



A avaliação da produtividade científica

The evaluation of scientific productivity

Marcos Antônio Mattedi

Professor, Programa de Pós-graduação em
Desenvolvimento Regional; coordenador, Núcleo de Estudos da
Tecnociência (NET)/Universidade Regional de Blumenau.
Rua Antônio da Veiga, 140, Bloco R
89012-900 – Blumenau – SC – Brasil
mattediblu@gmail.com

Maiko Rafael Spiess

Professor, Departamento de Ciências Sociais e Filosofia;
coordenador, NET/Universidade Regional de Blumenau.
Rua Antônio da Veiga, 140, Bloco R
89012-900 – Blumenau – SC – Brasil
mspiess@furb.br

Recebido para publicação em dezembro de 2015.

Aprovado para publicação em agosto de 2016.

<http://dx.doi.org/10.1590/S0104-59702017000300005>

MATTEDI, Marcos Antônio; SPIESS, Maiko Rafael. A avaliação da produtividade científica. *História, Ciências, Saúde – Manguinhos*, Rio de Janeiro, v.24, n.3, jul.-set. 2017, p.623-643.

Resumo

O texto trata da avaliação da produtividade científica. Analisa o processo de metrificacão da avaliação da produçao científica, bem como o processo histórico de construçao da avaliação científica e seus usos atuais. Argumenta que esse processo encerra um paradoxo: quanto mais impessoais se tornam as métricas, menor seu reconhecimento pelos cientistas. O estudo foi dividido em cinco partes: contextualizacão da problemática da avaliação científica; descriçao das principais etapas do processo de institucionalizacão da metrificacão; apresentacão do processo de concepçao dos principais indicadores de avaliação; exemplos da aplicacão desses indicadores; apresentacão das consequências analíticas e algumas recomendações para formulacão de uma nova agenda de avaliação.

Palavras-chave: avaliação; produtividade; metrificacão; bibliometria; indicadores.

Abstract

The paper examines the evaluation of scientific productivity. It analyzes the metrification of the evaluation of scientific production, as well as the historical construction, and current uses of scientific evaluation. It argues that this process contains a paradox: the more that metrics become impersonal, the less they are recognized by scientists. The study is divided into five sections: contextualization of the problematics of scientific evaluation; a description of the main stages in the institutionalization of metrification; an overview of the development of the main evaluation indexes; some examples of the application of these indexes; and analytical consequences and recommendations for the formulation of a new evaluation agenda.

Keywords: evaluation; productivity; metrification; bibliometrics; indicators.

Dizem que a melhor forma de conhecer uma instituição é descobrir suas obsessões. Por isso, se quisermos entender a comunidade científica temos que investigar a avaliação da produtividade, que constitui os mecanismos por meio dos quais a comunidade científica certifica e controla a produção do conhecimento. A avaliação da produtividade é ubíqua na atividade científica e costuma ser utilizada para fins muito diversos como, por exemplo, concessão de financiamentos ou bolsas, recrutamento ou promoções na carreira científica, atribuição de prêmios ou distinções etc. Portanto, a avaliação está nos ranqueamentos universitários, mas também está presente na análise dos programas de pesquisa, na classificação de revistas, na qualidade e padrões de citações dos artigos. Assim, a massificação da comunicação científica acabou convertendo a avaliação em classificação, e a classificação em controle da atividade científica (Gingras, 2014).

A comunidade científica constitui uma instituição reputacional (Whitley, 1982). Sua organização se estabeleceu com base em formas espontâneas de avaliação e, conseqüentemente, hierarquias implícitas do valor do conhecimento. Porém, com a massificação da atividade científica essas regras foram substituídas por um sistema de avaliação formal. Sabemos pelos estudos de estratificação social que toda avaliação endossa escalas de classificação que atribuem prestígio a pessoas ou grupos sociais, mas também a profissões, locais ou objetos (Constans, Rivoal, 2014). Na comunidade científica, o prestígio e a reputação estruturam a produção do conhecimento científico: quanto mais original a descoberta científica, maior o reconhecimento na comunidade científica (Hagstron, 1965). Em termos de avaliação, a literatura internacional (Donavan, 2007), diferencia duas estratégias: (a) abordagens qualitativas; (b) abordagens quantitativas.

A abordagem qualitativa de avaliação fundamenta-se no *peer review*. Por muito tempo a produtividade científica ficou restrita a avaliação por pares. Esse procedimento remonta ao final do século XVII, quando em 1665 o *Journal des Sçavants*, na França, e o *Philosophical Transactions*, da Royal Society, na Inglaterra, instituíram a prática da opinião especializada para avaliação da comunicação científica. Progressivamente ela foi sendo estendida para avaliação do desempenho de departamentos em universidades, programas de pesquisa, laboratórios, periódicos, disciplinas, pesquisadores etc. Contudo, o aumento excessivo do número de avaliações, agravado pela competência às vezes inadequada de alguns avaliadores, bem como conflitos de interesse, foi criando problemas na avaliação por revisores (Bornmann, 2008). Por isso, nos últimos anos o *peer review* tem sido objeto de críticas, e a centralidade de seu papel na avaliação científica tem sido progressivamente contestada do ponto de vista da eficácia no controle da qualidade dos artigos e da subjetividade, do corporativismo, conservadorismo e conflito de interesses dos pareceristas (Shatz, 2004; Smith, 2006; Manchikanti et al., 2015).

A abordagem quantitativa está associada ao desenvolvimento da bibliometria para avaliação da produtividade. A bibliometria constitui o resultado da progressiva convergência da estatística, da sociologia e da informática para avaliar pesquisadores, equipes ou instituições. Mais precisamente, constitui os procedimentos que contribuem para a avaliação da produtividade a partir do número de publicações, o prestígio das revistas em que foram realizadas e dos padrões de citações (Académie des Sciences, 2011). Esses procedimentos vêm ganhando cada vez mais importância porque oferecem um conjunto de informações sintéticas e factuais da dinâmica de funcionamento da comunidade científica. Muito embora

a bibliometria possa apresentar graves distorções quando utilizada isoladamente, sua aplicação na avaliação da produtividade científica vem suscitando grande entusiasmo e transformou o artigo científico em um fator central.

A progressiva substituição do *peer review* pelas “métricas” nos processos de recrutamento, formulação de políticas de financiamento, avaliação institucional etc. (Gingras, 2014) está associada a uma profunda transformação da autocompreensão e organização da comunidade científica. Por um lado, ela reflete o enfraquecimento da descrição epistemológica da ciência como uma atividade cognitiva racional (Popper, 2006) e o estabelecimento de uma visão sociológica da ciência como uma atividade social (Kuhn, 1989); por outro, reflete uma transformação de ordem sociológica, uma vez que o processo de massificação da ciência transforma a produção do conhecimento numa atividade coletiva com crescente investimento de recursos (Price, 1963). Portanto, o desenvolvimento de monitoramento e direção científica está associado a dois processos: avaliação e financiamento (Whitley, 2000). A relação prática entre esses dois processos encerra um paradoxo: quanto maior a excelência científica do conhecimento (originalidade), menor a acessibilidade social desse conhecimento (compreensão).

Dessa forma, é possível afirmar que a ênfase na quantificação da atividade científica possui inúmeras consequências inesperadas. Ao mesmo tempo em que proporciona formas de simplificação e objetivação de fatos sociais (a produtividade dos cientistas e instituições, suas relações de colaboração ou a dinâmica de um campo de conhecimento) por meio de sua “numerização” (Desrosières, 1998), torna-se um fim em si mesma. Em muitos casos, em vez de orientar a política científica e a distribuição de recursos, a lógica da avaliação da produtividade leva ao “produtivismo”, abrindo caminho para práticas como a autocitação espúria, a “ciência salame” (a duplicação ou “fatiamento” da publicação de resultados em diversos artigos), o crescimento de casos de plágio e de retratação científica (Castiel, Sanz-Valero, Red Mei-Cyted, 2007; De Bellis, 2014; Sguissardi, Silva, 2009). Ou seja: a centralidade das formas de avaliação parece pressionar os cientistas, introduzindo desvios e anomalias em suas práticas tradicionais.

Da percepção quanto às limitações da avaliação científica e aos impactos do produtivismo surgem críticas e reações, muitas vezes apresentadas em forma de manifestos de pesquisadores, organizações e associações: por exemplo, a San Francisco Declaration on Research Assessment (Dora), publicada em 2012, questiona a correlação entre medidas como o fator de impacto de uma publicação e o mérito dos pesquisadores (Alberts, 2013); o Leiden Manifesto (2015) parte da premissa de que as formas de avaliação são “normalmente bem intencionadas, nem sempre bem informadas e frequentemente mal aplicadas” (Hicks et al., 2015, p.429)¹ para então sugerir dez novos princípios para a avaliação de pesquisas; o Force11 Manifesto: Building the Future for Research Communications and e-Scholarship (2011) procura refletir sobre novas formas para a publicação científica. Em comum, essas iniciativas apontam para um esgotamento do modelo vigente de elaboração e aplicação da avaliação científica.

