Evaluation and Program Planning 60 (2017) 176–185

Contents lists available at ScienceDirect

Evaluation and Program Planning

journal homepage: www.elsevier.com/locate/evalprogplan

Hindsight is 20/20: Reflections on the evolution of concept mapping

William M. Trochim, Professor, Policy Analysis & Management

435 Kennedy Hall, Cornell University, Ithaca, NY 14853, United States[1,2]

ARTICLE INFO

Article history:
Received 4 August 2016
Accepted 22 August 2016
Available online 28 August 2016

Keywords:
Concept mapping
Construct validity
Pattern matching
Theory of conceptualization
Multidimensional scaling
Cluster analysis
Bridging analysis
Go-zone plot
Bibliometric analysis

ABSTRACT

This paper considers the origins and development of the concept mapping methodology, a summary of its growth, and its influence in a variety of fields. From initial discussions with graduate students, through the rise of the theory-driven approach to program evaluation and the development of a theoretical framework for conceptualization methodology, the paper highlights some of the key early efforts and pilot projects that culminated in a 1989 special issue on the method in *Evaluation and Program Planning* that brought the method to the attention of the field of evaluation. The paper details the thinking that led to the standard version of the method (the analytic sequence, "bridging" index, and pattern matching) and the development of the software for accomplishing it. A bibliometric analysis shows that the rate of citation continues to increase, where it has grown geographically and institutionally, that the method has been used in a wide variety of disciplines and specialties, and that the literature had an influence on the field. The article concludes with a critical appraisal of some of the key aspects of the approach that warrant further development.

This paper is decidedly a subjective one, designed to provide the kinds of reflections that one doesn't typically get to make in a refereed journal article. I take it as a bit of a personal prerogative, as the originator of the methodology that is the subject of this volume, to provide a biased portrayal of the origins of the approach as best I recollect it. It may be a fool's errand to attempt such retrospectives, for whom do they serve and to what end? But as a way of fleshing out the record and telling the story of the beginnings of a comparatively minor endeavor in the history of contemporary social research methods, perhaps it will be instructive or helpful to those who use the approach or to others who are embarking on their own methodological journeys, to see that such evolutions can occur and how from seemingly inconsequential interactions may come the substance of careers and the means to affect the work and lives of many more. This is a necessarily fallible and incomplete record, just the briefest sketch of a much richer experience, a retelling that if told again another time would undoubtedly highlight different events, include other actors and emphasize different story lines. I apologize in advance if this telling excludes something important to you or fails to mention a person who had an important impact on this evolution. It does not diminish the contributions of those who may be overlooked but rather reflects on the imperfect capabilities of the author.

1. The origins of concept mapping (1982–1986)

I remember the original conversation about concept mapping as clearly as I remember something that happened yesterday. At least, I think I do. It's always difficult with reconstructed memory to know for sure whether what is recalled is accurate or the recreation of what one wishes had occurred. That said, my recollection is that the initial conversation was in 1982 with a graduate student, Dorothy Torre, who was in the next department over from mine, the former Department of Human Development and Family Studies at Cornell University. Dorothy was one of a number of students who were working on the topic of "empowerment" with colleagues that included Mon Cochran and Steve Hamilton and who were all working in connection with Urie Bronfenbrenner, then late in his career. She was relatively early in her graduate work, and was trying to develop a Master's thesis project. She met with me to discuss some now unremembered aspect of research design, but I recall that I struggled to understand what she and her colleagues meant by the construct "empowerment". Was it a noun, a verb, or both? Did it mean "giving power to" or taking it directly? And what would either of those mean? How would we know empowerment if we saw it? I kept steering her methodological questions back to the issue of construct validity and the "preoperational explication"

[1] http://www.socialresearchmethods.net/
[2] http://www.human.cornell.edu/bio.cfm?netid=wmt1

http://dx.doi.org/10.1016/j.evalprogplan.2016.08.009
0149-7189/

(in Campbellian terms) of empowerment. In this hazy recreation of events, I recommended that Dorothy and I brainstorm what empowerment might mean, generating short snippets of ideas or thoughts about it. And, we were quite prolific. We must have generated several dozen ideas during our conversation. We adjourned that day with Dorothy having the "homework" assignment of generating as many more ideas as she could think of that described some aspect of empowerment. I think at the time my thought was that she might develop an empowerment scale. The first step of any such scaling attempt would be the generation of a large set of potential scale items.

Several days later we reconvened to look over her list. She must have had 50–60 items, conscientious graduate student that she was. We did what many people would naturally do if faced with a long list of items – we attempted to classify them into a fewer number of themes. Each of us did this separately and quickly discovered that while there was some overlap, our two thematic classifications were different. And, we were troubled by the fact that we could each have done several distinct thematic classifications, each one as conceptually legitimate as the previous. I believe I gave Dorothy more "homework" to do several more distinct classifications of her statements – as many as four or five. And I recall doing several more myself.

When we looked at our multiple "sorts" it was clear that we had a methodological and conceptual problem. How can we make sense of multiple classifications of the same set of items? Is there a way methodologically to integrate them? We had clearly been thinking about using the statements as the basis for developing a unidimensional scale, but we were struck by the multi-faceted nature of our multiple classifications of the same set of items. It seemed to me that a construct as complex as empowerment had to be multidimensional in nature. Our independent classifications supported that notion. So, that meant we needed to explore multidimensional scaling, not its unidimensional cousin. I had I think one or two class sessions in graduate school that covered the technique of multidimensional scaling (MDS), as part of a semester-long course at Northwestern on multivariate statistics that Will Shadish taught while doing a post doc there. I can't say that I remembered much about MDS except that it required a square matrix of similarities (much like factor analysis requires a matrix of correlations) and that it could generate an N-dimensional scale.

The big problem that we had was how to combine our separate classifications in a way that provided a matrix of similarities for MDS. At this point, my retrospective narrative becomes a blank. I don't remember how I came up with the idea of how to aggregate independent sorts of the same objects. I vaguely recall struggling with ways to compare our multiple sorts of the ideas, trying to set up tables of various types to summarize the similarities and differences. But to get from that struggle to the method of sort aggregation required seeing several things more or less simultaneously: that the square similarity matrix needed for MDS had to have as many rows and columns as there were elements (i.e., ideas); that one classification or sorting could be represented in this matrix with perfect accuracy using binary (0,1) values such that any cell would have a value of 1 if the row and column statements were placed together in a theme and a 0 otherwise; and, that these separate binary matrices could be added together to give a more variegated aggregate estimate across multiple sorts of the similarities among the statements. This was not exactly the most obvious thing to do with our multiple classifications. In effect I had created a method of aggregating sort data that would be so clearly articulated years later by Weller and Romney (1988) and that undoubtedly was independently reinvented by any number of others. I was unaware at the time of pioneering work done on the method of sorting (Rosenberg & Kim, 1975) but I doubt it would have made any difference (their method of aggregating sort data was distinctly different from the simpler approach I created and that Weller and Romney described). In any event, that was the key insight that led to concept mapping. The sequence of brainstorming, sorting, sort aggregation and MDS became and still remains the core of the methodology. Torre went on to complete a Master's thesis and then a doctoral dissertation that featured this methodology. To give you a flavor of the nature of the approach at that point, and for purposes of the historical record, her final point map is shown in full in Fig. 1.

