

Quem você conhece? Uma análise exploratória de redes sociais a partir de dados do Orkut

Matheus Albergaria de Magalhães
Instituto Jones dos Santos Neves (IJSN)
Victor Nunes Toscano
Instituto Jones dos Santos Neves (IJSN)

Resumo: O presente trabalho tem como objetivo realizar uma análise exploratória de padrões empíricos relacionados à formação de redes sociais na Internet. Para tanto, faz-se uso de uma amostra construída a partir de dados do *website* de relacionamentos *Orkut*. Os resultados obtidos a partir da análise conduzida demonstram que: (i) ocorre o fenômeno de “cauda longa” para dados da amostra, tanto no caso do número de contatos quanto no número de comunidades dos usuários do *Orkut*; (ii) existem nítidas diferenças entre homens e mulheres no que diz respeito a vários quesitos referentes a preferências dos usuários; (iii) resultados de um exercício econométrico demonstram que variáveis relacionadas à popularidade dos usuários (número de recados, fãs e classificações dadas por outros usuários) exercem um efeito quantitativo significativo no número de contatos estabelecido por cada usuário; (iv) por outro lado, não há coincidência entre usuários que possuem o maior número de contatos e aqueles que pertencem ao maior número de comunidades. Esses resultados são interessantes por constituírem evidência empírica inicial sobre a dinâmica de formação e consolidação de redes sociais entre indivíduos.

Palavras-Chave: Internet, *Orkut*, Redes Sociais.

Abstract: This paper aims to provide an exploratory analysis of empirical patterns related to social networks. In doing so, it is based on a sample of users from the *Orkut website*. Results obtained from the empirical analysis show that: (i) there is a “long tail” phenomenon in the data, both for contacts and communities per user; (ii) there are remarkably differences between men and women in terms of users’ preferences; (iii) results of econometric estimations point to the importance of popularity measures (scraps, fans and other users’ classifications) as statistically significant measures explaining the number of contacts of each user; (iv) on the other hand, there is no coincidence between users that have a large number of contacts and those who belong to a large number of communities. These results are important in the sense of providing initial empirical evidence related to the formation and consolidation of social networks.

Key-Words: Internet, *Orkut*, Social Networks.

Resumen: El presente trabajo tiene como objetivo realizar un análisis exploratoria de padrones empíricos relacionados a la formulación de redes sociales en la internet. Para eso, se hace uso de una muestra hecha de los datos del *website* de relacionamientos *Orkut*. Los resultados del análisis han demostrado que: (i) ocurre el fenómeno de “cola larga” para datos de la muestra, tanto en el caso del número de

contactos, cuanto en el número de comunidades de los usuarios del Orkut; (ii) existen nítidas diferencias entre los hombre y mujeres a respecto de variados temas referentes a las preferencias del los usuarios; (iii) resultados de um ejercicio econométrico han demostrado que las variables relacionadas a popularidad del los usuarios (números de fans y las calificaciones dadas por otros usuarios) ejercen un efecto cuantitativo significativo en el numero de contactos establecidos por cada usuario; (iv) de otra manera, no hay coincidencia entre usuarios que poseen el mayor numero de contactos y aquellos que poseen el mayor numero de comunidades. Este resultados son interesantes por constituir evidencias empíricas acerca de la dinámica de formación y consolidación de redes sociales entre individuos.

Palabra clave: Internet, Orkut, Redes sociales.

Introdução

A formação de redes equivale a um importante fenômeno, sendo alvo de estudo de distintas áreas do conhecimento. Nos últimos anos, vem ocorrendo considerável aumento da pesquisa sobre o tema. Por exemplo, em Biologia, o conceito de ecossistema pode ser modelado como uma intrincada rede de cadeias de alimentação entre distintas espécies. Similarmente, o funcionamento de uma célula pode ser vista como o resultado de uma rede de interações entre genes, proteínas e moléculas. Em termos sociais, o surgimento e consolidação da Internet nas últimas décadas pode ser tido como um exemplo patente de formação de uma rede virtual em escalas sem precedentes (Amaral *et alli*. 2004).

No período recente, tem crescido o interesse de diversos cientistas sociais acerca do tema específico de formação de redes sociais. Por exemplo, Currarini, Jackson e Pin (2008), analisando uma base de dados relacionada à formação de amizades entre adolescentes (*Adolescent Health Dataset*) constroem uma representação esquemática de uma rede social, conforme ilustrado na figura 1:

Figura 1
Amizades entre Adolescentes de uma Escola Norte-Americana



Fonte: Currarini, Jackson e Pin (2008, p.9).

Observações:

- (a) Cada nó da figura representa um indivíduo distinto, com segmentos de reta entre nós representando relações sociais estabelecidas entre esses indivíduos.
- (b) Distintas cores dos nós representam diferentes raças dos usuários. No caso, branco = brancos, cinza = negros, cinza claro = hispânicos e preto = outras raças.

No caso da figura, cada nó da rede equivale a um usuário distinto, com os segmentos de reta ligando cada nó equivalendo às relações sociais estabelecidas entre esses nós. As cores dos nós da figura representam as raças dos alunos em análise. Nós brancos representam alunos brancos, nós cinza representam alunos negros, nós cinza claro representam alunos hispânicos e nós pretos representam outras raças. No caso particular da representação esquemática considerada, nota-se a formação de agrupamentos mais numerosos (*clusters*) entre estudantes da mesma etnia, pelo menos no caso de estudantes brancos e negros (ou seja, brancos tenderiam, em média a estabelecer relações com brancos, enquanto que negros tenderiam a estabelecer relações com negros). Esse fenômeno é denominado de “homofilia” pelos autores, equivalendo à tendência de grupos de indivíduos se associarem com grupos com características similares. A homofilia tem importantes consequências sobre a dinâmica social entre indivíduos, podendo impactar na difusão de informações e conhecimento de membros de uma rede.

A formação e eventual consolidação desses padrões pode auxiliar na obtenção de respostas a diferentes problemas sociais. Por exemplo, o surgimento de *clusters* poderia estar relacionado a questões relevantes relacionadas a comportamentos de discriminação a minorias ou de preferências por produtos específicos de acordo com distintos grupos considerados. Além disso, a compreensão do processo de surgimento e consolidação de redes sociais pode ajudar a responder questões como a disseminação de linguagens, a difusão de novas tecnologias, e o aumento das desigualdades sociais, por exemplo.

Apesar de existir atualmente um crescente volume de referências relacionadas ao tema em Economia no contexto internacional¹, ainda parece haver poucas contribuições em termos nacionais. Por conta disso, o objetivo deste trabalho equivale a apresentar uma abordagem exploratória do tema de formação de redes sociais na Internet. Para tanto, será feito uso de uma amostra construída a partir de perfis de usuários do *website* de relacionamentos *Orkut*. A principal finalidade do trabalho é descritiva, no sentido de apresentar algumas das principais regularidades empíricas relacionadas à amostra em questão. Em particular, busca-se realizar uma descrição do perfil do usuário médio de uma rede social nos moldes do *Orkut* e, ao mesmo tempo, compreender as variáveis relacionadas à dinâmica de formação de contatos de cada usuário.

As vantagens de um empreendimento desse tipo são basicamente duas. Primeiro, é importante tentar entender a dinâmica de formação de contatos e redes sociais entre indivíduos, um tópico de interesse não apenas na área de Economia, mas também em outras ciências sociais aplicadas, como a Sociologia, a Comunicação Social e a Psicologia, por exemplo. Conforme citado acima, a compreensão desse tema pode vir a revelar importantes informações acerca do modo como pessoas interagem entre si socialmente, tendo consequências nas mais diversas áreas de conhecimento que tem as relações sociais como foco de estudo.

¹ Para uma resenha sobre o tema, ver Jackson (2008).

Em segundo lugar, vale notar que a compreensão das preferências expressas pelos usuários de *sites* de relacionamento como o *Orkut* pode vir a constituir uma informação valiosa em termos de políticas de *marketing* e vendas para distintas empresas, quer elas vendam seus produtos através da Internet ou não. Por exemplo, uma firma que possua informações detalhadas a respeito das preferências de seus potenciais consumidores pode poupar um significativo montante de recursos que seria, a princípio, gasto com políticas de *marketing* e propaganda.

Vale a ressalva inicial de que este trabalho não pretende apresentar uma visão definitiva sobre o tema, até mesmo por conta de sua natureza extremamente mutável ao longo do tempo. Ao invés disso, o trabalho busca apenas apresentar algumas evidências relacionadas ao fenômeno de formação de redes no caso específico do *site* de relacionamentos em análise (*Orkut*). Em termos gerais, espera-se que a pesquisa futura possa complementar alguns dos resultados aqui reportados e obtenha eventualmente novos resultados acerca de alguns dos temas aqui abordados.

O trabalho está dividido da seguinte maneira: na segunda seção, é feita uma revisão parcial da literatura relacionada à formação de redes sociais e Internet em Economia. A terceira seção faz uma breve descrição do *website* de relacionamentos *Orkut*, ao passo que a quarta seção apresenta a metodologia e a base de dados empregados no trabalho. A quinta seção reporta os principais resultados obtidos a partir da análise desenvolvida. Finalmente, a sexta seção apresenta conclusões e sugere algumas linhas de pesquisa futura.

Referencial Teórico

O interesse pelo estudo de formação de redes sociais em Economia vem crescendo nos últimos anos, com alguns dos estudos iniciais sobre o tema remontando pelo menos até a década de 70. O objetivo da presente seção equivale a realizar uma breve revisão parcial dessa literatura, que cresce à cada dia que passa.

Uma contribuição seminal sobre o tema equivale ao estudo de Becker (1974), onde o autor busca modelar as razões inerentes às interações sociais entre agentes econômicos maximizadores de utilidade em um contexto microeconômico. No caso, o autor analisa um modelo econômico onde os agentes preocupam-se não apenas com seu próprio bem-estar, mas também se importam com o bem-estar de outros agentes. O conceito central de análise desse autor equivale ao de “renda social”, correspondente à soma da renda de um indivíduo com o valor monetário de características relevantes de outros agentes, denominado “ambiente social”. A partir desses conceitos, o autor analisa os efeitos de mudanças em gastos nos diferentes tipos de renda, assim como mudanças nos preços relativos de distintos bens, na mesma linha da análise microeconômica tradicional. Os resultados obtidos demonstram que, em geral, indivíduos tendem a dar mais atenção para interações entre membros de suas respectivas famílias em comparação a outras famílias. Esse autor chega ainda a tratar rapidamente de temas como caridade e inveja, que podem vir a surgir a partir de interações sociais.

