Aplicada, 2019.
“Advisor: Prof. Dr. Jorge Luis Victória Barbosa.”

1. Doenças crônicas não transmissíveis. 2. Suporte Social. 3. Sistemas multi-agentes. 4. Modelo de Computação. I. Título.
CDU 004

Dados Internacionais de Catalogação na Publicação (CIP)
(Bibliotecária: Amanda Schuster – CRB 10/2517)

Henrique Damasceno Vianna (aluno)

Título: Pompilos: A Social Aware Model for Preventive Care of Non Communicable Diseases

Tese apresentada à Universidade do Vale do Rio dos Sinos – Unisinos, como requisito parcial para obtenção do título de Mestre/Doutor em Computação Aplicada.

Aprovado em 12 agosto 2019

BANCA EXAMINADORA

Prof. Cláudio Fernando Resin Geyer – UFRGS

Prof. José Palazzo Moreira de Oliveira – UFRGS

Prof. Kleinner Silva Farias de Oliveira – UNISINOS

Prof. Dr. Jorge Luis Victória Barbosa (Orientador)

Visto e permitida a impressão
São Leopoldo,

Prof. Dr. Rodrigo da Rosa Righi
Coordenador PPG em Computação Aplicada

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior Brasil (CAPES) - Código de Financiamento 001 /"This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001

To Grandma Alta (in memoriam).

ACKNOWLEDGEMENTS

Four years is a quite long time and it is hard to acknowledge every person that contributed to this work, so if you are not here, please forgive me. I will start with two key people that provided a lot of support for me to conclude this work. That two people are my wife Cristiane, who with a lot of patience understood my absences during that time and still always eager to help me be a better human, and my professor and friend Jorge Barbosa, who helped me conclude this work but also shaped me as a researcher.

I am deeply thankful to my friends Anelise, Bruno, Enedir, Lucas, Taís, and Suelen who helped me to forgive about the lots of work I had in delightful days at Rolante and Montenegro. Those days released me from the pressure I felt in regards to the deadlines and responsibilities during the conduction of this research. Thank you guys, I love you! =D

Gabriel and Sendy, thank you for the comprehension, patience, support, and meals. I will start to wash the dishes, I promise. ;)

A big thanks to my brother Igor, who is not just a friend and brother but who demonstrated being a great partner helping me with the English language reviews of some papers.

I am also very grateful to the Applied Computing Graduate Program Secretary staff, especially to Luciana Grimaldi Aquino, who always promptly and friendly helped me in my requests (and they were not few).

A very big thank you to all my professors, this work has a piece of your teachings.

This work was supported by Fapergs/Brazil (Foundation for the Supporting of Research in the State of Rio Grande do Sul - http://www.fapergs.rs.gov.br), CNPq/Brazil (National Council for Scientific and Technological Development - http://www.cnpq.br) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

“*El futuro es posible imaginarlo y no sólo aceptarlo*”.
(Eduardo Galeano)

“*La seule chose qu’on est sûr de ne pas réussir est celle qu’on ne tente pas*”.
(Paul-Emile Victor)

RESUMO

Segundo a Organização Mundial da Saúde, as doenças não transmissíveis foram responsáveis por 68% das mortes globais em 2012. Os cuidados com esse tipo de doença transcendem o envolvimento do paciente, estendendo-se à família, amigos e conhecidos que podem influenciar positivamente ou negativamente no tratamento. O apoio social pode ser entendido como a capacidade que as redes sociais têm para aliviar os efeitos nocivos do estresse e outros riscos à saúde. O apoio social no cuidado de doenças não transmissíveis é auxiliado pela computação, principalmente por meio de sistemas de informação, como fóruns de discussão, *blogs*, *wikis* e ferramentas de compartilhamento de vídeos, que visam aumentar a prevenção e o cuidado dessas doenças pelo paciente e suas famílias. No entanto, poucos estudos exploram a grande quantidade de dados gerados atualmente por *smartphones*, *wearables*, dispositivos inteligentes e mídias sociais. Tais dados, se devidamente explorados, podem servir como um meio para melhorar o apoio social na prevenção e tratamento de doenças não transmissíveis, recomendando novos contatos benéficos à saúde, ou apresentando a influência que os indivíduos exercem sobre a saúde dos outros como forma de conscientizá-los sobre seus hábitos e o impacto desses hábitos na vida de seus pares. Em outras palavras, tornar o apoio social auxiliado por computador ciente do contexto social. Esta tese apresenta o modelo Pompilos, que fornece cuidados preventivos através da consciência social para dar suporte ao cuidado de doenças não transmissíveis. Pompilos explora os dados gerados pelo uso de mídias sociais, smartphones, wearables e dispositivos inteligentes para inferir perfis de usuários e redes sociais. Perfis e redes sociais são utilizados para contabilizar a influência que os usuários exercem sobre os outros em aspectos relacionados ao cuidado e prevenção de doenças não transmissíveis, como consumo alimentar, atividade física ou tabagismo. Essa influência é usada para recomendar informações que possam melhorar o engajamento dos usuários na prevenção de doenças não transmissíveis, bem como para mostrar quais comportamentos influenciam a saúde de seus pares como um meio de conscientização social. Um protótipo do modelo Pompilos foi desenvolvido e testado em dois experimentos aleatórios por 61 usuários. Uma aplicação de assistente de saúde que implementou o modelo foi fornecida aos usuários que foram divididos em dois grupos de controle e intervenção. Esta aplicação teve características comuns de prevenção de doenças não transmissíveis como dieta, prática de atividade física, controle de peso e histórico. Para os usuários da intervenção, um recurso extra foi fornecido. Esse recurso extra permitiu que os usuários recebessem mensagens relacionadas à prevenção de doenças não transmissíveis coletadas de 15 perfis monitorados do Twitter. Os grupos de intervenção em ambos os experimentos aumentaram o uso dos históricos quando comparados ao grupo controle, indicando que eles estavam mais preocupados em acompanhar seus cuidados ao receber mensagens relativas à prevenção de doenças não transmissíveis. Por fim, dois perfis monitorados modificaram os comportamentos de postagem quando souberam do uso de suas mensagens para ajudar o usuário em seus cuidados.

Keywords: Doenças Crônicas Não Transmissíveis. Suporte Social. Sistemas Multi-Agentes. Modelo de Computação.

ABSTRACT

According to the World Health Organization, non-communicable diseases accounted for 68% of global deaths in 2012. Care for this type of diseases transcends patient engagement, extending to their family, friends and, acquaintances, who may influence their treatment positively or negatively. Social support can be understood as the ability that social networks have in alleviating the harmful effects of stress and other health risks. Social support in the care of non-communicable diseases is assisted by computing mainly through information systems such as discussion forums, blogs, wikis and video sharing tools, which aim for increasing patients' and families' prevention and care of these diseases. However, few studies explore the big amount of data that are currently generated by smartphones, wearables, smart devices, and social media. Such data, if properly explored, can serve as a means for improving social support in the prevention and care of non-communicable diseases by recommending new beneficial health contacts, or by presenting the influence that individuals make on the health of others as a way of aware them about their habits and the impact of these habits on the lives of their peers. In other words, make computer-aided social support computer aware of the social context. This thesis presents the Pompilos model, which provides social aware preventive care for non-communicable diseases. Pompilos explores the data generated by the use of social media, smartphones, wearables and smart devices to infer users' profiles and social networks. Profiles and social networks are used to account the influence that users exert on others in aspects related to the care and prevention of non-communicable diseases, such as food consumption, physical activity or smoking. This influence is used to recommend information that can improve users' engagement in the prevention of non-communicable diseases, as well as to show which behaviors influence the health of their peers as a mean of social awareness. A prototype of the Pompilos model was developed and tested in two random experiments by 61 users. A health assistant application which implemented the model was provided to the users who were divided into two groups control and intervention. This application had regular features of noncommunicable diseases prevention as diet, physical activity practice, weight management and, history charts. To the intervention users, an extra feature was provided. This extra feature allowed users to receive messages related to noncommunicable diseases prevention collected from 15 monitored Twitter profiles. The intervention groups on both experiments had improved the use of the history charts when compared to the control group, indicating that they were more concerned in following up their care when receiving messages relating to noncommunicable diseases prevention. Finally, two monitored profiles changed posting behaviors when aware of the use of their messages to aid user on their care.

Keywords: Non-communicable Diseases. Social Support. Multi-Agent Systems. Computing Model.

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

API	Application Programming Interface
FOAF	Friend of a friend
HTTP	Hypertext Transfer Protocol
JSON	JavaScript Object Notation
JSON-LD	JavaScript Object Notation for Linked Data
NCD	Non-Communicable Diseases
REST	Representational State Transfer
US	United States of America
WHO	World Health Organization

SUMMARY

1 INTRODUCTION

In 2014, the World Health Organization (WHO) defined the noncommunicable diseases (NCDs) as one of the greatest challenges of the twenty-first century health (WHO, 2014), due to the high death rate of those conditions. Only in 2012, the NCDs accounted for 68% of global deaths, and 40% of these deaths are considered premature. That is, deaths of individuals under 70 years old.

Most of chronic diseases are caused by habits such as sedentary lifestyle, smoking, among others, that result in “metabolic/physiological changes”, such as high blood pressure, overweight and obesity. Those habits and their results are know as risk factors, which must be controlled in order to prevent cases of those diseases (WHO, 2005).

The treatment of NCDs should be continuous, since most of these diseases does not have cure. Thus, the patient must be aware of his condition, follow the treatment determined by his physician and learn how to act when necessary. Even so, just the engagement of patients is not enough to cope with the challenges of their care. Sometimes, patients may not have the confidence to perform certain activities and need someone experienced to aid in their care (WAGNER et al., 2001; WAGNER; GROVE, 2002; BODENHEIMER; WAGNER; GRUMBACH, 2002). In this case, the participation of healthcare organizations, family, and community members in the assistance activities of these diseases is fundamental. These entities form the network of social relations of the patient (social network) (BARNES, 1954).

Social network has an important role in health as it regulates access to resources and opportunities to its members, as well as models their behavior, which may be of higher or lower risk. Hence, **social support** is the ability of that social network in alleviating the harmful effects caused by stress and other health risks through the provision of material, emotional and informational resources, and in the influence of behaviors such as eating, practicing physical activities, drug use and seeking medical follow-up (HOUSE; LANDIS; UMBERSON, 1988).

Researches about the influence of the social environment on health are not new, being already addressed by Emile Durkheim in the nineteenth century. Durkheim contributed significantly with his studies on the weight of the social effect on individuals’ morbidity. In his study on suicide, Durkheim analyzed particularly the influence that society has on the decision of an individual to commit suicide (DURKHEIM, 1897; BERKMAN et al., 2000). More recently, Christakis and Fowler used Framingham Heart Study (HISTORY OF THE FRAMINGHAM HEART STUDY, 2016) data to investigate the influence of social networking in individuals health. They found evidence of social network influence in weight gain (CHRISTAKIS; FOWLER, 2007), smoking cessation (CHRISTAKIS; FOWLER, 2008) and in the feeling of happiness (FOWLER; CHRISTAKIS, 2008).

Computing has been applied to support health care for decades and, its application to aid social support in NCDs has being done mainly by Internet driven platform such as social networks, forums, wikis, real-time chats, blogs and video channels (RICHARDSON et al., 2010; B et al., 2013; HOLTZ et al., 2014; WEYMANN et al., 2015; SMAHEL; ELAVSKY; MACHACKOVA, 2017). Most of these tools have an informational character, focusing on patients education to improve self-efficacy. Also, the exponential growth of data brought by popularization of social media gave opportunities to develop tools that analyze data shared among users to detect in advance diseases outbreaks (TUAROB et al., 2013; CULOTTA, 2014; LEE; AGRAWAL; CHOUDHARY, 2015; RAM et al., 2015; ABBAR; MEJOVA; WEBER, 2015), and so enabling health care organizations to create preventive strategies in order to diminish social and financial impacts related to these outbreaks. In addition, mobile applications are also being explored in order to provide social support in NCDs care due to their pervasive capabilities. That is, mobile devices, such as smartphones, are sensing platforms that acquire continuously users' data such as activities, health related variables (for example, heart rate), and locations. Furthermore, those devices are in major part of time with their owners, enabling continuous assistance (MARTIN et al., 2012; LAN et al., 2012; VIANNA; BARBOSA, 2014; ALSHURAFA et al., 2014a; SCHWARTZ et al., 2014; ALSHURAFA et al., 2016).

However, no computing model for NCDs care was found that integrates data generated everywhere like those from social media platforms, applications and personal activities[1] in order to, in a public level, improve preventive health policies and strategies, and in a personal level, patients awareness about implications of their behavior on health of others and to recommend resources that might favor health improvement by NCDs prevention.

1.1 Research Questions, Goals and Contributions

Social effects on health are well documented. Furthermore, the growth of social media platforms and mobile devices usage has brought an explosion of data which can be used to understand how individuals behaviors spread through social networks, and how these behaviors affect health. Such data can be used to promote health in different levels. In an individual level, people can receive contacts recommendations that might help them to tackle with health goals like, for example, engage in physical activities. In a collective level, health organizations may use these information to discover regions where a health behavior is common, and so, having means to spend health budget in actions that will have true effects.

For realizing this, three requirements might be achieved. Collecting patients' data is the first step. These data come from different sources, they can be collected from patients smartphone usage, for example, their location or type of physical activity (walking, running or still); social media interactions, for example, people who patients follow or information they share; and, information generated by the interaction with applications, for example, logging information about food consumed in diet diaries. Second, a great part of the acquired data, if not all, is raw. An approach is needed to summarize patients data to be used in later analysis, in other words, generate profiles (WAGNER; BARBOSA; BARBOSA, 2014). For example, it is well known

[1]For example, the act of purchasing an item may produce data related to users without their explicit interaction

that smoking habits, low physical activity, poor diet, high blood pressure, and total cholesterol are related to heart diseases, and these habits information can be used to summarize his coronary artery disease risk as low, medium or high (PITOLLI; VIANNA; BARBOSA, 2011). Third, it is necessary to understand how health behaviors spread in a social network. For this, it is possible to use social contagion models already defined.

Given the feasibility of the aforementioned requirements, this research focused on answer the following question:

- Could the reception of helpful health care information influence people's engagement in NCDs prevention and could the awareness of that influence by peers who disseminate that information, motivate them to keep the influence?

The hypothesis is that pervasive and social data can be used to infer the social influence that individuals apply to their peers when spreading health care information, and then aware the individuals about that influence in order to keep them spreading good health care information to improve the peers' engagement in preventive NCDs care. Hence, the main goal of this thesis is to provide a computing model that can ubiquitously detect social implications on individuals NCDs related factors (for example, social implications on weight gain or smoking behavior). A second goal is to use the model to supply social support in order to improve individuals engagement on preventive NCDs care. The following specific goals are presented in this work:

- An investigation of how influence is spreaded in social networks;
- An investigation of how computing aids social support on NCDs care;
- The design of a general computing model for aiding ubiquitous social care of NCDs;
- A prototype of the proposed model;
- The evaluation of the proposed model.

This work contributes in the area of health informatics by several ways. First it presents a systematic review of computer aided social support in NCDs care, which is available for on-line access in the Telematics and Informatics Journal (VIANNA; BARBOSA, 2017). It also shows a model that integrates social media and pervasive data to detect social influence as a mean to improve individuals engagement in preventive care of NCDs by the recommendation of beneficial connections and by the awareness of individual influence on health of others. Also, this research has already produced an ontology to detect the spreading of happiness, obesity ans smoking in social networks, based on the evidence described by Christakis and Fowler in (CHRISTAKIS; FOWLER, 2007, 2008; FOWLER; CHRISTAKIS, 2008), which was firstly published in the 42th Latin American Computing Conference (VIANNA et al., 2016) and extend to the International Journal of Metadata, Semantics and Ontologies (VIANNA et al., 2018). Finally, the general architectural model presented on this thesis was published on Information Processing Letters (VIANNA; BARBOSA, 2019).

1.2 Methodology

The aim of this thesis is to propose a computational model for aiding social support on non-communicable diseases care. In a first stage it was necessary to cope with wide scope of subjects from health, social science and informatics. Hence, a more flexible research approach was used and an exploratory research methodology had the required characteristics needed to cope with larger scopes (MALHOTRA; BIRKS, 2006). So, the first step of this research was to investigate in literature how information spreads in networks and how computing can aid social support on NCDs care.

After the conclusion of the literature review a computational model for social support on NCDs care was designed. A prototype of the designed model was developed, serving as a tool for hypothesis testing. A health care assistant application was built on top of the model. This application monitored Twitter profiles that produced health information. NCDs related messages were then forwarded to the application users. Two randomized experiments were done to check if recommended messages influenced users in keeping using application. Hence, the users of the application were divided in two groups: intervention and control. The intervention group received the recommended messages, while the control group did not. A rank showing users engagement in regards to each monitored profile was also built using the model. The monitored profiles were informed about the rank as a way to present them the influence they played on users health care behavior, and also to motivate them in producing more messages. Finally, acceptance of the application features was assessed by users. This assessment used the model proposed by Davis (DAVIS, 1989), enhanced by Yoon and Kim for acceptance of ubiqui- tous computing technology (YOON; KIM, 2007) is extensively used for technology acceptance (MARANGUNIC; GRANIC, 2015).

1.3 Thesis Structure

The thesis is organized as follows. Chapter 2 explains the foundations of this work, comprehending social influence in health and models for computing diffusion of influence or information in social networks. A detailed review about computer aided social support in NCDs care is given in chapter 3. Chapter 4 presents the Pompilos ontology, for detecting the spread of obesity, happiness and smoking cessation in social networks. The proposed model is then explained in chapter 5. Chapter 6 details the approaches for assessing the model. Finally, chapter 7 gives final remarks about this work.

2 SOCIAL NETWORKS, HEALTH AND INFLUENCE DIFFUSION

This chapter introduces the fundamentals of NCDs, Social Networks and Social Influence in Health subjects which are the foundation of the present work. A contextualization about NCDs, what are their causes and how risk factors prevention relates to them are presented in section 2.1. Section 2.2 defines social networks and explains concepts and properties related to network analysis. Finally, section 2.3 discusses the subject of social influence in health, its origins and exposes some works that could correlate social network influence in NCDs.

2.1 Non-Communicable Diseases and Risk Factors

Circulatory and respiratory diseases, cancers and diabetes, are considered the NCDs of highest incidence on the global level (WHO, 2005, 2008; CASTRO et al., 2008). The NCDs belong to the group of chronic conditions that involves a wider range of health problems, such as long-term mental disorders, permanent transmissible conditions and continuous physical disabilities (WHO, 2003). In general, these conditions have some common characteristics: they "require changes in lifestyle" and "health management for a long period of time", just as they are caused by lifestyles and risk behaviors, such as inappropriate diet, smoking, physical inactivity, alcohol abuse, high-risk sexual practices and social stress. In addition to behavioral factors, age is also a factor that can influence the incidence of some chronic conditions (WAGNER et al., 2001; CASTRO et al., 2008; WHO, 2003).

The management, prevention and control of NCDs can be made by preventing behavioral and biological risk factors related to these diseases. Inadequate diets, physical inactivity and smoking are major behavioral risks considered by WHO, while hypertension, high cholesterol and overweight are considered as the main factors of biological risks to be controlled (WHO, 2010) and according to WHO *"if the risk factors are eliminated, at least 80% of all heart disease, stroke and type 2 diabetes would be prevented; over 40% of cancers could be prevented"* (WHO, 2005).

2.2 Social Networks

One of the first quotes on the term social network was made by John Arundel Barnes in his article "Class and Committees in a Norwegian Island Parish" (BARNES, 1954). In this article the author presents his field study observations made at the parish of Bremen, Norway. By focusing their attention on relationships face to face the author arrived at the concept called social network. According to the author, a social network would be a representation of the relationships between people through a graph without limits. In this representation, people would be points and the lines indicate the interactions between people where each analyzed person is seen as the center of the network.

Table 1: Graph Basic Concepts

Concept	**Description**
Vertex (pl. vertices)	Network basic unit.
Edge	Connection between two vertices, represented graphically by a line
Direction	Graphs can be directed or undirected. The direction is usually represented by an arrow, which defines the direction of an edge connecting two vertices. Networks formed entirely by directed edges are called digraphs.
Degree	Number of vertices that link to the observed vertex.
Component	Set of vertices reachable by the observed vertex.
Geodesic Path	The smallest path between two vertices, measured in edges.
Diameter	Size of the largest geodesic path in the network.

Source: (NEWMAN, 2003)

A social network is type of natural network, i.e. networks that are not randomly generated. Other types of natural networks are information networks (e.g. article citation networks or the world wide web), networks of preferences (e.g. link people to things of their ownpreferences), technological networks (e.g. electric or Internet network) or Biological networks (e.g. neural network or chain feed) (NEWMAN, 2003). Social network, as all other networks, share graph basic concepts (see Table 1) and are formally expressed as $G = (V, E)$, where G is a graph formed by a set of vertices V and a set of edges E connecting these vertices.

Vertices represent different elements that connect, they can be a computer in a network, a web page or a person, and are also referenced as node. In the same way, edges (or links) represent different types of relationships, for example, a degree of kinship or the role in a computer network architecture. In turn, such connections carry weight, for example, the proximity of two individuals, or the communication latency between two computers. In an analytical perspective, connected nodes can be called egos or alters, so that *egos* are the nodes that are being analyzed and *alters* are the nodes connected to the *ego*, alters can be referenced also as neighbors (CHRISTAKIS; FOWLER, 2007, 2008; FOWLER; CHRISTAKIS, 2008)[1].

There is some properties that natural networks share, known as Small World Effect, Transitivity or Clustering, Degree distributions, Network resilience, Mixing patterns, Degree Correlations, Community Structure, Network Navigation and Betweenness Centrality (NEWMAN, 2003), which are explained in Table 2.

2.3 Social Influence in Health

The influence of social environment on health is not a new theme, being already addres- sed by Emile Durkheim in the nineteenth century. Durkheim contributed significantly with his

[1]By the rest of the term **node** will be used to denote a the **vertex being analyzed** and **neighbor** to denote its **first degree alters**

Table 2: Natural Network Properties

Property	**Description**
Small World Effect	Determines that the path between any two vertices of the network is small.
Transitivity or Clustering	States that there is a large number of connected vertices. For example, given three vertices **A**, **B** and **C**. **A** connects to **B** which connects to **C** which connects to A. Since **A** connects to **B** and **B** connects to **C** there is a **high probability that A and C will connect**.
Degree distributions	How network degrees are statistically distributed. For example, by Power Law distribution.
Network resilience	The behavior of the network in relation to the removal of vertices. For example, size of the paths between nodes of the network may increase, or nodes may become disconnected.
Mixing patterns	When vertices link according to their characteristics. For example, homophily in social networks, people with similar patterns tend to group together (e.g. smoking habits (CHRISTAKIS; FOWLER, 2008)).
Degree Correlations	When groups have high internal connection but low connection with other groups.
Community Structure	The ability of vertices to find the best path between peers, as observed in Small World Study (TRAVERS; MILGRAM, 1969).
Betweenness Centrality	Number of geodesic paths that pass through a node. It is also a measure of network resilience for indicating the effect caused by the removal of some node from the network.

Source: (NEWMAN, 2003)

studies on the weight of the social effect on individuals' morbidity (DURKHEIM, 1897; BERKMAN et al., 2000). In his study "suicide", Durkeheim analyzed particularly the influence that society has on the decision of an individual to commit suicide.

Social network has an important role in health as it regulates access to resources and opportunities to its members, as well as models their behavior, which may be of higher or lower risk. Hence, **social support** is the ability of that social network in alleviating the harmful effects caused by stress and other health risks through the provision of material, emotional and informational resources, and in the influence of behaviors such as eating, practicing physical activities, drug use and seeking medical follow-up (HOUSE; LANDIS; UMBERSON, 1988).

More recently, Nicholas Christakis and James Fowler investigated the spreading of NCDs and its risks through social network data collected from the Framingham Heart Study (CHRISTAKIS, 2004; CHRISTAKIS; FOWLER, 2007, 2008; FOWLER; CHRISTAKIS, 2008). Started in 1948, the Framingham Heart Study is a joint project of the National Heart, Lung and Blood Institute of the USA and Boston University, which aims to identify common features and factors that contribute to cardiovascular disease through long-term follow-up of participants who had not yet developed symptoms of cardiovascular disease (HISTORY OF THE FRAMINGHAM

HEART STUDY, 2016).

Originally, the project recruited 5,209 participants (men and women), returning to studies every two years. In 1971, a second generation of the study was done, involving wives and children of the original participants. In 2002 began the third generation of the study. As a result, the Framingham Heart Study has mapped the risk factors that may increase the incidence of cardiovascular disease.

According to Christakis and Fowler, social networks may influence the weight gain in individuals (CHRISTAKIS; FOWLER, 2007). The authors documented the use of social network with 12,067 people, to assess whether the weight gain of an individual may influence weight gain of others. For this, they used a regressive statistical model that allowed examine this influence among different social relationships (friends, siblings, spouses and neighbors).

According to the authors, the chance of becoming obese increases 57% if the individual has a friend who became obese; 40%, if one of the brothers became obese; 37%, if the spouse became obese.

Fowler and Christakis analyzed data from *Framingham Heart Study* to identify the spread of "happiness" through the social network of the participants. The feeling of happiness was measured according to the answers given by participants to CES-D questionnaires. The social network analysis identified that central nodes, i.e. those with more connections, are more likely to feel happy, and this feeling is transmitted to neighboring nodes (i.e. alters) until neighbors of neighbors (FOWLER; CHRISTAKIS, 2008).

Social networks may also influence individuals in their decision to quit smoking (CHRISTAKIS; FOWLER, 2008) as Christakis and Fowler report after analyzing data from the Framingham Heart Study project. The authors took into account contacts in the period from 1971 to 2003. Their results observed mainly that smokers tend to cluster in groups marginal to the social network, and that the act of quit smoking tends to be done in groups, suggesting that there is a partnership between smokers (homophily).

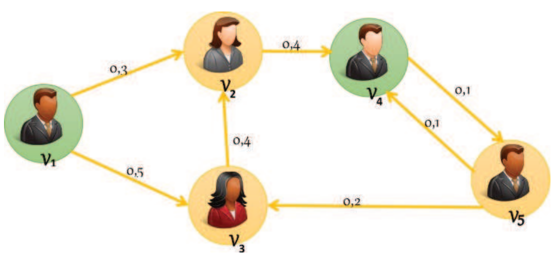
2.4 Influence Diffusion and Analysis on Social Networks

In social networks information and behaviors may spread from node to node. Moreover, there is a likelihood of a node "persuading" other nodes in transmitting information or behaviors. Thus, this chapter presents some models used for computing information diffusion in social networks, as also presents models used to compute the probability that a node has to influence another node. The influence detection model used in this thesis was inspired by the models for computing influence probability presented in this chapter.

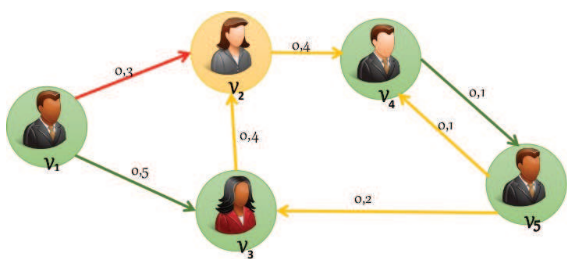
2.4.1 Independent Cascade and Linear Threshold Diffusion Models

Diffusion models are used to identify the way information is transmitted in a network, being applied in several areas such as marketing, health and social sciences. Among the models of information diffusion in social networks, the **Independent Cascade (IC)** and **Linear Threshold (LT)** models are the most commonly applied. In both models the information diffusion occurs by the activation of nodes in discrete steps, with the distinction that in the IC Model this activation is done in a single chance, whereas in the LT Model this activation is done according to the degree of influence that a certain node receives from its neighbors. This difference in functioning determines the type of application of the model, the IC being most suitable for viral diffusion, i.e. transmitted by the contact, while the LT best represents the change in behavior caused by constant exposure (CHEN; LAKSHMANAN; CASTILLO, 2013; KEMPE; KLEINBERG; TARDOS, 2003).

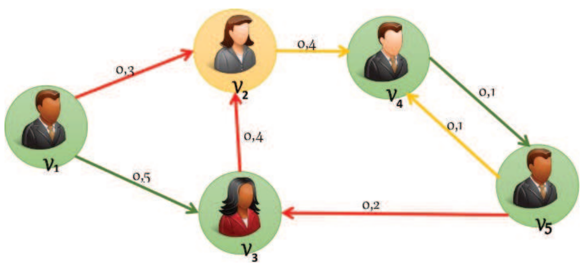
Figure 1: Independent Cascade Model Example



(a) Independent Cascade Model Example in *t=0*



(b) Independent Cascade Model Example in *t=1*

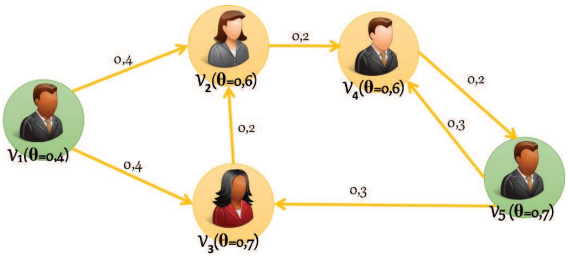


(c) Independent Cascade Model Example in *t=2*

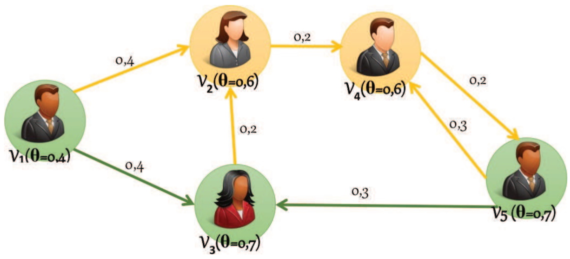
Source: Adapted from (CHEN; LAKSHMANAN; CASTILLO, 2013)

The network described in Figure 1 exemplifies the operation of the IC model. This network is composed of nodes v1, v2, v3, v4, and v5. Green circles represent activated nodes, yellow circles represent inactivated nodes. Yellow edges represent pending activation attempts, green represents success in the activation and red represents failures. Values above the edges indicate the probability required for the activation of the output node. Firstly, a set of active nodes that will initiate the diffusion process is selected, in case of example v1 and v4 (Figure 1a). In the next time (Figure 1b), each active edge will have a unique chance to activate each of its out-neighbors using a Bernoulli test (i.e., true or false, according to the probability of edge activation). Thus, when $t \geq 1$, v1 succeeds in successfully activating v3, but fails to activate v2, while v4 succeeds in activating v5. At the end of the process nodes v1, v3, v4 and v5 will be active, so that v3 and v4 have been activated by diffusion of influence (Figure 1c).

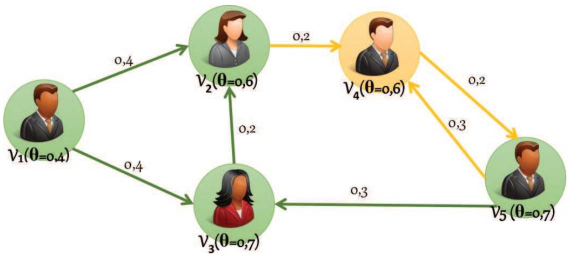
Figure 2: Linear Threshold Model Example



(a) Linear Threshold Model Example in *t=0*



(b) Linear Threshold Model Example in *t=1*



(c) Linear Threshold Model Example in *t=2*

Source: Adapted from (CHEN; LAKSHMANAN; CASTILLO, 2013)

The previous network was slightly modified to exemplify the LT model, being presented in the Figure 2. In each node was added the weight of its influence over its neighbors, denoted by the variable v, and the threshold required for its activation, defined by the variable θ, edges in green indicate the path for node activation. The activation of a node occurs when the sum of the incoming neighboring weights is greater than or equal to the defined threshold (i.e. $\sum_i w_i \geq \theta_v$). Thus at $t \geq 0$ (Figure 2a) nodes v1 and v5 are chosen to start the diffusion process. At $t \geq 1$ (Figure 2b) v3 will be active, since the sum of the weights of v1 and v5 is greater than or equal to the threshold set for v3. Finally, at $t \geq 2$ (Figure 2c) v2 will become activated due to the influence of neighbors v1 and v3. Hence, the process is terminated, since v2 and v5 do not have enough force to activate v4.

2.4.2 Christakis and Fowler Social Contagion Model

In *Social Contagion Theory: Examining Dynamic Social Networks and Human Behavior* (CHRISTAKIS; FOWLER, 2011), Christakis and Fowler explain the model in which they reached conclusions regarding the rule of three degrees of influence, related to the transmission of behaviors between nodes (FOWLER; CHRISTAKIS, 2008; CHRISTAKIS; FOWLER, 2008, 2007). The authors' approach first considers a permutation test to verify the validity of the hypothesis that some trait (for example, obesity or happiness) is diffused through the network, i.e., one node influences others to have its trait. In this test, the topology of the network is kept static and the nodes that are bound to a trait of interest (e.g. obese or not, smoker or not) are randomly distributed, while maintaining the prevalence of the trait, i.e., the number of nodes with a trait in relation to the population is the same of the original sample data.

The influence of a particular trait received by a node is evaluated through the use of a longitudinal regression model. The model considers two waves of observations, t and $t + 1$ where the relation between the current trait of an observed node (ego) is given by the observation of the same trait in the ego in a previous wave, and by observing alters traits in the previous and current waves. Thus, the regression function is given by the Equation 2.1, where β_n are regression parameters and ε is the error term.

$$Trait[ego, t+1] = \beta_1 Trait[ego, t] + \beta_2 Trait[alter, t+1] + \beta_3 Trait[alter, t] + \varepsilon \quad (2.1)$$

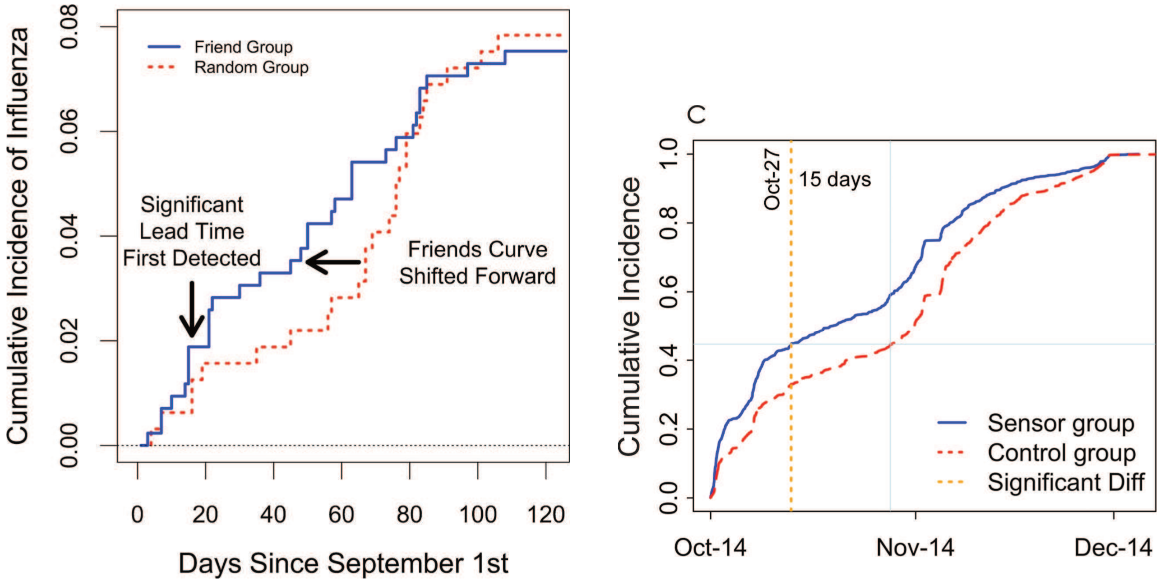
The model then expresses that the changing of trait in the observed node (ego) is given by the association (homophily) with alters with same trait at near moments. That is, by the observations in t and $t + 1$, assuming that the ego will be less susceptible to changes if association between nodes ends.

2.4.3 Social Network Sensors

Social Network Sensors is a model for early detection of outbreaks that uses nodes as sensors (GARCIA-HERRANZ et al., 2014; CHRISTAKIS; FOWLER, 2010). Nodes that have a higher number of connections (higher degree) and are more central, that is, integrate more paths to other nodes than the average, are more likely to be “infected” than those that are more peripheral. The selection of the sensor nodes is based on the so-called “friendship paradox”, which defines that neighboring nodes are usually more central and have a higher degree. Thus, given a set of randomly chosen vertices, those neighbors whose degree is greater than the mean of the randomly chosen population are selected as sensors.

The model was used in two experiments. The first monitored influenza spread at Harvard College from September to December 2009, where a group of selected students named their friends. The selected group and friends were monitored and the observations indicated an early influenza infection in the group of friends, as shown in Figure 3a (the solid line indicates infection in sensors and doted line in the random group, indicating an early infection in the sensor group). Already the second experiment used the “friendship paradox” to select a random sample of twitter users and their friends with more centrality to monitor outbreaks of *hashtags*. As in the first experiment, this experiment also showed an anticipation of hashtags use by the sensor group (Figure 3b).

Figure 3: Sensor Networks Cumulative Incidence



(a) Cumulative Incidence of Influenza (b) Cumulative Incidence of Hashtags

Source: (CHRISTAKIS; FOWLER, 2010; GARCIA-HERRANZ et al., 2014)

2.4.4 Topical Factor Graph for Influence Learning

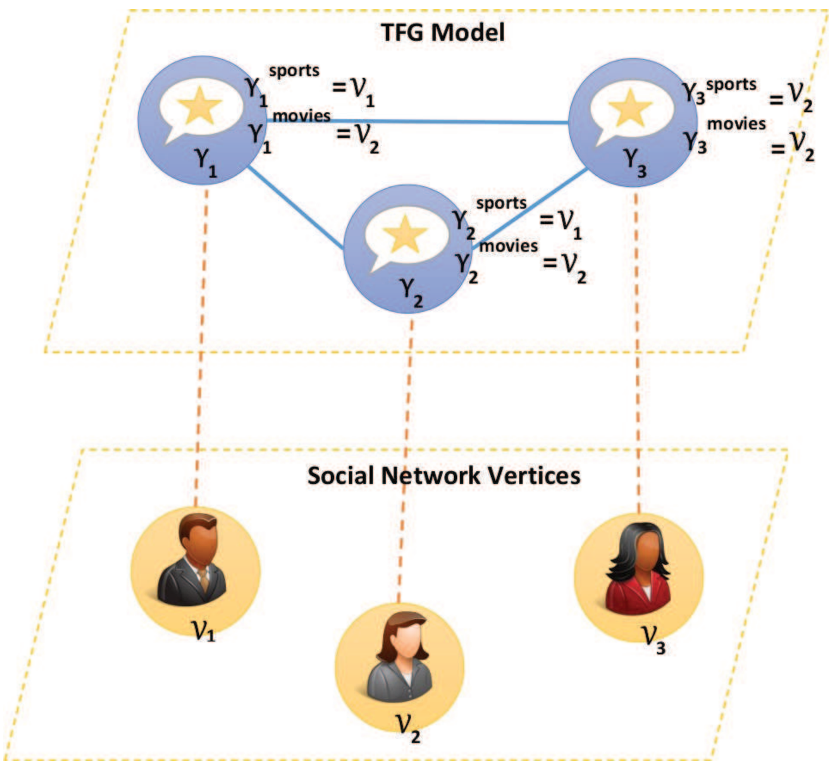
The Topical Factor Graph (TFG) model (TANG et al., 2009) understands that the influence received by a node on a particular topic (e.g., sports, gardening or travel) is given not only by the interest in the topic, but also by the influence received from the neighbors. In other words, the influence on a given topic is propagated by affinity in the network. Thus, given the nodes u and v, the model wants to find the weighted influence probability of node u in v for a given topic t (formally μ^t_{uv}). In order to compute such probability, the authors designed the TFG model, where each node v_i is related to a vector y_i, which stores the indexes of neighbors most likely to influence v_i in each topic t (or Y^t).$_i$The TFG is then used to learn, for each node, the neighbor that influences the most in each topic, by maximizing the result obtained in a probability function $P(v, Y)$, where $v = [v_1, ..., v_n]$ and $Y = [y_1, ..., y_n]$.

Figure 4 illustrates the TFG model. This example shows how the set of social network observed nodes, composed by $v1$, $v2$ and $v3$ spread influence in regards to topics $sports$ and $movies$. $v1$ is influenced by itself in topic $sports$ and by $v2$ in topic $movies$; $v2$ is influenced by $v1$ in topic $sports$ and by itself in topic $movies$; $v3$ is influenced by $v2$ in both topics. In this case, the set Y is composed by $\{\{v1, v2\}, \{v1, v2\}, \{v2, v2\}\}$.

2.4.5 Influence Probability Learning

The models of information diffusion described early do not define how to compute influence probabilities. Goyal, A. et al (GOYAL; BONCHI; LAKSHMANAN, 2010) present models to

Figure 4: TFG Model Topical Influence Example



Source: Adapted from (TANG et al., 2009)

compute the probability of influence between two nodes. That is, of a neighbor adopting the behavior of a given node. The basis of the models lies in the use of sets that store the actions performed by each node, such as sharing a link. The action by a node serves as a trigger for adopting the behavior by other nodes, so that the probability of taking action increases incrementally, i.e. monotonically, by the other neighboring nodes that did not take the action. A propagation graph denotes the diffusion of an action by neighbors. Goyal, A. et al demonstrated three statics methods to compute influence: Bernoulli distribution, Jaccard index and partial credits.

A Bernoulli Distribution defines the probability that a node has to activate an action on a neighbor, through a binomial attempt (e.g., throwing a coin). Thus, this estimate is given by the ratio between successful neighbor activation attempts (A_{v2u} in Equation 2.2) and the total number of actions performed by the node (A_v in Equation 2.2).

$$p_{u,v} = A_{v2u}/A_v \tag{2.2}$$

Goyal, A. et al present an adaptation of the Jaccard coefficient of similarity (Equation 2.3). In the presented version, the probability of a node activate a neighbor is given by the ratio between the activation attempts in the neighbor (A_{v2u} in Equation 2.3) and the union of the total actions performed by the node and the neighbor minus its intersection ($A_{u|v}$ in Equation 2.3).

$$p_{u,v} = A_{v2u}/A_{u|v} \tag{2.3}$$

Partial Credits takes into account that the execution of an action by a node was influenced by the neighbors, so the credit is given in relation to the occurrence of the action in previous moments. The probability of credit can be applied either with the use of Bernoulli (Equation 2.4) or Jaccard index (Equation 2.5).

$$p_{v,u} = (\sum_{a \in A} credit)/A_v \tag{2.4}$$

$$p_{v,u} = (\sum_{a \in A} credit)/A_{u|v} \tag{2.5}$$

The three models presented are static, that is, the probability does not change as a function of time, which is not real since the influence decreases as a function of time (GOYAL; BONCHI; LAKSHMANAN, 2010). These particularities are approached by the authors through a continuous model and a discrete one. The two models use the afore mentioned static probabilities equations, with the addition that in the continuous model the influence of a user will decrease as a function of time, and in the discrete model it will be valid for a period (by three iterations, for example). The continuous model is computationally more intense, since the probabilities must be recomputed with each increment of time. On the other hand, the discrete model can approximate the results of the continuous model, with a performance similar to the use of static models.

3 COMPUTER AIDED SOCIAL SUPPORT IN NCDS

As way to investigate how computing can aid social support on preventive NCDs care, this thesis uses systematic mapping study as methodology for its literature review[1] (BUDGEN et al., 2008; PETERSEN; VAKKALANKA; KUZNIARZ, 2015; COOPER, 2016). Systematic mapping study, as systematic literature review (SLR), are types of systematic review. Even though systematic reviews are not frequently used in computing, they are widely recognized and applied in other areas such as medicine (COOPER, 2016) and social sciences (PETTICREW; ROBERTS, 2006). In general, the main purpose of reviews is identifying the existent evidence and trends in collections of literary works related to a set of topics of interest as way of reducing the bias present when single references are used.

Based on the guidelines proposed by Petersen et al. (PETERSEN; VAKKALANKA; KUZNIARZ, 2015), this thesis will use the following steps for its systematic mapping study:

1. Define the research questions;
2. Define the search process;
3. Define the criteria for text selection;
4. Execute the analysis and classification of the selected texts.

3.1 Research Questions

Based on evidence reporting the influence of social relationships in health, this thesis wants to investigate the following questions:

- How computing can promote social support in NCDs care? (Q1);
- How computing can use social data to support NCDs care? (Q2);
- Is there any computing model for social support promotion in NCDs care? (Q3).

The study guided by these might help to map the current state of computer aided social support in NCDs care, its trends, gaps and how it is being addressed by computer applications.

