

ALMA MATER STUDIORUM · UNIVERSITÀ DI BOLOGNA

SCUOLA DI SCIENZE

Corso di Laurea magistrale in Matematica

Un modello su network
per malattie neurodegenerative

Relatore:
Chiar.ma Prof.ssa
Maria Carla Tesi

Presentata da:
Alessandra Piroddi

Sessione Unica
Anno Accademico 2021/2022

Indice

Introduzione	7
1 Prerequisiti matematici	11
1.1 Equazione di Fisher-KPP	11
1.2 Generalità e Laplaciano sul Grafo	16
1.2.1 Laplaciano sul Grafo e autovalori	19
1.3 Funzione di Lambert	23
2 Modello su network e metrica su un grafo per le stime dei tempi d'arrivo	27
2.1 Fisher-KPP su un grafo	28
2.2 Analisi lineare del modello	31
2.3 Distanza di Lambert	35
2.3.1 Luoghi di semina multipli	42
2.3.2 Discussione e confronto con la distanza effettiva	43
3 Applicazione a malattie neurodegenerative	45
3.1 Scale temporali di progressione della malattia	46
3.2 Stadiazione Braak	50
A Analisi qualitativa di un sistema bidimensionale	53
Bibliografia	57

Elenco delle figure

1.1	Piano delle fasi per il caso $0 < c < 4$	15
1.2	Piano delle fasi nel caso di $c \geq 2$	16
1.3	Grafico diramazioni della funzione di Lambert W	24
2.1	Grafico delle concentrazioni per il regime dominato dalla crescita e dalla diffusione	30
2.2	Confronto tra soluzioni esatte e lineari	34
2.3	Confronto tra soluzioni di Lambert, lineari ed esatte	41
2.4	Confronto distanza di Lambert ed effettiva	44
3.1	Grafico delle concentrazioni in funzione delle scale temporali .	49
3.2	Stadiazione Braak dell'AD	51
3.3	Grafico delle concentrazioni di τ e stadiazione Braak	52
A.1	Classificazione delle singolarità dei punti stazionari	55

Introduzione

Viviamo in un mondo connesso in cui persone, beni, informazioni e malattie viaggiano da una regione all'altra, senza barriere. Negli anni del *COVID-19* abbiamo assistito ad un fenomeno di propagazione che ne è la prova: i primi casi di coronavirus coinvolsero i lavoratori del mercato umido di Wuhan (in Cina) nel gennaio del 2020 e a già a fine febbraio si registrarono contagi in diversi Paesi europei. Un fenomeno simile si manifesta anche in un sistema molto più piccolo, il cervello umano. Si ritiene, infatti, che alcune proteine tossiche associate a malattie neurodegenerative come il morbo di Parkinson o di Alzheimer abbiano origine in una singola regione e che vengano poi trasportate al resto del cervello attraverso il *connettoma strutturale*, la rete di percorsi assonali che collegano le diverse aree dell'encefalo. Da un punto di vista matematico, si parla di propagazione di processi su una rete ed il nostro interesse è volto alla comprensione delle dinamiche di progressione della malattia di Alzheimer (AD). Nonostante sia la più comune forma di demenza senile, le cause ed i meccanismi evolutivi del morbo non sono ancora noti nella loro interezza e sono attualmente oggetto di ricerca scientifica. Tuttavia, si ritiene che l'accumulo di aggregati mal ripiegati di proteina *tau* nel cervello sia associato alla perdita delle funzioni cognitive del paziente. Pertanto, è importante comprenderne le dinamiche di trasporto.

Nel secondo Capitolo di questo elaborato, seguendo la ricerca di P. Putra, H. Oliveri, T. Thompson e A. Goriely in [1] (2022), presentiamo un modello matematico per un processo di invasione su networks ed analizziamo due differenti metodi utilizzati per ottenere stime dalla dinamica. Come vedre-

mo, il modello proposto approssima piuttosto fedelmente la propagazione dei grovigli neurofibrillari di τ nel cervello di un paziente affetto da AD. Nel loro lavoro, Putra & co, modellano il cervello del paziente con un *network cerebrale*, ossia un grafo, ed assumono che le dinamiche di invasione su di esso siano governate dall'equazione diffusione-reazione di *Fisher-KPP*. Quando la diffusione è sufficientemente piccola rispetto alla dinamica di crescita, come accade nel caso della propagazione degli aggregati di τ [16], il grafo viene sistematicamente invaso, nodo dopo nodo, ed il processo può essere caratterizzato dal tempo di arrivo, ad ogni nodo, di una concentrazione tossica non nulla. Pertanto, una questione importante è determinare una stima del tempo che impiega una quantità di interesse, in una regione di semina, a raggiungere un certo livello di concentrazione tossica in ogni nodo e quindi a portare alla demenza cerebrale. Un primo approccio al problema consiste nella linearizzazione dell'intero sistema di equazioni differenziali attorno ad un opportuno stato di equilibrio. La concentrazione di carico tossico in ogni nodo si ottiene come soluzione di questo sistema lineare per una data condizione iniziale. A partire da questa soluzione, i *tempi d'arrivo*, da un nodo seme ad un nodo *target*, si determinano risolvendo un'equazione trascendentale in ogni nodo. Come vedremo, il metodo lineare fornisce un limite inferiore per i *tempi d'arrivo* ed ha il vantaggio di catturare la topologia del *network* attraverso l'esponenziale del Laplaciano del grafo. Tuttavia, in assenza di buone ipotesi iniziali, ottenere le stime temporali potrebbe risultare piuttosto difficile.

Applicando, però, il metodo lineare ad un sottografo costituito da un'unica coppia di nodi vicini, cioè vertici collegati da un solo arco, è possibile ottenere una nozione di distanza tra nodi adiacenti. Tale distanza prende il nome di *distanza d'arco di Lambert*, in quanto ottenuta in termini di *funzione di Lambert*, la quale è definita essere l'inversa multivariata della funzione $\omega \rightarrow \omega e^\omega$ e di cui tratteremo più approfonditamente nel primo Capitolo, dedicato alle nozioni preliminari. A partire da questa nozione di distanza tra nodi vicini, muniamo il grafo di una nuova metrica, la *metrica di Lambert*. Definia-

mo quindi la *distanza di Lambert* tra due nodi del grafo (non necessariamente adiacenti) come la lunghezza del più corto cammino, rispetto alla *distanza d'arco di Lambert*, che li collega. Approssimando poi il tempo che impiega il carico tossico di una regione di semina a raggiungere un'altra regione del grafo con la *distanza di Lambert* tra questi due nodi, otteniamo nuove stime per i *tempi d'arrivo*. Come vedremo, quello di Lambert è un metodo particolarmente efficace che può essere utilizzato per fornire un'approssimazione complessiva delle dinamiche del processo.

Infine, nel terzo Capitolo, applichiamo il modello proposto al fenomeno di propagazione degli aggregati tossici di *tau* ed utilizziamo il metodo di Lambert per comprenderne le dinamiche, ottenendo così scale temporali caratteristiche, come il tempo d'arrivo al primo nodo invaso ed il *periodo pandemico* della malattia, in seguito al quale si registra una concentrazione tossica non nulla in tutte le regioni cerebrali. Inoltre, le simulazioni ottenute risolvendo il modello di *Fisher-KPP* sui connettomi cerebrali di alcuni pazienti, dimostrano che il metodo di Lambert è in grado di cogliere caratteristiche essenziali della malattia, come i corretti tempi d'arrivo ad ogni regione ed il corretto ordinamento dei sei *stadi di Braak*.

Capitolo 1

Prerequisiti matematici

1.1 Equazione di Fisher-KPP

In ambito biologico e biomedico è molto diffusa l'esistenza di fenomeni ondulatori. Sono numerosi, infatti, i processi di sviluppo biologici e i modelli di dinamica di popolazione caratterizzati dalla comparsa di onde viaggianti di segnali elettrici, concentrazioni chimiche, densità di popolazione e così via. Per poter descrivere in modo accurato fenomeni di questo tipo, è necessario approfondire lo studio delle *travelling waves* ed utilizzare modelli matematici che ammettono soluzioni di questo tipo (per un background più completo si veda il Capitolo 13 di [2]).

Con il termine *travelling wave* intendiamo una soluzione di un'equazione differenziale parziale (PDE) che si propaga con una velocità costante, senza cambiare forma. In particolare, se la soluzione $u(x, t)$ di una PDE può essere scritta come segue

$$u(x, t) = u(x - ct) = u(z) \quad \text{con } z = x - ct$$

allora u viene detta *travelling wave* e si propaga nello spazio con velocità c . In base alla forma d'onda possiamo distinguere in soluzioni di tipo

- *wavefront*, se $\lim_{z \rightarrow -\infty} u(z) = u_l$ e $\lim_{z \rightarrow \infty} u(z) = u_r$ con $u_l \neq u_r$;
- *pulse wave*, se $u_l = u_r$;

- *spatially periodic wave*, se $u(z + T) = u(z) \forall z$, con $T > 0$.

La classe di equazioni paraboliche di tipo diffusione-reazione,

$$\frac{\partial u}{\partial t} - D \frac{\partial^2 u}{\partial x^2} = f(u) \quad (1.1)$$

con D costante di *diffusione* e $f(u)$ *termine di reazione* (o *termine sorgente*), assume un ruolo di rilievo in questo ambito in quanto, a seconda della forma di $f(u)$, questo tipo di equazione può presentare soluzioni *travelling waves*.

Il nostro interesse sarà volto allo studio e all'analisi dell'equazione di Fisher–Kolmogorov–Petrovsky–Piskunov (*Fisher-KPP*), che è considerata il prototipo delle equazioni che ammettono soluzioni *travelling wavefront*. È la più semplice equazione di tipo *diffusione-reazione* con termine sorgente non lineare, della forma:

$$\frac{\partial u}{\partial t} = D \frac{\partial^2 u}{\partial x^2} + \lambda u (1 - u) \quad (1.2)$$

dove,

- $u = u(x, t) \geq 0$ è una funzione che dipende dalla posizione $x \geq 0$ e dal tempo $t > 0$;
- $D > 0$ è la costante di *diffusione* ;
- $\lambda > 0$ la costante di crescita.

Si tratta di un'estensione dell'*equazione logistica*, $\frac{\partial u}{\partial t} = \lambda u (1 - \frac{u}{k})$, ed è stata suggerita da Fisher come la versione deterministica di un modello stocastico per la diffusione di un gene all'interno di una popolazione [2].

Essendo un'equazione non lineare, non è semplice determinarne una soluzione numerica accurata. Forniamo, quindi, un'analisi qualitativa del modello e determiniamo per quali valori della velocità di propagazione c , la *Fisher-KPP* ammette soluzioni d'onda. Iniziamo sostituendo i valori

$$t^* = \lambda t \quad x^* = x \left(\frac{\lambda}{D} \right)^{\frac{1}{2}}$$

in (1.2) e, rinominando $t^* = t$ ed $x^* = x$, riscriviamo la *Fisher-KPP* nella forma seguente

$$\frac{\partial u}{\partial t} - \frac{\partial^2 u}{\partial x^2} = u(1 - u). \quad (1.3)$$

Come prima cosa osserviamo che, in caso di omogeneità spaziale, gli stati di equilibrio sono $u = 0$ ed $u = 1$ (zeri del termine sorgente $f(u) = u(1 - u)$) ed essendo $f'(u) = 1 - 2u$ con

$$f'(0) = 1 > 0, \quad f'(1) = -1 < 0$$

per il Teorema A.7 in Appendice, possiamo concludere che $u = 0$ è *instabile* mentre $u = 1$ è *stabile*.

Se esistono soluzioni u di tipo *travelling waves*, queste sono della forma $u(x, t) = U(z)$ con $z = x - ct$. Assumiamo $c \geq 0$ e, sostituendo in (1.3), si ha che U soddisfa la seguente equazione

$$U'' + cU' + U(1 - U) = 0 \quad (1.4)$$

dove, U' indica la derivata della soluzione U rispetto alla variabile z .

Tipicamente u rappresenta la concentrazione di una sostanza o la densità di una popolazione, dunque, per ottenere una soluzione fisicamente significativa, ha senso supporre che $U(z)$ sia non negativo e limitato.

Per esaminare il comportamento della soluzione, studiamo il piano delle fasi imponendo $U' = V$ e riscrivendo (1.4) come sistema di due equazioni del 1° ordine:

$$\begin{cases} U' = V \\ V' = -cV - U(1 - U). \end{cases} \quad (1.5)$$

Dalla definizione di *punto di equilibrio* (si veda A.2 in Appendice), è facile osservare che il sistema ammette due stati stazionari $(U, V) = (0, 0)$ e $(U, V) = (1, 0)$, di cui ci interessa studiarne la stabilità. Seguendo quanto

riportato in Appendice A, calcoliamo la matrice Jacobiana del sistema

$$J(U, V) = \begin{pmatrix} 0 & 1 \\ 2U - 1 & -c \end{pmatrix}$$

e, per i due stati di equilibrio, si ha

$$J(0, 0) = \begin{pmatrix} 0 & 1 \\ -1 & -c \end{pmatrix} \quad \text{e} \quad J(1, 0) = \begin{pmatrix} 0 & 1 \\ 1 & -c \end{pmatrix}$$

i cui autovalori sono,

$$(0, 0) : \lambda_{\pm} = \frac{1}{2} \left[-c \pm (c^2 - 4)^{\frac{1}{2}} \right]$$

$$(1, 0) : \lambda_{\pm} = \frac{1}{2} \left[-c \pm (c^2 + 4)^{\frac{1}{2}} \right]$$

e, sempre facendo riferimento ad A, possiamo concludere che

- $(1, 0)$ è *punto di sella* $\forall c \geq 0$
- $(0, 0)$ è $\begin{cases} \text{nodo stabile} & \text{se } c^2 > 4 \\ \text{fuoco stabile} & \text{se } c^2 < 4 \end{cases}$

Per poter raffigurare in modo accurato le orbite del sistema e quindi determinare quando esistono soluzioni d'onda, distinguiamo in due possibili range di valori per c .

