A Security Perspective on Publication Metrics

Hugo Jonker[1,2] and Sjouke Mauw[3]

[1] School of Computer Science, Open University of the Netherlands
hugo.jonker@ou.nl, http://www.open.ou.nl/hjo/
[2] Digital Security group, Sciences, Radboud University
[3] CSC/SnT, University of Luxembourg
sjouke.mauw@uni.lu, http://satoss.uni.lu/sjouke

1   Introduction

The importance of publication metrics, such as the h-index [1], has increased dramatically in recent years. Unfortunately, as Goodhart [2] already remarked: "when a measure becomes a target, it ceases to be a good measure". And indeed: hiring, grants and tenure decisions depend more and more on performing well in publication metrics. This leads to a perverse incentive for individual researchers and journals to "optimise" their perfomance. However, such behaviour undermines the utility of the measure itself, in the extreme case nullifying its value. The underlying cause is that besides the functional requirements on a measurement, there are also security requirements on them. As is often the case, these security objectives remain implicit. In this paper, we provide a much-needed security perspective on publication metrics.

There are different kinds of security issues that may arise. We distinguish between weaknesses in the design and weaknesses of a specific implementation. Moreover, problems can arise either without malicious behaviour, or due to a malicious actor. The case where a malicious actor abuses design weaknesses of a publication metric is overlooked in literature. We term such malicious actions "gaming". Gaming can be characterised as "those actions taken by a researcher that impact that researcher's publication metric, which he would not take in absence of that publication metric." For example, a journal editor is gaming when he insists that authors of submitted papers cite several previously published papers in the journal to game the journal's impact factor; or a citation ring of researchers is gaming when members of the ring systematically cite all other members, thereby artificially increasing their citation count.

There have been many discussions in literature on the use of and the methodology underpinning publication metrics. Concerning the h-index, for instance, an overview of discussions, weaknesses and alternatives can be found in [3]. This has led to an abundance of "improved" publication metrics and critiques, in turn, on these new metrics. The discussions mostly center around methodological shortcomings (e.g., skewing to young or old, to males or females, to researchers in western countries, etc.). The fact that the methodology can be deliberately exploited by a malicious agent is hardly touched upon in literature.

Contributions. We develop a formal model of the publication process. This enables us to model various publication metrics, including the h-index and impact factor. Using this model, we can derive a given metric's "gaming surface". The gaming surface of a publication metric identifies every potential source of change in the model of the publication process that results in a change in the publication metric. Ergo, any changes in the publication metric must originate from this gaming surface. Having identified all potential sources of change, one can design mitigation strategies based on the standard security triad of prevention / detection / correction. We propose several detection heuristics. The foremost detective strategy in our view is semi-automated anomaly detection. Finally, we advocate an open publication process, including not only reviews, but also communications between editors, reviewers, and authors.

2   Categorising weaknesses of publication metrics

Popular publication metrics have been avidly discussed and vivisected by researchers. For example, literature has strongly contested the h-index's suitability as an indication of a researcher's quality, see for instance [4,5,6] for specific critiques and [3] for an overview of discussions, weaknesses and alternatives. The perceived weaknesses of the h-index have led to a flood of alternative metrics, including the g-, e-, a-, and m-index (see e.g. [7] for an overview).

In order to make a clear distinction between the various types of weaknesses, we will order these along two independent axes. The first axis concerns the distinction between design and implementation, while the second axis concerns the distinction between weaknesses that can be observed already in a *friendly* environment and weaknesses that involve an *attacker*. These two axes imply the four types of weaknesses illustrated in Figure 1.

	Implementation	Design
Errors	Deficiencies	Methodological drawbacks
Attacks	Hacking	Gaming

Table 1. Four types of weaknesses related to publication metrics.

I. **Errors**
   a. **Implementation deficiencies.** Software design involves making design decisions. Moreover, implementation is an error-prone process. Taken together, it is likely that implementations of publication metrics contain specifics that cause them to produce different output than other implementations of the same publication metric. Such errors are not restricted to simple encoding or algorithmic errors. They could also be introduced at the software design level.

