Forthcoming in *Quality & Quantity*

MEASURING INNOVATION AND INNOVATIVENESS: A data-mining approach[1]

Bernard Sinclair-Desgagné[2]

Skema Business School –
Université Côte d'Azur (GREDEG)

ABSTRACT: Despite substantial advances over the past decades, measuring innovation and innovativeness remains a challenge for both academic researchers and management practitioners. To address several key concerns with current indicators – such as their specialization and consequent one-sidedness, their frequent lack of theoretical foundations, and the fact that they may not really foster creativity and invention – this paper introduces some new metrics via one data-mining approach – Formal Concept Analysis (FCA) – which is increasingly used to represent and treat knowledge. This approach can adapt to particular needs and goals, incorporate various kinds of information (qualitative or quantitative) from different sources, and cope with several types of innovations. It also uncovers a logical route to novelty, which might enhance the generation of ideas and is used here to support the measurement of innovativeness.

KEYWORDS: Knowledge representation and measurement; Innovation indicators; Formal Concept Analysis (FCA); Out-of-the-box thinking; Logic of invention

---

[1] **Acknowledgements:** Constructive comments from participants at the MOSAIC-HEC Montréal seminar, the CORE-UC Louvain research seminar, the 2017 Schumpeter Conference, and the 2019 KTO Workshop at Skema Business School are gratefully acknowledged. I owe special thanks to Marine Agogué, Thierry Bréchet, Patrick Cohendet, Michel Desmarais, Ludovic Dibiaggio, Soumitra Dutta, Armand Hatchuel, Pengfei Li, Paavo Ritala, Henry Tulkens, Gilda Villaran, and Isabelle Walsh for valuable conversations, remarks or suggestions. Two diligent and competent anonymous referees helped to improve significantly the content and presentation of this paper. All remaining errors and shortcomings are my own.

[2] **Corresponding author:** Bernard Sinclair-Desgagné, Skema Business School, 60 rue Dostoïevski – CS 30085, 06902 Sophia Antipolis, France. Ph: 33 4 93 95 44 85. E-mail: bsd@skema.edu

"Measuring innovation - how much is taking place and the value it's producing - remains a challenge for many companies."

- *The Wall Street Journal*, August 8th 2016 -

1. Introduction

Innovation is widely seen as key to economic growth, sustainable development, productivity improvement, firm competitiveness, and social cohesion. Governments and businesses are therefore putting considerable resources into fostering innovation. Since well-informed and effective policies must hinge on sound indicators and measurement, academic researchers, policy makers, business consultants, innovation scholars, engineers and managers have accordingly developed several means to measure innovation and innovativeness.[3] Despite some substantial advances, though, measuring innovation or innovativeness still remains a challenge for most government agencies, private businesses and social communities.[4]

At least two sets of reasons might explain this situation.

First, there is now a plethora of indicators to choose from. Some indicators – e.g., patent counts, ratio of sales of new products to total sales, invention disclosures – focus on the output of innovation efforts; others – e.g., R&D spending, information flows, use of instruments and tools – consider inputs. Some are 'hard' quantitative figures; others – e.g., managerial attitude, clarity of vision/mission – are based on 'soft', subjective, oftentimes qualitative, survey-gathered views and perceptions. Some – e.g., number of new product ideas – are direct; others – e.g., R&D budget – are partial and indirect. As repeatedly pointed out in reviews and surveys, all these measures have advantages and shortcomings. Patents (per se and through their citations) deliver valuable information on innovation efforts and trends (Oh et al. 2017), but they constitute only one type of value-appropriation mechanism (Zobel et al. 2017), they are often not converted into commercial products, and they can be irrelevant in certain industries (like computer software). R&D expenses show a commitment to innovate, but they do not necessarily reflect efficiency in doing so, and many firms (especially small ones) currently

[3] Good reviews and assessments of the academic and policy literatures include Adams et al. (2006), van Beers et al. (2015), Brattström et al. (2018), Chan et al. (2008), Dziallas and Blind (2019), Gamal et al. (2011), Garcia and Calantone (2002), Gault (2013, 2018), Goldsmith and Foxall (2003), Gupta (2009), Hall and Jaffe (2018), von Hippel (2017), Kleinknecht et al. (2002), Lhuillery et al. (2017), OECD (2010), Saunila (2017), Smith (2005).

[4] See, e.g., Dziallas and Blind (2019), Dewangan and Godse (2014), Cruz-Cázares et al. (2013), Adams et al. (2006), and the industry surveys conducted by McKinsey and the Boston Consulting Group.

innovate without doing R&D. Many indicators, moreover, might be useful for academic research or public policy, but their application to investment decisions or business strategy is not straightforward (Adams et al. 2006). In such a setting, the matter of choosing appropriate indicators that would support an entity's innovation strategy has become a rather complicated task (Kleinknecht et al. 2002; Richtnér et al. 2017).[5]

Secondly, as Brattström et al. (2018) warned about, there can be unintended consequences to implementing some innovation metrics in an organization: this might in fact channel people's attention in ways detrimental to innovativeness. To support creativity and invention, a measurement approach – using these authors' language – should be 'directional' (so be used "as a mechanism for control and direction") *as well as* 'conversational' (hence providing "organizational members with a stable, yet adaptable, language and frame of reference"). To be sure, not all current measurements would satisfy such provisos. Based on a survey of 169 companies, Raedersdorf Bollinger (2019) reports, for instance, that the definition of indicators which encourage creativity and facilitate innovation management faces difficulties in over 40% of the firms.

This paper's primary objective is to bring in a data-mining approach to measuring innovation and innovativeness that might help overcome these current hurdles.

The Oslo Manual, produced by the OECD (1997) to set a benchmark for empirical studies and policies, proposes the following definition of innovation.

> A technological product (good or service) innovation can involve either a new or improved product whose characteristics differ significantly from previous products.[6]

Examples of product innovation might include a new product's invention; technical specification and quality improvements made to a product; or the inclusion of new

[5] What complicates the matter further is that organizations often disagree on what should be measured (see Dziallas and Blind 2019, and the references therein), with the result that innovative project managers and finance department might often not use the same tools (Stefani et al. 2019). Addressing this important issue – which has to do with organizational design and incentives – is unfortunately beyond the scope of this paper.

[6] This definition fits the purpose of this paper. It is consistent with Joseph Schumpeter's initial definition of product innovation (OECD 1997, p. 28). The literature offers several alterrnate definitions, however. For an overview and useful discussion of these, the reader may look at Rogers (1998), Baregheh et al. (2009), and Quintane et al. (2011),

components, materials or desirable functions into an existing product. Quite naturally, this definition refers to some 'objects' (literally, here, goods and services, but one might extend it as well to technologies or processes) and their 'attributes' or characteristics (physical and intangible properties, technical specifications, quality, functions, and market value). To innovate then means having an existing desirable attribute newly associated with a current object, a novel attribute ascribed to an already known object, or a newly discovered object exhibiting valuable attributes (some known already, some not).

As it turns out, many of the methods used in data-mining can deal with this sort of knowledge, which they represent in tabular form – the rows corresponding to objects and the columns to their attributes. Central among these methods is *Formal Concept Analysis* (FCA), a mathematical field at the interface of lattice theory, machine learning and computer science, which recently entered the social sciences and is increasingly employed in industry.[7]

Building on this approach, this paper will introduce some new metrics which, in principle, can: (i) bring together and exploit various kinds of data, such as patent and survey data, objective and subjective figures, quantitative and qualitative information, as well as output, input, process, financial and environmental/social considerations; (ii) adjust to the needs, goals and priorities of a country (advanced or emerging), region, industrial sector, firm, or entrepreneur; (iii) reveal innovation patterns; and (iv) provide some rigorous guidance for how to think more effectively out of the box. These metrics are straightforward and user-friendly, amounting to manipulating spreadsheets. Yet, they lie on rigorous theoretical foundations. And they should meet many of the desiderata expressed by practitioners, such as pursuing continuous improvement in measurement (Ritchnér et al. 2017), not listening only to the voice

---

[7] Formal Concept Analysis (FCA) began with Rudolf Wille (1982)'s seminal article. Extensive introductions can be found in Ganter and Wille (1996), Davey and Priestley (2002), Bělohlávek (2008), and Ignatov (2015). Over the past decades, numerous applications have been found, including "(...) hierarchical organization of web search results into concepts based on common topics, gene expression data analysis, information retrieval, analysis and understanding of software code, debugging, data mining and design of software engineering, internet applications including analysis and organization of documents and e-mail collections, annotated taxonomies, (…)." (Bělohlávek 2008, p. 4) More recently, Poelmans et al. (2013ab) describe applications in linguistics, bioinformatics, and medicine, while Gardiner and Gillet (2015) cover many real and potential ones in chemistry. Specific uses in mechanical engineering, manufacturing and pharmaceuticals are outlined in Nanda et al. (2007), Tóth et al. (2014), and Quintero and Restrepo (2017), respectively.

of consumers but also to the voice of your product (Goldenberg et al. 2003), and seeking management tools which allow to integrate the different views of the stakeholders (Raedersdorf Bollinger 2019). As we move on, their development will be illustrated using a concrete example drawn from environmental management.

