Content-Oblivious Quality Measures and the Control of Academia

Oded Goldreich
Department of Computer Science
Weizmann Institute of Science
Rehovot, Israel.
oded.goldreich@weizmann.ac.il

July 20, 2015

Abstract

The essay examines the common practice of relying on content-oblivious quantitative measures for the evaluation of the quality of academic research. It starts with a definition of these measures, while distinguishing the raw bibliographic data and the way it is processed (to obtain a numerical value). It exposes the hidden decisions that determine which pieces of data are used and how they are processed (i.e., which statistics is taken of it). Indeed, content-oblivious quantitative measures are obtained by an automatic processing of superficial parameters of scholarly work, and their claim for objectivity hides the arbitrariness of the decisions on which they are based, and does not allow to discuss these decisions.

The direct consequence of relying on content-oblivious quantitative measures is reaching uninformed decisions, since professional judgment that relates to the content is replaced by a superficial measure whose relevance to the questions at stake is highly questionable. The indirect consequences of relying on these measures are even more dangerous. This practice neglects the actual content while fetishizing quantity, it replaces the academic vocation by preparation for accounting, it oppresses intellectual curiosity, encourages manipulations, and reproduces power relations within scientific disciplines and between them. The content-oblivious quantitative measures are compatible with shifting the focus of evaluation from verifying the satisfaction of threshold criteria to ranking and forming a rigid hierarchy.

The essay presents several explanations for the popularity of these content-oblivious quantitative measures: They fit the neo-liberal order and its preference for standardized regulation procedures; they carry a

> seductive promise of objectivity (which is extremely tempting to modern science); they serve opportunism (in the form of intellectual laziness and escaping responsibility); they empower the academic-managerial class by providing it with mechanisms for control of the academic works (by subjecting scientific knowledge to managerial knowledge); they facilitate diffusion of business attitudes to the academic world and the domination of scientific content by (bibliometric) technology. All these phenomena are related to the rise of a new political and scientific order in which a tighter control of academia plays an important role.

Following is an English translation of the introduction of the essay, which was written in Hebrew.

Introduction

Although all members of academia are supposed to be aware of the extreme complexity of the academic activities (of research and teaching) the high level of professionalism required for understanding their content, let alone to evaluate their quality, many of them rely on content-oblivious quantitative measures when trying to assess academic quality. It seems that confidence in these measures is rooted in their mathematical and objective (or impersonal) appearance. In particular, repeated applications the same measure regarding the same question will yield the same answer, no matter who invokes the application.[1]

But also the height of the member of the academia is an objective measure, so is their weight and age. Likewise, the total length of their publications is an objective measure, so is the weight of the books on their shelves, and the average age of their students. And likewise, the number of papers that they published and the number of citations that these papers received are objective measures. Of course there is a difference: Nobody uses the former measures whereas many use the latter. Indeed, the claim that there is any relation between the former measures and academic quality seems absurd, whereas such a claim regarding the latter measures sounds reasonable. *But is this claim actually correct? Was it established that this relation is significant enough to allow relying on it in determining the fate of individual academics, or of academic units, universities, or all universities in a state?*

These questions are indeed begging. But they are also premature. They presume that it is known what is academic quality, that it is agreed that it

[1] This assertion presumes that all users know how to apply the measure correctly, which is clearly a non-universal condition (and yet nobody seems aware of this condition).

should be evaluated in certain cases (i.e., in the very cases where one seeks to do so), and that the evaluation process is aimed at describing the state of affairs (rather than forming a new state of affairs). These assumptions are quite naive, and one should critically wonder about what they take for granted. Indeed, one should start with the following questions:

- What is academic quality? Who determines the answer and why is it determined?
- Why and when is it adequate to evaluate academic quality?
- Is the evaluation procedure aimed at describing a given situation or is it aimed to create a new situation? An in the latter case, what is the new situation that is sought and why is it sought; and how does the change in the situation relate to academic quality as discussed in the

This essay will not attempt to answer the first question, but rather postulates that a key principle that must underly any reasonable definition of quality (in any field) is that *the quality of activity in a field is internal to the field and can be evaluated only within the field*. Evaluating an activity in the field requires a good understanding of the field (i.e., expertise in the field). Full or partial expertise can be acquired, also by people outside the field, but the level of expertise in the field determines the level of reliability of the evaluation of the quality of activities in the field. In particular, for the purpose of evaluating the quality of an activity in the field, one cannot replace an understanding of the specific field by an understanding of the general dynamics of fields.[2]

The principles, which are all implied by the key principle (i.e., that quality is internal feature of contents of the field), invalidates any attempt to evaluate research in a discipline based on content-oblivious quantitative measures (while lacking any understanding of the content of the discipline). A detailed analysis of the unreliability of such attempts constitutes a main focus of this essay.

