How to do research on the societal impact of research? Studies from a semantic perspective

Andrea Bonaccorsi[1], Filippo Chiarello[1], Nicola Melluso[1] and Gualtiero Fantoni[2]

[1] *a.bonaccorsi@gmail.com, nicolamelluso@gmail.com, filippo.chiarello@unipi.it*
Department of Energy, Systems, Territory and Construction Engineering, University of Pisa, Pisa (Italy)

[2] *gualtiero.fantoni@unipi.it*
Department of Civil and Industrial Engineering, University of Pisa, Pisa (Italy)

Abstract
We review some recent works of our research lab that have applied novel text mining techniques to the issue of research impact assessment. The techniques are Semantic Hypergraphs and Lexicon-based Named Entity Recognition. By using these techniques, we address two distinct and open issues in research impact assessment: the epistemological and logical status of impact assessment, and the construction of quantitative indicators.

Introduction

In the latest few months, we have published four papers in various journals that contribute to the debate on research impact (Bonaccorsi et al. 2020; 2021a; 2021b; 2021c). The overall aim of this effort is to call the attention on two dimensions of the research impact that have been somewhat neglected. The first is a theoretical issue: What kind of statement is a statement of the type "research X has produced an impact on Y"? What is the logical nature of this statement? What is the semantic structure of a statement of this type? It turns out that addressing this apparently theoretical question has far reaching implications. The second is a methodological issue: Which methods can we use for research impact assessment? Can we build indicators, as it happens in other areas of S&T? In this paper we deepen the conceptual issues underlying these papers, we describe their findings, and illustrate a research program for future studies and practical applications.

The historical nature of research impact statements

What kind of statement is a statement of the type "research X has produced an impact on Y"? When asked to give account of the impact of their research, researchers make use of a narrative. This is dictated by the nature of the requirement: the impact of research is a process that unfolds over time, involving several actors, with many events, facts, accidents taking place at different points in time. The outcome is a change of state that does not take place abruptly but is prepared by a sequence of events and a plurality of actors. To make sense of the unfolding of the impact process the only type of writing is a narrative one. Researchers must persuade their readers that their research has indeed taken part to a historical process, whose reality can be demonstrated, in which some actors have directly or indirectly benefited from it, leading to some improvement. To persuade the readers the researchers must do two things: first, it must build a narrative that unfolds over time, is plausible and realistic, and ends up with the impact; second, it must demonstrate that within this narrative, researchers have had a role, that is, have produced the impact, or have contributed to the production of the impact.
The narrative style is shared by two fundamentally different types of writing: history and fiction. Historical narrative is expected to reconstruct sequences that represent historical facts, or events that have taken place in historical time, according to the best available documentation. Historical narratives may have a subjective flavour in the way in which the flow of events is

reconstructed and the importance is weighted, but they must comply with severe standards of control about the historical truthfulness of their statements. In contrast, fiction has no obligation whatsoever with respect to truth.
In order to build a flow of events that make sense, when writing a narrative text authors place of attention in giving details, offering a rich and contextualized description of what indeed happened. Vividness, detail, and richness of the description are essential component of narratives.
The narrative nature of research impact reports, however, may reasonably lead to the claim that they cannot include causal statements. To have a causal statement one must satisfy some general requirements that cannot be satisfied if the action takes place in the past. According to an influential tradition of research in post-structuralism, if the events narrated are placed in the past, then there is no difference between history and fiction, because the readers can never control the objectivity of the events.

