

Filippo Geraci



# **INTRODUZIONE ALLA SOCIAL NETWORK ANALYSIS**



# Le reti sociali (Social networks)

- Una rete sociale e' costituita da un insieme di individui che intrattengono tra loro relazioni di varia natura
- Le reti sociali possono essere modellate come grafi nei quali ad ogni nodo e' associato un individuo e ad ogni arco viene associata una relazione
- Si parla di reti sociali anche nei contesti in cui al posto degli individui vi sono elementi di altra natura

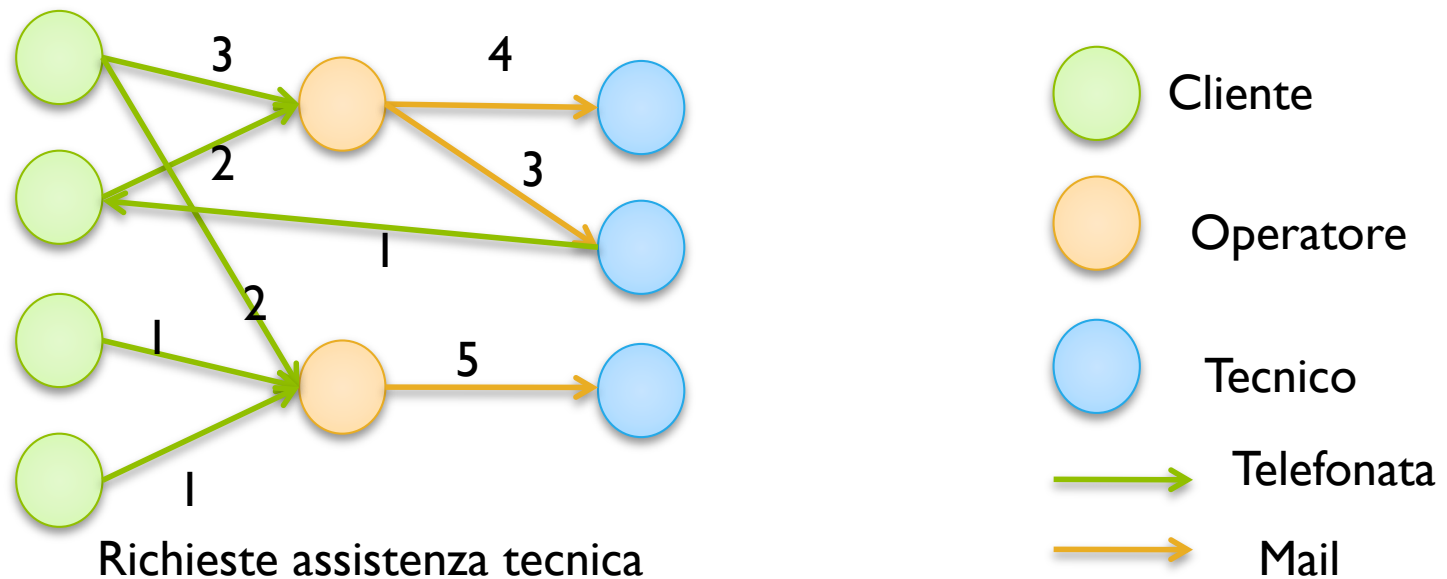


# Caratteristiche delle reti sociali

- Ad una rete sociale possono partecipare individui di tipo diverso
- Anche le relazioni possono essere di vari tipi
- Una relazione può anche avere un peso che indica in qualche modo l'intensità della relazione
- Le relazioni possono essere sia biunivoche sia univoche

# Reti sociali e grafi

- Una relazione univoca viene mappata in un arco diretto
- L'intensità della relazione viene espressa associando un peso all'arco
- La distinzione tra individui viene fatta associando un tipo ai nodi

