



Universidade de Brasília
IE – Departamento de Estatística

O Impacto da Atenção Visual nas Redes Sociais

Maria de Fátima Lira de Carvalho

Junho de 2014

Maria de Fátima Lira de Carvalho

O Impacto da Atenção Visual nas Redes Sociais

Relatório apresentado à disciplina Estágio Supervisionado II do curso de graduação em Estatística, Instituto de Ciências Exatas, Universidade de Brasília, como parte dos requisitos necessários para o grau de Bacharel em Estatística.

Orientador: Prof. Eduardo Yoshio Nakano

Junho de 2014

Dedico este trabalho aos meus pais, por acreditarem em mim; aos meus professores, pela dedicação e paciência; a UnB, por ser tão acolhedora e ter se tornado minha segunda casa; a minha irmã e meus amigos, por serem o motivo da minha alegria.

Maria de Fátima Lira de Carvalho

Agradecimentos

Agradeço primeiramente a Deus, que sempre me iluminou e me ajudou a concluir mais essa etapa em minha vida.

Ao professor Eduardo Nakano, pela orientação e dedicação e inestimável auxílio com este trabalho, foi uma honra ter seus conhecimentos e ideias transmitidas. A todos os professores do departamento de Estatística, que me passaram com carinho, os seus conhecimentos e sabedoria.

Aos meus amigos, Roberto Lazarte, Lucas Santos, Eduardo Augusto, Ana Luiza, Lucas Lima, Laura Teixeira e Eduarda Leão que me ajudaram na programação, design e conteúdo deste trabalho.

Resumo

O aumento da popularidade das redes sociais tem causado um impacto muito grande na forma com que os indivíduos se relacionam. Entender esse relacionamento é essencial para compreender o comportamento humano. Portanto, diversos estudos de atenção visual nas redes sociais têm sido feitos.

Visando entender melhor a atenção visual, foi feito o estudo - “Um Estudo Sobre o Impacto da Atenção Visual nas Redes Sociais” por Daniele V. Ferreira e Genáina Rodrigues e Wellington Y.L. Akamine e Mylene C. Q. Farias. O estudo usou algoritmos computacionais que detectam possíveis regiões atrativas para o olho humano e, através do algoritmo, foram criados mapas de atenção visual em 240 interfaces de quatro redes sociais: Facebook, Orkut, Twitter e LinkedIn.

Através dos mapas de atenção visual é possível calcular regiões de atratividade nas interfaces e saber quais áreas são as mais atrativas e tem maior navegação pelos usuários. Foram definidos scores para calcular uma média de atratividade de cada região. Dentro de um contexto bayesiano, usando distribuições a priori, é possível encontrar um estimador para a média de atratividade através de uma distribuição à posteriori.

Encontrados os estimadores é possível concluir as regiões das interfaces que tem maior atratividade pelos usuários e também comparar o estudo, com o semelhante, feito por Benevenuto et al. (2009), sobre o impacto na atenção visual nas redes sociais.

Lista de Figuras

Figura 3.1 - Imagens com seus respectivos mapas de saliência, gerados pelo GAFFE	7
Figura 5.1 - Imagens com seus respectivos mapas de saliência, gerados pelo GAFFE	9
Figura 6.1 - Regiões de atividade do Facebook.....	13
Figura 6.2 - Região de atividade do Twitter	14
Figura 6.3 - Regiões de atividade do Orkut.....	15
Figura 6.4 - Regiões de atividade do LinkedIn	16
Figura 8.1 - Densidades da atratividade das Regiões do Orkut	25
Figura 8.1 - Densidades da atratividade das Regiões do Facebook	26
Figura 8.3 - Densidades da atratividade das Regiões do Twitter	28
Figura 8.3 - Densidades da atratividade das Regiões do LinkedIn	29

Lista de Tabelas

Tabela 4.1. Resultados da atratividade das Redes Sociais Orkut, My Space, LinkedIn e Hi5.....	8
Tabela 8.1. Resultados da atratividade da Rede Social Orkut.....	21
Tabela 8.2. Resultados da atratividade da Rede Social Facebook.....	22
Tabela 8.3 Resultados da atratividade da Rede Social LinkedIn.....	22
Tabela 8.4 Resultados da atratividade da Rede Social Twitter	23
Tabela 8.5. Resultados da estimativa da atratividade da Rede Social Orkut.....	24
Tabela 8.6. Resultados da estimativa da atratividade da Rede Social Facebook	26
Tabela 8.7 Resultados da estimativa da atratividade da Rede Social Twitter.....	27
Tabela 8.8 Resultados da estimativa da atratividade da Rede Social LinkedIn	29

Sumário

1. Introdução.....	1
2. Redes Sociais.....	3
2.1. As redes sociais mais populares.....	4
3. Atenção Visual.....	6
3.1. Mapas de Saliência.....	6
4. Trabalhos Relacionados.....	6
5. Atratividade das Regiões.....	9
6. Regiões de Atividade.....	12
6.1. Regiões de Atividade do Facebook.....	13
6.2. Regiões de Atividade do Twitter.....	14
6.3. Regiões de Atividade do Orkut.....	15
6.4. Regiões de Atividade do LinkedIn.....	16
7. Inferência da Atratividade.....	17
8. Resultados.....	21
9. Conclusões.....	30
10. Bibliografia.....	32
Apêndice.....	33

1. Introdução

Com a expansão da tecnologia e o fácil acesso à internet, as redes sociais se tornaram muito populares. Além de possibilitar o relacionamento e interação entre as pessoas, têm influenciado bastante em seus comportamentos, ocorrendo uma mudança significativa do modo com que as pessoas se comunicam, se relacionam e até mesmo consomem produtos. Portanto, diversas pesquisas têm sido feitas a fim de entender o comportamento dos usuários e melhorar o layout das redes sociais.

O olho humano, instintivamente tende a se concentrar em regiões com grande quantidade de informação. Com base nisso, é possível construir mapas de saliência, que são feitos através de modelos computacionais, para as interfaces das redes sociais, ou seja, podem-se criar imagens de interfaces, destacando as regiões onde a atenção visual se manteve por mais tempo.

A fim de verificar como características intrínsecas dos objetos na cena (cor, formato, posição, etc.) podem motivar interações em redes sociais, o algoritmo GAFFE (Rajashekar et al., 2008) foi usado para construir mapas de saliência no estudo “Um Estudo Sobre o Impacto da Atenção Visual nas Redes Sociais” por Daniele V. Ferreira e Genáina Rodrigues do departamento de ciência da computação da UnB e Wellington Y.L. Akamine e Mylene C. Q. Farias do departamento de engenharia elétrica da UnB. Para cada imagem, os modelos computacionais geram os mapas de saliência, que são imagens em escala cinza. Para o estudo, foram usadas 80 interfaces, ou seja, perfis de usuários, das redes sociais Facebook e do Orkut e 40 das redes sociais Twitter e do LinkedIn.

O seguinte estudo busca calcular, através dos mapas de saliência construídos no estudo anterior, dentro de um contexto bayesiano, a atratividade das regiões de interface das redes sociais Orkut, Facebook, LinkedIn e Twitter e definir um score para cada região. O score será transformado em proporções dentro de cada região, desta forma, cada unidade amostral consistirá em um vetor de proporções cuja soma será 1. As estimativas serão obtidas dentro de um contexto bayesiano utilizando-se o modelo de Dirichlet (Balakrishnan, 2005). Os resultados serão comparados com o estudo de Benevenuto et al. (2009) sobre a atenção visual nas redes sociais.

2. Redes Sociais

Redes sociais se referem a um conjunto de pessoas ou organizações que se relacionam profissional ou efetivamente e que, de alguma forma compartilham interesses em comum. Segundo Marteleto, R. M. (2001) em *Análise de redes sociais: aplicação nos estudos de transferência da informação*, 2001, redes sociais representam um conjunto de participantes autônomos, unindo ideias e recursos em torno de valores e interesses compartilhados.

A tecnologia possibilitou o crescimento desenfreado das redes sociais, graças ao fácil acesso à internet. Além de possibilitar o relacionamento afetivo e profissional entre os usuários, as redes influenciam no comportamento das pessoas, na forma em que elas se comunicam e, inclusive, na forma que consomem produtos.

Entender o modo como os usuários interagem nas redes sociais é de fundamental importância para a melhoria das interfaces dessas redes, além de promover diversos estudos sobre interações sociais e distribuição de conteúdo, pois as redes se tornaram uma poderosa ferramenta de disseminação dos mais diversos conhecimentos, fundamentando diversos tipos de pesquisa.

As dinâmicas das redes sociais se desenvolvem através dos sites, que possuem os mais diversos tipos de layout. Esses visam destacar os objetivos que cada rede tem, seja troca de mensagem, visualização do perfil, comunidades, entre outros. A busca pela melhor interface é constante, e envolvem várias questões, como design, arranjo, cores, fontes, entre outros.

2.1 As redes sociais mais populares

Um estudo da Comscore (empresa de pesquisa de mercado dos Estados Unidos, que fornece dados de marketing e serviços para as maiores empresas da Internet), feito em 2012, mostrou um crescimento no número de usuários nas redes sociais e apontou as principais redes do Brasil entre os anos de 2010 e 2011, essas redes foram Facebook, Orkut, LinkedIn e Twitter.

- **Facebook**

Criada em 2004, a rede permite os usuários ou organizações criarem suas próprias páginas mostrando fotos, interesses e atualizações. Possibilita a troca de mensagens privada e pública entre os participantes da rede. Possui ferramentas mais elaboradas e detalhadas, como pesquisar usuários por cidade, ou local de trabalho ou estudo. O maior enfoque é a divulgação das atualizações de cadastrados. O Facebook se popularizou bastante nos últimos anos, como comparação, se essa rede fosse um país, seria o terceiro mais populoso do mundo, com 500 milhões de usuários registrados.