Caso $0 < c < 2$:

Per verificare l'esistenza di soluzioni d'onda, analizziamo il piano delle fasi riportato in Figura 1.1. La varietà stabile del punto $(1, 0)$ (in blu) divide il piano in due regioni. Le traiettorie alla sua sinistra si muovono a spirale verso l'origine (*fuoco stabile*) assumendo valori negativi, per qualche z . Dunque,

per le assunzioni fatte, le orbite in questa regione non sono fisicamente rilevanti. Per quanto riguarda l'altra regione, vi sono traiettorie che giacciono nel I quadrante ed orbite che dal semipiano negativo $U' < 0$ si muovono verso la porzione di piano ($U \geq 1, U' > 0$). Pertanto, le soluzioni corrispondenti a queste curve di fase saranno caratterizzate da una crescita illimitata e quindi non rilevanti nel nostro contesto.

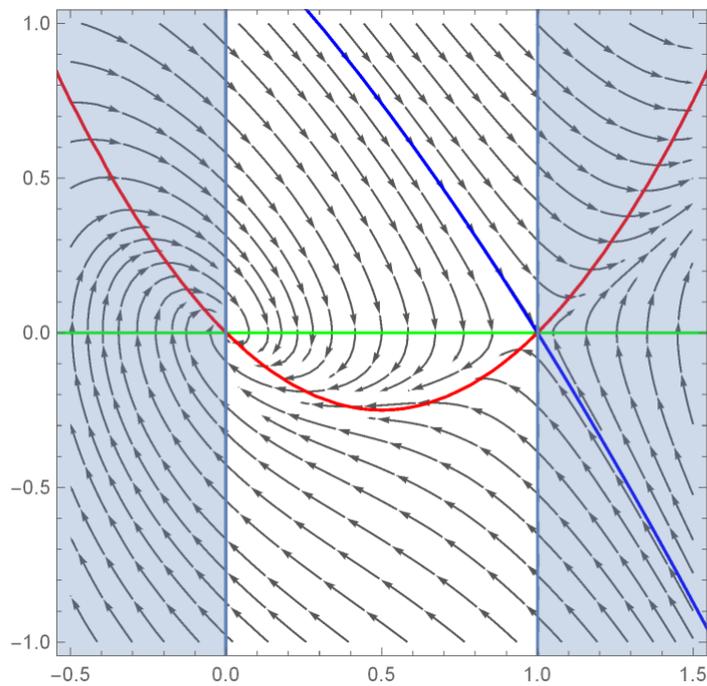


Figura 1.1: Piano delle fasi (U, V) per $c = 1$. In verde la U -nullocina, in rosso la V -nullocina ed in blu la varietà stabile di $(1, 0)$.

Caso $c \geq 2$:

La Figura 1.2 (a) mostra alcune traiettorie nel piano delle fasi (U, U') .

Osservando il grafico, vediamo che per $c \geq 2$ esiste una traiettoria da $(1, 0)$ a $(0, 0)$ che giace nel IV quadrante ($U \geq 0$ e $U' \leq 0$), con $0 \leq U \leq 1$. Ciò significa che la soluzione corrispondente sarà decrescente e assumerà valori nell'intervallo $[0, 1]$. Come si evince dal grafico riportato in Figura 1.2 (b) e

dalla classificazione in 1.1, si tratta di una soluzione *travelling wavefront*.

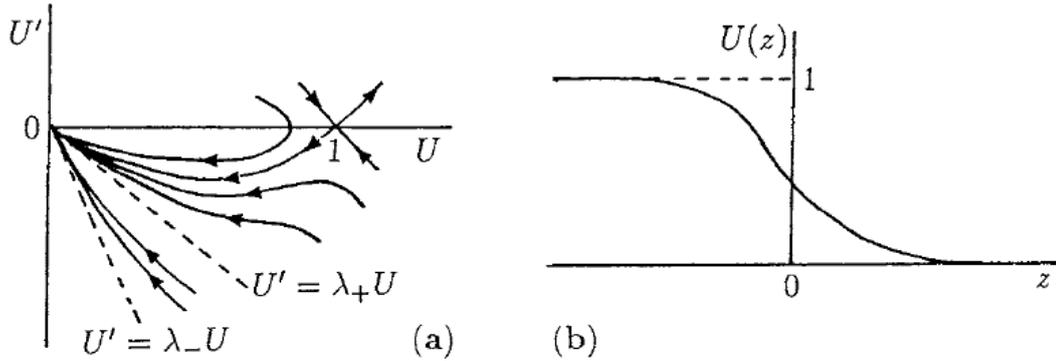


Figura 1.2: (a) Piano delle fasi nel caso di $c \geq 2$, con origine nodo stabile. (b) Grafico della soluzione *travelling wavefront* per l'equazione di *Fisher-KPP*.

Possiamo quindi concludere che $\forall c \geq c_{min} = 2$, l'equazione di *Fisher-KPP* (1.3) ammette soluzioni di tipo *travelling waves* fisicamente realistiche e tali che

$$\lim_{z \rightarrow -\infty} U(z) = 1 \text{ e } \lim_{z \rightarrow \infty} U(z) = 0 .$$

In termini dell'iniziale equazione (1.2) non riscalata, il range di valori di c per cui esistono soluzioni d'onda è $c \geq 2\sqrt{\lambda D}$.

Per una dimostrazione dettagliata sull'esistenza di soluzioni d'onda o per ulteriori approfondimenti sulla *Fisher-KPP* si fa riferimento a [2, 3].

1.2 Generalità e Laplaciano sul Grafo

In termini non rigorosi possiamo pensare il grafo come una struttura costituita da *vertici* (o *nodi*) collegati tra loro da linee (dette *archi*). Nel caso in cui questi ultimi siano dotati di un verso, si parlerà di *archi orientati*, altrimenti di *archi non orientati*. Generalmente, un grafo viene raffigurato

sul piano da cerchietti, i *vertici*, collegati da segmenti o curve, nel caso di *archi non orientati*, o da frecce, se il grafo è *orientato*.

Formalizziamo tali concetti dando le seguenti definizioni:

Definizione 1.1. Un grafo finito è una coppia $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, dove $\mathcal{V} = \{v_1, \dots, v_N\}$ indica l'insieme dei *vertici* ed \mathcal{E} l'insieme degli *archi*. Più precisamente, \mathcal{E} è formato da coppie di vertici collegati da un arco.

Definizione 1.2. Sia $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ un grafo finito.

\mathcal{G} viene detto *orientato* se $\mathcal{E} \subset \{(v_i, v_j) \in \mathcal{V} \times \mathcal{V} \text{ con } i \neq j\}$ con (v_i, v_j) coppie ordinate di vertici. Mentre, \mathcal{G} viene detto *non orientato* se le coppie di vertici $\{v_i, v_j\}$ sono non ordinate.

D'ora in avanti, se non specificato, considereremo solo grafi finiti e non orientati.

Definizione 1.3. Sia $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ un grafo. Definiamo *ordine* di \mathcal{G} il numero dei suoi *vertici*, $|\mathcal{V}|$, e *dimensione* di \mathcal{G} il numero dei suoi *archi*, $|\mathcal{E}|$.

Definizione 1.4. Sia $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ un grafo, con $\mathcal{V} = \{v_1, \dots, v_N\}$ insieme dei vertici. Si definisce *grado* del vertice v_i e si indica con d_{v_i} , il numero di archi ad esso collegati.

Se $d_{v_i} = 0$, v_i si dice *vertice isolato*.

Definizione 1.5. Un arco che ha due estremi coincidenti (cioè connessi allo stesso vertice) si dice *cappio*.

Definizione 1.6. Sia $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ un grafo. Due vertici v_i e v_j si dicono *adiacenti* se sono collegati da un arco, cioè se $\{v_i, v_j\} \in \mathcal{E}$.

In simboli: $v_i \sim v_j$.

Definizione 1.7. Sia $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ un grafo. Viene definito *cammino* sul grafo, una qualsiasi successione finita di vertici $v_1 v_2 \dots v_n$ tale che $v_i \sim v_{i+1}$.

Definizione 1.8. Un grafo $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ si dice *connesso* se per ogni coppia di vertici $\{u, v\}$ esiste un cammino $\Gamma = (v_0 v_1 \dots v_m)$ che collega u a v ($v_0 = u$ e $v_m = v$).

Definizione 1.9. Si dice che $\mathcal{G}' = (\mathcal{V}', \mathcal{E}')$ è un sottografo di $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ se $\mathcal{V}' \subset \mathcal{V}$ e $\mathcal{E}' \subset \mathcal{E}$.

Definizione 1.10. Sia $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ un grafo. Una *componente connessa* di \mathcal{G} è un suo sottografo connesso e massimale.

Ad ogni arco di un grafo può essere associato un valore reale, chiamato *peso*. In questo caso, il grafo si dice *pesato* e può essere rappresentato in modo compatto dalla matrice dei pesi ad esso associata. Più precisamente,

Definizione 1.11. Dato un grafo orientato $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, un *peso* su \mathcal{G} è una funzione $\omega: \mathcal{V} \times \mathcal{V} \rightarrow \mathbb{R}$ tale che:

1. $\omega(u, v) \geq 0 \quad \forall u, v \in \mathcal{V}$
2. $\omega(u, v) = \omega(v, u) \quad \forall u, v \in \mathcal{V}$
3. $\omega(u, v) > 0 \iff \{u, v\} \in \mathcal{E}$

Un grafo a cui viene assegnato un *peso* si dice *pesato*.

Osservazione 1.12. Per la simmetria di ω (prop. (2)), la definizione di *peso* continua a valere anche per un grafo non orientato.

Osservazione 1.13. Il peso $\omega: \mathcal{V} \times \mathcal{V} \rightarrow \mathbb{R}$ definito da

$$\omega(u, v) = \begin{cases} 1 & \text{se } \{u, v\} \in \mathcal{E} \\ 0 & \text{altrimenti} \end{cases}$$

viene detto *peso standard*.

Definizione 1.14. Sia $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \omega)$ un grafo pesato di ordine N . Ad esso, partendo dalla funzione peso ω , è possibile associare la matrice dei pesi

$$W = (\omega_{i,j})_{i,j=1,\dots,N}$$

detta *matrice di adiacenza ponderata*. Essendo il grafo non orientato, W è simmetrica. Nel caso di ω *peso standard* W si indica con A e si denomina *matrice di adiacenza*.

Definizione 1.15. In un grafo pesato $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \omega)$ di ordine N , il *grado* di un vertice v_i è dato da

$$d_{v_i} = \sum_{k=1}^N \omega(v_i, v_k)$$

1.2.1 Laplaciano sul Grafo e autovalori

Consideriamo $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \omega)$ grafo pesato di ordine N , non orientato, senza *cappi* e che non ammette vertici isolati. Sia $W = (\omega_{i,j})_{i,j=1,\dots,N}$ la sua matrice di adiacenza ponderata. Oltre a W , possiamo associare a \mathcal{G} la matrice diagonale D tale che $D_{i,i} = d_{v_i} \forall i = 1, \dots, N$. D è invertibile, in quanto \mathcal{G} non ammette vertici isolati ($d_{v_i} > 0 \forall i = 1, \dots, N$).

Definizione 1.16. Sia $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \omega)$ come sopra. Si definisce **Laplaciano** di \mathcal{G} la matrice $L = D - W$. Dati quindi $x, y \in \mathcal{V}$

$$L(x, y) = \begin{cases} d_x & \text{se } x = y \\ -\omega(x, y) & \text{se } x \neq y \end{cases}$$

In modo del tutto equivalente,

$$L_{i,j} = -\omega_{i,j} + \delta_{i,j} \sum_{k=1}^N \omega_{i,k}$$

Definizione 1.17. Sia $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \omega)$ come sopra. Si definisce **Laplaciano normalizzato** di \mathcal{G} la matrice $\mathcal{L} = D^{-1/2} L D^{-1/2}$,

$$\mathcal{L}(x, y) = \begin{cases} 1 & \text{se } x = y \\ -\frac{\omega(x,y)}{\sqrt{d_x d_y}} & \text{altrimenti.} \end{cases}$$

La definizione di \mathcal{L} è ben posta, poichè D è non singolare.

Possiamo pensare il Laplaciano L come un operatore definito sullo spazio di funzioni $\{g: \mathcal{V} \rightarrow \mathbb{R}\}$ tale che, data $f: \mathcal{V} \rightarrow \mathbb{R}$,

$$Lf(x) = \sum_{y \in \mathcal{V}} L(x, y)f(y) = d_x f(x) - \sum_{y \in \mathcal{V}} \omega(x, y)f(y) \quad \forall x \in \mathcal{V} .$$

Teorema 1.18. *Sia $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \omega)$ un grafo pesato ed L Laplaciano di \mathcal{G} . Sia $f: \mathcal{V} \rightarrow \mathbb{R}$, allora*

$$f^T Lf = \frac{1}{2} \sum_{x, y \in \mathcal{V}} (f(x) - f(y))^2 \omega(x, y) \quad (1.6)$$

Dimostrazione.

$$\begin{aligned} f^T Lf &= \sum_{x, y \in \mathcal{V}} L(x, y)f(x)f(y) \\ &= \sum_{x \in \mathcal{V}} f(x)^2 d_x - \sum_{x, y \in \mathcal{V}} f(x)f(y)\omega(x, y) \\ &\stackrel{\text{def. di } d_x}{=} \sum_{x, y \in \mathcal{V}} f(x)^2 \omega(x, y) - \sum_{x, y \in \mathcal{V}} f(x)f(y)\omega(x, y) \\ &= \frac{1}{2} \left(\sum_{x, y \in \mathcal{V}} f(x)^2 \omega(x, y) - 2 \sum_{x, y \in \mathcal{V}} f(x)f(y)\omega(x, y) + \sum_{x, y \in \mathcal{V}} f(y)^2 \omega(x, y) \right) \\ &= \frac{1}{2} \sum_{x, y \in \mathcal{V}} (f(x)^2 - 2f(x)f(y) + f(y)^2) \omega(x, y) \\ &= \frac{1}{2} \sum_{x, y \in \mathcal{V}} (f(x) - f(y))^2 \omega(x, y) \end{aligned}$$

□

Vogliamo studiare lo spettro del Laplaciano di \mathcal{G} . Per iniziare, osserviamo che dalla simmetria di W segue quella di L , dunque L possiede N autovalori reali contati con molteplicità e ammette una base ortonormale di autovettori. Inoltre, da (1.6) si deduce che L è semidefinita positiva quindi i suoi autovalori sono tutti positivi o nulli. Dunque indichiamo con

$$\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_N$$

gli autovalori di L contati con molteplicità.

