

healthcare-analytics

COURSE NAME

Big Data Analytics

Author:

Federica Galimberti (191676)

26 agosto 2025

UNIMORE
UNIVERSITÀ DEGLI STUDI DI
MODENA E REGGIO EMILIA



Indice

Presentazione del grafo	2
1 Exploratory Analysis	7
2 Prima reaserch question	8
2.1 CARBOPLATIN	17
3 Seconda Reaserch question	19
4 Conclusione	23

Presentazione del grafo

Il grafo scelto per l'analisi è Healthcare-analytics, disponibile al seguente link:

https://github.com/FedericaGalimberti1234/Big_data_exam/blob/main/Graph_analytics/. Tutto il codice utilizzato per ottenere i risultati è disponibile sia nella stessa cartella in cui è stato caricato il progetto, sia al seguente link: (link da inserire). Questo grafo è composto da 11.381 nodi e 6.145 relazioni. I nodi appartengono a 8 tipologie differenti, mentre le relazioni sono suddivise in 11 tipologie. Nella seguente tabella viene riportata la descrizione delle diverse tipologie di nodi presenti nel grafo.

Entità	Etichetta
Case	Caso clinico in esame
Drug	Farmaco coinvolto nell'evento.
Reaction	Reazione sviluppata dalla persona (Caso).
Outcome	Esito del caso
Report-Source	Soggetto che ha segnalato l'evento
Terapy	Terapia prescritta ad una certa persona (Case).
Manufacturer	Azienda produttrice del farmaco
Age-Group	Fascia di età del caso clinico in esame.

Tipo di Relazione	Descrizione	Nodi coinvolti
REGISTERED	Il caso è stato registrato dall'azienda.	Case Manufacturer
FALLS_UNDER	Il caso in esame ha una determinata fascia di età.	Case AgeGroup
RESULTED_IN	il caso ha portato a un esito (outcome).	Case Outcome
HAS_REACTION	Il caso ha sviluppato una reazione.	Case Reaction
REPORTED_BY	Il caso è stato segnalato da una determinata fonte (repourt source).	Case Report Source
IS_PRIMARY_SUSPECT	Il farmaco è sospettato come causa principale della reazione nel caso in esame (reaction).	Case Drug
IS_SECONDARY_SUSPECT	Il farmaco è sospettato come causa secondaria della reazione nel caso (reaction).	Case Drug
IS_CONCOMITANT	Il farmaco è stato assunto contemporaneamente ad un altro ma non è considerato sospetto diretto.	Drug Case
IS_INTERACTING	Il farmaco potrebbe aver interagito con altri farmaci causando effetti avversi.	Case Drug
RECEIVED	Il caso ha ricevuto una terapia con un determinato farmaco.	Case Terapy
PRESCRIBE	Il farmaco è stato prescritto per una specifica indicazione clinica.	Drug Terapy

Le due tabelle descrivono le tipologie di nodi e di relazioni presenti nel grafo. Il grafo contiene otto tipi di nodi che rappresentano i farmaci, le reazioni avverse, i casi clinici, il range di età dei pazienti, l'azienda produttrice dei farmaci, il soggetto che ha riportato l'evento avverso e le terapie a cui è sottoposto il caso. Sono presenti undici tipi di relazioni che, nella maggior parte dei casi, partono dal nodo relativo al caso clinico e lo collegano agli elementi che descrivono la sua situazione. Ad esempio, la relazione `RECEIVED` collega il caso alle terapie intraprese, mentre `IS_PRIMARY_SUSPECT` e `IS_SECONDARY_SUSPECT` lo collegano ai farmaci ritenuti rispettivamente causa principale o secondaria di una reazione avversa. La relazione `HAS_REACTION` associa il caso alle reazioni osservate. L'unica relazione che non parte dal nodo caso è `PRESCRIBE`, che collega direttamente i farmaci alle terapie che li coinvolgono. Questa strutturazione risulta evidente anche dal comando `db.schema` (riportato nella figura sottostante)

```
CALL db.schema.visualization()
```

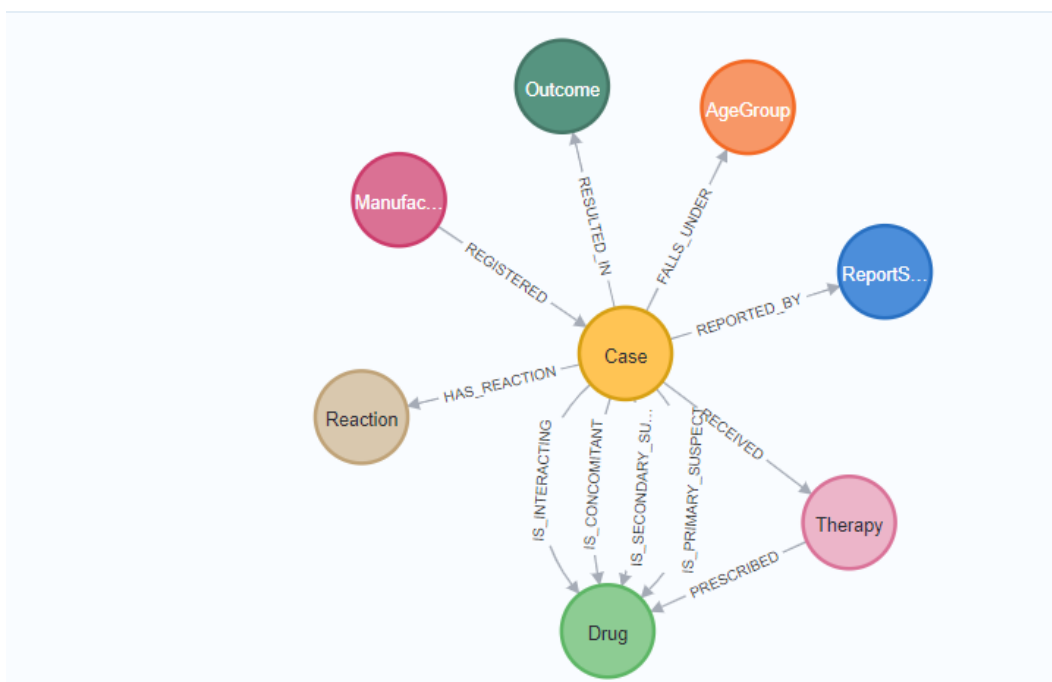


Figura 1: schema visualization

Di seguito sono riportate le proprietà associate ai nodi e alle relazioni presenti nel grafo. La maggior parte delle caratteristiche è riferita al nodo Case, per il quale vengono riportati attributi quali primaryid, age, ageUnit, gender, eventDate, reportDate e reporterOccupation. Per gli altri nodi, invece, le informazioni disponibili si limitano generalmente al nome e/o al codice identificativo. Per quanto riguarda le relazioni, solo quelle che collegano i nodi Case e Drug presentano proprietà aggiuntive, tra cui doseAmount, doseUnit, drugSequence, route e indication.

	relType	collect(propertyName)
0	:`FALLS_UNDER`	[]
1	:`HAS_REACTION`	[]
2	:`IS_CONCOMITANT`	[doseAmount, doseUnit, drugSequence, route, indication]
3	:`IS_INTERACTING`	[doseAmount, doseUnit, drugSequence, route, indication]
4	:`IS_PRIMARY_SUSPECT`	[doseAmount, doseUnit, drugSequence, route, indication]
5	:`IS_SECONDARY_SUSPECT`	[doseAmount, doseUnit, drugSequence, route, indication]
6	:`PRESCRIBED`	[drugSequence, startYear, endYear]
7	:`RECEIVED`	[]
8	:`REGISTERED`	[]
9	:`REPORTED_BY`	[]
10	:`RESULTED_IN`	[]

Figura 2: Proprietà delle relazioni

	nodeType	collect(propertyName)
0	:`AgeGroup`	[ageGroup]
1	:`Case`	[primaryid, age, ageUnit, gender, eventDate, reportDate, reporterOccupation, componentId]
2	:`Drug`	[name, primarySubstance, componentId]
3	:`Manufacturer`	[manufacturerName]
4	:`Outcome`	[code, outcome]
5	:`Reaction`	[description, componentId]
6	:`ReportSource`	[name, code]
7	:`Therapy`	[primaryid]

Figura 3: Proprietà dei nodi

Di seguito sono riportate le distribuzioni dei nodi e delle relazioni.

