

Not so simple graph visualization

Cucchi Matteo, Dileo Manuel, Pindaro Margherita, Uslenghi Filippo



[Repo Github](#)

Cos'è un grafo “non semplice”?

- Un grafo che “evolve”
- Un grafo che coinvolge entità astratte diverse
- Un grafo che scelti due nodi può avere più di due link
- Un grafo in cui gli archi coinvolgono più di due nodi

Agenda

1

Diffusione dell'informazione
Manuel Dileo

2

Reti Bipartite
Filippo Uslenghi

3

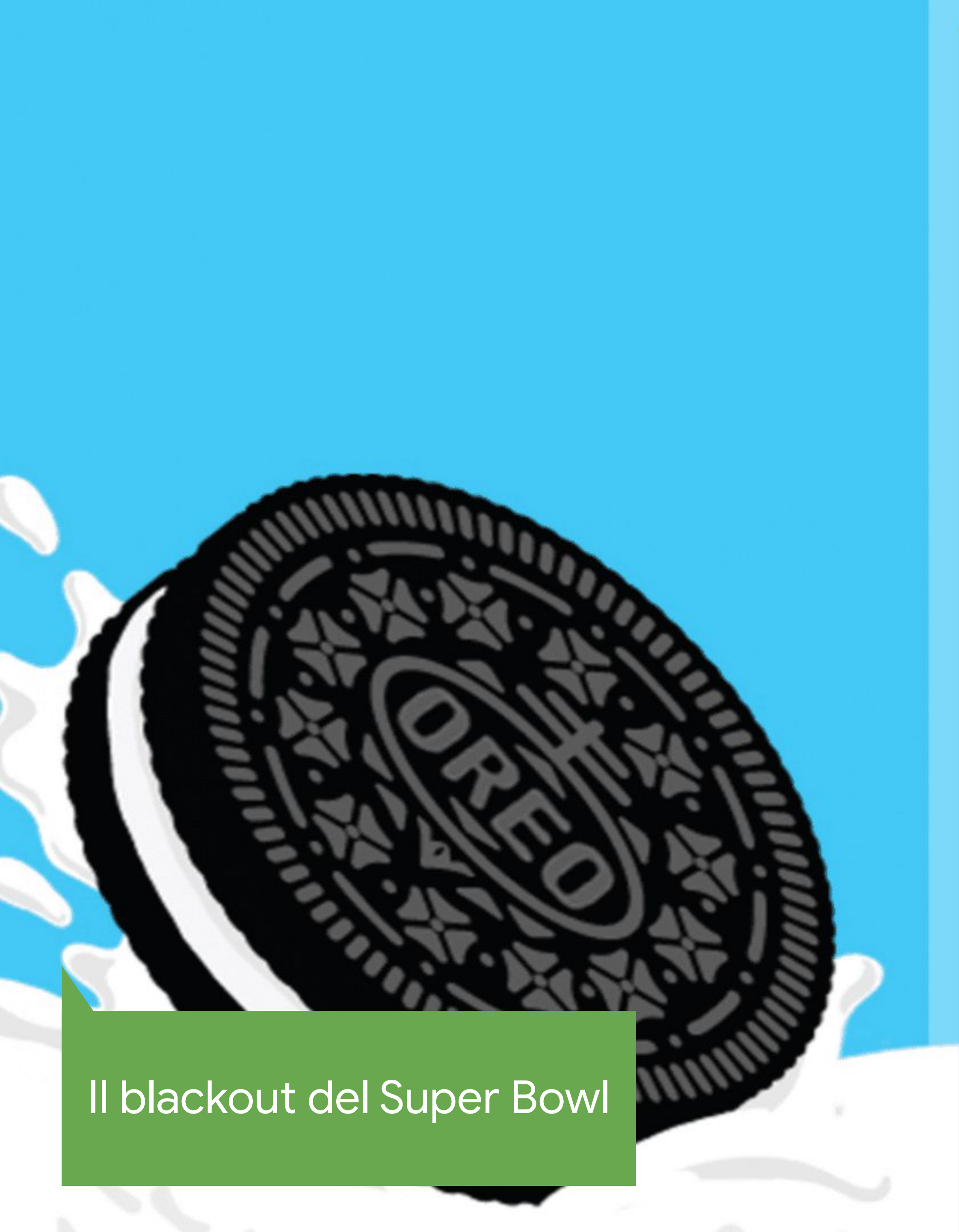
Multigrafi
Matteo Cucchi

4

**Multigrafi orientati eterogenei
applicati ai Database**
Margherita Pindaro

Information Diffusion

Manuel Dileo



Il blackout del Super Bowl



Oreo Cookie 
@Oreo



Following

Power out? No problem.
pic.twitter.com/dnQ7pOgC

 Reply  Retweet  Favorite  More



15,872
RETWEETS

6,381
FAVORITES

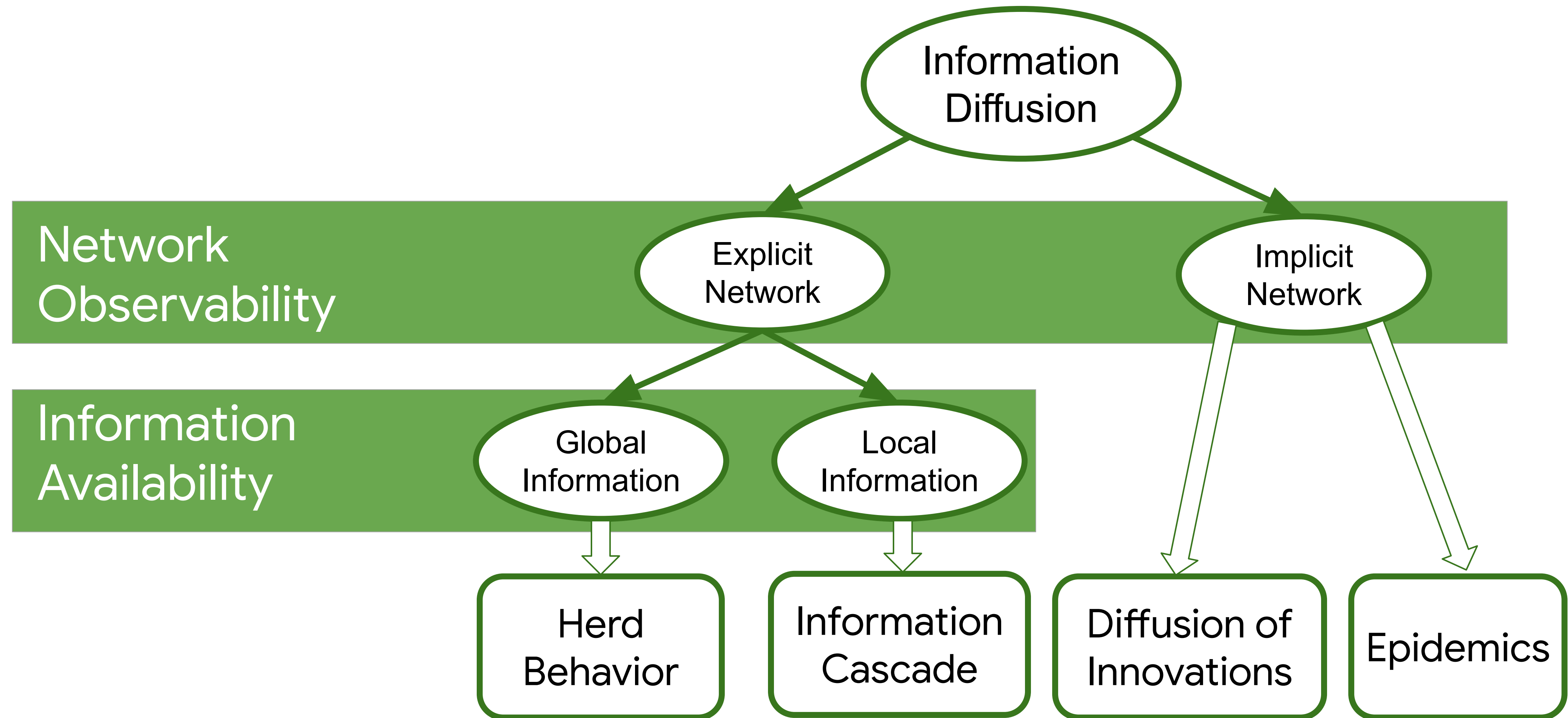


Cos'è la diffusione dell'informazione?

La **diffusione dell'informazione** è il processo con cui un pezzo di informazione (conoscenza) si diffonde raggiungendo individui attraverso le interazioni fra di essi.

Phenomena	Agent	Network
Venereal Disease	Pathogens	Sexual Network
Rumor Spreading	Information, Memes	Communication Network
Diffusion of Innovations	Ideas, Knowledge	Communication Network
Computer Viruses	Malwares, Digital viruses	Internet
Mobile Phone Virus	Mobile Viruses	Social Network/Proximity Network
Bedbugs	Parasitic Insects	Hotel - Traveler Network
Malaria	Plasmodium	Mosquito - Human network

Una possibile classificazione



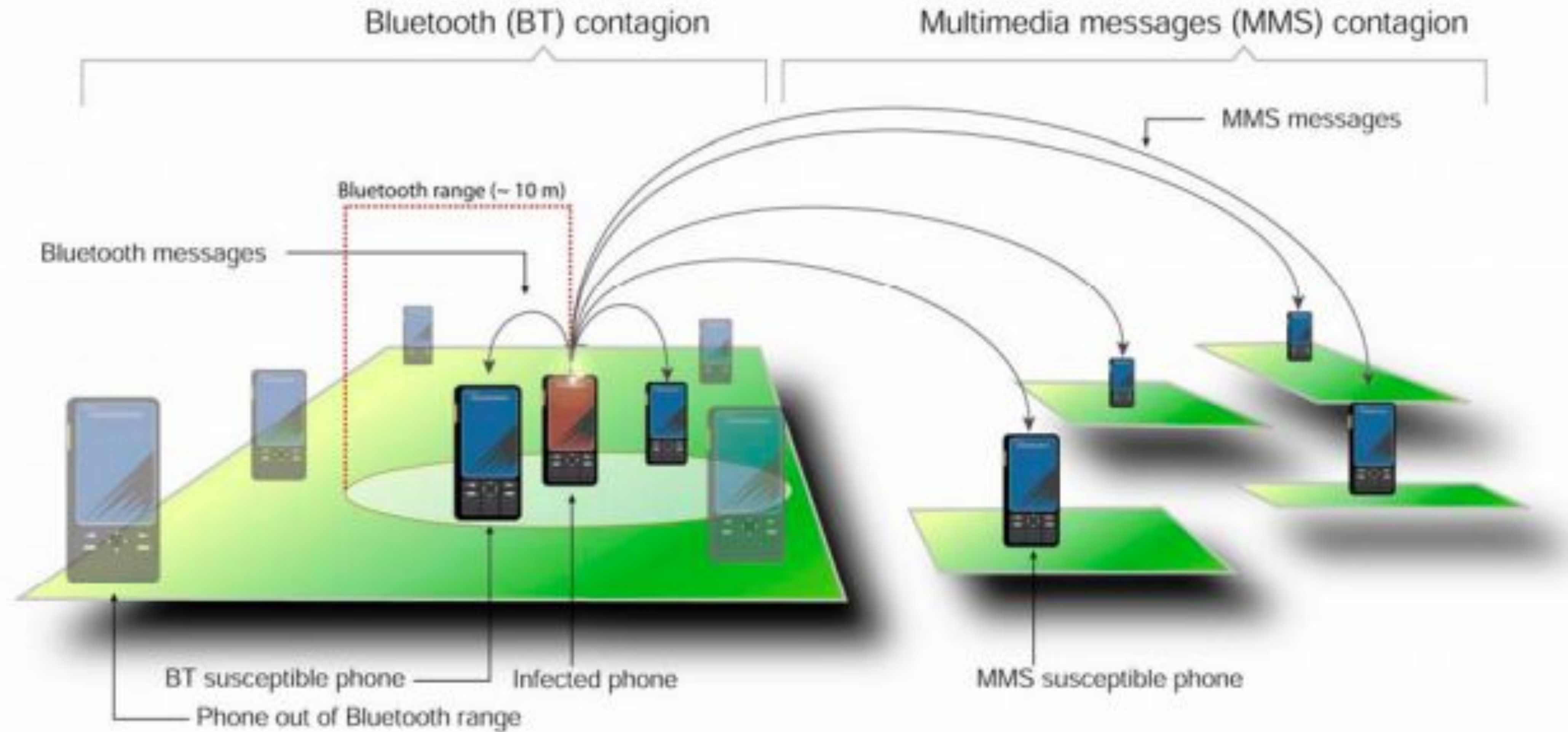
Guardiamo 3 esempi

1. Diffusione di virus su reti telefoniche
2. Diffusione di informazione su reti di collaborazione scientifica
3. Epidemie

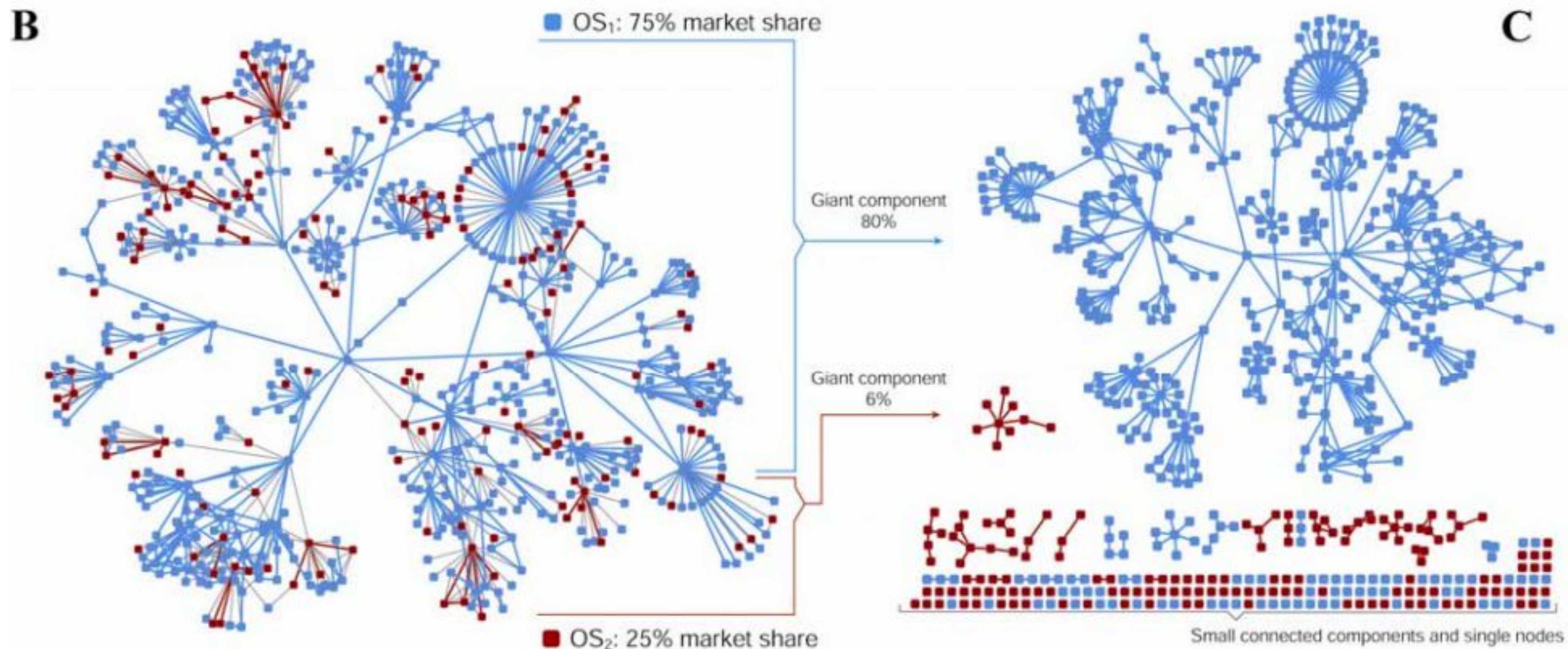
Guardiamo 3 esempi

- 1. Diffusione di virus su reti telefoniche**
2. Diffusione di informazione su reti di collaborazione scientifica
3. Epidemie

