# DHM: a Digital History of Macroeconomics interactive platform

true true

**Abstract**

Abstract goes here

## Contents



The Mapping Macroeconomics project is an online interactive platform displaying bibliometric data on a large set of macroeconomic articles. It aims at offering a better understanding of the history of macroeconomics through the navigation between the different bibliometric networks.

The point of departure of the project is the observation of an exponential increase in the number of articles published in academic journals in economics since the 1970s. This phenomenon makes it harder for historians of economics to properly assess the trends in the transformation of economics, the main topics researched, the most influential authors and ideas, etc. We consider that developing collective quantitative tools could help historians to confront this challenge. The opportunities that a quantitative history brings are particularly useful to the recent history of macroeconomics. Practicing macroeconomists are eager to tell narratives of the evolution of their field that serve the purpose of intervening on current debates, by giving credit to particular authors and weight to specific ideas. Historians who go into this area find plenty of accounts by macroeconomists and have to handle the vast increase in the macroeconomic literature since the last quarter of the past century. The Mapping Macroeconomics platform aims at helping historians to empirically check macroeconomists' narratives on the discipline, to explore interesting patterns on the evolution of macroeconomics, and eventually to write new histories of macroeconomics.

# 1 Methodology

## 1.1 Construction of the Corpus

Our corpus is composed of macroeconomic articles published in economics. We identified all articles published in macroeconomics using JEL codes related to macroeconomics (Econlit database). JEL codes are used in economics to classify articles into specialties, like “Microeconomics”, “Macroeconomics & Monetary Economics”, “Industrial Organization”, etc. An article can have multiple JEL codes and so can be identified as part of multiple specialties. The JEL nomenclature was radically altered in 1991, and while these results in some discontinuity between the two nomenclatures, there are some correspondence (see Cherrier (2017) for a history of the JEL codes). The contemporary list of JEL codes can be found on the AEA website[1] and the old JEL codes with old/new correspondence table can be found in the Journal of Economic Literature, volume 29(1) (JEL, 1991).

For our corpus, we consider that an article is a macroeconomics article if it has one of the following codes:

- For old JEL codes (pre-1991): 023, 131, 132, 133, 134, 223, 311, 313, 321, 431, 813, 824.
- For new JEL codes (1991 onward): all E, F3 and F4.2.

Two additional comments on the JEL classification are necessary:

- First, we had to use some pre-1991 JEL codes that are not considered in the new classification as totally belonging to macroeconomics. Consequently, many articles in our pre-1991 corpus are public finance/public economics articles. Nonetheless, this group is clearly identifiable in our networks and thus do not disturb the interpretation of our results.
- Second, in the recent classification, the letter E designates macroeconomics JEL code, while F designates International Economics. In this last sub-discipline, we decided that it would be important to have articles dealing with international macroeconomics and thus we integrated articles with F3 and F4.2 JEL codes.

Using these JEL codes, we match the articles extracted from Econlit with Web of Science articles using the following set of matching variables:

- Journal, Volume, First Page
- Year, Journal, First Page, Last Page
- First Author, Year, Volume, First Page
- First Author, Title, Year
- Title, Year, First Page

We expect that these matching procedure results in some false positive. However, two elements prevent false positive from having any importance on the platform. First, we applied a general threshold on edges by keeping links between articles that had at least two references in common, and our projected networks are only made of the main component of our corpus (i.e,the biggest connected network). In other words, false positive have a very high chance of being completely disconnected from our main component and therefore filtered out from our analysis. Second, even if false positive “articles” are present in some networks, these articles, when irrelevant, would be relegated at the margins of our networks and thus do not have any significant impact on our results.

[1] https://www.aeaweb.org/econlit/jelCodes.php

## 1.2 Network construction

Our networks are based on bibliographic coupling. In a bibliographic coupling network, a link is created between two articles when they have one or more references in common. The more references two articles have in common, the stronger the link. The idea is that articles sharing many references to gather are likely to share cognitive content (ideas, theories, methods, objects of study, etc.).