- **Orkut**

Criada em 2004 com o objetivo de ajudar os membros a conhecer pessoas e manter relacionamentos. Seu funcionamento se baseia em perfil, com fotos e informações, além das comunidades, que participam os usuários com interesses em comum. No início de sua implantação só era possível participar da rede quem recebesse o convite de um membro já ativo da rede, hoje em dia o cadastramento é livre. A maior parte dos

usuários é do Brasil e da Índia. O Orkut tem mais de 33 milhões de membros ativos no mundo.

- **LinkedIn**

Lançada em 2003, é principalmente usada por profissionais. Permite que os usuários possam manter uma lista detalhada de contatos de pessoas que conheça e em quem confiem nas empresas. A lista é chamada de conexão. Útil para HeadHunters (caça talentos de empresas) e analistas de RH, pois, com sua utilização é possível selecionar candidatos a empregos ou estágios. A rede conta com mais de 33,9 milhões de membros ativos.

- **Twitter**

Lançada em 2006, a rede permite aos usuários enviar e receber atualizações pessoais de outros contatos, de maneira dinâmica, chamadas de *tweets*. As atualizações são exibidas no perfil de um usuário em tempo real e enviadas a outros membros seguidores que tenham assinado para recebê-las. A ideia inicial era que o Twitter seria um “SMS da internet”, com a limitação de caracteres de uma mensagem de celular. Possui uma interface simplificada e os textos postados têm no máximo 140 caracteres. O Twitter conta com uma média de 200 milhões de usuários ativos mensais.

3. Atenção Visual

A atenção visual é uma característica do sistema ocular humano, que tem por objetivo reduzir a complexidade dos dados a serem processados e maximizar a quantidade de informações úteis a serem assimiladas em uma determinada cena. O olho humano instintivamente tende a se concentrar em regiões com grande quantidade de informação e “filtrar” o que considera mais importante.

A atenção visual envolve mecanismos utilizados para definir características presentes no campo visual de cada ser, pois o olho humano tende a selecionar a focar seu campo de visão em determinados aspectos no ambiente que o cerca. As informações consideradas importantes são concentradas em regiões chamadas saliências. As regiões salientes são mais visíveis que outras regiões.

3.1 Mapas de saliência

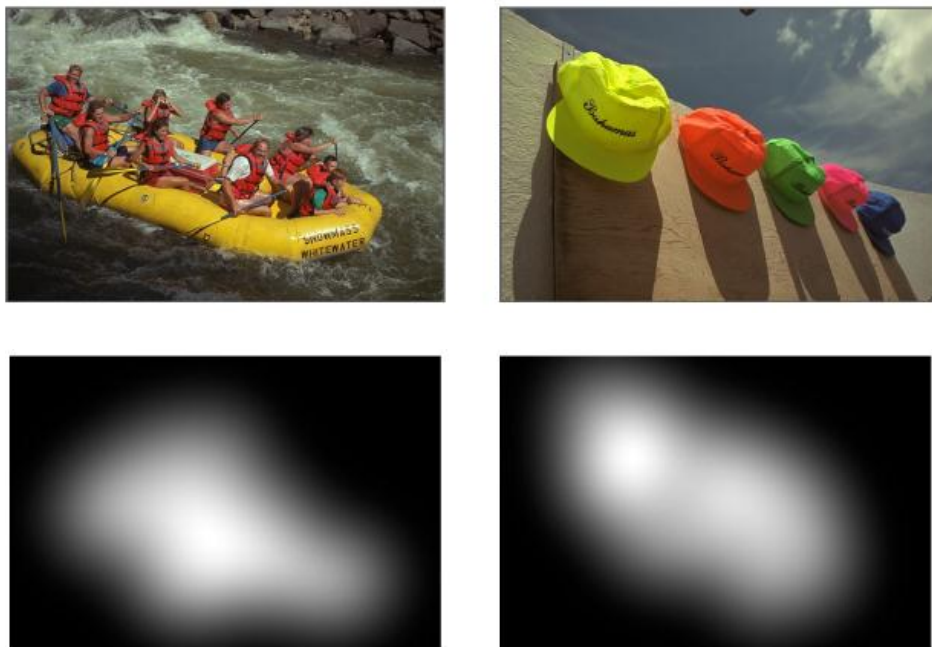
Com base na atenção visual é possível construir mapas de saliência, que são feitos através de modelos computacionais, para as interfaces das redes sociais, ou seja, podem-se criar imagens de interfaces, destacando as regiões onde a atenção visual se manteve por mais tempo.

Entre os mais populares algoritmos tem-se o GAFFE (Gaze-Attentive Fixation Finding Engine), proposto por Rajashekar et al, 2008. Esse algoritmo foi desenvolvido através de um experimento com 29 observadores, enquanto eles assistiam a 101 imagens. A partir desse estudo foi feita a estatística de quatro características locais de baixo nível desse conjunto de imagens: Luminância,

contraste e resposta passa-faixa da luminância e contraste. Descobriu-se que retalhos de imagens em torno das fixações oculares tinham valores maiores para cada uma destas características do que os valores correspondentes a retalhos da imagem em regiões aleatórias das imagens. O algoritmo recebe como entrada uma imagem e, em sua primeira etapa, decompõe ela nas características primitivas. Cada característica será geradora de um mapa que, por sua vez, gerará uma sequência crescente de imagens.

O GAFFE gera imagens em escala cinza, a partir de outras imagens. As regiões mais atrativas da imagem ficam numa escala mais clara, e as regiões onde a atenção menos se concentrou, em uma escala mais escura. Na imagem abaixo se observa duas fotos com seus respectivos mapas de saliência. Nota-se que os objetos em destaque, com cores mais fortes, têm uma atratividade maior.

Figura 3.1. Imagens com seus respectivos mapas de saliência, gerados pelo GAFFE



4. Trabalhos Relacionados

Em um trabalho anterior, Benevenuto et al. (2009) realizaram um estudo do comportamento dos usuários em redes sociais e verificaram que é possível identificar padrões nas interações nesses comportamentos. Foram analisadas as seguintes redes sociais: Orkut, MySpace, LinkedIn e Hi5. Foi usado o método *clickstream*, que permite que, além das interações visíveis, sejam capturadas interações consideradas “silenciosas”, como navegação do perfil, através de interações via dispositivo *mouse* e coletadas via um agregador de redes sociais. Portanto, cada clique dado pelo usuário em determinada região de atividade da rede, foi coletada pelo agregador de tráfego e foram analisadas as atividades das áreas. A tabela abaixo mostra alguns resultados do estudo.

Tabela 4.1. Resultados da atratividade das Redes Sociais Orkut, My Space, LinkedIn e Hi5

	Orkut	MySpace	LinkedIn	Hi5
Perfil e Amigos	41%	88%	51%	67%
Fotos	31%	3%	42%	18%
Scrap / Mensagens	20%	5%	4%	3%
Comunidades	4%	1%	<1%	6%
Busca / Outros	2%	1%	42%	4%

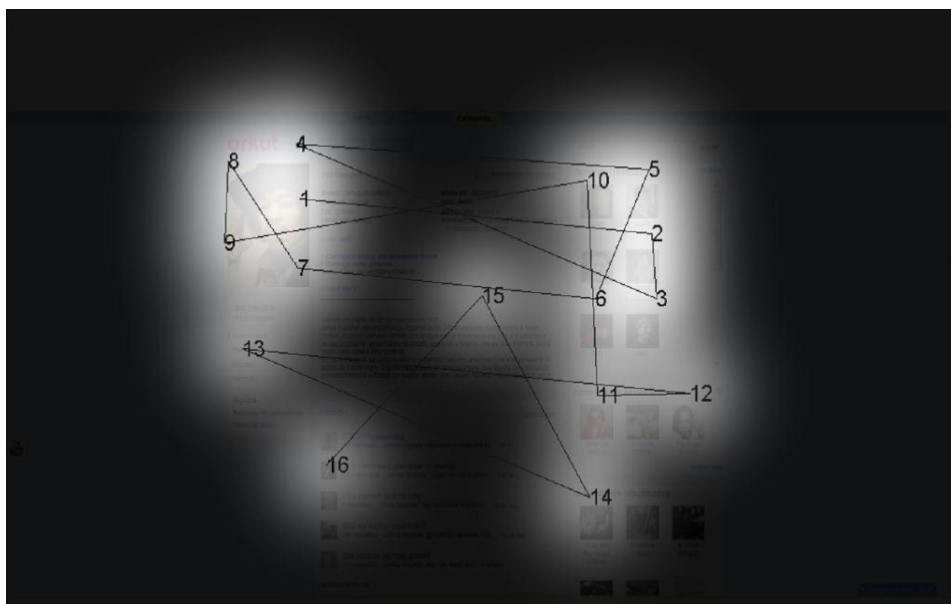
FONTE: Benevenuto et al (2009)

Observa-se que a atividade mais popular é navegar nos perfis dos amigos. Nota-se que a rede social LinkedIn apresentou um grau menor de interação entre os usuários por meio de scraps ou mensagens. Isso se deve ao fato de a rede ser principalmente para relacionamento profissional, portanto, é natural que os usuários primeiramente naveguem nos perfil e criem laços entre si, ao invés de trocarem mensagens. A rede social Hi5 se assemelhou ao Orkut e o MySpace obteve um score menor que as outras redes sociais na atividade “fotos”.

