UAB
Universitat Autònoma de Barcelona

Dipòsit digital de documents de la UAB

This is the **published version** of the bachelor thesis:

Termes Bòria, Pol. *Plataforma analítica per a l'estudi de la resiliència de cadenes de subministrament basada en Knowledge Graphs.* Treball de Final de Grau (Universitat Autònoma de Barcelona), 2026 (Grau en Enginyeria de Dades)

This version is available at https://ddd.uab.cat/record/337336

under the terms of the CC BY-SA license.

Plataforma analítica per a l'estudi de la resiliència de cadenes de subministrament basada en *Knowledge Graphs*

Pol Termes Bòria

Resum—*Les cadenes de subministrament globals són sistemes complexos i fràgils, on una sola incidència pot propagar-se per tota la xarxa. Aquest treball proposa el disseny i la implementació d'una plataforma analítica per a l'estudi de la seva resiliència, basada en una arquitectura de dades de Knowledge Graph (KG). A partir d'un dataset logístic redissenyat amb una lògica de generació coherent amb la realitat (13 ciutats, 25 parells origen–destinació amb rutes alternatives i un model causal de risc), es construeix un graf de coneixement sobre Neo4j que representa explícitament les dependències estructurals de la xarxa. Sobre aquest graf s'aplica una anàlisi de resiliència organitzada en sis blocs — des dels baselines operatius fins a la simulació d'escenaris de fallada what-if — que identifica punts de fragilitat estructural i quantifica el cost de les disrupcions. Tot el sistema s'integra en una aplicació interactiva desenvolupada amb Streamlit, que incorpora a més un agent d'intel·ligència artificial que permet realitzar consultes en llenguatge natural. Els resultats demostren que la representació en graf habilita anàlisis (centralitat, detecció de comunitats, simulació de fallades) inviables eficientment en un model relacional, validant la metodologia com a banc de proves reutilitzable per a l'estudi de la resiliència logística.*

Paraules clau—*Cadena de subministrament, resiliència, Knowledge Graph, Neo4j, teoria de grafs, anàlisi de risc, agent IA, simulació what-if.*

Abstract—*Global supply chains are complex and fragile systems in which a single incident can propagate across the entire network. This work proposes the design and implementation of an analytical platform for studying their resilience, based on a Knowledge Graph (KG) data architecture. Starting from a logistics dataset redesigned with a generation logic consistent with reality (13 cities, 25 origin–destination pairs with alternative routes and a causal risk model), a knowledge graph is built on Neo4j that explicitly represents the structural dependencies of the network. A resilience analysis organised in six blocks — from operational baselines to what-if failure-scenario simulation — is applied on top of this graph, identifying points of structural fragility and quantifying the cost of disruptions. The whole system is integrated into an interactive application developed with Streamlit, which additionally incorporates an artificial-intelligence agent for natural-language querying. The results show that the graph representation enables analyses (centrality, community detection, failure simulation) that are not efficiently feasible in a relational model, validating the methodology as a reusable testbed for the study of logistics resilience.*

Index Terms—*Supply chain, resilience, Knowledge Graph, Neo4j, graph theory, risk analysis, AI agent, what-if simulation.*

I. Introducció

LES cadenes de subministrament són sistemes complexos i inherentment fràgils davant de qualsevol errada, retard o mala gestió. Aquesta fragilitat implica que qualsevol incidència pot propagar-se al llarg de la xarxa i generar un impacte significatiu si no es gestiona adequadament. Per aquest motiu, és fonamental disposar de mecanismes d'auditoria i monitoratge que permetin detectar problemes amb antelació i actuar de manera eficient, minimitzant-ne la propagació i les conseqüències.

En les darreres dècades, les cadenes de subministrament globals han incrementat notablement la seva complexitat, donant lloc a xarxes extenses i interconnectades de proveïdors, rutes i clients distribuïts a escala global. Gestionar aquestes xarxes requereix sistemes d'informació capaços de representar no només els fluxos de dades transaccionals, sinó també les relacions i dependències estructurals entre els seus components, fet que incrementa la vulnerabilitat del sistema davant d'errors o disrupcions.

En aquest escenari, els models de dades tabulars o relacionals tradicionals presenten limitacions importants, ja que no són capaços de representar de manera natural l'estructura interconnectada d'una xarxa logística ni d'analitzar eficientment la propagació de pertorbacions. Com a alternativa, els *Knowledge Graphs* (KG) permeten modelar entitats, relacions i propietats mitjançant una estructura de xarxa semàntica, fet que els fa especialment adequats per a dominis amb dependències complexes com les cadenes de subministrament [4], [3].

El present TFG proposa el disseny i la implementació d'una aplicació interactiva desenvolupada amb *Streamlit*, orientada a analistes i usuaris de negoci. Aquesta eina permet la visualització de la xarxa mitjançant visualitzacions interactives i KPIs per donar suport a la presa de decisions. Per fer-ho possible,

• E-mail de contacte: Pol.Termes@autonoma.cat
• Treball tutoritzat per: Josep Lladós
• Curs 2025/26
• Codi i contingut del projecte disponibles a: https://github.com/poltermes21/Supply-Chain-KG

es dissenya i implementa una arquitectura de dades basada en *Knowledge Graphs* per representar les relacions i dependències de la xarxa i analitzar-ne la resiliència. Addicionalment, l'aplicació incorpora un agent d'intel·ligència artificial que permet realitzar consultes en llenguatge natural sobre la informació de la xarxa.

A partir d'aquest plantejament, el treball es concreta en quatre objectius principals:

- **O1.** Dissenyar i implementar una arquitectura de dades basada en *Knowledge Graphs* (Neo4j) que representi explícitament les dependències estructurals d'una cadena de subministrament global.
- **O2.** Desenvolupar una capa analítica de resiliència, estructurada en blocs, que aprofiti la representació en graf per identificar punts de fragilitat estructural i quantificar l'impacte de les disrupcions.
- **O3.** Integrar un agent d'intel·ligència artificial que permeti consultar el graf en llenguatge natural de manera segura, sense exposar l'esquema ni habilitar operacions d'escriptura.
- **O4.** Empaquetar tot el sistema en una aplicació interactiva, parametritzable i reutilitzable, orientada a analistes i usuaris de negoci.

El grau d'assoliment d'aquests objectius es valora explícitament a les conclusions (secció XI).

II. Planificació i metodologia

El projecte es va plantejar com un desenvolupament **incremental per fases**, on cada fase produeix un artefacte funcional que serveix de base per a la següent, fet que permet validar cada component abans de construir-hi a sobre i detectar incidències aviat. El seguiment es va fer amb un diagrama de Gantt revisat periòdicament.

Les fases van ser: (1) exploració i preparació del *dataset*; (2) disseny del model semàntic del graf; (3) *pipeline* d'ingestió i construcció del graf sobre Neo4j; (4) capa analítica de resiliència en sis blocs; (5) integració en l'aplicació *Streamlit*; (6) incorporació de l'agent IA; i (7) redacció de la memòria. Aquesta organització respon a una **arquitectura modular** que separa generació de dades, preparació, construcció del graf, anàlisi i presentació, de manera que cada component es pot substituir o ampliar de forma independent.

A. Desviacions respecte a la planificació inicial

El desenvolupament va comportar dues desviacions, totes dues absorbides dins el marge previst. La primera va ser la **substitució del *dataset* original per un de sintètic** (detallada al capítol IV), que va suposar aproximadament una setmana addicional no contemplada i va allargar les tasques d'anàlisi uns cinc dies.

La segona afecta l'**agent IA**, reprogramat cap al tram final com a funcionalitat integrada dins l'aplicació. Durant el disseny es va valorar una arquitectura basada en *embeddings* i reutilització de consultes prèvies, però es va **descartar** per una raó de domini: com que el sistema està pensat per ingerir dades noves periòdicament, reutilitzar respostes en funció de la similitud de la pregunta podria retornar resultats desactualitzats. Per garantir que cada resposta reflecteix les dades vigents es va optar per una arquitectura **ReAct** que genera i executa la consulta en cada interacció (secció IX).

III. Revisió de sistemes similars

Es van revisar sistemes orientats a la gestió del risc i la resiliència de cadenes de subministrament, tant plataformes comercials com treballs acadèmics, per situar el treball respecte a l'estat de l'art i identificar referències d'inspiració.

En l'àmbit comercial destaquen **Everstream Analytics** [1] i **Resilinc** [2], dues plataformes de referència que construeixen una representació digital (*digital twin*) de la xarxa de proveïdors a partir de dades propietàries, amb visibilitat *multi-tier* i alertes de risc en temps real. Resilinc, a més, ha incorporat recentment un agent d'IA conversacional que permet interrogar el sistema en llenguatge natural i obtenir recomanacions —una funcionalitat que valida de manera independent el concepte del mòdul *KG Chat* desenvolupat en aquest treball. Ambdues són sistemes **tancats i propietaris**: modelen la xarxa com una estructura de dependències però no exposen ni l'esquema ni el motor de base de dades subjacent, i s'orienten sobretot al monitoratge i l'alerta més que a l'anàlisi estructural oberta.