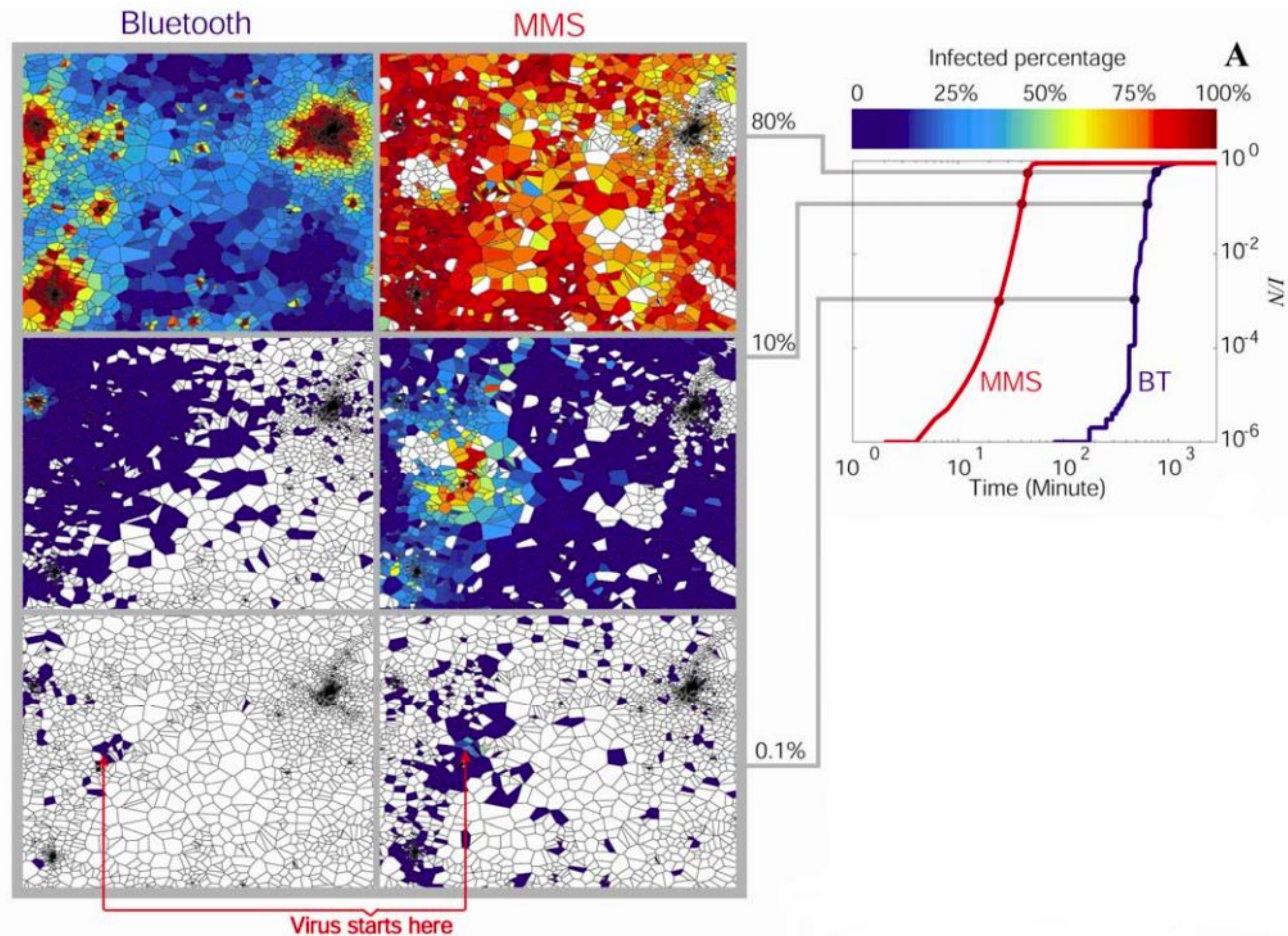
Pattern di diffusione di virus su cellulari



Il grafo delle chiamate



Spatial Spreading



Guardiamo 3 esempi

1. Diffusione di virus su reti telefoniche
- 2. Diffusione di informazione su reti di collaborazione scientifica**
3. Epidemie

Processo di diffusione da un singolo nodo

Spesso si vuole analizzare come un singolo nodo sia influenzato o stia influenzando la sua rete locale come fonte di un processo di diffusione dell'informazione.

Di seguito vedremo un **network layout algorithm** per grafi con milioni di nodi che visualizza i fenomeni di diffusione dalla prospettiva di un singolo nodo.

A Offline:

1. Global force-based layout in spherical space



Online:



Node selection

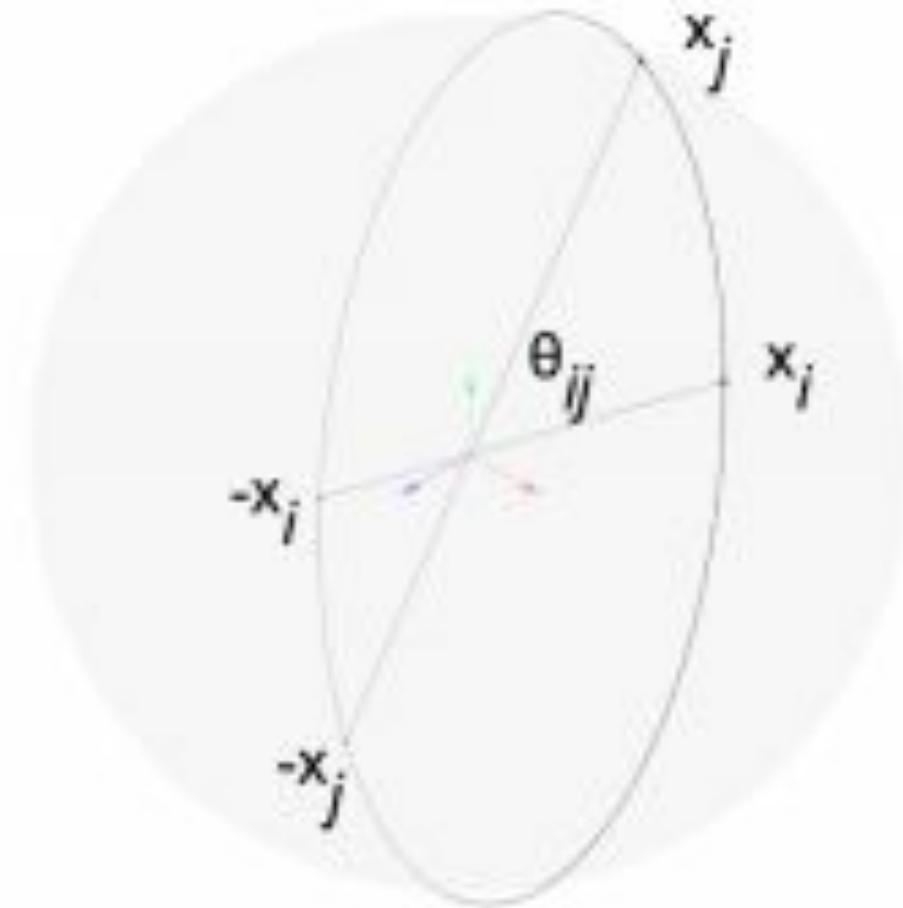
2. Local distance-based layout in spherical space



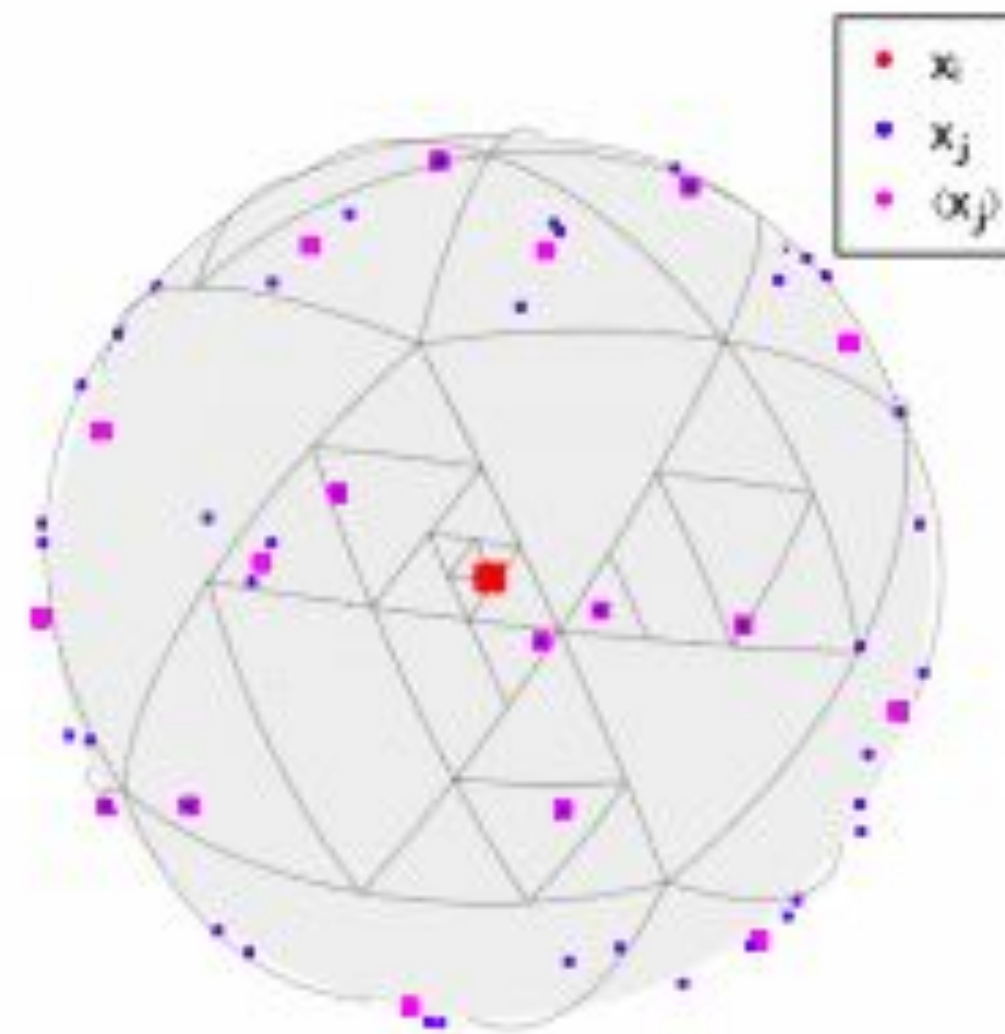
3. Rotation, map projection and 2D raster-based rendering



B

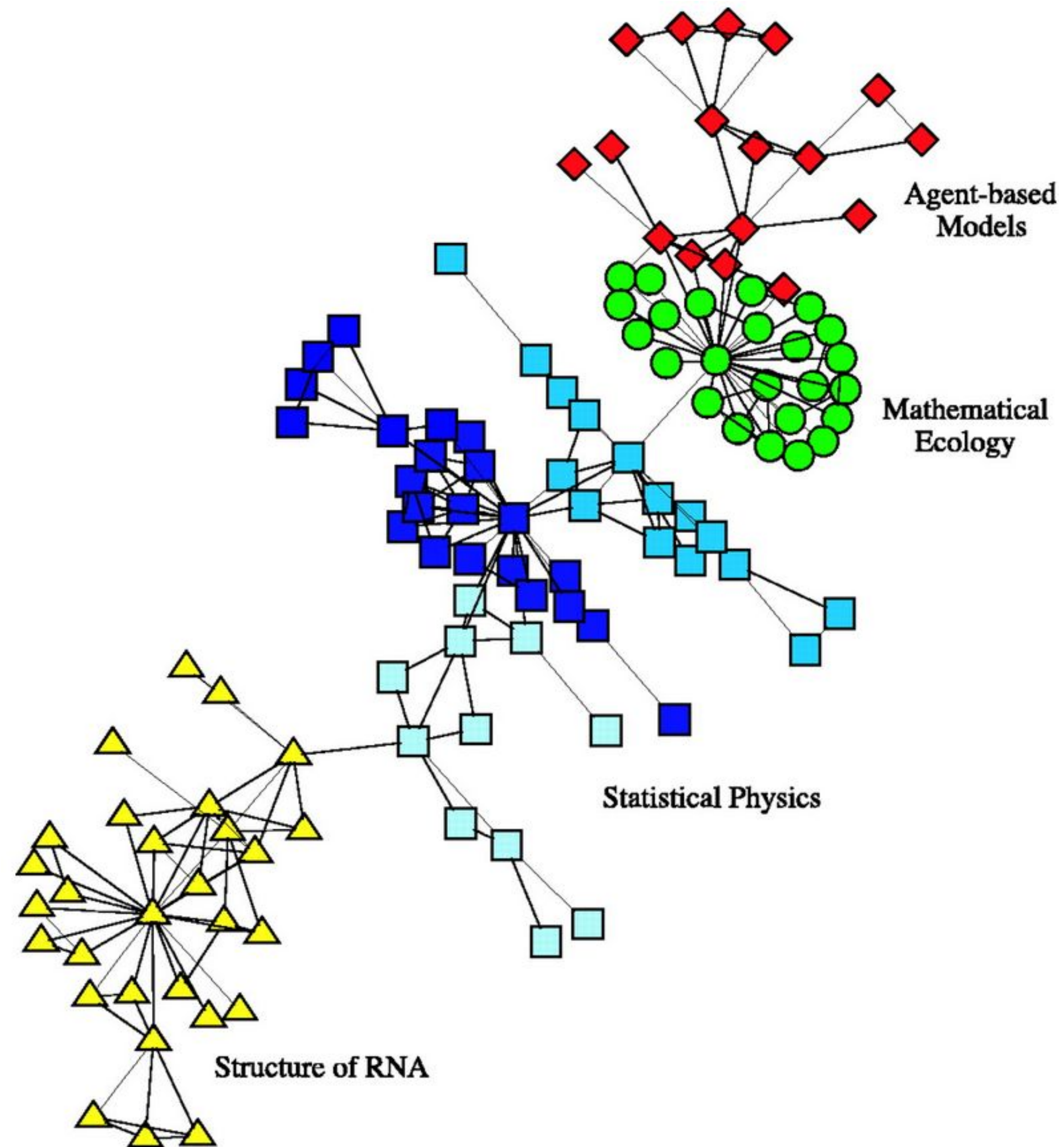


C



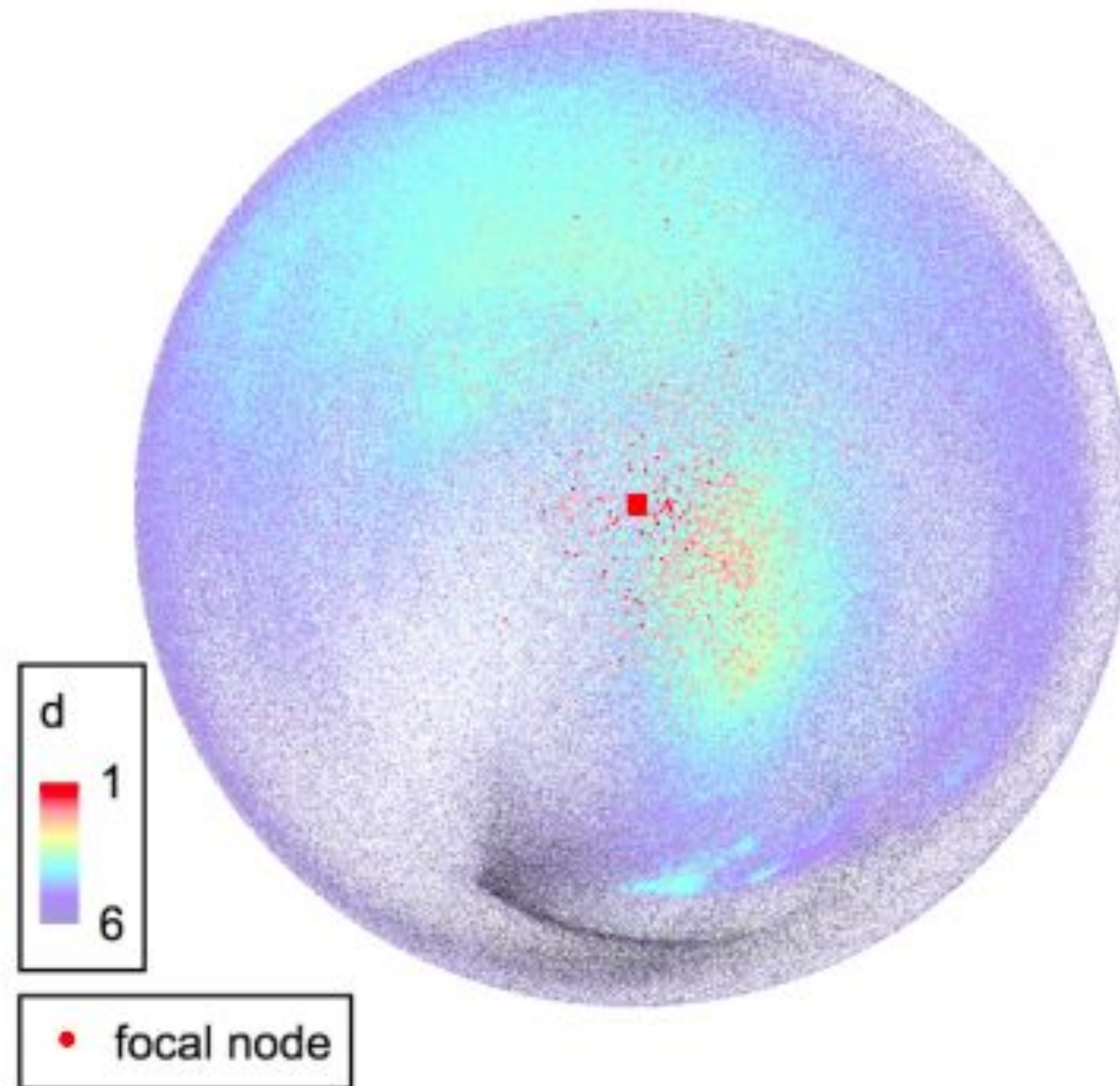
Come funziona
l'algoritmo?