2. Construct validity and pattern matching, 1983–1985

In the early 1980s, experimental and quasi-experimental approaches to evaluation were considered the "dominant paradigm" for the field. The randomized experiment was considered the most rigorous evaluation design, at least with respect to internal validity (Cook & Campbell, 1979). But an experiment typically told us only whether an operationalized program had a detectable effect on operationalized outcomes. It didn't typically tell us whether the program as operationalized reflected well the program that it was theoretically intended to reflect. It didn't typically tell us whether the measures reflected well the outcomes that we thought we were measuring. And it typically didn't tell us

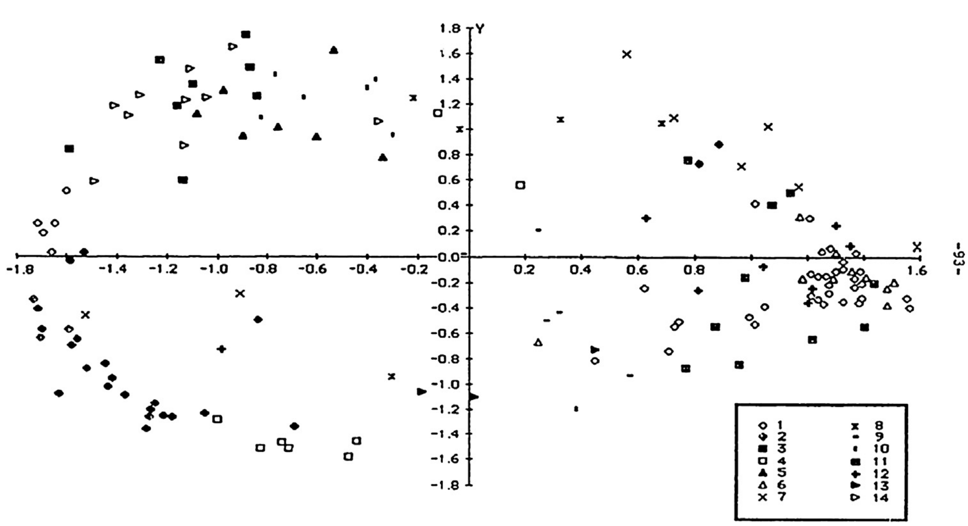

Fig. 1. One of the original concept maps, from Torre's 1986 dissertation (but produced several years earlier)(Torre, 1986 p. 93).

anything about the mechanism or process by which the purported program might have brought about the hypothesized effect. In short, experiments and many quasi-experiments were weak with respect to construct validity (Cook & Campbell, 1979).

I was not alone at the time in being troubled by the idea that the randomized experiment treated the program like a "black box" that functioned as a whole unit whose mechanisms we cared or knew little about. The reaction of the field to this "black box" aspect of experimentalism was most evidenced in the idea of program theory and the development of what has become known as the "theory-driven" approach to evaluation (Chen & Rossi, 1983, 1984, 1990). Huey Chen and Peter Rossi made the case that we needed to open up the "black box" and develop and test theories about how programs worked. But where were these theories going to come from? Chen and Rossi initially argued that "Often enough policymakers and program designers are not social scientists and their theories (if any) are likely to be simply the current folklore of the upper-middle-brow media" (Chen & Rossi, 1984 p. 339), a position whose rough and somewhat elitist edges they softened in subsequent discussions. They continued "The primary criterion for identifying theory in the sense used in this article is consistency with social science knowledge and theory. Indeed theoretical structures constructed out of social science concerns may directly contradict what may be the working assumptions of policymakers and program designers."

While I agreed with the central emphasis of the theory-driven approach, I resisted elevation of social science as the primary source of theory. It is true that most policymakers and program designers do not operate in the rarified academic milieu that is steeped in theoretical speculation. But, as a psychologist, I didn't think that was because they lacked the cognitive structures that reflected complex implicit theories; I thought it was because they typically don't have the time, interest in or experience with articulating the implicit theories that must necessarily drive their decisions about programs. And, I realized when reading Chen and Rossi's original work that approaches like the structured conceptualization method I had been developing and starting to use could enable program stakeholders to articulate their implicit complex theories, and their thinking about constructs, in a way that would help address the needs of program theory and enhance construct validity.

In response to the initial theory-driven work I authored the first publication (Trochim, 1985) that mentioned and described the concept mapping methodology. Most of the essential ingredients were already in place at this point: the participatory nature of the process; the basic steps for accomplishing it; the use of unstructured sorting; the central role of multidimensional scaling; and, even the idea that the maps might provide the foundation for a pattern matching approach that could be used to enhance construct validity and help address one of the major weaknesses in the dominant experimental and quasi-experimental framework.

3. The theory of conceptualization 1982–1986

While the Trochim (1985) paper was being drafted, I was also working with Rhoda Linton, then a graduate student, on a publication that would introduce the new methodology (Trochim & Linton, 1986), and it was only publication happenstance that led this introductory paper to come out after the one on pattern matching, validity and conceptualization. One of the most intriguing aspects of this effort, and the part that has been least fulfilled in subsequent work, was our overly-ambitious attempt to situate the newly emerging concept mapping method within a more general comprehensive framework of structured conceptualization approaches. We were attempting – without the necessary foundation in cognitive psychology and before the rise of the cognitive neurosciences – to articulate a general theory that would describe *any and all* conceptualization methods, from the kind of everyday thinking we do individually in our conscious minds to the type of group thinking we were implementing in concept mapping.

This theory imagined that all conceptualization necessarily involved three distinguishable processes, the generation of ideas (G), the structuring of the ideas (S) and the representation (R) of them. In most everyday thought, individuals don't enact these processes sequentially or even consciously. Configurations of ideas more or less spontaneously occur to us. But we postulated that even in these cases there had to be some underlying cognitive steps, however unconscious they were. And, we suspected that even as individuals we might be able to enhance how we think by making those unconscious steps more conscious, that is, by consciously employing a more "structured conceptualization" approach.

We also made a distinction between three types of "entities" that could be engaged in enacting each of these three steps: an individual (i) person; a group (g); or an algorithm (a). This last may seem at first a bit odd. We were suggesting that algorithms – structured sets of rules – could be used, however imperfectly, to accomplish some of the same tasks that individuals or groups could do. We had in mind here the kinds of emerging computer technologies that predated the internet by more than a decade but that already looked promising for generating or accumulating ideas (e.g., algorithmic machine-driven content analysis of texts (Krippendorf, 1980, 2004; Stone, Dunphy, Smith, & Ogilvie, 1966)) or for structuring or organizing them (e.g., the algorithms built into any cluster analysis (Anderberg, 1973; Everitt, 1980)).

We also postulated that conceptualizations could be represented in one or more of three forms: verbally (V), pictorially (P) in graphic form, or mathematically (M) in the form of an equation. For example, take the classic equation from Einstein's theory of relativity, $E = mc^2$. That's its mathematical form. We could describe it verbally as "energy is equal to mass multiplied by the square of the speed of light" and would of course want to add considerable verbal definition to each term. Or it is possible to represent the equation in a variety of graphic forms (all of which lead one to conclude that the apparently simple formula is far more subtle and less intuitive than it might initially seem).