Scheinkman (2008) apresenta uma breve resenha sobre o tema de interações sociais, com ênfase em formas alternativas de estimação desse fenômeno. Basicamente, segundo esse autor, interações sociais equivalem a um tipo particular de externalidades, onde ações de um grupo de referência afetam as preferências de um dado indivíduo. Em conjunto com a hipótese de complementariedades estratégicas (quando a utilidade de um indivíduo que segue com uma ação depende diretamente das ações tomadas, em média, por seus semelhantes), as interações sociais podem ajudar a explicar a ocorrência de significativas diferenças em termos de resultados obtidos, mesmo sem a ocorrência de diferenças nos fundamentos do modelo econômico em análise. O autor ainda fornece exemplos de aplicações relacionadas ao fenômeno, como criminalidade, emprego, formação de cidades, dentre outros

Marmaros e Sacerdote (2006) realizam um estudo empírico interessante, onde buscam entender o processo de origem e evolução de amizade entre estudantes universitários. Para tanto, baseando-se em uma

amostra composta pelo número de e-mails trocados entre estudantes de uma mesma universidade (Dartmouth), esses autores concluem que: (i) estudantes do primeiro ano de faculdade tendem a interagir com estudantes geograficamente próximos, formando laços de amizade de longo prazo com um subconjunto desses estudantes; (ii) fatores como proximidade geográfica e raça são mais importantes para a formação de amizades do que interesses em comum, curso de graduação ou *background* familiar; (iii) estudantes brancos apresentam quase três vezes mais chance de interação entre si do que estudantes de raças distintas; (iv) entretanto, caso um estudante branco e um estudante negro sejam colocados no mesmo dormitório da universidade durante o primeiro ano do curso de graduação, a probabilidade de ocorrência de interação entre eles triplica. Esses resultados são importantes não apenas no sentido de identificar possíveis padrões de homofilia e discriminação, podendo também ser úteis para a formulação de políticas de ação afirmativa, por exemplo.

Por sua vez, Costa e Kahn (2006) apresentam uma análise inovadora relacionada à mensuração da importância de redes sociais. No caso, a partir de uma amostra de prisioneiros da Guerra Civil norte-americana, os autores buscam quantificar a importância de laços sociais para a sobrevivência desses prisioneiros. Seus resultados demonstram que, em média, prisioneiros que tinham mais amigos na prisão também foram aqueles que apresentaram as maiores taxas de sobrevivência durante a guerra. Segundo os autores, fatores como etnia, camaradagem e origem dos prisioneiros foram fundamentais para aumentar suas chances de sobrevivência em meio a um ambiente hostil.

Hitsch, Hortaçsu e Ariely (2010) apresentam um interessante estudo baseado em dados referentes a um *website* de namoro *online*. Basicamente, esses autores tentam compreender a interação entre homens e mulheres em um ambiente virtual. Os resultados obtidos demonstram que, ao longo do processo de busca por um par, tanto homens quanto mulheres tendem a exagerar seus atributos físicos como forma de atrair um maior número de pretendentes. Além disso, os autores encontram resultados que demonstram

não haver evidências de comportamento estratégico entre homens e mulheres nesse tipo de ambiente, ao mesmo tempo em que há uma tendência de ambos os gêneros buscarem parceiros com atributos semelhantes, em média. Por outro lado, os autores constatam que, no caso da amostra analisada, as mulheres tendem, em média, a preferir atributos financeiros em relação a atributos físicos em uma proporção superior aos homens.

O estudo de Zentner (2006) equivale a uma análise dos impactos da introdução de mecanismos de compartilhamento de arquivos de música (*peer-to-peer*, P2P) sobre vendas da indústria fonográfica. Utilizando uma amostra referente a 15.000 consumidores europeus durante o ano de 2001, o autor estima os efeitos quantitativos de *downloads* de arquivos .mp3 sobre a probabilidade dos consumidores continuarem comprando músicas. Os resultados obtidos apontam para a existência de uma relação positiva e estatisticamente significativa entre essas variáveis. Ou seja, consumidores que baixam músicas na Internet, também tendem, em média, a comprar mais música. Esse resultado, apesar de surpreendente à primeira vista, demonstra a possibilidade de ocorrência de complementariedades entre *downloads* e compra efetiva de bens ofertados na Internet.

Jackson (2008) realiza uma resenha de parte da literatura teórica e empírica em Economia relacionada à formação de redes. São dois os objetivos básicos desse autor: primeiro, compreender como o ambiente social pode afetar decisões econômicas; segundo, entender como a análise econômica pode ser complementada a partir de análises de outra natureza como forma de se ampliar o escopo de estudos relacionados ao tema. Por exemplo, a formação de redes pode ser fundamental para o mercado de trabalho, uma vez que, dependendo das redes às quais um trabalhador tenha acesso, ele ou ela pode aumentar consideravelmente suas chances de obter um novo emprego. No caso, esse autor foca sua análise em dois tipos de modelos usualmente empregados na análise desse fenômeno: (i) modelos onde redes são vistas como originando de forma estocástica, sendo feito uso da teoria de grafos aleatórios; (ii) modelos onde as ligações de uma rede

(links) podem ser vistos como resultados de relações sociais ou econômicas escolhidos pelos indivíduos envolvidos, com uso de elementos de Teoria dos Jogos. O autor conclui seu trabalho ressaltando as diversas possibilidades de pesquisa relacionadas a interações sociais, assim como a necessidade de esforços multidisciplinares para sua melhor compreensão.

No caso brasileiro, pode-se citar o trabalho de Recuero (2004), que embora não esteja diretamente relacionado à área de Economia, constitui uma importante contribuição sobre o tema. Basicamente, essa autora realiza uma discussão sobre a insuficiência de alguns conceitos de rede para a explicação do fenômeno de formação de redes sociais mediadas via computador. Em particular, a autora discute os méritos e deficiências de três modelos relacionados ao tema: (i) “redes aleatórias” (a formação de conexões entre membros de uma mesma rede ocorre de forma aleatória); (ii) “mundos pequenos” (membros de uma mesma rede estariam ligados em algum grau)²; (iii) “redes sem escalas” (quanto maior o número de conexões de um membro de uma rede, maiores as chances de ele ter novas conexões)³. Para tanto, busca discutir a aplicabilidade desses modelos tendo como base de análise o *Orkut*, *Blogs* e *Flogs*. A autora conclui sua análise demonstrando que a maior parte dos modelos propostos, embora úteis para a análise de redes sociais, não apresentam um desempenho satisfatório quando da análise de redes formadas na Internet, dado o alto grau de complexidade dessas redes. Em particular, esses modelos apresentam significativas limitações na análise do fenômeno de formação de laços sociais entre indivíduos.

A partir das referências acima, buscar-se-á realizar uma análise exploratória de alguns dados relacionados ao *site Orkut* como uma primeira tentativa de análise de interações sociais mediadas via computador entre distintos indivíduos.

² A título de exemplo, Albert, Jeong e Barabási (1999) apresentam evidência favorável à formação de uma rede nesses moldes, no caso da *World Wide Web*.

³ No caso desse último tipo de modelo, as redes daí derivadas apresentariam uma distribuição em cauda longa, condizente com leis de potências (Huberman e Adamic 1999; Adamic 2002; Adamic e Huberman 2002). Esse ponto é explorado em maiores detalhes abaixo.

Sobre o Orkut

O *website* de relacionamentos *Orkut* foi lançado em janeiro de 2004 pelo engenheiro de *software* Orkut Büyükkökten, funcionário da *Google*. Originalmente, esse *site* havia sido criado para facilitar a comunicação entre membros da empresa, tendo sido disponibilizado em escala pública posteriormente. Atualmente, o *site* pertence à *Google*.

Apesar da administração do *site* ter sido localizada inicialmente no estado norte-americano da Califórnia, ocorreu uma transferência da sede oficial em agosto de 2008, com o escritório brasileiro da *Google* situado na cidade de Belo Horizonte passando a ser responsável pela administração do *site*, dado o grande número de usuários brasileiros⁴.

Na próxima seção do trabalho são detalhados a metodologia e a base de dados empregada na análise abaixo, ao mesmo tempo em que são apresentadas algumas das limitações relacionadas a esses dados.

Base de Dados

A base de dados utilizada neste trabalho equivale a uma amostra de 453 perfis de usuários do *Orkut*, coletada entre os meses de março e setembro do ano de 2005. Basicamente, os dados utilizados correspondem a informações detalhadas sobre os usuários desse *website* de relacionamentos, englobando características e preferências pessoais.

No intuito de se obter uma amostra aleatória, os dados foram obtidos a partir da ferramenta *Friend Finder*, disponível no próprio *Orkut*. No caso, optou-se por construir uma amostra de perfis a partir das sugestões disponíveis nessa ferramenta. O resultado desse processo foi a compilação de uma lista de perfis contendo as informações supracitadas.

Vale a ressalva inicial de que existem dois problemas relacionados a essa base de dados. Primeiro, mesmo com o objetivo de se obter uma amostra aleatória de dados do *website*, há a possibilidade de ocorrência de

⁴ Dados referentes ao ano de 2010 apontavam o Brasil como o país com o maior número de usuários no *Orkut*, com a Índia vindo em segundo lugar no *ranking*. Para maiores informações a esse respeito, ver <http://en.wikipedia.org/wiki/Orkut>. Maiores detalhes adiante.

viés de seleção, caso o *website* forneça sugestões automáticas de contatos aos usuários com base nos dados pessoais e/ou preferências de cada usuário⁵.

Segundo, há a possibilidade de surgimento de vieses originados a partir do fato de que as informações contidas nessa base de dados são, em sua grande maioria, fornecidas pelos próprios usuários. Esse fato pode levar a vieses relacionados a comportamentos e à descrição de atributos físicos dos usuários, assim como a construção de perfis falsos, por exemplo. Ainda assim, o uso dessa amostra pode vir a revelar importantes padrões referentes ao comportamento de usuários de redes sociais, com os problemas supracitados podendo ocorrer mesmo no caso de pesquisas de órgãos oficiais envolvendo dados de alcance universal⁶.