Des de la perspectiva acadèmica, Liu [4] proposa una perspectiva de *Knowledge Graph* per a la resiliència i Kumar [3] explora l'ús de bases de dades de grafs i estàndards RDF; ambdós aporten el fonament metodològic però se centren en el model o l'anàlisi, no en una eina interactiva accessible a usuaris no tècnics.

Aquest treball se situa com una alternativa **oberta i reproduïble**, que integra en un únic sistema el model en graf, l'anàlisi de resiliència i l'agent en llenguatge natural.

IV. Redisseny del *dataset*

L'anàlisi exploratòria del *dataset* original va revelar limitacions estructurals que el feien analíticament inviable per a l'estudi de la resiliència: topologia tancada sense rutes alternatives, correlació nul·la entre risc i resultats, i un model de costos pràcticament uniforme entre corredors molt diferents. Aquestes deficiències requerien reconstruir el *dataset* amb una lògica de generació coherent amb la realitat logística. Aquest capítol presenta les decisions de disseny del generador sintètic desenvolupat amb aquest objectiu; la parametrització completa es presenta a l'Annex A.

A. Redisseny de la xarxa logística

El *dataset* original presentava sis parells origen–destinació biunívocs tancats amb una única ruta per parell. Aquesta estructura, equivalent a sis línies paral·leles independents sense topologia real, inutilitzava els algorismes de teoria de grafs: la *betweenness centrality* de tots els nodes era pràcticament zero i no es podien detectar comunitats significatives.

El *dataset* sintètic modela **13 ciutats amb tres rols diferenciats** (origen pur, destinació pura i *hubs* mixtos que actuen com a punts de transbordament), introduint així *hubs* amb *betweenness centrality* elevada que es converteixen en punts de vulnerabilitat estructural detectables. Sobre vint-i-cinc parells origen–destinació, s'assignen **rutes alternatives**

amb probabilitats diferenciades, fet que habilita la simulació *what-if* de bloqueig de corredors.

Finalment, el *dataset* original contenia "Commodity" com a etiqueta de ruta, que en realitat era una categoria de càrrega i no un corredor geogràfic. Es substitueix per **CoGH (Cape of Good Hope)**, el corredor real que creua l'Atlàntic Sud i l'Oceà Índic, permetent modelar les seves vulnerabilitats específiques (tempestes al Cap, congestió als ports sud-africans) de manera independent a les de Suez.

B. Models de generació de comandes

La generació de cada comanda combina cinc decisions de modelatge basades en geografia econòmica i operativa real, en lloc d'assignacions uniformes.

La **distribució de productes per origen** reflecteix l'especialització productiva de cada regió (electrònica a la Xina, components d'automòbil a Tòquio i Busan, tèxtils i farmacèutics a Mumbai), fet crític perquè les consultes sobre productes exposats a una ruta retornin resultats rellevants. La **selecció del mode** depèn de la densitat de valor del producte (els d'alt valor unitari tendeixen a l'aeri; els pesats i de baix valor, al marítim). Els **temps de lliurament** es basen en distàncies reals de navegació, amb un *buffer* planificat proporcional al risc previst. El **model de costos** és multiplicatiu, combinant base per ruta, factor per mode i producte, escalat logarítmic pel pes i una component de soroll de mercat.

C. Model de risc, disrupcions i mitigació

El *dataset* original presentava una correlació pràcticament nul·la entre els índexs de risc i els retards ($-0{,}012$), cosa que els feia informativament inútils. El *dataset* sintètic assigna **índexs de risc per ruta amb distribucions diferenciades**: Suez té un risc geopolític elevat, el Pacífic un risc meteorològic alt, i CoGH una combinació de risc geopolític baix amb risc meteorològic alt. El **risc combinat** es calcula com una mitjana ponderada (60% geopolític, 40% meteorològic), atorgant més pes al risc geopolític perquè les disrupcions d'aquest tipus tendeixen a tenir un impacte més sever i durador.

El **model causal** opera en dues etapes: risc $\rightarrow$ probabilitat de disrupció $\rightarrow$ retard. El risc no impacta directament el retard, sinó que modula la probabilitat que es produeixi una disrupció, la qual al seu torn genera l'impacte temporal. Cada corredor és vulnerable a tipus específics de disrupció en proporcions diferents — Suez concentra conflictes geopolítics i congestió portuària, el Pacífic tifons, CoGH les tempestes del Cap (un tipus de disrupció afegit exclusivament per a aquesta ruta).

La **matriu de mitigació** assigna accions de resposta amb probabilitats condicionals al tipus de disrupció. Davant una congestió portuària, l'enviament estàndard domina perquè la resposta habitual és esperar; davant un conflicte geopolític, el transport aeri urgent és majoritari perquè és l'única alternativa ràpida; davant un tifó, el desviament de ruta domina perquè els tifons afecten zones delimitades. L'efectivitat de la mitigació es modela com una reducció del retard.

V. Exploració del *dataset*

Abans de la modelització i la construcció del *Knowledge Graph*, es presenta una exploració del conjunt de dades per comprendre'n l'estructura, la qualitat i les principals característiques estadístiques.

A. Estructura general del conjunt de dades

El conjunt de dades redissenyat conté informació detallada sobre ordres de transport dins d'una cadena de subministrament global, amb un total de **15.000 registres i 19 variables**. Les variables s'organitzen en cinc grups funcionals que capturen les diferents dimensions del problema logístic: les característiques de l'ordre i del producte, el context geogràfic i logístic (origen, destinació, ruta i mode), el rendiment del lliurament (temps planificats i reals, estat final), els factors externs de risc (disrupcions, índexs geopolítics, meteorològics i macroeconòmics) i l'impacte econòmic i la mitigació (cost de l'enviament i accions correctives aplicades).

Del total d'atributs, **9 són variables quantitatives** (temps, costos, índexs i pesos) i **10 són variables categòriques o identificadores** (ordres, ciutats, rutes, estats i accions). Aquesta combinació permet una anàlisi multidimensional del sistema logístic, integrant tant informació numèrica com semàntica, fet especialment rellevant per a la posterior representació mitjançant un *Knowledge Graph*. La Taula I presenta una visió sintètica de les variables.

Taula I
Variables del *dataset* agrupades per categoria funcional.

Variable	Tipus	Descripció breu
Característiques de l'ordre i del producte		
Order_ID	String	Identificador únic de l'ordre.
Order_Date	Date	Data d'emissió de l'ordre.
Product_Category	String	Categoria del producte transportat.
Order_Weight_Kg	Integer	Pes total de la comanda (kg).
Context geogràfic i logístic		
Origin_City	String	Ciutat i país d'origen.
Destination_City	String	Ciutat i país de destinació.
Route_Type	String	Corredor logístic principal.
Transportation_Mode	String	Mode de transport (Sea, Air).
Rendiment del lliurament		
Base_Lead_Time_Days	Integer	Temps base estimat.
Scheduled_Lead_Time_Days	Integer	Temps planificat amb marges.
Actual_Lead_Time_Days	Integer	Temps real de lliurament.
Delay_Days	Integer	Dies de retard.
Delivery_Status	String	Estat final (On Time, Late).
Factors externs de risc		
Disruption_Event	String	Tipus de disrupció, si escau.
Geopolitical_Risk_Index	Float	Índex de risc geopolític (0–1).
Weather_Severity_Index	Float	Índex de severitat meteo. (0–10).
Inflation_Rate_Pct	Float	Taxa d'inflació (%).
Impacte econòmic i mitigació		
Shipping_Cost_USD	Float	Cost total de l'enviament (USD).
Mitigation_Action_Taken	String	Acció de mitigació aplicada.

B. Tipus de dades i qualitat del conjunt

L'anàlisi dels tipus de dades mostra que les variables estan correctament tipificades segons la seva naturalesa: les quantitatives representen magnituds contínues o discretes, mentre que les categòriques representen estats logístics, classificacions de risc, ubicacions i identificadors únics. Pel que fa a la qualitat del conjunt, no s'han detectat files duplicades i només s'observen valors perduts a la variable Disruption_Event, amb 12.722 registres sense valor. Aquest fet és coherent amb el domini del problema, ja que la majoria de les ordres no experimenten cap esdeveniment de disrupció: els valors nuls en aquesta variable poden interpretar-se semànticament com a **absència de disrupció**, i no com a un problema de qualitat de les dades.

C. Caracterització estadística

Una anàlisi descriptiva preliminar de les variables del *dataset* permet validar la coherència del conjunt i identificar trets estructurals rellevants per a la modelització posterior. Les variables temporals i econòmiques presenten rangs amplis però realistes: el temps base de lliurament oscil·la entre 1 i 45 dies amb una mitjana propera als 18 dies, mentre que el cost d'enviament varia entre aproximadament 1.250 USD i 350.000 USD, reflectint diferències substancials en distància, pes, mode i urgència. La majoria de comandes no presenten retards, però existeixen casos extrems amb fins a 19 dies de demora. Els índexs de risc geopolític, severitat meteorològica i inflació es mantenen dins de rangs coherents amb el context econòmic i logístic internacional.