Dalla definizione di L segue che $\lambda_1 = 0$ con $(1, \dots, 1)$ autovettore relativo

all'autovalore nullo e la sua molteplicità è legata al numero di componenti connesse del grafo. Vediamo in che senso:

Lemma 1.19. *Sia $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \omega)$ un grafo pesato, che non ammette vertici isolati. Siano C_1, \dots, C_m con $m \geq 1$ le componenti connesse di \mathcal{G} , sia L Laplaciano su \mathcal{G} ed $f: \mathcal{V} \rightarrow \mathbb{R}$. Allora,*

$$Lf = 0 \iff f \text{ è costante sulle componenti connesse di } \mathcal{G}$$

cioè, $\text{Ker}(L) = \{f: \mathcal{V} \rightarrow \mathbb{R} \text{ costanti sulle componenti connesse di } \mathcal{G}\}$

Dimostrazione. (\Leftarrow)

$$\forall x \in \mathcal{V} \quad Lf(x) = d_x f(x) - \sum_{y \in \mathcal{V}} \omega(x, y) f(y) = \sum_{x, y \in \mathcal{V}} (f(x) - f(y)) \omega(x, y)$$

Siano, poi, C_1, \dots, C_m con $m \geq 1$ le componenti connesse di \mathcal{G} .

Se $x \in C_i$, $\omega(x, y) = 0 \forall y \notin C_i$ per definizione di componente connessa.

Dunque,

$$\forall x \in C_i, \quad Lf(x) = \sum_{x, y \in C_i} (f(x) - f(y)) \omega(x, y)$$

e se f è costante sulle componenti connesse, allora $Lf(x) = 0 \forall x \in \mathcal{V}$.

(\Rightarrow)

$$\text{Se } Lf = 0 \Rightarrow f(x) = \sum_{y \in \mathcal{V}} \omega(x, y) \frac{f(y)}{d_x} \quad \forall x \in \mathcal{V}$$

Sia $s \in C_i$ tale che $f(x) \leq f(s) \forall x \in C_i$. Se $z \in C_i$ è adiacente ad s ($z \sim s$), allora $f(s) = f(z)$. Infatti, se per assurdo $f(z) < f(s)$, essendo $\omega(s, z) > 0$ (z ed s adiacenti),

$$f(s) = \sum_{y \in \mathcal{V}} \omega(s, y) \frac{f(y)}{d_s} < f(s) \sum_{y \in \mathcal{V}} \frac{\omega(s, y)}{d_s} = f(s)$$

il che è assurdo.

Essendo C_i componente connessa, $\forall x \in C_i$ esiste un cammino $s \sim v_1 \sim \dots \sim v_t \sim x$ che collega x ad s . Dunque, f assume lo stesso valore su

tutta la componente connessa C_i . Ragionando allo stesso modo per le altre componenti connesse si ottiene la tesi. □

Teorema 1.20. *Sia $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \omega)$ un grafo pesato. Sia L Laplaciano di \mathcal{G} e $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_N$ i suoi autovalori contati con molteplicità. Allora,*

$$\lambda_2 > 0 \iff \mathcal{G} \text{ è connesso.}$$

In particolare, la molteplicità di λ_1 è pari al numero di componenti connesse di \mathcal{G} .

Dimostrazione. Siano C_1, \dots, C_m con $m \geq 1$ le componenti connesse di \mathcal{G} e definiamo per $i = 1, \dots, m$ le funzioni $f_i: \mathcal{V} \rightarrow \mathbb{R}$ tali che

$$f_i(x) = \begin{cases} 1 & \text{se } x \in C_i \\ 0 & \text{altrimenti} \end{cases}$$

Le f_i così definite sono, banalmente, linearmente indipendenti e, dal Lemma 1.19 segue che le f_i sono un sistema di generatori per il $Ker(L)$. Dunque, $\dim(Ker(L)) = m$ con m numero di componenti connesse di \mathcal{G} .

Essendo $\lambda_1 = 0$, la sua molteplicità è pari alla dimensione di $Ker(L) = m$ da cui segue la tesi (se \mathcal{G} è connesso $m = 1$). □

Gli stessi risultati sono stati provati anche per la versione normalizzata del Laplaciano del grafo [4], spesso preferito nello studio matematico dei grafi. Tuttavia, nel modello proposto nel capitolo successivo si è scelto di utilizzare il *Laplaciano standard*, in quanto è l'unico che, in assenza di un termine di reazione, soddisfa le proprietà di Fick e di conservazione globale della massa, fondamentali per modellare la diffusione [5].

Per ulteriori risultati sullo spettro del Laplaciano normalizzato del grafo si fa riferimento sempre a [4].

1.3 Funzione di Lambert

L'importanza delle funzioni esponenziale e logaritmica, in tutti i campi della scienza, è ormai nota da tempo. Senza di esse la nostra capacità di ricavare soluzioni esplicite a problemi fondamentali della biologia matematica sarebbe molto ridotta. Purtroppo, però, la semplice combinazione di una di queste due funzioni con il proprio argomento porta ad equazioni apparentemente impossibili da risolvere con il solo uso delle funzioni elementari. Soprattutto in ambito ecologico e biologico, sono tanti i modelli descritti da questo tipo di equazioni ed è questo, uno dei motivi che ha spinto i ricercatori a rivalutare e studiare in modo più approfondito quella che è stata, poi, denominata *funzione di Lambert*. Definita come l'inversa multivariata della funzione

$$\omega \mapsto \omega e^{\omega} ,$$

la funzione di Lambert è strettamente legata a quella logaritmica e il suo nome deriva dal fatto che si tratta del logaritmo di un caso speciale della serie di Lambert (vedi [6]).

Mentre il logaritmo naturale risponde alla domanda "*quale potenza di e ha prodotto questo numero?*", la funzione di Lambert risponde alla domanda "*quale potenza di e, moltiplicata per se stessa, ha prodotto questo numero?*". Dunque, in termini matematici, così come il logaritmo naturale $\ln(x)$ è soluzione dell'equazione $e^{\ln(x)} = x$, si definisce la funzione di Lambert $W(x)$ come soluzione di $W(x)e^{W(x)} = x$.

Restringendoci al caso di x reale, osserviamo che W è definita per $x \geq -\frac{1}{e}$ e, sull'intervallo $-\frac{1}{e} \leq x < 0$ assume due valori reali (come mostrato in Figura 1.3). Distinguiamo, quindi, il grafico di W in due diramazioni: denotiamo con $W_0(x)$ quella che soddisfa $W(x) \geq -1$ e con W_{-1} quella per cui vale $W(x) < -1$. W_0 viene detta *diramazione principale* ed è definita su tutto l'intervallo $x \geq -\frac{1}{e}$, al contrario di W_{-1} definita solo per x negative. Dunque, se siamo interessati a valori reali di W ed è chiaro che $x > 0$,

possiamo riferirci direttamente alla diramazione principale, W_0 . Mentre, nel caso in cui l'argomento di W sia negativo, è necessaria un'interpretazione biologica per poter scegliere quale delle due diramazioni è rilevante.

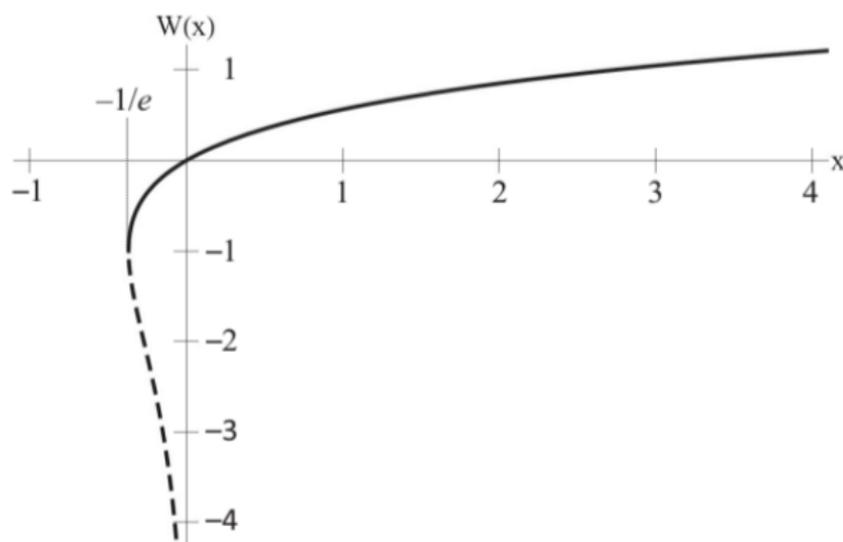


Figura 1.3: Valori *reali* della funzione W di *Lambert*. In nero la diramazione principale W_0 e con la curva tratteggiata la diramazione W_{-1}

Come per le altre funzioni elementari, anche W può essere derivata ed integrata (come mostrato in [6]), aspetto essenziale per l'applicabilità di tale funzione a modelli biologici. Inoltre, le soluzioni esplicite calcolate con la funzione di Lambert forniscono una visione più approfondita del problema biologico in questione e si prestano ad ulteriori operazioni numeriche quali differenziazione ed integrazione. Questi vantaggi fanno sì che questa funzione possa essere applicata ad un'ampia gamma di problemi che spaziano dall'ambito combinatorio a quello epidemico e di cinetica enzimatica. W viene anche utilizzata per risolvere equazioni del tipo $y = x \ln(x)$ o $y = x + e^x$, l'equazione di *Volterra* per la crescita di una popolazione ed equazioni lineari con ritardo (per approfondimenti sulle applicazioni di W si veda [6] e [7]).

Infine, la funzione di Lambert è, oggi, implementata in tutti i principali software matematici, rendendone la facilità d'uso paragonabile a quella di funzioni elementari come esponenziale e logaritmo.

Capitolo 2

Modello su network e metrica su un grafo per le stime dei tempi d'arrivo

La malattia di Alzheimer è caratterizzata dall'accumulo di grovigli neurofibrillari (NFT) di proteina *tau*. Si ritiene che l'attività di semina abbia inizio nella *corteccia entorinale* e che, da qui, si propaghi nel resto del cervello attraverso il *connettoma strutturale*, la rete di vie assonali che collegano le diverse regioni dell'encefalo [8]. Una volta che tutte le regioni cerebrali raggiungono un certo livello di concentrazione della proteina tossica, il paziente è considerato affetto da Alzheimer e, per le conoscenze attuali, non curabile. È quindi importante intervenire prima che la malattia invada tutto il cervello. A tal fine, è utile rispondere alla seguente domanda: *una volta che la regione di semina è infetta, quanto tempo ci vuole prima che si sviluppi un'invasione su larga scala, portando alla demenza cerebrale?*

Dal punto di vista matematico, possiamo rispondere a questo quesito modellando il cervello umano con un grafo e risolvendo un '*problema di tempo d'arrivo*', ovvero determinando il tempo che impiega una quantità di interesse a raggiungere un fissato livello in ogni nodo.

Uno dei possibili approcci a questo problema consiste nel considerare la pro-

pagazione tra due nodi vicini, il nodo seme ed il nodo *target*, trascurando tutte le altre connessioni. Linearizzando il sistema, è possibile estrarre una stima del tempo che il secondo nodo impiega per raggiungere il livello di concentrazione iniziale del nodo seme. Questo tempo fornisce una nozione di distanza tra nodi adiacenti, la *distanza d'arco effettiva* e, a partire da essa, si approssima il *tempo d'arrivo* tra due nodi distanti come la lunghezza del più breve cammino che li collega. In genere, però, le soluzioni ottenute utilizzando la *distanza effettiva* non sono molto precise ed è proprio dal tentativo di ottenere stime più significative per i tempi d'arrivo, che nasce l'idea di munire il grafo di una nuova metrica, quella di *Lambert*.

Presentiamo, quindi, il modello matematico su cui ci baseremo per definire questa nuova nozione di distanza su un grafo.

2.1 Fisher-KPP su un grafo

Assumiamo che le dinamiche di un processo di invasione su un network non orientato e connesso siano governate dall'equazione di *Fisher-KPP*, studiata nella Sezione 1.1.

Consideriamo, quindi, un grafo connesso e non orientato $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ di ordine N . Gli N vertici di \mathcal{G} rappresentano le regioni cerebrali di interesse ed \mathcal{E} l'insieme dei fasci assionali che connettono una regione all'altra. Indichiamo con $p_i = p_i(t)$ la quantità di interesse nella regione i , le cui dinamiche sono descritte dal seguente sistema di equazioni differenziali ordinarie

$$\frac{dp_i}{dt} = -\rho \sum_{j=1}^N L_{i,j} p_j + \alpha p_i (1 - p_i), \quad i = 1, \dots, N \quad (2.1)$$

$$p_i(0) = q_i, \quad i = 1, \dots, N \quad (2.2)$$

dove,

- $\rho, \alpha > 0$ sono la costante di diffusione e di crescita, rispettivamente;
- le concentrazioni iniziali $q_i \in [0, 1] \forall i = 1, \dots, N$;

- $\mathbf{L} = (L_{i,j})_{i,j=1,\dots,N}$ (matrice $N \times N$), è il Laplaciano del grafo \mathcal{G} , la cui definizione è riportata in 1.16.