Resultados

Análise Descritiva

Algumas estatísticas disponíveis no *Orkut* atualmente podem ajudar a esclarecer alguns padrões empíricos iniciais relacionados a esse *site* de relacionamentos. A tabela 1 contém as participações de distintos países nesse *site*, de acordo com as respectivas participações relativas de usuários dessas nacionalidades:

Tabela 1
Percentual de Usuários do
Orkut por País, Dados
Populacionais

Pais	Participação Relativa
Brasil	50,60%
Índia	20,44%

⁵ O *website* do *Orkut* não possuía em 2005 explicações acerca do processo de sugestões de contatos do *FriendFinder*, não tendo havido respostas dos responsáveis pelo *site* aos autores em relação a questionamentos nesses moldes. Para tentar contornar a possibilidade de surgimento de vieses, foram consideradas sugestões de contatos fornecidas pela ferramenta *Friend Finder* tanto em ordem ascendente quanto descendente.

⁶ No caso, usuários do *website* podem fornecer informações falsas e/ou enganosas sobre atributos físicos, como cor de pele, por exemplo. Apesar desta ser uma potencial fonte de viés dos resultados reportados no presente trabalho, vale lembrar que situação semelhante ocorre no caso de pesquisas oficiais como o Censo Demográfico do Instituto Brasileiro de Geografia e Estatística (IBGE), por exemplo.

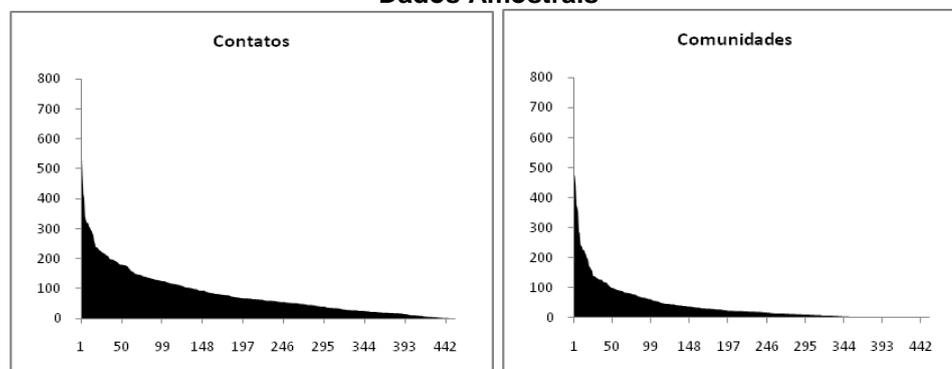
Estados Unidos	17,78%
Paquistão	0,86%
Paraguai	0,44%
Reino Unido	0,40%
Portugal	0,36%
Afeganistão	0,35%
Japão	0,34%
Canadá	0,33%

Fonte: *Orkut*
(www.orkut.com).

As informações contidas na tabela demonstram que o Brasil apresenta, de fato, a maior participação em termos de usuários registrados no *Orkut*, em torno de 51% do total. A Índia vem em segundo lugar, com uma participação equivalente a menos da metade desse valor (aproximadamente 20%), ao passo que os Estados Unidos apresentam uma participação de 18%. Por sua vez, os demais países da tabela apresentam participações inferiores a 1%.

O gráfico 1 equivale a um histograma descrevendo o número de contatos e comunidades dos usuários da amostra. O eixo vertical contém o número de contatos por usuário, enquanto o eixo horizontal apresenta os usuários da amostra. No caso, os usuários foram identificados a partir de códigos numéricos, como forma de proteger suas identidades. Adicionalmente, as escalas de ambos os gráficos foram padronizadas no sentido de facilitar a comparação de resultados.

Gráfico 1
Histogramas das Distribuições de Contatos e Comunidades dos Usuários do *Orkut*, Dados Amostrais



Fonte: Cálculos dos Autores, com base em informações do *Orkut*.

De acordo com os resultados reportados no gráfico, é possível notar que, no caso dos dados considerados, tanto o número de contatos quanto de comunidades exibem um padrão empírico de “cauda longa”, em conformidade com uma distribuição de Pareto. Ou seja, no caso do *Orkut*, nota-se que alguns poucos usuários possuem um grande número de contatos e/ou pertencem a um grande número de comunidades. Um padrão semelhante a esse também é verificado em outras situações relacionadas à Internet, como no caso da enciclopédia *online Wikipedia*, onde alguns poucos usuários fornecem diversas contribuições relevantes, enquanto que a grande maioria dos usuários fornece um número significativamente menor de informações (Huberman e Adamic 1999; Adamic e Huberman 2002; Ayres 2009)⁷.

Apesar de ambos os gráficos apresentarem o mesmo padrão qualitativo, nos moldes de uma distribuição de cauda longa, pode-se notar algumas diferenças, uma vez que esses gráficos possuem a mesma escala. Basicamente, nota-se que o gráfico referente a contatos dos usuários possui uma cauda mais “grossa” do que aquele referente ao número de comunidades às quais cada usuário pertence. Esses resultados provavelmente decorrem do fato de que, no caso do *Orkut*, os usuários tendem, em média, a valorizar mais o número de contatos que possuem do que o número de comunidades a que pertencem, sendo mais seletivos no último caso⁸. A identificação de padrões empíricos nesses moldes pode ser útil tanto em termos de políticas de *marketing* quanto de planejamento de esquemas de incentivo aplicados a contextos virtuais.

Conforme dito anteriormente, a amostra construída para o presente trabalho contém 453 perfis de usuários do *Orkut*. Desse total, 213 são reportados como sendo do sexo masculino (47% do total), ao passo que 240 são do sexo feminino (53%). No que segue, será feita uma análise dos itens

⁷ Para um resumo da evidência relacionada a distribuições de cauda longa e leis de potências, ver Adamic (2002) e Gabaix (2008). Anderson (2006) apresenta diversos exemplos relacionados a distribuições de cauda longa, com ênfase em fenômenos ocorridos na Internet.

⁸ Esse fato pode ser comprovado pela frase que aparecia no *Orkut* quando do início de suas atividades, no ano de 2004: “*Who do you know?*” (“Quem você conhece?”). Para uma breve análise de comportamentos relacionados à busca de popularidade de alguns usuários desse *site*, ver Recuero (2005).

referentes às características e preferências reportadas pelos usuários da amostra. Quando estiverem disponíveis informações referentes às estatísticas do próprio *Orkut*, será feita uma comparação entre resultados da amostra e do *site*. Caso contrário, quando essas informações não forem disponíveis, a análise será focada apenas nos itens disponíveis na amostra.

Hoje em dia, há pesquisas em Economia e Psicologia relacionadas à ocorrência de diferenças de comportamento entre homens e mulheres em distintos contextos (e.g., Antonovics, Arcidiacono e Walsh 2005). Uma questão interessante a ser investigada no presente contexto equivale à possível ocorrência de diferenças de comportamento em termos de gênero entre usuários do *Orkut*. Por conta disso, a análise desenvolvida abaixo procura checar essas diferenças a partir das preferências reportadas pelos próprios usuário do *site* de relacionamentos em análise.

A tabela 2 apresenta o percentual de usuários da amostra, de acordo com suas respectivas faixas etárias. As faixas etárias contidas na tabela foram escolhidas de acordo com as faixas etárias reportadas pelo *website* quando da divulgação de suas próprias estatísticas, tendo fins meramente comparativos. O número de usuários incluído na tabela foi menor do que o total da amostra original, uma vez que nem todos os usuários declaram idade quando do registro no *site*.

Tabela 2
Idade dos Usuários do *Orkut* (Participações Percentuais), Dados Amostrais

Faixa etária	Total		Masculino		Feminino	
	Nº	Partic.	Nº	Partic.	Nº	Partic.
18-25	129	28%	60	28%	69	29%
26-30	39	9%	20	9%	19	8%
31-35	9	2%	3	1%	6	3%
36-40	3	1%	3	1%		0%
41-50	6	1%	4	2%	2	1%
50+	2	0%		0%	2	1%
Sem declarar	265	58%	125	58%	140	59%
Total	453	42%	215	42%	238	100%

Fonte: Cálculos dos Autores, com base em informações do *Orkut*.

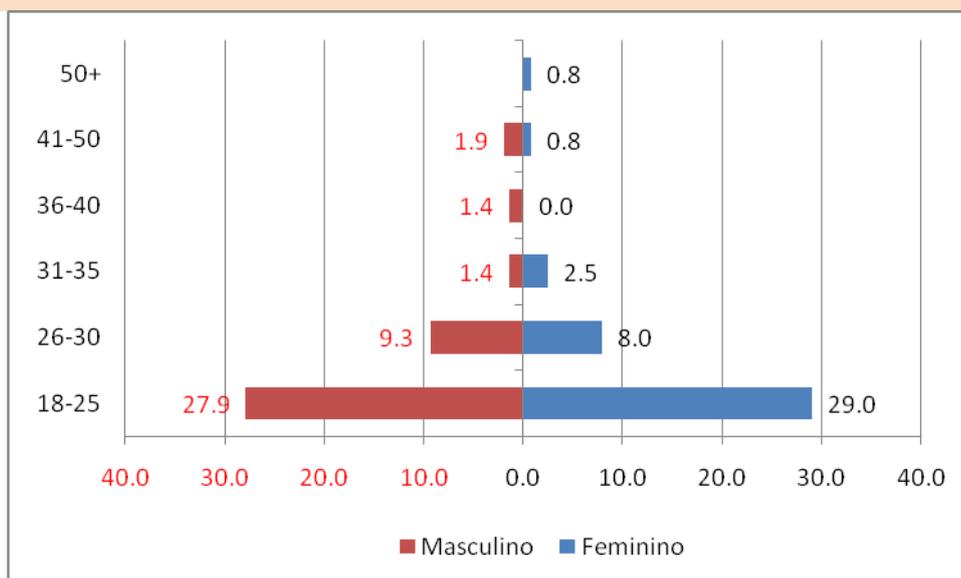
De acordo com os resultados reportados na tabela, nota-se que a maior parte dos usuários contidos na amostra e que reportam as respectivas idades

estão na faixa entre 18 e 25 anos (28%). Em seguida, vem um conjunto de usuários na faixa entre 26 e 30 anos, equivalente a menos de um terço do conjunto anterior (9%). No caso de faixas de idade mais elevadas (acima de 30 anos), nota-se uma ocorrência ainda menor de usuários, situada entre 1 e 3%, em geral. Interessante notar que, no caso de faixas mais jovens, os resultados referentes à amostra como um todo são praticamente replicados tanto para homens quanto para mulheres. Entretanto, no caso de idades mais avançadas, nota-se que há o triplo de mulheres na faixa entre 31 e 35 anos em comparação a homens da mesma idade (3% contra 1%), enquanto que, no caso da faixa entre 41 e 50 anos ocorre uma situação inversa, com um número de usuários do sexo masculino duas vezes superior ao número de usuários do sexo feminino (2% contra 1%).

