DigitCult | Scientific Journal on Digital Cultures

La rete degli editori

Modelli di text-mining e network analysis a partire dai dati di aNobii

Chiara Faggiolani
Università La Sapienza
Dipartimento di Scienze Documentarie, Linguistico-Filologiche e Geografiche

Lorenzo Verna
Independent data scientist

Maurizio Vivarelli
Università di Torino
Dipartimento di Studi storici

Abstract

Obiettivo di questo contributo è quello di esaminare e discutere presupposti, metodi e risultati dell'analisi di dati estratti dalla piattaforma di social reading aNobii (http://www.anobii.com/) nell'ambito del progetto "Leggere in rete. Analisi delle pratiche di lettura in ambiente digitale", in collaborazione tra Università degli Studi di Roma La Sapienza e Università degli Studi di Torino. Qui vengono presentati in particolare i risultati relativi all'analisi degli editori a partire non dai classici dati relativi sulla produzione editoriale rilevati annualmente da Istat ma a partire dalle recensioni dei libri inserite dai lettori sulla piattaforma aNobii. La ricerca è stata condotta secondo due prospettive tra loro integrate: una orientata a definire e visualizzare, in forma di grafo, la rete degli editori, e si avvale di strumenti ed euristiche situati nel campo della network science; l'altra, a partire dalla segmentazione degli editori realizzata attraverso le metriche di rete, analizza i vocabolari relativi a ciascun editore e ne individua le specificità, attraverso le tecniche dell'analisi automatica dei testi.

Publishers Network
Text-mining and network analysis models based on aNobii dataset

This paper aims to examine and discuss methods and results of the analysis of data extracted from the social reading platform aNobii (http://www.anobii.com/). This research is a part of the project "Read on the Net. Analysis of reading practices in a digital environment" (Leggere in rete. Analisi delle pratiche di lettura in ambiente digitale), in collaboration between the University of Rome La Sapienza and the University of Turin. Here we present in particular the results related to the analysis of publishers starting not from the classic data on publishing production reported annually by Istat but based on the reviews of the books left by readers on the platform aNobii. The research was conducted according to two integrated perspectives: one oriented to define and visualize, in graph form, the network of publishers using network science; the other, starting from the segmentation of the publishers realized through the network metrics, analyzes the vocabularies related to each publisher and identifies their specificities, through the techniques of text mining.

Published 29 September 2018

Correspondence should be addressed to Chiara Faggiolani, Dipartimento di Scienze Documentarie, Linguistico-Filologiche e Geografiche, Università La Sapienza, Piazzale Aldo Moro 5, 00185 Roma. Email: chiara.faggiolani@uniroma1.it

DigitCult, Scientific Journal on Digital Cultures is an academic journal of international scope, peer-reviewed and open access, aiming to value international research and to present current debate on digital culture, technological innovation and social change. ISSN: 2531-5994. URL: http://www.digitcult.it

Copyright rests with the authors. This work is released under a Creative Commons Attribution (IT) Licence, version 3.0. For details please see http://creativecommons.org/licenses/by/3.0/it/

 doi:10.4399/97888255181531

Premessa

Obiettivo di questo contributo è quello di esaminare e discutere presupposti, metodi e risultati dell'analisi di dati estratti dalla piattaforma di social reading aNobii (http://www.anobii.com/), secondo modalità che verranno dettagliatamente descritte nei paragrafi successivi; i dati presi in esame sono riferiti agli editori[1]. La ricerca effettuata è condotta secondo due prospettive, tra loro integrate; una, nella sua dimensione specifica, riguarda l'analisi delle parole utilizzate dai membri della community per descrivere, valutare, commentare la propria esperienza di lettura, e si colloca dunque nell'ambito del text-mining; l'altra è orientata a definire e visualizzare, in forma di grafo, la rete degli editori, e si avvale di strumenti ed euristiche situati nel campo della network science. Questa linea di indagine si inserisce all'interno di un percorso seguito nel corso degli ultimi anni dagli autori, le cui caratteristiche ed i cui esiti sono stati comunicati in numerose sedi editoriali (Faggiolani e Vivarelli 2016; Faggiolani, Verna e Vivarelli 2017). Come si accennava in precedenza oggetto specifico della ricerca presentata in questa sede sono gli editori, e le relazioni ad essi riferite presenti all'interno della piattaforma. La rete degli editori, che verrà presentata successivamente, va dunque ad aggiungersi ad altri grafi costruiti a partire dai dati di aNobii, come quello relativo ai libri, presentato con la Fig. 1.

Le ipotesi ed i risultati attesi sono dunque in primo luogo di natura descrittiva, e riguardano in senso stretto la rappresentazione delle entità oggetto dello studio. Su questa base si discutono alcune implicazioni dai caratteri più generali, riferite in senso più specifico e ristretto al campo degli studi sul social reading e, infine, alla riconfigurazione in atto della lettura in ambiente digitale. Questi elementi, di contesto e di scenario, sono delineati nel paragrafo successivo.

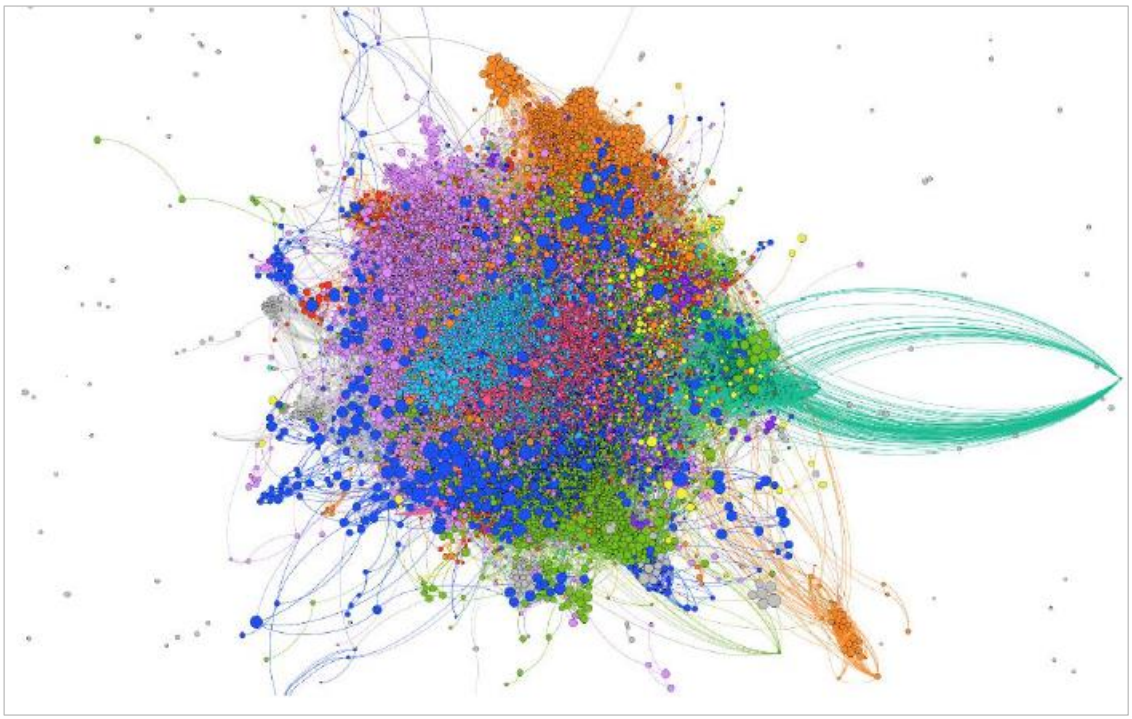

Figura 1. Una visualizzazione della rete dei libri di aNobii.

[1] Gli autori condividono i contenuti del contributo nel suo insieme. Si precisa che vanno attribuiti a Chiara Faggiolani i paragrafi *Il profilo emergente degli editori attraverso le parole dei lettori*; *Considerazioni di metodo*; a Lorenzo Verna il paragrafo *L'analisi della rete dei libri*; a Maurizio Vivarelli i paragrafi *I dati di aNobii ed i loro contesti* e *La natura sociale della lettura*. Data di ultima consultazione dei siti web: 20 maggio 2018.

I dati di aNobii ed i loro contesti

Sul tema del social reading esiste ormai una letteratura di riferimento ampia ed articolata, che rende disponibili una serie di metodi e strumenti di analisi e di comprensione, elaborati a partire da diversi punti di vista disciplinari (Social reading 2013). La categorizzazione delle diverse tipologie di social reading proposta da Bob Stein, fondatore dell'Institute for The Future of the Book (http://www.futureofthebook.org/) è spesso utilizzata come una sorta di snodo iniziale per accostarsi preliminarmente a questi argomenti (Stein 2018, Fig. 2). Ci limitiamo qui a segnalare che Stein è ben consapevole della natura schematizzata e semplificata delle categorie proposte, e lo dichiara in modo esplicito nell'*Introduction*: «I've opted instead not to address subtle nuances in the hope that drawing sharper lines will encourage a more vigorous discussion» (Stein 2018).

CATEGORY 1 **informal face-to-face discussion**	Offline	Synchronous	Informal	Ephemeral
CATEGORY 2 **informal online discussion**	Online	Asynchronous	Informal	Persistent
CATEGORY 3 **formal face-to-face discussion**	Offline	Synchronous	Formal	Ephemeral
CATEGORY 4 **formal discussion IN the margins**	Online	Synchronous or Asynchronous	Formal	Persistent

Figura 2. Matrice del social reading. Fonte: http://futureofthebook.org/social-reading/matrix/index.html.