Il Laplaciano del grafo tiene conto delle dinamiche di accoppiamento lineare tra i nodi, le quali influiscono molto sulla scala dei tempi d'arrivo [9]. Inoltre, \mathbf{L} modella la diffusione del processo e raccoglie le informazioni riguardanti le connessioni tra le regioni: $L_{i,j} = 0$ indica assenza di connessione tra il nodo i ed il nodo j .

Analizziamo il sistema appena descritto. Innanzitutto, denotiamo il *termine sorgente* dell' i -esima riga del sistema con $R_i := R(p_i) = \alpha p_i(1 - p_i)$ ed osserviamo che, come già mostrato nella Sezione 1.1, il sistema (2.1) supporta due stati d'equilibrio: lo *stato sano*, dato da $p_i = 0 \forall i$ e lo *stato tossico*, definito da $p_i = 1 \forall i$. Il primo instabile ($R'_i(0) = \alpha > 0 \forall i$) ed il secondo stabile ($R'_i(1) = -\alpha < 0 \forall i$).

Indichiamo, poi, con $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_N$ gli autovalori di \mathbf{L} contati con molteplicità. Facendo riferimento alla Sezione 1.2.1 si ha che $\lambda_1 = 0$ e λ_2 è il più piccolo autovalore non nullo. Inoltre, essendo una matrice simmetrica, \mathbf{L} è simile alla matrice diagonale Λ degli autovalori. Pertanto, fissata la riga i -esima di (2.1), si ha che il termine diffusivo dipende da $\rho\lambda_i$ mentre il termine sorgente dalla costante di crescita α . Dunque, le dinamiche del sistema sono controllate dal più piccolo autovalore non nullo λ_2 e questo ci permette di definire due diversi regimi di interesse: se $\rho\lambda_2 \gg \alpha \Rightarrow \rho\lambda_i \gg \alpha \forall i = 1, \dots, N$ e si parlerà di *regime dominato dalla diffusione*, mentre la condizione $\rho\lambda_2 \ll \alpha$ definisce il *regime dominato dalla crescita*.

Come si deduce dai grafici in Figura 2.1, nel regime dominato dalla diffusione il livello di concentrazione tende ad uniformarsi rapidamente su tutto il grafo, per poi crescere lentamente e simultaneamente in tutti i nodi. Nel regime dominato dalla crescita, invece, la concentrazione del nodo seme aumenta mentre avviene la diffusione. In questo caso, il grafo viene invaso sistematicamente, nodo per nodo, ed ogni volta che una nuova regione viene raggiunta, si assiste ad una crescita locale della sua concentrazione.

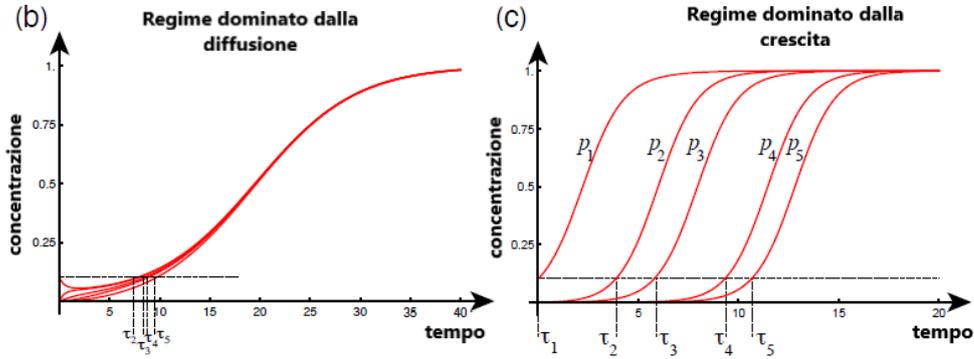


Figura 2.1: Grafici delle concentrazioni $p_i(t)$ per un grafo di ordine 5, con $\lambda_2 \approx 0.228$. In (b) un caso dominato dalla diffusione ($\alpha = \frac{1}{5}$, $\rho = 1$). In (c) caso dominato dalla crescita ($\alpha = 1$, $\rho = \frac{1}{100}$). Si considera un unico nodo seme con $q_1 = \frac{1}{10}$.

Infine, poiché siamo interessati ad ottenere delle stime per i *tempi d'arrivo*, distinguiamo in

- *Problema di tempo d'arrivo canonico*: al tempo iniziale $t = 0$ il nodo seme è unico ($q_s = \beta$, $q_i = 0 \forall i \neq s$) e l'obiettivo è ottenere il valore del tempo d'arrivo τ_i tale che $p_i(\tau_i) = \mu$ con $\beta \leq \mu < 1$ ($\tau_i = 0$ solo se $i = s$).
- *Problema generalizzato del tempo d'arrivo*: si considerano più nodi insemiati inizialmente e si assume $\beta = \max_i q_i > 0$. Come per il caso precedente vogliamo determinare il tempo d'arrivo τ_i tale che $p_i(\tau_i) = \mu$ con $\beta \leq \mu < 1$.

Osservazione 2.1. In assenza di trasporto ($\rho = 0$), le dinamiche del sistema seguono l'equazione logistica $\frac{dp}{dt} = \alpha p(1 - p)$, la quale presenta una soluzione sigmoide, cioè descrive una curva che inizialmente cresce quasi esponenzialmente, poi rallenta raggiungendo una posizione asintotica e rimanendo stabile su quel valore. Quindi, in questo caso, per tutti i valori di concentrazione iniziale $0 < \beta < 1$, il sistema evolverà verso lo stato tossico e i tempi d'arrivo saranno positivi e finiti.

2.2 Analisi lineare del modello

In questa sezione ci occuperemo di linearizzare il sistema (2.1) intorno ad uno stato di equilibrio ed ottenere una prima stima per i tempi d'arrivo. Come vedremo in seguito, questa approssimazione lineare ci fornirà una nozione di distanza tra nodi adiacenti.

Indichiamo con $R_i = R(p_i) = \alpha p_i(1 - p_i)$ l' i -esimo termine di reazione di (2.1) e linearizziamo il sistema intorno allo stato di equilibrio *sano* ($p_i = 0 \forall i$). Ovvero, $\forall i = 1, \dots, N$ sviluppiamo con Taylor R_i per $p_i \rightarrow 0$ e otteniamo

$$\frac{dp_i}{dt} = -\rho \sum_{j=1}^N L_{i,j} p_j + \alpha p_i, \quad i = 1, \dots, N \quad (2.3)$$

$$p_i(0) = q_i, \quad i = 1, \dots, N \quad (2.4)$$

che, in forma compatta, diventa

$$\begin{cases} \frac{d}{dt} p(t) = M p(t) \\ p(0) = q \end{cases} \quad (2.5)$$

con $p(t) = (p_1, \dots, p_N)$, $q = (q_1, \dots, q_N)$ ed $M = -\rho \mathbf{L} + \alpha \mathbf{I}$, dove \mathbf{I} denota la matrice Identità. Integrando, si determina la soluzione

$$p(t) = e^{tM} q$$

in termini di matrice esponenziale. Ricordiamone la definizione e vediamo sotto quali condizioni possiamo semplificarne il calcolo.

Definizione 2.2. Sia A una matrice quadrata ($N \times N$) a coefficienti reali o complessi. La matrice esponenziale di A , e^A , è una matrice ($N \times N$) ottenuta dal seguente sviluppo in serie di potenze

$$e^A = \sum_{k=0}^{\infty} \frac{A^k}{k!}$$

Osservazione 2.3. Sfruttando gli autovettori di A si riesce a scrivere la matrice esponenziale e^A come somma finita di termini [10]. In particolare, se A è una matrice diagonalizzabile,

$$e^A = \begin{bmatrix} t_1 & \dots & t_N \end{bmatrix} e^\Lambda \begin{bmatrix} s_1^T \\ \vdots \\ s_N^T \end{bmatrix} = \begin{bmatrix} t_1 & \dots & t_N \end{bmatrix} \begin{bmatrix} e^{\lambda_1} & & \\ & \ddots & \\ & & e^{\lambda_N} \end{bmatrix} \begin{bmatrix} s_1^T \\ \vdots \\ s_N^T \end{bmatrix} = \sum_{k=1}^N t_k \cdot s_k^T e^{\lambda_k}$$

dove,

- t_1, \dots, t_N sono autovettori destri di A ;
- s_1^T, \dots, s_N^T sono autovettori sinistri di A ;
- $\lambda_1, \dots, \lambda_N$ autovalori di A , con Λ matrice diagonale degli autovalori.

Essendo L simmetrica, esiste una base ortonormale di autovettori $v^{(1)}, \dots, v^{(N)}$ associati agli autovalori $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_N$ (non necessariamente distinti). Possiamo quindi calcolare la matrice esponenziale e^L seguendo quanto riportato nell'osservazione precedente e, tenendo conto del seguente risultato,

Osservazione 2.4. Sia A matrice simmetrica.

$$v \text{ autovettore destro di } A \Leftrightarrow v \text{ autovettore sinistro di } A$$

Dimostrazione. $Av = \lambda v \Leftrightarrow v^T A^T = \lambda v^T \underbrace{\Leftrightarrow}_{A=A^T} v^T A = \lambda v^T.$

□

possiamo scrivere,

$$e^L = \begin{bmatrix} v^{(1)} & \dots & v^{(N)} \end{bmatrix} \begin{bmatrix} e^{\lambda_1} & & \\ & \ddots & \\ & & e^{\lambda_N} \end{bmatrix} \begin{bmatrix} v^{(1)} \\ \vdots \\ v^{(N)} \end{bmatrix} = \sum_{k=1}^N v^{(k)} \cdot v^{(k)} e^{\lambda_k} .$$

Moltiplicando, poi, per il dato iniziale $q = (q_1, \dots, q_N)$ otteniamo

$$q \cdot e^L = \begin{bmatrix} q \cdot v^{(1)} & \dots & q \cdot v^{(N)} \end{bmatrix} \begin{bmatrix} e^{\lambda_1} & & \\ & \ddots & \\ & & e^{\lambda_N} \end{bmatrix} \begin{bmatrix} v^{(1)} \\ \vdots \\ v^{(N)} \end{bmatrix} = \sum_{k=1}^N (q \cdot v^{(k)} e^{\lambda_k}) v^{(k)} .$$

Possiamo quindi riscrivere la soluzione del sistema linearizzato (2.5) come combinazione lineare degli autovettori di \mathbf{L} :

$$\mathbf{p} = e^{\alpha t} \sum_{k=1}^N (q \cdot v^{(k)} e^{-\rho \lambda_k t}) v^{(k)} . \quad (2.6)$$

Imponendo la j -esima componente di \mathbf{p} pari a μ e risolvendo rispetto alla variabile temporale, si ottiene una stima del tempo che impiega il j -esimo nodo a raggiungere un livello di concentrazione pari a μ . Ovvero, il j -esimo *tempo d'arrivo* τ_j è soluzione dell'equazione

$$\mu = e^{\alpha \tau_j} \sum_{k=1}^N (q \cdot v^{(k)} e^{-\rho \lambda_k \tau_j}) v_j^{(k)} . \quad (2.7)$$

Se ci restringiamo al caso di un unico nodo seme, $q_s = \beta$ e $q_i = 0 \forall i \neq s$, possiamo semplificare il sistema indicando l' r -esima componente di q con $q_r = \beta \delta_{rs}$ e ottenendo il j -esimo *tempo d'arrivo canonico* come soluzione di

$$\frac{\mu}{\beta} = e^{\alpha \tau_j} \sum_{k=1}^N v_s^{(k)} v_j^{(k)} e^{-\rho \lambda_k \tau_j} . \quad (2.8)$$

Confrontando le soluzioni lineari con quelle numeriche in Figura (2.2), emergono alcune caratteristiche di questo metodo. In primo luogo, l'approssimazione data dai tempi d'arrivo lineari migliora al diminuire di $\mu = \beta$, come possiamo osservare dai grafici (b) e (c). Infatti, il Teorema della varietà stabile assicura che la soluzione lineare sia una fedele approssimazione della soluzione numerica, vicino allo stato di equilibrio sano ($\mu = \beta = 0$). Dunque, per $\mu \rightarrow 0$ con $\mu > 0$, la soluzione del sistema linearizzato (2.3)

2. Modello su network e metrica su un grafo per le stime dei tempi d'arrivo

converge alla soluzione esatta e, di conseguenza, i tempi d'arrivo lineari si avvicinano ai tempi esatti. In secondo luogo, dal grafico in (a), notiamo che le soluzioni del sistema linearizzato crescono più rapidamente delle soluzioni esatte, raggiungendo in meno tempo un fissato livello di concentrazione. Pertanto, le stime lineari forniscono un limite inferiore per i tempi d'arrivo. In conclusione, per calcolare i tempi d'arrivo lineari è necessario risolvere N equazioni trascendentali della forma (2.7) o (2.8) (a seconda che si tratti del caso *generalizzato* o *canonico*, rispettivamente), la cui risoluzione potrebbe risultare numericamente difficile senza delle buone ipotesi iniziali.

