UiT
THE ARCTIC
UNIVERSITY
OF NORWAY
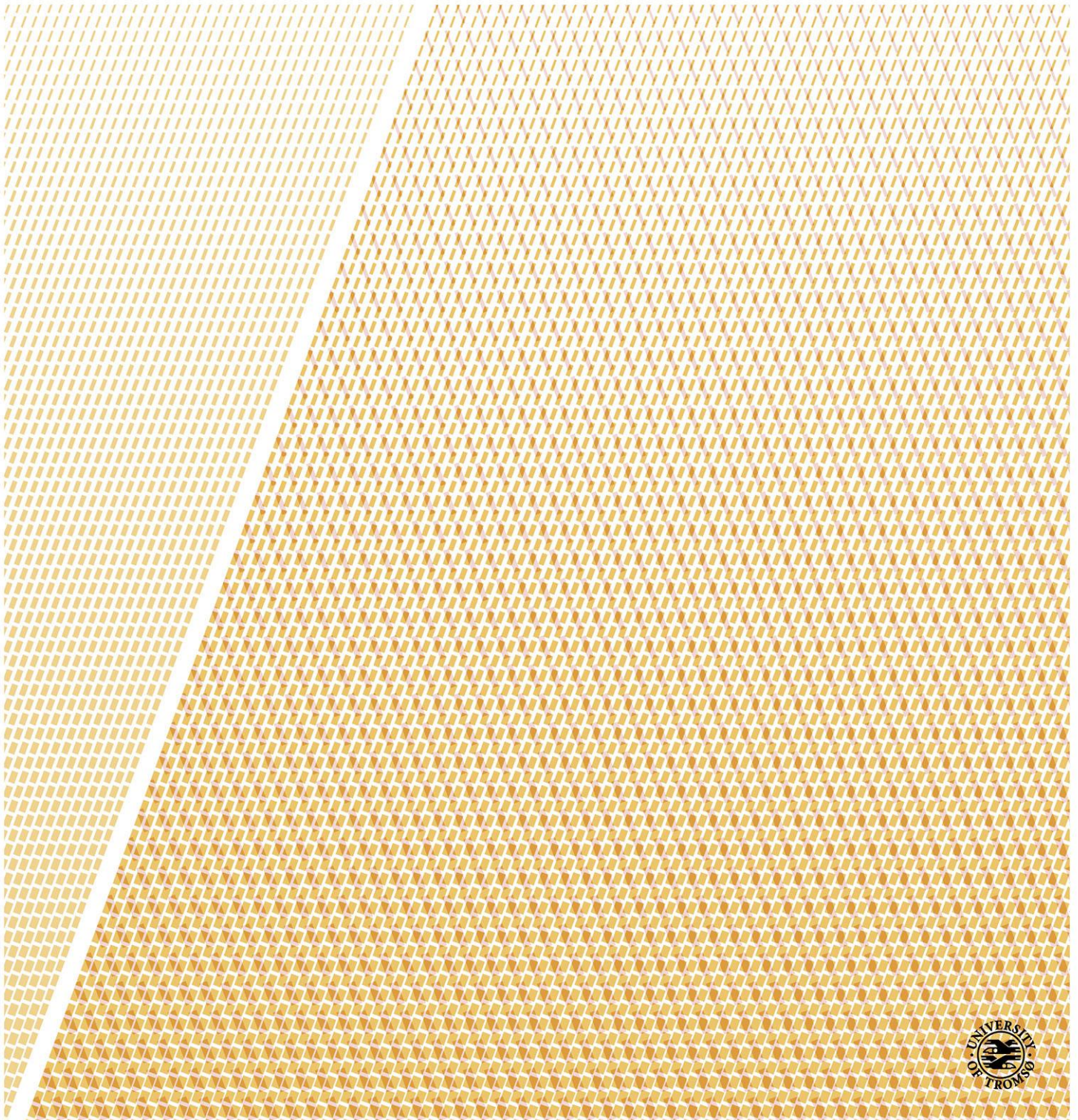
Department of Computer Science

# Complex Network Structure Patterns in Open Internet Communities for People with Diabetes

—

**Taridzo Chomutare**
*A dissertation for the degree of Philosophiae Doctor – December 2013*

UNIVERSITY OF TROMSØ

*" The thing that hath been, it is that which shall be; and that which is done is that which shall be done: and there is no new thing under the sun. Is there any thing whereof it may be said, See, this is new? it hath been already of old time, which was before us."*

*—Ecclesiastes*

# *Abstract*

Type 2 diabetes is one of the greatest challenges that continues to grow because of the ageing population and sedentary lifestyles, and consequently increasing morbid obesity. The emergence of usable mobile devices and the Internet has enabled the technologies for managing chronic illnesses such as diabetes, largely in an uncontrolled manner.

Social media such as Facebook and YouTube have transformed the way people interact in general and on the Internet, but the role social media play in healthcare is still not well-understood. The numbers of users in open Internet communities for patients now run into the millions, but current understanding of how online participation affects health outcomes or behaviour change is still limited.

In this dissertation a framework is presented, based on social network analysis, to explore the nature of patient interactions in online communities. Using recent advances in complex network analysis, and developing enhanced machine learning techniques, the community structures are articulated, showing how interaction behaviours correlate with health outcomes.

Results show that people with diabetes join online communities typically immediately following diagnosis, with over 80% of the patients having being diagnosed in under 2 years. The networks are very centralized with continually shrinking density and diameter as the the networks grow. These results directly contrast with current evidence about non-healthcare social networks.

Further, using this knowledge to enhance a classification method, it is shown that we can predict health outcomes, such as weight loss performance, based on how the patients interact online. Experimental data show that decision tree methods had superior performance on the healthcare datasets, reaching an F-score of 0.977, precision of 0.978 and AUC of 0.996. In addition, the evidence suggests that patient interaction data can be used to enhance user-similarity analysis when calculating top-N recommendations using collaborative filtering techniques.

These results have practical relevance for understanding the nature of patients interactions, as well as for designing personalized eHealth tools based on emergent social technologies. So far, little attention has been paid to these unregulated, open Internet communities, but the sheer numbers alone warrant some investigation. The findings in this dissertation build evidence supporting the significance of these online communities in disease management, and should provide the impetus for further research.

# *Preface and Acknowledgements*

During my MSc studies, I was intrigued by Eirik's work with mobile phones in diabetes self-management. When it was time to do my thesis I contacted him to ask if he had time to co-supervise my thesis work, and he agreed - my work in diabetes research had begun. The thesis went on to win the prestigious best master thesis award at the faculty, and I was highly motivated to start PhD studies. After important collaborations with colleagues like Johan Gustav Bellika, and a very supportive boss, Per Atle Bakkevoll, I was able to start research work.

There are many people who helped my success that I would like to thank; including my main supervisor, Gunnar Hartvigsen and co-supervisor, Eirik Årsand. The pair offered support during the highs and the lows, it's extraordinary!! My main supervisor supported my presentations and research in person – at conferences and research visits on several occasions and on more than two continents. I have heard many a PhD student complain about shoddy supervision, and I just cannot relate. For me, supervision meetings were fun for several reasons; besides the academic advice, sometimes having two (busy) experts listen while I whine about my research problems or difficulties during an episode of "writers block" – that was therapeutic by itself. In other times, they stood by me while pulling all-nighters under the threat of impending deadlines. You guys rock!

I'd also like to thank the Medical Informatics & Telemedicine (MI&T) Group at the University of Tromsø, and in particular Luis Fernandez-Luque for his support even before starting the PhD, but more so during the initial phases of the research, resulting in what turned out to be my most influential piece of scientific work. I thank my employer, NST @ UNN, for a supportive environment with amazing colleagues; Lone, Bente, Erlend, Gerd, Naoe, Randi, Line, Per, Geir, the developer team and others. I thank my supervisor during my visit at UTHealth, M Sriram Iyengar for the effective contribution to this work, and colleague Anna Xu, and Kathrine and Siv-Heidi at Diabetesforbundet.

I thank Helse Nord RHF for the unwavering and generous financial support for my research, and with Berit-Stine's fine work, I never wanted for anything. . . literary.

In the end, I would like to say that hard work and perseverance really pay off, but in my case many things were just in the right place and at the right time – a kind of ". . . *the race is not to the swift, nor the battle to the strong* . . ." scenario.

Thank you mom, Gill, Tanya and Unique for the moral support and prayers. Blessed art Thou, O Lord our God, King of the universe, who maketh wise the simple.

# Contents

# List of Figures

# List of Tables

# Publication List

**Paper 1**    Chomutare T, Fernandez-Luque L, Årsand E, Hartvigsen G. (2011) Features of mobile diabetes applications: Review of the literature and analysis of current applications against evidence-based guidelines. *Journal of Medical Internet Research*: 13(3);e65

**Paper 2**    Chomutare T, Årsand E, Fernandez-Luque L, Lauritzen J, Hartvigsen G. (2013) Inferring community structure in healthcare forums: An empirical study. *Methods of Information in Medicine*: 52(2)

**Paper 3**    Chomutare T, Årsand E, Hartvigsen G. (2013) Characterizing Development Patterns of Healthcare Social Networks. *Network Modeling Analysis in Health Informatics and Bioinformatics, Springer*: 3(2);147-157

**Paper 4**    Chomutare T, Årsand E, Hartvigsen G. (2012) Temporal community structure patterns in diabetes social networks. *HIBIBI Symposium in Proceedings of the International Conference on Advances in Social Networks Analysis and Mining*, pages 210–217 ACM/IEEE, Los Alamitos, CA

**Paper 5**    Chomutare T, Tatara N, Årsand E, Hartvigsen G. (2013) Designing a diabetes mobile application with social network support. *Studies in Health Technology and Informatics, IOS Press*: 188:58-64.

**Paper 6**    Chomutare T, Xu A, Iyengar MS. Social Network Analysis to Delineate Interaction Behaviour that Predicts Weight Loss Performance. *SUBMITTED*

**Paper 7**    Chomutare T. Collaborative Filtering with Community Structure Properties in Healthcare Social Networks. *MANUSCRIPT*

# Chapter 1

# Introduction

Diabetes is increasingly becoming a major public health problem, and is responsible for 8% of the global burden of disease for people between the ages of 20 and 79 (IDF, 2011). It is one of the greatest healthcare challenges that is expected to continue to grow because of the aging population, increasing obesity and sedentary lifestyles , and this is true for Type 2 Diabetes .

In 2003, 194 million people globally were estimated to have a form of diabetes, which is predicted to increase to 333 million in 2025, which constitute an increase of 72% (Lefebvre and Pierson, 2004). Recent reports show that the current number of people with diabetes is about 347 million, which is over the previous forecast for 2025 (IDF, 2011).

Now it is estimated diabetes will become the seventh leading cause of death by 2030 (IDF, 2011). The problem is seen not only in Europe and America, but on a global scale. WHO estimates that, 5-10% of the national healthcare budget in western countries is used on diabetes, which will increase with the increasing number of diabetes patients globally.

As alluded to, a closely related high risk factor for Type 2 diabetes is obesity, and the methods in this dissertation are also applied to the case. Obesity is a problem that is becoming increasingly difficult to ignore. In United States alone, recent surveys show more than a third of the population is considered obese (Ogden et al., 2012) with over 30 body mass index (BMI), while almost 7% is morbidly obese (Sturm and Hattori, 2012) with over 40 BMI. Although the obesity rates for Norway are comparatively modest at less than 10%, there still exists a long term trend towards increasing obesity rates.

While we have the burden of the disease on the one hand, we also have social media emerging as a platform for coping with the disease in online communities of similar patients. Social media such as Facebook and YouTube have transformed the way people interact in general and on the Internet, but the role social media play in disease management is still not well-understood.

Its emergence, and that of more usable and pervasive mobile devices , introduced a new dynamic for healthcare and self-management. The results in this dissertation will shed some light on the status quo as well as potential future trends.

## 1.1 Healthcare Terms and Concepts

There are some terms uses in this work that have been used in healthcare and disease management before the Internet. Some of these terms have connotations of their traditional meaning, but are used in this context to encourage new ways of thinking about disease self-management.

**Internet Patient Communities (IPC)**

Throughout this dissertation references are made to online communities and social media for people with different health conditions as *Internet Patient Communities* (IPC) . The phrase encompasses all the phenomena related to patient participation in health-related Internet communities.

**Users vs. Patients**

*Users* and *Patients* are interchangeably used to refer to people interacting in IPC, and occasionally *Participants*, in the case of the pilot study. This stems from the idea that people with different health conditions are also "users" of IPC, just like participants in other non-health forums.

**Health Outcomes**

In healthcare, *Health Outcomes* probably encompass objective findings and evaluation of the patient by a physician or expert. In this dissertation, however, the term is used to refer to the changes in health status measurements. The presented data analysis alone cannot evaluate how the patient is feeling, for example, but objective measurements can be viewed as limited surrogates for health status evaluation.

Throughout this dissertation reference is made to the improvements or deterioration of primary vital health sign measurements for diabetes as *Health Outcomes*:

- *(1) Glycated haemoglobin (HbA$_{1c}$)* – this is a long term measure of blood glucose levels. It is generally used to gauge how well a patient has long term blood glucose levels under control. This is sometimes reported in IPC.
- *(2) Blood Glucose Levels* (BG) – in the pilot study, the participants use a glucometer to measure blood glucose levels, and the data is transferred to their mobile device via Bluetooth. This is normally not reported in IPC.
- *(3) physical activity* and *(4) diet habits* – these are important since Type 2 diabetes is a lifestyle disease. People normally report the kind of food they eat or the calories, and also how much exercise they do, how strenuous and for how long. Normally unstructured reports in IPC.
- *(5) weight* – since obesity is an important risk factor for type 2 diabetes. Weight and Body Mass Index (BMI) are normally reported in IPC.
- *(6) blood pressure* – this is sometimes reported in IPC.

**eHealth Intervention**

The term *eHealth Intervention* is used in this dissertation to highlight IPC as a plausible tool for promoting healthy behaviours. Currently these communities are not organized, but the future ideal may that healthcare workers recognise the value of the tools, and thus also encourage more research and participation by the health service personnel.

## 1.2  Motivation

There is evidence of the number of IPC users running into the millions for various chronic illnesses, and more websites mushrooming to cater for different patient groups. The sheer numbers alone are indicative of a trend that is becoming increasingly difficult to ignore. The continued growth of these IPC suggest a new role that social media plays in disease management, although the nature of the role is still neither well-defined nor well-understood.

It is conceivable that users may not want to mix their normal social life on forums such as Facebook, with their obligations to self-manage chronic illnesses. This assertion is consistent with what has been reported in recent studies such as by van der Velden et al. (2013), who reported low usage of Facebook for managing disease by juveniles. Content analysis of breast cancer groups in Facebook done by Bender et al. (2011) showed that

although the total membership was more than one million, an overwhelming majority of the groups had under 25 wall-posts, indicating low utilization.

There has been some research where online patient portals connect to electronic records (EHR) at health institutions (Glasgow et al., 2012) or have GP support (Kummervold and Johnsen, 2011). Although the traditional *Doctor-Patient* social media is a promising trend that has gained acceptance by health service personnel, the numbers from the wider concept of unregulated, open, *patient-driven*, Internet communities are compelling.

However, so far little is known about the dynamics of user interactions in these open IPC. Current understanding of the association between user interaction behaviour and health outcomes or behaviour change is still limited. There is an obvious knowledge gap that can possibly be filled by multidisciplinary research, well-founded in both informatics and healthcare fields.

In this dissertation a framework is presented for abstraction of patient interaction patterns based on social network analysis (SNA) , in an effort to better understand their nature. Network analysis is one of the more practical methods for abstracting that patient-to-patient dialogue in IPC. By observing patient interactions we can develop connections or "networks" of inferred relationships , and use machine learning techniques to understand how these interactions correlate with health outcomes.

## 1.3   Research Questions

Given the continued growth of IPC, it is important to: (i) establish their relevance to disease management (ii) understand the nature of the interactions or patient-to-patient dialogue, and (ii) understand how the interactions potentially affect the health of participants. The main research question can be stated as:

*(MQ) - What is the nature of patient interactions patterns in Internet Patient Communities, and how do these interaction patterns affect health outcomes?*

To make the research easier to tackle, three sub-questions were developed to reflect the main thematic areas that emerge from the main research question, resulting in the following four sub-questions:

- **Question Q1 –** What is the relevance of Internet Patient Communities in diabetes self-management?

- **Question Q2 –** How can interaction patterns in Internet Patient Communities be modelled?
- **Question Q3 –** What interaction patterns characterize Internet Patient Communities?
- **Question Q4 –** How do these interaction patterns relate to health outcomes?

The first question (**Q1**) seeks to demonstrate and establish the relevance of eHealth and IPC for patients who have to self-manage illnesses such as diabetes. Therefore, some background work on eHealth self-management applications is required to put the research in context and demonstrate the practical relevance.

The second question (**Q2**) seeks to develop abstractions of patient interactions using informatics techniques. Using recent advances in complex network analysis , there is a potential to uncover previously unknown network structure patterns in IPC, allowing me to clearly articulate the nature of these social networks.

The third question (**Q3**) seeks to establish the fundamental developmental differences between these IPC and other types of non-health social networks. The knowledge could help us exploit the unique patterns that differentiate them.

The last question (**Q4**) seeks to apply these abstractions to assess correlations with health outcomes, and the hypothesis is that the way patients interact in IPC has an effect on their health.

### 1.3.1 Scope and Key Assumptions

It is important to note that work presented in this dissertation is based on IPC, this is, for patients only. Therefore, the scope does not extend to physician communities or other health service communities where patients communicate with health services, hospitals or GP offices using web applications.

Perhaps another important assumption to note is that the presented work assumes limited patient information. In order to have practical relevance, works has to be done under this assumption since we can only find limited patient information in IPC in practice. Although there is a lot of data generated in IPC, much of the information is scant and insufficient for many purposes.

There is no doubt that many people turn to IPC at one point or another if they have a health condition. This is in spite of the still very weak evidence to support IPC as an

eHealth tool. However, most users do not give away most of this important information, making it difficult to measure the improvements or how well the patients cope in IPC.

## 1.4 Significance of the Study

The findings from this dissertation have practical relevance for research practice in two seemingly disparate research fields; informatics and healthcare. A generalizable framework is developed for analysing IPC; both for understanding user interactions and how these interaction correlate with health outcomes.

Further, it is demonstrated that we can delineate the unique development patterns that characterize diabetes-related IPC. It is also shown that the analysis have practical relevance for understanding the relationships with health outcomes; also enhancing machine learning techniques in the design of personalized eHealth interventions .

### Informatics

The significance for informatics is based on the models and analyses that are based on collection of empirical data, and artefacts based on experimental work. The work results in a framework that is extensible and easily generalizable to most IPC and not just diabetes alone.

### Healthcare

The significance for healthcare is quite apparent since the analysis based on the framework results in practical and actionable information. Current work builds evidence for what was previously unknown about the impact of IPC on diabetes health outcomes. The results enhances our understanding of IPC as an eHealth intervention; both for influencing research practice and clinical guidelines in diabetes self-management.

### Patients

The study is also informative to patients because so far people have no idea how much IPC helps in their healthcare on the average. Our findings suggest that many people find IPC a good starting point once one is diagnosed with an unfamiliar health condition such as diabetes. Perhaps people turn to the web because it is highly available and mostly free of charge, with the promise of finding peers for support and empathy.

TABLE 1.1: Scientific Contributions (SC) related to the questions and scientific articles

| Contribution | | Description | Publication |
|---|---|---|---|
| **Q1** | SC1 | A systematic review of mobile applications in diabetes self-management. | Paper 1 |
| **Q1** | SC2 | Feasibility study for social mobile applications in diabetes | Paper 5 |
| **Q2** | SC3 | An evaluation of a network analysis abstraction of patient interaction patterns in IPC | Paper 2 |
| **Q3** | SC4 | An expansion of the abstraction for analysing how IPC for diabetes develop over time | Paper 4 |
| **Q3** | SC5 | Applying the abstraction to discover unique patterns that characterize IPC for diabetes | Paper 3 |
| **Q4** | SC6 | Based on the developed abstraction, a clustering-classification method for delineating interaction behaviours that affect health outcomes. | Paper 6 |
| **Q4** | SC7 | A collaborative filtering method for threads and users based on the developed abstraction | Paper 7 |

## 1.5 Contributions

This dissertation is based on several smaller studies, each making some scientific progress, resulting in a whole framework for analysing IPC. Based on the original research questions, it is shown what contributions were made, and how they are related to the published scientific articles. The separate scientific contributions (SC) discussed in the different articles are summarised in Table 1.1.

### 1.5.1 Weaving the Web of Publications

The overall objective of the project is to develop a framework for analysing IPC; to help us better understand the nature of patient-to-patient dialogue , and discovering how this sharing of experience online can affect both behaviour change and health outcomes.

Initially the work starts with a review of the literature, then goes on to explore an abstraction based on network analysis of IPC, and apply the abstraction to enhancing machine learning techniques before concluding. The work can be viewed as partitioned into four phases as discussed next.

**Phase #1 – Relevance of eHealth, mHealth and Patient Interactions in Diabetes – related to Q1**

The initial systematic review is a wide piece of work aimed at shedding light on research gaps in closely related fields of mobile devices, social media and education in healthcare *Paper 1* (Chomutare et al., 2011). This helps establish the relevance of eHealth in self-managing diabetes. One of the research gaps discovered was the limited knowledge about health social media. Further study with a social mobile application in a 12-month pilot study with diabetes patients *Paper 5* (Chomutare et al., 2013a) helped demonstrate practical relevance.

**Phase #2 – Modeling Patient Interactions – related to Q2**

Having had preliminary evidence of clinical benefits and feasibility to justify further work with social media, empirical evaluation of network analysis as an abstraction tool for patient interactions *Paper 2* (Chomutare et al., 2013b) was done.

**Phase #3 – Characterizing Interaction Patterns in IPC – related to Q3**

During the analysis of data, shortcomings were discovered in the literature regarding use of network analysis on empirical healthcare data sets. More elaborate work was needed to understand the development of the networks, and methods for analysing the temporal patterns in diabetes networks are explored *Paper 4* (Chomutare et al., 2012).

Further, it was then sought to explore any unique characteristics between these diabetes networks and other conventional social networks such as Facebook *Paper 3* (Chomutare et al., 2013c).

**Phase #4 – Interaction Patterns and Health Outcomes Correlation – related to Q4**

The last phase sought to show that the analysis and methods proposed could further enhance our understanding of IPC. One study explores the relationships between patient interaction patterns and weight loss performance in an online community *Paper 6* using classification techniques.

The other study develops a method for improving collaborative filtering by considering patient interaction patterns *Paper 7* when calculating user similarities in designing recommender systems for IPC.

## 1.5.2 Statements of Originality

**Paper 1** Chomutare T, Fernandez-Luque L, Årsand E, Hartvigsen G. (2011) Features of mobile diabetes applications: Review of the literature and analysis of current applications against evidence-based guidelines. *Journal of Medical Internet Research*: 13(3);e65

**Relevance** This was the first major study in the research project and was exploratory in nature. The study reviewed existing electronic applications for people with diabetes with a view to understand their compliance to clinical guidelines as well as to discover the research gaps. This paper is key to the dissertation because it reveals the gaps in knowledge of social media on mobile devices and the Internet. The dissertation is founded on the premises developed in this study.

**My Contribution** The initial idea for the review was coined by my co-author LFL and I developed the idea further into a scientific paper, designing the methods and analysis. I wrote the paper based on thorough reviews and discussions with my co-authors. Co-authors were also involved in independent checking for accuracy, consistency and for building consensus for the study.

**Paper 2** Chomutare T, Årsand E, Fernandez-Luque L, Lauritzen J, Hartvigsen G. (2013) Inferring community structure in healthcare forums: An empirical study. *Methods of Information in Medicine*: 52(2)

**Relevance** This study explores community structures from user interactions. Network analysis is one of the more practical ways of understanding the nature of connections among objects. I used this abstraction in this study and the whole dissertation to understand how patient-to-patient dialogue works in IPC for diabetes. This study provides a solid basis for much of the work in the dissertation, by revealing that we can infer useful community structures from user interactions.

**My Contribution** My role in this study was to design an abstraction upon which the study could be based. I did the data preparation and analysis, as well as writing the paper. My co-authors were instrumental in methodological deliberations and review iterations.

**Paper 3** Chomutare T, Årsand E, Hartvigsen G. (2013) Characterizing Development Patterns of Healthcare Social Networks. *Network Modeling Analysis in Health Informatics and Bioinformatics, Springer*: 3(2);147-157

**Relevance** This study succeeds the study in Paper 4, and was a result of an invitation to extend a paper for the HIBIBI 2012 Symposium. This paper sought to articulate the unique patterns that characterized the growth and development of diabetes social networks. At this stage, the goal was to be able to delineate the traits visible in diabetes networks as opposed to other non-health networks such as Facebook.

**My Contribution** My role in this paper was to design the data analysis and the experiments for the study. My co-authors had a role to advise whether the invited paper

constituted a more than 40% new work above the preceding symposium paper. My co-authors were involved in internal review iterations, making corrections on manuscript.

**Paper 4** Chomutare T, Årsand E, Hartvigsen G. (2012) Temporal community structure patterns in diabetes social networks. *HIBIBI Symposium in Proceedings of the International Conference on Advances in Social Networks Analysis and Mining*, pages 210–217 ACM/IEEE, Los Alamitos, CA

**Relevance** This study precedes that in Paper 3. During the preparation of the study in Paper 2, we discovered the limitations of the methodology in explaining some community structures phenomenon. This paper represents and extension of the abstraction to better understand how the networks form and develop over time. In this work I also explore the attributes of the users in the social networks in investigating forces behind the development patterns.

**My Contribution** I wrote the scientific paper, did the data preparation and analysis. My co-authors contributed to the methodological deliberations, and review iterations for the manuscript.

**Paper 5** Chomutare T, Tatara N, Årsand E, Hartvigsen G. (2013) Designing a diabetes mobile application with social network support. *Studies in Health Technology and Informatics, IOS Press*: 188:58-64.

**Relevance** This pilot study ran for 12 months, where actual patients used a diabetes mobile application with social network support. The aim of the study was to test if the social network concepts were feasible and if current work had any practical relevance for much older adults with Type 2 diabetes.

**My Contribution** My role was to design the study, coordinate development of the social application, and develop recruitment strategy and patient focus group meetings. I also did most of the data analysis with measurable help from my co-author NT. My co-authors were involved in methodological deliberations and planning for recruitment and user meetings, and manuscript review iterations.

**Paper 6** Chomutare T, Xu A, Iyengar MS. Social Network Analysis to Delineate Interaction Behaviour that Predicts Weight Loss Performance. *SUBMITTED*

**Relevance** This study develops tools for understanding the relation of user interaction behaviours to health outcomes. I use the example of weight loss and obesity since this is also a major risk factor for type 2 diabetes. Combining classification techniques, and the

now solid community structure analysis, the study articulates the interaction behaviours that affect weight loss.

