



Graduação em Ciência da Computação

Andreza Fabiola Vieira de Abreu

**SISTEMA ANAMNESE: VISUALIZAÇÃO DE DADOS PARA
DIAGNÓSTICO DIFERENCIAL**

Trabalho de Graduação



Universidade Federal de Pernambuco
secgrad@cin.ufpe.br
www.cin.ufpe.br/~secgrad

RECIFE
2018



Universidade Federal de Pernambuco
Centro de Informática
Graduação em Ciência da Computação

Andreza Fabiola Vieira de Abreu

**SISTEMA ANAMNESE: VISUALIZAÇÃO DE DADOS PARA
DIAGNÓSTICO DIFERENCIAL**

*Trabalho apresentado ao Programa de Graduação em
Ciência da Computação do Centro de Informática da Uni-
versidade Federal de Pernambuco como requisito parcial
para obtenção do grau de Bacharel em Ciência da Com-
putação.*

Orientador: *Dr. Nivan Roberto Ferreira Júnior*

RECIFE
2018

Resumo

Estudos têm comprovado que o diagnóstico tardio é uma realidade no Brasil. Um desafio que os médicos enfrentam é que várias doenças apresentam sintomas semelhantes dificultando o diagnóstico. Mas como as doenças sempre vão existir é necessário estar preparado para combatê-las. Os sintomas iniciais geralmente são os mesmos e, conseqüentemente, o diagnóstico diferencial (possíveis doenças) é amplo. Conforme o tempo é que vão surgindo sintomas mais característicos e parte das doenças integrantes da relação do diagnóstico diferencial pode ser descartada. Este trabalho propõe o Sistema Anamnese para acompanhar o médico na formulação do diagnóstico diferencial. O sistema funciona através da Visualização de uma Rede Complexa de relacionamentos entre doenças e sintomas, onde o médico insere alguns sintomas e obtém um diagnóstico diferencial relevante pois, além da interseção entre os conjuntos de doenças, a rede é otimizada por fatores de unicidade. O sistema lida com interseções de mais de três conjuntos, o que torna a sua representação não trivial. Diversas técnicas de Visualização de Dados têm sido empregadas para fornecer um sistema que possa representar todas as interações possíveis entre conjuntos de forma intuitiva, o que se revela um grande desafio. Porém, ao analisar a representação do sistema surgiu o fato de que a visualização da própria rede complexa poderia ser usada de forma simples e objetiva para resolver o problema. A ferramenta Anamnese foi implementada em Javascript utilizando a biblioteca de Visualização de Dados D3.js, sendo o Force Directed Graph o algoritmo para a Visualização da Rede Complexa, e também conta com a ferramenta JQuery Select2 que otimiza pesquisas em caixa de seleção.

Palavras-chave: Visualização de dados, Redes Complexas, Diagnóstico Diferencial.

Abstract

Several studies have shown that late diagnosis is a reality in Brazil. One of the challenges doctors face is that several diseases have similar symptoms that make diagnosis difficult. But since diseases will always exist, it is necessary to be prepared to cope with them. The initial symptoms are generally the same, and consequently the differential diagnosis (possible diseases) is broad. As time goes on, more characteristic symptoms appear and part of the diseases that are part of the differential diagnosis relationship can be ruled out. This work proposes the Anamnesis System to accompany the doctor in the formulation of the differential diagnosis. The system works through the Visualization of a Complex Network of relationships between diseases and symptoms, where the doctor inserts some symptoms and obtains a relevant differential diagnosis because, in addition to the intersection between the sets of diseases, the network is optimized by factors of uniqueness. The system handles intersections of more than three sets, which makes its representation not trivial. Several Data Visualization techniques have been employed to provide a system that can represent all possible interactions between sets in an intuitive way, which is a great challenge. However, when analyzing the representation of the system came the fact that the visualization of the complex network itself could be used in a simple and objective way to solve the problem. The Anamnese tool was implemented in Javascript using the D3.js Data Visualization library, Force DirectedGraph being the algorithm for Complex Network Visualization, and it also has the JQuery Select2 tool that optimizes searches in the check box.

Keywords: Data visualization, Complex Networks, Differential Diagnosis.

Lista de Figuras

2.1	Layout de Árvore (a) Vista hierárquica clássica para um tamanho moderado de árvore; (b) Vista radial e Layout de Árvore + Link (c) Árvore Clássica com links adicionados. Fonte: Cui e Qu, 2007.	13
2.2	(a) Layout da Divisão Espacial: Visualização de um diretório de arquivos usando SunBurst; (b) Layout aninhado no espaço: TreeMap para visualização de um diretório de arquivos e (c) Layout 3D: Cone tree layout. Fonte: Cui e Qu, 2007.	14
2.3	Construção de Venn (a) para representar 4 conjuntos com elipses; (b) método geral para 4 e (c) 5 conjuntos com curvas sucessivas. Fonte: Wikipédia, 2017.	15
2.4	Diagramas de Venn: (a) sem área proporcional com círculos usando Venn Diagram Plotter; com área proporcional com (b) polígonos usando 3 Circle e (c) elipses usando eulerAPE para os dados numéricos em (a). Fonte: Alsallakh, 2014.	16
2.5	Diagramas de Euler com glyphs: (a) iguais usando eulerGlyphs; (b) diferentes usando Braths. Fonte: Alsallakh, 2016.	16
2.6	Variantes do Diagrama de Euler com divisão dos conjuntos: sem repetições usando (a) e (b) ComED; (c) com repetições usando DupED. Fonte: Alsallakh, 2016.	17
2.7	Técnicas region-based overlay para representar: (a) areas of interest (AOI) e (b) clusters usando Vizster. Fonte: Alsallakh, 2016.	17
2.8	Técnicas line-based overlay usando: (a) Parallel Tag Clouds e (b) Kelp Diagrams. Fonte: Alsallakh, 2016.	17
2.9	Glyph e ícones overlays usando: (a) uma lista de ícones e (b) nós codificados por cores; Técnicas híbridas de overlay: (c) regiões hachuradas como colour-coded glyphs. Fonte: Alsallakh, 2016.	18
2.10	Node-link techniques usando: (a) Jigsaw, (b) anchored maps e (c) PivotPaths. Fonte: Alsallakh, 2016.	18
2.11	Matrix-based techniques usando: (a) ConSet, (b) similarity matrix, (c) UpSet e (d) the KMVQL layout. Fonte: Alsallakh, 2016.	19
2.12	Aggregation-based techniques usando: (a) linear diagram, (b) Mosaic displays, (c) Parallel Sets e Radial Sets. Fonte: Alsallakh, 2016.	20
3.1	Hetnet do Projeto Rephetio. Fonte: Thiele, 2013.	22
3.2	Layout do UpSet dividido em duas partes: interação com conjuntos e interação com elementos. Fonte: Lex, 2014.	23
3.3	Caso de uso do UpSet R: Comparação entre o genoma da banana com o de outras cinco espécies. Fonte: Lex, 2014.	23

5.1	Layout do Sistema Anamnese. Fonte: Autoria própria.	27
5.2	Interações no Sistema Anamnese. Fonte: Autoria própria.	30
5.3	Diagnóstico diferencial contendo Dengue e Febre de Chikungunya. Fonte: Autoria própria.	31
5.4	Diagnóstico diferencial a partir da seleção dos sintomas Prurido e Febre. Fonte: Autoria própria.	31
5.5	Tríade Febre, Cefálea e Mialgia, simulando o diagnóstico diferencial de Virose. Fonte: Autoria própria.	32
5.6	Parâmetros que geram Atratividade: Qualidade Pragmática e Qualidade Hedônica. Fonte: Attrakdiff, na conta do usuário.	34
5.7	Média dos parâmetros de satisfação. Fonte: Attrakdiff, na conta do usuário. . .	35
5.8	Respostas do Questionário de Satisfação. Fonte: Attrakdiff, na conta do usuário.	35

Sumário

1	Introdução	8
1.1	Objetivo	10
1.1.1	Objetivos Específicos	10
1.2	Estrutura do Trabalho	10
2	Referencial Teórico	11
2.1	Redes Complexas	11
2.2	Técnicas de Visualização de Layout de Grafos	11
2.2.1	Node-Link Layout	12
2.2.1.1	Layout de Árvore	12
2.2.1.2	Layout de Árvore + Link	12
2.2.1.3	Spring Layout or Force-Directed Layout	13
2.2.2	Layout da Divisão Espacial	13
2.2.3	Layout Aninhado no Espaço	14
2.2.4	Layout 3D	14
2.2.5	Matriz Layout	14
2.3	Técnicas de Visualização de Conjuntos	15
2.3.1	Diagramas de Euler e Venn	15
2.3.2	Overlays	16
2.3.3	Node-link Diagrams	18
2.3.4	Técnicas Baseadas em Matrizes	19
2.3.5	Técnicas Baseadas em Agregação	19
3	Trabalhos Relacionados e Similares	21
3.1	Rede Complexa de Doenças	21
3.2	Interactive Set Visualization	22
3.3	Aplicativos de Diagnóstico Diferencial	24
4	Base de Dados	26
5	Sistema Anamnese	27
5.1	Metodologia	28
5.2	Interface e Funcionalidades	29
5.3	Casos de Uso	30
5.3.1	Diagnóstico Diferencial Focado na Interseção	30
5.3.2	Diagnóstico Diferencial Focado na Unicidade	31

5.3.3	Diagnóstico Diferencial das Viroses	32
5.4	Avaliação Inicial	32
6	Conclusão	36
	Referências	37
A	Apêndice	40
A.1	Cenário das Tarefas	40

1

Introdução

O diagnóstico precoce pode diminuir sequelas da doença, facilitar o tratamento (consequentemente economizando recursos) e aumentar as chances de sobrevivência no caso de doenças graves (BRASIL, 2011; ADAW, 200-). Por isso, o diagnóstico precoce é uma necessidade vital. Segundo o Ministério da Saúde, o diagnóstico precoce aumenta em 80% as chances de cura de câncer infantil (BRASIL, 2017).

Estudos têm comprovado que o diagnóstico tardio é uma realidade no Brasil. Um relatório feito pelo Tribunal de Contas da União (TCU), apontou pelas análises das Autorizações de Procedimento Ambulatorial (APACs) de 2010, que 60,5% dos pacientes com câncer já estavam no estágio 3 ou 4 quando diagnosticados, sendo que, no caso de neoplasias de brônquios e pulmões, esse valor chega a 87,9% (BRASIL, 2011). Outro estudo, coordenado pelo pesquisador Alexandre Grangeiro da Universidade de São Paulo (USP), mostra que 40% da mortalidade por AIDS no Brasil está relacionada a falta de um diagnóstico precoce (ESTADÃO, 2011).

Uma das causas apontadas para o diagnóstico tardio, considerando-se a partir da primeira consulta, é o diagnóstico errôneo produto da falta de qualificação profissional. Por exemplo, um estudo sobre o diagnóstico de hanseníase, realizado com 21 pacientes, mostrou que a maioria considerou que o diagnóstico tardio estava relacionado à baixa resolutividade na confirmação dos casos, um deles somente tendo o diagnóstico após 3 anos (SANTOS et al., 2007). Outro estudo sobre o diagnóstico de sarcoidose, realizado com 100 pacientes, constatou que o diagnóstico da sarcoidose em aproximadamente 60 pacientes (60%) demorou 6 meses ou mais e que ele é tardio em muitos casos, mesmo quando há achados de imagem sugestivos. Apenas 11% foi diagnosticado na primeira consulta e 26% passou por cinco ou mais médicos, incluindo pneumologistas. Ainda 17% foi diagnosticado com tuberculose sem nenhuma confirmação bacteriológica (RODRIGUES et al., 2013).

Em caso de doenças que podem ser letais em torno de uma semana, como é o caso da Leptospirose, a vigilância deve ser redobrada. A taxa de letalidade média no Brasil é de 10,8% e, diante das dificuldades no diagnóstico clínico e à confirmação laboratorial (pois esta doença, na fase inicial, apresenta os mesmos sintomas de qualquer virose), a indicação do Ministério

da Saúde (MS) é de que se o paciente apresentar Febre, Cefaléia e Mialgia, declarando que num período de um mês antes do início dos sintomas, teve contato físico com áreas alagadas ou tem atividades como coleta de lixo, um algoritmo de atendimento para a leptospirose deve ser seguido, inclusive preenchendo-se uma Ficha de Notificação (BRASIL, 2014). Em Pernambuco, os números divulgados de 2017 (DIÁRIO DE PERNAMBUCO, 2018; BRASIL, 2018) parecem indicativos de que o Estado tem seguido as orientações do MS, dada a proporção entre casos notificados (801) e confirmados (197) e, sabendo-se que o número de óbitos foi de 24 pacientes, o que resulta em 12%, que é próximo da média de letalidade nacional.

