

## Análise macro das últimas atualizações dos Currículos Lattes

### **Luciano Antonio Digiampietri**

Doutor; Universidade de São Paulo (USP);  
digiampietri@usp.br

### **Rogério Mugnaini**

Doutor; Universidade de São Paulo (USP);  
mugnaini@usp.br

### **Jesús Pascual Mena Chalco**

Doutor; Universidade de São Paulo (USP);  
jesus.mena@ufabc.edu.br

### **Karina Valdivia Delgado**

Doutora; Universidade de São Paulo (USP);  
kvd@usp.br

### **José de Jesús Pérez Alcázar**

Doutor; Universidade de São Paulo (USP);  
jperezal1@gmail.com

**Resumo:** A completude e a atualização dos dados são fatores fundamentais para qualquer análise estatística. A Plataforma Lattes provê uma enorme quantidade de dados que podem ser utilizados em diferentes estudos bibliométricos, bem como na análise de redes sociais, podendo, por exemplo, auxiliar na tomada de decisões sobre políticas científicas. Considerando o grande volume e importância dos dados providos por esta plataforma ímpar, ainda não há estudos que evidenciem de forma macro a completude, corretude e atualização destes dados, já que os mesmos são preenchidos pelos seus usuários. Neste contexto, este trabalho visa a analisar as últimas atualizações dos currículos Lattes considerando um conjunto bastante representativo, composto de mais de três milhões de registros, de acordo com a formação acadêmica e a grande área de atuação profissional dos usuários da plataforma. Adicionalmente, foram utilizados dados históricos sobre a publicação de artigos completos para estimar a quantidade de artigos não registrados nos currículos Lattes. Dentre os resultados obtidos, um que merece destaque é a evidência de que, para determinados grupos, a falta de atualização possa representar mais de 20% do total de artigos publicados no último triênio.

**Palavras-chave:** Currículo Lattes. Atualização da informação. Análise bibliométrica.

## 1 Introdução

Ao longo dos últimos anos, estudos bibliométricos e cientométricos têm utilizado cada vez maiores volumes de dados. No Brasil existe uma base ímpar de informações bibliométricas, que é a Plataforma Lattes. Nesta plataforma, há mais de 3,2 milhões de registros cadastrados (CNPQ, 2014) contendo informações sobre formação, áreas de atuação, projetos de pesquisa, produções (bibliográficas, técnicas e artísticas), participação em eventos, bancas, orientações, dentre outras.

Ao longo dos anos e, em especial, na última década, diferentes pesquisas acadêmicas têm considerado a Plataforma Lattes como a principal fonte de dados. O espectro de trabalhos varia do nível macro ao micro. Por exemplo, trabalhos que visam a apresentar um panorama de toda a produção científica nacional (LEITE; MUGNAINI; LETA, 2011), incluindo a rede social acadêmica de coautorias (MENA-CHALCO et al., 2014), descrições sobre as informações presentes na plataforma (DIGIAMPIETRI et al., 2012a), ferramentas para a extração e/ou mineração dos dados da plataforma (ALVES et al., 2011; MENA-CHALCO; CESAR JUNIOR, 2009), estudos sobre grupos específicos de pesquisadores (DIGIAMPIETRI et al., 2012b; ARRUDA et al., 2009; WAINER; VIERA, 2013; COSTA et al., 2013), predição de relacionamentos (LIBEN-NOWELL; KLEINBERG, 2003; DIGIAMPIETRI et al. 2013), entre outros.

Apesar da abundância e relevância das informações contidas na Plataforma Lattes, existem diversas características que devem ser consideradas para sua utilização (CAÑIBANO; BOZEMAN, 2009). Entre elas: (a) o fato das informações não serem validadas, i.e., os dados são inseridos pelos possuidores dos currículos, sem posterior validação; (b) muitos campos são preenchidos manualmente, acarretando em erros de digitação e/ou falta de padronização; (c) a frequência de atualização dos dados depende dos possuidores dos currículos e varia bastante; e (d) muitos campos são opcionais, limitando alguns tipos de análise (MARQUES, 2010). Este alto grau de liberdade no registro das informações curriculares é pouco estudado pelos pares. No campo da Ciência da Informação, Silva e Smit (2009) alertam para o comprometimento da consistência dos dados para recuperação da

informação, o que acaba limitando o uso desta fonte curricular tão abrangente para uma análise mais profunda da produção científica nacional.

Em particular, o estudo da frequência de atualização dos dados é de extrema importância para trabalhos bibliométricos que pretendam identificar ou prever tendências na produção científica nacional, predição de relacionamentos ou de citações em redes acadêmicas e recomendação de trabalhos científicos. Este artigo tem por objetivo analisar a atualização dos currículos vitae (CVs) da Plataforma Lattes, segmentando-os pelas áreas de conhecimento de seus possuidores e também pela formação acadêmica máxima dos mesmos (e.g., doutorado). Além disso, ainda não tratada na literatura, segundo nosso entendimento, é realizada uma estimativa da quantidade de publicações ausentes nos CVs, possivelmente causada pela desatualização dos mesmos.

## 2 Metodologia

A metodologia está estruturada em quatro partes: obtenção dos dados, separação das informações de interesse; divisão das informações nos grupos de interesse; e cálculo de métricas.

### 2.1 Obtenção dos dados

Para este trabalho, foram obtidos os arquivos XML de 3.187.710 CVs da Plataforma Lattes durante o mês de julho de 2013. Para a obtenção destes currículos, as seguintes atividades foram executadas: (a) foi feita uma consulta no site de buscas por CVs da Plataforma Lattes de forma a solicitar a lista de todos os currículos cadastrados; esta consulta retornou múltiplas páginas Web de resposta; (b) cada uma das páginas de resposta foi copiada e os identificadores numéricos dos CVs (IDs Lattes) foram extraídos utilizando-se de um *script* de computador; (c) com o identificador dos currículos foi possível baixar cada CV Lattes em formato XML.

