



UNIVERSIDADE FEDERAL DO ESTADO DO RIO DE JANEIRO
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA

CARACTERIZAÇÃO DA PARTICIPAÇÃO DE USUÁRIOS EM SESSÕES
EDUCACIONAIS DE BATE-PAPO

Edmilson Barcelos Rocha

Orientadores
Mariano Pimentel
Morganna Carmem Diniz

RIO DE JANEIRO, RJ – BRASIL
JUNHO DE 2019

CARACTERIZAÇÃO DA PARTICIPAÇÃO DE USUÁRIOS EM SESSÕES
EDUCACIONAIS DE BATE-PAPO

Edmilson Barcelos Rocha

TESE APRESENTADA COMO REQUISITO PARCIAL PARA OBTENÇÃO DO
TÍTULO DE DOUTOR PELO PROGRAMA DE PÓS-GRADUAÇÃO EM
INFORMÁTICA DA UNIVERSIDADE FEDERAL DO ESTADO DO RIO DE
JANEIRO (UNIRIO). APROVADA PELA COMISSÃO EXAMINADORA ABAIXO
ASSINADA.

Aprovada por:

Mariano Pimentel, D.Sc. - UNIRIO

Morganna Carmem Diniz, D.Sc. - UNIRIO

Hugo Fuks, Ph.D - PUC-Rio

Marco Aurélio Gerosa, D.Sc. - Northern Arizona University

Sidney Cunha de Lucena, D.Sc. -UNIRIO

Sean Wolfgang Matsui Siqueira, D.Sc. - UNIRIO

RIO DE JANEIRO, RJ – BRASIL
JUNHO DE 2019

RESUMO

ROCHA, Edmilson Barcelos. **Caracterização da participação de usuários em sessões educacionais de bate-papo**. UNIRIO, 2019. 123 páginas. Tese de Doutorado. Departamento de Informática Aplicada, UNIRIO.

Nos últimos anos, tem-se observado o crescimento na oferta de cursos, entre eles cursos superiores, na modalidade a distância. A tecnologia proporcionou a criação de ambiente de educação on-line composto de uma gama de recursos cada vez mais sofisticados e atraentes para o usuário, como as redes sociais e uso de técnicas de jogos (gamificação). Embora o número de matrículas nessa modalidade seja crescente, a evasão desses cursos continua muito alta. Na literatura, encontram-se orientações para educadores diminuírem a evasão em cursos on-line como o uso de meios síncronos em conjunto com meios assíncronos de comunicação para aumentar o engajamento e possibilitar um novo modelo pedagógico, focado na interatividade. O uso desses meios de comunicação para educação estão suscetíveis a problemas como a dificuldade de acompanhamento de uma discussão on-line síncrona. Um dos fatores que ocasionam esse problema é o tamanho do grupo. Ao se consultar a literatura, são encontrados dados divergentes sobre a quantidade de participantes em bate-papo no contexto educacional. Garantir uma participação de qualidade é fundamental para o avanço da educação on-line e, por isso, a presente pesquisa tem por objetivo caracterizar matematicamente a participação dos usuários em sessões educacionais de bate-papo, para que seja possível melhor dimensionar os grupos de educação on-line. Para atingir esse objetivo, foi realizada uma pesquisa quantitativa com o uso do método modelagem matemática. Foram realizados estudos com turmas de educação on-line, o que permitiu tanto realizar a modelagem, quanto avaliar a adequação dos modelos desenvolvidos.

Palavras-chave: Bate-papo; Tamanho do grupo; Participação on-line; Modelagem matemática; Teoria de Filas.

ABSTRACT

In recent years, there has been an increase in the number of courses offered in distance learning, including those in higher education. Technology has provided the creation of an online education environment comprised of a range of resources increasingly sophisticated and attractive to the user, such as social networks and the use of gaming techniques (gamification). Although the number of enrollments in this modality is increasing, the evasion rates of these courses continue very high. In the literature, there are guidelines for educators to reduce evasion in online courses such as the use of combined synchronous and asynchronous means of communication to increase engagement and enable a new pedagogical model focused on interactivity. The use of these media for education are susceptible to problems such as difficulty in following a synchronous online discussion. One of the factors that causes this problem is the size of the group. When consulting the literature, we find divergent data on the amount of participants in chats in the educational context. Ensuring a quality participation is fundamental for the advancement of online education and, therefore, the present research aims to mathematically characterize the participation of users in educational chat sessions, so that it is possible to better dimension the groups of online education. To achieve this goal, a quantitative research was carried out using the mathematical modeling method. Studies were carried out with online education classes, which allowed both to perform the modeling, and to evaluate the adequacy of the developed models.

Keywords: Chat; Group size; Online participation; Mathematical modeling; Queuing Theory.

Sumário

1.	Introdução.....	1
1.1	Contexto: a expansão da EAD e o uso de bate-papo.....	1
1.2	Problema e questão de pesquisa	5
1.3	Modelo conceitual preliminar (conjecturas iniciais)	7
1.4	Metodologia desta pesquisa.....	8
1.5	Histórico da pesquisa.....	9
1.6	Organização da escrita.....	10
2.	Problema: até quantas pessoas devem participar de um bate-papo educacional?..	11
2.1.	Bate-papo	11
2.2.	Bate-papo na educação	12
2.3.	Participação em bate-papo educacional	14
2.4.	Pesquisas sobre fatores que influenciam a participação em bate-papo educacional	15
2.5.	Divergências sobre a quantidade de pessoas que devem participar de uma sessão de bate-papo	17
3.	Metodologia de pesquisa	19
3.1.	Paradigma epistemológico	19
3.2.	Método científico	21
3.2.1.	Definição do problema	22
3.2.2.	Construção do modelo.....	22
3.2.3.	Solução do modelo.....	23
3.2.4.	Validação do modelo.....	24
3.2.5.	Implementação da solução	24
3.3.	Modelos quantitativos: modelo da participação em bate-papo educacional....	26
3.3.1.	Modelo de programação matemática (determinístico).....	26
3.3.2.	Modelo estocástico (probabilístico)	27
3.3.3.	Modelos de simulação	27
3.4.	<i>Corpora</i>	27
4.	Caracterização das atividades elementares: ler, digitar e refletir	35
4.1	Leitura.....	35
4.1.1.	Tempo de leitura é diferente entre as pessoas	38
4.1.2.	Tempo de leitura é proporcional à quantidade de caracteres da mensagem	42
4.2.	Digitação	48
4.3.	Reflexão	56
4.3.1.	Aguardando resposta	56
4.3.2.	Tempo Total de Reflexão	59
5.	Relação entre a quantidade de participantes e a produção de mensagens	61
5.1.	Relação estatística entre a quantidade de participantes e a produção de mensagens.....	61
5.2.	Produção máxima de mensagens: lê e digita o tempo todo	64
5.2.1.	Modelo analítico da produção máxima de mensagens.....	64
5.2.2.	Modelo estocástico da produção máxima de mensagens (M).....	68
5.2.3.	Comparação entre os modelos de produção máxima de mensagens.....	74

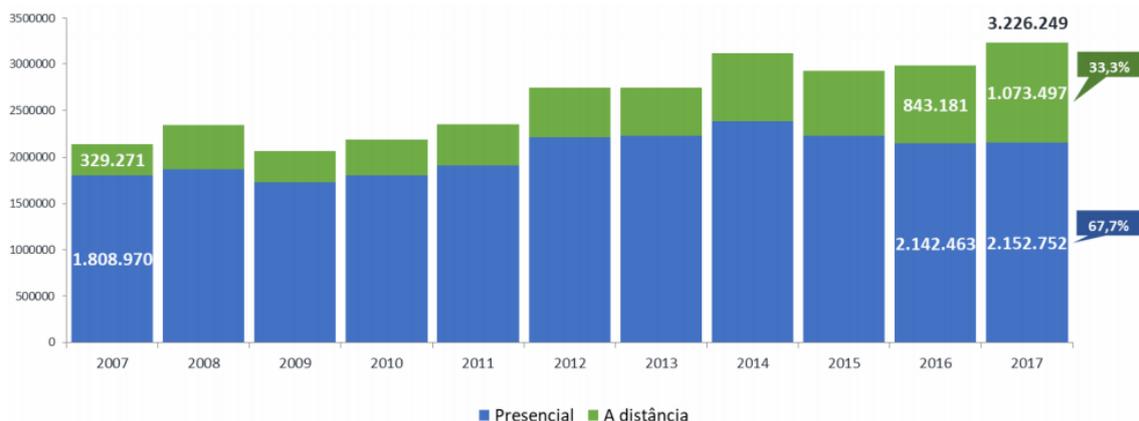
5.3.	Modelo em função do percentual de ocupação dos participantes (PO)	76
5.3.1.	Percentual de ocupação das turmas TIAE.....	79
5.3.2.	Percentual de ocupação da turma DSI.....	83
5.4.	Produção esperada de mensagens: Lê, Digita e Reflete (LDR).....	85
5.4.1.	Modelo analítico da produção esperada de mensagens.....	85
5.4.2.	Modelo estocástico da produção esperada de mensagens.....	89
5.4.3.	Comparação entre os modelos de produção esperada de mensagens.....	92
5.5.	Modelo Lê, Digita, Refletindo e Motivado (LDRM)	93
6.	Quantos alunos devem participar de uma sessão de bate-papo educacional?	96
7.	Conclusão	100
	Referências	104
	Apêndice 1 – Roteiro para entrevista com os alunos da disciplina DSI 2016.1	109
	Apêndice 2 – Tabela de parâmetros para cálculo da taxa de ocupação da turma no modelo Escrita-Leitura	110

1. Introdução

Neste capítulo, é apresentada uma visão geral da pesquisa, que objetiva caracterizar a participação de usuários em sessões de bate-papo educacional. Considerando o crescimento da educação a distância (EAD) e o uso de meios de conversação síncronos como o bate-papo, torna-se relevante destacá-lo como estratégia para complementar os demais métodos de ensino. No contexto educacional, espera-se que o aluno consiga acompanhar a discussão, lendo as mensagens enviadas pelos outros alunos e enviando mensagens que contribuam para a discussão do grupo. Para dimensionar turmas de EAD, parte-se do princípio de que é possível modelar o comportamento de usuários em sessões de bate-papo. Como forma de iniciar a investigação, é apresentado um modelo conceitual por meio do qual se espera obter estimativas de produção de mensagens a partir de uma quantidade de participantes.

1.1 Contexto: a expansão da EAD e o uso de bate-papo

O objetivo da presente pesquisa é produzir conhecimento sobre o uso do bate-papo na educação. Este é um objetivo importante na atualidade dado o grande crescimento da EAD no Brasil e o intenso uso desse meio de comunicação nessa modalidade. O número de alunos que ingressam na modalidade a distância continua crescendo: de 15,4%, em 2007, atingiu 33,3% em 2017, o que representa 1,07 milhão de estudantes universitários (INEP, 2017), como é ilustrado na Figura 1.



Número de Ingressos em Cursos de Graduação por Modalidade de Ensino – 2007-2017

Figura 1. Número de matrículas em cursos de graduação, por modalidade de ensino no Brasil – 2003-2014

Fonte: INEP, 2017, p. 11.

Dentre os canais de comunicação mais utilizados nos cursos na modalidade totalmente a distância, o bate-papo aparece em terceiro lugar. Ele é usado por 68% das instituições e perde apenas para o e-mail (88,2%) e fórum (87,2%) (CENSOEAD.BR, 2017, p. 189). Na Figura 2, são apresentados os canais de comunicação mais utilizados na EAD.

Tabela 8.4 – Canais de comunicação por tipo de curso, em percentual de instituições

	Totalmente a distância	Semipresenciais	Líves não corporativos	Líves corporativos
Rede social interna	29,1	28,2	14,74	31,08
SMS	31	35,9	24,36	21,62
Tutoria em vídeo	35	31,4	22,44	4,05
Videoconferência	39,9	33,3	21,15	28,38
Avisos automáticos	52,7	45,5	40,38	47,3
Ferramenta de avisos	67	66	46,79	51,35
Chat	68	64,1	46,79	52,7
Fórum	87,2	80,8	63,46	70,27
E-mail	88,2	82,7	78,21	79,73

Gráfico 8.4 - Canais de atendimento oferecidos, em percentual de instituições

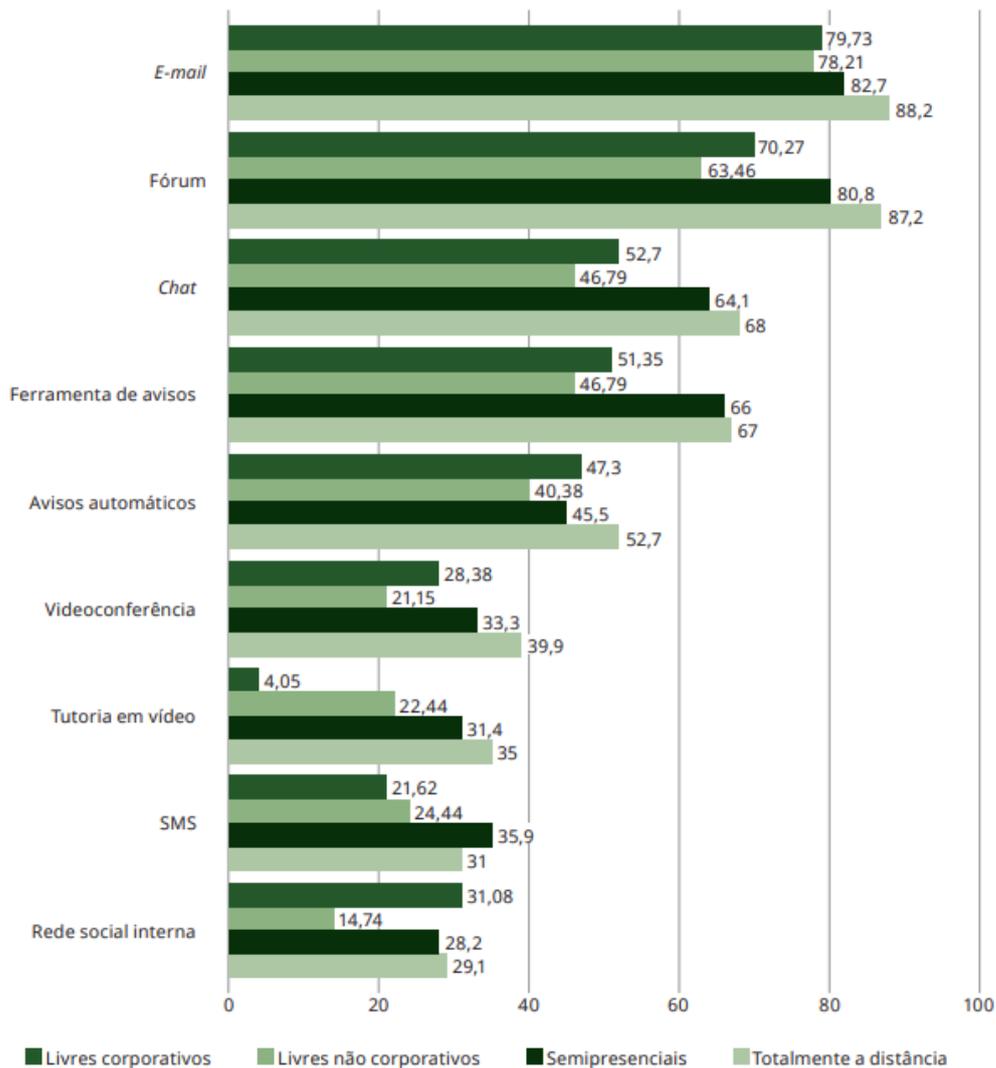


Figura 2. Canais de comunicação mais utilizados pelas instituições de EAD no Brasil

Fonte: CENSOEAD.BR, 2017, p. 189.

Os meios de comunicação baseados em computador são classicamente categorizados em dois grandes grupos: síncronos e assíncronos (CALVÃO; PIMENTEL; FUKS, 2014). Os meios assíncronos mais utilizados são o e-mail e o fórum, enquanto o bate-papo é o meio síncrono mais utilizado para a comunicação entre os alunos nas instituições de EAD.

É relevante posicionar que a presente pesquisa pretende apoiar a educação on-line e a EAD, no sentido de produzir conhecimento sobre bate-papo no contexto educacional visando promover mais participação e engajamento dos alunos nas aulas não presenciais. É preciso reconhecer que há muitas formas de realizar a EAD apoiada por tecnologias

computacionais. Uma das modalidades é o MOOC (*Massive Open Online Course* – Curso Online Aberto e Massivo), que pretende maximizar o número de usuários fazendo uso apenas de videoaulas gravadas e uma avaliação ao final do curso. Não é para esse tipo de curso que a presente pesquisa visa a contribuir. Os meios de conversação, tanto os assíncronos quanto os síncronos, potencializam uma educação voltada à aprendizagem colaborativa, dialógica, interativa, com autoria e construção compartilhada de conhecimento. É para apoiar a educação baseada nessas práticas que a presente pesquisa pretende contribuir.

Ao caracterizar a participação dos usuários em sessões de bate-papo educacional, a pesquisa visa mostrar que o bate-papo on-line requer mediação docente e que a quantidade de alunos que participam da sessão de bate-papo não pode ser muito grande, pois quanto maior o grupo de participantes, menos mensagens cada pessoa produz, o que limita o diálogo, a interatividade e a autoria. Dessa forma, a investigação tem o intuito de contribuir para uma educação de qualidade, e combater o sucateamento da educação pela adoção da modalidade a distância.

Em novembro de 2018, foram aprovadas, no Brasil, pelo Conselho Nacional de Educação (CNE), as novas diretrizes do ensino médio. Estas possibilitam que 20% do total da carga horária do ensino médio sejam ofertadas na modalidade a distância. No ensino médio noturno, a proporção pode chegar a 30% (BRASIL, 2018). Essa alteração enfrenta severas críticas por educadores brasileiros, que a consideram um sucateamento da educação pública, que vem sofrendo enfraquecimento desde 2016 com a aprovação da emenda constitucional de congelamento de gastos nos próximos 20 anos (PEC 241) (BRASIL, 2016). Cristina Maria D'Ávila, professora titular de Didática da Universidade Federal da Bahia (UFBA), em seu perfil do Facebook, alerta para os problemas da modalidade a distância no ensino médio:

Estão achando pouco o Brasil liderar o primeiro lugar no *Ranking* mundial de desqualificação da profissão docente? Pois aos que "acham" que entendem de educação e aplaudem as novas medidas, ontem foi aprovada a Reforma do ensino médio pelo CNE, dentre as belezas aprovadas estão: 30% do ensino médio deve ser a distância. 30% no turno noturno. Bom para nossos adolescentes brasileiros? Está bem assim? Isso não lhes incomoda uma vez que seus filhos estão salvaguardados em escolas da rede privada para as quais destinamos mensalidades absurdas de mais de 1000,00 a cada mês! Mas o filho do pobre que vai para a escola pública, saibam, este vai ter um ensino de pior qualidade ainda... a distância. Sem querer desprezar a EAD, sabemos que tipo de EAD será dispensada nessas escolas, ou não? com qual wifi? com qual tutoria? com qual projeto? quem será o responsável? ora ora, tenham paciência pois já perdi a minha! Como diz Calegari: As novas diretrizes abrem as escolas

para a educação a distância, permitem que trabalhos voluntários e comunitários, inclusive em igrejas, sejam contados como currículo e abrem caminho para a privatização da educação pública. Pois eu desprezo esta reforma, choro, reclamo e me desanimo, pois com 30 anos de educação estou assistindo à derrocada da educação pública. Vide BNCC, ESP, etc. Ah, me poupem de tanta vergonha! (D'ÁVILA, 2018)

Neste sentido, a pesquisa corrobora o desabafo da professora, pois mostra que o bate-papo on-line não deve ser realizado sem planejamento, sem docência e com um número excessivo de participantes.

1.2 Problema e questão de pesquisa

Quantos alunos devem participar de um bate-papo realizado no contexto educacional? Responder essa pergunta é importante, por exemplo, para planejar a carga horária necessária para realizar uma sessão de bate-papo numa turma de um curso a distância. Em sessões de bate-papo educacional, é necessário garantir que os alunos consigam participar e contribuir com a discussão, formando e construindo um conhecimento colaborativo. Em sessões de bate-papo recreativas, como BOL ou G1, não é necessário que o participante leia todas as mensagens ou mesmo envie mensagens. O participante, ou, neste caso, espectador, pode permanecer o tempo todo lendo as contribuições e respostas dos outros participantes.

Em uma sessão de bate-papo educacional, é necessário que o participante seja capaz de ler todas as mensagens enviadas pelos demais participantes e também digitar mensagens que contribuam com a discussão (PIMENTEL; FUKS; LUCENA, 2003a). Todos precisam ler e compreender todas as mensagens em um bate-papo educacional; isso é diferente quando se realiza um bate-papo em outros contextos, por exemplo, em sessões de bate-papo realizadas com muitas pessoas, em que as mensagens postadas são filtradas por uma equipe (PRIMO, 1997), e nesse caso, até podem participar centenas de pessoas. Em bate-papo recreativo, para socialização ou namoro, o participante não tem a necessidade de acompanhar a discussão nem se vê obrigado a digitar mensagens, não precisa ler tudo, pode descartar várias mensagens e manter-se calado. Assim, fora do contexto educacional, não há um claro fator limitante para a quantidade de pessoas no bate-papo. Só faz sentido discutir a quantidade máxima de participantes se for exigido um

nível mínimo de participação. No caso de a participação ser zero (o aluno não precisa ler nem digitar mensagens), tal como ocorre na transmissão de um vídeo em *broadcasting*, então não há um limite máximo de alunos que devem participar da sessão. O professor pode dar uma aula por bate-papo para milhares de alunos conectados ao mesmo tempo desde que somente o professor esteja digitando as mensagens e a participação dos alunos tenda a zero. O problema surge quando se dá a chance ao aluno de interagir com o professor e com os colegas – e é para isso que serve o bate-papo.

No contexto educacional, não é razoável colocar centenas de alunos conversando ao mesmo tempo no bate-papo, por isso deve-se determinar a quantidade máxima de participantes numa sessão de bate-papo para que seja possível alcançar uma qualidade definida em termos da quantidade de mensagens escritas e lidas pelos participantes. Saber se, em uma sessão de bate-papo educacional, podem participar ao mesmo tempo 100 ou apenas 10 alunos tem um grande impacto financeiro no planejamento das atividades de um curso a distância. O problema endereçado na presente pesquisa emergiu da divergência encontrada na literatura sobre a quantidade de participantes em bate-papos educacionais.

Em estudos anteriores, foi constatado que o tamanho do grupo influencia a participação on-line (BÖHLKE, 2003; CASPI; GORSKY; CHAJUT, 2003; HRASTINSKI, 2007; LÖBER; SCHWABE; GRIMM, 2007). Os estudos mostram que a produção de mensagens por participante é menor em grupos maiores. No entanto, quantas pessoas, no máximo, devem participar de uma sessão de bate-papo na educação? Como dimensionar a quantidade de alunos em sessões de bate-papo? Responder a essas questões é importante para dar subsídios aos gestores de EAD no dimensionamento das turmas.

Um fator que limita a quantidade máxima de participantes no bate-papo educacional é decorrente da expectativa de que todos os alunos consigam ler integralmente todas as mensagens enviadas na sessão, por isso deve ser limitada a quantidade de participantes de forma que a produção de mensagens não ultrapasse a capacidade média de leitura desses participantes. Além de ler todas as mensagens, um participante também precisa digitar mensagens de tempos em tempos, e também precisa de certo tempo livre para refletir sobre o que está sendo discutido antes de elaborar uma nova mensagem.

Neste contexto, para Cunha, Fuks e Lucena (2003), um fator influenciado pelo tamanho do grupo é a quantidade de trabalho produzida por seus membros; por exemplo, em grupos de discussão com muitos membros, a quantidade de mensagens por membro não é tão grande quanto em grupos menores, em que é preciso que cada participante envie mais mensagens para que haja um volume considerável destas.

Portanto, o objetivo da pesquisa é investigar o comportamento dos usuários de bate-papos educacionais de forma a caracterizar matematicamente a relação entre a quantidade de participantes e a quantidade de mensagens produzidas por cada participante durante a sessão de bate-papo. Para isso, foram investigadas as atividades elementares envolvidas em um bate-papo on-line: digitação de mensagens, leitura de mensagens e reflexão. Os estudos realizados forneceram modelos matemáticos que representam as atividades elementares dos usuários durante uma sessão de bate-papo.

1.3 Modelo conceitual preliminar (conjecturas iniciais)

A caracterização da participação em sessões de bate-papo educacional contribuirá para diferentes finalidades, dentre as quais estimar quantos alunos devem participar de uma sessão de bate-papo educacional de forma a garantir um nível de participação satisfatório na concepção do professor-tutor. Essa estimativa é obtida por meio de um modelo matemático que equaciona a quantidade de mensagens produzidas em função da quantidade de participantes.

O número total de mensagens produzidas na sessão aumenta proporcionalmente à quantidade de participantes. Conforme aumenta a quantidade de mensagens, o participante despende mais tempo para ler todas as mensagens e terá menos tempo para enviar novas mensagens, e assim sua produção de mensagens é reduzida com o aumento do número de participantes. Na Figura 3, é ilustrado um modelo idealizado da produção de mensagens por pessoa em função da quantidade de participantes na sessão de bate-papo educacional. Com a presente pesquisa, objetiva-se fornecer, aos educadores que atuam na EAD, um modelo capaz de produzir estimativas da produção de mensagens dos alunos por meio de dados característicos da turma e do tamanho da turma.

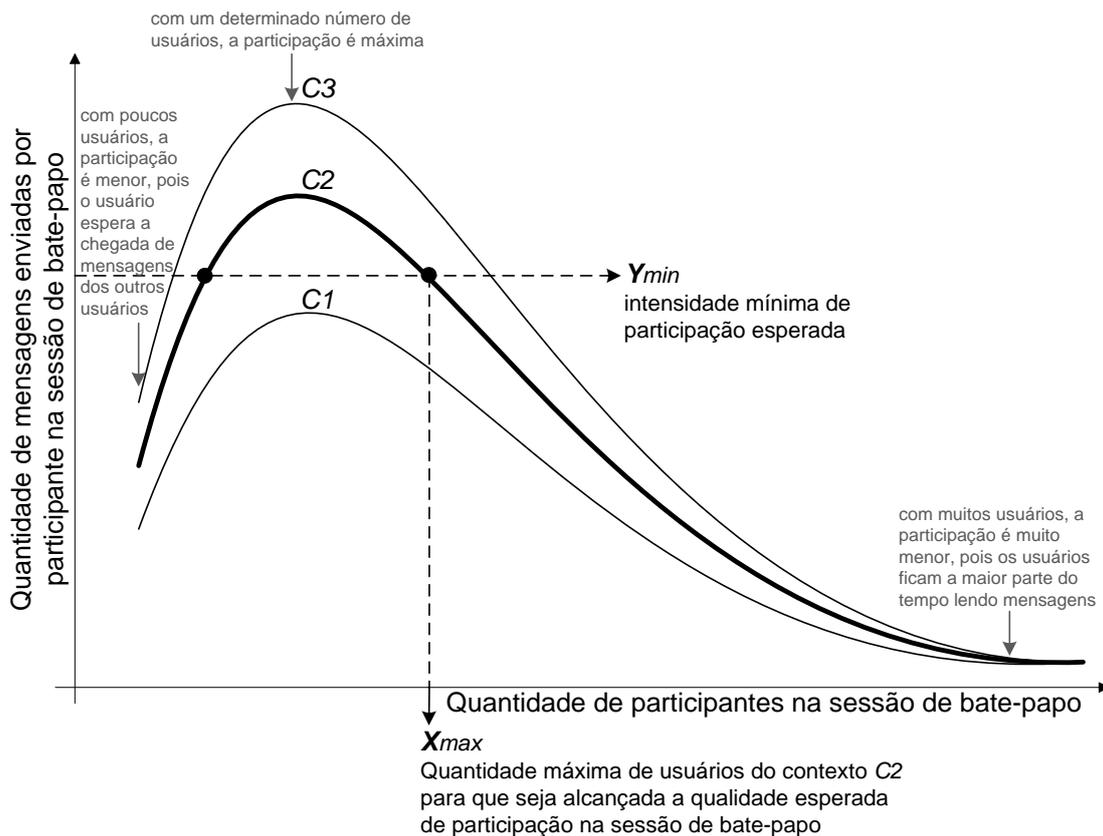


Figura 3. Modelo idealizado de participação em bate-papo educacional

Fonte: O autor, 2013.

Neste modelo, são apresentadas curvas diferentes – C1, C2 e C3 – que representam contextos distintos. Os contextos representam as características singulares de cada grupo de bate-papo tais como: idade média, nível de instrução, assunto debatido e distribuição dos participantes por gênero. Com a presente pesquisa, deseja-se determinar, para um dado contexto, a quantidade máxima de participantes (X_{max}) que pode estar na sessão para que seja possível alcançar uma determinada produção de mensagens (Y_{min}) por participante.

1.4 Metodologia desta pesquisa

No sentido de buscar o alcance do objetivo, optou-se por realizar uma pesquisa do tipo quantitativa seguindo o método de modelagem matemática. A análise quantitativa é realizada por meio de testes estatísticos, como o teste de hipótese, para avaliar se o tamanho do grupo influencia significativamente na produção de mensagens dos

participantes do bate-papo. Nesta pesquisa, são apresentadas duas abordagens de modelagem: modelo analítico e modelo estocástico.

1.5 Histórico da pesquisa

A pesquisa teve início no curso de Mestrado em Informática no Programa de Pós-Graduação em Informática da Universidade Federal do Estado do Rio de Janeiro (PPGI-UNIRIO), de 2011 a 2013, período em que, por meio de uma revisão de literatura, identificou-se disparidade sobre qual é o número de alunos que devem participar de uma sessão de bate-papo na educação. Equacionar esse valor tornou-se o objetivo daquela pesquisa. Ainda no mestrado, foi construído um modelo estocástico baseado em Teoria de Filas e implementado no ambiente de modelagem TANGRAM-II (SILVA; FIGUEIREDO; LEÃO, 2009). Para avaliar a adequação do modelo na solução do problema, foram usados *logs* de bate-papo reais.

Após a conclusão do mestrado, a pesquisa teve continuidade com a publicação dos resultados em meios acadêmicos, contudo foi preciso dar um “passo atrás” (ROCHA; PIMENTEL; DINIZ, 2015), porque foi constatado que o modelo elaborado (um modelo de quatro estados defendido na pesquisa de mestrado) gerava uma complexidade desnecessária, já que a diferença em relação ao modelo de três estados só é notada para os grupos pequenos de até cinco participantes (*small groups*), que não eram de interesse dessa pesquisa, cujo objetivo era verificar o número máximo de alunos que poderiam participar de um bate-papo sem prejudicar a qualidade do debate.

Além disso, um novo objetivo foi definido, o de tentar simplificar o problema e buscar uma equação que relacionasse todas as variáveis levantadas na fase exploratória da pesquisa (tempo de digitação, tempo de leitura, tempo de reflexão, quantidade de participantes e quantidade de mensagens produzidas). Após ciclos investigativos ao longo do doutorado, foi desenvolvida uma equação para realizar estimativas de produção de mensagens com base em parâmetros coletados da turma.

Outro ponto importante, que se apresentou como limitador da pesquisa de mestrado, foi a quantidade de participantes das turmas das quais os *logs* foram obtidos. Na *corpora* de dados utilizados na dissertação, a quantidade de participantes não apresentou uma grande variação no contexto de uma mesma turma, tendo média de 10

alunos. Para tentar contornar esse problema, no doutorado buscou-se uma variação maior sobre o número de participantes com sessões que variavam entre 2 a 15, variação mais expressiva do que os dados investigados no período do mestrado.

Os resultados obtidos durante o mestrado foram revisitados à luz daqueles obtidos com a pesquisa realizada no doutorado, e a comparação entre os resultados das diferentes abordagens dão uma percepção da razoabilidade/aceitabilidade dos resultados até agora obtidos.

1.6 Organização da escrita

Este trabalho possui a seguinte organização: neste capítulo, foi feita uma introdução à pesquisa e apresento brevemente o problema, as questões de pesquisa, os objetivos a serem alcançados e a metodologia. No capítulo 2, o problema da pesquisa é apresentado em mais detalhes, sendo abordado também um breve histórico dos sistemas computacionais de bate-papos e de seu uso para a educação on-line. No capítulo 3, são apresentados o paradigma epistemológico, a metodologia utilizada para resolver o problema e responder às questões de pesquisa, e também são apresentados a *corpora* de dados reais de sessões de bate-papo no contexto educacional. No capítulo 4, são apresentadas as atividades elementares desempenhadas pelos participantes de bate-papo durante uma sessão. No capítulo 5, apresentam-se as investigações entre o tamanho do grupo e a produção de mensagens. No capítulo 6, buscou-se responder à questão que instigou a realização desta tese: quantas pessoas, no máximo, devem participar de uma sessão de bate-papo? Por fim, no capítulo 7, é apresentada a conclusão da pesquisa.

2. Problema: até quantas pessoas devem participar de um bate-papo educacional?

2.1. Bate-papo

O computador, criado inicialmente para a realização de cálculos, após a interconexão em rede, tornou-se predominantemente um meio de comunicação humana (PIMENTEL; GEROSA; FUKS, 2011, p. 66).

O bate-papo é um meio de conversação síncrona para a discussão em turma (ou grupo pequeno) em que todos trocam mensagens textuais curtas. Denomina-se sala de bate-papo, ou canal, o ambiente virtual que reúne os interlocutores em função de um tema de interesse (CALVÃO; PIMENTEL; FUKS, 2014). Embora a troca de mensagens seja realizada por escrito, os participantes sentem-se como se estivessem falando. Os interlocutores até modificam a linguagem formal escrita para torná-la mais parecida com a conversação face a face; por exemplo, fazem uso de onomatopeias, alongamentos vocálicos e sobrecarga de pontuação. Esse fenômeno é denominado “reoralização da língua escrita” (HILGERT, 2000), indicando a adaptação de características da conversação oral para o texto produzido numa sessão de bate-papo.

Neste tipo de interação, interlocutores estão em contato por um canal eletrônico, o computador. Eles sentem-se falando, mas, pelas especificidades do meio que os põe em contato, são obrigados a digitar suas mensagens, ou seja, interagem construindo um texto “falado” por escrito (HILGERT, 2000, p. 17).

“Party Line”, lançado em 1971 (HILTZ; TUROFF, 1978), é reconhecido como o primeiro sistema computacional a ter implementado o meio de conversação bate-papo, desenvolvido para possibilitar a conversação simultânea por texto entre até 15 pessoas. *Party Line*, nos Estados Unidos, refere-se às linhas de telefone compartilhadas entre as

residências de uma região, sendo esta a forma típica (e não a exceção) de usar o serviço de telefonia local para fins não comerciais antes da Segunda Guerra Mundial.

Em 1971 considerávamos a funcionalidade de 'chat' como uma façanha menor em comparação com as outras coisas que estávamos fazendo. Hoje, em termos de uso, é provavelmente o modo de comunicação em grupo mais popular da net! (TUROFF, 2000).

Durante a década de 1980, muitos *bulletins boards systems* (BBS) disponibilizavam sistemas de bate-papo, o que contribuiu para a disseminação desse meio de comunicação. Já na segunda metade da década de 1990, com a popularização da internet, o *Internet Relay Chat* (IRC) se tornou um fenômeno social. No IRC, algumas pessoas atuavam como operadores responsáveis pela manutenção da sala de bate-papo, sendo capazes de expulsar usuários que desrespeitassem as regras daquela comunidade. A popularidade do IRC promoveu o estabelecimento de comunidades com usuários assíduos de determinadas salas de bate-papo. No final da década de 1990, o IRC caiu em desuso (PIMENTEL; GEROSA; FUKS, 2011).

Atualmente, os sistemas de bate-papo mais utilizados encontram-se integrados em sistemas de redes sociais. A conversação por bate-papo deixou de ser aberta ao público e realizada entre estranhos que se apresentavam por apelidos (*nickname*, como era praticado na época do IRC), e se tornou uma conversa contextualizada em comunidades entre participantes identificados por um perfil. Uma comunidade virtual pode ser uma turma de uma disciplina, um grupo de colegas de trabalho, um grupo de amigos, ou um agrupamento de pessoas interessadas num determinado tema. Além do uso para socialização e recreação, os sistemas de bate-papo passaram a ser usados para dar suporte a diferentes processos de trabalho em grupo: debates, reuniões e entrevistas.