To normalize and weight the link between two articles, we used the refined bibliographic coupling strength of Shen et al. (2019). This method normalized and weight the strength between articles by considering two important elements:

1. The size of the bibliography of the two linked articles. It means that common references between two articles with long bibliography are weighted as less significant since the likeliness of potential common references is higher. Conversely, common references between two articles with a short bibliography is weighted as more significant.
2. The number of occurrences of each reference in the overall corpus. When a reference is shared between two articles, it is weighted as less significant if it is a very common reference across the entire corpus and very significant if it is scarcely cited. The assumption is that a very rare common reference points to a higher content similarity between two articles than a highly cited reference.

For all macroeconomics articles published in the EER and in the Top 5, we build successive networks on 5-year overlapping windows (1969-1973; 1970-1974; . . . ; 2010-2014; 2011-2015). This results in 43 networks.

For each network, we apply a variety of treatments:

- We apply a general threshold on edges by removing links between articles that had only two references in common before weighting. We consider that it is more likely that the link between two articles is significant if they share at least two references. This is a great way to filter very common references such as econometrics hand book that might link two articles that have little in common. This is also a way to filter potential false positive from our matching method and exclude article that have little in common with the overall network beside one reference.[2]
- We only kept the main component of the network (thus ignoring singleton and secondary components). Nonetheless, we made sure that no other components represented more than 2% of the whole network.
- We place nodes in a 2-dimensional space using the Force Atlas 2 algorithm (Jacomy et al. 2014). Force Atlas relied on an attractive force - bringing closer articles which are linked - and a repulsive force - moving away the articles with no link, while minimizing the crossing between edges.
- The size of the nodes depends on the number of citations - coming from other macroeconomics papers - the article received during the time window.
- We identify relevant groups of articles using a cluster detection algorithm and colored nodes according to the cluster they belong to (see below for details).

## 1.3 Cluster detection

We use the Leiden detection algorithm (Traag, Waltman, and van Eck 2019) that optimize the modularity on each network to identify groups of articles that are similar to each other and dissimilar to the rest of the network. We use a resolution of 1 with 1000 iterations. This results in X clusters across all networks.

[2] There exists other recently developed methods to filter non-significant edges such as the Stochastic Degree Sequence Model from backbone approaches (Domagalski, Neal, and Sagan 2021). In addition to being costly computationally and unpractical in our case, backbone approaches are better suited for tightly knit networks for which there is not obvious weighting methods. This is not the case for your corpus. With a low threshold and the weighting method used, clear and stable clusters are easily identified.

Because networks have a lot of overlaps (three years are common to successive networks), many clusters between two successive periods are composed of the same articles. To identify these clusters that are very similar between two successive time-windows, we consider that two clusters can be merged if (i) at least 51% of the articles in a cluster of the first time-window were in the same cluster in the second time window and (ii) the cluster was also composed by at least 51% of articles of the first time-window. Simply put, between two successive time-windows, if two clusters share a high number of articles, and are both mostly composed by these shared articles, they are considered the same cluster.

Finally, we collect a certain number of information for each cluster (most cited nodes, most cited references, recurrent authors and journals, most "identifying" words in title) to name the clusters. We use two-level names:

- Meta-names are names shared across different clusters. They capture a broader characterization of the clusters: Business Cycles; Consumption, Investment & Production; Econometrics; Growth; International Macroeconomics; Keynesian Economics, Macro Policy; Monetary & Financial Economics; Public Finance.
- Sub-names are unique to clusters. They capture what makes the cluster specific to other clusters in the network, and other clusters with the same meta-name (e.g., International Macroeconomics: Exchange Rate Dynamics).

To make sure that the name chosen were not too biased by our individual subjective judgement, we submitted the list of clusters to **EXPLAIN HERE IN MORE DETAILED WHO DID WHAT**.

# 2 Features

## 2.1 Graph exploration and manual search

The platform and its results rely on data taken from Web of Science. Besides the underlying data used, we can distinguish two analytical layers to the platform. The first layer is made of the network nodes and edges and capture the way we project bibliographic information on the a two-dimensional space.

The main methodological elements that characterize this layers is the method of projection of the bipartite graph used (bibliographic coupling), the length of the time-window, the edge threshold, the filtering of nodes unconnected to the main component, the way we weighted the edges, and finally the algorithm used to placed nodes in a two-dimensional space.

Users of the platform can explore this first analytical layer freely using the main view of the platform. Using the timeline at the bottom of the platform, a particular time window can be chosen to display all macroeconomics articles published during that time frame.

A first obvious way to explore this graphical representation of the literature is to use the mouse to move (sustained click), zoom (mouse scroll), and interact with the network (left click). Different level of zoom capture different levels of information.