## 2.2 Separação das informações de interesse

Neste trabalho, quatro tipos de informação dos CVs Lattes foram considerados: (a) a data da última atualização (oriunda das informações gerais de cada CV); (b) as grandes áreas de atuação; (c) as formações acadêmicas/titulações; e (d) a quantidade de artigos completos publicados anualmente. Foi desenvolvido um *script* para extrair as informações de interesse de cada currículo.

## 2.3 Divisão das informações nos grupos de interesse

Além da análise conjunta da atualização de todos os CVs da Plataforma Lattes, também foram identificados grupos segundo as grandes-áreas de atuação e a maior formação presente em cada currículo. Cada CV pode manter registro de zero ou mais grandes-áreas de atuação, permitindo que determinado CV faça parte de mais de um grupo. Segundo a CAPES, são nove as grandes-áreas do conhecimento: Ciências Agrárias; Ciências Biológicas; Ciências da Saúde; Ciências Exatas e da Terra; Ciências Humanas; Ciências Sociais Aplicadas; Engenharias; Linguística, Letras e Artes; e Outros/Multidisciplinar. Também foi criado um grupo adicional formado por CVs que não continham esta informação. Quanto às formações acadêmicas/titulações, a CAPES permite o cadastro de 12 opções, das quais este estudo considerou seis: Ensino Fundamental/Primeiro Grau, Ensino Médio/Segundo Grau, Curso Técnico Profissionalizante, Graduação, Mestrado/Mestrado Profissionalizante e Doutorado. Além destes, foram criados dois grupos adicionais: um contendo os CVs que não apresentavam nenhuma destas formações/titulações e outro contendo os possuidores de bolsa de produtividade em pesquisa do CNPq (os CVs destes pesquisadores compõem tanto na conta do grupo Doutorado quanto do grupo Produtividade). Para a criação destes grupos só foram consideradas as formações/titulações máximas entre as seis destacadas (independente de estarem concluídas ou em andamento). Já que os possuidores de bolsa produtividade frequentemente atualizam seus currículos (ao menos) logo antes do pedido da bolsa, optou-se por selecionar os CVs daqueles que possuíam bolsa produtividade em 2010.

## 2.4 Cálculo de métricas

Tanto para o conjunto de todos os CVs quanto para cada uma das divisões feitas, foram calculadas as seguintes métricas referentes à atualização dos currículos: quantidade média de dias desde a última atualização; valor da mediana de dias desde a última atualização; distribuição das atualizações por meses; e quantidade de dias desde a última atualização para cada decil (cada decil representa 1/10 do total do conjunto de dados). Para a estimativa da quantidade de artigos faltantes devido à desatualização, foi utilizado o seguinte método: dividiu-se por 365 (número de dias do ano) a média de artigos completos publicados em cada CV nos 36 meses (três anos) anteriores a sua última atualização e multiplicou-se o valor resultante pelo número de dias desde a última atualização. Por exemplo, se um dado CV possuía ao todo 12 artigos completos nos últimos 36 meses (três anos) anteriores a atualização (média de quatro por ano) e o currículo estiver desatualizado há três meses (um quarto de ano) então pode-se estimar que esteja faltando o registro de um artigo em seu CV. Com base neste valor, foram calculados a média, a mediana e o número total de artigos faltantes estimados. É importante destacar que um currículo pode não ter sido atualizado simplesmente pelo fato de seu possuidor não ter nenhuma informação nova a cadastrar no currículo ou pela falta de interesse do possuidor de atualizar os seus dados por não estar mais atuante na área acadêmica, por exemplo. A estimativa de artigos faltantes proposta e utilizada neste artigo visa a fornecer apenas uma estimativa sobre a potencial quantidade de informações faltantes nos currículos devido à desatualização.

## 3 Resultados

Os resultados são apresentados em quatro subseções. Na primeira, são apresentados alguns dados gerais envolvendo todos os CVs analisados. Em seguida apresenta-se uma análise da atualização e da estimativa das informações faltantes dos CVs considerando-se as grandes-áreas de atuação. Na terceira parte, é descrita uma

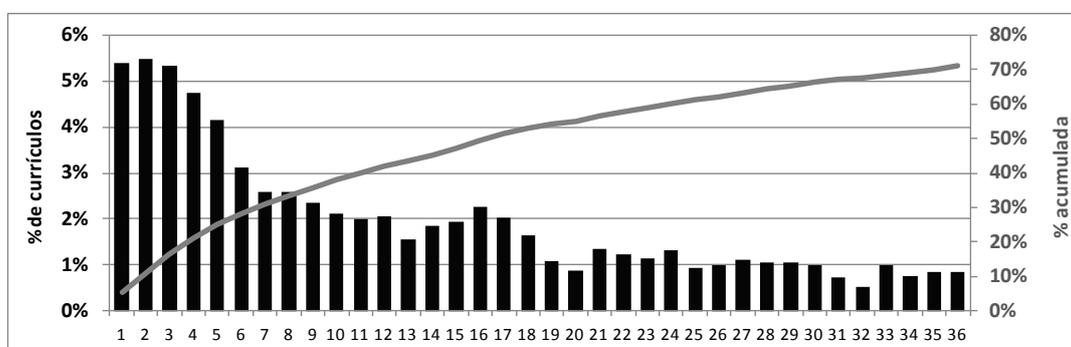
análise considerando-se as formações/titulações dos currículos registrados no conjunto de dados. Por fim, é apresentada uma análise que relaciona as informações das grandes áreas de atuação com as informações sobre formações/titulações acadêmicas.

### 3.1 Dados gerais

Dos 3.187.710 de CVs examinados, baixados da Plataforma Lattes em julho de 2013, as datas de atualização variaram de 22/08/1997 (CV com data de atualização mais antiga) a 30/07/2013 (currículo que foi atualizado no dia que foi baixado), este intervalo de datas corresponde a pouco mais de 16 anos (194 meses). Na média, os CVs foram atualizados 860 dias (mais de 28 meses) antes de terem sido baixados (correspondendo, na média, a 16/03/2011), já a mediana é de 486 dias (cerca de 16 meses, correspondendo a 24/03/2012). É possível observar que, considerando esse conjunto de dados, mais da metade dos currículos foram atualizados pela última vez há mais de um ano.