Reti di collaborazione scientifica

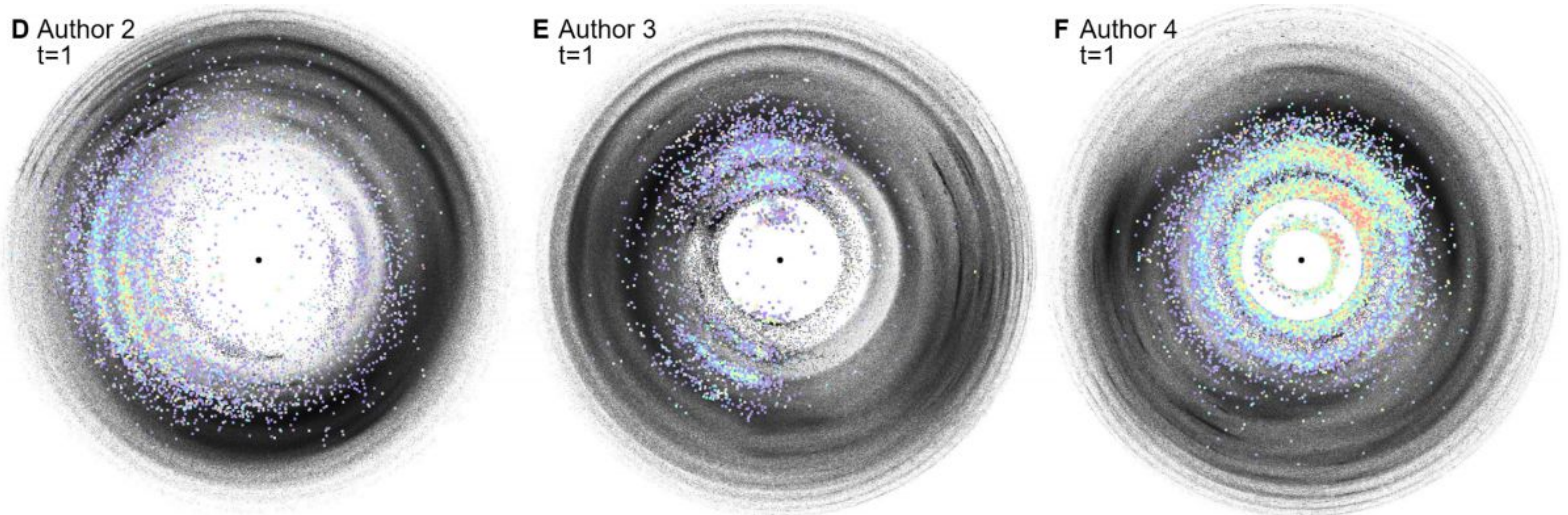


Un semplice esempio statico (gradi di separazione)

A WoS co-authorships



Visualizzare l'influenza di un certo autore



La prima volta in cui l'autore viene citato è mappata su uno schema di colori di raffreddamento (dall'inizio della carriera del **focal author** all'ultimo anno disponibile; $t: [0, 1] \rightarrow [\text{rosso}, \text{giallo}, \text{verde}, \text{blu}, \text{bianco}]$).

Guardiamo 3 esempi

1. Diffusione di virus su reti telefoniche
2. Diffusione di informazione su reti di collaborazione scientifica
- 3. Epidemie**

Diffusione di agenti patogeni

L'epidemiologia ha sviluppato un solido framework analitico e numerico per modellare la diffusione degli agenti patogeni.

Questo framework si basa su due ipotesi fondamentali:

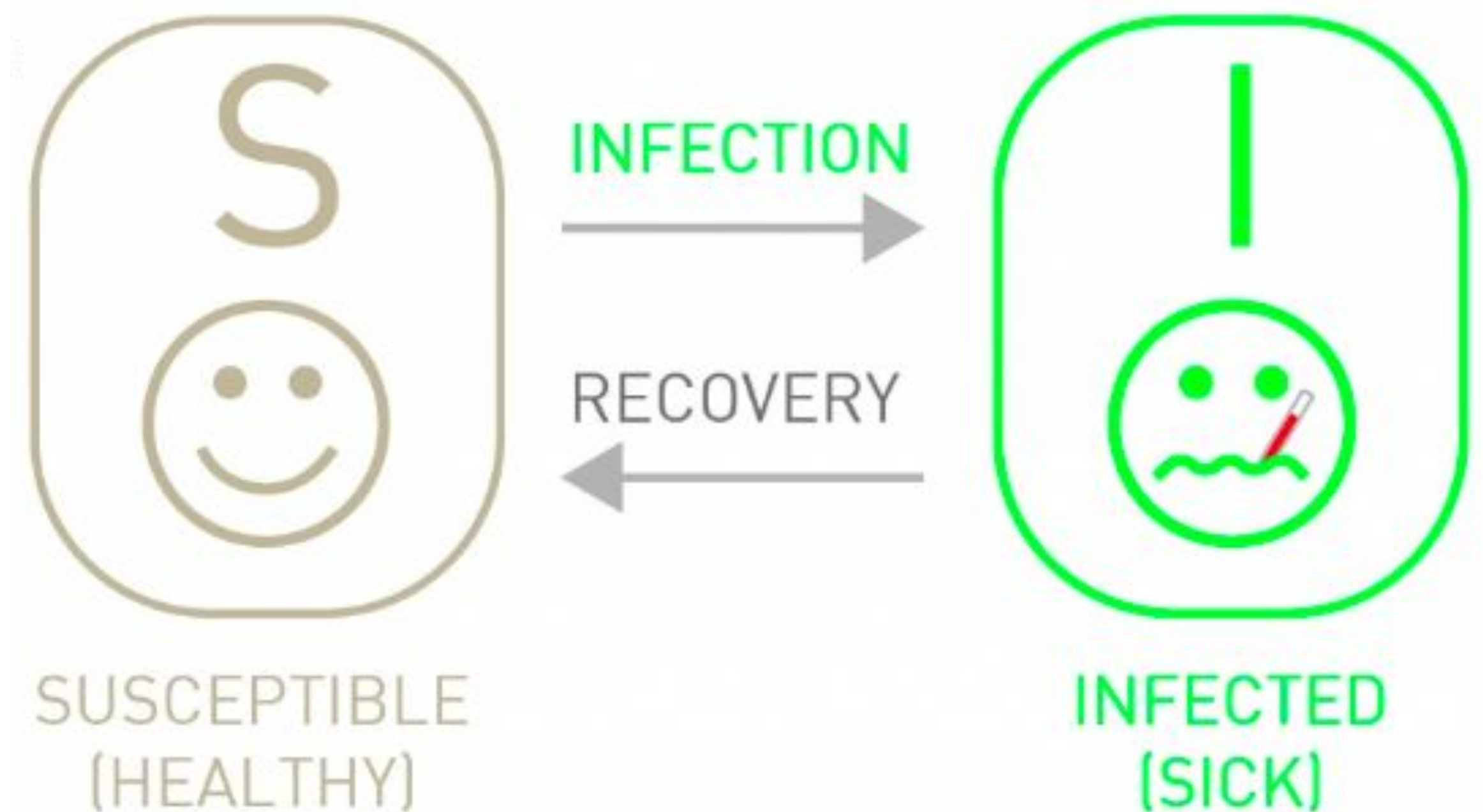
- L'ipotesi di compartimentalizzazione
- L'ipotesi di omogeneità

Modello SIS

Si consideri una malattia che si diffonde in una popolazione di N individui.

Indichiamo con $S(t)$ il numero di individui suscettibili (sani) al tempo t e con $I(t)$ il numero di individui che sono già stati infettati.

Se un singolo individuo viene infettato al tempo $t = 0$ (cioè $I(0)=1$), quanti individui saranno infettati in un momento successivo t' ?



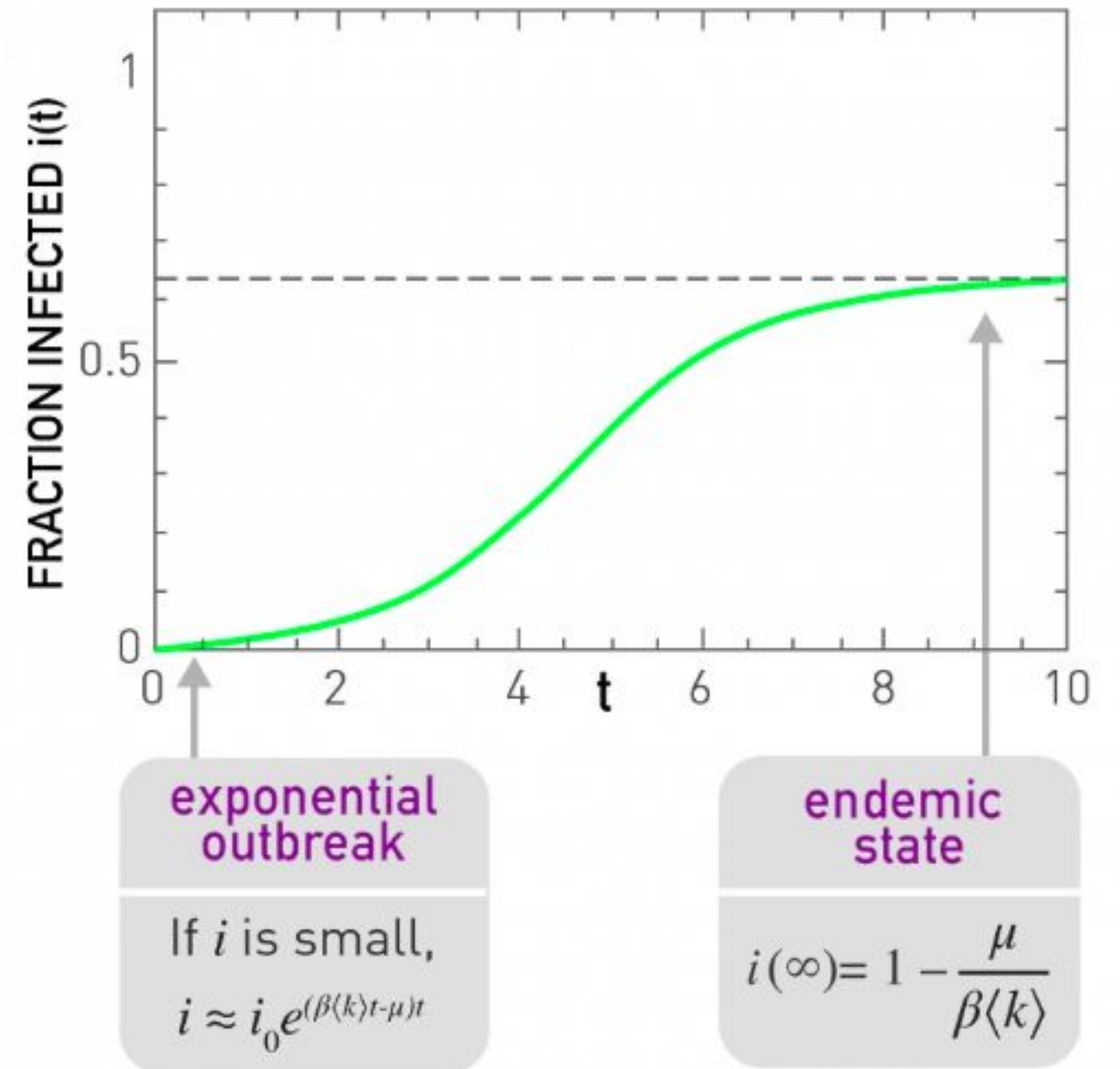
La frazione di individui infetti in funzione del tempo...

$$\frac{di}{dt} = \beta \langle k \rangle i(1 - i) - \mu i \longrightarrow i = \left(1 - \frac{\mu}{\beta \langle k \rangle}\right) \frac{C e^{(\beta \langle k \rangle - \mu)t}}{1 + C e^{(\beta \langle k \rangle - \mu)t}}$$

Dove:

- Un individuo ha mediamente $\langle k \rangle$ contatti.
- β è la probabilità che la malattia venga trasmessa da un individuo infetto a un soggetto suscettibile in un'unità di tempo.
- μ è il tasso di guarigione e μi cattura la velocità con cui la popolazione guarisce dalla malattia.

Don't panic!



Perchè non è così semplice?

- Le reti reali non sono statiche
- Le reti reali non sono omogenee
- Non abbiamo una conoscenza totale dei contatti tra le persone
- L'ipotesi di compartimentalizzazione viene violata

Bonus

Strategie di controllo - VAX | Gamifying Epidemic Prevention

Grafi bipartiti

Filippo Uslenghi

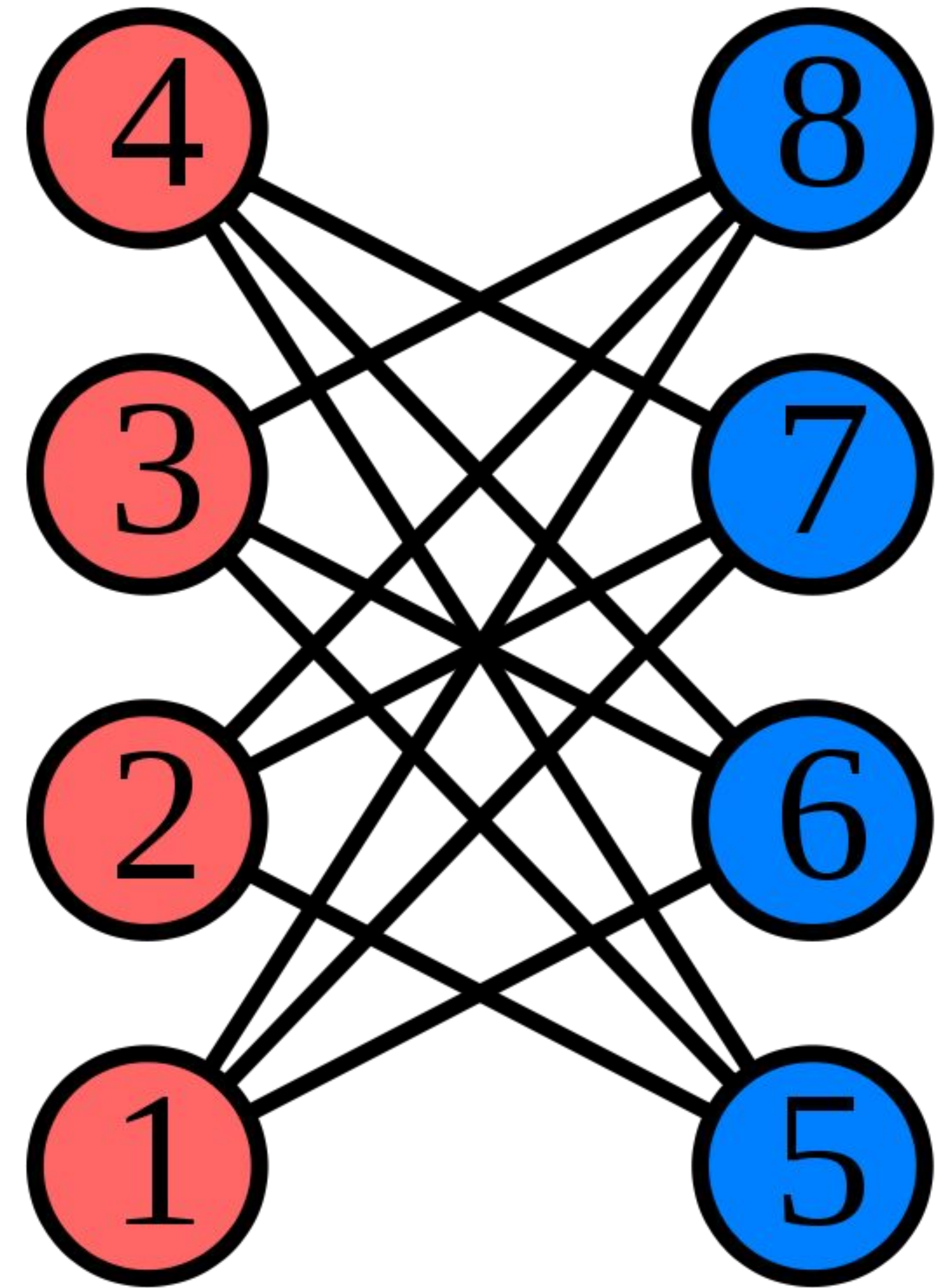
Cosa sono?