User can get an overview of the network while being completely zoomed out. A first obvious heuristics that catches the eye is to focus on big nodes. Big nodes in your bibliographic coupling networks are nodes that are very cited within the selected time window. In other words, it does not give a lot of information about long-term impact, but it captures what are the "hot" articles published during the selected time frame. For example, in the 1990-1994 network, a very big node at the bottom of the network catches the eye. Johansen and Juselius (1990) was published in 1990, and in the first year of its publications, it is already the most cited article in the network. It is not surprising to find that according to RePEc IDEAS ranking, this article rank 34th among the most cited economics articles.[3] Our network confirms the importance of this particular article in macroeconomics, but also as an article that within its first year was impactful.

[3]Source: https://ideas.repec.org/top/top.item.nbcites.html, consulted the 01/07/2022

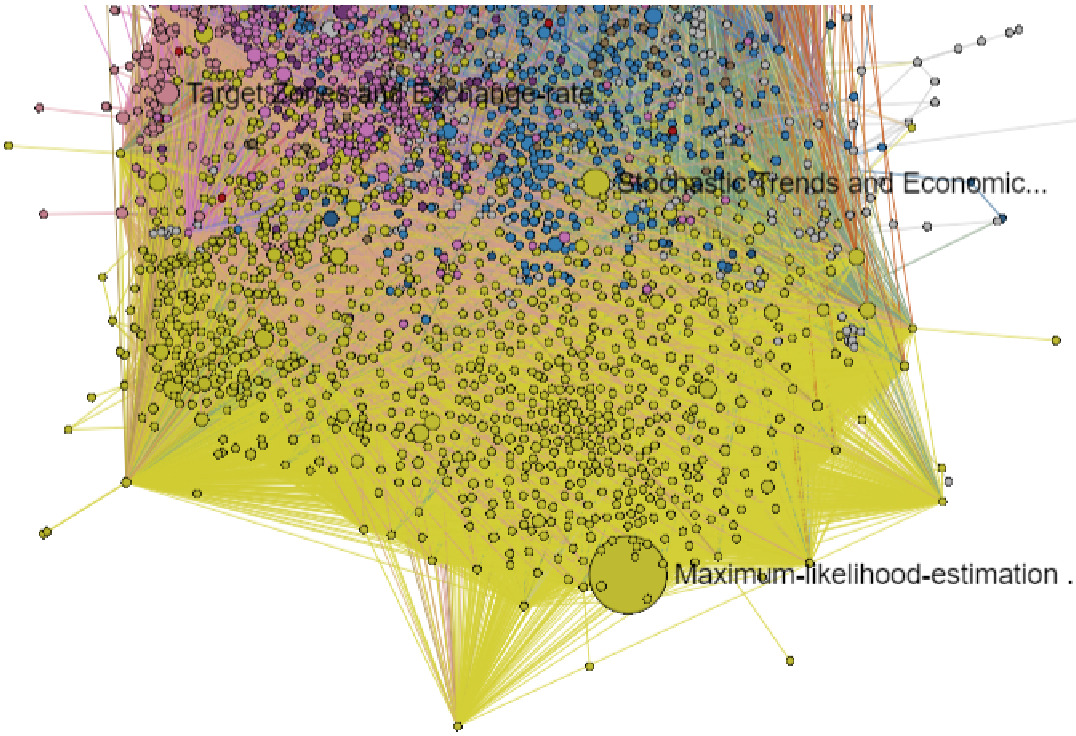
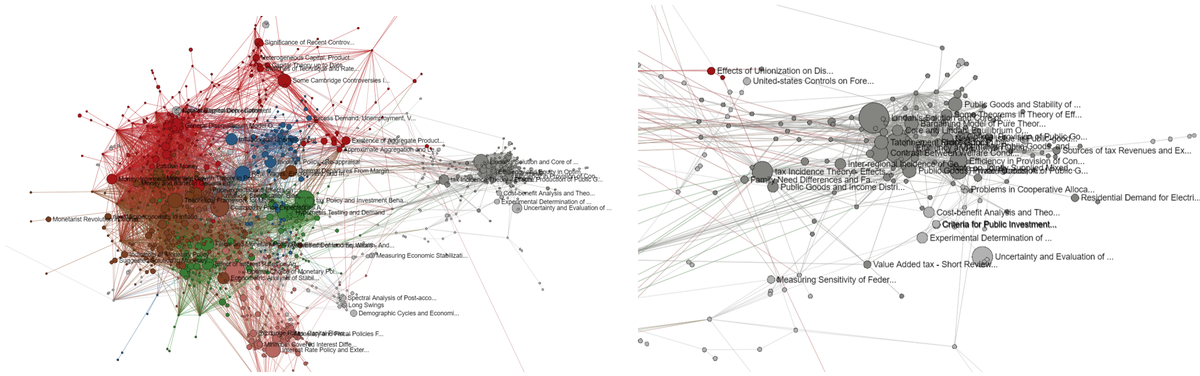



