RETI SOCIALI E DIFFUSIONE DI EPIDEMIE

di Carlo Piccardi

Carlo Piccardi

È docente di Automatica al Politecnico di Milano, dove insegna Dinamica dei sistemi e delle reti complesse.

Laureato in Ingegneria elettronica, nella sua attività di ricerca si è occupato di problemi di modellistica, controllo e identificazione di sistemi dinamici con applicazioni in settori diversi quali ad esempio le risorse idriche, l'epidemiologia, le scienze sociali ed economiche.

Il termine *rete sociale* (*social network*) è utilizzato da alcuni anni per designare le comunità virtuali degli utenti di piattaforme quali Facebook, Twitter o simili. Il legame tra gli individui ("amici" in Facebook o "followers" in Twitter) è pertanto virtuale, mediato dal *World Wide Web* e quindi veicolato dall'infrastruttura tecnologica di Internet. La Sociologia quantitativa, tuttavia, utilizza il termine rete sociale da ben prima dell'avvento di Internet per designare insiemi di individui tra loro connessi attraverso legami più tradizionali, in primo luogo la frequentazione reciproca che comporta una vera e propria prossimità fisica [1]. Ad esempio, un individuo adulto che vive e lavora in una città ha contatti sociali (seppure con frequenza diversa) con i propri familiari, i colleghi di lavoro, gli amici, gli utenti dei mezzi pubblici ecc. Com'è ovvio, è questa la rete sociale che ci interessa se vogliamo capire e modellizzare i fenomeni relativi alla diffusione di epidemie, la maggior parte delle quali si propagano per contatto tra individui: alcune per semplice prossimità (starnuti

o strette di mano possono trasmettere, ad esempio, influenza o morbillo), altre con contatti meno superficiali (si pensi alle malattie sessualmente trasmissibili o alla propagazione di epatiti e HIV causata dallo scambio di materiale infetto). Il connubio tra Epidemiologia e Matematica ha una lunga tradizione. Il primo significativo utilizzo di un modello matematico risale infatti al 1927, quando Kermack e McKendrick proposero un sistema di equazioni differenziali per spiegare i dati dell'evoluzione di un'epidemia di peste in India [2]. La loro assunzione fondamentale consisteva nell'immaginare la popolazione come una specie di gas perfettamente miscelato, in cui ogni individuo era equivalente a tutti gli altri e aveva, in un certo intervallo di tempo, la medesima probabilità di incontrare qualunque altro individuo. È proprio dall'incontro tra un individuo sano e non immune

“ È dall'incontro tra un individuo sano e non immune (detto *suscettibile*, S) e uno *infetto* (I) che, con una certa probabilità che dipende dalla virulenza del patogeno, si trasmette l'infezione: il suscettibile diventa a sua volta infetto. ”

(detto *suscettibile*, S) e uno *infetto* (I) che, con una certa probabilità che dipende dalla virulenza del patogeno, si trasmette l'infezione: il suscettibile diventa a sua volta infetto. Questo approccio modellistico ha dominato l'Epidemiologia matematica per circa 70 anni, pur con innumerevoli varianti ed estensioni che hanno permesso, ad esempio, di adattare il modello a differenti patogeni o di raffinarlo distinguendo classi d'età o sesso degli individui [3]. Vediamo un esempio, uno dei più semplici: nel modello *SIS* si ipotizza che ciascun individuo possa infettarsi e guarire, compiendo cioè le transizioni $S \rightarrow I \rightarrow S \rightarrow \dots$, per un numero illimitato di volte. Se indichiamo con $y(t)$ la frazione di infetti nella popolazione all'istante t , per cui $s(t) = 1 - y(t)$ è la frazione di suscettibili, abbiamo:

$$\dot{y}(t) = -\gamma y(t) + \beta y(t)s(t) = -\gamma y(t) + \beta y(t)(1 - y(t)).$$

Il primo termine del secondo membro rappresenta il flusso $I \rightarrow S$ di guarigioni ($1/\gamma$ è il tempo medio in cui un individuo rimane infetto), mentre il secondo termine è quello cruciale per descrivere il meccanismo $S \rightarrow I$ d'infezione: la probabilità che ciascun S incontri un I è proporzionale alla densità $y(t)$ di questi ultimi, per cui il flusso $S \rightarrow I$ è proporzionale al prodotto $y(t)s(t)$. Il coefficiente di trasmissione β dipende sia dalla struttura sociale, poiché caratterizza la frequenza degli incontri, che dalle caratteristiche di virulenza del patogeno. Cosa prevede il modello *SIS*? È facile verificare che esistono due soluzioni di equilibrio (Fig. 1): $y = 0$, che banalmente corrisponde all'assenza di infetti, e $y = 1 - \gamma/\beta$, che è positiva e asintoticamente stabile se $\beta > \beta_{cr} = \gamma$ mentre, per $\beta < \beta_{cr}$ è $y = 0$ ad essere asintoticamente stabile. Se ne deduce che il numero di infetti a regime cresce con β , come era intuibile, ma soprattutto che esiste un valore-soglia β_{cr} al di sotto del quale l'epidemia è destinata ad estin-