Un altro tentativo di sistematizzazione del campo del social reading è stato effettuato con il progetto *Social Reading in E-books and Libraries*, promosso dal Finnish Strategic Centre for Science, Technology and Innovation in the field of ICT (Heikkilä, Laine e Nurmi 2013; Heikkilä 2013; Di Giammarco 2016). Le tipologie di lettura praticabili sulle piattaforme di social reading sono classificate secondo una griglia che a partire dalla «lettura per me stesso», tipica delle «Book 1.0 Actions», approda alla «lettura collettiva», nella cornice delle «Book 2.0 Actions». Inoltre vengono ridotte a denominatore comune le azioni consentite dalle piattaforme, che oltre a differenziarsi per i diversi stili di lettura, si manifestano in attività connesse alla archiviazione dei propri libri (Scaffale), alla annotazione, alla valutazione o rating ed infine alla recensione (Heikkilä 2013, p. 52).

Passando ad un livello più specifico possiamo poi dar conto di opere che, secondo modalità diverse rispetto a quelle previste in questo contributo, descrivono funzionalità di specifiche piattaforme di social reading, come aNobii, Goodreads, Wattpad, Zazie (Aiello et al. 2010; Crippa e Akabochi de Carvalho 2013; Nakamura 2013; Franzoni, Poggioni e Zollo 2013; Dimitrov et al. 2015; Maity, Panigrahi e Mukherjee 2017; Burns 2017; Ramdarshan Bold 2018; Zanni 2018). La prospettiva di lavoro presentata in questa sede si muove invece secondo una linea che aspira ad essere, nello stesso tempo, microanalitica e panoramica. Con ciò si vuol affermare che il lavoro diretto sui dati, risultato di azioni dei lettori effettuate secondo le funzionalità delle diverse piattaforme, può risultare utile per due ordini principali di motivi. Il primo, più limitato e specifico, riguarda la descrizione e rappresentazione di ciò che accade all'interno delle piattaforme, e già in tal modo riesce a mettere in evidenza quei tratti delle esperienze di lettura consentiti dalla struttura degli ambienti entro i quali le interazioni vengono effettuate; a questo primo esito, ed applicando la network analysis, si aggiunge poi qualcosa di ulteriore, che non è di fatto *impresso* direttamente nella struttura informativa dei database.

 doi:10.4399/97888255181531

La natura sociale della lettura

Dalle considerazioni fin qui proposte emergono numerose possibili implicazioni; qui se ne sviluppano rapidamente solo alcune, riferite alla natura in senso lato "sociale" della lettura. Il filologo Jesper Svenbro, nel suo interessante saggio pubblicato in *Storia della lettura nel mondo occidentale*, ha individuato e censito i principali verbi utilizzati nella Grecia arcaica e classica per denotare e connotare l'atto del leggere, in una fase storica caratterizzata dall'uso prevalente della lettura ad alta voce (Svenbro 1995). Il primo dei verbi preso in esame è *nemein*, il cui significato base è "distribuire". Con questa base semantica, evidentemente, si intendeva fare riferimento alla "distribuzione" del testo sonorizzato da parte di chi ne effettuava la lettura ad alta voce. Una delle prime forme verbali utilizzate per riferirsi all'atto del leggere reca dunque in sé le tracce evidenti del contesto sociale e relazionale entro il quale la lettura veniva praticata, nel quale le forme sonore del testo, rese percepibili attraverso la voce, producevano i loro effetti di significazione; e tutto ciò avendo, a monte, la complessa fase, antropologica e cognitiva, che aveva gradualmente condotto all'"addomesticamento del pensiero selvaggio", secondo le linee di spiegazione tracciate da Jack Goody e Walter J. Ong (Goody 1981; Ong 1986; Ong 1989). Questa natura sociale della lettura, documentata già nella sua fase originaria e fondativa, ha sempre continuato ad essere presente, anche quando l'interiorizzazione dell'atto del leggere, divenuto prima "borbottante" e poi silenzioso, ha reso meno evidente la sua natura relazionale, accentuandone al contrario la dimensione privata ed intima. La lettura sociale in ambiente digitale, lavorando attraverso segni di nuovo esteriorizzati nelle interfacce, e percepiti in primo luogo attraverso la vista, rende di nuovo più spiccatamente esplicita la dimensione sociale della lettura, tuttavia sempre presente nella sua più che millenaria storia.

L'analisi della rete dei libri

In precedenti pubblicazioni (Faggiolani, Verna e Vivarelli 2017) abbiamo introdotto alcuni concetti teorici, attraverso i quali abbiamo affrontato l'analisi del dataset aNobii, e alcuni primi risultati, tra cui la rete degli utenti aNobii e la rete dei libri già citata nei paragrafi precedenti. L'approccio che abbiamo scelto di adottare in quello studio è stato di tipo olistico: abbiamo considerato i dati della piattaforma aNobii non come informazioni esplicite, ma come tracce non strutturate delle attività degli utenti che interagiscono sul social network. In particolare abbiamo considerato i commenti e le recensioni che gli utenti hanno scritto per i diversi libri, e l'obiettivo consisteva nel fare emergere le relazioni latenti e non esplicite tra lettori, parole e libri.

Per interpretare la complessità delle tracce digitali determinate dalle attività degli utenti abbiamo adottato il formalismo delle reti. Abbiamo scelto la scienza delle reti come strumento per rappresentare le informazioni disponibili, analizzarne le proprietà e i fenomeni emergenti, ritenendo il modello delle reti appropriato per la sua flessibilità, per la capacità di descrivere sistemi complessi e soprattutto per le caratteristiche specifiche del dato da analizzare. Le reti basano le loro proprietà matematiche e formali sulla teoria dei grafi. I grafi sono oggetti discreti che permettono di schematizzare una grande varietà di fenomeni e di processi, e di consentirne l'analisi quantitativa e lo studio attraverso funzioni e algoritmi. In sintesi un grafo è definito da un insieme di nodi e un insieme di archi che uniscono coppie di nodi. La teoria dei grafi definisce e indaga numerose loro proprietà, quali ad esempio la densità, la completezza e la modularità, e fornisce strumenti via via più complessi per descrivere il grafo e comprenderne le caratteristiche (Trudeau 1993).

Sulla base della teoria dei grafi, la recente disciplina della network science (o scienza delle reti) studia nel loro insieme le più diverse tipologie di fenomeni fisici, biologici e sociali (National Research Council 2005). Le reti sono uno strumento adatto a descrivere sistemi complessi in cui intervengono numerosi elementi che seguono regole non coordinate centralmente (Caldarelli e Catanzaro 2007). I "sistemi complessi", a loro volta, sono caratterizzati da fenomeni il cui comportamento non può essere previsto considerando solamente i singoli elementi che lo costituiscono. Rappresentati come reti, cioè come insiemi di nodi e archi, i fenomeni possono essere compresi attraverso la scienza delle reti che fornisce regole e proprietà per analizzarli.

Nel lavoro con i dati raccolti dalla piattaforma aNobii abbiamo utilizzato le reti per descrivere i comportamenti di lettura (Verna 2016; Faggiolani e Verna 2016). Le reti consentono a ogni frammento di informazione di relazionarsi agli altri in base a come è stato prodotto. Abbiamo generato una prima rete onnicomprensiva, che abbiamo denominato "rete plain" ("piatta"), in cui

non esiste ancora una gerarchia di relazioni; i suoi nodi sono di diverso tipo: libro, autore, commentatore, commento, testo, concetto, parola. Questa prima rete rappresenta i frammenti e gli atomi del dato sorgente. Tutti questi frammenti di informazione definiti dai dati che corrispondono a ciascun commento formano una rete molto estesa; al crescere del numero di oggetti che la alimentano, la rete andrà ad assumere una propria struttura e i suoi nodi avranno ruoli e dinamiche proprie.

Applicando algoritmi di network analysis abbiamo calcolato per ogni nodo-libro della rete plain il "peso" (importanza calcolata) delle relazioni verso ciascun altro nodo-libro presente sulla rete.

Nella Fig. 3 vediamo una esemplificazione degli elementi che contribuiscono al calcolo della forza della relazione tra il nodo libro A e il nodo libro B.

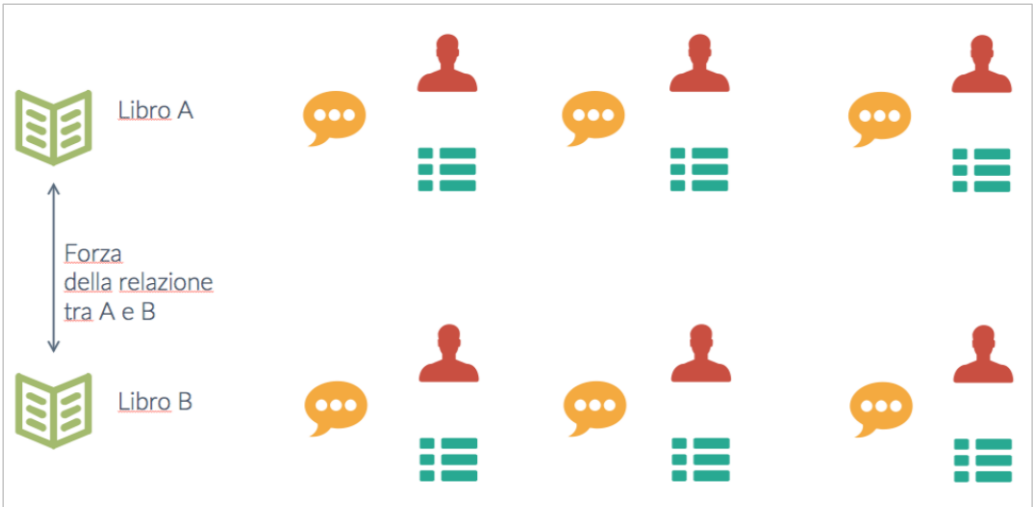

Figura 3. Visualizzazione delle relazioni tra libro A e libro B.

