Chapter IV
Diagonal Values in ACA

ABSTRACT

Diagonal values in the cocitation frequency counts matrix are a fundamental issue in ACA study. Diagonal values are the co-citation frequency counts between the author himself/herself excluding self-citation. Retrieving exact values of diagonal values in the co-citation matrix requires a manual and time consuming procedure. For that reasons, ACA researchers suggested many different approaches to create, not retrieving the real values, the diagonal cells in the cocitation matrix. They include the mean cocitation count, missing values, zeroes, highest off-diagonal counts, adjusted off-diagonal values, and the number of times cocited with himself/herself. The majority of ACA researchers seem to prefer to use either the adjusted value approach by adding three highest off-diagonal values and divided by two or the missing value approach. This chapter empirically examines the impact of these different approaches on the ACA outcomes. Based on the results of this study, if the pure cocitation counts are not used, the next best alternatives are as follows. They are the missing value approaches, mean cocitation value approach, and the highest off-diagonal value approaches in the order of the highest total variance explained.

INTRODUCTION

Diagonal values in the cocitation frequency counts matrix are a fundamental issue in ACA study. Diagonal values are the co-citation frequency counts between the author himself/herself, not including self-citations. Diagonal values are the total number of citing papers that include at least two contributions by the author in the

references, excluding self-citations. The retrieval of the true values of diagonal values in the co-citation matrix requires a time consuming procedure. For that reason, ACA researchers suggested many different approaches to artificially create the value of the diagonal cells in the cocitation matrix, rather than pure cocitation between the author himself/herself. The alternative approaches suggested by ACA researchers essentially create the diagonal values by arithmetic computation using non-diagonal values of each author.

The purpose of this chapter is to empirically examine the impact of many alternative ways of handling diagonal values in lieu of co-citation between the author himself/herself. Raw cocitation matrix is the prime input to the ACA process. In addition to the issue of how to compute co-citation frequencies, another issue occurring during the compilation of the raw cocitation matrix is the value of diagonal values. The off-diagonal $value_{ij}$ is the cocitation frequency between $author_i$ and $author_j$, either retrieved from ISI databases or generated from custom databases. It has been suggested that the number of times that an author has been cocited with himself (excluding self-citations) would yield a mathematically complete matrix (Ahlgren, Jarneving, & Rousseau, 2003). The problem is that retrieving pure cocitation counts is an extremely time consuming process that can be difficult to be programmed, whether we use the Thompson ISI citation index and Dialog Classic to retrieve the cocitation counts or custom databases and cocitation matrix generation systems such as developed by Eom (Eom & Farris, 1996) or McIntyre (McIntire, 2007), ACA researchers, therefore, have suggested several different possible ways of treating the diagonal values, in lieu of retrieving the true value. No prior study in this area exists that systematically compares the results of ACA using cocitation counts matrix with different diagonal values.

NUMBER OF TIMES COCITED WITH HIMSELF/HERSELF

A possible solution for the diagonal value problem is suggested by Ahlgren, Jarneving, and Rousseau (2003, p.551.) and White (Howard D. White, 2003). The diagonal values should be the number of articles in the bibliographic database that cite at least two (different) works authored or coauthored by the person. Ahlgren, Jarneving, and Rousseau suggest excluding self-citations. This number is very difficult to get when using ISI databases. The custom database and cocitation matrix generation system we have developed allowed us to retrieve the numbers so we can yield a co-citation frequency matrix, which Ahlgren, Jarneving, and Rousseau describe as "a mathematically complete matrix." One solution for the diagonal value problem is suggested by Ahlgren, Jarneving, and Rousseau (2003, p.551.)