

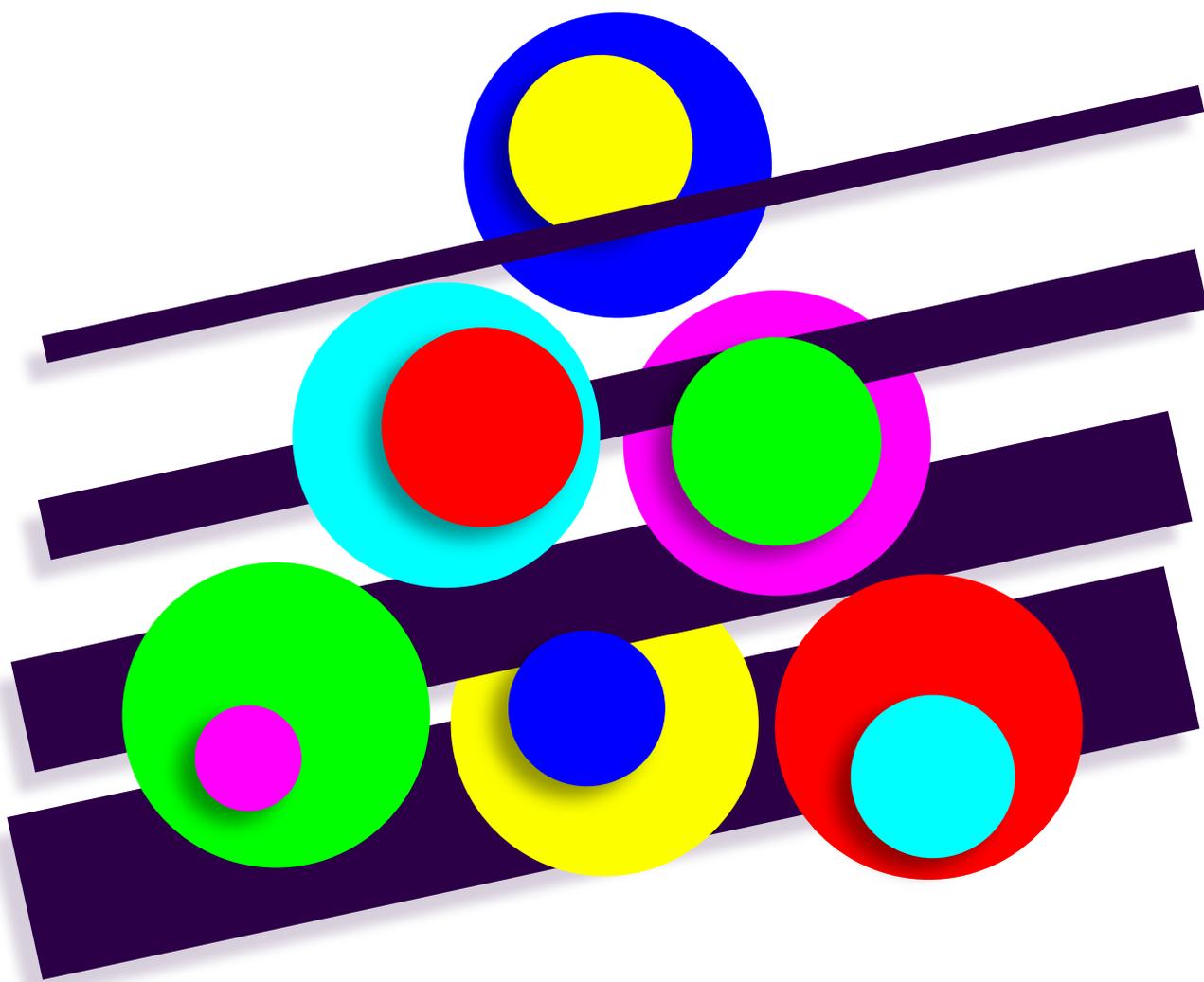
Numero XVII
Anno 2021


DIPARTIMENTO DI MATEMATICA E FISICA
"Ennio De Giorgi"


UNIVERSITÀ
DEL SALENTO

Viaggio nella Scienza

Ithaca



Dare i numeri: (ab)usi della Matematica (B)

Ithaca: Viaggio nella Scienza

Una pubblicazione del Dipartimento di Matematica e Fisica “Ennio De Giorgi” dell’Università del Salento.

Registrazione presso il Tribunale di Lecce n. 6 del 30 Aprile 2013.
e-ISSN: 2282-8079

Direttore Responsabile
Luigi Spedicato.

Ideatore
Giampaolo Co’.

Comitato di Redazione
Adriano Barra,
Rocco Chirivì,
Paolo Ciafaloni,
Maria Luisa De Giorgi,
Vincenzo Flaminio,
Luigi Martina,
Giuseppe Maruccio,
Marco Mazzeo,
Francesco Paparella,
Carlo Sempi.

Segreteria di Redazione
Daniela Dell’Anna.

© 2013-2023 Dipartimento di Matematica e Fisica “*Ennio de Giorgi*”.

© 2023 per i singoli articoli dei rispettivi autori.

Il materiale di questa pubblicazione può essere riprodotto nei limiti stabiliti dalla licenza
“*Creative Commons Attribuzione – Condividi allo stesso modo 3.0 Italia*” (CC BY-SA 3.0 IT).

Per il testo della licenza: <http://creativecommons.org/licenses/by-sa/3.0/it/deed.it>

Ithaca: Viaggio nella Scienza

è disponibile sul sito:

<http://ithaca.unisalento.it/>

Scriveteci all’indirizzo:

ithaca@unisalento.it

- 3** In questo numero
- 5** **Matematica epidemiologica per COVID-19**
Simone Bianco, Sara Capponi, James H. Kaufman
- 13** **L'arte e la scienza di imparare dai dati: la biostatistica, dalle mutazioni genetiche all'intelligenza artificiale.**
Serena Arima
- 21** **Meccanica statistica per reti complesse**
Tiziano Squartini
- 35** **Materiali fatti di molecole macroscopiche: dai granulari alla materia attiva**
Andrea Puglisi
- 47** **Le nuove frontiere del volo spaziale: Nanotecnologie e Biomedicina**
Giada Graziana Genchi, Gianni Ciofani
- 55** **Perché la Regina aveva torto, pur avendo ragione**
Michele Grillo

63 Dante e l'ipersfera

Giuseppe De Cecco

In questo numero

Questo è il secondo volume del numero XVII di **Ithaca**, anche questo centrato sul tema dell'applicazione della Matematica all'analisi di quanto osserviamo.

Il primo articolo, scritto da Bianco, Capponi e Kaufman, parla di un argomento di attualità: la modellizzazione matematica dello sviluppo dell'epidemia di Covid 19.

L'articolo di Arima tratta dell'applicazione della Matematica in generale, e più specificamente della Statistica, per analizzare sistemi biologici.

L'applicazione della Matematica per lo studio di reti complesse è l'argomento di cui tratta l'articolo di Squartini.

Un'altra applicazione della Matematica, declinata dalla Meccanica Statistica, è presentata da Puglisi, e tratta lo studio della materia organizzata in una maniera peculiare: in forma granulare.

Gli ultimi tre articoli del volume non sono direttamente collegati al tema generale. L'articolo di Genchi e Ciofani presenta un panorama degli studi di Biologia legati alla Nanotecnologia che sono stati, e saranno, sviluppati nelle missioni spaziali.

Nel suo articolo Grillo contestualizza all'ambito dell'Economia, e contesta, le definizioni di Scienza presentate da Boero e Co' in un loro articolo apparso su **Ithaca Educational**.

Conclude il volume un omaggio a Dante Alighieri nel 700-esimo anniversario della sua morte. De Cecco analizza l'universo come è presentato nella Divina Commedia riscontrandone una struttura che lo avvicina alle moderne descrizioni geometriche.

Buona lettura,
il Comitato di Redazione

Matematica epidemiologica per COVID-19

Simone Bianco

IBM Almaden Research Center, NSF Center for Cellular Construction

Sara Capponi

IBM Almaden Research Center, NSF Center for Cellular Construction

James H. Kaufman

IBM Almaden Research Center

Introduzione

L'epidemia di COVID-19, una malattia altamente contagiosa causata dal virus RNA SARS-CoV-2, ha causato finora oltre 110 milioni di infetti e più di 2.4 milioni di morti nel mondo. Fin dall'inizio, insieme all'enorme sforzo umanitario e medico, la comunità scientifica si è unita per ottenere il maggior numero di informazioni possibili sul virus e per identificare strategie di controllo del contagio. Il risultato è stato lo sviluppo rapido di diversi vaccini ad alta efficacia, alcuni dei quali sono stati già adottati, da diverse nazioni a livello mondiale, mentre altri sono negli ultimi stadi della sperimentazione clinica (ricordiamo ad esempio che il vaccino italiano sviluppato dall'azienda biotecnologica ReiThera in stretta collaborazione con lo Spallanzani è entrato nella fase due solamente a Marzo 2021).

Poiché la distribuzione del vaccino è attualmente in corso, lo studio dell'epidemiologia della malattia è fondamentale per capire come controllarne la diffusione e per prevenire o prepararsi o probabili pandemie future. I modelli matematici di diffusione delle malattie sono stati sviluppati nel primo periodo della pandemia, e hanno catturato ciò che si sapeva sulla diffu-

sione della malattia in quel preciso momento. I primi studi sono stati molto influenti nella realizzazione e attivazione di diversi interventi di tipo non farmaceutico (*Non pharmaceutical intervention*, NPI), messi in atto da tutti i paesi del mondo. Gli NPI variano in termini di rigore, durata e applicazione da un paese all'altro, determinando quindi distinti livelli di contenimento e controllo della pandemia a seconda della nazione a cui ci si riferisce. L'uso di rigorose strategie di quarantena in Cina, Colombia, Perù e altri paesi ha ridotto il numero di riproduzione di base, R_0 , definito come il numero di possibili contagiati dovuto ad un solo infetto, ad un valore inferiore ad 1 per diversi periodi, frenando fortemente l'epidemia [1]. Altri paesi hanno applicato regolamentazioni NPI meno rigide, come il distanziamento sociale e la restrizione di attività al chiuso, riducendo il tasso di trasmissione di circa la metà [1].

In questo articolo forniamo una breve panoramica dei vari modelli epidemiologici che sono stati proposti durante la prima fase della pandemia, ovvero quella di tipo strettamente epidemico. In particolare, discuteremo dei modelli matematici che hanno incorporato le caratteristi-

che specifiche del virus SARS-CoV-2 rispetto ad altri agenti patogeni come la trasmissione asintomatica e gli effetti delle regolamentazioni NPI non solo in termini della diffusione della pandemia ma anche rispetto alla mobilità umana. Infine esporremo gli effetti sulle previsioni fornite dai modelli matematici quando viene trascurata la diffusione locale del virus rispetto a quella nazionale.

Breve storia della modellizzazione matematica delle malattie infettive

I modelli matematici della diffusione delle malattie infettive sono stati per secoli un metodo efficace per controllarle e prevenirle. Il primo studio sistematico che ha incluso l'uso di modelli matematici per le malattie infettive risale a John Graunt (1620-1674) [2]. Circa un secolo più tardi, nel 1760, il fisico e matematico svizzero D. Bernoulli sviluppò il primo modello matematico che descrive la dinamica del vaiolo [3]. Oggi, i modelli epidemiologici basati su equazioni differenziali vengono applicati ordinariamente nello studio delle malattie infettive, trasmesse da vettori e zoonotiche in tutto il mondo [4].

Quando la diffusione di una malattia infettiva è abbastanza ampia da essere descritta dalla matematica continua, la maggior parte degli studi epidemiologici utilizza modelli compartimentali, cioè modelli che dividono la popolazione in compartimenti in base al loro stato di infezione e modellano la transizione tra compartimenti utilizzando equazioni matematiche. Ad esempio, la popolazione può essere suddivisa tra suscettibili alla malattia, cioè individui che non sono mai venuti in contatto con la malattia, persone in cui la malattia è in incubazione, cioè la malattia è presente ma non può ancora essere trasmessa individui infettivi ma asintomatici, ecc. A livello governativo, questi modelli sono stati per lungo tempo usati in modo efficace per fare delle previsioni sull'andamento delle malattie infettive e nella gestione pubblica delle epidemie. Tuttavia, non sono l'unico modo per studiare e caratterizzare le malattie infettive. Altri tipi di simulazioni usano modelli basati su agenti (*agent-based modeling*, ABM), ovvero modelli computazionali in cui

le interazioni tra le persone (agenti) di una popolazione sono esplicitamente definite e calcolate, oppure reti complesse, ovvero costrutti matematici che dividono le popolazioni nelle comunità locali che interagiscono tra loro in termini di nodi e *edge*. Rispetto ai modelli matematici compartimentali, queste sono le alternative praticabili specialmente nelle prime fasi di crescita esponenziale dell'epidemia, in cui è importante cogliere la natura stocastica della diffusione della malattia da persona a persona in piccole popolazioni e realtà locali, quali uffici, mezzi di trasporto, luoghi religiosi, ecc. In questo modo si è in grado di fornire informazioni di grande interesse sulla diffusione delle infezioni in una comunità locale. La granularità aggiuntiva offerta da questi metodi però ha un costo talvolta alto dovuto ad una maggiore complessità computazionale.

La pandemia di COVID ha evidenziato la necessità di una rapida modellazione della malattia. La disponibilità di *software* accessibili gratuitamente (*open source*) ha accelerato questo processo, consentendo trasparenza e responsabilità e promuovendo la collaborazione e la rapida diffusione della conoscenza. Lo *Spatio-Temporal Epidemiological Modeler* (STEM) è un esempio di *software open source* usato ampiamente nel passato per descrivere la diffusione di malattie infettive quali ad esempio SARS, Ebola, H1N15. STEM è parte della Eclipse foundation ed è guidato da IBM Research, dall'Istituto federale tedesco per la valutazione dei rischi (BfR) e da altri istituti [5]. Con la diffusione mondiale del COVID, altri consorzi stanno sviluppando modelli *open source* per catturare l'epidemiologia di questa pandemia [6].

Modelli di trasmissione asintomatica

Una caratteristica cruciale del COVID-19, che lo distingue da altre infezioni da Coronavirus come la SARS e la MERS, è il potenziale di causare una malattia così lieve che i suoi sintomi possono passare inosservati in alcune persone. Questi casi sono difficili da monitorare e contribuiscono a quella che viene comunemente chiamata trasmissione asintomatica. Nei casi in cui i sintomi invece sono gravi, spesso si richiedono

cure critiche e ospedalizzazione, causando un tasso di mortalità del 2-10% a seconda della regione e dei fattori di rischio individuali. Sebbene la ragione della trasmissione asintomatica sia ancora sconosciuta, l'effetto sulla prima fase dell'epidemia è stato devastante e probabilmente la trasmissione asintomatica è stata la causa principale della rapida diffusione mondiale di COVID-19 [7].

All'inizio della pandemia molti studi hanno pertanto cercato di affrontare gli effetti della trasmissione asintomatica. Inoltre, poiché i rapporti mostravano l'importanza della trasmissione sia asintomatica che pre-sintomatica, nonché il potenziale per gli asintomatici di rappresentare oltre il 50% tutti i casi infettivi [8], è stata raccomandata un'indagine diffusa attraverso i test come possibile strategia per controllare la malattia. Inizialmente si è pensato che una strategia che comprendesse test di asintomatici e il rafforzamento del distanziamento sociale sarebbe stata necessaria per mitigare la diffusione del virus [9].

Un piano del genere è stato applicato della Corea del Sud, dove un'ampia frazione della popolazione è stata sottoposta a screening continuo [10]. Tuttavia, i risultati presentati dal nostro e da altri gruppi hanno dimostrato chiaramente che anche una forte diminuzione della trasmissione ottenuta con strategie di regolamentazione NPI e test non avrebbe comunque impedito la diffusione globale del virus (vedi [9] e bibliografia citata), ma avrebbe avuto l'effetto di ridurre il picco di incidenza (riducendo l'impatto sulle infrastrutture mediche), estendendolo e ampliandolo. Questo effetto è dovuto al fatto che il distanziamento sociale ha lo scopo di rimuovere gli individui suscettibili dalla popolazione, ma non di proteggerli attivamente dall'eventuale infezione, per cui se rimangono piccoli focolai di infezione, la probabilità di avere una recrudescenza non è trascurabile, soprattutto se le regolamentazioni NPI vengono allentate troppo presto. Questo è proprio quello che è avvenuto nella seconda metà del 2020, quando, a seguito di un allentamento delle restrizioni, si è osservato un costante aumento del numero di casi.

Se misure restrittive come il distanziamento sociale fossero stati accompagnate da test diffusi, si sarebbe osservata una forte diminuzione dell'incidenza prevista dai modelli matematici

sviluppati in quel periodo anche se la rapida identificazione degli individui asintomatici è necessaria, ma non sufficiente per ottenere questa diminuzione. In realtà studi demografici delle comunità più colpite dalla pandemia rivelano che in assenza di aiuti economici diffusi gli individui sintomatici continuano ad essere l'elemento trainante della diffusione dell'infezione: in molte aree gli individui con un salario basso si trovavano di fronte alla scelta di isolarsi, perdendo tutto il reddito, o continuare a lavorare con il rischio reale del contagio [11].

Lockdown periodici, dispositivi di protezione individuale e distanziamento sociale per fermare la diffusione di COVID. Le misure più immediate di prevenzione della diffusione delle malattie infettive di tipo respiratorio come il COVID, molto prima della realizzazione di un vaccino o di farmaci antivirali, di solito coinvolgono la chiusura di scuole, luoghi di lavoro, chiese, uffici, fabbriche e altri luoghi pubblici, incoraggiano il mantenimento della distanza sociale, ovvero l'applicazione di una distanza minima di 2 m tra gli individui, e includono nei casi più gravi la quarantena totale e il blocco generalizzato della società, il cosiddetto *lockdown*. Queste misure mirano a diminuire il tasso di contatto effettivo della popolazione, che a sua volta riduce il numero di riproduzione della malattia, R_0 , ma presentano importanti limitazioni: è infatti impossibile imporre un *lockdown* completo per lunghi periodi, a causa del grave impatto sul tessuto sociale e sull'economia. Allo stesso tempo, la necessità di preservare categorie di lavoro essenziale mette comunque a rischio una parte consistente della popolazione a causa dei possibili e frequenti incontri tra queste categorie con individui asintomatici. La strategia di imporre *lockdown* periodici come un modo per controllare la diffusione della malattia mantenendo sano il tessuto sociale ed economico della società è stata formalmente studiata solo recentemente [12].

Laddove i *lockdown* sono stati allentati, la riattivazione di focolai di virus latenti ha aumentato la morbilità e la mortalità. Inoltre, la durata del *lockdown* è stata scelta empiricamente, monitorando il numero di nuovi casi entro un periodo simile al limite superiore del periodo di incubazione della malattia. Per prevenire l'insorgenza di nuove ondate infettive è dunque importante comprendere

in maniera quantitativa gli effetti della riduzione anticipata del *lockdown* e dell'allentamento dei controlli sulle popolazioni suscettibili.

Studi numerici preliminari suggeriscono che la chiusura periodica di 30-50 giorni di un *lockdown* rigoroso seguita da 30-50 giorni di restrizioni sociali non stringenti può aiutare a controllare il rischio di epidemia, attenuando efficacemente il picco di prevalenza del COVID-19 e riducendo i danni economici [13]. Un lavoro recentemente pubblicato introduce un approccio analitico generale per fornire una comprensione di base sul perché e sul quando tali strategie di minimizzazione del rischio sono efficaci [12]. In generale, il periodo di lockdown ottimale per il COVID-19 può dipendere in modo significativo dalla quantità di diffusione asintomatica, in particolare se c'è una grande differenza nei tassi di infezione rispetto ai casi sintomatici. Poiché la diffusione asintomatica è difficile da misurare direttamente, specialmente nelle prime fasi di un'epidemia di una malattia emergente come il COVID, può essere complicato stimare il controllo ottimale in modo sufficientemente accurato affinché il lockdown periodico sia una strategia attuabile. Una possibile soluzione è sviluppare precocemente test rapidi, efficaci, e diffusi all'interno di una popolazione, e quindi catturare la frazione di infezioni asintomatiche. Questa strategia fornirebbe previsioni affidabili sulla durata e sul picco dell'epidemia come menzionato sopra. Se i parametri epidemiologici di base per una malattia emergente sono noti, si prevede che la chiusura periodica sia efficace e produca riduzioni significative nella dimensione finale dell'epidemia se il carico epidemico non è troppo elevato.

Poiché i *lockdown* periodici incidono principalmente sulla frequenza di contatto, esiste un periodo ottimale che minimizza in modo naturale le dimensioni dell'epidemia della malattia. Le ipotesi di base importanti sono che il numero di riproduzione della malattia R_0 sia inferiore ad una soglia, che può essere calcolata analiticamente, e che il tempo di incubazione e di recupero siano simili, cioè che la malattia non sia cronica. Il presupposto di base per l'esistenza di un controllo così ottimale dell'epidemia attraverso lockdown periodici si basa sulla identificazione precoce della diffusione della malattia, ovvero su un numero di infetti ancora basso. Tale pre-

supposto di base consente di prevedere analiticamente il periodo ottimale e di fornire regioni nello spazio dei parametri che risultano in un controllo ottimale della diffusione. La logica di queste ipotesi è chiara: una malattia infettiva può essere controllata quando (1) l'intervento è attuato precocemente, (2) la malattia non è cronica o non è in grado di essere trasmessa per un periodo di tempo molto lungo e (3) il suo potenziale di diffusione è tale che le regolamentazioni NPI possono effettivamente aiutare. È stato osservato che, a seconda dei parametri, per il COVID-19 il periodo di *lockdown* deve essere compreso tra 2 e 4 volte la durata del periodo di incubazione, che è di circa 5 giorni¹⁴, quindi approssimativamente tra 10 e 20 giorni. Sebbene in generale sia stato suggerito che la chiusura periodica possa aiutare a frenare la diffusione di una malattia infettiva che si diffonde per via aerea come il COVID-19, l'applicazione di tali misure è stata, per quanto sia stato possibile appurare, principalmente basata su osservazioni del tempo di guarigione e assenza di nuovi casi per un determinato periodo. Per la pianificazione di un lockdown è necessaria una comprensione più quantitativa degli effetti di tali misure.

Oltre al distanziamento sociale e ai *lockdown*, dispositivi di protezione individuale come le mascherine per il viso sono ora necessari per la maggior parte delle attività che coinvolgono persone di nuclei familiari diversi. Le mascherine per il viso sono di gran lunga il metodo di prevenzione delle infezioni più ampiamente diffuse, disponibile ed economico, infatti esse agiscono come una barriera alla diffusione del virus sia per un individuo infetto (protezione esterna) sia per un individuo suscettibile di essere contagiato (protezione interna). Coprendo la zona del naso e della bocca, le mascherine chirurgiche per il viso forniscono un livello di filtrazione che blocca la trasmissione del virus, riducendola fino al 95%¹⁵, mentre le persone non infette che indossano una maschera chirurgica sono protette fino all'85% dalla contrazione dell'infezione [16]. L'efficacia delle mascherine nel prevenire la diffusione delle malattie è dovuta al controllo degli aerosol, la principale modalità di trasmissione del virus SARS-CoV-2 [17].

Poiché le mascherine per il viso sono state utilizzate in passato durante le epidemie di influen-

za e SARS e le pandemie di influenza, la loro utilità è stata studiata estensivamente usando modelli matematici, che hanno mostrato l'efficacia effettiva dell'utilizzo di mascherine nella prevenzione delle epidemie. D'altro canto, non sono molti gli studi che considerano la combinazione tra le varie forme di isolamento e allontanamento sociale e le mascherine o altri dispositivi di protezione individuale. Nel 2010 almeno due articoli hanno studiato l'effetto dell'uso della mascherina per contrastare la diffusione della nuova influenza A (H1N1) [18,19]. Più specificamente riguardo al COVID-19, l'efficacia di indossare la mascherina è stata recentemente studiata utilizzando un modello a equazioni differenziali ordinarie (ordinary differential equations, ODE), che considera tra gli altri parametri una percentuale variabile di individui asintomatici, il rispetto del mandato di indossare mascherine, e una diversa efficacia di protezione interna ed esterna [20].

In generale, il problema è stato studiato in diversi articoli utilizzando modelli basati su agenti, ciascuno focalizzato su una parte specifica della catena dell'infezione: in alcuni casi non è stata considerata la presenza di individui infetti asintomatici [21, 22], in altri l'efficacia di indossare maschere non è stata analizzata se combinata con altri NPI [23], infine, la differenza di protezione verso l'interno e verso l'esterno data da una mascherina è stata spesso trascurata o parametrizzata con un unico valore [24]. Hoertel et al. [25] hanno condotto un'indagine piuttosto completa basata sui dati di tutti questi effetti. Tuttavia, i loro risultati sono specificatamente riferiti alla Francia e l'elevata dimensionalità dello spazio dei parametri del modello rende difficile districare l'effetto dei vari interventi e applicare in modo diretto questo stesso modello ad altri paesi.

Una recente pubblicazione ha proposto un modello basato su agenti in cui è stata misurata l'efficacia relativa di indossare le mascherine per il viso per ridurre la diffusione di COVID-19 in presenza di individui asintomatici in combinazione con il distanziamento sociale [26]. Per sviluppare modelli che descrivono scenari di epidemia sempre più realistici sono necessarie simulazioni stocastiche di grandi dimensioni al computer e computazionalmente costose della diffusione dell'infezione. Non sorprende che il distanziamento so-

ciale e l'uso della mascherina siano sinergici nella prevenzione delle epidemie di COVID-19. In particolare, il distanziamento sociale è vantaggioso solo se accompagnato da una diffusa adesione della popolazione alla sua attuazione, al contrario, indossare le mascherine è molto efficace per ridurre la diffusione dell'infezione. Ancora una volta, questo è in parte una conseguenza dell'alto tasso di trasmissione asintomatica. Tuttavia, il distanziamento sociale rimane una misura necessaria per appiattire la curva epidemica in assenza di un vaccino.

Mobilità e *lockdown*

L'aggiunta di complessità a qualsiasi modello matematico per ottenere ulteriori informazioni deve essere bilanciata con l'incertezza associata all'aggiunta di parametri necessari per descrivere la domanda che si vuole affrontare scientificamente. Ad esempio, l'importanza di aggiungere la stratificazione per età è ben nota, ma un tale modello richiederebbe l'introduzione di parametri che descrivano la prevalenza della malattia nelle varie fasce d'età. Questo viene fatto di *routine* per malattie come il morbillo, ma diventa problematico nello studio di malattie emergenti come il COVID, in cui i dati demografici potrebbero non essere ancora disponibili. Studiando l'effetto di strategie NPI come il distanziamento sociale sulla diffusione di SARS-COV-2 è possibile suddividere esplicitamente la popolazione suscettibile al contagio tra gruppi che si mescolano e gruppi che invece sono in *lockdown* o quarantena e, quindi, non esposti al contatto con persone contagiose. Tale studio richiederebbe come *input* dei dati sulla mobilità della popolazione e sui cambiamenti nella mobilità di una popolazione per regione e nel tempo. Fortunatamente, tali dati indipendenti sono già disponibili per l'uso [27, 28, 29]. Incorporando questi dati in un modello appropriato, sarà possibile osservare una variazione dell'errore di predizione dell'incidenza, aumentando la nostra fiducia nelle previsioni.

Previsione iperlocale per l'allocazione delle risorse sanitarie

Un presupposto importante della maggior parte degli studi di sviluppo di modelli matematici basati su ODE è indicato come *homogenous mixing*. Secondo questo principio, ogni individuo in una popolazione ha una certa probabilità di incontrare tutti gli altri individui. Questa idea nasce dal campo delle reazioni chimiche, che storicamente ha fornito le basi matematiche per lo studio teorico delle malattie infettive: così come le specie chimiche in una miscela che entrano in contatto, possono reagire, e cambiare stato, in modo analogo gli individui suscettibili che entrano in contatto con gli individui infetti possono cambiare stato e essere contagiati. L'ipotesi di una popolazione in regime di *homogeneous mixing* è ovviamente un'idealizzazione e si realizza solo approssimativamente quando la densità di popolazione è molto alta.

La pandemia di COVID-19 ha portato molti sistemi sanitari sull'orlo del collasso. Le unità di terapia intensiva (ICU) e i letti ospedalieri sono stati assegnati dinamicamente per compensare i pazienti bisognosi di cure critiche durante le varie ondate di infezione. Poiché negli USA e in altre parti del mondo l'allocazione delle risorse è lasciata ai governi locali, i limiti dell'approssimazione dell'*homogeneous mixing* quando un modello matematico è usato per prevedere il numero di casi della malattia deve essere testato in un contesto realistico.

Una recente analisi ha dimostrato che se consideriamo come livello geografico una nazione o una provincia o macroregione (ad esempio, uno stato degli USA), le popolazioni non sono certamente ben mescolate. Negli USA è possibile ottenere una previsione accurata a livello di contea, sebbene questo richieda lo sviluppo di un modello iperlocale che esplicitamente tenga conto della risoluzione spaziale della densità abitativa [30]. Il modello di una regione della dimensione di uno stato americano non è in grado di prevedere con abbastanza precisione la diffusione della malattia a livello delle comunità nelle subregioni più piccole, con conseguenti grandi errori di previsione. Poiché una pianificazione efficace delle regolamentazioni NPI e di altre strategie

sociali ed economiche deve essere applicata dalle autorità responsabili su scala spaziale rilevante per le giurisdizioni locali, decisioni attuabili come la chiusura di ristoranti o l'annullamento di interventi chirurgici pianificati per aumentare la capacità di terapia intensiva devono riflettere le condizioni locali. Una pianificazione delle risorse basata su un modello di incidenza della malattia statale o nazionale è troppo approssimativo e impreciso e può portare a strategie di intervento o troppo lasche a livello locale o inutilmente rigide e gravose. Già stati degli USA come la California o paesi come l'Italia, ad esempio, hanno adottato strategie di restrizioni e blocchi regionali, che variano geograficamente in base ai valori regionali del numero riproduttivo della malattia R_0 .

Conclusioni

Sviluppare un modello di una pandemia di una malattia emergente come il COVID-19 richiede l'integrazione di più flussi di dati. Le previsioni sugli esiti della pandemia basate su modelli matematici rigorosi sono fondamentali, ma, come mostriamo in questo articolo, l'accuratezza delle previsioni dipende da diversi fattori legati a varie discipline come la biologia, l'epidemiologia, l'economia, ecc. Il potere predittivo dei modelli matematici cambia con un corpus crescente di informazioni disponibili e lo sviluppo di nuovi metodi basati sui dati con l'integrazione e l'analisi dei dati alimentati dall'intelligenza artificiale avrà sicuramente un impatto nel prossimo futuro, probabilmente prima che questa pandemia sia finita. I modelli matematici rimangono un potente strumento per comprendere una malattia e la sua diffusione nella popolazione. Le recenti notizie sull'esistenza di diverse varianti genetiche del virus SARS-CoV-2, ognuna con il potenziale di diffondersi più velocemente o di essere più mortale, rappresenta un'altra importante opportunità per i ricercatori di fornire previsioni tempestive e accurate per una migliore pianificazione delle risorse, purtroppo limitate, per combattere questa malattia.