Por fim, é possível criticar o estado atual da avaliação científica não apenas do ponto de vista prático, mas também ético (Furner, 2014). Todas as formas de avaliação são arbitrárias, representando diferentes visões sobre a atividade científica e escolhas relacionadas com a distribuição de recursos, mérito e visibilidade de pesquisadores e instituições. A

sistematização e o emprego de certas formas de avaliação, portanto, favorecem certas práticas, ao mesmo tempo que desencorajam outras. Isso significa uma homogeneização da atividade e o sufocamento da dissidência, isto é, a imposição de certas formas de atuação, valores e prioridades. Mais do que isso, representa um problema de justiça distributiva: quando certas práticas e perfis profissionais são mais valorizados, eles tendem a gerar acumulação de recursos e prestígio. Assim, a avaliação impõe obstáculos e dificuldades adicionais para os pesquisadores novatos, envolvidos com temas negligenciados ou localizados na periferia do sistema.

Portanto, em nosso ponto de vista, a metrificação constitui o efeito combinado desses processos sociotécnicos. Mais precisamente, a metrificação compreende a operação, ao mesmo tempo, cognitiva e normativa por meio da qual se procura transformar a avaliação da produtividade numa atividade imparcial e confiável (Porter, 1995). Trata-se de conceber e aplicar instrumentos de medição objetivos, ou seja, da standardização de medidas de avaliação. Esse processo está relacionado tanto à formação de *experts* para produção de dados como também ao estabelecimento de uma base social de autoridade dessas medidas. Nesse sentido, o artigo sustenta, por meio da análise do processo de formação e desenvolvimento da avaliação científica, que a metrificação exprime a transformação do artigo no principal produto da atividade científica. Essa transformação possui três implicações, que denominamos: (a) papirocentrismo; (b) produtivismo; (c) mimetismo.

Considerando essa caracterização do processo de metrificação da avaliação, o argumento do texto está dividido em três partes: a primeira trata da questão da institucionalização histórica da metrificação, destacando as principais etapas desse processo; a segunda parte trata dos indicadores para a avaliação científica e sua crescente diversidade e complexidade, explorando alguns dos principais indicadores de produtividade e relacionais; a terceira aborda a prática atual da avaliação científica, focando em seus possíveis usos e apontando suas vantagens e desvantagens. Finalmente, o artigo apresenta proposições para superar os dilemas colocados pela situação atual da discussão sobre a avaliação da atividade científica. Portanto, a partir da análise do processo de institucionalização das métricas e sua aplicação, o artigo demonstra o surgimento e consolidação da metrificação na avaliação científica.

A institucionalização da avaliação científica

A preocupação com a geração, comunicação e aplicação do conhecimento acompanha o processo de formação e desenvolvimento da comunidade científica (Mattedi, Spiess, 2010). Por um lado, está relacionada ao aumento do número de publicações e à percepção de que esse processo poderia ser descrito cientificamente; por outro, ao progressivo aumento de investimentos e ao redimensionamento dos limites acadêmicos e institucionais da pesquisa. Por isso a necessidade de avaliação está relacionada à aplicação de técnicas estatísticas à literatura científica, mas também à necessidade de gestão e controle. Mais precisamente, à necessidade de compreender o significado, as características e as diferenças do efeito combinado da produção individual dos cientistas em diferentes disciplinas. Nesse sentido, a preocupação da avaliação do desenvolvimento científico pode ser dividida em três fases principais: (a) fase de concepção; (b) fase de estabilização; (c) fase de disseminação.

A institucionalização do processo de avaliação tem início nas primeiras décadas do século XX, com a aplicação de técnicas quantitativas de análise das publicações, autores e referências bibliográficas. Fundamenta-se na forma canônica de comunicação científica baseada no artigo científico. O artigo científico possuía vantagens, pois apresentava um autor ou coautores facilmente identificáveis e uma bibliografia inerte, que se tornavam imutáveis a partir da publicação em papel. Ou seja, as preocupações iniciais com a avaliação constituem a extensão progressiva de técnicas de manipulação estatística de dados, tendo como base empírica o artigo científico. Isso envolve a emergência de uma atitude reducionista de avaliação da ciência a uma dimensão observável, facilmente colecionável, e a unidades contáveis num perfeito estilo positivista (De Bellis, 2014). Trata-se, portanto, de entender a ciência como uma instituição social pela análise de seus membros e seus produtos.

Essa perspectiva se consolidou em torno de diferentes leis bibliométricas. Uma das primeiras contribuições dessa natureza foi formulada pelo químico Alfred J. Lotka e publicada em 1926. A Lei de Lotka (ou Lei do Quadrado Inverso) indica que o número de autores que fazem “n” contribuições em uma área do conhecimento é aproximadamente “ $1/n^2$ ” daqueles que fazem apenas uma contribuição, ou seja, em uma dada área do conhecimento, a proporção dos autores que contribuem com apenas uma publicação é algo em torno de 60% do total dos autores de um campo (Coile, 1977). Em um sentido semelhante, a Lei de Bradford (proposta originalmente em 1934) se volta para os periódicos e procura “estabelecer o núcleo e as áreas de dispersão sobre um determinado assunto em um mesmo conjunto de revistas” (Vanti, 2002, p.153). Por fim, convém mencionar a Lei de Zipf, que analisa a frequência de uso de palavras em diferentes textos, criando uma ordem dos termos mais utilizados em um campo científico.

Essas leis foram praticamente ignoradas pela comunidade científica até Derek de Solla Price retomar a discussão proposta por Lotka em seus livros a respeito do crescimento da ciência (McRoberts, McRoberts, 1982). Desde então, suas influência e interpretações têm variado enormemente. Do ponto de vista conceitual e metodológico, sua reprodutibilidade vem sendo testada em diferentes áreas do conhecimento e bases de dados, com resultados ambíguos (Pinheiro, 1983; Urbizagástegui Alvarado, 2002). De forma particular, apresentam-se questionamentos sobre sua validade e seu alcance, quando consideradas diferentes áreas do conhecimento (como, por exemplo, as ciências humanas) e diferentes condições contextuais. Trata-se, portanto, de dúvidas acerca da universalidade dessas leis e de suas implicações para a organização da atividade científica. No final das contas, é possível compreender a atividade científica apenas a partir da medição de seus produtos?

Em resumo, as perspectivas de Lotka, Bradford e Zipf fundamentam-se na análise estatística das atividades dos indivíduos que compõem a comunidade científica. Baseiam-se na noção de autoria, conforme expressa pelas publicações em revistas especializadas. Por um lado, essa é uma visão demasiadamente restrita das funções da ciência, como aponta John D. Bernal (1967): considera apenas o produto final da atividade científica, excluindo a *grey literature*, a transferência de conhecimento tácito, ou seja, o cotidiano da ciência. Por outro lado, explicitam o efeito Mateus, conforme proposto por Robert K. Merton (1968): pesquisadores eminentes tendem a receber mais crédito do que colegas com menor reputação e, com isso, acumulam mais prestígio, acesso a recursos e visibilidade. Apesar dessas limitações, as

contribuições pioneiras de Lotka e Bradford permanecem exercendo grande influência nas estratégias de avaliação científica.

Após a Segunda Guerra Mundial, a proliferação e o desafio de avaliar a relevância da literatura científica se converteram num obstáculo para a realização de novas pesquisas. Assim, inspirado no Shepard's Citations (um índice empregado na área jurídica), Eugene Garfield propõe a criação do Science Citation Index (SCI). Como editor da revista *American Documentation*, Garfield encomenda um artigo para William Adair (1955), vice-presidente da Shepards Company, que descreve o funcionamento do utensílio jurídico. Alguns meses depois, Garfield (1955) apresenta na revista *Science* a proposta de uma base de dados que facilitasse a utilização de artigos citados para encontrar outros artigos. Mais precisamente, a proposta fundamentava-se na intuição de que existe uma ligação conceitual entre o artigo que cita e o artigo que é citado (Gingras, 2014). Assim, o SCI estabeleceu uma nova imagem da literatura científica, da mesma forma como uma lista telefônica cria uma imagem dos habitantes de uma cidade (Wouters, 1999).

