

# L'idea di complessità dall'informazione al mondo reale, con qualche difficoltà\*

Stefano Crespi Reghizzi  
Politecnico di Milano  
stefano.crespireghizzi@polimi.it

16 marzo 2012

## Introduzione

La parola complessità ricorre tanto spesso nel parlare scientifico che ci si aspetterebbe che i concetti che attorno ad essa ruotano fossero definiti con lo stesso rigore e sostenuti da eguale consenso di altri concetti fisici quali l'energia o la velocità. Questa aspettativa è rinforzata dai riferimenti di molti divulgatori scientifici alla teoria della complessità, come se fosse una compatta e solida branca delle scienze esatte.

Il quadro scientifico, come ci appare, è ben diverso: non c'è *una* teoria della complessità, ma esistono diverse definizioni, alcune consolidate altre mutevoli, proposte in ambiti diversi, che schematicamente possiamo identificare con la fisica e la teoria dei sistemi, la teoria dell'informazione e l'informatica. A conferma di ciò, mentre scienziati e tecnici concordano sulla misura delle grandezze fisiche (ad es. la potenza in watt), per alcune delle proposte è dubbia la possibilità di effettuare misure sperimentali o stime numeriche della complessità così definita e il dibattito su quale sia la migliore definizione di complessità di un sistema fisico prosegue stancamente.

In effetti il diverso stato di tali concettualizzazioni è confermato dall'osservazione seguente. Il progresso delle scienze esatte ha dato un significato preciso a certi termini del linguaggio naturale (energia, potenza, pressione, ecc.) ed ora, nell'ambito del discorso scientifico, non è più ammesso l'uso di tali termini nel loro significato pre-scientifico. Al contrario, gli scritti sulla complessità di molti scienziati contemporanei ricorrono spesso all'intuizione e al senso naturale dei termini tecnici. Basti uno tra i tantissimi esempi: l'anno scorso, Sean Carroll, fisico del CalTech e fecondo scrittore sul tema della natura del tempo, in una conferenza intitolata *Raddrizzando il tempo* chiede ai partecipanti: "Perché la *complessità* o l'*interesse* destato dai sistemi fisici danno l'impressione di crescere con il tempo, raggiungere un massimo e poi decrescere, in contrasto con l'entropia che ovviamente cresce monotonicamente?" Qui i termini in corsivo nell'originale si appellano al senso naturale prescientifico. Il parlare di crescita della

---

\*Versione preliminare destinata alla pubblicazione su *Nuova civiltà delle macchine*.

complessità o del grado di interesse di un sistema fisico o biologico presuppone che sia possibile assegnare una misura di tali qualità, analogamente al dire che una Ferrari è più potente di una Panda secondo una precisa definizione quantitativa della potenza. Con la sua domanda Carroll vuole stimolare un dibattito su quali, tra le diverse definizioni formali di complessità di un sistema, catturi meglio l'idea intuitiva o psicologica del termine. Il dibattito, rilanciato nel suo Blog da Scott Aaronson, informatico del MIT, di cui riparleremo, ha coinvolto diversi scienziati, è tuttora in corso e testimonia la molteplicità dei pareri spesso in stridente disaccordo.

Lasciamo esprimere a Cosma Shalizi, già nel Santa Fe Institute (il più famoso centro di studi interdisciplinari sulla complessità) il dubbio e il disorientamento di fronte al moltiplicarsi di proposte di misura della complessità:

*C'est magnifique, mais ce n'est pas de la science.* (Molte delle ricerche non sono neanche così splendide.) Nelle parole dello stimato Dave Feldman (che mi ha insegnato molto di quello che so in questo campo ma ha delle opinioni un po' meno acide), questo settore è un "campetto" all'interno del sedicente studio della complessità. Ogni pochi mesi appare un nuovo articolo che propone una nuova misura di complessità, che in generale è una quantità che non è computabile per nessuno dei casi che ci potrebbero interessare, se ve ne sono. Quasi sempre tali quantità non hanno rapporto con nessun'altra variabile, sicché non trovano posto in una teoria che ci spieghi quando e come le cose diventano complesse; tali quantità sono di solito delle quantificazioni [in senso logico] per il puro piacere della quantificazione.

Non diversamente sul portale ufficiale del governo francese per la scienza il dossier *Che cos'è la complessità* così inizia:

Questa scienza è ancora ai primi passi, pur se la necessità di studiare i fenomeni complessi è stata vista abbastanza presto. Il termine complessità è di gran moda ma non è definito in modo unico. In effetti, ciò che è complesso non si lascia facilmente afferrare; forse non ci sarà mai una definizione accettata da tutti. Ciononostante, il senso che si dà alla complessità è abbastanza in accordo con l'intuizione.[8]

**Quattro sensi comuni** Per le scienze esatte la parola complessità ha dunque più significati diversi e allo stesso tempo correlati tra di loro, in corrispondenza con definizioni proposte da fisici, matematici, ingegneri dell'informazione, e biologi. Le definizioni tentano di catturare e formalizzare certi aspetti dei sistemi fisici (naturali e artificiali) e dei prodotti culturali, che nel parlare comune sono considerati complessi. Per comprendere in pieno la portata e i limiti delle tante e controverse definizioni, occorrerebbe una cultura scientifica profonda e rara nella sua vastità e interdisciplinarietà. Nella speranza di facilitare l'accostamento a una materia piuttosto difficile, vorrei illustrare con parole comuni alcuni sensi di complessità su cui si è puntata l'attenzione degli specialisti.

1. Nella folla è *complesso* seguire i movimenti delle persone.
2. È *complesso* prevedere l'evoluzione dei mercati finanziari.  
Nel film "Jules et Jim" la relazione di coppia tra i protagonisti è *complessa*.