Em termos gerais, os resultados aqui reportados permitem inferir que, no caso da amostra em análise, a maioria dos usuários (cerca de 40%) encontra-se na faixa entre 18 e 30 anos, o que permite caracterizar esse conjunto como um grupo relativamente jovem.

O gráfico 2, por sua vez, confirma esses resultados, ao considerar as participações relativas de homens e mulheres na amostra, de acordo com as respectivas faixas etárias, nos moldes de uma pirâmide etária. De acordo com esse gráfico, nota-se que, enquanto há certo equilíbrio entre a proporção de homens e mulheres para faixas etárias mais jovens, o mesmo não ocorre para faixas mais velhas. Por exemplo, no caso da faixa entre 36 e 40 anos, nota-se um predomínio exclusivo de usuários do sexo masculino, com um padrão semelhante ocorrendo no caso da faixa acima de 50 anos, onde o total dos usuários pertence ao sexo feminino. Por outro lado, a pirâmide etária construída demonstra que os usuários da amostra constituem um grupo jovem, uma vez que a base da pirâmide é larga em comparação ao ápice.

Gráfico 2
Pirâmide Etária Relativa de Usuários do Orkut, Dados Amostrais



Fonte: Cálculos dos Autores, com base em informações do *Orkut*.

A tabela 3, por sua vez, apresenta resultados referentes aos interesses dos usuários do *Orkut*. No caso, são apresentados os percentuais de usuários que, ao se registrarem no *site*, marcam as alternativas relacionadas a seus principais objetivos de participação nessa rede social. Vale a observação de que, uma vez que cada usuário pode marcar mais de uma opção, há a possibilidade de que os totais de cada categoria possam vir a somar mais de 100%, em alguns casos.

Tabela 3
Interesses dos Usuários do *Orkut*, Dados Amostrais

Interesse	Total		Masculino		Feminino	
	Nº	Partic.	Nº	Partic.	Nº	Partic.
Amizade	344	99%	161	99%	183	99%
Namoro	51	15%	29	18%	22	12%
Parceiros de Atividades	106	31%	53	33%	53	29%
Contatos Profissionais	114	33%	57	35%	57	31%
Total	347	177%	162	185%	185	170%

Fonte: Cálculos dos Autores, com base em informações do *Orkut*.

Os resultados obtidos a partir da tabulação acima demonstram que a quase totalidade dos usuários (99%) registra-se no *site* com a intenção de fazer amigos, com esse resultado sendo válido tanto no caso de homens quanto de mulheres. Os resultados são semelhantes em termos de busca de parceiros de atividades assim como em termos de contatos profissionais (em torno de 30%), também válidos para ambos os sexos.

Por sua vez, a tabela 4 contém dados referentes ao total de usuários do *Orkut* (estatísticas do próprio *site*). Novamente, é possível notar que, apesar de algumas diferenças em termos de magnitudes, a ampla maioria dos usuários desse *site* registra estar interessada em amigos (91%), com os interesses relacionados a parceiros de atividades e contatos profissionais equivalendo a cerca de um terço desse valor (em torno de 30%), resultados em consonância com aqueles referentes à amostra utilizada neste trabalho.

Tabela 4
Interesses do *Orkut*, Dados Populacionais

Interesses	Partic.
Amigos	91%
Parceiros de Atividades	30%
Contatos Profissionais	33%
Namoro	18%
Total	172%

Fonte: *Orkut* (www.orkut.com).

Em relação às estatísticas supracitadas, chama atenção as diferenças relacionadas ao item “namoro”. No caso da amostra considerada, o número de homens que admite ter entrado no *site* com fins de conhecer outros usuários para namoro é cerca de 1,5 vezes superior ao número de mulheres (18% contra 12%). À primeira vista, esse resultado é interessante por apontar a ocorrência de diferenças em termos de preferências entre homens e mulheres que participam dessa rede social. No caso do total referente ao *website*, vale notar que há o registro de uma proporção de mesma magnitude que aquela referente aos usuários de sexo masculino da amostra.

As tabelas 5 e 6 expõem resultados referentes ao estado civil dos usuários da amostra e do *Orkut*, respectivamente.

Tabela 5
Estado Civil dos Usuários do *Orkut*, Dados Amostrais

Estado Civil	Total		Masculino		Feminino	
	Nº	Partic.	Nº	Partic.	Nº	Partic.
Solteiro	214	47%	110	61%	104	44%
Namorando	106	23%	48	27%	58	24%
Não há resposta	84	19%	35	19%	49	21%
Casado	39	9%	16	9%	23	10%
Relacionamento aberto	10	2%	6	3%	4	2%

Total	453	100%	180	100%	238	100%
--------------	------------	-------------	------------	-------------	------------	-------------

Fonte: Cálculos dos Autores, com base em informações do *Orkut*.

Tabela 6
Estado Civil dos Usuários do Orkut,
Dados Populacionais

Estado Civil	Partic.
Solteiro(a)	48%
Namorando	18%
Não há resposta	18%
Casado(a)	13%
Relacionamento Aberto	2%
Total	100%

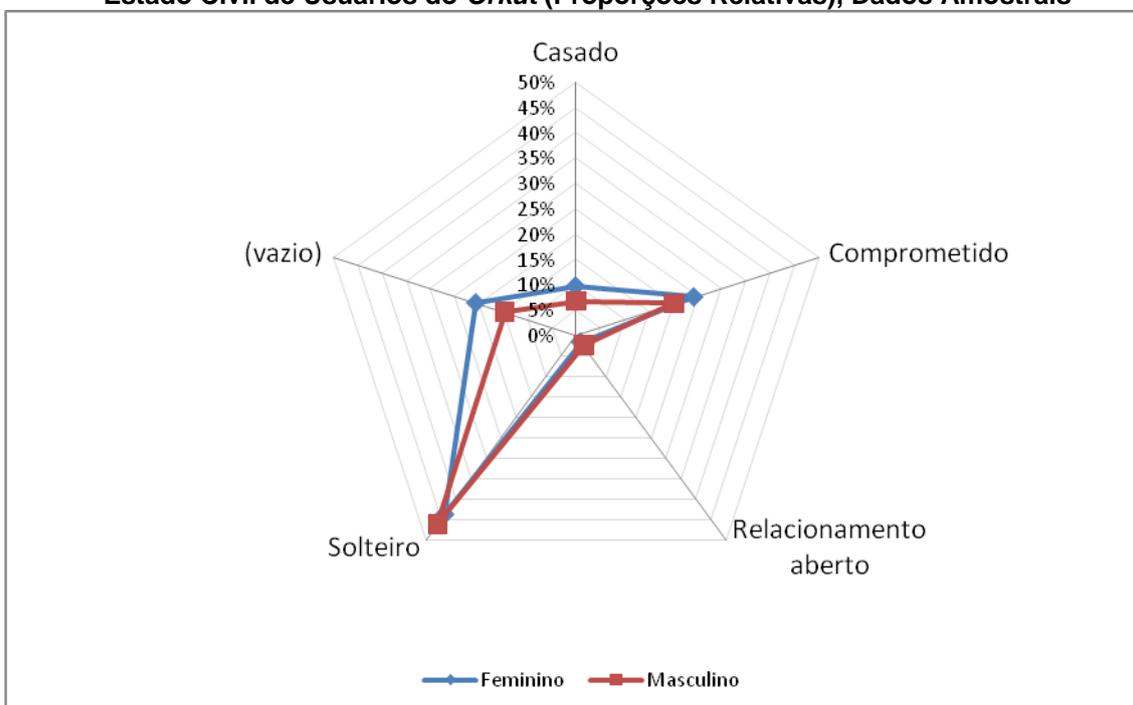
Fonte: Orkut (www.orkut.com).

De acordo com esses resultados, pode-se constatar que cerca de metade dos usuários (47%), tanto na amostra quanto na população do *website* são solteiros, um resultado decorrente provavelmente do fato de que a ampla maioria dos usuários pertence a faixas etárias relativamente jovens (entre 18 e 30 anos). Por outro lado, ocorrem diferenças entre a amostra e a população em relação às demais categorias. Enquanto a proporção de usuários namorando corresponde a 23% na amostra, essa proporção é menor na população, equivalendo a 18%. Do mesmo modo, ocorrem diferenças em termos de percentuais de usuários casados: na população, há cerca de 13% de usuários casados, enquanto que na amostra essa proporção fica em torno de 9%. Vale notar ainda que, em relação às demais categorias (“relacionamento aberto” e “não há resposta”), as proporções são praticamente idênticas.

Apesar da ocorrência de disparidades em relação a algumas categorias referentes ao estado civil dos usuários, é interessante comparar as diferenças entre homens e mulheres na amostra, mais uma vez. No caso, nota-se que os usuários de sexo masculino tendem, em média, a apresentar maiores magnitudes em relação às proporções referentes às categorias “solteiro” (61%) e “relacionamento aberto” (3%), ao passo que uma situação inversa ocorre nas categorias “namorando” (27%) e “casado” (9%), onde essas magnitudes são inferiores àquelas reportadas para usuários de sexo feminino. Esses resultados sugerem que, pelo menos no caso da amostra analisada, os homens tendem a fornecer informações que os identificam como disponíveis para relacionamentos em uma escala superior às mulheres.

O gráfico 3 corresponde a uma forma alternativa de reportar os resultados da tabela anterior, confirmando os padrões descritos. No caso, a linha azul representa o sexo feminino, ao passo que a linha vermelha representa o sexo masculino. O padrão gráfico descrito acima aponta para uma participação maior das mulheres nas categorias “comprometido” e “casado”, com um padrão semelhante ocorrendo para os homens no caso das categorias “solteiro” e “relacionamento aberto”, o que confirma os resultados descritos anteriormente.

