Visual Analytics for Health Monitoring and Risk Management in CARRE

Youbing Zhao[1], Farzad Parvinzamir[1], Hui Wei[1], Enjie Liu[1], Zhikun Deng[1], Feng Dong[1], Allan Third[2], Arūnas Lukoševičius[3], Vaidotas Marozas[3], Eleni Kaldoudi[4], and Gordon Clapworthy[1]

[1] University of Bedfordshire, Luton, UK, LU1 3JU
{youbing.zhao, farzad.parvinzamir, hui.wei, enjie.liu, zhikun.deng, feng.dong, gordon.clapworthy}@beds.ac.uk

[2] Knowledge Media Institute, the Open University, UK
allan.third@open.ac.uk

[3] Biomedical Engineering Institute, Kaunas University of Technology, Kaunas, Lithuania
{arunas.lukosevicius, vaimaro}@ktu.lt

[4] Physics of Medical Imaging & Telemedicine, School of Medicine, Democritus University of Thrace Dragana, Alexandroupoli, Greece
kaldoudi@med.duth.gr

Abstract With the rise of wearable sensor technologies, an increasing number of wearable health and medical sensors are available on the market, which enables not only people but also doctors to utilise them to monitor people's health in such a consistent way that the sensors may become people's lifetime companion. The consistent measurements from a variety of wearable sensors implies that a huge amount of data needs to be processed, which cannot be achieved by traditional processing methods. Visual analytics is designed to promote knowledge discovery and utilisation of big data via mature visual paradigms with well-designed user interactions and has become indispensable in big data analysis. In this paper we introduce the role of visual analytics for health monitoring and risk management in the European Commission funded project CARRE which aims to provide innovative means for the management of cardiorenal diseases with the assistance of wearable sensors. The visual analytics components of timeline and parallel coordinates for health monitoring and of node-link diagrams, chord diagrams and sankey diagrams for risk analysis are presented to achieve ubiquitous and lifelong health and risk monitoring to promote people's health.

Keywords: CARRE, visual analytics, wearable sensor, health monitoring, risk management

1 Introduction

The widespread use of wearable monitoring devices and mobile apps makes ubiquitous capture of life-logging personal health data a reality. Effective collection of long-term health-status data is valuable for clinical decisions and leads to strengthened interdisciplinary healthcare research and collaboration in supporting innovative medical care. With different types of wearable sensor the healthcare platform can aggregate heterogeneous health and medical data for health monitoring, disease prediction and risk management in a ubiquitous, personalised and continuous manner.

In this paper we present the ongoing work of visual analytics in the CARRE project [2] – Personalized Patient Empowerment and Shared Decision Support for Cardiorenal Disease and Comorbidities – funded by the 7th Framework Programme of the European Commission, which aims to provide innovative means for the management of cardiorenal diseases with the assistance of wearable sensors.

The target of CARRE is to provide personalised empowerment and shared decision support for cardiorenal disease, which is the condition characterised by simultaneous kidney and heart disease while the primarily failing organ may be either the heart or the kidney. In CARRE, sources of medical and other knowledge will be semantically linked with sensor outputs to provide clinical information personalised to the individual patient, so as to be able to track the progression and interactions of comorbid conditions. The ultimate goal is to provide the means for patients with comorbidities to take an active role in care processes, including self-care and shared decision-making, and also to support medical professionals in understanding and treating comorbidities via an integrative approach.

The CARRE repository acts as the central point of information storage for all CARRE applications. It conforms to the principles of the Semantic Web and the guidelines of Linked Data. The Linked Data guidelines can be summarized as follows:

Information stored in the CARRE repository consists of 'RDF triples'. RDF is a standard format for representing semantic data on the Web; an item of RDF data is a triple, which corresponds to a statement of the form 'subject predicate object'. Each term is a URI, often drawn from a standard vocabulary or ontology, making it easy to link triples from different sources – to allow Linked Data. RDF can be accessed through a SPARQL endpoint: SPARQL is a query language, much like SQL in syntax. The triples stored in the CARRE repository are either public or private. For private data, data privacy and security mechanisms have been deployed.

