

Storia degli automi cellulari

Una prospettiva bibliometrica

Carlo Debernardi

Relatore

Prof. Riccardo Bruni

Correlatore

Prof. Enrico Pasini



Università degli Studi di Firenze

A.A. 2019/2020

Abstract

Historiographical attention paid to cellular automata has so far been rather scarce, often limited to the introduction of technical monographs. One of the aims of this work is therefore to provide an account of the development of research work on cellular automata from a historical point of view. In particular, the first chapter describes the birth of cellular automata by directly analyzing the primary literature, consisting above all of von Neumann's contributions. This work also aims to map the theoretical and applicative legacy of cellular automata in the years following their diffusion, with particular attention to the late '80s. In this period cellular automata began to be studied systematically as models of complex systems. This is pursued in the second part of the work with the use of methods borrowed from bibliometrics and computational linguistics. The application of these techniques allowed a disciplinary characterization of papers concerning cellular automata in the period 1985-1995. The iterative use of the clustering process is also proposed to account for the evolution of research themes in this area over time.

Questo lavoro è giunto a buon fine grazie al supporto e ai consigli di molte persone. Un ringraziamento particolare va ai due relatori che hanno seguito il mio lavoro e ad Eugenio Petrovich che mi ha introdotto alle vie della bibliometria. Marco Viola ed Eleonora Priori mi hanno sostenuto nei primi passi della vita accademica. Un grazie enorme anche a Gio, alla mia famiglia, ai miei coinquilini e a tutte le compagne ed i compagni che in questi anni ho avuto al mio fianco.

Indice

Introduzione	4
0.1 Perché una storia degli automi cellulari	4
0.2 Automi cellulari: una definizione minimale	8
1 Le origini degli automi cellulari	13
1.1 I primi passi	14
1.2 Espansione e consolidamento della comunità di ricerca	21
2 Automi cellulari: uno studio quantitativo	29
2.1 Bibliometria: una breve panoramica	29
2.2 Gli sviluppi recenti sotto la lente bibliometrica	36
3 Conclusioni	59
Bibliografia	61

Introduzione

A volte gli scienziati si ingannano da soli, credendo che le idee filosofiche siano soltanto, nella migliore delle ipotesi, ornamento oppure osservazioni parassitiche dei trionfi, solidi e oggettivi, della scienza e pensando di non essere essi stessi soggetti a quegli equivoci alla cui dissipazione i filosofi dedicano la vita. Ma non esiste una scienza affrancata dalla filosofia; esiste soltanto una scienza il cui bagaglio filosofico viene caricato a bordo senza controllo.

(Dennett 1995)¹

Come nel processo di fabbricazione così anche nella scienza il cambiamento di strumenti è una stravaganza che va riservata per l'occasione che lo richiede. Il significato delle crisi sta nell'indicazione, da esse fornita, che l'occasione per cambiare strumenti è arrivata.

(Kuhn 1970)²

0.1 Perché una storia degli automi cellulari

Negli ultimi decenni, a cavallo tra discipline e ambiti di ricerca svariati, la cosiddetta “teoria della complessità” ha acquisito sempre più rilevanza. Sotto questa etichetta generale si trova un mondo multiforme, la cui varietà spazia da contributi teorici rigorosamente formalizzati ad applicazioni computazionali talvolta inspiegabilmente efficaci. Naturalmente, come per ogni moda nella scienza, sotto il cappello della complessità è spesso possibile trovare anche – avremo l'occasione di mostrare

¹“Scientists sometimes deceive themselves into thinking that philosophical ideas are only, at best, decorations or parasitic commentaries on the hard, objective triumphs of science, and that they themselves are immune to the confusions that philosophers devote their lives to dissolving. But there is no such thing as philosophy-free science; there is only science whose philosophical baggage is taken on board without examination.”

²“As in manufacture so in science—retooling is an extravagance to be reserved for the occasion that demands it. The significance of crises is the indication they provide that an occasion for retooling has arrived.”

che non è una novità – contributi controversi. È tuttavia fuor di dubbio che questo termine ricopra un ruolo centrale nella produzione scientifica degli ultimi anni. Il lavoro di ricerca che ha riguardato, e riguarda tuttora, gli automi cellulari rappresenta uno dei contributi sostanziali che hanno permesso lo sviluppo delle teorie che riguardano la complessità.

Dal punto di vista teorico, gli automi cellulari costituiscono una modalità alternativa – ma complementare – alle altre proposte per la descrizione dei processi computazionali, si pensi ad esempio alle macchine di Turing. Nella pratica scientifica hanno permesso di approssimare soluzioni a problemi altrimenti irrisolvibili, di sviluppare nuovi metodi di simulazione, di rendere conto di dinamiche evolutive e persino di immaginare nuove fondazioni per la fisica. Dal punto di vista tecnologico le proposte di loro applicazione vanno dalla generazione di numeri casuali al riconoscimento di *pattern* nelle immagini, per arrivare alla crittografia e alla progettazione di nuove architetture computazionali.

Malgrado questa loro indiscutibile importanza, l'attenzione storiografica rivolta agli automi cellulari è stata finora alquanto scarsa. È possibile immaginare motivazioni molteplici – dopotutto si tratta di storia alquanto recente, e gli anni e i protagonisti della loro nascita sono al centro dei riflettori per contributi scientifici ben più celebri – ma sta di fatto che la maggior parte delle opere che trattano gli automi cellulari consistono in introduzioni tecniche al loro studio, e limitano la trattazione storica allo spazio che – nelle introduzioni delle monografie – viene tradizionalmente riservato alla contestualizzazione.

Uno degli obiettivi di questo lavoro è dunque quello di fornire un resoconto dello sviluppo del lavoro di ricerca sugli automi cellulari dal punto di vista storico. In particolare, il primo capitolo descrive la nascita degli automi cellulari confrontandosi direttamente con la letteratura primaria, rappresentata soprattutto dai contributi di von Neumann. Tenteremo inoltre di ripercorrere le tappe fondamentali nella formazione della comunità di ricerca che li ha studiati e continua a farlo. Una traccia utile in questa ricostruzione è rappresentata dalla già citata contestualizzazione storica solitamente collocata nelle introduzioni delle monografie. L'analisi comparativa di questi testi fornisce infatti *i)* un'indicazione circa lo stato dell'arte della ricerca e le questioni aperte e *ii)* un tentativo di razionalizzazione del percorso storico – spesso svolto direttamente dagli esperti che vi hanno preso parte. Di entrambi questi aspetti è possibile cogliere le somiglianze e i punti di discontinuità in una prospettiva diacronica. In questo senso, svolgono un ruolo particolarmente rilevante le seguenti monografie:

- Toffoli e Margolus (1987), *Cellular Automata Machines: A New Environment for Modeling*
- Ilachinski (2001), *Cellular Automata: A Discrete Universe*
- Schiff (2007), *Cellular Automata: A Discrete View of the World*

Un'ulteriore fonte, importante soprattutto per quanto riguarda la ricostruzione storica, è rappresentata da prefazione e introduzione al volume *Theory Of Self Reproducing Automata* (1966); collezione di due diversi manoscritti di von Neumann pubblicata postuma e curata da Arthur Burks.

Questo lavoro si propone anche di mappare l'eredità teorica e applicativa degli automi cellulari negli anni successivi alla loro diffusione. L'utilizzo di metodi storiografici tradizionali e la lettura diretta dei testi rilevanti sarebbero risultati inadeguati a questo fine; in primo luogo a causa dell'insufficiente dimestichezza di chi scrive con teorie e metodi formali tanto diversi tra loro e distribuiti su ambiti disciplinari diversi, in secondo luogo per la dimensione stessa della letteratura scientifica sul tema.

La presenza di questi limiti è divenuta quindi l'occasione per l'applicazione di tecniche quantitative che negli ultimi anni si stanno diffondendo anche come strumenti di supporto nelle scienze umane. In particolare il secondo capitolo introduce e fa uso di metodi mutuati dalla bibliometria e dalla linguistica computazionale. Malgrado – o forse a causa – della versatilità di queste tecniche, non esiste al momento uno standard condiviso per la loro applicazione. Il lavoro proposto nel secondo capitolo prende dunque le mosse da alcuni dei metodi più diffusi, e si conclude con una proposta originale di applicazione combinata di tecniche già note. Sono liberamente disponibili online diversi *software* dedicati all'analisi bibliometrica e al *science mapping*, tuttavia – dal momento che sono sviluppati per consentire una certa semplicità di utilizzo – il loro funzionamento non è sempre chiaro e ben documentato. Per il presente lavoro si è dunque deciso di utilizzare prevalentemente codice scritto *ad hoc*³.

Nel seguito del lavoro faremo riferimento allo studio degli automi cellulari – inteso come insieme di aspetti epistemici (tesi, metodi, ...) e formali (pubblicazioni scientifiche) – in termini di *ambito di ricerca*, mentre parleremo di *comunità di ricerca* intendendo l'insieme delle persone che a questo studio hanno contribuito. Questa precisazione si rende necessaria poiché, sfortunatamente, la tassonomia delle ca-

³Il codice utilizzato per l'analisi quantitativa è pubblicamente disponibile all'indirizzo seguente: https://github.com/carlobebernardi/cellular-automata-bibliometrics/blob/master/ca_bibliometrics.py

tegorie epistemiche e sociologiche necessarie a un'analisi come quella in oggetto è tutt'altro che condivisa⁴.

Un aspetto su cui la maggior parte degli autori concorda è invece che le dimensioni delle unità fondamentali della ricerca siano ben inferiori a quelle disciplinari, e che in particolare debbano corrispondere a pubblicazioni, riviste e relazioni personali gestibili da un individuo o da un piccolo gruppo (Kuhn 1970; Price 1963). Nell'accingerci allo studio di una particolare comunità di ricerca terremo dunque presente la definizione data da Morris e Veer Martens (2009):

[...] *una specialità di ricerca è un network auto organizzato di ricercatori che tendono a studiare gli stessi argomenti, a partecipare alle stesse conferenze, a leggere e citare vicendevolmente i propri articoli e a pubblicare sulle stesse riviste scientifiche.* Una specialità di ricerca produce, nel tempo, un corpus cumulativo di conoscenze, che si concretizza in tesi di laurea, libri, atti di convegni e in una letteratura permanente su riviste. I membri di una specialità di ricerca tendono anche a condividere e utilizzare, a un certo livello, una cornice di conoscenze di base, che include la conoscenza di teorie, dati sperimentali, tecniche, standard di validazione, esempi paradigmatici, contraddizioni problematiche e controversie.⁵

Lo studio di queste unità della ricerca vanta una lunga tradizione nella sociologia e nella storia della scienza, per questo la scelta di utilizzare metodi quantitativi non largamente impiegati in questi ambiti disciplinari non deve essere interpretata come un rimpiazzamento. È invece opinione di chi scrive che la relativamente recente disponibilità di dati e mezzi computazionali non possa che arricchire di nuovi strumenti queste tradizioni di ricerca consolidate.

⁴Per una rassegna delle diverse proposte si rimanda a Morris e Veer Martens (2009) e Sugimoto e Weingart (2015)

⁵“[...] *a research specialty is a self-organized network of researchers who tend to study the same research topics, attend the same conferences, read and cite each other's research papers and publish in the same research journals.* A research specialty produces, over time, a cumulating corpus of knowledge, embodied in educational theses, books, conference papers, and a permanent journal literature. Members of a research specialty also tend to share and use, to some degree, a framework of base knowledge, which includes knowledge of theories, experimental data, techniques, validation standards, exemplars, worrisome contradictions, and controversies.”

0.2 Automi cellulari: una definizione minimale

Un tipico automa cellulare consiste in uno spazio regolare popolato di celle discrete, ognuna delle quali può assumere diversi stati. Le celle aggiornano il loro stato in funzione dello stato delle celle adiacenti e lo fanno *i)* a tempi discreti e *ii)* in parallelo. In particolare, in questo contesto, il tempo consiste in una successione di stati globali del sistema.

Formalmente un automa cellulare può essere inteso come la composizione di quattro elementi:

$$\langle V, S, A, R \rangle$$

Il primo di questi – V – è l'insieme dei possibili valori assunti dalle celle. Negli automi cellulari tradizionali questo insieme è numerabile (ovvero possiede al più la cardinalità di \mathbb{N}). Dal momento che la loro funzione è quella di “etichettare” le celle, questi valori possono essere rappresentati in modi svariati; oltre ai numeri naturali stessi vengono spesso impiegati simboli distinguibili per forma e/o colore come \square e \blacksquare .

Gli automi cellulari possono essere costruiti in uno spazio S con una o più dimensioni, inoltre – come si è detto – lo spazio in cui vengono definiti è discreto. Le celle di un automa in una dimensione possono quindi essere messe in corrispondenza con i numeri interi, quelle di un automa in due dimensioni con coppie di interi e così via. Se d è il numero di dimensioni scelto per lo spazio avremo dunque:

$$S = \mathbb{Z}^d$$

Possiamo ora introdurre la regola di aggiornamento R , una funzione in due variabili che, date le coordinate x di una cella ed il tempo t , restituisce il valore della cella al tempo dato. Avremo dunque:

$$R : \{ \langle x, t \rangle \mid x \in S, t \in \mathbb{N} \} \rightarrow V$$

Una definizione esplicita di R richiede la scelta di una relazione di adiacenza A tra le celle, nei termini di una funzione con dominio S e codominio l'insieme di ennuple di elementi di S . Fissato il numero n di celle adiacenti ad ognuna avremo dunque:

$$A : S \rightarrow \{ \langle x_0, x_1, \dots, x_n \rangle \mid x \in S \}$$

A questo punto possiamo definire R al passare del tempo t come:

$$R(x, t+1) = f(R(x_0, t), R(x_1, t), \dots, R(x_n, t)) \quad x_0, x_1, \dots, x_n \in A(x)$$

Dove f è la funzione che esplicita la relazione tra input – i valori delle celle adiacenti al tempo t – ed output – il valore della cella considerata al tempo $t+1$, ed ha la

seguinte forma:

$$f : \{ \langle R(x_0, t), R(x_1, t), \dots R(x_n, t) \rangle \mid x_0, x_1, \dots x_n \in A(x) \} \rightarrow V$$

Si noti che R deve essere ben definita su tutte le $|V|^n$ possibili combinazioni di valori delle celle nella relazione di adiacenza⁶ (per maggiore chiarezza si veda l'esempio di seguito).

Sulla base di questa definizione di R possiamo fissare lo stato s dell'automa al tempo t come segue:

$$s_t = \{ \langle x, R(x, t) \rangle \mid x \in S \} \quad t \in \mathbb{N}$$

Questa definizione degli stati rende particolarmente evidente la natura *parallela* della computazione operata dagli automi cellulari tradizionali. È infatti possibile notare come, nel passaggio dal tempo t al tempo $t + 1$ *tutte* le celle aggiornino il proprio valore secondo la regola di aggiornamento R . Grazie a questa caratteristica non troviamo traccia di interferenze date dall'ordine di aggiornamento delle celle, tutte le celle si aggiornano contemporaneamente ed il loro nuovo valore dipende unicamente dai valori delle celle adiacenti al tempo precedente.

Infine, i valori delle celle allo stato iniziale s_0 (e dunque il comportamento di $R(x, 0)$) devono essere fissati arbitrariamente.

Per chiarire maggiormente questa descrizione vedremo di seguito la definizione formale di un semplice automa cellulare binario in una dimensione.

Le celle potranno dunque assumere due soli valori:

$$V = \{\square, \blacksquare\}$$

Essendo in una sola dimensione, avremo:

$$S = \mathbb{Z}$$

Prenderemo poi come adiacenti le celle immediatamente precedente e successiva a quella data (ed avremo $n = 3$):

$$A(x) = \langle x - 1, x, x + 1 \rangle$$

La regola di aggiornamento dovrà dunque essere definita su tutte e 8 ($|V|^n = 2^3$) le possibili combinazioni di valori per le celle adiacenti ed avrà la forma funzionale seguente:

$$R(x, t + 1) = f(R(x - 1, t), R(x, t), R(x + 1, t))$$

⁶ $|V|$ indica la cardinalità dell'insieme V ed n il numero di celle definite dalla relazione di adiacenza. Date queste due quantità è possibile ricavare anche il numero di possibili regole di aggiornamento per l'automa, ovvero $|V|^{|V|^n}$.

A titolo di esempio considereremo una regola di aggiornamento rappresentata in questo modo:

$$R(x, t+1) = \begin{cases} (R(x-1, t) = \blacksquare \wedge R(x, t) = \blacksquare \wedge R(x+1, t) = \blacksquare) \implies \square \\ (R(x-1, t) = \blacksquare \wedge R(x, t) = \blacksquare \wedge R(x+1, t) = \square) \implies \blacksquare \\ (R(x-1, t) = \blacksquare \wedge R(x, t) = \square \wedge R(x+1, t) = \blacksquare) \implies \square \\ (R(x-1, t) = \blacksquare \wedge R(x, t) = \square \wedge R(x+1, t) = \square) \implies \blacksquare \\ (R(x-1, t) = \square \wedge R(x, t) = \blacksquare \wedge R(x+1, t) = \blacksquare) \implies \blacksquare \\ (R(x-1, t) = \square \wedge R(x, t) = \blacksquare \wedge R(x+1, t) = \square) \implies \square \\ (R(x-1, t) = \square \wedge R(x, t) = \square \wedge R(x+1, t) = \blacksquare) \implies \blacksquare \\ (R(x-1, t) = \square \wedge R(x, t) = \square \wedge R(x+1, t) = \square) \implies \square \end{cases}$$

Risulta evidente la scarsa maneggevolezza della regola di aggiornamento espressa in questi termini. Solitamente si preferisce riportarla in una forma grafica intuitiva:

$$\begin{array}{ll} \blacksquare \blacksquare \blacksquare \implies \square & \blacksquare \blacksquare \square \implies \blacksquare \\ \blacksquare \square \blacksquare \implies \square & \blacksquare \square \square \implies \blacksquare \\ \square \blacksquare \blacksquare \implies \blacksquare & \square \blacksquare \square \implies \square \\ \square \square \blacksquare \implies \blacksquare & \square \square \square \implies \square \end{array}$$

Nella notazione precedente la freccia \implies indica il passaggio dai valori di input (a sinistra) ai valori di output (sulla destra). Ad esempio:

$$(R(x-1, t) = \square \wedge R(x, t) = \blacksquare \wedge R(x+1, t) = \square) \implies \square$$

Significa che, se al tempo t le celle $x-1$ ed $x+1$ hanno valore \square mentre la cella x ha valore \blacksquare , al tempo $t+1$ la cella x avrà valore \square . Questo enunciato è espresso in forma grafica come:

$$\square \blacksquare \square \implies \square$$

Come detto in precedenza, è anche necessario selezionare arbitrariamente lo stato iniziale dell'automa. In questo caso fisseremo:

$$R(x, 0) = \begin{cases} x = 0 \implies \blacksquare \\ x \neq 0 \implies \square \end{cases}$$

Il nostro stato iniziale vedrà dunque la cella 0 con valore \blacksquare e tutte le altre \square :

$$\dots \square \square \square \square \square \blacksquare \square \square \square \square \square \dots$$

A questo punto possiamo osservare l'evoluzione nel tempo dell'automa cellulare ottenuto:

$$\begin{aligned}
 s_0 &= \cdots \square\square\square\square\square\square\square\square\square\square\square\square \cdots \\
 s_1 &= \cdots \square\square\square\square\square\square\square\square\square\square\square\square \cdots \\
 s_2 &= \cdots \square\square\square\square\square\square\square\square\square\square\square\square \cdots \\
 s_3 &= \cdots \square\square\square\square\square\square\square\square\square\square\square\square \cdots \\
 s_4 &= \cdots \square\square\square\square\square\square\square\square\square\square\square\square \cdots \\
 &\quad \dots
 \end{aligned}$$

Si noti che la generalizzazione di questo schema a più dimensioni richiede una certa accortezza nella definizione della relazione di adiacenza. Se, infatti, in una sola dimensione il numero di alternative è piuttosto contenuto, già il passaggio a due dimensioni introduce margini di scelta considerevoli. Le alternative più comuni sono l'adiacenza di von Neumann (*von Neumann neighborhood*) e quella di Moore (*Moore neighborhood*). Queste due relazioni di adiacenza si differenziano per il fatto di escludere (nel primo caso) o comprendere (nel secondo) anche le celle in posizione "diagonale". Se dunque abbiamo uno spazio bidimensionale definito come:

$$S = \{ \langle x, y \rangle \mid x, y \in \mathbb{Z} \}$$

Avremo le due possibili relazioni di adiacenza così definite:

$$A(x, y) = \left\langle \begin{array}{ccc} & \langle x, y-1 \rangle, & \\ \langle x-1, y \rangle, & \langle x, y \rangle, & \langle x+1, y \rangle, \\ & \langle x, y+1 \rangle & \end{array} \right\rangle$$

Adiacenza di von Neumann

$$A(x, y) = \left\langle \begin{array}{ccc} \langle x-1, y-1 \rangle, & \langle x, y-1 \rangle, & \langle x+1, y-1 \rangle, \\ \langle x-1, y \rangle, & \langle x, y \rangle, & \langle x+1, y \rangle, \\ \langle x-1, y+1 \rangle, & \langle x, y+1 \rangle, & \langle x+1, y+1 \rangle \end{array} \right\rangle$$

Adiacenza di Moore

Utilizzando l'adiacenza di von Neumann ogni cella avrà dunque un insieme di 5 celle adiacenti, nel caso dell'adiacenza di Moore saranno invece 9.