- [1] T. Hale, A. Noam, K. Beatriz, A. Petherick, T. Phillips, S. Webster: *Variations in Government Responses to COVID-19*, Version 7.0.; 2020. <https://www.bsg.ox.ac.uk/sites/default/files/2020-09/BSG-WP-2020-032-v7.0.pdf>.
- [2] J. Graunt: *Natural and Political Observations Made Upon the Bills of Mortality*, ?, ? (1662).
- [3] D. Bernoulli: *Essai d'une nouvelle analyse de la mortalité causée par la petite vérole*, In: *Mém. Math. Phys. Acad. Roy. Sci. Paris.*, 1 (1) 1766.
- [4] R. Anderson, R. May, B. Anderson: *Infectious diseases of humans: dynamics and control*. 1992 <http://onlinelibrary.wiley.com/doi/10.1111/j.1753-6405.1992.tb00056.x/abstract> Accessed September 29, 2012.
- [5] J. V. Douglas, S. Bianco, S. Edlund, et al.: *STEM: An open source tool for disease modeling*, *Heal. Secur.*, 17 (2019) 291.
- [6] Centers for Disease Control and Prevention *COVID-19 mathematical modeling* <https://www.cdc.gov/coronavirus/2019-ncov/covid-data/mathematical-modeling.html>
- [7] M. Chinazzi, J. T. Davis, M. Ajelli, et al.; *The effect of travel restrictions on the spread of the 2019 novel coronavirus (2019-nCoV) outbreak* medRxiv. 2020;9757(March)
- [8] R. Li, S. Pei, B. Chen, et al.: *Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2)*, *Science*, 3221 (2020) 1.
- [9] I. B. Schwartz, J. Kaufman, K. Hu, S. Bianco; *Predicting the impact of asymptomatic transmission, non-pharmaceutical intervention and testing on the spread of COVID19* 2020:1-30. doi:<https://doi.org/10.1101/2020.04.16.20068387>
- [10] J. Cohen, K. Kupferschmidt: *Countries test tactics in "war" against COVID-19*, *Science*, 367 (2020) 1287.
- [11] P. Reese; *High-Poverty Neighborhoods Bear the Brunt of COVIDs Scourge*. *California Healthline*; <https://khn.org/news/article/high-poverty-protect-discretionary-neighborhoods-bear-protect-discretionary-the-brunt-of-covids-scourge/> Published 2020.
- [12] J. Hindes, S. Bianco, I.B. Schwartz *Optimal periodic closure for minimizing risk in emerging disease outbreaks* C. Poletto, ed.; PLoS One. 2021;16(1):e0244706. doi:10.1371/journal.pone.0244706
- [13] R. Chowdhury, K. Heng, M. S. R. Shawon, et al.: *Dynamic interventions to control COVID-19 pandemic: a multivariate prediction modelling study comparing 16 worldwide countries*, *Eur. J. Epidemiol.*, 35 (2020) 389.
- [14] N. M. Ferguson, D. Laydon, G. Nedjati-Gilani, et al. *Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand* 2020;(March):20. doi:10.25561/77482
- [15] L. Bourouiba; *Turbulent Gas Clouds and Respiratory Pathogen Emissions* *JAMA* 2020 doi:10.1001/jama.2020.4756
- [16] N. H. L. Leung, D.K.W. Chu, E.Y.C. Shiu et al.: *Respiratory virus shedding in exhaled breath and efficacy of face masks*, *Nat Med.*, 26 (2020) 676.
- [17] CDC. National Center for Immunization and Respiratory Diseases (NCIRD) D of VD. Scientific Brief: SARS-CoV-2 and Potential Airborne Transmission | CDC. Cdc. <https://www.cdc.gov/coronavirus/2019-ncov/more/scientific-brief-sars-cov-2.html> Published 2020.
- [18] Tracht SM, Del Valle SY, Hyman JM. Mathematical modeling of the effectiveness of facemasks in reducing the spread of novel influenza a (H1N1). *PLoS One*. 2010;5(2). doi:10.1371/journal.pone.0009018
- [19] N.C.J. Brienen, A. Timen, J. Wallinga, J. E. Van Steenbergen, P.F.M. Teunis The effect of mask use on the spread of influenza during a pandemic Risk. *Anal.* 30 20101210
- [20] C. N. Ngonghala, E. Iboi, S. Eikenberry et al.; *Mathematical assessment of the impact of non-pharmaceutical interventions on curtailing the 2019 novel coronavirus*. medRxiv. 2020. doi:10.1101/2020.04.15.20066480
- [21] N. Mahdizadeh Gharakhanlou, N. Hooshangi: *Spatio-temporal simulation of the novel coronavirus (COVID-19) outbreak using the agent-based modeling approach (case study: Urmia, Iran)* *Informatics Med Unlocked*. 2020;20. doi:10.1016/j.imu.2020.100403
- [22] S.L. Chang, N. Harding, C. Zachreson, O. M. Cliff, M. Prokopenko *Modelling transmission and control of the COVID-19 pandemic in Australia*, arXiv. 2020.
- [23] D. Kai, G. P. Goldstein, A. Morgunov, v. Nangalia, A. Rotkirch; *Universal masking is urgent in the COVID-19 pandemic: SEIR and agent based models, empirical validation, policy recommendations*. arXiv. 2020.
- [24] Silva PCL, Batista PVC, Lima HS, Alves MA, Guimarães FG, Silva RCP.: *COVID-ABS: An agent-based model of COVID-19 epidemic to simulate health and economic effects of social distancing interventions*, *Chaos, Solitons and Fractals*, 139 (2020) ?.
- [25] N. Hoertel, M. Blachier, C. Blanco et al.: *A stochastic agent-based model of the SARS-CoV-2 epidemic in France.*, *Nat Med.*, 26 (2020) 1417.
- [26] A. Catching, S. Capponi, M. Te Yeh, S. Bianco, R. Andino; *Examining the interplay between face mask usage, asymptomatic transmission, and social distancing on the spread of COVID-19* medRxiv. 2020:1-30. doi:10.1101/2020.08.12.20173047
- [27] Apple; *Mobility Trends Reports Data*. COVID?19 (2020) <https://covid19.apple.com/mobility> Published 2020.
- [28] Google LLC; *COVID-19 Community Mobility Reports 2020* Google.com. <https://www.google.com/covid19/mobility> (2020)
- [29] D. Labs, *Data for Mobility Changes in Response to COVID-19* <https://github.com/descarteslabs/DL-COVID-19>

- [30] V. Gopalakrishnan, S. Pethe, S. Kefayati et al. *Globally Local: Hyper-local Modeling for Accurate Forecast of COVID-19* medRxiv. 2020:2020.11.16.20232686. <http://medrxiv.org/content/early/2020/11/18/2020.11.16.20232686.abstract>.



Simone Bianco: è un membro dello staff di ricerca presso l'IBM Almaden Research Center, dove dirige il laboratorio di ingegneria cellulare. Ha conseguito la laurea e la laurea specialistica in Fisica presso l'Università di Pisa, in Italia, e il dottorato in Fisica presso l'Università del Nord Texas. I suoi principali interessi di ricerca riguardano la biologia evolutiva teorica, in particolare l'evoluzione dei virus a RNA, l'epidemiologia e l'ingegneria cellulare.

Sara Capponi: è postdoc presso l'IBM Almaden Research Center ed ha una formazione in fisica, biofisica computazionale e biochimica. In IBM Research, sfrutta la sua esperienza per combinare la modellazione molecolare con l'ingegneria cellulare. Inoltre, utilizza approcci teorici per caratterizzare la diffusione, la dinamica e la predazione del virus.

James H. Kaufman: è uno scienziato nel settore Artificial Intelligence e software cognitivo presso l'IBM Almaden Research Center di San Jose, CA. Attualmente coordina la ricerca sull'utilizzo di dati molecolari per migliorare la nostra comprensione delle interazioni ospite-microbo e del fenotipo di malattia. È anche co-responsabile del progetto Eclipse per il modello open source SpatioTemporal Modeler (<http://www.eclipse.org/stem>).

L'arte e la scienza di imparare dai dati: la biostatistica, dalle mutazioni genetiche all'intelligenza artificiale.

La statistica: l'unica scienza che permette a esperti diversi, usando gli stessi numeri, di trarne diverse conclusioni.

Attribuita a Evan Esar

Serena Arima

Dipartimento di Storia, Società e Studi sull'Uomo - Università del Salento

La biostatistica è comunemente definita come l'applicazione di metodologie statistiche a dati di natura biologica o medica. In realtà, tale definizione è piuttosto riduttiva in quanto essa si occupa di sviluppare nuove metodologie che tengano conto della peculiarità dei dati sperimentali e che rispondano, in modo chiaro, alle domande poste dagli scienziati. In particolare, i metodi di biostatistica si sono particolarmente sviluppati per l'analisi di dati genetici che, per dimensione e per struttura di dipendenza, non possono essere analizzati con metodologie *standard*. La biostatistica nasce quindi come

materia interdisciplinare che formalizza con un linguaggio statistico-probabilistico problemi di natura applicata e cerca di dare delle risposte che riflettano sia una significatività statistica che biologica.

Introduzione

La natura umana ci porta a riconoscere nell'osservato strutture, gruppi o schemi che ci permettano di generalizzare e formalizzare i fenomeni concreti. Tipico esempio di tale processo di apprendimento è il concetto stesso di probabilità: infatti, la probabilità è un concetto primario, una intuizione innata nell'essere umano. Nella vita di tutti i giorni accade spesso di dover prendere

delle decisioni in condizioni di incertezza e, seppur a digiuno di nozioni specifiche, ciascuno di noi pondera le proprie scelte minimizzando una funzione di rischio che rappresenta null'altro che una quantificazione della casualità di ciascun evento. Mediante tale funzione, gli individui orientano le proprie scelte.

Tale intuizione è, per sua definizione, soggettiva e pertanto non univoca. Per essere generalizzata e automatizzata, se ne definisce una formalizzazione specifica che passa attraverso assiomi, definizioni e teoremi che impongono una sorta di oggettività il cui obiettivo è convertire le osservazioni in conoscenza. La biostatistica può quindi essere definita come lo strumento di quantificazione della conoscenza che deriva da esperimenti biologici e/o medici. Essa attua una formalizzazione in termini matematico-probabilistici del problema sperimentale riportando all'utente finale una valutazione del rischio, in condizioni di incertezza, delle proprie decisioni.

Mai come in questo momento storico la valutazione del rischio e dell'incertezza legata al dato campionario svolgono un ruolo cruciale. Un esempio più che attuale è quello relativo ai *test* diagnostici. Un *test* diagnostico per una particolare condizione è detto positivo se stabilisce che la condizione è presente e negativo se sancisce che la condizione è assente. Un modo per stabilire l'accuratezza dei *test* diagnostici è calcolare la probabilità associata a due tipi di possibili errori:

- **falso positivo:** il *test* afferma che la condizione è presente, ma essa è effettivamente assente;
- **falso negativo:** il *test* sancisce che la condizione è assente, ma essa è effettivamente presente.

Il tri-*test* del sangue fornisce una stima, per una donna in gravidanza, del rischio che il bambino nasca affetto dalla sindrome genetica di Down. Questa sindrome ha una prevalenza pari a 1/800 ma gli scienziati suppongono che tale prevalenza sia più elevata se le donne hanno più di 35 anni. Uno studio prospettico [1] condotto su 5282 donne con più di 35 anni confronta l'esito del *test* con l'effettivo stato di salute del nascituro, Tabella 1. La sensibilità del *test* diagnostico, ossia la probabilità che l'individuo positivo al *test*

Sindrome di Down	Pos	Neg	Tot
Si	48	6	54
No	1307	3921	5228
	1355	3927	5282

Tabella 1: *Esiti dei test diagnostici riguardanti l'emergenza della sindrome di Down eseguiti su 5282 donne con più di 35 anni [1].*

sia realmente affetto dalla patologia in esame è definita nel seguente modo

$$P(Si|Pos) = \frac{P(Si \cap Pos)}{P(Pos)} = \frac{48/5282}{1355/5282} = 0.035$$

In sintesi, delle donne con *test* positivo, meno del 4% ha effettivamente un feto con la sindrome di Down. Perché una donna, allora, dovrebbe sottoporsi a questo *test* visto che la maggior parte dei risultati positivi sono falsi positivi? In base alla tabella vista prima, si ricava che la probabilità che una donna abbia un figlio con la sindrome di down, $P(Si) = 54/5282 = 0.0102$, sia circa pari all'1%. Dalla stessa tabella si evince che la specificità, $P(No|Neg) = 0.0015$, poco più di 1 su 1000.

Il teorema di Bayes

La Statistica si evolve e si modifica ai fini di riflettere e spiegare il pensiero dello scienziato. Tipico esempio è il teorema di Bayes, punto cardine della Statistica moderna, che formalizza il ragionamento alla base delle Scienze sperimentali quali la Biologia e la Medicina. Tale teorema, che ha dato origine ad un intero filone di pensiero statistico, la statistica Bayesiana, formalizza in termini statistico-probabilistici la conoscenza a-prioristica dello scienziato sullo studio in esame. Infatti, è inimmaginabile che lo scienziato che sta per condurre un esperimento sia completamente cieco rispetto ai potenziali risultati che potrebbe ottenere. Nell'impostazione Bayesiana, la probabilità di un evento, risultato di un esperimento aleatorio, è la congiunzione tra l'idea pre-sperimentale dello scienziato (detta a-priori) e l'esito, tramutato in dato, dell'esperimento stesso (detto verosimiglianza). È pertanto possibile definire, secondo questo approccio, la probabilità di un evento come l'aggiornamento delle ipotesi a-prioristiche con il dato sperimentale. Tale

probabilità è detta probabilità a-posteriori. Si noti, tuttavia, che l'esperimento deve essere ovviamente condotto in modo assolutamente equo rispetto alle ipotesi di base.

Si consideri il seguente esempio: supponiamo che un paziente manifesti un dolore al petto e prima di recarsi dallo specialista consulti, telefonicamente, il proprio medico di base. Quest'ultimo, riferiti i sintomi e l'anamnesi del paziente che conosce bene, immagina una serie di possibili cause di tale dolore (infarto, infreddatura, rottura di una costola ...) attribuendo a ciascuna di essere un peso, ossia una probabilità. Dopo aver visitato il paziente e aver collezionato una serie di informazioni oggettive, il medico giunge ad una diagnosi a-posteriori ottenuta ponderando le ipotesi di base e la visita strumentale del paziente.

Si noti che il teorema di Bayes è una mera applicazione della legge delle probabilità condizionate e, pertanto, la sua validità è indipendente dal tipo di approccio alla probabilità che si intende seguire. È dalla sua interpretazione, in termini di a-priori, verosimiglianza e a-posteriori che prende le mosse l'impostazione Bayesiana della probabilità.

In ambito inferenziale, le differenze tra impostazione classica e Bayesiana della probabilità emergono in modo chiaro. In particolare, supponiamo di aver a disposizione delle osservazioni campionarie x_1, \dots, x_n e vogliamo fare inferenza sui parametri della popolazione da cui tale campione è stato estratto. Se assumiamo una forma parametrica per il fenomeno in esame (ad esempio un modello teorico come quello Normale o Binomiale), $f(x; \theta)$, utilizziamo il dato campionario per stimare il parametro (o i parametri) θ della popolazione. La differenza sostanziale tra le due interpretazioni risiede nella caratterizzazione di θ : in ambito frequentista, θ è una quantità incognita ma fissa. Tale quantità può essere stimata con diversi metodi, fra cui il più famoso è il metodo della massima verosimiglianza che consiste nel massimizzare la funzione di verosimiglianza $L(\theta : x) = \prod_{i=1}^n f(x_i; \theta)$ associata ai dati (supponendo un campione indipendente e identicamente distribuito). Nell'impostazione Bayesiana, il parametro θ è incognito ma non fisso: in particolare, θ è trattato come una variabile aleatoria con una sua distribuzione di probabilità $\pi(\theta)$ che

riassume la conoscenza a-priori sul parametro incognito della popolazione. Le procedure inferenziali sono basate sulla legge a-posteriori di θ che si ottiene come

$$\pi(\theta|x) = \frac{L(\theta; x)\pi(\theta)}{\int \pi(\theta)L(\theta; x)d\theta}$$

dove l'integrale può essere sostituito da una sommatoria se si ha a che fare con leggi di probabilità discrete. La quantità al denominatore svolge il ruolo di costante di normalizzazione (ossia fa sì che $\pi(\theta|x)$ sia una legge di probabilità) ma gioca un ruolo chiave nella statistica Bayesiana nel confronto tra modelli (fattore di Bayes). Tutte le procedure inferenziali si riducono quindi all'analisi della distribuzione a posteriori $\pi(\theta|x)$: gli intervalli di credibilità (controparte Bayesiana degli intervalli di confidenza) sono intervalli della legge di probabilità $\pi(\theta)$ che contengono una proporzione di osservazioni pari al livello di confidenza prefissato e i *test* di ipotesi si conducono valutando la plausibilità delle ipotesi che si vuole confrontare rispetto alla distribuzione a posteriori. In entrambi i casi non ci si rifà al principio del campionamento ripetuto, alla base delle procedure inferenziali classiche.

La presunta soggettività del metodo insieme alla complessità computazionale che si evince dalla formula precedente hanno reso la statistica Bayesiana meno comune rispetto agli altri approcci. Oggi tutte le metodologie sono ugualmente diffuse a livello di ricerca mentre in ambito applicato e soprattutto di statistica ufficiale i metodi classici sono più privilegiati (si veda, tra gli altri, [3] per una *review*).

Recentemente, in ambito biologico e medico, molti sono i lavori di Biostatistica che utilizzano la statistica Bayesiana anche perchè, come sarà chiaro nella sezione successiva, essa risulta particolarmente utile in situazioni in cui non si può avere un numero elevato di repliche sperimentali.

Modelli complessi

Dati genetici

Teorema di Bayes e studio di sensitività e specificità sono strumenti base essenziali per l'analisi dei dati. Tuttavia, i dati provenienti da studi

sperimentali, sono complessi e difficilmente analizzabili in modo esaustivo utilizzando tecniche preconfezionate. Ecco perché la Biostatistica può essere definita come l'arte e la scienza che ci permette di apprendere dai dati. Le metodologie devono comprendere nella loro formalizzazione la complessità del dato, riconoscendo che alcun modello teorico possa riprodurre in modo fedele l'incredibile aleatorietà della scienza. Una celebre frase attribuita a George Box, uno dei padri della statistica, che in un lavoro del 1976 pubblicato su *Journal of the American Statistical Association* [2], riassume il significato della statistica moderna:

"All models are wrong".

Tale aforisma è stato esteso in "*All models are wrong but some are useful*": in esso si riconosce la fallibilità dei modelli statistici, come strumenti utilizzati per formalizzare fenomeni talmente complessi da non poter essere ingabbiati in una qualsiasi formalizzazione matematica. Il modello è sempre da considerare come uno strumento di studio di una parte, a volte piccola, di un fenomeno talmente complesso che anche comprenderne una minuscola porzione è un passo avanti per la scienza.

Questo è ciò che accade quando la statistica si applica a fenomeni complessi come l'analisi delle mutazioni genetiche. La determinazione di quali geni o proteine si modificano in parti del DNA è alla base della diagnosi e, auspicabilmente in un prossimo futuro, della terapia di malattie genetiche che, nella maggior parte dei casi, rendono la vita dell'individuo particolarmente difficile.

Grazie all'avvento di nuove tecnologie, quali microarray genomici e/o proteomici, negli ultimi anni è stata prodotta una grande mole di dati la cui analisi risulta complessa da diversi punti di vista. In primis, i dati, che sono una quantificazione del segnale genico (si veda Figura 1), vengono collezionati in matrici con un numero di righe (tipicamente i geni) estremamente elevato. Inoltre, visti i costi piuttosto elevati di replicabilità dell'esperimento e, in alcuni casi, il numero esiguo di pazienti reclutabili, il numero di repliche per ciascuna espressione è piuttosto esiguo. Si è quindi tipicamente in una situazione in cui per uno stesso gene si hanno poche repliche in diverse condizioni sperimentali (*course*

of dimensionality). In tale contesto, qualsiasi *test* statistico è poco realistico perché la variabilità dovuta al campionamento, rumore sperimentale, domina sostanzialmente il segnale. È inoltre da considerare il problema della molteplicità e della struttura di correlazione naturalmente indotta dalla natura del dato genetico.

Introduco la notazione più utilizzata: sia y_{gcr} l'espressione genica del gene g -mo, r -ma replica (pazienti) e della c -ma condizione sperimentale ($g = 1, \dots, n$; $r = 1, \dots, r_s$, $c = 1, \dots, C$). Tipicamente, il numero di geni n è dell'ordine delle centinaia di migliaia mentre le repliche disponibili per ogni gene sono meno di una decina ($\max(r_s) \leq 10$). Circa le condizioni sperimentali, nel contesto più semplice si ha $s = 1, 2$ condizioni, ossia confrontano le sequenze geniche di soggetti malati e quelle di soggetti sani. L'obiettivo è quello di valutare se esistono geni per i quali l'espressione relativa ai soggetti malati risulta alterata, significativamente più elevata (*over-expression*) o significativamente più bassa (*under-expression*), rispetto ai soggetti sani. Pur volendo ascrivere questo problema al semplice confronto tra campioni, l'esiguo numero di repliche, l'elevato numero di osservazioni e la complessa struttura di correlazione indotta tra osservazioni rende necessario l'uso di metodi alternativi. Se si lavora in un contesto di Statistica parametrica, si assume che il dato sperimentale sia ottenuto come un campione casuale da un modello teorico descritto da una legge di probabilità $f(x; \theta)$, dove θ è il vettore di parametri. Nel contesto più comunemente utilizzato, la legge di probabilità di riferimento è la legge Gaussiana i cui parametri $\theta = (\mu, \sigma^2)$ ne definiscono la media e la varianza. I dati osservati sono quindi considerati un campione casuale da tale legge di probabilità, e vengono utilizzati per stimare i parametri incogniti. In particolare, per valutare l'espressione differenziale, si assume

$$y_{gcr} \sim N(\alpha_g + 0.5(-1)^c \delta_g + \beta_{gcr}, \sigma_{gc}^2)$$

dove α_g è l'effetto medio dovuto al gene g , δ_g è l'effetto differenziale tra due condizioni sperimentali; β_{gcr} cattura l'effetto *array*, ossia la parte del segnale osservato dovuto a rumore sperimentale, quali difetti del vetrino, quantifica-

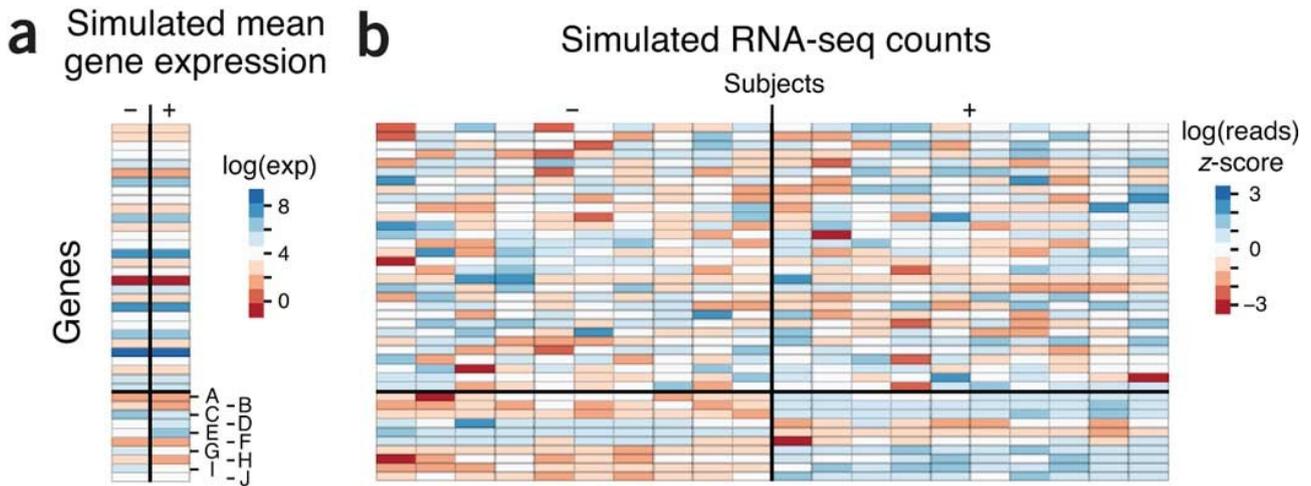


Figura 1: Un esempio di microarray genetico

zione errata del segnale di *background*, effetto dell'operatore. L'effetto *array* β_{gcr} si suppone agisca in modo **eteroschedastico**, ossia cresca in base al segnale corrispondente: è modellato, quindi, come $\beta_{gcr} = f_c(\alpha_g)$ dove f è una funzione tipicamente non parametrica, tipo *spline*, del livello di espressione stimato. Si noti che $\mu_{gc} = \alpha_g + 0.5(-1)^c \delta_g + \beta_{gcr}$ rappresenta la media della distribuzione Normale dalla quale supponiamo che i dati osservati siano stati selezionati. Per capire quindi se il gene g -mo è espresso in modo differenziale nelle due condizioni, bisogna confrontare μ_{g1} e μ_{g2} . Come chiaramente sottolineato in [4], la sovra/sottoespressione di un gene non è solo una questione di numeri, ossia dettata esclusivamente da fattori di natura statistica, ma deve tenere conto di una significatività prettamente biologica. Infatti, non è sufficiente che la differenza di espressione nelle due condizioni sia significativamente diversa per dichiarare un gene anomalo ma è anche necessario che tale gene superi un livello di espressione in ciascuna condizione che risulti significativo da un punto di vista biologico. Pertanto, un gene g è dichiarato differenzialmente espresso se

$$|\mu_{g1} - \mu_{g2}| Z \delta_{cut} \cap \alpha_g > \alpha_{cut}$$

dove δ_{cut} e α_{cut} sono soglie definite da esperti nel settore. In altre parole, α_{cut} rappresenta il livello medio minimo di espressione tale che un gene g possa essere considerato espresso, mentre δ_{cut} è la soglia tale che la differenza di espressione tra due condizioni sia biologicamente rilevante. Tipicamente δ_{cut} è posto pari a $\log(2)$ in quanto

da un punto di vista biologico un gene è differenzialmente espresso se la sua espressione in una condizione è il doppio che nell'altra. La soglia α_{cut} dipende dal tipo di *array* che si utilizza per la sperimentazione.

Questo tipo di modellistica, complicata nel corso degli anni, è un tipico esempio di come la Biostatistica coniuga la formalizzazione statistico-probabilistica con conoscenze e informazioni di natura biologica senza le quali le conclusioni, seppur corrette da un punto di vista metodologico, sarebbero vane nello specifico contesto di applicazione.

L'analisi di dati proteomici presenta una ulteriore complessità: il dato, che nei modelli precedenti viene linearizzato, quando si parla di proteine deve tenere in considerazione la correlazione tra i diversi siti e le relative distanze, nonché la conformazione tridimensionale della proteina come in Figura 2. (si veda, tra gli altri, [5]).

Filogenetica

La biostatistica trova larga applicazione in filogenetica, la scienza che si occupa di studiare il processo evolutivo degli organismi vegetali e animali dalla loro comparsa sulla Terra a oggi [6]. I metodi statistici permettono di costruire gli alberi filogenetici, diagrammi che mostrano le relazioni fondamentali di discendenza comune di gruppi tassonomici. Tali alberi vengono costruiti studiando le differenze relative tra sequenze geniche di diversi organismi e contando, mediante algoritmi complessi, tutte le possibili ricombina-

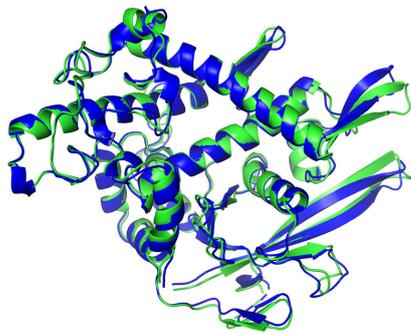


Figura 2: Un esempio come appare la struttura 3D di una proteina. Le distanze linearizzate non rispecchiano chiaramente le distanze nella conformazione tridimensionale.

zioni e relative distanze. In un albero filogenetico (si veda Figura 3), ciascun nodo rappresenta l'antenato comune più recente dei soggetti che si trovano ai nodi successivi e la lunghezza delle ramificazioni è proporzionale al tempo intercorso in termini evolutivi tra di essi.

A partire dal lavoro di Cavalli-Sforza [7] sono state proposte diverse metodologie di analisi che richiedono la definizione di processi stocastici (quali processi di Branching, processi Markoviani o semimarkoviani) che governano il meccanismo di mutazione delle sequenze geniche nel tempo. Questa tipologia di analisi è resa particolarmente complessa dalla grande mole di dati che coinvolge la cui analisi prevede l'utilizzo di algoritmi computazionali complessi anche da un punto di vista informatico. Algoritmi di percorso minimo o di massima parsimonia, come ad esempio l'algoritmo di Fitch [8] o di Sankoff [9], sono ancora attualmente oggetto di studio in diverse discipline per ottimizzarne la complessità matematica che computazionale.

Statistica e intelligenza artificiale

Il **Machine Learning** (ML), anche detto apprendimento automatico, è una delle applicazioni dell'intelligenza artificiale. Il ML è un insieme di algoritmi matematico/statistici che permettono ad un computer di apprendere dai dati e prendere decisioni senza ricevere istruzioni dirette. Il sistema informatico emula la logica del ragionamento umano e, mediante l'inserimento di nuovi dati, prende decisioni ottimizzando funzioni di rischio. La distinzione tra Statistica e

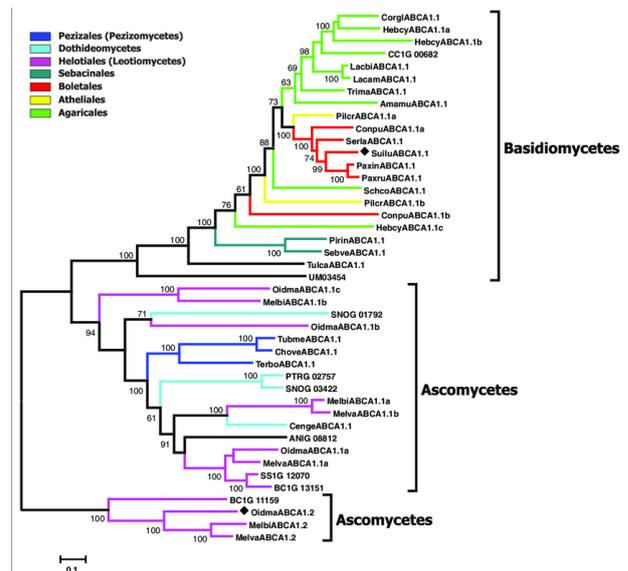


Figura 3: Un esempio di albero filogenetico.

ML è piuttosto oscura, ammesso che questa distinzione realmente esista. Nella letteratura, tale differenza è ascritta principalmente all'obiettivo delle due discipline. I metodi di ML sono basati su modelli ottimizzati dal punto di vista predittivo. I modelli statistici sono disegnati principalmente a fini esplicativi, ossia per valutare da un punto di vista inferenziale la significatività della relazione tra variabili. Sebbene tale definizione sia piuttosto condivisa, essa non è né univoca né esclusiva. Un recente articolo su Nature [10] evidenzia similarità e differenze tra Statistica e ML (così come *data science*) evidenziando che in realtà le differenze sono semplicemente di natura pratica e non concettuale. Il ML non esisterebbe senza la statistica, ma il ML è particolarmente utile in un'epoca come la nostra in cui la quantità di dati a nostra disposizione supera di gran lunga il concetto di campionamento tipico dei modelli statistici, nelle sue molteplici declinazioni.

Il ML ha trovato una grande applicazione in ambito medico, in particolare nelle Neuroscienze per lo studio degli stimoli cerebrali e le reazioni dei neuroni (si vedano recenti studi sull'Alzheimer fra cui [11]). Stesso dicasi per Data Science basato su metodi computazionali per l'analisi dei dati: tali metodi includono analisi esplorative che aiutano il *data scientist* a entrare nel dettaglio dei dati e del loro meccanismo generatore, a pulire e pre-processare i dati, intuire le prime relazioni tra variabili. Questo tipo di operazioni nell'epoca dei *big-data* richiede l'utilizzo di stru-

menti informatici specifici e più potenti dal punto di vista dell'esecuzione. Capire a quali aree della scienza attribuire ciascuna di queste definizioni è particolarmente complesso forse perché la scienza è per sua natura interdisciplinare.

Conclusioni

Molti ricercatori ritengono che la Biostatistica sia meramente un insieme di procedure da fare per soddisfare gli editori o *referee* di una rivista o le autorità regolatorie o le agenzie di *funding*. Ma la necessità della Biostatistica è chiara quando si vuole considerare la relazione tra l'intuizione del ricercatore e un metodo formale.

Concludendo, l'interdisciplinarietà e il ruolo della Biostatistica possono essere riassunti dall'affermazione di Laplace che ha osservato che la teoria della probabilità è [12]

"... at bottom only common sense reduced to calculus; it makes us appreciate with exactitude that which exact minds feel by a sort of instinct without being able oftentimes to give a reason for it. It leaves no arbitrariness in the choice of opinions and sides to be taken; and by its use can always be determined the most advantageous choice"

In questo lavoro si è voluto evidenziare come le metodologie statistiche nascono da problemi di natura concreta. Esse si sono sviluppate e continuano ad adattarsi ed evolversi grazie ai problemi che sorgono dai dati reali. Le evoluzioni in ambito biologico e medico sono il volano per le evoluzioni metodologiche di analisi dei dati: il biologo/medico spiega il problema in esame eviscerandone le problematiche inerenti, lo statistico lo formalizza in un linguaggio statistico-probabilistico e, insieme, discutono i risultati valutandone la significatività a tutto tondo.