The rest of the paper unfolds as follows. The upcoming section conveys the basics of knowledge representation via FCA. Section 3 next defines an innovation metric and discusses its properties. Section 4 brings out some new ramifications of FCA for creative thinking and the search for novelty. The upshot – and chief contribution of this paper – is a precise 'logic of discovery' (Simon 1973) which can support out-of-the-box thinking and the generation of ideas. On this foundation, Section 5 deals with the appraisal of innovativeness. Section 6 closes with concluding remarks and suggestions for future research. Throughout the paper, the presentation is meant to be self-contained and targeted at researchers as well as practitioners and teachers; readers eager to see the mathematical background can find it in the Appendix.

2. Knowledge representation

One major industrial sector in which technological and product innovation is widely on demand nowadays is the environmental goods and services industry (Sinclair-Desgagné 2017). A traditional and quickly expanding segment in this sector is the one offering groundwater remediation services. Table 1 shows six available techniques that providers of such services can use – air sparging, pump-and-treat, bioslurping, ultraviolet oxydation, biosparging, and natural attenuation – together with some relevant characteristics these techniques might have – the contaminants they can treat (VOC - volatile organic compounds, heavy hydrocarbons, insecticides, inorganic pollutants, and heavy metals), the depth at which they can be deployed (all of them work at depths of 30 feet or more), their cost range ($ - cheap, $$ - moderately expensive, or $$$ - expensive), their effectiveness (L - low, M - medium or H - high), and their duration (Fast, Average or Long). The symbol 'X' which appears in some cells of the table indicates that the corresponding 'row' technique possesses the associated 'column' characteristic: pump-and-treat, for example, can cope with insecticides; bioslurping cannot.

Insert Table 1 about here.

In the language of FCA, Table 1 constitutes what is called a *formal context*. In general terms, a formal context is denoted as a triplet K = (G,M;R), where G refers to a set of objects, M is a set of relevant attributes these objects may have, and R (the set of X-filled cells in the table) is a subset of the Cartesian product G×M such that (g,m) belongs to R, noted gRm, if object g possesses attribute m.[8]

This way of representing knowledge in matrix form encompasses quite a few cases. The objects of a formal context can be goods or technologies owned by a firm or overseen by a public policy body; the attributes can be important necessary inputs or technological specifications, features of the delivery process, characteristics that consumers or other stakeholders value, cost or price brackets, and so on. All this information can be objective or subjective, qualitative or quantitative. For sake of clarity throughout this paper, we will consider only the simplest (yet rather common) type of context – the binary context – in which an object either has an attribute or does not. More sophisticated kinds of context, dealing for instance with fuzzy or (what is rather key for market-oriented innovation) value-weighted attributes, have been introduced in the FCA literature and are applied in industry (Valverde-Albacete et al. 2016; Gardiner and Gillet 2015; Bělohlávek and Macko 2011). They are briefly brought up and discussed in this paper's concluding section.

The aim of FCA is to organize the information contained in a context. It does so by forming what can be seen as clusters of objects and attributes, which are called *formal concepts*. The following pair drawn from Table 1 is an example of a formal concept:

({Air sparging, Pump-and-treat, Natural attenuation} , {VOC, Medium effectiveness})

Notice that the objects {Air sparging, Pump-and-treat, Natural attenuation}, and only these ones, have the attributes {VOC, Medium effectiveness} in common. The way FCA defines a formal concept actually agrees with the International Standard Organization's ISO-704 definition: "In a concept, one distinguishes its 'intension' and 'extension'. The intension of a concept comprises all attributes thought with it, the extension comprises all objects for which

[8] Owing to the German origin of FCA, sets of objects and attributes are commonly denoted G and M. These letters stand for 'Gegenstand' and 'Merkmal', the respective German words for 'objects' and 'characteristics'.

the concept can be predicated. In general, the richer the intension of a concept is, the lesser is its extension, and vice versa."

If we were to reshuffle the rows and columns of Table 1 in an appropriate manner, the X-filled cells corresponding to a formal concept would form a maximal rectangle in this table.

3. An innovation index

As innovation takes place, knowledge grows. This gives rise to an expanded table or context. An example is displayed in Table 2. Compared with Table 1, it shows two new techniques – passive/reactive walls and groundwater circulation wells, one new (hypothetical) contaminant or stakeholder request ε, and several new relationships indicated by the small x's.

Insert Table 2 about here.

In general terms, from now on, the triplet K = (G,M;R) will refer to the initial context which precedes innovation, while the triplet $K^+=(G^+,M^+;R^+)$ denotes the augmented context which arises after innovation took place.[9] Clearly, the new set of objects G^+ contains G, the new set of attributes M^+ contains M, and the new set of relationships R^+ embeds R.

A straightforward way to measure innovation in this setting would be to count the new relationships (represented in Table 2 by the small x's), and divide the result by the total number of existing relationships (or the number of nonempty cells in Table 2). Using our mathematical notation, this metric is expressed in general form by the formula

$$\nabla = \frac{|R^+ \setminus R|}{|R^+|}$$

where the symbol \ stands for taking differences between sets – i.e. keeping all the relationships mentioned in R^+ which are not in R – and the two vertical bars |...| enclosing a set refer to computing this set's cardinality (or number of elements). The index's value in the current example is $\nabla = 18/67 \approx 0.27$.

[9] Considering the timing of innovation and its appraisal is out of the scope of this paper. In many cases, though, periodic reviews of innovation policies are already set a priori by contract, policy plans or rules of governance.

This indicator possesses the following properties:

- It gives percentages, since the value of ∇ rests between 0 and 1;
- It is equal to 0 when there is no innovation, i.e. when R^+ is the same as R;
- It is equal to 100% when all relationships are new, i.e. when R^+ is nonempty but R is;
- It grows as more innovation occurs, i.e. as the difference between R^+ and R increases.

The metric ∇ gives the proportion of total ex post knowledge captured by K^+ which is truly new. If one were to relate ex ante and ex post knowledge, trying for instance to track how each original concept in K developed into expanded ones in K^+, then ∇ would provide a summary statistic of these developments.[10]

An apparent caveat is that ∇ seems to weigh every new relationship the same, thereby ignoring an innovation's relative value. This is indeed true in the intendedly simple examples of Tables 1 and 2, but in general a formal context can easily assign (market/commercial or nonmarket/use) values to objects with certain attributes. One simple way is to qualify the relation R^+ (and, similarly, R as well) so that gR^+m means that object g has attribute m *and* 'sufficient' market value; in other words, a cell in the above tables would now be filled only if the corresponding pair meets a certain market test or value threshold.[11] This procedure, which the FCA literature calls 'scaling', has been further refined and made operational (see Ganter and Wille 1996, p. 36, for details).

Other criticisms – drawing notably on Griliches (1986)'s criteria for data quality – might point out that ∇ suffers from incomparability across contexts, as well as being exposed to manipulation or the so-called Hawthorne effect (whereby measurement affects people's behavior). One answer to these is to define the objects and attributes in the consensual terms of an official classification such as the one in EC et al. (2009), or survey such as the Oslo manual (Tanaka et al. 2005). In any case, the index ∇ does fulfill one crucial requirement of measurement: "(...) that different observers should obtain the same results by processing the same data by the same procedure." (Mari 2003, p. 28)

---

[10] Mathematical support for this statement can be found in subsection A.2 of the Appendix.

[11] Alternatively, the relation R^+ could be refined by indicating value ranges, so that a new object bearing certain new attributes but losing other valuable ones would be singled out. I thank one referee for raising this issue.

The upcoming section now investigates how to reach K^+ from K by thinking properly out of the box. This will provide a theoretical basis for the measurement of innovativeness.

4. Thinking out of the box

The challenge of innovation can be viewed as increasing the value of the index ∇ by expanding the existing table or context. A familiar piece of advice for achieving this is to 'think outside the box'. This is more easily said than done, though, for what the stipulated outside area concretely is, and *how* it should be explored, remain rather vague. This section will aim at more precision, by means of the present framework and the logical linkages it allows to establish between what is already known and what is yet to be discovered.

In our example, the additional knowledge that was produced is exhibited in Table 3, which is simply Table 2 with the cells previously filled with a capital X left empty.

Insert Table 3 about here.

This table forms another context. In general, we will designate the context consisting of additional knowledge only as the triplet $K^* = (G^+,M^+;R^*)$, where the relation R^* is precisely the 'difference' $R^+ \setminus R$ between the relations R^+ and R which appears in the numerator of the innovation index ∇. The innovator's challenge can be stated again as seeking to generate K^*.

Table 4 shows all the novel concepts for K^* in our example. Could these concepts be foreseen to some extent, *based solely on what was known initially*? In general terms, can a formal concept in K^* be somehow anticipated based on the original context K?

