

New CollEc: An Interactive Exploration of the Economic Literature's Co-Authorship Network

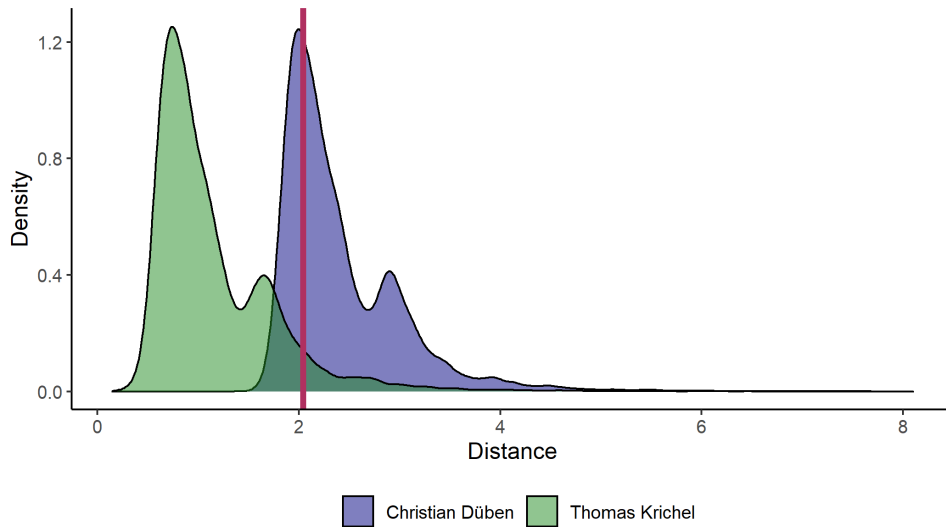
The economic literature is a field comprised of tens of thousands of authors. The American Economic Association alone has more than [20,000 members](#). In September, the RePEc Author Service passed [60,000 registered users with published research](#). Around 48,000 of them published at least one co-authored paper with another registered user.

[Co-authored research has been on the rise over the past decades](#) forming collaborations over vast geographic distances and fields of research. It is a network that interconnects the vast majority of published economists around the world. While many researchers are aware of collaborations between their close colleagues and prominent figures in their field, it is a challenge to even have a rough idea of what the overall network looks like.

When [Thomas Krichel](#) released the [CollEc](#) RePEc service in 2011, the co-authorship network's structure became accessible. With a few clicks users can evaluate which authors form the center of the discipline and who holds a more peripheral position. Each listed person is assigned a centrality value computed using methods from the field of graph theory.

Now in 2020, CollEc enters a new chapter of its existence. After years of maintaining the project and providing intriguing insights into the economic literature's co-authorship network, [Thomas Krichel](#) transferred the RePEc service to me, [Christian Dübén](#). I used the opportunity to come up with a completely new implementation, re-writing CollEc from scratch. The former network analysis written in Perl took the server hours at full capacity. Migrating it to C and C++ code wrapped in R functions boosted efficiency, cutting the required time and resources to a small fraction of what the previous implementation required, and facilitating extensions to the analysis. I added weighted edges, bilateral distances, and other results going beyond the centrality measures.

The interface through which users view the data changed from a static website to a web applications. Web applications are more complex and give me the necessary flexibility to fundamentally redefine how the data is presented. The [new CollEc](#) is highly interactive and puts results through combinations of plots and text into perspective. When a user inquires the distance between two authors, CollEc generates a figure comparing that bilateral distance to the distribution of distances to all other authors in the network. The following plot is the result of requesting the distance between Christian Dübén und Thomas Krichel with edges weighted by an inverse transition function. The concept of transition functions and the interpretation of the plot are outlined in the application.



With CollEc’s functionalities you can explore who someone’s co-authors are, how far two people are apart in the network, what the shortest path between them looks like, how centrally located a researcher is etc. All of the resulting plots are accompanied by a short text stating further information, e.g. on the network size. The web application evolves around the same approach as GraphEc, another recently developed but not yet publicly available RePEc service, does. It is an interactive tool focused on easily interpretable graphical output presenting results and facilitating comparisons.

Over the course of the past months, [Thomas Krichel](#) and [Christian Zimmermann](#) repeatedly reviewed the new CollEc and requested extensions and modifications. Thomas allowed me to host and test the application on his technical infrastructure from an early stage and did not withdraw his permission when I accidentally took his web server offline. Thanks to their great support, the application gradually improved and is now publicly available. To get started simply visit <http://app.collec.repec.org> and watch the tutorial or read the documentation. Either of the two options provides a brief, intuitive introduction into the basics of graph theory and the interpretation of CollEc’s results. Read the documentation on entry points, if you would like to generate a link to a certain output. Before you brag about your network centrality on Twitter, it should be noted, though, that author centrality is not a proxy for author quality. Successful authors can be central or remote. Consult the [IDEAS](#) website for a citation-based performance analysis of authors, journals, working paper series etc.

If you would like to contribute to CollEc, ask your colleagues to register with the [RePEc Author Service](#). CollEc’s network only entails authors listed in the RePEc Author Service’s data base. The vast majority of published economists is already registered. But some people are still missing. Fill the gaps in the network and ensure the reliability of CollEc’s results by promoting registration with the RePEc Author Service.

You can also decide to support RePEc more generally. IDEAS lists some [volunteering](#)

[options](#). RePEc is a non-commercial initiative run by volunteers providing openly accessible services. Small contributions like adding [RePEc Genealogy](#) entries already help in maintaining and improving this public good. I am a junior researcher who is going to be on the job market next year. Like the rest of my peers, I am under a lot of pressure to produce high quality research. Nonetheless, I do not regret having spent months on developing CollEc. Open science initiatives like RePEc are important contributors to an equitable research environment.