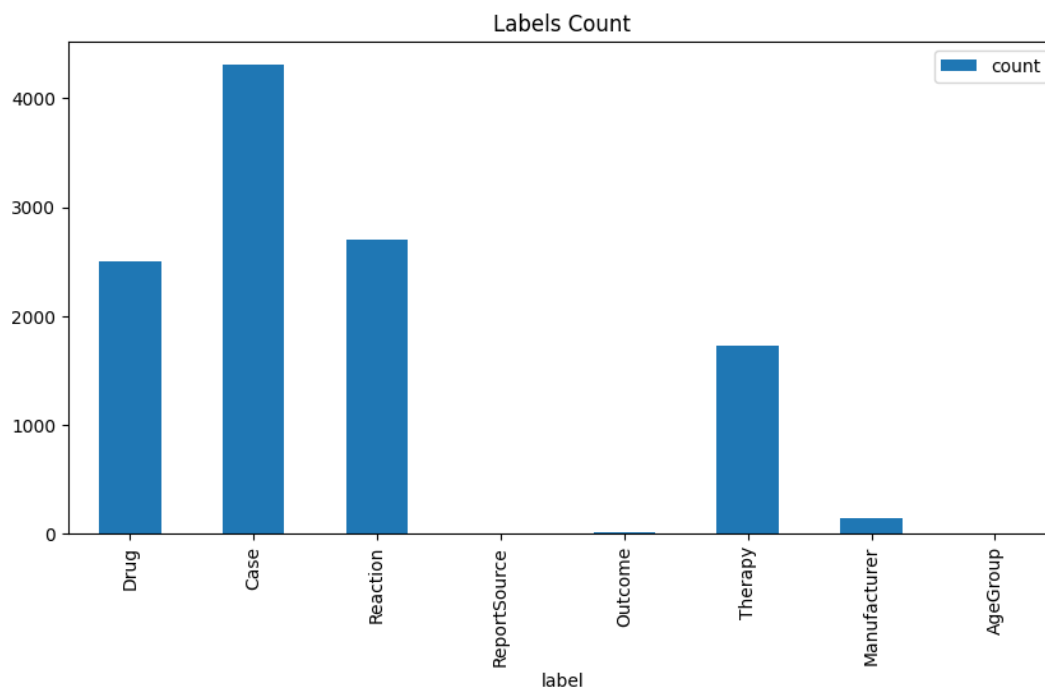


Figura 4: distribuzione dei nodi

I nodi presentano una distribuzione estremamente sbilanciata. Osserviamo infatti che i nodi Case, Reaction, Drug e Therapy risultano numericamente molto più presenti rispetto ad altri tipi di nodi, come ReportSource, Outcome, Manufacturer e AgeGroup, che invece compaiono con frequenza ridotta.

Tuttavia, una minore varietà di nodi non implica necessariamente una minore importanza. Ad esempio, il nodo Outcome, pur avendo soltanto sei possibili valori, risulta fortemente coinvolto nell'analisi: come si osserva nella figura sottostante, il numero di occorrenze dell'arco $RESULTED_1N$ mostra chiaramente come questo nodo sia comunque importante nel grafo

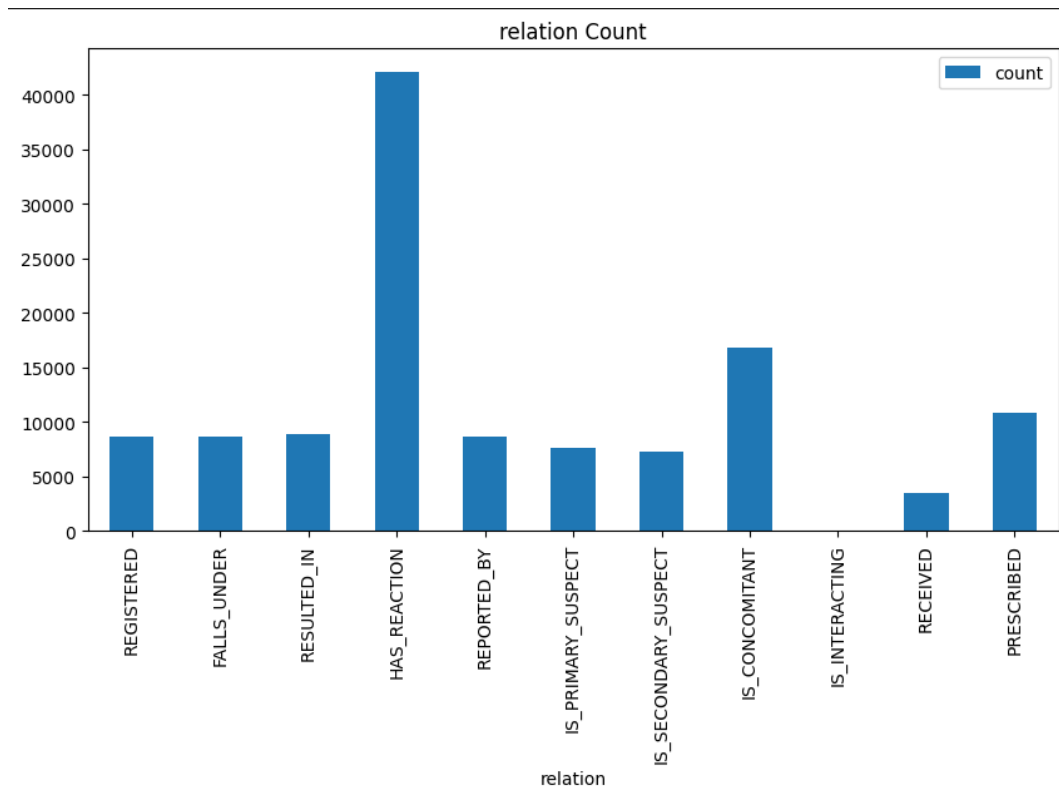


Figura 5: distribuzione delle relazioni

Per quanto riguarda invece le relazioni, si osserva che la relazione *HAS_RREACTION* presenta nettamente più occorrenze rispetto alle altre. Questo dato evidenzia come il grafo sia stato progettato principalmente per l'analisi delle reazioni ai farmaci. Infatti, mettendo insieme i risultati relativi alla distribuzione dei nodi e quelli delle relazioni, emerge chiaramente che l'obiettivo principale del grafo è lo studio delle reazioni e del loro legame con i farmaci.

Nell'analisi proposta l'attenzione sarà rivolta principalmente ai nodi Drug, Therapy, Case, Reaction, Manufacture e Outcome. Per quanto riguarda le relazioni, l'analisi si concentrerà su *HAS_REACTION*, *IS_PRIMARY_SUSPECT*, *RECEIVED*, *REGISTERED* e *RESULTED_IN*

1 Exploratory Analysis

Iniziamo a esplorare la struttura del grafo e, in generale, la distribuzione dei nodi e delle relazioni che costituiranno il focus delle research question sviluppate successivamente. Come primo passo, è stata verificata la connettività del grafo: esso risulta connesso, poiché presenta un'unica componente debolmente connessa (come mostrato in figura).

```
neo4j$ CALL gds.wcc.stats('myGraph') YIELD componentCount;
```

componentCount
1

Figura 6: numero componenti debolmente connesse

Nelle Research Question verrà analizzata in dettaglio la reazione ai farmaci, concentrandosi in particolare sui farmaci principali sospettati di causarla. Di seguito è riportato il grafico a barre che mostra il numero di Case collegati ai Drug tramite $IS_{PRIMARY_USPECT}$