Su questa base di partenza sono state elaborate innumerevoli variazioni, spesso in risposta ad esigenze specifiche sorte dal tentativo di applicare gli automi cellulari alla simulazione di particolari sistemi. Oltre ad aumentare le dimensioni dello spazio è ad esempio possibile considerare celle con forme diverse (ad esempio celle triangolari o esagonali comportano relazioni di adiacenza diverse).

L'aggiornamento degli stati può avvenire in maniera asincrona anziché in parallelo. In questi modelli solo una cella per volta (o comunque un numero limitato di esse) aggiorna il proprio valore nel passaggio dal tempo t al tempo $t + 1$, l'ordine di aggiornamento delle celle diventa dunque rilevante per la dinamica complessiva del sistema. Questo tipo di automi cellulari prevede spesso anche elementi stocastici: tipicamente, ad esempio, le celle che devono aggiornare il loro valore vengono di volta in volta selezionate in maniera casuale. Alcuni esempi di automi cellulari non tradizionali, con caratteristiche derivate dai sistemi che ambiscono a simulare, sono richiamati più avanti, nella sezione 1.2.

Capitolo 1

Le origini degli automi cellulari

Gli automi cellulari sono sistemi dinamici astratti che, nella matematica discreta, giocano un ruolo paragonabile a quello delle equazioni alle derivate parziali nella matematica del continuo. Sia in termini di struttura che di applicazione, sono per l'informatico la controparte del concetto di 'campo' – governato dalle 'equazioni di campo' – del fisico.

(Toffoli e Margolus 1990)¹

Nessuno dei paradigmi tradizionali, tuttavia, è in grado di gestire adeguatamente il comportamento collettivo di *sistemi complessi* i quali – pur presentando un elevato, spesso infinito, numero di gradi di libertà – possono comportarsi in modo tale che i teoremi del limite centrale non si applichino. In contrasto con la nostra comprensione dei movimenti irregolari nei sistemi continui, una trattazione analitica generale di sistemi complessi discreti – costituiti da un gran numero di componenti semplici che interagiscono localmente in modo non lineare – non è oggi disponibile.

(Ilachinski 2001)²

¹“Cellular automata are abstract dynamical systems that play a role in discrete mathematics comparable to that played by partial differential equations in the mathematics of the continuum. In terms of structure as well as applications, they are the computer scientist's counterpart to the physicist's concept of a 'field' governed by 'field equations'.”

²“Neither traditional paradigm, however, is capable of dealing adequately with the collective behavior of *complex systems*, which – although consisting of large, often infinite, number of degrees of freedom – can behave in ways such that central limit theorems do not apply. In contrast to our understanding of irregular motions in continuous systems, a comprehensive analytical description of discrete complex systems, constructed out of a large number of simple components undergoing nonlinear local interactions, is presently out of reach.”

1.1 I primi passi

Vari studiosi, in luoghi e tempi diversi, si sono affacciati allo studio di *sistemi dinamici discreti con interazioni locali*. Tra i precursori degli studi in questo ambito, per esempio, Toffoli e Margolus (1987) citano Konrad Zuse, mentre Ilachinski (2001) ricorda i matematici russi impegnati tra gli anni '30 e '40 allo sviluppo della *teoria dei sistemi*. Nessuna di queste esperienze è però riuscita a consolidarsi e a dotarsi di una terminologia tecnica comune al punto da costituire – in termini lakatosiani (Lakatos 1978) – un vero e proprio programma di ricerca duraturo.

Le condizioni intorno alle quali si sono sviluppati e diffusi quelli che oggi chiamiamo automi cellulari sono infatti piuttosto peculiari. Gli anni a cavallo tra la seconda guerra mondiale e il dopoguerra videro la nascita della cosiddetta *big science*³. In particolare negli Stati Uniti – e inizialmente con l'obiettivo di sostenere lo sforzo bellico – il governo avviò programmi di massiccio finanziamento della ricerca. È celebre il rapporto di Vannevar Bush (1945) che esorta il presidente Truman a intraprendere questa strada tramite l'istituzione di enti appositi.

Fu proprio in questo clima che John von Neumann, uno dei più celebri scienziati di quel periodo, formulò per la prima volta il concetto di automa cellulare. Il tentativo di fornire una biografia scientifica di von Neumann – anche volendo lasciare da parte il suo importante ruolo pubblico – sarebbe fuori tema e richiederebbe uno spazio eccessivo. La sua produzione scientifica è infatti amplissima e poliedrica. In questa sede dunque ci limiteremo a richiamare unicamente quegli aspetti che ebbero, più probabilmente, un'influenza rilevante per la nascita degli automi cellulari⁴.

Ilachinski (2001) riporta che “von Neumann [...] giunse agli automi cellulari tramite lo strano percorso di un interesse per la logica formale e i fondamenti della matematica”⁵. In effetti negli anni '20, von Neumann aderì al progetto di fondazione assiomatica della matematica iniziato da Hilbert – proponendo ad esempio una definizione degli ordinali che oggi porta il suo nome – ed era presente alla conferenza

³L'origine del termine è comunemente ricondotta a Weinberg (1961). Per gli scopi di questo lavoro è particolarmente notevole il fatto che sullo stesso terreno siano state gettate le basi per la nascita della scientometria come disciplina (Price 1963).

⁴Esistono innumerevoli biografie di von Neumann. Per questo lavoro si è scelto di far riferimento soprattutto alle fonti più dirette e vicine nel tempo, quali ad esempio l'articolo di commiato di Ulam (1958), la prefazione a *The Computer and the Brain* (von Neumann 1958) a cura della sig.ra von Neumann, e l'introduzione di *Theory Of Self Reproducing Automata* (von Neumann e Burks 1966) a cura di Burks. Per una bibliografia delle pubblicazioni di von Neumann si rimanda invece a Beebe (2019).

⁵“von Neumann [...] came to CA via the unlikely path of an interest in formal logic and the foundations of mathematics”

di Königsberg del 1930 in cui Gödel illustrò pubblicamente il suo teorema di incompletezza. Alla fine degli anni '30, a Princeton, ebbe inoltre modo di conoscere Turing e il suo lavoro sul problema della decisione. Sebbene il passaggio sintetizzato in una breve formula da Ilachinski sia stato tutt'altro che lineare e diretto, è ragionevole supporre che la familiarità di von Neumann con la logica, la matematica discreta e la teoria della dimostrazione abbia giocato un ruolo rilevante nel plasmare i futuri automi cellulari.

Un altro ingrediente fondamentale da considerare è quello della nascita di tecnologie applicate alla computazione, e soprattutto dei problemi sollevati dalla loro progettazione. Durante gli anni '40, von Neumann partecipò allo sviluppo dell'architettura dei primi computer. Inizialmente collaborò con il gruppo del progetto ENIAC – finanziato dall'esercito americano, diretto da Presper Eckert e John Mauchly, e di cui faceva parte anche Arthur Burks – per lo sviluppo di un computer in grado di effettuare rapidamente calcoli balistici. Successivamente, questo gruppo – di cui entrò a far parte anche von Neumann – diede vita al progetto EDVAC. Questa nuova macchina era superiore alla precedente sotto il profilo strettamente tecnico, ma soprattutto incorporava a più livelli le intuizioni di von Neumann circa la separazione tra logica di funzionamento e meccanismo di implementazione. Anzitutto, a partire dal modello computazionale di neurone formulato da McCulloch e Pitts (1943), von Neumann propose l'utilizzo di componenti logiche astratte, indipendenti dallo specifico circuito che ne realizza il comportamento. Come fa notare Burks “[...] l'uso di elementi computazionali idealizzati è un passo nella direzione di una teoria degli automi”⁶ (von Neumann e Burks 1966). In secondo luogo EDVAC rappresentò la nascita concreta della programmazione come oggi la conosciamo: con un linguaggio ad alto livello di astrazione scritto da esseri umani e poi implementato a livello più basso nel linguaggio macchina (*short code* e *complete code* nella terminologia di von Neumann), e con moduli di codice già pronti e conservati nella memoria della macchina in maniera analoga agli altri dati. Come si è detto in precedenza queste due macchine nascevano con l'obiettivo di una applicazione al calcolo balistico. Probabilmente von Neumann vi si avvicinò poiché fin dagli anni '30 si era dedicato a studi di dinamica dei fluidi e meteorologia, e i nuovi computer promettevano buoni risultati anche per l'approssimazione numerica di equazioni non lineari alle derivate parziali⁷.

Il nesso tra l'aspetto più applicativo dell'analisi numerica e le fondazioni teoriche

⁶ “[...] the use of idealized computing elements is a step in the direction of a theory of automata”

⁷In molti casi equazioni di questo tipo sono intrattabili dal punto di vista analitico, e anche oggi si ricorre alla soluzione numerica. Il caso più famoso è forse quello delle equazioni di Navier-Stokes per la dinamica dei fluidi nel loro caso più generale.

della matematica potrebbe non apparire evidente. Von Neumann era però fermamente convinto che la computazione non fosse semplicemente un modo per ottenere risultati, ma che dovesse svolgere un ruolo *euristico* per la matematica teorica e che, in qualche modo, rappresentasse per quest'ultima l'analogo dell'esperimento. Concretamente questo si traduce ad esempio nell'esplorazione numerica dello spazio dei parametri finalizzata alla comprensione del comportamento qualitativo di un sistema. Oggi questo modo di procedere è largamente diffuso in qualunque ambito di ricerca che abbia a che fare con dei sistemi complessi.

Più in generale, von Neumann teorizzava esplicitamente la necessità di uno stretto rapporto tra la matematica e i suoi ambiti di applicazione:

[...] a grande distanza dalla sua fonte empirica, o dopo molto *inbreeding* astratto, un tema matematico corre il rischio di degenerare. Solitamente all'inizio lo stile è classico; quando mostra i segni di diventare barocco, allora quello è il segnale di pericolo. [...] l'unico rimedio, ogni volta che si raggiunge questo stadio, mi sembra sia il ringiovanente ritorno alla fonte: la reiniezione di idee più o meno direttamente empiriche.

(von Neumann e Taub 1961) ⁸

Nel corso del suo lavoro sull'architettura dei computer, von Neumann si dedicò anche a uno studio delle analogie tra computer e cervello. È possibile trovare questo interesse espresso già a partire dall'Hixon Symposium del 1948 (von Neumann e Taub 1963, pp. 288-326) – a cui era presente anche il già citato Warren McCulloch – e dalle lezioni all'università dell'Illinois dell'anno successivo. In particolare evidenziò alcune differenze rilevanti: in primo luogo il cervello è provvisto di una tolleranza a errori e malfunzionamenti di gran lunga superiore a quella di sistemi artificiali, in secondo luogo i computer sono progettati sulla base di un paradigma di computazione sincrona (lavorano cioè al ritmo imposto da un'unità computazionale centralizzata), mentre il cervello funziona in maniera asincrona (con eventi e reazioni diffuse spazialmente e cronologicamente indipendenti). Sembra ragionevole sostenere che fu proprio la ricerca di meccanismi che permettessero la correzione degli errori (Shannon 1958) nelle macchine artificiali a portare von Neumann alla formulazione del problema in termini di sistemi auto-replicanti. Lo stesso Burks avanza questa tesi in forma ipotetica (von Neumann e Burks 1966, pp. 19-20). È dunque possibile che il

⁸ “[...] at a great distance from its empirical source, or after much “abstract” inbreeding, a mathematical subject is in danger of degeneration. At the inception the style is usually classical; when it shows signs of becoming baroque, then the danger signal is up. [...] whenever this stage is reached, the only remedy seems to me to be the rejuvenating return to the source: the reinjection of more or less directly empirical ideas.”

punto di partenza per lavoro sugli automi cellulari sia stato di carattere applicativo più che formale. Thatcher (1964) ad esempio afferma che:

von Neumann, in quanto fondatore di questo ambito di studi, non era interessato agli automi cellulari in quanto oggetti di studio matematici, stava invece cercando un modo praticabile per trattare nel dettaglio il problema di come costruire macchine in grado di auto-replicarsi [...] ⁹

Mentre Ilachinski (2001) sottolinea maggiormente l'obiettivo di modellizzazione di sistemi biologici:

Von Neumann era interessato a trovare un modello riduzionista dell'evoluzione biologica. Il suo ambizioso progetto era quello di astrarre un insieme di interazioni locali primitive necessarie per l'evoluzione di forme di organizzazione complesse essenziali per la vita. ¹⁰

In ogni caso, come si è visto in precedenza, nella produzione di von Neumann è difficile scindere l'aspetto puramente applicativo da quello teorico. Nel caso specifico degli automi cellulari è possibile osservare un percorso di astrazione graduale, analogo a quello che per i computer ha portato alla separazione dei circuiti logici dall'implementazione fisica. All'Hixon Symposium, ad esempio, von Neumann formulava il problema della replicazione di automi nei seguenti termini:

È possibile costruire un aggregato a partire da questi elementi in modo che, se posizionato in un serbatoio in cui galleggiano tutti questi elementi in gran numero, questo inizi a costruire altri aggregati, ognuno dei quali si rivelerà essere un altro automa identico all'originale?

(von Neumann e Taub 1963) ¹¹

L'argomento di von Neumann a favore di questa possibilità può essere riassunto come segue. Consideriamo:

- un elemento costruttore A in grado, data la descrizione di un automa, di

⁹“Von Neumann, as the originator of this area of study, was not interested in cellular automata as mathematical objects of study, but instead was attempting to find a manageable way of treating in detail the problem of how to make machines reproduce themselves [...]”

¹⁰“Von Neumann was interested in finding a reductionist model for biological evolution. His ambitious scheme was to abstract a set of primitive local interactions necessary for the evolution of complex forms of organization essential for life.”

¹¹“Can one build an aggregate out of such elements in such a manner that if it is put into a reservoir, in which there float all these elements in large numbers, it will then begin to construct other aggregates, each of which will at the end turn out to be another automaton exactly like the original one?”

riprodurlo utilizzando gli elementi costitutivi presenti nell'ambiente;

- un elemento copiativo B in grado, data la descrizione di un automa, di produrre un duplicato;
- un elemento di controllo C il quale, fornito della descrizione di un automa, la consegna ad A e B , e in seguito inserisca la copia prodotta da B nel nuovo automa prodotto da A .

Definiamo ora l'aggregato:

$$D = A + B + C$$

Se assumiamo di avere una descrizione I_D di questo aggregato e la forniamo a D , otteniamo l'automa E , il quale è in grado di replicarsi. Von Neumann prosegue facendo notare l'analogia con un sistema biologico astratto. La descrizione I_D , ad esempio, gioca all'incirca il ruolo di un gene¹². Sarà sufficiente introdurre in questo quadro un ulteriore elemento F che possa essere aggiunto alla descrizione – ottenendo così I_{D+F} – per ottenere un automa in grado di replicare sé stesso e al contempo di produrre una copia di F . Se consideriamo ora la possibilità che si verifichino errori nel processo di copia della descrizione abbiamo la distinzione tra mutazioni letali – che occorrono in D e impediscono la riproduzione ulteriore – e mutazioni non letali – che occorrono in F trasformandolo in F' e alterano il prodotto finale.

Durante una breve serie di lezioni all'Institute for Advanced Study nel giugno del 1948, von Neumann presentò una bozza di modello capace di auto-replicarsi¹³. Questo primo tentativo prevedeva la presenza di diversi organi con funzioni specifiche, ad esempio muscoli in grado di contrarsi per spostare altri elementi, oppure organi in grado di saldare o tagliare legami tra altri organi. Questo modello è stato in seguito denominato da Burks “modello cinematico”:

il modello iniziale di auto-riproduzione di von Neumann includeva problemi geometrico-cinematici di movimento, contatto, posizionamento, fusione e taglio, e ignorava aspetti propriamente meccanici e chimici come la forza e l'energia. Lo chiamerò *modello cinematico* di auto-riproduzione.

(von Neumann e Burks 1966, pp. 82)¹⁴

¹²Seppur con il limite che nei sistemi biologici i geni non costituiscono una descrizione *completa* del sistema e del processo per costituirlo.

¹³Una sintesi delle lezioni curata da Burks è stata pubblicata nella prima parte di *Theory Of Self Reproducing Automata*, ma si trova all'interno del corpo del testo della quinta lezione tenuta da von Neumann all'Università dell'Illinois nel dicembre del 1949. In particolare per quanto riguarda il modello cinematico si faccia riferimento a von Neumann e Burks (1966, pp. 81-82).

¹⁴“von Neumann's early model of self-reproduction deals with the geometrical-kinematic pro-

Il basso livello di astrazione di questo modello – testimoniato dall’esplicitazione dei dettagli ingegneristici necessari all’implementazione – è probabilmente all’origine del fatto che von Neumann non lo abbia in seguito ulteriormente sviluppato.

Questo modello è di grande interesse ma, dal momento che prevede complesse considerazioni sul movimento delle parti nello spazio reale euclideo sarebbe tremendamente difficile realizzarlo nel dettaglio, persino ignorando problemi di energia, rumore ambientale e simili.

(Shannon 1958)¹⁵

Fu il contributo di Stanislaw Ulam a segnare un cambio di rotta determinante. Ulam e von Neumann avevano già collaborato in precedenza – ad esempio allo sviluppo e all’implementazione del metodo Monte Carlo (Metropolis 1987) – e lavoravano entrambi ai laboratori di Los Alamos. Il loro contributo congiunto portò il passaggio a una specificazione puramente formale degli automi cellulari, alla loro collocazione in uno spazio discreto e alla loro definizione in termini di stati funzionali anch’essi discreti. La reciproca influenza tra i due è riportata da numerose fonti, e se ne trova testimonianza anche in alcune delle prime pubblicazioni sugli automi cellulari – si vedano ad esempio Ulam (1950) e Ulam (1958). È inoltre bene evidenziare un ulteriore elemento di affinità tra i due: se, da un lato, abbiamo visto come von Neumann fosse interessato al rapporto tra strutture biologiche e formali, dall’altro Ulam coltivò sempre un interesse per problemi matematici applicati alla biologia – che lo portò, a partire dal 1968, a ricoprire la posizione di professore di biomatematica – e in questo contesto fece uso degli automi cellulari (Ulam 1962).

Adottato questo punto di vista più astratto, von Neumann poté formulare un automa cellulare auto-replicante: si tratta di un sistema in due dimensioni, in cui ogni cella può assumere 29 possibili stati diversi. Dato il suo grado di complessità, che non lo rende certo maneggevole sotto il profilo applicativo né formale, questo sistema non trovò mai una diretta applicazione in studi successivi, e in seguito lo stesso risultato fu ottenuto più volte, via via con automi più semplici. Tuttavia sotto il profilo teorico questo episodio mantiene la sua importanza. Da un lato von Neumann ha dimostrato la possibilità di costruire formalmente automi auto-replicanti, dall’altro ha individuato un’analogia con la termodinamica: sotto una certa soglia la complessità e il grado di organizzazione di un sistema degenerano (ovvero l’entropia

blems of movement, contact, positioning, fusing, and cutting, and ignores the truly mechanical and chemical questions of force and energy. Hence I call it his *kinematic model* of self-reproduction.”

¹⁵ “This model is a very interesting one but, involving as it does complex considerations of motion of parts in a real Euclidean space, it would be tremendously difficult to carry out in detail, even if one ignored problems of energy, noise in the environment, and the like.”

del sistema aumenta), ma sopra a questa soglia possono crescere con il tempo.

Nel 1957 von Neumann morì lasciando incompiuti diversi manoscritti sugli automi cellulari; questi sarebbero poi stati raccolti – insieme ad altro materiale – e pubblicati da Burks alcuni anni dopo, sotto il titolo di *Theory Of Self Reproducing Automata*. Come von Neumann stesso ebbe a riconoscere nell’ultima parte della sua vita, l’ambizione di una teoria complessiva, in grado di rendere conto del comportamento tanto di automi artificiali quanto di sistemi biologici, non era stata raggiunta.

Sarebbe molto soddisfacente poter parlare di una ‘teoria’ di questi automi. Purtroppo ciò che esiste al momento [...] può essere descritto solo come come un ‘bagaglio di esperienze’ poco organizzate e scarsamente formalizzate.

(von Neumann 1958)¹⁶

Sebbene una teoria sistematica continuasse a mancare, von Neumann contribuì a sviluppare le condizioni che di lì in avanti avrebbero reso gli automi cellulari un campo di ricerca a sé stante in grado di attrarre studiosi da svariate discipline. Esisteva infatti una cornice teorica, in grado di rendere gli automi cellulari degli oggetti di studio riconoscibili e formalizzati, inoltre gli strumenti computazionali necessari a intraprendere tale studio erano stati sviluppati e nel tempo la loro disponibilità sarebbe andata espandendosi.

¹⁶“it would be very satisfactory if one could talk about a ‘theory’ of such automata. Regrettably, what at this moment exists [...] can as yet be described only as an imperfectly articulated and hardly formalized ‘body of experience’.”

1.2 Espansione e consolidamento della comunità di ricerca

In seguito alla morte di von Neumann – e alla pubblicazione postuma dei suoi lavori negli anni successivi – gli automi cellulari iniziarono a diffondersi in ambito accademico. Come però spesso accade ad ambiti di studio nascenti, in particolare se interdisciplinari, la mancanza di una terminologia tecnica e di metodi condivisi – oltre che l’assenza di una comunità di ricerca coesa in grado di curare riviste e organizzare conferenze – determinò un periodo di particolare confusione. Sistemi dinamici discreti analoghi agli automi cellulari furono studiati sotto nomi svariati come *cellular spaces*, *tessellation automata*, *homogeneous structures*, *cellular structures*, *tessellation structures* e *iterative arrays* (Wolfram 1983). Malgrado questa mancanza di comunicazione abbia comportato in più occasioni la duplicazione del lavoro (Toffoli e Margolus 1987), negli anni ’60 la ricerca sugli automi cellulari proseguì soprattutto sul versante teorico – complice la mancanza di strumenti computazionali che limitava il lavoro di ricerca a metodi per lo più analitici o di simulazione manuale.

