

Association for Information Systems

AIS Electronic Library (AISeL)

CAPSI 2019 Proceedings

Portugal (CAPSI)

10-2019

Information system for a digital method in social science

Filipe Montargil

Vitor Rodrigues

Branco Di Fátima

Follow this and additional works at: <https://aisel.aisnet.org/capsi2019>

This material is brought to you by the Portugal (CAPSI) at AIS Electronic Library (AISeL). It has been accepted for inclusion in CAPSI 2019 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

Sistema de informação para um método digital em ciências sociais

Information system for a digital method in social science

Filipe Montargil, ESCS-IPL, Portugal, fmontargil@escs.ipl.pt

Vitor Rodrigues, Innovation Makers, Portugal, vitor.rodrigues@inm.pt

Branco Di Fátima, CIES-IUL, Portugal, brancodifatima@gmail.com

Resumo

A utilização massiva de novas tecnologias, em especial da Internet, cria desafios para entender a sociedade contemporânea. De um lado, abordagens metodológicas tradicionais já apresentam sérias fragilidades para explicar as formas de vida em rede. De outro, métodos nativos digitais ainda carecem do aprofundamento de suas estruturas científicas. Este artigo apresenta e analisa o sistema de informação do *Living Lab on Media Content and Platforms (LLMCP)*, um painel online de utilizadores da Internet capaz de monitorizar, em tempo real, os comportamentos em rede. Com a combinação das áreas de estudos de painel e de big data, o objetivo do LLMCP consiste em desenvolver uma extensão do Google Chrome para recolher ações de navegação e mapear tendências de conectividade. Os resultados revelam que a complexidade da vida em rede exige a articulação de temas de pesquisa inovadores e ferramentas informáticas robustas, indicando a potencialidade do histórico de navegação como objeto digital para as ciências sociais.

Palavras-chave: Uso da Internet; Métodos Digitais; Estudos de Painel; Google Chrome; Big Data

Abstract

The massive use of new technologies, especially the Internet, creates many challenges for the understanding of contemporary society. On the one hand, traditional methodological approaches already have serious weaknesses in explaining networked life forms. On the other hand, native digital methods still need reinforcing its scientific structures. This paper presents and analyzes the information system of the Living Lab on Media Content and Platforms (LLMCP), an online panel of Internet users able to monitor, in real time, online behaviour of users. With the combination of the areas of Panel Studies and Big Data, LLMCP's goal is to develop an extension of Google Chrome to collect navigation actions and to map connectivity trends. Results reveal that the complexity of network life requires the articulation of innovative research themes and robust software tools, indicating the potential of browsing history as a digital object for the social sciences.

Keywords: Internet Usage; Digital Methods; Panel Studies; Google Chrome; Big Data

1. PROBLEMÁTICA: NOVOS OBJETOS, MÉTODOS TRADICIONAIS

A rápida disseminação da Internet, nos últimos anos, tem criado as estruturas para a emergência de formas de vida significativamente diferentes de períodos anteriores. Dados do Internet World Stats (2019), por exemplo, indicam que a utilização massiva dessa tecnologia já chega à grande maioria

dos europeus (86,8%) e norte-americanos (89,4%), embora ainda seja possível notar fossos digitais acentuados entre estados asiáticos (51,8%) e africanos (37,3%). Em todo o caso, práticas quotidianas, como o envolvimento em atividades políticas (Campante, Durante & Sobbrío, 2018) ou o consumo de conteúdos de entretenimento (Jenkins, 2013; Lessig, 2012), encontram-se fortemente associadas à nova paisagem mediática. As redes digitais de informação e de comunicação passaram a ser uma estrutura material das relações em comunidade (Rheingold, 2000), espaço dilatado de socialização e afirmação da identidade (Castells, 2002). Mas como é utilizada a Internet?

Na maioria dos casos, esta pergunta ainda é respondida através do recurso a abordagens metodológicas tradicionais, assentes na inquirição do utilizador. Centros de investigação públicos e privados, em diferentes latitudes, recorrem a entrevistas e questionários para caracterizar os comportamentos na Internet (Ibope, 2018; Pew Research Center, 2016; OberCom, 2014; Dubai School of Government, 2012; INE, 2010). A vida online, revelam essas pesquisas, é constituída pela articulação de hábitos diversos, como pesquisar notícias, falar com amigos, fazer negócios, publicar nas redes sociais ou ver um filme em *streaming*, correspondendo a uma profunda combinação de práticas informativas, comunicativas, utilitárias e de entretenimento. “Essa cultura digital tem propriedades emergentes com raízes em fenómenos online e offline, (...) mudando as formas como utilizamos e damos sentido à vida num ambiente cada vez mais interconectado” (Deuze, 2006: 63).