Por outro lado, dado que várias doenças apresentam sintomas semelhantes dificultando o diagnóstico, é grande o desafio que os médicos enfrentam na sua rotina profissional. Mas como as doenças sempre vão existir é necessário estar preparado para combatê-las. As ações estratégicas na prevenção e diagnóstico precoce de câncer (doença referência em diagnóstico tardio, conforme já foi relatado) indicam que o caminho é a capacitação e disseminação de informações dirigida aos profissionais de saúde que atuam na atenção primária em saúde de forma que os diagnósticos possam ser realizados menos tardiamente (BRASIL, 2011). O Movimento Todos Juntos Contra o Câncer (TJCC) firmou um compromisso de Educação Permanente, cuja proposta inclui disponibilizar capacitações gratuitas à distância e campanhas sobre a jornada de trabalho e estado de estresse dos profissionais da saúde (TJCC, 2014).

Para auxiliar os médicos durante a entrevista ao paciente, que tem como finalidade conhecer o caso clínico (anamnese), especificamente na obtenção dos sintomas e definição das possíveis doenças (diagnóstico diferencial), o presente trabalho propõe o Sistema Web Anamnese, para análise da coocorrência entre sintomas, através de uma Rede Complexa resultando no respectivo diagnóstico diferencial.

Note que, se trata de um problema de interseção das doenças relacionadas aos sintomas. As doenças são o conjunto e os sintomas são os elementos desse conjunto e, como muitos sintomas são recorrentes, esta interseção é de um número não trivial de conjuntos, o que dificulta a sua representação. Diversas técnicas de Visualização de Dados têm sido empregadas para fornecer um sistema que possa representar todas as interações possíveis entre conjuntos de forma intuitiva (ver seção 2.2), o que se revela um grande desafio. Porém, ao estudar como representar as interseções entre os conjuntos de doenças, surgiu o fato de que a visualização da própria rede complexa poderia ser usada de forma simples e objetiva para resolver o problema.

A ferramenta Anamnese foi implementada em Javascript utilizando a biblioteca de Visualização de Dados D3.js, sendo o Force Directed Graph o algoritmo para a visualização da Rede Complexa. O sistema é compatível com o navegador Chrome e também conta com a ferramenta JQuery Select2 que otimiza pesquisas em caixa de seleção. A base de dados foi contruída a partir do Guia de Vigilância em Saúde (BRASIL, 2016). Na fase inicial do projeto foram consultados uma enfermeira e um médico e, conseqüentemente, realizado um teste de usabilidade com médicos para analisar o protótipo.

1.1 Objetivo

Desenvolver um Sistema de Visualização de uma Rede Complexa para Diagnóstico Diferencial que dê suporte a capacitação técnica dos profissionais de saúde, visando a qualidade da saúde.

1.1.1 Objetivos Específicos

1. Auxiliar o médico no momento do diagnóstico diferencial, principalmente para o caso de doenças que não são habituais, mantendo o médico em contato com todos os sintomas (mesmo que raros) da doença e com doenças menos frequentes para sintomas habituais, a fim de que eles não cheguem a se esquecer dessas possibilidades.
2. Auxiliar o médico na indicação de sintomas característicos de doenças graves para casos suspeitos, em que o paciente fica sob observação, sendo orientado a voltar no caso de surgimento desses sintomas.
3. Ser acessível aos demais profissionais de saúde para que eles se mantenham atualizados sobre os sintomas e doenças mais frequentes, podendo assim auxiliar na prevenção de doenças e detecção precoce, orientando casos suspeitos a consultarem um médico.
4. Analisar a Usabilidade do Sistema.
5. Diante da diversidade de doenças e semelhanças de sintomas fomentar nos profissionais de saúde a necessidade da educação continuada.

1.2 Estrutura do Trabalho

Este capítulo introdutório contextualiza os desafios com a tarefa do diagnóstico diferencial, que na verdade se trata de várias interações entre conjuntos de doenças, um problema não trivial.

O capítulo 2 apresenta os conceitos básicos relacionados a Redes Complexas e as inovadoras técnicas de Visualização de Conjuntos, com destaque para os métodos que lidam com a interseção de mais de três conjuntos. O capítulo 3 apresenta os trabalhos com redes complexas de doenças a nível genético. Depois introduz a técnica Upset que permite vários tipos de interações, inclusive com mais de três conjuntos. Em seguida, observa-se avaliações de aplicativos de diagnóstico diferencial tanto para não-especialistas quanto para profissionais.

O capítulo 4 trata de como a Base de Dados foi construída. O capítulo 5 contém as etapas do projeto, análise dos dados e apresenta o sistema a nível de usuário. Finaliza com uma avaliação inicial da usabilidade. O capítulo 6 retoma os desafios enfrentados e o tipo de técnica do sistema resultante. Depois pode-se ver como os objetivos específicos foram atingidos.

2

Referencial Teórico

2.1 Redes Complexas

Uma rede nada mais é do que um grafo, e estes são compostos por um conjunto finito de pontos, denominados de nós ou vértices, e um conjunto de arestas que representam algum tipo de relação entre pares de vértices (LINDEN, 2009; METZ et al., 2007). Estas arestas podem receber um peso e o número de arestas do vértice é denominado grau.

A análise de Redes Complexas é uma área multidisciplinar da Ciência que, além de proporcionar o entendimento de grandes volumes de dados a partir da sua visualização, também possui uma consolidada teoria no estudo das ligações entre nós que, dependendo da quantidade de links em comuns com os vizinhos, podem ser agrupados em comunidades. Estudar as características desses links possibilita entender as interações entre eles e prever conexões, por exemplo, numa rede de proteínas quais as necessárias num evento bioquímico ou, no facebook, quem será amigo; quanto aos caminhos formados por esses links, possibilitam analisar a intensidade de impactos na rede, por exemplo, na disseminação de epidemias, o contágio de um indivíduo com muitos amigos é um fator crítico (PRUDÊNCIO, 2018; RODRIGUES, 2007).

2.2 Técnicas de Visualização de Layout de Grafos

Esta seção apresenta o trabalho realizado num survey de Cui e Qu (2007). Segundo os autores, os grafos node-link são os mais populares para a visualização de dados e incluem árvores e redes. Logo, para um melhor entendimento, é importante que os conceitos estejam bem definidos. Árvores são um tipo específico de grafo que não apresenta ciclos (WIKIPÉDIA, 2018b) e redes são mais genéricas (ver seção 2.1). A seguir, é mostrada a visão geral que foi dada para as representações gráficas de dados usando redes e seus algoritmos de layout.

2.2.1 Node-Link Layout

Descrever o problema do layout desse grafo é muito simples: calcular e representar as posições de um conjunto de nós com um conjunto de relações (arestas). No entanto, gerar layouts gráficos compreensíveis é bastante difícil. Existem algumas regras estéticas, a saber:

- Distribuir nós e arestas uniformemente.
- Evitar o cruzamento de arestas.
- Apresentar estruturas semelhantes da mesma forma.
- Minimizar as curvas ao longo das arestas.

Porém, na maioria das vezes, não é possível cumprir todas as regras. Há choque entre as regras e algumas são computacionalmente caras. Assim, um layout prático depende do compromisso com a estética. Outro ponto sobre o layout de grafos é a previsibilidade. Devido à tarefa de visualização gráfica, é essencial que dois resultados, executados no mesmo algoritmo, com a mesma ou semelhante entrada de dados, consequentemente, também sejam iguais ou semelhantes.

2.2.1.1 Layout de Árvore

Esse tipo de layout utiliza links entre os nós para representar os relacionamentos pai-filho numa árvore. Estruturas de árvores são mais simples que redes, por isso são mais tratáveis e fáceis de entender. Além disso, problemas que podem ser representados com o layout de árvore geralmente têm baixa complexidade.

O algoritmo tradicional é simples, rápido, previsível e esteticamente agradável, porém faz uso ineficiente do espaço da tela (Figura 2.1a). Outros algoritmos, para tornar esses layouts mais densos, têm sido desenvolvidos, porém eles são menos compreensíveis, como se pode ver no algoritmo do layout da Figura 2.1b, que posiciona os filhos de uma sub-árvore em forma radial de acordo com as suas profundidades.

2.2.1.2 Layout de Árvore + Link

Esta técnica utiliza a visualização de árvores em problemas com grafos grandes, isso é alcançado através de uma árvore que venha a cobrir todo o grafo. Depois disso, basta usar alguma das técnicas de visualização de árvores para mostrá-la e adicionar as arestas restantes do grafo (Figura 2.1c). O objetivo é ajudar os usuários a entender essas estruturas, que são mais difíceis de manipular que árvores.

Diversos problemas têm mais de dois tipos de relacionamentos. Caso um deles seja hierárquico, layouts de árvore + link são particularmente apropriados.

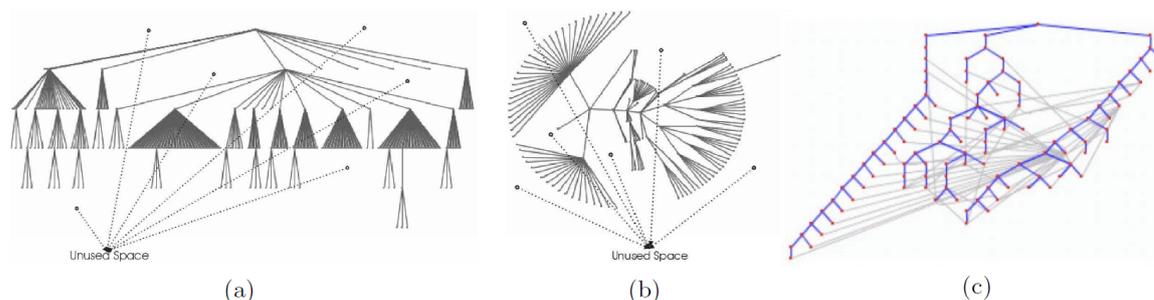


Figure 2.1 Layout de Árvore (a) Vista hierárquica clássica para um tamanho moderado de árvore; (b) Vista radial e Layout de Árvore + Link (c) Árvore Clássica com links adicionados. Fonte: Cui e Qu, 2007.

2.2.1.3 Spring Layout or Force-Directed Layout

Nesse layout os grafos são modelados como sistemas físicos de anéis ou molas. O Spring Layout é baseado em uma função de custo (energia), que mapeia diferentes layouts do mesmo gráfico com base em diferentes valores não-negativos de energia dos nós. Essa energia vai sendo reduzida e, a medida que se aproxima da energia mínima, o layout alcança resultados esteticamente mais agradáveis. Force-Directed layout foi o usado na rede do presente trabalho (Figura 5.1).

Uma vantagem sobre esse layout é que a analogia física pode ser naturalmente estendida para incluir informações estéticas adicionais para ajustar as forças entre nós. Sendo um dos primeiros algoritmos práticos para desenhar gráficos gerais, foi proposto por Eades (1984).

Os pesquisadores também tentam usar Spring Layouts para evitar problemas de sobreposição de nós. Além disso, existem algoritmos característicos dirigidos por força em iterações. Durante cada iteração, todas as posições dos nós são recalculadas com base em iteração. Porém, essas iterações aumentam o custo computacional do algoritmo.

Em geral, algoritmos direcionados à força produzem bons resultados para pequenos gráficos, mas grafos grandes normalmente dificultam que a função de energia alcance o mínimo. Além disso, esses algoritmos apresentam falta de previsibilidade, ou seja, duas execuções para um mesmo conjunto de dados num mesmo algoritmo podem ser diferentes. O que prejudica a navegação do usuário, pois esta depende da representação visual dos grafos.

2.2.2 Layout da Divisão Espacial

Nesse layout a hierarquia pai-filho é indicada pela anexação de nó (s) filho (s) para o nó pai. O Layout da Divisão Espacial deve ter uma orientação clara para diferenciar as adjacências entre os relacionamentos pai-filho e irmão, tendo como características aproveitar melhor a área limite (Figura 2.2a). Porém, como se pode ver no exemplo dado, alguns nós com muitos filhos são tão grandes que há muitos espaços em branco, enquanto outros são muito finos para serem rotulados ou coloridos (note que apenas as folhas da árvore são coloridas).

2.2.3 Layout Aninhado no Espaço

Layouts Aninhados, como Treemaps (Figura 2.2b), desenharam a estrutura hierárquica de maneira aninhada, com os nós filhos dentro de seu nó pai. Em um layout de Treemaps os nós são representados como retângulos, onde os sub-retângulos representam seus filhos. O tamanho de cada retângulo é proporcional a um atributo do nó. Treemaps tem o melhor aproveitamento de tela pois não deixam espaço livre no visor. Além disso, são indicados quando se quer mostrar o tamanho das folhas da árvore.

Os Treemaps clássicos têm como desvantagem o mesmo problema do layout anterior (presença de nós muito grandes ou muito finos) e estrutura hierárquica difícil de discernir, porque nós não-folha são representados implicitamente pelo encadeamento de nós filho em seus respectivos nós pai.

2.2.4 Layout 3D

Esse layout, por ter uma dimensão extra, pode dar mais espaço e deve ser mais fácil para exibir estruturas grandes. Muitos dos layouts 2D foram estendidos para 3D, a Cone Tree (Figura 2.2c) é o exemplo mais conhecido. A maioria dos layouts force-directed pode ser generalizado para 3D.

Outra vantagem desse layout é a familiaridade que o ser humano tem com o 3D, pois se assemelha ao mundo real. A desvantagem seria a oclusão dos objetos 3Ds. Portanto, é fundamental que visualizações 3Ds tenham rotação, para que as estruturas ocultas possam aparecer.