Un esempio di risultato rilevante conseguito in questo periodo è quello del teorema *Garden of Eden*, il quale afferma che – a determinate condizioni – esistono necessariamente delle configurazioni di celle (dette appunto *Garden of Eden*) che non possono essere raggiunte tramite l’applicazione iterata della regola di aggiornamento scelta, ma possono comparire solo se arbitrariamente impostate nello stato iniziale (Moore 1962; Myhill 1963). Questo teorema è particolarmente interessante dal momento che rivela un’attenzione per il tema della reversibilità degli automi, tema che – come vedremo – sarà particolarmente importante nel corso degli anni ’80, nell’ambito del dibattito sulla modellizzazione delle leggi fisiche.

All’inizio degli anni ’70, si ebbe un’ulteriore diffusione degli automi cellulari – questa volta anche fuori dagli ambienti accademici – tramite il successo di *Game of Life*. Si tratta di un automa cellulare sviluppato dal matematico John Conway e reso famoso da Martin Gardner su *Scientific American*. Le regole di *Life*, ispirate a una descrizione molto astratta dell’andamento di una popolazione biologica, possono essere così descritte:

- le celle con meno di due celle vive (nello stato ■) adiacenti muoiono (vanno allo stato □), come per un effetto di spopolamento;
- le celle con più di tre celle vive adiacenti muoiono, come per sovrappopolazione;
- le celle con tre celle vive adiacenti saranno anch’esse vive al tempo successivo.

Dentro *Life* è possibile osservare strutture stabili muoversi, interagire e riprodursi. In seguito fu inoltre dimostrata la possibilità di costruire una macchina di Turing all'interno di *Life*, il che lo rende un sistema in grado di computazione universale (Berlekamp, Conway e Guy 1982; Rendell 2002).

Intorno a *Life* si creò una vera e propria comunità di appassionati che si sfidavano a trovare configurazioni via via più complesse e con i comportamenti più svariati (Gotts e Callahan 1998). Fu forse proprio questa varietà di strutture e comportamenti ottenuta con regole intuitive a determinarne la fama, tanto che *Life* divenne un esempio paradigmatico richiamato in più ambiti in riferimento alle proprietà emergenti.

La popolarità di *Game of Life* contribuì ad attirare l'attenzione di scienziati e profani, da un lato diffondendo ulteriormente gli automi cellulari, dall'altro contribuendo alla confusione intorno al loro studio. Alcuni autori non mancano di sottolineare quest'ultimo aspetto. Riguardo agli automi cellulari Toffoli e Margolus (1990) diranno ad esempio:

Sono stati usati e abusati da scienziati interdisciplinari quanto da imbranati interdisciplinari. Sono stati l'argomento o la scusa per infinite tesi di dottorato. Si è molto parlato e scritto di loro, ma fino a poco tempo fa nessuno ne aveva visti molti.

...

Il gioco "life" di Conway [...] per un po' ha goduto di una popolarità simile a quella di un culto e ha trasformato gli 'automi cellulari' in una parola familiare per una generazione di giovani scienziati.¹⁷

Più tardi anche Ilachinski (2001) riconoscerà il carattere perlomeno ambivalente del contributo di *Game of Life* alla causa degli automi cellulari:

Sebbene questo gioco abbia fatto molto per pubblicizzare gli automi cellulari, l'effetto sui fisici, i matematici e la comunità scientifica generale fu abbastanza variegato. Che vi sia o meno una correlazione causale, il numero di articoli di matematica pura riguardo agli automi cellulari calò drasticamente dopo la pubblicazione di *Life* di Conway nel 1970.¹⁸

¹⁷"They have been used and abused by interdisciplinary scientists as well as interdisciplinary bumbler. They have been the topic or the excuse for countless doctoral theses. They have been much talked and written about, but until recently no one had actually *seen* much of them. [...] John Conway's game of "life," [...] for a while enjoyed a popularity close to a cult and turned 'cellular automata' into a household word for a generation of young scientists."

¹⁸"Although this game did a great deal to publicize CA, the effect on the physics, mathematics and general scientific community was rather mixed. Whether causally related or not, the number of

Parallelamente a questo processo di diffusione – e forse anche grazie a esso – gli automi cellulari divennero un oggetto di studio riconoscibile e iniziò a emergere una terminologia condivisa. A questa trasformazione contribuì, probabilmente, anche la nuova e crescente disponibilità di strumenti computazionali. L’attenzione per l’aspetto di implementazione a livello di singolo bit della computazione – così presente in tutti gli scritti di von Neumann – permane ancora negli anni ’80, testimoniata dallo sviluppo di macchine *ad hoc* per la simulazione di automi cellulari (Toffoli 1984a; Toffoli e Margolus 1987), per poi farsi via via più evanescente, superata dagli avanzamenti tecnici nei processori dei computer *general-purpose*.

Un importante contributo nella costruzione di una comunità intorno agli automi cellulari fu quello di Wolfram (1983) che – in un celebre articolo intitolato “Statistical mechanics of cellular automata” – fornì da un lato una visione d’insieme tramite un’ampia revisione della letteratura, dall’altro una nuova prospettiva grazie allo studio sistematico del comportamento di semplici automi cellulari in una dimensione. Ilachinski (2001) definisce questo articolo una “pietra miliare”, mentre per Schiff (2007) “ha dato inizio al primo studio serio degli automi cellulari”.

Wolfram fornì anzitutto un sistema di numerazione per automi cellulari binari in una dimensione (spesso definiti *elementari*), che associa a ogni possibile regola di transizione locale un numero naturale compreso tra 0 e 255. Dopo aver posto in ordine lessicografico tutte le possibili combinazioni di stati di una cella e delle sue adiacenti, si pone per ognuna di queste lo stato della cella al tempo successivo come uguale alla cifra del numero della regola scritto in base due nella posizione corrispondente (vedi figura 1.1).

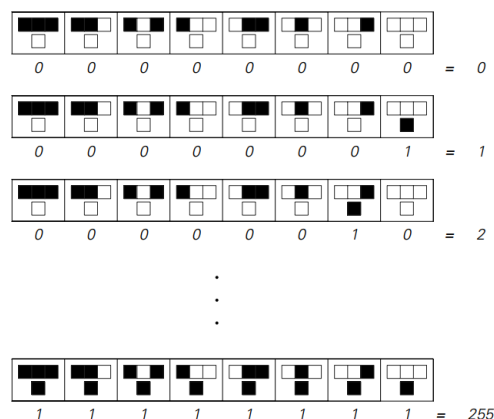
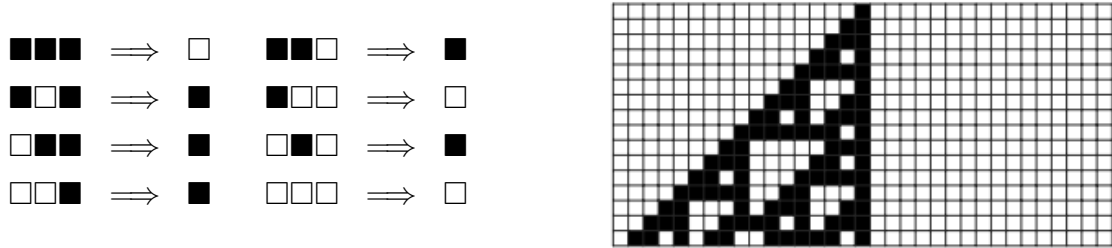


Figura 1.1 – Schema della numerazione di Wolfram (2002).

purely mathematical papers on CA dropped sharply after the publication of Conway’s *Life* game in 1970.”

Inoltre Wolfram compì uno studio sistematico dei 256 automi così ottenuti giungendo a dividerli in quattro classi sulla base della loro evoluzione nel tempo: uniforme, periodica, caotica o complessa. Di particolare interesse furono fin da subito gli automi appartenenti alla quarta classe, ovvero quelli in grado – come von Neumann aveva dimostrato – di operare fuori dall’equilibrio termodinamico e produrre un aumento della loro complessità strutturale. Di uno di questi infatti – quello corrispondente alla regola 110 (vedi figura 1.2) – è stata dimostrata la Turing-completezza.



(a) Rappresentazione nella notazione introdotta nella sezione 0.2.

(b) Esempio di evoluzione nel tempo da Wolfram (2002).

Figura 1.2 – La regola 110 di Wolfram.

La pubblicazione del decimo volume della rivista *Physica D* nel 1984 ebbe, se possibile, ancor più risonanza dell’articolo di Wolfram dell’anno precedente. Questo volume tematico raccolse infatti contributi da numerosi autori attivi nel campo e al suo interno è possibile trovare lavori rappresentativi di molte delle aree teoriche e applicative degli automi cellulari.

La categoria di articolo prevalente è – prevedibilmente – quella dei lavori teorici sugli automi cellulari in quanto tali. Tra gli altri (Gosper 1984; Grassberger 1984; Ingerson e Buvel 1984; Kauffman 1984; Lind 1984; Omohundro 1984; Waterman 1984; Willson 1984) troviamo anche un contributo di Wolfram (1984b), nel solco di quello del 1983, e uno di Langton (1984). Quest’ultimo avrebbe in seguito ideato un automa cellulare leggermente diverso da quelli tradizionali, divenuto poi noto come *formica di Langton* (Langton 1986). In questo sistema gli stati delle celle non sono aggiornati tutti simultaneamente, è invece una “formica” a spostarsi sul piano – in una direzione determinata dallo stato della cella in cui si trova – lasciando dietro di sé delle tracce (ovvero alterando lo stato delle celle). La versione più semplice della formica di Langton prevede una sola formica e due possibili stati delle celle, malgrado ciò il sistema presenta un comportamento piuttosto sorprendente: inizialmente il moto della formica è apparentemente caotico e l’area visitata può essere descritta come un processo diffusivo¹⁹, a un certo punto però la formica adotta un movimento

¹⁹In fisica si dicono processi diffusivi quei processi che sono descritti da equazioni differenziali stocastiche, uno degli esempi più noti è quello del moto browniano.

che la porta ad allontanarsi con una traiettoria lineare dal suo punto di partenza (vedi figura 1.3). Questo fenomeno è all'origine della popolarità di questo modello, dal momento che rappresenta un esempio di *transizione di fase* – ovvero di drastica variazione del comportamento qualitativo di un sistema.

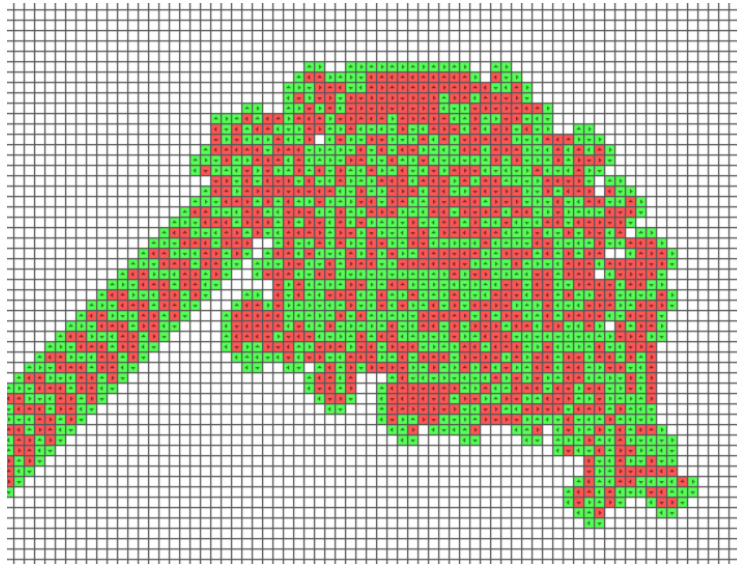


Figura 1.3 – Esempio di evoluzione nel tempo della *formica di Langton*. Le celle non visitate sono in bianco per evidenziare visivamente la dinamica.

Un altro ambito individuabile nel decimo volume di *Physica D* è quello dei modelli astratti di sistemi biologici a cui appartiene anche la formica di Langton (Burks e Farmer 1984; Smith, Watt e Hameroff 1984). I contributi di Toffoli (1984a), Carter (1984) e Hillis (1984) testimoniano invece il rapporto a doppio senso tra automi cellulari e implementazione di tecniche computazionali. In particolare Toffoli (1984a) documenta la progettazione di macchine dedicate specificamente all'esecuzione di automi cellulari, dall'altro lato invece Carter (1984) e Hillis (1984) rappresentano il tentativo di utilizzare gli avanzamenti teorici sugli automi cellulari per la progettazione di nuove architetture computazionali.

Parzialmente collegato a quest'ultimo aspetto vi è l'utilizzo degli automi cellulari come *trasduttori* (Toffoli e Margolus 1987), ovvero come dispositivi computazionali per il trattamento di informazioni. Applicazioni tipiche in questo senso sono il processamento di immagini e la codifica crittografica. Nel volume che stiamo considerando i contributi di Crutchfield (1984) e Preston (1984) sono rappresentativi di questo tipo di utilizzo. Inoltre Preston è tra gli autori di una corposa monografia dedicata a questo tema (Preston e Duff 1984).

L'ultimo filone individuabile nel volume è quello del dibattito sulla modellizzazione di fenomeni fisici tramite automi cellulari (Margolus 1984; Toffoli 1984b; Vichniac

1984). Si noti che in questo caso il tentativo non è quello di effettuare previsioni tramite un modello di approssimazione numerica, ma quello – ben più ambizioso – di trovare una qualche forma di corrispondenza diretta a livello microscopico tra sistemi fisici e automi cellulari. Questa aspirazione è ben sintetizzata nel seguente passaggio di Vichniac (1984):

[Gli automi cellulari] forniscono una terza alternativa alla dicotomia classica tra modelli che possono essere risolti in maniera esatta (per via analitica) ma sono molto stilizzati, e modelli più realistici che possono però essere risolti solo in maniera approssimata (per via numerica).

Infatti gli automi cellulari hanno abbastanza potere espressivo per rappresentare fenomeni di complessità arbitraria, e allo stesso tempo possono essere simulati in modo esatto da mezzi computazionali concreti: esecuzioni concrete effettuate da apparati di simulazione come CAM [si veda Toffoli (1984a)] sono un'implementazione esatta della matematica, e sono automaticamente libere dal rumore numerico poiché la discretezza è una componente essenziale del modello stesso.

In altre parole, siamo di fronte a una terza classe di modelli: quelli *calcolabili esattamente*. Per comprendere l'originalità di questo tipo di modello si deve tenere in considerazione che qui non vi è alcun tentativo di risolvere alcuna equazione data, infatti gli automi cellulari non svolgono alcun procedimento numerico, semplicemente compiono scelte logiche elementari, locali e spazialmente dipendenti. Questa terza classe di modelli dunque ambisce alla *simulazione* digitale *non-numerica* di fenomeni fisici.

(Vichniac 1984) ²⁰

Il tentativo di sviluppare automi cellulari in grado di svolgere simulazioni in questo senso – accompagnato talvolta dalla convinzione che a qualche livello l'universo *sia*

²⁰ “They provide a third alternative to the classical dichotomy between models that are solvable exactly (by analytical means) but are very stylized, and models that are more realistic but can be solved approximately (by numerical means) only. In fact cellular automata have enough expressive power to represent phenomena of arbitrary complexity, and at the same time they can be simulated exactly by concrete computational means: actual runs by a simulating device like CAM are an exact implementation of the mathematics, and they are automatically free of numerical noise since discreteness is a defining component of the model itself. In other words, we have here a third class: that of *exactly computable* models. To appreciate the originality of this type of modeling, one should keep in mind that there is no attempt here to solve any given equation, in fact cellular automata do not engage in any numerical processing, they merely perform simple spacedependent logical decisions. In other words, this third class aims at digital *non-numerical simulations* of physical phenomena.”

un automa cellulare (Zuse 1967; Schiff 2007; Wolfram 2002) – fornì la motivazione per numerosi studi su sistemi che conservassero invarianze e simmetrie fisiche. Una proprietà di particolare interesse era quella della reversibilità: già indagata sul versante computazionale per ragioni di efficienza energetica (Bennett 1973), fu tenuta in grande considerazione dal momento che è una delle proprietà salienti della fisica microscopica (Toffoli e Margolus 1987; Toffoli e Margolus 1990). L'intersezione tra gli aspetti fisici e la teoria dell'informazione nel contesto degli automi cellulari coinvolse anche l'ambito nascente della computazione quantistica e, a partire da una suggestione di Feynman (1982), vennero sviluppati anche automi cellulari basati su un paradigma computazionale quantistico (Deutsch 1985; Watrous 1995).

In anni più recenti il testimone di questa interpretazione forte degli automi cellulari è stato raccolto da Wolfram con una famosa – quanto controversa²¹ – monografia (Wolfram 2002) dal titolo evocativo: *A New Kind of Science*. Forse anche a causa di questo accento sugli aspetti di radicale discontinuità rispetto ai metodi tradizionali – riscontrabile fin dal contributo di Toffoli (1984b): “*Cellular automata as an alternative to (rather than an approximation of) differential equations in modeling physics*” – questa interpretazione forte non ha avuto una grande diffusione. Proprio Vichniac (1984), anzi, suggerisce una certa cautela sull'argomento, portando ad esempio l'impossibilità di una mappatura “ingenua” del modello di Ising²² in un automa cellulare.

Sebbene i tentativi di sistematizzare le leggi fisiche in termini di uno o più automi cellulari non abbia avuto grande seguito, la più pragmatica applicazione di questi ultimi a problemi svariati si è rivelata fruttuosa. Inoltre, a partire dagli automi cellulari, si è assistito a un'esplosione di nuovi sistemi e modelli computazionali con caratteristiche diverse. Fin dagli anni '80 i lavori sulla percolazione diretta (Domany e Kinzel 1984) hanno inaugurato l'applicazione degli automi cellulari – soprattutto nella loro variante probabilistica – a sistemi complessi come modelli di traffico

²¹La pubblicazione di *A New Kind of Science* ha suscitato un certo dibattito nella comunità scientifica. Numerose recensioni dell'opera hanno smentito alcune delle tesi sostenute da Wolfram (Georgiadis 2007) e sollevato critiche al suo stile, rimproverandogli uno scarso utilizzo dei riferimenti alle fonti. A scapito della popolarità dell'opera ha giocato anche la causa intentata da Wolfram al suo ex-collaboratore Matthew Cook, autore della dimostrazione di Turing-completezza della regola 110. A titolo di esempio si veda Giles (2002).

²²Il modello di Ising è un modello discreto che rappresenta il comportamento di un materiale ferromagnetico. È uno dei più noti modelli nell'ambito della meccanica statistica dal momento che è uno dei più semplici sistemi in cui sia possibile osservare delle transizioni di fase notevoli. Nel nostro caso Vichniac (1984) fa notare come la natura discreta del modello, che porterebbe a pensare ad una semplice trasposizione uno ad uno in automa cellulare, configga con la natura parallela della computazione operata dagli automi cellulari tradizionali.

urbano o epidemiologici. A cavallo tra la modellizzazione di sistemi biologici e problemi di ottimizzazione, gli automi cellulari sono stati usati per l'implementazione di algoritmi evolutivi (Alba e Dorronsoro 2008). Infine, negli ultimi anni, una loro filiazione particolarmente elaborata – gli *Agent-Based Models* (Epstein e Axtell 1996) – si è diffusa anche in ambiti disciplinari storicamente meno avvezzi a strumenti di simulazione numerica, come l'economia, la sociologia e talvolta persino la filosofia (Weisberg e Muldoon 2009).

Capitolo 2

Automi cellulari: uno studio quantitativo

[...] le discipline costituiscono il dominio di insegnamento della scienza, mentre unità intellettuali più piccole (annidate all'interno e a cavallo delle discipline) formano il dominio di ricerca. Nella sociologia della scienza queste unità sono state battezzate “*scientific specialties*”.

(Chubin 1976)¹

Perché non dovremmo applicare gli strumenti della scienza alla scienza stessa? Perché non misurare e generalizzare, fare ipotesi e derivare conclusioni?

(Price 1963)²

2.1 Bibliometria: una breve panoramica

La bibliometria – e in particolare la scientometria, ovvero la sua branca applicata al campo delle pubblicazioni scientifiche – consiste in un insieme di tecniche e metodi per lo studio quantitativo di documenti. Probabilmente il suo utilizzo più noto è quello a scopo valutativo – anche a causa della diffusione di esercizi di valutazione della ricerca su scala nazionale – che mira a fornire metriche e *ranking* nel tentativo di misurare la qualità dei prodotti della ricerca. Nel presente lavoro considereremo invece l'utilizzo descrittivo delle tecniche bibliometriche, il quale può fornire interes-

¹ “[...] disciplines form the teaching domain of science, while smaller intellectual units (nestled within and between disciplines) comprise the research domain. Within the sociology of science, these units have been termed “scientific specialties.””

² “Why should we not turn the tools of science on science itself? Why not measure and generalize, make hypotheses, and derive conclusions?”

santi elementi per l'analisi storica e sociologica di una comunità di ricerca (Petrovich e Tolusso 2019).