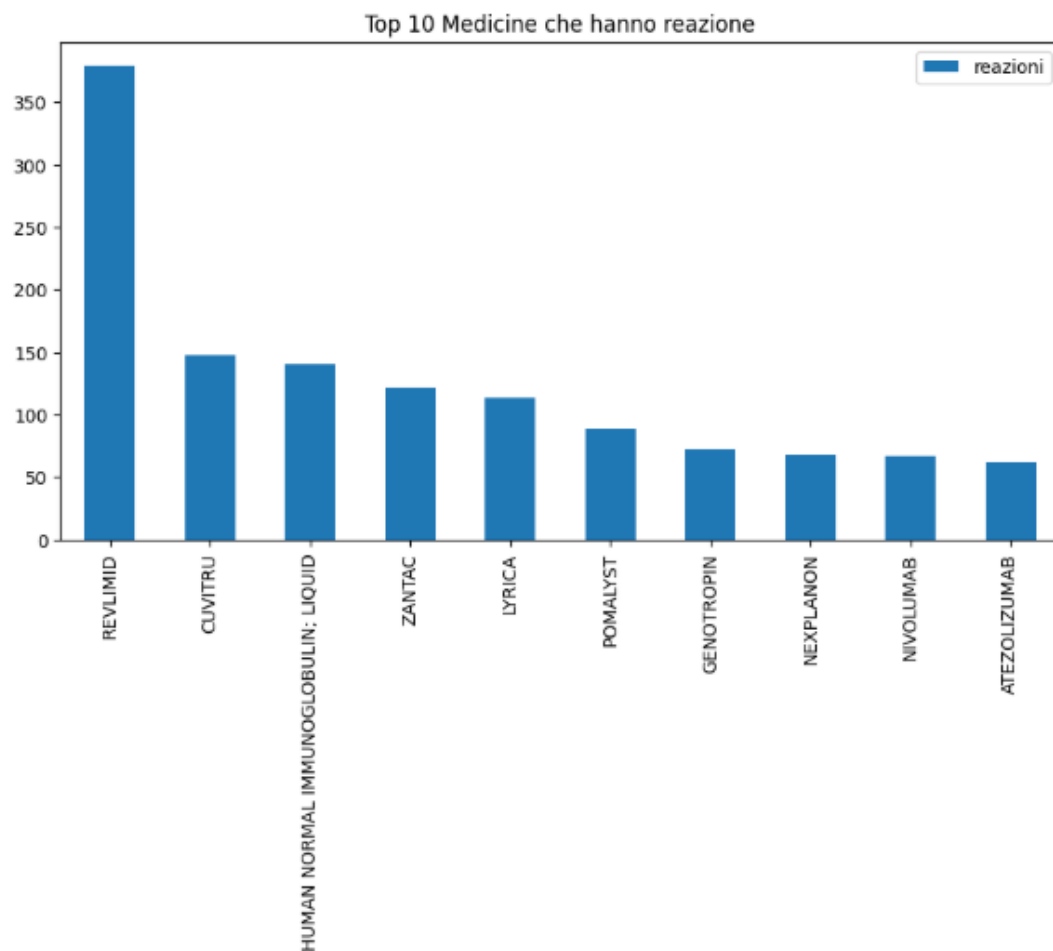


Figura 7: Il numero di casi per farmaco (primi 10)

Essendo fondamentale la relazione $IS_{PRIMARY_USPECT}$, per condurre un'analisi adeguata è stata effettuata una prima valutazione su come si relazionano tra loro la prescrizione dei farmaci e l'insorgenza di reazioni avverse. La prima domanda posta sul grafo riguarda proprio il legame tra le prescrizioni dei farmaci e la loro possibile responsabilità nell'indurre reazioni avverse. Infatti, nel caso in cui esistesse una relazione diretta, analizzare i risultati da un punto di vista assoluto potrebbe risultare fuorviante.

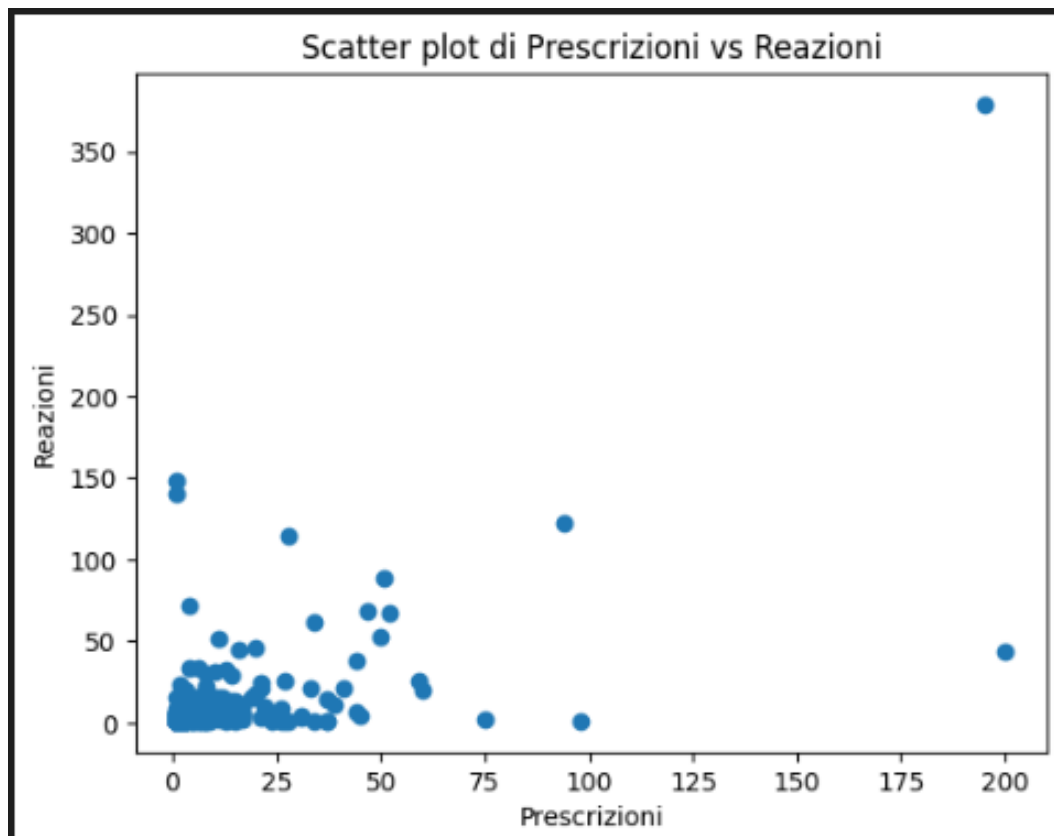


Figura 8: scatter plot prescrizioni contro reazioni

Dal grafico si evince che le reazioni ai farmaci possono essere considerate anche da un punto di vista assoluto, poiché non emerge una relazione lineare tra il numero di farmaci prescritti e il numero di reazioni riscontrate.

2 Prima reaserch question

La prima research question è articolata in due fasi. Nella prima fase l'obiettivo è indagare la relazione tra farmaco, reazione e outcome, con particolare attenzione all'esito Death. Nella seconda fase, invece, l'analisi si concentra sulla relazione tra farmaco, reazione e azienda produttrice.

Come primo passo, è stata effettuata un'osservazione preliminare sugli Outcome presenti nel dataset, individuando sei differenti tipologie: Death, Hospitalization – Initial or Prolonged, Other Serious (Important Medical Event), Life-Threatening, Disability e Congenital Anomaly. Poiché l'esito Death rappresenta il caso più critico e clinicamente rilevante, l'analisi si focalizzerà principalmente su di esso, per comprendere meglio quali farmaci e quali reazioni siano maggiormente associate a questo tipo di outcome.

	name
0	Death
1	Hospitalization - Initial or Prolonged
2	Other Serious (Important Medical Event)
3	Life-Threatening
4	Disability
5	Congenital Anomaly

Figura 9: Differenti tipi di outcome

In particolare, l'analisi si è concentrata sui farmaci che risultano essere i principali sospettati di aver portato a questo esito e sui tipi di reazioni coinvolte in tali casi.