5. Atratividade das Regiões

Usando o algoritmo GAFFE foi possível construir os mapas de saliência no estudo “Um Estudo Sobre o Impacto da Atenção Visual nas Redes Sociais” por Daniele V. Ferreira e Genáina Rodrigues do departamento de ciência da computação da UnB e Wellington Y.L. Akamine e Mylene C. Q. Farias do departamento de engenharia elétrica da UnB. Foram construídos os mapas de 80 interfaces das redes sociais Facebook e do Orkut e 40 das redes sociais Twitter e do LinkedIn. As imagens possuíam sua interface sobreposta pelo mapa com a escala em cinza, e além disso, foi possível numerar de 1 a 16 as regiões mais atrativas, sendo 1 a mais atrativa e 16 a menos atrativa das 16 regiões. A Figura 5.1 abaixo mostra um dos mapas construídos com uma interface da rede social Orkut. Observa-se visualmente que as regiões mais atrativas foram a foto do perfil, no canto superior esquerdo e os participantes da comunidade, na parte superior direita.

Figura 5.1. Imagens com seus respectivos mapas de saliência, gerados pelo GAFFE



À fim de calcular um score de atratividade para cada região, para cada local numerado do mapa será atribuído um peso. Para adaptação em cálculos futuros, cada região receberá um valor inicial de 1, em seguida serão distribuídos em cada área, o seu determinado valor de atratividade, variando de 1 a 16. A área numerada como 1 (como na Figura 5.1), que é a mais atrativa da interface da rede social, terá o maior peso entre todas as área, ou seja, lhe será atribuído o score de 16, a área 2, que é a segunda região mais atrativa entre as 16 regiões, deverá ter o segundo maior peso dentre as 16 áreas, que é 15, a área 3 terá um score de 14, a área 4, um score de 13, e assim por diante, até a área 16, que será atribuído o menor score, que é igual a 1. Para determinar o escore padronizado, dividem-se todos os valores do score pelo peso total dos scores, mais o 1 inicial, que por soma de progressão aritmética é 136, mais a quantidade de áreas da região. No caso de um rede social dividida em 6 áreas de atividade, o peso total dos scores será 136 + 6, como mostrado abaixo:

- Peso total dos scores para rede social dividida em 6 regiões de atratividade:

$$16 + 15 + 14 + \dots + 1 + (6) = \frac{(16 + 1) \cdot 16}{2} + 6 = 142$$

- Peso total dos scores para rede social dividida em 5 regiões de atratividade:

$$16 + 15 + 14 + \dots + 1 + (5) = \frac{(16 + 1) \cdot 16}{2} + 5 = 141$$

Portanto, supondo que a foto do perfil foi somente a primeira e a terceira região de visualização de algum usuário, nessa região de atividade, estarão os números 1 e 3, que recebem o score de 16 e 14, respectivamente. Sabendo que esta foto do perfil pertence a uma rede social que foi dividida em 6 regiões de atividade, o score padronizado desta região de atividade será:

$$\frac{16}{142} + \frac{14}{142} \cong 0,21$$

Será calculado o score padronizado para todas as regiões de atividade de cada uma das quatro redes sociais utilizadas. As redes sociais Facebook, Twitter e Orkut possuem 6 regiões de atividade e a rede LinkedIn possui 5 regiões de atividade.

6. Regiões de Atividade

Para calcular as atratividades, cada interface das quatro redes sociais foi dividida em regiões de atividades consideradas relevantes. A divisão das regiões para o Facebook, Orkut, Twitter e LinkedIn se encontram abaixo.

6.1 Regiões de Atividade do Facebook

Para o Facebook a interface da rede foi dividida da seguinte forma:

Região 1 – Foto do Perfil

Região 2 – Foto de Capa

Região 3 – Anúncios, Busca e Outros

Região 4 – Amigos em Comum

Região 5 – Mural

Região 6 – Sobre Mim

Figura 6.1. Regiões de atividade do Facebook



6.2 Regiões de Atividade do Twitter

Para o Twitter a interface da rede foi dividida da seguinte forma:

Região 1 – Foto do Perfil

Região 2 – Busca

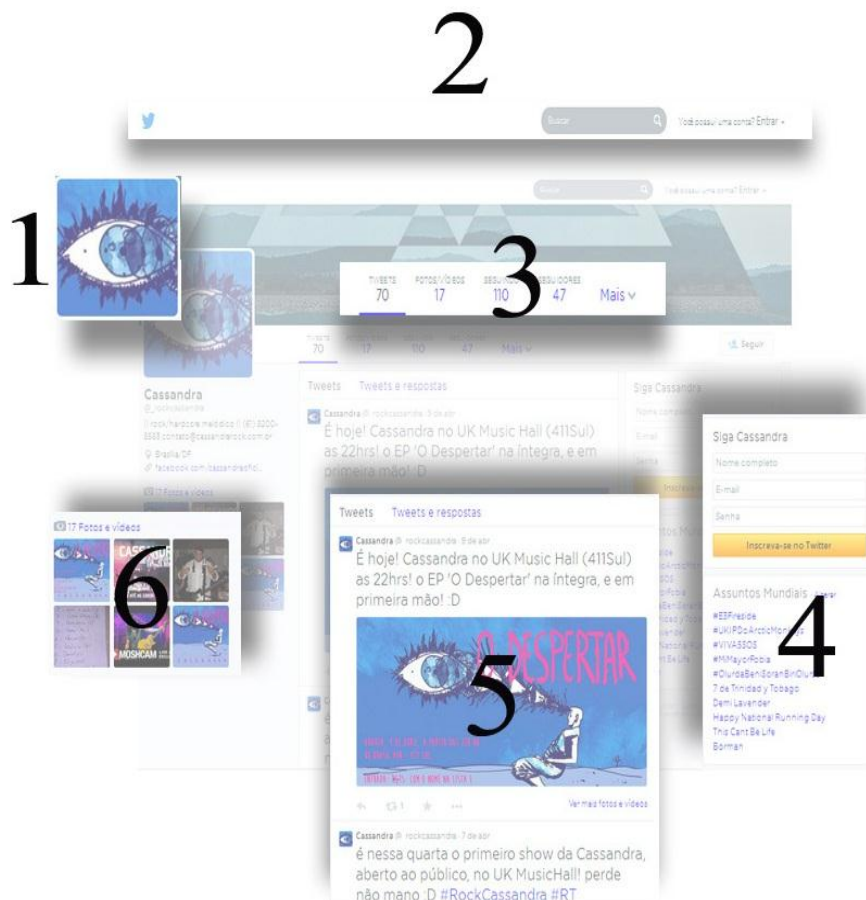
Região 3 – Informações de quantidade de twittes, seguidores e seguidos

Região 4 – Área de cadastro de e-mail

Região 5 – Mensagens twittadas

Região 6 – Fotos e Vídeos

Figura 6.2. Região de atividade do Twitter



6.3 Regiões de Atividade do Orkut

Para o Orkut a interface da rede foi dividida da seguinte forma:

Região 1 – Foto do Perfil

Região 2 – Sobre Mim

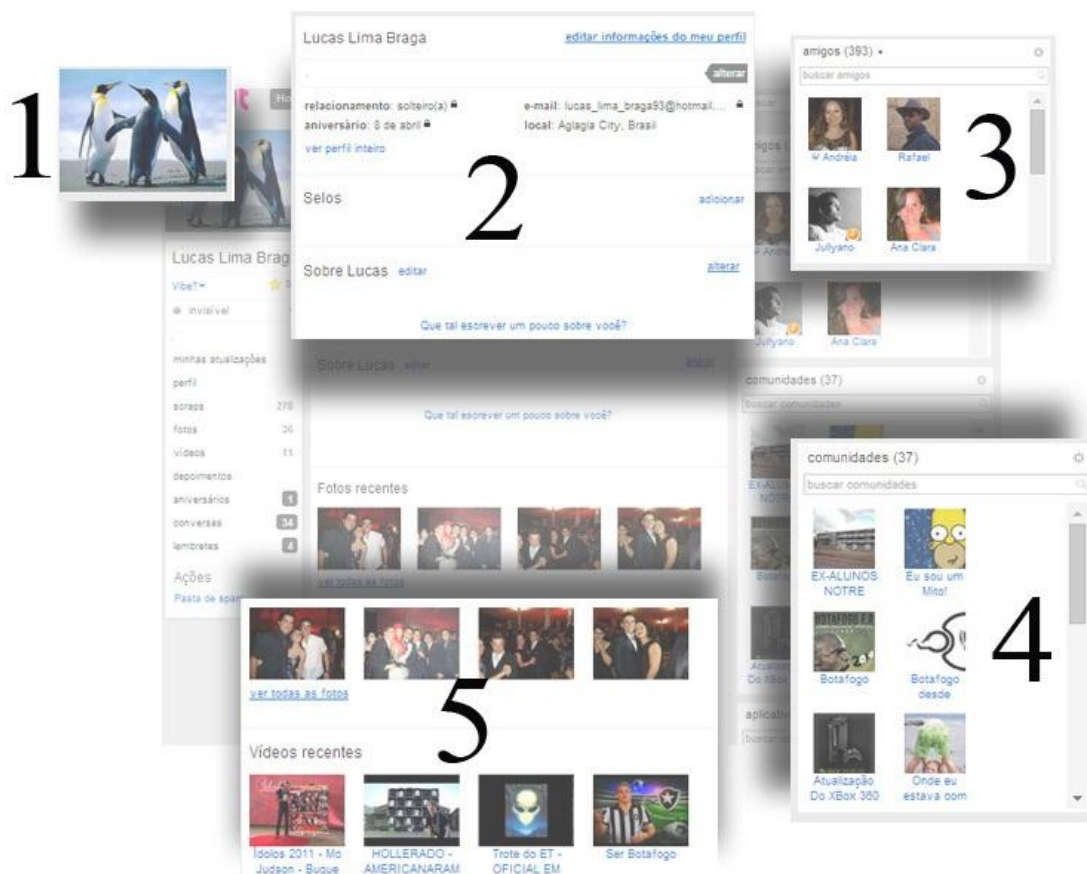
Região 3 – Amigos

Região 4 – Comunidades

Região 5 – Fotos e Vídeos

Região 6 – Outras regiões

Figura 6.3. Regiões de atividade do Orkut



6.4 Regiões de Atividade do LinkedIn

Para o LinkedIn a interface da rede foi dividida da seguinte forma:

