

TALK2ME

VISUALIZAÇÃO E ANÁLISE DE REGISTOS
PESSOAIS DE *INSTANT MESSAGING*

Victor Manuel Soares Gomes Silva
LICENCIADO

Dissertação para a obtenção do grau de Mestre em:
ARTE MULTIMÉDIA

Professor Auxiliar Convidado FBAUP Vítor Martins
ORIENTADOR

PORTO, Setembro de 2007

Agradecimentos

É com grande satisfação que expresso aqui o meu sincero agradecimento a todos aqueles que tornaram possível a realização deste trabalho.

Em primeiro lugar ao Professor Doutor Vítor Martins, orientador da presente dissertação, pelo empenho e disponibilidade demonstrados ao projecto.

À Adriana, Andreia, Eustáquio, Joana, José, Marco, Miguel, Raquel, Rita, Rui, Sérgio e Tiago pela sua “contribuição” nos registos analisados. À Maria João e à Cristina pela ajuda nas “galhas”, ao Pedro Faria pelo debugging e a todos os meus amigos que me acompanharam durante este processo.

Um muito especial agradecimento ao Luís Sarmento pelo acompanhamento integral de todo o processo, pelos aconselhamentos, sugestões, incentivos, raspanetes quer estivesse pessoalmente no Porto, Amesterdão ou Nova York via e-mail, MSN Messenger e Skype. Sem o seu especial contributo esta dissertação não teria sido possível.

Resumo

O uso de aplicações de mensagens instantâneas encontra-se bem disseminado hoje em dia. O volume de dados gerados aumenta de dia para dia mas, no entanto, as aplicações tradicionais não permitem uma consciência do modo em como comunicamos digitalmente ou como as relações se desenvolvem nesse meio. Numa era em que cada vez mais estamos conscientes de nós próprios como bases de dados, identificados por números de segurança social, fiscal, bancário, etc., é crítico produzir meios que ajudem a reflectir sobre a informação digital que produzimos. As mensagens instantâneas são o reflexo das interacções pessoais digitais com a nossa lista de contactos. Ao analisar informação temporal, de início e término de conversa, comprimento e conteúdo da mensagem, esta dissertação é uma reflexão pessoal acerca do uso das mensagens instantâneas, da caracterização e evolução das relações pessoais, do design das formas de visualização de tais sistemas e do possível conteúdo artístico que se pode retirar a partir de tais reflexões.

Abstract

The use of instant messaging applications is well disseminated nowadays. The volume of data generated increases day by day. However, the traditional applications do not allow to become aware of the way in which they communicate digitally, or how their personal relations evolve in this environment. In an time when we are increasingly aware of ourselves as data bases, identified by social security numbers, IRS, bank accounts numbers, etc., it is essential to produce methods for helping us reflect about the digital information that we produce. Instant Messaging is the reflection of personal interactions with our contact list. By analyzing temporal information, from the beginning and to the end of the conversation, the message length and content, this dissertation is a personal reflection about the use of instant messaging, the characterization and evolution of personal relations, the design of visualization forms for such systems, and about the possible artistic content that we might retrieve of such a reflection.

Índice

1 - Introdução - p15

- 1.1 Motivação - p16
- 1.2 Questões a colocar - p19

2 - As mensagens instantâneas - p21

- 2.1 IM - Instant Messaging -p21
- 2.2 O Boom - p22
- 2.3 As alternativas - p23
- 2.4 Cooperação - p24
- 2.5 MSN Messenger - p25
 - 2.5.1 - Revisões da aplicação cliente e do protocolo - p27
- 2.6 Os registos (histórico) do MSN Messenger - p28
- 2.7 Fora do circuito comercial - p30

3 State of the art- p31

- 3.1 Thread Arcs - p32
- 3.2 PostHistory - p34
- 3.3 Themail - p36
- 3.4 Growing Message Post - GMP - p38
- 3.5 ChatCircles - p40
- 3.6 CrystalChat - p42
- 3.7 Quadro Comparativo - p45
- 3.8 A Importância de um caso de estudo - p45

4 Análise Técnica - p47

- 4.1 Os ficheiros de Histórico - registo da comunicação - p47
- 4.2 O processo e fases - p50
- 4.3 Recolha de dados - p51
- 4.4 Identificação de conversas e contactos - p51
- 4.5 Estatísticas básicas, contagem de palavras - p53
- 4.6 Visualização Gráfica - p54
- 4.7 Os gráficos - p56
- 4.8 *Wordclouds* - p58
- 4.9 Navegação entre mensagens - p60
- 4.10 Cor - p61
- 4.11 caminhando para os resultados - p63

5 Resultados do caso de estudo - p65

- 5.1 Classificação inicial dos contactos - p66
- 5.2 Divisão temporal do período analisado -p67
- 5.3 Tabela de resultados obtidos - p68
- 5.4 Totais - p69
 - 5.4.1 Totais de conversas, mensagens e palavras por utilizador - p69
 - 5.4.2 Totais de conversas, mensagens e palavras recebidas - p70
 - 5.4.3 Totais de conversas, mensagens e palavras enviadas - p71
- 5.5 Médias - p72
 - 5.5.1 Médias: mensagens por conversa e palavras por mensagem - p72
 - 5.5.2 Médias: mensagens por conversa e palavras por mensagem recebidas - p73
 - 5.5.3 Médias: mensagens por conversa e palavras por mensagem enviadas - p74
- 5.6 Adriana - p75
 - 5.7.1 Gráfico comparativo entre os diversos itens enviados, recebidos e total - p75
 - 5.7.2 Gráfico comparativo de médias - p75
 - 5.7.3 Distribuição temporal das mensagens - p76
 - 5.7.4 Word cloud - p76
- 5.7 Andreia - p77
 - 5.7.1 Gráfico comparativo entre os diversos itens enviados, recebidos e total - p77
 - 5.7.2 Gráfico comparativo de médias - p77
 - 5.7.3 Distribuição temporal das mensagens - p78
 - 5.7.4 Word cloud - p78
- 5.8 Eustáquio - p79
 - 5.8.1 Gráfico comparativo entre os diversos itens enviados, recebidos e total - p79
 - 5.8.2 Gráfico comparativo de médias - p79
 - 5.8.3 Distribuição temporal das mensagens - p80
 - 5.8.4 Word cloud - p80
- 5.9 Joana - p81
 - 5.9.1 Gráfico comparativo entre os diversos itens enviados, recebidos e total - p81
 - 5.9.2 Gráfico comparativo de médias - p81
 - 5.9.3 Distribuição temporal das mensagens - p82
 - 5.9.4 Word cloud - p82
- 5.10 José - p83
 - 5.10.1 Gráfico comparativo entre os diversos itens enviados, recebidos e total - p83
 - 5.10.2 Gráfico comparativo de médias - p83
 - 5.10.3 Distribuição temporal das mensagens - p84
 - 5.10.4 Word cloud - p84
- 5.11 Marco - p85
 - 5.11.1 Gráfico comparativo entre os diversos itens enviados, recebidos e total - p85
 - 5.11.2 Gráfico comparativo de médias - p85
 - 5.11.3 Distribuição temporal das mensagens - p86
 - 5.11.4 Word cloud - p86
- 5.12 Miguel - p87
 - 5.12.1 Gráfico comparativo entre os diversos itens enviados, recebidos e total - p87
 - 5.12.2 Gráfico comparativo de médias - p87
 - 5.12.3 Distribuição temporal das mensagens - p88
 - 5.12.4 Word cloud - p88

- 5.13 Raquel - p89
 - 5.13.1 Gráfico comparativo entre os diversos itens enviados, recebidos e total -p89
 - 5.13.2 Gráfico comparativo de médias - p89
 - 5.13.3 Distribuição temporal das mensagens - p90
 - 5.13.4 Word cloud - p90
- 5.14 Rita - p91
 - 5.14.1 Gráfico comparativo entre os diversos itens enviados, recebidos e total - p91
 - 5.14.2 Gráfico comparativo de médias -p91
 - 5.14.3 Distribuição temporal das mensagens - p92
 - 5.14.4 Word cloud - p92
- 5.15 Rui - p93
 - 5.15.1 Gráfico comparativo entre os diversos itens enviados, recebidos e total -p93
 - 5.15.2 Gráfico comparativo de médias - p93
 - 5.15.3 Distribuição temporal das mensagens - p94
 - 5.15.4 Word cloud -p94
- 5.16 Sérgio - p95
 - 5.16.1 Gráfico comparativo entre os diversos itens enviados, recebidos e total -p95
 - 5.16.2 Gráfico comparativo de médias - p95
 - 5.16.3 Distribuição temporal das mensagens - p96
 - 5.16.4 Word cloud - p96
- 5.17 Tiago - p97
 - 5.17.1 Gráfico comparativo entre os diversos itens enviados, recebidos e total - p97
 - 5.17.2 Gráfico comparativo de médias - p97
 - 5.17.3 Distribuição temporal das mensagens -p98
 - 5.17.4 Word cloud -p98
- 5.18 O passo seguinte - p99

6 Discussão e análise dos resultados - p101

- 6.1 Uma primeira Análise – os resultados gerais - p102
- 6.2 Recebido vs Enviado - Assimetrias Gerais - p104
- 6.3 As Médias - p105
- 6.4 Adriana - p106
- 6.5 Andreia - p107
- 6.6 Eustáquio - p108
- 6.7 Joana - p108
- 6.8 José - p109
- 6.9 Marco - p109
- 6.10 Miguel - p110
- 6.11 Raquel - p110
- 6.12 Rita - p111
- 6.13 Rui - p111
- 6.14 Sérgio - p112
- 6.15 Tiago - p113
- 6.16 Antes e Depois - p113
- 6.17 Assentando ideias - p114

7 Conclusões Finais - p115

- 7.1 As condicionantes da análise - p116
- 7.2 Possíveis melhorias técnicas ao processo - p117
- 7.3 Caminho a seguir - p118

Referências Bibliográficas - p121

Bibliografia Geral - p125

Anexos - p127

- me&Adriana - p129
- me&Andreia- p131
- me&Eustáquio - p133
- me&Joana - p135
- me&José - p137
- me&Marco - p139
- me&Miguel - p141
- me&Raquel - 143
- me&Rita - 145
- me&Rui - 147
- me&Sérgio - 149
- me&Tiago - 151

Índice de Imagens

Figura 2.1 - Janela de contactos do MSN Messenger

Figura 2.2 - Janela de Conversação do MSN Messenger

Figura 2.3 - Log em XML visualizado numa janela do Internet Explorer

Figura 2.4 – Tabela com resultados obtidos

Figura 3.1 – Representação de uma malha de e-mail pelo Thread Arcs

Figura 3.2 – Visualização de um thread arc em acção no ReMail.

Figura 3.3 – Interface do PostHistory

Figura 3.4 – Pormenor do interface tipo calendário do PostHistory

Figura 3.5 – Exemplo de uma visualização com o Themail

Figura 3.6 - Pormenor do Themail

Figura 3.7 - Instalação no Nickle Art Museum na Universidade de Calgary.

Figura 3.8 - Pormenor Growing Message Post

Figura. 3.9 – Interface Chat Circles

Figura 3.10 - Vista de topo de uma visualização de CrystalChat

Figura 3.11 – Vista em perspectiva de uma visualização de CrystalChat

Figura 3.12 - Vista lateral de uma visualização de CrystalChat

Figura 3.13 – Visualização das mensagens de uma conversa no CrystalChat

Figura 4.1 – Exemplo da estrutura de um ficheiro de registo de MSN Messenger

Figura 4.2 – Janela de Mensagem do MSN Messenger

Figura 4.3 – Esquema representativo das diferentes fases e processos.

Figura 4.4 - Trecho de código de Processing e o seu resultado visual

Figura 4.5 – Ben Fry Isometric Blocks

Figura 4.6 – Esboço inicial do modelo de gráfico a utilizar

Figura 4.7 – Modelo representativo das 3 dimensões analisadas

Figura 4.8 – Exemplo de uma representação temporal da troca de mensagens com um contacto

Figura 4.9 – Exemplo de uma wordcloud gerada em www.iconinteractive.com/tools/wordcloud/ a partir do endereço web da faculdade de Belas Artes da Universidade do Porto a 2 de Julho de 2007

Figura 4.10 - Visualização de todas as mensagens trocadas com 1 dos contactos

Figura 4.11 – Representação tridimensional do espaço de cor CIELab

Figura 4.12 – Exemplo de 1 corte transversal no espaço de cor CIELab

Figura 4.13 – Cores usadas para representação dos diferentes contactos analisados

1 Introdução

Desde a invenção da escrita que a lógica material dos sistemas de memória, tal como arquivos históricos e arquivos administrativos tem prevalecido. Estes arquivos são ordenados hierarquicamente, quer linearmente ou em grelha, e foram criados para controlo, quer dos itens registados quer das pessoas e processos que estes itens significam. Ao lado destes arquivos rígidos e estáveis sempre existiram arquivos flexíveis e instáveis ou aquilo que se pode falar de informação imaterial que seguia uma racionalidade diferente, a labiríntica, enevoada lógica da cultura oral, que é a cultura sem informação escrita. Explorar informação é uma tarefa difícil, que se complica ainda mais, à medida que o volume e a diversidade crescem.

Com a introdução das bases de dados digitais assistiu-se a uma mudança. O que costumava ser “material” nos sistemas de arquivo passou a ser bancos de informação imaterial. Ao contrário das formas clássicas de arquivamento, as recentes bases de dados digitais não precisam de ser ordenadas hierarquicamente e com o surgimento de novas tecnologias computacionais verificou-se uma explosão de informação, quer no que respeita ao seu armazenamento quer à sua produção e distribuição.

A informação digital pode ser considerada como material em bruto que reside numa realidade digital, desta forma transforma o computador num instrumento, numa infra-estrutura e num ambiente de design. Seguindo o conceito de Louis Henry Sullivan's de que a forma segue a função, podemos então imaginar um conceito similar de que a forma segue os dados. Não apenas no contexto de exploração formalista de princípios de data *mining*, mas também para detectar e esteticamente traduzir as estruturas inerentes escondidas dentro dos conjuntos de dados, enquanto se explora as qualidades das novas linguagens visuais que estas representam.

Arquivos e bases de dados digitais oferecem ao artista um veículo para comentar as práticas culturais e institucionais através de uma intervenção directa. Trabalhar com bases de dados digitais é a forma mais eficiente para realizar este tipo de trabalho. Esta é uma das arenas mais promissoras para o desenvolvimento de trabalho artístico. São os lugares não só para comentar e intervir mas também para conceptualizar alternativas para a prática artística.

É necessário exprimir os valores e sentimentos da cultura e da época. É claro que nem toda a produção artística consegue resistir ao tempo, mas algumas mantêm-se contemporâneas, pois tratam de valores e problemas atemporais e universais.

Informação: mensagens instantâneas, e-mail, cotações de bolsa, dados meteorológicos, listas telefónicas, códigos postais, IRS, ADN, censos, a própria gramática. Estes são apenas alguns exemplos de informação que fazem parte do nosso dia a dia. Visualizar tais volumes de informação só o é possível devido ao uso do computador. Se por um lado este permitiu a explosão de informação de que se falou anteriormente, tais níveis de informação gerados escapam ao domínio da cognição humana. É claro que poderíamos referir representações gráficas datadas do Sec. XII [1], mas foi o computador que tornou facilmente estas repre-

sentações da excepção para a regra pois com ele facilmente podemos: visualizar conjuntos de dados gigantescos e incomportáveis para a compreensão humana e “alimentar” estes gráficos com dados em tempo real; basear representações gráficas de dados na sua análise matemática usando uma variedade de métodos, desde as estatísticas clássicas à *data mining*; “mapear” um tipo de representação noutra (imagem em som, som em espaço 2D/3D etc).

1.1 Motivação

A comunicação *online* produzida hoje em dia (e-mail, mensagens instantâneas, fóruns, etc.) ultrapassa claramente a nossa capacidade de compreensão e introspecção daquilo que produzimos. A velocidade a que se gera estes dados é tal que não há tempo e/ou capacidade de pensar sobre eles. No entanto devido à natureza “crua” dos dados gerados, estes podem revelar-se uma importante ferramenta de introspecção. Ao contrário de um diário tradicional, em que resumimos e escrevemos o nosso ponto de vista, os dados gerados através da comunicação digital são o próprio acto de comunicação.

Existe uma necessidade efectiva de aumento da nossa capacidade de introspecção sobre os dados que geramos, pois com o aumento do volume e velocidade cria-se uma distância cada vez maior entre o “passado” e o “presente”. Há um claro aumento do número de pessoas com quem interagimos no dia-a-dia. A denominada (por alguns) sociedade de informação em que vivemos e o factor globalização a ela associado obriga-nos (quase que) a uma multiplicidade crescente de contactos com outros indivíduos a nível digital, especialmente verdade no uso do e-mail a nível profissional. A nossa capacidade de produzir grandes volumes de dados aumentou drasticamente mas a nossa capacidade de “pensar” sobre eles não, devido principalmente ao facto de estes estarem tão disseminados e integrados na nossa vida (se vivermos numa sociedade “informatizada”) que quase nem questionamos a sua existência.

Este é o grande potencial introspectivo que este tipo de dados nos pode fornecer, o acesso directo aos dados originais da comunicação, a possibilidade de análise sobre “informação” concreta e não sobre memórias distantes.

A comunicação *online* é cada vez mais uma componente importante da nossa vida diária. As pessoas “entram” *online* para procurar emprego, manter contacto com amigos e familiares, conduzir negócios, discutir política, falar acerca de *hobbies* e até procurar eventuais parceiros. A maioria destas interacções sociais deixa para trás algum tipo de registo: mensagens trocadas de e-mail, históricos de mensagens instantâneas, *posts* em *newsgroups*, entradas de *blog*, etc. Escondidos nestes arquivos crescentes de interacções existem padrões sociais úteis, que se mais facilmente perceptíveis podem melhorar significativamente as dinâmicas sociais do mundo *online*.

A Internet adoptou ambientes que suportam interacções sociais numa escala sem precedentes. Centenas, milhares de pessoas agregam-se em espaços públicos *online* para trocar ideias, questionar, e comentar os eventos da vida diária. Uma só pessoa pode facilmente

entrar em contacto com várias centenas por todo o mundo através do *e-mail*. Estas trocas públicas e privadas deixam atrás de si grandes quantidades de traços persistentes que são representativos das relações que as pessoas mantêm. No entanto estes traços são geralmente invisíveis e inusáveis para os utilizadores de hoje.

Existem essencialmente dois tipos de arquivos *online*: as colecções públicas de interacção social (tais como as encontradas nas comunidades *online*) e os arquivos de comunicação pessoal (tal como os históricos de mensagens instantâneas ou arquivos de *e-mail*). Os arquivos colectivos são diferentes dos pessoais em várias questões importantes. Nos espaços públicos *online*, os utilizadores geralmente interagem com várias pessoas que nunca viram, pessoas que nunca conheceram no mundo físico. A participação em conversas públicas pode variar de duas, até centenas de pessoas. A maioria dos novos utilizadores entram e saem sem sequer deixar algum tipo de marca persistente nas conversas comuns, enquanto outros ficam e tornam-se elementos chave nas suas comunidades. Na maioria das vezes, os participantes não estão familiarizados com a colecção de mensagens que foi trocada na comunidade na qual participam. Os arquivos pessoais de comunicação por outro lado tendem a ser mais familiares aos seus “criadores”. Visto que as comunidades *online* são geralmente formadas à volta de um conjunto específico de interesses comuns (política, *hobbies*, etc.) o arquivo pessoal de mensagens instantâneas vai provavelmente agregar as várias facetas da vida pessoal, desde mensagens relacionadas com trabalho, vida familiar às conversas com amigos. A estrutura das conversas nas mensagens instantâneas é também diferente da maioria das interacções públicas porque tende a ser mais diádica (1 para 1) que as conversas orientadas para o grupo das comunidades *online*. Finalmente, o simples facto de que as mensagens instantâneas são na sua maior parte privadas interfere no tipo de trocas presentes nos seus arquivos. Resumindo, o propósito social dos arquivos públicos e dos pessoais de conversas *online* é significativamente diferente.

No geral esta dissertação, preocupa-se com a extracção de “informação” a partir de grandes conjuntos de dados (arquivos de comunicação pessoal). Não é propriamente um problema novo, pois a noção de visualização de dados para melhor compreensão destes já tem uma extensa história [2]. A diferença reside no “quê”, “como” e “porquê” da visualização. A incidência é na visualização de arquivos de mensagens instantâneas a partir dos meus registos pessoais do MSN Messenger. Existe um paradoxo na comunicação online: a quantidade de dados sociais persistentes existentes *online* parece ser inversamente proporcional à quantidade de uso que as pessoas retiram deste tipo de dados. Parece que apesar de sermos capazes de manter cada vez mais registos (*logs*) detalhados das nossas actividades *online*, não temos a capacidade de recuperar informação destes arquivos de uma forma útil e compreensível. E contudo, é evidente que estes arquivos podem ser fontes importantes de informação acerca das pessoas que os criam e das suas experiências e interacções como seres sociais.

A grande maioria dos sistemas de visualização, mesmo aqueles que representam dados sociais, são desenvolvidos para que cientistas, analistas e outros peritos possam analisar dados de outros. Ao desenvolver um sistema que está direccionado ao criador destes arquivos esta dissertação explora o nosso conhecimento em como estas visualizações podem ser usadas e que tipo de impacto estas têm no seu utilizador.

As mensagens instantâneas tornaram-se num espaço digital de interacção diária para muitos. De tal forma que o seu uso pode ser considerado um vício em muitos casos. Tornaram-se num autêntico habitat, numa casa fora (dentro) de casa. São usadas para trocar ideias de trabalho, marcar compromissos, mas principalmente para simplesmente interagir com outros. São uma ferramenta de comunicação para ligar pessoas. Existe uma grande quantidade de informação pessoal nos arquivos destas “ferramentas”, se bem que não seja perceptível através das ferramentas de visualização existentes. Escondida atrás dos “de” e “paras” está uma representação da nossa rede social, das nossas relações pessoais, e principalmente uma representação da nossa memória digital.

A familiaridade de um indivíduo com os arquivos visualizados determina como este pode usar as visualizações. De uma certa maneira o projecto aqui descrito nesta dissertação visualiza um arquivo que é supostamente familiar: as minhas mensagens instantâneas pessoais. Apesar de ser uma visualização com o poder da “descoberta” (quando os arquivos são tão longos que já não fazem parte da memória pessoal), o objectivo é explorar/reavivar a nossa memória, de ser um diário de referência visual. É uma exploração (de um retrato e/ou um auto-retrato) de relações pessoais entre um sujeito (eu) e os seus contactos do MSN Messenger, numa tentativa de visualizar as diferentes características da rede social formada por estas interacções.

“Important as the data mapping new media projects are, they miss something else. While modern art tried to play the role of “data-epistemology,” thus entering in completion with science and mass media to explain to us the patterns behind all the data surrounding us, it also always played a more unique role: to show us other realities embedded in our own, to show us the ambiguity always present in our perception and experience, to show us what we normally don’t notice or don’t pay attention to. Traditional “representational” forms - literature, painting, photography, and cinema – played this role very well. For me, the real challenge of data art is not about how to map some abstract and impersonal data into something meaningful and beautiful – economists, graphic designers, and scientists are already doing this quite well. The more interesting and at the end maybe more important challenge is how to represent the personal subjective experience of a person living in a data society. If daily interaction with volumes of data and numerous messages is part of our new “data-subjectivity,” how can we represent this experience in new ways? How new media can represent the ambiguity, the otherness, the multi-dimensionality of our experience, going beyond already familiar and “normalized” modernist techniques of montage, surrealism, absurd, etc.? In short, rather than trying hard to pursue the anti-sublime ideal, data visualization artists should also not forget that art has the unique license to portray human subjectivity – including its fundamental new dimension of being “immersed in data.” “

Lev Manovich, The Anti-Sublime Ideal in Data Art - 2002

1.2 Questões a colocar

Quantificar simplesmente as actividades produzidas no MSN Messenger não basta. Para que esta investigação entre também no campo artístico é necessário uma visão (...e interesse) pessoal sobre o objecto a produzir para que este seja um reflexo de uma introspecção da minha “comunicação” digital.

Para tal existem três questões às quais devem ser procuradas repostas:

1. A minha percepção social (concepção) das relações pode ser transcrita (visível) nos registos das conversas do MSN Messenger? Será que a “classificação” (a ideia) que tenho das pessoas num contexto social físico se reflecte no contexto digital de uma forma de comunicação específica, como o MSN Messenger?
2. Qual a relevância do factor “procura” numa determinada relação? Quem procura quem? Quem inicia um maior número de conversas? É possível caracterizar relações “dependentes”? E será que estas são facilmente reconhecíveis?
3. Qual a influência no discurso (contexto / frequência / tamanho) que o conhecimento do registo e arquivo das mensagens trocadas pode ter (antes e depois)? A ideia de sermos observados é um factor de influência ou facilmente nos abstrairmos do conceito do *“Big Brother is watching you”*?

2 As mensagens instantâneas

As pessoas que estão fisicamente mais próximas, mais propensas estão a comunicar frequentemente. Este foi um dos factores decisivos para o desenvolvimento de tecnologia que artificialmente possa trazer as pessoas que estão separadas pela distância, a uma proximidade artificial. Consequentemente, o interesse também surge nos anos mais recentes para desenvolver ferramentas de comunicação em tempo real de modo a promover interações entre colegas de trabalho. Algumas destas tecnologias incluem os sistemas de mensagens instantâneas (*Instant Messaging*).

2.1 *Instant Messaging*

O *instant messaging* [3] necessita do uso de uma aplicação cliente para efectuar a ligação ao fornecedor do serviço. Difere do e-mail no sentido em que as conversações acontecem em tempo real. A maior parte dos serviços proporciona uma informação de presença, que indica o “estado” dos contactos (se estão *online*, qual o tipo de indicador de presença como estar ocupado, ao telefone, etc). A maior parte dos sistemas permitem ao utilizador usar uma mensagem de status, permitindo aos seus contactos saber quando o utilizador se encontra ocupado, disponível, ausente, entre outras. Isto permite que não haja uma imposição de resposta tornando a comunicação menos intrusiva como por exemplo a comunicação via telefone.

O *e-mail*, considerado por muitos uma das mais importantes tecnologias que alguma vez apareceu desde que os computadores estão ligados à Internet proporcionou a esta o seu rápido crescimento e desenvolvimento. Hoje em dia é o *instant messaging* (mensagens instantâneas) a tecnologia que faz da Internet o que ela é em termos de contacto social.

Nos primeiros dias da Internet, o *e-mail* e as mensagens instantâneas eram principalmente usadas como uma forma alternativa mais barata de comunicação. Isto era verdade entre comunicações de longa distância e, de repente, o mundo ficou bem mais pequeno. Era bem mais fácil e barato entrar em comunicação com colegas e amigos no estrangeiro. Antes da Internet, ou tínhamos que viajar para o destino ou fazer uma chamada internacional para poder comunicar rapidamente com alguém. Não é fácil ou económico manter uma comunicação constante nestas condições. A Internet criou novas possibilidades de comunicação e principalmente, meios mais rápidos de comunicação. Por transmitir electronicamente texto e vários outros tipos de ficheiros via *e-mail* (ou via serviços de mensagens instantâneas), o emissor pode transmitir a sua mensagem através do globo em poucos segundos. O *e-mail* e as mensagens instantâneas também não possuem a desvantagem ou limitação da verbosidade ou alto custo que restringe formas tradicionais de mensagens instantâneas como o telegrama.

Este tipo de programas de conversação directa começou nos dias da inocência da Internet. Nessa altura, apareceu sob a forma do Talk no UNIX [4]. Os primeiros que adoptaram este tipo de programa eram na sua maioria engenheiros, cientistas e académicos de universidades, que o usavam para comunicar e discutir o seu trabalho através da Internet.

Foi apenas quando um programa chamado ICQ [5] surgiu em Novembro de 1996 que as mensagens instantâneas explodiram para o circuito comercial. Abreviatura para “I Seek You” (eu procuro-te), o ICQ entrou de rompante na Internet e eventualmente cresceu para um tamanho ainda maior do que a própria loucura do IRC (Internet Relay Chat) [6]. Desde que o ICQ apareceu em cena, dois outros gigantes fornecedores de mensagens instantâneas surgiram - nomeadamente o MSN Messenger [7] e o Yahoo! Messenger [8]. Hoje em dia (à data da escrita da presente dissertação) competem entre si para ganhar controlo do mercado e estabelecerem-se como o standard dominante. Contudo apesar desta batalha feroz, não é fora de comum o facto de um utilizador ser membro das três grandes redes de mensagens instantâneas.

Ao fornecer capacidades de conversação em tempo real, muito ao estilo do IRC, as mensagens instantâneas deram aos utilizadores da Internet o equivalente do sistema telefónico, apesar do uso do texto e não do áudio. As mensagens instantâneas foram ainda mais flexíveis ao permitir aos utilizadores terem múltiplas conversas simultâneas e mais tarde incorporar capacidades de conferência.

As mensagens instantâneas oferecem interacções curtas e ricas em contexto em tempo real. Quase todos os sistemas de mensagens instantâneas são iniciados por apenas uma pessoa, em que essa pessoa pode iniciar uma conversa directa com outra. Alguns sistemas como por exemplo o AOL Instant Messenger [9] e o MSN Messenger suportam *chat* em conferência que permite vários contactos na mesma janela de conversa.

2.2 O Boom

Os sistemas de mensagens instantâneas começaram como um pequeno fenómeno e apenas cresceram em popularidade relativamente tarde. Mas quando o fizeram, foi como um big bang. Em 1997 toda a gente falava no e-mail e no IRC. Eram as aplicações mais “fixes” de então. Continuaram a sê-lo até que o ICQ fez com que as mensagens instantâneas se tornassem *user friendly*, pois o programa existente até então de conversação do UNIX requeria um nível de compreensão tecnológica que tornava-o inacessível a leigos tecnológicos.

E assim a popularidade do ICQ explodiu, usando o método extremamente eficaz do boca a boca para ganhar popularidade. O número de pessoas que subscreviam o serviço grátis cresceu exponencialmente. Em apenas poucos meses, milhões estavam a mandar mensagens de texto entre si usando o ICQ.

O ICQ actualmente, já ultrapassou os 200 milhões de membros (inscrições). Tanto o MSN como o Yahoo suportam o mesmo número de subscritores e continuam a aumentar. Estes números podem parecer exagerados, mas se pensarmos que não demora muito para que um utilizador cative um grupo de amigos para estas redes, e estes por sua vez façam o mesmo a outros. Outro facto é que muitos utilizadores usam vários serviços de *instant messaging* ao mesmo tempo. Isto permite que possam estar contactáveis em várias redes.

As mensagens instantâneas começaram como simples programas de *chat* com amigos em tempo real. Contudo, para competirem com os seus rivais, as empresas que disponibilizam este tipo de serviços continuam a adicionar novas características e serviços a cada nova versão dos seus programas. Funções de multi-conferência permitem que mais do que duas pessoas tenham uma conversa instantânea, muito ao estilo do IRC. Esta simples, mas útil característica é uma das razões para que algumas empresas usem programas de mensagens instantâneas no seu ambiente de trabalho. Outras características que foram adicionadas aos *messengers* básicos, incluem a capacidade para personalizar os *emoticons* (pequenas imagens vulgarmente conhecidas por *smiles* que expressam emoções), partilha de ficheiros e imagens, conversas áudio e mais recentemente o uso de *webcams* para transmissão de imagem.

Os programas de *instant messaging* tiveram impacto em muitas vidas da sociedade moderna digital. Para os seus utilizadores é quase impossível passar um dia sem usar uma aplicação de mensagens instantâneas. Muitos acabam por as usar para manter conversas regulares com amigos, familiares e colegas no dia a dia de forma económica.

2.3 As alternativas

Hoje em dia, os programas de *instant messaging* estão basicamente divididos em 4 redes, nomeadamente o ICQ, MSN Messenger, Yahoo! Messenger e o AOL IM. Devido às várias alternativas existentes de diferentes *instant messengers*, é inevitável que alguns dos nossos contactos estejam numa outra rede. Assim muitas vezes é preciso instalar vários clientes diferentes de mensagens instantâneas.

Felizmente existem alternativas a instalar quatro sistemas diferentes de. Programas multi-plataformas tal como o Trillian [10], GAIM [11], Miranda IM [12] ou o Jabber [13] permitem consolidar num só programa os contactos das quatro principais redes. A vantagem é evidente no propósito de poder usar apenas um programa e de juntar numa só janela os vários contactos das diferentes redes. As desvantagens prendem-se principalmente no uso de características nativas a cada programa diferente de mensagens instantâneas, tal como *emoticons* personalizáveis, conversas áudio, vídeo, etc.

Algumas das características como a multi-conferência podem não funcionar correctamente apesar de serem suportadas em multi-cliente. Mas é quando existe uma mudança em algum protocolo das redes, que existe o maior problema. É necessário uma nova versão do programa para poder lidar com esta mudança. Por isso se o cliente não é constantemente actualizado, será difícil de acompanhar o desenvolvimento das várias redes sem ter problemas de ligação cada vez que a Yahoo! ou a MSN façam um upgrade aos seus protocolos.