Our general model for conceptualization could give rise to a large set of combinations, each of which stood for a distinct way to conceptualize that yielded a specific representational form. For example, to represent the instantaneous spontaneous thoughts that occur in everyday life, we might use the equation:

$$(GSR)_i \rightarrow V$$

which essentially means that the three presumptive process steps GSR occur simultaneously (i.e., with no conscious separation, indicated by grouping all three processes within single parentheses) in an individual (indicated by the subscript i) and yield a thought that can be expressed verbally. If an individual consciously first "brainstorms" a set of related ideas (as when making a grocery list) and then organizes that set (perhaps by aisle or section in the grocery store) and then represents it in an organized written list, we might show this as:

$$(G)_i(S)_i(R)_i \rightarrow V$$

In contrast, if a group of people are working on some task and brainstorm or generate in some way a list of ideas, and one of them subsequently organizes the ideas and represents them in a hand-drawn picture that shows the interrelationships, we could depict this as:

$$(G)_g(S)_i(R)_i \rightarrow P$$

What, one might ask, would be the value of this kind of conceptual theory and why would I be attempting yet again to foist it upon the world in this paper? One of the most compelling reasons for the development of this theory of conceptualization was the desire to situate the type of structured conceptualization I was creating within a potentially universal set of methodological approaches to conceptualizing. With that in mind, it was possible to describe the emerging concept mapping method with the equation:

$$(G)_g(S)_i(R)_a \rightarrow P$$

which, succinctly stated, means that a group of people generates the ideas, individuals structure or organize them, and that an algorithm (in this case a sequence of multivariate analyses) represents them in pictorial form, in this case as a map. The separation of each step within their own parentheses conveys that each is a distinct part of the whole process.

Needless to say, this general theory of conceptualization didn't survive past this initial publication, although I hold an unfounded optimism that this current attempt at resurrection may spur others to contemplate how it might be usefully applied and extended. It has value, as Donald T. Campbell used to say, for the "historical indexicality" it provides by showing how at the earliest stage of its development we construed concept mapping as a specific approach to a broader class of cognitive processes, something that tends to have gotten lost over the decades since its inception. Concept mapping was, from the beginning, integrally connected to the ideas of concepts and constructs, the psychology of cognition, and the conscious attempts to develop methodologies to enhance our ability to formulate and manipulate them.

4. First projects, 1983–1986

At the time concept mapping was developed, the Department of Human Service Studies at Cornell had what I immodestly believe was the finest PhD program in Evaluation of that era. My colleagues, Charles McClintock (who started the program), Jennifer Greene (who came around 1983) and I (who showed up in 1980) were fortunate to have a large cohort of graduate students who had an abiding interest in evaluation and an openness to looking at it from fresh perspectives. Almost immediately after working through the beginnings of the concept mapping approach with Dorothy Torre, I began introducing the approach to other students with whom I worked.

One of these graduate students, Rhoda Linton, was doing a small project with a group of local youth workers who had created a summer camp for high school students that was designed to encourage the youth to be more aware of different groups and cultures and, especially to raise issues of class, race, gender and sexual orientation. We decided to apply this new approach to see if it could help the staff better understand their goals for and the expected effects of the program. Rhoda brought to this endeavor an especially valuable perspective on participatory and collaborative evaluation and a background in community organizing that greatly enhanced the group process side of the fledgling method. The map that resulted looks incredibly crude by today's standards but provided the foundation for continuing work on the method that eventually led to the publication that first formally introduced concept mapping (Trochim & Linton, 1986). This paper laid out the GSR general theory of conceptualization described above, presented the steps in the concept mapping methodology (in terms of the three GSR steps), and provided two of the earliest examples of its use.

The first example occurred when I was approached by Cornell University officials who ran the Division of Campus Life (DCL), an entity that encompassed such impossibly disparate departments as the Campus Store, Religious Works, Public Safety, the Dean of Students, Transportation, Dining, Health Services, and the International Student Office, among others – a strangely nostalgic administrative amalgam that would be implausibly complex in a contemporary university. They asked if I would be willing to help them with a strategic planning process. Even though I had virtually no formal training in strategic planning, I was a faculty member in the concentration of Program Evaluation and Planning and presumably knew something about such things. I was also a young, tenure-track faculty member, being asked to volunteer for university service. So, my thinking was that if I was going to do this work for free I may as well try to turn it into something that would contribute to my teaching and research. For them the price was right. They were willing to let me and several graduate students use this as an opportunity to explore the new concept mapping method. Primary among these students was Rhoda Linton whose thoughtfulness regarding the process of engaging the different stakeholders had an important impact on this project and the development of the methodology generally.

The project was massive relative to the first two efforts with Dorothy Torre and the multicultural awareness camp (and even relative to many subsequent ones), involving 75 staff members over four major in-person meetings who brainstormed 876 original ideas that were ultimately synthesized to a final set of 137 items that were sorted by 43 people and ultimately mapped. When I look back on that project one of its most impressive features is the extensive and largely successful effort made to connect the map results to subsequent strategic planning and evaluation.

One of the staff who participated in the DCL study was an administrator at Cornell's University Health Service (UHS) which was also contemplating doing their own internal strategic planning and evaluation effort and were evidently impressed enough with the DCL project that they asked me to facilitate something comparable in their organization. In this project, again done collaboratively with a group of graduate students, 77 people brainstormed 315 ideas. Rather than synthesize these we simply randomly selected 100 of the ideas which were subsequently sorted by 69 participants and mapped.

One of the more striking memories I have of this project, and a particularly telling one given the participatory nature of the methodology, was of the physician-director of the UHS who approached the whole endeavor with fairly public skepticism. In the final massive interpretation session, virtually the entire UHS staff was present. We broke them into small subgroups to come up with potential labels for the clusters on the map. In the process of going around the room and asking about one of the clusters of statements that had to do with the rather pedestrian topic of staff meetings, one subgroup suggested the cluster label "Eat not Meet" and the entire room erupted with applause and laughter. We subsequently learned that the problem was that increasingly, under the guise of staff development, the UHS had been scheduling training sessions during lunchtime so staff would be able to attend. But this also meant that the staff had fewer and fewer opportunities to socialize during their brief lunchtime breaks. The director, who was sitting at a table on the raised stage, visibly reddened at the spontaneous eruption of the group and then, to his credit, stood up and said that this issue would be addressed immediately – and then added that every single idea on the map would also be addressed. I thought this was an incredibly overambitious promise until he subsequently explained what he meant. The strategic planning group, he said, would go through each statement on the map and make a determination about whether it could be done now, done later, or was not possible to address, and would provide a public explanation of each decision to the entire staff. I still see this as a model of how to stay true to the participatory collaborative nature of the method while acknowledging the realistic limitations on implementation.

5. Widespread field testing, 1983–1989

The years from 1983 to 1989 were the concept mapping equivalent of the Cambrian explosion with the graduate students in evaluation at Cornell. Beginning with Linton's (1985) and Torre's (1986) dissertations, a large number of students in the program engaged in special independent study projects, theses and dissertations that incorporated the method. In fact, we used to joke that the approach was like a "dissertation in a box" because it provided a general methodology that was useful for the early exploration of virtually any research topic. It was an excellent empirical complement to the traditional literature review and, since it was essentially new, required only a brief review of the methodological literature for justification. The fact that it was participatory, involved a standardized method for collecting data and utilized a scientifically credible sequence of sophisticated multivariate analyses made it especially good for at least a central chapter in a thesis write-up. This period is described well in the article by Donnelly included in this special issue (Donnelly, this issue).