Another central issue, which deserves attention, is the obsessive preoccupation with evaluating academic quality rather than making candid attempts to improve it. The hidden assumption that guides this preoccupation is that

[2]The term "dynamics of fields" is taken from Bourdieu, but is extended here to cover also general knowledge about the current state of the academia and the interactions between disciplines in it. This is the type of knowledge that serves as a basis for the most popular content-oblivious quantitative measures, and it stands in contrast to a deep and wide understanding of a specific discipline.

the evaluation of quality leads to is improvement via mechanisms that inflict positive and negative sanctions (i.e., benefits and punishments) based on the result of the evaluation. This assumption is valid in some situations, but invalid in others. Furthermore, over-preoccupation with quality evaluation may lead to a decline rather than improvement in quality.[3] In any case, quality evaluation cannot replace the allocation of resources. Furthermore, in some cases lower quality requires investing more resources rather than cutting resources.

The obsessive preoccupation with evaluating academic quality is closely related to the emergence of the "Audit Culture" as a part of the neo-liberal order [5]. This refers to the dominance of an accountancy-like mentality in the social sphere and its application to the management of governmental and public institutions. The said audit, which is typically external to the institute and lacks expertise in the content of its activity, results in ignoring the actual contents and its meaning. In the context of academia, this culminates in the increased dominance of content-oblivious quantitative measures, and the immediate result is a *significant devaluation in the importance attached to the actual research and concentration on improving the value of the measure*. In the long run, the result is neglecting (or even abandoning) the formative goals of the academic system, which are transformative with respect to the society, and replacing them by the conservative goals of serving and reproducing the current social order.[4] It is indeed ironic that the desire to serve the current social order leads to a desire to transform the academia.

There is an inherent connection between attempting to perform external quality evaluation and the dominance of content-oblivious quantitative measures in such an evaluation. The reduction of contents to quality eliminates the advantage of expertise in the field, since quantities are clear to all. Furthermore, non-experts gain an advantage since they feel comfortable with

[3] An analysis of the effects of over-preoccupation with quality evaluation in the context of pre-academic education is provided in Sahlberg's book on the Finnish education system [4]. In particular, it focuses on the harmful effect of focusing on improving the ranking with respect to external standardized tests, which comes at the expense of actually learning and understanding the material itself. Another harmful effect is discouraging the creativity and motivation of the teachers, which is the most important resource of a good education system. These points are all the more valid when it comes to academic education (let alone to academic research, under obvious modifications).

[4] Concrete examples in which the operative goals were modified in respond to external audits (which were based on content-oblivious quantitative measures) are presented in [2], which considers the context of admission to Law schools in the US. For wider perspective on the transformation of the academic system, see [3].

ignoring the contents of the field and treating the activity in it via the prism of content-oblivious quantitative measures. This enables the oppression of those who actually working in the field (i.e., the academic workers) by those who belong to the managerial class (and by those who control them via the external politics).

The last two paragraphs clarify that, in contrast to the pretense that the quality evaluation process reflects an existing situation, the external evaluation process actually constructs a new situation: The fact that this process changes the activity and the content of the evaluated field is not an accident caused by a specific implementation of the evaluation process but is rather inherent to this process. The change is a drastic one; it is a fundamental change in the general goal of the academic system: Turning a system that its central ethos is transforming society via the discovery and dissemination of new knowledge and enriching conceptualizations into a system that is aimed at serving and reproducing the social order, as culminating in viewing the students as interested solely in acquiring professional skills. The fact that the academic system was never sufficiently committed to its central ethos does not contradict the observation that the current process leads to the full abandoning of this ethos.