The CARRE repository [23] stores general medical knowledge relating to risk associations, evidence and observables, and is available for public querying without authentication, as it contains no personally identifying data for any patient and serves as a general-purpose resource for medical knowledge in a semantic format.

In CARRE the risk model is a large semantic graph structure data consisting of interlinked entities, such as risk elements and risk evidence, that are either related to ground knowledge in cardio-renal disease and comorbidities (symptoms, diseases, risk factors, treatments, medical evidence source data, educational content, etc.) or personalised to each patient (patient demographics, medical history, sensor data, lifestyle data, etc.) [19]. The data structure of the risk factor repository is shown in Figure 1.

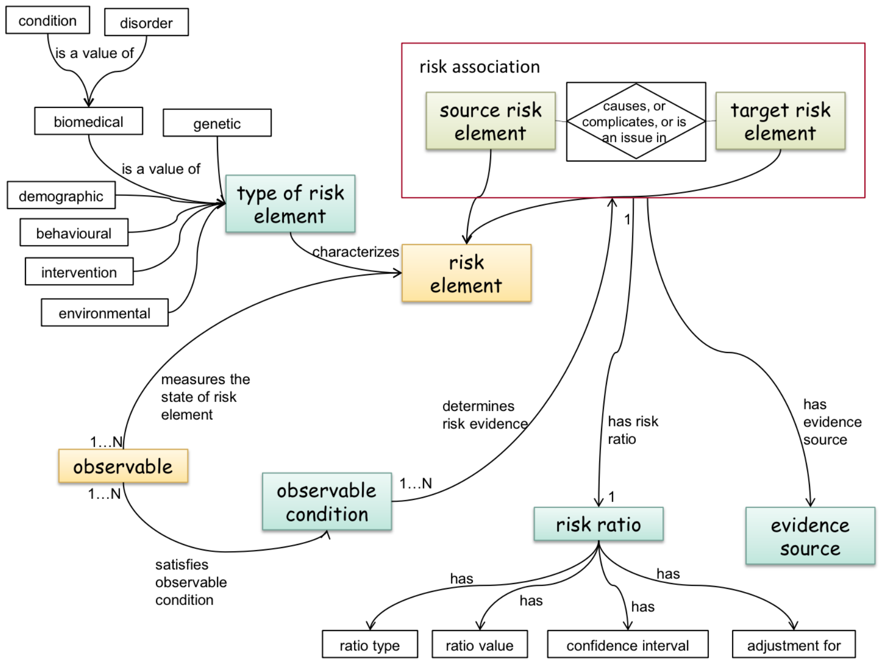

Figure 1. The risk model structure in the CARRE risk repository

The key concepts in the CARRE risk factor network are defined as follows:

- Risk Element: Risk elements include all the disorders/diseases involved in the comorbidity under discussion as well as any other risk causing agent, e.g. demographic (e.g. age, sex, race), genetic (gene polymorphisms), behavioural (e.g. smoking, physical exercise), environmental (e.g. air pollution, allergens) or even an intervention (e.g. pharmaceutical substances, contrast agents).
- Risk Association: The association of one risk element as the risk source with another risk element, which is the negative outcome under certain conditions, is a 'risk association'. A source risk element can be associated to a target risk element with more than one risk association.
- Risk Ratio: The association is always accompanied by the likelihood of the negative outcome occurring. This likelihood is expressed as a 'risk ratio', which is the ratio of the probability of the negative outcome when the person is exposed to the risk agent over the probability of the negative outcome when the person is not exposed to the risk agent.
- Risk Observable: In a risk association the prerequisite circumstances relate directly to the existence of the risk agent (source risk target) and/or its

severity, and/or any other specific conditions. These are reported via certain 'observables', that is, physical variables that can be measured or otherwise ascertained (e.g. biomarkers, biometric variables, biological signals and other non-biological factors, e.g. environmental).

