Topological analysis of healthcare data

Medical informatics draws increasingly from administrative healthcare datasets containing high-dimensional and heterogeneous data on large populations. This rich substrate fuels the ongoing development of predictive modeling for personalized medicine. Equally important are efforts to extract clinically interpretable features, such as distinct patient phenotypes, and to understand the web of relationships among them. Concepts and tools from network analysis are naturally suited to these purposes, and our entry into this research program begins with a systematic assessment of this work.

We performed a systematic review of applications of network analysis to administrative healthcare data (Brunson & Laubenbacher, 2018), in which social network analyses predominated. Clinical co-occurrence networks were the focus of a distinct subgenre that included analysis of *comorbidity networks*, constructed from patient-level diagnosis profiles much as species interaction networks from presence–absence data in ecology. The wide range of techniques and lack of cross-talk has produced a literature with no standard protocols and low comparability. Our most recent work has been to test the robustness of the network-analytic results this field has generated, and to suggest some steps to improve reproduction, validation, and interpretation (Brunson et al, *in preparation*).

Going forward, we are transitioning into the related framework of topological analysis. The powerful techniques of persistent homology and statistical Reeb graphs allow investigators to explore high-count and high-dimensional datasets with more flexibility and scalability than is possible using most dimension reduction techniques. Still, modeling data clouds as topological spaces represents only the first stage of an emerging paradigm. We are drawing upon concepts in algebraic topology to couple recent developments in personalized risk prognosis with multiple outcomes assessment. We expect to present some preliminary results in Summer 2019.

Jason Cory Brunson, Reinhard C Laubenbacher; Applications of network analysis to routinely collected health care data: a systematic review, Journal of the American Medical Informatics Association, Volume 25, Issue 2, 1 February 2018, Pages 210–221, https://doi.org/10.1093/jamia/ocx052

Jason Cory Brunson, Thomas P Agresta, and Reinhard C Laubenbacher; Sensitivities of comorbidity networks, *in preparation*.