FIGURA 1: NEL CLASSICO MODELLO *SIS*, LA QUANTITÀ DI INFETTI A REGIME CRESCE CON IL COEFFICIENTE DI TRASMISSIONE β MA L'EPIDEMIA SI ESTINGUE (CIOÈ $y \rightarrow 0$) SE $\beta < \beta_{cr}$

Interventi

guersi. Se β è troppo piccolo, infatti, i pochi infetti presenti nella popolazione non sono in grado di generare abbastanza nuovi infetti, a causa della scarsità dei loro contatti sociali e/o della scarsa virulenza del patogeno. L'esistenza di questa soglia è, da sempre, uno dei cardini della modellistica epidemiologica [3].

Dai modelli tradizionali ai modelli a rete

La complessità delle società odierne, specialmente nei Paesi avanzati, mette in crisi l'approccio modellistico sopra delineato. Se già è assai discutibile l'ipotesi di "miscela perfetta" della popolazione, visto che ciascuno ha una propria specifica rete di interazioni sociali, ancora meno fondata si rivela l'assunzione che tutti gli individui siano pressoché equivalenti in termini di numero di contatti: svariate indagini hanno mostrato che gli individui sono in realtà ampiamente differenziati, nel senso che tipicamente la popolazione comprende un ristretto gruppo di individui (i cosiddetti *hub*) caratterizzati da un elevato numero di contatti sociali, inseriti nella generalità della popolazione che è invece caratterizzata da connettività medio-bassa (si vedano gli esempi di Fig. 2). È lecito attendersi che gli *hub* giochino un ruolo cruciale nella diffusione di epidemie.

Un altro aspetto riguarda i meccanismi di propagazione spaziale. Il modello *SIS* sopra presentato non tiene conto

FIGURA 2: IL NUMERO DI CONTATTI DI PROSSIMITÀ DEGLI ABITANTI DI UNA GRANDE CITTÀ (IN ALTO [4]) O IL NUMERO DI PARTNER SESSUALI NEGLI ULTIMI 12 MESI DICHIARATI DAGLI INDIVIDUI DI UNA COMUNITÀ (IN BASSO [5]) VARIANO GRANDEMENTE, INDICANDO UNA FORTE ETEROGENEITÀ NELLA CONNETTIVITÀ DEGLI INDIVIDUI STESSI

Interventi

della dimensione spaziale, proprio perché immagina la popolazione idealmente confinata in un “contenitore” e lì perfettamente miscelata. Volendo descrivere la propagazione dell’epidemia sul territorio, dovremmo derivare un modello alle derivate parziali nella variabile $y(t,x)$, in cui x è la variabile spaziale. Le più semplici e intuitive generalizzazioni del modello *SIS* ipotizzano un termine diffusivo, per cui un’epidemia che si origina in un punto si allarga con un fronte d’onda ricoprendo aree sempre più vaste. È un fenomeno documentato in innumerevoli occasioni nel passato. In Fig. 3 è riportato l’esempio dell’epidemia di peste nera che decimò l’Europa a metà del XIV secolo: sbarcata in Sicilia nel 1347, l’epidemia si diffuse gradualmente verso nord ricoprendo l’intera Europa nel giro di tre anni. In concreto, la propagazione fu mediata dai lenti e quantitativamente modesti spostamenti di viaggiatori e mercanti tra villaggi e città tra loro adiacenti.

Ben diversa è la situazione odierna, come emerge in tutta evidenza dall’altro caso presentato in Fig. 3 e relativo all’epidemia di SARS dell’estate 2003. L’epidemia ebbe origine in estremo oriente (Hong Kong) ma nel giro di poche settimane si accesero importanti focolai in molte altre città sparse in tutto il mondo, specialmente in quelle ospitanti importanti comunità cinesi. Il meccanismo di diffusione fu duplice. Da un lato, analogamente al caso sopra citato, fu di tipo “locale”: la rete di contatti sociali comprende perlopiù individui a sé prossimi (familiari, colleghi, amici, ecc.) con i quali i contatti sono relativamente frequenti. È questo il motore primo di diffusione che causò, infatti, un’elevata incidenza nella popolatissima Hong Kong e nei dintorni. Ma a ciò si affiancò una diffusione a grande distanza veicolata dal trasporto aereo, nel caso specifico verso le aree di importante emigrazione. Oltre ai contatti locali, relativamente ab-

bondanti e frequenti, gli individui possiedono spesso contatti “remoti”, tipicamente pochi e rari, ma nel complesso sufficienti a propagare l’epidemia a grande distanza in tempi brevissimi: un fenomeno ovviamente impossibile non solo ai tempi della peste nera ma anche fino a pochi decenni or sono e che non può essere descritto con i tradizionali modelli diffusivi, richiedendo il ricorso a modelli a rete.