Região 1 – Foto do Perfil

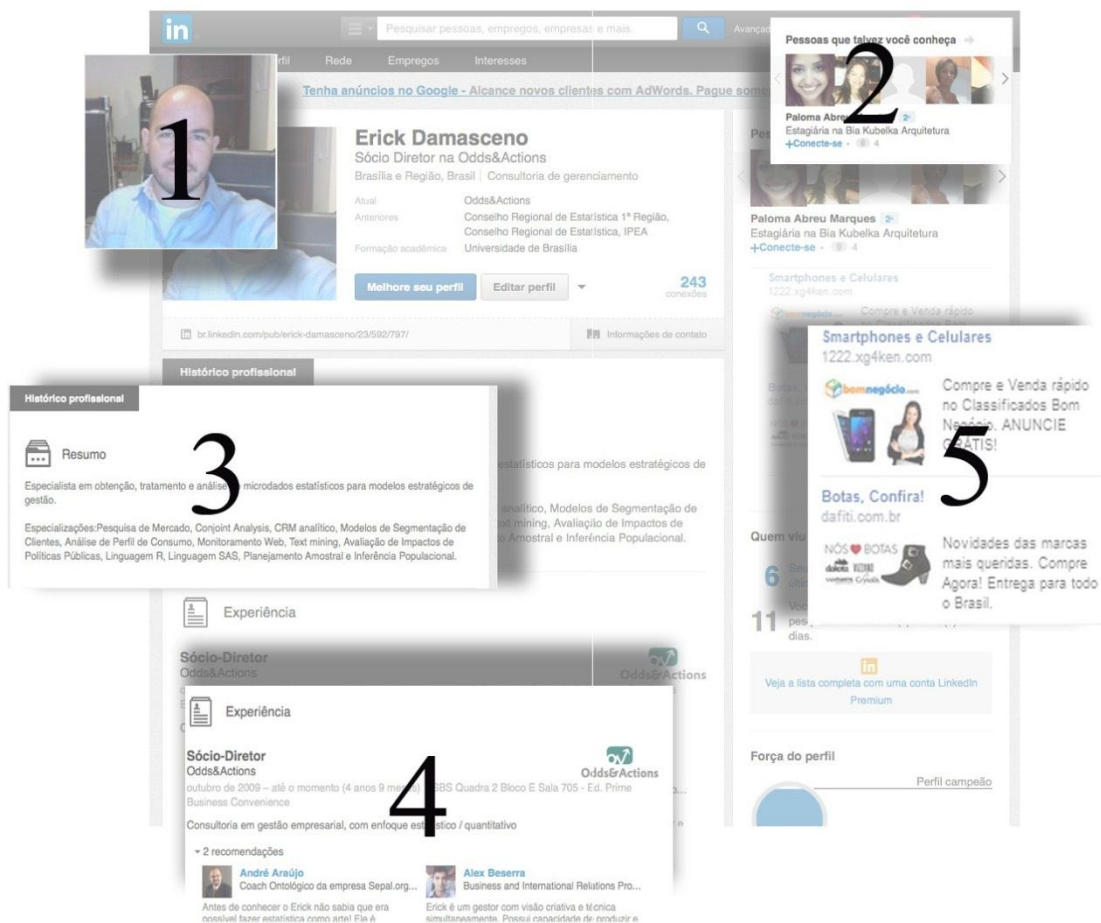
Região 2 – Pessoas que talvez você conheça

Região 3 – Resumo

Região 4 – Currículo

Região 5 – Anúncios

Figura 6.4. Regiões de atividade do LinkedIn



7. Inferência da Atratividade

Apresentamos aqui a inferência da atratividade das regiões das redes sociais. Toda inferência foi realizada dentro de um contexto bayesiano e através de técnicas de simulação.

Após dividir as interfaces em regiões de atividade, o interesse passa a ser calcular a atratividade de cada região. No caso do Orkut, por exemplo, a interface foi dividida em seis regiões de atividade: 1 - Foto do Perfil; 2 - Sobre Mim; 3 - Amigos; 4 - Comunidades; 5 - Fotos e Vídeos; 6 - Outros. Como foram coletadas 80 interfaces da rede social, cada interface terá um score padronizado para determinada região, que será chamado de X_{ij} , sendo $i=1,2,3,4,5,6$ indicando a região de atividade e $j = 1,2,3,\dots,80$ indicando a observação na amostra.

O interesse está em estimar a atratividade para cada uma das seis regiões, que serão denominados de $\theta_1, \theta_2, \theta_3, \theta_4, \theta_5$ e θ_6 . Essas estimativas serão calculadas através dos scores padronizados X_{ij} , que foram encontrados na amostra das interfaces.

Como $0 < X_{ij} < 1$ e $\sum_{i=1}^6 X_{ij} = 1$, assumiremos que $(X_{1j}, X_{2j}, X_{3j}, X_{4j}, X_{5j}, X_{6j} \mid \theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6) \sim \text{Dirichlet}(1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6)$. Assim, a função de verossimilhança é dada por:

$$p(\mathbf{X} \mid \Theta) = \prod_{j=1}^n \frac{1}{B(\Theta)} X_{1j}^{\theta_1-1} X_{2j}^{\theta_2-1} X_{3j}^{\theta_3-1} X_{4j}^{\theta_4-1} X_{5j}^{\theta_5-1} X_{6j}^{\theta_6-1}$$

$$= (B(\Theta))^{-n} \left(\prod_{j=1}^n X_{1j} \right)^{\theta_1-1} \left(\prod_{j=1}^n X_{2j} \right)^{\theta_2-1} \left(\prod_{j=1}^n X_{3j} \right)^{\theta_3-1} \left(\prod_{j=1}^n X_{4j} \right)^{\theta_4-1} \left(\prod_{j=1}^n X_{5j} \right)^{\theta_5-1} \left(\prod_{j=1}^n X_{6j} \right)^{\theta_6-1}$$

no qual $B(\Theta) = \frac{\prod_{i=1}^6 \Gamma(\theta_i)}{\Gamma(\sum_{i=1}^6 \theta_i)}$ é a constante normalizadora da distribuição Dirichlet e Θ

$$= (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6).$$

A distribuição Dirichlet é uma extensão multivariada da família Beta, a que se reduziria caso $i=1,2$. Nesta situação a interface da rede social teria apenas duas regiões de atividade.

Como $\theta_i > 0, i=1,2,\dots,6$, assumiremos, a priori, distribuições independentes *Gama* (a_i, b_i) .

$$p(\theta_i) = \frac{b_i^{a_i}}{\Gamma(a_i)} \theta_i^{a_i-1} e^{-b_i \theta_i}$$

Assim, pela Fórmula de Bayes a distribuição a posteriori de $\Theta|X$ é dada por:

$$p(\Theta|X) = \frac{p(X, \Theta)}{p(X)} = \frac{p(X|\Theta) p(\Theta)}{p(X)} = \frac{p(X|\Theta) p(\Theta)}{\int p(\Theta, X) d\Theta}$$

No caso, $1/p(X)$ não depende de Θ , e funciona como constante normalizadora. Para um valor fixo de X , a função $p(X|\Theta)$ fornece a plausibilidade ou verossimilhança de cada um dos possíveis valores de Θ enquanto que $p(\Theta)$ é distribuição à priori de Θ . Combinando estas duas fontes de informação, priori e verossimilhança, obtém-se a distribuição à posteriori de $\Theta, p(\Theta|X)$. Assim, pode-se reescrever a Fórmula de Bayes da seguinte forma:

$$p(\Theta|X) \propto p(\Theta; X)p(\Theta)$$

ou seja, distribuição a posteriori \propto verossimilhança \times distribuição a priori

Portanto, a distribuição à posteriori de $(\Theta|X)$ será dada por:

$$p(\Theta|X) \propto (B(\theta))^{-n} \left(\prod_{j=1}^n x_{1j} \right)^{\theta_1-1} \left(\prod_{j=1}^n x_{2j} \right)^{\theta_2-1} \left(\prod_{j=1}^n x_{3j} \right)^{\theta_3-1} \left(\prod_{j=1}^n x_{4j} \right)^{\theta_4-1} \left(\prod_{j=1}^n x_{5j} \right)^{\theta_5-1} \left(\prod_{j=1}^n x_{6j} \right)^{\theta_6-1}$$

$$\times \theta_1^{a_1-1} e^{-b_1\theta_1} \theta_2^{a_2-1} e^{-b_2\theta_2} \theta_3^{a_3-1} e^{-b_3\theta_3} \theta_4^{a_4-1} e^{-b_4\theta_4} \theta_5^{a_5-1} e^{-b_5\theta_5} \theta_6^{a_6-1} e^{-b_6\theta_6}$$

E as distribuições a posteriori condicionais de $(\theta_i|X)$, $i=1,2,\dots,6$ são dadas

por:

$$p(\theta_i|X, \Theta_{(-i)}) \propto \varphi(\theta_i) \times Gama \left(a_i, b_i - \log \prod_{j=1}^n x_{1j} \right)$$

com

$$\varphi(\theta_i) = \left(\frac{\Gamma(\theta_i)}{\Gamma(\sum_{u=1}^6 \theta_u)} \right)^{-n} \quad ,, \quad i=1,2,\dots,6.$$

Note que as distribuições à posteriori condicionais não possuem uma distribuição fechada. No entanto, os valores de θ_i podem ser gerados através do método MCMC - Markov Chain Monte Carlo (Geman, 1997) usando passos do algoritmo Metropolis Hastings (Metropolis, 1953; Hastings, 1970).