3. Le regole degli scacchi sono più *complesse* di quelle della dama.
4. È un lavoro *complesso* mettere in ordine questo archivio.

Il primo esempio ci avvicina al punto di vista della meccanica statistica cui si deve la classica definizione di *entropia*, una grandezza il cui valore caratterizza il disordine delle particelle elementari di un sistema. Gli esempi 2 portano all'analisi dei *sistemi dinamici* che evolvono nel tempo sotto il controllo di leggi espresse da equazioni non lineari, che costituiscono il modello matematico, più o meno astratto. L'esempio 3 guarda la lunghezza del testo delle istruzioni, quanto più corto è, tanto meno complesso è il gioco: è il punto di vista della *teoria algoritmica dell'informazione* nota anche come complessità di Kolmogorov. L'ultimo esempio porta alla *complessità computazionale*, la quale misura il tempo (e lo spazio) necessario per risolvere un problema: è un concetto centrale scientificamente consolidato per l'informatica.

**Mondi continui, discreti o simbolici** Le teorie della complessità si differenziano anche per la natura dei sistemi studiati, fisici o non, e per la natura *continua* o *discreta* del mondo studiato. In un mondo continuo le variabili sono misurate da numeri reali con una precisione sempre migliorabile ricorrendo a strumenti più precisi; in un mondo discreto esse sono numeri interi rappresentabili con i simboli d'un alfabeto che può ridursi a due lettere sole. Una terza distinzione attiene infine al fatto che le variabili misurate siano probabilistiche o deterministiche.

Mentre la meccanica statistica e i sistemi dinamici nascono dalla fisica del continuo e si appoggiano sul calcolo differenziale, la teoria algoritmica dell'informazione e quella della complessità computazionale sono di matrice logico-informatica e trattano entità simboliche, eppure anch'esse vengono sempre più spesso applicate ai sistemi il cui modello fisico tradizionale è continuo. Ciò avviene in due modi diversi così schematizzabili:

- Un sistema continuo, come ad es. l'atmosfera, si rende discreto, nel tempo e nello spazio, scegliendo una griglia spaziale di maglia abbastanza fitta, misurando o calcolando i valori delle variabili di stato soltanto sui punti della griglia, a intervalli di tempo prefissati. Questa tecnica è diffusissima e applicata con successo in ogni campo, dalla epidemiologia alla dinamica dei fluidi, ecc. Ma per certi sistemi dinamici non lineari complessi il passaggio dal continuo al discreto presenta difficoltà e richiede risorse di calcolo eccessive.
- Certi sistemi fisici continui, osservati a livelli macroscopici, sono descritti in modo assai più espressivo da sistemi simbolici, gli esempi interessanti non mancano cominciando da quelli biologici. In biologia molecolare il genoma è visto nel modello di Watson-Crick come un testo le cui lettere  $\{A, C, G, T\}$  indicano le quattro basi. Così pure, nello studio del cervello un segnale nervoso è rappresentato da una serie di impulsi elettrici, simbolizzati da 0 o 1 a seconda dell'assenza o presenza dell'impulso. Passando dal livello delle cellule a quello dell'intero organismo, il fenomeno del linguaggio umano non solo è rappresentato (anche se non esaurientemente) da modelli simbolici, ma sta all'origine del ragionamento logico-deduttivo stesso. Infine non può mancare, pro domo mea, l'esempio dell'elettronica

dei calcolatori: i potenziali elettrici dei transistori sono grandezze continue, nel cui campo di variabilità si identificano due livelli di tensione, che si fanno corrispondere ai simboli logici vero e falso.

Impiegando un termine abusato, si può dire che nei sistemi viventi e in quelli dell'ingegneria dell'informazione "emergono", a certi livelli di aggregazione della materia, dei fenomeni di natura simbolica. Però non si deve dimenticare che, anche ai livelli dove il fenomeno più interessante è simbolico, persistono gli aspetti fisici che fanno prediligere il modello continuo: basti pensare al metabolismo delle cellule del cervello o alla dissipazione termica dei circuiti digitali.

Dovendo per brevità fare una scelta tra le teorie citate, opto per quelle logico-informatiche che sono più vicine alla mia cultura, mentre altri possono esporre meglio di me le teorie fisiche della complessità. Dopo un breve inquadramento storico, toccherò alcuni aspetti interessanti per comprendere il contributo che queste teorie possono dare alla comprensione di certe realtà.

## Definizioni di complessità, cenni storici

**Complessità in fisica** Secondo una visione diffusa tra gli addetti ai lavori (come introduzione si può vedere la voce *Sistemi complessi* di [7]), lo studio della complessità si presenta come un paradigma post-newtoniano per affrontare da un punto di vista unificato molti fenomeni che si presentano in sistemi fisici, ambientali e biologici costituiti da diverse parti interagenti. Fino alla fine del '800 si pensava che l'impressione di complessità data da tali sistemi fosse puramente dovuta alla nostra limitata conoscenza, essendovi un numero di variabili da studiare troppo elevato per riconoscere le regolarità sottostanti. Ma una serie di difficoltà di natura sperimentale e teorica portò all'abbandono di questa opinione; esse sono ben descritte nel classico libro di Ruelle [10] e brevemente qui richiamate: l'abbandono della meccanica classica nello studio dei gas a favore di quella statistica di Boltzmann e Gibbs; l'insolubilità del problema dei tre corpi; il comportamento caotico dei sistemi dinamici non lineari, fino a giungere all'indeterminismo al livello delle particelle, dove la meccanica newtoniana non vale, e ad essa subentra la meccanica quantistica che combina probabilità e determinismo.