**My Contribution** My co-author AX contributed to identifying a problem domain and data search, and my role was defining the problem, and designing the analysis and experiments suitable for analysing the problem. My co-authors contributed to manuscript review iterations.

**Paper 7** Chomutare T. Collaborative Filtering with Community Structure Properties in Healthcare Social Networks. *MANUSCRIPT*

**Relevance** This study is another illustration of the value of the community structure analysis. The results demonstrate the relevance of user interaction behaviour in personalization of IPC, based on collaborative filtering.

**My Contribution** I initiated and developed the idea and also wrote the paper. At this time, the point was just to solidify the analysis by exploring multiple application areas.

## 1.6 Organization of the Dissertation

The next chapter deals with the theoretical background of the work, and Chapter 3 goes into the methods we used. Chapter 4 introduces network analysis in online communities; the design of the networks, development of the networks, variations and unique characteristics. Chapter 5 is an application of network analysis in enhancing other machine learning techniques in classification and collaborative filtering. The second to last chapter details a 12-month pilot study with actual diabetes patients, while the final chapter concludes the work.

# Chapter 2

# Background Theory and Literature

> *Chapter Synopsis - First the background for open participation of patients online is presented. Then as orientation to the broader context of eHealth in diabetes; a systematic review of mobile applications that highlights the state of social media in these applications. The latter part of the chapter reviews literature related to network analysis, community detection and its previous and potential applications in healthcare.*

## 2.1  Open Internet Communities for Patients

*"Open"* Internet communities or IPC in this dissertation refer to online communities that are *patient-driven*. They are usually not associated with established general practice, hospitals or health service organisations. These are places where patients voluntarily register to participate for support from peers. Therefore, they are not places where patients go with the expectation of getting professional advice from a physician. Rather, they are places where patients can share everyday experiences with the disease.

It naturally follows from this that advice about everyday experiences comes from other people with diabetes rather than trained professionals. Such advice is based on reports of mundane activities that cannot all be documented in medical books and clinical guidelines. However, it is also important to note that most of these patient-driven IPC do have *moderators* who ensure patients adhere to the community's guidelines for participation or code of conduct.

However, because these open communities are not regulated and do not have the support of trained health service professionals, it is hardly surprising that there maybe other non-monetary costs of participation, which are discussed next.

### 2.1.1 Case in Point

Currently, most patient communities do not offer personalized experiences although they may have the capacity to do so. They offer a façade of social vibrancy, but as revealed in this dissertation, some of the communities are not as vibrant as they initially appear. Fig. 2.1 is an illustration of the kind of emails one could get when they enroll on an IPC. The emails are not targeted or personalized, and in this instance, the email contains everything, from popular stories to pregnancy. The recipient is neither married nor pregnant nor female, but receives emails sometimes more than twice a week. The emails can be considered nuisance reminders or spam since the same information is on the website.
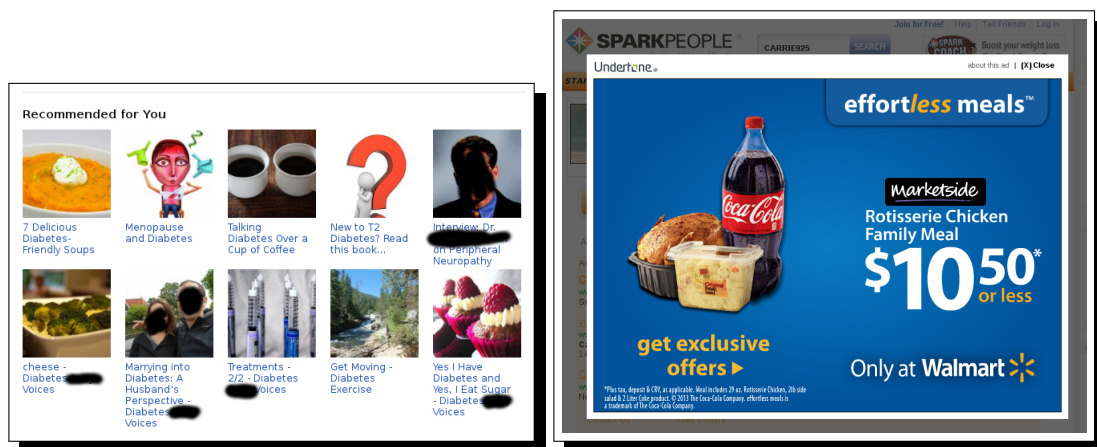


FIGURE 2.1: Email sent from my subscription to an online diabetes forum. The second figure shows a very obtrusive advert of junk food on a weight loss website.

Worse still, one weight loss website featured a very obtrusive advertisement featuring cheap junk food for "the whole family" (see Fig. 2.1 b.). This goes beyond the technical requirements, and presents significant ethical shortfalls. Such behaviour can be considered obscene and unethical; preying on the vulnerable patients who are struggling to lose weight.

The point is that a lot is going on in these open and unregulated social networks, and without the support of professionals in the health service, the potential of these communities may be harmed by unregulated commercial greed.

Perhaps a logical starting point might be to try to understand the interactions and how they affect the health of users; as means to both highlight the problems and potential, as well as accelerate the research. This dissertation aims to do both.

### 2.1.2  Research Mind Map

The whole dissertation work can be conceptualized in the mind map in Fig 2.2 where the main branch from the core is the work regarding social network analysis (SNA) . However, there is also an initial review work that had a larger scope, including a pilot study of a social mobile application for diabetes.

While the two minor branches ( the review and pilot) are loosely connected to the main SNA branch, they helped provide the background, the context and the demonstration of practical relevance of the work. The succeeding subsections describe the background theory and literature review of the branches.
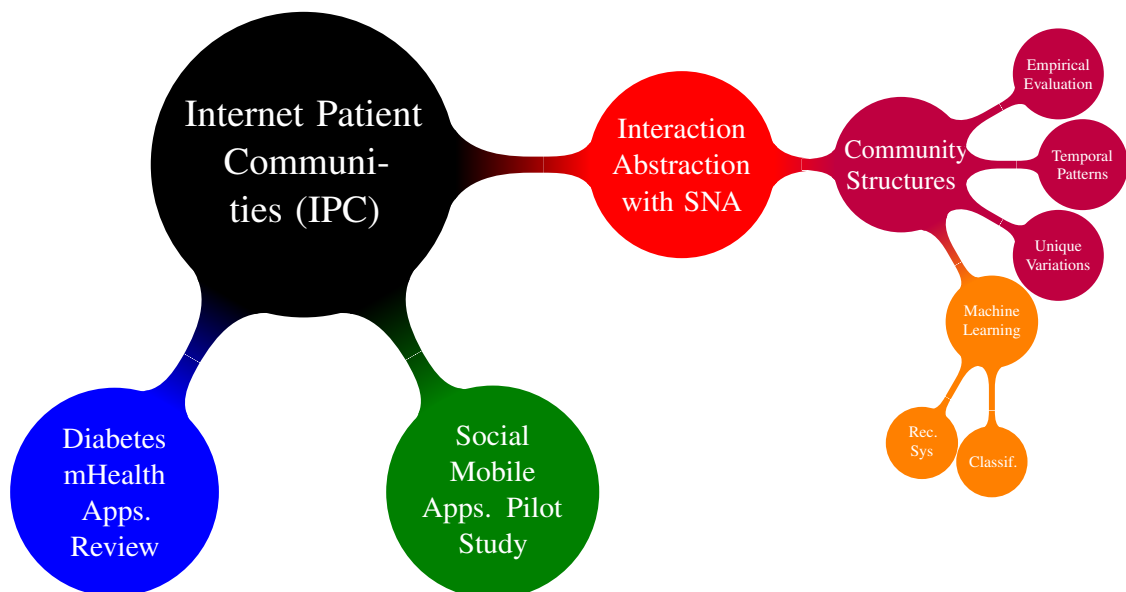


FIGURE 2.2: Mind map showing the minor branches (blue and green) and the major branch (red) of the research.

## 2.2  The Relevance of eHealth, mHealth and IPC in Diabetes

Although there is now a wide body of literature on the use of mobile health (mHealth) applications and the Internet in self-management of blood glucose (SMBG) , present knowledge about good practice in designing integrated health applications seems rather

limited. We have not found research focused on the gaps between the functional require-
ments (evidence-based recommendations in clinical guidelines) and the functionality
available in current tools.

Next is a systematic review conducted with a view to revealing research gaps, and the
potential role of social media tools as part of disease management regimes for chronic
illnesses, and this serves to establish the relevance (*see* Research Question **Q1**).

### 2.2.1  mHealth Applications for Diabetes

It is important to note that the review was initially done in early 2011. At the end of
the review (*see Update section 2.2.2*) a note is made on some of the changes that have
happened since, but the core methods designed for this review remain valid.

The goal was to review as many and as diverse diabetes mobile applications as possible,
both in the literature and in commercial markets, since the mobile platform has become
very popular. Many successful applications do not have any grounding in research, hence
the decision to include the online markets and grey literature; where novices showcase
their innovation, sometimes based on personal needs. While the literature typically
reflects emerging applications and new trends, the market gives a good indication of
mature applications and functionality.

**Inclusion Criteria**

The main inclusion criterion was that the application had a feature for blood glucose
self-monitoring. This filtered out applications intended exclusively for medical profes-
sionals rather than patients, as well as other general health and lifestyle applications.
Applications without English-language user interfaces were excluded. Also excluded
were hardware-based solutions geared toward blood glucose tracking or insulin pumps
only. Applications with their latest updates or publications prior to 2006 were excluded.

**Search Strategy**

The search was based on two main source types. The first source was online journal
databases, indexers, and reference lists. We searched for prototypes and work in progress
using the search terms *"diabetes," "mobile," "PDA," "cell," "phone,"* and *"application"*.
We constructed a search string using both the conjunction "AND" and the disjunction
"OR" logical operators:

```
(diabetes AND [mobile OR PDA OR cell OR phone OR application])
```

The search was based on the metadata — that is - title, abstract, and keywords. We targeted both original research papers and review articles indexed by Medline, ScienceDirect, ACM (Association for Computing Machinery) Digital Library, IEEE (Institute of Electrical and Electronics Engineers) Xplore Digital Library, Google Scholar, and DBLP (Digital Bibliography and Library Project) Computer Science Bibliography.

The databases reflect the multidisciplinary nature of the research involving both medical and computer science fields. We identified three recent relevant reviews by Årsand et al (2011), Tatara et al (2009), and Liang et al (2011), where we cross-checked descriptions. We also searched the grey literature: technical reports, Internet blogs, and portals.

The second source was online stores for mobile applications, using the search terms *"diabetes"*, *"glucose"* and *"sugar"* with the disjunction *"OR"* logical operator:

```
(diabetes OR glucose OR sugar)
```

We identified online stores for four leading platforms: Apple iPhone, Google Android, BlackBerry, and Nokia Symbian.

**Data Extraction**

We analyzed the following features:

1. Self-monitoring:
   (a) Blood glucose,
   (b) Weight,
   (c) Physical activity and Diet,
   (d) Insulin and medication,
   (e) Blood pressure,
2. Education,
3. Disease-related alerts and reminders,
4. Integration of social media functions,
5. Data export and communication,
6. PHR synchronization or portals.

These features are the result of iterated brainstorming sessions among the co-authors and discussions in focus group meetings with patients and physicians. The emphasis in these sessions was put on translating guideline recommendations into a requirements specification implementable on a mobile phone platform. We created a list with multiple features and in iteration reduced the list to six main features, which we believed had the most potential for enhancing future mobile applications.

These features are individually quite distinct, but they have the potential to work as an integrated self-management tool. For example, the user could log weight, physical

activity, meals, or carbohydrate intake, and have an easy-to-understand visual display to see how they correlate or affect the blood glucose. It should be noted that the *insulin* feature in most applications was part of a customizable *medication* feature for managing other medications as well.

We installed and tested most of the reviewed application, and where impossible, we cross-referenced the function descriptions in published articles. We noted whether each of the functions required manual interaction with the user, or whether wired or wireless sensors were used to import data into the application automatically. We then compared the prevalence of features with the recommendations in several clinical guidelines (see Discussion section for references to guidelines). Guideline recommendations can provide a good basis for requirements analysis and specification during the design and development of diabetes applications.

The process of extracting the data presented a major risk of error and uncertainty. For example, the literature is in most instances implicit about the functionality, and it is easy to miss or misunderstand feature descriptions within the text. To avoid potential problems, we enhanced the assessment process with independent verification. While we cannot claim the process we designed is entirely infallible, we avoided likely pitfalls by using building consensus and inter-rater agreement analysis.

**Results of the Review**

The breakdown of the search process from online journal databases, grey literature, and online markets is shown in Fig. 2.3. As illustrated in the figure, the total matches were 485 for literature and 488 for online markets, bringing the total matches for this study to 973. We went through a sifting process, with 36 applications from the literature and 101 from the online markets remaining, ending at a total of 137 mobile applications.

Of the selected 101 market applications, 40 were available for free. The mean and modal price for the rest of the applications was the equivalent of €2.50 and €1.50, respectively. Of the 40 free applications, 12 had some premium functionality available only at an additional cost.

Some applications were counted multiple times—that is, for each platform or source on which they appeared. Of the total 137 eligible applications, we installed 82 on mobile devices for further analyses and classified the rest as either work in progress or unavailable for installation. Three of the 82 installed applications—namely *Tag-It-Yourself*, *BANT* and *Few Touch*—were from the literature.
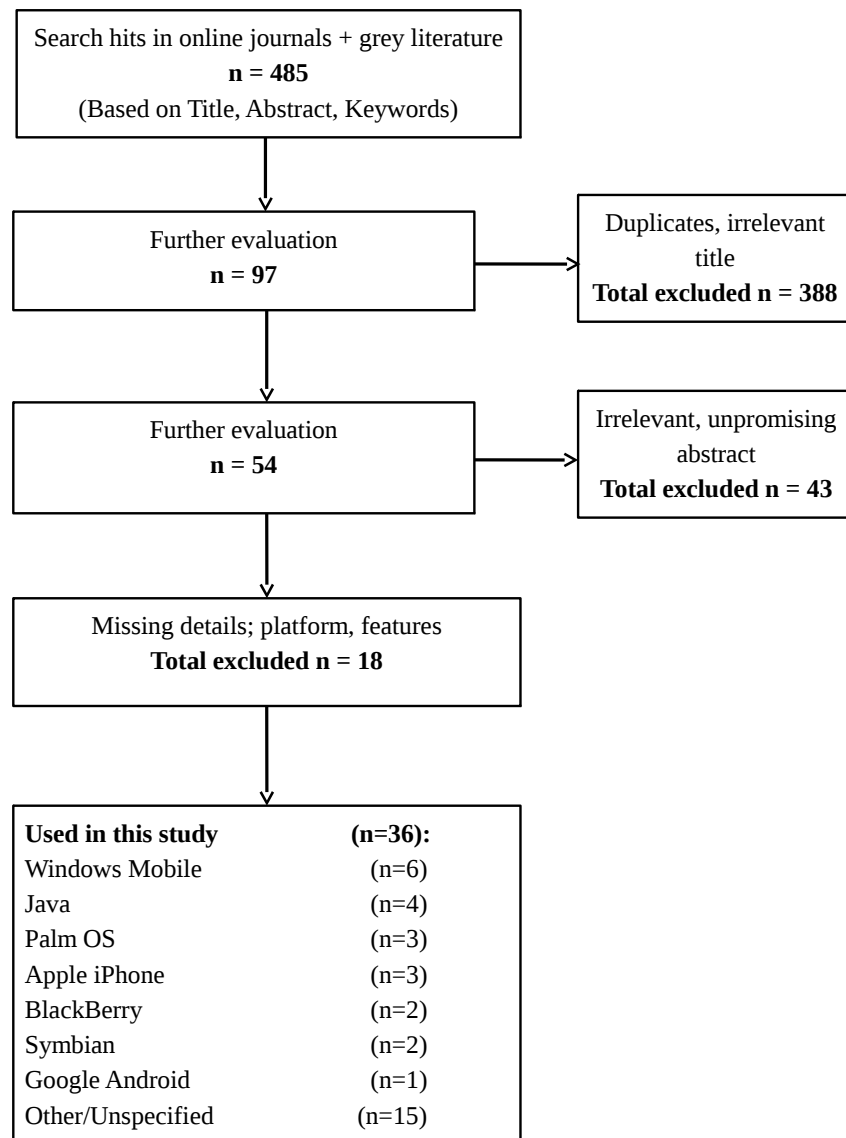
FIGURE 2.3: Selection process for online journal databases and online markets (Chomutare et al., 2011) Fig.1 (*see update section 2.2.2*).

It is important to note that some studies used commercially available applications but did not explicitly refer to the application names or features, and were thus excluded from this study. Our search was based on the title, abstract, or keywords, but even this streamlined search criterion is bound to yield many irrelevant articles. On the other hand, most of the articles that matched the search criteria in information and communication technology journals turned out to contain relevant data for this study. Abstracts that were judged to have low probability of containing relevant data were labelled as unpromising and excluded from this study.

The extracted features of the mobile applications per mobile platform and source are summarized in Table 2.1. The numbers include the total results from the online stores,

journal databases, and grey literature. Explanations of the features are given in Table 2.2. The blood glucose monitoring feature is not shown in Table 2.1 because it is a part of all applications as implied by the selection criteria.

Table 2.1 shows that tools for tracking insulin or other medication were present in 89 (65%) of the applications, although most online market applications did not specify whether the application was meant for type 1 or type 2 diabetes. Just over half of the applications had some form of diet management, either by tracking carbohydrate intake or by providing meal suggestions. Physical activity and weight tracking had 55 (40%) and 53 (39%) applications, respectively. A component for synchronizing with PHRs or Web portals was present in 40 (29%) of the applications. Only seven of the 27 applications with an educational module had personalized education, tips, feedback, or advice.

Few applications were sensitive to the age or gender of the users; important specific factors for special user groups such as pregnant women, for example, were largely ignored. Some form of lightweight integration with social media was present in 21 (15%) applications, while 16 (12%) had disease-related reminders. Of the applications randomly sampled for verification checking, 7 (5%) of the 130 features analysed were in disagreement. None of the disagreements concerned features related to our main findings.

The results in Table 2.1 are revealing in several ways. Perhaps the most significant outcome apparent in the table is that education is a feature present in only a few diabetes-related mobile applications. Second, we can observe that a small percentage of applications have social media, suggesting that the influence of social media on the development of diabetes mobile applications is so far negligible. In the remaining subsections, we discuss the details of these results.

**Functionality versus Requirements**

To discover whether the requirements from clinical guidelines were necessarily met, we turned to what was available on the online markets. However, it was impossible to accurately determine how many of the applications available on the commercial market were used in research or were founded on evidence-based principles. Most applications used in the literature integrated with a PHR, despite the intricacies associated with PHR integration. Outside well-controlled research, however, it is typically more difficult to offer PHR features for facilitating collaborative care or communication with healthcare facilities.

TABLE 2.1: Numbers and percentages of applications (n = 137) with the respective features of insulin, communication (Comm), diet, physical activity (PA), weight, blood pressure (BP), personal health record (PHR), education (Edu), social media (SM), and alerts

| Application Platform | Insulin | Comm | Diet | PA | Weight | BP | PHR | Edu | SM | Alerts |
|---|---|---|---|---|---|---|---|---|---|---|
| Apple iPhone (n = 49) | 35(71%) | 36(73%) | 26(53%) | 17(35%) | 19(39%) | 13 (27%) | 7(14%) | 8(16%) | 12(24%) | 7(14%) |
| Google Android (n = 33) | 19 (58%) | 17 (52%) | 15 (45%) | 10 (30%) | 16 (48%) | 16 (48%) | 7 (21%) | 3 (9%) | 0 (0%) | 0 (0%) |
| BlackBerry (n = 13) | 5 (38%) | 6 (46%) | 3 (23%) | 2 (15%) | 5 (38%) | 4 (31%) | 1 (8%) | 2 (15%) | 4 (31%) | 0 (0%) |
| Nokia Symbian (n = 6) | 3 (50%) | 2 (33%) | 4 (67%) | 4 (67%) | 4 (67%) | 3 (50%) | 2 (33%) | 2 (33%) | 1 (17%) | 1 (17%) |
| Avg. markets (n= 101) | 63 (62%) | 61 (60%) | 47 (47%) | 34 (34%) | 43 (43%) | 36 (36%) | 17 (17%) | 16 (16%) | 17 (17%) | 8 (8%) |
| Avg. literature (n = 26) | 17 (65%) | 16 (62%) | 17 (65%) | 15 (58%) | 7 (27%) | 6 (23%) | 18 (69%) | 10 (38%) | 3 (12%) | 7 (27%) |
| Avg. grey lit. (n= 10) | 9 (90%) | 4 (40%) | 7 (70%) | 5 (50%) | 3 (30%) | 2 (20%) | 5 (50%) | 2 (20%) | 0 (0%) | 1 (10%) |
| Total weighted average | 89 (65%) | 81 (59%) | 71 (52%) | 55 (40%) | 53 (39%) | 44 (32%) | 40 (29%) | 27 (20%) | 21 (15%) | 16 (12%) |

TABLE 2.2: Types of features and the comments about them, based on the way the features are currently implemented.

| Function | Input/Output | Summary |
| --- | --- | --- |
| **Blood Glucose** | Glucometer value, pre/postprandial tagging | Users enter values and can view graphs, with low, high and normal ranges well demarcated |
| **Weight** | weight scale value (kg,lbs) | Users enter values and can view graphs, can also get BMI if height is supplied |
| **Physical Activity** | Activity type, intensity, duration | Highly manual task. Users have to choose activity type, intensity, duration. Varied output in graphs and smileys, or other motivational output. |
| **Diet** | food eaten, estimate carbs, recommendations | Highly manual task. Users have to estimate carbs or navigate a comprehensive food menu. Output can be in graphs or other motivational output. |
| **Insulin** | insulin type, amount, pre/postprandial tagging | Users enter values and can view graphs |
| **Blood Pressure** | meter value | Users enter values and can view graphs |
| **Education** | tips, diabetes information | Some applications had link to another application for educational material |
| **Alerts** | reminder | Users enter reminders manually, and they also get reminders for postprandial testing |
| **Social Media** | social data | No real integration. Some applications have a link to their social networking websites or forums. |
| **Communication** | email, import/export csv, pictures, graphs | Most application allowed exporting data to a spreadsheet or emailing data or pictures |
| **PHR** | health record | Those that have PHR integration synchronize the PHR with the application seamlessly. |

Recent advances reflected in clinical guidelines (Funnell, 2010; Paulweber et al., 2010), including NICE (2009) recommend the following features (in random order) as part of important elements of diabetes self-management:

- Education and personalized feedback;
- Diet management; Weight management;
- Physical activity;
- Communication and patient monitoring by primary care providers;
- Insulin and medication management;

> - **Other therapeutics (foot, eye care);**
> - **Psychosocial care;**
> - **Immunization;**
> - **Complication management.**

It is important to note that current applications meet the functional requirements list only partially. The last four feature in the list has not sufficiently implemented in any of the reviewed applications. Of interest is the *psychosocial care* which also include the social support gap that social media might be able to fill.

**Classification of Functionality**

Fig. 2.4 illustrates an arbitrary classification of the surveyed mobile applications on the basis of prevalence. The *Core* class comprises the four major features. *Standard* class functionality comprises weight management, blood pressure monitoring, and PHR integration. These have a significantly higher prevalence than the *Premium* class features, which comprise education, social media integration, and alerts.

In the future, we expect the ideal application to have all the features available as part of the *Core* application, resulting in an integrated, feature-rich system. The presented classification may be useful for application developers and intervention designers when considering the features to implement. In addition, the classification is intended to draw attention to the least prevalent and less well-studied features.

**The Missing Link**

Current results reveal something completely unexpected: only 27 (20%) of the applications had an education module, and only 7 (26%) of these met our criteria for personalized education or feedback. It is somewhat surprising that education is conspicuously underrepresented, even when consistently recommended by clinical guidelines.
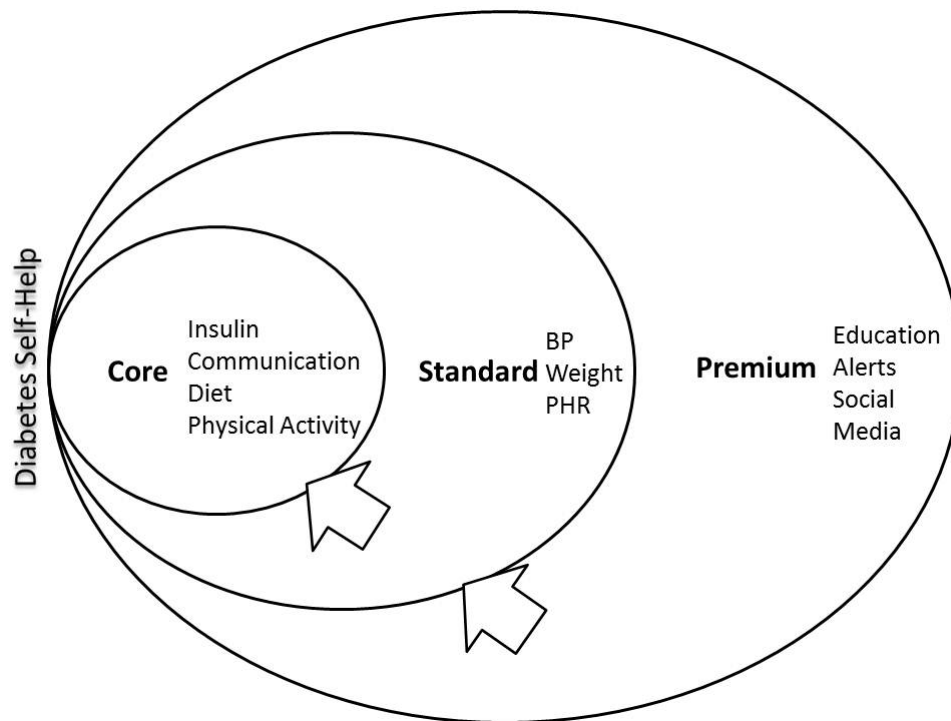
FIGURE 2.4: Arbitrary classification of functionality based on prevalence in the surveyed mobile applications. Adapted from (Chomutare et al., 2011). Fig.4.

Structured and personalized education and actionable feedback are widely suggested as the missing link for people with diabetes who do not use insulin.

A recent study (van Deursen and van Dijk JA., 2011) showed that, although Internet health information is growing rapidly, the average person lacks the skills for finding and using the health information strategically for his or her benefit. For people with diabetes who do not use insulin, personalized structured education may be the missing link to deriving benefits from SMBG (Clar et al., 2010).