3.2 Search Process

Before starting the search, the three research questions must be transformed in Boolean expressions to be used in search databases. For this, the research questions are broken into different question parts. Then, a set of keywords are related to each question part. These keywords

[1]The findings of this chapter are available online in the Journal of Telematics and Informatics (VIANNA; BARBOSA, 2017)

Table 3: Research questions to boolean expressions mapping

Research Question	Question Part	Boolean Disjunction	Boolean Expression
Q1 and Q2	**"How computing can promote (...)" or "How computing can use (...)"**	recommendation ∨ recommender systems ∨ system ∨ application ∨ computer aided	(recommendation ∨ recommender systems ∨ system ∨ application ∨ computer aided) ∧ (social data ∨ social network ∨ social support) ∧ (noncommunicable diseases ∨ chronic diseases ∨ risk factors ∨ chronic conditions)
	"(...) social support (...)" or "(...) social data to support (...)"	social data ∨ social network ∨ social support	
	"(...) in NCD care?" or "(...)NCD care?"	noncommunicable diseases ∨ chronic diseases ∨ risk factors ∨ chronic conditions	
Q3	**"Is there any computing model for promoting (...)"**	computational model ∨ informatics model ∨ framework ∨ architecture	(computational model ∨ informatics model ∨ framework ∨ architecture) ∧ (social data ∨ social network ∨ social support) ∧ (noncommunicable diseases ∨ chronic diseases ∨ risk factors ∨ chronic conditions)
	"(...) social support (...)"	social data ∨ social network ∨ social support	
	"(...) in NCD care?"	noncommunicable diseases ∨ chronic diseases ∨ risk factors ∨ chronic conditions	

Source: Own authorship

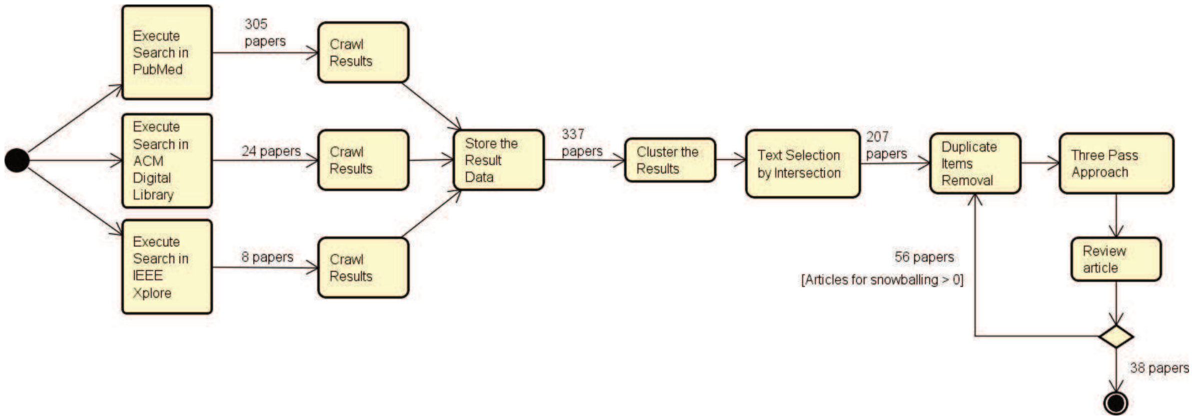
form a Boolean disjunction which represents part of the Boolean expression representing the research question. Finally, the Boolean expression is assembled by the conjunction of Boolean disjunction parts, as shown in Table 3.

Three scientific literature databases were chosen to run the mapped research questions. These were ACM Digital Library[2], IEEE Xplore[3] and PubMed[4]. The first two are recogni- zed digital libraries of computer science documents. The latter is a database of medical and life sciences literature that contains more than 26 million references.

Once the databases were defined, the two Boolean expressions were transformed in six queries, according to query format supported by each database. In addition to the query, the searches in the databases were set to return documents from 2012 to 2016. Documents before 2012 were mapped by following relevant citations from the original results (i.e. by snowballing) (SNOWBALLING, 2016).

[2] http://dl.acm.org/
[3] http://ieeexplore.ieee.org/Xplore/home.jsp
[4] http://www.ncbi.nlm.nih.gov/pubmed

Figure 5: Text Selection Process



Source: Own authorship

3.3 Text Selection

Figure 5 depicts the text selection process used. After querying each database, an automated process has been executed in order to **crawl results** from these databases. The crawling process then **stored the result data** in a XML document. The stored data consisted of the title, abstract and URI of each search result. The XML documents were later inputed in Carrot2 Workbench[5] in order to **cluster the results**, i.e., group similar results (HARTIGAN, 1975), with aid of **Suffix Tree Clustering** (STC) (ZAMIR; ETZIONI, 1998). This method groups similar documents based on phrases shared among them. Table 4 lists the resulting **shared phrases** for each **research question**, and also the **number of documents** sharing the same phrases.

The shared phrases were used as a criteria for text selection in a way that the selected documents have phrases closer to the semantics of the research questions. Consequently, the texts for both research questions were selected by intersecting the documents belonging to a set of shared phrases (Figure 5, text selection by intersection process). Table **??** shows the number of documents according to the intersection of shared phrases. The set of shared phrases was chosen due its relevance with the research questions, therefore phrases that are not related with the keywords used in the text selection queries were excluded.

Following, a filter was made to select only those documents that are more likely to answer the research questions. For this, the first pass of the three pass approach (KESHAV, 2007) was applied in the documents resulting from the intersection of shared phrases. Thus, for each document the following actions were executed:

- Read the title, abstract, introduction, section, sub-section headings and conclusions;
- Identify if the context of the paper is related with the research questions;

[5]http://project.carrot2.org

Table 4: Documents shared phrases

Research question	Shared phrases	Number of documents
Q1 and Q2	Study, Health, Social Support	239
	Disease, Chronic	120
	Risk Factor Surveillance System, Behavioral Risk Factor Surveillance	16
	Mental Health	58
	Associated	107
	Patients	97
	Identified	96
	Age	93
	Assessment	88
	Logistic Regression	33
	Social Network	31
	Effects	83
	Significant	83
	Stress Disorder PTSD, Posttraumatic Stress Disorder	14
	Results	82
Q3	Study, Health, Social Support	95
	Cardiovascular Disease CVD, CVD Risk Factors	8
	Chronic Conditions	23
	Social Network	20
	Mental Health	16
	Stress Disorder PTSD	8
	Effects	38
	Adults, Older	24
	Identified	37
	Model	37
	Participants	37
	Research	36
	Important	32
	Interventions	32
	Chronic	30

Source: Own authorship

- Accept or reject the paper for further review.

As result 19 papers were selected aiming to answer the questions Q1 and Q2, and 2 papers were found suitable to answer the question Q3.

A snowballing process was executed after the review of these 21 papers. Hence, another 56 papers were extracted from those initial 21 papers' references. Again, the three pass approach was applied in the 56 papers which resulted in more 17 papers suitable for review. In the end, the text selection process resulted in 38 papers proper for review.

Table 5: Number of intersected documents according with shared phrases

Research question	Intersection of shared phrases	Number of documents	Selected documents
Q1 and **Q2**	(Study, Health, Social Support ∪ Social Network) ∩ (Disease, Chronic ∪ Risk Factor Surveillance System, Behavioral Risk Factor Surveillance)	128	19
	Study, Health, Social Support ∩ Social Network	7	0
Q3	(Study, Health, Social Support ∪ Social Network) ∩ (Cardiovascular Disease CVD, CVD Risk Factors ∪ Chronic Conditions ∪ Chronic ∪ Model)	60	2
	Study, Health, Social Support ∩ Social Network	12	0

Source: Own authorship

3.4 Analysis and Classification

After the reviewing process each paper was categorized as **controlled trials**, **frameworks and systems**, **knowledge discovery**, **social media usage analysis** or **simulation models**, according to its perceived characteristics.

Appendix 28 abstracts the reviewed papers and shows their title, where they were published, type of publication (e.g. journal, conference or chapter), year of publication and how they were classified. Appendix 28 also indicates if papers **promote social support**, **use social data** or present a **computing model for social support promotion** as asked in questions Q1, Q2 and Q3.

The following sections will briefly discuss each category and also statistics regarding this study.

3.4.1 Controlled Trials

The papers classified as Controlled Trials present controlled experiments that assess the effects and outcomes of using social support interventions in conjunction with the use of information technology tools in the care of NCDs.

Due to population's low physical activity and unhealthy eating habits, obesity and overweight are common NCDs risk factors (WHO, 2005, 2008, 2003). Holtz et al. (HOLTZ et al., 2014) conducted a randomized cross-over study on American veterans which used the online social support tool SparkPeople for helping them to control their body mass index. The online tool allowed the participants to exchange experiences related to the management of their physical condition with other SparkPeople users.

In turn, Spring et al. (B et al., 2013) presented the results achieved in the randomized trial where a intervention group used a smartphone application as a tool to support weight reduction.

The intervention group received training calls every two weeks, and could participate in group sessions led by nutritionists, psychologists, and physicians. Compared with other participants, those who joined the group sessions and used the support application performed considerably better than those who only used the application, or who only joined the group sessions.

Richardson et al. (RICHARDSON et al., 2010) used a walk program available on the internet in a randomized trial to assess whether the intervention group (those who had access to the online walking program community) would have a greater increase in average daily walking. The use of the online community did not present an increase in the mean of walking in relation to the control group. However, the participants in the intervention group had greater engagement than the control group. In addition, participants reporting lower social support at baseline used the online community functionality often.

Depression in people with diabetes is twice as high when compared to people who do not have diabetes (BOND et al., 2010). As a way to understand if internet use would have positive effects in the treatment of diabetes, Bond et al. conducted a randomized six-month trial, where the intervention group had online services to support their care. The authors hypothesized that the intervention group would be less depressed and had gains in social support, quality of life and self-efficacy. A multivariate covariance analysis in the resulting data confirmed the hypothesis.

Finally, Weymann et al. (WEYMANN et al., 2015) report a trial to assess the effectiveness of their web interactive health communication application (IHCA) by patients with type 2 diabetes or chronic low back pain. Patients in the intervention group, which used the IHCA, had gains in disease knowledge and perceived emotional well-being.

3.4.2 Frameworks and Systems

The papers classified as frameworks and systems include those that present computational systems or models of computational systems integrating social support features or social data in the care of NCDs.

Accessible Telehealth (DHILLON; WüNSCHE; LUTTEROTH, 2013) and Virtual Aged Care System (ROBERTSON et al., 2014) focus on elderly care. Accessible Telehealth defines a conceptual framework which integrates the features of different types of health care systems (telehealth, health record, health portals and serious games). Yet, Virtual Aged Care System proposes the integration of health care service infrastructure and demographic and geolocation data for optimizing the care of elderly populations of Australia.

Patient Journey Record System (PaJR) framework (MARTIN et al., 2011, 2012) was designed as a metaphor that understands patients similar to travelers. Patients, just as travelers, will need guidance in some point of their journeys to cope with the challenges of the treatment. PaJR analyses patients' answers given in structured questionnaires to detect in advance the need for health intervention.

Schwartz et al. present a collaborative approach for helping patients that are in anaphylaxis emergency situations through a smartphone mobile application. With the application patients can send emergency notifications that are received by nearby collaborators (SCHWARTZ et al., 2014).

In turn, WANDA (LAN et al., 2012; ALSHURAFA et al., 2014a, 2016; SIDERIS et al., 2015; ALSHURAFA et al., 2014b) improves the care of heart failure patients with the use of non-invasive mobile platform. Data acquired in patients' smartphones are used to prospect knowledge and detect health variations in real-time. Social features embedded in the platform give patients means to collaborate with others to improve their care.

Finally, in Health Care 2020 (MILANI; LAVIE, 2016), Milani, and Lavie propose a new model of care to face the health challenges of the century. In their model, information and communication technologies play important roles, as they might enable continuous monitoring, real-time information access, communication between health professionals and provide social support to patients.

3.4.3 Knowledge Discovery

Today each person is capable of generate big quantities of data while shopping, using Internet and social media or simply by possessing a smartphone. All these data can be used to discover knowledge and to predict collectively or individually health trends. Thus, papers classified as Knowledge Discovery are those that address the usage of data from health records, social media, social networks or smartphones to predict, discover or reveal hidden information.

Khan et al. showed their approach to use social network analysis to predict the likelihood of a patient developing a non-communicable disease comorbidity (KHAN; UDDIN; SRINIVASAN, 2016).

Twitter[6] users generate thousands of messages each minute and many of these messages represent users' feelings, health or emotional statuses. Furthermore, the possibility of embedding users' location in Twitter messages enables not only the prediction of trends, but where these trends will be located. Paul and Dredze used ailment topic aspect model to identify diseases that are not explicit expressed in Twitter messages (PAUL; DREDZE, 2011). Similarly, Tuarob et al. used an ensemble method to classify health related messages (TUAROB et al., 2013). Twitter messages could also be used to identify what Twitter users eat, and by doing that predict the caloric levels and incidence of obesity and diabetes in a individual level, as described by Abbar et al. (ABBAR; MEJOVA; WEBER, 2015). Culotta addresses how language expressed through Twitter messages and demographic data can predict health statistics at the geographic level of American counties using regression models (CULOTTA, 2014).

Ram et al. and Zhang et al. (RAM et al., 2015; ZHANG et al., 2016) could predict the average of hospital visits related to asthma in 28 hospitals from the Dallas/Fort Worth area

[6]http://twitter.com

using Twitter messages, hospital visits data and US adult asthma prevalence data. In turn, Lee et al. (LEE; AGRAWAL; CHOUDHARY, 2015) analyzed Twitter messages to predict time and location of allergy incidence in the US. Finally, Weber and Mejova used a crowd sourcing services to annotate Twitter users profile picture as overweight or not (WEBER; MEJOVA, 2016). However, statistical analysis with US obesity data did not correlate to all regions.

Other social media can be used to identify health statuses, as Balani and De Choudhury report. They used reddit[7] forums messages as input to classify self-disclosure (BALANI; DE CHOUDHURY, 2015). According to the authors, messages with high self-disclosure in mental health communities have derogatory content, and the classification of such messages can be used by moderators or community to identify users who need support.

As part of the WANDA project, Sideris et al. present a model for predicting patients' care adherence after receiving automated social support interventions (SIDERIS et al., 2015). As an effect, the authors could detect those patients who did not initially adhere to the treatment and started to adhere after receiving the intervention. Yet in WANDA project, Alshurafa et al. used baseline social support survey answers to predict body mass reduction (ALSHURAFA et al., 2014a) and adherence to physical activity (ALSHURAFA et al., 2014b).

Using free tools available in the Internet D'Ambrosio et al. have built a cost effective system for surveillance the consistency of the information about preconceptions searched and shared in the Internet (D'AMBROSIO et al., 2015). Using the system the authors observed a lack of interest for chronic diseases genetic risk assessment.

3.4.4 Simulation Models

Simulation models may be used when is difficult or time consuming to reproduce the behaviors or mechanics of a system. Hence, two reviewed paper used simulation models in order to understand the relationships between social activities and the spread of NCDs (AZIZA et al., 2016; CHIêM; MACQ; SPEYBROECK, 2012).

SimNCD is an agent-based simulation model created by Aziza et al. (AZIZA et al., 2016) where activities and individuals are modeled as agents. Individuals engage in their goals through interaction with activities. The interaction between individual and activity results in an update of the internal status of the agent, which may be the reduction of the body mass index of the simulated agent. The proposed model was used to simulate childhood obesity among children between 6 and 18 years of age. Three scenarios were created using data from interventions already performed, and the results of the simulations were very close to the results found in the interventions.

Both social support and events occurring during the life of an individual are elements associated with depression in the elderly. Chiêm et al. investigate this link through rule-based simulations. The model was designed to simulate how life course events of individuals can

[7]http://www.reddit.com

impact in their received social support, and how this received social support can impact on individual's depression status. Specialists were consulted to assess the importance of the types of contact in the life of subjects and how much such contacts influence their lives. Four scenarios containing different demographic and social parameters were simulated. The results presented a good approximation to the existing trends in the empirical data and the authors believe that the proposed framework is a good alternative for exploring the relations between health and social aspects (CHIêM; MACQ; SPEYBROECK, 2012).

3.4.5 Social Media Usage Analysis

As verified by Gomez-Galvez et al. (GOMEZ-GALVEZ; MEJíAS; FERNANDEZ-LUQUE, 2015) in their review of the use of social media in diabetes care, integration with social media is done through the analysis of data from social networks and/or the use of the information sharing functionality of these media to support the care or to motivate users in following their treatments. Online communities are part of social media, they offer tools such as forums, wikis, realtime chats, blogs and libraries that serve to exchange experiences among patients, caregivers and physicians. They also help to disseminate knowledge effectively and improve social and psychological support to patients and families by increasing their self-efficacy (EIJK et al., 2013).

Social media may even improve the care of subjective matters like, for example, pain, which is an idea that cannot be transmitted through language. Becker (BECKER, 2012) presents that messages containing characteristics of validation and emotional support were sent the most by the 18 participants diagnosed with chronic pain who used an online platform for six weeks.

Libin et al. discuss the role of "how-to"videos available in social media as a tool to support and educate spinal cord injury (SCI) users. Therapists have rated the usefulness of 10 videos created to support SCI. The authors evaluated the categories of SCI-related videos on the platform and how users interacted and consumed the videos. Videos that demonstrated basic skills were the most accessed (LIBIN et al., 2014).

Family history has a strong influence on the aggravation of risks of some diseases, however filling a record of health information of another family member in a wrong way could lead to misdiagnoses. Welch et al. (WELCH et al., 2015) carried out a study to evaluate the acceptance of use of a tool to record family health history using social network metaphor. The result of the research revealed a good acceptance of the proposed approach.

There is a reasonable number of weight loss programs available on the Internet. DailyBurn[8], SparkPeople[9] and Nutracheck[10] are some examples. Ba and Wang (BA; WANG, 2013) have identified that paying members who have larges social network or more posts have a higher probability of engagement on the DailyBurn platform. Encouragement, motivation and ex-

[8]https://www.dailyburn.com/
[9]http://www.sparkpeople.com/
[10]http://www.nutracheck.co.uk/

change of experiences are the main types of social support perceived by users of SparkPeople as verified by Hwag et al. (HWANG et al., 2010, 2014). Johnson and Wardle (JOHNSON; WARDLE, 2011) have identified that postings in online forums were significant predictors of weight loss among women and that dietary diaries are a predictor of weight loss for both male and female Nutracheck users.

3.5 Results

This section demonstrates how the research questions are accomplished by the reviewed papers and also presents some statistics regarding this study.

3.5.1 How computing can promote social support in NCDs care? (Q1)

Internet driven platforms play a significantly role in promoting social support in NCDs care. Social networks, forums, wikis, realtime chats, blogs and video sharing platforms offer means to patients, caregivers and physicians share experiences which increases patients' confidence and self-efficacy.

Moreover, the increasing interest in machine learning (TRENDS, 2016) has also brought contributions to the promotion of social support in NCDs care, as seen in Martin et al. (2011, 2012) which analyzed patients' answers to questionnaires to detect the need of health intervention in advance. Mobile applications may also promote social support by fomenting cooperation among users or enabling users to receive notification in order to change their behaviors, as shown in Lan et al. (2012); Schwartz et al. (2014); Alshurafa et al. (2014); Sideris et al. (2015).

3.5.2 How computing can use social data to support NCDs care? (Q2)

The use of social data for detection of health related information and trends has become standard. Authors used data from Twitter to discover geographically located trends in diseases (PAUL; DREDZE, 2011; CULOTTA, 2014; LEE; AGRAWAL; CHOUDHARY, 2015; WEBER; MEJOVA, 2016), NCDs risk factors (ABBAR; MEJOVA; WEBER, 2015) or average asthma related hospital visits (RAM et al., 2015; ZHANG et al., 2016). Data from other social media can also be used to discover information. For example, reddit forum messages can be used to detect self-disclosure discourses (BALANI; DE CHOUDHURY, 2015).

Users' location data from smartphones may also be used to support cooperation in emergency situations (SCHWARTZ et al., 2014). Some works used social support group questionnaires for means of predicting treatment adherence and success (ALSHURAFA et al., 2014a, 2016, 2014b; SIDERIS et al., 2015). Finally, census information and resources location data were used to improve the care of Australia's elderly populations (ROBERTSON et al., 2014).

3.5.3 Is there any computing model for social support promotion in NCDs care? (Q3)

Few papers presented computing models for social support promotion. In essence, there are two main models, WANDA, which is shared by several papers (LAN et al., 2012; ALSHURAFA et al., 2014a, 2016; SIDERIS et al., 2015; ALSHURAFA et al., 2014b) and PaJR (MARTIN et al., 2011, 2012). Both models propose a data analytics module which uses data acquire from patients for predicting health variations, and so enabling the execution of health interventions on patients.

Although Health Care 2020 (MILANI; LAVIE, 2016) was not created as a computational model, it recognizes the need of computational aid as means for social support and beneficial relationships promotion.

3.5.4 Review Statistics

Most of the reviewed papers were from distinct publications titles (Figure 6a). However, Annual ACM Conference on Human Factors in Computing Systems and Journal of Medical Internet Research contrast. The first encompasses Controlled Trials and Social Media Usage Analysis papers, yet the latter has a focus on Knowledge Discovery (Figure 6b). Those publications could be a starting point for researching on these topics.

The even distribution of papers from journals and conferences (Figure 7a) was not forced and may reflect the text selection process. Most papers from the first phase of text selection, when the selected texts were resulting from search in repositories, were from conferences. Nevertheless, this inclination has changed in the second phase of text selection when the snowballing process was executed (Table 6). Possibly, this is due that major part of papers indexed in repositories are conference papers, but majority part of cited papers are from journals. However, this hypothesis was not tested and deserves further investigation.

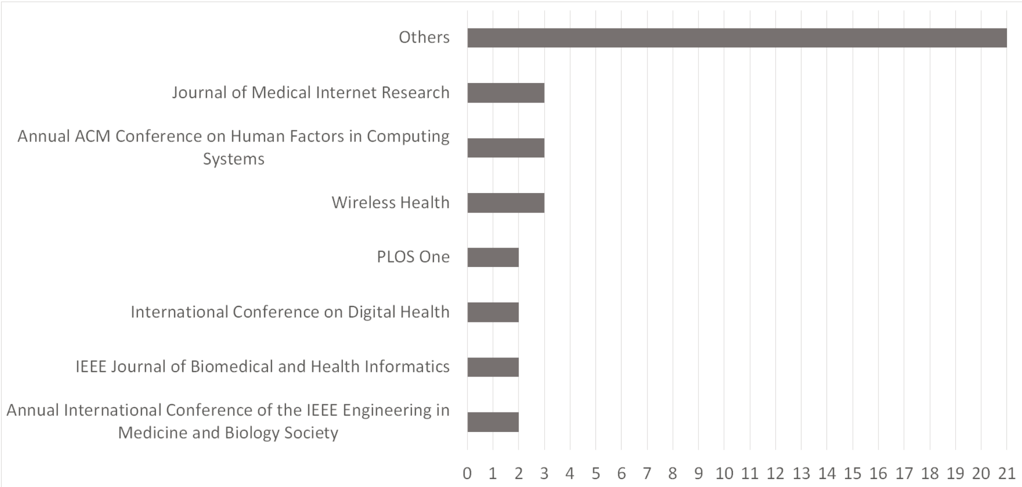
Table 6: Publication Type by Text Selection Phase

Phase	Number of Conference Papers	Number of Journal Papers
Repository Search	12 (63.16%)	7 (36.84%)
Snowballing	6 (37.50%)	10 (62.5%)

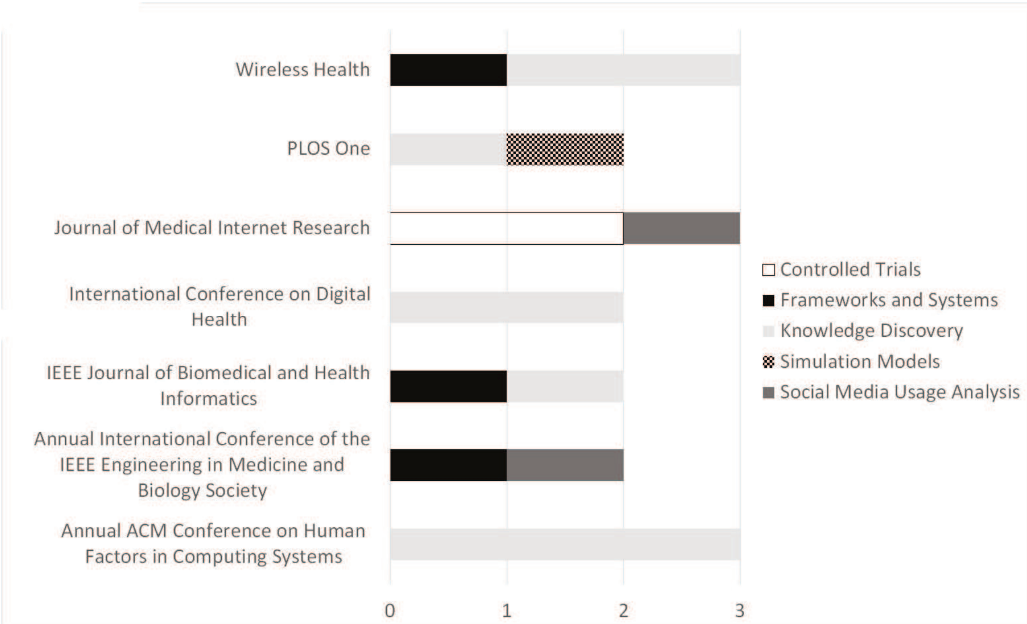
Source: Own authorship

The majority of reviewed papers were related to Knowledge Discovery, followed by Social Media Usage Analysis and Frameworks and Models (Figure 7b). It seems that the interest about Knowledge Discovery is growing through the years (Figure 8), particularly since 2013, possibly due a increased interest in related topics such as big data and machine learning (TRENDS, 2016). It is important to note that the data from 2016 do not reflect the entire bibliographical production of that year, since the preliminary digital libraries search were executed in august 2016.

Figure 6: Papers statistics by source



(a) Total of papers by publication source



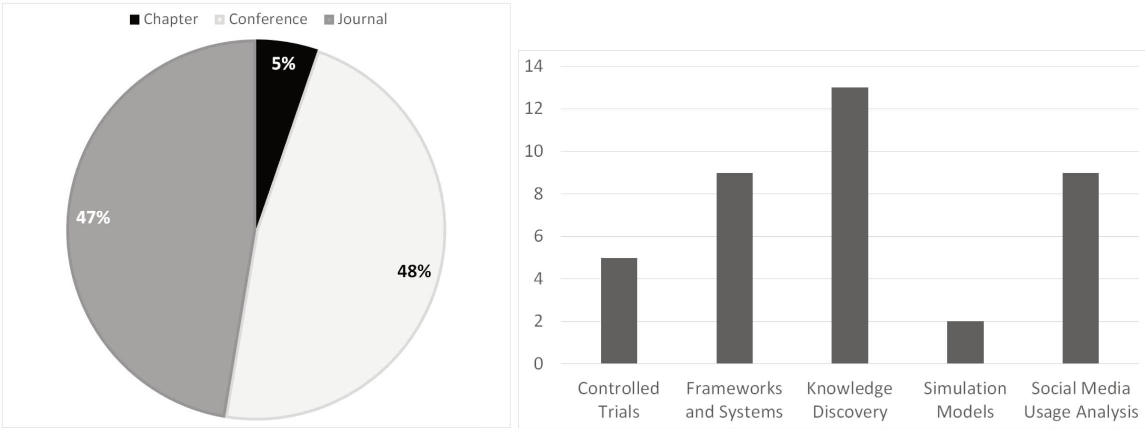
(b) Total of papers classification by publication source

Source: Own authorship

3.5.5 Related Works

This chapter presented a review of 38 works focused on social support for NCDs care. From these 38 works, 7 belong to two larger projects (WANDA (LAN et al., 2012; ALSHURAFA et al., 2014a,b; SIDERIS et al., 2015; ALSHURAFA et al., 2016)) and Patient Journey Record System (MARTIN et al., 2011, 2012)) which present models for social support promotion in NCDs care as asked by the research question *“Is there any computing model for social support*

Figure 7: Papers statistics by type and classification



(a) Total of papers by type of publication (b) Total of papers by classification

Source: Own authorship

promotion in NCDs care? (Q3)" (section 3.5.3). Other two works are related to systems that aid social support in NCDs care (DHILLON; WüNSCHE; LUTTEROTH, 2013; SCHWARTZ et al., 2014), and one other enables information discovery of NCDs related conditions in an individual level by the use of social media data (ABBAR; MEJOVA; WEBER, 2015; WEBER; MEJOVA, 2016). These works were selected due to their characteristics of being systems focused on improving social support on individual level or are techniques for identification of social awareness with social media data.

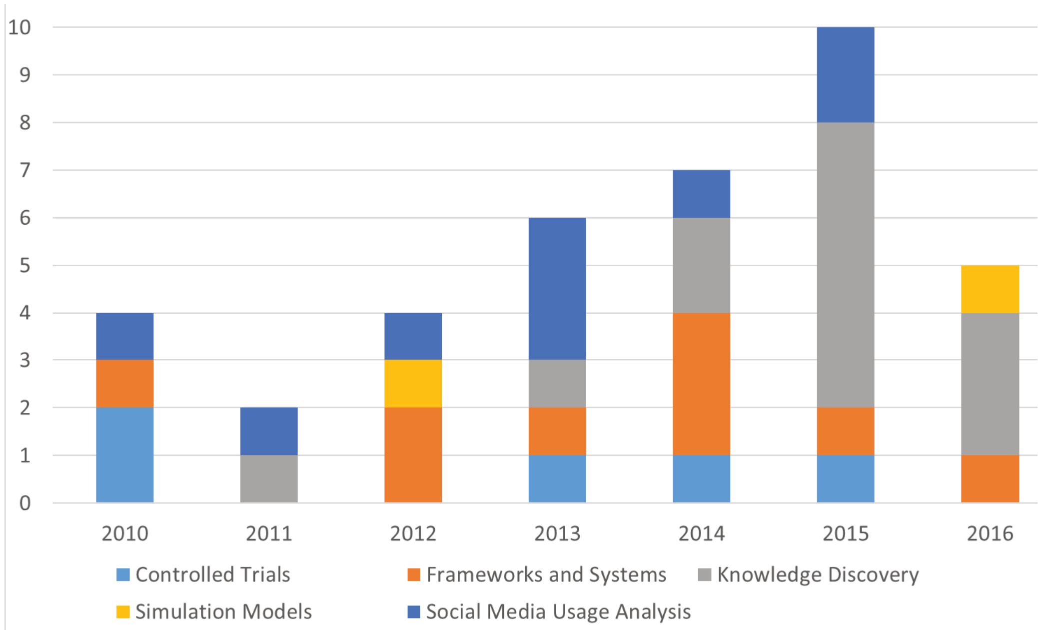
This section shows a closer look on these works and compares them in terms of pervasiveness, adaptability with patients behavior, social support availability and how they enhance awareness about social influence on Health. Those works belonging to large projects are grouped for the sake of objectivity.

3.5.6 WANDA

WANDA (LAN et al., 2012; ALSHURAFA et al., 2014a,b; SIDERIS et al., 2015; ALSHURAFA et al., 2016) is a remote health monitoring system that uses patients baseline data and contextual information (collected by patients' smarphones) to improve NCDs care. According to Lan et al., monitoring systems failed to achieve significant improvements in heart failure treatment. In general they tend to be invasive and reactive. WANDA is a proactive, non-invasive platform integrated with smartphones, which has an analytics engine and a social platform. The engine analyzes patients' data that are collected through automatic sensing to provide cus- tom monitoring and notifications, while the social platform adds cooperation and competition features among system users.

WANDA has gained great improvements since its first publication. WANDA remote monitoring system was already used to predict the success in care of certain risk factors related

Figure 8: Total of papers by classification through years



Source: Own authorship

to cardiovascular diseases, adherence of physical activity practice, daily questionnaire completion, blood pressure measurement of the Women's Heart Health Study participants, treatment adherence prediction after intervention by social support, and to identify patients who are most likely to be successful in their treatment to receive custom care plans in advance.

3.5.7 Patient Journey Record System

Patient Journey is a metaphor that understands that the patient, as well as a traveler, is on a journey, and he will need assistance to deal with situations he is not used to in different stages of the disease (MARTIN et al., 2011, 2012). Patient Journey Record System (PAJR) is a patient-centered framework proposal that aims to integrate information systems, social networks and digital democracy, so that different agents can construct a health support system collaboratively, taking into consideration that each patient has his own journey.

PAJR was already used to analyze the responses given by patients with chronic diseases to semi structured questionnaire questions asked by "care guides". The analysis performed by the PaJR verifies the severity of the reports made by the patients and classifies according to their severity, identifying in advance the need for intervention in the patients.

The system evaluation was took from November 2010 to December 2011 in a controlled experiment where 153 patients participated in the intervention group and 61 in the control group. Overall the intervention group had a 50% lower admission number than the control group. In addition, the model used by the system was able to classify 100% of cases of unplanned urgent events.

According to authors, although the system has achieved good results, there is still a need for further evaluation.

3.5.8 Accessible telehealth - Leveraging consumer-level technologies and social networking functionalities for senior care

Accessible Telehealth (DHILLON; WüNSCHE; LUTTEROTH, 2013) is a conceptual framework for elderly care. The framework design was elaborated by reviewing the existing absences between different types of care platforms, namely: telehealth, health record, health information web sites and serious games, and by the collection of requirements made by interviewing patients. Within the functionalities identified are social and emotional support. An prototype was developed implementing features of social support in the form of social network, where users could create groups and search for friends.

An evaluation of the prototype was carried out by 43 seniors. In the evaluation, 35% of patients agreed that the social functionalities motivated them to use the system. While 31% agreed that the involvement of friends helped them manage their health.

3.5.9 Towards chronic emergency response communities for anaphylaxis

Schwartz et al. (SCHWARTZ et al., 2014) propose a workflow approach to help chronic patients facing emergency situations. In this approach, patients who face emergency situations and do not have the resources required to deal with the situation are able to trigger an alert request. This request is passed to those individuals (members of the system) who have the resources needed to help and are within a distance range that allows for timely intervention. Members may accept or reject the aid request. This type of mechanism, where patients request each other help in health situations, was called by the authors as “social medicine”.

3.5.10 You Tweet What You Eat: Studying Food Consumption Through Twitter

Food consumption screening is done through questionnaires which has a high cost. So- cial networking users, such as twitter, often share information about foods consumed. Social network data analysis, such as Twitter data, makes it possible to identify diet trends at a geographical level (e.g., at a state or city level), as well as personal habits, and the relationship these habits have with the social network of individuals (ABBAR; MEJOVA; WEBER, 2015).

Using a Naive Bayes model, the authors classified the type of food consumed according to the text of messages shared in Twitter. Crossing the information about food consumption and caloric level of foods it was possible to correlate obesity and diabetes at the state level and to validate them with the results published by the American Centers for Disease Control and Prevention (CDC).

The authors were also able to predict obesity and diabetes at individual levels. To do so, a regression model inferred the risk of obesity based on the type of food shared by users in their messages. Finally, social network analysis performed by the authors found that the probability of being obese or having diabetes increases when one is connected to other obese or diabetic nodes, that is, nodes of the same network share messages containing similar foods.

3.5.11 Related Works Comparison

The previously presented works are compared in terms of pervasiveness, adaptation according to users profile, social support and awareness of social situation. Table 7 summarizes this comparison. The column Pervasive indicates if the work proposes the provision of assistance to users anytime and anywhere, that is, if it is pervasive (SATYANARAYANAN, 2001). Adaptive indicates if the work proposes the adaptation of its operation according to users' behaviors or situations they are inserted, in other words, if its supports context awareness (DEY; ABOWD; SALBER, 2001). Column Social Support Enabled indicates if the work proposes, in some manner, the participation of others in improving the health of its users. Finally, Social Aware indicates if the work proposes a mean for awareness the social influence of others on the health of its users, or the influence of its user on health of others, in some way that it can recommend connections that might benefit the health of its users, or improve the understanding of how users' behaviors influence the health of others.

Table 7: Related Works Comparison

Title	**Pervasive**	**Adaptive**	**Social Support Enabled**	**Social Aware**
WANDA (LAN et al., 2012; ALSHURAFA et al., 2014a,b; SIDERIS et al., 2015; ALSHURAFA et al., 2016)	Yes	Yes	Yes	No
Patient Journey Record System (MARTIN et al., 2011, 2012)	No	Yes	Yes	No
Accessible Telehealth (DHILLON; WüNSCHE; LUTTEROTH, 2013)	No	No	Yes	No
Towards chronic emergency response communities for anaphylaxis (SCHWARTZ et al., 2014)	Yes	Yes	Yes	No
You Tweet What You Eat (ABBAR; MEJOVA; WEBER, 2015)	No	No	No	Yes

Source: Own authorship

WANDA and Towards chronic emergency response communities for anaphylaxis are understood as pervasive once they provide a platform that can be used anytime and anywhere by its users, providing continuous care. They are also adaptive, as so is Patient Journey Record System. WANDA is able to predict engagement of users in their care according to users behaviors. Patient Journey Record System checks users' answers to questionnaires to identify in advance the need of intervention. By its turn, Towards chronic emergency response communities for anaphylaxis is aware of users location and resources possession to indicate the possible caregivers in emergency situations. With exception of You Tweet What You Eat, all other works offer some functionality of social support. WANDA detects the need of social intervention to improve engagement, Patient Journey Record System is based on contacts between patients and caregivers, Accessible Telehealth offers a social network to improve the collaboration among users, and in Towards chronic emergency response communities for anaphylaxis the users can request the help of others. Finally, just You Tweet What You Eat is social aware once its model considers the influence of others in the probability of having a health condition. Thus, different from the related works presented in this chapter, Pompilos aims on integrating pervasiveness, adaptability and social data to improve social support by giving user awareness of their social context. That is, provide Pompilos's users with social recommendations for the improvement of their health, as also to show how users behaviors impact on health of closest people.

3.6 Conclusions about the State of Computer Aided Social Support in NCDs

This chapter presented a mapping study with aims on understanding the state of the art field of computer aided social support for NCDs care. For this, three research questions were elaborated to guide the study. Afterwards, the questions were transformed in queries to be used in scientific repositories. The results were then clustered with suffix three method for selec- ting papers more closely related to the research questions. After that, the first pass of the three pass approach was executed for filtering papers to review. Import citations from the reviewed papers also passed by the first pass and reviewed, if found relevant. As result, 38 papers from journals, conferences and chapters from 2010 to 2016 were reviewed and classified as control- led trials, frameworks and systems, knowledge discovery, simulation models or social media usage analysis, according to their characteristics. Knowledge discovery was the predominant classification.

The use of clustering in the text selection process has the advantage of reducing the amount of texts needed to review, this is accomplished by selecting only those documents that share common topics that are under review. Another advantage is the discovering of latent topics related to the review which were not used in the search queries. However, the intersection of phrases should be done with care, as the use of a large number of topics may exclude relevant documents.

Some of the discarded papers treated social support marginally, despite the well known association between social support and health. Some papers that clearly report the use social support as an item for NCDs care improvement were not listed in the first search results. This may be due the lack of use of the specific name of the disease or risk factor, or even the type of process related with the care of a disease or risk factor like, for example, weight loss or physical activity engagement. However, those papers were included in the snowballing phase. Yet, the use of computer statistical tools for data analytics are becoming common and many papers address such use, regardless of cite the name of tools, languages used or how it may be applied. Thus, those papers focused only on data analytics which did not shown a clear application were not used for this study.

The reviewed papers were classified in one of this five classifications: controlledtrials, frameworks and systems, knowledge discovery, simulation models or social media usage analysis. Controlled trials papers present controlled experiments that assess the effects and outcomes of using social support interventions in conjunction with the use of information technology tools in the care of NCDs and its risk factors. Those papers classified as knowledge discovery used social data to detect health trends by location and to identify and predict users health statuses. While, frameworks and systems papers shown computational systems or models of computational systems that integrate social support features or social data in the care of NCDs or its risk factors. Simulation models papers are characterized by the use of simulation modeling to understand the relations between social activities and NCDs spreading. Finally, social media usage analysis papers address how social media or Internet driven platforms are used to support NCDs care.

Most of the reviewed papers treated about knowledge discovery, and the use of Twitter messages for this matter has become standard. This may be due the increase of interest in big data and machine learning since 2013 (TRENDS, 2016).

Much of the computing aided social support is made by Internet driven platforms such as blogs, instant messaging, forums, wikis or video sharing. These tools enable the exchange of knowledge between patients, caregivers and physicians. Moreover, they have the ability of increasing confidence and self-efficacy of patients. Today's smartphones and gadgets have a great capacity in terms of sensing patients, for example, they can be used to infer user activity type and frequency, sleep quality, determine user location or communicate with other people. However, the real power of these technologies in offering social support seems to be underused. Nonetheless, this effort can be seen in WANDA project (LAN et al., 2012; ALSHURAFA et al., 2014a, 2016; SIDERIS et al., 2015; ALSHURAFA et al., 2014b) which has used baseline and smartphone sensor data to detected in advance the need of intervention in patients. Or in Schwartz et al. paper (SCHWARTZ et al., 2014), which uses social collaboration to help people in emergency situations.

By the review was possible to glimpse that there is still room to research the use of social data, mobile technologies, smart devices and things for aiding social support in NCDs care. The use of these technologies may be used to turn people more aware of how their behaviors may impact on health of others, or to help them to find better partners to help them in their care ina way not yet seen.

4 POMPILOS ONTO

The concepts that exist in the social relations of patients with NCDs, i.e., the social relations of the patient and the implications that these relations exercise on their health can be described by an ontology. An ontology is a formal description of concepts of a domain of knowledge, being used to seek answers to questions formulated for this domain (GRIMM et al., 2011). Since any person (**ego node**) is virtually connected to someone (**alters nodes**), any intervention made in a ego node will generate an effect (negative or positive) on alters nodes, i.e. generates **an adverse effect on health**. Such adverse effects are also known as externalities (CHRISTAKIS, 2004).

Examples of such externalities are demonstrated by the influence of nodes in the cessation risk behavior on another node - quit smoking, for example (CHRISTAKIS; FOWLER, 2008). Furthermore, these externalities are also observed by the transmission of health effects from one node to another, for example, obesity, and happiness (CHRISTAKIS; FOWLER, 2007; FOWLER; CHRISTAKIS, 2008). Thus, these externalities observed in the works described by Christakis and Fowler were used to design an ontology for detecting the spread of happiness, obesity, and smoking on social networks. The role of this ontology in Pompilos is to act as a model for detecting influence between nodes and, consequently, improving social awareness on preventive care of NCDs.

The Pompilos Onto[1] extends the general data schema (section 5.1.5), mapping concepts related to the influence of social relations in the **spread of obesity, hapiness and the cessation of smoking** and is the first result acquired during this research. The design of this ontology was based on the Grüniger and Fox's methodology, which divides the development of an ontology in six parts, namely: motivation scenarios, informal competency questions, formal terminology, formal competency questions, formal axioms and completeness of the theorems (GóMEZ-PéREZ; FERNANDEZ-LOPEZ; CORCHO, 2004). Such processes will be described in more detail in the following subsections.

4.1 Motivation Scenarios

The motivation scenarios should describe those situations that are faced by applications that will use the ontology (GóMEZ-PéREZ; FERNANDEZ-LOPEZ; CORCHO, 2004). In this way, it is possible to recall two known characteristics of NCDs care. First, it is known that through the control of risk factors it is possible to prevent NCDs such as heart disease, stroke and cancers (WHO, 2005). Moreover, it is also known that some risk factors such as obesity and smoking can be generated by the influence of alters from the ego's social network (CH- RISTAKIS; FOWLER, 2007, 2008; FOWLER; CHRISTAKIS, 2008; MADAN et al., 2010a,b).

[1]The findings of this chapter are available in the annals of 2016 XLII Latin American Computing Conference (CLEI) and expanded to be published in the International Journal of Metadata, Semantics and Ontologies (VIANNA et al., 2016, 2018)

Thus the construction of the proposed ontology was guided by the following scenario:

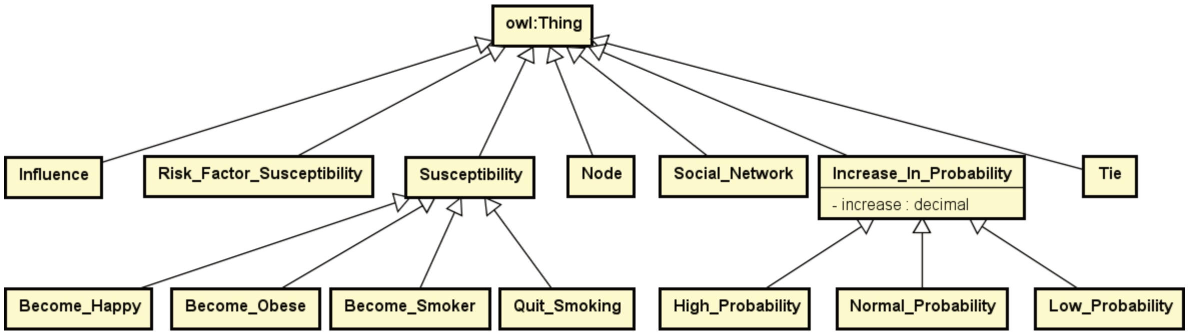
> "AmbiensDuctor is a mobile personal assistant, which uses the profile information and social network of its users to suggest new social connections that can benefit their health. In addition, the AmbiensDuctor must present to their users those existing connections that offer better health benefits.

4.2 Informal Competence Questions

Relying on the proposed scenario was prepared twelve questions that the ontology should be able to meet. Such questions are written in natural language, and will be used to assess whether the ontology will satisfy the proposed scenario.

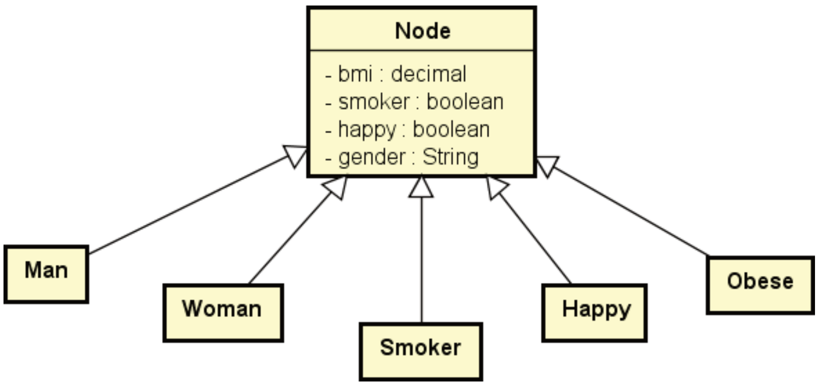
1. Given a social network, what kind of social relations exist?
2. Given a social network, what kind of distance relationships exist?
3. Given a social network, what kind of influences are there?
4. Given a social network, which nodes generate more influence?
5. Given a node linked to a social network, in which network nodes it has influence?
6. Given a node connected to a social network, from which network nodes it receives influence?
7. Given a node connected to a social network, with which nodes it is connected, according to each existing distance relationship?
8. Given a node connected to a social network, to what kind of risk factors it is susceptible?
9. Given a node linked to a social network, how it influences other nodes in the network?
10. Given a node associated with a social network, which nodes are suggested to form a new connection, in order to it obtains a lower likelihood of obesity?
11. Given a node associated with a social network, which nodes are suggested to form a new connection, in order to it obtains a lower likelihood of smoking?
12. Given a node associated with a social network, which nodes are suggested to form a new connection, in order to it obtains a greater likelihood of happiness?