Definitivamente tal como o *e-mail*, as mensagens instantâneas estão para ficar. Não existe dúvida de que mudaram a forma em como comunicamos digitalmente. É difícil hoje em dia imaginar a vida sem mensagens instantâneas para os que são digitalmente dependentes.

2.4 Cooperação

Existiram várias tentativas de criar um standard unificado para os serviços de mensagens instantâneas tais como: o SIP (Session Initiation Protocol) e o SIMPLE (SIP for Instant Messaging and Presence Leveraging Extensions) da IETF, APEX (Application Exchange), Prim (Presence and Instant Messaging Protocol), o XMPP (Extensible Messaging and Presence Protocol) aberto e baseado em XML mais comumente conhecido como Jabber e o IMPS (instant Messaging and Presence Service) da Opean Mobile Alliance criado especificamente para dispositivos móveis. A maioria das tentativas de criar um standard unificado para os maiores fornecedores de mensagens instantâneas (AOL, Yahoo! e a Microsoft) falhou e cada uma destas empresas continua a usar o protocolo proprietário próprio. Contudo em Setembro de 2003 a Reuters surpreendeu o mundo das mensagens instantâneas ao assinar o primeiro acordo de conectividade entre fornecedores destes serviços em Setembro de 2003. Este acordo histórico permitiu aos utilizadores do AIM, ICQ e do MSN Messenger pudessem comunicar com o serviço Reuters Messaging [14] e vice-versa. Depois deste importante acordo entre redes, a Microsoft, a Yahoo! e a AOL chegaram a um acordo em que os utilizadores do Microsoft Live Communication Server também tivessem a possibilidade de comunicarem com utilizadores públicos de *instant messaging*. Este acordo estabeleceu de uma vez por todas o protocolo de inter-conectividade do mercado o SIP/SIMPLE. Em Outubro de 2005 a Microsoft e a Yahoo! anunciaram que até ao final de 2006 iriam desenvolver as suas respectivas aplicações de mensagens instantâneas livres para que pudessem comunicar directamente entre si usando o SIP/SIMPLE. Em Dezembro desse mesmo ano a AOL e o Google anunciaram uma parceria em que permitia aos utilizadores do Google Talk comunicar com as redes do AIM e do ICQ.

Existem duas maneiras de combinar os vários protocolos existentes. Uma das maneiras é combinar esses protocolos dentro da aplicação cliente. Exemplos incluem o iChat, Trillian, GAIM, Fire, Proteus, Miranda IM, Adium, Everybuddy, Ayttn, Kopete, Centeriq, BitlBee e o IMVITE. A outra é combinar os diferentes protocolos na aplicação do lado do servidor. Esta solução faz com que a tarefa de comunicar entre os diferentes serviços passe para o lado da aplicação do servidor. Assim as aplicações cliente não necessitam de se preocupar com os diferentes protocolos. Esta é a solução dos servidores Jabber/XMPP, no entanto uma vez que se trata de um formato e protocolo fechado o desenvolvimento de aplicações para este tipo de servidores sofre de muitas dificuldades.

Algumas redes efectuaram algumas mudanças ao longo do tempo para tentar impedir que as suas redes fossem usadas por clientes multi-redes. As aplicações “oficiais” de cada rede apresentam muitas vezes publicidade, e o uso de aplicações que não a oficial contornam este aspecto. Por exemplo, a aplicação Trillian teve que implementar vários *patches* para permitir aos seus utilizadores acederem às redes MSN, AOL e Yahoo depois destas sofrerem mudanças nos seus protocolos de acesso.

Um dos novos conceitos a surgir é o conceito de sistemas descentralizados usando tecnologia de sistemas *peer-to-peer*. Esta solução tenta fazer com que os sistemas de *instant messaging* sejam independentes de uma autoridade central.

2.5 MSN Messenger

O MSN Messenger é um serviço gratuito gerido pela MSN [7] que inclui um serviço de mensagens instantâneas além de outros. O termo “MSN Messenger” é algo ambíguo uma vez que a Microsoft usa-o para se referir a diversas partes da sua solução de mensagens instantâneas. Conversamos através da rede de MSN Messenger, a aplicação mais popular é o MSN Messenger (apesar da Microsoft disponibilizar de raiz no seu sistema operativo Windows XP uma versão muito semelhante chamada Windows Messenger) e o protocolo usado na rede MSN Messenger é o MSN Messenger Protocol.

A rede MSN Messenger é uma rede de mensagens instantâneas da qual a Microsoft é a gestora e proprietária e entrou *online* em Julho de 1999. A rede da Microsoft é uma das quatro maiores existentes sendo em final de 2005 a segunda maior em termos de utilizadores “únicos” atrás da AOL Instant Messenger.

A aplicação que instalamos nos nossos computadores é o programa cliente da rede. Com ela, podemos ligar-nos ao servidor MSN Messenger através da Internet e através deste trocamos informação com outras aplicações clientes. A maioria do tempo a aplicação cliente encontra-se a comunicar com a aplicação de servidor, enviando-lhe informação de estado. A Microsoft lançou dois clientes diferentes para a rede MSN Messenger como já referi. Todos oferecem uma consciência de informação mínima em tempo real, usualmente ao indicar a presença estimada pela actividade do teclado ou do rato.

A fig. 2.1 mostra o interface do MSN Messenger. O interface apresenta 10 amigos ligados, 4 deles com mensagem de (Ausente) e 1 com a mensagem (Ocupado) e 25 que se encontram *offline*.



Figura. 2.1 - Janela de contactos do MSN Messenger

Ao “clicar” num dos nossos contactos na lista, é criado uma janela de *chat* (fig. 2.2). Aqui, podemos trocar mensagens com o nosso contacto. Desta janela, podemos também convidar mais contactos para a mesma conversa. Em adição a texto (e uso de *emoticons*), a janela pode ser usada para enviar ficheiros, assim como mostrar um “avatar” (imagem pessoal) ou ter uma conversa áudio ou vídeo (características mais recentes).

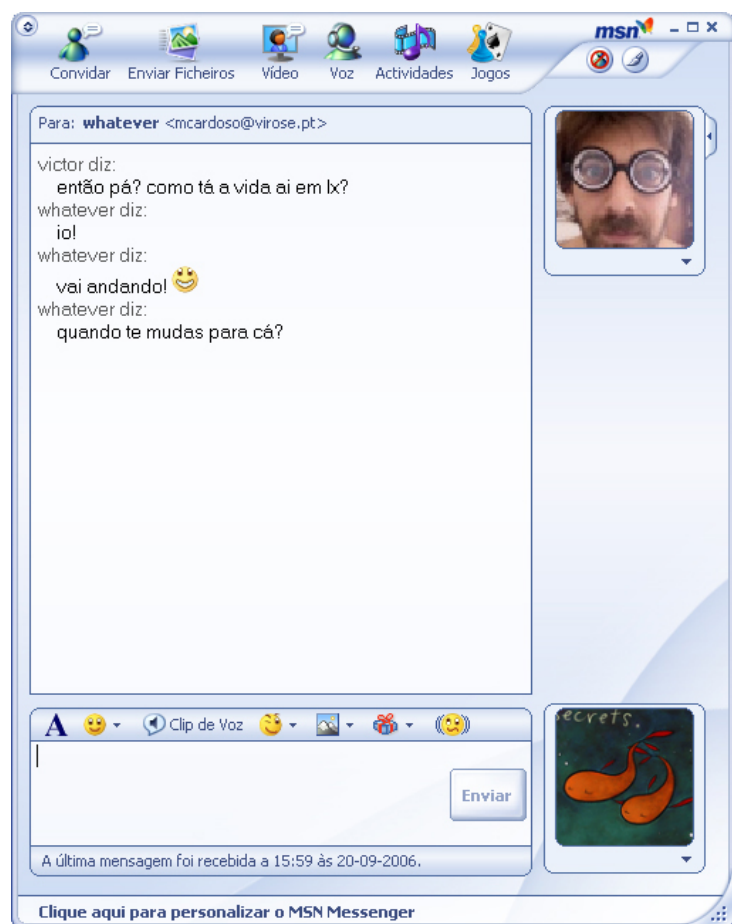


Figura 2.2 - Janela de Conversação do MSN Messenger

Estudos realizados [15], mostram que a maioria das pessoas usa as mensagens instantâneas como uma ferramenta para efectuar pequenas questões, para realizar simples clarificações, para combinar eventos sociais ou de trabalho e para cimentar relações pessoais. Dada a natureza espontânea, as mensagens instantâneas são muitas vezes usadas como o meio principal para descobrir a disponibilidade do contacto, ao consultar a lista de contactos ou usando uma mensagem curta tipo “podes falar?”. Isto é uma maneira rápida de localizar um “contacto” enquanto se minimiza a interrupção que uma chamada telefónica ou uma visita “*in loco*” pode causar, apesar do indicador de presença *online* de outros ser importante.

“As mensagens instantâneas permitem-nos a “negação” da presença do contacto, se ele (ou ela) não responder á mensagem enviada. Isto acontece porque os indicadores de presença das mensagens instantâneas são muitas vezes incorrectos, uma vez que medem a indicação de presença por uso do teclado/rato (se presumirmos que o indicador de presença que a pessoa decida usar seja o que realmente esteja a acontecer). Assim a pessoa que inicia a conversa está geralmente sem certeza que o recipiente da mensagem está ou não lá. Como resultado, a falha na resposta não é necessariamente interpretada como rude. Por outro lado, responder a uma mensagem requer muito pouco esforço relativamente ao requerido para receber e responder a um voicemail ou a um telefonema no momento certo. Em comparação, apenas cerca de 40% das chamadas de trabalho chegam ao recipiente certo à 1ª. As mensagens instantâneas também um registo temporário da conversa, desde que a janela de *chat* não seja fechada temos o registo das mensagens trocadas para nos ajudar a manter contexto da conversa.”

Nardi, *Interaction and outeraction: instant messaging in action*, 2000

2.5.1 Revisões da aplicação cliente e do protocolo

Como já referi o protocolo MSN Messenger, como qualquer outro protocolo de mensagens instantâneas consiste numa série de comandos enviados entre a aplicação cliente e a aplicação de servidor. Por exemplo, quando alguém da nossa lista de contactos fica *offline* o servidor envia uma mensagem a comunicar este facto às aplicações cliente do qual este contacto pertença. Ao receber esta mensagem, a aplicação cliente retira o contacto da listagem *online* para a listagem *offline*.

O protocolo MSN Messenger [16] sofreu várias revisões ao longo dos anos, sendo a mais recente a ser usada pelas versões 7.5 e 8 do MSN Messenger as versões de protocolo 11, 12 e 13. Estas são muitas vezes referidas como MSNP11, MSNP12 e MSNP13. Tal como o protocolo as aplicações clientes e de servidor também sofreram várias alterações e melhoramentos ao longo do tempo. Deixo aqui uma pequena linha temporal das alterações mais importantes feitas na aplicação cliente que é a aplicação usada para estudo nesta dissertação.

Julho de 1999 – Surge a primeira versão do MSN Messenger. Consiste num interface básico para troca de mensagens de texto online e uma simples lista de contactos.

Novembro de 1999 – Poucos meses depois de ser lançada a primeira versão a Microsoft lança a versão 2 do MSN Messenger que incluía um *banner* para uso comercial na caixa de listagem de contactos e a possibilidade de customizar o texto da mensagem em termos de cor e tipografia usada. Surge a possibilidade de uso de *emoticons*.

Mai de 2000 – A versão número 3 da aplicação é lançada. Oferece a possibilidade de partilha de ficheiros entre clientes assim como a possibilidade de uso de um sistema VOIP (Voice Over Internet Protocol). A partir desta versão é possível usar uma imagem de identificação pessoal (avatar).

Outubro de 2001 – A versão 4.6 lançada pela Microsoft sai com a possibilidade de agregar contactos num só janela (tipo *chatrooms*) e melhoramentos no seu serviço de conversação áudio (VOIP).

Julho de 2003 – Com o lançamento da versão 6 do MSN Messenger a velhinha caixa de mensagens que ainda resistia desde a versão 3 pode ser agora personalizada com diferentes tipos de fundos, *emoticons* e avatars personalizados. Uso de *webcam* também foi facultado nesta versão.

Agosto de 2005 – A versão 7.5 do MSN Messenger apresenta novos serviços como a Instant Voice Transporting, em que podemos gravar pequenos trechos de áudio para envio, uso de *emoticons* animados e *winks* e grandes melhoramentos nos serviços de vídeo via *webcam*.

2.6 Os registos (histórico) do MSN Messenger

O histórico de mensagens do MSN Messenger desenvolve-se da seguinte forma. As mensagens recebidas em tempo real e são guardadas num ficheiro de histórico que é ordenado consecutivamente de acordo de quando a mensagem de texto foi recebida e escrita. Estes ficheiros log são formatados como uma lista de XML e contêm os dados acerca da data da mensagem (de envio), da pessoa que a escreveu, da pessoa que a recebeu e o seu conteúdo. Existe um ficheiro como este para cada contacto com quem comunicamos. Este formato de dados pode revelar os padrões de chat entre o utilizador principal e os seus contactos mas a sua apresentação em modo texto dificulta este processo. Ao visualizar dois ficheiros separados através da janela de um browser (que consiga ler o ficheiro XML), torna-se difícil fazer a comparação com duas pessoas. A fig. 2.3 representa um extracto de um *log* em XML de uma conversa do MSN Messenger visualizada no Internet Explorer. Cada mensagem contém um timestamp com o tempo, dia, mês e ano. Cada mensagem também inclui um “de” e um “para” que revela o *nickname* que o utilizador está a usar no momento. A cor e o tipo de fonte usada na mensagem original também estão incluídas no ficheiro XML, mas não são apresentadas no browser.

O fim de cada mensagem individual é indicado quando uma pessoa carrega na tecla “enter” ou no botão de envio (fig. 2.2). O ficheiro XML também inclui informação de quando o utilizador fecha a aplicação cliente indicando assim o número de sessões.



Data	Hora	De	Para	Mensagem
16-06-2006	17:54:37	titi	victor	Bom tarde. Para podermos indicar o modelo exacto, vamos precisar do número de série do computador. Informo ainda que não estamos autorizados a vender peças de service - tentamos de ser nós a substituir.
16-06-2006	17:54:38	titi	victor	!
16-06-2006	17:54:58	victor	titi	????
16-06-2006	17:55:10	titi	victor	que não
16-06-2006	17:55:16	victor	titi	??
16-06-2006	17:55:31	titi	victor	a tempo do display q eu fodi com a bincadeira. P
16-06-2006	17:56:10	victor	titi	alinh
16-06-2006	17:56:11	victor	titi	ok
16-06-2006	17:56:13	titi	victor	no eBay anajo a parte de cima completa, com o LCD e tudo, por 45 euros. P
16-06-2006	17:56:35	titi	victor	mas o teclado é até anarã e n tenho cartão de crédito pra me registar no ebay.!
16-06-2006	17:56:57	victor	titi	;\$
16-06-2006	17:57:10	victor	titi	nibnet?
16-06-2006	17:57:34	titi	victor	é q nem preciso ter visa para depois pagar, basta o nibnet ou o paypal, mas exigem cartão para fazer o registo. @
16-06-2006	17:57:37	titi	victor	qm
16-06-2006	17:57:46	titi	victor	n aceita os dados dos cartões nibnet
16-06-2006	17:57:46	victor	titi	ok. \$

Figura - 2.3 log em XML visualizado numa janela do Internet Explorer

Ao visualizarmos directamente o ficheiro XML em bruto através do Internet Explorer, compreendemos que é praticamente impossível de distinguir variações na comunicação (troca de mensagens entre utilizadores). Para que seja possível analisar os padrões que possam existir nos registos, é necessário “mapear” os ficheiros de histórico do para uma plataforma que permita facilmente processar os dados guardados pelo MSN Messenger.

Não basta no entanto tabelar os dados (ver fig 2.4), é necessário construir ferramentas visuais que permitam identificar padrões e variações destes. Este assunto será aprofundado nos capítulos que se referem ao caso de estudo apresentado nesta dissertação (capítulos 4, 5, 6 e 7).

	Conversas iniciadas	Mensagens enviadas	Média de mensagens por conversa	Palavras usadas	Média de palavras por mensagem
Adriana	45	1434	15	4920	3.4
Victor	43	1258	14	4109	3.3
Total	89	2692	30	9029	3.4

	Conversas iniciadas	Mensagens enviadas	Média de mensagens por conversa	Palavras usadas	Média de palavras por mensagem
Andreia	20	671	15	2227	3.3
Victor	24	825	19	3685	4.5
Total	44	1496	34	5912	3.9

	Conversas iniciadas	Mensagens enviadas	Média de mensagens por conversa	Palavras usadas	Média de palavras por mensagem
Eustáquio	77	1300	9	5101	4.6
Victor	55	1807	13	4767	2.7
Total	132	3107	22	10568	3.2

	Conversas iniciadas	Mensagens enviadas	Média de mensagens por conversa	Palavras usadas	Média de palavras por mensagem
Joana	39	823	12	2794	3.4
Victor	31	730	10	3000	4.1
Total	70	1553	22	5794	3.7

	Conversas iniciadas	Mensagens enviadas	Média de mensagens por conversa	Palavras usadas	Média de palavras por mensagem
Jose	94	1593	12	8935	5.2
Victor	49	1705	12	5135	3.0
Total	143	3298	24	14070	4.1

	Conversas iniciadas	Mensagens enviadas	Média de mensagens por conversa	Palavras usadas	Média de palavras por mensagem
Marco	65	2530	11	11879	5.7
Victor	78	2940	20	9563	3.3
Total	144	5470	38	21442	2.9

	Conversas iniciadas	Mensagens enviadas	Média de mensagens por conversa	Palavras usadas	Média de palavras por mensagem
Isabel	13	196	8	757	3.9
Victor	12	200	8	702	3.5
Total	25	396	15	1459	3.7

	Conversas iniciadas	Mensagens enviadas	Média de mensagens por conversa	Palavras usadas	Média de palavras por mensagem
Isabel	0	110	12	1059	9.6
Victor	3	219	73	772	3.5
Total	3	329	110	1831	5.6

	Conversas iniciadas	Mensagens enviadas	Média de mensagens por conversa	Palavras usadas	Média de palavras por mensagem
Rita	2	85	21	224	2.6
Victor	2	86	21	229	2.6
Total	4	171	42	453	2.6

	Conversas iniciadas	Mensagens enviadas	Média de mensagens por conversa	Palavras usadas	Média de palavras por mensagem
Rui	20	358	11	1070	3.9
Victor	12	298	9	764	2.6
Total	32	656	20	1834	2.7

	Conversas iniciadas	Mensagens enviadas	Média de mensagens por conversa	Palavras usadas	Média de palavras por mensagem
Sérgio	64	972	12	3766	3.4
Victor	20	806	10	2151	2.7
Total	84	1778	22	5917	3.0

	Conversas iniciadas	Mensagens enviadas	Média de mensagens por conversa	Palavras usadas	Média de palavras por mensagem
Tiago	26	753	14	3336	4.4
Victor	28	945	10	1156	2.2
Total	54	1698	24	4492	3.4

Figura 2.4 – Tabela com resultados obtidos

2.7 Fora do circuito comercial

Até agora foi apresentado o que é a comunicação por mensagens instantâneas, as diferentes redes e os diferentes tipos de programas cliente que existem para usar mensagens instantâneas. Relatamos as suas origens, e como se disseminou até ser usada actualmente por milhões um pouco por todo o planeta. Com tal proliferação no uso desta tecnologia, é natural que o fenómeno seja alvo de análise e exploração pelo meio científico como pelo artístico. Aborda-se o conteúdo, (como se conversa, o que se conversa), questiona-se o espaço de *chat* e como nos relacionamos neste meio e como as conversas digitais pode funcionar como um espelho social de quem as usa.

No capítulo seguinte iremos abordar seis projectos que resultam de estudos sobre o uso do meio digital para troca de mensagens e conversas.

3 State of the Art

A visualização de conversas *online* é um assunto que já foi abordado de diversas perspectivas. Estas incluem o simples realce visual da conversa em si, a visualização da rede social formada através das actividades de *chat* e a representação do rasto deixado pelas actividades pela conversa ao longo do tempo (quantificação de palavras usadas, tempo de resposta, etc.). Neste contexto, a maioria dos trabalhos e investigações existentes reportam-se a três grandes grupos, o *e-mail*, os *newsgroups* e fóruns e finalmente o *chat online* em tempo real (mensagens instantâneas).

A grande maioria dos espaços de troca de mensagens são baseados em texto, apesar da inclusão de alguns símbolos gráficos (uns que representam cada pessoa, outros que representam sentimentos), e de os registos (*logs*) armazenarem também informação de tempo das actividades realizadas. É natural que as visualizações tomem o texto da mensagem como o principal elemento a observar, uma vez que este além de poder ser quantificado também pode ser também contextualizado.

Uma vez que esta dissertação se debruça principalmente na análise de interacções 1 para 1, (e principalmente pela caracterização do dono dos arquivos de mensagens) os projectos aqui apresentados pertencem principalmente aos grupos de análise de e-mail e de mensagens instantâneas. Excepção feita ao GMP (Growing Message Post) [17] que pela suas características de instalação artística é aqui incluído.

São seis os trabalhos (de que resultaram aplicações reais) aqui descritos e analisados. Em primeiro lugar iremos analisar visualizações de e-mail (malhas e conteúdos). Thread Arcs [18], PostHistory [19] e Themail [20] formam o grupo de projectos seleccionados para esta secção. O GMP é uma instalação que analisa mensagens deixadas na sua consola (ao estilo fórum/*newsgroup*). O ChatCircles [21] é uma aplicação/espço para troca de mensagens instantâneas, e finalmente o CrystalChat [22] que analisa históricos de mensagens instantâneas e que por este factor de grande proximidade ao assunto abordado nesta dissertação merece especial atenção.

Todos os projectos aqui descritos visualizam arquivos de interacções sociais. Isto significa que estes projectos lidam com indivíduos e as suas interacções com os espaços e as pessoas com quem estes entram em contacto. Não é uma visualização de fenómenos físicos, biológicos ou químicos. Em vez disso, são visualizações do tecido social da vida diária. São visualizações de eventos do dia a dia.

3.1 Thread Arcs

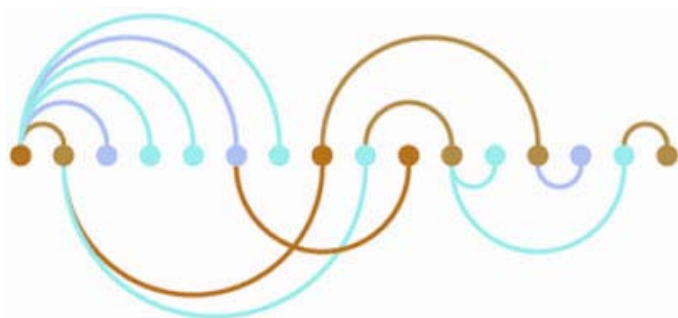


Figura 3.1 – Representação de um thread de e-mail pelo Thread Arcs

Thread Arcs [18] é uma visualização interactiva de malhas e-mail desenvolvida por Bernard Kerr no departamento de investigação da IBM em finais de 2003. Esta técnica de visualização representa as malhas encontradas no *e-mail* de maneira a ajudar os utilizadores a usar estas malhas. O Thread Arcs combina a cronologia das mensagens com a estrutura em árvore de uma malha de conversas num só modelo visual estável e compacto. Esta técnica de visualização foi desenvolvida para funcionar com o cliente protótipo de e-mail da IBM, o ReMail. O seu objectivo principal é ajudar na representação dos vários atributos das mensagens trocadas e encontrar facilmente mensagens específicas.

Uma das questões chave no desenvolvimento deste projecto prendia-se com o facto de as formas tradicionais de visualização de *e-mail* apenas o ordenaram numa ordem temporal. Assim, as mensagens de resposta a uma mensagem enviada eram representadas da mesma forma que mensagens que não eram de resposta. As vantagens de desenvolver uma representação visual destas malhas era clara: fornecer aos utilizadores uma maior contextualização na leitura das suas mensagens, memória do estado de uma determinada malha, ligação automática entre as mensagens que reduz o número de mensagens representadas na caixa de correio, o que por sua vez permite um melhor controlo dos utilizadores na selecção das mensagens quer seja para leitura ou para eliminação.

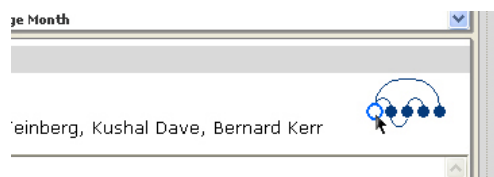


Figura 3.2 – Visualização de um *thread arc* em acção no ReMail.

A malha pode ser definida como uma colecção de mensagens individuais relacionadas umas com as outras pela função de resposta no *e-mail*. A primeira mensagem numa malha é a raiz (*root*), qualquer mensagem numa malha à qual é efectuada uma resposta é uma mensagem *parent* e as respostas a essa mensagem são denominadas de *children* dessa mensagem.

Foram consideradas seis atributos chave no desenvolvimento da visualização Thread Arcs: a cronologia das mensagens, as relações entre mensagens, a estabilidade, o factor de compactação visual, o atributo de selecção de um arco de mensagem e o factor escala.

O Thread Arcs usa um *layout* linear na representação das suas mensagens. Cada círculo representa uma mensagem da malha, e uma vez que a cronologia é um factor chave a sua posição representa a sua ordem de chegada (o começo da malha é representado pelo círculo mais à esquerda e os restantes círculos são ordenados cronologicamente). Cada arco representa a ligação da mensagem de resposta (*children*) à mensagem à qual está a responder (*parent*). O uso de diferentes cores nos arcos representa os diferentes utilizadores do *thread*. O uso de arcos em baixo e em cima deve-se ao facto de muitas vezes em malhas mais densas a leitura ser prejudicada com a acumulação de vários arcos. Assim, a solução foi usar arcos em cima e em baixo para clarificar melhor a representação.

Esta foi umas das primeiras soluções de visualização de malhas de *e-mail* para ajudar os utilizadores a terem uma melhor consciência e visualização das conexões existentes entre as suas mensagens de *e-mail* que de outra forma seriam invisíveis. Desde então, e principalmente com o surgimento do gmail [24] e a sua capacidade de agregar as malhas de mensagens que a importância de ordenamento visual das mensagens de *e-mail* tornou-se num factor crucial no desenvolvimento de clientes de *e-mail*.

3.2 PostHistory

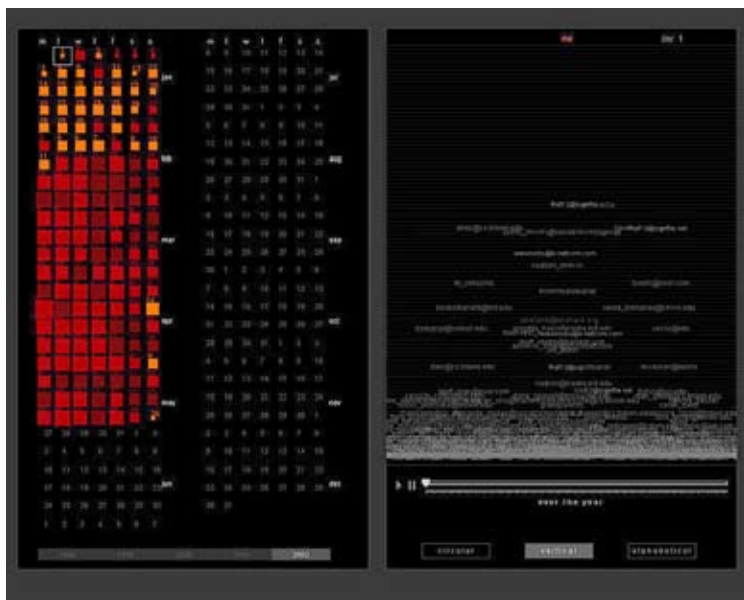


Figura 3.3 – Interface do PostHistory

O PostHistory [19] explora a noção de história no mundo digital do computador pessoal. Grande parte da nossa experiência do mundo tem a ver com a passagem do tempo e as mudanças que ocorrem enquanto este passa. Ficamos mais velhos, acumulamos coisas, os objectos com o tempo apresentam marcas de uso e de envelhecimento. No entanto no mundo virtual, o tempo não afecta os “objectos” da mesma maneira e não existe propriamente um sentido na sua acumulação. Estes “objectos” digitais ao contrário do mundo físico parecem eternos. O objecto que o PostHistory analisa é o e-mail. Lida-se com o *e-mail* a um ritmo diário e a maioria já o faz há alguns anos. No entanto é difícil perceber a acumulação desta actividade frenética, do número de mensagens enviadas e recebidas e das diferentes pessoas que contactamos. Em resumo, é difícil ter uma consciência da escala e dos padrões produzidos por esta actividade.

O objectivo do PostHistory é de proporcionar aos seus utilizadores uma nova experiência sobre as suas actividades de *e-mail*. Ao permitir a possibilidade de criar distância do uso rotineiro da leitura, resposta, envio e outras tarefas executadas diariamente no *e-mail*, o PostHistory oferece a oportunidade de reflexão, consciência e visualização sobre os padrões fundamentais da (inter)actividade do *e-mail*. Este projecto representa visualmente a acumulação diária, o tamanho e a escala dos arquivos de *e-mail*, com o objectivo de permitir olhar para trás nas acções efectuadas no mundo digital para assim entender a escala, intensidade e forma das interacções neste meio.

Todas as análises no PostHistory acontecem ao nível do cabeçalho da mensagem. Isto significa que o PostHistory não analisa o conteúdo do que está a ser falado nas mensagens de e-mail que está a analisar, só “sabe” que pessoas estão a falar entre si.

A aplicação diferencia os seguintes cabeçalhos: de, para, CC (*carbon copy*) e BCC (*blind carbon copy*). De maneira a visualizar os padrões de uso de uma maneira simples, o PostHistory têm que interpretar o que estes cabeçalhos significam em termos de construtores da rede social assim como estes integram-se na sua estrutura. A aplicação necessita de determinar:

O que é uma mensagem de *e-mail* “directa” e como se diferencia essa mensagem de uma menos directa.

Quem conhece quem numa rede de *e-mail*.

Como determinar e medir o peso de um contacto na rede de *e-mail*.

Como determinar a relação do utilizador com os diferentes contactos da sua rede de *e-mail*.

No PostHistory as visualizações descrevem as actividades do uso de *e-mail* numa base diária abordando questões como:

Quando é que surge um novo contacto no histórico de *e-mail*?

Quando é que um laço formado desvanece?

Quantas mensagens são-nos enviadas pessoalmente em oposição aquelas que são enviadas para uma lista de contactos?

Quais os laços mais fortes formados na rede? (frequência de contacto).

As visualizações no PostHistory obedecem à metáfora visual do calendário, uma vez que um dos seus objectivos é a visualização das mudanças ao longo do tempo.



Figura 3.4 – Pormenor do interface tipo calendário do PostHistory

O objectivo do PostHistory, é de criar uma ferramenta informativa que fornece uma visualização ao nível do utilizador das interacções sociais ao longo do tempo para que este possa reflectir sobre elas. Mas conseguiu ir além disto, pois o seu uso permitiu que funcione como um objecto de partilha de memórias como se de um álbum fotográfico se trata-se. Esta foi a principal razão do desenvolvimento do projecto seguinte, o Themail, um projecto que também analisa arquivos de *e-mail*, mas que incide essencialmente na questão da análise do conteúdo

3.3 Themail

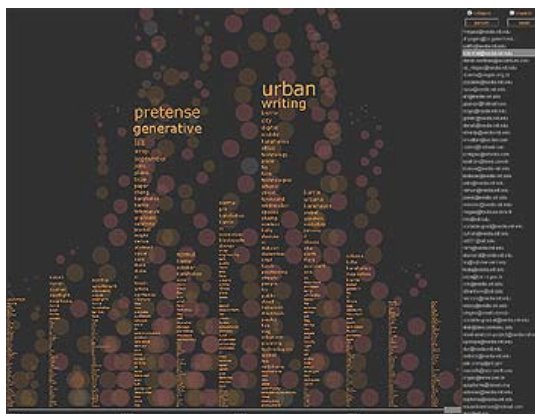


Figura 3.5 – Exemplo de uma visualização com o Themail

Se o Thread Arcs [18] foi desenvolvido com a análise e visualização de malhas em mente, o Themail [20] foi pensado para responder primariamente a duas perguntas:

1. Que tipo de coisas falamos com os contactos de *e-mail*?
2. Em como é que as conversas de *e-mail* com o sujeito X diferem das conversas com outras pessoas?

O Themail é uma ferramenta de visualização do conteúdo de arquivos de *e-mail* para ser usada pelo dono desses arquivos. A maioria dos utilizadores de *e-mail* tem a tendência de guardar grande parte das mensagens que recebem. O Themail usa essas mensagens (conteúdo das caixas de *e-mail*) para retratar as relações individuais entre os contactos com a premissa que essa visualização constitui um meio viável para retratar relações pessoais e de extrair algum significado sobre estas.