Le più semplici e intuitive generalizzazioni del modello *SIS* ipotizzano un termine diffusivo, per cui un’epidemia che si origina in un punto si allarga con un fronte d’onda ricoprendo aree sempre più vaste.

Elementi di teoria delle reti

Una rete è definita da un *grafo* composto da N nodi e L *collegamenti* o *link* [8-11]. Nel caso delle reti sociali, i nodi rappresentano gli individui mentre i *link* descrivono il legame sociale esistente tra coppie di individui. Come già evidenziato, questo può essere un legame virtuale, se stiamo considerando i moderni *social network* basati sul WWW, o un vero e proprio legame fisico se la rete modella relazioni di frequentazione tra familiari, amici, colleghi ecc. Poiché ci occupiamo di diffusione di epidemie, possiamo senz’altro assumere che la rete sia *non diretta*, visto che il patogeno può potenzialmente trasmettersi sia nella direzione $i \rightarrow j$ che $j \rightarrow i$. Inoltre, limitandoci al caso più semplice, trascuriamo la possibilità di dotare ciascun *link* $i \leftrightarrow j$ di un pe-

FIGURA 3: LA PROPAGAZIONE TERRITORIALE DELLA PESTE NERA DEL XIV SECOLO (A SINISTRA [6]) È GOVERNATA DAL CLASSICO MECCANISMO DIFFUSIVO. NEL CASO DELL’EPIDEMIA DI SARS DEL 2003 (A DESTRA [7]), L’EPIDEMIA SI PROPAGA RAPIDAMENTE A GRANDE DISTANZA GRAZIE AL TRASPORTO AEREO

FIGURA 4: UNA RETE CON $N = 7$ NODI E $L = 14$ LINK E LA CORRISPONDENTE MATRICE DI ADIACENZA A

so $w_{ij} > 0$ che potrebbe, ad esempio, modulare la frequenza dei contatti sociali tra diverse coppie di individui. Con queste ipotesi, la rete è completamente descritta dalla *matrice di adiacenza* $A = [a_{ij}]$, la quale è una matrice $N \times N$ in cui $a_{ij} = a_{ji} = 1$ se esiste il *link* $i \leftrightarrow j$, mentre $a_{ij} = a_{ji} = 0$ in caso contrario: A è quindi una matrice simmetrica.

La rete si dice *connessa* se, per qualunque coppia (i, j) di nodi, esiste un percorso nel grafo che collega i a j e viceversa. La *distanza* d_{ij} è allora definita come la lunghezza (espressa in numero di *link*) del più breve percorso (i, j) . Da qui, possiamo introdurre i concetti di *diametro* D della rete, cioè la più grande delle distanze d_{ij} , e di *distanza media* d , vale a dire la media di tutte le d_{ij} delle $N(N-1)/2$ coppie distinte di nodi. Come vedremo più avanti, una caratteristica rilevante delle reti reali, e di quelle sociali in particolare, è di avere distanza media sorprendentemente piccola anche quando la loro dimensione N risulta molto grande.

Il *grado* k_i è il numero di contatti del nodo i , vale a dire il numero di *link* collegati a i : risulta quindi $k_i = \sum_j a_{ij}$. La più importante caratterizzazione di una rete è data dalla *distribuzione di grado* $P(k)$ la quale, per ogni $k = 1, 2, \dots$, specifica l'abbondanza relativa dei nodi di grado k presenti nella rete:

$$P(k) = \frac{\text{numero nodi di grado } k}{N}.$$

Poiché $\sum_k P(k) = 1$ e $0 \leq P(k) \leq 1$ per ogni k , la distribuzione di grado è formalmente analoga a una distribuzione di probabilità discreta. In questo senso, il valore $P(k)$ può essere interpretato come la probabilità che un nodo scelto a caso abbia grado pari a k . Il *grado medio* m della rete risulta dato da:

$$m = \frac{1}{N} \sum_{i=1}^N k_i = \sum_k k P(k) = \frac{2L}{N}$$

dove l'ultima espressione deriva dal fatto che ognuno degli L *link* contribuisce per un'unità al grado di entrambi i nodi a cui è connesso.