Sebbene con diversi gradi e sfumature, una delle assunzioni che hanno accompagnato la bibliometria fin dalle sue origini, con il lavoro di Eugene Garfield (1955), è che le citazioni siano una sorta di *moneta corrente* per la comunità scientifica. Che sia utilizzata per riconoscere un debito intellettuale, per orientare il lettore, come parte di una strategia di pubblicazione o per criticare un altro lavoro, la citazione rappresenta in ogni caso l'atto *formale* di creazione di un legame tra due documenti (Bornmann e Daniel 2008; Martyn 1975).

Questo carattere di formalità della citazione – garantito da un lato dalle norme sociali che regolano il funzionamento della comunità scientifica (Merton 1968) e dall'altro dal processo editoriale – ci restituisce un'immagine codificata e cristallizzata della dinamica di ricerca. Questo comporta, sì, la perdita di informazioni interessanti per uno studio storico³ e un sistematico ritardo cronologico⁴, ma fornisce anche la struttura necessaria per la trattazione computazionale di grandi moli di informazioni.

Naturalmente un approccio che non tenga conto della funzione specifica e dell'importanza relativa delle singole citazioni va incontro a numerosi problemi (Martyn 1964; Martyn 1975; Edge 1979; MacRoberts e MacRoberts 1989). Sebbene esistano metodi per tentare di mitigarli – ad esempio l'utilizzo di tecniche di analisi testuale per caratterizzare il contesto delle citazioni – questi limiti determinano una difficoltà sostanziale per quanto riguarda l'analisi di singoli documenti. Per questo da uno studio bibliometrico potremo aspettarci non tanto di ricevere una nuova chiave di lettura di singoli testi, quanto piuttosto di ottenere la visione d'insieme dello stato – e dell'evoluzione – di un *corpus* corrispondente alla produzione scientifica di una determinata comunità di ricerca.

Prima di proseguire con una presentazione delle tecniche che possono essere utilizzate per lo studio quantitativo di un ambito di ricerca è bene soffermarsi a definire precisamente i concetti che utilizzeremo.

Gli oggetti testuali che analizzeremo saranno prevalentemente *articoli scientifici* – provvisti dunque di uno o più *autori* – pubblicati su *riviste* e, come si è detto in

³Si tralasciano, ad esempio, tutte le informazioni diffuse tramite comunicazione informale tra ricercatori che sono a lungo state oggetto di studio da parte della sociologia della scienza degli anni '70 (Edge 1979).

⁴Una causa evidente di questo ritardo sono i tempi necessari a revisione e pubblicazione degli articoli su rivista, che possono essere lunghi e spesso variano considerevolmente a seconda di riviste, editori e fattori contingenti.

precedenza, le loro relazioni reciproche saranno stabilite sulla base delle citazioni. Nella lingua italiana il termine “citazione” è però molto versatile e si presta a utilizzi svariati. Nel seguito di questo lavoro impiegheremo dunque una terminologia che, per quanto possibile, non dia adito ad ambiguità.

Dato un testo A chiameremo:

- **riferimento bibliografico** (*cited reference*) un riferimento da A a un secondo testo B;
- **citazione** (*citation*) un riferimento da un secondo testo B in direzione di A.

Rimane fuori da questo quadro la citazione testuale (*quotation*) che, non essendo formalizzata al pari della bibliografia degli articoli scientifici, non è frequentemente oggetto di studio quantitativo.

La prima questione da sciogliere nell’accingersi a uno studio scientometrico è quella della selezione del *corpus*. Oltre a individuare un *database* sufficientemente fornito per i propri scopi – spesso vengono utilizzati i dati messi a disposizione da *Web of Science*, *Scopus* e *JSTOR* – è necessario selezionare un insieme di dati che sia allo stesso tempo rappresentativo dell’ambito oggetto di indagine e trattabile dal punto di vista delle dimensioni.

Una volta definito il *corpus*, ottenuti i dati relativi e presi gli accorgimenti necessari alla loro pulizia⁵, è necessario scegliere il tipo di analisi che si intende operare. Ad un livello generale l’analisi scientometrica consiste, nella maggior parte dei casi, nella costruzione e nello studio quantitativo di un grafo sulla base di alcune relazioni citazionali. Questa definizione generale lascia però ampi margini di scelta nella sua implementazione, ad esempio per quanto riguarda la relazione citazionale su cui definire il grafo.

La scelta apparentemente più semplice sarebbe quella di prendere come relazione di base la citazione pura e semplice. Tuttavia questa alternativa – sebbene sia stata storicamente la prima a comparire (Garfield 1955) – è perseguita di rado per due motivi principali: dal punto di vista delle risorse computazionali e di memoria richieste è molto onerosa, in secondo luogo l’accessibilità dei dati rappresenta un problema. Come vedremo in seguito, infatti, le politiche dei principali gestori di *database* citazionali prevedono limitazioni alla quantità e alla tipologia di informazioni

⁵Nell’analisi quantitativa ci si riferisce alla “pulizia” (*data cleansing*) dei dati nel senso di un passaggio preliminare volto a standardizzare il materiale oggetto di analisi, ad esempio scartando elementi mancanti di informazioni vitali. Nel nostro caso un passaggio elementare è, ad esempio, quello dell’eliminazione di articoli privi di una data di pubblicazione.

disponibili.

Una prima alternativa possibile è quella del *bibliographic coupling*, introdotto da Kessler (1963). Questa tecnica si basa sull'idea che il numero di riferimenti bibliografici in comune tra due articoli sia un buon indicatore della loro somiglianza, o che comunque possa essere usato come misura dell'intensità del loro legame. Di conseguenza il *bibliographic coupling* ci fornisce una definizione *statica* e *retrospettiva* del grafo citazionale; infatti – dal momento che i riferimenti bibliografici di un articolo non variano nel tempo – i legami tra i testi vengono specificati completamente al momento della loro pubblicazione. Questo implica anche che la lista dei riferimenti bibliografici degli articoli A e B è necessaria e sufficiente a calcolare il loro *coupling* bibliografico, e nessuna informazione aggiuntiva potrà fare la differenza.

Un'alternativa in grado di tener conto dell'evoluzione nel tempo della percezione di un prodotto della ricerca è rappresentata dalla *co-citation analysis*, introdotta da Small (1973). Secondo questa tecnica l'intensità del legame tra due articoli è data dalla frequenza con cui questi compaiono assieme tra i riferimenti bibliografici di articoli terzi. Risulta evidente che in questo caso i legami tra i nodi hanno una dipendenza temporale; la peculiarità di questa tecnica risiede nel fatto che la misura dei legami – proprio perché determinata da articoli altri rispetto a quelli da mettere in relazione – è strettamente dipendente dalla scelta del *corpus*.

La rappresentazione grafica in figura 2.1 può aiutare a chiarire le relazioni citazionali fin qui definite. Assumendo di aver compiuto una ricerca su un *database* citazionale abbiamo individuato tre *citing papers* – F, C ed E – e quattro *cited references* – A, B, G e D. L'articolo C cita tutte e quattro le *cited references*, mentre E cita G e D ed F cita solamente A. In figura le frecce in grigio rappresentano le citazioni dirette, mentre i legami in verde rappresentano il *bibliographic coupling* (che tra C ed E è più forte in quanto i due articoli condividono due riferimenti bibliografici) e quelli in rosso la *co-citation* (qui il legame tra G e D è più forte poiché entrambi sono citati sia da C che da E).

Esistono poi altre modalità composite per la costruzione del grafo, similmente basate sulla co-occorrenza di proprietà degli articoli. Dalla co-occorrenza di autori si ottiene ad esempio l'analisi della *co-authorship* (Beaver e Rosen 1978), mentre da quella di autori citati si ottiene la *author co-citation analysis* (White e Griffith 1981; White e Griffith 1982). L'uso di proprietà diverse dalla semplice citazione tra documenti richiede però spesso accorgimenti tecnici più raffinati, dal momento che i dati di riferimento non posseggono lo stesso grado di formalizzazione. Le pubblicazioni –

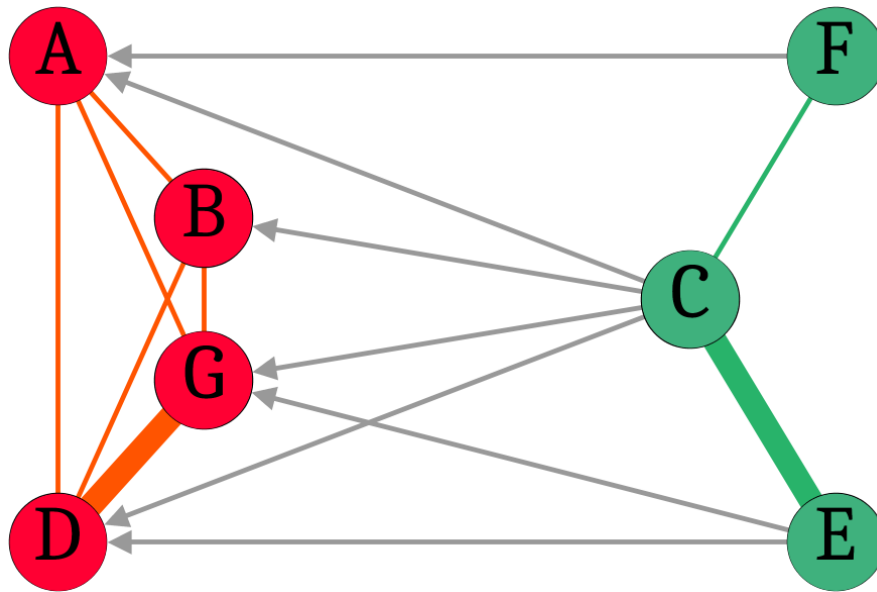


Figura 2.1 – Rappresentazione grafica di alcune relazioni citazionali. In grigio la citazione diretta, in verde il *bibliographic coupling* e in rosso la *co-citation*.

specie quelle recenti – sono dotate di codici identificativi unici⁶, mentre ad esempio la disambiguazione degli autori e la soluzione dei casi di omonimia sono problemi ancora aperti (D’Angelo, Giuffrida e Abramo 2010; Tang e Walsh 2010).

Dalla ricerca operata sui database di *Web of Science* si ottengono due diversi tipi di oggetto. Da un lato avremo gli articoli direttamente rispondenti ai criteri di ricerca definiti (che chiameremo *citing papers*), correlati di tutti i loro metadati, ove disponibili – autori con relative affiliazioni istituzionali, rivista di pubblicazione, ente finanziatore, ecc. Dall’altra avremo invece la lista dei riferimenti bibliografici contenuti nei *citing papers* (ovvero la lista delle *cited references*); questo secondo tipo di oggetti è significativamente meno informativo, dal momento che i dati qui compresi sono spesso parziali e rappresentati in forma abbreviata. Nelle tabelle 2.1 e 2.2 è possibile osservare la sproporzione tra le informazioni disponibili.

Questa distinzione è ben rappresentata in figura 2.1: i nodi in rosso sono le *cited references*, quelli in verde i *citing papers*. Come in figura, la struttura tipica dei dati è bipartita, le *cited references* sono significativamente più numerose dei *citing papers* e le uniche informazioni citazionali presenti vanno dall’insieme dei *citing papers* a quello delle *cited references*. In linea di principio sarebbe possibile reperire informazioni ulteriori sulle citazioni entro i due insiemi, e anche reperire i riferimenti bibliografici delle *cited references*. Questo tuttavia richiederebbe operazioni di ricerca

⁶Per gli articoli scientifici è comunemente utilizzato il *DOI* (*Digital object identifier*), le riviste scientifiche sono abitualmente provviste di un *ISSN* (*International Standard Serial Number*), mentre alle monografie è solitamente associato un *ISBN* (*International Standard Book Number*).

ID	781
AU	POLITI, A; LIVI, R; OPPO, GL; KAPRAL, R
C1	[; IST NAZL OTTICA,FLORENCE,ITALY.]; [...]
TI	UNPREDICTABLE BEHAVIOR IN STABLE SYSTEMS
PY	1993
SO	EUROPHYSICS LETTERS
AB	An unusual type of deterministic chaos found in chains of coupled maps is investigated. [...]
DT	Article
WC	['Physics, Multidisciplinary']
SC	['Physics']
KP	['ATTRACTORS']

Tabella 2.1 – Esempio di *citing paper* estratto dai database di *Web of Science*. Per ragioni di spazio è rappresentata solo una parte dei dati disponibili e alcuni campi sono troncati.

CR	Von Neumann J., 1966, THEORY SELF REPRODUC
----	--

Tabella 2.2 – Esempio di *cited reference* estratta dai database di *Web of Science*.

iterativa sui database, spesso non sono consentite dalle *policy* dei loro gestori.

Una volta definito il grafo – sulla base del *bibliographic coupling* se i nodi sono *citing papers*, o sulla *co-citation* se i nodi sono *cited references* – inizia il vero e proprio lavoro di analisi. Un primo passaggio utile per familiarizzare con il grafo è quello di effettuare delle misure standard; dalle più semplici (come il numero di nodi e legami) a quelle più articolate, come la distribuzione del grado dei nodi⁷ o la misura della loro centralità⁸. Tuttavia il cuore di un'analisi di questo tipo sta nel processo di *clustering*, ovvero nell'individuazione automatica di comunità di nodi a partire dalla struttura dei loro legami. Il *clustering* applicato al grafo del *bibliographic couplig* è, ad esempio, spesso utilizzato per l'individuazione dei cosiddetti *fronti di ricerca* (Price 1965). Sebbene esistano svariati algoritmi di *clustering* la loro idea di fondo è sempre quella di tentare di suddividere la popolazione dei nodi in comunità che

⁷Il grado di un nodo è il numero – eventualmente pesato – dei suoi legami. In molte dinamiche sociali la distribuzione del grado dei nodi segue una legge di potenza e presenta dunque code lunghe; per le pubblicazioni scientifiche questo fatto è stato rilevato da Lotka (1926). Per una recente revisione delle evidenze disponibili sulle leggi di potenza si rimanda a Newman (2005).

⁸Esistono in letteratura decine di misure di centralità diverse, ognuna adatta a una particolare applicazione. Una particolarmente intuitiva è la *betweenness centrality*, che assegna un valore di centralità a ogni nodo contando i cammini che lo attraversano (Freeman 1977).

abbiano al loro interno legami più forti rispetto a quelli che hanno con i nodi esterni.

Per l'interpretazione delle comunità così ottenute è però fondamentale il lavoro umano: la semplice relazione formale che lega i nodi appartenenti a una comunità, infatti, non ci fornisce immediatamente le sue ragioni. A questo livello l'analisi si sposta sulla composizione delle comunità e consiste principalmente nel dare una loro interpretazione nei termini di un principio unificante (Petrovich e Tolusso 2019). Un particolare *cluster* potrebbe infatti corrispondere a una specifica tematica di ricerca, a una o più riviste del settore, a uno specifico gruppo di ricerca, a una particolare istituzione ecc. Questo tipo di analisi richiede solitamente tante più competenze specialistiche nell'ambito studiato quanto più fine è il livello di dettaglio che si vuole ottenere. È però possibile facilitare lo sforzo interpretativo con l'impiego tecniche quantitative che aiutino ad evidenziare le caratteristiche salienti dei *cluster*.

Se poi, come nel nostro caso, l'obiettivo del lavoro bibliometrico è quello di considerare non solo la composizione delle comunità “fotografate” in un dato momento, ma anche il processo della loro formazione e diversificazione nel tempo, è necessario condurre l'analisi in più passaggi. Alcuni esempi di analisi diacronica possono essere rintracciati nella tecnica della *reference publication year spectroscopy* – solitamente abbreviato in *RPYS* – (Marx et al. 2013) o nei lavori di Small (Small e Greenlee 1989) e Morris (Morris et al. 2003; Morris e Boyack 2005). Tuttavia non esiste una procedura standard consolidata, e le varie tecniche possono risultare più o meno appropriate a seconda delle dinamiche che si intendono indagare.

2.2 Gli sviluppi recenti sotto la lente bibliometrica

Per il presente lavoro si è deciso di utilizzare come fonte il portale di *Web of Science*, una grande piattaforma che raccoglie dati citazionali le cui collezioni spaziano dal 1985 a oggi. Per individuare la letteratura di riferimento per gli automi cellulari si è deciso di utilizzare una ricerca per termini su titolo, *abstract* e parole chiave degli articoli – limitatamente al periodo 1985-1995 – secondo la seguente *query*:

“cellular automat*” OR “automata theory” OR “tessellation automat*” OR “tessellation structure” OR “iterative array”

Tabella 2.3 – *Query* della ricerca per termini.

Tra i termini di ricerca compaiono i diversi nomi che sono stati utilizzati per designare gli automi cellulari⁹, a eccezione di quelli che avrebbero rischiato di introdurre un numero eccessivo di falsi positivi (come ad esempio *homogeneous structure* o *cellular structure*). Per il periodo di riferimento individuato la ricerca ha restituito un *corpus* di quasi 2000 *citing papers* e più di 24000 *cited references*.

La pulizia dei dati è stata effettuata con un passaggio e successiva esportazione tramite *CRExplorer*¹⁰. Come primo passaggio di analisi si è deciso di costruire il grafo da studiare sulla base del *bibliographic coupling* secondo la procedura seguente.

- **Costruzione della matrice citazionale C**

Si è definita una matrice $p \times r$, con p uguale al numero di *citing papers* e r al numero di *citing references*. Dati due indici i, j tali che $1 \leq i \leq p$ e $1 \leq j \leq r$, gli elementi $C_{i,j}$ della matrice hanno valore uguale a 1 se il *paper* i cita la *reference* j , altrimenti hanno valore 0. Si noti che la matrice C è una delle possibili rappresentazioni di un grafo bipartito analogo a quello definito dalle frecce grigie in figura 2.1, la matrice C per l'esempio in figura sarebbe la seguente.

$$\begin{array}{c} \begin{array}{cccc} & A & B & G & D \\ \begin{array}{l} F \\ C \\ E \end{array} & \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix} \end{array}$$

⁹Per una rassegna della terminologia si rimanda a Wolfram (1983).

¹⁰Per esteso: *CitedReferencesExplorer*. Si tratta di uno strumento software per il tipo di analisi proposto da Marx et al. (2013). Maggiori informazioni sono reperibili sulla pagina dedicata al seguente indirizzo: <https://andreas-thor.github.io/cre/>.

- **Costruzione della matrice di adiacenza A**

La matrice di adiacenza basata sul *bibliographic coupling* è stata ottenuta dal prodotto di C con la sua trasposta: $A = CC^T$. Abbiamo dunque:

$$A_{x,y} = \sum_{j=1}^r C_{x,j} C_{j,y}^T \quad 1 \leq x \leq p, \quad 1 \leq y \leq p$$

Informalmente possiamo descrivere A come una matrice simmetrica $p \times p$. Ogni suo elemento $A_{x,y}$ contiene il numero di riferimenti bibliografici in comune tra l'articolo x e l'articolo y , ovvero il valore di *bibliographic coupling* tra i due articoli. La matrice A così ottenuta è una delle possibili rappresentazioni formali del grafo del *bibliographic coupling*. Nel caso dell'esempio riportato in figura 2.1 avremo dunque:

$$\begin{matrix} & F & C & E \\ \begin{matrix} F \\ C \\ E \end{matrix} & \begin{pmatrix} - & 1 & 0 \\ 1 & - & 2 \\ 0 & 2 & - \end{pmatrix} \end{matrix}$$

Si noti che i valori sulla diagonale della matrice non vengono considerati. Rappresenterebbero infatti il valore di *coupling* bibliografico di ogni articolo con sé stesso, ovvero il semplice numero dei suoi riferimenti bibliografici.

- **Riduzione della matrice A**

Solitamente prima di procedere all'analisi di un grafo di questo genere si eliminano gli elementi meno rilevanti, sia per alleggerire il carico computazionale che per ridurre il rumore da questi introdotto. In questo caso si è optato per la rimozione di tutti gli articoli con un valore di *coupling* complessivo inferiore a 10, ovvero quegli articoli n per cui vale la seguente condizione:

$$\sum_{x=1}^p A_{n,x} < 10$$

Dal punto di vista del grafo questa operazione può essere descritta come l'eliminazione dei nodi con grado pesato inferiore a 10.

Il grafo risultante da queste operazioni è composto da 1499 nodi e da più di 116000 legami tra essi. Ogni nodo ha in media 77.46 legami (*average degree*) e un *coupling* complessivo di 121.27 (*average weighted degree*).

Il passo successivo è stato quello di individuare le comunità di nodi. Per fare questo si è utilizzato il cosiddetto “metodo di Lovanio” (Blondel et al. 2008), un algoritmo di *clustering* che tenta – in maniera euristica – di massimizzare la modularità della

partizione del grafo. Informalmente si consideri la modularità come una misura (dal valore compreso tra $-1/2$ e 1) della “bontà” di una partizione – ovvero di una suddivisione del grafo in comunità. Valori di modularità elevati significano che le comunità hanno, al loro interno, legami più forti rispetto a quelli che hanno con i nodi esterni. La scelta di questo algoritmo in particolare è dettata dal fatto che, oltre ad essere computazionalmente più efficiente di molte delle alternative, il suo risultato dipende strettamente dalla struttura del grafo e non richiede di stabilire a priori il numero di comunità da individuare. Nel nostro caso la partizione risultante ha modularità pari a 0.47 e sono state individuate 11 diverse comunità di nodi, la taglia delle comunità è riportata nella tabella 2.4. Nel seguito escluderemo il *cluster* 10 dalla nostra analisi, in quanto composto da due soli articoli, degli stessi autori, e con legami deboli con il resto del grafo.