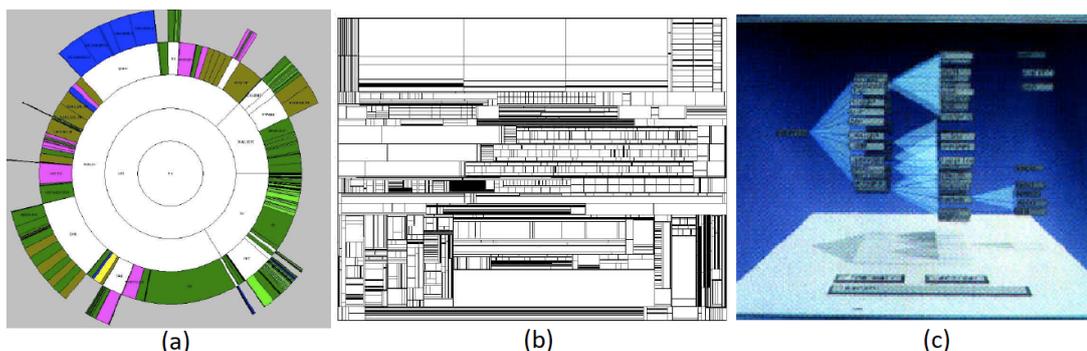


Figure 2.2 (a) Layout da Divisão Espacial: Visualização de um diretório de arquivos usando SunBurst; (b) Layout aninhado no espaço: TreeMap para visualização de um diretório de arquivos e (c) Layout 3D: Cone tree layout. Fonte: Cui e Qu, 2007.

2.2.5 Matriz Layout

Esse layout representa os grafos por suas matrizes de adjacência. Cada linha e cada coluna corresponde a um nó e um glifo na célula (i,j) codifica a respectiva aresta. Atributos da aresta podem ser codificados variando a cor, forma e tamanho do glifo. O principal benefício desse layout é a escalabilidade, podendo facilmente mostrar grafos com milhares de nós.

2.3 Técnicas de Visualização de Conjuntos

Uma infinidade de problemas podem ser resolvidos com o agrupamento dos elementos com base em algum critério de similaridade entre eles. Por exemplo, o diagnóstico clínico nada mais é do que um problema de interseção entre doenças que possuem sintomas em comum. O método mais popular de visualização de conjuntos é o Diagrama de Venn. Segundo a Wikipédia (2017; 2018a) este foi baseado no Diagrama de Euler e a diferença é que o diagrama de Venn tem que mostrar todas as possíveis relações entre os conjuntos. Ainda consta que a interseção entre mais de três conjuntos não é trivial, dessa forma Venn tentou resolver o problema com uma representação para quatro conjuntos (Figura 2.3a) e também desenvolveu um método geral com curvas sucessivas (Figura 2.3b-c).

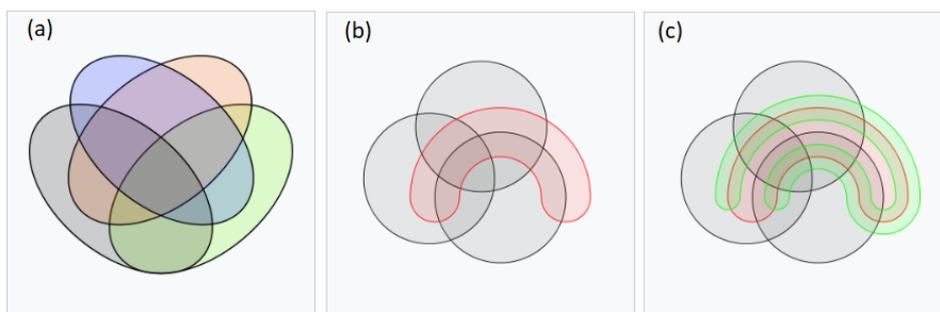


Figure 2.3 Construção de Venn (a) para representar 4 conjuntos com elipses; (b) método geral para 4 e (c) 5 conjuntos com curvas sucessivas. Fonte: Wikipédia, 2017.

Um survey de Alsallakh (2014; 2016) forneceu uma visão sistemática das técnicas de última geração para visualizar conjuntos e analisar suas relações. Os autores classificaram essas técnicas em 7 categorias de acordo com a representação visual, escalabilidade e tarefas relacionadas ao conjunto que elas suportam. A seguir, estas categorias são apresentadas com foco nas características da visualização para um número de conjuntos maior que três (a técnica de Scatter Plots não foi considerada representativa para esse critério, por isso foi omitida).

2.3.1 Diagramas de Euler e Venn

A maioria das técnicas ainda usa círculos porque são simétricos e isso auxilia a compreensão. Porém círculos têm graus de liberdade limitados, o que dificulta representar as proporções entre as áreas. Note que na Figura 2.4a a região com 3 é muito menor que aquela com 4. Outras implementações buscaram tratar esse problema dos diagramas tradicionais usando polígonos, curvas convexas de 4 ou 5 lados e curvas irregulares (Figura 2.4b). O uso de elipses por serem parecidas com os círculos, mas tendo grau de liberdade maior, proporciona diagramas estéticos e geralmente precisos. A Figura 2.4c representa os mesmos dados utilizados no diagrama da Figura 2.4a, porém com áreas proporcionais.

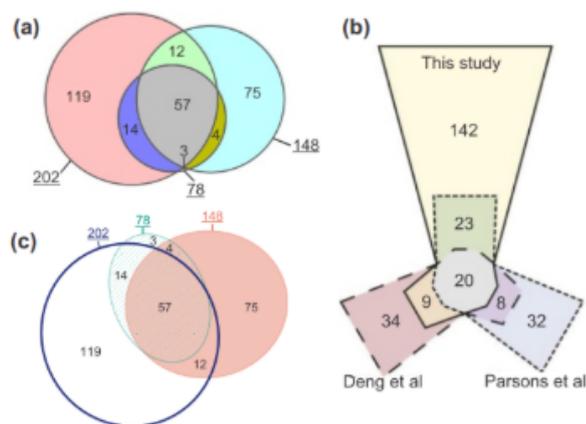


Figure 2.4 Diagramas de Venn: (a) sem área proporcional com círculos usando Venn Diagram Plotter; com área proporcional com (b) polígonos usando 3 Circle e (c) elipses usando eulerAPE para os dados numéricos em (a). Fonte: Alsallakh, 2014.

Para representar a cardinalidade da relação existem técnicas que utilizam glyphs e não as áreas das regiões. Os glyphs podem ser iguais (Figura 2.5a) ou de tamanhos e atributos diferentes (Figura 2.5b).

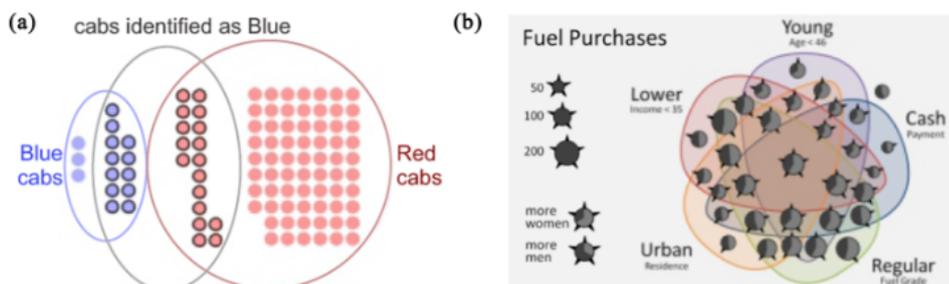


Figure 2.5 Diagramas de Euler com glyphs: (a) iguais usando eulerGlyphs; (b) diferentes usando Braths. Fonte: Alsallakh, 2016.

Outras técnicas variantes dos diagramas de Euler usam regiões fechadas para representar os conjuntos ou subconjuntos dos mesmos. Na Figura 2.6a temos um exemplo de método que divide ou duplica determinados conjuntos em partes disjuntas, conectando essas partes com arestas e na Figura 2.6b-c são representadas idéias semelhantes com os conjuntos divididos sem e com repetições, respectivamente.

2.3.2 Overlays

As técnicas Overlays são usadas quando os relacionamentos são parte de um cenário e precisam ser contextualizados. Por exemplo, dados espaciais georreferenciados num mapa. Podem ser usados elementos visuais como regiões (Figura 2.7), linhas (Figura 2.8), glyphs e ícones (Figura 2.9a-b) ou uma combinação deles (Figura 2.9c). Os conjuntos têm cores semi-transparentes para destacar as suas sobreposições e manter a visualização do contexto visível.

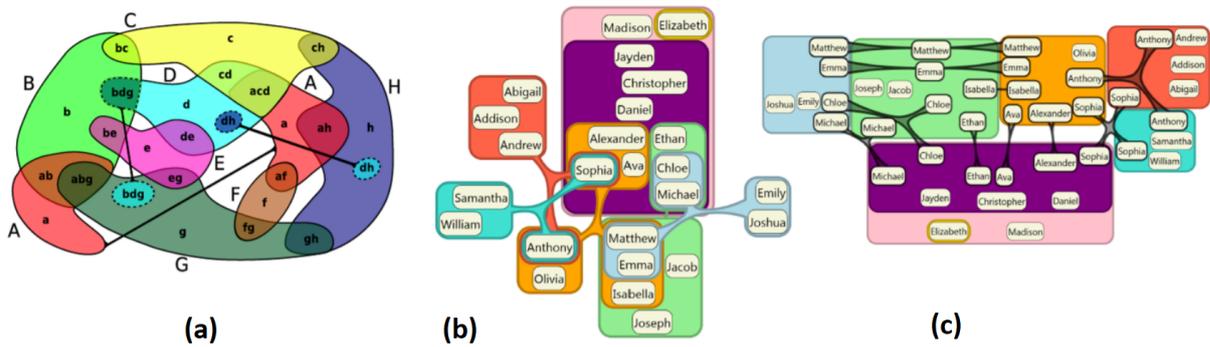


Figure 2.6 Variantes do Diagrama de Euler com divisão dos conjuntos: sem repetições usando (a) e (b) ComED; (c) com repetições usando DupED. Fonte: Alsallakh, 2016.

As técnicas Line-based overlays tentam evitar sobreposições, facilitando o acompanhamento da linha. Na Figura 2.8b, as sobreposições reais são marcadas com anéis concêntricos aos elementos, sendo codificadas por cores de acordo com o conjuntos e as linhas têm espessuras diferentes para reduzir a confusão visual.

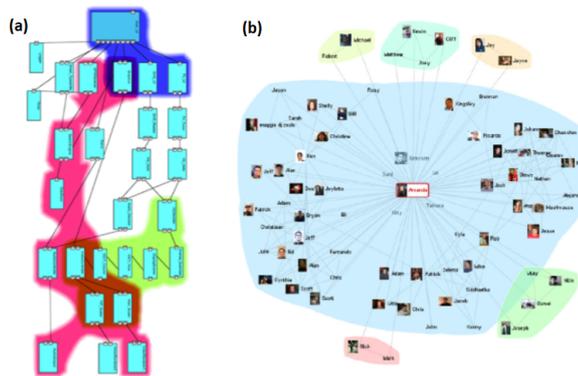


Figure 2.7 Técnicas region-based overlay para representar: (a) areas of interest (AOI) e (b) clusters usando Vizster. Fonte: Alsallakh, 2016.

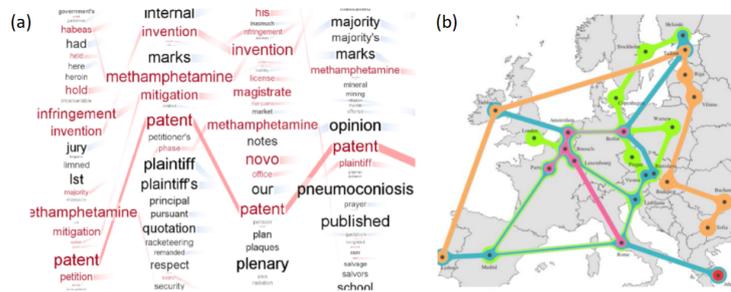


Figure 2.8 Técnicas line-based overlay usando: (a) Parallel Tag Clouds e (b) Kelp Diagrams. Fonte: Alsallakh, 2016.

Quando é suficiente representar a associação definida para o elemento individualmente, sem a necessidade de representar cada um dos conjuntos como um objeto conectado, os glyphs e ícones podem ser usados como sobreposições simples para definir associações (Figura 2.9a).

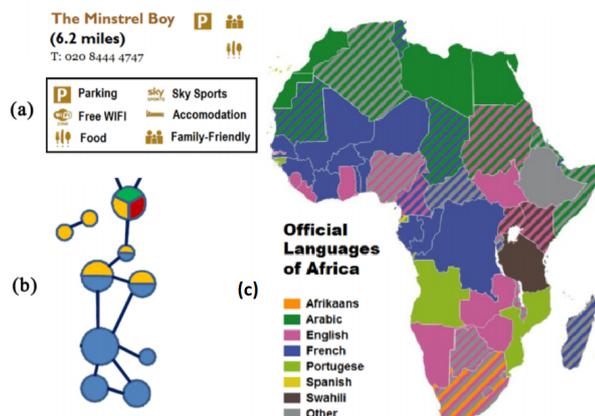


Figure 2.9 Glyph e ícones overlays usando: (a) uma lista de ícones e (b) nós codificados por cores; Técnicas híbridas de overlay: (c) regiões hachuradas como colour-coded glyphs. Fonte: Alsallakh, 2016.