Figure 9: Ontology Main Classes



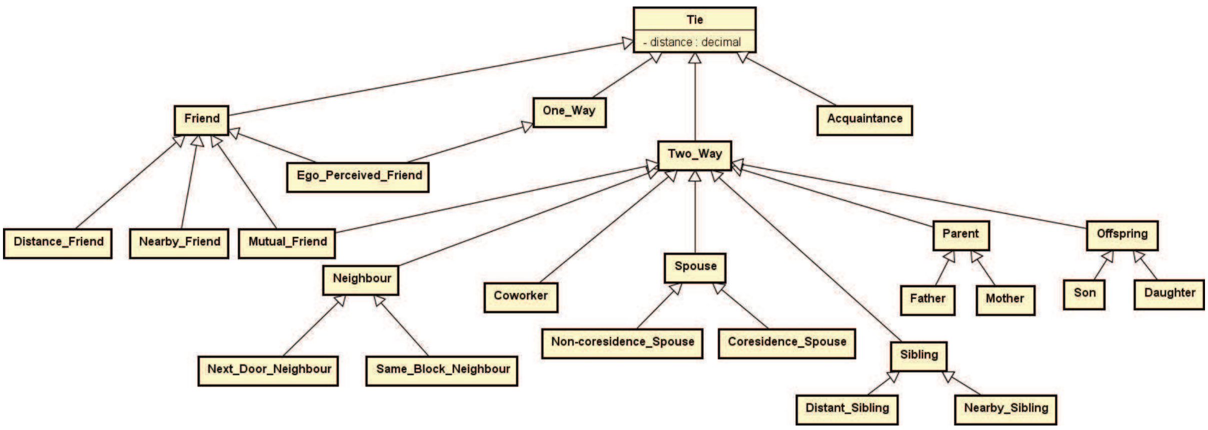
Source: Own authorship

Figure 10: Node Subclasses



Source: Own authorship

Figure 11: Tie Subclasses



Source: Own authorship

4.3 Formal Terminology

In this phase, the knowledge gained in the preparation of competence questions are used to formally define the concepts that represent the ontology (GóMEZ-PéREZ; FERNANDEZ-

LOPEZ; CORCHO, 2004). For simplicity in presenting the basic concepts of the ontology, this thesis will use UML class diagrams notation [2]. The Manchester syntax for OWL2 [3] will be used to demonstrate the equivalence, which are more complex concepts.

Figures 9, 10, 11 and 12 use the UML generalization notation to represent the formal concept of subsumption. The figure 13 uses the UML association notation to describe the formal concepts of domain and range of object properties. Data properties are presented by UML class attributes.

Figure 9 shows the higher level subsumptions from the ontology. Class *Risk_Factor_Susceptibility* represents the probability of a influence become a risk factor. Thus, this class is equivalent to *Be- come_Obese*, formally described by the axiom *Risk_Factor_Susceptibility* == *Become_Obese*. Class *Influence* sets out the types of influence that a node can have on another. Thus, accor- ding to the scope of the ontology, an influence can be *Become_Happy*, *Become_Obese*, *Be- come_Smoker* or *Quit_Smoking*. *Node* is a person connected to a social network, represented by *Social_Network*. Class *Increase_In_Probability* indicates the occurrence of an influence, which may be high (class *High_Probability*), normal (class *Normal_Probability*) or low (class *Low_Probability*).

The subsumptions and attributes of a Node are shown in Figure 10, so that a Node can be smoker (class *Smoker*), happy (class *Happy*), obese (class *Obese*), woman (class *Woman*) or man (class *Man*). Figure 11 presents the subsumptions and attributes related to *Tie* class. This class represents the knowledge about the connection between two nodes, for example, father (*Father*), son (*Son*), friend (*Friend*). The attribute distance is used to establish the physical dis- tance in meters between two nodes and is used to infer relationships that vary according to the distance (*Distance_Friend*, *Nearby_Friend*, *Next_Door_Neighbour*, *Same_block_Neighbour*, *Non-residence_Spouse*, *Coresidence_Spouse*, *Distant_Sibling*, *Nearby_Sibling*). Class *Two_-Way* defines those relationships that are reciprocal. The class *One_Way* is a disjunction with the *Two_Way* class, i.e. those relationships that are not understood as reciprocal. Thus, the classes Friend and Acquaintance, are direct subsumptions of *Tie*, since the definition of reciprocity for these classes varies for each individual.

The ontology object properties are shown in Figure 13. The attribute *belongsTo* establishes the participation of a node in a social network, while attribute *hasNode* denotes its inversion. The attribute *hasDegreeOfSeparation* defines the degree of separation of one node to another. The *hasTie* attribute indicates the perception of an ego being connected to an alter. Thus, the with attribute sets the alter node of this connection. The attribute spreads describes the influence one alter has on an ego, described by the attribute *on*. Finally, the *hasProbability* attribute defines the degree of this probability, i.e., high, normal or low.

Figure 12 shows the existing types of degree of separation. *has1Degree* indicates that two nodes have a direct connection. *has2Degree* indicates that the path between two nodes is se-

[2]http://www.uml.org/what-is-uml.htm

[3]https://www.w3.org/TR/owl2-manchester-syntax/

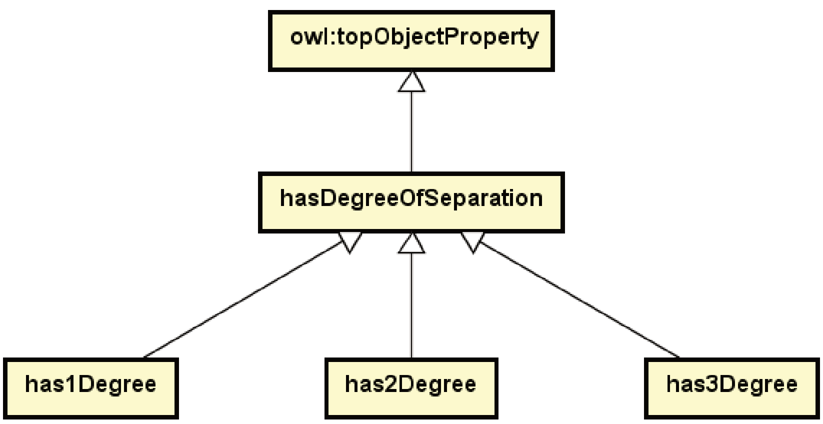
parated by an intermediate node, and *has3Degree* indicates the connection between two nodes have two intermediate nodes. Finally, table 8 shows the axioms of equivalence of the classes from this section.

Table 8: Axioms of Equivalence

Class	**Axiom of Equivalence**
High_Increase	Increase_In_Probability and (increase some xsd:decimal[>= 50])
Low_Increase	Increase_In_Probability and (increase some xsd:decimal[<25])
Normal_Increase	Increase_In_Probability and (increase some xsd:decimal[>= 25 , <50])
Risk_Factor_Susceptibility	Become_Obese or Become_Smoker
Woman	Node and (gender value ”female”)
Man	Node and (gender value ”male”)
Happy	Node and (happy value true)
Obese	Node and and (bmi some xsd:decimal[>= 30])
Smoker	Node and (smoker value true)
One_Way	(Acquaintance or Friend) and (not (Two_Way))
Ego_Preceived_Friend	(Friend) and (not (Two_Way))
Next_Door_Neighbour	Neighbour and (distance some xsd:decimal[<25])
Same_Block_Neighbour	Neighbour and (distance some xsd:decimal[>= 25])
Distant_Sibling	Sibling and (distance some xsd:decimal[>= 1600])
Nearby_Sibling	Sibling and (distance some xsd:decimal[<1600])
Coresidence_Spouse	Spouse and (distance some xsd:decimal[<25])
Non-residence_Spouse	Spouse and (distance some xsd:decimal[>= 25])

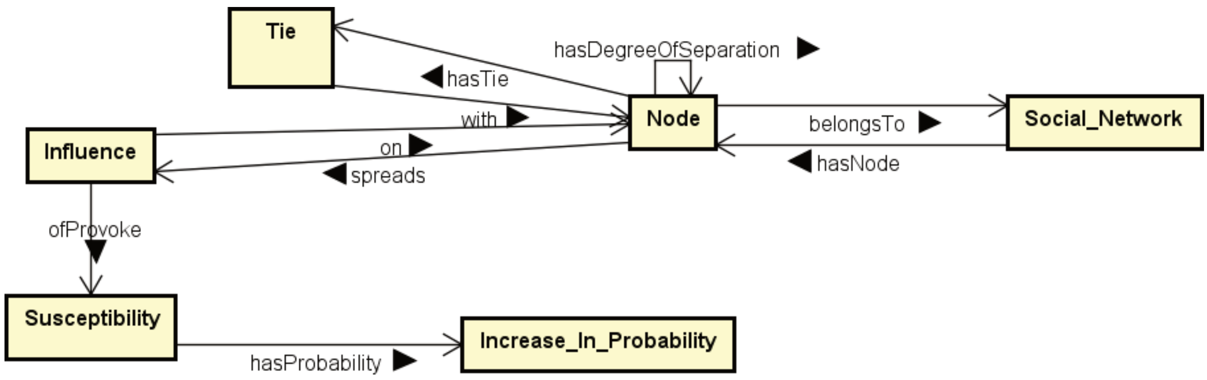
Source: Own authorship

Figure 12: Sub-properties of Degree_of_Separation



Source: Own authorship

Figure 13: Object Properties



Source: Own authorship

4.4 Formal Axioms

Formal axioms should define the conditions for questions of competence become complete (GóMEZ-PéREZ; FERNANDEZ-LOPEZ; CORCHO, 2004). The axioms presented in this section focused on the description of how nodes in a social network can influence the spread of obesity. According to the findings described in (CHRISTAKIS; FOWLER, 2007), the increase in risk of obesity is 45% higher for alters directly connected to an ego (i.e. separation of 1 degree); 20% greater for alters' alters (i.e. separation of 2 degrees) and 10% greater for alters' alters' alters (i.e. separation of 3 degrees). Besides the risk defined by the social distance (CHRISTAKIS; FOWLER, 2007) describes the increased risk of an Ego become obese, if an Alter became obese, according to the connection type among pairs. These probabilities are shown in Table 9.

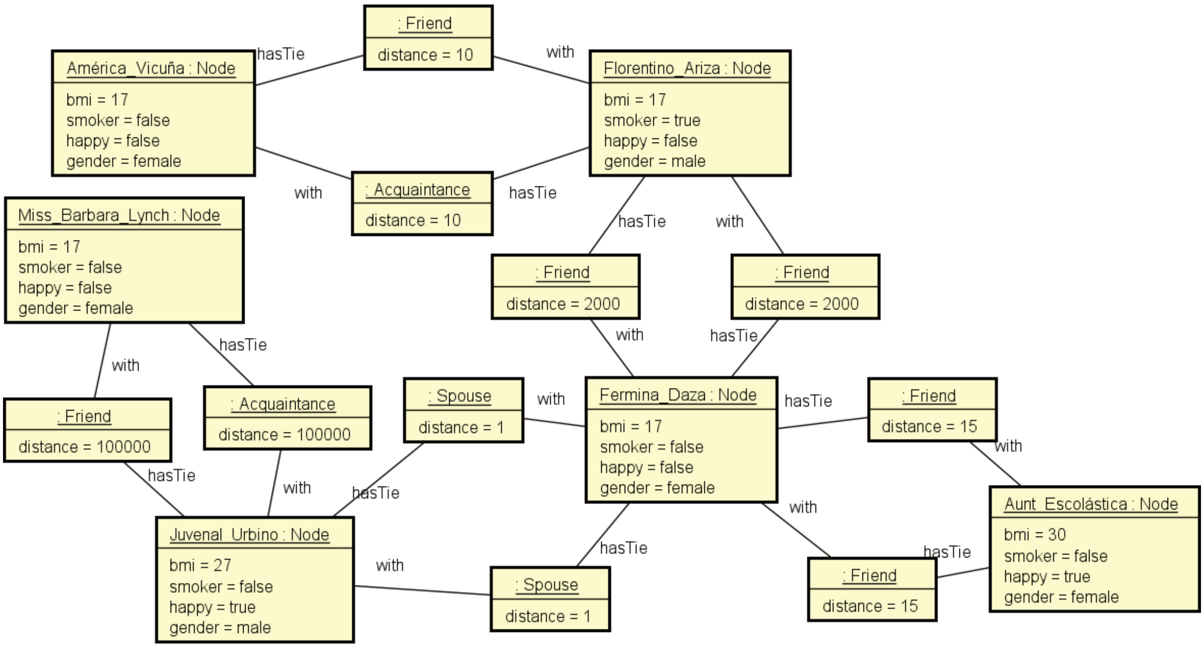
Table 9: Increase in Probability of an Ego Becoming Obese by Tie Type

Tie Type	**Increase in Probability of an Ego Becoming Obese**
Ego perceived friendship	57%
Mutual friendiship	171%
Friendship where both nodes are men	100%
Adults siblings	40%
Same sex siblings (i.e. brother-brother)	44%
Same sex siblings (i.e. sister-sister)	67%
Spouse	37%

Source: (CHRISTAKIS; FOWLER, 2007)

In order to test the concepts that will be expressed through the axioms described in this section, a small social network of six individuals was created. Figure 14 presents a small social network represented by a UML class diagram, where individuals and links are presented as instances of classes Node and Tie, respectively.

Figure 14: Social Network Individuals



Source: Own authorship

The axioms that involve transformations necessary to resolve the question of competence as well as the solution of the question itself will be made with the use of query commands and axioms of the *PROWLOG* environment [4].

The PROWLOG is an environment for programming applications using OWL ontologies, allowing the definition of axioms and running queries through syntax similar to SPARQL[5]. Furthermore, PROWLOG is based on the Prolog language, allowing programs written in that language to use ontologies.

The axioms presented in this and the next section use the change command *add* and *delete* (addition and removal of axioms, respectively), the query command *select* and the assertion axioms of PROWLOG . The command *add* has the following syntax:

$$add\{< Axiom_1 >, ..., < Axiom_n >\}\ where$$

$$\{< Query_1 >, ..., < Query_n >\}$$

Where $Axiom_i$ is a valid PROWLOG axiom (as in Table 10) and $Query_i$ is a query expression. Similarly, a *delete* command has the following syntax:

$$delete\{< Axiom_1 >, ..., < Axiom_n >\}\ where$$

[4]http://obaa.unisinos.br/drupal7/?q=node/8
[5]https://www.w3.org/TR/rdf-sparql-query/

$$\{< Query_1 >, ..., < Query_n >\}$$

The *select* query command has the following syntax:

$$select < Var_1 >, ..., < Var_n > where$$

$$\{< Query_1 >, ..., < Query_n >\} => Results$$

Where Var_i is a Prolog variable, $Query_i$ is query expression and *Results* is a variable that unifies the query results.

Table 10: Axioms of Assertion

Axiom	**Meaning**
Indiv:Class	Individual *Indiv* of type *Class*
Subj:Prop(Obj)	Individual *Subj* that has a object property *Prop* of range *Obj*
prop(Subj) – Value	Individual *Subj* that has a data property *prop* with a value equals to *Value*
I_1 \= I_2 ... \= I_n	Different individuals

Source: Own authorship

Figure 15: Addition of tie Juvenal_Urbino x Florentino_Ariza as mutual friends

```
add {
  'Juvenal_Florentino':'Friend',
  'Juvenal_Urbino':hasTie('Juvenal_Florentino'),
  'Juvenal_Florentino':with('Florentino_Ariza'),
  'Florentino_Juvenal':'Friend',
  'Florentino_Ariza':hasTie('Florentino_Juvenal'),
  'Florentino_Juvenal':with('Juvenal_Urbino')
}
```

Source: Own authorship

To identify which nodes are most suitable for connecting to an ego, so this get a lower likelihood of obesity, the following strategy will be used:

1. Choosing an existing node on the network as ego and create new connections in order to test the ones that best benefits it (in this example the node **Juvenal_Urbino** will be used as ego);

2. Creating connections of **Friend** type with the three nodes that the ego is not connected. That is, **América_Vicuña**, **Florentino_Ariza** and **Aunt_Escolástica**;

3. The connection **Juvenal_Urbino** with **Florentino_Ariza** will be bidirectional, the others connections will be perceived only by the ego.

Figure 16: Addition of individuals of “Influence” class

```
add {
  'Increase_45_Percent':'Increase_In_Probability',
  increase('Increase_45_Percent') — 45^^float,
  'Influence_1Degree'
  :'Become_Obese',
  'Influence_1Degree'
  :hasProbability('Increase_45_Percent'),

  'Inc_57_Percent':'Increase_In_Probability',
  increase('Inc_57_Percent') — 57^^float,
  'Influence_Friendship_Perceived_From_Ego'
  :'Become_Obese',
  'Influence_Friendship_Perceived_From_Ego'
  :hasProbability('Inc_57_Percent')
}.
```

Source: Own authorship

Figure 17: Addition of social distance relations between nodes

```
add {
  Ego:has1Degree(AlterDegree1),
  Ego:has2Degree(AlterDegree2),
  Ego:has3Degree(AlterDegree3)
} where {
  Ego:hasTie(TieDegree1),
  TieDegree1:with(AlterDegree1),
  AlterDegree1:hasTie(TieDegree2),
  TieDegree2:with(AlterDegree2),
  AlterDegree2:hasTie(TieDegree3),
  TieDegree3:with(AlterDegree3),
  Ego \= AlterDegree1,
  AlterDegree1 \= AlterDegree2,
  Ego \= AlterDegree2,
  Ego \= AlterDegree3,
  AlterDegree1 \= AlterDegree3,
  AlterDegree2 \= AlterDegree3
}.
```

Source: Own authorship

Figure 15 illustrates an addition of axioms to define connections between the ego Juvenal_Urbino and the alter Florentino_Ariza. Connections to other two alters are not displayed, since they have the same format.

Figure 16 presents the definition of an axiom that represents an influence generated by social distance between nodes, and an axiom that represents the influence generated according to the type of connection between nodes (Table 9). The creation of other axioms was omitted since they follow the same pattern.

After the addition of axioms that represent the influences, the social distance between network nodes is defined as shown in Figure 17.

Figure 18: Addition of friendship connections according to their type

```
add {
  EgoFriendTie: 'Mutual_Friend'
} where {
  Ego:hasTie(EgoFriendTie),
  EgoFriendTie: 'Friend',
  EgoFriendTie:with(AlterFriend),
  AlterFriend:hasTie(AlterFriendTie),
  AlterFriendTie: 'Friend',
  AlterFriendTie:with(Ego)
}.

add {
  EgoFriendTie: 'Ego_Perceived_Friend'
} where {
  Ego:hasTie(EgoFriendTie),
  EgoFriendTie: 'Friend'
}.

delete {
  EgoFriendTie: 'Ego_Perceived_Friend'
} where {
  EgoFriendTie: 'Mutual_Friend'
}.
```

Source: Own authorship

Once the social distances between nodes is defined, the axioms that represent links between nodes are added, which are mutual friendship and ego perceived friendship, as shown in Figure 18. For this case, the connection type spouses was defined preliminarily (Figure 14). The link types adult brothers and brothers of the same sex were not part of the preliminary definition, and is understood that the examples in this section already cover the situations faced by individuals of these classes.

The next step will be to add the influences that egos get from their alters. The Figure 19 illustrates the command for axiom addition that indicates the influence of an obese alter on an ego, where the ego perceives this alter as a friend, but the inverse does not occur (e.g. friendship perceived by the ego). Other types of axioms have been omitted due to similarity they have with this command.

Figure 19: Addition of the alter on ego influence axiom

```
add {
  Alter:'spreads'('Influence_Friendship_Perceived_From_Ego'),
  'Influence_Friendship_Perceived_From_Ego':'on'(Ego)
} where {
  Ego:hasTie(Tie),
  Tie:'Ego_Perceived_Friend',
  Tie:with(Alter),
  Alter:'Obese',
  Alter:hasTie(AlterTie),
  AlterTie:with(Ego)
}.
```

Source: Own authorship

Finally, after the addition of the axioms of influence, it will be possible to formally solve the competence question. This solution is presented in the next section.

4.5 Formal Competence Questions

The formal competence questions denote the solutions of the questions presented in section 4.2, being written using axioms from the ontology. The complete solution of all competence questions is very extensive, since each solution needs to be given the necessary changes in the knowledge base that has condition to answer the question. In addition, it is necessary to present theoretical concepts that underlie the solution of each question.

As proof of concept of the validity of ontology, this thesis presents the solution to the competence question number 10, **"Given a node associated with a social network, which nodes are suggested to form a new connection, in order to it obtains a lower likelihood of obesity?"**. The choice for this question is due to its involving the changes necessary for the solution of other existing questions. It is also similar to the questions 11 and 12, and its definition encompasses the purpose of ontology, which is the detection of influence dissemination of happiness, obesity and smoking on social networks.

So the question 10, considering individuals created through the commands listed in the previous section, can be solved by Figure 20. In other words, Figure 20 defines that the most appropriate nodes for connection are those that are less likely to affect other nodes to become obese.

Figure 20: Solution Competence Question Number 10

```
select RecommendedAlter
where {
  HighInfluencer:spreads(HighInfluence),
  HighInfluence:hasProbability(HighProbability),
  HighInfluence:'Become_Obese',
  HighProbability:'High_Increase',
  RecommendedAlter:'Node',
  RecommendedAlter \= HighInfluencer
}.
```

Source: Own authorship

4.6 Completeness of Theorems

In completeness theorems are defined the conditions necessary for the competence questions to be complete. Thus, the question of competence 10 will be complete when the influences that the nodes have on other nodes are defined, i.e., the generation of individuals from the class Influence and their relations *on* and *spreads* with individuals of the Node class. Moreover, it is necessary to define the degree of influence exercised through the relation hasProbability that exists between individuals of classes Influence and Increase_In_Probability.

These rules can also be applied to the questions 11 and 12, as they have similar meaning, although address influences of different types (happiness and smoking). Questions 1 to 4 are also solvable by the ontology proposed, since its solution depends on terminological knowledge (e.g. the concepts defined by the classes of the ontology), with no need to create new individuals or relationships definition.

4.7 Conclusions About the Ontology Development

This chapter presented an ontology proposal for detecting the spreading of happiness, obesity and smoking in social networks. The concepts addressed in this ontology were based on three studies that analyzed social networks of individuals participating in the Framingham Heart Study project, in order to predict the influence of the social network on weight gain, on the dissemination of happiness and on smoking cessation (CHRISTAKIS; FOWLER, 2007, 2008; FOWLER; CHRISTAKIS, 2008). Furthermore, the construction of the ontology was guided by the methodology Grüniger and Fox's (GóMEZ-PéREZ; FERNANDEZ-LOPEZ; CORCHO, 2004).

Although the chapter has described the detailed solution of question number 10, others questions can be also solved using the same strategy used for that question. Finally, the proposed ontology is capable of inferring which nodes are indicated for a new connection and which exert a greater influence on the health of the patients, enabling a better health management for patients with NCDs. Next chapter will explain how this ontology could be part of a model to awareness users about their role in the health of others at the same time that users could change behaviors to improve their health.

5 POMPILOS: A SOCIAL AWARE MODEL FOR PREVENTIVE CARE OF NON COMMUNICABLE DISEASES

Social Cognitive Theory (SCT) states that behaviors are learned and reinforced by social interactions (BANDURA, 2001). Self awareness of how one can influence the health of others might also lead, not just to improvement of one's well-being, but also to the improvement of well-being of people surrounding. For example, a person aware that his eating habits are influencing people close may start to make better choices to help them in obtaining a better health.

Additionally, according to Milani, and Lavie (2016), the current model of care can no longer cope efficiently with the challenges of this century. The aging of the population, the increasing incidence of chronic diseases and the imbalance between the demand for doctors and the formation of new professionals, created the need to elaborate a new model of care. In this new model, technology plays an important role in providing continuous monitoring, access to real-time information, communication with health professionals, support in disease management, and fostering social support (and beneficial social relationships).

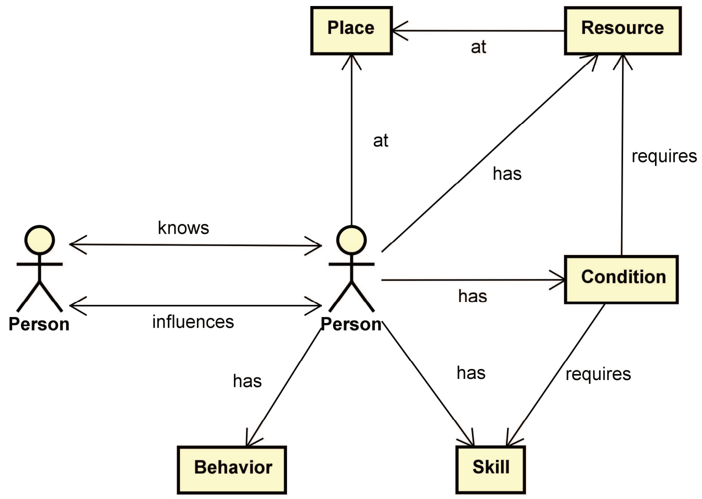
Hence, computer aided healthcare must be aware that sometimes people do not have the right partner that helps them having beneficial health gains. Thus, that help can come from people existent in their social network. Direct connected nodes of an ego could be a match for a certain health related activities, sometimes these nodes could serve as bridges to aid the formation of new beneficial ties.

In essence, social support is the beneficial influence of social relations in health of people (HOUSE; LANDIS; UMBERSON, 1988; BERKMAN et al., 2000). In a computing context, social support is successfully achieved by Internet applications such as chats, blogs, forums, wikis or video sharing (VIANNA; BARBOSA, 2017). However, few studies deal with the possibilities and opportunities of integrating the ever growing amount of computational devices and data for improvement of social support.

Today we are immersed in a world of computing devices and data. Data generated by personal devices, such as smartphones, mimic the behaviors of their owners, and then shall be applied to reveal aspects that influence in health, or to find people which the contact can influence the improvement of health. Applications are successful in providing features for continuous monitoring and caring, like diet diaries and physical activities tracking (B et al., 2013; JOHNSON; WARDLE, 2011). Also, users that participate in online communities tend to be more engaged in care activities (RICHARDSON et al., 2010), social media usage was already associated with the reception of encouragement for weight loss behaviors (HWANG et al., 2014) and tailored content has the capacity to improve users' knowledge (WEYMANN et al., 2015). However, there is no documentation of how these features can be integrated in a system nor a computing model which integrates them (VIANNA; BARBOSA, 2017).

According to the results of the literature review, the key aspects required for applications

Figure 21: Schematic model for NCDs prevention



Source: Own authorship

focused on preventive care of NCDs are **Activity tracking**, **Localized community formation**, and **Automatic content curation** (LAN et al., 2012; ALSHURAFA et al., 2014a,b; SIDERIS et al., 2015; ALSHURAFA et al., 2016; MARTIN et al., 2011, 2012; DHILLON; WüNSCHE; LUTTEROTH, 2013; SCHWARTZ et al., 2014; ABBAR; MEJOVA; WEBER, 2015). Figure 21 shows a basic schematic model containing the actors, objects and their relations regarding NCDs prevention. Person, is the central element of the model as its goal is to improve people's health. Social network is expressed in the model by the relation "Person knows Person". This knowledge may represent the influence of a person in another person. Influence may be beneficial or not. For example, the meeting of two people may represent improvement of physical activity practice, but can represent alcohol consumption. A Person has behaviors, which means any type of behavior related or not with health. An example would be, going to sleep every night around 10 PM. Condition is another relation inherent to people and, in this case, models the needs and constraints people might have. To illustrate, imagine a person with dietary restrictions. This condition requires some categories of food for consumption and skills for combining them in pleasant way. Finally, people and resources are available in places and these relations bind all the concepts. For example, an encounter of people that share the same place and conditions but have different levels of skills could be suggested as way for aiding people accomplish their care activities.

5.1 Pompilos Model

The next subsections will present Pompilos[1] which focuses in the relations between people to provide and recommend resources useful for preventive care of non-communicable diseases, as also to aware people about their influence on the care activities of others. Thus, sections 5.1.1, 5.1.2, and 5.1.3 present a **conceptual model** in which its elements are needed to build social aware software able to improve social support. These elements were extracted from the observations made on the results of the literature review (Chapter 3). Furthermore, section 5.1.4 presents the generic architecture parts that can implement the elements described in the conceptual model. This architecture was based in part on the U'Ductor model, which provides the infrastructure required to follow the dynamics between people and their underlying pla- ces, behaviors, and resources (VIANNA; BARBOSA, 2014; VIANNA; BARBOSA; PITTOLI, 2017; PITTOLI et al., 2018).

5.1.1 General Model

The core concerns of Pompilos conceptual model are expressed in Figure 22 as an UML activity diagram (GROUP, 2015). This diagram exposes a black box model showing the hierarchy of activities that must be realized in order to accomplish Pompilos' goals of improving social support by means of revealing influence on health of others and suggesting beneficial health resources.

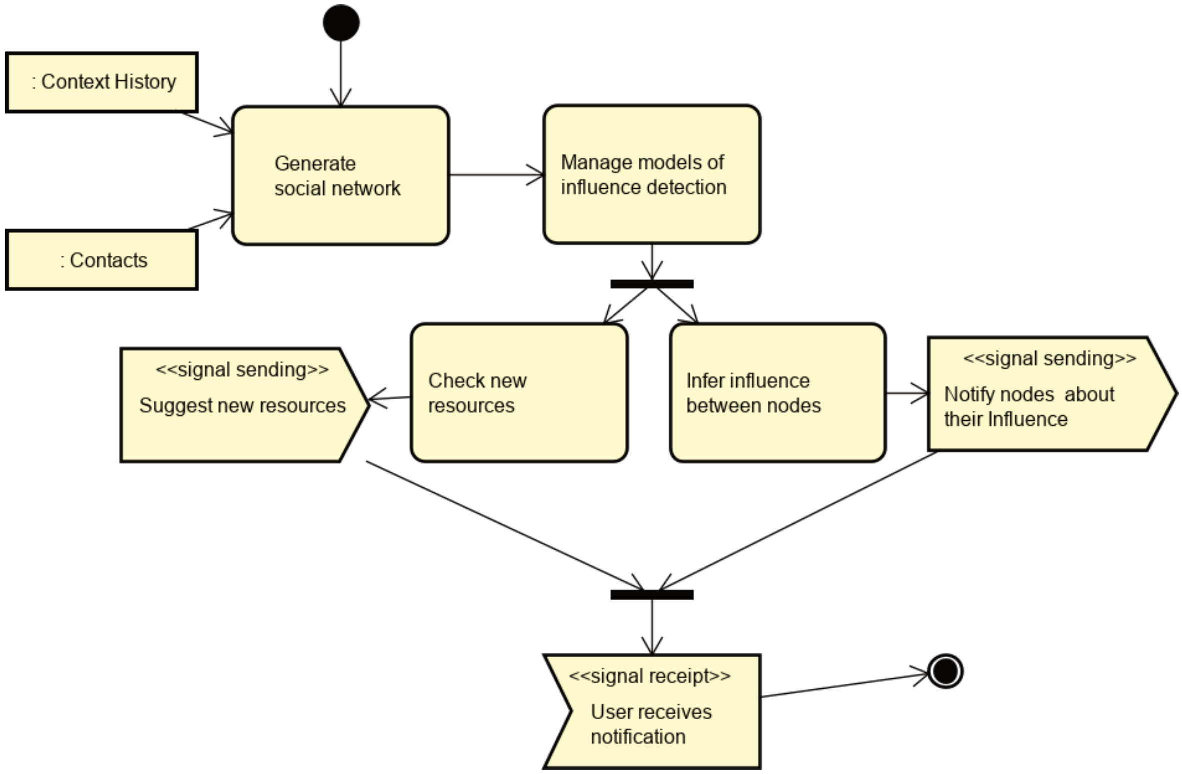
Contacts and Context History data are used as input to generate social network of users. Social networks and context information are then used to manage models of influence detection, that is, create and train models that can identify the influence of someone behaviors on health of others. Test new social resources and infer influence between nodes are executed in different nodes' data, in order to suggest new social resources that may benefit users' health and notify nodes about their Influence. Finally, the users receive notification regarding the new social resources and node influence information generated to them.

5.1.2 Social Network Generation and Models of Influence Management Dynamics

The participants and dynamics responsible for accomplish the actions of **Generate social networks** and **Manage models of influence detection** and their interactions are expressed as a communication diagram (GROUP, 2015) in Figure 23. In this communication diagram and following, the actors represent types of software agents, the objects represent data storage com-

[1]The name of the model is based on the legend of the pilot-fish. Pompilos was the name of the fisherman that protected a nymph and was later turned into a fish by Apollon (https://www.theoi.com/Nymphe/NympheOkyrhoe1.html). The U'Ductor model name, which is the base of the Pompilos model, is constructed based on the scientific name of the pilot-fish (Naucrates Ductor, or N'Ductor - https://en.wikipedia.org/wiki/Pilot_fish). U'Ductor is the Ubiquitous Ductor, as it follows their users everywhere in the same way that pilot-fishes guide ships. Yet, the Pompilos legend represents the effects of social relations.

Figure 22: Pompilos General Model



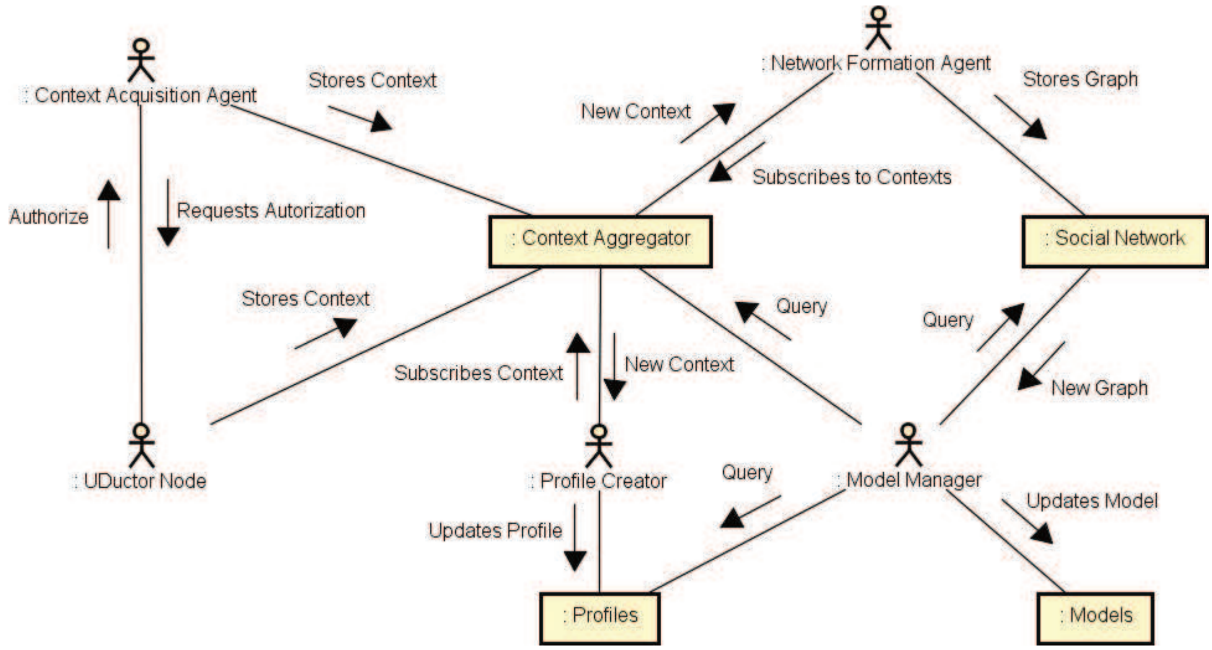
Source: Own authorship

ponents and, messages represent the actions performed by agents on data storages or other agents. Software agents types are supposed to have many instances that have autonomous behaviors, while data storage components are responsible to proxy the storage and retrieval of large amounts of data. Both, agents and data storage components, must interact with each other in order to realize the actions defined in the general model.

The *U'Ductor Node* represents instances of the U'Ductor middleware (VIANNA; BARBOSA, 2014). This instances may represent people, that generate contexts from smartphone activities, or places, that generate contexts from users visitation, resource sharing or location related status (e.g. temperature or air humidity). *Context Acquisition Agent* instances may impersonate U'Ductor's users as a mean to collect context data of user's activities on different applications. For example, a Context Acquisition Agent may be authorized to grab the Twitter activity generated by its user, which can be a like or re-tweet on others tweet. UDuctor Node and Context Acquisition Agent store the collected context data in a *Context Aggregator*. *Network Formation Agent* instances use contexts data, like contacts, phone calls, visited locations or any other relevant information received from the Context Aggregator to form social networks which are stored by Social Network storage component.

Different models to predict influence between nodes exist. So, distinct instances of *Model Manager* will co-exist, aiming to discover the influences of nodes by testing different models

Figure 23: Social Network and Model Generation Dynamics



Source: Own authorship

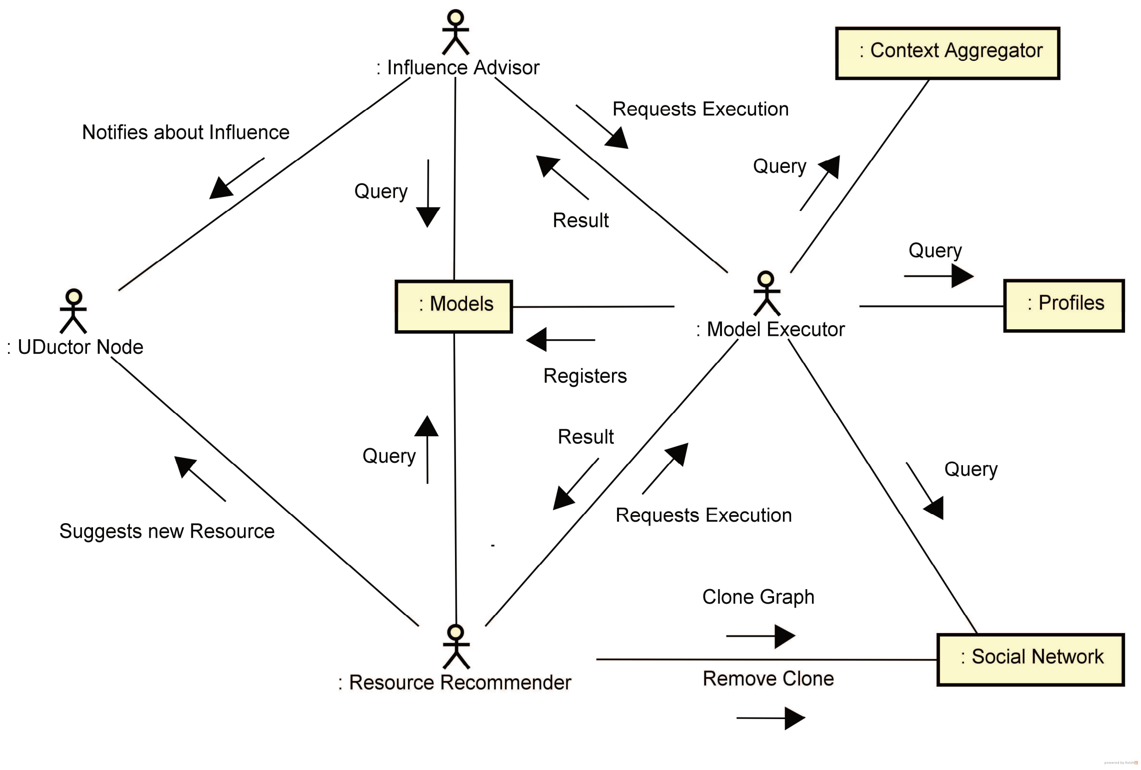
with data from context history, profiles and social network. The findings of Model Manager are then stored in the Models data storage component to be used later to infer influence on nodes. Finally, *Profile Creator* instances are responsible for creating summarizations of users' context histories. For example, this summarization can indicate the engagement of users in a particular treatment or care plan. Profile information is stored in the Profile data storage component and may be useful to the creation of new models of causal influence.

5.1.3 Influence Notification and Connection Suggestion Dynamics

Figure 24 shows a communication diagram where is expressed the participants and dynamics of the General Model's activities **Test New Connections** and **Infer Causal Influence between Nodes** (Figure 22). The interaction of Influence Advisor and Connection Recommender with others components are very similar. Both select the most appropriate model for their tasks in Models data storage. After the selection of the model, a request for execution is made to some instance of a Model Executor. These instances of Model Executor run the requested model using data from Context Aggregator, Profiles and Social Network, according to the parameters of the request.

After receiving the result from a Model Executor, Influence Advisor and Connection Recommender might interpret that result and execute further processing in order to send to Personal Nodes the influence that they exercise on others nodes and to recommend new beneficial health connections. In particular Connection Recommender may want to clone social network graphs in order to simulate new connections and infer its outcomes.

Figure 24: Influence Notification and Connection Suggestion Dynamics



Source: Own authorship

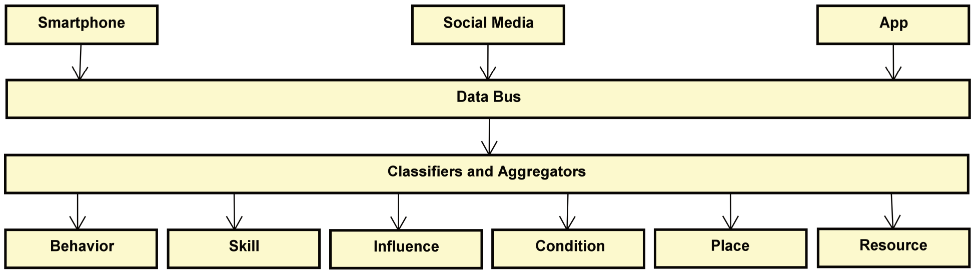
5.2 Conceptual Architecture

An architecture which the goal is to improve people's health must, above all, be built for staying aware of the relations between people and the environment where they are inserted, as this regards to NCDs prevention[2]. It must capture the relationships between people as they are exposed to mutual influence due their behaviors. For example, the meeting of two people may represent an improvement of physical activity practice but can represent alcohol consump- tion. The model must also help applications adapt to people constraints as they might have particular conditions. Finally, the model must know that people are available in places and use this knowledge to support the encounter of people for helping them to accomplish their goals collaboratively.

NCDs can be prevented by controlling their risk factors like obesity, overweight, hyperten- sion or diabetes. Those, by their turn, have relation with routine activities like diet and physical

[2]The architecture described in this section was published on the Information Processing Letters with the title "A scalable model for building context-aware applications for noncommunicable diseases prevention" - https://doi.org/10.1016/j.ipl.2019.03.010

Figure 25: Data flow schema



Source: Own authorship

activities practice. So, the key aspects required for applications focused on preventive care of NCDs are Activity tracking, Localized community formation and Automatic content curation. Activity tracking is the planning, notification, and accompaniment of preventive care activities, walk every day by 30 minutes, for example. Localized community formation is the automatic content generation of communities based by the localization and conditions of people. Automatic content curation is the automatic collection of content related to the condition of the users from social media platforms.

As computing, data is also becoming more ubiquitous and coming from different sources and sometimes unstructured. So, a strategy is necessary to streamline those different sources of data for distributing among different processors, as shown in Figure 25. The Data Bus implements the publish-subscribe model, it captures data from different sources as smartphones, social medias, and applications to distribute to classifiers and aggregators that subscribed to them. The main goal of a classifier is to semantically annotate a set of data. For example, messages shared in social networks may be used to identify the skills a person has. Yet, aggregators consolidate data from different sources in a single set. The generation of medical records is one example of use of aggregators. In this way, data from medical screenings could be used to generate derived data, for example, using cholesterol and glycemic index to determine cardiovascular disease risk information.

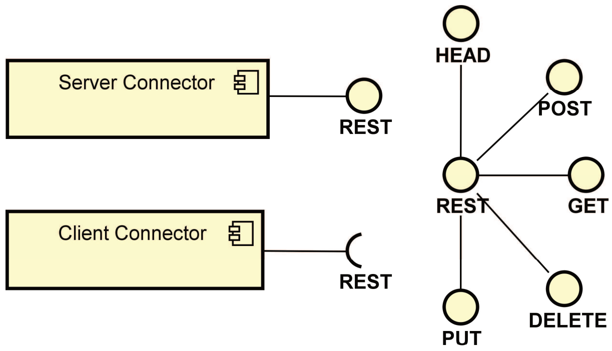
Having the classified and aggregated data is just a part for designing and building applications for preventive care of NCDs. A conceptual architecture is necessary to guide this development which requires information scalability, support for near real-time notification, access control, physical integration, and customization, so the architecture style is the first decision made in the designing phase. REST (Representational State Transfer) style was chosen as it was designed to ensure high scalability, so one of the requirements in this style is not to keep client sessions stored on servers. This feature guarantees some advantages such as the possibility of replacing a component, or the parallelization of requests through multiple servers. Portability is another feature of this style. In a REST application the same resource can be presented in different formats, so if the client cannot display the default format it can use an alternative format available from the server (FIELDING; TAYLOR, 2000). The ease of service implementation is another positive aspect of this style, since the number of architectural decisions is lower in relation to the WSDL/SOAP-based service implementations (PAUTASSO; ZIMMERMANN; LEYMANN, 2008).

In REST, a component is an abstraction of some client or server entity. Clients and servers communicate with each other through connectors that abstract the understanding of therequest and response protocol used between them, that is, they abstract their uniform interface. The server connector is responsible for interpreting the request made by the client for a given resource and transferring a representation of that resource to the client. In this way, client and server connectors can be encapsulated to offer different functionalities, such as security, caching, or tunneling information. The representation determines how a resource can be offered. For example, a book may be offered in an audio or text format. Finally, a resource is defined as "any information that can be named" (FIELDING; TAYLOR, 2000). In summary, Pompilos uses a REST architectural style to interface the communication between its components and clients.

Near real-time notification is brought by two components, which work in conjunction: **Context** and **Notification**. The Context component allows clients to register their interest in attributes that identify the status of an entity represented by the conceptual architecture, and to receive notifications of changes to those attributes. If the entity represents a person these attributes could be his temperature, blood pressure, glycemic index, shared messages on a social media or any other relevant attribute that the person shares. In the case of a location, these attributes may be air quality index, air relative humidity, ultraviolet index, or other pertinent information shared by the site. The operation of the Context component is based on the conceptual framework of context management defined by Dey et al. (DEY; ABOWD; SALBER, 2001). In the case of Pompilos the Context component works with a data aggregator, in charge of recei- ving new context information, which arrives through clients. In other words, clients may update context information of an entity, if they are authorized. The Notification component is responsible for informing subscribers about contextual updates, enabling near real time information transmission.