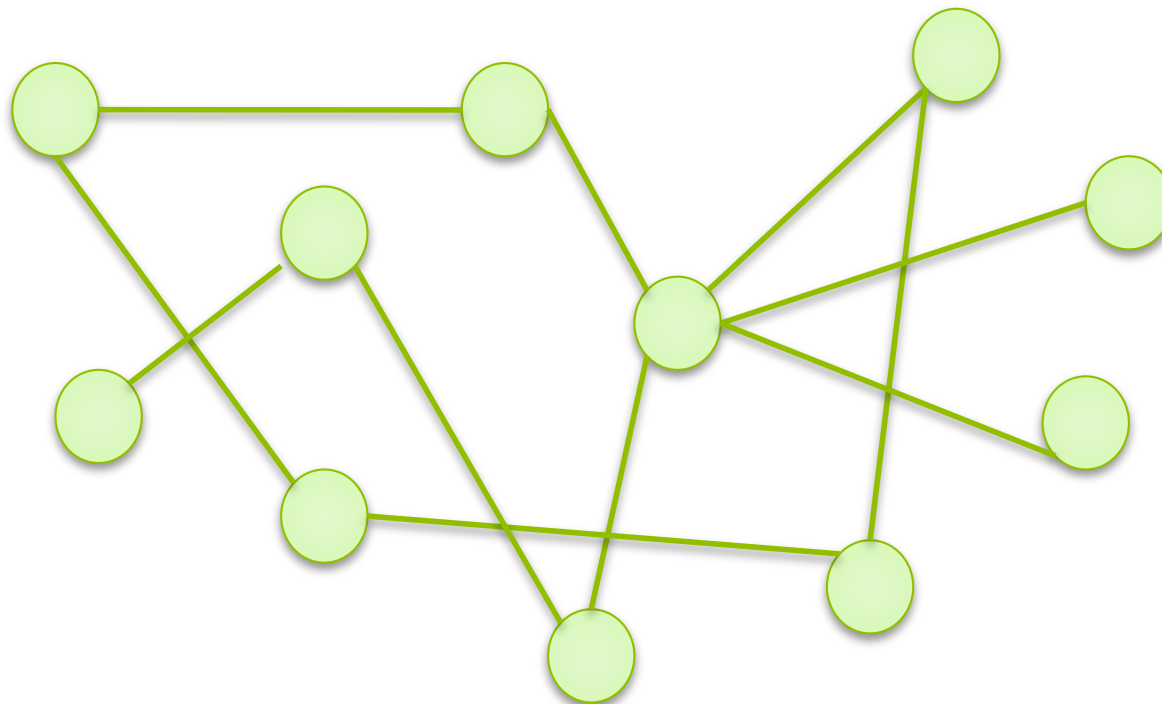




# Elementi di teoria dei grafi

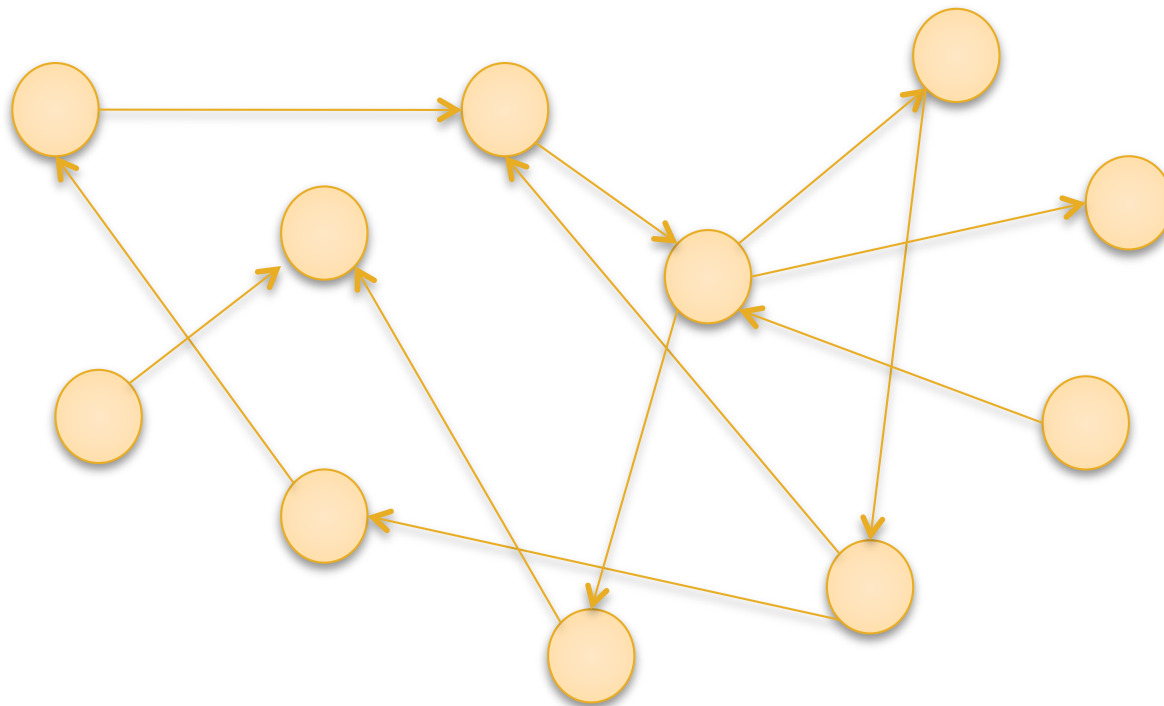
# Definizione

- Un grafo e' costituito da un certo numero di elementi detti nodi collegati tra loro da archi



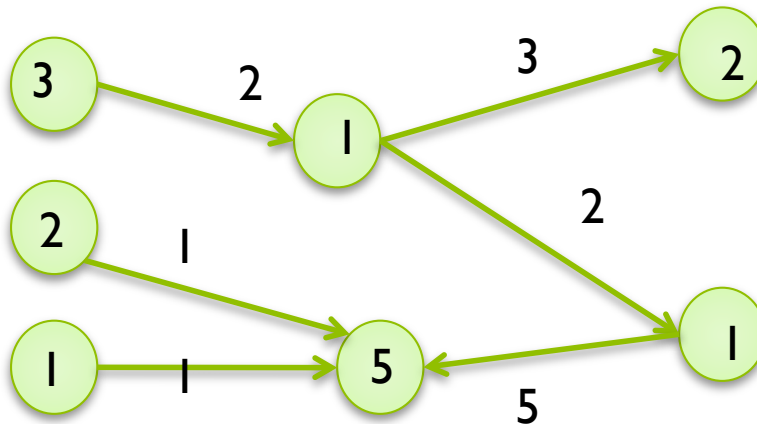
# Grafi diretti ed indiretti

- In un grafo diretto ad ogni arco viene associata una direzione



# Grafi pesati

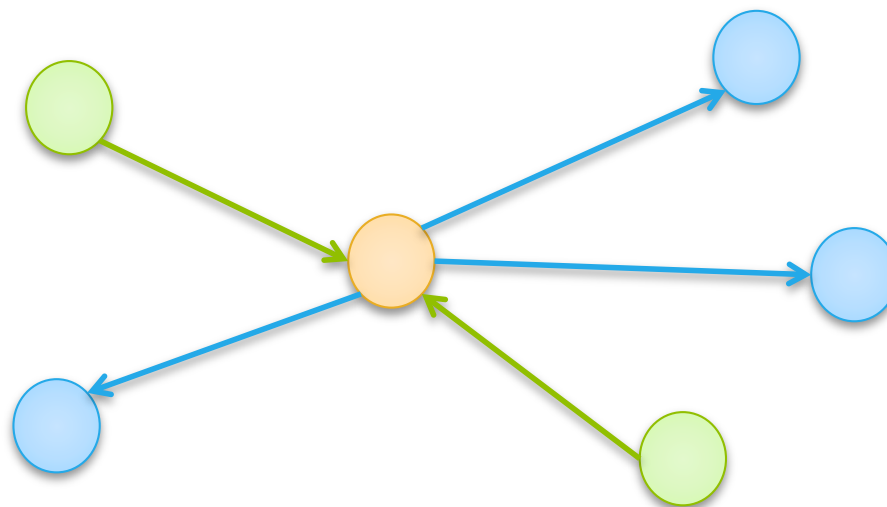
- Sui nodi
  - Ad ogni nodo viene associato un peso
  - Indica l'importanza del nodo
  - **Esempio:** ranking
- Sugli archi
  - Ad ogni arco viene associato un peso
  - Indica l'intensità della relazione tra due nodi





# Grado dei nodi

- Il grado (degree) di un nodo e' dato dal numero di archi in esso incidenti
- Se il grafo e diretto si distingue il grado degli archi entranti (indegree) ed il grado degli archi uscenti (outdegree)
  - Vale che  $\text{degree} = \text{indegree} + \text{outdegree}$