Cluster	0	1	2	3	4	5	6	7	8	9	10
Numero elementi	308	263	139	63	378	269	5	16	23	33	2

Tabella 2.4 – Numero di elementi presenti in ogni *cluster*.

Uno strumento utile per l’esplorazione del grafo nel dettaglio è quello della sua visualizzazione. In figura 2.2 è possibile osservare una rappresentazione del grafo considerato ottenuta con algoritmo per il posizionamento dei nodi denominato *Kamada-Kawai* (Kamada e Kawai 1989)¹¹. È bene sottolineare che la posizione dei nodi è, in questo caso, un semplice strumento per rendere agevole la visualizzazione, e non ha un valore direttamente interpretativo. Il colore dei nodi è invece determinato dal loro *cluster* di appartenenza, e la dimensione dal loro valore di *coupling* complessivo. Negli ultimi anni il tema della visualizzazione dei dati in ambito bibliometrico si è sviluppato notevolmente, per un approfondimento sul tema si rimanda a Vargas-Quesada e Moya-Anegón (2007).

Una volta individuate le comunità ha inizio il lavoro interpretativo. Quali sono le loro caratteristiche? Sono connotate dal punto di vista disciplinare? O si distinguono sulla base dei temi di ricerca trattati? Per rispondere a queste domande possiamo, come primo passo, considerare le etichette disciplinari attribuite alle riviste da *Web of Science*. Queste etichette sono assegnate con criteri archivistici che sono cambiati nel corso del tempo e sono finalizzate al recupero di informazioni, non sono dunque da considerarsi completamente affidabili per l’analisi della composizione disciplinare

¹¹L’algoritmo *Kamada-Kawai* fa parte della famiglia degli algoritmi di posizionamento *force-directed*. Questi algoritmi, che hanno la loro ispirazione in modelli fisici, trattano i nodi come corpi dotati di una massa, e li posizionano nello spazio secondo forze attrattive e repulsive definite dai loro legami.

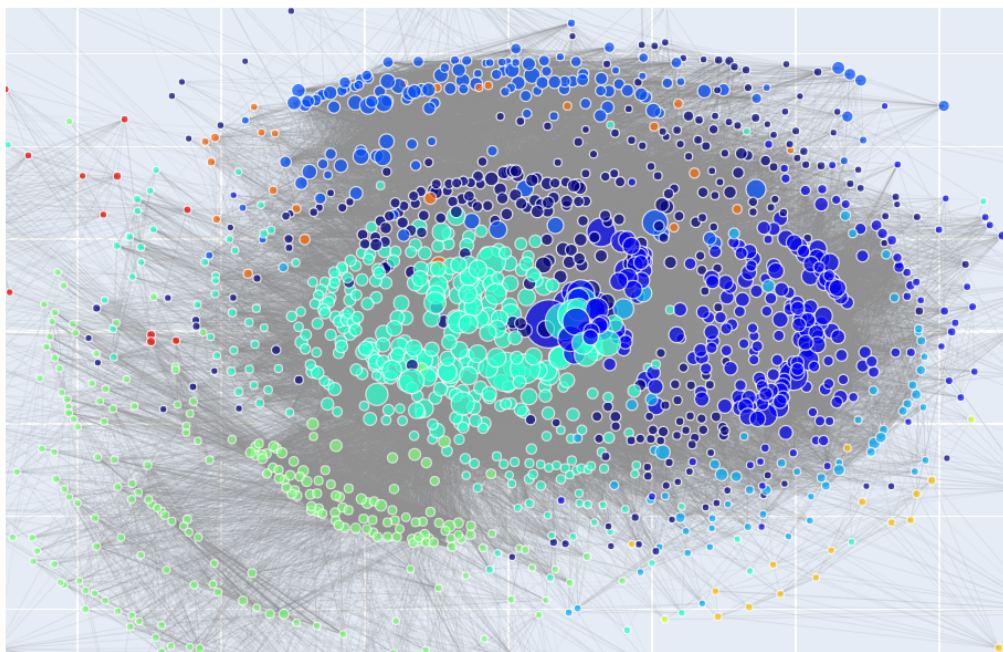


Figura 2.2 – Visualizzazione del grafo del *bibliographic coupling*.

dei *cluster* che stiamo considerando. Tuttavia possono fornire un buon punto di partenza per formulare delle ipotesi da verificare sulla base di altri dati.

In figura 2.3 è possibile osservare la somiglianza tra i *cluster* – calcolata con il metodo della similarità del coseno (*cosine similarity*)¹² – sulla base di due diversi attributi. In figura 2.3a la similarità è calcolata sulla base delle *Research Areas* (SC), mentre in figura 2.3b si sono considerate le *Categories* (WC). Entrambi questi attributi sono propri di riviste e monografie e servono alla loro catalogazione disciplinare. Il primo (SC) più generale, il secondo (WC) a grana più fine. Questo tipo di misura ci consente di osservare la prossimità disciplinare tra le comunità. Dalla figura 2.3 possiamo notare che: *i*) i *cluster* 0, 1, 2 e 8 (in misura minore anche il 3 ed il 4) presentano una forte affinità disciplinare, *ii*) i *cluster* 9 e 5 sono affini tra loro e piuttosto distanti dagli altri, *iii*) i *cluster* 6 e 7 sono piuttosto isolati e non presentano valori di similarità elevati con alcun altro *cluster*.

Individuate le similarità tra i *cluster*, è possibile osservare la loro composizione sulla base dei diversi attributi per evidenziare quali siano gli elementi che li accomunano. Come appare evidente in figura¹³ 2.6, la caratteristica che lega così strettamente i *cluster* 0, 1, 2, 4 e 8 è la presenza preponderante di articoli di ambito fisico.

¹²La similarità del coseno è una misura euristica ampiamente utilizzata nell'analisi quantitativa di testi ed in svariati altri ambiti. Il suo valore è calcolato come il coseno dell'angolo compreso tra due vettori in uno spazio n -dimensionale, dove n è il numero di attributi considerati.

¹³I grafici sulla composizione dei *cluster* sono collocati alla fine della sezione per ragioni di spazio.

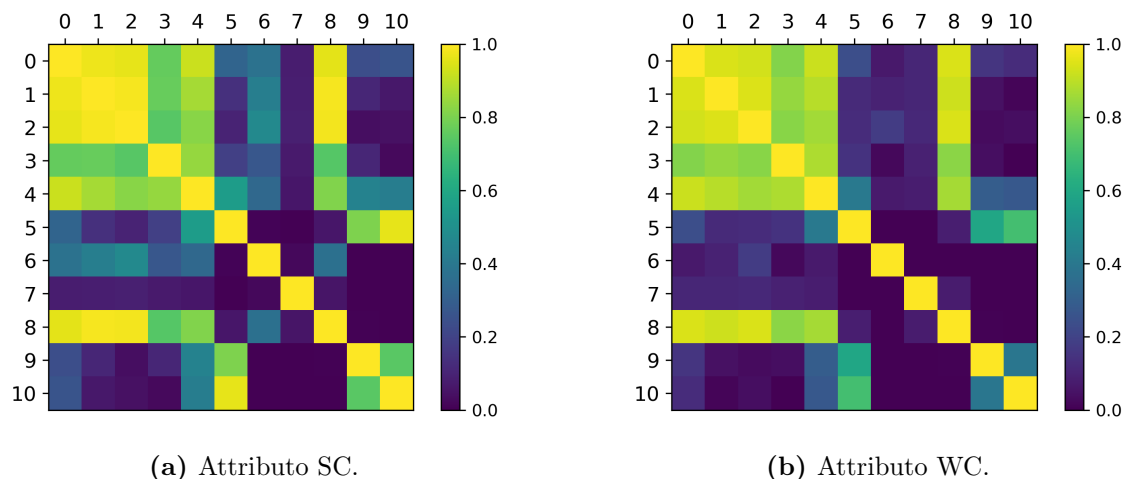


Figura 2.3 – *Cosine similarity* tra i *cluster*. Ogni casella riporta il valore di similarità (da 0 a 1) tra i due *cluster* riportati su riga e colonna.

Non è invece altrettanto chiaro l'elemento che fa di loro *cluster* distinti, in questo caso dovremo evidentemente indagare più a fondo per individuare le caratteristiche salienti di queste comunità.

La figura 2.8 permette di osservare la presenza preponderante dell'informatica nei *cluster* 5 e 9. I rimanenti *cluster* riguardano invece ambiti di applicazione particolari degli automi cellulari. La composizione disciplinare del *cluster* 3 mette in luce le applicazioni alla scienze biologiche, soprattutto nella loro branca computazionale. Il *cluster* 6 è invece composto unicamente da lavori legati alle applicazioni in ambito ottico, mentre il *cluster* 7 è formato da articoli nell'ambito della scienza dei materiali e della metallurgia.

Da questa prima indagine possiamo supporre che la maggior parte della produzione scientifica sugli automi cellulari del decennio considerato si collochi disciplinarmente negli ambiti della fisica e dell'informatica. Tuttavia, come si è detto, è necessario un criterio di verifica dell'attribuzione disciplinare degli articoli. In questo caso si è deciso di procedere analizzando le riviste di pubblicazione maggiormente rappresentative dei *cluster*. Il risultato, rappresentato in figura 2.9, conferma in questo caso le attribuzioni disciplinari delle etichette di *Web of Science*.

Dall'analisi delle etichette disciplinari e delle riviste di pubblicazione possiamo quindi iniziare a caratterizzare le comunità. La distinzione tra alcune di queste è evidente fin da questa prima analisi, in quanto presentano caratteristiche peculiari anche a questo livello di analisi superficiale. I *cluster* appartenenti alle discipline più rappresentate sono invece ancora indistinti a questo livello, sarà necessaria un'analisi più approfondita per caratterizzarli ulteriormente, distinguendoli – ad esempio – sulla

base degli specifici temi di ricerca trattati. Le attribuzioni disciplinari finora discusse possono essere ricapitolate nella tabella 2.5.

Cluster	0	1	2	3	4
Disciplina	Fisica	Fisica	Fisica	Biologia computazionale	Fisica

Cluster	5	6	7	8	9
Disciplina	Informatica	Ottica	Scienza dei materiali	Fisica	Informatica

Tabella 2.5 – Caratterizzazione disciplinare dei *cluster* dall’analisi dei parametri SC, WC e SO.

Per individuare i tratti distintivi dei *cluster* appartenenti alle medesime aree disciplinari si è deciso di operare dei test statistici sulla distribuzione delle parole nei titoli (TI) e negli *abstract* (AB) degli articoli. A questo fine si è considerata la concatenazione di tutti i titoli degli articoli di ogni *cluster* come un unico testo, si sono poi costituiti gli insiemi delle coppie ordinate date da ogni parola e dal suo numero di occorrenze. Lo stesso è stato fatto per gli *abstract*. Nel tentativo di individuare la presenza di terminologia tecnica specifica si è deciso di operare lo stesso tipo di analisi anche sui bigrammi¹⁴ presenti nei titoli (TI_bi).

Dato l’insieme P delle parole presenti, il numero dei *cluster* n , e la funzione *count* che ne ritorna il numero di occorrenze di una parola nel cluster considerato, avremo:

$$W_i^a = \langle \text{count}(w, i, a) \mid w \in P \rangle \quad a = TI, AB, TI_bi \quad 1 \leq i \leq n$$

La prima misura statistica utilizzata è la distanza dalla distribuzione media. Per calcolarla si è utilizzato il seguente procedimento.

- **Definizione della distribuzione media**

Si è definito un ulteriore insieme W_{tot} per ogni attributo, i valori dei suoi elementi sono definiti come il numero medio di occorrenze di ogni parola tra i *cluster*.

$$W_{mean}^a = \left\langle \frac{\sum_{j=1}^n W_j^a[w]}{n} \mid w \in P \right\rangle \quad a = TI, AB, TI_bi$$

Si noti che con l’espressione $W_i^a[w]$ si fa riferimento all’elemento di W_i^a corrispondente alla parola w .

¹⁴In linguistica computazionale i bigrammi sono coppie di parole che in un dato *corpus* compaiono adiacenti.

- **Normalizzazione**

La composizione degli insiemi W (compreso quello medio) è stata normalizzata, in modo i valori di ognuno sommassero ad 1. I nuovi insiemi normalizzati \overline{W} sono così definiti:

$$\overline{W}_i^a = \left\langle \frac{W_i^a[w]}{\sum_j W_i^a[j] \forall j \in P} \middle| w \in P \right\rangle \quad a = TI, AB, TI_bi \quad 1 \leq i \leq n$$

$$\overline{W}_{mean}^a = \left\langle \frac{W_{mean}^a[w]}{\sum_j W_{mean}^a[j] \forall j \in P} \middle| w \in P \right\rangle \quad a = TI, AB, TI_bi$$

Avremo dunque:

$$\sum_j \overline{W}_i^a[j] \forall j \in P = 1 \quad a = TI, AB, TI_bi \quad 1 \leq i \leq n$$

$$\sum_j \overline{W}_{mean}^a[j] \forall j \in P = 1 \quad a = TI, AB, TI_bi$$

- **Calcolo della distanza dalla media**

Per ogni elemento di ogni distribuzione media \overline{W}_i si è calcolata la differenza con l'elemento corrispondente della distribuzione media \overline{W}_{mean} .

$$D_i^a = \langle \overline{W}_i^a[w] - \overline{W}_{mean}^a[w] \mid w \in P \rangle \quad a = TI, AB, TI_bi \quad 1 \leq i \leq n$$

Una volta svolto questo passaggio abbiamo a disposizione, per ogni parola, la differenza tra la sua frequenza in un dato *cluster* e nel *corpus* complessivo. Maggiore questa differenza, maggiore è la peculiarità della parola considerata riguardo un determinato *cluster*.

La seconda misura statistica utilizzata è il test binomiale. Per ogni parola assumeremo la sua frequenza nel *corpus* complessivo come termine di paragone, e per ogni *cluster* calcoleremo la probabilità di ottenere una frequenza uguale a quella della parola scelta nel *cluster* considerato. Minore la probabilità ottenuta, maggiormente caratteristica del *cluster* sarà la parola considerata. Per l'esecuzione del test binomiale si è seguito il procedimento seguente.

- **Calcolo delle frequenze**

Si è calcolata la frequenza di apparizione di ogni parola nel *corpus* complessivo.

$$\overline{W}_{tot}^a = \left\langle \frac{\sum_{i=1}^n W_i^a[w]}{\sum_{i=1}^n \left(\sum_j W_i^a[j] \forall j \in P \right)} \middle| w \in P \right\rangle \quad a = TI, AB, TI_bi$$

La frequenza di apparizione di ogni parola è stata calcolata in precedenza, e corrisponde ai valori \overline{W}_i^a (si veda il passo *normalizzazione* nel calcolo della distanza dalla distribuzione media).

- **Esecuzione del test binomiale**

Nella sua forma generale il test binomiale è espresso come segue:

$$p(x|X, \pi) = \frac{X!}{x!(X-x)!} \pi^x (1-\pi)^{X-x}$$

Nel nostro caso π rappresenta la frequenza assunta a riferimento, x le occorrenze della parola considerata osservate nel *cluster* dato e X il numero complessivo di parole nel *cluster*. I nostri risultati saranno dunque:

$$B_i^a = \langle p(W_i^a[w] | X, \overline{W}_{tot}^a) | w \in P \rangle \quad a = TI, AB, TI_bi \quad 1 \leq i \leq n$$

Ottenuti questi dati, si sono anzitutto escluse le parole con meno di 10 occorrenze: queste infatti possono facilmente risultare rilevanti per i criteri utilizzati, ma sono scarsamente rappresentative delle comunità. In seguito si sono selezionate solo le 5 parole più significative secondo ciascuno dei due criteri. Per i *cluster* la cui caratterizzazione era finora ambigua, il risultato di questo ulteriore processo di selezione è riportato nelle tabelle alla fine del capitolo, dalla 2.7 alla 2.13.

Anzitutto possiamo osservare una certa coerenza tra i due test statistici utilizzati: nella maggior parte dei casi la stessa parola compare tra le più rilevanti per entrambi i criteri. I limiti di questo strumento stanno invece nella scarsa qualità dei risultati per *cluster* di taglia piccola (si veda ad esempio la tabella 2.12). Per le comunità di dimensione sufficiente abbiamo però a disposizione elementi interessanti, grazie ai quali la tabella 2.5 può essere aggiornata come segue.

Cluster	0	1	2	3	4
Disciplina	Fisica, sistemi complessi	Fisica, gas e fluidi	Fisica, <i>self-organized criticality</i>	Biologia computazionale	Fisica, teoria degli automi

Cluster	5	6	7	8	9
Disciplina	Informatica	Ottica	Scienza dei materiali	Fisica	Informatica, <i>learning automata</i>

Tabella 2.6 – Caratterizzazione disciplinare dei *cluster* aggiornata grazie all’analisi statistica delle parole più rilevanti nei campi TI e AB.

In seguito a questa analisi possiamo caratterizzare più nel dettaglio anche le comunità che finora erano indistinte dal punto di vista disciplinare. Il *cluster* 0 (tabella 2.7) è composto da articoli che si collocano nell’ambito dei sistemi complessi e spaziano dai modelli sui flussi di traffico urbano a quelli epidemiologici, pur comprendendo

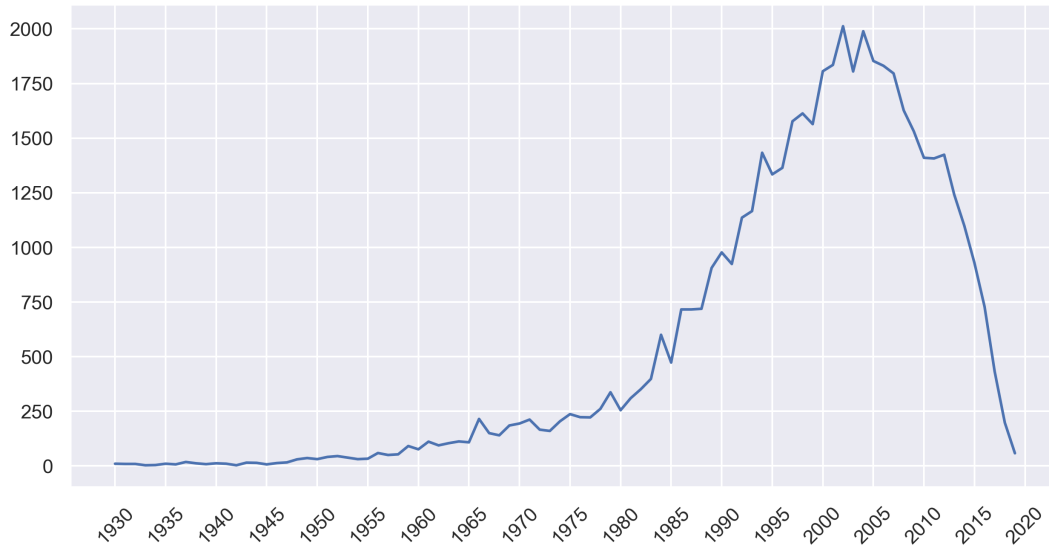
anche lavori più tradizionali sulle transizioni di fase in sistemi fisici. Il *cluster* 1 invece (tabella 2.8) rappresenta uno dei temi tradizionali di applicazione degli automi cellulari: lo studio della dinamica di gas e fluidi. Il *cluster* 2 (tabella 2.9) testimonia l'incontro tra gli automi cellulari e la teoria del caos. In particolare è composto da articoli riguardanti il fenomeno della *self organized-criticality* (Bak, Tang e Wiesenfeld 1987), il tratto tipico dei fenomeni fisici che presentano questa caratteristica è quello di avere il punto critico di una transizione di fase come attrattore della loro dinamica; di conseguenza questi sistemi evolvono spontaneamente verso una transizione di fase. Il *cluster* 4 (tabella 2.10) è invece composto prevalentemente da articoli riguardo allo studio degli automi cellulari in quanto tali e dei loro aspetti computazionali, un'indagine diretta del suo contenuto rivela la presenza di diversi lavori di autori discussi in precedenza come Wolfram e Toffoli. Come si è detto in precedenza la taglia ridotta del *cluster* 8 (tabella 2.12) non ha permesso di individuarne i termini salienti tramite analisi quantitativa, il *cluster* 9 invece (tabella 2.13) si differenzia dal 5 – che ha una composizione disciplinare di informatica senza una connotazione specifica in termini di argomenti di ricerca – per la presenza preponderante di articoli riguardo i *learning automata* (Narendra e Thathachar 1974): una delle prime forme di *machine learning* studiate fin dai primi anni '70.

Fino a questo punto la nostra analisi si è concentrata sul grafo complessivo costruito a partire da tutti gli articoli pubblicati tra il 1985 ed il 1995. Questo tipo di indagine ci ha fornito una fotografia degli ambiti di ricerca legati agli automi cellulari, ma come è possibile recuperare una descrizione dinamica della loro evoluzione?