Embora sejam lentes importantes para ler a sociedade, os métodos com base na inquirição também apresentam fragilidades identificadas, desde a década de 1930, como a diferença entre as práticas e as suas representações ou a própria limitação da memória humana (LaPiere, 1934; Roberts, 1985; Foddy, 1996). Num contexto que os indivíduos estão permanentemente conectados (Vorderer & Kohring, 2013) e o volume de informações disponíveis tem crescido de maneira exponencial (boyd & Crawford, 2012), é importante apurar até que ponto as abordagens metodológicas tradicionais ainda oferecem dados fiáveis sobre os hábitos online. Assumindo essas limitações como uma força modular das investigações, estarão as instituições públicas e privadas a desenhar estratégias de atuação com base em informações efetivamente relevantes? Quando os utilizadores afirmam ter consumido um determinado conteúdo online, estarão a falar sobre as suas ações concretas de navegação ou sobre as suas representações?

A utilização massiva das novas tecnologias, em especial da Internet e suas redes, também criou um terreno fecundo para o desenvolvimento de métodos nativos digitais (Rogers, 2013). Mais do que transpor modelos tradicionais, como as entrevistas feitas por videoconferência ou os questionários aplicados online (Bryman, 2012), os métodos nativos digitais trabalham com objetos que existem apenas em bits e ecrãs. Entre os novos métodos estariam, por exemplo, a análise de redes sociais (Barabási, 2017), a netnografia (Kozinets, 2010) ou os estudos do hipertexto (Fragoso, Recuero &

Amaral, 2011). Bryman (2012) fala sobre o florescimento da *e-Pesquisa* e dos impactos da Internet na construção do saber em sociedades complexas. Ainda, a partir da sociologia do conhecimento, é importante pensar a influência que planos empíricos para recolha de dados exercem na criação da realidade e nas estratégias dos diferentes *stakeholders* frente à inquirição (Di Fátima, Montargil & Miranda, 2019; Montargil *et al.*, 2019).

Embora seja consensual que as ciências sociais precisam de ajustar as suas lentes para ler as sociedades informacionais altamente dinâmicas, também parece verdade que os métodos nativos digitais carecem ainda de aprofundar a sistematização de técnicas de recolha, tratamento e visualização de dados. Trabalhos inaugurais como o de Jones (1999) e, depois, os de Fragoso, Recuero & Amaral (2011) e de Rogers (2013), esforçaram-se justamente para sistematizar os métodos digitais. Nesse ponto, os sistemas de informação assentes na Internet necessitam de ser entendidos, ao mesmo tempo, como ferramentas e como objetos de estudo capazes de criar os seus próprios dados (Jensen, 2011).

Nesse sentido, as próximas páginas apresentam e analisam o sistema de informação do *Living Lab on Media Content and Platforms* (LLMCP)¹, um painel online de utilizadores da Internet capaz de monitorizar, em tempo real, os comportamentos e tendências de conectividade dos seus membros. Com recolha de dados a partir de uma extensão do Google Chrome, especialmente desenvolvida para este propósito, o histórico de navegação é visto como ponto de partida para a sistematização de um novo método digital. Basicamente, a extensão conecta-se à API do Chrome e, a partir de um pedido REST, extrai e grava as informações de navegação numa base MySQL. Com a proliferação de ferramentas de extração, tratamento e visualização de dados, como lembram boyd & Crawford (2012), é prudente questionar como os sistemas estão a condicionar as atividades de investigação.

Estratégias metodológicas assentes num painel online não são exatamente novas. Iniciativas como *Computer and Mobile Panel* (Nielsen)², *Panel Network* (comScore)³ e *NetPanel* (Marktest)⁴ fazem a recolha de dados de utilização da Internet, como a página visitada ou o tempo de navegação. Em comum, essas plataformas nasceram numa lógica empresarial e estão condicionadas por, pelo menos, dois fatores. De um lado, pela ausência de controlo dos mecanismos de extração de dados por parte do investigador. De outro, pelo alto custo associado à aquisição dos serviços. O LLMCP apresenta-se como uma iniciativa pioneira. Pela primeira vez, um centro público de ensino em Portugal e, possivelmente, na União Europeia, lança um painel online para fins académicos (Di Fátima, Montargil & Miranda, 2019). Os resultados preliminares do LLMCP apontam para o potencial do histórico de navegação como objeto digital para as ciências sociais. Já as limitações