2.2. Bate-papo na educação

Na educação, o bate-papo possibilita a constituição de um espaço para realizar atividades educacionais diferentes da aula tradicional. No bate-papo, geralmente não há exposição de conteúdos, mas sim muito diálogo. O professor não é considerado como o detentor do conhecimento e da palavra, pois a palavra é de todos, e todos compartilham

seus conhecimentos e dúvidas. A conversação informal, típica do bate-papo, possibilita ao aluno perceber melhor o outro e perceber-se como parte do grupo, e proporciona um espaço para emoções que diminuem a sensação de impessoalidade e isolamento. Em cursos a distância, o uso do bate-papo contribui para manter os alunos motivados e engajados na disciplina, diminuindo a evasão (PIMENTEL; FUKS; LUCENA, 2003b).

Quando o bate-papo é usado para realizar atividades educacionais, os participantes precisam acompanhar a conversação e compreender cada mensagem. A necessidade de acompanhar a conversação é menor em bate-papo de socialização e recreação, pois os participantes não estão comprometidos em entender a conversação e podem até deixar de ler algumas mensagens. Contudo, em “bate-papo sério”¹ (não recreativo), há o comprometimento em entender o que está sendo discutido, portanto é preciso ler todas as mensagens. Nesta situação, pode ocorrer a “confusão em bate-papo” (FUKS; PIMENTEL; LUCENA, 2006).

A confusão em bate-papo é decorrente de vários problemas sobrepostos, entre os quais, destacam-se: dificuldade em acompanhar um fluxo muito intenso de mensagem (sobrecarga de mensagem), dificuldade de identificar quem está falando com quem sobre o quê (perda de cotexto²), dificuldade para identificar o contexto da discussão (descontextualização), excesso de participantes, entre outros. Nesta investigação, busca-se equacionar a quantidade de pessoas que devem participar de um bate-papo educacional para evitar o excesso de participantes e, assim, potencialmente diminuir a sobrecarga de mensagens.

Apesar da possibilidade de ocorrer confusão num bate-papo, diversos autores reconhecem sua adequação como meio para a realização de atividades educacionais (SILVA, 2010; STAHL, 2009). Portanto, é preciso realizar pesquisas para melhorar os sistemas e os usos do bate-papo para evitar a ocorrência da confusão, por exemplo, com a adoção de dinâmicas de interação mais organizadas ou com o desenvolvimento de sistemas de bate-papo mais específicos para educação. Com isso, busca-se estimar a quantidade de pessoas que devem participar de um bate-papo educacional de forma a conseguir realizar um planejamento adequado do dimensionamento do grupo de participantes.

¹ Essa expressão é baseada no termo “jogo sério” (*serious game*), que são os jogos cujo objetivo é transmitir um conteúdo além do entretenimento <http://pt.wikipedia.org/wiki/Serious_game>.

² O termo cotexto designa texto ao redor, o que está escrito antes ou após um enunciado e que fornece elementos para compreendê-lo. Difere-se de contexto que designa fatores externos ao texto, também necessários para a compreensão do texto (PIMENTEL, 2002).

2.3. Participação em bate-papo educacional

O que é participar de uma sessão de bate-papo educacional? O que é uma participação de qualidade? Apenas “estar na sala” e não enviar mensagem pode ser considerado participação? De que participação, no contexto educacional, estamos falando?

A participação em ambientes virtuais pode ser caracterizada de diferentes formas, como: a quantidade de vezes que o aluno acessou o ambiente de aprendizagem, a quantidade de tarefas on-line realizadas e a duração de permanência no ambiente virtual. Um aluno estar on-line durante uma sessão de bate-papo configura que ele participou, porém é uma participação com qualidade diferente de um outro aluno que, além de estar on-line, também leu todas as mensagens enviadas pelos outros colegas e enviou várias mensagens ao longo da sessão, discutindo o tema com os demais participantes.

Em um bate-papo em que os alunos apenas assistem à aula proferida pelo professor, centenas ou milhares de alunos-ouvintes poderiam estar presentes já que não gerariam sobrecarga, pois não influenciariam a quantidade de mensagens produzidas na sessão. Segundo a classificação proposta por Hrastinski (2008a), acessar o ambiente é o nível 1 de participação e o autor especifica outros níveis conforme listados abaixo:

- a) nível 1: participação como acesso aos ambientes virtuais de aprendizagem;
- b) nível 2: participação como escrita;
- c) nível 3: participação como qualidade da escrita;
- d) nível 4: participação como escrita e leitura;
- e) nível 5: participação como escrita real e percebida;
- f) nível 6: participação como se juntando a um diálogo.

O primeiro nível de participação é medido pela simples contagem do número de acessos aos ambientes virtuais de aprendizagem – num bate-papo educacional, esse nível de participação ocorre quando o aluno apenas entra na sala de bate-papo. O nível 2 ocorre quando o usuário digita algo e é medido pelo número de produções do aluno – em bate-papo, seria equivalente a contar a quantidade de mensagens enviadas por um aluno na sessão. Já o nível 3 é classificado pela qualidade da escrita – no bate-papo, não basta enviar mensagens de socialização, é esperado que o aluno esteja produzindo conteúdo de

qualidade sobre o assunto em discussão. O nível 4 diz respeito à participação como escrita e leitura: um aluno que digita e lê muitas mensagens é considerado mais participativo que outros. O nível 5 de participação é definido como escrita real e percebida, ou seja, quando o aluno digita muitas mensagens e essas são percebidas pelos outros como mensagens importantes. E, finalmente, o nível 6, participação como parte de um diálogo, ocorre quando o aluno sente que está fazendo parte da discussão e esta se torna prazerosa – no bate-papo, este nível é equivalente ao aluno conseguir acompanhar toda a discussão, engajar-se nela e se sentir corresponsável pela qualidade da discussão.

Na presente pesquisa, por “participação na sessão de bate-papo” será considerado o nível 4 da classificação de Hrastinski: o participante deve ler todas as mensagens e ser capaz de juntar-se ao diálogo que está ocorrendo na sessão de bate-papo, digitando mensagens durante a sessão. Esse nível caracteriza quantitativamente a participação que é esperada nesta pesquisa em relação às sessões de bate-papo educacional. Cabe ressaltar que os níveis 1 a 4 são considerados como “participação real”, enquanto os níveis mais altos de participação, 5 e 6, são classificados como “participação percebida” (HRASTINSKI, 2008a). Só é possível caracterizar se a participação alcançou os níveis 5 e 6 por meio de entrevistas diretas com os participantes, exigindo uma abordagem qualitativa sobre a participação, o que está fora do escopo da presente pesquisa.

2.4. Pesquisas sobre fatores que influenciam a participação em bate-papo educacional

A participação on-line pode ser influenciada por diversos fatores. Carr *et al.* (2004) concluíram que existe diferença entre a participação de **homens e mulheres** no bate-papo. Hsieh e Tsai (2012) concluíram que **o moderador** influencia a discussão on-line. No estudo de Bolliger e Martin (2018), são apresentadas as percepções de alunos e professores com relação à importância das estratégias usadas para o engajamento na educação a distância (EAD).

Hrastinski (2007) identifica três grupos de fatores que influenciam a participação on-line: fatores demográficos, fatores comportamentais e fatores de contexto (Figura 4). Os fatores demográficos são aqueles relacionados à população como: nível de educação, tipo de residência (urbana ou rural), gênero, idade, tamanho do grupo. Os fatores

comportamentais estão relacionados às características individuais das pessoas como: atitudes dos alunos, estratégias dos professores e senso de comunidade. Já os fatores contextuais, como o nome sugere, estão relacionados ao contexto em que se dá a conversação: dispersão geográfica dos participantes, disponibilidade dos participantes, acessibilidade do meio de conversação e tipo do curso.

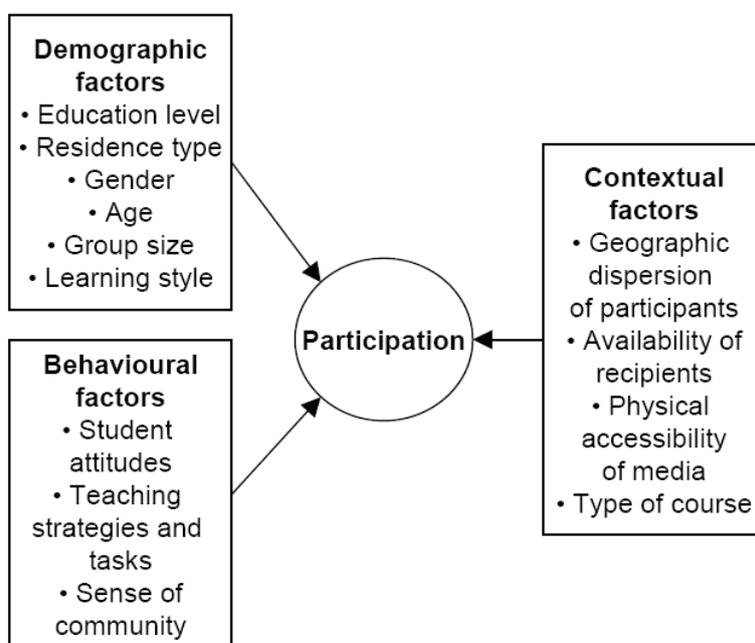


Figura 4. Fatores que influenciam a participação on-line

Fonte: HRASTINSKI, 2007, p. 116.

O tamanho do grupo, identificado como um dos fatores demográficos por Hrastinski, é muito discutido na literatura. Na presente pesquisa, investiga-se quantitativamente a relação entre o tamanho do grupo e a produção de mensagens numa sessão de bate-papo educacional. Conforme apresentado anteriormente, vários fatores influenciam a participação on-line: o moderador, o gênero, o nível de escolaridade, a idade, o meio de conversação e a complexidade do assunto são alguns desses fatores. Esta pesquisa está direcionada especificamente para o fator tamanho do grupo (*group size*) e sua influência na produção de mensagens na sessão de bate-papo.

2.5. Divergências sobre a quantidade de pessoas que devem participar de uma sessão de bate-papo

Bump (1990) considera a velocidade em uma conversa de bate-papo como uma grande desvantagem. Uma vez que uma sala de bate-papo não requer tomada de turno, cada aluno pode digitar a qualquer momento, e um enorme corpo de texto é produzido. Para amenizar tais problemas, Bump e Beauvois propõem criar pequenas conferências (BEAUVOIS, 1992) e limitar o tamanho a não mais de **quatro** ou **cinco** pessoas (BUMP, 1990).

Na pesquisa de BÖHLKE (2003), foram comparadas discussões face a face e as mediadas pelo bate-papo. Nesse estudo, a participação, em geral, foi distribuída de forma mais uniforme entre os grupos de bate-papo, e foi observado que os grupos com apenas **quatro** membros se beneficiaram do efeito de igualdade da sala de bate-papo.

Löber, Schwabe e Grimm (2007) realizaram um experimento comparando grupos de discussão por meio de audioconferência e por meio de bate-papo com a intenção de escolher o meio mais apropriado para as discussões. As conclusões do experimento foram que grupos com **sete ou mais** membros deveriam usar a comunicação por bate-papo e que grupos de até cinco ou seis membros se comunicariam melhor por audioconferência, tendo em vista alcançar mais produtividade (LÖBER; SCHWABE; GRIMM, 2007).

Hrastinski (2008b) realizou um estudo de caso sobre participação real (quantitativa) e participação percebida (qualitativa), comparando discussões por meio de bate-papo de duas turmas de EAD com diferentes tamanhos: **8** e **19** participantes. A conclusão foi que a diferença nos níveis de participação (real e percebida) foi especialmente verificada na classe menor (8 participantes) (HRASTINSKI, 2008b).

Autores indicam que 10 é um tamanho bom para um grupo de participantes no contexto educacional (PIVA *et al.*, 2011). Pimentel, Fuks e Lucena (2003b), que investigaram debates realizados por meio de bate-papo, apontaram problemas como a “confusão da conversação”, “sobrecarga de mensagens” e “excesso de participantes” em uma disciplina de Informática ofertada totalmente a distância. Em uma das turmas investigadas, em média estavam presentes 19 participantes por debate. Ao final das sessões de bate-papo, alguns participantes foram entrevistados e, quando um entrevistado mencionava o excesso de participantes, perguntava-se qual deveria ser a quantidade ideal:

a maioria indicou quantidade entre **9** a **12** participantes, sendo que dois entrevistados foram ainda mais restritos e indicaram **5** a **7** participantes.

Uma pesquisa realizada com 82 alunos do curso de Ciências Biológicas a distância da Universidade Estadual do Ceará constatou que, entre as dificuldades enfrentadas durante as atividades com o uso de bate-papo, destaca-se o grande número de participantes interagindo ao mesmo tempo: em média 30 alunos por sessão. Essa grande quantidade de participantes provocou dificuldades na moderação, inclusive desconcentrando os alunos. Visando minimizar as dificuldades encontradas por discentes e docentes no uso do bate-papo educacional, a coordenação do curso implementou diretrizes de boas práticas para a condução e a avaliação da participação no bate-papo; uma delas foi limitar a participação a no máximo **18** participantes por sessão (CASTRO; PANTOJA; PAIXÃO, 2016).

Alguns sistemas de bate-papo limitam o número de participantes a 30 usuários por sessão, como é o caso do Bate-papo UOL. E há outros sistemas que não estabelecem limites.

Como pode ser verificado, as pesquisas científicas divergem com relação à quantidade máxima de pessoas que devem participar de uma sessão de bate-papo educacional, mas todas elas indicam uma quantidade menor que 18 participantes. Dessa maneira, a pesquisa objetiva caracterizar matematicamente a participação das pessoas em sessões de bate-papo, em especial a relação entre o tamanho de grupo e sua influência na produção individual de mensagens.

3. Metodologia de pesquisa

No presente capítulo, é apresentado o paradigma epistemológico em que a pesquisa se embasa (3.1), o tipo da pesquisa realizada e a justificativa para esse tipo de pesquisa. Em seguida, é apresentado o método científico utilizado para alcançar os objetivos da pesquisa (3.2). Por fim, serão apresentadas as possíveis abordagens para o método escolhido (3.3) e uma breve introdução para os demais capítulos onde essas abordagens foram desenvolvidas.

3.1. Paradigma epistemológico

De acordo com Simon (1969), a ciência vem se desenvolvendo por meio da condução de pesquisas que seguem determinados paradigmas epistemológicos. Um dos paradigmas epistemológicos ou correntes de pensamento é o positivismo. A hipótese central do positivismo sociológico é de que a sociedade humana é regulada por leis naturais que atingem o funcionamento da vida social, econômica, política e cultural de seus membros (MINAYO, 2008, p. 81). No positivismo, os cientistas sociais, quando analisam determinado grupo ou comunidade, têm de descobrir as leis invariáveis e independentes de seu funcionamento.

Daí decorre que os métodos e técnicas para se conhecer um sociedade ou determinado segmento dela devam ser da mesma natureza que os empregados nas ciências naturais. E ainda mais, da mesma forma que as ciências naturais defendem um conhecimento **objetivo, neutro, livre de juízo de valor**, de implicações político-sociais, também as ciências sociais devem buscar, para sua cientificidade, este “conhecimento objetivo” (MINAYO, 2008, p. 82).

A ciência positivista tem raízes na filosofia das luzes no século 18 (LÖWY, 1986). Para Löwy, o pai do positivismo é Condorcet, um enciclopedista que formulou de forma clara e precisa a ideia de que a ciência da sociedade deveria ser uma Matemática Social baseada em estudos quantitativos rigorosos e probabilísticos.

Para Augusto Comte, o pensamento teria de ser totalmente positivo. Isto é, dele deveria ser eliminado todo o conteúdo crítico para que os cientistas descobrissem as leis da sociedade. Seu “método positivo” propunha que o cientista social se consagrasse teórica e praticamente à defesa da ordem e fosse contrário ao que considerava “negativismo” perigoso nas doutrinas críticas, destrutivas, subversivas e revolucionárias da Revolução Francesa e do Socialismo (COMTE, 1978, p. 44).

É nesta corrente de pensamento positivista, de enxergar os fenômenos sociais de forma precisa, quantitativa, rigorosa e probabilística, ou seja, como uma matemática social conforme formulação de Condorcet, que a presente pesquisa se insere. Considera-se que é possível representar a participação de alunos em sessões de bate-papo educacionais por meio de um modelo quantitativo, ou seja, acredita-se que é possível representar, por meio de uma função matemática, a relação entre as variáveis quantidade de mensagens produzidas por participante e quantidade de participantes no bate-papo. Por meio dessa função matemática, seria possível prever a quantidade de mensagens possíveis de serem produzidas por um aluno durante uma sessão de bate-papo com base na quantidade de participantes da sessão e em outras variáveis complementares. Por meio do modelo, é possível fornecer respostas que auxiliam na gestão e na tomada de decisões no gerenciamento de recursos.

Para alcançar o objetivo proposto, foi utilizado o Método de Modelagem Matemática que é oriundo da Pesquisa Operacional.

Na pesquisa científica, existem técnicas para usar o computador para imitar ou simular a operação de vários processos do mundo real. As duas técnicas mais conhecidas são a Modelagem e Simulação (LAW; KELTON, 2000). Se a relação que compõe o modelo é simples o bastante, então é possível usar métodos matemáticos (álgebra, cálculo, teoria de probabilidade) para obter a informação exata da questão de interesse; isso é chamado solução analítica. A maioria dos sistemas do mundo real é muito complexa para ser analisada analiticamente; esses modelos devem ser estudados por meio de simulação. Uma simulação usa um computador para avaliar o modelo numericamente, os dados são recolhidos de modo a estimar as características desejadas do modelo.

Algumas das áreas de aplicação da simulação: Projetos e análise de sistemas de manufatura; Avaliação de projetos de serviços organizacionais como *call centers*, restaurantes *fast-food*, hospitais e postos de correio; Análise de sistemas financeiros ou econômicos.

Sistemas são coleções de entidades, pessoas ou máquinas que agem e interagem juntos para realização de um mesmo objetivo. Estado de um sistema é o conjunto de variáveis necessárias para digitar o sistema em um tempo específico (LAW; KELTON, p. 3). Categorias de sistemas: discretos e contínuos em relação ao tempo. Há uma necessidade de estudar os sistemas para tentar obter alguns *insights*.

3.2. Método científico

A pesquisa operacional (*operational research*) é uma abordagem científica para auxiliar no processo de tomada de decisões, que procura determinar como melhor projetar, planejar e operar sistemas, usualmente sob condições que requerem alocações diferentes de recurso escassos (ARENALES *et al.*, 2007).

No diagrama apresentado na Figura 5 e extraído de Arelanes *et al.* (2007), é ilustrado o processo simplificado da abordagem de solução de problema utilizando a Modelagem Matemática.

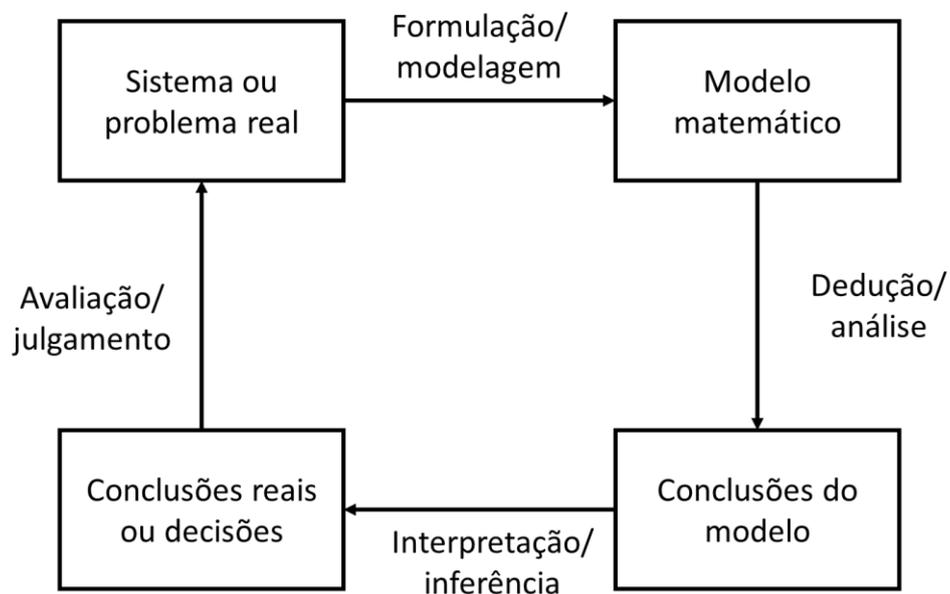


Figura 5. Processo de modelagem
Fonte: ARENALES *et al.*, 2007, p. 4.

Essa abordagem de resolução de um problema envolve várias fases a serem descritas em seguida.

3.2.1. *Definição do problema*

Nesta fase são definidos o escopo do problema em estudo, as decisões de interesse e os objetivos envolvidos, e o modelo conceitual do problema. Segundo Arenales *et al.* (2007), o modelo conceitual deve digitar as alternativas de decisões e as limitações sob as quais o sistema modelado funciona.

Conforme já apresentado no capítulo 1, o problema que se deseja resolver é saber quantos alunos devem participar de uma sessão de bate-papo. O objetivo é produzir um ou mais modelos que equacionem a quantidade de participantes de uma sessão de bate-papo (p) e a quantidade de mensagens produzidas por participante (m) para que se possa dimensionar o sistema de forma a maximizar a participação dos alunos.

O modelo conceitual é baseado em que os alunos permanecem em um dos três estados descritos no capítulo 4: Lendo, Digitando ou Refletindo. Outra conjectura sobre o modelo é de que todos os participantes desempenham o mesmo papel, podendo enviar uma mensagem a qualquer momento. O modelo conceitual foi apresentado na capítulo 1 (Figura 3) e é mais detalhado no capítulo 5.

3.2.2. *Construção do modelo*

Com as informações coletadas na primeira fase, é possível desenvolver e avaliar um modelo matemático do problema por meio de relações matemáticas ou lógicas de simulação. Na presente pesquisa, foram desenvolvidas duas abordagens para a modelagem. No capítulo 5, é apresentado o modelo analítico da participação dos alunos em bate-papos educacionais em que ao se atribuir os valores das variáveis independentes, obtém-se diretamente o valor da variável dependente. Também no capítulo 5 é aprofundado o modelo estocástico baseado em um ambiente de modelagem

computadorizado onde também é possível fazer simulações por meio da manipulação das variáveis de interesse.

3.2.3. Solução do modelo

Nesta fase, são utilizados métodos de solução e algoritmos para resolver o modelo construído na segunda fase. Esses algoritmos podem ser conhecidos da literatura ou desenvolvidos especificamente para tratar o modelo da segunda fase. Nesta fase também busca-se por métodos de solução mais efetivos para resolver o modelo e são realizados experimentos comparando as soluções obtidas com as soluções de outros métodos da literatura (*benchmarking*). Alguns dos métodos de solução dos modelos são modelos de programação linear, modelos heurísticos e modelos de otimização realistas. Algoritmos, como simplex, são bastante conhecidos na literatura. Uma tendência recente é a incorporação de modelos de pesquisa operacional em *softwares* de apoio à decisão (*Decisions Support Systems - DSS*) e sistemas de informação gerencial (*Management Information Systems - MIS*) e planejamentos de recursos (*Enterprise Resource Planning - ERP*).

A proposta de aplicação para o modelo desenvolvido na presente pesquisa é de integração com bate-papos no intuito de subsidiar dois tipos de decisões. Ao se integrar o modelo em um bate-papo, o professor da disciplina on-line poderá realizar uma primeira sessão de bate-papo diagnóstica em que é possível parametrizar o modelo para o contexto daquela turma e posteriormente obter as medidas de interesse. O professor poderá ser capaz de dimensionar sua turma on-line para realizar sessões de bate-papo educacional que possam produzir uma maior quantidade de mensagens por aluno e seja possível o acompanhamento das mensagens enviadas pelos participantes.

Partindo para o nível estratégico de decisão, a integração do modelo em portais, redes sociais ou ambientes virtuais de aprendizagem possibilitará um volume maior de dados coletados ao longo das realizações das sessões de bate-papo e informações a respeito de contextos distintos. Por exemplo, diferentes contextos relativos a nível de escolaridade (fundamental, médio e superior), idade (crianças, jovens, adultos, idosos), curso (informática, pedagogia, matemática). Essa grande base de informações será útil aos

tomadores de decisão dos níveis estratégicos da educação a distância (EAD), que terão subsídios para melhor dimensionarem as turmas e cursos a distância.

3.2.4. *Validação do modelo*

Nesta fase é verificado se o modelo proposto representa apropriadamente o problema, ou seja, se o modelo digita adequadamente o comportamento do sistema real. Um método comum para verificar a validade de um modelo é comparar seus resultados com dados históricos do problema. A qualidade da solução do modelo também depende da qualidade dos dados de entrada do modelo; a conhecida expressão inglesa “*garbage in, garbage out*” (entra lixo, sai lixo) sintetiza a questão.

Para a validação do modelo, foram utilizados dados históricos das turmas Tecnologia da Informação Aplicada à Informação (TIAE) ministradas na Universidade Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio) nos anos de 1999 a 2008. Utilizou-se parte dos *logs* de bate-papo obtidos para parametrizar o modelo e a outra parte para comparar os dados estimados pelo modelo como os dados observados no contexto real.

Como limitação dos dados históricos das turmas TIAE, o número de participantes das sessões de bate-papo se mantinha quase que inalterado, salvo se algum aluno faltasse. Para resolver esse problema, um novo estudo foi realizado em uma turma de pós-graduação em Informática na disciplina Docência em Sistemas de Informação (DSI) da Universidade Federal do Estado do Rio de Janeiro (UNIRIO). Naquela turma, as sessões de bate-papo foram realizadas com uma maior variação do número de participantes: 2, 3, 6, 10 e 15 alunos.

3.2.5. *Implementação da solução*

Nesta fase, a preocupação é com a implementação da solução na prática da empresa ou organização de modo a traduzir os resultados do modelo em conclusões ou decisões. A realização dessa fase pode requerer bastante tempo, principalmente se ela

envolver a implantação de um sistema computacional integrado ao sistema de informações gerenciais e bases de dados da empresa para apoiar as decisões na prática. Essa fase é muito importante na pesquisa empírica.

Como contribuição para a área gerencial, espera-se que os órgãos reguladores de EAD utilizem o modelo para dimensionar melhor as turmas dos cursos on-line.

Outra proposta para implementação do modelo é incorporá-lo no Tagarelas (ESTRUC, 2012). O Tagarelas é uma rede social educacional que promove os mais variáveis ambientes de conversação e artefatos de apoio à decisão para docentes. A implementação no Tagarelas será na forma de um relatório que poderá ser emitido de acordo com a abrangência selecionada pelo usuário, seja a nível de seção de bate-papo, turma, curso ou instituição de ensino, seja por diferentes focos: escolaridade, natureza da instituição (pública ou privada), idade etc.

Na Figura 6, é apresentada a proposta de relatório a ser implementado no Tagarelas para subsidiar a tomada de decisões a nível operacional ou tático, ou seja, dentro do contexto de uma turma específica. Após uma primeira sessão de bate-papo, o relatório produzirá um diagnóstico da turma e fará prescrições para uma maior produtividade de mensagens.

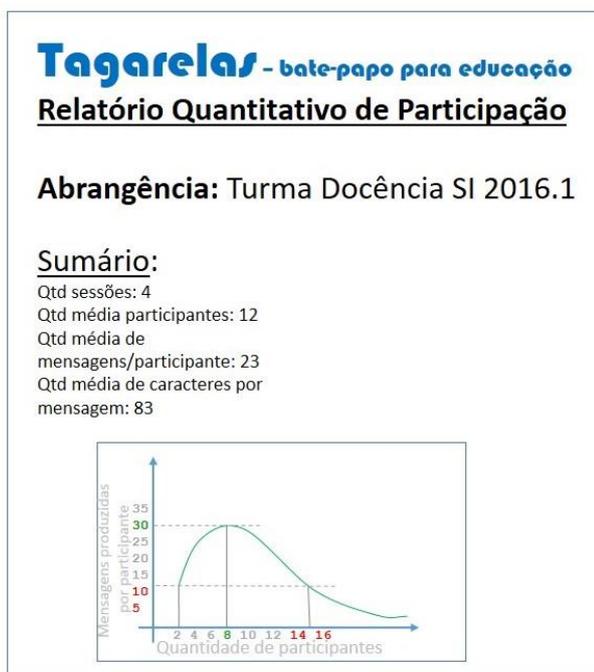


Figura 6. Relatório quantitativo da participação em bate-papo educacional
Fonte: O autor, 2018.

Por meio dos dados produzidos pelo modelo e evidenciados pelo relatório quantitativo da participação em bate-papo educacional, o professor poderá dimensionar melhor as próximas sessões de bate-papo.

3.3. Modelos quantitativos: modelo da participação em bate-papo educacional

Modelos quantitativos são modelos abstratos descritos em linguagem matemática e computacional, que utilizam técnicas analíticas (matemáticas, estatísticas) e experimentais (simulação) para calcular valores numéricos das propriedades do sistema em questão, podendo ser usados para analisar os resultados de diferentes ações possíveis no sistema (NETO; PUREZA, 2012).

3.3.1. Modelo de programação matemática (determinístico)

O modelo de programação matemática (determinístico) é um modelo analítico que representa alternativas ou escolhas desse problema como variáveis de decisão, e procura por valores dessas variáveis de decisão que minimizam ou maximizam uma função dessas variáveis – denominada *função objetivo*.

No modelo da presente pesquisa, busca-se maximizar a produção de mensagens por participante por meio da variação da quantidade de participantes da sessão. A *função objetivo* se apresenta da seguinte forma:

Maximizar (m) para um maior (p) possível

Dessa forma, pode-se concluir que existe um número máximo de participantes pelos quais uma quantidade de mensagens é produzida. O modelo analítico desta pesquisa é apresentado no capítulo 5.

3.3.2. *Modelo estocástico (probabilístico)*

Diferente dos modelos determinísticos, modelos de programação estatística (otimização probabilística) consideram incertezas nos parâmetros do problema. Segundo Neto e Pureza (2012, p. 191), modelos de programação estocástica são obtidos para obter uma solução que maximize, por exemplo, a probabilidade de atendimento das demandas ou a probabilidade de completar as rotas no tempo máximo estabelecido. Modelos estocásticos requerem, em geral, grande quantidade de dados históricos e tratamento estatístico adequado. Os modelos estocásticos que consideram incertezas são aqueles baseados em **teoria de filas**, modelos de previsão baseados em séries temporais (modelos de **regressão linear** e suavização exponencial, modelos de Box-Jenkins e inferência bayesiana). Outros modelos são os modelos de otimização robusta, modelos de teoria de jogos, modelos baseados em cadeias de Markov e processos markovianos de decisão e os modelos de simulação. Os modelos estocásticos também são descritos no Capítulo 5.

3.3.3. *Modelos de simulação*

Uma classe importante de modelos de pesquisa operacional são os de simulação (NETO; PUREZA, 2012, p. 191). Esses modelos são poderosos e amplamente utilizados para analisar sistemas complexos. Em geral, eles imitam as operações do sistema real à medida que este evolui no tempo. Exemplos de modelos de simulação são a simulação de Monte-Carlo e modelos de simulação dinâmicos.

Não está planejado na presente pesquisa realizar simulações, porém pode-se utilizar uma abordagem como a Simulação de Monte-Carlo.

3.4. Corpora

Na presente pesquisa, formando o *corpus* de análise, foram obtidos *logs* de sessões de bate-papo das turmas de uma disciplina do Departamento de Informática da PUC-Rio,

lecionada totalmente a distância pelo ambiente AulaNet, denominada “Tecnologias de Informação Aplicadas à Educação” (TIAE) (FUKS *et al.*, 2002). Ao todo, foram obtidos 142 logs de bate-papo ocorridos em 18 turmas ministradas entre os anos de 1999 e 2008 (em cada turma foram realizadas, em média, 8 sessões de bate-papo).

Nas sessões de bate-papo daquela disciplina, foram promovidos debates entre os alunos visando alinhar as ideias discutidas sobre o tema estudado a cada semana. Todos os alunos da turma deveriam estar conectados das 12h às 13h nas quintas-feiras, devendo preferencialmente conectar-se 10 minutos antes do debate. Em cada sessão, um dos alunos atuava como moderador do debate, sendo o responsável por manter o foco do debate, evitando que a discussão tomasse um rumo inadequado ou que um dos participantes ficasse calado por muito tempo (LUCENA; FUKS, 2002).

Um dos objetivos da pesquisa é verificar a influência do tamanho do grupo (número de participantes de uma sessão de bate-papo) na participação desses usuários (leitura de todas as mensagens dos demais participantes e produção de mensagens). Contudo, dentro de uma mesma turma, não houve uma variação significativa do número de participantes. Como exemplo, as turmas TIAE apresentaram a seguinte distribuição visualizada na Tabela 1.