Figure 1: caption

A second important heuristics is to find close-knit groups of articles and blank spaces in the networks. For example, in the 1969-1973 a large group of article in the right part of the network is very detached from the rest (Figure 2a). Generally, when a close-knit group of articles is detached from most of the networks it means that its articles are bibliographically different from the rest of the network. User can zoom on relevant part of the network and investigate more closely the composition of such groups (Figure 2b).

From the zoomed in view, users can get a quick glance of the titles of the nodes. A better way to investigate the nature of the group is to click on individual nodes to bring forth the Node information panel. This panel will give you information about the authors names, the year of publication of the article, journal in which it is published, and the number of citations in the selected time window. By clicking on a node, you can also get information about its relationship to other nodes. For example, by clicking on one of the big node of the previously highlighted groups of articles (Figure 3a), we can see that the clicked node (McGuire and Aaron 1969) has only bibliographic relationship with other nodes from the group confirming the idea that it is part of a specialized literature dedicated to public finance and public good. Besides this big nodes at the center of the group, users can also investigate articles that bridges these different literatures. For example, while heavily bibliographically related to the public good groups, Mieszkowski (1969) on the effect of taxes bridge the public finance group with others groups such as exchange rate (on tariff and tax rates), investment behavior (corporate taxes), and Keynesian economics (Figure 3b). Articles bridging different parts of the literature means that they cite similar references as these different groups. It is not surprising to find that Mieszkowski (1969) was published in the *Journal of Economic Literature* as literature review tend to bring together sub-fields of economics that rarely cite the same references otherwise.

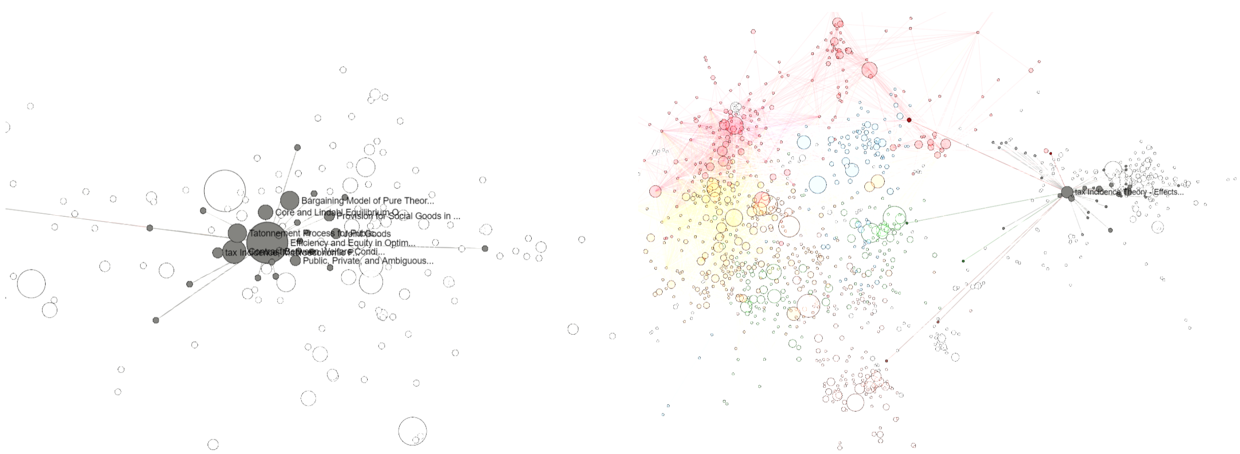
Finally, a third important heuristic tool is the search bar which allows users to search articles with particular authors, published in specific journals or whose titles contain defined keywords. After a search query, nodes matching the query will be highlighted with a red circle. For example, by search for occurrences of the keywords “Cambridge Journal” in the Journals of publications, we find that most articles published in the *Cambridge Journal of Economics* between 1989-1993 are situated in a very specific part of the network that