A Figura 1 apresenta a porcentagem de currículos que foram atualizados dentro de períodos mensais. Só são exibidos os 36 primeiros meses dentro dos quais 70,96% dos currículos foram atualizados. 5,4% dos CVs foram atualizados há, no máximo, um mês da data que foram baixados. A curva cinza apresenta os valores acumulados. É possível observar que 41,90% dos CVs foram atualizados nos últimos 12 meses e 60,16% nos últimos 24 meses.

**Figura 1** – Porcentagem de CVs atualizados ao longo do tempo (eixo horizontal: meses)



Fonte: os autores.

Os 3.187.710 de CVs possuem ao todo 7.135.143 artigos completos registrados. Somando-se as publicações de apenas os três anos anteriores à atualização de cada CV, existem 2.020.455 artigos completos registrados. Estima-se, segundo a medida adotada, que estejam faltando nos CVs 658.705 artigos devido à falta de atualizações. Na média, estaria faltando cerca de 0,207 artigos por currículo, isto é, ao se analisar os artigos completos de um grupo arbitrário de cinco CVs, estima-se que estará faltando um artigo completo na análise. A mediana para esta estimativa é zero, isto é, para mais de metade dos currículos, não é esperado que esteja faltando nenhum artigo completo. Vale destacar que do total de CVs analisados, 2.483.062 não possuem nenhum artigo completo registrado (77,9% do total de currículos).

### 3.2 Grandes áreas

O Quadro 1 mostra os decis da variável dias desde a última atualização. Por exemplo, ao se analisar os 10% CVs mais atualizados em Ciências Biológicas, estaremos olhando para CVs atualizados até 28 dias antes da data em que os CVs foram baixados. Ao se analisar 50% de todos os currículos que declararam esta mesma grande área, estaremos olhando para CVs atualizados há até 247 dias antes da data em que foram baixados. A coluna *Total* contém o número total de currículos pertencentes ao respectivo grupo.

**Quadro 1** – Número de dias desde a última atualização dos CVs por grande-área

Grande Área	Total CVs	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
<b>Sem Grande Área</b>	1.006.088	124	328	637	978	1.173	1.425	1.718	2.121	2.631	5.782
<b>Ciências Agrárias</b>	141.730	31	63	103	168	278	441	661	1.043	2.066	5.779
<b>Ciências Biológicas</b>	175.384	28	59	97	151	247	398	613	1.001	1.944	5.783
<b>Ciências da Saúde</b>	459.825	45	95	158	257	397	530	741	1.093	1.921	5.780
<b>Ciências Exatas e da Terra</b>	266.693	40	78	130	218	344	501	747	1.187	2.251	5.796
<b>Ciências Humanas</b>	417.334	37	78	130	201	316	478	670	975	1.783	5.779
<b>Ciências Sociais Aplicadas</b>	439.330	47	99	168	275	418	561	776	1.146	2.009	5.795
<b>Engenharias</b>	192.495	44	90	151	258	403	566	851	1.450	2.633	5.814
<b>Linguística, Letras e Artes</b>	157.597	40	83	137	218	333	487	684	1.012	1.834	5.599
<b>Outros/ Multidisciplinar</b>	337.017	49	86	125	179	242	330	420	469	708	5.780
<b>Mundo Lattes</b>	3.187.710	55	112	196	333	486	715	1.045	1.523	2.293	5.814

Fonte: os autores.

No Quadro 1 é possível observar que os CVs da área de Ciências Biológicas são aqueles atualizados mais recentemente (exceto pela última coluna). Já os currículos que não têm nenhuma área de atuação declarada são aqueles mais desatualizados (exceto, novamente, pela última coluna). A última coluna do Quadro representa a quantidade de dias passados desde a última atualização do currículo com data de atualização mais antiga de cada grupo. Por esse motivo, esta informação não é muito representativa.

O Quadro 2 apresenta a estimativa de artigos faltantes de acordo com a grande área de atuação. Ao se observar a quantidade de artigos completos publicados nos três anos anteriores a atualização de cada currículo, observa-se que as duas grandes áreas com maior quantidade são Ciências da Saúde e Ciências Humanas. Porém, ao se considerar a média dos artigos publicados nos últimos três anos por currículo, destacam-se *Ciências Agrárias* (média 1,8 artigos completos por CV no triênio) e *Ciências Biológicas* (1,6 artigos). A coluna *Estimativa de Artigos Faltantes* contém a estimativa do número total de artigos completos faltantes em cada uma das grandes áreas. Considerando-se todos os CVs, este valor é superior a 658 mil artigos. As colunas *Média* e *Mediana* contêm, respectivamente, a média e a mediana de artigos faltantes por currículo. Observa-se que a média é maior em *Ciências Agrárias* e *Ciências Biológicas*, que são justamente as áreas com maior média de artigos publicados no triênio anterior à última atualização de cada currículo. As menores médias de artigos faltantes estimados ocorrem nos CVs com área de atuação igual a *Outros/Multidisciplinar* e *Sem Grande Área*. Já a mediana tem valor igual a zero para todas as grandes áreas.