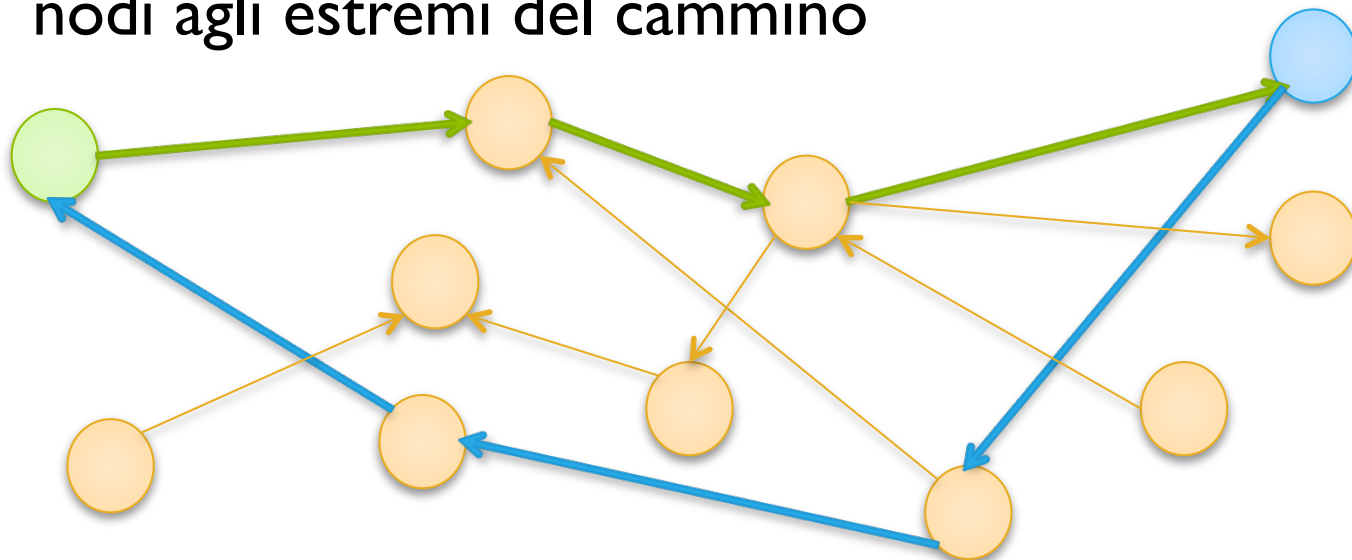
Grado 5

Outdegree 3

Indegree 2

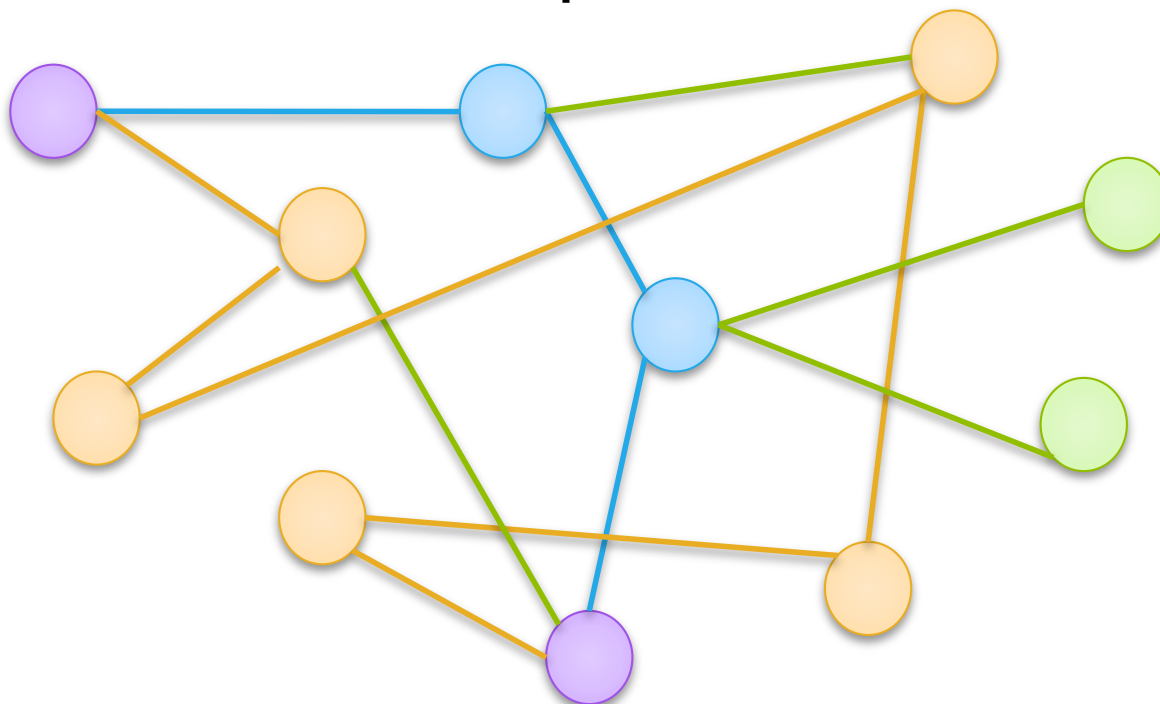
# Cammini

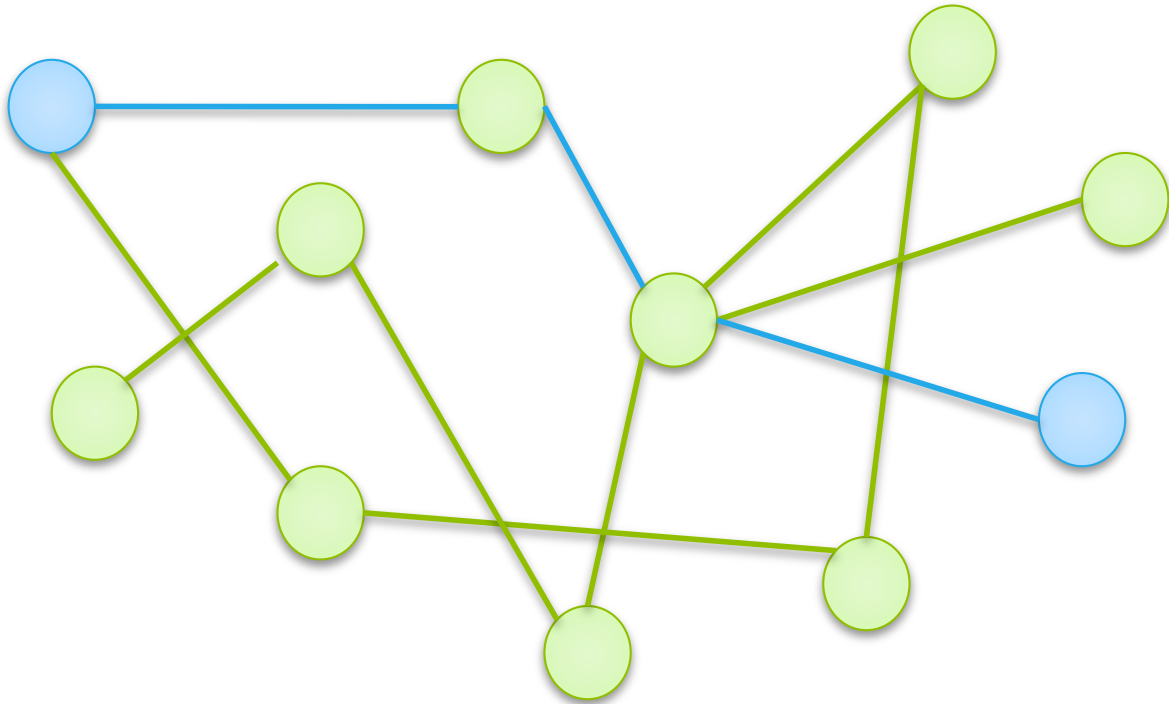
- Un cammino e' un percorso che va da un nodo sorgente ad una destinazione
  - In un grafo diretto il percorso deve seguire il verso degli archi
  - In un grafo diretto il percorso da A a B e' diverso da quello tra B ed A.
  - Uno dei due potrebbe non esistere
- La lunghezza del cammino e' il numero di archi tra i nodi agli estremi del cammino