The Access Control component is responsible for managing the access permissions of the clients to the functionalities of the other components belonging to the Pompilos. For exam- ple, information that indicates that a client is authorized to receive temperature updates from an entity is stored by this component. Technically this accomplished using OAuth 2 Client Credentials Flow (HARDT, 2012). Three parts compose this flow: application registration, authentication request and authentication response. The first part is not defined in the OAuth 2.0 scope and is customized to Pompilos case, where the entity administrator allows or denies registration requests. Once authorized, client applications will receive a client id and secret to authenticate on the entity. To not violate the stateless concept of REST, the secret is formed with a JSON Web Token (JWT) (JONES; BRADLEY; SAKIMURA, 2015) that encapsulates the permissions a client has. As JWT is digitally signed, so a central unity for authentication is not required, provided that all components use the same key that generated the JWT.

Figure 26: REST Components



Source: Based on (FIELDING; TAYLOR, 2000)

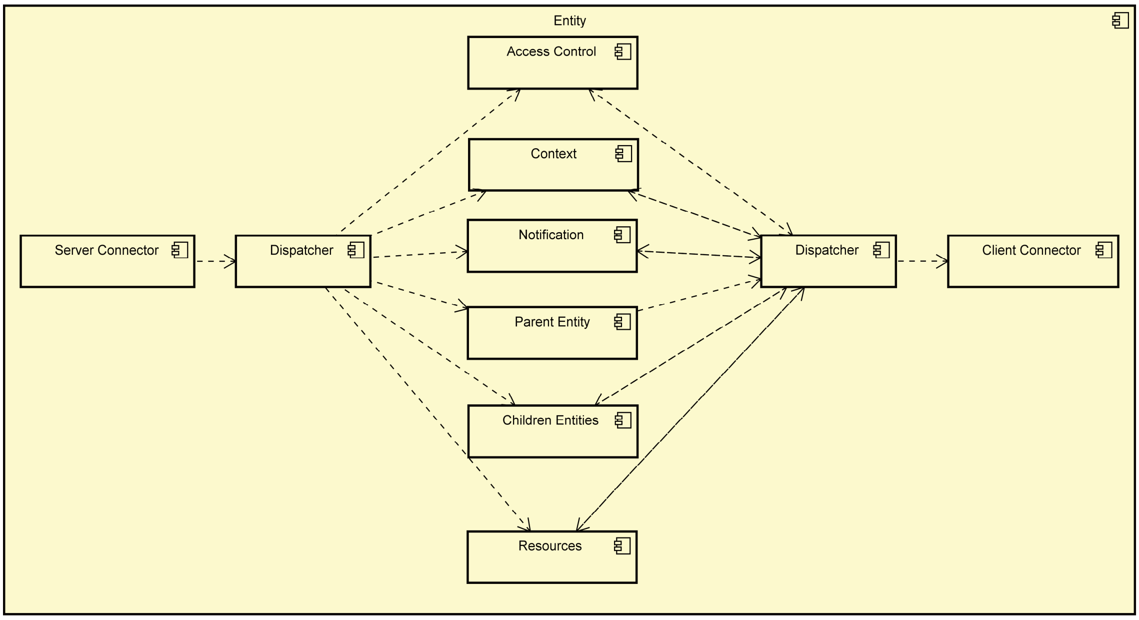
Each entity has a reference (link) to a parent entity. In addition, parent entities have a list of links to entities that are connected to it. These reference types are used so that entities can access the shared features, thus being able to find opportunities of interest. This structure facilitates the hierarchical composition of environments, providing physical integration, since the existing computing infrastructure in one place can be used to improve the quality of use of that environment by someone. For example, a house entity would have entities linked to it like, for example, its inhabitants and objects belonging to it. Moreover, some entities could unlink from a parent entity to become linked to another entity. From the work to the house, for example. The obtaining of the link depends on the implementation of the conceptual architecture, but one possible way to do it is by using service location protocol (GUTTMAN et al., 1999).

In the end, customization is possible by applications that use the afore mentioned components. Applications could follow context updates of an entity to discover new data, or to improve user experience of one place. That is, being context aware.

Figure 26 expresses the REST architectural style components. The rest interface is composed by the verbs HEAD, POST, GET, DELETE, and PUT. The server connector exposes the REST interface, which is then consumed by the client connector. Figures 27 and 28 show the architectural alternatives of the model usage. Figure 27 shows an example a more self-contained implementation. In this case, the entity receives request by the Server Connector (FIELDING; TAYLOR, 2000) which uses the Dispatcher component to delegate the request to the appropriate component. Hence, the dispatcher is responsible for identifying if the request will be redirect to the internal components or if the request will be redirect to remote components using the client connector (FIELDING; TAYLOR, 2000).

The three bottom components from Figure 28 show a distributed example of conceptual architecture usage. In this example the request for each component is redirected to different instances. For example, a phone application redirects the context and resource requests to cloud services, and it is connected to an entity that offers services belonging to a location. Figure 28 also illustrates the hierarchical distribution of the entities, where one entity (Phone App) is connected to a parent, which is connected to another entity. Entities can navigate between them and use services from other entities. This feature enables different types of application like, for example, people and resource location, communication between entities, instant social network formation, localized scalability, and remote monitoring.

Figure 27: Conceptual architecture self-contained implementation

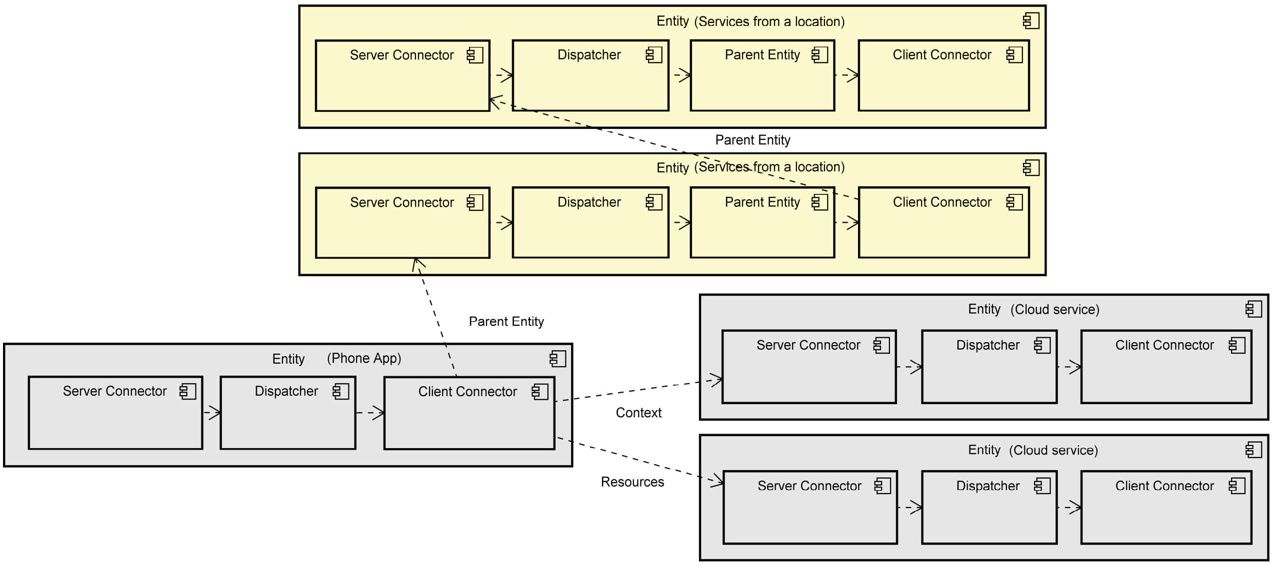


Source: Own authorship

As shown in Figure 29, the Pompilos's conceptual architecture has two categories of services, which use the afore mentioned components, exposing them as interfaces: Pompilos Node (Figure 29b) and Storage (Figure 29a). Thus, these services are able to provide features for access control, resource and context sharing, notification, location, application and node lookups. Both resource and context interfaces allow the query and storage of data, while the context interface allows also context subscription. Hence, these interfaces allow other services or Software Agents to respectively manage access control, store, fetch and receive automatically updates from the data maintained by the Pompilos Nodes. It is important to notice that the Pompilos context interface is a proxy for the Context Aggregator (Figure 30). This guarantees that users' data will be retrieved or stored only with their consent.

Storage service is an extension of the Pompilos Node, being responsible for the management of data repositories, exposing implementations of the context or resource interfaces. Models and Social Network services extend the Storage service definition by exposing new interfaces. Models services expose the register interface that allows the registration of Model Executor agents. Yet, the Social Network service provides interfaces for cloning and removing social network graphs that are used in the training of new knowledge discovery models. The register, clone and unclone interfaces are provided as node resources.

Figure 28: Hierarchical and distributed usage examples



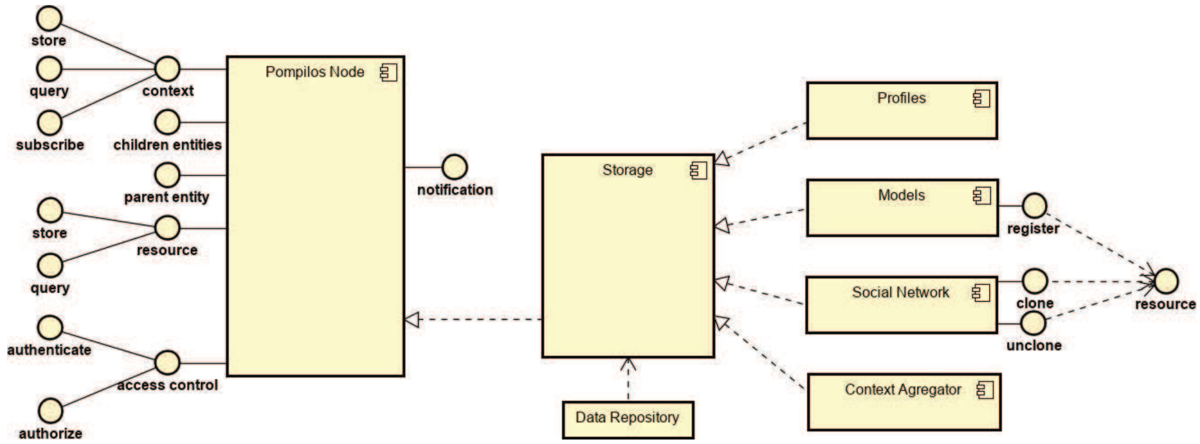
Source: Own authorship

Pompilos Node services are also aware of their environment to actuate when necessary. This is achieved by exposing the notification interface that is used to receive data updates of subscribed interests from Storage services, or other Pompilos Nodes (BIRMAN; JOSEPH, 1987; DEY; ABOWD; SALBER, 2001). The Resource Recommender and Influence Advisor extend the Pompilos Node service by adding a result interface. This interface is invoked by the Model Executor service to deliver results of knowledge discovery model executions requested with the execute interface. Both the result and the execute interface rely on the resource interface from the Pompilos Nodes, as they are offered as resources resident in nodes. Finally, Context Acquisition Agents use the access control interface to negotiate authorizations for collecting contexts on behalf of Pompilos nodes.

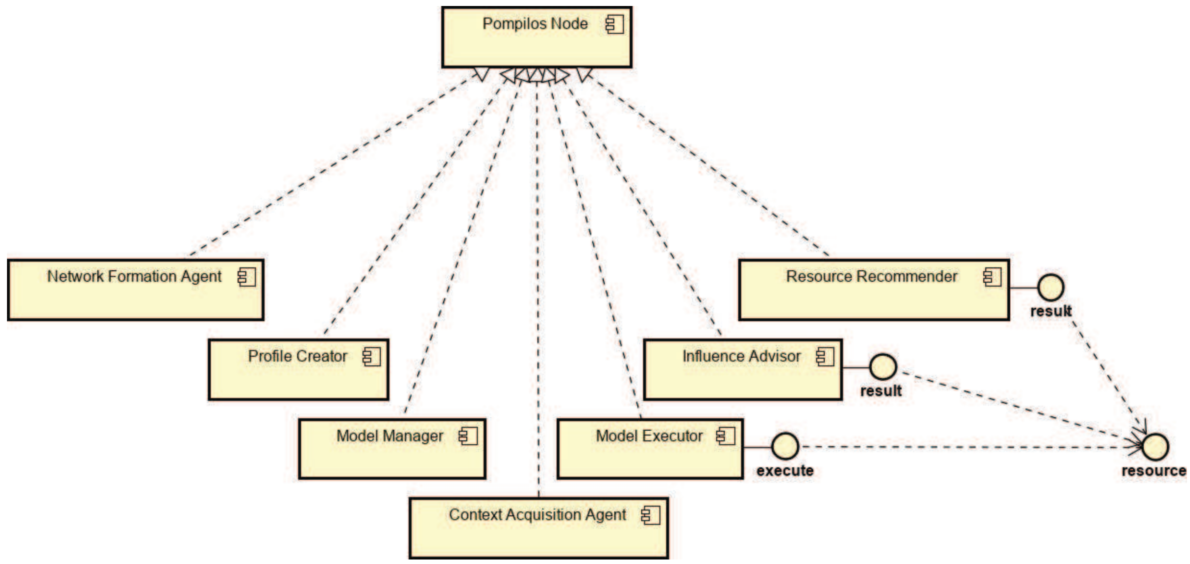
The following use case is presented to illustrate the application of the Pompilos model:

> Given a Twitter user that is willing to have a better food consumption, the Pompilos model could offer him relevant healthy diet messages, as identify which followed account influences the user in food consumption topics. Thus, a Network Formation Agent could use the followed account to form its network. Messages from the followed accounts would be used to create the message topic profile of each followed account. Also, messages to which the user interacted could be used to generate his user profile. Based on the user profile and the message topics profile is possible to identify the profiles that influenced the user the most by message topic. Finally, messages from healthy diets could be recommended to the user by monitoring accounts that publish this type of message.

Figure 29: Conceptual Architecture Specialized Services



(a) Pompilos Storage Services



(b) Software Agents Services

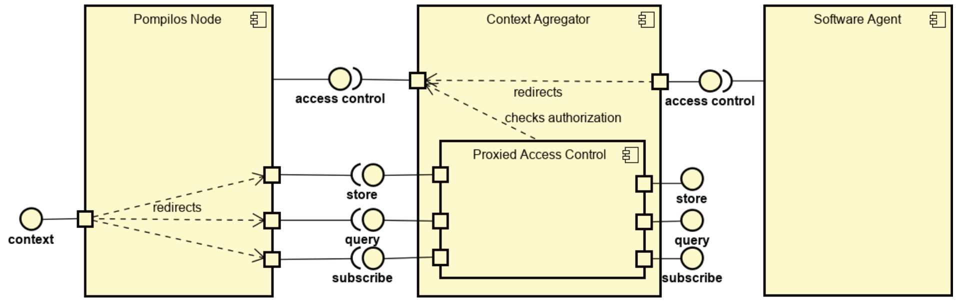
Source: Own authorship

5.3 General Data Schema

Figure 31 shows the model's **general data schema**, represented as UML class diagram (GROUP, 2015). The schema encompasses the concepts regarding the model, which are used in the communication between components. In other words, the data transmitted in components' communication are instances of the schema's classes.

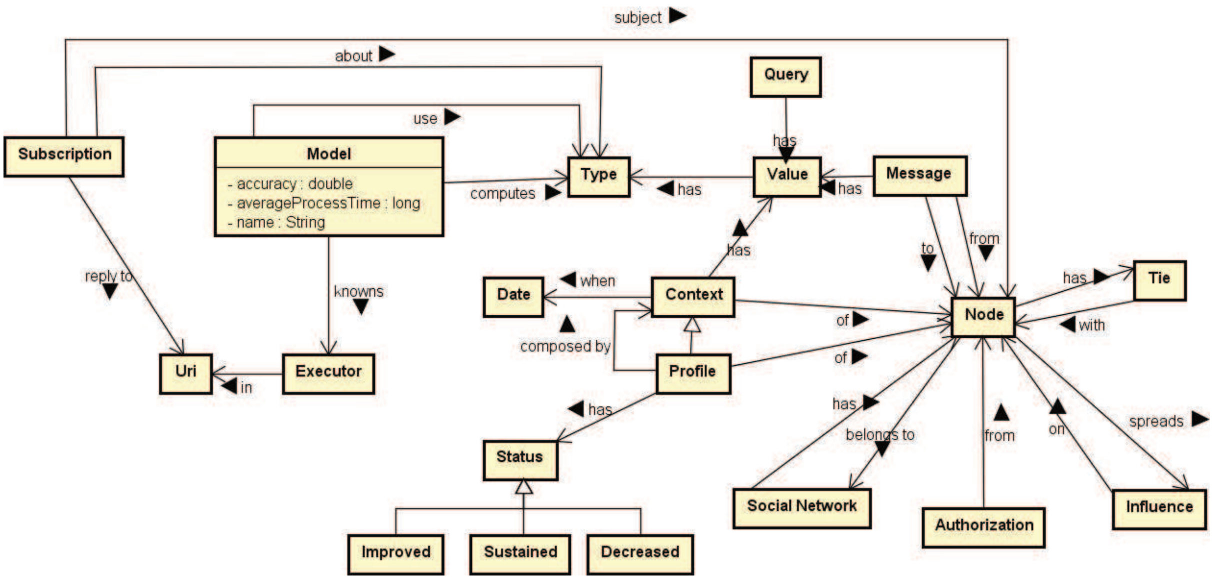
Node is the central class of the schema. It represents any entity that generates contexts, or those entities that are socially connected. Context is used to map those attributes and values that can identify the situation of a Node. In this sense, each context is composed by a Value (which by its turn has a Type) and a Date. Type refers to the value semantics, for example a

Figure 30: Context Proxy



Source: Own authorship

Figure 31: General Data Schema



Source: Own authorship

weight, Value refers to the raw data, for example 176 lbs, and Date refers to when the context was acquired. The Profile class summarizes Nodes behaviors. It extends Context, in order that it has a Value, and a Date, which define, respectively, the semantics, information and generation time of a Profile. Furthermore, the Profile is composed by a collection of Contexts that were used for its generation and a Status indicating if the node has Improved, Sustained or Decreased the status of its behavior compared to an older Profile.

Nodes may authorize the acquisition of contexts by software agents in their behalf, this concept is described by the relation Authorization *from Node*. The Message class denotes the communication between two or more Nodes, and has the relationships from, which describes the origin node of the message, and to, describing the destiny node. Social networks which the Nodes participate are described by classes Social Network and Tie. The former represents the various social groups that a individual may participate, as the later indicates qualified relationships between individuals, i.e., friend, co-worker or sibling.

The Subscription class is used when some software agent needs to subscribe for receiving updates about changes in some Node's attribute, and is used in communications with the subscribed interface of Storage components. The update data will be send to the Uri informed in the subscription, whenever a change is detected on the subject. Furthermore, Query is used in communications with query interfaces. Instances of this class must define the value of query that will be used as instances of Value class, which have a type denoting the query type, for example SPARQL or SQL. Models for computation of social influence in health are described by the Model class. The relation computes Type of this class indicates the health variables that the model is able to detect influence, for example, obesity. Finally, the relation *knows Executor* indicates those known components which are able to compute the Influence that one Node spreads over other, using the described model.

5.4 Conclusions about Pompilos Model

This chapter presented an architectural model for detecting the influence of nodes on the health of others. Applications can be built based on this model to aware influencer nodes about how their behavior impacts others, providing a way for them to enhancement the content they share. Also, influenced nodes could benefit from the model as they will be receiving content with continuous increased quality. The next chapter will present the application of the proposed model and its ontology to support the preventive care of NCDs.

6 INTEGRATING POMPILOS WITH TWITTER FOR PROMOTING NON-COMMUNICABLE DISEASES PREVENTION

This Chapter exposes the strategies used to evaluate the possibility of solving the research question defined in Chapter 1 with the deployment of the conceptual model elements and architecture presented in Chapter 5. In essence, the goal is to apply ubiquitous computing as way to stimulate social support on NCDs prevention for the improvement of engagement in care activities. Therefore, the evaluation took the following two use cases:

- The users of an online health assistant application will receive NCDs preventive messages collected from the Twitter[1] social media to foment NCDs prevention behavior
- Twitter profiles, who are used to collect messages to be forwarded to users, will be notified about a rank which could be used to check their influence in NCDs prevention behavior of users

This chapter explains the implementation of a prototype of the conceptual architecture and the conceptual model elements. Also, this chapter explains how the model elements were technically used to materialize the two early mentioned use cases. Analysis of the data captured in experiments was also done in this chapter to determine if there was a behavior change in users, what were their feelings in using the application and if the Twitter profiles have changed their attitude after being aware of their influence in the users' health. The Pompilos model was used in two experiments run in different periods, in order to assess the engagement of the application's users and the behavior of social network profiles in regards to the posting of messages related to non-communicable diseases prevention.

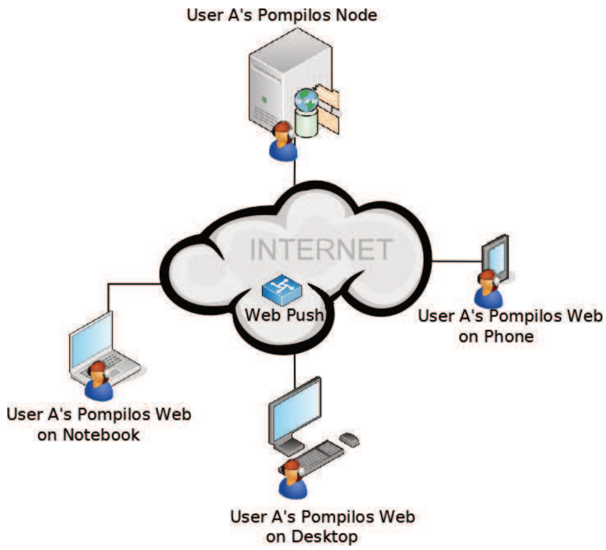
It is worth citing that the experiments needed an appreciation of the Research Ethics Committee of Unisinos, once the research involved humans and its thematic area was oriented to the development of health inputs. The Research Ethics Committee is a collegiate body with the objective of ensuring compliance with ethical principles in research involving human beings. The research was registered in the Brazil Platform[2] with the title "My U'Ductor: Assistant for the prevention of risk factors for chronic non-communicable diseases with automatic social support", where were attached the descriptive text of the project, informed consent form, and the authorization document of the coordinator of the Graduate Program in Applied Computing. Once the research project was completed, it was submitted to the ethics committee for approval, and was approved on October 3rd, 2018, under No. 97397118.8.0000.5344.

Finally, this chapter ends by explaining how location data were used to infer real social networks and how the ontology, presented in Chapter 4, can be used to know the influence received by nodes of a social network generated from the data collected in the application usage of the first experiment.

[1]Twitter was used by the easiness of acquiring messages through its application programming interface. Other social media platforms could be used, as long they allow to acquire data from the profiles.

[2]http://aplicacao.saude.gov.br/plataformabrasil/

Figure 32: Pompilos Architectural Components Example



Source: Own authorship

6.1 Implementation Aspects

Before evaluating the model's applicability in changing NCDs prevention behavior, a prototype with the core features of the architecture was built. This prototype was responsible for managing contexts and resources, was able to notify contextual changes, and could guarantee authorized access to data. These features allowed the creation of services for the generation of aggregated contexts, social network generation, resources recommendation, and influence detection that are expected by the conceptual model.

This section first introduces the implementation aspects of the architecture core features and then approaches the modeling of the model elements in the form of twitter messages recommendation, real social network formation and influence assessment. It then explains how these parts are assembled to provide an online health assistant application for NCDs prevention.

6.1.1 Architectural Aspects

Figure 32 presents the three main components of the prototype: **Pompilos Node**, **Pompilos Web** and **Web Push**. The division between Pompilos Web and Pompilos Node is important, as the Pompilos Web may have different instances and be distributed on different platforms (televisions, phones, tablets or desktops). However, the web applications are linked to just one Pompilos Node representing their entity (a person, a place or a thing).

The Pompilos **architectural stack**, implemented by the prototype used to evaluate this thesis's hypothesis, is shown in Figure 33. The Pompilos Node represents an entity managing the model core features (authorization, contexts, resources and notification management) and providing services for other software agents to interact with its contexts and resources and to be aware of the entity's context changes. So, this component is composed by six other parts. The **HTTP Server** provides a uniform interface enabling software agents to access the **Pompilos API** which exposes the entity's contexts, resources and also enables the notification about events or context changes occurred in other nodes. That, by its turns, delegates the messages to each other component: **Authorization Manager**, **Context Manager**, **Resource Manager**, and **Notification Manager**.

The Authorization Manager is responsible for authenticating software agents by using the OAuth 2 protocol (HARDT, 2012) and to authorize the access to one entity's context, resources or notification by checking the validity of the JWT Token (JONES; BRADLEY; SAKIMURA, 2015) issued by the software agent. The Context Manager and Resources Manager store their data in a PostgreSQL database. Each context or resource is formed by a custom mime type (FREED; BORENSTEIN, 1996) and a document in the JSON format (INTERNATIONAL, 2017). Each new context or each change in a resource causes an invocation of the Notification Manager which publishes these events using the **Notification BUS**. The Notification BUS uses the Redis Pub/Sub implementation[3] to notify the changes to software agents that have previously subscribed to the events using the Context component subscription interface. Pompilos Web is also notified about these events by **Web Push** notifications, so the web push client keys are stored in a PostgreSQL database. Finally, the Notification Manager enables theregistration of **endpoints** that will handle the notifications in the server.

The Pompilos Web is a web application linked to one Pompilos Node, in this way multi- ple web applications instances linked to one node may exist. The **application** is implemen- ted using W3C standards technologies as ECMA Script (INTERNATIONAL, 2018), Cascade Style Sheets (CSS) (JR.; ETEMAD; RIVOAL, 2018) and Hypertext Markup Language (HTML) (FAULKNER et al., 2018). The **Client Authentication Manager** is used to establish an identifiable connection between the application and its node. **Context Local Manager** and **Resource Local Manager** replicates the Context and Resource Global Store from the nodes by storing the data in IndexedDB repositories (ALABBAS; BELL, 2018). The **Synchronization Manager** is used to keep the local stores up to date by receiving data web push notification updates from **Notification Endpoint** or by pushing the data to the node as it is saved. If no Internet connection is detected, the data will be kept locally and upload once the connection is regained.

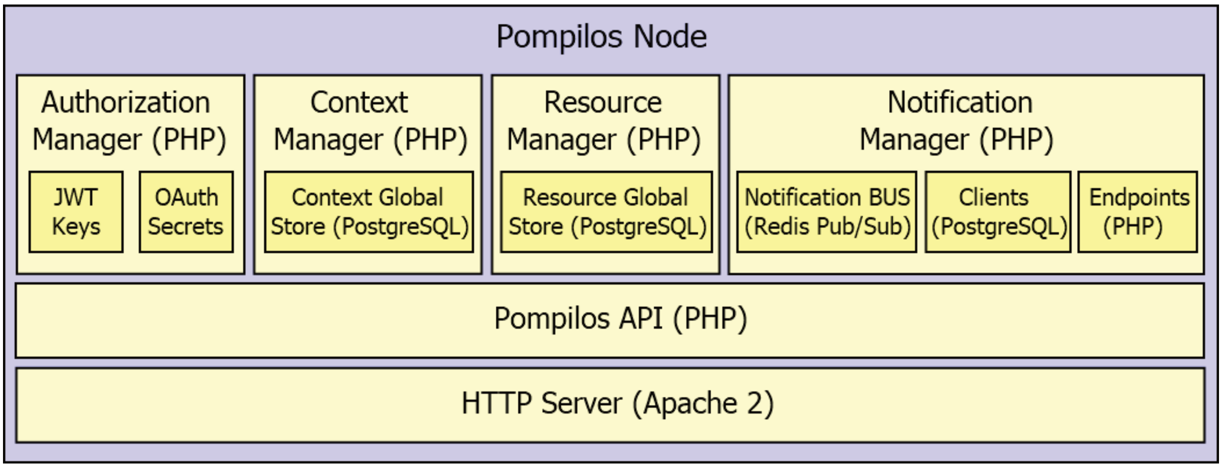
6.1.2 Detecting Non-Communicable Diseases Prevention Messages

Profiles related to physical activity, healthy eating, and weight control subjects were identified in the Twitter social network before recommending messages regarding these topics. Figure 34 synthesizes the process used to select these profiles.

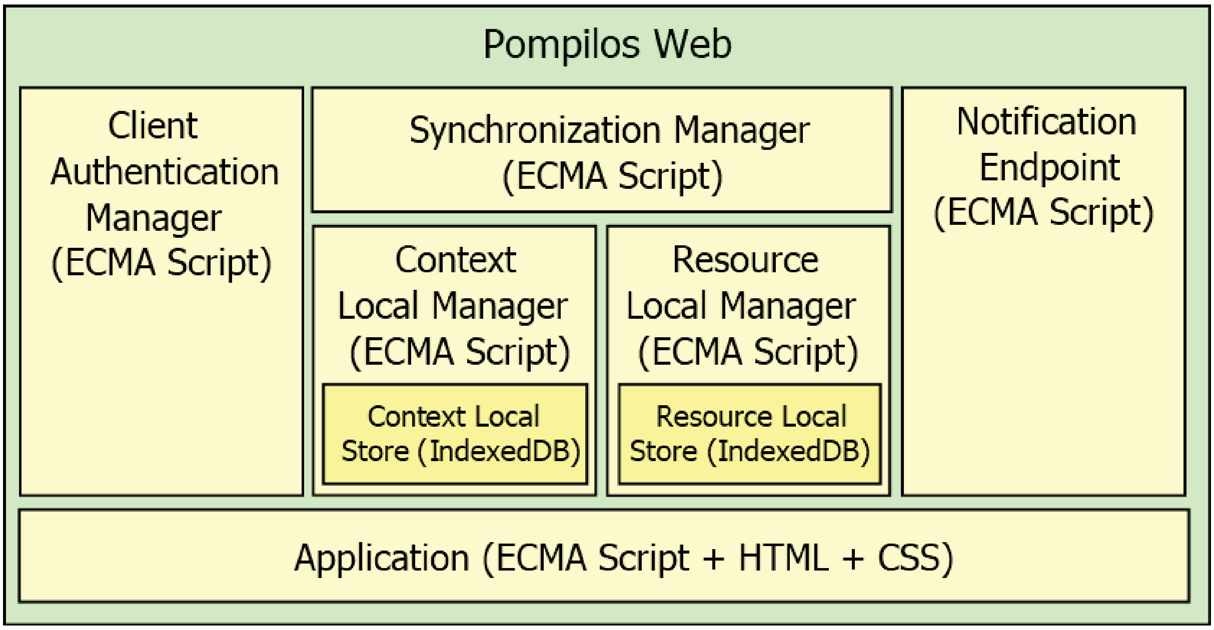
First, a seed profile was defined to be examined. To do so, the profile of the ministry of health (@minsaude) was chosen. Then a review of the messages sent by the profile related to

[3]https://redis.io/topics/pubsub

Figure 33: Pompilos Architectural Stack



(a) Pompilos Architectural Stack - Server

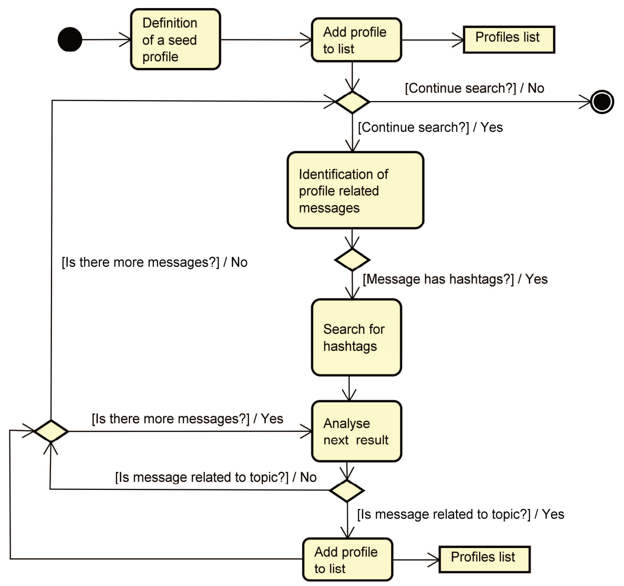


(b) Pompilos Architectural Stack - Web

Source: Own authorship

healthy eating and physical activity was reviewed. If the related message had hashtags, these hashtags were used to search for Twitter messages to identify more related profiles. Hence, the following hashtags were used: #saudenacozinha (health in the kitchen), #blogdasaude (health blog), #vidasaudavel (healthy life), #dicasparaemagrecer (weight loss tips), #alimentacaosaudavel (healthy eating) and #emagrecersaudavel (healthy weight loss). The process stopped when 15 profiles were found due to management reasons, as more profiles would overweight the communication with these profiles. In the end, the following profiles were selected: @BuscarSaude, @saudavelebarato, @realfoodbrasil, @rederms, @saudavelcomida, @BlogSaudeBetter, @WorldBoaForma, @fitcomdaniele, @CamilaGramelick, @DetoxReceitas, @opiniaomedica, @juntoemisturadi, @belagil, @SES_RS, @SaudeMG, and @minsaude.

Figure 34: Data Acquisition Process



Source: Own authorship

Table 11: Training Performance

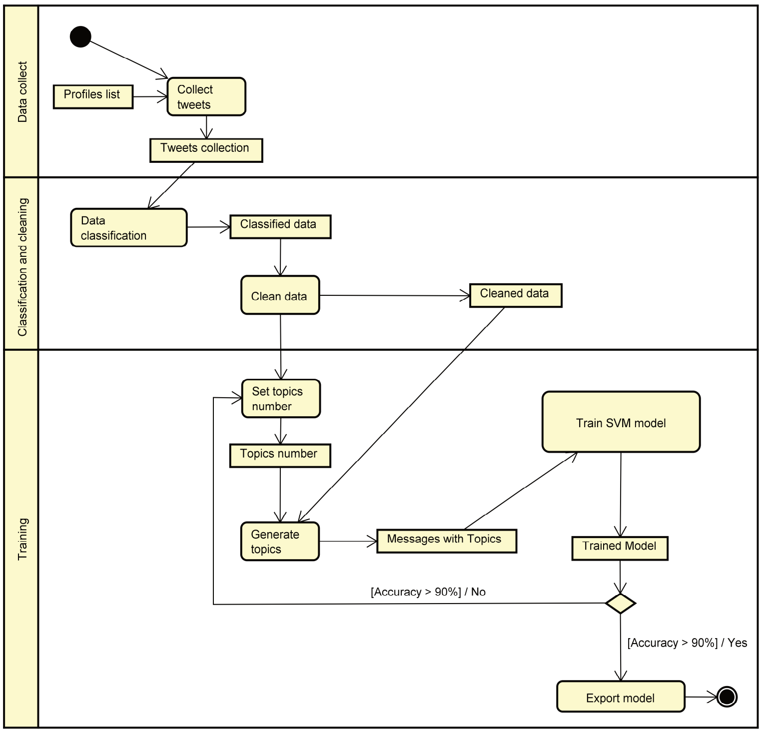
Number of topics	Confusion Matrix				Accuracy	Precision	Recall	F1 Score
	Positive		Negative					
	True	False	True	False				
32	340	39	3,379	989	78.34%	89,71%	25.58%	39.81%
64	811	145	3,273	518	86.03%	84,83%	61.02%	70.98%
96	979	92	3,326	350	90.69%	91,41%	73.66%	81.58%

Source: Own authorship

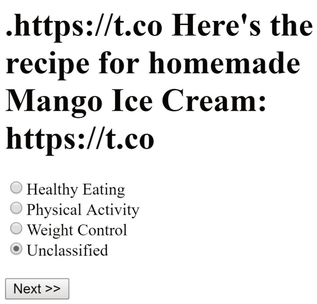
Figure 35a shows the activities done in training the model for detection of NCDs preventive messages. First, all messages sent by the selected profiles with up to two years of publication were registered in a database. In the end, 24,971 messages were stored in the **tweets collection**. A **data classification** process was then started. 7,040 from the tweets collection were classified as, “Healthy Eating”, “Physical Activity”, “Weight Control” or “Unclassified”. To perform the classification a simple web page application was made available (Figure 35b). After classification, the data were cleaned for impurity removal like stop-words, URLs, hash symbols (#) and duplicate messages. In the end, 4,747 remained for classification.

The model training consisted of two main activities: **topics generation** and **model training**. Given that each tweet can be understood as being composed by a mixture of topics, and that each word from that tweet has a probability of pertaining to a topic, the Latent Dirichlet allocation algorithm was used to map the topical probability of each tweet from the tweets collection (ROBINSON; SILGE, 2017). The topical probability of each tweet was used as features to train a Support Vector Machine (SVM) model for classifying the tweets as NCDs prevention messages (healthy eating, physical activity practice or weight control) or not related with NCDs prevention (CHANG; LIN, 2011). The number of generate topics was increased until an accuracy greater than 90% was obtained. This was reached when the SVM model was trained with 96 topics. The training performance by number of topics is summarized in Figure 35c and is detailed in Table 11.

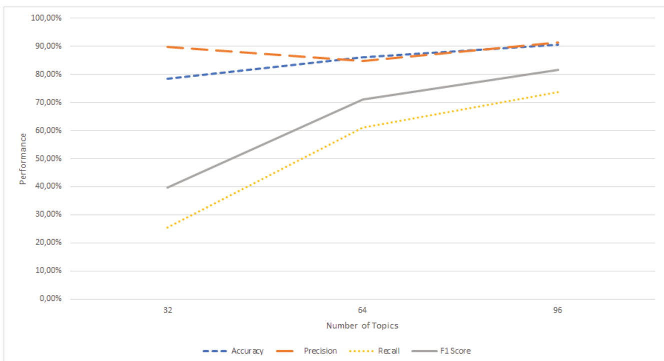
Figure 35: Training Process and Performance



(a) Training Process



(b) Message Tagger



(c) Training Performance

Source: Own authorship

In regards to generate resource recommendations to users, three independent services were created following the Pompilos model definition, as shown in Figure 36a. The **Twitter Monitor** is a Context Acquisition Agent responsible for monitoring the activity of the selected Twitter profiles and saving them as contexts. The **Health Message Classifier** is a Model Executor that stores the trained model and the classification app, enabling the classification of messages and the improvement of the model. The **Health Message Checker** implements the Context Acquisition Agent, Profile Creator and Resource Recommender services definition. That service subscribe for message contexts from Pompilos nodes representing Twitter profiles. Each new message from a profile is checked and added in the moderation list for further approval if classified as a NCD prevention message (Figure 36b). Moderation is desired to prevent the sending of false positives to users, and to prevent that users will not receive messages that stimulate the worship of body image (SMAHEL; ELAVSKY; MACHACKOVA, 2017).

The sequence diagram presented in Figure 36c is an example that better illustrate these dynamics. First, the **System Administrator** uploads the trained model to the Health Message Checker service. This service then subscribes for contexts of type message from the @minsaude profile's Pompilos node[4]. Following the Twitter Monitor uses the **Twitter Stream Api**[5] to monitor messages from selected Twitter profiles. Each new message from the @minsaude profile is added as new context in their representing Pompilos node which notifies all its subscribers about the new context. When the Health Message Checker receives the message context it invokes the Health Message Classifier to check if the message is a NCD preventive message. In positive case, the Health Message Checker will notify the System Administrator about a message waiting for release. By releasing it, the message is delivered to the application users.

6.1.3 My U'Ductor, an Assistant for Non-Communicable Diseases Prevention

My U'Ductor is a mobile and online assistant for diets, weight management and physical activity practice (which are the tasks needed to prevent and control risk factors for NCDs) that implements the Pompilos Web architectural stack and is connected to a Pompilos Node. Like any other assistants, it allowed the scheduling and recording of care activities, as well as allowing the user to visualize the progress of their care. However, different from the existing care applications, My U'Ductor collects, in social networks, messages related to the practice of physical activities and healthy eating. Such messages are presented to users as a form of social support, which is recognized as a motivational factor for reinforcing user engagement in their care activities.

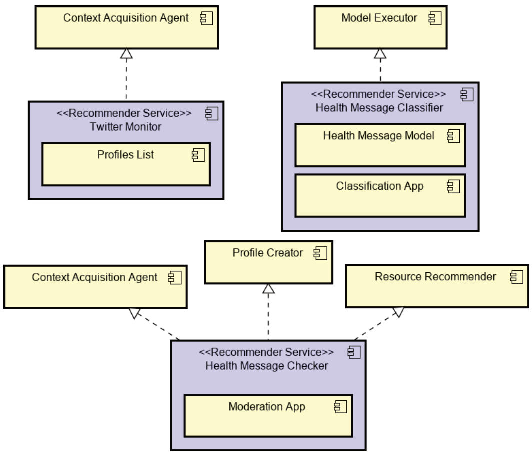
My U'Ductor was publicly available at https://app1.uductor.com. To access the application, the users should have a Google account. This restriction was made to ensure the authenticity of the users' e-mail address since some of the communications the application were made through electronic messages.

The first time the users authenticated, the application displayed the free and informed consent form (Figure 37a). In this form, the users could also configure the notifications receiving

[4]As in the prototyped implementation all nodes shared the same server, services subscribed for contexts and resources notification globally. That is, they issued their subscription for a particular context type or resource once, instead of creating a subscription for each new node on the system.

[5]Consuming streaming data - https://developer.twitter.com/en/docs/tutorials/consuming-streaming-data.html

Figure 36: Recommendation Process

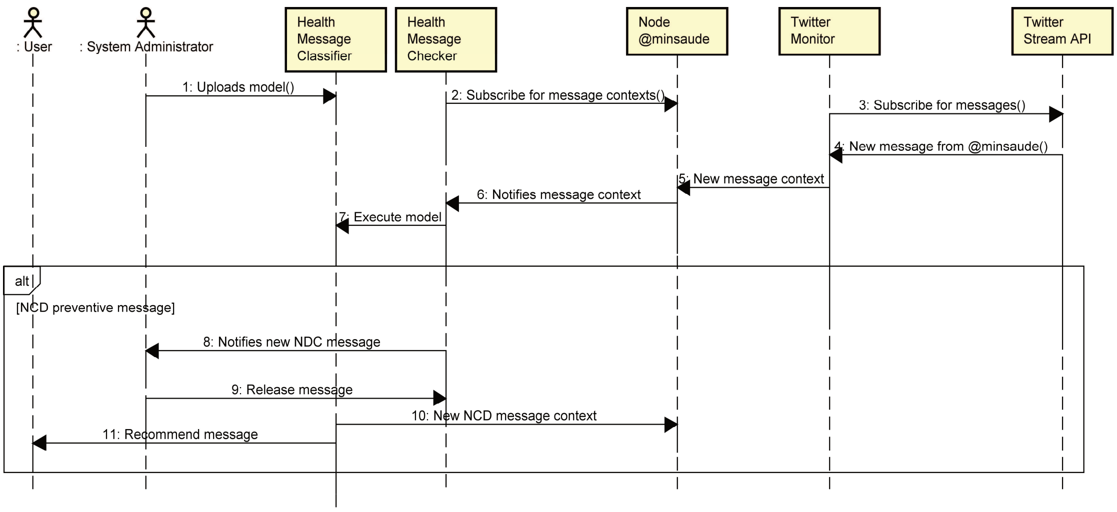


(a) Recommendation Services

Messages waiting to be released

Profile	Message	Action
twitter:opiniaomedica	It's simple (maybe difficult is "falling apart") to know what healthy eating is. Help spread this video (on Facebook, Whatsapp, etc.). Copy this link and pass it to your friends:https://t.co/tvUsgRo7dF https://t.co/L4JaYIPOHI	Release Repeated No quality Delete
twitter:BuscarSaude	When the exercises are well done, the results are very favorable. What are the best exercises for glutes? How to do step up? How to do squats? https://t.co/wZsKpAyD34	Release Repeated No quality Delete

(b) Waiting Messages



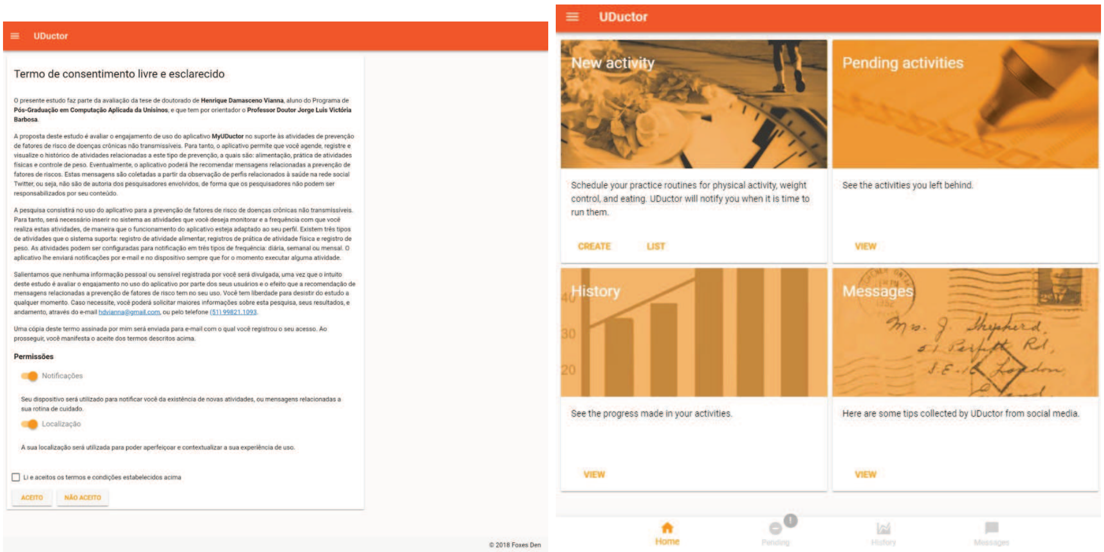
(c) Recommendation Sequence Example

Source: Own authorship

permissions and access to location data. The use of the application was only allowed after the users checked the option “I have read and I accept the terms and conditions set forth above” and pressed the “I accept” button. In this way, the free and informed consent term was accepted electronically and online, allowing users from different locations to participate in the experiment. A copy of the informed consent form signed by the person in charge of the research was sent by e-mail to the users after the acceptance of the term in the application.