Per grafo bipartito si intende un grafo $G=(V, E)$ il cui insieme di nodi V può essere partizionato in due sottoinsiemi V_1, V_2 tali che:

- $V_1 \cap V_2 = \emptyset$
- $V = V_1 \cup V_2$
- Un arco può connettere nodi di V_1 con nodi di V_2

Vengono spesso rappresentati in questa forma:

- $G = (V_1, V_2, E)$



Reti di affiliazione

Le reti di affiliazione sono reti nelle quali vi sono due tipi differenti di nodi:

- Attori
- Gruppi

Le reti di affiliazione modellano sistemi in cui gli attori appartengono a determinati gruppi.

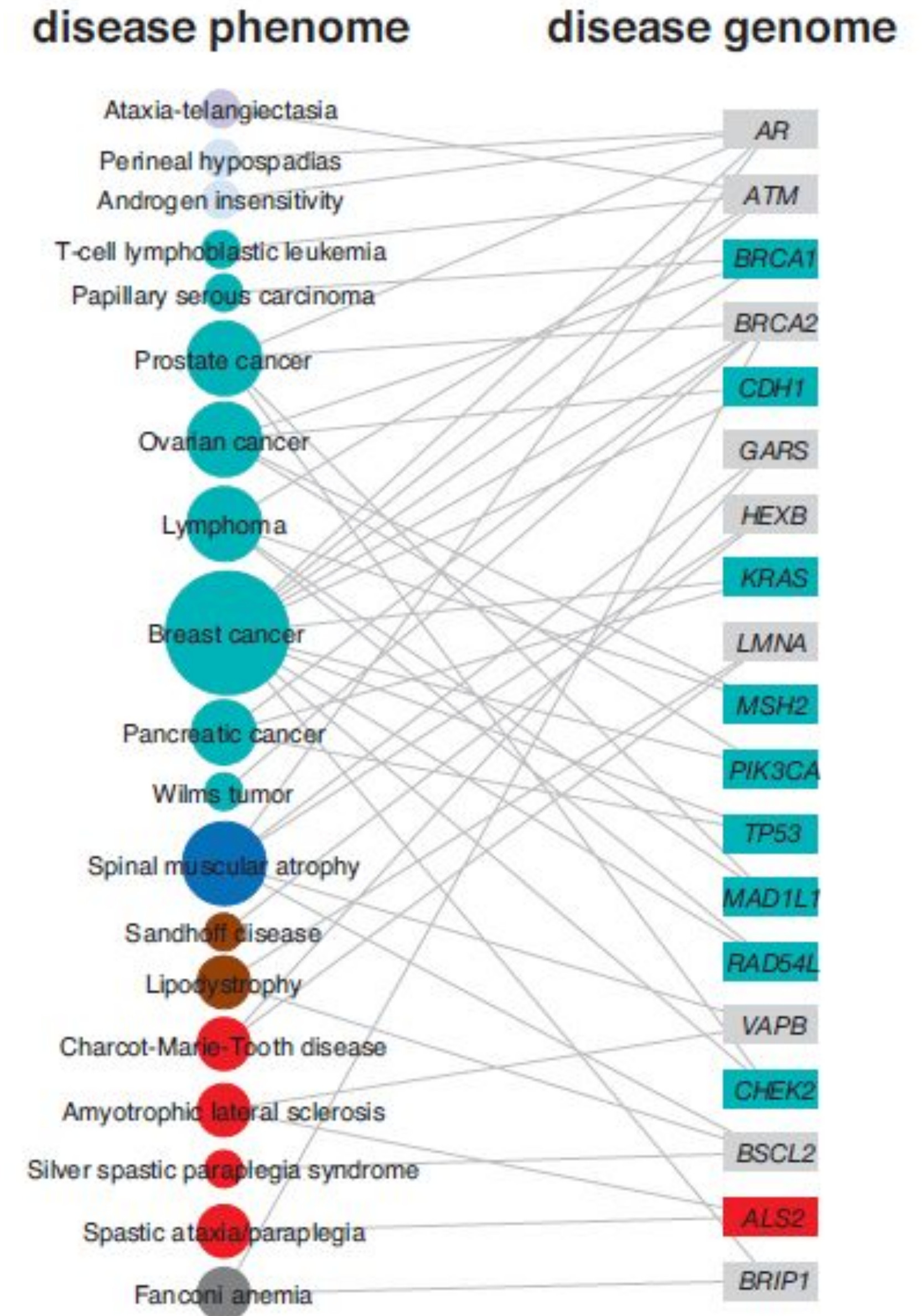
Un link può esistere solo tra nodi di tipo diverso e indica l'appartenenza di quell'attore a quel gruppo.

I grafi bipartiti ben modellano le reti di affiliazione

Un famoso esempio

- [The human disease network](#)
 - Kwang-Il Goh, Michael E. Cusick, David Valle, Barton Childs, Marc Vidal, Albert-László Barabási

I nodi sono divisi in malattie e geni e un link collega una malattia ad un gene se mutazioni di quel gene concorrono all'insorgere di quella malattia



Proiezione di una rete bipartita

È un'operazione che permette di scoprire le strutture relazionali tra nodi di uno stesso tipo riportando la rete bipartita ad una rete semplice.

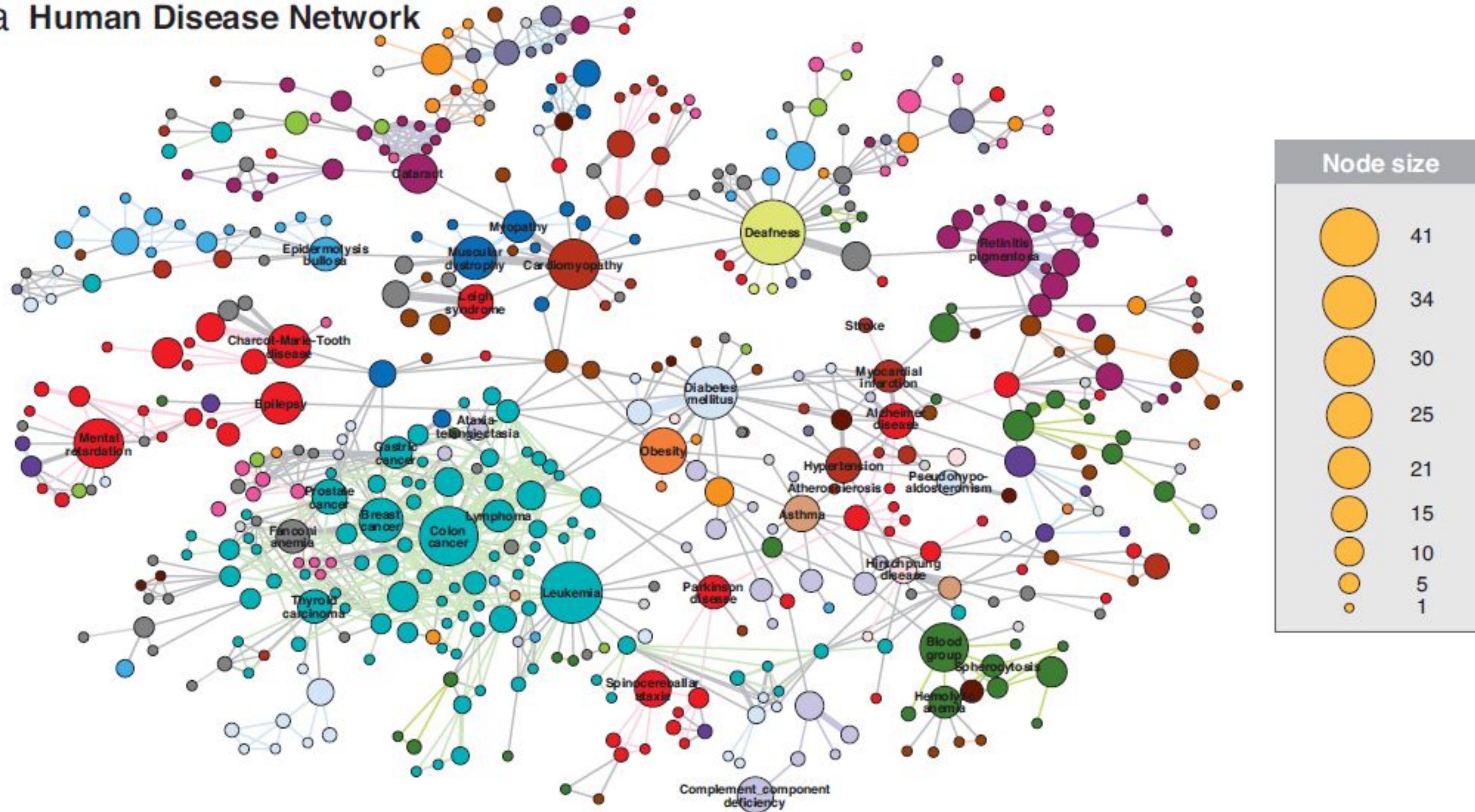
Il risultato è un grafo semplice in cui un arco collega due nodi se questi sono in relazione con lo stesso nodo dell'altra partizione.

Per evitare la perdita di informazioni gli archi devono essere pesati secondo un criterio:

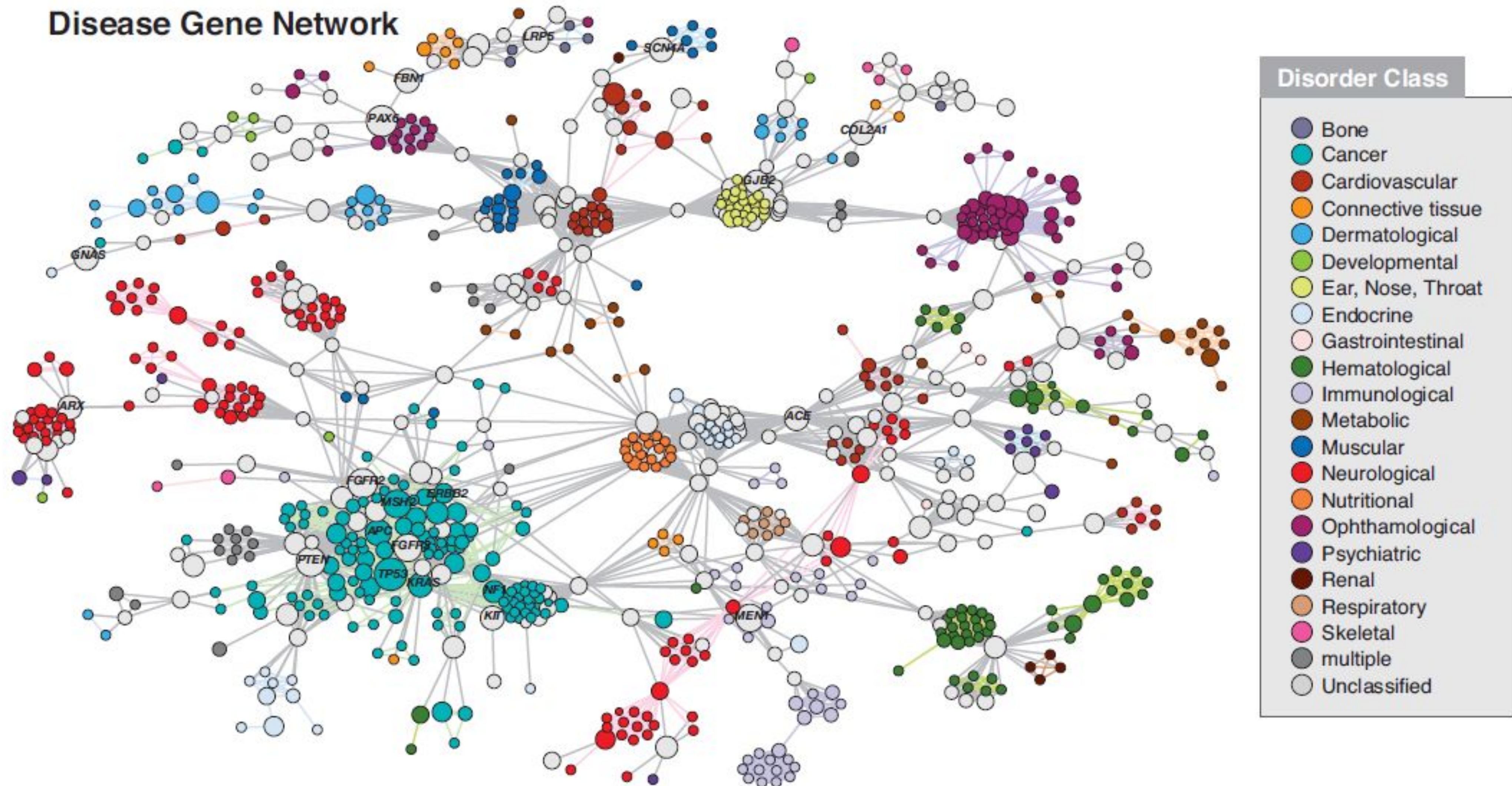
- Semplice (numero di gruppi in comune)
- Iperbolico (fattore moltiplicativo $1/(n-1)$)

Proiezione sulle malattie

a Human Disease Network



Proiezione sui geni

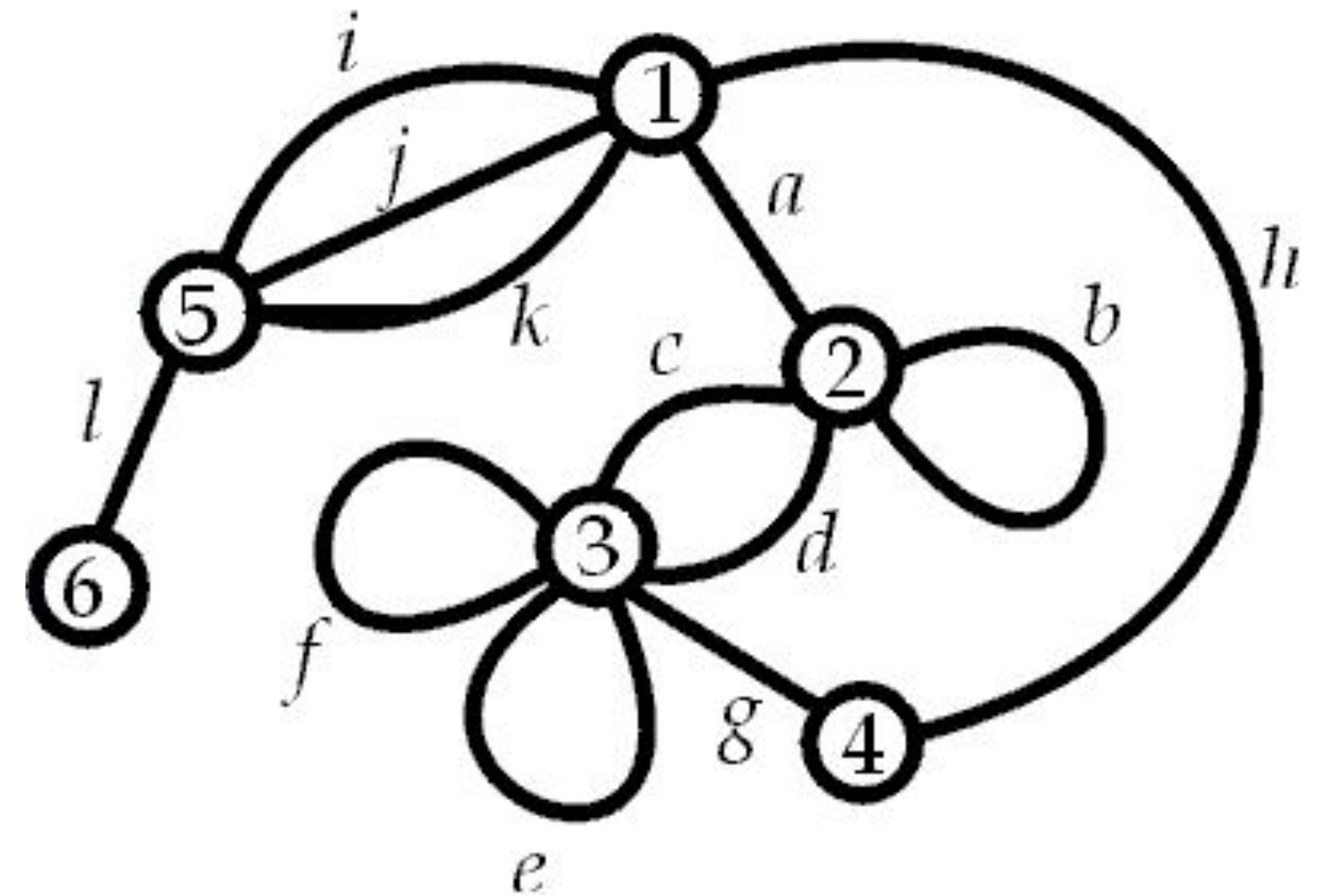


Multigrafi

Matteo Cucchi

Cos'è un Multigrafo

Con multigrafo si intende una struttura che può dirsi costituita da un insieme finito di vertici e da spigoli che collegano due vertici o un vertice con sé stesso, con la possibilità che due vertici siano collegati da più spigoli distinti.