2.3.3 Node-link Diagrams

As técnicas de visualização de dados centradas em elementos e centradas em conjuntos se baseiam em node-links diagrams. Aquelas centradas em elementos modelam os relacionamentos entre elementos e conjuntos como arestas de um grafo bipartido cujos nós representam os elementos e os conjuntos. Técnicas centradas em conjuntos focam nas relações entre os conjuntos (interseções, frequência, etc.). Um layout simples é o de listas paralelas, útil para mostrar coocorrência entre termos em documentos (Figura 2.10a), onde atributos adicionais dos elementos podem ser explorados usando cor ou colunas adicionais. Um layout circular dispõe os nós em torno de um círculo, um exemplo dessa técnica pode ser visto na Figura 2.10b, onde os nós representativos de conjuntos são colocados em torno do círculo e os que são compartilhados entre conjuntos dentro do círculo.

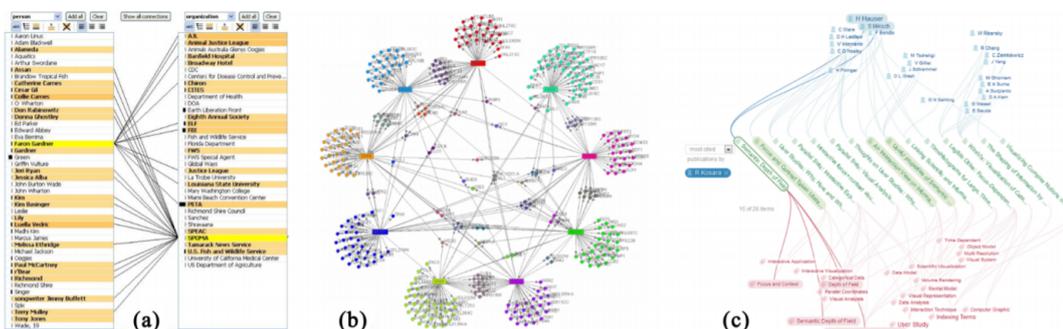


Figure 2.10 Node-link techniques usando: (a) Jigsaw, (b) anchored maps e (c) PivotPaths. Fonte: Alsallakh, 2016.

Existem diversos layouts, o da Figura 2.10c foi projetado para navegar entre várias informações. Node-link diagrams são usados para facilitar o raciocínio sobre a Análise Formal de Conceitos e mostrar similaridades, sendo geralmente fáceis de entender para uma quantidade pequena de nós.

2.3.4 Técnicas Baseadas em Matrizes

Essas técnicas são centradas em elementos ou centradas em conjuntos e aproveitam o conceito de linhas e colunas das matrizes. Várias interações e recursos visuais são possíveis para a agregação de elementos ou conjuntos. Elementos agregados podem ser indicados visualmente usando células mais escuras ou barras adicionais e as relações de pertinência podem ser codificadas por cores. As matrizes também podem ser reordenadas para analisar padrões.

As técnicas centradas em elementos representam os conjuntos como linhas e os elementos como colunas, cada célula codifica um elemento do conjunto (Figura 2.11a). Já as técnicas centradas em conjuntos apenas os representam como linhas e colunas, assim cada célula contém uma medida de similaridade entre os respectivos conjuntos, geralmente codificada em cores como em um mapa de calor (Figura 2.11b). A Figura 2.11c é uma forma de representar a agregação de conjuntos com barras e a Figura 2.11d mostra uma matriz reordenada por clusters, codificada por combinações booleanas, que também mostra a frequência dos elementos.

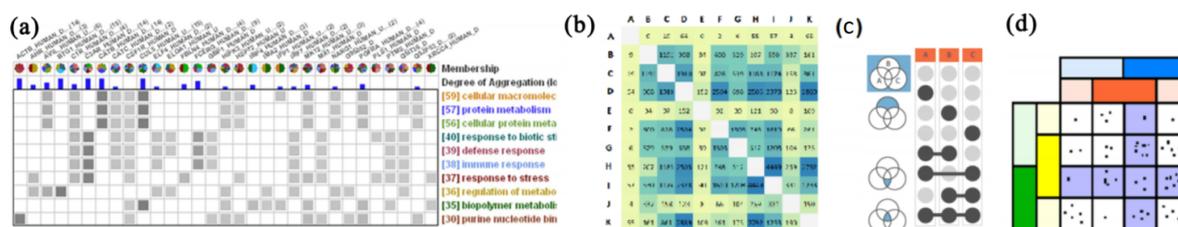


Figure 2.11 Matrix-based techniques usando: (a) ConSet, (b) similarity matrix, (c) UpSet e (d) the KMVQL layout. Fonte: Alsallakh, 2016.

Matrizes são livre de desordem e podem representar diferentes aspectos dos dados de conjuntos assim como os node-link diagrams. No entanto, são sensíveis à ordenação das linhas e colunas, o que tem grande impacto nos insights que elas podem revelar.

2.3.5 Técnicas Baseadas em Agregação

Quando o número de elementos é grande, torna-se menos viável para descrever e investigar como os elementos únicos pertencem aos conjuntos. Por isso, essas técnicas trabalham com a representação da frequência dos elementos e os acessam por demanda, por exemplo, como um gráfico de barras interativo. A técnica da representação dos conjuntos linear permite divisões mais genéricas e fáceis de entender (Figura 2.12a). Algumas técnicas dividem o espaço por variáveis categóricas para mostrar como os conjuntos se correlacionam (Figura 2.12b). Na técnica de conjuntos paralelos, o conjunto é representado num dos eixos horizontais de forma proporcional à quantidade total, assim as associações de conjuntos são tratadas como variáveis categóricas binárias (Figura 2.12c). Já a técnica de conjuntos radiais, representa os conjuntos em regiões não sobrepostas em posições radiais, geralmente usam cores para representar as associações entre conjuntos (Figura 2.12d).

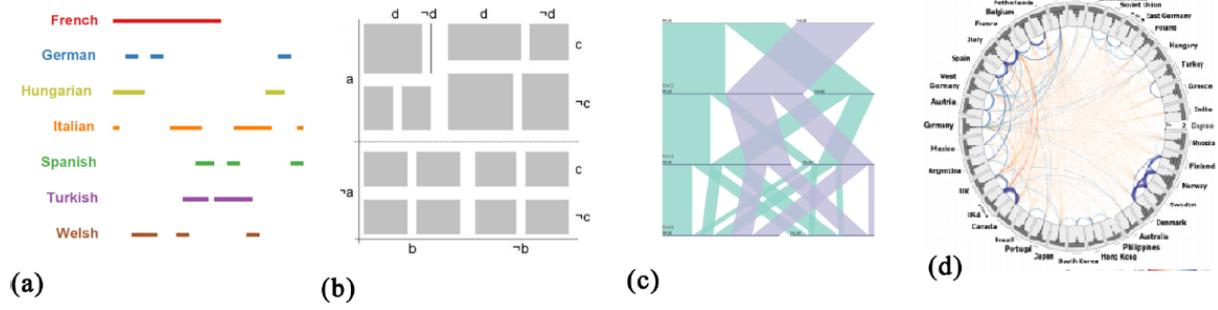


Figure 2.12 Aggregation-based techniques usando: (a) linear diagram, (b) Mosaic displays, (c) Parallel Sets e Radial Sets. Fonte: Alsallakh, 2016.

3

Trabalhos Relacionados e Similares

3.1 Rede Complexa de Doenças

No estudo da genética das doenças, uma importante base de dados é a Online Mendelian Inheritance in Man - OMIM (2018) que existe há 50 anos, tendo o mais completo repositório de genes de doenças. A base é pública e os usuários podem enviar dados que serão curados manualmente por especialistas antes da possível publicação.

O trabalho de Goh (2007) aparentemente foi o primeiro que analisou uma rede complexa de doenças, pois é bastante citado na literatura e relata que a maioria dos estudos usando a nova abordagem de redes centrou-se numa única doença. A base de dados utilizada foi a da OMIM (2018). A rede foi construída de modo que uma ligação é colocada entre uma doença e um gene da doença se mutações nesse gene levarem a respectiva doença. Também foram analisadas as projeções dessa rede formando uma rede de doenças e outra de genes, onde duas doenças estão conectadas, se houver um gene que esteja relacionado com ambas e dois genes estão conectados se eles estão envolvidos na mesma doença. Foram consideradas em torno de 300 doenças. A análise mostrou que as origens da maioria das doenças são compartilhadas com outras doenças e que vários genes da doença estão envolvidos em até 10 doenças, ou seja, a rede possui alguns hubs de genes relacionados às causas de doenças.

O trabalho anterior mostrou que distúrbios no metabolismo são a causa da maioria das doenças. Tendo em vista a importância do metabolismo para o estudo de doenças, surgiu o Recon 2, um consenso da comunidade de reconstrução metabólica, que é o mais amplo mapeamento metabólico (THIELE et al., 2013). Recon 2 é estruturado no formato padrão Systems Biology Markup Language (SBML) que é um meio de representação de modelos de redes bioquímicas (HUCKA et al., 2003).

O Rephetio (disponível em: <<https://het.io/>> e <<https://github.com/topics/rephetio>>) é um dos projetos que se fundamentou nos avanços da rede metabólica do Recon (THIELE et al., 2013; DUARTE et al., 2007). O projeto visa identificar computacionalmente a eficácia de medicamentos usando uma hetnet (Figura 3.1). A hetnet do projeto (rede com vários tipos de nós e/ou relacionamentos) consiste de 47.031 nós de 11 tipos e 2.250.197 relacionamen-

tos de 24 tipos, sendo que as doenças são representadas em 137 nós e os sintomas em 438 (think-lab). Os termos para as doenças foram selecionados da Disease Ontology (disponível em: <<http://disease-ontology.org/>>) e para os sintomas foi usado o vocabulário Medical Subject Headings - MeSH (disponível em: <<https://www.ncbi.nlm.nih.gov/mesh>>), considerando os descendentes de sinais e sintomas. A coocorrência de doenças e sintomas foi calculada a partir da mineração de tópicos dos artigos MEDLINE da National Library of Medicine (HIMMELSTEIN, 2015).

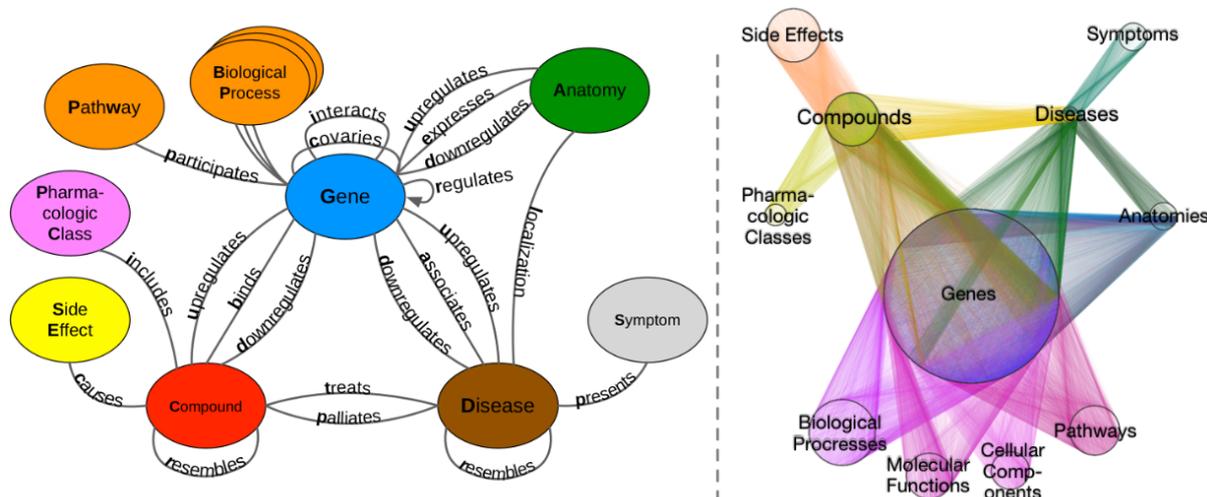


Figure 3.1 Hetnet do Projeto Rephetio. Fonte: Thiele, 2013.

3.2 Interactive Set Visualization

O UpSet é uma técnica centrada em conjuntos focada na análise dos dados através da combinação de interseções (agregados) acionadas por tarefas (LEX et al., 201-, 2014). Tendo em vista o crescimento exponencial dessas combinações, a ferramenta foi projetada baseada em matrizes (ver seção 2.2.4) com uma abordagem diferente da tradicional (matriz de elemento-conjunto ou conjunto-conjunto). No UpSet as colunas são os conjuntos e cada célula indica a presença ou ausência do conjunto em uma dada relação de interseção, marcadas pelas cores preto e cinza, respectivamente, e para reforçar a idéia da ligação, entre os elementos que participam de uma interseção, foram usadas barras que conectam os conjuntos presentes na operação (Figura 2.11c).