Como a atratividade média da região i é dada por:

$$E(X_i|\Theta) = \frac{\theta_i}{\sum_{u=1}^6 \theta_u}$$

Então, a distribuição a posteriori de $E(X_i|\theta)$ pode ser facilmente obtida através dos valores gerados da posteriori $p(\theta|\mathbf{X})$. Logo, a estimativa pontual e seu respectivo Intervalo de Credibilidade da atratividade média da região i é dada, respectivamente, pela esperança e quantis de $\left(\frac{\theta_i}{\sum_{u=1}^6 \theta_u} \mid \mathbf{X}\right)$, $i=1,2,\dots,6$.

8. Resultados

Das 80 interfaces do Facebook e Orkut, e 40 do Twitter e LinkedIn, sobrepostas com suas mapas de saliência e enumeradas de 1 a 16, foram calculados os escores para cada região de atividade, sendo possível encontrar uma média amostral para a atratividade das áreas. Foi calculado também o desvio padrão da atratividade e a porcentagem para cada região de atividade. Estes resultados são apresentados nas Tabelas 8.1, 8.2, 8.3 e 8.4.

Os resultados da rede social Orkut mostraram que a região que tem maior atratividade média é a de “Amigos”, com 62,5% da atenção concentrada nessa área. A região de menor atratividade do Orkut foi a “Foto do Perfil”, com 4,7% da atenção visual, seguida pela região “Sobre Mim”, com 5,4% da atenção visual. Os demais resultados se encontram na tabela abaixo:

Tabela 8.1. Resultados da atratividade da Rede Social Orkut

	Atratividade Média	Desvio Padrão	Porcentagem da Atratividade média
Foto do Perfil	0,04	0,08	4,7%
Amigos	0,62	0,22	62,5%
Comunidades	0,09	0,10	9,8%
Sobre Mim	0,05	0,09	5,4%
Fotos	0,15	0,23	15,8%
Outros	0,01	0,4	1,8%

Os resultados da rede social Facebook mostram que a região de atividade de maior atenção visual é a “Sobre Você”, onde o usuário se descreve, com 37% da atenção visual, seguida pela região de “amigos em comum” com 34%. A região de

atividade de menor atratividade média foi a “Busca, Anúncios e Outros”, com 7,4% da atratividade média.

Tabela 8.2. Resultados da atratividade da Rede Social Facebook

	Atratividade Média	Desvio Padrão	Porcentagem da Atratividade média
Foto do Perfil	0,03	0,04	2,91%
Sobre Você	0,37	0,15	37%
Mural	0,06	0,13	6,3%
Amigos em comum	0,34	0,2	34%
Capa	0,13	0,13	13%
Busca, Anúncios e Outros	0,07	0,1	7,4%

Os resultados da rede social LinkedIn mostram que a região de atividade de maior atenção visual é a “Resumo” (com 30% da atratividade média), onde o usuário se descreve. Seguida de perto, a segunda região de maior atratividade foi a “Pessoas que talvez você conheça” (com 24,4% da atratividade média), que se assemelha aos “Amigos” do Orkut, região de maior atratividade. A “Foto do Perfil”, assim como no Facebook, recebeu a menor atratividade média, com 17,3% da atenção visual. Por ser uma rede social mais profissional é compreensível que a atenção visual do LinkedIn se volte para a “análise de currículo” e o “Resumo”.

Tabela 8.3 Resultados da atratividade da Rede Social LinkedIn

	Atratividade Média	Desvio Padrão	Porcentagem da Atratividade média
Foto do Perfil	0,17	0,06	17,3%
Pessoas que talvez você conheça	0,24	0,12	24,4%
Resumo	0,3	0,09	30%
Currículo	0,17	0,12	17,6%
Anúncios	0,1	0,08	10,5%

Os resultados da rede social Twitter mostram que a região de atividade de maior atenção visual são as “informações Twittadas”, ou seja, as mensagens que os usuários postam na rede. A segunda maior região de atratividade foram as “informações sobre a quantidade de Twittes”, quantidade de pessoas seguidas e quantidade de seguidores do usuário, com 35,5% da atratividade média. A menor região de atratividade foi a “busca”, com 1,2% da atenção visual.

Tabela 8.4 Resultados da atratividade da Rede Social Twitter

	Atratividade Média	Desvio Padrão	Porcentagem da Atratividade média
Busca	0,01	0,03	1,2%
Foto do Perfil	0,07	0,12	6,7%
Cadastro de e-mail	0,06	0,07	6,5%
Fotos e Vídeos	0,10	0,13	10,7%
Informações	0,35	0,23	35,5%
Mensagens Twittadas	0,39	0,23	39%

A atratividade foi estimada segundo o modelo Dirichlet como apresentado no capítulo 7. Calculado o estimador pontual da atratividade de cada região, foi calculado também o intervalo de credibilidade da estimativa, que foi definido entre os quantis de 2,5% e 97,5%, e comparado com a média amostral da atratividade apresentadas pelas tabelas 8.1 a 8.4.

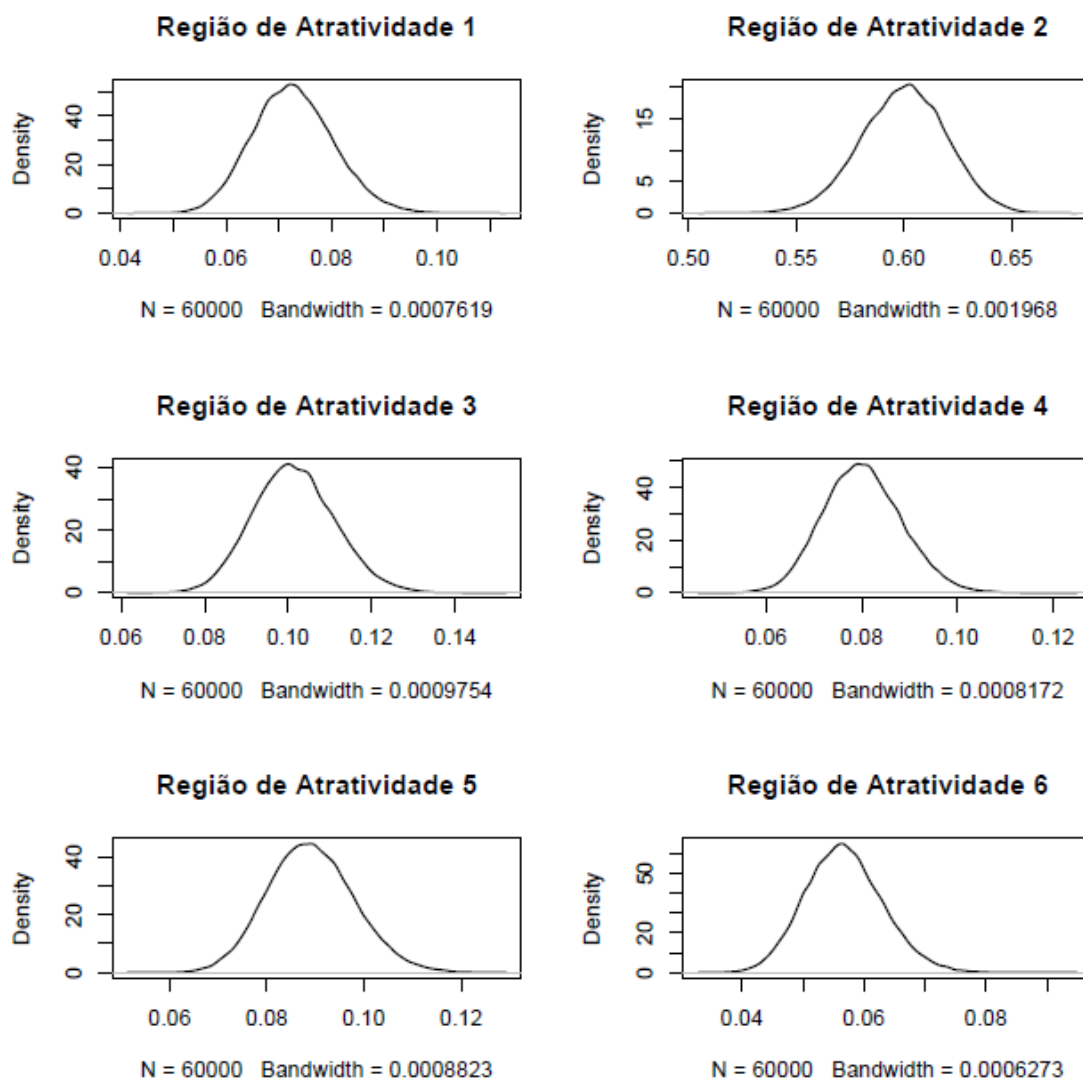
A tabela 8.5 abaixo mostra os resultados obtidos para a rede social Orkut. As regiões de Amigos e Comunidades se encontram dentro do intervalo de credibilidade. A atratividade média da Foto do Perfil, Sobre Mim, Fotos e Outros se encontrou fora do intervalo de credibilidade.