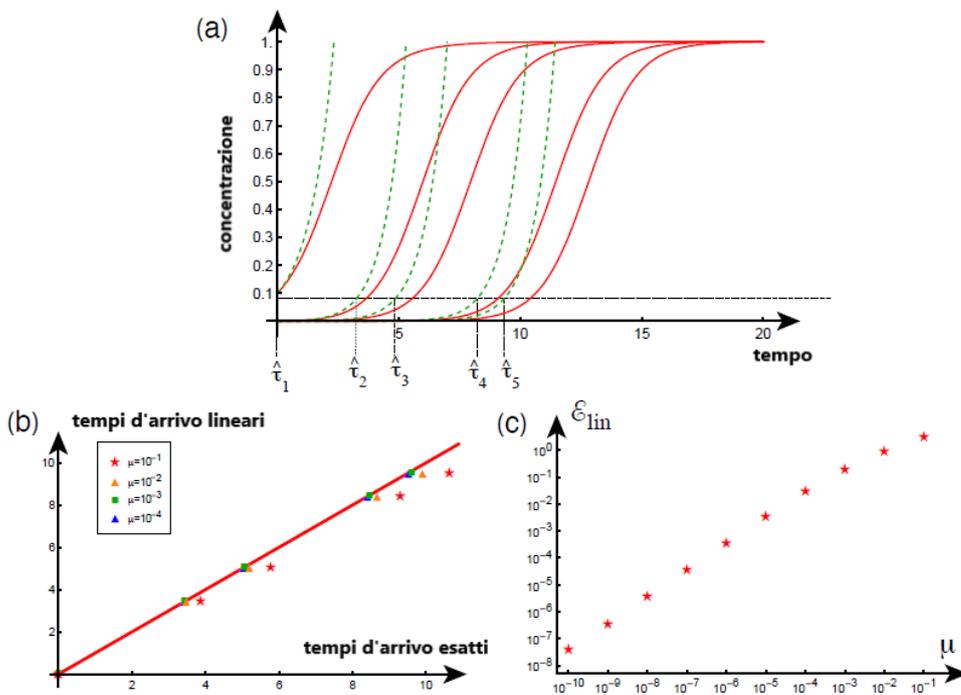


Figura 2.2: In (a) un confronto tra le soluzioni esatte (in rosso) e quelle lineari (in verde) in un caso dominato dalla crescita ($\alpha = 1$, $\rho = 1/100$, $\beta = 1/10$). (b) confronta i tempi d'arrivo lineari con quelli esatti (linea continua, rossa) per diversi valori di concentrazione $\mu = \beta$. In (c) il grafico dell'errore dei tempi d'arrivo, $\mathcal{E}_{lin} = \sum_{i=1}^N |\tau_i(\mu) - t_i(\mu)|$ con t_i tempo d'arrivo esatto.

Osservazione 2.5. Il metodo lineare è sufficiente per la piena comprensione del problema nel caso di un regime dominato dalla diffusione (pag.6 di [1]). Il nostro interesse è, però, volto al caso dominato dalla crescita. Pertanto, d'ora in avanti assumiamo α di ordine 1 e ρ parametro piccolo.

2.3 Distanza di Lambert

Con l'obiettivo di migliorare le approssimazioni lineari, cambiamo approccio al problema e muniamo il grafo di una nuova metrica, la metrica di *Lambert*. Questa nozione di distanza tra nodi è molto potente, in quanto conferisce al grafo una struttura aggiuntiva da cui è possibile ottenere molte stime. In particolare, il tempo che impiega una quantità di interesse a passare da una regione all'altra del cervello sarà approssimato con la distanza tra queste due regioni. Vediamo, quindi, come costruire questa metrica e ottenere stime significative per i tempi d'arrivo.

In primo luogo, definiamo una nozione di distanza tra nodi adiacenti lavorando con un'unica coppia di vertici ed ignorando le connessioni con le restanti regioni del grafo. Come vedremo, saranno i tempi d'arrivo lineari a fornirci questa prima nozione di distanza.

Indichiamo con $A = (A_{i,j})_{i,j=1,\dots,N}$ la matrice di adiacenza ponderata di \mathcal{G} e consideriamo il sottografo \mathcal{G}' di \mathcal{G} costituito da due nodi adiacenti i e j ($A_{i,j} \neq 0$). \mathcal{G}' è un grafo pesato di ordine 2, non orientato e connesso, il cui Laplaciano \mathbf{L}' è il surrogato del Laplaciano \mathbf{L} di \mathcal{G} . Dunque, ricordando che $L_{i,j} = -A_{i,j} + \delta_{i,j} \sum_{k=1}^N A_{i,k}$ e restringendoci al sottografo \mathcal{G}' si ha che

$$\mathbf{L}' = \begin{bmatrix} L_{i,i} & -L_{i,j} \\ -L_{i,j} & L_{j,j} \end{bmatrix} = \begin{bmatrix} A_{i,j} & -A_{i,j} \\ -A_{i,j} & A_{i,j} \end{bmatrix}.$$

Assumiamo che le dinamiche di propagazione della proteina *tau* su \mathcal{G}' siano

governate dal sistema linearizzato (2.3), qui ridotto al sistema

$$\frac{dp_k}{dt} = -\rho \sum_{r=i,j} L'_{k,r} p_r + \alpha p_k, \quad k = i, j \quad (2.9)$$

$$p_k(0) = q_k, \quad k = i, j \quad (2.10)$$

con $q_i = \beta$ e $q_j = 0$.

Vogliamo poi stimare il tempo $t_{i,j}$ che impiega il nodo j a raggiungere un livello di concentrazione pari a $\mu = \beta$. Trattandosi di un caso *canonico* del problema dei tempi d'arrivo, otterremo $t_{i,j}$ come soluzione dell'equazione (2.8). Calcoliamo, quindi, gli autovalori e gli autovettori di \mathbf{L}' . Da

$$\det(\mathbf{L}' - \lambda) = 0 \iff \lambda(\lambda - 2A_{i,j}) = 0$$

determiniamo gli autovalori $\lambda_i = 0$ e $\lambda_j = 2A_{i,j}$, i cui relativi autovettori sono $v^{(i)} = (1, 1)$ e $v^{(j)} = (1, -1)$, rispettivamente. Dunque, sostituendoli in (2.8), con $\mu = \beta$ otteniamo l'equazione

$$e^{\alpha t}(1 - e^{-2\rho A_{i,j}t}) = 1 \quad (2.11)$$

ed essendo interessati al caso dominato dalla crescita (stiamo assumendo ρ piccolo), sviluppiamo con Taylor $e^{-2\rho A_{i,j}t}$ per $\rho \rightarrow 0$ e scriviamo

$$2\rho A_{i,j}te^{\alpha t} = 1. \quad (2.12)$$

Moltiplicando entrambi i membri per $\alpha/(2\rho A_{i,j})$ ricaviamo l'equazione

$$\alpha te^{\alpha t} = \frac{\alpha}{2\rho A_{i,j}}$$

la cui soluzione è espressa in termini di *funzione di Lambert* (definita nella Sezione 1.3). In particolare, essendo $\frac{\alpha}{2\rho A_{i,j}} \geq 0$, si ha

$$\alpha t = W_0\left(\frac{\alpha}{2\rho A_{i,j}}\right)$$

da cui,

$$t_{i,j} = \frac{1}{\alpha} W_0\left(\frac{\alpha}{2\rho A_{i,j}}\right) \quad i \neq j \quad (2.13)$$

con W_0 *diramazione principale* della funzione di Lambert W .

Il grafo \mathcal{G} su cui stiamo lavorando è non orientato, dunque $A_{i,j} = A_{j,i}$ da cui segue la simmetria dei tempi d'arrivo, $t_{i,j} = t_{j,i}$.

Definizione 2.6. Sia $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ un grafo pesato, non orientato e connesso. Sia $A = (A_{i,j})$ la sua matrice di adiacenza ponderata e siano $i, j \in \mathcal{V}$ con $i \neq j$ due nodi adiacenti, $i \sim j$. Definiamo la distanza tra i nodi i e j come il tempo d'arrivo

$$t_{i,j} = \frac{1}{\alpha} W_0 \left(\frac{\alpha}{2\rho A_{i,j}} \right)$$

e ci riferiamo ad essa in termini di *distanza d'arco di Lambert*.

A partire da $t_{i,j}$, definiamo la *distanza di Lambert* tra i nodi i e j (non necessariamente adiacenti) come la lunghezza del più corto cammino, rispetto alla *distanza d'arco di Lambert*, che li collega. Formalmente,

Definizione 2.7. Sia $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ un grafo pesato, non orientato e connesso. Sia $A = (A_{i,j})$ la sua matrice di adiacenza ponderata e sia $\Gamma_{i,j} = \gamma_0, \gamma_1, \dots, \gamma_n$ con $\gamma_0 = i$ e $\gamma_n = j$ il più corto cammino, rispetto alla *distanza d'arco di Lambert*, tra i nodi i e j . Allora, la *distanza di Lambert* W_{ij} tra questi due nodi è definita da

$$W_{ij} = \begin{cases} \sum_{k=0}^{n-1} t_{\gamma_k, \gamma_{k+1}} = \frac{1}{\alpha} \sum_{k=0}^{n-1} W_0 \left(\frac{\alpha}{2\rho A_{\gamma_k, \gamma_{k+1}}} \right), & i \neq j \\ 0, & i = j \end{cases} \quad (2.14)$$

e chiamiamo *matrice della distanza di Lambert*, la matrice $\mathbf{W} = (W_{ij})$.

Proposizione 2.8. Sia $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ un grafo pesato di ordine N , non orientato e connesso. Sia $A = (A_{i,j})$ la sua matrice di adiacenza ponderata e $\mathbf{W} = (W_{ij})$ la matrice della distanza di Lambert. Allora,

\mathbf{W} è una metrica su \mathcal{G} .

Ovvero, \mathbf{W} soddisfa le seguenti proprietà:

(a) $W_{ij} \geq 0 \quad \forall i, j \in \mathcal{V} \quad e \quad W_{ij} = 0 \Leftrightarrow i = j;$

(b) $W_{ij} = W_{ji} \quad \forall i, j \in \mathcal{V};$

(c) $W_{ij} \leq W_{ik} + W_{kj} \quad \forall i, j, k \in \mathcal{V}.$

Dimostrazione. (a). $W_{ii} = 0 \quad \forall i \in \mathcal{V}$ segue dalla definizione.

Se mostriamo che $\forall i \neq j$ l'argomento di W_0 è positivo per ogni elemento della sommatoria, abbiamo concluso (vedi 1.3). Sia, quindi, $i \neq j$ e sia $\Gamma_{i,j} = \gamma_0, \gamma_1, \dots, \gamma_n$ il più corto cammino tra questi due nodi. Dalla definizione di cammino su un grafo (1.7) segue che $\gamma_k \sim \gamma_{k+1} \quad \forall k = 0, \dots, n-1$ da cui, $A_{\gamma_k, \gamma_{k+1}} > 0 \quad \forall k = 0, \dots, n-1$.

(b). Sia $\Gamma_{i,j} = \gamma_0, \gamma_1, \dots, \gamma_n$, con $\gamma_0 = i$ e $\gamma_n = j$ il minimo cammino di Lambert tra i nodi i e j . Essendo il grafo \mathcal{G} non orientato, il minimo cammino di Lambert tra i nodi j e i è $\Gamma_{i,j}$ percorso al contrario. Dunque, per la simmetria della *distanza d'arco di Lambert* ($t_{\gamma_k, \gamma_{k+1}} = t_{\gamma_{k+1}, \gamma_k} \quad \forall k = 0, \dots, n-1$) e per commutatività della somma, concludiamo che $W_{ij} = W_{ji} \quad \forall i, j$.

(c). Sia $\Gamma_{i,j} = \gamma_0, \gamma_1, \dots, \gamma_n$, con $\gamma_0 = i$ e $\gamma_n = j$ il più corto cammino, rispetto alla *distanza d'arco di Lambert*, tra i nodi i e j . Allora, per definizione,

$$W_{ij} \leq \sum_{s=0}^{n-1} t_{\tilde{\gamma}_s, \tilde{\gamma}_{s+1}}$$

per un qualsiasi cammino $\tilde{\Gamma}_{i,j} = \tilde{\gamma}_0, \tilde{\gamma}_1, \dots, \tilde{\gamma}_n$ tra i e j (con $\tilde{\gamma}_0 = \gamma_0 = i$ e $\tilde{\gamma}_n = \gamma_n = j$).

Fissato un terzo nodo k , indichiamo con $\Gamma_{i,k} = \gamma_0, \dots, \gamma_m$ e $\Gamma_{k,j} = \gamma_m, \dots, \gamma_n$ i minimi cammini di Lambert i nodi $i-k$ e $k-j$ rispettivamente ($\gamma_m = k$). La concatenazione di questi due cammini è un cammino che tra i nodi i e j .

Dunque,

$$W_{ij} \leq \underbrace{\sum_{s=0}^{m-1} t_{\gamma_s, \gamma_{s+1}}}_{W_{ik}} + \underbrace{\sum_{s=m}^{n-1} t_{\gamma_s, \gamma_{s+1}}}_{W_{kj}}$$

da cui si ha la tesi. □

Abbiamo dotato il grafo di una nuova metrica, la quale ci fornisce un'utile stima dei tempi d'arrivo. In particolare, la *distanza di Lambert* tra un nodo seme ed un generico nodo di \mathcal{G} , approssima il tempo che impiega quest'ultimo a raggiungere lo stesso livello di concentrazione della regione di semina. Generalmente però, siamo interessati a casi in cui la soglia critica di concentrazione μ è diversa dalla concentrazione iniziale β (più precisamente $\mu > \beta$). Pertanto, indichiamo con t_{ii} il tempo che impiega la concentrazione iniziale β , del nodo i , a raggiungere un valore pari a μ , in assenza di connessioni con altri nodi. Una stima di t_{ii} si ottiene studiando le dinamiche di crescita locale del nodo i , descritte dal seguente Problema di Cauchy

$$\begin{cases} \frac{dp_i}{dt} = \alpha p_i (1 - p_i) \\ p_i(0) = \beta \end{cases} \quad (2.15)$$

Più precisamente, determinata la soluzione del problema $p_i(t)$, cerchiamo il tempo t_{ii} tale che $p_i(t_{ii}) = \mu$. Essendo le dinamiche di crescita locale governate dall'equazione logistica, la soluzione del problema è della forma

$$p_i(t) = \frac{\beta e^{\alpha t}}{1 - \beta + \beta e^{\alpha t}}$$

ed otteniamo t_{ii} come soluzione dell'equazione

$$\mu = \frac{\beta e^{\alpha t}}{1 - \beta + \beta e^{\alpha t}}.$$

Dunque,

$$t_{ii} = \frac{1}{\alpha} \log \left(\frac{(1 - \beta)\mu}{(1 - \mu)\beta} \right). \quad (2.16)$$

Nel caso *canonico* di un unico nodo seme i , i *tempi d'arrivo di Lambert*, ad ogni nodo j , sono dati da

$$\tilde{\tau}_{ij} = t_{ii} + W_{ij} . \quad (2.17)$$

Allora, il j -esimo *tempo d'arrivo di Lambert* può essere interpretato come la somma del tempo necessario al nodo seme i per raggiungere la concentrazione μ e del tempo che impiega la concentrazione μ nel nodo i a raggiungere il nodo j .