Insert Table 4 about here.

Systematic searches and forecasts often involve probabilistic beliefs. Alternatively, the data-mining and FCA literatures offer other, non-probabilistic, methods (see, e.g., Ganter and Kuznetsov 2000). All these approaches, however, envision an already given, more or less specific, set of possible future outcomes. We want to avoid this common assumption here. While many of the objects and attributes comprised respectively in G^+ and M^+ can belong to the initial sets G and M, we suppose that the overall composition of the former (which may bring in new objects and attributes) is never settled in advance, and that *nothing* whatsoever is known a priori concerning the new relationships featuring R^*.

An intuitive way to then start searching/building the concepts of a context is to look for their possible preconcepts (Denniston et al. 2013). In FCA, a *preconcept* is a list of objects and attributes so that more objects in the context can share the listed attributes, and the given objects can have more attributes in common than the ones listed. A preconcept *anticipates* a concept if the objects and attributes it displays are respectively included in the concept's objects and attributes.

In accordance with our assumption, preconcepts of a concept in K^* should only be made of the initially-known objects and attributes in K. We refer to such preconcepts as *seeds*. In the groundwater remediation services example which has been considered so far, a seed is the pair ({Air sparging},{$$}). One can check that it anticipates in the above sense the novel concept ({Air sparging, Wells},{ε, $$}) drawn from Table 3.

Seeds might be quite useful in fostering innovation.

First, a seed can be seen as a suggestive low-hanging fruit to be picked early on. As Table 4 shows, not all new concepts emanate from a seed; this is the case, for instance, for the concept ({P/R walls, Wells},{VOC, HH Carb, M eff}). But after finding the seed ({Air sparging},{$$}), the new concept ({Air sparging, Wells},{ε, $$}) might have come from searching for additional uses of the now moderately expensive air-sparging technique (which leads to treatment of contaminant ε) and for new similarly-expensive technologies (which leads to groundwater circulation wells).

Second, and most importantly, *seeds can be located ex ante* to some degree. To see this, start with noticing that a seed exhibits some relationships between objects and attributes which did not exist before. These relationships are to be found in the occupied cells of Table 5, which is just the negative picture of Table 1: the filled cells are the ones that were empty in Table 1, and the empty cells are those that were filled in Table 1.

Insert Table 5 about here.

Table 5 constitutes a tangible 'outbox' to launch creative thinking, for it relies on a priori knowledge only.[12] It also defines a context (the fourth and last one to be used in this paper). In general terms, the FCA literature calls the context $\overline{K}$ = (G,M;$\overline{R}$) made of the objects and attributes in K, with $\overline{R}$ saying that an object in G relates to an attribute in M if and only if it does *not* (yet) possess this attribute according to *R*, the *complementary* context of K.

As a context, $\overline{K}$ naturally has its own formal concepts, which we may call the *anticoncepts* of K. The full list of anticoncepts in our example is displayed in Table 6.

Insert Table 6 about here.

Now, somewhat surprisingly, *a seed is not only the preconcept of a yet-to-be-discovered novel concept in* K^**, it is also the preconcept of an anticoncept*. In other words, seeds happen to point, not only at concepts in K^* (which are a priori unknown), but also at concepts in $\overline{K}$ (which can all be obtained from initial data!). This fact suggests a concrete and systematic procedure for thinking effectively out of the box:

- Starting with the initial knowledge represented by the context K, consider the complementary context $\overline{K}$;
- From $\overline{K}$, list the anticoncepts of K;
- Examine the preconcepts of these anticoncepts;
- If the listed objects of a given preconcept end up possessing this preconcept's listed attributes, then a seed has been found, which can develop into a novel concept in K^*.

****

From an epistemological viewpoint, the above course of action, based on seeds seeking and exploitation, might be seen as an instance of *abduction* – a mode of reasoning associated with creativity and the generation of ideas (Simon 1973; Thomé and Crespo 2013; Mabsout 2015; Pietarinen and Belucci 2015; Bruscaglioni 2016). Unlike deduction, which draws the logical ramifications of previously given assertions, or induction, which infers general laws from the observation of recurrent facts, abduction (also called *retroduction*) looks for the best

[12] This outbox might be seen as the 'creative reservoir', or repertoire of creative opportunities, previously identified by Cohendet and Simon (2015).

'justification' after hitting a singular event. Quoting philosopher Charles Peirce's founding work on the subject (from Pierarinen and Bellucci 2015, p. 355):

> By Retroduction I mean that kind of reasoning by which, upon finding ourselves confronted by a state of things that, taken by itself, seems almost or quite incomprehensible, or extremely complicated if not very irregular, or at least surprising, we are led to suppose that perhaps there is, in fact, another definite state of things, because, though we do not perceive any unequivocal evidence of it, nor even a part of it (or independently of such evidence if it does exist), we yet perceive that this supposed state of things would shed a light of reason upon the state of facts which we are confronted (…).

This description matches the above procedure, which says to look for the encompassing concept in K^* once a seed (the 'surprise') has been found.

More recently, Bruscaglioni (2016, p. 2019) distinguishes three types of abduction. In the first one, the justification for an observed singular event (the seed, here) "is given in automatic or semi-automatic way." In the second one, it can be "identified through a selection within the available encyclopaedia." In the third one, which the author associates with truly *creative abduction*, "the richness of knowledge at your disposal is not sufficient to interpret the result, so it is necessary to formulate a new hypothesis." This corresponds precisely to the last step of the procedure outlined in this section, whereby the encounter of a seed (the 'result') must be imputed to a yet-to-be-uncovered new concept (the 'new hypothesis').

The latter interpretation suggests, furthermore, that innovativeness might somewhat actually be linked to the presence and fertility of seeds. This point will now be explored.

5. Innovativeness assessments

The literature gives the term 'innovativeness' at least two different meanings (see, e.g., Goldsmith and Foxall 2003): one refers to a given entity's ability, capability or 'propensity to innovate', another to the 'newness' actually delivered. Both connotations can be separately captured in this paper's framework.

To begin with, let us introduce two indispensable pieces of mathematical notation. For a given set S of objects from the initial context K, let $\gamma(S)$ be the set of attributes (old or new) in the augmented context K^+ which are newly associated with these objects. In Table 2, for instance, γ(pump-and-treat, bioslurping) = ∅, the empty set, but γ(air sparging) = {ε, $$}, where ε is a new attribute and $$ is an old one. Similarly, for a subset of initial attributes T in K, let μ(T) denote the set of old or new objects in K^+ that possess these attributes. In Table 2, for instance, μ(VOC, $$) = {groundwater circulation wells}.[13]

5.1 *Measuring the propensity to innovate*

The previous section put forward that innovation might begin with finding seeds, while seeds rest in anti-concepts. A consistent indication[14] of an entity's propensity to innovate could then be the fertility of the anticoncepts it was initially endowed with.

Let $\Gamma(G,M;\bar{R})$ regroup all the anticoncepts (A,B) of K, with A denoting the objects and B the attributes. The following pair

$$T(\bar{R}) = \sum_{(A,B)\in\Gamma(G,M;\bar{R})} \frac{|\gamma(A)|/|M^+|}{|\Gamma(G,M;\bar{R})|}$$

$$D(\bar{R}) = \sum_{(A,B)\in\Gamma(G,M;\bar{R})} \frac{|\mu(B)|/|G^+|}{|\Gamma(G,M;\bar{R})|}$$

gives the *average* number of attributes T(R) and objects D(R) newly associated with, respectively, the objects and the attributes of an anticoncept. Meanwhile, the pair

$$\tilde{T}(\bar{R}) = \sum_{(A,B)\in\Gamma(G,M;\bar{R})} \frac{(|\gamma(A)|/|M^+| - T(\bar{R}))^2}{|\Gamma(G,M;\bar{R})|}$$

$$\tilde{D}(\bar{R}) = \sum_{(A,B)\in\Gamma(G,M;\bar{R})} \frac{(|\mu(B)|/|G^+| - D(\bar{R}))^2}{|\Gamma(G,M;\bar{R})|}$$

provides the respective variances of such associations. The former measures the yield that can be expected from each anticoncepts. The latter one quantifies the randomness of the harvest.

[13] In the Appendix, γ and μ are suggestively called *innovation mappings*.

[14] Following Mairesse and Mohnen (2002, p. 226): “Innovativeness is conditional on a model of an innovation function (...).”

In our example, $T(\bar{R}) \approx 0.05$, $D(\bar{R}) \approx 0.0655$ and $\tilde{T}(\bar{R}) \approx 0.04$, $\tilde{D}(\bar{R}) \approx 0.047$. These relatively small values signal a consistently low anticoncept fertility.[15] In fact, only two anti-concepts have seeds here: ({Air sparging},{HHC, Insect, Inorg, Heavymet, Low depth, $, $$, L eff, H eff, Avg, Long}) with seed ({Air sparging},{$$}), and ({Nat atten},{Inorg, Heavy met, $$, $$$, L eff, Fast, Avg}) with seed ({Nat atten},{Avg}).