**Social Media and mHealth for People with Diabetes**

In a recent survey, Chen (2010) showed the importance of social aspects and experience-sharing among people with diabetes. Chen's findings underscore the importance of individuality and the need for tailored social interactions, which resonate with the concept of *PatientsLikeMe* (Frost and Massagli, 2008; Wicks et al., 2010), which has recently received enormous attention.

Findings from this review suggest very little influence of social media on current diabetes mobile applications. Most applications that claim to include social media features only provide a link to their groups in well-known social networking sites such as Facebook and Twitter. Some applications also provide the user with an account to a forum. However,

there are no functional links or integration between information in the mobile application and the social media application. For instance, it is not easy to share graphs and data in the mobile applications with friends or relatives in social networks.

In terms of design and development of personalized education and social media, the task is challenging and the research field is still undeveloped. There is considerable scope for personalization because the mobile applications have access to some data about the users and their health status. Analysis of social media in healthcare is a rich and interesting field of inquiry that deserves urgent attention.

### 2.2.2 Review Update

The review was done about three years ago in early 2011 and there have since been several new developments. The number of mobile applications available to patients have increased dramatically, to more than 600 on the Apple platform alone. A new search is not necessary since the underlying methodology for analysing these applications remains valid.

Additionally, many of the mobile platforms considered in the review have become almost obsolete. For example, the winding up or mergers of Nokia with Microsoft and Blackberry. Since then, Google Android seemed to have gained a significant market share and remains has become a major player together with Apple.

Also important to note is that the succeeding work focuses one of the identified knowledge gaps in the review – and that is the nature and potential role of social media or Internet patient communities in disease management.

## 2.3 Patient Interaction Behaviours Online

In addition to missing personalized education, the preceding review revealed important research gaps regarding social media in healthcare. This dissertation will explore social media, with a view to understand social interactions among patients, and the impact they have on diabetes-related health status measurements or outcomes, and this forms background work for answering the Research Questions **Q2** and **Q3**.

The illustration in Fig 2.5 shows two important directions that this research could have taken. From an informatics point of view, social network analysis (SNA) and natural language processing (NLP) are important directions to understand the nature of IPC.
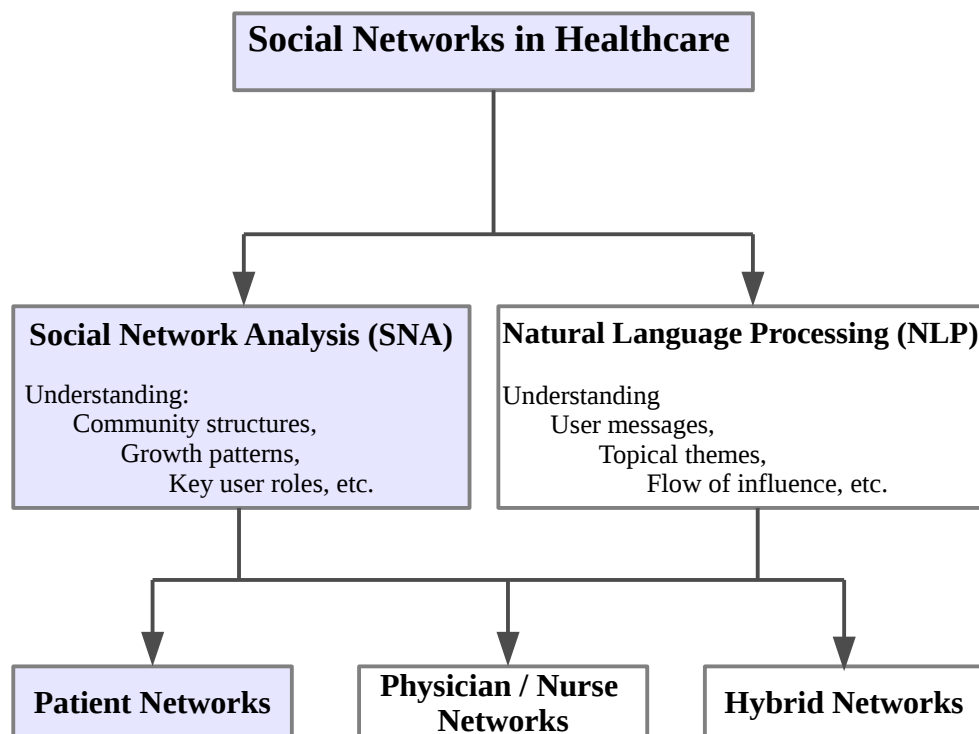
FIGURE 2.5: Important research directions from analysis of IPC from an informatics point of view.

Using NLP, we can detect topics and "hot" themes (Lu et al., 2013) to help automate content and sentiment analysis. In this dissertation however, the SNA route is taken because it seemed more natural since the basic understanding of interactions is key to unravelling the true nature of these open communities. NLP remains a natural next or complementary step.

### 2.3.1 Network Analysis Overview

Networks have now become common occurrence in scientific research. The exponential growth of interest in social media has resulted in increased interest in network analysis. This interest cuts across several domains in the recent past. In healthcare, much of the research has been largely uncoordinated and fragmented, making it difficult to positively identify scientific progress in the field.

Research on networks remains a rich field of enquiry, with potential for interesting new discoveries. One might ask why networks are so important. The spread of rumours or diseases, for example, depend on the underlying social connectivity of the communities. Without proper tools to abstract some of these complex social activities, we would
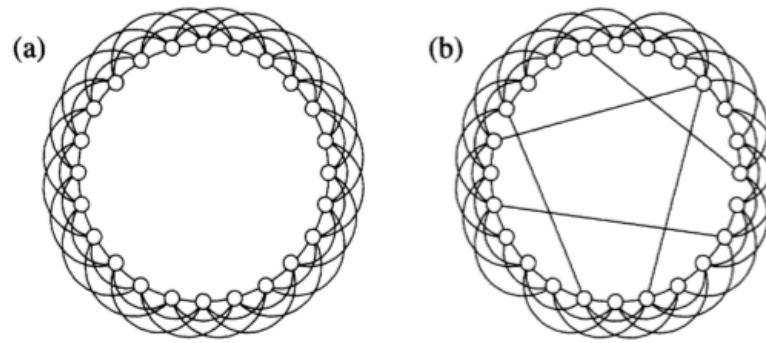
FIGURE 2.6: A one-dimensional lattice with connections between all vertex pairs separated by k or fewer lattice spacing, with k = 3 in this case, (b) The small-world model is created by choosing at random a fraction p of the edges in the graph and moving one end of each to a new location, also chosen uniformly at random. Source: adapted from (Newman, 2003a) Fig.11.

be overwhelmed by the amount of data. Even with abstraction tools such as network analysis, the networks can be quite complex; with morphing structures, and connections that have weight variations and directions. Next discussed are two properties that also have important implications in healthcare; small world networks and scale-free networks.

**Small World Networks**

First proposed by Watts and Strogatz (1998), a model of non-random networks. For instance, randomly replacing some of the links on a regular network such as a lattice, where each node has a link to every neighbour; see Fig 2.6 (a). The small-world model is created by taking a small fraction of the edges in this graph and reconnecting them as illustrated in Fig 2.6 (b). The procedure involves going through each edge in turn and, with probability $p$, moving one end of that edge to a new location chosen uniformly at random from the lattice, except that no double edges or self-edges are ever created.

This can have important implications for vaccination strategies for a particular infection (Polgreen et al., 2010). For example, we may want to vaccinate the most connected individuals rather than choosing individuals at random because the infection factors are usually not consistent with a random network.

Thus the findings indicated that it was easy to transform a network into a "small-world" network with short paths between any pair of nodes, and hence also more clustering. Watts and Strogatz (1998) conjectured that these short paths and high clustering held also for many natural and technological networks. Obviously, short paths provide faster connectivity between distant nodes in the network, thereby facilitating global coordination or information flow.

Many examples of small-world networks have been identified from biology to economy(Adamic, 1999; Jeong et al., 2000; Sporns et al., 2000; Stephan, 2000; Wagner and Fell, 2001). For instance, in epidemiology, small world networks have been proposed to explain how the clustering and contact networks influence the spread of infections(Boots and Sasaki, 1999; Jeong et al., 2000; Keeling, 1999).

**Scale-free Networks**

Studies of empirical networks have shown that some nodes are more highly connected than others, quite different from random graph models. This model predicts that the number of links of each node, *Pk*, will follow a Poisson distribution, but for many real-world networks *Pk* is highly skewed and decays much slower than a Poisson.

Barabási et al. (Barabasi and Albert, 1999; Barabasi et al., 1999) dubbed these networks "scale-free", where power laws arise and no single characteristic scale can be defined, but this was somewhat of a rediscovery (SIMON, 1955). They showed that such structures emerge in a model where new nodes are added randomly but attach preferably to existing nodes with a probability proportional to the degree of the existing node, that is, "preferential attachment".

These networks show interesting behaviour - that they are resistant to random failure if some nodes were removed, but may be highly susceptible to removal of nodes with high degrees, that is, hubs. Again, the implication in disease management communities could be to ensure that influential hub nodes (users) have adequate training and guard against misinformation.

### 2.3.2 Network Analysis in Health Informatics

Making a distinction between general social media and IPC is important for several reasons. First, because relationship dynamics in healthcare social media are still not well understood. Relationships in most general social media (such as Facebook) can easily be analysed because relationships are explicitly stated, and the origins of such relationships can easily be traced offline, for example, old school or college mates, childhood friends, conference connections.

Second, we find that most social media are not as personal as IPC because health information is generally considered private. Therefore, information is usually only released on a need-to-know basis, and most people are sceptical about releasing identifying information.

Third, in healthcare, we could easily assume that all relationships are based on the common interest in the disease, like diabetes, but no further information about specific relationships can easily be obtained. Unlike conventional social media, conversions in IPC is mostly about the disease, new drugs, diagnosis, symptoms, and related talk. Some users are bound to be more knowledgeable about the disease than others, and this has implications on the resulting networks. The succeeding paragraphs discuss potential key uses of network analysis in healthcare.

**Identifying hubs**

As alluded to in previous sections, hubs have a comparatively high number of edges connecting to them, which implies they are literal information hubs. In social media for managing disease, users who exhibit hub characteristics can be targeted for proper training to support peers. Further analysis could also help identify experts in the communities (Zhang et al., 2007).

**Identifying isolated peers**

With clustering and visualization algorithms, it is possible to identify users that may be struggling with the disease and those that are disconnected from main communities. It is important to identify potentially vulnerable users to ensure they receive enough attention and support by a trained mentor.

**Identifying temporal patterns**

Network analysis can also be used to examine community development over a period of time. Taking snapshots of networks at intervals, we can examine how the communities develop, and infer factors that sustain community development. For example, we could introduce a service or clinical intervention and then analyse the changes in communities before and after the intervention. We can also estimate users' knowledge development as the users attach and detach from communities.

**Social Networks for Patients or IPC**

One growing source of support for chronic illnesses is social media, which is emerging as a promising platform for "online network therapy" (Coiera, 2013). The current number of users are in the millions, with websites ranging from a few thousand users to hundreds of thousands.

Despite the fact that little is known about its effect on health outcomes and behaviour change, IPC seem to continue to grow in popularity. In a recent study, Lau at al.(2013)

found social features, such as the forum and polls, in a Personally controlled health management systems, were the most popular among participants. The continued growth is indicative of a new role that social medial plays in healthcare, although the nature of the role is still not well-understood.

Christakis et at.'s (Christakis and Fowler, 2007) influential study concluded obesity can be contagious through physical social networks. Although these findings were later criticised for ignoring other phenomenon such as homophily or assortativity (Newman, 2002b), they provided the impetus for further studies on online social networks. More recently however, advances in complex network analysis have allowed us to consider previously unknown empirical network structure properties.

Thus, on the one hand we have an approach based on understanding the structural properties of social networks (Cobb et al., 2010; Durant et al., 2011; Ma et al., 2010), while another approach, such as Weitzman et al.'s (Weitzman et al., 2011) work with a diabetes online community, have viewed healthcare social media in terms of its public health surveillance value. These approaches have potential for interesting new results in healthcare, but only a few studies have examined how online interaction patterns relate to behaviour change and health outcomes.

### 2.3.3 Community Detection in Networks

One classical example of a community within a network is Zachary's (Zachary, 1977) network of karate club members, illustrated in Fig. 2.7. It consists of 34 vertices, representing members of a karate club in the United States. The members were observed over a period of three years. Edges connect individuals who were observed to interact outside the activities of the club.

At some point, a conflict between the club president and the instructor led to the fission of the club into two separate groups, one supporting the instructor and the other, the president. The question is whether we can derive the composition of the two groups form the original networks structure. The split was over communities of 33 and 34 edges, with 3 edges in between the two, that are commonly misclassified by algorithms.

Most current algorithms are based on the iterative bisection of the network into increasingly smaller groups. Two bisection algorithms are the most common approach: spectral bisection (FIEDLER, 1973; Pothen et al., 1990) and Kernigham-Lin (B. W. Kernighan, 1970). The first depends on the spectral of Lapalacian and the second starts with a initial separation that is improved using a greedy approach.
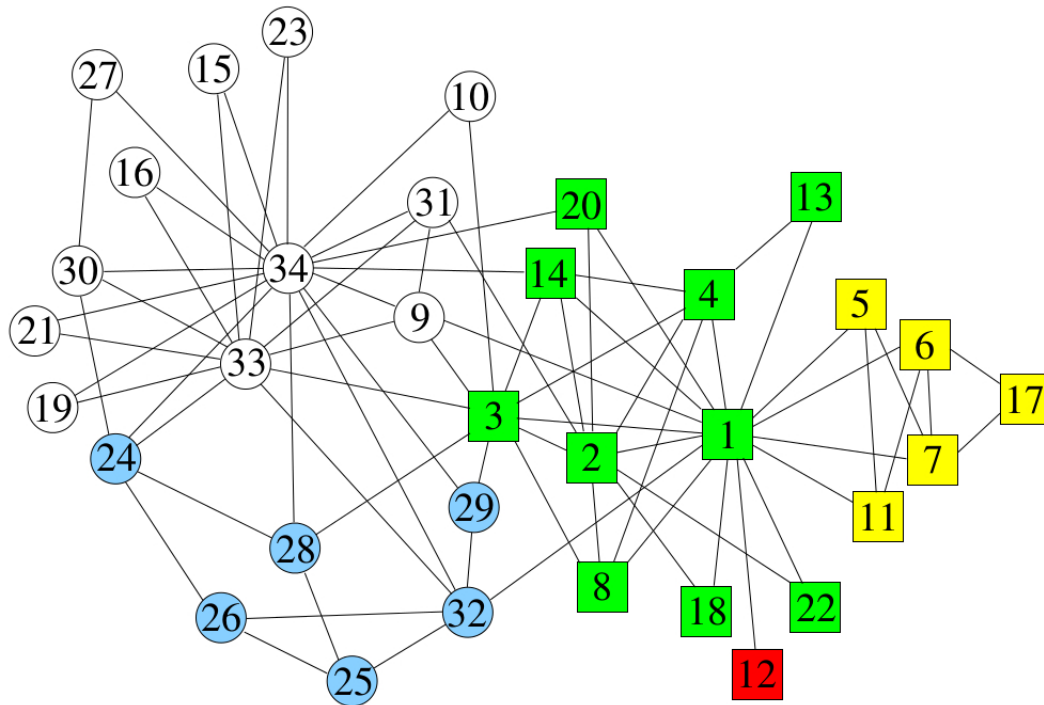
FIGURE 2.7: Zachary's karate club, a standard benchmark in community detection. Source: (Donetti and Muñoz, 2004) Fig.4.

More recently, algorithms have emerged based on the concept of hierarchical clustering (Newman, 2004). The idea behind this technique is to develop a measure of similarity $x_{ij}$ between pairs $(i, j)$ of vertices. Starting with an empty network of $n$ vertices and no edges, edges are added between pairs of vertices in order of decreasing similarity, starting with the pair with strongest similarity.

Another commonly used algorithm for finding communities is the Girvan-Newman algorithm (Girvan M, 2002), where edges in a network that lie between communities are removed, leaving behind just the communities themselves. These edges are identified using a graph-theoretic measure, *betweenness*, which assigns a number to each edge. The number is larger for edges that lie "between" many pairs of nodes. However, the algorithm has high complexity, making it impractical for networks of more than a few thousand nodes. Other approaches are based on removing vertices rather than edges (Holme et al., 2003).

**Empirical Analysis in Healthcare**

The idea of discovering communities from forum interactions is not new; researchers in a number of fields have always been fascinated by the prospect (Dawson, 2006; L'Huillier et al., 2010). Different methods have been discussed extensively in the literature, but work by L'Huillier (L'Huillier et al., 2010) enhanced our understanding of how forum

discussions can be analysed and connected using network analysis and text mining. While the work focused on terrorism and used a comparatively far smaller number of users, it nonetheless shed light on modeling networks from forum interactions for the general case.

In healthcare, network analysis has been used in many different unrelated scenarios (Durant et al., 2011; Ma et al., 2010; Pfeil and Zaphiris, 2009; Scott et al., 2005; Wegrzyn-Wolska et al., 2011). Ma (Ma et al., 2010) analysed a healthcare forum for weight changes and the influence it had on the weight of people in relationship circles, and Burton (Burton et al., 2012a) extended the analysis to video social network (YouTube) interactions in public health. Cancer forums have also been analysed to identify temporal patterns and influential topics that promote community growth (Durant et al., 2011).

Previous research on temporal trends has included group evolution dynamics (Bródka et al., 2012; Chakrabarti et al., 2006; Lin et al., 2008; Palla et al., 2007). Other studies focused on content popularity and predicting social ties (Almansoori et al., 2012) or links (Ahmad et al., 2010) and information flow (Yang and Leskovec, 2011). Although many previous studies substantially enhanced our understanding of group evolution dynamics, far less attention has been paid to healthcare networks.

Ma et al.(2010) analysed temporal weight changes over a five month period. The study reported positive correlations between the user neighbourhood size and the weight changes in the user's neighbourhood. Although the study was done for only a short period, and therefore difficult to say if the noted correlations are sustained, it nonetheless enhances our understanding of online influence and its propagation over time. The only drawback could be that no references to known temporal models were made.

Another major study by Durant et al.(2011) analysed data from six cancer forums and identified growth stages for the different online communities as well as topics that promote growth, using a new phase detection algorithm and a response function. The study concludes that treatment discussions rather than diagnosis discussions are more engaging to cancer patients, and thus also promote growth.

## 2.4 Linking Interaction Behaviour to Health Status or Outcomes

This section prepares the background work for answering the last Research Question **Q4** – by introducing the idea of predicting health outcomes as well as influencing the outcomes through recommendation – all based on online interaction patterns or behaviour.

Previous work thus far has not linked IPC activities to health status measurements or outcomes, and studies that made this attempt had several shortcomings that make their findings not easily generalizable. First these studies have only considered a small number of patients because it is inherently difficult to recruit the number of participants and sufficiently diverse to simulate real-life IPC.

For example Newton and Ashley's (2013) study consisted of less than 59 participants using an IPC in a closely controlled fashion. Thus, the evidence so far suggests that patients who participate in IPC tend to perform better health-wise than those who do not (Newton and Ashley, 2013; Wijsman et al., 2013). It is conceivable that this approach to studying social media can easily result in a more homogeneous recruitment pool than what is apparent in real-life IPC.

This dissertation is based on analysis of large offline empirical datasets, which precludes some of the problems with real-time studies such as recruitment, the Hawthorn Effect and participant homogeneity . This makes it possible to retrospectively analyse, without influencing interaction behaviour or the resulting health outcomes.

To link interaction behaviour to health outcomes, enhancements of two machine learning techniques (classification and collaborative filtering ) are developed and it is shown how an understanding of patient interactions can enhance these techniques in the current context.

**Predicting Health Status or Outcomes**

Classification is a versatile technique for solving many problems in several domains, including healthcare. One of the most studied areas in healthcare is pattern recognition in medical imaging (Aguilar et al., 2013; Folkesson et al., 2005; Lebedev et al., 2013). In this dissertation, the problem of predicting health outcomes from patient interaction patterns is modeled as a binomial classification task. The underlying nature of the data will dictate the most appropriate classifier for the problem, determined through experiments with empirical datasets.

**Mining IPC for Recommendation**

Related to prediction of health outcomes is prediction of user interests as a means to influencing model behaviour. One of the most popular methods for developing recommender systems is collaborative filtering (Herlocker et al., 2004), where recommendations are based on a collective user or item data or both. This has become an even more popular method since the emergence of social media, because recommendations can be made by

user-based:

(i) first determining the similarity among users, and then

(ii) using that data to make predictions

or item-based:

(i) first determining the relationship between items, and then

(ii) predict items based on the user's current data

or a hybrid of several methods.

In this dissertation, collaborative filtering enhanced with patient interaction information is used to predict top-N threads. If we can predict user interests, then we may be able to influence the interaction in the direction most suitable for that particular patient or group of patients.

## 2.5 Theoretical and Conceptual Limitations

There are some limitations to the theoretical and conceptual approaches taken for this dissertation. Some of the limitations are described from the main thematic areas of the presented background work:

(i) the review of mHealth applications for diabetes self-management,

(ii) abstracting patient interaction behaviours using network analysis, and

(iii) linking patient interaction behaviours to health outcomes.

**Review of diabetes eHealth Applications**

Many of the applications found outside the official online stores were not available for installation. As a result, some of the functionality was recorded from only the description or from published articles. There may be discrepancies between the text description and the actual features, and some functionality is not apparent until the application is installed and tested.

Additionally, this study only analysed the availability of applications and their features, it lacks information about the users and usage statistics. However, a follow-up pilot study tested the aspects related to practical feasibility for the user group.

**Abstractions of Patient Interactions**

Although analysing Internet data offline may reduce or even remove the Hawthorne Effect, there are some other limitations. Perhaps one obvious limitation is that Internet data

cannot be relied on entirely. As in any other system that relies on human input including research surveys, the data is susceptible to inaccuracies; intentional and unintentional.

However, network analysis allow us to analyse "Big Data" based on user interaction patterns, and this is more reliable since users do not make conscious effort to create the patterns and it becomes much easier to identify significant trends or events.

**Linking Interaction Behaviour to Health Outcomes**

An important limitation related to health outcomes, such as blood glucose levels or weight, may be that there are more complex factors (eg. family situation, depression symptoms, etc.) that may affect health outcomes than just online interaction.

Unfortunately, it is impossible to capture all these factors that happen outside electronic applications. However, because we are dealing with an enormous number of patients, it is possible to reasonably isolate the role of online interactions, while exceptional circumstances can be treated as unimportant outliers.

## 2.6 Chapter Summary

The chapter gave some background on why it may be important to pay attention to what is happening in IPC, and accelerate the research. Then a wide-scoping systematic review of diabetes self-management using eHealth applications was presented, revealing a gap in knowledge of social media or online communities for managing diseases.

The latter half of the chapter went on to provide the literature related to network analysis and community detection, examining potential use cases such as identifying key people for mentorship roles or for public health purposes. The chapter also discusses exploitation of network analysis to understand how participation in health online communities may affect health or behaviour outcomes.

# Chapter 3

# Materials and Methods

*Chapter Synopsis -* *This Chapter discusses the common tools and methods used in the different sub-studies of the dissertation. The details include collection of data, pre-processing and designing the abstractions common to many of the individual studies. Naturally, the finer methodology and experimental details of the individual studies are discussed in their respective sections and chapters.*

## 3.1 Multidisciplinary Research

Research in medical informatics is often multidisciplinary ; using methodologies from both the informatics and clinical research. This often presents a challenge in terms of the audience because the research then targets both clinicians and researchers with informatics backgrounds. However, multidisciplinary research has been shown to solve important problems and often yields more interesting findings (Glattre, 1991; Hettne et al., 2007). This is not a surprise because these, seemingly disparate fields of research, are intertwined at some abstract level, resulting in some synergy.

It is important to not only apply existing informatics methods to healthcare data, but rather to develop new methods based on the unique characteristics of healthcare data. This dissertation develops new methods, from a unique ensemble of informatics tools, to enhance our understanding of this relatively new phenomenon - Internet Patient Communities.

**Approach Overview**

The approaches used for scientific discovery are twofold:

(i) Systematic review

(ii) Collection, modelling and analysis of empirical data,

(iii) Controlled experiments and feasibility pilot test

These approaches are consistent with scientific methods for acquiring new knowledge. Systematic reviews are important for synthesizing evidence, while the second approach, empirical data are collected from real-world IPC, and a model designed to abstract complex social elements of the data, and then analysis of the results. In the third approach, results obtained using the second approach are used to design experiments for enhancing machine learning techniques, as well as pilot testing an IPC on a mobile phone with real patients. The result was a generalizable framework for analysing IPC.

## 3.2 Overview of Sub-Studies and Progression Phases

This dissertation can be viewed as a collected of several small sub-studies, each of which contributed to knowledge. These individual studies can be grouped into four phases, as illustrated in Fig. 3.1, and these are discussed next.

The phases follow a chronological order, it should be noted that the pilot in Phase #1 was conducted over a 12-month period and the related scientific article (Paper 5) was published near the end of that period, with only the preliminary results. The rest of the results are reported in Chapter 6.

### Phase #1 – Relevance of eHealth, mHealth and Patient Interactions in Diabetes – related to Q1

Related Papers – `Paper 1, Paper 5`

There are two studies in this phase; `Paper 1` is a wide-scope systematic review mHealth applications for diabetes and `Paper 5` is a 12-month feasibility study of IPC on mobile devices.

The review was part of the initial work in the project to identify research gaps and establish relevance for further work. The pilot study was also part of initial work with IPC on mobile phones, to test the feasibility of use in typical living environments for older adults with Type 2 diabetes.
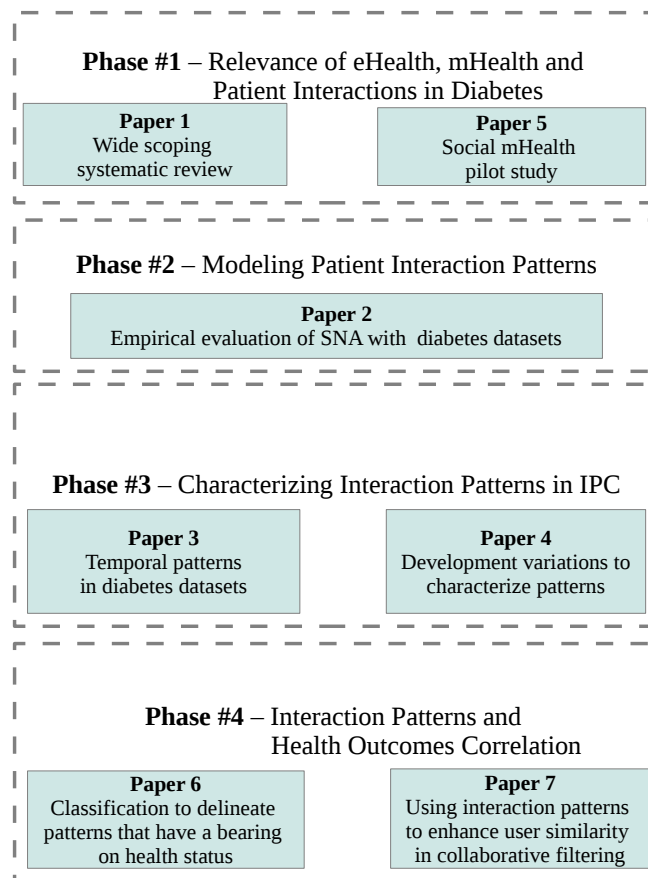
FIGURE 3.1: Phases of the research progression for the dissertation

**Phase #2 – Modeling Patient Interaction Patterns – related to Q2**

Related Papers – Paper 2

After research gaps are noted, the third study attempts to validate a usable abstraction of user interaction patterns in IPC, and Paper 2 starts the empirical evaluation of network analysis for diabetes datasets.