Currently there are 98 risk factors, 53 risk elements, 253 risk evidences, 63 observables and 60 evidence sources in the CARRE risk data repository.

The risk model provided by CARRE can also been applied to other platforms for effective health monitoring and disease diagnosis. For example, the bridging of the CARRE risk model and the life-logging platform MyHealthAvatar [12,11] is under investigation.

To view and understand both the health status data and the risk data, visual analytics is indispensable in CARRE to provide patients and clinicians the ability to visualise, understand and interact with this linked knowledge and also take advantage of personalised empowerment services. The aim is to help medical professionals to better understand the individual patient's disease development and help patients to understand their own disease development, which in turn assists them to adhere to the self-management plan.

This paper presents the first stage of the CARRE Visual Analytics implementation: timeline and parallel coordinates are employed for health monitoring and correlation analysis, node-link diagrams, chord diagrams, sankey diagrams etc. are chosen for risk factor visualisation and risk analysis. A preliminary version of personalised risk visualisation and disease propagation simulation is demonstrated.

The paper is organised as follows: section 2 introduces related mobile fitness sensors and apps; section 3 discusses major visual analytics components of CARRE for health monitoring and risk management and section 4 concludes with the summary and future work.

2 Related Sensors and Apps

This section reviews existing life-logging health data collection tools and technologies that relate to CARRE, including wearable sensors and device-based health data collection tools, mobile app-based health data collection tools and health information sharing platforms.

1. Wearable devices for health data collection
   Wearable device based health data collection tools traditionally refers to use of medical devices to monitor medical data, such as heart rate, blood pressure, glucose, etc. Recently, the use of wearable devices in life-logging data collection mainly indicates the record of some personal physical activity data. In particular, prior work has shown that wearable sensors can benefit an individual patient's health and individual personal fitness. The most popular products are listed below:
   - Fitbit [3] provides wearable devices that record steps, distance, calories, etc. These devices communicate with a host computer using Bluetooth that sends their data directly to a user's account on the Fitbit website.

- Withings [14] is a consumer-level activity device providing steps, distance, calories, heart rate, etc. Devices include wristband, watch, scales, blood pressure monitor, etc.
- iHealth Labs [8] offers a range of connected health products: blood pressure monitors, activity trackers, glucometers, body analysis scales, etc.

There are also other fitness and health tracking wearable devices such as Apple Watch [1], Samsung Gear [13], Huawei wearables [7], etc which monitor fitness activities including heart rate, steps, activity mode and sleep quality.

2. Mobile app-based health data collection
Mobile apps have recently turned out to be a great source of user empowerment in healthcare fields. The most well-known mobile apps are based on observing GPS signal information for tracking user movement activities outdoors, including location, speed and distance. Some mobile apps explore the further use of mobile phone sensors for improving accuracy of tracking physical activities and observing other types of health information. Currently, the type of health data collected by mobile apps includes location, distance, speed, calories, heart rate, emotion and other manually recorded health data.
   - Moves [9] is an app for fitness and activity recording. Moves automatically records the step number and location of the user and calculates calories burned and distance of movement. It automatically recognises the activity type, such as walking, running, cycling, transport, etc. The user can either view the distance, duration, steps, and calories data on the mobile phone or export the data from the Moves server. The daily activities are visualised in a storyline in the app. The daily route can be visualised on the map.
   - Google MyTracks [5] is also based on the use of GPS to record the user's path, speed, distance and elevation while they are walk, run, and bike or do any activities outside. Google MyTracks is supported by Google with a comprehensive API documentation. The daily tracks can be shown on the map. The main drawback of GPS-based mobile apps is short battery lifetime and usability only for outdoor tracking. This might limit accuracy and continuities of life-logging captured personal activity data.
3. Health information sharing platforms: Lastly, health information sharing platforms have come with the emergence of web-enabled healthcare services. Due to the great evolution of Internet technology, this is emerging as a new healthcare delivery trend. These web-based healthcare platforms provide a multi-functional server for users to store, manage and and make basic visualisation of health data from various third party devices.
   - Microsoft HealthVault [6] is intended to enable users to gather, store, and use and share personal health information through many medical devices. It enables a connected ecosystem with privacy and security-enhanced foundation including more than 300 applications and more than 80 connected health and fitness devices.
   - MyHealthAvatar [12,11] is a European Commission funded project aimed to design a lifetime companion for people to collect, track and store