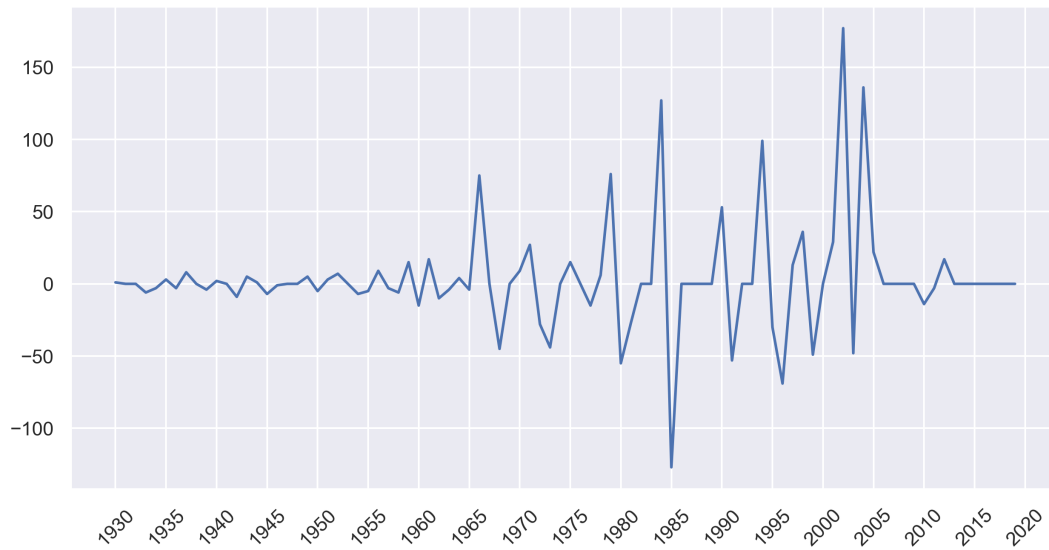
Una prima possibile risposta viene dalla già citata tecnica proposta da Marx et al. (2013), la *Reference Publication Year Spectroscopy* (RPYS). Questo metodo prevede il semplice conteggio delle citazioni ottenute da ognuna delle *cited references* presenti nel *corpus* selezionato e la loro aggregazione per anno di pubblicazione. Gli anni in cui si osservano picchi particolarmente consistenti nelle citazioni rappresentano, solitamente, momenti fondativi o di svolta per un ambito di ricerca, e spesso sono riconducibili alla pubblicazione di uno o più contributi particolarmente rilevanti. In figura 2.4 è possibile osservare il risultato della RPYS applicata al *corpus* finora considerato, esteso però al periodo 1985-2019. La rappresentazione della deviazione dalla mediana degli anni immediatamente precedenti (figura 2.4b) rende poi i picchi particolarmente evidenti.

Tramite questo strumento possiamo anzitutto confermare alcuni degli elementi precedentemente analizzati. Il primo picco consistente, nel 1966, è ad esempio da ricondurre alla pubblicazione di *Theory Of Self Reproducing Automata*. Quello del 1984 invece è dovuto all'uscita del decimo volume della rivista *Physica D*, mentre quello

del 2002 è un eco del dibattito sulla discussa monografia di Wolfram *A New Kind of Science*.



(a) Numero di citazioni per anno.



(b) Deviazione dalla mediana degli ultimi 5 anni.

Figura 2.4 – *Reference Publication Year Spectroscopy* (RPYS) degli articoli ottenuti da *Web of Science* tramite la *query* in tabella 2.3 nel periodo 1985-2019.

Oltre a pubblicazioni citate in precedenza, però, questo metodo rivela anche articoli e monografie influenti per lo sviluppo degli automi cellulari anche laddove questi ultimi non siano l'oggetto di studio principale. È ad esempio il caso del picco del 1979, riconducibile ad una monografia molto influente in ambito informatico: *Introduction to automata theory, languages, and computation* (Hopcroft 1979). Allo stesso modo, il picco del 1994 ci permette di individuare – oltre a un'ulteriore pubblicazione di Wolfram (1994) – un lavoro che propone l'applicazione di automi cellulari quantistici

alla costruzione di porte logiche (Tougaw e Lent 1994). Il vantaggio di questo tipo di analisi sta principalmente nella sua semplicità, che lo rende un ottimo strumento di indagine preliminare. Il suo limite è però quello di operare a livello aggregato e di non distinguere in maniera approfondita le relazioni tra gruppi di pubblicazioni correlate.

Per tentare di recuperare queste informazioni sulle relazioni di “parentela” tra le comunità che abbiamo analizzato, si è deciso di applicare iterativamente il processo di *clustering* descritto in precedenza. Inizialmente si sono selezionati i soli lavori pubblicati nel 1985. Poi, ad ogni iterazione, si sono aggiunti quelli dell’anno successivo. Formalmente, possiamo descrivere il risultato di questa operazione come una serie di insiemi C_y^i – uno per ogni *cluster* individuato – i cui indici fanno riferimento all’anno fino al quale si sono considerate le pubblicazioni (y) e al numero arbitrario con cui si individua il *cluster* (i). Date queste strutture, uno dei possibili modi di intendere la parentela che stiamo cercando di descrivere è quello dell’intersezione tra questi insiemi. In particolare la natura cronologica della relazione che stiamo studiando ci permette di considerare unicamente le coppie di insiemi collocati in anni consecutivi. Prenderemo quindi la catena di relazioni definita nel modo seguente:

$$C_y^i \cap C_{y+1}^{i'} \quad 1985 \leq y \leq 1994$$

Il risultato di questa operazione è visualizzabile intuitivamente tramite un diagramma di Sankey (talvolta chiamato anche *flowchart* o *flow diagram*) come quello in figura 2.5. I nodi colorati rappresentano i *cluster* e sono ordinati in colonne corrispondenti agli anni considerati (y). I valori dei flussi in grigio sono invece determinati dalla cardinalità dell’intersezione tra gli insiemi che collegano, ovvero:

$$|C_y^i \cap C_{y+1}^{i'}| \quad 1985 \leq y \leq 1994$$

Per caratterizzare anche visivamente le relazioni tra le comunità si sono attribuiti di colori ai nodi secondo il procedimento seguente.

- **Individuazione dei nodi di origine**

Si sono individuati i nodi di origine, ovvero quelli privi di flussi in ingresso. Formalmente questi sono i nodi per cui è valida la seguente condizione:

$$O(C_y^i) \iff \neg \exists C_{y-1}^{i'}, |C_y^i \cap C_{y-1}^{i'}| \neq \emptyset$$

Nel caso considerato questa condizione è banalmente verificata per tutti i *cluster* dell’anno 1985, tuttavia anche nei periodi successivi emergono occasionalmente *cluster* composti unicamente da articoli di nuova pubblicazione. A tutti i nodi di origine individuati viene poi assegnato un colore arbitrario.

• Colorazione dei nodi restanti

Tutti i nodi restanti ereditano, almeno in parte, gli articoli di cui sono composti da nodi cronologicamente precedenti. Si definirà quindi il loro colore come la media dei colori dei nodi da cui ricevono articoli, pesata per il loro numero.

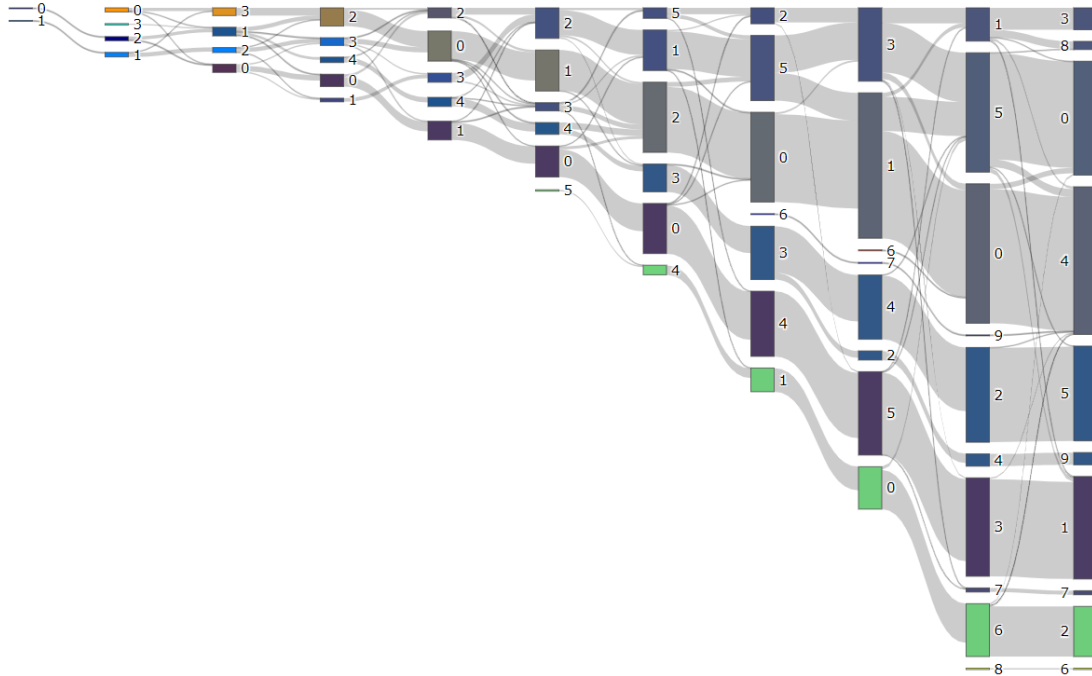


Figura 2.5 – Visualizzazione delle relazioni tra i *cluster* tramite diagramma di Sankey

L’attribuzione dei colori eseguita secondo questo procedimento, unita al tipo di visualizzazione utilizzata, rende particolarmente intuitive le relazioni tra i *cluster* individuati e permette di individuarne le “parentele”. Nel caso in esame la figura 2.5 permette di osservare una certa continuità cronologica, soprattutto per quanto riguarda i *cluster* di taglia considerevole. Con il passare del tempo vediamo infatti che i flussi incrociati tra *cluster* rappresentano una quota del totale via via decrescente, mentre si creano delle strutture stabili ed omogenee che assorbono una frazione crescente dei flussi di articoli. Dal punto di vista interpretativo, questo fenomeno suggerisce che, fin dalla metà degli anni ’80, gli ambiti di studio ed applicazione degli automi cellulari fossero già distinti in maniera piuttosto marcata. L’andamento generale descritto non è però esente da eccezioni. Una particolarmente rilevante è quella del *cluster* C_{1992}^5 , il quale presenta due flussi in uscita consistenti e di taglia paragonabile. Abbiamo infatti:

$$|C_{1992}^5 \cap C_{1993}^3| = 82 \quad |C_{1992}^5 \cap C_{1993}^1| = 61$$

Allargando lo sguardo all’anno ancora successivo osserviamo che i due flussi, separati nel 1993, si riuniscono nel 1994:

$$|C_{1992}^5 \cap C_{1993}^3 \cap C_{1994}^5| = 68 \quad |C_{1992}^5 \cap C_{1993}^1 \cap C_{1994}^5| = 51$$

Questo fenomeno sembra suggerire una certa vicinanza tra quelli che saranno i *cluster* C_{1995}^0 e C_{1995}^4 . Vicinanza che, alla luce dell'analisi disciplinare illustrata in precedenza, appare pienamente giustificata dallo stretto rapporto tra la tematiche trattate nei due *cluster*: le applicazioni degli automi cellulari allo studio di sistemi complessi da un lato, la teoria degli automi in quanto tali dall'altro.

Possiamo ora ricapitolare brevemente l'analisi quantitativa svolta. Per prima cosa si è proceduto alla generazione del grafo da analizzare sulla base della relazione di *bibliographic coupling*. In seguito si è utilizzato un algoritmo di *clustering* per individuare le comunità dei nodi più strettamente legati tra loro. L'analisi delle etichette assegnate da *Web of Science* è stata utilizzata per fornire una prima caratterizzazione disciplinare dei *cluster*, poi confermata da un'indagine sulle riviste di pubblicazione. Laddove la caratterizzazione disciplinare era insufficiente a porre un discrimine tra i *cluster* si è optato per l'utilizzo di test statistici sulle distribuzioni delle parole nei titoli e negli *abstract* degli articoli. Questa ulteriore analisi ha permesso di individuare criteri di distinzione tra i *cluster* ad un livello sub-disciplinare, evidenziando temi di ricerca specifici. Si è infine tentato di recuperare la dimensione temporale del processo di ricerca, dapprima utilizzando la tecnica della RPYS. In secondo luogo si è proposta una nuova combinazione di tecniche esistenti, basata sull'applicazione iterativa dell'algoritmo di *clustering* e sulla visualizzazione delle intersezioni tra i *cluster* di anni successivi tramite diagramma di Sankey.

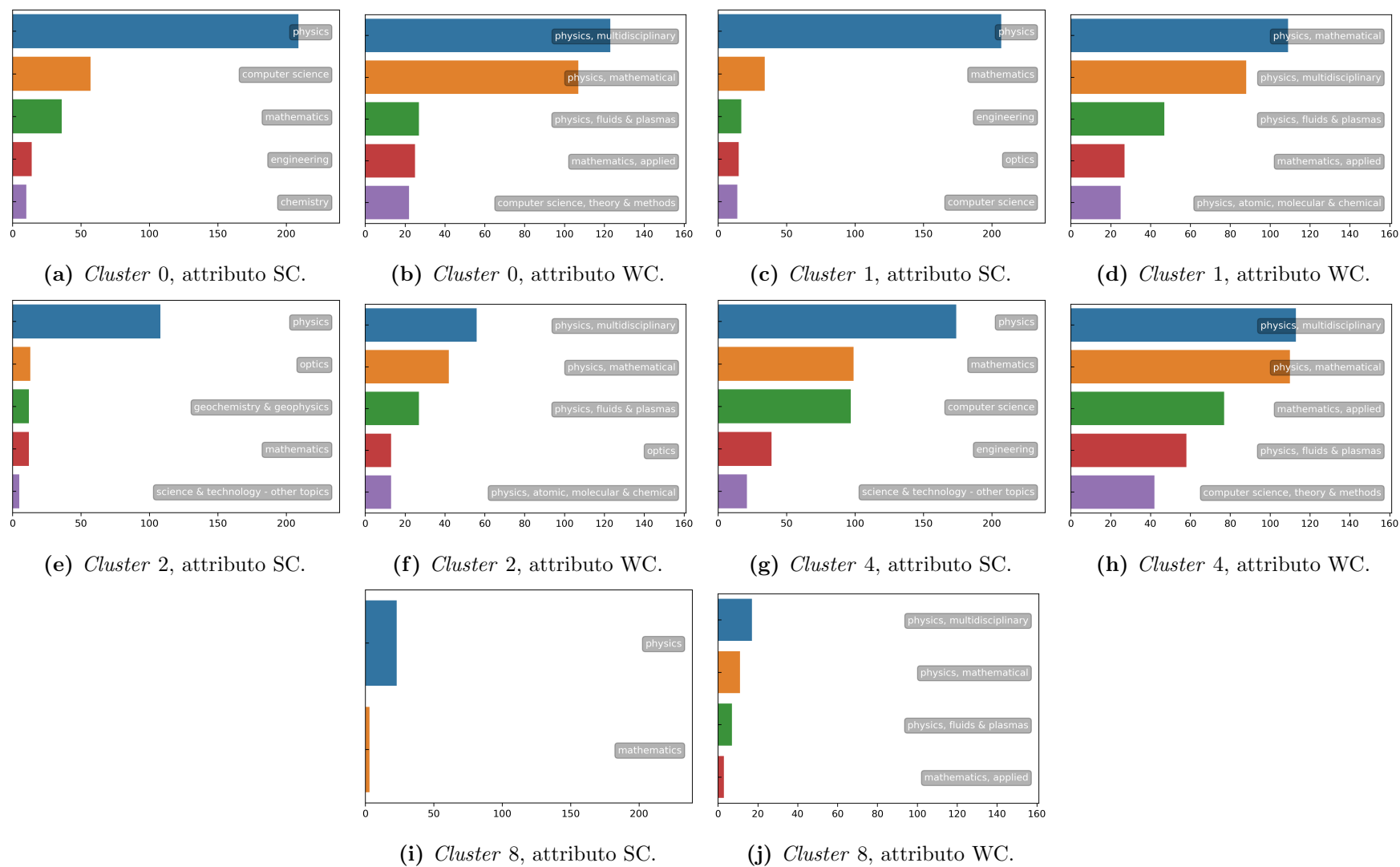


Figura 2.6 – Composizione dei *cluster* 0, 1, 2, 4 e 8. Sono rappresentati solo i cinque valori più numerosi per ogni *cluster*.

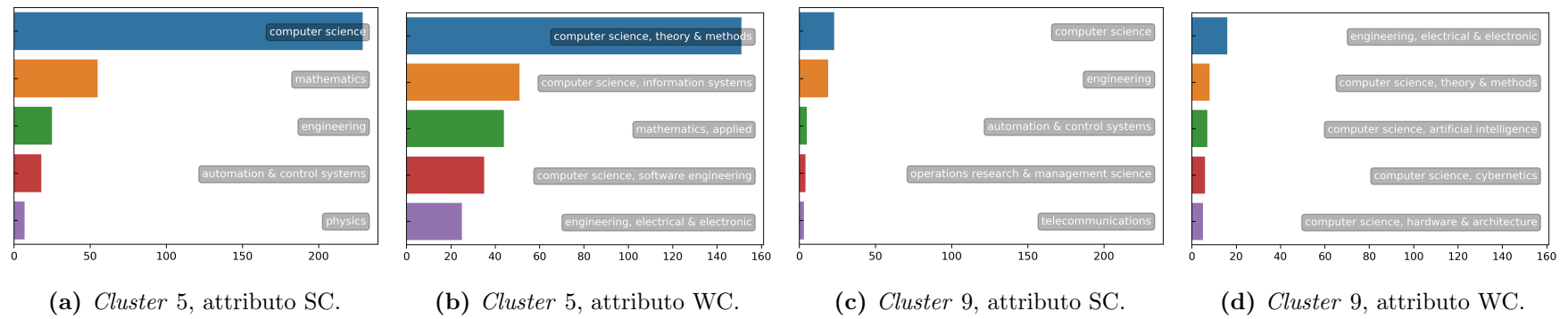


Figura 2.7 – Composizione dei *cluster* 5 e 9. Sono rappresentati solo i cinque valori più numerosi.

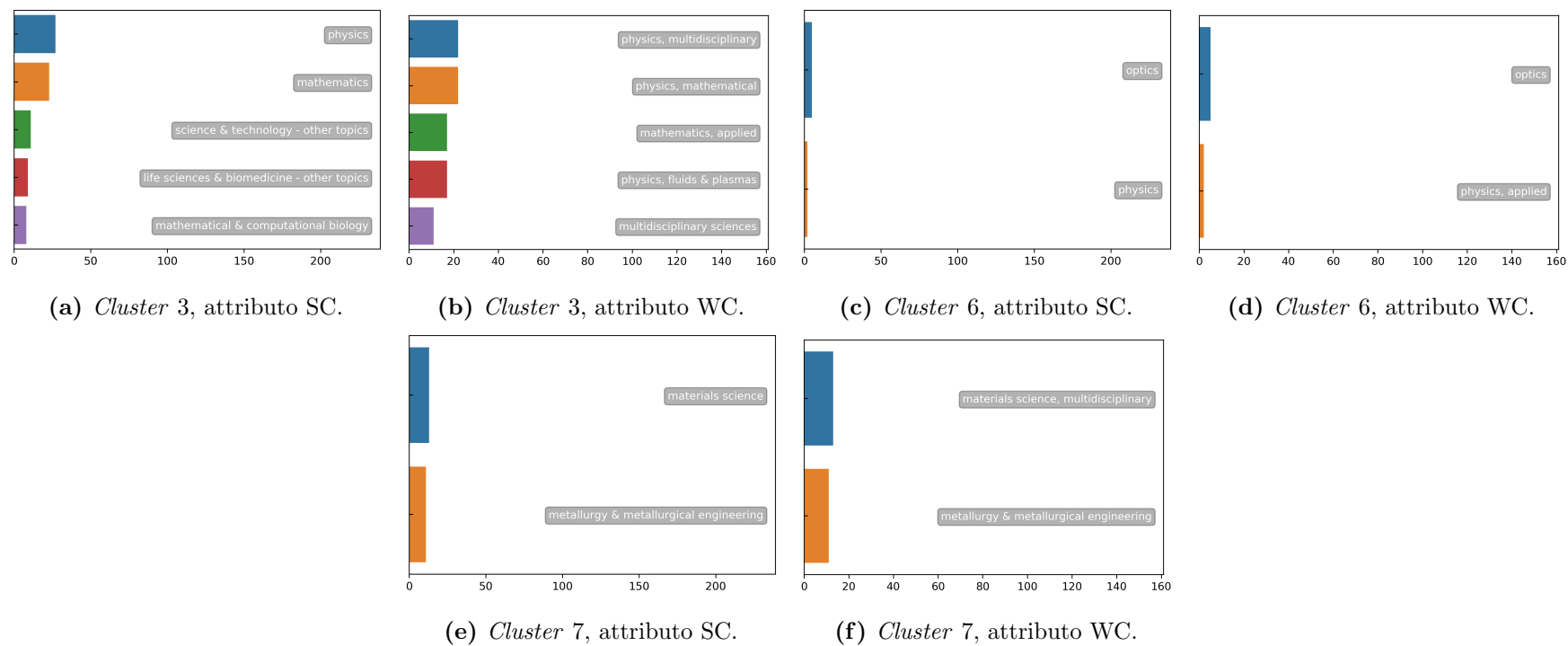


Figura 2.8 – Composizione dei *cluster* 3, 6 e 7. Sono rappresentati solo i cinque valori più numerosi.