Causality and credibility

In our first paper (Bonaccorsi et al. 2021a) we argue that research impact reports do include propositions that have a causal value. We defend this argument in two ways: making reference to the philosophical debate on the nature of historical knowledge, and by applying to impact statements a new text mining technique that makes it possible to identify the semantic structure of complex text structures, such as entire sentences, called Semantic Hypergraphs. By applying this technique to the collection of REF impact case studies we achieve some interesting results.
If we agree that research impact statements are historical statements, then we can move forward and ask some questions that have been the object of a passionate debate in the '50s and '60s, which has been renewed in recent times. The questions can be formulated as follows: Do historical narratives include explanatory statements? Can historical explanation be considered a scientific explanation? If the answer is positive, under which formal conditions do historical statements have explanatory value? If the answer is negative, how can historians aim at scientific objectivity?
The debate was sparked by the celebrated article by Carl Hempel. The argument was sharp: for a statement to be explanatory, it must include reference to a law-like proposition and a specific, contextualized event, given some boundary conditions. The explanation is the logical process by which we demonstrate that the specific event (explanandum) is logically entailed by the general law (explanans), given the boundary conditions. We perform this task in a logical way, by subsuming the individual event into a general category, following a nomological-deductive approach.
This formulation had an enormous impact. We might say it is, explicitly or implicitly, at the root of all arguments that sharply separate between hard science and humanities, negating the scientific value of the latter. Interestingly, it has been rejected with a number of strong arguments by philosophers and historians. We summarize these arguments here, before showing why they are relevant to the debate on research impact.
First of all, the nomological-deductive model of explanation is not the only possible model of causality. It is a model of necessary and sufficient conditions, which requires for any explanation the existence of general laws that treat the specific event as a member of a class. In other words, it requires the existence of a class of entities. The class of entities is demonstrated by repetition, that is, by producing experimentally manipulated pieces of evidence that reproduce with regularity the same outcomes.
This is not the only logically valid type of explanation. Another logically valid type is a statement which claims that a given condition X has been partially sufficient for Y. This statement would not treat X as a case in a class of entities that can be reproduced. By reasoning

on X historians can build up a collection of partial explanations which together produce a sufficient reason for the manifestation of Y. Each of them has local and circumstantial value.
Second, historians do make use of general laws in their formulations. As stated by Nicholas Rescher, historians are consumers of general laws, not producers. They do not ignore the physical or chemical general laws that dominate the working of nature and they clearly see their application in specific cases. It is rare, however, that the implication of a general law is the crucial element to be invoked to formulate a historical judgment. As the historian Marc Bloch stated, it is certainly true that it is the law of gravity that governs the fact that King X fell from his horse and died. But it is not so interesting to know. Other less general explanatory factors, such as the speed at which he was riding or the threat he was addressing are more important. As the philosopher noted, such general laws very often do not exist, or are trivial.
Third, the validity of historical statements is not predicated on the logical strength of a deductive reasoning, but on the completeness and accuracy of historical documentation, and the plausibility of the narrative reconstruction of causal linkages. It is certainly true that it is not possible to validate experimentally all causal linkages between events. However, historians collect all possible causal explanations (including those that are generated by true general laws) and select among them those that can create a plausible chain of events. In doing so they make appeal to the largest available documentation and to patterns of reasoning that are not deductive but abductive, following the "method of clues" illustrated by Carlo Ginzburg.

On the causality value of research impact statements

From this purely theoretical discussion we move towards another question: if research impact statements have historical nature, what kind of explanation do they include?
Here it is important to call into the debate the influential line of thinking that has proposed the notion of contribution, as opposed to attribution, as the logical foundation for impact assessment. According to this argument, it is impossible to control for all potentially influencing factors that may lead to a research impact. The social impact of research takes place within complex and multidimensional processes that extended over long, often unpredictable, time horizons.
Consequently, the notion of attribution, or the process by which a specific event may be logically demonstrated to be dependent upon a specific condition, must be rejected, in favour of a weaker notion of contribution. We agree with this argument with a qualification. It is one thing to reject the notion of attribution if we assume that the only causal model underlying the attribution is the nomological-deductive causality suggested by Hempel, or variants of this model that fit the way in which causality is assumed in hard sciences. It is another thing to suggest that impact statements do not include any kind of causality. We show that they indeed include causal statements, but these statements must be understood in the light of a theory of historical, not nomological, causality.
In order to examine this issue we adopt a recently developed technique in text mining called Semantic Hypergraphs (Menezes and Roth, 2019). This technique overcomes one of the most important limitation of Natural Language Processing techniques, that is, the ability to examine individual word, or short sequences of words (n-grams), due to computational limits of the algorithms. With Semantic Hypergraphs the meaning of the text is not reconstructed via the statistical analysis of frequency and clustering of words, but by the construction of higher-level topological structures in which the meaning of words is derived from their relational position with all other words. The unit of analysis is a sentence, that is a chunk of text included between two periods. The authors develop a systematic language that allows the automatic processing of sentences. We have applied this technique to the REF collection of impact case studies, following the flowchart in Figure 1 (source: Bonaccorsi et al. 2021a).