**Quadro 2** – Número de artigos completos publicados no triênio anterior à atualização e estimativa dos artigos faltantes por grande área

Grande Área	Total CVs	Artigos triênio	Artigos por currículo	Estimativa de Artigos Faltantes	Média	Mediana
Sem Grande Área	1.006.088	191.483	0,190	106.577	0,106	0
Ciências Agrárias	141.730	257.613	1,818	76.531	0,540	0
Ciências Biológicas	175.384	287.778	1,641	80.998	0,462	0
Ciências da Saúde	459.825	417.687	0,908	130.525	0,284	0
Ciências Exatas e da Terra	266.693	340.060	1,275	107.406	0,403	0
Ciências Humanas	417.334	414.949	0,994	99.713	0,239	0
Ciências Sociais Aplicadas	439.330	317.635	0,723	86.639	0,197	0
Engenharias	192.495	297.011	1,543	95.737	0,497	0
Linguística, Letras e Artes	157.597	114.446	0,726	28.477	0,181	0
Outros/ Multidisciplinar	337.017	45.317	0,134	9.173	0,027	0
Mundo Lattes	3.187.710	2.020.455	0,634	658.705	0,207	0

Fonte: os autores.

### 3.3 Formações acadêmicas/titulações

O Quadro 3 está organizado da mesma maneira que o Quadro 1, porém os CVs foram agrupados de acordo com sua maior formação/titulação acadêmica. É interessante notar que mais de dois terços dos CVs têm a graduação (completa ou em andamento) como maior formação. Dentre os grupos, aqueles de maior

formação são atualizados mais recentemente para a grande maioria das colunas. É possível observar, por exemplo, ao se analisar metade dos CVs dos doutores (ou doutorandos) da plataforma, que a última atualização foi realizada há no máximo 114 dias (menos de 4 meses). Já para os 20% dos currículos mais atualizados dos graduados (ou graduandos) a última atualização foi realizada há 141 dias.

**Quadro 3 – Número de dias desde a última atualização dos CVs por formação**

Grande Área	Total CVs	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
<b>Sem Formação Declarada</b>	108.758	621	793	1.202	2.021	2.290	2.442	2.584	2.804	3.202	5.814
<b>Ensino Fundamental Primeiro Grau</b>	15.508	123	176	434	561	697	783	841	921	1.156	4.157
<b>Ensino Médio Segundo Grau</b>	121.554	112	196	368	485	621	697	781	849	1.114	4.284
<b>Curso Técnico Profissionalizante</b>	30.940	205	552	678	781	878	1.044	1.329	1.762	2.604	5.786
<b>Graduação</b>	2.203.076	70	141	247	384	523	791	1.138	1.561	2.239	5.796
<b>Mestrado</b>	428.264	34	74	124	187	299	478	747	1.195	2.093	5.787
<b>Doutorado</b>	279.610	19	34	55	79	114	167	281	533	1.314	5.780
<b>Produtividade</b>	13.787	10	18	22	32	42	58	78	111	134	2.534

Fonte: os autores.

Já para os pesquisadores possuidores de bolsa produtividade, 80% dos currículos atualizados mais recentemente foram atualizados há menos de quatro meses da data em que os currículos foram baixados (111 dias). Por outro lado, os

CVs de pesquisadores que não possuem nenhuma formação declarada são aqueles que foram atualizados há mais tempo.

O Quadro 4 está organizada da mesma maneira que o Quadro 2, porém os CVs foram agrupados de acordo com a maior formação/titulação acadêmica.

**Quadro 4** – Número de artigos completos publicados no triênio anterior à atualização e estimativa dos artigos faltantes por formação

<b>Maior Formação/Titulação</b>	<b>Total CVs</b>	<b>Artigos triênio</b>	<b>Artigos por currículo</b>	<b>Estimativa de Artigos Faltantes</b>	<b>Média</b>	<b>Mediana</b>
<b>Sem Formação Declarada</b>	108.758	3.895	0,036	7.041	0,065	0,000
<b>Ensino Fundamental Primeiro Grau</b>	15.508	21	0,001	28	0,002	0,000
<b>Ensino Médio Segundo Grau</b>	121.554	971	0,008	444	0,004	0,000
<b>Curso Técnico Profissionalizante</b>	30.940	639	0,021	669	0,022	0,000
<b>Graduação</b>	220.3076	249.058	0,113	170.416	0,077	0,000
<b>Mestrado</b>	428.264	402.657	0,940	172.771	0,403	0,000
<b>Doutorado</b>	279.610	1.363.214	4,875	307.337	1,099	0,205
<b>Produtividade</b>	13.787	252.124	18,287	15.321	1,111	0,616
<b>Mundo Lattes</b>	3.187.710	2.020.455	0,63383	658.705	0,207	0,000

Fonte: os autores.

Dos mais de dois milhões de artigos completos publicados nos três anos anteriores às atualizações de cada CV, 87,4% destes artigos (1.765.871) pertencem

aos currículos que possuem como maior formação graduação, mestrado ou doutorado e, conseqüentemente, menos de 13% desses artigos estão registrados nos currículos sem a declaração de formação ou cuja maior formação é o ensino fundamental, médio ou técnico/profissionalizante. Conforme era de se esperar, entre os CVs que possuem alguma formação declarada, quanto maior a formação/titulação, maior a média de artigos publicados nos três anos anteriores à última atualização, variando de 0,001 artigos para os CVs cujo Ensino Fundamental corresponde à maior formação, passando por 0,113 para os CVs cuja maior titulação é a Graduação e chegando a 4,875 para os CVs com Doutorado. Para os CVs dos pesquisadores que possuem bolsa produtividade em pesquisa do CNPq, esta média chega a 18,287 artigos.

Ao se observar a média estimada de artigos faltantes, este mesmo padrão se observa: quanto maior a formação, maior a média. É interessante notar que a média de artigos estimados faltantes é bastante semelhante para os doutores e os bolsistas produtividade. Isto se dá pelo fato de, apesar dos bolsistas produtividade possuírem uma média de artigos por CV muito maior, seus CVs são atualizados mais frequentemente, então a relação entre artigos publicados no último triênio e tempo decorrido desde a última atualização é semelhante para estes dois grupos. A mediana do número estimado de artigos faltantes é igual a zero para todos os grupos, exceto para o grupo dos Doutores (0,205) e para o grupo dos bolsistas de produtividade (0,616).