> *Example:* a researcher's h-index as computed by Google Scholar is (almost always) higher than his/her h-index as computed by Web of Science. This is because Web of Science uses a stricter definition of "publication" than Google Scholar.
>
> b. **Methodological drawbacks.** A publication metric will have methodological limitations, due to the discrepancy between what the publication metric actually measures (e.g. citations) and its purpose to measure research quality or quantity. These limitations are already present in the system in regular usage, with honestly behaving agents.
> *Examples for the h-index:* women score lower than men [5], citations in negative context contribute positively [8], retracted papers continue to contribute to the h-index [9], biased against developing countries [10], etc.

II. **Attacks**

> a. **Hacking.** A publication metric's implementation can be affected by security vulnerabilities, such as buffer overflows, vulnerable cryptography, flawed protocols, insecure authentication, etc. Any such vulnerability may be exploited to change a publication metric.
>
> b. **Gaming.** A publication metric is intended to reflect behaviour as done in absense of the metric. In essence, gaming weaknesses are possibilities to apply Goodhart's Law [2]. By crafting input specifically for the publication metric process, the process is unduly affected.
> *Examples:* Publishing pacts [11], reviewing one's own papers[4], create bogus papers citing existing papers[5], create bogus papers all citing each other [12], asking authors of drafts to include a citation to your related paper [11], etc.

We pose that methodological arguments are to be discussed in the field of bibliometrics, while attacks are within the purview of the field of security.

Recall the notion of gaming, described as "*those actions taken by a researcher that impact a researcher's publication metric, which he would not take in absence of that publication metric*". Thus, whether or not some activity is regarded as gaming clearly depends on an interpretation of the researcher's activities and underlying intent. The goal of this paper is not to determine intent, but to provide a method to determine the possibilities for gaming inherent in a publication metric, and to suggest countermeasures to prevent, detect or correct gaming if and when it should occur.

Various instances of gaming publication metrics are described in literature. For example, researchers have found evidence of gaming journal impact factors (see e.g. [13]); several researchers have set up sting operations to uncover sloppy review practices, e.g. [14,15]; and various researchers performed experiments to game the publication databases directly, bypassing the normal publication process completely (e.g. [12]).

---

[4] http://retractionwatch.com/2012/08/24/korean-plant-compound-researcher-faked-email-addresses-so-he-
[5] López-Cózar et. al. 2012 – deliberately not cited here as the researchers used their study to increase their own h-index.

All these works found a specific gaming attack on the publication process. However, a uniform approach to modeling and analyzing such gaming attacks is missing. To understand which gaming attacks can be executed, below we construct a model of the publication process, a so-called *publication structure*. This will help us to understand and possibly reduce the "gaming surface" of publication metrics.

3 Publication structures

In this section we define the notion of a *publication structure* using simple set theory, which describes the relations between the basic notions of *author*, *paper*, *publication venue* and *citation*. These notions are considered at two levels of abstraction: the *data view*, which concerns raw publication data, and the *publication view*, which defines the abstract categories used to calculate the *h-index* and other publication metrics.

3.1 Structure of the publication process

Even if the definition of a publication metric seems clear, the devil is in the details. Take, for example, the h-index. Hirsch [1] describes the h-index as follows: *"A scientist has index h if h of his or her papers have at least h citations each, and the other papers have ≤ h citations each."* This informal definition seems clear, but it does not define clearly what a citation is nor, more pressingly, where these citations originate. Indeed, different *publication data processors*, like Google Scholar and Scopus, often calculate significantly different h-indexes for the same author, even though they use exactly the same definition of h-index. We observe that the calculation of an h-index (and, similarly, any other citation-based bibliographical metric) is relative to a particular interpretation of the notions of *author*, *paper*, *publication venue* and *citation*, and the relations between these. We call such an interpretation of reality a *publication structure*. From an abstract point of view, a publication structure for a given publication data processor describes the set of rules, algorithms and heuristics that the publication data processor uses when building, updating and interpreting its publications database.

An important observation is that publication data processors base their calculations on a particular selection and interpretation of *raw data*. For example, a publication data processor may consider only digitally available papers, and consider only the author names as they literally appear in a given pdf file. Indeed, an author's name can be represented in many different ways, e.g., by using initials, by leaving out accents, or by variations in its transcription to the latin alphabet. Consequently, we will make a distinction between the various representations of an author's name and the author itself. Similarly, we consider the abstract notion of a paper as opposed to its various representations, allowing us, for instance, to consider the pre-proceedings and the post-proceedings version

of a paper as the same publication. Likewise, we make a distinction between publication venues and their representations.