O Themail é uma visualização tipográfica do conteúdo de uma caixa de *e-mail* individual ao longo do tempo. O interface (fig. 3.5) mostra uma série de colunas de palavras-chave ordenadas ao longo de uma linha temporal. Os diferentes tamanhos e intensidade de cor representam a sua frequência e distinção ao longo do tempo. O Themail foi desenvolvido para uma exploração didáctica de relações, ao visualizar uma relação de cada vez entre o dono da caixa de *e-mail* e um dos seus contactos. Múltiplas camadas de informação são representadas usando várias técnicas diferentes para o *parsing* dos conteúdos. São representadas as palavras tópico anuais e mensais. As palavras anuais são representadas em último plano e as mensais em primeiro plano. Os contactos referidos na caixa de *e-mail* são listados à direita no interface, e ao selectivamente escolher cada um temos representada o conteúdo dos *e-mails* trocados com esse contacto. A linha temporal em baixo serve como filtro de conteúdo, permitindo ao utilizador ajustar a visualização, quer por estendendo a linha ao longo de todo o período analisado para obter uma perspectiva geral, ou focando em determinado período de tempo para obter uma visualização mais específica.

1. Themail Processor que lê os arquivos de *e-mail* (formato mbox) e a partir destes gera uma série de ficheiros de texto que contêm os dados que vão ser usados pelo programa de visualização.
2. Themail Visualization que vai ler os dados gerados pela aplicação de fundo e gerar a representação visual desses dados.

[illegible]

37

3.4 Growing Message Post - GMP

Growing Message Post [17] é um projecto que aborda a visualização dos aspectos temporais no envio de mensagens digitais. O GMP foi criado para visualizar mensagens de texto no ambiente de fóruns de discussão. Os autores referem que as mensagens de um fórum assemelham-se ao tipo de escrita das paredes de casa de banho. Um utilizador escreve uma mensagem, essa mensagem pode ser ou não obter uma resposta de outros utilizadores que por sua vez podem iniciar outros tópicos.

O GMP resultou numa instalação (fig.3.7) exposta no Nickle Art Museum na Universidade de Calgary. Permitia a diferentes indivíduos enviarem interactivamente mensagens (uma de cada vez), e de poderem responder a uma mensagem anterior.



Figura 3.7 - Instalação no Nickle Art Museum na Universidade de Calgary

Este acto contínuo eventualmente tornou-se numa conversa de grandes proporções. O grande objectivo deste projecto consiste na visualização destas conversas ao longo do tempo ao produzir padrões inspirados na natureza que tentam ser apelativos artisticamente e transmitir a informação histórica existente. Este tipo de padrões inspirados pela natureza também é usado pelo GMP para transmitir a complexidade das palavras e da leitura. Duas motivações chave foram tidas em conta na elaboração deste projecto:

1. A exploração da visualização de um diálogo baseado em tempo/texto na tentativa de tornar as conversas baseadas em texto mas graficamente informativas.
2. Um profundo interesse em como as metáforas baseadas na natureza, tal como o crescimento e deterioração das plantas pode ser usado para visualizar e expressar aspectos temporais do diálogo humano.

A abordagem usada por este projecto foi a de criar uma visualização que “mapeasse” os dados temporais de uma mensagem numa representação metafórica de uma planta. A natureza de uma planta foi usada devido ao seu carácter visível de crescimento e deterioração ao longo do tempo.

Os dados do GMP originam do processo contínuo de escrita de mensagens numa caixa de texto no fundo do ecrã. Eventualmente o acumular de mensagens dá forma à conversa. No entanto, existe sempre a possibilidade como já foi referido de iniciar outra conversa (novo tópico).

As mensagens baseadas em texto numa conversa, contêm componentes de tempo que são contemplados na visualização do GMP, tais como:

Quem se encontra correntemente a interagir com a peça
Quem interagiu anteriormente com a peça
O período de tempo com que se interage com a peça
O numero de vezes que se interagiu com a peça

Além deste componentes de tempo foram considerados factores de legibilidade da mensagem usando testes Flesch [23], com escala entre 1-100, sendo 100 a mensagem mais facilmente lida e o nível de grau da mensagem, com escala de 1-12 com 12 a ser o grau mais complexo.

Diferentes tipos de pistas visuais são usadas para representar quer os componentes de tempo quer os de complexidade. A metáfora baseada na natureza usada é a de uma planta estilizada. As folhas são “desenhadas” como um círculo para representar uma palavra e o seu diâmetro depende do comprimento dessa mesma palavra. As hastes representam a mensagem e o seu comprimento está relacionado com o comprimento da mensagem. O uso da cor no GMP está relacionado mais uma vez com a metáfora da natureza. As cores das folhas são inicialmente verdes e com o passar do tempo amarelecem, tal como o efeito de passagem de estações. (Fig.3.8) Com o tempo algumas das palavras caem tal como as folhas de uma árvore.



Figura 3.8 - Pormenor Growing Message Post

As conversas no mundo físico estão cheias de pistas temporais que podem ser informativas e enriquecer a conversa. Contudo a falta do factor “visual” no texto digital pode fazer com que uma conversa se torne seca e de menor interesse. Esta informação temporal está presente nas mensagens deixadas em fóruns mas é geralmente ignorada na sua forma original de uma simples listagem de texto. A razão para isto, é que não existe um padrão visual que capture o olhar imediatamente. Os dados temporais podem tornar-se mais perceptíveis se forem “mapeados” para uma representação gráfica. O GMP é um projecto que usa técnicas de “mapeamento” destes dados, para criar uma representação visual dos dados e da informação temporal das conversas digitais.

3.5 Chat Circles

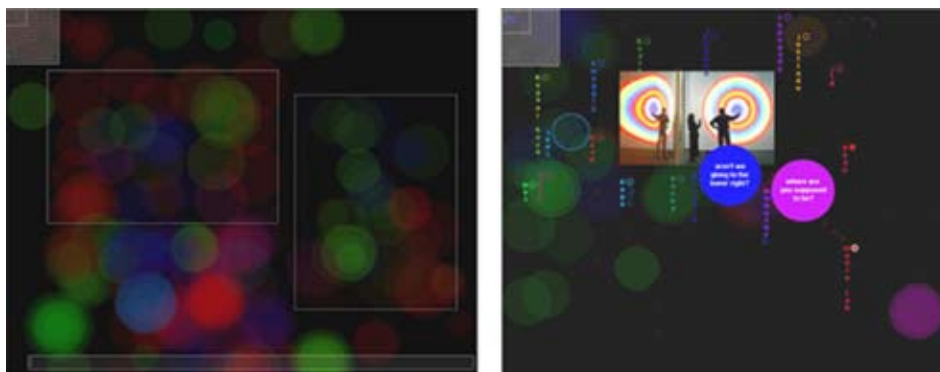


Figura 3.9 – Interface Chat Circles

Chat Circles [21] é um sistema/espço visual abstracto (correntemente na sua versão 2) no qual a conversa (*chat*) toma lugar. Cada indivíduo é representado como um círculo que se expande para mostrar a mensagem de texto. Para poder “ouvir” ou ver as palavras das mensagens de outro utilizador o nosso círculo tem que estar perto do deles. A nossa próxima mensagem substitui a mensagem anterior. O agrupamento social e ajustamento espacial indica quem está a “ouvir” quem pela proximidade. Desenvolvido no MIT Media Lab, o Chat Circles é uma tentativa de capturar (transportar para o domínio digital) as *nuances* subtis de expressão do nosso discurso do dia a dia. Ao contrário das formas tradicionais de *chat* que são bastante limitadas na forma de transmitir peças importantes de informação social, o Chat Circles representa a presença e actividades do utilizador usando a cor e forma, e a filtragem por proximidade forma agrupamentos de utilizadores intuitivamente com o objectivo de criar um ambiente mais rico para troca de mensagens online.

O interface gráfico do Chat Circles usa formas abstractas para representar a identidade e actividade. Foi criado com o objectivo de usar estas formas gráficas para representar as dinâmicas de uma conversação assim como desmascarar os padrões de actividade que emergem da interacção entre utilizadores. São empregadas formas bidimensionais simples que mudam em forma, tamanho e cor para comunicarem assim o ritmo de uma conversa.

O filtro de proximidade permite aos utilizadores concentrarem-se na conversa da qual fazem parte (aproximando-se do resto dos intervenientes dessa conversa), mas sem nunca perderem a noção do que se passa no resto do sistema.

Cada participante é representado como um círculo colorido no ecrã no qual as suas palavras vão aparecer (fig. 3.9). O círculo aumenta e torna-se mais brilhante com cada mensagem, e diminui de tamanho e desvanece nos períodos de silêncio, mas sem nunca desaparecer completamente. Este tipo de abordagem é uma tentativa de imitar as conversas orais do dia a dia, em que o foco é a pessoa que pronunciou a última palavra e progressivamente, essas palavras dissipam-se no decorrer da conversa.

O Chat Circles aponta claramente no sentido de revelar o nível de actividade, ou falta desta de cada participante. Difere dos ambientes baseados em texto, na medida em que a presença no *chat* é contínua. Continua com a tradição de interfaces baseados em avatars na medida em que a sua presença é constante, mas o nível de participação é constantemente sentido através dos gráficos, em contraste com os ambientes tradicionais baseados em avatars em que a presença é estática.

Mas o Chat Circles não é apenas um interface/espço de chat. O Chat Circles oferece também ao seu utilizador registos das suas actividades únicos. Ao contrário dos registos (como por exemplo do MSN Messenger) em que estes são ficheiros baseados em texto, o Chat Circles apresenta os relatórios da actividade com um interface visual. Esta escolha permite que seja mais fácil a descoberta de padrões de comportamento. O Chat Circles regista as mensagens do utilizador, com referências do tempo e tamanho da mensagem (além da própria mensagem), e as nossas movimentações no próprio espaço entre os vários *clusters* de utilizadores.

3.6 CrystalChat



Figura 3.10 - Vista de topo de uma visualização de CrystalChat

CrystalChat [22] visualiza redes pessoais de mensagens instantâneas tendo como objectivo principal revelar a interactividade temporal e fornecer informação acerca da quantidade e frequência das mensagens. Revela as actividades pessoais de *chat* usando os dados disponíveis dos registos do histórico de *chat*.

Ao fornecer um interface gráfico que representa as conversas através de um ambiente gráfico, o objectivo por detrás do CrystalChat é encorajar as pessoas a ver as suas conversas passadas e examinar o seu comportamento social (no *chat*).

O CrystalChat representa as nossas conexões para as pessoas que fazem parte da lista de contactos de *chat* como um eixo e raios de uma roda. Este diagrama de “eixo e raios” pode ser visto na fig. 3.10, onde o eixo (o vazio no centro) representa o dono dos arquivos de *chat* e cada raio representa um contacto de *chat*. Contudo existem mais dados no registo de mensagens do que aqueles que transparecem nesta vista de topo. Incluem-se dados acerca da iniciação e terminação de conversa, comprimento da conversa e da mensagem, padrões de repetição e o tom emocional representado por *emoticons*.

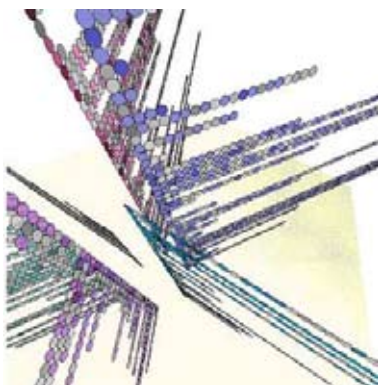


Figura 3.11 – Vista em perspectiva de uma visualização de CrystalChat

Uma das abordagens usada no CrystalChat para resolver esta questão, foi designar uma maior quantidade de dados possível para diferentes variáveis visuais e representar as “margens” do raio de acordo com essas variáveis. Esta abordagem envolve fazer uso das variáveis visuais que ainda não foram usadas no diagrama básico que caracteriza os raios. Uma vez que a localização e orientação criam o diagrama de eixo e raios, isto deixa as variáveis visuais tais como cor, tamanho, forma, valor de luminosidade e textura. Para incluir o maior número de detalhes possíveis acerca de padrões temporais e para incorporar o texto actual da mensagem dentro da visualização, os raios foram “estendidos” na terceira dimensão. A fig. 3.11 mostra uma vista aérea de modo a revelar como estes raios representam conversas com cada pessoa que se tornam uma “face” na qual os padrões temporais podem ser revelados.

Foram desenvolvidos os seguintes objectivos de design para a criação de uma estrutura que pudesse revelar os padrões temporais dos registos pessoais de mensagens instantâneas:

- Revelar as actividades de um individuo usando os dados que estão disponíveis nos registos de *chat* (histórico).
- Tornar as mensagens individuais visíveis e manter a ligação entre o símbolo visual da mensagem e o seu conteúdo de texto.
- Indicar o começo e fim da conversa como definido pela abertura da caixa de diálogo.
- Representar as interacções com cada pessoa como um grupo visual distinto numa conversa de *chat*.
- Ser capaz de mostrar a temporalidade, incluindo a ordem que indica a sequência na qual a mensagem é seguida de outra mensagem, e frequência para indicar o comprimento comparativo do tempo entre mensagens.
- Fornecer uma indicação do conteúdo emocional de um diálogo como indicado pelos *emoticons* usados.

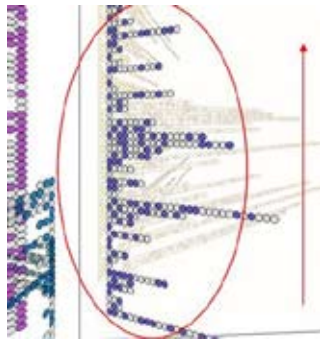


Figura - 3.12 Vista lateral de uma visualização de CrystalChat

Quando visto de “cima” cada raio representa todas as conversas com um contacto. Quando visto de lado, estes raios são revelados como faces planas que contem todos os detalhes das conversações com determinada pessoa. Como demonstrado na fig.3.12, as conversas são sequenciadas começando pelo fundo, de maneira que a conversa mais recente seja representada sempre no topo. Esta ordem faz mais sentido quando vista do topo, uma vez que a conversa mais recente é sempre a mais próxima.

Cada pessoa na lista de contactos tem uma “face” nesta visualização. O número de faces representa o número de pessoas que se escolhe representar. Todas as imagens usadas para esta explicação representam seis pessoas diferentes. Figura 3.10 e 3.12 mostram algumas evidências destas múltiplas faces ao longo do lado esquerdo. Cada face começa do eixo central e aponta para o exterior como os raios numa roda. A fig. 3.11 mostra a vista de topo dos seis raios. A região central aberta é pequena mas importante para a leitura de cada face individual. Cada face é composta por uma série de conversas que são construídas da seguinte maneira. Um disco colorido representa uma mensagem única e cada cor indica um contacto diferente. A leitura é da esquerda para a direita nas mensagens. Isto é, as mensagens são ordenadas de maneira que a primeira mensagem na conversa comece na esquerda. Uma conversa é definida, tal como nos dados, pela abertura e fecho da janela de chat. Uma fila de discos de mensagens representa uma conversa.

Como declarado nos objectivos de design, todos os detalhes de texto estão relacionados com os discos de mensagem. É possível explorar cada conversa em detalhe ao seleccionar cada um dos discos. A conversa “abre” Fig.3.13 diagonalmente para manter a ordem temporal e providenciar espaço para o texto de cada mensagem ser mostrado. A cor do texto corresponde à cor usada para representar a pessoa que enviou a mensagem.

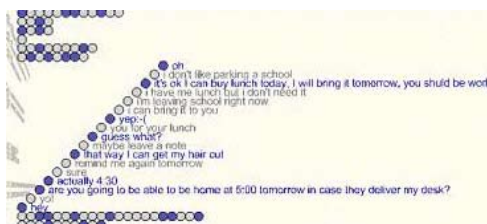


Figura 3.13 –Visualização das mensagens de uma conversa no CrystalChat

Todas as ilustrações até agora mostram as mensagens ordenadas por tempo e as conversas também. O tempo foi considerado como a variável de ordenamento. Para representar mais precisamente a sequência de informação, os tempos actuais de conversação podem ser incluídos. As “falhas” (espaços) entre mensagens representam a quantidade de tempo que ocorreu entre as conversas. Este modo de representação ocupa mais espaço visual mas revela o “espaço” de tempo das conversas com diferentes contactos, ordenando a visualização relativo ao tempo. A linha de tempo ainda é vertical e aumenta para “cima”. Quando os dados temporais são usados explicitamente, cada fila de mensagens corresponde às conversas do dia. Uma maior distância entre as filas, corresponde a dias de intervalo entre conversas.

Ao fornecer um interface gráfico que representa as conversas de mensagens instantâneas, o Crystalchat é uma tentativa de encorajar as pessoas a ver as suas conversas passadas e examinar o seu comportamento social no meio digital. Enquanto os padrões emergem visualmente através de uma visualização social, as memórias e reminiscências do passado podem ressurgir e os hábitos pessoais consciencializados e talvez influenciar comportamentos futuros de interacção.

3.7 Quadro Comparativo

	Thread Arcs	PostHistory	Themail	GMP	ChatCircles	CrystalChat
Tipo de Arquivos analisados	<i>e-mail</i>	<i>e-mail</i>	<i>e-mail</i>	entradas de texto próprias (tipo fórum)	mensagens instantâneas próprias	mensagens instantâneas (MSN Messenger)
Análise do protocolo da mensagem	Emissor CC BCC	Emissor CC BCC	Emissor CC BCC Assunto	Data Hora	Emissor Receptor Data Hora	Emissor Receptor Data Hora
Análise do conteúdo da mensagem	—	—	Texto da mensagem	—	Texto da mensagem	Texto da mensagem e <i>emoticons</i>
Método de análise	Tempo real / automático (modo vigilância)	Iniciado pelo utilizador (modo introspectivo)	Iniciado pelo utilizador (modo introspectivo)	Tempo real / automático (modo vigilância)	Iniciado pelo utilizador (modo introspectivo)	Iniciado pelo utilizador (modo introspectivo)

3.8 A Importância de um caso de estudo

Os dados analisados pelos projectos descritos anteriormente, representam actos de comunicação (envio e recepção de mensagens) entre várias pessoas e como tal, a privacidade e importância destes actos não pode ser relativizada. Para melhor entender, analisar e agir sobre dados que contêm actos de comunicação entre duas pessoas foi necessário desenvolver um projecto específico para estudo.

Os capítulos seguintes são dedicados à apresentação, descrição técnica e análise dos dados recolhidos para o caso de estudo desta dissertação.

4 Análise Técnica

Para que seja possível analisar a comunicação produzida via mensagens instantâneas é necessário construir ferramentas capazes de dissecar os vários componentes dos ficheiros de histórico, vulgarmente conhecidos como *logs*.

O caso de estudo que se segue, diz respeito a 12 semanas de mensagens trocadas com 12 contactos via MSN Messenger. Neste capítulo iremos analisar como estes ficheiros se compõem e os métodos usados para arquivar, contabilizar, pesquisar e visualizar os dados recolhidos.

4.1 Os ficheiros de Histórico - o registo da comunicação

O MSN Messenger (versão 7.5) não tem como opção predefinida a recolha automática das mensagens trocadas com os contactos. Esta função tem que ser activada pelo utilizador nas opções da aplicação. Só então passam a ser guardados automaticamente numa pasta definida pelo utilizador os registos da troca de mensagens.

O MSN Messenger guarda o registo das mensagens trocadas no formato XML [25] e cria um ficheiro distinto para cada contacto.

XML é a sigla de EXtensible Markup Language, uma linguagem de marcação (“Markup Language”), um subconjunto da SGML – Standard Generalized Markup Language. Mais do que uma linguagem, XML é uma metalinguagem que se constitui como um conjunto de especificações que permitem definir novas linguagens de marcação. A XML foi projectada para descrever e organizar documentos. Uma linguagem XML não tem um conjunto de marcadores (*tags*) predefinidos (como acontece, por exemplo, no HTML), mas, para cada caso, a XML permite que seja possível especificar e criar as *tags* que melhor se adaptam à situação.

A XML foi concebida para ser fácil de usar na Web, para suportar o ciclo de vida de um documento (edição, validação, armazenamento e transformação) e ser um subconjunto simples da SGML, simplificando a sua utilização.

“Com <marcadores> tipo XML podemos não só descrever o conteúdo, como também fixar a estrutura lógica desse mesmo conteúdo. A flexibilidade da XML provém da faculdade de transportar variadíssimos tipos de dados e de mantê-los estruturalmente coesos. A XML serve magnificamente para a estruturação de qualquer tipo de dados e para descrevê-los sem dúvidas ou ambiguidades em formato de texto. Isto porque a XML permite definir a linguagem de marcação mais adequada a um dado tipo de documentos. Documentos em formato XML podem conter, por exemplo, Transacções Comerciais, Catálogos de Produtos, Gráficos Vectoriais, Equações Matemáticas, Anúncios Publicitários, enfim: todos os dados que necessitem de uma representação estruturada...”

Paulo Heitlinger, O Guia Prático da XML, 2001

Alguns factos a sublinhar:

- A XML não foi pensada para substituir o HTML;
De facto, enquanto que o XML descreve e organiza documentos, de forma flexível e adaptável, o HTML apenas pode permitir a organização de um único tipo de documento;
- A XML permite separar a sintaxe e a semântica da visualização, especificando apenas os mecanismos que permitem a estruturação dos documentos;
- A XML especifica como criar e usar marcadores, ou etiquetas, como forma de definir tipos de documentos; ou seja a XML é um mecanismo usado para criar novas linguagens de marcadores;
- A XML pode ser usada em qualquer plataforma e qualquer país (ou seja, suporta o UNI-CODE) Qualquer que seja a aplicação onde um documento XML seja usado, ele vai ser processado (analisado) e só depois usado. Muitas vezes esse processamento pode significar transformação.

Um processador XML não é mais do que um analisador sintáctico (parser) que, reconhecendo e validando os elementos básicos XML, efectua uma determinada acção sobre eles.

Um ficheiro XML criado pelo MSN Messenger têm a seguinte estrutura:

```
<Message Date="10-05-2006" Time="20:21:41" DateTime="2006-05-10T19:21:41.314Z" SessionID="1">
  <From>
    <User FriendlyName="Rita"/>
  </From>
  <To>
    <User FriendlyName="Victor"/>
  </To>
  <Text>hello victor... :)</Text>
</Message>

<Message Date="10-05-2006" Time="20:21:55" DateTime="2006-05-10T19:21:55.034Z" SessionID="1">
  <From>
    <User FriendlyName="Victor"/>
  </From>
  <To>
    <User FriendlyName="Rita"/>
  </To>
  <Text>olá :)</Text>
</Message>
```

Figura 4.1 – Exemplo da estrutura de um ficheiro de registo de MSN Messenger

Cada *tag* de mensagem (<Message> </Message>) contém vários tipos de elementos:

- Date** – Informação de data (dia, mês, ano)
- Time** – Informação de tempo (hora, minuto, segundo)
- DateTime** – Timestamp com a informação de data e tempo agregadas
- SessionID** – Informação sobre o número de sessão.
- From** – Quem envia a mensagem
- To** – Quem recebe a mensagem
- Text** – Mensagem enviada

Convém referir que a informação sobre o número de sessão é criada da seguinte maneira: cada vez que se inicia a aplicação com sucesso (ligação ao servidor), é contabilizada como uma sessão diferente. O número de sessões de cada contacto pode diferir, pois esta informação de sessão só é recolhida cada vez que existe troca mensagens com um contacto (activação de uma janela de conversação fig.4.2).



Fig. 4.2 – Janela de Mensagem do MSN Messenger

Se durante uma troca de mensagens com um contacto, a ligação ao servidor, por exemplo, for reiniciada três vezes, o ficheiro de registo terá indicado três sessões diferentes com esse contacto, mesmo que a conversa tenha sido contínua e com intervalos de tempo muito reduzidos (poucos segundos).

Por outro lado podemos ter uma troca de mensagens com um determinado contacto com intervalo de tempo grande entre conversas (várias horas, dias), e se a ligação ao servidor não for perdida a sessão continua a ser a mesma. O acto de “fecho” da janela de conversação não influencia o factor sessão.

A maneira em como foi definida uma conversa será discutida no ponto 4.4.

A XML é óptima para a estruturação de dados e para descrevê-los sem dúvidas ou ambiguidades em forma de um texto. Isto porque a XML permite ao autor do documento definir os seus próprios <elementos>, criando e usando os <marcadores> (=tags) mais adequados à descrição e à estruturação dos dados em causa.

No entanto por ser um “formato de texto”, não é a solução mais eficaz para efectuar operações de pesquisa (e manipulação) do seu conteúdo. Por esta razão os dados recolhidos dos diversos contactos foram “mapeados” para uma base de dados MySQL [26].

4.2 O processo e fases

O processo, representado na fig. 4.3, descreve as várias fases: recolha/aquisição, processamento e visualização da informação.

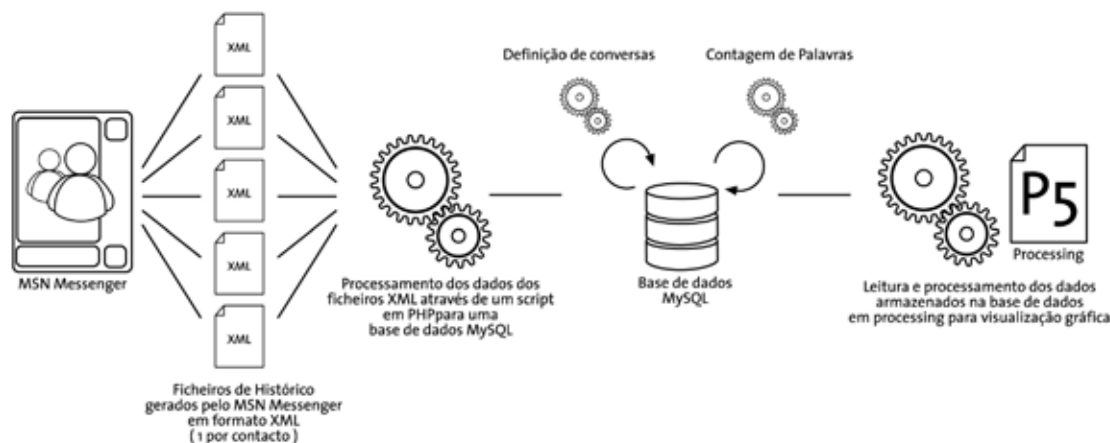


Figura 4.3 – Esquema representativo das diferentes fases e processos.

Para uma maior eficácia no processamento dos dados, foi escolhido como suporte de armazenamento dos dados, intermédios e finais, uma base de dados relacional, neste caso a MySQL. As bases de dados relacionais permitem o armazenamento de grandes quantidades de dados de forma eficiente, quer na adição de novos dados quer na consulta dos mesmos de forma rápida, um requisito importante para a questão em estudo (*logs* de conversas), já que a quantidade de dados a ser processada pode ascender rapidamente a dezenas de MB (*megabytes*).

De entre os software livres (*free* – sem custo de aquisição) para gerir bases de dados, foi escolhido o software MySQL, porque é dos que tem a maior divulgação (importante quando se procura soluções/respostas para dúvidas que aparecem) e tem um baixo consumo de recursos (*memória* e *cpu*), permitindo assim trabalhar com quantidades razoáveis (muitos MB) de dados sem ser muito exigente para o computador onde executa.

Os dados armazenados na BD (Base de dados) contêm informação sobre: emissor, receptor, da data e hora (h:m:s) de envio, mensagem enviada e nº da conversa respectiva.

Foram criadas 3 tabelas na base de dados, uma para recolher a informação inicial, outra com a informação processada (definição de “conversa” e homogeneização dos *nick's* usados) e finalmente uma para contabilização de palavras.

Para o estudo foi necessário guardar as conversas originais, os dados intermédios (necessários para o “cálculo” dos dados finais) e os dados finais. Estas fases são descritas de forma mais pormenorizada nos pontos seguintes.

4.3 Recolha de dados

Na fase de aquisição dos dados, os vários ficheiros usados pelo MSN Messenger para armazenar os dados dos históricos das conversas são reunidos e transformados, através de um “mapeamento” simples, e finalmente, guardados numa tabela na base de dados.

```
inicio
  abrir ficheiro xml
  se encontrar mensagem
    guardar elementos da mensagem (de, para, data, mensagem) num array
  ray
    passar à próxima mensagem
  se não encontrar mais mensagens
    introduzir elementos do array na base de dados
fim
```

Para uma maior eficácia no processamento dos dados armazenados nos ficheiros de histórico, foi construído um *parser* em PHP [27] que analisasse e inserisse cada *tag* e atributo do ficheiro de XML numa base de dados MySQL [26]. O pseudo código acima é a estrutura base do *parser* criado para ler os vários ficheiros originais (históricos das conversas), decompor o XML nas várias componentes, “mapeá-los” e inseri-los na base de dados.

A estrutura da tabela que recolheu a informação inicial está ordenada da seguinte maneira:

```
| msn_1 | (
  `id` int(11),
  `datahora` datetime,
  `de` varchar(50),
  `para` varchar(50),
  `mensagem` text,
  `contacto` varchar(50),
)
```

Para cada entrada (mensagem) na tabela existe um id específico e além dos atributos que constam no ficheiro XML (de, para, data, hora), foi definido um campo “contacto” que representa o contacto com o qual foi efectuado o acto comunicativo.

4.4 Identificação de conversas e contactos

Um dos projectos revistos no capítulo anterior, o Crystal Chat [22], tem também como base a análise de registos de mensagens instantâneas. A definição de conversa usada é dada através da *tag* “session” que se encontra no ficheiro XML do registo. No entanto como já tinha sido explicado anteriormente (4.2), este item não é um indicador fiel do que pode ser considerado uma conversa. Quer seja que por falha na ligação (o que implica o reinício da aplicação e acrescento de sessões no que seria considerado uma só conversa) ou por uma ligação prolongada mas com grandes intervalos de tempo entre troca de mensagens.

Dada a natureza espontânea e casual do uso das mensagens instantâneas [28] foi decidido usar um intervalo máximo de 15 minutos como separador de conversas. Ou seja, se existir um intervalo superior entre duas mensagens é contabilizado uma nova conversa. Whittaker [29] refere que no estudo efectuado foi usado como separador de conversas um intervalo máximo de 7 minutos. O uso de um intervalo superior deveu-se a uma análise manual prévia de várias amostras e consequente contabilização de intervalos entre assuntos abordados.

A diferença entre estes dois intervalos centra-se principalmente no estilo de conversação, caracterizado nesta dissertação mais facilmente pois a troca de mensagens é de 1 para n (autor e os seus contactos) enquanto que o estudo levado a cabo por Whittaker [28] tem em conta várias dezenas de participantes (n para n). Este factor é determinado por uma “média” dos vários participantes do estudo. Uma vez que neste projecto “eu” sou a “constante” foi possível determinar um período que reflectisse mais este caso de estudo.

```
inicio
    ler dados da base de dados ordenados pela data da mensagem

    iniciar contador conversa = 0
    iniciar contador mensagem = 0

    se tempo da mensagem - tempo mensagem seguinte for > 15 minutos
        conversa = conversa + 1
        inserir numero da conversa na base de dados
    se tempo da mensagem - tempo mensagem seguinte for < 15 minutos
        conversa = conversa
        inserir numero da conversa na base de dados

    passar à mensagem seguinte

fim
```

O código PHP (exemplo da estrutura em pseudo código acima), percorre todas as conversas, de cada utilizador, com o objectivo de segmentar as mensagens trocadas em “sessões” de conversas em que o intervalo entre elas seja superior a 15 minutos. Depois de identificada uma nova “conversa” é-lhe atribuída um identificador, a mensagem é novamente guardada na BD mas com o respectivo id da “conversa”, este id será usado em posteriores processamentos. Os dados são inseridos numa nova tabela que contém um campo para contabilização de conversas.

```
| msn | (
    `id` int(11),
    `datahora` datetime,
    `de` varchar(50),
    `para` varchar(50),
    `mensagem` text,
    `conversa` int(11),
    `contacto` varchar(50),
)
```

Os dados iniciais contêm informação da “alcunha” (*nickname*) que os diversos contactos podem estar a usar em determinado momento. Durante todo o processo de recolha de dados o meu *nickname* manteve-se o mesmo para evitar confusões de pertença da “mensagem”. Como já referido o MSN Messenger armazena um ficheiro para cada contacto, no seu interior os “nomes” usados para identificar o emissor (e receptor) podem mudar, sendo a única constante o que usei durante todo o processo (ver trecho de XML abaixo).

```
<From>
  <User FriendlyName="Victor"/>
</From>
```

Ao serem inseridos na BD os *nickname's* foram “mapeados” ao seu respectivo contacto para uma melhor identificação.