Convencido da aplicabilidade de sua proposta, Garfield funda em 1959 o Institute for Science Information (ISI). O termo substitui a denominação Eugene Garfield Associates Inc., para dar a impressão de se tratar de uma agência governamental e poder competir com organizações sem fins lucrativos. Em 1961, obtém o apoio de US\$ 300.000 da National Science Foundation (NSF) e do National Institute of Health (NIH) para criação de um *index* automatizado, uma ferramenta de divulgação e recuperação de informações (Garfield, 2007). Após esse suporte inicial, o NIH foi impedido de subsidiar empresas. Então, o NSF negociou um contrato com o ISI por mil cópias de um Genetics Citation Index (GCI). Nesse sentido, o SCI fornece uma nova representação da ciência em termos de produção e consumo da informação científica.

O propósito inicial do SCI foi facilitar a busca de literatura. Como assinala Garfield (2007, p.65): "O banco de dados multidisciplinar da SCI tem duas finalidades: em primeiro lugar, identificar o que cada cientista publicou e, em segundo, onde e quantas vezes os textos por cientista são citados". Portanto, o SCI visava resolver o problema do acesso rápido e automático, reduzindo a complexidade da linguagem científica a um grupo de unidades manejáveis. Ou seja, um grupo particular de relacionamentos metatextuais, obtidos por meio do vínculo entre os artigos das revistas e as referências bibliográficas. O interesse na análise do SCI decorre de que, como uma ferramenta do produtivismo, ele efetivamente oblitera o conteúdo da literatura ao focar em suas propriedades formais. Portanto, o efeito desse processo estabelece uma nova representação da ciência, diferente da descrição do processo cognitivo ou comportamental dos cientistas (Wouters, 1999).

O SCI imediatamente despertou atenção da sociologia da ciência, pois permite investigar o funcionamento da comunidade científica. Ao agregar as propriedades de publicações, referências e citações, permite testar as hipóteses do programa mertoniano do comportamento dos cientistas (Mattedi, 2006), o que desencadeou uma disputa e posterior aproximação entre a Columbia School (de orientação mertoniana) e a Philadelphia School (influenciada por Price, com acentos quantitativos) (Elkana et al., 1978; Wouters, 1999). Derek de Solla Price foi o primeiro autor a utilizar o índice criado pelo SCI: examinando os dados quantitativos do desenvolvimento da ciência, Price (1978a) indica um padrão de distribuição desigual na

produção científica. Isso significa que o crescimento da informação científica se produz a um ritmo muito superior ao de outros fenômenos sociais, mas muito similar a outros fenômenos observáveis em contextos naturais. Essa relação ficou conhecida como Price's Law: 25% dos autores científicos são responsáveis por 75% dos artigos publicados (Price, 1963).

Com os avanços da eletrônica e da ciência da computação, novas possibilidades se abriram para a medição e análise da produção científica. Desde a Segunda Guerra Mundial, progressivas inovações tecnológicas (circuitos integrados, microprocessadores) permitiram uma gradual miniaturização e um aumento da capacidade de processamento dos computadores. Esse processo possibilitou o surgimento dos microcomputadores, nos anos 1970, e dos computadores pessoais, na década de 1980 (Mowery, Rosenberg, 2005). Em particular, a disseminação dessas tecnologias implicou uma “revolução” que atingiu várias áreas administrativas e produtivas das sociedades ocidentais, significando um ganho de tempo e escala na realização de tarefas, ou seja, grandes volumes de dados puderam ser analisados e processados em um curto intervalo de tempo. Assim, a aplicação dessas tecnologias possibilitou mudanças significativas também na área da avaliação científica.

Inicialmente, convém salientar que pesquisadores e iniciativas como o SCI encontravam nas tecnologias computacionais o suporte necessário para a implementação de formas cada vez mais complexas de análise estatística. Essa correlação entre a tecnologia disponível e a avaliação da atividade científica foi destacada por Solla Price (1978b) no editorial da primeira edição da revista *Scientometrics* e se consolidou, por exemplo, com a medição do fator de impacto. Originalmente proposto por Garfield em 1955, o fator de impacto – uma forma de avaliação baseada não apenas na quantidade de publicações individuais dos autores, mas em sua influência, ou na importância de um periódico em um dado campo do conhecimento – era produzido a partir da análise de todos os periódicos cobertos pelo SCI. Em 1969, isso significava analisar milhares de referências contidas em mais de 2.200 revistas (Bensman, 2007), uma tarefa impossível de ser realizada sem a ajuda de tecnologias de computação.

Mais recentemente, com a popularização dos computadores pessoais e da internet, a divulgação científica passou a ser cada vez mais eletrônica. Em um primeiro momento, isso significou uma multiplicação e fragmentação das bases de dados, em função de campos do conhecimento e abrangência geográfica ou linguística. Além disso, gradualmente se estabeleceram novas formas de acesso ao conhecimento científico e, conseqüentemente, novas práticas e problemas relacionados com avaliação da atividade científica. Todavia, uma das mais importantes inflexões ocorre em 1992, com a aquisição do ISI e seus produtos (o SCI e índices semelhantes) pela Thomson Corporation e a criação da plataforma Web of Science (De Bellis, 2014). Foi a partir desse momento que os avanços conceituais e os meios técnicos se complementaram e se consolidaram em torno de um modelo de negócios, dando origem a outras bases, como a Scopus, da Elsevier, e o Google Scholar.

A associação das Tecnologias de Informação e Comunicação (TICs) com a atividade científica modificou profundamente a forma de comunicação e avaliação da atividade científica. Mais precisamente, a publicação de artigos *on-line* ou em periódicos científicos híbridos, relacionados ao longo de uma miríade de outros *papers* e objetos digitais de relevância potencial, conectados em tempo real e *hiperlinks*, estabeleceu um novo patamar de avaliação.

O efeito combinado desse processo foi a constituição de um caleidoscópio de domínios de avaliação como, por exemplo, a infometria, a cientometria, a cybermetria, a webmetria, a influentometria, a digimetria, entre outros neologismos que exprimem esse processo. Ao mesmo tempo, verifica-se o surgimento de uma parafernália de revistas, sociedades profissionais, conferências, prêmios, currículos, centros de pesquisa relacionados à avaliação científica que envolvem fundos de investimento, corporações empresariais e universidades (Cronin, 2014).

Nesse sentido, a instituição da cultura da avaliação exprime a institucionalização do artigo e a incorporação das referências como parâmetro de avaliação da atividade científica. Portanto, a adoção de medidas de produtividade (contagem de publicações) ou indicadores de impacto (contagem de citações) indica a passagem da *little science* para a *big science* (Price, 1963). Expressa também as relações de cooperação e competição entre Merton, Bernal, Price e Garfield (Elkana et al., 1978). Finalmente, reflete a mudança do suporte de divulgação das informações do suporte de papel para o suporte eletrônico. A internet associada à expansão das técnicas estatísticas possibilitou a manipulação de bases de dados enormes; as consequências foram a sofisticação das técnicas de medida, monitoramento de vários aspectos do processo de comunicação científica em geral e da avaliação científica em particular, em nível macro (países), meso (disciplinas) e micro (programas).

A profusão dos índices de avaliação científica

Ao longo das últimas duas décadas muitos esforços têm sido feitos para criação de novos indicadores. Por um lado, esse processo está associado ao aumento da demanda de informações por parte de pesquisadores, agências de fomento e avaliação, editores e revistas; por outro, a disponibilidade de grandes bancos de dados internacionais e a ampliação da capacidade de manipulação estatística, sociológica e informática. O resultado foi a multiplicação de indicadores como, por exemplo, indicadores de atividade baseados em publicações indexadas; indicadores de especialização temática; indicadores de visibilidade e difusão; indicadores de colaboração institucional e internacional; indicadores de uso e recomendação. Esses indicadores variam segundo o tipo de contagem, a natureza do cálculo e o padrão de medida. Segundo Callon, Courtial e Penan (1995), podem ser divididos em dois grupos principais: (a) indicadores de produtividade; (b) indicadores relacionais.

Os indicadores de produtividade partem da premissa de que a ciência e a tecnologia são atividades produtivas que, portanto, podem ser medidas e compreendidas em termos de seus *inputs* e *outputs*: de um lado, entrariam os recursos financeiros, os materiais e a mão de obra; de outro, sairiam os produtos da atividade científica, isto é, os artigos, patentes, instrumentos e profissionais formados. Nesse sentido, a tarefa de avaliação da atividade científica deve ser a medição do volume de produção e seu impacto em dada área do conhecimento, de forma a determinar seu dinamismo e evolução, assim como a contribuição e produtividade individual dos pesquisadores. Trata-se de uma forma de avaliação intrinsecamente numérica, estatística, que descreve e analisa a atividade científica em termos instrumentais, de rendimento e impacto. Essas medições podem ser divididas em dois tipos: a contagem de publicações e a contagem de citações.