FIGURA 5: A SINISTRA: UNA RETE QUASI-OMOGENEA (ERDŐS-RÉNYI) CON $N = 100$ E $L = 300$. A DESTRA: UNA DISTRIBUZIONE DI GRADO POISSONIANA CON GRADO MEDIO m (DA [12])

Una rete si dice (strettamente) *omogenea* quando tutti i nodi hanno lo stesso grado $k_i = m$, per cui $P(m) = 1$ e $P(k) = 0$ per $k \neq m$. È una situazione che mai si riscontra nelle reti reali dove è possibile trovare invece esempi di reti *quasi omogenee*, in cui cioè il grado dei nodi si discosta non di molto dal valore medio m . Un esempio importante è costituito dalle cosiddette *reti Erdős-Rényi*: fissati N ed L (e quindi il grado medio $m = 2L/N$), la rete si costruisce assegnando i *link* a L coppie di nodi sorteggiate in modo casuale tra le $N(N-1)/2$ possibili. Per N e L sufficientemente grandi (a rigore, per $N \rightarrow \infty$ con m fissato), si può dimostrare che $P(k)$ tende ad una distribuzione di Poisson con media m : in pratica, significa che la gran parte dei nodi ha grado poco distante da m che, pertanto, può essere preso come valore rappresentativo (o "scala tipica") del grado dei nodi. In termini di reti sociali, una rete Erdős-Rényi rappresenta quindi una comunità di individui in cui, approssimativamente, tutti hanno lo stesso numero di conoscenti.

Negli ultimi quindici anni, la massiccia analisi di dati rappresentativi di reti reali ha mostrato con chiarezza che distribuzioni di grado (quasi) omogenee costituiscono l'eccezione, piuttosto che la regola. Molte reti hanno infatti distribuzioni di grado fortemente *eterogenee*, in cui pochi nodi di grado molto elevato (i cosiddetti *hub*) coesistono con la maggioranza dei nodi, dotati di grado medio-basso. Esempi in tal senso sono stati riscontrati pressoché in tutti i settori, dalla Biologia al WWW, dalle reti di trasporto a quelle di scambi economici (solo per citare alcuni esempi). In una rete sociale, ciò significa che esistono nella popolazione individui che, a causa dell'insieme delle loro attivi-

“ Molte reti hanno distribuzioni di grado fortemente eterogenee, in cui pochi nodi di grado molto elevato (i cosiddetti *hub*) coesistono con la maggioranza dei nodi, dotati di grado medio-basso. ”

Interventi

FIGURA 6: A SINISTRA: UN ESEMPIO STILIZZATO DI RETE ETEROGENEA. A DESTRA: IN SCALA LOGARITMICA, UNA DISTRIBUZIONE DI GRADO A LEGGE DI POTENZA (TRONCATA A k FINITO); IL GRADO DEI NODI DELLA RETE CORRISPONDENTE VARIA TRA 1 E 100 (DA [12])

tà, sono in contatto con un numero elevatissimo di persone. Questi individui coesistono con altri che, all'estremo opposto, hanno pochissimi contatti. Per una rete di questo tipo, il grado medio m (sebbene matematicamente ben definito) non ha alcun valore informativo visto che gli scostamenti rispetto ad esso sono enormi: la rete è priva di una “scala caratteristica” e viene detta infatti *scale-free*. In un famoso articolo del 1999, Barabási e Albert [13] pro-

“ In un famoso articolo del 1999, Barabási e Albert proposero un semplice ed elegante algoritmo per creare artificialmente una rete *scale-free*. ”

Il numero di Erdős

Una particolare rete sociale è quella definita dalle *collaborazioni tra matematici*: ogni nodo rappresenta uno studioso e due nodi sono collegati da un *link* se i due studiosi hanno pubblicato almeno un articolo insieme. Sebbene il nucleo centrale connesso della rete (tralasciando cioè nodi isolati e piccoli spezzoni non connessi) comprenda ben 268 mila nodi, la distanza media tra due studiosi qualunque risulta decisamente bassa, pari circa a 7: un evidente indizio della proprietà *small-world*. Un nodo molto speciale è quello relativo a Paul Erdős (1913-1996), matematico ungherese estremamente prolifico (più di 1400 articoli pubblicati) ma soprattutto molto collaborativo (509 diversi coautori). Nel gergo delle reti, significa che il nodo Erdős ha 509 vicini o, in modo equivalente, che ha grado pari a 509. Preso un qualunque studioso, il suo *numero di Erdős* E è la distanza tra il nodo corrispondente allo studioso in questione e il nodo Erdős. Per quanto detto, esistono quindi 509 studiosi con $E = 1$. Grazie alla già citata proprietà *small-world* della rete, E può risultare sorprendentemente basso: chi scrive, sebbene pubblichi solo occasionalmente in riviste classificate di area matematica, ha $E = 4$, leggermente al di sotto del valore medio che risulta essere pari a 4.65. Per maggiori dettagli, nonché per il calcolo del proprio numero di Erdős, si consulti la pagina web del *Erdős Number Project* (<http://www.oakland.edu/enp/>).

