

Influence dynamics of adjacent academic disciplines. The case of economics and sociology.

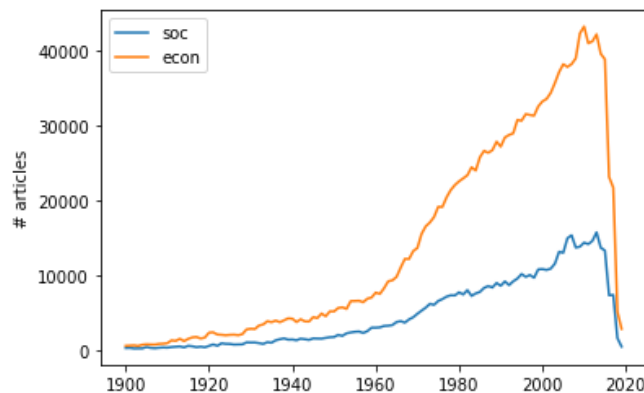
Benjamin Rosche – CS6742

Economics and sociology are often described as two fundamentally different social sciences. Economics is characterized as the study of rational action, focusing on the efficient allocation of scarce resources. Sociology, by contrast, is characterized as the study of nonrational behavior, focusing on topics, such as norms and biases.

In this project, I use data from the digital library JSTOR to study how research topics in economics and sociology co-evolve over time. Specifically, I analyze the titles of published research in the two disciplines to assess whether sociology and economics become more or less similar over time. In next steps, I will switch to a full-text analysis in which I will draw on the introduction and conclusion of each article.

From Constellate, I download metadata of most articles that JSTOR offers in sociology and economics between 1900 to 2020. The metadata consists of the title, abstract, author, journal, and publication year of an article. For most articles, the full text is also available. Figure 1 plots the number of articles in each discipline by year. The figure shows that economics is the bigger discipline and that the JSTOR lags behind with adding the most recently published articles.

Figure 1: Number of articles by discipline



Data cleaning

From each article title, I extract terms (unigrams) that are longer than 3 characters, excluding stop-words. I end up with 45,289 distinct terms in economics and 29,830 distinct terms in sociology.

The intersection of economics and sociology

To find terms that are common to economics and sociology, I first take the top 1000 most frequent terms in each discipline and then get the intersection of those two sets. Here are examples from the resulting set:

['american', 'care', 'case', 'change', 'development', 'economic', 'education', 'effects', 'employment', 'family', 'health', 'impact', 'income', 'international', 'labor', 'life', 'market', 'model', 'national', 'policy', 'political', 'problems', 'public', 'role', 'rural', 'social', 'state', 'states', 'structure', 'theory', 'united', 'urban', 'welfare', 'women', 'work', 'world']

Research objective and methodological strategy

The objective of this project is to understand how terms in economics and sociology are co-evolving over time. While each discipline has domains of research with little overlap from the other one, they also share a several research areas. I pursue three strategies to examine who is influencing whom in those intersections over time.

1. Fighting words analysis

In an influential article, Monroe, Colaresi, and Quinn (2008) develop a method to identify terms that distinguish between two sets of documents. Like PMI and TF-IDF, their approach is based on term frequencies. I employ their method to analyze how whether terms at the intersection of economics and sociology change the relative frequency with which they appear in both disciplines over time. This relative frequency is expressed as a z-score. If this score is positive, the term appears more frequently in sociology (proportionally to the size of its corpus). If the score is negative, the term appears more frequently in economics.

2. Co-occurrence networks analysis

My second strategy to understand the co-evolution of economics and sociology is based on co-occurrence networks. For each discipline, I identify term pairs (tuples) within the top 1000 most frequent words that appear together more often than we would expect by chance. I identify the pairs

using pointwise mutual information (PMI), which measures the association between two terms. PMI is defined as $PMI = \log \frac{P(AB)}{P(A)P(B)}$, where $P(A)$ refers to the frequency of term A and $P(B)$ refers to the frequency of term B in the entire corpus. If PMI is larger (smaller) than 0, the terms A and B appear more (less) frequently together than we would expect by chance. For each term, I then create egonets of terms with which they frequently appear together (i.e., term + neighbors). I use a PMI of 0.5 as cutoff point. I use this procedure to create the co-occurrence networks by decade.