Tabela 1. Dados dos debates das turmas TIAE

Turma	Sessão de bate-papo na turma	Quantidade de mensagens na sessão de bate-papo	Quantidade de mensagens de debate (descartando mensagens da sessão antes e após debate, e as mensagens automáticas)	Média de caracteres / mensagens de debate	Média de duração dos debates (min)	Média de participantes nos debates	Média de mensagens de debate / participantes	Média de mensagens / participantes projetada em 60 minutos de debate
1999.1	1	150	148	81,2	60,0	6,0	24,7	24,7
	2	257	226	78,7	60,0	9,5	23,9	23,9
	3	190	178	104,3	60,0	11,0	16,2	16,2
	5	281	257	67,2	60,0	10,9	23,5	23,5
	6	236	209	63,0	60,0	10,9	19,2	19,2
	7	233	216	77,8	60,0	10,6	20,3	20,3
	8	150	129	94,6	60,0	9,9	13,1	13,1
	9	230	205	77,8	60,0	9,0	22,8	22,8
	10	206	194	75,5	60,0	8,2	23,5	23,5
	Média	-	214,8	195,8	80,0	60,0	9,6	20,8
1999.2	1	367	286	50,4	60,0	7,0	41,2	41,2
	2	344	283	47,7	60,0	7,7	36,6	36,6
	3	325	298	49,9	60,0	11,1	26,9	26,9

	4	418	377	49,5	60,0	10,9	34,7	34,7
	5	582	412	44,3	60,0	9,3	44,3	44,3
Média	-	407,2	331,2	48,4	60,0	9,2	36,7	36,7
2000.1	1	310	273	55,5	60,0	8,8	31,1	31,1
	2	304	281	53,0	60,0	7,1	39,9	39,9
	3	276	252	56,3	60,0	5,0	50,4	50,4
	4	342	317	57,4	60,0	5,9	53,5	53,5
	5	348	339	55,4	60,0	7,7	43,9	43,9
	6	368	331	54,2	60,0	8,0	41,6	41,6
	7	356	336	59,9	60,0	8,0	42,0	42,0
	8	415	367	55,8	60,0	7,9	46,6	46,6
	9	351	345	49,0	60,0	6,7	51,4	51,4
	10	464	380	44,8	60,0	7,9	48,0	48,0
	11	375	360	57,1	60,0	8,0	45,0	45,0
	12	402	377	47,5	60,0	6,2	60,6	60,6
	13	245	218	55,6	60,0	4,2	51,7	51,7
Média	-	350,5	321,2	53,9	60,0	7,0	46,6	46,6
2001.1	1	367	339	56,6	60,0	16,6	20,4	20,4
	2	426	352	55,3	60,0	16,7	21,0	21,0
	3	378	333	52,0	60,0	15,8	21,1	21,1
	4	346	264	47,1	60,0	14,0	18,8	18,8
	5	399	359	47,9	60,0	13,8	26,0	26,0
	6	156	135	53,0	60,0	12,7	10,6	10,6
	7	276	237	64,4	60,0	11,7	20,2	20,2
Média	-	335,4	288,4	53,8	60,0	14,5	19,7	19,7
2001.2	1	417	300	75,1	60,0	16,0	18,8	18,8
	2	526	459	70,9	60,0	16,1	28,6	28,6
	3	538	463	74,9	60,0	18,0	25,8	25,8
	7	529	425	66,2	60,0	16,0	26,6	26,6
	5	573	522	77,9	60,0	14,6	35,7	35,7
	6	560	540	77,9	60,0	14,9	36,3	36,3
	7	574	488	75,7	60,0	15,9	30,8	30,8
Média	-	531,0	456,7	74,1	60,0	15,9	28,9	28,9
2002.1	1	528	515	56,1	60,0	14,5	35,5	35,5
	2	504	494	71,9	60,0	19,7	25,1	25,1
	3	692	683	59,2	60,0	18,4	37,2	37,2
	4	609	598	58,9	60,0	18,1	33,0	33,0
	5	699	675	65,3	60,0	18,9	35,7	35,7
	6	641	622	39,1	60,0	16,8	37,0	37,0
	7	582	561	54,0	60,0	20,4	27,5	27,5
	8	607	583	64,4	60,0	15,8	36,9	36,9
Média	-	607,8	591,4	58,6	60,0	17,8	33,5	33,5
2002.2	1	399	374	82,2	60,0	11,0	34,0	34,0
	2	517	489	67,3	60,0	11,0	44,6	44,6

	3	440	438	81,1	60,0	11,0	39,8	39,8	
	4	462	458	75,1	60,0	11,0	41,6	41,6	
	5	313	272	88,8	46,4	11,0	24,7	32,0	
	6	200	134	93,5	24,1	10,0	13,4	33,3	
	7	352	323	83,5	50,9	8,6	37,5	44,2	
	8	384	332	78,4	51,5	9,8	33,8	39,4	
	Média	-	383,4	352,5	81,2	51,6	10,4	33,7	38,6
2003.1	1	407	393	80,0	60,0	16,9	23,3	23,3	
	2	455	357	87,2	56,8	13,2	27,0	28,5	
	3	446	415	84,9	55,1	16,0	25,9	28,3	
	4	507	495	74,8	57,9	17,8	27,8	28,8	
	5	593	566	81,2	56,0	17,8	31,8	34,1	
	7	763	596	80,2	53,1	16,0	37,3	42,1	
	8	524	501	92,8	55,2	15,9	31,6	34,4	
	Média	-	527,9	474,7	83,0	56,3	16,2	29,2	31,3
2003.2	1	293	249	65,6	57,4	9,6	26,0	27,1	
	2	480	339	72,1	58,2	11,5	29,5	30,4	
	3	504	388	62,7	57,9	10,2	38,0	39,5	
	4	608	411	67,3	57,3	11,4	36,1	37,8	
	5	597	380	73,1	49,6	11,9	31,9	38,6	
	6	646	478	72,4	51,7	12,7	37,6	43,6	
	7	387	287	71,6	54,2	7,3	39,1	43,3	
	8	482	318	71,4	42,8	10,9	29,2	41,0	
	Média	-	499,6	356,3	69,5	53,6	10,7	33,4	37,7
2004.1	1	202	165	108,2	53,5	7,6	21,6	24,3	
	2	258	188	85,7	53,4	7,4	25,4	28,5	
	3	268	266	113,8	54,6	10,8	24,6	27,0	
	4	299	296	116,4	57,2	11,6	25,5	26,8	
	5	227	222	85,3	50,3	6,6	33,4	39,9	
	6	261	239	98,3	55,4	8,8	27,2	29,5	
	7	269	248	96,8	53,6	10,0	24,8	27,8	
	8	209	202	105,8	47,2	8,0	25,3	32,1	
	Média	-	249,1	228,3	101,3	53,1	8,9	26,0	29,5
2004.2	1	269	241	88,2	59,1	7,9	30,5	31,0	
	2	331	203	87,8	48,9	8,7	23,4	28,7	
	3	312	277	88,1	57,8	7,0	39,5	41,0	
	4	436	318	73,6	55,2	6,9	46,2	50,2	
	5	493	303	67,9	53,2	4,9	61,7	69,6	
	6	471	322	78,9	55,2	6,0	53,7	58,4	
	7	522	299	81,4	51,3	7,8	38,2	44,7	
	8	648	308	76,9	52,8	7,0	44,0	50,0	
	Média	-	435,3	283,9	80,4	54,2	7,0	42,2	46,7
2005.1	1	397	376	96,8	58,3	13,7	27,5	28,3	
	2	376	354	100,2	56,3	14,0	25,4	27,0	

	3	438	402	91,9	59,2	12,8	31,5	31,9
	4	336	303	96,0	58,5	10,0	30,3	31,1
	5	274	249	113,2	60,0	9,0	27,7	27,7
	6	341	323	88,5	57,6	10,0	32,4	33,8
	7	322	276	89,1	56,8	8,9	31,0	32,7
	8	362	310	95,1	53,9	10,3	30,0	33,4
Média	-	355,8	324,1	96,4	57,6	11,1	29,5	30,7
2005.2	3	290	241	92,4	57,0	6,9	35,0	36,8
	4	224	174	76,8	57,3	5,0	34,8	36,4
	5	302	246	71,3	58,2	5,0	49,2	50,7
	6	297	248	60,4	59,6	3,0	82,7	83,2
	7	292	235	53,8	52,3	4,0	58,5	67,1
	8	272	141	90,5	40,5	4,7	29,9	44,4
Média	-	279,5	214,2	74,2	54,1	4,8	48,3	53,1
2006.1	1	276	270	90,2	54,2	8,3	32,7	36,2
	2	274	270	104,1	60,7	9,9	27,4	27,0
	3	300	290	94,4	57,5	10,9	26,5	27,7
	4	255	255	96,2	53,0	9,0	28,3	32,1
	5	288	275	98,8	53,0	8,5	32,4	36,6
	6	242	240	120,5	60,3	8,4	28,7	28,5
	7	283	281	99,5	61,1	10,0	28,1	27,6
	8	221	212	103,5	55,3	10,3	20,5	22,2
Média	-	267,4	261,6	100,9	56,9	9,4	28,1	29,7
2006.2	1	365	348	80,7	59,7	8,1	43,0	43,2
	2	329	321	80,9	59,1	8,0	40,2	40,8
	3	331	316	96,5	56,8	8,9	35,3	37,4
	4	298	291	70,3	57,9	4,9	59,5	61,6
	5	406	382	62,9	58,2	7,6	50,1	51,7
	6	366	351	61,7	57,2	6,0	58,8	61,7
	7	329	324	81,7	59,4	5,7	56,8	57,4
	8	339	275	75,2	54,5	6,9	39,7	43,7
Média	-	345,4	326,0	76,2	57,8	7,0	47,9	49,7
2007.1	1	268	194	116,4	59,4	7,9	24,6	24,8
	2	250	236	92,5	59,8	8,6	27,6	27,6
	3	323	303	102,2	58,2	9,7	31,4	32,3
	4	326	271	105,2	59,6	9,4	28,8	29,0
	5	223	222	90,5	60,0	5,9	37,4	37,4
	6	278	265	88,9	51,1	8,0	33,1	38,9
	7	251	241	89,4	58,7	7,9	30,6	31,3
	8	235	227	67,4	58,3	5,7	40,2	41,4
Média	-	269,3	244,9	94,1	58,1	7,9	31,7	32,9
2007.2	1	305	250	106,8	53,9	11,0	22,8	25,3
	2	388	233	62,7	42,1	11,2	20,8	29,7
	3	278	275	88,8	56,7	12,0	22,9	24,2

	4	308	250	94,5	48,7	11,7	21,4	26,4
	5	247	201	91,2	52,7	6,0	33,5	38,1
	6	283	276	67,8	60,0	7,8	35,6	35,6
	7	232	197	67,4	31,2	11,8	16,7	32,2
	8	295	273	78,0	59,2	12,8	21,4	21,7
Média	-	292,0	244,4	82,2	50,6	10,5	24,4	29,1
2008.1	1	373	302	73,4	54,9	7,7	39,1	42,7
	2	330	245	89,0	55,2	6,5	37,5	40,7
	3	288	251	88,1	55,4	6,7	37,4	40,5
	4	298	245	95,8	59,6	5,2	47,6	47,9
	5	301	265	77,5	59,6	5,5	48,1	48,4
	6	233	197	79,6	58,0	4,9	40,2	41,6
	7	321	280	83,4	58,3	6,9	40,9	42,1
	8	442	410	60,8	61,1	6,6	62,3	61,2
Média	-	323,3	274,4	80,9	57,7	6,2	44,1	45,6
Média Geral	-	370,8	320,5	77,2	56,8	10,2	33,6	35,6

Fonte: O autor, 2013.

Em média, as turmas TIAE apresentaram a quantidade de 10 participantes por sessão de bate-papo. Para confirmar que as estimativas produzidas pelos modelos elaborados possuem precisão aceitável, foi necessário realizar um segundo estudo sobre uso de bate-papo na educação on-line conforme apresentado a seguir.

Para complementar o *corpus* de análise da pesquisa, foi realizado um estudo de caso na disciplina DSI da UNIRIO no primeiro semestre de 2016. O estudo teve a intenção tanto de aprimorar o entendimento sobre as atividades elementares realizadas pelos participantes de bate-papo, quanto confirmar a influência do tamanho do grupo na participação dos usuários.

Como discutido na sessão anterior, nas turmas TIAE não houve uma variação significativa entre o número de participantes (tamanho do grupo) nas sessões de bate-papo. No estudo realizado, buscou-se a obtenção de três informações importantes: a taxa de leitura dos alunos da disciplina, a taxa de digitação e a média de mensagens produzidas por alunos em sessões de bate-papo com diferentes configurações de tamanho do grupo.

A disciplina DSI 2016.1 teve 32 alunos matriculados. Com esse cenário, foram realizadas quatro sessões de bate-papo. Para o estudo, esperava-se a distribuição de participantes da seguinte forma: 5, 10, 15 e 30 alunos. Porém, a real participação foi: 6, 8, 7, 15. O assunto dos debates foi Bate-papo na Educação. Um dos alunos da disciplina atuou como moderador, porém a moderação se limitou a apresentar tópicos para debate e coordenar uma votação a cada tópico discutido. Os tópicos foram sobre pontos positivos

e negativos do uso do bate-papo na educação on-line. Os debates tiveram duração de 60 minutos. Ao final, foram feitas entrevistas (bate-papo entre 2 pessoas) com 11 alunos voluntários. Dentre essas entrevistas, uma teve participação simultânea de 3 pessoas (Edmilson, Tadeu e Thiago); e outra, de 4 pessoas (Edmilson, Mônica, Maria Teresa, Luiza).

Na Tabela 2, é apresentado o sumário das 4 sessões de bate-papo e 8 entrevistas realizadas na turma DSI 2016.1.

Tabela 2. Sumário das sessões de bate-papo realizadas na turma DSI 2016.1

Sessão	duração (min)	caracteres	mensagens	participantes	msg/partic	msg/part projetada (60 min)	carac/msg
1	57	16716	499	6	83,17	87,54	33,50
2	38	12359	205	8	25,63	40,46	60,29
3	33	8931	179	6,5	27,54	50,07	49,89
4	60	29566	581	15	38,73	38,73	50,89
5	18,9	3676	40	2	20	63,5	91,9
6	24,8	4420	77	2	38,5	93,1	57,4
7	19,9	3165	41	2	20,5	61,8	77,2
8	30,8	7395	145	2	72,5	141,2	51,0
9	70,9	11647	184	2,8	65,7	55,6	63,3
10	30,8	5039	78	2	39	76,0	64,6
11	45,4	8808	194	3,1	62,6	82,7	45,4
12	44,8	5583	55	2	27,5	36,8	101,5
Média	39,5	9775,4	189,8	4,5	43,4	69,0	62,2

Fonte: O autor, 2016.

Conforme já relatado anteriormente, foi projetado um estudo para 5, 10, 15 e 30 participantes. Contudo, alguns alunos tiveram indisponibilidade para participar do dia e horário definidos para as sessões de bate-papo. Na Figura 7, pode-se observar que o valor expresso na coluna msg/part projetada (60 minutos) para a sessão 4 de bate-papo (15 participantes) ficou bem abaixo das sessões 1 e 3, com 6 e 6,5 participantes. É importante esclarecer que o valor em decimal para participante significa que algum participante chegou atrasado na sessão ou saiu antes do seu término. Embora a sessão 1 apresente um número elevado de produção de mensagens por participante (83,17), observa-se que naquele debate as mensagens produzidas foram mais curtas (33,5 caracteres). Nesse caso, apenas o número de participantes não é suficiente para inferir que a produção de mensagens é maior em pequenos grupos (até 5 ou 6 participantes).

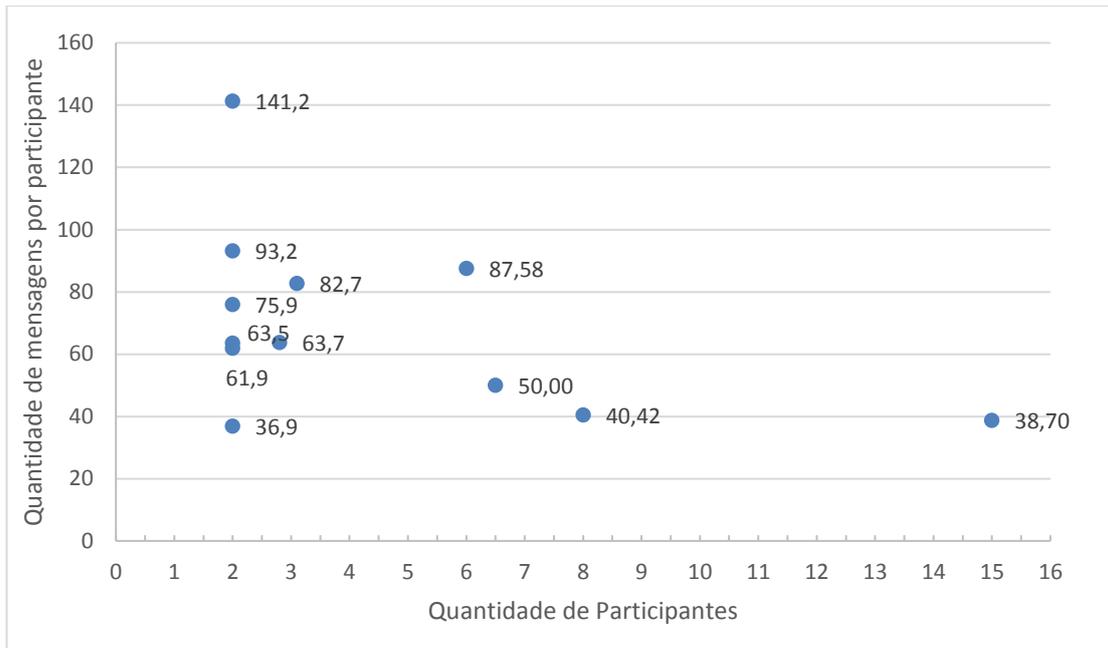


Figura 7. Distribuição dos participantes nas sessões da turma DSI 2016.1
 Fonte: O autor, 2016.

4. Caracterização das atividades elementares: ler, digitar e refletir

Durante uma sessão de bate-papo, o participante lê as mensagens publicadas na sessão, digita algumas mensagens e fica refletindo em alguns momentos. A realização dessas atividades depende de alguns fatores como a velocidade de digitação do participante, e também depende de alguns eventos como a chegada de uma mensagem postada por um outro participante. Conforme detalhado na seção 4.1, os participantes do bate-papo leem em velocidades diferentes e a velocidade de leitura de cada participante é afetada em função do tamanho da mensagem. Os participantes do bate-papo também digitam em velocidades diferentes e o tempo de digitação é proporcional à quantidade de caracteres da mensagem digitada, como abordado na seção 4.2. Os participantes de um bate-papo não ficam apenas lendo e digitando mensagens uma após a outra; eles também permanecem por determinados momentos sem realizar nenhuma dessas atividades – esse comportamento foi aqui denominado “reflexão”, descrito na seção 4.3.

4.1 Leitura

Uma pessoa não lê igual à outra pessoa. A velocidade de leitura depende de algumas características do indivíduo como o nível de escolaridade, a experiência em bater papo e o grau de interesse e de conhecimento sobre o assunto em discussão. O tempo de leitura de uma mensagem também depende de fatores externos ao sujeito, como o tamanho da mensagem, o sistema de bate-papo usado e a complexidade do assunto em discussão.

Para investigar que fatores externos influenciam o processo de leitura, nesta pesquisa foram realizados experimentos para verificar a relação entre o tempo de leitura e certos fatores externos. Segundo Filippo, Pimentel e Wainer (2011, p. 379), em um experimento, o pesquisador busca identificar relações entre variáveis a fim de comprovar ou refutar hipóteses que levam à formulação de leis e teorias gerais. O experimento se desenvolve num ambiente laboratorial, artificial ou semiartificial, onde o pesquisador tem

controle sobre as variáveis: fixa algumas e varia outras. As variáveis são medidas com precisão e os dados coletados são analisados estatisticamente.

Nesta pesquisa, durante o mestrado, foi realizado um experimento com a turma da disciplina Sistemas Colaborativos 2011.2 (SisColab 2011.2), e esse experimento foi replicado na turma Docência em Sistemas de Informação 2016.1 (DSI 2016.1) durante o doutorado. No primeiro experimento (SisColab 2011.2), foram avaliados os tempos de leitura de 6 alunos voluntários: 4 homens e 2 mulheres, com idade entre 24 e 55 anos. No segundo experimento (DSI 2016.1), também foram avaliados 6 alunos: 2 homens e 4 mulheres, com idade entre 30 e 60 anos. Todos aceitaram ter seu nome real divulgado nos resultados desta pesquisa.

Para medir o tempo de leitura, cada indivíduo leu, pela tela do computador, mensagem a mensagem de um *log* real de bate-papo educacional – à medida que pressionavam uma tecla, uma nova mensagem era apresentada e era registrado o tempo de leitura decorrido entre uma tecla e a seguinte. Dessa maneira, foi possível registrar o tempo empregado por cada aluno ao ler cada mensagem do *log*, como exemplificam os dados apresentados na Tabela 3.

Tabela 3 Registro do tempo de leitura (em segundos) de cada aluno ao ler cada mensagem do *log*

Mensagem a ser lida:	Tempo de leitura (em segundos)											
	SisColab 2011.2						DSI 2016.1					
	Fernando	Joanne	Juliana	Manuel	Rafael	Ronaldo	Aline	Evamilda	Gustavo	Lilian	Pedro	Silvia
14:08 <Renato> ok	0,84	0,78	1,42	0,70	0,92	1,76	2,77	7,05	1,50	3,22	0,87	1,24
14:09<Bianca> vamos	0,97	0,69	1,11	0,67	0,69	0,80	2,52	0,93	0,36	4,95	0,93	1,22
14:09<Filipe> Sem problemas por mim...	1,25	0,98	1,17	0,85	0,66	1,13	4,47	0,83	0,20	1,18	0,94	1,46
14:09<Viviane> Bem, como conversamos antes de cair (rs) a maioria leu o texto da Claudia Dias, certo?	3,78	3,48	4,61	3,03	4,56	3,73	3,98	6,88	0,79	5,23	3,70	4,79
14:09<André> ok	1,06	0,70	2,09	0,64	0,63	1,19	2,26	0,79	0,53	1,15	0,71	0,88
14:09<Casimiro> certo	0,91	0,70	0,78	0,56	1,59	0,60	1,05	0,63	0,23	0,82	0,64	1,24
14:09<Renato> isso	0,91	0,50	0,91	0,57	0,87	0,53	0,82	0,58	0,23	0,88	0,52	0,86
14:09<Filipe> sim	1,00	0,54	0,53	0,41	0,77	0,52	1,05	0,56	0,22	1,43	0,50	1,01
14:10<Bianca> isso mesmo	1,34	0,92	0,65	1,68	0,62	0,55	2,38	1,15	0,22	1,58	0,47	0,88
14:10<Filipe> sim	0,87	0,65	0,47	0,64	1,18	0,60	1,98	0,61	0,22	0,71	0,46	2,82
14:10<Viviane> então vamos começar com o conceito de usabilidade	2,23	3,91	1,89	1,94	2,00	1,79	2,81	2,05	0,22	2,36	1,52	1,87
14:10<Raphael> boa tarde	1,33	1,28	1,04	1,46	1,13	4,25	2,06	0,76	0,23	1,13	0,87	0,89
14:11<Raphael> desculpe o atraso	1,25	2,44	0,85	0,81	1,00	0,03	1,58	1,06	0,26	1,78	0,86	0,90
14:11<Viviane> o que vocês consideram usabilidade depois de ler esse texto? Tinham algum conceito anterior ou nunca leram nada sobre o assunto?	6,31	5,04	5,06	8,69	4,76	4,54	5,27	1,47	3,21	6,20	4,38	7,38
14:11<Renato> acho que é quando um sistema pode ser utilizado por diversos tipos de usuários sem deixar a desejar	5,02	4,72	4,77	3,28	4,29	3,11	10,8	6,91	1,90	5,86	3,72	7,68

...	
Média:	60,2caracteres	/	4,03	3,91	3,33	2,01	2,96	3,46	3,45	4,13	1,41	3,04	2,51	3,27
mensagem														
Desvio Padrão:	59,6 caracteres		2,98	3,71	2,80	2,02	2,58	3,38	2,65	6,66	1,48	3,54	2,14	2,93
Tempo médio de leitura:			3,13s											

Fonte: O autor, 2016.

Nesta pesquisa, buscou-se investigar se os alunos têm velocidade diferente de leitura, e se o tempo de leitura é influenciado pelo tamanho da mensagem. Portanto, objetivou-se testar as seguintes hipóteses:

(H1) Os alunos têm velocidades diferentes de leitura;

(H2) O tempo de leitura é proporcional à quantidade de caracteres da mensagem.

Para testar uma hipótese, tenta-se rejeitar a hipótese nula, ou seja, uma hipótese inversa ao que se deseja chegar (WAINER, 2007, p. 230):

Testes estatísticos são procedimentos que fazem uma particular pressuposição sobre os dados, a chamada **hipótese nula**, e mais uma série de outras pressuposições (que chamaremos de **condições do teste**) sobre os dados, e calculam a probabilidade que alguma **propriedade relacionada** aos dados seja verdadeira, dadas as pressuposições. Essa probabilidade calculada é chamada **valor p** ou (*p-value*). Se as condições do teste são verdadeiras, e o *p-value* é suficientemente baixo, então o pesquisador pode assumir que a hipótese nula é falsa, ou como é normalmente fraseado, “há evidências suficientes para rejeitar a hipótese nula”. O valor do *p-value* abaixo do qual se assume que a hipótese é falsa é usualmente 0.05 ou 0.01. (WAINER, 2007, p. 230)

Dessa forma, nesta pesquisa foi preciso testar as seguintes hipóteses nulas (relacionadas às hipóteses listadas anteriormente), onde T_l representa o tempo médio de leitura das mensagens.

(1) $H_{01} = T_l \text{ Aluno1} = T_l \text{ Aluno2} = T_l \text{ Aluno3} = T_l \text{ Aluno4} = T_l \text{ Aluno5} = T_l \text{ Aluno6}$

(2) $H_{02} = T_l \text{ MensagemPequena} = T_l \text{ MensagemMédia} = T_l \text{ MensagemGrande}$

Para testar essas hipóteses, cada aluno deveria ler 3 *logs* reais de bate-papo. As hipóteses H_{01} e H_{02} foram testadas a partir do tempo de leitura dos alunos nos três *logs*.

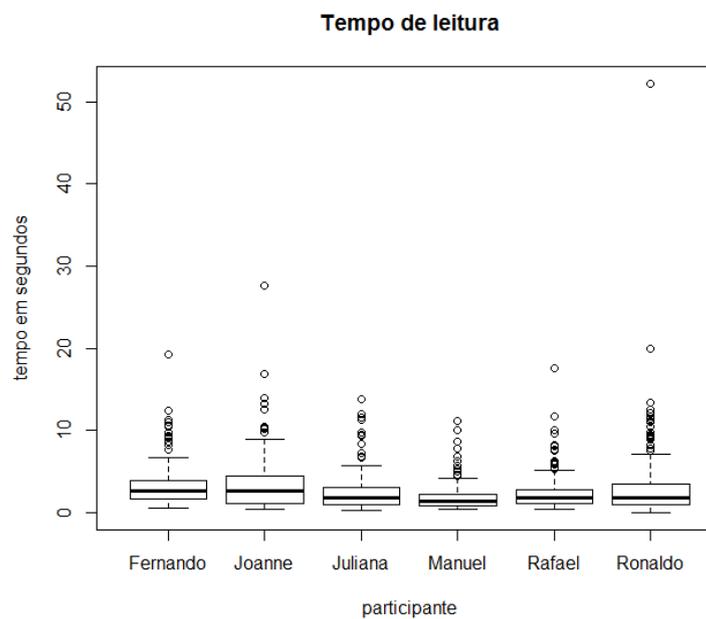
Os dados coletados no experimento foram categorizados e agrupados em uma base de dados para a realização dos testes estatísticos. As categorias em que os dados foram distribuídos são: Participante (pessoa que leu as mensagens do *log*), *Log* e Tamanho.

Todas as análises foram realizadas com o uso do *software* estatístico R³ e o nível de significância adotado foi de 5%.

Com a intenção de verificar qual melhor conjunto de testes a ser usado, testou-se inicialmente se os dados seguiam a distribuição normal por meio do teste Shapiro-Wilk. Como o valor p obtido pelo teste foi inferior a 0,05, a distribuição dos tempos de leitura não é normal em nenhuma das duas turmas: SisColab 2011.2 ($W = 0.64217$, $p\text{-value} < 2.2e-16$) e DSI 2016.1 ($W = 0.61033$, $p\text{-value} < 2.2e-16$). Dado que foi violado o pressuposto de normalidade, deve ser utilizado um método não paramétrico.

4.1.1. Tempo de leitura é diferente entre as pessoas

Para testar se o tempo de leitura difere de pessoa para pessoa, foram considerados os tempos medidos de cada pessoa (voluntária da turma SisColab 2011.2 para participar do experimento) ao ler todas as mensagens dos três *logs*. Com isso, obteve-se a média e a dispersão dos tempos caracterizados no diagrama de caixa (*boxplot*) da Figura 8.



³ <<http://www.r-project.org>>

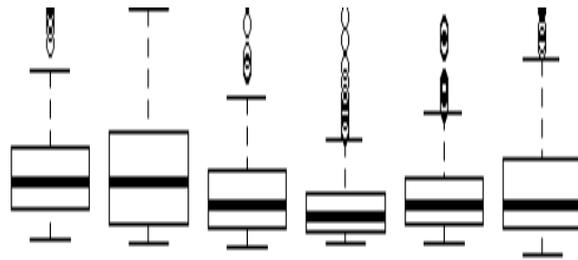


Figura 8 *Boxplot* dos tempos de leitura dos três *logs* na turma SisColab 2011.2

Fonte: O autor, 2019.

Para verificar se os alunos leem de forma diferente, é preciso testar se os dados produzidos por cada aluno vêm de uma população diferente dos demais. Para esta finalidade, foi usado o teste de Kruskal-Wallis, que é o teste não paramétrico utilizado na comparação de três ou mais amostras independentes (BLACK, 2004); ele nos indica se há diferença entre pelo menos duas amostras. A aplicação do teste utiliza os valores numéricos transformados em postos e agrupados num só conjunto de dados. A comparação dos grupos é realizada por meio da média dos postos (posto médio).

O teste de Kruskal-Wallis sobre os tempos de leitura dos alunos SisColab 2011.2 (chi-squared = 125.96, df = 5, *p-value* < 2.2e-16) resultou em um valor *p* inferior a 0,05. Isso significa que os tempos de leitura dos alunos da turma SisColab 2011.2 não têm a mesma distribuição, ou seja, foi possível rejeitar a hipótese nula e afirmar que as pessoas leem em tempos diferentes.

Procedimento semelhante foi repetido para avaliar o tempo de leitura na turma DSI 2016.1, contudo nessa turma foi pedido que os alunos voluntários lessem apenas o *log* 3 (não leram os outros dois *logs*). Do resultado do teste (chi-squared = 68.716, df = 5, *p-value* = 1.895e-13), em função do valor *p* inferior a 0,05, pode-se afirmar que aquelas pessoas também leem em tempos diferentes.

Para obter um resultado mais detalhado, foram comparados os tempos de leitura entre cada duas pessoas por meio de um teste *a posteriori* não paramétrico para comparações múltiplas, disponível para o R no pacote nparcomp⁴. Verificou-se que, para a turma SisColab 2011.2, não há diferença significativa no tempo de leitura entre os alunos Fernando e Joanne (eles leem com a mesma velocidade), como também não há diferença

⁴ O teste de Kruskal-Wallis só consegue detectar uma diferença significativa entre as seis amostras (participantes) como um grupo. Podemos dizer que existe pelo menos uma diferença significativa entre as amostras com a maior e a menor soma de postos. Mas podem ocorrer outras diferenças. A interface de fórmula permite rodar um teste *post-hoc* usando o pacote nparcomp, para determinar entre quais amostras ocorrem diferenças específicas. <<http://labmолuscos.files.wordpress.com/2012/04/rtutorial.pdf>>

significativa entre os alunos Juliana, Rafael e Ronaldo (esses alunos também leem com a mesma velocidade, mas esse grupo lê numa velocidade diferente do grupo anterior).

Tabela 4 Resultado da comparação entre pares de alunos da SisColab 2011.2

	Comparação	Estimador	Limite Inferior	Limite Superior	Estatística	valor p
1	p(Fernando , Joanne)	0,484	0,415	0,553	-0,66	0,990
2	p(Fernando , Juliana)	0,362	0,297	0,426	-6,127	0,000
3	p(Fernando , Manuel)	0,247	0,191	0,302	-12,987	0,000
4	p(Fernando , Rafael)	0,346	0,283	0,409	-6,942	0,000
5	p(Fernando , Ronaldo)	0,384	0,317	0,45	-5,013	0,000
6	p(Joanne , Juliana)	0,408	0,341	0,475	-3,946	0,001
7	p(Joanne , Manuel)	0,315	0,252	0,378	-8,403	0,000
8	p(Joanne , Rafael)	0,394	0,327	0,462	-4,498	0,000
9	p(Joanne , Ronaldo)	0,416	0,35	0,483	-3,6	0,005
10	p(Juliana , Manuel)	0,39	0,324	0,455	-4,81	0,000
11	p(Juliana , Rafael)	0,491	0,423	0,559	-0,377	1,000
12	p(Juliana , Ronaldo)	0,497	0,429	0,565	-0,134	1,000
13	p(Manuel , Rafael)	0,603	0,538	0,669	4,489	0,000
14	p(Manuel , Ronaldo)	0,587	0,519	0,654	3,684	0,004
15	p(Rafael , Ronaldo)	0,506	0,437	0,574	0,244	1,000

Fonte: O autor, 2019.

Para a turma DSI 2016.1, foi encontrada a mesma distribuição para o tempo de leitura dos seguintes alunos: Aline, Lilian e Silvia; Evanilda, Lilian, Pedro e Silvia; Lilian, Pedro e Silvia (conforme o valor p maior que 0,05 apresentado na Tabela 5).

Tabela 5 Resultado da comparação entre pares de alunos na turma DSI 2016.1

	Comparação	Estimador	Limite Inferior	Limite Superior	Estatística	valor p
1	p(Aline , Evanilda)	0,36	0,242	0,477	-3,452	0,009
2	p(Aline , Gustavo)	0,177	0,094	0,261	-11,122	0,000
3	p(Aline , Lilian)	0,385	0,267	0,502	-2,834	0,057
4	p(Aline , Pedro)	0,363	0,25	0,476	-3,505	0,008
5	p(Aline , Silvia)	0,44	0,322	0,558	-1,466	0,699
6	p(Evanilda , Gustavo)	0,328	0,22	0,437	-4,57	0,000
7	p(Evanilda , Lilian)	0,484	0,364	0,604	-0,375	1,000
8	p(Evanilda , Pedro)	0,512	0,392	0,632	0,29	1,000
9	p(Evanilda , Silvia)	0,584	0,466	0,702	2,051	0,322
10	p(Gustavo , Lilian)	0,656	0,544	0,769	4,012	0,001
11	p(Gustavo , Pedro)	0,7	0,595	0,806	5,491	0,000

12	p(Gustavo , Silvia)	0,767	0,672	0,861	8,15	0,000
13	p(Lilian , Pedro)	0,509	0,389	0,63	0,225	1,000
14	p(Lilian , Silvia)	0,58	0,462	0,699	1,953	0,378
15	p(Pedro , Silvia)	0,584	0,468	0,7	2,086	0,301

Fonte: O autor, 2019.

Portanto, mesmo que não exista diferença estatisticamente significativa no tempo de leitura entre algumas pessoas, ao considerar todas as pessoas do experimento, a conclusão é que o tempo de leitura é diferente entre as pessoas em geral.

Para comparar os tempos de leitura entre as pessoas das duas turmas, deve-se considerar somente os tempos de leitura das mensagens do *log 3*, porque foi o único *log* lido pelos dois grupos de voluntários. A Figura 9 apresenta o *boxplot* sobre os tempos de leitura dos voluntários das turmas SisColab 2011.2 e DSI 2016.1.

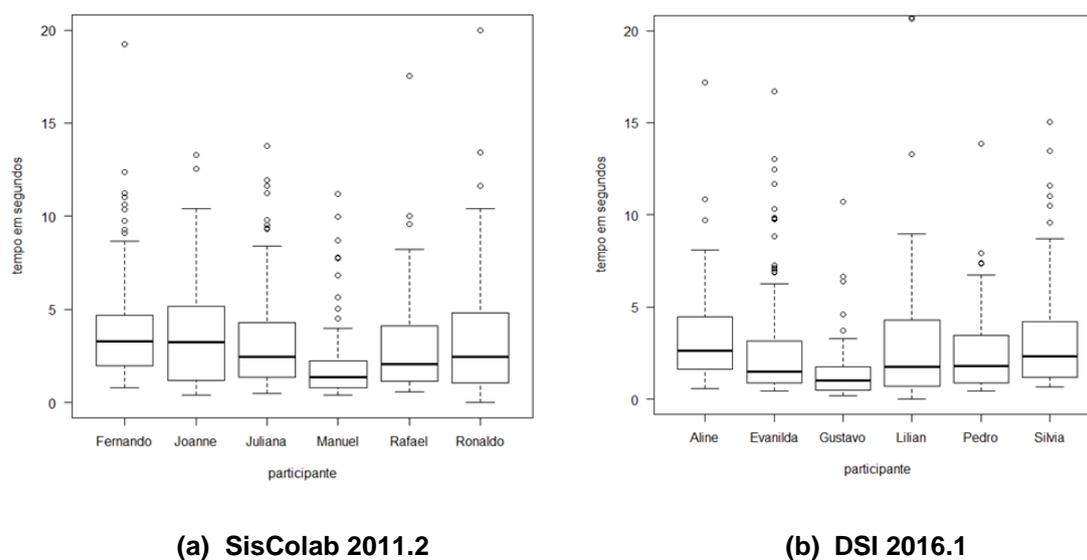


Figura 9 *Boxplot* do tempo de leitura do *log 3*

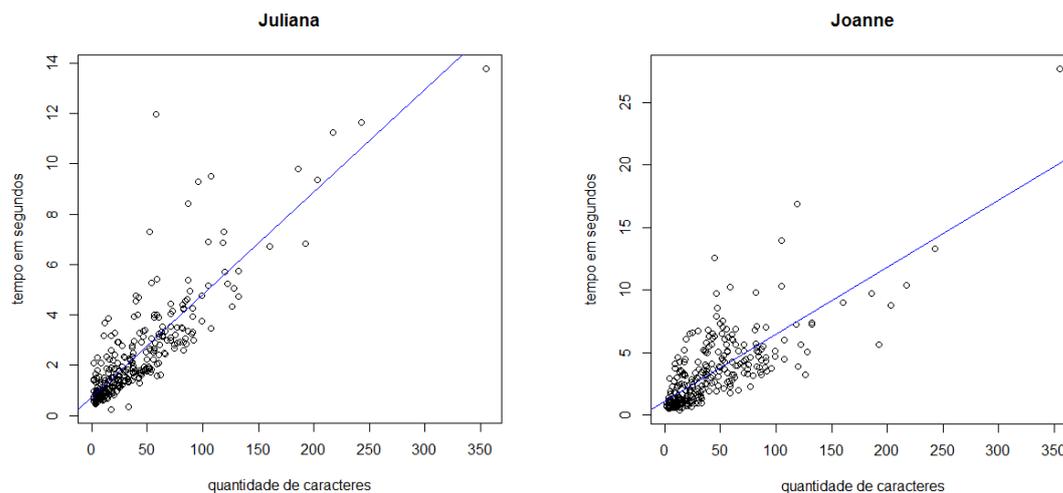
Fonte: O autor, 2019.