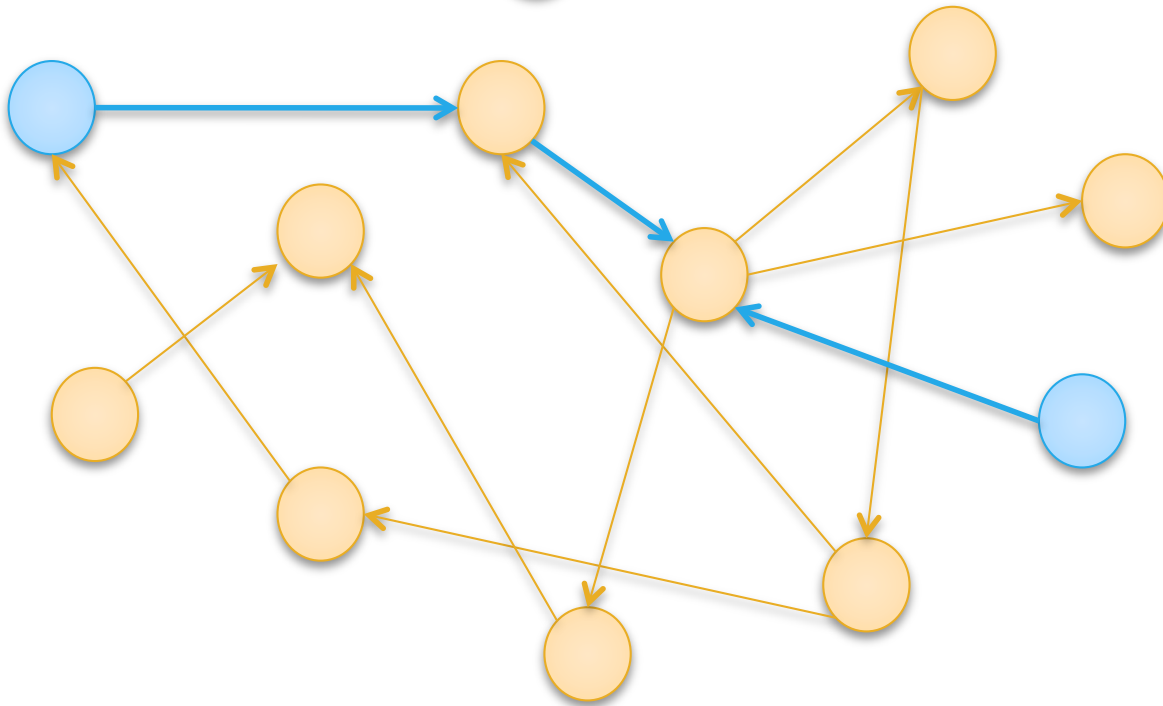
# Cammini minimi

- Un cammino tra due nodi e' minimo quando nessun altro cammino tra i nodi e' composto da numero di archi inferiore
  - Possono esistere più cammini minimi





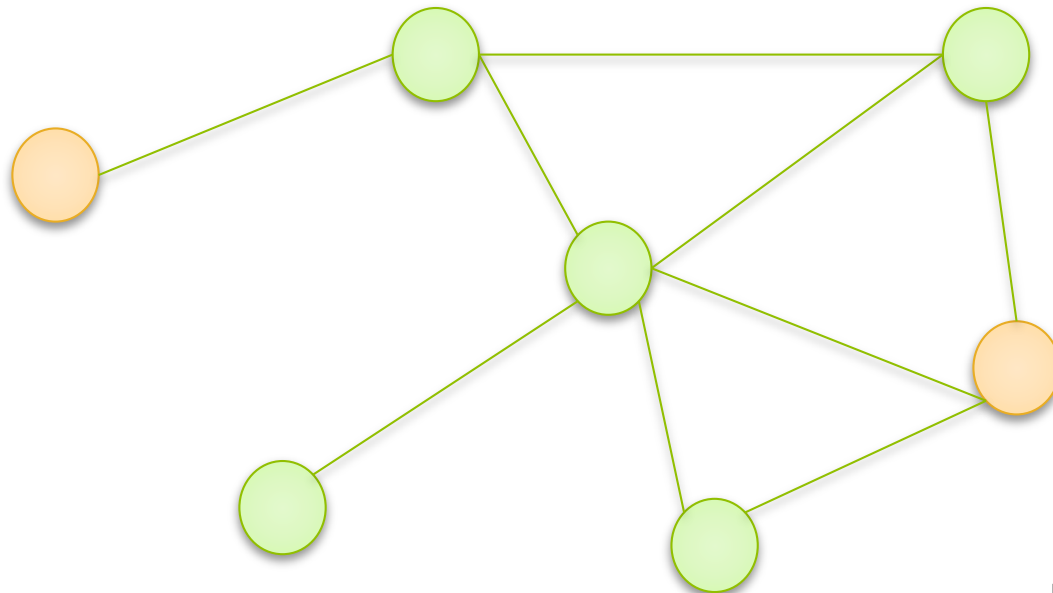
**SI**



**NO**

# Diametro del grafo

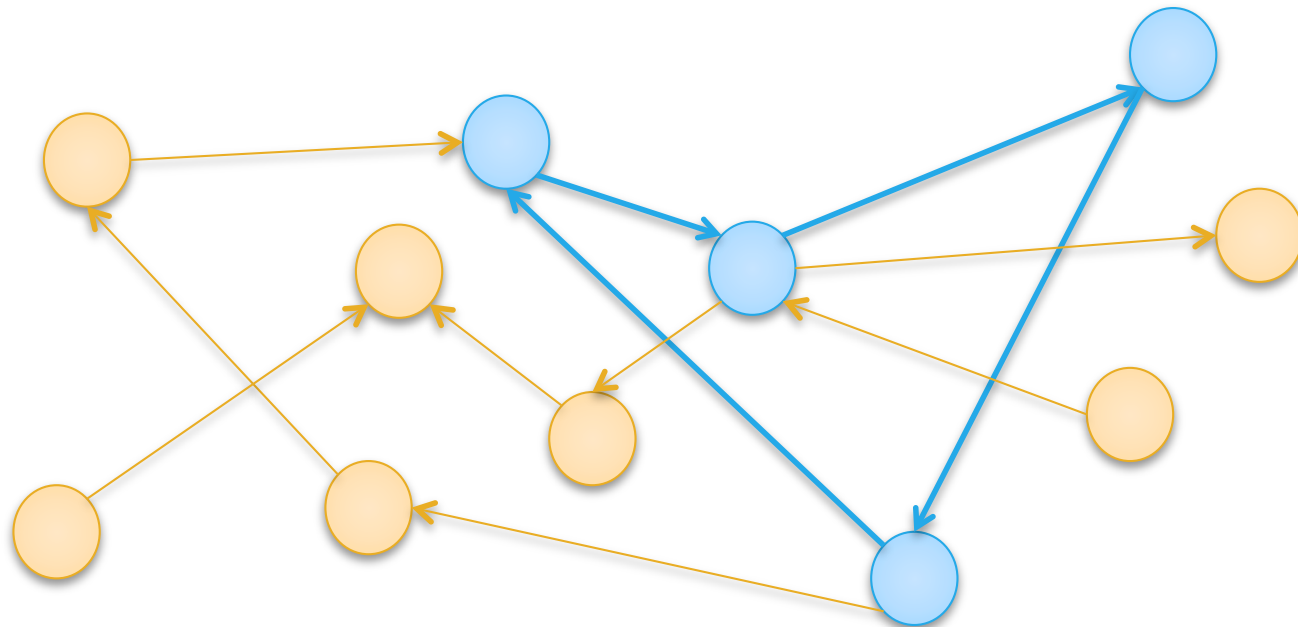
- Il diametro di un grafo e' costituito dal numero di archi che compongono il cammino minimo più lungo presente nel grafo