lifestyle and health data to promote personal well-being by collecting and aggregating life-logging data from wearable devices and mobile apps such as Fitbit, Moves, Withings, etc. and recording, storing and reusing the unified and structured personal health information in the long term, including activities, location, exercise, etc.
- Fluxtream [4] is an open-source non-profit personal data visualisation framework to help users make sense of their life and compare hypotheses about what affects their well-being.
- MyFitnessCompanion [10] is a healthcare platform for users to manage their personal health data, including metrics like weight, heart rate and heart rate variability (HRV), blood pressure, food intake, blood glucose, insulin, asthma, etc. The functionalities are highly similar to Microsoft HealthVault. It has a real-time visualisation mode, which keeps track of and visualises all user measurements with simple time graphs and can share these graphs with others.

Most of the health data collection devices, apps and health data repository platforms only provide very basic visualisation of the data they collect or import. They lack a visualisation and data analysis strategy on a systematic level, especially for analysis of wide time range data from heterogeneous sources. In addition, the employment of user interactions have not been widely studied in data visualisation and analysis in those applications. Moreover, none of the wearable devices or health information platforms can be directly associated with the user's personal health records to make health predictions due to lack of a verified risk model. CARRE can not only directly access personal health and lifestyle data from devices such as Fibit, Withings and iHealth, but also access data from multiple heterogeneous data sources via Microsoft HealthVault [6]. Based on risk models [19]extracted from medical literatures, CARRE supports personalised risk management and analysis, which is a key difference from the current available fitness sensors and health data platforms.

3 CARRE Visual Analytics

Visual analytics combines automated analysis techniques with interactive visualisations for an effective understanding, reasoning and decision-making on the basis of very large and complex datasets [20]. It is designed to promote knowledge discovery and utilisation of big data via effective visual paradigms and well-designed user interactions. In visual analytics, visualisation becomes the medium of an interactive analytical process, where humans and computers cooperate using their respective distinct capabilities for data processing and visual recognition for the most effective results. Thus, in visual analytics, user interaction constitutes the key for a user-centred data exploration. Visual analytics is an indispensable technology for information processing in the big data era.

In CARRE the data can be generally categorised into fitness data collected from sensors/apps, biomarker data from personal electronic health records (PHR) and risk factor data extracted from medical literatures. The role of visual analytics is to visualise health data, risk factor data and the integrated visual analysis of health data and risk factor data. In the current first stage implementation, CARRE provides web-based components for interactive health data visualisation and risk analysis, including healthline and parallel coordinates for fitness and biomarker data, node-link diagram, chord diagram and sankey diagram for risk factor data visual analysis and a preliminary experiment on personalised risk visualisation and disease progression simulation.

3.1 Fitness and biomarker data visualisation

3.2 Healthline

Lifestyle, health, fitness and medical data are inherently time dependent. To visualise time-varying data, a linear form timeline is a natural choice. A healthline is a special form of timeline to visualise multiple variables of continuous fitness statistics and biomedical markers which may cover a long period. Data trends can be observed from the variable curves and data correlations may be discovered by comparison of the data curves of the multi-variables. As the data records may cover a long time range, interactive techniques such as zooming and overview+details [15] are employed in the healthline visualisation. The users can also select the variables they are interested in the variable legend list to perform a user defined variable filtering. Figure 2 shows multiple biomarkers visualised in the interactive healthline in CARRE.