After accepting the informed consent form, users were directed to the main application screen (Figure 37b). This screen described all the features in the application, as well as provided links so that users could access these actions. Optionally users could access these same features in the options menu or by the shortcut bar. The application provided five main features: **New Activity**, **Activity List**, **Pending Activities**, **History**, and **Messages**.

Figure 37: Consent Form and Application main screen



(a) Consent Form (b) Application main screen

Source: Own authorship

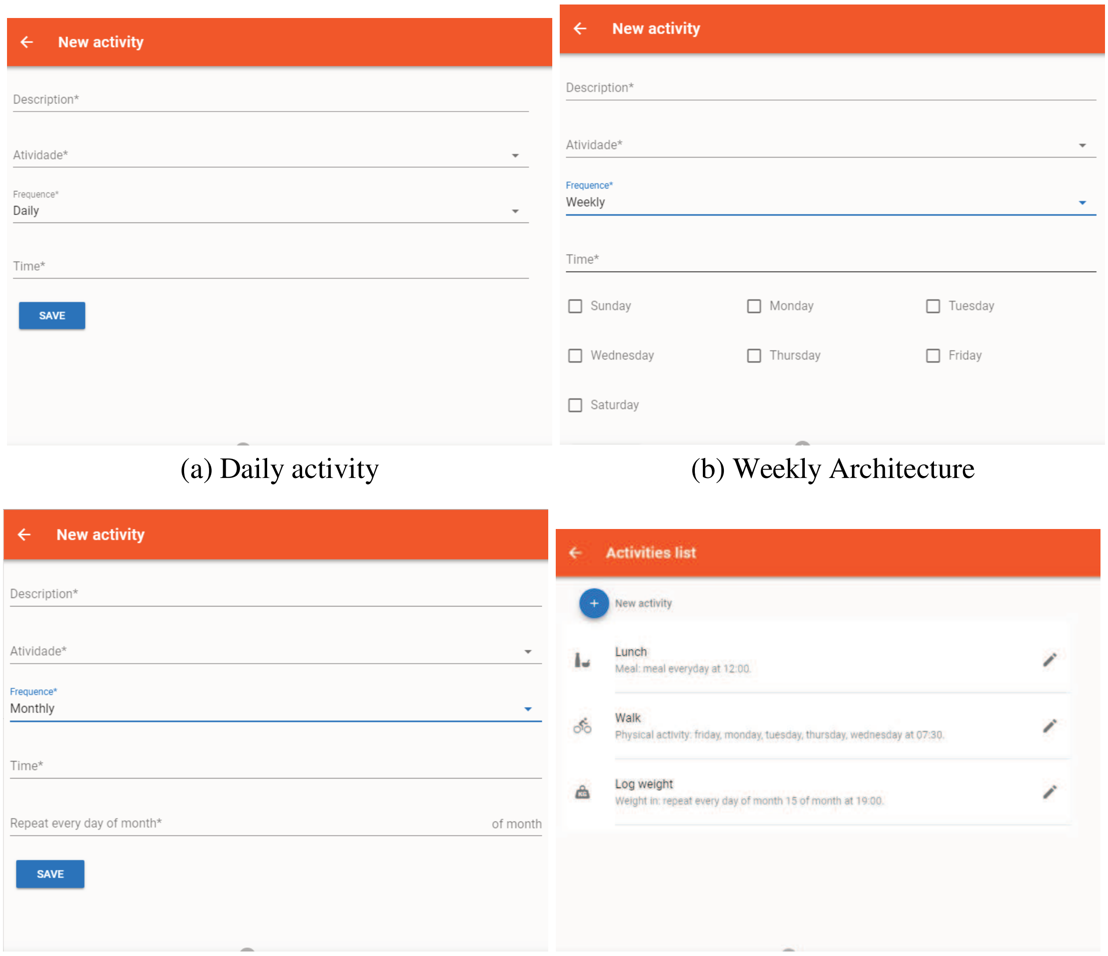
The “New activity” feature allowed the users to define the recurrence with which they would be advised of the need to carry out their activities. There were three types of possible activities, all related to the prevention of risks of noncommunicable diseases: **Physical activity**, **Meal**, and **Weight**. Users could configure activities to be performed daily in a certain hour (Figure 38a), weekly with a repetition rate per day of the week (Figure 38b), or monthly repetition with a repetition rate per day of the month (Figure 38c). In turn, the “Activity List” feature allowed the users to edit the recurrence of previously registered activities (Figure 38d).

The users received reminders for those activities that they had scheduled. Reminders for activities that the user was required to perform were emailed to them (Figure 39a). Optionally, users were able to receive notifications via smartphone as long as the permission to receive notifications had been accepted by the users (Figure 39b).

Activities that the user did not register appeared in the “Pending Activities” action (Fi- gure 38d). When users clicked on any items in the pending activities list, the application dis- played a report form according to the type of activity. The “Physical activity report” allowed the users to record what types of activity they performed, the pace of these activities, and for how long time the activities were performed (Figure 40b). For example, in the same report the user could indicate that he walked for 20 minutes at a moderate pace and ran for 10 minutes at a low pace.

The “Weight Report” allowed the users to enter their height and weight (Figure 40c), in this way it was possible to calculate the user’s minimum and maximum weight limits (WHO, 2018a). The “Meal Report” allowed the users to enter the portions that they consumed in

Figure 38: Activities editing

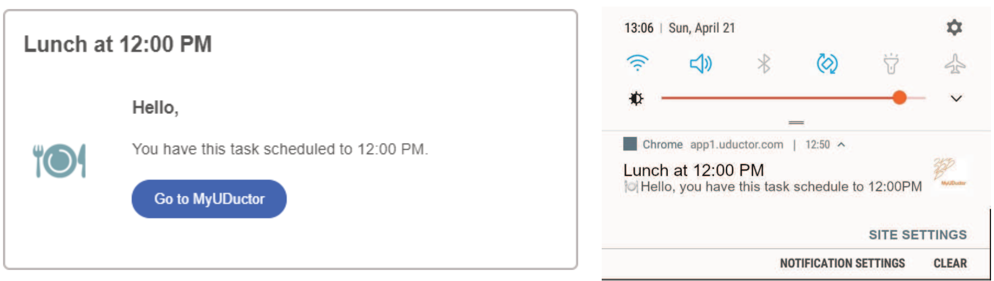


(a) Daily activity (b) Weekly Architecture

(c) Monthly monthly (d) Activities list

Source: Own authorship

Figure 39: Activities Notification



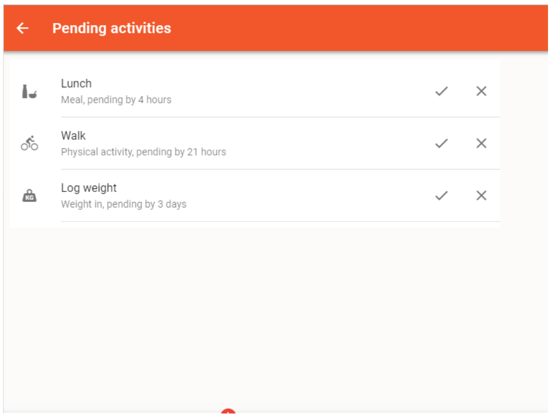
(a) Mail Notification (b) Activity Notification

Source: Own authorship

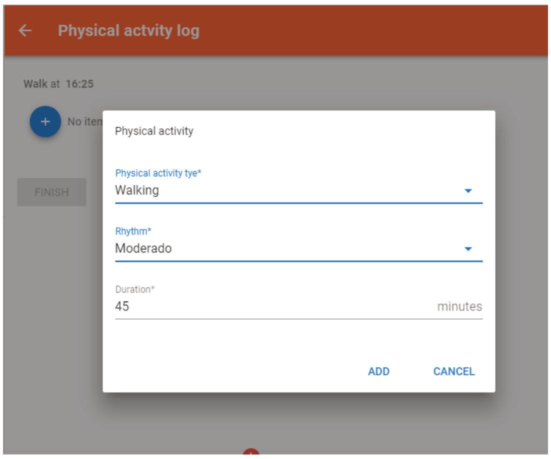
the meal for each food category (Figure 40d). When inserting a food portion, the application indicated the appropriate number of portions per day and presented some portion equivalences for some types of foods according to the one proposed by Philippi et. al (1999).

The “History” feature allowed the users to graphically view records made for each type of

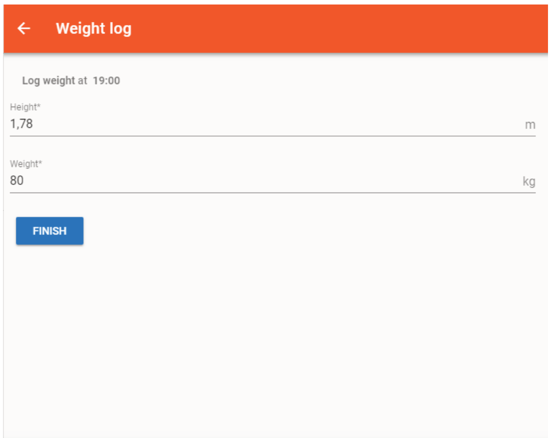
Figure 40: Pending activities



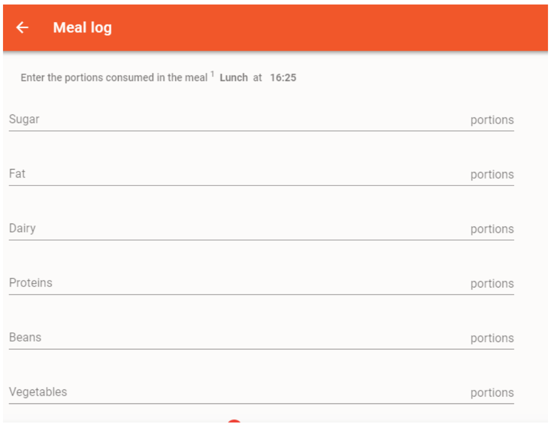
(a) Pending activities



(b) Pending physical activity log



(c) Pending weight log



(d) Pending meal log

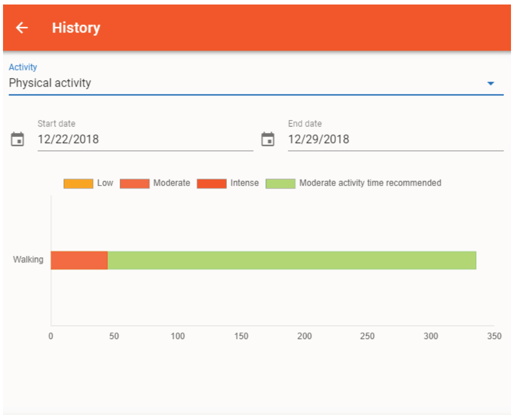
Source: Own authorship

activity in a period. The history of physical activity (Figure 41a) indicated the activity summary performed by the user and the green-time recommended activity time for adults by (WHO, 2018b). The weight history (Figure 41b) presented, in addition to the user's weight records, the minimum and maximum limits according to their height (WHO, 2018a). The meal his- tory (Figure 41c) presented the summary of portions consumed by the user by type and the recommended amount for the same period, as proposed by (PHILIPPI et al., 1999).

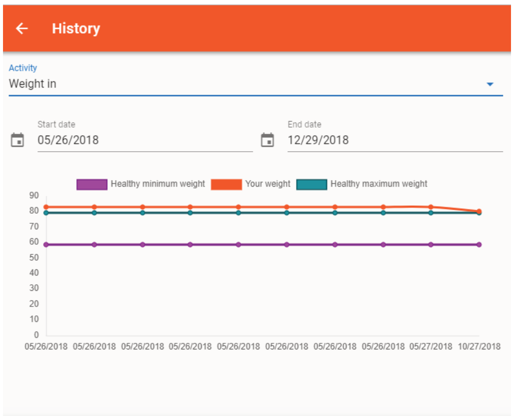
The application notified its users whenever a new message relevant to the prevention of ch- ronic disease risk factors was identified. The list of recommended messages could be accessed by the user through the "Messages" action (Figure 42). Users had the possibility to "like" or go to the source of each message received.

Figure 43a highlights the components that participate in the realization of the My U'Ductor Application features. **My U'Ductor App** is properly the software in which the users interact and is executed usually in a web browser. Each My U'Ductor Application instance is connected with the Pompilos Node that represents the user in order to synchronize its contexts and resources.

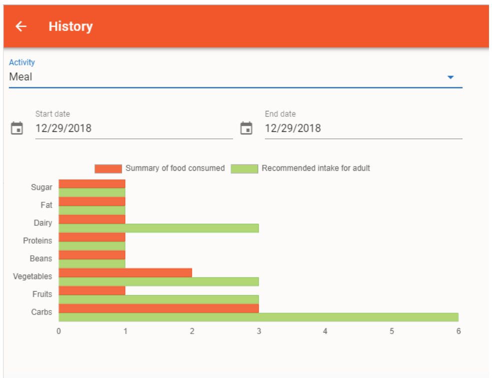
Figure 41: History



(a) Physical activity history



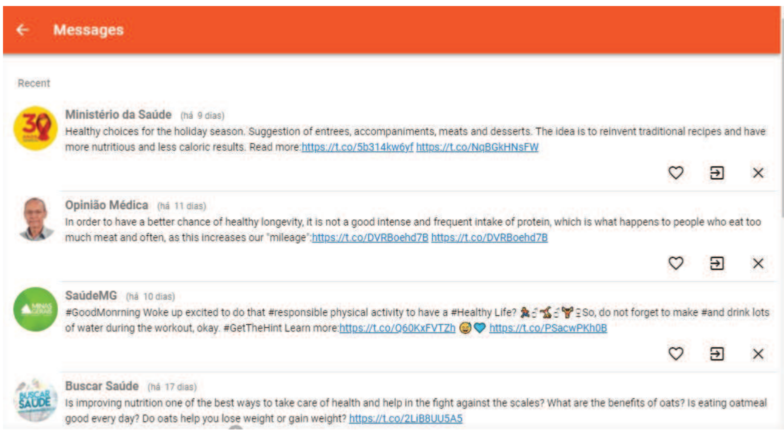
(b) Weight history



(c) Meal history

Source: Own authorship

Figure 42: Messages



Source: Own authorship

In My U'Ductor Application the care plan, the pending activities and the received messages are treated as resources. Log activities such as physical activity, weight and diet, and other user actions such as like in messages or links followed are contexts. The main distinction between resources and contexts resides in their application. Contexts are pieces of information used to compute states of a entity (DEY; ABOWD; SALBER, 2001), while resources are self-explained pieces of information. For example, an image is a resource, as its self-explained, but the act of generating this image or changing one of its attributes is represented with a context.

The **My U'Ductor Node** extends Pompilos Node and it is responsible for the management of particularities of the My U'Ductor App, such as registering users in a specific research group (i.e. intervention or control) and managing acceptance terms. **Planner** is responsible for scheduling the reminders of activities that must be executed by the users in the future like, weight in, practice physical activity or eat. The schedule management was done with aid of the Agenda job schedule library[6]. Finally, the **Activity Mailer** is responsible for sending e-mail notifications of activities that must be executed to users.

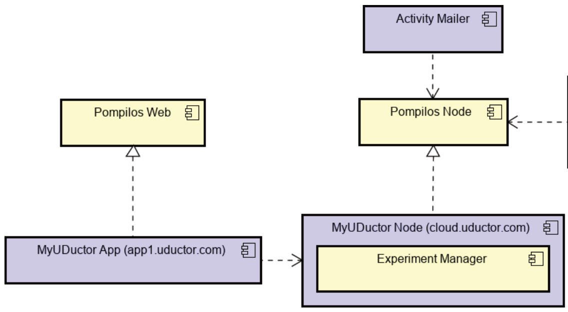
Figure 43b shows an example of how the components may interact for the registering of new plans and to notify users about activities that must be completed. The recording of log activities and interaction with messages is omitted, but it will follow a similar approach. First the user request the application that is downloaded and executed in a browser. To access the application the user first needs to authenticate. The authentication is managed by the My U'Ductor Node, which register the user node and send a **App Token** to the web application. Concomitantly, the Activity Mailer **subscribes for activity contexts updates** and the Planner subscribes for plans updates. Then, the user creates a plan, like for example, "walk every day at 10AM". The plan is sent by the My U'Ductor app to the **Users Node** which notifies the Planner about this new plan. The planner schedules a job representing the plan to be executed. When the job runs, it **creates an activity** related to plan in the User Node, that notifies the My U'Ductor App via web push protocol, and the Activity Mailer via Redis Pub/Sub, which respectively alerts the user and send an e-mail to him.

6.1.4 Assessing Influence from Twitter's Profiles

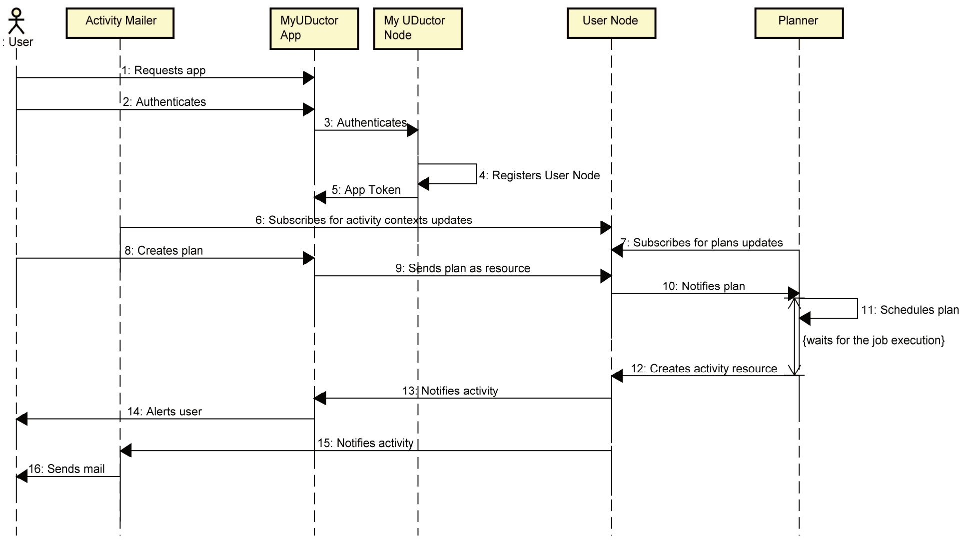
A rank was designed to present the influence that the monitored Twitter's Profiles yield on the application users. The score of each profile took into account a variation of a Partial Credits with Bernoulli distribution (GOYAL; BONCHI; LAKSHMANAN, 2010) of users' engagement in terms of likes, followed links and the use of application's logging features (diet, physical activity or weight), in a 24 hours window by the intervention users after receiving the message. The equation used to calculate the score is presented in 6.1 and its parts are explained in the glossary Table 12. As expressed in the equation, engagement of users in the application was only considered if they interacted with the message by liking it or following its link.

[6]https://github.com/agenda/agenda

Figure 43: My U'Ductor Application Components and Activation and Plan Creation Sequence



(a) My U'Ductor Application Components



(b) My U'Ductor Application Activation and Plan Creation Sequence Example

Source: Own authorship

$$Score_p = \frac{\sum_{i=1}^{M(p)} \sum_{u=1}^{U} (L(p,i,u) + LF(p,i,u)) * (L(p,i,u) + LF(p,i,u) + E(p,i,u))}{\sum_{i=1}^{M(p)} 1} \tag{6.1}$$

Source: Own authorship

A service was created to compute the rank of the Twitter profiles (Figure 44a). The **Rank Manager** service implements the Model Executor and Profile Creator Pompilos services definitions. This service has three internals components: **Influence Equation**, **Rank** and **Rank Web Page**. The Influence Equation is invoked to calculate the rank of the monitored pro- files based on the event subscribed from the application users which is stored and managed by the Rank component. The Rank Web Page is a publicly available webpage located at https://app1.uductor.com/rank that presents the calculated rank for the monitored Twitter profiles (Figure 44b).

Table 12: Glossary of the score equation

Element	Description
p	Profile Index
Score p	Calculated score for p
M(p)	Collection of p messages
i	Message index
U	Users collection
u	User index
L	Identity function that returns 1 if u has generated a like in i of p
LF	Identity function that returns 1 if "u"has followed a link presented in "i"of "p
E	Function that returns the number of interactions with the application during the 24 hours after u has received the message i of p

Source: Own authorship

Figure 44c presents an example of the interaction of the Rank Manager and other components of the model. First the the Rank Manager will start to follow updates of like and links followed in messages from the Twitter profiles, as well to actions of engagement executed in the application from all application users. Whenever the Rank Manager receives a context of those types it will recalculate the score of each Twitter profile, update the rank and finally update the rank in the Twitter profile node.

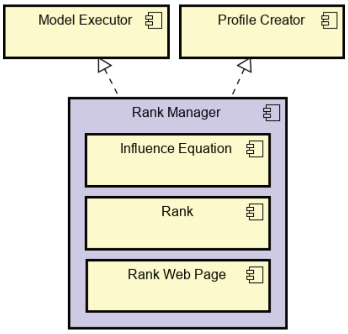
6.1.5 Automatic Formation of Real Social Networks

Real social network were formed with aid of hierarchical clustering (MüLLNER, 2013). This strategy was used to group sites that are located nearby each other. Global Position System (GPS) coordinates were used as input to form new locations, which were considered by the group of GPS points distant 25 meters from a common center. Two individuals were considered to know each other if their presence is overlapped in space and time. In other words, if they were in the same group of GPS points sharing an intersection of time between the minimum and maximum hours of presence in a place in a day.

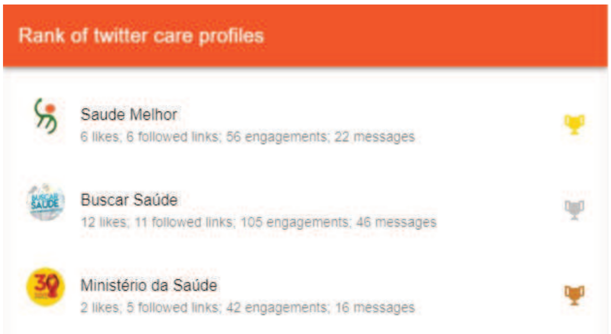
The **Real Social Network Formation** service, Figure 45a, realizes the Context Acquisition Agent and Network Formation Agent interfaces. In this way, this service is responsible for creating social contacts for users nodes and for managing the real social network of all nodes. The **Scheduler**, **Location Clustering** and **Collocation Storage** components collaborate to create new location clusters (Figure 45b). The Scheduler component is setup to trigger the forma- tion of new cluster locations in the Location Clustering component once a day. The generated locations are then stored in the Collocation Storage.

Real social contacts are generated when the Real Network Formation Service receives a new location context from (Figure 45b). The received context is forwarded to the **Location Recorded** which stores the location in the Collocation Storage. Then, the **Collocation Detector**

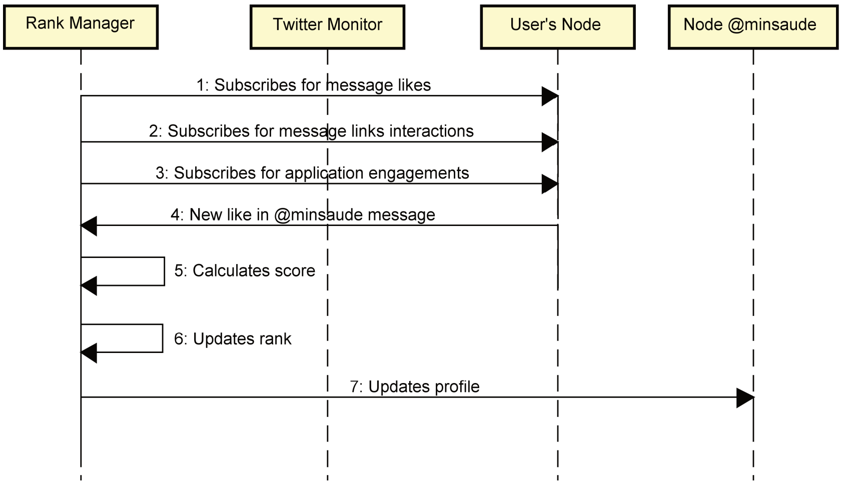
Figure 44: Influence Assessment Process



(a) Influence Assessment Services



(b) Profiles Rank Web Page



(c) Influence Assessment Sequence Example

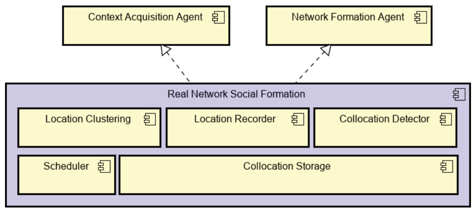
Source: Own authorship

is used to query for overlapped presence in time and space for the received context. If the overlap is detected the collocation is stored in the Collocation Storage and sent as a context to be stored in the Users' Nodes. It is worth noting that this process is also scheduled for execution as batch for the detection of collocation in new cluster locations.

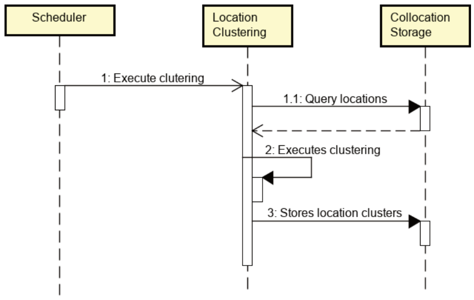
6.2 Findings of the application's usage on Author's Contacts

A randomized experiment was designed to assess engagement of the application's users. For this, two variants of the application were made available in https://app1.uductor.com, and were namely **Control** and **Intervention**. The health message recommendations feature was enabled on the Intervention variation and disabled on the Control variation. Users were randomly assigned to one of the variations as they registered in the application and were not informed about the group they were assigned.

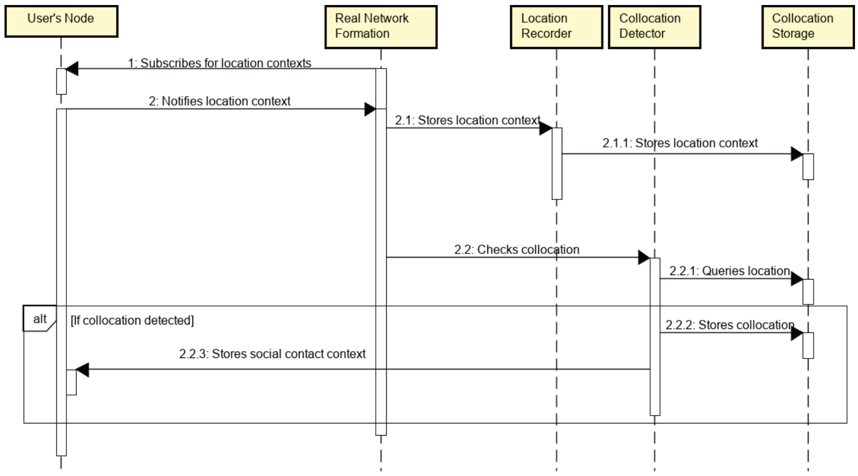
Figure 45: Real Social Network Formation Components and Process



(a) Real Social Network Formation Components



(b) Real Social Network Formation Components Training



(c) Real Social Network Formation Components Checking

Source: Own authorship

The application was promoted by direct messages in instant message platforms to the author's contacts, through posts on the author's social platforms, and to four different university discussion lists at the dates 10/12/2018, 10/15/2018 and 10/16/2018. After one month of application usage, those users who accepted the application's terms and conditions were invited to answer a survey about the application acceptance in terms of usefulness and ease of use. Next sections will expose the findings of application usage data analysis and users' perception.

6.2.1 Usage Data Analysis on Author's Contacts

Users subscribed to the application from 12th October 2018 to 30th October 2018. A total of 45 users have registered in the application, 23 (8 women and 15 men) were assigned to the Intervention version and 22 (3 women and 19 men) in the Control version.

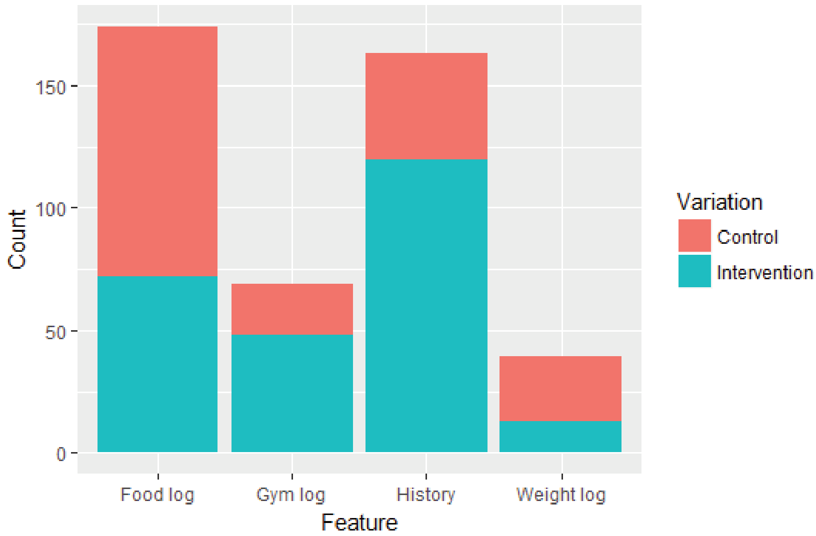
Figure 46a shows the number of times each type of feature was accessed by the users. In overall, the Control variation users had a greater usage of food and weight logging features, while the users from the Intervention group had a greater usage the History and gym logging features. Figure 46b shows the distribution of features usage through time. The Messages and Messages Interaction features were added to that plot, where it is possible to observe that

Table 13: My U'Ductor Usage Statistics of Author's Contacts

Variable	Gender	Control		Intervention		t Test p-value
		Mean	Standard Deviation	Mean	Standard Deviation	
Days General	Female	0.00	0.00	2.13	2.10	0.12
	Male	2.74	5.27	3.47	5.53	0.70
	All	2.36	4.97	3.00	4.61	0.66
Interactions General	Female	0.00	0.00	18.00	34.38	0.40
	Male	11.11	28.55	22.87	52.15	0.41
	All	9.59	26.72	21.17	45.96	0.31
Days General (Except Message Interactions)	Female	0.00	0.00	1.63	1.85	0.17
	Male	2.74	5.27	2.87	4.97	0.94
	All	2.36	4.97	2.43	4.14	0.96
Interactions General (Except Message Interactions)	Female	0.00	0.00	15.50	34.80	0.47
	Male	11.11	28.55	10.27	16.53	0.92
	All	9.59	26.72	12.09	23.78	0.74
Days (Food Log)	Female	0.00	0.00	0.75	1.39	0.39
	Male	1.16	3.85	0.13	0.52	0.32
	All	1.00	3.59	0.35	0.93	0.40
Interactions (Food Log)	Female	0.00	0.00	8.50	21.72	0.53
	Male	5.37	21.99	0.27	1.03	0.38
	All	4.64	20.45	3.13	12.92	0.77
Days (Weight Log)	Female	0.00	0.00	0.00	0.00	-
	Male	0.95	3.01	0.73	1.79	0.42
	All	0.82	2.81	0.48	1.47	0.61
Interactions (Weight Log)	Female	0.00	0.00	0.00	0.00	-
	Male	1.37	4.15	0.87	2.10	0.67
	All	1.18	3.87	0.57	1.73	0.49
Days (Gym Log)	Female	0.00	0.00	0.63	1.19	0.40
	Male	0.84	3.02	1.33	3.68	0.67
	All	0.73	2.81	1.09	3.03	0.68
Interactions (Gym Log)	Female	0.00	0.00	3.13	8.06	0.53
	Male	1.11	4.15	1.53	4.19	0.77
	All	0.95	3.86	2.09	5.70	0.44
Days (History Log)	Female	0.00	0.00	1.00	1.51	0.30
	Male	1.05	1.51	2.07	3.28	0.24
	All	0.91	1.44	1.70	2.80	0.25
Interactions (History Log)	Female	0.00	0.00	2.50	3.93	0.31
	Male	2.16	2.87	6.47	11.19	0.12
	All	1.86	2.77	5.09	9.40	0.13
Days (Message Interaction Log)	Female	-	-	0.88	0.99	-
	Male	-	-	1.20	1.97	-
	All	-	-	1.09	1.68	-
Interactions (Message Interaction Log)	Female	-	-	2.50	3.66	-
	Male	-	-	12.60	39.16	-
	All	-	-	9.09	31.69	-

Source: Own authorship

Figure 46: Plots of Application Usage



(a) Total features usage of Author's Contacts

(b) Distribution of features usage through time of Author's Contacts

Source: Own authorship

Intervention users have used the application for longer. Users of the intervention variation accessed the application for more days, on average 27% more.

Table 13 presents comparisons about mean usage information between the two groups controlled by gender. The variables starting by the description **"Days"** indicate the mean quantity of days which the application or one of its features were used. The variables starting by the description **"Interactions"** indicate the mean quantity of interactions the users did with the application or one of its features. The variables having **"General"** in their description present the sum of all features: food log, weight log, gym log, history chart query or interaction with the received messages (by liking, following links or closing them, i.e., marking them as read). The text **“Except Message Interactions”** in variables indicate that intervention users interaction with messages were not counted, this is particularly special for comparing means of the same features available for the two groups.

All variables from Table 13 were tested with a two tailed T-student test, assuming in the null hypothesis that means between the two groups are equals with a confidence interval of 95% ($\alpha = 0.05$, $p - \text{value} < \alpha$). Although the null hypothesis was not rejected, there were two cases where the p-value was lower than 0.15 (i.e. a confidence interval of 85%). These two cases were the mean number of days the application was used by women and the mean history chart query done by men and women and suggest that intervention had a positive effect in these cases. However, standard deviation values were very high indicating that usage behavior were very unequal among users. Other relevant values were general used days without counting interactions with messages among women ($p - \text{value} = 0.17$) and history chart query used days among all users ($p - \text{value} = 0.25$).

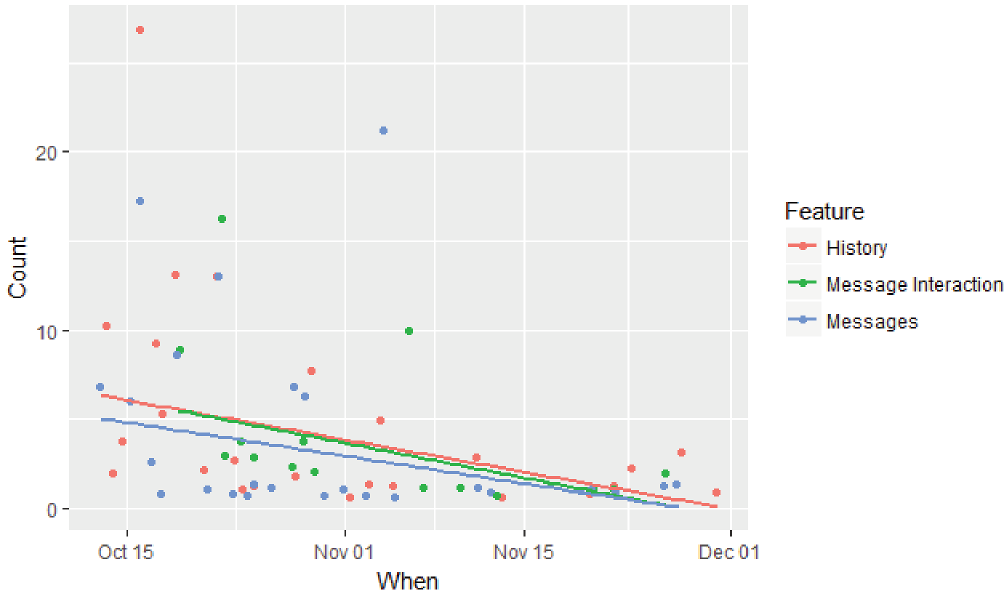
Somehow, the distributions of history chart query and message interactions showed in Figure 46b, indicate a possible connection between these two variables. To check this, a robust fitting of linear model was applied to check if these feature correlated in some way. Therefore, a strong relationship was detected as Figure 47 demonstrates by the three lines that have the same trend over the distribution. Following a Wald test for multiple coefficients was applied in each feature distribution to check the statistical significance of the linear model. The test resulted in the following p-values: History $p - \text{value} = 0.013$, $p - \text{value} < 0.05$; Message $p - \text{value} = 0.022$, $p{-}\text{value} < 0.05$; and Message Interaction $p{-}\text{value} = 0.028$, $p{-}\text{value} < 0.05$. Hence, it was possible to reject the null hypothesis and state that Messages and Message Interactions features have possibly influenced the use of history feature by the Intervention users. The regression linear model was chosen due to its ability to explain an independent variable, in this case the history variable, according to its dependents variables, which are use of the message feature and number of interaction with the messages (FREEDMAN, 2005).

6.2.2 My U’Ductor Acceptance by Author’s Contacts

After one month of application usage, users were invited to answer a survey which aimed in evaluate users acceptance about application features. Thus, two surveys were elaborated and made available for on-line access[7,8]. A total of 14 users anonymously answered the survey, 6 users from the control group and 8 users from the intervention group. The surveys questions were almost the same and differed in eight question which were addressed to the intervention group.

[7]https://forms.gle/VHJnjhyioXrDpkGV7 (Control survey)
[8]https://forms.gle/ogBfGYeNq9fgk1RN6 (Intervention survey)

Figure 47: Distribution of History, Message Interactions, and Messages features of Author's Contacts



Source: Own authorship

All survey questions are listed in Appendix B and are grouped according with the application assessed feature (General goals, Activities Schedule, Activities Report, Activities History and Recommended Messages). Questions took the form of five points Likert scale (LIKERT, 1932), single choice (i.e. user could choose one answer from a list), multiple choice (i.e. users could choose more then one answer from a list), boolean (i.e. Yes or No) and open (i.e. allowed the users to write their answers). All open questions responses are listed in Appendix C. Finally, each survey question was related to one acceptance subject which can be **Technical Aspects**, **User Report**, **Usefulness** and **Ease of Use**.

6.2.2.1 Technical Aspects

The technical aspects subject relates to the platforms and software which users used to access the application. Table 14 summarizes the browsers and operating systems used while running the application. Most users used the same browser. In regards to operating system, is worth noticing that intervention users prioritized mobile access (most users reported using Android), while control users were more platform eclectic.

6.2.2.2 User Report

Questions belonging to the user report subject were used to check the conformity of users profile in relation to the application goals and if the provided features operated as expected. The feeling of relatedness with application goals in users routine was proportional in both groups. However, the control group had more users feeling unrelated or less related with application

Table 14: Technical Aspects Answers of Author's Contacts

Which browser do you use to run the application? (Q3)			
Answer	**Control**	**Intervention**	**Total**
Chrome	6 (100,00%)	8 (100,00%)	14 (100,00%)
Firefox	2 (33,33%)	2 (25,00%)	4 (28,57%)
Safari	0 (0,00%)	1 (12,50%)	1 (7,14%)
Opera	1 (16,67%)	0 (0,00%)	1 (7,14%)
Internet Explorer	0 (0,00%)	0 (0,00%)	0 (0,00%)
Edge	0 (0,00%)	0 (0,00%)	0 (0,00%)
What operating system did you use while running the application? (Q4)			
Answer	**Control**	**Intervention**	**Total**
Windows	5 (83,33%)	3 (37,50%)	8 (57,14%)
Android	4 (66,67%)	5 (62,50%)	9 (64,29%)
Linux	3 (50,00%)	0 (0,00%)	3 (21,43%)
iOS	1 (16,67%)	1 (12,50%)	2 (14,29%)
MacOS	0 (0,00%)	0 (12,50%)	0 (7,14%)

Source: Own authorship

goals, while intervention had more users who felt neutral (Table 15). Physical activity management was the type of activity that interest users the most. Control group users were more interested in food consumption management then Intervention group, but less interested in weight management activities.

Table 16 presents users reports in regards to functional compliance of the application. Three users reported having some problem in using the application. Based on the answers given to the open question number seven (see Appendix C), one of the problem seems to be user specific, as it was reported that *"It does not let me complete the activity"*. Even so, application successfully recorded activities from other users. Other user reported difficulties on using the interface, *"The interface is a bit confusing for the user, I think you should rethink the application design a bit."*, and one user reported a bug in the user interface behavior, *"When opening the application, from the notification, the menu was always open and not the expected screen."*. In this last case, the expected screen was opened, however the application menu was in front of it.

As seen by the answers to questions 11, 28, 29 and 30, all respondent users have received notifications from the application. Most of the intervention users read the recommended messages once a day. However, two users did not know that was possible to follow message links, and three users did not know that was possible to like messages. These two features followed common user interface design patterns, and its possible that these users read the messages by their smartphone notification tray, where is not possible to follow links or like the messages.

Table 16: User Reported Application Functional Compliance of Author’s Contacts

Did you have any problems using the app? (Q6)			
Answer	**Control**	**Intervention**	**Total**
Yes	2 (33,33%)	1 (12,50%)	3 (21,43%)
No	4 (66,67%)	7 (87,50%)	11 (78,57%)
Did you receive activity notifications on your device? (Q11)			
Answer	**Control**	**Intervention**	**Total**
Yes	6 (100,00%)	8 (100,00%)	14 (100,00%)
No	0 (0,00%)	0 (0,00%)	0 (0,00%)
How often did you read the recommended messages? (Q28)			
Answer	**Control**	**Intervention**	**Total**
More than twice a day	-	1 (12,50%)	-
Once a day	-	3 (37,50%)	-
More than twice a week	-	2 (25,00%)	-
Once a week	-	1 (12,50%)	-
Did not read	-	1 (12,50%)	-
I did not receive any messages	-	0 (0,00%)	-
Did you follow the links in the messages? (Q29)			
Answer	**Control**	**Intervention**	**Total**
Yes	-	4 (50,00%)	-
No	-	2 (25,00%)	-
I did not know it was possible to follow the links	-	2 (25,00%)	-
I did not receive any messages	-	0 (0,00%)	-
Did you usually like the received messages? (Q30)			
Answer	**Control**	**Intervention**	**Total**
Yes	-	3 (37,50%)	-
No	-	2 (25,00%)	-
I did not know it was possible to like the messages	-	3 (37,50%)	-
I did not receive any messages	-	0 (0,00%)	-

Source: Own authorship

6.2.2.3 Usefulness

Perceived usefulness is one of the two dimensions assessed by the technology acceptance model (TAM) proposed by Fred D. Davis (DAVIS, 1989) and expanded by Yoon and Kim to enable its application in ubiquitous computing environments (YOON; KIM, 2007). In essence,

Table 15: User Reported Relatedness to Application Goals of Author’s Contacts

Indicate the degree of relationship of the overall goals of the application to your routine (Q1)			
Answer	**Control**	**Intervention**	**Total**
Unrelated	1 (16,67%)	0 (0,00%)	1 (7,14%)
Little related	1 (16,67%)	1 (12,50%)	2 (14,29%)
Neutral	1 (16,67%)	3 (37,50%)	4 (28,57%)
Related	2 (33,33%)	3 (37,50%)	5 (35,71%)
Very Related	1 (16,67%)	1 (12,50%)	2 (14,29%)
Which of these activities interest you most? (Q2)			
Answer	**Control**	**Intervention**	**Total**
Managing the practice of physical activities	5 (83,33%)	5 (62,50%)	10 (71,43%)
Food consumption management	3 (50,00%)	3 (37,50%)	6 (42,86%)
Weight Management	1 (16,67%)	3 (37,50%)	4 (28,57%)
Water consumption	1 (16,67%)	0 (0,00%)	1 (7,14%)

Source: Own authorship

perceived usefulness is used to determine if the proposed technology can help its users to do a better job, in this case, help managing preventive NCDs activities. The elaborated survey took into account 10 questions related to perceived usefulness, three of these were specific to intervened users. Each question was in the form of five points Likert scale, beginning from unhelpful (1) to very helpful (5). Users answers to each the question are presented in Table 17.

Users from the control found the feature for scheduling activities more helpful than intervention users, who were more neutral about it. Also, control users found the notification of scheduled activities more helpful than intervention users, although most intervened have found this notification helpful (62% of the users). By contrast, more intervened users found activity notifications by e-mail helpful, while a major part of the control users found less helpful or unhelpful (50% of the users).