### **3.4 Grandes-áreas *versus* formações acadêmicas/titulações**

Nesta seção são apresentados os dados referentes às atualizações dos currículos, aos artigos publicados e à estimativa de artigos faltantes cruzando-se as informações das grandes áreas de atuação com a maior formação/titulação acadêmica.

O Quadro 5 apresenta o número total de CVs de acordo com a grande área de atuação e maior formação/titulação declarada. Vale lembrar que em cada CV é possível declarar mais de uma área de atuação. Neste quadro foi destacada uma célula por coluna, indicando o maior valor de cada coluna.

**Quadro 5** – Total de CVs agrupados pela grande área de atuação e maior formação declarada

Grande Área/ Maior Formação/Titulação	Sem Formação Declarada	Ensino Fundamental	Ensino Médio	Curso Técnico Profissionalizante	Graduação	Mestrado	Doutorado	Produtividade
Sem Grande Área	71.894	3.382	19.997	13.132	78.8749	86.719	22.228	0
Ciências Agrárias	677	374	5.366	2.477	72.911	27.393	32.532	2.140
Ciências Biológicas	1367	181	2.849	755	85.395	36.608	48.229	3.489
Ciências da Saúde	10.390	163	9.502	2.833	315.776	68.519	52.642	2.164
Ciências Exatas e da Terra	2.743	2.610	8.205	3.333	147.004	52.092	50.707	4.049
Ciências Humanas	9.406	290	6.222	1.009	241.335	95.890	63.183	2.226
Ciências Sociais Aplicadas	9.115	210	8.852	2.044	286.147	89.910	43.054	1.136
Engenharias	1.533	253	5.303	3.097	106.968	40.311	35.030	2.491
Linguística, Letras e Artes	3.434	129	3.072	535	96.210	32.716	21.501	649
Outros/Multidisciplinar	1.174	7.971	53.525	3.063	251.615	13.265	6.404	138

Fonte: os autores.

É possível observar no Quadro 5 que, para as cinco primeiras formações, há prevalência de CVs cadastrados ou *Sem Grande Área* ou *Outros/Multidisciplinar*. Para CVs cuja formação máxima é a graduação, destaca-se a área de Ciências da Saúde, registrada em mais de 315 mil CVs. O mesmo ocorre para a formação máxima Doutorado (em mais de 52 mil CVs). Para os CVs cuja maior formação é o Mestrado, destaca-se a grande área *Ciências Humanas* (95.890 CVs). Por fim,

dentre os CVs dos possuidores de bolsa produtividade em pesquisa do CNPq, a grande área mais frequente é *Ciências Exatas e da Terra*.

O Quadro 6 apresenta as médias da variável dias desde a última atualização dos CVs para cada um dos grupos analisados. É possível observar que as maiores médias ocorrem para os currículos que não declararam grande área de atuação, exceto para aqueles cuja maior formação é o *Mestrado* (neste caso, a maior média ocorreu nas *Engenharias*).

As menores médias, isto é, os CVs na média mais atualizados foram aqueles que cadastraram a atuação na grande área *Outros/Multidisciplinar*, exceto para a formação máxima *Ensino Fundamental*, na qual os CVs mais atualizados, na média, foram os de *Ciências Exatas e da Terra*.

Destaca-se também no Quadro 6 o maior valor encontrado em todo o quadro: a média de 2.383 dias que ocorre nos 71.894 CVs que não declararam nenhuma grande área de atuação e nenhuma formação/titulação. Essa média corresponde a mais de 6,5 anos.

**Quadro 6** – Média de dias desde a última atualização dos CVs agrupados pela grande área de atuação e maior formação declarada

Grande Área/ Maior Formação/Titulação	Sem Formação Declarada	Ensino Fundamental	Ensino Médio	Curso Técnico Profissionalizante	Graduação	Mestrado	Doutorado	Produtividade
Sem Grande Área	2.383	1.170	1.303	1.397	1.243	828	768	-
Ciências Agrárias	1.861	772	468	985	752	713	362	44
Ciências Biológicas	1.574	651	591	1.363	759	582	346	44
Ciências da Saúde	1.268	703	617	841	713	624	367	39
Ciências Exatas e da Terra	1.462	260	495	1274	831	694	416	63
Ciências Humanas	1.088	654	604	855	700	557	330	49
Ciências Sociais Aplicadas	1.182	731	655	1.054	762	626	378	57
Engenharias	1.864	673	532	1.666	922	905	457	56
Linguística, Letras e Artes	1.152	529	613	957	702	563	344	62
Outros/Multidisciplinar	699	554	434	610	264	420	263	33

Fonte: os autores.

O Quadro 7 apresenta os valores da mediana da variável número de dias desde a última atualização dos CVs agrupados de acordo com a grande área de atuação e a formação. O comportamento dos valores deste quadro é bastante similar aos presentes no Quadro 6. Destaca-se o menor valor de todo o quadro, o qual indica que a mediana dos dias desde a última atualização dos 138 CVs dos bolsistas

produtividade em pesquisa do CNPq que declararam atuar em *Outros/Multidisciplinar* é de apenas 9 dias.

**Quadro 7** – Mediana de dias desde a última atualização dos CVs agrupados pela grande área de atuação e maior formação declarada

Grande Área/ Maior Formação/Titulação	Sem Formação Declarada	Ensino Fundamental	Ensino Médio	Curso Técnico Profissionalizante	Graduação	Mestrado	Doutorado	Produtividade
Sem Grande Área	2.440	1.126	1.124	1.210	1.143	410	236	-
Ciências Agrárias	1.703	702	417	703	362	267	69	13
Ciências Biológicas	976	611	479	759	369	215	72	16
Ciências da Saúde	826	680	654	710	419	249	87	11
Ciências Exatas e da Terra	843	116	436	748	428	272	88	26
Ciências Humanas	746	691	638	684	395	211	82	17
Ciências Sociais Aplicadas	774	697	666	731	444	258	95	19
Engenharias	1.089	677	492	817	437	424	101	24
Linguística, Letras e Artes	761	531	624	707	394	212	88	23
Outros/Multidisciplinar	459	649	432	487	194	187	83	9

Fonte: os autores.