Pel que fa a les variables categòriques, el *dataset* es concentra en una xarxa relativament petita de **9 ciutats d'origen i 10 de destinació**, amb **Shanghai** destacant com a node bidireccional i el **corredor de Suez** com a eix estratègic dominant. Les set categories de producte es distribueixen de manera força equilibrada i el transport marítim (Sea) domina clarament amb més del 80% de les ordres. El conjunt cobreix una finestra temporal completa de **730 dies** (del 4 de gener de 2024 al 3 de gener de 2026), suficient per analitzar l'evolució del risc i els retards al llarg del temps.

L'aplicació d'un criteri *z-score* ($|z| > 3$) identifica valors extrems en variables com Delay_Days i Shipping_Cost_USD. Aquests valors no es consideren errors sinó **esdeveniments rellevants per a l'anàlisi de resiliència**.

VI. Model Semàntic del KG

El model semàntic constitueix el nucli conceptual del *Knowledge Graph* desenvolupat en aquest treball. El seu objectiu és transformar les observacions operatives del *dataset* original en una representació semàntica estructurada de la cadena de subministrament, on les relacions entre entitats són tan rellevants com els seus atributs.

Un *Knowledge Graph* és una estructura de dades formada per **nodes** (entitats del domini) i **arestes** (relacions entre elles), on tant els uns com les altres poden incorporar **propietats** i **tipus específics**. A diferència del model relacional, que organitza la informació en taules connectades per claus, el *Knowledge Graph* representa explícitament les relacions com a elements de primera classe. Aquesta diferència és substancial: les consultes de travessia de múltiples nivells (per exemple, *quins productes crítics viatgen per rutes vulnerables a conflictes geopolítics?*) s'expressen de manera natural i s'executen amb eficiència. En un model relacional, en canvi, requeririen múltiples *joins* aniuats amb un cost computacional significatiu.

A. Components del Graf

El *Knowledge Graph* transforma el *dataset* tabular en una xarxa d'entitats interconnectades. No totes les entitats tenen la mateixa naturalesa: algunes representen estructures persistents del sistema logístic (rutes, ciutats, productes), d'altres representen esdeveniments individuals (els enviaments concrets), i d'altres encapsulen informació contextual multidimensional.

Els nodes del model es classifiquen en **tres categories**. Els **nodes transaccionals** capturen esdeveniments operatius individuals: l'entitat central és Order, que representa cada enviament concret amb les seves propietats temporals, econòmiques i d'estat. Els **nodes mestres** (*master data*) representen les estructures persistents del sistema: el model n'inclou set (Route, City, Country, ProductCategory, TransportMode, DisruptionType i MitigationAction). Finalment, els **nodes de context** encapsulen informació multidimensional mitjançant la tècnica de **reificació**: l'entitat RiskAssessment agrupa les diferents dimensions del risc (geopolític, meteorològic i macroeconòmic) en lloc d'afegir-les com a propietats directes de Order.

Les relacions del *Knowledge Graph* també es classifiquen en tres famílies funcionals. Les **relacions operatives** són *Order*-cèntriques i defineixen les característiques operatives de cada enviament. Les **relacions de risc i resposta** capturen els esdeveniments disruptius i les estratègies de mitigació activades. Les **relacions estructurals** defineixen la topologia de la xarxa logística amb independència dels enviaments individuals, incloent la relació CITY_FLOW, una capa agregada origen–destinació derivada dels enviaments.

B. Esquema conceptual

Havent definit els components bàsics del model, la Figura 1 presenta l'esquema conceptual. Les entitats centrals són els enviaments (Order), que es connecten amb les rutes que utilitzen (SHIPPED_VIA), les ciutats d'origen i destí (ORIGIN_FROM, DESTINATION_TO), el producte transportat (TRANSPORTS), els esdeveniments que els afecten (AFFECTED_BY), les accions de mitigació aplicades (MITIGATED_WITH) i les avaluacions de risc associades (HAS_RISK).

El catàleg detallat de cada entitat i cada relació del model, amb les seves propietats, cardinalitats i justificacions de disseny, es presenta a l'Annex B, lliurat com a document complementari a aquesta memòria.

C. Decisions de modelatge i dependències estructurals

Dues decisions de modelatge condicionen de manera significativa les capacitats analítiques del model.

La primera és la **reificació de l'avaluació de risc** com a entitat pròpia (RiskAssessment) en lloc de modelar-la com

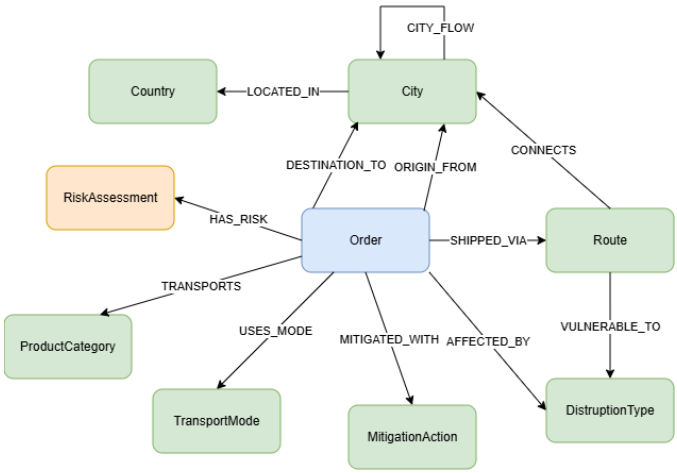

Figura 1. Esquema conceptual del *Knowledge Graph*. Els colors identifiquen categories de nodes: blau (transaccionals), verd (mestres), taronja (context).

a conjunt de propietats directes de Order. El risc associat a un enviament és intrínsecament multidimensional (geopolític, meteorològic, macroeconòmic) i agregar-lo en una entitat dedicada permet consultes que filtren simultàniament per múltiples dimensions sense sobrecarregar l'entitat Order amb desenes de propietats heterogènies, alhora que facilita l'extensió futura del model amb noves dimensions de risc.

La segona és la introducció de la relació **CITY_FLOW** com a **capa agregada origen–destinació** derivada dels enviaments individuals. A diferència de les relacions transaccionals, CITY_FLOW consolida per a cada parell de ciutats el volum d'enviaments, els indicadors de rendiment mitjà i la distribució per ruta. Aquesta agregació estructural permet aplicar de manera eficient algorismes de la llibreria *Neo4j Graph Data Science* [7] (centralitat, *community detection*, *shortest path*) sobre la topologia de la xarxa, sense necessitat de recórrer els 15.000 enviaments individuals, i és el que fa viable la simulació d'escenaris *what-if* del capítol VIII.

D. Cardinalitat del Graf

El model complet consta dels components recollits a la Taula II.

Taula II
CARDINALITAT DELS COMPONENTS DEL *Knowledge Graph*.

Component	Quantitat
Nodes totals	30.046
Enviaments (transaccionals)	15.000
Avaluacions de risc (RiskAssessment)	15.000
Entitats mestres	46
Relacions totals	120.100
Relacions operatives	75.000
Relacions de risc	45.000
Relacions estructurals	100

VII. IMPLEMENTACIÓ DEL KG

A. Arquitectura tecnològica

Per a la implementació del *Knowledge Graph* s'ha seleccionat Neo4j com a sistema de gestió de base de dades. Es tracta d'una base de dades de grafs nativa basada en el model de graf de propietats, adequada per representar estructures altament relacionals com la cadena de subministrament modelada en aquest treball.

La seva elecció es fonamenta en tres criteris principals: alineació directa amb el model semàntic definit (nodes i relacions), eficiència en consultes de travessia de múltiples nivells i disponibilitat d'una edició gratuïta suficient per al context acadèmic del projecte. Alternatives com Amazon Neptune o ArangoDB van ser considerades, però no aportaven avantatges rellevants per a l'abast i escala del TFG.

La instància s'ha executat en local mitjançant Neo4j Desktop, facilitant el desenvolupament i la validació del model.

B. Preparació del dataset

1) Neteja i normalització: A partir de l'exploració del conjunt de dades redissenyat, la primera etapa del *pipeline* s'ocupa de netejar i normalitzar les dades: tractament de valors nuls, normalització de formats i validació de la consistència.

El primer pas adreça els valors nuls a Disruption_Event: tal com s'ha establert a la secció anterior, l'absència de valor indica semànticament que no s'ha produït cap incidència. Per tant, aquests registres s'omplen amb la categoria explícita No Disruption, convertint un valor nul en informació semànticament vàlida i consistent. A continuació es du a terme la normalització geogràfica: les variables Origin_City i Destination_City, emmagatzemades en format compost (Ciutat, Codi de país), es descomponen en columnes independents per a la ciutat, el codi de país, el nom complet del país i la regió geogràfica (Àsia, Europa o Amèrica).

La fase de detecció d'inconsistències aplica regles de negoci per identificar anomalies semàntiques (coherència entre temps planificat i base, entre estat de lliurament i retard, entre mitigació i pertorbació, costos atípics per IQR i consistència ruta–regions). Els registres que les incompleixen es marquen amb una bandera descriptiva sense eliminar-los, ja que els casos excepcionals són rellevants per a l'anàlisi de resiliència.

2) Transformació i enriquiment de dades: La fase de transformació enriqueix el conjunt netejat amb camps derivats necessaris per a la representació en forma de graf. En primer lloc, s'assignen identificadors numèrics únics a les entitats categòriques (categories de producte, tipus de pertorbació i accions de mitigació) i es normalitzen els noms de les pertorbacions a format *snake_case* estable.

