
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...

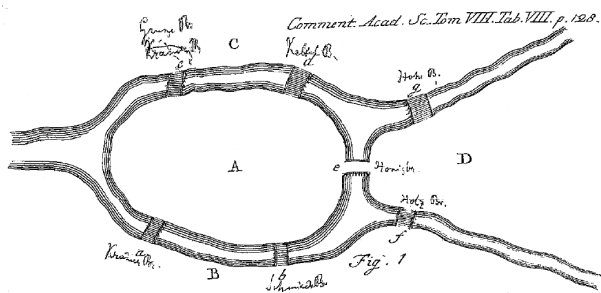


Figura 1: Pianta della città di Königsberg nel 1736. Essendo divisa in quattro zone in ciascuna delle quali afferisce un numero di ponti dispari, una passeggiata che attraversi tutti i ponti una sola volta non esiste [1].

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].

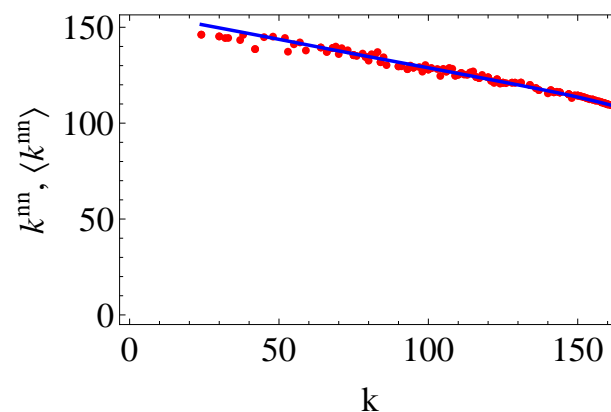


Figura 2: World Trade Web nell'anno 2002: andamento del numero medio dei vicini di ciascun vertice, in funzione del grado degli stessi (in rosso, l'andamento empirico; in blu, l'andamento atteso sotto il Configuration Model) [13].

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*

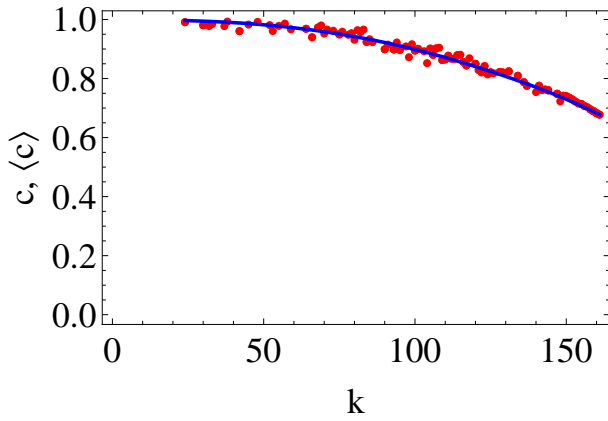


Figura 3: World Trade Web nell'anno 2002: andamento della frazione di triangoli di cui ciascun nodo è anche un vertice, in funzione del grado degli stessi (in rosso, l'andamento empirico; in blu, l'andamento atteso sotto il Configuration Model) [13].

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 $\{\text{GDP}_i\}_{i=1}^N$ 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,

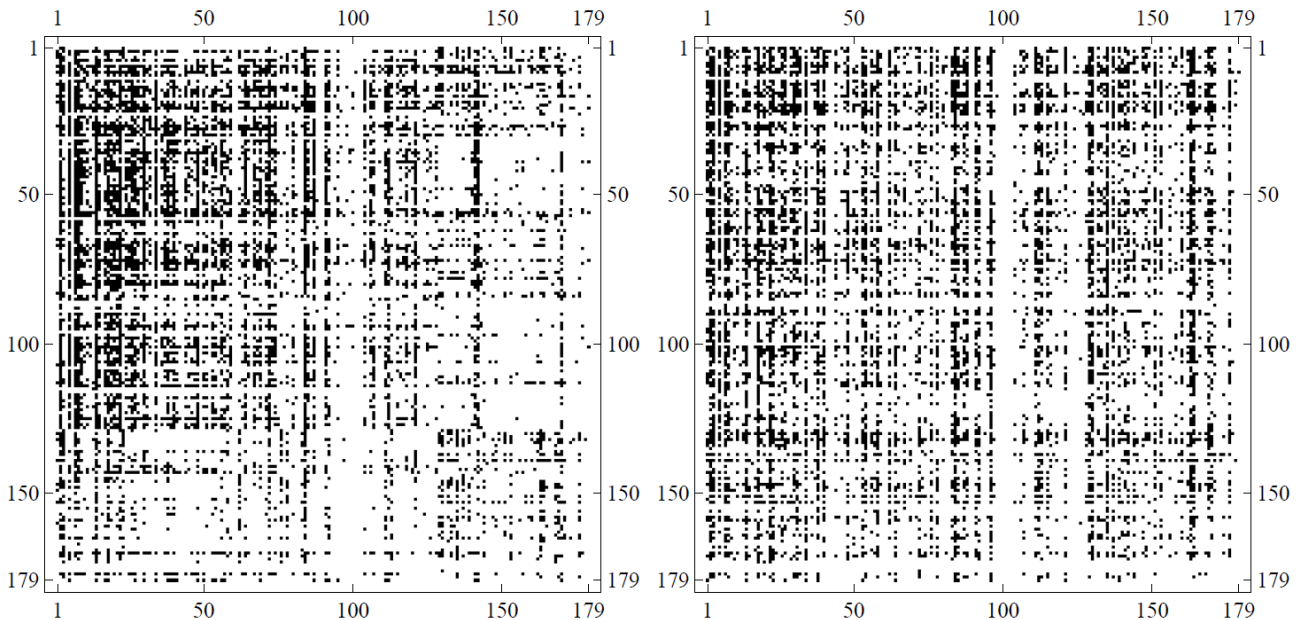


Figura 4: Sulla sinistra: matrice di adiacenza della rete interbancaria eMID, nel 1999. Sulla destra: matrice di adiacenza ricostruita sulla base del fitness model finanziario [13].

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.

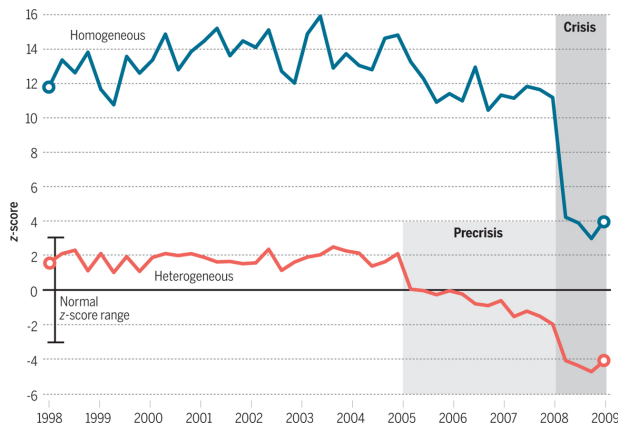


Figura 5: Analisi di significatività della reciprocità, misurata sulla rete di scambi tra banche olandesi nella decade 1998-2008. L'andamento in blu è ottenuto impiegando il modello di Erdős come benchmark; l'andamento in rosso è ottenuto impiegando il Configuration Model come benchmark [14].

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.