Interestingly, the metrics T(R) and D(R) might also provide another useful piece of information. Using a well-known taxonomy, the function $\gamma(\cdot)$, which confers new attributes to objects, and the function $\mu(T)$, which seeks objects that newly share some attributes, respectively capture what can be interpreted as *'technology-pushed'* and *'demand-pulled'* innovations.[16] This suggests that the ratio $T(\bar{R})/D(\bar{R})$ could measure the relative impact of technology versus demand on innovation. In our example, demand factors appear to dominate since the ratio equals 0.76. Indeed, Table 3 shows that no new attributes have been found for 3 of the 6 initial techniques, while only 5 out of 15 initial attributes were not newly satisfied.

5.2 *Measuring newness*

Intuitively, the word 'newness' evokes an unforeseen event. The term would well apply to novel concepts that could not be anticipated. In the language of this paper, this could mean concepts drawn from the context K^* of incremental knowledge that do not have seeds in the context K of initial knowledge. The proportion of such concepts is formally given by

$$\Delta = \frac{|\ \Gamma(G^+, M^+; R^* \setminus G \times M)\ |}{|\ \Gamma(G^+, M^+; R^*)\ |},$$

where $\Gamma(G^+,M^+;R^*\setminus G\times M)$ refers to the set of all concepts in the context K^* deprived of the previous relationships involving objects and attribute which already belonged to the original context K. This formula might be seen as an index of newness.

---

[15] Some anticoncepts might have been discarded right away, which would make the assessments more accurate. This is the case for those exhibiting an empty set, like ({Airsparg, Bioslurp, UV oxyd, Biosparg, Nat atten}, ∅}.

[16] An exhaustive survey of the literature and debates surrounding this taxonomy can be found, for instance, in Di Stefano et al. (2012).

Considering the concepts listed in Table 4, the incremental knowledge context of the example has newness value $\Delta = 0.75$. Such a relatively high figure was to be expected, in view of the relative scarcity of seeds in this case.

6. Concluding remarks

This paper submitted new measures of innovation and innovativeness. Since innovation measurement appears to be rather infrequent across firms (Adams et al. 2006; Dewangan and Godse 2014; Dziallas and Blind 2019), these metrics are based on an intuitive, yet rigorous, data-mining approach to represent and structure knowledge, namely Formal Concept Analysis (FCA), which is increasingly used by engineers and managers involved in R&D projects (Nanda et al. 2007; Tóth et al. 2014; Quintero and Restrepo 2017). They are in principle widely applicable, sensible and user-friendly, boiling down to using only spreadsheets. The underlying approach should also allow to incorporate various kinds of data – quantitative and qualitative (as stressed notably by Park 2020), objective and subjective, financial and non-financial –, thereby providing business strategists with an integrative tool to cope with the multiple facets of knowledge production and the myriad of innovation types.

Exploiting this framework also generated insights and prescriptions for thinking more effectively outside the box. This meets a major criticism of innovation measurement, that it often hinders creativity and invention (Brattström et al. 2018). It might also fill an important practical gap, as it seems that it is at the stage of exploring and conceptualizing new ideas that managerial tools as most needed (Raedersdorf Bollinger 2019).

Several research avenues can now be taken from here.

First, the metrics proposed in this paper seem to have little to do with the widely used ones prescribed in the Oslo manual (Tanaka el al. 2005) or the Global Innovation Index (Dutta 2012). They focus on old and new objects (in a broad sense) and their characteristics, while the latter emphasize institutions, regulations, education, infrastructures, market conditions, and linkages. They stick to innovation as it happens while the latter highlight the supporting ecosystem. However, many of the categories constituting, for instance, the Global Innovation Index – especially those pertaining to the 'Innovation Input' Sub-Index, like R&D expenses, tertiary education and credit – could well figure as additional attributes in an exhaustive context. Although some necessary adjustments would have to be made (to what the negative or 'outbox' context $\overline{K}$ is, notably), this mixing of perspectives could deliver welcome policy insights and prescriptions.

Second, building on Brattström et al. (2018)'s conceptual framework, further ramifications for attention management of the procedure proposed in Section 4 should be explored. In the same vein, parallels and connections should be established between this procedure and some of the tools currently used in industry to boost inventiveness, such as TRIZ (Ilevbare et al. 2013; Goldenberg et al. 2003), design thinking (Brown 2009), CK-theory (Hatchuel and Weil 2009), ACE and Big Data (Hua Tan and Zhang 2017), and innovation process models more generally (Barros Bagno et al. 2017). As good ideas seem to have become harder and harder to find (Bloom et al. 2017), while coping with climate change and sustainable development raise major challenges, finding practical devices that will concretely enhance creativity and invention should constitute a major pursuit.

Third, in Section 4 we made minimal assumptions about the use of a priori knowledge, ignoring issues of context and timing, and forbidding the use of probabilities. In practice, however, one should draw from the analysis of adaptive evolutionary systems – as Munier (2013) did, for instance, using Viability Theory (Aubin 1990) – and tap on probabilistic beliefs based on science, generalized information theory (Palm 2012), predictive models (Sood et al. 2012), or sound experience, in order to figure out the plausibility of new relationships between objects and attributes. This endeavor will enhance the search for seeds, hence the discovery of new concepts.

Fourth, on a more technical note, to convey the main substantive points without burdening the exposition, this paper stuck to a simple, yet frequent, form of knowledge representation, using discrete attributes and binary relationships. The FCA literature provides more sophisticated representations, though, drawing for example on stochastic Galois lattices (Diday and Emilion 2003), rough sets (Pawlak and Showron 2007), and fuzzy sets (Poelmans et al. 2014). Adapting the above metrics to these refinements might be worthwhile, if not necessary, for *coping with some of the main challenges for innovation studies* (highlighted, for example, in Martin 2016, Edwards-Schacter 2018, and Christensen et al. 2018), such as assessing innovation in services, non-R&D related innovation, non-patented innovation, *disruptive innovation* and 'dark innovation.'

Still on a technical note, finally, listing all the concepts of a formal context is generally burdensome.[17] Yet, the search for seeds and the assessment of innovativeness that we propose require this exercise. Research and development on the issue is very much ongoing.[18] Several algorithms and softwares already exist: many (mentioned in Singh et al. 2016, for instance) are subject to a patent but others – GALICIA and JALABA, for example – can be freely downloaded. Two promising trends are *to take full advantage of negative information* (i.e. the information contained in $\overline{K}$), as in Rodriguez-Jimenez et al. (2016), and *to assign weights to attributes*, as in Belohlavek and Macko (2011). The former could help systematize further the process of thinking out of the box, especially when dealing with very big sets; the latter allows to introduce what can be individual, business or social preferences, and might thereby help in fostering 'responsible' innovation (Martin 2016).

[17] An upper bound on the number of concepts in the context K = (G,M;R) is $\frac{3}{2}\,2^{\sqrt{|R|+1}} - 1$. See Ganter and Wille 1996, p. 94.

[18] For surveys and explanations, see, e.g., Dias and Vieira (2015), Ignatov (2015), Valtchev et al. (2004), Kuznetszov and Obiedkov (2003), and Godin et al. (1995).

References

Aubin, Jean-Pierre (1990), "A survey of viability theory," *SIAM Journal of Control and Optimization* 28 (4), 749-788.

Adams Richard, John Bessant, and Robert Phelps (2006), "Innovation management measurement: a review," *International Journal of Management Reviews* 8 (1), 21-47.

Baregheh, Anahita, Jennifer Rowley, and Sally Sambrook (2009), "Towards a multidisciplinary definition of innovation," *Management Decision* 47 (8), 1323-1339.

Barros Bagno, Raoni, Mario Sergio Salerno, and Débora Oliveira da Silva (2017), "Models with graphical representation for innovation management: a literature review," *R&D Management* 47 (4), 637-653.

van Beers, Cees, Attila Havas, and Enrica Chiappero-Martinetti (2015), "Overview of existing innovation indicators," CRESSI Working Paper no. 24/2015.

Bělohlávek, Radim (2008), *Introduction to Formal Concept Analysis*, Department of Computer Science, Palacký University, Olomouc, Czeh Republic.

Bělohlávek, Radim, and Juraj Macko (2011), "Selecting important concepts using weights," in *Proceedings of the International Conference on Formal Concept Analysis – ICFCA*, p. 65-80, Springer-Verlag.

Bloom, Nicholas, Charles I. Jones, John Van Reenen, and Michael Webb (2017), "Are ideas getting harder to find?," Centre for Economic Performance Discussion Paper No. 1496, London School of Economics and Political Science.

Brattström, Anna, Johan Frishammar, Anders Richtnér, and Dane Pflueger (2018), "Can innovation be measured? A framework of how measurement of innovation engages attention in firms," *Journal of Engineering and Technology Management* 48, 64-75.

Brown, Tim (2009), *Change by Design - How design thinking transforms organizations and inspires innovation*, HarperCollins Publishers.

