David Bawden and Lyn Robinson

Information's Magic Numbers: The Numerology of Information Science

1 Introduction

Two themes were presented at the Libraries in the Digital Age (LIDA) conference, held in Zadar, Croatia, in June 2014. The first, chaired by David Bawden, focused on "qualitative assessment". Blaise Cronin chaired the second theme, focused on "altmetrics". They, as well as several other conference speakers, emphasized the complementary nature of qualitative and quantitative methods; while quantitative data is of unarguable importance, it must be interpreted insightfully and used with care.[1] Applying numbers sensibly has always been a major concern for Blaise Cronin, as is attested by his publication list. His academic webpage[2] notes that "much of his research focuses on collaboration in science, scholarly communication, citation analysis, the academic reward system and cybermetrics—the intersection of information science and social studies of science", while his Wikipedia entry[3] describes him as being jointly an "information scientist and bibliometrician". Despite this strong informetric focus, Cronin has had a long-standing concern about the potential descent of this aspect of the information science discipline into a "new age of numerology," due to over-use and misuse and of bibliometrics and altmetrics; see, for example, Cronin (1998; 2000), Cronin and Sugimoto (2015), and Priego (2012). It is therefore appropriate to include in this volume a chapter on the numerology of information science; to ask to what extent we are able to identify a few numbers which may helpfully encapsulate important aspects of the subject.

Numerology, roughly the belief that numbers in general, and integers in particular, have their own nature and properties, and can of themselves influence events, is rather out of favor nowadays, being regarded as a pseudoscience. The impeccable scientific belief that the regularities of nature can be captured by simple mathematical relationships is a long way from Blair's (1976, p. 81) notion that "numbers, quite distinct from their empirical use, become a language, as full of

1 The presentations may be found on the conference website at http://ozk.unizd.hr/proceedings/index.php/lida

2 http://www.soic.indiana.edu/all-people/profile.html?profile_id=4

3 http://en.wikipedia.org/wiki/Blaise_Cronin

DOI 10.1515/9783110308464-012,

metaphor and dimension as poetry". However, before sneering at the idea that numbers in themselves can have a significance, we should remember that the long-standing, and still influential, Platonic tradition within science views numbers as having their own objective existence, and indeed that the physical universe and everything in it is, at root, a mathematical structure made of numbers (see Tegmark [2014] for a recent and accessible account of this position).

As well as numbers *per se*, numerology is often taken, usually critically, to mean an enthusiasm for simple numerical formulae, usually involving integers, capturing some significant aspect of reality. These have been seen in both the sciences and the social sciences: notoriously, the British physicist and astronomer Sir Arthur Eddington spent many years seeking simple integer relationships as the clue to the universe (Kilmister, 2005). It is clear that there are strong relations between numbers, the physical world and cultural issues, as is clearly shown by the sequence of "kissing numbers", the number of spheres which in any space exactly bound a further identical sphere (Weisstein, n.d.); two points on a one dimensional line bound a third point, six circles circumscribe a seventh, and twelve balls circumscribe a thirteenth. The resultant sequence of "kissing number plus one"—three, seven, thirteen—captures the principal significant/lucky/unlucky numbers in numerous cultures, and is numerologically present in the 'leader with twelve followers' meme of Christ, Osiris, King Arthur, and others (Blair, 1976).[4] Therefore, despite the dangers of slipping into a facile numerology, simple numbers and integer relations may still be worth investigating.

There are, in fact, relatively few such simple numbers and number relations in information science, and what exists was imported from adjacent disciplines. In truth, they are not all very simple: one is very large, some have alternatives, one is a sequence, and one is infinite. These numbers encapsulate a variety of issues: how much information there is, or could be; the optimal size of communicating groups; the structure of information networks; the distribution of information activities; and the limits to the growth of knowledge. We find that sometimes, but not always, the actual number is less important than the theoretical perspective to which it points. We begin by considering the big picture: how much information there is, or could be, in the human context and in the universe. Then we move to the smallest scale, the information associated with the conscious attention of a single person. From there, we move up the scale, to information associated with

4 It would have been nice if the four-dimensional kissing number, which was not known until 2003 (Pfender & Ziegler, 2004) and cannot be intuitively grasped like the small dimension equivalents, had also related to some culturally significant number. Disappointingly, it was shown to be 24, and 25 does not appear to have significance in any culture.

groups, with networks, and with disciplines; and finally up to the largest scale, to the infinity of possible recorded information.

2 The Universal Number: 8×10^{21}

The most fundamental number-related question we can ask about information is simply: *How much information is there*? This leads to a spin-off question: *How much information could there be*? Both, perhaps not surprisingly, are difficult to answer accurately. Attempts to answer such questions have been reviewed by Bawden and Robinson (2012a), Gleick (2011), Davis and Shaw (2011), and Floridi (2014).

Before the advent of widespread digital information, the "How much information is there?" question was generally answered in terms of counts of documents: how many books, articles, reports, etc. had been published. For example, Jinha (2010) suggested that the total number of scholarly articles had reached fifty million. More recent attempts have had to include the much larger amount of born-digital information—an intrinsically more difficult process—with results that can only be approximate.