¹ Living Lab on Media Content and Platforms (LLMCP): <<http://prd.llmcp.pt>>. Acesso: 17/07/2019

² Computer and Mobile Panel: <computermobilepanel.nielsen.com/cmp/landingeng.jsp>. Acesso: 15/07/2019

³ Panel Network: <www.comscore.com/comScore-Panel-Services>. Acesso: 13/07/2019

⁴ NetPanel: <<http://netpanel.marktest.pt>>. Acesso: 10/07/2019

prendem-se à dimensão da base para uma análise agregada e à dificuldade de alguns candidatos em completar o processo de ativação da extensão.

2. SISTEMA DE INFORMAÇÃO DO LLMCP

O LLMCP desenvolveu, no final de 2018, um sistema de informação como estrutura organizativa de um painel online de utilizadores da Internet. De acordo com as avaliações de custo-benefício do projeto, liderado por investigadores da Escola Superior de Comunicação Social (ESCS/IPL), uma extensão do Google Chrome era a solução informática mais eficiente para cumprir a expectativa de recolha do histórico de navegação. Embora também apresente limitações, expectáveis num período de implementação de métodos digitais (Rogers, 2013), o Chrome é o *browser* mais utilizado pelos portugueses e tem a maior quota no mercado global de navegadores (StatCounter, 2019).

Na primeira etapa do projeto, os investigadores do LLMCP pesquisaram experiências semelhantes que pudessem sustentar as decisões metodológicas do painel. Já o desenvolvimento do sistema foi planeado em três grandes fases: (i) desenho do projeto; (ii) elaboração de formulário de utilizador, (iii) programação informática da extensão. O desenvolvimento do sistema consumiu cerca de dois meses, que também contou com testes unitários e integrados da recolha de informações. Os testes foram particularmente importantes para garantir que os dados armazenados refletissem o histórico de navegação do utilizador e pudessem validar a ferramenta de extração. Instalada em computador portátil ou desktop, a extensão permite a recolha das atividades realizadas no Google Chrome.

O candidato à participação no painel regista-se pelo formulário Web, através do preenchimento de seus dados sociodemográficos em uma página HTML, com componentes de JavaScript e de CSS. A conclusão do registo cria um pedido REST (Representational State Transfer) para o servidor desenvolvido em Java, que processa as informações e comunica-se com a base de dados MySQL e sistema Linux. Esse procedimento garante o armazenamento seguro das informações, nunca tendo sido identificadas perdas ou fuga de dados pelo sistema (Figura 1).

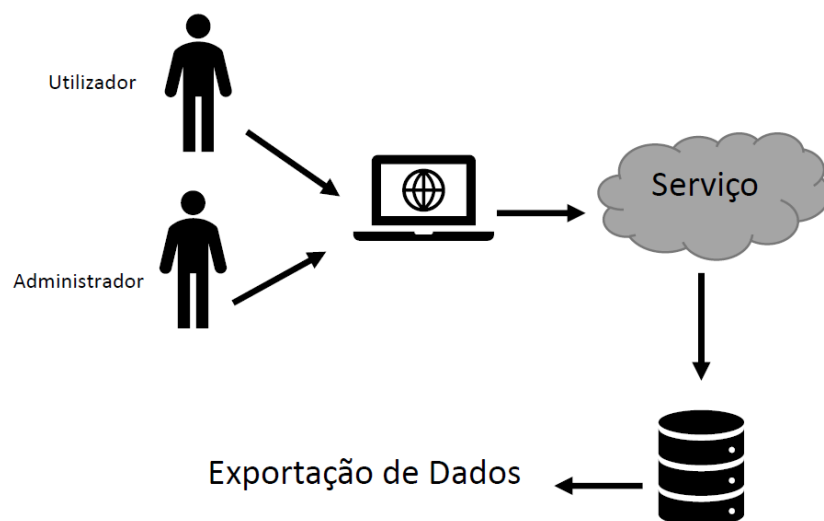


Figura 1 – Sistema de informação do LLMCP

O sistema permite que a equipa do LLMCP acesse à plataforma por um BackOffice. Através de uma operação de *login* com credenciais (utilizador/password) é novamente chamado um serviço REST, que valida as informações do administrador. Se estiverem corretas, o procedimento gera um Token que dá acesso, por um tempo limitado, ao BackOffice. Para garantir o controlo da amostra, indispensável às ciências sociais (Ander-Egg, 1978), os candidatos têm os dados de seu formulário analisados pela equipa. É importante destacar que esse formulário está organizado em quatro áreas: i) informações pessoais, (ii) situação laboral, (iii) questões sobre o acesso à Internet e (iv) dados de autenticação da conta.