Figure 2. The healthline visualises personal fitness and biomarker data

3.3 Parallel coordinates

The technique of parallel coordinates is an approach for visualising multiple quantitative variables using multiple axes which are placed parallel to each other in the most common case [18]. The advantage of parallel coordinates is that it supports visualisation of multiple variables and correlation between attributes can be discovered by certain visualisation patterns. It is a common technique of visualising high-dimensional data and analyzing multivariate data. In CARRE there are a number of fitness and biomarker variables and both patients and medical practitioners like to view or study correlations among different variables. Consequently, parallel coordinates is chosen for multi-variable correlation visualisation analysis of fitness and biomarker data. An example view of the parallel coordinate view is shown in Figure 3 where negative correlations can be found between walking minutes and blood pressure as well as BMI (Body Mass Index).

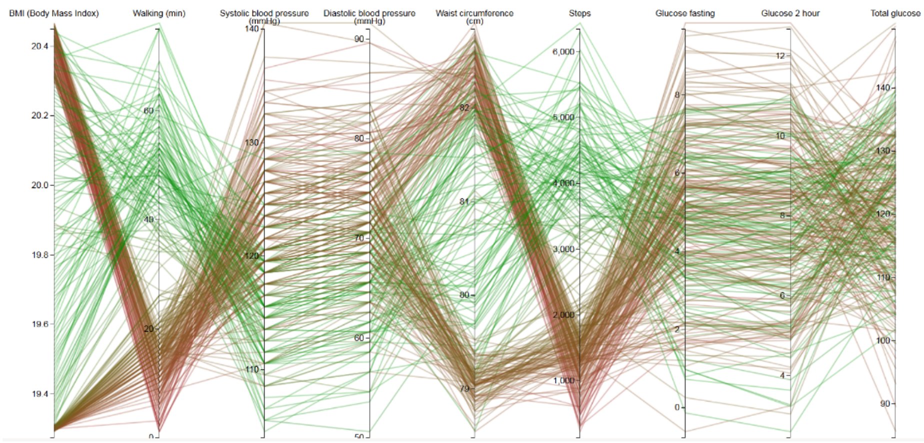

Figure 3. Fitness and biomarker visualisation and correlation analysis based on Parallel Coordinates

3.4 Risk factor visualisation

As introduced in section 1, the risk factor data in the CARRE repository is essentially a graph whose nodes are risk elements with multiple attributes attached, such as risk observables. Each directed edge represents a risk factor directed from the source to the target risk element. In CARRE node-link diagrams as well as other graph visualisation techniques, such as chord diagrams [17], sankey diagrams [22], etc., are used to visualise the risk factor graph. In this paper we focus on the node-link diagram and chord diagram which can be used by professionals and patients to view the risks.

3.5 Node-link diagram

Node-link diagram [21] is one traditional technique to visualise graph data structure visually, where point markers represent nodes in the graph and line segments connecting nodes represent edges in the graph. Node-link diagrams can be used to visualise trees and general network data structure. Different layout algorithms can be applied to minimize the number of distracting artifacts such as edge crossings. Force-directed layout algorithms [16] have been proved to be flexible and effective for dynamic and interactive graphs. Figure 4 is a force-directed node-link diagram visualisation of all 53 risk elements and 98 risk factors in the current CARRE repository. The node-link diagram clearly visualises risk associations and promotes studying and understanding of the the risk factor data base. Through the visual analytic interface, users are able to explore different diseases by dragging them to the centre to clearly see the relationships. By interactively viewing the graph, a patient also understands their risks in a more intuitive manner.

Figure 4. Force-directed layout based node-link diagram of risk elements and risk factors in CARRE: the whole risk factor database

3.6 Chord diagram

The disadvantage of the node-link diagram is that without proper handling, when the number of nodes and links increase, the visualisation will become increasingly messy for effective recognition by human beings, as obvious in Figure 4. Though filtering may be applied based on the conditions of a particular patient, it does not help for the visual analysis of the whole risk factor database. Fortunately there are some network visualisation techniques to alleviate the problem, such as the chord diagram [17], as shown in Figure 5.