Using the created co-occurrence networks, I then analyze how the egonets of relevant terms at the intersection of sociology and economics have changed over time. In particular, I calculate the Jaccard similarity of egonets from time t to time $t+1$. More importantly, not only do I analyze the similarity of egonets over time separately by discipline, but I also analyze the similarity of a term's egonet in economics at time t to its egonet in sociology at time $t+1$ (and vice versa). In doing so, I am able to assess who is influencing whom.

Figure 2: Egonet of the term “family” in sociology (left) and economics (right) in 2000-2010



3. Word mover distance analysis

My third strategy is to calculate the distance of titles in economics and sociology using Word2Vec (Mikolov et al. 2013), which represents words in vector form based on the context of the word using a shallow neural network. Rather than training the model myself, I make use of Google's pre-trained model, which extracts 300 features from 100 billion words from Google News. I then draw on the word mover distance technique (Kusner et al. 2015) to compute the distance between terms in economics and sociology titles. Since I am interested in observing change over time, I first extract all unique words across titles per decade per discipline, and then compute the distance of the list of terms in economics and sociology.

Results

1. Fighting words analysis

From the list of terms at the intersection of economics and sociology, I identify three term fields that would be interesting to analyze:

1. Subfields of sociology that have an equivalent in economics.

A fighting words analysis is indicative how strongly these subfields dominate the overall discourse in each discipline and, thus, how well they separate sociology from economics articles.

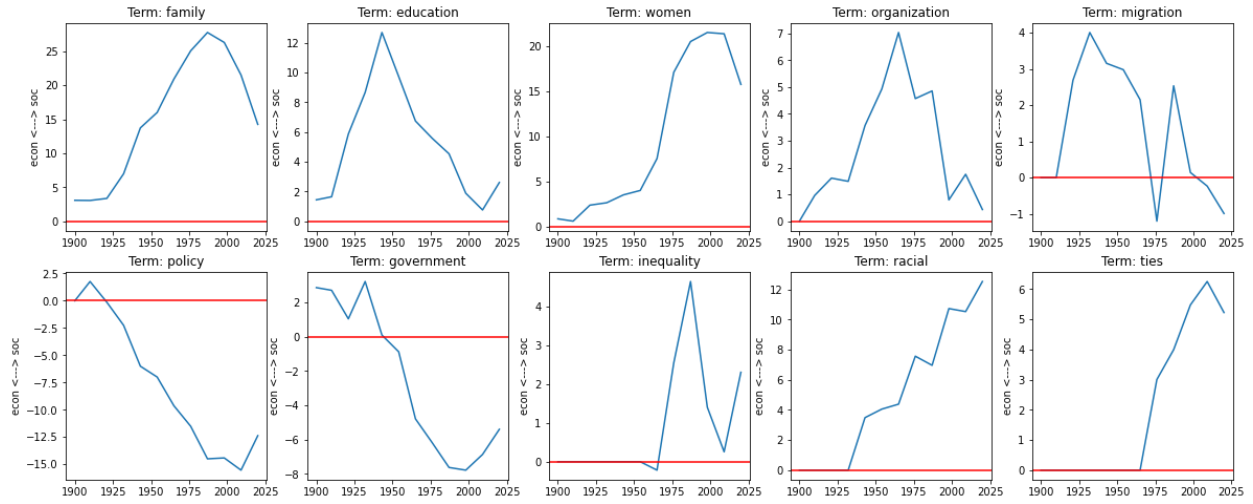
2. Words associated with certain research practices.

A fighting words analysis will give insight in how common such practices in each discipline are.

I analyze the following subfields: family sociology/economics (family, household), educational sociology/economics (education, school), political sociology/economics (policy, governance, institutional, capital), network sociology/economics (networks, graph), gender/race sociology/economics (women, men, gender, race), international sociology/economics (comparative, development).