On Quality, Impact, and Visibility

While the conceptual gap between quality and quantity is self-evident, one may confuse the actual content with context-oblivious quantitative measures when referring to notions such as impact and visibility. Nevertheless, also when considering these notions, a second thought reveals a gap between their actual meaning, which is a moment of quality, and content-oblivious measures that claim to reflect this meaning.

When discussing the impact (or influence) of a specific scientific work, the intended meaning is that of the influence of this work on the relevant area of research as reflected in its contribution to the world-view of experts in the area (which changes to some (possibly small) extent their world-view of the area). The contribution may amount to additional factual information, and/or to a new conceptual perspective on known facts, and/or to a refutation of known beliefs about facts and/or to a critique of a known conceptualization. Such a contribution may be more or less significant, have a wide or narrow scope, have huge or small potential, etc. Hence, *influence in this essential sense* is linked to the content of the relevant discipline, and evaluating its magnitude requires an understanding of that discipline. This

essential meaning of the influence (or impact) of a work is closely related to the quality of the work, and there is a huge conceptual gap between it and the number of citations of the said work. In particular, the number of citations is not of intrinsic interest, but is rather supposed to serve as an indicator of the essential influence of the work. However, for reasons to be discussed in this essay, this indicator is highly unsound: A strong correlation between the qualitative measure and the essential notion of influence exists only in extreme cases, whereas the common use of this measure is in normal cases.

Unfortunately, influence (or impact) in the essential sense is easily confused with quantitative measures that bear a name containing the work 'impact' (such as the "impact factor" of journals). Similarly, 'visibility' in an essential sense refers to the way specific researchers are perceive in their research community, but this sense is easy to confuse with quantitative measures such as the number of appearances of researchers in certain scientific venues. Given that the essential meaning of impact and visibility is closely related to academic quality, I chose to conduct the discussion in terms of quality (rather than in terms of impact and/or visibility), since the notion of quality causes less conceptual confusion.

Organization

Following is a brief description of the subsequence sections of the essay. First, we shall define and clarify the notion of *content-oblivious quantitative measures*, describe a few popular measures of that type, and briefly review the history of their development and usage. Next, we shall uncover a sequence of *problematic decisions that are hidden under the objective appearance of the content-oblivious quantitative measures*. These decisions include the very decision to ignore the actual content of the publications and view all publications that appear in the same venue as equal; the decision of what counts as publication and consequently what is included in the bibliographic database; and the decision of what statistics is to be extracted from the database (i.e., determining the mathematical formula that assigns numerical values to collections of publications). We shall ask who determines the answers to these questions, based on what considerations, and how does this reflect the power relations between different disciplines and between different approaches within a discipline.

As just said, the content-oblivious quantitative measures are not "neutral" but rather reflect hierarchies between between disciplines and within

disciplines; that is, reflect the hierarchies held as natural by the designers of the measures. In particular, measures that are developed by engineers and scientists who are familiar with some scientific disciplines are likely to fit the publication culture in these disciplines, and may be highly inappropriate for other disciplines in which the publication norms are very different. Hence, the *legitimate variety of scientific cultures is not respected*, since the measures fit publication norms that hold in some scientific communities but not in others.[5] One acute example is the preference, given by all popular measures, to publications in international venues (especially, in English), where in some cases publications in local venues (and/or in other languages) are totally ignored. Needless to say, if these measures influence the allocation of resources, then they *create a strong bias in favor of some disciplines at the expense of oppressing others*. One of the consequences is the discouragement of research that is focused on the local society and culture, which is *especially harmful to peripheral societies like the Israeli society*.

We then turn to the direct and indirect effects of using content-oblivious quantitative measures for the evaluation of scientific research. *The direct consequence of dominant usage of content-oblivious quantitative measures is uninformed and often wrong decisions*. The decisions are uninformed by definition, since these measures ignore the content (i.e., the information). The decisions are often wrong because the correlation between quality and these measures is too low to justify relying on the measures. In particular, we point out the huge quality variance among publications, even among those that appear in the same venue, and the big variance among the publications that are cited in the same scientific paper (regarding both the contributions of the cited publications to the current paper, let alone their contributions to the area at large). We stress that this phenomenon occurs also when comparing publications within any discipline, but is indeed stronger when referring to publications in different disciplines (where inter-disciplinary studies may suffer the fate of comparisons between different disciplines).