[1] J. Haddow et al.: *Reducing the Need for Amniocentesis in Women 35 Years of Age or Older with Serum Markers for Screening*, New England Journal of Medicine, 330 (1994) 1114.

[2] G. E. P. Box: *Science and Statistics*, Journal of American Statistical Association, 71 (1976) 791.

- [3] C. P. Robert: *The Bayesian choice*, Springer, Berlino (2006).
- [4] A. Lewin, S. Richardson, C. Marshall, A. Glazier, T. Aitman: *Bayesian modeling of differential gene expression*, Biometrics, 62 (2006) 1.
- [5] P. Baldi, G.W. Hatfield: *DNA Microarrays and Gene Expression*, Cambridge University Press, Cambridge (UK) (2011).
- [6] Rita Levi Montalcini: *La Galassia Mente*, Baldini & Castoldi, Milano (2001).
- [7] L. L. Cavalli-Sforza, A.W.F. Edwards: *Phylogenetic analysis. Models and estimation procedures*, Am. J. Hum. Genet., 19 (1967) 233.
- [8] W. M. Fitch: *The molecular evolution of Cytochrome c in Eukaryotes*, Journal of Molecular Evolution, 8 (1976) 13.
- [9] D. Sankoff, R.J. Cedergren, G. Lapalme: *Frequency of insertion-deletion, transversion and transition in the evolution of 5S Ribosomal RNA*, Journal of Molecular Evolution, 7 (1976) 133.
- [10] D. Bzdok, N. Altman, M. Krzywinski: *Statistics versus machine learning*, Nature, 15 (2018) 233-234.
- [11] T. Jo, K. Nho e A.J. Sayking: *Deep Learning in Alzheimer's Disease: Diagnostic Classification and Prognostic Prediction Using Neuroimaging Data*, Frontier Aging Neuroscience, 20 (2019) .<https://doi.org/10.3389/fnagi.2019.00220>
- [12] P.S. Laplace: *A Philosophical Essay on Probabilities*, Dover, New York (1951).



Serena Arima: è professore associato di Statistica per la ricerca sperimentale e tecnologica presso l'Università del Salento. Si occupa di statistica bayesiana e di metodi computazionali per la stima di modelli complessi.

Meccanica statistica per reti complesse

John Von Neumann [...] mi disse che avrei dovuto chiamarla 'entropia', per due motivi: "Innanzitutto, la tua funzione d'incertezza è già nota nella meccanica statistica con quel nome. In secondo luogo, e più significativamente, nessuno sa cosa sia con certezza l'entropia, così in una discussione sarai sempre in vantaggio".

— Claude Shannon

Tiziano Squartini

Scuola IMT Alti Studi Lucca (Lucca, Italy)

Institute for Advanced Study (IAS), University of Amsterdam (Amsterdam, The Netherlands)

Una rete non è altro che un insieme di nodi collegati da link. Questa definizione, apparentemente banale, cattura l'essenza di una miriade di fenomeni di natura diversa, quali il commercio globale, i prestiti interbancari, le interazioni proteiche e quelle tra gli utenti di Twitter. Una volta riconosciuta la natura complessa di tali sistemi, si pone il problema di quale sia il miglior *framework* concettuale per analizzarli. Tale inquadramento è fornito dalla Meccanica Statistica, branca della Fisica nata per studiare il comportamento dei gas perfetti ed oggi impiegata per modellizzare grafi reali.

Come nasce un grafo?

Si racconta che, nel 1736, i cittadini di Königsberg (al tempo, in Prussia Orientale; oggi in Russia, dopo aver cambiato il nome in Kaliningrad) fossero ossessionati da un problema: "è possibile, con una passeggiata, seguire un percorso che attraversi ogni ponte della città

una ed una volta soltanto?" La soluzione a tale quesito, noto come problema dei sette ponti di Königsberg, segna l'inizio della cosiddetta teoria dei grafi: Eulero affrontò il problema come se fosse un quesito di un ramo della Geometria noto, al tempo, col nome di Geometria situs (oggi, Topologia), disegnando il primo grafo della storia e dimostrando che la configurazione della città non consentiva l'esistenza di una tale passeggiata [1].

Eulero considerò il problema nella sua generalità, dividendo le zone della città in due classi, quelle toccate da un numero pari di ponti (zone pari) e quelle toccate da un numero dispari di ponti (zone dispari); dopodiché, considerò che le zone pari dovevano essere zone di passaggio (visto che ad ogni ponte in entrata sarebbe corrisposto un ponte in uscita) e che le zone dispari dovevano essere zone di partenza oppure di arrivo (una volta accoppiati i ponti afferenti ad esse, ne sarebbe sempre avanzato uno: quello tramite cui arrivare o da cui uscire). Il terzo passo del ragionamento fu quello di contare il numero di zone dispari: visto che queste sono sempre presenti

in numero pari (la piena giustificazione di questa asserzione, nota come *handshaking lemma*, richiederebbe qualche tecnicismo), ne possiamo avere 0, 2, 4...

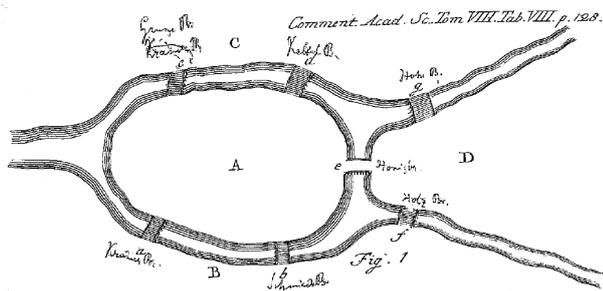


Figura 1: Pianta della città di Königsberg nel 1736. Essendo divisa in quattro zone in ciascuna delle quali afferisce un numero di ponti dispari, una passeggiata che attraversi tutti i ponti una sola volta non esiste [1].

Se sono presenti 0 zone dispari, allora la passeggiata è possibile, partendo da qualsiasi nodo e terminando su di esso; se sono presenti 2 zone dispari, allora la passeggiata è possibile, partendo da una delle due zone dispari e terminando sull'altra; se sono presenti più di due zone dispari, allora la passeggiata è impossibile.

La pianta della città di Königsberg nel 1736 è mostrata in figura 1: come si vede, vi sono sette ponti e quattro zone (indicate dalle lettere A, B, C, D), tutte dispari: quindi, la passeggiata cercata non esiste. Il problema dei ponti di Königsberg era stato risolto ed il primo grafo della storia era nato.

In principio fu Erdős

La Teoria dei grafi, dunque, nacque con Eulero. Tale branca della matematica ha continuato a svilupparsi indipendentemente da altri campi fino al 1959, anno in cui Erdős ne affrontò lo studio combinandola con un'altra branca, apparentemente molto distante da essa: la Teoria della probabilità. Se con Eulero era nato il primo grafo, con Erdős nacque il primo grafo probabilistico.

In poche parole, Erdős studiò le proprietà di grafi i cui nodi stabilivano connessioni con una probabilità, p , fissata¹. Nonostante tali oggetti

¹Per essere esatti, furono presentati contemporaneamente due modelli diversi, rispettivamente da Erdős e Rényi

mostrassero un comportamento molto ricco, la richiesta che il valore di p fosse lo stesso per tutte le coppie di nodi costituiva la principale limitazione di quello che potremmo definire il primo modello probabilistico in teoria dei grafi.

Di ciò ci si rese conto (solo) nel 1999, anno in cui il World Wide Web (WWW, nel seguito) venne descritto come un grafo [4]: non appena si confrontò la struttura di tale sistema con quella predetta dal modello di Erdős, ci si rese conto che quest'ultimo falliva completamente nel predire anche le più semplici proprietà del WWW: una su tutte, la distribuzione di grado, ovvero la distribuzione del numero di vicini di ciascun nodo. Mentre il modello di Erdős prevedeva una curva a campana, caratterizzata da un valor medio ben definito ed una varianza (ovvero la scala tipica di variazione della variabile in esame) altrettanto ben definita, il WWW mostrava una distribuzione di grado a legge di potenza, con varianza non definita. Le implicazioni di tale risultato sono immediatamente chiare: poiché non esiste una scala di variazione tipica, il grado dei nodi può estendersi per intere decadi, da valori prossimi all'unità a valori prossimi al milione; in termini meno matematici, questo significa che esistono pagine web con pochissimi *link* (la stragrande maggioranza) accanto a veri e propri *hub*, ovvero (poche) pagine da milioni di connessioni.

La scoperta di leggi di potenza ha costituito il primo segno tangibile dell'esistenza sistemi che si discostavano dalle previsioni dell'unico modello disponibile all'epoca (il modello di Erdős, appunto) facendo guadagnare al WWW, ed a sistemi simili, il nome di reti complesse.

L'entropia di Shannon

La teoria delle reti, quindi, riprende il concetto matematico di grafo e lo applica per descrivere sistemi reali. Oggetti complessi, tuttavia, richiedono modelli meno semplicistici di quello di Erdős: negli anni, quindi, si è cercato un *framework* capace di produrre modelli probabilistici più raffinati di quello proposto attorno alla metà del secolo scorso. Tale *framework* viene fornito dalla Meccanica statistica, nell'interpretazione che

[2] e da Gilbert [3]: il modello a cui ci riferiremo in tutto il contributo è quello dovuto a Gilbert ma continueremo a chiamarlo col nome di modello di Erdős.

ne dette Jaynes: ovvero, un modo nuovo di fare inferenza statistica, a partire da pochi principi primi [5]. In effetti, uno soltanto: il principio di massima entropia.

Il principio di massima entropia altro non è che la versione 2.0 del buon vecchio principio di ragione insufficiente (o principio di indifferenza) introdotto da Bernoulli e studiato, poi, da Laplace: in sostanza, è il più semplice principio a cui appellarsi per fare inferenza, recitando "In mancanza di informazione in base alla quale preferire qualcuna delle ipotesi alternative, occorre assegnare a tutte la stessa probabilità". Detto altrimenti, il principio di ragione insufficiente sancisce che, in mancanza di informazione ulteriore, la distribuzione migliore per descrivere un *set* di ipotesi mutuamente esclusive (ciascuna delle quali, cioè, esclude le altre) ed esaustive (quelle considerate, cioè, sono tutte e sole le ipotesi plausibili) è la distribuzione uniforme.

E se, invece, si possiede un po' d'informazione sul sistema? Come gestirla al meglio? La risposta viene fornita dal principio di massima entropia. In pratica, si considera la quantità

$$S = - \sum_{\mathbf{G}} P(\mathbf{G}) \ln P(\mathbf{G}) \quad (1)$$

e la si massimizza sotto certi vincoli che quantificano l'informazione che si ha sul sistema in esame. La somma corre sul supporto della distribuzione di probabilità P che regola il manifestarsi degli eventi appartenenti al proprio supporto e di cui \mathbf{G} indica un generico elemento - grafo, nel nostro caso (si veda anche il box "Quale ensemble?" per una spiegazione più approfondita). L'entropia è una misura della nostra ignoranza sul sistema: massimizzarla significa ammettere di essere massimamente incerti sul sistema - eccezion fatta per le informazioni che si possiedono sullo stesso, appunto.

La quantità S si chiama entropia di Shannon [7] ed il suo significato appare più chiaro alla luce del concetto di autoinformazione di un evento \mathbf{G} , definita come

$$I(\mathbf{G}) = - \ln P(\mathbf{G}); \quad (2)$$

tale quantità ci dice che l'informazione contenuta in un evento \mathbf{G} è funzione della probabilità di

verificarsi dell'evento stesso, secondo una legge logaritmica [6]. Tale definizione implica che il realizzarsi di un evento impossibile (ovvero tale che la probabilità di realizzarsi è $P(\mathbf{G}) = 0$) porterebbe una quantità di informazione infinita, mentre l'accadere di un evento certo (ovvero tale che la probabilità di realizzarsi è $P(\mathbf{G}) = 1$) non aggiunge niente a quanto già si sa. Il concetto di autoinformazione è stato descritto anche come sorpresa, termine che ci aiuta a capire meglio i due esempi appena fatti: mentre saremmo infinitamente sorpresi di vedere realizzato un evento impossibile, non siamo sorpresi affatto nel vedere realizzato un evento certo!

Mediando l'autoinformazione su tutti gli eventi che possono accadere, si ottiene

$$S = \sum_{\mathbf{G}} P(\mathbf{G}) I(\mathbf{G}), \quad (3)$$

precisamente l'entropia di Shannon. Da notare che nell'articolo originale di Shannon, l'esempio con cui viene introdotta S riguarda una sorgente probabilistica di simboli: ecco perché, spesso, S viene descritta come entropia di sorgente.

Perché proprio Shannon?

In termini vaghi ma intuitivi, si può dire che l'entropia di Shannon rappresenta una buona misura dell'incertezza contenuta in una distribuzione di probabilità. Per essere più precisi, occorre procedere in modo assiomatico e richiedere alcune proprietà che una buona misura di incertezza dovrebbe avere. A tale proposito, l'articolo di Jaynes del 1957 è illuminante [6]. Definendo

$$S(p_1, p_2 \dots p_n) \quad (4)$$

come l'entropia della distribuzione di probabilità il cui supporto è il *set* di eventi $1, 2 \dots n$, si richiede che essa

- sia continua nelle $p_i, \forall i$;
- sia una funzione monotona di n , nel caso in cui $p_i = \frac{1}{n}, \forall i$;
- soddisfi la cosiddetta legge di

composizione, ovvero

$$S(p_1, p_2 \dots p_n) = S(q_1, q_2 \dots q_m) + \sum_{i=1}^m q_i S\left(\frac{p_1}{q_i}, \frac{p_2}{q_i} \dots \frac{p_n}{q_i}\right). \quad (5)$$

Il significato della prima legge è abbastanza evidente: piccoli cambiamenti nelle probabilità dei singoli eventi non possono condurre a variazioni troppo grandi dell'informatività degli stessi. La seconda legge, invece, completa il principio di ragione insufficiente, stabilendo che, in caso di equiprobabilità, l'incertezza legata ad insiemi di eventi via via più numerosi deve crescere col numero di eventi. La terza legge, infine, stabilisce che l'informatività di un *set* di eventi non deve dipendere da come tali eventi siano suddivisi. L'entropia di Shannon è l'unico funzionale che soddisfa le tre leggi di cui sopra.

Naturalmente, sono possibili variazioni nel numero e nel tipo di leggi richieste. Ad esempio, spesso si incontra la richiesta che un evento di probabilità nulla non debba influenzare l'informatività del set: se l'evento k ha probabilità 0, allora

$$S(\dots p_{k-1}, p_k, p_{k+1} \dots) = S(\dots p_{k-1}, p_{k+1} \dots). \quad (6)$$

Un'altra richiesta che si incontra di frequente è la cosiddetta additività: dati due sottosistemi indipendenti, si deve avere che

$$S(1 \cup 2) = S_1 + S_2 \quad (7)$$

ovvero che l'entropia della loro unione si possa scrivere come somma di entropie. Tale richiesta conduce alla forma logaritmica immediatamente, visto che l'entropia dell'unione di due sistemi indipendenti altro non è che $S(W_1 \cdot W_2)$ (ovvero l'entropia del numero totale di configurazioni, ottenibile moltiplicando il numero di configurazioni in cui si può trovare il sottosistema 1, W_1 , per il numero di configurazioni in cui si può trovare il sottosistema 2, W_2) mentre l'entropia della somma si scrive come $S(W_1) + S(W_2)$: la funzione che trasforma un prodotto in una somma, garantendo che $S(W_1 \cdot W_2) = S(W_1) + S(W_2)$, è proprio il logaritmo. Da notare che la richiesta di additività viene naturalmente inglobata

Quale ensemble?

S è definita da una somma su \mathbf{G} . Ma a quale insieme appartiene \mathbf{G} ? Si indica con \mathcal{G} ed è l'insieme dei grafi con un numero di nodi fissato (diciamo N) ed un numero di connessioni variabile, da 0 al numero massimo di coppie di nodi collegabili con un link. Nel caso di reti indirette, ad esempio, tale numero coincide con $N(N-1)/2$ (ovvero, il numero di combinazioni di N oggetti presi a due a due). Tale insieme gioca il ruolo chiave di supporto della distribuzione canonica.

Da notare che in teoria delle reti la distinzione tra ensemble canonico ed ensemble gran canonico è meno marcata che in meccanica statistica: di fatto, si usa la dicitura ensemble canonico per sottolineare che si stanno considerando dei vincoli che devono essere soddisfatti in media; la dicitura ensemble microcanonico, invece, è intesa indicare l'insieme di configurazioni che soddisfano un certo insieme di vincoli esattamente.

dalla legge di composizione, una volta che il set iniziale di eventi sia suddiviso opportunamente (per questo, però, si rimanda all'articolo di Jaynes [6]).

Inferenza statistica: un paio di casi semplici...

È il momento di fare esempi concreti di inferenza. Immaginiamo di non imporre altro vincolo che quello di normalizzazione, ovvero di non richiedere altro se non che la distribuzione di probabilità sia normalizzata (sul nostro ensemble canonico). È possibile imporre tale richiesta massimizzando la funzione

$$\mathcal{L} = S - \psi \left[\sum_{\mathbf{G}} P(\mathbf{G}) - 1 \right] \quad (8)$$

chiamata *Lagrangiana* del problema, rispetto a $P(\mathbf{G})$. La soluzione è

$$P(\mathbf{G}) = \frac{1}{|\mathcal{G}|}, \quad \forall \mathbf{G} \quad (9)$$

ovvero la distribuzione uniforme (sull'ensemble canonico): in pratica, il principio di massima entropia consente di ritrovare il principio di ragione insufficiente che non ne costituisce altro che un caso particolare. Naturalmente, ψ non è altro che il moltiplicatore di Lagrange corrispondente al vincolo desiderato.

Il passo successivo è quello di imporre un vincolo ulteriore rispetto a quello (banale) rappresentato dalla normalizzazione della probabilità. Per procedere, occorre decidere che tipo di reti considerare: in questo caso, consideriamo reti binarie, indirette. Il più semplice vincolo da imporre è rappresentato dal numero totale di *link*, L , e la Lagrangiana diventa

$$\mathcal{L} = S - \psi \left[\sum_{\mathbf{G}} P(\mathbf{G}) - 1 \right] - \theta \left[\sum_{\mathbf{G}} P(\mathbf{G})L(\mathbf{G}) - \langle L \rangle \right] \quad (10)$$

la cui massimizzazione rispetto a $P(\mathbf{G})$ porta all'espressione

$$P(\mathbf{G}) = \frac{e^{-\theta L(\mathbf{G})}}{Z}, \quad \forall \mathbf{G}. \quad (11)$$

(di nuovo, ψ e θ altro non sono che i moltiplicatori di Lagrange corrispondenti ai vincoli considerati). È istruttivo calcolare esplicitamente la funzione di partizione

$$Z = \sum_{\mathbf{G}} e^{-\theta L(\mathbf{G})}, \quad (12)$$

espressione che si può risolvere analiticamente. Considerando che il numero totale di link si può riscrivere come $L = \sum_{i < j} a_{ij}$, ovvero come somma esplicita degli elementi della matrice di adiacenza che rappresenta un generico grafo, si ottiene

$$\begin{aligned} Z &= \sum_{\mathbf{G}} e^{-\theta L(\mathbf{G})} = \sum_{\mathbf{G}} e^{-\theta \sum_{i < j} a_{ij}} \\ &= \sum_{\mathbf{G}} \prod_{i < j} e^{-\theta a_{ij}} = \prod_{i < j} \sum_{a_{ij}=0}^1 e^{-\theta a_{ij}} \\ &= (1 + e^{-\theta})^{\frac{N(N-1)}{2}}. \end{aligned} \quad (13)$$

Come si rappresenta un grafo?

Un grafo si può rappresentare mediante una matrice che assume il nome di matrice di adiacenza. Nel caso più semplice di un grafo binario, indiretto con N nodi, tale matrice è di ordine N , si indica con il simbolo \mathbf{A} e la generica *entry* a_{ij} è 1 se i nodi i e j sono collegati da un *link* e 0 altrimenti. In tal caso, la matrice di adiacenza è anche simmetrica. Mediante la matrice di adiacenza, tutte le quantità di interesse possono essere calcolate compattamente: un esempio è fornito dal numero totale di *link*, L , che può essere calcolato come $L = \sum_{i < j} a_{ij}$, ovvero come somma su (la metà de) le entries della matrice di adiacenza; un altro esempio è fornito dal grado di un nodo, che viene definito come $k_i = \sum_j a_{ij}$ (e che conta il numero di vicini del nodo stesso).

Inserendo l'espressione di Z dentro l'espressione per $P(\mathbf{G})$ si trova il risultato

$$P(\mathbf{G}) = p^{L(\mathbf{G})} (1-p)^{\frac{N(N-1)}{2} - L(\mathbf{G})}, \quad \forall \mathbf{G} \quad (14)$$

dove p è stato definito come $p \equiv e^{-\theta} / (1 + e^{-\theta})$, espressione che indica proprio il modello probabilistico di Erdős!

Calcoliamo esplicitamente la distribuzione di grado predetta dal modello di Erdős. La probabilità che un certo numero di nodi, n , possieda grado $k_i \equiv k$ altro non è che la frazione attesa di nodi con tale grado. Quindi, si può scrivere

$$f(k_i = k) = \frac{n_k}{N} = \frac{\sum_{i=1}^N I(k_i = k)}{N} \quad (15)$$

(facendo uso della variabile indicatore I che assume il valore 1 se il grado del nodo i -esimo è pari a k e 0 altrimenti) e

$$\begin{aligned} \langle f(k_i = k) \rangle &= \frac{\langle n_k \rangle}{N} = \frac{\sum_{i=1}^N \langle I(k_i = k) \rangle}{N} \\ &= p(k_i = k) \end{aligned} \quad (16)$$

dove $p(k_i = k)$ altro non che è la probabilità che il grado k_i sia uguale a k . Ricordando che ciascun

link ha la stessa probabilità, p , di comparire, si ottiene

$$p(k_i = k) = \binom{N-1}{k} p^k (1-p)^{N-1-k} \quad (17)$$

espressione che ci dice che il singolo grado è distribuito binomialmente - quindi, secondo una distribuzione con un valore medio ben definito e pari a $\langle k_i \rangle = p(N-1)$. Averne trovato l'espressione esplicita ci consente di capire perché tale modello è inadeguato a spiegare le reti reali: imporre che una qualsiasi coppia di nodi stabilisca una connessione con la stessa probabilità, p , non consente di rendere conto dell'eterogeneità, inevitabilmente presente, dei nodi che costituiscono la rete.

...ed il caso generale

Abbiamo visto che il principio di massima entropia consente di ritrovare due casi semplici: la distribuzione uniforme predetta dal principio di ragione insufficiente ed il modello di Erdős. Chiediamoci, ora, se sia anche in grado di generalizzare quest'ultimo, producendo, quindi, modelli più raffinati. La risposta è affermativa e la ricetta è la seguente: per produrre modelli più raffinati occorre imporre più vincoli, meno banali. In termini del tutti generali si può considerare il caso seguente

$$\begin{aligned} \mathcal{L} = & S - \psi \left[\sum_{\mathbf{G}} P(\mathbf{G}) - 1 \right] \\ & - \sum_i \theta_i \left[\sum_{\mathbf{G}} P(\mathbf{G}) X_i(\mathbf{G}) - \langle X_i \rangle \right] \end{aligned} \quad (18)$$

dove si è vincolata tutta una serie di grandezze, genericamente indicate con X_i . In termini puramente formali, massimizzare la Lagrangiana di cui sopra conduce all'espressione

$$P(\mathbf{G}) = \frac{e^{-\sum_i \theta_i X_i(\mathbf{G})}}{Z} \equiv \frac{e^{-H(\mathbf{G})}}{Z}, \quad \forall \mathbf{G} \quad (19)$$

dove, adesso, $Z = \sum_{\mathbf{G}} e^{-H(\mathbf{G})}$. La formula sopra è quella che definisce i cosiddetti Exponential Random Graphs, ovvero grafi la cui probabilità assume una forma funzionale esponenziale. Tale formalismo è il cuore della Meccanica statistica delle reti e consente di definire una vasta gamma

di modelli diversi. Da notare che la quantità che riassume i vincoli imposti, $H(\mathbf{G})$, assume il nome evocativo di hamiltoniana del grafo: mentre in Meccanica statistica tradizionale, coinciderebbe con l'energia (ed eventualmente il numero di particelle) di un gas, in teoria delle reti, l'hamiltoniana elenca i vincoli topologici che si desiderano imporre [8, 9].

Un passo oltre Erdős

A questo punto ci potremmo chiedere quale set di vincoli definisca il modello più semplice, ma meno banale, dopo quello di Erdős. Tale set di vincoli coincide con la sequenza di grado e definisce il cosiddetto Configuration Model. In termini formali, si trova che

$$\begin{aligned} Z &= \sum_{\mathbf{G}} e^{-\sum_{i=1}^N \theta_i k_i(\mathbf{G})} = \sum_{\mathbf{G}} e^{-\sum_{i<j} (\theta_i + \theta_j) a_{ij}} \\ &= \sum_{\mathbf{G}} \prod_{i<j} e^{-(\theta_i + \theta_j) a_{ij}} = \prod_{i<j} \sum_{a_{ij}=0}^1 e^{-(\theta_i + \theta_j) a_{ij}} \\ &= \prod_{i<j} (1 + e^{-(\theta_i + \theta_j)}) \end{aligned} \quad (20)$$

e che la probabilità di un generico grafo \mathbf{G} assume la forma

$$P(\mathbf{G}) = \prod_{i<j} p_{ij}^{a_{ij}} (1 - p_{ij})^{1-a_{ij}}, \quad \forall \mathbf{G} \quad (21)$$

dove si è definito

$$p_{ij} \equiv \frac{e^{-(\theta_i + \theta_j)}}{1 + e^{-(\theta_i + \theta_j)}}, \quad \forall i < j .$$

È interessante notare come il generico coefficiente di probabilità predetto dal Configuration Model, riscritto come

$$p_{ij} \equiv \frac{x_i x_j}{1 + x_i x_j}, \quad \forall i < j$$

(avendo definito $x_i \equiv e^{-\theta_i}$), contenga quantità specifiche di nodo. Tali quantità sono state chiamate *fitness* e riassumono una qualche proprietà intrinseca del vertice che si suppone guidi la sua tendenza a stabilire connessioni.

Anticipando qualche applicazione che verrà discussa in seguito, si può, qui, citare il caso della rete del commercio globale: come alcuni lavori hanno messo in luce [10] i paesi del mondo sta-

biliscono relazioni commerciali con una probabilità che dipende dal loro prodotto interno lordo (o GDP, dall'inglese *Gross Domestic Product*), secondo la forma funzionale

$$p_{ij} = \frac{z \text{GDP}_i \text{GDP}_j}{1 + z \text{GDP}_i \text{GDP}_j}, \quad \forall i < j; \quad (22)$$

ignorando, per il momento, z (che gioca il ruolo di parametro libero), tale formula stabilisce che paesi economicamente più massivi stabiliscono più connessioni. Una considerazione analoga vale per le banche, sostituendo, al posto di GDP_i , il capitale della banca i -esima (se si vuole, la sua massa finanziaria, continuando l'analogia con la teoria della gravitazione newtoniana): in pratica, le banche più ricche sono anche quelle più connesse.

La potenza del Configuration Model risiede proprio in questo: permettere che coppie di nodi diverse stabiliscano connessioni con probabilità diverse, in modo da tenere conto dell'eterogeneità dei nodi della rete - a differenza del modello di Erdős.

Il secondo pezzo del puzzle: verosimiglianza

La massimizzazione dell'entropia fornisce una ricetta per ottenere la forma funzionale di una distribuzione di probabilità. Tuttavia, tale distribuzione è funzione di alcuni parametri: i moltiplicatori di Lagrange che vincolano l'entropia stessa a soddisfare le quantità che definiscono ciascun particolare modello. Poiché una procedura di inferenza statistica non è tale se non definisce un modo per stimare numericamente i parametri di una distribuzione, è arrivato il momento di introdurre il secondo pezzo del puzzle: il principio di massima verosimiglianza [9].

Poiché tale principio prescrive di massimizzare la probabilità che un evento si manifesti (rispetto ai parametri che entrano nella definizione della probabilità stessa), risponde anche ad una seconda esigenza: legare il formalismo degli Exponential Random Graphs all'analisi di reti reali. Immaginiamo, infatti, di voler analizzare la rete del commercio globale in un particolare anno e chiamiamola \mathbf{G}^* : come stimare i parame-

Reti fermioniche e reti bosoniche

Non sarà sfuggito che il calcolo della funzione di partizione richiede che siano definiti i numeri di occupazione di ogni singola coppia di nodi in un grafo. Nel caso binario (indiretto o diretto), tali numeri possono essere solo $a_{ij} = 0, 1$: questo implica che al più un link possa collegare la generica coppia $i - j$. Naturalmente, è possibile generalizzare tale schema: nel caso di reti pesate, di solito si lascia che i numeri di occupazione varino fino all'infinito, ovvero $w_{ij} = 0, 1 \dots + \infty$. Mentre nel primo caso, la regola di occupazione binaria conduce ad un valore atteso del numero di *link* per stato, pari a $\langle a_{ij} \rangle = p_{ij} = [1 + e^{(\theta_i + \theta_j)}]^{-1}$, nel secondo caso si trova che $\langle w_{ij} \rangle = [1 - e^{-(\theta_i + \theta_j)}]^{-1}$. Nel primo caso, l'essersi limitati a considerare reti binarie conduce alla statistica di Fermi-Dirac; nel secondo caso, invece, non aver posto alcun limite al numero di occupazione dei link per stato, conduce alla statistica di Bose-Einstein.

tri del modello che vogliamo testare? Beh, prima occorre immaginare che tale rete sia un evento possibile del modello e, poi, occorre massimizzare la probabilità che tale rete si manifesti sotto tale modello: in termini matematici

$$\max_{\vec{\theta}} \{ \ln P(\mathbf{G}^* | \vec{\theta}) \}. \quad (23)$$

Proviamo a testare esplicitamente tale prescrizione in un paio di casi specifici. Riprendiamo Erdős e immaginiamo che la rete del commercio globale (il *World Trade Web*, in inglese) provenga da esso. Sostituendo eq. (11) in eq. (23), troviamo

$$\max_{\theta} \{ -\theta L(\mathbf{G}^*) - \ln Z(\theta) \} \quad (24)$$

prescrizione che conduce a risolvere l'equazione

$$L(\mathbf{G}^*) = \sum_{\mathbf{G}} P(\mathbf{G}) L(\mathbf{G}) \quad (25)$$

e che stabilisce che il valore *osservato* del numero totale di link deve essere posto uguale al valore *atteso* (sull'ensemble di reti considerato,

evidentemente) dello stesso.

È possibile trovare una condizione più esplicita? Certo, non appena si considera che il valore atteso del numero totale di *link* non è altro che il valore atteso del numero di connessioni tra coppie di nodi e che ciascuna di esse stabilisce una connessione con la stessa probabilità, p : quindi

$$L(\mathbf{G}^*) = p \frac{N(N-1)}{2}; \quad (26)$$

invertendo questa relazione si ottiene

$$p = \frac{2L(\mathbf{G}^*)}{N(N-1)}. \quad (27)$$

Quindi, per il modello di Erdős, tarato sulla rete reale \mathbf{G}^* , il principio di massima verosimiglianza stabilisce che la probabilità che una qualsiasi coppia di nodi sia connessa da un link è uguale al (buon vecchio) "numero di casi favorevoli sui casi possibili".

Più in generale, il principio di massima verosimiglianza applicato ai modelli esponenziali conduce alla prescrizione

$$X_i(\mathbf{G}^*) = \langle X_i \rangle, \quad \forall i \quad (28)$$

ovvero, ad imporre che il valore osservato di ciascuno dei vincoli imposti coincida col valore atteso dello stesso vincolo.

Da notare che il principio di massima verosimiglianza completa il formalismo della Meccanica statistica tradizionale, secondo il quale il valore atteso di un qualsiasi vincolo si può calcolare derivando il logaritmo della funzione di partizione rispetto al moltiplicatore corrispondente, grazie alla celebre relazione $\nabla_{\theta_i} \ln Z(\vec{\theta}) = \langle X_i \rangle$; tuttavia, tale formalismo non fornisce alcuna prescrizione su come valutare i parametri che vi compaiono.