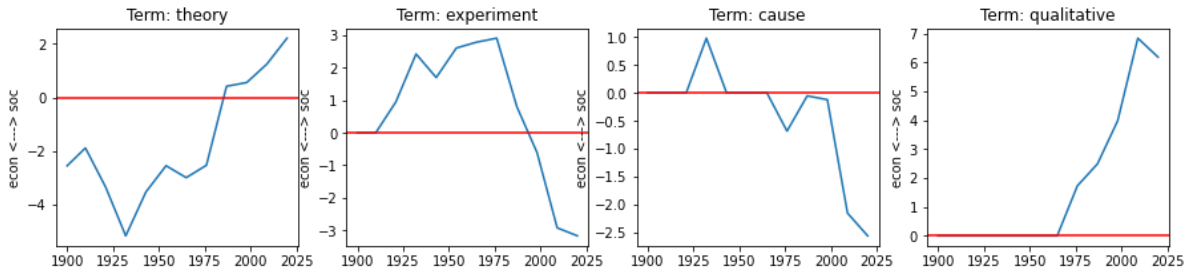
Figure 3 displays the results of a fighting word analysis of subfields. We can see several patterns. First, economics is an imperialist discipline. Except for race and networks, all here considered subfields lean less toward sociology over time. Take family sociology for instance. Until 1980, the term “family” has become an ever-better separator of sociology and economics articles. In 1981, Gary Becker published his influential book “A treatise on the family”, which founded family and household economics, and won him a Nobel prize. Since then, “family”, while still being associated with sociology, has become more and more frequent in economics. This reversal is also visible for education (reversal started around 1950), for gender (reversal started around 2000), for organization (reversal started around 1970), and for migration (reversal started around 1940). Second, economics took the spot to inform policy as can be seen from the trajectory of “policy” and “government”. Third, race and networks are still firmly in the hand of sociologists.

Figure 3: A fighting word analysis of subfields.



Regarding research practices, figure 4 shows that causal research is more commonly mentioned in economics and qualitative research more frequently mentioned in sociology. Experimental research, while more frequently mentioned in sociology until 1980, is now more common in economics. Finally, articles mentioning the word “theory” have been strongly associated with economics until about 1940, which around the time the foundations of neoclassical economics have been established. Since then, the term has become leaning more and more towards sociology.

Figure 4: A fighting words analysis of research practices.



2. Co-occurrence networks analysis

Figure 5 plots the results the co-occurrence networks similarity analysis for the term “family”. On the y-axis, the Jaccard similarity between the egonets of the term “family” in economics and sociology is displayed. Rather than comparing the egonets at same point in time, I compute the similarity of the egonet in sociology at time t (e.g., 1960) to the egonet in economics at time $t+1$ (i.e., 1965), which is a measure of whether sociology has influenced economics, and vice versa.

We can see that the similarity of egonets in the same discipline is higher than the similarity of egonets across disciplines. Moreover, we can also see that article titles in economics on family have become more similar to article titles in sociology on family up until 1985. Since then, the similarity is about the same.

Figure 5: Co-occurrence networks similarity analysis for the term “family”.

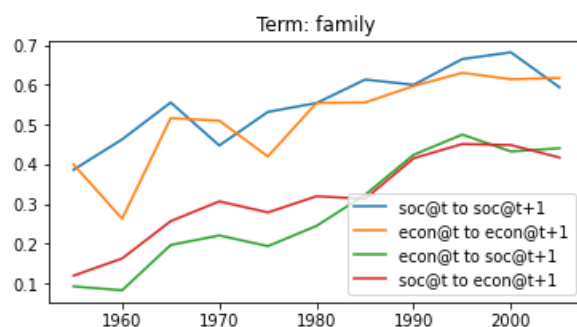
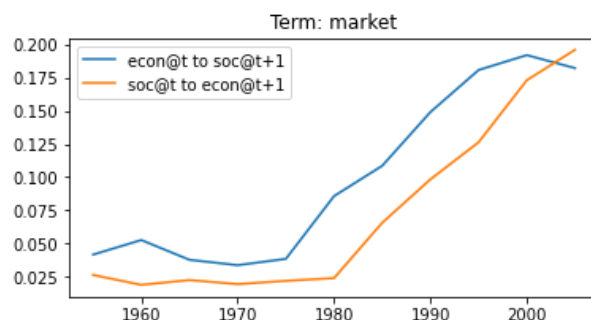


Figure 6 shows the same analysis for the term “market” but without the similarity of the disciplines to themselves. We clearly see that for a topic close to economics, the influence dynamic is the other way around: economics influences sociology.

Figure 6: Co-occurrence networks similarity analysis for the term “market”.



3. Word mover distance analysis

Lastly, I'm presenting the results for the word mover analysis. The advantage of the word mover technique is that an all-encompassing test can be performed (i.e., are terms overall in sociology and economics becoming more alike?). Moreover, the word2vec technique captures different aspects of neighboring words.