Bruscaglioni, Livia (2016), "Theorizing in Grounded Theory and creative abduction," *Quality & Quantity* 50, 2009-2024.

Chan, Vanessa, Chris Musso, and Venkatesh Shankar (2008), "Assessing innovation metrics: McKinsey Global Survey results," *The Mckinsey Quarterly*.

Christensen, Clayton M., Rory McDonald, Elizabeth J. Altman, and Jonathan E. Palmer (2018), "Disruptive innovation: An intellectual history and directions for future research," *Journal of Management Studies* 55 (7), 1043-1078.

Cohendet, Patrick, and Laurent Simon (2015), "Introduction to the special issue on creativity in innovation", *Technology Innovation Management Review* 5 (7), 5-13.

Cruz-Cázares, Claudio, Cristina Bayona-Sáez, and Teresa García-Marco (2013), "You can't manage right what you can't measure well: Technological innovation efficiency," *Research Policy* 42, 1239-1250.

Davey, Brian A. and Hilary. A. Priestley (2002), *Introduction to Lattices and Order*, Cambridge University Press (2nd edition).

Denniston, Jeffrey T., Austin Melton, and Stephen E. Rodabaugh (2013), "Formal contexts, formal concept analysis, and Galois connections," in *Proceedings Festschrift for Dave Schmidt*, accessed 10 April 2015 at http://arxiv.org/pdf/1309.5134v1.pdf.

Dewangan, Vikas, and Manish Godse (2014), "Towards a holistic enterprise innovation performance system," *Technovation* 34, 536-545.

Di Stefano, Giada, Alfonso Gambardella, and Gianmario Verona (2012), "Technology push and demand pull perspectives in innovation studies: Current findings and future research directions," *Research Policy* 41, 1283-1295.

Dias, Sérgio M., Newton J. Vieira (2015), "Concept lattices reduction: definition, analysis and classification," *Expert Systems with Applications* 42, 7084-7097.

Diday, Edwin, and Richard Emilion (2003), "Maximal and stochastic Galois lattices," *Discrete Applied Mathematics* 127, 272-284.

Doignon, Jean-Paul, and Jean-Claude Falmagne (1999), *Knowledge Spaces*, Springer-Verlag.

Dutta, Soumitra, editor (2012), *The Global Innovation Index 2012 – Stronger innovation linkages for global growth*, INSEAD and World Intellectual Property Organization (WIPO).

Dziallas, Marisa, and Knut Blind (2019), "Innovation indicators throughout the innovation process: An extensive literature analysis," *Technovation* 80-81, 3-39.

EC, IMF, OECD, UN, The World Bank (2009), *Systems of National Accounts, 2008*. United Nations, New York.

Edwards-Schacter, Mónica (2018), "The nature and variety of innovation," *International Journal of Innovation Studies* 2, 65-79.

Gamal, Dalia, Tarek Salah, and Nesreen Elrayyes (2011), "How to measure organization Innovativeness? – An overview of innovation measurement frameworks and innovation audit/management tools," Technology Innovation and Entrepreneurship Center (TIEC), Egypt.

Ganter, Bernhard, and Sergei O. Kuznetsov (2000), "Formalizing hypotheses with concepts," in Bernhard Ganter and Guy W. Mineau (eds.), *ICCS 2000*, Springer Verlag.

Ganter, Bernhard and Rudolf Wille (1996), *Formal Concept Analysis - Mathematical Foundations*, Springer-Verlag.

Garcia, Rosanna, and Roger Calantone (2002), "A critical look at technological innovation typology and innovativeness terminology: a literature review," *Journal of Product Innovation Management* 19, 110-132.

Gardiner, Eleanor J., and Valerie J. Gillet (2015), "Perspectives on knowledge discovery algorithms recently introduced in chemoinformatics: rough set theory, association rule mining, emerging patterns, and formal concept analysis," *Journal of Chemical Information and Modeling* 55, 1781-1803.

Gault, Fred, ed. (2013), *Handbook of Innovation Indicators and Measurement*, Edward Elgar.

Gault, Fred (2018), "Defining and measuring innovation in all sectors of the economy," *Research Policy* 47 (3), 617-622.

Godin, Robert, Rokia Missaoui, and Hasssan Alaoui (1995), "Incremental concept formation based on Galois (concept) lattices," *Computational Intelligence* 11 (2), 246-267.

Goldenberg, Jacob, Roni Horowitz, Amnon Levav, and David Maxursky (2003), "Finding the sweet spot of innovation," *Harvard Business Review* 81, 120– 129.

Goldsmith, Ronald E., and Gordon R. Foxall (2003), "The measurement of innovativeness," in Larisa V. Shavinina (editor), *The International Handbook on Innovation*, Elsevier.

Griliches, Zvi (1986), "Data issues in econometrics," in Zvi Griliches and Michael D. Intriligator (eds.), *Handbook of Econometrics*, vol. III, chapter 21, North-Holland.

Gupta, Abhishek (2009), "A study of metrics and measures to measure innovation at firm level & at national level," IMRI research paper 2009-03.

Hall, Bronwyn H., and Adam B. Jaffe (2018), "Measuring science, technology, and innovation: a review," *Annals of Science and Technology Policy* 2 (1), 1-74.

Hatchuel, Armand, and Benoit Weil (2009), "C-K design theory: an advanced formulation," *Research in Engineering Design* 19, 181-192.

von Hippel, Eric (2017), *Free Innovations*, MIT Press.

Hua Tan, Kim, and Yuanzhu Zhan (2017), "Improving new product development using big data: a case study of an electronics company," *R&D Management* 47 (4), 570-582.

Ignatov, Dmitry (2015), "Introduction to Formal Concept Analysis and its applications to information retrieval and related fields," in Pavel Braslavski, Nikolay Karpov, Marcel Worring, Yana Volkovich, and Dmitry I. Ignatov (eds.), *Information Retrieval*, Springer.

Ilevbare, Imoh M., David Probert, and Robert Phaal (2013), "A review of TRIZ, and its benefits and challenges in practice," *Technovation* 33 (2-3), 30-37.

Khan, Faisal I., Tahir Husain, and Ramzi Hejazi (2004), "An overview and analysis of site remediation technologies," *Journal of Environmental Management* 71, 95-122.

Kleinknecht, Alfred, Klees van Montfort, and Eric Brouwer (2002), "The non-trivial choice between innovation indicators," *Economics of Innovation and New Technology* 11, 109-121.

Kuznetsov, Sergei O., and Sergei A. Obiedkov (2003), "Comparing performance of algorithms for generating concept lattices," *Journal of Experimental & Theoretical Artificial Intelligence* 14 (2-3), 189-216.

Lhuillery, Stéphane, Julio Raffo, and Intan Hamdan-Livramento (2017), "Measurement of innovation," in Harald Bathelt, Patrick Cohendet, Sebastian Henn, and Laurent Simon (eds.), *The Edward Elgar Companion to Innovation and Knowledge Creation*, Edward Elgar.

Mabsout, Ramzi (2015), "Abduction and economics: the contributions of Charles Peirce and Herbert Simon," *Journal of Economic Methodology* 22 (4), 491-516.

Mairesse, Jacques, and Pierre Mohnen (2002), "Accounting for innovation and measuring innovativeness: an illustrative framework and an application," *American Economic Review* 92 (2), 226-230.

Mari, Luca (2003), "Epistemology of measurement," *Measurement* 34, 17-30.

Martin, Ben R. (2016), "Twenty challenges for innovation studies," *Science and Public Policy* 43 (3), 432-450.

Munier, Fraancis (2013), "Creativity and uncertainty in the act of work: the contribution of the viability theory," in Thierry Burger-Helmchen (ed.), *The Economics of Creativity, Ideas, Firms and Markets*, Routledge.

Nanda, Jyotirmaya, Henri Thevenot, Timothy W. Simpson, and Roger B. Stone (2007), "Product family design knowledge representation, aggregation, reuse, and analysis," *Artificial Intelligence for Engineering Design, Analysis and Manufacturing* 21, 173-192.

OECD (1997), *The Oslo Manual: Proposed Guidelines for Collecting and Interpreting Technological Innovation Data*, OECD Publishing, Paris.

OECD (2010), *Measuring Innovation, A New Perspective*, OECD Publishing, Paris.

Oh, Gabjin, Ho-Yong Kim, and Ayoung Park (2017), "Analysis of technological innovation based on citation information," *Quality & Quantity* 51, 1065-1091.

Ore, Oystein (1944), "Galois connexions," *Transactions of the American Mathematical Society* 55, 493-513.

Palm, Günther (2012), *Novelty, Information and Surprise*, Springer-Verlag.

Park, Han Woo (2020), "A new era of *Quality & Quantity: International Journal of Methodology*," *Quality & Quantity* 54, 1-2.

Pawlak, Zdzislaw, and Andrzej Showron (2007), "Rudiments of rough sets," *Information Sciences* 177, 3-27.

