S8 | Steps for the semantic network analysis of abstracts

The term maps for neuroeconomics (Fig. 4, main text) and for economics (Box S9) were created following the same process. First, some specific abbreviations were standardized (e.g., “pfc” was turned to “prefrontal cortex”). Then expressions composed of several terms (called “n-grams”) were identified across all abstracts (an expression could be as long as four terms). This step ensured that expressions such as “functional magnetic resonance imaging” could be included in the final map, along with potentially distinct single-term expressions such “functional” or “magnetic”. Next, the plural and singular forms of terms were grouped together following simple heuristics. Finally, the most frequent words of the English language and expressions typical of academic writing were removed, such as “the”, “a”, or “we conclude that”. Two words (“game” and “risk”) were retained because of their special relevance for the corpora studied, even though their high frequency in the English language would have otherwise eliminated them from the analysis.

For the network analysis, we selected the most frequent terms identified in the previous steps and determined which co-occur most frequently within abstracts. We visualized the resulting network with VosViewer1. This software includes a clustering algorithm identifying groups of words which co-occur most frequently2. These groups of words – coded by different colors on the map – can be interpreted as key topics in the corpus under study3. All code for the semantic and network analysis is publicly available: https://github.com/seinecle/Cowo.

References