(a) Overview

(b) Zoomed

Figure 2: caption

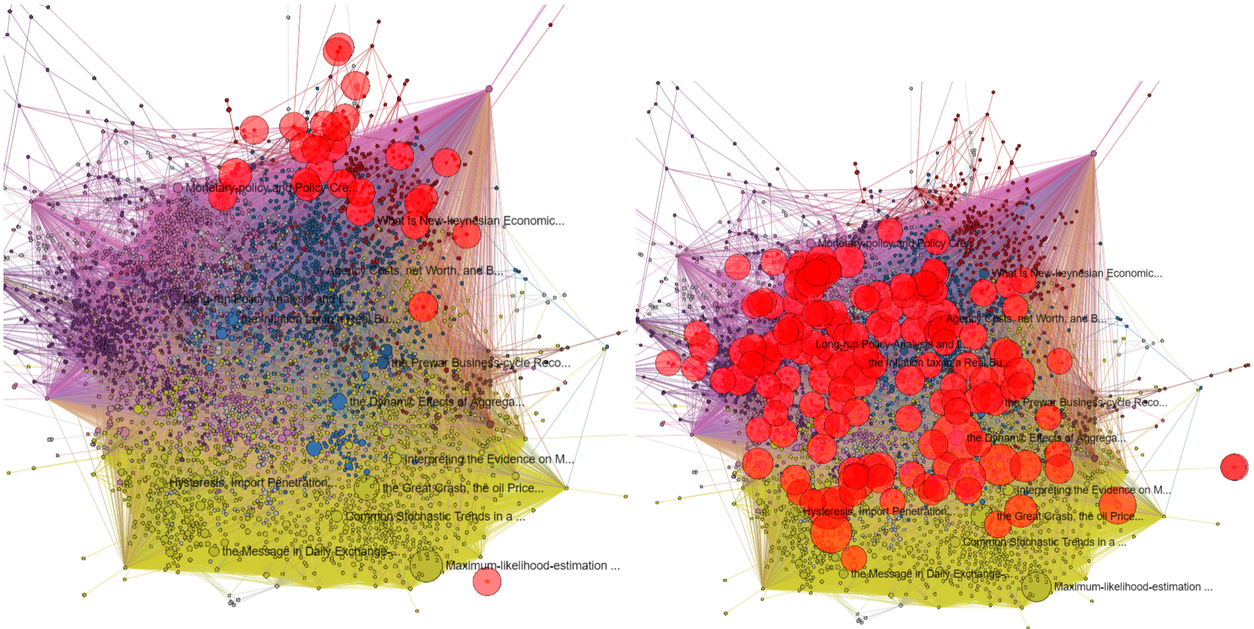



(a) Clicked node in group

(b) Bridging node

Figure 3: caption

we identified through clustering techniques as Keynesian economics (Figure 4a). This results contrasts with a query looking for articles published in the *American Economic Review* which highlight many different parts of the network (Figure 4b).




(a) Cambridge Journal query

(b) American Economic Review query

Figure 4: caption

Taken together, these three heuristics tools allow users to freely explore the structure of macroeconomics without concerns for methodological considerations regarding clustering techniques. The different time windows allows user to follow dynamically how groups of nodes, individual articles or particular keywords queries change over time. One can follow two close-knit group of articles to investigate the articles bridging them, or follow the number of citations of one particular articles by other macroeconomics articles within the first five years of its publications. In the case of keywords search, query remain active as time windows are selected, allowing user to easily track particular trends. For example, one can search for publications in the *Cambridge Journal of Economics* and change time windows. As soon as the network is loaded, relevant articles will be highlighted thus making changes in structure easily noticeable.

While these features highlight important ways to use the platform, its full power come when using these heuristics tools in combination with the cluster analysis made available on the platform.