posero un semplice ed elegante algoritmo per creare artificialmente una rete *scale-free*. Fissato il grado medio m (intero e pari, per semplicità), la rete viene costruita iterativamente aggiungendo ad ogni passo un nuovo nodo dotato di $m/2$ *link*, da connettere ad $m/2$ nodi della rete preesistente. L'ingrediente cruciale dell'algoritmo sta nella selezione di questi ultimi, che sono sorteggiati assegnando loro una probabilità di essere scelti proporzionale al loro grado corrente k_i . In altre parole, il nuovo nodo si collegherà preferenzialmente a nodi già molto connessi, con un effetto di reazione positiva che, in ultima analisi, risulta essere proprio il responsabile della creazione degli *hub*. Via via che si aggiungono nodi (quindi per $N \rightarrow \infty$), la distribuzione di grado tende ad una *legge di potenza* del tipo $P(k) \sim k^{-3}$. È importante notare che questa distribuzione ha varianza divergente, vale a dire $\sigma^2 = \sum_k (k - m)^2 P(k) \rightarrow \infty$ quando $N \rightarrow \infty$. Questa peculiarità non sussiste in una rete reale, caratterizzata da N finito, ma per reti di questo tipo, se N è molto grande, anche σ^2 tenderà ad essere molto grande con conseguenze interessanti, come vedremo, anche sui meccanismi di diffusione delle epidemie.

Un'altra caratteristica che emerge chiaramente dall'analisi dei dati rappresentativi di reti reali riguarda la distanza media d , spesso sorprendentemente piccola in confronto alla dimensione N della rete. Nell'ambito delle reti sociali, questa proprietà fu resa evidente fin dagli anni '60 del secolo scorso dall'ormai famoso esperimento di Milgram [14] che ha dato origine alla “teoria” dei “sei gradi di separazione”, entrata a pieno titolo nella cultura popolare. Secondo tale teoria, nella rete sociale costituita da tutti gli abitanti della Terra (in cui N vale quindi alcuni miliardi), la distanza tra due individui qualunque è mediamente pari a sei. In altre parole, assumendo collegati da un *link* due individui se legati da relazioni di familiarità, amicizia ecc., è possibile connettere due abitanti qualunque della Terra con una catena formata mediamente da sei *link*. Sebbene questa credenza sia basata sull'extrapolazione dei risultati di esperimenti effettuati su scala ben più ridotta, non è difficile credere che la “vera” distanza media d non sia mol-

“ Nel 1998, Watts e Strogatz fornirono una semplice spiegazione sperimentale della proprietà *small-world*, mostrando che d è piccola quando esiste una quantità anche molto esigua di connessioni “di lunga distanza”.

”

to superiore e comunque risulti di molti ordini di grandezza più piccola rispetto a N . Una rete con tale proprietà è detta *small-world*. Più precisamente, considerando un insieme di reti parametrizzato in N , si dice che l'insieme possiede la proprietà *small-world* se d cresce con $\log N$ e quindi molto più lentamente di N .

Nel 1998, Watts e Strogatz [15] fornirono una semplice spiegazione sperimentale della proprietà *small-world*, mostrando che d è piccola quando esiste una quantità anche molto esigua di connessioni “di lunga distanza”. L'esperimento è così congegnato: partendo da una rete a geometria regolare (Fig. 7), in cui cioè tutte le connessioni sono di tipo “locale”, si sorteggia una (piccola) frazione $0 < p < 1$ di *link*, si scollega uno a caso dei loro estremi e lo si riconnette ad un nodo della rete pure estratto a caso (“rewiring”). Come si può dedurre dal grafico in figura, è sufficiente $p = 0.001$ (solo un millesimo dei *link* è soggetto a *rewiring*) per dimezzare la distanza media d o $p = 0.01$ per abbatterla fino al 20% del valore originale. Si può dimostrare che, mentre $d \sim N$ nella rete regolare, risulta $d \sim \log N$ in quella soggetta a *rewiring*. Le connessioni di lunga distanza trasformano quindi la rete da regolare a *small-world* e, nel caso specifico della diffusione di epidemie, permettono di trasmettere rapidamente il patogeno in tutte le zone della rete: è ciò che è avvenuto nel caso della SARS citato in precedenza.