Disponendo di queste nuove relazioni tra i nodi-libro abbiamo costruito la rete dei libri, rappresentata nella Fig. 1. L'analisi delle proprietà della rete dei libri così ottenuta permette di raggruppare in modo inedito i libri, sulla base di come vengono letti e commentati dai lettori. È possibile identificare i gruppi più coesi, quelli più centrali, quelli più connessi e quelli più periferici. Da una parte otteniamo alcune conferme, come per esempio la naturale emersione di piccole comunità di libri molto specifici accumunati dal genere (es. graphic novel), dall'altra scopriamo gerarchie di relazioni che costituiscono interessanti e nuove correlazioni tra opere e autori.

Come accennato, in questo caso abbiamo utilizzato un dato non strutturato, cioè non abbiamo utilizzato metadati formalmente definiti e le relazioni esplicite ad essi correlate. Le informazioni relative all'oggetto "libro" sono state raccolte così come erano rappresentate nella base dati di aNobii, molto spesso in forma non rigorosa e con pochissimi attributi di natura in senso stretto catalografica[2].

Nel voler costruire una nuova rete per rappresentare una mappa delle collane e una mappa degli editori, abbiamo provveduto a normalizzare e aumentare il dato descrittivo dei libri per attribuire a ciascun oggetto un corredo di informazione più completo. Per fare ciò abbiamo attinto da fonti esterne, quali un database dei prefissi editori assegnati dalla codifica ISBN e le informazioni fornite dall'OPAC del Servizio Bibliotecario Nazionale (http://opac.sbn.it/opacsbn/opac/iccu/free.jsp) interrogato attraverso il protocollo Z35.90. Attraverso questo secondo repertorio abbiamo raccolto, quando disponibile, l'informazione relativa alla collana di cui il libro fa parte.

A valle di un processo di cura e pulizia del dato è stato possibile normalizzare una buona parte dei nodi-libro corredandoli con le corrette codifiche e riferimenti a collana ed editore.

Partendo dalla rete dei libri, costruita in precedenza, e sulla base dei nuovi attributi, abbiamo affrontato un processo di aggregazione dei singoli nodi-libro in nuovi macro-nodi che rappresentano la collana a cui ciascun libro appartiene. Tale processo è stato predisposto con il

[2] Nel dataset di dati estratti da Anobii la tabella relativa ai "libri" è popolata in modo approssimativo con le informazioni inserite dagli utenti, e per questo motivo accade quindi di trovare lo stesso libro con indicazioni circa editore, autore, talvolta anche titolo e sottotitolo non uniformi.

 doi:10.4399/97888255181531

fine di ottenere nuove relazioni (archi) pesate che collegano questi nodi-collana. I nodi collana sono collegati tra loro da archi-relazioni la cui intensità (il cui peso) riflette la forza delle relazioni esistenti tra i libri che appartengono a ciascuna collana.

Vediamo nella Fig. 4 una rappresentazione della rete delle collane generata con Gephi, un software open-source e free di network analysis. La disposizione dei nodi sul piano è stata ottenuta a seguito di numerose iterazioni e configurazioni dell'algoritmo di network layout ForceAtlas2 (Jacomy et al. 2014). Come indicazione generale per la lettura del grafico, consideriamo che:

- la dimensione del nodo è proporzionale al numero di libri afferenti alla collana;
- il colore del nodo rappresenta la classe a cui il nodo appartiene; la classe è un insieme di nodi maggiormente connessi tra loro;
- la posizione dei nodi nel piano è scelta da algoritmi di layout di rete che cercano di disporli secondo un criterio di equilibrio tra le forze di attrazione date dai numerosi legami che ciascuno nodo ha con gli altri nodi, definito attraverso l'uso di ForceAtlas2. La vicinanza di due nodi sul piano non è indicativa di un legame più o meno forte tra i nodi stessi, ma del miglior equilibrio tra tutti i legami.

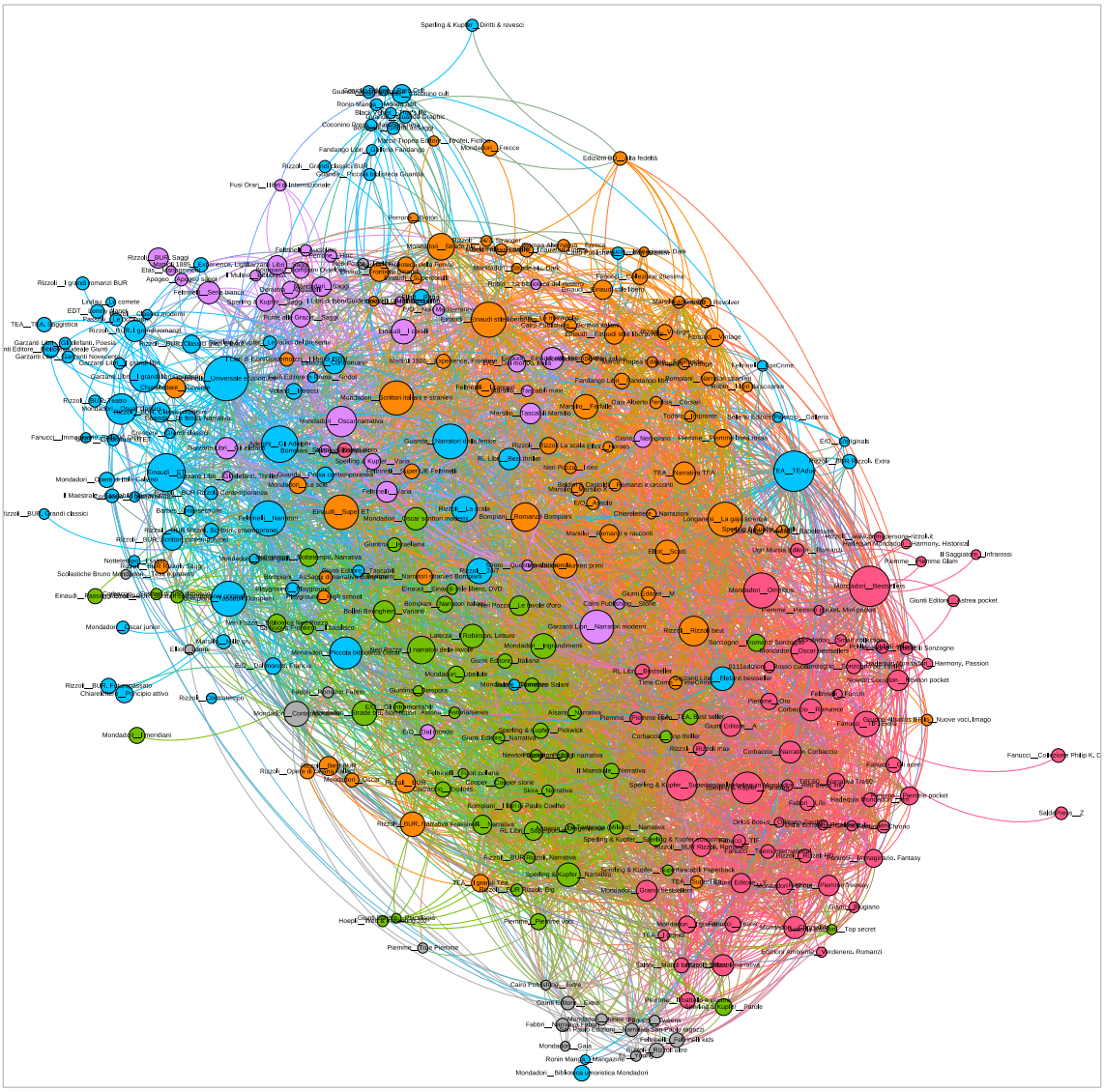

Figura 4. Visualizzazione della rete delle collane.

Seguendo un processo di aggregazione simile a quello che ha condotto alla creazione della rete delle collane, abbiamo aggregato i libri sulla base dell'editore che li ha pubblicati. Ovvero, partendo dalla rete dei libri, abbiamo costruito una nuova rete, la rete degli editori, in cui ogni nodo rappresenta un editore. In questo caso abbiamo aggregato i nodi-libro in nuovi nodi-editore.

La dimensione di ogni nodo-editore è proporzionale al numero dei suoi libri presenti in aNobii. Un nodo-editore è collegato a un altro nodo-editore da una relazione pesata che sintetizza i legami istituiti tra i nodi-libro dei due editori considerati. Nella Fig. 5 proponiamo una rappresentazione della rete degli editori, esito del processo di aggregazione eseguito a partire dalla rete dei libri.