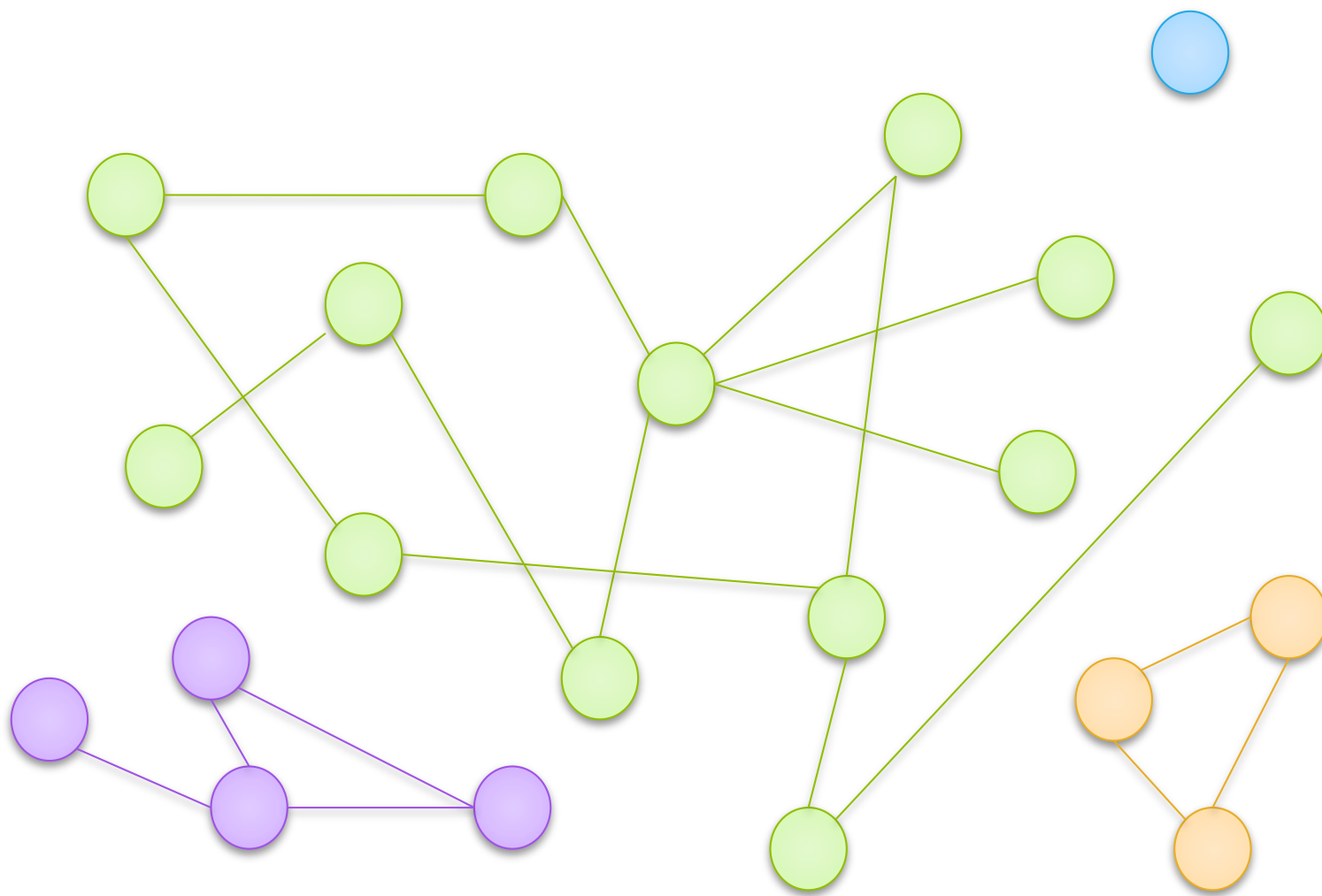
Diametro = 3

# Cicli

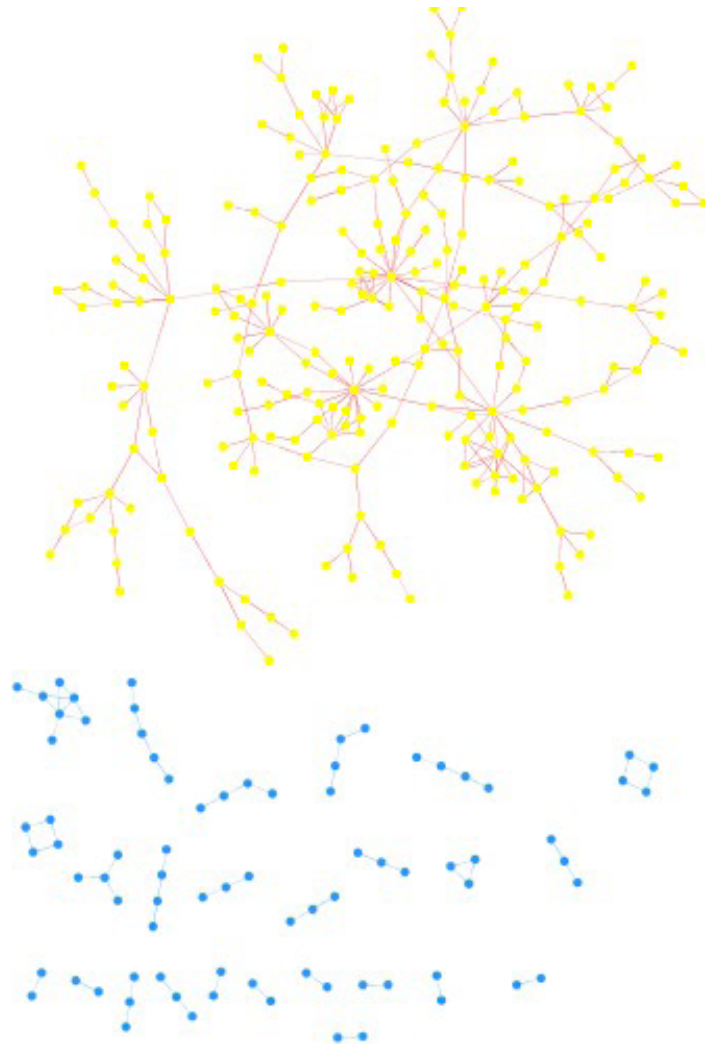
- Un ciclo e' un cammino che parte da un nodo per tornare su se stesso
- La lunghezza del ciclo e' data dal numero di nodi (o archi) di cui e' composto