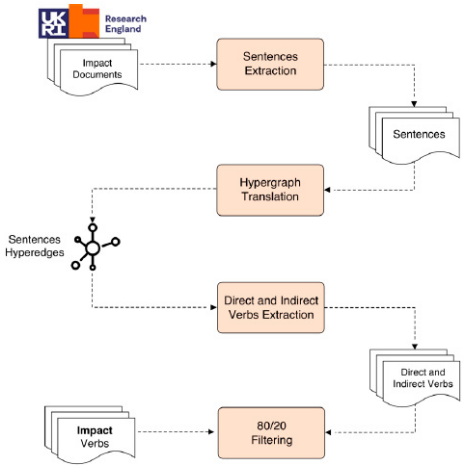

Figure 1. Workflow for the extraction of impact verbs from the REF corpus.

By looking at the structure of sentences we have identified the following elements

i) Direct impact verb
ii) Indirect impact verb
iii) Agent
iv) Topic

The technique allows a clear distinction between impact verbs and non-impact verbs. Impact verbs imply an action of something over something else. This structure is systematically discovered in sentences included in the REF reports. Impact verbs can be direct or indirect, depending on their position in the sentence (technically speaking, the rank in the depth of sentences). On the basis of this distinction and of statistical definitions, it is possible to clearly identify impact sentences and non-impact sentences. At this point we have a powerful tool to ask questions about the distribution of impact vs non-impact sentences across the collection of REF documents. We build up an indicator (IR) as the ratio between impact and non-impact sentences. We find a number of interesting insights:

- Impact sentences have a clear causal structure: they try to demonstrate that X has produced a change in Y
- To achieve the goal of showing the causal effect there is a preparation, mostly of descriptive type, that is done by means of non-impact sentences
- Reports in SSH make larger use of non-impact sentences, meaning that they need to establish a larger and more detailed descriptive evidence before claiming for impact
- The verbal structure of reports, that is, the pattern of utilization of direct and indirect verbs, is not different between SSH and STEM

We interpret these findings as suggesting that research impact statements do have a causal structure, with no difference between SSH and STEM. At the same time we find an interesting difference, in the sense that the construction of the causal statements is longer, more articulated, more complex in SSH reports.

We deepen this issue by introducing the notion of credibility of historical statements in our second paper (Bonaccorsi et al. 2021b). While for professional historians we may assume that rigorous methodological standards are adopted, as it is clear from the methodological and epistemological debate discussed above, the same cannot be said for REF reports. They are written by researchers, or university administrators, or consultants. They have inevitably a rhetorical and instrumental value: the must persuade the evaluators, call their attention to the importance of the impact, and obtain a high score. Remember that within the framework of REF, no less than 20% of funding is allocated on the basis of the impact assessment.

Nevertheless, the authors of the REF reports clearly understand that all these instrumental and practical goals cannot be achieved at all if the statements are not credible. Since they are placed in the past, the readers do not have any experimental control on them. But they mentally reason about the plausibility of the narrative reconstruction. By knowing the pragmatic orientation of the authors, the readers must challenge the credibility of the arguments.
We examine this issue by applying to the causal argumentation the criteria for historical explanation proposed by the philosopher Carl Hammer (2008). According to this philosopher, historical statements have causal power if they include "partial sufficient conditions that are normative, identifiable, manipulable, and not easily replaceable" (Hammer, 2008, 198). We refer the readers to our paper for a full scale, but non-technical discussion of these criteria. By applying these criteria to the main areas of SSH research (social sciences, humanities) and STEM (medicine, technology) and to their main fields of social impact, we derive a kind of theory of credibility of research impact statements. According to this theory, the aim of credibility is more difficult in SSH than in STEM. This is because STEM causality statements may make reference to an established repository of highly structured and formalized processes, in some cases standardized in the public regulation or business practice. During the pathways of impact, the achievement of causal effects is witnessed by the production of formal and socially identified intermediate outputs (e.g., clinical trial, patent, prototype). The description of these chains by the authors of REF reports activates in the mind of evaluators a pattern of recognizability and familiarity. The causal linkages suggested at each of the junctures of the pathway are highly credible. This is not the case for SSH. Here the chain is longer and more fragile. We elaborate on this issue in the final sections of this paper.