Il primo passo è stato quello di costruire una proiezione Cypher al fine di generare il grafo su cui condurre le analisi successive.

```
query_create = '''
CALL gds.graph.project.cypher(
  'DrugCaseReactionGraphDeath',
  '
    MATCH (d:Drug)-[:IS_PRIMARY_SUSPECT]-(c:Case)-[:HAS_REACTION]-(r:Reaction),
      (c)-[:RESULTED_IN]-(o:Outcome)
    WHERE o.outcome = "Death"
    RETURN DISTINCT id(d) AS id
  UNION
    MATCH (d:Drug)-[:IS_PRIMARY_SUSPECT]-(c:Case)-[:HAS_REACTION]-(r:Reaction),
      (c)-[:RESULTED_IN]-(o:Outcome)
    WHERE o.outcome = "Death"
    RETURN DISTINCT id(c) AS id
  UNION
    MATCH (d:Drug)-[:IS_PRIMARY_SUSPECT]-(c:Case)-[:HAS_REACTION]-(r:Reaction),
      (c)-[:RESULTED_IN]-(o:Outcome)
    WHERE o.outcome = "Death"
    RETURN DISTINCT id(r) AS id
  ',
  // Archi
  '
    MATCH (d:Drug)-[:IS_PRIMARY_SUSPECT]-(c:Case)-[:HAS_REACTION]-(r:Reaction),
      (c)-[:RESULTED_IN]-(o:Outcome)
    WHERE o.outcome = "Death"
    RETURN id(d) AS source, id(c) AS target, "IS_PRIMARY_SUSPECT" AS type
  UNION
    MATCH (d:Drug)-[:IS_PRIMARY_SUSPECT]-(c:Case)-[:HAS_REACTION]-(r:Reaction),
      (c)-[:RESULTED_IN]-(o:Outcome)
    WHERE o.outcome = "Death"
    RETURN id(c) AS source, id(r) AS target, "HAS_REACTION" AS type
  '
)
'''
```

Figura 10: Proiezione cypher

Il grafo risultante dalla proiezione rappresenta la rete dei farmaci sospettati, dei casi e delle reazioni esclusivamente per gli eventi che hanno portato a un esito fatale.

```

query_components = """
CALL gds.graph.list()
YIELD graphName, nodeCount, relationshipCount
WHERE graphName = 'DrugCaseReactionGraphDeath'
RETURN graphName, nodeCount, relationshipCount
"""

with driver.session(database="neo4j") as session:
    result = session.run(query_components)
    data = [record.data() for record in result]

import pandas as pd
data_df = pd.DataFrame(data)
print(data_df)

```

✓ 0.2s

C:\Users\Utente\AppData\Local\Temp\ipykernel_7292\330043038.py:8:

```

with driver.session(database="neo4j") as session:
    graphName  nodeCount  relationshipCount
0 DrugCaseReactionGraphDeath      870      1362

```

Figura 11: Elementi del grafo

Il grafo risultante risulta avere 870 nodi e 1362 relazioni rimanenti.

Come prima analisi è stata eseguita una Weakly Connected Components (WCC), che permette di individuare le componenti debolmente connesse e di verificare il grado di connettività tra i nodi, in modo tale da comprendere la struttura del grafo appena costruito.

```

query = ''' CALL gds.wcc.stream('DrugCaseReactionGraphDeath')
YIELD nodeId, componentId
WHERE componentId IN [18,22, 60, 84,113,114]
RETURN componentId,
    COLLECT(coalesce(gds.util.asNode(nodeId).name, gds.util.asNode(nodeId).description)) AS nodes
ORDER BY componentId '''

with driver.session(database="neo4j") as session:
    result = session.run(query)
    data = [record.data() for record in result]
    pd.set_option('display.max_colwidth', None)
    data_df = pd.DataFrame(data)
    data_df.head(10)

```

	componentId	nodes
0	18	[CARBOPLATIN, Hepatitis B reactivation]
1	22	[MYCOPHENOLATE MOFETIL, Death neonatal, Maternal exposure during pregnancy]
2	60	[ALUNBRIG, Central nervous system infection]
3	84	[MYCAPSSA, Embolic stroke]
4	113	[AUTOGENE CEVUMERAN, Systemic immune activation]
5	114	[COSELA, Swelling face, Epistaxis]

Figura 12: elementi per componente

Il grafo risultante presenta 7 componenti debolmente connesse: sei di esse sono costituite da meno di sei nodi ciascuna, mentre una singola componente ne contiene oltre 800. Questa distribuzione ha reso necessario condurre due tipi di analisi distinte: una dedicata alle componenti di dimensioni ridotte e un'altra focalizzata sulla componente di grandi dimensioni.

Osserviamo che, all'interno delle componenti più piccole, è presente un solo farmaco e, al massimo, due reazioni diverse. Questo indica che, per quei farmaci, le uniche reazioni che portano all'esito fatale ("Death") sono proprio quelle collegate all'interno di tali componenti. Un'analisi più approfondita di uno di questi casi, relativo al Carboplatin, è stata condotta nella sottosezione dedicata.

Per quanto riguarda invece la componente principale, quella con oltre 800 nodi, l'attenzione si concentra sullo studio del grado sia dei nodi farmaco sia dei nodi reazione, al fine di individuare i farmaci più connessi e le reazioni più comuni associate all'esito fatale. Per effettuare questa analisi è stato usato l'algoritmo della Degree sia sui nodi di tipo Drug sia sui nodi di tipo Reaction.

	drug	score
0	REVLIMID	24.0
1	ATEZOLIZUMAB	18.0
2	IMBRUVICA	13.0
3	ELIQUIS	11.0
4	NIVOLUMAB	9.0
5	HUMAN NORMAL IMMUNOGLOBULIN; LIQUID	7.0
6	LONSURF	6.0
7	Niraparib	6.0
8	POMALYST	6.0
9	LYRICA	4.0
10	AMG 510	4.0
11	CUVITRU	4.0
12	OPDIVO	4.0
13	NUPLAZID	4.0
14	KYMRIAH	4.0
15	OBINUTUZUMAB	3.0
16	XL184	3.0
17	CABOMETYX	3.0
18	NINLARO	3.0
19	TYSABRI	3.0

Figura 13: Degree dei farmaci nei casi che hanno come outcome Death

L'analisi del degree dei farmaci evidenzia come REVLIM presenti un valore particolarmente elevato (24), suggerendo che sia coinvolto in un numero molto alto di connessioni nel grafo relative ai casi con outcome Death. Si tratta infatti di un farmaco utilizzato per il trattamento di specifiche patologie del sangue, il che potrebbe indicare che venga prescritto in numerose terapie e spesso in contesti clinici caratterizzati da condizioni gravi.

Anche ATEZOLIZUMAB mostra un degree rilevante (18) ed essendo un farmaco oncologico è plausibile che sia frequentemente associato a esiti fatali. Allo stesso modo, altri farmaci come IMBRUVICA, ELIQUIS e INVOLUMAB appartengono anch'essi all'ambito oncologico, rafforzando l'idea che i farmaci impiegati nel trattamento di patologie tumorali siano maggiormente rappresentati nei casi con outcome Death.

	reaction	score
0	Death	95.0
1	Pneumonia	22.0
2	Dyspnoea	21.0
3	Sepsis	19.0
4	Acute kidney injury	14.0
5	Fall	13.0
6	Cardiac arrest	13.0
7	Pyrexia	13.0
8	Fatigue	12.0
9	Cerebrovascular accident	11.0
10	Respiratory failure	10.0
11	Septic shock	10.0
12	Cough	10.0
13	Malignant neoplasm progression	9.0
14	Hypotension	9.0
15	Acute respiratory failure	9.0
16	Atrial fibrillation	9.0
17	Hypoxia	9.0
18	Off label use	8.0
19	Nausea	8.0

Figura 14: Degree delle reazioni farmaci nei casi che hanno come outcome Death

Osserviamo ora il degree dei nodi di tipo Reaction.

Le reazioni innescate dai farmaci analizzati risultano essere variegata, come emerge chiaramente dalla distribuzione dei gradi associati ai nodi di reazione. Un aspetto particolarmente interessante è che il nodo con il punteggio più alto corrisponde proprio a Death: questo suggerisce che, nella maggior parte dei casi in cui l'outcome finale è stato la morte, essa viene registrata anche come reazione diretta al farmaco. Quindi questo evidenzia come nella maggior parte dei casi death è una reazione diretta al farmaco.

Tra le altre reazioni più frequenti nei casi con esito fatale troviamo Pneumonia, Dyspnea e Sepsis, che sono purtroppo eventi comuni nei pazienti in condizioni cliniche critiche. Tuttavia, emergono anche alcune reazioni inattese tra quelle più rilevanti, come Fall e Cough. La loro presenza potrebbe essere spiegata dal fatto che si tratta di sintomi generici, che vengono riportati spesso e in contesti differenti, assumendo così un peso sproporzionato nell'analisi.