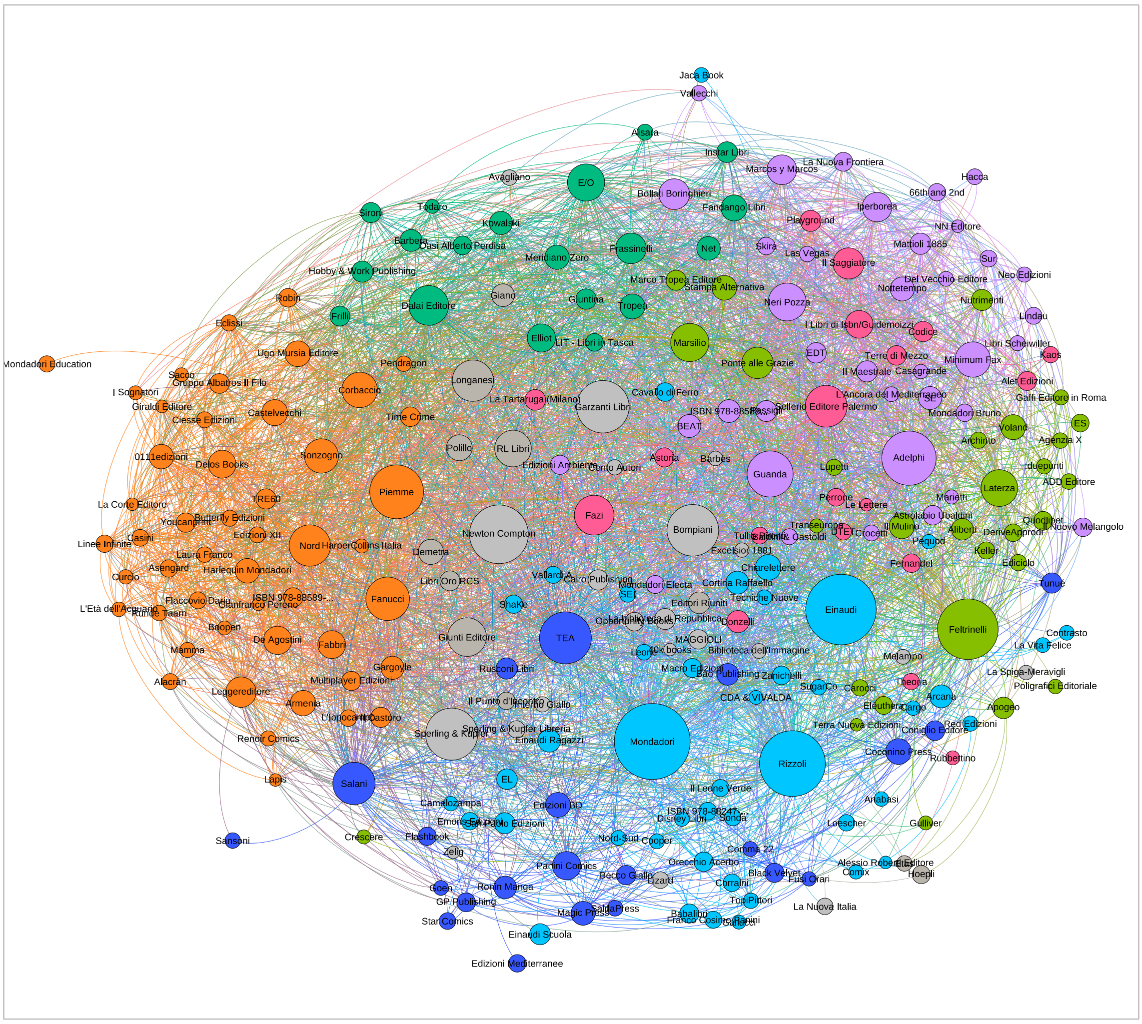

Figura 5. Visualizzazione della rete degli editori.

Come nelle visualizzazioni delle reti precedenti:

- la dimensione del nodo è proporzionale al numero di libri considerati per quell'editore;
- il colore del nodo rappresenta la classe a cui il nodo appartiene;
- la posizione del nodo nel piano è determinata dall'algoritmo di network layout ForceAtlas2. La vicinanza di due nodi sul piano non è indicativa di un legame più o meno forte tra i nodi stessi, ma del miglior equilibrio calcolato tra tutti i legami.

La rete degli editori è composta da circa 300 nodi con un degree medio di 33 e 10 classi di modularità. Nelle tabelle che seguono è registrata una selezione dei nodi principali per ciascuna classe.

 doi:10.4399/97888255181531

Classe 9

Editori	Libri
Adelphi	1528
Guanda	892
Neri Pozza	466
Minimum Fax	375
Bollati Boringhieri	231

Classe 8

Editori	Libri
Mondadori	7030
Einaudi	4153
Rizzoli	3082
Chiarelettere	118
Arcana	76

Classe 7

Editori	Libri
Piemme	1486
Fanucci	740
Nord	626
Corbaccio	380
Sonzogno	336

Classe 6

Editori	Libri
Feltrinelli	2248
Marsilio	480
Laterza	423
Ponte alle Grazie	250
Il Mulino	111

Classe 5

Editori	Libri
Longanesi	623
Giunti Editore	504
RL Libri	421
Polillo	124
Demetra	111

Classe 4

Editori	Libri
Dalai Editore	551
E/O	439
Frassinelli	229
Elliot	171
Fandango Libri	125

Classe 3

Editori	Libri
Sellerio	645
Fazi	548
Il Saggiatore	245
I Libri di Isbn	160
Donzelli	68

Classe 2

Editori	Libri
Sperling & Kupfer	1339
Bompiani	1291
Cairo Publishing	47
Sperling & Kupfer Libreria	23
Lizard	20

Classe 1

Editori	Libri
Newton Compton	2005
Garzanti Libri	1370
La biblioteca di Repubblica	53
Maggioli	39
Opportunity Books	32

Classe 0

Editori	Libri
TEA	1324
Salani	675
Panini Comics	173
Fandango/Coconino Press	120
Edizioni BD	107

Il profilo emergente degli editori attraverso le parole dei lettori

Prima di continuare questo percorso che dalla rete degli editori – che possiamo considerare un primo livello di elaborazione per una ipotesi integrativa di segmentazione – porta alle parole dei lettori, riteniamo sia interessante aprire una brevissima parentesi rispetto alla diversa “visione” che del sistema editoria emerge attraverso questo tipo di analisi. Cosa è possibile conoscere in più o di diverso rispetto a quanto conosciamo oggi?

La fonte alla quale facciamo riferimento non può che essere quella rappresentata dall'indagine Istat sulla produzione libraria che ogni anno dal 1951 (attraverso interviste a tutte le case editrici italiane e agli altri enti sia pubblici che privati che svolgono attività editoriale) raccoglie dati statistici che consentono di descrivere la quantità e le principali caratteristiche dei libri pubblicati nel corso dell'anno[3]. La Fig. 6 riporta la tabella pubblicata da Istat a dicembre 2017 e relativa all'anno 2016.

TIPI DI EDITORE	Editori attivi		Opere pubblicate		Copie stampate		Numero medio di opere pubblicate per editore	Numero medio di copie stampate per editore
	N.	%	N.	%	(in migliaia)	%		
Piccoli editori	825	54,8	3.380	5,5	3.536	2,7	4,1	4.286
Medi editori	476	31,6	11.272	18,4	14.809	11,5	23,7	31.111
Grandi editori	204	13,6	46.536	76,1	110.481	85,8	228,1	541.574
Totale	**1.505**	**100,0**	**61.188**	**100,0**	**128.825**	**100,0**	**40,7**	**85.599**

Figura 6. Editori attivi, opere pubblicate e copie stampate per tipo di editore (Fonte: Produzione Libraria, Istat, 2016)[4]. I valori si riferiscono agli editori "attivi", ovvero quelli che hanno pubblicato almeno un'opera libraria nell'anno 2016.

Ad ogni casa editrice si chiede la descrizione di tutte le opere librarie pubblicate, in termini di numero di volumi, genere, materia trattata, numero di pagine, tiratura, presenza o meno di una versione e-book ecc.[5] Assistiamo, dunque, ad una segmentazione del mercato editoriale a partire dalla materialità dell'oggetto libro (numero di libri, numero di copie, genere, materia trattata), dove il lettore, la sua percezione, i suoi concreti atti di lettura non vengono presi in esame. Ciò consente di definire la visione che ne emerge "bidimensionale", espressione con la quale intendiamo porre l'attenzione in particolare sull'assenza di legami e relazioni entro il contesto dato, che rappresentano un possibile valore aggiunto che l'approccio basato sulle metriche di rete evidenzia.

Lo spazio a disposizione e le diverse finalità di questo articolo non ci consentono un approfondimento in questa direzione, ma è utile quanto meno rilevare l'opportunità di una integrazione di questi diversi strumenti e approcci alla segmentazione prospettica del mercato editoriale. Come si dirà meglio nel paragrafo conclusivo, i nostri studi e l'approccio che proponiamo vogliono promuovere quando possibile l'integrazione/triangolazione metodologica e l'utilizzo di fonti di dati diverse. Proprio a tal fine, la nostra analisi procede con l'applicazione delle metriche dell'analisi automatica dei testi – AAT[6] alle recensioni scritte dai lettori su aNobii. Per

[3] Istat effettua con cadenza annuale l'Indagine sulla produzione libraria, una rilevazione censuaria (su tutte le case editrici e gli altri enti che svolgono attività editoriale) con l'obiettivo di descrivere le principali caratteristiche della produzione di libri nel nostro Paese. L'indagine si rivolge a circa 2.000 unità, registrate in un archivio informatizzato degli editori che viene aggiornato annualmente da Istat. L'intervista alle case editrici e enti che svolgono attività editoriale viene effettuata con un questionario online auto-compilato dai rispondenti senza l'intervento dell'intervistatore. Il 27 dicembre 2017 sono usciti i dati relativi all'anno 2016. La situazione descritta è la seguente: oltre l'86% dei circa 1.500 editori attivi pubblica non più di 50 titoli all'anno; oltre la metà (54,8%) sono "piccoli editori", che producono al più 10 opere in un anno, e il 31,6% sono "medi" editori, che producono in un anno da 11 a 50 opere. I "grandi editori", con una produzione libraria superiore alle 50 opere annue, rappresentano il 13,6% degli operatori attivi nel settore e pubblicano più di tre quarti (76,1%) dei titoli sul mercato, producendo quasi l'86% delle copie stampate. Si veda Produzione e lettura di libri in Italia 2016, su https://www.istat.it/it/files/2017/12/ReportEditoriaLettura.pdf.

[4] Ivi, p. 1.