Nella modellazione matematica dei sistemi dinamici il comportamento caotico è definito come l'evoluzione temporale di un sistema che dipende fortemente dalle condizioni iniziali. Il comportamento caotico fu poeticamente chiamato dal matematico Edward Lorenz *effetto farfalla*, perché perfino il lieve frullio delle ali può produrre enormi conseguenze sul futuro d'un sistema vasto come l'atmosfera. Già in un sistema newtoniano, purché dissipativo, si possono manifestare oscillazioni irregolari di apparenza aleatoria, ma prodotte da meccanismi deterministici. La conoscenza matematica di tali comportamenti risale a Poincaré, ma lo sviluppo maggiore dei modelli avvenne verso la metà del secolo scorso ed è legato, tra altri, ai nomi dei matematici Edward Lorenz e René Thom e del premio Nobel Ilya Prigogine.

L'impronta caratteristica della complessità sarebbe la possibilità per il sistema di evolvere in modi qualitativamente diversi, mostrando così la capacità di adattarsi all'ambiente. Tali comportamenti sono difficilmente prevedibili perché in certi momenti (biforcazioni) il sistema diventa instabile e piccolissime variazioni ambientali possono farlo propendere verso una o l'altra scelta.

La comparsa di comportamenti nuovi difficilmente prevedibili mediante le leggi elementari della fisica è una caratteristica di tali sistemi. Tale idea di *emergenza* è, come scrive Corning in un bel inquadramento storico [3], un concetto venerando che tuttora manca d'una definizione teorica rigorosa.

L'analisi dei sistemi dinamici si avvale di metodi matematici ormai consolidati che sono argomento di insegnamento universitario, in particolare nelle scuole di ingegneria.<sup>1</sup> Le applicazioni allo studio di sistemi reali non mancano e passano attraverso una modellazione dei fenomeni più o meno aderente alla realtà e l'uso di programmi di simulazione o di risoluzione di equazioni; a seconda del grado di astrazione, cambia la finalità dell'analisi. Ad un estremo la meteorologia ha raggiunto una buona e crescente affidabilità delle previsioni, all'altro estremo non sembra che finora questi metodi siano stati in grado di predire i comportamenti umani individuali o sociali, i crolli della borsa, o gli attentati terroristici, anche se spesso si leggono notizie giornalistiche di senso contrario. Un'insolita applicazione [9] di tali metodi matematici allo studio della relazione talvolta burrascosa tra una coppia di amanti, nelle trame dei film e nei sonetti del Canzoniere di Petrarca, ha destato curiosità e interesse nel pubblico.

**Trasmissione e complessità dei messaggi** Nel corso della seconda guerra mondiale, lo studio del miglior modo di codificare i messaggi trasmessi nelle reti di telecomunicazione portò alla definizione quantitativa di informazione, solitamente riferita al lavoro di Claude Shannon. La quantità di informazione presente in un messaggio dipende dalla varietà dei messaggi che la sorgente informativa potrebbe trasmettere e dalla loro probabilità. La storia racconta che venne dato il nome di entropia alla misura dell'incertezza associata a una lettera del messaggio perché la formula matematica ricordava quella della entropia termodinamica. In seguito il nesso concettuale tra l'entropia di Shannon e quella della meccanica statistica fu così spiegato: per una particella che raggiunge un certo microstato partendo dalla condizione iniziale, l'entropia misura l'informazione necessaria a descrivere la traiettoria percorsa dalla particella.

Si deve però osservare che la teoria di Shannon si applica a entità discrete (i messaggi sono codificati con le lettere d'un alfabeto), non alle variabili continue della fisica classica. Per passare dal mondo fisico a quello dell'informazione abbiamo visto che occorre discretizzare le variabili e dunque le misure delle grandezze fisiche, con ripercussioni profonde sui fondamenti stessi della fisica classica, come ci spiegano Rossella Fabbrichesi e Giuseppe Longo [4]:

Da Newton e Leibniz abbiamo infatti costruito una nuova oggettività fisico-matematica grazie al continuo. Con Poincaré, il continuo delle dinamiche classiche ha permesso di distinguere fra una determinazione globale (le equazioni di un sistema) ed una locale (la sensibilità del sistema a fluttuazioni al di sotto della misura possibile). E' il continuo, anche solo quello di Aristotele, che ci permette di concepire attività (fluttuazioni, perturbazioni ...) al di sotto di ogni intervallo di misura. Infatti, una organizzazione del mondo

---

<sup>1</sup>Un esempio è il corso tenuto da C. Piccardi al Politecnico di Milano: *Dinamica dei sistemi complessi*. L'insegnamento si rivolge a tutti coloro che sono interessati a studi di carattere generale riguardanti i modelli matematici e le loro applicazioni, in particolare a temi quali la dinamica non lineare, il caos, le reti complesse, i sistemi adattativi. Si tratta di temi, oltre che culturalmente significativi, anche di notevole attualità.

con la matematica del discreto, imporrebbe una taglia minima, una distanza fra i punti, senza permettere di concepire nulla, matematicamente, fra essi; un assoluto, quello della topologia discreta e della griglia soggiacente al mondo, limite della misura perfetta. Invece, con la geometria dei sistemi dinamici siamo arrivati a dare centralità a fenomeni possibilmente al di sotto dell'intervallo della misura fisica: una variazione, fluttuazione non misurabile, un non-nulla al di sotto della misura, può determinare l'evoluzione di una dinamica. Solo il continuo, non ponendo limiti inferiori alla misura (classica) ha consentito di sviluppare queste analisi, che, a partire da Poincaré, hanno rivoluzionato la meccanica classica.