Conforme apresentado na Figura 9, pode-se identificar que o tempo médio de leitura das mensagens compreende entre 0 e 5 segundos. Quanto ao limite superior, para dois alunos na turma SisColab 2011.2 ultrapassa 10 segundos, enquanto que na turma DSI 2016.1 todos ficam abaixo de 10 segundos. Os pontos representados acima do limite superior são os *outliers*. Também por meio da Figura 9, é possível identificar que a linha que representa a mediana aparece praticamente na mesma altura para a dupla de alunos Fernando e Joanne, assim como está praticamente na mesma altura a linha que representa

a mediana dos alunos Juliana, Rafael e Ronaldo na turma SisColab 2011.2. Ainda sobre a Figura 9, pode-se observar que os alunos Manuel (SisColab 2011.2) e Gustavo (DSI 2016.1) são os alunos cujo tempo de leitura mais se diferencia nas suas respectivas turmas e também são os alunos que leem mais rápido.

4.1.2. *Tempo de leitura é proporcional à quantidade de caracteres da mensagem*

Como pode ser visualizado na Figura 10, nota-se uma tendência linear⁵ diretamente proporcional entre o tempo de leitura e a quantidade de caracteres da mensagem na turma SisColab 2011.2.



⁵ A regressão linear é uma equação para estimar-se a condicional (valor esperado) de uma variável y, dados os valores de algumas outras variáveis x. <https://pt.wikipedia.org/wiki/Regress%C3%A3o_linear>.

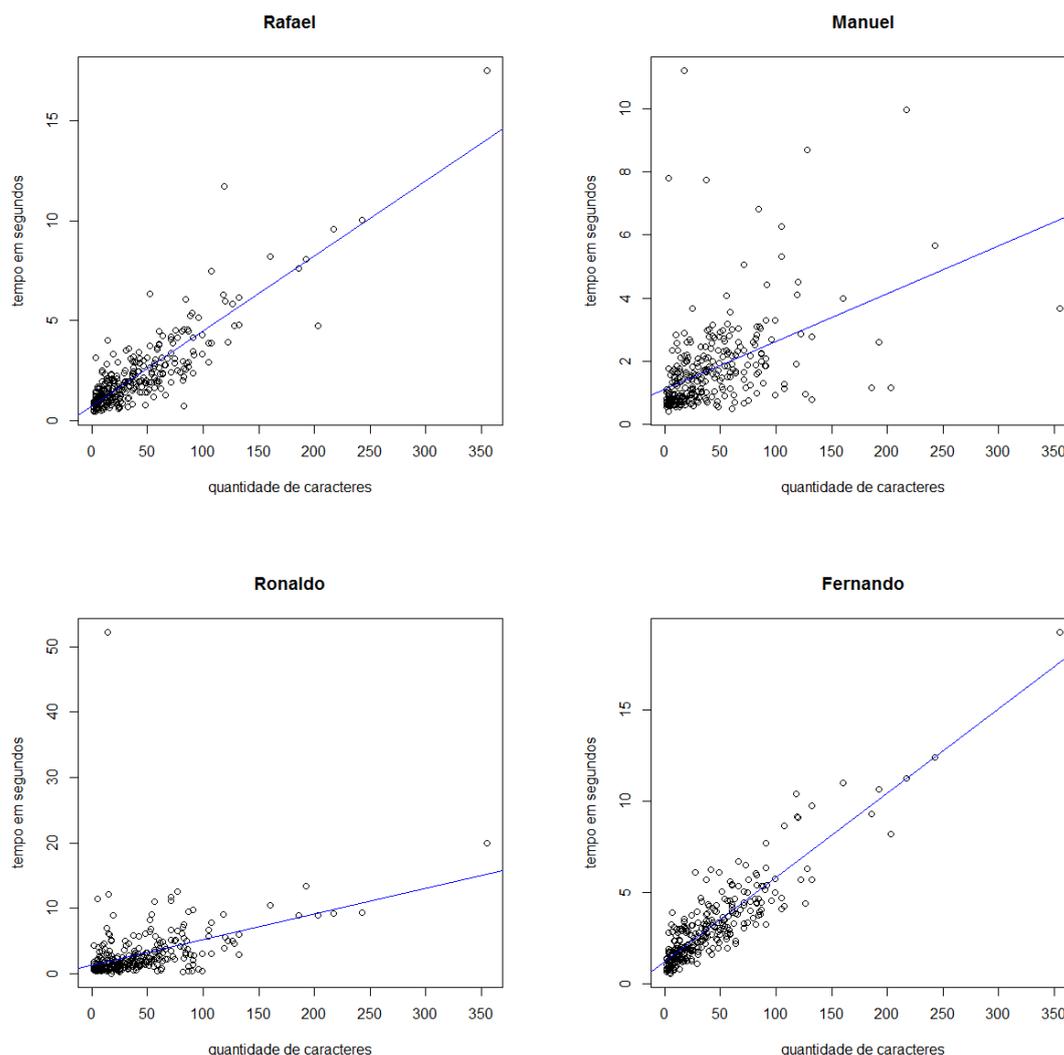


Figura 10. Correlação entre o tempo de leitura e a quantidade de caracteres das mensagens dos três logs na turma SisColab 2011.2

Fonte: O autor, 2019.

Para verificar esta relação, é preciso testar a segunda hipótese nula (H_02) que diz que o tempo de leitura é diferente para diferentes tamanhos de mensagem. Para realizar este teste, foi necessário classificar as mensagens em três categorias (pequena, média e grande). A divisão do tamanho das mensagens foi feita pela frequência com que cada quantidade de caracteres apareceu nos logs de bate-papo que os alunos leram. As categorias foram assim divididas: mensagens pequenas entre 2 e 15 caracteres (594 ocorrências); mensagens médias, entre 16 e 45 caracteres (594 ocorrências); mensagens grandes, entre 46 e 355 caracteres (611 ocorrências). Foi realizado o teste de Kruskal-Wallis para identificar se há diferença significativa entre as três categorias de tamanho de mensagens. O resultado do teste ($\chi^2 = 748.66$, $df = 2$, $p\text{-value} < 2.2e-16$) apontou

um valor $p < 0,05$, em que se rejeita a hipótese nula (H_0) de que os tempos de leitura para diferentes tamanhos de mensagem pertencem a uma mesma distribuição. O teste *a posteriori* de comparação entre pares de categorias (pacote `nparcomp`) confirma que nenhuma das três categorias vem da mesma distribuição conforme o valor p apresentado na Tabela 6.

Tabela 6 Resultado da comparação entre pares de tamanhos de mensagem na turma SisColab 2011.2

	Comparação	Estimador	Limite Inferior	Limite Superior	Estatística	Valor p
1	p (pequena , média)	0,755	0,722	0,788	18,1	0
2	p (pequena , grande)	0,911	0,89	0,932	46,5	0
3	p (média , grande)	0,788	0,757	0,819	22	0

Fonte: O autor, 2019.

Buscou-se identificar a força da correlação entre o tamanho da mensagem e o tempo de leitura. Como o pressuposto de normalidade não foi atendido, conforme constatado pelo teste de Shapiro-Wilk, o método de correlação a ser utilizado é o método de Spearman ($S = 293072043$, $p\text{-value} < 2.2e-16$). Para esse teste, o nível de correlação encontrado foi forte (0,698)⁶. Esse resultado mostra que, quanto mais caracteres na mensagem, maior é o tempo de leitura.

Considerando os resultados dos testes, foram reunidos os dados de todos os participantes para calcular a média das equações lineares, conforme apresentado na Figura 11.

⁶ De acordo com a classificação de Shimakura (2006), tem-se que: para valores de 0,00 a 0,19, a correlação é bem fraca; de 0,20 a 0,39, a correlação é fraca; de 0,40 a 0,69, a correlação é moderada; de 0,70 a 0,89, a correlação é forte; e de 0,90 a 1,00, a correlação é muito forte.

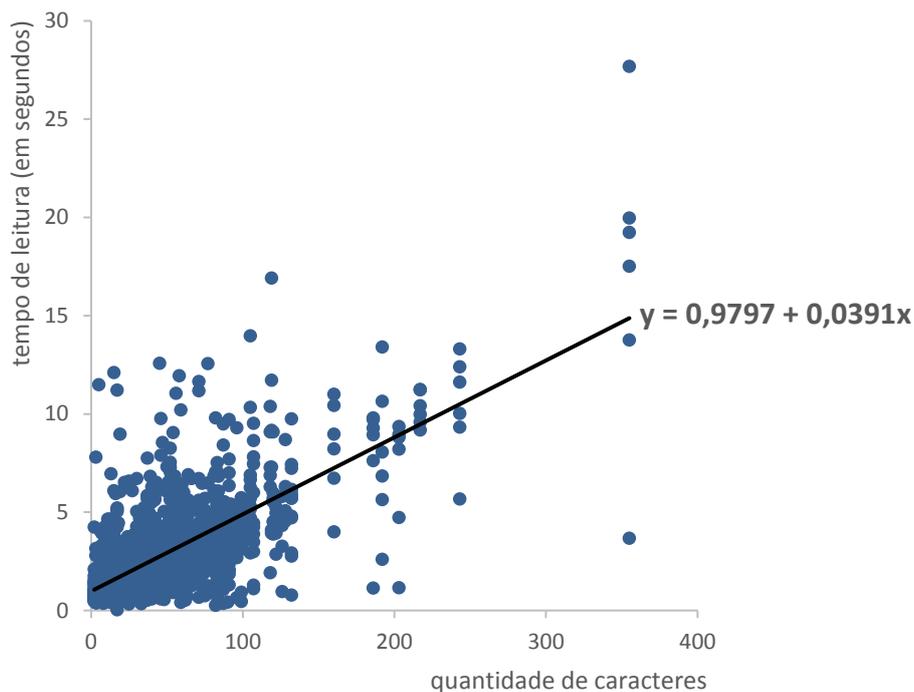


Figura 11. Equação linear considerando os tempos de leitura dos 6 voluntários da turma SisColab 2011.2 em relação à quantidade de caracteres das mensagens dos três logs

Fonte: O autor, 2019.

A partir desse resultado, é assumido que o tempo médio de leitura T_l pode ser modelado por uma equação linear em função da quantidade de caracteres da mensagem dada pela Equação (1).

$$T_l = 1,0 + (0,04 * \text{número caracteres por mensagem}) \quad (1)$$

onde T_l é o tempo médio de leitura medido em segundos.

É preciso ressaltar que essa equação foi obtida considerando o tempo de leitura de 6 alunos da turma SisColab 2011.2, e que, para outro perfil de alunos, os coeficientes da equação serão diferentes. Para avaliar os coeficientes em outro contexto, os mesmos estudos foram replicados para os dados da turma DSI 2016.1.

Como pode ser visualizado na Figura 12, sobre os dados de voluntários da turma DSI 2016.1, também se nota uma tendência linear diretamente proporcional entre o tempo de leitura e a quantidade de caracteres da mensagem.

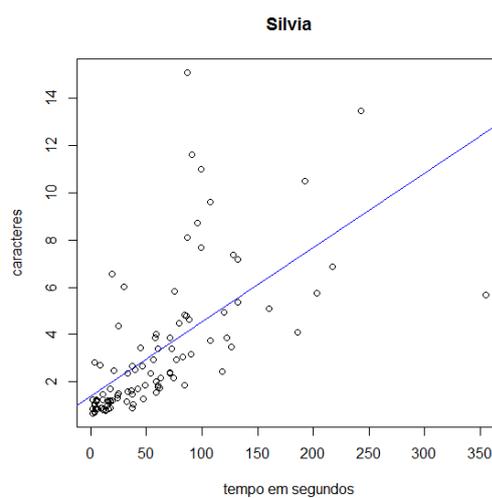
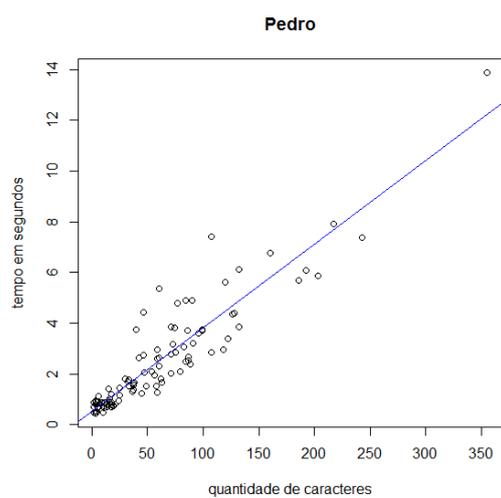
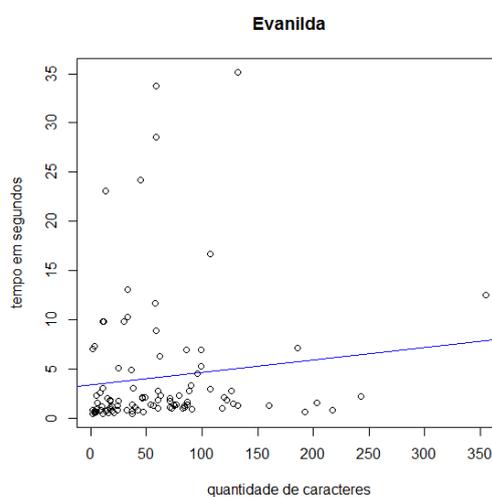
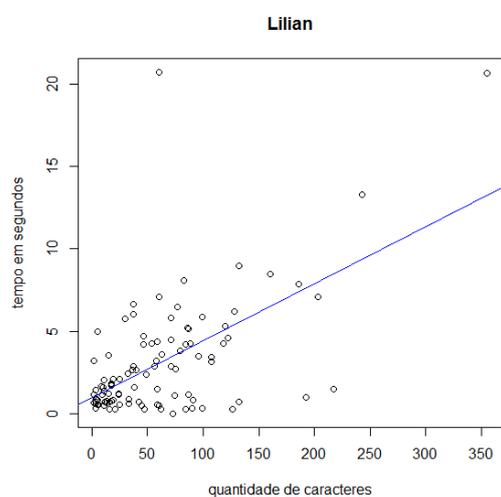
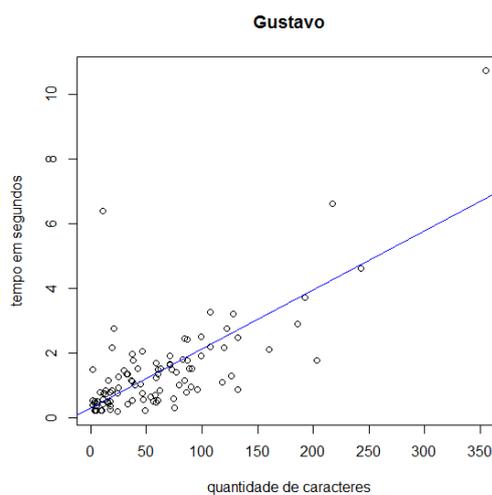
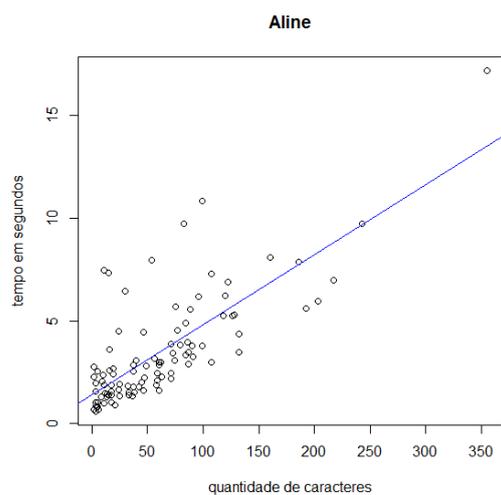


Figura 12 Gráfico de dispersão do tempo de leitura da turma DSI 2016.1
 Fonte: O autor, 2019.

Para verificar a relação, foram replicados os mesmos testes já explicados anteriormente. Nessa turma, contudo, os alunos leram apenas as mensagens do *log* 3. As categorias de tamanhos de mensagens também foram divididas em três, conforme a frequência de ocorrências: mensagens pequenas, entre 2 e 21 caracteres (198 ocorrências); mensagens médias, entre 22 e 71 caracteres (204 ocorrências); mensagens grandes, entre 72 e 355 caracteres (198 ocorrências). Pelo teste de Kruskal-Wallis, foi possível concluir que as distribuições dos tempos de leitura das três categorias de tamanho de mensagem são diferentes ($\chi^2 = 180.08$, $df = 2$, $p\text{-value} < 2.2e-16$) e o teste *a posteriori*, na comparação por pares, também mostrou que há diferença significativa entre as distribuições dos tempos de leitura das categorias de mensagens, conforme dados apresentados na Tabela 7.

Tabela 7 Resultado da comparação entre pares de tamanhos de mensagem na turma DSI 2016.1

	Comparação	Estimador	Limite Inferior	Limite Superior	Estatística	Valor <i>p</i>
1	p(pequena , média)	0,751	0,697	0,805	10,93	0
2	p(pequena , grande)	0,855	0,801	0,91	15,5	0
3	p(média , grande)	0,688	0,606	0,769	5,43	4,27E-07

Fonte: O autor, 2019.

Sobre a correlação entre as variáveis “tempo de leitura” e “tamanho da mensagem”, considerando os dados da turma DSI 2016.1, pelo teste de correlação de Spearman ($S = 15291067$, $p\text{-value} < 2.2e-16$) foi possível concluir que há diferença significativa entre os tempos de leitura entre as três categorias de tamanho de mensagem, bem como há uma correlação moderada (0,575) entre essas variáveis.

Considerando os resultados dos testes, foram reunidos os dados de todos os participantes para calcular a média das equações lineares, conforme apresentado na Figura 13.

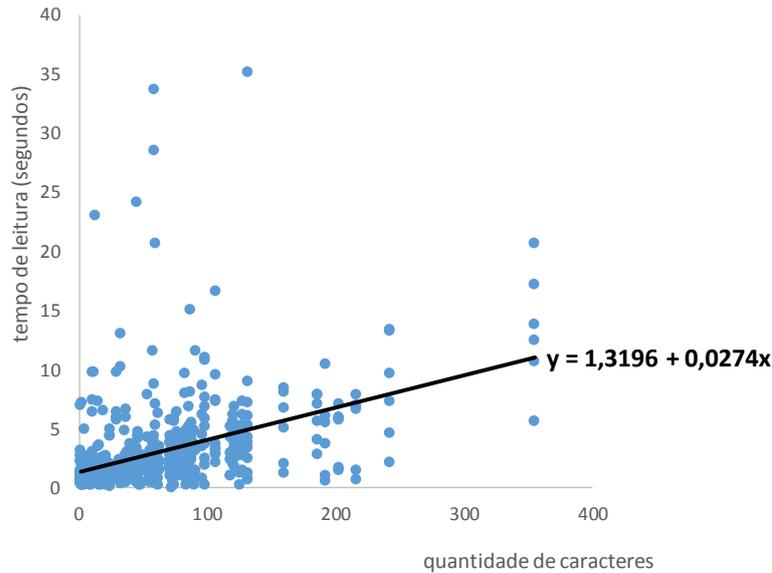


Figura 13. Equação linear considerando os tempos de leitura dos 6 voluntários da turma DSI 2016.1 em relação à quantidade de caracteres das mensagens

Fonte: O autor, 2019.

A partir desse resultado, é assumido que o tempo médio de leitura T_l da turma DSI 2016.1 pode ser modelado por uma equação linear em função da quantidade de caracteres da mensagem dada pela Equação(2).

$$T_l = 1,3 + (0,03 * \text{número caracteres por mensagem}) \quad (2)$$

onde T_l é o tempo médio de leitura medido em segundos.

É preciso ressaltar que as Equações (1) e (2) apresentam coeficientes parecidos porque ambas foram calculadas em função de dados de estudantes de pós-graduação em Informática. Como referência para outros estudos da presente pesquisa, serão considerados os coeficientes da Equação (1) por terem sido calculados a partir de uma quantidade maior de dados (tempos de leitura sobre 3 logs, em vez de apenas 1 log).

4.2. Digitação

Estudos na área de Biometria da Digitação concluíram que cada pessoa possui uma velocidade de digitação específica; não só a velocidade de digitação, mas também os padrões de digitação diferem de um indivíduo para outro: o intervalo entre o pressionar

de uma tecla e a outra, e o tempo em que uma tecla permanece pressionada, são características únicas de cada indivíduo e contribuem para o estabelecimento de um padrão de digitação individual (BUZING, 2003; GAINES *et al.*, 1980; GARCIA, 1986). Dados esses estudos na área de Biometria da Digitação, na presente pesquisa foi assumido que não era preciso investigar se as pessoas digitam em tempos diferentes; isso foi assumido como premissa.

Para a presente pesquisa, o importante é estimar o tempo médio de digitação de mensagens durante a participação em sessões educacionais de bate-papo, sendo necessário medir a velocidade de digitação da mensagem como um todo. Para investigar o processo de digitação de mensagens em bate-papo, na pesquisa do mestrado foi desenvolvido um *script* para capturar o instante em que cada tecla foi pressionada durante a digitação de uma mensagem de bate-papo, conforme ilustrado na Figura 14. Com esse *script*, é possível analisar o comportamento do usuário ao longo da sessão de bate-papo, conforme proposto por Pimentel (2002, p. 26). A partir dos registros dos tempos de digitação, é possível caracterizar o processo de digitação de mensagens em bate-papo.

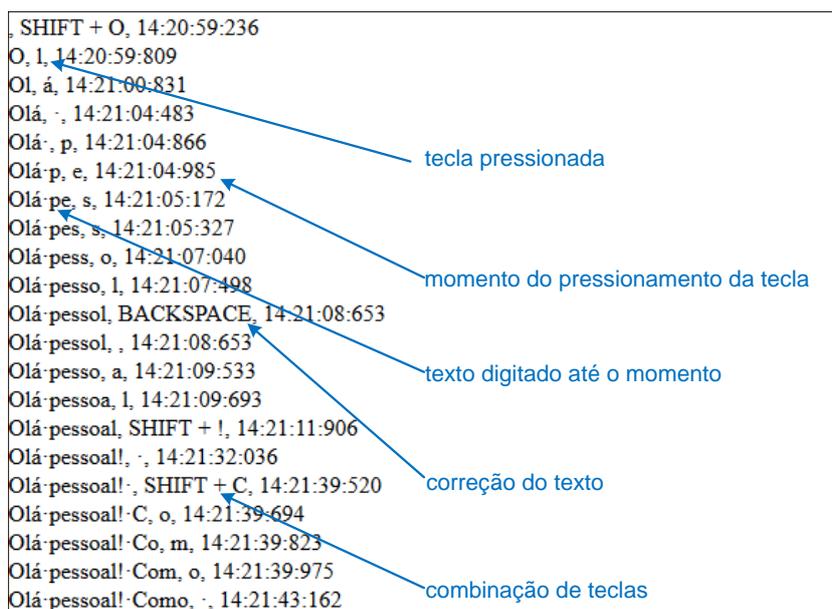


Figura 14. Registro dos tempos com que cada caractere foi pressionado ao longo da digitação da mensagem “Olá pessoal! Como vocês vão?”

Fonte: O autor, 2013.

O *script* de rastreamento de teclas pressionadas foi implementado inicialmente em uma página *web* que simula um bate-papo (Figura 15). Essa implementação possibilitou um estudo em que alguns indivíduos digitaram mensagens e o tempo de digitação foi

medido. Quando o usuário pressionava a tecla *Enter*, o sistema publicava a mensagem digitada na área de mensagens juntamente com informações sobre o processo de digitação.

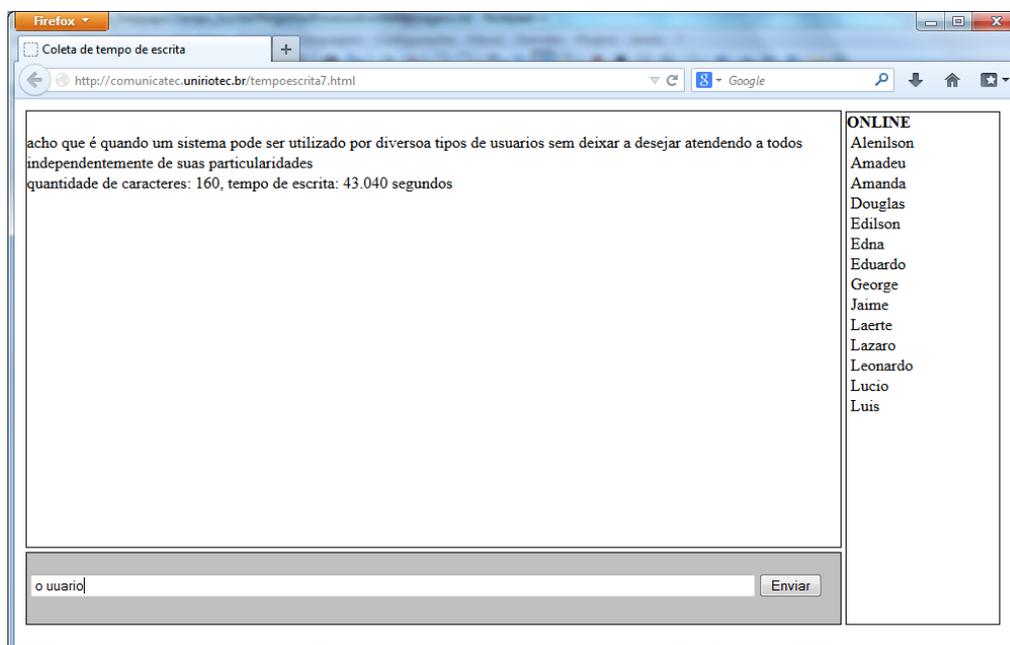


Figura 15. Tela de captura do tempo de digitação de mensagens

Fonte: O autor, 2013.

Dois usuário intensivos em bate-papo participaram desse estudo, ambos do sexo masculino, com 31 e 38 anos. Cada participante digitou três mensagens. Os dados produzidos estão listados na Tabela 8. O tamanho médio das mensagens digitadas foi de 85,2 caracteres por mensagem, e o tempo médio de digitação foi de 24 segundos.

Tabela 8. Dados sobre o processo de escrita de mensagens em bate-papo

Indivíduo	Mensagem	Quantidade de Caracteres	Tempo de Digitação (segundos)
1	Alguma sugestão para melhorarmos a interface proposta?	54	16
	Sim Thiago, essa é a minha proposta de interface para o bate-papo típico do comunicatec.	88	21
	Observe que diminui o tamanho do avatar para caber mais texto na tela.	70	19
2	Muito mal, Dominick. Achei que ia tirar no máximo 3, mas a Michelle me deu 5,5 para eu passar direto.	101	27
	Validar o modelo para podermos simular o bate-papo com mais usuários.	69	19

Quando o Carlos começou a perguntar coisas sobre modelagem de sistemas. Achei que ele ia perguntar alguma coisa que eu não sabia.	129	44
MÉDIA	85,2	24

Fonte: O autor, 2013.

De forma semelhante à realizada para a leitura, buscou-se avaliar a correlação entre o tempo de escrita de mensagens e o tamanho da mensagem.

Ao analisar os dados produzidos, verificou-se uma correlação muito forte (0,95) entre a quantidade de caracteres da mensagem e o intervalo de tempo empregado em sua digitação, sendo em média 1 caractere / 0,28 segundos. Com base nesse estudo, foi elaborada a Equação (3) para a estimativa do tempo médio de escrita de uma mensagem:

$$T_d = 0,28 * \text{número de caracteres da mensagem} \quad (3)$$

onde T_d é o tempo estimado de digitação da mensagem, medido em segundos.

Esse estudo inicial envolveu poucos usuários e foram medidas poucas mensagens, e as mensagens não foram digitadas durante uma sessão real de bate-papo. Para melhor caracterizar a digitação de mensagens em bate-papo, durante a pesquisa de doutorado, o *script* foi implementado no sistema De Bate Papo (MORAES; PIMENTEL, 2010; SILVA; PIMENTEL, 2017), apresentado na Figura 16. Esse sistema foi utilizado em um experimento com a turma DSI 2016.1.



Figura 16. Imagem do sistema De Bate Papo usado nas entrevistas da pesquisa

Ao final da disciplina DSI 2016.1, os alunos que haviam participado de algum dos debates via bate-papo do Facebook foram convidados a participar de uma entrevista pelo De Bate Papo. Nessas entrevistas, foram feitas 5 perguntas abertas (documentadas no Apêndice 1). Além do pesquisador, participaram desse estudo 12 dos 32 alunos inscritos na disciplina, sendo 4 homens e 7 mulheres, com idades entre 25 e 60 anos. No total, foram digitadas 751 mensagens.

As entrevistas foram realizadas nos horários disponíveis de cada participante e deveriam ocorrer individualmente. Entretanto, alguns participantes se juntaram em algumas entrevistas, o que ocasionou a incidência de grupos de 3 e até 4 participantes, incluindo o entrevistador. Na Tabela 9, é apresentado o sumário dos dados sobre a digitação dos voluntários da turma DSI 2016.1.

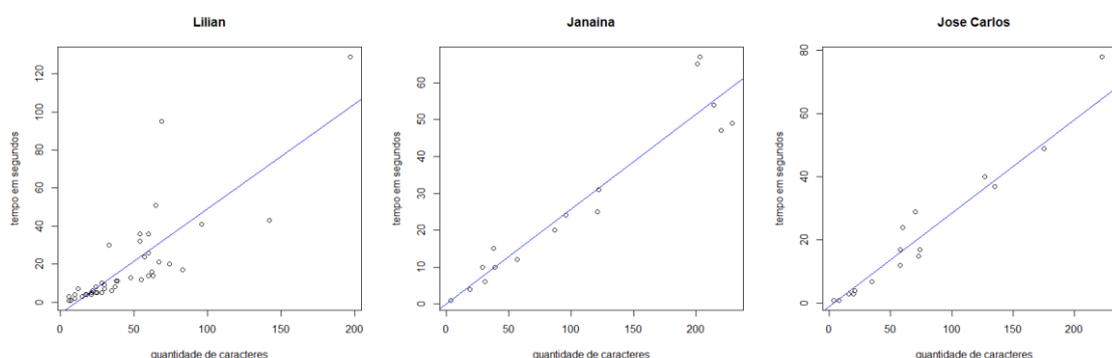
Tabela 9 Sumário dos tempos de digitação na turma DSI 2016.1

Aluno	Quantidade de Mensagens	Caracteres	Tempo (segundos)	Taxa de digitação (seg/car)
Edmilson	293	21913	6256	0,29
Lilian	44	1956	811	0,41
Janaina	18	1718	442	0,26
Jose Carlos	17	1389	392	0,28
Gustavo	87	3895	1334	0,34
Mirian	40	2685	1104	0,41
Maria Evanilda	24	2311	1185	0,51
Tulio	43	2744	646	0,24
Tiago	32	1441	530	0,37
Tadeu	27	2110	471	0,22
Monica	54	2056	859	0,42
Maria Tereza	9	671	272	0,41
Luiza	63	2380	763	0,32
MÉDIA: 0,34s				

Fonte: O autor, 2016.

Cabe ressaltar que esses tempos foram produzidos em situações reais de bate-papo educacional. O bate-papo fluíu naturalmente como uma conversa, sendo que o *script*, durante a digitação das mensagens, registrava também o tempo em que cada tecla estava sendo pressionada, o que possibilitou medir o tempo total em segundos gasto para digitar cada mensagem produzida por cada participante.

Da mesma forma como foi feito para o processo de leitura, buscou-se analisar a correlação entre o tamanho da mensagem (em caracteres) e o tempo de digitação, cujos gráficos são apresentados na Figura 17.



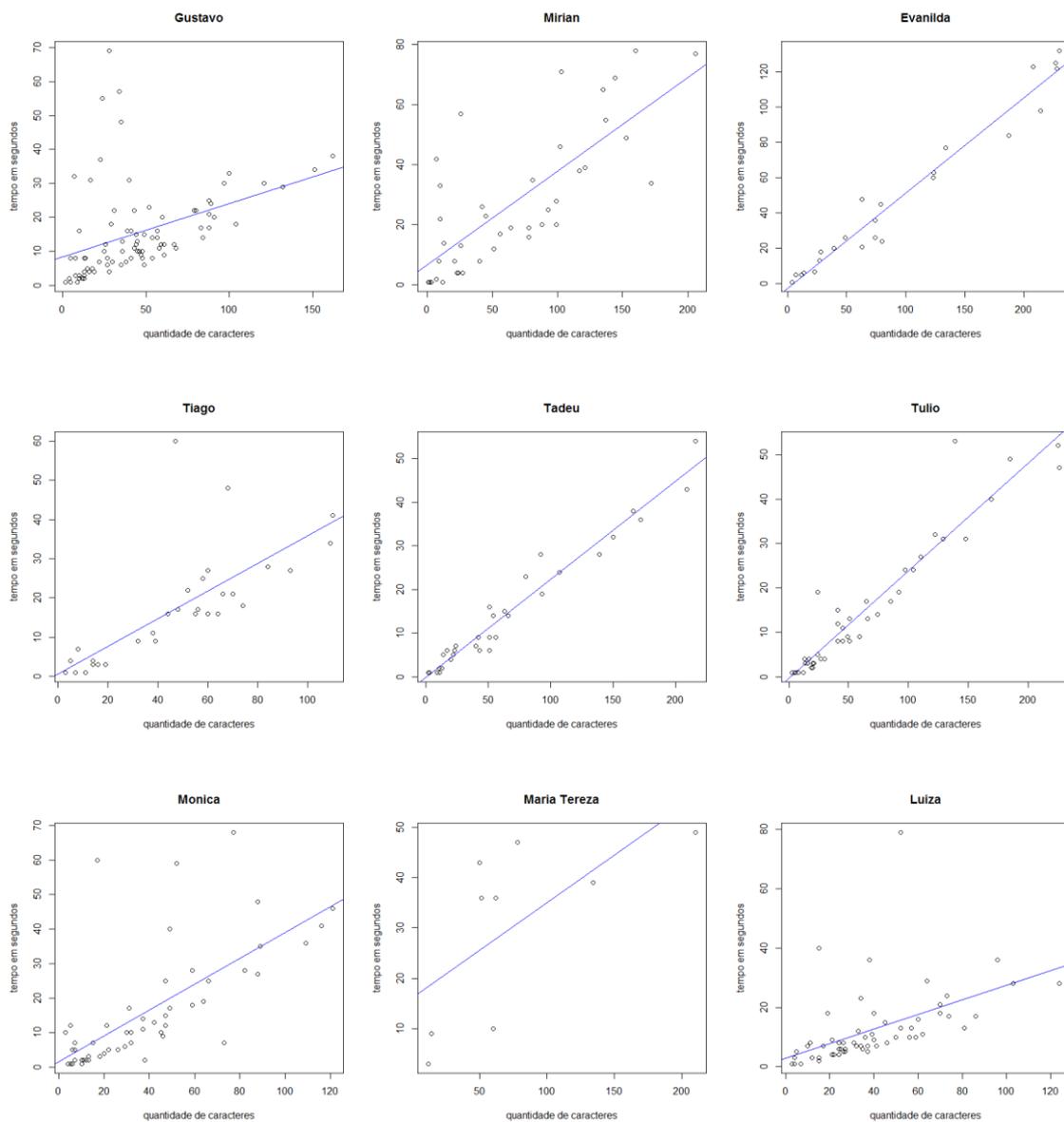


Figura 17 Gráfico de dispersão do tempo de digitação em função da quantidade de caracteres da mensagem

Fonte: O autor, 2019.

Testou-se a distribuição do tempo de digitação para a distribuição normal por meio do teste de Shapiro-Wilk ($W = 0.75989$, $p\text{-value} < 2.2e-16$). O pressuposto de normalidade não foi atendido para o tempo de digitação, então o teste de Pearson não pôde ser utilizado para estabelecer a correlação, mas sim o método de Sperman.

O resultado do teste de correlação pelo método de Sperman ($S = 2829544$, $p\text{-value} < 2.2e-16$) apontou uma correlação forte (0,83). Na Tabela 10, são apresentadas as correlações entre o tempo de digitação e o tamanho da mensagem de cada voluntário da turma DSI 2016.1. Em todos os casos, a classificação para a correlação foi forte, o que

possibilita concluir que a quantidade de caracteres influencia fortemente o tempo de digitação.

Tabela 10. Correlação entre o tempo de escrita e o tamanho da mensagem

Participante	Correlação
Maria Evanilda	0,820
Gustavo	0,864
Janaina	0,822
José Carlos	0,824
Lilian	0,824
Luiza	0,835
Maria Tereza	0,831
Mirian	0,844
Monica	0,833
Tadeu	0,823
Tiago	0,824
Tulio	0,819

Fonte: O autor, 2016.