**Phase #3 – Characterizing Interaction Patterns in IPC – related to Q3**

Related Papers – Paper 3, Paper 4

There are two studies in this phase; Paper 4 develops temporal analysis of community structure in diabetes networks, while Paper 3 delineates development patterns that characterize IPC. This phase enhances our understanding of the nature of user interactions in IPC and how they develop in relation to other non-healthcare online communities.

TABLE 3.1: Datasets used in the studies and their sources. Note that only subsets of the datasets are used on some sub-studies.

| Type | Disease or Name | Description/Website[2] |
|---|---|---|
| Healthcare | Diabetes$_1$ – [D$_1$] | General diabetes community |
| | Diabetes$_2$ – [D$_2$] | General diabetes community |
| | Diabetes$_3$ – [D$_3$] | General diabetes community |
| | Diabetes$_4$ – [D$_4$] | Diabetes community for children |
| | Diabetes$_5$ – [D$_5$] | Diabetes community for juveniles (in Spanish) |
| | Obesity$_1$ – [Ob$_1$] | Obesity and weight loss community |
| | Obesity$_2$ – [Ob$_2$] | Obesity and weight loss community |
| | Obesity$_3$ – [Ob$_3$] | Obesity and weight loss community |
| Non-healthcare | Facebook | http://www.facebook.com |
| | Slashdot | http://www.slashdot.org |
| | Yahoo! Movies | http://www.yahoo.com |
| | Flixter | http://www.flixster.com |
| | MovieLense | http://www.movielens.org |

2.To protect the privacy of the patients and the websites, the URLs are omitted, but they have been included in the non-public Appendix C.

**Phase #4 – Interaction Patterns and Health Outcomes Correlation – related to Q4**
Related Papers – Paper 6, Paper 7

There are two studies in this phase; one study develops an improved classification method while the other develops an improved collaborative filtering method. In this phase, the aim is to show that the analyses of IPC are useful and that we can enhance machine learning techniques based on the proposed abstraction of patient interactions.

## 3.3 Data Sources and Extraction

The experiments are based on both healthcare and non-healthcare data sources, and the datasets are also available in anonymized form on my dissertation website [1] or by request.

**Crawling and HTML parsing**

---

[1]http://www.diabetesbuddy.org/datasets

The healthcare datasets were crawled from publicly available data on the Internet, while the non-healthcare datasets were obtained from related and unrelated studies from the literature. A well-behaved Python program was developed to download the relevant web pages by simulating user clicks over long periods at a time. The structure of the crawler program can be divided into three distinct modules:

*Module 1. Data Download*

This modules is responsible for the actual download of hypertext, and Listing 3.1 shows a key statement for requesting HTML pages, by modifying the headers to simulate browser clicks. Using a lower-level protocol monitor application Wireshark[3], it is possible to observe packet and header information such as cookies in order to get the `'COOKIE'` to appended to the headers as part of the HTTP request. It is possible to find the `User-Agent` string, which is based on the emulated browser environment, in this instance using Firefox browser the string could be as follows:

```
'Mozilla/5.0 (Windows; U; Windows NT 5.1; en-US; rv:1.9.2.13) Gecko/20101203
Firefox/3.6.13'
```

.

With the cookie and user-agent, the HTTP request could be made (*see* Listing 3.1), and the downloaded HTML data was appended to a text file, to form one large file for easier compression and handling.

```
1 request = urllib2.Request(url,headers={'COOKIE':cookietxt,
2   'User-Agent':user_agent,'Accept-Encoding':'gzip'}, data=data)
```

Listing 3.1: HTML Request for the web crawler

*Module 2. HTML Parsing*

With the data downloaded, the next step is parsing the semi-structured HTML text into a more structured format. The HTML in this instance was fairly structured and predictable, so I just used regular expressions to parse the data.

```
1 # page detection
2 re_pages=re.compile('">Page <strong>1</strong> of
    <strong>(\\d+)</strong>')
```

---

[3]http://www.wireshark.org

```
3
4  # page detection on forum pages
5  re_forum_pages=re.compile('">Page <strong>1</strong> of
      <strong>(\\d+)</strong>')
6
7  # get the page views and replies numbers
      re_topicviews=re.compile(r'<dl
8  class="icon"style="background-image:url\([^"]*?/([^/]*)\);
9  [^"]*">\s*<dt[^>]*>.*<a[^>]*href="[^"]*t=?(\d+)[^"]*"[^>]*
      class="topictitle"> [\w\W]*?<dd class="views">([\d,]+)
      <dfn>Views</dfn>')
```

Listing 3.2: Regular expression in Python for parsing HTML

In Listing 3.2, using regular expressions (regex) , we can obtain certain data elements
in the HTML, page numbers in this instance. The parsed data were stored in MySQL
databases and in structured text formats such as delimited files for easier pre-processing.

*Module 3. Data Formatting and Storage*

The main types of data stored in MySQL and structured text are:

- *User Topics* –  This database relation contains the connections among the users as
  they create threads and comments. A typical structure includes
  `|user-id-hash | post-date | post-id | . . .`
- *User Profiles* –  This database relation contains the profile data of the users, based
  on the public data that they shared on the Internet
  `|user-id-hash | join-date | age | gender | BMI | . . .`
- *Text Formats (ARFF, CSV)* –  These are examples of formats required by some
  machine learning and data mining APIs, but conversion can be done easily using
  several existing wrappers.

## 3.4  Data Modelling and Network Abstractions

The data modelling objective is to describe an abstraction of the complex social interac-
tions that happen in IPC among patients, and show that we gain new knowledge from its
use, being the building blocks for answering Research Question **Q2**.

TABLE 3.2: Description of key concepts used in the abstraction of patient interactions using network analysis.

| Key Term | Description |
| --- | --- |
| User | Someone who interact on the IPC website. A user can create a thread or comment on threads, and is also referred to as a "Patient". |
| Post | A message written (or object posted) by a user on a thread. |
| Thread | Series of posts (messages) written by users on a specific subject line. A thread has (i) a creator, (ii) a title/subject, (iii) a message and (iv) comments. |
| Thread creation | Is when a user initiates a subject line by posting a message or object. |
| Commenting | Is when users respond by posting a message or object to an existing thread, thereby creating a "comment". |
| Lurking | Is when a user does not actively participate in any thread. Lurkers generally view the threads and comments, without posting any message. |
| Out degree | Total number of posts that a user creates. |
| In degree | All the comments on threads started by a user. I consider this an act of "connecting to" the thread creator. |
| Degree | In degree + Out degree, also used as a description of the level of a user's activity in the IPC. |
| Self-loops | Comments written on thread that was originally created by the same user, ie. commenting on one's own thread. |
| Complex networks | Other complex networks terms used in this paper can be found in (Newman, 2003b). |

### 3.4.1 Key Definitions of Concepts

Table 3.2 provides a summary of the key terms used in the network design, which is an abstraction of patient interactions. These definitions are important to provide context to the abstraction, and to show that the abstraction method is suitable to apply to IPC.

### 3.4.2 Abstraction of Patient Interactions

We can think of patient interactions as one patient connecting to one or more other patients. Premised on this basic concept, the creation of threads and comments can also be thought of as one patient (the thread creator) reaching out to other patients. The patient that responds to the thread (commenter) in effect connects to the thread creator.
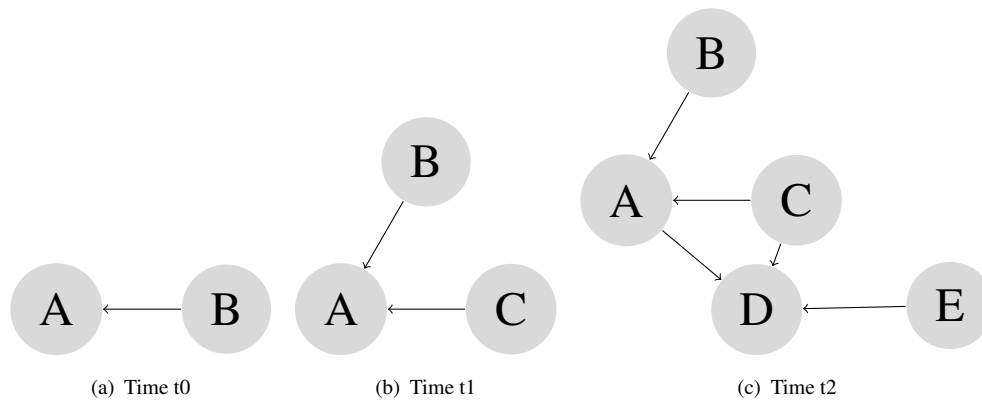
FIGURE 3.2: A network of thread creation and comments developing over time.

As an illustration, from the Figure 3.2 we can observe how a typical forum network could develop over time. For instance in Figure 3.2(a), a network is established at time $T_0$ when user *B* comments on a topic that was created by user *A*. At an arbitrary future time $T_1$, the topic creator can post a comment on his/her topic , and a new user *C* posts a comment on *A*'s topic as well.

At another even further time $T_2$, the user *D* creates a new topic and users *A*, *C* and *E* comment on the topic, thereby creating a small network. As this process progresses with many subsequent topics and comments, networks and sub-networks emerge, and the details are discussed further in the chapter.

### 3.4.3 Abstraction With $k$-Partite Networks

The network that emerges can be described in terms of graph theory , resulting in:

**Definition 3.1 — Network Model.** Graph $G = (V, E)$
where $V$= set of nodes and $E$= set of edges

This results in a bipartite network where the set of nodes contains both *Users* and *Threads* . The pseudocode in Algorithm. 3.4.1 specifies how the network is created. The process involves looping through the whole *Forum f*, where the forum contains *Threads t* and the threads contain *Posts p* in this hierarchical order.

The graph is then developed by creating a set of users and threads, and a set of edges. It is possible to work with such an abstraction, but a transformation to one dimension can

be made without losing much information, and this is discussed next.

---

**Algorithm 3.4.1:** CREATEBIPARTITENETWORK(*Forum*)

**comment:** Forum f contains Thread t which has Post p

**comment:** Returns a multidimensional network MdN

**for each** $t \in f$

$\mathbf{do} \begin{cases} \textbf{for each } p \in t \\ \\ \quad \mathbf{do} \begin{cases} Record\ r \\ r.user \leftarrow p.user \\ r.thread\_id \leftarrow p.thread\_id \\ network.append(r) \end{cases} \end{cases}$

**return** (*MdN*)

---

### 3.4.4 Reduction to One-dimensional Networks

A transformation of the network can reduce the dimensionality to one, as illustrated in Fig.3.3, where users $U_i$ and topics/threads $T_i$ are two different kind of nodes before the transformation. The transformation forms a one-dimensional network, with only users $U_i$ as nodes connected by threads.

First, an appropriate threshold that describes a valid or legitimate connection has to be determined. For instance, the strongest connection is between user $U_2$ and user $U_3$ because they have more than one thread in common, and we can choose to sever the rest of the connections. If we had more information such as user ratings on the thread, then the decision to keep or discard a network would be more clear.

The transformation is formally described in the pseudocode in Alg. 3.4.2. The result is a $GraphG = (V, E)$, where the set of nodes $V$ only contains the users and the set $E$
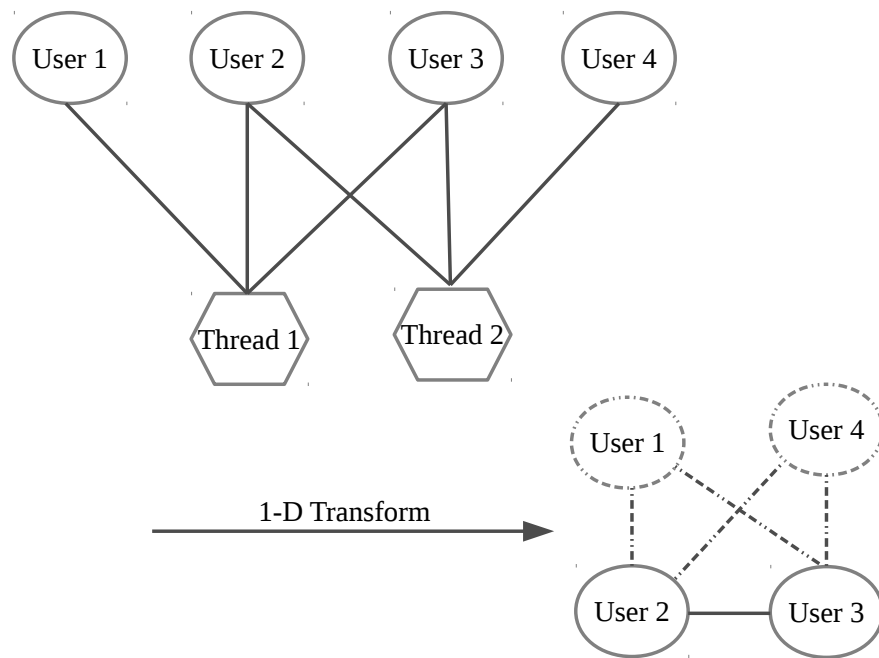
FIGURE 3.3: A two dimensional network comprising the users $U_i$ and the topics $T_i$ as nodes, and $W$ indicates the weight or rating. This bipartite network can be reduced to a one dimensional network as illustrated.

contains the edges.

---

**Algorithm 3.4.2:** REDUCETOONEDOMENSION(*MdN*)

**comment:** Multidimensional network MdN (user_id, thread_id)

**comment:** Returns a one-dimensional network (creator, commenter)

**for each** *thread_id* ∈ *MdN*

**do** {
    **comment:** Loop through all records with the same thread_id

    **while** *thread_id*

    **do** {
        **comment:** First poster in thread is creator, rest are commenters

        **if** *first post*
            **then** *creator* ← *MdN.user_id*
            **else** *commenter* ← *MdN.user_id*

        **comment:** Ignore redundent pairs in thread, and self-loops

        **if** ((*commenter, creator*) *pair exists* **or** *commenter* == *creator*)
            **then** *continue*
            **else** *network.append*(*commenter, creator*)
}

**return** (*network*)

From the pseudocode , a person who creates a *Thread t* is the *Creator*, and is connected to a person who comments on the thread, *Commenter*. The edges $E$ are directed, from a *Commenter* to *Creator*. It should be noted that an edge is immediately established at the first connection between a unique pair, and that no edges are created among the people that comment.

In addition, the *Weight* of edges increase when a unique pair of nodes appear on a different thread. The code loops through the multidimensional network *MdN* to identify these connections within different threads.

### 3.4.5  Abstraction with Dense and Sparse Graphs

The literature has not been specific how the used network abstractions were designed or developed. The type of design will affect the complexity of the process and the quality of the results. In this subsection, I compare different potential designs and show why I used the design used in this dissertation.

**Creator-Commenter Cycles**

Revisiting the multidimensional illustration in the previous subsections, I illustrate two scenarios in Fig 3.4, where we now have information on the user that created the threads. In this illustration, I assumed *User* 2 created both threads, and therefore all the other users are *Commenters* , meaning they all connect to the thread *Creator* .

The basic differences between the two is that the dense *Graph* $G = (V, E)$, where the number of edges $|E| = O(|V|^2)$, while the number of edges $|E| = O(|V|)$ in the sparse graph.

**Connection Types**

Although in this project I looked at inferred connections based on interaction in the social networks, there are multiple types of connections. For example, there are private messages and sharing of objects that could be defined as a legitimate network.

For privacy reasons, we decided not to explore private communications. This does not diminish the value of the networks we looked at because in most instances, connections are first established publicly before private communication, and the network is established immediately.
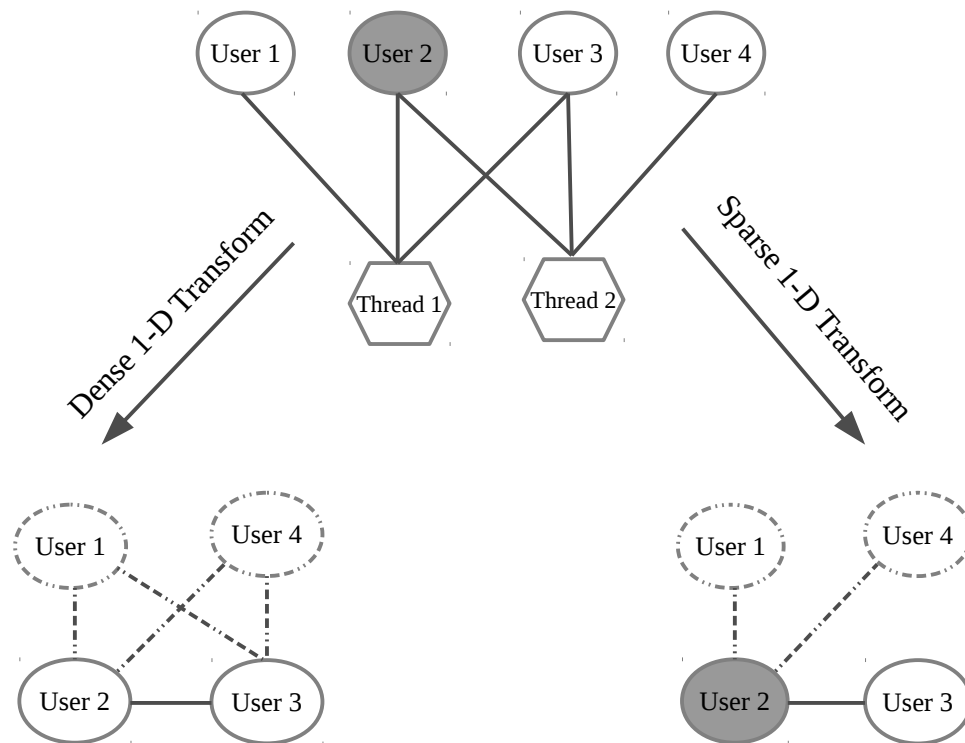
FIGURE 3.4: Illustration of two designs for abstraction user interactions

### 3.4.6 Community Detection

Community detection is a useful tool for delineating the community structure and potentially revealing hidden patterns encoded in massive amounts of user interaction data. Most community detection algorithms are based on the iterative partitioning of the network into increasingly smaller groups. One popular approach in common use recently is hierarchical clustering (Newman, 2004).

The idea behind this technique is to develop a measure of similarity $x_{ij}$ between pairs $(i, j)$ of vertices, based on a given network structure. Starting with an empty network of $n$ vertices and no edges, edges between pairs of vertices are added in order of decreasing similarity, starting with the pair with strongest similarity.

**Greedy Optimization**

The first community detection algorithm used was a hierarchical clustering algorithm, the greedy optimization (GO) (Clauset et al., 2004). This is a well-studied community detection algorithm, which is extremely fast and suitable for large networks. The algorithm is based on modularity maximization, where the number of edges within a

community are preferred to edges between communities, as illustrated in the algorithmic summary in pseudocode Algorithm 3.4.3 (Chomutare et al., 2013c).

---

**Algorithm 3.4.3:** GREEDYOPTIMIZATION($G = (V, E)$)

**comment:** The input is a network G = (V,E)

**comment:** Returns clustering C of the graph

$C \leftarrow$ *singletons* **comment:** initial clustering C of single node clusters

*matrix M*
**repeat**
  *find clusters $C_i, C_j \in M$ with highest $\uparrow$ modularity*
  *merge $C_i$ with $C_j$*
  *update matrix M*
  **comment:** has to stop when no improvements on modularity are possible
**until** $|C| \leq 0$

---

### Affinity Propagation

The second community detection algorithm used in experiments was the Affinity Propagation (Frey and Dueck, 2007); based on message-passing, where an initial data set is chosen at random and then refined in iterations. Obviously, the success of this algorithm is predicated on a good initial selection. The algorithm has a complexity of O($N^2$ ), where $N$ = size of the network.

Although there is arguably several ways of detecting communities (Leung et al., 2013), the two community detection algorithms were selected for the experiments based on trial and error to see the best performance and results. The Much of this work is described in Dias et al. (2012).

### Network-level Metrics

At the network level there are a number of measurements that are interesting in this context. Many of these are discussed in the literature (Newman, 2006), for example:
(i) the number of clusters,
(ii) average size of the clusters,
(iii) maximum/minimum size of a cluster,
(iv) modularity ,

(v) clustering coefficient,

(vi) network diameter and,

(vii) network density.

**Node-level Metrics**

Then at the node level I examined the centrality measures (Wasserman and Faust, 1994), which measure how important the individual nodes are, and these have important implication for practice in IPC;

1. *Betweenness*

   **Definition 3.2 — Betweenness.** Betweenness formally defined as

   $$C_B(v) = \sum_{s \neq v} \sum_{t \neq v} \frac{\sigma_{st}(v)}{\sigma_{st}} \tag{3.1}$$

   where $\sigma_{st}$ is the number of shortest paths from vertex $s$ to vertex $t$, and $\sigma_{st}(v)$ is the number of shortest paths from $s$ to $t$ that go through vertex $v$. In practice, users with high betweenness values indicate that they can be information conduits in the community or between communities. Depending on the network topology, we can identify points of failure for information dissemination. They are also important for testing the resilience of a network.

2. *Closeness*

   **Definition 3.3 — Closeness.** Closeness formally defined as

   $$C_C(v) = \frac{1}{\sum_{t \in V} d_G(v,t)} \tag{3.2}$$

   where $d_G(v,t)$ is the distances between vertices $v$ and $t$. The closeness depicts how much the user is assimilated in the network, that is, how close the user is to the rest of the users.

3. *Degree*

   **Definition 3.4 — Degree.** Degree formally defined as

   $$C_D(v) = deg(v) \tag{3.3}$$
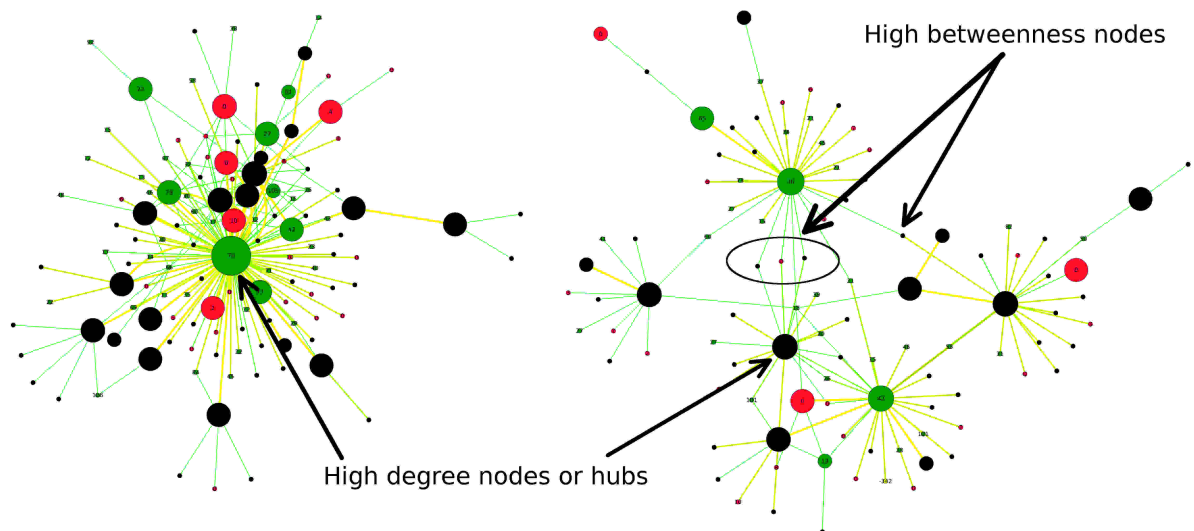
FIGURE 3.5: Empirical network showing the community structure and some key users. In the zoom-in image, Green = high performance, Red=Low performance and Black = no data. The nodes are sized by degree

where $deg(v)$ is the degree on the node $v$. This is the number of links that the user has. Given a directed graph, the degree becomes even more important in denoting authorities (high in-degree) and hubs (high out-degree). In practice, users with high degree are literary information hubs. The can act as central figures for social support in IPC.

### 3.4.7 Community Structure Visualization

There are currently several tools that can be used to visualize the community structure of networks. These visualizations provide unique perspectives that would otherwise be impossible.

Fig. 3.5 shows a zoomed-in visualization of a weight loss community structure. The figure shows how the hub nodes (*Users*) are organized and also illustrate the position of nodes with high *betweenness* values. In practice, nodes with high *betweenness* means that a lot of messages between sub-communities pass through the nodes. They act as some sort of conduits for connecting disparate communities.

The illustrated centralized structure, as explained later in this dissertation, is typical of most healthcare social networks. Contrasting with the very decentralized structure of non-healthcare social networks such as Facebook, I found the structure a plausible

discriminator of the two. I also explore this structure and develop methods to analyse how they develop over time.

## 3.5 Methodology Critique

### Web Crawling Ethics

Web crawlers can pose ethical concerns for several reasons. First, the crawling process can take up resources otherwise meant for surfing users. Second, while most crawlers are used by legitimate users for search engines, some crawlers copy website data for malicious use, such as creating clone websites. The crawler I developed was well-behaved; sending only a few HTTP request per period, as opposed to potentially causing denial of service attacks.

### Internet Data Use Ethics

Many ethical issues surrounding use of publicly available data on the Internet have not been dealt with in depth. For instance, international law is currently not clear on the ownership of data generated in online communities, especially for disease management.

In IPC, the situation is much different. We have users divulging a lot of information that is normally considered quite private. The average Internet user appear ignorant of the privacy concerns regarding the data they publish online.

Some studies, such as the seminal study on the topic by Eysenbach and Till (2001) and more recently (McKee, 2013), have raised concern over the ethical issues. Eysenbach and Till suggested that the rule of thumb is that if the online community is secured by a username and password, then patients' expectation of privacy can be reasonably assumed, and therefore permission or consent must be sought.

In this dissertation, only publicly available data is used and care was taken to anonymize and/or pseudonymize all data before the analysis. All the published data was based on aggregated and non-identifying data. Borrowing from the US "Fair Use" law, the rules of thumb for using publicly available data are easily universally applicable:

- Non-profit educational purposes; current purposes are purely for non-commercial research.
- The nature of the data; in this instance, it's connection of user-aliases without any actual content.