Seguidament es construeixen les classificacions derivades: la gravetat del retard s'estratifica en cinc nivells en funció dels dies de demora acumulats, i el risc combinat de cada comanda es calcula com una combinació ponderada de l'índex de risc geopolític (pes 0,6) i l'índex de severitat meteorològica (pes 0,4), classificat posteriorment en quatre nivells (*low, medium, high, critical*).

S'afegeixen mètriques d'eficiència per a cada ordre: la desviació percentual del temps real respecte al planificat, el

cost per quilogram de càrrega i el ràtio de retard relatiu al temps programat. Per a l'anàlisi de resiliència, es determina l'efectivitat de la mitigació de cada ordre disruptiva en funció de si el lliurament va ser puntual, va acumular un retard lleu o un retard significatiu, i es calcula la desviació del cost real respecte al cost de referència del mateix segment de ruta en condicions normals (cost vs. *baseline*). Finalment, s'atribueix a cada ordre un identificador d'avaluació de risc que la vincularà al seu node RiskAssessment al graf.

C. Construcció del graf de coneixement

1) Extracció de nodes: El *pipeline* de construcció del graf extreu del conjunt de dades transformat les entitats que formaran els nodes de Neo4j, distingint entre nodes d'instància i nodes mestres.

Els nodes Order (15.000 en total) concentren tots els atributs operatius i les mètriques derivades: dates, temps de lliurament, costos, estat de lliurament i els indicadors calculats durant la transformació. Per a cada comanda es genera paral·lelament un node RiskAssessment que agrupa exclusivament els indicadors de risc (índex geopolític, índex meteorològic, taxa d'inflació i puntuació de risc combinada), mantenint-los separats per facilitar les consultes centrades en risc.

Els nodes mestres s'obtenen per deduplicació sobre els valors únics de cada camp categòric: Route per als cinc tipus de ruta, City per a cada localitat única amb atributs de grau d'entrada i sortida, Country per als onze països representats, ProductCategory per a les set categories de producte, TransportMode per als modes de transport disponibles, DisruptionType per a cada tipus de pertorbació i MitigationAction per a cada acció de resposta, aquestes dues últimes enriquides amb mètriques agrupades sobre el seu impacte mitjà en cost i retard.

2) Modelització de relacions: Seguint les tres famílies de relacions definides al model semàntic, la construcció del graf n'implementa el detall tècnic i els atributs que carrega cada relació.

Les **relacions operatives** connecten cada comanda amb les entitats que hi intervenen directament: la ciutat de procedència, la de destinació, la ruta emprada, la categoria de producte transportada (amb el pes com a atribut) i el mode de transport.

Les **relacions de risc i resposta** associen cada comanda al tipus de pertorbació que ha patit, a l'acció de mitigació aplicada i al seu node d'avaluació de risc individual, separació que permet analitzar l'efectivitat de les mitigacions sense accedir als atributs de cada ordre.

Les **relacions estructurals** defineixen la topologia de la xarxa amb independència de les comandes individuals. La relació LOCATED_IN situa cada ciutat dins del seu país; CONNECTS enllaça cada ruta amb les ciutats que serveix, indicant si la ciutat hi actua com a origen, destinació o ambdós rols; i VULNERABLE_TO associa cada ruta amb els tipus de pertorbació que hi han ocorregut, quantificant-ne la freqüència i la gravetat mediana observada. Dins d'aquesta família, la relació agregada **CITY_FLOW** consolida els fluxos entre cada parell de ciutats, agregant sobre totes les comandes el nombre d'ordres, el pes total, les rutes i modes emprats, els costos i temps de lliurament mitjans, i les taxes de retard i pertorbació.

3) Càrrega a Neo4j i validació: La càrrega es realitza amb operacions MERGE sobre l'identificador únic de cada node, fet que la fa idempotent i re-executable sense duplicats, amb restriccions d'unicitat prèvies sobre cada etiqueta. Un cop completada, es verifica la integritat del graf: absència de relacions amb extrems pendents, cardinalitats esperades per cada tipus de relació i capacitat de respondre les consultes analítiques per a les quals va ser dissenyat.

VIII. Anàlisi de Resiliència

Un cop carregat i validat, el *Knowledge Graph* queda preparat per a la seva explotació analítica mitjançant l'aplicació web, que integra tant les anàlisis estructurades en blocs com l'agent IA en llenguatge natural. Aquest capítol presenta la capa analítica, organitzada en **sis blocs** independents que responen, de manera quantitativa, a les preguntes que es plantejaria un analista en l'estudi de la resiliència d'una xarxa logística. Cada bloc se centra en una pregunta clau amb una metodologia diferenciada (agregacions descriptives, perfils de risc, mètriques de teoria de grafs, *clustering*, anàlisi econòmic i simulació d'escenaris) i s'exposa com una pàgina independent dins de l'aplicació.

La seqüència dels blocs no és arbitrària: respon a una progressió que va de la **caracterització descriptiva** del sistema (Blocs 1–2), passa per l'**anàlisi estructural i geogràfica** de la fragilitat (Blocs 3–4) i l'impacte econòmic de les disrupcions (Bloc 5), i culmina en la **simulació d'escenaris de fallada** (Bloc 6), que explota el potencial diferencial del graf. A les seccions següents es presenta cada bloc amb les seves visualitzacions, resultats i interpretació.

A. Baselines de rendiment operatiu

L'anàlisi de la resiliència requereix establir prèviament una línia base del comportament del sistema en condicions normals: sense aquesta referència, cap observació posterior sobre risc o fragilitat té un punt de comparació clar. Aquest bloc respon, doncs, a la pregunta fonamental de tot analista: *com es comporta la xarxa quan no passa res excepcional?*

El bloc parteix d'una caracterització global mitjançant un conjunt de KPIs operatius (volum total d'ordres, taxa de retard i de disrupció, *lead time* mitjà i percentil 95, cost mitjà, desviació respecte al planificat) que sintetitzen en una sola lectura l'estat del sistema. Sobre aquest *baseline*, es descompon el tràfic en les dimensions estructurals més rellevants per a la presa de decisions logística — ruta, mode de transport i categoria de producte —, i s'observa la distribució de la severitat dels retards i la seva evolució temporal. Aquesta descomposició permet detectar concentracions, comportaments asimètrics i potencials biaixos en el funcionament normal de la xarxa. Finalment, el bloc introdueix una anàlisi de **redundància origen–destinació** basada en la relació agregada CITY_FLOW, que mesura, per a cada parell de ciutats, en quantes rutes diferents es distribueixen els seus enviaments i amb quin grau de concentració. Aquesta visió, que va més enllà de la perspectiva tradicional ordre-a-ordre, identifica els parells OD més dependents d'un únic corredor i serveix com

a pont conceptual cap a l'anàlisi de vulnerabilitat estructural (Bloc 3) i la simulació d'escenaris de fallada (Bloc 6).

Els resultats confirmen una xarxa amb un comportament operatiu globalment estable però amb una **concentració estructural elevada**: el corredor de Suez canalitza una proporció dominant del tràfic global i un nombre significatiu de parells origen–destinació depèn d'una única ruta, situació que prefigura la fragilitat que s'analitzarà en els blocs posteriors. La Figura 3 (Annex C) mostra la distribució del tràfic per ruta.

B. Exposició al risc i correlació amb resultats

Un cop caracteritzat el comportament normal, la pregunta natural és *on es concentra el risc i fins a quin punt es tradueix en resultats operatius adversos.* Aquest bloc separa explícitament l'**exposició al risc** (potencial) dels seus **efectes operatius** (realitzats), una distinció imprescindible per identificar zones de vulnerabilitat latent abans no es manifestin com a problemes.

L'anàlisi s'articula al voltant de la relació HAS_RISK, que vincula cada ordre amb una avaluació multidimensional del seu risc (geopolític, meteorològic i agregat). El bloc parteix d'una distribució global dels nivells de risc i la seva correlació amb les taxes de retard i disrupció, que permet validar empíricament si els nivells assignats prediuen efectivament els resultats. A continuació, l'anàlisi es descompon per les dimensions estructurals més rellevants — **ruta** i **categoria de producte** —, separant les contribucions geopolítica i meteorològica per identificar corredors amb perfils de risc qualitativament diferents. Aquesta descomposició es complementa amb una perspectiva geogràfica que avalua l'exposició al risc des de cada ciutat d'origen i de destí, així com una consulta d'exposició conjunta que identifica els escenaris on els riscos geopolític i meteorològic actuen simultàniament. Finalment, el bloc tanca amb la identificació de ***lanes* OD crítics per concentració de risc**, una vista directa de quins parells origen–destinació concentren la major part de l'exposició i que enllaça conceptualment amb l'anàlisi de fragilitat estructural del bloc següent.

Els resultats evidencien que **el risc no es distribueix uniformement per la xarxa**: el corredor Suez concentra la major part del risc geopolític, mentre que el risc meteorològic afecta de manera més transversal el conjunt de rutes; aquesta asimetria justifica que les estratègies de mitigació, analitzades al Bloc 5, hagin de ser específiques per tipus de disrupció i no genèriques. La Figura 4 (Annex C) mostra com la taxa de disrupció i de retard creix amb el nivell de risc assignat.

