

# Quantifying creative symbiosis: a lexical analysis of the evolution of technology as reflected in patent records

Noah Pepper, Charles Francis, Mark Bedau, Devin Chalmers and Bryan Head

Reed College  
peppern@reed.edu

The similarities and differences between adaptive dynamics in biological and cultural evolution is an important and controversial open question about evolutionary processes in the real world. One way to address this issue is by studying empirical data from biological and cultural evolution. Technology is itself an important part of culture, and one that is rather amenable to empirical investigation, and an excellent window on the evolution of technology is to study patterns in the citations among patent records (Jaffee & Trajtenberg, 2005, *Patents, Citations, and Innovations*, MIT Press). Each patent must describe a novel concept and must cite previous related works and each patent is filed at a distinct point in time. This provides us with a time series corpus of formally written text that records all patented technological innovations, an ideal platform for studying the creativity of one form of cultural evolution.

Bedau and Skusa (2002, *ALife VIII*, p. 431, MIT Press) analyzed the dynamics of adaptation in the evolution of technology, by looking at patent citations and measuring evolutionary activity statistics over time (Bedau & Packard, 1991, *ALife II*, p. 431, Addison-Wesley; Bedau et al., 1998, *ALife VI*, p. 233, MIT Press). We here present a novel, complementary method of analyzing technological evolution using the textual content of patent records. Our analysis takes linguistic tokens as the unit of cultural adaptation and measures their occurrence and relations using several linguistic tools. This can reveal latent connections between conceptual or technological innovations.

We have analyzed a corpus containing thirty years of patents (over 4 million) with the WORDSPACE model for quantification of word relatedness within a corpus (Widdows, 2004, *Geometry and Meaning*, CSLI). This produces a fuzzy set of correlated terms based on concurrence within a text. The change in relatedness of n-grams over time provides a movie of a part of cultural evolution. We also analyze regressions on time series frequency counts of n-grams and groups of n-grams. (N-grams are linguistic tokens n words in length which represent technologies or concepts in the corpus.) Patterns in n-gram frequencies provide another window into the evolution of culture.

These two tools reveal and map the cross-temporal relationships between technologies. One can see classes of technologies share in significant evolutionary success, and then eventually decline (for example, the recent technology bubble). The tools can reveal when different technologies have symbiotic relationships (Margulis & Fester, 1991, *Symbiosis as a Source of Evolutionary Innovation*). These results help illuminate the nature of cultural change and open-evolution, and in particular, whether there is a fundamental difference in the evolution of cultural and biological systems.