Il (meta)teorema di Gibbs

Prima di passare a discutere le applicazioni del formalismo di massima entropia, è utile rimarcare ancora una volta cosa rende potente tale formalismo: dato un *set* di vincoli, esso consente di derivare la distribuzione che minimizza l'impatto di assunzioni non giustificate.

Questo viene sancito formalmente dal teorema di Gibbs che stabilisce che, tra tutte

le distribuzioni di probabilità che soddisfano un *set* di vincoli dati, quella esponenziale è l'unica che massimizza l'entropia di Shannon [6]. Detto in termini più intuitivi, una volta convinti della bontà dell'entropia di Shannon quale misuratore dell'incertezza caratterizzante una certa distribuzione, massimizzare tale grandezza consente di derivare la distribuzione di probabilità minimamente distorta da assunzioni non giustificate. La relazione di Gibbs è, in effetti, un (meta)teorema perché coinvolge l'intera classe di distribuzioni che soddisfano uno stesso set di vincoli.

Ciò che può essere ricostruito...

Le applicazioni di maggior successo del formalismo appena descritto sono due: la *network reconstruction* e la *network validation*. In questo paragrafo affronteremo la discussione della prima applicazione.

Tipici sistemi di cui si ha soltanto conoscenza parziale sono le reti economiche e finanziarie. Ragioni che hanno a che fare con la privacy, infatti, non consentono di avere accesso alla matrice di adiacenza di tali sistemi nella sua interezza ma soltanto ad una parte di essa. L'informazione accessibile è, tipicamente, a livello di singolo nodo ed aggregata. Questo significa che si può conoscere il numero totale di relazioni commerciali di un paese (le esportazioni/importazioni totali, in un *framework* diretto) o il valore totale dei prestiti di una banca ma non "chi commercia con chi" oppure "quanto ha prestato la banca i alla banca j ". Il modo migliore per inferire l'informazione mancante è proprio quello di utilizzare il principio di massima entropia, considerando le informazioni in nostro possesso come vincoli.

Riprendiamo il World Trade Web e supponiamo di conoscere il numero di vicini di ciascun paese (ovvero, i primi vicini topologici e, cioè, i nodi direttamente collegati ai paesi stessi - che, quindi, si trovano ad appena un link di distanza da essi). La probabilità che i paesi i e j stabiliscano un legame commerciale è quello predetto dal Configuration Model, ovvero

$$p_{ij} = \frac{x_i x_j}{1 + x_i x_j}; \quad (29)$$

Inferenza statistica per reti complesse (in a nutshell)

In termini molto generali, la ricetta presentata in questo articolo per fare inferenza su reti complesse consta di due passi. In ciascuno di essi, entra in gioco una grandezza. Il primo passo prevede la stima della forma funzionale di una probabilità. A questo proposito, occorre massimizzare l'entropia di Shannon

$$S = - \sum_{\mathbf{G}} P(\mathbf{G}) \ln P(\mathbf{G})$$

vincolata, dove le grandezze che si vincolano rappresentano l'informazione che si ha sul sistema in esame. Il risultato è una distribuzione esponenziale, la cui forma funzionale è $P(\mathbf{G}) = e^{-H(\mathbf{G})}/Z, \forall \mathbf{G}$.

Una volta assunto che la probabilità di osservare la rete in esame \mathbf{G}^* sia descritta dalla forma funzionale di cui sopra, occorre stimare i parametri che definiscono $P(\mathbf{G}^*)$. A tal fine, si può invocare il principio della massimizzazione della verosimiglianza, ovvero della quantità

$$\mathcal{L}(\vec{\theta}) = \ln P(\mathbf{G}^* | \vec{\theta});$$

tale principio formalizza l'assunzione che la probabilità di osservare la configurazione che effettivamente si è realizzata, è massima. Una volta trovato il valore numerico della probabilità per la rete \mathbf{G}^* , è possibile calcolare i valori attesi di tutte le quantità di interesse (in generale, non osservabili) [11, 12].

a questo punto, potremmo assumere il punto di vista di i e chiederci con quanti paesi commerciano i nostri vicini. Tale numero è dato da

$$k_i^{(2)} = \sum_j \sum_l a_{ij} a_{jl} = \sum_j a_{ij} k_j \quad (30)$$

ma non è calcolabile, visto che non conosciamo \mathbf{A} , la matrice dei coefficienti a_{ij} . Tuttavia, possiamo calcolare il numero di paesi con cui ci aspettiamo che i nostri vicini commercino: tale numero è dato da

$$\langle k_i^{(2)} \rangle = \sum_j \sum_l p_{ij} p_{jl} = \sum_j p_{ij} k_j. \quad (31)$$

L'esempio di cui sopra ci mostra che è possibile calcolare i valori attesi di quantità di interesse mediante il modello che si è scelto di implementare. Considerare il World Trade Web, poi, è particolarmente utile perché è uno dei pochi sistemi di cui, effettivamente, conosciamo tutto: quindi, possiamo addirittura confrontare le previsioni di un certo modello con le osservazioni vere e proprie. Ciò che emerge da questo tipo di analisi è che la conoscenza della sequenza di grado con-

sente di ricostruire buona parte delle proprietà della rete del commercio globale [13].

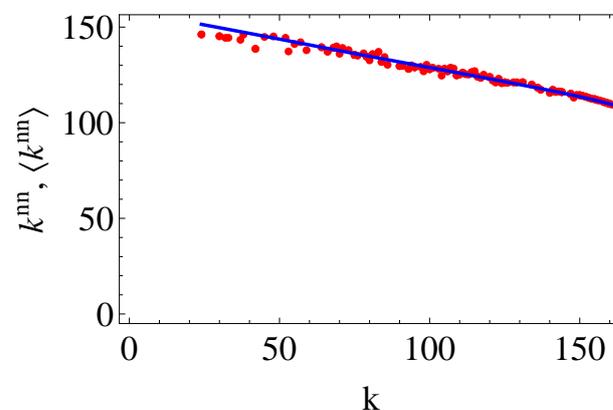


Figura 2: World Trade Web nell'anno 2002: andamento del numero medio dei vicini di ciascun vertice, in funzione del grado degli stessi (in rosso, l'andamento empirico; in blu, l'andamento atteso sotto il Configuration Model) [13].

Esempi di tale grandezze sono mostrate nelle figure 2 e 3 [13]: figura 2 mostra l'andamento della media aritmetica del numero di vicini di un certo nodo (ANND, da *average nearest neighbors*

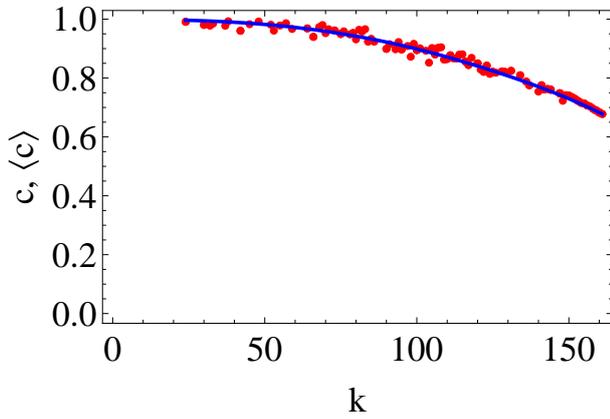


Figura 3: World Trade Web nell'anno 2002: andamento della frazione di triangoli di cui ciascun nodo è anche un vertice, in funzione del grado degli stessi (in rosso, l'andamento empirico; in blu, l'andamento atteso sotto il Configuration Model) [13].

degree, in inglese), in funzione del grado degli stessi, mentre figura 3 mostra l'andamento della frazione di triangoli di cui ciascun nodo è anche un vertice (CC, da *clustering coefficient*, in inglese), in funzione del grado degli stessi. Utilizzare (solamente) l'informazione contenuta nella sequenza di grado per stimare il valore atteso di ANND e di CC consente (già) di ottenere stime molto soddisfacenti.

Più in generale, gli andamenti mostrati in figura 2 e figura 3 rivelano che il World Trade Web è una rete disassortativa e gerarchica. Questo significa che nodi con molti vicini (ad esempio, gli USA) sono circondati da nodi che, a loro volta, hanno molti meno vicini (e viceversa) e che i nodi con grado molto elevato fanno parte di meno triangoli dei nodi con grado basso.

Finora, abbiamo utilizzato l'informazione contenuta nell'intera sequenza di grado. Adesso chiediamoci: riusciamo a ricostruire ancora bene la rete in esame, facendo a meno di tale informazione (ovvero rinunciando a conoscere il numero di vicini di ciascun nodo)? La risposta è affermativa ed il procedimento passa per l'identificazione di grandezze esterne alla rete che, tuttavia, abbiano la stessa potenza descrittiva dei gradi. Come anticipato, tali grandezze altro non sono che i valori del GDP dei singoli paesi. Ciò che si trova [12] è che il nostro set di incognite $\{x_i\}_{i=1}^N$ correla con il set di valori $\{GDP_i\}_{i=1}^N$ secondo

una legge che può essere formalizzata come

$$x_i = \sqrt{z} \text{GDP}_i, \quad \forall i \quad (32)$$

e che conduce alla riscrittura dei coefficienti di probabilità mostrata in Eq. (22).

Tale modello semplificato riduce enormemente la complessità numerica della procedura di stima dei parametri, richiedendo che venga fissato soltanto il valore di un parametro. A tal proposito, si può invocare di nuovo il principio di massimizzazione della verosimiglianza che, adesso, assume la forma più semplice

$$L^* = \langle L \rangle = \sum_{i < j} \frac{z \text{GDP}_i \text{GDP}_j}{1 + z \text{GDP}_i \text{GDP}_j} \quad (33)$$

(dove L^* indica il numero totale di connessioni che caratterizza il World Trade Web nell'anno scelto).

Tale prescrizione è particolarmente utile per studiare le reti finanziarie: ad esempio, quelle che rimangono definite dai prestiti tra banche. In tal caso, infatti, il numero di vicini di ciascun nodo non costituisce un'informazione disponibile: tutto ciò che si sa di ciascuna banca è il capitale, oppure la quantità di denaro prestata e ricevuta. Questo è, tuttavia, sufficiente a definire dei coefficienti di probabilità in modo del tutto analogo a quanto visto per il World Trade Web. Considerando che la rete di scambi interbancari è diretta, l'analogo finanziario del fitness model introdotto per studiare la rete del commercio globale diventa

$$p_{ij} \equiv \frac{z a_i l_j}{1 + z a_i l_j}, \quad \forall i \neq j \quad (34)$$

dove, adesso, a_i indica la quantità di denaro prestata dalla banca i -esima ed l_j indica quantità di denaro presa in prestito dalla banca j -esima. Naturalmente, il parametro z rimane ancora da definire: a tal fine, occorre conoscere la quantità totale di link della rete oppure stimarla sulla base di informazioni pubbliche di reti di banche in paesi finanziariamente simili a quello considerato.

L'efficacia di tale modello nel riprodurre le strutture finanziarie reali è evidente guardando figura 4. Sulla sinistra, è mostrata una fotografia della rete di banche denominata eMID (*Electronic Italian Interbank Market*, in inglese); sulla destra,

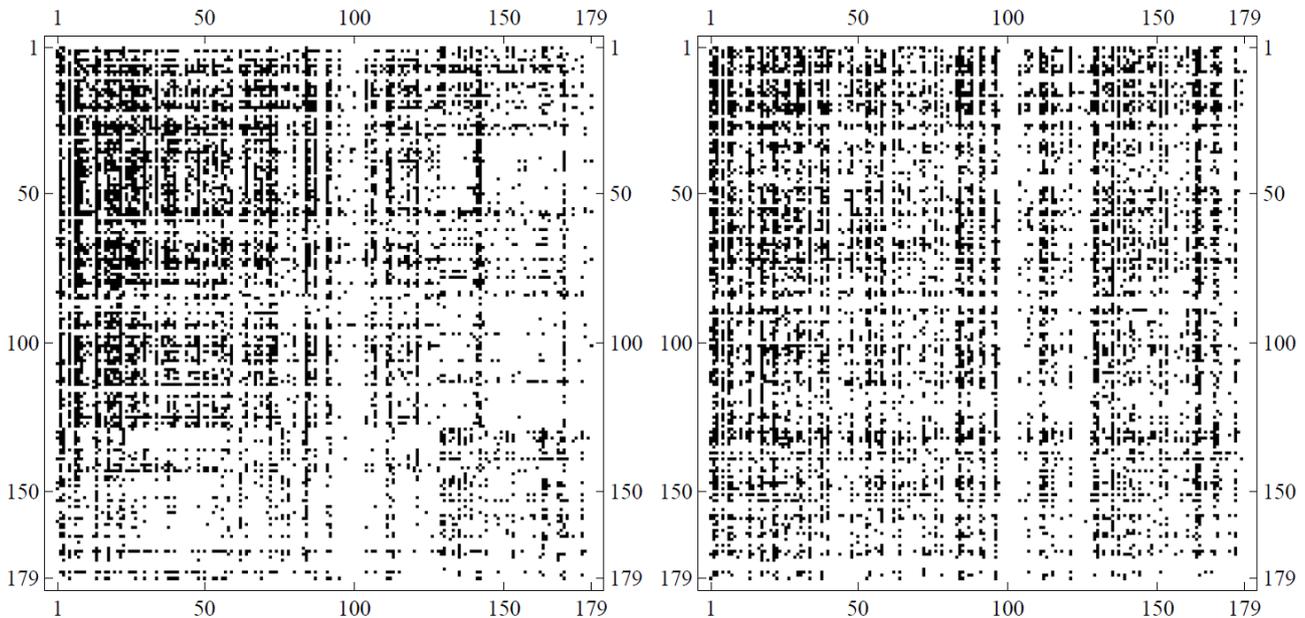


Figura 4: Sulla sinistra: matrice di adiacenza della rete interbancaria eMID, nel 1999. Sulla destra: matrice di adiacenza ricostruita sulla base del fitness model finanziario [13].

è mostrata una realizzazione tipica dell'ensemble di reti indotto dal modello appena descritto (generata campionando esplicitamente l'ensemble). Esercizi di questo tipo sono utilissimi per avere un'idea realistica della rete di scambi interbancari su cui girare i cosiddetti *stress test*, ovvero simulazioni di *shock* finanziari (e della propagazione delle perdite conseguenti), al fine di testare la robustezza di una rete in conseguenza di eventi (più o meno) catastrofici. Naturalmente, maggiore è l'accuratezza della rete ricostruita, più realistici saranno gli *stress test*: di conseguenza, maggiore sarà l'efficacia delle misure prese per contrastare gli *shock* che potrebbero manifestarsi in seguito a crisi bancarie.

... e ciò che non può esserlo

Nella sezione precedente ci siamo concentrati sul problema della *network reconstruction*, ovvero dell'inferenza di dettagli strutturali di reti di cui solo parte della topologia è nota. Tuttavia, il framework derivante dalla massimizzazione dell'entropia consente di fare anche *hypothesis testing* su strutture pienamente accessibili: in altre parole, consente di testare la significatività di certe grandezze di interesse contro l'ipotesi (nulla) che le stesse vengano spiegate dal modello scelto.

Ci si può chiedere quale sia l'interesse per tale tipo di analisi. La risposta è duplice:

- in primo luogo, questo è un modo per quantificare la complessità del sistema in esame: identificare il minimo *set* di vincoli in grado di replicare la struttura di una rete consente di individuare il livello minimo di informazione che bisogna conoscere per rendere conto degli andamenti osservati;
- in secondo luogo, questo tipo di analisi serve per rilevare e monitorare cambiamenti strutturali di una rete.

Sviluppiamo questo secondo punto, discutendo un esempio particolare. La DNB (ovvero la Banca Nazionale Olandese) possiede un *dataset* con tutte le transazioni tra le banche olandesi che sono occorse durante la decade 1998-2008. Per verificare l'impatto del primo anno della recente crisi economico-finanziaria (e, cioè, il 2008 stesso) sul sistema olandese si possono scegliere alcuni indicatori topologici e studiarne l'evoluzione. È quello che è stato fatto in [15], in cui gli autori hanno scelto di analizzare l'evoluzione della quantità detta *reciprocità* (ovvero la percentuale di link che hanno un compagno che punta in direzione opposta). Tale analisi rivela che la *reciprocità* diminuisce drasticamente in corrispondenza del 2008, fornendo un indicatore

topologico della crisi - evidentemente, meno coppie di banche hanno stabilito legami reciproci, a partire dai primi 3 mesi del 2008, rispetto a tutto il periodo precedente.

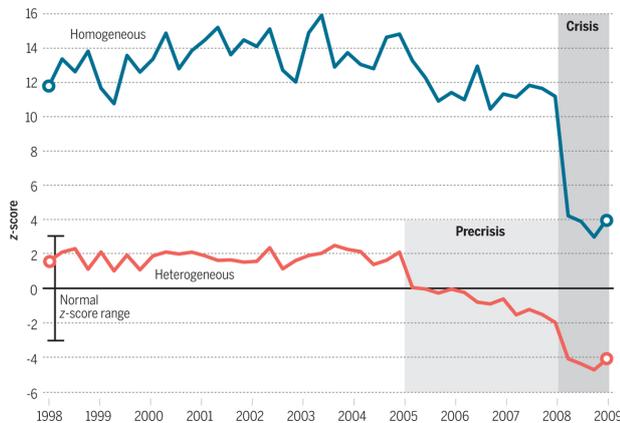


Figura 5: Analisi di significatività della reciprocità, misurata sulla rete di scambi tra banche olandesi nella decade 1998-2008. L'andamento in blu è ottenuto impiegando il modello di Erdős come benchmark; l'andamento in rosso è ottenuto impiegando il Configuration Model come benchmark [14].

Ora, confrontando l'evoluzione della reciprocità con le previsioni del Configuration Model, si nota che una decrescita del valore della reciprocità è osservabile già dal 2005. Come mostra figura 5, tale *early-warning signal* (ovvero, segnale anticipatore di un qualche pericolo), emerge nel momento in cui si valuta la significatività statistica della reciprocità empirica con la sua predizione sotto il Configuration Model: in altre parole, la rete di prestiti interbancari è caratterizzata da una reciprocità via via minore di quella predetta vincolando semplicemente i gradi. Quindi, non solo la sequenza di grado non riesce a riprodurre il valore atteso ma rivela una tendenza delle banche ad evitare di stabilire connessioni reciproche, a partire già dal 2005.

Da notare che il modello di Erdős fallisce nel rivelare tale tendenza: l'andamento decresce in modo netto solo nel 2008, rendendo impossibile capire se la rete in esame stia subendo un qualche tipo di cambiamento strutturale. La ragione della minore efficacia del modello di Erdős risiede nella totale incapacità di quest'ultimo di rendere conto dell'eterogeneità dei nodi della rete.

Infine, uno sguardo a Bayes

Nel presente contributo abbiamo descritto il *framework* derivante dalla massimizzazione vincolata dell'entropia di Shannon, mostrando un paio di esempi della sua applicazione. Tale formalismo può essere generalizzato introducendo un qualche tipo di informazione a priori (ad esempio, sulle strutture di rete da analizzare).

Questo tipo di informazione può essere inserito modificando il principio primo da cui partire: invece di massimizzare l'entropia, ovvero l'incertezza sul sistema *tout court*, si può minimizzare la distanza rispetto alla *prior information* in nostro possesso. In termini matematici, questa prescrizione si realizza minimizzando la quantità nota come Kullback-Leibler divergence, definita come

$$D_{KL}(P||Q) = \sum_{\mathbf{G}} P(\mathbf{G}) \ln \left(\frac{P(\mathbf{G})}{Q(\mathbf{G})} \right) \quad (35)$$

e che misura quanto differiscano tra loro due generiche distribuzioni di probabilità - qui indicate con P e Q . La minimizzazione vincolata di tale quantità conduce alla definizione di una Lagrangiana analoga a quella vista in precedenza, derivando la quale (rispetto a $P(\mathbf{G})$) si ottiene la distribuzione

$$P(\mathbf{G}) = \frac{Q(\mathbf{G})e^{-H(\mathbf{G})}}{\sum_{\mathbf{G}} Q(\mathbf{G})e^{-H(\mathbf{G})}} \quad (36)$$

dove $Q(\mathbf{G})$ gioca il ruolo di informazione a priori sulle strutture di rete. Supporre di non avere alcuna informazione a priori sul sistema (ovvero, che tutte le configurazioni ammesse siano equiprobabili) consente di ritrovare la forma funzionale classica; d'altra parte, possedere qualche informazione in più consente di piccare la distribuzione *a posteriori* sulle configurazioni compatibili con essa.

Esempi pratici dell'applicazione di tale formalismo comportano la specificazione di strutture mesoscopiche (ad esempio, le comunità) anche se non ne sono state ancora esplorate appieno le potenzialità come è avvenuto per il formalismo indotto dalla massimizzazione dell'entropia di Shannon.



- [1] L. Euler, *Solutio problematis ad geometriam situs pertinentis*, Commentarii Academiae Scientiarum Petropolitanae 8 (1736) 128.
- [2] P. Erdős, A. Rényi, *On random graphs. I*, Publicationes Mathematicae 6 (1959) 290.
- [3] E. N. Gilbert, *Random graphs*, Annals of Mathematical Statistics 30 (1959) 1141.
- [4] A.-L. Barabási, R. Albert, *Emergence of scaling in random networks*, Science 286 (1999) 509.
- [5] E. T. Jaynes, *Information theory and statistical mechanics*, The Physical Review 106 (1957) 620.
- [6] T. M. Cover, J. A. Thomas, *Elements of information theory*, Wiley-Interscience, New York (2006).
- [7] C. E. Shannon, *A mathematical theory of communications*, The Bell System Technical Journal 27 (1948) 379.
- [8] J. Park, M. E. J. Newman *The statistical mechanics of networks*, Physical Review E 70 (2004) 066117.
- [9] T. Squartini, D. Garlaschelli, *Analytical maximum-likelihood method to detect patterns in real networks*, New Journal of Physics 13 (2011) 083001.
- [10] D. Garlaschelli, M. I. Loffredo, *Fitness-dependent topological properties of the World Trade Web*, Physical Review Letters 93 (2004) 188701.
- [11] G. Cimini et al., *The statistical physics of real-world networks*, Nature Reviews Physics 1 (2019) 58.
- [12] T. Squartini, D. Garlaschelli, *Maximum-Entropy Networks: pattern detection, network reconstruction and graph combinatorics*, Springer, Cham (2017).
- [13] T. Squartini, G. Fagiolo, D. Garlaschelli, *Randomizing world trade. I. A binary network analysis*, Physical Review E 84 (2017) 046117.
- [14] S. Battiston et al., *Complexity theory and financial regulation*, Science 351 (2016) 818.
- [15] T. Squartini, I. van Lelyveld, D. Garlaschelli, *Early-warning signals of topological collapse in interbank networks*, Scientific Reports 3 (2013) 3357.



Tiziano Squartini: è RTD-B (Ricercatore a Tempo Determinato di tipo B) presso la Scuola IMT Altissimi Studi Lucca. Si occupa di meccanica statistica delle reti complesse e della sua applicazione allo studio di sistemi economici e finanziari.

Materiali fatti di molecole macroscopiche: dai granulari alla materia attiva

Hopeless thing sand. Nothing grows in it. All fades.

J. Joyce. Ulysses

Andrea Puglisi

*Istituto dei Sistemi Complessi - Consiglio Nazionale delle Ricerche
Dipartimento di Fisica, Università di Roma Sapienza, P.le Aldo Moro 2, 00185, Rome, Italy*

La Meccanica statistica è nata per descrivere sistemi fatti di molte molecole, provando a dedurre le leggi della materia che emergono dalle proprietà dei singoli costituenti. Negli ultimi decenni questa disciplina ha imparato a trattare sistemi in cui le unità elementari non sono più molecole ma particelle macroscopiche: l'esempio principale che descriviamo qui è la cosiddetta materia granulare, formata da grani di dimensioni millimetriche. Un materiale granulare può presentarsi in fasi simili a quelle di un materiale molecolare: solida, liquida o gassosa, ma le leggi che governano questi stati e le transizioni tra esse sono ancora poco comprese e riservano numerose sorprese. Un ruolo essenziale nella Fisica dei materiali granulari è giocato dagli attriti, ovvero l'ine-

vitabile conversione dell'energia dalla scala macroscopica a quella microscopica. Un granulare sottoposto a vibrofluidizzazione diventa quindi un esempio naturale di sistema stazionario fuori dall'equilibrio termodinamico, in cui si può osservare l'emergere spontaneo di ordine sotto forma di fenomeni convettivi, separazione di fase, segregazione, e effetti motore. In questa breve rassegna di fenomeni e teorie granulari, accennerò anche alle numerose analogie con i sistemi attivi, formati da particelle dotate di auto-propulsione, quali i batteri e gli uccelli.

Introduzione

Che differenza c'è tra un gas di molecole e un gas di palline? Perché rovesciare un secchio di granaglie ci sembra tanto simile a versare dell'acqua

da una bottiglia, eppure il flusso di uscita dal fondo di un silo può interrompersi senza preavviso? Come mai agitando una miscela di diverse sabbie possiamo arrivare a separare i diversi materiali, mentre una miscela di liquidi diversi non si separa? Queste domande vicine alla vita di tutti i giorni e a molte applicazioni industriali sono l'occasione per parlare di una direzione recente della Fisica teorica, ovvero la Meccanica statistica di non-equilibrio.

Un po' di storia. La Meccanica statistica è una delle branche principali della Fisica teorica, nata ancora prima della Relatività generale e della Meccanica quantistica ad opera di alcuni padri fondatori nella seconda metà del diciannovesimo secolo, principalmente James Clerk Maxwell e Ludwig Boltzmann. L'intento della teoria è costruire una descrizione statistica di un sistema formato da moltissime molecole, in modo da poter dedurre le proprietà macroscopiche del sistema a partire da quelle microscopiche dei singoli costituenti. La Meccanica statistica ha avuto i suoi primi successi nello spiegare l'origine microscopica della pressione e della temperatura di un gas, e della relazione (equazione di stato) che lega queste quantità. L'opera di Boltzmann ha permesso poi di dare una definizione microscopica dell'entropia e di affrontare la questione dell'evoluzione temporale, ovvero di cosa accade quando le proprietà statistiche di questi sistemi composti da molte molecole cambiano nel tempo, arrivando a comprendere il significato microscopico (puramente probabilistico) del secondo principio della termodinamica, secondo il quale in un sistema isolato l'entropia di un sistema non può diminuire. Nei decenni successivi la Meccanica statistica ha raccolto i suoi maggiori successi sul vasto terreno dei sistemi all'equilibrio termodinamico, dove l'entropia è già massima e questo permette di dedurre più facilmente le proprietà statistiche di un sistema conoscendo solo la sua Hamiltoniana ovvero la funzione che ne determina l'energia. Il ventesimo secolo è stato quindi dominato dall'applicazione della ricetta fondamentale della Meccanica statistica di equilibrio ai sistemi più disparati, arrivando a comprendere in profondità gli affascinanti meccanismi delle transizioni di fase, anche in situazioni estreme come temperature molto basse o forti correlazioni, grazie anche al-

l'introduzione, in questo schema teorico, delle leggi della Meccanica quantistica.

La conservazione dell'energia. Fin verso gli anni '80 del ventesimo secolo, in tutta la trattazione della Meccanica statistica di equilibrio, come anche nelle numerose sortite di questa teoria verso i fenomeni di non-equilibrio (dovuti cioè a rilasciamenti del sistema da condizioni iniziali atipiche, oppure alla presenza di forze esterne che ne perturbano lo stato di equilibrio), un principio è sempre stato tenuto fermo, ovvero la conservazione dell'energia, che produce due fondamentali conseguenze: la presenza di una ben definita Hamiltoniana e l'esistenza, nella maggior parte dei casi, di un bagno termico, ovvero un sistema molto più grande di quello considerato, che svolge il ruolo di riserva o tampone di energia.

La conservazione dell'energia è, ovviamente, un principio che non è mai messo in discussione in Fisica. Al di fuori della Meccanica statistica viene a volte trascurato per lasciar posto ad una descrizione effettiva di alcuni sistemi macroscopici, spesso per scopi applicativi al confine tra la Fisica e l'Ingegneria. Ad esempio la conservazione dell'energia è trascurata nei modelli di attrito, fin dall'epoca dei primi studi ad opera di Guillaume Amontons e Charles Augustin Coulomb, risalenti all'infanzia della rivoluzione industriale. Ugualmente la si trascura nei modelli di fluidodinamica con viscosità, come per esempio nell'equazione di Navier-Stokes. Anche la Meccanica statistica è stata costretta a trascurare la conservazione dell'energia quando si è rivolta allo studio di alcuni sistemi inizialmente interessanti solo per gli ingegneri, ovvero i materiali granulari, le cui interazioni coinvolgono - per l'appunto - l'attrito e la viscosità. La rinuncia di uno dei principi fondamentali della teoria, è stata in realtà l'occasione per aprire la disciplina a un mondo molto più ampio di fenomeni (tutti macroscopici) che non comprendono solo i materiali granulari ma anche materiali viventi ovvero formati da molte unità elementari di interesse biofisico, quali i tessuti cellulari, le colonie di batteri e persino i banchi di pesci e gli stormi di uccelli. È bene però raccontare questa storia dall'inizio, spiegando prima di tutto cosa sono i materiali granulari.



Figura 1: Alcuni esempi di materiali granulari che si trovano in natura, nell'industria e... nello spazio.

Materiali emergenti, nella vita quotidiana

I sistemi granulari ci circondano [1]. Si tratta di qualunque insieme di molti grani con un diametro di circa un millimetro, anche un po' più piccoli (decine di micron) o un po' più grandi. Lo zucchero, il sale, la farina, il riso, sono i sistemi granulari che troviamo nelle nostre cucine. La neve, la grandine, la sabbia, sono granulari presenti in natura. Possiamo considerare sistemi granulari anche gli anelli planetari, formati da rocce e polveri di dimensioni anche molto maggiori del millimetro. L'interesse applicativo per questi sistemi, che precede la Fisica di molti decenni, riguarda numerosi processi minerari e industriali, essendo considerati materiali granulari i suoli da cui si estraggono le risorse energetiche naturali, le polveri utilizzate in molte manifatture, nella produzione di alimenti e medicinali, in molti aspetti della chimica industriale. Ottimizzare questi e altri processi, come il trasporto e lo stoccaggio di granaglie o sabbie per il cemento, può dar luogo a risparmi considerevoli. La comprensione della dinamica dei mezzi granulari può avere conseguenze anche per lo studio del rischio geologico, in particolare per i terremoti e le valanghe, ma anche per il trasporto dei detriti nei fiumi.

Sebbene per un granulare si capisca l'uso del termine sistema è meno chiaro (ma molto frequente il suo uso) il termine materiale. I singoli grani sono composti da materiali usuali, sostanze ordinarie (molecolari) in fase solida. Perché allora si parla di materiali granulari? Il termine deriva dall'osservazione che un insieme di molti grani solidi ha un comportamento piuttosto peculiare, diverso da quello dei solidi o di altre fasi o materiali esistenti. È il caso quindi di parlare di materiale emergente, che di fatto esiste solo come conseguenza dell'aver preso molte unità elementari interagenti, ciascuna composta da un materiale solido ordinario.

Quali sono gli ingredienti fisici che rendono strano, o anomalo, il comportamento di un materiale granulare? L'ingrediente fondamentale è la macroscopicità delle unità elementari. Un granello di un millimetro di diametro, a differenza di un singolo atomo o una singola molecola (le unità elementari dei materiali ordinari), è composto a sua volta di un numero enorme di atomi e molecole. Per via delle sue dimensioni, il moto del centro di massa di un granello non risente (se non in maniera del tutto trascurabile) dell'agitazione termica a temperatura ambiente: l'invisibile agitazione molecolare che permea tutte le sostanze è responsabile della dinamica di atomi e molecole ma certamente non può mettere in moto dei granelli di sabbia. Per lo stesso motivo, un granello di sabbia lasciato cadere da una certa altezza su un piano solido dissiperà rapidamente tutta l'energia cinetica acquisita durante la caduta: questa energia non sparisce ma viene trasformata in agitazione termica (dapprima all'interno del grano e nel piano solido, per finire poi dispersa in tutto l'ambiente circostante), che per lo stesso motivo di prima, è invisibile alla scala del granello e non può tornare ad animare il suo moto macroscopico. Queste due conseguenze della macroscopicità, due facce della stessa medaglia, ovvero l'irrelevanza della temperatura e la dissipazione di energia negli urti, sono le caratteristiche essenziali di un materiale granulare, che ne determinano tutte le anomalie e rappresentano il... ruvido guanto di sfida lanciato alla Meccanica statistica.