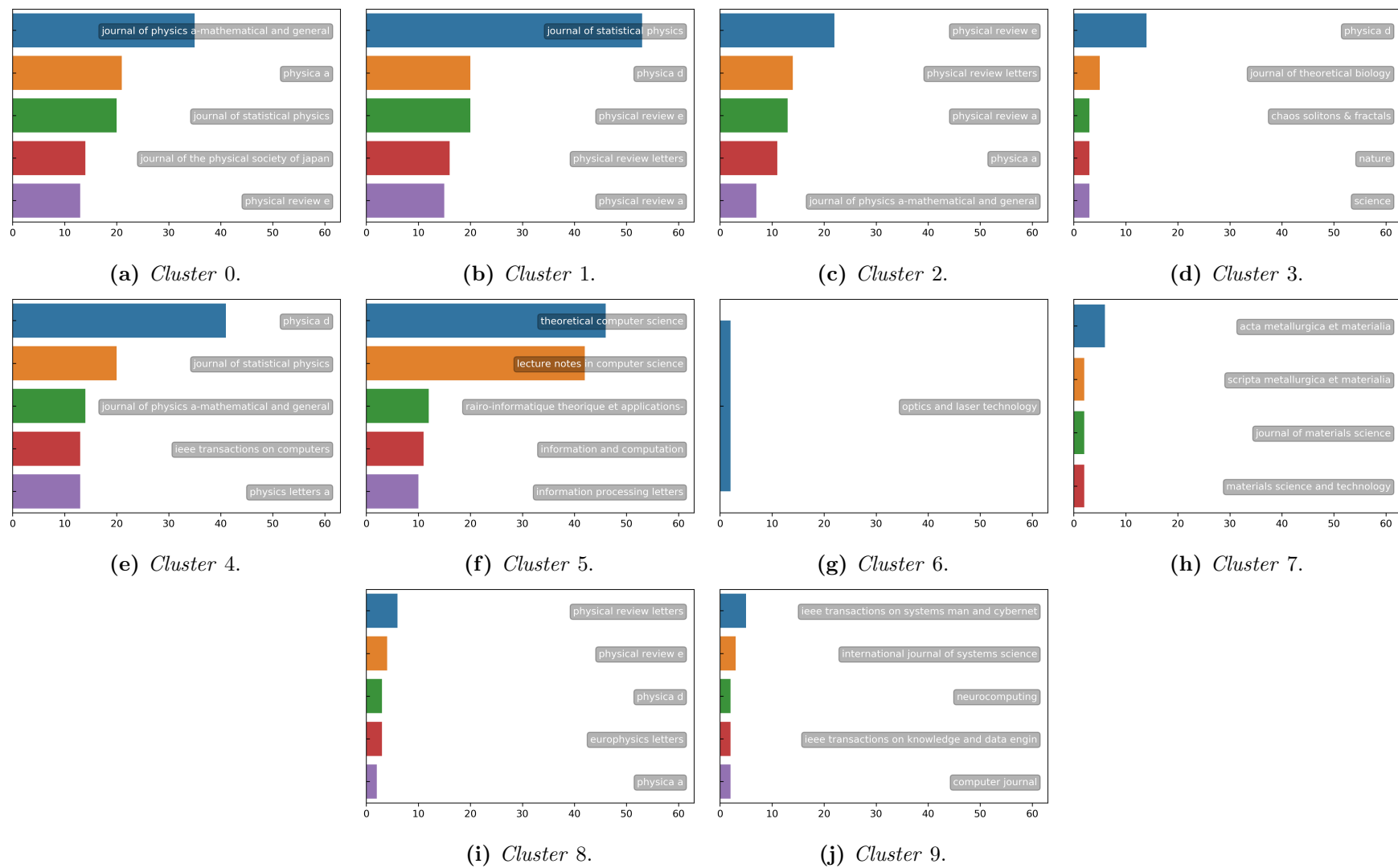


Figura 2.9 – Composizione dei *cluster* da 0 a 9 sulla base del parametro SO (riviste di pubblicazione). Sono rappresentati solo i cinque valori più numerosi per ogni *cluster*.

Parola o bigramma	Attributo	Test binomiale	Differenza dalla media
cars	AB	2.0022914291753427e-20	-
traffic	AB	2.860632058838363e-19	0.0031562509744026404
cell	AB	3.3128682328292003e-21	0.003734966111935245
phase	AB	8.801440931545118e-21	0.005134656740726249
jamming	AB	9.148904300607376e-15	-
cellular	AB	-	0.0050812496701602445
model	AB	-	0.0031551591867753463
model	TI	0.00010228816010450739	0.012323834728650777
critical	TI	0.00014902285585593577	0.005922270906415261
traffic	TI	1.1293152770082973e-09	0.010586450010892872
immune	TI	3.0782383455884686e-06	-
ising	TI	3.1554763112321e-05	-
cellular	TI	-	0.02012678267880247
automaton	TI	-	0.005288385300189852
flow model	TI_bi	0.0001469438421046962	0.0038407446130497265
cellular automaton	TI_bi	0.002848631023232703	0.007579116991936221
cellular automata	TI_bi	0.009704447275417218	0.011457851120067274
2 dimensional	TI_bi	0.018717485202792906	0.003222197380347157
traffic flow	TI_bi	4.613199533546739e-07	0.0069470252678981124

Tabella 2.7 – Test di rilevanza di parole e bigrammi, *cluster* 0.

Parola o bigramma	Attributo	Test binomiale	Differenza dalla media
gas	AB	1.6893946687127202e-63	0.009502121552856244
boltzmann	AB	2.6923566298160816e-33	0.004706575864599529
flow	AB	4.312545312080118e-22	0.004817599669476365
lattice	AB	4.545671857854336e-90	0.01621461338992594
fluid	AB	5.263084542332889e-29	0.004363298844953448
gas	TI	1.1374604067097525e-33	0.0366931491143713
fluid	TI	2.545828156467802e-08	-
fluids	TI	2.9818345811644285e-08	-
lattice	TI	3.243729934357042e-43	0.052141216611649686
boltzmann	TI	5.551750393036834e-11	0.01066749083141024
diffusion	TI	-	0.010166318324237294
flow	TI	-	0.008724192005116137
porous media	TI_bi	1.3348658837656123e-06	0.005780658811007802
gas cellular	TI_bi	5.612323870202806e-11	0.010671985497245174
lattice gas	TI_bi	7.2901768538820355e-31	0.033595813584218745
lattice boltzmann	TI_bi	8.426349269338987e-08	0.00711465699816345
lattice gases	TI_bi	8.426349269338987e-08	0.00711465699816345

Tabella 2.8 – Test di rilevanza di parole e bigrammi, *cluster* 1.

Parola o bigramma	Attributo	Test binomiale	Differenza dalla media
organized	AB	1.1821761217374131e-29	0.005273876007733063
earthquakes	AB	1.225985465666988e-24	-
self	AB	5.549278785453598e-22	0.006140338894115016
criticality	AB	7.960613992850794e-21	-
sandpile	AB	9.260425658111789e-26	-
model	AB	-	0.007573381682717919
critical	AB	-	0.005867368662785838
models	AB	-	0.004093873200849454
self	TI	1.1982992176290211e-17	0.03308022474394077
granular	TI	1.7288722820404673e-09	-
sandpile	TI	2.044626391664886e-14	0.018978460477777622
organized	TI	7.593140065756636e-23	0.031314459788333075
criticality	TI	7.69923919445138e-19	0.026371254657966227
model	TI	-	0.018925065993709134
cellular automaton	TI_bi	0.7314353424410142	-0.00187698572468575
cellular automata	TI_bi	0.9999999999634034	-0.03766227399506146
organized criticality	TI_bi	1.1743956799900303e-17	0.023747859120200242
self organized	TI_bi	7.599491442266531e-23	0.03134717403866432

Tabella 2.9 – Test di rilevanza di parole e bigrammi, *cluster 2*.

Parola o bigramma	Attributo	Test binomiale	Differenza dalla media
test	AB	1.1657309166269965e-10	-
local	AB	1.1998482163381497e-07	0.001937899911095923
cellular	AB	2.0087186254203477e-21	0.00826094744927939
ca	AB	5.033258919445798e-14	0.003601934770764585
rule	AB	5.156894063825126e-13	0.0025105091759466667
automata	AB	-	0.003291293663885546
additive	TI	0.00024666943013742213	-
linear	TI	0.0002923839623697978	0.004486594913698574
one	TI	0.0008063330599094246	0.005544886453535661
cellular	TI	1.3895389690437135e-16	0.044798162356915736
automata	TI	9.943055817035421e-11	0.03375937986040034
dimensional	TI	-	0.004425222021748313
one dimensional	TI_bi	0.00010741415067176747	0.0058225766563786115
automata based	TI_bi	0.00024301720520269354	-
linear cellular	TI_bi	0.0002481863143853524	0.0034548765564931076
deterministic cellular	TI_bi	0.00045540578143010826	-
cellular automata	TI_bi	1.7524161474396605e-20	0.04489862119689252
dimensional cellular	TI_bi	-	0.004300208442399132
automata cellular	TI_bi	-	0.003579331763859656

Tabella 2.10 – Test di rilevanza di parole e bigrammi, *cluster 4*.

Parola o bigramma	Attributo	Test binomiale	Differenza dalla media
languages	AB	1.4394098368064e-37	0.005083542417115332
automaton	AB	3.4132766266213916e-32	0.009903194977679677
finite	AB	5.228417581288219e-44	0.008919174489779996
language	AB	6.75395361829395e-51	0.007039968753214052
k	AB	8.278289007672256e-26	-
automata	AB	-	0.005143296468552367
language	TI	1.0613547073075168e-08	-
tree	TI	1.0613547073075168e-08	-
languages	TI	2.560885552509211e-13	0.014319169866085103
finite	TI	3.5463881406259134e-14	0.020579660219911814
regular	TI	6.781019982459438e-09	0.010183869321987456
systems	TI	-	0.011442067212594425
automata	TI	-	0.01102099098930942
cellular automata	TI_bi	0.9999999999999999	-0.037797927265302726
discrete event	TI_bi	2.484616319584412e-07	0.006946029778238123
finite automata	TI_bi	2.587648922614667e-13	0.014326965960888819

Tabella 2.11 – Test di rilevanza di parole e bigrammi, *cluster* 5.

Parola o bigramma	Attributo	Test binomiale	Differenza dalla media
time	AB	0.03467361368069196	0.005705289067662033
cellular	AB	0.1422543757569775	0.0043515987456458054
automata	AB	0.439257407697488	0.0009274524528250054

Tabella 2.12 – Test di rilevanza di parole e bigrammi, *cluster* 8.

Parola o bigramma	Attributo	Test binomiale	Differenza dalla media
control	AB	1.0810121070906872e-13	0.008039985158106637
performance	AB	1.2772802648020638e-13	-
virtual	AB	1.3725955379368707e-15	-
learning	AB	1.9299418556878264e-68	0.02679486293506945
estimator	AB	5.384933706207679e-16	-
algorithm	AB	-	0.009361940040647294
problem	AB	-	0.009277696070108605
stochastic	AB	-	0.007347658322103593
automata	TI	0.6852252601942539	-0.005735104930814361
learning	TI	1.6926522582250991e-31	0.09898948236749851
using	TI	1.7790448166770688e-06	0.03156599135151422
control	TI	1.952323334384078e-10	0.034743431526273384
learning automata	TI_bi	1.0994296641316946e-16	0.04650187112618414

Tabella 2.13 – Test di rilevanza di parole e bigrammi, *cluster* 9.

Capitolo 3

Conclusioni

Nel corso di questo lavoro si è fornito un resoconto storico del lavoro di ricerca sugli automi cellulari. Infatti, malgrado il ruolo fondamentale giocato da quest'ambito di ricerca nella nascita e nello sviluppo delle teorie della complessità, non è ad oggi disponibile una trattazione storica ad esso dedicata. È speranza di chi scrive che questo lavoro possa rappresentare un primo passo in tal senso.

Si è, in primo luogo, fornita una possibile definizione formale degli automi cellulari (§0.1). In seguito (§1.1) si sono richiamati i momenti salienti della nascita degli automi cellulari e dei primi anni della ricerca sul tema, a partire dai contributi di von Neumann di fine anni '40. La sezione 1.2, che riguarda soprattutto la prima parte degli anni '80, documenta invece l'inizio dello studio sistematico degli automi cellulari come modelli di sistemi complessi e la loro applicazione a cavallo tra diverse discipline. Il periodo successivo, in particolare il decennio 1985-1995, è invece stato oggetto di studio quantitativo (2.2).

La scelta di applicare tecniche bibliometriche allo studio di questo lasso di tempo ha permesso di aggirare alcuni dei limiti propri di un'analisi tradizionale, fondata sulla lettura diretta delle fonti (*close reading*). Grazie agli strumenti bibliometrici (introdotti nella sezione 2.1) è stato possibile caratterizzare la letteratura scientifica nel periodo di riferimento sia dal punto di vista disciplinare che da quello delle tematiche di ricerca. Ottenere un risultato analogo con mezzi tradizionali avrebbe comportato il coinvolgimento di diversi esperti della materia oggetto di studio, la lettura di una mole considerevole di letteratura specialistica, e non avrebbe eliminato il rischio di tralasciare ambiti di ricerca collegati – si pensi ad esempio ai *cluster* che testimoniano applicazioni degli automi cellulari in ottica o nelle scienze dei materiali. Naturalmente la lettura diretta dei testi, oltre a richiedere tempo e lavoro considerevoli, consente un livello di approfondimento decisamente più elevato

di quello ottenuto nell'ultima parte della nostra analisi. Questo conferma quanto affermato nell'introdurre le tecniche bibliometriche: queste rappresentano un ottimo strumento di indagine preliminare, forniscono una buona panoramica dell'argomento in questione e consentono di lavorare su scale altrimenti inaccessibili. Rappresentano, dunque, un utile supporto al lavoro interpretativo di indagine storica e non una sua sostituzione. È inoltre opinione di chi scrive che l'uso di tecniche bibliometriche per questo tipo di analisi – che mira alla ricostruzione ragionata di un panorama intellettuale complesso – possa contribuire ad avvicinare le scienze umane a metodi di indagine quantitativa, rappresentando un'alternativa costruttiva e arricchente alle applicazioni meramente valutative degli strumenti bibliometrici.

Per quanto riguarda la ricerca più recente sugli automi cellulari è possibile notare un'ultima cosa. La RYPS in figura 2.4 evidenzia una diminuzione considerevole dei riferimenti bibliografici a lavori successivi al 2002. Sebbene questo tipo di tendenza generale sia comune per quanto riguarda le pubblicazioni scientifiche (la distribuzione di qualunque loro attributo risente pesantemente di effetti cronologici), è probabile che nello specifico caso degli automi cellulari vi sia di più. Come si è detto in precedenza, negli ultimi decenni il loro studio dal punto di vista strettamente formale si è ridotto considerevolmente, mentre hanno acquisito sempre più rilevanza le loro applicazioni modellistiche – in particolare nella loro variante stocastica. Se nei primi decenni di vita degli automi cellulari abbiamo visto difficoltà e duplicazione del lavoro causate dalla mancanza di una terminologia comune, è probabile che osservando gli ultimi anni ci troviamo di fronte alla loro esplosione in una miriade di loro filiazioni – più o meno dirette – sotto nomi diversi. È dunque ragionevole pensare che per indagare più a fondo i legami tra automi cellulari e tecniche modellistiche diverse sia necessario districare la molteplicità di nomi e tecniche di modellizzazione emerse negli ultimi anni.

Bibliografia

- Alba, Enrique e Bernabe Dorronsoro (6 giu. 2008). *Cellular Genetic Algorithms*. Springer-Verlag GmbH. ISBN: 0387776095.
- Bak, Per, Chao Tang e Kurt Wiesenfeld (1987). “Self-organized criticality: An explanation of the $1/f$ noise”. In: *Physical Review Letters* 59.4, pp. 381–384. DOI: [10.1103/physrevlett.59.381](https://doi.org/10.1103/physrevlett.59.381).
- Beaver, D. e R. Rosen (1978). “Studies in scientific collaboration”. In: *Scientometrics* 1.1, pp. 65–84. DOI: [10.1007/bf02016840](https://doi.org/10.1007/bf02016840).
- Beebe, Nelson H. F. (2019). “A Complete Bibliography of Publications of John von Neumann”. Version 1.184.
- Bennett, C. H. (1973). “Logical Reversibility of Computation”. In: *IBM Journal of Research and Development* 17.6, pp. 525–532. DOI: [10.1147/rd.176.0525](https://doi.org/10.1147/rd.176.0525).
- Berlekamp, Elwyn R., John H. Conway e Richard K. Guy (1982). *Winning Ways for Your Mathematical Plays, Vol. 2*. London: Academic Press.
- Berto, Francesco e Jacopo Tagliabue (2017). “Cellular Automata”. In: *The Stanford Encyclopedia of Philosophy*. A cura di Edward N. Zalta. Fall 2017. Metaphysics Research Lab, Stanford University.
- Blondel, Vincent D. et al. (2008). “Fast unfolding of communities in large networks”. In: *Journal of Statistical Mechanics: Theory and Experiment* 2008.10, P10008. DOI: [10.1088/1742-5468/2008/10/p10008](https://doi.org/10.1088/1742-5468/2008/10/p10008).
- Bornmann, Lutz e Hans-Dieter Daniel (2008). “What do citation counts measure? A review of studies on citing behavior”. In: *Journal of Documentation* 64.1, pp. 45–80. DOI: [10.1108/00220410810844150](https://doi.org/10.1108/00220410810844150).
- Burks, Arthur W. (1969). *Von Neumann’s Self-Reproducing Automata*. Rapp. tecn. University of Michigan.
- Burks, Christian e Doyn Farmer (1984). “Towards modeling DNA sequences as automata”. In: *Physica D: Nonlinear Phenomena* 10.1, pp. 157–167. ISSN: 0167-2789. DOI: [https://doi.org/10.1016/0167-2789\(84\)90258-6](https://doi.org/10.1016/0167-2789(84)90258-6).

- Bush, Vannevar (1945). *Science: The Endless Frontier. A Report to the President on a Program for Postwar Scientific Research*. Available at: <https://www.nsf.gov/od/lpa/nsf50/vbush1945.htm>.
- Carter, Forrest L. (1984). "The molecular device computer: Point of departure for large scale cellular automata". In: *Physica D: Nonlinear Phenomena* 10.1, pp. 175–194. ISSN: 0167-2789. DOI: [https://doi.org/10.1016/0167-2789\(84\)90260-4](https://doi.org/10.1016/0167-2789(84)90260-4).
- Chubin, D. E. (1985). "Beyond invisible colleges: Inspirations and aspirations of post-1972 social studies of science". In: *Scientometrics* 7.3-6, pp. 221–254. DOI: [10.1007/bf02017148](https://doi.org/10.1007/bf02017148).
- Chubin, Daryl E. (1976). "State of the Field. The Conceptualization of Scientific Specialties". In: *The Sociological Quarterly* 17.4, pp. 448–476. DOI: [10.1111/j.1533-8525.1976.tb01715.x](https://doi.org/10.1111/j.1533-8525.1976.tb01715.x).
- Crutchfield, James P. (1984). "Space-time dynamics in video feedback". In: *Physica D: Nonlinear Phenomena* 10.1, pp. 229–245. ISSN: 0167-2789. DOI: [https://doi.org/10.1016/0167-2789\(84\)90264-1](https://doi.org/10.1016/0167-2789(84)90264-1).
- D'Angelo, Ciriaco Andrea, Cristiano Giuffrida e Giovanni Abramo (2010). "A heuristic approach to author name disambiguation in bibliometrics databases for large-scale research assessments". In: *Journal of the American Society for Information Science and Technology* 62.2, pp. 257–269. DOI: [10.1002/asi.21460](https://doi.org/10.1002/asi.21460).
- Das, Debasis (2012). "A Survey on Cellular Automata and Its Applications". In: *Communications in Computer and Information Science*. Springer Berlin Heidelberg, pp. 753–762. DOI: [10.1007/978-3-642-29219-4_84](https://doi.org/10.1007/978-3-642-29219-4_84).
- Dennett, D. C. (1995). *Darwin's dangerous idea : evolution and the meanings of life*. New York: Simon & Schuster. ISBN: 9780684802909.
- Deutsch, David (1985). "Quantum theory, the Church–Turing principle and the universal quantum computer". In: *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences* 400.1818, pp. 97–117. DOI: [10.1098/rspa.1985.0070](https://doi.org/10.1098/rspa.1985.0070).
- Domany, Eytan e Wolfgang Kinzel (1984). "Equivalence of Cellular Automata to Ising Models and Directed Percolation". In: *Physical Review Letters* 53.4, pp. 311–314. DOI: [10.1103/physrevlett.53.311](https://doi.org/10.1103/physrevlett.53.311).
- Edge, David (1979). "Quantitative Measures of Communication in Science: A Critical Review". In: *History of Science* 17.2, pp. 102–134. DOI: [10.1177/007327537901700202](https://doi.org/10.1177/007327537901700202).
- Eom, Sean B. (11 mag. 2011). *Author Cocitation Analysis: Quantitative Methods for Mapping the Intellectual Structure of an Academic Discipline*. Information Science Reference. 368 pp. ISBN: 1599047381.
- Epstein, Joshua M. e Robert L. Axtell (11 ott. 1996). *Growing Artificial Societies*. MIT Press Ltd. 224 pp. ISBN: 0262550253.