The first attempt in the digital era to address this question in a rigorous way was the "How much information?" study from the School of Information Management and Systems at University of California Berkeley, United States (U.S.), first carried out in 2000, and repeated in 2003 (SIMS, 2003). The study estimated that approximately 12 exabytes of information had been recorded by humanity before the general use of computers, but was being dwarfed by the amounts now being generated and stored. About 5 exabytes of new information was stored during 2002; equal to 37,000 times the information in the Library of Congress, or 800 megabytes/30 feet (10 metres) of bookshelf content per person on the planet. However, more than three times this amount of information was communicated through electronic channels but never stored.

Later studies (Hilbert & Lopez 2011; Ganz & Reinsel 2011; Hilbert, 2014) have suggested that the amount of information broadcast each year, in increasingly varied formats, was approaching 2 zetabytes by 2007, that current capacity of all information storage devices approaches 300 exabytes, and that the total grew to over 1600 exabytes between 2007 and 2011. Floridi (2014, p. 13) points out that this amounts to enough information being generated each day to fill all the libraries in the U.S. eight times over, and that the figure is likely to grow threefold every four years, so that there may be expected to be 8 zetabytes (8 times 10^{21} bytes) of information by 2015; this figure is taken as 'the number' for this section.

One consequence of this, as Floridi (2014) points out, is that since 2007 information has been produced at a faster rate than have storage devices to handle it; this despite Kryder's Law, which shows that the capacity of storage devices is increasing at an even faster rate than is processing capability, the latter obeying Moore's Law. That we cannot therefore store all our information arguably does not, in fact, matter; the great proportion now is data generated by machines and used by machines, without any need for longer-term storage for human intervention or reflection.

The actual value of these numbers is immaterial—what is of importance is their scale and order of magnitude, and the ways in which they are changing, to create what Floridi (2014) terms the "infosphere," an entirely new form of information environment.

The second question, how much information could there be, is answered by considering the capacity of the physical universe to hold bits of information; a rough estimate, subject to many approximations, is reported by Gleick (2011) to be about 10^{90} bits. This figure, though of no practical significance for information science, reminds us that information is always physically instantiated, and its processing is limited by the constraints of the physical universe. Of course, there are those who would go further, and say that information is physical *per se*, but that is a topic for a different discussion; see, for instance, Bawden and Robinson (2013), the contributors to Davies and Gregersen (2010), and Hjørland (2007).

3 The Personal Number: 7 (or 4)

One of the most cited papers in the human sciences is American psychologist George Miller's "Magical number 7, plus or minus 2" (Miller, 1956). This paper drew attention to the significance of the number in human information processing. The main finding was that the number of concepts of items which an adult can hold in conscious attention, or short-term memory, at one time is about seven. Miller, an early enthusiast for the application of information theory in psychology, used information theory concepts to speculate about what this meant for the mechanism of memory. This limit has been widely tested, and was generally regarded as correct. However, more recent studies, summarized by Cowan (2001), have suggested that the limit may be lower, between three and five rather than between five and nine. On the basis of these findings, Cowan recommends a "magical number 4".

Whatever the exact number may be, it is clearly small, and has implications for the way in which information is handled and should be presented. However, there seems to have been relatively little explicit recognition of this in information science. It is tempting to ascribe to this factor the well-known tendency for users of search engines to attend only to the items presented on the first page. This seems likely to be more a matter of disinclination to spend the time necessary to consider more items, rather than an inability to hold them in conscious attention at once; but it may be that some underlying mechanism, associated with the number limit, accounts for both factors.

Knowledge organization systems appear to respect this feature. Decimal classifications may be guided to their ten main sections by a desire for a pleasing notation, and others, most notably the Library of Congress Classification, follow the twenty six letters of the Roman alphabet. But the general tendency, following Ranganathan's five fundamental facets (Hedden, 2010; Broughton, 2006), to have between four and ten main sections or facets in the great majority of taxonomies and thesauri may be seen as an unconscious recognition that this is a number which enables the user, or at least the compiler, to hold the whole structure in mind. Miller's number holds up at more detailed levels of taxonomy design: "A popular rule of thumb is to go only three levels deep and have only six to eight concepts per level. These numbers are based on user experience tests, which have shown that users have the patience to click down only to a third level and can scan only six to eight term entries at once" (Hedden, 2010, p. 236). This, of course, reflects the similar experience with search engines noted above.

4 The Group Number: 150

The group number stems from the work of the British evolutionary psychologist Robin Dunbar, initially inspired by the study of the correlation between brain size and the size of social groups in primates. This led to the idea that there is a natural group size for humans; stable communicative relationships can be maintained with about 150 people (Dunbar, 1993; Dunbar, 2008; Dunbar, 2012). The number derived from the correlation was actually 148, but it has generally, and sensibly, been rounded to 150. This idea is of evident importance for information science, since it is well-known that close acquaintances are a major source of information in most, if not all, contexts (Case, 2012). Further, shared knowledge is a major factor in the maintenance of social groups (Dunbar, 2012; McPherson, Smith-Lovin, & Cook, 2001).