Gráfico 3
Estado Civil de Usuários do Orkut (Proporções Relativas), Dados Amostrais



Fonte: Cálculos dos Autores, com base em informações do Orkut.

As próximas tabelas correspondem a tabulações referentes a características pessoais fornecidas pelos usuários da amostra. No caso dessas informações, não há possibilidade de comparação com dados populacionais oriundos do Orkut, uma vez que, na época de construção da base de dados, o website não compilava estatísticas referentes aos itens descritos abaixo. Basicamente, as tabelas fazem referência a itens como etnia, religião, orientação política e hábitos de bebida e fumo dos usuários.

A tabela 7 apresenta resultados relacionados à etnia dos usuários da amostra. Conforme ressaltado acima, esses resultados devem ser vistos com

cautela, uma vez que equivalem a informações fornecidas diretamente pelos usuários do *site* de relacionamentos, podendo apresentar vieses por conta disso.

Tabela 1
Etnia dos Usuários do Orkut, Dados Amostrais

Etnia:	Total		Masculino		Feminino	
	Nº	Partic.	Nº	Partic.	Nº	Partic.
Caucasiano (Branco)	115	50%	47	42%	68	57%
Hispanico/Latino	52	23%	33	30%	19	16%
Outro	26	11%	12	11%	14	12%
Multiétnico	21	9%	11	10%	10	8%
Afro-Brasileira	8	3%	5	5%	3	3%
Asiático	7	3%	3	3%	4	3%
Oriental	1	0%		0%	1	1%
Americano	1	0%		0%	1	1%
Total	231	100%	111	100%	120	100%

Fonte: Cálculos dos Autores, com base em informações do Orkut.

Os resultados contidos nessa tabela demonstram que cerca de metade dos usuários da amostra considerada se reportam como caucasianos (50%). Em seguida, vem um grupo que se assume como hispânico/latino, com a proporção referente a esse grupo equivalendo a menos da metade da etnia anterior (23%). As categorias “outro” e “multiétnico” aparecem com proporções na faixa entre 9 e 11%⁹, ao passo que as demais etnias apresentam valores inferiores a 5%. Em particular, chama atenção o fato de que, dentre os usuários pesquisados, há uma proporção de apenas 3% que se registra como afro-brasileira (negra).

Em relação a diferenças entre gêneros, nota-se que, no caso de etnia, homens e mulheres apresentam distintas informações. Assim, tem-se que, dentre os usuários do sexo masculino, cerca de 42% se dizem caucasianos, 30% se dizem hispânicos/latinos, com cerca de 23% se identificando como pertencentes às categorias “outro” e “multiétnico”, em conjunto. No caso desse conjunto de usuários, as frações pertencentes às etnias afro-brasileira e asiática equivalem a 5% e 3%, respectivamente.

⁹ É possível que usuários que registrem essas duas categorias pertençam, na realidade, a uma mesma categoria, referente a pessoas de raça mestiça. Uma situação semelhante parece ocorrer em relação a outras opções de preferências do Orkut, o que pode vir a enviesar resultados de análises relacionadas a fatores como etnia, religião e visão política, por exemplo. Embora exploratória no presente contexto, essa hipótese merece uma investigação mais detalhada, servindo de sugestão para a pesquisa futura.

Por outro lado, no caso de usuários do sexo feminino, nota-se que mais da metade da amostra corresponde a usuárias de etnia branca (57%), com 16% desses usuários pertencendo à etnia hispânica/latina. Interessante notar que, enquanto a proporção de usuários caucasianos é maior nesse grupo, a proporção de latinos é nitidamente menor do que no grupo masculino. A proporção de negros também é menor nesse grupo (3%), correspondendo a quase metade da proporção referente ao grupo masculino. Por sua vez, a proporção de asiáticos no grupo feminino é superior ao masculino, estando em torno de 3%.

A título de comparação, a tabela 8 apresenta dados referentes à população brasileira, oriundos do Censo Demográfico do ano de 2000. Embora as categorias referentes à etnia da amostra não sejam as mesmas do Censo, o intuito básico desse exercício de comparação equivale a verificar a ocorrência de diferenças entre as preferências reportadas pelos usuários selecionados e as características reportadas por membros da população.

Tabela 8
População residente por cor ou raça - Brasil - ano 2000

Cor ou raça	População residente (Percentual)		
	Total	Homens	Mulheres
Total	100,00	49,21	50,79
Branca	53,74	25,80	27,95
Preta	6,21	3,23	2,98
Parda	38,45	19,40	19,05
Amarela	0,45	0,22	0,23
Indígena	0,43	0,22	0,22
Sem declaração	0,71	0,35	0,36

Fonte: Instituto Brasileiro de Geografia e Estatística (IBGE), Censo Demográfico.

Quando da comparação dos resultados, nota-se que, embora as proporções de homens e mulheres na amostra sejam semelhantes àquelas reportadas no censo (47% contra 49% e 53% contra 51% para homens e mulheres, respectivamente), há diferenças consideráveis em relação aos dados referentes a etnia ou raça. Por exemplo, enquanto a proporção de pessoas de cor branca na população brasileira é próxima de 54%, a proporção na amostra equivale a 50%. Entretanto, as diferenças entre homens e mulheres tendem a ser maiores em relação a essa categoria: na amostra, 42% dos usuários de sexo masculino se dizem brancos, enquanto

que, na população, a proporção equivalente corresponde a 26%, aproximadamente. Situação semelhante ocorre para usuários de sexo feminino, cuja proporção de etnia caucasiana na amostra equivale a 57%, enquanto que na população brasileira, essa proporção corresponde à metade do último valor (28%).

Por outro lado, em relação à categoria afro-brasileira, percebe-se uma maior semelhança entre os resultados referentes à amostra e ao Censo, embora o percentual de usuários do sexo masculino pertencentes a essa categoria seja maior na amostra (5% contra 3%). Resultados referentes a outras categorias são razoavelmente semelhantes entre a amostra e os dados populacionais, embora a ocorrência de diferenças entre categorias não permita uma comparação direta, no caso.

Esses resultados chamam atenção para o fato de que, em termos de raça/etnia, podem ocorrer diferenças entre características referentes à população como um todo e as características reportadas pelos próprios usuários da amostra, um resultado em consonância com estudos relacionados a outros contextos (e.g., Hitsch, Hortaçsu e Ariely 2010). Desde já, fica a sugestão que a pesquisa futura busque compreender as razões inerentes a essas diferenças.

A tabela 9 contém dados referentes à religião dos usuários da amostra. No caso desses resultados, nota-se que cerca de metade (53%) dos usuários se reporta como “cristão/católico”, com essa proporção sendo superior no caso do grupo feminino (61%) do que no caso masculino (44%). Em seguida, aparece a categoria “cristão/protestante”, com uma participação cerca de quatro vezes inferior à primeira categoria (13%). Entretanto, ao contrário da categoria anterior, nota-se neste caso uma preponderância do grupo masculino sobre o feminino, com o primeiro apresentando uma proporção em torno de 17%, ao passo que o segundo apresenta uma proporção em torno de 10%, apenas. Embora as demais categorias referentes à religião dos usuários também apresentem diferenças em termos de magnitudes, esses resultados são interessantes no sentido de demonstrarem a ocorrência de diferenças quantitativas em relação às distintas religiões, bem como diferenças em

termos de gênero. Nesse sentido, fica a sugestão de elaboração de estudos adicionais que buscassem explorar esses padrões empíricos de forma mais detalhada, assim como a evolução das preferências religiosas dos usuários ao longo do tempo.

Tabela 9
Religião dos usuários do Orkut, Dados Amostrais

Religião:	Total		Masculino		Feminino	
	Nº	Partic.	Nº	Partic.	Nº	Partic.
Cristão/Católico	155	53%	54	44%	101	61%
Cristão/Protestante	38	13%	21	17%	17	10%
Lado espiritual independente de religião	37	13%	14	11%	23	14%
Outro	23	8%	15	12%	8	5%
Cristão/Outro	19	7%	9	7%	10	6%
Ateu	7	2%	7	6%		0%
Agnostico	5	2%	3	2%	2	1%
Judeu	3	1%		0%	3	2%
Budista	1	0%		0%	1	1%
Jainismo	1	0%	1	1%		0%
Humanista religioso	1	0%		0%	1	1%
Total	290	100%	124	100%	166	100%

Fonte: Cálculos dos Autores, com base em informações do Orkut.

A tabela 10 expõe resultados relacionados à visão política dos usuários.

Neste caso, nota-se que 37% dos usuários contidos na amostra declaram-se “não-políticos”, com essa proporção mantendo-se aproximadamente tanto no caso de homens (37%) quanto de mulheres (38%). Em seguida, vem a categoria “depende”, com uma proporção em torno de 29% no caso da amostra total e de 27% e 31% para homens e mulheres, respectivamente. As diferenças entre ambos os sexos são menores no caso das demais alternativas, exceto pelas categorias “centro” e “libertário”. Em última instância, esses resultados demonstram que, no caso da amostra considerada, mais da metade dos usuários (proporção superior a 60%) não tem uma visão política definida.

Tabela 10
Visão Política dos Usuários do Orkut, Dados Amostrais

Visão Política:	Total		Masculino		Feminino	
	Nº	Partic.	Nº	Partic.	Nº	Partic.
Não político	48	37%	26	37%	22	38%
Depende	37	29%	19	27%	18	31%
Esquerda liberal	17	13%	10	14%	7	12%
Centro	10	8%	5	7%	5	9%
Libertário	9	7%	6	8%	3	5%
Direita conservador	4	3%	3	4%	1	2%
Esquerda muito liberal	2	2%	1	1%	1	2%
Autoritário	1	1%		0%	1	2%
Muito Autoritário	1	1%	1	1%		0%

Total	129	100%	71	100%	58	100%
--------------	------------	-------------	-----------	-------------	-----------	-------------

Fonte: Cálculos dos Autores, com base em informações do *Orkut*.

As tabelas 11 e 12 expõem resultados relacionados a hábitos de bebida (tabela 11) e cigarros (tabela 12).