A partir dos dados de digitação de voluntários da turma DSI2016.1, identificou-se que, em média, aquelas pessoas digitaram a uma taxa de 0,34 segundos por caractere (conforme dados da última coluna da Tabela 9), o que possibilita estabelecer a Equação (4). Essa equação é utilizada nesta pesquisa para estimar o tempo médio de digitação em função da quantidade média de caracteres das mensagens de uma sessão de bate-papo.

$$T_d = 0,34 * \text{número de caracteres da mensagem (segundos)} \quad (4)$$

Dessa forma, uma vez que o tempo de digitação é fortemente correlacionado com a quantidade de caracteres, ambas as Equações (3) e (4) são válidas para estimar o tempo médio de digitação de uma mensagem, considerando-se o número de caracteres da mensagem, sendo que a taxa de digitação depende do perfil médio das pessoas de cada contexto (0,28 foi a taxa encontrada no primeiro estudo e 0,34 foi a taxa encontrada no segundo estudo).

4.3. Reflexão

Durante o bate-papo, o participante não fica o tempo todo lendo e digitando mensagens. Em alguns momentos, o participante pode estar aguardando uma resposta para a mensagem que ele acabou de postar. O participante também pode estar pensando em uma resposta para o debate antes de efetivamente digitar uma mensagem. Ou, ainda, pode estar simplesmente ocioso. Nesta pesquisa, será considerado “tempo de reflexão” o tempo em que o participante não está lendo nem digitando mensagens.

Nesta pesquisa, foi considerada a possibilidade de investigar o tempo ocioso com o uso de um *eyetracking* para medir o tempo em que a pessoa permanece olhando para cada ponto da tela, e assim inferir os momentos em que a pessoa não se encontra nem digitando nem lendo as novas mensagens do bate-papo. Pela ausência desse equipamento e pelo alto custo para adquiri-lo à época em que esta pesquisa foi iniciada, não foram realizados estudos para investigar e caracterizar empiricamente o tempo de reflexão.

Ainda que nesta pesquisa não tenha sido possível realizar estudos empíricos para caracterizar objetivamente o tempo de reflexão, é possível ao menos modelar, pela Teoria de Filas, o tempo que uma pessoa fica aguardando uma resposta (que nesta pesquisa é considerado como parte do tempo de reflexão), como detalhado na próxima subseção.

4.3.1. *Aguardando resposta*

Na Figura 18 é representada a troca de mensagens entre participantes para diferentes tamanhos de grupo. Quando há apenas dois participantes, após uma pessoa enviar uma mensagem, ela provavelmente ficará aguardando uma resposta (representado na Figura por T_{ar} [Tempo Aguardando Resposta]).

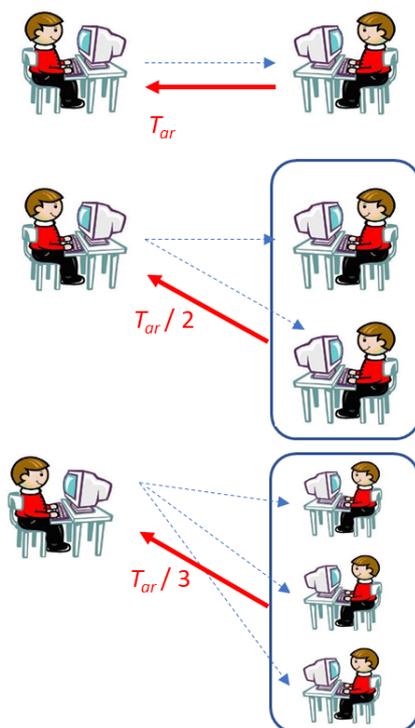


Figura 18. Representação da troca de mensagens em bate-papo

Fonte: O autor, 2019.

Quando há três participantes, após a pessoa enviar uma mensagem, os outros dois participantes irão ler a mensagem, refletir e respondê-la. As mensagens não têm tamanho padronizado; às vezes uma pessoa digita uma mensagem mais longa enquanto a outra digita uma mensagem mais curta. Quanto mais pessoas estiverem conversando no bate-papo, é de se esperar que uma pessoa fique menos tempo aguardando uma resposta. Esse comportamento pode ser modelado pela Teoria de Filas (FOGLIATTI; MATTOS, 2007; KLEINROCK, 1975).

A participação no bate-papo educacional pode ser modelada pela Teoria de Filas (KLEINROCK, 1975) (Figura 19), considerando cada participante como um servidor que precisa ler (servir) todas as mensagens (clientes) digitadas pelos outros participantes, sendo que as mensagens vão sendo enfileiradas na ordem em que vão chegando.

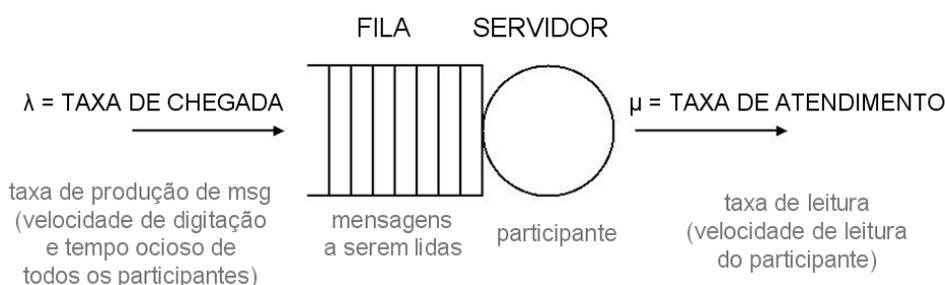


Figura 19 Sistema com uma fila

Fonte: O autor, 2013.

Cada participante, além de ser um servidor, também desempenha o papel de uma fonte geradora de mensagens para as filas dos demais participantes. A taxa de leitura de mensagens μ é o inverso do tempo médio de leitura (T_l), e a taxa de produção de mensagens λ é o inverso do tempo médio de digitação (T_d).

Em um modelo markoviano M/M/1 (que é o modelo ilustrado na Figura 19), os intervalos entre chegadas sucessivas e os tempos de atendimento seguem distribuições exponenciais (FOGLIATTI; MATTOS, 2007, p. 51). Na pesquisa realizada no mestrado (ROCHA; PIMENTEL; DINIZ, 2013, p. 76), foi apresentado um estudo sobre a distribuição de probabilidade da taxa de chegada de mensagens de bate-papo nas turmas TIAE. Com esse estudo, foi possível caracterizar o processo de chegada de mensagens em bate-papo como tendo distribuição de Poisson, portanto o intervalo entre chegadas sucessivas de mensagens é exponencial, o que foi confirmado em 51 das 62 sessões de bate-papo (82%) das turmas de Tecnologia da Informação Aplicada à Informação (TIAE) 2004.2 à 2008.1 pelo teste Kolmogorov-Smirnov⁷ (para teste de hipótese sobre a distribuição do intervalo de chegada de mensagens). Por outro lado, sobre os tempos de atendimento (leitura de mensagens do bate-papo), pelo estudo feito no mestrado, foi possível concluir que o tempo de leitura não é exponencial (sendo a distribuição binomial negativa a que melhor explicou os dados). Contudo, segundo Fogliatti e Mattos (2007, p. 219), “cabe ressaltar que, sempre que possível, deve ser escolhida a distribuição exponencial mesmo que não seja o melhor ajuste, pois a utilização em muito facilita a obtenção das medidas de desempenho do sistema”.

A partir dos estudos empíricos realizados anteriormente e das considerações encontradas na literatura, nesta pesquisa foi assumido que a taxa de leitura de mensagens μ e a taxa de produção de mensagens λ de uma sessão de bate-papo educacional podem ser caracterizadas como uma distribuição exponencial e, portanto, os processos de chegada e produção de mensagens podem ser modelados pelo modelo M/M/1.

Ainda pela Teoria de Filas, conforme esquematizado na Figura 20, a taxa de chegada λ na fila de leitura do participante é igual ao somatório da taxa de produção de mensagens dos $(p - 1)$ participantes.

⁷o teste Kolmogorov-Smirnov é usado para determinar se duas distribuições de probabilidade subjacentes diferem uma da outra ou se uma das distribuições de probabilidade subjacentes difere da distribuição em hipótese, em qualquer dos casos com base em amostras finitas. No R, a função `ks.test` (do pacote `stats`) faz o teste de Kolmogorov Smirnov. < <http://stat.ethz.ch/R-manual/R-patched/library/stats/html/ks.test.html> >

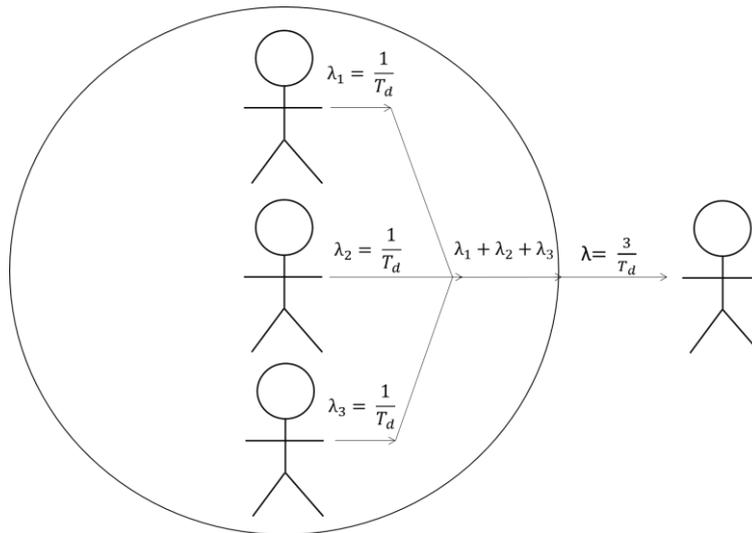


Figura 20 Representação do tempo de resposta
 Fonte: O autor, 2019.

Se um participante demora T_d para digitar uma mensagem, então esse participante produz mensagens a uma taxa de $1/T_d$. Com um grupo de 4 participantes, um participante receberá mensagens dos 3 outros participantes a uma taxa de chegada igual ao somatório das taxas de produção de mensagens dos 3 participantes, que é $3/T_d$. Como o tempo médio que o participante fica aguardando uma mensagem é o inverso da taxa de chegada de mensagens ($3/T_d$), então o tempo médio é $T_d/3$ (conforme esquematizado na Figura 18). Se o grupo fosse de 5 participantes, uma pessoa ficaria aguardando uma resposta, em média, durante $4/T_d$ segundos. Quanto mais participantes houver na sessão de bate-papo, menos o participante ficará aguardando uma resposta, o que é modelado na Equação (5).

$$T_{ar} = \frac{1}{\lambda} = \frac{T_d}{p - 1} \quad (5)$$

4.3.2. Tempo Total de Reflexão

O Tempo de Reflexão T_R inclui o Tempo Aguardando Resposta T_{ar} (modelado na subseção anterior) e outros tempos, como o tempo em que fica pensando em uma resposta, pensando sobre o que acabou de ler, revendo as mensagens anteriores, ou mesmo o tempo em que fica ocioso ou distraído com outras coisas. Esses outros tempos não serão modelados na presente pesquisa dada a ausência de estudos empíricos disponíveis.

Por simplificação, o Tempo de Reflexão T_R é caracterizado, nesta pesquisa, como sendo todo o tempo que sobra de uma sessão de bate-papo, desconsiderando todo o tempo em que o usuário fica digitando suas mensagens (T_D) e todo o tempo em que o usuário fica lendo as mensagens dos outros (T_L). Dessa forma, o tempo total de reflexão é caracterizado pela Equação (6):

$$T_R = \text{duração da sessão} - (T_L + T_D) \quad (6)$$

5. Relação entre a quantidade de participantes e a produção de mensagens

O objetivo deste capítulo é demonstrar as modelagens da participação em bate-papo educacional por meio das atividades elementares apresentadas no capítulo anterior e por meio das variáveis identificadas na pesquisa. Durante toda a pesquisa desenvolvida no mestrado e agora com uma outra abordagem no doutorado, foram identificados alguns cenários de aplicação para a modelagem da participação.

Na Seção 5.1, é apresentada a relação entre o tamanho do grupo e a produção de mensagens pelos participantes. Na Seção 5.2, um cenário mais simples onde o participante é modelado por apenas dois estados: lendo e digitando. Na Seção 5.3, é apresentado um estudo sobre o percentual de ocupação dos participantes baseado no modelo de dois estados. Na Seção 5.4, um novo estado é introduzido no modelo, o que gera um segundo cenário de aplicação. Na Seção 5.5, um terceiro cenário, com um diagrama contendo quatro estados em que um participante pode se encontrar em uma sessão de bate-papo, é apresentado. Esse último cenário foi estudado apenas na pesquisa de mestrado, pois conforme é apresentado, só existe diferença nas estimativas produzidas entre o modelo de três estados para pequenos grupos de até cinco participantes.

5.1. Relação estatística entre a quantidade de participantes e a produção de mensagens

A pesquisa iniciou com um modelo conceitual, apresentado no capítulo 1, que parte do princípio de que existe uma relação entre a variável independente “quantidade de participantes” (p) e a variável explicativa “quantidade de mensagens produzidas por cada participante” (m). Conjecturou-se que, para grupos pequenos (com duas, três ou poucas pessoas), os participantes podem ficar muito tempo aguardando uma resposta e, dessa forma, a produção de mensagens aumenta conforme o tamanho do grupo aumenta. Supõe-se, ainda, que a produção individual de mensagens alcança um ápice e depois decai continuamente conforme aumenta o tamanho do grupo. Portanto, para grupos não

pequenos, a hipótese é de que existe uma relação inversamente proporcional entre as duas variáveis: conforme aumenta p , diminui m .

A relação entre as duas variáveis pode ser visualmente percebida nos gráficos com os dados reais das turmas de Tecnologia da Informação Aplicada à Informação (TIAE) e Docência em Sistemas de Informação (DSI) (Figura 21).

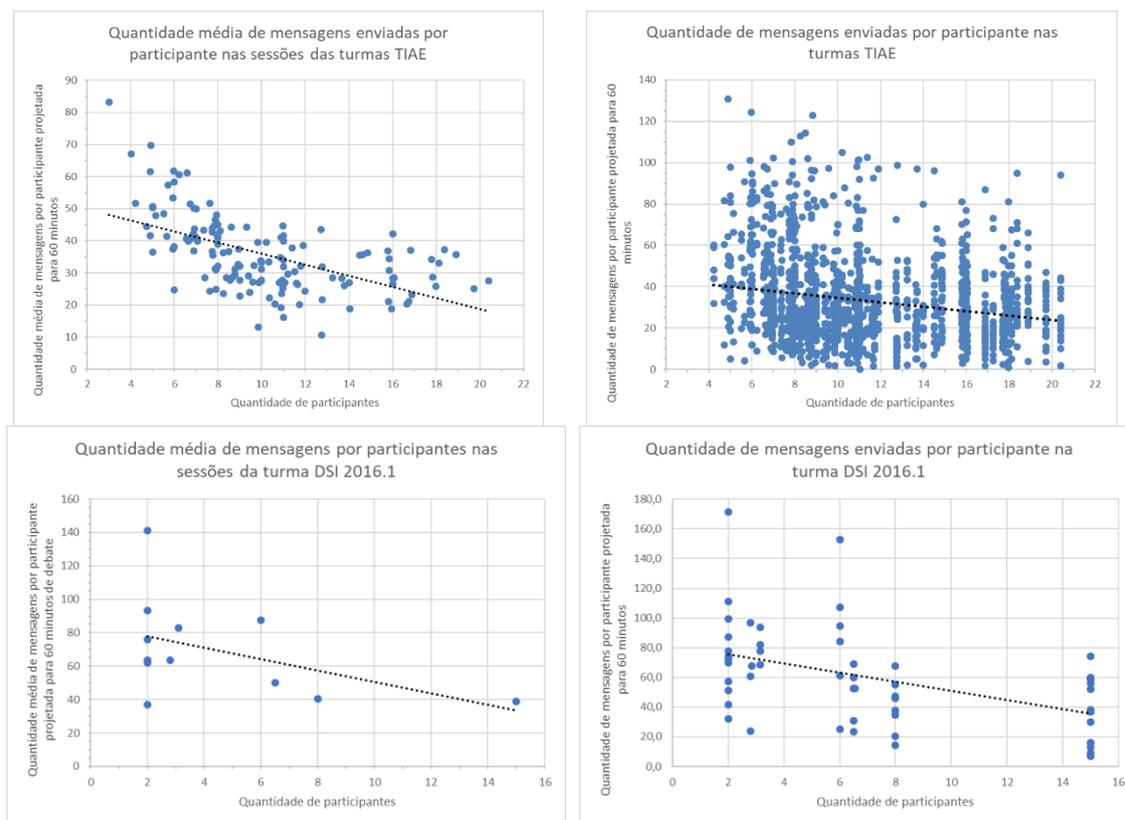


Figura 21 Gráfico de dispersão das turmas TIAE e DSI

Fonte: O autor, 2019.

Além da análise gráfica, para se comprovar a relação entre as duas variáveis de interesse da pesquisa, foi realizado o teste de hipótese:

H0: Não existe relação entre as variáveis p e m (hipótese nula)

H1: Existe relação entre as variáveis p e m

Para o teste de hipótese, foram usados os dados das turmas TIAE. Com os dados desse *corpus* agrupados por sessão de bate-papo (considerando a média da quantidade de mensagens produzidas pelos participantes daquela sessão, quando a sessão é projetada para durar uma hora), não foi possível rejeitar a hipótese nula.

Foi formada então uma base de dados composta pelos microdados, que são os dados de produção de mensagens individualizados por participante em cada uma das

sessões de bate-papo. Ou seja, o dado deixa de ser a produção total de mensagens na sessão de bate-papo para ser a produção de mensagens de um participante na sessão.

Foi utilizado o *software* R e a função *gam* para o teste de hipótese. A função *gam* se refere ao uso de um modelo aditivo generalizado (*generalized additive model*).

Um modelo aditivo generalizado (HASTIE, 2017) é um modelo linear com um preditor linear envolvendo a soma de funções suavizadas das covariáveis. Para parametrizar o modelo, foram utilizadas as variáveis: mensagens e num_partic (número de participantes) e as covariáveis moderador, duracao_sessao, tempo_part, além da família Poisson e da função de ligação logarítmica. A covariável moderador é um marcador que informa se o participante foi o moderador daquela sessão, pois essa figura costuma enviar mais mensagens que os demais participantes. Outras duas covariáveis que também influenciam na produção de mensagens são a duração da sessão de bate-papo e o tempo que o participante permaneceu na sessão de bate-papo. No

Quadro 1, é apresentado o resultado do teste de hipótese:

Quadro 1 Teste de hipótese para relação entre as variáveis de interesse

```
gam(formula = mensagens ~ num_partic + moderador + duracao_sessao +
tempo_part, family = poisson(link = "log"), data = MergedDataset)

Anova for Parametric Effects
      Df Sum Sq Mean Sq F value    Pr(>F)
num_partic      1    821.1    821.1  93.930 < 2.2e-16 ***
moderador       1   4628.0   4628.0 529.399 < 2.2e-16 ***
duracao_sessao  1    259.7    259.7  29.704 5.894e-08 ***
tempo_part      1    855.7    855.7  97.883 < 2.2e-16 ***
Residuals     1470 12850.8      8.7
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> coefficients(GAM1)
(Intercept)      num_partic  moderadorsim duracao_sessao tempo_part
 0.71592677   -0.02398191    0.77011514    0.01674612    1.96741454
```

Fonte: O autor, 2019.

Com relação à variável mensagens, a variável num_partic apresenta um valor de Pr próximo a zero (< 2.2e-16), o que rejeita a hipótese nula de não haver relação entre as variáveis. Conclui-se que existe relação entre a produção de mensagens e a quantidade de participantes com intervalo de confiança superior a 99%.

Com relação às demais covariáveis, também é possível observar que todas influenciam na produção de mensagens, com a diferença de que o número de participantes influencia inversamente a produção de mensagens (-0.02398191) conforme apresentado no Quadro 1.

5.2. Produção máxima de mensagens: lê e digita o tempo todo

Como um primeiro estudo, nesta pesquisa buscou-se modelar o comportamento dos participantes numa situação hipotética em que eles ficariam o tempo todo lendo e digitando mensagens, sem refletir momento algum. Com este cenário, é possível obter valores limites da produção de mensagens do participante e do grupo; sabe-se que os participantes não conseguem produzir mais mensagens do que os valores obtidos com este modelo, dada a capacidade de leitura e de digitação dos participantes.

5.2.1. Modelo analítico da produção máxima de mensagens

Nesta pesquisa, considera-se que, no contexto de bate-papo educacional, um participante lê todas as mensagens dos demais e digita mensagens de tempos em tempos. Esse ciclo de digitação e leitura de mensagens é repetido, pelo participante, durante todo o tempo da sessão de bate-papo (T_s). Portanto, considera-se que cada participante, ao longo de todo o tempo da sessão (T_s), passa uma parte desse tempo lendo as mensagens (T_{Lendo}) e outra parte digitando mensagens ($T_{Digitando}$). Nesse caso, pode-se representar o tempo total da sessão T_s por meio da Equação (7):

$$T_s = T_{Digitando} + T_{Lendo} \quad (7)$$

Sabe-se que, durante uma sessão de bate-papo, são enviadas mensagens com diferentes quantidades de caracteres. Contudo, para desenvolver a Equação (7), é assumida uma “mensagem típica” contendo a média de caracteres de todas as mensagens emitidas na sessão de bate-papo. Deseja-se calcular M (variável de interesse que se deseja modelar), que é a quantidade máxima de mensagens que cada participante produz em média durante a sessão de bate-papo.

Sabe-se que cada participante digita mensagens de uma maneira particular, conforme discutido na Seção 4.2. Contudo, para modelar o tempo total que o participante

passa digitando mensagens ($T_{Digitando}$), por simplificação, nesta modelagem é adotado um T_d , que representa o tempo médio de que cada participante do grupo precisa para digitar uma “mensagem típica” daquela sessão. Com essa simplificação, é possível modelar o tempo total em que o participante permanece digitando na sessão de bate-papo (Equação (8)): é a quantidade M de mensagens escritas pelo participante ao longo da sessão multiplicada pelo tempo médio de digitação de uma mensagem típica, T_d .

$$T_{Digitando} = MT_d \quad (8)$$

Além de digitar mensagens, o participante também lê as mensagens enviadas por outros participantes do grupo de bate-papo. Considerando que cada participante digita o máximo de M mensagens, e considerando que o grupo é formado por p participantes, então a quantidade total de mensagens a serem lidas por cada participante é Mp . É preciso ressaltar que, nesta pesquisa, é assumido que o participante lê a própria mensagem que envia; se fosse considerado que o participante não lê as próprias mensagens que envia, então a quantidade total de mensagens a serem lidas por cada participante seria $M(p-1)$; contudo, foi identificado que ler ou não ler a própria mensagem enviada não influencia muito os resultados encontrados, e assim o fator $(p-1)$ criaria uma complexidade desnecessária nas fórmulas a seguir.

Sabe-se, também, que o tempo de leitura de uma mensagem é diferente para cada participante e é fortemente correlacionado com o tamanho da mensagem, conforme discutido na Seção 0. Por simplificação, é adotado um T_l , que representa o tempo médio de leitura de uma “mensagem típica” pelos participantes do grupo. Assim, o tempo médio que cada participante permanece lendo mensagens durante o bate-papo é dado pela Equação (9): a quantidade M de mensagens produzidas por cada participante, multiplicada pelos p participantes da sessão (já que cada participante produz M mensagens), multiplicada pelo tempo médio de leitura de uma mensagem típica, T_l .

$$T_{Lendo} = Mp T_l \quad (9)$$

Substituindo os termos das Equações (8) e (9) na Equação (7), e isolando a variável M de interesse desta pesquisa (a média de mensagens produzidas por cada participante durante a sessão de bate-papo), chega-se à Equação (10).

$$T_s = T_{Digitando} + T_{Lendo}$$

$$T_s = MT_d + MpT_l$$

$$T_s = M(T_d + pT_l)$$

$$M = \frac{T_s}{T_d + pT_l} \quad (10)$$

Esta equação estabelece a relação da quantidade máxima de mensagens M produzida por uma pessoa em função do tamanho p do grupo de participantes do bate-papo.

Pode-se exemplificar a Equação (10) com os parâmetros obtidos da turma TIAE 2007.1 (FUKS; PIMENTEL; LUCENA, 2006). Nessa turma, as mensagens produzidas apresentaram uma média de 94,1 caracteres (tamanho da “mensagem típica”). Utilizando a fórmula da Equação (1) apresentada na Seção 0, calcula-se o $T_l = 1 + 0,04 * 94,1 = 4,764s = 0,08min$. Pela Equação (3) apresentada na Seção 4.2, calcula-se o $T_d = 0,28s * 94,1 = 26,348s = 0,44min$. Considerando o tempo total da sessão de 60 minutos ($T_s = 60$ minutos), da Equação (10) resulta: $M = 60 / (0,44 + p * 0,08)$. Esta equação encontra-se traçada na Figura 22, considerando o tamanho p do grupo entre [3, 40].

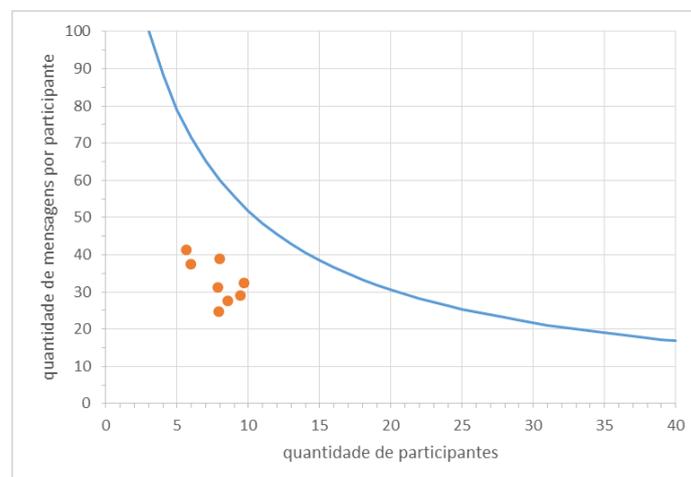


Figura 22. Curva M (máximo de mensagens em função do tamanho p do grupo)

Fonte: O autor, 2019.

A curva M representa a produção de mensagens que um grupo de tamanho p poderia produzir se ficasse o tempo todo lendo e digitando mensagens, sem nunca ficar ocioso ou refletindo. Essa curva modela o limite máximo de produção de mensagens em um bate-papo educacional; dada a velocidade de leitura e de digitação dos participantes, sabe-se que o participante não seria capaz de produzir mais mensagens do que o limite apresentado por essa curva. Os oito pontos abaixo da curva representam a quantidade de mensagens produzidas por participante em cada uma das oito sessões de bate-papo realizadas na turma TIAE 2007.1. Como nem toda sessão durou exatamente 60 minutos, a quantidade de mensagens apresentada na Figura 22 foi calculada considerando a quantidade de mensagens realmente produzidas na sessão multiplicando-a por $60/\text{duração_sessão}$ (ajuste por regra de três). Observa-se que as quantidades de mensagens realmente produzidas pelos participantes da turma 2007.1 (projetadas como se a sessão tivesse durado 60 min) ficaram abaixo da curva M de produção máxima de mensagens, como era esperado.

Considerando a Equação (10), é possível equacionar a produção máxima de mensagens que o grupo produz, M_G (máximo de mensagens do grupo). Esta equação é modelada multiplicando M (a produção máxima de mensagens de um participante) pelo tamanho p do grupo:

$$M_G = Mp$$

Substituindo M pela Equação (10), resulta na Equação (11) a seguir:

$$M_G = \frac{T_s}{T_d + pT_l} p \quad (11)$$

Observa-se que, quando p tende a infinito, a produção máxima de mensagens do grupo (M_G) tende a um valor constante, aqui rotulado como H, calculado pelo limite:

$$\lim_{p \rightarrow \infty} f_{M_G}(p) = \lim_{p \rightarrow \infty} \left(\frac{T_s}{T_d + pT_l} p \right) = \frac{T_s}{T_l} = H \quad (12)$$

Por exemplo, considerando os dados obtidos da turma TIAE 2007.1 ($T_s = 60$ minutos; $T_d = 0,44$ minutos; $T_l = 0,08$ minutos), $H = T_s / T_l = 60 / 0,08 = 750$ mensagens. Esse valor significa que, quando a quantidade de participantes tende ao infinito, o grupo consegue produzir, no máximo, 750 mensagens por hora; esse limite é calculado como se cada participante só ficasse lendo mensagens o tempo todo da sessão ($H = T_s / T_l$).

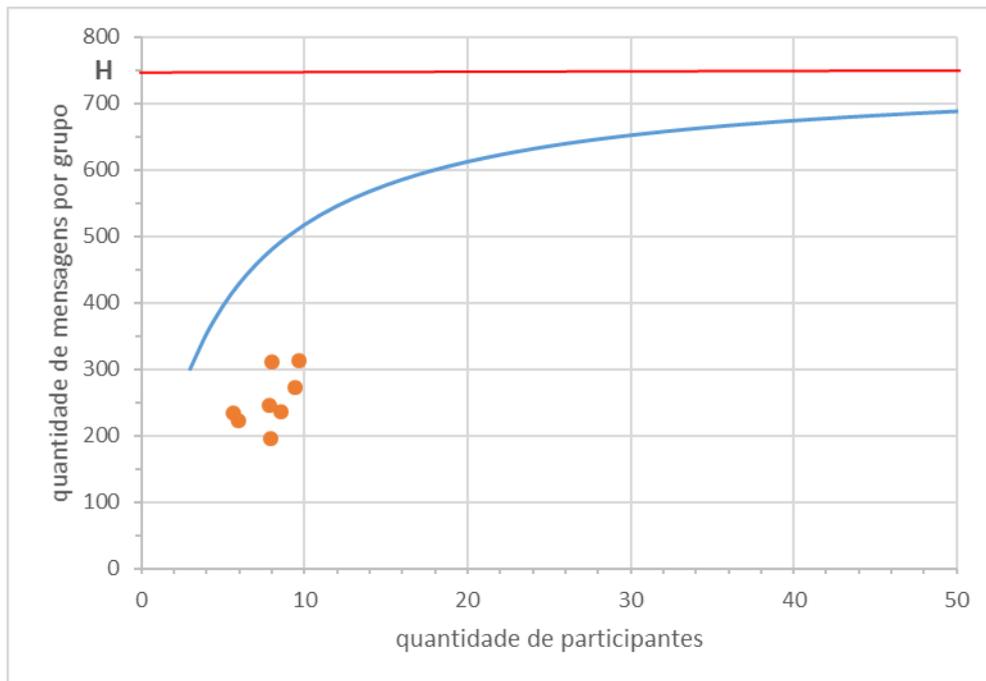


Figura 23. Curva M_G (máximo de mensagens do grupo), o limite H quando o número de participantes p tende a infinito, e a quantidade real total de mensagens produzidas pelo grupo TIAE 2007.1

Fonte: O autor, 2019.

Na Figura 23, é traçada a curva M_G (máximo de mensagens de grupo) da turma TIAE 2007.1 e a constante H (limite máximo da quantidade de mensagens produzidas pelo grupo por hora). Também é apresentada a quantidade de mensagens produzidas pelo grupo em cada sessão (projetadas como se cada sessão tivesse durado exatamente 60 minutos).

5.2.2. Modelo estocástico da produção máxima de mensagens (M)

A produção máxima de mensagens já havia sido investigada nesta pesquisa durante o mestrado por meio de um modelo estocástico. É relativamente mais complexo obter

participante a leia somente após terminar de digitar a mensagem que está elaborando. Nesse caso, após enviar a mensagem elaborada, passa para o estado Lendo e permanece nesse estado enquanto houver alguma mensagem a ser lida. Após ter lido todas as mensagens que chegaram, o participante volta para o estado Digitando. O indivíduo permanece nesse ciclo escrita-leitura até o final da sessão de bate-papo.

A transição de estados é modelada em função de taxas. O participante permanece lendo uma determinada mensagem por um tempo médio (T_l), e a taxa de leitura é calculada como o inverso do tempo médio. Da mesma forma, o participante permanece no estado Digitando durante um tempo médio de digitação (T_d), a partir do qual se calcula a taxa de escrita.

Para tornar possível a obtenção da medida de interesse M (produção máxima de mensagens por participante por hora de sessão de bate-papo) em função do tamanho do grupo, o modelo representado no Diagrama Lendo-Digitando (Figura 24) foi implementado no ambiente Tangram-II, que é um ambiente de modelagem para fins de pesquisa e ensino (SILVA; FIGUEIREDO; LEÃO, 2009), cuja interface é apresentada na Figura 25.

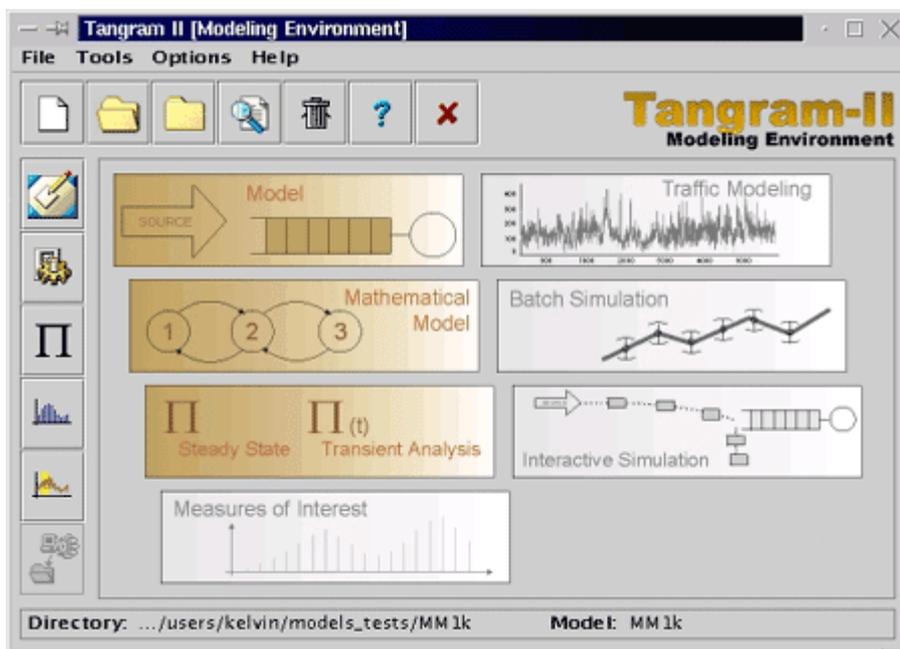


Figura 25 Interface do ambiente de modelagem Tangram-II

Fonte: Silva, 2009.

O ambiente de modelagem Tangram-II é organizado em módulos, em que cada módulo apoia uma atividade específica do processo de modelagem: *Model Specification*,

Mathematical Model, Analytical Model Solution, Measures of Interest, Traffic Descriptors, Simulation e Hidden Markov Models.

Pelo módulo *Model Specification*, ilustrado na Figura 26, representa-se o modelo por meio de objetos que interagem enviando e recebendo mensagens, e o estado de cada objeto é representado por um conjunto de variáveis de valores inteiros.

```

name=Participante

Events=
event=Ler (EXP, taxaLeitura
condition= (Estado == 1)
action= { int q,s;
s = 1;
q = Fila - 1;
if(q == 0)
s = 0;
set_st("Fila",
set_st("Estado"

Watches=
Declaration=
Var:
/* Estado = 0 (OciosoMotivadoParaEscrever)
Estado = 1 (Lendo)
Estado = 2 (Escrevendo)
Estado = 3 (OciosoDesmotivadoParaEscrever)*/
Const
Integer: tamanhoMaximoFila;
Port : portaEntrada;
EFloat : taxaEscrita, taxaLeitura, taxaIniciativaParaEscrever;

Initialization=
/* Constants Initialization */
event=Escrever (EXP, taxaEs
condition= (Estado == 2)
action= { int s;
s = 3;
if(Fila > 0)
s = 1;
set_st("Estado"

Envio)
Fila = 0;
Estado = 0;
tamanhoMaximoFila = 100
portaEntrada = wire
taxaEscrita = 2.2
taxaLeitura = 12.9
taxaIniciativaParaEscrever = 1.2

Messages=
msg_rec=portaEntrada
action= { int q,s;
q = Fila;
s = Estado;
if (Fila < tamanhoMaximoFila)
q = Fila + 1;

Rewards=
impulse_reward=mensagens:ei
event=Escrever, 1
value=1;
impulse_reward=mensagens:l:
event=Ler, 1
value=1;

```

Figura 26 Interface de especificação do modelo no Tangram-II

Fonte: Silva, 2009.