- Amount or proportion of data used; current studies only use a small proportion of the available data.
- The whole picture, including the potential market value or impact; the current work at best influences research practice and any impact on commercial markets is substantially unrelated nor can it be traced to any of the datasets.

**Approach Limitations**

One limitation may lie with the use of network analysis on threads and comments or replies. The method only uses data based on an active expression of interest, that is, when a user actually replies or responds to threads. The 'lurkers', that is, people who only view threads without being actively involved in conversions are excluded from the analysis. Although this does not seem like a major limitation, perhaps we could get more insight if the data on lurkers where used.

Another limitation may be that this dissertation has not considered several psychology and social theories that might have been relevant to the work. These theories might help explain many issues that some of the presented quantitative analysis might not be able to explain.

Therefore many factors that influence the patient's decision to engage in online conversations are not reflected in the current work. Additionally, there are many personal psychological and social factors that influence not only how patients interact, but also the impact on their health.

## 3.6 Chapter Summary

The chapter initially provides an overview of the progression of research in the four phases, which also correspond to the research questions. The chapter provided the basic building blocks for the sub-studies included in this dissertation. Definitions of key terms and metrics were provided as well as the data crawling process. Also detailed are the algorithms for building the networks.

Also included in the chapter is a critical evaluation of the methodology used in the dissertation, highlighting the ethical issues as well as the limitations of the approach.

# Chapter 4

# Empirical Analysis of Community Structure

*Chapter Synopsis - Detecting community structures in complex networks is a problem interesting to several domains. In healthcare, discovering communities may enhance the quality of web offerings for people with chronic diseases. Understanding the social dynamics and community attachments is key to predicting and influencing interaction and information flow to the right patients. The goal of the first half of the chapter is to empirically assess the extent to which we can infer useful community structures from implicit networks of peer interaction in IPC. Then the chapter analyses the temporal patterns of diabetes online communities. The last part of the chapter compares and contrasts findings with non-health social networks; with a view to delineate the development patterns that characterize IPC. The work presented in the chapter provides answers for Research Questions Q2 and Q3*

## 4.1  Introduction and Background

Empirical analyses are important for informing and validating existing general models. The process may yield unique scenarios that require new methods or approaches for solving problems. The next sections explore network analysis, and community detection, on empirical IPC datasets.

A major distinguishing feature between interaction in forums (message boards) and interaction in other social media is that healthcare forums normally do not have explicit relationships. Unlike most popular social networking websites (Traud et al., 2011), relationships in healthcare forums are mostly implied or inferred based on some criteria.

**Inferred or Implied Relationships**

These relationships are encoded in large datasets of forum threads-and-comments dynamics, and network analysis is a practical tool for deciphering these relationships. Abstracting user interactions using network analysis hold a potential to reveal characteristics that help us identify important peers (Kleinberg, 1999), as well as identify those users who may be in need of help. The possibilities are limitless; we could also be able to predict community attachments and analyse the flows of influence (Subbian et al., 2013) in these IPC.

**Discovering Communities**

Discovering the community structure is important for understanding how users interact with peers in different IPC. However, previous studies have consistently shown that assessing whether the discovered communities are meaningful or good is rather application-specific (Boutin and Hascoet, 2004; Schaeffer, 2007). Therefore, it is important to ascertain the usefulness of community structures in the context of IPC.

Many studies analyse several unrelated networks in pursuit of thoroughness and generalization. Although this comprehensive approach seems to work quite well for analysis of computing complexity(Leskovec et al., 2010), formal validation(Boutin and Hascoet, 2004) and other quantitative measures, it seems to suffer significant drawbacks when it comes to quality assessment of the discovered communities, which requires some level of homogeneity in the datasets. As a result, there have been many quite simplistic heuristic quality evaluations (Girvan M, 2002).

The sub-study aims for this chapter are three-fold:

(i) First, the nature of networks in IPC for diabetes is articulated, then
(ii) Explore new techniques for analysing development patterns, and
(iii) Delineate the development patterns that characterize IPC

## 4.2 Nature of Diabetes Social Networks

As discussed in previous Chapters, the main activities in a forum or IPC can be summarized as (i) a user creates a topic or thread, and (ii) several users, including the creator, can comment on the thread. Through the analysis of this topic creation and commenting cycles we are able to infer relationships within the IPC.

Although clean data is not always readily available for research (Rowe et al., 2007), there are several techniques that can be used to crawl publicly available data on the Internet. For this sub-study, the Python program crawled public data on five diabetes forums on the Internet. The forums had totals in excess of 140k registered users and over 1.6 million posts.

The dataset comprised Spanish and English forums, and two of them were dedicated to juvenile diabetes. As the data was extracted, usernames were changed to anonymized identification numbers (IDs) from the semi-structured HTML data. The extracted data was stored in a relational database, MySQL.

### 4.2.1 Reply-View (RV) Ratios

Analysing the data; the RV ratio is interesting in the sense that it shows how many views actually result in a reply to any thread in the IPC. This measure may be relevant in recommender systems, because a high RV ratio means the users take interest in the threads that they view. A very low RV ratio may be indicative of a Question and Answer type of social network, rather than a network with a real social experience with vibrant, non-superfluous, peer interactions.

In this work, however, it is important to note that replies are considered an *active expression of interest*. Consequently all the presented sub-studies are based on the assumption that relationships can only be formed if users actively express interest.

In an example from weight loss and diabetes social networks, plotted are the RV ratios of the sub-forums as illustrated in Fig. 4.1. A pattern is easy observe by mere visual inspection in the figure. The RV ratio seems to be highest in general discussions and lowest in research activities. These results suggest the type of topic or theme affects the activity or inactivity.
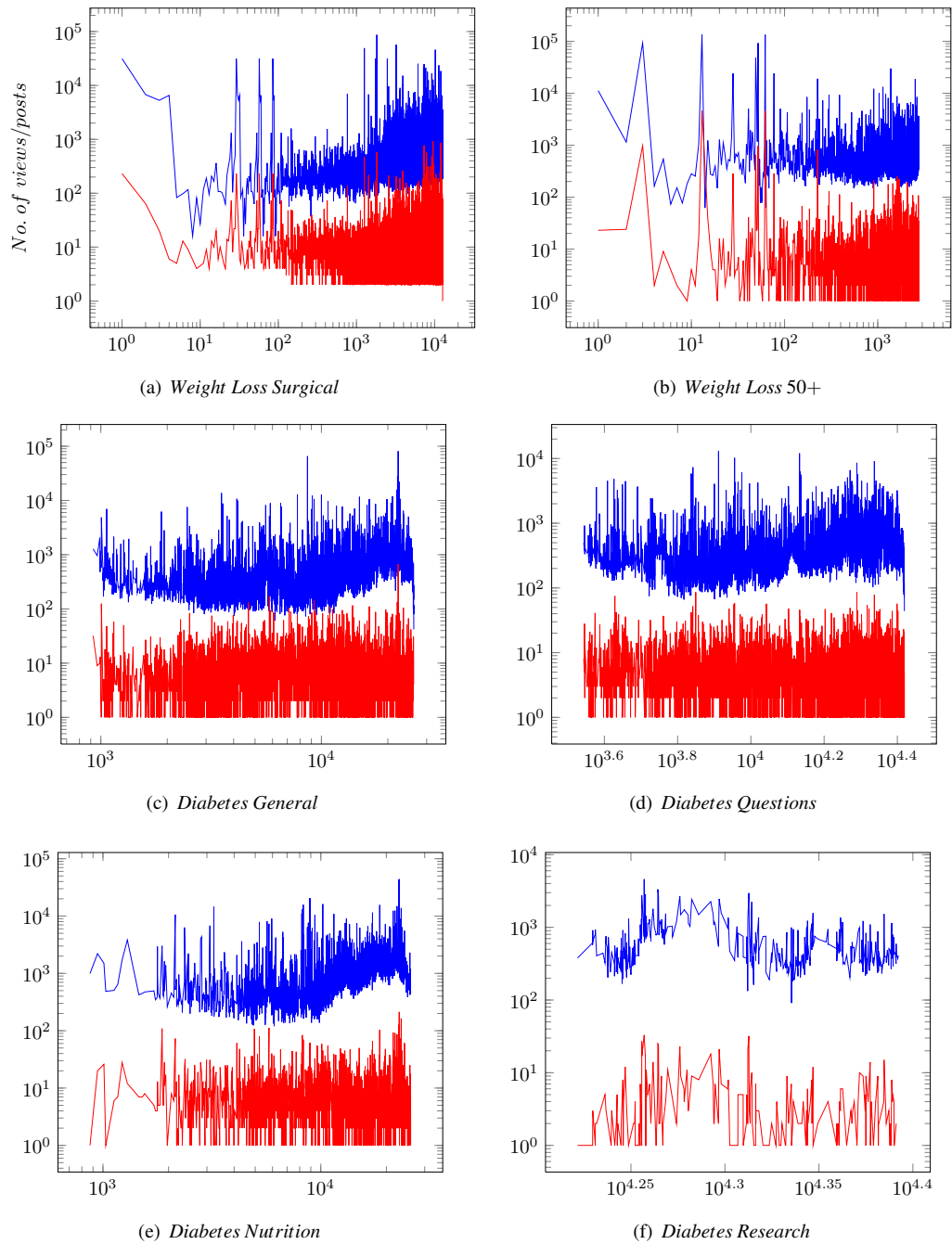
FIGURE 4.1: The difference between the number of views (blue) and the number of replies (red) on the posts over different periods. The average view/replies ratio for (a)=0.014, (b)=0.012, (c)=0.007, (d)=0.003 and (e)/(f) are less than 0.001

It seems apparent from the results that people with diabetes are more interested in practical topics such as insulin pump usage rather than in what may be regarded as secondary issues such as news and research.

(a)

(b)

(c)

(d)
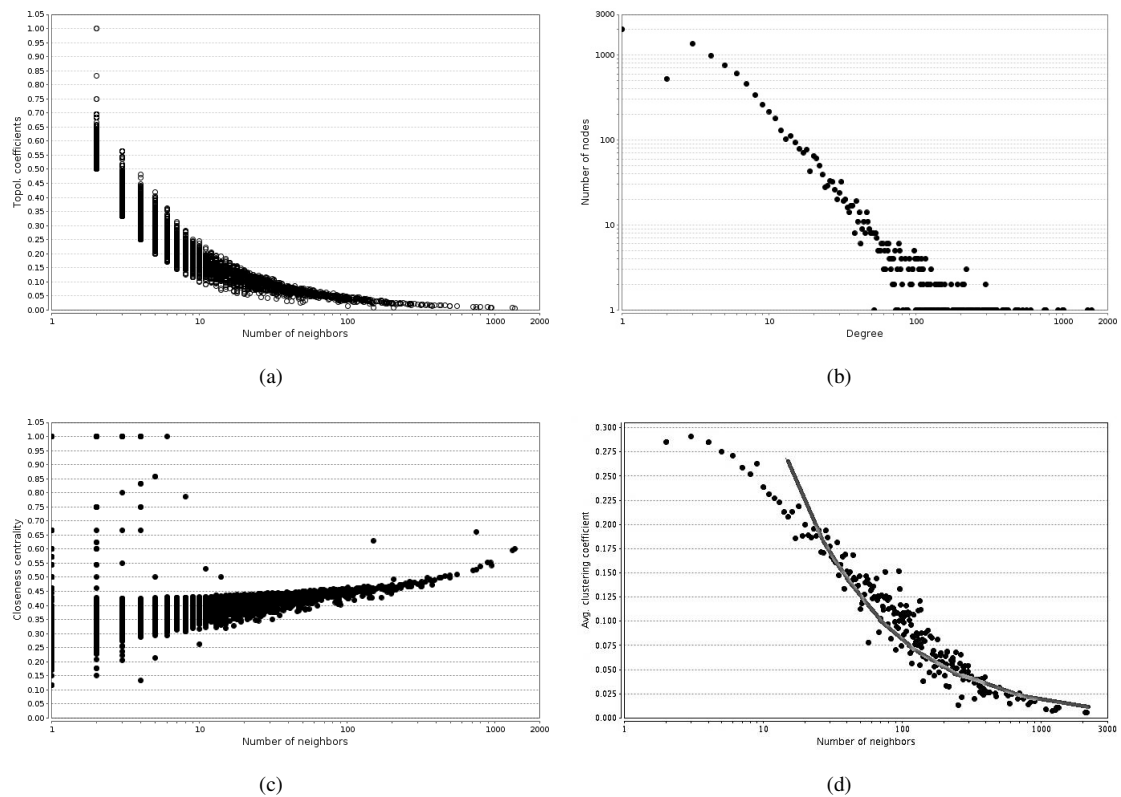
FIGURE 4.2: Network property plots of the diabetes social networks analysed in this sub-study.

## 4.2.2  Network Topology

Table 4.1 shows the summaries of basic network structure properties for all the datasets that were analysed in this sub-study. One interesting observation is the vast differences in the average number of neighbours, ranging from 10 to 62. Perhaps the latter can be considered an outlier since all but one has over 20 average number of neighbours.

However, this does not impact the Characteristic path length, which is rather short, ranging between 2.9 and 3.9, perhaps reflecting the centralised nature of the networks.

## 4.2.3  Scale-Free Tendencies of IPC

The images in Fig. 4.2 offers clues as to the scale-free nature of the diabetes social networks. This means only a few nodes have very high degree while the rest, and the majority have a very small degree distribution. In Fig. 4.2(d) I superimpose a scale-free line on the clustering coefficient/number of neighbours plot.

TABLE 4.1: Basic network characteristics from the five datasets and the community detection results. AP = Affinity Propagation, and GO = greedy Optimization

|  | $D_1$ | $D_2$ | $D_3$ | $D_4$ | $D_5$ |
|---|---|---|---|---|---|
| Number of nodes: | 9679 | 16404 | 5553 | 470 | 2948 |
| Number of edges: | 109695 | 467677 | 767413 | 42924 | 75027 |
| Users who posted (%): | 27 | 23 | 53 | 28 | 13 |
| Clustering coefficient: | 0.181 | 0.297 | 0.232 | 0.154 | 0.233 |
| Connected components: | 97 | 197 | 89 | 15 | 5 |
| Network diameter: | 11 | 9 | 9 | 8 | 13 |
| Network centralization: | x | x | x | x | x |
| Shortest paths (%): | 35 | 31 | 52 | 39 | 23 |
| Characteristic path length: | 3.6 | 3.3 | 2.9 | 3.2 | 3.9 |
| Average no. of neighbours: | 13 | 19 | 62 | 10 | 12 |
| Network heterogeneity: | x | x | x | x | x |
| Isolated nodes: | 93 | 192 | 84 | 13 | 2 |
| **AP (GO) algorithms** |  |  |  |  |  |
| Number of clusters: | 360(171) | 2600(242) | 1290(115) | 193(23) | 670(22) |
| Average size: | 27(57) | 6(68) | 4(48) | 2(20) | 4(134) |
| Maximum size: | 1976(3911) | 2163(8450) | 470(2597) | 239(131) | 1351(1372) |
| Modularity: | 0.07(0.34) | 0.14(0.35) | 0.02(0.20) | 0.40(0.28) | 0.35(0.35) |

## 4.3 Temporal Structure Patterns

For this sub-study, networks of user interactions are designed in two diabetes and two non-healthcare forums, and applied an existing community detection algorithm to time-partitioned datasets (snapshots) . To better understand the periodical changes, similarity analysis are used based on Jaccard similarity index, and cohesion analysis based on centrality measures and user attributes.

We use real-world data from large forums to explore the temporal nature of interaction networks with a view for enhancing our knowledge on the (i) temporal patterns of the communities, (ii) attributes that influence temporal community cohesion and (iii) salient patterns that characterize the networks. Empirical observations based on real-world data are important for validating and informing existing general models.

In this study community developments are not considered.

A recent study by Bródka et al. (2012), although not health-related, proposed a group evolution discovery (GED) method for analysing evolution of group structures or communities. The study provides a complete synthesis of temporal patterns of community structures to date. However, this method, as well as other approaches that presume group overlaps (Palla et al., 2005), seem to suffer a weakness when node sets from one period to the next have a consistently small number of elements in the intersection set, as is the case with some real-world networks. The algorithm consistently results in *formation* and *dissolution* patterns of evolution, and requires additional supporting information to make more real-world sense.

The proposal is to augment the framework by Bródka et al. with similarity measures to quantify the development patterns at both the network and community levels. In addition, we do cohesion analysis to reason about the development patterns in the context of diabetes, and show that we gain new perspectives of the evolution even when the networks are extremely volatile. We also attempt to discover the unique development patterns that distinguishes healthcare from closely similar general social networks.

### 4.3.1 Network Time-slices and Partitioning

Although there does not seem to be a unified framework, in most recent studies, temporal analysis have been based on partitioning of the networks into arbitrary time periods or snapshots . To start off, we present some of the major failing points for static analysis of healthcare social networks in Fig. 4.3, as the network grows from period $T_0$ to $T_2$.

(a) $T_0$        (b) $T_1$        (c) $T_2$
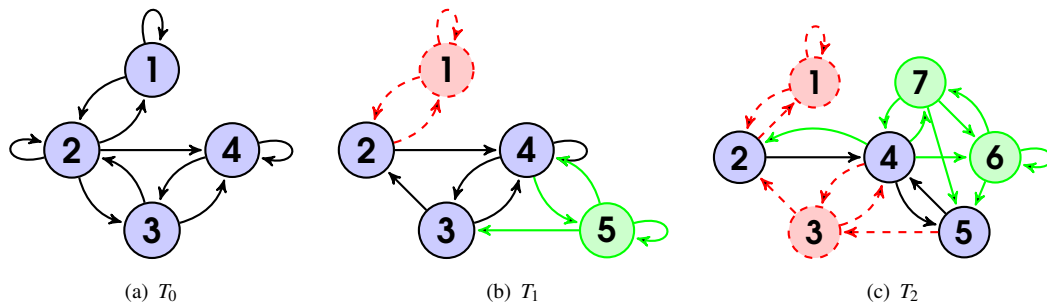
FIGURE 4.3: Evolution of social ties through time $t_0$ to arbitrary future time $t_2$. The nodes and edges with dotted red lines are dissolved social ties. The nodes and edges with solid green lines are the new social ties in the period.

In this instance, a static network is the absolute representation of the network from the beginning up to the cut-off period.

In the figure, static analysis of the network makes less and less sense as time progresses and the network changes because all nodes, including both new nodes and retired nodes, are treated as active. For example, Fig. 4.3(c) shows all the data from the beginning (with nodes 1,2,3 and 4) up to period $T_2$, and when looked at statically, without the distinguishing colouring, a lot of evolutionary details are obscured. For instance, it may seem as if, at time $T_2$, the network is invigorated into a dense network when in fact node 1 and 3 are retired nodes.

In this work, we partitioned the datasets for each forum into periodical sub-datasets, in order to be able to isolate activity in specific time periods. The networks can *continue*, *shrink*, *grow*, *split*, *merge*, *dissolve* or *form* completely new ones (Bródka et al., 2012; Palla et al., 2007). Although we used annual time-slices, we should highlight the problem of determining the optimal time-partitions or slices. The next three subsections describe our methodology for temporal analysis to some detail.

## 4.3.2 Experimental Approach

Online social forums and networks are emerging as platforms for healthcare interventions and convenient healthcare information access and support tools (Burton et al., 2012b). Present understanding of temporal development patterns and the factors that influence interaction in healthcare online communities seem quite limited. In the presented work, we explore the development patterns of diabetes online communities and seek to understand the factors that characterize and influence community development in this domain. Network analysis (and community detection (Fortunato, 2009)) is one of

the most practical ways of facing the challenges of mining (Wegener et al., 2013) the growing data for meaningful information.

An overview of the methodology is illustrated in Fig. 4.4, where networks are designed from user interaction data crawled from two diabetes forums. It is important to note that relationships in forum-like communities can be difficult to ascertain since there are no explicit relationships. An alternative is to form implied relationships from how the users interact with each other; forming bonds and ties from exchange of objects and through social discourse. To analyse the community structures we applied an established community detection algorithm on the networks; the greedy optimization (GO) algorithm (Clauset et al., 2004). In addition, we formulated similarity and cohesion analysis using a blend of Java machine learning libraries and network visualization tools. We also explore some properties such as the temporal density and diameter to distinguish healthcare development patterns from other general social networks.

### 4.3.3 Network Similarity

Next, we compare the communities for the different years. In this context, communities are coherent sub-networks in the time-sliced network, that is, clusters of nodes with dense connections. We used the Jaccard Similarity index to compare the networks and communities. Whereas the index has been used to compare the two datasets as a form of external validation, in this work we explore its use for analysing two datasets from two periods, where the community $C_a$ can have $n$ nodes $\{x_1, \ldots, x_n\}$ at time $T_0$ and the community $C_b$ is the similar community at time $T_1$ with $m$ nodes $\{x_i, \ldots, y_m\}$, where $x_i$ can be a subset of $C_a$. The aim of the analysis is two-fold:

- first, for quantifying the similarity at the network level (see Fig. 4.5), declaring the first (or preceding) network as the benchmark:

  **Definition 4.1 — Jaccard Similarity Coefficient.** Jaccard Similarity Coefficient defined as

  $$J(C_a, C_b) = \frac{|C_a \cap C_b|}{|C_a \cup C_b|} \tag{4.1}$$

  Where $C_a$ is the benchmark network at time $T_0$, and $C_b$ is the network at an arbitrary future time $T_n$.

- second, for quantifying the similarity for the communities in each time period, where each network connection is annotated with a community identifier; for
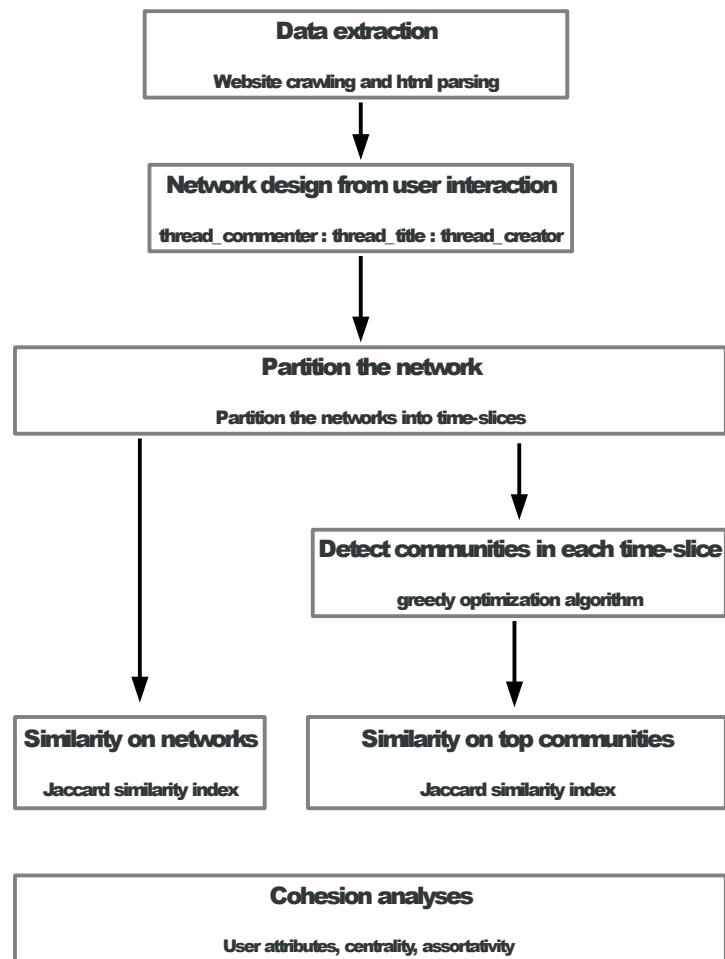
FIGURE 4.4: The methodology for the study, summarizing the flow of the steps for each diabetes forum.

identification in the subsequent period. Communities in each period are compared with communities in the subsequent period. In this case, each of the top three communities is compared to each of the top three in the next period (see Fig. 4.5). These Comparisons allow us to gauge the stability of the communities from one period to the next period or periods, providing both a course and detailed overview of how the communities are forming or dissolving over time.

### 4.3.4 Community cohesion heuristics

Finally, we analysed community cohesion to understand the bonding factors. Several types of attributes were available for this analysis; (i) years-since-diagnosis, (ii) type-of-diabetes, (iii) $HbA_{1c}$, (iv) Age and (v) gender.
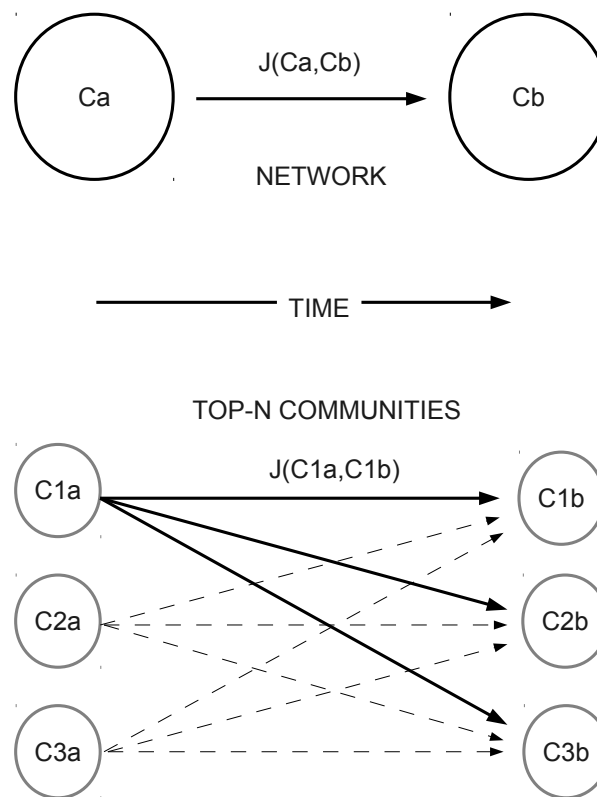
FIGURE 4.5: The top large circles resembles the network and the bottom circles represent the top communities in one period to the next, as they are compared using the Jaccard Index; illustrated by both the solid and dotted lines, source (Chomutare et al., 2013c)

We further looked at (i) degree assortativity, (ii) diameter (iii) density and (iv) average degree.

In the succeeding subsections we explore some of the attributes that are highly relevant for analysing cohesion in diabetes networks. There are potentially quite many attributes that we could discuss, but we highlight just a few based on expert opinion and what the forum users could have provided in their public profiles. In the process, we also explore the discussion and debate around use of personal health information, and it's availability for information processing.

