Evaluation and Program Planning xxx (2016) xxx–xxx

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 xxx

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

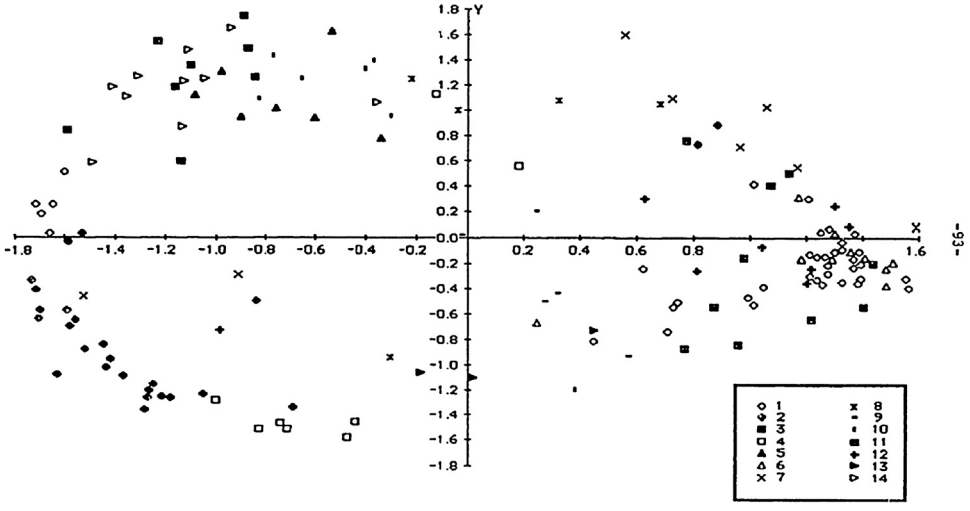

Fig. 1. One of the original concept maps, from Torre's 1986 dissertation (but produced several years earlier)(Torre, 1986 p. 93).

Please cite this article in press as: W.M. Trochim, Hindsight is 20/20: Reflections on the evolution of concept mapping, Evaluation and Program Planning (2016), http://dx.doi.org/10.1016/j.evalprogplan.2016.08.009