Various practical and even standardized approaches have been developed that address the distinction between concrete representations in the publication process and their abstract entities. Examples are the ISBN number for books, the ISSN number for journals and the DOI for electronic documents. However, these notions do not necessarily line up with a publication data processor's notion of a publication. For instance, two editions of the same book may be considered as instances of the same publication for the purpose of citation counting, while they have different ISBN numbers. A promising initiative to define abstract author entities is the ORCID initiative[6], but it's still too early to observe its impact in practice. Consequently, publication data processors will continue to use their own proprietary abstract representations of authors, publications, and publication venues. We formalise this as a *data view*, expressing raw data, a *selection and interpretation function*, which represents a publication data processor's view on the raw data, and a *publication view*, which represents the publication data processor's selection and interpretation of the data view.

3.2 Formalisation

We assume the existence of *raw data* available to a publication data processor on the publication process – papers, venues, authors, citations. On this raw data we define a *data view*, which provides a selection σ of the raw data.

First we will introduce some standard mathematical notation used in the definitions below. Let $p\colon X \times Y$ and $q\colon Y \times Z$ be relations, then their *composition* is defined by $q \circ p = \{(x,z) \in X \times Z \mid \exists y \in Y\colon (x,y) \in p \wedge (y,z) \in q\}$. Let $\sim\colon X \times X$ be an equivalence relation on a set X, then the *quotient set* of X by $\sim$ is denoted by $X/\sim$.

The relation between the sets of author representations and paper representations is defined by $authored_\sigma$, where $authored_\sigma(a,p)$ means that author representation a (co-)authored paper representation p. The set of citations is denoted by C_σ, containing all expressions that, according to the publication data processor's standard, describe a reference to a paper. The citations occurring in a paper representation are expressed by the relation $contains_\sigma$, where $contains_\sigma(p,c)$ means that citation c occurs in paper representation p. Similarly, $refersto_\sigma(c,p)$ expresses that citation c refers to paper representation p. The relation $at_\sigma(p,v)$ states that publication representation p was published in venue representation v.

A publication data processor considers a specific subset of the raw data. Some papers may not be in its view – e.g., Google contains many more "publications" than Scopus. Moreover, the publication data processor will interpret its subset of raw data to unify double entries. We call this the *data view* of a publication data processor on the raw data.

---

[6] https://orcid.org/
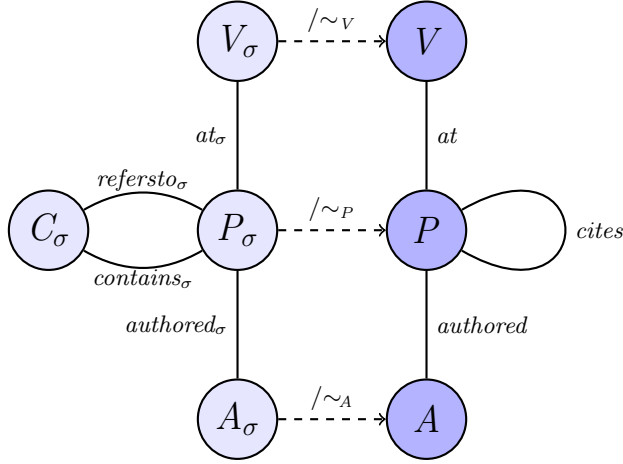
Fig. 1. Data view (left) and its induced publication view (right).

Definition 1 (Data view). *A data view dv is a selection σ of the raw data, i.e. $dv = (A_\sigma, P_\sigma, V_\sigma, C_\sigma, authored_\sigma, contains_\sigma, refersto_\sigma, at_\sigma)$, where*

- A_σ, P_σ, V_σ, *and* C_σ *are sets of author representations, paper representations, publication venue representations and citation representations, respectively,*
- $authored_\sigma : A_\sigma \times P_\sigma$ *is a relation defining which authors wrote which papers,*
- $contains_\sigma : P_\sigma \times C_\sigma$ *is a relation defining which citations occur in which papers,*
- $refersto_\sigma : C_\sigma \times P_\sigma$ *is a relation defining which papers are referred to in which citations.*
- $at_\sigma : P_\sigma \times V_\sigma$ *is a relation defining at which venue a publication appeared.*

The components of a data view are graphically depicted in Figure 1 (left).