A contagem de publicações é a forma mais simples de indicador de produção científica (Callon, Courtial, Penan, 1995). Baseia-se no princípio de que a atividade de pesquisadores ou de grupos dentro de um campo do conhecimento, especialidade ou divisão geográfica pode ser medida por meio da identificação e contagem dos artigos publicados em revistas acadêmicas. Do ponto de vista individual, dos pesquisadores, possibilita analisar a evolução quantitativa de sua produção e a comparação dos currículos de diferentes profissionais. Portanto, a contagem de publicações é o ponto de partida e o mecanismo de verificação de certas proposições sobre a capacidade e contribuição dos autores como, por exemplo, a Lei de Lotka. Quando analisados de forma agregada, esses indicadores também ajudam a medir a taxa de produtividade em torno de uma disciplina e a participação de uma instituição ou país no conjunto da produção científica, em um dado período de tempo.

Por um lado, essa forma de avaliação procura racionalizar a aplicação de recursos e a elaboração de políticas públicas para a atividade científica, em um contexto cada vez mais competitivo (Velho, 1985; Leta, 2011). Seu uso foi sistematizado e disseminado por meio de iniciativas como o *Manual di Frascati*, publicado originalmente em 1963, que procurava estabelecer formas padronizadas para a avaliação científica. Por outro lado, no entanto, se baseia em uma imagem reducionista da atividade científica: o foco em um *output* específico – o conhecimento escrito, representado pelo artigo científico – torna invisíveis diversos outros resultados e produtos dessa atividade; mais do que isso, naturaliza a imagem da ciência como uma atividade cumulativa, submetida às leis estatísticas. Nesse aspecto, a contagem do número de publicações de um autor, instituição ou país pode se apresentar como autoevidente, reificada, dispensando problematização ou contextualização.

Outro indicador de atividade muito utilizado para avaliação científica constitui a contagem de citações. A ideia básica da contagem de citações é simples: os trabalhos e cientistas mais influentes são os mais citados, e a contagem de citações é um indicador de recepção (Glänzel, 2008). Assim, as conexões existentes entre os documentos permitem avaliar o quanto um artigo é útil para outro pesquisador e dessa forma estimar sua importância. Afinal, existe uma correlação entre altos índices de citação e os juízos de pares sobre a excelência científica das contribuições (Garfield, 1979). Isso significa que, quanto maior o número de citações, mais importantes o trabalho e o cientista. Nesse sentido, quando um artigo é citado é possível estabelecer duas hipóteses sobre sua importância: (a) o artigo é suficientemente visível para servir de referência; (b) produz um impacto que só pode ser considerado pelo documento em si mesmo (Callon, Courtial, Penan, 1995). Contar citações permite avaliar grau de utilização e, portanto, determinar o impacto de artigos e revistas.

A operacionalização da técnica de contagem de citações é simples. A base da técnica são os elementos formais do texto: autores, instituição, título, periódico, local, número de páginas, data, referências bibliográficas etc. Como quase todos os documentos científicos possuem referências, pega-se o total de artigos, comunicações, cartas, relatórios etc. publicados nas revistas científicas e cruza-se com as notas de rodapé e as referências bibliográficas. Forma-se um banco de dados que permite a contagem de todas as referências da revista J no ano Y, resultando num fator de impacto (Garfield, 1972). A contínua sofisticação dessas técnicas possibilitou a criação de *softwares* como o Publish or Perish (PoP) pela Microsoft Academic Search em 2006: além de estatísticas simples (número de documentos, número de citações

etc.), o *software* PoP calcula métricas de citação individual como o *h-index* proposto por Jorge E. Hirsch e o *g-index* proposto por Leo Egghe (Harzing, 2011).

A adoção de indicadores de produtividade para avaliação científica tem sido objeto de muitas controvérsias (Cozzens, 1981; Leydesdorff, 2001). Por isso, não existe consenso sobre adequação de sua aplicação. Afinal, enquanto um grupo de pesquisadores está preocupado com o que contagens de citações medem, outro grupo está preocupado com o que elas não medem. Assim, se, por um lado, indicadores de produtividade são vistos como um poderoso instrumento para mapear o impacto intelectual de cientistas, revistas, disciplinas, programas, por outro, a validade desses dados é questionada pela adequação das bases de dados para a avaliação da pesquisa. Por exemplo, muitas vezes o trabalho num pequeno campo e não publicado em inglês, ou em livros, não é capturado, como indicam as ciências sociais. Dito de outra forma, indicadores de produtividade não identificam e até subestimam quem não é reconhecido pela comunidade científica.

Os indicadores relacionais foram concebidos para detectar as ligações entre elementos bibliográficos. Eles descrevem o grau de semelhança/diferença existente entre documentos, autores, revistas, conceitos e permitem medir a intensidade da relação. Isso significa que é possível posicionar e agregar as interações entre esses elementos. As referências bibliográficas não são mais consideradas entidades isoladas e desconectadas, mas um todo, com base em regras de relação intrínsecas como *co-word analysis* e *softwares* como o T-LAB. Os indicadores relacionais se diferenciam em função da análise do conteúdo: (a) indicadores relacionais de primeira geração, que não penetram no conteúdo como análise: (i) coautoria; (ii) redes de citações; e (b) indicadores relacionais de segunda geração, que analisam o conteúdo dos documentos: (i) co-ocorrência de termos; (ii) coclassificação de publicações; (iii) cocitações (Callon, Courtial, Penan, 1995).

O principal indicador relacional de primeira geração é a análise de coautoria. A coautoria é utilizada para determinar a cooperação entre instituições ou grupo de pesquisadores. A ideia de base é que o número de artigos produzidos conjuntamente exprime a cooperação nas atividades de pesquisa. Isso significa que a coautoria indica a profissionalização e a especialização da comunidade de autores em termos de colaboração e financiamento. Portanto, quanto maior a quantidade de artigos firmados conjuntamente, maior a dotação econômica que esses autores recebem para suas investigações. Entre os principais aspectos pesquisados destacam-se quantidade de autores, sua hierarquia e os autores principais, assim como a questão da heterogeneidade entre pesquisadores universitários e da indústria. Portanto, a análise de coautoria constitui redes de cooperação entre pesquisadores e possibilita entender a dinâmica da comunidade científica.

Para além do foco no binômio autoria-citação, existem indicadores relacionais que buscam descrever a dinâmica e a densidade das contribuições em torno de um tema ou especialidade. Um primeiro exemplo é a análise da co-ocorrência de termos. Trata-se de uma técnica estatística que analisa pares de palavras ou trechos de textos, de forma a identificar padrões de recorrência que indiquem a associação entre conceitos em um determinado conjunto de textos. Também conhecida como Leximappe, foi desenvolvida pelo Centre de Sociologie de l'Innovation, da École Nationale Supérieure des Mines de Paris, e o Centre National de La Recherche Scientifique (CNRS) nos anos 1980 (He, 1999). Seu produto é a representação

de uma rede e das ligações entre conceitos, problemas e ideias. Em resumo, busca criar indicadores relacionais que possibilitem mapear e compreender a evolução da ciência e da tecnologia. Portanto, ainda que o método seja estatístico, sua finalidade é intrinsecamente *policy-oriented* (Courtial, Law, 1989).

Em sentido semelhante, é possível analisar a relação entre as revistas científicas. O ponto de partida é a contagem de citações de uma revista para outra, sugerida inicialmente por Katherine W. McCain (1990). Esse índice de “relacionalidade” (*relatedness*) possibilita evidenciar, por um lado, quais publicações estão relacionadas com os mesmos temas, dentro de um campo ou área de especialidade. Além disso, implica a possibilidade de construir uma cartografia ou rede na qual a quantidade de citações cruzadas denota a intensidade da ligação entre as revistas e, indiretamente, a densidade de uma disciplina. Tal postura analítica procura complementar a análise de caráter basicamente quantitativo como, por exemplo, a classificação pelo total de artigos publicados ou o fator de impacto (calculado a partir da razão entre as citações recentes de artigos publicados em uma revista e o total de artigos publicados em um dado período de tempo).

Outro indicador relacional de segunda geração constitui a análise de cocitações. A contagem de cocitações compreende a técnica de acoplamento de pares de dados ou autores cocitados com base em metodologias estatísticas como a análise de *clusters*, escalas multidimensionais e análise fatorial. Foi concebida por Howard White e Bewar Griffith em 1981 e parte do pressuposto de que quando duas citações são colocadas no mesmo texto elas apresentam uma relação que indica proximidade em termos de conteúdo. Isso aponta que um número de referências comuns entre dois ou mais textos constitui um indicador de proximidade cognitiva. A consideração das cocitações permite a formação de classes de área e, portanto, conceber *clusters* (Andrés, 2009). A análise de cocitações tem sido estendida a vários níveis de agregação, como revistas, autores e assuntos, para assim examinar e traçar a estrutura de comunidade de pesquisadores e de disciplinas.