4.5 Estatísticas básicas, contagem de palavras

Para a contagem de palavras foi usado um script em PHP (ver pseudo código abaixo), para “ler” todas as mensagens, separar as respectivas palavras, retirando os símbolos de pontuação (?,!,”,/, etc) e contabilizar o número de vezes que cada palavra é usada.

```
inicio
  ler dados da base de dados // mensagens
  remover pontuação (.,?!; ...etc)
  substituir duplos, triplos, etc. espcementos por 1 espcamento

  separar mensagem pelos espaços em varias palavras
  contabilizar o número de vezes que a palavra é repetida

  inserir na nova tabela (palavra , nº de vezes que aparece)
fim
```

Por fim os dados apurados são guardados numa tabela criada para este efeito. Cada palavra continua associada à mensagem correspondente (de quem, para quem, hora, etc) através do item id que está relacionado com o id da tabela principal.

```
| palavras | (
  `id` int(11),
  `palavra` varchar(50),
  `contador` int(11),
)
```

4.6 Visualização Gráfica

“Processing is an open source programming language and environment for people who want to program images, animation, and sound. It is used by students, artists, designers, architects, researchers, and hobbyists for learning, prototyping, and production. It is created to teach fundamentals of computer programming within a visual context and to serve as a software sketchbook and professional production tool. Processing is developed by artists and designers as an alternative to proprietary software tools in the same domain.”

www.processing.org, Julho 2007

O Processing é uma linguagem e um ambiente de programação vocacionada para o ensino dos fundamentos da programação em computador. É usado em vários cursos [30] (com especial relevância nos cursos ligados à arte digital e *new media*) um pouco por todo o mundo e foi criado como uma ferramenta para prototipagem (criação de esboços) de *software* mas com capacidade para criar um produto final.

O carácter *open-source*, a facilidade de aprendizagem e uso da sua linguagem, as capacidades gráficas 2D e 3D e integração com base de dados MySQL foram as características chave que fizeram para que a aplicação/linguagem Processing fosse a ferramenta usada na visualização dos dados recolhidos.

O ambiente Processing é escrito em JAVA. Programas escritos em Processing são também traduzidos para Java e depois corridos (*run*) como programas de Java. Programas escritos em JAVA e Processing geralmente correm mais depressa que programas baseados em linguagem de *scripting* como o ActionScript ou Lingo, o que é importante para o volume de dados acumulado deste projecto.

As grandes distinções entre o Processing e Java residem na biblioteca gráfica do Processing e no estilo mais simplificado de programação que não requer entender conceitos mais avançados como classe, objectos, double-buffering (enquanto mesmo assim deixando-os acessíveis a utilizadores mais avançados) tornando os programas mais curtos e fáceis de ler (e entender).

O ambiente de desenvolvimento do Processing (PDE – Processing Development Environment) consiste num simples editor de texto para escrita do código, uma área de mensagem para *debugging* e uma série de botões básicos para fazerem correr os programas criados neste ambiente.

O software criado usando o Processing é designado como *sketch*. Estes são escritos no editor de texto e para os fazer “correr” (e parar) basta usar os botões de acção (*Play* e *Stop*).

```

size(200, 200);
smooth();
background(0);
noStroke();
fill(226);
triangle(10, 10, 10, 200, 45, 200);
rect(45, 45, 35, 35);
quad(105, 10, 120, 10, 120, 200, 80, 200);
ellipse(140, 80, 40, 40);
triangle(160, 10, 195, 200, 160, 200);

```

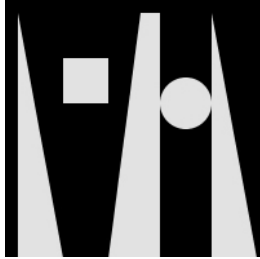


Figura 4.4 - Trecho de código de Processing e o seu resultado visual

“When comparing the genome of two different people, you’ll see single letter changes (called SNPs, pronounced “snips”) every few thousand letters. An interesting feature of SNPs is that their ordering has distinct patterns, where sets of consecutive changes are most often found together. There are many methods for looking at this data, so this piece combines several of them into a single visual display.”

Ben Fry, <http://benfry.com/isometricblocks>, Julho 2007

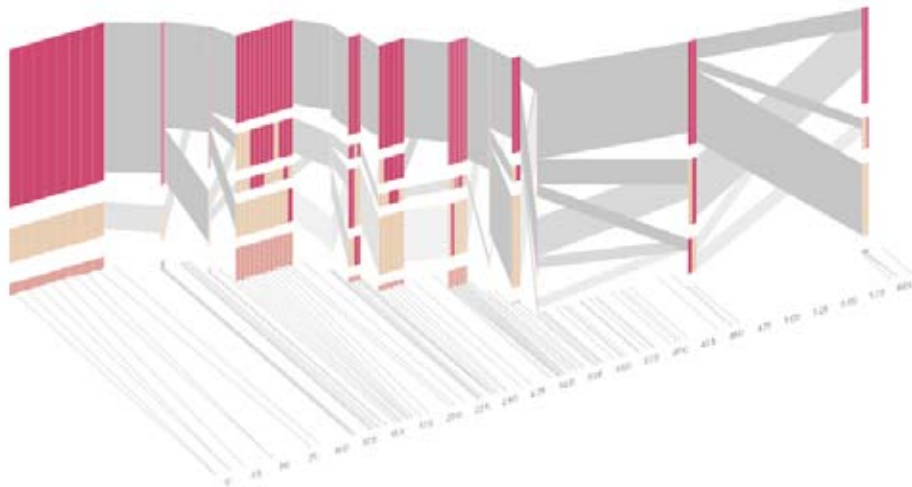


Fig.ura 4.5 – Ben Fry Isometric Blocks

Este é um exemplo do “bom” comportamento do Processing para lidar com grandes quantidades de dados e das suas capacidades de representação gráfica, qualidades essenciais para o desenvolvimento deste caso de estudo.

O Processing é *open-source*, fazendo com que o seu desenvolvimento seja constante, com novas características a surgirem regularmente devido à grande comunidade que se aliou ao projecto Processing. O facto de usar bibliotecas externas foi um factor crucial para a sua utilização neste projecto. Uma das bibliotecas desenvolvidas contempla a comunicação directa do Processing com bases de dados MySQL. As bibliotecas gráficas do Processing proporcionam a possibilidade de exportação de formas 2D para ficheiros que suportem formatos vectoriais de desenho tal como Illustrator [32] ou PDF [32]. Este factor fez do Processing o companheiro ideal para a visualização gráfica dos dados neste projecto pois possibilita que as visualizações produzidas sejam facilmente transportadas do meio digital para o físico. O formato vectorial (escalável) permite expor fisicamente (quase sem limite de dimensões e em suporte físico) uma experiência inerentemente digital.

4.7 Os Gráficos

A representação estatística feita com gráficos encontra-se presente em quase todos os meios de divulgação de informação, designadamente em jornais e revistas, manuais escolares, apresentações públicas, etc. É necessário comunicar ideias complexas com clareza, precisão e eficiência e para isso os gráficos devem: mostrar os dados; induzir o observador a pensar na substância e não na forma; evitar distorcer os dados; representar muitos dados em pouco espaço.

O melhor gráfico é aquele que dá ao observador o maior número de ideias no mais curto intervalo de tempo usando o menor espaço. A grande vantagem de um gráfico reside na sua capacidade de contar uma história de forma interessante e atractiva permitindo compreender rapidamente fenómenos que dificilmente seriam percebidos de outra forma. Contudo, tal não implica que este processo seja feito de forma simples, sendo necessário muito trabalho e cuidado.

Segundo Edward Tufte [1] a excelência em gráficos de estatística consiste em comunicar ideias complexas com clareza, precisão e eficiência. Contudo, criar um gráfico ou um mapa que de facto informe e seja, simultaneamente, apelativo, legível e coerente com os dados não é tarefa fácil.

Tufte delineou algumas regras básicas para a construção de um bom gráfico:

- A representação dos números, medida na superfície dos gráficos, deve ser proporcional à quantidade representada
- Devem usar-se legendas claras e detalhadas para minorar ambiguidade e distorção
- Deve mostrar-se a variação dos dados e não do design
- O número de dimensões usadas na representação não deve ser maior que o número de dimensões dos dados

- Não se deve representar informação fora de contexto

“Graphical elegance is often found in simplicity of design and complexity of data.”

Edward Tufte, *The Visual Display of Quantitative Information*, 1992

O objectivo deste estudo foi criar gráficos que simultaneamente condensassem vários níveis de informação (conversas iniciadas, mensagens, palavras usadas) e permitissem identificar padrões (e consequentemente casos que se destacassem da norma) de modo a facilitar a análise dos dados.

Para esse propósito foi criado um modelo adequado às dimensões observadas: conversas, mensagens e palavras. A fig.4.9 representa o modelo de gráfico criado para representar os dados recolhidos neste projecto. Com esta solução é possível condensar os dados e mais facilmente identificar padrões que na forma de tabela (fig. 2.4) seria uma tarefa quase impossível.

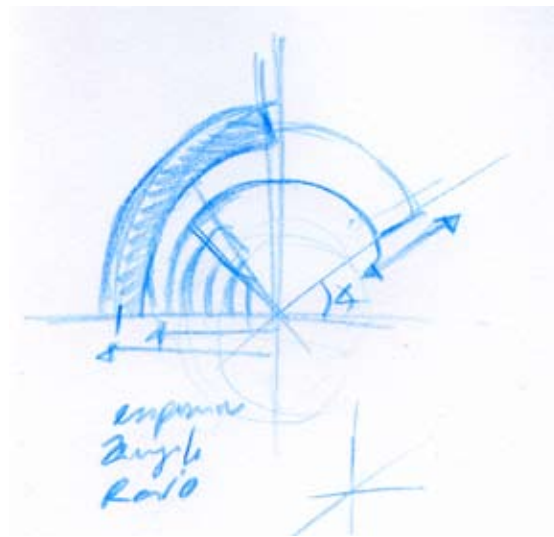


Figura 4.6 – Esboço inicial do modelo de gráfico a utilizar

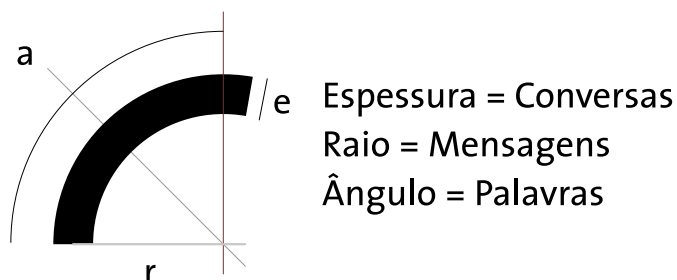


Figura 4.7 – Modelo representativo das 3 dimensões analisadas

Com a utilização de diferentes propriedades de um arco (espessura, ângulo e raio) foi possível concentrar 3 dimensões de análise (conversas, mensagens e palavras). A organização dos dados recolhidos teve um papel fundamental na criação deste modelo pois influencia de uma forma decisiva o tipo de solução encontrado. Este modelo resulta bem na identificação de padrões, mas para uma leitura integral dos dados cada gráfico é acompanhado de uma tabela (legenda) com os valores numéricos observados.

Um segundo tipo de gráfico foi criado para responder à exigência de uma sequenciação temporal da troca de mensagens com os diferentes contactos. Este é uma planificação bi-dimensional em que o eixo vertical representa as horas e o horizontal os dias. Cada mensagem é representada por um círculo com a cor do contacto respectivo (com o cinzento a representar-me a mim) e o diâmetro deste varia de acordo com o tamanho da mensagem (número de caracteres).

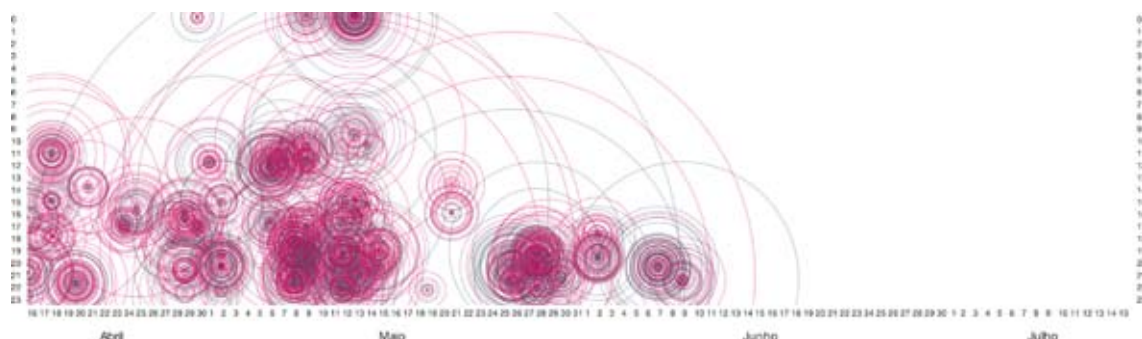


Figura 4.8 – Exemplo de uma representação temporal da troca de mensagens com 1 contacto

Desta maneira é possível observar a flutuação da distribuição das mensagens através dos dias e das horas.

4.8 – Wordclouds

Uma das possibilidades interessantes de análise é a gramática e as palavras usadas na troca de mensagens. As visualizações do tipo *wordcloud* [33] (nuvem de palavras que diferem no seu tamanho por ordem de importância / uso / popularidade) permitem dar uma perspectiva sobre quais as palavras mais usadas por cada pessoa, estilos de escrita, etc.

"A word cloud is a visual depiction of frequently used words in your web site. It illustrates keyword density using font size. The more often a word appears on your site, the larger it appears within the word cloud."

www.iconinteractive.com/tools/wordcloud/, Julho 2007

4.9 Navegação entre mensagens

A fig. 2.3 (capítulo 2) mostra a visualização que o Internet Explorer permite do ficheiro de histórico em formato XML. Como referido é praticamente impossível de discernir padrões (e variações) da comunicação através desta simples visualização. No entanto ela mostra-nos de uma forma clara e sequencial as mensagens trocadas, com indicação das várias características da mensagem: de, para, data, a própria mensagem, etc. Quando procedemos à identificação de variações (fig. 4.11) ao longo do tempo é necessário ter uma ferramenta auxiliar que permita navegar pelas mensagens trocadas para uma melhor contextualização dessas variações. Foi construída uma “ferramenta” que mostra todas as mensagens trocadas com um determinado contacto (fig. 4.13).

Cada “quadrado” representa uma mensagem (os quadrados marcados com um ponto interior representam as minhas mensagens) e estes estão ordenados sequencialmente (esquerda para a direita e de cima para baixo). Os agrupamentos de vários quadrados com intensidade de cor diferente, representam as conversas (segundo as especificações de diferença de 15min descritas anteriormente).

É possível ao seleccionar com o cursor um determinado quadrado aceder às suas propriedades: de, para, hora, data e a mensagem.



Figura 4.10 - Visualização de todas as mensagens trocadas com 1 dos contactos

Esta ferramenta é uma ajuda preciosa quando queremos “recordar” o que se passou em determinado momento, podendo assim aceder às mensagens trocadas para uma melhor contextualização. Ao “ligar” cada mensagem com o seu emissor é possível através desta ferramenta mostrar a frequência de envio o que poderá ser um factor indicador do estilo de escrita de cada contacto.

4.10 Cor

Para uma melhor distinção dos diferentes contactos, foi necessário associar cores a cada um. Estas foram seleccionadas através de um espaço de cor CIE Lab [34] devido à uniformidade perceptual deste. De maneira a tentar cobrir o máximo de espectro as cores foram escolhidas o mais dispersas possível no espaço, para que a selecção resultasse em cores distintas de representação para cada um dos contactos.

O espaço e cores CIE $L^*a^*b^*$ (também conhecido por CIE Lab) é actualmente um dos espaços de cor mais populares para a medição de cores e é amplamente utilizado em praticamente todos os campos de investigação. Ele é um dos espaços uniformes de cor definido pela CIE [34] (*Commision Internationale L'Eclairage*) em 1976 com o objectivo de reduzir os problemas do espaço de cor original Yxy no qual as distâncias do diagrama de cromaticidade, não representavam as diferenças visuais das cores.

No espaço de cores $L^*a^*b^*$, o L^* indica a luminosidade, enquanto que a^* e b^* , representam as coordenadas das cores: sendo $+a^*$ a direcção do vermelho, $-a^*$ a direcção do verde, $+b^*$ a direcção do amarelo e $-b^*$ a direcção do azul. O centro é acromático. Com o aumento dos valores a^* e b^* , o ponto distancia-se do centro e a saturação da cor aumenta. A figura 4.6 é a representação de um sólido de cores do Espaço de Cor $L^*a^*b^*$. A figura 4.7 mostra um corte horizontal deste sólido num valor constante de L^* .

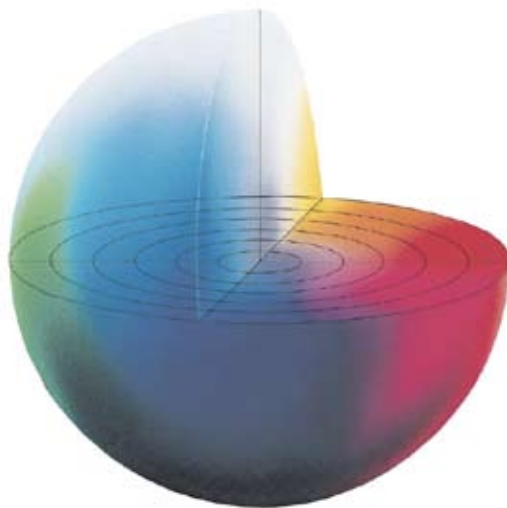


Figura 4.11 – Representação tridimensional do espaço de cor CIE Lab

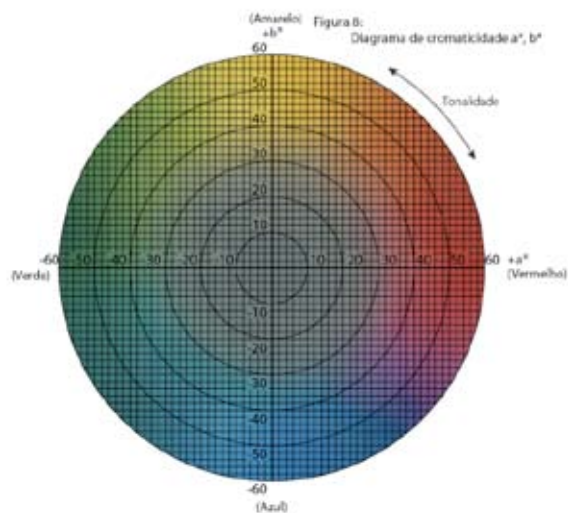


Figura 4.12 – exemplo de 1 corte transversal no espaço de cor CIE Lab

Usando como referencia o trabalho de Michael B. Dillencourt, David Eppstein e Michael T. Goodrich [35] sobre selecção de cores distintas para melhor percepção (no espaço CIE Lab) foram seleccionadas 12 cores para representação dos diferentes contactos.

A imagem seguinte representa as cores seleccionadas para cada contacto no seu valor em RGB, um dos sistemas usados pelo Processing para atribuição da cor.

ADRIANA - RGB(223,0,94)	
ANDREIA - RGB(255,160,0)	
EUSTÁQUIO - RGB(97,201,242)	
JOANA - RGB(123,90,54)	
JosÉ - RGB(83,117,40)	
MARCO - RGB(106,71,75)	
MIGUEL - RGB(179,21,25)	
RAQUEL - RGB(122,26,29)	
RITA - RGB(101,20,100)	
RUI - RGB(128,140,148)	
SÉRGIO - RGB(174,255,22)	
TIAGO - RGB(0,98,113)	

Figura 4.13 – Cores usadas para representação dos diferentes contactos analisados

4.11 Caminhando para os resultados

Neste capítulo foi descrito métodos e ferramentas usadas para recolher, armazenar, processar e visualizar os dados relativos à troca de mensagens via MSN Messenger. No capítulo seguinte serão apresentados os resultados deste caso de estudo, expondo totais, médias e distribuição temporal das conversas, mensagens e palavras usadas na troca de mensagens com os diversos contactos.

5 Resultados do caso de estudo

Como referido no capítulo 4, a recolha de mensagens prolongou-se por um período de 12 semanas entre 16 de Abril a 15 de Julho e não existiu nenhum processo de selecção de contactos a analisar. O total de contactos na lista (durante o período de recolha) era 35. Os 12 ficheiros de histórico representam os contactos com quais existiu troca de mensagens durante o período de recolha. Não houve um “contacto” (ou recusa de) deliberado (pelo menos conscientemente) da minha parte com nenhum contacto em especial, numa tentativa de não influenciar a amostra. Durante todo o processo de recolha tentei fazer uma utilização “regular” do MSN Messenger. Pode-se contra-argumentar que segundo o princípio da incerteza de Heisenberg [36] (numa versão rude e popular da física quântica) que o problema do observador reside no facto que o acto de observar interfere e altera, inevitavelmente o comportamento do que está a ser observado, por isso refiro que “pelo menos conscientemente” não interfeiri nos “contactos” que foram realizados durante este estudo.

Nos diversos trabalhos analisados anteriormente (que analisam a troca de mensagens), os autores dos projectos não constavam como participantes activos. Os dados que analisaram eram pertença de terceiros que os disponibilizaram para o estudo. Pelos dados que possuo desses trabalhos e estudos, a possível influência do “saber que estou a ser observado” não é referenciada ou contabilizada como possível factor de influencia dos dados.

Sendo este um caso de estudo com algumas características de auto-retrato/diário pessoal, é possível desde já identificar 3 factores potencialmente subversivos dos dados:

1. **Promiscuidade do sistema** – é um sistema “teocêntrico” em que a referência comum entre os diversos contactos sou “eu” o criador, o administrador, o observador e o avaliador do sistema.
2. **Problema da observação** – o acto de observar não é um acto paralelo sem impacto no sistema. O conhecimento da observação (meu desde o início da recolha e dos diversos contactos a partir de determinado período descrito mais detalhadamente no ponto 5.3) é um factor a ter em consideração.
3. **Subjectividade da análise** – a “classificação” do grau da relação é uma questão pessoal, dificilmente quantificável e a análise (quer inicial quer final) é relativa, e depende essencialmente do indivíduo que avalia e não de uma concertação geral do que é uma “relação”.

No final do período de observação e recolha registou-se comunicação com 12 contactos de um total de 35, e os dados resultantes da troca de mensagens foram integralmente “mapeados” para a base de dados sem recurso a nenhum processo de censura.

5.1 Classificação inicial dos contactos

Uma das questões mais complicadas que podemos encontrar é a de classificar as nossas relações. Mesmo que se consiga num determinado momento dividir as nossas relações em grupos (tais como, amigos íntimos, amigos distantes, relações laborais, família, etc), essa classificação é sempre relativa a um determinado momento no tempo. As relações estão em constante mutação/evolução, quer seja devido a uma zanga ou afastamento, ou uma mudança na direcção contrária (evolução de relação laboral para amizade ou de amizade para relação amorosa por exemplo).

Uma classificação simples da minha visão sobre o tipo de relação que mantinha com os 12 contactos foi realizada antes da recolha e análise dos dados. Estes 12 contactos foram divididos inicialmente entre quatro grupos genéricos: círculo íntimo, colegas, círculo profissional e amigos distantes. Estes grupos não são estanques e podem se sobrepor-se.

A minha perspectiva inicial era de que iria existir uma distinção clara entre certos contactos. Esta distinção não se reflectia apenas na quantidade de mensagens trocadas, mas principalmente na relação entre mensagens enviadas e mensagens recebidas. O facto de haver grande quantidade de comunicação (reflectida através do numero de mensagens) não significa que determinado contacto possa ser considerado mais “importante” (numa escala de intimidade) do que outro. Muitos factores externos podem condicionar a quantidade. O modo como os “outros” se relacionam com o MSN Messenger é um destes factores, pois a quantidade de tempo que passam “ligados” condiciona a quantidade ou simplesmente o tipo de relação: uma relação laboral têm tendência a gerar grandes quantidades de dados, mas a relação pode ser considerada irrelevante em termos sociais (íntimo).

Os ficheiros de histórico do MSN Messenger não revelam nada sobre o que os outros contactos estão a fazer quando não estão a interagir comigo. Seria errado considerarmos que estes estão constantemente “ligados”. Os hábitos de uso diferem bastante de pessoa para pessoa. Algumas usam-no ocasionalmente ao fim de semana, outras apenas em horário pós laboral e os utilizadores mais intensivos podem facialmente estar constantemente *online*. Devido a esta impossibilidade de os ficheiros de histórico não registarem o tempo de ligação dos contactos (nem o meu) o factor da intensidade de troca de mensagens não pode ter a relevância que teria se todos os contactos estivessem com o mesmo tempo de ligação.

Neste capítulo, numa tentativa de ultrapassar o problema do simples somatório, serão apresentados gráficos de médias que indicam a média de mensagens por conversa e de palavras por mensagem. Assim poderá verificar-se se em contactos com um menor número de conversas, se verifica uma maior intensidade de comunicação (troca de mensagens).

A cada utilizador foi atribuído um factor de “conectividade”, factor que indica a minha percepção do tempo que cada utilizador passou ligado durante o tempo da recolha. Defini 5 níveis, sendo que o nível 5 representa um utilizador muito frequente e o nível 1 um utilizador pouco frequente. Um utilizador muito frequente é aquele que passa praticamente todos os dias ligado com uma média diária de 12 ou mais horas. Considerando que a minha

média de ligação diária era de cerca de 10 horas, o meu factor de conectividade é de 4 (C4). Só é possível observar a conectividade dos contactos, se nós próprios estivermos ligados, mas a simples observação de que um contacto já está ligado a partir do momento em que nos ligamos e continua ligado depois de nos desligarmos é a razão de existir um C5, um factor de conectividade superior ao que atribui a mim próprio.

A minha percepção inicial seria de que os contactos categorizados como íntimos (e com uma relação profissional associada) seriam obviamente aqueles com quem iria existir um maior número de conversas e uma troca equilibrada de mensagens (enviadas vs recebidas) por conversa. Este “equilíbrio” de que falo deveria também reflectir-se no número de conversas iniciadas e na sua distribuição temporal. Os contactos pertencentes a este grupo são contactos com uma conectividade elevada (C4 ou C5). Os grupos denominados de Colegas ou Círculo Profissional viriam logo a seguir ao grupo “Círculo Íntimo” em termos de número de mensagens, no entanto a minha visão inicial seria que este grupo poderia mostrar desequilíbrios. Uma vez que algumas das relações “profissionais” tendem a ser relações com hierarquias, a tendência natural é que um dos contactos seja mais “procurado” pelo outro. Para último o grupo dos “amigos distantes” seria um grupo com poucos contactos, mas quando aconteciam seriam contactos mais prolongados, dando origem a uma média maior de mensagens por conversa.

A minha percepção inicial dos contactos seria então representada deste modo:

Adriana – Círculo íntimo / Círculo profissional - C4

Andreia - Círculo íntimo – C5

Eustáquio - Círculo íntimo – C5

Joana - Círculo íntimo - C4

José - Círculo íntimo – C5

Marco - Círculo íntimo / Círculo profissional - C5

Miguel – Colegas - C3

Raquel – Amigos distantes / Círculo íntimo (distância física) – C1

Rita – Amigos distantes (distância física) – C1

Rui – Amigos distantes (distância física) – C2

Sérgio - Círculo profissional – C4

Tiago – Colegas / Círculo profissional – C4

5.2 Divisão temporal do período analisado

As 12 semanas de recolha podem ser divididas em 2 períodos de igual duração. Durante as 6 semanas iniciais a comunicação processou-se sem que ninguém fosse “avisado” da recolha e armazenamento das mensagens trocadas. Findo este período inicial, foi comunicado a todos os contactos da lista (via *e-mail*, SMS e no próprio MSN Messenger) que iria proceder à recolha de todas as conversas que tinha via MSN Messenger. Nenhum contacto foi informado da existência de uma recolha inicial. A divisão em 2 períodos (não informado / informado) foi deliberada e a razão por detrás desta divisão é a procura de (e se pode ser determinada) uma diferença de comunicação quando existe uma consciência de monitorização do processo comunicativo.

Nos pontos seguintes irei apresentar os resultados finais (totais e médias) quer na sua forma inicial tabelar (ponto 5.4) quer na forma de gráficos (segundo o modelo descrito previamente) dos resultados de conjunto e individuais. Para cada contacto observado será incluído um gráfico de progressão temporal da troca e intensidade das mensagens e duas *wordclouds* que obedecem às regras descritas no capítulo 4.

5.3 Tabela de Resultados Obtidos

	Conversas iniciadas	Mensagens enviadas	Média de mensagens por conversa	Palavras usadas	Média de palavras por mensagem
Adriana	46	1434	16	4920	3.4
Victor	43	1258	14	4109	3.3
Total	89	2692	30	9029	3.4

	Conversas iniciadas	Mensagens enviadas	Média de mensagens por conversa	Palavras usadas	Média de palavras por mensagem
Andreia	20	671	15	2327	3.3
Victor	24	825	19	3685	4.5
Total	44	1496	34	5912	3.9

	Conversas iniciadas	Mensagens enviadas	Média de mensagens por conversa	Palavras usadas	Média de palavras por mensagem
Eustáquio	77	1300	9	5301	4.0
Victor	65	1807	13	4767	2.7
Total	142	3107	22	10068	3.2

	Conversas iniciadas	Mensagens enviadas	Média de mensagens por conversa	Palavras usadas	Média de palavras por mensagem
Isana	39	823	12	2794	3.4
Victor	31	730	10	3000	4.1
Total	70	1553	22	5794	3.7

	Conversas iniciadas	Mensagens enviadas	Média de mensagens por conversa	Palavras usadas	Média de palavras por mensagem
Jose	94	1693	12	8935	5.2
Victor	49	1705	12	5135	3.0
Total	143	3398	24	14070	4.1

	Conversas iniciadas	Mensagens enviadas	Média de mensagens por conversa	Palavras usadas	Média de palavras por mensagem
Marco	66	2530	18	11879	4.7
Victor	78	2940	20	9563	3.3
Total	144	5470	38	21442	3.9

	Conversas iniciadas	Mensagens enviadas	Média de mensagens por conversa	Palavras usadas	Média de palavras por mensagem
Miguel	13	195	8	757	3.9
Victor	12	200	8	702	3.5
Total	25	395	16	1459	3.7

	Conversas iniciadas	Mensagens enviadas	Média de mensagens por conversa	Palavras usadas	Média de palavras por mensagem
Raquel	0	110	37	1059	9.6
Victor	3	219	73	772	3.5
Total	3	329	110	1831	5.6

	Conversas iniciadas	Mensagens enviadas	Média de mensagens por conversa	Palavras usadas	Média de palavras por mensagem
Rita	2	85	21	224	2.6
Victor	2	86	21	229	2.6
Total	4	171	42	453	2.6

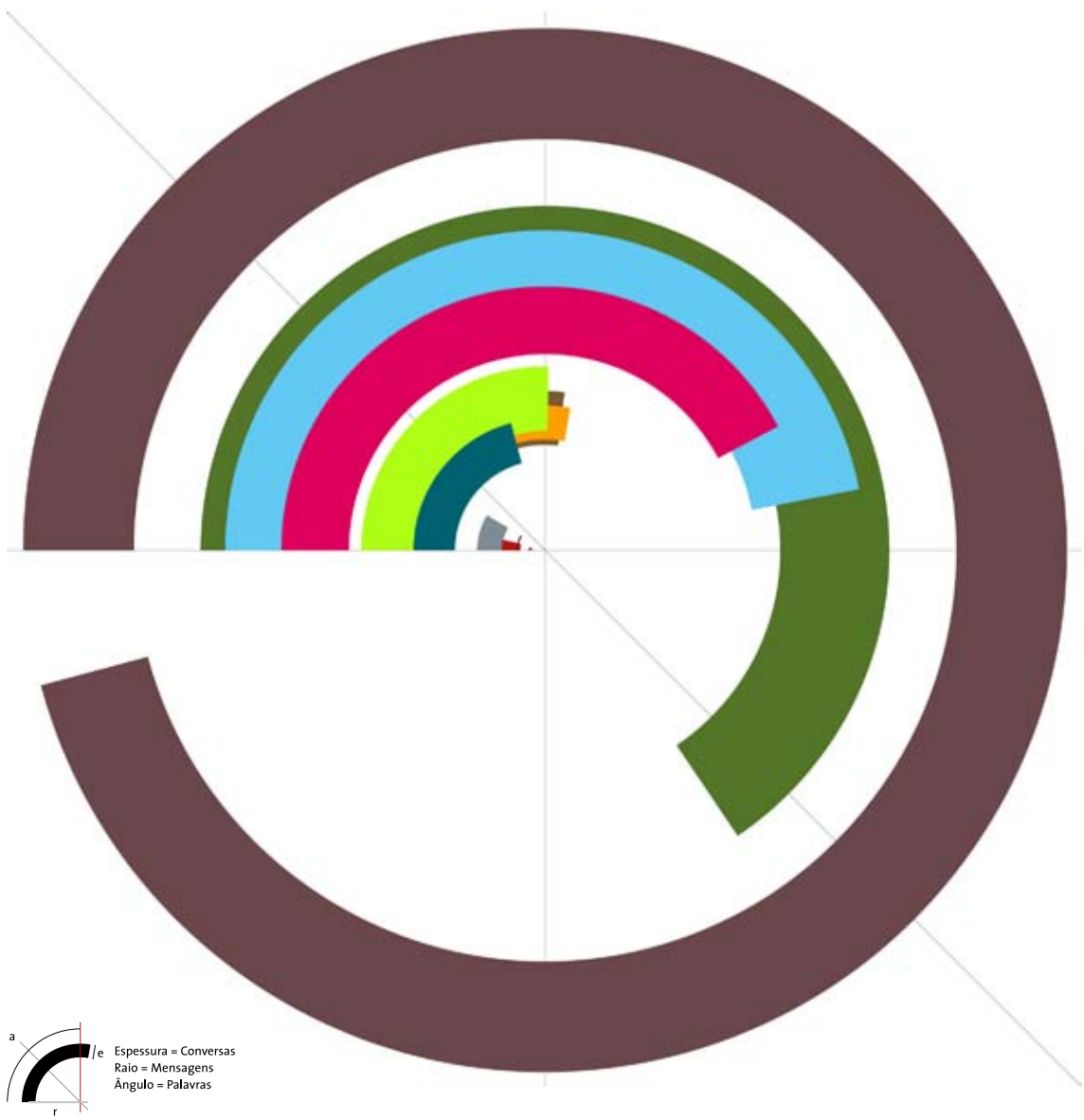
	Conversas iniciadas	Mensagens enviadas	Média de mensagens por conversa	Palavras usadas	Média de palavras por mensagem
Rui	20	358	11	1070	2.9
Victor	12	298	9	764	2.6
Total	32	656	20	1834	2.7

