The Cosmos as a Colloid

A new methodology for analyzing the 3D distribution of galaxies borrows techniques from the study of colloids and other disordered materials.

By **Charles Day**

Scheduled for launch this summer, the European Space Agency's Euclid space observatory seeks to elucidate the expansion of the Universe and the formation of galaxies, clusters of galaxies, and larger structures. To meet that goal, Euclid's instruments will determine the 3D positions of millions of cosmic bodies. Fully exploiting that bounty of information will take more than prodigious computing resources. New methods will be needed that distill astrophysical and cosmological insights from the observations. With that aim in mind, Salvatore Torquato of Princeton University and Oliver Philcox of Columbia University have developed a novel methodology for analyzing cosmic structures that draws from techniques used in disordered materials [1].
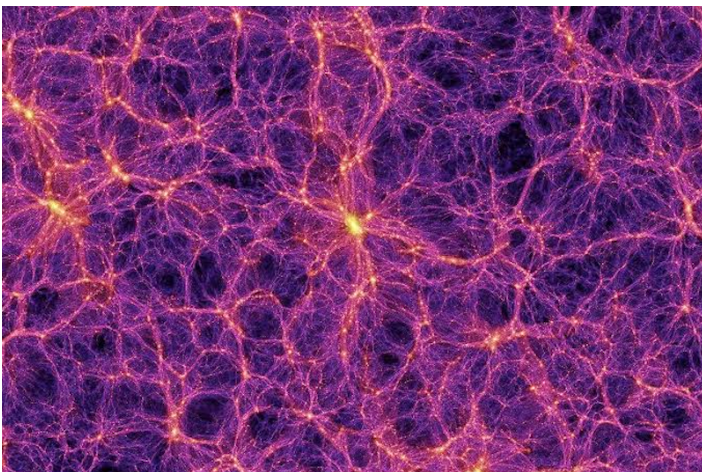
While Torquato works in statistical mechanics and soft condensed matter, Philcox is a cosmologist. The pair began their collaboration in the autumn of 2021, when Torquato heard

The distribution of dark matter in the Universe at the present epoch as derived from the Millennium Simulation.
Credit: P. Volker Springel/Max Planck Institute for Astrophysics

a talk on galaxy distributions that Philcox gave at the Institute for Advanced Study in Princeton, New Jersey. The talk roused Torquato's curiosity. He wondered if the theoretical and computational ideas he had developed for disordered heterogeneous materials might be profitably applied to characterize the structure and topology of galaxy distributions.

Observational cosmology makes use of many statistical tools such as the two-point correlation function. This parameter quantifies by how much the probability of finding a pair of galaxies separated by a given distance exceeds that expected for a uniform, random distribution of galaxies. Torquato and Philcox introduced additional statistical methods and concepts that are new to cosmology and complement the two-point correlation function. Among them are hyperuniformity and pair-connectedness functions.

To evaluate the usefulness of their approach, the duo avoided testing it with real data—such as those coming from the Sloan Digital Sky Survey or from other observational surveys—because such data come with systematic biases that complicate the analysis. Instead, they turned to Quijote—a suite of 45,000 N-body simulations that are freely available to cosmologists for training and testing machine-learning algorithms. The particular Quijote simulations used by Torquato and Philcox tracked the formation of 10^5 dark matter halos to the present day inside a cube whose sides measured 3 billion light years. Since dark matter is over 5 times more abundant than baryonic matter, and galaxies form in dark matter halos, the evolution of cosmic structure on the largest scales down to single galaxies can be approximated by modeling just dark matter.

At scales between 10 and 100 million light years, galaxies congregate in a network of filaments that snake between voids.

The assemblage looks like a foam of soap bubbles—that is, a disordered medium. On larger scales, the Universe becomes uniform and isotropic. Torquato and Philcox characterized that transition from a lumpy to a uniform medium, which astronomers call the "End of Greatness," with one of the statistics in their methodology, hyperuniformity. The duo showed that hyperuniformity provides a unified description of cosmological structures across length scales, capturing both the large-scale uniformity and the small-scale heterogeneity. The finding opens the prospect of using galaxy surveys to investigate the origin of that structural transition, which is related to primordial density fluctuations imprinted on the Universe during cosmic inflation.

On smaller scales, astronomers face the challenge of efficiently extracting from surveys the most important features of how galaxies group together. This extraction is often done with the two-point correlation function, which is computationally straightforward, but misses some details. The three-point correlation function, on the other hand, is more sensitive to small-scale details, but its computation is extremely demanding. Torquato and Philcox found that one of their statistics, the pair-connectedness function, captured as much detail as the three-point correlation function while easing the computational cost.

Matteo Biagetti of the International School for Advanced Studies (SISSA) in Italy is developing an approach called persistent homology. Similar in spirit to that of Torquato and Philcox, it identifies topological features in discrete sets of points. "These are exciting times for observational cosmology," he says. The findings by Torquato and Philcox give a strong indication that researchers should devote time and energy to further developing new statistical methods for efficiently analyzing the wealth of data that observations will deliver, says Biagetti.

Charles Day is a is a Senior Editor for *Physics Magazine*.

REFERENCES

1. O. H. E. Philcox and S. Torquato, "Disordered heterogeneous Universe: Galaxy distribution and clustering across length scales," **Phys. Rev. X 13, 011038 (2023).**