In conclusione, i farmaci coinvolti in esiti fatali sono spesso impiegati per trattamenti di patologie con situazioni cliniche particolarmente compromesse. Inoltre, l'analisi ha evidenziato che, oltre all'outcome più ovvio rappresentato dalla morte, emergono anche altre reazioni più generiche, indicando una varietà di effetti avversi associati a questi farmaci

Un'altra domanda che ci poniamo riguarda l'analisi, per ciascun produttore, dei farmaci che generano il maggior numero di effetti indesiderati. Come primo passo, è stato realizzato un grafico che mostra il numero totale di reazioni avverse segnalate dalle diverse aziende produttrici. Questo permette di identificare rapidamente quali produttori sono maggiormente coinvolti nei casi di eventi avversi e fornisce una base per successivi approfondimenti sui singoli farmaci responsabili.

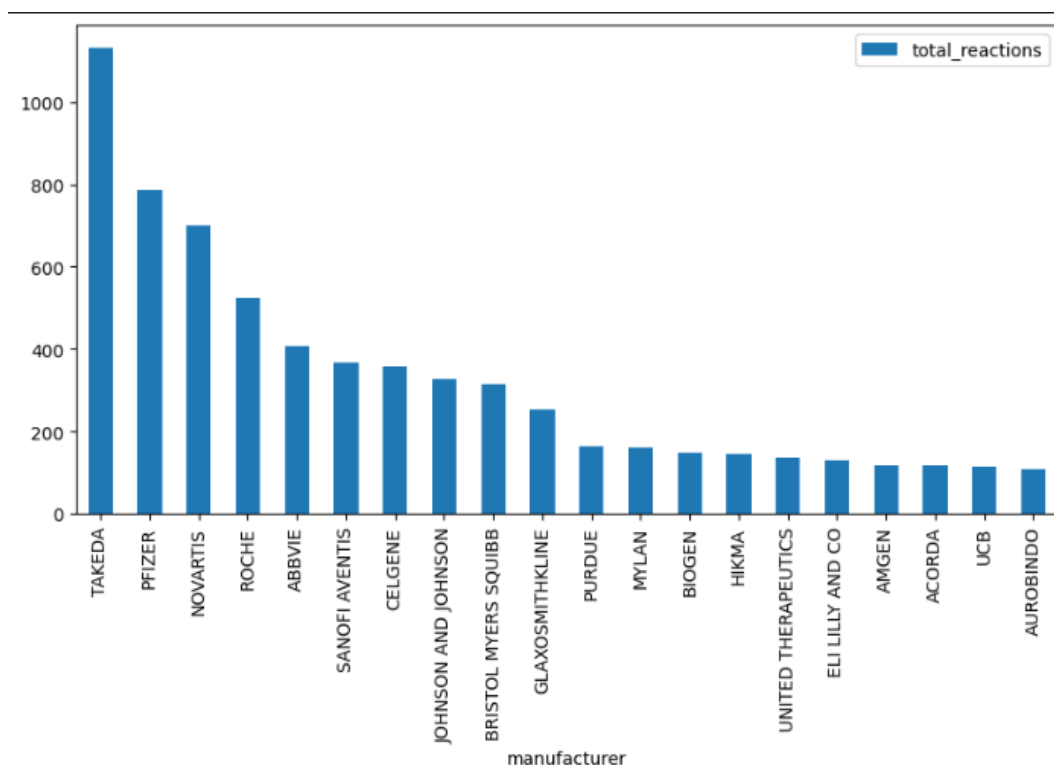


Figura 15: Totale reazioni ai farmaci per azienda

Dall'analisi emerge che Takeda e Pfizer sono le due aziende che, in termini assoluti, presentano il maggior numero di segnalazioni di reazioni avverse. L'obiettivo, quindi, è quello di concentrarsi su queste due aziende, analizzando in dettaglio quali siano i farmaci maggiormente coinvolti e come le reazioni avverse si distribuiscono per ciascun farmaco.

Va inoltre sottolineato che, per entrambi i grafici, la leggibilità è stata migliorata mostrando in legenda solo i primi 10 farmaci per volume di reazioni. Infatti, Takeda conta oltre 20 farmaci e Pfizer addirittura 34: rappresentarli tutti avrebbe compromesso la chiarezza visiva dell'analisi.

reazioni avverse ai farmaci prodotti da TAKEDA

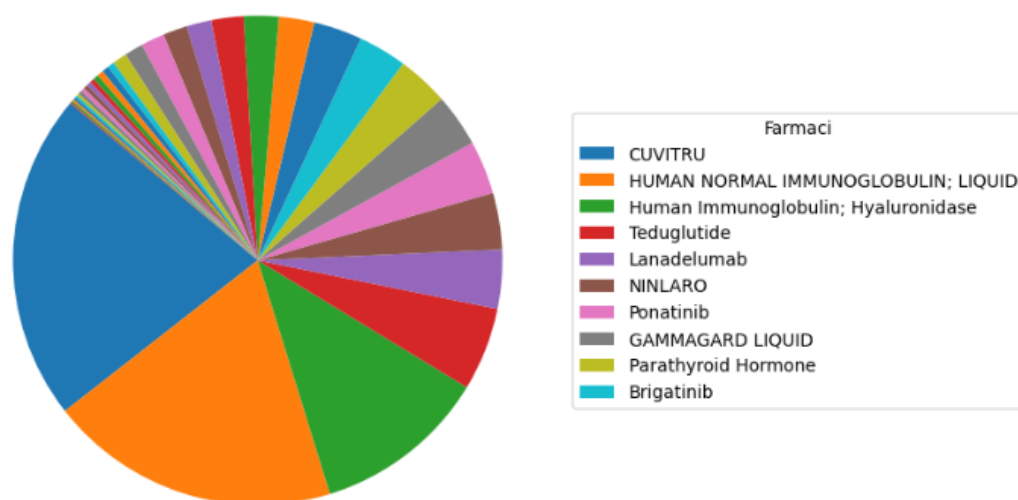


Figura 16: Totale reazioni ai farmaci per TAKEDA

Notiamo dal grafico come il 50% delle reazioni ai farmaci è stata riscontrata nell'assunzione di soli 3 farmaci CUVITRU, HUMAN NORMAL IMMUNOGLOBULIN LIQUID e HUMAN IMMUNOGLOBULIN: HYALURONIDASE. Questi tre farmaci sono parte di terapie immunologiche sostitutive basate sulla somministrazione di anticorpi (immunoglobuline), utilizzate generalmente per supportare pazienti con deficit nella risposta immunitaria. Considerando la tipologia di trattamenti che eseguono il fatto che essi siano ritenuti responsabili di molte reazioni è un risultato attendibile.

Verifichiamo se questi farmaci, per la loro tipologia, vengono prescritti frequentemente in numerose terapie oppure se il loro utilizzo è più limitato.

Numero di terapie per farmaco prodotto da TAKEDA

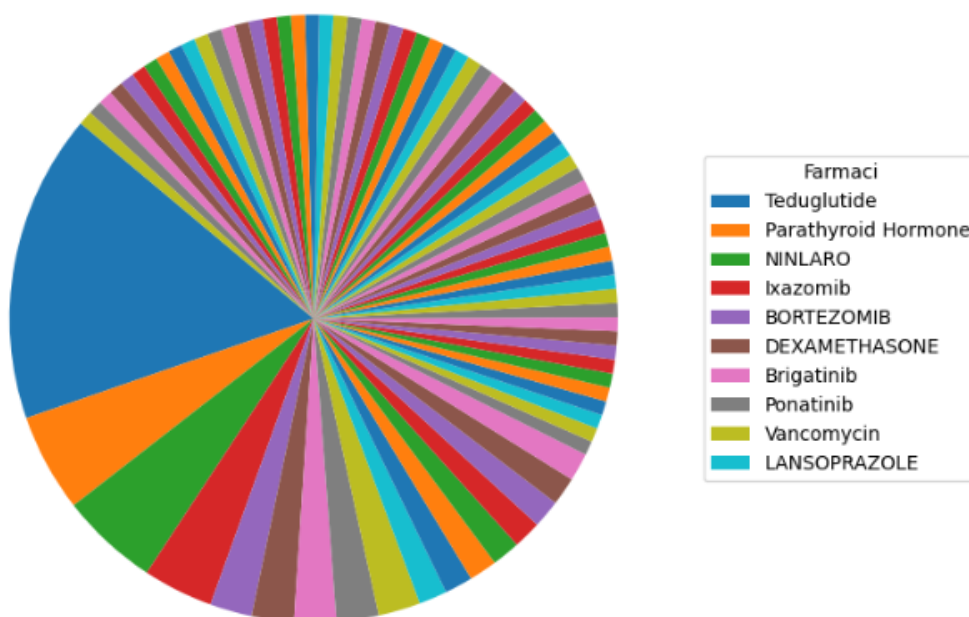


Figura 17: Numero di terapie che prescrivono il tipo di farmaco prodotto da TAKEDA

I farmaci che presentano il numero più elevato di reazioni, come si evince dal grafico, non risultano coinvolti in un ampio numero di terapie e, di conseguenza, non sono associati a un alto volume di casi trattati. Al contrario, i farmaci più frequentemente prescritti sono Teglutide e Parathyroid Hormone, che appartengono a categorie terapeutiche più generiche e quindi trovano applicazione in un numero maggiore di trattamenti.