Em síntese, as diferenças e semelhanças entre os indicadores de produtividade e os indicadores relacionais assinalam que a avaliação científica tem muitos usos. Por um lado, a passagem dos indicadores de produtividade para os relacionais reflete o progressivo deslocamento do interesse dos elementos inter- para os elementos intrabibliográficos, assinalando o desenvolvimento de metodologias que visam capturar a tendência de indexação automática; por outro, reflete também a entrada de novos atores e o deslocamento do debate dos EUA para a Europa, o que indica o aparecimento de novas necessidades. O efeito combinado dessas duas transformações foi não somente a proliferação de indicadores, mas também a substituição dos métodos deliberativos por métodos indicadores quantitativos (De Bellis, 2014; Cronin, Sugimoto, 2014). Em termos práticos, isso significa que se tornou mais difícil escolher os instrumentos de medida apropriados para a atividade científica.

Aplicabilidade dos indicadores na avaliação da atividade científica

Indicadores quantitativos não são apenas produtos científicos (Van Raan, 2004), mas constituem também instrumentos para avaliação, regulação e formulação de políticas (Narin,

1976). Começaram a ser utilizados para fins de avaliação de produtividade desde o final do século XIX (Godin, 2009) e estão frequentemente associados ao financiamento da pesquisa. Atualmente, entre os principais instrumentos para a avaliação da capacidade científica de pessoas ou instituições estão os indicadores bibliométricos (Hicks et al., 2015). Um dos principais indicadores é o *impact factor*, proposto por Eugene Garfield em 1952 para determinar a qualidade de periódicos científicos, mas também outros indicadores menos conhecidos como, por exemplo, o *immediacy index*, o *prestige factor*, o *usable factor* etc. Diante disso, nesta seção consideramos a aplicação de indicadores para avaliação de três unidades de análise da atividade científica: (a) micro (pesquisadores); (b) meso (periódicos); (c) macro (organizações).

Provavelmente, a avaliação de pesquisadores constitui o tema que mais tem chamado atenção da comunidade científica na última década. Por um lado, esse processo está associado à massificação da atividade científica; por outro, aos limites da avaliação baseada no *peer review*. Por isso, a maior preocupação dos pesquisadores se concentra no aumento da quantidade de publicações e no número de citações de suas publicações. Nesse sentido, existem duas formas de efetuar o cálculo: (a) manual: feito por meio da análise individual dos trabalhos; (b) automática: utilizam-se as informações de bancos de dados como, por exemplo, Publish or Perish, que permitem avaliar os membros de uma reunião, do conselho editorial de uma revista, participantes de um evento científico etc. Para exemplificar esse processo podemos considerar em especial: (a) o *h-index*; (b) o *g-index*.

O *h-index* foi proposto por Jorge E. Hirsch em 2005 para avaliar a produtividade de um cientista. A ideia por trás desse índice combina a medida da quantidade (publicações) com a medida do impacto (citações). Em suas palavras: “um cientista tem índice *h* se *h* de seus textos N_p tiver pelo menos *h* citações cada e os outros $(N_p - h)$ textos tiverem $\leq h$ citações cada” (Hirsch, 2005, p.16569). Portanto, se um cientista possui um *h-Index* igual a cinquenta, é porque escreveu cinquenta artigos que têm pelo menos cinquenta citações cada. Se o *h-index* constitui um indicador de reconhecimento da comunidade científica, a carreira acadêmica consistiria em ir obtendo progressivamente reconhecimento para superar os limiares *h* (Grupo Scimago, 2006). Contudo, distorce a avaliação individual porque não permite a comparação entre disciplinas (número de referências e quantidade de artigos) e, sobretudo, penaliza os cientistas que publicam seletivamente (Costas, Bordons, 2007).

O *g-index* criado por Leo Egghe, em 2006, se apresenta como uma proposta de aperfeiçoamento da lógica aplicada para o *h-index*. Dado um conjunto de artigos, classificados em ordem decrescente de número de citações que eles receberam, o *g-index* é o maior valor *g* em que os primeiros *g* artigos receberam juntos ao menos g^2 citações (Egghe, 2006). Isso significa, por exemplo, que um pesquisador que possua quatro publicações cujos números de citações sejam, em ordem decrescente: cinco, três, um, um, possuirá um *g-index* igual a três. Nesse sentido, o *g-index* torna mais visível a diferença de impacto entre os autores, ao mesmo tempo em que evidencia a importância de seus principais artigos. Portanto, um *g-index* maior representaria mais e melhores artigos (Tol, 2008). Todavia, assim como o *h-index*, trata-se de um índice representado em números inteiros, isto é, vários autores podem ter o mesmo score, dificultando a diferenciação entre eles (Huang, Chi, 2008).

Analisados em conjunto, os índices *h* e *g* se apresentam como indicadores quantitativos que buscam diferenciar os autores científicos, em termos do impacto de suas contribuições.

Representam formas relativamente recentes de avaliação da produção dos pesquisadores, ao lado de outras formas de medição, como o *a-index*, o *4-index* ou o *r-index*, que ganham crescente visibilidade como mecanismos para a classificação de profissionais em um campo, instituição ou departamento (Schreiber, 2008; Selek, Saleh, 2014). Procuram prover alternativas aos métodos de avaliação baseados nas contagens de produção bruta, que podem gerar avaliações sujeitas às distorções do produtivismo, tais como os pesquisadores *freeriders* ou os *papers* de baixíssima qualidade. No entanto, ainda não conseguem superar distinções geradas pela posição do pesquisador, seu acesso a recursos e redes de colaboração, e até mesmo a forma como os pesquisadores usam as medições de produtividade a seu favor.

A aplicação das métricas em periódicos científicos estabelece uma equivalência entre avaliação e classificação. Se em cada disciplina sempre existiu uma hierarquia tácita de revistas, com as medidas quantitativas, esse processo se institucionaliza (Gingras, 2009). Por exemplo, o *impact factor* incide sobre a notoriedade das revistas à medida que atrai mais submissões e melhores artigos. Portanto, a avaliação converte-se numa instância de atribuição de prestígio que afeta a estrutura da comunicação científica. Pode-se ilustrar isso recorrendo a várias classificações como, por exemplo, a Listes de Revues en Sciences Humaines et Sociales da Agence d'Évaluation de la Recherche et de l'Enseignement Supérieur (Aeres) da França, ou a classificação Excellence in Research for Australia do Australian Research Council (ARC). Vamos, porém, considerar apenas três modalidades: (a) o *Journal Citation Reports* (JCR); (b) o European Reference Index; (c) e o Qualis Capes.

O *Journal Citation Reports* (JCR) é um produto bibliométrico do grupo Thomson Reuters, que oferece formas para “avaliar criticamente as principais revistas científicas” (Thomson Reuters, 2015, s.p.). Sua origem está vinculada ao Science Citation Index proposto originalmente pelo ISI, mas sua lógica é distinta: se inicialmente o foco estava nos autores, no segundo a chave para a organização e classificação dos dados passou a ser uma lista de revistas e sua produtividade (Garfield, 2007). Da análise dos dados compilados no JCR surgiram o *impact factor*, o *immediacy index* (o número de artigos de uma revista citados no mesmo ano de sua publicação) e outros indicadores similares. Assim, a ferramenta buscava oferecer critérios para julgar a relevância de uma publicação, especialmente para orientar a pesquisa bibliográfica (Pendlebury, Adams, 2012). Na prática, porém, vem gradualmente promovendo diferenciação e concentração de importância entre as publicações.

O European Reference Index for the Humanities (ou Erih Plus) constitui um índice de revistas na área de ciências humanas. Foi estabelecido pela European Science Foundation (ESF), em 2005, e transferido para a Norwegian Social Science Data Service (NSD) em 2014. Todas as revistas incluídas no Erih Plus são classificadas segundo três categorias, de acordo com o nível de escopo e o público, com base na realização de painéis *experts* e consultas interativas: A (nível de publicação internacional com forte reputação entre os pesquisadores); B (nível de publicação internacional com boa reputação entre os pesquisadores); C (nível de reputação regional, com importância local). A diferença entre a concepção e a execução indica que a classificação se converteu num processo de alocação de qualidade que atribui um valor superior a quem publica na categoria A (Editorial, 2009).