Remark that an author can be represented in a data view in different ways: with initials (A.N. Example), with first name (Alfred Example), full name (Alfred Nicholas Example), etc. Similarly, different objects may represent the same paper (PDF version, HTML version), and the same venue may also be represented in different ways (abbreviation, full name). Unifying these representations constitutes an *interpretation* of the data view.

Definition 2 (interpretation). *An interpretation ι is a tuple $(\sim_A, \sim_P, \sim_V)$, where the equivalence relations interpret which representations are referring to the same abstract object in the classes A_σ, P_σ, and V_σ, respectively.*

By interpreting a given data view, we arrive at a *publication view* – the abstraction of a data view.

Definition 3 (Publication view). *Consider a data view*
$dv = (A_\sigma, P_\sigma, V_\sigma, C_\sigma, authored_\sigma, contains_\sigma, refersto_\sigma, at_\sigma)$, *and an interpretation* $\iota = (\sim_A, \sim_P, \sim_V)$. *The induced* publication view *$pv \in PV$ is a tuple* $(A, P, V, authored, cites, at)$, *where*

- *A, P, and V are sets of authors, papers, and publication venues, defined by $A = A_\sigma/\sim_A$, $P = P_\sigma/\sim_P$, and $V = V_\sigma/\sim_V$, respectively.*
- *authored: $A \times P$ is a relation defined by authored $= \{ (a, p) \mid authored_\sigma(a_\sigma, p_\sigma) \}$.*
- *cites: $P \times P$ is a relation defined by cites $= \{ (p, q) \mid refersto_\sigma \circ contains_\sigma(p_\sigma, q_\sigma) \}$.*
- *at: $P \times V$ is a relation defined by at $= \{ (p, v) \mid at_\sigma(p_\sigma, v_\sigma) \}$.*

The components of a publication view are graphically depicted in Figure 1 (right). The sets A, P, and V lift the sets A_σ, P_σ, and V_σ to an abstract level by dividing out their equivalences. Thus, the set A contains all authors, irrespective of the possible variety in representing their names in print. The relation *authored* lifts the author-publication relation from the representation level to the abstract level, so *authored*(a, p) means that author a authored publication p. Similarly, the relation *at* lifts the relation at_σ. The relation *cites* interprets the citations as a relation between citing papers and cited papers. By *cites*(p, q), we mean that paper p cites paper q – abstracting away from the representation of p, q and from the representation of the citation. Note that in the case of paper r containing a citation c that links to two papers p, q, this results in both *cites*(r, p) and *cites*(r, q). Additionally, for the case where r contains two citations c, c' that both point to the same paper p, this results only in one citation *cites*(r, p).

Using the notion of a publication view, we can formally define a *publication metric*.
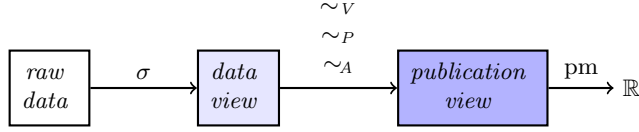
Definition 4 (Publication metric). *A* publication metric pm *is a function* pm: $PV \times Args \to \mathbb{R}$, *where Args is defined as*

$$Args ::= V \mid P \mid A \mid \mathbb{N} \mid Args \times Args.$$

The first argument is the *publication view* of the publication data processor. The remaining arguments allow us to define for which entity the metric is evaluated. For instance, the *h-index* has two arguments: a publication view and the author for which the *h-index* is calculated. When discussing publication metrics and their definition, we will often leave the publication view unspecified.

Some publication metrics make use of metadata that is not represented explicitly in our model. An example is the year of publication. Whenever necessary, we will assume that such metadata is implicitly represented in the model as an *attribute* of the concerned notion. For instance, we denote the publication year of a publication p as $p.year$.