Diffusione di epidemie in rete

Una volta descritta la popolazione con un modello a rete, la propagazione di un'epidemia può essere vista come un processo dinamico che utilizza la rete come supporto. La classe più diretta ed efficace di modelli utilizzabili a questo scopo è quella degli *automi cellulari probabilistici*. Nella loro versione più semplice, si tratta di modelli a tempo discreto, in cui cioè i fenomeni avvengono in modo sincrono negli istanti $t = 0, \Delta, 2\Delta, \dots$ avendo scelto Δ come intervallo di discretizzazione. All'istante t , ogni nodo i si trova in uno stato x_t^i appartenente a un insieme finito $\Sigma = \{0, 1, \dots, q-1\}$, tipicamente di piccola cardinalità q . All'istante $t + \Delta$, il nodo i passerà in uno stato $x_{t+\Delta}^i$ dipendente, secondo regole probabilistiche da specificare, da x_t^i ma anche dallo stato dei nodi direttamente connessi a i (i cosiddetti “vicini” di i). L'automa cellulare è completamente definito una volta specificata la rete, attraverso la matrice di adiacenza A e le regole di interazione tra nodi vicini. Fissato lo stato iniziale x_0^i di tutti i nodi e lasciando evolvere il sistema, ciò che otteniamo è la realizzazione di un processo stocastico, visto che le regole di interazione sono probabilistiche. Il primo quesito che ci poniamo è più o meno ovvio: quali sono le proprietà collettive di tutte le possibili realizzazioni? Significa chiedersi se l'epidemia è in grado di propagarsi oppure no e, in caso positivo, a quali valori globali si attesta, quali nodi sono colpiti più di frequente ecc. Il secondo quesito è più intrigante: la risposta al primo quesito dipende dalle caratteristiche topologiche della rete? Ci chiediamo, cioè, se le caratteristiche strutturali della rete (la distribuzione di grado, ad esempio) abbiano un ruolo nell'esito del processo di propagazione.

Come abbiamo fatto in precedenza, per semplicità fissiamo la nostra attenzione sul processo *SIS*. Due sono gli stati possibili, suscettibile ed infetto, per cui poniamo $\Sigma = \{0, 1\}$ definendo $x^i = 0$ se i è suscettibile, $x^i = 1$ se i è infetto. Per il processo di guarigione $I \rightarrow S$, supponiamo che se il nodo i è infetto all'istante t allora con probabilità $\gamma\Delta$ sarà

FIGURA 7: CON UNA PICCOLA FRAZIONE p DI COLLEGAMENTI “A LUNGA DISTANZA”, UNA RETE REGOLARE (A) ASSUME LA PROPRIETÀ *SMALL-WORLD* (B): LA DISTANZA MEDIA $d(p)$ DECRESCE RAPIDAMENTE CON p (ELABORAZIONE DA [15])

Interventi

guarito (e quindi di nuovo suscettibile) all'istante $t + \Delta$, mentre con probabilità $(1 - \gamma\Delta)$ resterà infetto. Quanto al processo di infezione $S \rightarrow I$, assumiamo che la probabilità che il nodo i suscettibile diventi infetto sia pari a

$$\rho\Delta \sum_{j=1}^N a_{ij}x_t^j,$$

cioè proporzionale al numero di vicini infetti (si ricordi che $a_{ij} = 1$ se j è un vicino di i , altrimenti $a_{ij} = 0$). Questo insieme di regole è facilmente trasferibile in istruzioni in un ambiente di programmazione (Matlab, ad esempio): definita la rete attraverso la matrice di adiacenza, è possibile lanciare numerose simulazioni e analizzarne poi i risultati in vario modo. Ci chiediamo tuttavia se non sia possibile descrivere proprio a livello aggregato, eventualmente a prezzo di qualche approssimazione, i risultati che è lecito attendersi da tali simulazioni. Ciò significa scrivere un *modello di campo medio* di un processo per sua natura distribuito. Partiamo scrivendo un bilancio del numero di individui infetti $Y \in [0, N]$ tra due istanti (discreti) successivi:

$$Y(t + \Delta) = Y(t) - \gamma\Delta Y(t) + \rho\Delta\Phi(t)(N - Y(t))$$

dove $N - Y(t)$ è il numero di suscettibili, mentre $\Phi(t)$ è una stima del numero medio di vicini infetti di ciascun suscettibile. Si noti il passaggio dalla descrizione *microscopica* propria dell'automa cellulare, in cui il numero di vicini infetti è a priori diverso per ogni suscettibile, a quella *macroscopica* in cui siamo obbligati a considerare valori medi. Ora, se la distribuzione di grado è (quasi) *omogenea*, possiamo senz'altro assumere che ogni nodo abbia m vicini (il grado medio); inoltre, trascurando le eventuali disomogeneità spaziali, possiamo assumere che, di questi, sia infetta una frazione pari a $y(t) = Y(t)/N$, la stessa frazione cioè che caratterizza la rete nella sua globalità. Incorporando queste ipotesi nell'equazione di bilancio, dividendo entrambi i membri per N , e passando al limite per $\Delta \rightarrow 0$, otteniamo:

$$\dot{y}(t) = -\gamma y(t) + \rho m y(t)(1 - y(t)).$$

È immediato notare che questa equazione è la stessa del modello *SIS* "classico" discusso in precedenza, che quindi è valido non solo nell'ipotesi di popolazione perfettamente miscelata ma anche nel caso a rete in cui la struttura dei contatti è fissa purché tutti gli individui abbiano (approssimativamente) il medesimo numero di contatti. Confrontando le due scritture, notiamo che $\beta = \rho m$: la derivazione attraverso il modello a rete rende evidente come il coefficiente di trasmissione β sia composto da due fattori: il primo (ρ) prettamente biologico e rappresentativo della virulenza del patogeno, il secondo (m) legato alla struttura della rete sociale. Così, come predetto dal modello *SIS* "classico", anche in una rete omogenea l'epidemia potrà quindi

propagarsi e sopravvivere solo se $\beta > \beta_{cr} = \gamma$ quindi solo se $\rho > \rho_{cr} = \gamma/m$ mentre si estinguerà in caso contrario.