There were several developments in the methodology during this period that warrant discussion. In the first few concept mapping projects we used the sorting-based similarity matrix as input independently to both the MDS and cluster analysis routines. I struggled with cluster analysis (and still do) as a method because, unlike with the concept of Euclidean distance in MDS, there was in cluster analysis no clear agreed-upon definition of the underlying concept of a cluster and, consequently, there were a wide variety of different clustering algorithms that led to sometimes considerably disparate results. But the major problem the initial analyses posed was that all clustering algorithms yielded results that were at odds with the placement of the ideas on the map by MDS. Because both analyses took the same similarity matrix as input, it was frequently the case that points with the same cluster could wind up in very different locations on the map, complicating the interpretation considerably. Two decisions resolved this problem. The first was the insight that we could make the analyses sequential, using the two-dimensional coordinates from MDS, instead of the similarity matrix as the input to the clustering algorithm. In effect, this made the cluster analysis dependent upon or conditional on the MDS. This made sense because, as mentioned earlier, MDS was the stronger analysis statistically. The other decision, which flowed from that, was to use Ward's method (Ward, 1963) as the clustering algorithm, simplifying considerably the confusing choice of clustering methods. Ward's method was an especially appropriate complement under these circumstances because it assumes Euclidean distances as the input (it is based on the sum of squares of the distances between all points in potential candidate cluster merges).

A second development was the creation of the "bridging index". This index, described in detail elsewhere (Kane & Trochim, 2006), was created in order to help in the interpretation of the results of a concept map. Since MDS had to place a point somewhere on a map, we needed an algorithm that would tell us whether any given point was positioned where it was because that's where it belonged geographically (it was sorted with other points that wound up located near it) or because it had to be located in some intermediate location when the points it was sorted with were in two or more disparate locations. Because an index in Euclidean distance units made little sense interpretively, I decided to relativize or normalize the index so that in every map there would always be at least one point that had a 0 and a 1 value. We created the index so that low values signified a point that described its local area of the map and higher values indicated that the point "bridged" between multiple areas. We could have reversed that by subtracting each value from 1 (and in contemporary analyses this option is routinely offered) and the index could be considered a relative weight where higher values indicate that the point is more representative of its locational area on the map (what we have since labeled an "anchor" index). Regardless of the order, there are two things that warrant attention about this index. First, it is unique to this particular sequence of analyses which uses sorting data (and the resulting similarity matrix) as input because the index calculation integrates the sort input with the MDS map results. That is, this index would not make sense for standard MDS dissimilarity matrix input. Second, it really did help in the interpretation of clusters. By listing statements within cluster in ascending order of bridging, one could subjectively weight low bridging (i.e., high "anchor") points in determining the content label for that area of the map.

The third innovation that occurred in this period was the inclusion of pattern matching, and particularly the pattern matching graph, as an integral part of the analysis whenever ratings were done. This graphic and its associated correlational analysis came from work that I had done as a graduate student with Donald T. Campbell on regression artifacts. Campbell preferred to use what he termed a "pair-link diagram" (Campbell & Kenny, 1999) to the traditional bivariate plot because it illustrated well the nature of regression to the mean. I saw this graphic tool as especially useful for exploring relationships in concept mapping results that a traditional bivariate plot could not easily achieve. Although we could do pattern matching at the statement level in concept mapping, it was much more instructive to use the cluster-level rating averages. The averages for clusters constitute a two-level hierarchical aggregation (averaging across participants for each statement and across statement averages for each cluster) and are by definition more stable (less error) than the averages for statements. As with the bridging value, because this analysis relies on both the rating data and the clustering results (which themselves are conditional on the MDS coordinates and, in turn, on the sorting data), this approach only makes sense for the relatively unusual data structure in concept mapping. I believe Campbell thought that the pair-link diagram was a unique invention of his (Campbell & Kenny, 1999, p. 12) and have learned only in the past few months that his was an independent recreation and that it originated in the 1880s (d'Ocagne, 1885; Hewes & Gannett, 1883) where it was called a "parallel coordinates" graph. It has become quite popular in the contemporary discussions of the fields of data visualization and data mining (Inselberg, 1997, 2009) where it is seen as a useful tool for conveying multidimensional information. The method continues to undergo considerable development there, including the exploration of three-dimensional pattern representations (Achtert, Schubert, Schubert, & Zimek, 2013).

6. The software, 1982–1989

No discussion of the origin of concept mapping would be complete without consideration of the development and availability of software for accomplishing the methodology. Since manual computation of most multivariate statistical analyses is all but impossible (although I do remember hearing of the legendary pioneers in factor analysis who did their calculations using manual calculators), the development of concept mapping was dependent on the availability of suitable software. The first few concept mapping projects made use of the ALSCAL module in the SPSS statistical package (SPSS Inc., 2005) but in those days this still required a fair amount of effort and some special programming, especially for the computation of the similarity matrix from the sorting data. But concept mapping had the good fortune of beginning at the advent of the microcomputer revolution, and I quickly determined that a stand-alone program that enabled the entire sequence of calculations and the creation of the graphic results would be desirable.

In 1982 I wrote the first version of The Concept System software in Apple Basic on an Apple II microcomputer equipped with the full complement of 64 K of built-in memory and two 5.25" floppy disk drives (by comparison, my current iPhone has 64GB of built-in memory which, by my calculation is one million times larger). There was so little memory to work with that the entire data for a concept map could not fit in active memory. In order to compute a map I had to read and write the data to floppy disk for every iteration of the MDS algorithm. The simplest of maps took about 12–14 hours to compute. By the mid-1980s I had gotten one of the first IBM Personal Computers (PCs) that used Microsoft's new Disk Operating System (DOS) and the Microsoft Basic programming language and ported the new software to that platform. It wasn't much faster or bigger than the Apple II but at that point the PC was an open-platform architecture that was more promising for developers to use.

The early versions of the software had several notable characteristics. Initially they included not just idea maps but also person maps. These were computed in a rather simplistic fashion by calculating a square similarity matrix of people based on the similarity of their sorts and then running an MDS on that matrix. I employed it on a single project that involved 12 directors of major departments in the Cornell DCL. The person map that resulted had 10 of these people clustered together on one side of the map, one person on the other, and one in the center. When I showed the unlabeled map to the group they all laughed because they immediately knew who was located where. The lone person on one side was the strongest personality in the group and had been trying to persuade and, to some extent, to dominate the discussion so that things would work out the way he wanted. No one was comfortable standing up to him in the conversation. The person in the middle was the chair and diplomat of the group. All of the rest of the group were clustered on the opposite side. The good news was that, without even having a discussion about it, the 10 leaders who clustered together realized that they had the majority and didn't subsequently worry about the potential conflict with the dominant leader. The bad news, from my point of view, was the powerful nature of these types of maps and the potential for them to be misused, especially in identifying people who were "outliers" or who were in the minority or didn't think like the majority of the group. I thought the potential for this would be so irresistible, and the potential consequences so negative, that I removed the analysis.