C. Topologia de la xarxa i fragilitat estructural

Més enllà del comportament operatiu i de risc, la resiliència depèn d'una propietat estructural: *quins nodes i corredors són crítics, és a dir, la fallada dels quals comprometria el funcionament del conjunt del sistema?* Aquesta propietat és independent del risc assignat —una ciutat pot operar amb normalitat i risc baix però ser un *single point of failure*. Aquest bloc hi respon aplicant mètriques de teoria de grafs amb la llibreria *Neo4j Graph Data Science* (GDS) sobre la projecció no dirigida de CITY_FLOW.

L'anàlisi s'estructura en dues perspectives. A nivell de **ciutats**, primer es caracteritza el rol estructural de cada node segons el balanç entre fluxos sortints i entrants, classificant-les en *hubs* balancejats, *hubs* d'exportació o d'importació i nodes purs d'origen o destí. Sobre aquesta caracterització s'apliquen quatre mètriques de centralitat complementàries: la ***degree centrality***, que compta el nombre de connexions directes; la ***weighted degree***, que pondera aquestes connexions pel volum d'enviaments; el **PageRank**, que mesura la importància d'una ciutat segons la dels seus veïns; i la ***betweenness centrality***, que captura el paper d'una ciutat com a connector irreemplaçable entre regions. Aquesta darrera s'ha destacat com a mètrica principal perquè és l'única que mesura **irreemplaçabilitat estructural**, propietat directament rellevant per a la resiliència. A nivell de **rutes**, el bloc complementa l'anàlisi amb un mapa de vulnerabilitat basat en la relació VULNERABLE_TO, que identifica quines rutes són sensibles a quins tipus de disrupció.

Els resultats revelen una xarxa amb **una asimetria estructural marcada**: només Shanghai presenta una centralitat de pas significativa, derivada del seu rol bidireccional únic, mentre que a nivell de rutes, Suez és l'única vulnerable a conflictes geopolítics — un perfil que reforça les observacions del Bloc 2 i prepara la simulació d'escenaris del Bloc 6. La Figura 5 (Annex C) mostra els rols estructurals de les ciutats.

D. Comunitats logístiques i exposició territorial

A escala global, una pregunta amb implicacions estratègiques és: *existeixen agrupacions naturals dins la xarxa, i fins a quin punt els fluxos depenen d'unes poques connexions entre elles?* Una xarxa aparentment diversificada pot estar internament estructurada en **comunitats** amb dependències crítiques entre si. Identificar-les —i sobretot els fluxos que les connecten— revela concentracions de risc que les mètriques de centralitat individuals del bloc anterior no capturen.

L'anàlisi s'estructura en dues escales. La primera adopta una perspectiva **territorial**, agregant els fluxos per país i regió mitjançant la relació LOCATED_IN, per quantificar la concentració del tràfic en zones geogràfiques concretes i identificar dependències nacionals que podrien comprometre la resiliència en cas de fallada d'un únic territori. La segona aplica l'algorisme **Louvain Community Detection** [6] sobre la projecció ponderada de CITY_FLOW, que detecta agrupacions de ciutats fortament interconnectades pels fluxos d'enviaments. Sobre aquesta partició, el bloc analitza el comportament intern de cada comunitat (volum de tràfic, taxa de retard i risc agregat) i, com a aportació diferencial, els **fluxos inter-comunitat**: enllaços OD que travessen els límits d'una comunitat i que sovint constitueixen punts crítics de dependència, ja que la fallada d'un únic enllaç pot aïllar regions senceres del sistema. La consulta d'OD *pairs* **intra-comunitat** completa l'anàlisi al nivell més granular, identificant els parells dominants dins de cada agrupació.

L'aplicació de Louvain detecta **tres comunitats logístiques** clarament definides. La més gran, articulada al voltant de Shanghai, agrupa la majoria dels grans corredors transoceànics (Pacific, Atlantic i les rutes Suez/CoGH cap a nord d'Europa); la segona, vertebrada per Singapore i Rotterdam, concentra

els fluxos intra-asiàtics i les connexions amb Europa central; la tercera, més petita i amb perfil de risc clarament superior (cost mitjà i taxes de retard i disrupció més elevades), agrupa essencialment els fluxos amb origen a Mumbai. L'anàlisi inter-comunitat revela dependències crítiques en uns pocs enllaços: destaca singularment el cas de Mumbai cap a Shanghai, que actua com a **únic pont rellevant entre la tercera comunitat i la resta de la xarxa** — una propietat estructural amb implicacions directes per a la diversificació logística i que connecta amb la simulació de fallades del Bloc 6. La Figura 6 (Annex C) mostra un exemple de una comunitat.

E. Cost de les disrupcions i efectivitat de mitigació

Falta encara una pregunta directament accionable: *quin cost econòmic generen les disrupcions quan es materialitzen, i quines respostes són efectivament eficaces per contenir-les?* Saber **on** és vulnerable la xarxa no n'hi ha prou; cal saber **quant costa** una disrupció i **quines mitigacions compensen**. Aquest bloc quantifica l'impacte econòmic de cada tipus de disrupció i avalua l'efectivitat de les tres accions de mitigació disponibles.

L'anàlisi s'articula en dues perspectives complementàries. La primera quantifica el **cost de les disrupcions per tipus**, descomponent l'impacte econòmic (sobrecost respecte al *baseline*, retard mitjà i percentil 95) per a cada categoria — *Geopolitical Conflict*, *Port Congestion*, *Severe Weather (Cape Storms)* i *Severe Weather (Typhoon/Storm)* —, vinculades a les ordres mitjançant la relació AFFECTED_BY. Aquesta descomposició permet identificar quines disrupcions són més costoses no només en termes absoluts sinó també en termes de severitat per ordre afectada. La segona perspectiva avalua les **accions de mitigació** (*Standard Shipping*, *Re-routing* i *Expedited Air Freight*) connectades a través de la relació MITIGATED_WITH, mesurant per a cadascuna la taxa d'efectivitat, el retard residual, el sobrecost i la proporció d'enviaments recuperats dins de la finestra planificada. L'anàlisi es completa amb una **matriu disrupció $\times$ acció** que avalua l'efectivitat condicional de cada resposta segons el tipus de disrupció, ja que una mateixa acció pot ser òptima en un context i ineficaç en un altre.

Els resultats revelen un panorama **asimètric**. El *Geopolitical Conflict* és la disrupció més costosa amb diferència (sobrecost mitjà del 92,6% respecte al *baseline*), mentre que la *Port Congestion*, malgrat ser la més freqüent, té un impacte unitari molt més contingut. L'anàlisi condicional confirma que **no hi ha una acció universalment òptima**: l'*Expedited Air Freight* és imprescindible davant escenaris geopolítics, mentre que davant disrupcions portuàries el *Standard Shipping* ja resulta efectiu amb un sobrecost pràcticament nul — un resultat que justifica un enfocament de mitigació diferenciat per tipus de disrupció. La Figura 7 (Annex C) recull l'efectivitat per acció.

F. Simulacions d'escenari de fallada

Tot el capítol anterior és **observacional**: descriu el que ha passat o el que pot passar segons patrons històrics. La pregunta crítica de la resiliència, però, és prospectiva: *què passaria si una ruta o una ciutat clau quedés fora de servei?* Aquest bloc hi respon amb la simulació interactiva d'escenaris *what-if*, explotant la flexibilitat consultiva del *Knowledge Graph* per quantificar l'impacte de fallades i identificar camins alternatius en temps real —l'aportació diferencial més rellevant del treball respecte a un enfocament relacional.

L'anàlisi s'estructura en tres tipus d'escenaris. El primer simula la **fallada d'una o més rutes**: a partir d'un conjunt de rutes bloquejades, el sistema quantifica el volum d'enviaments afectats, classifica cada parell OD segons el seu estat de xoc (*fully_blocked*, *primary_loss* o *partial_loss*) i estima la penalització de cost i *lead time* de redirigir el tràfic cap a rutes alternatives observades històricament. El segon escenari simula la **fallada de nodes**: quan una ciutat queda fora de servei, l'anàlisi quantifica les *lanes* sortints i entrants afectades i identifica ciutats substitutes que històricament han pogut absorbir fluxos equivalents del mateix producte. El tercer escenari aplica l'algorisme **Dijkstra** sobre la projecció dirigida de CITY_FLOW per trobar el camí òptim entre qualsevol parell de ciutats segons el criteri d'optimització seleccionat (*lead time*, cost o risc), permetent comparar diferents estratègies de re-encaminament davant d'un mateix escenari de fallada.

Els resultats demostren la utilitat operativa d'aquesta capa. La simulació de **bloqueig de Suez** revela que el 33,64% del tràfic global quedaria afectat, amb diversos parells OD havent de redirigir-se cap a la **ruta del Cape of Good Hope**, una alternativa més llarga però sovint més econòmica que les rutes aèries d'urgència. La simulació de **fallada de Shanghai** quantifica un impacte de 4.661 ordres afectades en 7 *lanes*, confirmant el seu paper de *hub* bidireccional irreemplaçable detectat al Bloc 3. Finalment, el càlcul de **camins òptims** mostra com una mateixa ruta origen–destí pot canviar segons el criteri (per exemple, entre Tokyo i Nova York el camí mínim en cost és diferent del mínim en *lead time* o en risc), una funcionalitat que dóna a l'analista una eina concreta de suport a la decisió i tanca el capítol amb la materialització pràctica de tot l'anàlisi de resiliència. La Figura 8 (Annex C) visualitza l'escenari de bloqueig de Suez.