# Componenti connesse e nodi isolati



# Giant Component

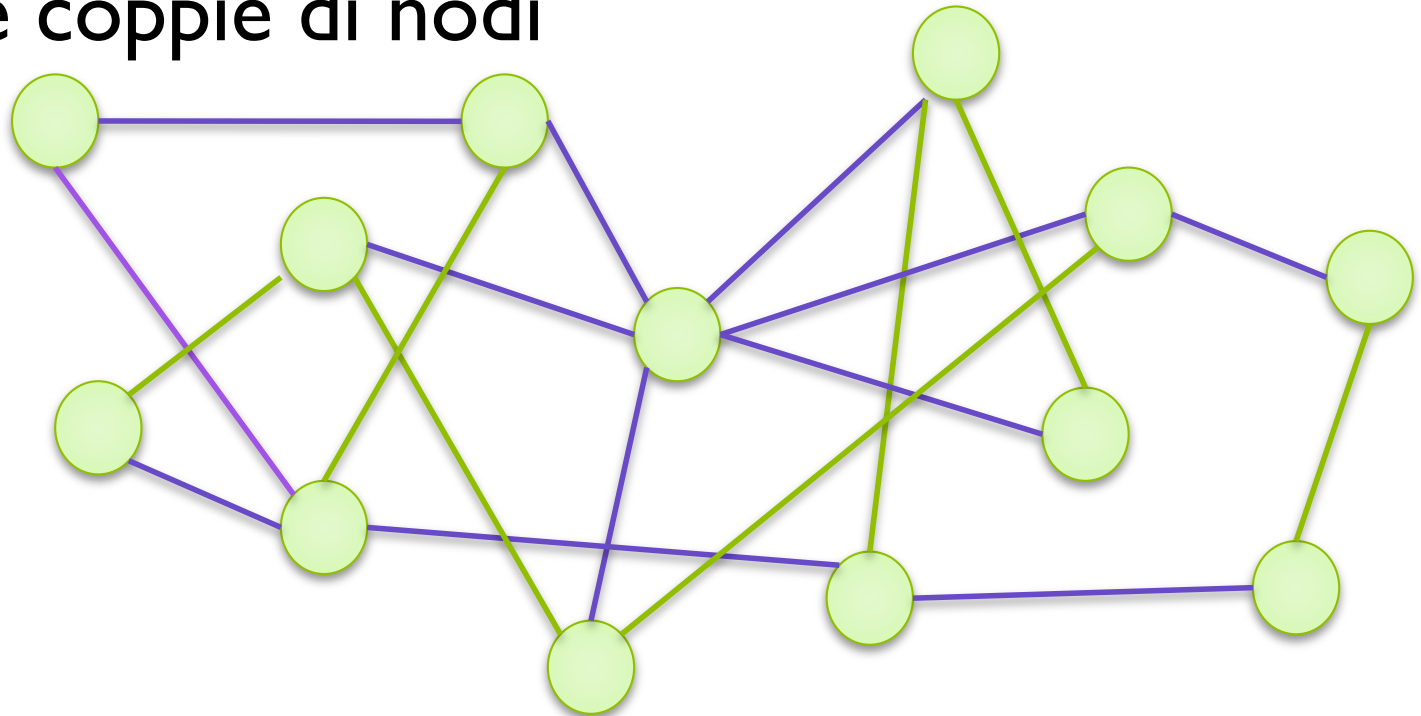


- In una rete sociale la maggior parte dei nodi sono connessi
  - Tipicamente tra il 60% ed il 100%
- Questo fenomeno e' noto come giant component



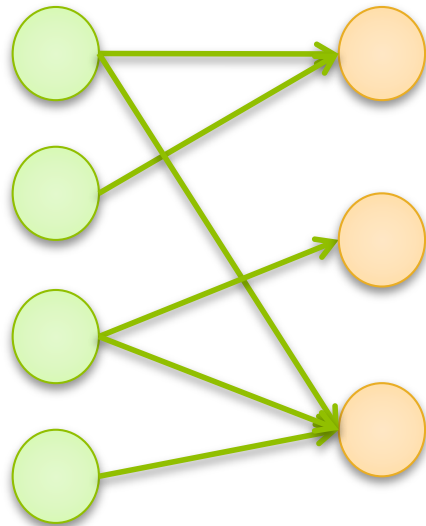
# Alberi di copertura

- Un albero di copertura di un grafo e' un albero che contiene tutti i nodi del grafo ed un sottoinsieme degli archi tale che esista uno ed un solo cammino tra tutte le coppie di nodi



# Strutture notevoli: grafi bipartiti

- Due insiemi di nodi disgiunti
  - Nelle reti sociali possono essere di tipi diversi
- Gli archi collegano nodi del primo insieme a nodi del secondo

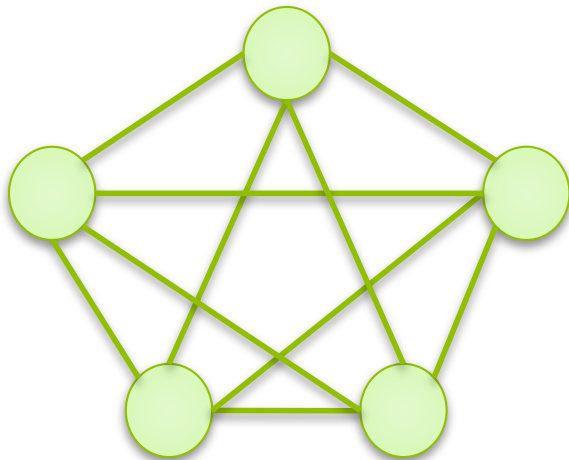


## Esempi

- **Clienti / prodotti**
  - Permette di stabilire se ci sono gruppi di clienti con gusti simili
- **Mansione / salario**
  - Permette di stabilire l'equità delle retribuzioni in azienda

# Strutture notevoli: clique

- Insieme di nodi tali che ogni nodo e' collegato con tutti gli altri
  - Nelle reti sociali si fa riferimento anche alle comunità che sono delle strutture simili alle clique in cui alcuni collegamenti sono assenti