Tabela 11
Hábitos de Bebida dos Usuários do *Orkut*, Dados Amostrais

Bebe:	Total		Masculino		Feminino	
	Nº	Partic.	Nº	Partic.	Nº	Partic.
Socialmente	151	42%	68	40%	83	43%
Não Bebe	117	32%	49	28%	68	36%
Eventualmente	70	19%	35	20%	35	18%
Regularmente	16	4%	13	8%	3	2%
Excessivamente	9	2%	7	4%	2	1%
Total	363	100%	172	100%	191	100%

Fonte: Cálculos dos Autores, com base em informações do *Orkut*.

No caso da primeira tabela, nota-se que 42% dos usuários contidos na amostra admite beber apenas socialmente, com 32% assumindo não consumir bebidas alcoólicas¹⁰. Por outro lado, 19% dos usuários admite beber eventualmente, enquanto que apenas 4% admitem fazê-lo regularmente. Apenas 2% dos usuários admite beber excessivamente.

As diferenças em termos desse hábito social ficam mais evidentes quando ocorre uma divisão por gênero. Neste caso, nota-se que, enquanto uma proporção razoavelmente semelhante de homens e mulheres admite beber social ou eventualmente (proporções aproximadas de 40% e 20%, respectivamente), as diferenças são quantitativamente maiores no caso de categorias equivalentes a extremos da amostra. Assim, em relação à categoria “não bebe”, nota-se que, enquanto essa proporção equivale a 28% no caso dos homens, ela chega a 36% no caso das mulheres, uma diferença de sete pontos percentuais (p.p.). Por outro lado, no caso da categoria “regularmente”, a proporção masculina equivale a um valor quatro vezes superior à proporção feminina (8% contra 2%). No caso da categoria “excessivamente”, tem-se um resultado semelhante, onde 4% dos homens admitem pertencer a essa categoria, com apenas 1% das mulheres admitindo o mesmo. Esses resultados são interessantes por apontarem para a

¹⁰ Um problema com essa divisão de categorias equivale ao fato de que um usuário qualquer pode considerar as alternativas “socialmente” e “eventualmente” como categorias substitutas entre si. Ou ainda, pode vir a marcar ambas simultaneamente, com tais possibilidades podendo enviesar os resultados obtidos.

existência de nítidas diferenças em termos de hábitos sociais entre homens e mulheres ou, pelo menos, por demonstrar diferenças em termos de auto-imagem dos usuários do *site* de relacionamentos em análise.

Em relação ao hábito social de fumar, nota-se que, no caso da amostra considerada, a ampla maioria dos usuários (87%) admite não fumar, com essa proporção sendo menor para homens (84%) do que para mulheres (90%). Por outro lado, quando da comparação dentre as demais categorias, nota-se um padrão onde as proporções dos homens tendem a ser maiores do que aquelas referentes às mulheres. Em termos gerais, esses resultados demonstram que, apesar do hábito de fumar não ser admitido como parte da rotina da maioria dos usuários analisados, esse parece ser um hábito mais disseminado entre usuários de sexo masculino, pelo menos no caso da amostra analisada.

Tabela 12
Hábitos de Fumo dos Usuários do *Orkut*, Dados Amostrais

Cigarro:	Total		Masculino		Feminino	
	Nº	Partic.	Nº	Partic.	Nº	Partic.
Não	314	87%	143	84%	171	90%
Regularmente	14	4%	8	5%	6	3%
Socialmente	13	4%	9	5%	4	2%
Ocasionalmente	8	2%	5	3%	3	2%
Bastante	4	1%	1	1%	3	2%
Parou	4	1%		0%	4	2%
Tentando parar	4	1%	4	2%		0%
Total	361	100%	170	100%	191	100%

Fonte: Cálculos dos Autores, com base em informações do *Orkut*.

Em termos gerais, a partir dos resultados descritos nesta seção, pode-se notar que o típico usuário do *Orkut*, de acordo com a amostra analisada, pode ser classificado da seguinte maneira:

- (i) Possui entre 18 e 30 anos;
- (ii) Tem como principal interesse no *site* de relacionamentos a formação de novas amizades;
- (iii) É solteiro(a), com essa proporção sendo maior entre usuários do sexo masculino;

- (iv) Declara-se como pertencente à etnia caucasiana/branca, mesmo ocorrendo diferenças em relação a estatísticas oficiais referentes à população brasileira como um todo;
- (v) É católico(a);
- (vi) Declara não ter visão política definida;
- (vii) Não fuma, mas bebe socialmente.

Enquanto esses resultados equivalem a uma descrição genérica inicial do perfil da maioria dos usuários da amostra em questão, faz-se, em seguida, uma análise econométrica dos principais fatores relacionados ao aumento do número de contatos por parte dos usuários do *Orkut*. A tabela 13 contém resultados de estimativas relacionando variáveis específicas ao número de contatos de cada usuário da amostra considerada. O intuito básico desse exercício empírico equivale a verificar o grau de associação quantitativa entre o número de contatos dos usuários da amostra e variáveis que possam afetar esse número. Em termos gerais, quer-se verificar quais fatores afetam a popularidade dos usuários do *Orkut*, com o fator popularidade sendo representado pelo número de contatos de cada usuário.

No caso dessa tabela, não há a pretensão de se identificar padrões causais entre as variáveis, mas, conforme dito acima, apenas fornecer primeiras evidências relacionadas ao grau de associação entre as mesmas. Em particular, as especificações foram estimadas para variáveis em escala logarítmica natural, o que faz com que os coeficientes daí resultantes equivalham a medidas de elasticidade¹¹. Adicionalmente, os erros-padrão das estimativas foram obtidos a partir do método de White (White 1980) de forma a se evitar problemas de heterocedasticidade, comuns a dados de cortes seccionais, conforme o presente caso. Os tamanhos das amostras consideradas para cada especificação variam, uma vez que nem todos os usuários apresentam informações referentes a todas as variáveis em questão.

¹¹ Para um exemplo de uso de especificações econométricas com sentido meramente ilustrativo, relacionado ao contexto de cidades, ver Glaeser (1998, p.143, Tabela 1).

Tabela 13
Estimações Econométricas: Determinantes do Número de Contatos dos Usuários do Orkut, Dados Amostrais

Variável dependente:
Log(Contatos)

Variáveis	Coeficientes (Erro padrão)		
	Equação 1	Equação 2	Equação 3
Constante	1,426*** (0,097)	1,215*** (0,109)	-5,808 (4,067)
Log(Fãs)	0,934*** (0,032)	0,596*** (0,048)	0,652*** (0,079)
<i>Dummy Sexo</i>	0,190*** (0,049)	0,175*** (0,047)	0,191*** (0,070)
Log(comunidades)		0,109*** (0,026)	0,119*** (0,038)
Log(Recados)		0,211*** (0,041)	0,202** (0,081)
Log(Idade)			4,210* (2,467)
Log(Idade) ²			-0,645* (0,373)
R ² Ajustado	0,8115	0,8153	0,8569
Número de Observações	430	371	163

Fonte: Cálculos dos Autores, com base em informações do *Orkut*.
 Observações:

- (a) Erros-padrão estimados são reportados entre parênteses. Os erros-padrão foram obtidos a partir do método de White (White 1980).
 (b) Os símbolos (*), (**) e (***) denotam significância estatística dos coeficientes estimados aos níveis de 10%, 5% e 1%, respectivamente.

De acordo com os resultados obtidos, nota-se que, mesmo no caso de distintas especificações econométricas, alguns resultados são robustos, no sentido de que as variáveis incluídas nas regressões em questão mostram-se estatisticamente significativas e com valores aproximadamente semelhantes. Adicionalmente, o ajuste quantitativo das especificações consideradas pode ser tido como bom, uma vez que os valores dos coeficientes de determinação ficam na faixa entre 0,81 e 0,86 (os modelos em questão explicam mais de 80% da variação observada nos dados).

No caso do número de fãs dos usuários, os resultados obtidos apontam para uma relação positiva e estatisticamente significativa, embora possam ocorrer variações nas magnitudes dos coeficientes estimados. Em particular, um aumento de 1% no número de fãs tende a gerar, em média, um aumento inferior no número de contatos (em torno de 0,6% no caso da maioria das especificações consideradas).

Quando considerada uma variável *dummy* representando o sexo dos usuários (que assume valor 1 para usuários do sexo masculino e valor 0 em caso contrário), nota-se que essa variável é estatisticamente significativa no caso de todas as especificações consideradas. Esse resultado demonstra que, em média, usuários do sexo masculino tendem a possuir um maior número de contatos. Este relacionado pode estar relacionado com as preferências expressas por usuários desse sexo do *Orkut*, conforme citado acima, quando da análise descritiva da amostra.

A variável “recados” também mostra-se como estatisticamente significativa. No caso, um aumento de 1% nos recados recebidos por um usuário do *Orkut* tende, em média, a levar a um aumento de 0,2% no número de contatos. Vale a ressalva de que esse resultado pode refletir problemas de endogeneidade, uma vez que a presente análise não distingue a direção de causalidade entre contatos e recados. A princípio, é possível que usuários que tenham um maior número de contatos também recebam um maior número de recados. De qualquer maneira, fica a sugestão de pesquisa futura relacionada à estimação de especificações utilizando variáveis instrumentais como forma de identificar efeitos causais entre as variáveis supracitadas.

A variável “idade” possui significância estatística nas especificações consideradas. No caso dessa variável, nota-se que o número de contatos tende a aumentar com a idade dos usuários: um aumento de 1% na idade dos usuários tende a gerar, em média, um aumento de 4% no número de contatos. Por outro lado, nota-se que esse aumento no número de contatos ocorre a taxas decrescentes, o que fica evidenciado pelo sinal negativo da variável “idade ao quadrado”. Essa última variável foi incluída nas estimações como forma de captar efeitos não-lineares do ciclo de vida dos usuários.

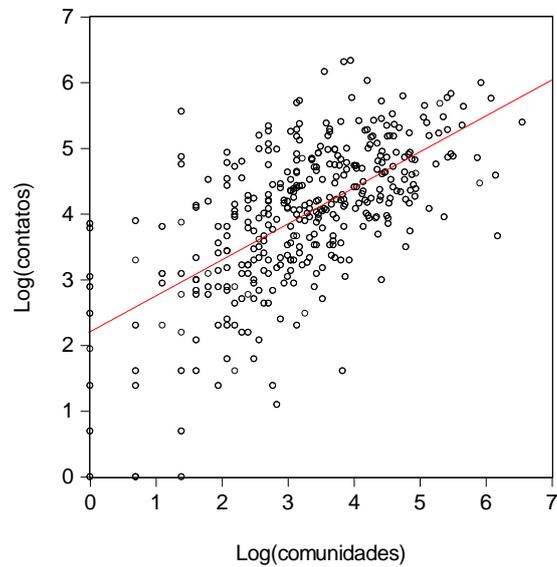
No caso da variável “comunidades”, os resultados obtidos demonstram que seu sinal aparece como positivo em todas as especificações consideradas. Basicamente, no caso das especificações onde essa variável é estatisticamente significativa, nota-se que um aumento de 1% no número de comunidades que um usuário do *Orkut* participa tende a aumentar, em média, seu número de contatos em 0,11%.