O Quadro 8 apresenta o total de artigos completos publicados nos três anos

anteriores à atualização de cada CV, considerando a grande área de atuação e maior formação.

**Quadro 8** – Total de artigos completos nos três anos anteriores à última atualização de cada currículo agrupados pela grande área de atuação e maior formação declarada.

Grande Área/ Maior Formação/Titulação	Sem Formação Declarada	Ensino Fundamental	Ensino Médio	Curso Técnico Profissionalizante	Graduação	Mestrado	Doutorado	Produtividade
Sem Grande Área	1.622	0	187	168	75.186	61.671	52.649	0
Ciências Agrárias	73	5	111	144	18.910	35.278	203.092	48.773
Ciências Biológicas	258	1	81	60	16.914	33.196	237.268	62.198
Ciências da Saúde	1.115	0	93	34	54.667	70.730	291.048	52.263
Ciências Exatas e da Terra	204	3	79	127	19.238	47.779	272.630	67.824
Ciências Humanas	340	1	129	37	37.855	103.780	272.807	27.188
Ciências Sociais Aplicadas	311	0	65	17	32.384	88.280	196.578	18.097
Engenharias	117	0	38	100	16.353	39.377	241.026	62.508
Linguística, Letras e Artes	152	0	60	21	11.551	31.036	71.626	4.913
Outros/Multidisciplinar	60	11	203	31	8.292	11.135	25.585	2.938

Fonte: os autores.

O Quadro 9 apresenta a média de artigos completos publicados por CV, isto é, o número de artigos presente no Quadro 8 dividido pela quantidade de pesquisadores presente em cada um dos conjuntos de grande área e maior formação (Quadro 5). De um modo geral, as médias crescem de acordo com a formação, com exceção apenas dos dados nos quais não há declaração de formação. Para cada coluna, foi destacada a maior média. Tanto para o *Ensino Fundamental*, *Ensino*

*Médio e Profissionalizante* essa média é inferior a 0,1 artigos por CV em qualquer uma das grandes áreas. Para a *Graduação*, a maior média ocorre em *Ciências Agrárias* (0,259). Esta média sobe para 1,288 artigos por CV para os pesquisadores cuja maior formação é o *Mestrado* e que declararam a área de atuação *Ciências Agrárias*. Para os pesquisadores com *Doutorado*, a maior média de artigos completos no triênio ocorre nas *Engenharias* (6,881 artigos). Por fim, para os bolsistas produtividade, essa média varia de 7,570 artigos (para *Linguística, Letras e Artes*) até 25,094 (para *Engenharias*).

**Quadro 9** – Média de artigos completos publicados por CV nos três anos anteriores à última atualização de cada CV

Grande Área/ Maior Formação/Titulação	Sem Formação Declarada	Ensino Fundamental	Ensino Médio	Curso Técnico Profissionalizante	Graduação	Mestrado	Doutorado	Produtividade
Sem Grande Área	0,023	0,000	0,009	0,013	0,095	0,711	2,369	-
Ciências Agrárias	0,108	0,013	0,021	0,058	0,259	1,288	6,243	22,791
Ciências Biológicas	0,189	0,006	0,028	0,079	0,198	0,907	4,920	17,827
Ciências da Saúde	0,107	0,000	0,010	0,012	0,173	1,032	5,529	24,151
Ciências Exatas e da Terra	0,074	0,001	0,010	0,038	0,131	0,917	5,377	16,751
Ciências Humanas	0,036	0,003	0,021	0,037	0,157	1,082	4,318	12,214
Ciências Sociais Aplicadas	0,034	0,000	0,007	0,008	0,113	0,982	4,566	15,930
Engenharias	0,076	0,000	0,007	0,032	0,153	0,977	6,881	25,094
Linguística, Letras e Artes	0,044	0,000	0,020	0,039	0,120	0,949	3,331	7,570
Outros/Multidisciplinar	0,051	0,001	0,004	0,010	0,033	0,839	3,995	21,290

Fonte: os autores.

O Quadro 10 apresenta o número total de artigos completos faltantes, estimados segundo a metodologia apresentada, agrupados segundo grande área e maior formação.

**Quadro 10** – Total de artigos completos faltantes estimados agrupados pela grande área de atuação e maior formação declarada

Grande Área/ Maior Formação/Titulação	Sem Formação Declarada	Ensino Fundamental	Ensino Médio	Curso Técnico Profissionalizante	Graduação	Mestrado	Doutorado	Produtividade
Sem Grande Área	3.690	0	175	146	51.390	24.845	26.330	-
Ciências Agrárias	120	13	29	102	13.757	19.004	43.506	2.741
Ciências Biológicas	420	3	46	72	12.424	15.588	52.446	3.427
Ciências da Saúde	1.575	0	52	41	40.484	31.157	57.215	2.800
Ciências Exatas e da Terra	344	8	30	178	14.406	24.317	68.122	4.665
Ciências Humanas	463	1	32	12	20.157	33.268	45.780	1.523
Ciências Sociais Aplicadas	411	0	17	24	20.409	31.947	33.831	1.102
Engenharias	265	0	33	168	14.603	25.688	54.980	3.999
Linguística, Letras e Artes	243	0	17	13	5.883	9.712	12.609	297
Outros/Multidisciplinar	41	3	32	9	2.471	2.899	3.719	141

Fonte: os autores.