Dunbar argues that the size of the group of communicative relationships which can be maintained at any one time is constrained in part by cognitive factors, and hence ultimately by some aspect of brain size and structure, and partly by available time. Direct evidence in humans is provided by correlations of individual differences in social network size and volumes of social cognition areas in the cortex and amygdala brain structures (Kanai, Bahrami, Roylance, & Rees, 2012; Bickart, Wright, Dautoff, Dickerson, & Feldman Barrett, 2011; Powell, Lewis, Roberts, García-Fiñana, & Dunbar, 2012). Empirically, this has been tested by the observations of groupings in a wide variety of contexts, including hunter-gatherers, farming communities, military formations, industrial and commercial workforces, Christmas card lists, online social networks, and academic disciplines (Dunbar, 1992; Dunbar, 2008; Hill & Dunbar, 2003; Roberts, Dunbar, Pollet, & Ruppens, 2009).

Objections to the 150 value have been raised by those who argue that groupings of around 30–50 people are commonly found in hunter-gatherer populations, arguably the most "natural" form of human grouping (de Ruiter, Weston, & Lyon, 2011). Others have suggested that the number must be much larger because of the evidence that many people have several hundred contacts on social media (Wellman, 2012). However, as Dunbar (2012, p. 2195) asserts, "there is now considerable evidence that groupings of this size [around 150 individuals] occur frequently in human social organization, and that this is the normative limit on the size of personal social networks among adults."

A more nuanced viewpoint, rather than seeking to insist on a single number to encapsulate the complexities of social interaction, is to see a series of numbers, reflecting different strengths of social ties, and of shared knowledge and perspectives. These groups exhibit the kind of "small world" network structure and behavior which will discussed later. Typically, these represent groups with rough sizes 5, 15, 50, 150, 500, 1500, which can be seen as circles, each including those inner, with a scaling factor of about three (Dunbar, 2012; Hamilton, Milne, Walker, Burger, & Brown, 2007; Roberts, Dunbar, Pollet, & Ruppens, 2009; Zhou, Sornette, Hill, & Dunbar, 2005). Groups of these sizes may be characterized roughly as follows:

- **5:** a core social group, or "support clique," to whom an individual would refer very frequently for support, assistance, information and advice;
- **15:** a "sympathy group," with whom there are special ties and frequent contact;
- **50:** typically a temporary grouping, formed for a particular period or task;
- **150:** the stable inter-communicating group, with regular interaction and knowledge sharing;
- **500:** the "megaband," again typically a temporary or pragmatic grouping; and
- **1500:** the "tribe"—acquaintances at best, with whom any relationship, or communication of information, is typically one-way, and there is no little or no sharing of knowledge.

Of these, the 5 and 150 levels seem particularly significant: 150 for the reasons set out by Dunbar, supported by a good deal of evidence, and in keeping with Shirky's (2003, n.p.) recommendation for effective online group size (i.e., "larger than a dozen, smaller than a few hundred"); five because it appears to be a natural small group equivalent, related to the idea that spontaneous conversation and information sharing almost always occurs in groups of not more than four individuals (Dunbar, Duncan, & Nettle, 1995).

It seems evident that an understanding of this group interaction structure, if indeed it is valid and omnipresent, is important for several areas within information science, perhaps most notably in knowledge management. However, there seems to have been little examination of the significance of this group structure with respect to the communication of information. Studies have established that the smaller, and more information-intensive and knowledge-sharing, groups require an investment of time, and ideally substantial face-to-face contact, if their members are not to slip into the larger, and less effective groupings (Dunbar, 2012; Roberts, Dunbar, Pollet, & Ruppens, 2009). This seems to be a warning against reliance on purely digital information sharing, particularly with an assumption that its scale can be increased by technological means, and typifies the value that such theoretical concepts can bring to information practice.

5 The Linking Number: 6

The idea that everyone in the world is connected to everyone else by no more than "six degrees of separation" has become entrenched in popular consciousness through newspaper and magazine articles, plays, TV series, films, and games (Six degrees of separation, n.d.). The concept was introduced by the Hungarian writer Frigyes Karinthy (1929), in his short story *Láncszemek* (*Chains*), but became well-known only with the classic paper of American psychologist Stanley Milgram (1967). This initiated a research program in what became known as "small world" phenomena; for a detailed review, from Karinthy onwards, see Schnettler (2009).

In Milgram's study, randomly chosen participants in the Midwest (U.S.), were asked to try to send a printed message to a target in New England (U.S.), by sending it to a person with whom they were personally acquainted, asking that it be forwarded in the same way. Only about 30 % succeeded, and those that did varied between two and ten intermediaries, with a median of five. This was the basis for the idea of "six degrees of separation," although Milgram did not use this phrase in his paper. Focusing on the number of nodes, rather than links, in the chain, he wrote of "five circles of acquaintances" (Milgram, 1967, p. 65). The rather more memorable "six degrees of separation" phrase was introduced two decades later, in a play with that name (Guare, 1990).