Le fasi granulari. Un materiale granulare può trovarsi in stati all'apparenza simili a quello solido, liquido o gassoso, ma che nella maggior par-

te dei casi sono stati intermedi o ambigui [2, 3]. Un cumulo di sabbia può essere in uno stato di quiete, del tutto simile a un solido, se gli angoli delle pendici sono minori dell'angolo critico per la formazione di valanghe. Al contrario, se l'angolo delle pendici supera l'angolo critico, una parte considerevole del cumulo scorre in maniera simile a un fluido, sotto l'effetto della gravità, formando per l'appunto valanghe. Versando un pacco di farina in una scatola si ha l'impressione di versare un liquido, ad esempio osservando come questo prende la forma del contenitore in cui si versa. È facile però avere una sensazione di solidità provando a comprimere la superficie della farina nella scatola, ad esempio con una mano o un pugno. Sensazione sostituita da quella di fluidità, se si immerge un dito o un cucchiaino, o da quella di un gas, se si soffia all'interno della scatola, sollevando in aria una nuvola di farina. Si tratta, ovviamente, di comportamenti che conseguono a perturbazioni esterne (versare, premere, scavare, soffiare), ma che comunque configurano una grande varietà di risposte per un sistema che è sostanzialmente sempre nelle stesse condizioni iniziali. I materiali molecolari raramente mostrano questa particolare sensibilità al tipo di perturbazione. Qualche esempio di questa variabilità è presente in materiali solidi con particolari strutture cristalline che danno luogo a risposte differenti (duttilità, malleabilità, esfoliazione): si tratta comunque di proprietà che caratterizzano particolari materiali e quindi giustificano ancora di più l'uso del termine *materiale* o *stato della materia* per i sistemi granulari.

Granulari statici. Uno dei fenomeni più sorprendenti osservati nei materiali granulari in assenza di perturbazioni esterne è la cosiddetta **legge di Janssen**. Questa legge descrive il comportamento della pressione all'interno di un materiale granulare in un contenitore cilindrico (con possibili estensioni anche ad altre geometrie). Così come in un liquido la pressione varia con la profondità, così ci si aspetta che succeda per un granulare. In una colonna di liquido di densità ρ la pressione cresce con la profondità h seguendo la legge di Stevino $p = \rho gh$ (g è l'accelerazione di gravità). In un granulare a basse profondità si osserva una legge analoga, ma quando la profondità supera una lunghez-



Figura 2: Silos crollati per un'imprevisto sbilanciamento della pressione interna sulle pareti rispetto al fondo. I crolli sono in parte dovuti alla legge di Janssen (spiegata nel testo) che complica la progettazione, sia alle grandi fluttuazioni interne delle forze tra i grani che implicano deviazioni imprevedibili da questa stessa legge.

za critica dello stesso ordine del diametro della sezione del cilindro, la pressione satura a un valore costante. Una buona approssimazione del comportamento osservato è data dalla formula

$$p = p_s [1 - \exp(-h/h^*)]$$

dove p_s è la pressione di saturazione e h^* l'altezza critica che separa il regime di crescita lineare della pressione, a piccole profondità $h \ll h^*$, e il regime a pressione costante per $h \gg h^*$. Un argomento semplice che spiega questa legge e ne predice anche le (a volte catastrofiche) conseguenze è il seguente: si aggiunge all'usuale bilancio di forze che determina la statica di una fetta orizzontale di materiale, oltre alla forza di gravità e alla differenza di pressione tra la superficie superiore e quella inferiore della fetta, anche le forze di attrito statico esercitate dalla parete del contenitore, che si assumono proporzionali alla pressione stessa. Questo argomento, appunto, individua nelle pareti del contenitore l'origine della forza che sostiene il peso del materiale oltre l'altezza critica. Una sottovalutazione di questa legge e del ruolo delle pareti comporta un errato dimensionamento dei silos per lo stoccaggio di granulari quali grano e cemento, e la loro conseguente - spesso improvvisa - rottura.

Granulari dinamici. L'applicazione di perturbazioni ripetute nel tempo, esempio tipico una vibrazione con caratteristiche stazionarie, può indurre ulteriori anomalie nel comportamento dei materiali granulari, dando luogo a sta-

ti con proprietà ancora differenti dai granulari statici. Una vibrazione continua del contenitore del materiale granulare tipicamente induce uno stato parzialmente o completamente fluido che presenta una notevole ricchezza e complessità di fenomeni. Un primo parametro utilizzato per caratterizzare l'intensità della vibrazione è l'accelerazione massima della vibrazione normalizzata con l'accelerazione di gravità. Ad esempio se la vibrazione è sinusoidale con legge per la quota del contenitore $z(t) = A \sin(\omega t)$, l'accelerazione massima è $\ddot{z}_{max} = \omega^2 A$ e il parametro adimensionale della vibrazione è $\Gamma = \ddot{z}_{max}/g$. Nelle situazioni in cui $\Gamma \ll 1$ si hanno fenomeni molto lenti di deformazione e trasformazioni interne al granulare, mentre quando $\Gamma \gg 1$ si ha una dinamica rapida. Per **vibro-fluidizzazione** si intendono, in genere, entrambi i regimi, ma la differenza tra di essi è paragonabile a quella tra un liquido sopra e sotto la temperatura di fusione (quando lo si osserva su tempi più brevi del tempo di congelamento), con tempi di rilassamento che possono essere notevolmente differenti, anche di diversi ordini di grandezza.

Fenomeni collettivi. La vibro-fluidizzazione di un materiale granulare può indurre nel materiale fenomeni di ordinamento di molti generi differenti. Onde stazionarie possono formarsi sulla superficie del granulare o coinvolgere un intero strato di materiale. In certi regimi si può sviluppare il cosiddetto **effetto Leidenfrost**, effetto inizialmente osservato per una goccia di liquido ordinario su una superficie molto calda, tale da sviluppare tra la superficie e la goccia un cuscino di vapore del liquido stesso, che può avere una vita particolarmente lunga. Nel caso di un granulare questo effetto si realizza con la formazione di una regione del granulare più densa (simile alla goccia) sovrastante una regione in fase gassosa (simile al vapore) a sua volta a contatto diretto con la superficie vibrante (simile alla superficie calda).

In entrambe i regimi di vibrazione $\Gamma \ll 1$ e $\Gamma \gg 1$ (fluido lento e fluido veloce) si possono osservare movimenti convettivi che coinvolgono gran parte del granulare, trascinandolo periodicamente in percorsi circolari all'interno del contenitore. Questi movimenti, che sono piuttosto comuni, possono avvenire in direzioni diverse a seconda delle proprietà geometriche del

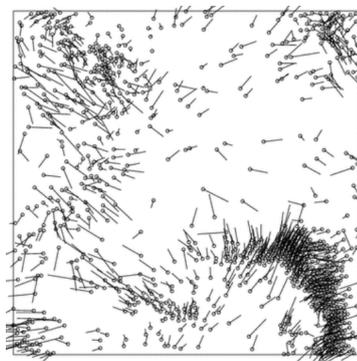


Figura 3: Uno "sciame" granulare osservato in una simulazione di dischi duri anelastici, preparati inizialmente in una configurazione con posizioni omogeneamente casuali e velocità estratte da una distribuzione di Maxwell. Durante l'evoluzione l'omogeneità spaziale si perde, per essere sostituita da una forte polarizzazione locale della velocità a cui segue una evidente coagulazione spaziale delle posizioni.

contenitore e dei singoli grani, e sono spesso difficili da prevedere. Nei regimi lenti questi moti convettivi possono portare alla segregazione spontanea di una miscela, il caso più eclatante e famoso è quello del cosiddetto effetto di noce brasiliana. Nei regimi veloci la convezione acquista le sembianze della termoconvezione nota nei fluidi, il cui esempio tipico è la convezione di Rayleigh-Bénard in presenza di un gradiente termico (fluido riscaldato da sotto, come nel caso di una pentola di acqua sul fuoco). L'aspetto teorico interessante, nel caso del granulare, è che questo tipo di convezione si sviluppa in assenza di un gradiente termico imposto dall'esterno: è sufficiente un solo bagno termico (cioè una piastra vibrante a contatto con la superficie inferiore del granulare), sarà la dissipazione interna dovuta agli urti anelastici tra grani a produrre spontaneamente un gradiente termico che, oltre ad una certa soglia critica, induce la rottura della simmetria per traslazioni (nella direzione parallela alla piastra) e la formazione di celle convettive, con un meccanismo la cui compressione teorica è del tutto analogo a quello della termoconvezione di Rayleigh-Bénard.

Un'altra categoria di fenomeni collettivi si osserva principalmente nelle simulazioni numeriche di modelli di gas granulari, nel cosiddetto regime di raffreddamento: si immagina, cioè, di sottoporre il granulare a vibro-fluidizzazione

e poi spegnere improvvisamente il sistema vibrante, osservando ad altissima risoluzione temporale la dinamica di spegnimento. Durante il raffreddamento, un gas granulare evolve attraverso una dinamica altamente correlata, con veri e propri sciame di grani che si muovono coerentemente. Si intuisce che fenomeni di questo tipo, come anche i fenomeni termoconvettivi descritti prima, sono particolarmente adatti ad un approccio di Meccanica statistica, che andiamo ora a descrivere.

La Meccanica statistica dei materiali granulari

Le osservazioni discusse qui sopra sottolineano, oltre alle differenze, anche le forti analogie con i materiali molecolari, suggerendo la necessità di una nuova Meccanica statistica che descriva il comportamento dei materiali granulari. In questa Meccanica statistica granulare, però, le unità elementari sono i grani, mentre è ragionevole trascurare tutti i gradi di libertà molecolari interni al singolo grano (così come quelli dell'aria o dei fluidi circostanti i grani). E per questo motivo si deve rinunciare alla conservazione dell'energia e alla sua fondamentale conseguenza, la distribuzione di Gibbs-Boltzmann $\sim e^{-H/k_B T}$. In effetti per un materiale granulare non esiste un Hamiltoniana H ed è molto sfuggente il concetto stesso di temperatura T . Di fronte al cedimento delle fondamenta della Meccanica statistica, la nuova teoria ha seguito diverse strade. Ne descriviamo brevemente due.

La Meccanica statistica di Edwards. Questo approccio è stato proposto da Sir Sam Edwards negli anni '80 del ventesimo secolo per descrivere sistemi granulari compatti e perturbati molto debolmente. Un esperimento tipico che può venir descritto da questa teoria è il cosiddetto esperimento di tapping in cui il materiale viene perturbato periodicamente e tra una perturbazione e l'altra passa abbastanza tempo da far raggiungere al materiale una nuova configurazione a riposo. Il materiale quindi esplora una sequenza di configurazioni di equilibrio meccanico. L'osservazione empirica (in laboratorio o nelle simulazioni) dimostra che il volume occupato dal granulare diminuisce sempre più lentamente: in

effetti il granulare ottimizza lo spazio occupato, ma questa ottimizzazione diventa via via più difficile. L'idea di Edwards è di considerare un nuovo *ensemble* statistico, ovvero la collezione di tutti gli stati di equilibrio meccanico ad un dato volume occupato. Questo *ensemble* sarebbe l'analogo del microcanonico nella Meccanica statistica usuale, in cui quindi l'energia viene sostituita dal volume occupato. A partire da questo *ensemble* è possibile costruire un *ensemble* canonico coniugato e quindi una temperatura (la compattività) associata alle fluttuazioni di volume. L'idea di Edwards è una delle poche generalizzazioni esistenti della teoria di Boltzmann-Gibbs ai sistemi fuori dall'equilibrio, e si basa su un'ipotesi ergodica molto coraggiosa, ovvero l'idea che in un esperimento di tapping il sistema possa esplorare con uguale probabilità tutti gli stati con lo stesso volume. Questa idea è ancora oggi molto dibattuta, esistono numerosi esempi a favore e contro le ipotesi di Edwards [4]. La teoria di Edwards ha avuto fortuna anche per la forte connessione tra il problema della compattazione granulare e numerosi problemi di Fisica dei sistemi disordinati, vetrosi e soffici. Lo stato bloccato (*jammed*) di un materiale granulare mostra forti analogie non solo con lo stato solido della materia molecolare, ma anche con gli stati bloccati osservati in certe sospensioni colloidali (ad esempio dopo una centrifuga), emulsioni o schiume compresse, vetri strutturali e vetri di spin al di sotto della temperatura critica per la transizione vetrosa, e materiali biologici come cellule, DNA e *packing* di proteine. In tutti questi sistemi ricorre l'ingrediente della ottimizzazione ostacolata da barriere energetiche molto alte che comportano tempi di rilassamento estremamente lunghi.

La Meccanica statistica dei gas e liquidi granulari. Come abbiamo visto discutendo dei fenomeni convettivi, i materiali granulari sottoposti a intensa vibro-fluidizzazione possono somigliare molto a fluidi in rapida evoluzione, particolarmente quando $\Gamma \gg 1$. Per questi sistemi è più naturale un approccio teorico slegato dalla Fisica dei fenomeni lenti e vetrosi, ispirato invece alla usuale Meccanica dei fluidi. Dopo una serie di teorie idrodinamiche per fluidi granulari basate su principi macroscopici e di simmetria, dalla fine degli anni '90 sono state proposte vere e proprie teorie cinetiche granulari, ovvero dedu-

zioni della fluidodinamica (compresi i cosiddetti coefficienti del trasporto) a partire da modelli di gas con interazioni inizialmente trascurabili e via via più importanti [5].

Il modello principe di gas granulare è il gas di sfere dure anelastiche. Le sfere si muovono liberamente fino a toccarsi. Quando due sfere urtano, un'interazione istantanea trasforma i vettori velocità dei centri di massa delle due sfere da $\mathbf{v}_1, \mathbf{v}_2$ a $\mathbf{v}'_1, \mathbf{v}'_2$ con una legge che conserva il momento e riduce l'energia cinetica associata al moto relativo:

$$\mathbf{v}'_1 = \mathbf{v}_1 - \frac{1-\alpha}{2}[(\mathbf{v}_1 - \mathbf{v}_2) \cdot \hat{n}]\hat{n} \quad (1a)$$

$$\mathbf{v}'_2 = \mathbf{v}_2 + \frac{1-\alpha}{2}[(\mathbf{v}_1 - \mathbf{v}_2) \cdot \hat{n}]\hat{n}. \quad (1b)$$

In questa regola di urto \hat{n} è la direzione che unisce i centri delle due sfere, mentre il parametro fondamentale è $\alpha \in (0, 1]$ che determina la frazione di energia cinetica che rimane nel sistema. Se $\alpha = 1$ la collisione è elastica, altrimenti si ha dissipazione. Modelli più realistici possono includere le rotazioni e lo scambio di momento angolare (con eventuale dissipazione di energia rotazionale) dovuto ad asperità superficiali, o l'elasticità e quindi la deformazione delle sfere, incorporando quindi un tempo finito - non istantaneo - per l'interazione.

Il primo passo di una teoria cinetica granulare è la scrittura di un'equazione di Boltzmann che descrive l'evoluzione nel tempo t della distribuzione di probabilità nello spazio delle fasi di singola particella $P(\mathbf{r}, \mathbf{v}, t)$. Questa equazione la si può derivare da equazioni di pseudo-Liouville (cioè equazioni per l'evoluzione dell'intero sistema, nello spazio delle posizioni e velocità per N particelle, identiche a quelle originali di Liouville per la Meccanica delle molecole, ma con l'aggiunta della dissipazione). Si possono effettuare operazioni formalmente simili al limite di Grad-Boltzmann per i sistemi elastici, oppure lavorare euristicamente per analogia con l'equazione di Boltzmann originale [6, 5]. La cinematica dell'urto viene sostituita con la regola di interazione anelastica in Eq. (1). Il risultato è un'equazione

di questo tipo

$$\frac{\partial P}{\partial t} + \mathbf{v} \cdot \nabla_{\mathbf{r}} P + \nabla_{\mathbf{v}} \cdot (\mathbf{F}P) = Q(P), \quad (2a)$$

$$Q(P) = \int_{\mathcal{R}^3} d\mathbf{v}^* \int_{S^2} d\hat{n} |\mathbf{v} - \mathbf{v}^*| \left(\frac{1}{\alpha^2} P' P^{*'} - P P^* \right) \quad (2b)$$

dove abbiamo introdotto molte quantità importanti: \mathbf{F} è una forza esterna, non necessaria ma per i granulari essenziale se si vuole raggiungere uno stato stazionario non banale, Q il cosiddetto operatore di collisione, e infine le probabilità P^* , P' e $P^{*'}$ che sono la $P(\mathbf{r}, \mathbf{v}, t)$ calcolata rispettivamente in $\mathbf{v} = \mathbf{v}^*$, $\mathbf{v} = \mathbf{v}'$ e $\mathbf{v} = \mathbf{v}^{*'}$ dove \mathbf{v}' , $\mathbf{v}^{*'}$ sono le **velocità precollisionali** cioè quelle che inserite nella equazione (1) al posto di $\mathbf{v}_1, \mathbf{v}_2$ restituiscono $\mathbf{v}'_1 = \mathbf{v}$ e $\mathbf{v}'_2 = \mathbf{v}^*$. L'equazione di Boltzmann granulare ha una fondamentale differenza rispetto al caso elastico: non esiste una soluzione stazionaria che non sia quella banale in cui tutte le velocità sono nulle, proprio come un esperimento senza vibrazioni esterne evolve fino all'esaurimento di tutta l'energia cinetica (convertita in calore e sparita dalla nostra descrizione). Per tenere vivo un granulare bisogna fornire continuamente energia dall'esterno e questo si riflette, nei modelli, nell'aggiunta di termini forzanti \mathbf{F} nell'equazione di Boltzmann, eventualmente stocastici. Ad esempio, modelli non troppo idealizzati includono dei termini di diffusione nello spazio delle velocità, simili a forzanti di Langevin che possono riprodurre anche risultati sperimentali con un buon grado di approssimazione [7, 8]. L'equazione di Boltzmann granulare con termini forzanti può prevedere soluzioni stazionarie. Queste soluzioni, anche nel caso spazialmente omogeneo, non sono delle distribuzioni di Gauss e per esse non vale l'usuale teorema H. Nell'ultimo decennio è stato proposto, e dimostrato entro certe condizioni, un teorema H granulare, con minor potere predittivo [9, 10].

Procedure di proiezione più o meno standard (come ad esempio la cosiddetta espansione di Chapman-Enskog) per derivare la fluidodinamica usuale sono state applicate all'equazione di Boltzmann granulare per derivare una fluidodinamica granulare, con equazioni tipo Navier-Stokes per i campi densità, velocità e temperatura granulare. Quest'ultimo campo merita una breve discussione. Si tratta di una temperatura ci-

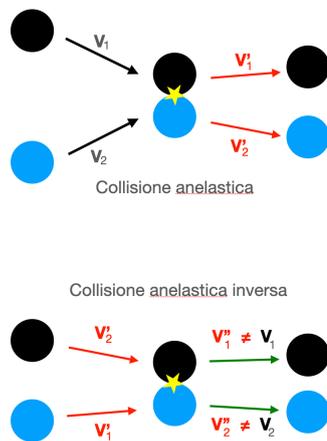


Figura 4: Esempio di una collisione anelastica che, per effetto della dissipazione, risulta in velocità uscenti leggermente più parallele di quelle entranti. In basso si vede cosa accade quando due particelle granulari si incontrano con le due velocità uscenti: il risultato sono due velocità ancora più allineate e certamente differenti da quelle iniziali. Nel caso elastico invece la seconda collisione riprodurrebbe le velocità iniziali. Per questo motivo si dice che le collisioni anelastiche sono irreversibili temporalmente.

netica, definita sulla base dell'energia cinetica media locale dei grani, e quindi non ha nulla a che vedere con la temperatura dell'ambiente e, soprattutto, non ci si aspetta che abbia proprietà termodinamiche. In molti modelli ed esperimenti si osserva la violazione dell'equipartizione dell'energia.

La fluidodinamica granulare ha conseguito molti successi. Essa è riuscita a spiegare le instabilità osservate nelle simulazioni di gas di sfere dure anelastiche, quando durante il regime di raffreddamento omogeneo il campo di velocità si increspa dando luogo a onde trasversali (*shear*) e vortici, e conseguentemente il campo di densità perde in omogeneità formando aggregati molto densi. Simili fenomeni sono osservati anche negli esperimenti. Le teorie cinetiche granulari hanno anche permesso di comprendere i fenomeni termo-convettivi dovuti alla sola anelasticità.

Un aspetto fondamentale della Meccanica statistica dei granulari è la rottura intrinseca della simmetria per inversione temporale. Per un sistema fatto di molecole questa simmetria può essere rotta solo da forze esterne, ad esempio preparando il sistema lontano dalla sua condizione di equilibrio, oppure sottoponendo il sistema ad

una forza stazionaria nel tempo. Esempio classico di questa rottura stazionaria dall'esterno in Meccanica dei fluidi, è la turbolenza: il fluido viene sottoposto a forzanti esterne (ad esempio degli sforzi di taglio ai bordi del contenitore) e si genera un flusso di energia dalle grandi alle piccole scale, flusso che certamente determina una direzione temporale privilegiata, anche se le singole molecole del fluido obbediscono ad equazioni del moto simmetriche per inversione temporale. Per i materiali granulari sono le interazioni stesse tra grani a determinare la rottura della simmetria per inversione temporale. Ogni collisione trasferisce energia dalla scala del moto del centro di massa alla scala dei costituenti elementari del singolo grano. Nella figura qui sopra si capisce come una collisione anelastica è temporalmente irreversibile.

Questa osservazione permette di costruire delle apparenti violazioni del secondo principio della termodinamica. Si può concepire infatti un modello di macchina termica granulare che trasforma calore in lavoro, seguendo il progetto del motore molecolare di Smoluchowski-Feynman: un meccanismo rotante asimmetrico viene posto al centro di un gas granulare diluito, ad una ben determinata temperatura granulare. Si può dimostrare con conti analitici di teoria cinetica, simulazioni e anche esperimenti, che il meccanismo si mette in moto in una direzione media ben definita e così facendo è in grado di sollevare un peso. Apparentemente la macchina assorbe calore da una sorgente termica, senza essere in contatto con altri bagni. Si tratta ovviamente di un'illusione, il gas granulare non è un bagno termico, la dissipazione delle collisioni è a tutti gli effetti analoga ad un contatto con un bagno a temperatura ambiente, molto minore della temperatura granulare. Nel sistema c'è una corrente di energia che scorre dallo *shaker* che agita il materiale granulare fino al calore dissipato nell'ambiente. Il fatto che in situazioni molto diluite e fortemente agitate il gas granulare può sembrare - guardando ad esempio la statistica delle velocità dei grani - all'equilibrio, aumenta ovviamente la sensazione di un paradosso [5].

Ci sembra importante un cenno ad un ulteriore aspetto concettuale della Meccanica statistica granulare, che riguarda le fluttuazioni. Sebbene

un sistema granulare sia formato da molti grani, questo molti non è lontanamente paragonabile al numero di molecole coinvolto dalla Meccanica statistica usuale, dell'ordine di grandezza del numero di Avogadro e per questo spesso sostituito da infinito nel cosiddetto limite termodinamico. In molti sistemi di laboratorio il numero di grani si aggira tra le centinaia e le migliaia. Raramente raggiunge le centinaia di migliaia. Con taglie (nel senso del numero di gradi di libertà) così piccole, i granulari si possono considerare sistemi statistici con fluttuazioni apprezzabili. Questo ha motivato la ricerca teorica verso un'ulteriore sfida: aggiungere alle descrizioni tipo Teoria cinetica e Fluidodinamica anche dei termini stocastici appropriati (nei campi e nelle correnti associate), per poter descrivere le proprietà di queste fluttuazioni. La sfida è particolarmente stimolante in quanto non ci si aspetta la validità delle relazioni che legano le fluttuazioni e le forze deterministiche all'equilibrio, relazioni scoperte da Einstein e da altri grandi scienziati della prima metà del ventesimo secolo, quali ad esempio Onsager e Kubo. La Teoria delle fluttuazioni idrodinamiche di Landau, per esempio, è in grado di rispondere a questa domanda per sistemi molecolari, ma non può essere direttamente adattata all'Idrodinamica granulare, se non a prezzo di approssimazioni che si è visto essere valide solo nel limite diluito. Altre approssimazioni ed altre teorie sono in corso di sviluppo negli ultimissimi anni.

Altri materiali costituiti da particelle macroscopiche

Lo studio dei materiali granulari ha fatto crescere l'interesse dei fisici teorici per i fondamenti della Meccanica statistica e per i molti strumenti e approcci che precedono la teoria degli *ensembles* di Gibbs valida all'equilibrio. Ovviamente questi approcci non erano mai stati dimenticati, essendo cruciali anche per la comprensione di alcuni fenomeni di forte non-equilibrio in fluidi molecolari, ad esempio per lo studio della turbolenza e della transizione vetrosa.

Ancora più forte è stato l'impulso dato dalla Fisica granulare a considerare nuovi sistemi

da porre sotto la lente della Meccanica statistica, senza più la necessità di considerare unità microscopiche e interazioni conservative. L'esempio più evidente di questa nuova direzione è lo studio della cosiddetta *materia attiva*. Anche se alcuni modelli seminali sono stati introdotti già negli anni '90 del ventesimo secolo, quando la Fisica statistica granulare era agli inizi, è opinione condivisa che la *materia attiva* sia storicamente un'erede della Fisica statistica granulare. Ne aumenta infatti la complessità e la ricchezza, oltre ovviamente a coinvolgere un numero molto maggiore di classi di sistemi a cui applicare la teoria.

La **materia attiva** comprende tutti quei sistemi composti da molte unità elementari che sono dotate di auto-propulsione [11]. Gli esempi più frequenti sono di natura biologica, in particolare un'ampia classe di materiali attivi sono insiemi di micro-nuotatori cioè di esseri viventi alla scala microscopica (dimensioni intorno al micron, come colonie di organismi unicellulari, batteri, spermatozoi etc.) che convertono energia chimica - ad esempio riserve di ATP - in movimento attraverso il fluido ospitante, principalmente tramite l'attuazione di flagelli o pseudopodi. Si trovano esempi di *materia attiva* biologica anche a scale più piccole (all'interno della cellula si possono trovare reti di filamenti attivi, ad esempio formati da actina, che tra le varie funzioni conferisce motilità alla cellula), e ovviamente a scale più grandi: colonie di insetti, banchi di pesci, stormi di uccelli e folle di uomini [12]. Esistono poi molti sistemi attivi non biologici, quali ad esempio i colloidali attivi, che sviluppano un'auto-propulsione dovuta a reazioni chimiche con il fluido circostante, o insiemi di (nano, micro e macro) robot, o persino gli stessi granulari, in certe configurazioni, sono a pieno titolo considerati sistemi attivi.

Chiaramente ognuno di questi esempi ha proprietà peculiari che rendono a volte profondamente diversi i fenomeni osservati. I biologi difficilmente considerano i sistemi elencati qui sopra come parte di un'unica classe. L'idea di creare una categoria così ampia è dovuta al potere di astrazione della Fisica teorica, che certamente si basa su una forte semplificazione delle caratteristiche dei sistemi, allo scopo di ridurre la complessità del reale in modelli maneggevoli per i



Figura 5: Alcuni esempi di materia attiva tratti dal mondo dei micro-nuotatori (come spermatozoi e batteri), dal mondo animale e dal mondo sociale.

calcoli e la comprensione qualitativa.

In questo articolo, ovviamente, non abbiamo l'obiettivo di raccontare la materia attiva con lo stesso dettaglio di quella granulare. Ma ci sembra interessante sottolineare alcuni punti in comune. Il primo elemento di continuità tra Meccanica statistica attiva e Meccanica statistica granulare è la rinuncia alla conservazione dell'energia, in questo caso ancora più radicale: una particella granulare libera (senza interazioni) infatti segue la Fisica delle molecole (cioè conserva l'energia o a limite può risentire di forze potenziali esterne, come la gravità); una particella attiva libera invece converte costantemente energia chimica in energia Meccanica, eventualmente dissipata dal fluido circostante, quindi ad esempio se una perturbazione la arresta temporaneamente, al termine della perturbazione la particella accelera e riprende il suo moto alla sua velocità caratteristica (batteri e spermatozoi possono correre decine o anche centinaia di micron al secondo). Questo fatto implica ovviamente che la materia attiva è intrinsecamente fuori dall'equilibrio, proprio come la materia granulare. Per alcuni materiali attivi, inoltre, si osserva un tipo di interazione che ricorda il meccanismo di allineamento della collisione anelastica tra le particelle granulari, anche se - nel caso attivo - può avvenire più morbidamente (ad esempio in maniera non istantanea, è raro infatti avere urti duri tra particelle attive) o addirittura a distanza, come capita per molti animali che interagiscono a vista. Le instabilità osservate nei fluidi granulari, dovute alla dissipazione negli urti, si osservano - in forme qualitativamente simili ma anche con molte differenze quantitative - anche nella materia attiva: in diversi modelli attivi si osserva la cosiddetta *motility induced phase separation* (separazione di fase indotta dalla motilità) che consiste nell'apparire di una coesistenza tra fase diluita e fase

densa all'aumentare del parametro di attività (la velocità o il tempo di persistenza dell'auto-propulsione), instabilità che appare analoga - ed è guidata da meccanismi simili - al cosiddetto *clustering* (aggregazione) osservato nei gas granulari; nei modelli attivi con interazioni di allineamento, è possibile osservare l'emergere di una fase fortemente polarizzata, in cui cioè tutto il sistema si muove coerentemente in una direzione, fase analoga all'instabilità di *shear* dei gas granulari in raffreddamento. Esiste poi una serie molto interessante di esperimenti su materiali granulari con particelle allungate o sbilanciate, sottoposti a vibrazione, tali per cui la singola particella ha una propria auto-propulsione, cioè converte l'energia della vibrazione esterna in moto proprio. Anche in questi sistemi è possibile riconoscere i fenomeni aggregativi e di polarizzazione menzionati poco fa. Un'ultimo notevole punto in comune tra sistemi attivi e granulari riguarda il piccolo numero di elementi in un insieme. Anche per un materiale attivo i numeri si aggirano sulle centinaia e migliaia di unità, e quindi le fluttuazioni sono rilevanti anche per la Meccanica statistica attiva. Le differenze tra materiali granulari e materia attiva riguardano le simmetrie in gioco, tipicamente le particelle auto-propulse hanno un asse privilegiato e possono essere polari (mentre i granulari sono spesso sferici), come anticipato le interazioni tra particelle attive sono tipicamente a distanza o mediate dal fluido, in alcuni casi possono addirittura essere topologiche (vedi il caso degli stormi di uccelli, in cui la distanza conta molto poco).

Molti degli approcci teorici basati su equazioni idrodinamiche per i campi rilevanti (i cosiddetti campi lenti, che nel caso dei sistemi attivi includono - tipicamente - anche il campo di polarizzazione), del tutto analoghi alle idrodinami-



Figura 6: Resti fossili di un banco di pesci della specie (estinta) *Erismatopterus levatus*, il cui comportamento collettivo è stato studiato - a partire proprio da questi resti - in un recente articolo pubblicato su *Proceedings of the Royal Society B* [14].

che granulari, sono stati sviluppati per i sistemi attivi [13]. La maggior varietà e complessità delle interazioni in gioco, oltre che la giovinezza della teoria, hanno privilegiato approcci fenomenologici, in cui cioè le equazioni idrodinamiche vengono scritte sulla base delle simmetrie e di argomenti di ragionevolezza, ma non sono dedotte dal modello microscopico, cioè dalle interazioni dei singoli elementi del sistema. Questo modo di scrivere le equazioni idrodinamiche ha due svantaggi fondamentali: non permette di avere (neanche a livello di stime) i valori dei coefficienti che compaiono nelle equazioni, e spesso è costretto a tener conto di molti termini non necessari. In alcuni casi, poi, queste equazioni vengono scritte ispirandosi a modelli di equilibrio, rischiando quindi di compiere degli errori concettuali. Un modello molto fortunato, e certamente fondamentale nella storia della Meccanica statistica attiva, è l'idrodinamica di Toner e Tu per la descrizione dei fenomeni di *flocking*. Questo modello è sostanzialmente una variazione del modello "XY" nella Meccanica statistica dei sistemi di spin. È chiaro che assimilare delle velocità, ad esempio di uccelli, a spin è un'astrazione estrema, ad esempio l'inerzia (importante al di fuori della classe dei micro-nuotatori) non viene considerata. In conclusione la sfida centrale, in questo momento, posta dalla Meccanica statistica attiva riguarda proprio la derivazione rigorosa delle teorie idrodinamiche, terreno sul quale i materiali granulari possono dare numerosi spunti e rappresentare una guida importante.