**Teoria algoritmica dell'informazione o complessità di Kolmogorov** Nella telecomunicazione il messaggio è codificato dal trasmettitore, diffuso attraverso il mezzo trasmissivo e poi decodificato dal ricevitore. Codifica e decodifica sono svolte da dispositivi elettronici relativamente economici che realizzano certe semplici trasformazioni. Se al posto del codificatore/decodificatore poniamo un computer, ovvero più astrattamente un algoritmo, le trasformazioni possono essere molto più profonde e ci portano verso il paradigma della *teoria algoritmica dell'informazione* (come introduzione si rimanda alla voce *Algorithmic information theory* [5]), sviluppata negli anni 1960 da G. Chaitin, A. Kolmogorov e R. Solomonoff, indipendentemente.

Si parla ora di testo (o stringa o serie di cifre) anziché di messaggio, perché il contesto telecomunicativo non è più necessario. La *complessità algoritmica* CA di un testo è la lunghezza della più corta *descrizione* che se ne può dare. Un testo è semplice se, pur essendo molto lungo o al limite infinito, può essere descritto in poche parole, ad es. il testo (serie numerica) 3,1415... è semplicemente il rapporto (pi greca) tra la circonferenza e il diametro di un cerchio; all'altro estremo, una successione di cifre ottenute lanciando i dadi ha la massima complessità, non essendo possibile descriverla in altro modo che elencando tutte le cifre. Occorre precisare in quale linguaggio si scrive la descrizione: non essendo i linguaggi naturali abbastanza chiari ed esenti da ambiguità, la descrizione si dà con un algoritmo, ossia un programma che calcola il testo. Inoltre, seguendo una prassi diffusa nell'informatica teorica, si preferisce codificare l'algoritmo, non in un linguaggio di programmazione come ad es. Java, ma sotto forma di istruzioni di una macchina astratta di calcolo, la macchina di Turing che sta alla base della teoria del calcolo. Queste scelte sono in effetti abbastanza secondarie e irrilevanti, quasi come le scelte delle unità di misura nella meccanica.

Prendiamo dunque il programma, nella versione per macchina di Turing, che calcola le lettere o cifre del nostro testo, il numero pi greca, con il numero di cifre desiderato. Contiamo il numero di bit o di byte del programma, ecco che abbiamo la complessità algoritmica del testo. Non proprio, perché un programmatore più abile potrebbe inventare un algoritmo più piccolo che effettua lo stesso calcolo. La lunghezza di questo secondo programma abbasserebbe allora la CA del testo. La CA è definita in modo puramente ideale, come la lunghezza del più piccolo programma, pur sapendo che non c'è limite alla furbizia dei futuri programmatori.

Dunque la CA di un testo si misura in bit se l'alfabeto è binario (o in byte se l'alfabeto ha più caratteri). Naturalmente la stessa unità, il bit, misura la

lunghezza del testo stesso.

Immaginiamo di voler trasmettere ad un amico un lunghissimo testo di un milione di bit, che si può generare con un algoritmo di soli mille bit. Certo ci costerebbe meno tempo e denaro trasmettere l'algoritmo invece del testo, avvertendo l'amico che per leggerlo deve far eseguire l'algoritmo sul computer. Il rapporto tra la lunghezza del testo e quella dell'algoritmo ha un significato importante: è la *compressione* ossia la riduzione di memoria che si ottiene rappresentando il testo per mezzo del programma che lo genera.

Ma la complessità algoritmica classica trascura del tutto un aspetto certo non secondario del problema: immaginiamo che l'amico, ricevuto l'algoritmo e lanciata l'esecuzione, scopra con disappunto che dopo giorni e giorni in cui il suo computer non ha fatto altro, il testo non è ancora stato ricostruito; oppure che la memoria della sua macchina non è sufficiente per l'algoritmo! In entrambi i casi, l'algoritmo ha esigenze eccessive, richiede più *risorse* di tempo di calcolo o di capacità di memoria di quelle realisticamente disponibili. Nell'informatica si dice che la *complessità di calcolo* dell'algoritmo è eccessiva. Ritourneremo su questo punto, per presentare alcuni correttivi recenti alla teoria classica della complessità algoritmica.

Questa non è la sola critica che si muove al concetto classico di CA, ve ne è un'altra più fondamentale. E' impossibile calcolare o misurare la CA di un testo a causa del celebre teorema di indecidibilità di Turing (a sua volta derivabile da quello di Gödel) che esclude l'esistenza di un algoritmo capace di dire se due programmi sono equivalenti.

In effetti, il problema di calcolare la CA di una serie di lettere o numeri ricorda da vicino quello della *inferenza induttiva*, studiato in questo contesto da Solomonoff: data una serie scoprire la regola che permette di prevedere quale sarà il prossimo numero; ad es. dalla serie 3, 5, 7 si può indurre che i numeri debbano essere dispari, e quindi il prossimo sarà 9, ma altrettanto valida è l'induzione che la serie contenga i numeri primi, con la scelta 11 per il prossimo. Tra le due ipotesi la più semplice è la prima, perché, come anche il profano può intuire, il programma che genera i numeri pari è più semplice e breve di quello che genera i numeri primi. Si osserva allora un buon accordo tra la scelta della CA ed il criterio classico, noto come *rasoio di Occam*, che tra più possibili spiegazioni dei dati sperimentali conviene scegliere la più semplice; ma l'accordo non c'è sempre.