Some limited empirical research in the social sciences investigated this idea over the next thirty years, until the subject was revitalized by formal mathematical modeling of network connectivity in all kinds of contexts, not just social (Caldarelli & Catanzaro, 2012; Mitchell 2009; Schnettler, 2009). The formal modeling results tend to support empirical studies in various contexts, in confirming commonly occurring short paths through extensive networks, though they do not support the idea that there is anything special about the number six (or five); median chain lengths can vary from three to fifteen, according to the nature of the network. However, a study aiming to replicate Milgram's work on a much larger scale using e-mail gave quite similar results, of between five and seven steps for the minority of messages which were completed, suggesting that this may be a natural scale for social information networks (Dodds, Muhamad, & Watts, 2003).

As Stock and Stock (2013, p. 384–385) note, the "six degrees of separation" concept has become synonymous with the idea of "small worlds". This expresses, in the social context, the idea that "people are not only linked to their immediate friends, family, and acquaintances, but they are embedded in a larger structure of direct and indirect contacts" (Schnettler, 2009, p. 166). More formally, the "small-world effect" denotes the fact that most nodes in most networks are joined by relatively small paths; a specific "small-world network" has been identified as one with a structure intermediate between highly regular and totally random, with nodes highly clustered, as in regular graphs, and yet with a short path length between any two nodes, as is typical in random graphs (Schnettler, 2009; Watts & Strogatz, 1998).

However, despite this theoretical support for short paths, empirical work on social networks, typically carried out in the sociology domain, have tended to show that, although extended chains of social contacts were available, they were used infrequently for finding information (Schnettler, 2009). For example, in a study of how people found information about job prospects, most used one intermediary, or none, and no chain was more than four links (Granovetter, 1995). The only example of longer chains, with up to nine links and a median of five, was found in a study of women in the U.S. seeking a doctor willing to perform an abortion at a time when legal abortion was severely restricted (Lee, 1969).

Björneborn and Ingwersen (2001, p. 74) noted that small world metrics were potentially relevant to several topics within information science including webometrics, citation analysis, semantic networks, and thesauri, but that there was a lack of research in these areas. Since then there has been some usage in webometric studies, a typical example being the demonstration that the typical path link between sites in the United Kingdom (U.K.) academic web network is three or four (Björneborn, 2006), and in bibliometrics, for example, a study of the co-occurrence of keywords in databases, where the number reflects the distance between papers measured by the keywords in common (Zhu, Wang, Hassan & Haddawy, 2013). The only specific mentions of the "six degrees" idea in the recent information science literature appears to be James' (2006) reflections on the relevance of the idea to information literacy instruction, and Dennie and Cuccia's (2014) application to a chemical literature search assignment.

While considerable research has been carried out within information science using the "small worlds theory", this has largely been detailed qualitative studies of information interactions between groups and networks in limited spaces, physical or virtual (Savolainen, 2009). Concepts such as 'density' from network theory may be applied (see, for example, Huotari & Chatman, 2001), but generally in an informal and semi-quantitative way. Even within these caveats, Schultz-Jones (2009, p. 626) found that "library and information service settings [are] a largely undeveloped context for the application of social network theory and social network analysis." It may be that there is scope for better integration of qualitative and quantitative methods, as Schnettler (2009) advocates in general for small world research, and for a greater focus on contexts closer to our own (disciplinary) home. The number, whether it be 6 or not, is not, in this case, as important as the "network thinking" (Mitchell 2009) to which it points.

Finally, we might note that the "six degrees of separation" idea has launched metrics such as the Bacon number, the closeness of the Hollywood actor Kevin Bacon to any other actor, based on the actors who have worked with actors who have worked with Kevin Bacon, and, perhaps more seriously, the Erdös number, based on how many links of co-authorship link anyone to the Hungarian mathematician Paul Erdös (Grossman, 2014). Perhaps we should establish an analogous Cronin number: one of us [DB] would be 2, since he has not co-authored with Cronin, but has co-authored with at least one person who has, LR would have a Cronin number of 3, on the same basis.

6 The Network Number: 59

As we have just seen, experiments have shown that messages across small world networks fail to get through a majority of the time. This may be due to a variety of context-specific factors, depending on the nature of the network, and the pattern and strength of its connections (Dodds, Muhamad & Watts, 2003; Schnettler, 2009). Milgram (1967) noted a specific, and fairly obvious, point that two groups within a network may be cut off, if there is no link path joining them, so that there is no possibility of information passing between them. Mathematical analysis of networks by the American complexity scientist Stuart Kauffman has shed an interesting light on network behavior in this respect.

Kauffman has shown that, for any network of nodes which are all initially isolated, adding links randomly between nodes causes a pattern of connections to build up, steadily and linearly, so that linked groups are created within the overall network. This may be seen as an instantiation of Ramsey theory, which posits the unavoidable emergence of regularity in large structures, such as networks. It is often expressed as the 'party problem'; how many guests must be invited to a party (or people invited to link to a social media site) so that a minimum number (the "Ramsey Number") will know each other (Gould, nd).