O Sistema é implementado em JavaScript e usa a biblioteca D3.js para visualização. Todo o layout da ferramenta foi projetado de modo a separar a visualização da interação dos usuários com os conjuntos e os elementos desses conjuntos, facilitando a análise dos dados, como se pode ver na Figura 3.2.

O UpSet foi desenvolvido pela Caleydo, uma comunidade de pesquisadores da Johannes Kepler University (JKU Linz), da Universidade de Utah e da Harvard Medical School, que tem

como principal objetivo a visualização de dados biológicos (acesso em: <<http://caleydo.org>>). A técnica surgiu com o objetivo de representar melhor a relação entre genomas de espécies, o primeiro caso de uso pode ser visto na Figura 3.3. Apesar da imagem original estar pequena, os autores fazem uma breve leitura do gráfico, mostrando como é simples observar tendências: a maioria dos genes é compartilhado entre todas as plantas, as três primeiras espécies [de baixo para cima](*Oryza_sativa*, *Sorghum_bicolor* e *Brachypodium_distachyon*) parecem estar altamente relacionadas, enquanto a sexta espécie (*Phoenix_dactylifera*) é a mais diferente" (LEX et al., 201-).

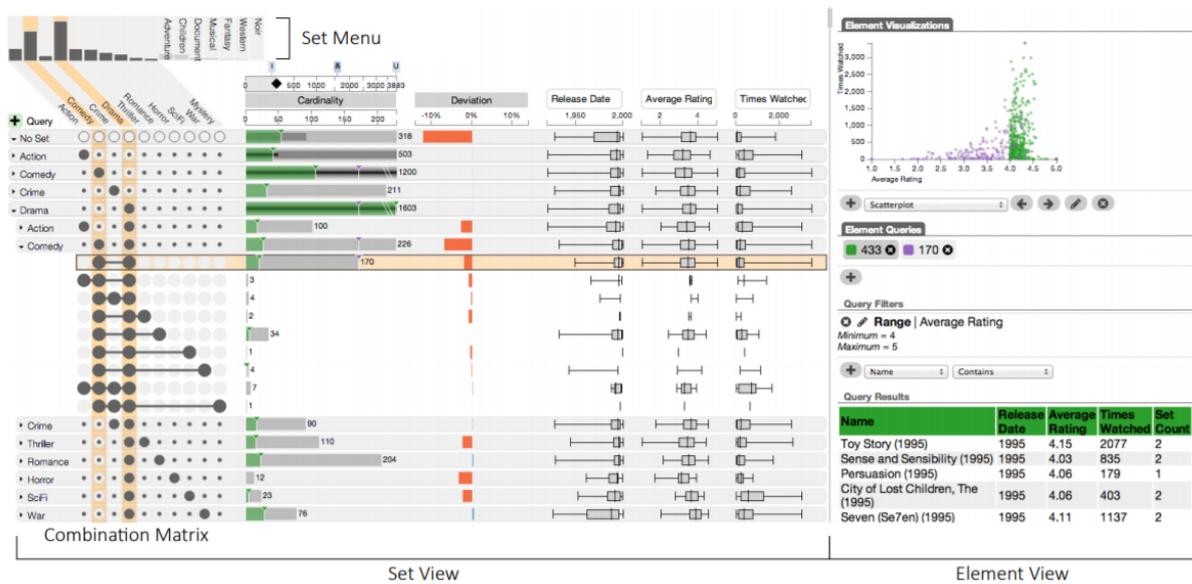


Figure 3.2 Layout do UpSet dividido em duas partes: interação com conjuntos e interação com elementos. Fonte: Lex, 2014.

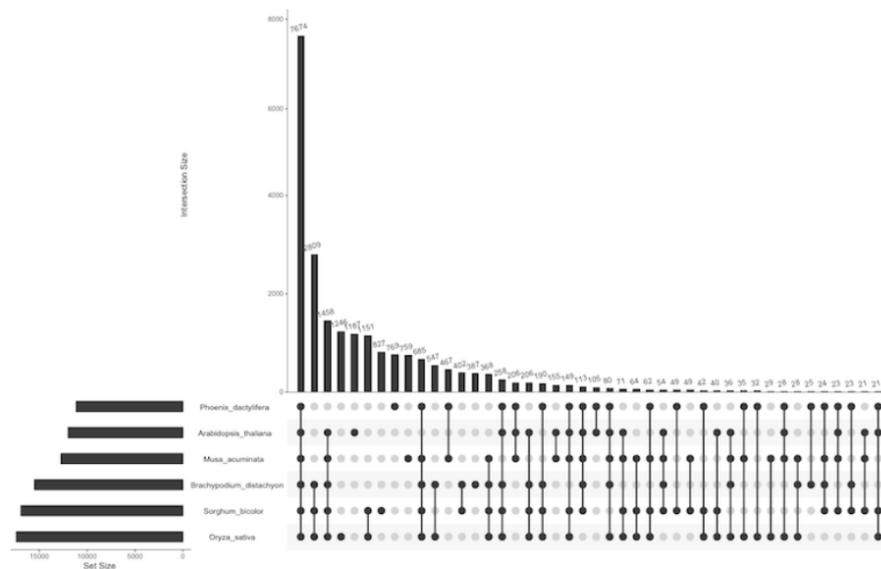


Figure 3.3 Caso de uso do UpSet R: Comparação entre o genoma da banana com o de outras cinco espécies. Fonte: Lex, 2014.

3.3 Aplicativos de Diagnóstico Diferencial

Diversas pessoas usam a internet para tirar dúvidas na área de saúde. Querer entender melhor como seu corpo funciona, alimentação saudável, obter mais detalhes da sua doença como tratamentos e os possíveis sintomas é natural. O que não se deve fazer é deixar de consultar um médico pois só ele é um profissional habilitado para tratar cada caso.

Neste contexto, o Google em parceria com o hospital Israelita Albert Einstein, criou um sistema (que não foi nomeado) que exibe um box com informações gerais da doença, como sintomas e tratamentos, quando a pessoa faz a busca sobre a doença. O UOL (2017) pediu para médicos avaliarem o sistema que teve resultado positivo, com a vantagem de ter uma linguagem simples e objetiva.

Um aplicativo para não-especialistas, desenvolvido pela Harvard Medical School com Inteligência Artificial, Buoy (disponível em: <<https://www.buoyhealth.com/technology/>>), precisa em geral de 20 perguntas antes dar o resultado. A proposta é simular o raciocínio de um médico, buscando ter uma visão individual de cada usuário, calcula quais doenças são mais prováveis.

Quanto aos aplicativos para especialistas, o blog Raciocínio Clínico (201-) avaliou 4 aplicativos de diagnóstico diferencial para profissionais de saúde (todos pagos), apresentados a seguir:

- **DXplain** (disponível em: <<http://www.mghlcs.org/projects/dxplain>>): aplicativo somente para instituições, projetado pelo Massachusetts General Hospital (também conhecido como MassGen) em 1987. Sua base de dados tem cerca de 2.500 doenças e outras 6.000 entradas, entre sintomas, sinais, resultados laboratoriais etc. O médico insere os dados clínicos e recebe o diagnóstico diferencial por ordem de importância com justificativas. Depois o médico ainda pode acrescentar mais dados para filtrar a lista. Uma característica [nesta lista do blog] única é que ele permite a entrada de ausência de sintomas (ausência de edema, sem febre) o que [neste aplicativo] ajuda a excluir algumas possibilidades da lista.
- **VisualDx** (disponível em: <<https://www.visualdx.com/>>): inicialmente projetado para diagnóstico em Dermatologia, constrói o diagnóstico diferencial usando sua biblioteca de imagens médicas [ainda tem versões na área da dermatologia para pacientes: os aplicativos Aysa (o paciente pode inclusive enviar uma foto da condição da pele) e Skinsight].
- **Isabel** (disponível em: <<https://www.isabelhealthcare.com/>>): destaca os resultados mais relevantes com red flags. O aplicativo funciona com processamento de linguagem natural, tendo uma base de 10 mil condições. Uma característica única [nesta lista do blog] é que apresenta os diagnósticos por faixa etária, sexo e localização geográfica do paciente. "(Probleminha [um grande problema]: percebemos

que, ao informar que nosso paciente mora no Brasil, sempre aparece malária entre as hipóteses - independente de quais sejam os sintomas!)"

- **Xebra** (disponível em: <<https://www.xebrapro.com/>>): usa estatísticas e análises de dados para processar entradas de sintomas, sinais, exames complementares, medicações e antecedentes para gerar uma lista de resultados ordenados por probabilidade.

Por fim, o autor citou um aplicativo para diagnóstico em português, que na verdade é apenas um coletor focado, a saber, o Dr. Diagnóstico, da PEBMED, maior produtora de aplicativos médicos do Brasil.

4

Base de Dados

Como o sistema de diagnóstico diferencial Anamnese é baseado numa Rede Complexa não direcionada, a base de dados principal é uma lista de arestas, no caso duplas de sintomas que têm coocorrência em doenças, para gerar uma rede de sintomas. Sendo o diagnóstico diferencial um conjunto de possíveis doenças, que tem sintomas semelhantes, no início também havia a intenção de avaliar a coocorrência de doenças.

Apesar do interesse não ser nas doenças genéticas, como análise preliminar foi utilizada a base de dados do Rephetio (ver seção 3.1), disponível em: <https://github.com/dhimmel/medline/tree/gh-pages/data>, para observar como seria a estrutura de uma rede de doenças-sintomas, especialmente a densidade da rede (quantidade de arestas em relação à quantidade de nós). Construindo-se a rede de doenças com os dados em "disease-disease-cooccurrence.tsv", resultaram 132 nós e 2.000 arestas, ou seja, uma rede com alta densidade. Logo, essa seria a tendência.

Na falta de uma base de dados pronta, a solução era construir a base a partir de um dicionário de doenças, mas para isso também era necessário definir um escopo de quais doenças abordar. Consultando alguns profissionais de saúde, estes afirmaram não ter conhecimento de tal dicionário. Então, uma enfermeira indicou o Guia de Vigilância em Saúde (Brasil, 2016), que por tratar apenas de 41 doenças e na sua maioria epidemiológicas, sua listagem foi adotada como escopo.

Partindo para a busca na web, foram acessados mais de 10 sites de saúde, porém nenhum tinha todas as páginas com as informações das doenças e sintomas de interesse com conteúdo html estruturado, de forma que fosse possível uma mineração de texto estruturada, restando apenas aplicar técnicas de linguagem natural que são mais avançadas, não havendo tempo hábil para a sua implementação. Então, o dicionário de doenças foi construído a partir da curagem manual do Guia de Vigilância citado (ver seção 5.1).

5

Sistema Anamnese

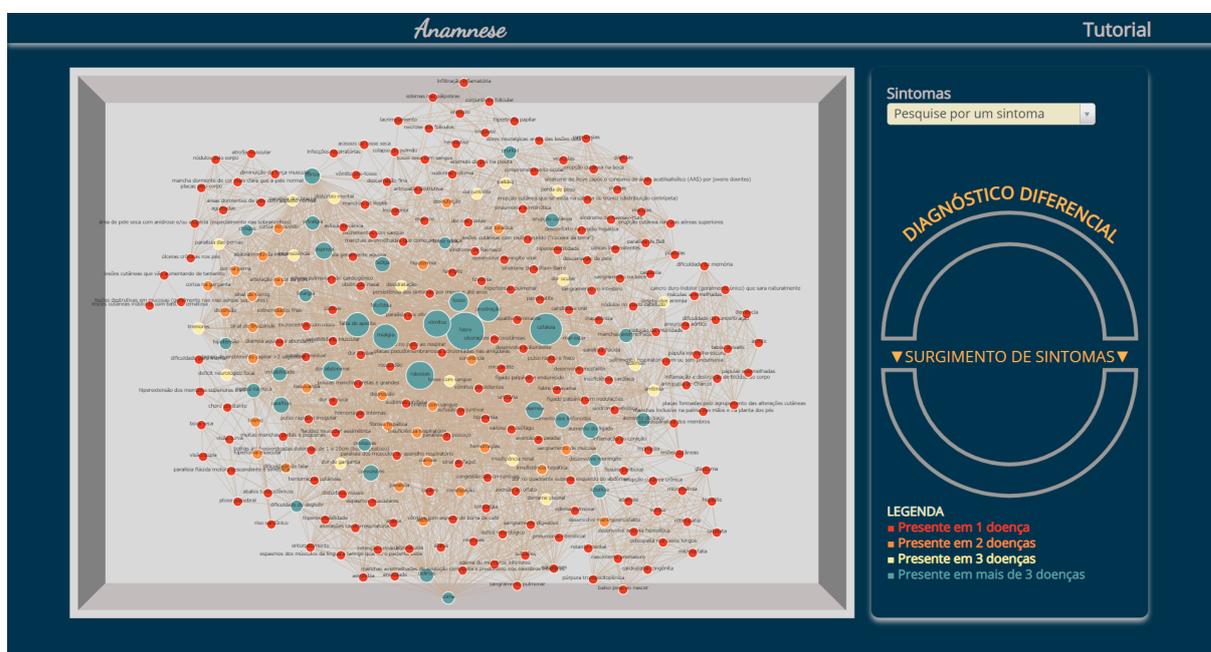


Figure 5.1 Layout do Sistema Anamnese. Fonte: Autoria própria.