Mapping users of research

In two other papers (Bonaccorsi et al. 2020; 2021c) we make use of another technique from machine learning, i.e., Named Entity Recognition. We identify all names in REF documents that refer to social groups, or groups that may be the intended or unintended, direct or indirect beneficiaries of the research of universities. These groups are mentioned in a variety of ways in REF reports. We use a lexicon-based approach, by filtering the REF texts with a lexicon of users with 76,857 entries, developed after extensive research.
With the help of this lexicon, it is possible to saturate the semantic space of research beneficiaries, identifying all social groups that are implied in the impact process. After the extraction we are able to obtain two applications. The first is a mapping exercise, in the tradition of science maps. We draw the entire map of users of research of UK universities, modulating the granularity of the representation. We then cluster the user groups by using community detection techniques. A very informative map is reproduced in Figure 2 (source: Bonaccorsi et al. 2020). All identified clusters are consistent and well delineated. By modulating the granularity we can zoom further in the representation.
The second application is the construction of indicators. Since our lexicon saturates the semantic space, we are in a position to define indicators with appropriate statistical properties. We define the following indicators:

- Frequency
- Diversity
- Specificity

We run the calculation on the entire REF collection and then separately by discriminating between SSH and STEM. This offers some interesting insights.

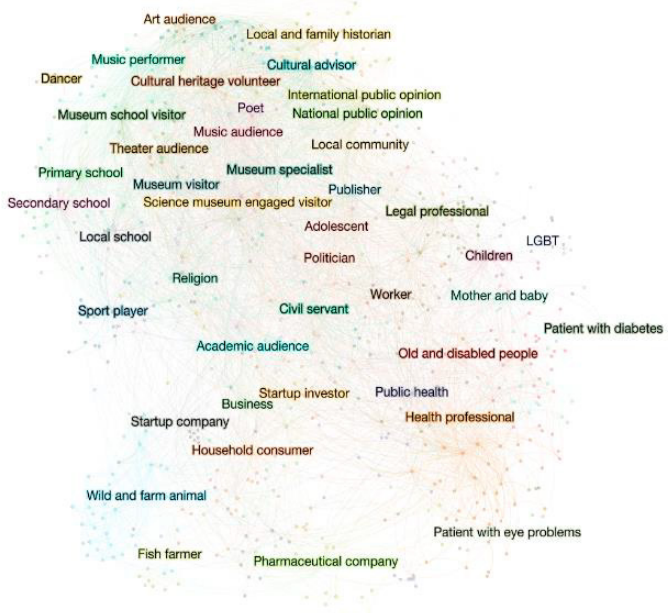

Figure 2. Map of users of research of UK universities.

Lesson learnt and future developments

In the next sub-sections, we develop some implications from this research program and discuss some extensions and new applications.

Towards the construction of indicators of research impact

The construction of indicators in S&T is a complex process, with epistemic, social and political dimensions. With respect to the epistemic dimension, there are several requirements that have been clearly outlined by the literature. Among them, we call the attention on the role of completeness.

When using statistical indicators, researchers and policy makers implicitly rely on the ability of statistical authorities to produce data that correctly refer to the universe under analysis. It is interesting to observe that in the field of S&T indicators this assumption is not warranted. This is why the overall use of data from, say, publications and patents is predicated after a standard methodological caution, of the form "we know that patents do not capture all innovation activities, but.." or "we know that indexed publications do not represent all scientific production, but…". After a while, such clauses are omitted and often forgotten, with the exception of a stream of critical studies that remind us about the limitations of statistical indicators and of simplified input-output models.

The community of S&T indicators is reasonably cautious in accepting the use of data from Machine learning sources, such as text mining. One important reason is that it is not possible to define the reference set and establish its representativeness.

We suggest that this state of affairs can be greatly improved with the use of lexicons. Lexicons are human-made, controllable, improvable cognitive structures. If they are left open to public consultation and correction, they can incorporate corrections and updates, almost in real time. We argue that with well-designed lexicons it is possible to achieve the saturation of a semantic field. If this is true, then there are no conceptual obstacles to the construction of indicators.

If a lexicon saturates a semantic field, then the automatic extraction of whatever semantic structure in a corpus of text will deliver consistent results. They might differ depending on the specific algorithm, but the difference can be clearly explained and made transparent.