	Conversas iniciadas	Mensagens enviadas	Média de mensagens por conversa	Palavras usadas	Média de palavras por mensagem
Sérgio	64	972	12	3266	3.4
Victor	20	805	10	2151	2.7
Total	84	1778	22	5417	3.0

	Conversas iniciadas	Mensagens enviadas	Média de mensagens por conversa	Palavras usadas	Média de palavras por mensagem
Tiago	26	753	14	3336	4.4
Victor	28	545	10	1156	2.2
Total	54	1298	24	4492	3.4

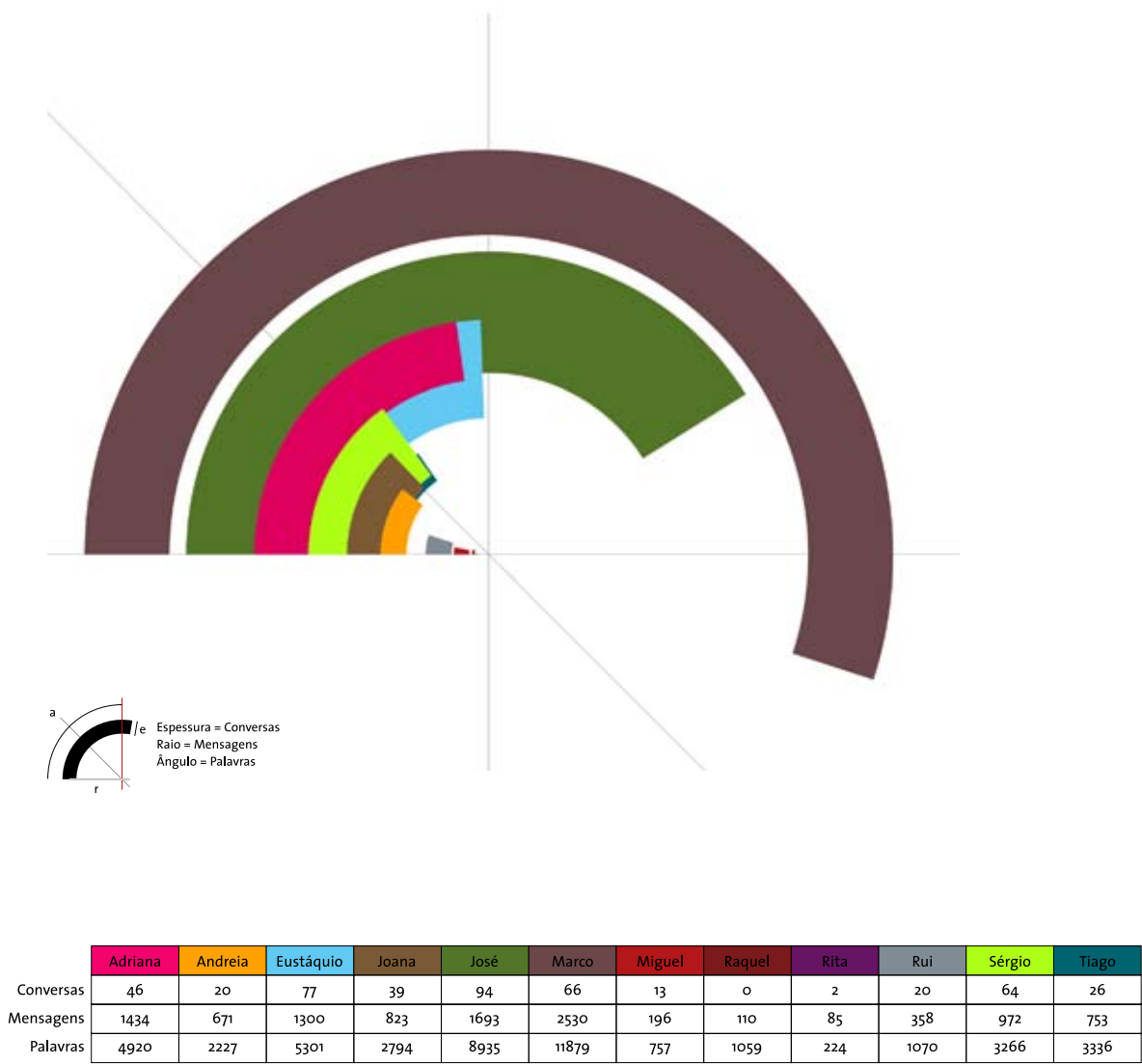
5.4 Totais

5.4.1 Totais de conversas, mensagens e palavras por utilizador (recebido e enviado)

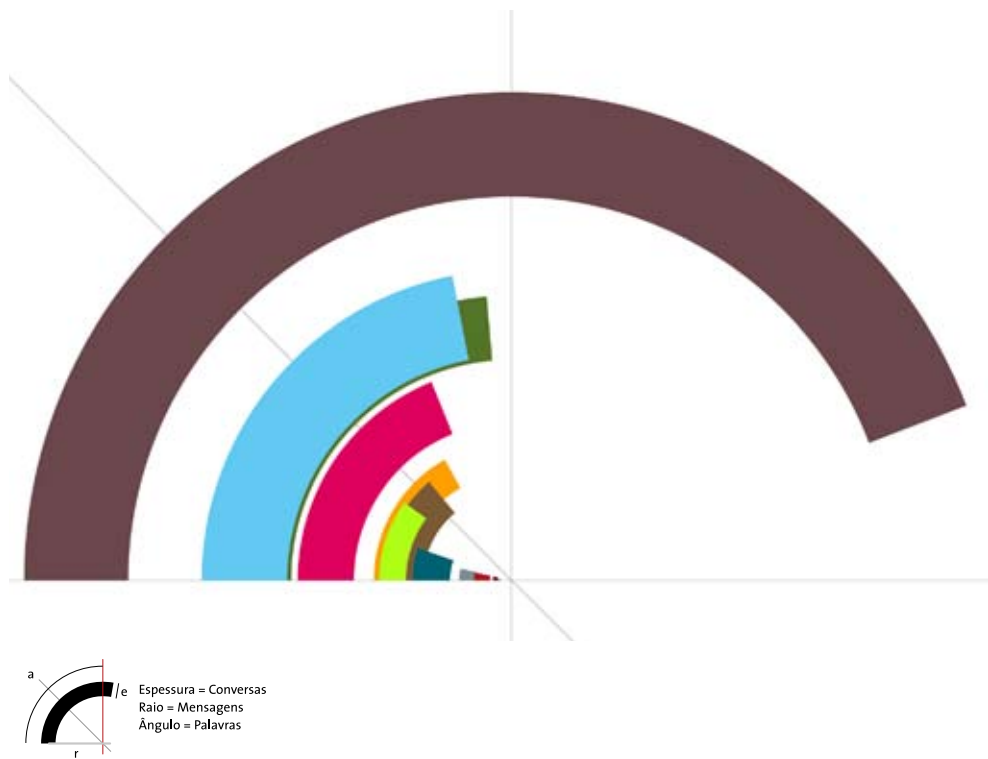


	Adriana	Andreia	Eustáquio	Joana	José	Marco	Miguel	Raquel	Rita	Rui	Sérgio	Tiago
Conversas	89	44	142	70	143	144	25	3	4	32	84	54
Mensagens	2692	1496	3107	1553	3398	5470	396	329	171	656	1778	1298
Palavras	9029	5912	10068	5794	14070	21442	1459	1831	453	1834	5417	4492

5.4.2 Totais de conversas, mensagens e palavras por utilizador recebidas



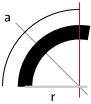
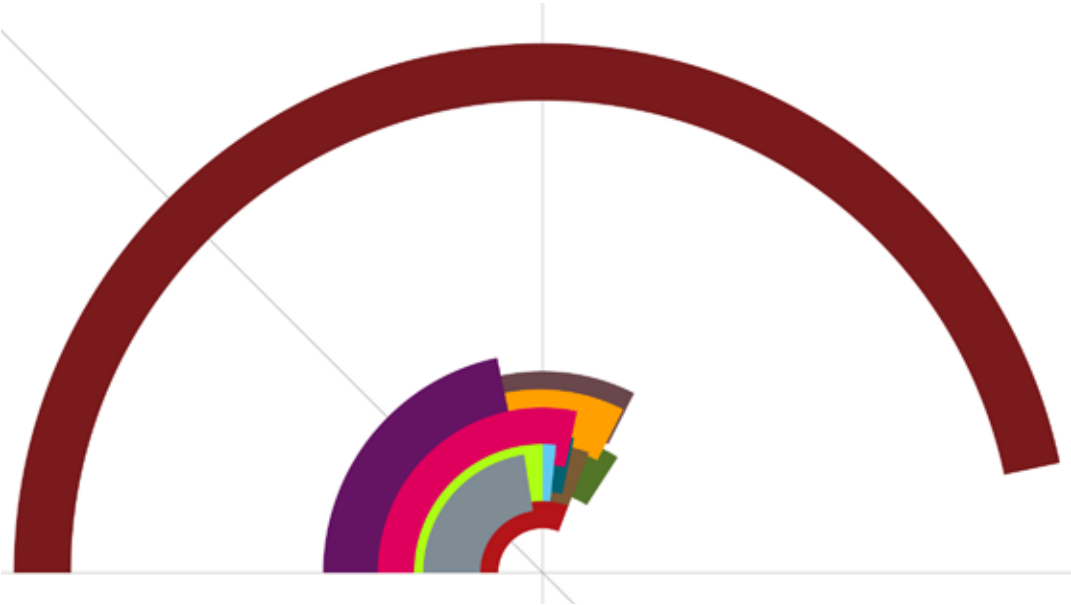
5.4.3 Totais de conversas, mensagens e palavras por utilizador enviadas



	Adriana	Andreia	Eustáquio	Joana	José	Marco	Miguel	Raquel	Rita	Rui	Sérgio	Tiago
Conversas	43	24	65	31	49	78	12	3	2	12	20	28
Mensagens	1258	825	1807	730	1705	2940	200	219	86	298	806	545
Palavras	4109	3685	4767	3000	5135	9563	702	772	229	764	2151	1156

5.5 Médias

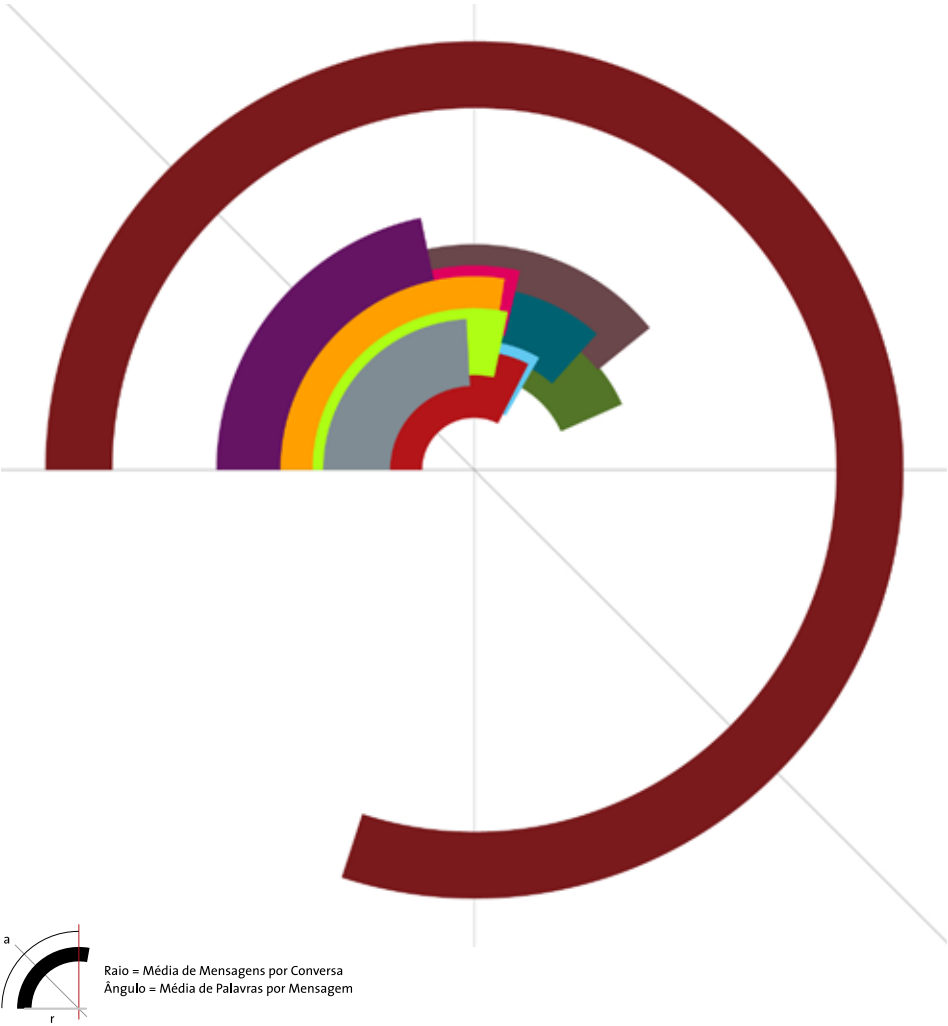
5.5.1 Médias Totais: Mensagens por Conversa e Palavras por Mensagem



Raio = Média de Mensagens por Conversa
Ângulo = Média de Palavras por Mensagem

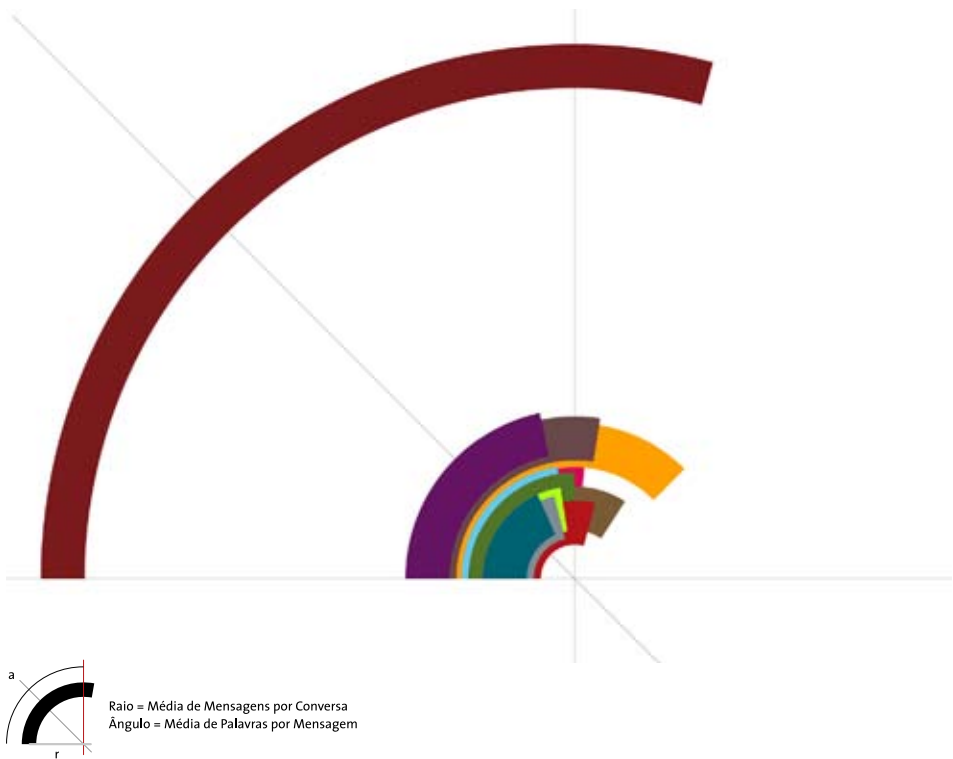
	Adriana	Andreia	Eustáquio	Joana	José	Marco	Miguel	Raquel	Rita	Rui	Sérgio	Tiago
Mensagens por Conversa	30.2	34	21.9	22.9	23.8	38	15.8	119.7	42.8	20.5	22.2	24
Palavras por Mensagem	3.35	3.95	3.24	3.73	4.14	3.91	3.68	5.56	2.64	2.79	3.04	3.46

5.5.2 Médias de Mensagens por Conversa e Palavras por Mensagem Recebidas



	Adriana	Andreia	Eustáquio	Joana	José	Marco	Miguel	Raquel	Rita	Rui	Sérgio	Tiago
Mensagens por Conversa	16.1	15.2	9.1	11.7	11.8	17.5	7.8	36.6	21.2	11.2	11.5	13.9
Palavras por Mensagem	3.43	3.32	4.08	5.28	5.28	4.7	3.86	9.63	2.63	2.99	3.36	3.43

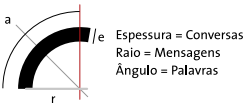
5.5.3 Médias de Mensagens por Conversa e Palavras por Mensagem Enviadas



	Adriana	Andreia	Eustáquio	Joana	José	Marco	Miguel	Raquel	Rita	Rui	Sérgio	Tiago
Mensagens por Conversa	14.1	18.7	12.7	10.4	11.9	20.4	8	73	21.5	9.3	9.6	10.1
Palavras por Mensagem	3.16	4.47	2.64	4.11	3.01	3.25	3.51	3.52	2.66	2.56	2.67	2.12

5.6 Adriana

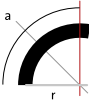
5.6.1 Gráfico comparativo entre itens enviados, recebidos e totais



Espeçura = Conversas
Raio = Mensagens
Ângulo = Palavras

	Adriana	Victor	Total
Conversas	46	43	89
Mensagens	1434	1258	2692
Palavras	4920	4109	9029

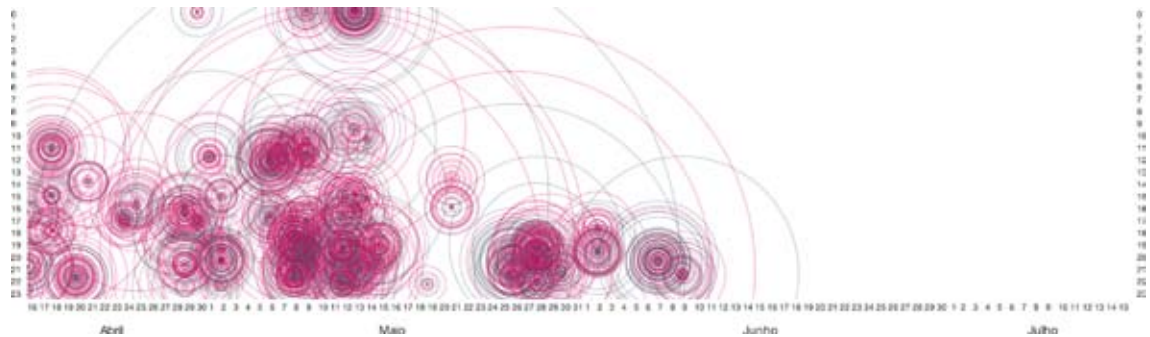
5.6.2 Gráfico Comparativos de médias



Raio = Média de Mensagens por Conversa
Ângulo = Média de Palavras por Mensagem

	Adriana	Victor	Total
Mensagens por Conversa	16.1	14.1	30.2
Palavras por Mensagem	3.43	3.16	3.35

5.6.3 Distribuição temporal das mensagens

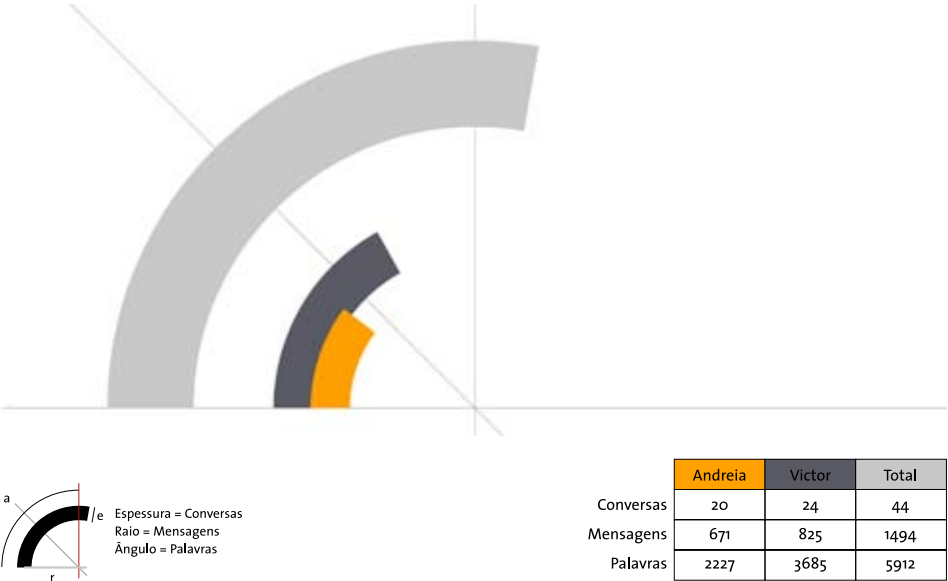


5.6.4 Wordcloud

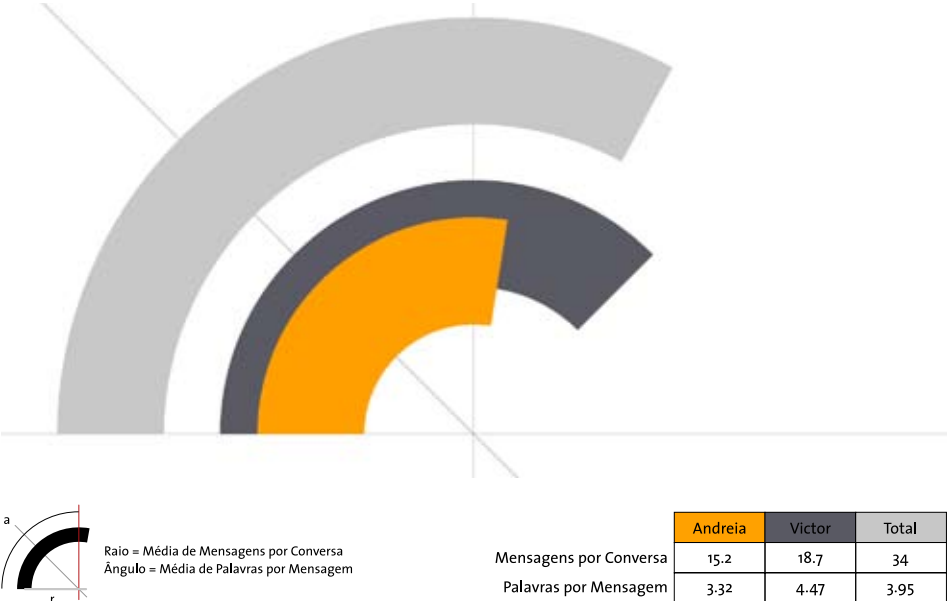
coisa site mail tempo boa tipo
beijinhos vik casa coisas euros dia noite
computador xico miguel envio som ideia fica sneakers
desculpa porto trabalhar qc proposta fds trabalho cena beijo volta
mestrado jantar mig parte email memso manha hora exacto trbalho ideias
semana ficheiros seca irma horas orfama amnha tenis orçamento pdf fontes falta
3d etc noticias tratar verdade menu semna amanha imagem perguntar funcionar
pc bj novidades investigar pensar flash escrever certeza falar novo mac existe
super empresa min dados abri linha catalogo pena tsc volto ligado jeito ligar disco
calma tentar sites saber heheh sorry pessoal novos ocupado procurar ano filme
emprego material inicio barra meio baixo andar responder abrir msn acabar falamos cheia
minimo preço design miuda voltar lisboa modo refazer dormir primeira victor freehand tiff img
nome \$ vale ligou-me swf aparece falou dito arrnjar colocar consultora dominio estra sistema
precisamos palavra final carro fazermos ecri cheguei comprar tenha combinar quiseses atelier temso
valor altura nozelos concorrência mudar mes inteiro basicamente musica procura files rapariga
mandar formato explicar mini chegar fazemos esboço dias nike online mnha pro diferentes sai faço
contrato vir resto serve perfeitamente po sms tao simples problema altamente empresas sergio facil

.7 Andreia

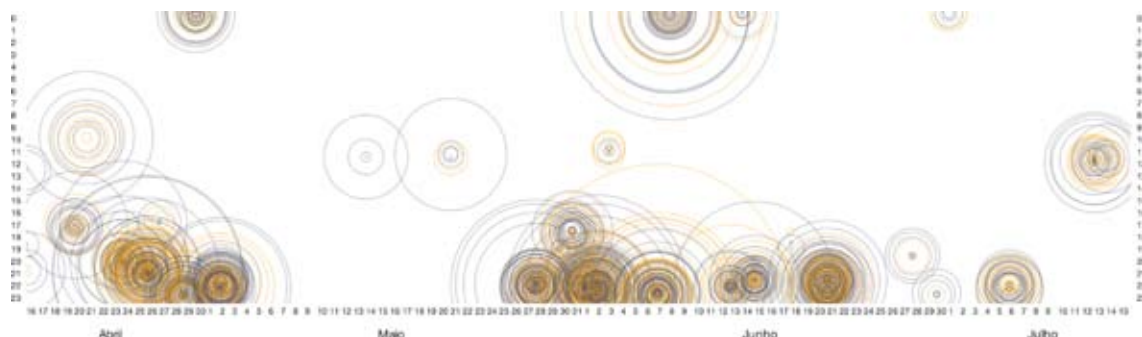
5.7.1 Gráfico comparativo entre itens enviados, recebidos e totais



5.7.2 Gráfico Comparativos de médias



5.7.3 Distribuição temporal das mensagens

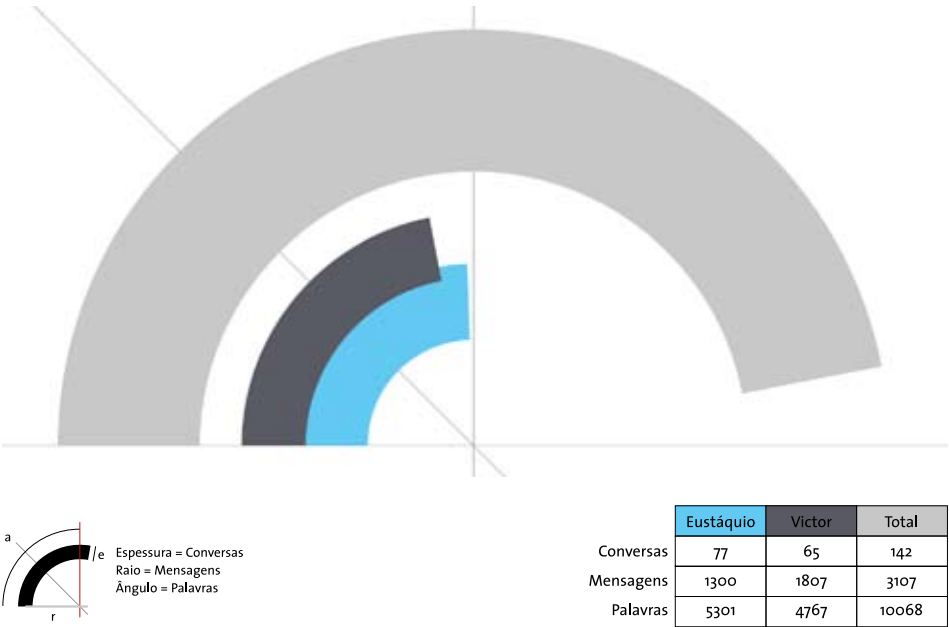


5.7.4 Wordcloud

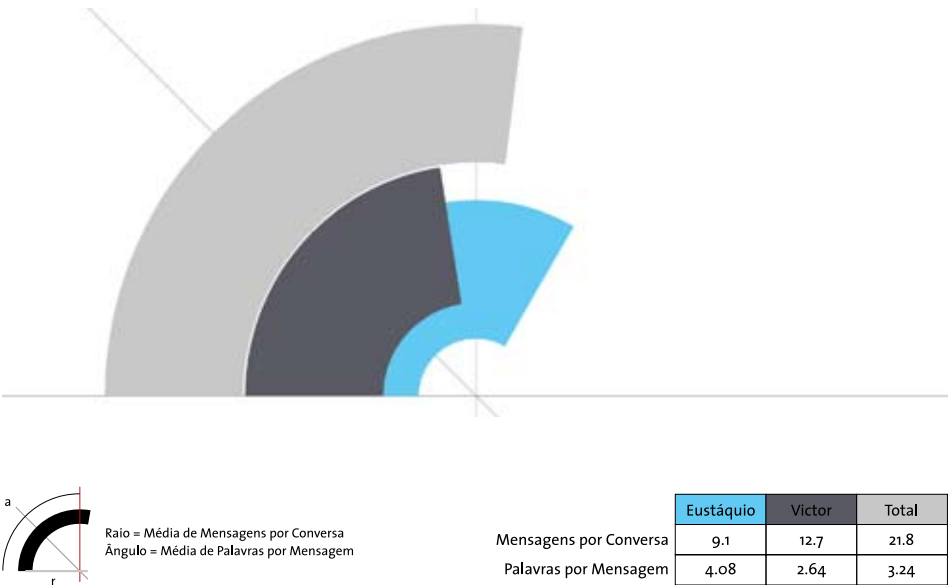
tempo coisa coisas fica gajo dia aldeia master boa novo
pessoal ano tipo trabalho meio interessa mac trbalho falar horas etc
tropas processo sorry luis faço dias empresa noite segunda manha entregar
capitulo site FEUP mestrado pedir computador final fico projecto mes euros
trabalhar marco aumentar seca cavalos ataque conseguir motivação dissertação parte
marta novos devem certeza imenso meses heitor area artes miguel cena normal mandei
merda muro campos tentar precisas soldados mnha amanha contas qures computer aliança
metro dizes tese exemplo list setembro carta anda passado Maio correr cool daqui grande recibos
causa abrir pessoas fazes andar pagar heheh po states enquanto momento mandar tratar uam ve
jogo saber travian mensagens ataques contumil super nivel conta montes diferentes bairro pass dele
definição pontos cabeça gaja sal entregou sucesso FBAUP explicar pro comprar saíram book betariz
vou-te pergunta gente usar episodio tipos esperar dá-me jeito começo livros pos contente poder anos teses
apresentar vida engenharia tempos meter vantagens vista percebes desenhar paper princio mundo icon
escrever fuck hummm fase mail atacada gato builders saia tv chegam obrigada resto atacar ler hora vizinho
are there sério ha aldeias opções memso problema fores nome

5.8 Eustáquio

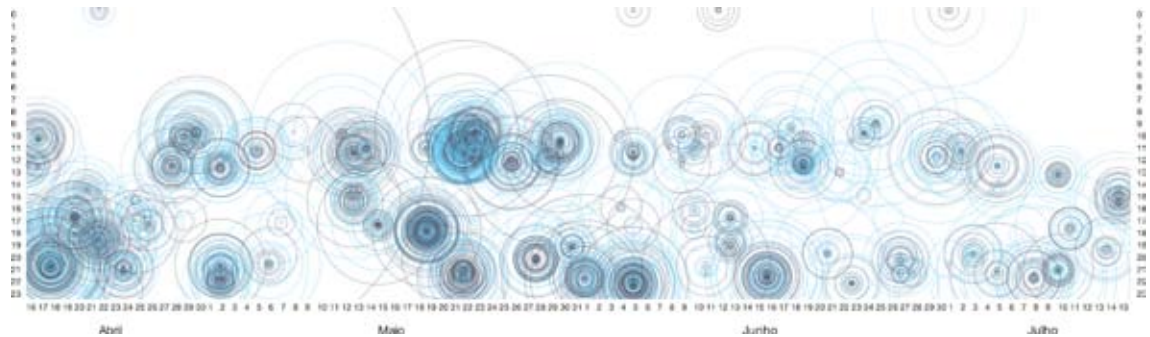
5.8.1 Gráfico comparativo entre itens enviados, recebidos e totais