[5] Il questionario propone anche dei quesiti sulla percezione che i lettori hanno degli e-book e sulla quota di vendita di prodotti digitali. Le domande alle quali si fa riferimento sono così formulate nell'ultimo questionario in corso di somministrazione: "A suo parere, quali sono le caratteristiche degli e-book maggiormente apprezzate dal pubblico nel nostro paese?" e "Quali sono i fattori che tendono ad ostacolare la diffusione degli e-book in Italia?". Il questionario è scaricabile da https://www.istat.it/it/archivio/6899.

[6] Il trattamento automatico dei testi secondo un approccio di tipo metrico (analisi automatica del testo - AAT), effettuata attraverso software dedicati con l'obiettivo di rappresentare il contenuto dei testi oggetto di analisi e di estrarre informazioni di interesse attraverso misure quantitative, è l'approccio necessario quando si ha a disposizione una imponente molte di dati testuali per i quali non è possibile applicare

 doi:10.4399/97888255181531

chiarezza l'idea non è quella di confrontare due diversi approcci alla conoscenza e la validità delle loro metriche ma al contrario restituire la complessità di un sistema che, come detto in precedenza, non può essere descritto da un punto di vista unico e soltanto attraverso la somma delle sue parti.

Abbiamo così costruito 10 diversi *corpora* testuali[7], uno per ciascuna classe individuata attraverso l'analisi della rete descritta nel paragrafo precedente. Dopo una prima normalizzazione e lemmatizzazione dei testi[8] abbiamo realizzato una descrizione statistica dei *corpora*, così da poter comparare le classi dal punto di vista quantitativo: per numero di testi/recensioni (UCI - unità di contesto iniziali) (Fig. 7), numero di forme grafiche (*type*) (Fig. 8) e numero di occorrenze (*token*) (Fig. 9). Osserviamo come la classe 8 – nella quale si collocano editori come Rizzoli, Mondadori, Einaudi – risulti essere ovviamente decisamente più corposa delle altre soprattutto in termini di numero di recensioni scritte (UCI).

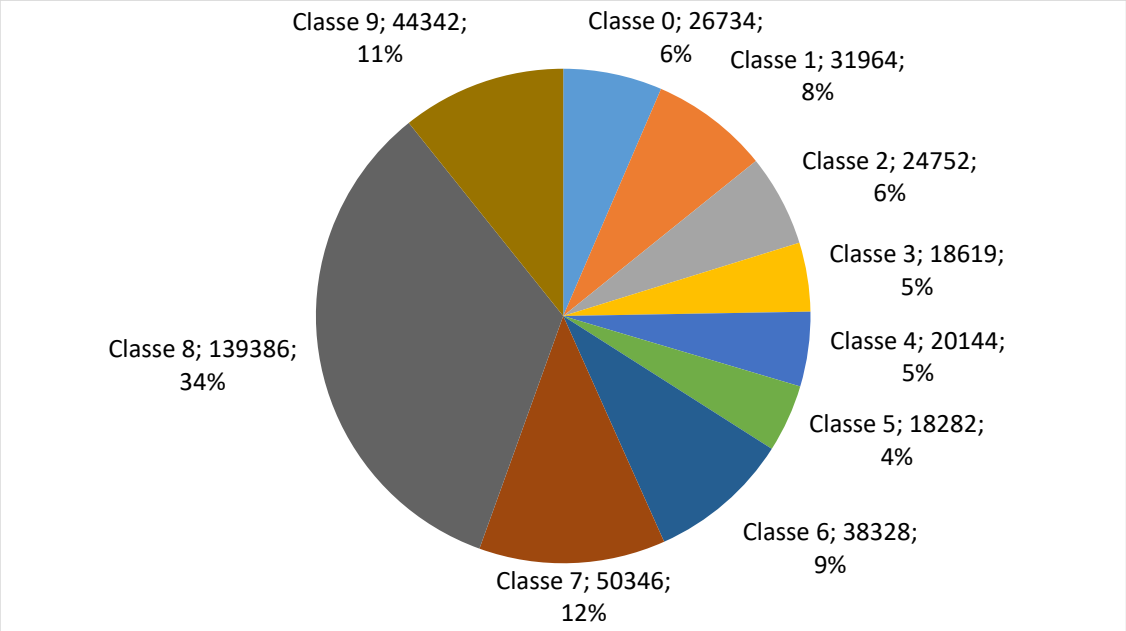

Figura 7. Percentuale di recensioni (UCI) per classe di editori.

analisi del contenuto di tipo interpretativo. Si tratta di un complesso ambito di studi al quale sono ascrivibili le tecniche di estrazione delle informazioni da materiali espressi in linguaggio naturale – Information Retrieval (IR) e Information Extraction (IE) – utili per avere accesso alla conoscenza nascosta dentro le tracce digitali lasciate dagli utenti, per estrarre e visualizzare informazioni rilevanti. Si veda Faggiolani, Verna e Vivarelli (2017) e Bolasco (2013). Tra i software di maggior rilievo possiamo segnalare TaLTaC2, Alceste, T-LAB, IRaMuTeQ, Lexico3: Cfr. (Giuliano 2013). Le esemplificazioni che seguono sono frutto di elaborazioni condotte con IRaMuTeQ (http://www.iramuteq.org).

[7] Al momento dell'estrazione dei dati (giugno 2016) in aNobii erano presenti 2.552.955 recensioni, di cui 1.740.394 in italiano, per un totale di 80 milioni di parole circa. In questo caso abbiamo estratto (con campionamento casuale) il 10% delle recensioni di ogni libro per ciascun editore di ogni classe.

[8] Per *corpus* si intende una collezione di testi o frammenti, che chiameremo unità di contesto iniziali (UCI) fra loro coerenti e pertinenti per essere studiate sotto un qualche punto di vista: in questo caso le recensioni. I testi che costituiscono il *corpus* devono essere prodotti in condizioni di enunciazione simili e devono avere caratteristiche confrontabili in merito alla ricchezza del vocabolario e alla lunghezza. Chiamiamo le parole del *corpus* 'forma grafiche' – sequenze di caratteri delimitate da due separatori – intese come unità elementari del testo (*type*). Esse sono l'unità statistica sulla quale vengono operate le analisi. Il numero di volte in cui il *type* appare nel *corpus* determina le sue *occorrenze* (*tokens*). Il *lemma* è costituito dalla forma corrispondente all'entrata del termine nel dizionario e rappresenta tutte le flessioni con cui quell'unità lessicale può presentarsi nel discorso. Ad esempio, le occorrenze <leggevo> e <ho letto> sono due forme grafiche distinte, due flessioni appartenenti allo stesso *lemma*: <leggere>.

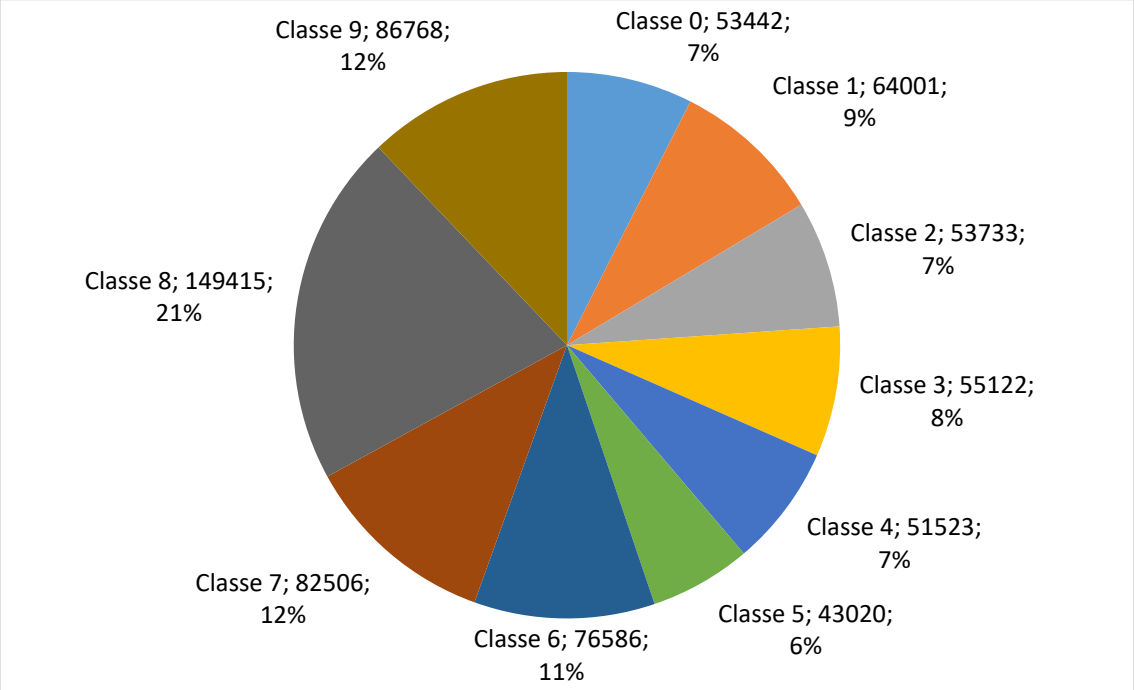

Figura 8. Percentuale di *type* (forme grafiche) per classe di editori.

