Библиографический анализ дискуссии

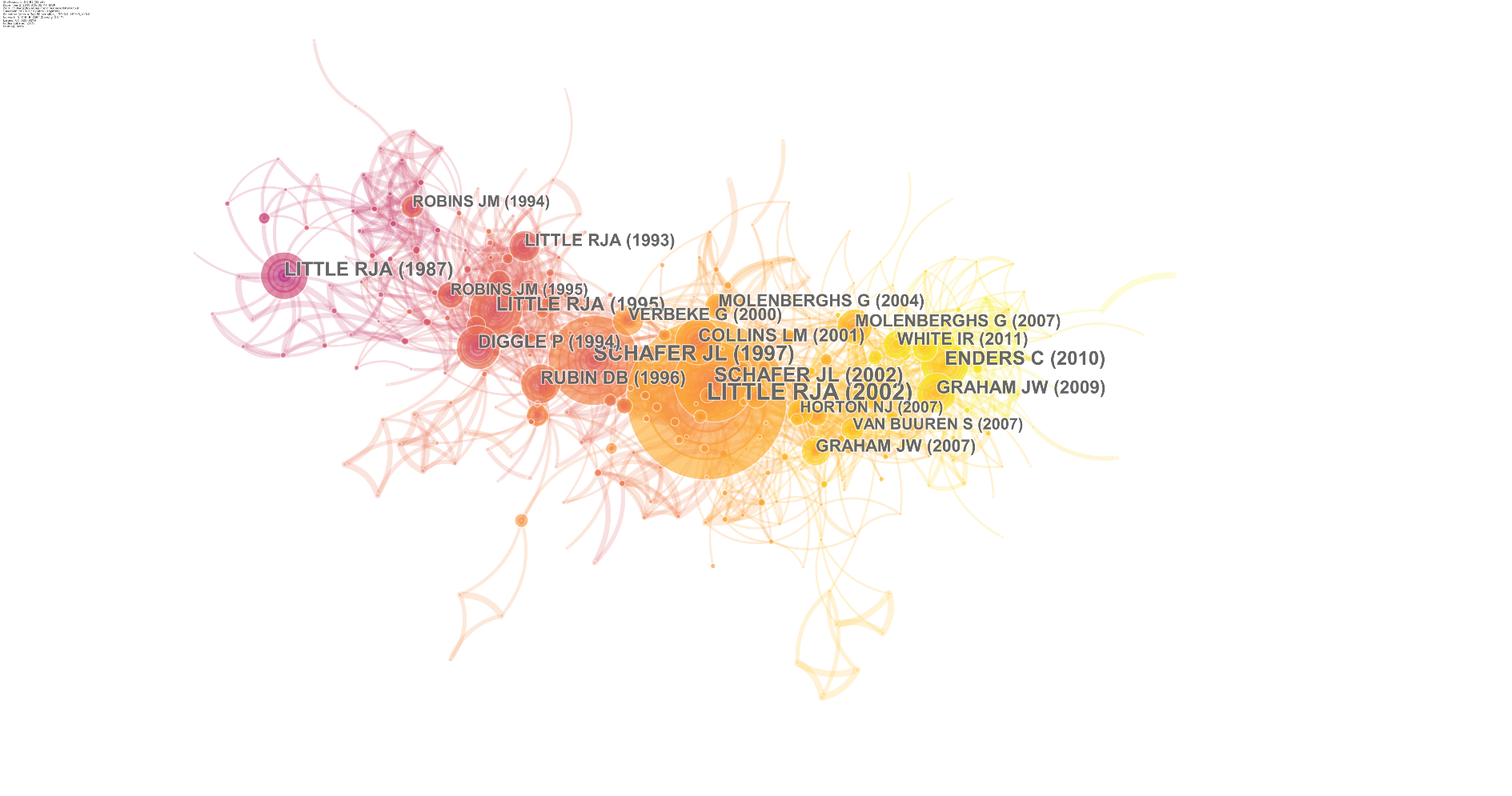
*Резников Дмитрий*

*09.10.19*

В докладе представлен актуальный способ работы с дискуссией, выстраиваемой вокруг некоторой статьи. Современные подходы составления литературного обзора и анализа литературы в целом включают в себя количественные методы, направленные на работу с цитированием (как значением), разного рода индексами, цитирующими и цитируемыми статьями и другими агрегируемыми значениями. Популярность набирает такой тип исследования, в котором целостно освещена работа с библиографическим методом – а точнее, конкретным его применением, построением сетей ко-цитирований – по анализу некоторой академической сферы (Aleixandre-Benavent и др. 2018; Chabowski, Samiee, и Hult 2013; Maia и др. 2019; do Prado и др. 2016). Такого рода исследования, выбирая произвольное академическое направление, например, кредитные союзы (Maia и др. 2019), в качестве входной базы данных для анализа используют, как правило, список публикаций и их атрибутов, полученных c помощью обоснованно составленного поискового запроса в базах данных цитирования Scopus и / или WOS. Основное предположение, раскрытое в докладе, - это вопрос о возможности проведения библиографического анализа при использовании такой базы данных, которая была получена не с помощью уточнения поискового запроса, а с помощью агрегирования всех статей, *цитирующих* произвольно взятую публикацию.

Этой публикацией, в связи с предстоящей над ней работой научно-учебной группы, стало исследование Рубин о свойствах пропусков в данных (Rubin 1976). Из WOS было скачано 4400 статей, *цитирующих* публикацию Рубин, после чего они были предобработаны на предмет наличия дубликатов. Итоговая сеть ко-цитирований (см. рис. 1) представляет собой специфическую визуализацию дискуссии вокруг одной статьи, в которой размер узла связан с его значимостью, а цвет узлов и мостов между ними связан с годом публикации. Предполагается, что при первичном анализе дискуссии возможно построение подобных сетей ко-цитирований с целью дальнейшего их применения: описание наиболее значимых узлов; кластеризация для выявления дискуссионных полей; поиск ключевых слов, отражающих основное направление дискуссии.