A Figura 8.1 mostra as densidades a posteriori estimadas da atratividade média para cada região de atratividade do Orkut. As densidades foram baseadas em uma amostra de tamanho 100.000 considerando um período de aquecimento (burn-in) de 10.000. Aparentemente, todas as distribuições apresentaram uma curva simétrica.

Tabela 8.5. Resultados da estimativa da atratividade da Rede Social Orkut

	Atratividade Média	Estimativa da Atratividade	Intervalo de Credibilidade 95%	
			Limite Inferior	Limite Superior
1. Foto do Perfil	0,04	0,072	0,058	0,088
2. Amigos	0,62	0,599	0,56	0,637
3. Comunidades	0,09	0,101	0,083	0,121
4. Sobre Mim	0,05	0,08	0,065	0,096
5. Fotos	0,15	0,089	0,072	0,107
6. Outros	0,01	0,056	0,045	0,069

Figura 8.1. Gráfico de densidades das Regiões do Orkut



A tabela 8.6 mostra os resultados obtidos para a rede social Facebook. As regiões “Sobre Você”, “Mural”, “Amigos em Comum”, “Busca”, “anúncio e Outros” se encontraram dentro do intervalo de credibilidade. A atratividade média da “Foto do Perfil”, e “capa” se encontrou fora do intervalo de credibilidade.

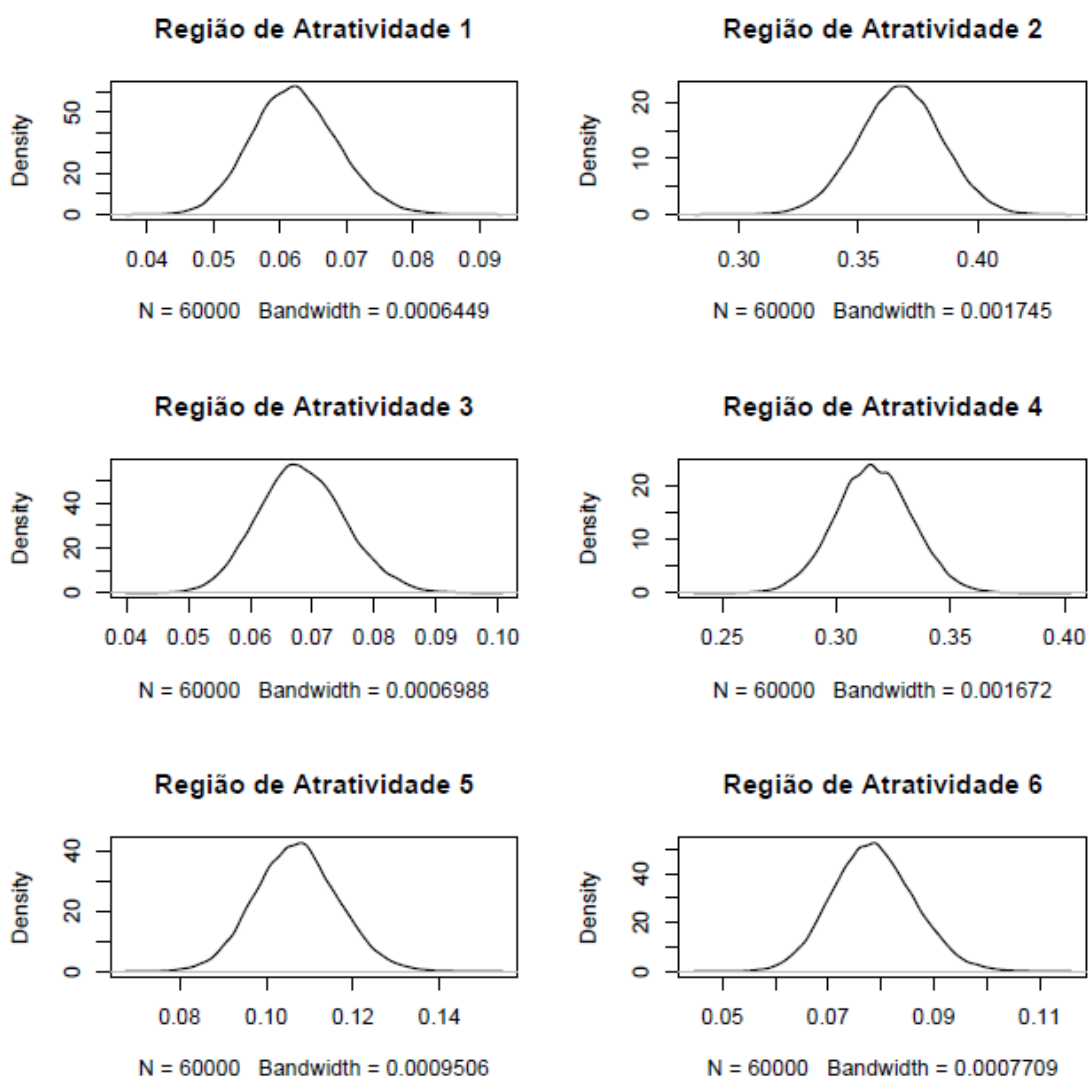
A Figura 8.2 mostra as densidades a posteriori estimadas da atratividade média para cada região de atratividade do Facebook. As densidades foram baseadas em uma amostra de tamanho 100.000 considerando um período de

aquecimento (burn-in) de 10.000. Aparentemente, todas as distribuições apresentaram uma curva simétrica.

Tabela 8.6. Resultados da estimativa da atratividade da Rede Social Facebook

	Atratividade Média	Estimativa da Atratividade	Intervalo de Credibilidade 95%	
			Limite Inferior	Limite Superior
Foto do Perfil	0,03	0,062	0,050	0,075
Sobre Você	0,37	0,367	0,332	0,401
Mural	0,06	0,068	0,552	0,082
Amigos em comum	0,34	0,316	0,283	0,349
Capa	0,13	0,107	0,089	0,126
Busca, Anúncios e Outros	0,07	0,078	0,064	0,094

Figura 8.2. Gráfico de densidades das Regiões do Facebook



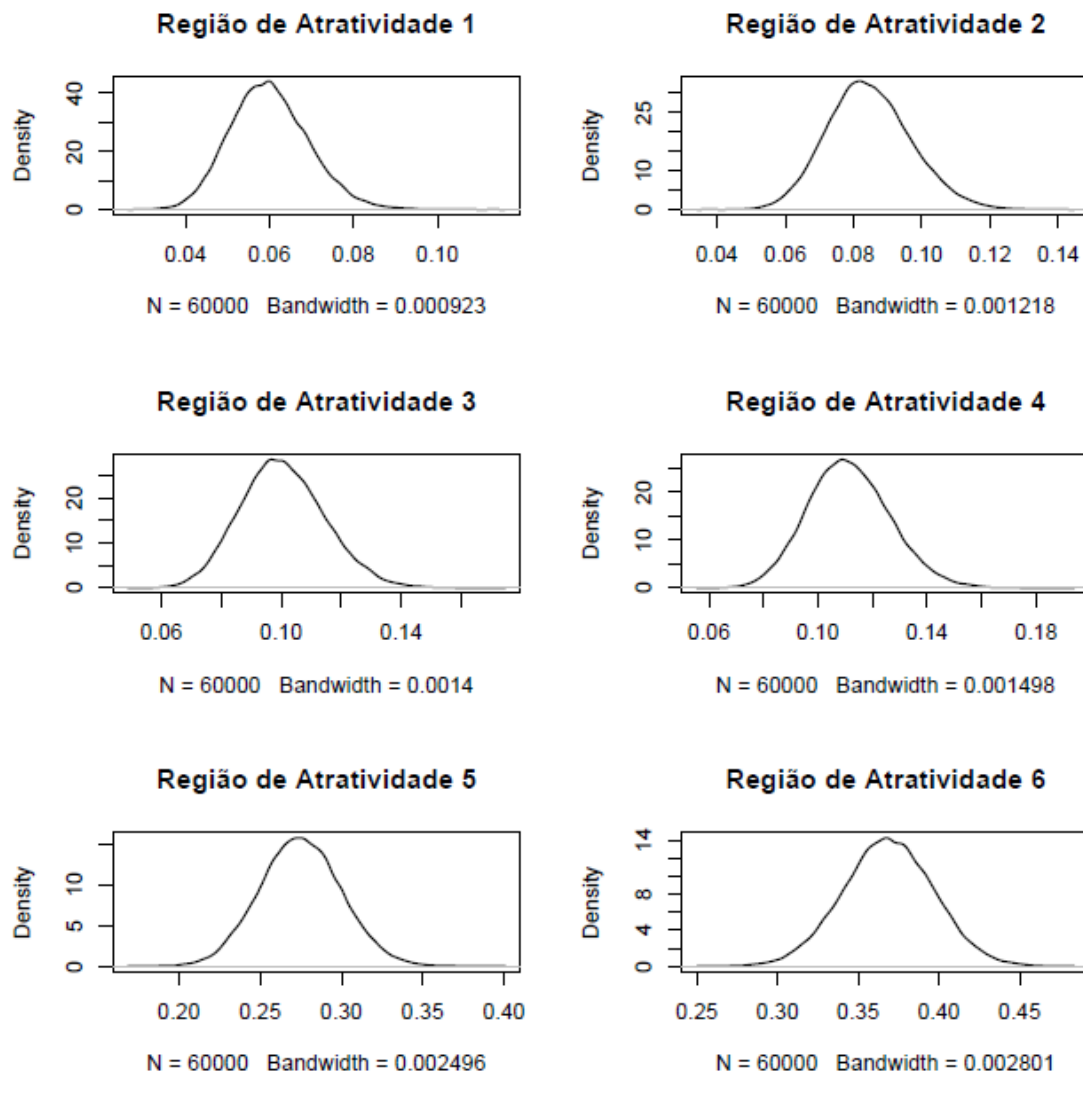
A tabela 8.7 mostra os resultados obtidos para a rede social Twitter. As regiões “Foto do Perfil”, “Fotos e Vídeos” e “mensagens twittadas” se encontraram dentro do intervalo de credibilidade. A atratividade média da “Busca”, “cadastro de e-mail”, “informações” se encontrou fora do intervalo de credibilidade.