The other analysis that was in the original software was cognitive mapping (Axelrod, 1976), essentially a form of directed causal graph. It used the mathematics of reachability matrices and operated only on the cluster level. Essentially the user would tell the software which clusters "caused" which other clusters and, from this input, it was possible to compute a binary square asymmetric matrix where a 0 meant there was no causal relationship and a 1 meant that the row cluster caused the column cluster. This information could be directly overlaid as a causal map (directed or with arrows) on top of a cluster concept map. In the end, I omitted this analysis, not because it was not interesting or useful but because it was an analysis distinct from the concept map itself, added significantly to the burden of describing the entire methodology to participants, and made the entire process more complex. Ironically, in the past decade I have circled back to that approach in my recent work developing a Systems Evaluation Protocol (Cornell Office for Research on Evaluation (CORE), 2009; Trochim et al., 2012) that focused on causal pathway modeling (Urban, Hargraves, & Trochim, 2014; Urban & Trochim, 2009).

7. The 1989 *Evaluation and program planning* special issue

By 1986, the Cornell graduate students in evaluation and I had accomplished at least twenty concept mapping projects and were exploring a wide range of methodological issues and potential substantive applications of the method. For the Annual Conference of the American Evaluation Association in 1986 in Kansas City we presented a panel entitled "Conceptualization for Evaluation and Planning" that described our collective work to that point. A unique feature of that panel was that not only did we have all of the panel papers actually written before the presentation but we also bound them together into a single document which we called "Structured Conceptualization: A Resource Guide and Progress Report" (Trochim, 1986) that included a draft bibliography of the topic, the in press copy of the Trochim & Linton (1986) paper and descriptions of twenty structured conceptualization examples we had done. The editor (then and now) of the journal *Evaluation and Program Planning* (EPP), Jon Morell, was in the audience for that presentation and approached me immediately afterwards to suggest that it would be a relatively simple matter to turn our resource guide into a special issue of the journal. Over the subsequent two years we worked on the special issue which was ultimately published in 1989 (Trochim, 1989b), incorporating five new project summaries (and removing several of the older ones that had been published elsewhere), and I wrote an introduction (Trochim, 1989c) and concluding paper (Trochim, 1989a). This issue, more than any other event to that point, marked the introduction of the methodology to the field. It was the memory of that publication, twenty-five years later, which led Dan McLinden and me to embark on the development of the retrospective special issue that you hold in your hands (or more likely view on an electronic device that could not have existed in 1989).

8. Beyond 1989

At this point, I've accomplished the primary purpose of this retrospective – describing the origins and early years of the methodology. In some sense, the method has not evolved much from its 1989 form. Certainly the graduate students of that era would readily recognize the current method and the products of concept mapping. With relatively few exceptions (which will be discussed below) there has been little in the way of major innovation in the methodology itself that could compare with the first seven years (and especially the first two years).

One of the methodological innovations since 1989 worth noting was the introduction of the "go-zone" graph, essentially a specific type of bivariate plot, usually at the statement level, that enables one to examine a cluster or the entire set of statements relative to any two ratings (either two different ratings or one rating by two different subgroups). One of the earliest uses, and the one that gave the graph its name, was a cluster-level bivariate plot of statements that shows importance and feasibility ratings on the two axes. Because the go-zone plot also overlays "crosshair" axes, it enables one to subdivide the space into quadrants that show the statements above or below the average importance and feasibility. The quadrant representing statements above average in both importance and feasibility represents what we termed the "go-zone" (from which the graph gets its name) – the area you might focus on that would be most likely most immediately productive or fruitful. Even though bivariate plots are among the most widely used in research, the go-zone special case illustrates a nice adaptation that is unique to this methodology because it typically is built on the clusters that emerge from the map.

One of the major changes that has occurred over the life of the concept mapping methodology was the evolution of the software, and particularly the introduction of the internet. Until the mid-1990s, the method was accomplished primarily either through face-to-face meetings or by distributing materials to participants through mail and, later, email. Data collection was entirely manual. Brainstorming typically was done with groups in a room, sorting required the printing of card decks and ratings were distributed essentially as paper surveys. The data were manually entered into the software. In 1993, I did a sabbatical year in Chicago and somehow made my way out to the training campus for Anderson Worldwide (then the umbrella organization for Anderson Consulting and Arthur Anderson). There I met Dan McLinden, ever since my collaborator on all things concept mapping (and the co-editor of this volume) and, over a two-year period we worked with the Anderson programmers to put together the first version of the software for Windows. This software enabled multiple participants over a local area network to brainstorm, sort and rate. It was written in Visual Basic and had a visual interface for the sorting that enabled one, for the first time, to conduct something approximating a sort on a computer. By the late 1990s the internet had become widespread and I wrote the first web-based version of the software that enabled people to participate synchronously or asynchronously from anywhere in the world. While it remains to this day challenging with a computer interface to approximate anything like what one experiences when brainstorming in face-to-face groups or manually sorting cards, the advantages of being able to have participation at times convenient for them from people anywhere in the world has generally outweighed the potential loss in direct interpersonal contact.

9. The influence of concept mapping

One way to look at the post 1989 history of concept mapping is through a bibliometric analysis of the publications associated with it. To do this I did a search in the Scopus bibliometric database for all publications that cited the introductory article in the 1989 Evaluation and Program Planning special issue on concept mapping (Trochim, 1989c). This article was used as the reference point because it has been the standard citation for the methodology. Even though there were a few articles published prior to this one, once this special issue was published, this lead article became the single definitive reference for the method. As of the date of the search (February 12, 2016) there were 478 publications in Scopus that cited the introductory article. The rate of citations has continued to increase over time, suggesting that the article (and, by inference, the method) continues to grow in influence on the literature. The Scopus analysis provides some interesting insights about where concept mapping is used most heavily and who the primary users are. In terms of countries, most of the citing articles were produced in the United States as shown in Fig. 2. A bit more surprising (until one discovers the level of enthusiasm for and the unique evolution of the method there) is the high rate of use in the Netherlands which produced the second highest number of citing articles. Four of the next five countries are former British Commonwealth countries (Canada, Australia, the U.K., and New Zealand).

If we look at the top ten institutions (Fig. 3) associated with citing publications we get a more precise sense of where the center of activity has been. It's not surprising that the leading institution is Cornell University, given that's where the method was developed, or that the third highest is Concept Systems Incorporated which holds the license for the software. Beyond that, we see that the primary users in Canada were at the University of Western Ontario and the University of Manitoba. In the Netherlands, four institutions are in the top ten in terms of citing publications: the Open University of the Netherlands, the Trimbos Institute, the University of Amsterdam and Utrecht University.

The vast majority of Scopus documents that cited the article were journal articles (82.8%). Most of the documents were in the broad fields of medicine (46.7%), the social sciences (41.8%) or psychology (22.6%). The 478 documents were published in 355 separate journals with the most (N = 21) being published in *Evaluation and Program Planning*, followed by the *American Journal of Evaluation* (N = 9) and *Social Science and Medicine* (N = 8).