Osserviamo come le stesse osservazioni non posso essere ripetute per Pfizer perchè infatti riscontriamo una corrispondenza tra il numero di terapie che prescrivono il farmaco.

Numero Reazioni prodotti da PFIZER

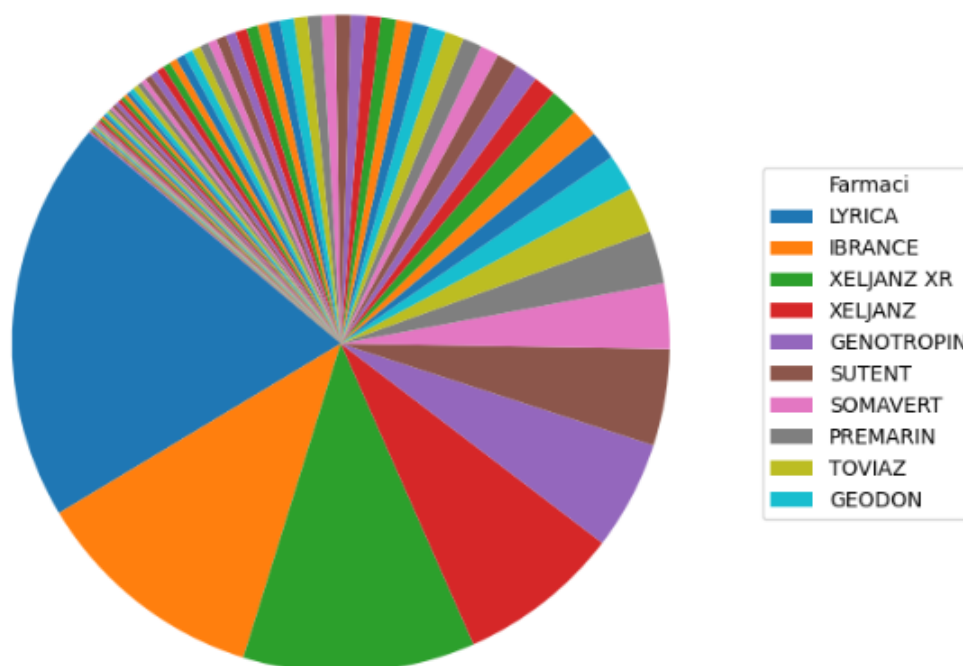


Figura 18: Totale reazioni ai farmaci per Pfizer

Nel grafico sopra si osserva come, per Pfizer, circa il 50% delle reazioni avverse sia provocato da quattro farmaci principali: Lyrica, Ibrance, Xeljanz e Xeljanz XR. Questi farmaci hanno funzioni piuttosto diverse tra loro: Lyrica è utilizzato per il trattamento del dolore neuropatico, della fibromialgia e delle convulsioni parziali; Ibrance è un farmaco antitumorale; mentre Xeljanz/Xeljanz XR è impiegato per patologie autoimmuni.

Osserviamo quindi che, come tipologia, si tratta di farmaci relativamente eterogenei, ma nel complesso possono essere considerati abbastanza generici, in quanto coinvolti in diverse tipologie di terapie

Numero di terapie per farmaco prodotto da PFIZER

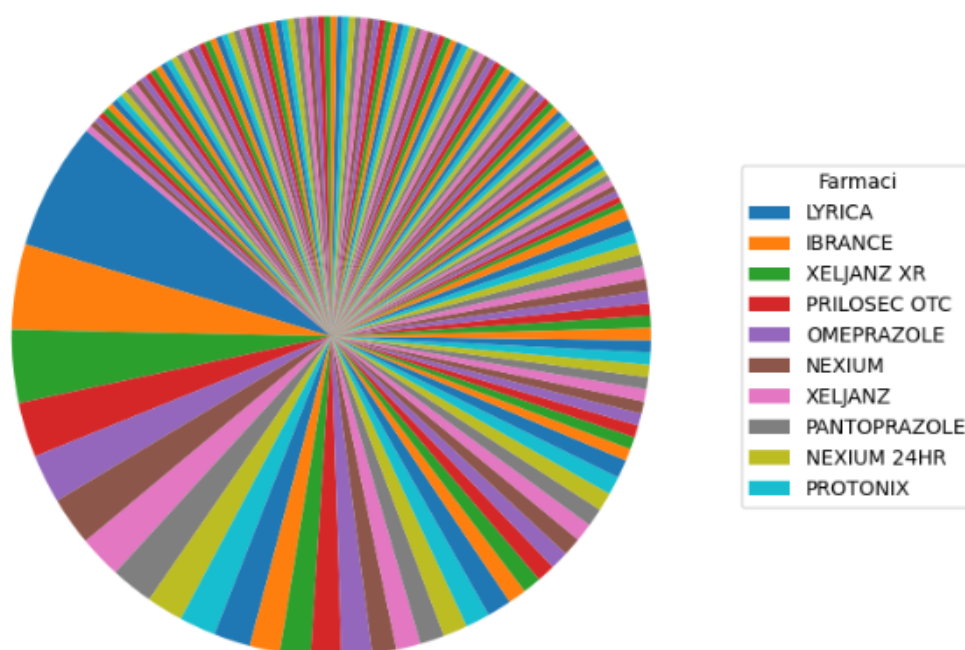


Figura 19: Numero di terapie che prescrivono il tipo di farmaco prodotto da Pfizer

Per quanto riguarda invece il numero di prescrizioni dei farmaci, osserviamo che quelli di Pfizer risultano coinvolti in un maggior numero di terapie rispetto al caso di Takeda, anche se nel grafico a torta non si riscontra la stessa marcata differenza tra le fette osservata in precedenza.

Come conclusione di questa analisi, possiamo affermare che le reazioni ai farmaci a livello aziendale sono spesso concentrate in un piccolo gruppo di farmaci. Non emerge un legame diretto tra il numero di prescrizioni e le reazioni avverse; piuttosto, l'incidenza delle reazioni sembra essere più strettamente correlata al tipo di patologie per le quali i farmaci sono impiegati.

2.1 CARBOPLATIN

CARBOPLATIN emerge dall'analisi delle componenti debolmente connesse dei nodi associati all'outcome Death. Per contestualizzare, Carboplatin è un farmaco utilizzato nel trattamento di diversi tipi di tumori, in particolare quelli che interessano gli apparati riproduttivi femminili.

Questo dato spiega perché la reazione predominante associata a Carboplatin sia proprio la Hepatitis B reactivation: infatti, tra le principali cause di alcuni tumori dell'apparato genitale femminile (ad esempio, al collo dell'utero) vi sono diversi ceppi del virus dell'Epatite B.