IX. AGENT IA

L'agent IA s'integra a l'aplicació com a pàgina addicional, oferint una via d'accés en llenguatge natural complementària a les pàgines analítiques. La seva motivació funcional és permetre que un usuari de negoci pugui interrogar el *Knowledge Graph* sense necessitat de conèixer Cypher ni l'esquema del model.

A. Arquitectura i flux de raonament

L'agent s'ha implementat seguint el patró **ReAct** (*Reasoning + Acting*) [5] sobre la llibreria *LangGraph*, que estructura el processament d'una pregunta com un cicle iteratiu de raonament, acció i observació. A diferència d'una crida única al LLM, aquest patró permet a l'agent descompondre preguntes complexes en múltiples consultes successives al graf.

L'arquitectura combina **dos models LLM amb rols diferenciats**: un model d'interfície ràpid i econòmic (Gemini 2.5 Flash) per a tasques lleugeres, i un model de raonament més capaç (Gemini 2.5 Pro) que dirigeix el bucle ReAct i genera

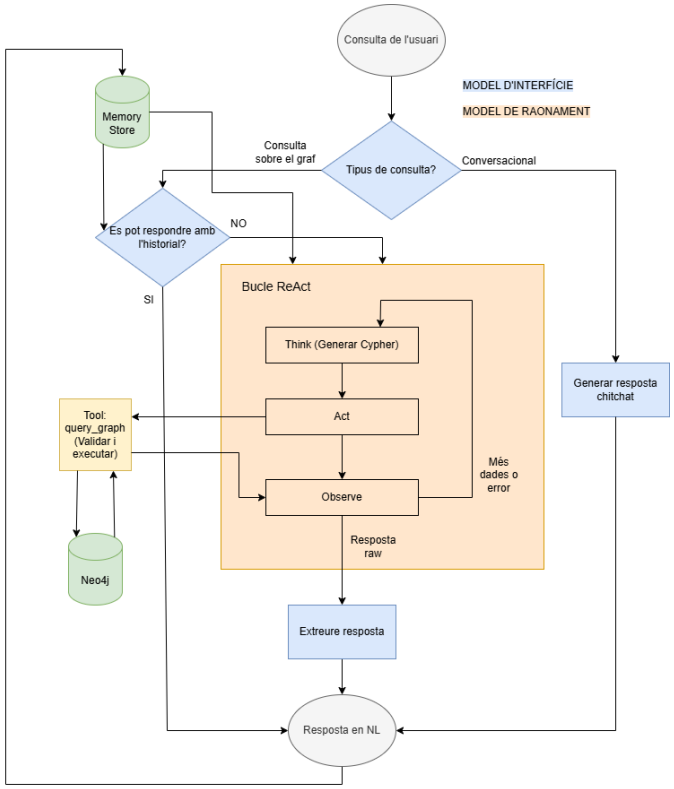

Figura 2. Flux de raonament de l'agent ReAct. Els dos camins ràpids eviten consultes innecessàries; el bucle principal incorpora una capa de validació determinista sobre cada Cypher generat.

les consultes Cypher. Abans d'entrar al bucle, l'agent aplica **dos camins ràpids** que eviten consultes innecessàries al graf: respostes a preguntes conversacionals (*chitchat*) i respostes basades en el context conversacional ja disponible.

El punt clau de l'arquitectura és que la **seguretat no depèn del comportament del LLM**, sinó d'una **capa de validació determinista** independent que intercepta cada consulta Cypher abans d'executar-la a Neo4j. Aquesta capa rebutja operacions d'escriptura (MERGE, CREATE, DELETE, SET, REMOVE, DROP), restringeix els procediments APOC a una llista blanca, valida els noms d'entitats i relacions contra l'esquema permès, i exigeix la presència d'una clàusula LIMIT. A més, els *prompts* del sistema imposen dues regles addicionals: respondre sempre en l'idioma de l'usuari i no exposar mai termes interns del model.

Opcionalment, l'agent pot generar visualitzacions a partir dels resultats obtinguts del *Knowledge Graph*. Un cop produïda la resposta i els *DataFrames* associats, un model d'interfície proposa fins a tres gràfics, cadascun definit mitjançant una especificació estructurada que passa per una validació de coherència amb les dades disponibles.

B. Validació del comportament

Per validar el funcionament de l'agent s'ha dissenyat una bateria de *prompts* organitzats en sis famílies, cadascuna destinada a verificar un aspecte específic del comportament esperat.

Les **consultes simples** i **de complexitat mitjana** validen la capacitat de l'agent de traduir correctament preguntes en llenguatge natural a consultes Cypher, des de casos directes (rànquings, agregacions per ruta o categoria) fins a casos amb múltiples filtres i agregacions condicionals. Les **consultes complexes** validen l'aprofitament del valor diferencial del *Knowledge Graph*, incloent simulacions *what-if* (bloqueig de rutes, eliminació de *hubs*) i càlcul de camins òptims amb Dijkstra. Les **consultes sobre informació inexistent** o inconsistent amb el model validen que l'agent reconeix els límits de la seva informació i ofereix alternatives en lloc d'inventar respostes. Els *prompts* que **forcen operacions de modificació** (*delete* i *write*) validen que el comportament defensiu funciona tant a nivell de capa de validació com a nivell d'instruccions del *prompt*. El *prompt* sobre **l'arquitectura del graf** valida la confidencialitat de l'esquema, és a dir, que l'agent no exposa noms d'entitats, relacions ni detalls tècnics del model en la seva resposta. Finalment, els *prompts* **conversacionals i de seguiment** (*follow-up*) validen el correcte funcionament del camí ràpid de *chitchat* i la persistència contextual a través de la memòria de sessió.

L'objectiu d'aquesta validació no és establir una mètrica quantitativa global de qualitat, sinó **comprovar de manera dirigida que cada decisió de disseny té efecte observable** en el comportament de l'agent davant dels escenaris més rellevants i de risc.

X. Discussió i Limitacions

A. Interpretació dels resultats

L'anàlisi desenvolupat revela un missatge transversal: la xarxa combina una **concentració estructural elevada** en alguns nodes i corredors (Suez com a eix dominant, Shanghai com a *hub* bidireccional irreemplaçable) amb una **organització comunitària** que, lluny de diversificar el risc, introdueix dependències inter-comunitàries crítiques en uns pocs enllaços. Aquests dos vectors **es reforçen mútuament**: la concentració de tràfic en pocs corredors fa que la fallada d'un sol enllaç pugui aïllar comunitats senceres del sistema.

Des d'un punt de vista metodològic, aquests resultats confirmen que certes preguntes —irreemplaçabilitat de Shanghai, enllaç crític Mumbai–Shanghai, camins òptims amb pesos variables— només es responen eficientment mitjançant una representació en graf.

B. Limitacions del treball

Tot i la viabilitat demostrada del plantejament, el treball presenta limitacions que cal explicitar per emmarcar correctament l'abast dels resultats i orientar treballs futurs.

Cobertura d'actors i topologia. El model representa la xarxa logística de transport entre ciutats, però no incorpora tots els actors d'una cadena de subministrament real: queden fora **proveïdors** (origen físic dels productes), **magatzems i centres de distribució** (punts amb capacitats d'emmagatzematge) i **clients finals**. Aquesta simplificació implica que el model captura el flux de transport però no el flux d'aprovisionament integral, i tampoc pot quantificar la propagació d'una disrupció més enllà del moviment físic (com l'impacte d'una ruptura

d'estoc al magatzem central). A més, la topologia es limita a **13 ciutats** que representen els *hubs* logístics principals, deixant fora desenes de ports regionals i punts de transbordament que existirien en una xarxa real.

Cobertura i naturalesa de les dimensions de risc. El model captura tres dimensions de risc — geopolítica, meteorològica i macroeconòmica —, però una cadena de subministrament real està exposada a moltes més: regulatori (canvis duaners, sancions), laboral (vagues, conflictes sindicals), energètic (volatilitat del preu dels combustibles), tecnològic (ciberatacs) o reputacional. A més, el risc es modela de manera **estàtica**: cada enviament té uns valors de risc fixos que no s'actualitzen en funció d'esdeveniments dinàmics ni de l'evolució de la xarxa. La incorporació de noves dimensions, així com d'un model de risc dinàmic, seria una extensió natural facilitada per l'estructura reificada de RiskAssessment descrita anteriorment.

Naturalesa sintètica de les dades i abast temporal. Els 15.000 registres del *dataset* han estat generats sintèticament mitjançant un model probabilístic reproduïble. Això comporta que els percentatges concrets dels resultats (33,64% del tràfic via Suez, 4.661 ordres afectades per la caiguda de Shanghai...) reflecteixen les decisions del generador i no observacions empíriques: **el que el treball valida és la metodologia, no la quantificació específica d'una xarxa real**. A més, el *dataset* cobreix un horitzó temporal de dos anys, suficient per analitzar la dinàmica d'esdeveniments puntuals però insuficient per capturar cicles plurianuals com estacionalitats llargues, transicions estructurals o efectes macroeconòmics persistents.