The 478 Scopus publications were then imported into the SciVal (Colledge & Verlinde, 2014) system for conducting bibliometric analyses. Because SciVal only goes back as far as 1996, only 426 of the publications had records in that system. One way to gauge the impact of the introductory publication is to look at the disciplinary diversity of the cited articles. The analysis showed that the 426 citing articles occurred in 316 separate journals (journal count) and that these journals were listed in 22 of the 27 main Scopus journal categories and in 149 of the 334 Scopus sub-categories, suggesting that there is high diversity of disciplines represented in this literature. In terms of secondary reach, the 426 publications were in turn cited by 6402 other publications. In addition, the 426 citing publications represented 116 separate countries, suggesting

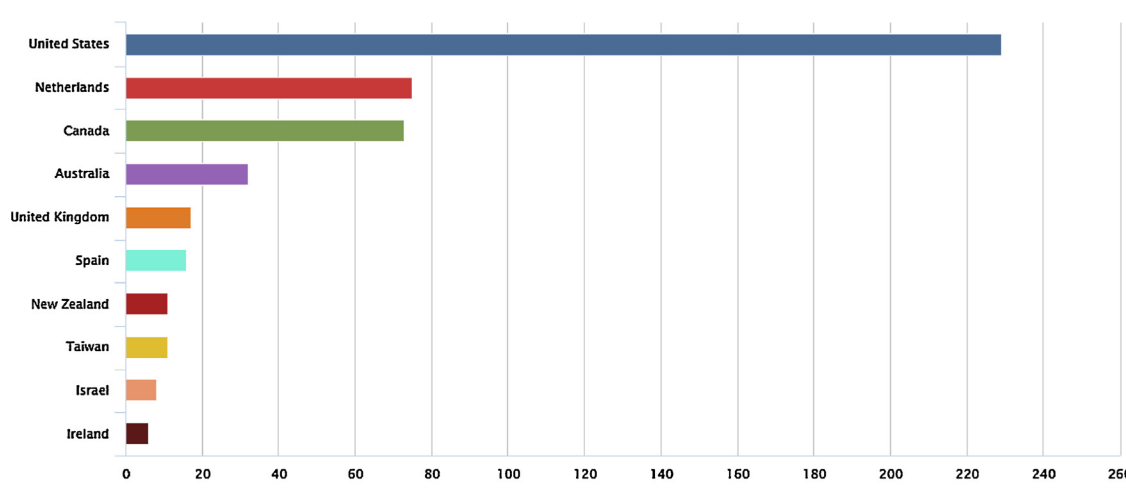

Fig. 2. Top 10 country affiliations of publications citing introductory article (Trochim, 1989c).

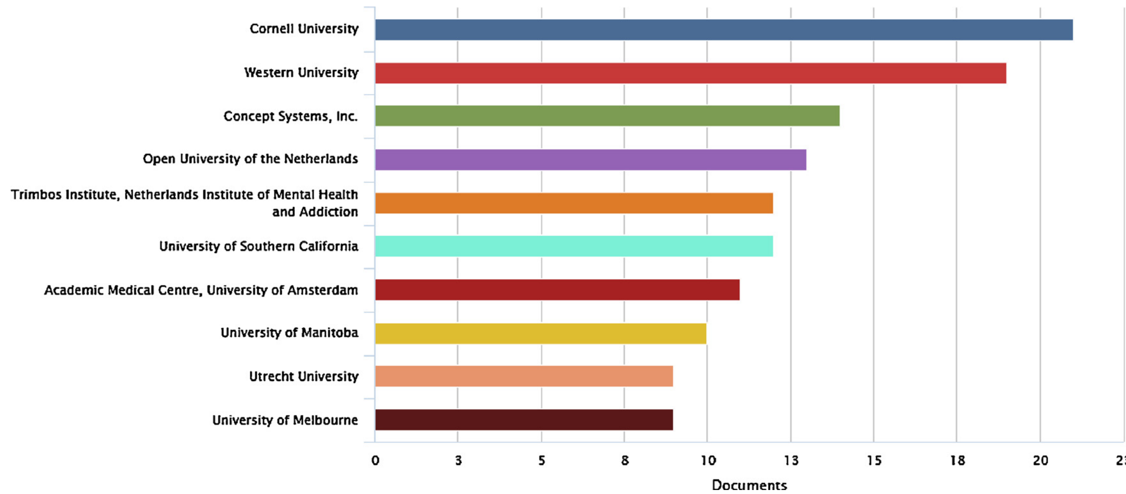

Fig. 3. Top 10 institutional affiliations of publications citing introductory article (Trochim, 1989c).

considerable geographic diversity in the recognition of the method. Overall, 85% of the citing publications received at least one citation.

One of the most important bibliometric indicators is the Field Weighted Citation Impact (FWCI) measure which "indicates how the number of citations received by an entity's publications compares with the average number of citations received by all other similar publications" (Colledge & Verlinde, 2014 p. 61). To compute the FWCI, the SciVal system estimates the number of citations one would expect for a similar document type (e.g., journal article) in the field that journal is associated with (weighted for multiple fields if the journal spans across them). A FWCI of 1.0 indicates that the set of articles has received the number of citations expected for articles like them in their fields. The results show that overall the set of 426 articles received 1.46 times the number of expected citations and in only 3 of the 18 years from 1996 to 2013 did this index fall below 1.0. 13.4% of the citing articles are in the top 10% of the journals worldwide. What all of these bibliometrics analyses suggest is that the concept mapping methodology, as reflected in citations of this foundational publication, is widely and increasingly cited across diverse disciplines and geographical areas, suggesting that the influence of the methodology continues to grow over time.

10. Summary of the history of concept mapping

We can summarize the history of the concept mapping methodology concisely with the timeline provided in Fig. 4. In the figure, each vertical line indicates a major or notable milestone or event in the method's evolution, beginning with its initial development in 1982. The figure shows the "incubation" or field testing period from 1983 to about 1988 when the method was used in a wide variety of settings largely by an insider group or faculty and graduate students at Cornell. It shows the first mention of the method in a research publication (Trochim, 1985), the first major introduction where the method was presented as part of a larger class of approaches to structured conceptualization (Trochim & Linton, 1986), its introduction in a major panel at the 1986 American Evaluation Association conference, the publication of the 1989 special issue of *Evaluation and Program Planning* that was devoted to the method, the introduction of the software in DOS (1989), for Windows (1993), and finally for the web (1998), and the production of the major book on concept mapping (Kane & Trochim, 2006).

Fig. 4 also shows the graph of the citations of the introductory article of the 1989 special issue of *Evaluation and Program Planning* as indicated through the Scopus analysis. It is clear from the figure that the number of citations continues to rise over time.

11. Hindsight is 20/20

In this history of the concept mapping method I have attempted to provide a personal recollection of the origins of the approach, a very simple bibliometric analysis that provides some evidence about its dissemination and influence, and a brief chronology of how it was developed. Clearly, the method has become an established one that has been used in a wide variety of settings and disciplines and one could argue that its recognition and influence continue to grow. But I would prefer to end this article by noting some of the limitations of the method because it is in recognizing those that we might most be able to consider how it could be advanced even more in the future. Several issues are worth emphasizing.

First, there is the somewhat disappointing lack of use of the method in evaluation. When I originally developed it, and especially through the first several years, I thought its primary value would be as a conceptual framework for conducting evaluations. I was taken with the notion that if an evaluator had a conceptual structure like a concept map, it would be possible to make evaluative inferences that were simultaneously stronger in both internal *and* construct validity than in an evaluation that lacked such a framework. You can see this hinted at in the earliest writing such as the 1985 *Evaluation Review* paper (Trochim, 1985) that emphasized the value of concept mapping as a foundation for a pattern matching approach to causal inference. For instance, if we had a concept map that charted the conceptual domain in some context, and we decided to take action or intervene in one specific area or location on this map, then it stands to reason that we would hypothesize that if this intervention is successful we would see greater effect sizes on measures that were "closer" to this point of intervention on the map than to ones further away. The map becomes a Euclidean basis for hypothesizing patterns of treatment effects.