Após fazer *login* no sistema, é possível para o administrador transferir a informação armazenada, resultado direto da monitorização em tempo real das atividades de navegação no Google Chrome, através de um serviço especialmente desenvolvido para esta finalidade. O sistema também permite que a equipa de investigadores consulte a lista dos utilizadores da extensão e, sempre que necessário, segmente uma subamostra por género, idade, estado civil, condição laboral e hábitos de partilha do computador.

O BackOffice do LLMCP conta com uma ficha de participante, na qual é possível consultar todas as informações dos membros do painel, ordenados alfabeticamente pelo *username*. Este serviço de administrador, desenvolvido em HTML, JavaScript e CSS, também permite aprovar ou rejeitar as candidaturas ao painel, na sequência da verificação do correto preenchimento dos campos do formulário. Neste caso, é possível exportar apenas uma lista de membros aprovados ou rejeitados, com dados armazenados em Excel e informações detalhadas sobre cada candidatura (Figura 2).

Nome	Email	Estado
#feeltheinfluencers	@hotmail.com	Aprovado
1FoY^okC	@gmail.com	Aprovado
282000	@gmail.com	Aprovado
4589910	@gmail.com	Aprovado
88	@gmail.com	Aprovado
93	@gmail.com	Aprovado
ARG	@hotmail.com	Aprovado
ARPandinha	@hotmail.com	Aprovado
A_Mafiosa	@gmail.com	Aprovado
Ana Cardoso	@gmail.com	Aprovado

Figura 2 – BackOffice do formulário de candidatura ao painel

Em caso de aprovação da candidatura, o sistema envia um email de boas-vindas ao candidato, com um texto de apoio (“Obrigado pelo vosso registo. Foi aceite para participar do painel de utilizadores da Internet. Por favor, instale a extensão para Google Chrome, autentifique o email e password para fazer parte da nossa investigação. Em caso de dúvida, não hesite em contactar-nos”) e um link para aceder à Chrome Web Store e transferir a extensão para o computador pessoal. Já em caso de rejeição da candidatura, a equipa de investigação elabora uma mensagem personalizada, em que explica o motivo da recusa em integrar o painel.

Do lado do utilizador, após instalar a extensão, é necessário fazer *login* com os dados fornecidos no momento do registo. Quando o *login* é efetuado com sucesso, como no caso do administrador, o sistema produz um Token que dura até o utilizador fazer *logout* da extensão ou limpar o *cache* do navegador. É possível ativar e desativar a monitorização a qualquer momento. Mas, estando ativa, a extensão conecta-se à API (Application Programming Interface) do Google Chrome. A partir daí, com uma comunicação JavaScript, extrai e guarda as informações de navegação na base de dados. Já se a monitorização for desativada, a extensão suspende a recolha do histórico (Figura 3). Tanto a extensão como o formulário foram desenvolvidos em consonância com o *Regulamento Geral de Proteção de Dados* (RGPD), em vigor na União Europeia desde 25 de maio de 2018.

