

Big Data, la Sissa scopre come trovare le città senza mappa

Categorizzare e rappresentare in maniera sintetica enormi quantità di dati (si parla di petta, o addirittura esa-byte di informazione) è la sfida del futuro. Una ricerca della Scuola Internazionale Superiore di Studi Avanzati (Sissa) di Trieste, pubblicata sulla rivista *Science*, propone una procedura efficiente per affrontare questa sfida.

Gli addetti ai lavori usano l'espressione *big data* per indicare grandissime quantità di in-

formazioni, 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. Quello attuale 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 mo-

le 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 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.

«Pensate a un database di fotografie di volti -», spiega Alessandro Laio, professore di Fisica e Statistica Biologica della Sissa -, nell'archivio ci possono essere più fotografie di una stessa persona, la Ca serve a raggruppare tutti gli scatti relativi allo stesso individuo. Questo tipo di analisi viene fatto per esempio dai sistemi automatici di riconoscimento dei volti. Noi abbiamo cercato di ideare un algoritmo più effi-

ciente di quelli attualmente usati, che risolva alcuni dei problemi classici della Ca». «Il nostro approccio si basa su un modo nuovo di individuare il centro dei cluster, cioè i sottoinsiemi - spiega Alex Rodriguez, autore insieme a Laio della ricerca -: provate a immaginare di dover individuare tutte le città del mondo, senza avere a disposizione una mappa. Un compito immane», spiega Rodriguez. «Abbiamo perciò individuato un'euristica, cioè una regola semplice, una sorta di scorciatoia per ottenere il risultato».