Table 17: Users Reported Usefulness Answers of Author’s Contacts

	Unhelpful				**Very Helpful**
Evaluate the usefulness of scheduling activities (Q8)	**1**	**2**	**3**	**4**	**5**
Control	1 (16,67%)	0 (0,00%)	0 (0,00%)	4 (66,67%)	1 (16,67%)
Intervention	0 (0,00%)	1 (12,50%)	4 (50,00%)	2 (25,00%)	1 (12,50%)
Total	**1 (7,14%)**	**1 (7,14%)**	**4 (28,57%)**	**6 (42,86%)**	**2 (14,29%)**
Evaluate the usefulness of activity notifications received on your device (Q12)	**1**	**2**	**3**	**4**	**5**

Control	1 (16,67%)	0 (0,00%)	0 (0,00%)	3 (50,00%)	2 (33,33%)
Intervention	0 (0,00%)	0 (0,00%)	3 (37,50%)	3 (37,50%)	2 (25,00%)
Total	**1 (7,14%)**	**0 (0,00%)**	**3 (21,43%)**	**6 (42,86%)**	**4 (28,57%)**
Evaluate the usefulness of activity notifications received via email (Q13)	**1**	**2**	**3**	**4**	**5**
Control	2 (33,33%)	1 (16,67%)	1 (16,67%)	1 (16,67%)	1 (16,67%)
Intervention	0 (0,00%)	2 (25,00%)	2 (25,00%)	3 (37,50%)	1 (12,50%)
Total	**2 (14,29%)**	**3 (21,43%)**	**3 (21,43%)**	**4 (28,57%)**	**2 (14,29%)**
Evaluate the usefulness of meal reporting in your routine (Q14)	**1**	**2**	**3**	**4**	**5**
Control	2 (33,33%)	1 (16,67%)	0 (0,00%)	3 (50,00%)	0 (0,00%)
Intervention	0 (0,00%)	2 (25,00%)	2 (25,00%)	3 (37,50%)	1 (12,50%)
Total	**2 (14,29%)**	**3 (21,43%)**	**2 (14,29%)**	**6 (42,86%)**	**1 (7,14%)**
Evaluate the usefulness of reporting physical activity in your routine (Q17)	**1**	**2**	**3**	**4**	**5**
Control	1 (16,67%)	0 (0,00%)	1 (16,67%)	4 (66,67%)	0 (0,00%)
Intervention	2 (25,00%)	1 (12,50%)	1 (12,50%)	3 (37,50%)	1 (12,50%)
Total	**3 (21,43%)**	**1 (7,14%)**	**2 (14,29%)**	**7 (50,00%)**	**1 (7,14%)**
Evaluate the usefulness of weight reporting in your routine (Q20)	**1**	**2**	**3**	**4**	**5**
Control	2 (33,33%)	0 (0,00%)	1 (16,67%)	3 (50,00%)	0 (0,00%)
Intervention	0 (0,00%)	2 (25,00%)	0 (0,00%)	4 (50,00%)	2 (25,00%)
Total	**2 (14,29%)**	**2 (14,29%)**	**1 (7,14%)**	**7 (50,00%)**	**2 (14,29%)**
Evaluate the usefulness of your activity history in your routine (Q23)	**1**	**2**	**3**	**4**	**5**
Control	1 (16,67%)	0 (0,00%)	1 (16,67%)	2 (33,33%)	2 (33,33%)
Intervention	0 (0,00%)	0 (0,00%)	3 (37,50%)	3 (37,50%)	2 (25,00%)
Total	**1 (7,14%)**	**0 (0,00%)**	**4 (28,57%)**	**5 (35,71%)**	**4 (28,57%)**

Evaluate, in general, the usefulness of the recommended messages (Q31)	**1**	**2**	**3**	**4**	**5**
Control	-	-	-	-	-
Intervention	1 (12,50%)	0 (0,00%)	1 (12,50%)	5 (62,50%)	1 (12,50%)
Evaluate the usefulness of the recommended messages related to the practice of physical activities (Q32)	**1**	**2**	**3**	**4**	**5**
Control	-	-	-	-	-
Intervention	1 (12,50%)	0 (0,00%)	2 (25,00%)	4 (50,00%)	1 (12,50%)
Evaluate the usefulness of the recommended messages related to healthy eating (Q33)	**1**	**2**	**3**	**4**	**5**
Control	-	-	-	-	-
Intervention	1 (12,50%)	0 (0,00%)	1 (12,50%)	3 (37,50%)	3 (37,50%)
Evaluate the usefulness of the recommended messages related to weight control (Q34)	**1**	**2**	**3**	**4**	**5**
Control	-	-	-	-	-
Intervention	1 (12,50%)	0 (0,00%)	2 (25,00%)	4 (50,00%)	1 (12,50%)

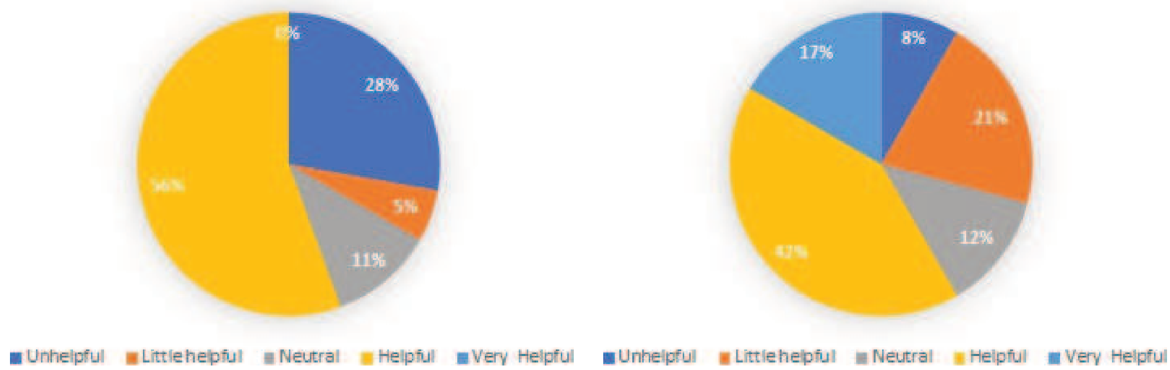
Source: Own authorship

In regards to activities log, the feature for reporting consumed meals was the worst evaluated among the two groups. In general, 35.5% of users (control and intervention) evaluated the feature as less helpful or unhelpful. Users from the control group were more likely to found physical activity reporting useful (66.67% of the users), while 50% of intervened users found this reporting useful. By contrast, the weight report was the best log feature evaluated by users of intervention group (75% of the users), compared to 50% from users of the control group. The activities history chart query was the best evaluated feature by both groups. This feature was understood as useful by 64.29% of users (66.67% of control users and 62.50% of intervention users).

The recommended messages feature was evaluated by users from the intervention group. In general, 75% of the users found the recommended messages useful. Messages related to healthy eating were the most appreciated by the users (75%), while messages about physical activity practice and weight control were little less appreciated (62.5% in both types of messages).

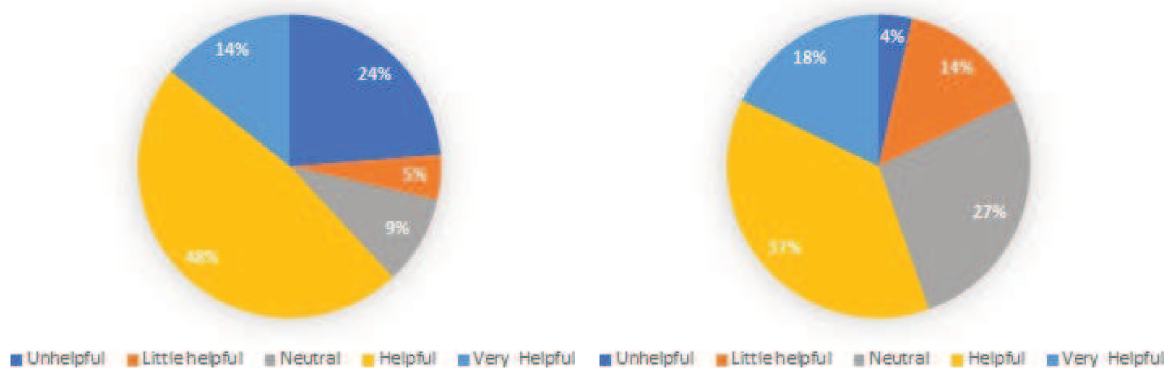
Figure 48 shows the general perceived usefulness for all types of activities reporting features and for the application as a whole in both groups. In general, the application and the activities reporting features were found useful in both groups. However, users from the control group were more likely to find the application or the activities reporting features less helpful or unhelpful. Users from the intervention group were more likely to be neutral in relation to all evaluated application features. Finally, T-student tests were executed for each question related to usefulness, but no statistical significance was found, indicating that answers given by the two groups were not significantly different (WINTER; DODOU, 2010).

Figure 48: General Perceived Usefulness of Author's Contacts



(a) General Perceived Usefulness of the Activities Report by the Control Group of Author's Contacts

(b) General Perceived Usefulness of the Activities Report by the Intervention Group of Author's Contacts



(c) General Perceived Usefulness by the Control Group of Author's Contacts

(d) General Perceived Usefulness by the Intervention Group of Author's Contacts

Source: Own authorship

6.2.2.4 Ease of Use

Perceived ease of use is another of the two dimensions assessed in TAM evaluations, aiming in evaluating whether the technology can be used with a minimum of effort. The elaborated survey took into account 7 questions related to perceived ease of use, each of these was in the form of five points Likert scale, beginning from very difficult (1) to very easy (5). Users answers to each the question are presented in Table 18.

Table 18: Users Reported Ease of Use Answers of Author's Contacts

	Very Difficult				**Very Easy**
Evaluate the degree of ease of use of the activity schedule (Q9)	**1**	**2**	**3**	**4**	**5**
Control	0 (0,00%)	0 (0,00%)	2 (33,33%)	2 (33,33%)	2 (33,33%)
Intervention	0 (0,00%)	1 (12,50%)	1 (12,50%)	4 (50,00%)	2 (25,00%)
Total	**0 (0,00%)**	**1 (7,14%)**	**3 (21,43%)**	**6 (42,86%)**	**4 (28,57%)**
Evaluate the degree of ease of use of the meal reporting (Q15)	**1**	**2**	**3**	**4**	**5**
Control	1 (16,67%)	3 (50,00%)	1 (16,67%)	1 (16,67%)	0 (0,00%)
Intervention	1 (12,50%)	1 (12,50%)	2 (25,00%)	2 (25,00%)	2 (25,00%)
Total	**2 (14,29%)**	**4 (28,57%)**	**3 (21,43%)**	**3 (21,43%)**	**2 (14,29%)**
Evaluate the degree of ease of use of the physical activity report (Q18)	**1**	**2**	**3**	**4**	**5**
Control	0 (0,00%)	1 (16,67%)	3 (50,00%)	1 (16,67%)	1 (16,67%)
Intervention	1 (12,50%)	1 (12,50%)	0 (0,00%)	4 (50,00%)	2 (25,00%)
Total	**1 (7,14%)**	**2 (14,29%)**	**3 (21,43%)**	**5 (35,71%)**	**3 (21,43%)**
Evaluate the degree of ease of use of the weight reporting (Q21)	**1**	**2**	**3**	**4**	**5**
Control	1 (16,67%)	0 (0,00%)	1 (16,67%)	3 (50,00%)	1 (16,67%)
Intervention	1 (12,50%)	1 (12,50%)	0 (0,00%)	3 (37,50%)	3 (37,50%)
Total	**2 (14,29%)**	**1 (7,14%)**	**1 (7,14%)**	**6 (42,86%)**	**4 (28,57%)**
Evaluate the ease of understanding of the food consumption history chart (Q24)	**1**	**2**	**3**	**4**	**5**
Control	0 (0,00%)	0 (0,00%)	1 (16,67%)	3 (50,00%)	2 (33,33%)
Intervention	0 (0,00%)	0 (0,00%)	3 (37,50%)	3 (37,50%)	2 (25,00%)
Total	**0 (0,00%)**	**0 (0,00%)**	**4 (28,57%)**	**6 (42,86%)**	**4 (28,57%)**

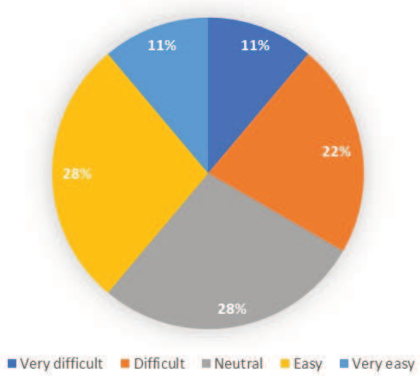
Evaluate the ease of understanding of the history chart of physical activity practice (Q25)	**1**	**2**	**3**	**4**	**5**
Control	0 (0,00%)	0 (0,00%)	1 (16,67%)	4 (66,67%)	1 (16,67%)
Intervention	0 (0,00%)	0 (0,00%)	2 (25,00%)	3 (37,50%)	3 (37,50%)
Total	**0 (0,00%)**	**0 (0,00%)**	**3 (21,43%)**	**7 (50,00%)**	**4 (28,57%)**
Evaluate the ease of understanding of the weight history chart (Q26)	**1**	**2**	**3**	**4**	**5**
Control	0 (0,00%)	0 (0,00%)	1 (16,67%)	4 (66,67%)	1 (16,67%)
Intervention	0 (0,00%)	0 (0,00%)	2 (25,00%)	3 (37,50%)	3 (37,50%)
Total	**0 (0,00%)**	**0 (0,00%)**	**3 (21,43%)**	**7 (50,00%)**	**4 (28,57%)**

Source: Own authorship

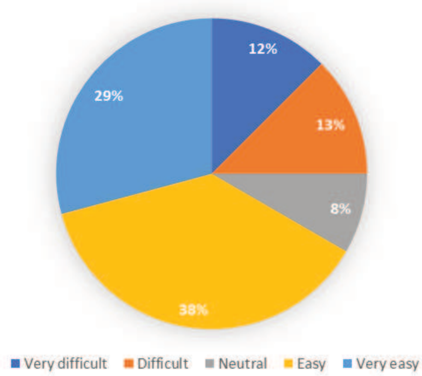
The scheduling activities feature was found easy by most of the users from both the control and intervention groups. Answers relating the ease of use of meal reporting feature have diverged among the two groups. A T-student test executed with the answers samples resulting in a $p - value$ equals to 0.15, which may be not statistically significant, but was low in relation to the values of tests executed in the samples of other question answers, and considering the scale length the questions. In this case, 66.67% of control users found the feature very difficult or difficult to use. This issue with meal reporting usability was also addressed by two users in the open question related to this feature (question 16 from Appendix C). Users from the control group found the physical activity reporting feature more easy to use (75% of the respondent users) than control group users who tend to be more neutral. The weight reporting feature was found easy to use by users from both groups (71% of all respondent users, 66% of control group users, and 75% intervention group users).

All types of history charts were found easy to use by most of the respondent users of both groups. Moreover, control group users slightly evaluated history charts better than intervened users, who felt neutral in regards to food consumption charts. In general, users from the intervention group found activity report features more useful than control users (Figure 49a), while history charts were best evaluated by control users than intervened users (Figure 49a). Finally, overall features were found easy to use by the respondent users from both groups (Figure 49f). T-student tests were executed to check if was significant differences in the answers given by users from the control and intervention groups. No statistical significance was found, indicating that answers given by the two groups were not significantly different.

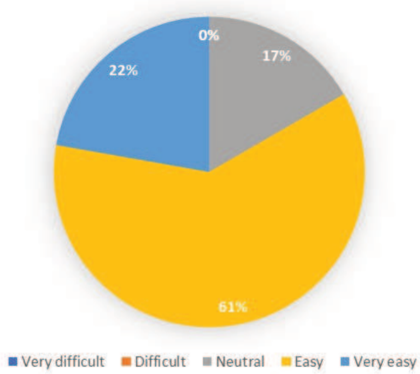
Figure 49: General Perceived Ease of Use of Author's Contacts



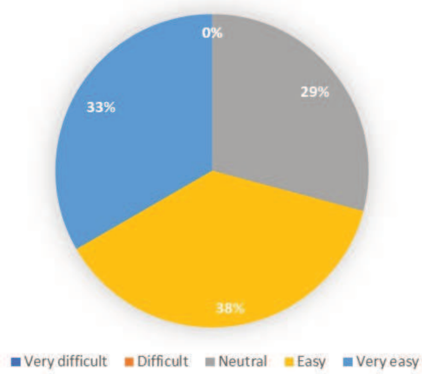
(a) General perceived ease of use of activity report by the Author's Contacts Control Group



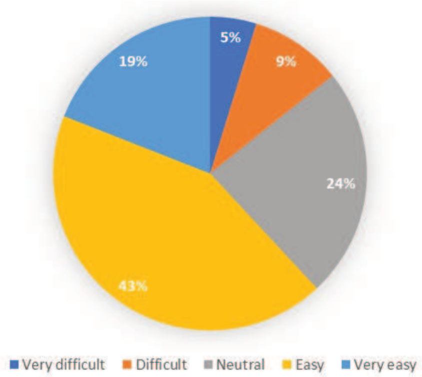
(b) General perceived ease of use of activity report by the Author's Contacts Intervention Group



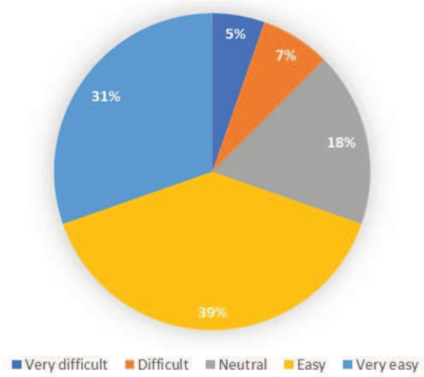
(c) General perceived ease of use of history charts by the Author's Contacts Control Group



(d) General perceived ease of use of history charts by the Author's Contacts Intervention Group



(e) General perceived ease of use by the Author's Contacts Control Group



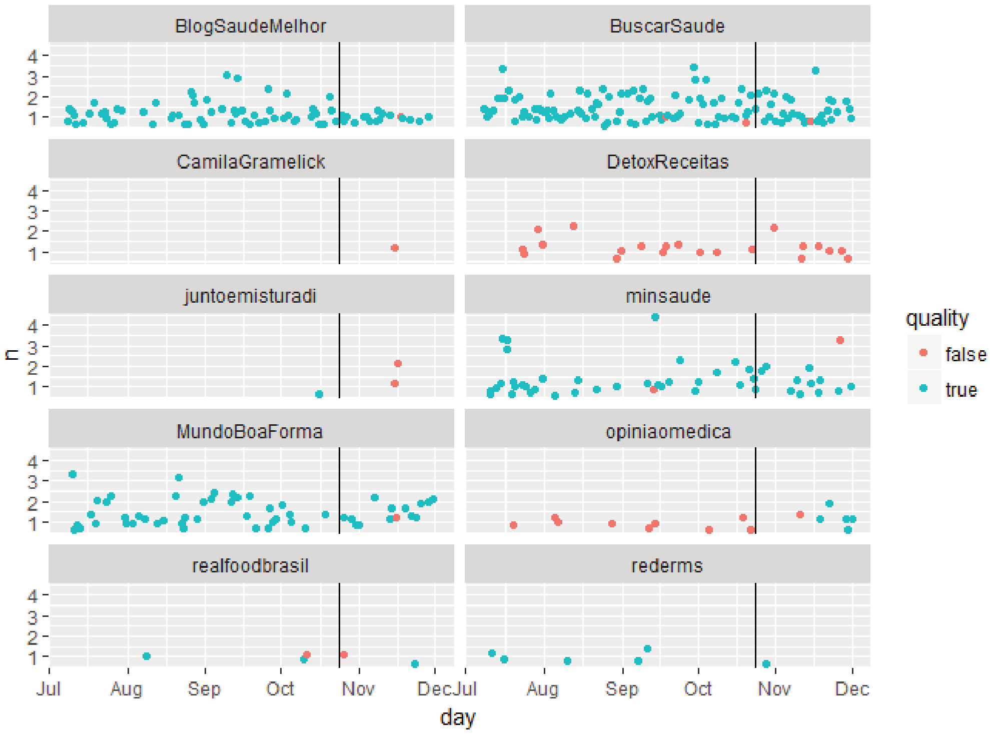
(f) General perceived ease of use by the Author's Contacts Control Group

Source: Own authorship

6.2.3 Twitter Profiles Behavior's Analysis on Author's Contacts Usage

The selected Twitter profiles were monitored from 07/08/2018 to 12/01/2018. In total, 7,011 messages were processed, from that 470 from 10 profiles were classified as non-communicable

Figure 50: Distribution of the sending of non-communicable diseases prevention message by the monitored Twitter profiles on author's contacts experiment



Source: Own authorship

diseases prevention messages and 420 passed the quality check. Quality check was done to ensure the delivery of messages that did not stimulate the cult of body image (BOEPPLE et al., 2016; SIMPSON; MAZZEO, 2017), so that the messages sent by the platform had, for the most part, educational bias (SMAHEL; ELAVSKY; MACHACKOVA, 2017). Also, messages should be self-contained, or at least had a link pointing to more information.

By the day 10/23/2018 a contact with each monitored Twitter profile was made through direct messaging in order to encourage them to improve the posting of non-communicable diseases prevention-related messages. The message presented the My UDuctor application and had a link pointing to the Twitter's Health Profiles Rank. The profiles @opiniaomedica and @BuscarSaude accessed the rank once, as the others profiles did not accessed the rank. Fi- gure 50 presents the distribution non-communicable diseases prevention messages sent by the monitored Twitter profiles. The vertical line indicates the moment when the profiles were notified about the application, and red dots represent messages that did not pass the quality check, while the blue dots did.

Before the first contact, @minsaude was on the top of the rank followed by @BlogSaudeMelhor and @BuscarSaude. By the day 11/15/2018, the profile @BuscarSaude took the first position, followed by @BlogSaudeMelhor and @minsaude. At the end of the experiment, @BlogSaudeMelhor was in the first position, followed by @BuscarSaude. The profile @opiniaomedica jumped from the tenth position to the fifth.

Table 19: Causal Inference Test on Author's Contacts Usage

Profile	Before Intervetion			After Intervention			Causal Test (p-value)
	Messages	Daily Mean	Standard Deviation	Messages	Daily Mean	Standard Deviation	
BlogSaudeMelhor	**72**	**0.67**	**0.71**	**17**	**0.44**	**0.55**	**0.062**
BuscarSaude	121	1.13	0.80	43	1.10	0.72	0.417
CamilaGramelick	0	0.00	0.00	1	0.03	0.16	0.000
DetoxReceitas	16	0.15	0.41	8	0.21	0.47	0.248
juntoemisturadi	**1**	**0.01**	**0.10**	**3**	**0.08**	**0.35**	**0.001**
minsaude	48	0.45	0.83	18	0.46	0.85	0.477
MundoBoaForma	57	0.53	0.78	20	0.51	0.76	0.488
opiniaomedica	**9**	**0.08**	**0.28**	**7**	**0.18**	**0.45**	**0.051**
realfoodbrasil	3	0.03	0.17	2	0.05	0.22	0.302
rederms	5	0.05	0.21	1	0.03	0.16	0.357

Source: Own authorship

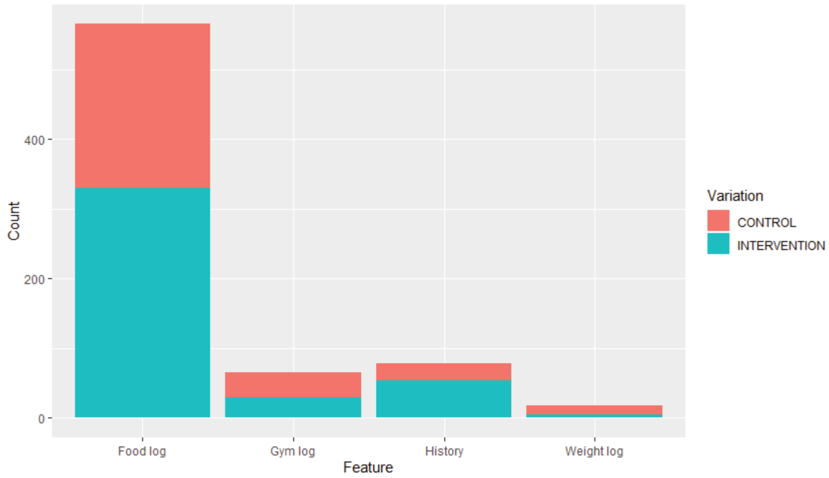
A causal inference with Bayesian structural time-series model test (BRODERSEN et al., 2015) was made to check if the contact made had an effect in the tweeting behavior of Twitters' profile. Table 19 summarizes the results of the test. In general, tests results were not statistically significant and could not have meaningful interpretations. However, three profiles were exceptions and they are highlighted in Table 19. It seems that contact did a positive effect, in the profiles @juntoemisturadi and @opiniaomedica, and a negative effect on @BlogSaudeMelhor. Particularly the profile @opiniaomedica increased the number of messages sent as also had an improvement in the quality of the content of the messages after the intervention, as shown by the blue dots in Figure 50 that only started to exist after the intervention. Furthermore, the p-value result for this profile was close to statistical significance, suggesting that the intervention had a positive effect on him.

6.3 Findings of the application's usage on Physical Education Students

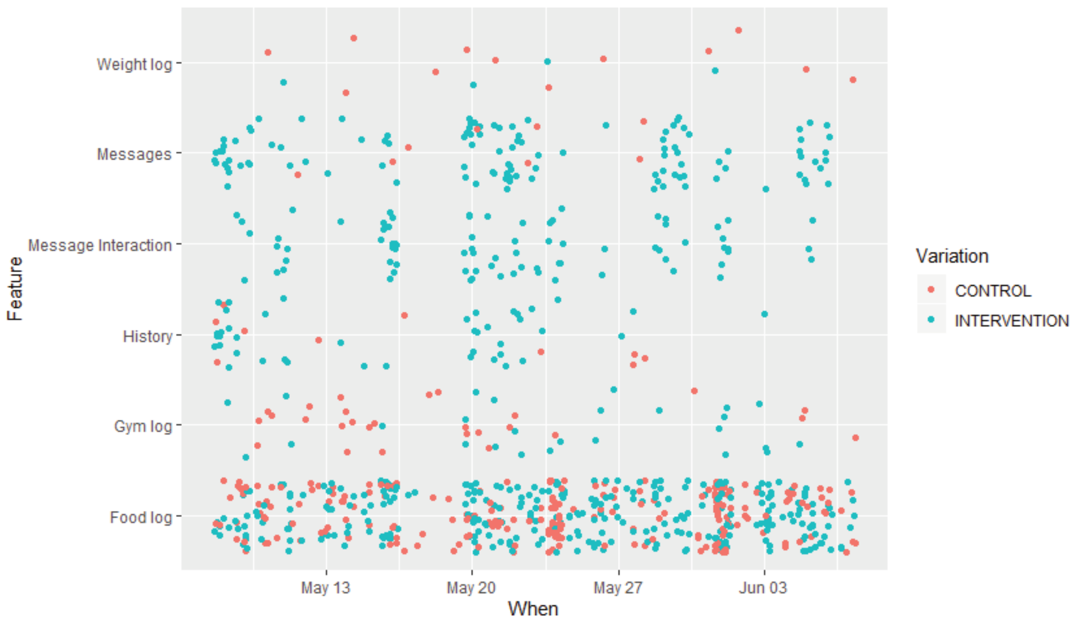
The experiment took with the author's contacts was repeated with students from the Physical Education course of the University of Vale do Rio dos Sinos from 05/08/2019 to 06/06/2019. The application, its features, and restrictions were presented to two classes of physical education students in the days 05/08/2019 and 05/09/2019. A total of 16 students subscribed the application, each being randomly assigned to the control or intervention group. The health message recommendations feature was enabled on the Intervention variation and disabled on the Control variation. Again, the application was available at https://app1.uductor.com, the users' acceptance was evaluated in terms of usefulness and ease of use, and are explained in next as are the findings of application usage data analysis.

6.3.1 Usage Data Analysis on Physical Education Students

Figure 51: Plots of Application Usage of Physical Education Students



(a) Total features usage of physical education students



(b) Distribution of features usage of physical education students through time

Source: Own authorship

The users were assigned to a group as long as they subscribed to the application. The 16 users were equally distributed to the groups, that is, eight users were assigned to the control group, whereas the eight remaining users were assigned to the intervention group. Groups were equally distributed in terms of gender, having each group three women and five men.

Figure 51a shows the number of times each type of feature was accessed by the users. The overall feature usage distribution was similar to the first experiment, as the food log and history were the most used features. Differences reside on usage by groups. Different from the later experiment, at this time intervention users made larger use of the food log feature. The gym log

Table 20: My U'Ductor Usage Statistics of Physical Education Students

Variable	Gender	Control		Intervention		t Test p-value
		Mean	Standard Deviation	Mean	Standard Deviation	
Days General	Female	11.67	14.36	9.33	12.10	0.84
	Male	3.60	5.86	4.20	6.57	0.88
	All	6.63	9.80	6.13	8.58	0.92
Interactions General	Female	67.00	59.02	63.33	78.39	0.95
	Male	13.00	19.94	54.60	95.27	0.37
	All	33.25	44.76	57.88	83.44	0.47
Days General (Except Message Interactions)	Female	11.67	14.36	9.33	12.10	0.84
	Male	3.60	5.86	4.20	6.57	0.88
	All	6.63	9.80	6.13	8.58	0.92
Interactions General (Except Message Interactions)	Female	67.00	59.02	62.00	12.10	0.93
	Male	13.00	19.94	40.00	64.07	0.39
	All	33.25	44.76	48.25	64.28	0.60
Days (Food Log)	Female	11.33	14.74	9.33	12.10	0.86
	Male	2.80	5.17	4.00	6.52	0.76
	All	6.00	9.84	6.00	8.59	1.00
Interactions (Food Log)	Female	58.33	51.81	49.00	75.54	0.87
	Male	7.00	12.00	30.00	49.48	0.34
	All	26.25	39.43	37.13	55.91	0.66
Days (Weight Log)	Female	2.33	4.04	0.00	0.00	0.37
	Male	0.80	1.79	0.80	1.79	1.00
	All	1.38	2.67	0.50	1.41	0.43
Interactions (Weight Log)	Female	2.33	4.04	0.00	0.00	0.37
	Male	0.80	1.79	0.80	1.79	1.00
	All	1.38	2.67	0.50	1.41	0.43
Days (Gym Log)	Female	2.67	4.62	1.33	2.31	0.68
	Male	1.60	3.58	3.00	4.80	0.62
	All	2.00	3.70	2.38	3.93	0.85
Interactions (Gym Log)	Female	3.33	5.77	1.33	2.31	0.61
	Male	3.20	7.16	4.60	6.39	0.75
	All	3.25	6.23	3.38	5.26	0.97
Days (History Log)	Female	1.33	1.53	4.33	4.51	0.34
	Male	0.20	0.45	0.80	1.79	0.49
	All	0.63	1.06	2.13	3.31	0.24
Interactions (History Log)	Female	2.00	2.65	9.00	8.19	0.23
	Male	0.60	1.34	3.20	7.16	0.45
	All	1.13	1.89	5.38	7.58	0.15
Days (Message Interaction Log)	Female	-	-	1.00	1.73	-
	Male	-	-	2.40	5.37	-
	All	-	-	1.88	4.22	-
Interactions (Message Interaction Log)	Female	-	-	1.33	2.31	-
	Male	-	-	14.60	32.65	-
	All	-	-	9.63	25.65	-

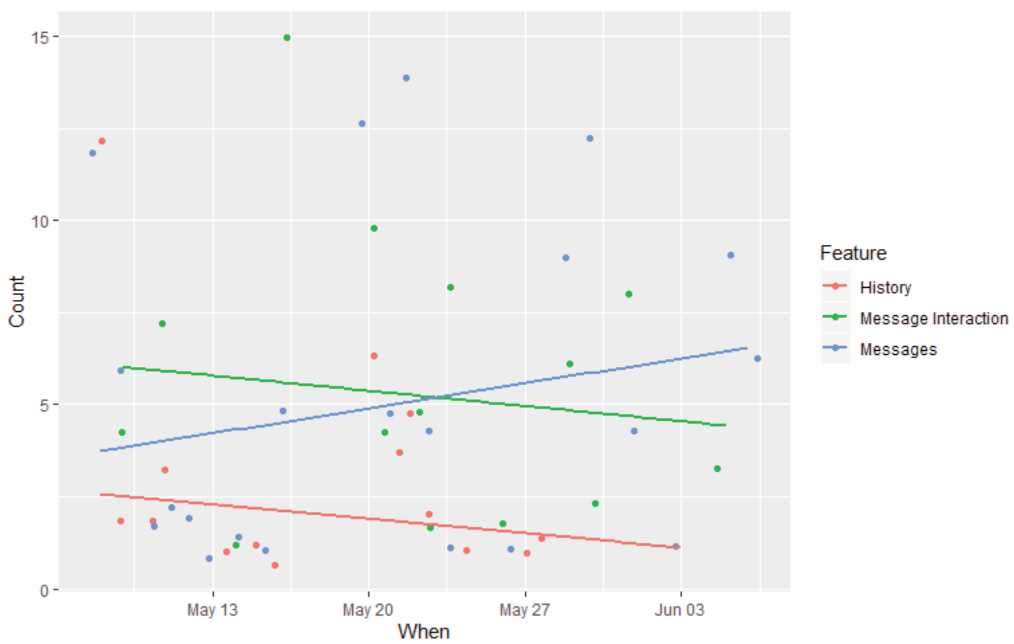
Source: Own authorship

feature usage was slightly larger by the control group. Again, the history feature was used most of the time by the intervention group, while the weight log feature was used in most part by the control group. Distribution of the features usage through time is shown in Figure 51b, pointing that application was regularly used by both groups.

The mean usage comparison controlled by gender is presented in Table 20. Again, the variables starting by the description "Days" indicate the mean quantity of days which the application or one of its features was used, and the variables starting by the description "Interactions" indicate the mean quantity of interactions the users did with the application or one of its features. The "General" description means the aggregation of all features, and the "Except Message Interactions" means that log of users interaction with messages were discarded.

All variables of Table 13 were tested by the application of a T-student test. This test assumed that means between the two groups are equals with a confidence interval of 95% (α = 0.05, $p - value < \alpha$). The null hypothesis was not rejected on any test. However, the assessed $p - value$ of overall interactions with history log was again the lower value which gives a strong indication of the likelihood of a positive effect of the intervention on users in following their health progress. This insight is reinforced by the robust fitting of linear model presented in Figure 52 that shows a similar trend of message interaction and history usage through time. Nevertheless, the Wald test results on these variables were not statistically significant, possibly due to the sample size.

Figure 52: Distribution of History, Message Interactions, and Messages features of Physical Education Students



Source: Own authorship

Table 21: Technical Aspects Answers of Physical Education Students

Which browser do you use to run the application? (Q3)			
Answer	**Control**	**Intervention**	**Total**
Chrome	6 (100.00%)	4 (100.00%)	10 (100.00%)
Firefox	0 (0.00%)	0 (0.00%)	0 (0.00%)
Safari	0 (0.00%)	0 (0.00%)	0 (0.00%)
Opera	0 (0.00%)	0 (0.00%)	0 (0.00%)
Internet Explorer	0 (0.00%)	0 (0.00%)	0 (0.00%)
Edge	0 (0.00%)	0 (0.00%)	0 (0.00%)
What operating system did you use while running the application? (Q4)			
Answer	**Control**	**Intervention**	**Total**
Windows	0 (0.00%)	2 (50.00%)	2 (20.00%)
Android	5 (83.33%)	2 (50.00%)	7 (70.00%)
Linux	0 (0.00%)	0 (0.00%)	0 (0.00%)
iOS (iPhone/iPad)	1 (16.67%)	0 (0.00%)	1 (10.00%)
MacOS	0 (0.00%)	0 (0.00%)	0 (0.00%)

Source: Own authorship

6.3.2 My U'Ductor Acceptance by Physical Education Students

After 06/06/2019 users were invited to answer a survey which aimed to evaluate their acceptance of application features. The same surveys answered by the author's contacts, and listed in Appendix B, were electronically sent to the physical education students[9,10]. Six users answered the control survey, while four users answered the intervention survey. At this time, the survey was not anonymous and users required to identify themselves. All open questions responses are listed in Appendix D. Again, each survey question was related to one acceptance subject like Technical Aspects, User Report, Usefulness, and Ease of Use.

6.3.2.1 Technical Aspects

Table 21 shows information about the platforms used by users to access the application. The same browser was used by all users, diverging on operational systems, whereas Android was the most used operating system, followed by Windows and iOS. The intervention group was more platform eclectic, while the control group was predominantly composed of windows users. It is worth to mention that the application was not almost compatible with iOS, so its possible that one user has accidentally mistaken the question, and probably used a desktop operating system instead. Furthermore, this incompatibility was presented in class to the students, which reinforces this interpretation.

[9]https://forms.gle/KVyzMiyUC7H38fhn6 (Control survey)
[10]https://forms.gle/btRj2E6ySRhsTQrG6 (Intervention survey)

Table 22: User Reported Relatedness to Application Goals of Physical Education Students

Indicate the degree of relationship of the overall goals of the application to your routine (Q1)			
Answer	**Control**	**Intervention**	**Total**
Unrelated	0 (0.00%)	0 (0.00%)	0 (0.00%)
Little related	1 (16.67%)	1 (25.00%)	2 (20.00%)
Neutral	2 (33.33%)	1 (25.00%)	3 (30.00%)
Related	2 (33.33%)	1 (25.00%)	3 (30.00%)
Very Related	1 (16.67%)	1 (25.00%)	2 (20.00%)
Which of these activities interest you most? (Q2)			
Answer	**Control**	**Intervention**	**Total**
Managing the practice of physical activities	3 (50.00%)	3 (75.00%)	6 (60.00%)
Food consumption management	4 (66.67%)	2 (50.00%)	6 (60.00%)
Weight Management	2 (33.33%)	1 (25.00%)	3 (30.00%)

Source: Own authorship

6.3.2.2 User Report

Table 22 shows the perception users had about conformance of the application with their routine. Half of the users agreed that application is related or very related with their routine, as the other half was composed of users who felt neutral (30%) or little related (20%) with the applications goals. The majority of users reported being more interested in managing physical activities and food consumption than weight. Intervention users reported a bigger interest in physical activity management than the control group, while control group users reported a bigger interest in food consumption. Curiously, the reported interest was not reflect in the features usage.

Functional compliance of the application is shown in Table 23. Two control group users reported having problems with the application, as shows the answers to question 6. One of these problems was not technical being related to a usability issue (see Appendix D). The other problem was related to instability in the cloud service provider which affected the Planner's scheduler component. This problem is related to question 11 (Q11) answers since notifications were not able to be sent while the scheduler component was not working. Intervention users reported reading messages at lest once a week. However, one user seemed not to be interested in messages. Also, two intervened users reported not knowing the possibility of following links, and one user reported not knowing the possibility of linking messages. Theses features were presented in class and followed common user interface design patterns.

Table 23: User Reported Application Functional Compliance of Physical Education Students

Did you have any problems using the app? (Q6)			
Answer	**Control**	**Intervention**	**Total**
Yes	2 (33.33%)	0 (0.00%)	2 (20.00%)
No	4 (66.67%)	4 (100.00%)	8 (80.00%)
Did you receive activity notifications on your device? (Q11)			
Answer	**Control**	**Intervention**	**Total**
Yes	5 (83.33%)	3 (74.00%)	8 (80.00%)
No	0 (0.00%)	0 (0.00%)	0 (0.00%)
Sometimes there were scheduled activities but notifications were not received	1 (16.67%)	1 (25.00%)	2 (20.00%)
How often did you read the recommended messages? (Q28)			
Answer	**Control**	**Intervention**	**Total**
More than twice a day	-	1 (25.00%)	-
Once a day	-	0 (0.00%)	-
More than twice a week	-	0 (0.00%)	-
Once a week	-	2 (50.00%)	-
Did not read	-	1 (25.00%)	-
I did not receive any messages	-	0 (0.00%)	-
Did you follow the links in the messages? (Q29)			
Answer	**Control**	**Intervention**	**Total**
Yes	-	0 (0.00%)	-
No	-	2 (50.00%)	-
I did not know it was possible to follow the links	-	2 (50.00%)	-
I did not receive any messages	-	0 (0.00%)	-
Did you usually like the received messages? (Q30)			
Answer	**Control**	**Intervention**	**Total**
Yes	-	1 (25.50%)	-
No	-	2 (50.00%)	-
I did not know it was possible to like the messages	-	1 (25.00%)	-
I did not receive any messages	-	0 (0.00%)	-

Source: Own authorship

6.3.2.3 Usefulness

The usefulness of the application perceived by users is described in Table 24 with answers to five points Likert scale questions, beginning from unhelpful (1) to very helpful (5). The scheduling activities was found helpful by the majority of users of both groups. Intervention users found the receiving of notification less helpful than control group users. This may be related to the users that accessed the application by desktop, where notifications do not play the same function as in smartphone. E-mail notifications were found useful by just 30% of the users of both group, 70% of users found this feature unhelpful.

Table 24: Users Reported Usefulness Answers of Physical Education Students

	Unhelpful				**Very Helpful**
Evaluate the usefulness of scheduling activities (Q8)	**1**	**2**	**3**	**4**	**5**
Control	0 (0.00%)	1 (16.67%)	1 (16.67%)	3 (50.00%)	1 (16.67%)
Intervention	0 (0.00%)	1 (25.00%)	0 (0.00%)	3 (75.00%)	0 (0.00%)
Total	**0 (0.00%)**	**2 (20.00%)**	**1 (10.00%)**	**6 (60.00%)**	**1 (10,00%)**
Evaluate the usefulness of activity notifications received on your device (Q12)	**1**	**2**	**3**	**4**	**5**
Control	1 (16.67%)	0 (0.00%)	2 (33.33%)	0 (0.00%)	3 (50.00%)
Intervention	1 (25.00%)	0 (0.00%)	2 (50.00%)	0 (0.00%)	1 (25.00%)
Total	**2 (20.00%)**	**0 (0.00%)**	**4 (40.00%)**	**0 (0.00%)**	**4 (40.00%)**
Evaluate the usefulness of activity notifications received via email (Q13)	**1**	**2**	**3**	**4**	**5**
Control	1 (16.67%)	3 (50.00%)	0 (0.00%)	1 (16.67%)	1 (16.67%)
Intervention	2 (50.00%)	1 (25.00%)	0 (0.00%)	0 (0.00%)	1 (25.00%)
Total	**3 (30.00%)**	**4 (40.00%)**	**0 (0.00%)**	**1 (10.00%)**	**2 (20.00%)**
Evaluate the usefulness of meal reporting in your routine (Q14)	**1**	**2**	**3**	**4**	**5**
Control	0 (0.00%)	2 (33.33%)	2 (33.33%)	2 (33.33%)	0 (0.00%)
Intervention	1 (25.00%)	1 (25.00%)	2 (50.00%)	0 (0.00%)	0 (0.00%)
Total	**1 (10.00%)**	**3 (30.00%)**	**4 (40.00%)**	**2 (20.00%)**	**0 (0.00%)**

Evaluate the usefulness of reporting physical activity in your routine (Q17)	**1**	**2**	**3**	**4**	**5**
Control	0 (0.00%)	2 (33.33%)	1 (16.67%)	3 (50.00%)	0 (0.00%)
Intervention	1 (25.00%)	0 (0.00%)	1 (25.00%)	1 (25.00%)	1 (25.00%)
Total	**1 (10.00%)**	**2 (20.00%)**	**2 (20.00%)**	**4 (40.00%)**	**1 (10.00%)**
Evaluate the usefulness of weight reporting in your routine (Q20)	**1**	**2**	**3**	**4**	**5**
Control	0 (0.00%)	1 (16.67%)	2 (33.33%)	3 (50.00%)	0 (0.00%)
Intervention	2 (50.00%)	0 (0.00%)	1 (25.00%)	0 (0.00%)	1 (25.00%)
Total	**2 (20.00%)**	**1 (10.00%)**	**3 (30.00%)**	**3 (30.00%)**	**1 (10.00%)**
Evaluate the usefulness of your activity history in your routine (Q23)	**1**	**2**	**3**	**4**	**5**
Control	0 (0.00%)	2 (33.33%)	2 (33.33%)	2 (33.33%)	0 (0.00%)
Intervention	0 (0.00%)	2 (50.00%)	0 (0.00%)	0 (0.00%)	2 (50.00%)
Total	**0 (0.00%)**	**4 (40.00%)**	**2 (20.00%)**	**2 (20.00%)**	**2 (20.00%)**
Evaluate, in general, the usefulness of the recommended messages (Q31)	**1**	**2**	**3**	**4**	**5**
Control	-	-	-	-	-
Intervention	0 (0.00%)	0 (0.00%)	3 (75.00%)	0 (0.00%)	1 (25.00%)
Evaluate the usefulness of the recommended messages related to the practice of physical activities (Q32)	**1**	**2**	**3**	**4**	**5**
Control	-	-	-	-	-
Intervention	0 (0.00%)	0 (0.00%)	2 (50.00%)	1 (25.00%)	1 (25.00%)
Evaluate the usefulness of the recommended messages related to healthy eating (Q33)	**1**	**2**	**3**	**4**	**5**
Control	-	-	-	-	-
Intervention	0 (0.00%)	1 (25.00%)	1 (25.00%)	1 (25.00%)	1 (25.00%)

Evaluate the usefulness of the recommended messages related to weight control (Q34)	1	2	3	4	5
Control	-	-	-	-	-
Intervention	1 (25.00%)	0 (0.00%)	1 (25.00%)	1 (25.00%)	1 (25.00%)

Source: Own authorship

Usefulness perception of logging activities features is presented in the answers to questions 14, 17 and 20. The perception of usefulness in regards to the meal reporting was uniform among the control group users, varying from little helpful to helpful. Intervention users did not find this feature helpful, most of them perceiving it neutrally. Physical activity reporting was found useful by 50% of users in both groups. In its turn, weight reporting was found useful in most part by control group users, while half of the intervention group users found it unhelpful. The usefulness of activity history (question 23) reflects in part the usage of this feature in relation to the groups. In this case, 50% intervention group users found the feature helpful, in contrast to 33% of control group users.

Intervention group users felt neutral in relation to the overall usefulness of recommended messages. However, this trend changed when users were asked to evaluate the usefulness of messages according to their topics. In this case, half of the users evaluate recommended messages regarding physical activity practice, healthy eating, and weight control as helpful.

Overall perceived usefulness of features is shown in Figure 53. Control group users perceived most features of activity reporting helpful (Figure 53a) contrasting with the intervention users who found those features unhelpful. The control group users evaluated the overall application's features as helpful (Figure 53c), as opposed to the intervention group users who felt application features as little helpful or unhelpful (Figure 53d). T-student tests were executed for each question related to usefulness, and no statistical significance was found.