O Qualis Capes constitui um conjunto de procedimentos de estratificação da produção científica brasileira. Foi concebido para avaliar e disponibilizar uma lista com uma classificação

para divulgação da produção da pós-graduação. A qualidade das publicações é aferida de forma indireta, a partir da análise da qualidade dos periódicos científicos. Como indica a Capes (2014, s.p.): “A classificação de periódicos é realizada pelas áreas de avaliação e passa por processo anual de atualização. Esses veículos são enquadrados em estratos indicativos da qualidade – A1, o mais elevado; A2; B1; B2; B3; B4; B5; C – com peso zero.” Essa classificação estratifica os periódicos em: A1 e A2 (excelência internacional), B1 e B2 (excelência nacional), B3, B4 e B5 (relevância média), C (baixa relevância), com base no *impact factor* (Ferreira, Antoneli, Briones, 2013), opondo-se às recomendações internacionais como a San Francisco Declaration on Research Assessment. Em termos práticos, esse sistema de avaliação e classificação afeta as decisões de publicações dos pesquisadores, as rotinas editoriais e de processos de qualidade das revistas (Frigeri, Monteiro, 2014) e até mesmo suas possibilidades de captação de financiamentos oficiais (Silva, 2009).

A avaliação de laboratórios ou programas de pesquisa com ferramentas bibliométricas constitui uma operação controversa. Embora a produtividade científica de uma organização científica possa ser concebida como o produto de investimentos, não existe um protocolo de como isso deve ser feito (Okrasa, 1987). Mesmo assim, um conjunto de iniciativas vem sendo desenvolvido para avaliar e, sobretudo, justificar a alocação de recursos como o *Manual di Frascati* (OCDE, 2007). Isso tem desencadeado conflitos, uma vez que não existe um modelo linear (administrativa, econômica ou bibliometricamente) (Godin, 2009) que permita comparação de organizações com perfis institucionais, administrativos e financeiros muito distintos. Apesar de não existir um algoritmo capaz de avaliar o desempenho de um corpo de pesquisadores e que sirva como um padrão indicativo da produtividade científica, pode-se considerar o caso de programas de pesquisa e universidades.

Desde o estudo seminal publicado por Martin e Irvine (1983), a avaliação de grupos de pesquisa (programas, laboratórios, faculdades etc.) fundamenta-se na influência internacional das publicações. Isso significa que os cientistas que têm algo importante para dizer empenham-se vigorosamente para publicar suas descobertas em periódicos internacionais (Vinkle, 2010). A extensão desse pressuposto para o estudo de grupos de pesquisa pode ser ilustrada por meio da consideração das práticas de publicação da Leiden School (De Bellis, 2014). Essa abordagem está relacionada ao Centre for Science and Technology Studies da Leiden University e para isso pode-se tomar o exemplo do CPP/FCSm Indicator, que relaciona as citações por publicação e a média do escore no campo da citação (Van Raan, 2004). Esse indicador permite comparar as citações de todos os artigos publicados em todas as revistas com a média mundial e, portanto, estabelecer a produtividade do instituto.

Os *rankings* de universidades se apresentam como formas de avaliar a qualidade e relevância do ensino e pesquisa universitária. Atualmente, dois *rankings* rivais dominam o cenário mundial: o chinês Shanghai Jiao Tong University Academic Ranking of World Universities, criado em 2003, e o inglês Times Higher Education World University Rankings, criado em 2004. Em ambos os casos, a classificação é obtida por meio da avaliação de indicadores como a qualidade da educação e do corpo funcional, citações em revistas e índices reconhecidos, *performance per capita*, internacionalização e atração de recursos, considerando diferentes pesos para o cálculo. Uma racionalidade semelhante está presente no Ranking Universitário

Folha que, desde 2012, analisa as universidades brasileiras considerando cinco indicadores: pesquisa, internacionalização, inovação, ensino e mercado.

Em um contexto de massificação do ensino superior e pesquisa científica, e de aumento da competição entre as instituições, os *rankings* de grupos de pesquisa, institutos e universidades servem para orientar a alocação de recursos, prestígio pessoal e regular a competição entre instituições pelos alunos e pesquisadores mais brilhantes e promissores (Altbach, 2006). Os problemas dessa forma de avaliação não estão em seus princípios, mas na prática: atualmente, diferentes *rankings* costumam apresentar resultados consistentes para as instituições mais bem colocadas, mas pouca consistência para aquelas que estão mais abaixo na classificação (Saisana, D'Hombres, 2008; Usher, Savino, 2007). Portanto, essa forma de avaliação parece certificar as instituições que já possuem prestígio, uma orientação científica clássica e maior acesso a recursos, ao mesmo tempo em que avaliam negativamente perfis institucionais periféricos, seja do ponto de vista geográfico ou disciplinar.

A consideração da aplicação dos indicadores na avaliação da produtividade científica revela que existe mais informação sobre a produção científica do que sobre a sua aplicação. Por isso, pode-se dizer que, atualmente, um pesquisador produtivo não é apenas um cientista que publica, mas um pesquisador que publica em determinadas revistas e uma quantidade específica de artigos por ano; um periódico com credibilidade não é apenas aquele que é reconhecido pelos membros de uma disciplina científica, mas aquele que apresenta uma preferência ranqueada em sistemas de classificação. O aumento da produtividade de organização científica vem sempre acompanhado pelo distanciamento da comunidade regional. O que deve ser construído a partir dos dados não é um número, mas um padrão de elementos significativos que permita perceber o processo de mudança. Portanto, pode-se dizer que a assimetria entre a quantidade de métricas e sua aplicação indica que as práticas são adaptadas aos critérios (Gingras, 2014).

Subsídios para uma nova agenda na avaliação científica

A metrificação da avaliação científica constitui o efeito combinado da integração progressiva da estatística, da sociologia e da informática. Reflete o aperfeiçoamento das ferramentas estatísticas, do padrão de organização da comunidade científica e do suporte tecnológico. Está associada às contribuições pioneiras de Lotka, Bradford e Zipf, mas também de Merton, Bernal e Price, além de Garfield, e às necessidades de gestão e controle da atividade científica. Esse processo de quantificação da informação, massificação da comunicação científica e formação de bancos de dados orienta a produção científica e redefine o seu significado. Portanto, o processo de metrificação surge num contexto muito específico, mas acaba se disseminando por toda a comunidade científica. Mas o que o exame do processo de metrificação revela é que existem mais informações sobre a produção de indicadores do que sobre a sua aplicação.

A metrificação encerra um paradoxo intrigante. Quanto mais a avaliação da produtividade científica se aperfeiçoa tecnicamente, menor a confiança dos cientistas nessas ferramentas. Mais precisamente, quanto maior a objetividade, menor a credibilidade. A chave para a interpretação desse paradoxo é que, na metrificação, a objetividade não deriva do conhecimento adquirido ao longo da carreira, mas da aplicação de regras desconhecidas pela comunidade

científica. Nesse sentido, a profusão de métricas indica que cada grupo acaba desenvolvendo seus próprios parâmetros e indicadores para justificar suas práticas científicas. Por isso, na maior parte dos casos, as métricas são aceitas quando são convenientes e descartadas quando são desfavoráveis. O apelo da metrificação à imparcialidade dos números padroniza competências locais em regras gerais: transforma um padrão de comunicação científica em parâmetro de avaliação para toda a produção científica.

As controvérsias sobre esse paradoxo podem ser observadas por diversos tipos de reações institucionais. Por um lado, do ponto de vista metodológico, verifica-se um contínuo processo de aperfeiçoamento e multiplicação das técnicas de avaliação, indicando o caráter contingente dessas ferramentas. Por outro, do ponto de vista político, verifica-se também a resistência de certas áreas de conhecimento à incorporação desses instrumentos, como demonstra o caso das ciências humanas, em geral, e das ciências sociais em particular. Nesse sentido, o surgimento de manifestos como o Force11 Manifesto: Building the Future for Research Communications and e-Scholarship (2011), o San Francisco Declaration on Research Assessment (2012) e o Leiden Manifesto (2015) e a discussão em torno deles constituem apenas o ponto mais evidente de um mal-estar generalizado na comunidade científica. Portanto, o paradoxo parece se converter em um conflito de interesses entre o caráter público e privado da ciência.

Além disso, a estratégia da quantificação da produtividade individual e da avaliação de dados agregados (ou “grandes números”) tem gerado inúmeras distorções na atividade científica, do ponto de vista normativo. Em outras palavras, a ênfase passa a ser o *output* de um pesquisador, grupo ou instituição, em detrimento de valores tradicionalmente associados com a comunidade científica, como autonomia intelectual e independência política. Assim, entre as consequências indesejadas das métricas da ciência, se verificam casos frequentes de plágio (e, portanto, de retratação de artigos), autocitação, publicações redundantes, atribuição de autoria indevida e pesquisadores *freeriders*. Em resumo, a centralidade da avaliação em torno dos artigos e de medidas de produtividade, somada aos problemas de atribuição de recursos e prestígio, tem gerado anomalias na comunicação entre os cientistas e para a definição de políticas de ciência e tecnologia.

