

# Politecnico di Torino

Corso di Laurea in Ingegneria Gestionale A.a. 2021/2022

# Analisi del network collaborativo nella pubblicazione scientifica tra gli istituti italiani mediante Social Network Analysis

Relatore: Prof. Federico Caviggioli Candidato: Giuseppe Caputo

# **Abstract**

Lo studio portato avanti con la seguente tesi ha l'intento di analizzare la rete di collaborazioni in Italia tra gli istituti, didattici e non, in merito alla pubblicazione di articoli nel mondo accademico. L'approccio analitico della rete e delle sue collaborazioni è stata affrontata con una Social Network Analysis (SNA), strumento di analisi in continua crescita e diffusione in un mondo sempre più interconnesso e che offre una serie di metriche ben definite per poter focalizzare e porre l'accento su diversi aspetti della rete, sia a livello di singoli attori coinvolti nella rete che di network aggregato. Dopo aver analizzato e descritto il contesto odierno in Italia in termini di presenza di istituti e loro disposizione geografica, approfondito la loro attività sotto l'aspetto della pubblicazione di articoli e l'andamento nel tempo arricchendone l'analisi sfruttando l'informazione sul numero di citazioni ricevute, lo studio si concentra sull'analisi del network collaborativo mediante la SNA. Dopo aver analizzato e discusso la disposizione spaziale della rete secondo due diversi algoritmi di visualizzazione si passa ad un'analisi delle principali metriche che la SNA offre per poter caratterizzare e studiare in dettaglio una rete sociale, individuandone i principali attori, quelli maggiormente collaborativi e in una posizione più strategica rispetto alla rete.

# Indice

1. I	Ricero	a	5
1.1	1 R	icerca nel mondo	5
1.2	2 R	icerca in Europa	7
1.3	3 R	icerca in Italia	8
2	Collab	orazioni nella ricerca: vantaggi e svantaggi	g
3	Social	Network Analysis	10
3.1	1 C	os'è la Social Network Analysis	10
3.2	2 T	eoria dei grafi: concetti base	13
3.3	3 M	etriche	13
;	3.3.1	Network Density	14
;	3.3.2	Modularity	14
3.4	1 M	isure di centralità	15
;	3.4.1	Degree centrality	15
;	3.4.2	Betweenness Centrality	16
;	3.4.3	Closeness centrality	18
;	3.4.4	Altre misure di centralità	19
3.5	5 V	isualizzazione del network	19
3.6	6 E	sempi di studi e contesti applicativi	21
4	Istituti	in Italia e pubblicazione di articoli	25
4.1	1 Is	tituti osservati	28
4.2	2 P	ubblicazione articoli	32
4	4.2.1	Analisi citazioni	36
4.3	3 C	o-pubblicazioni	39
5 I	Netwo	rk Visualization	44
6 I	Metric	he	50
6.1	1 N	etwork Density	50
6.2	2 M	odularity	51
6.3	3 D	egree Centrality	53
6.4	4 B	etweenness Centrality	55
6.5	5 C	loseness Centrality	57
7 Co	nclusi	oni	59
Biblio	ografia	1	61

# 1. Ricerca

Nuove sfide, ricerca di risposte e curiosità sono aspetti che hanno da sempre accompagnato l'uomo nella storia. Innumerevoli sono le scoperte e le invenzioni dell'uomo che per pura curiosità o per bisogno e necessità hanno migliorato, e continuano a migliorare, le condizioni di vita dell'uomo sulla Terra. Dell'importanza della ricerca ne è stato un lampante esempio la situazione pandemica dovuta al Covid che ha evidenziato la necessità di una collaborazione internazionale volta alla sperimentazione, della scoperta, produzione di vaccini efficaci a centinaia di migliaia di individui nel più breve tempo possibile.

La ricerca scientifica può dare un grande contributo a trovare soluzioni che supportino lo sviluppo economico di oggi con i bisogni delle generazioni future, che attenuino lo squilibrio fra la ricchezza di pochi e la povertà di molti. [1] Ai fini quindi di una migliore qualità della vita e risolvere problemi di rilevanza sociale è dunque fondamentale che l'uomo non smetta di essere curioso o di porsi quesiti ed è altrettanto necessario che la sete di curiosità venga opportunamente incentivata dalle politiche locali con finanziamenti, progetti di ricerca e incentivi in quanto i meccanismi spontanei del mercato spesso non sono sufficienti.

#### 1.1 Ricerca nel mondo

Un indicatore spesso utilizzato per misurare quanto un paese investe in ricerca e sviluppo è dato rapportando la spesa totale in ricerca e sviluppo al numero di ricercatori nel paese, calcolando quanto il paese, di fatto, investe per ogni singolo ricercatore. Analizzando questo dato per ogni area geografica è facile farsi un'idea sulla situazione generale della ricerca nel mondo.

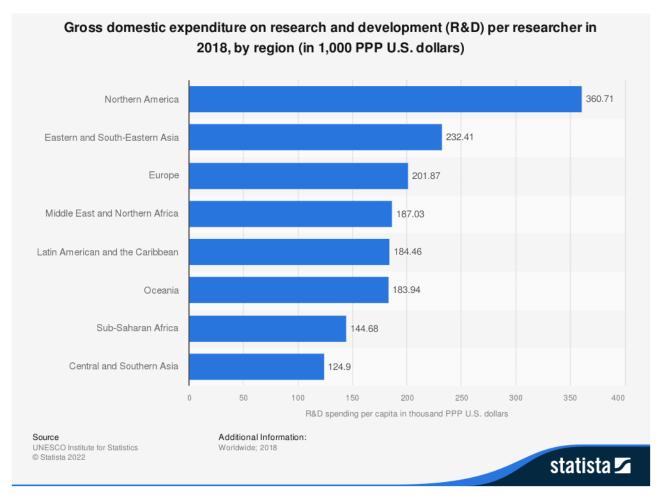


Figura 1- Spesa interna lorda investita per ricercatore nel 2018 per macroregioni geografiche (fonte Statista)

Analizzando la **Figura 1** si nota come il principale investitore in ricerca e sviluppo risulta essere la regione del nord America con un totale di 360\$ pro capite dando un notevole distacco alla seconda classificata, Asia dell'est e del sudest, con un totale di 232\$ pro capite investito in ricerca. A seguire si trova l'Europa con 201\$.

Gli Stati Uniti sono quindi i leader mondiali nella promozione del cambiamento tecnologico e dell'innovazione alimentando l'economia e lo sviluppo di prodotti e servizi. La cultura dell'innovazione degli Stati Uniti ha superato di gran lunga le controparti europee, gli Stati Uniti infatti, con la loro cultura imprenditoriale, danno modo agli imprenditori delle aziende più innovative di mandare avanti il loro business e investimento, rendendolo sostenibile economicamente grazie all'accesso ai finanziamenti di capitale di rischio e protezione della concorrenza che incoraggia anche le aziende di maggior successo a continuare a innovare per rimanere davanti ai rivali che desiderano superarle. [2]

Inoltre, la storia e il tessuto economico favorevole degli Stati Uniti ha da sempre attirato maggiormente i fondi di investimento di venture capital andando ad alimentare le prospettive di crescita delle start up più promettenti, future leader mondiali nel loro settore. Non a caso, infatti, nella classifica per il 2020 delle 50 imprese più innovative del mondo condotta da BCG, con 14 delle prime 20 aziende (70%) le aziende americane dominano sulle altre, mentre le restanti aziende della top 20 risultano essere tutte asiatiche (seconda regione al

mondo per gli investimenti in ricerca e sviluppo) andando a confermare la correlazione positiva tra investimento e innovazione.

Analizzando le aziende nel complesso, gli Stati Uniti rappresentano 25 delle prime 50 aziende (50%). Solo 14 (28%) hanno sede Europea e nessuna di queste rientra in quello che generalmente viene considerato il settore tecnologico, ma rappresentano piuttosto industrie come la produzione di automobili, la vendita al dettaglio, i prodotti farmaceutici e i beni di consumo.

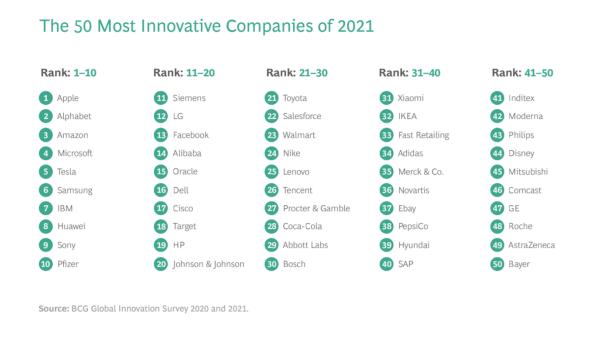


Figura 2 - Prime 50 compagnie col più alto livello di innovazione nel 2021 secondo una ricerca condotta da Boston Consulting Group

## 1.2 Ricerca in Europa

Secondo le stime di Eurostat, nel 2019 l'Europa contava 1 milione e 853 mila ricercatori. In quell'anno la prima nazione europea è stata la Germania con oltre 450 mila ricercatori, seguita dalla Francia con 313 mila e dall'**Italia** con **160.824 ricercatori**. Nel 2020 gli Stati membri dell'UE hanno speso circa 311 miliardi di euro in ricerca e sviluppo, con un decremento di 1 miliardo rispetto all'anno precedente. Il livello di percentuale del PIL speso in ricerca si è attestata attorno al 2,3% nel 2020, rispetto al 2,2% del 2019. Tuttavia, è da tener conto della situazione pandemica che nel 2020 ha stravolto la normale quotidianità, progetti di ricerca compresi. Guardando più indietro nel tempo, infatti, si assiste ad una crescita degli investimenti nella ricerca, dovuta anche grazie agli obiettivi stilati da *Europa 2020*. Tuttavia questo aumento non è bastato a superare la quota di PIL destinata all'innovazione negli altri paesi. [3] Tra gli obiettivi stilati da *Europa 2020* (Europe 2020), una strategia decennale proposta dalla Commissione europea il 3 marzo 2010 per promuovere una "crescita intelligente, sostenibile e inclusiva" dopo la recessione iniziata nel 2008, vi è l'aumento degli investimenti in ricerca e sviluppo al 3% del PIL dell'UE. Avere negli obiettivi

l'aumento della spesa nella ricerca e nell'innovazione è da considerare un obiettivo strategico volto a colmare il divario tecnologico dell'Europa con le altre potenze mondiali. [3]

Nel 2020, secondo una stima dell'Eurostat, le nazioni europee che ha più investito in ricerca e sviluppo, in termini di percentuale del PIL, sono state Belgio e Svezia (3,5% del PIL), seguite da Austria (3,2% del PIL) e Germania (2,1% del PIL), Danimarca (3% del PIL) e Finlandia (2,9% del PIL). Le nazioni che invece hanno investito meno in ricerca sono state la Romania (0,5% del PIL), Malta e Lettonia (0,7% del PIL), Cipro, Bulgaria e Slovacchia (0,9% del PIL). L'Italia, invece, nel 2020 ha speso l'1,53% del PIL in ricerca e sviluppo.

#### 1.3 Ricerca in Italia

Le politiche della ricerca in Italia sono definite dal Programma nazionale per la ricerca (PNR), ovvero un documento che programma e coordina le attività e i finanziamenti della ricerca, previsto dal D.Lgs. 204/1998. Il PNR definisce un quadro definito di interventi e una serie di interventi mirati a rafforzare la presenza dell'Italia e dei suoi ricercatori nel panorama europeo e mondiale.

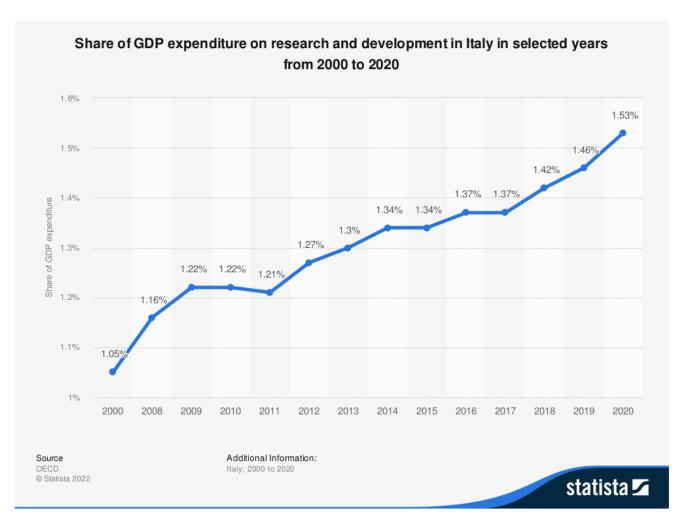


Figura 3- Investimento in ricerca e sviluppo in Italia come percentuale del prodotto interno lordo (fonte Statista)

A fare da traino per la crescita dell'investimento in programmi di ricerca e sviluppo in Italia visibile in **Figura 3**, oltre che le direttive europee stilate del programma *Europe 2020* ha influito anche il PNR 2015-2020.

Il Programma nazionale per la ricerca 2015-2020 ha definito una serie di programmi innovativi volti a far lievitare il numero di ricercatori in Italia, investimenti sulle infrastrutture, la definizione di quattro aree prioritarie su cui puntare per la ricerca applicata (Aerospazio, Agrifood, Salute e Industria 4.0), più sinergia tra sistema pubblico e privato e un programma strategico ad hoc per il Mezzogiorno. Il programma ha stanziato fondi per quasi 2,5 miliardi di euro di risorse nei primi tre anni, che si aggiungono al finanziamento che il Ministero dell'Università e della Ricerca destina a Università ed Enti Pubblici di Ricerca, pari a 8 miliardi ogni anno. Gran parte delle risorse, circa il 40%, sono destinate all'accrescimento del numero di ricercatori e del capitale umano cercando anche di attrarre i migliori talenti da fuori i confini nazionali con misure di sostegno per i giovani prevedendo un ingresso di 6000 dottori e ricercatori in aggiunta a quelli ordinari. Come rilevato nelle relazioni annuali della Commissione europea, per il nostro paese un maggior investimento in innovazione potrebbe avere ricadute particolarmente positive, in particolare per il meridione. È per questo che vengono poste in essere azioni prioritarie per il sostegno alla ricerca e innovazione in quest'area dell'Italia, ponendo in sinergia Programma Operativo Nazionale, Programmi Operativi Regionali e risorse ordinarie. [4]

# 2 Collaborazioni nella ricerca: vantaggi e svantaggi

Al fine di migliorare la ricerca, contribuire maggiormente alla divulgazione scientifica, trovare soluzioni più efficienti in tempi più brevi un obiettivo fondamentale è quello della collaborazione tra ricercatori ed enti di ricerca. Lo sviluppo di una rete di collaborazione porta con sé infatti numerosi vantaggi, tra questi vi sono: un maggior impatto della pubblicazione, c'è infatti una diretta correlazione tra il numero di autori e i fattori d'impatto della pubblicazione, incoraggia la creatività, giungendo a proposte innovative grazie anche al criticismo accendendo il dibattito delle idee e delle proposte. Meno lavoro individuale porta a delle conclusioni più precise e ben fatte senza compromettere quindi il risultato della ricerca. Lavorare su progetti di ricerca con un team opportunamente costruito inoltre, porta più esperienza e conoscenza sul tavolo, aumentano quindi le chance di risoluzioni di problemi più efficientemente e risparmiando tempo. Lo scambio di idee e punti di vista differenti infatti porta a soluzioni più impattanti, come dimostra anche uno studio condotto in Giappone su come un network di collaborazione internazionale di ricerca impatta sull'innovazione dimostrando come studi di ricerca svolti in un contesto internazionale portano a soluzioni 2.7 volte più efficaci rispetto a contesti di ricerca nazionali. [5]

Osservando colleghi e collaboratori lavorare e affrontare criticità porta ad un apprendimento più efficiente del singolo. Tra i vantaggi, c'è anche quella della condivisione, oltre che della conoscenza del singolo, anche degli strumenti. Inoltre, gli studi di ricerca collaborativi hanno una maggiore probabilità di ottenere fondi, negli US e in Europa, specie se l'output della ricerca ripercuote positivamente nel sociale. Per esempio, la Commissione Europea, organo

esecutivo dell'Unione Europea, ha creato un programma di finanziamento che va sotto il nome di *Horizon 2020* stanziando circa 80 miliardi di euro tra il 2014 e Il 2020 per sostenere e promuovere la ricerca nello spazio europeo della ricerca, prioritizzando progetti di ricerca collaborativi. [6]

Possono esserci tuttavia anche delle criticità e aspetti negativi nella collaborazione. Ad esempio, si potrebbe vanificare il vantaggio collaborativo se non c'è un giusto clima tra i ricercatori, specie se essi non sono caratterialmente predisposti alla collaborazione e al lavoro in team. Potrebbero nascere delle discordie per l'accredito di meriti e di authorship e inoltre, specie in campi in cui la ricerca è vasta e ci sono un grosso numero di autori attivi, potrebbe essere difficile trovare il giusto collaboratore.

