



Olimpiadi di Astronomia, pronta la squadra italiana

A Teramo sono stati scelti i sei studenti che parteciperanno alle Olimpiadi Internazionali Astronomia 2013, in programma dal 6 al 14 settembre a Vilnius, in Lituania. L'importante manifestazione è organizzata dall'Inaf in collaborazione con la Società astronomica italiana e il Miur. I sei selezionati tra 11 finalisti per cinque giorni hanno risolto problemi vari di Astronomia, Astrofisica, Cosmologia

Scienza

COMPLESSITÀ

Big data: sì, ma quanto?

di **Alessandro Vespignani**

Da circa un paio d'anni, *Big data* è sicuramente uno degli "slogan" più usati e abusati per riassumere in due sole parole la rivoluzione digitale che stiamo vivendo. Si è generata intorno al concetto di *Big data* un'aura addirittura mitologica, come se i *Big data* da soli potessero incarnare una forma di intelligenza superiore. Ovviamente, una delle prime domande che viene in mente conoscendo un po' di numeri è cosa significa "big". A Ginevra, il Cern produce da anni petabyte (10^{15} bytes) di dati in pochi minuti. Gli esperimenti astronomici fanno altrettanto, e il radio telescopio Ska che entrerà in funzione nel 2016 produrrà un Exabyte (10^{18} bytes) di dati al giorno. Al confronto, i 500 terabyte giornalieri che vengono postati su Facebook sono briciole. È in tale ottica che molti editorialisti del mondo

tecnologico si cimentano nel definire quale sia la soglia oltre cui i *Big data* possono essere definiti come veramente "grandi". Come se si potesse assegnare un bollino "Doc" dei *Big data* in funzione della loro grandezza. Per alcuni, bisogna avere almeno un terabyte, per altri il petabyte, per altri ancora la definizione deve cambiare con la tecnologia che permette di trattare questi dati.

Purtroppo tutta questa discussione è assolutamente futile. La definizione di *Big data* non può essere racchiusa in un numero, ma è invece la definizione di un nuovo modo di analizzare i sistemi. E tre ne sono i fattori cruciali.

La crescita relativa. Uno dei fattori rilevanti è la variazione relativa della dimensione dei dati che possono essere raccolti. Ad esempio, fino a dieci anni fa gli esperimenti che mappavano le interazioni sociali (chi è amico di chi, per dirne una) si limitavano a questionari destinati a poche decine di persone. Uno dei lavori più imponenti tracciava la rete di amicizia di una decina di classi

scolastiche negli Stati Uniti. Pochi kilobyte di informazione su cui si lavorava rosciando tutta l'informazione possibile per capire i meccanismi alla base della formazione delle reti sociali. Oggi, le compagnie di telefonia mobile tracciano le reti sociali di milioni di utenti, insieme alla loro posizione geografica, alla lunghezza delle chiamate telefoniche, e con la precisione temporale del secondo. Per questi dati, la cui dimensione è modesta rispetto ai petabyte ed exabyte generati in altri contesti, il "big" si riferisce al fattore di crescita relativo dei dati disponibili che aumenta di 100mila volte rispetto alle poche decine di individui che venivano studiate nel passato.

L'abbondanza. Un altro parametro fondamentale è l'abbondanza dei dati. Le stime ci dicono che nel 1986 il 92% dei dati venivano immagazzinati in forma analogica. Nel 2007 questo dato si inverte, e la digitalizzazione rappresenta il 94% del totale. Per digitalizzazione si intende che i dati possono essere integrati, assimilati e analizzati velocemente. I dati sono finalmente a portata di mano e una moltitudine di dati, anche se non troppo grandi, possono essere incrociati, fusi, confrontati, amplificando il loro potere di conoscenza. Le mappe di popolazione, anche le più dettagliate del mondo,

sono qualche gigabyte. I dati del trasporto aereo mondiale hanno più o meno la stessa dimensione. Possiamo poi aggiungere i dati di mobilità locale e le tracce che arrivano dalla telefonia cellulare, aggiungere i dati demografici e localizzarli geograficamente. Alla fine di questo processo di integrazione, inserendo tutto nel computer, possiamo creare modelli della mobilità sociale di notevole precisione. In questo caso il "big" deriva dalla forza combinata dei dati disponibili che si moltiplicano l'uno con l'altro per darci delle fotografie nuove, globali e dinamiche della realtà circostante.

La scienza dei dati. In fondo la scienza è da sempre basata sui dati. Ma quello a cui la "scienza dei dati" (dall'inglese *data science*) si riferisce è il fatto che si stanno sviluppando nuovi approcci scientifici per estrarre conoscenza da questo nuovo flusso di dati. Dati grandi e piccoli che sono il "prodotto" delle nostre attività quotidiane: prenotazioni di viaggio, telefonate, ricerche sul Web, email, instant messaging (Im), microblogging, transazioni di carte di credito eccetera. Questo mare di briciole digitali che noi tutti produciamo quotidianamente non è il frutto di esperimenti definiti o controllati in laboratorio. È un mare, invece, che contiene informazioni prevalentemente inutili, ciò che i tecnici defi-

PREMIO LAGRANGE

È il fisico e sociologo australiano Duncan J. Watts, ricercatore presso i laboratori Microsoft Research di New York, il vincitore del Premio Lagrange - Fondazione Crt 2013. Intitolato a Joseph-Louis Lagrange che nacque a Torino e di cui qui si ricordano i duecento anni dalla scomparsa, il premio è il primo ambito riconoscimento internazionale nel campo della scienza della complessità.

Lo scienziato lo riceverà giovedì 27 giugno, alle ore 17.30, presso il Teatro Vittoria di Torino nell'ambito di un evento in cui sarà consegnato anche il Premio Lagrange - Fondazione Crt per la comunicazione al giornalista Luna e in concomitanza con il trentennale della Fondazione (Istituto per l'Interscambio Scientifico di Torino) del cui presidente Alessandro Vespignani pubblichiamo un intervento.