Cada objeto possui cinco áreas: *Declaration*, *Initialization*, *Events*, *Messages* e *Rewards*. Em *Declaration*, são listadas as variáveis e constantes do objeto, usadas para armazenar, por exemplo, o estado atual do objeto ou a quantidade de mensagens na fila. Em *Initialization* são atribuídos os valores das constantes e o valor inicial das variáveis. As ações realizadas pelos objetos são codificadas na forma de eventos na área *Events*. Esses eventos possuem uma assinatura composta pelo nome do evento, distribuição de probabilidade e taxa com que ocorrem. A área *Messages* é usada para criar mensagens de comunicação entre os objetos. Nas mensagens, são informados o nome da porta em que a mensagem é recebida e a ação tomada no recebimento da mensagem. Por fim, a área *Reward* é usada para implementar recompensas. As recompensas são relacionadas com a ocorrência de eventos e servem, entre outras coisas, para quantificar a ocorrência desses eventos.

Para implementar o Diagrama Lendo-Escrevendo no Tangram-II, foi definido um objeto para representar individualmente cada participante de uma sessão de bate-papo, conforme ilustrado na Figura 27.

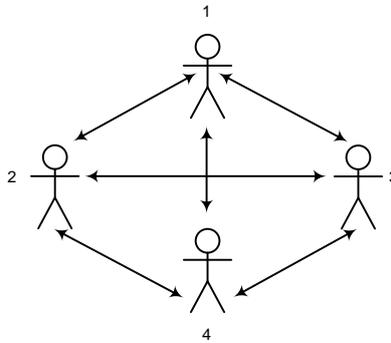


Figura 27. Estratégia de modelagem: interação entre indivíduos
Fonte: O autor, 2013.

Nessa estratégia de modelagem, cada objeto participante tem os estados definidos na Figura 24: lê e digita mensagens. Cada participante também é caracterizado por suas taxas de leitura e escrita de mensagens. Os participantes enviam mensagens uns para os outros e também leem as mensagens enviadas de acordo com suas respectivas taxas. As mensagens recebidas são enfileiradas. Cada participante é representado por duas variáveis: a fila de mensagens, que contém valor inteiro; e o estado de participação, que contém um dos dois valores: Lendo e Digitando.

Após implementar esse modelo por meio dos objetos do Tangram-II, é preciso gerar os estados da cadeia de Markov, o que é feito no módulo *Mathematical Model* (Figura 28).

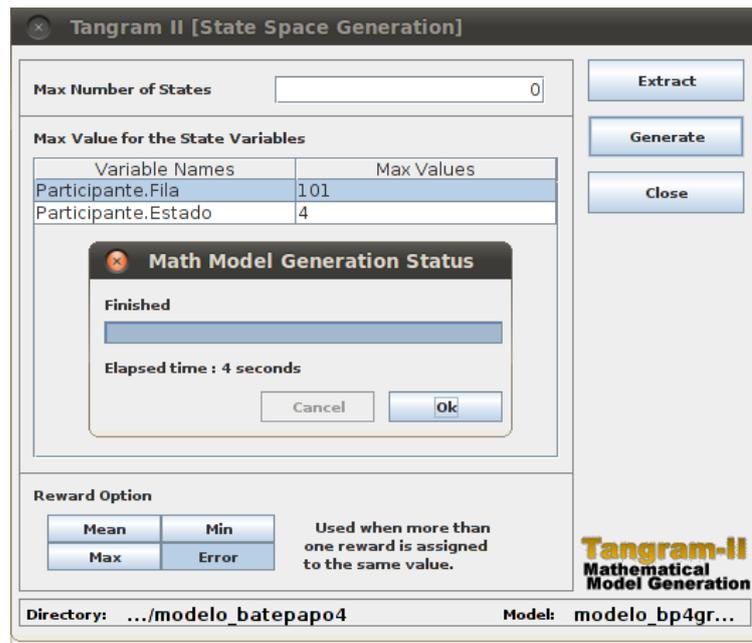


Figura 28 Módulo *Mathematical Model*

Fonte: Silva, 2009.

Após a geração dos estados da cadeia de Markov, são extraídas as medidas de interesse pelo módulo *Analytical Model Solution* (Figura 29). Nesse módulo, são especificados os valores dos parâmetros e é escolhida a técnica de solução apropriada para o modelo. Para solução do modelo e posterior obtenção das medidas de interesse, na presente pesquisa foram usadas duas técnicas de análise transiente: *Expected Cumulative Rate Reward* e *Expected Cumulative Impulse Reward*.

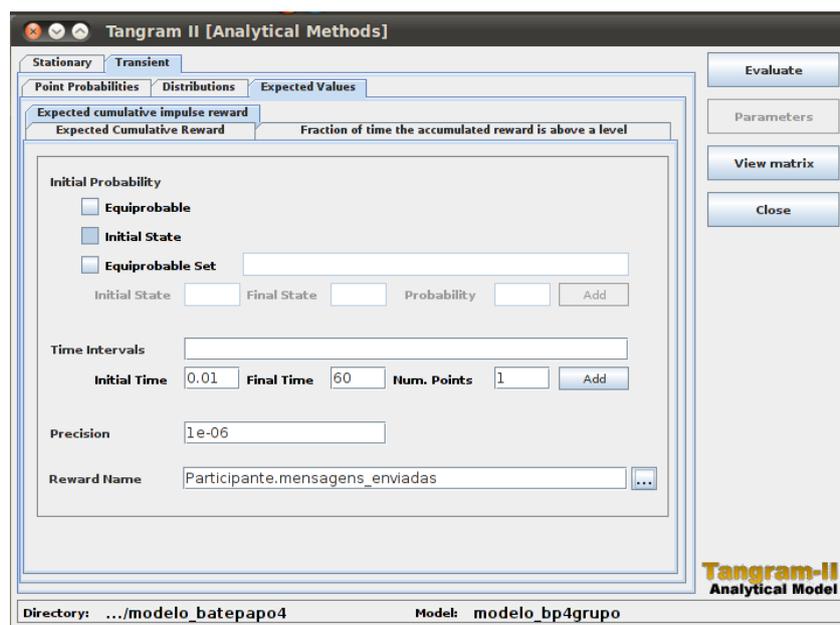


Figura 29 Interface do Módulo de Solução Analítica do Tangram-II

Fonte: Silva, 2009.

Na figura do módulo *Analytical Model Solution*, o tipo de análise escolhido é a análise transiente com a técnica de recompensa acumulada de impulso. A probabilidade inicial é selecionada como estado inicial. Para o tempo inicial é atribuído o valor 0.01, pois não é aceito valor zero; e para o tempo final, o valor 60 (duração do bate-papo igual a 60 minutos). A precisão usada para todas as soluções foi 1 e -06.

Na Figura 30, são apresentados os dados reais da turma TIAE 2007.1 e as estimativas obtidas pelo modelo de produção máxima do Tangram-II. Nesta turma, obteve-se um T_l igual a 0,08 minutos e um T_d igual a 0,44 minutos.

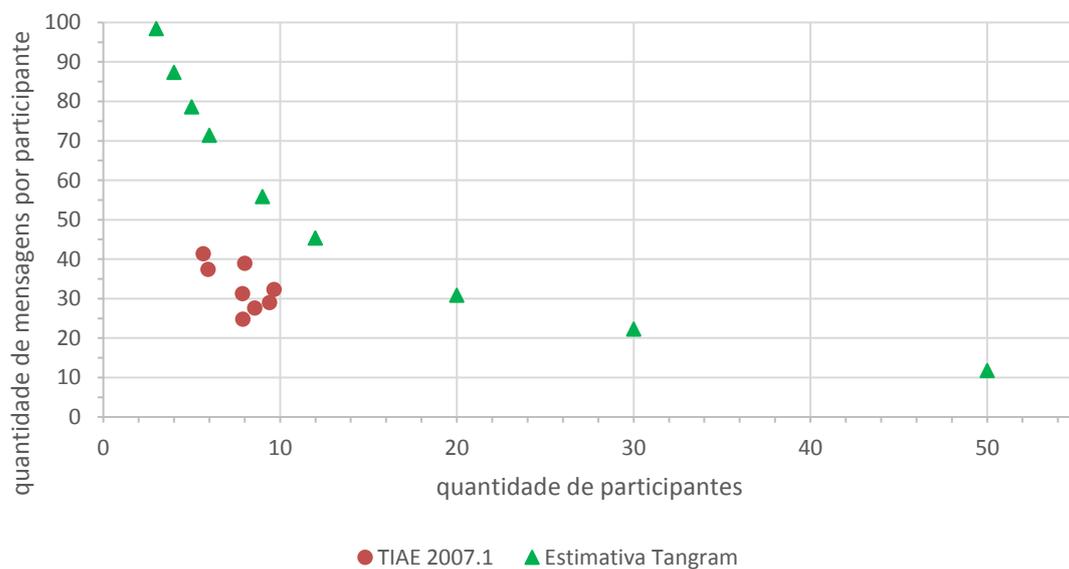


Figura 30 Estimativas produzidas pelo Tangram-II para a produção máxima de mensagens no contexto da Turma TIAE 2007.1

Fonte: O autor, 2019.

Da mesma forma que o modelo analítico apresentado na subseção anterior, todos os dados reais observados nas sessões de bate-papo da turma TIAE 2007.1 ficam abaixo das estimativas produzidas pelo modelo estocástico de produção máxima de mensagens.

5.2.3. Comparação entre os modelos de produção máxima de mensagens

O modelo estocástico, apresentado na subseção anterior, torna complexa e trabalhosa a produção de estimativas (em comparação com a Equação (10)), pois requer

saber usar o Tangram-II, modelar um participante, acrescentar cópias do objeto “participante” até alcançar a quantidade de pessoas do grupo, rodar o *Solver* para obter uma estimativa, e repetir esse processo para cada estimativa desejada visando a traçar uma curva como a ilustrada na Figura 30. Foi por causa dessas dificuldades que no doutorado buscou-se desenvolver um modelo analítico (Equação (10)), que produz estimativas de maneira bem simples.

Dadas as duas abordagens (modelo estocástico e modelo analítico), é possível comparar as estimativas produzidas por ambos os modelos, como apresentado na Figura 31.

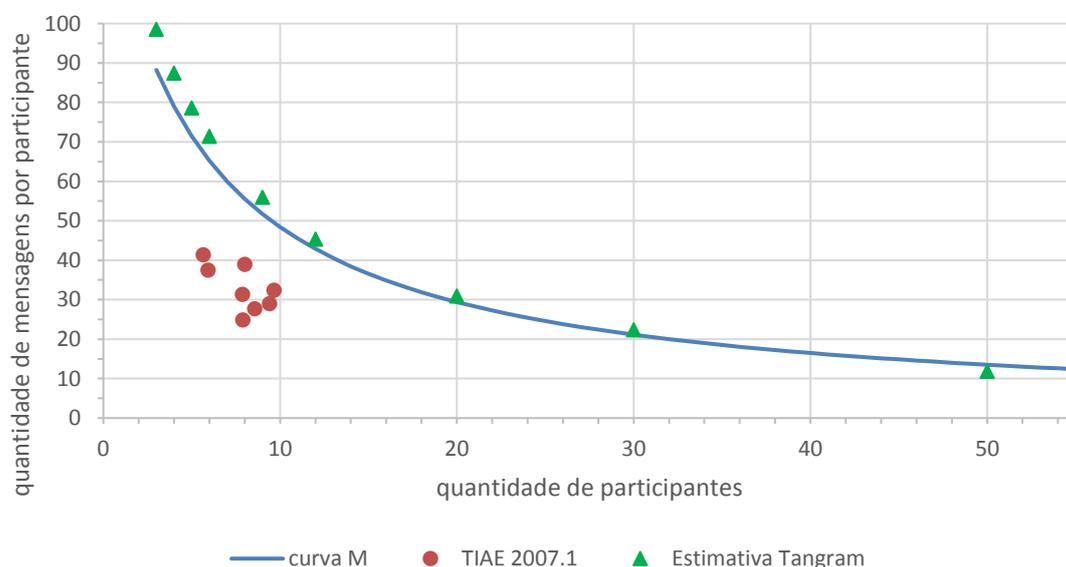


Figura 31 Curva M: comparação das estimativas produzidas a partir da Equação (10) e as obtidas pelo Tangram-II no contexto da turma TIAE 2007.1

Fonte: O autor, 2019.

Conforme pode ser observado na Figura 31, a curva *M* produzida pela equação e os pontos produzidos pelo Tangram-II são bastante próximos, o que indica que a Equação (10) pode ser utilizada no lugar do modelo estocástico para a produção de estimativas sobre a produção máxima de mensagens. A proximidade dos resultados também é um indicativo de que ambas as abordagens parecem adequadas (seria um problema se as abordagens resultassem em estimativas muito distintas).

5.3. Modelo em função do percentual de ocupação dos participantes (PO)

A produção máxima de mensagens, modelada na seção anterior, estabelece uma curva que decai conforme aumenta o tamanho do grupo. Essa é a tendência esperada com relação ao fenômeno real, contudo a curva estimada de produção máxima de mensagens passa muito acima dos dados reais observados. Isso é razoável porque os participantes não ficam o tempo todo lendo-e-digitação mensagens; os participantes também realizam outras ações durante parte do tempo da sessão de bate-papo: refletem, aguardam uma resposta, releem as mensagens já publicadas, dentre outras ações. É razoável supor, então, que os participantes ficam uma parte do tempo ocupados com as ações de leitura-e-digitação de mensagens, e outra parte do tempo ocupado com outras coisas. Esse é o raciocínio empregado para elaborar o modelo Percentual de Ocupação (PO) apresentado nesta seção.

Em Teoria de Filas, chama-se taxa de utilização (ρ) a relação entre o ritmo médio de chegada (λ) e o ritmo médio de atendimento (μ) (PRADO, 2009, p. 67):

$$\rho = \frac{\lambda}{\mu}$$

Para um sistema estável, é preciso que λ seja menor que μ , pois o ritmo de atendimento precisa ser maior que o ritmo de chegada, o que resulta em $\rho < 1$. Quando ρ tende a 1 (a taxa de utilização do sistema tende a 100%), a fila de espera tende a aumentar infinitamente. Na literatura, encontra-se a informação de que computadores do tipo *mainframe* são instalados em empresas para trabalhar a uma taxa de utilização abaixo de 60% para atender adequadamente à demanda de processamento. Esses equipamentos são substituídos quando a taxa de utilização chega a 90%, pois o tempo de resposta (e o tamanho da fila de espera) geralmente já não é adequado (PRADO, 2009, p. 68).

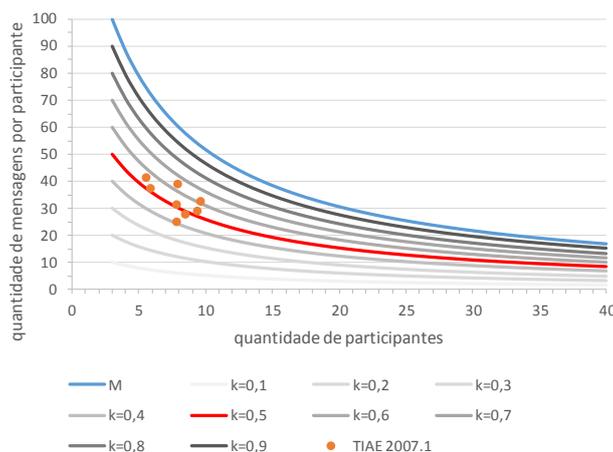
Embora a referência ao uso de *mainframe* esteja relacionada à limitação tecnológica (enquanto a presente pesquisa lida com limitações humanas relacionadas à leitura e digitação), esse raciocínio da taxa de utilização é interessante para pensar o modelo de produção de mensagens em função do tamanho do grupo. A curva M (Equação

(10)), que representa a produção máxima de mensagens por participante, equivale à utilização de 100% do sistema. Contudo, esse não é um estado “aceitável”. Deve ser razoável supor que um participante do bate-papo educacional produza mensagens a um ritmo confortável, de tal maneira que fique ocupado a um percentual k de seu potencial de produção máxima modelado pela curva M .

Analisando a Figura 22, há um determinado PO k que, ao multiplicar a curva M , passa pelo centro dos pontos reais observados naquela turma. Deve-se encontrar um k de tal maneira que o erro entre a curva kM e os pontos reais seja o menor possível. Uma forma para medir o erro entre um conjunto de pontos e uma curva é o erro quadrático médio (EQM). Em estatística, o EQM de um estimador mede a média dos quadrados dos erros ou desvios, isto é, a diferença entre o estimador e o que é estimado. O EQM serve para medir a qualidade de um estimador. Dado \hat{Y} um vetor com valores estimados, e Y um vetor de valores observados (dados reais), o EQM é calculado da seguinte maneira:

$$EQM = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

A partir dessa métrica foi realizado um estudo com os dados da turma TIAE 2007.1 em comparação com a curva kM , variando-se a constante k , conforme apresentado na Figura 32.



	A	B	C	D			
p	TIAE 2007.1	M(k=0,52)	M(k=0,53)	M(k=0,54)	(A-B) ²	(A-C) ²	(A-D) ²
5,6	41,4	38,6	39,4	40,1	7,8	4,2	1,7
5,9	37,4	37,5	38,2	38,9	0,0	0,7	2,4
7,9	31,3	31,5	32,1	32,7	0,0	0,6	1,9

7,9	24,8	31,5	32,1	32,7	44,2	52,7	61,8
8	38,9	31,2	31,8	32,4	59,3	50,4	42,2
8,6	27,6	29,8	30,3	30,9	4,7	7,5	11,0
9,4	29,0	28,1	28,6	29,1	0,9	0,2	0,0
9,7	32,3	27,5	28,0	28,5	23,4	18,6	14,3
EQM(Σ)					140,3	134,7	135,3

Figura 32. Identificação do fator k de ocupação do participante na sessão

Fonte: O autor, 2019.

Inicialmente, variou-se o k em intervalos de 0,1 e foi calculado o EQM para cada k visando identificar qual k melhor aproxima a curva kM dos dados reais tem $k=0,5$ (o produz o menor EQM). Depois, esse procedimento foi repetido para calcular a segunda casa decimal para k . O resultado, para aqueles dados, é $k=0,53$. Esse resultado significa que aquele grupo entrou em equilíbrio quando alcançou aproximadamente 53% do seu limite máximo de produção de mensagens na sessão de bate-papo.

Com esse raciocínio tem-se um primeiro modelo analítico para a quantidade de mensagens produzidas por participante (m_p) em função do tamanho do grupo de participantes (p) em uma sessão de bate-papo educacional, dado pela equação a seguir:

$$m_p = kM \quad (13)$$

Substituindo M da Equação (13) pelo M da Equação (10), tem-se:

$$m_p = kM = k \frac{T_s}{T_d + pT_l} \quad (14)$$

A Equação (14) é aqui denominada de modelo PO. Estudo equivalente também pode ser realizado sobre a curva MG [máximo de mensagens de grupo, Equação (11)], substituindo a produção individual de mensagens pela quantidade de mensagens produzidas pelo grupo (m_g):

$$m_g = kMG = k p \frac{T_s}{T_d + pT_l} \quad (15)$$

Como exemplo da aplicação da Equação (15), diferentes curvas kMG (com k variando de 0,1) foram traçadas considerando o contexto da turma TIAE 2007.1, apresentadas na Figura 33. Novamente, a constante $k=0,5$ se mostrou a mais adequada para aqueles dados, como era de se esperar.

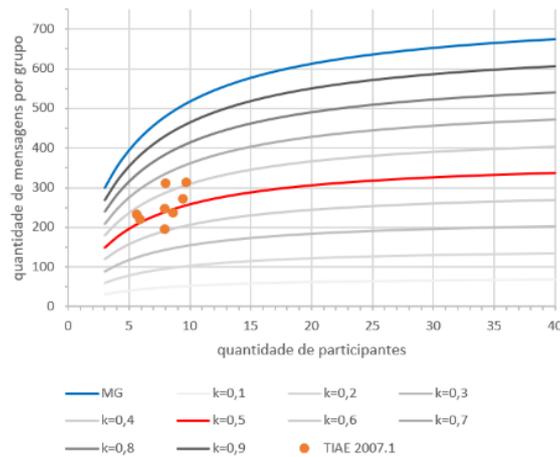


Figura 33. Identificação do fator k de ocupação do grupo TIAE 2007.1
 Fonte: O autor, 2019.

As Equações (14) e (15) são os primeiros modelos analíticos, elaborados nesta tese, para estimar a produção de mensagens m em função do tamanho p do grupo de participantes da sessão de bate-papo. Essas equações são bem simples, e dependem apenas do tempo médio de digitação de uma mensagem típica (T_d), do tempo médio de leitura de uma mensagem típica (T_l), da duração da sessão de bate-papo (T_s) e de um fator de ocupação (k) – parâmetros que podem ser obtidos empiricamente. Os parâmetros T_d e T_l já foram empiricamente estudados no Capítulo 4; o parâmetro T_s é arbitrário, que é o tempo que deve durar uma sessão de bate-papo, a ser definido pelo professor ou pelo curso; deseja-se, agora, obter valores médios para o parâmetro k (Percentual de Ocupação), considerando diferentes contextos (diferentes níveis de curso, perfil de alunos, complexidade da discussão etc.). Nesta tese, foi realizado o estudo do k considerando os dados empíricos obtidos das turmas TIAE e DSI, conforme apresentado nas subseções a seguir.

5.3.1. Percentual de ocupação das turmas TIAE

Com base nos *logs* de bate-papos das turmas TIAE, é possível calcular um percentual de ocupação médio dos participantes nas sessões de bate-papo daquela disciplina (e não apenas de uma turma, como visto anteriormente). Com os tempos médios de digitação e leitura, pode-se estimar o limite máximo de mensagens por participante por

hora de bate-papo, e, em seguida, calcular a porcentagem alcançada daquele valor com base nos valores realmente observados na turma. Por exemplo, considerando os dados da turma TIAE 2007.1 apresentados na Tabela 11, conclui-se que aquela turma, em média, ficou ocupada 52,7% do tempo, o que corresponde ao k estimado na seção anterior ($k=0,53$).

Tabela 11 Dados da turma TIAE 2007.1

C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11
Sessão de bate-papo na turma	Duração dos debates (min)	Número de participantes antes nos debates	Quantidade de mensagens de debate (descartando mensagens da sessão antes e após debate, e as mensagens automáticas)	Qtd msg por 1h de debate	Média de caracteres / mensagens de debate	Tempo médio de leitura (min)	Tempo médio de escrita (min)	Estimativa do Limite Máximo de mensagens por participante por hora [dada pela Equação (10)]	Produção de mensagens por participante durante 1h de debate (C5/C3)	Porcentagem do Limite Máximo atingido (C10/C9)
1	59,4	7,9	194	196,1	116,4	0,10	0,55	49,3	24,8	50,3
2	59,8	8,6	236	236,8	92,5	0,08	0,44	58,8	27,5	46,8
3	58,2	9,7	303	312,4	102,2	0,08	0,49	49,2	32,2	65,5
4	59,6	9,4	271	272,8	105,2	0,09	0,50	48,8	29,0	59,5
5	60,0	5,9	222	222,0	90,5	0,07	0,43	75,3	37,6	50,0
6	51,0	8,0	265	311,8	88,9	0,07	0,42	64,0	39,0	60,8
7	58,7	7,9	241	246,3	89,4	0,07	0,43	64,2	31,2	48,6
8	58,3	5,7	227	233,6	67,4	0,06	0,32	103,0	41,0	39,8
Média	58,1	7,9	244,8	254,0	94,1	0,08	0,45	64,1	32,8	52,7

Fonte: O autor, 2019.

O mesmo cálculo pode ser feito para as demais turmas TIAE, conforme dados apresentados na Tabela 12.

Tabela 12 Cálculo das porcentagens médias de atingimento do limite máximo

C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13
Turma	Sessão de bate-papo na turma	Duração dos debates (min)	Número de participantes nos debates	Quantidade de mensagens na sessão de bate-papo	Quantidade de mensagens de debate (descartando mensagens da sessão antes e após debate, e as mensagens automáticas)	Qtd msg por 1h de debate	Média de caracteres / mensagens de debate	Tempo médio de leitura (min)	Tempo médio de escrita (min)	Estimativa do Limite Máximo de mensagens por participante por hora dado pela Equação (10)	Produção de mensagens por participante durante 1h de debate (C7/C4)	Porcentagem do Limite Máximo atingido
2006.1	1	54,2	8,2	276	270	298,9	90,2	0,07	0,43	62,6	36,5	58,3
	2	60,7	9,9	274	270	266,9	104,1	0,09	0,50	54,7	27,0	49,4
	3	57,8	10,9	300	286	296,9	95,4	0,08	0,45	55,8	27,2	48,7
	4	53,0	9,0	255	255	288,7	96,2	0,08	0,46	78,2	32,1	41,0
	5	53,0	8,5	288	275	311,3	98,8	0,08	0,47	59,5	36,6	61,5
	6	60,3	8,4	242	240	238,8	120,5	0,10	0,57	56,0	28,4	50,7
	7	61,1	10,0	283	281	275,9	99,5	0,08	0,47	69,8	27,6	39,5
	8	55,3	10,3	221	212	230,0	103,5	0,09	0,49	60,2	22,3	37,0
2006.2	1	59,7	8,1	365	348	349,7	80,7	0,07	0,38	70,0	43,2	61,7
	2	59,1	8,0	329	321	325,9	80,9	0,07	0,39	70,4	40,7	57,8
	3	56,8	8,9	331	316	333,8	96,5	0,08	0,46	55,1	37,5	68,1
	4	57,9	4,9	298	291	301,6	70,1	0,06	0,33	107,3	61,6	57,4
	5	58,2	7,6	406	382	393,8	62,9	0,05	0,30	93,5	51,8	55,4
	6	57,2	6,0	366	350	367,1	61,8	0,05	0,29	109,2	61,2	56,0
	7	59,4	5,7	329	324	327,3	81,7	0,07	0,39	85,0	57,4	67,5
	8	54,5	6,9	339	275	302,8	75,0	0,06	0,36	83,1	43,9	52,8
2007.1	1	59,4	7,9	268	194	196,1	116,4	0,10	0,55	49,3	24,8	50,3

	2	59,8	8,6	250	236	236,8	92,5	0,08	0,44	58,8	27,5	46,8
	3	58,2	9,7	323	303	312,4	102,2	0,08	0,49	49,2	32,2	65,4
	4	59,6	9,4	326	271	272,8	105,2	0,09	0,50	48,8	29,0	59,4
	5	60,0	5,9	223	222	222,0	90,5	0,07	0,43	75,3	37,6	49,9
	6	51,0	8,0	278	265	311,8	88,9	0,07	0,42	64,0	39,0	60,9
	7	58,7	7,9	251	241	246,3	89,4	0,07	0,43	64,2	31,2	48,6
	8	58,3	5,7	235	227	233,6	67,4	0,06	0,32	103,0	41,0	39,8
2007.2	1	53,9	11,0	305	234	260,5	114,1	0,10	0,54	39,6	23,7	59,8
	2	42,1	11,2	388	222	316,4	65,8	0,06	0,31	67,9	28,3	41,7
	3	56,7	10,0	278	275	291,0	88,8	0,08	0,42	54,4	29,1	53,5
	4	48,7	11,7	308	249	306,8	94,9	0,08	0,45	45,6	26,2	57,5
	5	52,7	5,9	247	201	228,8	91,2	0,08	0,43	73,7	38,8	52,6
	6	60,0	7,8	283	276	276,0	67,8	0,06	0,32	83,9	35,4	42,2
	7	31,2	10,8	232	197	378,8	67,4	0,06	0,32	68,0	35,1	51,6
	8	59,1	8,8	295	273	277,2	78,0	0,07	0,37	67,5	31,5	46,7
2008.1	1	54,9	7,7	373	302	330,1	73,4	0,06	0,35	77,9	42,9	55,1
	2	55,2	6,5	330	245	266,3	88,8	0,08	0,42	71,4	41,0	57,4
	3	55,3	6,7	288	251	272,3	88,1	0,08	0,42	70,7	40,6	57,4
	4	59,6	5,1	298	245	246,6	95,7	0,08	0,46	75,8	48,4	63,9
	5	59,6	5,5	301	265	266,8	77,5	0,07	0,37	89,9	48,5	53,9
	6	58,0	4,9	233	197	203,8	79,6	0,07	0,38	93,1	41,6	44,7
	7	58,2	6,9	321	280	288,7	83,4	0,07	0,40	73,4	41,8	56,9
	8	61,1	6,6	442	410	402,6	60,8	0,05	0,29	103,4	61,0	59,0
Média	-	56,2	8,0	299,5	269,4	288,8	87,1	0,07	0,41	71,0	37,8	53,5
DP	-	5,5	1,9	50,6	48,2	48,5	15,1	0,01	0,07	17,2	10,5	7,9

Fonte: O autor, 2019.

Pelos cálculos apresentados na Tabela 12, o Percentual de Ocupação médio de todas as turmas TIAE foi de 53,5% (com um desvio padrão de 7,9%). Esse valor é um pouco abaixo dos 60% utilizados para o dimensionamento de *mainframes* (conforme apresentado no início desta seção).

5.3.2. *Percentual de ocupação da turma DSI*

Assim como foi calculado o percentual médio de ocupação dos participantes nas sessões de bate-papo de todas as turmas TIAE ($k=0,57$), é de interesse calcular também esse percentual no contexto de uma outra disciplina para fins de comparação. Foram extraídos os dados dos *logs* da turma DSI 2016.1, levando em consideração os mesmos parâmetros extraídos das turmas TIAE. Na turma DSI 2016.1, foram realizadas 13 sessões de bate-papo e os dados produzidos estão listados na Tabela 13. Naquela disciplina, o percentual médio de ocupação foi de 49%, um pouco abaixo do percentual médio de ocupação das turmas TIAE. Como se pode constatar, esse dado depende do contexto: em algumas turmas, os participantes se mantêm mais ocupados que em outras turmas (por diferenças na dinâmica, na mediação, pelas capacidades dos participantes, dentre outros motivos).

Tabela 13 Cálculo do percentual de ocupação para as sessões da turma DSI 2016.1

C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13
Turma	Sessão de bate-papo na turma	Duração dos debates (min)	Número de participantes nos debates	Quantidade de mensagens na sessão de bate-papo	Quantidade de mensagens de debate (descartando mensagens da sessão antes e após debate, e as mensagens automáticas)	Qtd msg por 1h de debate	Média de caracteres / mensagens de debate	Tempo médio de leitura (min)	Tempo médio de escrita (min)	Estimativa do Limite Máximo de mensagens por participante por hora dado pela Equação (10)	Produção de mensagens por participante durante 1h de debate (C7/C4)	Porcentagem do Limite Máximo atingido
2016.1	1	57,0	6,0	627	499	525,3	33,5	0,03	0,18	187,7	87,6	46,6
	2	38,0	8,0	309	205	323,7	60,3	0,05	0,33	88,9	40,5	45,5
	3	33,0	6,5	221	179	325,5	49,9	0,04	0,27	120,8	50,1	41,5
	4	60,0	15,0	925	581	581,0	50,9	0,04	0,27	69,3	38,7	55,9
	Janaína	18,9	2,0	57	40	127,0	91,9	0,08	0,41	123,2	63,5	51,5
	Lilian	24,8	2,0	99	77	186,3	57,4	0,05	0,32	163,2	93,2	57,1
	José Carlos	19,9	2,0	55	41	123,6	77,2	0,06	0,36	140,7	61,8	43,9
	Gustavo	30,8	2,0	162	145	282,5	51,0	0,04	0,26	197,0	141,3	71,7
	Thiago e Tadeu	39,2	3,0	207	107	163,8	61,9	0,05	0,29	153,8	54,6	35,5
	Tulio	26,6	2,0	207	74	166,9	66,7	0,06	0,28	176,7	83,5	47,2
	Mirian	30,8	2,0	96	78	151,9	64,6	0,05	0,36	145,5	76,0	52,2
	Monica, Tete, Luiza	45,4	3,1	223	194	256,4	45,4	0,04	0,26	176,0	82,7	47,0
Maria Evanilda	44,8	2,0	68	55	73,7	101,5	0,08	0,61	86,1	36,9	42,8	
Média	-	36,1	4,3	250,5	175,0	252,9	62,5	0,05	0,32	140,7	70,0	49,1
DP	-	12,5	3,7	243,2	165,6	148,8	18,0	0,0	0,1	39,3	28,9	8,7

Fonte: O autor, 2019.

5.4. Produção esperada de mensagens: Lê, Digita e Reflete (LDR)

Um participante no bate-papo não fica apenas lendo as mensagens que chegam e digitando novas mensagens o tempo todo. Ao longo da sessão, o participante também fica ora relendo mensagens anteriores do *log* da sessão de bate-papo, ora aguardando uma resposta, ora pensando no que pode escrever ou no que acabou de ler. Nesta seção, busca-se modelar esse cenário.

5.4.1. Modelo analítico da produção esperada de mensagens

Nesta pesquisa, considera-se como Tempo de Reflexão (T_R) todo o tempo em que o participante não está digitando novas mensagens nem lendo as mensagens que acabaram de chegar. Esse tempo está caracterizado pela Equação (6) da Subseção 4.3.2, que pode ser reescrita da seguinte maneira:

$$T_S = T_L + T_D + T_R \quad (16)$$

O Tempo de Reflexão inclui o tempo empregado para realizar diferentes atividades, uma delas é ficar aguardando respostas (T_{ar}) (conforme discutido na Subseção 4.3.1), o que, pela Teoria de Filas, pode ser modelado pela Equação (5): $T_{ar} = T_d / (p-1)$.

Enquanto uma parte do tempo de reflexão é para aguardar uma resposta, a parte que sobra será aqui denominada Tempo Livre (T_{LV}). Portanto:

$$T_R = T_{AR} + T_{LV} \quad (17)$$

Substituindo a Equação (17) na Equação (16), tem-se:

$$T_S = T_L + T_D + T_{AR} + T_{LV} \quad (18)$$

A Equação (18) é relativa ao tempo total da seção. Neste tempo, um participante enviou m_p mensagens. Entre o envio de uma mensagem e o envio da seguinte, o participante passa pelo ciclo de produção de mensagem (T_{cpm}), um período em que o participante passa um tempo digitando uma mensagem (T_d), lendo (T_l) as p mensagens que os demais participantes também enviaram naquele ciclo, e também passa um tempo refletindo (T_r). O tempo refletindo, por sua vez, inclui um tempo aguardando a resposta sobre a mensagem recém-enviada (T_{ar}) e um tempo em que fica livre (T_{lv}). Dessas considerações, resulta a equação:

$$\begin{aligned}
 T_s &= m_p T_{cpm} \\
 T_s &= m_p (T_d + pT_l + T_r) \\
 T_s &= m_p (T_d + pT_l + T_{ar} + T_{lv}) \\
 m_p &= \frac{T_s}{T_d + pT_l + T_{ar} + T_{lv}} \tag{19}
 \end{aligned}$$

Substituindo a Equação (5) na Equação (19), tem-se Equação (20), que é a equação que representa o modelo LDR:

$$m_p = \frac{T_s}{T_d + pT_l + \frac{T_d}{p-1} + T_{lv}} \tag{20}$$

Para exemplificar a aplicação dessa equação, considere os dados do contexto TIAE 2007.1, em que as mensagens produzidas tiveram em média 94,1 caracteres, o que possibilita calcular o tempo médio de digitação de uma mensagem pela Equação (3), $T_d = 0,44$ minutos; e calcular o tempo médio de leitura de uma mensagem pela Equação (1), $T_l = 0,08$ minutos. Com esses parâmetros aplicados na Equação (20), e considerando que a sessão dura 60 minutos (T_s), tem-se: $m_p = 60 / (0,44 + p \cdot 0,08 + (0,44 / (p-1)) + T_{lv})$. Nesta pesquisa não há uma equação para se calcular o fator T_{lv} . Na Figura 34, é apresentada a plotagem dessa curva (m_p) considerando diferentes tempos livres: $T_{lv} = 0,4; 0,8; 1,2$.