O processo de diagnosticar uma doença não é tarefa fácil. Os sintomas iniciais geralmente são os mesmos, quem nunca passou pela tríade febre, dor de cabeça e dor no corpo? Conforme o quadro clínico vai evoluindo é que vão surgindo os sintomas característicos da doença, mas infelizmente, para alguns casos isso é tarde demais. É uma corrida contra o tempo. Por isso, a proposta do Sistema Anamnese (Figura 5.1) é que o médico possa estabelecer um diagnóstico diferencial o mais rápido possível. Para isso despreza os sintomas genéricos (hubs), buscando características específicas.

Observou-se nos aplicativos mencionados (ver seção 3.3) que a maioria fornece um ranqueamento de doenças, mas a amostra do Anamnese é norteada apenas pelo surgimento de sintomas.

5.1 Metodologia

Para construir o dicionário de doenças (ver cap. 4) foi necessário uniformizar os termos usados para definir os sintomas. Desde as diferenças mais simples como singular e plural, para que um mesmo sintoma não fosse considerado como dois nós, assim como sinônimos dos termos técnicos e na medida do possível procurar um meio termo entre um nome técnico e um mais popular, para que o sistema fosse acessível também a estudantes. Por exemplo, hepatomegalia foi substituído por aumento do fígado, porque o termo é mais difícil de lembrar, mas ataxia foi mantido. O dicionário resultou em 41 doenças e 273 sintomas.

Montado o dicionário, o passo seguinte foi criar as arestas e analisar a rede. Para isso, dado que os nós são os sintomas, foi considerado que uma aresta existe entre todos os sintomas de uma doença, ou seja, os sintomas seguem um ao outro. Por causa disso, não faz sentido considerar a quantidade de ligações que o nó possui (grau), assim foi considerado um grau modificado, que é a quantidade de doenças que o sintoma aparece. Se dois sintomas ocorrem em doenças diferentes a aresta se repete. As repetições foram consideradas como peso das arestas. Estes pesos foram usados apenas para tentar otimizar o layout da rede. A rede resultou em 3.800 arestas.

Quanto à análise da rede, foi estudada a frequência das relações. Os sintomas que ocorrem em apenas uma doença são 189 e os links únicos mais de 75% e, acrescento-se ao fato de que 37 doenças tem pelo menos um sintoma único, isso garantiu que o diagnóstico seria possível por unicidade.

Os nós da rede tem tamanho proporcional a quantidade de doenças (grau modificado) e as cores destacam os nós associados a no máximo três doenças. Assim, com a experiência o usuário poderá obter os resultados do sistema rapidamente. O diagnóstico funciona através da redução da sub-rede de sintomas que o paciente apresenta, considerando-se a interseção entre as doenças dos respectivos sintomas.

Para utilizar o sistema, o usuário apenas precisa visualizar todos os elementos dos conjuntos de doenças em que o sintoma se manifesta, logo essa quantidade é relativamente pequena. Por isso, a técnica escolhida para a visualização foi uma rede complexa, porque esta pertence a categoria das técnicas Node-link diagrams (como se pode ver na Figura 2.10 que nada mais é do que uma rede complexa), e essas técnicas geralmente são fáceis de entender para uma quantidade pequena de nós (ver seção 2.2.3). Outro fator determinante, foi a familiaridade que os profissionais de saúde têm com a estrutura de dados Árvore, utilizadas no estudo de Genealogia - Doenças Hereditárias, Genética (MYHERITAGE, 2010). Como árvores são um tipo de grafo (WIKIPÉDIA, 2018b) e redes também são grafos (ver seção 2.1), os conceitos básicos são os mesmos e, portanto, do conhecimento desses profissionais. Desta forma, o sistema pode ser facilmente entendido por esses usuários, pois para utilizá-lo é necessário conhecer o mínimo, ou seja, o que é um nó e qual o relacionamento que as arestas representam. E depois, com a experiência, ele poderá usar o grau modificado para otimizar a busca.

Quanto ao layout da rede do sistema (rede grande com 273 sintomas), sabe-se que a desordem visual é um dos principais problemas em grafos grandes e, também, que tentar encontrar um layout ideal para reduzir essa desordem não é prático, pois este problema de otimização geralmente é difícil para tais grafos (CUI; QU, 2007). Então, simplesmente, foi escolhido o layout disponível na biblioteca de visualização d3.js, que seria a utilizada, a saber, o Force Directed Graph, através da chamada `d3.forceSimulation()`. No entanto, como o sistema opera com redes egocêntricas (ver seção 5.2), representar apenas as arestas entre o primeiro sintoma escolhido e os outros nós (ou seja, não mostrar as ligações das outras arestas com as demais), seria suficiente para tratar a desordem visual, visto que isso elimina o cruzamento de arestas, geralmente considerado a principal causa da desordem visual (CUI; QU, 2007).

Apesar da densidade da rede, a estrutura foi relativamente otimizada, apresentando alguns clusters bem definidos (Figura 5.1). Isso foi possível após estudar os valores default da simulação, a amplitude desses fatores e levar em conta o peso e o grau modificado. Consequentemente, os sintomas que tem coocorrência maior ficaram relativamente mais próximos.

5.2 Interface e Funcionalidades

A interface do Anamnese (Figuras 5.1 e 5.2) foi projetada de modo a criar empatia com os profissionais de saúde:

- O próprio nome do sistema: anamnese é uma entrevista que o médico faz, em que entre outras coisas, vai obter do paciente os sintomas que ele está sentindo.
- A fonte cursiva do logo: os médicos escrevem muito manualmente.
- A saída de doenças do arco superior para o inferior, simulando a redução do conjunto de possíveis doenças após o surgimento de sintomas.
- A visualização da rede complexa, estrutura que seria facilmente compreendida (ver seção 5.1).

No sistema Anamnese (Figura 5.2) o usuário pode:

- Consultar um breve tutorial.
- Dar zoom e, conseqüentemente, pan na rede.
- Arrastar nós (útil se houver algum label sobreposto).
- Resetar para a tela inicial.
- Navegar pelos sintomas, visualizando a sub-rede do sintoma (rede egocêntrica) e as doenças relacionadas no diagrama.

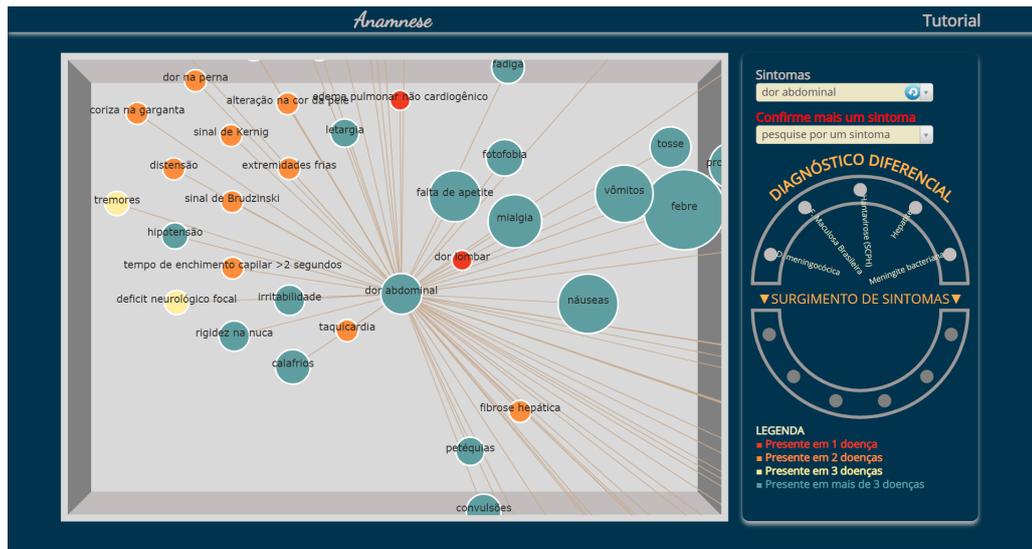


Figure 5.2 Interações no Sistema Anamnese. Fonte: Autoria própria.

- Pesquisar por um sintoma por qualquer termo contido na descrição do mesmo (não por ordem alfabética), isso ajuda no match da pesquisa. A pesquisa abre a mesma visualização anterior.
- Nos dois passos anteriores, se o sintoma escolhido tiver grau modificado maior que 10, então se abre um segundo select2, solicitando a pesquisa de um segundo sintoma. Funciona como um filtro para não se ter muitas doenças no diagrama (Esta abordagem ainda deve ser estudada, pois causou confusão desde os testes informais com não-especialistas).
- Obter o diagnóstico diferencial, confirmando sintomas até onde o profissional achar necessário. Para isso ele deve clicar no primeiro sintoma e confirmar os seguintes segurando Ctrl+clique, assim tanto a quantidade de possíveis doenças pode se reduzir quanto a sub-rede da associação dos sintomas confirmados.

5.3 Casos de Uso

A seguir, serão feitas algumas simulações no sistema, para mostrar a teoria do seu funcionamento (ver seção 5.1). Note que a ordem de seleção dos sintomas não importa e que alguns nós foram arrastados visando uma representação melhor.

5.3.1 Diagnóstico Diferencial Focado na Interseção

Neste caso, para simplificar a representação das interseções, foi escolhido um diagnóstico diferencial com apenas dois conjuntos de doenças e um deles contido no outro, a saber Dengue e Febre de Chikungunya. Para chegar nessa amostra basta selecionar os sintomas

"erupção cutânea" e "prurido". A relação entre os conjuntos foi desenhada para um melhor entendimento, a saber, o conjunto Dengue (círculo) está contido dentro do conjunto Febre de Chikungunya (elipse), conforme a Figura 5.3.

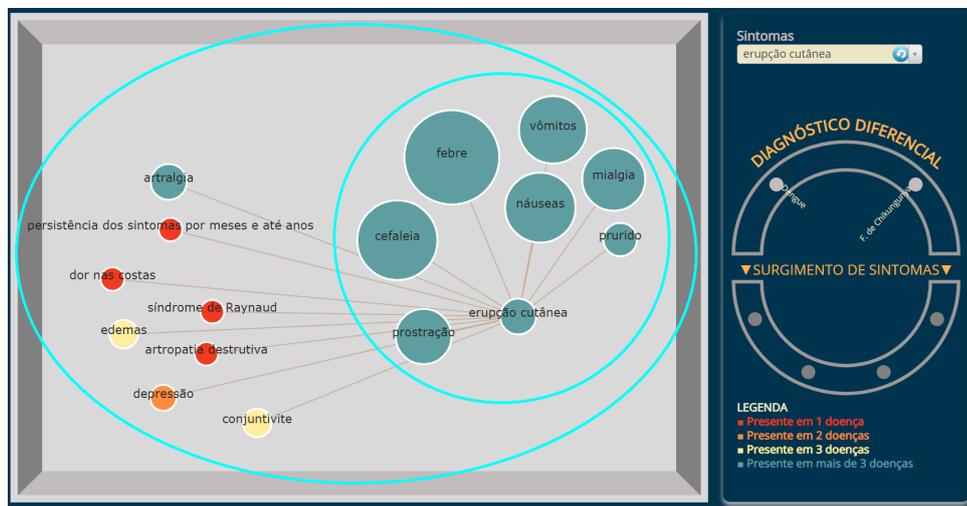


Figure 5.3 Diagnóstico diferencial contendo Dengue e Febre de Chikungunya. Fonte: Autoria própria.

Observe que, enquanto os sintomas manifestados estiverem dentro da área de interseção (círculo) as duas doenças são possíveis, mas surgindo um sintoma fora, independente da quantidade de doenças em que ele ocorre (grau modificado do novo nó), a doença só pode ser Febre de Chikungunya, ficando excluída a possibilidade de ser Dengue.

5.3.2 Diagnóstico Diferencial Focado na Unicidade

Este caso explora o "grau" da combinação de sintomas, ou seja, qual a quantidade de doenças relacionadas as duplas, triplas, quádruplas (e assim por diante) de sintomas. Por simplificação, foi escolhido o diagnóstico diferencial para uma dupla, a saber, "prurido" e "febre". O sintoma "prurido" esta relacionado a cinco doenças (Figura 5.4a) e, ao ser confirmado o sintoma "febre", ainda que este esteja relacionado com 35 doenças, é possível reduzir o diagnóstico diferencial porque só há três doenças em que aparecem "prurido" e "febre" juntos (Figura 5.4b).