In Figure 2, we provide a schematic representation of the role of a publication structure in determining the value of a publication metric. It expresses that a publication data processor's *data view* is constructed by collecting and selecting raw data. The *publication view* is then derived by equating the various equivalent names of the sets of venues, publications and authors. Based on this publication view, the publication data processor calculates the publication metric.

Fig. 2. From data to a publication metric.

3.3 Examples of publication metrics

In this section we show that a publication structure is not only rich enough to define the *h-index*, but also to express most publication metrics with a similar scope as the *h-index*. We start by defining two auxilliary functions. The first function expresses the number of times that a publication has been cited. For $p \in P$ this metric is defined by

$$\#citing(p) = |\{ q \in P \mid cites(q, p) \}|. \tag{1}$$

The second auxiliary function determines the set of all publications of author a. It is defined by

$$pubs(a) = \{p \in P \mid authored(a, p)\}. \tag{2}$$

Using these functions, we define the h-index of a as the largest subset of a's publications such that all its elements have at least as many citations as there are elements in the subset:

Definition 5 (*h-index*). *The h-index for a publication structure $S \in \Sigma$ is a function h-index: $A \to \mathbb{N}$, defined by*

$$h\text{-}index(a) = \max\{ i \in \mathbb{N} \mid \exists_{T \subseteq pubs(a)} |T| = i \wedge \forall_{p \in T} \#citing(p) \geq i \}.$$

In a similar way, we can define other publication metrics, like the g-index [16], or Google's i10-index. The i10-index is the number of articles (co-)authored by a with at least 10 citations. Formally:

$$i10\text{-}index(a) = |\{ p \in P \mid authored(a, p) \wedge \#citing(p) \geq 10 \}|.$$

The g-index of an author a is the largest number i such that the i most cited papers of a together have collected at least i^2 citations. Formally:

$$g\text{-}index(a) = \max\{ i \in \mathbb{N} \mid \exists_{T \subseteq pubs(a)} |T| = i \wedge \sum_{p \in T} \#citing(p) \geq i^2 \}.$$

Finally, we give an example of a metric for publication venues: the *impact factor*. For this publication metric we use the attribute *p.year*, which determines the year of a publication.

8

The impact factor of a publication venue v is the number of citations in a given year to articles published in v in the preceding two years, divided by the number of articles published in v in the preceding two years. Informally:

$$\frac{(\#\text{this year's citations to articles published in } v \text{ in last 2 years})}{(\#\text{articles published in } v \text{ in last 2 years})}.$$

We can make this more precise as follows. Consider a publication structure S. The set of publications that appeared in venue $v \in V$ in any period $[y, y']$ is then given by

$$app(v, y, y') = \{q \in P \mid at(q, v) \wedge y \leq q.year \leq y'\}.$$

The impact factor IF of a specific venue $v \in V$ for year y is then given by

$$\text{IF}(v, y) = \frac{|\{ (p,q) \in P \times P \mid q \in app(v, y-2, y-1) \wedge cites(p,q) \wedge p.year = y \}|}{|app(v, y-2, y-1)|}.$$

3.4 Extending the model

Note that some publication metrics make use of information that is not represented in our definition of a publication structure. An example is the *acceptance rate* of a venue – the percentage of submitted papers that was accepted. It is a simple matter to extend the model to be able to also capture metrics that incorporate slightly more data than available in the basic definition. We illustrate this by showing how to extend the model to define the metric *acceptance rate*.

We extend the data view with a relation *submitted-to*$_\sigma$: $P_\sigma \times V_\sigma$, which defines the venue to which a paper has been submitted. After extending the corresponding publication view with generalized relation *submitted-to*, the acceptance rate AR of a venue $v \in V$ is defined as

$$\text{AR}(v) = \frac{|\{ p \in P \mid at(p, v) \}|}{|\{ p \in P \mid \textit{submitted-to}(p, v) \}|}. \tag{3}$$

4 Gaming surface

To determine where a change in a publication metric originates, we identify every potential source of change in publication structures – the *gaming surface* of publication structures. The gaming surface of a publication metric consists of exactly those sets and functions of the data view on which the outcome of the publication metric depends. Therefore, a specific change in a publication metric must originate from a change in the publication structure's gaming surface. More precisely, a change in a publication metric's output can only be the result of a change in the underlying publication view, and a change in a publication view must originate either from changes in the underlying data view, or from changes in the equivalence relations used to derive a publication vier.