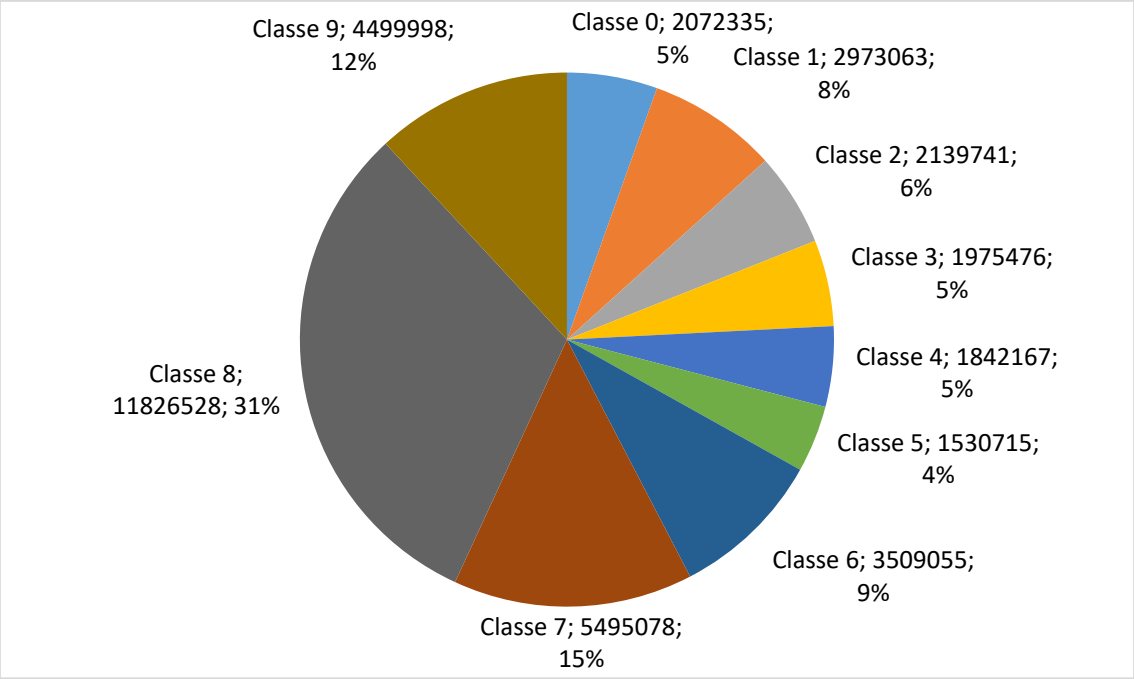

Figura 9. Percentuale di *token* (occorrenze) per classe di editori.

Ogni analisi testuale basata su criteri statistici assegna alla frequenza delle parole un ruolo estremamente importante, anche se non sempre questo costituisce un criterio decisivo di estrazione di conoscenza, come dimostra la nuvola di parole in Fig. 10**,** che rappresenta le 'parole tema' della classe 0 (presa in considerazione solo a titolo esemplificativo). Anche le parole incontrate poche volte o una sola (*hapax*) – perfino le parole assenti, talvolta – possono avere un valore rilevante.

 doi:10.4399/97888255181531

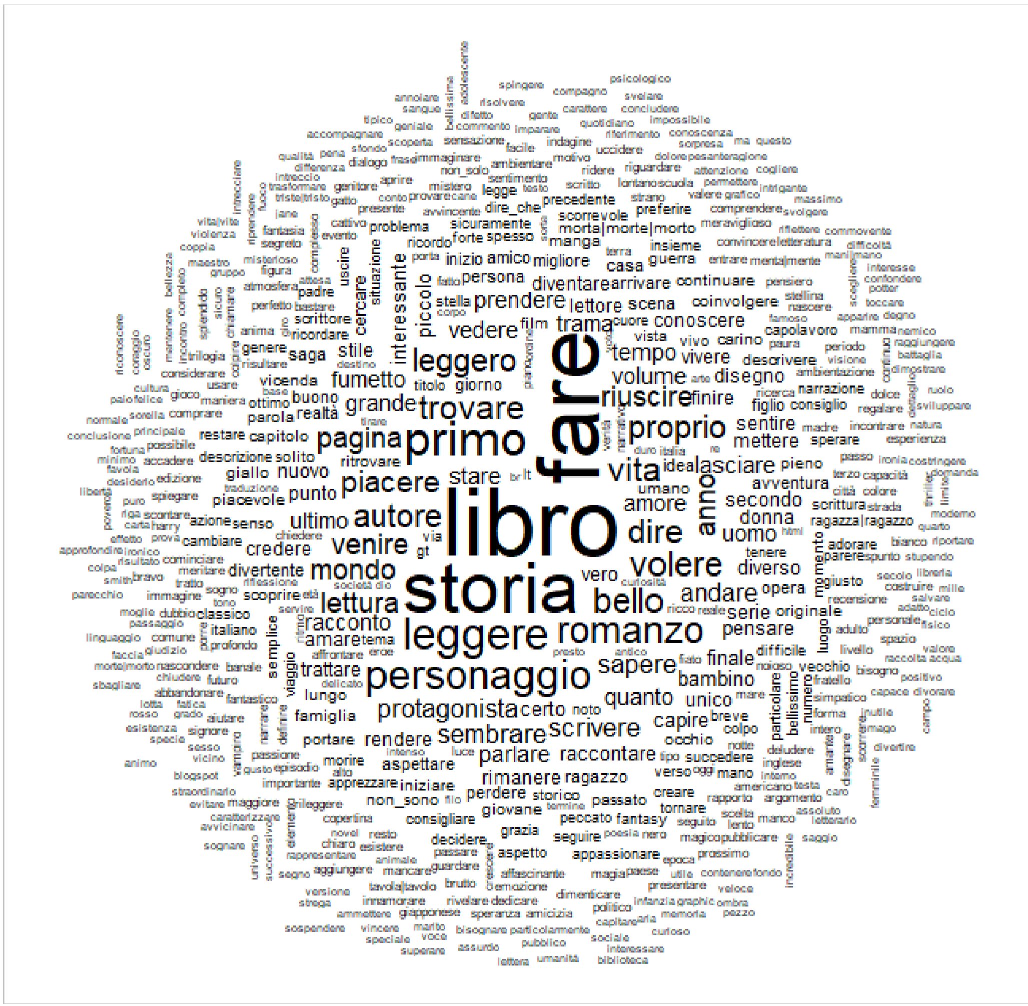

Figura 10. Nuvola di parole della Classe 0 (forme grafiche con almeno 50 occorrenze).

Procediamo dunque con la costruzione dei 10 vocabolari[9] (uno per ogni classe) e con l'analisi delle "parole piene" e delle "parole vuote": le prime fanno riferimento ai termini che hanno un senso in sé e comprendono le forme verbali, i nomi, gli aggettivi e gli avverbi; le seconde sono tra le parole più frequenti e non sono portatrici di significato autonomo. Tra queste gli articoli, le parole finalizzate a funzioni grammaticali ecc.

A partire dai vocabolari dei 10 *corpora* abbiamo osservato le prime 100 parole piene in ordine decrescente (per numero di occorrenze): la maggior parte sono abbastanza scontate per un *corpus* testuale costituito da recensioni di libri e sono trasversalmente presenti: "libro", "leggere", "romanzo", "personaggio", "storia". Altre forme si distinguono per essere presenti in modo specifico solo in alcune classi: ad esempio le forme "fumetto", "manga", "disegno" nel vocabolario della classe 0, dove troviamo editori come TEA e Salani.

Per ciascuna classe è stata condotta una analisi di specificità[10] per osservare eventuali differenze significative nei lessici caratterizzanti le recensioni di libri editi da editori appartenenti alla medesima classe. Senza entrare nel merito dei singoli risultati ottenuti dalle comparazioni è interessante notare come alcuni editori abbiano nel loro *brand* un elemento decisivo di specificità

[9] Il "vocabolario" è definibile come l'insieme delle parole diverse (o forme grafiche) che compongono il corpus e può essere utilizzato al fine di selezionare una serie di parole significative che consentono di interpretarne il contenuto.

[10] Si definisce "specifica" in un testo ogni parola o espressione sovra/sotto utilizzata rispetto ad una media attesa. Ogni specificità positiva (sovra-utilizzo) di una parola o di una espressione equivale ad un uso superiore a quello atteso. Ogni specificità negativa di un termine, equivale ad un sotto-utilizzo (o rarità del termine, fino ad arrivare anche alla sua totale assenza) rispetto al valore "atteso". Questo scarto è valutato in termini probabilistici.

– questo vale per editori come Laterza, Einaudi, Adelphi – e altri per i quali la propria specificità nelle parole dei lettori si riflette in personaggi, generi, ambientazioni o collane – per esempio nella classe 3 per Sellerio sono specifiche le forme “Camilleri”, “Montalbano”, “Sicilia”; nella classe 1 per Newton Compton sono specifiche le forme “Vampiro”, “Saga”, “Trilogia”; “Horror”; nella classe 8 per Rizzoli le forme “BUR”, “Fallaci”, “Maraini”. Queste possono essere informazioni interessanti sia in un’ottica di marketing strategico, per esempio nella definizione del target di riferimento, nell’analisi del posizionamento percepito e della propria *brand image*, ma anche nella direzione del marketing tattico, per esempio per le decisioni che riguardano la comunicazione nell’ambito del marketing mix.

Procedendo nella esplorazione dei *corpora* ad un maggiore livello di dettaglio, per ogni classe e poi per editore è stata realizzata una classificazione gerarchica discendente che consente di osservare i “mondi lessicali” soggiacenti, ovvero le classi lessicali in cui ricorrono, con maggiore frequenza, alcune espressioni che sono, quindi, individuate come tipiche delle porzioni di testo analizzate[11]. È come rispondere alla domanda: quali temi vengono affrontati dai lettori di un certo editore trasversalmente ai titoli letti?[12]

Possiamo considerare queste prime analisi descritte come assaggi o esplorazioni del *corpus* animate da un obiettivo di conoscenza che potremmo definire induttiva, dove “si trova ciò che ci trova e non ciò che si cerca”; possiamo accostarci però all’analisi anche con un obiettivo di conferma di ipotesi formulate *a priori* attraverso un approccio più di carattere deduttivo, in cui cioè “si trova ciò che ci cerca e non ciò che si trova”.