6.3.2.4 Ease of Use

Table 25: Users Reported Ease of Use Answers of Physical Education Students

	Very Difficult				Very Easy
Evaluate the degree of ease of use of the activity schedule (Q9)	**1**	**2**	**3**	**4**	**5**
Control	0 (0.00%)	1 (16.67%)	2 (33.33%)	1 (16.67%)	2 (33.33%)
Intervention	0 (0.00%)	0 (0.00%)	1 (25.00%)	0 (0.00%)	3 (75.00%)

Total	**0 (0.00%)**	**1 (10.00%)**	**3 (30.00%)**	**1 (10.00%)**	**5 (50.00%)**
Evaluate the degree of ease of use of the meal reporting (Q15)	**1**	**2**	**3**	**4**	**5**
Control	0 (0.00%)	2 (33.33%)	1 (16.67%)	3 (50.00%)	0 (0.00%)
Intervention	1 (25.00%)	0 (0.00%)	1 (25.00%)	1 (25.00%)	1 (25.00%)
Total	**1 (10.00%)**	**2 (20.00%)**	**2 (20.00%)**	**4 (40.00%)**	**1 (10.00%)**
Evaluate the degree of ease of use of the physical activity report (Q18)	**1**	**2**	**3**	**4**	**5**
Control	0 (0.00%)	0 (0.00%)	3 (50.00%)	2 (33.33%)	1 (16.67%)
Intervention	0 (0.00%)	0 (0.00%)	0 (0.00%)	2 (50.00%)	2 (50.00%)
Total	**0 (0.00%)**	**0 (0.00%)**	**3 (30.00%)**	**4 (40.00%)**	**3 (30.00%)**
Evaluate the degree of ease of use of the weight reporting (Q21)	**1**	**2**	**3**	**4**	**5**
Control	0 (0.00%)	1 (16.67%)	2 (33.33%)	2 (33.33%)	1 (16.67%)
Intervention	0 (0.00%)	1 (25.00%)	0 (0.00%)	1 (25.00%)	2 (50.00%)
Total	**0 (0.00%)**	**2 (20.00%)**	**2 (20.00%)**	**3 (30.00%)**	**3 (30.00%)**
Evaluate the ease of understanding of the food consumption history chart (Q24)	**1**	**2**	**3**	**4**	**5**
Control	1 (16.67%)	2 (33.33%)	0 (0.00%)	2 (33.33%)	1 (16.67%)
Intervention	0 (0.00%)	0 (0.00%)	0 (0.00%)	2 (50.00%)	2 (50.00%)
Total	**1 (10.00%)**	**2 (20.00%)**	**0 (0.00%)**	**4 (40.00%)**	**3 (30.00%)**
Evaluate the ease of understanding of the history chart of physical activity practice (Q25)	**1**	**2**	**3**	**4**	**5**
Control	0 (0.00%)	1 (16.67%)	2 (33.33%)	2 (33.33%)	1 (16.67%)
Intervention	0 (0.00%)	1 (25.00%)	0 (0.00%)	1 (25.00%)	2 (50.00%)
Total	**0 (0.00%)**	**2 (20.00%)**	**2 (20.00%)**	**3 (30.00%)**	**3 (30.00%)**
Evaluate the ease of understanding of the weight history chart (Q26)	**1**	**2**	**3**	**4**	**5**
Control	0 (0.00%)	1 (16.67%)	2 (33.33%)	1 (16.67%)	2 (33.33%)

Intervention	0 (0.00%)	1 (25.00%)	0 (0.00%)	1 (25.00%)	2 (50.00%)
Total	**0 (0.00%)**	**2 (20.00%)**	**2 (20.00%)**	**2 (20.00%)**	**4 (40.00%)**

Source: Own authorship

The ease of use of the application perceived by users is described in Table 25 with answers to five points Likert scale questions, beginning from very difficult (1) to very easy (5). The major part of users found the scheduling activities feature easy to use, however, there were more neutral control group users than intervention group users. Users found the mealreporting feature easy to use, 50% in both groups, which was the lowest ease of use evaluation from all questions. Weight reporting was found easy to use by the majority of both groups, with 75% of the intervention group users understanding this feature as easy to use. All users of the intervention group found the physical report activity and food consumption history chart easy to use. These were the best-evaluated features from all questions, having overall ease of use agreement of 70%. Physical activity practice and weight history charts were also found easy to use by both groups. More intervention group users found these features easy to use when compared with the control group users.

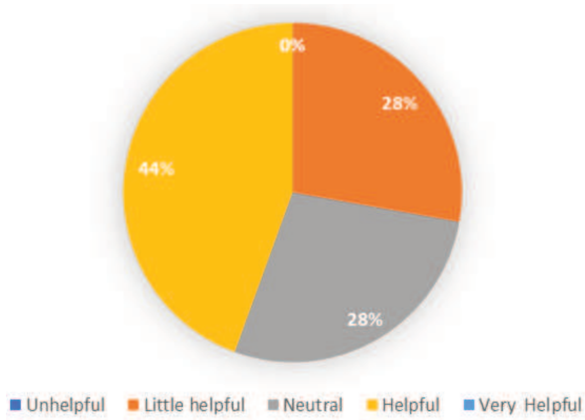
Intervention users perceived the general easiness of use of activity reports, history charts and all features better than the control group users (Figure 54). In this case, 83% of features were found easy to use by the intervention group users, while control group users found 50% of the features easy to use. T-student tests were executed to check if was significant differences in the answers given by users from the control and intervention groups. With exception to general perceived easy to use, no statistical significance was found, indicating that in most part answers given by the two groups to different features were not significantly different in most cases. However, when analyzed as a whole, easiness perception from the intervention group users is significantly greater than the control group users with a $p - value = 0.02$.

6.3.3 Twitter Profiles Behavior's Analysis on Physical Education Students Usage

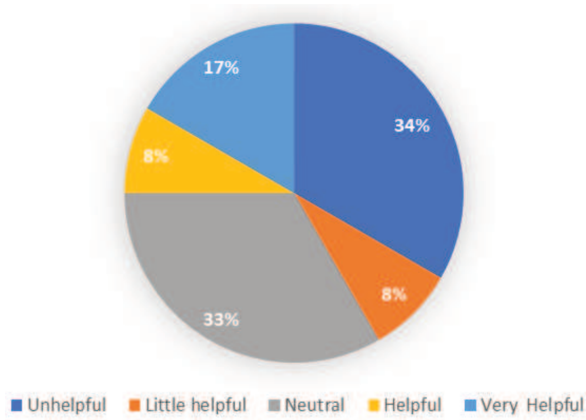
In this experiment the selected Twitter profiles were monitored from 04/16/2019 to 06/10/2019, totalizing eight weeks of observation. The first observation was done four weeks before con- tacting the profiles. A total of 1,800 messages were processed, 141 messages were identified as NCDs prevention messages and 139 passed the quality check. Profiles were contacted on four occasions on days 05/14, 05/21, 05/28, and 06/05. The contacts were done to inform about the new experiment, updates on the influence ranking, and also to encourage the selected Twitter profiles in posting NCDs prevention-related messages.

Figure 50 shows the distribution of the messages sent by Twitter profiles throughtime. The solid vertical line on the plot indicates the day when the first contact occurred, the dashed verti- cal lines indicate the subsequent contacts, the red dots represent messages that did not pass the

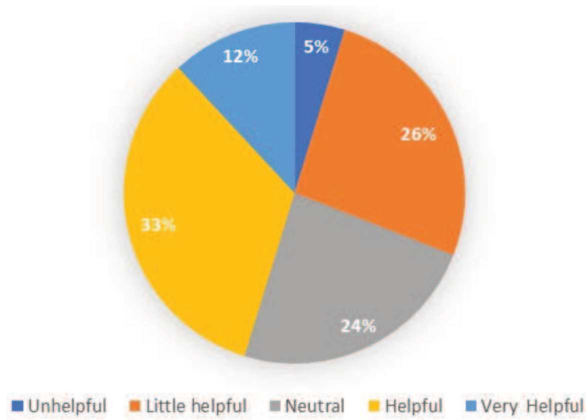
Figure 53: General Perceived Usefulness of Physical Education Students



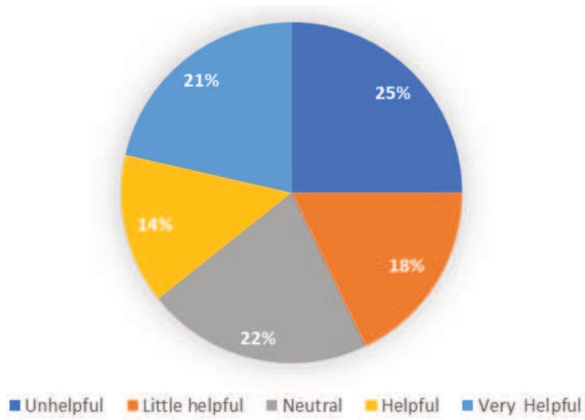
(a) General Perceived Usefulness of the Activities Report by the Physical Education Students Control Group



(b) General Perceived Usefulness of the Activities Report by the Physical Education Students Intervention Group



(c) General Perceived Usefulness by the Physical Education Students Control Group



(d) General Perceived Usefulness by the Physical Education Students Intervention Group

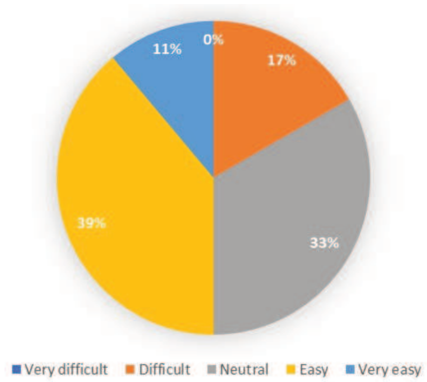
Source: Own authorship

quality check, and the blue dots represent messages that passed the quality check. Again just the profiles @opiniaomedica and @BuscarSaude accessed the rank. The profile @opiniaomedica accessed the rank on three occasions (05/14, 05/15 and 05/21), while @BuscarSaude accessed only on one occasion (05/22).

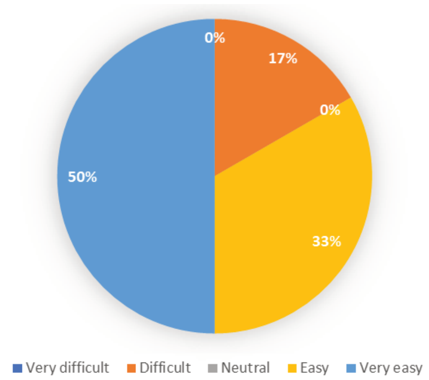
Five changes occurred in the rank position after contacting the profiles. At the second week, the profile @SaudeMG passed from the sixth position to the fifth position, and @realfoodbrasil passed from the seventh to the sixth position to. At the third week, @realfoodbrasil passed to the fifth position, and the profile @BuscarSaude has caught the first position from the profile @minsaude. At the last week, @opiniaomedica passed from the seventh position to the sixth position.

Table 26 presents the message count and the daily message sent by each profile before and after the first intervention. Table 26 also presents the results of the causal inference with Bayesian structural time-series model test aiming to detect changes on the posting behavior of

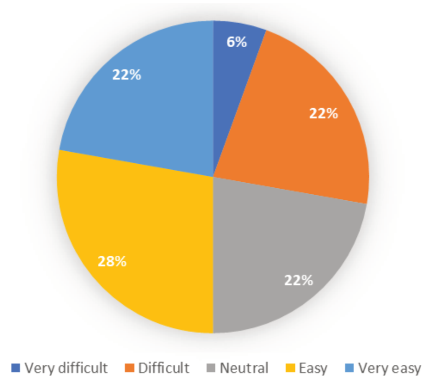
Figure 54: General Perceived Ease of Use by Physical Education Students



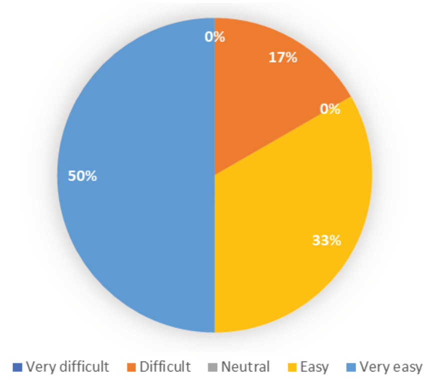
(a) General perceived ease of use of activity report by the Physical Education Students Control Group



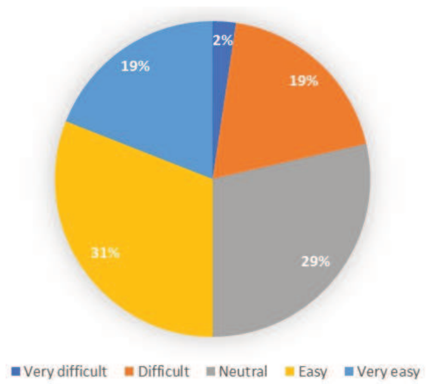
(b) General perceived ease of use of activity report by the Physical Education Students Intervention Group



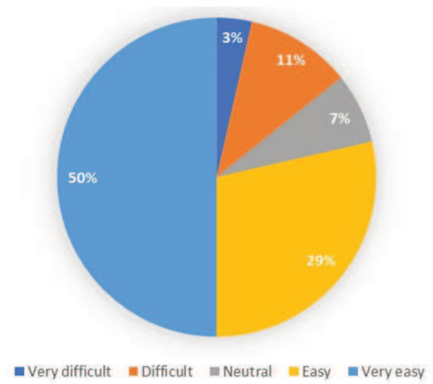
(c) General perceived ease of use of history charts by the Physical Education Students Control Group



(d) General perceived ease of use of history charts by the Physical Education Students Intervention Group



(e) General perceived ease of use by the Physical Education Students Control Group

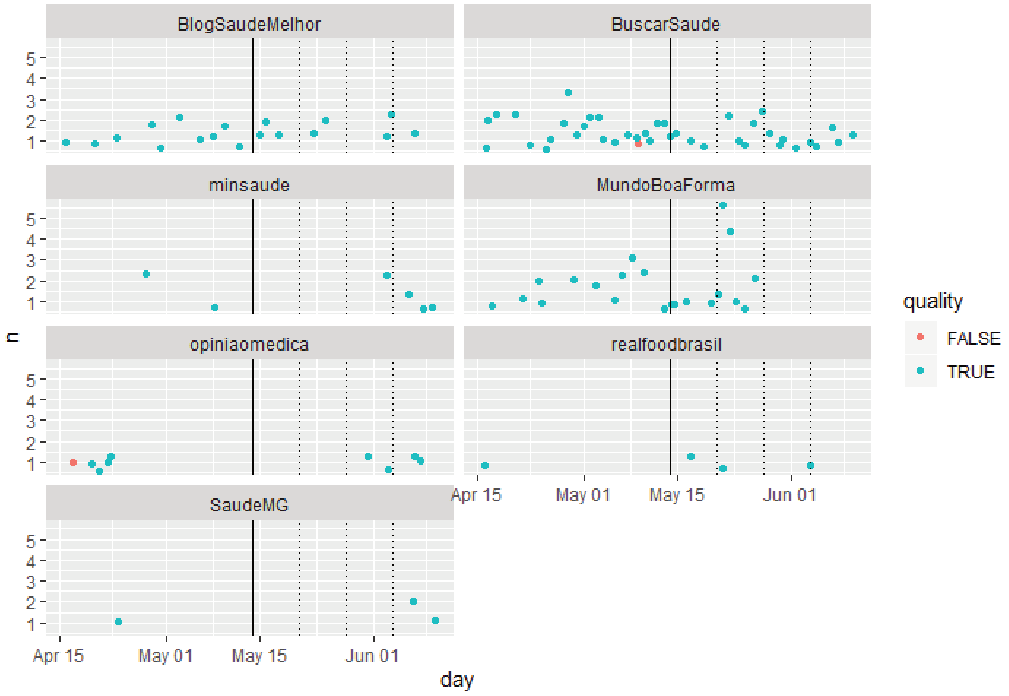


(f) General perceived ease of use by the Physical Education Students Intervention Group

Source: Own authorship

each profile. In general, most of the test results did not show statistical significance. The only exception was the profile @BuscarSaude in which test resulted in a p-value lower than 0.05, showing evidence that the intervention played an effect on this profile. It is worth noting that

Figure 55: Distribution of the sending of non-communicable diseases prevention message by the monitored Twitter profiles on the physical education students experiment



Source: Own authorship

rather than increasing message posting, this profile posted fewer messages on average afterthe intervention. Although the test result on the @SaudeMG profile did not demonstrate statistical significance it was very close to it. However, the number of messages sent by this profile was low in comparison to others profiles like @BlogSaudeMelhor, @BuscarSaude, and @Mundo-BoaForma, hence the low p-value result may be due to random fluctuations unrelated to the intervention.

6.4 Using the Pompilos Onto for Connection Recommendation

The ontology presented in Chapter 4 was used to find suitable nodes for connection recommendation. With the aid of the model presented in Section 6.1.5 a real social network with four nodes was inferred from the data collected in the first experiment (Figure 56). This social network is formed by two clusters: a coworker cluster composed by three individuals and a residence cluster. Individuals gender and body mass index (bmi) were acquired from data informed in the My U'Ductor application, whereas smoking habits, happiness information and tie types were acquired by interviews with the individuals conducted by the researcher. Happiness was inferred by answers gave by individuals to CES-D questionnaire (RADLOFF, 1977).

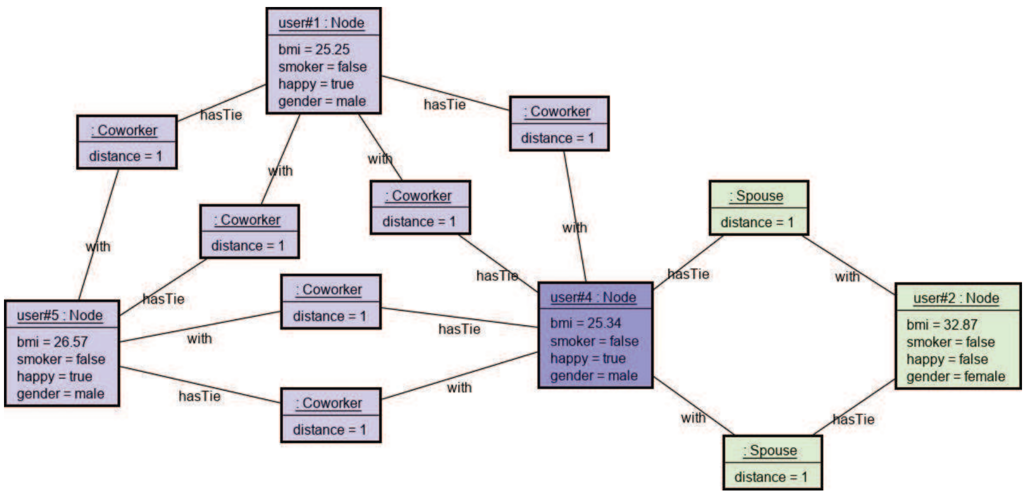
Axioms for detecting influence of noncommunicable diseases risk factors spreading were based on the works from Christakis and Fowler (CHRISTAKIS; FOWLER, 2007, 2008; FOWLER; CHRISTAKIS, 2008). As the network was composed mainly by coworkers and spouse

Table 26: Causal Inference Test on Physical Education Students Usage

Profile	Before Intervention			After Intervention			Causal Test (p-value)
	Messages	Daily Mean	Standard Deviation	Messages	Daily Mean	Standard Deviation	
BlogSaudeMelhor	13.00	0.45	0.69	11.00	0.41	0.69	0.311
BuscarSaude	**34.00**	**1.17**	**0.85**	**21.00**	**0.78**	**0.70**	**0.049**
DetoxReceitas	0.00	0.00	0.00	0.00	0.00	0.00	0.000
juntoemisturadi	0.00	0.00	0.00	0.00	0.00	0.00	0.000
minsaude	3.00	0.10	0.41	5.00	0.19	0.48	0.268
MundoBoaForma	19.00	0.66	0.90	18.00	0.67	1.39	0.422
opiniaomedica	5.00	0.17	0.38	4.00	0.15	0.36	0.391
realfoodbrasil	1.00	0.03	0.19	3.00	0.11	0.32	0.231
rederms	0.00	0.00	0.00	0.00	0.00	0.00	0.000
saudavelcomida	0.00	0.00	0.00	0.00	0.00	0.00	0.000
saudavelebarato	0.00	0.00	0.00	0.00	0.00	0.00	0.000
SaudeMG	**1.00**	**0.03**	**0.19**	**3.00**	**0.11**	**0.42**	**0.074**

Source: Own authorship

Figure 56: Generated Real Social Network



Source: Own authorship

ties, the following axioms were considered and converted to SPARQL 1.1 Update expressions[11]:

- 37% of increased probability in becoming obese if spouse is obese (Figure 57a);
- 15% of increased probability in becoming happy if alters are happy (Figure 57b);
- 8% of increased probability in becoming happy if spouse is happy (Figure 57c);
- 61% of increased probability in becoming smokers if alters are smokers (Figure 57d);

[11]https://www.w3.org/TR/2013/REC-sparql11-update-20130321/

- 34% of increased probability in quitting smoking if spouse is nonsmoker (Figure 57e).

Complementarily, Table 27 shows the axiom of equivalence created to aid the addition of influence spreading axioms.

Figure 57: Influence Axiom Additions

```
INSERT INTO GRAPH
<https://cloud.uductor.com/social-network>  {
   $influence a pompilos:Influence .
   $alter pompilos:spreads $influence  .
   $influence pompilos:on $ego .
   $influence pompilos:ofProvoking $susceptibility .
}
FROM <https://cloud.uductor.com/social-network>
WHERE {
   $ego pompilos:hasTie $tie .
   $tie a pompilos:ObeseSpouse .
   $tie pompilos:with $alter .
   $susceptibility pompilos:hasProbability $probability .
   $susceptibility a pompilos:Become_Obese.
   $probability pompilos:increase 37 .
   BIND(UUID()) AS $influence) .
}
```

(a) Obesity Spreading Influence by Spouse

```
INSERT INTO GRAPH
<https://cloud.uductor.com/social-network>  {
   $influence a pompilos:Influence .
   $alter pompilos:spreads $influence  .
   $influence pompilos:on $ego .
   $influence pompilos:ofProvoking $susceptibility .
}
FROM <https://cloud.uductor.com/social-network>
WHERE {
   $ego pompilos:has1Degree $alter .
   $alter a pompilos:Happy .
   $susceptibility pompilos:hasProbability $probability .
   $susceptibility a pompilos:Become_Happy .
   $probability pompilos:increase 15 .
   BIND(UUID()) AS $influence) .
}
```

(b) Happiness Spreading Influence by First Degree Alters

```
INSERT INTO GRAPH
<https://cloud.uductor.com/social-network>  {
   $influence a pompilos:Influence .
   $alter pompilos:spreads $influence  .
   $influence pompilos:on $ego .
   $influence pompilos:ofProvoking $susceptibility .
}
FROM <https://cloud.uductor.com/social-network>
WHERE {
   $ego pompilos:hasTie $tie .
   $tie a pompilos:HappySpouse .
   $tie pompilos:with $alter .
   $susceptibility pompilos:hasProbability $probability .
   $susceptibility a pompilos:Become_Happy .
   $probability pompilos:increase 8 .
   BIND(UUID()) AS $influence) .
}
```

(c) Happiness Spreading Influence by Spouse

```
INSERT INTO GRAPH
<https://cloud.uductor.com/social-network>  {
   $influence a pompilos:Influence .
   $alter pompilos:spreads $influence  .
   $influence pompilos:on $ego .
   $influence pompilos:ofProvoking $susceptibility .
}
FROM <https://cloud.uductor.com/social-network>
WHERE {
   $ego pompilos:has1Degree $alter .
   $alter a pompilos:Smoker .
   $susceptibility pompilos:hasProbability $probability .
   $susceptibility a pompilos:Become_Smoker .
   $probability pompilos:increase 61 .
   BIND(UUID()) AS $influence) .
}
```

(d) Smoking Spreading Influence by First Degree Alters

```
INSERT INTO GRAPH
<https://cloud.uductor.com/social-network>  {
   $influence a pompilos:Influence .
   $alter pompilos:spreads $influence  .
   $influence pompilos:on $ego .
   $influence pompilos:ofProvoking $susceptibility .
}
FROM <https://cloud.uductor.com/social-network>
WHERE {
   $ego pompilos:hasTie $tie .
   $ego a pompilos:Smoker .
   $tie a pompilos:NonSmokerCoworker .
   $tie pompilos:with $alter .
   $susceptibility pompilos:hasProbability $probability .
   $susceptibility a pompilos:Quit_Smoking .
   $probability pompilos:increase 34 .
   BIND(UUID()) AS $influence) .
}
```

(e) Quit Smoking Spreading Influence by Coworker

Source: Own authorship

The suitable nodes for connection recommendation are find by answering which nodes are suggested to form a new connection, in order to it obtains a lower likelihood of obesity, to lower the likelihood of smoking, or to obtain a greater likelihood of happiness. To answer these, each question was converted to a SPARQL 1.1. query [12] as presented in Figure 58.

[12] https://www.w3.org/TR/2013/REC-sparql11-query-20130321/

Table 27: Axioms of Equivalence

Class	Axiom of Equivalence	Description
SpouseObeseInfluencer	Obese and (hasTie some Coresidence_Spouse)	Obese coresidence spouse.
ObeseSpouse	Spouse and (with some SpouseObeseInfluencer)	Tie referring to an obese spouse.
SpouseHappyInfluencer	Happy and (with some SpouseHappyInfluencer)	Happy coresidence spouse.
HappySpouse	Coresidence_Spouse and (with some SpouseHappyInfluencer)	Tie referring to a happy spouse.
NonSmoker	Node and (smoker value false)	A nonsmoker node.
CoworkerQuitSmokingInfluencer	NonSmoker and (hasTie some Coworker)	A nonsmoker coworker.
NomSmokerCoworker	Coworker and (with some CoworkerQuitSmokingInfluencer)	Tie referring to a nonsmoker coworker

Source: Own authorship

Recommended nodes for lowering the probability of obesity are obtained by selecting those second degree nodes of an ego that do not spreads obesity influence (Figure 58a). Recommended nodes for increasing the probability of becoming happy, or the increasing the probability in quitting smoking are obtained by selecting those second degree nodes that have greater probability in spreading happiness or nonsmoking behaviors (Figure 58b and Figure 58c). As a result, user#1 and user#5 are recommended to user#2 for both lowering the probability of obesity and for increasing the probability of happiness.

Figure 58: Recommendation Queries

```
SELECT $ego $recommendedAlter
FROM <https://cloud.uductor.com/social-network>
WHERE {
  $ego pompilos:has2Degree $recommendedAlter .
  FILTER NOT EXISTS {
    $recommendedAlter pompilos:spreads $influence .
    $influence pompilos:ofProvoking $susceptibility .
    $susceptibility a pompilos:Become_Obese .
  }
}
```

(a) Recommended Nodes Query for Lowering Obesity Probability

```
SELECT $ego $recommendedAlter $increase
FROM <https://cloud.uductor.com/social-network>
WHERE {
  $ego pompilos:has2Degree $recommendedAlter .
  $recommendedAlter pompilos:spreads $influence .
  $influence pompilos:ofProvoking $susceptibility .
  $susceptibility a pompilos:Become_Happy .
  $susceptibility pompilos:hasProbability $probability .
  $probability pompilos:increase $increase .
}
ORDER BY DESC($increase)
```

(b) Recommended Nodes Query for Increasing Happiness Probability

```
SELECT $ego $recommendedAlter $increase
FROM <https://cloud.uductor.com/social-network>
WHERE {
  $ego pompilos:has2Degree $recommendedAlter .
  $recommendedAlter pompilos:spreads $influence .
  $influence pompilos:ofProvoking $susceptibility .
  $susceptibility a pompilos:Quit_Smoking .
  $susceptibility pompilos:hasProbability $probability .
  $probability pompilos:increase $increase .
}
ORDER BY DESC($increase)
```

(c) Recommended Nodes Query for Increasing Quit Smoking Probability

Source: Own authorship

7 CONCLUSIONS

Social influence on health is not a new theme, being addressed already by Émile Durkheim in the nineteenth century (DURKHEIM, 1897; BERKMAN et al., 2000). More recently, Christakis and Fowler used Framingham Heart Study (HISTORY OF THE FRAMINGHAM HEART STUDY, 2016) data to investigate influence of social networking in the health of individuals. They found evidence of social network influence in weight gain (CHRISTAKIS; FOWLER, 2007), smoking cessation (CHRISTAKIS; FOWLER, 2008) and in the feeling of happiness (FOWLER; CHRISTAKIS, 2008).

Social support may be understood as the beneficial social influence on health given by peers **structurally**, through the availability of resources, or **functionally**, by the generation of perception of trust and encouragement (HWANG et al., 2014). In computing, social support on health is promoted mainly by the use of Internet driven platforms such as blogs, chats, fo- rums, wikis, or video sharing. These tools enable the exchange of knowledge between patients, caregivers, and physicians as they have the ability of increasing patients' confidence and self-efficacy. However, the power of personal sensing platforms such smartphones and gadgets in promoting social support seems to be underused.

This thesis proposed **Pompilos**, a model for social aware preventive care of NCDs. Part of Pompilos goals is to help individuals percept the influence generated by them in health of others, and so, prompting them to have better health choices and so influence their peers in doing the same. A prototype of the model was developed to provide continuous monitoring, access to real-time information, communication with health professionals, support disease management, and fostering social support (for example, by recommending beneficial health resources aiming to improve users' health behaviors).

For evaluating the model a mobile and online assistant for diets, weight management, and physical activity practice was developed and tested in two experiments. The first experiment took 45 users from the author's contacts for one month and a half. In the second experiment, 16 students from the University of Vale do Sinos Physical Education course participated. Also, 16 Twitter profiles, who main focus was to share messages regarding well-being were monitored to provide social support to the platform users in both experiments.

7.1 Contributions

Models for detecting the diffusion of information in social networks, as also to identify influence that nodes exert on others already exists (CHEN; LAKSHMANAN; CASTILLO, 2013; CHRISTAKIS; FOWLER, 2011; GARCIA-HERRANZ et al., 2014; CHRISTAKIS; FOWLER, 2010; TANG et al., 2009; GOYAL; BONCHI; LAKSHMANAN, 2010). Social data is used for knowledge discovery, and was successfully applied in studies related to NCDs to detect diseases outbreaks (LEE; AGRAWAL; CHOUDHARY, 2015; RAM et al., 2015; ZHANG et al., 2016), health risks (PAUL; DREDZE, 2011; CULOTTA, 2014; WEBER; MEJOVA, 2016) or self-disclosure discourses (BALANI; DE CHOUDHURY, 2015). Internet is the default platform for offering social support for NCDs prevention and care. This support is generally offered by forums, social networks, wikis, blogs, chats and shared videos. Some studies used data collected from patients to detect the need of interventions (MARTIN et al., 2011; LAN et al., 2012; SCHWARTZ et al., 2014; ALSHURAFA et al., 2014b; SIDERIS et al., 2015). For example, Alshurafa et al. (2014) used data collected from patients answers in a smartphone app to send them notifications for behavior change. However, none of the studied works used social data or data automatically collected from patients to provide social support based on the influence received from their social network.

The Pompilos model is supported by the idea that social behaviors spread through people social networks. In this way, the model dynamics relies on certain activities, which are: collecting data from users, generating users' social networks, inferring users' profiles, and training models for detecting and computing social influence on spreading NCDs risk factors on people. By accomplishing these activities it is possible to recommend resources for health improvement, as also to aware people about their influence on the health of others. The model based on the achievement of these activities is the main contribution of this thesis since this concept was not explored in any of the studied works. The presented model is also able to integrate the ever growing amount of computational devices and data for improvement of social support.

Moreover, this work has some other contributions. First, a mapping study explained how computing and social data supports NCDs care and which computing models are used for social promotion. Then, it presented an ontology for detecting the spread of obesity, happiness and smoking cessation (which was used as a model for influence detection on social networks). Both works were already shared with the scientific community (VIANNA; BARBOSA, 2017; VIANNA et al., 2018). A prototype of the proposed model was developed, in which its main advantage is the easiness in distributing its components. A machine learning model for detecting messages related to preventive NCDs messages was trained and developed. Finally, a web application for the management of preventive NCDs activities was provided publicly. This application does not only allow the user to manage preventive NCDs activities but also sends to them messages related to a healthy diet, physical activities practice and weight management.

7.2 Final Considerations

As shown in this thesis, pervasive software can be used to build solutions that positively enhance well-being feelings and social participation in health care, which is improved when the different types of data are applied to empower users' capabilities. The proposed archi- tecture is technically feasible and the model is not complex to understand. Furthermore, it can scale, guarantee privacy, and provide near real time updates which eases the creation of context aware applications. The model was designed for accomplish the requirements of preventive care of NCDs but can be extended to other domains. As the communication between the model's elements is based on the REST architecture style, a great part of its implementation can be realized using existent web technologies. However, other architectures can be used to accomplish the requirements like, for example, Enterprise Service Bus (ESB), which is an architectural framework that provides a common interface for integrating applications to different services (BHADORIA; CHAUDHARI, 2018; BHADORIA; CHAUDHARI; VIDANAGAMA, 2018; SHARMA; BHADORIA; DIXIT, 2017).

It is worth adding that for applications built on the model to succeed it is desirable that Internet information and services be semantically annotated to facilitate the understanding and finding of these resources by software agents. This will alleviate the time needed to design, built, train, test and validate probabilistic models for information classification necessary for influence detection, profile generation and context aggregation. These models usually need large datasets, are empirically designed, and have limitations in their accuracy. On the other hand, semantic annotations (once they exist) are logically inferred and are easily understood by people, as also enable the creation of more compatible applications.

The main question which this thesis tries to answer is if the recommendation of beneficial social resources and the awareness of the influence of individual behaviors in the health of others improve the engagement in NCDs prevention. To achieve this a web health assistance application was provided to users who agreed in collaborate with the research. The prototype was accessed by a total of 61 users in two experiments. The first experiment participated 45 users, while the second experiment had 16 users. Each user was assigned to one of the two variants of the application. In the control variant, users could schedule to receive reminders of activities related to preventive care of NCDs like weight control, food intake and physical activity practice. The users could also log what they did in the scheduled activities and access history chart related to their progress. The intervention variant had all features from the control variant but also provided to its users with relevant messages about the prevention of NCDs. These messages were collected from 16 monitored Twitter profiles that were previously analyzed for determining if they shared messages related to physical activity practice, healthy diet or weight control.

In a practical perspective, general days of application usage and the number of interactions with the application were greater among the users from the intervention groups on the first experiment, but this usage pattern changed in the second experiment. The feature for reporting meals was used for longer by control group users in the first experiment, but users from the intervention group used this feature for more times in the second experiment. Control group users from both experiments used the weight log feature for more time. Intervention users have used physical activity logs more and for longer on both experiments. The query of history charts was accessed more times by intervention users in both experiments. However, statistical tests do not show significant differences between mean access related to these features. This may be due to the differences in the usage profile of users, that is, some users were more focused on using the application as a tool for their daily care, while others were just curious to see how the application worked, as demonstrated by the high standard deviation for the number of used days and interactions.

It is important to add, even that history chart queries were below expected significance, this value was much lower than others and somehow very close to the expected significance (i.e. $p - value < 0.05$) on both experiments. The repetition of this phenomenon may indicate that the fact of receiving messages played a different effect on users, indicating that they were more concerned in following their activity history, but it is not possible to guarantee that this effect happened with high confidence. To reinforce this idea, a connection between activity history usage and interaction with messages was found by the application of a robust fitting linear model in the time series data of these variables. That test result was observed in both experiments, showing a possible trend that users who interacted with messages were more concerned in following a healthy behavior.

After one month of application usage, users were invited to answer a survey. A total of 14 users from the first experiment and 10 from the second completed the survey. Few users have reported technical problems in using the application. Some users reported not knowing about the possibility of liking messages or to follow its links. The liking feature is similar to Twitter like feature, and the links are presented with the default browser behavior. Possibly, users whom unknown these features read the message by smartphone notification tray, where these features were not present. Add links and like button in the notification tray is a possible enhancement of the application.

Based on survey reports, application access on the first experiment from intervention users was made in most of the part by smartphones, when compared to control group users. However, this relationship did not occur in the second experiment. Further investigations must be carried to understand if the proximity with devices can explain an improvement in the number of interactions with the application. As part of the surveys were answered unanimously is not possible to relate the platform type to the number of user interaction. The feeling of relatedness with the application seems to have little effect on the intervention group in both experiments. On the first experiment, this group had a great proportion of neutral users than the control group, although positive feelings proportions were the same in both groups. Still, in the second experiment, relatedness showed uniformity in both groups. It is important to note that answers reporting that the application was related or very related by Physical Education students did not diverge from normal users. In both experiments, the sum of these answers was equal to 50%, showing that the goals of the application are related to the general public.

The general perceived usefulness of application was greater in the control group than in the intervention group in both experiments. Nonetheless, the distribution of answers regarding usefulness diverged on the experiments. Most of the users from the first experiment tended to look at the application features as useful, while Physical Education students tended to look at the application from neutral to unhelpful. The professional profile may have influenced this difference in that case, once the activities managed by the application might be incorporated in the routine of the students.

It is possible that receiving preventive NCDs care messages played a negative effect on the feeling of the perceived usefulness of some features. For example, in the first experiment, the answers about the perceived usefulness of the scheduling activities feature had 83.3% of positive answers by the control group in contrast to just 37.5% of positive answers by the intervention group. By its turn, in the second experiment, the positive general perceived usefulness of activities report features was of 44% on the control group but just 25% on the intervention group.

Furthermore, answers to activity reporting features were, in general, similar over the groups of the first experiment, with distinction to the type of reporting activity. In this experiment, the control group users were more positive about physical activity report and intervention users were more positive about weight report. Curiously, this does not reflect part of the reported user activities interest where 83.3% of control users reported interest in managing physical activities practice, and just 16.6% of intervention users reported interest in weight management. On the second experiment, this dissociation between intention and usefulness perception was related to the meal reporting which was of interest to 60% of users, but just 20% of users understood it as useful. The understanding of this dissociation will deserve further investigations.

Survey answers about perceived usefulness in querying activity history from users of the first experiment were answered more positively by the control group in spite of the intervention group have had greater usage of this feature. This may be due to the low rate of users who answered the survey, about 30% of the registered users. Nevertheless, survey results may not be invalidated as its bias is related to comparisons with usage data. In comparison, in the second experiment, 50% of users from the intervention group understood this feature as useful, while 33.33% of control group users felt the same. Hence, the results were more related to this feature usage in the second experiment.

In the first experiment, recommended messages were found useful by 75% of the intervention users, but there were distinctions about which type of message has more relevance, tending to a preference towards healthy diets messages. Nonetheless, the overall perception of usefulness was taken as neutral by users of the second experiment. However, the perception of the usefulness of messages by topic was of 50%. This may be an indication that the content of messages received by the users was not interpreted as associated with topics such as healthy diets, physical activity practice, or weight management. Nevertheless, this does not invalidate the usefulness of this feature, as 52.78% of users from both experiments, indicating that the receiving of message relating to topics is useful.

Perceived ease of use took into account the following features: activity scheduling, activities reporting (meal, physical activity practice, and weight) and the query to history charts (consumed meals, realized physical activities and weight). With exception to activities reporting, all other features were positively evaluated by the respondent of both groups of the two experiments. Users from the control group of the first experiment found the meal and physical activities reporting difficult to use especially the meal reporting feature. This perception may be reinforced by the fact that these users had greater usage of this feature. Besides, some open answers given by users from both experiments about the meal reporting seem to reflect that users found its approach (based on food portions) overly complex affecting their evaluation, for example:

> "*The app could contain a facility to fill in the information (lunch photo for example); In my meal follow-up, it was much higher than the daily portion intakes so I believe it is not correct, it was confusing filling.*" (Control user #3, first experiment)

> "*It would be more interesting if the reporting process were more friendly. The tool provides tips (captions) based on total consumption / day and not per meal, which ends up creating confusion at the time of filling. If the process is guided (at the beginning it is informed which meals are by default and the tool already shows examples of foods for category - ex: breakfast - carbohydrates, meats, etc.*" (Control user #5, first experiment)

> "*It was cool. But the issue of portions I got in the way a bit. and does kind of activity. Maybe if an interface were more visual (where only this choice) and if I had set a pattern myself and then just confirming ... That would be easier.*" (Intervention user #5, first experiment)

> "*It is difficult to put the food in portions.*" (Control user #6, second experiment)

Finally, contact with each monitored Twitter profile was made by instant message or e-mail. The content of the first contact message explained the project and presented a link to the profiles rank. Subsequent messages to the profiles provided information about ranking changes and a link to the profiles rank. The contact served as an intervention on the Twitter profiles aiming at evaluating if their behaviors would change by awareness their influence on NCDs prevention care activities of others. After the intervention, two Twitter profiles have accessed the rank, and the two have changed their behaviors with high statistical significance.

The first change was done by the profile @opiniaomedica during the first experiment. Different from the other profiles, which in great part were related to health portals or health departments from governments, this was a personal profile of physician. The second change was done by the profile @BuscarSaude during the second experiment. This profile also belongs to a health professional, in this case, a physical educator. It is possible that this personal distinction has powered the intervention, suggesting that health professionals are more concerned about their influence in the health of others. Unlike the profile @opiniaomedica which increased the number and quality of messages related to NCDs prevention, @BuscarSaude decrease the number of messages. This could be a strategy of the profile to raise its position in the rank, as the number of messages could penalize the rank position if users do not engage. However, this evidence is small and may be addressed in future works.

7.3 Future works

One of the findings of this thesis relies on the association between social media influence and preventive care follow-up of users in the form of queries of activities history. This finding relates to the improvement of engagement by the recommendation of beneficial social resources. Another finding is that at least two Twitter profiles have changed their behaviors after being aware of their influence in the preventive care activities of others. This second finding relates to the awareness of the influence of individual behaviors in the health of others.

Nevertheless, this study may not be understood as generally conclusive and is open to further improvements and following researches. The experiment presented in chapter 6 with the author's contacts was very loosely in the sense that users were not controlled. This has some advantages, as it elevates the users' perception of blindness, that is, their perception in forgiven being observed by the researcher (DALE et al., 2016), but hinders the collection of more cohesive data. Hence, the second experiment was done with a more controlled group with the support of health professionals. However, information about the effectiveness of the model, which is desirable, is being collected as an extension of the second experiment and will provide the interlink between application acceptance, usage, and its effectiveness. Additionally, the model could be adopted by public health departments as a way to improve NCDs preven- tion policies and check their effectiveness. This last proposed work is in the design phase in partnership with the Department of Health of the State of Rio Grande do Sul.

Finally, results on user perception of usefulness and ease of use pointed to a need of user experience improvement in regards to features of activity reporting, mainly about meal report which was poorly evaluated by the users.

7.4 Publications

Part of the findings reported in this thesis was already published. The mapping study documented in Chapter 3 is available for online access in the Journal of Telematics and Informatics (VIANNA; BARBOSA, 2017). The ontology described in Chapter 4 was published and presented in 2016 XLII Latin American Computing Conference (CLEI) (VIANNA et al., 2016) and extend for publication in International Journal of Metadata, Semantics and Ontologies (VIANNA et al., 2018). The model motivations and insights were published in IEEE Software (VIANNA; BARBOSA; PITTOLI, 2017). The general model design and architecture were published in Information Processing Letters (VIANNA; BARBOSA, 2019).