In previous research we have developed a lexicon that includes all social groups that have been hitherto considered in all social worlds in which people can be aggregated (e.g. work, profession, health, economic, social and demographic conditions, mobility, hobby, sport, entertainment, culture and education, crime), combining hundreds of published sources. We have no formal demonstration of completeness. At the same time we ask- do we have completeness, say, in patents as indicators of technology, or indexed publications as indicators of science? Certainly not, but we know reasonably well the limitations of these indicators, so that we can safely use them for a number of relevant applications.
We suggest that lexicons can do the same job for us. If we saturate the semantic field in terms of textual descriptions of relevant phenomena, then we can start to build up indicators that have proper statistical properties.

Research impact assessment in SSH

Another area in which we see clear implications from our approach is the hot debate on research impact assessment in SSH. It has been repeatedly and convincingly argued that asking SSH research to demonstrate impact in the same way as it is done for STEM is a serious mistake, with potentially far reaching implications. In particular, it might be used to justify a differential treatment of SSH and STEM in the funding of research. Under the pressure of government demand for impact and demonstrable results, decision makers and funding agencies may find it safer to maintain support for STEM research and to cut, or postpone, the support to SSH. After all, it is often said, researchers in SSH work for long time horizons and do not compete for discoveries. This means that if funds for SSH are delayed or reduced, there is no risk for the national scientific competitiveness.
With our studies we establish the following points:

- Researchers in SSH are able to identify their audiences and talk extensively (i.e. frequently) and intensively (i.e. with a variety of names) about social groups that may benefit from their research. Hence it is not true that SSH research underestimate the importance of addressing social groups as target for the impact.
- At the same time, they find more difficult to identify specifically and in a granular way their target groups, given the generality of their research.
- In order to claim that their research has produced an impact they use the same semantic structure than STEM (impact sentences with direct and indirect impact verbs + agent + mode) and the same set of verbs, implying an effort to build up causal statements.
- They however make use of a larger share of non-impact sentences in the articulation of their impact reports, given a stronger need to introduce the complex social context for the impact.
- They build longer causal chains.
- They make use of a larger number of agents.

From these quantitative findings we obtain a relatively clear picture of the differences between the ways in which research produces an impact in STEM and SSH. In order to build up a credible reconstruction of the historical pathway that has led to the impact, researchers in SSH must invoke many more agents (two times than in STEM) and describe a longer chain of causal linkages. Now from the theory of historical explanation suggested by Hammer (2008) we derive the implication that longer chains and chains with more agents are *more fragile*. This means that it is more difficult to satisfy the requirements of normativity, identifiability, manipulability and non-replaceability that are needed to establish causality in historical statements. At each juncture of the causal chain it is more difficult to credibly argue that X has indeed been a partial

sufficient condition for Y, because there might be many other factors at place or many independent agents whose actions and motivations may not well known.
Given this pattern, we advocate an approach to research impact assessment that makes full justice for these differences.

Impact mapping

Finally, we suggest a new area of application of text mining methods, i.e. mapping research institutions from the perspective of research users. This area would be complementary to the well established area of science mapping, in which institutions are represented in 2D maps on the basis of the disciplines they cultivate, or the topics that their research are addressing. We give an example below, by showing the user maps of two UK universities.