- [1] R. Munroe: *What Makes Sand Soft?*, The New York Times, p.4 10 Nov. (2020). <https://www.nytimes.com/2020/11/09/science/what-makes-sand-soft.html>
- [2] H. M. Jaeger, S. R. Nagel, R. P. Behringer: *Granular solids, liquids, and gases*, *Reviews of Modern Physics*, 68 (1996) 1259.
- [3] B. Andreotti, Y. Forterre, O. Pouliquen: *Granular media: between fluid and solid.*, Cambridge University Press, Cambridge (2013).
- [4] A. Baule, F. Morone, H. J. Herrmann, H. A. Makse: *Edwards statistical mechanics for jammed granular matter*, *Reviews of Modern Physics*, 90 (2018) 015006.
- [5] A. Puglisi: *Transport and fluctuations in granular fluids: From Boltzmann equation to hydrodynamics, diffusion and motor effects*, Springer, Berlino (2014).
- [6] C. Villani: *Mathematics of granular materials*, *Journal of Statistical Physics*, 124 (2006) 781.
- [7] A. Puglisi, V. Loreto, U. Marini Bettolo Marconi, A. Petri, A. Vulpiani: *Clustering and non-gaussian behavior in granular matter*, *Physical Review Letters*, 81 (1998) 3848.
- [8] A. Puglisi, A. Gnoli, G. Gradenigo, A. Sarracino and D. Villamaina: *Structure factors in granular experiments with homogeneous fluidization*, *Journal of Chemical Physics* (136) 2012. 014704
- [9] U. Marini Bettolo Marconi, A. Puglisi and A. Vulpiani: *About an H-theorem for systems with non-conservative interactions*, *J. Stat. Mech*, 8 (2013) P08003.
- [10] C. A. Plata, A. Prados: *Global stability and H theorem in lattice models with nonconservative interactions*, *Physical Review E*, 95 (2017) 052121.
- [11] C. Bechinger, R. Di Leonardo, H. Löwen, C. Reichhardt, G. Volpe, G. Volpe: *Active particles in complex and crowded environments*, *Reviews of Modern Physics*, 88 (2016) 045006.
- [12] A. Cavagna, I. Giardina, T. Grigera: *The physics of flocking: Correlation as a compass from experiments to theory*, *Physics Reports*, 728 (2018) 1.
- [13] M. C. Marchetti, et al.: *Hydrodynamics of soft active matter*, *Reviews of Modern Physics*, 85 (2013) 1143.
- [14] N. Mizumoto, S. Miyata, S. C. Pratt: *Inferring collective behaviour from a fossilized fish shoal*, *Proceedings of the Royal Society B*, 286 (2019) 20190891.



Andrea Puglisi: è dirigente di ricerca presso l'Istituto dei Sistemi Complessi del Consiglio Nazionale delle Ricerche, a Roma. Si occupa di Meccanica statistica di non-equilibrio, materiali granulari e materia attiva, con qualche appassionante deviazione verso la dinamica dei linguaggi.

Le nuove frontiere del volo spaziale: Nanotecnologie e Biomedicina

Giada Graziana Genchi

Istituto Italiano di Tecnologia, Pontedera

Gianni Ciofani

Istituto Italiano di Tecnologia, Pontedera

Il rinnovato interesse per l'esplorazione dello spazio ha evidenziato la necessità di una comprensione profonda dei rischi per l'equipaggio e delle opportunità commerciali insite nel volo, sia in bassa orbita terrestre che nello spazio profondo. L'esplorazione spaziale rappresenta una sfida complessa per gli organismi di origine terrestre, comportando per essi notevoli *stress* di natura fisica (transizioni gravitazionali, permanenza in microgravità ed assorbimento di radiazioni cosmiche), ma anche chimica, microbiologica e psicologica. Fortemente ostile alla vita, l'ambiente spaziale rappresenta inoltre un prezioso banco di prova di strumenti innovativi per il contrasto di numerosi fenomeni degenerativi associati all'invecchiamento, nonché all'insorgenza di malattie a Terra. Fra questi strumenti, quelli nanotecnologici iniziano a mostrare evidenze promettenti di un valido supporto alla vita nello spazio.

Il volo spaziale

Inizialmente sostenuta da ragioni di carattere geopolitico, l'esplorazione spaziale ha ormai assunto da diversi decenni la veste di strumento di espansione della conoscenza umana nello spirito della cooperazione internazionale e a beneficio dell'umanità, producendo numerose tecnologie con larga applicazione terrestre [1]. Con il riaccendersi dell'interesse all'esplorazione umana dello spazio profondo, alla colonizzazione di corpi celesti prossimi alla Terra e al turismo spaziale in bassa orbita terrestre (*low Earth orbit*, LEO) promosso da enti privati, è diventata di fondamentale importanza la comprensione profonda delle alterazioni biologiche che potrebbero interessare gli astronauti o i viaggiatori spaziali non professionisti. Altrettanto importante è l'elaborazione delle strategie che prevenivano tali alterazioni in particolar modo sul lungo termine, sia durante il volo, che potrebbe durare anni, che al rientro a Terra. Ma quali sono gli agenti nocivi per la vita umana nello spazio? Essi sono di natura prevalentemente fisica e chimica: i principali

rischi per la salute umana originano infatti dalla microgravità e dalle radiazioni cosmiche, ma anche dalle polveri fini a varia composizione come il regolith lunare e dalle popolazioni microbiche a bordo dei veicoli spaziali. Rischi secondari ma non trascurabili scaturiscono inoltre a livello biologico e psicologico (fra loro strettamente connessi) dal confinamento in ambienti angusti, dall'isolamento e dalla distanza dalla Terra.

Il volo spaziale comporta *in primis* per gli organismi terrestri l'esposizione a transizioni gravitazionali, di cui l'ipergravità (3-6g) rappresenta una fonte primaria benché temporanea di *stress*. Successivamente, esso si caratterizza per una permanenza in condizioni di riduzione del carico gravitazionale, come succede agli astronauti in orbita o potrebbe loro succedere per via dell'esposizione ad una gravità notevolmente inferiore a quella terrestre (ad esempio quella lunare, circa un 1/6 di quella terrestre, o marziana, circa un 1/3 di quella terrestre) [2].

La variazione della forza di gravità tuttavia si accompagna anche ad un'altra formidabile fonte di *stress* fisico per gli organismi viventi nello spazio: la radiazione cosmica [3]. Allontanandosi dalla magnetosfera terrestre, gli astronauti sono infatti esposti a ioni pesanti contenuti in raggi cosmici galattici e a particelle secondarie generate dall'impatto degli ioni pesanti con le strutture e gli schermi a bordo dei veicoli spaziali (che danno quindi luogo a spallazione), nonché a protoni ad alta energia prodotti da eventi solari. Di questa radiazione, la componente protonica risulta essere quella maggioritaria (85%) nell'ambiente intraveicolare spaziale, seguita dalle particelle alfa (14%) ed infine dagli ioni pesanti ed i loro prodotti di frammentazione. La complessità degli agenti fisici di *stress* attivi sul personale nello spazio ha impedito ad oggi una completa comprensione dei fenomeni biologici adattativi innescati dal volo spaziale ed ha inaugurato un ampio settore di ricerca volto alla riproduzione a Terra delle condizioni di microgravità e radiazioni cosmiche. Questa attività di ricerca mira a:

- 1) colmare l'ampia lacuna circa gli effetti di una esposizione prolungata nel tempo alla microgravità ed alla radiazione cosmica per gli organismi che si sono evoluti a Terra,
- 2) fornire elementi utili al volo interplanetario ed al futuro allestimento di una base lunare o

marziana.

Alterazioni indotte dallo spazio

Cosa comportano microgravità e radiazioni a livello biologico? Fra i principali effetti del volo spaziale a livello molecolare ed intracellulare, vi è, insorgenza di *stress* ossidativo, danni al DNA e disregolazione mitocondriale [4, 5]. A livello superiore a quello cellulare, il volo induce effetti deleteri a carico del sistema cardiocircolatorio, immunitario, muscolare e nervoso, determinando tra l'altro un aumento del rischio di insorgenza di tumori [6, 7].

Gli astronauti sono infatti solo parzialmente protetti contro la radiazione cosmica dal campo magnetico terrestre e sono quindi esposti a circa 100-200 mSv/anno, assorbiti in basse dosi dell'ordine dei μS /giorno. Gli effetti delle radiazioni sono strettamente legati al trasferimento di energia dalle particelle al corpo che esse attraversano, con dipendenza dalla carica, dalla massa e dall'energia cinetica delle particelle. La carica è in particolar modo determinante ai fini del trasferimento energetico e gli ioni pesanti che compongono la radiazione cosmica contribuiscono in maniera preponderante alla dose assorbita dal personale di bordo. Oltre alla radiazione cosmica, anche gli eventi solari, benché estremamente rari, rappresentano una minaccia importante per la salute degli astronauti [8].

Quali sono le alterazioni biologiche indotte dall'esposizione agli agenti di *stress* fisici che caratterizzano l'ambiente spaziale? A livello molecolare, il fenomeno distintivo dell'esposizione alle condizioni di volo spaziale consiste nell'instaurarsi di un elevato *stress* ossidativo. L'insorgenza di quest'ultimo è legata allo sbilanciamento fra produzione e smaltimento di specie chimiche altamente pro-ossidanti, fra cui rientrano i cosiddetti radicali liberi. Di tali composti, i più noti e studiati sono le specie reattive dell'ossigeno (*reactive oxygen species*, ROS) e dell'azoto, che hanno numerosissimi bersagli citoplasmatici, mitocondriali e nucleari [9].

I radicali liberi possono essere prodotti dall'interazione della radiazione con l'acqua di cui sono ricchi gli organismi viventi, determinando danno per ionizzazione indiretta delle macromolecole biologiche, fra cui il DNA. Quest'ultimo

può essere inoltre danneggiato per effetto della ionizzazione diretta prodotta dai raggi cosmici. Quando il danno al DNA consiste in rotture che interessano un singolo filamento, quest'ultimo può essere efficacemente risaldato da meccanismi endogeni di riparazione, ma quando le rotture colpiscono entrambi i filamenti, la riparazione può avvenire in modo deficitario e risultare in mutazioni stabili, riarrangiamento dei cromosomi, morte cellulare o insorgenza di tumori. I raggi cosmici a elevato *linear energy transfer* contenenti ^{12}C , ^{56}Fe , ^{20}Ne , ^{48}Ti sono i più efficaci nell'indurre tumori attraverso la ionizzazione diretta del DNA [10].

Strettamente correlata allo stress ossidativo è la disregolazione mitocondriale, identificata dalla riduzione dell'espressione di geni coinvolti nella fosforilazione ossidativa del mitocondrio e codificati da DNA nucleare, nonché dall'attivazione di meccanismi compensatori di geni codificati da DNA mitocondriale. Il mitocondrio è infatti la principale sede di produzione endogena di ROS, da cui deriva la sua maggiore vulnerabilità (anche a carico del suo DNA) all'insorgenza di stress ossidativo [11]. La microgravità e le radiazioni cosmiche svolgono un ruolo importante nella disregolazione mitocondriale, determinando la sovrapproduzione di ROS attraverso la perturbazione della catena di trasporto elettronico che ha luogo in questo organello [5].

A livello di organismo, il volo spaziale produce innanzitutto uno spostamento dei fluidi biologici verso la parte alta del corpo: tale spostamento risulta in un incremento nell'escrezione e nell'attivazione di meccanismi di compensazione da parte del sistema cardiocircolatorio. Ha quindi luogo il cosiddetto decondizionamento del sistema cardiocircolatorio, molto simile a quello legato all'invecchiamento, caratterizzato dalla riduzione del volume del sangue, dalla variazione della contrattilità cardiaca, dall'irrigidimento dei vasi arteriosi e dallo sviluppo della resistenza all'insulina [15]. Tali meccanismi confluiscono nella cosiddetta intolleranza ortostatica esperita dagli astronauti al rientro a Terra, spesso accompagnata da altri fenomeni quali tachicardia e sincope, contro cui recenti studi sembrano dimostrare l'effetto benefico dell'esercizio fisico

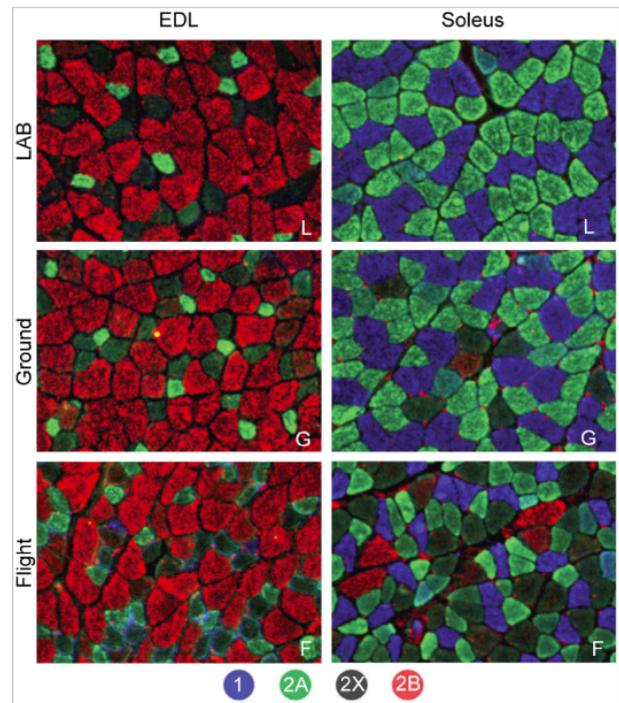


Figura 1: *Variazione della composizione molecolare delle fibre dei muscoli extensor digitorum longus-EDL e soleus di topi esposti a microgravità reale di lunga durata: in particolare, dopo il volo compare de novo nel soleus l'isoforma veloce 2B della miosina (colorata in rosso). L'immagine di immunostochimica evidenzia inoltre l'atrofia del muscolo antigravitazionale soleus dei campioni di volo rispetto a quelli di Terra attraverso la diminuzione delle dimensioni della sezione delle fibre identificate dalle isoforme 1 e 2A della miosina. I numeri corrispondono alle diverse isoforme della miosina, cui corrispondono le diverse tipologie di fibre muscolari. Da [14].*

a bordo dei veicoli spaziali [16].

Il sistema nervoso è negativamente influenzato dal volo spaziale: gli astronauti dimostrano infatti segni di deterioramento cognitivo, motorio e comportamentale che durante il volo possono avere conseguenze molto rischiose per la loro vita. Anche in questo caso, lo stress ossidativo sembra giocare un ruolo preponderante nell'insorgenza di tali alterazioni. I deficit neurologici nello spazio sono intrinsecamente legati alla disregolazione immunitaria indotta dal volo spaziale, che ne determina un inasprimento mediante fenomeni di natura infiammatoria [12].

Deficit motori si verificano durante il volo spaziale anche per la riduzione della massa muscolare e la conversione delle fibre muscolari da lente a veloci. L'atrofia muscolare colpisce in partico-

lare i muscoli preposti al mantenimento della postura (ad esempio: il soleus, come riportato in Figura 1 e può essere in parte contrastata con l'esercizio fisico [13, 14].

La riduzione del carico gravitazionale durante il volo spaziale determina inoltre una riduzione della densità ossea prossima all'1.5 % per ogni mese di volo, analogamente a quanto succede con l'invecchiamento femminile post-menopausa. Tale riduzione è il risultato della promozione di fenomeni coinvolti nel resorbimento osseo, che normalmente a Terra è bilanciato dalla deposizione ossea. Fra le strategie volte al contrasto di tale decremento osseo in cui un ruolo importante è altresì giocato dallo stress ossidativo, è risultata essere valida l'attività fisica e la somministrazione di antiossidanti, che però deve essere ripetuta nel tempo a causa della loro breve emivita plasmatica [17]. La disregolazione immunitaria già precedentemente menzionata consiste in una variazione della funzione di alcune cellule del sistema immunitario (cellule T e natural killer), nonché dell'espressione di alcuni mediatori della risposta immunitaria (citochine) [18]. La persistente infiammazione di moderata entità che ne consegue è alla base della riattivazione di virus latenti nell'organismo degli astronauti, con incremento del rischio di insorgenza di tumori [19].

Da questa breve rassegna degli effetti biologici del volo spaziale, appare evidente come approcci anche ibridi di esercizio fisico e di impiego di antiossidanti possano avere un ruolo benefico determinante per consentire una prolungata permanenza umana nello spazio.

Simulazione del volo spaziale

Gli elevati costi e gli stringenti vincoli del volo spaziale hanno motivato lo sviluppo di diverse piattaforme terrestri di simulazione dello stesso allo scopo di approfondire la natura dei fenomeni adattativi deleteri della salute dell'uomo e di altri organismi viventi, nonché di elaborare strategie utili alla prevenzione e al contenimento di tali fenomeni.

Fra i fattori ambientali di più difficile riproduzione a Terra, vi sono la microgravità e le radiazioni cosmiche. Il primo viene tradizionalmente riprodotto ruotando opportunamente i sistemi

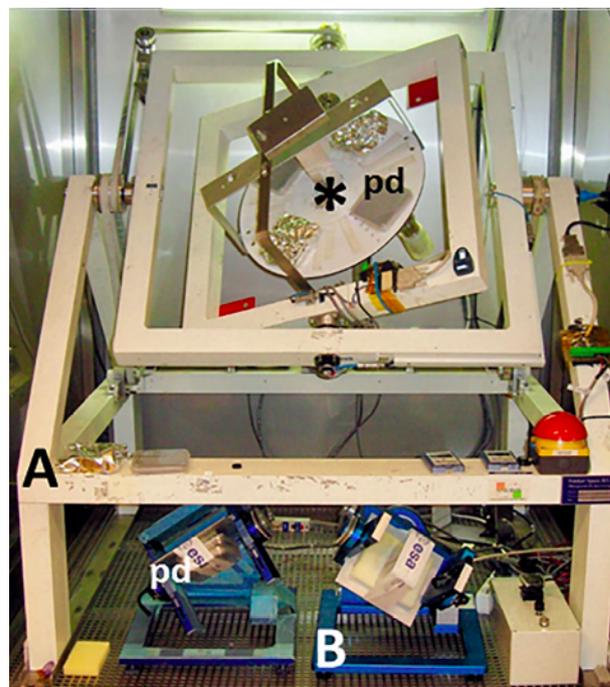


Figura 2: *Random positioning machine (A, B) per la simulazione della microgravità presso la European Space Agency-ESA. Da [20].*

biologici oggetto di investigazione (colture cellulari, fette di tessuto etc.) attraverso dispositivi denominati **clinostati**, **random positioning machine** o **rotating walled vessel** (Figura 2). Questi sistemi non annullano la forza di gravità terrestre, ma rendono semplicemente casuale la direzione di un corpo rispetto a tale forza nel tempo e sono particolarmente indicati per modelli biologici semplici come colture cellulari. Sistemi di più recente implementazione invece forniscono una forza contraria a quella di gravità terrestre attraverso la levitazione magnetica [21].

La necessità di condurre esperimenti su sistemi biologici di crescente complessità in elevato grado di sicurezza ha motivato l'elaborazione di strategie volte a mimare la microgravità a Terra con diverso livello di successo, fra cui la sospensione degli arti (per i *test* su animali) e il riposo a letto (per i *test* su soggetti umani). Ciascuna modalità di simulazione offre numerosi vantaggi e svantaggi, per cui l'attività sperimentale di Terra rimane ad oggi non esaustiva, ma utile a supportare e complementare l'attività sperimentale di volo.

I **clinostati** sono dei dispositivi che consentono la rotazione di un corpo lungo uno o più assi. Quando gli assi possono essere ruotati a diversa

velocità, allora i clinostati prendono il nome di **random positioning machine**: in entrambi i casi, i dispositivi si caratterizzano per l'elevata semplicità di impiego, richiedendo prevalentemente stabilizzazione del carico sulle cornici rotanti ed attenta valutazione della velocità di rotazione. Eventuali accelerazioni centrifughe residue rimangono dell'ordine dei 10^{-3} g [20, 22].

I **rotating walled vessel** consistono in due cilindri coassiali in rotazione orizzontale, ove il cilindro interno rappresenta la fonte di scambio gassoso con il cilindro esterno che invece ospita la coltura cellulare. Il disegno originario di questi dispositivi è stato condotto dalla NASA che ne ha curato in particolare la fluidodinamica, minimizzando sforzi e turbolenze e massimizzando l'ossigenazione delle colture. Essi sono particolarmente indicati per le colture cellulari in sospensione, ma sono di impiego meno immediato, richiedendo un lungo lavoro di ottimizzazione delle condizioni colturali più adeguate in termini di velocità terminale, di bilanciamento della forza di Coriolis e delle forze centrifughe e di trasferimento di massa [23].

I **sistemi di levitazione magnetica** si basano invece su solenoidi attraverso i quali sono applicati campi magnetici estremamente intensi (15 T) a dei corpi sfruttandone le proprietà diamagnetiche. Questi sistemi sono estremamente complessi e dispendiosi (richiedendo fra l'altro anche opportuni sistemi di raffreddamento) ed hanno trovato minore applicazione in ambito biologico, date le piccole dimensioni della cavità magnetica e l'entità dei campi magnetici applicati, potenzialmente in grado di interferire con la biologia del campione [24, 25].

Ben più difficile è la simulazione delle radiazioni cosmiche a Terra, che normalmente sono a bassa dose e con durata prolungata: in numerosi studi, per praticità è stato scelto di testare la somministrazione della stessa dose di radiazioni in condizioni acute. Tipicamente, sono impiegati raggi γ , raggi X, fasci di neutroni o a singolo ione monoenergetici. La radiazione cosmica si compone tuttavia di una molteplicità di ioni a diversa energia. Particelle cariche pesanti simili alla radiazione cosmica galattica in termini di energia e spettro sono state prodotte presso il *NASA Space Radiation Laboratory* (NSRL) al *Brookhaven National Laboratory* di New York. Presso questo

centro è possibile ottenere fasci ionici a rapido *switch*, maggiormente adeguati alla simulazione dell'ambiente spaziale, in particolar modo quello all'interno dei veicoli spaziali o attivo direttamente sull'organismo degli astronauti. Sussistono tuttavia ancora delle principali limitazioni nelle attività sperimentali che è possibile condurre presso questo centro. Il simulatore infatti produce solo alcuni degli ioni facenti parte della radiazione cosmica galattica, non esaurendone la varietà originaria e prodotta anche a seguito di spallazione, e non produce neanche pioni e neutrini che a loro volta caratterizzano la radiazione intraveicolare [26].

Nanotecnologie a supporto della vita nello spazio

Le attività umane in LEO richiedono numerosi approvvigionamenti da Terra, ma la dispendiosità e l'impraticabilità di questi ultimi in caso di esplorazione dello spazio profondo rendono indispensabile lo sviluppo di composti con proprietà stabili ed azione prolungata nel tempo. Alcuni preparati con dimensioni nanometriche studiati negli ultimi anni per il contrasto di numerose patologie geriatriche fra cui l'osteoporosi e la sarcopenia sono stati di recente proposti proprio quali contromisure -rispettivamente- alla perdita di massa ossea e muscolare che colpiscono gli astronauti già dopo un volo spaziale di breve durata [27, 28].

Questi fenomeni degenerativi associati all'invecchiamento sembrano essere infatti accelerati dalla permanenza in condizioni di microgravità ed esposizione a radiazioni, rendendo così la sperimentazione in volo utile all'elaborazione di strumenti terapeutici anche per la popolazione a Terra [29]. Per la loro elevata bioattività e biocompatibilità, nanoparticelle di idrossiapatite arricchite con stronzio sono state ad esempio testate come contromisure alla perdita di massa ossea nello spazio.

Nell'ambito del progetto NATO svolto durante la biomissione Futura dell'Agenzia Spaziale Italiana-ASI, queste nanoparticelle sono state somministrate a cellule staminali mesenchimali, promuovendo in condizioni di gravità terrestre la sovraespressione di geni marca-

tori del processo di maturazione (anche definito differenziazione) osteogenica, assieme all'attività di un marcatore del differenziamento (ovvero la fosfatasi alcalina) [27].

Le colture esposte a nanoparticelle hanno inoltre prodotto una matrice extracellulare più abbondante e mineralizzata. In microgravità simulata mediante *random positioning machine*, le nanoparticelle hanno protetto l'attività della fosfatasi alcalina, che si è mantenuta costante invece di andare incontro a declino come nel caso dei campioni non trattati con nanoparticelle. L'analisi condotta su colture esposte a microgravità simulata ha supportato la sperimentazione in condizioni di microgravità reale, ovvero a bordo della Stazione Spaziale Internazionale (*International Space Station*, ISS), ove le nanoparticelle hanno dimostrato di favorire la deposizione di matrice inorganica in idrossiapatite da parte delle colture.

Questo studio dimostra come le nanoparticelle di idrossiapatite arricchite fungano da adeguato vettore di ioni Sr^{2+} con elevato potenziale osteogenico utile alla rigenerazione tissutale in circostanze ambientali ostili alla vita. Prossimi studi saranno volti alla valutazione delle modalità di somministrazione di tali nanoparticelle ad organismi interi attraverso opportune formulazioni farmacologiche o alimentari. Nella prima parte di questo articolo, abbiamo inoltre visto come la fonte primaria di danno biologico indotto dall'ambiente spaziale sia rappresentata dallo stress ossidativo: per questo motivo, ogni innovazione introdotta nell'ambito dei composti ad azione antiossidante è estremamente desiderabile al fine di favorire la permanenza umana in condizioni di microgravità ed esposizione a radiazione cosmica. Nell'ambito del progetto NANOROS svolto durante la biomissione Vita di ASI, sono state somministrate nanoparticelle di ossido di cerio a colture muscolari in differenziazione a bordo dell'ISS [28].

Queste particelle redox-attive si caratterizzano per proprietà antiossidante autorigenerante che potrebbe contrastare efficacemente e in modo duraturo (a differenza di antiossidanti tradizionali quali le vitamine) lo stress ossidativo a livello cellulare e mitocondriale durante il volo spaziale di lunga durata. Le colture muscolari sono state sottoposte a studi di trascrizione ge-

nica che hanno rivelato come le nanoparticelle regolino favorevolmente l'espressione di marcatori coinvolti nei processi di invecchiamento, con potenziale attività protettiva contro lo stress ossidativo indotto dal volo spaziale. In particolare, è risultata essere significativa la regolazione della lamina e dell'istone H2fax, che cooperano strettamente in processi di riparazione del DNA a seguito di danno ossidativo.

La somministrazione di nanoparticelle a colture muscolari in volo è sembrata avere effetti opposti a quelli indotti dalla semplice permanenza in condizioni di microgravità, meritevoli di approfondimento a livello proteico. La stessa tipologia di particelle è stata impiegata in organismi interi per valutarne la capacità di protezione contro basse dosi di radiazioni: in particolare, sono state impiegate planarie della specie *Dugesia japonica* quale modello di rigenerazione tissutale, per via dell'abbondanza di cellule staminali. I plattelminti sono stati trattati con nanoparticelle e successivamente sono stati irradiati con una dose di 7 Gy (*rate*: 3 Gy/min) prodotta da una sorgente di raggi X con fascio da 15 MV. Dopo irradiazione, gli animali trattati con nanoparticelle hanno dimostrato un minor numero di cellule in apoptosi (morte cellulare programmata) ed una minore frammentazione del DNA rispetto agli animali non trattati. Al contempo, la somministrazione di nanoparticelle ha promosso una maggiore staminalità e capacità di rigenerazione tissutale negli animali, con potenziale rilevanza anche per organismi superiori e con applicabilità in ambiente spaziale [30].

Conclusioni

L'avversità dell'ambiente spaziale alla vita presuppone lo sviluppo di approcci sinergici e multidisciplinari alla prevenzione e alla cura di quei fenomeni degenerativi di natura complessa e cumulativa che attualmente preven- gono una prolungata permanenza umana nello spazio. L'esplorazione spaziale richiede certamente sforzi internazionali congiunti per lo sviluppo di metodologie e strutture raffinate di simulazione a Terra delle condizioni associate al volo, in modo tale che l'accesso allo spazio possa essere garantito, sia a personale altamente specializzato come gli astronauti che a potenziali turisti nel

prossimo futuro. In questo contesto, le nanotecnologie rappresentano una nuova frontiera della conoscenza umana che passa dall'infinitamente piccolo all'esplorazione dell'Universo attraverso una approfondita comprensione dei meccanismi molecolari che regolano la vita.

Le nanotecnologie applicate alla biologia e alla medicina permettono infatti di studiare e potenzialmente governare in modo specializzato dei bersagli di crescente complessità, da quella molecolare a quella dei compartimenti intracellulari, fino a quella di singoli distretti anatomici, sulla base delle dimensioni dei nanomateriali (compatibili con quelle cellulari) e delle loro proprietà chimico-fisiche, in alcuni casi controllabili mediante opportuna sorgente di stimolazione.

Affinché gli sforzi della comunità scientifica siano fruttuosi a tal proposito, vi è ormai consenso unanime sulla necessità di effettuare una transizione tempestiva ad una raccolta di dati omici (ovvero che interessino l'intero corredo genico, di trascritti, di proteine e di lipidi di una singola cellula o di diverse popolazioni cellulari) dagli esperimenti in microgravità e con esposizione a radiazioni cosmiche sia simulate che reali, nonché ad una sostenuta condivisione di questi dati in banche dati con libero accesso (come ad esempio: il GeneLab NASA).

Per affrontare la sfida della comprensione profonda degli effetti degli agenti di *stress* fisico nello spazio e di come sia possibile contrastarli, i ricercatori dovranno affrontare delle tematiche urgenti come l'interazione di breve e di lunga durata di nanomateriali con sistemi biologici modello (sia sani che tumorali, ottenuti innanzitutto in microgravità simulata e poi in quella reale), prima di approdare al loro impiego sull'uomo.

Dispositivi automatizzati in grado di ospitare un'ampia varietà di tipologie cellulari e di piccoli animali, somministrare nanomateriali in modo controllato e valutarne il destino dovranno essere sviluppati per garantirne sicurezza ed efficacia di azione in un ambiente ostile quale quello spaziale. Le strutture di Terra volte alla simulazione di microgravità e radiazioni cosmiche svolgeranno un ruolo ineludibile nel processo di disegno e validazione di dispositivi nanotecnologici a scopo biomedicale. La diminuzione dello stress ossidativo alla base di diversi processi degenerativi rimarrà una tematica prioritaria nella ricerca del

prossimo futuro, in grado di apportare numerosi benefici alla comunità astronautica e globale.

Ringraziamenti

Si ringraziano Agenzia Spaziale Italiana, European Space Agency e Fondazione Cariplo per il supporto finanziario ai progetti NANOROS, NOEMI, InterGravity, NanOxSpace e Nichoid in Space. Si ringraziano anche Kayser Italia S.r.l. ed Andrea Degl'Innocenti per il prezioso supporto all'attività sperimentale.



- [1] D. Dietrich et al.: *Applications of Space Technologies to Global Health: Scoping Review*, J. Med. Internet Res., 20 (2018) e230.
- [2] G. Clément: *Fundamentals of space medicine*, The Space Technology Library, Berlino (Springer). 2005
- [3] G. A. Nelson: *Fundamental space radiobiology*, Gravit. Space Biol. Bull., 16 (2003) 29.
- [4] P. Bradbury et al.: *Modeling the Impact of Microgravity at the Cellular Level: Implications for Human Disease*, Front. Cell Dev. Biol., 8 (2020) 96.
- [5] W. A. da Silveira et al.: *Comprehensive Multi-omics Analysis Reveals Mitochondrial Stress as a Central Biological Hub for Spaceflight Impact*, Cell, 183 (2020) 1185.
- [6] E. Afshinnekoo et al.: *Fundamental Biological Features of Spaceflight: Advancing the Field to Enable Deep-Space Exploration*, Cell., 183 (2020) 1162.
- [7] M. H. Barcelos-Hoff et al.: *Concepts and challenges in cancer risk prediction for the space radiation environment*, Life Sci. Space Res., 6 (2015) 92.
- [8] S. Hu et al.: *Modeling the acute health effects of astronauts from exposure to large solar particle events*, Health Phys., 96 (2009) 465.
- [9] S. K. Powers et al.: *Reactive oxygen species: Impact on skeletal muscle*, Compr. Physiol., 1 (2011) 941.
- [10] M. Moreno-Villanueva et al.: *Interplay of space radiation and microgravity in DNA damage and DNA damage response*, npj Microgravity, 3 (2017) 14.
- [11] C. Richter: *Reactive oxygen and DNA damage in mitochondria*, Mutat. Res. DNAGing, 275 (1992) 249.
- [12] G. R. Clément et al.: *Challenges to the central nervous system during human spaceflight missions to Mars*, J. Neurophysiol, 123 (2020) 2037.
- [13] R. H. Fitts et al.: *Microgravity and skeletal muscle*, J. Appl. Physiol., 89 (2000) 823.
- [14] D. Sandonà et al.: *Adaptation of Mouse Skeletal Muscle to Long-Term Microgravity in the MDS Mission*, PLOS One, 7 (2012) e33232.