Definizione di Multigrafo

Formalmente definiamo multigrafo una struttura relazionale della forma

$$M = \langle Q, A, \sigma \rangle$$

dove:

- Q è un insieme finito chiamato insieme dei vertici di M
- A è un insieme finito chiamato insieme delle etichette degli spigoli di M
- $\sigma : A \mapsto B_{1,2}(Q)$ qui con $B_{1,2}(Q)$ denotiamo la collezione dei sottoinsiemi di Q costituiti da uno o due vertici

Esempio di funzione σ dove:

- b è uno spigolo che collega il vertice 2 a se stesso
- c e d sono due spigoli paralleli che collegano i vertici 2 e 3

$$\sigma = \begin{array}{|c|c|c|} \hline b & c & d \\ \hline 2 & 23 & 23 \\ \hline \end{array}$$

Cosa ha di speciale un multigrafo?

$$M = \langle Q, A, \sigma \rangle$$

Quando si studiano i multigrafi per le loro proprietà generali, gli elementi di A sono oggetti semplici che servono solo per distinguere i diversi spigoli.

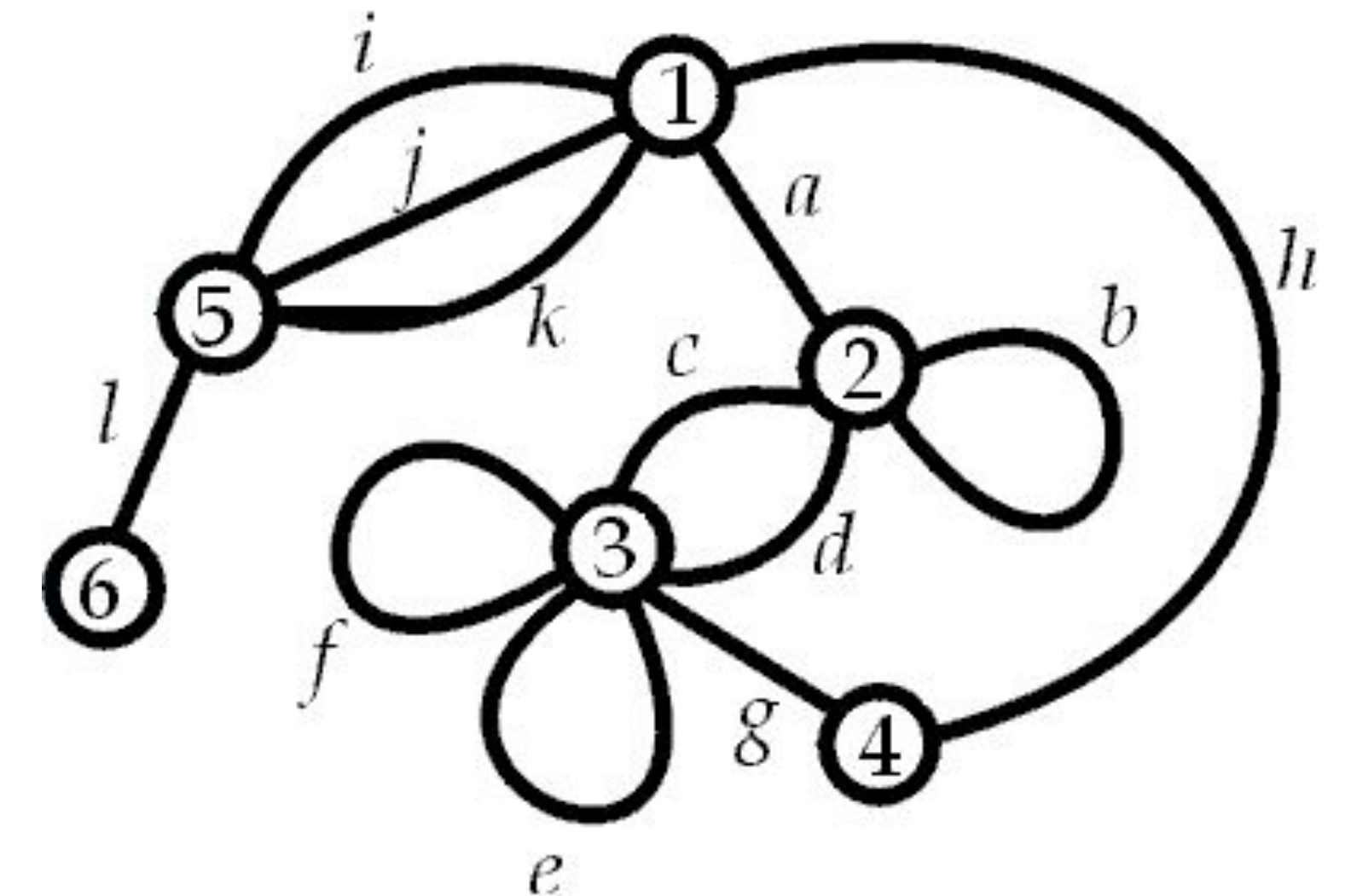
- Le coppie $\langle a, \sigma(a) \rangle$ si dicono spigoli del multigrafo;
- due spigoli $\langle a, \sigma(a) \rangle$ e $\langle b, \sigma(b) \rangle$ con $\sigma(a) = \sigma(b)$ si dicono spigoli paralleli del multigrafo;
- questa relazione di parallelismo è chiaramente una relazione di equivalenza;
- uno spigolo $\langle c, \sigma(c) \rangle$ con $|\sigma(c)| = 1$ si dice cappio del multigrafo.
- Uno spigolo privo di altri spigoli paralleli si dice spigolo semplice.

Esempio Multigrafo

Consideriamo il multigrafo $M = \langle Q, A, \sigma \rangle$ corrispondente a:

$$Q = \{1, 2, 3, 4, 5, 6\} \quad A = \{a, b, c, \dots, l\},$$

$$\sigma = \begin{array}{c|cccccccccccc} & a & b & c & d & e & f & g & h & i & j & k & l \\ \hline & 12 & 2 & 23 & 23 & 3 & 3 & 34 & 14 & 15 & 15 & 15 & 56 \end{array}$$



Nella indicazione tabellare della funzione abbiamo abbreviato $[i,j]$ con ij e $[i]$ con i .

La rappresentazione tramite insieme di nodi, insieme di archi e sigma, è l'implementazione più comune di un multigrafo.

Possiamo facilmente notare dalla raffigurazione che:

- gli spigoli etichettati da **c** e **d** sono paralleli;
- gli spigoli etichettati da **i**, **j**, **k** sono paralleli;
- gli spigoli etichettati da **e** e **f** sono cappi paralleli;

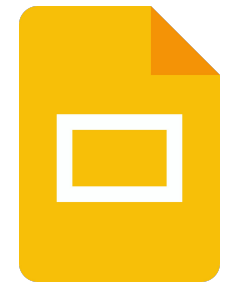
Visualizzatore di un multigrafo in Gephi



Gephi è un software open source per l'analisi e la visualizzazione delle reti sociali, scritto in Java e basato sulla piattaforma NetBeans, inizialmente sviluppato dagli studenti di 'The University of Technology of Compiègne in Francia.

Gephi è stato usato in una serie di progetti di ricerca in ambito universitario, giornalistico e in vari altri campi, per esempio per visualizzare le connessioni globali ai contenuti del New York Times o per esaminare la rete di traffico su Twitter in occasioni di disordini sociali, ma anche per altre tematiche solitamente oggetto di analisi di rete

Tutorial Gephi



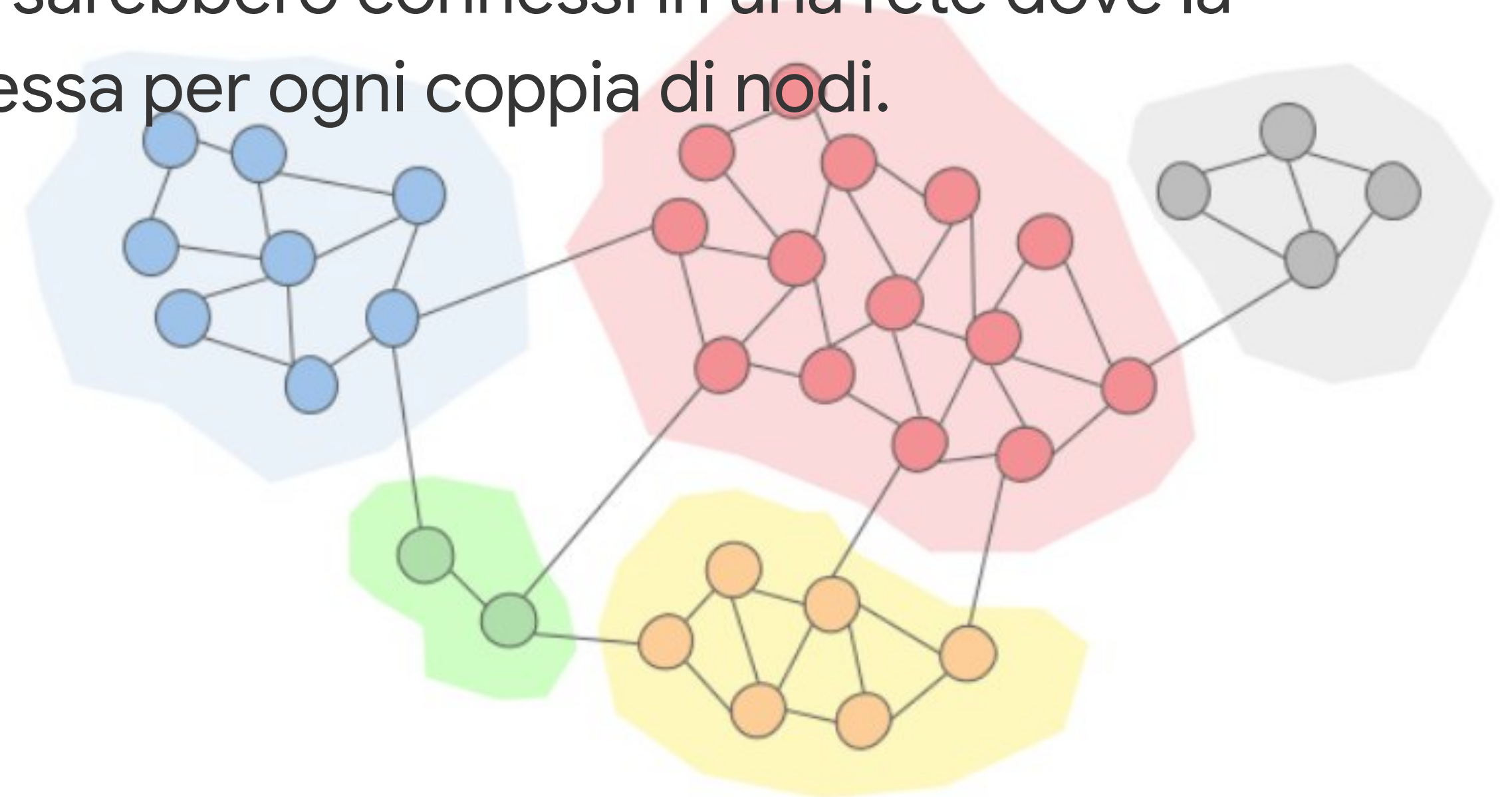
[Link alla presentazione](#)

Metodo di Louvain: Modularity

Per dividere il grafo in comunità su gephi abbiamo lanciato il comando “modularity”.

L'algoritmo di Louvain è un algoritmo che effettua Community Detection, ovvero partiziona il grafo in community, cioè sottoinsiemi di nodi che hanno “tanti legami fra loro”.

Esso si basa su massimizzare la modularity, una misura che quantifica la qualità di un'assegnazione di nodi alle community valutando quanto più densamente sono connessi i nodi all'interno di una community rispetto a quanto sarebbero connessi in una rete dove la probabilità di formazione di link tra due nodi è la stessa per ogni coppia di nodi.



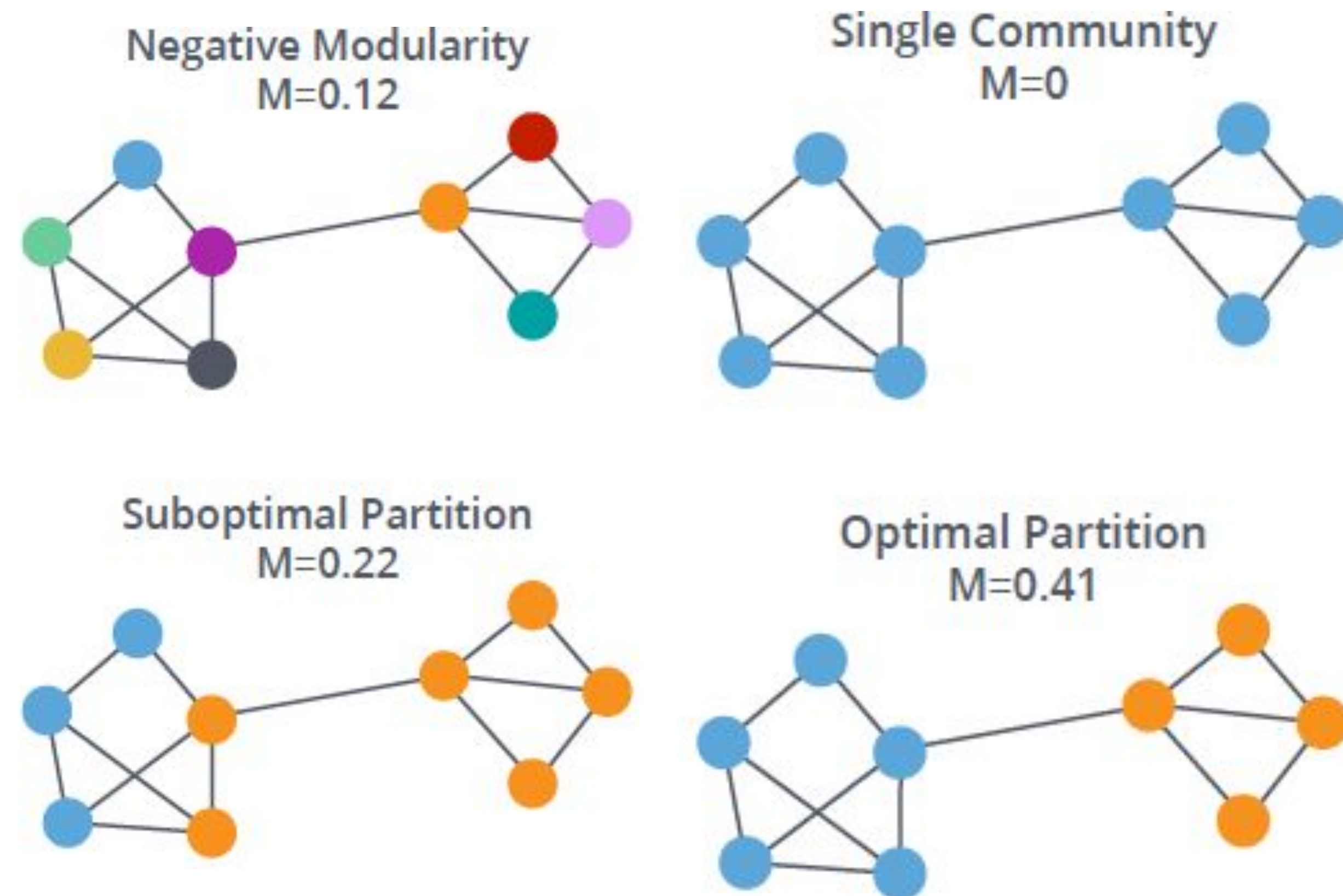
Metodo di Louvain: Modularity

La modularity è un indice dell'intera rete e assume un valore tra -0.5 e 1.