5.8.2 Gráfico Comparativos de médias



5.8.3 Distribuição temporal das mensagens

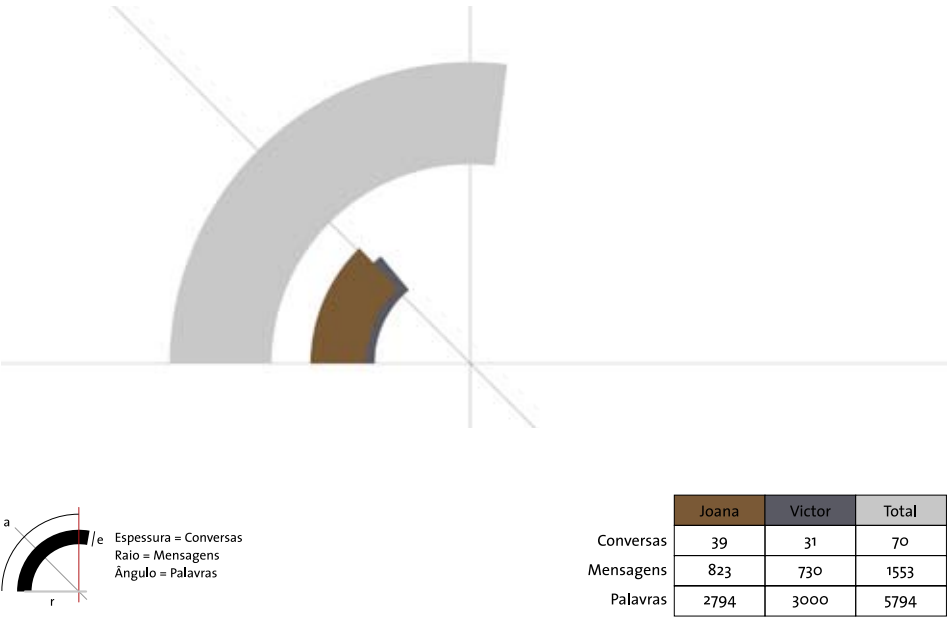


5.8.4 Wordcloud

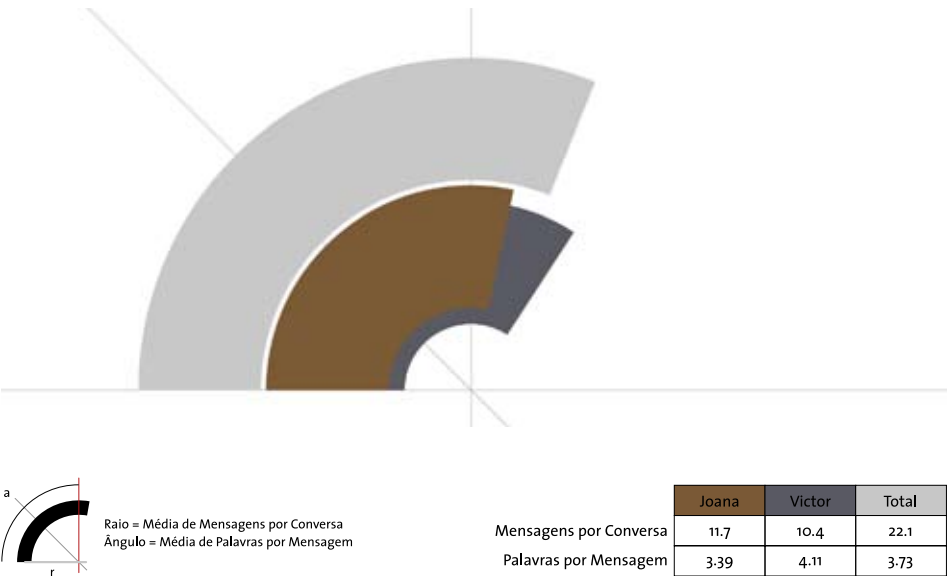
boa coisa aldeia nivel marco gajo horas
francesinha cool aliança Pro casa tropas belas aldeias
euros campos hora amial cena etc fixes upgrade fica dia tipo dias
po semna Mac fds amnha exacto merda sério disco the filme ataques qc
dvd crop bike conheço tempos funciona mandei pena memso power começar
iBook anos meio segunda daqui video coisas preto sangue ataque produção link
jeito vale preço pasta manana molho merdas levar raids precisas resto caralho
point tempo hasta quiseses noite PC fodido baixo correr ram max aumentar yeah
meia gauleses master mês travian frente barato gravar final conta hostel sexo violencia
boss trailer spider match chave Tiger montes sucks site online takio parte mail contas
precisar andam windows pessoal problema gostei novo fuck tiver ve raio of gajos divx defesa
altura metade demais semana continuo minimo tv toast tiveres mapa montar ano quadro sitio
bocado havia continua nome blacksmith meses UZ0 minima caixa GB FBAUP inté glossy sorte
sacar falta ha user takiomaker nevogilde minimo grande embaixada G4 nº resultados muda bootar
Macintel fosse flash dobro like intel modo corre cenas performance pensar caro

5.9 Joana

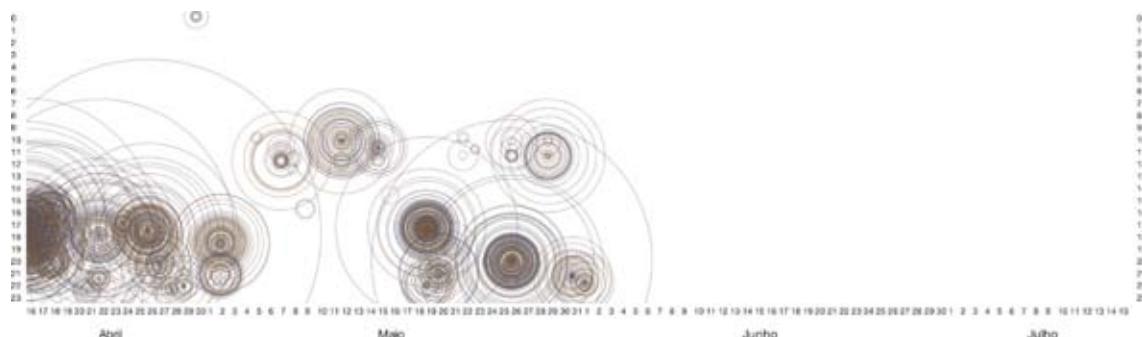
5.9.1 Gráfico comparativo entre itens enviados, recebidos e totais



5.9.2 Gráfico Comparativos de médias



5.9.3 Distribuição temporal das mensagens

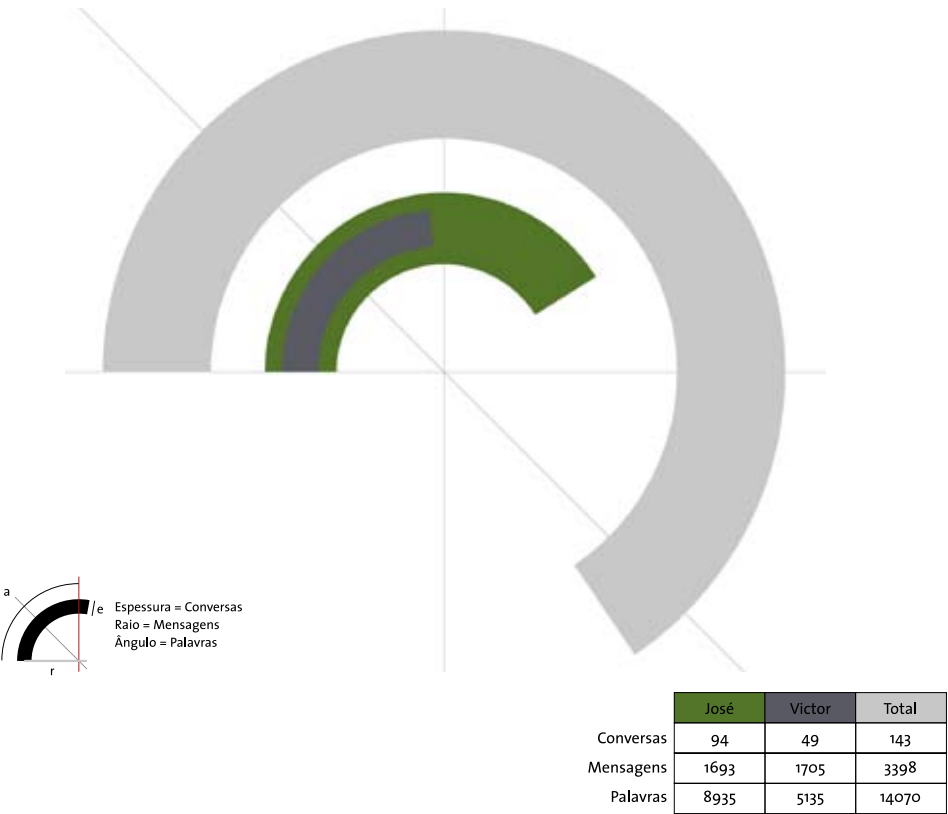


5.9.4 Wordcloud

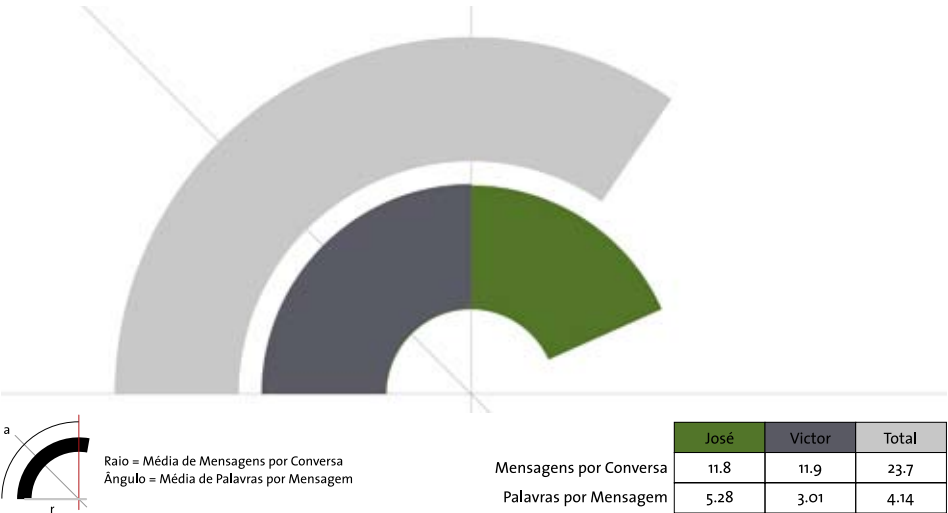
dia vitor tempo coisa falar site zé exacto sério coisas
pessoas noite beijinhos tentar semana revista rita trabalho estranho
desligar bj marco industria penso ideia segunda questão portfolio sinceramente
obrigada lisboa dominios desculpa euros cerca mail verdade tardes amigos cena
entrevista emprego boas pensar reunião dias pag meio anuncio ano preço fica orçamento
porto saber pessoa bolas almoçar zezito marcar sexta termos feira faço falr razão pessoal será
certeza hummm cara jorge trbalhar adriana tempos nova João final semna gente volta lx começar
beijo aep programa inov selecção disse-lhe andar etc seguinte web rio dever falo complicado diferente
parte patroao precisa manha empresa gaia interessante acredito almoço lugar sitio faças pagina abril tipo
iniciar mês mandar fazr montes dinheiro sites minimo funcionário contactar conheço material livro gosta pedir
jantar trbalhos anos pensei orçamneto online texto terça proxima vir uam dito pouquinho factores empresas
atende daqui usar toca novo crer investigar serralves pagar dani dormir perguntar trbalho Maio concurso rede
esperar vejo vontade pedi cabeça janela deixa-me andas dei princio registo alojamento opcoes sol grafico
fraco rancor gaveta mnha roupa chegar capaz engraçado bocado lembrei masi resto ficava solange custa
anuncios relax

5.10 José

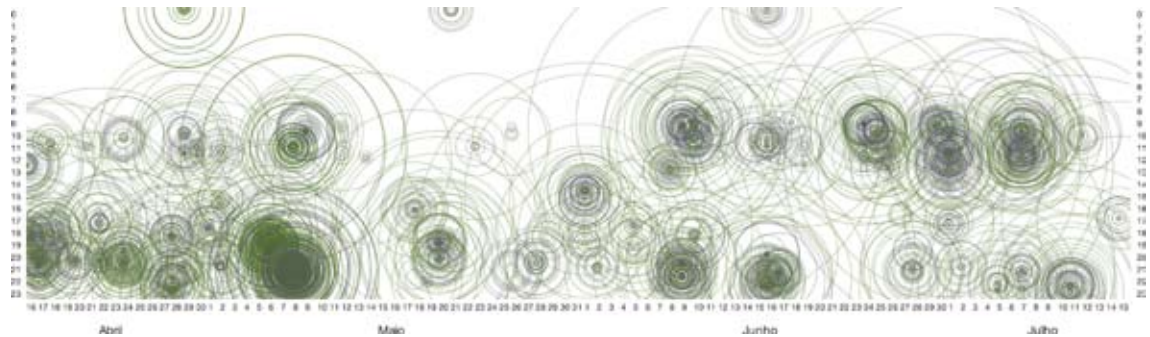
5.10.1 Gráfico comparativo entre itens enviados, recebidos e totais



5.10.2 Gráfico Comparativos de médias



5.10.3 Distribuição temporal das mensagens

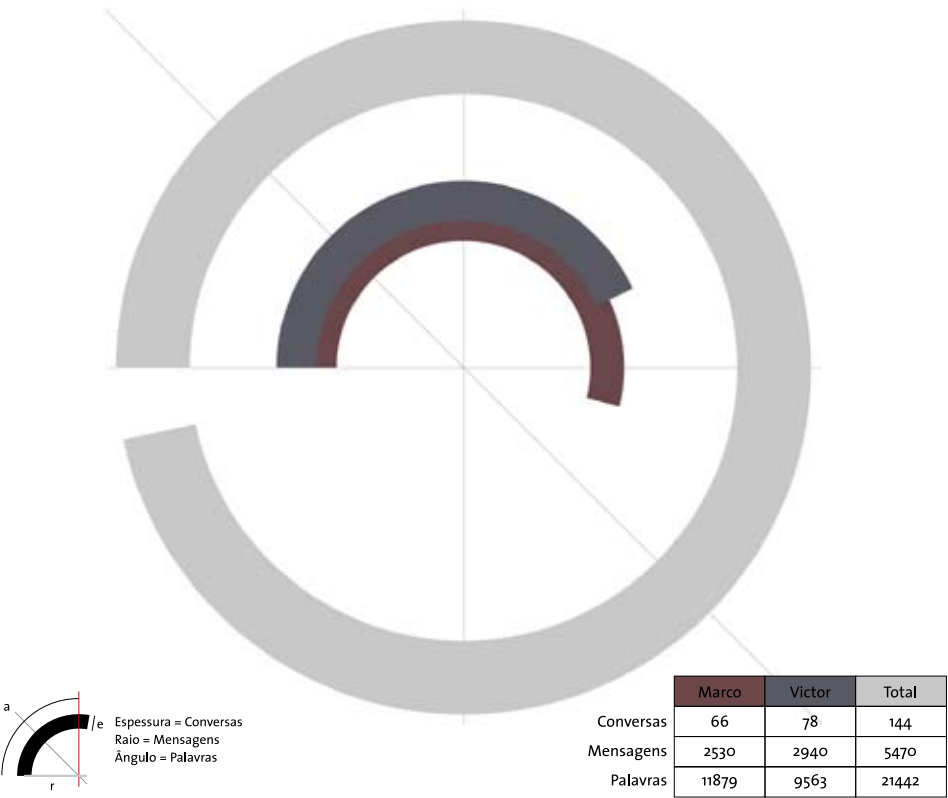


5.10.4 Wordcloud

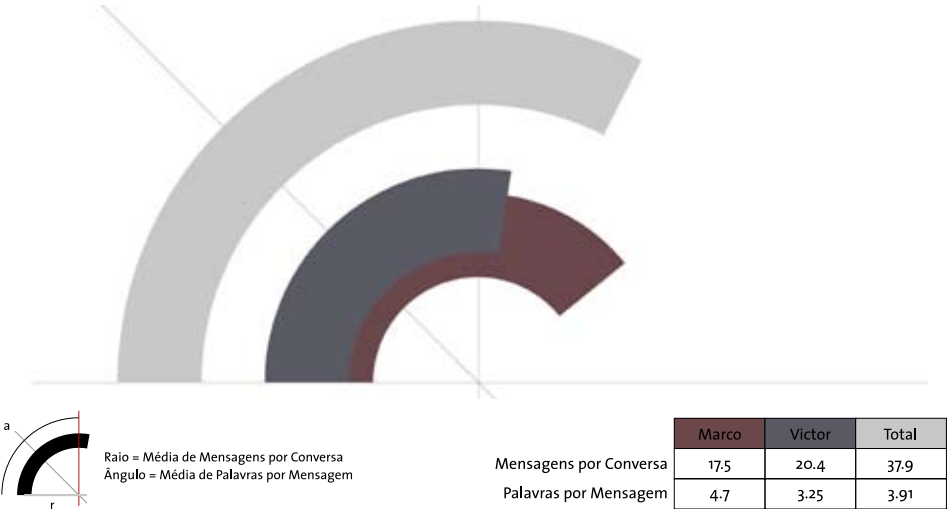
pah cena falar hasta vic preços
tempo tiago coisa marco dia hora boas euros
abraço casa coisas fica certeza pasta pagar semana
tipo saber segunda merda novo fotolitos dias jantar
impressão orçamento mail catálogo ab ligar amanhã cenas site
xl sexta levar etc som cesar preço sergio faço pessoal normal tentar
mandar programa crany volto margem parte manana coto tratar mês
dizes zé ligo almoço cores imagem campos azul tishs café mañana
desligar volta ficheiro dinheiro falta horas shirts amnha sitio resto computer
joana messenger online fico sozinho restaurante noite dados barrigas pensar
belas aparece broas andar verdade levo t-shirt sério cool luego fodido vens bazar
causa dvd manhã fumar semna sabe quantidades tamanho a4 contas print final
tamanhos grande feup guito encomenda combinar falo chegar voltar trabalho
combinado puderes prob sacar trabalhar caralho fazes pensei pequeno manda enquanto
esperar iva mes link fds numero exportar primeira aldeia recursos legionarios t-shirts
quinta fala quarta ar luz pergunta combinamos às hummm portes amigo conversa minima
daniela pessoas merdas começar problema tomar decidir vale daqui poe pera euninos fabrica
pedir encomendas deixei mj festa tamos maximo vinil direito joão material dei sweats

5.11 Marco

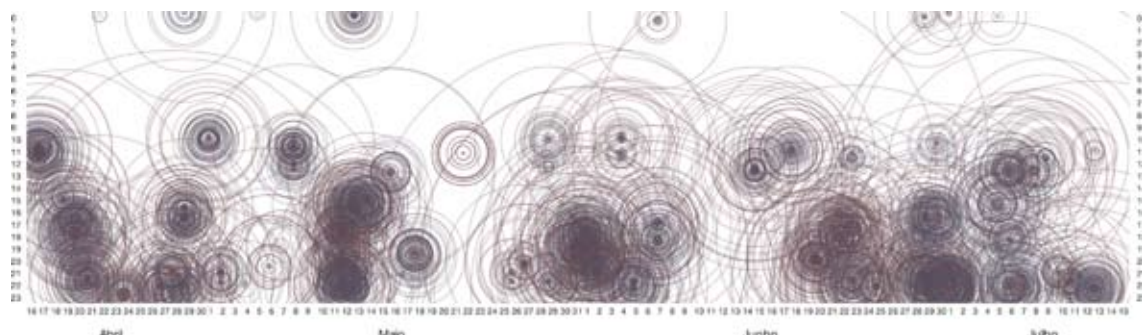
5.11.1 Gráfico comparativo entre itens enviados, recebidos e totais



5.11.2 Gráfico Comparativos de médias



5.11.3 Distribuição temporal das mensagens

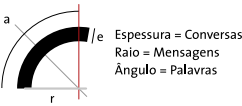


5.11.4 Wordcloud

gajo aldeia tropas
nível ataque gajos boa
mandar crop campos atacar treinar aliança
exacto tempo 2ª pah aldeias daqui amanhã tipo horas
hora segunda coisa ataques dele dia legionarios dias faço
recursos principal noite fica cavalos aumentar merda semana
manha coisas manana checa cena uvd bike raids manda rams pop
muro bombar pessoal amnha conseguir cool guiador certeza xupa falta iron
zona mandei resto fazes fdx euros travian desenvolver conta site curto catas
mensagem normal vale bico consegui volta poder upgrade vukojebina fds defesa
sapatilha mail vir mens meio cabo mando Bounty pretorianos avanço qc ninguem
minutos grande andar atque pasta podia puta demora caralhu min diferença ha
imperions imperians catapultas chegam esperar espera atacado blacksmith caro academia
tiver forum casa perdi maior programa moessa parte oilha clay wood arrnjar cesaries caesaris
momento começar cenas ataca memso chegar acabar enquanto nice ficou coto boas precisas
falar exercito pena problema protecção cagar minimo meia pensar tenta começa gauleses final
guerra foder research fico posta tentar madeira demorar campo armazen scouts barato jogar
bou fuck etc meter cranny semna contas sábado comprar 2º merdas pontape scout to php

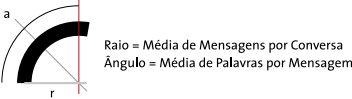
5.12 Miguel

5.12.1 Gráfico comparativo entre itens enviados, recebidos e totais



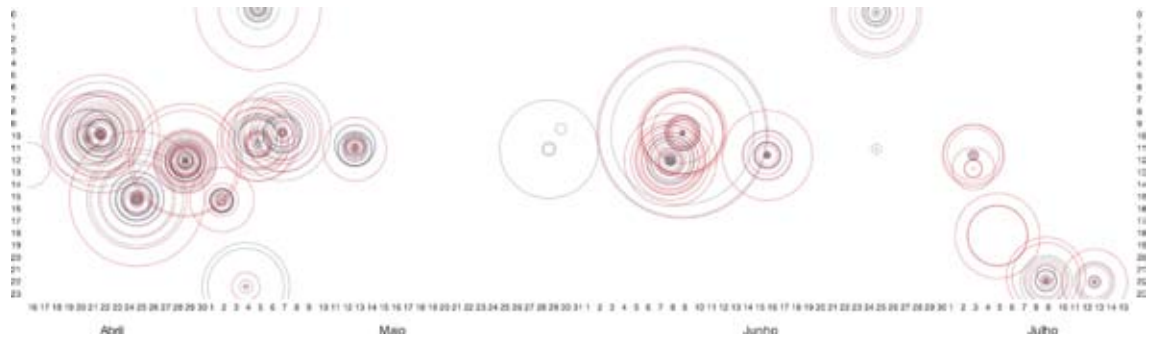
	Miguel	Victor	Total
Conversas	13	12	25
Mensagens	196	200	396
Palavras	757	702	1459

5.12.2 Gráfico Comparativos de médias



	Miguel	Victor	Total
Mensagens por Conversa	7.8	8	15.8
Palavras por Mensagem	3.86	3.51	3.68

5.12.3 Distribuição temporal das mensagens

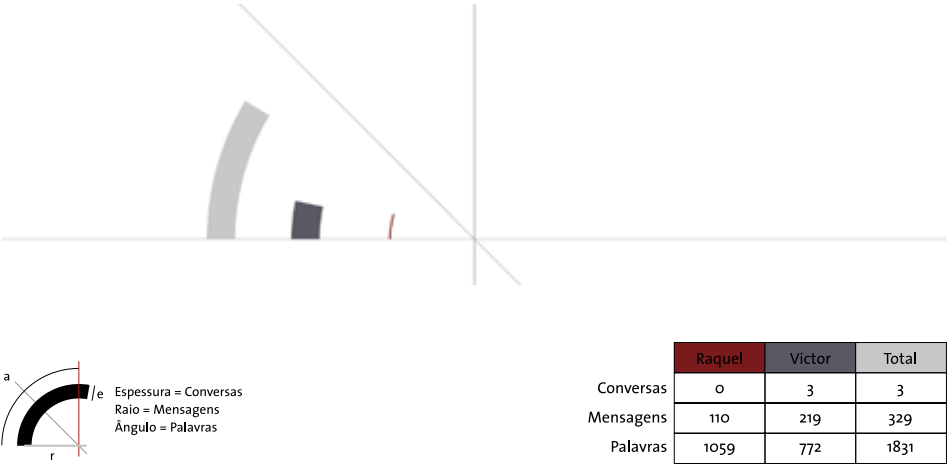


5.12.4 Wordcloud

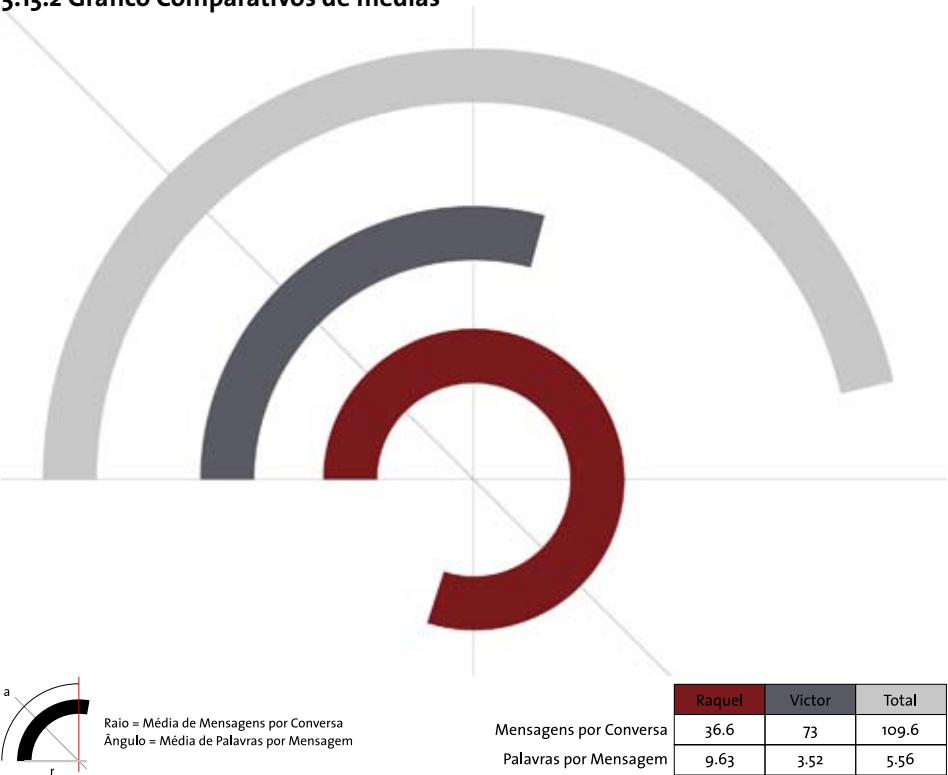
dias mail cena boa xml baixa linha ocupado trabalho metro fica verde tempo
falar coisa mandar thks parte online horas estação oriente chiado telefone tuti santiago
processing ficou cenas som master super boas problem livro cd almoço combinado numeros
momento escadas exacto vale projects lx tempos surgir junho tinhas euros are andado curtir pure data
suporta demorar pass fontlab flash tipo actimel podia altura desligar ftp serviu contacto adriana sacas falas-te
consegui envia hernani semana barcelona arrnjo custom porto benetton pedir tratar gráfica FEUP sonar meo
precisava dia consegues posar morada custumas quarta feira chego volta apnhar podiamos almoças chegaeres toq
vens passa-me telemovél dani telem emprestei-o daniela contactar-me ligo receber telef maneira chegar saio estação
apanhas vermelha alameda trocas descer etic faço sigo então sobes rolantes praça camoes cais sotre chiadao epenultima
baixa-chiado penultima saís chegares chegartes maias estiver agoara andas desenvolver interessado voces falarem
manda-lhe vieres combinar ás arnhas código caro duvido proximos oportunidade sm aviso relação fxas proxmso difícil pegar
concentrar 30 razão expo Frida Khalo memsmo pena ricos you recording nsa nacional altamente SIS cool nível farto
cracks marados po max idem open

5.13 Raquel

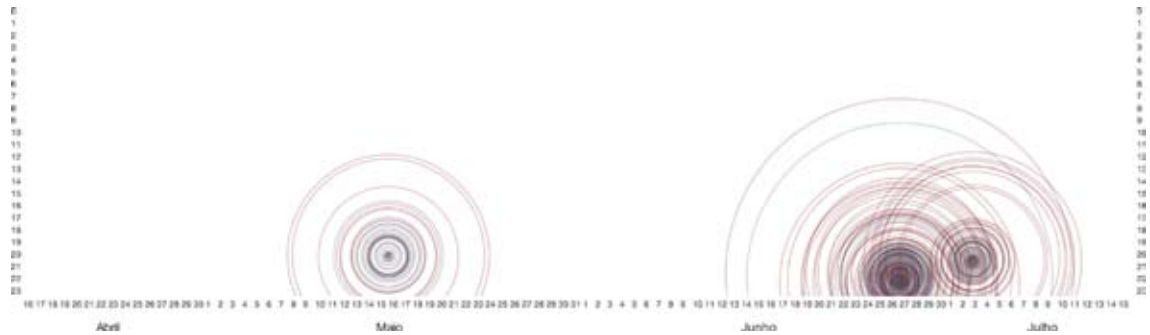
5.13.1 Gráfico comparativo entre itens enviados, recebidos e totais



5.13.2 Gráfico Comparativos de médias



5.13.3 Distribuição temporal das mensagens

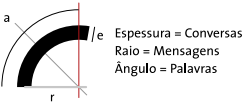


5.13.4 Wordcloud

tempo coisas ano coisa desenho dia anos pessoal começar boa mestrado aulas FBAUP resto meio casa pensei entrar combinar café verdade gente prof ha msn candidaturas exacto trbalho imagino estamos parte contactos interessa teoria putos organizar heheh noite queixar sério problema pessoas pensar vejo sabe cena semna palavras sorte miuda dias FEUP foram proximo pintura curso organização pós-graduação abrir vila real agosto inicio tivesse faço andado bocado fotografar mês dinheiro meses master vagas paper visuais falta mail momento pratica junho tipo tsc teclado maior trocar letras design imensa saudades assunto \$ adoro arrnjar normal fica entregar candidatura resultados concorrer técnicas impressão multimédia Jorge festa quebrar quiseses deixo acordos são feitos serem quebrados irremediável melhor envolve morte prisão fuga fisco drogas interessante ve fiques bués forma relaciono remédio há-de conselho ouviste conselhos cor infelizmente aprende semanas estado férias ferias lixa temso jeito apita terças quartas sábados domingos quilometros terreno neutro refiro-me aproximação Porto referência combinemos facilites querer cortar-me menina cortou-se combinarmos passe arrncar força voltar mecanica volto customizar bicicleta bj andando quae ouvi reomdelar cursos licenciaturas unicas entregram dossiers bolonha atempo fores site descrição escultura fundidos método diria profissionalização disseram-me telefonar saberes tosta

5.14 Rita

5.14.1 Gráfico comparativo entre itens enviados, recebidos e totais



Espessura = Conversas
Raio = Mensagens
Ângulo = Palavras

	Rita	Victor	Total
Conversas	2	2	4
Mensagens	85	86	171
Palavras	224	229	453

5.14.2 Gráfico Comparativos de médias



Raio = Média de Mensagens por Conversa
Ângulo = Média de Palavras por Mensagem

	Rita	Victor	Total
Mensagens por Conversa	21.2	21.5	42.7
Palavras por Mensagem	2.63	2.66	2.64

5.14.3 Distribuição temporal das mensagens

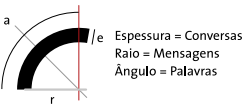
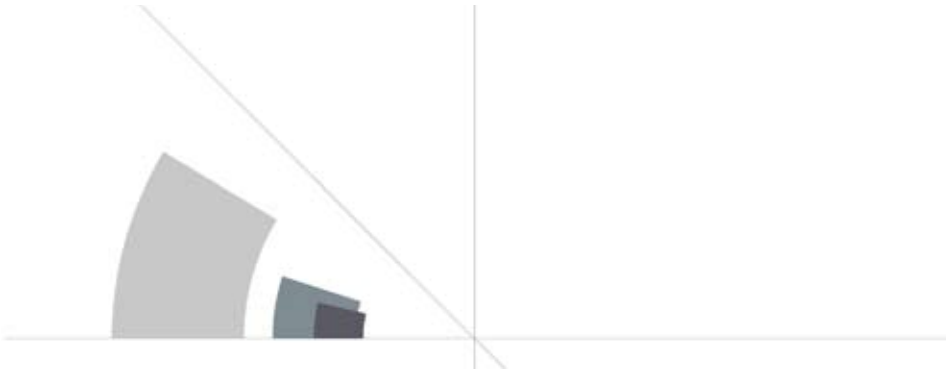


5.14.4 Wordcloud

fontes boa fontbook trabalho contos trbalho cabelos semana duvidas acabar editora saido dia livros
horas vida mes coisas alicia estra a4 A6 acaba perguntar gestao falar aplicacao gerir fonte mac uso suitcase gratuita
encontro problemas embebe fonts torna frete pasta sistema esquisito gente recomenda gratuito cracks serials quiseses
tirar toq precisares software novo avisa photosho cs2 demias fisteus ajudar porto tempo sozinha meia noite encontrei nuno
vocs andam cheios trabalhar sabados dose escolares grafica vale pagam extra rastros porra acredito treco nova aguento
ganhar brncos detesto robo miuda hora cinderela dormir mudar vem feup fica beijo grande parabens confirmou mnha
boneco lindo monstinho parecez estamos tao vergonha estado rato caseiro entregaste tese adiar setembro conseguir
entregar foste feira livro estavam trabalhos serio avisas crom cromos ilustracoes banca barracas passou-me
completamnete lado ano catalogo mostro casa gostar revista devolver aliaice desenhos esboco sara amiga bd devem
guardados perdi fotocopias merda capa procurar comprar wacom fazes fds A5 altura fori net custa valem pena fis
experencia usa-las tomar banho