The direct damage of the dominant usage of content-oblivious quantitative measures is amplified by their tendency to hide the problematic nature of the decisions on which they are based, which prevents an open discussion

[5]Indeed, attempts to enforce norms of some disciplines on other disciplines existed also before the rise of the use of content-oblivious quantitative measures, but the use of these measures strengthens this tendency. In particular, in the past these attempts took place at the level of disciplines and involved open confrontations between opposing programmatic views, whereas nowadays publication norms are enforced at the level of individual researchers (via the usage of these measures) and involve decisions that are not open to debate (see below).

of the principles that underly these decisions. The preference of publications of a specific type is always rooted in positions regarding the research agenda of the discipline, whereas these positions are often in dispute within the discipline. Needless to say, such disputes are legitimate and actually healthy, but the dispute should be open to discussion, whereas these measures are based on one position in the dispute, which is not stated explicitly and is not debatable. In addition, an increasing reliance on content-oblivious quantitative measures leads to neglecting the only reliable source for the evaluation of academic quality – reviews that are based on understanding of the content of the academic activity and judgment based on expertise in the relevant discipline.

The *indirect consequences of the usage of content-oblivious quantitative measures* are even more harmful than the direct consequences. The most clear damage is significant erosion of the academic commitment to candid and ambitious research, which is the hallmark of the quality of content, to preoccupation with the quantity of publications. Needless to say, the purpose of academic research is not only to discover new knowledge or new perspectives but also to share these discoveries with the relevant scientific community, whereas this sharing takes place via the media of publications. Yet, this fact does not change the perception of research discoveries as the goal, and of publications as means for sharing it. The dominance of content-oblivious quantitative measures stands in contrast to this perception and erodes it: *Publications become the goal, whereas research is viewed as means for obtaining them.*

Transforming the status of academic publications from a mean to a goal is not a novelty of the content-oblivious quantitative measures, but is rather a historical process that goes along with the transformation of academic research from an aristocratic hobby to a (prestigious) profession. Hence, we are talking about a significant amplification of an existing process: The dominant usage of content-oblivious quantitative measures strengthens the change in the status of publications and the relation to them. The dominant message that resonates all around calls for optimizing various quantitative parameters of publications (e.g., their number, citation counts, venues, etc), which is fundamentally different from calling for an improvement of the quality of research. In particular, even when one hears a reference to quality, it is in reference to the quality of the publication venue (rather than in reference to the content of the research itself). All this may even lead to attempts to *manipulate the measures.*

These phenomena *encourage research of a conformistic character*, which evolves along well-established and popular directions, since its results are

easier to publish in the "prestigious" venues of the mainstream, *and discourages new and/or critical research directions*, which tends to be excluded from the "prestigious" venues and get published only in venues labeled "marginal" by the mere fact that they publish non-mainstream works. The chilling effect of these facts is more acute on academic institutes and disciplines that are weakened by the current academic mangers, but also institutes and disciplines that posses a stronger position are not immune to it.

The next section deals with common *defenses of the use of content-oblivious quantitative measures*. We distinguish between defenses that argue that this usage does not change the essence of the evaluation process and defenses that support this usage in hope that it will significantly change the essence of the evaluation process in a progressive manner. Our opinion is that the first type of defense over-estimates the reliability of the measures and under-estimates the direct' and indirect damages caused by their usage, whereas the second type of defense unjustifiable believes in the "egalitarian promise" of these measures and fails to notice their oppressive nature.

The topic of the next section is the *compatibility of the use of content-oblivious quantitative measures with a shift in the goals of the evaluation process*. The original goal of quality evaluation (in academia) was to verify the satisfaction of threshold criteria for the purpose of resource allocation. Archetype examples include the hiring and promotion of academics and the acceptance of paper for publication in various venues. Needless to say, also such threshold decisions are relative, but their focus is the specific body of work that is being evaluated. In contrast, the goal of much of the current preoccupation with quality evaluation is to *create a ranking and form a rigid hierarchy*, especially among academic units and institutes.[6] Forming such a ranking requires ignoring the different content and different visions of different units and institutes, and content-oblivious quantitative measures are just perfect for ignoring these elements of content. Indeed, the begging question is what goals are being promoted and served by a full ranking and a rigid hierarchy, and what are the consequence of forming such a hierarchy.