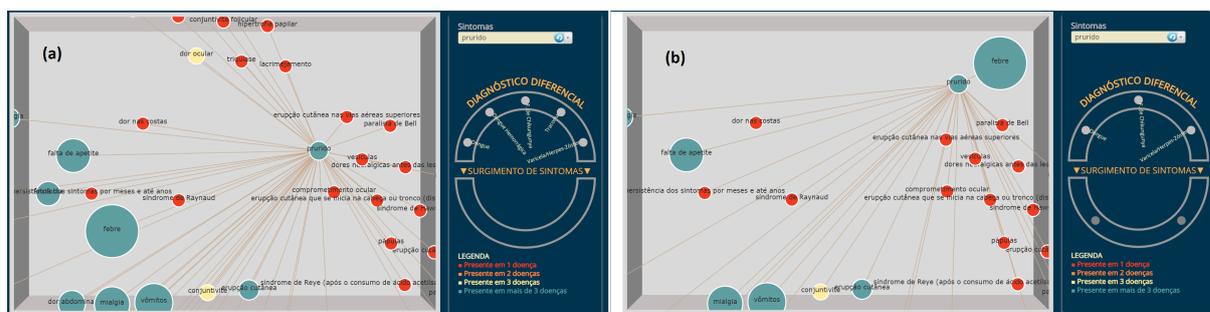


Figure 5.4 Diagnóstico diferencial a partir da seleção dos sintomas Prurido e Febre. Fonte: Autoria própria.

5.3.3 Diagnóstico Diferencial das Víruses

Este caso aborda a grande semelhança entre os sintomas iniciais de várias doenças, por isso o filtro de não apresentar muitas doenças no diagrama foi ignorado (ver seção 5.2). Para esta simulação foi escolhida a tríade "febre", "cefálea" e "mialgia", o que resultou em 15 doenças relacionadas, entre elas Dengue, as Meningites, Febre de Chikungunya, Hepatite, Influenza (Gripe), Leptospirose e Rubéola (Figura 5.5). Mesmo para este quadro, com base na legenda, note pela distribuição das cores, que há uma grande quantidade de sintomas únicos (presentes em apenas uma doença) que a partir do seu surgimento vão definir o diagnóstico, também há muitos sintomas de grau modificado 2 e 3, tudo isso, contando com duplas, triplas e etc. (ver seção 5.3.2) que podem reduzir essa amostra, fornecendo um diagnóstico diferencial com poucas doenças para investigar.

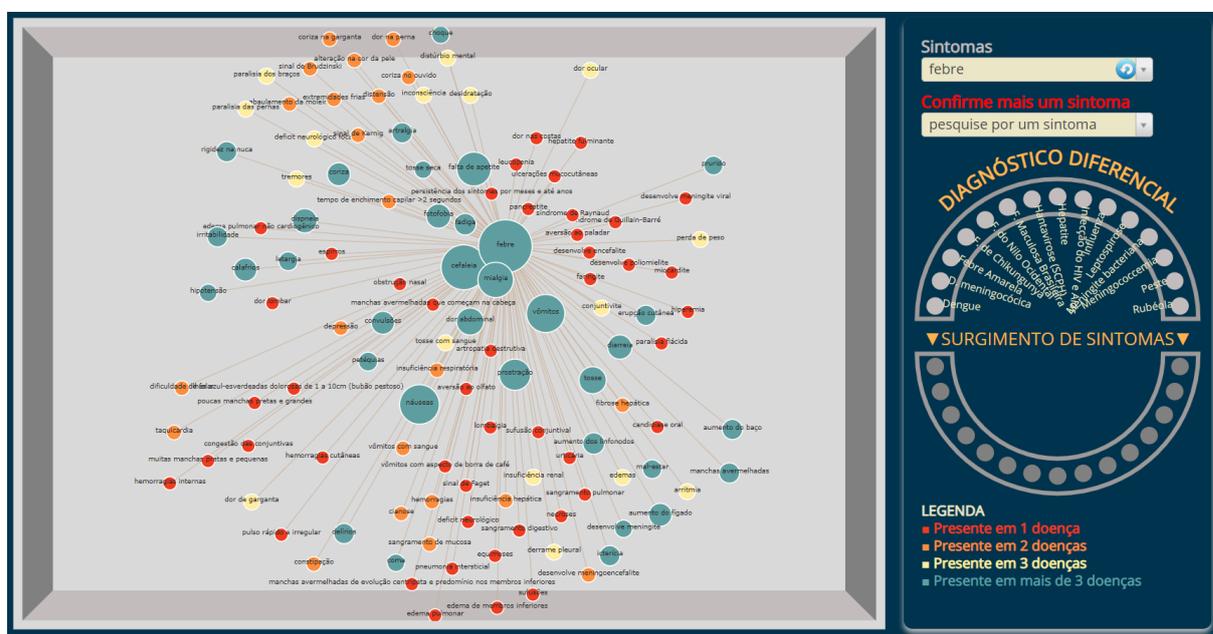


Figure 5.5 Tríade Febre, Cefálea e Mialgia, simulando o diagnóstico diferencial de Vírose. Fonte: Autoria própria.

5.4 Avaliação Inicial

Usabilidade "é a medida pela qual um produto pode ser usado por usuários específicos para alcançar objetivos específicos com efetividade, eficiência e satisfação em um contexto de uso específico" (ISO 9241-11:2018).

Segundo um estudo de Nielsen (2012; 2000), para identificar os principais problemas de usabilidade de um projeto, cinco usuários geralmente é suficiente, mas ele complementa dizendo que três usuários já representam o comportamento dentro do grupo. Então, a meta era fazer o teste de usabilidade com três médicos, mas só foi possível com uma médica. Segundo

Nielsen (2000), um único usuário de teste representa 1/3 das informações. Como o valor é representativo, o teste foi incluído no trabalho.

A técnica aplicada foi o Think Aloud protocol que consiste em observar o usuário enquanto ele executa algumas tarefas ao usar o sistema. Para obter mais informações também é necessário que ele fale suas impressões, ou seja, o que está fazendo, para onde está olhando, o que pretende fazer, se está confuso.

O Cenário das Tarefas consta no Apêndice A.1, a lista de doenças relatada são as do Guia de Vigilância em Saúde (BRASIL, 2016) do escopo do trabalho. A seguir o objetivo, resultado e considerações da execução de cada tarefa:

■ **Tarefa 1:**

Fazer o usuário ter uma visão geral. Para isso ele deveria ler a legenda que indica o "grau" dos nós, clicar num nó de "grau" 1 e encontrar o botão Reset. Como é o primeiro contato do usuário com o sistema, esperava-se a leitura do tutorial.

A médica pensou em um sintoma único e pesquisou (tendo resultado positivo), depois foi no reset.

Embora tenha conseguido executar a tarefa em segundos não seguiu o caminho esperado, o que dificultou um pouco a execução das tarefas seguintes. E como agravante não abriu o tutorial. Chegou a comentar que o reset estava muito escondido, que só tinha visto porque iria consultar outro sintoma.

■ **Tarefa 2:**

Fazer o usuário se familiarizar com a confirmação de sintomas (Ctrl+clique). Como não há nada mais genérico que virose, a seleção do primeiro sintoma retorna mais de cinco doenças e o usuário poderia ir diminuindo o resultado confirmando sintomas.

A médica tentou usar apenas o select2 para a seleção de todos os sintomas. Somente após ser informada que o sistema possuía ajuda foi que ela abriu o tutorial.

Devido ao uso do Ctrl+clique, essa tarefa só seria possível após a leitura do tutorial. Tendo lido o tutorial a médica conseguiu executar a tarefa e fez outras simulações, mas não conseguiu entender o segundo select2. Foi observado que a médica não percebeu que poderia clicar nos nós. A médica já havia usado outros aplicativos de diagnóstico nos quais é necessário digitar uma lista de sintomas e tentou fazer o mesmo. Talvez essa também tenha sido a causa de não ter buscado outra forma e percebido a possibilidade de usar a rede.

■ **Tarefa 3:**

A médica fez algumas simulações por conta própria após terminar a tarefa 2, sempre comentando os resultados.

O teste foi bastante produtivo pois a pessoa que usou o sistema era bastante observadora. Ela comentou que o Guia usado era muito restrito (no sentido de "bitolado") e por isso não encontrou algumas correspondências e perguntou se havia a intenção de aumentar o número de doenças. A mesma ficou satisfeita com os resultados, mas indicou a procura de um médico para completar a base de dados. Questionada se viu o tutorial, ela respondeu que sim, mas que preferia descobrir como usar. Esse comportamento foi observado desde os testes préliminares com não-especialistas.

Apesar da pesquisa de satisfação precisar de um número maior de pessoas, por ser um fator subjetivo, ela foi realizada para obter mais dados sobre as impressões que a médica teve em relação ao sistema. O Questionário escolhido foi o Attrakdiff (disponível em: <http://attrakdiff.de/>) porque havia o interesse de medir tanto a satisfação com o uso quanto a empatia com o sistema e ele atende a esses requisitos.

Segundo a Attrakdiff (2005), Qualidade Pragmática mede o sucesso em alcançar seus objetivos e Qualidade Hedônica descreve o valor global do produto baseado na percepção da qualidade, as duas são independentes e contribuem igualmente para o ranking da Atratividade (Figura 5.6).

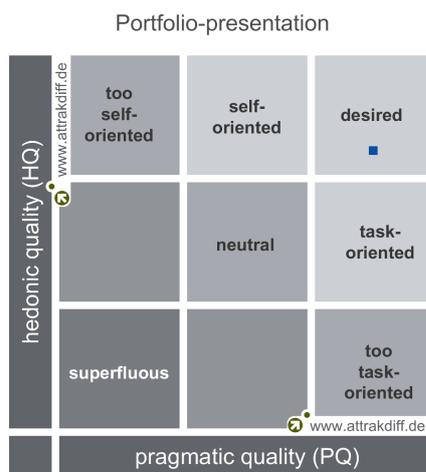


Figure 5.6 Parâmetros que geram Atratividade: Qualidade Pragmática e Qualidade Hedônica. Fonte: Attrakdiff, na conta do usuário.

O resultado do questionário de satisfação (Figura 5.7) para as respostas da médica (Figura 5.8) mostra que o sistema foi aprovado pela mesma, com Atratividade de 2.57 sendo próxima do valor máximo (a escala é de -3 a 3).

Note que o valor da Qualidade Hedônica foi dividido em HQ-I e HQ-S, que segundo a Attrakdiff (2005), significam:

- **HD-S (Estímulo):** funções, conteúdo e interações inovadores, interativos e estimulantes.
- **HD-I (Identificação):** indica em que medida o usuário se identifica.

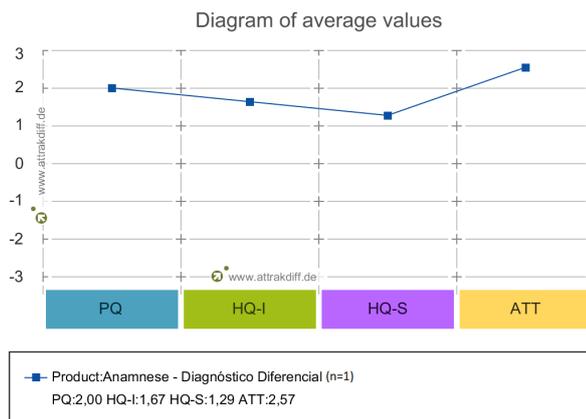


Figure 5.7 Média dos parâmetros de satisfação. Fonte: Attrakdiff, na conta do usuário.

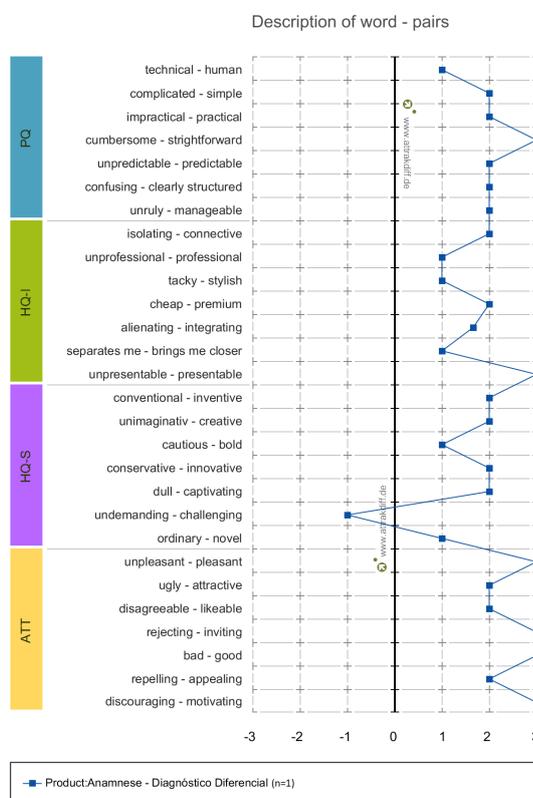


Figure 5.8 Respostas do Questionário de Satisfação. Fonte: Attrakdiff, na conta do usuário.

A única resposta negativa foi no par de palavras Pouco_exigente-Desafiador (undemanding-challenging) que integra o HD-S. Por estar ligado ao estímulo, esse par diz respeito ao quanto a pessoa se sente desafiada a fazer algo, mas uma possibilidade é de que a médica tenha interpretado no sentido de fácil-difícil, e como pelo resultado da Qualidade Pragmática, ela considerou o sistema fácil de usar, marcou o valor mais próximo da palavra Pouco exigente (undemanding), o que resultou no valor negativo. De todo modo, esse par Pouco_exigente-Desafiador (undemanding-challenging) é irrelevante para o sistema, porque a sua utilização não é motivada pelo desafio, mas pela necessidade de obter um diagnóstico diferencial significativo.