Osservazione 2.9. Abbiamo determinato W_{ij} assumendo implicitamente $\mu = \beta$ e, in effetti, se ci troviamo in questo caso, il termine t_{ii} si annulla. Inoltre, dalla simmetria di W_{ij} segue la simmetria dei tempi d'arrivo, $\tilde{\tau}_{ij} = \tilde{\tau}_{ji}$.

Chiarimento 1. Le stime di Lambert non tengono conto dei seguenti fatti:

- (1) tutti i nodi prima del nodo *target* sono connessi ad altri nodi. Di conseguenza, la quantità di interesse viene diluita lontano dal nodo target;
- (2) il nodo *target* può ricevere contributi da altri percorsi;
- (3) il sistema linearizzato su cui ci si basa per ottenere i *tempi di Lambert*, trascura il termine non lineare (il quale agisce rallentando la crescita della soluzione). Di conseguenza, le soluzioni lineari impiegano meno tempo a raggiungere un fissato livello di concentrazione.

Questi tre aspetti impediscono ai *tempi di Lambert* di essere un limite superiore o inferiore per i reali tempi d'arrivo. Infatti, se non si tiene conto di (1) e (3), si ottengono dei tempi d'arrivo ridotti rispetto a quelli reali; ma trascurando (2) si determinano dei tempi sovrastimati.

Nonostante non tenga conto di questi importanti aspetti, il metodo di *Lambert* fornisce un'approssimazione notevolmente buona, come mostreremo in seguito. Inoltre, presenta una caratteristica degna di nota: la *distanza di Lambert* tra due nodi distinti è una proprietà intrinseca del sistema, in

quanto gli elementi W_{ij} della *matrice di distanza di Lambert* non dipendono dalla concentrazione iniziale β . Quest'ultima, infatti, agisce perturbando solo il termine t_{ii} . Ciò implica che il dato iniziale non ha un impatto apprezzabile sulla propagazione complessiva.

Assumendo che le dinamiche, dopo il tempo d'arrivo $\tilde{\tau}_{ij}$, siano controllate dalla sola crescita locale, a partire dai *tempi di Lambert* è possibile ottenere un'approssimazione della soluzione d'onda di (2.1). Più precisamente, se scegliamo il nodo i come regione di semina, stiamo supponendo che le dinamiche in j , trascorso il tempo $\tilde{\tau}_{ij}$, siano governate da

$$\begin{cases} \frac{d\tilde{p}_j}{dt} = \alpha\tilde{p}_j(1 - \tilde{p}_j) \\ \tilde{p}_j(\tilde{\tau}_{ij}) = \mu \end{cases} \quad (2.18)$$

e risolvendo il problema ottiamo la *soluzione di Lambert*

$$\tilde{p}_j = \frac{\mu e^{\alpha t}}{\mu e^{\alpha t} - (\mu - 1)e^{\alpha\tilde{\tau}_{ij}}} \quad j = 1, \dots, N \quad (2.19)$$

Come possiamo osservare dai grafici in Figura 2.3, il metodo di *Lambert* fornisce delle buone stime sia in termini di soluzione che di tempi d'arrivo, migliorando le approssimazioni lineari.

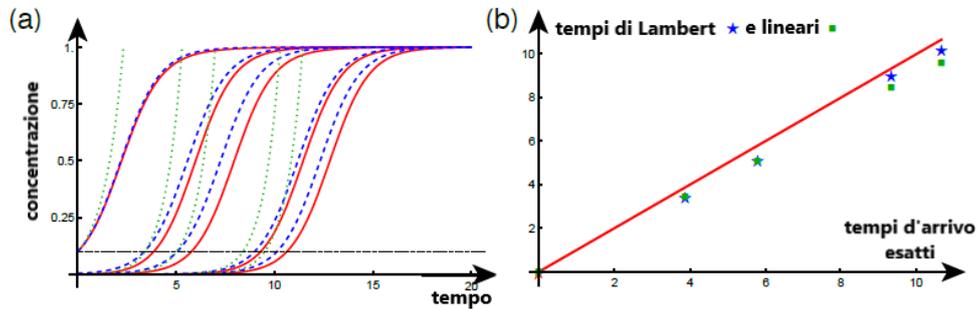


Figura 2.3: (a) mostra un confronto tra le soluzioni esatte (in rosso), di *Lambert* (in blu) e lineari (in verde), nel caso di $\alpha = 1$, $\rho = 1/100$ e $p_1(0) = \beta = \mu$ (1 nodo seme). In (b) un confronto tra i tempi d'arrivo di *Lambert* (stelle blu) e lineari (quadrati verdi).

2.3.1 Luoghi di semina multipli

La *distanza di Lambert* presuppone che il nodo seme sia unico. Potrebbe però succedere che le regioni di semina siano multiple. In tal caso, il metodo di *Lambert* va leggermente modificato. Studiamo, quindi, il caso *generalizzato* del problema dei tempi d'arrivo, supponendo che la concentrazione iniziale del sistema sia data da $\beta = \max_i(q_i) < \mu \ll 1$, con $p(0) = q = (q_1, \dots, q_N)$. Se scegliamo j come nodo *target*, siamo interessati a determinare il tempo \tilde{t}_j tale che $p_j(\tilde{t}_j) = \mu$. L'idea è quella di ottenere il tempo d'arrivo come soluzione dell'equazione $p_j(t) = \mu$. Dunque l'obiettivo che ci poniamo è di trovare un'approssimazione della concentrazione p_j . A tal fine, dobbiamo tenere conto del fatto che ogni regione di semina contribuisce alla crescita della concentrazione del nodo target j . Pertanto, fissiamo una regione di semina i ($\Rightarrow q_i > 0$) e, seguendo l'analisi precedente, utilizziamo (2.17) con $\beta = q_i$ per ottenere una stima del j -esimo tempo d'arrivo

$$\tau_{ij} = \frac{1}{\alpha} \log \left(\frac{(1 - q_i)\mu}{(1 - \mu)q_i} \right) + W_{ij} .$$

dovuto al solo contributo della concentrazione iniziale q_i . Grazie a questa stima, possiamo costruire un'approssimazione del contributo che il nodo seme i fornisce alla concentrazione del nodo j , come segue

$$\tilde{p}_j^{(i)}(t) = \mu \frac{te^{\alpha t}}{\tau_{ij}e^{\alpha\tau_{ij}}} \quad (2.20)$$

La $\tilde{p}_j^{(i)}$ è tale che $\tilde{p}_j^{(i)}(0) = 0$ e $\tilde{p}_j^{(i)}(\tau_{ij}) = \mu$.

Sommando, poi, i contributi di tutti i siti di semina si ottiene la concentrazione del nodo j

$$\tilde{p}_j(t) = \sum_{i=1}^N \tilde{p}_j^{(i)} = \mu t e^{\alpha t} \sum_{i=1}^N \frac{1}{\tau_{ij}} e^{-\alpha\tau_{ij}} . \quad (2.21)$$

A questo punto, possiamo risolvere l'equazione $\tilde{p}_j(t) = \mu$ ed il j -esimo tempo

d'arrivo sarà dato da

$$\tilde{t}_j = \frac{1}{\alpha} W_0 \left(\frac{\alpha}{\sum_{i=1}^N \frac{1}{\tau_{ij}} e^{-\alpha \tau_{ij}}} \right) \quad (2.22)$$

2.3.2 Discussione e confronto con la distanza effettiva

L'idea di definire una distanza tra nodi adiacenti per stimare i tempi d'arrivo era già stata proposta nel lavoro [11], dove la *distanza effettiva*

$$t_{\text{effettiva}} = \frac{1}{\alpha} \left(\log \left(\frac{\alpha}{\rho A_{ij}} \right) - \gamma \right) \quad (2.23)$$

sostituisce la *distanza d'arco di Lambert* (2.6).

Nello stesso lavoro, si avanza la proposta di utilizzare la funzione di Lambert W per approssimare i tempi d'arrivo. Nonostante la migliore corrispondenza ai dati, la *distanza di Lambert* è stata ignorata dagli autori ed in seguito al lavoro di Brockman, Helbing e coautori [12, 13] la distanza effettiva ha guadagnato ulteriore popolarità, probabilmente grazie alla familiarità con la funzione logaritmica.

Successivamente si tentò di mettere in relazione queste due nozioni di distanza [14] ma la differenza tra le due rimane, tutt'ora, molto evidente, come possiamo osservare dai grafici mostrati in Figura 2.4. Non solo, è stato riscontrato che la distanza effettiva non ha le stesse prestazioni della distanza di Lambert, soprattutto su percorsi lunghi.

Inoltre, la presenza della costante di Eulero γ in (2.23) è legata all'approccio stocastico di [11], in cui è stato dimostrato che, nel quadro probabilistico, i tempi d'arrivo seguono una distribuzione di Gumbel. Tuttavia, nel contesto del modello deterministico di nostro interesse, non c'è alcuna giustificazione a priori per questa costante.

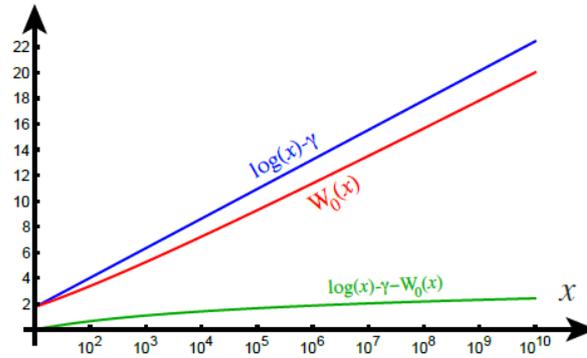


Figura 2.4: Confronto tra *distanza effettiva* (blu) e *distanza d'arco di Lambert* (rosso) in funzione del loro argomento, in un caso dominato dalla crescita. In verde la differenza tra le due.

Per concludere, notiamo che l'utilizzo della funzione logaritmo fornisce tempi d'arrivo sovrastimati. Pertanto, su lunghi cammini l'uso di questa funzione può portare ad un errore sistematico.

Capitolo 3

Applicazione a malattie neurodegenerative

È ormai noto che malattie neurodegenerative come il Parkinson e l'Alzheimer siano associate alla formazione di grandi aggregati di proteine tossiche. Nel caso della malattia di Alzheimer, l'accumulo della proteina *tau* è particolarmente importante per la comprensione della malattia, in quanto è associato a regioni di atrofia e disfunzione cognitiva. Il carico tossico ha origine in aree cerebrali localizzate intorno alla corteccia entorinale [15] e, a partire da questa regione, invade sistematicamente il cervello. Dunque, possiamo interpretare il fenomeno di propagazione della proteina *tau* come un processo di invasione su networks. Inoltre, da studi istopatologici post-mortem è emerso che l'invasione avviene nel regime dominato dalla crescita [16]. Pertanto, possiamo utilizzare il modello proposto nel capitolo precedente per comprenderne le dinamiche. Il network cerebrale di un paziente è quindi definito da un insieme di vertici che rispecchiano le regioni corticali e sottocorticali di interesse ed un insieme di archi che rappresentano i fasci assonali tra queste regioni. Mediante immagini di risonanza magnetica, le quali permettono la ricostruzione dei fasci assonali, è possibile ottenere una mappatura completa delle connessioni neurali (*the structural connectome*) e la matrice di adiacenza ponderata $A = (A_{ij})$ ad essa associata. I *connettomi*

sono disponibili a cinque livelli di risoluzione [17]. Quelli a più bassa risoluzione sono rappresentati da grafi pesati di ordine $N = 83$, mentre grafi di ordine $N = 1015$ modellano i *connettomi* a più alta risoluzione.

Generalmente, il peso A_{jk} dell'arco tra il nodo j ed il nodo k è definito come il rapporto tra il numero di fasci assonali tra i due nodi e la lunghezza media dei fasci al quadrato. Dunque, a pesi maggiori corrispondono coppie di nodi maggiormente connessi, ossia collegati da un maggior numero di fasci assonali.

Osservazione 3.1. La *metrica di Lambert* caratterizza la locale connettività dei networks cerebrali. Più precisamente, per la *distanza di Lambert*, due nodi adiacenti sono tanto più vicini quanto più sono connessi.

Dimostrazione. Sia $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ un grafo pesato ed $A = (A_{ij})$ la sua matrice di adiacenza ponderata.

Ricordiamo che la *distanza di Lambert* tra due nodi adiacenti i e j è data da

$$t_{ij} = \frac{1}{\alpha} W_0 \left(\frac{\alpha}{\rho A_{ij}} \right)$$

con W_0 *funzione di Lambert*, $\alpha > 0$ costante di crescita e $\rho > 0$ costante di diffusione. Fissiamo quindi un nodo i e consideriamo due nodi j e k ad esso adiacenti. Banalmente, se $A_{ij} \leq A_{ik} \Rightarrow \frac{\alpha}{\rho A_{ik}} \leq \frac{\alpha}{\rho A_{ij}}$ e, poiché la diramazione principale della funzione di Lambert W_0 è monotona crescente (grafico in Figura 1.3), concludiamo che $A_{ij} \leq A_{ik} \Rightarrow t_{ik} \leq t_{ij}$. Ossia, maggiore è il numero di connessioni assonali tra coppie di nodi adiacenti, minore è la *distanza di Lambert* tra questi due nodi.