O Quadro 11 apresenta a média estimada de artigos faltantes por CV, isto é, os dados presentes no Quadro 10 divididos pelos dados do Quadro 5. Encontra-se destacado o maior valor para cada uma das colunas do quadro. Dos CVs *Sem*

*Formação Declarada* até os com maior formação igual ao *Mestrado*, as maiores médias encontram-se todas nas grandes áreas *Ciências Agrárias e Ciências Biológicas*, que são justamente as grandes áreas nas quais a média de artigos publicados no triênio anterior à atualização são maiores (ver Quadro 9). Já para os CVs cuja maior formação/titulação é o *Doutorado* e para aqueles que possuem bolsa produtividade, as maiores médias ocorreram nas *Engenharias*, seguindo novamente o comportamento identificado no Quadro 9. Isso indica que, apesar do número de dias desde a última atualização dos currículos variar entre as áreas (Quadros 6 e 7), o valor médio estimado de artigos faltantes apresentou um comportamento bastante parecido com o apresentado pela variável média de artigos publicados no triênio anterior à atualização.

Também é importante observar pelo Quadro 11 que se estima que, na média, ao se analisar um grupo de pesquisadores cuja maior formação/titulação é o *Doutorado*, estará faltando de 0,581 artigos por CV (*Outros/Multidisciplinar*) até 1,569 artigos (*Engenharias*). Essa estimativa pode corresponder a mais de 20% do total das publicações (ver Quadro 11 e Quadro 9), podendo assim interferir significativamente em pesquisas que não considerem a provável desatualização dos currículos.

O Quadro 12 apresenta a mediana do número estimado de artigos faltantes por CV, agrupada de acordo com a grande área de atuação e a formação. Para a grande maioria dos grupos, a mediana vale zero, isto é, é estimado que para mais de metade dos currículos desses grupos não esteja faltando nenhum artigo. Isto se dá porque a maioria dos CVs não possui nenhum artigo completo publicado. Há medianas diferentes de zero apenas para os doutores e para os possuidores de bolsa produtividade. Os maiores valores para essa mediana ocorrem nas *Engenharias*, chegando a 0,995 artigos para os CVs que possuem bolsa produtividade, ou seja, estima-se que para metade dos CVs deste grupo esteja faltando cerca de um artigo ou mais.

**Quadro 11** – Média de artigos completos faltantes estimados por CV agrupados pela grande área de atuação e maior formação declarada

Grande Área/ Maior Formação/Titulação	Sem Formação Declarada	Ensino Fundamental	Ensino Médio	Curso Técnico Profissionalizante	Graduação	Mestrado	Doutorado	Produtividade
Sem Grande Área	0,051	0,000	0,009	0,011	0,065	0,287	1,185	-
Ciências Agrárias	0,178	0,035	0,005	0,041	0,189	0,694	1,337	1,281
Ciências Biológicas	0,307	0,015	0,016	0,095	0,145	0,426	1,087	0,982
Ciências da Saúde	0,152	0,000	0,006	0,015	0,128	0,455	1,087	1,294
Ciências Exatas e da Terra	0,126	0,003	0,004	0,053	0,098	0,467	1,343	1,152
Ciências Humanas	0,049	0,002	0,005	0,012	0,084	0,347	0,725	0,684
Ciências Sociais Aplicadas	0,045	0,000	0,002	0,012	0,071	0,355	0,786	0,970
Engenharias	0,173	0,000	0,006	0,054	0,137	0,637	1,569	1,606
Linguística, Letras e Artes	0,071	0,000	0,005	0,024	0,061	0,297	0,586	0,458
Outros/Multidisciplinar	0,035	0,000	0,001	0,003	0,010	0,219	0,581	1,021

Fonte: os autores.

**Quadro 12** – Mediana do número estimado de artigos faltantes por CV agrupados pela grande área de atuação e maior formação declarada

Grande Área/ Maior Formação/Titulação	Sem Formação Declarada	Ensino Fundamental	Ensino Médio	Curso Técnico Profissionalizante	Graduação	Mestrado	Doutorado	Produtividade
Sem Grande Área	0	0	0	0	0	0	0,043	-
Ciências Agrárias	0	0	0	0	0	0	0,271	0,748
Ciências Biológicas	0	0	0	0	0	0	0,217	0,589
Ciências da Saúde	0	0	0	0	0	0	0,252	0,773
Ciências Exatas e da Terra	0	0	0	0	0	0	0,232	0,628
Ciências Humanas	0	0	0	0	0	0	0,173	0,384
Ciências Sociais Aplicadas	0	0	0	0	0	0	0,153	0,537
Engenharias	0	0	0	0	0	0	0,329	0,995
Linguística, Letras e Artes	0	0	0	0	0	0	0,137	0,252
Outros/Multidisciplinar	0	0	0	0	0	0	0,121	0,592

Fonte: os autores.

#### 4 Considerações finais

Alguns dos principais fatores considerados na análise de dados, como a completude, correte e atualização dos dados (CAÑIBANO; BOZEMAN, 2009), são pouco tratados na literatura. Apesar dos dados contidos na Plataforma Lattes serem de grande valia para pesquisas bibliométricas e cientométricas, nenhum destes fatores são garantidos, pois os três dependem dos usuários que registraram as informações

em seus currículos e do recorte utilizado para selecionar os CVs. Mesmo com estas limitações, a quantidade e a riqueza da informação disponível nessa plataforma são tão grandes que justificam sua ampla utilização.

Neste trabalho foi analisada a data da última atualização dos CVs considerando-se as diferentes áreas de atuação e maior formação/titulação acadêmica. Com a análise apresentada, pretende-se deixar mais claro quais grupos mantêm os CVs mais atualizados (visão macro) de forma a auxiliar os trabalhos futuros na seleção dos grupos de currículos a serem utilizados e/ou dos recortes a serem feitos. Por exemplo, supondo que foram obtidos os CVs da Plataforma Lattes durante o mês de julho de 2013 e deseja-se realizar um estudo sobre a área de *Engenharias* considerando apenas doutores. Para ter alguma garantia que os dados usados estejam atualizados, pode-se consultar a média e a mediana da variável dias desde a última atualização dos CVs, que são 457 e 101, respectivamente. Baseando-se nesses dados, pode-se decidir fazer um recorte no estudo, considerando, por exemplo, apenas a informação dos CVs até 2011.