**Years-since-diagnosis**

*Years − since − diagnosis* was an obvious cohesion factor because almost 80% of the registered users have been diagnosed less than two years ago in any of the periods. This is indicative of how online communities have become the preferred source of support for newly diagnosed patients. While some newly diagnosed patients also supported other new patients, the majority only acted as information hubs. The authoritative role
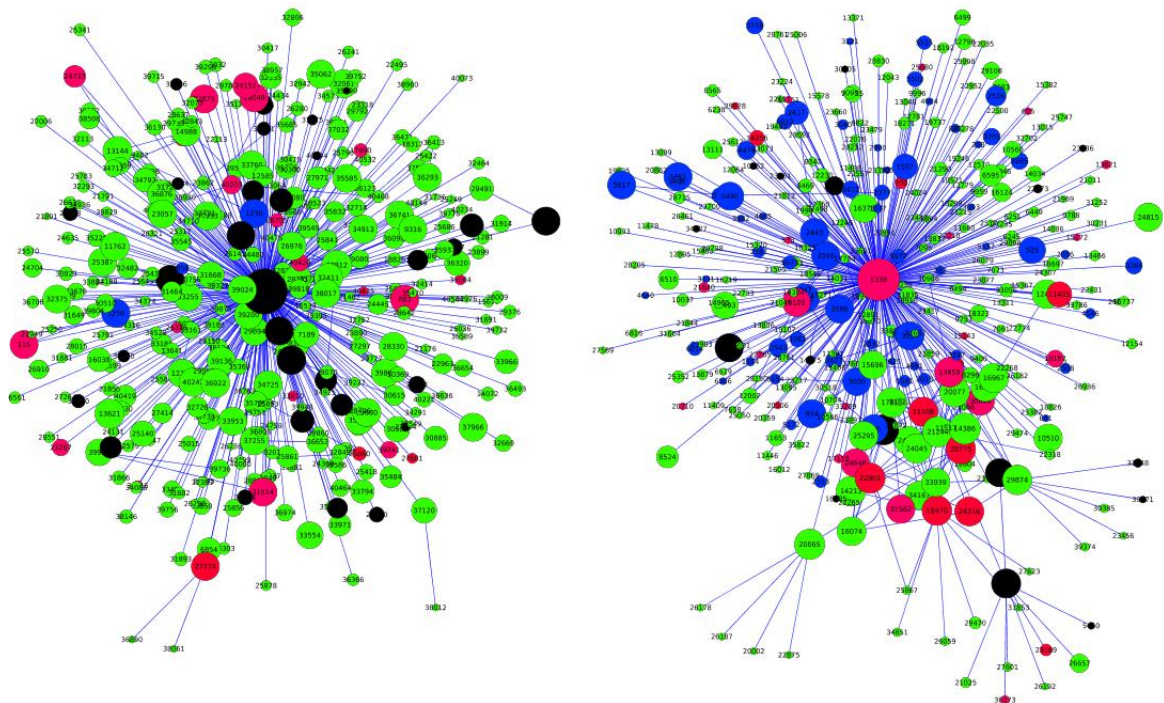
FIGURE 4.6: A zoomed-in figure of some of forum F1 detected communities based on the Greedy Optimization algorithm. The node size is related to the node's in-degree, and the colours are: blue = no data provided by the user, green = 0-1 year after diagnosis, red = 2 -10 years after diagnosis, and black = more 10 years after diagnosis, adapted from (Chomutare et al., 2013b). NOTE: the higher resolution figures can be obtained on
http://www.diabetesbuddy.org

(Kleinberg, 1999) was assumed by patients with 2-10 years' experience after diagnosis as can be seen in Fig. 4.6, a zoomed-in figure representative of the communities. From the figure, we can observe that a huge majority of the nodes are the newly diagnosed users (green), which are connected to central figures who have more experience with diabetes (red,black).

**Type-of-diabetes**

*Type-of-diabetes* is an intuitive attribute for cohesion because the main types of diabetes (Type 1 and Type 2) have several lifestyle and behavioural goals in common; mostly blood glucose management, dietary and physical activity goals. Only about 5%-10% of patients with diabetes have type 1 diabetes, and this likely obfuscate some community patterns unique to type 1 diabetes.

**HbA$_{1c}$**

*HbA$_{1c}$* is a measure (in percentage) of long term blood glucose levels and is an important outcome for people with diabetes. It is not certain why forum users did not disclose

their HbA$_{1c}$. Only 3% and 5% of the users of *Diabetes$_1$* and *Diabetes$_2$*, respectively, disclosed their values. The majority of the disclosed $HbA_{1c}$ values where between 6 and 9, which is considered an OK range for people with diabetes (values around 7 are more desirable). This suggests people that manage their disease well are more likely to disclose their $HbA_{1c}$.

**Age and Gender**

Age was disclosed by only 5% and 10% in *Diabetes$_1$* and *Diabetes$_2$*, respectively. Therefore, it is difficult to explore its impact. The fact that users do not want to provide their age may be suggestive of its irrelevance. On the other hand, just over half disclosed their gender.

## 4.4 Characterizing Unique Patterns in IPC for Diabetes

Further analysis of the network properties will show that there are some differences between healthcare and non-healthcare networks. Comparison in the preceding Table 4.1 reveal an interesting contrasts among the network and community structures. Visualizing the structural differences, as illustrated in Fig. 4.7, we observe the decentralised Facebook network and a very centralised community structure for the diabetes network, while the weight loss community is much less centralized.

This may be an important observation for two reasons; (i) the first is that we do not know which structure offers the best health outcomes, and (ii) we do not know what influences the structures. However, what we can infer is that decentralised networks may be a sign of persistent and stronger social bonds. This inference is based on the fact the most Facebook wall-post networks are decentralized, and most friends in Facebook have met at some point, and have more intimate awareness of each other than in general healthcare online communities. Although much of this reasoning may be considered speculative at this point, these findings warrant further investigation.

### 4.4.1 Assortativity and other Network Attributes

In this section we form generalizations about the salient characteristics that distinguish healthcare datasets from other social networks. We base the analysis on the trends of (i) assortativity, (ii) network diameter, (iii) network density and (iv) average degree, over the studied period.

(a) This is a zoom-in for a Facebook network, showing a very decentralized community structure

(b) Cluster zoom-in for an obesity network, and shows the semi-centralized nature of the communities

(c) A zoom-in for a diabetes network, showing a very centralized community structure.
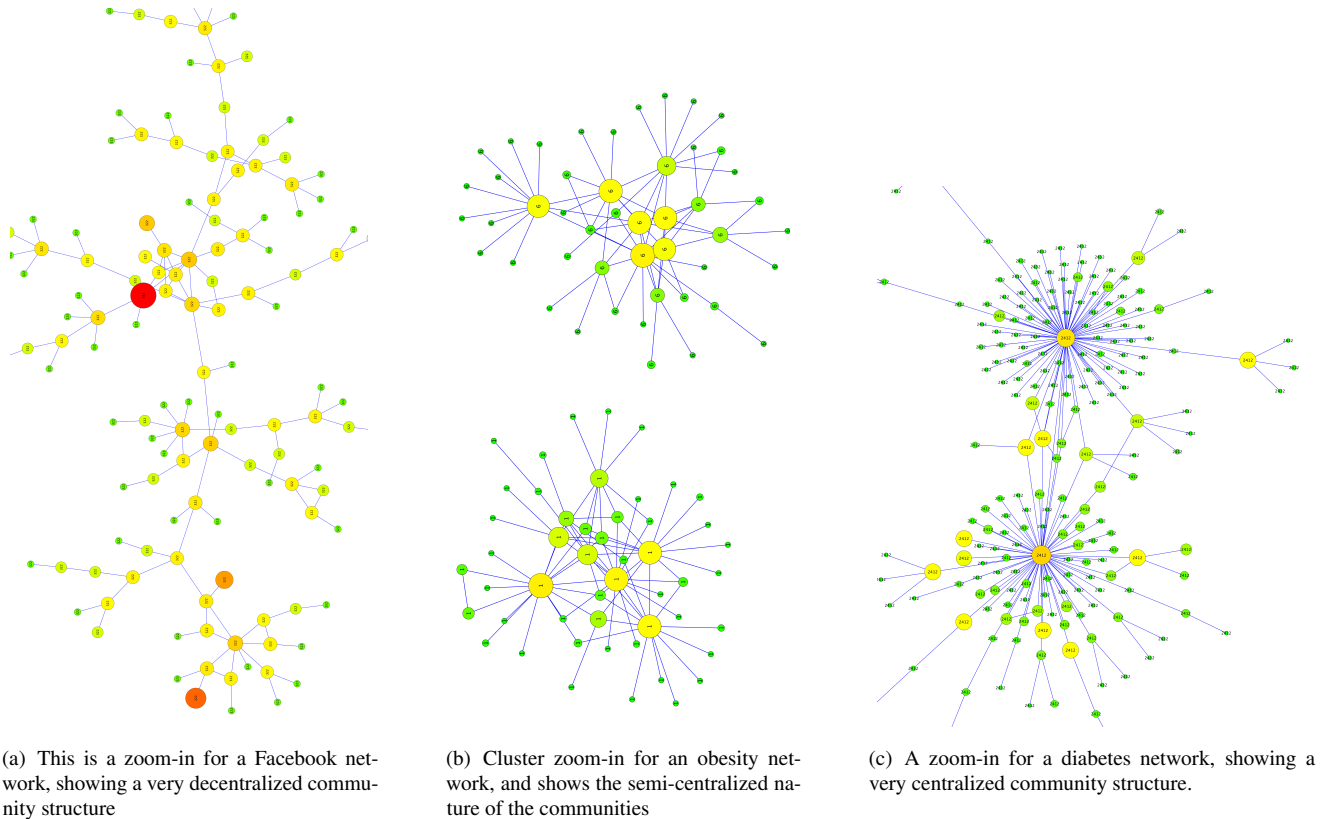
FIGURE 4.7: Some of the visualizations of the communities found in the networks using the GO community detection algorithm at different zoom levels. NOTE: the higher resolution figures can be obtained on http://www.diabetesbuddy.org

Assortativity as described by Newman (2002a) (also called Homophily in the literature (McPherson et al., 2001)) is the tendency for similar or dissimilar nodes to connect to each other. Degree assortativity describes the extent to which nodes of similar degree cluster together. For example, people with many connections in popular social networks tend to connect to other people with many connections. In this study we used a Java machine learning library that implements the assortativity formula, which is merely the Pearson correlation coefficient, see Definition 4.2 and Equation 21 in (Newman, 2002a).

**Definition 4.2 — Assortativity.** Assortativity based on Pearson Correlation Coefficient defined as

$$r = \frac{\sum_{xy} xy(e_{xy} - a_x b_y)}{\sigma_a \sigma_b} \tag{4.2}$$

where $a_x = \sum_y e_{xy}$ and $b_y = \sum_x e_{xy}$, and $e_{xy}$ represents the fraction of edges between the vertices $x$ and $y$, and $\sigma_a$ and $\sigma_b$ are standard deviations.

Finally, we studied diameter and density, which are key to understanding networks because they describe the inter-connectedness of the nodes, and they can be a plausible basis for distinguishing network characteristics. We also focused on understanding the changes in the average degree over time.

We can observe from Fig. 4.8 that there is significantly higher assortative mixing in the Facebook network, while there is mostly disassortative mixing in the healthcare datasets. This result is not surprising because Facebook networks are more decentralised, and it is easy for users with high degree to be connected to other users with high degree, as can be seen in Fig. 4.7(a).

We further observe from Fig. 4.7(c) that healthcare network structure, as sampled here, has a far more centralized star topology. This means users with very high degree connect to several users with very low degree, hence the mostly negative assortative mixing. Perhaps this reflects the very core of diabetes forums, where a few experienced and knowledgeable users tend to support a large number of newly diagnosed users as seen in Fig. 4.6. Information dissemination becomes vital as it is placed in the hands of a few central nodes that have a very short path to a large number of nodes. These findings can be contrasted to (Newman, 2002a) results that indicate most social networks exhibit assortative traits.

**Diameter and Density**

It follows from the network structure argument in the preceding paragraph, that the diameter for healthcare datasets is much lower than non-healthcare datasets. In terms of the temporal patterns, it seems from Fig. 4.8 that the network diameter falls with time in healthcare datasets while it actually increases in the other datasets. The density of the non-healthcare datasets was extremely low.

On the other hand, the density in the healthcare datasets, while also very low but higher than in non-healthcare, exhibited a diminishing trend over time, from 0.010 in period one for $Diabetes_1$ to 0.002 in the last period. Although some recent studies like (Leskovec et al., 2005) have shown that density increases and diameter shrinks over time for most networks, our results suggest both density and diameter shrink in healthcare forum data. This finding may be partially explained by the tendency to attach to the central and more experienced node rather than for novices to interconnect among themselves, resulting in shrinking of both density and diameter.

**Average Degree**
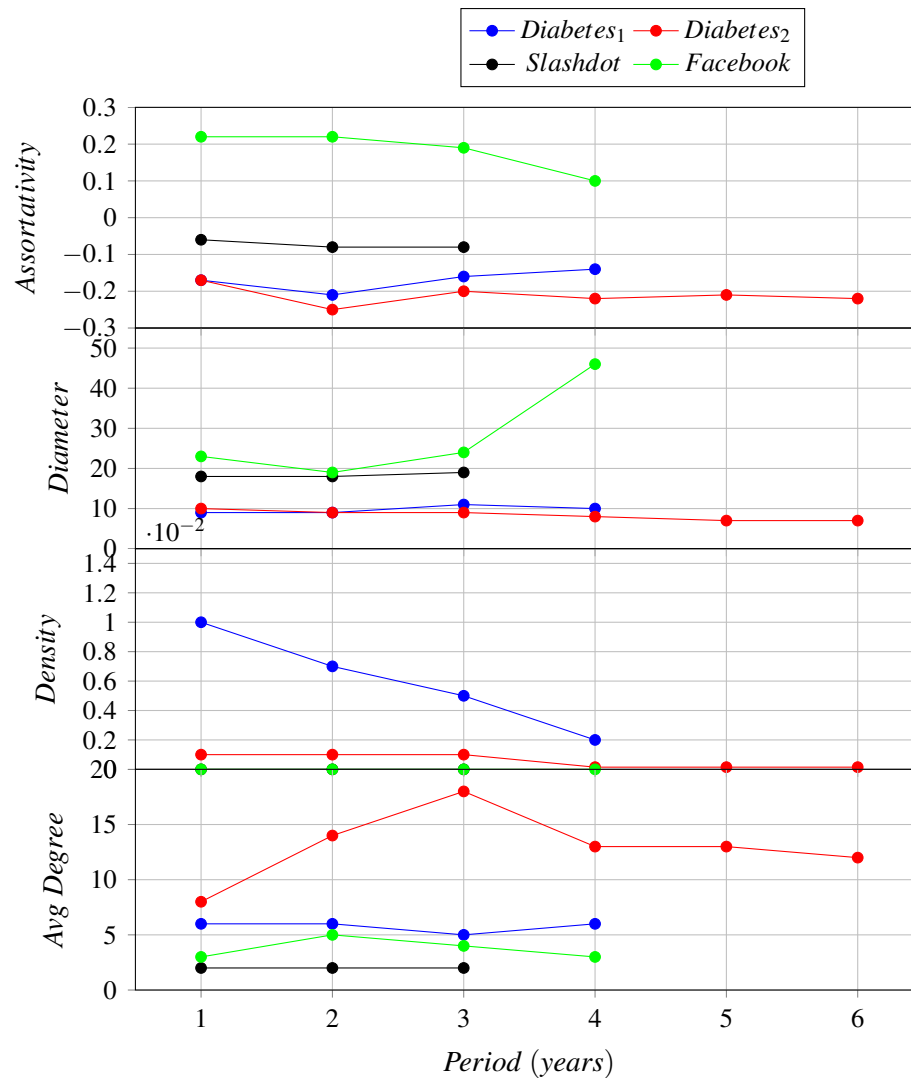
FIGURE 4.8: Comparison of the temporal networks in terms of the average degree, network diameter and degree assortativity. It is interesting to note that the diabetes networks are always on the same side of the spectrum.

The average degree is higher in diabetes networks than the non-healthcare networks, and perhaps this reflects the information needs of many newly diagnosed users as they try to get to grips with diabetes, while they communicate with only a few experienced users.

## 4.4.2 Limitations

Perhaps this sub-study might have benefited more if several other IPC other than just diabetes were included, as well as other diverse non-health social networks. This might have enhanced the basis for the conclusions. However, one aim of the sub-study was to highlight the differences with the popular media such as Facebook, and closely similar support forums such as Slashdot.

### 4.4.3 Knowledge Summary

The knowledge added is here summarized in point form in relation to the original research questions.

*What is already known on the topic*

- There is a lot of activity in IPC for diabetes, with complex interactions and relationships.
- Relationships in IPC are not explicit and people meet peers solely for the purpose of managing their illnesses.

*What this chapter added to our knowledge*

- Network analysis is a plausible tool for abstracting complex social interaction in IPC – resulting in networks that exhibit scale-free tendencies like most non-health social networks and other real-world, non-random networks – helping answer Research Question **Q2**.
- A method for multi-level analysis of time-sliced partitions helped understand the development of these IPC, revealing the dynamic nature of the communities and superficial relationships – helping partially answer Research Question **Q3**.
- Clear structural differences were revealed between diabetes and obesity, and non-health social networks, eg. the degree disassortative mixing, shrinking of both diameter and density. Further attributes such as *Year-Since-Diagnosis* were identified as important factors for community cohesion – helping partially answer Research Question **Q3**.

## 4.5 Chapter Summary

This chapter has shown how implicit networks can be modelled from forum interaction. Further, showing that the resulting networks actually do make sense and correlate with some attribute-based homophily measures. Current empirical observations strengthen the basic science that supports future analysis and reasoning about diabetes patients as users of the medical web. Further studies are required before the nature of these healthcare online networks is more clearly understood; to foster adoption of healthy lifestyles among people with diabetes and other chronic illnesses.

It was shown how existing methods may fail to meaningfully describe extreme development patterns where communities constantly dissolve and form. Extending existing

temporal models with quantifiable similarity measures and reasoning about community cohesion seemed to reveal potentially hidden details in the real-world networks.

More important, the empirical findings in this study provide a new understanding of social engagement in diabetes social networks. The most surprising finding to emerge from this study is that diabetes communities are very dynamic and short-lived, implying users engage only for short periods, and do not sustain any noteworthy networks or communities. Perhaps the lack of will to invest themselves in online communities is reflected in their reluctance to disclose personal data. Finally, we observed the shrinking diameter and density, and the disassortative mixing in the diabetes networks.

Current work informs future intervention strategies for promoting health behaviour and lifestyle changes among people with diabetes, but further research is required before much of the implications of the discovered patterns and cohesion trends are more clearly understood.

# Chapter 6

# Internet Patient Community – Diabetes Pilot Study

***Chapter Synopsis -** Although mobile applications and social media have emerged as important facets of the Internet, their role in healthcare is still not well-understood. Design artefacts are presented, inspired by persuasive technology concepts, from a study of IPC as part of a diabetes mHealth application – and the work seeks to partially answer the first Research Question **Q1** – by demonstrating relevance of eHealth, mHealth and patient interactions in real-life diabetes self-management scenarios. The design science approach was used for mobile application design, and user testing and focus group meetings to test the application over a 12-week period with 7 participants, with additional followups at 6 months and 12 months.*

## 6.1 Social Mobile Apps in Group Education Programs

Recently, there has been an exponential growth of interest in diabetes mobile application as documented in a review by Holtz and Lauckner (2012), as well as healthcare social media (Chomutare et al., 2011) - for ***Self-Management of Blood Glucose (SMBG)***.

Although there is increasing concern over the use of personal health information and trustworthiness of user-generated health content, far less attention has been paid to the impact these emergent tools have on health outcomes (Baron et al., 2012). In this sub-study, a mobile application is designed with social features, and the application tested with diabetes patients over a 12-week group education course.

### 6.1.1 Individual Versus Group Education and Counselling

Current evidence is clear about the superiority of group education programs (Adlerberth et al., 1992; Steinsbekk et al., 2012) over usual care. However, the literature has consistently shown that these benefits are often in the short term only (Sperl-Hillen et al., 2013) while in other longer-term studies, the benefits have been shown to rescind (Sarkadi and Rosenqvist, 2001).

There seems to be conflicting evidence on the long term benefits of group education for diabetes patients. Many studies have found no significant benefits (Rosenbek et al., 2011; Rygg et al., 2012) while a review by (Deakin et al., 2005) reported some benefits. Mobile devices and social media may have a role to play in exploiting the latent benefits of structured group learning.

More recently however, studies have emerged that report individual education as more efficacious than group education programs (Sperl-Hillen et al., 2011, 2013). Despite these new findings, group programs seem to continue to be attractive, perhaps primarily for their affordability and social aspects.

### 6.1.2 New Roles of Smartphones and Social Features

A significant shortcoming in most of the previous studies is that no patient support tools are offered once participants complete the group education programs. Smartphones are now pervasive and hold a potential for enhancing group education programs, and help patients maintain model behaviour into the long term.

Although education can generally be thought of as key to empowering the patient to adopt a good lifestyle (Davis et al., 2007), proper context must be given to its delivery. While group education programs provide the basis for life-long learning, further effort is required to persuade the patient to sustain model behaviour, and smartphone applications are suited for this role.

Social networking holds a potential as a motivational agent for actioning the knowledge people acquire while self-managing diabetes. Studies have shown that knowledge about a disease does not necessarily translate into action (Curtis et al., 1993); patients normally require further nudges and reminders (Mulvaney et al., 2012).

Unlike structured education, informal social interaction provides random learning material from real users candidly. A concept that has not been fully explored in the literature

is that social media could also be a social platform for fostering better health behaviour based on social comparison theory (Mueller et al., 2010), for example, people comply with certain behaviours as a way of seeking validation or 'fitting-in'.

### 6.1.3 Mobile Application Design Artefacts

The mobile application comprises the FewTouch application (FTA), a well-studied mobile application (Årsand, 2009; Årsand et al., 2012, 2010), and a semi-integrated social forum. Fig.6.1(a) shows the blood glucose measuring kit with a Bluetooth module for data communication with the mobile application. Blood glucose measurements from the glucose meter are sent by wireless transfer to the mobile phone without the need for user input.

The data in the forum is totally decoupled from the FTA data to enhance security by allowing users to determine the data they want to share with peers on the Internet. The forum represents the public interface that communicates with other Web forums, independent of mobile platform, using Internet Protocols and browser. It is important to note that the forum connects to a social engine in the cloud (at `http://www.diabetesbuddy.org`).

**Persuasive Technology**

Our design of the mobile application is not based on any persuasive technology philosophy such as the Persuasive system Design model (PSD) (Oinas-Kukkonen and Harjumaa, 2009), but is based on experience from previous user studies on similar patient groups. Modern mobile platforms already provide native support for core persuasive design concepts such as:

- *Convenience;* easy to carry everywhere
- *Attractiveness;* most mobile platforms natively support attractive user interfaces and appealing device hardware
- *Trustworthiness;* use of sensors minimizes user error and tampering
- *Simplicity;* mobile devices have now simple touch-screens and simple device hardware

Some more effort is required to exploit these persuasive qualities, by adding complementary persuasive features in applications such as:

- *Tailoring, suggestion;* through the recommender system
- *Self-monitoring, surveillance;* users self- monitored blood glucose levels, and researchers review these measurements

(a) BG Measuring Kit
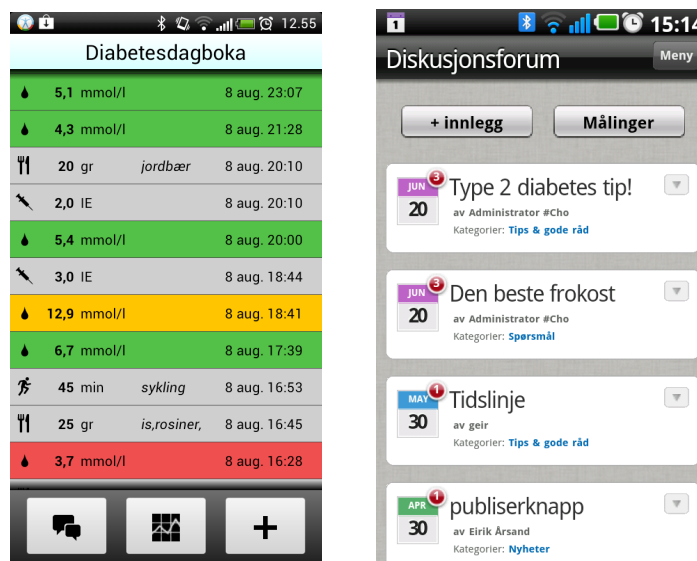


(b) BG tracking



(c) Forum posts

FIGURE 6.1: Overview of the mobile application architecture. Fig.6.1(a) shows the blood glucose measuring kit that has an attached Bluetooth module to send readings to the mobile phone as discussed in (Årsand et al., 2010). Fig. 6.1(b) and 6.1(c) shows the Android platform screenshots for FTA's blood glucose tracking and the personalized social media posts updates.

- *Social facilitation, social comparison, normative influence, competition, cooperation;* all done through sharing information on the web forum with peers in IPC.

## 6.2 Pilot Study with Diabetes Patients

The pilot with diabetes patients was part of a group education and motivation program (*Motivasjonsgrupper* ) run by the Diabetes Association (*Diabetesforbundet* [1] ). The groups are run independently by area leaders who are motivated to do so. In this project, we engaged an area leader to help with the recruitment.

The group ran for an initial 12-week period between September and November 2012, where participants met once a week for some education, group activities, as well as mixing and mingling with peers. I introduced the mobile application during this meeting period. Participants also filled out surveys at baseline and at each follow-up session.

Participants continued to use the mobile phone at the 6-month and 12-month follow-up. Although the pilot study was not properly powered to assess the significant of the health outcomes, the results were promising, serving as preliminary evidence of the relevance of patient interactions, both on- and off-line.

**Survey Administration**

All the people who filled in the questionnaires had used the diabetes mobile application for at least 12 weeks. We used the System Usability Score (SUS) (Bangor et al., 2008) to assess the usability at the end of the study. The Health Education Impact Questionnaire (HeiQ) and the Diabetes Empowerment Scale-Short Form (DES-SF) were used to assess self-efficacy. Some of the information we wanted to gather was very specific and not part of any recently validated questionnaire. We brainstormed a set of questions for gauge at the familiarity with and usage of social media among the participants.

**Extending the FewTouch Mobile Application**

The social feature extension of FewTouch is based on a simple WordPress application, installed and configured on the project website `http://www.diabetesbuddy.org`. Main FewTouch features include:

- Blood glucose monitoring
- Physical activity recording
- Diet and food habit recording
- Setting goals

The social extension included:

---

[1]`http://www.diabetes.no`

- Polls and voting
- Forum and discussion platform

The full application allowed participants to self-monitor their daily blood glucose levels, and also be able to discuss their experiences with peers. This meant that participants could vote on issues relating to everyday experiences, as well as post their statuses such as having healthy food or when exercising. They could also share tips on simple things such as recipes and exercise routines based on home chores, just as they would in any open IPC.