# 3 Social Network Analysis

# 3.1 Cos'è la Social Network Analysis

La Social Network Analysis, o SNA, offre la soluzione a numerosi problemi di ricerca con un approccio teorico e metodologico basato sullo studio, l'analisi e la visualizzazione delle relazioni, legami e contatti tra attori di una determinata comunità o network. I contesti analizzabili mediante uno studio di social network analysis quindi, possono essere svariati. Di fatto, ogni contesto che coinvolge relazioni tra enti e individui può essere soggetto a uno studio basato sulla SNA trovando applicazioni in analisi a livello di società, comunità e organizzazioni in ambiti come relazioni inter-organizzative, nella diffusione di malattie contagiose, nella salute mentale, nel sostegno sociale, nella diffusione delle informazioni.

Esempi di informazioni che si possono dedurre da uno studio di SNA e ad un'analisi delle sue metriche possono essere diverse. Ad esempio, la Social Network Analysis è diventato un metodo ampiamente diffuso nella ricerca e nel business per la ricerca e l'analisi delle relazioni a livello di individuo, all'interno di un'organizzazione o a livello di società. Identificando gli attori "leader" nel network in base al numero di connessioni o posizioni strategiche; esaminare eventuali gruppi e cluster che si sono definiti e la loro correlazione con gli altri moduli della rete; l'intensità dei collegamenti può mettere in evidenza le relazioni più importanti e eventuali pattern ricorrenti, tra quali attori del network e gli effetti che essi determinano.

Possiamo far risalire l'origine dello studio delle reti sociali al lavoro condotto da Stanley Milgram. Nel suo famoso esperimento del 1967, Milgram condusse un test per capire in che modo le persone sono collegate agli altri esaminando l'average path length (concetto della topologia di rete, definito come il numero medio di passi lungo i percorsi più brevi per tutte le possibili coppie di nodi della rete e che aiuta a dare una misura dell'efficienza del trasporto di informazioni su una rete) di un campione di popolazione degli Stati Uniti. I risultati del test hanno portato alla conclusione che la società è un network caratterizzato da "short pathlenghts" (connessioni corte), il numero di legami che intercorrono tra due persone qualsiasi de campione, infatti, risulta essere pari a 6. [7]

Un altro passo importante nello sviluppo dell'analisi delle reti sociali è stato il lavoro di Mark Granovetter. Nel suo ampiamente citato articolo del 1973 "The Strength of Weak Ties" (La forza dei legami deboli), Granovetter sostiene che i "legami deboli", ovvero le relazioni con i conoscenti, sono più importanti dei "legami forti" ovvero le relazioni con la famiglia e gli amici più stretti, quando si cerca di trovare un impiego. [8]

Lo studio di Granovetter e ricerche successive hanno ampliato questa tesi sostenendo che le reti aperte, più disperse e non ridondanti, hanno un maggiore accesso alle informazioni rispetto alle reti più piccole, più dense e più interconnesse, in quanto riescono a fornire una maggiore diversità di conoscenze e informazioni nel network.

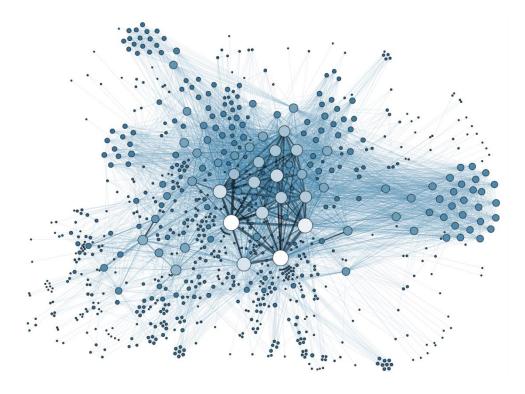


Figura 4 - Esempio di visualizzazione grafica di un network sociale

Il metodo della SNA trova sempre più spazio nella ricerca e il suo utilizzo sta diventando anno dopo anno sempre più diffuso. I fattori che incoraggiano la diffusione di questo strumento sono diversi.

- Interconnessione. La complessità delle connessioni della società moderna, data da fenomeni come internet, crisi finanziarie o epidemie è data dal comportamento aggregato di gruppi di persone le cui azioni hanno conseguenze sul comportamento di tutti gli altri. Il crescente interesse verso lo studio di tale complessità ha reso la Social Network Analysis uno degli strumenti di visualizzazione e rappresentazione dei sistemi complessi più utilizzati. [9]
- Big Data. Lo sviluppo tecnologico e il costante utilizzo di internet nei più disparati settori genera ogni giorno una mole di informazioni e collezione di dataset sempre più grandi. Strumenti di data analysis sono necessari e

- sempre più richiesti da compagnie e organizzazioni al fine di estrapolare utili informazioni dai dati.
- Sviluppo Tecnologia. Altro elemento che ha favorito la diffusione del metodo SNA è stata la crescente disponibilità di calcolo computazionale affiancato allo sviluppo di linguaggi di programmazione e apposite librerie, elementi quindi, in grado di elaborare grandi quantità di dati in poco tempo e offrire calcolo e misure in breve tempo.

Lo studio delle reti sociali, quindi, ha cominciato a decollare come specialità interdisciplinare andando incontro ad una rapida crescita e diffusione del suo utilizzo. Nell'ambito della ricerca, infatti, si può notare nella **Figura 5** come anno dopo anno, dal 2000 al 2021, si è assistito ad una continua crescita del numero di articoli e pubblicazioni riportanti la voce "social network analysis" nel titolo, nell'abstract o nelle parole chiave definite dagli autori. Nel 2022, ad esempio, assistiamo già ad un totale di 1196 articoli pubblicati al mese di luglio è quindi facile aspettarsi una certa crescita degli articoli connessi alla SNA anche rispetto al 2022.

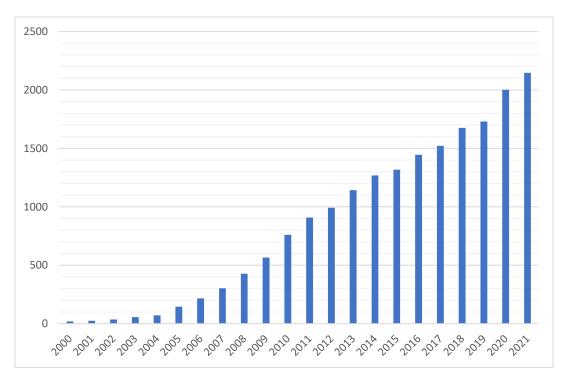


Figura 5 - Pubblicazioni e articoli riportanti "social network analysis" in titolo o in abstract. Ricerca condotta dallo studente tramite Scopus.

Dal 1990, la network analysis è stata istituzionalizzata. Diversi sono i corsi di network analysis sparsi per le università di tutto il mondo. È stato inoltre fondato *l'International Network of Social Network Analysis* per creare una community di ricercatori appassionati ed esperti di network analysis coinvolgendoli in congressi, iniziative e pubblicazioni di articoli scientifici nel suo stesso journal "*Connections*".

## 3.2 Teoria dei grafi: concetti base

La SNA è uno strumento che si basa principalmente sulle scienze sociali e sui concetti chiave della teoria dei grafi, ampliando questa branca della matematica con metriche e analisi proprie.

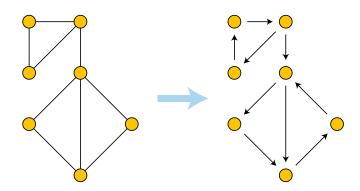


Figura 6 - Esempio di grafo non orientato (a sinistra) e grafo orientato (a destra)

I grafi sono strutture matematiche, composta da nodi e archi che uniscono coppie di nodi esprimendone la relazione, che possono essere utili in una moltitudine di contesti applicativi, tra cui appunto la Social Network Analysis. Un grafo G è dunque generalmente definito come G := (V, E) con |V| vertici e |E| archi.

Gli archi, nonché i collegamenti tra i diversi nodi, possono essere orientati con un verso di percorrenza da uno specifico nodo all'altro, oppure possono essere privi di verso. Nel primo caso avremo quindi un grafo orientato (grafo a destra della **Figura 6**), nel secondo un grafo non orientato (grafo a sinistra della **Figura 6**). A seconda del contesto, agli archi, possono essere associati valori e pesi, rappresentando una qualche caratteristica del collegamento. Ad esempio, nell'ambito della ricerca operativa, in un grafo dove i nodi rappresentano dei centri logistici, un valore numerico associato al collegamento tra due centri potrebbe essere il numero di merci che deve transitare da un centro all'altro.

## 3.3 Metriche

Le reti sociali sono sistemi complessi e come tali necessitano di una moltitudine di metriche e strumenti di analisi. Le principali metriche sono quelle di centralità, dove l'analisi si focalizza a livello di singolo nodo all'interno della rete. La Social Network Analysis con le metriche di centralità permette di studiare le relazioni di ogni attore nella rete mostrandone le gerarchie. Ma vi sono anche analisi che guardano alla rete nel suo insieme analizzandone la topologia, la sua struttura generale e nello specifico individuando eventuali cluster e gruppi definiti oppure si può adottare un approccio dinamico guardando l'evoluzione della rete nel tempo.

In questo capitolo andremo ad analizzare nello specifico le diverse metriche principalmente adottate in uno studio di network analysis che rispondono a diverse esigenze fornendo valori e contenuti informativi utili a leggere la rete e la sua struttura.

#### 3.3.1 Network Density

La network density (o densità) rappresenta una delle più importanti statistiche descrittive quando si parla di SNA, esso infatti è un indicatore del livello generale di coesione della rete. È quindi un indicatore di quanto la rete e il network è unito, coeso e collaborativo.

La network density, è definita dal rapporto tra il numero di legami presenti nella rete e il numero totale di legami possibili data la numerosità dei nodi e può quindi assumere valori compresi tra 0 e 1.

Dato un grafo G := (V, E) con |V| vertici e |E| archi la formula per il calcolo della densità per un grafo non orientato è:

$$D = \frac{\frac{|E|}{|V| * (|V| - 1)}}{2} = \frac{2|E|}{|V| * (|V| - 1)}$$

Un valore di densità vicina allo 0 denota un livello di coesione basso, il network risulta quindi essere scarsamente connesso. Al contrario, valori di densità vicini e prossimi all'1 denotano un alto livello di coesione nella rete. La densità, infatti, è pari ad uno in caso di reti "fully-connected", quelli cioè dove tutti i nodi sono collegati agli altri, come il grafo raffigurato in **Figura 7**.

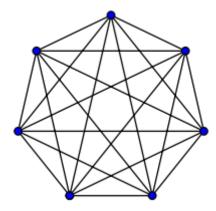


Figura 7 - Esempio di grafo completamente connesso o "fully connected"

## 3.3.2 Modularity

La modularity (modularità) è una metrica che da informazioni sulla struttura del network o del grafo e il suo valore quantifica la qualità della divisione della rete in moduli o comunità. Una suddivisione netta e definita in cluster e moduli determina alti valori di modularità della rete, troveremo quindi un quanto più denso numero di relazioni e collegamenti tra nodi dello stesso cluster e meno collegamenti tra nodi appartenenti a diversi moduli man mano che il valore della modularità aumenta. Il suo utilizzo è quindi utile a determinare le comunità presenti su una rete. La sua misura può assumere valori che vanno da -1 a 1. Inoltre, con l'ausilio di appositi algoritmi e software, si è in grado di mettere in evidenza eventuali moduli della rete, come per esempio l'algoritmo *Force Atlas* di cui si discuterà in seguito.

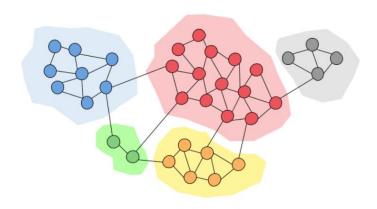


Figura 8 - Esempio di network in cui sono ben identificabili alcuni cluster

## 3.4 Misure di centralità

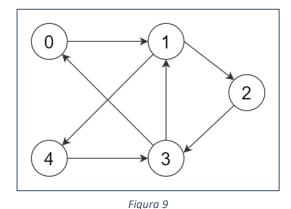
Le misure di centralità permettono di definire il posizionamento di un attore nella propria rete in termini relazionali identificando, quindi, i nodi maggiormente influenti e più coinvolti dalla rete oppure quelli che fungono in maggior misura da "ponte" e attraverso i quali confluisce conseguentemente un maggior numero di informazioni. Le misure di centralità, di fatto, sono misure di "importanza" di un nodo all'interno di un network. Possiamo dunque avere diverse metriche che misurano il grado di centralità di un nodo in base all'aspetto su cui ci si va a focalizzare. I criteri di importanza possono essere infatti differenti, ad esempio, l'influenza che il nodo ha piuttosto che il numero di informazioni che passano attraverso esso o altre. Diverse metriche analizzano quindi differenti aspetti del nodo in analisi.

## 3.4.1 Degree centrality

La degree centrality è la misura di centralità più semplice, misura semplicemente il numero di collegamenti che ogni nodo possiede, un maggior valore di degree centrality corrisponde ad un maggior numero di legami e connessioni diretti che il nodo ha instaurato nella rete.

Si può utilizzare il grado di degree centrality per individuare gli attori maggiormente connessi, gli individui più popolari del network e che probabilmente portano con sé un numero di informazioni maggiori. Identificare attori della rete più interconnessi può essere molto utile in determinati contesti applocativi. Ad esempio, identificando gli individui che hanno un numero di connessione maggiore può essere utile a prevenire catene di infezione da virus ed evitare una propagazione. Al contrario, può essere utile a diffondere in modo più efficace e rapido un'informazione all'interno della rete, oppure ancora può identificare l'individuo a cui chiedere qualche tipo di informazione per primo, in un'ottica di riduzione degli sforzi. Erroneamente si può indurre a pensare che l'attore maggiormente interconnesso sia quello più influente sugli altri, ma non necessariamente. Ad esempio, infatti, se prendiamo in analisi la gerarchia all'interno di un contesto aziendale, l'attore più influente, ovvero il CEO o amministratore delegato dell'azienda, generalmente è l'individuo che ha meno connessioni.

Quando si ha un grafo orientato, ovvero quando è possibile identificare una direzionalità dei collegamenti tra i nodi, può essere utile identificare il grado di degree centrality separato per connessioni in entrata e connessioni in uscita, identificando così due sotto misure della degree centrality: in-degree centrality, ovvero il conteggio del numero di legami diretti al nodo e out-degree centrality, il numero di legami che il nodo dirige verso altri.



Nodo	Degree-	In-degree	Out-	
	centrality	centrality	degree	
			centrality	
0	2	1	1	
1	4	2	2	
2	2	1	1	
3	4	2	2	
4	2	1	1	
T-1 II 4				

Tabella 1

#### 3.4.2 Betweenness Centrality

Il calcolo del livello di betweenness centrality aiuta ad identificare gli elementi della rete che fungono in maggior misura da tramite tra i nodi del network. Esso, infatti, misura il numero di volte che un determinato nodo si interpone sul percorso più breve tra altri nodi mettendo in evidenza gli individui che influenzano il flusso informativo all'interno del sistema, utile quindi quando è necessario analizzare le sue dinamiche comunicative. Per due qualsiasi nodi della rete, infatti, è possibile trovare il percorso più breve che li connette, ovvero quello che li mette in comunicazione con un minor numero di nodi e archi comunicanti.