Resta la difficoltà di identificare la legge induttiva, o che è lo stesso, di trovare l'algoritmo generatore più piccolo. Potremmo scoprire che una serie di cifre che ci appare misteriosa e complicata, di colpo riveli il suo semplice segreto, causando così un crollo nella misura della CA che finora le avevamo assegnato. Ciononostante, vedremo ora una via, alquanto discutibile, per stimare sperimentalmente e grossolanamente la CA di un testo.

Le applicazioni dirette della teoria della CA non sono numerose (vedasi la voce *Applications of Algorithmic Information Theory* [12]), ma stanno diffondendosi certe utili tecniche di classificazione e analisi dei dati ispirate anche da tale teoria.

## Una misura universale di somiglianza tra oggetti simbolici

L'ingente quantità di testi o più in generale di *dati* scambiati attraverso la Rete ha stimolato l'invenzione di efficienti algoritmi di compressione, come il popolare programma *zip*, al fine di ridurre i tempi di trasmissione e le dimensioni delle memorie in cui immagazzinare i dati. Per certi tipi di dati, come quelli televisivi e musicali, è tollerabile che la compressione faccia perdere qualche dettaglio e peggiori la qualità dell'immagine o del suono, ma per i testi si richiede di solito assoluta fedeltà.

L'idea di usare i programmi di compressione come strumento per misurare seppure rozzamente la CA di un testo ha destato molto interesse e avuto successo. L'idea è semplice. Se schiaccio con il compressore *zip* un romanzo lungo, mettiamo, un milione di byte ottengo, mettiamo, un testo di mezzo milione di byte. Per rigenerare il romanzo userò l'algoritmo di decompressione del pacchetto *zip*, mettiamo di dimensione centomila byte. Dunque, facendo le somme, posso descrivere il romanzo con due pezzi: il testo compresso e il decompressore, per una dimensione totale di seicentomila byte. Questa viene da molti proposta come una misura della CA del romanzo, seppure approssimativa e dipendente dalle qualità del compressore utilizzato.

Si cita come esempio di successo la definizione d'una metrica di carattere universale per misurare la *somiglianza tra due oggetti* codificati da un testo. Vediamo come gli inventori del metodo (Vitanyi e Li [12]) lo descrivono:

Nella fisica classica sappiamo misurare la distanza fisica tra due oggetti. Nell'epoca dell'informazione è importante misurare la "distanza informativa" tra due oggetti: due documenti, lettere, messaggi, spartiti musicali, linguaggi, programmi, figure, sistemi o geni. Tale misura non dovrebbe dipendere dall'applicazione né essere arbitraria. La metrica universale di somiglianza è probabilmente il risultato pratico più rilevante della teoria algoritmica dell'informazione. E' ragionevole definire la somiglianza tra due oggetti come la difficoltà con cui uno di essi può essere trasformato nell'altro. Formalmente, si può definire la somiglianza tra due testi  $x$  e  $y$  come la lunghezza del più piccolo programma che partendo da  $x$  calcola  $y$  e viceversa.

Per stimare la somiglianza occorrerebbe calcolare la CA della trasformazione di un oggetto nell'altro ma, non essendo essa computabile, si ricorre alla tecnica di compressione. Un esempio può chiarire le idee: immaginiamo di voler confrontare due appartamenti opportunamente descritti da due testi (come quelli dei programmi CAD di disegno automatico usati dagli architetti); gli appartamenti possono assomigliare al livello macroscopico della pianta, o al livello, di scala inferiore, degli arredi. Un buon compressore è in grado di estrarre gli elementi comuni nelle descrizioni dei due appartamenti (la pianta o l'elenco degli arredi) e di scriverli una sola volta nella descrizione compressa. Essa di conseguenza sarà molto più breve della somma delle due descrizioni separate, proprio a causa del fatto che vi sono somiglianze.

La distanza informativa tra due testi si calcola dunque così: l'unione dei due testi si comprime ottenendo un testo di cui si nota la lunghezza. Con lo



stesso compressore si trattano i due testi separatamente e, tra le due misure di lunghezza così ottenute, si nota la minore. Si confrontano poi le due misure annotate, dopo certi aggiustamenti su cui sorvoliamo. La differenza tra le due misure viene presa come la misura di somiglianza tra i due testi; essa è nota come *distanza comune normalizzata* (NCD normalized common distance). Questo metodo è stato applicato in genetica per ricostruire in modo automatico gli alberi evolutivi delle specie in base alla distanza tra i genomi, oltre che in altri problemi di classificazione automatica, per esempio per attribuire la paternità d'uno scritto a uno tra più autori.

Ma non per tutti gli oggetti è disponibile una codifica testuale, come per il genoma di un topo o per un romanzo di Buzzati. Per altri oggetti, come ad esempio i vocaboli “casa” e “rosso”, si può pensare che il testo esista ma in senso virtuale, essendo costituito da tutti i documenti contenenti tali parole. Tale corpus sterminato è una base comune di conoscenze del genere umano e riflette i contesti in cui si usano le due parole. Usando un motore di ricerca (come Google o Yahoo) si possono contare i documenti che contengono entrambi i vocaboli,  $x$  e  $y$  o soltanto l'uno o soltanto l'altro. Il motore di ricerca può essere visto come un compressore di documenti che da essi estrae le parole che servono agli scopi per cui lo strumento stesso è stato costruito: comprensione e classificazione dei testi, ricerche tematiche, traduzione, ecc. Applicando lo stesso metodo NCD visto sopra, si calcola così la “distanza semantica tra due parole  $x$  e  $y$ .”<sup>2</sup>

Da altri specialisti viene però una critica a questi metodi: la misura della complessità algoritmica di un oggetto, e dunque anche della somiglianza tra oggetti, dipende eccessivamente dal particolare compressore utilizzato. Un compressore può essere o meno capace di riconoscere in un testo certe regolarità o sottostrutture, anche quando esse sono facili da scoprire per chi ha occhio. Un esperimento condotto da Shalizi mostra l'incapacità dei compressor attuali di riconoscere una pur semplice legge matematica usata per generare una serie di numeri (i quadrati delle cifre del numero  $\pi$  greco). Ciononostante, queste e altre misure di distanza informativa o semantica sono spesso usate per ricercare documenti simili nei campi più vari.

