539g Structure and Evolution of Organic Chemistry

Kyle J.M. Bishop, Marcin Fialkowski, Christopher J. Campbell, and Bartosz A. Grzybowski

For almost two centuries, chemists all over the world have applied their expertise and creativity to the synthesis of new molecules. Since each individual chemist – or a collaborating group of chemists – tries to select unique synthetic targets and come up with a maximally original method of making them, it might appear that the activities of such independent "agents" should be largely uncorrelated and that no generalizations about the evolution of chemistry en large could be made. As we show here, however, there exist universal statistical laws that describe how molecules are made and interconverted.

We analyze organic chemistry (as recorded in the Beilstein Database) at the level of an abstract graph representation whereby molecules correspond to nodes characterized by molecular masses, and reactions to directed edges connecting these nodes. As a directed graph, organic reactions form a scale free network – interestingly, of topology very similar to that of the Internet. Tracking the evolution of this network allows one to predict how many molecules will be made in the future, how many reactions these molecules will participate in (either as reactants or products), which molecular masses have been/will be most commonly used, and more.

Molecular masses of all organic molecules ever made are governed by one, Kesten-type stochastic equation. This equation allows one to predict the most likely masses of products in any organic reaction to be performed; it allows construction of optimal fragment libraries for combinatorial chemistry such that they would produce most "druggable" molecules; it reveals interesting facts about existing drugs (e.g., that mass distribution of drugs is no different from that of "ordinary" molecules)

Finally, there are some very important industrial predictions we can make based on the laws we discovered. We can estimate a price of a chemical based on its network connectivity. We can predict whether an existing molecule is an important one (pharmaceutically or industrially) based on the correlation between its in- and out- network connectivities. We show that it is less probable (by ~48%) to make molecules with odd molecular masses than with even ones.

On the fundamental level, our results provide yet another demonstration of how seemingly independent/uncorrelated activities of individual agents fall into a larger scheme of things and obey – or, as proponents of emergence would argue, "give rise to" – universal laws.