Abbiamo selezionato una serie di parole che riteniamo significative rispetto alle condizioni di enunciazione che caratterizzano i testi oggetto di analisi: la fabbricazione del lettore, ossia cosa la lettura del testo genere in termini di emozione, riflessione, arricchimento. Ogni nuova lettura dipende dalle esperienze precedenti del lettore, dal *set* e dal *setting* della lettura stessa[13].

Di seguito, nelle Figg. 11, 12 e 13 si riportano alcuni dei risultati ottenuti, da intendersi per la loro portata esemplificativa rispetto ad un approccio possibile e al tipo di conoscenza generata. Abbiamo creato un nuovo *corpus* che include le recensioni scritte sui libri editi dal primo editore di ciascuna classe e lo abbiamo sottoposto ad una analisi di specificità, in questo caso osservando il posizionamento dell’editore rispetto all’uso di quella specifica parola. Osserviamo che la forma “prezioso” (Fig. 11) e “poetico” (Fig. 12) connotano fortemente le recensioni scritte dai lettori di Adelphi e Feltrinelli e sono sotto-rappresentate per tutti gli altri (eccetto qualche eccezione); la forma “rileggere” (Fig. 13) è molto specifica nelle recensioni dei libri Mondadori.

[11] «Noi chiamiamo “mondi lessicali” le impronte lessicali di questi luoghi nell’enunciazione, mondi che sono visualizzati tecnicamente, dal vocabolario specifico delle classi». Max Reinert, *Mondes lexicaus et topoi dans l’approche Alceste, in Mots chiffrés et déchiffrés, eds. Sylvie Mellet y Marc Vuillaume*, Paris: Honoré Champion, 1998, p. 292.

[12] A questo scopo IRaMuTeQ utilizza il metodo ALCESTE – *Analyse des Lexemes Cooccurrents dans les Enoncés Simplifiés d’uni Texte* – che si basa sulla logica della ricerca delle similitudini, rintracciando nel testo la presenza co-occorrente delle stesse forme grafiche (parole o lessemi) (Reinert 1990).

[13] Le espressioni *set* e *setting* si riferiscono rispettivamente «a quell’insieme di attitudini mentali e di atteggiamenti personali che influenzano ciò che facciamo: le nostre aspettative, le nostre precedenti esperienze e conoscenze, il nostro stato d’animo, la nostra relazione con gli altri […]» e a «l’ambiente fisico e la sua adeguatezza rispetto all’attività proposta» (Chambers 2015).

 doi:10.4399/97888255181531

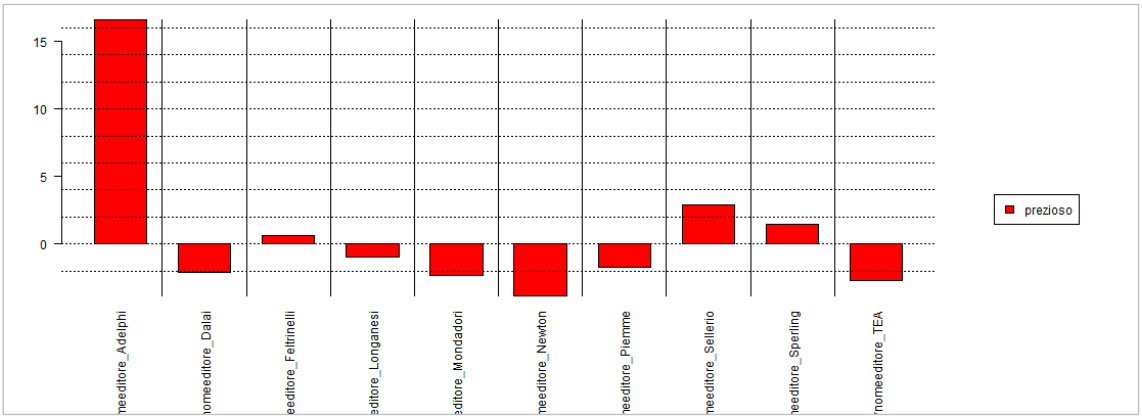

Figura 11. Specificità rispetto alla forma “prezioso”.

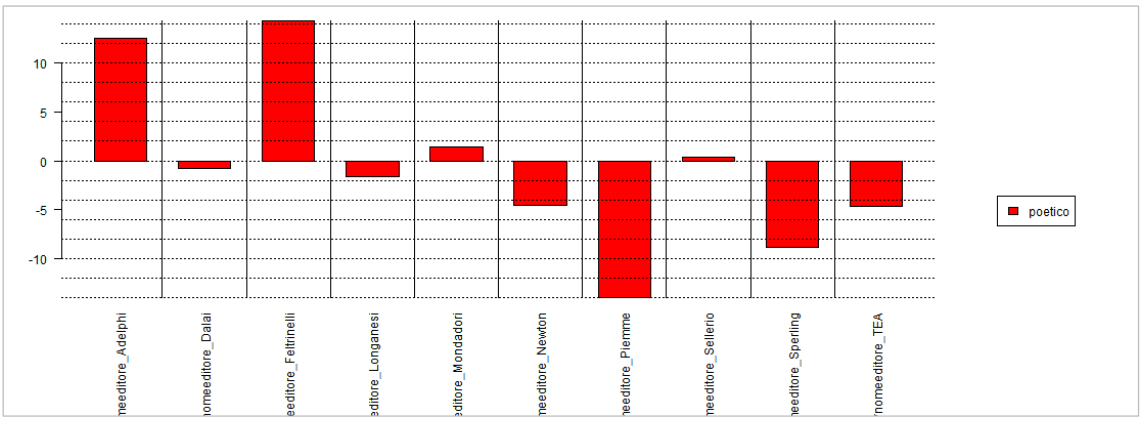

Figura 12. Specificità rispetto alla forma “poetico”.

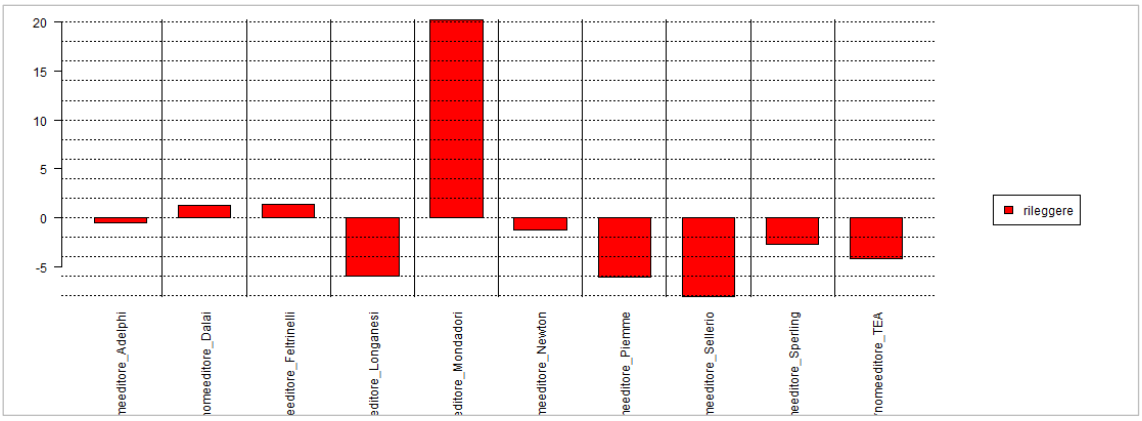

Figura 13. Specificità rispetto alla forma “rileggere”.

Considerazioni di metodo

Se guardiamo ai dati sulla lettura di libri e sulla produzione libraria pubblicati ogni anno da Istat la sensazione è che le cose stiano cambiando molto lentamente e che soprattutto rispetto alla lettura di libri negli ultimi 20 anni sia cambiato poco e niente: il 40% circa degli italiani leggeva almeno un libro l'anno nel 1996, il 40% circa degli italiani ha letto almeno un libro l'anno nel 2016[14]. Eppure in questi venti anni sappiamo che è cambiato quasi tutto: i supporti, le interfacce, il mercato, la testualità, le abitudini delle persone nella partecipazione e fruizione culturale a 360 gradi.

È una sorta di ossimoro questo: la statistica ufficiale ci offre indagini che guardano al passato – per ovvi motivi di confronto in serie storica – e noi sentiamo che tutto intorno sta cambiando ad una velocità difficile da cogliere con gli strumenti tradizionali. Questo approccio metodologico alla conoscenza dei comportamenti culturali è figlio di un paradigma che sentiamo comodo ma che forse non basta più perché ciò di cui abbiamo bisogno oggi è entrare dentro i significati attribuiti dalle persone alle azioni che compiono (Faggiolani 2016).

La lettura è una di queste e sicuramente la percezione del lettore è un tema poco battuto nelle indagini empiriche tradizionali, anche se l'urgenza di un approccio olistico veniva denunciata già tanti anni fa e in diverse sedi:

> «Poiché […] la lettura non è ipotizzabile come travaso o iniezione di informazioni ma, secondo approcci fenomenologici ed ermeneutici, quale incontro lettore-testo, fusione di orizzonti, interrogazione e risposta, la conoscenza sociologica dovrebbe […] farsi carico sia di una analisi del testo (dei testi) letti sia delle modalità di ricezione da parte di coloro che ne attivano i significati. Di fatto assai scarse solo le ricerche di questo tipo […]» (Pagliano 1986).

L'analisi della lettura è una attività complessa, perché la lettura, come già si detto in precedenza, è un sistema complesso (Faggiolani Verna e Vivarelli, 2017). Su questa assunzione si basano le attività e la vocazione interdisciplinare del nostro gruppo di ricerca.