Portanto, o processo de metrificação da avaliação da produtividade constitui a história de como o artigo científico se converteu em expressão da atividade científica. Nesse sentido, uma nova agenda de avaliação deve contornar três obstáculos estabelecidos pela metrificação:

- (a) papirocentrismo: o artigo científico não pode ser considerado o centro da avaliação científica;
- (b) produtivismo: um bom pesquisador não é apenas aquele que é um bom pontuador nos *rankings* existentes;
- (c) mimetismo: reconhecimento internacional não pode ser considerado parâmetro para certificação do conhecimento.

Desse modo, uma nova agenda de avaliação pressupõe uma postura “antirreducionista”: não reduzir a avaliação científica à análise da literatura científica; não reduzir a utilidade de um artigo científico à sua visibilidade na comunidade científica; não reduzir a excelência da produção científica à similaridade internacional.

Tal agenda pode ser articulada por meio da investigação das causas e impactos de novos fenômenos relacionados ao cotidiano da atividade científica. Assim, por exemplo, a investigação pode ser direcionada não apenas para as mudanças na noção de autoria e crédito entre os cientistas (por meio da divisão de autoria entre grandes grupos de pesquisa), mas também em relação à *crowd* ou *networked science*, em que profissionais e amadores trabalham em torno de um mesmo problema, muitas vezes coordenados por plataformas computacionais complexas. Além disso, é possível pesquisar casos como as organizações de pacientes e familiares que se tornam “especialistas leigos”, colaborando com médicos e cientistas para descobertas de novos tratamentos, redefinindo as fronteiras da produção científica e o papel do artigo como método de divulgação.

Portanto, a questão da avaliação da atividade científica deve ser reavaliada por meio de duas estratégias: por um lado, é necessário compreender a formação histórica das formas de avaliação científica, o processo de “metrificação” da ciência, suas implicações epistemológicas e políticas e, em particular, suas atuais limitações. Por outro lado, é imperativo compreender as novas modalidades de produção de conhecimentos científicos, para propor novas formas de avaliação; nesse aspecto, se as métricas e os indicadores existentes já demonstram limitações para o acompanhamento da ciência “normal”, é óbvio que serão ainda mais restritos em sua capacidade avaliativa nos casos em que a atribuição de autoria e as formas de divulgação forem radicalmente novas ou heterodoxas. Torna-se necessário, afinal, redefinir o problema, construindo novas formas e novos sentidos para as métricas de avaliação científica, superando a atual obsessão produtivista na ciência.

A quebra do monopólio cognitivo deve progressivamente modificar a avaliação da produtividade científica. Por muito tempo, o controle da qualidade científica ficou restrito a avaliação *ex-ante*: primeiro o conhecimento é certificado na comunidade científica por meio da avaliação por pares e depois transborda para a sociedade. Porém, com o surgimento da internet e a perda do monopólio dos cientistas na produção do conhecimento, novas experiências de avaliação começam a ser experimentadas como, por exemplo, a avaliação pós-publicação. Nesse sentido, seguindo a tendência de experiências como a Wikipédia e outras formas de hipertexto, o conhecimento é avaliado *ex-post*: a revisão por pares pós-publicação, ativado por *sites* como o PubPeer. Isso significa que o conhecimento é tornado público e aberto para todo tipo de contribuição. Trata-se, portanto, da associação da avaliação interna da comunidade científica com a avaliação externa da sociedade.

NOTA

¹ Nessa e nas demais citações de textos em outros idiomas, a tradução é livre.

REFERÊNCIAS

ACADÉMIE DES SCIENCES.
Du bon usage de la bibliometrie pour l'évaluation individuelle des chercheurs: rapport remis le 17 janvier 2011 à Madame la Ministre de l'Enseignement Supérieur et de la Recherche. Disponível em: http://www.academie-sciences.fr/archivage_site/active/rapport/avis170111_synthese.pdf. Acesso em: 1 dez. 2016. 2011.

ADAIR, William.
Citation indexes for scientific literature?
American Documentation, v.6, n.1, p.31-32. 1955.

ALBERTS, Bruce.
Impact factor distortions, *Science*, v.340, n.6134, p.787. 2013.

- ALTBACH, Philip G.
International higher education: reflections on policy and practice. Chestnut Hill: Boston College Center for International Higher Education. 2006.
- ANDRÉS, Ana.
Measuring academic research: how to undertake a bibliometric study. Oxford, Cambridge, Nova Déli: Chandos Publishing. 2009.
- BERNAL, John D.
The social function of science. Cambridge: The MIT Press. 1967.
- BORNMANN, Lutz.
Scientific peer review: an analysis of the peer review process from the perspective of sociology of science theories. *Human Architecture: Journal of the Sociology of Self-knowledge*, v.6, n.2. p.23-38. 2008.
- BENSMAN, Stephen J.
Garfield and the impact factor. *Annual Review of Information Science and Technology*, v.41, n.1, p.93-155. 2007.
- CALLON, Michel; COURTIAL, Jean-Pierre; PENAN, Hervé.
Cienciometria: el estudio cuantitativo de la actividad científica: de La bibliometría a la vigilancia tecnológica. Oviedo: Trea. 1995.
- CAPES.
Coordenação de Aperfeiçoamento de Pessoal de Nível Superior. *Classificação da produção intelectual*. Disponível em: <http://www.capes.gov.br/avaliacao/instrumentos-de-apoio/classificacao-da-producao-intelectual>. Acesso em: 1 dez. 2016. 2014.
- CASTIEL, Luis David; SANZ-VALERO, Javier; RED MEI-CYTED.
Entre fetichismo e sobrevivência: o artigo científico é uma mercadoria acadêmica? *Cadernos de Saúde Pública*, v.23, n.12, p.3041-3050. 2007.
- COILE, Russell C.
Lotka's frequency distribution of scientific productivity. *Journal of the American Society for Information Science*, v.28, n.6, p.366-370. 1977.
- CONSTANS, Carine; RIVOAL, Isabelle.
Le prestige des revues scientifiques et la logiques de classement. In: Hurllet, Frédéric; Rivoal, Isabelle; Sidéra, Isabelle. *Le prestige: autour des forms de la différenciation sociale*. Paris: Bocard. p. 255-270. 2014.
- COSTAS, Rodrigo; BORDONS, María.
The h-index: advantages, limitations and its relation with other bibliometric indicators at the micro level. *Journal of Informetrics*, v.1, n.3, p.193-203. 2007.
- COURTIAL, Jean-Pierre; LAW, John.
A co-word study of artificial intelligence. *Social Studies of Science*, v.19, n.2, p.301-311. 1989.
- COZZENS, Susan.
Taking the measure of science: a review of citation theories. *Newsletter of the International Society for the Sociology of Knowledge*, v.7, n.1, p.16-21. 1981.
- CRONIN, Blaise.
Scholars and scripts, spoors and scores. In: Cronin, Blaise; Sugimoto, Cassidy. *Beyond bibliometrics: harnessing multidimensional indicators of scholarly impact*. Cambridge: The MIT Press. p.3-21. 2014.
- CRONIN, Blaise; SUGIMOTO, Cassidy.
Beyond bibliometrics: harnessing multidimensional indicators of scholarly impact. Cambridge: The MIT Press. 2014.
- DE BELLIS, Nicola.
History and evolution of (biblio)metrics. In: Cronin, Blaise; Sugimoto, Cassidy. *Beyond bibliometrics: harnessing multidimensional indicators of scholarly impact*. Cambridge: The MIT Press. p.23-44. 2014.
- DESROSIÈRES, Alain.
The politics of large numbers: a history of statistical reasoning. Cambridge: Harvard University Press. 1998.
- DONAVAN, Claire.
The qualitative future of research evaluation. *Science and Public Policy*, v.34, n.8, p.585-597. 2007.
- EDITORIAL.
Editorial. *Medical History*, v.53, n.1, p.1-4. 2009.
- EGGHE, Leo.
Theory and practice of the g-index. *Scientometrics*, v.69, n.1, p.131-152. 2006.
- ELKANA, Yehuda et al. (Ed.).
Toward a metric of science: the advent of science indicators. Brisbane: John Wiley. 1978.
- FERREIRA, Renata C.; ANTONELLI, Fernando; BRIONES, Marcelo R.S.
The hidden factors in impact factors: a perspective from Brazilian science. *Frontiers Genetics*, v.4, n.130, p.1-2. 2013.
- FRIGERI, Mônica. MONTEIRO, Marko Synésio Alves.
Qualis Periódicos: indicador da política científica no Brasil? *Estudos de Sociologia*, v.19, n.37, p. 299-315. 2014.