Figure 5. Chord diagram of risk elements and risk factors in CARRE

The benefits of the chord diagram are that all the nodes are arranged on a circle and the edges from one node are grouped and bundled, which reduces the hairball problems which occur in the node-link diagram. With proper mouse hover interactions all the edges from or to one node can be highlighted, thus making the observation of the connections from or to one node much easier. The chord diagram clearly visualises the relationships of all risk elements in the repository and is particularly useful when professionals check and insert new risks.

3.7 Visual analytics of personalised risks and disease progressions

The ultimate goal of CARRE is to integrate the sensor data and the risk factor database to promote risk discovery, prediction and management, and individual health monitoring. In the first stage of CARRE visual analytics, we also made a prototype of interactive risk visualisation based on individual conditions and inputs. As shown in Figure 6, the node-link diagram only shows the related risk elements and risk factors based on the patient's profile, thus greatly reducing the complexity the original diagram. Moreover, the risk associations update dynamically based on changes in the patient's conditions. Currently, this is achieved by the interactive adjustment of some of the fitness and biomarker data. For example, if the blood pressure drops to the normal range, the hypertension risk element will disappear. In another example if the user walks more, the obesity risk element and all risk factors related to obesity will disappear.

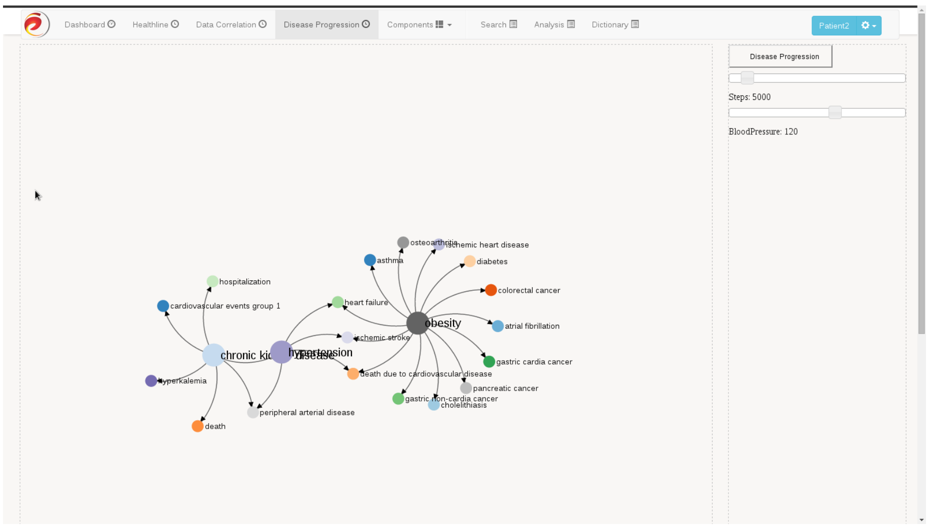

Figure 6. Interactive risk visual analysis: risks filtered and changed according to the conditions of a particular patient

4 Conclusions and Future Work

With the increasing popularity of wearable sensors, people are more interested in studying the applications of the sensors for lifestyle tracking and decision support for professional clinicians. While the amount of data collected by wearable sensors is huge, without visual analytics it is almost impossible to carry out any effective data analysis. This paper introduces in particular the role of visual analytics for exploiting sensor data for health monitoring and risk management in CARRE. Multiple variable time-dependant data visualised in a linear healthline helps to study and analyse fitness and biomarker data in CARRE, especially when proper user interaction techniques are incorporated. Parallel coordinates are very useful for correlation analysis of the fitness and biomarker samples. The network of risk elements and risk factors can be visualised with node-link diagrams and chord diagrams, with the latter preferable when the number of nodes becomes large. Moreover, a prototype of individual patient based interactive risk visualisation, prediction and management is presented to show the future directions of our work. In conclusion, interactive visual analytics is critical and effective in sensor-assisted health risk management and analysis; the future work will focus on integrating the risk model more closely with the process of the individualised and dynamic risk management, prediction and visualisation.