- [15] S. Patel: *The effects of microgravity and space radiation on cardiovascular health: From low-Earth orbit and beyond*, Int. J. Cardiol. Heart Vasc., 30 (2020) 100595..
- [16] Q. Fu et al.: *Impact of Prolonged Spaceflight on Orthostatic Tolerance during Ambulation and Blood Pressure Profiles in Astronauts*, Circulation, 140 (2019) 729.
- [17] Y. Tian et al.: *The Impact of Oxidative Stress on the Bone System in Response to the Space Special Environment*, Int. J. Mol. Sci., 18 (2017) 2132.
- [18] B. Crucian et al.: *Immune system dysregulation during spaceflight: clinical risk for exploration-class missions*, J. Leukoc. Biol., 86 (2009) 1017.
- [19] S. K. Mehta et al.: *Latent virus reactivation in astronauts on the international space station*, npj Microgravity, 3 (2017) 11.
- [20] E. Z. Kiss et al.: *Comparison of Microgravity Analogs to Spaceflight in Studies of Plant Growth and Development*, Front. Plant Sci., 10 (2019) 1577.
- [21] R. Herranz et al.: *Ground-Based Facilities for Simulation of Microgravity: Organism-Specific Recommendations for Their Use, and Recommended Terminology*, Astrobiology, 13 (2013) 1.
- [22] S. L. Wuest et al.: *Simulated Microgravity: Critical Review on the Use of Random Positioning Machines for Mammalian Cell Culture*, Biomed. Res. Intern, 2015 (2015) 971474.
- [23] T. G. Hammond et al.: *Optimized suspension culture: the rotating-wall vessel*, Am. J. Physiol. Renal Physiol., 281 (2001) F12.
- [24] A. Paul et al.: *High magnetic field induced changes of gene expression in arabidopsis*, Biomagn. Res. Technol, 4 (2006) 7.
- [25] G. R. Souza et al.: *Three-dimensional tissue culture based on magnetic cell levitation*, Nat. Nanotechnol., 5 (2010) 291.
- [26] J. W. Norbury et al.: *Galactic cosmic ray simulation at the NASA Space Radiation Laboratory*, Life Sci. Space Res., 8 (2016) 38.
- [27] F. Cristofaro et al.: *The NATO project: nanoparticle-based countermeasures for microgravity-induced osteoporosis*, Sci. Rep., 9 (2019) 17141.
- [28] G. G. Genchi et al.: *Modulation of gene expression in rat muscle cells following treatment with nanoceria in different gravity regimes*, Nanomed., 13 (2018) 2821.
- [29] S. Amselem: *Remote Controlled Autonomous Microgravity Lab Platforms for Drug Research in Space*, Pharm. Res., 36 (2019) 183.
- [30] A. Salvetti et al.: *Stem cell and tissue regeneration analysis in low-dose irradiated planarians treated with cerium oxide nanoparticles*, Mat. Sci. Eng. C, 115 (2020) 111113.
- [31] S. Blue et al.: *Limitations in predicting radiation-induced pharmaceutical instability during long-duration spaceflight*, npj Microgravity, 5 (2019) 15.

Giada Graziana Genchi: (H-index: 21) è Post-Doc presso il laboratorio di Smart Bio-Interfaces dell'Istituto Italiano di Tecnologia-IIT, ove svolge la sua attività di ricerca in micro- e nanofabbricazione di materiali intelligenti, in ingegneria tissutale assistita da materiali intelligenti e DNA ricombinante e in biologia cellulare in condizioni di gravità alterata. è attualmente attiva nell'ambito dei progetti InterGravity e Nichoid in Space promossi dalla European Space Agency-ESA.

Gianni Ciofani: (H-index: 39) è Research Director del laboratorio di Smart Bio-Interfaces (SB-I). Attualmente coordina il Center for Materials Interfaces di IIT e svariati progetti nazionali ed internazionali in qualità di titolare (PI), fra cui due European Research Council Grant e numerosi progetti spaziali promossi dall'Agenzia Spaziale Italiana e dall'ESA, rivolti allo studio dei nanomateriali intelligenti, della nanomedicina, delle interazioni bio/non-bio e della biologia in gravità alterata.



Perché la Regina aveva torto, pur avendo ragione

Michele Grillo *Università Cattolica del Sacro Cuore di Milano*

Nella discussione avviata su **Ithaca**, in tema di confronto tra statuto scientifico della Fisica e delle altre discipline non può mancare l'Economia che, più di altre discipline sociali, afferma di guardare al modello della Fisica. La maggior parte degli economisti, pur riconoscendo che ci sono diversità tra il mondo fisico e quello della convivenza umana associata, non ha obiezioni alla definizione di Scienza offerta da Giampaolo Co' nel suo contributo su "Riduzionismo ed Olismo nelle scienze" [1]:

"Considero scopo della Scienza quello di comprendere il comportamento del sistema studiato ... Ritengo che si comprenda un sistema se si è capaci di descriverne il comportamento passato e prevederne il comportamento futuro. Queste due fasi ... sono essenziali ... per definire lo scopo dello studio scientifico".

L'Economia affida all'Econometria la capacità di elaborare descrizione del comportamento passato e previsione del comportamento futuro. Spesso si avvale anche di un ulteriore elemento di assimilazione alle scienze dure, che può essere espresso nella proposizione secondo cui "la scienza non conosce le proprie premesse". Non perché l'Economia abbia obiezioni di principio a una indagine sulle premesse (vi sono campi dell'analisi economica che se ne occupano), ma perché, alla fine della giornata, la sua preoccupazione è che le deduzioni logiche, anche da premesse che non

sono oggetto di analisi, siano formulate in maniera tale da essere asseguibili a un test empirico; di modo che, se non confutate, possano essere utilizzate come previsione.

Nello stesso numero di **Ithaca**, Ferdinando Boero [1] offre buone ragioni per diffidare di una estensione acritica della definizione di Scienza data da Giampaolo Co' a discipline diverse dalla Fisica. Boero utilizza con arguzia l'economista come lo scemo del villaggio, facendone l'esempio negativo del proprio argomento ("Dove ci manderebbe uno storico se gli chiedessi di prevedere il futuro dei sistemi che studia? Soprattutto con formule matematiche? Beh, risponderebbe: non sono mica un economista!").

Chi scrive è un economista che vorrebbe essere più severo di Boero nei confronti dello statuto epistemologico oggi prevalente nella propria professione. Accettando l'invito su **Ithaca**, mi auguro di offrire, alla perplessità di Boero, ragioni anche più gravi delle sue. Peraltro, non vorrei limitarmi solo a considerazioni negative, ma vorrei anche mostrare che, senza rinnegare i suoi risultati, l'Economia può formulare con coerenza domande ambiziose e suggerire risposte importanti, per la cui comprensione non è necessario imitare la Fisica. Al contempo temo che il mio resoconto non sarà in grado di rispondere alla domanda "ma che Scienza è allora l'Economia?". In altre parole, sono consapevole di imbarcarmi su **Ithaca** un po' come clandestino. Spero però di pagare il biglietto con qualche curiosità verso elementi, oggi a mio avviso più in

ombra, di una disciplina proteiforme.

Anni fa la Regina Elisabetta II rimproverò gli economisti: "Come avete fatto a non prevedere la crisi?". Era nel diritto del suo ruolo prendere per buono ciò che la professione degli economisti raccontava di sé e, per questo, aveva ragione. Minori ragioni ebbero coloro che si affrettarono a controbattere che il rimprovero non era fondato, che c'era ben stato chi la crisi l'aveva prevista. Coloro che avevano previsto la crisi (non furono pochi e grande ne ebbero merito) non avevano affatto utilizzato, dal punto di vista logico e metodologico, l'analisi economica alla luce di un paradigma di descrizione di comportamenti passati per la previsione di comportamenti futuri. Avevano piuttosto anticipato che proposizioni scientificamente consolidate della teoria economica non avrebbero offerto adeguata comprensione delle novità introdotte dalla innovazione finanziaria e del modo in cui individui e società vi avrebbero fatto fronte. Un chiaro esempio è offerto dalla riabilitazione, dopo la crisi, di Hyman Minsky, la cui teoria dell'instabilità finanziaria, proposta all'inizio degli anni Ottanta, era stata a lungo considerata, dall'Economia *mainstream*, poco più di una stravaganza. A usare scrupolosamente, dal punto di vista logico e metodologico, l'analisi economica, erano stati invece proprio coloro che avevano escluso una crisi dietro l'angolo. Tra costoro c'erano peraltro i grandi sacerdoti dell'economia come scienza che "descrive il comportamento passato e prevede il comportamento futuro". Un nome per tutti è il premio Nobel Robert Lucas, a sua volta interprete ed erede di Milton Friedman. Negli anni Sessanta, Friedman aveva dato un fondamentale contributo per esplicitare i presupposti metodologici che avrebbero permesso di annoverare l'economia tra le scienze dure (cioè: (i) che l'economia non conosce le sue premesse; e (ii) che le deduzioni logiche da quelle premesse devono essere formulate in modo da essere confutabili).

Alla conoscenza scientifica non chiediamo una soddisfazione estetica, le chiediamo di aiutarci a plasmare il mondo in cui viviamo. Affinché tale obiettivo abbia senso, dobbiamo però riuscire a individuare una qualche separazione tra ricercatore e oggetto della ricerca. Non ne sto escludendo l'interazione. La Fisica sa che l'esito di un'osservazione è influenzato in qualche

misura dalla presenza dell'osservatore. La consapevolezza di questa influenza rende più complessa la comprensione, e la previsione, del comportamento del sistema, ma non la trasforma in un'aporia: è comunque sempre possibile assumere che esista un sistema studiato che è oggetto dell'osservazione. Boero ci mette in guardia sul fatto che la complessità dei sistemi che studiamo può essere grandemente vasta. E che, in fondo, la Fisica si confronta con sistemi semplici, mentre altre scienze si confrontano con sistemi ben più complessi. Boero sembra suggerire che sia proprio questa complessità a ostacolare l'applicazione a quei sistemi della definizione di Scienza come capacità di "descrivere il comportamento passato e prevedere il comportamento futuro". Prima di suffragare l'argomento con riflessioni suggeritegli dalla Biologia, Boero evoca però un altro sistema complesso, diverso dalla Biologia. Poiché in esso operano elementi non estranei alla riflessione dell'economia, mi ci soffermo brevemente. Boero suggerisce infatti di guardare alla storia e, chiedendosi: "Possiamo prevedere il corso futuro della storia?", aggiunge:

"Certo, gli economisti ci provano con i loro modelli. Ma abbiamo visto con quale accuratezza. La storia è fatta di contingenze (la sottolineatura è mia), altrimenti ... non ci sarebbe storia. Basterebbe un bel sistema di equazioni, e il futuro verrebbe fuori da lì. Ma se riuscissimo ad avere ... quelle equazioni, probabilmente ci comporteremmo in un certo modo, e influenzeremmo il sistema, inficiando le nostre previsioni".

L'assunto di Boero della Storia come sistema mi lascia perplesso ma, per altro verso, può facilitare l'argomento che voglio mettere in rilievo. La mia perplessità è, innanzi tutto, la seguente. So che esistono approcci metodologici che guardano con questa prospettiva alla Storia (a grandi linee, si può pensare alla storiografia marxista). Però in prevalenza la Storia non ha di sé la concezione di essere un sistema (per quanto complesso) del quale descrivere il comportamento passato, meno che mai come presupposto per una previsione del comportamento futuro. L'etimologia della parola suggerisce piuttosto il con-

trario. La radice di *historia* è $\iota\delta$, il cui significato (anche se pressoché perduto nell'italiano corrente, a parte forse idiosincratico) vuole cogliere un attributo qualitativo proprio di ciascun singolo individuo che Musil evoca in "L'uomo senza qualità". La storia è individualità, irriducibile individualità: esperienza individuale dell'osservatore (che vede - e infatti anche il verbo vedere ha la stessa radice); e, soprattutto, dell'oggetto dell'osservazione. È alla composizione di un variegato insieme di elementi individuali che chiediamo di raccontarci una storia; che vogliamo maestra di vita, ma sapendo bene che "non si ripete".

Alcuni economisti, peraltro in prevalenza non marxisti, ci hanno messo del loro per confondere le acque. Chiamandola con il nome della musa, forse per evitare il rischio di fare i conti con l'etimologia, hanno messo al mondo una *Cliometria*. La *Cliometria* assume il resoconto degli eventi passati e ne espunge tutti gli elementi individuali per trattare il residuo come una regolarità. Se l'operazione ha un senso, l'argomento di Boero sulla complessità perderebbe mordente (la riduzione della complessità sarebbe parte del metodo di analisi) e la strada per concepire la storia come un sistema, di cui comprendere il comportamento passato per prevedere il futuro, sarebbe facilitata.

A me sembra, invece, che ci sia una difficoltà intrinseca, che ha poco a che fare con la complessità, nel ricondurre le discipline che hanno come oggetto di studio la società umana a un paradigma scientifico che si propone di descrivere per prevedere. Innanzi tutto, in Economia, non mi sembra di poter cogliere una complessità che ostacolerebbe l'accoglimento della definizione di Scienza offerta da Co'. Gli ambiti dell'Economia sono molto più ristretti di quelli di una Storia universale. La loro definizione è abbastanza semplice. Facendo leva su un approccio di individualismo metodologico, l'Economia analizza le decisioni di produzione e di consumo dei singoli individui, assumendole guidate da una razionalità allo scopo. Le decisioni di produzione e di consumo si attagliano facilmente al comportamento individuale (sono la maggior parte delle preoccupazioni di Robinson Crusoe nella sua isola deserta). Poi 'uscendo dall'isola' l'Economia vuole capire come si comporta l'ar-

ciipelago (cioè, la società) nel suo complesso, in base a come gli individui interagiscono tra di loro. È qui che, a mio avviso, si annida invece la difficoltà a estendere all'Economia il paradigma di una Scienza che si propone di descrivere per prevedere. L'interazione tra i individui si regge infatti su istituzioni che gli stessi individui si sono dati (per taluni, queste istituzioni sono l'esito di una riflessione costruttiva; per altri, l'esito di una evoluzione spontanea - il tema è importante, ma non ho modo di svilupparlo in questa sede). Ciò che mi preme mettere in evidenza è invece che, in entrambi i casi, quelle istituzioni non piovono da un'altura esogena, ma sono il risultato dello stesso comportamento degli individui, cioè proprio di ciò che l'Economia si propone di studiare.

Se, per un verso, guardiamo alla conoscenza scientifica non per ragioni estetiche, ma affinché ci aiuti a plasmare il mondo in cui viviamo e, per altro verso, chiediamo all'Economia come Scienza di formulare leggi che descrivano il sistema sociale in base al comportamento degli individui, al fine di prevederne (prima) ma per condizionarne (poi) un comportamento che, se tali leggi esistono, non potrebbe esserne indipendente, la sola proposizione scientifica cui l'Economia può pervenire, per evitare un ricorso logico all'infinito, è che "viviamo nel migliore dei mondi possibili". Non perché nel mondo in cui viviamo riscontriamo (in base a una valutazione, per dirla con Hume, fondata su un *ought* diverso da un *is*) particolari caratteristiche di desiderabilità, ma perché il mondo in cui viviamo non può che essere compreso da noi che come il solo mondo possibile. Voltaire aveva ben chiara l'inevitabilità di questa conclusione. Per renderla evidente a tutti costruì il personaggio di Pangloss. E, per mostrarne l'insipienza, lo accoppiò a un aristotelismo di accatto: ogni cosa che esiste al mondo ha una sua funzione; e, infatti, abbiamo il naso per reggere gli occhiali.

Tra la caduta del muro di Berlino e la crisi finanziaria che la Regina Elisabetta II ha rimproverato agli economisti di non aver previsto, l'Economia ha vissuto un'età dell'oro di un paradigma scientifico che l'aveva portata a concludere, come Pangloss, che "viviamo nel migliore dei mondi possibili". Quel paradigma scientifico la portò a convincersi, e a diffondere la tesi, che

non aver saputo cogliere l'ineludibilità di questa conclusione era stata la causa di una sbornia di costruttivismo istituzionale che aveva coinvolto, per ampia parte del secolo XX, non solo la discussione accademica, ma anche progetti politici che sulla analisi economica facevano grande affidamento e che però erano ineluttabilmente condannati al fallimento. Ad affermare che, solo quando tutto questo fu compreso, diventava facile vedere che i problemi erano risolti (come affermò Lucas nel 2003, sia pure specificamente in tema di controllo del ciclo economico). Ciò che mi preme sottolineare è che, nelle sue tendenze dominanti, l'Economia ha rivendicato queste conclusioni proprio rivendicando per sé l'ambizione di avere, a imitazione della Fisica e più di altre discipline sociali, lo scopo di descrivere il comportamento passato del sistema economico per prevederne il futuro. Questa premessa metodologica è stata un ingrediente essenziale nel portarla a prevedere che il mondo in cui viviamo è il migliore dei mondi possibili; e a elaborare, su queste premesse, nozioni come tasso naturale di disoccupazione; o a suggerire a Margaret Thatcher l'acronimo T.I.N.A. (There Is No Alternative). Naturalmente questa prospettiva non ha portato a negare l'evidenza della crisi, quando questa si è manifestata nel 2008. Ha portato però a imputarla (come le precedenti, meno devastanti) a uno *shock* esogeno. Detto altrimenti, la crisi del 2008 semplicemente non era prevedibile: se lo fosse stata, non si sarebbe concretizzata.

Provo a riassumere. Il paradigma della Fisica conduce l'Economia a conclusioni inutili oppure, per chi è restio a convincersi che viviamo nel migliore dei mondi possibili, a interpretazioni fuorvianti. Ma non a causa di una particolare complessità del sistema, bensì perché, proponendosi di descrivere il comportamento passato di un sistema per prevederne il comportamento futuro, lo studio che ha come contenuto il comportamento degli individui e della società in cui essi vivono non può non includere nel suo oggetto il comportamento degli individui e della società che utilizzano la comprensione dei propri comportamenti e la previsione degli esiti associati a quei comportamenti per plasmare il mondo in cui vivono.

Una prospettiva diversa

L'Economia ha cominciato a riflettere esplicitamente sul proprio statuto metodologico a partire dalla metà del secolo XIX. Una delle prime evidenze la troviamo in un passo dei *Principles of Political Economy* in cui John Stuart Mill afferma che a determinare i prezzi dei fattori produttivi (la cui somma forma il prezzo naturale di ogni bene scambiato nel mercato) possono contribuire due principi: la concorrenza e la consuetudine. Tuttavia, "solo attraverso l'applicazione del principio di concorrenza l'economia politica può pretendere di caratterizzarsi come Scienza". In altri termini, l'Economia è Scienza solo subordinatamente a un particolare contesto istituzionale di interazione tra gli individui.

La specificazione non è irrilevante. Nel ritenere che l'Economia possa descrivere scientificamente il comportamento passato del sistema economico per prevederne il comportamento futuro solo condizionatamente a uno specifico assetto istituzionale (il mercato concorrenziale), l'Economia come Scienza è per John Stuart Mill ancillare a uno scopo gerarchicamente superiore, al quale però non è chiaro quali contributi possa offrire la Scienza. In effetti, a metà del secolo XVIII, l'Economia politica si era staccata dalla Filosofia politica e morale portandosi dietro una diversa finalità. Il tema dell'Economia politica è la libertà o, meno pomposamente, l'autonomia dell'individuo (peraltro in un ambito limitato alla divisione sociale del lavoro). Al timore che affidarsi all'autonomia dei singoli soggetti sociali avrebbe portato a una anarchia inefficiente, a un contesto di guerra di tutti contro tutti dove la vita dell'uomo sarebbe stata "povera, odiosa, brutale e breve" (Hobbes, *Leviathan*, XIII, 9), l'Economia politica oppose che l'esito sociale dell'interazione concorrenziale tra soggetti economici è efficiente. A dire il vero, la prospettiva analitica si complicò presto. Per gli economisti classici era condizione sufficiente per l'operare della concorrenza che ciascun soggetto economico fosse libero di entrare e uscire dal mercato (cioè, che il potere politico non interferisse con le scelte private). Gli economisti neoclassici del secolo XIX sostennero invece che è possibile garantire che la concorrenza produca gli effetti desiderati solo se è perfetta, cioè

se nessun soggetto economico gode di potere di mercato (analiticamente: un soggetto non ha potere di mercato se, modificando la propria domanda o la propria offerta di ciascun bene, non è in grado di influire sul prezzo al quale quel bene è scambiato nel mercato).

Il modello di concorrenza perfetta descrive un mondo ideale che, però, si prestò molto bene alle ambizioni scientifiche dell'Economia. In concorrenza perfetta il comportamento del sistema può essere rappresentato mediante un equilibrio (più specificamente: un vettore di prezzi e un vettore di quantità dei diversi beni, simultaneamente domandate e vendute a quei prezzi) associato a ogni assetto dei fondamentali del mercato (i gusti dei consumatori e le tecnologie per produrre i beni). L'equilibrio si modifica quando i fondamentali si modificano. Entrando più in dettaglio, nel modello di concorrenza perfetta nessun soggetto è in grado di influire sui prezzi, così da incidere su come ogni altro soggetto affronta il proprio problema di scelta. In questo contesto, i singoli individui sono in grado di associare a ogni propria azione una unica conseguenza (che, in condizioni di certezza, consiste di un singolo elemento; mentre, in un mondo incerto, in cui intervengono fattori aleatori purché indipendenti dai comportamenti dei soggetti sociali, può essere comunque rappresentata mediante una unica, ben determinata, lotteria). Sotto tali ipotesi di comportamento razionale (che gli economisti chiamano, nel loro gergo, parametrico), è possibile individuare una unica soluzione razionale al problema di scelta di ciascun individuo. Su queste premesse, la teoria economica esplicita le condizioni alle quali esiste un equilibrio; le condizioni, più restrittive, alle quali l'equilibrio è stabile (le nozioni di equilibrio e di stabilità sono interamente prese a prestito dalla Meccanica) e assicura che l'equilibrio, se esiste, è efficiente (Primo Teorema dell'Economia del Benessere).

Studiando i modelli di concorrenza perfetta, gli economisti si sono principalmente preoccupati analiticamente dell'esistenza di (almeno) un equilibrio; altrimenti come prevedere prezzi e quantità se l'esistenza dell'equilibrio non era assicurata? L'ipotesi che, affinché la concorrenza fosse perfetta, nessun soggetto doveva godere di alcun potere di mercato destava una preoccupazione di natura differente. A parte (forse) il

caso del monopolio, si intravedevano mondi in cui molte cose potevano essere possibili; peraltro, anche a causa della mancanza di una teoria che consentisse di analizzare rigorosamente il comportamento di individui razionali dotati di potere di mercato. Per molto tempo, la risposta consolatoria fu che le imperfezioni della concorrenza dovessero essere trattate come l'attrito in Fisica. Ma già dalla prima metà del secolo XIX era chiaro a molti (che si rifacevano all'economista francese Cournot) che l'approccio andava invertito: che era la concorrenza perfetta che doveva essere interpretata come l'esito dell'operare di un meccanismo concorrenziale che, però solo a certe condizioni, eliminava il potere di mercato dei soggetti economici.

Molta letteratura provò a rimediare all'ipotesi di un contesto ideale, caratterizzato da assenza di potere di mercato, avanzando congetture disparate per descrivere il comportamento razionale di individui dotati invece di potere di mercato. Il Gotha della scuola economica, preoccupato di difendere lo statuto scientifico dell'economia, si diede allora da fare per allontanare lo spettro che tutto, e pertanto nulla, si potesse dire. Sfruttando il presupposto metodologico che "la scienza non conosce le proprie premesse" Milton Friedman inferì che, anche se non è vero che nella realtà l'interazione nel mercato avvenga tra soggetti che non godono di alcun potere di mercato, e forse non è neppure vero che gli individui si comportino in modo rigorosamente razionale, alla fine della giornata l'ipotesi che i mercati reali producono esiti simili a quelli dei mercati ideali della concorrenza perfetta non può essere scientificamente confutata; per lo meno in contesti nei quali una società, per rimediare alle discrepanze, non si lasci abbindolare da ingenua velleità di costruttivismo istituzionale.

A partire dagli anni Sessanta fu però possibile, con la Teoria dei Giochi, disporre di una teoria della razionalità strategica. Ciò consentì all'analisi economica di analizzare il comportamento razionale di soggetti che, dotati di potere di mercato, condizionano con le proprie scelte le conseguenze che i comportamenti degli altri soggetti hanno sul rispettivo benessere. Ne emerse un quadro inatteso. La nozione di equilibrio (presa a prestito dalla Meccanica) si rivelò inadeguata: per la Teoria dei Giochi, una situazione sociale

è in equilibrio quando nessun soggetto sociale ha incentivo a modificare il proprio comportamento, dati i comportamenti degli altri soggetti. La nozione di stabilità dell'equilibrio (la capacità del sistema di tornare all'equilibrio se sottoposto a uno *shock* esogeno) si rivelò inservibile e fu ignorata. Ma, soprattutto, nell'analizzare un sistema sociale nel quale gli individui sono guidati da razionalità parametrica (in cui le conseguenze dell'azione ricadono solo sul soggetto agente), la principale preoccupazione del ricercatore era l'esistenza di un equilibrio. Invece, in condizioni di razionalità strategica la preoccupazione si capovolge: il problema diventa quello di una elevata molteplicità degli equilibri.

Se si escludono esempi semplici, che hanno solo uno scopo didattico, gli equilibri di un gioco sono in genere numerosi, spesso innumerevoli. La molteplicità degli equilibri si presenta in modo esplosivo quando sono coinvolti aspetti distributivi. Un primo esempio è quello di due bambini che devono dividere una torta, ciascuno nominando la quota che vuole per sé: se la somma delle pretese è maggiore della torta, nessun bambino avrà nulla; altrimenti ciascuno avrà la quota che ha richiesto; ogni coppia di numeri reali t_1, t_2 tali che $t_1 + t_2 = 1$ è un equilibrio del gioco. Un secondo risultato (noto come *folk-theorem* del 'dilemma del prigioniero' ripetuto indefinitamente) è più sofisticato: si consideri un contesto di interazione strategica nel quale esiste una situazione sociale di equilibrio, s' , che attribuisce ai due giocatori un profilo di utilità $[U_1(s'), U_2(s')]$. La situazione s' , però, è dominata da almeno un'altra situazione del gioco, nella quale entrambi i giocatori stanno meglio. Rappresentando il gioco con l'aiuto di un piano cartesiano con ordinate U_1 e U_2 , se l'interazione tra i giocatori si ripete indefinitamente, la ripetizione indefinita di qualsiasi situazione del gioco che attribuisce ai due giocatori un profilo di utilità che si colloca nel quadrante di nord-est rispetto al punto $[U_1(s'), U_2(s')]$ rappresenta un equilibrio.

È difficile, in queste condizioni, inquadrare l'Economia nella definizione di Scienza offerta da Giampaolo Co', come descrizione a fini di previsione. È possibile darsi una ragione a posteriori del perché è successo quello che è successo, del perché è emerso un certo equilibrio (l'analisi economica delle istituzioni tende a in-

terpretare gli assetti istituzionali come differenti equilibri di un gioco), ma non si capisce cosa possa volere dire prevedere il futuro, in che modo affermare che in futuro, a quelle stesse condizioni, succederebbe la stessa cosa. Detto altrimenti, non è chiaro cosa possa voler dire sottoporre la teoria a una confutazione nel senso di Popper. Può essere di interesse osservare, a questo riguardo, che negli ultimi 30 anni un approccio microeconomico ha cercato di far progredire l'analisi di falsificazione in economia. Implicitamente o esplicitamente, l'attenzione si è focalizzata sulla capacità di prevedere i comportamenti individuali a fronte di variazioni dell'ambiente o delle scelte politiche, in ambienti istituzionali dati. Questo richiama alla mente l'intuizione di John Stuart Mill, secondo cui è possibile prevedere il comportamento dei prezzi solo subordinatamente a un contesto istituzionale di interazione concorrenziale tra gli individui. Centocinquanta anni dopo Mill sentiamo però il bisogno di fare analisi economica delle istituzioni. Ma le cose si complicano notevolmente quando vogliamo interrogarci sugli effetti di discontinuità del disegno istituzionale (che è però proprio quello a cui ci sfidano le crisi).

Diversamente dalla preoccupazione di Boero, le difficoltà descritte hanno a mio avviso poco a che vedere con un problema reso complesso da innumerevoli contingenze; la loro origine è in elementi intrinseci negli stessi presupposti della razionalità umana. Mentre il modello di concorrenza perfetta si era rivelato compatibile con un modello meccanico del funzionamento della società - e ciò aveva permesso a Milton Friedman e ai suoi eredi di fondare la propria analisi su una metodologia compatibile con la definizione di scienza richiamata sopra - la necessità di fare i conti con una razionalità strategica, lungi dal potere essere trattata come un attrito, costringe l'analisi economica ad aprirsi a una vasta molteplicità di equilibri relazionali.

Nel 1994, John Harsanyi e Reinhard Selten sono stati insigniti congiuntamente del premio Nobel per l'Economia. Il Nobel fu loro riconosciuto per contributi separati e distinti, benché entrambi nel campo della Teoria dei Giochi (quell'anno il Nobel fu assegnato anche a John Nash). Meno noto è forse che, per un lungo tempo, Harsanyi e Selten lavorarono su un ambizioso progetto co-

mune che alla fine però abbandonarono: una *tracing procedure* in grado di selezionare sempre un unico equilibrio in qualsiasi gioco che ammette una molteplicità di equilibri.

Se avesse avuto successo, la ricerca congiunta di Harsanyi e Selten avrebbe aperto una via per ricondurre a unità comprensione e previsione nelle discipline sociali. Non credo che l'abbandono del progetto sia da imputare a difficoltà tecniche. Lo vedo piuttosto come segnale della necessità di lasciar cadere un modello analitico - lasciatici in eredità, talvolta anche con ricadute proficue, da mondo classico ed età della ragione - che, per evitare che il processo di conoscenza scientifica si avviti in un circolo infinito, si costringe a fare leva sul presupposto immotivato (pseudoscientifico, per Isaiah Berlin) di un determinismo della natura umana da cui far discendere un determinismo delle forme in cui si risolve l'interazione sociale.



[1] F. Boero, G. Co': *Riduzionismo e Olismo nelle Scienze*, Ithaca Educational, II (2020) 5.



Michele Grillo: è professore di Economia Politica presso la Facoltà di Giurisprudenza dell'Università Cattolica del Sacro Cuore di Milano. I suoi interessi di ricerca concernono la teoria della concorrenza imperfetta, la disciplina della concorrenza e del mercato, la teoria economica dell'organizzazione e l'analisi economica delle istituzioni democratiche. Tra il 1997 e il 2004 è stato Componente dell'Autorità Garante della Concorrenza e del Mercato.

Dante e l'ipersfera

... o se del mezzo cerchio far si puote
triangol sì ch'un retto non avesse ...
(Divina Commedia, Paradiso XII, 101-102)

Dante Alighieri

Giuseppe De Cecco

già membro del Dipartimento di Matematica e Fisica "Ennio De Giorgi", Unisalento

Quest'anno, come è noto, ricorrono i settecento anni dalla morte di Dante Alighieri (1265-1321). Qui si vuole ricordare la sua visione dell'Universo, collegandola alla Geometria che, come egli dirà nell'ultimo canto della Divina Commedia, è scienza che più di ogni altra può avvicinarsi a una pallida rappresentazione degli attributi trinitari:

"Qual è 'l geometra che tutto s'affige
per misurar lo cerchio, e non ritrova,
pensando, quel principio ond'elli indige,"
(Par XXXIII, 133-135)

Non si parlerà invece del collegamento con la Fisica (in particolare con la Teoria della Relatività), che pure ha suscitato interesse in studiosi [1].