XI. Conclusions i Treball Futur

A. Conclusions principals

Aquest treball ha demostrat la viabilitat de modelar una cadena de subministrament global com un *Knowledge Graph* per a l'anàlisi de resiliència, integrant-ho en una aplicació interactiva amb un agent IA. Els quatre objectius plantejats s'han assolit.

Pel que fa a *O1* i *O2*, **la representació en *Knowledge Graph* ha habilitat preguntes que els enfocaments tabulars no responen eficientment**: els algorismes de teoria de grafs (centralitat, detecció de comunitats, *shortest path*) han permès identificar punts de fragilitat estructural i quantificar dependències crítiques.

Pel que fa a *O3*, **la integració de l'agent IA amplia l'accessibilitat de l'eina** als usuaris no tècnics, però només resulta segura amb mecanismes deterministes de control de l'abast (rebuig d'escriptures, ocultació de l'esquema i respostes acotades al domini), validats amb *prompts* adversarials.

Pel que fa a *O4*, l'aplicació transforma la presa de decisions sobre la resiliència de la xarxa: en lloc de dependre de la intuïció, l'analista disposa d'evidència quantitativa per fonamentar-les. A més, és reutilitzable, ja que la seva parametrització permet adaptar-la a altres xarxes sense modificar el codi.

Com a apreciació transversal, el **redisseny del *dataset*** no ha estat un detall tècnic sinó una condició necessària: sense topologia real, rutes alternatives i un model de risc coherent, els algorismes de graf no haurien aportat valor. El generador sintètic reproduïble és, per tant, una contribució metodològica del treball, en la línia d'una alternativa **oberta i reproduïble** a les solucions comercials propietàries del sector.

Finalment, la revisió de sistemes similars (secció III) confirma que la direcció del treball és coherent amb l'evolució de la indústria —plataformes com Resilinc ja incorporen agents conversacionals sobre representacions de xarxa—, amb el valor diferencial que aquest treball ofereix una alternativa **oberta, transparent i reproduïble** a unes solucions comercials potents però propietàries.

B. Compliment de la planificació

La planificació establerta (secció II) s'ha cobert en la seva totalitat. Les dues desviacions principals —el redisseny del *dataset* i la reorientació de l'arquitectura de l'agent— es van absorbir dins el marge temporal previst, sense afectar l'abast ni cap dels objectius. La metodologia incremental per fases va permetre, precisament, detectar aviat la inviabilitat del *dataset* original i reaccionar-hi sense comprometre la resta del calendari.

C. Línies de treball futur

Les principals línies d'extensió que es deriven directament d'aquest treball són:

- **Validació amb dades reals**: aplicar la mateixa arquitectura a un *dataset* operatiu d'una empresa logística per quantificar l'aplicabilitat externa dels resultats.
- **Inclusió de la dimensió d'inventari**: modelar *buffers* als nodes (capacitats portuàries, estocs intermedis) i estendre la simulació *what-if*.
- **Detecció proactiva d'esdeveniments**: integrar fonts externes en temps real (notícies, alertes meteorològiques, índexs geopolítics) per actualitzar dinàmicament el risc i les probabilitats de disrupció del graf.

Agraïments

Vull expressar el meu agraïment al meu tutor, Josep Lladós, per la seva orientació i suport al llarg del desenvolupament d'aquest treball. També agraeixo a la meva família i als meus amics pel seu suport i la seva companyia, que han fet aquests anys de grau molt més fàcils de recórrer.

Referències

[1] Everstream Analytics, "Supply Chain Risk Management Platform," 2026. [En línia]. Disponible: https://www.everstream.ai/

[2] Resilinc, "Agentic AI Supply Chain Risk Platform," 2026. [En línia]. Disponible: https://resilinc.ai/

[3] V. Kumar, "Interoperable Knowledge Graphs for Localized Supply Chains: Leveraging Graph Databases and RDF Standards," *Logistics*, vol. 9, no. 4, p. 144, 2025. https://doi.org/10.3390/logistics9040144

[4] Y. Liu *et al.*, "A Knowledge Graph Perspective on Supply Chain Resilience," *arXiv preprint arXiv:2305.08506*, 2023. https://doi.org/10.48550/arXiv.2305.08506

[5] S. Yao *et al.*, "ReAct: Synergizing Reasoning and Acting in Language Models," in *Proc. 11th Int. Conf. on Learning Representations (ICLR)*, Kigali, Rwanda, 2023. https://arxiv.org/abs/2210.03629

[6] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, i E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, p. P10008, 2008. https://doi.org/10.1088/1742-5468/2008/10/P10008

[7] Neo4j Inc., *Neo4j Graph Data Science Library Manual*, v. 2.x, 2024. [En línia]. Disponible: https://neo4j.com/docs/graph-data-science/

APÈNDIX A
PARAMETRITZACIÓ DEL GENERADOR DE DADES SINTÈTIQUES

A. Introducció

Aquest annex documenta els paràmetres principals del generador descrit conceptualment al capítol IV. S'inclouen únicament els valors essencials per entendre la lògica del *dataset* i reproduir-ne la generació. La parametrització completa es troba accessible al fitxer `config.py` del repositori del projecte. El generador genera 15.000 comandes amb dates entre l'1 de gener de 2024 i el 31 de gener de 2026, utilitzant una llavor aleatòria fixa (SEED = 42).

B. Xarxa logística

El *dataset* modela 25 parells origen–destinació amb rutes alternatives. La Taula III mostra els parells amb més d'una ruta possible, on les probabilitats reflecteixen el comportament observat dels operadors el 2024. Els nou parells restants tenen una única ruta determinista.

Taula III
PARELLS ORIGEN–DESTINACIÓ AMB RUTES ALTERNATIVES I PROBABILITATS.

Origen	Destinació	Probabilitats per ruta
Shenzhen	Rotterdam	Suez 75% · CoGH 25%
Shanghai	Rotterdam	Suez 70% · CoGH 30%
Shanghai	Hamburg	Suez 70% · CoGH 30%
Mumbai	Felixstowe	Suez 85% · CoGH 15%
Mumbai	Antwerp	Suez 85% · CoGH 15%
Mumbai	Shanghai	Intra-Asia 70% · CoGH 30%
Busan	Rotterdam	Suez 65% · CoGH 35%
Busan	Hamburg	Suez 70% · CoGH 30%
Singapore	Hamburg	Suez 80% · CoGH 20%
Singapore	Rotterdam	Suez 80% · CoGH 20%
Rotterdam	Singapore	Suez 80% · CoGH 20%
Shanghai	Los Angeles	Pacific 95% · CoGH 5%
Busan	Los Angeles	Pacific 95% · CoGH 5%
Shanghai	New York	Pacific 80% · Suez 20%
Santos	Rotterdam	Atlantic 60% · CoGH 40%

C. Especialització productiva per origen

La distribució completa per categoria es troba al paràmetre PRODUCT_BY_ORIGIN del codi.

Taula IV
CATEGORIES DE PRODUCTE DOMINANTS PER CIUTAT D'ORIGEN.

Origen	Categories dominants
Shanghai, Shenzhen	Consumer Electronics i Semiconductors (55–65%)
Tokyo, Busan	Auto Parts i Semiconductors (50–55%)
Singapore	Pharmaceuticals, Consumer Electronics i Semiconductors (70%)
Mumbai	Textiles i Pharmaceuticals (65%)
Rotterdam, Hamburg	Auto Parts dominant (30–35%) amb mix balancejat
Santos	Raw Materials i Perishable Foods (70%)

D. Lead times per ruta

El temps base de lliurament es mostra d'una distribució uniforme dins del rang específic per cada ruta i mode. La Taula V recull els rangs, que reflecteixen les distàncies reals de navegació.

Taula V
RANGS DE TEMPS BASE (DIES) PER RUTA.

Ruta	Mode marítim	Mode aeri
Intra-Asia	5–14	1–2
Atlantic	8–18	1–3
Pacific	14–24	2–4
Suez	18–32	2–4
CoGH	25–45	3–5

E. Model de risc per ruta

El risc geopolític ($[0, 1]$) i la severitat meteorològica ($[0, 10]$) segueixen distribucions normals amb paràmetres per ruta.

Taula VI
MITJANA DEL RISC GEOPOLÍTIC I SEVERITAT METEOROLÒGICA PER RUTA.

Ruta	Risc geopolític	Severitat meteo.
Suez	0,65 (elevat)	3,5 (baix)
Pacific	0,35 (moderat)	7,0 (alt)
Intra-Asia	0,35 (moderat)	6,0 (alt)
CoGH	0,22 (baix)	6,5 (alt)
Atlantic	0,30 (moderat)	4,5 (moderat)

F. Matriu de mitigació condicional

L'acció de mitigació es mostra a partir d'una distribució condicional al tipus de disrupció. L'efectivitat de cada acció es modela com una reducció proporcional del retard: **0% per Standard Shipping**, **25% per Re-routing** i **55% per Expedited Air Freight**. Aquestes reduccions són parcials per disseny.