## Risorse di calcolo nella complessità algoritmica

**Una serie aleatoria è complessa?** Nel senso comune un testo generato in modo casuale lanciando dei dadi alfabetici non è da considerarsi complesso forse perché per noi non ha un senso: “orrida casualità dominio del caos, assenza del senso”, dice Nietzsche. Al contrario, la CA di tale testo è massima perché non si può dare una regola che permetta di rappresentare il testo in modo più breve. Il valore della CA di un testo casuale è quindi maggiore di quello di un'opera letteraria! Questa conclusione può lasciare perplessi e sono in molti a pensare che la teoria algoritmica dell'informazione risponda bene alla domanda “che cosa è la casualità”, ma caratterizzi meno bene ciò che intuitivamente si considera complesso. Invece di questo termine, la comunità scientifica che si occupa di questi problemi ama parlare di ciò che è “interessante” usando

<sup>2</sup>Per chi apprezza le formule, tale distanza è data da  $\frac{\max\{\log f(x), \log f(y)\} - \log f(x,y)}{\log N - \min\{\log f(x), \log f(y)\}}$  dove  $f(x)$  è il numero di pagine contenenti  $x$ ,  $f(x,y)$  è il numero di pagine contenenti  $x$  e  $y$  e il numero  $N$  dà il numero totale di pagine indicizzate dal motore di ricerca.

una parola ancora più carica di connotati antropologici e psicologici. Molti di loro fanno notare che un testo casuale è di solito privo di interesse allo stesso modo di un sistema fisico che si trova in uno stato di massimo disordine delle particelle, ossia di massima entropia. Sarebbero invece interessanti i sistemi che presentano un elevato grado di organizzazione interna.

Molte tra le troppe proposte di correttivi alla teoria della CA, di cui si lamentava Shalizi, mirano a riformare la definizione in modo che la complessità di una serie casuale risulti piccola anziché massima. La maggior parte di queste proposte combina la CA con qualche elemento della complessità computazionale. Il terreno concettuale si fa ancora più arduo per i non addetti ai lavori, e non senza esitazione nel prossimo paragrafo schematizzo alcune proposte che hanno avuto qualche successo. Come detto, una lacuna nella teoria della complessità di Kolmogorov è di non prendere in considerazione il tempo e lo spazio di memoria necessari all'algoritmo per ricostruire la serie di simboli. Immaginiamo due testi del tutto diversi ma aventi la stessa complessità algoritmica, ossia descritti da due algoritmi di pari lunghezza. Può succedere che il tempo necessario al primo algoritmo per generare il primo testo sia assai maggiore di quello necessario al secondo algoritmo per il secondo testo. In tale caso, la definizione CA non sarebbe in accordo con la nostra intuizione: d'accordo che i programmi hanno le stesse dimensioni, ma se il primo richiede un'eternità di tempo o un universo di memoria per fare il suo lavoro di ricostruzione, saremmo portati a considerarlo più complesso del secondo. Alla ricerca di una definizione più vicina all'intuizione, gli specialisti hanno approfondito questi aspetti, ottenendo però risultati assai meno eleganti e essenziali di quelli della teoria originale (si può vedere la voce *Complessità* curata da Sporns [11]). In genere tali definizioni si propongono due obiettivi: far sì che la complessità di una serie aleatoria ovvero di un sistema disordinato sia minima, e permettere un migliore raccordo tra la teoria della complessità algoritmica e la teoria dei sistemi fisici dinamici.

Le misure della complessità più note possono essere classificate in due categorie principali. Alla prima appartengono le proposte (contenuto informativo algoritmico e profondità logica [discussa dopo]) che catturano la casualità, l'informazione contenuta o la lunghezza della descrizione di un sistema o processo; i processi casuali hanno la complessità più alta perché sono i più resistenti alla compressione. La seconda categoria (comprendente la complessità statistica, quella fisica e quella neuronale) concettualizza la complessità in modo diverso dalla casualità. Qui i sistemi complessi sono quelli ricchi di struttura o di informazione, spesso disposta a più livelli lungo la scala temporale o spaziale. All'interno di questa categoria la posizione dei sistemi complessi sta in mezzo tra i sistemi molto ordinati (regolari) e quelli molto disordinati (casuali). ... si deve sottolineare che al momento non esiste una definizione quantitativa di complessità e di disordine che sia generalmente accettata.

Non essendo possibile descriverle neppure sommariamente tutte queste misure di complessità, ne scegliamo una delle meno effimere, a titolo di esempio.

La cosiddetta *profondità logica* (logical depth, Bennett, 1988) si fonda sulla CA e la completa con una misura classica della teoria della complessità computazionale: il tempo di calcolo. La profondità logica di un testo è il tempo

necessario a generare il testo utilizzando il più piccolo programma che lo descrive. Come la CA, anche questa nuova definizione è una misura del processo generativo e non si applica direttamente a un sistema fisico, a un organismo o a un processo dinamico, come ben spiegava Arecchi [2]:

C. Bennett ha proposto una diversa definizione di organizzazione chiamata la *logical depth* di un sistema. Questa è “la difficoltà di generare una descrizione del sistema a partire dal più corto algoritmo”. Un sistema di complessità organizzativa elevata, come un organismo vivente, richiederebbe un calcolo lungo ed elaborato per essere descritto. Ciò riflette la lunga ed elaborata serie di passi nell’evoluzione dell’organismo. D’altra parte, possiamo descrivere un cristallo con un un calcolo semplice svolto da un algoritmo singolo.