Rather than showing similarity, figure 7 gives the distance between sociology and economics. Accordingly, the y-axis can be interpreted as a form of predictive error – the lower, the more predictive. Two things can be seen from the figure. First, the language in sociology and economics articles has become more distinct between 1900 and 1950 and more similar between 1950 and nowadays. Second, there is no clear trend in who is influencing whom.

Figure 7: Overall distance between sociology and economics

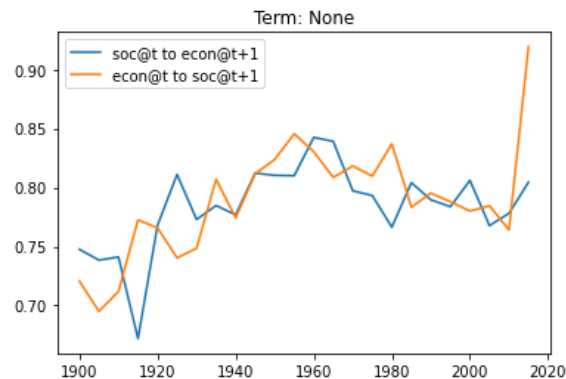


Figure 8 shows a term central to economics. We can see that the word embedding of “firm” in economics is more predictive of the word embedding in sociology than the other way around. The opposite is visible in figure 9 for the word “network”. Here the sociology lingo is more predictive of the econ lingo - up until 1990.

Figure 8: Word mover distance of the word “firm”

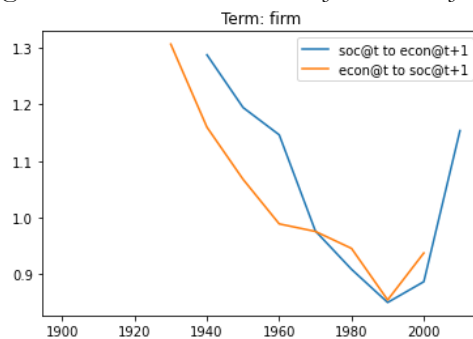
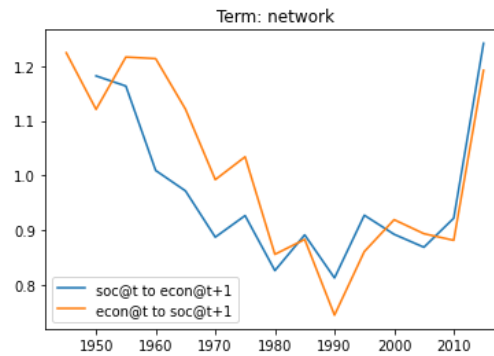


Figure 9: Word mover distance of the word “network.”



Conclusion and next steps

The word mover distance analysis has shown that sociology and economics overall, after a phase of self-discovery that lasted until 1960, have become more similar over time. The analysis also showed that there is not one discipline driving this convergence.

The co-occurrence analysis and the word mover analysis show that, in domains central to sociology, such as family, race, gender, and education, sociology seems to have indeed influenced economics. The fighting word analysis is picking up this convergence, showing that research articles that contain those terms are less clearly associated with sociology over time. The opposite is true for terms central to economics, such as “market” or “firm”. It is also noticeable from the co-occurrence and the word mover analysis that such convergence processes seem to have leveled off in the last decade.

This project was helpful not only for generating substantive insight but also for exploring methods to analyze meta data on research articles to examine co-evolution dynamics between disciplines. The fighting word analysis, for instance, is able to show convergence of language in articles but has not way of discerning who is influencing whom. Using co-occurrence networks or word embeddings to analyze the distance of language in articles, by contrast, is better able to tackle this question. The downside of those approaches is that they are computationally more intensive. Because of that, I had to sample articles and words, which creates instability in the results.

In next steps, I would like to train (finetune) the word2vec model on the data itself, switch to full-text analysis, and choose specific terms, such as “family”, for which I have enough substantive oversight to map the results on to events in the literature. Finally, I would like to take a more data driven approach to finding interesting results. For instance, I would like to bring back the k-shape algorithm to cluster time trends and thereby identify terms for which we see surprising results.

References

- Kusner, M., Sun, Y., Kolkin, N., & Weinberger, K. (2015, June). From word embeddings to document distances. In International conference on machine learning (pp. 957-966). PMLR.
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.
- Monroe, B. L., Colaresi, M. P., & Quinn, K. M. (2008). Fightin' words: Lexical feature selection and evaluation for identifying the content of political conflict. *Political Analysis*, 16(4), 372-403.