Taula VII
PROBABILITAT DE CADA ACCIÓ DE MITIGACIÓ SEGONS EL TIPUS DE DISRUPCIÓ.

Disrupció	Std. Ship.	Re-rout.	Exp. Air
Port Congestion	60%	25%	15%
Geopolitical Conflict	10%	35%	55%
Severe Weather (Typhoon)	35%	50%	15%
Severe Weather (Cape)	50%	40%	10%

APÈNDIX B
MODEL SEMÀNTIC DEL *Knowledge Graph*

El catàleg complet del model semàntic es lliura com a document independent (`TFG_Annex_B_Model_Semantic.docx`), que documenta sistemàticament, per a cada element, la definició,

les propietats —distingint entre originals (provinents del *dataset*) i derivades (calculades al *pipeline* ETL)—, la cardinalitat i la justificació de la decisió de modelatge. A continuació se'n resumeix l'estructura, ja comentada prèviament.

A. Catàleg d'entitats

El model defineix **nou entitats**. L'entitat transaccional central és **Order** (15.000 nodes), que concentra els atributs operatius de cada enviament i nombroses propietats derivades clau per a l'anàlisi (entre d'altres, *delay_severity*, *cost_vs_baseline_pct* i *mitigation_effectiveness*, aquesta darrera amb quatre nivells que capturen matisos operatius més enllà d'una avaluació binària). Les set entitats mestres representen les estructures persistents del sistema: **Route** (5), **City** (13, enriquida amb mètriques de centralitat GDS i rol estructural), **Country** (9), **ProductCategory** (7), **TransportMode** (2), **DisruptionType** (4) i **MitigationAction** (3, amb mètriques d'impacte agregades). Finalment, **RiskAssessment** (15.000) és el node reificat que encapsula les dimensions de risc de cada enviament i en deriva el *combined_risk_score* i el *risk_level*.

B. Catàleg de relacions

Les **dotze relacions** s'agrupen en tres famílies funcionals. Les **operatives** (*Order*-cèntriques: ORIGIN_FROM, DESTINATION_TO, SHIPPED_VIA, TRANSPORTS i USES_MODE) connecten cada enviament amb el seu origen, destí, ruta, producte i mode; totes són de cardinalitat N:1 i només TRANSPORTS porta una propietat (*weight_kg*). Les **de risc i resposta** (AFFECTED_BY, MITIGATED_WITH i HAS_RISK) vinculen cada enviament amb la disrupció patida, l'acció de mitigació i la seva avaluació de risc; HAS_RISK és 1:1, materialitzant la reificació del risc. Les **estructurals** defineixen la topologia amb independència dels enviaments: LOCATED_IN (jerarquia ciutat–país), CONNECTS (ruta–ciutat, amb la propietat *role* que indica si la ciutat hi actua com a origen, destí o ambdós) i VULNERABLE_TO (ruta–disrupció, amb *frequency* i *severity*, aquesta calculada com la mediana de la gravetat del retard de les ordres afectades). La capa agregada CITY_FLOW, ja descrita al cos, enriqueix cada parell de ciutats amb mètriques de volum, rendiment i diversificació de rutes, incloent-hi un índex de concentració HHI (*route_concentration*).

C. Propietats derivades

Una característica transversal del model és la distíncia entre propietats **originals** (provinents del *dataset*) i **derivades** (calculades al *pipeline* ETL i materialitzades als nodes). Aquestes últimes —com *delay_severity*, *cost_vs_baseline_pct*, *mitigation_effectiveness* o *combined_risk_score*— es precalculen per dos motius: optimitzar les consultes (filtrar per una categoria és immediat, mentre que recalcular-la sobre milions de registres seria costós) i oferir categories interpretables que faciliten l'agregació i la comunicació de patrons. És el cas de *mitigation_effectiveness*, que en lloc d'una avaluació binària distingeix quatre nivells (*not_applicable*, *fully_effective*, *partially_effective* i *not_effective*) per capturar matisos operatius, com ara que un retard lleu pugui considerar-se un resultat acceptable.

APÈNDIX C
RESULTATS VISUALS DE L'ANÀLISI DE RESILIÈNCIA

Aquest annex recull les visualitzacions principals generades per l'aplicació per a cadascun dels sis blocs de l'anàlisi de resiliència (capítol VIII).



Figura 3. Bloc 1 (*baseline*). Volum d'ordres per ruta. Suez concentra el 33,6% del tràfic global.

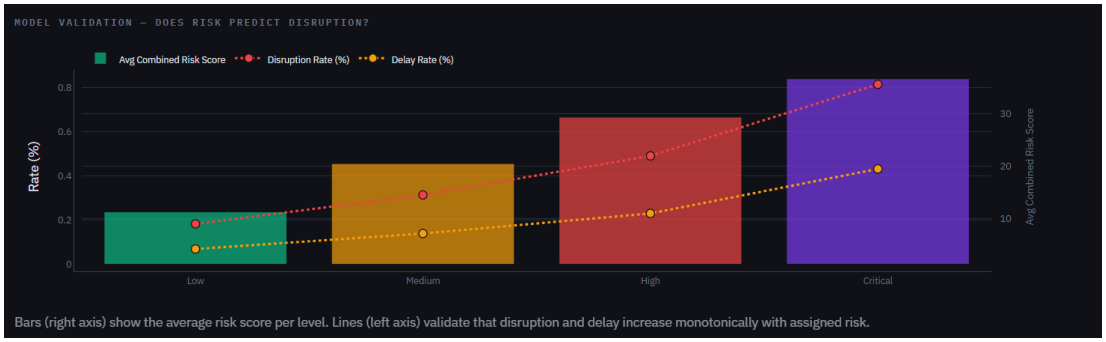

Figura 4. Bloc 2 (risc). Validació del model: la taxa de disrupció i de retard creixen monòtonament amb el risc combinat assignat per nivell (*low/medium/high/critical*).



Figura 5. Bloc 3 (topologia). Balanç de flux enfront del volum total de cada ciutat: -1 indica importador pur, $+1$ exportador pur i ≈ 0 *hub* bidireccional. La mida és proporcional al volum.



Figura 6. Bloc 4 (comunitats). Indicadors clau i xarxa interna d'una de les comunitats logístiques detectades per Louvain, amb el seu perfil de risc i concentració de ruta.



Figura 7. Bloc 5 (mitigació). Distribució de l'efectivitat (*fully/partially/not effective*) per a cada acció de mitigació, mesurada sobre el nombre de casos.

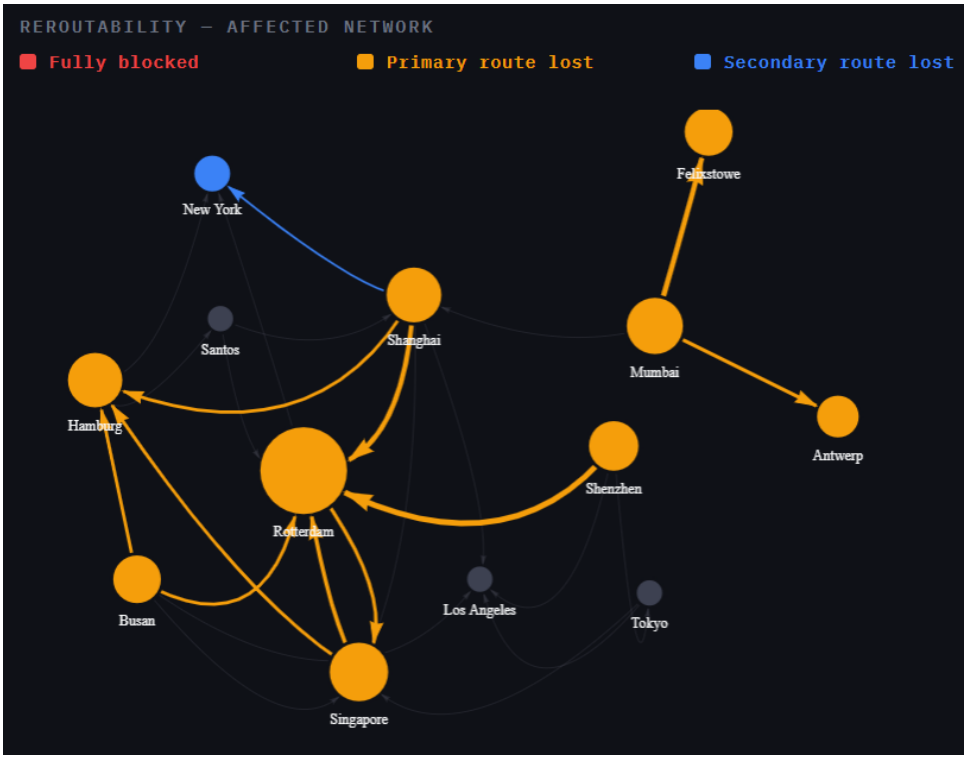

Figura 8. Bloc 6 (*what-if*). Xarxa afectada per la simulació de bloqueig de Suez: nodes segons l'estat de cada *lane* (*fully blocked*, *primary route lost*, *secondary route lost*).