Se la modularity = 0 allora siamo nel caso in cui il grafo ha 1 sola community

Se la modularity assume un valore vicino a 1 possiamo dire che il grafo è stato diviso bene in community, i nodi “simili” sono stati messi nella stessa community

Se la modularity assume un valore negativo possiamo dire che il grafo è diviso in community che però non hanno nodi simili, infatti nodi “simili” sono posizionati in community diverse



Metodo di Louvain

Qui la descrizione dei passi che caratterizzano l'algoritmo di Louvain.

All'inizio tutti i nodi appartengono a una community diversa.

La condizione di stop non necessariamente è quella di massimizzazione della modularità.
(e.g. posso fermarmi dopo un certo numero di iterazioni)



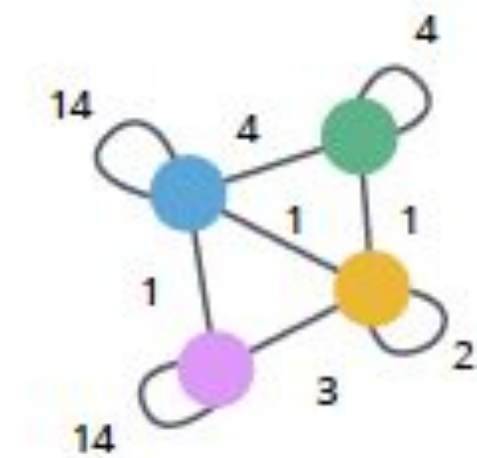
Step 0

Choose a start node and calculate the change in modularity that would occur if that node joins and forms a community with each of its immediate neighbors.



Step 1

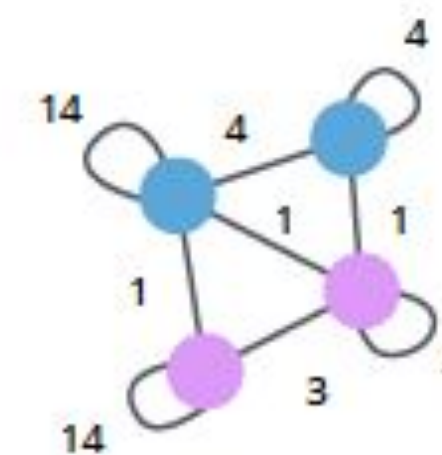
The start node joins the node with the highest modularity change. The process is repeated for each node with the above communities formed.



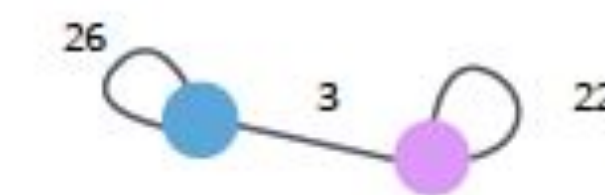
Step 2

Communities are aggregated to create super communities and the relationships between these super nodes are weighted as a sum of previous links. (Self-loops represent the previous relationships now hidden in the super node.)

Pass 2



Step 1



Step 2

Steps 1 and 2 repeat in passes until there is no further increase in modularity or a set number of iterations have occurred.

Bonus: applicazioni del metodo di Louvain

- Il metodo Louvain è stato proposto per fornire raccomandazioni agli utenti di Reddit per trovare subreddit simili in base al comportamento generale degli utenti. Per maggiori dettagli, vedi [Subreddit Recommendations within Reddit Communities](#).
- Il metodo Louvain è stato utilizzato per estrarre argomenti da piattaforme social online, come Twitter e YouTube, sulla base del grafico di ricorrenza dei termini nei documenti come parte del processo di modellazione degli argomenti. Questo processo è descritto in [Topic Modeling based on Louvain method in Online Social Networks](#).
- Il metodo Louvain è stato utilizzato per indagare sul cervello umano e trovare strutture di comunità gerarchiche all'interno della rete funzionale del cervello. Lo studio citato è [Hierarchical Modularity in Human Brain Functional Networks](#).

Multigrafi Orientati Eterogenei applicati ai Database

Margherita Pindaro

PORTAL 7.3 ENTITY RELATIONSHIP DIAGRAM

INDEX

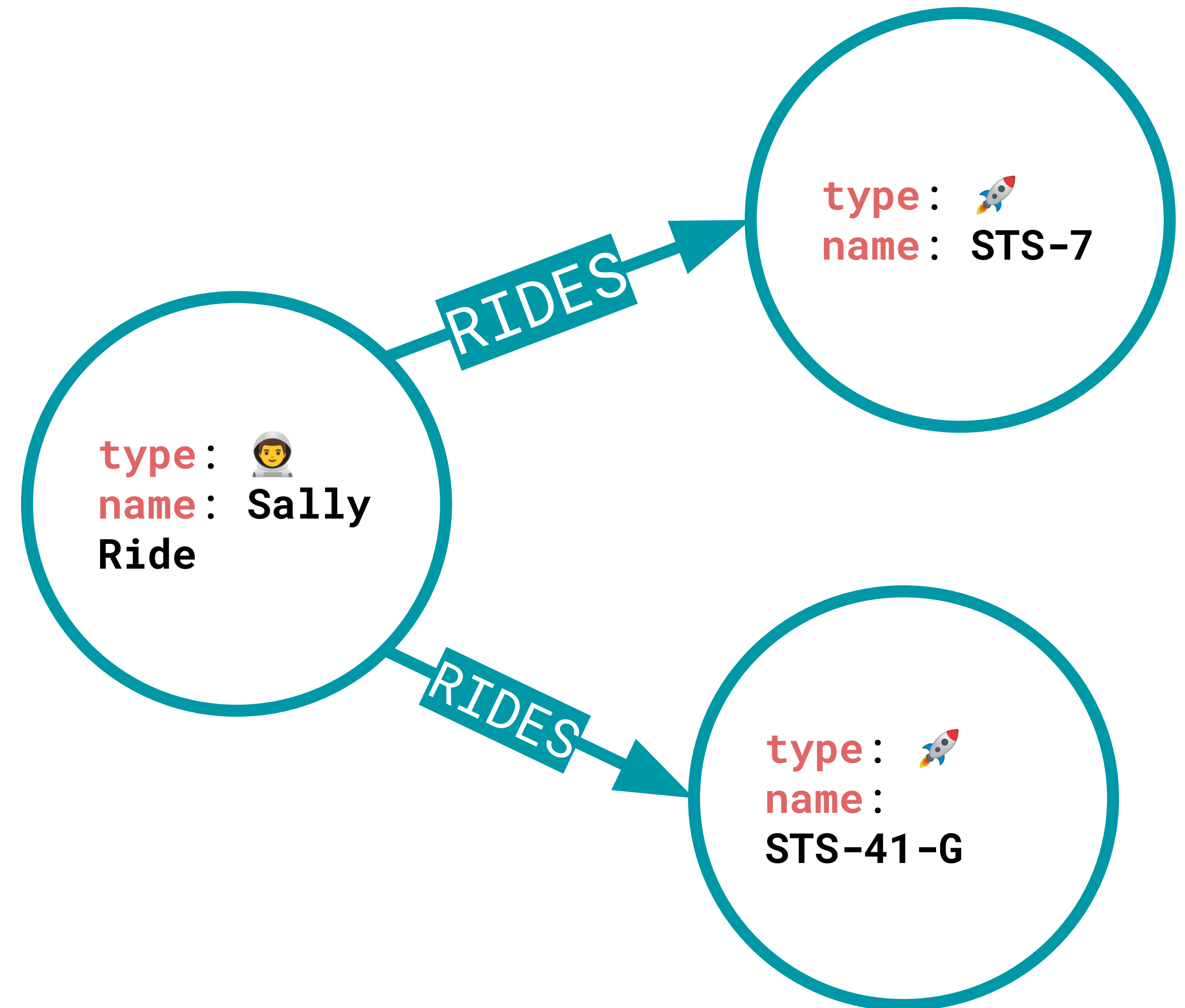


I Graph DB

Definibili come un **multigrafo orientato** (*multidigraph*) **eterogeneo** in cui:

- I nodi rappresentano istanze delle entità
- Gli archi le relazioni tra di esse
- Gli attributi sono metadati di archi e nodi basati su *chiave - valore*.

👉 Un approccio ai DB più intuitivo e **scalabile**. Infatti sono maggiormente usati per database in **larga scala**, con molte entità differenti.



Alcuni esempi di DBMS

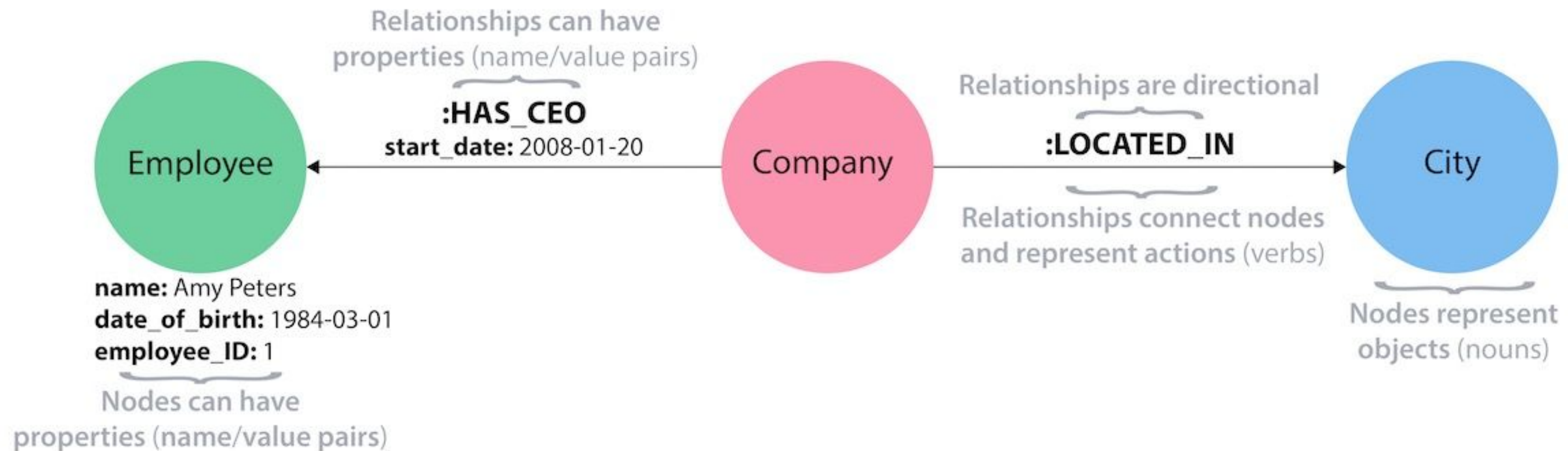


II DBMS Neo4j

Neo4j è un software per basi di dati a grafo open source sviluppato interamente in Java.

I suoi DB sono *schemeless*, ovvero non necessitano che venga specificato uno schema come per il relazionale (*diagramma ER*). E' NoSQL ma differisce dalla maggior parte delle tecnologie di quella categoria nel fatto che è **ACID-compliant**.

Linguaggio di querying utilizzato: **Cypher**



E quindi?

Prendiamo in esempio una query:

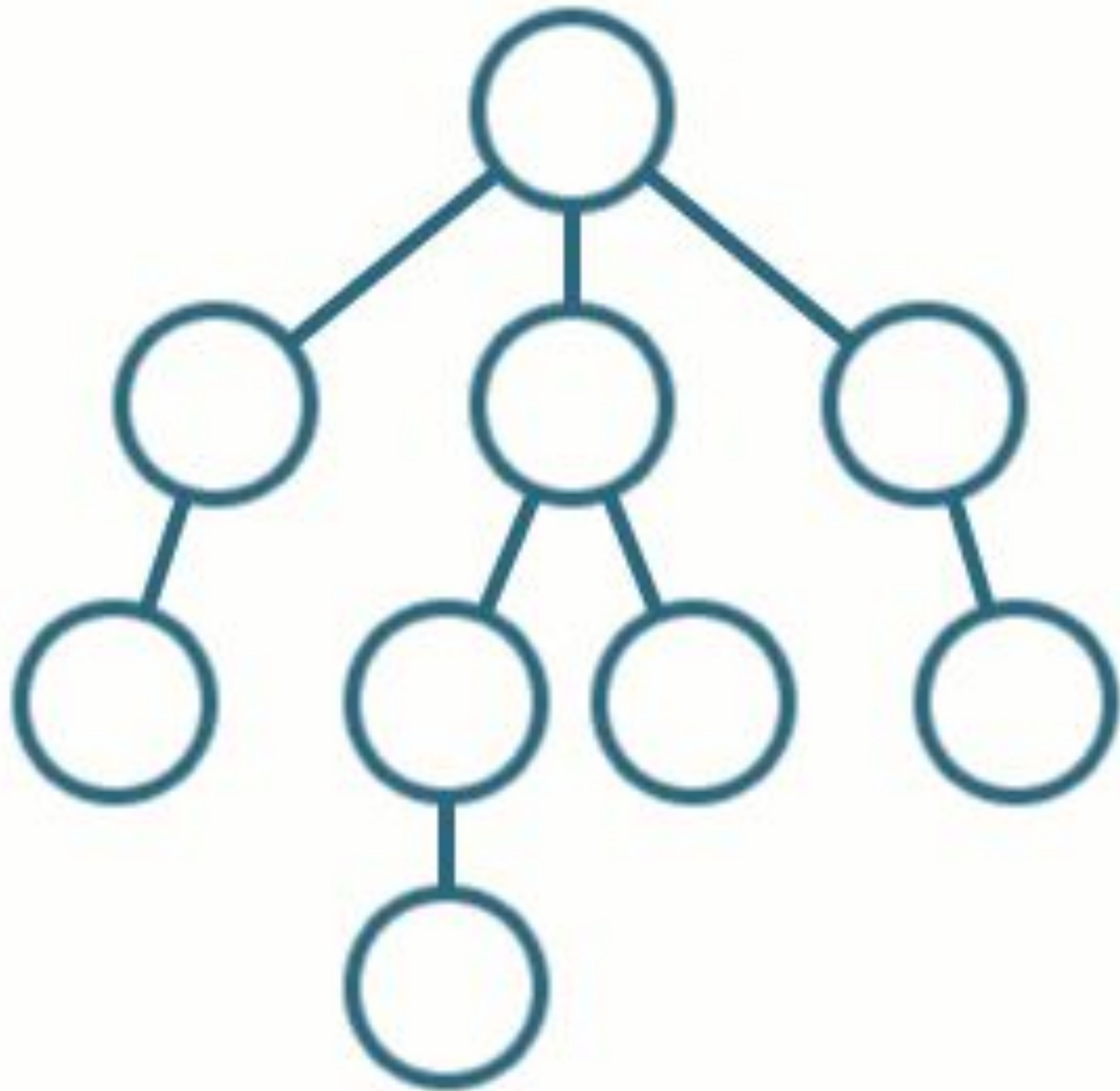
Restituire i potenziali amici di un utente fino al terzo grado di separazione

```
select count(distinct uf3.*)  
from t_user_friend uf1  
inner join t_user_friend uf2  
on uf1.user_1 = uf2.user_2  
inner join t_user_friend uf3  
on uf2.user_1 = uf3.user_2  
where uf1.user_1 = ?
```