This means for publication structures S and S':

- $\text{pm}(pv_S) \neq \text{pm}(pv_{S'}) \implies pv_S \neq pv_{S'}$,
- $pv_S \neq pv_{S'} \quad\quad\quad\quad \implies dv_S \neq dv_{S'} \vee (\sim_A, \sim_P, \sim_V) \neq (\sim_{A'}, \sim_{P'}, \sim_{V'})$.

Thus, someone attempting to game a publication metric should craft or manipulate raw data in such a way that it influences the sets V_σ, P_σ, A_σ, C_σ, their relations $contains_\sigma$, $refersto_\sigma$, $authored_\sigma$, at_σ, and/or the equivalences $\sim_{A'}$, $\sim_{P'}$, $\sim_{V'}$. These items form the generic *gaming surface*. Since a particular publication metric may not depend on all of these, their actual gaming surface may be reduced. We illustrate how to determine the gaming surface of a particular metric by considering two example metrics: the number of publications of an author and the *h-index*.

4.1 Gaming surface for number of publications

Consider the metric "number of publications of an author", $a \in A_\sigma$, defined by

$$\text{numpub}(a) = |\{p \in P \mid authored(a, p)\}|.$$

Thus, this metric is only affected by changes to the publication view's A, *authored*, and P. These, in turn, are affected by the following changes in the data view:

- A: A_σ, $\sim_A$.
- P: P_σ, $\sim_P$.
- *authored*: $authored_\sigma$.

Table 2 lists the gaming surface of this metric, including the effects of possible gaming actions. Some actions have the desired effect of improving an author's number of publications, while others have no effect or an adverse effect. Concrete examples of the actions with a positive effect are: produce and publish bogus papers under your name, an author with a common name can claim publications of other authors with a similar or identical name, slightly change the title of a paper between its publication in pre- and post-proceedings, the editor of a journal accepts a paper under the condition that he becomes co-author.

4.2 Gaming surface of the h-index

Recall the definition of the h-index (Def. 5):

$$h\text{-}index(a) = \max\{\ i \in \mathbb{N} \mid\ \exists_{T \subseteq pubs(a)} |T| = i \wedge \forall_{p \in T} \#citing(p) \geq i\ \}.$$

This definition depends on three aspects: the given author $\bar{a} \in A$, the definition of the function *#citing* (Eq. 1), and the definition of the function *pubs* (Eq. 2). These two functions depend on the set P, and the relations *authored* and *cites*. This means that any change in an author's h-index must be triggered by changes to the underlying data view elements, i.e.:

	action	Poss. effect
A_σ:	add an author	none
A_σ:	remove an author	−
P_σ:	add a paper	+
P_σ:	remove a paper	−
$\sim_A$:	unify two author representations	+
$\sim_A$:	distinguish between two author representations	−
$\sim_P$:	unify two paper representations	−
$\sim_P$:	distinguish between two paper representations	+
$authored_\sigma$:	add an author to a paper	+
$authored_\sigma$:	remove an author from a paper	−

Table 2. Gaming surface for "number of publications".

- From A: A_σ, $\sim_A$,
- From P: P_σ, $\sim_P$,
- From *authored*: $authored_\sigma$, $\sim_P$, $\sim_A$,
- From *cites*: $contains_\sigma$, $refersto_\sigma$.

In Table 3, we list all update actions that could affect one's h-index, keeping in mind the type of component. The table identifies what the potential immediate effect (if any) of the update is. As can be seen, not all updates have an immediate effect. For example, a newly introduced author is not linked to any paper, and therefore has no h-index. On the other hand, some effects require multiple updates. For example, to increase one's h-index with +1, a researcher needs to publish a new paper, and that paper needs to be cited. In the model, there is also a formal constraint: to extend *cites* with p, p', there must be a pair p, c to $contains_\sigma$ and a pair p', c to $refersto_\sigma$. An update that adds only one of these pairs will not propagate to the publication view.

In the table, we find that some updates may lower a person's h-index. This implies that not only can an attacker aim to improve his own h-index, he can also work to lower his competitors' h-indices.