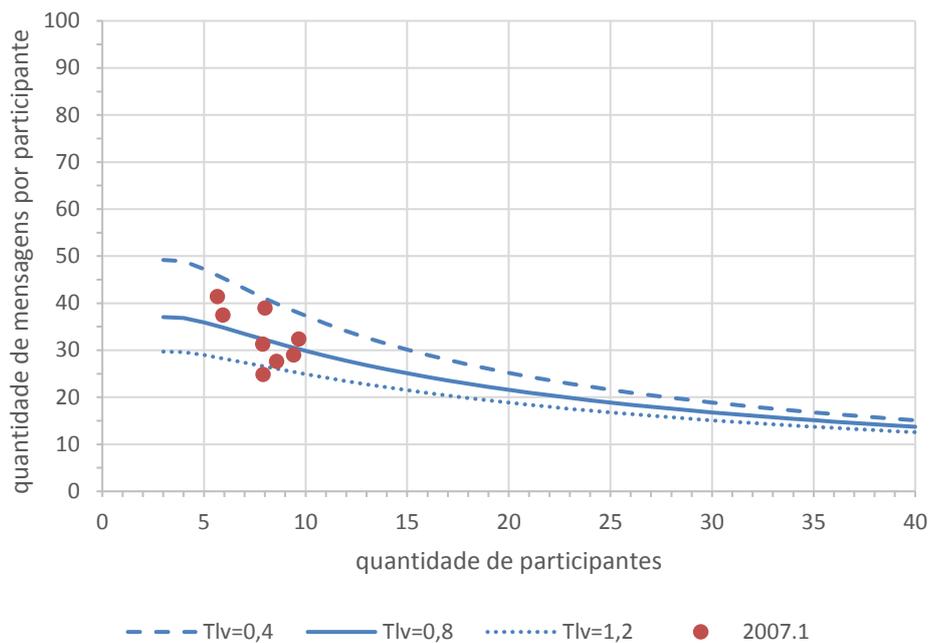


Figura 34 Representação das curvas m_p com diferentes T_{LV}
 Fonte: O autor, 2019.

O parâmetro T_{lv} deve ser ajustado para aproximar a curva aos dados reais. Verifica-se que, dentre as curvas traçadas, o parâmetro $T_{lv} = 0,8$ é o que mais aproxima a curva aos dados reais. Ajuste semelhante também foi realizado no modelo PO (fator k de ocupação).

Uma vez que, nesta pesquisa, foram propostas duas equações para o fenômeno em estudo, é importante compará-las: a Equação (14) relacionada ao modelo PO, e a Equação (20) relacionada ao modelo LDR.

A Equação PO parte de um modelo simples, e resulta em uma equação mais simples que a Equação LDR, pois a Equação PO contém apenas dois fatores no denominador, enquanto a equação LDR possui quatro fatores. A Equação LDR, por sua vez, é mais complexa porque parte de um modelo com mais conhecimento sobre o fenômeno, tendo sido modelado o tempo de resposta-aguardada (que é inversamente proporcional à quantidade de participantes, assumindo que o processo de chegada de mensagens tem distribuição de Poisson). A forma da curva produzida pela Equação LDR é mais condizente com a expectativa inicial da curva idealizada para o fenômeno em estudo, e por isso parece estar mais correta que a Equação PO [ver Seção 1.3 Modelo conceitual preliminar (conjecturas iniciais), Figura 3].

Ambas as equações apresentam os mesmos parâmetros: T_l , T_d , T_s , e uma constante de ajuste, sendo que na Equação PO a constante de ajuste foi denominada k (percentual de ocupação), e na Equação LDR a constante de ajuste foi denominada T_{lv} (tempo livre).

Em ambas as equações, o fator de ajuste só pode ser determinado em função dos dados reais observados, pois o ajuste é uma representação de tudo o que não foi considerado no modelo, mas que de algum modo influencia a produção de mensagens, tal como a dinâmica de conversação, a complexidade do tema em discussão etc.

Na Figura 35 e na Figura 36, as duas curvas foram traçadas considerando os ajustes dessas curvas para os dados da turma TIAE 2007.1 e para a turma DSI 2016.1, respectivamente.

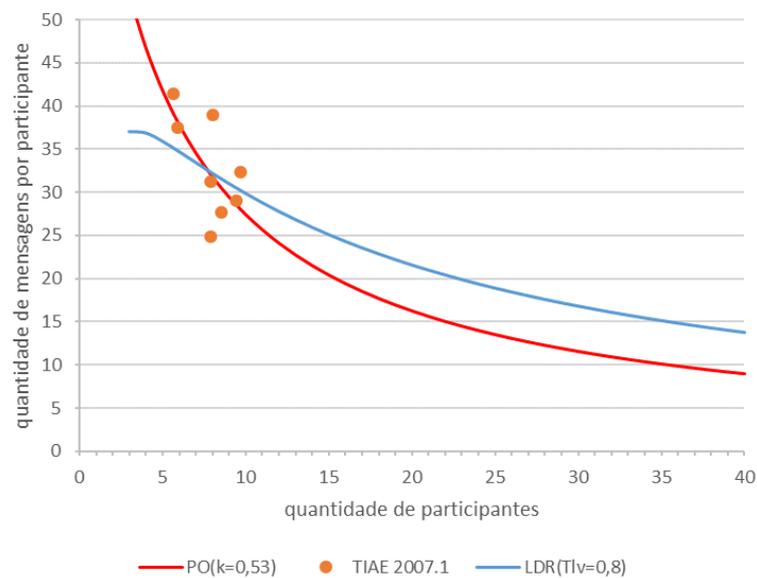


Figura 35. Comparação das curvas da Equação PO e da Equação LDR, considerando os dados reais da turma TIAE 2007.1

Fonte: O autor, 2019.

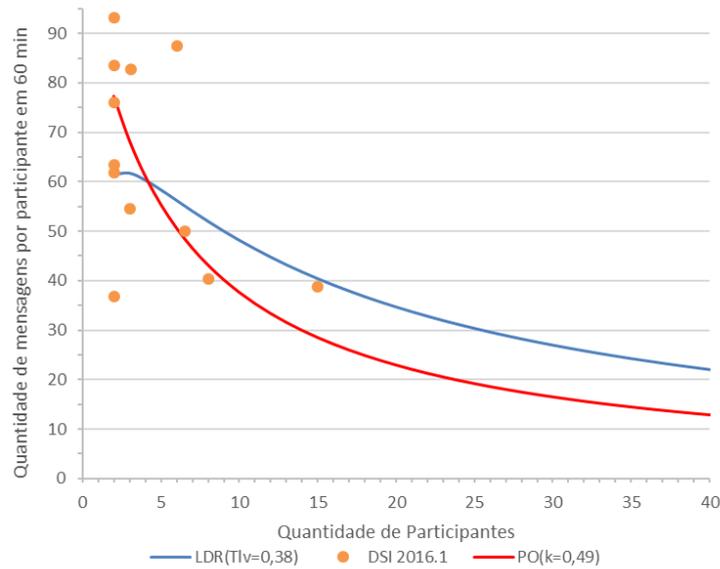


Figura 36. Comparação das curvas da Equação PO e da Equação LDR, considerando os dados reais da turma DSI 2016.1

Fonte: O autor, 2019.

A partir da análise visual de ambas as curvas, três aspectos são observados.

O primeiro é com relação ao início das curvas: a curva resultante da Equação LDR apresenta um crescimento inicial para depois começar a decair (como se fosse uma onda), o que está de acordo com as expectativas para a curva do fenômeno em estudo (ver Seção 1.3); já a curva resultante da Equação PO é sempre decrescente, não apresenta um comportamento inicial diferenciado.

Outro aspecto a ser analisado é com relação às extremidades das curvas. Ambas as curvas tendem a zero conforme aumenta a quantidade de participantes do grupo.

O terceiro aspecto é a diferença nas estimativas produzidas conforme aumenta a quantidade de participantes: a curva da Equação LDR produz estimativas mais altas do que a curva da Equação PO.

5.4.2. Modelo estocástico da produção esperada de mensagens

Para desenvolver o modelo estocástico da produção esperada de mensagens, foi elaborado o diagrama de estados com 3 estados apresentado na Figura 37, considerando

que um participante, além de ler e digitar mensagens, também reflete em alguns momentos.

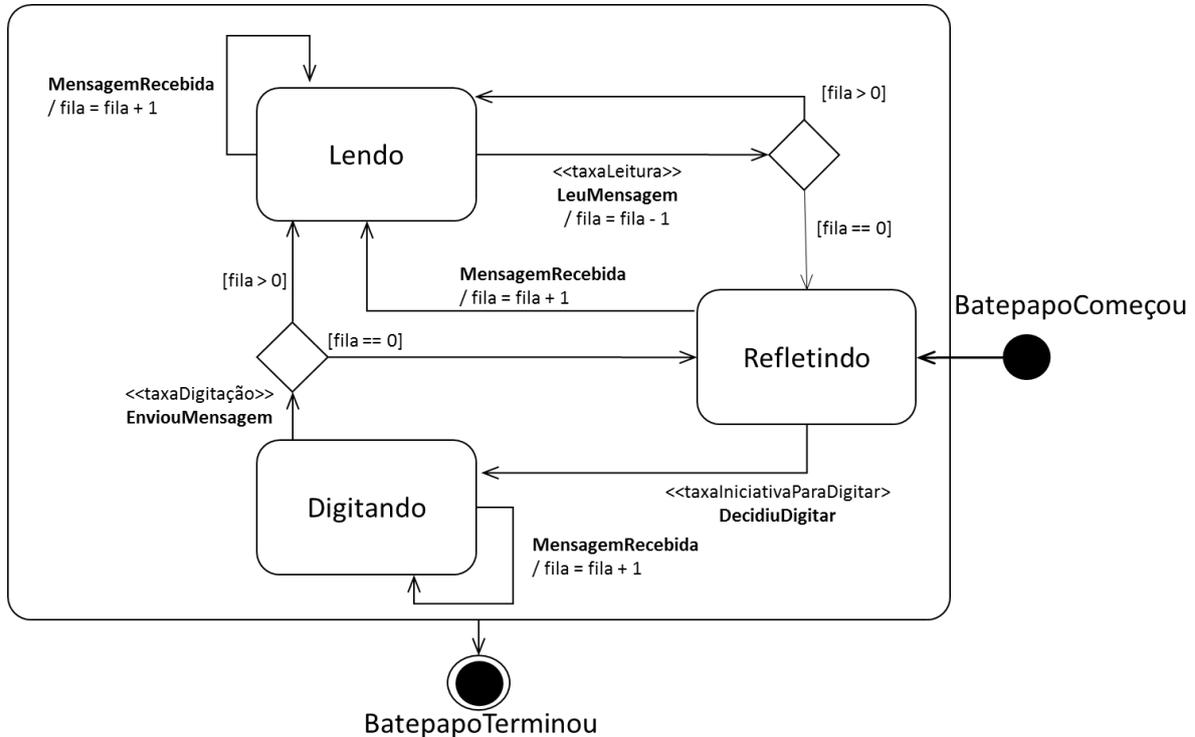


Figura 37. Diagrama de estado de 1 indivíduo, contendo três estados
 Fonte: O autor, 2019.

Neste diagrama, o indivíduo passa para o estado Refletindo quando não tem mais mensagens a serem lidas. Ele fica refletindo por um tempo até que: chega uma nova mensagem a ser lida e, imediatamente, o usuário volta para o estado Lendo; ou fica refletindo por algum tempo até que decide digitar uma nova mensagem, indo para o estado Digitando.

Para se obter as medidas de interesse, o modelo de três estados foi implementado no Tangram-II, porém ao se resolver o modelo para mais de 6 participantes, ocorria uma explosão de estados possíveis na Cadeia de Markov. Isso se dá pois cada um dos objetos participantes, além de possuir uma variável que guarda a quantidade de mensagens na sua fila, também guarda um estado, que neste modelo podem ser três. Para contornar o problema de explosão de estados do ambiente Tangram-II, foi elaborada uma segunda estratégia de modelagem com apenas 2 objetos: um Participante com o comportamento idêntico ao da estratégia de modelagem anterior, e um Grupo de Participantes modelado

como uma fonte Poisson que produz mensagens emitidas para o Participante a uma taxaEnvioMsg (Figura 38).

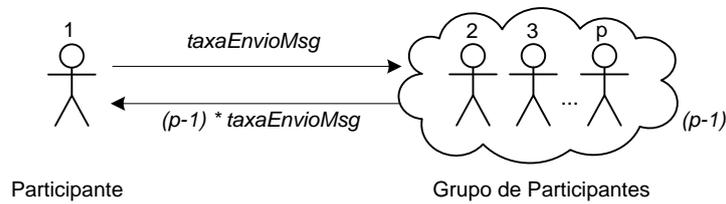


Figura 38 Estratégia de modelagem: interação grupo-indivíduo

Fonte: O autor, 2013.

Nesta estratégia, cada participante do Grupo deve produzir mensagens numa taxa semelhante à do indivíduo isolado, pois assume-se que todos os participantes do bate-papo têm um comportamento similar. Isto significa que a taxa de envio do Grupo dividida por $(p-1)$ precisa ser próxima ao número de mensagens enviadas pelo indivíduo isolado. O valor da taxa de mensagens enviadas pelo indivíduo isolado é a medida de interesse que se quer encontrar.

Para obter essa taxa, o primeiro passo é atribuir um valor aleatório para a taxa de produção de mensagens pelo Grupo e, pelo Tangram-II, calcular a taxa de envio de mensagens do indivíduo isolado. Em seguida, compara-se a taxa atribuída ao Grupo e a produzida pelo indivíduo. A diferença entre as taxas deve ser a menor possível; na presente pesquisa, foi definido que a diferença deve ser inferior a 1 mensagem para se obter maior precisão. Vai-se ajustando a estimativa da taxa de envio de mensagens do Grupo até que a taxa dos indivíduos do grupo fique próxima da taxa do indivíduo isolado. Desta forma, iterativamente, obtém-se uma estimativa para a medida de interesse da presente pesquisa para qualquer tamanho de grupo.

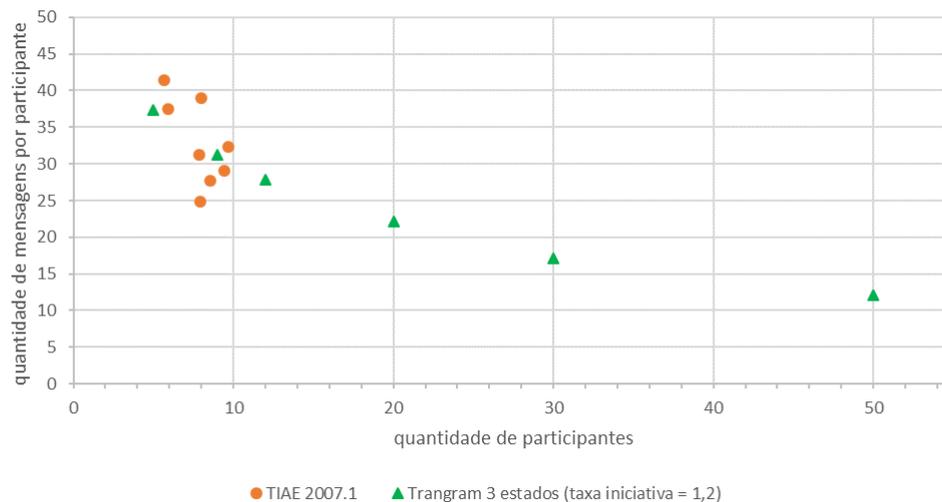


Figura 39 Estimativas do modelo de três estados obtidas pelo Tangram-II no contexto da turma TIAE 2007.1
 Fonte: O autor, 2019.

Com a segunda estratégia de modelagem, foi possível obter as medidas de interesse por meio do *Solver* do Tangram-II. As estimativas para a turma TIAE 2007.1 são apresentadas na Figura 39. É possível observar que, para os parâmetros daquela turma ($\text{taxaLeitura} = 12,9$; $\text{taxaEscrita} = 2,2$ e $\text{taxaIniciativa} = 1,2$), as estimativas para 5 participantes foi de 37 mensagens e para 9 participantes foi de 31 mensagens. Essas estimativas dividem ao meio os oito pontos que representam as produções de mensagens das sessões reais de bate-papo realizadas naquela turma.

5.4.3. Comparação entre os modelos de produção esperada de mensagens

Dadas as duas abordagens (analítica e estocástica), é de interesse comparar as estimativas produzidas pelos distintos modelos, como apresentado na Figura 40.

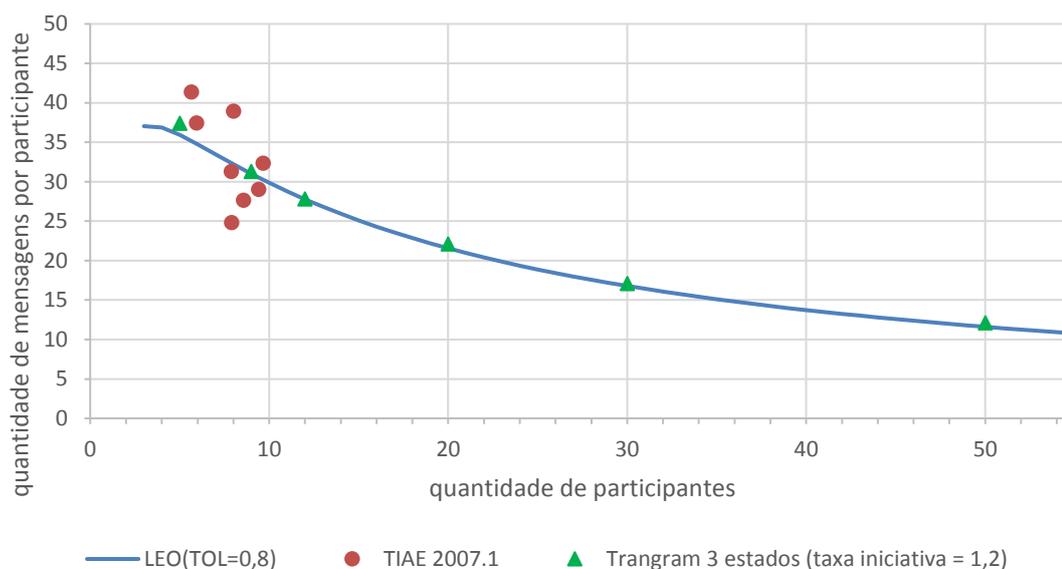


Figura 40. Modelo LDR: comparação da curva m_p produzida a partir da equação (20) e as estimativas obtidas pelo Tangram-II, com 3 estados, no contexto da turma TIAE 2007.1

Fonte: O autor, 2019.

Conforme pode ser observado na Figura 40, a curva m_p produzida pela Equação (20) e os pontos produzidos pelo Tangram-II são bastante próximos, o que indica que a Equação (20) pode ser utilizada no lugar do modelo estocástico para a produção esperada de mensagens. A proximidade dos resultados também é um indicativo de que ambas as abordagens parecem adequadas (seria um problema se as abordagens resultassem em estimativas muito distintas).

Uma vez que a modelagem pelo processo estocástico produziu estimativas próximas das obtidas pela abordagem analítica do modelo LDR, e como as estimativas do modelo LDR foram diferentes das obtidas pelo modelo PO, nesta pesquisa será assumido que o modelo LDR é mais adequado para representar o fenômeno em investigação.

5.5. Modelo Lê, Digita, Refletindo e Motivado (LDRM)

Com a intenção de refinar ainda mais o diagrama de estados da participação numa sessão de bate-papo, foi considerado que o participante não digita mensagens

aleatoriamente, mas sim busca responder a alguma mensagem enviada anteriormente, dando continuidade à conversação. Esse comportamento foi representado pelo estado “Motivação” para digitar novas mensagens: o participante precisa estar motivado para digitar uma nova mensagem, e essa motivação só ocorre quando o participante lê alguma mensagem de outro participante. Considerando a motivação do participante para digitar mensagens e a ausência dessa motivação, foi elaborado o diagrama de estados apresentado na Figura 41.

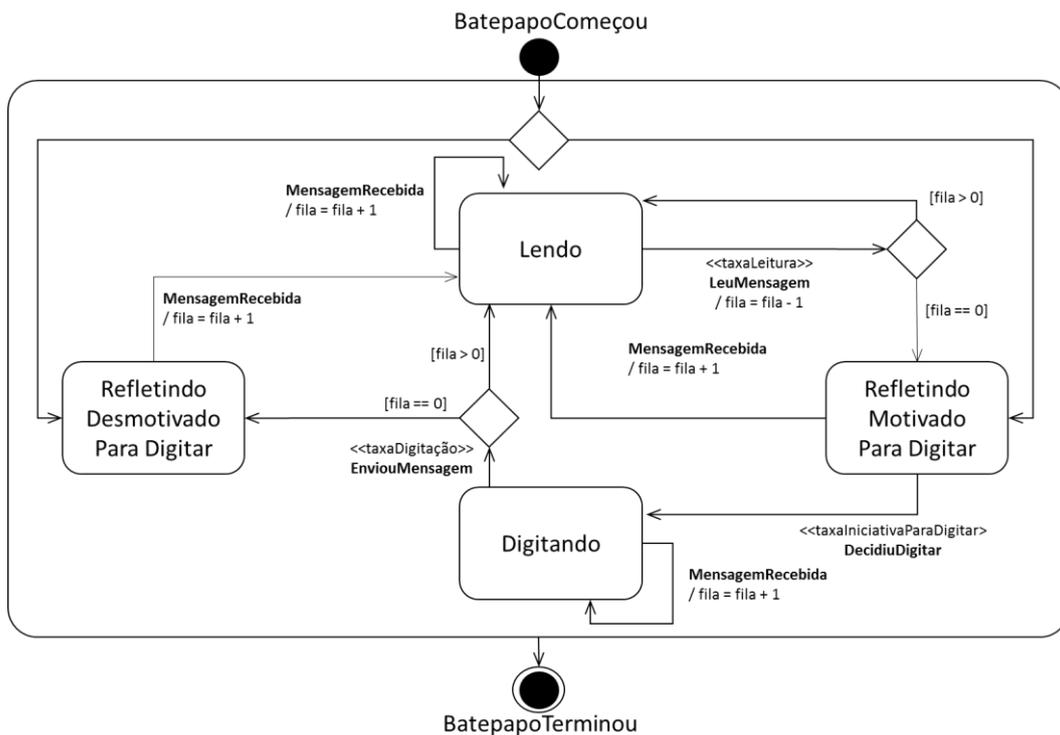


Figura 41. Diagrama de estado de 1 indivíduo, contendo quatro estados
 Fonte: O autor, 2019.

Nesse diagrama, o estado Refletindo é desmembrado em dois estados: Refletindo Motivado Para Digitar ocorre quando o participante acabou de ler alguma mensagem e, portanto, encontra-se motivado para enviar alguma resposta ainda que não tenha tomado a iniciativa de começar a digitar; e Refletindo Desmotivado Para Digitar, quando o participante acabou de enviar uma mensagem e fica aguardando até chegar alguma mensagem de algum colega no bate-papo.

É preciso ressaltar que, numa sessão real de bate-papo, um participante chega a enviar duas mensagens mesmo sem ter recebido uma mensagem de algum colega. Por outro lado, ninguém envia várias mensagens numa sessão se nenhum outro participante

envia uma resposta, pois nesse caso o usuário supõe que os demais participantes não estão dispostos a interagir e para de enviar mensagens. Todas essas considerações implicam diagramas de estados cada vez mais elaborados e complexos.

Nesta pesquisa, não foi desenvolvido um modelo analítico para o modelo que considera um 4º estado (o da Motivação). Segundo apresentado em Rocha, Pimentel e Diniz (2015), a modelagem da participação com um modelo de quatro estados se mostrou desnecessária, já que apresentava diferenças com o modelo de três estados apenas para grupos pequenos (*small groups*) com até cinco ou seis participantes. Como o objetivo é obter estimativas para grupos maiores, para as finalidades desta pesquisa, o modelo LDR, com apenas 3 estados, foi considerado suficiente por ser mais simples e conter um modelo analítico apto a ser usado no lugar do modelo estocástico.

6. Quantos alunos devem participar de uma sessão de bate-papo educacional?

Neste capítulo, pretende-se responder à questão inicial desta pesquisa: quantos alunos, no máximo, devem participar de uma sessão de bate-papo educacional? Responder essa questão é importante para apoiar professores e gestores educacionais no dimensionamento dos grupos de alunos em sessões de bate-papo no contexto da educação on-line, dado que o fórum e o bate-papo são os meios de conversação mais utilizados na educação a distância (EAD) e dada a consolidação da importância da educação on-line (não somente a EAD) no cenário cibercultural contemporâneo.

A relação entre o tamanho do grupo e a quantidade de mensagens produzidas por cada participante individualmente é o que nos possibilita esboçar uma resposta para a pergunta em questão com base nos modelos matemáticos aqui elaborados e investigados.

A primeira conclusão é que a literatura científica, identificada nesta pesquisa, está equivocada. Não existe um número ideal, um número idealizado ou mágico, como às vezes se encontra na literatura. Conforme discutido na Seção 2.5, a literatura científica apresenta números divergentes para esse número ideal: 4, 5, 7, 9, 12, 18 participantes. Conforme havíamos conjecturado no modelo idealizado, o que se pode dizer é que a produção de mensagens fica prejudicada à medida que aumenta o tamanho do grupo. Sendo assim, o que o gestor ou professor pode buscar, ao dimensionar o tamanho do grupo, é definir a qualidade da participação esperada em termos de quantidade de mensagens que os alunos devem ser capazes de enviar ao longo da sessão de bate-papo.

Se, para um determinado gestor ou professor, o aluno não precisasse enviar mensagens e pudesse apenas ficar lendo as mensagens, então não haveria um limite na quantidade de participantes, podendo-se incluir até milhares de “ouvintes”, como frequentemente ocorre nas sessões de bate-papo relacionadas aos programas televisivos.

Nas sessões educacionais de bate-papo, partimos do pressuposto de que o aluno deve ser capaz de ler todas as mensagens e ainda digitar uma mensagem de vez em quando. Mesmo supondo uma quantidade muito grande de participantes, a modelagem desenvolvida nesta pesquisa alerta para o fato de que há um limite na quantidade de mensagens que os participantes podem ler durante a sessão de bate-papo. Conforme visto no Modelo MG (Máximo de Mensagens do Grupo), representado pela Equação (12), há

um limite para a quantidade de mensagens que o participante é capaz de ler, e depende apenas da duração da sessão de bate-papo dividida pelo tempo médio de leitura de uma mensagem típica daquela sessão (no limite, é como se o participante ficasse apenas lendo as mensagens). No caso da turma de Tecnologia da Informação Aplicada à Informação (TIAE) 2007.1, esse limite seria de 750 mensagens; já para a turma de Docência em Sistemas de Informação (DSI) 2016.1, em que o tempo médio de leitura de uma mensagem típica foi de 0,05min (menor que a turma TIAE, pois as mensagens da turma DSI em média foram menores que as da turma TIAE), o limite máximo seria de $H = 60/0,05 = 1200$ mensagens.

Para os professores e gestores que objetivam uma educação baseada na interatividade (SILVA, 2000), em que o aluno deve ser capaz de intervir, espera-se que o aluno seja capaz de enviar uma mensagem de tempos em tempos na sessão de bate-papo. A quantidade de mensagens que o professor espera que o aluno seja capaz de enviar (um parâmetro da qualidade da participação na sessão de bate-papo) é o que nos serve como parâmetro para determinar a quantidade máxima de participantes no bate-papo. Por exemplo, pode ser que um professor espere que cada aluno envie uma mensagem a cada 2 minutos, ou talvez espere que os alunos não fiquem mais que 5 minutos apenas lendo mensagens.

Considerando o contexto da turma TIAE 2017.1, de acordo com o modelo analítico Lê, Digita e Reflete (LDR) da produção esperada de mensagens (Figura 42), se o gestor ou professor tiver a expectativa de que o aluno envie uma mensagem a cada 2 minutos, significa que cada aluno deve ser capaz de enviar 30 ou mais mensagens por hora de sessão de bate-papo, o que implica que o grupo deve conter 9 ou menos pessoas para que os participantes consigam produzir ao menos uma mensagem a cada 2 minutos. Relaxando esse limite para 3 minutos, então os participantes devem enviar 20 mensagens ao longo da sessão de 60 minutos, o que, pelo modelo, implica não ter mais que 22 participantes.

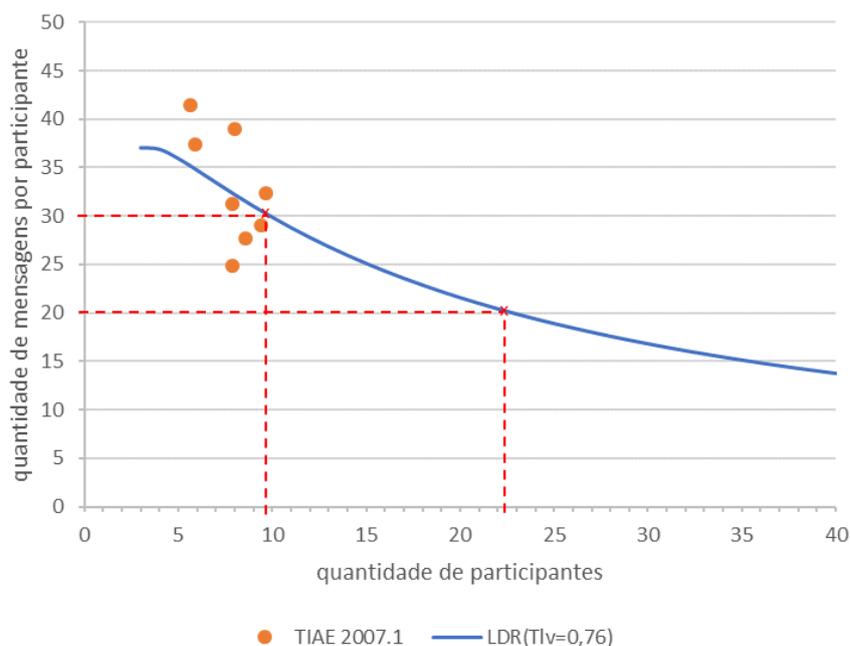


Figura 42. Estimativas para a quantidade máxima de alunos nas sessões de bate-papo no contexto da turma TIAE 2007.1

Fonte: O autor, 2019.

O mesmo tipo de estudo pode ser feito para o contexto DSI 2016.1, e os resultados obtidos serão diferentes. Conforme os dados apresentados na Figura 43, para a produção de 30 mensagens durante 1h de sessão de bate-papo (1 mensagem a cada 2 minutos), o grupo não deveria conter mais que 25 participantes.

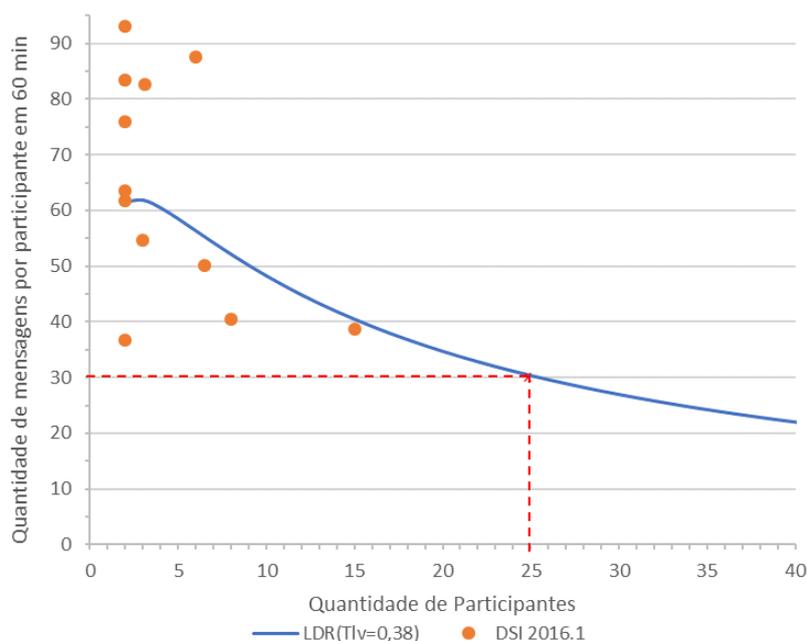


Figura 43. Estimativas para a quantidade máxima de alunos nas sessões de bate-papo no contexto da turma DSI 2016.1

Fonte: O autor, 2019.

Estes exemplos ilustram que a quantidade de participantes impacta a quantidade de mensagens produzidas. Não existe um número ideal, mas é possível estimar o tamanho do grupo em função da qualidade de participação desejada em termos de produção de mensagens.

Outra conclusão importante que esses exemplos ilustram é que a quantidade de participantes depende do contexto. Depende especialmente da velocidade de leitura e de escrita dos participantes da sessão (podendo variar muito se o grupo for formado por crianças, universitários ou idosos), bem como da quantidade média de caracteres por mensagem produzida no grupo, dentre outros parâmetros que também influenciam o modelo (ainda que não tenham sido discriminados no modelo, estão encapsulados no fator de ajuste do modelo). Portanto, mesmo que o professor-gestor delimite um parâmetro de qualidade como “o aluno deve ser capaz de enviar uma mensagem a cada 2 minutos”, como visto nos exemplos anteriores, o limite da quantidade de participantes irá variar muito em função do contexto.

7. Conclusão

Esta pesquisa teve como objetivo investigar o comportamento dos usuários de bate-papos educacionais de forma a caracterizar matematicamente o comportamento de participação: a leitura de mensagens, a digitação de mensagens, a reflexão ao longo da sessão, e especialmente a quantidade de mensagens produzidas por cada participante durante a sessão de bate-papo em função do tamanho do grupo de participantes. Para alcançar o objetivo, foi utilizado o método Modelagem Matemática, tendo como premissa a existência de uma relação inversamente proporcional entre as variáveis de interesse, produção de mensagens por participante “ m ” e quantidade de participantes “ p ” na sessão de bate-papo. Dentre as abordagens possíveis, foram usados o modelo analítico e o modelo estocástico.

O modelo analítico representa alternativas do problema como variáveis de decisão, e procura valores dessas variáveis que minimizem ou maximizem uma função dessas variáveis denominada função objetivo. Partindo das duas principais variáveis p e m , investigou-se quais outras variáveis deveriam compor a função objeto. Para isso, analisaram-se as atividades elementares dos participantes: leitura, digitação e reflexão. A partir da identificação dessas atividades, foram caracterizados os tempos médios necessários para realizar cada uma dessas atividades.

Em uma primeira modelagem, foi considerado que o participante permanece o tempo todo ocupado lendo e digitando mensagens. Ao serem comparadas as estimativas produzidas pela primeira equação, observaram-se valores de produção de mensagens por participante muito superiores aos valores observados nos *logs* reais de bate-papo. Essas estimativas são resultantes de uma ocupação de 100% do participante. Considerando que os participantes não ficam o tempo todo ocupados, o modelo analítico foi incrementado por uma variável que representa um percentual de ocupação (PO) dos participantes. Nas turmas investigadas, esse PO foi em média 53%. A equação da relação entre as variáveis de interesse p e m , os tempos médios de leitura e digitação e o percentual de ocupação compõem o primeiro modelo analítico, nomeado “Percentual de Ocupação”.

A segunda modelagem analítica é baseada no fato de que o participante não fica o tempo todo lendo e digitando mensagens. O participante aguarda o recebimento de novas mensagens, relê algumas mensagens, reflete sobre o que já leu e sobre o conteúdo que irá

digitar. Esse modelo foi denominado, na pesquisa, “Modelo LDR” (Lê, Digita e Reflete). Esse modelo produz estimativas de produção de mensagens próximas do real.

O modelo estocástico parte da identificação dos estados em que um participante de bate-papo pode se encontrar em um determinado momento da sessão. Nessa abordagem, utiliza-se a Teoria de Filas, e, no caso da conversação pelo bate-papo, o participante foi modelado como uma fonte geradora de mensagens, que envia mensagens para as filas dos demais participantes, e como um atendente, que lê as mensagens que são publicadas na sessão. No processo de leitura, pode-se formar uma fila: enquanto o participante está lendo uma mensagem, outras podem ser publicadas na sessão e são enfileiradas para o participante ler cada uma delas.

Foi desenvolvido um modelo baseado em Teoria de Filas considerando que o participante permanece o tempo todo ocupado Lendo e Digitando mensagens. Por meio da implementação do modelo no ambiente Tangram-II, foi possível obter as medidas de interesse, que são as quantidades de mensagens produzidas variando-se a quantidade de participantes. Esse modelo também produziu estimativas distantes dos valores de produção de mensagens observados em sessões reais de bate-papo. Os resultados obtidos com a abordagem estocástica foram muito próximos dos obtidos com o modelo analítico, o que corrobora a sua adequação.