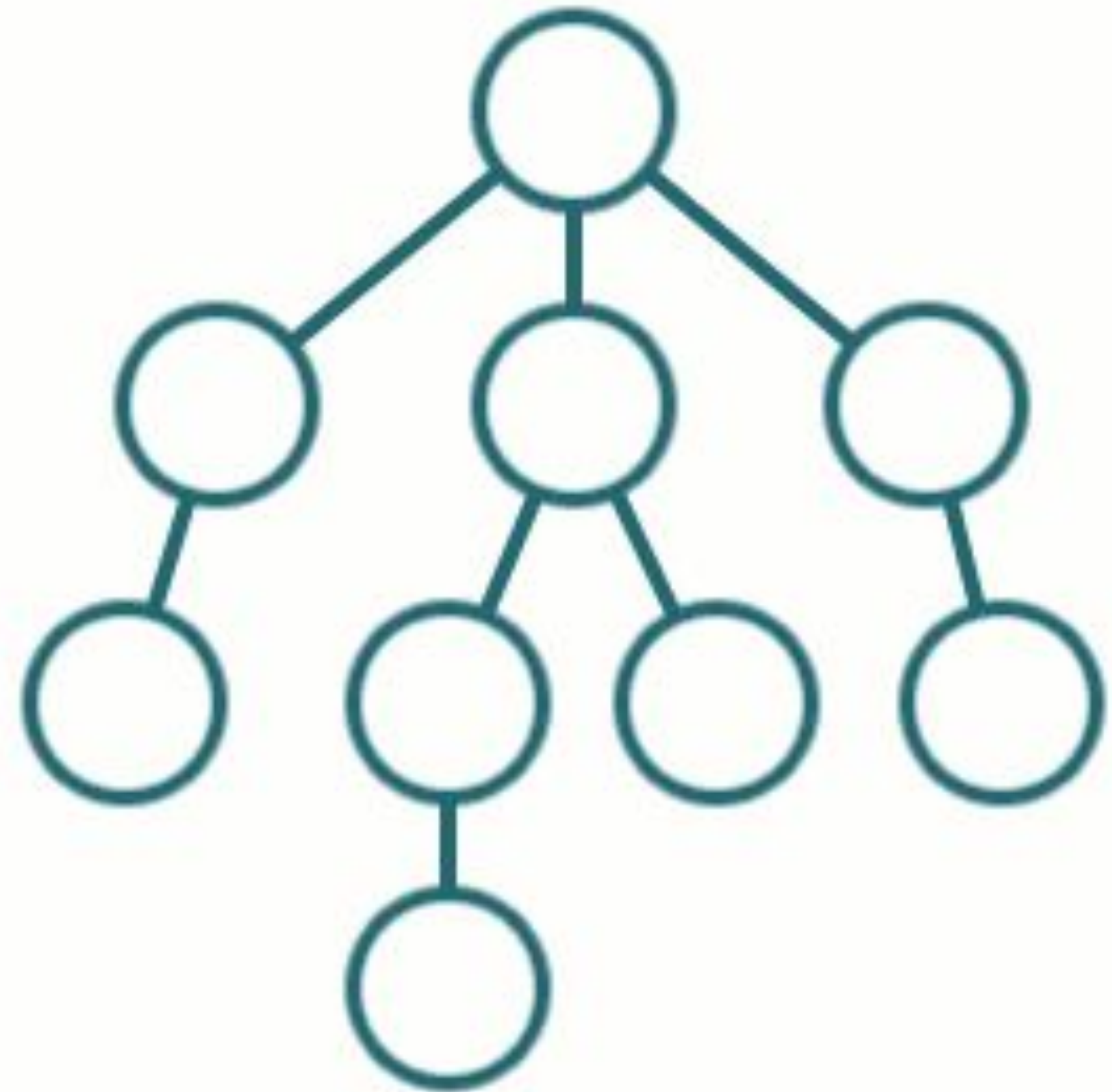
```
TraversalDescription t =  
Traversal.description()  
    .relationships("IS_FRIEND_OF",  
        Direction.OUTGOING)  
    .evaluator(Evaluators.atDepth(3))  
    .uniqueness(Uniqueness.NODE_GLOBAL);  
Iterable<nodes> =  
traversalDescription.traverse(nodeById).nodes()
```

Graph traversal

DFS



BFS



Tirando le somme

✓ Vantaggi

- Pensati per dati in larga scala
- Flessibili
- ACID-compliant
- Query significativamente più performanti quando coinvolgono delle stringhe
- Organizzati nel modo naturale di rappresentare le informazioni

✗ Svantaggi

- Ancora poco utilizzati
- Manca un linguaggio di querying unificato
- Non supportano ambienti multiutente
- Mancanza di meccanismi di sicurezza
- E un'ultima cosa che tra poco vediamo...

Hands-on



SECTIONS ▼

NIGHTLY NEWS

MSNBC

MEET THE PRESS

DATELINE

 TODAY



TECH > SOCIAL MEDIA

GADGETS INTERNET SECURITY INNOVATION MOBILE

TECH DEC 20 2017, 11:11 AM ET

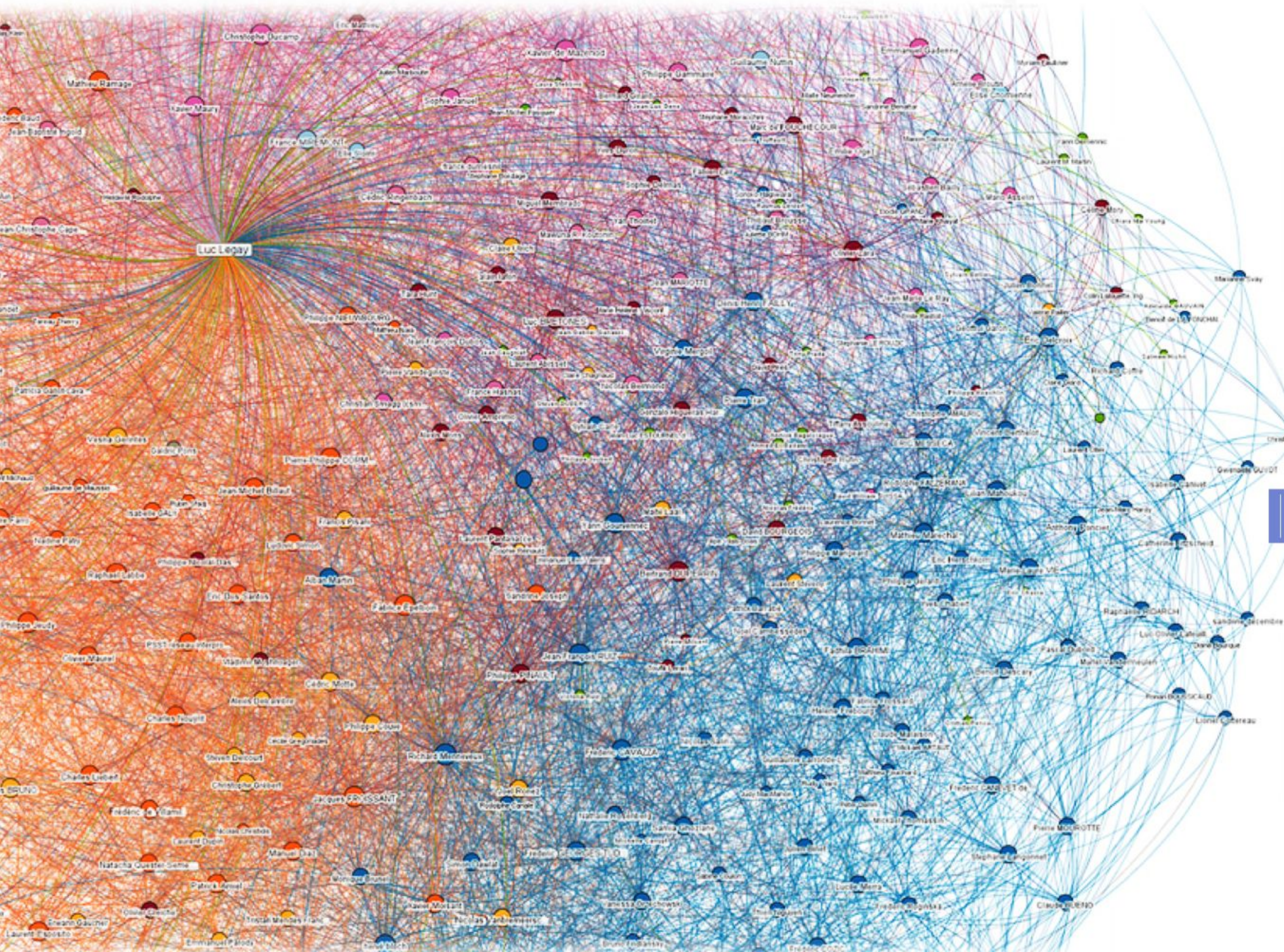
Russian trolls went on attack during key election moments

by BEN POPKEN



[Link alla demo](#)

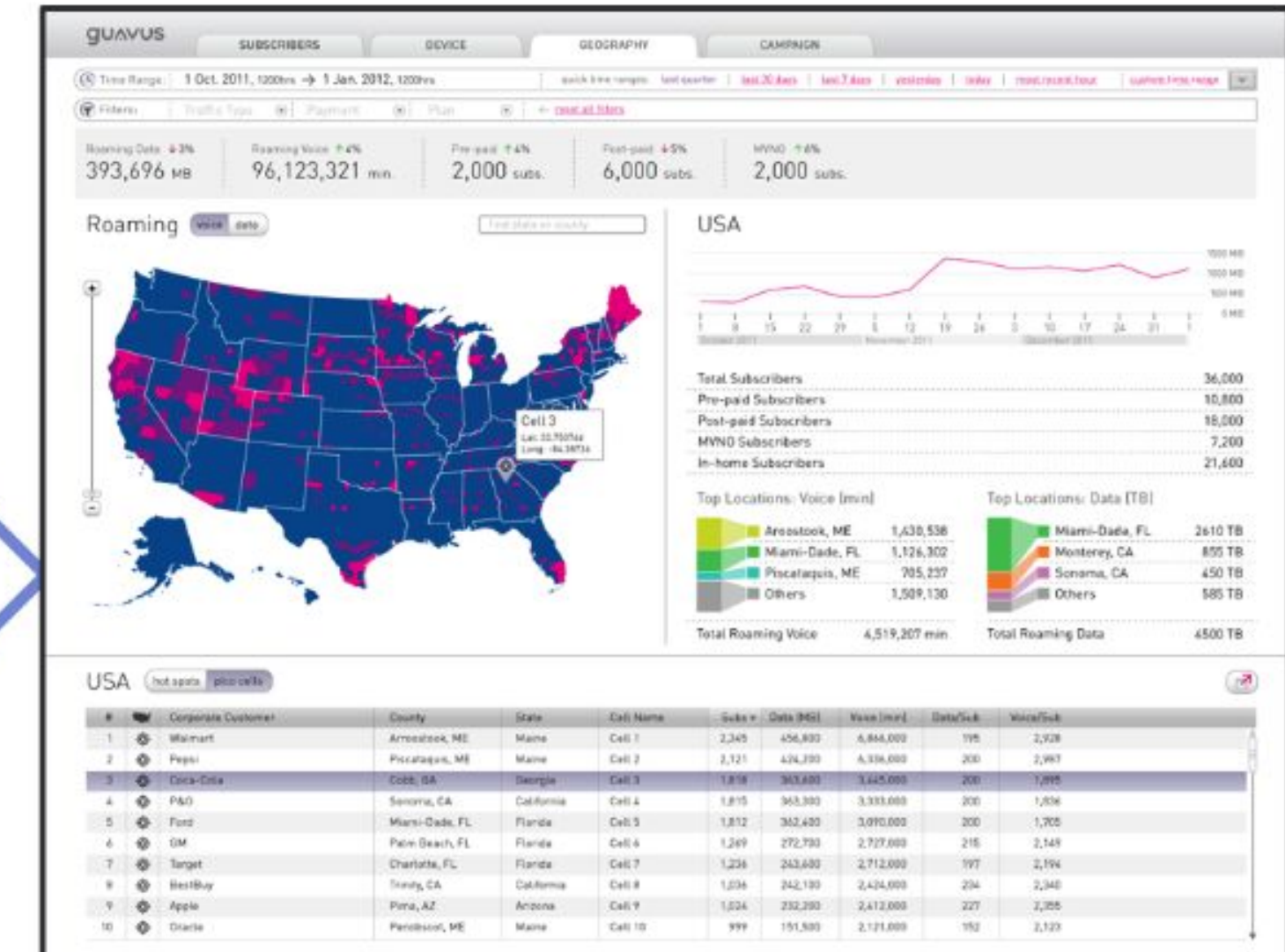
E per quanto riguarda la visualizzazione?



 Data-Driven Documents

 KeyLines

 Cytoscape



 sigma.js

 vis.js

 LINKURIOUS

Astrazione

Il bisogno di modi per semplificare e alleggerire la visualizzazione di questi grafi ha portato alla creazione di tecniche dette **astrazioni** per multidigraph e grafi eterogenei.

Banalizzato, riassumere più elementi in un solo.

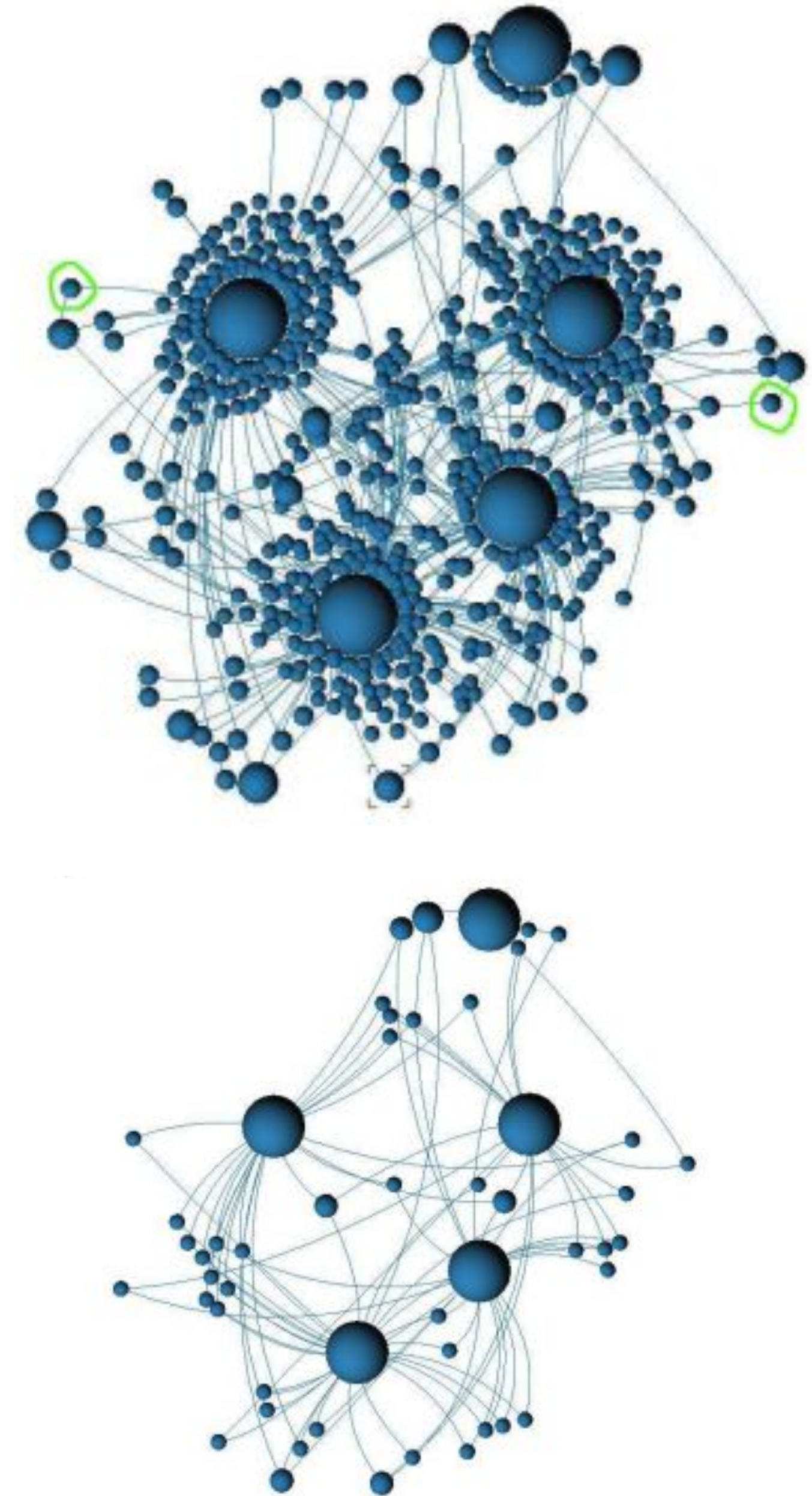
💡 L'operazione di astrazione **dovrebbe essere anche reversibile**: nel momento in cui si esprime una interrogazione sul grafo astratto dove è stata effettuata un'astrazione, il sistema deve saperla tradurre sul grafo con un livello di dettaglio maggiore dove le entità sono ancora distinte.