Moreover, we also find some actions which by themselves will not alter a h-index: adding an author to A_σ, adding a paper to P_σ and extending the relation $contains_\sigma$. Such alterations need to be followed by further alterations (linking the author to a paper, linking the paper to citations) to influence a subject's h-index.

Updates to $\sim_P$ deserve special mention, as the same type of update may have a positive, a negative effect, or no effect. Unifying two lowly-cited paper representations may give the resulting combination enough citations to let it contribute towards the h-index. Conversely, if two highly-cited paper representations are unified, this may lower the h-index as there is one less paper contributing to the h-index. Distinguishing two paper representations can have the opposite effect: breaking a paper that counts for the h-index into two that do not count for the

11

	action	Poss. effect
A_σ:	add an author	none
A_σ:	remove an author	$-$
P_σ:	add a paper	none
P_σ:	remove a paper	$-$
$\sim_A$:	unify two author representations	$+$
$\sim_A$:	distinguish between two author representations	$-$
$\sim_P$:	unify two paper representations	$+, -$
$\sim_P$:	distinguish between two paper representations	$+, -$
$authored_\sigma$:	add an author to a paper	$+$
$authored_\sigma$:	remove an author from a paper	$-$
$refersto_\sigma$:	cite a paper	$+$
$refersto_\sigma$:	remove a citation to a paper	$-$
$contains_\sigma$:	inject a reference in a paper	none
$contains_\sigma$:	remove a reference from a paper	$-$

Table 3. The gaming / update surface of the *h-index*

h-index will lower the h-index, while breaking such a paper into two papers that both count will increase the h-index.

Finally, we point out that Table 3 identifies the effect on an individual's h-index. Group effects are not accounted for. For instance, getting into the top 5% of authors can easily be achieved by adding many (bogus) authors without publications.

5   Mitigation strategies

We develop mitigation strategies based on an analysis of the involved parties and their incentives. We recognise the following parties:

– Targets of evaluation: researchers, journals, workshops, but also universities, departments, etc.
  Incentives: to score as well as possible on publication metrics.
– Stakeholders in the publication process: reviewers, editors, program committee members, publishers, etc.
  These may be evaluated by publication metrics as well. Their incentives therefore are the same as targets of evaluation.
– Publication data processors: Google Scholar, Thomson, etc.
  Incentives: provide reliable data.
– Evaluators: employers, evaluation committees, funding agencies, etc.
  Incentives: support the decision process based on an evaluation that correctly uses the right publication metrics.

Given the incentives and the gaming surface, it is clear where an actor can act to further his goals. This helps guide where mitigation can take place and where to place security controls.

Mitigation must be based on the data available. Curiously, in scientific publishing, details such as reviews, communication between authors and reviewers, communications between reviewers and editors, etc. are *not* publicly available. However, there are types of gaming that can only be detected in such data. One example is a reviewer who requires a paper under review to add a citation of his work.

Manipulations of the publication process often escape the attention of quality control. For example, quality control levels proposed by the *Transparency and Openness Promotion* group [17] extensively detail quality controls for the research process, but lack any considerations of the reporting process.

We therefore advocate that data concerning the *reporting process* must be publicly available.

We believe that the most relevant mitigation strategies are based on analysing the aggregated data to detect possible gaming. Borrowing from the field of intrusion detection, we discern two types of detective strategies:

– Signature-based detection.
  Signature-based detection identifies patterns in the data that are common to a particular gaming attack. An example of such a detector is the SciGen detector by Labbé et al. [18].
– Anomaly-based detection.
  Anomaly-based detection identifies deviations from the expected norm in the data. An example is the following check by Thomson for detecting a journal that is gaming: a very quick increase in impact factor of the journal, coupled with a high number of self-citations.

Currently, a few publishers are taking some initial measures in this direction. However, current efforts are far behind what is possible. We advocate the development and use of tools to automatically analyse data and identify anomalies. Remark that not all detected cases will be instances of gaming – some will be benign. For example, a detector that detects papers that are cited before their official publication can also trigger on a paper for which the publisher makes a webversion available prior to official publication. As such, the detection process can never be fully automated. Human evaluation will be required, but the effort necessary to identify gaming can be significantly reduced.