Il nostro progetto di ricerca, i cui ultimi esiti sono stati descritti in queste pagine, ha questa ambizione: individuare punti di forza e criticità di dati "nuovi" sulla lettura di libri tentando di rispondere in estrema sintesi a questa domanda: in che modo i cosiddetti *User Generated Content* - UGC[15] possono contribuire alla conoscenza che abbiamo del mercato e dei comportamenti di lettura?

Siamo consapevoli di essere solo agli inizi ma siamo anche convinti che sia questa la strada giusta.

Sarà auspicabile in futuro un allargamento dell'analisi ad altre basi di dati simili ad aNobii, sulla quale per il momento ci siamo concentrati – GoodReads, Amazon, IBS, ad esempio – e soprattutto un confronto con tutti gli operatori della filiera del libro per orientare in modo più strategico le sollecitazioni a cui sottoporre i dati.

Il valore dei dati in termini conoscitivi dipende anche dal progetto e dal processo di analisi cui vengono sottoposti. Non solo la scelta delle domande, come ovvio, è discriminante rispetto all'avanzamento della conoscenza su certi temi e non su altri ma, determinando anche l'accesso a certi dati e non ad altri, le domande stesse permettono al ricercatore di confrontarsi con questioni importanti anche dal punto di vista metodologico.

[14] Il dato arriva a circa il 60% se consideriamo non solo la lettura "per piacere", ovvero non per motivi strettamente professionali.

[15] Non possiamo infatti trascurare che ben il 30,2% delle persone di 6 anni e più negli ultimi 3 mesi ha pubblicato sul web contenuti di propria creazione: testi, fotografie, musica ecc. Si veda l'indagine Istat, *Aspetti della vita quotidiana*, 2017 – Report "Cittadini, imprese e ICT" che si occupa di fornire il quadro informativo integrato sull'utilizzo delle tecnologie ICT da parte di cittadini e imprese in Italia. Si veda https://www.istat.it/it/archivio/207825.

 doi:10.4399/97888255181531

Bibliografia

Aiello, Luca Maria, Alain Barrat, Ciro Cattuto, Giancarlo Ruffo e Rossano Schifanello. “Link Creation and Profile Alignment in the aNobii Social Network.” *Social Computing (SocialCom)*, *2010 IEEE Second International Conference on Social Computing*, 20-22 Aug. 2010. Disponibile all’URL: http://www.di.unito.it/~aiello/papers/socialcom10.pdf.

Bolasco, Sergio. L’analisi automatica dei testi. Fare ricerca con il text mining. Roma: Carocci, 2013.

Burns, Dylan. “How to Rate a Book: Goodreads, Taste, and Reading in the 21st Century.” *Library Faculty & Staff Presentations*, Paper 106. Disponibile all’URL: https://digitalcommons.usu.edu/libpresent/106.

Caldarelli, Guido e Michele Catanzaro. *A Very Short Introduction to Networks*. Oxford: Oxford University Press, 2007.

Chambers, Aidan. *Il lettore infinito. Educare alla lettura tra ragioni ed emozioni*. Modena: Equilibri, 2015.

Crippa, Giulia e Larissa Akabochi de Carvalho. “A mediação da informação através da comunidade virtual Anobii: um estudo de caso.” *Encontros Bibli: revista electrônica de biblioteconomia e ciência de informação* 17 (2012). doi: 10.5007/1518-2924.2012v17n35p97.

Di Giammarco, Fabio. “Social reading and eBooks.” *Cultura digitale* 2 (dicembre 2016). Disponibile all’URL: http://www.culturadigitale.it/wp/ebook/419/social-reading-and-ebooks/.

Dimitrov, Stefan, Faiyaz Zamal, Andrew Piper e Derek Ruths. “Goodreads versus Amazon: The Effect of Decoupling Book Reviewing and Book Selling.” *Proceedings of the Ninth International AAAI Conference on Web and Social Media* (2015). Disponibile all’URL: https://www.aaai.org/ocs/index.php/ICWSM/ICWSM15/paper/download/10557/10452.

Faggiolani, Chiara. “Morfologia dei dati sulla lettura (di libri).” In *I percorsi della conoscenza. Dialogando con Giovanni Solimine su biblioteche, lettura e società*, a cura di Giovanni Di Domenico, Giovanni Paoloni, Alberto Petrucciani, 169-183. Milano: Editrice Bibliografica, 2016.

Faggiolani, Chiara e Lorenzo Verna. “La lettura sul lettino: primi tentativi di data analysis.” In *Le reti della lettura. Tracce, modelli, pratiche del social reading*, a cura di Chiara Faggiolani e Maurizio Vivarelli, 169-183. Milano: Editrice Bibliografica, 2016.

Faggiolani, Chiara, Lorenzo Verna e Maurizio Vivarelli. “Text mining e network science per analizzare la complessità della lettura. Principi, metodi, esperienze di applicazione.” *JLIS.it* 8.3 (2017): 115-136. doi 10.4403/jlis.it-12414.

Faggiolani, Chiara e Maurizio Vivarelli. “Leggere in rete. La lettura in biblioteca al tempo dei Big Data”. In *Bibliotecari al tempo di Google. Profili, competenze, formazione*, a cura dell’Associazione Biblioteche oggi, 101-126. Milano: Editrice Bibliografica, 2016.

Faggiolani, Chiara e Maurizio Vivarelli (a cura di). *Le reti della lettura. Tracce, modelli pratiche del social reading*. Milano: Editrice Bibliografica, 2017.

Franzoni, Valentina, Valentina Poggioni e Fabiana Zollo. “Automated Classification of Book Blurbs According to the Emotional Tags of the Social Network Zazie.” *First International Workshop on Emotion and Sentiment in Social and Expressive Media: Approaches and Perspectives from AI (ESSEM 2013), AI*IA 2013 Conference, CEUR – WS, Turin* 1096 (2013): 83-94.

Giuliano, Luca. *Il valore delle parole. L'analisi automatica dei testi in Web 2.0*. Roma: Dipartimento di Scienze statistiche, 2013.

Goody, Jack. *L'addomesticamento del pensiero selvaggio*. Milano: Franco Angeli, 1981 (*The Domestication of the Savage Mind*, 1977).

Heikkilä, Harri. "Il social reading incontra l'e-book. Uno sguardo sulla storia del social reading alla luce delle prospettive della nuova lettura digitale." *Biblioteche oggi* 34 (2016): 51-4. doi: 10.3302/0392-8586-201602-051-1.

Heikkilä, Harri, Janne Laine e Olli Nurmi. *D1.3.7.4 Social Reading in E-books and Libraries* (2013). Disponibile all'URL: http://virtual.vtt.fi/virtual/nextmedia/Deliverables2013/D1.3.7.4eReading%20Social%20reading%20in%20e-books%20and%20Libraries.pdf.

Istat. *Produzione e lettura di libri in Italia* (2016). Disponibile all'URL: https://www.istat.it/it/files/2017/12/ReportEditoriaLettura.pdf.

Jacomy, Mathieu, Tommaso Venturini, Sebastien Heymann e Mathieu Bastian. "ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software." *Plos One* (June 10, 2014). doi: 10.1371/journal.pone.0098679.

Maity, Suman Kalyan, Abishek Panigrahi e Animesh Mukherjee. "Book Reading Behavior on Goodreads Can Predict the Amazon Best Sellers." *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* (2017): 451-454.

Nakamura, Lisa. "'Words with Friends': Socially Networked Reading on Goodreads." *PMLA* 128 (2013): 238–43.

National Research Council. *Network Science*. Washington, DC: The National Academies Press. 2005. doi: 10.17226/11516.

Ong, Walter J. *Interfacce della parola*. Bologna: Il Mulino, 1989 (*Interfaces of the Word*, 1977).

Ong, Walter J. *Oralità e scrittura: le tecnologie della parola*. Bologna: Il Mulino, 1986 (*Orality and Literacy: The Technologizing of the Word*, 1982).

Pagliano, Graziella. "La finzione del leggere." In Marino Livolsi, *Almeno un libro. Gli italiani che (non) leggono*. Firenze: La Nuova Italia editrice, 1986.

Ramdarshan Bold, Melanie. "The Return of the Social Author: Negotiating Authority and Influence on Wattpad." *Convergence: The International Journal of Research into New Media Technologies* 24 (2018): 117-36. doi: 10.1177/1354856516654459.

Reinert, Max. "ALCESTE: Une méthodologie d'analyse des données textuelles et une application: Aurélia de Gérard de Nerval". *Bulletin de méthodologie sociologique* 26.1 (1990): 24-54.

Stein, Bob. "A Taxonomy of Social Reading: a proposal" (2018). Disponibile all'URL: http://futureofthebook.org/social-reading/index.html.

Svenbro, Jesper. "La Grecia arcaica e classica: l'invenzione della *lettura* silenziosa". In *Storia della lettura nel mondo occidentale*, a cura di Guglielmo Cavallo e Roger Chartier, 3-56. Roma-Bari: Laterza, 1995.

Trudeau, Richard J. *Introduction to Graph Theory (Corrected, enlarged republication)*. New York: Dover Pub, 1993.

 doi:10.4399/97888255181531

Verna, Lorenzo. “Prospettive di analisi dei dati”. In *Le reti della lettura. Tracce, modelli, pratiche del social reading*, a cura di Chiara Faggiolani e Maurizio Vivarelli, 219-229. Milano: Editrice Bibliografica, 2016.

Zanni, Andrea. “All the Books I’ve Read in the Last Ten Years (2008-2017).” *Medium* (2 aprile 2018). Disponibile all’URL: https://medium.com/@aubreymcfato/all-the-books-ive-read-in-the-last-10-years-2008-2017-b3396416c13.