Uma vez que agências de fomento, centros de pesquisa e departamentos das universidades utilizam os dados da Plataforma Lattes para análise dos CVs, especialmente para conhecerem a trajetória e produção, é de se esperar que alunos de graduação que pretendem seguir dentro da comunidade acadêmica, pesquisadores e professores atualizem seus CVs com mais frequência, e que aqueles CVs dos pesquisadores com maior formação sejam os CVs atualizados mais recentemente. Adicionalmente, foram apresentados os números totais de artigos completos publicados nos três anos anteriores à última atualização dos CVs e, combinando-se essa informação com as datas de atualização, foi possível estimar a quantidade de artigos faltantes nos CVs devido a potencial desatualização da informação cadastrada.

Estima-se que uma quantidade significativa de informação esteja ausente dos currículos devido à falta de atualização. Segundo a estimativa realizada neste trabalho, para artigos completos, foi possível observar que, para alguns grupos de pesquisadores, essa falta possa representar mais de 20% do total de artigos publicados no último triênio.

Finalmente, considerando que as análises bibliométricas são comumente pautadas em bases de dados de revistas científicas, e sendo os currículos uma fonte de informação curricular alternativa, o questionamento sobre a atualidade dos dados também deve ser observado naquele caso, uma vez que as revistas muitas vezes falham na pontualidade de publicação.

Como trabalhos futuros, pretende-se estudar a qualidade e completude da informação presente nos currículos da Plataforma Lattes de forma micro.

## Referências

ALVES, A. D. et al. LattesMiner: a multilingual DSL for information extraction from Lattes platform. In: SPLASH'11 WORKSHOPS, 11. 2011, Portland. **Proceedings...** New York, 2011, p. 85-92.

ARRUDA, D. et al. Brazilian computer science research: gender and regional distributions. **Scientometrics**, Amsterdam, v. 79, n. 3, p. 651-665, jun. 2009.

CAÑIBANO C.; BOZEMAN B. Curriculum vitae method in science policy and research evaluation: the state-of-the-art. **Research Evaluation**, Guildford, v. 18, n. 2, p. 86-94, 2009.

CONSELHO NACIONAL DE DESENVOLVIMENTO CIENTÍFICO E TECNOLÓGICO (CNPq). **Dados e estatísticas da Plataforma Lattes**. 2014. Disponível em: <<http://www.cnpq.br/web/portal-lattes/dados-e-estatisticas>>. Acesso em 20 ago. 2014.

COSTA, B. G. et al. Scientific collaboration in biotechnology: the case of the northeast region in Brazil. **Scientometrics**, Amsterdam, v. 95, n. 2, p. 571-592, maio 2013.

DIGIAMPIETRI, L. A. et al. Dinâmica das relações de coautoria nos programas de pós-graduação em computação no Brasil. In: BRAZILIAN WORKSHOP ON SOCIAL NETWORK ANALYSIS AND MINING, 1., 2012, Curitiba. **Anais...** Curitiba, 2012b.

DIGIAMPIETRI, L. A. et al. Minerando e caracterizando dados de Currículos Lattes. In: BRAZILIAN WORKSHOP ON SOCIAL NETWORK ANALYSIS AND MINING, 1., 2012, Curitiba. **Anais...** Curitiba, 2012a.

DIGIAMPIETRI, L. A. et al. Predição de coautorias em redes sociais acadêmicas: um estudo exploratório em Ciência da Computação. In: BRAZILIAN WORKSHOP ON SOCIAL NETWORK ANALYSIS AND MINING, 2., 2013, Maceio. **Anais...** Maceio: BraSNAM, 2013.

LEITE, P., MUGNAINI, R.; LETA, J. A new indicator for international visibility: exploring brazilian scientific community. **Scientometrics**, Amsterdam, v. 88, n. 1, p. 311-319, jul. 2011.

LIBEN-NOWELL, D.; KLEINBERG, J. The link prediction problem for social networks. In: INTERNATIONAL CONFERENCE ON INFORMATION AND KNOWLEDGE MANAGEMENT, 12., 2003, New York. , **Anais...** New York: ACM, 2003. p. 556.

MARQUES, K. A plataforma Lattes e a organização da informação. **Gestão & Planejamento**, Salvador, v. 11, n. 2, p. 250-266, 2010.

MENA-CHALCO, J. P. et al. Brazilian bibliometric coauthorship networks. **Journal of the Association for Information Science and Technology**, Hoboken, v. 65, n. 7, p. 1424-1445, jul. 2014.

MENA-CHALCO, J. P.; CESAR JUNIOR, R. M. ScriptLattes: an open-source knowledge extraction system from the Lattes platform. **Journal of the Brazilian Computer Society**, Rio de Janeiro, v. 15, n. 4, p. 31-39, dez. 2009.

SILVA, F. M.; SMIT, J. W. Organização da informação em sistemas eletrônicos abertos de Informação Científica & Tecnológica: análise da Plataforma Lattes. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 14, n. 1, jan./abr. 2009 .

WAINER, J.; VIERA, P. Correlations between bibliometrics and peer evaluation for all disciplines: the evaluation of Brazilian scientists. **Scientometrics**, Amsterdam, v. 96, n. 2, p. 395-410, ago. 2013.

## Macro analysis of Lattes Curricula latest updates

**Abstract:** The completeness and the updating of data are key factors to any scientific analysis. The Lattes Platform provides a huge amount of data that can be used in different bibliometric studies as well as in social network analysis. Despite the volume and importance of the data provided by this platform, there are no guarantees regarding the completeness, correctness and update situation of these data, as they are registered by the platform users. This work aims to analyze the data update of the Lattes Platform considering over three million curricula and analyzing the data according to the area of interest and the last update. Additionally, historical information about the publication of full papers was used to estimate the number of missing items due to the lack of updates.

**Keywords:** Lattes Curricula. Information update. Bibliometric analysis.

Recebido: 28/07/2014

Aceito: 24/11/2014