Engaging with homologies in discourses: towards a sociological appropriation of topic models

Pierre Benz   Radim Hladík   Yann Renisio

Topic modeling is a suite of information retrieval algorithms that has grown popular in sociological applications. Various implementations, including Bayesian inference, semantic networks and neural network-based embeddings have flourished in recent literature. As a dimensionality reduction technique, topic modeling assumes that the content of documents can be described by a finite set of topics, or distributions over vocabulary, which are typically interpretable by domain experts. Sociological applications of topic modeling tend to adopt its operational definition from computer science and information retrieval and, therefore, predominantly revolve around interpreting or discovering particular topics in large corpora. There is consequently a notable scarcity of studies that would develop this method from a distinctively sociological standpoint. Here, we propose to shift the perspective and argue for a sociological appropriation of the method that would focus on the structure of discourse that surfaces in the latent topic space. By highlighting the often overlooked feature of topic models, namely the representation of documents as distribution over topics, we emphasize the intrinsically relational character of topic models. In doing so, we aim to demonstrate how topic models empower sociologists to study the creation of symbolic values and their structuration according to the specific logic of different linguistic markets. The framework of the collective book lends itself perfectly to such task, namely to demonstrate the sociological potential of topic models to uncover homological structures. In this context, we propose to illustrate how the relationality of topic models renders them useful tools for examining homologies between discursive and social dimensions of fields, or between different discursive spaces, with empirical evidence from the scientific field.