Il grado di betweenness centrality di un nodo v di un grafo G := (V, E) con |V| vertici e |E| archi si calcola come:

$$C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

Dove  $\sigma_{st}$  è il numero di percorsi più brevi che intercorrono da s a t, mentre  $\sigma_{st}(v)$  è il numero di quei percorsi che passano per v.

Quindi un nodo  $\Delta$  diventa fondamentale per l'efficienza tra il flusso che c'è tra i nodi  $\epsilon$  e  $\beta$  se esso si interpone sul percorso più breve che intercorre tra i nodi  $\epsilon$  e  $\beta$ . Contando di fatto il numero di volte in cui un determinato nodo si trova su un percorso critico tra altri nodi, la betweenness centrality diventa un indicatore importante per la stima dei nodi fulcro fondamentali della rete. I nodi con elevata centralità giocano un ruolo cruciale nel flusso di

informazioni e della coesione della rete e sono considerati centrali e indispensabili per la rete per il loro ruolo nel flusso comunicativo. [10]

Diventa inoltre importante conoscere quali sono i nodi con i più alti valori di betweenness centrality perché oltre ad essere nodi strategici, in caso di loro malfunzionamento o problematica, si può predire facilmente quale sarà il nodo di successiva importanza per la propagazione del flusso della rete.

Generalmente i nodi aventi alta betweenness sono quelli aventi anche alta degree centrality, ma non sempre è vero, infatti è possibile avere nodi di connessione tra due regioni di network ben distinte e collegate soltanto da uno o pochi nodi che fungono da ponte per le due regioni, e che di conseguenza assumono un basso valore di degree centrality. Ad esempio, in **Figura 10** è evidenziato in rosso il nodo avente un maggior valore di betweennees centrality, in questo caso esso porta con sé anche un alto valore di degree centrality. L'esempio riportato invece in **Figura 11** dimostra come possono esserci casi in cui nodi aventi grandi valori di betweennees centrality possono avere un basso valore di degree centrality, ovvero quando abbiamo tipicamente nodi che fanno da "ponte" e collegamento tra due regioni di network ben distinte, in questo caso infatti, i percorsi che collegano i nodi delle due diverse regioni devono necessariamente passare per i nodi ponte che automaticamente assumono quindi alti valori di betweennees centrality.

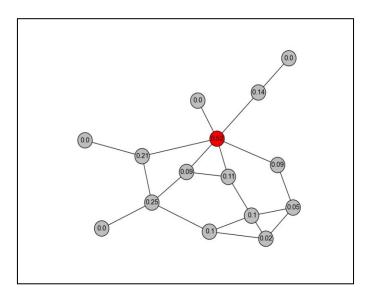


Figura 10

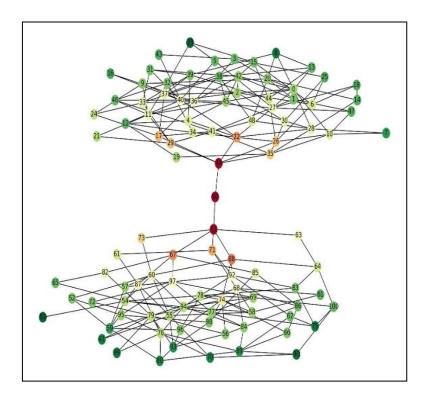


Figura 11

#### 3.4.3 Closeness centrality

Un'altra metrica di centralità molto utilizzata nella SNA la closeness centrality. Questa metrica è utilizzata per individuare i nodi della rete che sono nella posizione migliore per raggiungere tutti gli altri e capaci quindi di "influenzare" la rete più rapidamente. La metrica, infatti, associa ad ogni nodo un punteggio basandosi sulla somma dei percorsi più brevi per il raggiungimento di tutti gli altri nodi della rete calcolando la lunghezza media del percorso più breve da un nodo verso qualunque altro. In altre parole, rappresenta la velocità con cui l'informazione può raggiungere altri nodi da un dato nodo di partenza [9]

In un grafo connesso, la centralità normalizzata della vicinanza (o prossimità) di un nodo è la lunghezza media del percorso più breve tra il nodo e tutti gli altri nodi del grafo. Pertanto, più un nodo è centrale, più è vicino a tutti gli altri nodi.

Quando si parla di closeness centrality di norma si fa riferimento alla formula che prevede la normalizzazione per il numero dei nodi della rete in modo da rendere possibile la comparazione di questa misura anche con altri grafi di grandezza differente.

La closeness centrality di un nodo si calcola con la seguente:

$$C(v) = \frac{N-1}{\sum_{U} d(u, v)}$$

dove d(u, v) è la distanza tra due nodi  $u \in v$ , mentre N è il numero di nodi totale della rete.

Nella figura sottostante, il nodo ad avere maggior closeness centrality è il nodo C.

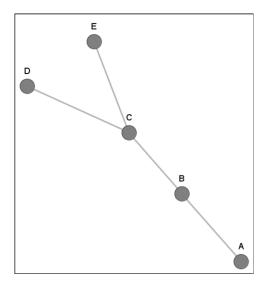


Figura 12

#### 3.4.4 Altre misure di centralità

Esistono inoltre, altre metriche di centralità dei nodi più sofisticate. Ad esempio, vi è l'eigenvector centrality ovvero la centralità degli autovettori che tiene conto non solo del numero dei collegamenti incidenti sul nodo in questione ma anche della qualità di tali collegamenti ovvero vedendo quanti collegamenti hanno essi a loro volta e così via. Un nodo con alta eigenvector centrality quindi è un nodo connesso a nodi aventi molte connessioni a loro volta. Un'altra misura di centralità è stata introdotta da Leo Katz, uno statistico americano, nel 1953. La centralità di Katz può essere considerata una variante dell'eigenvector centrality in quanto essa associa un peso alle connessioni e ai percorsi, più la connessione è vicina a nodo di cui si calcola le centralità più essa avrà associato un peso maggiore, al contrario per quanto riguarda invece i collegamenti più lontani dal nodo.

## 3.5 Visualizzazione del network

Grazie all'utilizzo di software specializzati nello studio e visualizzazione di network analysis siamo in grado oggi di utilizzare algoritmi di visualizzazione della rete in grado di darci ulteriori informazioni in base alla disposizione dei nodi e della rete. Introdotti per avere una visione più limpida del network e senza l'incrocio di connessioni, gli algoritmi di VNA (Visual Network Analysis) sono basati su principi di minimizzazione dell'energia del sistema bilanciando un meccanismo di forze facendo analogia con un sistema fisico. I nodi sono carichi di una forza repulsiva che li allontana, mentre i bordi introducono una forza attrattiva tra i nodi che collegano. Lanciato l'algoritmo negli appositi software, i vettori forza variano la posizione dei nodi cercando di bilanciare la repulsione dei nodi e l'attrazione dei bordi tendendo verso il raggiungimento di un equilibrio. All'equilibrio, i layout orientati dalle forze producono una disposizione dei nodi che è visivamente significativo.

Utilizzando algoritmi di spazialità basata sull'equilibrio di forze, in VNA, è importante non focalizzarsi sull'esatta posizione dello specifico nodo all'interno della rete quanto piuttosto ad identificare strutture significative in termini di vicinanze e spazialità relative tra gruppi e cluster di nodi. Non conta tanto la posizione del nodo quanto la densità strutturale dell'area analizzata nella rete e di eventuali polarizzazioni rappresentativi di raggruppamenti e collaborazioni maggiori.

Un esempio di disposizione del grafo rappresentante un network determinata dall'algoritmo di Fruchterman-Reingold è rappresentato in **Figura 13**, L'effetto di questo algoritmo è quello di avvicinare tra loro i nodi aventi una correlazione maggiore tra essi e distanziare maggiormente quelli meno collegati verso l'esterno. L'effetto finale è quindi quello di un raggruppamento dei nodi della rete tra loro meglio collegati e che trovano una disposizione centrale nel grafo e quelli meno collegati alle estremità della rete.

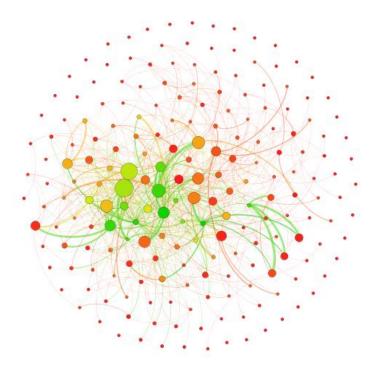


Figura 13 - Esempio di network disposto utilizzando l'algoritmo di Fruchterman-Reingold

In questo modo si evidenzia immediatamente quali membri si stanno impegnando bene e quali sono completamente scollegati dagli altri membri. Tuttavia, quando si deve analizzare e visualizzare un network composto da un maggior numero di nodi, questo algoritmo può diventare meno efficace nell'identificazione di eventuali cluster e raggruppamenti. Il grafico successivo in **Figura 14** dispone i nodi utilizzando un algoritmo differente, come può essere per esempio Force Atlas 2.

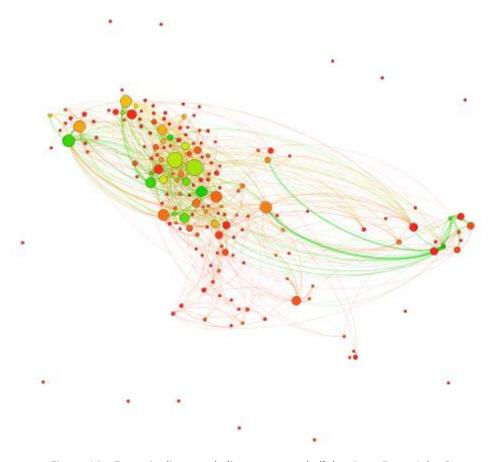


Figura 14 – Esempio di network disposto secondo l'algoritmo Force Atlas 2

Una migliore lettura del network e delle sue caratteristiche, inoltre, può essere raggiunta associando agli algoritmi di VNA sopra citati ad un'opportuna scelta grafica e di dimensionamento dei nodi. Ad esempio, nel caso in cui si ha un insieme di nodi appartenenti a categorie diverse o accomunabili per qualche caratteristica, si può optare per una colorazione dei nodi basata sulla categoria di appartenenza. Oppure ancora, se ad ogni nodo è associato un valore, i software offrono una opzione di visualizzazione che consente di ridimensionare la grandezza dei nodi sulla base dei valori ad essi associati.

# 3.6 Esempi di studi e contesti applicativi

La SNA trova quindi applicazione in tutte quelle che sono le scienze sociali, ovvero quell'insieme di discipline che si occupano di studiare l'essere umano e la società attraverso l'impiego del metodo scientifico, quindi contesti come quelli psicologici, economici, sociali. L'analisi delle relazioni e della rete potrebbe portare l'enorme vantaggio predittivo del comportamento di alcuni attori e delle reazioni delle loro azioni sulla rete. È importante, infatti, il contributo che la SNA ha dato e continua a dare nel campo investigativo per quanto riguarda la prevenzione di attacchi terroristici. Di seguito sono riportati due studi condotti sulla base della social network analysis.

Uno studio condotto da Raquel Recuero, Gabriela Zago e Felipe Soares e pubblicato su un articolo con il titolo di "Using Social Network Analysis and Social Capital to Identify User Roles on Polarized Political Conversations on Twitter", ha analizzato un dataset estratto da Twitter per studiare il ruolo che giocano diverse categorie di utenti (attivisti politici, influencer, giornalisti ecc...) sulle discussioni politiche su Twitter. Il caso studio si basa su dati estratti durante 3 giorni cruciali riguardanti un processo per corruzione che ha coinvolto l'ex presidente Brasiliano Luiz Inácio Lula da Silva. I dati raccolti sono quindi così suddivisi nelle tre giornate: 22 Gennaio (10,115 nodi, 22,883 archi) 23 gennaio (9,288 nodi, 18,123 archi) 24 gennaio (17,710 nodi, 34,936 archi). I nodi sono rappresentativi di account Twitter, mentre gli archi possono rappresentare retweets o menzioni tra utenti.

Lo studio si è concentrato sull'analisi di tre metriche riguardanti la SNA. In-degree centrality, out-degree centrality e modularity. L'in-degree centrality, associando un valore al nodo basato sul numero di nodi entranti in esso, determina nel caso di studio specifico, quali sono gli utenti a ricevere il maggior numero di retweet e mansioni. Calcolando la metrica dell'in-degree centrality quindi si è in grado di determinare quali sono gli utenti dotati di maggiore visibilità, credibilità o livello di conoscenza. La metrica inoltre, fa emergere, associando un numero di in-degree centrality maggiore, gli attori che meglio riescono ad influenzare l'opinione pubblica sulla base del contenuto pubblicato e la loro reputazione/autorità, associandogli quindi l'appellativo di "opinion leaders". [11]

L'out-degree centrality, al contrario, calcolando il numero di archi uscenti dal nodo associa un alto valore ad utenti che hanno relazionato con tante persone mediante retweet o menzioni. Il valore dell'out-degree centrality è quindi associato, nello studio specifico, al livello di attività e ad aspetti e caratteristiche che associano all'utente una tendenza alla comunicazione, partecipazione e a stabilire un grande numero di relazioni.

Con il calcolo della modularità invece misura la tendenza a creare moduli o gruppi strettamente connessi in un network e può assumere valori da 0 a 1. Maggiore è la modularità, più densamente saranno connessi i nodi all'interno di un cluster e meno dense invece le connessioni con gli altri gruppi. La modularità, in altre parole, determina in che misura è disconnesso un network e rappresenta un'importante metrica in questo caso studio in quanto mette in evidenza eventuali gruppi di pensiero. Tanto più un network ha alta modularità e tanto minori saranno i confronti e le interazioni tra gruppi, mettendo in chiara luce la polarizzazione della rete, evitando così scambio di idee e confronti tra posizioni differenti e accesso a idee diverse dalle proprie.

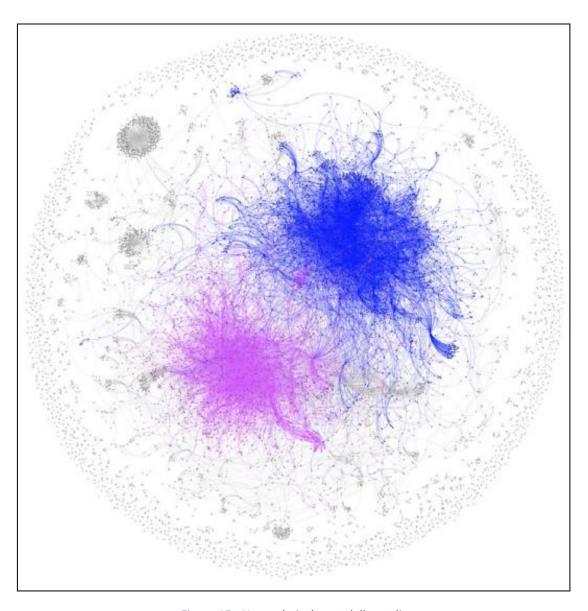


Figura 15 - Network risultante dello studio

In Figura 15 è rappresentata la visualizzazione grafica del network studiato corrispondente al giorno 23 gennaio. Ogni nodo rappresenta un utente, mentre i collegamenti sono dati dall'interazione osservata nel dataset tra due nodi attraverso una menzione dell'account o un retweet. Analizzando i tweet, gli autori, hanno determinato la posizione politica del tweet catalogandolo come Pro-presidente o Contro-presidente e associando una colorazione rosa alle interazioni favorevoli all'ex presidente, mentre con il colore blu sono stati messi in evidenza le posizioni favorevoli alla condanna o contro l'ex presidente. La modularità calcolata per i 3 dataset risulta essere relativamente alta, riportando valori di 0.621, 0.685 e 0.652 per i giorni 22,23 e 24 gennaio rispettivamente denotano un alto valore di polarizzazione politica. I due gruppi, come emerge anche dalla rappresentazione della rete in Figura 15, infatti, sono scarsamente connessi. Lo studio mette anche in evidenza come i tweet aventi una maggiore visibilità e condivisione nei due gruppi sono quelli più politicamente schierati sfavorendo il confronto e circolo di idee contrastanti.