Esse último resultado é interessante por demonstrar que, embora a participação em comunidades do *site* de relacionamentos em análise tenda a aumentar o número de contatos de cada usuário, esse aumento ocorre em uma escala inferior ao aumento no número de comunidades. A princípio, um resultado nesses moldes pode demonstrar que a maior parte dos usuários da amostra considerada estabelece novos contatos a partir de conhecimento pessoal prévio, embora esta seja uma hipótese exploratória no momento. Também vale a ressalva de que, assim como no caso da variável “recados”, as estimativas relacionadas à variável “comunidades” também podem apresentar problemas de endogeneidade, uma vez que usuários com muitos contatos podem tender a participar de muitas comunidades ou vice-versa.

Os gráficos 4 a 6 apresentam resultados relacionados às variáveis descritas nas especificações acima, a partir de diagramas de dispersão contendo as variáveis analisadas em pares. Adicionalmente, cada diagrama apresenta uma reta de regressão estimada a partir do Método de Mínimos Quadrados Ordinários (MMQO), como forma de se verificar o ajuste dessa reta aos dados.

Gráfico 4

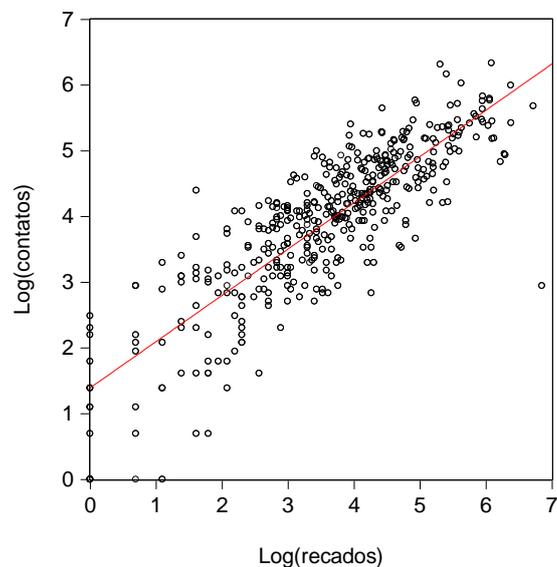
Diagrama de Dispersão: Contatos x Comunidades, Dados Amostrais



Fonte: Cálculos dos Autores, com base em informações do *Orkut*.

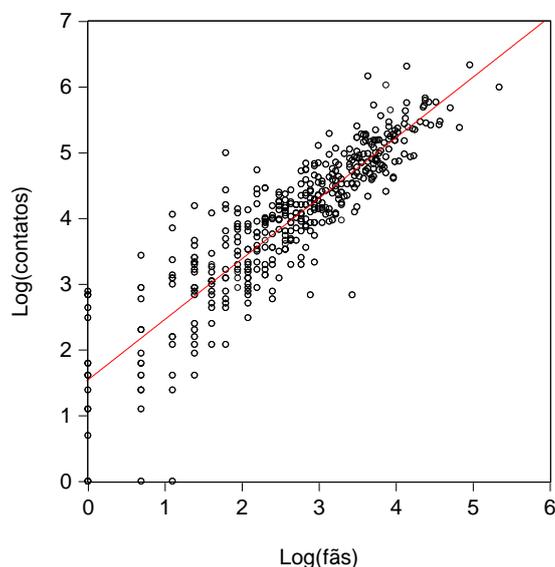
Gráfico 5

Diagrama de Dispersão: Contatos x Recados, Dados Amostrais



Fonte: Cálculos dos Autores, com base em informações do *Orkut*.

Gráfico 6
Diagrama de Dispersão: Contatos x Fãs, Dados Amostrais



Fonte: Cálculos dos Autores, com base em informações do *Orkut*.

Conforme exposto nesses gráficos, nota-se uma relação positiva entre contatos e comunidades, recados e fãs, com o ajuste da reta de regressão estimada sendo melhor no caso da última variável (fãs). Esses resultados são confirmados a partir da inspeção da tabela 14, que apresenta resultados de regressões relacionando as variáveis supracitadas, tomadas duas a duas.

Tabela 14
Estimativas de Elasticidade para Número de Contatos dos Usuários do Orkut, Dados Amostrais

Variável dependente: Log(contatos)			
	1	2	3
Log(Comunidades)	0,547*** (0,042)		
Log(Recados)		0,704*** (0,029)	
Log(Fãs)			0,921*** (0,032)
R ² ajustado	0,386	0,712	0,782
Observações	382	440	434

Fonte: Cálculos dos Autores, com base em informações do *Orkut*.

Uma questão interessante a ser respondida no presente contexto é a seguinte: há alguma relação entre os usuários que mais tem contatos e

aqueles que pertencem ao maior número de comunidades? As tabelas 15 e 16 buscam responder essa questão. No caso, ambas as tabelas expõem *rankings* contendo resultados relacionados aos dez usuários com os maiores números de contatos e comunidades, respectivamente. A título de comparação, cada tabela reporta o número referente a outra variável em análise. Assim, por exemplo, a primeira linha da tabela 15 contém o primeiro lugar do *ranking* em termos de contatos (562 contatos), acompanhado de sua respectiva posição no *ranking* de comunidades (52 comunidades). Um raciocínio semelhante vale para os resultados contidos na tabela 16.

Tabela 15
10 Maiores Usuários em Número de Contatos do Orkut,
Dados Amostrais

<i>Ranking</i> Contatos	Total de Contatos	<i>Ranking</i> Comunidades	Total de Comunidades
1	562	111	52
2	552	119	47
3	474	155	35
4	414	92	67
5	401	5	376
6	340	12	238
7	328	44	115
8	319	15	225
9	319	110	53
10	316	4	438

Fonte: Cálculos dos Autores, com base em informações do *Orkut*.

Tabela 16
10 Maiores Usuários em Número de Comunidades do Orkut,
Dados Amostrais

<i>Ranking</i> Comunidades	Total de Comunidades	<i>Ranking</i> Contatos	Total de Contatos
1	702	27	220
2	482	291	39
3	469	138	98
4	438	10	316
5	376	5	401
6	368	154	87
7	355	94	128
8	288	16	279
9	282	32	210
10	246	89	131

Fonte: Cálculos dos Autores, com base em informações do *Orkut*.

De acordo com os resultados reportados nessas tabelas, é possível notar que não há uma relação direta entre as variáveis consideradas. Ou seja, os usuários que possuem o maior número de contatos não são necessariamente aqueles que pertencem ao maior número de comunidades, pelo menos no caso do *ranking* dos dez maiores usuários de cada tipo.

A tabela 17, por sua vez, busca retratar a relação entre as variáveis supracitadas no caso dessas amostras específicas, assim como no caso da amostra completa. Para tanto, são reportados os coeficientes de correlação linear estimados para essas variáveis.

Tabela 17
Coefficientes de Correlação Linear entre
Contatos e Comunidades dos Usuários do
Orkut, Dados Amostrais

Estratos da Amostra	Coefficiente de correlação
Total	0,42
Dez primeiros no <i>Ranking</i> de Contatos	-0,05
de Comunidades	-0,52

Fonte: Cálculos dos Autores, com base em informações do *Orkut*.

Neste caso, nota-se a ocorrência de uma relação positiva entre as variáveis citadas (coeficiente de correlação linear de 0,42), no caso da amostra como um todo, resultado que demonstra, a princípio, que a participação em um maior número de comunidades possui uma associação (linear) positiva com o número de contatos de cada usuário. Os resultados reportados nas tabelas anteriores demonstram, porém, que essa relação não é válida no caso das subamostras considerando os dez maiores usuários de cada categoria (coeficientes de correlação de -0,05 e -0,52). Esses resultados são interessantes por demonstrarem que padrões inerentes a subconjuntos de usuários do *Orkut* não equivalem necessariamente a padrões vigentes para a amostra como um todo. Mais do que isso, esses resultados apontam para a inexistência de uma relação empírica robusta entre número de contatos e comunidades dos usuários do *Orkut*.

Conclusões e Agenda de Pesquisa Futura

A origem e consolidação das interações sociais entre diferentes indivíduos correspondem a temas atuais em diversas áreas de estudo, como Sociologia, Psicologia, Comunicação, Ciência da Computação e Economia. Neste trabalho, foram apresentados alguns resultados relacionados a esse tema, tendo a teoria econômica como base de análise. No caso, foram expostos resultados relacionados a padrões empíricos de formação de redes sociais a partir de dados do *site* de relacionamentos *Orkut*.

Os principais resultados obtidos a partir desse esforço de pesquisa foram os seguintes:

- i. Ocorre o fenômeno de “cauda longa” tanto no caso do número de contatos quanto no caso de comunidades dos usuários do *Orkut*. Ou seja, existe um número relativamente reduzido de usuários com alto número de contatos e/ou comunidades, com a maioria dos usuários desse *site* de relacionamentos tendo um número de contatos menor;
- ii. Há nítidas diferenças entre homens e mulheres no que diz respeito a vários quesitos referentes a preferências expressas pelos usuários do *Orkut*.
- iii. O usuário típico do *Orkut* apresenta, em média, as seguintes características: possui entre 18 e 30 anos, tem como principal interesse no *site* a formação de amizades, é solteiro(a), branco(a), católico(a), não tem visão política definida, não fuma, mas bebe socialmente.
- iv. Resultados de um exercício econométrico demonstram que variáveis relacionadas à popularidade dos usuários (número de recados, fãs e classificações dadas por outros usuários) exercem um efeito quantitativo significativo no número de contatos estabelecido por cada usuário.
- v. Não há coincidência entre usuários que possuem o maior número de contatos e aqueles que pertencem ao maior número de comunidades.