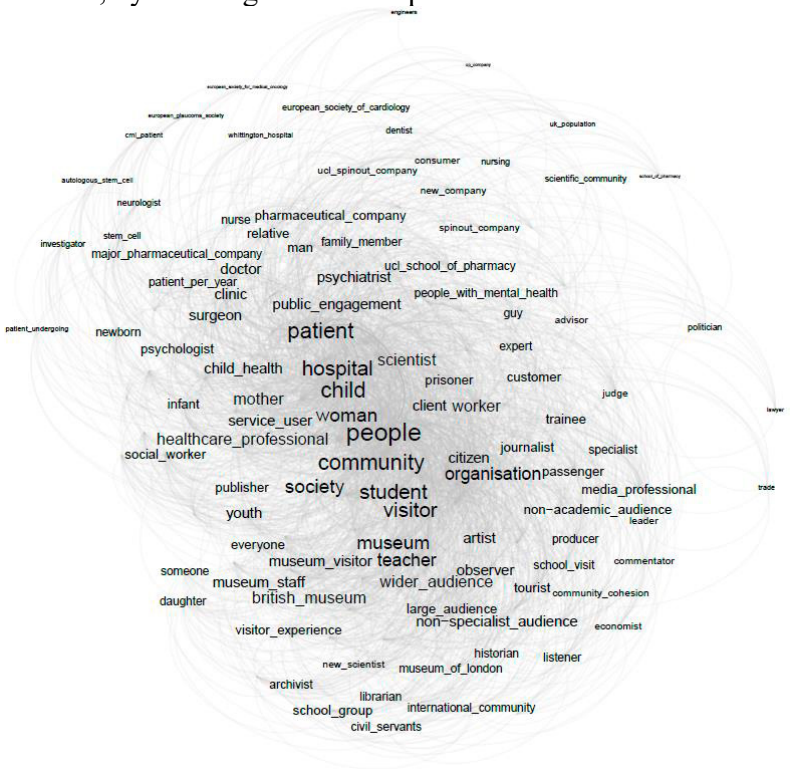

Figure 3. Map of user groups of research of University College, London.

In Figure 3 we see the map of UCL, a top research-intensive university with an international visibility. It can be observed a strong orientation towards user groups in the health sector (patient, hospital, child, woman, mother, healthcare professional community), as well an orientation towards users of cultural heritage (museum, museum visitor, student, visitor, British Museum), while on the contrary the business audience does not seem prominent.

Figure 4 shows the same map for the University of Sheffield. It seems that the main orientations here are largely different: one is towards a local audience (local school, local community, person, community, young people), another is directed towards business actors (organization, company, leader, customer, manager). Much less prominent, contrary to UCL, are the health sectors and the cultural heritage.

There will be a need to refine the analysis and perhaps to develop quantitative indicators after controlling for the robustness of classification obtains by clustering the names of user groups.

We hope there will be opportunities in the near future to extend and refine the methodologies, with an aim to improve the theoretical foundations and the methodological sophistication of research impact assessment.

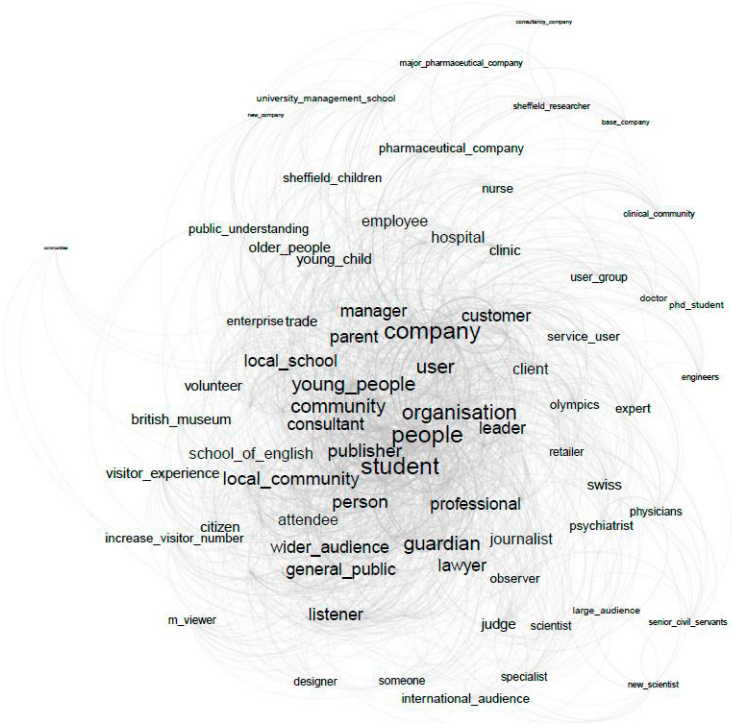

Figure 4. Map of users of research of University of Sheffield.

References

Angrist, J.D. & Pischke, J.S. (2015). *Mastering metrics. The path from cause to effect*. Princeton, Princeton University Press.

Bai, S., Zhang, F. & Torr, P. (2020). Hypergraph Convolution and Hypergraph Attention. arXiv:1901.08150.