Astrazione strutturale

L'astrazione strutturale si basa sull'utilizzo di informazioni topologiche come il grado di nodo, la connettività, etc.

In questa operazione non si fa mai riferimento al significato che è di interesse invece per l'astrazione semantica.

I sottografi vengono collassati in base al significato o ad etichette specifiche.



Astrazione semantica

Introduciamo il concetto di **ontologia**: una rappresentazione formale, condivisa ed esplicita di una concettualizzazione di un dominio di interesse.

La presenza di ontologie permette di effettuare le operazioni di astrazione e di filtraggio basandosi sui due grafi (ontologia e grafo in sé) in modo tale da ridurre la complessità di una rete, evidenziare relazioni o individuarne nuove in modo tale da creare nuova conoscenza nascosta

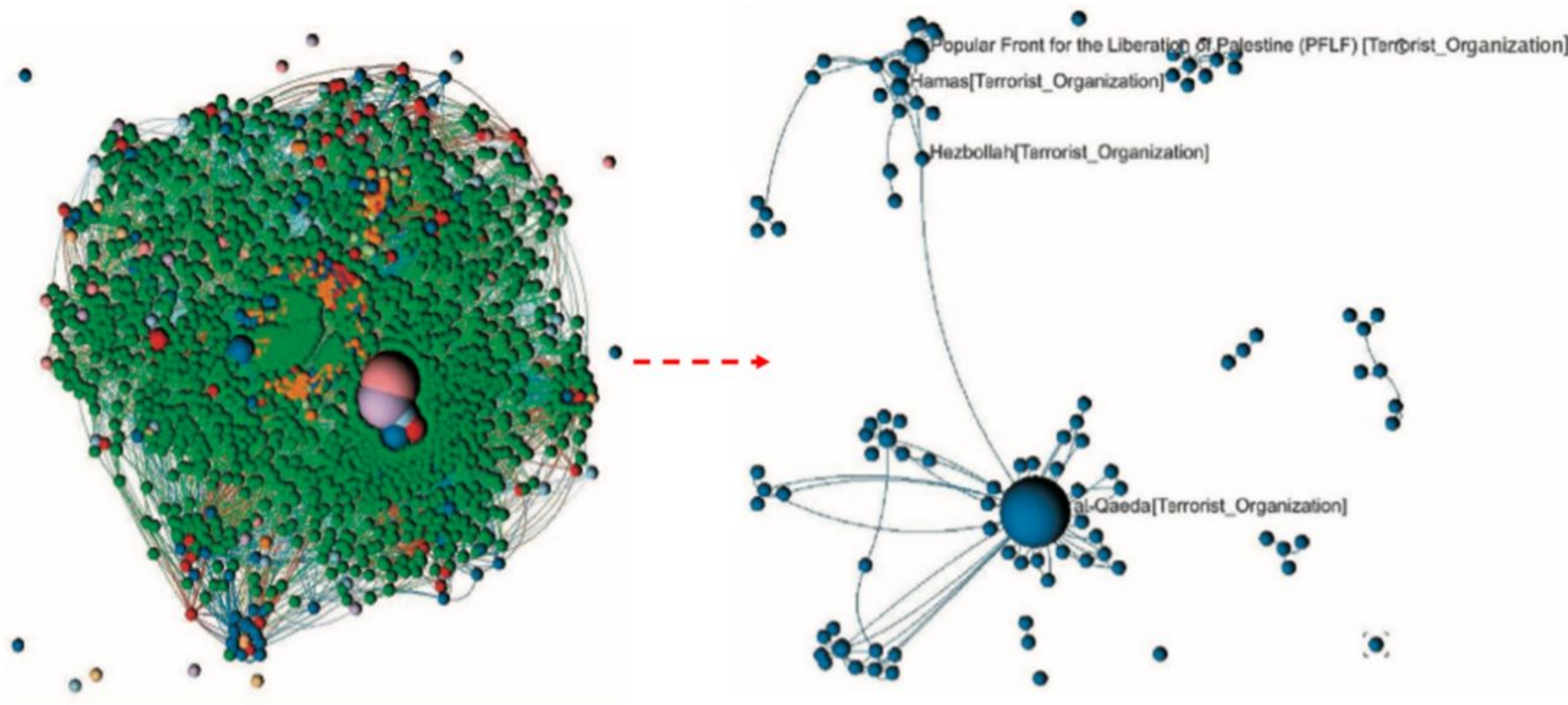


(a) A semantic abstraction that only contains terrorist organizations. Nodes with types of country/area and classification become attributes of their neighbors in the original graph.



Astrazione basata su clustering

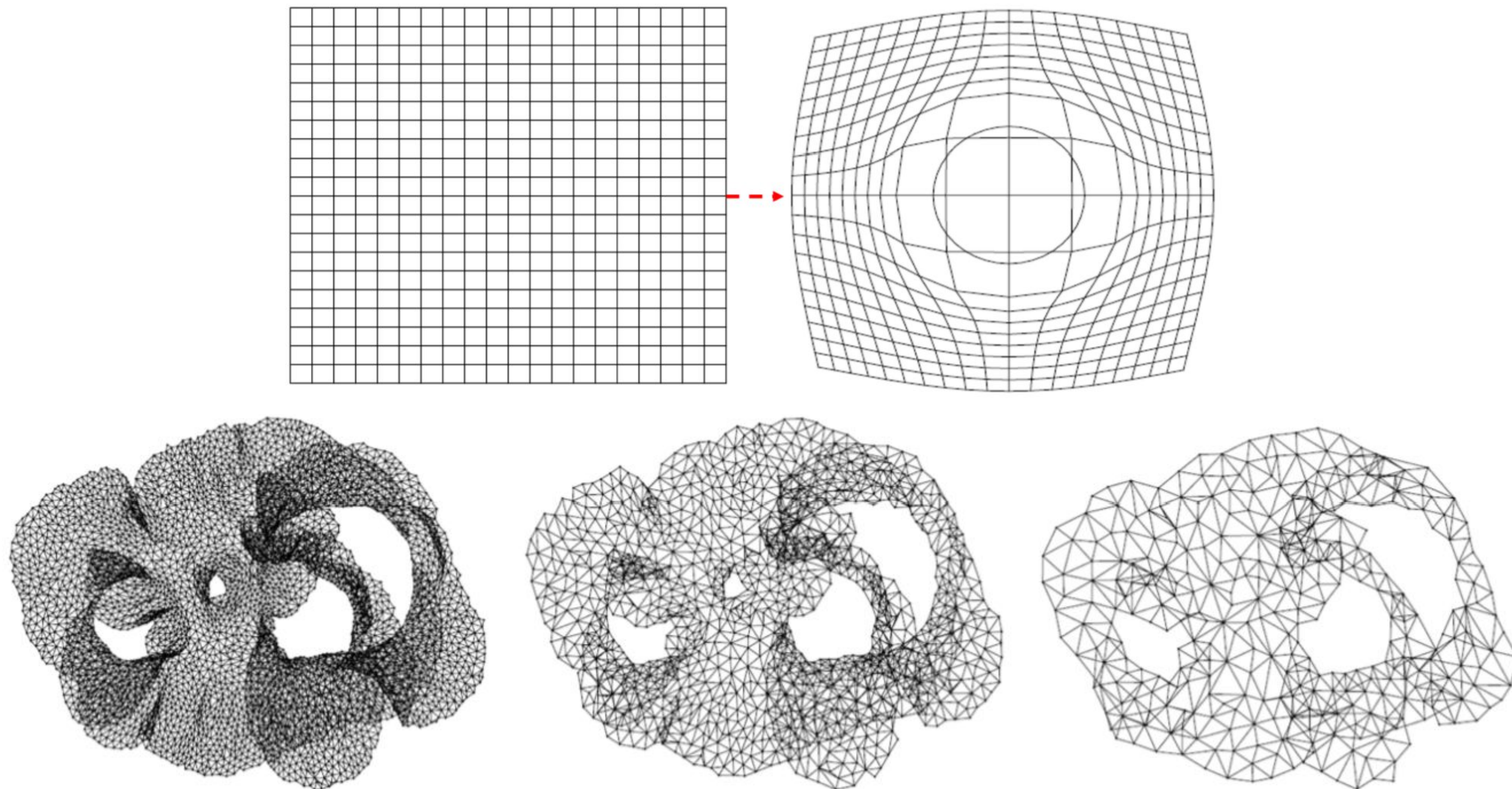
Effettuare una operazione di clustering significa raggruppare i nodi del grafo semantico secondo ad una determinata partizione. Il nodo risultante da questa unione ha la necessità di mantenere le etichette dei nodi che ha inglobato.



Diversi tipi:

- Semantico
- Gerarchico
- Coesivo
- Strutturalmente connesso
- Per categoria
- Per comunicazione
- Basato sulla frequenza

Astrazioni fish-eye



Graph Sketches

💡 Lo schermo e le risorse del pc sono limitate, sono una limitata quantità di dati può essere rappresentata alla volta su un display.

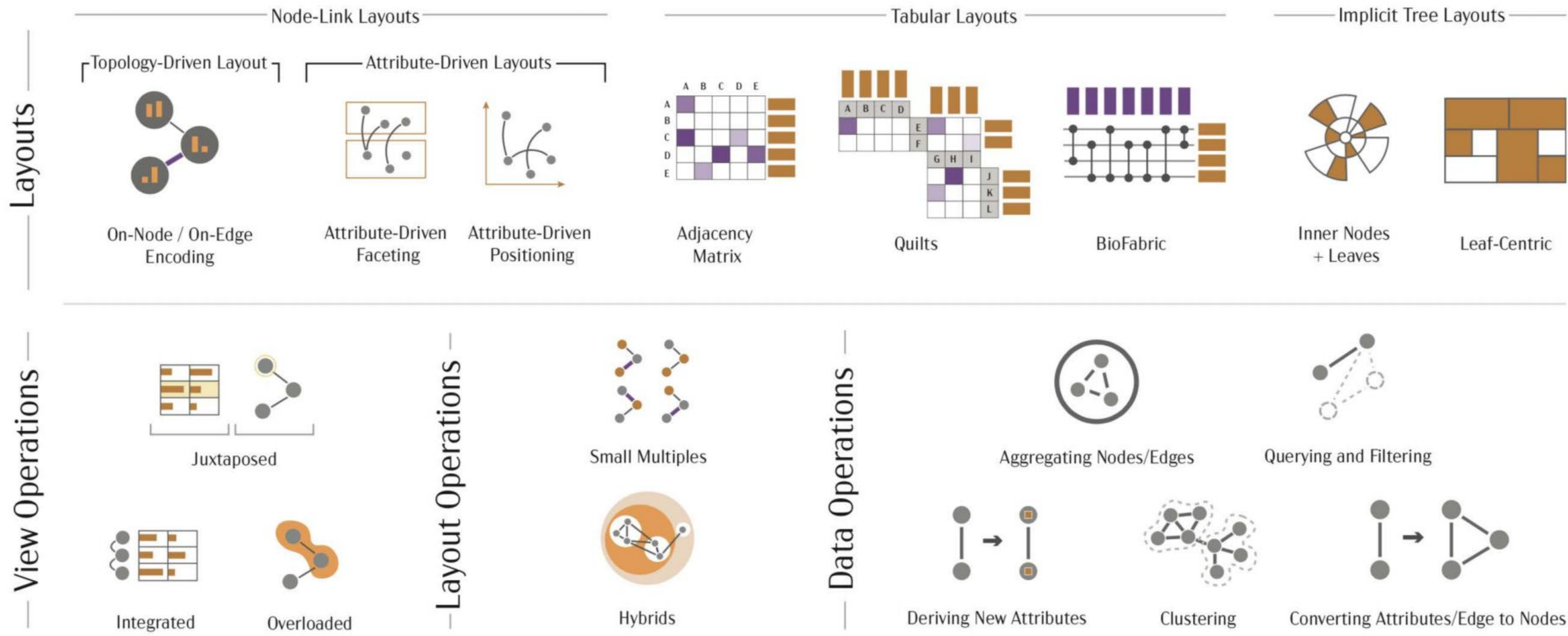
Uno sketch per un grafo G come un multi-digraph definito su una partizione V_0, V_1, \dots, V_k di $V(G)$ attualmente visualizzabili (il calcolo viene fatto sui pixel).

Un multi-edge da V_i a V_j , rappresenta il set di edge in G compresi tra V_i e V_j e la molteplicità è il numero di quegli edge.

Dato un algoritmo che computa uno sketch per il grafo G , può essere usato ricorsivamente per generare un albero T tale che le foglie di T rappresentino un raffinamento della partizione originale e questo albero rappresenta una partizione gerarchica di $E(G)$. Ciò significa che una visualizzazione dettagliata si può raggiungere zoomando nell'albero



Riassumendo





Ma non finisce qui!

- Reti tripartite
- Reti temporali
- Ipergrafi

References

Diffusione dell'informazione

- [Wang, Pu & Gonzalez, Marta C. & Hidalgo, Cesar & Barabasi, Albert-Laszlo. \(2009\). Understanding the Spreading Patterns of Mobile Phone Viruses. Science \(New York, N.Y.\). 324. 1071-6. 10.1126/science.1167053.](#)
- [Schulz, Christian. \(2018\). Visualizing spreading phenomena on complex networks.](#)
- [Barabási, A.-L., Pósfai, M. \(2016\). Network science. Cambridge: Cambridge University Press. ISBN: 9781107076266 1107076269](#)
- [G. Rossetti, L. Milli, S. Rinzivillo, A. Sirbu, D. Pedreschi, F. Giannotti. “NDlib: Studying Network Diffusion Dynamics” IEEE International Conference on Data Science and Advanced Analytics, DSAA. 2017.](#)

References

Reti Bipartite

- [Goh, K & Cusick, Michael & Valle, David & Childs, Barton & Vidal, Marc & Barabasi, Albert-Laszlo. \(2007\). The human disease network. Proceedings of the National Academy of Sciences of the United States of America. 104. 8685-90. 10.1073/pnas.0701361104.](#)
- [Krzywinski, Martin & Birol, Inanc & Jones, Steven & Marra, Marco. \(2011\). Hive plots—Rational approach to visualizing networks. Briefings in bioinformatics. 13. 627-44. 10.1093/bib/bbr069.](#)

References

Multigrafi

- [Škrlj, Blaž & Kralj, Jan & Lavrac, Nada. \(2019\). Py3plex toolkit for visualization and analysis of multilayer networks. Applied Network Science. 4. 10.1007/s41109-019-0203-7.](#)
- [Bastian, Mathieu & Heymann, Sebastien & Jacomy, Mathieu. \(2009\). Gephi: An Open Source Software for Exploring and Manipulating Networks. 10.13140/2.1.1341.1520.](#)

References

Multigrafi orientati eterogenei applicati ai Database

- [Shen, Zeqian & Ma, Kwan-Liu & Eliassi-Rad, Tina. \(2006\). Visual Analysis of Large Heterogeneous Social Networks by Semantic and Structural Abstraction. IEEE transactions on visualization and computer graphics. 12. 1427-39. 10.1109/TVCG.2006.107.](#)
- [Vicknair, Chad & Macias, Michael & Zhao, Zhendong & Nan, Xiaofei & Chen, Yixin & Wilkins, Dawn. \(2010\). A comparison of a graph database and a relational database: A data provenance perspective. The 48th ACM Southeast Conference. 10. 42. 10.1145/1900008.1900067.](#)
- [Vukotic, Aleksa & Watt, Nicki & Abedrabbo, Tareq & Fox, Dominic & Partner, Jonas. \(2014\). Neo4j in Action.](#)
- [Abello, James & Finocchi, Irene & Korn, Jeffrey. \(2003\). Graph Sketches. 10.1109/INFVIS.2001.963282.](#)
- [Abello, James & Korn, J.. \(2002\). MGV: A system for visualizing massive multidigraphs. Visualization and Computer Graphics, IEEE Transactions on. 8. 21-38. 10.1109/2945.981849.](#)
- [Russian trolls went on attack during key election moments](#)