- Feynman, Richard P. (1982). “Simulating physics with computers”. In: *International Journal of Theoretical Physics* 21.6-7, pp. 467–488. DOI: [10.1007/bf02650179](https://doi.org/10.1007/bf02650179).
- Freeman, Linton C. (1977). “A Set of Measures of Centrality Based on Betweenness”. In: *Sociometry* 40.1, p. 35. DOI: [10.2307/3033543](https://doi.org/10.2307/3033543).
- Garfield, E. (1955). “Citation Indexes for Science: A New Dimension in Documentation through Association of Ideas”. In: *Science* 122.3159, pp. 108–111. DOI: [10.1126/science.122.3159.108](https://doi.org/10.1126/science.122.3159.108).
- Georgiadis, Evangelos (2007). “A Note on Minimal Boolean Formula Size of One-Dimensional Cellular Automata”. In: *J. Cell. Autom.* 2, pp. 351–353.
- Giles, Jim (2002). “What kind of science is this?”. In: *Nature* 417.6886, pp. 216–218. DOI: [10.1038/417216a](https://doi.org/10.1038/417216a).
- Gosper, R.Wm. (1984). “Exploiting regularities in large cellular spaces”. In: *Physica D: Nonlinear Phenomena* 10.1, pp. 75–80. ISSN: 0167-2789. DOI: [https://doi.org/10.1016/0167-2789\(84\)90251-3](https://doi.org/10.1016/0167-2789(84)90251-3).
- Gotts, Nicholas Mark e Paul B. Callahan (1998). “Emergent structures in sparse fields of Conway’s “Game of Life””. In: *Proceedings of the Sixth International Conference on Artificial Life (Complex Adaptive Systems)*.
- Grassberger, P. (1984). “Chaos and diffusion in deterministic cellular automata”. In: *Physica D: Nonlinear Phenomena* 10.1, pp. 52–58. ISSN: 0167-2789. DOI: [https://doi.org/10.1016/0167-2789\(84\)90248-3](https://doi.org/10.1016/0167-2789(84)90248-3).
- Hillis, W. Daniel (1984). “The connection machine: A computer architecture based on cellular automata”. In: *Physica D: Nonlinear Phenomena* 10.1, pp. 213–228. ISSN: 0167-2789. DOI: [https://doi.org/10.1016/0167-2789\(84\)90263-X](https://doi.org/10.1016/0167-2789(84)90263-X).
- Hopcroft, John (1979). *Introduction to automata theory, languages, and computation*. Reading, Mass: Addison-Wesley. ISBN: 020102988X.
- Ilachinski, Andrew (2001). *Cellular Automata: A Discrete Universe*. World Scientific Publishing Company. 808 pp. ISBN: 981238183X.
- Ingerson, T.E. e R.L. Buvel (1984). “Structure in asynchronous cellular automata”. In: *Physica D: Nonlinear Phenomena* 10.1, pp. 59–68. ISSN: 0167-2789. DOI: [https://doi.org/10.1016/0167-2789\(84\)90249-5](https://doi.org/10.1016/0167-2789(84)90249-5).
- Kamada, Tomihisa e Satoru Kawai (1989). “An algorithm for drawing general undirected graphs”. In: *Information Processing Letters* 31.1, pp. 7–15. DOI: [10.1016/0020-0190\(89\)90102-6](https://doi.org/10.1016/0020-0190(89)90102-6).
- Kauffman, Stuart A. (1984). “Emergent properties in random complex automata”. In: *Physica D: Nonlinear Phenomena* 10.1, pp. 145–156. ISSN: 0167-2789. DOI: [https://doi.org/10.1016/0167-2789\(84\)90257-4](https://doi.org/10.1016/0167-2789(84)90257-4).
- Kessler, M. M. (1963). “Bibliographic coupling between scientific papers”. In: *American Documentation* 14.1, pp. 10–25. DOI: [10.1002/asi.5090140103](https://doi.org/10.1002/asi.5090140103).

- Kuhn, Thomas (2000). *The road since Structure : philosophical essays, 1970-1993, with an autobiographical interview*. Chicago: University of Chicago Press. ISBN: 0226457982.
- Kuhn, Thomas S. (1970). *The structure of scientific revolutions*. Chicago: University of Chicago Press, pp. xii, 210.
- Lakatos, Imre (1978). *The Methodology of Scientific Research Programmes*. A cura di John Worrall e Gregory Currie. Cambridge University Press. DOI: [10.1017/cbo9780511621123](https://doi.org/10.1017/cbo9780511621123).
- Langton, Christopher G. (1984). "Self-reproduction in cellular automata". In: *Physica D: Nonlinear Phenomena* 10.1, pp. 135 –144. ISSN: 0167-2789. DOI: [https://doi.org/10.1016/0167-2789\(84\)90256-2](https://doi.org/10.1016/0167-2789(84)90256-2).
- Langton, Christopher G (1986). "Studying artificial life with cellular automata". In: *Physica D: Nonlinear Phenomena* 22.1-3, pp. 120–149. DOI: [10.1016/0167-2789\(86\)90237-x](https://doi.org/10.1016/0167-2789(86)90237-x).
- Laurens, Patricia, Michel Zitt e Elise Bassecoulard (2010). "Delineation of the genomics field by hybrid citation-lexical methods: interaction with experts and validation process". In: *Scientometrics* 82.3, pp. 647–662. DOI: [10.1007/s11192-010-0177-9](https://doi.org/10.1007/s11192-010-0177-9).
- Leahey, Erin e Ryan C. Reikowsky (2008). "Research Specialization and Collaboration Patterns in Sociology". In: *Social Studies of Science* 38.3, pp. 425–440. DOI: [10.1177/0306312707086190](https://doi.org/10.1177/0306312707086190).
- Lind, D.A. (1984). "Applications of ergodic theory and sofic systems to cellular automata". In: *Physica D: Nonlinear Phenomena* 10.1, pp. 36 –44. ISSN: 0167-2789. DOI: [https://doi.org/10.1016/0167-2789\(84\)90246-X](https://doi.org/10.1016/0167-2789(84)90246-X).
- Lotka, Alfred J. (1926). "The frequency distribution of scientific productivity". In: *Journal of the Washington Academy of Sciences* 16.12, pp. 317–323. ISSN: 00430439.
- MacRoberts, Michael H. e Barbara R. MacRoberts (1989). "Problems of citation analysis: A critical review". In: *Journal of the American Society for Information Science* 40.5, pp. 342–349. DOI: [10.1002/\(sici\)1097-4571\(198909\)40:5<342::aid-asi7>3.0.co;2-u](https://doi.org/10.1002/(sici)1097-4571(198909)40:5<342::aid-asi7>3.0.co;2-u).
- Margolus, Norman (1984). "Physics-like models of computation". In: *Physica D: Nonlinear Phenomena* 10.1, pp. 81 –95. ISSN: 0167-2789. DOI: [https://doi.org/10.1016/0167-2789\(84\)90252-5](https://doi.org/10.1016/0167-2789(84)90252-5).
- Martyn, John (1964). "Bibliographic Coupling". In: *Journal of Documentation* 20.4, pp. 236–236. DOI: [10.1108/eb026352](https://doi.org/10.1108/eb026352).
- (1975). "Citation Analysis". In: *Journal of Documentation* 31.4, pp. 290–297. DOI: [10.1108/eb026610](https://doi.org/10.1108/eb026610).

- Marx, Werner et al. (2013). "Detecting the historical roots of research fields by reference publication year spectroscopy (RPYS)". In: *Journal of the Association for Information Science and Technology* 65.4, pp. 751–764. DOI: [10.1002/asi.23089](https://doi.org/10.1002/asi.23089).
- McCulloch, Warren S. e Walter Pitts (1943). "A logical calculus of the ideas immanent in nervous activity". In: *The Bulletin of Mathematical Biophysics* 5.4, pp. 115–133. DOI: [10.1007/bf02478259](https://doi.org/10.1007/bf02478259).
- Merton, Robert (1968). *Social theory and social structure*. New York: Free Press. ISBN: 0029211301.
- Metropolis, N., J. Hewlett e Gian-Carlo Rota, cur. (1980). *History of Computing in the Twentieth Century*. Academic Press. ISBN: 0124916503.
- Metropolis, Nicholas (1987). "The beginning of the Monte Carlo method". In: *Los Alamos Science (Special Issue dedicated to Stanislaw Ulam)*, 125–130.
- Moore, Edward F. (1962). "Machine models of self-reproduction". In: *Proceedings of the Fourteenth Symposium in Applied Mathematics of the American Mathematical Society. Mathematical Problems in the Biological Sciences*.
- Morris, Steven A. e Kevin W. Boyack (2005). "Boyack Visualizing 60 Years of Anthrax Research". In: *Proceedings of ISSI 2005 - the 10th International Conference of the International Society for Scientometrics and Informetrics*. A cura di B. Larsen P. Ingwersen. Vol. 1. Stockholm, Sweden: Karolinska University Press, pp. 45–55.
- Morris, Steven A. e Betsy Van der Veer Martens (2009). "Mapping research specialties". In: *Annual Review of Information Science and Technology* 42.1, pp. 213–295. DOI: [10.1002/aris.2008.1440420113](https://doi.org/10.1002/aris.2008.1440420113).
- Morris, Steven A. et al. (2003). "Time line visualization of research fronts". In: *Journal of the American Society for Information Science and Technology* 54.5, pp. 413–422. DOI: [10.1002/asi.10227](https://doi.org/10.1002/asi.10227).
- Mulkay, M. J. (1975). "Three Models of Scientific Development". In: *The Sociological Review* 23.3, pp. 509–526. DOI: [10.1111/j.1467-954x.1975.tb02231.x](https://doi.org/10.1111/j.1467-954x.1975.tb02231.x).
- Myhill, John (1963). "Shorter Note: The Converse of Moore's Garden-of-Eden Theorem". In: *Proceedings of the American Mathematical Society* 14.4, p. 685. DOI: [10.2307/2034301](https://doi.org/10.2307/2034301).
- Narendra, Kumpati S. e M. A. L. Thathachar (1974). "Learning Automata - A Survey". In: *IEEE Transactions on Systems, Man, and Cybernetics* SMC-4.4, pp. 323–334. DOI: [10.1109/tsmc.1974.5408453](https://doi.org/10.1109/tsmc.1974.5408453).
- Newman, M. E. J. (2005). "Power laws, Pareto distributions and Zipf's law". In: *Contemporary Physics* 46.5, pp. 323–351. DOI: [10.1080/00107510500052444](https://doi.org/10.1080/00107510500052444).
- Omohundro, Stephen (1984). "Modelling cellular automata with partial differential equations". In: *Physica D: Nonlinear Phenomena* 10.1, pp. 128–134. ISSN: 0167-2789. DOI: [https://doi.org/10.1016/0167-2789\(84\)90255-0](https://doi.org/10.1016/0167-2789(84)90255-0).

- Petrovich, Eugenio (2020). "Science mapping". In: *Encyclopedia of Knowledge Organization (ISKO)*.
- Petrovich, Eugenio e Emiliano Tolusso (2019). "Exploring Knowledge Dynamics in the Humanities. Two Science Mapping Experiments". en. In: *Journal of Interdisciplinary History of Ideas* Vol 8, No 16 (2019)–. DOI: [10.13135/2280-8574/4304](https://doi.org/10.13135/2280-8574/4304).
- Preston, Kendall (1984). "Four-dimensional logical transforms: Data processing by cellular automata". In: *Physica D: Nonlinear Phenomena* 10.1, pp. 205–212. ISSN: 0167-2789. DOI: [https://doi.org/10.1016/0167-2789\(84\)90262-8](https://doi.org/10.1016/0167-2789(84)90262-8).
- Preston, Kendall e Michael J. B. Duff (1984). *Modern Cellular Automata*. Springer US. DOI: [10.1007/978-1-4899-0393-8](https://doi.org/10.1007/978-1-4899-0393-8).
- Price, Derek (1963). *Little science, big science*. New York: Columbia Univ. Press. ISBN: 0231085621.
- (1965). "Networks of Scientific Papers". In: *Science* 149.3683, pp. 510–515. DOI: [10.1126/science.149.3683.510](https://doi.org/10.1126/science.149.3683.510).
- (1986). *Little science, big science– and beyond*. New York: Columbia University Press. ISBN: 0231049560.
- Rendell, Paul (2002). "Collision-Based Computing". In: a cura di Adamatzky A. Springer. Cap. Turing Universality of the Game of Life.
- Sarkar, Palash (2000). "A brief history of cellular automata". In: *ACM Computing Surveys* 32.1, pp. 80–107. DOI: [10.1145/349194.349202](https://doi.org/10.1145/349194.349202).
- Schiff, Joel L. (2007). *Cellular Automata: A Discrete View of the World*. John Wiley & Sons. 288 pp. ISBN: 047016879X.
- Shannon, Claude E. (1958). "Von Neumann's contributions to automata theory". In: *Bulletin of the American Mathematical Society* 64.3.
- Small, H. G. e D. Crane (1979). "Specialties and disciplines in science and social science: An examination of their structure using citation indexes". In: *Scientometrics* 1.5-6, pp. 445–461. DOI: [10.1007/bf02016661](https://doi.org/10.1007/bf02016661).
- Small, Henry (1973). "Co-citation in the scientific literature: A new measure of the relationship between two documents". In: *Journal of the American Society for Information Science* 24.4, pp. 265–269. DOI: [10.1002/asi.4630240406](https://doi.org/10.1002/asi.4630240406).
- (2006). "Tracking and predicting growth areas in science". In: *Scientometrics* 68.3, pp. 595–610. DOI: [10.1007/s11192-006-0132-y](https://doi.org/10.1007/s11192-006-0132-y).
- Small, Henry e Edwin Greenlee (1989). "A Co-Citation Study of AIDS Research". In: *Communication Research* 16.5, pp. 642–666. DOI: [10.1177/009365089016005006](https://doi.org/10.1177/009365089016005006).
- Smith, Steven A., Richard C. Watt e Stuart R. Hameroff (1984). "Cellular automata in cytoskeletal lattices". In: *Physica D: Nonlinear Phenomena* 10.1, pp. 168–174. ISSN: 0167-2789. DOI: [https://doi.org/10.1016/0167-2789\(84\)90259-8](https://doi.org/10.1016/0167-2789(84)90259-8).

- Sugimoto, Cassidy R. e Scott Weingart (2015). “The kaleidoscope of disciplinarity”. In: *Journal of Documentation* 71.4, pp. 775–794. DOI: [10.1108/jd-06-2014-0082](https://doi.org/10.1108/jd-06-2014-0082).
- Tang, Li e John P. Walsh (2010). “Bibliometric fingerprints: name disambiguation based on approximate structure equivalence of cognitive maps”. In: *Scientometrics* 84.3, pp. 763–784. DOI: [10.1007/s11192-010-0196-6](https://doi.org/10.1007/s11192-010-0196-6).
- Thatcher, J. W. (1964). *Universality in the von Neumann cellular model*. Rapp. tecn. University of Michigan.
- Toffoli, Tommaso (1984a). “CAM: A high-performance cellular-automaton machine”. In: *Physica D: Nonlinear Phenomena* 10.1, pp. 195–204. ISSN: 0167-2789. DOI: [https://doi.org/10.1016/0167-2789\(84\)90261-6](https://doi.org/10.1016/0167-2789(84)90261-6).
- (1984b). “Cellular automata as an alternative to (rather than an approximation of) differential equations in modeling physics”. In: *Physica D: Nonlinear Phenomena* 10.1, pp. 117–127. ISSN: 0167-2789. DOI: [https://doi.org/10.1016/0167-2789\(84\)90254-9](https://doi.org/10.1016/0167-2789(84)90254-9).
- Toffoli, Tommaso e Norman Margolus (1987). *Cellular Automata Machines: A New Environment for Modeling*. The MIT Press. ISBN: 0262200600.
- Toffoli, Tommaso e Norman H. Margolus (1990). “Invertible cellular automata: A review”. In: *Physica D: Nonlinear Phenomena* 45.1-3, pp. 229–253. DOI: [10.1016/0167-2789\(90\)90185-r](https://doi.org/10.1016/0167-2789(90)90185-r).
- Tougaw, P. Douglas e Craig S. Lent (1994). “Logical devices implemented using quantum cellular automata”. In: *Journal of Applied Physics* 75.3, pp. 1818–1825. DOI: [10.1063/1.356375](https://doi.org/10.1063/1.356375).
- Ulam, Stanislaw (1950). “Random processes and transformations”. In: *Proceedings of the International Congress of Mathematics. Vol. 2. Cambridge, Mass.*
- (1958). “John von Neumann 1903–1957”. In: *Bulletin of the American Mathematical Society* 64.3.
- (1962). “On some mathematical problems connected with patterns of growth of figures”. In: *Proceedings of the Fourteenth Symposium in Applied Mathematics of the American Mathematical Society. Mathematical Problems in the Biological Sciences.*
- Vargas-Quesada, Benjamín e Félix de Moya-Anegón (14 mag. 2007). *Visualizing the Structure of Science*. Springer Berlin Heidelberg. 320 pp. ISBN: 3540697276.
- Vichniac, Gérard Y. (1984). “Simulating physics with cellular automata”. In: *Physica D: Nonlinear Phenomena* 10.1, pp. 96–116. ISSN: 0167-2789. DOI: [https://doi.org/10.1016/0167-2789\(84\)90253-7](https://doi.org/10.1016/0167-2789(84)90253-7).
- von Neumann, John (1958). *The Computer and the Brain*. Yale Univ Pr. ISBN: 0300007930.

- von Neumann, John e Arthur W. Burks (1966). *Theory Of Self Reproducing Automata*. University of Illinois Press. ISBN: 0252727339.
- von Neumann, John e A. H. Taub (1961). *John von Neumann Collected Works v. 1: Design of Computers, Theory of Automata and Numerical Analysis*.
- (1963). *John von Neumann Collected Works v. 5: Design of Computers, Theory of Automata and Numerical Analysis*. ISBN: 0080095712.
- Waterman, Michael S. (1984). “Some applications of information theory to cellular automata”. In: *Physica D: Nonlinear Phenomena* 10.1, pp. 45 –51. ISSN: 0167-2789. DOI: [https://doi.org/10.1016/0167-2789\(84\)90247-1](https://doi.org/10.1016/0167-2789(84)90247-1).
- Watrous, J. (1995). “On one-dimensional quantum cellular automata”. In: *Proceedings of IEEE 36th Annual Foundations of Computer Science*. IEEE Comput. Soc. Press. DOI: [10.1109/sfcs.1995.492583](https://doi.org/10.1109/sfcs.1995.492583).
- Weinberg, A. M. (1961). “Impact of Large-Scale Science on the United States: Big science is here to stay, but we have yet to make the hard financial and educational choices it imposes”. In: *Science* 134.3473, pp. 161–164. DOI: [10.1126/science.134.3473.161](https://doi.org/10.1126/science.134.3473.161).
- Weisberg, Michael e Ryan Muldoon (2009). “Epistemic Landscapes and the Division of Cognitive Labor”. In: *Philosophy of Science* 76.2, pp. 225–252. DOI: [10.1086/644786](https://doi.org/10.1086/644786).
- White, Howard D. e Belver C. Griffith (1981). “Author cocitation: A literature measure of intellectual structure”. In: *Journal of the American Society for Information Science* 32.3, pp. 163–171. DOI: [10.1002/asi.4630320302](https://doi.org/10.1002/asi.4630320302).
- (1982). “Authors As Markers of Intellectual Space: Co-citation in Studies of Science, Technology and Society”. In: *Journal of Documentation* 38.4, pp. 255–272. DOI: [10.1108/eb026731](https://doi.org/10.1108/eb026731).
- Willson, Stephen J. (1984). “Growth rates and fractional dimensions in cellular automata”. In: *Physica D: Nonlinear Phenomena* 10.1, pp. 69 –74. ISSN: 0167-2789. DOI: [https://doi.org/10.1016/0167-2789\(84\)90250-1](https://doi.org/10.1016/0167-2789(84)90250-1).
- Wolfram, Stephen (1983). “Statistical mechanics of cellular automata”. In: *Reviews of Modern Physics* 55.3, pp. 601–644. DOI: [10.1103/revmodphys.55.601](https://doi.org/10.1103/revmodphys.55.601).
- (1984a). “Cellular automata as models of complexity”. In: *Nature* 311.5985, pp. 419–424. DOI: [10.1038/311419a0](https://doi.org/10.1038/311419a0).
- (1984b). “Universality and complexity in cellular automata”. In: *Physica D: Nonlinear Phenomena* 10.1, pp. 1 –35. ISSN: 0167-2789. DOI: [https://doi.org/10.1016/0167-2789\(84\)90245-8](https://doi.org/10.1016/0167-2789(84)90245-8).
- (1994). *Cellular Automata And Complexity: Collected Papers*. Basic Books. ISBN: 0-201-62716-7.
- (2002). *A New Kind of Science*. Wolfram Media inc. 1197 pp. ISBN: 1579550088.

- Wolfram, Stephen (1984c). “Preface”. In: *Physica D: Nonlinear Phenomena* 10.1, pp. vii–xii. ISSN: 0167-2789. DOI: [https://doi.org/10.1016/0167-2789\(84\)90244-6](https://doi.org/10.1016/0167-2789(84)90244-6).
- Wray, K. Brad (2005). “Rethinking Scientific Specialization”. In: *Social Studies of Science* 35.1, pp. 151–164. DOI: [10.1177/0306312705045811](https://doi.org/10.1177/0306312705045811).
- Zitt, Michel e Elise Bassecoulard (2006). “Delineating complex scientific fields by an hybrid lexical-citation method: An application to nanosciences”. In: *Information Processing & Management* 42.6, pp. 1513–1531. DOI: [10.1016/j.ipm.2006.03.016](https://doi.org/10.1016/j.ipm.2006.03.016).
- Zuse, Konrad (1967). *Rechnender Raum*. Schriften zur Datenverarbeitung. Traduzione: Calculating Space, MIT Technical Translation AZT-70-164-GEMIT, MIT (Proj. MAC), Cambridge, Mass. 02139, Feb. 1970. Vieweg & Sohn.