6

Conclusão

O Sistema Anamnese funciona através da Visualização de uma Rede Complexa que, além da interseção entre conjuntos de doenças, é otimizada por fatores de unicidade para fornecer o diagnóstico diferencial. Durante o projeto se conseguiu ultrapassar o desafio de representar as interações entre mais de três conjuntos, o resultado foi utilizar o sistema centrado no elemento de forma simples e intuitiva, assim o usuário interage com os elementos (sintomas) na visualização da rede complexa e os conjuntos (doenças) são tratados de forma secundária, com os seus valores surgindo apenas como resultado.

Como uma das causas do diagnóstico tardio é falta de capacitação profissional, o presente trabalho se preocupou não apenas em fornecer um sistema de diagnóstico diferencial para médicos, mas também uma forma de contribuição à educação continuada desses profissionais para a melhoria da qualidade dos serviços de saúde.

No sistema, o médico visualiza sintomas que só aparecem em uma doença, ou seja, o seu surgimento fecha o diagnóstico e, pesquisando por um único sintoma, tem disponível toda a sub-rede de sintomas relacionados. Assim, o profissional sem nenhum esforço extra, permanece em contato com todos os sintomas (mesmo que raros) das doenças e com doenças menos frequentes para sintomas habituais que também são retornadas no diagnóstico diferencial.

Dado que a qualidade em saúde abrange desde a prevenção de doenças até a manutenção da qualidade (CM Tecnologia, 2016), para abordar a base, a rede foi projetada com um meio termo entre sintomas técnicos e do senso comum, para que fosse acessível aos demais profissionais de saúde e, para que estes pudessem se manter informados sobre os sintomas e doenças mais frequentes, podendo assim auxiliar na prevenção de doenças e detecção precoce, orientando casos suspeitos a consultarem um médico.

Como trabalhos futuros a prioridade é concluir o teste de Usabilidade. Em seguida, avaliar a sua aplicação na comunicação remota entre profissionais de saúde (Telessaúde), com o intuito de agilizar o atendimento. Além disso, é necessário ampliar as funcionalidades do Sistema Anamnese, para cadastrar pacientes, assim será possível atualizar o diagnóstico diferencial conforme o surgimento de sintomas. Também seria interessante que o médico pudesse inserir doenças e sintomas, especialmente para suprir a falta de informações regionais.

Referências

- ADAW, Alzheimer's Dementia Alliance of Wisconsin. **Importance of an early diagnosis.** 200-. Disponível em: <<https://www.alzisc.org/Importance%20of%20an%20early%20diagnosis.htm>>. Acesso em: 28 nov. 2018.
- ALSALLAKH, B. et al. Visualizing sets and set-typed data: state-of-the-art and future challenges. **In: Eurographics conference on Visualization (EuroVis)–State of The Art Reports**, [S.l.], p.1–21, 2014.
- ALSALLAKH, B. et al. The State-of-the-Art of Set Visualization. **In: Computer Graphics Forum**, [S.l.], v.35, n.1, p.234–260, 2016.
- ATTRAKDIFF. **Evaluation Report for Product "Demo - B"**. 2005. Disponível em: <http://www.attrakdiff.de/files/demoproject_results.pdf>. Acesso em: 27 ago. 2018.
- BRASIL, Ministério da Saúde (MS). Coordenação-Geral de Desenvolvimento da Epidemiologia em Serviços. **Guia de Vigilância em Saúde**. Brasília: MS, 2016. 1 ed. atual, 773p. Disponível em: <http://bvsms.saude.gov.br/bvs/publicacoes/guia_vigilancia_saude_led_atual.pdf>. Acesso em: 20 ago. 2018.
- BRASIL, Ministério da Saúde (MS). Departamento de Vigilância das Doenças Transmissíveis. **Leptospirose: diagnóstico e manejo clínico**. Brasília: MS, 2014. Disponível em: <<http://portalarquivos.saude.gov.br/images/pdf/2014/dezembro/02/Miolo-manual-Leptospirose-17-9-2014.pdf>>. Acesso em: 22 ago. 2018.
- BRASIL, Ministério da Saúde (MS). Departamento de Vigilância das Doenças Transmissíveis. **Diagnóstico precoce aumenta a 80% chances de cura de câncer infantil**. Brasília: MS, 2017. Disponível em: <<http://www.brasil.gov.br/editoria/saude/2017/11/diagnostico-precoce-aumenta-a-80-chances-de-cura-de-cancer-infantil>>. Acesso em: 18 ago. 2018.
- BRASIL, Ministério da Saúde (MS). Departamento de Vigilância Epidemiológica. **Óbitos por Leptospirose. Brasil, Grandes Regiões e Unidades Federadas, 2000 a 2017**. Brasília: MS - Sistema de Informação de Agravos de Notificação, 2018. Disponível em: <<http://portalarquivos2.saude.gov.br/images/pdf/2018/marco/09/Leptospirose-2000-2017-obitos.pdf>>. Acesso em: 30 ago. 2018.
- BRASIL, Tribunal de Contas da União (TCU). **Política Nacional de Atenção Oncológica**. Brasília: TCU - Secretaria de Fiscalização e Avaliação de Programas de Governo, 2011. Disponível em: <<https://portal.tcu.gov.br/lumis/portal/file/fileDownload.jsp?inline=1&fileId=8A8182A14D6E85DD014D7327C1CB5497>>. Acesso em: 27 ago. 2018.
- CM Tecnologia. **As medidas essenciais da gestão de qualidade em saúde**. 2016. Disponível em: <<https://www.cmtecnologia.com.br/qualidade-saude>>. Acesso em: 25 ago. 2018.

CUI, W.; QU, H. A survey on graph visualization. **Computer Science Department, Hong Kong University of Science and Technology**, Kowloon, 2007.

DIÁRIO DE PERNAMBUCO. **Após suspeita de morte por leptospirose, saiba como identificar os sintomas da doença**. 2018. Disponível em:

<<http://www.diariodepernambuco.com.br/app/noticia/vida-urbana/2018/02/21/interna{ }vidaurbana,742506/apos-suspeita-de-morte-por-leptospirose-saiba-como-identificar-sintom.shtml>>. Acesso em: 30 ago. 2018.

DUARTE, N. C. et al. Global reconstruction of the human metabolic network based on genomic and bibliomic data. **Proceedings of the National Academy of Sciences**, [S.l.], v.104, n.6, p.1777–1782, 2007.

EADES, P. A heuristic for graph drawing. **Congressus numerantium**, [S.l.], v.42, p.149–160, 1984.

ESTADÃO. **40% da mortalidade de pacientes com aids está ligada a diagnóstico tardio**. 2011. Disponível em: <<https://www.estadao.com.br/noticias/geral,40-da-mortalidade-de-pacientes-com-aids-esta-ligada-a-diagnostico-tardio-imp-,679150>>. Acesso em: 27 ago. 2018.

GOH, K.-I. et al. The human disease network. **Proceedings of the National Academy of Sciences**, [S.l.], v.104, n.21, p.8685–8690, 2007.

HIMMELSTEIN, D. **Mining knowledge from MEDLINE articles and their indexed MeSH terms**. 2015. Disponível em: <<https://think-lab.github.io/d/67/>>. Acesso em: 13 ago. 2018.

HUCKA, M. et al. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. **Bioinformatics**, [S.l.], v.19, n.4, p.524–531, 2003.

ISO 9241-11. **Ergonomics of human-system interaction - Part 11: usability: definitions and concepts**. 2018.

LEX, A. et al. 201-. UpSet: Visualizing Intersecting Sets **In: Caleydo**. Disponível em: <<http://caleydo.org/tools/upset/>>. Acesso em: 20 nov. 2018.

LEX, A. et al. UpSet: visualization of intersecting sets. **IEEE transactions on visualization and computer graphics**, [S.l.], v.20, n.12, p.1983–1992, 2014.

LINDEN, R. Técnicas de agrupamento. **In: Revista de Sistemas de Informação da FSMA**, [S.l.], v.4, p.18–36, 2009.

METZ, J. et al. Redes Complexas: conceitos e aplicações. **In: Revista de Sistemas de Informação da FSMA**, São Carlos, n.4, 2007. ISSN 0103-2569.

MYHERITAGE. **A importância da genealogia médica**. 2010. Disponível em: <<http://blog.myheritage.com.br/2010/09/a-importancia-da-genealogia-medica/>>. Acesso em: 28 nov. 2018.

NIELSEN, J. **Why You Only Need to Test with 5 Users**. 2000.

NIELSEN, J. **Usability 101**: introduction to usability. 2012.

OMIM, Online Mendelian Inheritance in Man. **An Online Catalog of Human Genes and Genetic Disorders**. 1970-2018. Disponível em: <<https://www.omim.org/>>. Acesso em: 30 set. 2018.

PRUDÊNCIO, R. B. C. **Notas de aulas**. Recife: UFPE - Centro de informática (CIN).

RACIOCÍNIO CLÍNICO. **Aplicativo para diagnóstico diferencial**: escolhemos os 4 melhores! 201-. Disponível em: <<http://raciocinioclinico.com.br/blog/novidades/aplicativo-para-diagnostico-diferencial/>>. Acesso em: 25 nov. 2018.

RODRIGUES, F. A. **Caracterização, classificação e análise de redes complexas**. 2007. Tese (Doutorado em Ciência da Computação) — Instituto de Física de São Carlos.

RODRIGUES, M. M. et al. Diagnóstico tardio da sarcoidose é comum no Brasil. **In: J Bras Pneumol**, [S.l.], v.39(5):539-546, 2013.

SANTOS, D. C. M. et al. A hanseníase e o seu processo diagnóstico. **In: Hansen. Int.**, [S.l.], v.32(1):19-26, 2007.

THIELE, I. et al. A community-driven global reconstruction of human metabolism. **Nature biotechnology**, [S.l.], v.31, n.5, p.419, 2013.

TJCC, Movimento Todos Juntos Contra o Câncer. **Declaração para Melhoria da Atenção ao Câncer no Brasil**. 2014. Disponível em: <<https://todosjuntoscontraocancer.com.br/downloads/declaracaoTJCC.pdf>>. Acesso em: 30 ago. 2018.

UOL. **Sites e apps ouvem seus sintomas e falam sobre sua doença. Pode, Arnaldo?** 2017. Disponível em: <<https://noticias.uol.com.br/saude/ultimas-noticias/redacao/2017/03/31/sites-e-apps-ouvem-seus-sintomas-e-falam-sobre-sua-doenca-pode-arnaldo.htm>>. Acesso em: 25 nov. 2018.

WIKIPÉDIA, a enciclopédia livre. **Diagrama de Venn**. Flórida: Wikimedia Foundation, 2017. Disponível em: <https://pt.wikipedia.org/w/index.php?title=Diagrama_de_Venn&oldid=49779956>. Acesso em: 22 nov. 2018.

WIKIPÉDIA, a enciclopédia livre. **Diagrama de Euler**. Flórida: Wikimedia Foundation, 2018a. Disponível em: <https://pt.wikipedia.org/w/index.php?title=Diagrama_de_Euler&oldid=50936929>. Acesso em: 22 nov. 2018.

WIKIPÉDIA, a enciclopédia livre. **Árvore (grafo)**. Flórida: Wikimedia Foundation. 2018b. Disponível em: <[https://pt.wikipedia.org/w/index.php?title=%7BC3%7D81rvore_\(grafo\)&oldid=52304583](https://pt.wikipedia.org/w/index.php?title=%7BC3%7D81rvore_(grafo)&oldid=52304583)>. Acesso em: 28 nov. 2018.

A

Apêndice

A.1 Cenário das Tarefas

Teste de Usabilidade - Cenário das tarefas

Apresentação

Obrigada por ter vindo hoje, sua participação é muito importante. Este sistema é parte de um Trabalho de Conclusão de Curso (TCC) de Ciência da Computação na área de Saúde.

Estamos testando o quanto é fácil ou difícil chegar nos resultados fornecidos pelo Anamnese, que aborda os sintomas apresentados pelo paciente e as possíveis doenças. O protótipo funciona com a lista de doenças do Anexo I.

Papel do Moderador de Teste

Não poderei ajudar na utilização do sistema. Mas posso pedir algum esclarecimento sobre o que disse, tentava fazer ou esperava que acontecesse.

Papel do Participante

Durante a execução das tarefas é importante conhecer as suas impressões sobre o sistema. Então, por favor, diga tudo que estiver fazendo, para onde está olhando, o que pretende fazer, se está confuso.

Após as tarefas também lhe pedimos para responder um breve questionário de satisfação.

Descrição das Tarefas

- **Tarefa 1**
Selecione um sintoma que se manifesta em apenas uma doença e depois volte para a tela inicial.
- **Tarefa 2**
Simule um caso em que a maioria dos sintomas que o paciente apresenta é característico de virose. Para isso selecione sintomas até onde achar necessário. Note que, neste caso pode haver mais de uma doença relacionada.
- **Tarefa 3**
Pense numa doença e tente chegar no seu diagnóstico através da seleção de alguns de seus sintomas.