E per un sistema disordinato come un gas? Non si può fare altro che partire dal testo lunghissimo che elencherebbe gli stati delle particelle elementari e darlo come dati di ingresso al cortissimo programma “Print”, che eseguirebbe un numero di passi eguali al numero di stati descritti. Si tratta dunque di un programma di complessità computazionale bassa, più precisamente lineare. La CA del gas (somma della lunghezza del testo e del programma) è grande, mentre è piccola la profondità logica.

La *logical depth* di un oggetto – la sua complessità – è misurata dal tempo necessario a un computer per simulare lo sviluppo completo dell’oggetto cominciando con l’algoritmo elementare e senza prendere scorciatoie. La complessità, in questo senso è una misura di quanto sia difficile mettere insieme una cosa partendo da pezzi elementari.

Anche questa misura di complessità non è effettiva perché, come per la CA, la scelta della descrizione minima è indecidibile.

Arecchi mette poi in evidenza che, anche se il comportamento di un sistema fisico può concettualmente essere simulato passo-passo nella sua evoluzione, il calcolo potrebbe essere troppo lungo. In effetti molta parte della fisica ha studiato delle scorciatoie per calcolare lo stato finale senza tracciare ogni passo. Ma ciò può riuscire soltanto se le operazioni del calcolatore sono più potenti di quelle che l’oggetto dello studio (un sistema fisico) può esso stesso eseguire. Per esempio il sistema solare può essere pensato come un calcolatore che esegue le equazioni di Newton. Tornando al calcolo, esso è eseguito da un computer, che è un esempio di sistema fisico. Di conseguenza “esso può determinare l’esito della propria evoluzione solo eseguendola completamente passo a passo: non ci sono scorciatoie. Tale irriducibilità computazionale ha luogo ogni volta che un sistema fisico può agire da computer”. Secondo Arecchi i casi in cui la complessità del calcolo può essere ridotta attraverso le scorciatoie delle leggi fisiche sarebbero l’eccezione e non la regola: “è possibile che a molte domande fisiche non si possa rispondere che attraverso quantità irriducibili di calcolo.” Se tali calcoli si basano su limiti idealizzati per quanto riguarda il tempo, lo spazio (memoria) o la precisione numerica, essi potrebbero richiedere computazioni arbitrariamente lunghe tanto da essere praticamente indecidibili (Wolfram).

Dopo Bennett, vari raffinamenti sono stati proposti, alcuni dei quali da Seth Lloyd, professore di meccanica quantistica al MIT, che ne parla nel piacevole

libricino *L'universo come programma* [6]. Da fisico qual'è egli ha proposto, insieme ad altri, varie combinazioni dei concetti della complessità algoritmica, della complessità o profondità logica di Bennett, e dei concetti fisici di energia ed entropia. Il nome tecnico della prima delle sue definizioni è “thermodynamic depth” o *profondità termodinamica*, ma l'autore concede che

la profondità logica e quella termodinamica non sono le sole misure per quantificare qualche aspetto della complessità. A seconda delle caratteristiche di un sistema complesso che si vogliono studiare, esistono altre misure altrettanto o più utili. Una di esse è la quantità chiamata “effective depth”, *profondità effettiva*, che misura la quantità di regolarità presente in un sistema.

Veniamo allora alla “profondità effettiva”, un'idea dovuta a Murray Gell-Man e Lloyd stesso, non priva di fascino. Ogni sistema fisico contiene “una certa quantità di informazione, quella che descrive lo stato fisico del sistema con la precisione concessa dalla meccanica quantistica”. Ma soltanto una piccola parte di questa informazione è interessante per chi deve interagire con il sistema, la rimanente è fatta di bit che si possono qualificare (soggettivamente) “spazzatura”. Ad esempio, un aeroplano è descritto dai disegni costruttivi e dagli schemi dei motori, dalla composizione chimica delle leghe, dalle lavorazioni meccaniche per costruire le ali, ecc. La forma delle ali o la composizione delle leghe descrivono delle regolarità del progetto, indispensabili affinché l'aereo funzioni come voluto. Ma i disegni tecnici non specificano la posizione di ogni atomo nel metallo delle ali! I tantissimi bit che specificano queste posizioni sono accidentali, non servono per capire se l'aereo può volare e non contribuiscono alla sua complessità *effettiva*.

Anche questa definizione nasconde una petizione di principio perché si tira dietro “il giudizio su che cosa sia o non sia una regolarità”, ovvero se un certo bit che descrive lo stato del sistema sia importante o meno, e per chi. Per un aereo la risposta può essere data dal progettista, ma per un organismo vivente non si sa da chi, forse da un principio di selezione naturale.

Le proposte non finiscono qui e ci troviamo di fronte la pletora di definizioni lamentata da Shalizi, ognuna delle quali aspira a cogliere uno dei tanti sensi intuitivi della complessità. Una fresca ma non assestata discussione su questi temi è ospitata nel Blog di Scott Aaronson, studioso della complessità e professore del MIT [1].