Kauffman shows that when 59 % of the nodes are linked to at least one other, the pattern suddenly and dramatically changes, and the great majority of the nodes are connected. This is referred to as a network phase transition. An accessible account of the phenomenon is given by Kauffman (1996), and its significance is described for computer networks, such as the World Wide Web (Tetlow, 2007) and for social networks such as the financial system (Beinhocker, 2006).

The importance of this number for information science is that it should instill an awareness that the behavior of information networks of all kinds may change, suddenly and dramatically, as their interconnectivity increases. It is easy to assume that overall connectivity within a network, and hence the ability to pass information between any two of its nodes, will increase in a regular manner, as more individual interconnections are added, depending on the number of connected nodes. This is the basis for "laws" relating the value of a network, specifically a computer network, to the number of nodes connected. Metcalfe's Law, for example, states that the value of a network increases as the square of the numbers of nodes connected (Floridi, 2014), while a variant due to Briscoe, Odlyzko, and Tilly (2006) argue for a less rapid growth of n(log n), with n nodes connected. Kauffman's number shows us that this kind of continuous growth in network value is only valid up to a point. Beyond this, rather precisely specifiable, point, a qualitative change in the nature, and value, of the network occurs, leading to an essentially new information environment.

7 The Distribution Numbers: 90, 9, and 1

The numbers 90-9-1 have been found to represent the distribution of activity among users of social media sites, including microblogs, such as Twitter, and wikis, most notably Wikipedia. For every regular contributor, or "superuser", there are nine occasional contributors, and ninety "lurkers", who take information but do not contribute with any regularity; as an example, see van Mierlo (2014). This is an instantiation of a very widespread distribution in information areas. From our days as information practitioners, we recall it being an article of faith, stated anecdotally though never written down so far as we know, that in any complex search for information requiring high recall it was easy to get 90 % of the material, very difficult to get 99 % and impossible to get 100 %.

These are examples of the ubiquitous power law distributions that govern the information world, including those of Bradford, Lotka, Pareto, and Zipf (Bawden & Robinson 2012a; Egghe, 2005; Rousseau, 2010). As such, they are better known within the information science community than the other numbers described in this chapter, and need less exposition. An appreciation of these laws, and the numbers which come from them, informs practice in areas such as collection management, information retrieval, institutional bibliometrics, and the assessment of impact of social media; see, as examples, Corby (2003), Nicolaisen and Hjørland (2007), Bhavnani and Peck (2010), Åström and Hansson (2012), and Hoffman and Doucette (2012). These are thus among the few "magic numbers" which are used widely and directly in the practice of the information disciplines, and particularly in scientometrics.

8 The Knowledge Number: ∞

The knowledge number is generally termed the Champernowne number, after the British mathematician David Champernowne, who derived and published it while still an undergraduate student before going on to a career as an economics professor (Champernowne, 1933; Pickover, 2012, p. 364–365). While he derived his number simply as a mathematical curiosity, it has interesting implications for the information world (von Baeyer, 2003, p. 101–102).

We first choose a base for our number, say binary or decimal. Then we enumerate all the symbols that constitute that number set, then all the pairs, then all the triplets, and so on, for as long as we wish. In decimal base 10, as Champernowne originally presented it, we would write 0.12345678910111213141516 ... or, in the binary system, we would write 0 1 00 01 10 11 000 001 010 100 ... Since we can always continue adding to this number, it must necessarily be infinite in magnitude.

Then we choose a code to convert the number to characters—something like ASCII or Unicode—and convert our potentially infinite number to a potentially long infinite text string. In this infinite character string there will be found everything that has ever been written using the chosen character set, embedded in the (literally) infinitely larger set of everything could be written. We will find the text of Shakespeare's *Midsummer Night's Dream*, in all its editions, in all possible languages, and with all possible misprints and errors. We will find a copy of this paper, with all these variants, and a copy of all the works which Blaise Cronin has written, or might have written. This is an instantiation of Borges' (1998) Library of Babel,

> [Whose] bookshelves contain all possible combinations of [symbols] – that is, all that is able to be expressed, in every language. *All* – the detailed history of the future, the autobiographies of the archangels, the faithful catalog of the Library, thousands and thousands of false catalogs, the proof of the falsity of those false catalogs, a proof of the falsity of the *true* catalog, the Gnostic gospel of Basilides, the commentary upon that gospel, the commentary on the commentary on that gospel, the true story of your death, the translation of every book into every language, the interpolations of every book into all books, the treatise Bede could have written (but did not) on the mythology of the Saxon people, the lost books of Tacitus.
> (Borges, 1998, p. 115)

And Champernowne gives us this in a number.

The number is of no practical value, but it is a striking formal indication of the idea that creativity, and growth of knowledge, are unlimited. While our first number indicated that the amount of information that can be held within the physical universe must be finite, creativity is unlimited, and knowledge can grow indefinitely (see, for example, Deutsch, 2011; Kauffman, 2010).