Benneworth, P. (2015). Putting impact into context: The Janus face of the public value of Arts and Humanities research. *Arts and Humanities in Higher Education*, 14(1), 3–8.

Bertin, M., Atanassova, I., Sugimoto, C. R. & Larivière, V. (2016). The linguistic patterns and rhetorical structure of citation context: an approach using n-grams. *Scientometrics*, 109(3), 1417-1434.

Bonaccorsi A., Melluso, N., Chiarello, F. & Fantoni, A. (2021a). The semantic of research impact. Examining societal impact of research with Semantic Hypergraphs. Submitted for publication.

Bonaccorsi, A., Melluso, N., Chiarello, F. & Fantoni, G. (2021b). The credibility of research impact statements: A new analysis of REF with Semantic Hypergraphs. Science and Public Policy. https://doi.org/10.1093/scipol/scab008

Bonaccorsi, A., Chiarello, F. & Fantoni, G. (2021c). SSH researchers make an impact differently. Looking at public research from the perspective of users. *Research Evaluation*, 2021, 1-21, doi: 10.1093/reseval/rvab008.

Bonaccorsi, A., Chiarello, F. & Fantoni, G. (2020). Impact for whom? Mapping the users of public research with lexicon-based text mining. *Scientometrics*, https://doi.org/10.1007/s11192-020-03803-z

Bornmann, L. (2013). What is societal impact of research and how can it be assessed? A literature survey. *Journal of the Association for Information Science and Technology,* 64(2), 217-233.

Bornmann, L. & Haunschild, R. (2017). Does evaluative scientometrics lose its main focus on scientific quality by the new orientation towards societal impact? *Scientometrics*, 110(2), 937-943.

Bornmann, L. & Marx, W. (2014). How should the societal impact of research be generated and measured? A proposal for a simple and practicable approach to allow interdisciplinary comparisons. *Scientometrics*, 98(1), 211-219.
Colinet L., Joly P-B., Gaunand A., Matt M., Larédo P. & Lemarié S. (2014). ASIRPA. Analyse des Impact de la Recherche Publique Agronomique. Rapport final. Rapport préparé pour l'Inra. Paris, France.
Collier, A. (2005). Philosophy and critical realism. In G.Steinmetz (ed.) *The politics of method in the Human Sciences. Positivism and its epistemological others.* Durham, Duke University Press.
De Jong, S. P., Van Arensbergen, P., Daemen, F., Van Der Meulen, B. & van den Besselaar, P. (2011). Evaluation of research in context: an approach and two cases. *Research Evaluation*, 20(1), 61-72.
De Jong, S., Barker, K., Cox, D., Sveinsdottir, T. &van den Besselaar, P. (2014). Understanding societal impact through productive interactions: ICT research as a case. *Research Evaluation*, 23(2), 89-102.
Derrick, G. E. (2014). Intentions and strategies for evaluating the societal impact of research. Insights from REF 2014 evaluators. In: P. Wouters (Ed.), *Proceedings of the science and technology indicators conference 2014 Leiden "Context Counts: Pathways to Master Big and Little Data"* (pp. 136-144). Leider, the Netherlands: University of Leiden.
Derrick, G. E., Meijer, I., van Wijk, E. (2014). Unwrapping "impact" for evaluation: A co-word analysis of the UK REF2014 policy documents using VOSviewer. In: P. Wouters (Ed.), *Proceedings of the science and technology indicators conference 2014 Leiden "Context Counts: Pathways to Master Big and Little Data"* (pp. 145-154). Leider, the Netherlands: University of Leiden.
Devlin, J., Chang, M., Lee, K. & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv:1810.04805.
Digital Science (2015). *REF 2014 Impact Case studies and the BBSRC*. Available at www.bbsrc.ac.uk/documents/1507-ref-impact-case-studies-pdf/. Accessed December 3, 2019.
Digital Science (2016). *The societal and economic impacts of academic research. International perspectives on good practice and managing evidence.* Digital Research Reports, March.
Donovan, C. (2011). State of the art in assessing research impact: introduction to a special issue. *Research Evaluation,* 20(3), 175-179.
Ernø-Kjølhede, E. & Hansson, F. (2011). Measuring research performance during a changing relationship between science and society. *Research Evaluation*, 20(2), 131-143.
Gibson, A. G. & Hazelkorn, E. (2017). Arts and humanities research, redefining public benefit, and research prioritization in Ireland. *Research Evaluation*, 26(3), 199-210.
Godfrey-Smith, P. (2003). *Theory and reality. An introduction to the philosophy of science.* Chicago, University of Chicago Press.
Green, N. (2018). Towards mining scientific discourse using argumentation schemes. *Argument & Computation,* 9(2), 121-135.
Green, N.L: (2020). Recognizing rhetoric in science policy arguments. *Argument & Computation*, 1-12.
Heffernan, K. & Teufel, S. (2018). Identifying problems and solutions in scientific text. *Scientometrics*, 116, 1367-1382.
Honnibal, M. Montani, I., Van Landeghem, S. & Boyd, A. (2020). spaCy: Industrial-strength Natural Language Processing. 10.5281/zenodo.1212303
Joly, P.B., Gaunand, A., Colinet, L., Larédo, P., Lemarié, S. & Matt, M. (2015). ASIRPA: A comprehensive theory-based approach to assessing the societal impacts of a research organization. *Research Evaluation,* 24, 440-453.
Kaufmann M., van Kreveld M. & Speckmann B. (2009). Subdivision Drawings of Hypergraphs. In: *Graph Drawing*, Vol. 5417, pages 396–407. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009. ISBN 978-3-642-00218-2 978-3-642-00219-9.
King's College, Digital Science (2015). *The nature, scale and beneficiaries of research impact. An initial analysis of REF (2014) impact case studies.* Research Report 2015/01. London, HEFCE.
Lawrence, J. & Reed, C. (2019). Argument mining: a survey. *Computational Linguistics*, 45(4), 765-818.