The most thorough and ambitious statement of this idea is in the 1989 paper (published in a different issue of *Evaluation and Program Planning* than the volume devoted to concept mapping) that focused on outcome pattern matching and program theory (Trochim, 1989d). This paper attempted to describe a generalized

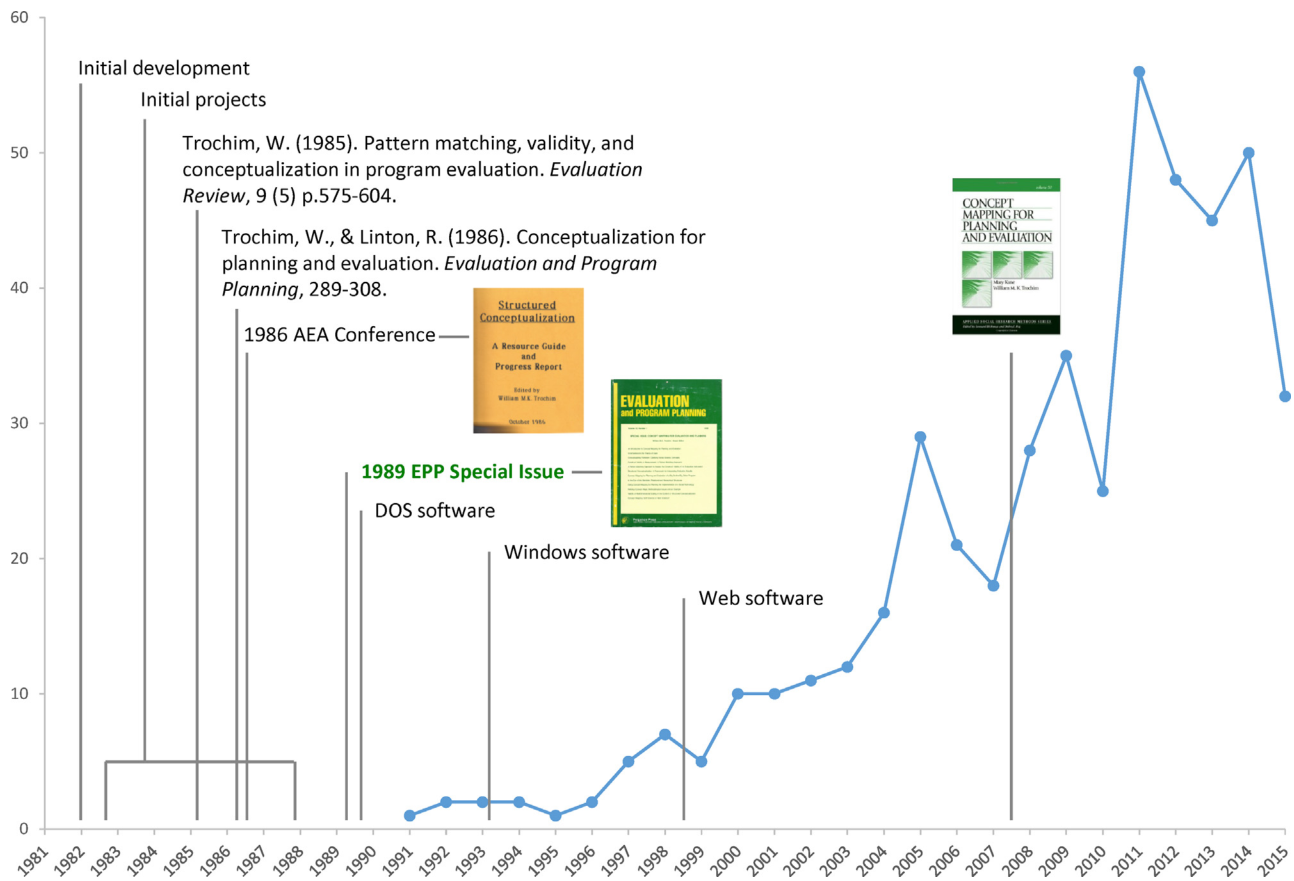

Fig. 4. Timeline of major events in the history of concept mapping and graph of citations (as of Fall 2013) of the introductory article in the 1989 Evaluation and Program Planning Special Issue.

approach to causal inference based on patterned expectations, an idea I still find to be generative and promising. But, alas, this approach has not been successfully pursued in practice probably because of the sheer complexity of it and of the considerable demands it makes on the evaluator. It assumes that one has a concept map, that the map can be operationalized through multiple measures of different elements of the map, that we can articulate the nature of the program or intervention in terms of this structure, that we can collect data with measures linked to points throughout the map, and that we can test the multivariate hypothesis that results. In most evaluation contexts in order to accomplish this it would be necessary to use concept mapping up front to plan the intervention and its measurement and to carry that through to an outcome assessment which in typical cases would be years (and many program revisions) later. As the subsequent literature attests, we had no trouble finding numerous situations where concept mapping was useful in the conceptualization and planning of interventions. But it was virtually impossible to carry this through the life-course of an intervention to its subsequent outcome evaluation. That said, the idea of patterned hypotheses that could utilize such structures is at least an implicit part of some still-evolving approaches that have promise, such as the nonequivalent dependent variables design in its pattern-matching variation, and I still have some hope that it will make its way into more mainstream evaluation.

Second, the dependence of concept mapping on specialized software has always been and remains a major limiting factor to the more widespread use of the methodology. As I've discussed here, the development of the software for the method co-evolved with the method itself; the two were largely inseparable. In the earliest days of concept mapping we did use existing software packages like SPSS and SAS to do the analysis for concept mapping, and it certainly remains possible to do that to this day. But the analysis is only one step of the method. The more challenging part of the process is the collection of the data, and particularly of the sorting data. In the earliest days we collected this manually, by having people actually sort cards. This required either having them present in a room or sending them a packet of instructions, cards and a sort recording sheet. This limited the use of the method only to circumstances where such data could be collected. With the development of a Microsoft Windows program in the early 1990s, we finally had a software interface that enabled people to sort ideas on a computer screen in a way that was manageable. But this too required the participant to download and install specialized software if they were to do this on their own, a challenge that proved daunting. When the internet arrived, we were finally able to develop an interface that could be easily accessible from anywhere at any time through virtually any modern web browser. This greatly increased the reach of the data collection. But all of the electronic screen sorting approaches tend to fall short of the actual manual sorting of ideas. There simply is not enough screen real estate to achieve verisimilitude with the manual approach. Even so, the web-based versions have arguably extended the feasibility of the method. The problem is that the challenge of web-based electronic sorting has not to date had enough applications to warrant the availability of a free and generally accessible web-based sorting platform that is useable for concept mapping. The only solutions that exist are proprietary and, at least for some, costly enough to be discouraging. We are beginning to see more widespread accessibility to free software that can accomplish the analysis (such as the paper in this volume by Haim Bar that uses the R programming language), but until we have an accessible and user-friendly low or no-cost web-based approach to collecting sorting data, the method is liable to remain limited in its applicability to a relatively small group of specialist facilitators.