Altro esempio di studio basato sull'utilizzo della SNA e delle sue metriche è dato dallo studio condotto da Alireza Abbasi, Jörn Altmannb e Liaquat Hossain intitolato "Identifying the effects of co-authorship networks on the performance of scholars: A correlation and regression analysis of performance measures and social network analysis measures". Lo studio determina la correlazione tra le principali metriche adottate nella Social Network Analysis e le performance di successo degli autori di articoli pubblicati sulle riviste scientifiche. In particolare, sono stati raccolti dati di cinque università: University of Pittsburgh, University of Berkley, University of Maryland, University of Michigan, and Syracuse University. La scelta di queste specifiche università è data dalla condivisione di corsi di laurea simili tra loro nel campo dell'information management and systems coprendo un arco temporale di cinque anni che va dal 2001 al 2005 e collezionando informazioni riguardanti circa 2139 pubblicazioni, 1806 autori e 5310 collaborazioni.

I risultati dello studio dimostrano una alta correlazione tra le metriche calcolate (normalized degree centrality, normalized betweenness centrality, efficiency, and average ties strength) e le misure di performance degli articoli come per esempio il numero di citazioni che tale articolo ha registrato. Gli autori che riportano un maggior valore di degree-centrality, ovvero quelli connessi maggiormente con altri studenti, mostrano delle performance migliori rispetto a quelli con un basso valore di degree-centrality. Allo stesso modo, autori che hanno consolidato col tempo dei legami più "forti", come per esempio co-authorship ripetute, performano meglio rispetto a chi hanno una singola collaborazione con diversi autori. [12] Un altro aspetto interessante emerso dall'analisi delle metriche è la conclusione che autori e ricercatori che mantengono una forte collaborazione con solo un autore di un determinato cluster mostra misure di performance superiori rispetto a chi ha stabilito diverse relazioni collaborative con più studenti dello stesso gruppo di autori tra loro fortemente connessi. La studio afferma quindi che un ricercatore dovrebbe evitare collaborazioni con autori dello stesso cluster. Inoltre, lo studio mette in luce una correlazione positiva tra alti valori di metriche di centralità e produttività nel tempo. I ricercatori che hanno più collaborazioni con singoli autori o gruppi risultano essere più produttivi.

# 4 Istituti in Italia e pubblicazione di articoli

Il database utilizzato per la realizzazione dello studio, generato da **Web of Science** è composto da un totale di 112 file csv (comma separated values) che riportano la lista di pubblicazioni dell'istituto di riferimento avvenute nel corso del tempo. Le osservazioni sono state fatte su un totale di **98** istituti tra università statali, private, telematiche e cliniche associate, su un intervallo di tempo che può variare da quaranta a sei anni per un numero complessivo di **1.389.312** osservazioni.

La struttura dei file csv è quindi composto da un totale di 1.389.312 osservazioni di venti diverse variabili (Accession.Number, DOI, Pubmed ID, ArticleTitle, Link, Authors, Souce, Research Area, Document Type, Volume, Issue, Pages, Publication Date, Times Cited, Journal Expected citations, Category Expected Citations. Journal Normalized Citation Impact, Category Normalized Citation Impact, Percentile in Subject Area, Journal Impact Factor) di cui è stata fatta una selezione con il principale fine di diminuire il peso del file finale e filtrare le informazioni non utili per lo studio del network di collaborazione tra gli enti e delle sue caratteristiche.

#### Le variabili selezionate sono:

Accession. Number: codice alfanumerico identificativo univoco della pubblicazione.

Article Title: titolo della pubblicazione

Authors: lista autori e collaboratori della pubblicazione

Research Area: Area di ricerca dell'articolo

Document.Type: Tipologia di documento (Article, Proceeding Paper, meeting Abstract

ecc...)

Publication.Date : anno di pubblicazione del documento

Times.Cited: Numero di volte che la pubblicazione di riferimento è stata citata

Le tipologie di documenti (document type) che si trovano all'interno del database sono in totale 30 e le loro occorrenze nel dataset sono le seguenti:

Article	948436	68.27%	
Meeting Abstract	155887	11.22%	
Proceedings Paper	114370	8.23%	
Review	75637	5.44%	
Letter	37133	2.67%	
Editorial Material	33652	2.42%	
Note	6953	0.50%	
Correction	6756	0.49%	
Book Review	5379	0.39%	
Book Chapter	2793	0.20%	
Biographical-Item	728	0.05%	

News Item	451	0.03%	
Discussion	226	0.02%	
Retracted Publication	158	0.01%	
Data Paper	156	0.01%	
Correction, Addition	101	0.01%	
Poetry	99	0.01%	
Reprint	98	0.01%	
Retraction	93	0.01%	
Item About An			
Individual	90	0.01%	
Art Exhibit Review	36	0.00%	
Software Review	28	0.00%	
Bibliography	19	0.00%	
Fiction, Creative			
Prose	10	0.00%	
Database Review	7	0.00%	
Film Review	6	0.00%	
Theater Review	5	0.00%	
Excerpt	2	0.00%	
Record Review	2	0.00%	
TV Review, Radio			
Review	1	0.00%	
TOTALE	1389312	100%	

Tabella 2 - Conteggio occorrenze delle tipologie di documenti del dataset e relative percentuali

Si può notare come la grande maggioranza delle pubblicazioni universitarie sono rappresentate da *Articoli*, tipologia di documento su cui si focalizzerà lo studio, presenti nel dataset con 948.436 occorrenze su un totale di 1.389.312 e componendo quasi il **70%** del dataset. Un articolo, nella letteratura scientifica, è un lavoro scritto e pubblicato e può avere diverse finalità come lo scopo di diffusione di notizie, risultati di ricerche e analisi oppure dibattiti. A seguire abbiamo i *Meeting Abstract* con l'11,22% presenti con un totale di 155.887 occorrenze. Anch'essi sono dei documenti pubblicati con una struttura definita che deve contenere un'introduzione, l'obiettivo del paper, il metodo utilizzato e infine risultati raggiunti e conclusioni dell'autore. Proseguendo nella lista dei documenti più numerosi del dataset, si hanno i *Proceedings Papers* per un totale di 114.370 occorrenze componendo l'8,23% delle osservazioni. I proceedings papers consentono un setting meno formale rispetto ai journals sottoposti a peer review. Sono tipicamente utilizzati per presentare nuovi concepts e tecniche in campi ancora non totalmente sviluppati e diffusi. A seguire *Review, Letter, Editorial Material, Note, Correction, Book Review, Book Chapter,* e altri.

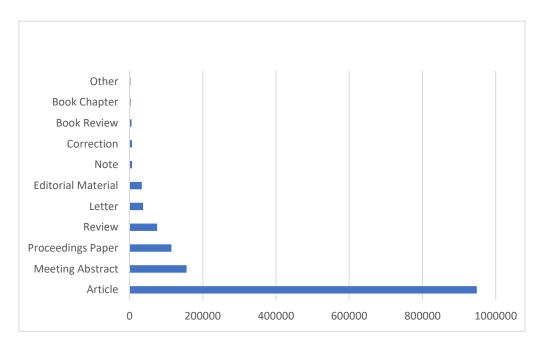


Figura 16 – Occorrenze di "document types" presenti nel dataset

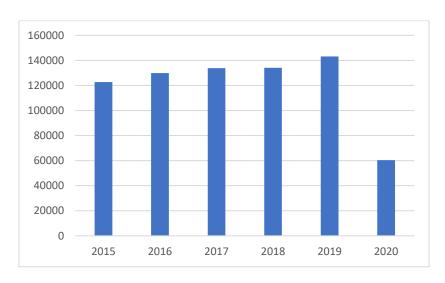


Figura 17 – Totale delle pubblicazioni suddivise per anno

Il dataset risulta tuttavia essere frammentato e non uniforme nella raccolta dati rispetto agli anni di osservazione tra i vari istituti. Non tutte le università, infatti, hanno dati che arrivano fino al 1980 come per altre. Soltanto 16 enti osservati su 98 infatti hanno un record di dati risalente al 1980.

Per poter effettuare statistiche comparative e confronti tra gli enti universitari si è deciso quindi di focalizzare l'attenzione su un intervallo di tempo congruo con i dati a disposizione. Il focus si ha quindi sull'intervallo 2015-2019, frangente di tempo in cui si hanno osservazioni e dati a disposizione per tutte le università del database. Si è scelto di omettere altresì il 2020, anno dello scoppio della pandemia dovuta al Covid-19, anno di blocco e rallentamento

generale che ha coinvolto anche il mondo didattico e accademico influenzando il numero di pubblicazioni e ricerche in via del tutto eccezionale. Inoltre, per poter approfondire l'analisi del network, la disponibilità di dati in merito al numero di citazione avute dalle pubblicazioni aiuta a determinare una sorta di misura della "qualità" e importanza dell'articolo in base alla risonanza che esso ha avuto nel mondo accademico di riferimento. Ci si focalizza quindi sui documenti di tipo "articoli" in quanto includere tipologie di pubblicazioni che non hanno una particolare diffusione nel mondo accademico andrebbe a invalidare il calcolo di indici e metriche come, per esempio, il numero medio di citazioni di una specifica università. Escludere dall'analisi i documenti poco diffusi, popolari per poche università, e che non inglobano al 100% le misure offerte dal dataset come numero di citazioni, ci porta ad una visione anche qualitativa delle pubblicazioni italiane e di conseguenza anche del livello d'impatto che ha una co-pubblicazione nel mondo accademico.

#### 4.1 Istituti osservati

Gli istituti osservati dal dataset sono in tutto 98, sparsi per tutto il suolo italiano.

Essi sono classificati come:

- Università: istituto scientifico di ordine superiore comprendente varie facoltà, ripartite secondo le specializzazioni, e alle quali si accede per il conseguimento di un titolo legalmente riconosciuto.
- **Politecnico** : istituto di studi superiori le cui facoltà sono di ambito tecnologico, tipicamente ingegneria e architettura.
- Scuola superiore ad ordinamento speciale: si tratta di Scuole Superiori Universitarie che offrono percorsi di alta formazione e ricerca che si affiancano ai corsi di laurea "comuni".
- **Università telematica**: l'università telematica è un istituto di istruzione superiore di livello universitario che eroga corsi con modalità a distanza, prevalentemente mediante l'impiego di tecnologie di comunicazione.
- Aziende ospedaliere costituite mediante trasformazione dei presidi ospedalieri nei quali insiste la prevalenza del corso di laurea in medicina e chirurgia, anche operanti in strutture di pertinenza dell'università, denominate aziende ospedaliere integrate con l'università.
- **Ente ricerca:** Gli organismi di ricerca, sono enti no profit indipendenti, pubblici o privati, il cui fine, da statuto, consiste in attività di ricerca, sviluppo tecnologico con l'obiettivo di diffusione della conoscenza.

Gli enti individuati dal dataset e analizzati per lo studio sono così suddivisi:

Tipologia	n
Università	74
Azienda Ospedaliera Universitaria	8
Scuola superiore ad ordinamento	
speciale	6
Università telematica	5
Politecnico	3
Ente Ricerca	2

Tabella 3 – Tipologie di enti presenti nel dataset con relativo conteggio delle occorrenze

Le università sono gli istituti più presenti con un totale di 74 università (~75%), a seguire ci sono le aziende ospedaliere universitarie (~8%), scuole ad ordinamento speciale (~6%), università telematiche (~5%), politecnici (~3%) e enti di ricerca (~2).

Ulteriore distinzione fra gli istituti può essere fatta in base alla classificazione tra ente Statale o non (libera). Risultano essere presenti un totale di **73** (~75%) **enti statali** e **25** (~25%) **enti privati**. Analizzando più nel dettaglio lo statuto dell'ente e sua tipologia si nota che la totalità delle università telematiche ed enti di ricerca risultano essere privati. Istituti privati, tuttavia, compaiono anche negli Ospedali e Università. Su un totale di 74 Università, 58 sono università pubbliche e 16 private costituendo così circa il 22% delle università in Italia.

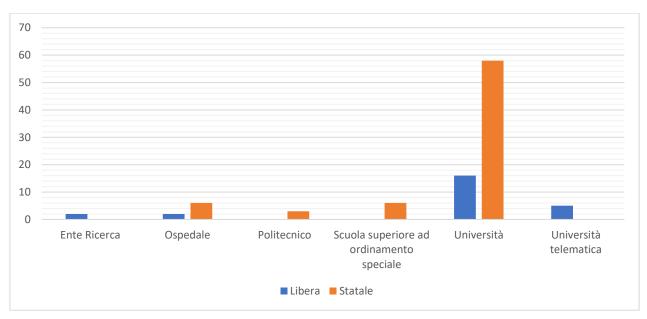


Figura 18 – Suddivisione istituti del dataset per ordinamento dell'ente

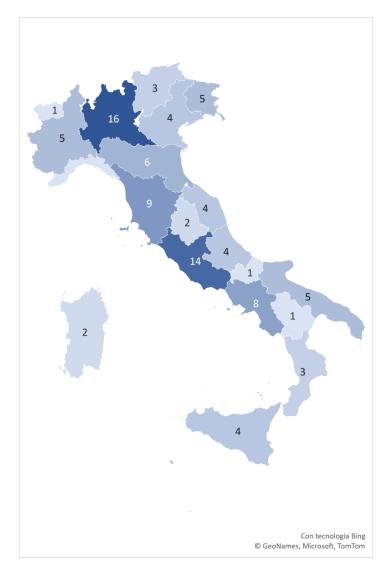


Figura 19 – Rappresentazione geografica con occorrenze degli istituti presenti nel dataset per appartenenza a regione

In totale la maggioranza degli istituti sono situati nel nord Italia con 23 enti nel Nord-Ovest, 18 Nord-Est totalizzando così 41 istituti, 42% del totale delle osservazioni. A seguire il Centro con 29 istituti, Sud con 22 e le isole Sardegna e Sicilia con un totale di 6 istituti.

Mentre la tipologia di istituto maggiormente presente in tutte e cinque le zone sono le università, entrambi gli enti di ricerca privati osservati nel database sono situati a nord est. La presenza dei 3 politecnici invece è distribuita in Lombardia, Piemonte e Puglia con il Politecnico di Milano, Politecnico di Torino e Politecnico di Bari rispettivamente.

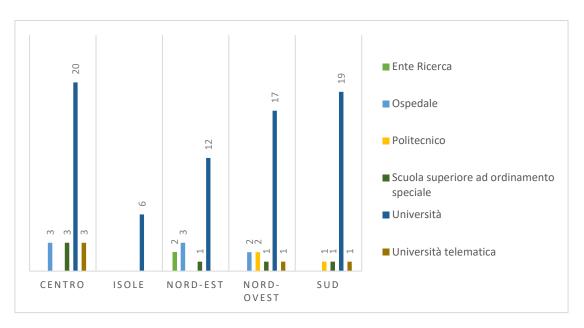


Figura 20 – Suddivisione deli istituti del dataset per macroregione

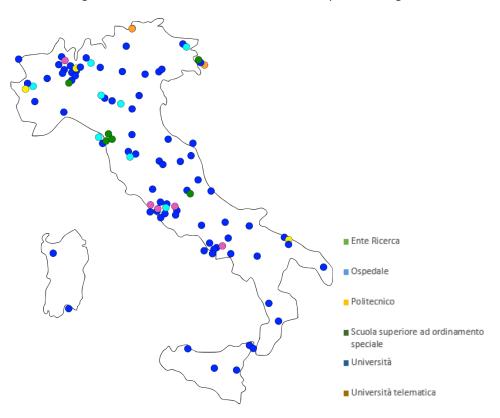


Figura 21 – Disposizione geografica degli istituti analizzati

#### 4.2 Pubblicazione articoli

Focalizzando l'analisi solo sulle pubblicazioni di soli articoli avvenute tra i 2015 e il 2019 per poter fare delle analisi comparative tra i diversi istituti, il dataset passa da un totale di 1.389.312 osservazioni di varie tipologie di documenti a **439.486 articoli**. Analizzando l'andamento della ripartizione degli articoli per anni si denota una crescita del 24% delle pubblicazioni dal 2015 al 2019 che passano da un totale di 78.898 del 2015, con una media di 805 articoli per istituto, ad un totale di 97.979 del 2019, con una media di pubblicazioni per istituto che passa così a 1000. In media ogni ente ha pubblicato 200 articoli in più nel 2019 rispetto al numero di articoli pubblicati cinque anni prima.