L'analisi precedente aveva evidenziato una componente isolata composta unicamente da Carboplatin e dalla reazione Hepatitis B reactivation. Per approfondire questo aspetto, ho proposto una digressione mirata volta a capire sia quali altri farmaci possono provocare la stessa reazione, sia quali tipi di outcome risultano maggiormente associati all'uso di Carboplatin

```
query = '''
MATCH (d:Drug)-[:IS_PRIMARY_SUSPECT]-(c:Case)-[:HAS_REACTION]-(r:Reaction),
      (c)-[:RESULTED_IN]-(o:Outcome)
WHERE d.name = 'CARBOPLATIN'
WITH o.outcome AS outcome, collect(DISTINCT r.description) AS reactions
RETURN outcome, reactions
'''

with driver.session(database="neo4j") as session:
    result = session.run(query)
    data = [record.data() for record in result]
    data = pd.DataFrame(data)
    data.head(10)
```

0.1s

C:\Users\Utente\AppData\Local\Temp\ipykernel_7292\1557471628.py:10: DeprecationWarning: Using a driver after it has been with driver.session(database="neo4j") as session:

	outcome	reactions
0	Hospitalization - Initial or Prolonged	[Asthenia, Malaise, Pyrexia, Hepatitis B reactivation, Atrial flutter, Cytokine release syndrome, Sepsis]
1	Life-Threatening	[Hepatitis B reactivation, Sepsis]
2	Death	[Hepatitis B reactivation]
3	Other Serious (Important Medical Event)	[Hepatitis B reactivation]

Figura 20: Carboplatin diverse reazioni in base agli outcome

Si osserva che, nei casi in cui l'outcome risulta essere più grave — come Death, Life-Threatening e Other Serious (Important Medical Event) — è coinvolta quasi esclusivamente la reazione Hepatitis B reactivation. Al contrario, quando l'outcome riguarda l'ospedalizzazione, le reazioni associate risultano molto più variegata.

```
query = '''
MATCH (d:Drug)-[:IS_PRIMARY_SUSPECT]-(c:Case)-[:HAS_REACTION]-(r:Reaction),
      (c)-[:RESULTED_IN]-(o:Outcome)
WHERE r.description = 'Hepatitis B reactivation'
WITH d.name AS farmaco, collect(o.outcome) AS outcomes
RETURN farmaco, outcomes
'''

with driver.session(database="neo4j") as session:
    result = session.run(query)
    data = [record.data() for record in result]
    data = pd.DataFrame(data)
    data.head(10)
```

0.1s

C:\Users\Utente\AppData\Local\Temp\ipykernel_7292\1213319689.py:10: DeprecationWarning: Using a driver with driver.session(database="neo4j") as session:

	farmaco	outcomes
0	CARBOPLATIN	[Life-Threatening, Death, Other Serious (Important Medical Event), Hospitalization - Initial or Prolonged]
1	VENETOCLAX	[Other Serious (Important Medical Event)]

Figura 21: I diversi farmaci e i rispettivi outcome che portano all' Hepatitis B reactivation

Per quanto riguarda l'analisi dei farmaci che portano come reazione Hepatitis B reactivation, il risultato evidenzia soltanto due farmaci coinvolti: Carboplatin e Venetoclax (entrambi impiegati in terapie oncologiche). È interessante osservare che, mentre per Carboplatin la reazione è associata prevalentemente a esiti fatali, Venetoclax mostra come unico outcome l'ospedalizzazione. Inoltre, Venetoclax risulta meno specifico rispetto a Carboplatin per i trattamenti dei tumori che interessano l'apparato riproduttore femminile.

3 Seconda Reaserch question

La seconda Reaserch question si pone invece l'obiettivo di andare ad analizzare quelli che sono farmaci con reazioni simili.

Il primo passaggio è quella di creare un grafo che contiene solo i nodi Drug e la unica relazione tra loro è rappresentata dal fatto di aver provocato su un caso una reazione comune.

```
query = '''CALL gds.graph.project.cypher(
  'DrugSimilarityGraph',
  'MATCH (d:Drug)-[:IS_PRIMARY_SUSPECT]-(:Case)-[:HAS_REACTION]-(:Reaction)
  RETURN id(d) AS id, labels(d) AS labels',
  'MATCH (d1:Drug)-[:IS_PRIMARY_SUSPECT]-(:Case)-[:HAS_REACTION]-(:Reaction)-[:HAS_REACTION]-(:Case)-[:IS_PRIMARY_SUSPECT]-(:d2:Drug)
  WHERE id(d1) < id(d2)
  RETURN id(d1) AS source, id(d2) AS target'
)
YIELD graphName, nodeCount, relationshipCount'''
```

Figura 22: Proiezione del grafo

Il grafo risultante implementando l'algoritmo della wcc ha 8 componenti connesse.

componentid	num_component_in_component	nodes
0	0	593 [AFINITOR, CLONAZEPAM, DEXAMETHASONE, KEPRA, LISINAPRIL, Belimumab, PREDNISONE, REVLMID, LYRICA, FABRAZYME, METOPROLOL, CARVEDILOL, Velagluferase alfa, ERIVEDGE, BORTEZOMIB, VELCADE, XYREM, PLAVIX, ATORVASTATIN CALCIUM, DEXAMETHASONE ORAL, METFORMIN, THALOMID, AMLODIPINE, MYRBETRIQ, CLOPIDOGREL, DEXTROAMPHET SACCHARAT, OMEPRAZOLE, MIRCERA, GILENYA, PREMARIN, brentuximab vedotin, Vincristine, NEURONTIN, AMPYRA, TECFIDERA, KYPROLIS, Teduglutide, ENTYVIO, GABAPENTIN, LIPITOR, PROTONIX, HUMIRA, IBRANCE, POMALYST, MORPHINE SULFATE, Metreleptin, VISMODEGIB, MINOXIDIL, TYLENOL, DILAUDID, RAPAMUNE, TERIFLUNOMIDE, CELEBREX, MYCOPHENOLATE MOFETIL, NEXIUM, METHOTREXATE, CYTARABINE, GAMUNEX C, ELIQUIS, XELJANZ, EXTAVIA, VINCRISTINE SULFATE, MERCAPTOPYRINE, TOVIAZ, RELPAX, BOSULIF, Ixazomib, XELJANZ XR, ESTRADIOL, APXABAN, PANTOPRAZOLE, DABRAFENIB, GSK1120212, MEDROL, POLYETHYLENE GLYCOL 3350, NUPLAZID, carfilzomib, OTEZLA, EFFEXOR, Niraparib, CELLCEPT, Ponatinib, REMODULIN, ADCIRCA, FENTANYL CITRATE, AGALSIDASE BETA, DEXAMETHASONE IV, VELTASSA, TOPAMAX, OXYCODONE AND ACETAMINOPHEN, NAGLAZYME, LINAGLIPTIN, LOSARTAN, TOCILIZUMAB, ZYKADIA, KYMRIAH, NINLARO, SUTENT, PREMPRO, OCREVUS, ...]
1	141	1 [TRASTUZUMAB]
2	196	1 [AUTOGENE CEVUMERAN]
3	216	1 [Fedratinib]
4	361	1 [talimogene laherparepvec]
5	375	1 [MESNA]
6	392	1 [LIDOCAINE HYDROCHLORIDE]
7	514	1 [Semaglutide B 30 mg/ml PDS290]

Figura 23: Componenti connesse grafo

Notiamo come ci sono 7 nodi isolati e una componente con più di 500 elementi. I nodi isolati segnalano farmaci che non hanno reazioni in comune con altri farmaci. Ora osserviamo quali di questi hanno reazioni simili utilizzando l'algoritmo similarity.

	Drug1	Drug2	similarity
0	ERIBULIN	KUVAN	1.000000
4	KUVAN	ERIBULIN	1.000000
8	ONUREG	ERIBULIN	1.000000
12	Scopolamine Transdermal System	ERIBULIN	1.000000
16	ZYPREXA RELPREVV	ERIBULIN	1.000000
20	BELVIQ	SILDENAFIL CITRATE	0.965517
21	SILDENAFIL CITRATE	BELVIQ	0.965517
22	XELJANZ	XELJANZ XR	0.913551
23	XELJANZ XR	XELJANZ	0.913551
24	LYRICA	REVLIMID	0.897533
25	REVLIMID	LYRICA	0.897533
26	ALTEPLASE	UPADACITINIB	0.888889
27	UPADACITINIB	ALTEPLASE	0.888889
28	NUPLAZID	XELJANZ XR	0.885442
30	BEOVU	TIGAN	0.875000
31	TIGAN	BEOVU	0.875000
32	ATEZOLIZUMAB	NIVOLUMAB	0.865229
33	NIVOLUMAB	ATEZOLIZUMAB	0.865229
35	TOBI	SILDENAFIL CITRATE	0.862069
38	NINLARO	OCREVUS	0.847458

Figura 24: I 20 farmaci che hanno gli score più alti di similarità

Avere alti score di similarità indica che due farmaci condividono un numero elevato di vicini comuni (altri farmaci collegati tramite reazioni simili). Questo suggerisce che, pur non avendo necessariamente le stesse reazioni, i farmaci tendono a essere coinvolti in contesti clinici simili e, di conseguenza, possono presentare tipi di reazione analoghi.