A Figura 8.3 mostra as densidades a posteriori estimadas da atratividade média para cada região de atratividade do Twitter. As densidades foram baseadas em uma amostra de tamanho 100.000 considerando um período de aquecimento (burn-in) de 10.000. Aparentemente, todas as distribuições apresentaram uma curva simétrica.

Tabela 8.7 Resultados da estimativa da atratividade da Rede Social Twitter

	Atratividade Média	Estimativa da Atratividade	Intervalo de Credibilidade 95%	
			Limite Inferior	Limite Superior
1. Busca	0,01	0,062	0,051	0,075
2. Foto do Perfil	0,07	0,084	0,061	0,110
3. Cadastro de e-mail	0,06	0,1	0,074	0,129
4. Fotos e Vídeos	0,10	0,111	0,083	0,142
5. Informações	0,35	0,274	0,226	0,324
6. Mensagens Twittadas	0,39	0,368	0,313	0,423

Figura 8.3. Gráfico de densidades da atratividade das Regiões do Twitter



A tabela 8.8 abaixo mostra os resultados obtidos para a rede social LinkedIn. Diferente das outras redes, o LinkedIn, teve cinco regiões de atratividade escolhidas. Todas as atratividades médias estão dentro do intervalo de credibilidade.

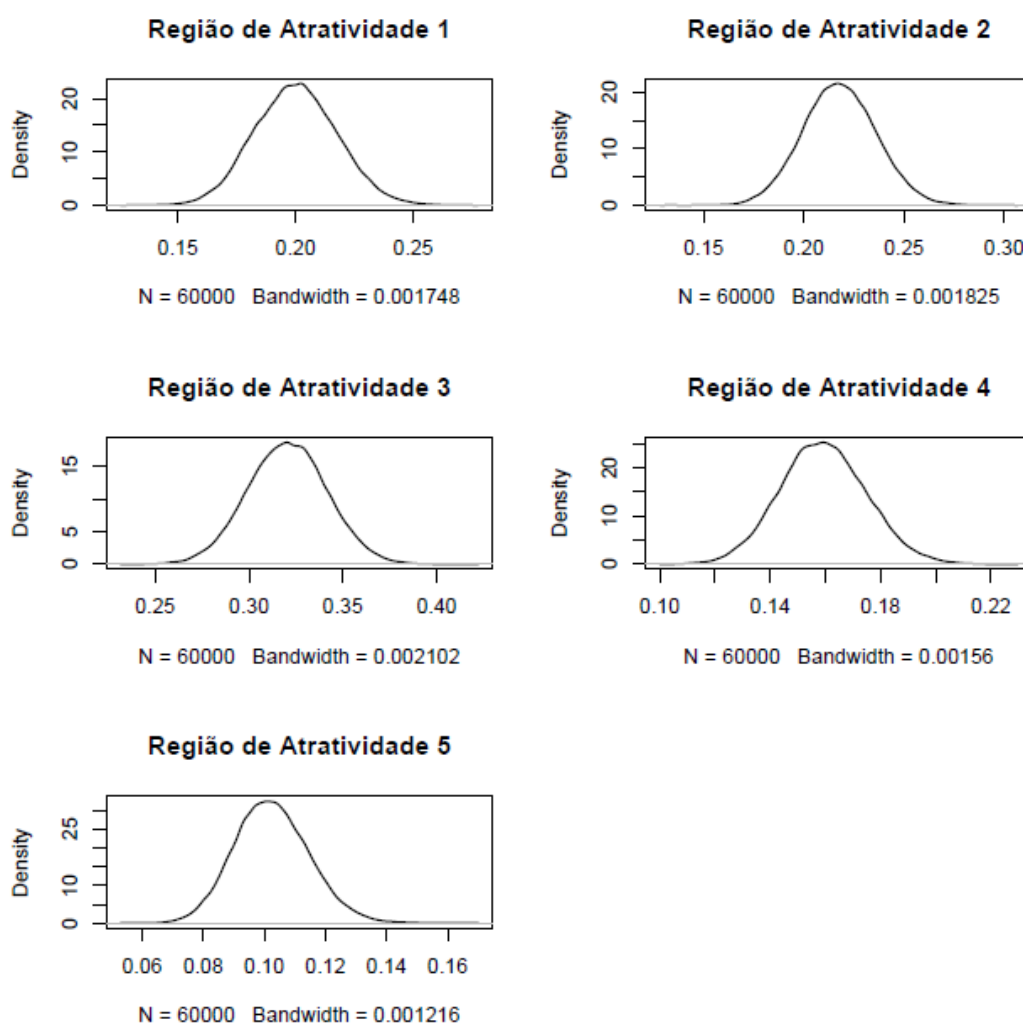
A Figura 8.4 mostra as densidades a posteriori estimadas da atratividade média para cada região de atratividade do LinkedIn. As densidades foram baseadas em uma amostra de tamanho 100.000 considerando um período de aquecimento

(burn-in) de 10.000. Aparentemente, todas as distribuições apresentaram uma curva simétrica.

Tabela 8.8 Resultados da estimativa da atratividade da Rede Social LinkedIn

	Atratividade Média	Estimativa da Atratividade	Intervalo de Credibilidade 95%	
			Limite Inferior	Limite Superior
1. Foto do Perfil	0,17	0,200	0,166	0,234
2. Pessoas que talvez você conheça	0,24	0,234	0,182	0,254
3. Resumo	0,3	0,320	0,279	0,361
4. Currículo	0,17	0,159	0,129	0,190
5. Anúncios	0,1	0,102	0,079	0,127

Quadro 8.4. Gráfico de densidades da atratividade das Regiões do LinkedIn



9. Conclusões

Neste trabalho foi descrito o procedimento, baseado na distribuição Dirichlet, para achar um estimador para a média de atratividades em regiões específica de interfaces das quatro redes sociais mais utilizadas no Brasil.

Através distribuição à posteriori encontrada, foi possível encontrar os estimadores de atratividades para as regiões. Foram usadas 80 interfaces do Facebook e do Orkut e 40 do Twitter e do LinkedIn.

O modelo Dirichlet apresentou um bom ajuste dos dados, apresentando uma tendência de superestimar a atratividade das regiões menos atrativas e subestimar as regiões mais atrativas. Mesmo assim, a maioria dos estimadores se encaixou no intervalo de credibilidade encontrado. Vale lembrar que os valores da atratividade apresentam a restrição da soma ser igual a 1. Assim, qualquer desvio na estimativa de uma determinada região de atratividade reflete diretamente nos demais valores. Esse comportamento apresenta um maior reflexo nas regiões que apresentam baixas médias de atratividade e grande variância.

Observou-se que a região Amigos, foi bastante atrativa nas redes sociais Orkut e Facebook, semelhante a região Pessoas que Talvez você conheça, do LinkedIn, que também foi bastante atrativa. A região Sobre você, a mais atrativa do Facebook, onde o usuário se descreve, se assemelha ao Resumo do LinkedIn, e também foi uma região de muita atenção. No twitter, as mensagens twittadas são as que chamam mais a atenção. A região de busca e anúncios são as menos atrativa nas redes sociais.

Comparando os resultados obtidos com os resultados de Benevenuto et al (2009) para a rede social Orkut, observa-se que foram parecidos. Por as regiões de atratividade serem mais específicas neste trabalho, a comparação do LinkedIn com os resultados de Benevenuto foi mais limitada, mas também apresentou leve semelhança

10. Bibliografia

FERREIRA, D; RODRIGES, G; AKAMINE, W. “Um Estudo Sobre o Impacto da Atenção Visual nas Redes Sociais” In: II Brazilian Workshop on Social Network Analysis and Mining, 2013, Maceió. BraSNAM 2013. p. 1.

BENEVENUTO, F. (2009). Characterizing user behavior in online social networks: In Proceedings of the 9th ACM SIGCOMM conference on Internet Measurement Conference, p. 49–62.

HAMMERSLEY, J; HANDSCOMB, D. (1964). *Monte Carlo methods*. London: Methuen.

PAULINO, C; TURKMAN, M; MURTEIRA, B. (2003). *Estatística Bayesiana*. Fundação Calouste Gulbenkian, Lisboa.

BALAKRISHNAN V. (2005). Dirichlet Distribution. Chapter 27. *A Primer on Statistical Distributions*. Hoboken, NJ: John Wiley & Sons, Inc. p. 274.

RAJASHEKAR, U; Van der Linde I; Bovik AC; Cormack LK. (2008) Gaffe: *Agaze-attentive fixation finding engine*. IEEE Transactions on Image Processing, p. 564–573.

R Development Core Team (2012). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.