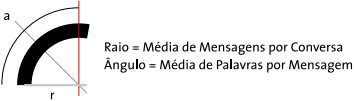
5.15 Rui

5.15.1 Gráfico comparativo entre itens enviados, recebidos e totais



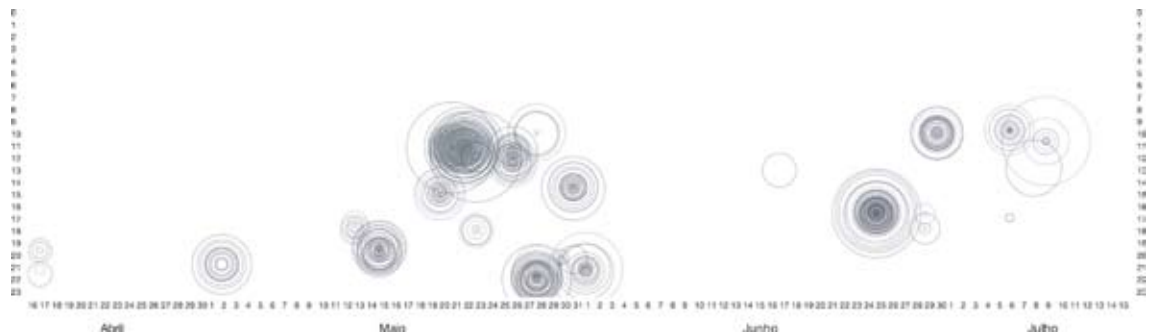
	Rui	Victor	Total
Conversas	20	12	32
Mensagens	358	298	656
Palavras	1070	764	1834

5.15.2 Gráfico Comparativos de médias



	Rui	Victor	Total
Mensagens por Conversa	11.2	9.3	20.5
Palavras por Mensagem	2.99	2.56	2.79

5.15.3 Distribuição temporal das mensagens

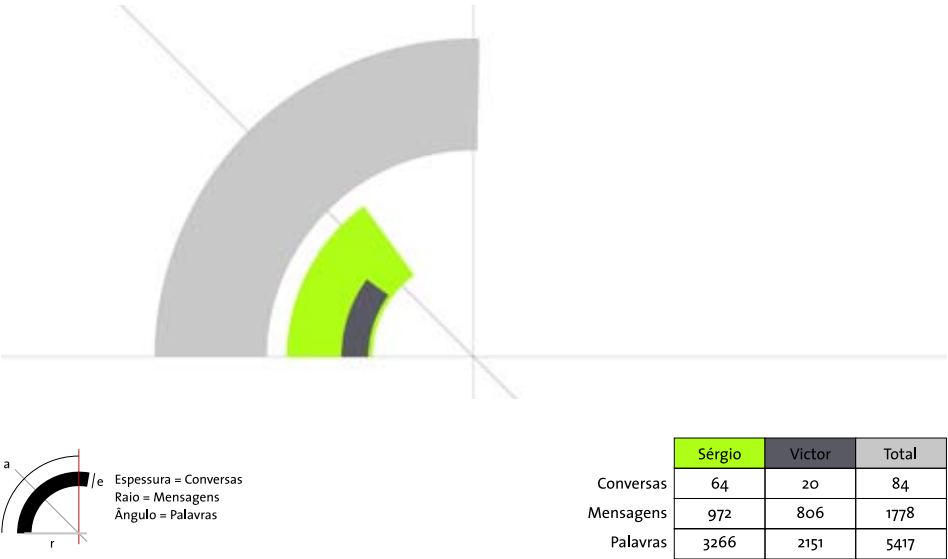


5.15.4 Wordcloud

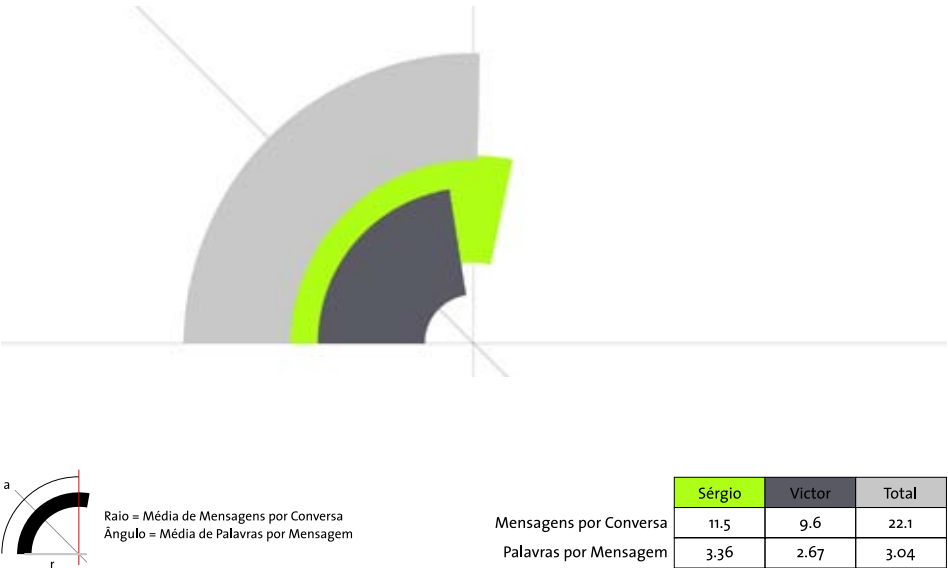
boa marco casa site ha belas gajo pessoal francesinha horas filme fiche
mosso biba curriculum mandar eustaquio y manda meia comé conheço bocado telefona
link gore tao point victor ibook dia graça acabou foto montes jantar atom FEUP semana sexta
sitio mando numero sms escolhe mão conservador voto vamos yeah ten ho saudades molho Knorr
ovo cavalo the works hot pepper fome pão saio prás ucranianas telefone fixo sounds like job hostel
viram cronenberg demais sexo sangue violencia fã inveterado mandaram-me tempos havia utilitários
boss key zombie trailer trailler spider diferente gostei be gone demon witch match chungu boua livros
banda desenhada aqueles epico-heróticos etc bazar pah boas Joana Campos parabens marcamos foste
cenas merda vem emprego maus minima mensagens luego cabeça anos rita ninguem andrea ricardo
confirmados poder disse broa pensei porto nivel zona comprar im apple coreduo ibooks passado mail tentar
conheces daniela ruiva fbaup adriana mandou pagar Londres irmão pensar sousa sigarra ano problema
freehand imagem broas jantarada cinema batalha reabriu sala bebé concerto jornal festa artes multiplas entrada merdas
acabar amnhã serralves moosso quinta tempo combinar bue atrasado mestrado acbar junho começo novo tecnico design
continua part-time

5.16 Sérgio

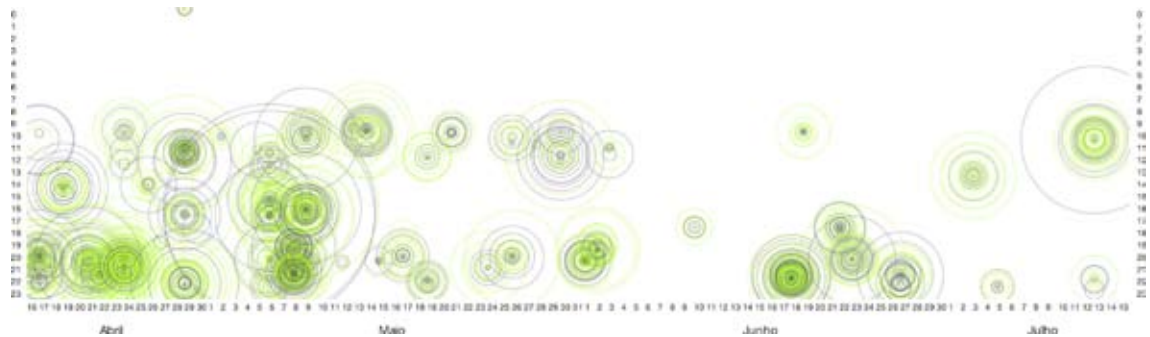
5.16.1 Gráfico comparativo entre itens enviados, recebidos e totais



5.16.2 Gráfico Comparativos de médias



5.16.3 Distribuição temporal das mensagens

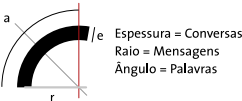


5.16.4 Wordcloud

dia boa boas coisa video noite euros director zé
coisas fica pc amanhã victor apra café tempo mac sistema novo gozar casa
faculdade filme noites senhor peças dias pasta faço manana daqui master pah tipo
comprar tiver metade fodido depos pensar ideia inté merda fodasse minutos preto anda
gajo caixa pago calma cena pergunta joana desligar directa correr fazes ajuda pagar volto
zona tasca copo pessoal pdf only dvd po semana precisa andar freehand disco read site
bigode layout resposta beat cool fico tenta solução bue max applescript dever sofá tardes trabalhar
apple script txt comunicação texto parte vidro semna volta \$ anos ligado trabalho queres começa
normal jantar francesinha flash luz perdi net entrega system cs2 xico funcionar motion ram som silva
ar sexta ocupado monitor venda baixo aparecer lcd certeza falo tinhas receber tiago preocupes abro manda
genero mestrado computador trbalhar pratica cortar ideia vento caralho xuta podre carro qc conjunto vale ano
engenharia partir maio digo-te semnas FESTA ULTIMO cerveja programa juntar masi depende maior codigo
falamos rato velocidade resto princio irmão final perguntei botão mudança resolução forma uam levar projector
hora pifou pastas trbalho

5.17 Tiago

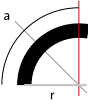
5.17.1 Gráfico comparativo entre itens enviados, recebidos e totais



Espeçura = Conversas
Raio = Mensagens
Ângulo = Palavras

	Tiago	Victor	Total
Conversas	26	28	54
Mensagens	753	545	1298
Palavras	3336	1156	4492

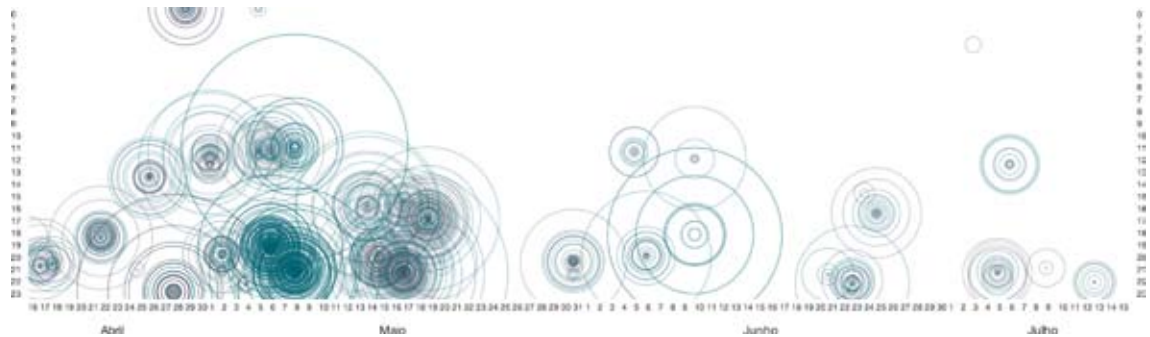
5.17.2 Gráfico Comparativos de médias



Raio = Média de Mensagens por Conversa
Ângulo = Média de Palavras por Mensagem

	Tiago	Victor	Total
Mensagens por Conversa	13.9	10.1	24
Palavras por Mensagem	3.43	2.12	3.46

5.17.3 Distribuição temporal das mensagens



5.17.4 Wordcloud

cena vic zé hasta boas abraço preços certeza falar fotolitos
xl mapa cores merda quarta catálogo orçamento porto t-shirts velhota fica
casa volto preço conta tiago coisa site desenho semana noite sozinho exacto tirar
fac tempo pessoal restaurante princio dia pena cesar cenas barrigas causa margem
tamanho a4 almoço sweats levo tamanhos azul shirts ta-se prob falta saber manana
photoshop trabalho horas jantar hora euros segunda pensei coisas dinheiro levar tipo afinal
havia min tuti semanas pos contas falo mundo rapaz vale pdf link deixei andar tentou ligar tanga
shirt maior liga forem yellow magenta a3 voltar penso convém barato prints caros fosse comida vinil
tishs rosa back etiqueta novo vens etc nome rest combinado ligo piada parte ninguém faço sério
cartão virtual importante sentido norte mandei frente proxima fixes roupa parede medo tratar imagem
querias ausente confusão testes sabia jogo sporting jeito debes memso encomenda desligar inicio limpei podia
tnx metro fodido baixa combinamos compra devem ganhar fazia people olhem tivesse tomar pequeno imagina
vê poe cara mta tripar conheço online ctg funciona minima credito ebay vontade estva bike tv bigode continua
convite taça seguidos trabalhar

5.18 O passo seguinte

Neste capítulo identificamos factores potencialmente subversivos dos dados, categorizamos a minha visão inicial dos contactos e apresentamos os dados recolhidos na forma de gráficos de 3 dimensões circulares, na forma de tabela, representamos a sua distribuição temporal e a *wordcloud* formada.

No capítulo seguinte iremos analisar e discutir os resultados aqui apresentados.

6 Discussão e Análise dos resultados obtidos

A actual versão do MSN Messenger lida com os vários contactos do seu utilizador sem atribuir preferências, pois cada contacto recebe a mesma quantidade de atenção/informação (tal como a informação presença e mensagem de estado), quer estes pertençam ao nosso círculo íntimo, familiar, laboral ou qualquer outro. O MSN Messenger não permite um controlo específico de como estas actualizações são transmitidas, e num certo sentido força o seu utilizador a tratar todos os seus contactos da mesma maneira.

No mundo físico, no entanto variamos o nosso nível de distância pessoal com os outros baseados nas nossas relações sociais e mantemos uma consciência das actividades dos outros, da sua localização e estado ao interagirmos com eles (face-a-face ou mediado). A frequência e o detalhe do qual esta consciência é mantida varia e depende essencialmente do tipo de relação que mantemos com determinado indivíduo.

Um problema detectado quando da análise foi a constatação de que não possuímos dados sobre a localização física de determinado contacto nem do contexto em que este está inserido. Este factor pode ser decisivo e de grande influência sobre o discurso que produzimos.

Em interacções do dia-a-dia podemos gerir a maneira e a quantidade de informação que outros podem recolher sobre nós, como por exemplo, fornecendo indicações físicas da nossa presença, ao selectivamente informar algumas pessoas da nossa localização/actividades, ao escolher atender (ou não) uma chamada, etc.

No entanto não existe nada que distinga directamente a nossa interacção com os vários contactos nos ficheiros de histórico. Para tal é necessário analisar a frequência da comunicação, o rácio entre mensagens enviadas e recebidas e o seu conteúdo.

6.1 Uma primeira Análise – os resultados gerais

Uma das primeiras impressões que podemos retirar do gráfico representativo dos totais gerais (conversas, mensagens, palavras) (ponto 5.3) são três agrupamentos (relativamente ao número de mensagens trocadas representado pela distância ao centro – raio) e dois casos de excepção (um por uma maior troca de mensagens e outro por uma menor) relativamente aos três agrupamentos referidos.

Podemos assim dividir o gráfico em cinco secções partindo de uma maior troca de mensagens para uma menor:

- 1 – Marco (CI+CP, C5)
- 2 – José (CI, C5), Eustáquio (CI, C5) e Adriana (CI+CP, C4)
- 3 – Sérgio (C+CP, C4), Joana (CI, C4), Andreia (CI, C5) e Tiago (C+CP, C4)
- 4 – Rui (AD, C2), Miguel (C, C3) e Raquel (AD+CI, C1)
- 5 – Rita (AD, C1)

considerando que:

CI = Círculo Íntimo

CP= Círculo profissional

C = Colegas

AD = Amigos Distantes

C* = Factor de Conectividade descrito no ponto 5.2

Considerando a distinção inicial dos contactos (5.2), verifica-se que a previsão inicial que os contactos classificados como pertencentes ao círculo íntimo seriam os contactos com os quais a comunicação seria mais intensa estava correcta. Os dois contactos com pertencentes a este “círculo íntimo” com menor troca de mensagens estão ao nível de dois contactos classificados como “colegas” e pertencentes também ao “círculo profissional. Estes contactos também revelam uma classificação “alta” ao nível da sua conectividade variando entre os dois níveis mais altos (o 4 e 5).

Os contactos classificados fora do círculo íntimo apresentam uma comunicação menos “intensa” e podem-se ordenar pela seguinte ordem decrescente: CP, C e AD. Quando existe uma sobreposição de agrupamentos (como por exemplo CP+C) a intensidade da troca de mensagens é maior.

Os dois casos que se situam fora destes agrupamentos (Marco e Rita) pertencem cada um deles aos extremos da classificação. O Marco representa uma sobreposição dos dois agrupamentos principais (ao nível de intensidade), o círculo íntimo e o círculo profissional e o seu factor de conectividade é o mais alto (C5). A Rita representa o outro extremo, é uma “amiga distante” com um factor de conectividade extremamente baixo.

Uma observação importante neste gráfico é que os três itens representados (conversas, mensagens e palavras) não têm propriamente uma correlação directa entre si. Um “qua-

se” igual número de conversas não indica a mesma intensidade ao nível da comunicação efectiva (uso de palavras). Basta comparar por exemplo o número de conversas ocorridas com o Marco e o Eustáquio (144 e 142 respectivamente) e podemos verificar que a diferença de palavras usadas é mais do dobro (10068 para o Eustáquio e 21442 para o Marco). Tal diferença apresentada foi surpreendente para mim uma vez que considero os dois contactos a um mesmo “nível”. Usando a ferramenta descrita no ponto 4.9 para “navegar” por entre as mensagens destes dois contactos, facilmente descobre-se qual o motivo desta diferença. Podia-se sem antes “observar” as mensagens trocadas que a diferença poderia estar relacionada com o facto de o Marco pertencer a dois agrupamentos (CI e CP) mas não, o motivo é bem mais trivial. Observando a *wordcloud* do Marco, destacam-se claramente as palavras “gajo”, “aldeia”, “nível”, “ataque”, “tropas”, que aparecem como as palavras mais usadas. Estas têm uma razão de ser: no início do mês de Abril ambos ficamos “viciados” no jogo *online* Travian [37]. O Travian é um jogo de estratégia medieval, joga-se a partir de um *web browser* e conta com milhares de jogadores a competir simultaneamente. Desenvolvem-se aldeias e tropas, faz-se alianças e combina-se estratégias de ataque e defesa com outros jogadores. Escusado será dizer que tanto eu como Marco pertencíamos a uma mesma “aliança” e estávamos completamente viciados na altura no jogo. Semanas mais tarde conseguimos “chamar” ao jogo outros contactos aqui representados (Eustáquio, José e Andreia) mas que nunca mostraram o entusiasmo e dedicação ao jogo que nós tínhamos.

Este pequeno “caso” serve para demonstrar que a “classificação” que temos das pessoas que nos rodeiam pode não ser directamente representada na comunicação que efectuamos online. Pode bastar um simples interesse (neste caso algo que só podia ser “jogado” online) para fomentar a comunicação e desequilibrar a balança. O gráfico representativo das “médias” neste caso não ajuda, pois a combinação de estratégias *online* implica uma maior atenção de parte a parte e por isso dá azo a conversas mais longas com um maior número de mensagens e palavras por conversa.

Se por um lado o gráfico de médias não pode ajudar a entender este “caso”, ele é uma ferramenta importante para ajudar a homogeneizar os contactos relativamente à sua conectividade. Contactos com um factor de conectividade mais baixo (C2 e C1) poderiam ser facilmente considerados menos importantes (devido ao menor número de conversas, mensagens e palavras) que presumivelmente teriam em comparação com contactos que apresentem uma conectividade superior.

Analisando o número de conversas (espessura), chegamos facilmente também a um divisão em cinco secções aqui descritas por ordem decrescente:

- 1 – Marco (CI+CP, C5), José (CI, C5) e Eustáquio (CI, C5)
- 2 – Adriana (CI+CP, C4), Sérgio (C+CP, C4) e Joana (CI, C4),
- 3 – Tiago (C+CP, C4) e Andreia (CI, C5)
- 4 – Rui (AD, C2) e Miguel (C, C3)
- 5 – Rita (AD, C1) e Raquel (AD+CI, C1)

O factor conectividade parece ser o factor de maior influência no ordenamento dos contactos relativamente ao número de conversas. Excepção feita à Andreia que com um factor

de conectividade máximo (C5) aparece apenas a meio da tabela com um número próximo das conversas do Rui (44 e 32 respectivamente), mas este com um nível de conectividade bastante inferior (C2). À primeira vista tal facto parecia-me estranho. Considero a Andreia alguém pertencente ao círculo mais próximo, amiga de longa data com quem foram partilhados muitos momentos. O contacto no mundo “físico” é mantido regularmente, mas tal parece não ter correlação no uso do MSN Messenger. Existem períodos (ver gráfico temporal) de inexistência de troca de mensagens seguidos de períodos com troca intensa e maioritariamente os contactos ocorrem durante o período nocturno (depois as 21h). Analisando a *wordcloud* e usando a ferramenta de navegação de mensagens, verifica-se que os assuntos abordados são “banais” passando essencialmente por situações vividas no emprego e assuntos relacionados com mestrado. Este é também um dos raros contactos com tendência para que seja eu que procure e efectue contacto um maior numero vezes. Sem que conseguisse explicar tal facto com os dados registados indaguei directamente a Andreia sobre o que ela achava sobre este assunto. A resposta dela veio clarificar um factor muito importante que dificilmente pode ser quantificado. Ela respondeu-me que o “uso” que dá MSN Messenger é o de minimizar a ausência de contacto com determinadas pessoas, ou seja, quanto menos contacto no mundo físico (considerando o seu círculo íntimo) ela tinha, mais ela tentava compensar “digitalmente”. Esta forma de uso do MSN Messenger é bastante diferente da minha, que essencialmente (penso que) é um espelho das minhas relações: quanto mais contacto no mundo físico, mais este é reflectido digitalmente.

6.2 Recebido vs Enviado - Assimetrias Gerais

Os gráficos 5.4 e 5.5 representam respectivamente os totais (conversas, mensagens e palavras) recebidos e os totais enviados. Estes gráficos servem essencialmente para revelar assimetrias gerais que possam existir entre o que recebo comparado com o que envio. Essencialmente verifica-se que existe uma tendência geral para que “envie” menos do que “recebo”.

Muitos de nós estão familiarizados com relações “com que temos de lidar” simplesmente para manter uma boa impressão. Estas relações também se manifestam em sistemas de mensagens instantâneas: tais contactos, deixados em paz, iniciam muitas conversas apesar de tipicamente não serem contactados. Este tipo de relações podem incluir relações de “hierarquia”, do tipo “professor” e “aluno” (pergunta / resposta a duvidas).

Globalmente é difícil de retirar dados relevantes destes gráficos. Para uma classificação de relações com tendências assimétricas é necessário analisar os gráficos gerais e de médias de 1 para 1. Tal análise será feita para cada contacto individualmente mais à frente neste capítulo.

6.3 As Médias

O gráfico 5.6 é representativo da média de mensagens por conversa e de palavras por mensagem. Uma vez que o tipo de uso e o tempo *online* difere de contacto para contacto, é necessário calcular as médias da intensidade da comunicação quando esta ocorre.

A primeira observação que podemos fazer deste gráfico é o obvio destaque que o contacto Raquel tem em relação ao agrupamento geral que formam o resto dos contactos. Apesar de no agrupamento principal a média variar entre as 15,8 e as 42,8 mensagens por conversa (uma diferença significativa) as 119,7 de média com o contacto Raquel “esconde” qualquer outra diferença significativa. De certa este factor não foi surpreendente, sempre achei que sempre que possível dava mais atenção aos “amigos distantes” (principalmente se este têm um factor de conectividade bastante baixo) para me actualizar da sua situação.

Isto parece fazer sentido na medida que as interacções com os nossos “amigos distantes” são incomuns, mas compridas quando ocorrem. Em particular no conjunto de dados, as sessões mais longas de conversa estão associadas com amigos que se mudaram ou estão simplesmente longe fisicamente.

O caso de excepção neste caso parece ser o contacto Rui que apesar de o considerar pertencente ao grupo amigos distantes revela uma média das mais baixas no total dos contactos.

A média referente ao número de palavras é bastante homogénea por todos os contactos (excepção feita à Raquel) mas para uma melhor compreensão dos gráficos das médias é necessário analisar os gráficos referentes ao “enviado” e ao “recebido” separadamente.

Se o número de mensagens por conversas (proporcionando conversas mais longas ou curtas) pode ser relacionado com distância física e factor de conectividade, o número de palavras por mensagem pertence mais ao estilo pessoal de cada um em relação ao uso do MSN Messenger.

Quem usa regularmente serviços de *instant messaging* geralmente consegue identificar um ou outro contacto com tendência para escrever “testamentos” enquanto desesperamos pela resposta. Sempre preferi pessoalmente dividir uma frase que ocupasse três linhas na janela do MSN Messenger por 3 ou 4 mensagens diferentes. Acho que este tipo de procedimento dá origem a uma conversa mais ritmada, mas com uma maior possibilidade de interrupção de pensamento.

Sendo utilizador regular do MSN Messenger com alguns anos de uso, já vivi várias situações que podemos de apelidar “mal entendidos *à la messenger*”. Obviamente uma conversa escrita não possui a mesma fluidez de uma oral não mediada e interpretações erróneas são mais frequentes. Se numa conversa face a face é possível guiarmo-nos através de expressões faciais e corporais, tom e ritmos da voz, numa conversa “textual” o ambiente é homogeneizado e não temos acesso a tais características.

Uma das soluções que uso frequentemente para esclarecer algum mal entendido é o acto de “desacelerar” a troca de mensagens agrupando o que usualmente espalho por três ou quatro mensagens numa só. Desta forma ganho algum tempo de processamento mental na construção da mensagem e permite-me também ter um cuidado superior no que respeita ao uso da gramática.

Esta é a minha solução, e não quero de um caso específico retirar elações sobre estilos de escrita via MSN Messenger. Tais observações e análise de estilo de escrita não fazem parte do foco principal desta dissertação. Este “problema” aliás é pertinente e dava ele próprio razão de ser a uma dissertação dedicada ao assunto. Assim, mais do que analisar o gráfico de médias do que recebo é bem mais pertinente analisar o que envio, uma vez que eu sou a constante e apenas do meu estilo de escrita posso tirar elações devido à possibilidade de comparar possíveis flutuações entre os diferentes contactos.

Observando o gráfico referente às medias das mensagens enviadas (ponto 5.5.3) (e no que se refere à media de palavras por mensagem) pode-se destacar o caso da Andreia e da Joana com as duas médias mais altas, 4.47 e 4.11 respectivamente e do caso do Tiago à qual pertence a média mais baixa 2.12. É necessário neste ponto ver a possível relação da média de envio de palavras por mensagem com o número de mensagens por conversa e com as médias recebidas. Refiro isto porque só assim é possível a tentativa de determinar um factor pelo qual todos nós passamos no dia: o factor de interesse na conversa.

Para determinar o “interesse”, a possível adaptação ao estilo e intensidade de comunicação do contacto, a “procura” (assimetria) da relação é necessário analisar os gráficos 1 para 1 para uma análise mais detalhada que será efectuada nos pontos seguintes deste capítulo.

6.4 Adriana

A Adriana foi categorizada inicialmente como pertencente ao meu “círculo íntimo” e ao “círculo profissional”. Nos resultados gerais (totais) foi o quarto contacto com quem ouve mais “comunicação” quer seja ao nível do número de conversas, de mensagens ou de palavras. Estes resultados são conseguidos mesmo observando (ver gráfico 5.9.3 sobre a distribuição temporal das mensagens) que depois de 9 de Junho a comunicação com este contacto simplesmente não existe. Sendo a localização desta data bastante próxima do início do segundo período de análise (1 de Junho a 15 de Julho), aquele em que os contactos têm conhecimento de que a comunicação está a ser “arquivada”, usei a ferramenta de navegação de “mensagens” para ver que tipo de mensagens foram trocadas neste período de tempo e facilmente encontrei a minha resposta. Na última conversa, parte das mensagens trocadas referiam-se ao período de férias que a Adriana iria iniciar a 12 desse mês (Junho) que se prolongaria até Agosto.

Observando os gráficos comparativos de totais e de médias (ponto 5.6.1 e 5.6.2 respectivamente) verifica-se que existe uma grande equilíbrio entre o “enviado” e o “recebido”.

A qualquer nível observado (conversas, mensagens, palavras) os valores observados são bastante próximos mas com uma predominância maior para que a comunicação efectuada seja maior da parte da Adriana em qualquer dos itens. As médias também se revelam bastante semelhantes (comparativamente), representando conversas fluidas e “equilibradas” e revelam que as conversas com a Adriana são das mais longas observadas (30.2 mensagens por conversa) e acontecem principalmente a partir das 17 horas até ao final da noite (24 horas) e o seu factor de conectividade é semelhante ao meu (C4).

As palavras mais representativas da *wordcloud* fizeram-me recordar os projectos profissionais que mantinha com a Adriana (2 projectos para design e construção de páginas *web*) e curiosamente (ou não) a data de entrega de tais projectos coincide com períodos de comunicação mais intensa (segunda e última semana do mês de Maio).

6.5 Andreia

Como já foi referido anteriormente a Andreia faz parte do meu “círculo íntimo” e o seu factor de conectividade é dos mais elevados. Apesar destes factores os seus totais gerais são bastante inferiores a outros contactos classificados inicialmente ao mesmo nível. Em termos de periodicidade da troca de mensagens esta não é regular e alterna períodos em que a troca de mensagens é inexistente com períodos de troca intensa que ocorrem maioritariamente depois das 21 horas.

Analisando os gráficos comparativos, verifica-se que a “procura” parte maioritariamente da minha parte (apesar de não ser evidente no número de conversas iniciadas) mas é-o na diferença final de palavras usadas (2227 para 3685). Acho que estes números revelam o meu modo de uso do MSN Messenger, o de usar esta ferramenta como um prolongamento ou espelho das relações que mantenho no mundo físico, que neste caso contrastam com o depoimento que a Andreia me forneceu acerca do uso que dá ao MSN Messenger que o utiliza para colmatar a distância que o mundo físico às vezes apresenta.

Estes modos contrastantes podem revelar um dado curioso de que a média de palavras por mensagem que envio para a Andreia é a mais alta de todos os contactos. Posso especular que eventualmente isto se deve a uma tentativa minha de compensar um menor “interesse” na conversa por parte da Andreia.

A *wordcloud* revela palavras relacionadas com a vida profissional, mestrado e curiosamente sobre o facto de que a Andreia foi um dos contactos que ficaram viciados temporariamente no Travian ao aparecem por exemplo as palavras “aldeia” e “tropas” em lugares cimeiros da *wordcloud*.

6.6 Eustáquio

A distribuição temporal (ponto 5.8.3) da troca de mensagens com o Eustáquio é a mais homogénea e regular de todos os contactos. É facilmente distinguível duas linhas temporais: uma que se situa ao fim da manhã (a rondar as 12h) e outra ao início da noite (entre as 21 e as 22 horas). A variação do diâmetro dos círculos que representam as mensagens trocadas é bastante pequena comparada com outros contactos, o nos pode indicar uma troca de mensagens bastante homogénea no que respeita ao seu tamanho. Os círculos azuis tendem a ser maiores que os cinzentos revelando uma média de palavras por mensagem maior por parte do Eustáquio.

Os números mostram que apesar de eu usar menos palavras no total (4767 contra 5301 por parte do Eustáquio) eu produzo um número bastante superior de mensagens (1807 contra 1300). O estilo de escrita que descrevi no ponto 6.3 pode ser facilmente observado neste números.

O Eustáquio é outro contacto pertencente ao “círculo íntimo”, detêm um dos totais mais altos e a comunicação é mantida com regularidade. Mais uma vez a *wordcloud* revela-o como outro utilizador do Travian e relembra-me os jantares de “francesinhas” que regularmente combinamos no “belas” (Café Belas Artes). As conversas com este contacto são tipicamente curtas (comparadas com outros contactos) mas extremamente regulares quer na sua distribuição horizontal (ao longo dos dias) quer na distribuição horária.

6.7 Joana

A comunicação efectuada entre mim e a Joana parece obedecer quase que a uma simetria quando observamos inicialmente os gráficos. A Joana detêm um maior número de conversas iniciadas, mas sou eu que “comunico” mais ao utilizar um maior número de palavras, mas as diferenças existentes podem ser consideradas mínimas como se pode observar pela espessura, diâmetro e ângulo no gráfico (ponto 5.9.1). A Joana faz parte dos contactos considerados pertencentes ao círculo íntimo e o seu factor de conectividade foi considerado semelhante ao meu (C4), o número de conversas é relativamente elevado, mas a média de mensagens por conversa é bastante baixa.