Pietarinen, Atti-Veikko, and Francesco Bellucci (2015), "New light on Pierce's conception of retroduction, deduction, and scientific reasoning," *International Studies in the Philosophy of Science* 28 (4), 353-373.

Poelmans, Jonas, Dimitri I. Ignatov, Sergei Kuznetzov, and Guido Dedene (2014), "Fuzzy and rough formal concept analysis: a survey," *International Journal of General Systems* 43 (2), 105-134.

Poelmans, Jonas, Sergei O. Kuznetsov, Dimitri I. Ignatov, and Guido Dedene (2013a), "Formal concept analysis in knowledge processing: a survey on models and techniques," *Expert Systems with Applications* 40 (16), 6601--6623.

Poelmans, Jonas, Dimitri I. Ignatov, Sergei O. Kuznetsov, and Guido Dedene (2013b), "Formal concept analysis in knowledge processing: a survey on applications,".*Expert Systems with Applications* 40 (16), 6538--6560.

Quintane, Eric, R. Mitch Casselman, B. Sebastian Reiche, and Petra A. Nylund (2011), "Innovation as a knowledge-based outcome," *Journal of knowledge management* 15 (6), 928-947

Quintero, Nancy Y., and Guillermo Restrepo (2017), "Formal concept analysis applications in chemistry: from radionuclides and molecular structure to toxicity and diagnosis," in Marco Fattore and Rainer Bruggemann (eds.), *Partial Order Concepts in Applied Sciences*, Springer International Publishing.

Raedersdorf Bollinger, Sophie (2019), "Creativity and forms of managerial control in innovation processes: tools, viewpoints and practices," *European Journal of Innovation Management* 23 (2), 214-229.

Richtnér, Anders, Anna Brattström, Johan Frishammar, Jennie Björk, and Mats Magnusson (2017), "Creating better innovation management practices," *MIT Sloan Management Review* 59 (1), 45-53.

Rodriguez-Jimenez, J. M., P. Cordero, M. Enciso, and S. Rudolph (2016), "Concept lattices with negative information: a characterization theorem," *Information Sciences* 369, 51-62.

Rogers, Mark (1998), "The definition and measurement of innovation," Melbourne Institute Working Paper 10/98.

Saunila, M. (2017), "Innovation performance measurement: a quantitative systematic literature review," *Proceedings of the European Conference on Innovation and Entrepreneurship*, 596–601.

Simon, Herbert A. (1973), "Does scientific discovery have a logic," *Philosophy of Science* 40 (4), 471-480.

Sinclair-Desgagné, Bernard (2017), "Innovation and the global eco-industry," in Harald Bathelt, Patrick Cohendet, Sebastian Henn and Laurent Simon (eds.), *Companion to Innovation and Knowledge Creation*, Edward Elgar.

Singh, Prem Kumar, Cherukuri Aswani Kumar, and Abdullah Gani (2016), "A comprehensive survey on formal concept analysis, its research trends and applications," *International Journal of Applied Mathematics and Computer Science* 26 (2), 495-516.

Smith, Keith (2005), "Measuring innovation," in Jan Fagerberg, David G. Mowery, and Richard R. Nelson (eds.), *The Oxford Handbook of Innovation*, Oxford University Press.

Sood, Ashish, Gareth M. James, Gerard J. Tellis, and Ji Zhu (2012), "Predicting the path of technological innovation: SAW vs. Moore, Bass, Gompertz, and Kryder," *Marketing Science* 31 (6), 964-979.

Stefani, Ulrike, Francesco Schiavone, Blandine Laperche, and Thierry Burger-Helmchen (2019), "New tools and practices for financing novelty: a research agenda," *European Journal of Innovation Management* 23 (2), 314-328.

Tanaka, Nobuo, Michel Glaude, and Fred Gault (2005), *Oslo Manual: Guidelines for Collecting and Interpreting Innovation Data - 3rd edition*, OECD/OCDE and Eurostat.

Tohmé, Fernando, and Ricardo Crespo (2013), "Abduction in economics: a conceptual framework and its model," *Synthese* 190, 4215-4237.

Tóth, Tibor, Sándor Radeleczki, Laura Veres, and Attila Körei (2014), "A new mathematical approach to supporting group technology," *European Journal of Industrial Engineering* 8 (5), 716-737.

Valtchev, Petko, Rokia Missaoui, and Robert Godin (2004), "Formal concept analysis for knowledge discovery and data mining: the new challenges," in *Proceedings of the International Conference on Formal Concept Analysis (ICFCA'2004)*, Sydney, Australia.

Valverde-Albacete, Francisco J., José María González-Calabozo, and Anselmo Peñas (2016), "Supporting scientific knowledge discovery with extended, generalized Formal Concept Analysis," *Expert Systems with Applications* 44, 198-216.

Wille, Rudolf (1982), "Restructuring lattice theory: an approach based on hierarchies of concepts," in Ivan Ridal (editor), *Ordered Sets*, Springer Verlag.

Zobel, Ann-Kristin, Boris Lokshinb, and John Hagedoorn (2017), "Formal and informal appropriation mechanisms: The role of openness and innovativeness," *Technovation* 59, 44-54.

Table 1. The initial context ***K*** of groundwater remediation technologies[1]

Attributes ..... Techniques	VOC	Heavy Hydro Carb	Insec ticide	Inorg anic	Heav meta	Low dep	$	$$	$$$	L eff	M Eff	H eff	Fast	Avg	Long
Air sparging	X								X		X		X		
Pump-and-treat	X	X	X	X	X			X	X	X	X				X
Bioslurping		X				X		X	X		X	X		X	X
Ultraviolet Oxidation	X	X	X	X		X	X	X				X	X	X	
Biosparging	X	X				X		X	X		X	X		X	X
Natural attenuation	X	X	X			X	X				X	X			X

Air sparging: involves injecting atmospheric air, under pressure, into the saturated zone to volatize contaminants and promote biodegradation
Pump-and-treat: extraction wells are introduced at various locations and the contaminants are removed with the pumped water
Bioslurping: consists of wells into which an adjustable ‘slurp structure’ is installed
Ultraviolet oxydation: UV bulbs are placed in a reactor where an oxidant comes in contact with the contaminants
Biosparging: air and nutrients are injected into the soil in order to enhance contaminants degradation by naturally occurring organisms
Natural attenuation: uses natural processes to contain the spread of contamination

[1] Adapted from Table 4 in Khan et al. (2004), p. 117.

Table 2. The ex post 'augmented' context $\boldsymbol{K}^{+}$ of groundwater remediation technologies[2]

Attributes ..... Techniques	VOC	Heavy Hydro Carb	Insec ticid	Inorg anic	Heav meta	ε	Low dep	$	$$	$$$	L eff	M eff	H eff	Fast	Avg	Long
Air sparging	X					x			x	X		X		X		
Pump-and-treat	X	X	X	X	X				X	X	X	X				X
Bioslurping		X					X		X	X		X	X		X	X
Ultraviolet Oxidation	X	X	X	X			X	X	X				X	X	X	
Biosparging	X	X				x	X		X	X		X	X		X	X
Natural attenuation	X	X	X				X	X				X	X		x	X
Passive/reactive walls	x	x		x			x					x				x
Groundwater circulation wells	x	x				x		x	x			x	x		x	

Passive/reactive walls: rely on the natural movement of water to carry the contaminants through an installed underground wall structure
Groundwater circulation wells: draws water into and pump through a well, then reintroducing it without reaching the surface

2 Adapted from Table 4 in Khan et al. (2004), p. 117.

Table 3. The ‘incremental-knowledge’ context ***K**** for groundwater remediation technologies

Attributes ..... Techniques	VOC	Heavy Hydro Carb	Insec ticid	Inorg anic	Heav meta	ε	Low Dep	$	$$	$$$	L eff	M eff	H eff	Fast	Avg	Long
Air sparging						x			x							
Pump-and-treat																
Bioslurping																
Ultraviolet Oxidation																
Biosparging						x										
Natural attenuation															x	
Passive/reactive walls	x	x		x			x					x				x
Groundwater circulation wells	x	x				x		x	x			x	x		x	

Table 4. The concepts of the incremental-knowledge context ***K**** of groundwater remediation technologies

	Concept (A,B)
1	({Airsparg, GW walls} , {ε, $$})
2	({Airsparg, Biosparg, GW walls} , {ε})
3	({Nat atten, GW walls} , {Avg})
4	({P/R walls} , {VOC, HHC, Inog, LowD, M eff, Avg})
5	({GW walls} , {VOC, HHC, ε, $, $$, M eff, H eff, Avg})
6	({P/R walls, GW walls} , {VOC, HHC, M eff})
7	({Airsparg, Biosparg, Nat atten, P/R walls, GW walls} , ∅)
8	(∅ , {VOC, HHC, Inorg, ε, Low D, $, $$, M eff, H eff, Avg, Long})

Table 5. The 'complementary' initial context $\overline{\boldsymbol{K}}$ of groundwater remediation technologies