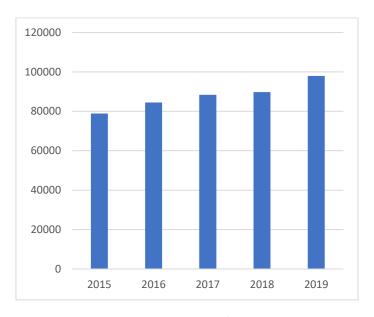


Figura 22 – Pubblicazioni articoli relativi all'intervallo di anni 2015-2019

Nell'intervallo di analisi, quindi, il numero di pubblicazioni in Italia è in crescita anno dopo anno, presentando un particolare aumento tra il 2018 e il 2019 dove il numero di pubblicazioni cresce di 8218 articoli (+10%). La crescita più contenuta invece è avvenuta tra il 2017 e il 2018 con un aumento di sole 1410 unità.

L'aumento, anno dopo anno, del numero di articoli pubblicati dalle università ed enti italiani è indice della loro attività e testimonia l'aumento della ricerca sostenuta anche da diversi progetti e finanziamenti. Ne è un esempio il *PNR* (*Programma nazionale per la ricerca 2015-2020*) con lo stanziamento di fondi pubblici per la ricerca volti ad incentivare programmi innovativi per aumentare il numero di ricercatori in Italia e dall'estero di cui si è analizzato il programma nelle sezioni antecedenti.

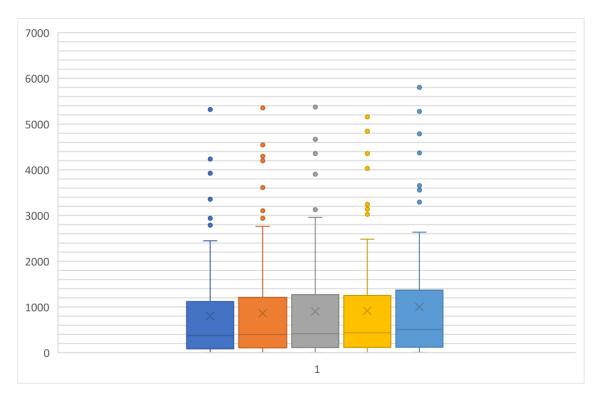


Figura 23 — Boxplot raffiguranti le pubblicazioni di articoli suddivisi per anni. Da sinistra verso destra: 2016,2016,2017,2018.

Il trend di crescita della pubblicazione di articoli è confermato anche dai diagramma a scatola in **Figura 23** che consente di analizzare visivamente e con semplicità la dispersione dei dati. La mediana, indicata con una linea orizzontale all'interno della scatola testimonia la non simmetricità dei dati in quanto essa non posizionata esattamente al centro di essa. Il gran numero di articoli prodotti annualmente, dunque, si trova al di sopra della mediana. Il leggero spostamento della x (rappresentante la media) durante gli anni conferma la crescita, seppur contenuta, della media di articoli pubblicati annualmente. Gli *outliers*, ovvero i punti che ricadono al di fuori dei "baffi" del grafico, ovvero al di sopra dell'1,5 \* range interquartile, e il loro numero costante testimoniano la presenza di poche università "leader" nelle pubblicazioni di articoli individuabili nelle prime 7 università in **Tabella 4** che si riconfermano negli anni.

La media di articoli pubblicati annualmente da un ente dal il 2015 e il 2019 passa da 805 a 1000. Gli istituti più prolifici nella pubblicazione di articoli sono riportati nella tabella sottostante dove sono riportate le prime 10 Università con un maggior numero di pubblicazioni nei cinque anni di riferimento. Tra le università che hanno rilasciato più articoli, sul podio abbiamo alcune delle università più longeve in Italia: "La Sapienza" con un totale di 5804 articoli pubblicati nel 2019, seguita dall'Università di Padova e Università di Bologna rispettivamente con 5276 e 4786 articoli in media rispettivamente nel 2019.

	2015	2016	2017	2018	2019
Roma La Sapienza		5355	5371	5158	5804
Padova	4237	4545	4667	4844	5276
Bologna	3924	4294	4414	4362	4786
Milano	3942	4196	4354	4354	4841
Napoli Federico II	3358	3613	3903	4029	4371
Torino	2938	3103	3166	3142	3557
Firenze	2788	2941	3130	3246	3654
Pisa	2446	2762	2960	3026	3292
Genova	2204	2276	2352	2482	2635
Roma Tor Vergata	2229	2342	2359	2351	2498

Tabella 4 - Primi 10 istituti per numero di articoli pubblicati negli anni

Nonostante gli istituti privati in Italia abbiano una notevole presenza rappresentando il 25% delle occorrenze del dataset, dalla **Figura 24** sottostante si denota chiaramente come il maggior numero delle pubblicazioni di articoli è dovuta agli istituti pubblici (in arancione) rispetto ad istituti privati (in blu). Più nello specifico, prendendo il 2019 come anno di riferimento, il totale di articoli pubblicati da istituti privati ammonta soltanto a 6150 costituendo il 6,3% del totale degli articoli pubblicati (97979).

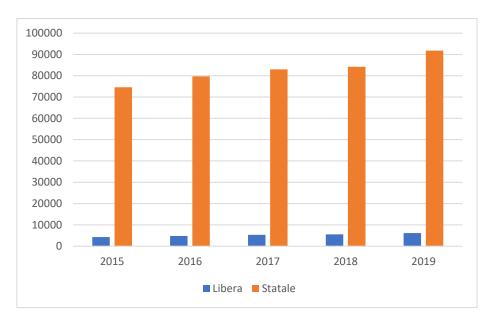


Figura 24 – numero di pubblicazioni di articoli suddivisi per anni e per ordinamento

Analizzando l'andamento delle pubblicazioni di articoli nell'intervallo di anni che va dal 2015 al 2019, si nota una crescita di pubblicazioni sia da parte di enti pubblici che privati. Ad aver avuto una maggiore crescita nelle popolazioni, però, sono stati gli enti privati passando da un totale di 4297 a 6150 con una crescita del 43% contro il 23% degli enti pubblici. Nello specifico la crescita delle pubblicazioni degli istituti privati è da imputare maggiormente alle università con +1651 articoli rispetto al 2015. Anche nel caso degli enti pubblici, a far crescere in modo più pronunciato il numero delle pubblicazioni sono le università con +15.634 articoli (+23%) ma a crescere maggiorente sono state le pubblicazioni delle aziende ospedaliere cresciute del 38%.

Per avere idea di come sia cambiato il panorama italiano in marito al numero di pubblicazione di articoli si può fare riferimento alla **Figura 25**. Gli enti sono riportati sulla mappa italiana con cerchi la cui dimensione, grandezza del carattere e intensità di colore sono direttamente proporzionali al numero di articoli pubblicati nell'anno di riferimento, 2015 a sinistra e 2019 a destra.

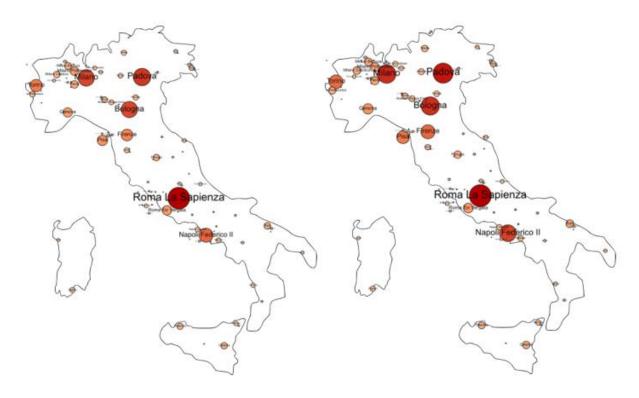


Figura 25 – Rappresentazione grafica della pubblicazione di articoli in Italia nel 2015 (a sinistra) e nel 2019 (a destra)

Non emergono particolari crescite e sproporzioni da un anno all'altro. Si nota una maggiore intensità di colore e grandezza del nodo generalmente per tutte le università del nord Italia e per l'università di Napoli Federico II, inoltre, troviamo conferma di quanto riportato nella **Tabella 4** in merito alle università con un grande numero di articoli pubblicati.

Tra gli enti che hanno avuto un maggiore incremento percentuale di articoli pubblicati nell'intervallo 2015-2019 troviamo principalmente università telematiche, università private e aziende ospedaliere universitarie. Si trovano infatti, tra i primi posti, la Libera Università Mediterranea Jean Monnet, (+1100%), Rozzano Humanitas University (+948%), Università di Siena per stranieri (+500%), Unitelma Sapienza(+250%). E' importante evidenziare però che queste elevate crescite percentuali sono dovute al fatto che le pubblicazioni in numero assoluto di questi enti è relativamente basso se confrontato con le istituzioni del dataset, pertanto anche pochi articoli in più rispetto all'anno precedente si riflettono in un aumento in termini percentuali elevato. Ad esempio, analizzando la crescita della Libera Università Mediterranea Jean Monnet del 1100% troviamo u numero di articoli che passa da 1 a 12 oppure quella dell'università telematica Unitelma Sapienza che in cinque anni ha avuto una crescita di soli 20 articoli confrontando 2015 e 2019. Da sottolineare invece a crescita di pubblicazioni di Rozzano Humanitas University che ha quasi decuplicato le sue pubblicazioni passando da 33 a 346.

#### 4.2.1 Analisi citazioni

Disporre del numero di citazioni avute per ogni articolo ci consente di determinare, facendo una media di tutti gli articoli, quale sia l'università/ente italiano con un maggiore impatto accademico. Un articolo che approfondisce tematiche attuali, che porta sul campo della

ricerca nuove scoperte che possono aprire la porta a svariati contesti e applicazioni ha più probabilità ad essere citato un gran numero di volte.

Al contrario del numero di articoli pubblicati, in continua crescita anno dopo anno, il numero medio di citazioni annualmente scende drasticamente come messo in evidenza dalla **Figura 26**. Il numero medio di citazione va a decimarsi in 5 anni passando da una media di 20.78 a 2.74 C'è da evidenziare, però, come il grado di "maturazione" dell'articolo possa influenzare il numero di citazioni registrate. Di fatti, è più probabile che un articolo più datato abbia registrato nel corso del tempo una maggiore visibilità, e quindi un numero maggiore di possibili citazioni, rispetto a quanto ne abbia potuto avere invece un articolo più recente.

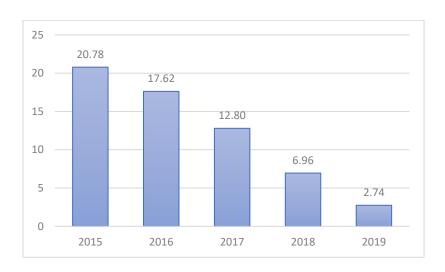


Figura 26 – numero di citazioni medie per articolo per anni

Focalizzando l'attenzione sul numero di citazioni medie attribuite per istituto nei vari anni si assiste ad una maggiore variabilità. Alcuni nomi, tuttavia, continuano a comparire tra le università ed enti più citati.

Top cit. 2015	cit. medie	Top cit. 2016	cit. medie
Napoli Pegaso - telem	127.83	Novedrate e-Campus	81.11
University Hospital of Siena	91.10	L'Aquila GSSI	79.59
Roma Marconi - telem	49.40	Trieste SISSA	47.31
Bari Politecnico	33.81	Napoli Pegaso - telem	43.56
Perugia	32.20	Urbino	39.03
L'Aquila GSSI	30.13	Sannio	35.12
Siena	29.44	Camerino	34.48
		European Academy of Bolzen-	
Napoli Parthenope	29.21	Bolzano	32.13
Trieste SISSA	29.08	Rozzano Humanitas Univerisity	29.62
Pisa Normale	28.57	University Hospital of Parma	26.92
Top cit. 2017	cit. medie	Top cit. 2018	cit. medie
L'Aquila GSSI	41.43	Napoli Pegaso - telem	14.13

Sannio	39.18	Trieste SISSA	13.25
Napoli Pegaso - telem	37.61	L'Aquila GSSI	12.80
Urbino	35.56	Roma Marconi - telem	12.15
Camerino	29.47	Rozzano Humanitas Univerisity	11.85
Pisa Normale	26.43	Casamassima - J.Monnet	11.20
University Hospital of Siena	25.22	Milano San Raffaele	11.06
Rozzano Humanitas Univerisity	22.40	University Hospital of Siena	10.01
Roma Marconi - telem	21.30	Roma UNITELMA - telem	10
Salerno	20.56	Milano Bicocca	9.57
Top cit. 2019	cit. medie		
Humanitas Hospital Gradenigo	5.10		
L'Aquila GSSI	5.05		
Rozzano Humanitas Univerisity	4.68		
Roma UNITELMA - telem	4.40		
Napoli Pegaso - telem	4.22		
Urbino	4.17		
Istituto Clinico Humanitas			
Gavazzeni	4.02		
Camerino	3.95		
Perugia Stranieri	3.88		
Sannio	3.80		

Tabella 5 – Primi 10 istituti per citazioni

Ad esempio, osservando i dati della **Tabella 5**, dove per ogni anno dell'arco temporale di osservazione sono stati presi gli enti e le università più citate in media possiamo notare come ad esser più frequente nella classifica delle 10 università più citate in media ci sia una grande occorrenza di università telematiche e aziende ospedaliere universitarie. Il motivo per cui compaiono enti e università non particolarmente produttive in termini di numeri di pubblicazioni tra le università/enti più citati in media è dovuto, probabilmente, al grande numero di pubblicazioni che i maggiori istituti hanno e conseguentemente fanno abbassare il numero di citazioni medie per articolo. Inoltre, una focalizzazione su un determinato tema o area di ricerca comporta una concentrazione degli sforzi e una ricerca più intensa su campi specifici, alzando, di conseguenza, le probabilità che la pubblicazione abbia un impatto maggiore nella comunità scientifica e letteratura di riferimento. Non deve sorprendere quindi la presenza di enti come Scuole ad ordinamento speciale, Enti di ricerca, Aziende ospedaliere universitarie (Università ospedaliera di Siena, Università ospedaliera di Rozzano), focalizzate sulla ricerca medica.

Il GSSI Gran Sasso Science Institute, ad esempio, una scuola superiore universitaria ad ordinamento speciale nata come istituto di ricerca principalmente incentrata sullo studio di

fisica nucleare, pur avendo una media di 291 articoli all'anno rientra 5 anni consecutivi nella lista dei 10 enti più citati.

# 4.3 Co-pubblicazioni

Le co-pubblicazioni tra gli istituti, ovvero le pubblicazioni in cui hanno contribuito più di due istituti, sono stati identificate sulla base del codice *Accession.Number*, univoco per ogni articolo e pubblicazione. Di conseguenza, uno stesso codice identificativo presente in due dataset appartenenti a istituti differenti definisce una collaborazione tra i due enti.

In totale sono state identificate 12.978 collaborazioni nei cinque anni di analisi, intendendo per collaborazione una coppia di due istituti che hanno co-pubblicato almeno un articolo nell'anno. Suddividendo le 12.978 coppie in base all'anno di pubblicazione degli articoli possiamo analizzare come, negli anni osservati, il grado di collaborazione tra gli istituti sia cambiato e se si osservano eventualmente tendenze crescenti, decrescenti o stazionarie negli anni.