- FURNER, Jonathan.
The ethics of evaluative bibliometrics. In: Cronin, Blaise; Sugimoto, Cassidy. *Beyond bibliometrics: harnessing multidimensional indicators of scholarly impact*. Cambridge: The MIT Press. p.85-108. 2014.
- GARFIELD, Eugene.
The evolution of the Science Citation Index. *International Microbiology*, v.10, n.1, p.65-69. 2007.
- GARFIELD, Eugene.
Is citation analysis a legitimate evaluation tool? *Scientometrics*, v.1, n.4, p.359-375. 1979.
- GARFIELD, Eugene.
Citation analysis as a tool in journal evaluation. *Science*, v.178, n.4060, p.471-479. 1972.
- GARFIELD, Eugene.
Citation indexes for science: a new dimension in documentation through association of ideas. *Science*, v.122, n.3159, p.108-111. 1955.
- GINGRAS, Yves.
Les derives de l'évaluation de la recherche: du bon usage de la bibliométrie. Paris: Raisons d'Agir. 2014.
- GINGRAS, Yves.
Les systèmes d'évaluation de la recherche. *Sciences de l'Information*, v.46, n.4, p.34-35. 2009.
- GLÄNZEL, Wolfgang.
Seven myths in bibliometrics about facts and fiction in quantitative science studies. *Issi Newsletter*, v.4, n.2, p.24-32. 2008.
- GODIN, Benoît.
The making of science, technology and innovation policy: conceptual frameworks as narratives, 1945-2005. Quebec: Centre Urbanisation Culture Société. 2009.
- GRUPO SCIMAGO.
El índice h de Hirsch: aportaciones a un debate. *El Profesional de la Información*, v.15, n.4, p.304-306. 2006.
- HAGSTRON, Warren O.
The scientific community. New York: Basic Books. 1965.
- HARZING, Anne-Wil.
The Publish or Perish book. Melbourne: Tama Software Research. 2011.
- HE, Qin.
Knowledge discovery through co-word analysis. *Library Trends*, v.48, n.1, p.133-159. 1999.
- HICKS, Diana et al.
The Leiden Manifesto for research metrics. *Nature*, v.520, p.429-431. 2015.
- HIRSCH, Jorge E.
A index to quantify an individual's scientific research output. *Proceedings of the National Academy of Science*, v.102, n.46, p.16569-16572. 2005.
- HUANG, Mu-hsuan; CHI, Pei-shan Chi.
A comparative analysis of the application of h-index, g-index, and a-index in institutional-level research evaluation. *Journal of Library and Information Studies*, v.8, n.2, p.1-10. 2008.
- KUHN, Thomas.
A estrutura das revoluções científicas. São Paulo: Perspectiva. 1989.
- LETA, Jacqueline.
Indicadores de desempenho, ciência brasileira e a cobertura das bases informacionais. *Revista USP*, n.89, p.62-67. 2011.
- LEYDESDORFF, Loet.
The challenge of scientometric: the development, measurement, and self-organization of scientific communications. New York: Universal. 2001.
- MANCHIKANTI, Laxmaiah et al.
Medical journal peer review: process and bias. *Pain Physician*, v.18, n.1, p.E1-E14. 2015.
- MARTIN, Ben R.; IRVINE, John.
Assessing basic research: some partial indicators of scientific progress in radio astronomy. *Research Policy*, v.12, n.2, p.61-90. 1983.
- MATTEDI, Marcos A.
Sociologia e conhecimento: introdução à abordagem sociológica do problema do conhecimento. Chapecó: Argos. 2006.
- MATTEDI, Marcos A.; SPIESS, Maiko R.
Modalidades de regulação da atividade científica: uma comparação entre as interpretações normativa, cognitiva e transacional dos processos de integração da comunidade científica. *Educação e Sociedade*, v.31, n.110, p.73-92. 2010.
- MCCAIN, Katherine W.
Mapping authors in intellectual space: a technical overview. *Journal of the American Society for Information Science*, v.41, n.6, p.433-443. 1990.
- MCCROBERTS, Michael H.; MCCROBERTS, Barbara R.
A re-evaluation of Lotka's Law of scientific productivity. *Social Studies of Science*, v.12, n.3, p.443-450. 1982.
- MERTON, Robert K.
The Matthew Effect in science. *Science*, v.159, n.3810, p.56-63. 1968.
- MOWERY, David C.; ROSENBERG, Nathan.
Trajéórias da inovação. Campinas: Unicamp. 2005.

NARIN, Francis.

Evaluative bibliometrics: the use of publication and citation analysis in the evaluation of scientific activity. Washington: National Science Foundation. 1976.

OCDE.

Organização para a Cooperação e Desenvolvimento Econômico. *Manual di Frascati.* Coimbra: Gráfica de Coimbra. 2007.

OKRASA, Włodzimierz.

Differences in scientific productivity of research units: measurement and analysis of output inequality. *Scientometrics*, v.1, n.3, p.221-239. 1987.

PENDLEBURY, David A.; ADAMS, Jonathan. Comments on a critique of the Thomson Reuters journal impact factor. *Scientometrics*, v.92, n.2, p.395-401. 2012.

PINHEIRO, Lena V.R.

Lei de Bradford: uma reformulação conceitual. *Ciência da Informação*, v.12, n.2, p.59-80. 1983.

POPPER, Karl R.

A lógica da pesquisa científica. São Paulo: Cultrix. 2006.

PORTER, Theodore M.

Trust in numbers: the pursuit of objectivity in science and public life. New Jersey: Princeton University Press. 1995.

PRICE, Derek de Solla.

Toward a model for science indicators. In: Elkana, Yehuda et al. (Ed.). *Toward a metric of science: the advent of science indicators.* Brisbane: John Wiley. p.69-95. 1978a.

PRICE, Derek de Solla.

Editorial statements. *Scientometrics*, v.1, n.1, p.3-8. 1978b.

PRICE, Derek de Solla.

Little science, big science. New York: Columbia University Press. 1963.

SAISANA, Michaela; D'HOMBRES, Beatrice.

Higher education rankings: robustness issues and critical assesment. Luxembourg: Office for Official Publications of the European Communities. (JRC Scientific and Technical Reports). 2008.

SCHREIBER, Michael.

An empirical investigation of the g-index for 26 physicists in comparison with the h-index, the 4-index and the r-index. *Journal of the American Society for Information Science and Technology*, v.59, n.9, p.1513-1522. 2008.

SELEK, Salih; SALEH, Ayman.

Use of h-index and g-index for American academic psychiatry. *Scientometrics*, v.99, n.2, p.541-548. 2014.

SHATZ, David.

Peer review: a critical inquiry. New York: Rowman and Littlefield. 2004.

SGUISSARDI, Valdemar; SILVA JÚNIOR, João dos Reis.

Trabalho intensificado nas federais: pós-graduação e produtividade acadêmica. São Paulo: Xamã. 2009.

SILVA, Antonio O.

A sua revista tem Qualis? *Mediações*, v.14, n.1, p.117-124. 2009.

SMITH, Richard.

Peer review: a flawed process at the heart of science and journals. *Journal of the Royal Society of Medicine*, v.99, n.4, p.178-182. 2006.

THOMSON REUTERS.

2015 Journal Citation Reports. Disponível em: http://wokinfo.com/products_tools/analytical/jcr/. Acesso em: 26 out. 2015. 2015.

TOL, Richard S.J.

A rational, successive g-index applied to economics departments in Ireland. *Journal of Informetrics*, v.2, n.2, p.149-155. 2008.

URBIZAGÁSTEGUI ALVARADO, Rubén.

A Lei de Lotka na bibliometria brasileira. *Ciência da Informação*, v.31, n.2, p.14-20. 2002.

USHER, Alex; SAVINO, Massimo.

A global survey of university ranking and league tables. *Higher Education in Europe*, v.32, n.1, p.5-15. 2007.

VAN RAAN, Anthony F.J.

Measuring science: capita selecta of current main issues. In: Moed, Henk F.; Glänzel, Wolfgang; Schmoch, Ulrich. *Handbook of quantitative science and technology research: the use of publication and patent statistics in studies of S&T systems.* Dordrecht: Kluwer. 2004.

VANTI, Nadia A.P.

Da bibliometria à webometria: uma exploração conceitual dos mecanismos utilizados para medir o registro da informação e a difusão do conhecimento. *Ciência da Informação*, v.31, n.2, p.152-162. 2002.

VELHO, Lea M.S.

Como medir ciência? *Revista Brasileira de Tecnologia*, v.16, n.1, p.35-41. 1985.

VINKLE, Peter.

The evaluation of research by scientometric indicators. Oxford: Chandos. 2010.

WHITLEY, Richard.
The intellectual and social organization of the sciences. Oxford: Oxford University Press. 2000.

WHITLEY, Richard.
The establishment and structure of science as reputational organization. In: Elias, Norbert;

Martins, Hermínio; Whitley, Richard. *Scientific establishments and hierarchies*. London: D. Reidel. p.313-358. 1982.

WOUTERS, Paul.
The citation culture. Tese (doutorado em Estudos da Ciência e Tecnologia) – University of Amsterdam, Amsterdam. 1999.