## 2.2 Clusters analysis

When exploring particular networks, nodes are colored according to particular algorithmically-identified clusters. These clusters groups particular articles into discrete aggregates. As an overview of the network, users can use two important tools. First, the "Meta-name" and "Sub-name" buttons allows user to display directly on the network the name of the clusters (see section 1.3 for the difference between meta- and sub-names). Second, by clicking on the "Size of communities" users can obtain the list of all clusters and their relative importance in the network. Doing so will open a window with the share of nodes in the networks identified as part of the different clusters. This allows users to easily track how different clusters evolved over time in term of number of publications.

By clicking on any nodes in the networking belonging to a particular cluster, the "Node information" window opens and gives access to a new button called "Community information". By clicking on this new button, users can explore specific information that relate to the cluster whose name is the sub-name in the “Node information” window.

From this new windows, user can access a variety of table that summarize the nature of the cluster during this particular time window. It is possible to find out which are the biggest nodes of the community in the network and use the “Show” button to zoom and select a chosen node. It is also possible to have a summary of the venue of publications of articles in the cluster, or which authors have authored the most articles in the clusters.

In addition to these information which can be found by exploring the network, user can also find a very new information such as with the “Most cited references” table which rank the most cited references by articles in the clusters. As we said previously, the size of the nodes is a great way to investigate short-time influence. However, for long term impact, the “Most cited references” table is more useful as you can track how the importance of particular references increases or decreases across multiple decades for particular communities in macroeconomics.

The TFIDF table will give you information about the most distinct words (bigrams and unigrams) that appear in the title of the articles in the clusters. Finally, the “Evolution communities” and “Origin communities” tables allows you to investigate how the clustered chosen relates to clusters in the following and previous time windows respectively. This table is a way to quantitatively assess how clusters evolve in relationship to others clusters beyond heuristics impressions such as visually identifiable close-knit groups of articles.

While we gave name to particular cluster as a heuristics tools, it is important to explore communities of interest using these tables so that one can better understand the nature of the clusters. The way we named clusters capture important features of the clusters, but is might also play down potential heterogeneity in the clusters or smaller clusters within the clusters that capture relevant specialties of economics.

% The appendix command is issued once, prior to all appendices, if any.

# A Mathematical Appendix


Cherrier, Beatrice. 2017. “Classifying Economics: A History of the JEL Codes.” *Journal of Economic Literature* 55 (2): 545–79. https://econpapers.repec.org/article/aeajeclit/v_3a55_3ay_3a2017_3ai_3a2_3ap_3a545-79.htm.

Domagalski, Rachel, Zachary P. Neal, and Bruce Sagan. 2021. “Backbone: An R Package for Extracting the Backbone of Bipartite Projections.” Edited by Celine Rozenblat. *PLOS ONE* 16 (1): e0244363. https://doi.org/10.1371/journal.pone.0244363.

Jacomy, Mathieu, Tommaso Venturini, Sebastien Heymann, and Mathieu Bastian. 2014. “ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software.” Edited by Mark R. Muldoon. *PLoS ONE* 9 (6): e98679. https://doi.org/10.1371/journal.pone.0098679.

Johansen, Søren, and Katarina Juselius. 1990. “MAXIMUM LIKELIHOOD ESTIMATION AND INFERENCE ON COINTEGRATION - WITH APPLICATIONS TO THE DEMAND FOR MONEY: INFERENCE ON COINTEGRATION.” *Oxford Bulletin of Economics and Statistics* 52 (2): 169–210. https://doi.org/10.1111/j.1468-0084.1990.mp52002003.x.

McGuire, Martin C., and Henry Aaron. 1969. “Efficiency and Equity in the Optimal Supply of a Public Good.” *The Review of Economics and Statistics* 51 (1): 31. https://doi.org/10.2307/1926945.

Mieszkowski, Peter. 1969. “Tax Incidence Theory: The Effects of Taxes on the Distribution of Income.” *Journal of Economic Literature* 7 (4): 1103–24. https://www.jstor.org/stable/2720302.

Shen, Si, Danhao Zhu, Ronald Rousseau, Xinning Su, and Dongbo Wang. 2019. “A Refined Method for Computing Bibliographic Coupling Strengths.” *Journal of Informetrics* 13 (2): 605–15. https://doi.org/10.1016/j.joi.2019.01.012.

Traag, V. A., L. Waltman, and N. J. van Eck. 2019. “From Louvain to Leiden: Guaranteeing Well-Connected Communities.” *Scientific Reports* 9 (1): 5233. https://doi.org/10.1038/s41598-019-41695-z.