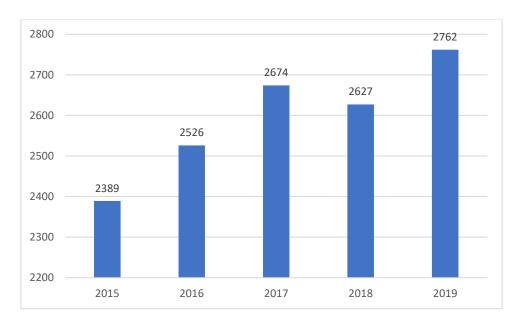


Figura 27 – Collaborazioni nella pubblicazione di articoli suddivise per anno

Analizzando la **Figura 27** si può identificare una crescita generale del numero di collaborazioni per la pubblicazione di articoli complessiva dal 2015 al 2019. Il numero di collaborazioni cresce anno dopo anno dal 2015 al 2017 per poi avere un decremento nel 2018 di 47 collaborazioni rispetto all'anno precedente e per poi aumentare nuovamente superando quelle del 2017 arrivando ad un totale di 2762. Dal 2015 al 2019 non si assiste quindi ad un grande incremento di collaborazioni che restano in grandi linee stabili con un piccolo incremento passando da 2389 a 2762 aumentando del 15%.

Mettendo in relazione il numero di articoli pubblicati collaborando insieme ad altri istituti con gli articoli totali si riesce a determinare il grado di collaborazione di ciascun ente del dataset.

Calcolando le percentuali ed estrapolandone la media, il valore massimo e il minimo si nota una grande variabilità. I valori massimi di collaborazione pari al 100% (vale a dire che ogni articolo pubblicato dall'istituto sia stato frutto di collaborazione con almeno un altro ente) si trovano in tutti e cinque gli anni per le aziende ospedaliere integrate ad un'università come quella di Udine, Modena e Reggio Emilia, di Siena, Pisana, della Sapienza e Parma. Ragionevolmente, tutti articoli co-pubblicati con almeno l'università di riferimento. Alti livelli di collaborazione si trovano anche nelle università telematiche come Napoli Pegaso, Roma Marconi, Novedrate, per l'ospedale Humanitas Gradenigo, con una media negli anni dell'88%. L'ipotesi più accreditata per spiegare un gran numero di università telematiche tra gli istituti con il più alto numero di co-pubblicazioni tra i propri articoli può è dovuta alla scarsa tecnologia, personale e laboratori a disposizione per poter sviluppare in autonomia uno studio, perlomeno per ciò che concerne l'ambito scientifico.

Tra gli enti meno collaborativi invece si trovano istituti politecnici come quello di Milano e Torino, con una media collaborativa dal 2015 al 2019 rispettivamente del 27% e 29% degli articoli in co-pubblicazione. L'università Orientale di Napoli e Bocconi di Milano oltre che l'università per stranieri di Siena.

In linea generale, i 98 istituti hanno in media una collaborazione che si mantiene stabile al 58%. Volendo mettere in evidenza gli istituti che hanno avuto una crescita accentuata in termini di collaborazioni negli anni, è da citare: l'Università di Napoli Orientale, l'Università di Scienze Gastronomiche di Bra, Università di Brescia, Cattolica di Milano, Università di Foggia, l'istituto clinico Humanitas Gavazzeni. Tra quelle invece che hanno subito un decremento maggiore nel rapporto tra articoli in collaborazione e propri, propendendo quindi verso una maggiore autonomia e un minor effort collaborativo, si hanno: Unitelma Sapienza, Università della Valle d'Aosta, Università Carlo Cattaneo, Università degli Studi Suor Orsola Benincasa di Napoli, la Libera Università Mediterranea Jean Monnet, Istituto universitario di studi superiori di Pavia, Scuola Internazionale Superiore di Studi Avanzati (SISSA).

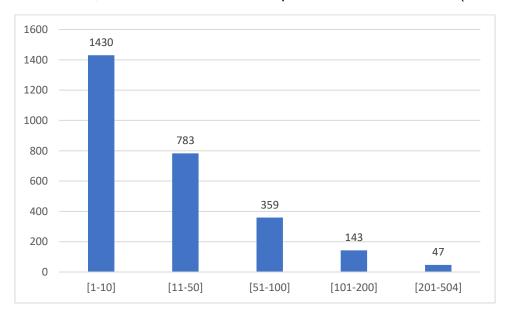


Figura 28 – Numero di collaborazioni dell'anno 2019 suddiviso per quantità di articoli pubblicati.

Analizzando il numero di articoli co-pubblicati per ciascuno delle 2762 collaborazioni avute nell'anno 2019 emerge che, sul totale delle coppie che nel 2019 hanno avuto una co-pubblicazione di articoli 1430 ovvero poco più della metà (51%) abbiano pubblicato un numero di articoli che va da 1 a 10. Il range va da 11 a 504 è da imputare quindi alle restanti 1332 coppie, ovvero al 49% delle collaborazioni totali avute nell'anno. Risulta quindi evidente come il numero di articoli pubblicati tra le coppie sia altamente disomogeneo. Ovvero poche coppie contribuiscono con un grande numero di pubblicazioni.

Ponendo l'attenzione sulla natura delle diverse collaborazioni in merito alla tipologie di istituti coinvolti e calcolandone il numero di volte due istituti si siano messi a cooperare negli anni si scopre, senza sorpresa, che gli istituti che tra essi collaborano di più sono le università, con un numero di collaborazioni nel 2019 pari a 1782 (64% del totale). A seguire abbiamo le collaborazioni tra ospedali e università, che ha un netto aumento negli anni, arrivando ad un totale di 312 collaborazioni contate nel 2019. Poco probabile è la collaborazione tra due università telematiche. Guardando la **Tabella 6** infatti, questo tipo di collaborazione si trova all'ultimo posto con un solo caso nel 2019.

Collaborazione	2015	2016	2017	2018	2019
Università - Università	1564	1639	1714	1702	1782
Ospedale - Università	266	287	313	283	312
Scuola superiore ad ordinamento speciale - Università	192	203	217	221	226
Politecnico - Università	144	153	160	157	170
Università - Università telematica	109	114	136	122	125
Università - Ente Ricerca	53	60	59	69	66
Ospedale - Ospedale	14	19	22	17	22
Politecnico - Scuola superiore ad ordinamento speciale	14	13	14	15	15
Scuola superiore ad ordinamento speciale - Scuola superiore ad					
ordinamento speciale	7	9	10	8	11
Scuola superiore ad ordinamento speciale - Università telematica	6	6	4	5	2
Politecnico - Università telematica	5	4	5	7	6
Ente Ricerca - Politecnico	3	2	3	3	2
Ente Ricerca - Scuola superiore ad ordinamento speciale	3	3	2	3	4
Ospedale - Scuola superiore ad ordinamento speciale	3	5	5	4	7
Politecnico - Politecnico	3	3	3	3	3
Ente Ricerca - Ospedale	1	1	2	1	3
Ente Ricerca - Università telematica	1	1	1	1	
Ospedale - Politecnico	1	3	1	2	2
Ospedale - Università telematica		1	2	3	3
Università telematica - Università telematica			1	1	1
Totale complessivo	2389	2526	2674	2627	2762

Analizzando il numero di collaborazioni e pubblicazioni tra le diverse tipologie di istituti ed enti nel corso dei diversi anni, si nota una generale crescita tra tutte le diverse collaborazioni. In particolare, le collaborazioni tra le Università, sono incrementate del 14% circa passando da un totale di 1564 collaborazioni nel 2015 ad un totale di 1782 nel 2019.

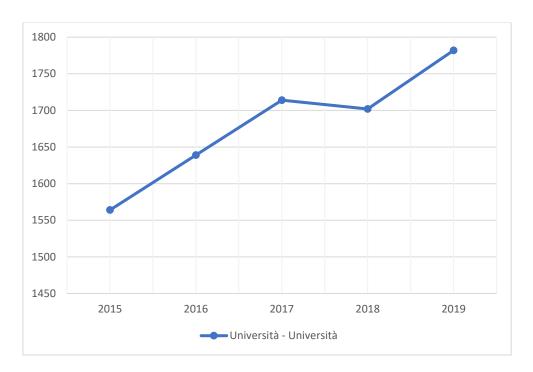


Figura 28 – Andamendo collaborazioni tra università negli anni

Focalizzando l'attenzione sulla giurisdizione degli istituti, invece, si evince una grande disparità tra il numero delle collaborazioni in cui partecipa un ente pubblico rispetto alle collaborazioni con almeno un ente privato. Questa disparità non sorprende dal momento che gli enti pubblici in Italia sono più numerosi e produttivi in termini di pubblicazioni.

	2015	2016	2017	2018	2019
Libera - Libera	144	170	252	215	343
Statale - Libera	6106	7410	8247	8588	9700
Statale - Statale	63169	68654	71484	69968	71262
Totale complessivo	69419	76234	79983	78771	81305

Tabella 7 - Numero di articoli pubblicati per anno in base al tipo di collaborazione

In particolare, è interessante osservare come anno dopo anno gli enti pubblici collaborano sempre più con gli enti privati. Gli articoli pubblicati in co-pubblicazione tra enti pubblici sono cresciuti dal 2015 al 2019 di 8093 unità (+13% circa) mentre le collaborazioni tra enti pubblici

e privati sono cresciute in valore assoluto di 3594 unità, numero più basso della crescita di articoli in collaborazioni pubbliche-pubbliche, ma se analizziamo la crescita in termini percentuali rispetto al 2015 si ha un incremento del +58,8%. Nei cinque anni si assiste anche ad una sostanziale crescita di collaborazioni tra enti privati, passando da un totale di 144 co-pubblicazioni nel 2015 ad un totale di 343 nel 2019 con una crescita del +138%.

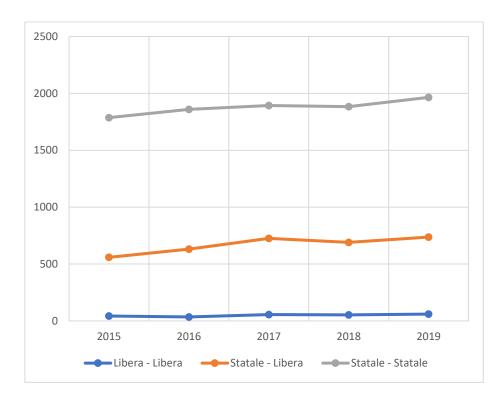


Figura 29 - Andamento nel tempo del numero di collaborazioni tra tipologie di ordinamenti

In termini di efficacia della collaborazione possiamo analizzare la media delle citazioni per articolo per ogni tipologia di collaborazione. Quello che si osserva è che oltre ad essere più numerose, le collaborazioni tra enti pubblici danno un distacco di quasi 16 citazioni in più ad articolo in media.

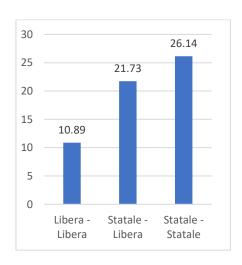


Figura 30 – Media delle citazioni relative alle diverse collaborazioni

### 5 Network Visualization

Disponendo, mediante l'ausilio del software Gephi, gli enti (individuati dai nodi) secondo la loro localizzazione geografica<sup>1</sup>, inserendo nel dataset l'indicazione di latitudine e longitudine, e individuando le collaborazioni tra essi è possibile visualizzare graficamente il network collaborativo tra gli istituti italiani in termini di co-pubblicazione di articoli. La dimensione del singolo nodo (istituto) è direttamente proporzionale al numero di articoli pubblicati dall'istituto, mentre la sua colorazione indica la categoria di cui fa parte. Indicati col colore blu abbiamo le Università, in azzurro gli ospedali convenzionati con le università, in verde le scuole superiori ad ordinamento speciale, in marroncino le università telematiche, giallo per i politecnici e infine il verde per gli istituti di ricerca. Dando una rapida occhiata alla **Figura 31**, è possibile confermare anche visivamente, quanto mostrato in precedenza. Ovvero si noti come i nodi dai quali partono più collegamenti, nonché collaborazioni, siano quelli di dimensione maggiore, ovvero, maggiormente produttivi in termini di articoli pubblicati (in questo caso nell'anno 2019).

\_

<sup>&</sup>lt;sup>1</sup> Si è optato, durante la creazione grafica del network, si applicare la funzione "Label Adjustment" evitando così che i nodi si sovrappongano l'uno sull'altro non dando la giusta idea in merito alla numerosità degli istituti. Di conseguenza, i nodi del grafo (rappresentanti i vari istituti analizzati) non occupano la loro reale posizione geografica precisa.

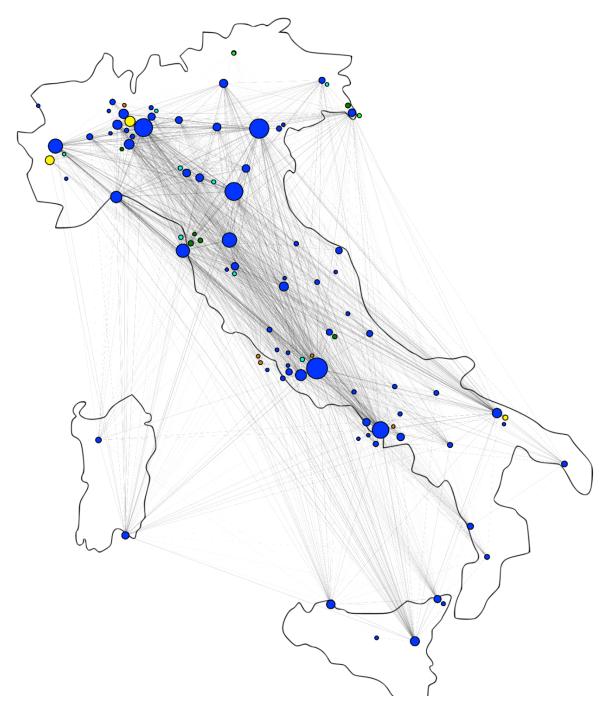


Figura 31 - Grafo delle collaborazioni per l'anno 2019 rappresentato in cartina

Per avere una visione complessiva del network e delle sue principali caratteristiche è possibie disporre il grafo delle collaborazioni, sempre mediante il software *Gephi*, secondo diverse disposizioni e algortmi differenti, volti a far emergere visivamente alcuni tratti distintivi e metriche di un network.

Disponendo il grafo delle collaborazioni avute nell'anno 2019 secondo l'algoritmo che va sotto il nome di *Force Atlas* 2 si ottiene ciò che è mostrato in Figura32. ForceAtlas2 è un layout diretto dalle forze: simula un sistema fisico per spazializzare una rete, i nodi si

respingono come particelle cariche, mentre i bordi attraggono i loro nodi, come molle. Tutto questo sistema di forze converge dopo qualche secondo di elaborazione e calcoli verso uno stato di equilibrio aiutando l'interpretazione dei dati mettendo visivamente in risalto ipotetici moduli nell rete, gruppi di nodi più collaborativi, nodi di particolare importanza in termini di metriche di centralità. L'algoritmo Force Atlas 2, è spesso utilizzato quando si ha la necessità di analizzare le caratteristiche di una rete con numerosi nodi.

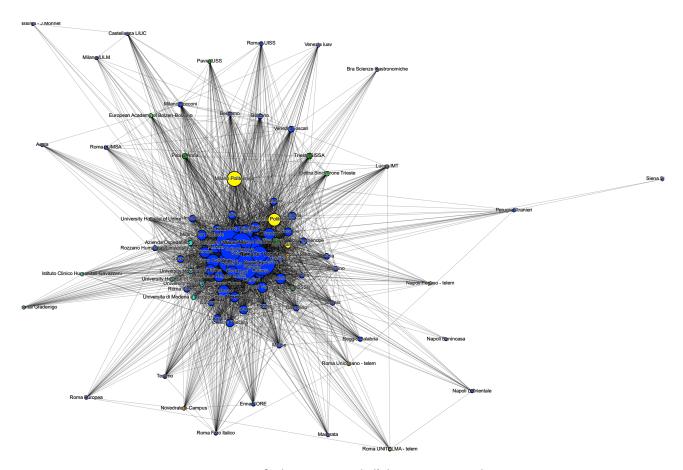


Figura 32 - Grafo disposto secondo l'algoritmo Force Atlas 2

Nonostante la numerosità dei nodi della rete dalla disposizione del grafo, ottenuto mediante l'algoritmo Force Altlas 2, non emergono cluster e moduli ben definiti, come riportato anche nella sezione dedicata alla metrica Modularity, bensì il grafo si dispone quasi come un'unica nube di nodi occupata centralmente dalle università (in blu) con più articoli pubblicati. Di fatto, impostando il dimensionamento dei nodi in base al numero di articoli pubblicati nell'anno, i nodi centrali al grafo hanno una dimensione maggiore e a giudicare dalla fitta rete hanno tra loro una maggiore intensità collaborativa. Occupano una posizione più esterna e marginale alla rete, invece, oltre alle piccole università anche tutto il resto di istituti ed enti presenti nel dataset, come gli enti di ricerca, i politecnici, istituti ospedalieri e scuole

ad ordinamento speciale. La marginalità degli istituti come enti di ricerca, istituti ospedalieri, università telematiche e scuole ad ordinamento speciale è un risultato che non sorprende in quanto, avendo riscontrato come seppur negli anni abbiano avuto una notevole crescita sia in termini di pubblicazioni che collaborazioni, il loro contributo nella letteratura non è paragonabile agli istituti universitari.