Um factor curioso é de que a média de palavras por mensagem que envio para a Joana é a segunda mais elevada de todos os contactos. A minha relação pessoal com a Joana sempre sofreu de “altos” e “baixos” em que a inconstância era palavra de ordem. Observando as últimas mensagens trocadas precisamente no dia em que eu comunicava que iria “arquivar as mensagens”, “brincávamos” sobre que a partir desse momento ela iria deixar de me “falar” via Messenger. A comunicação cessou realmente a partir desse dia mas por motivos exteriores e de índole pessoal. A média alta de palavras por mensagens trocadas com este contacto penso que se deve a muitas vezes ter que produzir/apresentar pensamentos (mensagens) de uma forma mais pausada e mais “completa” na tentativa de evitar más interpretações.

A *wordcloud* representativa das palavras trocadas (ponto 5.9.4) faz-me recordar um projecto em comum para a implementação de um website que muitas das vezes falávamos do pouco “tempo” que disponha-mos para o finalizar.

6.8 José

O José é o segundo contacto com mais conversas, mensagens e palavras trocadas durante o período de recolha deste projecto sendo que para tal contribui o facto de lhe ter sido atribuído o factor de conectividade mais elevado (C5) e de pertencer ao “círculo íntimo”. Ao analisarmos os gráficos comparativos dos totais (ponto 5.10.1) podemos verificar facilmente uma grande diferença entre o “enviado” e o “recebido”. Essa diferença é evidente em termos de conversas iniciadas (94 contra 49) e na quantidade de palavras usadas (8935 contra 5135). Curiosamente apesar da grande diferença ao nível da quantidade de palavras o meu número de mensagens enviadas é ligeiramente superior. O gráfico temporal mostra um padrão semelhante à verificada com o contacto Eustáquio, duas linhas temporais que ocorrem ao final da manhã e ao início da noite.

A assimetria dos números revelados foi de certa forma inesperada. Sabia que a comunicação com o José não era propriamente “equilibrada” e sempre tive consciência que ele “falava” muito mais do que eu por MSN Messenger. A sua média de palavras enviada por mensagem é a segunda mais alta recebida só atrás do contacto Raquel e parece indicar uma verbosidade maior nas suas comunicações. O José é amigo com quem mantenho um contacto bastante regular no mundo físico tal como saídas nocturnas e aparentemente “cá fora” a relação não parece espelhar o que se passa no MSN Messenger pois a minha visão é de que “eu” é que sou o contacto mais verboso no mundo físico.

Curiosamente as palavras mais destacadas da *wordcloud* são palavras relacionadas com um projecto pessoal que desenvolvi com o José e o Tiago e que por esta altura estava no seu pico e era necessário marcar reuniões saber preços e orçamentos, fazer encomendas, etc.

6.9 Marco

O Marco é o contacto com quem existiu uma maior comunicação no total. É o contacto pessoal mais antigo com mais de dez anos e desde então passamos por imensos projectos a nível da faculdade e profissional juntos. A estes factores alia-se a partilha de vários interesses em comum. Um em particular já foi mencionado anteriormente, o vício que se desenvolveu pela altura da recolha destes dados pelo jogo online Travian. De tal forma este “vício” é relevante que domina totalmente a *wordcloud* gerada. A palavra “aldeia” aparece N vezes e a palavra “tropas” X vezes no total e outras palavras relacionadas com o jogo surgem nas posições mais destacadas da *wordcloud* (ex: atacar, treinar, aliança, crop, etc).

Esta intensidade da comunicação gerada pelo interesse no jogo, foi de certa forma uma surpresa para mim, pois só com a observação da *wordcloud* é que consegui produzir uma consciência do enorme vício que o Travian representava para mim e para o Marco.

Analisando os gráficos comparativos (ponto 5.11.1 e 5.11.2) verifica-se um certo equilíbrio na “comunicação” que a nível da troca de mensagens quer de palavras com um predominância maior por parte do Marco, mas sendo eu aquele que inicia mais conversas. A mancha da distribuição das mensagens é bastante uniforme ao longo dos dias e é mais intensa a partir do meio da tarde prosseguindo até ao final da noite.

6.10 Miguel

A classificação de “colega” dada ao Miguel inicialmente, espelha o que é a nossa relação. Não temos um contacto social grande, não partilhamos projectos profissionais e existe uma grande distância física a separar-nos (Porto – Lisboa). Conheci o Miguel na faculdade e partilhávamos alguns interesses em comum mas sempre foi uma relação distante.

Os resultados observados revelam isso mesmo: poucas conversas, mensagens e palavras trocadas, com um média de mensagens por conversa muito baixa. Estas trocas no entanto revelam um equilíbrio extremo com os números a mostrarem-se muito próximos. A distribuição temporal (ponto 5.12.3) mostra que o contacto é esporádico com conversas curtas. Uma vez que o total de palavras trocadas foi bastante pequeno a *wordcloud* é demasiado uniforme para que possa retirar alguma conclusão sobre algum tema dominante, pois apenas ajuda-me a recordar algumas situações como o planeamento de uma ida a Lisboa.

6.11 Raquel

A minha relação pessoal com a Raquel sempre foi bastante inconstante, principalmente devido ao facto de sermos ex-namorados. A proximidade física existiu apenas durante a duração da licenciatura e finda esta em 2003, que não mais a vi pessoalmente. O contacto que mantemos no MSN Messenger é extremamente irregular, mas quando o fazemos é bastante prolongado. A isto provavelmente ajudará o facto de pouco termos notícias um do outro sem ser desta forma. A *wordcloud* (ponto 5.13.4) recorda “notícias” de cada um, um possível ingresso num mestrado por parte da Raquel, factos sobre o ano que passou a dar aulas no interior, etc.

Das três conversas observadas (ponto 5.13.1) durante o período de recolha, todas foram por mim iniciadas. Este factor aliado aos temas da conversa parece indicar que era “eu” que mais necessitava de “notícias” e que para tal iniciava conversas e fazia um maior número de questões para actualizar-me do que se passava com ela. Esta é a relação mais assimétrica observada nesta recolha. Um maior número de palavras (1059 vs 772) usadas pela Raquel indica que era ela que transmitia mais informação. Os números gerais bastante baixos em

relação a outros utilizadores são um reflexo também do baixo nível de conectividade que a Raquel têm: C1. A troca de mensagens ocorreu sempre ao final da tarde princípio da noite (ponto 5.13.3) que é o horário habitual durante o qual a Raquel às vezes se liga.

A Raquel sempre teve um “português” muito correcto aliado a um uso de gramática exemplar. Estas características reflectem-se na comunicação produzida entre nós. Como se pode observar pelos resultados obtidos (ponto 5.13.1) a Raquel produziu um número bastante superior de palavras usando cerca de metade das mensagens que eu usei. Navegando pelas mensagens raramente encontro o uso de abreviações por parte da Raquel e as mensagens longas são sempre bastante descritivas e com pouco espaço para confusões ou mal entendidos gerados quando habitualmente “cortamos” nas frases que produzimos.

6.12 Rita

Tal como o Marco a Rita é um dos contactos mais antigos. Sempre fomos bons “amigos” mas sem nunca pertencermos ao círculo de intimidade mais próximo um do outro mas o suficiente para sermos mais do que “colegas”. Isso reflecte-se socialmente, pois os contactos sociais sempre foram esporádicos, mas aconteciam. A distância aumentou também a partir do momento em que a Rita se mudou para Lisboa. O contacto entre nós passou apenas a ser efectuado por MSN Messenger e devido à baixa conectividade da Rita (C1) estes foram bastante esporádicos à semelhança da Raquel.

Comunicar com a Rita online sempre foi muito estranho. Quando estamos juntos fisicamente a verbosidade partilhada por nós é bastante grande, mas online a comunicação parece sempre “encravar”. As frases são bastante curtas, os assuntos parecem não desenvolver mas as conversas são “compridas”. Isto reflecte-se num elevado número de mensagens por conversa (ponto 5.14.2) mas com um número bastante baixo de palavras por mensagem. Estes números são partilhados pelos dois. A relação é de um equilíbrio extremo pelos números analisados: cada um de nós iniciou 2 conversas, a Rita enviou 85 mensagens e eu 86 usando 224 palavras e eu 229. As conversas aconteceram maioritariamente ao final da noite.

A *wordcloud* não mostra nada de relevante a não ser a reminiscência de uma dúvida da Rita sobre gestor de fontes tipográficas em Macintosh [38] principalmente devido ao baixo número de palavras trocadas.

6.13 Rui

O Rui é mais um amigo que se mudou e armas e bagagens para Lisboa. Conhecemo-nos durante a licenciatura em design e trabalhamos juntos em vários projectos académicos e profissionais. Durante esses anos fomos amigos bastante próximos e com contacto regular socialmente (saídas nocturnas, jantares, etc.). A partir do momento em que o Rui se mudou para Lisboa foi evidente que este contacto social tinha que diminuir. A *wordcloud* mostra

várias palavras relacionadas com algumas vindas de fim de semana ao Porto por parte do Rui e a consequente marcação de algumas idas às “francesinhas” e trocas de ideias e opiniões sobre “informática”. Lembro-me particularmente de falarmos algumas vezes sobre a compra de um novo portátil por parte do Rui.

Os números mostram que é o Rui o contacto com maior iniciativa para abrir uma janela de conversação (ponto 5.15.1) (20 conversas iniciadas vs 12). Durante o total das conversas produziu mais mensagens e um maior número de palavras mas a diferença não é significativa. Considerando o factor bastante baixo (C2) de conectividade do Rui este é um dos contactos com que mantenho um contacto regular sempre que possível.

A distribuição em termos horários das mensagens mostra que não existe uma janela temporal predilecta, e devido à baixa conectividade do Rui presumo que a comunicação acontece quase sempre que este se ligava independentemente da hora.

6.14 Sérgio

A primeira impressão sobre este contacto ao observar os números é o extremo desequilíbrio existente na comparação do número de conversas iniciadas. Todos nós temos uma relação com a qual não sentimos necessidade em manter contacto, mas este vai acontecendo. Conheci o Sérgio na faculdade. Eu já tinha iniciado o mestrado e o Sérgio estava num dos últimos anos da sua licenciatura. O que nos juntou foi o interesse de ambos por “novas tecnologias”, pelo vídeo e pelo *Vjing* [39].

A *wordcloud* não parece revelar ao observador comum, mas as palavras “vídeo”, “noite” e “director” (*software* na altura usado em algumas das noites em que fizemos de VJ [32]) são palavras-chave para recordar-me de alguns dos assuntos abordados. Ao usar a “ferramenta” descrita no ponto 4.9 para navegar entre as mensagens só veio recordar a limitação de assuntos abordados com o Sérgio.

O desequilíbrio mencionado anteriormente relaciona-se com o facto de a grande maioria das conversas se relacionar com dúvidas da parte do Sérgio sobre aspectos técnicos e de design relativos aos projectos de *Vjing* que tanto eu como ele produzíamos. Assim consigo explicar o grande desequilíbrio na “procura” e á comunicação total (palavras usadas).

A distribuição temporal revela a tendência para que as mensagens trocadas se produzam maioritariamente em dois períodos horários: o final da manhã e o final da tarde / princípio da noite.

6.15 Tiago

O que me chamou mais à atenção inicialmente ao olhar inicialmente para os gráficos foi a grande diferença no número de palavras usadas. Os números do Tiago chegam a ser quase três vezes mais no número de palavras usadas comparativamente comigo. No entanto a diferença entre conversas iniciadas é mínima (26 iniciadas pelo Tiago, 28 iniciadas por mim). Na totalidade das 54 conversas o Tiago efectuou um maior número de mensagens, mas a diferença mais relevante centra-se mesmo na média de palavras por mensagem (ponto 5.17.2).

Caracterizei inicialmente a minha relação com o Tiago como “colega” e pertencente também ao círculo profissional. No entanto nada me indicava estas diferenças de quantidade no uso de palavras. A minha visão inicial era de uma relação equilibrada e nunca senti diferenças no tipo de verbosidade. Ao olhar para a *wordcloud* criada algumas das palavras (preços, fotolitos, cores, etc.) do topo fizeram-me recordar um projecto em conjunto que efectuamos em Abril e Maio. Era um projecto no qual eu não tinha uma participação muito activa no que respeita à gestão do projecto (pedidos de orçamento, produção de fotolitos, reuniões com o cliente, etc.). O Tiago era o elemento do projecto que tratava desses assuntos. Navegando pelas mensagens trocadas verifico que muitas vezes questiono o Tiago sobre o andamento destas questões. As suas respostas eram bastante descritivas e completas e creio ser essa a principal razão da grande diferença no número de palavras usadas. A distribuição temporal parece destacar isso mesmo, mostrando uma maior intensidade até meados de Maio, altura em que o projecto que tivemos em comum findava. Navegando um pouco mais à frente no tempo a conversas parecem “esmorecer” um pouco relacionando-se com assuntos mais banais e sem a intensidade inicial.

6.16 Antes e Depois

Como já foi referido no capítulo 4, a recolha pode ser dividida em dois períodos de tempo iguais, um em que a recolha dos dados não foi comunicada aos contactos e um segundo que estes estavam conscientes da recolha e do seu propósito. O meu objectivo inicial seria tentar analisar se existiria alguma diferença de atitude por parte dos contactos a partir do momento em que estes estavam informados sobre este projecto. A conclusão a que cheguei é que não posso retirar/formular nenhuma conclusão sobre este assunto devido principalmente a dois factores que passo a enunciar:

1. Os ficheiros de histórico do MSN Messenger não dão a indicação de meu tempo de ligação (a que hora que eu ligo e desligo a aplicação) e dos meus contactos. Mesmo que estes ficheiros registassem o meu tempo de “entrada” e saída” seria necessário cruzar esses dados com os dos meus contactos.
2. Mesmo que os dados mencionados no ponto anterior estivessem disponíveis, o ideal seria um tempo de recolha bastante superior (para tentar evitar casos como o da Adriana que foi para férias) e seria necessário recordar constantemente aos contactos que TODA

a comunicação estaria a ser guardada para posterior análise. Refiro este facto pois em conversas com alguns dos contactos, passado algum tempo do anúncio da recolha estes referiam que se esqueciam simplesmente que estavam a ser “gravados”.

6.17 Assentando ideias

Neste capítulo analisamos os números e dados obtidos e o modo em como podemos relacionar estes para compreender se conseguem (ou não) confirmar a classificação inicial dada aos contactos analisados. No capítulo seguinte iremos expor as conclusões que podemos retirar desta análise, os pontos fortes e as falhas deste projecto e o modo em como prosseguir no futuro.

7 Conclusões finais

Quando iniciei este projecto a minha visão inicial dos resultados seria que estes iriam espelhar as interações do mundo físico, considerando que este tipo de comunicação digital (*Instant Messaging*) é realizado geralmente entre pessoas que se conhecem no mundo físico e que as pessoas que fazem parte da minha lista de contactos são todas, sem excepção, pessoas que conheci previamente. Não quero dizer que este facto seja regra, longe disso, mas é suficiente relevante para distinguir os sistemas de *Instant Messaging* de uma tradicional sala de *chat* em que a maioria dos seu intervenientes não se conhecem ou conheceram-se previamente no mundo físico. Assim na maioria das vezes a nossa lista de contactos de MSN Messenger é a “lista” de amigos, familiares, colegas, etc. do mundo físico.

No entanto, foi curioso verificar que nem sempre aquilo que pensava sobre uma determinada relação se reflectia totalmente na comunicação online via MSN Messenger. Como referido anteriormente as relações são o que são num determinado momento de tempo pois encontram-se constantemente em evolução. Mas mais do que esta constante evolução, uma relação é essencialmente a interacção entre duas pessoas e a avaliação que fazemos dela é apenas nossa. É impossível determinar (quantificar) o grau de intimidade que o outro retém sobre nós. A fase de análise dos resultados fez com que reflectisse “bastante” sobre as minhas relações pessoais, numa tentativa de explicar algumas das surpresas que os números parecem apresentar. Deste modo além de funcionar como um diário avivando memórias e criando reminiscências, a quantificação da comunicação (transposição para gráficos) proporcionou-me uma nova perspectiva sobre as minhas relações ao conseguir distanciar-me das minhas ideias iniciais, ideias que se calhar significam o que procuro e não o que “existe”.

Voltando às questões colocadas inicialmente e descritas no primeiro capítulo, os resultados obtidos permitiram-me concluir que a minha concepção das relações que mantenho é reflectida nos dados, mas não de uma forma clara e sem margens para dúvidas. A minha percepção inicial encontra-se mais próxima dos “resultados” obtidos quando se analisa o meu comportamento (comunicação enviada) mas mais dúbia quando se analisa o comportamento dos contactos (comunicação recebida). Estas dúvidas e *nuances* surgem quando analisamos a questão da “procura” (conversas iniciadas) e a relação entre o “enviado” e “recebido”. Se as classificações iniciais de contacto (íntimo, profissional, colega, etc.) é reflectido no “geral” e enquadram-se nessas categorias, quando analisadas ao pormenor estas podem apresentar algumas variações e por em causa o “grau” que atribuíamos à relação.

Quando falo aqui de “grau” não falo de uma medida quantificável e universal que se possa facilmente atribuir. Classificar uma relação é algo extremamente complexo e difícil, mas as observações finais conseguem descrever as diferenças entre as percepções iniciais de “grau” da relação e algumas diferenças observadas como podemos verificar em alguns casos no capítulo anterior.

Em relação ao factor “observação” os dados recolhidos (como explicado mais detalhadamente no ponto 16 do capítulo anterior) não me permitem tirar quaisquer conclusões

acerca da influência que o conhecimento da recolha dos dados possa ter tido sobre o discurso.

7.1 As condicionantes da análise

Ao guardar toda a comunicação produzida no período observado, de certo modo estava a criar o diário “perfeito” da minha vida social digital no MSN Messenger. A percepção inicial era de que se quantificasse e analisasse os dados teria uma boa consciência do tipo de relações que desenvolvo, mas existem demasiados factores condicionantes para que tal aconteça. O mais importante, na minha opinião (e que foi referido anteriormente), é o facto de apenas dispor dos meus *logs* e de não ter acesso aos *logs* dos outros contactos. Uma vez que os dados analisados são a representação da interacção de 1 interlocutor para *n* contactos, a comparação da comunicação que os contactos “enviam” contém mais limitações do que a comparação dos dados que eu envio, principalmente no que diz respeito à análise das médias. Um estudo comparativo em “rede”, em que cada interveniente fornecesse os seus *logs* teria uma validade bastante superior devido à possibilidade de cruzamento de dados.

As conclusões que se podiam retirar sobre o tipo de discurso (assuntos abordados, gramática, etc.), a sua frequência e intensidade estão extremamente condicionadas pelo factor da localização física. Obviamente que o uso do MSN Messenger no trabalho terá diferenças em relação ao uso deste num ambiente mais privado como o aconchego da nossa casa. Estas diferenças podem ser mais ou menos relevantes consoante o tipo de trabalho que desenvolvemos, o ambiente em que estamos inseridos, etc. Devido ao facto de o MSN Messenger não revelar a situação/localização “física” em que estamos inseridos num determinado momento torna mais difícil a análise sobre as várias características do discurso e a comparação entre vários indivíduos. Quanto mais “fechado” e controlado for o factor localização e contexto, uma melhor análise poderá ser efectuada. Um exemplo concreto seria a possível análise de relações em equipa no ambiente de trabalho. Os *logs* produzidos pelos vários elementos teriam em comum o contexto físico e social que tornaria as conclusões e análises sobre estes mais fiáveis.

O período de recolha é um elemento importante para projectos deste género. Quanto maior for o período, mais elementos (dados) são produzidos para produzir um retrato mais fiel. Este projecto centra-se sobre os dados produzidos durante quatro meses. Uma vez que o factor de localização e contexto referido aqui não pode ser identificado, era necessário um maior período de recolha para evitar dados que pudessem causar conclusões erróneas como os que podem ser produzidos devido ao período de férias. O ideal seria a análise de 1 ano reflectindo o ciclo anual que muitas vezes a nossa vida sofre. Este período “ideal” de análise ajudaria também a combater outros factores condicionantes como a disponibilidade psicológica. No entanto, como referido várias vezes anteriormente, as relações encontram-se sempre numa constante evolução e qualquer dado produzido (ou falta deste) é sempre relevante para a análise de um determinado momento.

Outros problemas a enfrentar neste tipo de análise de relações pessoais são os três problemas iniciais que referi no capítulo 5: a promiscuidade do sistema, o problema da observa-

ção e a subjectividade da análise que são quase inultrapassáveis em projectos essencialmente introspectivos como este de 1 para n em que a promiscuidade do sistema não pode ser repartida uma vez que pode ser considerado um sistema “teocêntrico”. O problema da observação pode provavelmente ser minimizado com o prolongamento do período de observação (quanto mais tempo passar menor será a probabilidade de controlarmos as nossas acções numa tentativa de subversão de dados), assim como evitar desenvolver as ferramentas de análise de *logs* sobre dados em período de recolha uma vez que uma observação prévia do que “está a decorrer” poderá influenciar o restante tempo de recolha. O ideal seria desenvolver as ferramentas de análise sobre dados “falsos” ou sem relação com os dados observados.

7.2 Possíveis melhorias técnicas ao processo

Para evoluir o processo de análise o primeiro item a modificar seria o modo em como o MSN Messenger produz os seus registos de comunicação. Os registos só se produzem quando existe uma comunicação activa, mas o ideal seria que os registos estivessem constantemente a guardar dados como entradas e saídas dos contactos (e nosso) e a indicação da disponibilidade destes (se estão ocupados, ausentes, etc.) através do item de “estado” que o MSN Messenger permite mostrar. Mesmo sem haver um cruzamento de dados (acesso aos registos dos outros contactos) estes elementos seriam uma grande ajuda para a determinação de “conectividade” e disponibilidade dos diferentes contactos assim como da nossa. Teríamos duas opções a contemplar, criar um *plugin* para o MSN Messenger (ou qualquer que seja o programa de Instant Messaging observado) ou desenvolver um sistema *opensource* com essa função e que mais facilmente poderia ser modificado e ajustado sem estar dependente de *upgrades* de clientes comerciais por nós não controlados.

A evolução seguinte seria no outro extremo das fases processuais, as visualizações. Um único interface dinâmico e personalizável ao nível da representação dos diversos gráficos e da sua “plotagem” final. Seria de mais fácil utilização em relação à sua presente forma em que estas funções se encontram todas separadas. O dinamismo que falo, seria por exemplo, a possibilidade de especificar os gráficos de distribuição temporal de uma forma dinâmica em conjugação com a ferramenta de “navegação” de mensagens produzida neste projecto. Este upgrade iria facilitar assim a identificação das mensagens produzidas nas quebras do padrão comunicacional para posterior análise.

Estas modificações seriam relevantes no contexto a que este projecto se propôs trabalhar: os registos de uma só pessoa para uma possível caracterização das suas relações pessoais.

7.3 Caminho a seguir

A premissa de “caracterização de relações pessoais” a partir da quantificação da comunicação é apenas um dos muitos temas que podemos trabalhar a partir dos dados recolhidos. Um possível ponto de partida (que foi apenas abordado ligeiramente neste projecto) seria o uso da gramática para caracterização de relações assim como da própria evolução da escrita digitalmente. No projecto que realizei foquei a minha atenção na quantificação da comunicação e menos no seu conteúdo. As referências que efectuei a “palavras-chave” das *wordclouds* produzidas serviam apenas como “recordações” de situações vividas que me permitiam contextualizar determinados períodos mais ou menos intensos da comunicação.

Num possível *spin-off* deste projecto, como seria a possível caracterização de relações num ambiente de trabalho em equipa, a análise das *wordclouds* não poderia apenas limitar-se ao seu uso como “recordações” e teria essencialmente de ter um papel mais activo na determinação de relações e a sua evolução. Falo especificamente da possibilidade de procura de palavras-chave específicas e a determinação de que outras palavras surgem associadas mais vezes a esta.

Um exemplo concreto e contemporâneo a esta dissertação é o projecto MoodViews [40] que trabalha sobre dados de *blogs* mas que essencialmente sobre associação e procura de palavras-chave para determinar, prever e contextualizar o “humor” dos conteúdos de texto existentes em *blogs*.

“Our research aim is to develop novel methods for searching, discovering and retrieving blogs. We believe that non-factual aspects of blog entries such as moods are an important part of what makes people read and navigate around blogs. In addition to Moodsignals we are currently working on a number of new tools to track, explore, and analyze moods.”

www.moodviews.com, Agosto 2007

A recolha destes dados teria sempre que passar pelo cruzamento dos diversos *logs* produzidos por cada elemento a observar para uma melhor caracterização destes.

À medida em que escrevo estas últimas palavras sobre um possível *spin-off* apercebo-me de que algo me inquieta. Vêm-me à memória o 1984 [41] de George Orwell pois “assusta-me” a facilidade com que hoje em dia podemos exercer vigilância e monitorização sobre outros. Esta facilidade não deixa apesar de tudo de ser especialmente tentadora.

Referências Bibliográficas

- [1] - **Tufte, Edward R.** - The Visual Display of Quantitative Information [Graphics Press 1992]
- [2] - **Ware, Colin** - Information Visualization: Perception for Design [Morgan Kaufman Publishers 2000]
- [3] - http://en.wikipedia.org/wiki/Instant_messaging [10/09/2006]
- [4] - www.avaloncity.com/info/fprefect/inet/desc/talk.htm [08/05/2007]
- [5] - www.icq.com [10/09/2006]
- [6] - www.mirc.com/irc.html [10/09/2006]
- [7] - www.msn.com [02/08/2006]
- [8] - <http://messenger.yahoo.com/webmessengerpromo.php> [10/09/2006]
- [9] - www.aim.com [10/09/2006]
- [10] - www.ceruleanstudios.com [10/09/2006]
- [11] - <http://sourceforge.net/projects/gaim/> [10/09/2006]
- [12] - www.miranda-im.org [10/09/2006]
- [13] - www.jabber.org [10/09/2006]
- [14] - <http://about.reuters.com/productinfo/messaging/> [10/09/2006]
- [15] - **Nardi, B. A. / Whittaker, S. / Bradner, E.** - Interaction and outeraction: instant messaging in action. In Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work (Philadelphia, Pennsylvania, United States). CSCW '00. ACM Press.
- [16] - Comunicação pessoal via e-mail com os serviços técnicos do MSN em Novembro de 2006
- [17] - **Tat, Annie / Kruger, Russell / Carpendale, Sheelagh / Dunning, Alan** - Growing Message Post - Visualizing Temporal Aspects of Postings - <http://pages.cpsc.ucalgary.ca/~annie/PeopleVisWebsite/Annie2/annie2.htm> [10/04/2007]
- [18] - **Kerr, Bernard** - Thread Arcs: An Email Thread Visualization. In Proceedings of the 2003 IEEE Symposium on Information Visualization, IEEE Computer Science Press, Los Alamitos, CA, 2003.

[19] – **Viégas, Fernanda B. / Donath, Judith S.** - Digital Artifacts for Remembering and Storytelling: PostHistory and Social Network Fragments. In Proceedings of the 37th Hawaii International Conference of System Science, IEEE Computer Society, Los Alamitos, CA, 2004

[20] - **Viégas, Fernanda B. / Golder, S. / Donath, J.** - Visualizing email content: portraying relationships from conversational histories. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Montréal, Québec, Canada, April 22 - 27, 2006

[21] – **Viégas, Fernanda B. / Donath, Judith S.** - Chat Circles, In Proceedings of CHI 1999, ACM Press.

[22] - **Tat , Annie / Carpendale , Sheelagh** - CrystalChat: Visualizing Personal Chat History. In Proceedings of the 39th Annual Hawaii International Conference on System Sciences 2006.

[23] - <http://csep.psyc.memphis.edu/cohmetrix/readabilityresearch.htm> [10/05/2007]

[24] – www.gmail.com [10/05/2007]

[25] - <http://www.w3.org/XML/> [22/03/2007]

[26] - www.mysql.com [17/03/2007]

[27] - www.php.net [17/03/2007]

[28] - Nardi, B. / Whittaker, S. / Bradner, E. - Interaction and Outeraction: Instant Messaging in Action. In Proceedings CSCW'2000, ACM Press

[29] – Isaacs, Ellen / Walendowski , Alan / Whittaker, Steve / Schiano , Diane J. / Kamm, Candace - The Character, Functions, and Styles of Instant Messaging in the Workplace, In Proceedings CSCW 2002, ACM Press.

[30] - <http://www.processing.org/courses.html> [15/04/2007]

[31] - <http://benfry.com/isometricblocks/> [17/05/2007]

[32] - www.adobe.com [09/07/2007]

[33] - www.iconinteractive.com/tools/wordcloud/ [02/07/2007]

[34] - <http://www.cie.co.at/cie/> [21/05/2007]

[35] - **Dillencourt, Michael B. / Eppstein , David, / Goodrich, Michael T.** - Choosing Colors for Geometric Graphs via Color Space Embeddings - Lecture Notes in Computer Science, Michael Kaufmann and Dorothea Wagner, ed., Springer-Verlag, 2006
ACM Computing Research Repository.

[36] - <http://www.ajc.pt/cienciaj/n33/atomo2.php> [18/07/2007]

[37] - <http://www.travian.com> [19/07/2007]

[38] - www.apple.com [05/08/2007]

[39] - <http://pt.wikipedia.org/wiki/VJ> [14/08/2007]

[40] - <http://moodviews.com/> [01/09/2007]

[41] - **Orwell, George** - 1984 [Plume, 2003]

Bibliografia Geral

Abrams, Janet / Hall, Peter - Else/Where: Mapping new cartographies of networks and territories [University of Minnesota Press 2006]

Appan, Preetha / Shevade, Bageshree / Sundaram, Hari - Interactive Visualization and Content Analysis of Instant Messaging Networks, Arts Media and Engineering Program, ASU, Dec. 2004.

Brouwer, Joke / Mulder, Arjen - Information is Alive - Art and theory on archiving and retrieving data [V2_/Nai Publishers 2003]

Card, Stuart / Mackinlay, Jock D. / Schneiderman, Ben - Readings in Information Visualization: Using Vision to Think [Morgan Kaufmann Publishers, 1999]

Dieberger, A. / Erickson, T. - Visualizing Chat: Issues and Prospects, submission to CHI 2006 workshop on social visualization, Montreal, Canada, April 2006.

Dieberger, A. / Svensson, M. - From Social Navigation to the Social Web, The Adaptive Web: Methods and Strategies of Web Personalization [Springer 2006]

Few, Stephen - Information Dashboard Design, [O'Reilly 2006]

Fry, Benjamin Jotham - Computational Information Design, Submitted to the Program in Media Arts and Sciences, School of Architecture and Planning, in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the Massachusetts Institute of Technology, April 2004

Greene, Rachel - Internet Art [Thames & Hudson 2004]

Hatje, Cantz – Code, the language of our time, [Ars Electronica 2003]

Heitlinger, Paulo - O Guia Prático da XML XML Conceito, exemplos, prática e aplicações da linguagem universal” [Edições Centro Atlântico 2001]

Lunenfeld, Peter - The Digital Dialectic, [MIT Press 2000]

Maeda, John - Creative Code [Thames & Hudson 2004]

Manovich, Lev - The Language of New Media, [MIT Press 2001]

Manovich, Lev - www.manovich.net [26/05/2007]

Mutton, Paul - “Inferring and Visualizing Social Networks on Internet Relay Chat,” Eighth International Conference on Information Visualisation (IV'04), 2004

Negroponte, Nicholas - Ser Digital [Caminho 1996]

Paul A. Fishwick - Aesthetic Computing [MIT Press 2006]

Saper, Craig J. - Networked Art [University of Minnesota Press, 2001]

Spence, Robert - Information Visualization [Addison Wesley 2000]

Tufte, Edward R. - Envisioning Information [Graphics Press 1990]

Tufte, Edward R. - The Visual Display of Quantitative Information [Graphics Press 1992]

Wilson, Stephen - Information Arts, intersections of art, science and technology [MIT Press 2002]

Anexos

Toda esta dissertação assenta sobre um projecto de introspecção desenvolvido para armazenar, pesquisar e visualizar os dados recolhidos. Para tal, foram desenvolvidas ferramentas específicas que cumprissem com estas funcionalidades. Partiu-se assim para um “entendimento” da tecnologia, das relações entre o homem e a “máquina” e da forma como estas afectam o ser humano na sua personalidade e forma de estar.

O novo desafio é desenvolver a estrutura de informação e espaço que habitamos em formas visuais, um desafio semelhante ao que os cartógrafos enfrentaram alguns séculos atrás. A base para a visualização de informação está na maneira como a organizamos visualmente: em forma de gráficos. Se os elementos estão organizados segundo uma relação estruturada, o gráfico serve para mostrar esse relacionamento. No campo do design, existe uma aproximação ao problema sob o ponto de vista do desenvolvimento do conceito e refinamento visual e por outro lado no campo da engenharia a procura da resolução concentra-se na construção de algoritmos complexos de modo a poder reunir, sintetizar e estruturar a informação.

Mas, e no campo da arte?

As páginas seguintes representam “estruturas de comunicação” produzidas através do “mapeamento” temporal das mensagens sem a constrição de uma grelha. São “imagens” autónomas, que representam a dicotomia de uma ferramenta capaz de processar, organizar informação e simular o mundo através da utilização do algoritmo, e do ser humano, um ser sensível com atitude estética, e sem o qual a primeira não teria qualquer função.

PORTO, Setembro de 2007