□

3.1 Scale temporali di progressione della malattia

Applicando il modello proposto a fenomeni di propagazione su networks, emerge una chiara separazione delle scale temporali di progressione della

malattia. Come già specificato, il modello da noi sviluppato descrive e comprende le dinamiche di invasione in regimi dominati dalla crescita, in cui la diffusione ha luogo solo in seguito ad un'iniziale aumento della concentrazione nelle regioni di semina. Pertanto, si può ottenere un primo tempo significativo studiando le dinamiche di crescita locale e stimando dopo quanto tempo una concentrazione iniziale $\beta < \frac{1}{2}$ raggiunge, localmente, un valore pari ad $\frac{1}{2}$. Sostituendo quindi $\mu = \frac{1}{2}$ in (2.16) otteniamo

$$T_0 = \frac{1}{\alpha} \log \left(\frac{1 - \beta}{\beta} \right)$$

e volendo ottenere una scala temporale tipica, indipendente dal livello di concentrazione iniziale del sistema, indichiamo come *scala temporale locale* il tempo

$$T_{\text{loc}} = \frac{1}{\alpha} \quad (3.1)$$

che governa la crescita locale di ogni nodo seme. Possiamo quindi assumere che dopo un lasso temporale pari a T_{loc} abbia inizio il processo diffusivo.

Fissato un nodo seme s , una seconda importante scala temporale si ottiene come tempo d'arrivo al primo nodo invaso. Tra i nodi adiacenti al nodo seme, il primo ad essere raggiunto è quello ad esso più connesso. Ossia, il nodo corrispondente al punto di massimo j tale che $\max_i A_{is} = A_{js}$. Dunque, indicando con $a_s = \max_i A_{is}$, per la regione di semina s si avrà

$$T_1(s) = \frac{1}{\alpha} W_0 \left(\frac{\alpha}{\rho a_s} \right) \quad , \quad a_s = \max_i A_{is} \quad (3.2)$$

a cui ci riferiremo in termini di *denial period locale*. Durante questo intervallo di tempo, la malattia sembra essere localizzata alla fonte e la sua propagazione ai nodi vicini non è ancora misurabile. Se si conosce o si sospetta un sito di semina, questa scala temporale risulta molto utile, in quanto fornisce una stima del tempo che trascorrerà prima che le regioni vicine siano invase. Ad esempio, nel caso della malattia di Alzheimer queste stime possono facilitare la comprensione dell'iniziale progressione della malattia, in quanto è ormai risaputo che la corteccia entorinale è la regione in cui essa ha inizio.

Se normalizziamo la matrice di adiacenza ponderata A dividendo ogni suo elemento per l'elemento massimo, per tutti i nodi seme si avrà $\max_i A_{is} = 1$. In tal caso, è possibile definire, per l'intero grafo, la scala temporale

$$T_1 = \min_s T_1(s) = \frac{1}{\alpha} W_0\left(\frac{\alpha}{\rho}\right). \quad (3.3)$$

T_1 viene detto *denial period globale* e quantifica l'iniziale ritardo di invasione della malattia da qualsiasi regione di semina. Più precisamente, rappresenta una stima dell'intervallo di tempo minimo necessario affinché una quantità di interesse (la proteina tossica *tau* nel caso della malattia di Alzheimer), si diffonda da un qualsiasi sito di semina all'interno del cervello.

Mentre il *denial period locale* dipende solo dalla connettività locale, quello *globale* (3.3) è legato ai pesi della matrice di adiacenza A e, quindi, al grafo del connettoma.

Per poter determinare il cosiddetto *periodo pandemico* della malattia, in seguito al quale il grafo risulta quasi completamente invaso, è utile stimare i tempi di invasione successivi alla prima. A tal fine, introduciamo la seguente nozione di distanza:

Definizione 3.2. Sia $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ un grafo pesato di ordine N e sia W la metrica di Lambert. Definiamo *distanza step* del grafo \mathcal{G} il vettore di lunghezza N la cui j -esima componente indica il numero di archi del cammino più breve, rispetto alla *distanza di Lambert* dal nodo di semina al nodo j .

Per $i > 1$ definiamo $T_i(s) = iT_1(s)$ come il tempo in cui i nodi che hanno valori i rispetto alla *distanza step*, vengono invasi da una concentrazione iniziale nel nodo seme s . Dunque, determiniamo il *periodo pandemico* come la scala temporale

$$T_{\max} = (l + 1)T_1$$

con l il massimo numero di archi tra due nodi qualsiasi del grafo, ossia la massima componente del vettore *distanza step*.

Abbiamo quindi individuato tre significative scale temporali per la progressione di una malattia che evolve seguendo le dinamiche di *Fisher-KPP su networks* (2.1):

- il periodo di crescita locale T_{loc} che determina l'istante in cui ha inizio la diffusione;
- il tempo T_1 in cui avviene la prima invasione;
- il *periodo pandemico* T_{max} in seguito a cui gran parte dei nodi del grafo risultano raggiunti.

Si tratta di intervalli temporali tipici, indipendenti dalla regione di semina e dalla concentrazione iniziale del sistema.

In Figura 3.1, mostriamo la soluzione dell'equazione di *Fisher-KPP* sul connettoma a più bassa risoluzione con $N = 83$ nodi, per fissate condizioni iniziali e valori dei parametri.

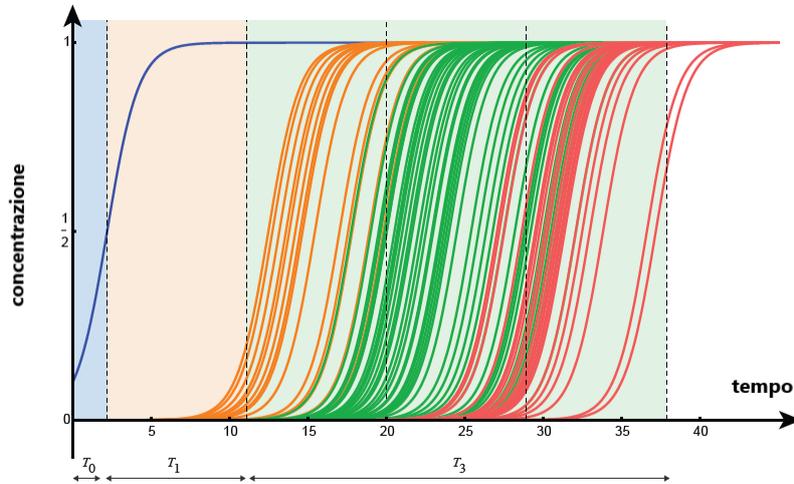


Figura 3.1: Stime delle scale temporali per il connettoma $N = 83$, con unico nodo seme $q_{27} = 1/10$ e $q_i = 0 \forall i \in 1, \dots, N$ e parametri: $\alpha = 1/\text{anno}$, $\rho = 0.001/\text{anno}$. In seguito alla crescita locale del nodo di semina con una scala temporale T_0 , il primo nodo viene invaso al tempo T_1 . I nodi successivi vengono invasi progressivamente con una scala temporale che dipende dalla *distanza step* dal nodo seme. In blu l'unico nodo seme (valore 0 rispetto alla *distanza step*); in arancione la concentrazione dei nodi che si trovano a *distanza step* pari a 1; in verdi quelli a distanza 2 ed in rosso i nodi a distanza 3.

3.2 Stadiazione Braak

In diverse aree della medicina, per valutare l'accuratezza della diagnosi clinica e guidare la gestione del paziente è necessaria la misurazione della gravità della malattia, che generalmente avviene attraverso sistemi di stadiazione. La malattia di Alzheimer (AD) è caratterizzata dalla presenza di grovigli neurofibrillari (NFT) di proteina *tau* e placche di proteina *beta-amiloide*. Entrambe queste proteine iniziano ad accumularsi nel cervello nell'arco dei due decenni precedenti alla comparsa dei sintomi clinici e la possibilità di quantificarne il carico, oggi permessa grazie allo sviluppo di nuove tecnologie, ha cambiato di molto l'approccio alla diagnosi di AD, consentendo di individuare ad uno stadio molto precoce la patologia. Inoltre, si ritiene che la diffusione dei grovigli neurofibrillari di *tau* sia correlata alla progressione della malattia. In particolare, la loro propagazione all'interno del cervello rispecchia il declino cognitivo del paziente e quindi la progressiva perdita di funzionalità delle regioni cerebrali. Pertanto, la localizzazione anatomica dei grovigli di tau può permetterci di identificare gli stadi di questa malattia. Studi istopatologici post-mortem hanno osservato che gli aggregati di proteina *tau* si diffondono seguendo una sequenza a sei stadi [18, 19], chiamati *i sei stadi di Braak* e descritti da Braak nel 1991 [20]. Più precisamente, questo sistema di stadiazione classifica l'AD in base alla presenza dei grovigli neurofibrillari nelle seguenti regioni cerebrali:

- corteccia transentorinale ed entorinale (**Stadio I**);
- lombo limbico ed ippocampo (**Stadio II e III**);
- corteccia cingolata anteriore (**Stadio IV**);
- area di associazione visiva (**Stadio V**);
- corteccie visiva primaria (**Stadio VI**).

Il passaggio di aggregati di *tau* tra queste regioni può essere spiegato dalla capacità di questa proteina di attraversare gli assoni e di spostarsi da un

neurone all'altro, come suggerito da diversi studi in vitro e in vivo [20].

Ciascuno stadio è associato a sintomi clinici differenti e quindi a fasi diverse della malattia. Per guidare la gestione dei pazienti si fa generalmente riferimento alla seguente gerarchia:

- l'AD asintomatico corrisponde allo stadio I;
- l'AD incipiente fa riferimento agli stadi II e III;
- si parla invece di AD sintomatico dal IV stadio in poi, quando gli aggregati tossici di proteina *tau* raggiungono la neocorteccia.

In particolare, se un paziente presenta un deposito di *tau* nella corteccia visiva primaria (ultimo stadio di Braak) significa che si trova nella fase finale della malattia e che grovigli di questa proteina sono presenti anche in tutte le altre aree coinvolte.

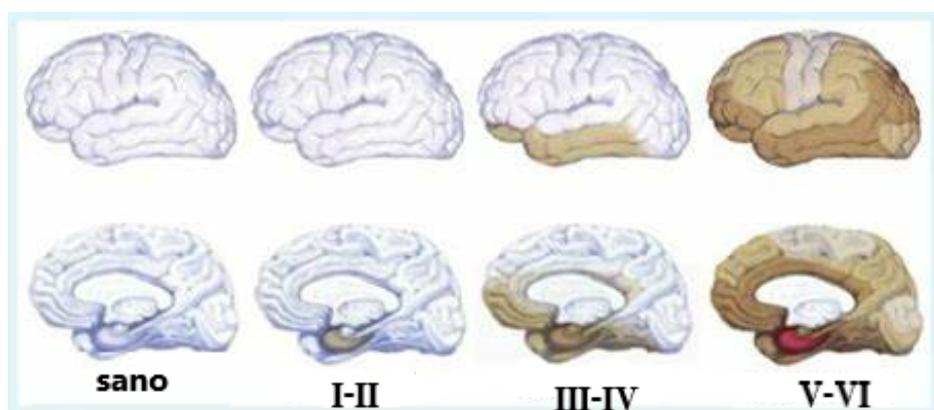


Figura 3.2: Aree dell'encefalo coinvolte nei diversi stadi di Braak

Nonostante la semplicità del modello di *Fisher-KPP* su networks, la sua dinamica descrive bene la progressione generalizzata dei grovigli neurofibrillari di *tau* osservati negli studi post-mortem. Infatti, simulazioni ottenute risolvendo il sistema (2.1) su un connettoma cerebrale con sito di semina

iniziale la regione I di Braak mostrano che gli aggregati di tau raggiungo le regioni successive alla corteccia entorinale seguendo la gerarchia della stadiazione di Braak. In Figura 3.3 un esempio di queste simulazioni.

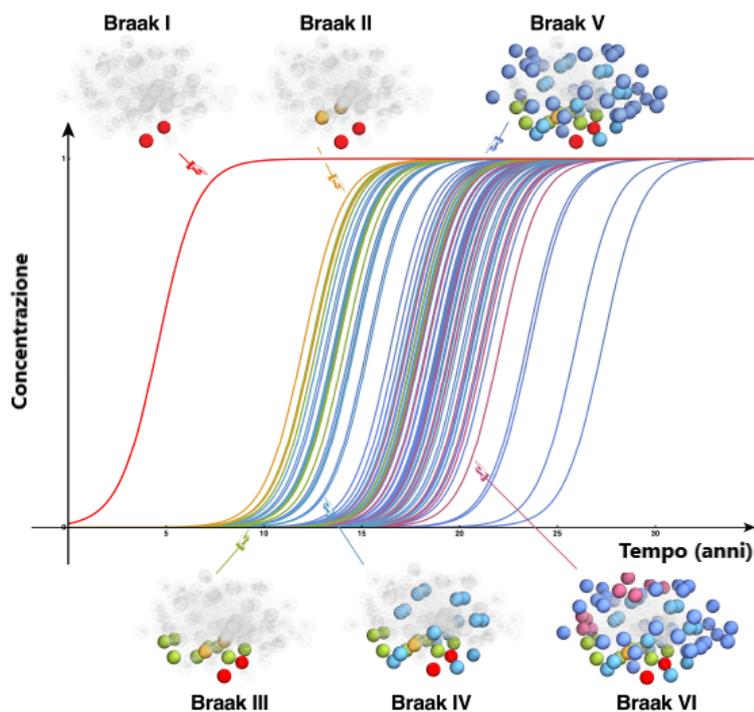


Figura 3.3: Grafico delle soluzioni dell'equazione di *Fisher-KPP* sul connettoma a più bassa risoluzione $N = 98$. Le regioni di semina, corrispondenti allo stadio I di Braak, sono rappresentate dai nodi 27 e 68. Parametri: $\alpha = \frac{1}{\text{anno}}$, $\rho = \frac{5.64 \cdot 10^{-3}}{\text{anno}}$. Condizioni iniziali: $p_{27} = p_{68} = 1/100$ e $p_i = 0 \forall 1, \dots, N$. Ognuno dei sei colori corrisponde alle regioni corticali che compongono uno stadio di Braak (I: rosso; II: giallo; III: verde; IV: azzurro; V: blu; VI: rosa).