Le cose vanno però diversamente se la rete sociale ha distribuzione di grado fortemente *eterogenea* (rete *scale-free*), come messo in evidenza da Pastor-Satorras e Vespignani in un famoso articolo nel 2001 [16]. L'ipotesi che il numero atteso di vicini infetti sia descrivibile con $\Phi(t) = m y(t)$ non è più sostenibile, poiché il numero di vicini si discosta grandemente, da nodo a nodo, dalla media m . È allora necessario sostituire all'unica equazione in $y(t)$ un sistema di equazioni differenziali nelle variabili $y_k(t)$, $k = 1, 2, \dots$, una per ogni grado presente nella rete, in cui y_k descriva la frazione di infetti *nell'insieme dei nodi di grado k* :

$$\dot{y}_k(t) = -\gamma y_k(t) + \rho \Phi_k(t)(1 - y_k(t)), \quad k = 1, 2, \dots$$

Il termine $\Phi_k(t)$ rappresenta ora il numero atteso di vicini infetti di un nodo di grado k e, come si può mostrare, è proporzionale a k avendo una forma del tipo $\Phi_k(t) = k F(y_1(t), y_2(t), \dots)$. Infine, la frazione complessiva di nodi infetti nella rete si ottiene come $y(t) = \sum_k P(k) y_k(t)$.

“ Qualunque approccio si usi, i risultati concordano nell'evidenziare la scomparsa del valore-soglia ρ_{cr} per reti di dimensione N arbitrariamente grande: l'epidemia è in grado di sopravvivere e propagarsi nella rete anche se la sua virulenza è arbitrariamente bassa. ”

Le proprietà del modello appena introdotto possono essere studiate in parte per via analitica e in parte mediante un'analisi numerica: i risultati si affiancano alle simulazioni direttamente effettuate mediante automa cellulare probabilistico. Qualunque approccio si usi, i risultati concordano nell'evidenziare la scomparsa del valore-soglia ρ_{cr} per reti di dimensione N arbitrariamente grande: l'epidemia è in grado di sopravvivere e propagarsi nella rete anche se la sua virulenza è arbitrariamente bassa. Più precisamente, si può dimostrare (mediante analisi di stabilità) che la soglia si colloca, per una qualunque distribuzione di grado, al valore:

$$\rho_{cr} = \frac{\gamma m}{\sigma^2 + m^2}$$

dove m e σ^2 sono, rispettivamente, media e varianza della distribuzione di grado. In una rete strettamente omogenea $\sigma^2 = 0$ e quindi $\rho_{cr} = \gamma/m$ come già visto. In una rete *scale-free* del tipo sopra discusso, invece, quando $N \rightarrow \infty$ abbiamo $\sigma^2 \rightarrow \infty$ e quindi $\rho_{cr} \rightarrow 0$. In concreto, in reti di grandi dimensioni, il valore-soglia ρ_{cr} risulterà estremamente piccolo, seppur non nullo.

FIGURA 8: IL VALORE-SOGLIA ρ_{cr} , SOTTO IL QUALE L'EPIDEMIA SI ESTINGUE, È TANTO MINORE QUANTO PIÙ È ELEVATA LA VARIANZA DELLA DISTRIBUZIONE DI GRADO $P(k)$. QUINDI IN RETI ETEROGENEE SI PUÒ AVERE $\rho_{cr} \rightarrow 0$ QUANDO $N \rightarrow \infty$: L'EPIDEMIA È IN GRADO DI SOPRAVVIVERE E PROPAGARSI ANCHE CON VIRULENZA ARBITRARIAMENTE BASSA (ELABORAZIONE DA [16])

Come appare evidente dalla Fig. 8, per ρ molto piccolo il valore y a cui si assesta a regime l'epidemia in una rete fortemente eterogenea è non nullo, ma decisamente basso. Inoltre, con alcuni passaggi algebrici si può dimostrare che, sempre all'equilibrio, la frazione y_k di nodi di grado k infetti vale:

$$y_k = \frac{1}{1 + \alpha/k}$$

dove α è funzione dei dati del problema ma non dipende da k . Quindi la frazione di nodi infetti cresce monotonamente con k e $y_k \rightarrow 1$ per k arbitrariamente grande. In sintesi, tanto più consideriamo valori elevati del grado k , tanto meno saranno i nodi che troviamo ($P(k) \rightarrow 0$) ma tanto più sarà elevata la probabilità che essi siano infetti $y_k \rightarrow 1$. È questo il meccanismo alla base dell'infezione in una rete eterogenea: se l'epidemia è presente nella rete, grazie al loro elevatissimo numero di connessioni gli *hub* hanno alta probabilità di infettarsi. In qualche modo, sono loro i "custodi" dell'epidemia che quindi riesce a non estinguersi seppure sopravvivendo a livelli modesti. È in effetti questo lo scenario riscontrabile in alcuni tipi di patologie [17] nonché, passando ad ambiti tecnologici, la fenomenologia tipica della diffusione e persistenza di virus informatici nella rete Internet [16].