L'assenza, nel grafo, di moduli e gruppi di istituti a sé stanti non è una sorpresa, in quanto, ulteriore conferma della disposizione della metrica degree centrality dei vari nodi. La maggioranza dei nodi infatti (circa il 40%) ha avuto un numero di collaborazioni che va da 51 a 75. Il numero medio di degree centrality individuato dal software, infatti, ammonta a 56. Ovvero, in media, ogni istituto ha collaborato con 56 enti differenti nel 2019. Tenendo conto che il dataset comprende la totalità di 98 istituti risulta chiaro il motivo per il quale non emergano particolari informazioni dalla disposizione del grafo secondo l'algoritmo Force Atlas 2.

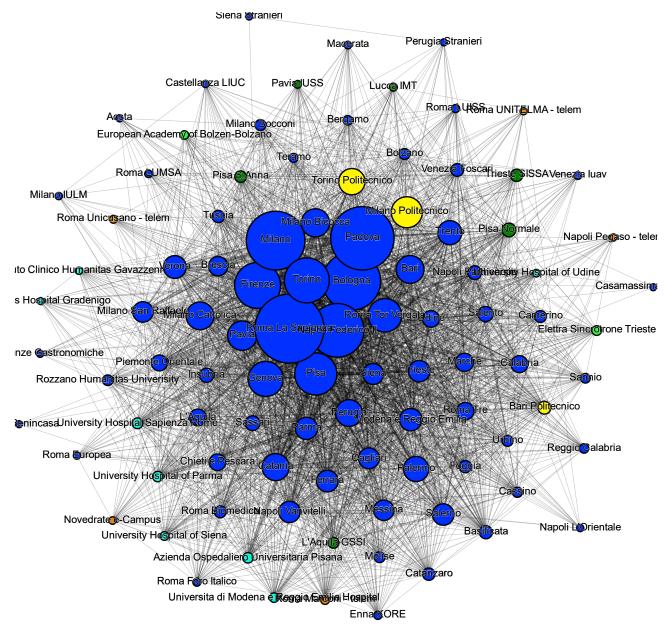


Figura 33 - Grafo disposto secondo l'algoritmo Fruchterman-Reingold.

Altro algoritmo utilizzato nei contesti di Social Network Analysis, è l'algoritmo Fruchterman-Reingold. Esso, dispone gli attori particolarmente coinvolti e attivi nella rete centralmente mentre gli attori meno relazionati occuperanno una posizione periferica rispetto al grafo. Il risultato conferma ulteriormente quanto analizzato precedentemente. Ad avere una maggior propensione alla collaborazione sono gli istituti universitari, rappresentati in blu, occupanti il cuore del grafo disposto. Maggiore è il loro apporto in termini di pubblicazioni rappresentato dalla dimensione del nodo stesso, maggiori sono gli articoli pubblicati in cooperazione. Ecco perché le università con una più grande pubblicazione annuale (in grafo è rappresentato l'anno 2019) occupano una posizione più centrale rispetto alle altre università. Si noti inoltre, come la rete di collaborazioni diventa più fitta man mano che ci si avvicina al cuore del grafo. Non solo le collaborazioni aumentano, rendendo, di fatto, il grafo più denso e numeroso di collegamenti, ma aumenta anche lo spessore dell'arco stesso. Un arco, infatti, è tanto più spesso quanto maggiore è stato il numero di articoli pubblicati tra i due enti del collegamento stesso.

Alle marginalità del grafo troviamo, oltre alle piccole università, la maggior parte degli istituti come enti di ricerca, università telematiche, politecnici, scuole superiori ad ordinamento speciale e ospedali. Istituti che, in quanto di piccola entità, hanno registrato un piccolo numero di articoli pubblicati e, di conseguenza, con un minor apporto collaborativo rispetto ad istituti più centrali.

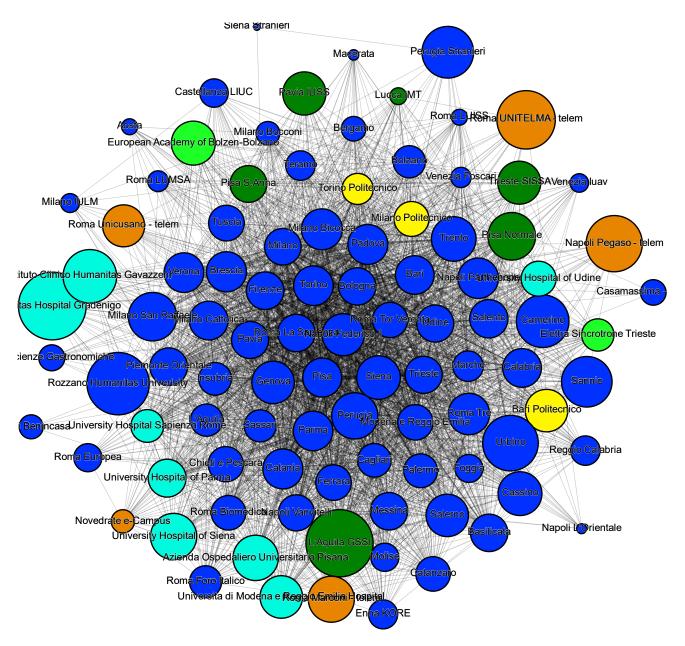


Figura 34

Utilizzando sempre l'algoritmo di Fruchterman-Reingold, ma applicando, questa volta, un criterio di dimensionamento dei nodi proporzionale alle citazioni medie ricevute dagli articoli (pubblicati nel 2019) si ottiene il grafo in **Figura 34**. La differenza sostanziale che emerge confrontando con il grafo in **Figura 33**, è il posizionamento dei nodi più grandi. Gli istituti con una pubblicazione minore, con una più bassa collaborazione per la pubblicazione di articoli e meno integrati nella rete, infatti, hanno un numero di citazioni medie per articolo maggiore rispetto gli istituti e le università centrali al grafo e quindi più "immersi" nella rete collaborativa. Vale, ovviamente, quanto già detto in merito al numero di citazioni medie. La ragione del contrasto tra le due figure è dovuto, in prima battuta, al grande numero di pubblicazioni degli istituti universitari facendo abbassare la media di citazioni per articolo. Inoltre, gli istituti speciali, enti di ricerca e ospedali hanno una focalizzazione tematica e in

termini di ricerca che porta, probabilmente, ad una pubblicazione scientifica più risonante nel mondo accademico, che aumenterebbe quindi il numero di citazioni medie per articolo.

### 6 Metriche

In questo capitolo analizzeremo alcune delle metriche che l'approccio Social Network Analysis è in grado di offrire e discusse nel capitolo dedicato, per comprendere meglio e analizzare nel dettaglio le caratteristiche della rete. Si analizzeranno, in particolare, due metriche a livello network (*Network Density, Modularity*) e tre metriche di centralità (*Degree Centrality, Betweenness Centrality, Closeness Centrality*) relative alla rete così come si presentano per il 2019, anno più recente, per un'analisi il quanto più puntuale della rete di collaborazione oggi. In seguito, focalizzeremo meglio l'attenzione sull'intervallo di anni 2015-2019 determinando quali siano stati i principali cambiamenti, se presenti, nei cinque anni analizzati dallo studio.

## 6.1 Network Density

Per poter fare delle considerazioni in merito al grado di collaborazione tra gli istituti italiani possiamo confrontare il numero di coppie totale possibili dato un insieme con il numero di coppie effettivamente risultanti dai dati a disposizione, calcolando quello che di fatto si chiama la *Network Density* di un grafo, o la sua densità.

Utilizzando il software *Gephi*, infatti, il livello di Network Density per l'anno 2019 definito, risulta essere pari a **0,581**. Calcolando la metrica anche per gli altri anni di analisi, i valori risultanti sono: 0,503 (2015), 0,531 (2016), 0,563 (2017), 0,553 (2018) e 0,581 (2019).

Risulta evidente, quindi, come su una totalità di circa 4000 possibili combinazioni di coppie realizzabili, ne viene effettivamente registrata circa la metà. Il numero di nuove collaborazioni stipulate tra gli enti è tuttavia in crescita, come osservabile dalla **Figura 35** dove sono riportati i valori di network density per ogni anno dell'osservazione individuando un possibile trend di crescita, nonostante la leggera contrattura registrata nell'anno 2018.

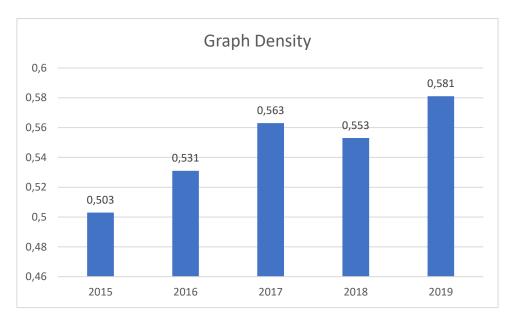


Figura 35 - Andamento graph density negli anni

## 6.2 Modularity

Il valore di *modularity* della rete collaborativa associato all'anno 2019 risulta pari a **0,086** individuando un totale di 4 moduli. Il valore della metrica è stato calcolato tenendo conto anche del peso degli archi, ovvero del numero di articoli in collaborazione tra due istituti nell'anno di riferimento.

I moduli, ovvero i gruppi di nodi tra loro più interconnessi, risultano non essere messi particolarmente in evidenza anche con la disposizione del grafo con l'algoritmo *Force Atlas* 2 che dovrebbe mettere in risalto i vari moduli della rete, dando ulteriore conferma del basso valore di modularità ad essa associato. La grandezza di queste comunità è variabile, infatti il software determina (in ordine decrescente di grandezza) un modulo più grande (evidenziato in arancio in **Figura36**) con il 40,82% dei nodi, il secondo modulo (in azzurro) riporta il 32,65% dei nodi totali, a seguire il terzo modulo (in verde) con il 16.33% e il quarto (viola) con il 10.2%.

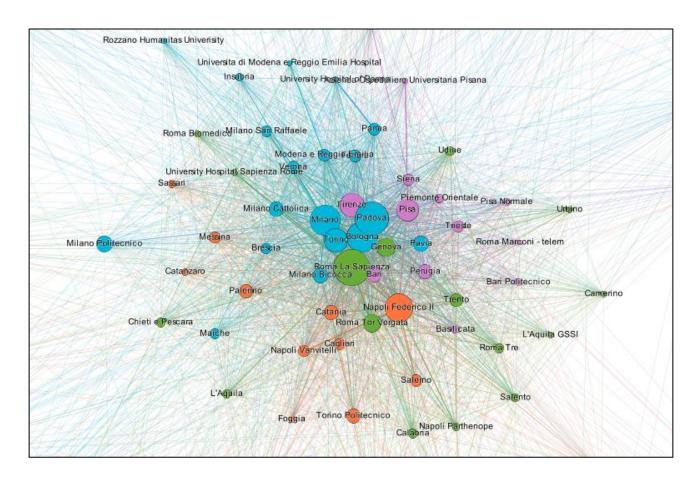


Figura 36

Avendo riportato un valore di modularity molto basso tendente allo zero, come anticipato, pur utilizzano un apposito algoritmo di disposizione non emergono alla vista moduli particolarmente ben distinti e separati. Aiutandoci con una colorazione differente dei moduli possiamo evidenziare quelli che sono i principali nodi dei diversi moduli. Disponendo i nodi geograficamente ed evidenziando i diversi moduli individuati dal software possiamo trovare una correlazione tra localizzazione geografica e appartenenza al modulo come mostrato in figura. Possiamo individuare infatti moduli collocati sulla parte più nordica del paese come il modulo in viola e azzurro. Un modulo situato nel sud Italia, in arancione e che mette in relazione principalmente le università campane oltre che alcuni istituti situali nelle isole e del nord Italia. Infine, si individua un modulo, quello evidenziato col colore verde, situato in centro Italia che abbraccia le molteplici università della regione Lazio e alcuni istituti del sud e nord Italia.

È bene comunque tenere in mente, il basso valore di *modularity* (tendente allo zero) che il software ha calcolato per la rete in analisi non individuando quindi una suddivisione netta e definita dei moduli. Infatti, si noti, come molte delle università e degli enti rappresentati abbiano numerose collaborazioni anche con istituti non appartenenti al proprio cluster e quasi della stessa entità. Particolarmente "contaminate" sono le zone del Nord Ovest. Inoltre, non vi è una netta distinzione tra il modulo in verde e arancio. Di fatto, molte degli enti in arancio risultano essere presenti e sparsi nella regione in verde (oltre che nel Nord) e viceversa sfumando e rendendo difficile contornare un'area geografica definita.

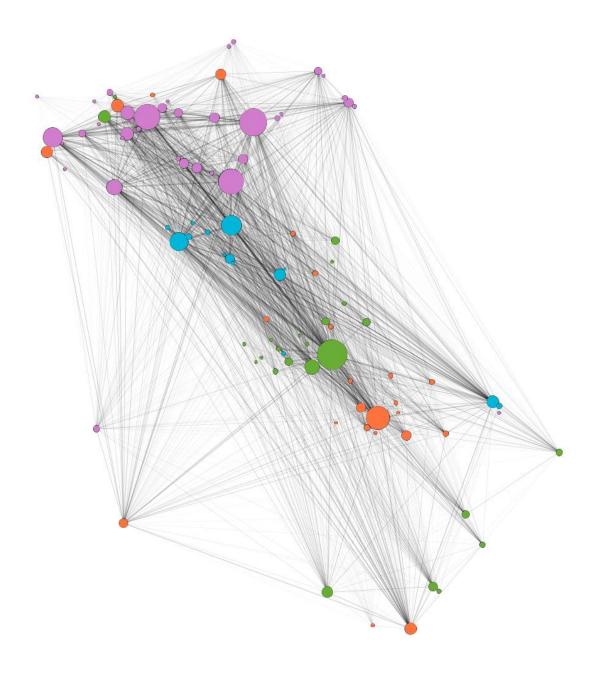


Figura 37

# 6.3 Degree Centrality

Passando in rassegna le metriche di centralità, analizziamo in prima battuta la metrica di *Degree Centrality*. Essa associa un valore (o grado) ad ogni attore della rete in base al numero di nodi entranti/uscenti. Nel nostro caso, quindi, il "grado" associato ad un ente determina il numero di collaborazioni avute nell'anno, per la pubblicazione di un articolo, indipendentemente dal numero di articoli co-pubblicati, permettendoci di conoscere con quanti altri enti ha collaborato nell'anno in questione.

Relativamente all'anno 2019, il livello medio di *degree centrality* è pari a **56,36**. Un'università italiana quindi, in media, collabora con più di 50 enti differenti per la co-pubblicazione di articoli in un anno. Come si può notare dalla distribuzione del valore della metrica riportata

in **Figura 38**, si ha un range abbastanza ampio di valori, che va da un minimo di 2 per l'Università per gli stranieri di Siena a 91 per l'università di Torino. La maggior parte delle università e degli enti, corrispondente al 40.8% del totale riporta un valore di degree centrality che va dai 51 a 75.