Appendice A

Analisi qualitativa di un sistema bidimensionale

Conoscere il comportamento della soluzione di un sistema di equazioni differenziali, è spesso più utile che ricavarla esplicitamente. Inoltre, nel caso di problemi non lineari, determinare una soluzione sia analitica che numerica può risultare alquanto complicato, se non impossibile. In molti casi, uno studio diretto dell'equazione, permette di studiare l'andamento delle curve integrali senza conoscerne l'espressione analitica. Si parla, quindi, di analisi qualitativa del sistema e, di seguito, ne discutiamo brevemente.

Consideriamo un sistema bidimensionale della forma

$$\begin{cases} x' = f_1(x, y) \\ y' = f_2(x, y) \end{cases} \quad (\text{A.1})$$

con $f = (f_1, f_2): D \rightarrow \mathbb{R}^2$, $f \in C^1(D, \mathbb{R}^2)$ e D aperto di \mathbb{R}^2 .

Definizione A.1. Sia $\phi: I \rightarrow \mathbb{R}^2$, con $I \subset \mathbb{R}$, soluzione di (A.1). Allora,

- (i) il grafico di ϕ è una curva parametrica in \mathbb{R}^3 ed è detto *curva integrale* del sistema;
- (ii) il piano contenente le variabili x e y è detto *piano delle fasi*;

(iii) la proiezione della curva integrale sul piano delle fasi è una curva piana e orientata detta *orbita*.

Definizione A.2. Il punto $(x_0, y_0) \in D$ si dice *stato stazionario* (o *punto di equilibrio*) per il sistema (A.1) se $f_1(x_0, y_0) = f_2(x_0, y_0) = 0$.

Lo studio qualitativo del problema presuppone un'analisi della stabilità dei suoi punti stazionari. Diamo una classificazione prendendo come modello di riferimento un sistema bidimensionale lineare, $y' = f(y)$, che ha come unico punto di equilibrio l'origine (se $\tilde{y} = (x_0, y_0) \neq (0, 0)$ basta porre $z = y - \tilde{y}$ e studiare il sistema $z' = g(z) := f(z + \tilde{y})$).

Sia, quindi, $\begin{pmatrix} x \\ y \end{pmatrix}' = A \begin{pmatrix} x \\ y \end{pmatrix}$ un sistema lineare bidimensionale ed indichiamo con $\lambda_{1,2}$ gli autovalori della matrice A . Allora,

$$\lambda_{1,2} = \frac{TrA \pm \sqrt{(TrA)^2 - 4detA}}{2} \quad \text{con } \Delta = TrA^2 - 4detA .$$

Classifichiamo il tipo di stabilità di $(0, 0)$ in base al segno di Δ e, quindi, al segno di TrA e $detA$. Suddividiamo lo studio degli autovalori in tre casi:

$\Delta > 0$:

In questo caso gli autovalori sono reali e distinti quindi studiamo la stabilità in base al loro segno:

- se $\lambda_1 < \lambda_2 < 0 \Rightarrow TrA < 0$, $detA > 0$ e l'origine è un **nodo stabile**;
- se $\lambda_2 > \lambda_1 > 0 \Rightarrow TrA > 0$, $detA > 0$ e l'origine è un **nodo instabile**;
- se $\lambda_1 < 0 < \lambda_2 \Rightarrow detA < 0$ ($TrA < 0$ o $TrA > 0$) e l'origine è un **nodo a sella**.

$\Delta = 0$:

L'autovalore λ è doppio, dunque suddividiamo i casi in base alla sua regolarità:

- se $ma(\lambda) = mg(\lambda)$ allora $(0, 0)$ è un **nodo stabile** se $\lambda < 0$, un **nodo instabile** se $\lambda > 0$;
- se $ma(\lambda) \neq mg(\lambda)$ allora $(0, 0)$ è un **nodo a stella stabile** se $\lambda < 0$, un **nodo a stella instabile** se $\lambda > 0$.

$\Delta < 0$:

Gli autovalori sono complessi coniugati $\lambda_1 = \alpha + i\beta$ e $\lambda_2 = \alpha - i\beta$ e il tipo di stabilità di $(0, 0)$ dipende dal segno di $Re(\lambda_{1,2}) = \alpha$:

- se $\alpha > 0$ l'origine è un **fuoco instabile**;
- se $\alpha = 0$ l'origine ha una stabilità di tipo **centro**;
- se $\alpha < 0$ l'origine è un **fuoco stabile**.

Per poterli visualizzare, riassumiamo graficamente i tipi di stabilità appena descritti:

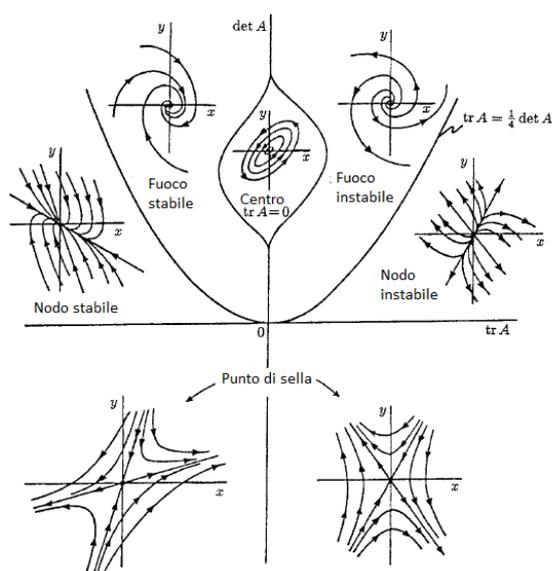


Figura A.1: Schema riassuntivo che mostra come $\text{tr} A$ e $\det A$ determinano il tipo di singolarità del sistema lineare

Nel caso non lineare, si linearizza il sistema intorno al punto di equilibrio. In questo modo si ottiene un problema lineare la cui matrice associata è la Jacobiana di f ($A = J_f$) e si studia la stabilità del punto a partire da questo sistema. Come mostra il seguente Teorema, nel passaggio da lineare a non lineare, l'asintotica stabilità e l'instabilità del punto si mantengono, mentre non si può dire nulla per gli stati stazionari stabili:

Teorema A.3. *Sia $y' = f(y)$ un sistema non lineare, con $f \in C^1$ tale che $f(0) = 0$ e $\det(J_f(0)) \neq 0$. Sia $Z' = AZ$ il sistema linearizzato intorno all'origine, con $A = J_f(0)$. Allora,*

- (i) *se 0 è asintoticamente stabile per il sistema linearizzato, lo è anche per il sistema non lineare;*
- (ii) *se la matrice A possiede un autovalore λ con $\operatorname{Re}(\lambda) > 0$ (ovvero l'origine è un punto instabile) allora è instabile anche per il sistema non lineare.*

In conclusione, dal sistema lineare a quello non lineare si trasportano i nodi, i fuochi e le selle.

Un altro importante risultato per la teoria qualitativa locale delle equazioni differenziali ordinarie è dato dal *Teorema della varietà stabile* [21], il quale afferma che, intorno ad un punto di equilibrio iperbolico, il sistema non lineare può essere approssimato con il suo linearizzato. Per poterlo enunciare ricordiamo le seguenti definizioni:

Definizione A.4. y^* si dice *punto di equilibrio iperbolico* per il sistema di equazioni differenziali autonomo $y' = F(y)$ se è un punto di equilibrio del sistema e nessun autovalore della Jacobiana del sistema in y^* , $J_F(y^*)$, ha parte reale nulla.

Definizione A.5. Sia $y' = F(y)$ con $F: A \rightarrow \mathbb{R}^n$, A aperto di \mathbb{R}^n ed $F \in C^1(A)$ un sistema di equazioni differenziali. $S \subset A$ si dice positivamente (negativamente) invariante per il sistema se $\forall \phi$ soluzione del sistema t.c.

$\phi(t_0) \in S$ (per un certo t_0) si ha che $\phi(t) \in S \quad \forall t \geq t_0 \quad (\forall t \leq t_0)$.

S viene detta invariante se è sia positivamente che negativamente invariante.

Teorema A.6. *Dato il sistema di equazioni differenziali $y' \stackrel{(*)}{=} F(y)$ con $F \in C^1(A)$ ed $A \in \mathbb{R}^n$, sia $y^* \in A$ un punto di equilibrio iperbolico del sistema. Se la matrice Jacobiana di F in y^* , $J_F(y^*)$, ha $k \leq n$ autovalori con parte reale negativa ed $n-k$ con parte reale positiva, allora esiste una varietà E k -dimensionale, differenziabile, positivamente invariante, avente come spazio tangente in y^* il sottospazio stabile del linearizzato del sistema $(*)$ in y^* e*

$$\forall p \in E \quad \phi(t, p) \xrightarrow[t \rightarrow \infty]{} y^* .$$

Esiste inoltre una varietà U $(n-k)$ -dimensionale, differenziabile, negativamente invariante, avente come spazio tangente in y^ il sottospazio instabile del linearizzato del sistema $(*)$ in y^* e*

$$\forall p \in U \quad \phi(t, p) \xrightarrow[t \rightarrow -\infty]{} y^* .$$

Per concludere, diamo un utile criterio per determinare il tipo di stabilità di un punto stazionario nel caso di equazioni differenziali autonome unidimensionali:

Teorema A.7. *Sia $x'(t) = f(x(t))$ un'equazione differenziale con x_0 punto di equilibrio ($f(x_0) = 0$). Allora,*

- se $f'(x_0) < 0$, x_0 è punto di equilibrio stabile (asintoticamente);
- se $f'(x_0) > 0$ x_0 è punto di equilibrio instabile.

Dimostrazione. si veda [22].

□

Bibliografia

- [1] Prama Putra, Hadrien Oliveri, Travis Thompson, and Alain Goriely. Front propagation and arrival times in networks with application to neurodegenerative diseases. *Mathematical Institute, University of Oxford*. January 4, 2022.
- [2] Murray JD. 2002, *Mathematical biology I. An introduction*. New York, NY: Springer.
- [3] Jonkhout 2016, *Traveling wave solutions of reaction-diffusion equations in population dynamics*, Mathematical Institute, University of Leiden.
- [4] Fan RK Chung and Fan Chung Graham. *Spectral graph theory*. Number 92. American Mathematical Soc., 1997
- [5] Prama Putra, Travis B Thompson, Pavanjit Chaggar, and Alain Goriely. Braiding Braak and Braak: Staging patterns and model selection in network neurodegeneration. *Network Neuroscience*, 5(4):1-41, 2021.
- [6] Corless, R.M., Gonnet, G.H., Hare, D.E., Jeffrey, D.J. & Knuth, D.E. (1996) *On the Lambert W function*. *Advances in Computational Mathematics*, 5, 329–359.
- [7] Evolution and Ecology Research Centre, School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, New South Wales 2052, Australia; 2016. *The Lambert W function in ecological and evolutionary models*

-
- [8] Sarah K Kaufman, Kelly Del Tredici, Talitha L Thomas, Heiko Braak, and Marc I Diamond. Tau seeding activity begins in the transentorhinal/entorhinal regions and anticipates phospho-tau pathology in Alzheimer's disease and PART. *Acta neuropathologica*, 2018.
- [9] Chittaranjan Hens, Uzi Harush, Simi Haber, Reuven Cohen, and Baruch Barzel. *Spatiotemporal signal propagation in complex networks*. Nature Physics, 2019.
- [10] https://it.wikipedia.org/wiki/Matrice_esponenziale, consultato il giorno 02/03/2023.
- [11] Aurélien Gautreau, Alain Barrat, and Marc Barthelemy. Global disease spread: statistics and estimation of arrival times. *Journal of theoretical biology*, 251(3):509-522, 2008.
- [12] Flavio Iannelli, Andreas Koher, Dirk Brockmann, Philipp Hovel, and Igor M Sokolov. Effective distances for epidemics spreading on complex networks. *Physical Review E*, 95(1):012313, 2017.
- [13] Dirk Brockmann and Dirk Helbing. The hidden geometry of complex, network-driven contagion phenomena. *Science*, 342(6164):1337-1342, 2013.
- [14] Alastair Jamieson-Lane and Bernd Blasius. Calculation of epidemic arrival time distributions using branching processes. *Physical Review E*, 102(4):042301, 2020.
- [15] Johannes Weickenmeier, Ellen Kuhl, and Alain Goriely. Multiphysics of Prionlike Diseases: Progression and Atrophy. *Physical Review Letters*, 121(15):158101, 2018.
- [16] Amelie Schafer, Elizabeth C Mormino, and Ellen Kuhl. Network diffusion modeling explains longitudinal tau pet data. *Frontiers in Neuroscience*, 14:1370, 2020.

-
- [17] Csaba Kerepesi, Balázs Szalkai, Bálint Varga, and Vince Grolmusz. The braingraph.org database of high resolution structural connectomes and the brain graph tools. *Cognitive Neurodynamics*, 11: 483-486, 2017.
- [18] Heiko Braak and Eva Braak. Neuropathological staging of Alzheimer-related changes. *Acta Neuropathologica*, 82(4):239-259, 1991.
- [19] Heiko Braak, Irina Alafuzoff, Thomas Arzberger, Hans Kretschmar, and Kelly Del Tredici. Staging of Alzheimer disease-associated neurofibrillary pathology using paraffin sections and immunocytochemistry. *Acta Neuropathologica*, 112(4):389-404, 2006.
- [20] Sarah L DeVos, Bianca T Corjuc, Derek H Oakley, Chloe K Nobuhara, Riley N Bannon, Alison Chase, Caitlin Commins, Jose A Gonzalez, Patrick M Dooley, Matthew P Frosch, et al. Synaptic tau seeding precedes tau pathology in human Alzheimer's disease brain. *Frontiers in Neuroscience*, 12:267, 2018.
- [21] Corrado Mascia, Eugenio Montefusco, "Un invito alla biomatematica", edizioni La Dotta, 2014.
- [22] Bramanti, Pagani, Salsa. *Analisi matematica 2*, Zanichelli, 2009.