Также в ходе доклада было обозначено значимое различие между анализом базы *цитирующих* публикаций, который реализуется непосредственно в WOS и Scopus, и анализом сети ко-цитирований. При анализе базы исследователь непременно сталкивается с проблемой "ложной" значимости публикации, когда количество цитирований *цитирующей* статьи вовсе не говорит о её включённости в дискуссию. Так, например, статьи по исследованиям в клинической медицине могут включать в список литературу публикацию Рубин, лишь мельком ссылаясь на проведённую ей работу. Напротив, в сети ко-цитирований, чем более значимой (цитируемой) является статья, тем с большей вероятностью она включена в дискуссию вокруг публикации. Статьи, *цитирующие* Рубин, цитируют также и публикации, включённые сеть. Это значит, что чем больше значимость включённой публикации, тем больше статей одновременно цитируют эту публикацию и Рубин.



*Рисунок 1.* Сеть ко-цитирований в WOS

Кроме того, в докладе обсуждалась возможность поиска конкретных интересующих исследователя тем во всём массиве заданных данных. Предполагается, что если исследователю необходимо отобразить отсутствие обсуждения эффектов взаимодействия в дискуссии о статье Рубин, то это возможно сделать с помощью поиска по ключевым словам в атрибутах названия и абстракта *цитирующих* публикаций. Так, было найдено тридцать статей, использующих словосочетание «missing at random» или его сокращение «mar» (см. табл. 1), одна статья, использующая словосочетание «missing indicator method» или «mim», и одна статья, использующая словосочетание «interaction effects», что говорит о возможной их недоисследованности.

*Таблица 1*

Статьи, *цитирующие* Рубин



Список литературы

Aleixandre-Benavent, Rafael, José Luis Aleixandre-Tudó, Lourdes Castelló-Cogollos, и José Luis Aleixandre. 2018. «Trends in global research in deforestation. A bibliometric analysis». *Land use policy* 72:293–302.

Chabowski, Brian R., Saeed Samiee, и G. Tomas M. Hult. 2013. «A Bibliometric Analysis of the Global Branding Literature and a Research Agenda». *Journal of International Business Studies* 44(6):622–34.

Maia, Saulo Cardoso, Gideon Carvalho de Benedicto, José Willer do Prado, David Alastair Robb, Oscar Neto de Almeida Bispo, и Mozar José de Brito. 2019. «Mapping the Literature on Credit Unions: A Bibliometric Investigation Grounded in Scopus and Web of Science». *Scientometrics* 120(3):929–60.

do Prado, José Willer, Valderí de Castro Alcântara, Francisval de Melo Carvalho, Kelly Carvalho Vieira, Luiz Kennedy Cruz Machado, и Dany Flávio Tonelli. 2016. «Multivariate analysis of credit risk and bankruptcy research data: a bibliometric study involving different knowledge fields (1968–2014)». *Scientometrics* 106(3):1007–1029.

Rubin, Donald B. 1976. «Inference and missing data». *Biometrika* 63(3):581–592.