Tra i primi dieci posti, in ordine di degree centrality decrescente, troviamo solo università statali. Al primo posto Università di Torino (91) classificandosi quindi come università più collaborativa, a seguire, Università di Bologna (90), Università di Firenze (90), Università di Napoli Federico II (90), Università di Milano (89), Università di Pisa (88), Università di Roma "La Sapienza" (88), Università di Padova (86), Università Tor Vergata di Roma (86). Negli ultimi posti troviamo, come ci si può aspettare, la maggior parte delle università telematiche e private.

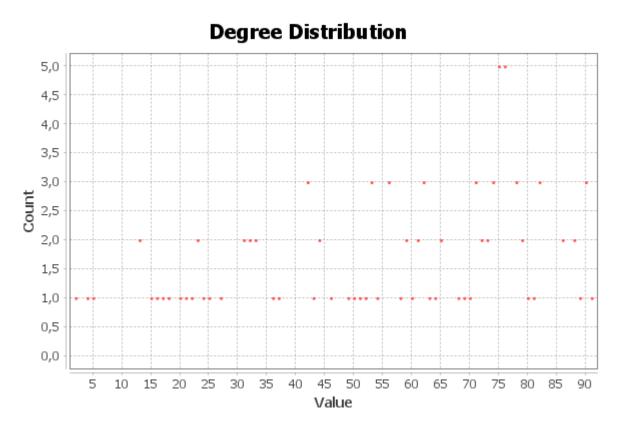


Figura 38 – Distribuzione grado di degree centrality relativa all'anno 2019

Degree Centrality	[0-1]	[2 – 10]	[11 – 25]	[26 – 50]	[51 – 75]	> 75
#	0	3	13	18	40	24
%	0.0%	3.1%	13.3%	18.4%	40.8%	24.5%

Analizzando l'andamento della metrica degree centrality negli anni e calcolandone la crescita percentuale per ogni ente e istituto del dataset si denota come la grande maggioranza di essi abbiano aumentato in cinque anni il numero di enti con cui collaboravano rispetto al 2015, fanno eccezione soltanto Novedrate e-campus (-23%),

Università degli Studi Suor Orsola Benincasa di Napoli (-20%) e Trieste SISSA (-7%) che riportano una diminuzione di collaborazioni e l'Università di Ferrara e Milano IULM che hanno mantenuto lo stesso numero di istituti con cui collaborano.

Dall'analisi del grafico sotto riportato si vede bene come ad avere una crescita più pronunciata in termini di collaborazioni stipulate con enti terzi siano in minor numero. Troviamo L'università per stranieri di Perugia che passa da 3 collaborazioni nel 2015 ad un totale di 16 cinque anni dopo (+433%), Libera Università Mediterranea (+400%) e l'Università di Scienze Gastronomiche di Bra (+333%). Queste grandi percentuali sono tuttavia giustificate anche dal basso numero di co-pubblicazioni che esse già avevano nel 2015, pochi articoli co-pubblicati in più determinano una grande crescita percentuale. Generalmente, circa l'80% degli enti hanno registrato una crescita di collaborazioni che va da 0% a 45% provando che la rete di cooperazione tra le università italiano stia crescendo.

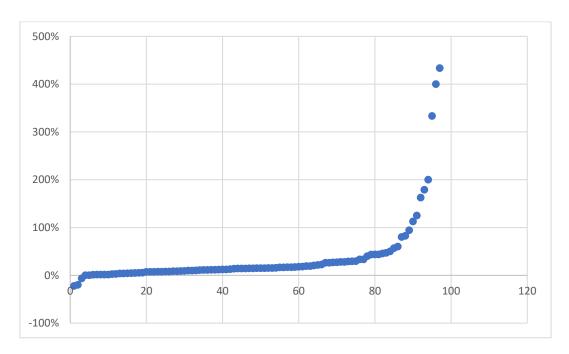


Figura 39 – Plot della crescita percentuale dei degree centrality, sull'asse y, relativi ai 98 istituti del dataset, sull'asse x

### 6.4 Betweenness Centrality

Calcolando la metrica di betweeness centrality siamo in grado di avere una misura di quanto e di quali nodi fungano da "tramite". Essendo anch'essa una metrica di centralità, i nodi che riportano un valore più elevato di betweeness centrality sono quelli che hanno una centralità e importanza più elevata rispetto agli altri della rete stessa in quanto la metrica, tiene conto di quante volte il nodo si interpone nei vari cammini da un nodo ad un altro che il grafo e la sua struttura consente.

#### **Betweenness Centrality Distribution**

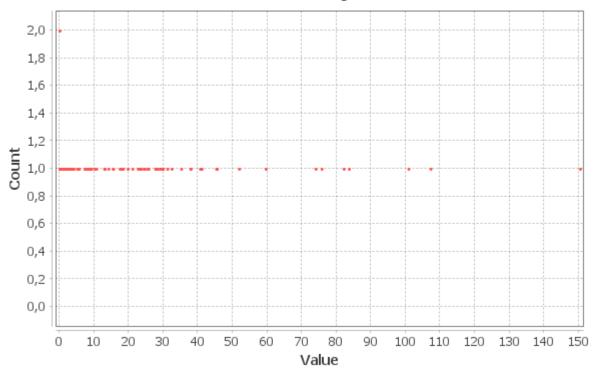


Figura 29 - Distribuzione grado di Betweenness Centrality relativa all'anno 2019

Betweeness Centrality	[0-10]	[11 – 25]	[26 – 50]	[51 – 75]	[76 – 100]	> 100
#	50	20	19	4	3	2
%	51.02%	20.41%	19.39%	4.08%	3.06%	2.04%

Dalla distribuzione della metrica che risulta per l'anno 2019, vediamo come la maggior parte dei nodi, ovvero degli enti, riportino un valore di betweeness centrality basso e man mano che il valore della metrica sale diminuiscono i nodi. Quindi, come ci si può aspettare, pochi nodi ricoprono una "posizione strategica" rispetto alla rete.

Ad esempio, tra le università più collaborative e che hanno stipulato un maggior numero di collegamenti con gli altri enti, abbiamo, tra i primi cinque posti per più alto valore di betweeness centrality: Università di Firenze (150.34), Università di Torino (107.15), Università di Napoli "Federico II" (100.75), Università di Milano (83.57), Università di Pisa (82.06). Viceversa, gli enti che riportano un numero più basso di betweeness centrality sono per lo più istituti privati con una bassa pubblicazione annuale di articoli, come per esempio, l'Università di Aosta, Bra e università telematiche come quella di Napoli Pegaso.

#### 6.5 Closeness Centrality

La misura di closeness centrality consente di individuare gli attori della rete sociale che sono nella posizione migliore per raggiungere gli altri, quindi capaci di "influenzare" la rete più rapidamente. In sostanza è una misura di come velocemente l'informazione può viaggiare attraverso la rete da uno specifico nodo. Analizzandone la distribuzione calcolata con Gephi, si nota come la maggior parte degli enti, circa il 70% del totale, abbiamo registrato, per l'anno 2019, un valore di closeness centrality che va tra i 0.57 e 0.76

A registrare il più alto valore di closeness centrality è l'Università di Torino, grazie al suo elevato livello di collaborazione di cui si è discusso in precedenza, con un valore di 0.941. Seguito dall'Università di Bologna (0.932), Napoli Federico II (0.932) e Firenze (0.932). La classifica sembra quasi rispecchiare quella basata sul livello di degree centrality, riscontrando tra i primi nomi, riorganizzato in ordine decrescente, enti come Roma La Sapienza, Padova, e Pisa.

#### Closeness Centrality Distribution

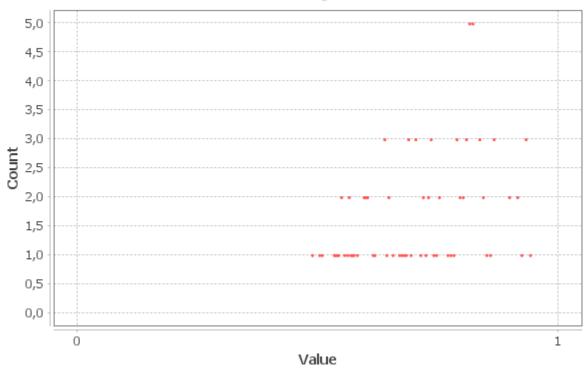


Figura 41 – Distribuzione del grado di closeness centrality relativo all'anno 2019

Closeness Centrality	[0.48 - 0.6)	[0.6 - 0.7)	[0.7 - 0.8)	[0.8 - 0.9)	> 0.9
#	21	19	24	27	7
%	21.4%	19.4%	24.5%	27.6%	7.1%

Passando ad un'analisi dell'evoluzione temporale della metrica rispetto alla rete universitaria, troviamo, tra gli enti che hanno avuto una maggiore crescita del livello di closeness centrality elevando così la propria centralità e importanza rispetto a tutto il network sono: Libera Università Merditerranea (+29%), Rozzano Humanitas University

(+25%), Università di Bolzano (16%), Università di Torino (+12%). In media, tenendo conto di tutti e 98 gli enti, si ha avuto una crescita del 6%. Enti che hanno registrato, invece, una decrescita del proprio valore di closeness centrality sono stati: Milano IULM (-1%), Università di Ferrara (-1%), Trieste SISSA (-2%), Novedrrate e-Campus (-6%).

# 7 Conclusioni

Il presente lavoro ha come obiettivo l'analisi del panorama accademico italiano in termini di pubblicazioni di articoli, approfondendo lo studio con un focus mirato sul grado di collaborazione tra i diversi enti nella co-pubblicazione degli articoli. Il lavoro si è basato su un database composto da 112 file csv che coprono le diverse pubblicazioni (articoli, meeting abstract, proceedings paper, review ecc...) di 98 istituti italiani su un intervallo di tempo variabile. Per garantire un opportuno livello di coerenza dei dati al fine di avere un corretto confronto tra i diversi istituti si è analizzato la pubblicazione di soli articoli tra un intervallo temporale di cinque anni che va dal 2015 al 2019. Si è scelto di omettere il 2020 (anno più recente del dataset), anno di blocco e rallentamento generale dovuto alla pandemia che ha coinvolto anche il mondo didattico e accademico influenzando il numero di pubblicazioni e ricerche in via del tutto eccezionale. Inoltre, per poter approfondire l'analisi del network, la disponibilità di dati in merito al numero di citazione avute dalle pubblicazioni aiuta a determinare una sorta di misura della "qualità" e importanza dell'articolo in base alla risonanza che esso ha avuto nel mondo accademico di riferimento.

I 98 istituti analizzati risultano essere geograficamente sparsi per tutto il territorio italiano generalmente in egual modo con una leggera concentrazione nel centro e nord Italia. Tra i cinque anni che vanno dal 2015 al 2019 sono stati pubblicati un totale di 439.486 articoli. Si denota una crescita del 24% delle pubblicazioni dal 2015 al 2019 che passano da un totale di 78.898 del 2015 ad un totale di 97.979 del 2019, con una media di pubblicazioni per istituto che passa così a 1000 pubblicando in media 200 articoli in più rispetto a cinque anni prima. Tra le prime università per numero di pubblicazioni si trova "La Sapienza" di Roma, l'Università di Padova, l'Università di Bologna, l'Università di Milano e l'Università Federico II di Napoli. Passando invece ad un'analisi del numero di citazioni registrate per ogni ente, tra gli istituti ad aver registrato una media più alta di citazioni per articolo rientrano università telematiche come Napoli Pegaso, Novedrate e-campus e istituti ad ordinamento speciale come L'Aquila GSSI, Trieste SISSA e aziende ospedaliere universitarie come Rozzano Humanitas University, University hospital of Siena.

Passando all'analisi del numero di collaborazioni, anche queste hanno subito una crescita, passando da un totale di 2389 nel 2015 a 2762 nel 2019, registrando un aumento del 15%. Dallo studio e dall'analisi eseguita mediate il software per lo studio dei network sociali *Gephi* non risultano particolari enti di spicco o in particolare rilievo in rapporto alle metriche analizzate nel dettaglio. I valori ad essi associati si presentano quando uniformemente distribuiti, come per il caso della *closeness centrality*, quando gradualmente crescenti o decrescenti, come per il caso di *degree centrality* e *betweenees centrality*. Dall'analisi della metrica di degree centrality si deduce che in media un'istituto collabora con 56 enti diversi (prendendo l'anno 2019 come riferimento) ovvero con più della metà degli istituti presenti sul suolo italiano. La conferma è data anche dal livello di Network Density che si assesta ad un valore pari a 0.581, determinando un network idealmente rappresentabile come un grafo a metà tra sconnesso e totalmente connesso (fully connected). Si da evidenza, tuttavia, di come questa metrica sia cresciuta negli anni presi in analisi dallo studio passando da 0.503 nel 2015 ad appunto 0.581, con una crescita del 15,5% testimoniando una crescita del

numero di collaborazioni tra enti negli anni così come il numero di co-pubblicazioni che subiscono un aumento. Rispetto alle metriche di centralità tra gli enti che riportano i valori più elevati, come ci si aspettava, si trovano principali università italiane come l'Università di Torino, L'Università Federico II di Napoli, l'Università di Pisa, l'Università di Bologna, l'Università di Milano e l'Università La Sapienza di Roma. Università longeve, con un grande numero di pubblicazioni e collaborazioni con gli altri istituti rendendole figure centralizzanti nel network collaborativo. Prosequendo con una Network Visualization è possibile confermare anche visivamente quanto rilevato dalle metriche. La rappresentazione di Fruchterman-Reingold nello specifico, posizionando le università sopra elencate al centro del grafo conferma quanto detto sulla loro posizione nevralgica all'interno del network. Altre università minori, enti di ricerca e università telematiche, invece, occupando una posizione periferica rispetto al centro del grafo evidenziano un minore coinvolgimento nella rete collaborativa in merito alla co-pubblicazione di articoli. Non emerge, invece, alcuna rappresentazione significativa disponendo il grafo secondo l'algoritmo Force Atlas 2, che dovrebbe metter in evidenza eventuali moduli, ovvero comunità di enti particolarmente collaborativi tra loro, della rete. Il grafo che l'algoritmo elabora non dà la chiara e distinta presenza di moduli.

# Bibliografia

- [1] B. d'Italia, «La ricerca scientifica fondamento dello sviluppo economico».
- [2] J. Rybnicek, «Innovation in the United States and Europe».
- [3] «openpolis.it,» 27 Novembre 2019. [Online]. Available: https://www.openpolis.it/ricerca-e-sviluppo-guanto-investono-litalia-e-i-paesi-ue-in-guesto-settore/.
- [4] «Ministero dell'università e della ricerca,» [Online]. Available: https://www.mur.gov.it/it/areetematiche/ricerca/programmazione/pnr-anni-2015-2020.
- [5] H. I. Y. U. S. Y. T. Takashi lino, «How does the global network of research collaboration,» The Japanese Economic Review, 2020.
- [6] «https://www.springernature.com/,» 11 Nov 2019. [Online].
- [7] S. Milgram, «The Small World Problem,» Psychology Today, May 1967.
- [8] M. Granovetter, «The Strength of Weak Ties,» American Journal of Sociology 78, 1973.
- [9] G. Ceccaroni, «orbyta.it,» 22 June 2021. [Online].
- [10] T. Arif, «The Mathematics of Social Network Analysis: Metrics for Academic Social,» *International Journal of Computer Applications Technology and Research*, 2015.
- [11] G. Z. F. S. Raquel Recuero, «Using Social Network Analysis and Social Capital to Identify User Roles on Polarized Political Conversations on Twitter,» *Social Media + Society*, 2019.
- [12] J. A. L. H. Alireza Abbasi, «Identifying the effects of co-authorship networks on the performance of scholars: A correlation and regression analysis of performance measures and social network analysis measures,» *Journal of Informetrics*, 2011.
- [13] G. Z. e. F. S. Raquel Recuero, «Using Social Network Analysis and Social Capital to Identify user Roles on Polarized Political Conversation on Twitter,» *Social Media + Society*, 2019.
- [14] «wikipedia,» [Online].