A shortcut to the application was put on the home screen for easy access, and the Bluetooth data communication with the blood glucose system worked in the background. In the main screen of the application (Fig.6.1(b)), the users could track their blood glucose values and be able to switch to the social forum (Fig.6.1(c)) or other features such as the blood glucose graphical plot. Users also received personalized tips and information using email alerts, which work like short messaging service (SMS) implemented in several studies, except we used Internet Protocols.

**Minimal Input**

In social media, user input was minimized by use of polls and ratings, so that the users only needed to press an option or a button to contribute. This circumvented the traditional pitfalls of awkward input modalities on small mobile phone screens. Overall, it may be difficult to understand why a fairly elderly group thought the application had high usability, but perhaps this could be partially explained by the user-centered design, coupled with the fact that people in general have become used to smart mobile phones.

**Feature Usage**

It is interesting to note that most of the features of the mobile application were not used, for example, carbohydrate counting and physical activity recording. Perhaps this is due to the nature of the features that require the user to input data manually.

The findings from feature usage is really not surprising. It seems patients are only interested in a few features on the mobile application. While large suites of features may certainly be useful if they were used, perhaps providing just a few, simple features is a better approach for avoiding cognitive overload.

**Patient Interactions On- and Off-Line**

It seemed that the participants preferred physical contact or off-line interactions to online contact. This is not a surprise, and indeed this would be ideal if it were possible in all circumstances. Online interaction can be a next best alternative, and the participants used online interaction to share recipes for low carbohydrate meals with fish or bread. They also used it to coordinate meetings, and wish each other well during holidays.

### 6.2.1 BG, HbA$_{1c}$ and other Clinical Outcomes

The results in Table 6.2 show that there was a marginal decrease in the average follow-up HbA$_{1c}$ to 6.79%, $\sigma = 0.68\%$. The decrease may seem insignificant but considering the baseline average HbA$_1 c$ was below seven and many studies have observed improvements only in people with high values (Cooper et al., 2008; Rygg et al., 2012), these results are noteworthy.

TABLE 6.1: Blood glucose Kendall's tau statistic ($\tau$) compared with the respective baseline and follow-up HbA$_{1c}$ of the participants. Mann-Kendall tests are simple and robust non-parametric tests to help reason about the changes in HbA$_{1c}$ based on the trend of the BG values over the study period.

| User | Gender | 3 Month Follow-up | | 12 Month Follow-up | |
|---|---|---|---|---|---|
| | | tau($\tau$) | $p$ | tau($\tau$) | $p$ |
| 1 | female | -0.156 | 5.55E-05 | 0.119 | 6.0263E-09 |
| 2 | female | -0.058 | 0.357 | -0.0728 | 0.068842 |
| 3 | female | -0.031 | 0.676 | 0.0143 | 0.81718 |
| 4 | male | -0.077 | 0.030 | -0.0314 | 0.071134 |
| 5 | female | -0.027 | 0.572 | -0.0835 | 0.0016069 |
| 6 | male | -0.037 | 0.342 | 0.13 | 1.7448E-08 |
| 7 | male | -0.146 | 0.001 | -0.0155 | 0.50029 |

To further reflect on the blood glucose (BG) values that resulted in the lowered HbA$_1 c$, we compute linear regression on the daily blood glucose levels as shown in Table 6.1. It is easy to observe that users with a steeper negative gradient resulted in a reduction in their HbA$_1 c$. Fig 6.2 shows the scatter plots for the participants over the 12 month period. For many of the participants it is easy to observe how the BG levels tended within the therapeutic range (the green zone) during the pilot.
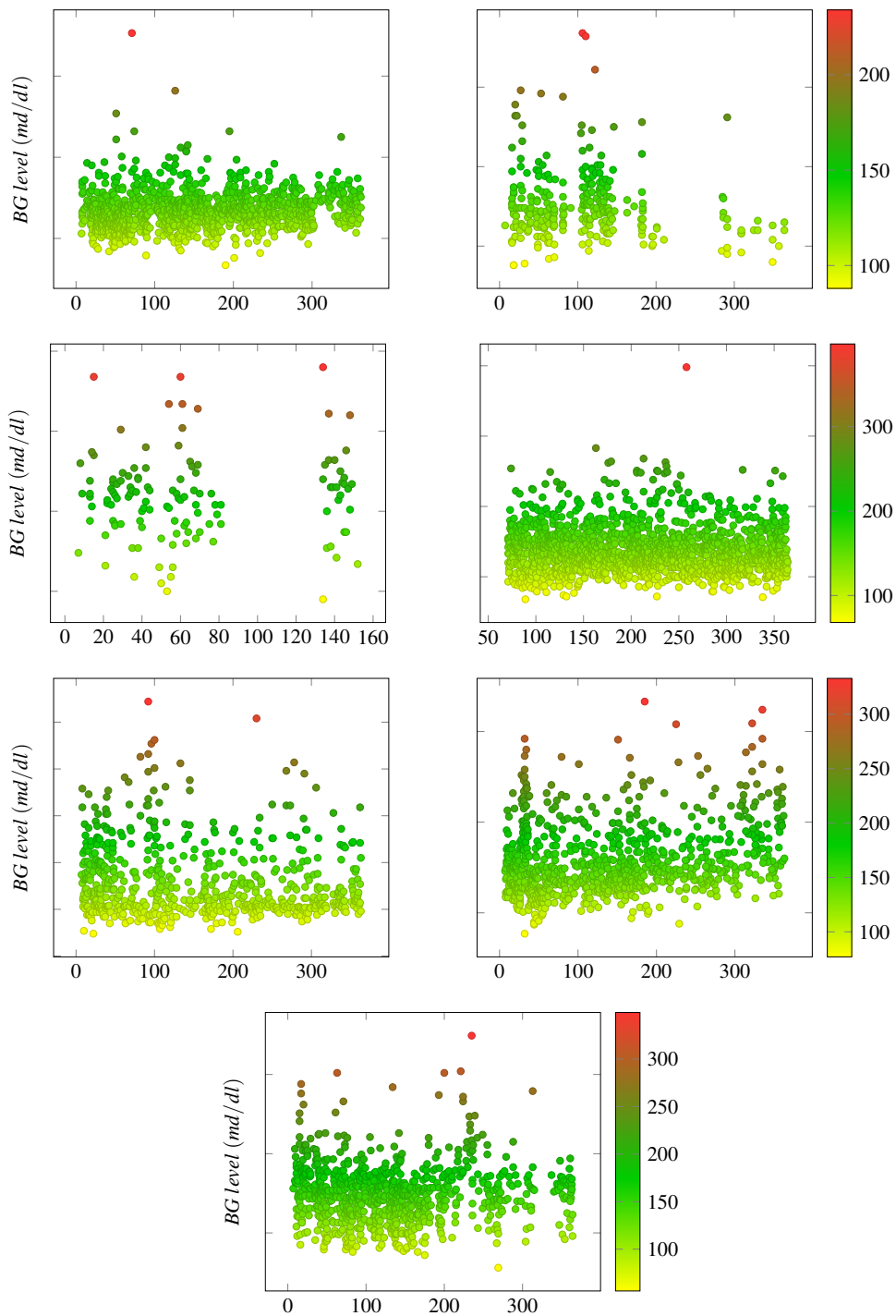
FIGURE 6.2: The BG levels for the study period, where time (in *days*) is plotted on the *x − axis* (Sept 2012 to Sept 2013). The graphs provides an overview of the intensity of the BG measurements.

However, there are some cases where there is a small negative gradient that results in an increase in the HbA$_1$c, eg. user 3 and user 6, who are also the only participants whose HbA$_1$c increased. Further analysis of blood glucose variability and the time of measurement could be done to more clearly understand these outcomes.

There was an increase of self-efficacy based on the HeiQ and DES-SF, respectively (see Table 6.1). Self-efficacy is an important outcome that can facilitate change in health behaviour in self-management (Bandura, 1997; Sturt et al., 2010).

### 6.2.2 Self-Efficacy and other Lifestyle Indicators

Results show that social interaction by regular meetings and online can influence self-management outcomes positively. First, we report some of the characteristics of the focus group participants as given in a short demographic questionnaire. The 7 participants were 4 females and 3 males between the ages of 46 and 70, with average age of 62.71 ($\sigma$=8.99) and 4 of them had been diagnosed with diabetes more than ten years ago. Only 4 of them reported using the Internet for finding health information and for social media activities.

On the question of sharing health data, 2 of them said they would share their health data with neither their families, friends nor other people with diabetes. Something interesting to note is that 3 of the 7 participants said they did not use the Internet to search for healthcare information.

TABLE 6.2: The results of $HbA_{1c}$ (in %) and the different self-efficacy and usability surveys.

| Description | $HbA_{1c}$ | SUS | HeiQ[2] | DES-SF |
|---|---|---|---|---|
| Baseline | 6.97,$\sigma$=0.69 | - | 2.96,$\sigma$=0.27 | 4.14,$\sigma$=0.62 |
| Follow-up (3month) | 6.79,$\sigma$=0.68 | 84.6,$\sigma$=13.2 | 3.10,$\sigma$=0.23 | 4.57, $\sigma$=0.21 |
| Follow-up (6month) | 6.6,$\sigma$=0.48 | - | - | - |
| Follow-up (12month) | 6.78,$\sigma$=0.54 | - | 3.06,$\sigma$=0.31 | 4.25, $\sigma$=0.65 |

1.Detailed domain analysis using the HeiQ survey might reveal more information

The primary endpoints for this study are shown in Table 6.2, where all the endpoints show an improvement between the baseline and follow-up. The overall picture is that there are improvements between the baseline and the last follow-up at 12-months, although the most gains were realised at the 3-month follow-up period.

Fig 6.3 illustrates the changes in $HbA_{1c}$ from the baseline to the 12-month follow-up period. Although nothing can be said about the significance, the $HbA_{1c}$ dropped for most users.

What is interesting to note is that the participants with the most self-efficacy are the ones who did not perform as well, and those who performed quite well did not have the best self-efficacy levels according to the used scales. Perhaps this indicates over-confidence about self-management, which is not matched in practice.

What is interesting to note is that the participants with the most self-efficacy are the ones who had increases in $HbA_{1c}$. Perhaps this indicates over-confidence about self-management, which is not matched in practice.
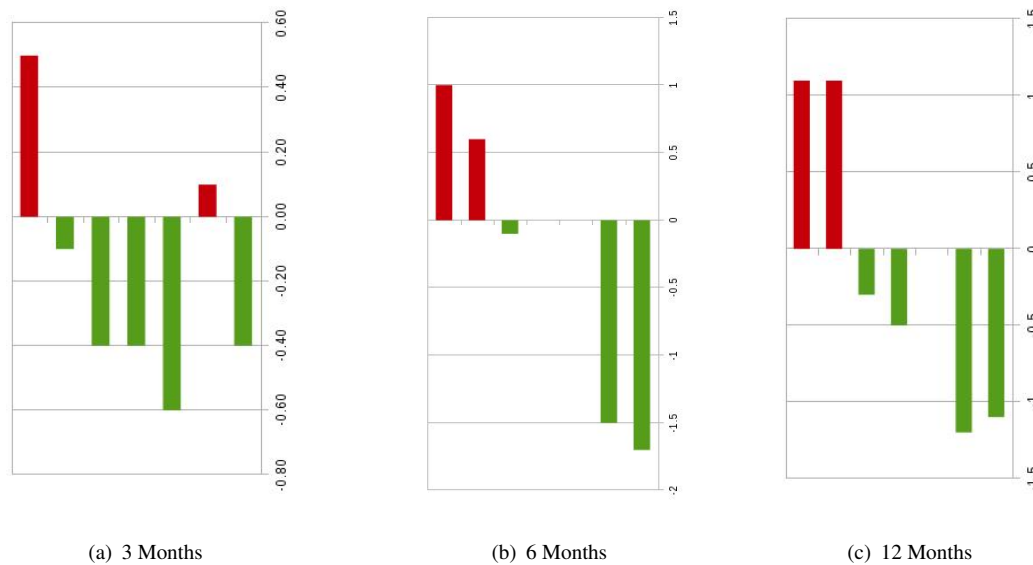


(a) 3 Months  (b) 6 Months  (c) 12 Months

FIGURE 6.3: Glycosylated Haemoglobin ($Hba_{1c}$) changes between the baseline and the follow-ups.

### 6.2.3  Limitations

The pilot study is subject to at least two limitations. First, our participants are people that have been recruited using newspaper adverts and direct phone calls based on their membership in the Diabetes Association. Therefore, it is conceivable that these individuals are more highly motivated to use mHealth technology than the average person with diabetes.

Second, this study was based on a comparatively small sample for studying social media. Although it is difficult to reason about the significance of the health outcomes and self-efficacy with this small sample, our qualitative analysis provides useful feedback. Additionally, I could have done more domain analysis of self-efficacy based on HeiQ as is recommended, but I used the DES-SF to check consistency in the outcomes.

## 6.3 DiabetesBuddy.org Demo

I have put up a website to demonstrate most of the algorithms developed in this dissertation. The site will become public once the related scientific articles are published, but privileged access can be obtained using:

- User : duranta
- Passwd: hoitLoGioL

The conceptual overview of the demo in shown in Fig 6.4. On the one hand there is the expansion-reduction algorithm, demonstrating how we can derive the important features for any selected outcome. The other part demonstrates the collaborative filtering algorithm on the stored datasets.

Most of the information will continue to evolve as more research is done in the field and on varied datasets, and updated on the project website.
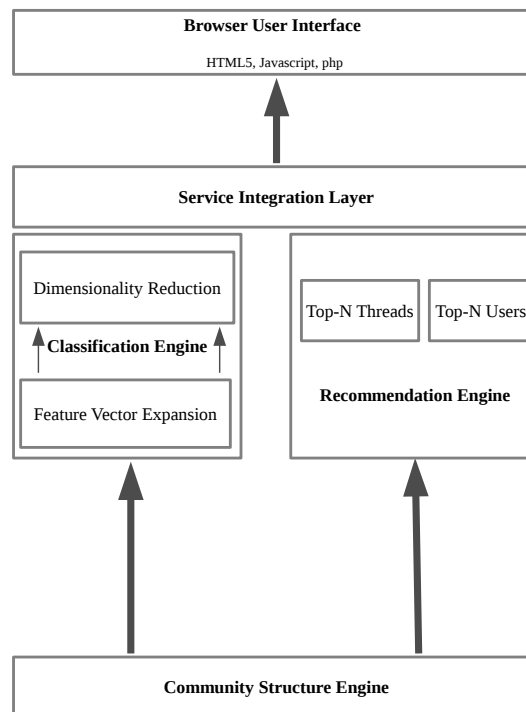


FIGURE 6.4: DiabetesBuddy.org Architecture

## 6.4 Knowledge Summary

The work in this chapter helped build evidence to support the significance and practical relevance emerging technologies in disease management, thereby helping partially answer the Research Question **Q1**.

*What is already known on the topic*

- There is some evidence of mHealth benefits in diabetes self-management.
- Group education in diabetes self-management can have short-term benefits.

*What this chapter added to our knowledge*

- The pilot study provides preliminary collective evidence of the potential relevance of emerging technology concepts – eHealth, mHealth and Social Media – in disease management and in particular, diabetes self-management.
- mHealth after group education can have beneficial effects in the longer term.

## 6.5 Chapter Summary

The pilot study shows that social interaction by regular meetings and online interaction can influence self-management outcomes positively. The use of a mobile application contributed to persuading the participants to measure their BG values more frequently, and social networking seemed to have influenced the motivation of just a few of the participants to remain engaged. Summed up, these results provide some evidence that warrants further investigation because larger studies are required before much of the results can be more fully exploited.

# Chapter 7

# Conclusion

## 7.1  Scientific Contributions

In this section scientific contributions are revisited, but with more detail about how these contributions have been achieved vis-à-vis the original research questions and objectives of the dissertation.

> **Contribution 7.1 — SC1 - Paper 1.**  A systematic review of mHealth applications in diabetes self-management.  ∎

This contribution is related to the first Research Question (**Q1**) and the point was to survey the literature before much of the work begun, to establish relevance of emerging tools to diabetes self-management. What came out of this contribution is a scientific article that has arguably become my most influential piece of work.

The contribution is a framework for analysing mobile applications for diabetes self-management. A categorical list of *Features* or functions was developed for existing mobile applications. The core of the contribution lies in raising of important questions such as whether the existing applications are consistent with evidence-based guidelines. I highlighted these clinical guidelines as concrete *Requirements* documents upon which specifications should be built, thereby revealing gaps between what is available and the requirements.

This work provided the basis for exploring IPC for diabetes self-management, because there seemed to be very limited knowledge about something so popular, and whose numbers were quite compelling.

Much of the work is described in *Chapter 2*.

> **Contribution 7.2 — SC2 - Paper 5.** Feasibility study design for social mobile applications in diabetes. ∎

This contribution is also related to the first Research Question (**Q1**) and the point was to put to the test concepts available in some of the mobile applications that were reviewed. This is because much of the data obtained about mobile applications did not say anything about the users and usage. For Type 2 diabetes the user group tends towards much older adults, and it was important to demonstrate the feasibility and relevance to the user group.

The contribution includes simplicity in the design of the social interfaces and assimilating the mobile application in group education settings. This resulted in users measuring their blood glucose more frequently, and they continued to use the mobile phone for several months after the group education program ended.

Although this pilot did not directly analyse patients interactions, the observation was that first having met in person, social media solidified the friendships that persuade people to maintain good lifestyles.

Although the pilot was not properly powered, we saw improvements in $HbA_{1c}$, blood glucose levels and self-efficacy, and usage in the long term.

Much of this work is described in the first part of *Chapter 6*.

> **Contribution 7.3 — SC3 - Paper 2.** An evaluation of a network analysis abstraction of patient interaction patterns in IPC. ∎

This contribution is related to the second Research Question (**Q2**) and the point was to develop an abstraction of patient interactions, and show that it could be used to better understand the complexities of IPC interactions.

Previous related work had only been exploratory and much of the descriptions about designing the networks were ambiguous. The contribution was a formalised method for abstracting patient interactions in IPC, and showing that the resulting networks are consistent with evidence in the literature about the structure of social networks such the scale-free nature.

Much of this work is partially covered in *Chapter 3* and also described in *Chapter 4*.

> **Contribution 7.4 — SC4 - Paper 4.** An expansion of the abstraction for analysing how IPC for diabetes develop over time. ∎

This contribution partially addresses the third Research Question (**Q3**) and the point was to apply the abstraction to observing temporal trends of IPC. This work was important for generating new theories and hypothesis about how diabetes patients form communities over time. The symposium paper associated with this contribution resulted in an invitation for a full article in a new Springer Journal (Paper 3), with at least 40% new work.

Using time-slicing method with clustering in each partition, it was shown that the diabetes IPC were volatile and comprised mostly newly diagnosed patients. What initially seemed like vibrant communities were mostly superficial networks that emulated "Question and Answer" rather than active social experiences among peers. People joined typically immediately following diagnosis and left with a period of less than a year. This gave us a better understanding of the IPC in the case of diabetes.

Much of this work is described in *Chapter 4*.

> **Contribution 7.5 — SC5 - Paper 3.** Applying the abstraction to discover unique patterns that characterize IPC for diabetes. ∎

This contribution also addresses the third Research Question (**Q3**) and the point was to develop further the temporal analysis to discover unique patterns that characterize IPC for diabetes, and uncover the drivers behind the patterns.

This work was important for generating new theories and hypothesis about how the essential unique qualities of IPC differ from the general and non-health social networks such as Facebook. Support interventions have been used on Facebook, many of which have largely been unsuccessful, but this work was important for highlighting that there are important differences to consider when designing a social healthcare intervention.

The communities rallied behind a central figure, most of the time it was someone more knowledgeable than the rest. In one community someone who seemed to know a lot about the new insulin pumps generated a lot of traffic towards herself. Although all the networks exhibited a scale-free nature, the health networks had shrinking diameter and density as the communities grew. This contrasts with current evidence about other types of social networks.

Much of this work is described in the latter part of *Chapter 4*.

> **Contribution 7.6 — SC6 - Paper 6.** Based on the developed abstraction, a clustering-classification method for delineating interaction behaviours that correlate with health measurements. ∎

This contribution is related to the final Research Question (**Q4**) and the point was to develop further experimental evidence to show that the abstraction was useful in understanding how online participation may affect health status.

This work delineated patient interaction behaviours that had a bearing on health outcomes(in this instance, weight loss). The problem was modelled as a binomial classification task, and an expansion-reduction method was developed for the patient feature vector. The expansion is based on the network analysis abstraction, and the reduction based on feature subset selection.

Using the method, it was possible to show that the way patients interact had a stronger bearing on weight loss performance than their basic demographic data, with the classification reaching an F-score of 0.977, precision of 0.978 and AUC of 0.996.

Much of this work is described in the first part of *Chapter 5*.

> **Contribution 7.7 — SC7 - Paper 7.** A collaborative filtering method for threads and users based on the developed abstraction. ∎

This contribution partially addresses the last Research Question (**Q4**) and the point was to develop further experimental evidence to show that the abstraction used had practical relevance for designing personalized eHealth interventions – to later help influence patients interact in the most effective way.

This part of the work developed a method that used the network analysis abstraction to improve user-similarity analysis in collaborative filtering. With controlled experiments, it was shown that community structure can improve top-N thread recommendations.

Much of this work is described in the last part of *Chapter 5*.

## 7.2 Main Conclusions

The main conclusions will be discussed in the context of the original Research Questions. For each research question I briefly describe the results, which were obtained by applying the methods described in the preceding section on scientific contributions.

> **Question 7.1 — Q1.** What is the relevance of Internet Patient Communities in diabetes self-management?

Tackling this question began with some background work to understand the context of eHealth, mHealth and patient interaction applications for diabetes self-management – and actually testing the concepts in a 12-month pilot study. Although the applications were usable in practice, many of the existing applications were not consistent with the requirements outlined in the clinical guidelines, but the pilot study showed that there was potential for improving health with use of the tools – thus establishing their relevance.

> **Question 7.2 — Q2.** How can interaction patterns in Internet Patient Communities be modelled?

Having established the relevance, an empirical evaluation of network analysis as a useful tool for abstracting complex interaction in IPC was done. Methods were tested for modeling the network, starting with bipartite networks, and reducing them to one dimension. Further analysis of alternative designs based on dense and sparse networks was also explored, and algorithms also provided. the resulting network structures were consistent with basic properties such as scale-free reported in the literature for other non-health social networks.

> **Question 7.3 — Q3.** What interaction patterns characterize Internet Patient Communities?

Analysis revealed how IPC self-organize into community structures with very centralized network topology. This may be partially explained by the fact that most interaction occurs around a knowledgeable person, who inevitably receives most of the communication.

As these communities continue to grow, the density shrinks, indication that less and less proportions of the communication happens among the patients. The diameter also shrinks, an indication that the patients continued to rally around a knowledgeable person

as the communities grew, and this suggests few friendships blossom among patients in these interactions.

The evidence seem to suggest that a mentor kind of relationship is apparent, and perhaps the focus of IPC should be on fostering meaningful mentorships. Current IPC seem to just encourage participation and ordinary friendships.

> **Question 7.4 — Q4.** How do these interaction patterns relate to health outcomes?

Based on the abstraction, methods were developed for relating interaction behaviours to health outcomes using ensembles of machine learning techniques. The result was that patient interaction behaviours were more important to predicting health outcomes than the demographic information.

There was a significant difference between the most active and the least active in the case of weight loss. The most active patients lost significantly more weight than the least active. Patients who connected to more than one sub-community tended to perform significantly better than those attached to singular sub-communities.

An understanding of how interactions affect health outcomes is important to place us in a better position to tailor eHealth interventions for the best possible effect. Predicting user interests was also possible in terms of threads, based on the interaction behaviour. This means we can encourage and foster online behaviours or patterns that promote better health.

### 7.2.1 Main Research Question - MQ

*(MQ) - What is the nature of patient interactions patterns in Internet Patient Communities, and how do these interaction patterns affect health outcomes?*

There is a large amount of unregulated IPC that continue to thrive, and this dissertation has enhanced our understanding of previously unknown elements of their nature. However, the evidence seems to suggest that health benefits from participating in online communities are quite limited for the majority, and in most cases, modest. For example, in most of the analysed weight loss communities (excluding surgery), most people lost less than 10lbs (4.5kg). Perhaps other measurements of success are warranted.

It's conceivable that the online camaraderie with peers and dispelling of initial fears immediately following diagnosis may be important elements to consider in addition to objective health measurements such as glycosylated haemoglobin or weight.

However, in overall, the dissertation shed light on the patterns that yield the most benefits, such as active participation in more than one sub-community.

**Involvement of Trained Health Service Professionals**

The potential of IPC in disease management has been shown. However, more is required from trained health service professionals to establish controlled environments where patients can share everyday experiences, but with the security of due caution.

**Fostering Mentor-Driven IPC Environments**

Current evidence seems to support the fostering of mentor-driven IPC because we have seen the network structures are very centralised, while the majority of the patients are newly diagnosed, at least in the diabetes case.

**Self-Monitoring in IPC**

Self-monitoring and public declaration of $HbA_{1c}$ or weight loss goals can sometimes be beneficial because a patient makes a public commitment. Divulging health information is frowned upon, but the paradox is that sharing experiences with peers is as personal as it gets, but perhaps security and anonymity are the key.

Obviously, there are many complex issues to consider when examining these emerging new trends in self-management of diabetes and other chronic illnesses, and I certainly hope my work provides the impetus for further research in the field.