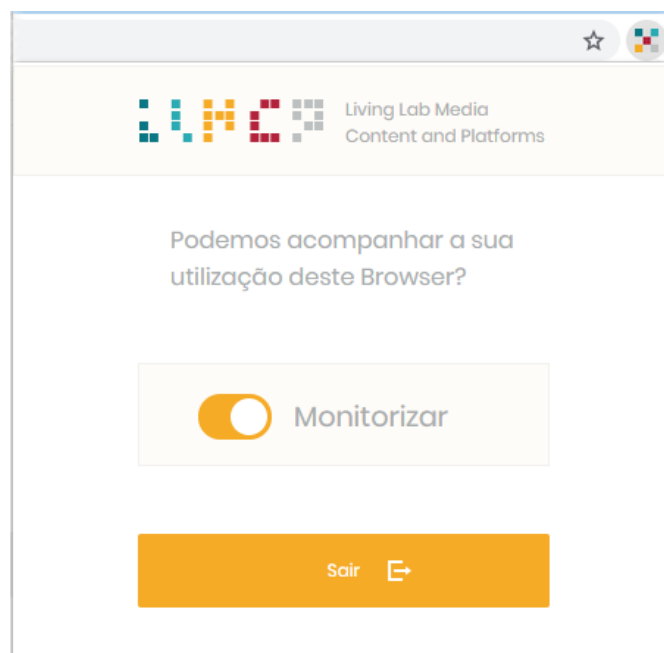


Figura 3 – Extensão com botão de monitorização

O serviço de transferência de dados encaminha um pedido REST para o servidor, que realiza uma pesquisa na base de dados MySQL. Este procedimento gera um conjunto de ficheiros em formato CSV (Comma Separated Values). Cada um dos arquivos armazena 50.000 ações de navegação, em ordem cronológica, que somam cerca de 15 megabytes. Na base de dados, cada ação de navegação representa um website, com o endereço URL, número da TAB, atividade da página, hora e data de acesso ao conteúdo. Para reduzir o tempo de comunicação com o servidor e minimizar os erros de TimeOut, comuns na extração de grandes bases (Dai *et al.*, 2018), os ficheiros são comprimidos e enviados por email. Em condições normais de conexão à Internet banda larga, o procedimento de extração dura cerca de cinco minutos.

O painel do *Living Lab on Media Content and Platforms* (LLMCP), ainda em formação, reunia em julho de 2019 cerca de 160 utilizadores, sobretudo estudantes da Escola Superior de Comunicação Social (ESCS). Destes, 90 haviam tido a sua navegação no Google Chrome monitorizada pelo menos uma vez, desde novembro de 2018 (55,9%). Atualmente, a base de dados do projeto agrega cerca de 680.000 ações de navegação, com um peso aproximado de 190 megabytes em formato CSV. Neste sistema, as limitações mais frequentes prendem-se, por um lado, ao tamanho da base para a análise agregada da informação e, por outro, à dificuldade de alguns candidatos em completar o processo de ativação da extensão no computador.

3. CONCLUSÃO

O LLMCP pretende contribuir para sistematização de novos métodos digitais, numa paisagem mediática assinalada pelo crescimento do volume de dados produzidos a cada segundo, na Internet (boyd & Crawford, 2012). Diante de objetos digitais, como o histórico de navegação do Chrome, abordagens tradicionais, assentes na inquirição do utilizador, aparentam revelar um conjunto significativo de limitações. Um desafio é perceber qual o peso das condicionantes nas estratégias de atuação de instituições públicas e privadas, na maneira como projetos com molduras tecnológicas são geridos em diversos contextos.

A nossa vida em rede exige o desenvolvimento de abordagens inovadoras, com base em sistemas de informação robustos. Neste sentido, os testes de funcionalidade da extensão foram particularmente importantes. Para entender sociedades complexas, altamente dinâmicas (Castells, 2002), não basta constituir *datasets* com dezenas ou centenas de milhar de linhas. É também necessário validar todo o processo de recolha, tratamento e análise de dados. Os resultados devem refletir justamente a diversidade e a complexidade da vida humana e das suas comunidades. A própria seleção do Chrome como *browser*, por exemplo, revelou-se adequada (Di Fátima, Montargil & Miranda, 2019): como navegador mais utilizado em Portugal (StatCounter, 2019), tem maior capacidade de atração de participantes para o projeto.

As próximas fases de desenvolvimento do LLMCP devem focar-se, sobretudo, na simplificação do sistema de registo do utilizador. Os dados empíricos do painel têm indicado a dificuldade que uma parcela dos candidatos enfrenta para completar o processo. De um lado, a informação pode denotar os níveis de literacia informática de parte da amostra. De outro, é uma prova que os sistemas mais complexos de registo, mesmo com interface gráfica amigável, acabam por construir barreiras para um estrato significativo do universo. Os índices de participação parecem estar em parte associados, assim, à própria estrutura do sistema.

AGRADECIMENTO

O projeto *Living Lab on Media Content and Platforms* é financiado pela Fundação para a Ciência e a Tecnologia (FCT), com apoio da União Europeia. LISBOA 01-0145 - FEDER 023937.

A Bruno Duarte e Ricardo Sequeira, da Innovation Makers, pelas informações sobre o formulário e a extensão do Google Chrome.

REFERÊNCIAS

- Ander-Egg, E. (1978). *Introducción a las técnicas de investigación social*. Buenos Aires: Editorial Humanitas.
- Barabási, A. L. (2017). *Network science*. Cambridge: Cambridge University Press.