Ainda com base na Teoria de Filas, o modelo foi incrementado com o estado Refletindo. Esse estado representa as outras atividades que não sejam as de leitura e digitação. As estimativas produzidas pela abordagem estocástica relacionada ao Modelo LDR foram próximas do modelo analítico LDR, o que contribui para a sua adequação.

Uma vez que o modelo estocástico exige um conhecimento de programação orientada a objeto e do ambiente de modelagem Tangram-II, a solução analítica mostrou-se mais vantajosa para a utilização por professores e gestores de educação. Sabendo-se os tempos médios de leitura e digitação dos alunos de uma turma, e estimando o parâmetro Tempo Livre (T_{LV}), a equação que representa o modelo analítico LDR pode ser usada para obter estimativas de produção de mensagens em função do número de participantes.

Em uma investigação preliminar, desenvolvida na dissertação de mestrado, foi modelado um quarto estado da participação, que representava a motivação ou não para digitar uma nova mensagem. Essa modelagem produz estimativas muito próximas às do modelo LDR com apenas 3 estados, apenas produzindo estimativas diferentes para grupos pequenos, com até cinco participantes. Contudo, esse modelo com 4 estados adiciona uma

complexidade desnecessária para investigar grupos maiores, que foi o objetivo da presente pesquisa.

Além de caracterizar matematicamente a participação dos usuários em sessões de bate-papo educacional e desenvolver modelos de produção de mensagens em função da quantidade de participantes, por meio dos estudos realizados nas turmas de Tecnologia da Informação Aplicada à Informação (TIAE) e de Docência em Sistemas de Informação (DSI), foi possível confirmar a relação inversamente proporcional entre as duas variáveis de interesse da pesquisa.

Como limitação da pesquisa, ressalta-se a pouca variância no número de participantes dentro de uma mesma turma de educação on-line. Nas turmas TIAE, embora o número de participantes variasse de uma turma para outra, de 4,8 participantes na turma 2005.1 para 17,8 participantes na turma 2002.1, dentro da mesma turma essa variação não ocorria, pois a participação nas sessões de bate-papo era, naquelas turmas, requisito parcial para aprovação na disciplina. Para contornar esse problema, realizou-se um estudo com a turma de DSI, forçando uma variação maior, de 2 a 15 participantes, porém não foi possível realizar uma sessão de bate-papo com todos os 30 alunos conforme havia sido inicialmente planejado, o que impossibilitou investigar uma amplitude ainda maior sobre os dados reais, de tal modo que se pudesse avaliar melhor as estimativas produzidas pelo modelo. Sem dados reais para grupos grandes, de 30 ou mais participantes, ainda que o modelo seja capaz de realizar estimativas de produção de mensagens para tais grupos, não foi possível validar as estimativas produzidas.

No sentido de generalização dos resultados da pesquisa, os estudos empíricos estão limitados a disciplinas de pós-graduação em informática. Não foram estudados outros contextos, como turmas de humanas, sociais ou de saúde. Não foram objeto de investigação contextos diferentes do nível superior (essa pesquisa não objetivou investigar o contexto do ensino fundamental e médio). Também não se investigou a influência, no contexto, de diferentes idades, como crianças e idosos.

Ainda como limitação da pesquisa, ressalta-se a necessidade de aprofundar as investigações sobre o que foi aqui denominado “tempo refletindo”. Foram feitos estudos sobre o processo de leitura e o processo de digitação de mensagens, porém é desejável que, em pesquisas futuras, seja mais bem caracterizado o comportamento do usuário quando não está digitando nem lendo uma mensagem. Identifica-se que, em trabalhos futuros, possam ser realizadas pesquisas com apoio de técnicas de rastreamento de retina

(*eyetracking*). Esses estudos aprimorariam os modelos desenvolvidos nos cálculos do tempo refletindo.

Ainda como trabalhos futuros, aponta-se uma proposta de implementar a equação do modelo desenvolvido nesta tese em um sistema de bate-papo, de forma que a partir de uma sessão prévia, o modelo pudesse ser parametrizado gerando relatórios ou um painel de controle (*dashboard*) sobre aquele determinado contexto. A utilização desse sistema de bate-papo geraria uma base de conhecimento que poderia melhorar as estimativas produzidas pelo modelo e identificar macrocontextos.

Por fim, esta pesquisa pretende reforçar a recomendação de inúmeros pesquisadores quanto ao uso de meios de conversação síncronos de comunicação, em conjunto com assíncronos, para possibilitar uma sala de aula mais interativa, com colaboração e construção de conhecimento (diferente das aulas com apenas transmissão de conhecimento). Pelas estimativas produzidas pelos modelos desenvolvidos nesta pesquisa, mostra-se que não é adequado realizar um bate-papo participativo com um número muito grande de alunos, pois quanto mais pessoas participam da sessão de bate-papo, menos mensagens conseguem enviar (além de outros problemas conhecidos na literatura, como a maior dificuldade para acompanhar a discussão).

Os resultados desta pesquisa devem ser divulgados de modo que professores e gestores envolvidos com a educação on-line e educação a distância (EAD) possam melhor dimensionar as sessões de bate-papo a fim de garantir o nível de qualidade desejada.

Referências

- ARENALES, M. *et al.* (2007). *Pesquisa Operacional*. São Paulo. Elsevier.
- BEAUVOIS, M. H. (1992). *Computer-assisted classroom discussion in the foreign language classroom: Conversation in slow motion*. *Foreign Language Annals*, 25, 455-464.
- BLACK, Ken. *Business Statistics Contemporary Decision Making*. Wiley, 2004.
- BÖHLKE, O. (2003). *A comparison of student participation levels by group size and language stages during chatroom and face-to-face discussions in German*. *CALICO Journal*, 21 (1), 67-87.
- BOLLIGER, Doris U.; MARTIN, Florence. *Instructor and student perceptions of online student engagement strategies*. *Distance Education*, p. 1-16, 2018.
- BRASIL (2018). Ministério da Educação. Diretrizes Curriculares Nacionais para o Ensino Médio. Available from:
http://portal.mec.gov.br/index.php?option=com_docman&view=download&alias=102371-proposta-apresentacao-dcnems-coletiva-de-imprensa-rev&category_slug=novembro-2018-pdf&Itemid=30192
- BRASIL (2016). Proposta de Emenda Constitucional 241, de 15 de jun. de 2016. *Altera o Ato das Disposições Constitucionais Transitórias, para instituir o Novo Regime Fiscal*. Brasília, DF, jun 2016. Disponível em:
<<https://www.camara.leg.br/proposicoesWeb/fichadetramitacao?idProposicao=2088351>>
- BUMP, J. (1990). *Radical changes in class discussion using networked computers*. *Computer and the Humanities*, 24, 49-65.
- BUZING, P., (2003). *Comparing different keyboard layouts: aspects of qwerty, dvorak and alphabetical keyboards*. Delft University of Technology Articles, Delft, Mai, 2003. Disponível em: <http://www.st.ewi.tudelft.nl/~buzing/Articles/keyboards.pdf>. Acessado em 09 de setembro de 2013.
- CALVÃO, L. D.; PIMENTEL, M.; FUKS, H. *Do email ao Facebook: uma Perspectiva Evolucionista sobre os meios de Conversação da Internet*. Rio de Janeiro: UNIRIO, 2014.

- CARR, T. *et al.* (2004). *From peripheral to full participation in a blended trade bargaining simulation*. British Journal of Educational Technology, 35(2), 197–211.
- CASPI, A., GORSKY, P., & CHAJUT, E. (2003). *The influence of group size on non-mandatory asynchronous instructional discussion groups*. The Internet and Higher Education, 6(3), 227–240.
- CASTRO, Laura Helena Pinto; PANTOJA, Lydia Dayanne Maia; PAIXÃO, Germana Costa. Diretrizes de Boas Práticas de Condução e Avaliação do Chat Educacional/Good Practices Guidelines for Conduction and Assessment of Educational Chat Rooms. Revista EducaOnline, v. 10, n. 2, p. 88-103, 2016.
- CENSOEAD.BR (2017) *Relatório analítico da aprendizagem a distância no Brasil*. Associação Brasileira de Educação a Distância. São Paulo: Pearson Education do Brasil http://abed.org.br/arquivos/CENSO_EAD_BR_2018_digital_completo.pdf>.
- COMTE, A. Discurso sobre o espírito positivo. Comte. Os Pensadores. São Paulo: Abril, 41-115, 1978.
- CUNHA, Leonardo Magela; FUKS, Hugo; LUCENA, CJP de. A Adaptação do Ambiente AulaNet para dar Suporte a Grupos de Aprendizagem e sua Formação Utilizando os Conceitos de Agentes de Software. Revista Brasileira de Informática na Educação, ISSN, p. 1414-5685, 2003.
- D'ÁVILA, Cristina Maria. Reforma do ensino médio pelo CNE. Bahia, 09 nov. 2018. Facebook: Cristina D'Ávila. Disponível em: <https://www.facebook.com/cristina.davila.372>. Acesso em: 15 jan. 2019.
- ESTRUC, M., PIMENTEL, M. (2012). “Portal Tagarelas: bate-papo para educação”. In: 23 Simpósio Brasileiro de Informática na Educação, Rio de Janeiro, Novembro, 2012.
- FILIPPO, D., PIMENTEL, M., WAINER, J. (2011). *Metodologia de pesquisa científica em sistemas colaborativos*. Sistemas Colaborativos, cap.23. Rio de Janeiro: Elsevier.
- FOGLIATTI, M. C., MATTOS, N. M. C. *Teoria de filas*. Rio de Janeiro: Interciência, 2007.
- FUKS, H., GEROSA, M.A., LUCENA, C.J.P. (2002), “*The Development and Application of Distance Learning on the Internet*”, The Journal of Open and Distance Learning, Vol. 17, N. 1, ISSN0268-0513, Fevereiro 2002, pp. 23-38.
- FUKS, H., PIMENTEL, M., LUCENA, C.J.P. (2006) *R-U-Typing-2-Me? Evolving a chat tool to increase understanding in learning activities*. International Journal of Computer-Supported Collaborative Learning, v. 1, p. 117-142.

GAINES, R. *et al.* (1980) *Authentication by keystroke timing: Some preliminary results*. Rand Report R-256-NSF. Rand Corporation, Santa Monica, CA, 1980.

GARCIA, J. *Personal identification apparatus*. (1986) Patent Number 4.621.334. U.S. Patent and Trademark Office, Washington, D.C., 1986.

HASTIE, Trevor J. Generalized additive models. In: *Statistical models* in S. Routledge, 2017. p. 249-307.

HILGERT, J.G. (2000) *A construção do texto “falado” por escrito na Internet. Fala e escrita em questão*. Dino Preti (org). São Paulo: Humanistas/FFLCH/USP.

HILTZ, S. R., TUROFF, M. *The Network Nation: Human Communication via Computer*. New York: Addison-Wesley, 1978. [Edição revisada. Cambridge, MA: MIT Press, 1993]

HRASTINSKI, S. (2007). *Participating in synchronous online education*. (Doctoral dissertation, Lund University).

HRASTINSKI, S. (2008b). *The potential of synchronous communication to enhance participation in online discussions: A case study of two e-learning courses*. *Information & Management*, 45, 499-506.

HRASTINSKI, S. (2008a). *What is online learner participation? A literature review*. *Computers & Education*.

HSIEH, Y., TSAI, C., (2012). *The effect of moderator’s facilitative strategies on online synchronous discussions*. *Computers in Human Behavior*, 28 (5) (2012), pp. 1708–1716.

INEP (2017). *Censo da Educação Superior 2017*. MEC, 2017. Disponível em <http://download.inep.gov.br/educacao_superior/censo_superior/documentos/2018/censo_da_educacao_superior_2017-notas_estatisticas2.pdf>.

KLEINROCK, L., (1975), *Queueing Systems - Theory*, Vol. I. Wiley-Interscience.
Gordon Pask, (1975), *The Cybernetics of Human Learning and Performance*.
Hutchinson

LAW, Averill M.; KELTON, W. D. *Simulation Modeling and Analysis* (2000).

LÖBER, A, SCHWABE, G, GRIMM S (2007). *Audio vs. chat: The effects of group size on media choice*. Proceedings of the 40th HICCS Hawaii International Conference on System Sciences.

LÖWY, Michael. Ideologia e ciências sociais. Elementos para uma análise marxista. 7ª ed. São Paulo: Cortez, 1991, p. 35, 1986.

LUCENA, C.J.P., FUKS, H. *Tecnologias de Informação Aplicadas à Educação (TIAE): Manual do Aprendiz*. Monografias em Ciência da Computação n.07/02, PUC-Rio, 2002.

MINAYO, María C. S. O desafio do conhecimento. Pesquisa qualitativa em saúde. 2008.

MORAES, E.; PIMENTEL, Mariano. Debatepapo: Sistema de bate-papo com Estruturação de Sequências e Visualização do Co-texto. In: Anais IADIS Conferência Ibero-Americana WWW/Internet (CIAWI 2010). Algarve, Portugal. 2010.

NETO, M. R.; PUREZA, V. Modelagem e simulação. Metodologia de pesquisa em engenharia de produção e gestão de operações, v. 2, p. 169-198, 2012.

PIMENTEL, M. G. (2002). *Hiperdiálogo: Ferramenta de Bate-papo para Diminuir a Perda de Co-texto*. Dissertação de Mestrado. IM-NCE/UFRJ.

PIMENTEL, M. G., FUKS, H., LUCENA, C. J. P. (2003a) “*Avaliação da Participação dos Aprendizes em Debates Síncronos*”, XIV Simpósio Brasileiro de Informática na Educação - SBIE 2003, Rio de Janeiro, UFRJ, p. 140-149.

PIMENTEL, M. G., FUKS, H. e LUCENA, C. J. P. (2003b) “*Debati, debati... aprendi? Investigações sobre o papel educacional das ferramentas de bate-papo*”, IX Workshop sobre Informática na Escola - WIE2003, Agosto.

PIMENTEL, M. GEROSA, M.A., FUKS, H. (2011) *Sistemas de comunicação para colaboração*. Sistemas Colaborativos, cap.5. Rio de Janeiro: Elsevier.

PIVA, D.J. et al. (2011). *EAD na Prática – Planejamento, métodos e ambientes de educação online*. Elsevier Editora Ltda.,2011, v.,p 164-170.

PRADO, D. *Teoria das Filas e da Simulação*. v.2, 4a ed. Nova Lima: INDG - Tecnologia e Serviços LTDA, 2009.

PRIMO, Alex Fernando Teixeira. A emergência das comunidades virtuais. Intercom, v. 20, 1997.

ROCHA, E. B., PIMENTEL, M., DINIZ, M. C. (2013). “*MODELO DE PARTICIPAÇÃO EM BATE-PAPO EDUCACIONAL*”. Dissertação (Mestrado em Informática) – Universidade Federal do Estado do Rio de Janeiro – UNIRIO. Rio de Janeiro, p. 123. 2013.

ROCHA, Edmilson Barcelos; PIMENTEL, Mariano; DINIZ, Morganna Carmem. Quantidade de Participantes em Bate-papo Educacional: um Modelo Baseado em Teoria de Filas. *Revista Brasileira de Informática na Educação*, 2015.

SHIMAKURA, S. (2006). CE003 - Estatística II. <<http://leg.ufpr.br/~silvia/CE003/>>, acessado em 20 jul 2013.

SILVA, Adilson Ricardo; PIMENTEL, Mariano. (2017) O Processo de Evolução de Regras de Gamificação para Emergir uma Inteligência Coletiva. Disponível em: <<http://csbc2017.mackenzie.br/public/files/14-sbsc/completos/14.pdf>>

SILVA, E. S., FIGUEIREDO, D.; LEÃO, R. (2009). *The TangramII integrated modeling environment for computer systems and networks*. SIGMETRICS Perform.Eval. Rev., 36(4):64–69. Disponível em: <http://www.land.ufrj.br/tools/tangram2/tangram2.html>. Acesso em: 15 mai. 2019.

SILVA, M. *Avaliação da Aprendizagem em Educação Online*. 2010. Entrevista publicada no YouTube: <<http://youtu.be/S7uUd6afEYE>>, acessado em 12 ago 2013.

SILVA, M. Sala de aula interativa. 2. ed. Quartet: Rio de Janeiro, 2000. [6 ed., São Paulo: Edições Loyola, 2012].

SIMON, Herbert A. *The sciences of the artificial*. Cambridge, MA, 1969.

STAHL, G. *Studying Virtual Math Teams*. New York, NY: Springer, 2009.

TUROFF, M. *IRC History – EMISARI*. 2000. Documento publicado em <http://www.livinginternet.com/r/ri_emisari.htm>, acessado em 29 abr 2013.

WAINER, J. (2007). *Métodos de pesquisa quantitativa e qualitativa para a Ciência da Computação*, em JAI 2007 - Jornada de Atualização em Informática, Anais do XXVII Congresso da Sociedade Brasileira de Computação. p.221-262.

Apêndice 1 – Roteiro para entrevista com os alunos da disciplina DSI 2016.1

1 - Experiência prévia

Você já tinha participado de alguma aula por bate-papo?

<se sim, peça para contar a experiência, o que faziam – se era para tirar dúvida, debater ou o quê? com que frequência se reuniam? Quantos participavam?>

2 – Experiência na disciplina DSI

E aí, o que achou do debate online que seu grupo realizou, foi boa a discussão?

3 – Finalidades educacionais do bate-papo

O que você aprendeu durante o debate online?

Muitos acham que não dá para aprender pelo bate-papo, pois é tudo muito corrido e frenético... o que você acha?

Além de discutir os conteúdos disciplinares, você acha que o bate-papo teria alguma outra finalidade para a educação a distância?

4 – Tamanho do grupo

No seu grupo, participaram XX pessoas. Você achou que foram muitas pessoas, poucas pessoas ou estava bom?

5 – feedback da aula do Marcelo

Ao final de cada aula, o Pimentel sempre pede o feedback... Mas como a aula do Marcelo foi online, sem o Pimentel, isso não foi possível. Sobrou pra mim fazer esse levantamento... Vamos começar falando dos pontos negativos sobre a aula. O que você não gostou e que o Marcelo poderia melhorar numa próxima aula online?

Apêndice 2 – Tabela de parâmetros para cálculo da taxa de ocupação da turma no modelo Escrita-Leitura

Tabela 14 – Tabela de parâmetros para cálculo da taxa de ocupação da turma no modelo Escrita-Leitura

Turma	Sessão de bate-papo na turma	Quantidade de mensagens na sessão de bate-papo	Quantidade de mensagens de debate (descartando mensagens da sessão antes e após debate, e as mensagens automáticas)	Média de caracteres / mensagem de debate	Média de duração dos debates (min)	Média de participantes nos debates	Tempo médio de leitura (min)	Tempo médio de escrita (min)	Estimativa do Limite Máximo de mensagens por participante	Estimativa do Limite Máximo de mensagens do grupo	Porcentagem do Limite Máximo atingido
1999.1	1	150	148	81,2	60,0	6,0	0,07	0,38	82,8	496,9	29,8
	2	257	226	78,7	60,0	9,5	0,07	0,37	64,0	604,8	37,4
	3	190	178	104,3	60,0	11,0	0,08	0,49	45,2	497,0	35,8
	5	281	257	67,2	60,0	10,9	0,06	0,31	66,0	721,0	35,6
	6	236	209	63,0	60,0	10,9	0,06	0,29	69,8	759,1	27,5
	7	233	216	77,8	60,0	10,6	0,07	0,36	59,6	633,0	34,1
	8	150	129	94,6	60,0	9,9	0,08	0,44	53,1	523,1	24,7
	9	230	205	77,8	60,0	9,0	0,07	0,36	66,8	601,1	34,1
	10	206	194	75,5	60,0	8,2	0,07	0,35	72,6	598,6	32,4

Média	-	214,8	195,8	80,0	60,0	9,6	0,07	0,37	64,4	603,8	32,4
DP	-	42,5	37,1	12,0	0,0	1,6	0,01	0,1	10,4	87,0	4,0
1999.2	1	367	286	50,4	60,0	7,0	0,05	0,24	113,8	790,8	36,2
	2	344	283	47,7	60,0	7,7	0,05	0,22	110,9	857,0	33,0
	3	325	298	49,9	60,0	11,1	0,05	0,23	82,9	917,4	32,5
	4	418	377	49,5	60,0	10,9	0,05	0,23	84,5	918,9	41,0
	5	582	412	44,3	60,0	9,3	0,05	0,21	103,1	960,2	42,9
Média	-	407,2	331,2	48,4	60,0	9,2	0,05	0,23	99,0	888,9	37,1
DP	-	92,8	53,1	2,2	0,0	1,6	0,00	0,01	13,0	59,1	4,2
2000.1	1	310	273	55,5	60,0	8,8	0,05	0,26	90,0	790,1	34,6
	2	304	281	53,0	60,0	7,1	0,05	0,25	108,2	762,7	36,8
	3	276	252	56,3	60,0	5,0	0,05	0,26	126,5	632,3	39,9
	4	342	317	57,4	60,0	5,9	0,05	0,27	112,7	668,4	47,4
	5	348	339	55,4	60,0	7,7	0,05	0,26	98,2	759,3	44,6
	6	368	331	54,2	60,0	8,0	0,05	0,25	98,1	780,6	42,4
	7	356	336	59,9	60,0	8,0	0,06	0,28	90,0	720,3	46,6
	8	415	367	55,8	60,0	7,9	0,05	0,26	96,4	759,7	48,3
	9	351	345	49,0	60,0	6,7	0,05	0,23	119,1	799,1	43,2
	10	464	380	44,8	60,0	7,9	0,05	0,21	114,6	907,6	41,9
	11	375	360	57,1	60,0	8,0	0,05	0,27	93,7	749,3	48,0
	12	402	377	47,5	60,0	6,2	0,05	0,22	128,2	797,3	47,3
	13	245	218	55,6	60,0	4,2	0,05	0,26	140,0	591,0	36,9
Média	-	350,5	321,2	54,0	60,0	7,0	0,05	0,25	108,9	747,5	42,9
DP	-	56,3	48,6	4,2	0,0	1,3	0,00	0,02	15,5	77,9	4,5
2001.1	1	367	339	56,6	60,0	16,65	0,05	0,26	54,8	911,9	37,2
	2	426	352	55,3	60,0	16,75	0,05	0,26	55,5	929,3	37,9
	3	378	333	52,0	60,0	15,81	0,05	0,24	60,9	962,9	34,6

	4	346	264	47,1	60,0	14,05	0,05	0,22	72,1	1012,6	26,1
	5	399	359	47,9	60,0	13,8	0,05	0,22	72,2	996,2	36,0
	6	156	135	53,0	60,0	12,75	0,05	0,25	71,1	906,4	14,9
	7	276	237	64,4	60,0	11,74	0,06	0,30	64,8	761,1	31,1
Média	-	335,4	288,4	53,8	60,0	14,5	0,05	0,25	64,5	925,8	31,1
DP	-	85,3	75,9	5,4	0,0	1,8	0,0	0,0	7,1	77,1	7,6
2001.2	1	417	300	75,1	60,0	16,0	0,07	0,35	45,3	722,8	41,5
	2	526	459	70,9	60,0	16,1	0,06	0,33	47,2	758,2	60,5
	3	538	463	74,9	60,0	18,0	0,07	0,35	41,3	742,2	62,4
	7	529	425	66,2	60,0	16,0	0,06	0,31	50,0	800,4	53,1
	5	573	522	77,9	60,0	14,6	0,07	0,36	47,0	687,8	75,9
	6	560	540	77,9	60,0	14,9	0,07	0,36	46,4	690,2	78,2
	7	574	488	75,7	60,0	15,9	0,07	0,35	45,2	716,9	68,1
Média	-	531,0	456,7	74,1	60,0	15,9	0,06	0,35	46,1	731,2	62,8
DP	-	50,0	73,5	3,9	0,0	1,0	0,00	0,02	2,5	36,8	11,9
2002.1	1	528	515	56,1	60,0	14,5	0,05	0,26	61,6	892,7	57,7
	2	504	494	71,9	60,0	19,7	0,06	0,34	39,6	780,4	63,3
	3	692	683	59,2	60,0	18,4	0,05	0,28	48,8	897,3	76,1
	4	609	598	58,9	60,0	18,1	0,05	0,27	49,6	898,6	66,6
	5	699	675	65,3	60,0	18,9	0,06	0,30	44,2	835,7	80,8
	6	641	622	39,1	60,0	16,8	0,04	0,18	71,1	1197,4	51,9
	7	582	561	54,0	60,0	20,4	0,05	0,25	48,0	979,4	57,3
	8	607	583	64,4	60,0	15,8	0,06	0,30	51,7	815,9	71,5
Média	-	607,8	591,4	58,6	60,0	17,8	0,05	0,27	51,8	912,2	65,6
DP	-	65,5	63,9	9,1	0,0	1,9	0,0	0,0	9,4	122,0	9,3
2002.2	1	399	374	82,2	60,0	11,0	0,07	0,38	55,5	610,4	61,3
	2	517	489	67,3	60,0	11,0	0,06	0,31	65,8	720,7	67,9

	3	440	438	81,1	60,0	11,0	0,07	0,38	56,1	617,4	70,9
	4	462	458	75,1	60,0	11,0	0,07	0,35	59,9	658,7	69,5
	5	313	272	88,8	46,4	11,0	0,07	0,41	52,0	571,5	47,6
	6	200	134	93,5	24,1	10,0	0,08	0,44	53,1	530,8	25,2
	7	352	323	83,5	50,9	8,6	0,07	0,39	64,7	557,3	58,0
	8	384	332	78,4	51,5	9,8	0,07	0,37	62,5	614,1	54,1
Média	-	383,4	352,5	81,2	51,6	10,4	0,07	0,38	58,7	610,1	56,8
DP	-	91,6	107,8	7,5	11,5	0,8	0,00	0,04	5,0	56,2	14,1
2003.1	1	407	393	80,0	60,0	16,9	0,07	0,37	41,1	694,5	56,6
	2	455	357	87,2	56,8	13,2	0,07	0,41	46,1	610,8	58,4
	3	446	415	84,9	55,1	16,0	0,07	0,40	40,8	653,5	63,5
	4	507	495	74,8	57,9	17,8	0,06	0,35	41,6	741,8	66,7
	5	593	566	81,2	56,0	17,8	0,07	0,38	39,0	693,4	81,6
	7	763	596	80,2	53,1	16,0	0,07	0,37	42,8	685,1	87,0
	8	524	501	92,8	55,2	15,9	0,08	0,43	38,2	605,2	82,8
Média	-	527,9	474,7	83,0	56,3	16,2	0,07	0,39	41,4	669,2	71,0
DP	-	111,2	83,0	5,4	2,1	1,4	0,00	0,03	2,4	45,5	11,6
2003.2	1	293	249	65,6	57,4	9,6	0,06	0,31	73,8	708,0	35,2
	2	480	339	72,1	58,2	11,5	0,06	0,34	60,0	690,0	49,1
	3	504	388	62,7	57,9	10,2	0,06	0,29	73,4	748,3	51,8
	4	608	411	67,3	57,3	11,4	0,06	0,31	63,9	728,4	56,4
	5	597	380	73,1	49,6	11,9	0,06	0,34	57,9	688,5	55,2
	6	646	478	72,4	51,7	12,7	0,06	0,34	55,5	706,1	67,7
	7	387	287	71,6	54,2	7,3	0,06	0,33	81,9	601,0	47,8
	8	482	318	71,4	42,8	10,9	0,06	0,33	63,0	684,9	46,4
Média	-	499,6	356,3	69,5	53,6	10,7	0,06	0,32	66,2	694,4	51,2
DP	-	111,1	68,4	3,6	5,0	1,6	0,00	0,02	8,6	40,7	8,7

2004.1	1	202	165	108,2	53,5	7,6	0,09	0,50	55,6	424,1	38,9
	2	258	188	85,7	53,4	7,4	0,07	0,40	69,7	515,7	36,5
	3	268	266	113,8	54,6	10,8	0,09	0,53	42,3	457,7	58,1
	4	299	296	116,4	57,2	11,6	0,09	0,54	39,5	458,4	64,6
	5	227	222	85,3	50,3	6,6	0,07	0,40	74,7	496,2	44,7
	6	261	239	98,3	55,4	8,8	0,08	0,46	55,4	486,5	49,1
	7	269	248	96,8	53,6	10,0	0,08	0,45	51,5	515,0	48,2
	8	209	202	105,8	47,2	8,0	0,09	0,49	55,1	440,6	45,8
Média	-	249,1	228,3	101,3	53,1	8,9	0,08	0,47	55,5	474,3	48,2
DP	-	31,2	40,1	11,1	2,9	1,7	0,01	0,05	11,3	32,0	8,7
2004.2	1	269	241	88,2	59,1	7,9	0,07	0,41	65,2	515,3	46,8
	2	331	203	87,8	48,9	8,7	0,07	0,41	61,7	534,8	38,0
	3	312	277	88,1	57,8	7,0	0,07	0,41	70,3	492,8	56,2
	4	436	318	73,6	55,2	6,9	0,06	0,34	83,2	573,0	55,5
	5	493	303	67,9	53,2	4,9	0,06	0,32	108,4	532,4	56,9
	6	471	322	78,9	55,2	6,0	0,07	0,37	85,0	509,7	63,2
	7	522	299	81,4	51,3	7,8	0,07	0,38	70,4	550,6	54,3
	8	648	308	76,9	52,8	7,0	0,07	0,36	79,3	555,0	55,5
Média	-	435,3	283,9	80,4	54,2	7,0	0,07	0,37	77,9	533,0	53,3
DP	-	117,8	39,1	7,0	3,1	1,1	0,00	0,03	13,9	24,7	7,1
2005.1	1	397	376	96,8	58,3	13,7	0,08	0,45	41,2	563,5	66,7
	2	376	354	100,2	56,3	14,0	0,08	0,47	39,3	550,2	64,3
	3	438	402	91,9	59,2	12,8	0,08	0,43	45,3	578,7	69,5
	4	336	303	96,0	58,5	10,0	0,08	0,45	51,9	518,8	58,4
	5	274	249	113,2	60,0	9,0	0,09	0,53	48,1	432,8	57,5
	6	341	323	88,5	57,6	10,0	0,07	0,41	55,8	556,0	58,1
	7	322	276	89,1	56,8	8,9	0,07	0,42	59,8	532,9	51,8

	8	362	310	95,1	53,9	10,3	0,08	0,44	51,2	528,3	58,7
Média	-	355,8	324,1	96,4	57,6	11,1	0,08	0,45	49,1	532,7	60,6
DP	-	46,6	47,8	7,4	1,8	1,9	0,00	0,03	6,6	42,0	5,4
2005.2	3	290	241	92,4	57,0	6,9	0,08	0,43	68,1	469,2	51,4
	4	224	174	76,8	57,3	5,0	0,07	0,36	96,2	481,2	36,2
	5	302	246	71,3	58,2	5,0	0,06	0,33	102,8	514,2	47,8
	6	297	248	60,4	59,6	3,0	0,06	0,28	152,7	458,0	54,2
	7	292	235	53,8	52,3	4,0	0,05	0,25	147,8	594,1	39,6
	8	272	141	90,5	40,5	4,7	0,08	0,42	85,6	403,0	35,0
Média	-	279,5	214,2	74,2	54,1	4,8	0,06	0,35	108,9	486,6	44,0
DP	-	26,5	41,4	14,3	6,5	1,2	0,01	0,07	31,2	58,4	7,5
2006.1	1	276	270	90,2	54,2	8,3	0,07	0,42	62,2	513,3	52,6
	2	274	270	104,1	60,7	9,9	0,08	0,49	48,7	481,1	56,1
	3	300	290	94,4	57,5	10,9	0,08	0,44	49,5	541,2	53,6
	4	255	255	96,2	53,0	9,0	0,08	0,45	55,6	500,1	51,0
	5	288	275	98,8	53,0	8,5	0,08	0,46	56,3	478,8	57,4
	6	242	240	120,5	60,3	8,4	0,09	0,56	47,6	398,6	60,2
	7	283	281	99,5	61,1	10,0	0,08	0,46	50,3	502,8	55,9
	8	221	212	103,5	55,3	10,3	0,08	0,48	47,5	490,9	43,2
Média	-	267,4	261,6	100,9	56,9	9,4	0,08	0,47	52,2	488,3	53,8
DP	-	24,5	23,6	8,6	3,2	0,9	0,01	0,04	4,9	38,7	4,8
2006.2	1	365	348	80,7	59,7	8,1	0,07	0,38	69,4	561,6	62,0
	2	329	321	80,9	59,1	8,0	0,07	0,38	69,8	557,9	57,5
	3	331	316	96,5	56,8	8,9	0,08	0,45	55,7	497,6	63,5
	4	298	291	70,3	57,9	4,9	0,06	0,33	105,4	515,3	56,5
	5	406	382	62,9	58,2	7,6	0,06	0,29	89,2	680,0	56,2
	6	366	351	61,7	57,2	6,0	0,06	0,29	105,6	630,2	55,7

	7	329	324	81,7	59,4	5,7	0,07	0,38	84,8	483,3	67,0
	8	339	275	75,2	54,5	6,9	0,07	0,35	81,3	563,7	48,8
Média	-	345,4	326,0	76,2	57,8	7,0	0,07	0,36	82,7	561,2	58,4
DP	-	30,6	32,0	10,7	1,6	1,3	0,01	0,05	16,4	62,4	5,3
2007.1	1	268	194	116,4	59,4	7,9	0,09	0,54	50,9	402,4	48,2
	2	250	236	92,5	59,8	8,6	0,08	0,43	59,4	508,7	46,4
	3	323	303	102,2	58,2	9,7	0,08	0,48	50,3	485,6	62,4
	4	326	271	105,2	59,6	9,4	0,08	0,49	49,9	469,2	57,8
	5	223	222	90,5	60,0	5,9	0,08	0,42	75,7	448,8	49,5
	6	278	265	88,9	51,1	8,0	0,07	0,41	64,3	514,1	51,5
	7	251	241	89,4	58,7	7,9	0,07	0,42	64,6	508,7	47,4
	8	235	227	67,4	58,3	5,7	0,06	0,31	101,0	570,5	39,8
Média	-	269,3	244,9	94,1	58,1	7,9	0,08	0,44	64,5	488,5	50,4
DP	-	35,7	31,6	13,5	2,7	1,4	0,0	0,1	16,1	46,9	6,5
2007.2	1	305	250	106,8	53,9	11,0	0,09	0,50	44,3	486,5	51,4
	2	388	233	62,7	42,1	11,2	0,06	0,29	68,6	767,9	30,3
	3	278	275	88,8	56,7	12,0	0,07	0,41	48,8	585,8	46,9
	4	308	250	94,5	48,7	11,7	0,08	0,44	47,2	550,9	45,4
	5	247	201	91,2	52,7	6,0	0,08	0,43	74,7	447,9	44,9
	6	283	276	67,8	60,0	7,8	0,06	0,32	82,8	642,4	43,0
	7	232	197	67,4	31,2	11,8	0,06	0,31	62,3	734,0	26,8
	8	295	273	78,0	59,2	12,8	0,07	0,36	52,0	664,5	41,1
Média	-	292,0	244,4	82,2	50,6	10,5	0,07	0,38	60,1	610,0	41,2
DP	-	44,1	29,7	14,6	9,1	2,2	0,01	0,07	13,3	106,0	7,9
2008.1	1	373	302	73,4	54,9	7,7	0,06	0,34	77,6	599,6	50,4
	2	330	245	89,0	55,2	6,5	0,07	0,42	72,6	474,9	51,6
	3	288	251	88,1	55,4	6,7	0,07	0,41	72,1	484,5	51,8

	4	298	245	95,8	59,6	5,2	0,08	0,45	77,6	399,6	61,3
	5	301	265	77,5	59,6	5,5	0,07	0,36	90,6	499,0	53,1
	6	233	197	79,6	58,0	4,9	0,07	0,37	94,2	461,6	42,7
	7	321	280	83,4	58,3	6,9	0,07	0,39	74,8	512,5	54,6
	8	442	410	60,8	61,1	6,6	0,06	0,28	100,8	663,1	61,8
Média	-	323,3	274,4	81,0	57,7	6,2	0,07	0,38	82,5	511,9	53,4
DP	-	58,3	58,7	10,1	2,2	0,9	0,0	0,0	10,3	77,4	5,8
MÉDIA DAS TURMAS	-	370,8	320,5	77,2	56,8	10,2	0,07	0,36	68,6	637,2	50,8
DESVIO PADRÃO	-	127,6	112,8	17,8	5,3	3,8	0,01	0,08	23,8	156,0	13,3