L'esempio qui analizzato, sebbene relativamente semplice, è sufficiente a mettere in evidenza come, passando dai modelli "classici" ai modelli "a rete", si possano ottenere risultati inaspettati e in ogni caso più complessi ed articolati. Innumerevoli sono le direzioni in cui si è mossa la ricerca negli ultimi dieci anni e in cui si muoverà nell'immediato futuro. Dal lato dell'epidemia, modelli più complessi del semplice processo *SIS* sono utilizzati, come è facilmente intuibile, per descrivere le diverse caratteristiche delle molteplici patologie esistenti. Dal lato della rete sociale, è evidente come una descrizione "statica" come quella qui presentata, in cui cioè gli individui sono sempre quelli e ciascuno di essi mantiene esattamente gli stessi contatti per sempre, costituisca un modello troppo semplificato. Descrizioni più realistiche devono necessariamente includere i processi di

nascita/morte degli individui (con conseguente (s)collegamento dei corrispondenti *link*), la loro mobilità (la rete diventa variante nel tempo) o, nel caso più complesso, l'interazione reciproca tra epidemia e topologia: le connessioni variano nel tempo in funzione dello stato epidemiologico degli individui coinvolti (reti adattative). Tutte queste estensioni, come è facile intuire, rendono via via più difficile e complesso lo studio. Ma ne vale sicuramente la pena: nel loro complesso, le malattie infettive causano in tutto il mondo più di 15 milioni di morti l'anno (il 25% del totale delle morti [17]). Lavoriamo affinché lo studio congiunto di epidemie e reti, mediato dai modelli matematici, possa contribuire ad abbassare questa drammatica cifra. ■

Bibliografia

- [1] Wasserman S., Faust K., *Social network Analysis*, Cambridge University Press, 1994.
- [2] Kermack W.O., McKendrick A.G., "A contribution to the mathematical theory of epidemics", *Proceedings of the Royal Society A*, 115, pp. 700-721, 1927.
- [3] Anderson R., May R., *Infectious Diseases of Humans: Dynamics and Control*, Oxford University Press, 1991.
- [4] Eubank S., Guclu H., Kumar V.S.A., Marathe M.V., Srinivasan A., Toroczkai Z., Wang N., "Modelling disease outbreaks in realistic urban social networks", *Nature*, 429, pp. 180-184, 2004.
- [5] Liljeros F., Edling C.R., Amaral L.A.N., Stanley H.E., Åberg Y., "The web of human sexual contacts", *Nature*, 411, pp. 907-908, 2001.
- [6] *Oxford Atlas of World History*, Oxford University Press, p. 105, 1999.
- [7] Banos A., Lacasa J., "Spatio-temporal exploration of SARS epidemic", *Cybergeog: European Journal of Geography* [online], URL: <http://cybergeog.revues.org/12803>, 2007.
- [8] Newman M.E.J., "The structure and function of complex networks", *SIAM Review*, 45, pp. 167-256, 2003.
- [9] Boccaletti S., Latora V., Moreno Y., Chavez M., Hwang D.H., "Complex networks: structure and dynamics", *Physics Reports*, 424, pp. 175-308, 2006.
- [10] Barrat A., Barthélemy M., Vespignani A., *Dynamical Processes on Complex Networks*, Cambridge University Press, 2008.
- [11] Newman, M.E.J., *Networks: An Introduction*, Oxford University Press, 2010.
- [12] Wang X.F., Chen G., "Complex networks: small-world, scale-free and beyond", *IEEE Circuits and Systems Magazine*, pp. 6-20, 2003.
- [13] Barabási A.L., Albert R., "Emergence of scaling in random networks", *Science*, 286, pp. 509-512, 1999.
- [14] Milgram S., "The small world problem", *Psychology Today*, 2, pp. 60-67, 1967.
- [15] Watts D.J., Strogatz S.H., "Collective dynamics of 'small-world' networks", *Nature*, 393, pp. 440-442, 1998.
- [16] Pastor-Satorras R., Vespignani A., "Epidemic spreading in scale-free networks", *Physical Review Letters*, 86, pp. 3200-3203, 2001.
- [17] Morens D.M., Folkers G.K., Fauci A.S., "The challenge of emerging and re-emerging infectious diseases", *Nature*, 430, pp. 242-249, 2004.