These questions are related to an attempt to explain the *popularity of these content-oblivious quantitative measures*. Such attempt is called for in light of the critique on the reliability of these measures and in light of

[6] One may claim that hierarchy and inequality are inherent in the academia, but this claim refers to a hierarchy that is created based on academic achievements. Hence, in principle, this is a flexible hierarchy, which changes frequently as a function of new achievements. The hierarchy formed by content-oblivious quantitative measures tends to be self-perpetuating and its reaction to new achievements, especially revolutionary ones, tends to be smaller and slower (for reasons stated above).

the direct and indirect damages of their usage. The following explanations arise from different perspectives and combined consistently to form a strong explanation.

One explanation, which was already hinted and seems most important, is that the use of content-oblivious quantitative measures perfectly fits the *Audit Culture*, which is a central aspect of the neo-liberal order [5]. A key component of this order is a weakening of all public institutions by subjecting them to external audit that ignores their social function and the content of their activity while analyzing them in pure quantitative terms of economic flavor. In the case of academic institutions, the scientific content is ignored and replaced by (or reduced to) an analysis of a business management flavor. Moving the focus from the scientific content to the business (or managerial) content fits well in the transition of power from the actual researchers to the managers.

The use of content-oblivious quantitative measures in the evaluation of research allows to *eliminate the inherent dominance of expertise* in such an evaluation, and even creates an advantage for those who specialize in using these measures (of feel more comfortable with them). Effectively denying the importance of the scientific content diminishes the academic authority of the actual researchers (i.e., the academic workers), puts down their occupation, glorifies administrative techniques, and provides legitimization for the policing of the academic workers by the academic administration (and the class that controls politics).

Content-oblivious quantitative measures fit well in the modern tendency to *objectify knowledge* and present it in a way that eliminate its meaning and the subject altogether (cf. [1]). In fact, the dominant usage of content-oblivious quantitative measures of academic quality is enabled by an intellectual atmosphere in which the actual content and meaning of the academic activity are being weakened and diminished. Needless to say, the materialization of this potential depends on the existence of a suitable technology, which appeared at the late 1990s when the use of content-oblivious quantitative measures based on a bibliographic database became available to all.

On the personal level, using content-oblivious quantitative measures presents an *opportunistic temptation*: It allows to replace the intellectual effort involved in reading and understanding material that is rich in content by glancing at a single numerical value. In addition, it offers a way of escaping responsibility for decisions, by claiming that they were determined by the quantitative measures. Needless to say, the opportunistic behavior requires legitimization or at least the non-existence of norms that condemn this be-

havior. This condition is met by the social and intellectual atmospheres described above (i.e., the Audit Culture and the fetish of objectivity, resp.).

The neo-liberal regime relies on "free" (of visible force) collaboration of its subjects, where this collaboration is far more vital to this regime than it was in prior regimes. This phenomenon explains the need of a deeper control of the regime in the consciousness of its subjects, and indeed the regime has established such a deep control (via TINA and other rhetorical claims). This state of affairs also holds an emancipatory potential: The regime will be quite helpless if opposed by a refusal of a significant part of an important institution, like academia. (That is, it will not be able to force collaboration.) In particular, changing the evaluation procedures in the academia and abolition of the uses of content-oblivious quantitative measures can be undertaken by the academia itself (and cannot be opposed by the regime).

References

[1] L. Daston and P. Galison. *Objectivity*. Zone Books, New York, 2007.

[2] W.N. Espeland and M. Sauder. Rankings and reactivity: how public measures recreate social worlds. *American journal of sociology*, 113 (1), 1-40, 2007.

[3] W. Halffman and H. Radder. The academic manifesto: from an occupied to a public university. *Minerva*, Springer, DOI 10.1007/s11024-015-9270-9, Brief Communication, 2015.

[4] P. Sahlberg. *Finnish lessons*. Teachers College Press, 2011.

[5] C. Shore. Audit culture and illiberal governance (universities and the politics of accountability). *Anthropological Theory*, 8 (3), 278-29, 2008.