The final limitation that I'll mention is one that could probably also be construed as a virtue: the fact that the concept mapping methodology has remained relatively static since its inception in the early 1980s. While there has been evolution in terms of the application of the method, and one can argue that it has been considerably fine-tuned through decades of practice and refinement, it would still be readily recognizable and comprehensible to the typical user of 1989. One can argue that this is an advantage, that it makes possible a cumulativity of knowledge that enables us to develop a deeper understanding of the method. But at some level it is a bit disquieting that there have not been more breakthroughs and advances that have evolved the method into a more sophisticated incarnation. We need better ways to generate the content for concept mapping than the method of brainstorming that has been primarily used (including methods for automatically generating potential statement sets from existing web content). We need software engineering that links the basic idea of concept mapping to a more contemporary "big data" view of the web. We need the next generation of multivariate statisticians to extend MDS and to develop better ways to link it to cluster analysis methods (in fact, we need to push beyond the MDS/CA model altogether into the next evolution of multivariate analysis). We need to be able to create adaptive maps that can be continually revised and added to without requiring a return to the original sorting of all statements. We need to be developing ways to accomplish truly massive maps that go far beyond the traditional 80–100 statement constraint, and ways to integrate across maps (meta-mapping) to enable the construction of conceptual structures that are compatible in scope and ambition to the breadth of the internet. In some sense, the concept mapping method still feels very much like a product of the 1980s pre-internet era from which it emerged, a somewhat static multivariate analysis grafted onto a fairly complex and circumscribed method of data collection. This volume indexes historically that view of concept mapping in the middle of the second decade of the twenty first century and, in the summation article of this special issue, Dan McLindon hints at some of the possible evolutions into the future. But what we have learned from the past several decades of doing concept mapping is that there is something inherently complex, dynamic and thoroughly contemporary about its products and impacts, something that we intuitively sense can potentially be extremely powerful as it becomes integrated into the increasingly transformative electronic cyberinfrastructures the internet makes possible.

Acknowledgements

The author wishes to acknowledge the extraordinary contributions of Tiffany Williams a graduate student at Cornell University, who played a critical role in helping coordinate the production of this volume, and Melida D. Busch, MLIS, Manager, Edward L. Pratt Medical Research Library, Cincinnati Children's Hospital Medical Center who provided assistance in the bibliometric analysis.

References

Achtert, E., Schubert, H.-P., Schubert, E., & Zimek, A. (2013). Interactive data mining with 3D-parallel-coordinate-trees. *Paper presented at the proceedings of the 2013 ACM SIGMOD international conference on management of data*.

Anderberg, M. R. (1973). *Cluster analysis for applications*. New York, NY: Academic Press, Inc.

Axelrod, R. (1976). *Structure of decision: the cognitive maps of political elites*. Princeton, NJ: Princeton University Press.

Campbell, D. T., & Kenny, D. A. (1999). *A primer on regression artifacts*. New York: Guilford.

Chen, H., & Rossi, P. (1984). Evaluating with sense: The theory-driven approach. In R. F. Conner (Ed.), Beverly Hills, CA: Sage.

Chen, H., & Rossi, P. (1983). Evaluating with sense: The theory-driven approach. *Evaluation Review*, *7*, 283–302.

Chen, H., & Rossi, P. (1990). *Theory-driven evaluations*. Thousand Oaks, CA: Sage.

Colledge, L., & Verlinde, R. (2014). *SciVal metrics guidebook*. Elsevier. https://www.elsevier.com/research-intelligence/resource-library/scival-metrics-guidebook.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: design and analysis for field settings*. Boston: Houghton Mifflin Company.

Cornell Office for Research on Evaluation (CORE) (2009). *The evaluation facilitator's guide to the systems evaluation protocol.* https://core.human.cornell.edu/documents/SEP_I.pdf Retrieved from https://core.human.cornell.edu/documents/SEP_I.pdf.

d'Ocagne, M. (1885). *Coordonnées parallèles et axiales: Méthode de transformation géométrique et procédé nouveau de calcul graphique déduits de la considération des coordonnées parallèles*. Paris: Gauthier-Villars.

Everitt, B. (1980). *Cluster analysis*, 2nd ed. New York, NY: Halsted Press, A Division of John Wiley and Sons.

Hewes, F., & Gannett, H. (1883). *Scribner's statistical atlas of the United States showing by graphic methods their present condition and their political, social and industrial development*. New York: Charles Scribner's Sons.

Inselberg, A. (1997). Multidimensional detective. *Paper presented at the IEEE symposium on information visualization*. http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=636793.

Inselberg, A. (2009). *Parallel coordinates: visual multidimensional geometry and its applications*. London: Springer.

Kane, M., & Trochim, W. (2006). *Concept mapping for planning and evaluation*. Thousand Oaks, CA: Sage Publications.

Krippendorf, K. (1980). *Content analysis: an introduction to its methodology*. Newbury Park, CA: Sage Publications.

Krippendorf, K. (2004). *Content analysis: an introduction to its methodology*, 2nd ed. Newbury Park, CA: Sage Publications.

Linton, R. (1985). *Conceptualizing feminism: a structured method (Ph.D. dissertation)*. Cornell University.

Rosenberg, S., & Kim, M. P. (1975). The method of sorting as a data gathering procedure in multivariate research. *Multivariate Behavioral Research*, *10*, 489–502.

SPSS Inc. (2005). SPSS (http://www.spss.com/). 233 S. Wacker Drive, 11th floor, Chicago, Illinois 60606: SPSS Inc.

Stone, P. J., Dunphy, D. C., Smith, M. S., & Ogilvie, D. M. (1966). *The general inquirer: a computer approach to content analysis*. Cambridge MA: The Massachusetts Institute of Technology.

Torre, D. (1986). *Empowerment: structured conceptualization and instrument development (Ph.D. dissertation)*. Ithaca, NY: Cornell University.

Trochim, W. (1989b). *Concept mapping for evaluation and planning: a special issue, Vol. 12*, New York: Pergamon Press.

Trochim, W., & Linton, R. (1986). Conceptualization for planning and evaluation. *Evaluation and Program Planning*, *9*(4), 289–308.

Trochim, W., Urban, J. B., Hargraves, M., Hebbard, C., Buckley, J., Archibald, T., Johnson, M., & Burgermaster, M. (2012). *The guide to the systems evaluation protocol*. Ithaca, NY: Cornell Digital Print Services.

Trochim, W. (1985). Pattern matching, validity, and conceptualization in program evaluation. *Evaluation Review*, *9*(October 5), 575–604.

Trochim, W. (1986). Structured conceptualization: A resource guide and progress report. In W. Trochim (Ed.), *Panel presentation to the annual conference of the American evaluation association*.

Trochim, W. (1989a). Concept mapping – soft science or hard art. *Evaluation And Program Planning*, *12*(1), 87–110.

Trochim, W. (1989c). An introduction to concept mapping for planning and evaluation. *Evaluation and Program Planning*, *12*(1), 1–16.

Trochim, W. (1989d). Outcome pattern matching and program theory. *Evaluation and Program Planning*, *12*(1), 355–366.

Urban, J. B., & Trochim, W. (2009). The role of evaluation in research-practice integration: Working toward the "Golden spike". *American Journal of Evaluation*, *30*(4), 538–553.

Urban, J. B., Hargraves, M., & Trochim, W. (2014). Evolutionary Evaluation: Implications for evaluators, researchers, practitioners, funders and the evidence-based program mandate. *Evaluation and Program Planning*, *45*, 127–139. http://dx.doi.org/10.1016/j.evalprogplan.2014.03.011.

Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, *58*, 236–244.

Weller, S. C., & Romney, A. K. (1988). *Systematic data collection*. Newbury Park, CA: Sage Publications.