## 1 Conclusione

Da oltre mezzo secolo si assiste allo sforzo degli scienziati per caratterizzare in modo preciso le proprietà che ai nostri occhi rendono un sistema complesso e per dare un senso quantitativo, una misura, alla complessità. Per gli oggetti testuali, la più bella proposta è la complessità algoritmica (o di Kolmogorov) che ben esprime la maggiore o minore semplicità della legge matematica che permette di generare il testo per mezzo di un algoritmo. Se tale legge matematica non è nota, l'oggetto testuale viene considerato (provvisoriamente) casuale. Questa definizione di complessità ha influito molto su tutte le teorie della complessità e anche sullo sviluppo di tecniche empiriche di confronto e di classificazione di oggetti testuale, fondate sui programmi per la compressione dei dati.

Se oltre alla semplicità della legge matematica si considera anche la difficoltà di applicarla, si arriva al concetto di Bennett, la profondità logica, che si focalizza sul tempo, misurato come numero di passi elementari, necessario per produrre l'oggetto ovvero per condurre il sistema nel suo stato attuale.

Al di là di queste due eleganti definizioni, la scena delle proposte più recenti appare confusa e instabile. Circa vent'anni fa, al congresso sulla complessità dell'Accademia Pontificia delle Scienze, la mancanza di consenso tra gli scienziati poteva ancora essere attribuita allo stadio infantile di tali ricerche [2]:

Gli studi sulla complessità, benché allo stadio infantile, sono controversi. Attraverso la visione di sistemi diversi, nell'ambito delle scienze biologiche, fisiche e persino sociali, come una serie di "agenti" interattivi, i ricercatori impegnati sulla complessità esplorano un insieme ricco e variegato di fenomeni quali l'autoorganizzazione e l'adattamento. Essi usano tali fenomeni per modellare e spiegare avvenimenti quali la formazione delle città, il comportamento delle borse, e l'origine della vita. In questo volume alcuni dei protagonisti scientifici a livello mondiale discutono il ruolo della complessità attraverso tutte le discipline scientifiche. Le opinioni sono differenti: per alcuni la complessità è la chiave per giungere a una comprensione più profonda e completa del mondo attorno a noi; per altri, essa non è altro che una versione moderna della pietra filosofale. Al lettore il libro dà sprazzi e vedute su un modo, tra i più discutibilmente stimolanti, per accostarsi alla scienza. (dall'introduzione di Pullman)

Vent'anni dopo, il quadro generale non sembra cambiato, la scusante dello stadio ancora infantile delle ricerche è meno convincente e l'eccitazione per questo tipo di indagini è molto diminuita. Ma, diversamente da allora, oggi gli specialisti stessi si dicono convinti dell'impossibilità di trovare una definizione di complessità che vada bene per tutto e per tutti. Lloyd, dopo aver recensito e classificato qualche anno fa oltre quaranta diverse definizioni tra cui anche le sue, fa così il punto:

Chiedere come si misura la complessità è come chiedere come si misura la fisica. Nelle leggi fisiche molte sono le quantità misurabili, energia, distanza . . . , ma la "fisica" stessa non è una quantità misurabile. Similmente, nell'identificare le leggi della complessità dovremmo aspettarci che in esse compaiano tante quantità misurabili che compongono un sistema complesso.

Ma anche questa posizione potrebbe essere troppo ottimistica: nessuno ci assicura che si possano formulare delle "leggi della complessità" che abbiano la stessa utilità delle leggi della fisica.

Sarebbe però sbagliato negare l'interesse delle ricerche, anche in corso, che affrontano il problema della complessità in ambiti specialistici diversi; ad esempio, sono attuali gli studi sulla complessità delle reti che, come Internet, mettono in relazione un vasto insieme di "agenti" sociali, persone o computer.

**Ringraziamenti** a Francesco Bruschi per le letture consigliate e le vivaci discussioni; a Carlo Piccardi e Sergio Rinaldi per i riferimenti ai sistemi dinamici complessi; a Calogero Caltagirone per il riferimento a Nietzsche.

## Riferimenti bibliografici

- [1] Scott Aaronson, *Philosophy and theoretical computer science, in defense of Kolmogorov complexity*, 2011, <http://www.scottaaronson.com/blog/?p=762>.
- [2] (Ed.) Bernard Pullman, *The emergence of complexity in mathematics, physics, chemistry, and biology*, Pontificia Accademia delle Scienze, Princeton University Press, Princeton, 1997.
- [3] P. A. Corning, *The re-emergence of emergence*, *Complexity* **7** (2002), no. 6, 18.
- [4] Rossella Fabbrichesi and Giuseppe Longo, *Il mito del continuo tra filosofia e scienza : introduzione al convegno*, *Dedalus*, rivista di scienza, filosofia, cultura **2** (2007), no. 2.
- [5] M. Hutter, *Algorithmic information theory*, *Scholarpedia* **2** (2007), no. 3, 2519.
- [6] Seth Lloyd, *Programming the universe, a quantum computer scientist takes on the cosmos*, Vintage books, London, 2006.
- [7] G. Nicolis and C. Rouvas-Nicolis, *Complex systems*, *Scholarpedia* **2** (2007), no. 11, 1473.
- [8] Le portail de la science, *Qu'est-ce que la complexité ?*, 2001, <http://www.science.gouv.fr/fr/dossiers/bdd/res/3205/qu-est-ce-que-la-complexite>.
- [9] Sergio Rinaldi, *Laura and Petrarch: an intriguing case of cyclical love dynamics*, *SIAM Journal on Applied Mathematics* **58** (1998), no. 4.
- [10] David Ruelle, *Caso e caos*, Bollati Boringhieri, Torino, 1995.
- [11] O. Sporns, *Scholarpedia* **2** (2007), no. 10, 1623.
- [12] P. M.B. Vitanyi and M. Li, *Applications of algorithmic information theory*, *Scholarpedia* **2** (2007), no. 5, 2658.