Estes resultados são importantes no sentido de representarem uma primeira tentativa de quantificação de interações sociais surgidas no contexto de um *site* de relacionamentos em âmbito nacional. Mais do que isso, esses resultados retratam a ocorrência de interações entre indivíduos, pelo menos no caso de um ambiente estático onde essas interações são mediadas via computador¹².

Em termos de sugestões de pesquisa futura, há muitas possibilidades relacionadas ao tema. Especificamente, algumas dessas possibilidades de pesquisa podem ser diretamente derivadas das limitações da análise proposta neste trabalho, conforme discutido abaixo.

Em primeiro lugar, um problema relacionado ao presente trabalho equivale ao fato de que o contexto em análise possui um caráter estático, por conta da própria natureza dos dados (*cross-section* de perfis de usuários do *Orkut*). Por conta disso, não foi possível responder questões relacionadas à evolução temporal dos usuários desse *site* de relacionamentos. A princípio, seria interessante realizar um estudo onde se comparasse, por exemplo, a evolução do número de contatos de cada usuário entre os anos inicial e final da amostra. Um empreendimento nesses moldes permitiria testar a ocorrência ou não do fenômeno de convergência no *Orkut*, qual seja, se usuários com um menor número de contatos em um dado momento inicial tendem a apresentar uma maior taxa de crescimento no número de contatos com o passar do tempo.

Uma segunda limitação do trabalho relaciona-se à impossibilidade de mensuração quantitativa do processo de formação de redes, no caso do presente contexto. Enquanto o trabalho forneceu alguns resultados relacionados a padrões empíricos referentes aos usuários do *Orkut*, não foi capaz de quantificar o processo de formação de nódulos de uma rede social,

¹² Um questionamento pertinente relacionado ao *Orkut* equivale ao fato de a maior parte das conexões estabelecidas através desse *site* de relacionamentos representar efetivamente interações sociais, uma vez que a formação de redes de relacionamento virtuais não pressupõe o estabelecimento efetivo de conexões sociais entre os indivíduos que as formam. Ver, a esse respeito, Recuero (2008, pp.8-9).

conforme ressaltado por alguns dos trabalhos teóricos citados acima. Essa limitação também decorre da impossibilidade de se avaliar a evolução temporal do número de contatos de cada usuário.

Nesse sentido, seria interessante a elaboração de um estudo que acompanhasse o processo de aquisição de contatos por parte de cada usuário ao longo do tempo, o que tenderia a enriquecer significativamente a análise proposta. Um estudo nesses moldes poderia, a princípio, permitir a identificação de padrões empíricos de aglomeração de usuários em comunidades baseadas em interesses comuns, conforme proposto por Currarini, Jackson e Pin (2008), ao tratar do fenômeno de homofilia. Por exemplo, pode-se pensar na mensuração da qualidade dos laços formados entre indivíduos pertencentes a uma mesma rede social. Também é possível se perguntar por que indivíduos tendem a formar redes baseadas em interesses semelhantes comuns ou se ocorrem padrões de discriminação entre indivíduos pertencentes a esse tipo de rede, conforme descrito acima. Do mesmo, o estudo de redes sociais pode permitir responder questões relacionadas à influência que um indivíduo sofre de seus pares, a partir da estimação de efeitos entre pares (*peer effects*) (Marmaros e Sacerdote 2006).

Alternativamente, a formação de redes sociais pode ser extremamente útil ao potencializar ações individuais. Por exemplo, relatos relacionados a atividades de voluntariado demonstram que a formação e consolidação de redes voltadas para essa atividade tende a aumentar significativamente a eficácia de ações voluntárias (Ayres 2010). Do mesmo modo, a interação via redes pode ser útil para potencializar as ações de distintas firmas e/ou instituições. Em termos gerais, é provável que a utilização de estruturas em rede seja útil tanto em termos mercadológicos quanto políticos¹³.

A disponibilidade de bases de dados mais amplas pode ser útil para a construção de novos indicadores quantitativos e testes estatísticos de hipóteses específicas relacionadas ao tema de redes sociais. Nesse sentido, a cooperação de órgãos e instituições como *Google*, *Facebook*, *Twitter* e

¹³ Um exemplo de uso inovador da estrutura de redes para fins políticos equivale à campanha eleitoral do último Presidente norte-americano, Barack Obama. No caso brasileiro, há o exemplo recente da campanha de Fernando Gabeira. Ver, a esse respeito, Falcão (2009).

organizações similares, que possuam dados dessa natureza pode ser fundamental à pesquisa aplicada, conforme já vem ocorrendo em países como os Estados Unidos, por exemplo (e.g., Marmaros e Sacerdote 2006; Zentner 2006; Hitsch, Hortaçsu e Ariely 2010).

Entender a formação de redes sociais pode vir a ser um dos temas mais importantes de pesquisa em ciências sociais aplicadas nos próximos anos. Contribuições de distintas disciplinas poderão ser fundamentais à compreensão dos diversos aspectos relacionados a esse fenômeno. Em termos econômicos, é possível que, pelo lado teórico, análises baseadas em Teoria dos Jogos sejam capazes de explicar as razões subjacentes à formação de redes entre distintos indivíduos. Pelo lado empírico, análises envolvendo amostras longitudinais (dados em painel) poderão vir a esclarecer questões relacionadas à dinâmica temporal da formação de redes. Do mesmo modo, análise utilizando técnicas econométricas destinadas a estimar os efeitos causais entre as variáveis analisadas (e.g., variáveis instrumentais), nos moldes propostos por Angrist e Krueger (2001), por exemplo, poderão ser úteis no sentido de identificar de maneira mais precisa as eventuais relações empíricas existentes entre essas variáveis. Apesar da presente análise possuir um caráter exploratório, espera-se que alguns dos resultados aqui reportados possam constituir evidência empírica inicial que permita uma melhor compreensão da formação, dinâmica e eventual estabilidade de redes sociais¹⁴.

Referências Bibliográficas

ADAMIC, L.A. *Zipf, Power-laws, and Pareto – a ranking tutorial*. Information Dynamics Labs, manuscrito, 2002.

ADAMIC, L.A.; HUBERMAN, B.A. Zipf's Law and the Internet. *Glottometrics*, v.3, n.1, p.143-150, 2002.

ALBERT, R.; JEONG, H.; BARABÁSI, A.L. Diameter of the world-wide web. *Nature*, v.401, p.130, Sept.1999.

¹⁴ Para uma lista de diversas questões remanescentes relacionadas ao estudo de formação de redes, ver Amaral *et alli*. (2004). Em particular, esses autores ressaltam a necessidade da pesquisa relacionada ao tema mudar seu foco de problemas teóricos para questões práticas e aplicadas.

AMARAL, L.A.N.; BARRAL, A.; BARABÁSI, A.L.; CALDARELLI, G.; DE LOS RIOS, P.; ERZAN, A.; KAHNG, B.; MANTEGNA, R.; MENDES, J.F.F.; PASTOR-SALORRAS, R.; VESPIGNANI, A. Virtual round table on ten leading questions for network research. *The European Physical Journal B*, v.38, n.1, p.143-145, 2004.

ANDERSON, C. *A cauda longa: do mercado de massa para o mercado de nicho*. Elsevier, 2006, 256p.

ANGRIST, J.D.; KRUEGER, A.B. Instrumental variables and the search for identification: from supply and demand to natural experiments. *Journal of Economic Perspectives*, v.15, n.4, p.69-85, Fall 2001.

ANTONOVICS, K.; ARCIDIACONO, P.; WALSH, R. Games and discrimination: evidence from the *Weakest Link*. *Journal of Human Resources*, v.40, n.4, p.918-947, Fall 2005.

AYRES, B. *Voluntariado e política 2.0*. Apresentação feita no Seminário *Política 2.0*. Vitória, mar.2010, 45p.

BECKER, G.S. *A theory of social interactions*. NBER Working Paper n.42, Jun.1974, 55p.

COSTA, D.L.; KAHN, M.E. *Surviving Andersonville: the benefits of social networks in POW camps*. UCLA, manuscrito, Jun.2006, 40p.

CURRARINI, S.; JACKSON, M.O.; PIN, P. *An economic model of friendship: homophily, minorities and segregation*. Stanford University, manuscrito, Jan.2007, 46p.

FALCÃO, J. *Mobilização social online*. Apresentação feita no Seminário *Política 2.0*. Vitória, mar.2010, 22p.

GABAIX, X. Power Laws. In: DURLAUF, S.N.; BLUME, L.E. (Eds.). *The New Palgrave Dictionary of Economics*, Second Edition, London: McMillan, 2008.

GLAESER, E.L. Are cities dying? *Journal of Economic Perspectives*, v.12, n.2, p.139-160, Spring 1998.

HITSCH, G.J.; HORTAÇSU, A.; ARIELY, D. What makes you click? Mate preferences in online dating. *Quantitative Marketing and Economics*, forthcoming, 2010.

HUBERMAN, B.A.; ADAMIC, L.A. Growth dynamics of the world-wide web. *Nature*, v.401, p.131, Sept.1999.

JACKSON, M.O. Network formation. In: NEWMAN, P. (Org.). *The New Palgrave Dictionary of Economics and the Law*, Macmillan, 2008.

LOPES, A.G. *Uma análise econômica envolvendo interações sociais no Orkut* (Monografia de Conclusão do Curso de Ciências Econômicas). Faculdade de Ciências Econômicas de São Paulo da Fundação Escola de Comércio Álvares Penteado (FACESP/FECAP), nov.2005, 32p.

MARMAROS, D.; SACERDOTE, B. How do friendships form? *Quarterly Journal of Economics*, v.121, n.1, p.79-119, Feb.2006.

RECUERO, R.C. Redes sociais na Internet: considerações iniciais. *E-Compós*, v.9, n.1, p.1-15, 2005.

SCHEINKMAN, J.A. Social interactions. In: DURLAUF, S.N.; BLUME, L.E. (Eds.). *The New Palgrave Dictionary of Economics*, Second Edition, London: McMillan, 2008.

WHITE, H. A heteroskedasticity-consistent covariance matrix and a direct test for heteroskedasticity. *Econometrica*, v.48, n.4, p.817-838, 1980.

ZENTNER, A. Measuring the effect of music *downloads* on music purchases. *Journal of Law and Economics*, v.49, n.1, p.63-90, Apr.2006.