6   Conclusions

Publication metrics are playing an ever-increasing role in scientific careers and scientific standing of journals and other publication venues. This creates a perverse incentive for "gaming" a publication metric: taking actions that improve a metric, which would not have been taken in absence of the metric. Publication metrics suffer from various drawbacks, both in implementation and in design. We showed that gaming is a category in itself, which has received scant attention in literature.

Gaming of publication metrics is a disservice to scientific advancement. We argue that steps are needed to prevent flagrant gaming. To this end, we introduced the notion of a *gaming surface* of publication metrics, and developed a formal framework to rigorously define the gaming surface of any specific publication metric.

Defences against gaming are then built upon understanding the given publication metric's gaming surface. We discussed potential security controls and classified these.

Such measures can help to not only safeguard publication metrics, but can also help to identify some forms of attempted fraud at an early stage. As such, we advocate ongoing research into designing security controls for publication metrics.

References

1. Hirsch, J.E.: An index to quantify an individual's scientific research output. Proceedings of the National Academy of Sciences of the United States of America **102**(46) (2005) 16569–16572
2. Goodhart, C.: Problems of monetary management: the UK experience. In: Papers in Monetary Economics I. Reserve Bank of Australia (1975)
3. Alonso, S., Cabrerizo, F., Herrera-Viedma, E., Herrera, F.: h-index: A review focused in its variants, computation and standardization for different scientific fields. Journal of Informetrics **3**(4) (2009) 273 – 289
4. Guilak, F., Jacobs, C.: The h-index: Use and overuse. Journal of Biomechanics **44** (2011) 209–209
5. Kelly, C.D., Jennions, M.D.: H-index: age and sex make it unreliable. Nature **449**(7161) (2007) 403–403
6. Waltman, L., van Eck, N.J.: The inconsistency of the h-index. Journal of the American Society for Information Science and Technology **63**(2) (2012) 406–415
7. Bornmann, L., Mutz, R., Daniel, H.D.: Are there better indices for evaluation purposes than the h index? a comparison of nine different variants of the h index using data from biomedicine. Journal of the American Society for Information Science and Technology **59**(5) (2008) 830–837
8. Wendl, M.C.: H-index: however ranked, citations need context. Nature **449**(7161) (2007) 403–403
9. Glänzel, W.: What is the impact of fraudulent literature? International Society for Scientometrics and Informetrics Newsletter **6**(2) (2010) 44–57
10. Mishra, D.C.: Citations: rankings weigh against developing nations. Nature **451**(244) (January 2008)
11. Parnas, D.L.: Stop the numbers game. Commun. ACM **50**(11) (November 2007) 19–21
12. Labbé, C.: Ike Antkare one of the great stars in the scientific firmament. International Society for Scientometrics and Informetrics Newsletter **6**(2) (2010) 48–52
13. Arnold, D.N., Fowler, K.K.: Nefarious numbers. Notices of the AMS **58**(3) (2011) 434–437
14. Sokal, A.D.: A physicist experiments with cultural studies. Lingua franca **6**(4) (1996) 62–64
15. Bohannon, J.: Who's afraid of peer review? Science **342**(6154) (2013) 60–65

16. Egghe, L.: Theory and practise of the g-index. Scientometrics **69**(1) (2006) 131–152

17. Nosek, B.A., Alter, G., Banks, G.C., Borsboom, D., Bowman, S.D., Breckler, S.J., Buck, S., Chambers, C.D., Chin, G., Christensen, G., Contestabile, M., Dafoe, A., Eich, E., Freese, J., Glennerster, R., Goroff, D., Green, D.P., Hesse, B., Humphreys, M., Ishiyama, J., Karlan, D., Kraut, A., Lupia, A., Mabry, P., Madon, T., Malhotra, N., Mayo-Wilson, E., McNutt, M., Miguel, E., Paluck, E.L., Simonsohn, U., Soderberg, C., Spellman, B.A., Turitto, J., VandenBos, G., Vazire, S., Wagenmakers, E.J., Wilson, R., Yarkoni, T.: Promoting an open research culture. Science **348**(6242) (2015) 1422–1425

18. Labbé, C., Labbé, D.: Duplicate and fake publications in the scientific literature: how many scigen papers in computer science? Scientometrics **94**(1) (2013) 379–396