# Bibliography

Abbassi Z, Mirrokni VS (2007) A recommender system based on local random walks and spectral methods in Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis WebKDD/SNA-KDD '07 pages 102–108 ACM, New York, NY, USA

Adamic L (1999) The small world web Research and Advanced Technology For Digital Libraries, Proceedings 1696:443–452

Adlerberth A, Rosengren A, Bresater L, Waller L, Ehnberg S, Welin L (1992) Group education in a diabetes school improves the metabolic control. Lakartidningen 89(23)

Aguilar C, Westman E, Muehlboeck JS, Mecocci P, Vellas B, Tsolaki M, Kloszewska I, Soininen H, Lovestone S, Spenger C, Simmons A, Wahlund LO (2013) Different multivariate techniques for automated classification of {MRI} data in alzheimer's disease and mild cognitive impairment Psychiatry Research: Neuroimaging 212(2):89 – 98

Ahmad MA, Borbora Z, Srivastava J, Contractor NS (2010) Link prediction across multiple social networks. in W Fan, W Hsu, GI Webb, BL 0001, C Zhang, D Gunopulos, X Wu, editors, ICDM Workshops pages 911–918 IEEE Computer Society

Almansoori W, Gao S, Jarada T, Elsheikh A, Murshed A, Jida J, Alhajj R, Rokne J (2012) Link prediction and classification in social networks and its application in healthcare and systems biology. Network Modeling Analysis in Health Informatics and Bioinformatics 1:27–36

B W Kernighan SL (1970) An efficient heuristic procedure for partitioning graphs The Bell system technical journal Vol. 49:291–307

Bandura A (1997) Self-efficacy. The Exercise of Control W. H. Freeman and Company, New York

Bangor A, Kortum PT, Miller JT (2008) An Empirical Evaluation of the System Usability Scale International Journal of Human-Computer Interaction 24(6):574–594

Barabasi A, Albert R (1999) Emergence of scaling in random networks Science 286(5439):509–512

Barabasi A, Albert R, Jeong H (1999) Mean-field theory for scale-free random networks Physica a 272(1-2):173–187

Baron J, McBain H, Newman S (2012) The impact of mobile monitoring technologies on glycosylated hemoglobin in diabetes: a systematic review. J Diabetes Sci Technol 6(5):1185–96

Bender JL, Jimenez-Marroquin MCC, Jadad AR (2011) Seeking support on facebook: a content analysis of breast cancer groups. Journal of medical Internet research 13(1)

Boots M, Sasaki A (1999) 'small worlds' and the evolution of virulence: infection occurs locally and at a distance Proceedings of the Royal Society of London Series B-Biological Sciences 266(1432):1933–1938

Boutin F, Hascoet M (2004) Cluster validity indices for graph partitioning in Information Visualisation, 2004. IV 2004. Proceedings. Eighth International Conference on pages 376 – 381

Bródka P, Saganowski S, Kazienko P (2012) Ged: the method for group evolution discovery in social networks. Social Network Analysis and Mining pages 1–14

Burton S, Morris R, Dimond M, Hansen J, Giraud-Carrier C, West J, Hanson C, Barnes M (2012a) Public health community mining in youtube in Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium IHI '12 pages 81–90 ACM, New York, NY, USA

Burton S, Tanner K, Giraud-Carrier C (2012b) Leveraging social networks for anytime-anyplace health information. Network Modeling Analysis in Health Informatics and Bioinformatics 1:173–181

Chakrabarti D, Kumar R, Tomkins A (2006) Evolutionary clustering in Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining. KDD '06 pages 554–560 ACM, New York, NY, USA

Chen Y (2010) Take it personally: accounting for individual difference in designing diabetes management systems in Proceedings of the 8th ACM Conference on Designing Interactive Systems DIS '10 pages 252–261 ACM, New York, NY, USA

Chomutare T, Fernandez-Luque L, Årsand E, Hartvigsen G (2011) Features of mobile diabetes applications: Review of the literature and analysis of current applications against evidence-based guidelines. Journal of Medical Internet Research 13(3):e65

Chomutare T, Tatara N, Årsand E, Hartvigsen G (2013a) Designing a diabetes mobile application with social network support. Stud Health Technol Inform 188:58–64

Chomutare T, Årsand E, Fernandez-Luque L, Lauritzen J, Hartvigsen G (2013b) Inferring community structure in healthcare forums: An empirical study. Methods of Information in Medicine 52(2)

Chomutare T, Årsand E, Hartvigsen G (2012) Temporal community structure patterns in diabetes social networks. in Proceedings of the International Conference on Advances in Social Networks Analysis and Mining. International Symposium on Network Enabled Health Informatics, Biomedicine and Bioinformatics pages 210–217 ACM/IEEE, Los Alamitos, CA

Chomutare T, Årsand E, Hartvigsen G (2013c) Characterizing development patterns of health-care social networks. Network Modeling Analysis in Health Informatics and Bioinformatics 2(3):147–157

Christakis NA, Fowler JH (2007) The spread of obesity in a large social network over 32 years New England Journal of Medicine 357(4):370–379 pMID: 17652652

Clar C, Barnard K, Cummins E, Royle P, Waugh N, Aberdeen Health Technology Assessment Group (2010) Self-monitoring of blood glucose in type 2 diabetes: systematic review. Health technology assessment (Winchester, England) 14(12):1–140

Clauset A, Newman MEJ, Moore C (2004) Finding community structure in very large networks. Physical Review E 70(6)

Cobb NK, Graham AL, Abrams DB (2010) Social network structure of a large online community for smoking cessation. American Journal of Public Health 100(7):1282–1289

Coiera E (2013) Social networks, social media, and social diseases BMJ : British Medical Journal 346

Cooper H, Booth K, Gill G (2008) A trial of empowerment-based education in type 2 diabetes–global rather than glycaemic benefits. Diabetes Res Clin Pract 82(2):165–71

Curtis V, Cousens S, Mertens T, Traore E, Kanki B, Diallo I (1993) Structured observations of hygiene behaviours in Burkina Faso: validity, variability, and utility. Bull World Health Organ 71(1):23–32

Davis W, Bruce D, Davis T (2007) Does self-monitoring of blood glucose improve outcome in type 2 diabetes? the fremantle diabetes study. Diabetologia 50(3):510–5

Dawson SP (2006) Online forum discussion interactions as an indicator of student community Australasian Journal of Educational Technology 22(4):495–510 the contents of this journal can be freely accessed online via the journal's web page (see hypertext link).

Deakin T, McShane CE, Cade JE, Williams RD (2005) Group based training for self-management strategies in people with type 2 diabetes mellitus. Cochrane Database Syst Rev 2(CD003417)

Dias A, Chomutare T, Botsis T (2012) Exploring the community structure of a diabetes forum. Stud Health Technol Inform 180:833–7

Donetti L, Muñoz M (2004) Detecting network communities: a new systematic and efficient algorithm J Stat Mech P10012

Douketis J, Macie C, Thabane L, Williamson D (2005) Systematic review of long-term weight loss studies in obese adults: clinical significance and applicability to clinical practice. Int J Obes (Lond) 29(10):1153–67

Durant KT, McCray AT, Safran C (2011) Identifying temporal changes and topics that promote growth within online communities: A prospective study of six online cancer forums. IJCMAM 2(2):1–22

Eysenbach G, Powell J, Englesakis M, Rizo C, Stern A (2004) Health related virtual communities and electronic support groups: systematic review of the effects of online peer to peer interactions BMJ 328(7449):1166

Eysenbach G, Till J (2001) Information in practice. ethical issues in qualitative research on internet communities British Medical Journal 323(2001):1103–1105

Fazel-Zarandi M, Devlin HJ, Huang Y, Contractor N (2011) Expert recommendation based on social drivers, social network analysis, and semantic data representation in Proceedings of the 2Nd International Workshop on Information Heterogeneity and Fusion in Recommender Systems HetRec '11 pages 41–48 ACM, New York, NY, USA

FIEDLER M (1973) Algebraic connectivity of graphs Czechoslovak Mathematical Journal 23(2):298–305

Folkesson J, Dam E, Olsen OF, Pettersen P, Christiansen C (2005) Automatic segmentation of the articular cartilage in knee mri using a hierarchical multi-class classification scheme. in JS Duncan, G Gerig, editors, MICCAI volume 3749 of Lecture Notes in Computer Science pages 327–334 Springer

Fortunato S (2009) Community detection in graphs. CoRR abs/0906.0612

Fouss F, Pirotte A, Renders JM, Saerens M (2007) Random-walk computation of similarities between nodes of a graph with application to collaborative recommendation IEEE Trans on Knowl and Data Eng 19(3):355–369

Frey BJJ, Dueck D (2007) Clustering by passing messages between data points. Science

Frost J, Massagli M (2008) Social Uses of Personal Health Information Within PatientsLikeMe, an Online Patient Community: What Can Happen When Patients Have Access to One Another's Data J Med Internet Res 10(3)

Funnell M (2010) Standards of care for diabetes: what's new? Nursing 40(10):54–6

Girvan M NM (2002) Community structure in social and biological networks Proc Natl Acad Sci U S A 99(7821-6)

Glasgow R, Kurz D, Dickman J, Osuna D, Strycker L, King D (2012) Linking internet-based diabetes self-management to primary care: lessons learned and implications for research translation and practice implementation Translational Behavioral Medicine 2(3):313–321

Glattre E (1991) Rationale for the multidisciplinary approach in epidemiological research Science of The Total Environment 106(1–2):147 – 153 <ce:title>Environmental Health </ce:title>

Glenny A, O'Meara S, Melville A, Sheldon T, Wilson C (1997) The treatment and prevention of obesity: a systematic review of the literature. Int J Obes Relat Metab Disord 21(9):715–37

Goldberg D, Nichols D, Oki BM, Terry D (1992) Using collaborative filtering to weave an information tapestry Commun ACM 35(12):61–70

Guy I, Zwerdling N, Ronen I, Carmel D, Uziel E (2010) Social media recommendation based on people and tags in Proceedings of the 33rd International ACM SIGIR

Conference on Research and Development in Information Retrieval SIGIR '10 pages 194–201 ACM, New York, NY, USA

Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH (2009) The weka data mining software: an update SIGKDD Explor Newsl 11(1):10–18

Herlocker JL, Konstan JA, Terveen LG, Riedl JT (2004) Evaluating collaborative filtering recommender systems ACM Trans Inf Syst 22(1):5–53

Hettne KM, de Mos M, de Bruijn AG, Weeber M, Boyer S, van Mulligen EM, Cases M, Mestres J, van der Lei J (2007) Applied information retrieval and multidisciplinary research: new mechanistic hypotheses in complex regional pain syndrome. Journal of biomedical discovery and collaboration 2:2+

Holme P, Huss M, Jeong H (2003) Subnetwork hierarchies of biochemical pathways Bioinformatics 19(4):532–538

Holtz B, Lauckner C (2012) Diabetes management via mobile phones: a systematic review. Telemed J E Health 18(3):175–84

IDF (2011) Idf diabetes atlas. International Diabetes Federation 1(5)

Jeong H, Tombor B, Albert R, Oltvai Z, Barabasi A (2000) The large-scale organization of metabolic networks Nature 407(6804):651–654

John GH, Kohavi R, Pfleger K (1994) Irrelevant features and the subset selection problem in MACHINE LEARNING: PROCEEDINGS OF THE ELEVENTH INTERNATIONAL pages 121–129 Morgan Kaufmann

Keeling M (1999) The effects of local spatial structure on epidemiological invasions Proceedings of the Royal Society of London Series B-Biological Sciences 266(1421):859–867

Kleinberg JM (1999) Authoritative sources in a hyperlinked environment J ACM 46:604–632

Kohavi R, John GH (1997) Wrappers for feature subset selection Artificial Intelligence 97(1–2):273 – 324 <ce:title>Relevance</ce:title>

Kummervold PEE, Johnsen JAA (2011) Physician response time when communicating with patients over the Internet. Journal of medical Internet research 13(4)

Lau A, Dunn A, Mortimer N, Proudfoot J, Andrews A, Liaw S, Crimmins J, Arguel A, Coiera E (2013) Consumers' online social network topologies and health behaviours Stud Health Technol Inform 192:77–81

Lebedev A, Westman E, Beyer M, Kramberger M, Aguilar C, Pirtosek Z, Aarsland D (2013) Multivariate classification of patients with alzheimer's and dementia with lewy bodies using high-dimensional cortical thickness measurements: an mri surface-based morphometric study Journal of Neurology 260(4):1104–1115

Lefebvre P, Pierson A (2004) The global challenge of diabetes. World hospitals and health services : the official journal of the International Hospital Federation 40(3):37–40

Leskovec J, Kleinberg J, Faloutsos C (2005) Graphs over time: densification laws, shrinking diameters and possible explanations in Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining KDD '05 pages 177–187 ACM, New York, NY, USA

Leskovec J, Lang KJ, Mahoney M (2010) Empirical comparison of algorithms for network community detection in Proceedings of the 19th international conference on World wide web WWW '10 pages 631–640 ACM, New York, NY, USA

Leung A, Dron W, Hancock J, Aguirre M, Purnell J, Han J, Wang C, Srivastava J, Mahapatra A, Roy A, Scott L (2013) Social patterns: Community detection using behavior-generated network datasets in Network Science Workshop (NSW), 2013 IEEE 2nd pages 82–89

L'Huillier G, Ríos SA, Alvarez H, Aguilera F (2010) Topic-based social network analysis for virtual communities of interests in the dark web in ACM SIGKDD Workshop on Intelligence and Security Informatics ISI-KDD '10 pages 9:1–9:9 ACM, New York, NY, USA

Liang X, Wang Q, Yang X, Cao J, Chen J, Mo X, Huang J, Wang L, Gu D (2011) Effect of mobile phone intervention for diabetes on glycaemic control: a meta-analysis. Diabet Med 28(4):465–9

Lin YR, Chi Y, Zhu S, Sundaram H, Tseng BL (2008) Facetnet: a framework for analyzing communities and their evolutions in dynamic networks. in Proceedings of the 17th international conference on World Wide Web. WWW '08 pages 685–694 ACM, New York, NY, USA

Liu Q, Chen E, Xiong H, Ding CHQ, Chen J (2012) Enhancing collaborative filtering by user interest expansion via personalized ranking. IEEE Transactions on Systems, Man, and Cybernetics, Part B 42(1):218–233

Liu Z, He JL, Kapoor K, Srivastava J (2013) Correlations between community structure and link formation in complex networks PLoS ONE 8(9):e72908

Lu Y, Zhang P, Liu J, Li J, Deng S (2013) Health-related hot topic detection in online communities using text clustering PLoS ONE 8(2):e56221

Ma X, Chen G, Xiao J (2010) Analysis of an online health social network. in Proceedings of the 1st ACM International Health Informatics Symposium IHI '10 pages 297–306 ACM, New York, NY, USA

Malucelli F, Cremonesi P, Rostami B (2012) An application of bicriterion shortest paths to collaborative filtering. in M Ganzha, LA Maciaszek, M Paprzycki, editors, FedCSIS pages 423–429

McKee R (2013) Ethical issues in using social media for health and health care research HEALTH POLICY 110(2-3):298–301

McPherson M, Lovin L, Cook J (2001) Birds of a feather: Homophily in social networks. Annual Review of Sociology 27(1):415–44

Mueller AS, Pearson J, Muller C, Frank K, Turner A (2010) Sizing up peers: adolescent girls' weight control and social comparison in the school context. J Health Soc Behav 51(1):64–78

Mulvaney SA, Anders S, Smith AK, Pittel EJ, Johnson KB (2012) A pilot test of a tailored mobile and web-based diabetes messaging system for adolescents. J Telemed Telecare 18(2):115–8

Newman M (2003a) The structure and function of complex networks Siam Review 45(2):167–256

Newman MEJ (2002a) Assortative mixing in networks Physical Review Letters 89(20):208701

Newman MEJ (2002b) Mixing patterns in networks Cornell University Library

Newman MEJ (2003b) The structure and function of complex networks SIAM REVIEW 45:167–256

Newman MEJ (2004) Detecting community structure in networks Eur Phys J B 38:321–330

Newman MEJ (2006) Modularity and community structure in networks Proceedings of the National Academy of Sciences 103(23):8577–8582

Newton K, Ashley A (2013) Pilot study of a web-based intervention for adolescents with type 1 diabetes. J Telemed Telecare

Ogden C, Carroll M, Kit B, Flegal K (2012) Prevalence of obesity in the united states, 2009-2010. NCHS Data Brief 82:1–8

Oinas-Kukkonen H, Harjumaa M (2009) Persuasive Systems Design: Key Issues, Process Model, and System Features Communications of the Association for Information Systems 24(28)

Palla G, Derényi I, Farkas I, Vicsek T (2005) Uncovering the overlapping community structure of complex networks in nature and society. Nature 435(7043):814–818

Palla G, lászló Barabási A, Vicsek T, Hungary B (2007) Quantifying social group evolution. Nature 446:2007

Paulweber B, Valensi P, Lindström J, Lalic NM, Greaves CJ, McKee M, Kissimova-Skarbek K, Liatis S, Cosson E, Szendroedi J, Sheppard KE, Charlesworth K, Felton AMM, Hall M, Rissanen A, Tuomilehto J, Schwarz PE, Roden M, Paulweber M, Stadlmayr A, Kedenko L, Katsilambros N, Makrilakis K, Kamenov Z, Evans P, Gilis-Januszewska A, Lalic K, Jotic A, Djordevic P, Dimitrijevic-Sreckovic V, Hühmer U, Kulzer B, Puhl S, Lee-Barkey YH, AlKerwi A, Abraham C, Hardeman W, Acosta T, Adler M, AlKerwi A, Barengo N, Barengo R, Boavida JM, Charlesworth K, Christov V, Claussen B, Cos X, Cosson E, Deceukelier S, Dimitrijevic-Sreckovic V, Djordjevic P, Evans P, Felton AMM, Fischer M, Gabriel-Sanchez R, Gilis-Januszewska A, Goldfracht M, Gomez JL, Greaves CJ, Hall M, Handke U, Hauner H, Herbst J, Hermanns N, Herrebrugh L, Huber C, Hühmer U, Huttunen J, Jotic A, Kamenov Z, Karadeniz S, Katsilambros N, Khalangot M, Kissimova-Skarbek K, Köhler D, Kopp V, Kronsbein P, Kulzer B, Kyne-Grzebalski D, Lalic K, Lalic N, Landgraf R, Lee-Barkey YH, Liatis S, Lindström J, Makrilakis K, McIntosh C, McKee M, Mesquita AC, Misina D, Muylle F, Neumann A, Paiva AC, Pajunen P, Paulweber B, Peltonen M, Perrenoud L, Pfeiffer A, Pölönen A, Puhl S, Raposo F, Reinehr T, Rissanen A, Robinson C, Roden M, Rothe U, Saaristo T, Scholl J, Schwarz PE, Sheppard KE, Spiers S, Stemper T, Stratmann B, Szendroedi J, Szybinski Z, Tankova T, Telle-Hjellset V, Terry

G, Tolks D, Toti F, Tuomilehto J, Undeutsch A, Valadas C, Valensi P, Velickiene D, Vermunt P, Weiss R, Wens J, Yilmaz T (2010) A European evidence-based guideline for the prevention of type 2 diabetes. Hormone and metabolic research = Hormon- und Stoffwechselforschung = Hormones et métabolisme 42 Suppl 1

Pfeil U, Zaphiris P (2009) Investigating social network patterns within an empathic online community for older people Comput Hum Behav 25:1139–1155

Polgreen P, Tassier T, Pemmaraju S, Segre A (2010) Prioritizing healthcare worker vaccinations on the basis of social network analysis. Infect Control Hosp Epidemiol 31:893–900

Pothen A, Simon H, Liou K (1990) Partitioning sparse matrices with eigenvectors of graphs Siam Journal On Matrix Analysis and Applications 11(3):430–452

Rosenbek LK, Wagner L, Lønvig EM, Hjelmborg J, Henriksen JE (2011) The effect of motivational interviewing on glycaemic control and perceived competence of diabetes self-management in patients with type 1 and type 2 diabetes mellitus after attending a group education programme: a randomised controlled trial. Diabetologia 54(7):1620–9

Rowe R, Creamer G, Hershkop S, Stolfo SJ (2007) Automated social hierarchy detection through email network analysis in Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis WebKDD/SNA-KDD '07 pages 109–117 ACM, New York, NY, USA

Rygg L Rise MB, Grønning K, Steinsbekk A (2012) Efficacy of ongoing group based diabetes self-management education for patients with type 2 diabetes mellitus. a randomised controlled trial. Patient Educ Couns 86(1):98–105

Sarkadi A, Rosenqvist U (2001) Field test of a group education program for type 2 diabetes : measures and predictors of success on individual and group levels Patient Education and Counseling 44(2):129–39

Schaeffer S (2007) Graph clustering Computer Science Review 1(1):27–64

Scott J, Tallia A, Crosson J, Orzano A, Stroebel C, DiCicco-Bloom B, O'Malley D, Shaw E, Crabtree B (2005) Social network analysis as an analytic tool for interaction patterns in primary care practices. Ann Fam Med 3(5):443–448

Sibal L, Home PD (2009) Management of type 2 diabetes: NICE guidelines Clin Med 9(4):353–7+

SIMON H (1955) On a class of skew distribution functions Biometrika 42(3-4):425–440

Sperl-Hillen J, Beaton S, Fernandes O (2011) Comparative effectiveness of patient education methods for type 2 diabetes: A randomized controlled trial Archives of Internal Medicine 171(22):2001–2010

Sperl-Hillen J, Beaton S, Fernandes O, Von Worley A, Vazquez-Benitez G, Hanson A (2013) Are benefits from diabetes self-management education sustained? The American journal of managed care 19(2):104–12

Sporns O, Tononi G, Edelman G (2000) Theoretical neuroanatomy: Relating anatomical and functional connectivity in graphs and cortical connection matrices Cerebral Cortex 10(2):127–141

Steinsbekk A, Rygg L, Lisulo M, Rise M, Fretheim A (2012) Group based diabetes self-management education compared to routine treatment for people with type 2 diabetes mellitus. a systematic review with meta-analysis BMC Health Services Research 12(1):213

Stephan KEea (2000) Computational analysis of functional connectivity between areas of primate visual cortex Phil Trans R Soc Lond B 355

Strogatz SH (2001) Exploring complex networks Nature 410(6825):268–276

Sturm R, Hattori A (2012) Morbid obesity rates continue to rise rapidly in the united states. Int J Obes (Lond)

Sturt J, Hearnshawn H, Wakelin M (2010) Validity and reliability of the dmses uk: a measure of self-efficacy for type 2 diabetes self-management Primary Health Care Res and Devpt 11(4):374–381

Su X, Khoshgoftaar TM (2009) A survey of collaborative filtering techniques Adv in Artif Intell 2009:4:2–4:2

Subbian K, Aggarwal C, Srivastava J (2013) Content-centric flow mining for influence analysis in social streams in Proceedings of the 22Nd ACM International Conference on Conference on Information &#38; Knowledge Management CIKM '13 pages 841–846 ACM, New York, NY, USA

Tatara N, Årsand E, Nilsen H, Hartvigsen G (2009) A review of mobile terminal-based applications for self-management of patients with diabetes. in eTELEMED pages 166–175 IEEE Computer Society

Traud AL, Mucha PJ, Porter MA, Porter MA (2011) Social structure of facebook networks

van der Velden M, Emam KE (2013) "not all my friends need to know": a qualitative study of teenage patients, privacy, and social media. JAMIA 20(1):16–24

van Deursen A, van Dijk JA (2011) Internet skills performance tests: are people ready for ehealth? J Med Internet Res 13(2):e35

Vasudevan M, Tchoua U, Gillard BK, Jones PH, Ballantyne CM, Pownall HJ (2013) Modest diet-induced weight loss reduces macrophage cholesterol efflux to plasma of patients with metabolic syndrome Journal of Clinical Lipidology 7(6):661 – 670

Wagner A, Fell D (2001) The small world inside large metabolic networks Proceedings of the Royal Society of London Series B-Biological Sciences 268(1478):1803–1810

Wasserman S, Faust K (1994) Social Network Analysis: Methods and Applications Cambridge University Press

Watts D, Strogatz S (1998) Collective dynamics of small world networks Nature 393(6684):440–442

Wegener D, Rossi S, Buffa F, Delorenzi M, Rüping S (2013) Towards an environment for data mining based analysis processes in bioinformatics and personalized medicine Network Modeling Analysis in Health Informatics and Bioinformatics 2(1):29–44

Wegrzyn-Wolska K, Bougueroua L, Dziczkowski G (2011) Social media analysis for e-health and medical purposes in Computational Aspects of Social Networks (CASoN), 2011 International Conference on pages 278 –283

Weiss JB, Berner ES, Johnson KB, Giuse DA, Murphy BA, Lorenzi NM (2013) Recommendations for the design, implementation and evaluation of social support in online communities, networks, and groups Journal of Biomedical Informatics (0):–

Weitzman ER, Adida B, Kelemen S, Mandl KD (2011) Sharing Data for Public Health Research by Members of an International Online Diabetes Social Network PLoS ONE 6(4):e19256+

Wen M, Kowaleski-Jones L, Fan JX (2013) Ethnic-immigrant disparities in total and abdominal obesity in the us American Journal of Health Behavior 37(6):807–818

Wicks P, Massagli M, Frost J, Brownstein C, Okun S, Vaughan T, Bradley R, Heywood J (2010) Sharing health data for better outcomes on patientslikeme J Med Internet Res 12(2):e19

Wijsman C, Westendorp R, Verhagen E, Catt M, Slagboom P, de Craen A, Broekhuizen K, van Mechelen W, van Heemst D, van der Ouderaa F, Mooijaart S (2013) Effects of a web-based intervention on physical activity and metabolism in older adults: randomized controlled trial. J Med Internet Res 15(11):e233

Yang J, Kim J, Kim W, Kim YH (2012) Measuring user similarity using electric circuit analysis: Application to collaborative filtering PLoS ONE 7(11):e49126

Yang J, Leskovec J (2011) Patterns of temporal variation in online media. in Proceedings of the fourth ACM international conference on Web search and data mining WSDM '11 pages 177–186 ACM, New York, NY, USA

Zachary W (1977) An information flow model for conflict and fission in small groups J Anthropol Res 33:452–473

Zeigler-Johnson C, Weber A, Glanz K, Spangler E, Rebbeck T (2013) Gender- and ethnic-specific associations with obesity: individual and neighborhood-level factors. J Natl Med Assoc 105(2):173–82

Zhang J, Tang J, Li JZ (2007) Expert finding in a social network. in K Ramamohanarao, PR Krishna, MK Mohania, E Nantajeewarawat, editors, DASFAA volume 4443 of Lecture Notes in Computer Science pages 1066–1069 Springer

Zheng W, Zou L, Feng Y, Chen L, Zhao D (2013) Efficient simrank-based similarity join over large graphs Proc VLDB Endow 6(7):493–504

Zhou N, Cheung WK, Qiu G, Xue X (2011) A hybrid probabilistic model for unified collaborative and content-based image tagging IEEE Transactions on Pattern Analysis and Machine Intelligence 33(7):1281–1294

Årsand E (2009) The few touch digital diabetes diary : user-involved design of mobile self-help tools for peoplewith diabetes. Ph.D. thesis University of Tromsø Norway

Årsand E, Frøisland D, Skrøvseth S, Chomutare T, Tatara N, G Hartvigsen JT (2012) Mobile health applications to assist patients with diabetes: Lessons learned and design implications. Journal of Diabetes Science and Technology 6(5):1197–1206

Årsand E, Tatara N, Østengen G, Hartvigsen G (2010) Mobile phone-based self-management tools for type 2 diabetes: the few touch application. Journal of diabetes science and technology 4(2):328–336

Årsand E, Tatara N, Østengen G, Hartvigsen G (2011) Wireless and mobile technologies improving diabetes self-management in Handbook of Research on Mobility and Computing: Evolving Technologies and Ubiquitous Impacts. pages 155–160 IGI Global

# Index

## V

## W

## Y