- boyd, d. & Crawford, K. (2012). Critical questions for Big Data. *Information, Communication & Society*, 15(5), 662-679.
- Bryman, A. (2012). *Social research methods*. 4^a Ed. Oxford: Oxford University Press.
- Campante, F., Durante, R. & Sobbrío, F. (2018). Politics 2.0: The multifaceted effect of broadband Internet on political participation. *Journal of the European Economic Association*, 16(4), 1094-1136.
- Castells, M. (2002). *A sociedade em rede*. A era da informação: economia, sociedade e cultura, Vol. 1. Lisboa: Calouste Gulbenkian.
- Dai, T. et al (2018). "Understanding real-world timeout problems in cloud server systems". *Atas da International Conference on Cloud Engineering (IC2E)*, Orlando, In: IEEE Xplore Digital Library, p. 1-11.
- Deuze, M. (2006). Participation, remediation, bricolage: considering principal components of a digital culture. *The Information Society*, 22(2), 63-75.
- Di Fátima, B., Montargil, F. & Miranda, S. (2019). Estudando os comportamentos online: premissas e desafios no desenvolvimento de um painel de utilizadores da Internet. *Texto Livre: Linguagem e Tecnologia*, 12(1), 123-137.
- Dubai School of Government (2012). Social media in the Arab World: Influencing societal and cultural change?, *Arab Social Media Report*, 2(1), 1-29.
- Foddy, W. (1996). *Como perguntar: teoria e prática da construção de perguntas em entrevistas e questionários*. Lisboa: Celta.
- Fragoso, S., Recuero, R. & Amaral, A. (2011). *Métodos de pesquisa para internet*. Porto Alegre: Sulina.
- Ibope (2018). *95% dos internautas brasileiros assistem TV enquanto usam Internet*. São Paulo: Ibope Conecta.
- INE (2010). *Sociedade da Informação e do Conhecimento: inquérito à utilização de Tecnologias da Informação e da Comunicação pelas famílias*. Lisboa: Instituto Nacional de Estatística.
- Internet World Stats (2019). *World Internet usage and population statistic*, March 2019, Updated. Disponível em: <http://www.internetworldstats.com/stats.htm> (10 de maio de 2019).
- Jenkins, H. (2013). *Cultura da Convergência*. São Paulo: Aleph.
- Jensen, K. B. (2011). New media, old methods – Internet methodologies and the online/offline divide, In: Consalvo, M. & Ess, C. (Ed.). *The handbook of Internet Studies*. Chichester: John Wiley & Sons, 43-58.
- Jones, S. (1999). Studying the Net: intricacies and issues, In: Jones, S. (Ed). *Doing Internet research: critical issues and methods for examining the Net*. Londres: Saga, 1-28.
- Kozinets, R. V. (2010). *Netnography: doing ethnographic research online*. Thousand Oaks: SAGE.
- LaPiere, R. (1934). Attitudes vs. Actions. *Social Forces*, 13(2), 230-237.
- Lessig, L. (2012). *Remix*. Cultura de la remezcla y derechos de autor en el entorno digital. Icaria: Barcelona.
- Montargil, F. et al. (2019). "Medir a sociedade de informação: sistema para um painel online de utilizadores da Internet". *Atas do 14th Iberian Conference on Information Systems and Technologies (CISTI)*, Coimbra, In: IEEE Xplore Digital Library, p. 1-4.
- Obercom (2014). *A Internet em Portugal: sociedade em rede 2014*. Lisboa: Observatório da Comunicação.
- Pew Research Center (2016). *Social media update 2016*. Internet & Technology, 2016. www.pewinternet.org/2016/11/11/social-media-update-2016/ (10 de maio de 2019).
- Rheingold, H. (2000). *The virtual community: homesteading on the electronic frontier*. Cambridge: The MIT Press.
- Roberts, J. V. (1985). The attitude–memory relationship after 40 years: a meta-analysis of the literature. *Basic and Applied Social Psychology*, Abingdon-on-Thames, 6(3), 221-241.
- Rogers, R. (2013). *Digital methods*. Cambridge: The MIT Press.
- StatCounter (2019). *Top desktop, tablet & console browsers per country*. Nova York: Global Stats Tool. Disponível em: <http://gs.statcounter.com/> (10 de maio de 2019).
- Vorderer, P. & Kohring, M. (2013). Permanently online: a challenge for media and communication research. *International Journal of Communication*, 7, 188-196.