Martin, B. R. (2011). The Research Excellence Framework and the 'impact agenda': are we creating a Frankenstein monster? *Research Evaluation*, 20(3), 247-254.
Matt, M., Gaunand, A., Joly, P.B. & Colinet, L. (2017) Opening the black box of impact. Ideal-type impact pathways in a public agricultural research organization. *Research Policy*, 46, 207-218.
Menezes, T. & Roth, C. (2019). Semantic Hypergraphs. arXiv:1908.10784.
Mihalcea, R. & Tarau, P. (2004). TextRank: Bringing Order into Texts. *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, 404-411.
Miettinen R., Tuunainen J. & Esko T. (2015) Epistemological, artefactual and interactional. Institutional foundations of social impact of academic research. *Minerva*, 53, 257-277.
Mikolov, T., Corrado, G., Chen, K. & Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv:1301.3781
Molas-Gallart, J. & Tang, P. (2011). Tracing 'productive interactions' to identify social impacts: an example from the social sciences. *Research Evaluation*, 20(3), 219-226.
Morton, S. (2015). Progressing research impact assessment: A 'contributions' approach. *Research Evaluation,* 24(4), 405-419.
Olmos- Peñuela, J., Benneworth, P. & Castro-Martinez, E. (2014). Are 'STEM from Mars and SSH from Venus'? Challenging disciplonary stereotypes of research's social value. *Science and Public Policy,* 41, 384-400.
Olmos- Peñuela, J., Benneworth, P. & Castro-Martinez, E. (2015). Are sciences essential and humanities elective? Disentangling competing claims for humanities' research public value. *Arts and Humanities in Higher Education,* 14(1), 61-78.
Pearl, J. & Mackenzie, D. (2018). *The book of why. The new science of cause and effect.* New York, Basic Books.
Penfield, T., Baker, M.J., Scable, R. & Wykes, M.C. (2014) Assessment, evaluations, and definitions of research impact. A review. *Research Evaluation,* 23 (1), 21-32.
Samuel, G. N. & Derrick, G. E. (2015). Societal impact evaluation: Exploring evaluator perceptions of the characterization of impact under the REF2014. *Research Evaluation*, 24(3), 229-241.
Spaapen, J. & Van Drooge, L. (2011). Introducing 'productive interactions' in social impact assessment. *Research Evaluation,* 20(3), 211-218.
Teufel, S. & Moens, M. (2002). Summarising scientific articles --- Experiments with relevance and rhetorical status. *Computational Linguistics* 28(4), 409-446.
Woodward, J. (2003). *Making things happen. A theory of causal explanation*. Oxford, Oxford University Press.