Dante

Dante vuole conciliare la cosmologia aristotelica, che pone la Terra al centro, con la visione cristiana, che pone al centro Dio, quindi cercare di conciliare la visione geocentrica con quella teocentrica, il visibile con l'invisibile, la materia con lo spirito. Questo tentativo, secondo alcuni studiosi, sembra che abbia portato Dante a superare la geometria euclidea concependo un universo curvo. A mio avviso, questa visione geometrica rimane però una interpretazione di

una mirabile intuizione poetica, che in ogni caso ci sorprende.

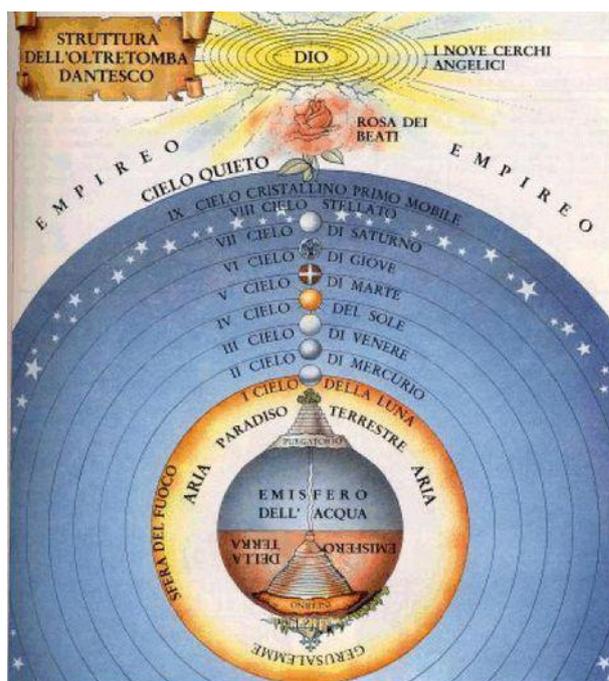


Figura 1: Struttura dell'oltretomba dantesco.

Chi per primo ha individuato nell'opera dantesca la geometria non-euclidea è stato il matematico (scienziato, filosofo e prete ortodosso, il "Leonardo russo") Pavel Alexandrovič Florenskij (1882-1937)¹ che nel 1921, proprio in occasione dell'anniversario della morte di Dante, nel-

¹Per una breve biografia e per una sintesi della sua visione della matematica si può consultare la referenza. [2]

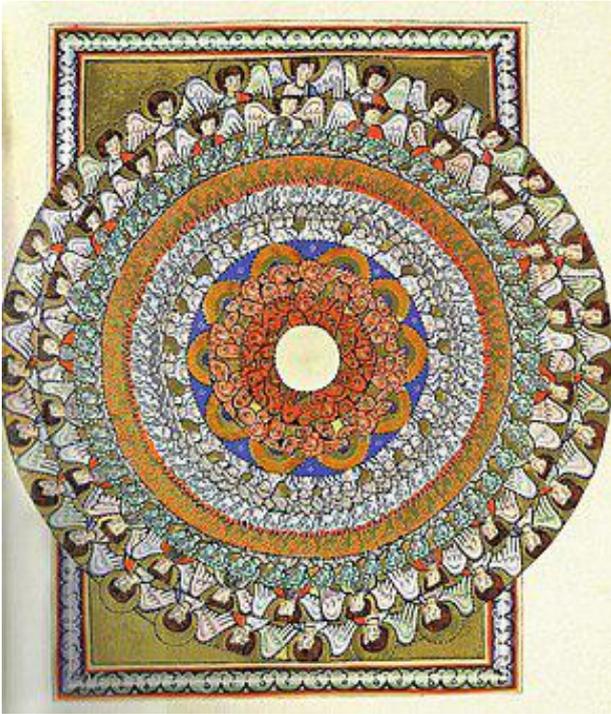


Figura 2: Hildegard von Bingen (1098-1179) da un Breviario



Figura 3: Piero di Puccio. Affreschi sul muro nord del Camposanto di Pisa (1389-1391)

l'opera "Gli immaginari in Geometria"² così si

²La parte finale di questo saggio, tradotta dall'originale russo ("Mnimosti v geometrii"), è contenuta in [3]. Il

esprime [3] (279-280, 282-283):

"È mia opinione che l'esegesi degli immaginari qui proposta, in relazione ai principi particolari e generali della relatività, getti nuova luce e argomenti la rappresentazione del mondo aristotelico-tolemaico-dantesco, quanto mai compiutamente cristallizzata nella Divina Commedia [...] Il quadro di tale universo non è raffigurabile come da schemi euclidei, così come la metafisica di Dante non è commensurabile con la filosofia di Kant. I matematici (Halsted nel 1905, Weber nel 1905, Simon nel 1912)³ hanno già rilevato in Dante un presentimento della geometria non euclidea.[...] la superficie su cui Dante si muove è tale che, [...] primo, in quanto contiene rette chiuse, è un piano di Riemann, e, secondo, in quanto capovolge la perpendicolare che su di essa si muove, è una superficie unilatera. Tali condizioni sono sufficienti a caratterizzare geometricamente lo spazio di Dante come conformato alla geometria ellittica. [...] Con ciò si getta una luce inattesa sulla concezione medievale della finitezza del mondo. Col principio di relatività, tuttavia, tali considerazioni geometriche generali hanno trovato di recente una sorprendente interpretazione concreta, e dal punto di vista della fisica moderna lo spazio del mondo va inteso proprio come spazio ellittico e si considera come finito così come finito e chiuso in sé il tempo."

saggio del 1920, censurato subito dalle autorità sovietiche, pare sia stato il principale capo di accusa per la condanna definitiva nel 1933 di Florenskij, che viene fucilato nel 1937 nei pressi di Leningrado, dopo aver trascorso quattro anni in un gulag nelle isole Solowki nel Mar Bianco. Per approfondimenti della sua visione dei numeri immaginari (molto diversa dalla tradizionale rappresentazione di Argand- Gauss) si veda [4]. Per un approfondimento anche dal punto di vista filosofico, con collegamenti con la Relatività, si veda [5].

³Georg Bruce Halsted, matematico americano fra i primi a introdurre le geometrie non euclidee negli Stati Uniti; Heinrich Martin Weber, matematico tedesco che fece ricerche di tipo puramente algebrico sui fondamenti dei piani di Riemann; Max Simon, matematico tedesco che si occupò prevalentemente di Storia della Matematica.

Molto probabilmente gli studi di Florenskij non furono conosciuti subito in Occidente, per cui generalmente viene indicato come scopritore dell'uso della geometria non-euclidea in Dante il matematico svizzero Andreas Speiser (1885-1970), che aveva studiato a Gottinga con David Hilbert e Hermann Minkowski. Egli ne parla (nel 1925) nel libro "Classici frammenti della Matematica" in questi termini [6]:

"Dante possiede una chiara visione globale della complessa struttura spaziale nella sua totalità. Per le nove sfere del cielo, Dante recupera la rappresentazione di Aristotele, apportando un cambiamento fondamentale che riguarda la fine dello spazio: come può essere che la sfera più distante, che appare la più grande, abbia in realtà le più piccole dimensioni? [...] Lo spazio di Dante è una varietà di Riemann con una fonte di energia che imprime ad esso la metrica."

(si vedano le figura 4 e 5).

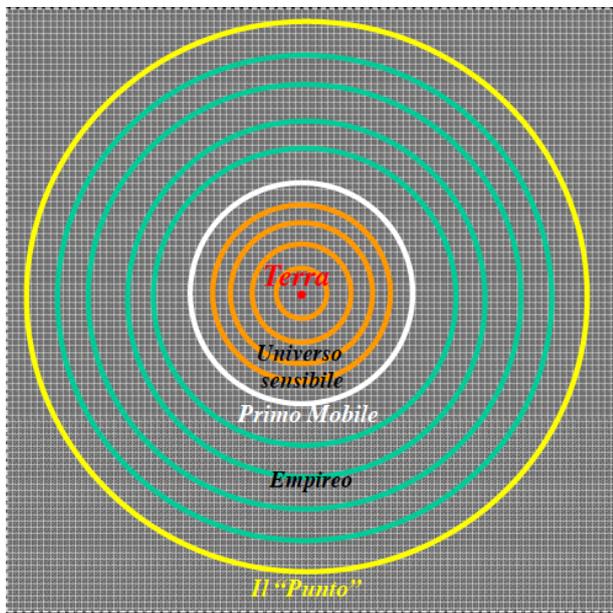


Figura 4: *Universo dantesco in uno spazio bidimensionale euclideo*

Il fisico-matematico e filosofo Hermann Weyl (1885-1955) invece nel 1932 scrive [7] (33-34):

La Divina Commedia di Dante non è solo un poema di grande potenza immaginativa, ma contiene anche una ben

precisa costruzione teologica e geometrica del cosmo, attraverso la quale la filosofia cristiana adatta la cosmologia aristotelica alle sue proprie esigenze. [...] Dante fa convergere i raggi che si diffondono dal centro della terra, la sede di Satana, verso un polo opposto, la fonte della potenza divina, proprio come sulla sfera terrestre i meridiani che si irradiano dal Polo Sud si riuniscono al Polo Nord. La potenza del Dio-persona deve irradiarsi dal centro, non può abbracciare la sfera del mondo rimanendo in quiete spaziale come il "primo motore immobile" di Aristotele. [...] I cerchi più interni, che cingono più da vicino la divina fonte di luce, per il fatto di contenere in massima misura la potenza divina, diventano quelli di massima estensione spaziale e racchiudono i cerchi più lontani. Nel linguaggio della matematica moderna, si potrebbe dire che Dante propone una dottrina che è stata ripresa da Einstein, anche se con motivazioni interamente differenti: si tratta della dottrina della chiusura dello spazio tridimensionale, ma dal polo della potenza divina s'irradia un campo metrico tale che la misura spaziale si riconnette alle condizioni descritte da Aristotele.

Chi per primo ha notato poi una somiglianza tra l'universo dantesco e la 3-sfera di uno spazio quadrimensionale è stato il matematico americano Mark Peterson nel 1979 nell'articolo "Dante e la 3-sfera" [8]. Egli così si esprime:

"La convinzione che la terra debba essere rotonda risale almeno ad Aristotele, la cui dottrina del luogo naturale richiedeva una terra rotonda al centro dell'universo. Questo stesso modello è diventato centrale per la teologia cristiana con il lavoro di Tommaso d'Aquino, e forma la struttura cosmologica della Divina Commedia di Dante. La convinzione che l'universo nel suo insieme possa essere rotondo (o più in generale curvo) è molto più recente. Sembra richiedere la matematica del XIX

secolo (geometria non euclidea) anche per formulare la nozione. È quindi una notevole sorpresa scoprire, a una lettura più attenta, che la cosmologia di Dante non è geometricamente semplice come appare a prima vista, ma in realtà sembra trattarsi di un universo cosiddetto chiuso, la 3-sfera, un universo che emerge anche come una soluzione cosmologica delle equazioni di Einstein nella teoria della relatività generale. Mi sono imbattuto in questa suggestione su Dante e la 3-sfera quando Dante si è posto di trattare una caratteristica evidentemente insoddisfacente della cosmologia aristotelica, quando egli, come narratore nel Paradiso, arriva al "bordo" o "vetta" "dell'universo. Come descriverebbe il bordo? È lo stesso problema che ogni bambino si trova davanti: a meno che l'universo non sia infinito, e quindi si sostiene avere un bordo, allora cosa c'è oltre il bordo? Dante affronta proprio questo problema alla fine della Divina Commedia, quando deve descrivere l'Empireo non in termini di principi o astrazioni, come faceva la cosmologia standard, ma come qualcosa effettivamente presente. L'Empireo viene visto e descritto per la prima volta nel Canto XXVIII del Paradiso in un brano che ho trovato, e ancora trovo, sorprendente. L'immagine è una 3-sfera, una buona descrizione di una che non ho mai visto da nessuna parte. Dante risolve così il problema del limite e allo stesso tempo completa la sua intera metafora cosmologica in un modo molto sorprendente e soddisfacente. Da allora ho capito che questo passaggio è considerato oscuro dalla critica, ma un relativista vedrà subito cosa sta succedendo⁴ [...]. Que-

⁴Dante fa coincidere la periferia dell'universo con un punto, curvando lo spazio su se stesso (vedi la figura 5). Rifarendosi al Punto, che è Dio, egli dice "parendo incluso da quel ch'elli include" (XXX,11-12), il verso oscuro a cui si allude. Per approfondimenti si possono leggere i libri [9], cap. 3, e [10], cap. 5, che trattano l'argomento anche in modo affascinante. Sull'intuizione di Dante di uno spazio sferico, il fisico Carlo Rovelli in [11] parla di

sta caratteristica, inaspettata in una cosmologia medievale, costituisce un'aggiunta interessante a qualsiasi discussione sullo spazio curvo, con evidenti ramificazioni interdisciplinari. "

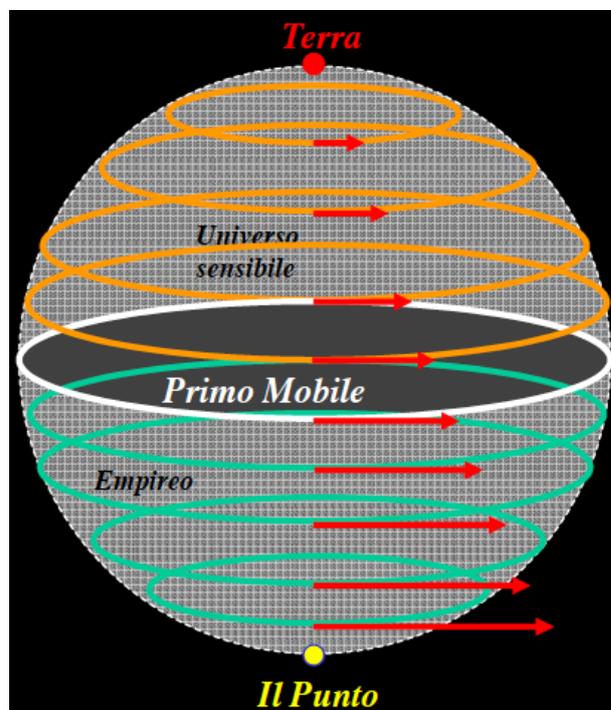


Figura 5: Universo dantesco in uno spazio sferico.

L'articolo, dopo le considerazioni matematiche che vedremo più avanti, così termina:

"Ci sono molti modi in cui questa cosmologia funziona nella Divina Commedia, ed è uno dei piaceri leggere la poesia per scoprirli. Non priverei il lettore di questi piaceri, ma ne menzionerò solo uno: la simmetria di riflessione della 3-sfera rispetto al suo equatore. Non è solo che Dio e Satana sono rappresentati come poli opposti, ma anche i loro intorni, per usare il termine matematico, sono il riflesso l'uno dell'altro: i nove cerchi angelici trovano una grottesca parodia nei cerchi dell'inferno. Considerazioni sulla simmetria devono sicuramente aver avuto un ruolo nella creazione dell'immagine dell'universo da parte di Dante. In effetti, la

una possibile influenza di Brunetto Latini, maestro di Dante.

ricomparsa di S^3 nella cosmologia moderna è in gran parte avvenuta attraverso argomenti di simmetria, sebbene ora l'omogeneità e l'isotropia dell'universo sono considerate proprietà più convincenti. Ci sono persone che non crederanno mai che la fisica sia bella e che forse ancora risentono del modo in cui ha demolito la visione del mondo medievale. Saranno sorpresi di apprendere che la fisica moderna può anche illuminare la ricchezza dell'immaginazione medievale."

Nel 2006 approfondisce l'analogia il fisico (scrittore, filosofo e saggista) rumeno Horia-Roman Patapievici (n. 1957) nel libro *Ochii Beatricei. Cum arăta cu adevărat lumea lui Dante*, pubblicato in italiano [12]. Nel commentare questo libro, l'italianista Geo Vasile così si esprime:

L'autore rumeno prende in considerazione, malgrado il loro ingenuo immaginario, i disegni che lungo i secoli hanno contornato una visuale dell'universo dantesco. Se ci riferiamo solo all'immagine lasciata da Michelangelo Cactani (1855), a prescindere dal cielo con le stelle fisse, conforme all'armoniosa cosmologia ellenica, appare una specie di inestetica escrescenza rappresentante l'Empireo e le gerarchie angeliche intorno a Dio. Horia-Roman Patapievici non ammette che Dante fosse capace di tale deformazione, tanto più che lo stesso poeta non poteva mancare l'istante unico nella sua vita: immortalare la visione abbagliante di Dio. Lo fa valendosi degli occhi di Beatrice come di uno specchio, anche se l'immagine nello specchio viene rovesciata. Il mondo invisibile diventa così una copia rovesciata del mondo visibile: l'Empireo è dio-centrico, mentre la Terra è demonocentrica, i cori degli angeli girano intorno a Dio con una velocità sempre più grande, mentre i cieli rallentano i loro motori man mano che s'avvicinano alla Terra; l'invisibile sottostà a delle norme in opposizione con quelle del mondo visibile. E per spiegare queste simmetrie,

all'autore non rimane altro che concepire l'universo visibile (avendo nel centro la Terra) e l'Empireo (avendo nel centro Dio) come due sfere che condividono la stessa superficie, cioè il "primo mobile". Possiamo chiamarlo l'equivalente di un'ipersfera, oggetto della geometria di Riemann, adottato da Einstein per descrivere l'universo della relatività dello spazio e del tempo. (Si vedano le figure 1 e 4, 5, 6 [13])"

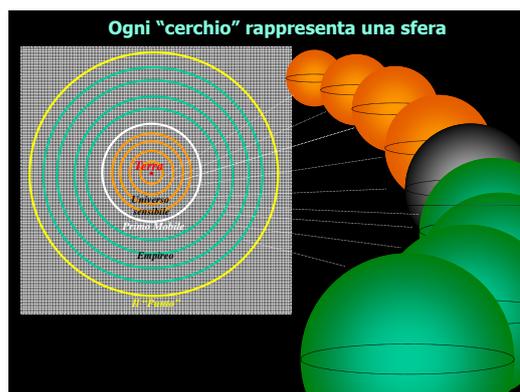


Figura 6: La corrispondenza tra circonferenze e sfere.

L'ipersfera

Come è usuale, indichiamo con S_r^n la sfera (superficie sferica) n -dimensionale di raggio r dello spazio euclideo \mathbf{R}^{n+1} , cioè

$$S_r^n = \{(x_1, \dots, x_{n+1}) \in \mathbf{R}^{n+1} | x_1^2 + \dots + x_{n+1}^2 = r^2\}.$$

Se $r = 1$, scriveremo semplicemente S^n . Indichiamo con D_r^n la n -sfera piena o n -disco di \mathbf{R}^n , cioè

$$D_r^n = \{(x_1, \dots, x_n) \in \mathbf{R}^n | x_1^2 + \dots + x_n^2 \leq r^2\}.$$

La sfera S_r^n è il bordo di D_r^{n+1} . In simboli :

$$\partial D_r^{n+1} = S_r^n ;$$

la circonferenza è bordo del cerchio, la superficie sferica è bordo della sfera piena.

Per comprendere bene la costruzione attribuita a Dante, partiamo dalle dimensioni più basse:

$$S_r^0 = \{-r, +r\} \subset \mathbf{R} ,$$

è una coppia di punti sulla retta;

$$D_r^1 = [-r, +r] \subset \mathbf{R},$$

è il segmento di estremi $-r$ e $+r$; quindi

$$\partial D_r^1 = S_r^0$$

Ora, è possibile deformare D_r^1 (di lunghezza $2r$) in una semicirconferenza, avente bordo la coppia $\{-r', +r'\}$, dove $r' = 2r/\pi$. Tecnicamente si dice che D_r^1 è omeomorfo⁵ alla semicirconferenza. Ora incollando specularmente due semicirconferenze si ha: $S_{r'}^1 \subset \mathbf{R}^2$.

Analogamente il disco D_r^2 si può deformare in una superficie semisferica (di opportuno raggio r') e, incollando due copie di questa tramite il loro bordo, si ha $S_{r'}^2 \subset \mathbf{R}^3$; intuitivamente è come gonfiare un palloncino di gomma.

Così procedendo, incollando due copie di D_r^3 tramite il bordo S_r^2 , si ha uno spazio omeomorfo a $S^3 \subset \mathbf{R}^4$. Naturalmente, non possiamo fare la figura di S^3 nello spazio quadridimensionale, ma possiamo immaginarcela⁶ considerando che una qualsiasi intersezione (non vuota) di S^3 con un iperpiano (spazio tridimensionale) dà la sfera S^2 .

La costruzione delle sfere sopra indicata, in Topologia, si chiama **sospensione** di uno spazio topologico (si veda la figura 8).

La sospensione di uno spazio topologico X , in simboli ΣX , è ottenuta da $[0, 1] \times X$ identificando ogni elemento del sottoinsieme $\{0\} \times X$ e $\{1\} \times X$ ad un punto.

Lo spazio topologico $[0, 1] \times X$ è chiamato **cilindro su X** , perciò ΣX è detta anche **doppio cono su X** .

Nella figura allegata $\Sigma S^n = S^{n+1}$.

⁵Un omeomorfismo $f : X \rightarrow Y$ tra due spazi topologici X e Y è una corrispondenza biunivoca e continua insieme alla sua inversa. Per es. un segmento privato degli estremi è omeomorfo ad una retta, ma anche ad una spirale; il bordo di un poligono è omeomorfo ad una circonferenza. In Topologia due spazi omeomorfi sono indistinguibili, perciò si può confondere S^n con S^n .

⁶Lo studio delle sezioni in una certa dimensione dà informazioni su fenomeni che avvengono in dimensioni superiori, come ha ben messo in evidenza E.A. Abbott (1838-1926) nel suo romanzo fantastico "Flatlandia" [14], che può considerarsi una divertente (ma rigorosa) introduzione al concetto di dimensione. In un mondo piatto, ad esempio, la sfera ordinaria si rivela dinamicamente come un insieme di cerchi concentrici.

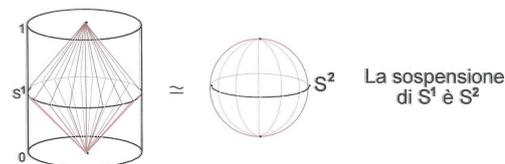
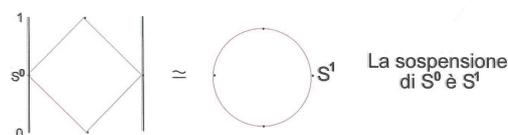
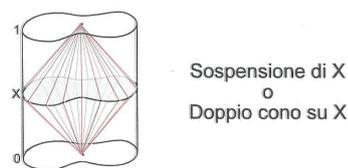


Figura 7: Sospensione di uno spazio topologico.

Inoltre, il matematico Peterson nota che la sospensione di S^2 rispetta anche quanto dice Dante nel Paradiso, XXVIII 41-42

"...Da quel punto/ dipende il cielo e tutta la natura."

Infatti, etimologicamente, la parola italiana dipende significa proprio che l'oggetto pende, cioè è sospeso come un lampione, esattamente secondo la costruzione sopra indicata. Nel caso nostro, si tratta della sospensione della Terra con vertici Dio e Satana.

Si osservi che, almeno nel caso bidimensionale, l'idea di ridurre ad un punto il bordo di un rettangolo (o di un disco) proviene dall'esperienza quotidiana di formare con un fazzoletto un fagottino (omeomorfo alla superficie sferica). In matematica troviamo un'analogia costruita, chiamata "compattificazione di uno spazio tramite l'aggiunta di un solo punto" (usualmente detto **punto all'infinito**). Essa è dovuta al matematico russo P. S. Alexandroff (1924) [15].

Ebbene, la compactificazione di Alexandroff dello spazio euclideo \mathbf{R}^n (o del disco senza bordo) è proprio la sfera S^n . Nel caso $n = 2$, la costruzione può essere resa visibile proiettando la sfera S^2 dal polo nord sul piano equatoriale (proiezione stereografica)[16], cap. 3. Si ottiene così una rappresentazione vicina a quella classica dell'Universo dantesco, dove il primo mobile è l'equatore.

Interessanti (e sorprendenti) proprietà di S^3 , con ricadute sulla possibile topologia dell'Universo (non solo dantesco!), si hanno considerando la fibrazione di Hopf di S^3 (dove S^2 è lo spazio base e S^1 la fibra); ma non ci fermiamo su questo concetto avanzato. Questa fibrazione fu scoperta nel 1931 [17] da Heinz Hopf (1894-1971), matematico svizzero, pioniere della Topologia algebrica.



Figura 8: *Fibrazione di Hopf.*

Ricordiamo, infine, che S^3 è anche lo spazio dei quaternioni⁷ di norma unitaria in R^4 , come S^1 è lo spazio dei numeri complessi di norma unitaria in R^2 . Ne segue che S^1 e S^3 hanno anche una struttura algebrica, con moltiplicazione commutativa nel caso dei numeri complessi e non commutativa nel caso dei quaternioni (per un approfondimento si veda [19]).

Conclusione

Concludo con due osservazioni di studiosi dell'argomento. Il critico e filosofo letterario William Egginton (n. 1969) osserva [20]:

"Non sto proponendo di interpretare ingenuamente il passato a nostra im-

⁷I quaternioni sono numeri complessi a 3 unità immaginarie, che hanno notevoli applicazioni in Fisica e in Robotica: infatti come i numeri complessi sono adatti a rappresentare i movimenti nel piano, così i quaternioni sono adatti a rappresentare i movimenti nello spazio. Per una semplice introduzione si veda [18].

agine, affermando che Dante era un genio matematico prima del suo tempo. Sto, tuttavia, suggerendo che esiste un'eccellente spiegazione del motivo per cui si deve guardare al XIV secolo per modelli di espressione adeguati alle sfide poste dai progressi del XIX e del XX secolo in matematica: vale a dire, quelle nel corso della Rivoluzione Scientifica, in cui furono poste le basi epistemologiche per le scoperte della scienza moderna. Certe possibilità di pensiero e immaginazione furono scartate, dimenticate e certe abilità andarono, anche se solo temporaneamente, perse. Il Medioevo è stato a lungo dipinto come un periodo la cui cultura era ostile al libero pensiero e alla nuova conoscenza; forse è ora il momento di riconoscere che era anche un tempo in cui i fenomeni potevano avere spiegazioni che si escludono a vicenda e contraddittorie, quando la battaglia tra fede e scienza non era stata ancora vinta in modo definitivo da nessuna delle due parti, e quando gli sforzi per colmare le storie raccontate potrebbero produrre edifici concettuali la cui mera possibilità sarebbe stata impensabile perfino 200 anni dopo. Questo saggio parla di uno di questi edifici concettuali e di come l'apertura alla sua prospettiva abbia permesso a un poeta di esprimere a parole ciò che non poteva essere immaginato dai suoi discendenti e ciò che ancora oggi ci sfugge."

L'astrofisico Marco Bersanelli (n. 1960) conclude così un suo articolo sull'argomento [21]:

"Naturalmente Dante non era uno scienziato moderno⁸. Grazie ai progressi della scienza la nostra conoscenza del cosmo e della natura oggi è immensamente più vasta e dettagliata di quella

⁸Tuttavia, come lo stesso Bersanelli afferma in [9] pag. 96: "L'immagine cosmica di Dante, qualunque sia stata la via per la quale il poeta vi giunse, merita a pieno titolo di essere considerata una delle vette più alte nella storia del pensiero cosmologico, al pari delle più mirabili intuizioni degli antichi Greci. Anche perché [...] essa risuona in modo sorprendente con la nostra attuale rappresentazione dello spazio-tempo."

dei medievali (chissà quanto avrebbe goduto Dante a conoscere anche solo una piccola parte di quello che abbiamo compreso oggi sulla struttura dell'universo e sulla simmetria delle leggi della fisica!). Ma forse noi moderni rischiamo di perdere la cosa più preziosa: quella gratitudine, quell'ampiezza della ragione, quella tensione all'unità, quel senso del mistero che doveva arde- re nello sguardo e nel cuore di Dante Alighieri e che, come diceva Einstein, "è il seme di ogni vera arte e di ogni vera scienza" ".

[...] Le cose tutte quante hanno ordine tra loro e questo è forma che l'universo a Dio fa somigliante.

Qui veggion l'alte creature l'orma de l'eterno valore, il qual è fine al quale è la toccata norma.

(Paradiso, I, 103-108)

Ringraziamenti

Ringrazio i colleghi Giampaolo Co', Stefano Marchiafava e Rocco Chirivì per i suggerimenti atti a migliorare il testo.



- [1] D. Costamagna, *Dante e Einstein*, https://lalucedellafisica.wordpress.com/2019/11/03/dante_e_einstein
- [2] R. Betti, *Pavel Florenskij: fede e matematica* <http://matematica.unibocconi.it/articoli/pavel-florenskij-fede-e-matematica>
- [3] P. A. Florenskij: *Il simbolo e la forma, scritti di filosofia della forma*, a cura di N. Valentini e A. Gorelov, Bollati Boringhieri, Torino (2007).
- [4] R. Betti: *La matematica come abitudine del pensiero. Le idee scientifiche di Pavel Florenskij*, Centro Pristem Eleusi, Milano (2009).
- [5] A. Oppo: *Se la misura di un corpo è un numero immaginario. Florenskij e il concetto di spazio in Dante*, Annali della Pontificia Facoltà Teologica della Sardegna, XXIV (2015) 171.

- [6] A. Speiser: *Klassische Stücke der Mathematik*, Orell Füssli, Zurich (1925).
- [7] H. Weyl: *Il mondo aperto*, Bollati Boringhieri, Torino (1981).
- [8] M. A. Peterson: *Dante and the 3-sphere*, Am. J. Phys., 47 (1979) 1031.
- [9] M. Bersanelli: *Il grande spettacolo del cielo. Otto visioni dell'Universo dall'antichità ai nostri giorni*, Sperling & Kupfer, Milano (2018).
- [10] R. Osserman: *Poesia dell'Universo. L'esplorazione matematica del cosmo*, Longanesi, Milano (1995).
- [11] C. Rovelli: *La realtà non è come ci appare*, Raffaello Cortina, Milano (2014).
- [12] H. R. Patapievici: *Gli occhi di Beatrice. Com'era davvero il mondo di Dante*, Bruno Mondadori, Milano (2006).
- [13] M. Bersanelli, *L'Universo di Dante* https://www.iasf-milano.inaf.it/Astro-Siesta/astro_marcob_1_Dante.pdf
- [14] E. A. Abbott: *Flatlandia*, Adelphi, Milano (1966).
- [15] P. S. Aleksandroff: *Über die Metrisation der im kleinen kompakten topologischen Räumen*, Math. Ann., 92 (1924) 294.
- [16] G. De Cecco, E. Mangino: *La sfera in Geometria e Geografia*, Quaderno, Dip. Mat. "E. De Giorgi", Univ. Studi Lecce, 1 (2001) Cap. 3.
- [17] H. Hopf: *Über die Abbildungen der dreidimensionalen Sphäre auf die Kugeloberfläche*, Math. Ann., 101 (1931) 637.
- [18] G. De Cecco, R. Vitolo, *Note di calcolo matriciale, Anno acc. 2003-2004*, Fac. Ing. Univ. Studi Lecce (reperibile anche in rete)
- [19] L. Zulli: *Charting the 3-Sphere-An Exposition for Undergraduates*, Am. Math. Montly, 103 (1996) 221.
- [20] W. Egginton: *On Dante, Hyperspheres and the Curvature of the Medieval Cosmos*, Journal of the History of Ideas, 60 (1999) 195.
- [21] M. Bersanelli: www.scienzainrete.it/articolo/l-universo-di-dante/marco-bersanelli/2011-10-26



Giuseppe De Cecco: è stato professore ordinario di Geometria nell'Università del Salento, dove ha insegnato per circa quaranta anni. Ha tenuto numerosi corsi di tipo diverso, privilegiando sempre l'aspetto interdisciplinare, anzi transdisciplinare. Le sue ricerche sono in Geometria differenziale, Topologia algebrica ed Analisi globale, in particolare sulle varietà riemanniane con singolarità.

Numero XVII Anno 2021

Viaggio nella Scienza

Ithaca



Dare i numeri: (ab)usi della Matematica (B)