9 Conclusions

It is difficult to state concisely where these numbers fit into our understanding of the information world, and more specifically in our understanding of informetrics and scholarly communication; though it would be difficult to deny their potential significance. All these numbers are interesting, and some are of immediate use for practice; they take us into the areas where the information sciences overlap with the human sciences, especially psychology, with the physical sciences, and even with philosophy. It is not evident that there is any metatheory which could encapsulate them all, and it may be unrealistic to think of anything of the sort. However, the links between the numbers, for example between Dunbar's social groups and Milgram's small world networks, may serve as a basis for building a modest theoretical framework.

It is still more unrealistic to seek for a single magic number for information. Though, if we had to do so, it would probably be 5, since this appears in several contexts, including cognitive scope, small world links, and optimal group size for information exchange. The numbers measure attributes of people and groups, cognition and networks, collections and activities; all three of Popper's Worlds, for those who like that ontology as a basic for the subject (Bawden, 2002; Bawden & Robinson, 2012B).

The numbers themselves appear rather fluid, and usually their exact value does not matter. It is the general magnitude that is important; it does not matter exactly what volume of information is produced daily, but it does matter, for practical purposes, that it is very large, and getting much larger very rapidly. Nor does it matter, for our purposes, whether the optimal group size for information interaction is exactly Dunbar's 150; though it does matter that it is about 150 rather than the suggested alternative values of 30 or 500.

We may do better to forget numerological relations, and think of qualitative patterns, with the numbers acting as a kind of aide-memoire: "statistical regularities, observed in a context where social influences play an important role", as Rousseau (2010, p. 2747) puts it. Or we may take the numbers as a clue, or introduction, to new theoretical perspectives, in the same way that Milgram's small world of 6 connections opens the way to the much wider idea of scale-free networks following power laws (Mitchell, 2009).

In their LIDA2014 presentations, both Blaise Cronin and David Bawden cited a quotation about the limitations of metrics. His quotation was Albert Einstein's remark that "not everything that can be counted counts, and not everything that counts can be counted," while David's mentioned Václav Havel's recommendation that we should have "a humble reverence for everything that we shall never measure". They amount to the same thing. Numbers will never tell the whole story, in information or in any other context. But that should not prevent us from continuing to seek for numbers, magic or otherwise, which capture the structures and patterns of the information world.

Cited References

Åström, F. & Hansson, J. (2012). How implementation of bibliometric practice affects the role of academic libraries. *Journal of librarianship and information Science*, *45*(4), 316–322.

Bawden, D. (2002). The three worlds of health information. *Journal of Information Science*, *28*(1), 51–62.

Bawden, D., & Robinson, L. (2012A). Informetrics. In D. Bawden and L. Robinson, *Introduction to information science* (pp. 163–185). London: Facet.

Bawden, D., & Robinson, L. (2012B). Basic concepts of information science. In D. Bawden and L. Robinson, *Introduction to information science* (pp. 63–89). London: Facet.

Bawden, D. & Robinson, L. (2013). "Deep down things": in what ways is information physical, and why does it matter for LIS?. *Information Research*, *18*(3), paper C03. Retrieved from http://InformationR.net/ir/18-3/colis/paperC03.html
Beinhocker, E. D. (2006). *The origin of wealth: evolution, complexity, and the radical remaking of economics*. Cambridge, MA: Harvard Business School Press.
Bhavnani, S. K., & Peck, F.A (2010). Scatter matters: Regularities and implications for the scatter of healthcare information on the Web. *Journal of the American Society for Information Science and Technology*, *61*(4), 659–676.
Blair, L. (1976). *Rhythms of Vision*, St. Albans: Paladin.
Bickart, K. C., Wright, C. I., Dautoff, R. J., Bradford, C. D., & Feldman Barratt, L. (2011). Amygdala volume and social network size in humans. *Nature Neuroscience*, *14*(2), 163–164.
Björneborn, L. (2006). "Mini small worlds" of shortest link paths crossing domain boundaries in an academic Web space. *Scientometrics, 68*(3), 395–414.
Björneborn, L. & Ingwersen, P. (2001). Perspectives of webometrics. *Scientometrics*, *50*(1), 65–82.
Borges, J. L. (1998), The Library of Babel. In J. L. Borges, *Collected fictions*, London: Allen Lane, pp. 112–118.
Briscoe, B., Odlyzko, A., & Tilly, B (2006). Metcalfe's Law is wrong. *IEEE spectrum*, July 2006, pp. 26–31. Retrieved from http://spectrum.ieee.org/computing/networks/metcalfes-law-is-wrong
Broughton, V. (2006). The need for a faceted classification as the basis of all methods of information retrieval. *Aslib Proceedings*, *58*(1/2), 49–72.
Caldarelli, G. & Catanzaro, M. (2012). *Networks: a very short introduction*. Oxford: Oxford University Press.
Case, D. O. (2012). *Looking for information: a survey of research on information seeking, needs and behavior (3rd ed.)*, Bingley: Emerald.
Champernowne, D. G. (1933). The construction of decimals normal in the scale of ten. *Journal of the London Mathematical Society*, *8*(4), 254–260.
Corby, K. (2003), Constructing core journal lists: mixing science and alchemy. *portal: Libraries and the academy*, *39*(2), 207–217.
Cowan, N. (2001). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behavioural and Brain Sciences*, *24*(1), 87–114.
Cronin, B. (1998). New Age numerology: a gloss on Apostol. *Science Tribune*, June 1998. Retrieved 1 July 2014 from http://www.tribunes.com/tribune/art98/cron.htm
Cronin, B. (2000). Bibliometrics and beyond: some thoughts on webometrics and influmetrics. *Biomed central*. Retrieved 1 July 2014 from http://www.biomedcentral.com/meetings/2000/foi/transcripts/cronin
Cronin, B., & Sugimoto, C. R. (Eds.). (2015). *Scholarly metrics under the microscope: From citation analysis to academic auditing.* Medford, NJ: Information Today, INC/ASIST, pp. 976.
Davies, P. & Gregersen, N. H. (2010). *Information and the nature of reality: from physics to metaphysics*. Cambridge: Cambridge University Press.
Davis, C. H. & Shaw, D. (eds.) (2011). *Introduction to information science and technology*. Medford NJ: Information Today.
de Ruiter, J., Weston, G., & Lyon, S. M. (2011). Dunbar's Number: group size and brain physiology in humans re-examined. *American Anthropologist*, *113*(4), 557–568.