Nel nostro caso, notiamo un primo gruppo composto da ERIBULIN, KUVAN, ONUREG, Scopolamine Transdermal System e ZYPREXA RELPREVV, che hanno tutti score di similarità pari a 1. Questo significa che ogni farmaco ha esattamente lo stesso insieme di vicini nel grafo proiettato.

Per comprendere meglio la natura di questa relazione, andiamo ora ad analizzare le Reaction associate a questo gruppo di farmaci con similarità massima

	Drug	Reactions
0	ERIBULIN	[Pneumonia]
1	KUVAN	[Weight increased, Weight increased]
2	ONUREG	[Decreased appetite, Fatigue]
3	Scopolamine Transdermal System	[Dizziness, Nausea, Vomiting]
4	ZYPREXA RELPREVV	[Depressed level of consciousness, Sedation, Post-injection delirium sedation syndrome, Disorientation, Hypertension, Confusional state, Hypotension, Delirium, Agitation]

Figura 25: Le reaction del gruppo con similarità 1

Osserviamo che, come detto in precedenza, i farmaci appartenenti a questo gruppo presentano reazioni diverse. Tuttavia, è possibile fare due osservazioni importanti. La prima è che ciascun farmaco all'interno del gruppo ha un numero limitato di reazioni associate. La seconda riguarda la natura delle reazioni stesse, che, nella maggior parte dei casi, risultano essere abbastanza generiche.

Proprio per questa combinazione, poche reazioni e tipologie generiche, i farmaci finiscono per condividere gli stessi vicini nel grafo, cioè gli stessi nodi Reaction collegati ai casi. Questo spiega perché il loro punteggio di similarità risulti pari a 1, nonostante le descrizioni delle reazioni possano apparire diverse.

Un'osservazione simile può essere fatta per BELVIQ e SILDENAFIL CITRATE che hanno una similarità di 0.96. Notiamo dalla figura sotto come essi condividano una reazione al farmaco ovvero Drug ineffective

	Drug	Reactions
0	BELVIQ	[Drug ineffective, Invasive lobular breast carcinoma, Invasive lobular breast carcinoma]
1	SILDENAFIL CITRATE	[Suspected counterfeit product, Drug ineffective]

Figura 26: Le reaction del gruppo con similarità 0.96

[illegible]

Ricordiamo che il numero di nodi nel grafo sono 601 quindi circa il 90 % degli elemnti appartengono ssa componente ovvero hanno delle reazione simili o comunque henno delle reazioni in comune.

Le reazioni con 1 elemento sono invece indicano che per tutti quei farmaci hanno reazioni particolari comuni e specifici per il tipo di farmaco.

communityd Drug

70	543	KUVAN
79	543	ERIBULIN
83	543	ONUREG
87	543	ZYPREXA RELPREVV
88	543	Scopolamine Transdermal System
238	543	BELVIQ
250	543	SILDENAFIL CITRATE
516	554	XELJANZ
523	554	XELJANZ XR

Nella tabella sottostante è riportata la similarità di tutti gli elementi presenti, ordinata in modo crescente rispetto al valore. Si osserva che, per gli elementi in esame, il valore minimo di similarità è pari a 0

	Drug1	Drug2	similarity
3521	SILDENAFIL CITRATE	HETLIOZ	0.307692
3419	BELVIQ	HETLIOZ	0.318681
2823	BELVIQ	KYLEENA	0.378378
2728	SILDENAFIL CITRATE	KYLEENA	0.388889
1974	SILDENAFIL CITRATE	BUPROPION HYDROCHLORIDE XL	0.482759
...
10	ONUREG	Scopolamine Transdermal System	1.000000
9	ONUREG	KUVAN	1.000000
8	ONUREG	ERIBULIN	1.000000
7	KUVAN	ZYPREXA RELPREVV	1.000000
6	KUVAN	Scopolamine Transdermal System	1.000000

Figura 29: Score degli Elementi con score di similitudine superiore a 0.90

Ora analizziamo quelli di 6 elementi e andiamo ad osservare le reazioni.

	Drug	Reactions
0	BUPRENORPHINE ACTAVIS	[Drug withdrawal syndrome neonatal, Fetal hypokinesia, Fetal biophysical profile score abnormal, Tonicus, Apnoea, Developmental delay, Fetal exposure during pregnancy, Poor feeding infant, Crying]
1	Betmumab	[Etioblastoma multiforme, Kaposiella virus, Pajocophaly, Fetal exposure during pregnancy, Congenital torticollis, Haemangioma of skin]
2	BRUGA	[Uppse increased]
3	EMTRICITABINE/TENOFOVIR/EMTRICITABINE/ALAFENAMIDE	[Abnormal aortic arch, Fetal exposure during pregnancy, Patent ductus arteriosus]
4	EMTRICITABINE/TENOFOVIR/ALAFENAMIDE	[Melanocytic nevus, Rash macular, Fontanelle depressed, Fetal hypoplasia, Fetal exposure during pregnancy]
5	EMTRICITABINE	[Excessive glutathione transferase increased, Uppse increased, Gastroenteritis]

Figura 30: Le reazioni ai farmaci presenti nella componente di 6 elementi riscontrata nella Louvain

Dall'analisi dei risultati si osserva che una parte significativa delle reazioni riportate riguarda la fase neonatale. Questo suggerisce che l'elevata similitudine tra i farmaci identificata tramite l'algoritmo Louvain ha portato alla formazione di un cluster specifico, caratterizzato proprio dalla condivisione di reazioni avverse di questo tipo

Osserviamo i tipi di reazioni che hanno i cluster che sono risultanti composti da un solo elemento con l'algoritmo Louvain.

	Drug	Reactions
0	ALDURAZYME	[Cartilage operation, Productive cough]
1	LIDOCAINE HYDROCHLORIDE	[Foreign body in gastrointestinal tract, Removal of foreign body from oesophagus]
2	LUPKYNIS	[Angioedema]
3	ORILISSA	[Abnormal uterine bleeding]
4	VUMERITY	[Seizure like phenomena, Multiple sclerosis]

Figura 31: Le reazioni di 6 differenti cluster composti da un solo elemento

Osserviamo che tutti i farmaci riportati nella tabella sopra presentano poche reazioni, e che queste reazioni non rientrano tra quelle generiche, risultando quindi più specifiche e distintive per ciascun farmaco.

In conclusione, i farmaci che mostrano alta similarità e appartengono allo stesso cluster individuato dall'algoritmo Louvain tendono a condividere caratteristiche simili: possono avere numerose reazioni o comunque reazioni generiche. Al contrario, i farmaci che formano cluster singoli presentano generalmente poche reazioni, che risultano più specifiche e distintive per ciascun farmaco.

4 Conclusione

La prima research question ha evidenziato che i farmaci coinvolti in esiti fatali sono spesso impiegati per trattamenti di patologie con situazioni cliniche particolarmente compromesse. L'analisi ha mostrato che, oltre all'outcome più ovvio rappresentato dalla morte, emergono anche altre reazioni più generiche, indicando una varietà di effetti avversi associati a questi farmaci. Inoltre, le reazioni a livello aziendale risultano concentrate in un piccolo gruppo di farmaci, senza un legame diretto tra il numero di prescrizioni e la frequenza di reazioni; l'incidenza delle reazioni sembra invece correlata al tipo di patologie trattate.

Dal punto di vista pratico, analizzare quali farmaci portano alla morte del paziente può essere utile. I risultati mostrano che molti dei farmaci coinvolti sono farmaci oncologici o comunque utilizzati in contesti clinici già compromessi, e che esistono reazioni avverse di impatto elevato che non sono immediatamente evidenti.

La seconda research question ha invece analizzato le similarità tra le reazioni dei farmaci. Circa il 90% dei farmaci presenta reazioni abbastanza simili tra loro, mentre il restante 10% mostra reazioni specifiche e distintive, portando l'algoritmo Louvain a generare cluster costituiti da singoli farmaci, evidenziando la loro unicità nel profilo di reazioni avverse.

Analizzare le reazioni a livello aziendale è importante perché un numero eccessivo di reazioni per un singolo farmaco può essere dannoso per l'azienda. In entrambi i casi, i farmaci che generano maggiori reazioni appartengono a un piccolo sottogruppo e sono spesso impiegati in patologie con situazioni cliniche già compromesse.

Infine, la valutazione delle similarità tra le reazioni offre spunti per analisi incrociate, anche se queste similarità non sono sempre evidenti, poiché i tipi di reazioni coincidono poco tra loro