

# Concurrency Based Unidirectional Correlations

Lu Yang

## The Question

I wanted to look at pairs of bibliographical items that were borrowed together and measurement that can be established based on this concurrency. I believe this is interesting as it will review circulation history and correlate items even though they have irrelevant metadata. From an agent-based modelling perspective, I'm more interested in creating unidirectional correlations and measure concurrence frequency based on Dewey class and subjects.

## The Approach

In order to find items that were borrowed together, I assumed that items that had the same checkout and checkin timestamp were likely to be checked out by the same person. It should be noted that it is possible that

- 1) two individuals checked in/out books at the exact same time
- 2) librarians hold books at a rotation interval, so batches of books checked in/out by different person at different time may have same timestamp
- 3) processing time by the same librarian can lead to a few minutes difference on the timestamp
- 4) someone checked out multiple books but did not return them all at the same time

## The Query

In order to perform the search, I looked at the cartesian product of the inraw table with itself and selected bibliographical items that had the same checkout and checkin times as pairs. To measure concurrence frequency, I also looked at 3-digit Dewey class, which contains 1) main class; 2) division; 3) section; and subjects, which are retrieved as a concatenated string from the subject table.

```
SELECT
  t1.bibNumber AS A_bib,
  FLOOR(t1.deweyClass) AS A_dewey,
  t3.subject AS subject,
  t2.bibNumber AS B_bib,
  FLOOR(t2.deweyClass) AS B_dewey,
  t4.subject AS subject
```

```
FROM
  (SELECT
    bibNumber,
    GROUP_CONCAT(subject
      SEPARATOR ';') AS subject
  FROM
    spl_2016.subject
  WHERE
    spl_2016.subject.subject != ''
  GROUP BY bibNumber) AS t3,
```

```
(SELECT
```

WHERE  
t1.bibNumber = t3.bibNumber  
AND t2.bibNumber = t4.bibNumber

## Findings

I have queried the entire dataset and retrieved 3,207,172 rows in 307 seconds. Attached is a sample of the data. The data shows correlation between two identical bibliographical items A and B with unique bibNumbers (copies with different format are treated as one bibliographical item). From this spreadsheet, further concurrence frequency can be counted such as A\_dewey to B\_dewey, A\_dewey to B\_subject, A\_subject to B\_subject.

[illegible]

These three frequencies are important to me because I'm interested in visualizing these correlations in an agent-based swarm simulation. The aggregation of these frequencies will determine the attraction/repulsion force from A to B, not necessarily from B to A, when two random bibliographical items meet.

One thing to notice is that Unicode character in subjects is creating duplicated information(such as below), and I will try to clean up the subject lists.

Cookery French;Cookery French Proven al style;Cooking French;Cooking French Proven al style;Cooking French Proven al style