Dennie, D. & Cuccia, L. A. (2014). "Six degrees of separation"—Revealing a "small-world phenomenon" through a chemistry literature search activity. *Journal of Chemical Education*, *91*(4), 546–549.

Deutsch, D. (2011). *The beginnings of infinity: explanations that transform the world*. London: Allen Lane.

Dodds, P. S., Muhamad, R. & Watts, D. J. (2003). An experimental study of search in global social networks. *Science*, *301*(5634), 827–829.

Dunbar, R. I. M. (1993). Coevolution of neocortex size, group size and language in humans. *Behavioral and Brain Sciences*, *16*(4), 681–684.

Dunbar, R. I. M. (2008), Mind the gap: or why humans aren't just great apes. *Proceedings of the British Academy*, vol. 154 (2007 lectures), 403–423.

Dunbar, R. I. M. (2012), Social cognition on the Internet: testing constraints on social network size. *Philosophical transactions of the Royal Society B*, *367*(1599), 2192–2201.

Dunbar, R. I. M., Duncan, N. & Nettle, D. (1995), Size and structure of freely forming conversational groups. *Human Nature*, *6*(1), 67–78.

Egghe, L. (2005). *Power laws in the information production process: Lotkaian informetrics*. Amsterdam: Elsevier.

Floridi, L. (2014). *The 4th revolution: how the infosphere is reshaping human reality*. Oxford: Oxford University Press.

Ganz, J. & Reinsel, D. (2011). Extracting value from chaos. Retrieved 30 June 2014 from http://uk.emc.com/collateral/analyst-reports/idc-extracting-value-from-chaos-ar.pdf

Gleick, J. (2011). *The information: a history, a theory, a flood*. London: Fourth Estate.

Gould, M. (nd). Ramsey theory. Retrieved 12 September 2014 from http://people.maths.ox.ac.uk/~gouldm/ramsey.pdf

Granovetter, M. S. (1995). *Getting a job: a study of contacts and careers (2nd ed.)*. Chicago IL: University of Chicago Press.

Grossman, J. (2014). The Erdös number project. Oakland University. Retrieved 12 September 2014 from http://www.oakland.edu/enp

Guare, J. (1990). *Six degrees of separation*. New York NY: Vintage.

Hamilton, M. J., Milne, B. T., Walker, R. S., Burger, O., & Brown, J. H. (2007). The complex structure of hunter-gatherer social networks. *Proceedings of the Royal Society B*, *274*(1622), 2195–2202.

Hedden, H. (2010). *The accidental taxonomist*. Medford NJ: Information Today.

Hilbert, M. (2014). What is the content of the world's technologically mediated information and communication capacity: how much text, image, audio and video? *The Information Society*, *30*(2), 127–143.

Hilbert, M. & Lopez, P. (2011). The world's technological capacity to store, communicate and compute information. *Science, 332*(6025), 60–65.

Hill, R. A. and Dunbar, R. I. M. (2003). Social network size in humans. *Human Nature*, *14*(1), 53–72.

Hjørland, B. (2007). Information: objective of subjective/situational? *Journal of the American Society for Information Science and Technology*, *58*(10), 1448–1456.

Hoffman, K. & Doucette, L. (2012), A review of citation analysis methodologies for collection management. *College and Research Libraries*, *73*(4), 321–335.

Huotari, M-L. & Chatman, E. (2001), Using everyday life information seeking to explain organizational behaviour, *Library and Information Science Research*, *23*(4), 351–366.

