

# Catalogare miliardi di dati: metodo scoperto alla Sissa

Gli addetti ai lavori usano l'espressione big data per indicare grandissime quantità di informazioni, come quelle (foto, video, testi, ma anche altri tipi di dati più tecnici) che vengono condivise da miliardi di persone in ogni momento attraverso computer, smartphone e altri dispositivi elettronici. È uno scenario che offre prospettive senza precedenti: tracciare le epidemie di influenza, per esempio, o monitorare il traffico stradale in tempo reale, o ancora gestire l'emergenza in caso di catastrofi naturali. Per usare questa enorme mole di dati, però bisogna capirli, e prima ancora bisogna categorizzarli in maniera efficace, veloce e automatizzata. Uno dei sistemi più usati è un insieme di tecniche statistiche chiamate Cluster analysis (Ca), in grado di raggruppare i set di dati secondo la loro "somiglianza". Due ricercatori della Sissa, Alessandro Laio e Alex Rodriguez hanno ideato un tipo di Ca basato su principi semplici e potenti, che si è dimostrata molto efficiente e in grado di risolvere alcuni dei problemi più tipici in questo ambito di analisi. La loro ricerca è stata pubblicata ieri sulla rivista Science.