Attributes ..... Techniques	VOC	Heavy Hydro Carb	Insec ticide	Inorg anic	Heav meta	Low dep	$	$$	$$$	L eff	M Eff	H eff	Fast	Avg	Long
Air sparging		Z	Z	Z	Z	Z	Z	Z		Z		Z		Z	Z
Pump-and-treat						Z	Z					Z	Z	Z	
Bioslurping	Z		Z	Z	Z		Z			Z			Z		
Ultraviolet Oxidation					Z				Z	Z	Z				Z
Biosparging			Z	Z	Z		Z			Z			Z		
Natural attenuation				Z	Z			Z	Z	Z			Z	Z	

Table 6. The *anti-concepts* of the initial context ***K*** of groundwater remediation technologies

Anti-concept (A,B)	γ(A)	μ(B)	Anti-concept (A,B)	γ(A)	μ(B)
({Airsparg} , {HHC, Insect, Inorg, Heavy met, Low depth, $, $$, L eff, H eff, Avg, Long})	{ε , $$}	∅	({Bioslurp, Biosparg, Nat atten} , {Inorg, Heavy met, L eff, Fast})	∅	∅
({P&T} , {Low depth, $, H eff, Fast, Avg})	∅	∅	({P&T, Nat atten} , {Fast, Avg})	∅	∅
({Bioslurp} , {VOC, Insect, Inorg, Heavy met, $, L eff, Fast})	∅	∅	({Airsparg, P&T, Bioslurp, Biosparg} , {$})	∅	{GW wells}
({UV oxyd} , {Heavy met, $$$, L eff, M eff, Long})	∅	∅	({Airsparg, P&T, Nat atten} , {Avg})	∅	{Nat atten, GW wells}
({Bioslurp, Biosparg}, {Insect, Inorg, Heavy met, $, L eff, Fast})	∅	∅	({Airsparg, Bioslurp, UV oxyd, Biosparg, Nat atten} , {Heavy met, L eff})	∅	∅
({Nat atten} , {Inorg, Heavy met, $$, $$$, L eff, Fast, Avg})	{Avg}	∅	({P&T, Bioslurp, Biosparg} , {$, Fast})	∅	∅
({Airsparg, P&T} , {L depth, $, H eff, Avg})	∅		({P&T, Bioslurp, Biosparg, Nat atten} , {Fast})	∅	∅
({Airsparg, Biospurg, Biosparg} , {Insect, Inorg, Heavy met, $, L eff})	∅	∅	({Airsparg, Bioslurp, Biosparg, Nat atten} , {Inorg, Heavy met, L eff})	∅	∅
({Airsparg, UV oxyd} , {Heavy met, L eff, Long})	∅	∅	({UV oxyd, Nat atten} , {Heavy met, $$$, L eff})	∅	∅
({Airsparg, Nat atten} , {Inorg, Heavy met, $$, L eff, Avg})	∅	∅	({Airsparg, P&T, Bioslurp, UV oxyd, Biosparg, Nat atten} , ∅)	∅	{Airsparg, P&T, Bioslurp, UV oxyd, Biosparg, Nat atten, P/R walls, GW wells}
			(∅ , {VOC, HHC, Insect, Inorg, Heavy met, L depth, $, $$, $$$, L eff, M eff, Fast, Avg, Long})	{VOC, HHC,Insect, Inorg, Heavy met, ε, L depth, $, $$, $$$, L eff, M eff, Fast, Avg, Long}	∅

APPENDIX

This appendix presents the mathematics which underlie this paper's new metrics and intended grasp at invention. Section A.1 revisits the basics of Formal Concept Analysis (FCA). Section A.2 next introduces the notion of 'innovation mappings' and shows how these mappings can provide further support to our innovation index. Section A.3 finally justifies the process for thinking out of the box that was outlined in Section 4. The treatment is self-contained. Only set-theoretic arguments are used throughout.

A.1 Basic FCA notions

A *formal context* is referred to as a triplet $K = (G, M; R)$, where G is a set of *objects*, M a set of *attributes* these objects may have, and R is a *relation* between G and M, i.e. a subset of the Cartesian product $G \times M$ with the interpretation that $(g, m) \in R$, or gRm, if object g has attribute m.

Denote $\wp(G)$ and $\wp(M)$ the respective power sets (or sets of all subsets) of G and M. Set inclusion $\subseteq$ provides a partial order on the elements of these sets.[1] The following set-to-set functions I_R and E_R defined as

$$\text{for } S \subseteq G, \; I_R(S) = \{m \in M : gRm \text{ for all } g \in S\}$$

$$\text{for } T \subseteq M, \; E_R(T) = \{g \in G : gRm \text{ for all } m \in T\}$$

are called the *Birkhoff Operators* for G and M respectively. For a set of objects S, $I_R(S)$ gives all the attributes in T which these objects have in common. For a given set of

[1] A set Q is a *partially ordered set* (or *poset*) if there is a relation $\leq$ on Q (called a *partial order*) such that: (i) for $q \in Q$, $q \leq q$ (reflexivity property); (ii) for q_1, $q_2 \in Q$, $q_1 \leq q_2$ and $q_2 \leq q_1$ implies $q_1 = q_2$ (antisymmetry); for q_1, q_2, $q_3 \in Q$, $q_1 \leq q_2$ and $q_2 \leq q_3$ implies $q_1 \leq q_3$ (transitivity).

attributes T, $E_R(T)$ gives all the objects in S that share these attributes. In the context displayed in Table 1, I_R(air sparging, pump-and-treat, bioslurping) = {$$$, M eff} and E_R(heavy metals) = {pump-and-treat}.

A well-known property of the Birkhoff Operators is that of *duality*: knowing $I_R(\cdot)$ completely determines $E_R(\cdot)$, and vice-versa, specifying $E_R(\cdot)$ also defines $I_R(\cdot)$.

A *formal concept* in the context $K = (G, M; R)$ is now a pair (A, B), with $A \subseteq G$ and $B \subseteq M$, such that $I_R(A) = B$ and $E_R(B) = A$. The *extent* of a concept (A, B) is A, while its *intent* is B (hence the notation chosen for the Birkhoff operators).

A *preconcept* in K, finally, is a pair (C, D), with $C \subseteq G$ and $D \subseteq M$, such that $C \subseteq E_R(D)$ or, equivalently, $D \subseteq I_R(C)$. Preconcepts can be ordered as follows (Denniston et al. 2013, p. 108): $(C, D) \sqsubseteq (C', D')$, meaning that (C, D) is less extensive than (C', D'), if $C \subseteq C'$ and $D \subseteq D'$.

A.2 Innovation mappings

From now on, as in the paper, $K = (G, M; R)$ will denote the context which precedes innovation, and $K^+ = (G^+, M^+; R^+)$ the augmented context generated by innovation.[2]

Let us call *innovation mappings* the functions $\gamma : \wp(G^+) \to \wp(M^+)$, $\mu : \wp(M^+) \to \wp(G^+)$ such that[3]

$$\text{for } S^+ \subseteq G^+,\ \gamma(S^+) = I_{R^+}(S^+) \setminus \bigcup_{g \in S^+} I_R(g)$$

$$\text{for } T^+ \subseteq M^+,\ \mu(T^+) = E_{R^+}(T^+) \setminus \bigcup_{m \in T^+} E_R(m)$$

[2]Power sets, Birkhoff Operators, formal concepts, and preconcepts are similarly defined on their context of reference, be it K, K^+, or any other context.

[3]Let's agree that $I_R(g) = \varnothing$ when $g \notin G$, and $E_R(m) = \varnothing$ when $m \notin M$.

If one takes a set $S \subseteq G$ of objects from the initial context K, $\gamma(S)$ delivers the set of attributes (old or new) in M^+ which are newly associated with these objects. In Table 2, for instance, γ(pump-and-treat, bioslurping) $= \varnothing$ and γ(air sparging) $= \{\varepsilon$, \$\$$\}$, where ε is a new attribute and \$\$ is an old one. Similarly, for a subset of initial attributes $T \subseteq M$, $\mu(T)$ gives all (and only) the old or new objects that now possess these attributes. For example, μ(VOC, \$\$) $= \{$groundwater circulation wells$\}$.

As for the Birkhoff operators, there is a *duality property* between $\gamma(\cdot)$ and $\mu(T)$: each one uniquely characterizes the other. These functions also hold additional features which are spelled out in the upcoming propositions.

First, say that a function $\pi : P \to Q$ between two sets P and Q, partially ordered by $\leq$ and $\sqsubseteq$ respectively, is *antitone* (or order-reversing) if, for $p_1, p_2 \in P$, $p_1 \leq p_2$ implies $\pi(p_2) \sqsubseteq \pi(p_1)$. A first statement is now at hand.

PROPOSITION 1: The innovation mappings γ and μ are antitone.