James, K. (2006), Six degrees of information seeking: Stanley Milgram and the small world of the library, *Journal of Academic Librarianship*, *32*(5), 527–532.
Jinha, A. (2010). Article 50 Million: an estimate of the number of scholarly articles in existence. *Learned Publishing*, *23*(3), 258–263.
Kanai, R., Bahrami, B., Roylance, R., & Rees, G. (2012). Online social network size is reflected in human brain structure. *Proceedings of the Royal Society B*, *279*(1732), 1327–1334.
Karinthy, F. (1929). Láncszemek (Chains). In F. Karinthy. *Minden masképpen van (Everything is different)*. Budapest: Atheneum Press. Reprinted in English in M. E.J Newman, A. Barabási & D. J. Watts. *The structure and dynamics of networks*. Princeton NJ: Princeton University Press, 2006, pp. 21–26.
Kauffman, S. (1996), *At home in the universe: the search for laws of self-organization and complexity*. Oxford: Oxford University Press.
Kauffman, S. (2010), *Reinventing the sacred: a new view of science, reason and religion*, New York NY: Basic Books.
Kilmister, C. W. (2005). *Eddington's search for a fundamental theory: a key to the universe*. Cambridge: Cambridge University Press.
Lee, N. H. (1969). *The search for an abortionist*, Chicago IL: University of Chicago Press.
McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: homophily in social networks. *Annual Review of Sociology*, *27*, 415–444.
Milgram, S. (1967). The small world problem. *Psychology Today*, *1*(1), 61–67.
Miller, G. A. (1956). The magical number 7 plus or minus 2: some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81–97.
Mitchell, M. (2009). *Complexity: a guided tour*. New York NY: Oxford University Press.
Nicolaisen, J. & Hjørland, B. (2007). Practical potentials of Bradford's Law: a critical examination of the received view. *Journal of Documentation*, *63*(3), 359–377.
Pfender, F., & Ziegler, G. M. (2004). Kissing numbers, sphere packings and some unexpected proofs. *Notices of the American Mathematical Society*, *51*(8), 873–883.
Pickover, C. A. (2012). *The math book*. New York NY: Sterling.
Powell, J., Lewis, P. A., Roberts, N., García-Fiñana, M., & Dunbar, R. I. M. (2012). Orbital prefrontal cortex volume predicts social network size: an imaging study of individual differences in humans. *Proceedings of the Royal Society B*, *279*(1736), 2157–2162.
Priego, E. (2012). Insights from "The Numbers Game", lecture by Blaise Cronin. *Altmetric blog*. Retrieved 1 July 2014 from http://www.biomedcentral.com/meetings/2000/foi/transcripts/cronin
Roberts, S. B. G., Dunbar, R. I. M., Pollet, T., & Ruppens, T. (2009). Exploring variations in active network size: constraints and ego characteristics. *Social Networks*, *31*(2), 138–146.
Rousseau, R. (2010). Informetric laws, *Encyclopedia of library and information sciences (3rd ed.)*. London: Taylor and Francis, 1:1, 2747–2754.
Savolainen, R. (2009), Small world and information grounds as contexts of information seeking and sharing. *Library and Information Science Research*, *31*(1), 38–45.
Schnettler, S. (2009). A structured overview of 50 years of small-world research. *Social Networks*, *31*(3), 165–178.
Schultz-Jones, B. (2009). Examining information behaviour through social networks: an interdisciplinary review, *Journal of Documentation*, *65*(4), 592–631.
Shirky, C. (2003). A group is its own worst enemy. *Clay Shirky's writings about the Internet*. Retrieved 1 July 2014 from http://www.shirky.com/writings/group_enemy.html

Six degrees of separation. (n.d). In *Wikipedia*. Retrieved 29 June 2014 from http://en.wikipedia.org/wiki/Six_degrees_of_separation

SIMS (2003). *How much information?* School of Information Management and Systems, University of California Berkeley. Retrieved June 30 2014 from http://www2.sims.berkeley.edu/research/projects/how-much-info-2003

Stock, W. G., & Stock, M. (2013). *Handbook of information science*. Berlin: Walter de Gruyter.

Tegmark, M. (2014). *Our mathematical universe*, London: Allen Lane.

Tetlow, P. (2007). *The Web's awake: an introduction to the field of web science and the concept of web life*. Hoboken NJ: Wiley.

van Mierlo, T. (2014). The 1 % rule in four digital health social networks: an observational study. *Journal of Medical Internet Research*, *16*(2), e33. Retrieved 30 June 2014 from http://www.jmir.org/2014/2/e33

von Baeyer, H. C. (2003). *Information: the new language of science*. London: Weidenfeld & Nicolson.

Watts, D. J., & Strogatz, S. H. (1998), Collective dynamics of 'small world' networks. *Nature*, *393*(6684), 440–442.

Wiesstein, E. W. (n.d.). Kissing number. *Mathworld*. Retrieved 1 July 2014 from http://mathworld.wolfram.com/KissingNumber.html

Wellman, B. (2012). Is Dunbar's number up?. *British Journal of Psychology*, *103*(2), 174–176.

Zhou, W-X., Sornette, D., Hill, R. A., & Dunbar, R. I. M. (2005). Discrete hierarchical organization of social group sizes. *Proceedings of the Royal Society B*, *272*(1561), 439–444.

Zhu, D., Wang, D., Hassan, S., & Haddawy, P. (2013). Small-world phenomenon of keywords network based on complex network. *Scientometrics*, *97*(2), 435–442.