

MAR16-2015-001424

Abstract for an Invited Paper
for the MAR16 Meeting of
the American Physical Society

Clustering means geometry in networks

DMITRI KRIOUKOV, Northeastern University

Using maximum-likelihood estimation techniques, any real network data can be fit to essentially any network model, inferring the most likely values of the model parameters for the network. However there is one caveat. The results of such fitting are not spurious but meaningful and predictive, only if the network is a typical network in the unbiased ensemble of random graphs with the inferred values of model parameters. Therefore, given a particular combination of a real network and a model, the first question one has to answer is what structural properties of the network ensure that this network is a typical element in the ensemble of random graphs defined by the model. This question is usually highly intractable, explaining why it is almost never answered before the fitting/inference task is performed. Inspired by recent observations that random geometric graphs reproduce many structural and dynamical properties of a variety of real networks, we find the network structural properties that guarantee that networks that have these properties are in fact geometric, meaning that latent space network models are their true models. Specifically we prove that peculiar organization of clustering observed in real networks is one of the main such properties. In other words, maximum-entropy random graphs with specific clustering properties, which are quite different from the clustering properties of random graphs in the Strauss model, are actually soft random geometric graphs with a specific form of the connection probability function. Using this function we can then infer the coordinates of nodes in a latent space for any given network, and reliably check if the network is a typical network in the resulting ensemble of soft random geometric graphs. If it is, then the inferred coordinates are meaningful and real, and can be used for prediction tasks with proved guarantees that the results of such predictions are reliable and not just transient artifacts.