PROOF:

First, consider γ. Take two sets S_1^+, $S_2^+ \in \wp(G^+)$ such that $S_1^+ \subseteq S_2^+$; we must show that $\gamma(S_2^+) \subseteq \gamma(S_1^+)$. If $m \in \gamma(S_2^+)$, then $m \in I_{R^+}(S_2^+)$ so gR^+m for all $g \in S_2^+$. Since $S_1^+ \subseteq S_2^+$, we have that gR^+m for all $g \in S_1^+$, hence $m \in I_{R^+}(S_1^+)$. Now, if $m \notin M$, $m \notin I_R(g)$ for any $g \in S_1^+$; it follows that $m \in I_{R^+}(S_1^+) \setminus \bigcup_{g \in S_1^+} I_R(g) = \gamma(S_1^+)$. Suppose, alternatively, that $m \in M$. Since $m \in \gamma(S_2^+)$, it must be the case that *not*(gRm) for all $g \in S_2^+$, hence *not*(gRm) as well for all $g \in S_1^+$ since $S_1^+ \subseteq S_2^+$; it follows again that $m \in \gamma(S_1^+)$. This shows that $\gamma(S_2^+) \subseteq \gamma(S_1^+)$.

The same line of reasoning works for μ (as it can be expected from duality). Q.E.D.

This property of innovation mappings means that, the more objects or attributes one starts with, the more demanding it is to find new relationships that fit them all. This intuitive result is also instrumental in deriving other important characteristics of innovation mappings.

A key notion to introduce at this point is that of a Galois connection.[4] Let P and Q be two sets partially ordered by $\leq$ and $\sqsubseteq$ respectively. A (antitone) *Galois connection* (π, θ) on P and Q is a pair of functions $\pi : P \to Q$ and $\theta : Q \to P$ such that the following equivalent properties are satisfied.

(i) For each $p \in P$, $p \leq \theta\pi(p)$ and for each $q \in Q$, $q \sqsubseteq \pi\theta(q)$.

(ii) For $p \in P$ and $q \in Q$, $p \leq \theta(q)$ if and only if $q \sqsubseteq \pi(p)$.

It is well-known that the Birkhoff operators (I_R, E_R), (I_{R^+}, E_{R^+}) are antitone Galois connections on, respectively, the power sets $\wp(G)$, $\wp(M)$ and $\wp(G^+)$, $\wp(M^+)$ ordered by set inclusion (see, e.g., Ganter and Wille 1996, p. 13-14). In this case, property (i) means that the attributes common to a given set of objects might be shared by more objects, while the objects that share a given set of attributes might have more attributes in common. Property (ii), on the other hand, says that some objects are among those sharing a given set of attributes if and only if these attributes are among those common to these objects.

[4] Since at least Ore (1944)'s seminal article, Galois connections have been increasingly employed throughout mathematics and computer science. To go beyond the very short primer offered in this paper, the reader may look at Davey and Priestley (2002), Doignon and Falmagne (1999), Ganter and Wille (1996), and some of their common references.

As it turns out, the pair (γ, μ) forms a Galois connection as well.

PROPOSITION 2: The pair of innovation mappings (γ, μ) is a Galois connection on the power sets $\wp(G^+)$ and $\wp(M^+)$ partially ordered by inclusion.

PROOF:

To see this, take two sets $S^+ \in \wp(G^+)$ and $T^+ \in \wp(M^+)$, and notice that

$$\begin{array}{lrcl} & S^+ & \subseteq & \mu(T^+) \\ \text{if and only if} & \forall g & \in & S^+, \forall m \in T^+: \;\; gR^+m \;\; \text{and} \;\; \mathit{not}(gRm) \\ \text{if and only if} & \forall m & \in & T^+, \forall g \in S^+: \;\; gR^+m \;\; \text{and} \;\; \mathit{not}(gRm) \\ \text{if and only if} & T^+ & \subseteq & \gamma(S^+)\ . \end{array}$$

Q.E.D.

Proposition 2 underlies a central result. Like any Galois connection (Ganter and Wille 1996, p. 14), (γ, μ) establishes a relation, noted $R^+_{(\gamma,\mu)}$, between the set of objects G^+ and the set of attributes M^+. This relation is defined as

$$\begin{aligned} R^+_{(\gamma,\mu)} &= \left\{(g,m) \in G^+ \times M^+ \mid g \in \mu(m)\right\} \\ &= \left\{(g,m) \in G^+ \times M^+ \mid m \in \gamma(g)\right\} \end{aligned}$$

We can show that $R^+_{(\gamma,\mu)}$ coincides with $R^+ \setminus R$, the set of all new relationships that was used to define the above innovation metric.

PROPOSITION 3: $R^+_{(\gamma,\mu)} = R^+ \setminus R$.

PROOF: Observe that $(g,m) \in R^+_{(\gamma,\mu)}$ if and only if gR^+m and $\mathit{not}(gRm)$, if and only if $(g,m) \in R^+ \setminus R$. Q.E.D.

This statement establishes a rigorous foundation for the innovation index ∇. Recall that

$$\nabla = \frac{\mid R^+ \setminus R \mid}{\mid R^+ \mid} .$$

Since $R^+_{(\gamma,\mu)} = R^+ \setminus R$, it appears that ∇ *summarizes the output of a pair of innovation mappings.*

A.3 Thinking out of the box

From now on, let $R^+_{(\gamma,\mu)} = R^+ \setminus R$ be referred to as R^*. The latter relation defines another formal context, noted $K^* = (G^+, M^+; R^*)$, which represents *incremental* knowledge. An innovator's challenge is to produce K^* starting with K.

The ordered pair (X, Y) with $X \neq \varnothing$, $Y \neq \varnothing$ is called a *seed* in K for K^* if it is a preconcept in K^* while $X \subseteq G$ and $Y \subseteq M$. As the next statement confirms, the existence of a seed is guaranteed when the initial context harbors a new relationship between the original objects and attributes.

Proposition 4: If $R^* \cap (G \times M) \neq \varnothing$, then there is at least one seed in K for K^*.

Proof:

The assumption implies that there is at least one concept (A, B) in K^* such that $A \cap G \neq \varnothing$ and $B \cap M \neq \varnothing$. Since $A \cap G \subseteq A = I_{R^*}(B) \subseteq I_{R^*}(B \cap M)$ and $B \cap M \subseteq B = E_{R^*}(A) \subseteq E_{R^*}(A \cap G)$, the pair $(A \cap G, B \cap M)$ is a preconcept in K^*. Q.E.D.

As argued in the paper, looking for seeds might be a reasonable first step for an innovator. One major reason is that *it is possible to characterize the location of seeds.*

According to the following proposition, a seed must combine objects and attributes which are a priori unrelated.

PROPOSITION 5: No preconcept (a fortiori concept) in K can be a seed for K^*.

PROOF:

Let (C, D) be a preconcept in K. By definition, $\gamma(C) = I_{R^+}(C) \setminus \bigcup_{g \in C} I_R(g)$. But $D \subseteq I_R(C) = \bigcap_{g \in C} I_R(g) \subseteq \bigcup_{g \in C} I_R(g)$. It follows that $D \nsubseteq \gamma(C)$, hence (C, D) is not a preconcept in K^*. Q.E.D.

A corollary to this assertion is that a seed in K for K^* must be a pair (X, Y), with $X \subseteq G$ and $Y \subseteq M$, such that $X \cap (\bigcup_{m \in Y} E_R(m)) = \varnothing$ and $Y \cap (\bigcup_{g \in X} I_R(g)) = \varnothing$. This suggests working with the so-called *complementary* context of K, noted $\overline{K} = (G, M; \overline{R})$, where $\overline{R} = G \times M \setminus R$ refers to the *negative relation* $g\overline{R}m$ which holds when object g *does not* have attribute m. The next (key, and somewhat surprising!) proposition shows that $\overline{K}$ (which can be obtained from initial data) provides the ground in which innovation might begin.

PROPOSITION 6: A seed is a preconcept in $\overline{K}$.

PROOF:

Let (C, D) be a seed for K^*. Then $D \subseteq \gamma(C) \cap M = M \cap I_{R^+}(C) \setminus \bigcup_{g \in C} I_R(g) \subseteq M \setminus \bigcup_{g \in C} I_R(g) = \bigcap_{g \in C} (I_R(g))^c = I_{\overline{R}}(C)$. Q.E.D.

Seeds for K^* - which are a priori unknown relationships between objects in G and attributes in M - thus happen to point, not only at concepts in K^*, but also at anticoncepts of K.

Interestingly, Proposition 5 suggests that anticoncepts can be constructed using the mappings $\widetilde{\gamma} : \wp(G) \rightarrow \wp(M)$, $\widetilde{\mu} : \wp(M) \rightarrow \wp(G)$ defined respectively as

$$\text{for } S \subseteq G,\ \widetilde{\gamma}(S) = M \setminus \bigcup_{g \in S} I_R(g)$$

$$\text{for } T \subseteq M,\ \widetilde{\mu}(T) = G \setminus \bigcup_{m \in T} E_R(m)$$

Clearly, $\widetilde{\gamma}$ and $\widetilde{\mu}$ are approximations for the innovation mappings γ and μ.