APPENDIX A CLASSIFICATION OF THE REVIEWED PAPERS

Table 28: Classification of the reviewed papers

Title	In	Type	Year	Classification	Social Support	Social Data	Computing Model
The Effects of a Web-Based Intervention on Psychosocial Well-Being Among Adults Aged 60 and Older With Diabetes (BOND et al., 2010)	The Diabetes Educator	Journal	2010	Controlled Trials	-	-	-
An Online Community Improves Adherence in an Internet-Mediated Walking Program. Part 1: Results of a Randomized Controlled Trial (RICHARDSON et al., 2010)	Journal of Medical Internet Research	Journal	2010	Controlled Trials	-	-	-
Integrating technology into standard weight loss treatment: A randomized controlled trial (B et al., 2013)	JAMA Internal Medicine	Journal	2013	Controlled Trials	-	-	-
Comparison of Veteran experiences of low-cost, home-based diet and exercise interventions (HOLTZ et al., 2014)	Journal of Rehabilitation Research & Development	Journal	2014	Controlled Trials	-	-	-
Effectiveness of a Web-Based Tailored Interactive Health Communication Application for Patients With Type 2 Diabetes or Chronic Low Back Pain: Randomized Controlled Trial (WEYMANN et al., 2015)	Journal Of Medical Internet Research	Journal	2015	Controlled Trials	-	-	-

Patient Journey Record Systems (PaJR): The Development of a Conceptual Framework for a Patient Journey System (MARTIN et al., 2011)	IGI Global	Chapter	2010	Frameworks and Systems	Yes	No	Yes
Implementation of complex adaptive chronic care: the Patient Journey Record system (PaJR) (MARTIN et al., 2012)	Journal of Evaluation in Clinical Practice	Journal	2012	Frameworks and Systems	Yes	No	Yes
WANDA: An End-to-end Remote Health Monitoring and Analytics System for Heart Failure Patients (LAN et al., 2012)	Wireless Health	Conference	2012	Frameworks and Systems	Yes	No	Yes
Accessible telehealth - Leveraging consumer-level technologies and social networking functionalities for senior care (DHILLON; WüNSCHE; LUTTEROTH, 2013)	International Conference on Human System Interaction	Conference	2013	Frameworks and Systems	Yes	No	Yes
Remote health monitoring: Predicting outcome success based on contextual features for cardiovascular disease (ALSHURAFA et al., 2014a)	Annual International Conference of the IEEE Engineering in Medicine and Biology Society	Conference	2014	Frameworks and Systems	Yes	Yes	Yes
Towards chronic emergency response communities for anaphylaxis (SCHWARTZ et al., 2014)	International Conference on Information Reuse and Integration	Conference	2014	Frameworks and Systems	Yes	No	Yes
A virtual aged care system: When health informatics and spatial science intersect (ROBERTSON et al., 2014)	Investing in E-Health: People, Knowledge and Technology for a Healthy Future	Chapter	2014	Frameworks and Systems	No	Yes	No

Health Care 2020: Re-engineering Health Care Delivery to Combat Chronic Disease (MILANI; LAVIE, 2016)	The Journal of American Medicine	Journal	2015	Frameworks and Systems	Yes	Yes	No
Remote Health Monitoring Outcome Success Prediction using Baseline and First Month Intervention Data (ALSHURAFA et al., 2016)	IEEE Journal of Biomedical and Health Informatics	Journal	2016	Frameworks and Systems	Yes	Yes	Yes
You Are What You Tweet: Analyzing Twitter for Public Health (PAUL; DREDZE, 2011)	International Conference Weblogs Social Media	Conference	2011	Knowledge Discovery	No	Yes	No
Discovering Health-related Knowledge in Social Media Using Ensembles of Heterogeneous Features (TUAROB et al., 2013)	ACM International Conference on Information & Knowledge Management	Conference	2013	Knowledge Discovery	No	Yes	No
A Framework for Predicting Adherence in Remote Health Monitoring Systems (ALSHURAFA et al., 2014b)	Wireless Health	Conference	2014	Knowledge Discovery	Yes	Yes	Yes
Estimating County Health Statistics with Twitter (CULOTTA, 2014)	Annual ACM Conference on Human Factors in Computing Systems	Conference	2014	Knowledge Discovery	No	Yes	No
Mining Social Media Streams to Improve Public Health Al- lergy Surveillance (LEE; AGRAWAL; CHOUDHARY, 2015)	International Conference on Advances in Social Networks Analysis and Mining	Conference	2015	Knowledge Discovery	No	Yes	No
Effects of Coaching on Adherence in Remote Health Monitoring Systems: Analysis and Prediction of Participant Adherence (SIDERIS et al., 2015)	Wireless Health	Conference	2015	Knowledge Discovery	Yes	Yes	Yes

Web-Based Surveillance of Public Information Needs for Informing Preconception Interventions (D'AMBROSIO et al., 2015)	PLOS One	Journal	2015	Knowledge Discovery	No	Yes	No
Predicting Asthma-Related Emergency Department Visits Using Big Data (RAM et al., 2015)	IEEE Journal of Biomedical and Health Informatics	Journal	2015	Knowledge Discovery	No	Yes	No
You Tweet What You Eat: Studying Food Consumption Through Twitter (ABBAR; MEJOVA; WEBER, 2015)	Annual ACM Conference on Human Factors in Computing Systems	Conference	2015	Knowledge Discovery	No	Yes	No
Detecting and Characterizing Mental Health Related Self-Disclosure in Social Media (BALANI; DE CHOUDHURY, 2015)	Annual ACM Conference on Human Factors in Computing Systems	Conference	2015	Knowledge Discovery	Yes	Yes	No
Adapting Graph Theory and Social Network Measures on Healthcare Data: A New Framework to Understand Chronic Disease Progression (KHAN; UDDIN; SRINIVASAN, 2016)	Proceedings of the Australasian Computer Science Week Multiconference	Conference	2016	Knowledge Discovery	No	No	No
Extracting Signals from Social Media for Chronic Disease Surveillance (ZHANG et al., 2016)	International Conference on Digital Health	Conference	2016	Knowledge Discovery	No	Yes	No
Crowdsourcing Health Labels: Inferring Body Weight from Profile Pictures (WEBER; MEJOVA, 2016)	International Conference on Digital Health	Conference	2016	Knowledge Discovery	No	Yes	No
Rule-Based Modeling of Chronic Disease Epidemiology: Elderly Depression as an Illustration (CHIêM; MACQ; SPEYBROECK, 2012)	PLOS One	Journal	2012	Simulation Models	-	Yes	-

SimNCD: An agent-based formalism for the study of noncommunicable diseases (AZIZA et al., 2016)	Engineering Applications of Artificial Intelligence	Journal	2016	Simulation Models	-	Yes	-
Social support in an Internet weight loss community (HWANG et al., 2010)	International Journal of Medical Informatics	Journal	2010	Social Media Usage Analysis	-	-	-
The association between weight loss and engagement with a web-based food and exercise diary in a commercial weight loss programme: a retrospective analysis (JOHNSON; WARDLE, 2011)	International Journal of Behavioral and Physical Activity	Journal	2011	Social Media Usage Analysis	-	-	-
Structural social support predicts functional social support in an online weight loss programme (HWANG et al., 2014)	Health Expectations	Journal	2012	Social Media Usage Analysis	-	-	-
Cyberhugs: Creating a Voice for Chronic Pain Sufferers Through Technology (BECKER, 2012)	Cyberpsychology, Behavior, and Social Networking	Journal	2013	Social Media Usage Analysis	-	-	-
Digital health communities: The effect of their motivation mecha- nisms (BA; WANG, 2013)	Decision Support Systems	Journal	2013	Social Media Usage Analysis	-	-	-
Using Online Health Communities to Deliver Patient-Centered Care to People With Chronic Conditions (EIJK et al., 2013)	Journal of Medical Internet Research	Journal	2013	Social Media Usage Analysis	-	-	-
Youtube How-to-Videos (HtV) (LIBIN et al., 2014)	International Conference on Computer Supported Education	Conference	2014	Social Media Usage Analysis	-	-	-

Social media for empowering people with diabetes: Current status and future trends (GOMEZ-GALVEZ; MEJíAS; FERNANDEZ-LUQUE, 2015)	Annual International Conference of the IEEE Engineering in Medicine and Biology Society	Conference	2015	Social Media Usage Analysis	-	-	-
Collecting Family Health History using an Online Social Network: a Nationwide Survey among Potential Users (WELCH et al., 2015)	AMIA Annual Symposium	Conference	2015	Social Media Usage Analysis	-	-	-

Source: Own authorship

APPENDIX B MY U'DUCTOR SURVEY QUESTIONS

Table 29: Survey Questions

Question	Description	Feature	Subject	Type	Accepted Values	Target
Q1	Indicate the degree of relationship of the overall goals of the application to your routine	General Goals	User report	Five points likert scale	Unrelated (1) to Very related (5)	All
Q2	Which of these activities interest you most?	General Goals	User report	Multiple Choice	Managing the practice of physical activities, Food consumption management, Weight Management, Water consumption	All
Q3	Which browser do you use to run the application?	General Goals	Technical	Multiple Choice	Chrome, Firefox, Safari, Opera, Internet Explorer, Edge	All
Q4	What operating system did you use while running the application?	General Goals	Technical	Multiple Choice	Windows, Android, Linux, iOS, MacOS	All
Q5	Please, indicate the device that you use to run the application (for example, Asus Zenfone 5)	General Goals	Technical	Open		All
Q6	Did you have any problems using the app?	General Goals	User report	Boolean	Yes or No	All
Q7	If you had any problems, please describe them so we can better understand them.	General Goals	User report	Open		All
Q8	Evaluate the usefulness of scheduling activities	Activities Schedule	Usefulness	Five points likert scale	Unhelpful (1) to Very useful (5)	All
Q9	Evaluate the ease of use of the activity schedule	Activities Schedule	Ease of use	Five points likert scale	Very difficult (1) to Very easy (5)	All

Question	Description	Feature	Subject	Type	Accepted Values	Target
Q10	If desired, please provide more details about your experience using scheduling activities	Activities Schedule	User report	Open		All
Q11	Did you receive activity notifications on your device?	Activities Report	User report	Boolean	Yes or No	All
Q12	Evaluate the usefulness of activity notifications received on your device	Activities Report	Usefulness	Five points likert scale	Unhelpful (1) to Very useful (5)	All
Q13	Evaluate the usefulness of activity notifications received via email	Activities Report	Usefulness	Five points likert scale	Unhelpful (1) to Very useful (5)	All
Q14	Evaluate the usefulness of meal reporting in your routine	Activities Report	Usefulness	Five points likert scale	Unhelpful (1) to Very useful (5)	All
Q15	Evaluate the ease of use of meal reporting	Activities Report	Ease of use	Five points likert scale	Very difficult (1) to Very easy (5)	All
Q16	If desired, please provide more details about your experience with meal reporting	Activities Report	User report	Open		All
Q17	Evaluate the usefulness of reporting physical activity in your routine	Activities Report	Usefulness	Five points likert scale	Unhelpful (1) to Very useful (5)	All
Q18	Evaluate the degree of ease of use of the physical activity report	Activities Report	Ease of use	Five points likert scale	Very difficult (1) to Very easy (5)	All
Q19	If desired, please provide more details about your experience with reporting physical activity	Activities Report	User report	Open		All
Q20	Evaluate the usefulness of weight reporting in your routine	Activities Report	Usefulness	Five points likert scale	Unhelpful (1) to Very useful (5)	All
Q21	Evaluate the degree of ease of use of weight reporting	Activities Report	Ease of use	Five points likert scale	Very difficult (1) to Very easy (5)	All
Q22	If desired, please provide more details about your experience with weight reporting	Activities Report	User report	Open		All

Question	Description	Feature	Subject	Type	Accepted Values	Target
Q23	Evaluate the usefulness of your activity history in your routine	Activities History	Usefulness	Five points likert scale	Unhelpful (1) to Very useful (5)	All
Q24	Evaluate the ease of understanding the food consumption history chart	Activities History	Ease of use	Five points likert scale	Very difficult (1) to Very easy (5)	All
Q25	Evaluate the ease of understanding of the history chart of physical activity practice	Activities History	Ease of use	Five points likert scale	Very difficult (1) to Very easy (5)	All
Q26	Evaluate the ease of understanding the weight history chart	Activities History	Ease of use	Five points likert scale	Very difficult (1) to Very easy (5)	All
Q27	If desired, please provide more details about your experience with activity history charts	Activities History	User report	Open		All
Q28	How often did you read the recommended messages?	Recommended Messages	User report	Single Choice	More than twice a day, Once a day, More than twice a week, Once a week, Did not read, I did not receive any messages	Intervention
Q29	Did you follow the links in the messages?	Recommended Messages	User report	Single Choice	Yes, No, Did not know it was possible to follow the links, I did not receive any messages	Intervention
Q30	Did you usually like the received messages?	Recommended Messages	User report	Single Choice	Yes, No, I did not know it was possible to like the messages, I did not receive anymessages	Intervention
Q31	Evaluate, in general, the usefulness of the recommended messages	Recommended Messages	Usefulness	Five points likert scale	Unhelpful (1) to Very useful (5)	Intervention
Q32	Evaluate the usefulness of the recommended messages related to the practice of physical activities	Recommended Messages	Usefulness	Five points likert scale	Unhelpful (1) to Very useful (5)	Intervention

Question	Description	Feature	Subject	Type	Accepted Values	Target
Q33	Evaluate the usefulness of recommended messages related to healthy eating	Recommended Messages	Usefulness	Five points likert scale	Unhelpful (1) to Very useful (5)	Intervention
Q34	Evaluate the usefulness of recommended messages related to weight control	Recommended Messages	Usefulness	Five points likert scale	Unhelpful (1) to Very useful (5)	Intervention
Q35	If desired, please provide more details about your messaging recommendation experience.	Recommended Messages	User report	Open		Intervention

Source: Own authorship

APPENDIX C MY U'DUCTOR OPEN SURVEY ANSWERS OF AUTHOR'S CONTACTS

Table 30: Author's Contacts Answers to Survey Open Questions

Question	User	Group	Answer
Q5	1	Control	I only used it on the PC
	2	Control	Moto G5, PC
	3	Control	Asus Zenfone 4
	4	Control	LG K 10
	5	Control	Xiomi Redmi 2
	6	Control	iPhone 7
	1	Intervention	RedMi 4
	2	Intervention	Sansung J5
	3	Intervention	Desktop
	4	Intervention	Samsung S8+
	5	Intervention	Motorola G5, MacBook air
	6	Intervention	I used it on the desktop
	7	Intervention	Moto G5 S Plus
	8	Intervention	iPhone6
Q7	2	Control	It does not let me complete the activity.
	3	Control	The interface is a bit confusing for the user, I think you should rethink the application design a bit.
	1	Intervention	When opening the application, from the notification, the menu was always open and not the expected screen.
Q10	1	Control	Exercise and good nutrition activities were already fully incorporated into my routine, I do not believe this app is for people on my profile.
	3	Control	I had a hard time canceling schedule activities when I was testing.
	2	Intervention	It would be good to schedule medication
	3	Intervention	- The app could contain a facility to fill in the information (lunch photo for example). - In my meal follow-up, it was much higher than the daily portion intakes so I believe it is not correct, it was confusing to fill it. - I would not always do the activity, and it would remain pending.
	5	Intervention	Was cool. But the question of portions and the type of activity disturbed me a bit. Maybe if the interface was more visual (where it only tightens the choice) and if I could set a pattern and then just confirm it ... it would be easier.
	6	Intervention	I did not schedule activities, I just logged on to the system and got warnings.

Question	User	Group	Answer
Q16	1	Control	I already have an application, Dietbox, that I use in partnership with my nutritionist.
	2	Control	I did not use the resource
	4	Control	I scheduled activity to play tennis but I can not work on it. This accumulates and I do not have access to complete it
	5	Control	It would be more interesting if the reporting process were more friendly. The tool provides tips (captions) based on total consumption / day and not per meal, which ends up creating confusion at the time of filling. If the process is guided (at the beginning it is informed which meals are by default and the tool already shows examples of foods for category - ex: breakfast - carbohydrates, meats, etc.
	2	Intervention	It would be nice to have a space to attach to the diet that was designed for the person
	4	Intervention	I did not use a meal report
	5	Intervention	See answer given earlier
	6	Intervention	In fact, I never registered my meals.
Q19	4	Control	I schedule an activity to play tennis but I can not work on it. It accumulates and I do not have access to complete it. Maybe I did something wrong.
	5	Intervention	It was not so complicated but I think that more visual and being able to register a pattern would be good.
	6	Intervention	I have also never reported physical activity.
Q22	2	Control	I did not use it.
	5	Intervention	I did not use it much.
	6	Intervention	I also did not report the weight.
Q27	2	Control	I did not use it because I could not complete the activities.
	5	Intervention	I thought it was nice but I did not use it much.
	6	Intervention	I did not use the history chart.
Q 35	5	Intervention	I think my assessment may be compromised because I wear a FitBit bracelet. So I had in mind the way of capturing the data of this wearable. I wondered how good it would be if the two of them could talk ... But if I did not have the bracelet, I would certainly use Uductor more. Also, I do not have much discipline to follow routine and keep registering activities ... so that influences as well.

APPENDIX D MY U'DUCTOR OPEN SURVEY ANSWERS OF PHYSICAL EDUCATION STUDENTS

Table 31: Physical Students Answers to Survey Open Questions

Question	User	Group	Answers
Q5	1	Control	Galaxy j7
	2	Control	Asus
	3	Control	Xiaomi POCOPHONE F1
	4	Control	iPhone 8
	5	Control	Moto Z2 Play
	6	Control	j5
	1	Intervention	notebook
	2	Intervention	Motorola
	3	Intervention	Moto G5
	4	Intervention	Desktop
Q7	1	Control	Difficulty in viewing diet information
	5	Control	There was a day that did not appear to inform the activities and meals at the time, but then entered all at the same time. That's ok!
Q10	4	Control	Although I did not have app for ios for all of this
	6	Control	There is no choice of drinks in it.
Q16	6	Control	There is no choice of drinks in it.
	4	Intervention	It is difficult to put the food in portions
Q19	6	Control	It lacks the option to be able to mark that the activity was not executed, that the activity can not be performed.
Q27	6	Control	The position of the graph makes it difficult to understand, against the dates of the choice option.

REFERENCES

ABBAR, S.; MEJOVA, Y.; WEBER, I. You Tweet What You Eat: studying food consumption through twitter. In: ANNUAL ACM CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS, 33., 2015, New York, NY, USA. **Proceedings. . .** ACM, 2015. p. 3197–3206. (CHI '15).

ALABBAS, A.; BELL, J. **Indexed Database API 2.0**. [S.l.]: World Wide Web Consortium, 2018. W3C Recommendation, https://www.w3.org/TR/IndexedDB/.

ALSHURAFA, N.; EASTWOOD, J. A.; POURHOMAYOUN, M.; LIU, J. J.; SARRAFZADEH, M. Remote health monitoring: predicting outcome success based on contextual features for cardiovascular disease. In: ANNUAL INTERNATIONAL CONFERENCE OF THE IEEE ENGINEERING IN MEDICINE AND BIOLOGY SOCIETY, 2014., 2014. **Anais. . .** [S.l.: s.n.], 2014. p. 1777–1781.

ALSHURAFA, N.; EASTWOOD, J.; POURHOMAYOUN, M.; LIU, J. J.; NYAMATHI, S.; SARRAFZADEH, M. A Framework for Predicting Adherence in Remote Health Monitoring Systems. In: WIRELESS HEALTH 2014 ON NATIONAL INSTITUTES OF HEALTH, 2014, New York, NY, USA. **Proceedings. . .** ACM, 2014. p. 8:1–8:8. (WH '14).

ALSHURAFA, N.; SIDERIS, C.; POURHOMAYOUN, M.; KALANTARIAN, H.; SARRAFZADEH, M.; EASTWOOD, J. A. Remote Health Monitoring Outcome Success Prediction using Baseline and First Month Intervention Data. **IEEE Journal of Biomedical and Health Informatics**, [S.l.], v. PP, n. 99, p. 1–1, 2016.

AZIZA, R.; BORGI, A.; ZGAYA, H.; GUINHOUYA, B. SimNCD: an agent-based formalism for the study of noncommunicable diseases. **Engineering Applications of Artificial Intelligence**, [S.l.], v. 52, p. 235 – 247, 2016.

B, S.; JM, D.; E, J.; AL et. Integrating technology into standard weight loss treatment: a randomized controlled trial. **JAMA Internal Medicine**, [S.l.], v. 173, n. 2, p. 105–111, 2013.

BA, S.; WANG, L. Digital health communities: the effect of their motivation mechanisms. **Decision Support Systems**, [S.l.], v. 55, n. 4, p. 941 – 947, 2013. 1. Social Media Research and Applications 2. Theory and Applications of Social Networks.

BALANI, S.; DE CHOUDHURY, M. Detecting and Characterizing Mental Health Related Self-Disclosure in Social Media. In: ANNUAL ACM CONFERENCE EXTENDED ABSTRACTS ON HUMAN FACTORS IN COMPUTING SYSTEMS, 33., 2015, New York, NY, USA. **Proceedings. . .** ACM, 2015. p. 1373–1378. (CHI EA '15).

BANDURA, A. Social Cognitive Theory of Mass Communication. **Media Psychology**, [S.l.], v. 3, n. 3, p. 265–299, 2001.

BARNES, J. **Class and Committees in a Norwegian Island Parish**. [S.l.]: Plenum, 1954. (Human relations. [Offprint]).

BECKER, K. L. Cyberhugs: creating a voice for chronic pain sufferers through technology. **Cyberpsychology, Behavior, and Social Networking**, [S.l.], v. 16, n. 2, p. 123–126, 2012.

BERKMAN, L. F.; GLASS, T.; BRISSETTE, I.; SEEMAN, T. E. From social integration to health: durkheim in the new millennium. **Social science & medicine (1982)**, [S.l.], v. 51, n. 6, p. 843–857, Sept. 2000.

BHADORIA, R. S.; CHAUDHARI, N. S. Provisioning for Sensory Data Using Enterprise Service Bus: a middleware epitome. In: **Enhancing CBRNE Safety & Security: proceedings of the sicc 2017 conference**. [S.l.]: Springer, Cham, 2018. p. 197 – 203.

BHADORIA, R. S.; CHAUDHARI, N. S.; VIDANAGAMA, V. T. N. Analyzing the role of interfaces in enterprise service bus: a middleware epitome for service-oriented systems. **Computer Standards & Interfaces**, [S.l.], v. 55, p. 146 – 155, 2018.

BIRMAN, K.; JOSEPH, T. Exploiting Virtual Synchrony in Distributed Systems. **SIGOPS Oper. Syst. Rev.**, New York, NY, USA, v. 21, n. 5, p. 123–138, Nov. 1987.

BODENHEIMER, T.; WAGNER, E. H.; GRUMBACH, K. Improving Primary Care for Patients With Chronic Illness. **JAMA**, Family and Community Medicine, University of California, San Francisco, USA. tbodie@earthlink.net, v. 288, n. 14, p. 1775–1779, Oct. 2002.

BOEPPLE, L.; ATA, R. N.; RUM, R.; THOMPSON, J. K. Strong is the new skinny: a content analysis of fitspiration websites. **Body Image**, [S.l.], v. 17, p. 132 – 135, 2016.

BOND, G. E.; BURR, R. L.; WOLF, F. M.; FELDT, K. The Effects of a Web-Based Intervention on Psychosocial Well-Being Among Adults Aged 60 and Older With Diabetes. **The Diabetes Educator**, [S.l.], v. 36, n. 3, p. 446–456, 2010.

BRODERSEN, K. H.; GALLUSSER, F.; KOEHLER, J.; REMY, N.; SCOTT, S. L. Inferring causal impact using Bayesian structural time-series models. **Annals of Applied Statistics**, [S.l.], v. 9, p. 247–274, 2015.

BUDGEN, D.; TURNER, M.; BRERETON, P.; KITCHENHAM, B. Using Mapping Studies in Software Engineering. In: PPIG 2008, 2008. **Proceedings. . .** Lancaster University, 2008. p. 195–204.

CASTRO, A. M. de; SIMONI, C. L. de; GONçALVES, C. C. M.; GOSCH, C. S.; MALTA, D. d. C.; SARDINHA, L. M. V. **Diretrizes e Recomendações para o Cuidado Integral de Doenças Crônicas Não-Transmissíveis**. Brasília: Ministério da Saúde. Secretaria de Vigilância à Saúde. Secretaria de Atenção à Saúde., 2008.

CHANG, C.-C.; LIN, C.-J. LIBSVM: a library for support vector machines. **ACM Transactions on Intelligent Systems and Technology**, [S.l.], v. 2, p. 27:1–27:27, 2011. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

CHEN, W.; LAKSHMANAN, L. V. S.; CASTILLO, C. **Information and Influence Propagation in Social Networks**. [S.l.]: Morgan & Claypool Publishers, 2013. 9-20 p.

CHIêM, J.-C.; MACQ, J.; SPEYBROECK, N. Rule-Based Modeling of Chronic Disease Epidemiology: elderly depression as an illustration. **PLOS ONE**, [S.l.], v. 7, n. 8, p. 1–14, 08 2012.

CHRISTAKIS, N. A. A.; FOWLER, J. H. H. The Spread of Obesity in a Large Social Network over 32 Years. **New England Journal of Medicine**, [S.l.], v. 357, n. 4, p. 370–379, July 2007.

CHRISTAKIS, N. A. Social networks and collateral health effects. **BMJ**, [S.l.], v. 329, n. 7459, p. 184–185, 2004.

CHRISTAKIS, N. A.; FOWLER, J. H. The Collective Dynamics of Smoking in a Large Social Network. **New England Journal of Medicine**, [S.l.], v. 358, n. 21, p. 2249–2258, May 2008.

CHRISTAKIS, N. A.; FOWLER, J. H. Social Network Sensors for Early Detection of Contagious Outbreaks. **PLOS ONE**, [S.l.], v. 5, n. 9, p. 1–8, 09 2010.

CHRISTAKIS, N. A.; FOWLER, J. H. Social Contagion Theory: examining dynamic social networks and human behavior. **CoRR**, [S.l.], v. abs/1109.5235, 2011.

COOPER, I. D. What is a "mapping study? **Journal of the Medical Library Association : JMLA**, [S.l.], v. 104, n. 1, p. 76–78, 2016.

CULOTTA, A. Estimating County Health Statistics with Twitter. In: SIGCHI CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS, 2014, New York, NY, USA. **Proceedings. . .** ACM, 2014. p. 1335–1344. (CHI '14).

DALE, L. P.; DOBSON, R.; WHITTAKER, R.; MADDISON, R. The effectiveness of mobile-health behaviour change interventions for cardiovascular disease self-management: a systematic review. **European Journal of Preventive Cardiology**, [S.l.], v. 23, n. 8, p. 801–817, 2016. PMID: 26490093.

D'AMBROSIO, A.; AGRICOLA, E.; RUSSO, L.; GESUALDO, F.; PANDOLFI, E.; BORTOLUS, R.; CASTELLANI, C.; LALATTA, F.; MASTROIACOVO, P.; TOZZI, A. E. Web-Based Surveillance of Public Information Needs for Informing Preconception Interventions. **PLOS ONE**, [S.l.], v. 10, n. 4, p. 1–12, 04 2015.

DAVIS, F. D. Perceived usefulness, perceived ease of use, and user acceptance of information technology. **MIS Q.**, Minneapolis, MN, USA, v. 13, n. 3, p. 319–340, Sept. 1989.

DEY, A. K.; ABOWD, G. D.; SALBER, D. A Conceptual Framework and a Toolkit for Supporting the Rapid Prototyping of Context-aware Applications. **Hum.-Comput. Interact.**, Hillsdale, NJ, USA, v. 16, n. 2, p. 97–166, Dec. 2001.

DHILLON, J. S.; WüNSCHE, B. C.; LUTTEROTH, C. Accessible telehealth - Leveraging consumer-level technologies and social networking functionalities for senior care. In: INTERNATIONAL CONFERENCE ON HUMAN SYSTEM INTERACTIONS (HSI), 2013., 2013. **Anais. . .** [S.l.: s.n.], 2013. p. 451–458.

DURKHEIM, E. **Suicide**: a study in sociology. [S.l.]: Taylor & Francis e-Library, 2005, 1897. 404 p. ; p.

EIJK, M. van der; FABER, J. M.; AARTS, W. J.; KREMER, A. J.; MUNNEKE, M.; BLOEM, R. B. Using Online Health Communities to Deliver Patient-Centered Care to People With Chronic Conditions. **J Med Internet Res**, [S.l.], v. 15, n. 6, p. e115, Jun 2013.

FAULKNER, S.; EICHOLZ, A.; LEITHEAD, T.; DANILO, A.; MOON, S. **HTML 5.2**. [S.l.]: World Wide Web Consortium, 2018. W3C Recommendation, https://www.w3.org/TR/html/.

FIELDING, R. T.; TAYLOR, R. N. Principled design of the modern Web architecture. In: SOFTWARE ENGINEERING, 22., 2000, New York, NY, USA. **Proceedings. . .** ACM, 2000. p. 407–416. (ICSE '00).

FOWLER, J. H.; CHRISTAKIS, N. A. Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the framingham heart study. **British Medical Journal**, [S.l.], v. 337, n. dec04_2, p. a2338+, December 2008.

FREED, N.; BORENSTEIN, N. **Multipurpose Internet Mail Extensions (MIME) Part One**: format of internet message bodies. United States: RFC Editor, 1996. RFC, https://tools.ietf.org/html/rfc2045.

FREEDMAN, D. **Statistical Models** : theory and practice. [S.l.]: Cambridge University Press, 2005.

GARCIA-HERRANZ, M.; MORO, E.; CEBRIAN, M.; CHRISTAKIS, N. A.; FOWLER, J. H. Using Friends as Sensors to Detect Global-Scale Contagious Outbreaks. **PLOS ONE**, [S.l.], v. 9, n. 4, p. 1–7, 04 2014.

GóMEZ-PéREZ, A.; FERNANDEZ-LOPEZ, M.; CORCHO, O. Methodologies and Methods for Building Ontologies. In:______. **Ontological Engineering**. [S.l.]: Springer-Verlag London, 2004. p. 107–197.

GOMEZ-GALVEZ, P.; MEJíAS, C. S.; FERNANDEZ-LUQUE, L. Social media for empowering people with diabetes: current status and future trends. In: ANNUAL INTERNATIONAL CONFERENCE OF THE IEEE ENGINEERING IN MEDICINE AND BIOLOGY SOCIETY (EMBC), 2015., 2015. **Anais. . .** [S.l.: s.n.], 2015. p. 2135–2138.

GOYAL, A.; BONCHI, F.; LAKSHMANAN, L. V. Learning Influence Probabilities in Social Networks. In: THIRD ACM INTERNATIONAL CONFERENCE ON WEB SEARCH AND DATA MINING, 2010, New York, NY, USA. **Proceedings. . .** ACM, 2010. p. 241–250. (WSDM ’10).

GRIMM, S.; ABECKER, A.; VÖLKER, J.; STUDER, R. Ontologies and the Semantic Web. In:______. **Handbook of Semantic Web Technologies**. [S.l.]: Springer Berlin Heidelberg, 2011. p. 507–579.

GROUP, O. M. **OMG Unified Modeling Language (OMG UML), Version 2.5**. [S.l.: s.n.], 2015.

GUTTMAN, E.; PERKINS, C.; VEIZADES, J.; DAY, M. **Service Location Protocol, Version 2**. [S.l.]: RFC Editor, 1999. RFC, `http://www.rfc-editor.org/rfc/rfc2608.txt`. (2608).

HARDT, D. **The OAuth 2.0 Authorization Framework**. [S.l.]: RFC Editor, 2012. RFC, `http://www.rfc-editor.org/rfc/rfc6749.txt`. (6749).

HARTIGAN, J. A. **Clustering Algorithms**. 99th. ed. New York, NY, USA: John Wiley and Sons, Inc., 1975.

HISTORY of the Framingham Heart Study. Accessed: 2016-04-30, `https://www.framinghamheartstudy.org/about-fhs/history.php`.

HOLTZ, B.; KREIN, S. L.; BENTLEY, D. R.; HUGHES, M. E.; GIARDINO, N. D.; RICHARDSON, C. R. Comparison of Veteran experiences of low-cost, home-based diet and exercise interventions. **Journal of Rehabilitation Research & Development**, [S.l.], v. 51, n. 1, p. 149–160, 2014.

HOUSE, J.; LANDIS, K.; UMBERSON, D. Social relationships and health. **Science**, [S.l.], v. 241, n. 4865, p. 540–545, 1988.

HWANG, K. O.; ETCHEGARAY, J. M.; SCIAMANNA, C. N.; BERNSTAM, E. V.; THOMAS, E. J. Structural social support predicts functional social support in an online weight loss programme. **Health Expectations**, [S.l.], v. 17, n. 3, p. 345–352, 2014.

HWANG, K. O.; OTTENBACHER, A. J.; GREEN, A. P.; CANNON-DIEHL, M. R.; RICHARDSON, O.; BERNSTAM, E. V.; THOMAS, E. J. Social support in an Internet weight loss community. **International Journal of Medical Informatics**, [S.l.], v. 79, n. 1, p. 5 – 13, 2010.

INTERNATIONAL, E. **The JSON Data Interchange Syntax**. [S.l.]: ECMA International, 2017. Standard, http://www.ecma-international.org/publications/files/ECMA-ST/ECMA-404.pdf.

INTERNATIONAL, E. **ECMAScript® 2018 Language Specification**. [S.l.]: ECMA International, 2018. Standard,

https://www.ecma-international.org/publications/files/ECMA-ST/Ecma-262.pdf.

JOHNSON, F.; WARDLE, J. The association between weight loss and engagement with a web-based food and exercise diary in a commercial weight loss programme: a retrospective analysis. **International Journal of Behavioral Nutrition and Physical Activity**, [S.l.], v. 8, n. 1, p. 83, 2011.

JONES, M.; BRADLEY, J.; SAKIMURA, N. **JSON Web Token (JWT)**. [S.l.]: RFC Editor, 2015. RFC, `http://www.rfc-editor.org/rfc/rfc7519.txt`. (7519).

JR., T. A.; ETEMAD, E. J.; RIVOAL, F. **CSS Snapshot 2018**. [S.l.]: World Wide Web Consortium, 2018. W3C Recommendation, https://www.w3.org/TR/CSS/#css.

KEMPE, D.; KLEINBERG, J.; TARDOS, E. Maximizing the Spread of Influence Through a Social Network. In: NINTH ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING, 2003, New York, NY, USA. **Proceedings. . .** ACM, 2003. p. 137–146. (KDD ’03).

KESHAV, S. How to Read a Paper. **SIGCOMM Comput. Commun. Rev.**, New York, NY, USA, v. 37, n. 3, p. 83–84, July 2007.

KHAN, A.; UDDIN, S.; SRINIVASAN, U. Adapting Graph Theory and Social Network Measures on Healthcare Data: a new framework to understand chronic disease progression. In: AUSTRALASIAN COMPUTER SCIENCE WEEK MULTICONFERENCE, 2016, New York, NY, USA. **Proceedings. . .** ACM, 2016. p. 66:1–66:7. (ACSW ’16).

LAN, M.; SAMY, L.; ALSHURAFA, N.; SUH, M.-K.; GHASEMZADEH, H.; MACABASCO-O’CONNELL, A.; SARRAFZADEH, M. WANDA: an end-to-end remote health monitoring and analytics system for heart failure patients. In: CONFERENCE ON WIRELESS HEALTH, 2012, New York, NY, USA. **Proceedings. . .** ACM, 2012. p. 9:1–9:8. (WH ’12).

LEE, K.; AGRAWAL, A.; CHOUDHARY, A. Mining Social Media Streams to Improve Public Health Allergy Surveillance. In: IEEE/ACM INTERNATIONAL CONFERENCE ON

ADVANCES IN SOCIAL NETWORKS ANALYSIS AND MINING 2015, 2015., 2015, New York, NY, USA. **Proceedings. . .** ACM, 2015. p. 815–822. (ASONAM ’15).

LIBIN, A.; SCHLADEN, M. M.; LJUNGBERG, I.; TSAI, B.; DANFORD, E.; GROAH, S. Youtube How-to-Videos (HtV). In: INTERNATIONAL CONFERENCE ON COMPUTER SUPPORTED EDUCATION - VOLUME 1, 6., 2014, Portugal. **Proceedings. . .** SCITEPRESS - Science and Technology Publications: Lda, 2014. p. 607–613. (CSEDU 2014).

LIKERT, R. A technique for the measurement of attitudes. **Archives of Psychology**, [S.l.], v. 22, n. 140, p. 1–55, 1932.

MADAN, A.; CEBRIAN, M.; LAZER, D.; PENTLAND, A. Social Sensing for Epidemiological Behavior Change. In: ACM INTERNATIONAL CONFERENCE ON UBIQUITOUS COMPUTING, 12., 2010, New York, NY, USA. **Proceedings. . .** ACM, 2010. p. 291–300. (UbiComp ’10).

MADAN, A.; MOTURU, S. T.; LAZER, D.; PENTLAND, A. S. Social Sensing: obesity, unhealthy eating and exercise in face-to-face networks. In: WIRELESS HEALTH 2010, 2010, New York, NY, USA. **Anais. . .** ACM, 2010. p. 104–110. (WH ’10).

MALHOTRA, N.; BIRKS, D. **Marketing Research**: an applied approach. [S.l.]: Pearson Education, 2006.

MARANGUNIĆ, N.; GRANIĆ, A. Technology acceptance model: a literature review from 1986 to 2013. **Universal Access in the Information Society**, [S.l.], v. 14, n. 1, p. 81–95, Mar 2015.

MARTIN, C. M.; BISWAS, R.; JOSHI, A.; STURMBERG, J. **Patient Journey Record Systems (PaJR): the development of a conceptual framework for a patient journey system**. [S.l.]: IGI Global, 2011. 75-92 p.

MARTIN, C. M.; VOGEL, C.; GRADY, D.; ZARABZADEH, A.; HEDERMAN, L.; KELLETT, J.; SMITH, K.; O’ SHEA, B. Implementation of complex adaptive chronic care: the patient journey record system (pajr). **Journal of Evaluation in Clinical Practice**, [S.l.], v. 18, n. 6, p. 1226–1234, 2012.

MILANI, R. V.; LAVIE, C. J. Health Care 2020: reengineering health care delivery to combat chronic disease. **The American Journal of Medicine**, [S.l.], v. 128, n. 4, p. 337–343, 2016/12/17 2016.

MüLLNER, D. fastcluster: fast hierarchical, agglomerative clustering routines for r and python. **Journal of Statistical Software, Articles**, [S.l.], v. 53, n. 9, p. 1–18, 2013.

NEWMAN, M. E. J. The Structure and Function of Complex Networks. **SIAM Review**, [S.l.], v. 45, n. 2, p. 167–256, 2003.

PAUL, M.; DREDZE, M. **You Are What You Tweet**: analyzing twitter for public health. 2011.

PAUTASSO, C.; ZIMMERMANN, O.; LEYMANN, F. Restful web services vs. "big"’ web services: making the right architectural decision. In: WORLD WIDE WEB, 17., 2008, New York, NY, USA. **Proceedings. . .** ACM, 2008. p. 805–814. (WWW ’08).

PETERSEN, K.; VAKKALANKA, S.; KUZNIARZ, L. Guidelines for conducting systematic mapping studies in software engineering: an update. **Information and Software Technology**, [S.l.], v. 64, p. 1 – 18, 2015.

PETTICREW, M.; ROBERTS, H. **Systematic Reviews in the Social Sciences**: a practical guide. [S.l.]: Blackwell Pub., 2006.

PHILIPPI, S. T.; LATTERZA, A. R.; CRUZ, A. T. R.; RIBEIRO, L. C. Adapted food pyramid: a guide for a right food choice. **Brazilian Journal of Nutrition**, [S.l.], v. 12, 1999.

PITOLLI, F.; VIANNA, H. D.; BARBOSA, J. L. **An Education Driven Model for Non-Communicable Diseases Care**. [S.l.]: IGI Global, 2011. 391-418 p.

PITTOLI, F.; VIANNA, H. D.; BARBOSA, J. L. V.; BUTZEN, E.; GAEDKE, M. Ângela; COSTA, J. S. D. da; SANTOS, R. B. S. dos. An intelligent system for prognosis of noncommunicable diseases' risk factors. **Telematics and Informatics**, [S.l.], v. 35, n. 5, p. 1222 – 1236, 2018.

RADLOFF, L. S. The CES-D Scale: a self-report depression scale for research in the general population. **Applied Psychological Measurement**, [S.l.], v. 1, n. 3, p. 385–401, 1977.

RAM, S.; ZHANG, W.; WILLIAMS, M.; PENGETNZE, Y. Predicting Asthma-Related Emergency Department Visits Using Big Data. **IEEE Journal of Biomedical and Health Informatics**, [S.l.], v. 19, n. 4, p. 1216–1223, July 2015.

RICHARDSON, R. C.; BUIS, R. L.; JANNEY, W. A.; GOODRICH, E. D.; SEN, A.; HESS, L. M.; MEHARI, S. K.; FORTLAGE, A. L.; RESNICK, J. P.; ZIKMUND-FISHER, J. B.; STRECHER, J. V.; PIETTE, D. J. An Online Community Improves Adherence in an Internet-Mediated Walking Program. Part 1: results of a randomized controlled trial. **J Med Internet Res**, [S.l.], v. 12, n. 4, p. e71, Dec 2010.

ROBERTSON, H.; NICHOLAS, N.; ROSENFELD, T.; GEORGIOU, A.; JOHNSON, J.; TRAVAGLIA, J. **A virtual aged care system**: When health informatics and spatial science intersect. [S.l.]: IOS Press Ebooks, 2014. 137-142 p. v. 204.

ROBINSON, D.; SILGE, J. **Text Mining with R - A Tidy Approach**. [S.l.]: O'Reilly Media, 2017.

SATYANARAYANAN, M. Pervasive computing: vision and challenges. **IEEE Personal Communications**, [S.l.], v. 8, n. 4, p. 10–17, Aug 2001.

SCHWARTZ, D. G.; BELLOU, A.; GARCIA-CASTRILLO, L.; MURARO, A.; PAPADOPOULOS, N. G. Towards chronic emergency response communities for anaphylaxis. In: IEEE 15TH INTERNATIONAL CONFERENCE ON INFORMATION REUSE AND INTEGRATION (IEEE IRI 2014), 2014., 2014. **Proceedings. . .** [S.l.: s.n.], 2014. p. 98–103.

SHARMA, U.; BHADORIA, R. S.; DIXIT, M. Featured Analysis of Enterprise Service Bus. In: **Exploring Enterprise Service Bus in the Service-Oriented Architecture Paradigm**. [S.l.]: IGI Global, 2017. p. 14 – 25.

SIDERIS, C.; ALSHURAFA, N.; KALANTARIAN, H.; SARRAFZADEH, M.; EASTWOOD, J.-A. Effects of Coaching on Adherence in Remote Health Monitoring Systems: analysis and prediction of participant adherence. In: CONFERENCE ON WIRELESS HEALTH, 2015, New York, NY, USA. **Proceedings. . .** ACM, 2015. p. 10:1–10:8. (WH '15).

SIMPSON, C. C.; MAZZEO, S. E. Skinny Is Not Enough: a content analysis of fitspiration on pinterest. **Health Communication**, [S.l.], v. 32, n. 5, p. 560–567, 2017. PMID: 27326747.

SMAHEL, D.; ELAVSKY, S.; MACHACKOVA, H. Functions of mHealth applications: a user's perspective. **Health Informatics Journal**, [S.l.], 2017. PMID: 29121831.

SNOWBALLING. Accessed: 2016-11-11, http://hlwiki.slais.ubc.ca/index.php/Snowballing.

TANG, J.; SUN, J.; WANG, C.; YANG, Z. Social Influence Analysis in Large-scale Networks. In: ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING, 15., 2009, New York, NY, USA. **Proceedings. . .** ACM, 2009. p. 807–816. (KDD '09).

TRAVERS, J.; MILGRAM, S. An Experimental Study of the Small World Problem. **SOCIOMETRY**, [S.l.], v. 32, n. 4, p. 425–443, 1969.

TRENDS, G. **Google Trends Search on Big Data and Machine Learning From 2010 to 2016**. [Online; accessed 26-December-2016], `https://www.google.com/trends/explore?date=2010-01-01%202016-12-26&q=big%20data,machine%20learning`.

TUAROB, S.; TUCKER, C. S.; SALATHE, M.; RAM, N. Discovering Health-related Knowledge in Social Media Using Ensembles of Heterogeneous Features. In: ND ACM INTERNATIONAL CONFERENCE ON INFORMATION & KNOWLEDGE MANAGEMENT, 22., 2013, New York, NY, USA. **Proceedings. . .** ACM, 2013. p. 1685–1690. (CIKM '13).

VIANNA, H. D.; BARBOSA, J. L. V. A Model for Ubiquitous Care of Noncommunicable Diseases. **IEEE Journal of Biomedical and Health Informatics**, [S.l.], v. 18, n. 5, p. 1597–1606, Sept 2014.

VIANNA, H. D.; BARBOSA, J. L. V. In search of computer-aided social support in non-communicable diseases care. **Telematics and Informatics**, [S.l.], 2017.

VIANNA, H. D.; BARBOSA, J. L. V. A scalable model for building context-aware applications for noncommunicable diseases prevention. **Information Processing Letters**, [S.l.], 2019.

VIANNA, H. D.; BARBOSA, J. L. V.; GLUZ, J. C.; MARQUES, E. B. Pompilos onto: an ontology for detecting the spreading of happiness, obesity and smoking in social networks. In: XLII LATIN AMERICAN COMPUTING CONFERENCE (CLEI), 2016., 2016. **Anais. . .** [S.l.: s.n.], 2016. p. 1–9.

VIANNA, H. D.; BARBOSA, J. L. V.; GLUZ, J. C.; SANTOS, R. B. S. D. Design of an ontology for detecting the social influence on non-communicable diseases risk factors. **International Journal of Metadata, Semantics and Ontologies**, [S.l.], v. 13, n. 2, p. 11, 2018.

VIANNA, H. D.; BARBOSA, J. L. V.; PITTOLI, F. In the Pursuit of Hygge Software. **IEEE Software**, [S.l.], v. 34, n. 6, p. 48–52, November 2017.

WAGNER, A.; BARBOSA, J. L. V.; BARBOSA, D. N. F. A model for profile management applied to ubiquitous learning environments. **Expert Systems with Applications**, [S.l.], v. 41, n. 4, Part 2, p. 2023 – 2034, 2014.

WAGNER, E. H.; AUSTIN, B. T.; DAVIS, C.; HINDMARSH, M.; SCHAEFER, J.; BONOMI, A. Improving Chronic Illness Care: translating evidence into action. **Health Aff**, W.A. MacColl Institute for Healthcare Innovation at the Center for Health Studies, Group Health Cooperative of Puget Sound, Seattle, USA., v. 20, n. 6, p. 64–78, Nov. 2001.

WAGNER, E. H.; GROVE, T. Care for chronic diseases - The efficacy of coordinated and patient centred care is established, but now is the time to test its effectiveness. **British Medical Journal**, [S.l.], v. 325, p. 913–914, 2002.

WEBER, I.; MEJOVA, Y. Crowdsourcing Health Labels: inferring body weight from profile pictures. In: INTERNATIONAL CONFERENCE ON DIGITAL HEALTH CONFERENCE, 6., 2016, New York, NY, USA. **Proceedings. . .** ACM, 2016. p. 105–109. (DH'16).

WELCH, B. M.; O'CONNELL, N. S.; QANUNGO, S.; HALBERT-HUGHES, C.; SCHIFFMAN, J. D. Collecting Family Health History using an Online Social Network: a nationwide survey among potential users. **AMIA Annu Symp Proc**, [S.l.], v. 2015, p. 1316–1325, 2015.

WEYMANN, N.; DIRMAIER, J.; WOLFF, A. von; KRISTON, L.; HäRTER, M. Effectiveness of a Web-Based Tailored Interactive Health Communication Application for Patients With Type 2 Diabetes or Chronic Low Back Pain: randomized controlled trial. **J Med Internet Res**, [S.l.], v. 17, n. 3, p. e53, Mar 2015.

WHO. **Cuidados inovadores para condições crônicas: componentes estruturais de ação: relatório mundial**. Brasília: World Health Organization, 2003.

WHO. **Preventing CHRONIC DISEASES a vital investment**. Geneva: World Health Organization, 2005.

WHO. **WHO global strategy on diet, physical activity and health**: a framework to monitor and evaluate implementation. 2008.

WHO. **Global status report on noncommunicable diseases 2010**. 20 Avenue Appia, 1211 Geneva 27, Switzerland: World Health Organization, 2010.

WHO. **Global Status Report On Noncommunicable Diseases 2014**. Geneva: World Health Organization, 2014.

WHO. **Obesity and overweight**. 2018.

WHO. **Physical Activity and Adults**. 2018.

WINTER, J. C. F. D.; DODOU, D. Five-point Likert items: t test versus mann-whitney-wilcoxon. **Practical Assessment, Research & Evaluation**, [S.l.], p. 11, 2010.

YOON, C.; KIM, S. Convenience and TAM in a ubiquitous computing environment: the case of wireless lan. **Electronic Commerce Research and Applications**, [S.l.], v. 6, n. 1, p. 102–112, 2007.

ZAMIR, O.; ETZIONI, O. Web Document Clustering: a feasibility demonstration. In: ANNUAL INTERNATIONAL ACM SIGIR CONFERENCE ON RESEARCH AND DEVELOPMENT IN INFORMATION RETRIEVAL, 21., 1998, New York, NY, USA. **Proceedings. . .** ACM, 1998. p. 46–54. (SIGIR '98).

ZHANG, W.; RAM, S.; BURKART, M.; PENGETNZE, Y. Extracting Signals from Social Media for Chronic Disease Surveillance. In: INTERNATIONAL CONFERENCE ON DIGITAL HEALTH CONFERENCE, 6., 2016, New York, NY, USA. **Proceedings. . .** ACM, 2016. p. 79–83. (DH ’16).