Marteleto, R. M. (2001). *Análise de redes sociais: aplicação nos estudos de transferência da informação*. Programa de Pós-Graduação em ciência da Informação, MCT/IBICT – UFRJ/ECO

Apêndice A – Programação em R da geração dos estimadores e intervalos de credibilidade

```
#### 6 regiões
set.seed(123)

### prioris ###
a1<-a2<-a3<-a4<-a5<-a6<- .0001
b1<-b2<-b3<-b4<-b5<-b6<- .0001

#####
### Entrar com os dados aqui
### dados - X1 = prod(x1)
#####
X1<-exp("Soma dos logarítimos de cada imagem para a região 1")
X2<-exp("Soma dos logarítimos de cada imagem para a região 2")
X3<-exp("Soma dos logarítimos de cada imagem para a região 3")
X4<-exp("Soma dos logarítimos de cada imagem para a região 4")
X5<-exp("Soma dos logarítimos de cada imagem para a região 5")
X6<-exp("Soma dos logarítimos de cada imagem para a região 6")

### n igual a 80, para o Facebook e o Orkut e 40 para o Twitter e o LinkedIn

n<-80

#####
### LOG POSTERIORI ###
#####
log.post<-function(theta,X1,X2,X3,X4,X5,X6,a1,a2,a3,a4,a5,a6,b1,b2,b3,b4,b5,b6,n){

if ( (theta[1]>.01) && (theta[2]>.01) && (theta[3]>.01) && (theta[4]>.01) &&
(theta[5]>.01) && (theta[6]>.01))
{
soma<-sum(theta)
print(soma)
Log1<-log(gamma(theta[1]))
Log2<-log(gamma(theta[2]))
Log3<-log(gamma(theta[3]))
Log4<-log(gamma(theta[4]))
Log5<-log(gamma(theta[5]))
Log6<-log(gamma(theta[6]))
LogGammaSoma<-log(gamma(soma))

(a1-1)*log(theta[1]) - theta[1]*(b1-log(X1)) +
(a2-1)*log(theta[2]) - theta[2]*(b2-log(X2)) +
```

```

(a3-1)*log(theta[3]) - theta[3]*(b3-log(X3)) +
(a4-1)*log(theta[4]) - theta[4]*(b4-log(X4)) +
(a5-1)*log(theta[5]) - theta[5]*(b5-log(X5)) +
(a6-1)*log(theta[6]) - theta[6]*(b6-log(X6)) -
n * ( Log1 + Log2 + Log3 + Log4 + Log5 + Log6 - LogGammaSoma) }
else return (-Inf)}

```

```

M<-100000
b<-10000

```

```

funcao<-rep(3,1,6)
require(MCMCpack)
THETA <- MCMCmetrop1R(log.post, theta.init=chute,thin=5,mcmc=M,
burnin=b,force.samp=T, logfun = TRUE,
X1=X1,X2=X2,X3=X3,X4=X4,X5=X5,X6=X6,a1=a1,a2=a2,a3=a3,a4=a4,a5=a5,a6=a6,
b1=b1,b2=b2,b3=b3,b4=b4,b5=b5,b6=b6,n=n)

```

```

### estimativa atratividade 1
atrat1<-
THETA[,1]/(THETA[,1]+THETA[,2]+THETA[,3]+THETA[,4]+THETA[,5]+THETA[,6])
mean(atrat1)      ## estimativa pontual
quantile(atrat1,0.025)  ## LI IC95%
quantile(atrat1,0.975)  ## LS IC95%

```

```

### estimativa atratividade 2
atrat2<-
THETA[,2]/(THETA[,1]+THETA[,2]+THETA[,3]+THETA[,4]+THETA[,5]+THETA[,6])
mean(atrat2)      ## estimativa pontual
quantile(atrat2,0.025)  ## LI IC95%
quantile(atrat2,0.975)  ## LS IC95%

```

```

### estimativa atratividade 3
atrat3<-
THETA[,3]/(THETA[,1]+THETA[,2]+THETA[,3]+THETA[,4]+THETA[,5]+THETA[,6])
mean(atrat3)      ## estimativa pontual
quantile(atrat3,0.025)  ## LI IC95%
quantile(atrat3,0.975)  ## LS IC95%

```

```

### estimativa atratividade 4
atrat4<-
THETA[,4]/(THETA[,1]+THETA[,2]+THETA[,3]+THETA[,4]+THETA[,5]+THETA[,6])
mean(atrat4)      ## estimativa pontual
quantile(atrat4,0.025)  ## LI IC95%
quantile(atrat4,0.975)  ## LS IC95%

```

```

### estimativa atratividade 5
atrat5<-
THETA[,5]/(THETA[,1]+THETA[,2]+THETA[,3]+THETA[,4]+THETA[,5]+THETA[,6])
mean(atrat5)      ## estimativa pontual
quantile(atrat5,0.025)  ## LI IC95%
quantile(atrat5,0.975)  ## LS IC95%

```

```

### estimativa atratividade 6
atrat6<-
THETA[,6]/(THETA[,1]+THETA[,2]+THETA[,3]+THETA[,4]+THETA[,5]+THETA[,6])
mean(atrat6)      ## estimativa pontual
quantile(atrat6,0.025)  ## LI IC95%
quantile(atrat6,0.975)  ## LS IC95%

```

```

par(mfrow=c(3,2))
plot(density(atrat1),main="Região de Atratividade 1")
plot(density(atrat2),main="Região de Atratividade 2")
plot(density(atrat3),main="Região de Atratividade 3")
plot(density(atrat4),main="Região de Atratividade 4")
plot(density(atrat5),main="Região de Atratividade 5")
plot(density(atrat6),main="Região de Atratividade 6")

```

```

#####
#### 5 regiões
set.seed(123)

```

```

### prioris ###
a1<-a2<-a3<-a4<-a5<- .0001
b1<-b2<-b3<-b4<-b5<- .0001

```

```

#####
### Entrar com os dados aqui
### dados - X1 = prod(x1)
#####
X1<-exp("Soma dos logarítimos de cada imagem para a região 1")
X2<-exp("Soma dos logarítimos de cada imagem para a região 2")
X3<-exp("Soma dos logarítimos de cada imagem para a região 3")
X4<-exp("Soma dos logarítimos de cada imagem para a região 4")
X5<-exp("Soma dos logarítimos de cada imagem para a região 5")

```

```

### n igual a 80, para o Facebook e o Orkut e 40 para o Twitter e o LinkedIn

```

```

n<-80

```

```
#####
### LOG POSTERIORI ###
#####
log.post<-function(theta,X1,X2,X3,X4,X5,a1,a2,a3,a4,a5,b1,b2,b3,b4,b5,,n){

if ( (theta[1]>.01) && (theta[2]>.01) && (theta[3]>.01) && (theta[4]>.01) &&
(theta[5]>.01))
{
soma<-sum(theta)
print(soma)
Log1<-log(gamma(theta[1]))
Log2<-log(gamma(theta[2]))
Log3<-log(gamma(theta[3]))
Log4<-log(gamma(theta[4]))
Log5<-log(gamma(theta[5]))
LogGammaSoma<-log(gamma(soma))

(a1-1)*log(theta[1]) - theta[1]*(b1-log(X1)) +
(a2-1)*log(theta[2]) - theta[2]*(b2-log(X2)) +
(a3-1)*log(theta[3]) - theta[3]*(b3-log(X3)) +
(a4-1)*log(theta[4]) - theta[4]*(b4-log(X4)) +
(a5-1)*log(theta[5]) - theta[5]*(b5-log(X5)) -
n * ( Log1 + Log2 + Log3 + Log4 + Log5 - LogGammaSoma) }
else return (-Inf)}

M<-300000
b<-100000

funcao<-rep(3,1,5)
require(MCMCpack)
THETA <- MCMCmetrop1R(log.post, theta.init=chute,thin=5,mcmc=M,
burnin=b,force.samp=T, logfun = TRUE, X1=X1,X2=X2,X3=X3,X4=X4,X5=X5
,a1=a1,a2=a2,a3=a3,a4=a4,a5=a5,b1=b1,b2=b2,b3=b3,b4=b4,b5=b5,n=n)

### estimativa atratividade 1
atrat1<-THETA[,1]/(THETA[,1]+THETA[,2]+THETA[,3]+THETA[,4]+THETA[,5])
mean(atrat1)      ## estimativa pontual
quantile(atrat1,0.025)  ## LI IC95%
quantile(atrat1,0.975)  ## LS IC95%

### estimativa atratividade 2
atrat2<-THETA[,2]/(THETA[,1]+THETA[,2]+THETA[,3]+THETA[,4]+THETA[,5])
mean(atrat2)      ## estimativa pontual
quantile(atrat2,0.025)  ## LI IC95%
quantile(atrat2,0.975)  ## LS IC95%
```

```
### estimativa atratividade 3
atrat3<-THETA[3]/(THETA[1]+THETA[2]+THETA[3]+THETA[4]+THETA[5])
mean(atrat3)      ## estimativa pontual
quantile(atrat3,0.025)  ## LI IC95%
quantile(atrat3,0.975)  ## LS IC95%
```

```
### estimativa atratividade 4
atrat4<-THETA[4]/(THETA[1]+THETA[2]+THETA[3]+THETA[4]+THETA[5])
mean(atrat4)      ## estimativa pontual
quantile(atrat4,0.025)  ## LI IC95%
quantile(atrat4,0.975)  ## LS IC95%
```

```
### estimativa atratividade 5
atrat5<-THETA[5]/(THETA[1]+THETA[2]+THETA[3]+THETA[4]+THETA[5])
mean(atrat5)      ## estimativa pontual
quantile(atrat5,0.025)  ## LI IC95%
quantile(atrat5,0.975)  ## LS IC95%
```

```
par(mfrow=c(3,2))
plot(density(atrat1),main="Região de Atratividade 1")
plot(density(atrat2),main="Região de Atratividade 2")
plot(density(atrat3),main="Região de Atratividade 3")
plot(density(atrat4),main="Região de Atratividade 4")
plot(density(atrat5),main="Região de Atratividade 5")
```