Acknowledgments CARRE project is funded by the European Commission's 7th Framework Programme – ICT under agreement FP7–ICT–611140. The related MyHealthAvatar project is funded by the European Commission's 7th Framework Programme – ICT under agreement FP7–ICT–2011–9. We would like to thank the European Commission for the funding and thank the project officers and reviewers for their indispensable support for both of the projects.

References

1. Apple Watch. http://www.apple.com/watch/, accessed: 2016-01-19
2. CARRE Project. http://www.carre-project.eu/, accessed: 2016-01-06
3. Fitbit. http://www.fitbit.com/, accessed: 2016-01-06
4. Fluxtream. https://fluxtream.org/, accessed: 2016-01-06
5. Google My Tracks. https://play.google.com/store/apps/details?id=com.google.android.maps.mytracks, accessed: 2016-01-20
6. HealthVault. https://www.healthvault.com/, accessed: 2016-01-06
7. Huawei wearables. http://consumer.huawei.com/en/wearables/index.htm, accessed: 2016-01-19
8. iHealth. http://www.ihealthlabs.com/, accessed: 2016-01-06
9. Moves. https://www.moves-app.com/, accessed: 2016-01-06
10. myFitnessCompanion. http://myfitnesscompanion.com/, accessed: 2016-01-20
11. MyHealthAvartar Platform. http://www.myhealthavatar.org/, accessed: 2016-01-06
12. MyHealthAvartar Project. http://www.myhealthavatar.eu/, accessed: 2016-01-06
13. Samsung wearables. http://www.samsung.com/us/mobile/wearable-tech, accessed: 2016-01-19

14. Withings. http://www.withings.com/, accessed: 2016-01-06
15. Cockburn, A., Karlson, A., Bederson, B.B.: A review of overview+detail, zooming, and focus+context interfaces. ACM Computing Surveys (CSUR) Surveys 41(1) (2009)
16. Fruchterman, T., Reingold, E.: Graph drawing by force-directed placement. Software - Practice & Experience(Wiley) 21(11), 1129–1164 (1991)
17. Holten, D.: Hierarchical edge bundles: Visualization of adjacency relations in hierarchical data. IEEE Transactions on Visualization and Computer 12(5) (2006)
18. Inselberg, A., Dimsdale, B.: Parallel coordinates: A tool for visualizing multidimensional geometry. In: Proc. the 1st IEEE Symposium On Visualization. pp. 361–378. IEEE (1990)
19. Kaldoudi, E., Third, A., Gotsis, G., Roumeliotis, S., Karvelyte, N., Rimseviciu, L., Pafili, K., Papazoglou, D., Juozalenaite, G., Semertzidou, E., Visockiene, Z., K., Z.: CARRE Deliverable D.2.2: Functional Requirements & CARRE information model". https://www.carre-project.eu/project-info/deliverables/ (July 2015), CARRE: Personalized Patient Empowerment and Shared Decision Support for Cardiorenal Disease and Comorbidities, FP7-ICT-61140
20. Keim, D., Kohlhammer, J., Ellis, G., et al.: Mastering the Information Age - Solving Problems with Visual Analytics. Eurographics Association (2010)
21. Liu, S., Cui, W., Wu, Y., Liu, M.: A survey on information visualization: recent advances and challenges. The Visual Computer 30(12), 1373–1393 (2014)
22. Riehmann, P., Hanfler, M., Froehlich, B.: Interactive sankey diagrams. In: Proc. IEEE Symposium on Infomation Visualization, InfoVis 2005. pp. 233–240. IEEE (2005)
23. Third, A., Kaldoudi, E., Gotsis, G., Roumeliotis, S., Pafili, K., Domingue, J.: Capturing scientific knowledge on medical risk factors. In: K-CAP2015: 8th International Conference on Knowledge Capture. ACM (2015)