## Esempi

- Mail tra membri di un team
  - La mancanza di un link potrebbe denunciare problemi relazionali tra i due membri

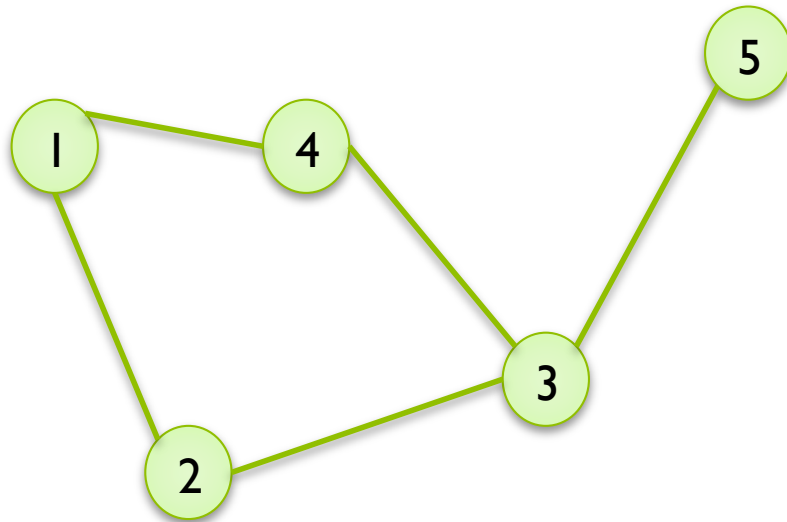
# Memorizzazione dei grafi

- Un grafo può essere memorizzato in due forme
  - Matrice di adiacenza
  - Lista di adiacenza
    - Per ogni nodo conservo la lista dei suoi vicini
    - Occupa tanto spazio quanto la somma di nodi più archi
    - Per accedere ad un arco bisogna prima accedere al nodo corrispondente

# Matrici di adiacenza

- Matrici quadrate  $n \times n$  in cui l'elemento in posizione  $i,j$  vale:
  - 1 se c'è un arco tra il nodo  $i$  ed il nodo  $j$ ,
  - 0 altrimenti
- Per grafi diretti la matrice è simmetrica
  - $A_{i,j} = A_{j,i}$
- Per grafi pesati al posto del valore 1 si mette il peso dell'arco
- Occupa spazio quadratico nel numero dei nodi
- Accesso immediato ad un arco

# Esempio



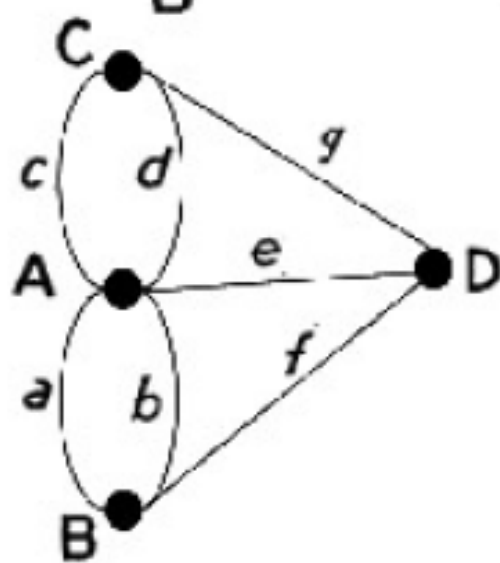
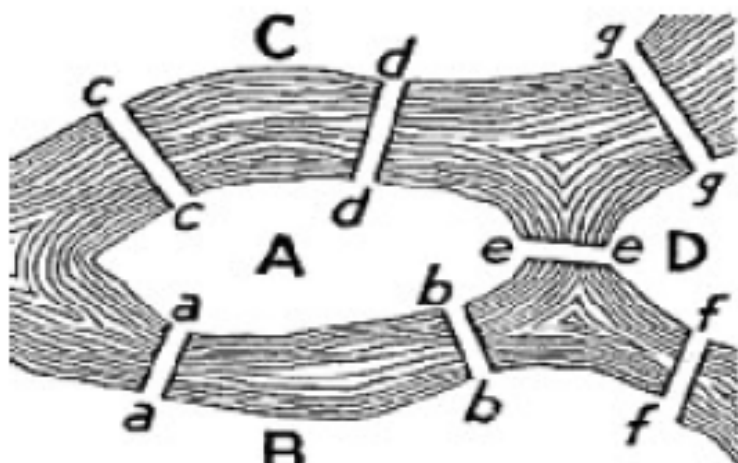
- 1 → 2, 4
- 2 → 1, 3
- 3 → 2, 4, 5
- 4 → 1, 3
- 5 → 3

|  |  |  |  |  |
|--|--|--|--|--|
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |



# Le reti sociali

# Un cenno storico



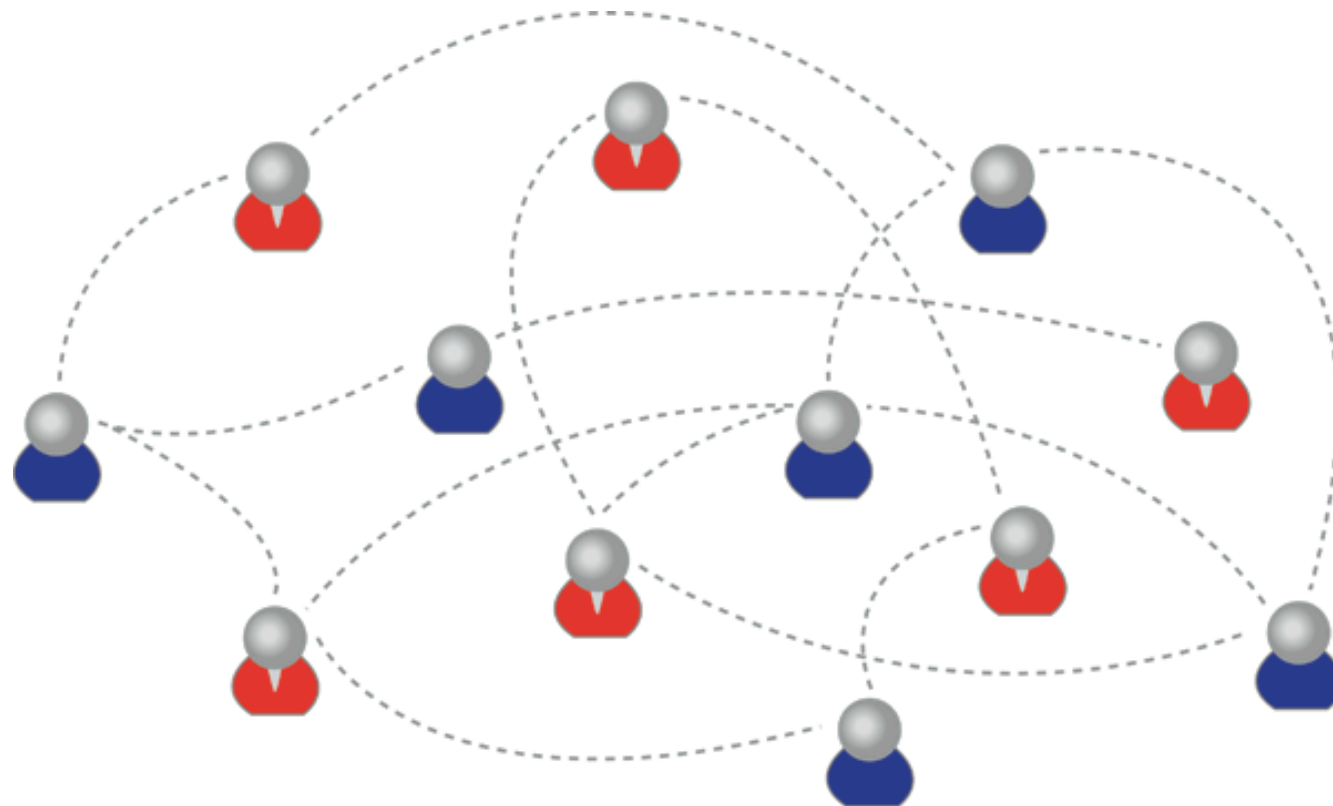
**Euler's Graph**

- La prima rete sociale viene presentata da Moreno nel 1934 e riguarda le relazioni di amicizia tra compagni di scuola
- Nel 1735 Eulero risolve il problema dei ponti di Königsberg utilizzando per la prima volta tecniche di social network analysis



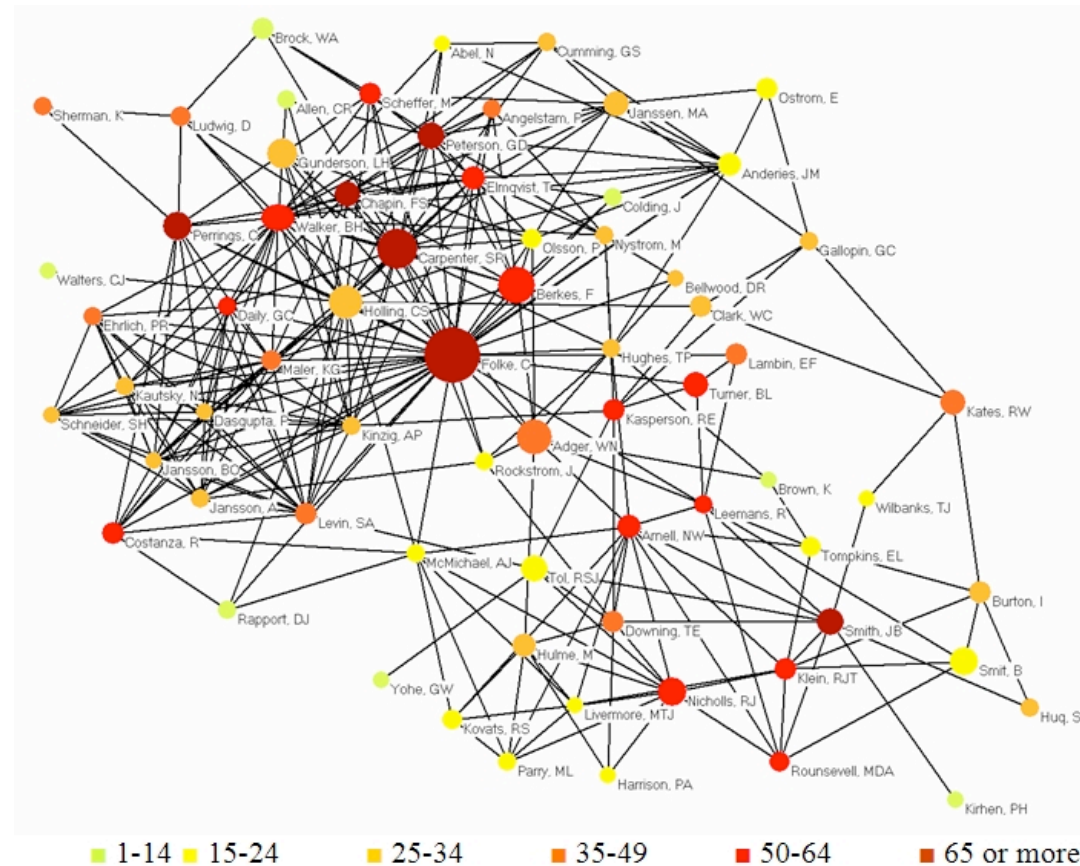
# Reti sociali nel mondo reale: la società

- Nodi: gli individui
- Archi relazioni tra essi



# Reti sociali nel mondo reale: le collaborazioni scientifiche

- Nodi: gli autori
- Archi: la condivisione di un articolo





## Caratteristiche delle reti sociali: il piccolo mondo (small world)

- Scoperto sperimentalmente, dal sociologo Milgram.
- La distanza tra le coppie dei nodi in una rete sociale e' molto bassa rispetto alla dimensione generale della rete.
- In presenza di Small World si hanno reti:
  - Di grandi dimensioni
  - Sparse
  - Bassa distanza media tra i nodi



# L'esperimento di Milgram e i sei gradi di separazione

- L'esperimento ideato da Milgram consisteva nel tentare di far pervenire, ad una persona residente a Boston, delle lettere, senza inviarle direttamente a questa persona.
- Per far ci`o il sociologo:
  - Inviò 160 lettere ad altrettante persone.
  - Ogni persona avrebbe dovuto rispedire la propria copia, ad una persona di proprio conoscenza, che avrebbe avuto maggiori possibilità di far arrivare la missiva a destinazione.
  - Giunsero a destinazione circa un terzo delle missive.
  - La media dei passaggi, conteggiata sulle lettere arrivate, si rivelò pari a 6.



# Esempi di reti sociali in azienda

- Rapporti di collaborazione tra aziende
- I prodotti venduti contemporaneamente
  - Sugli archi metto come peso il numero di volte che l'evento si e' verificato
  - Diverso dal grafo bipartito dei prodotti acquistati dai clienti
- I macchinari usati contemporaneamente per evadere una richiesta produttiva
- Il passaggio delle pratiche fra persone/uffici
  - Grafo diretto perché la relazione e' tra chi passa la pratica e chi la riceve
- I dipendenti che svolgono la stessa mansione
  - Mi permette di sapere se l'indisponibilità di un dipendente mi crea un gap di conoscenza



## Strumenti di analisi delle reti sociali

- **Indici di centralità:** mi dice quanto un nodo sia importante all'interno della rete
- **Indici di flusso:** misura quanto è coinvolto un nodo nel fluire dell'informazione attraverso la rete
- **Ricerca di pattern:** individua fenomeni che si ripetono all'interno della rete
- **Sottografi densi:** indicano forti correlazioni tra fenomeni

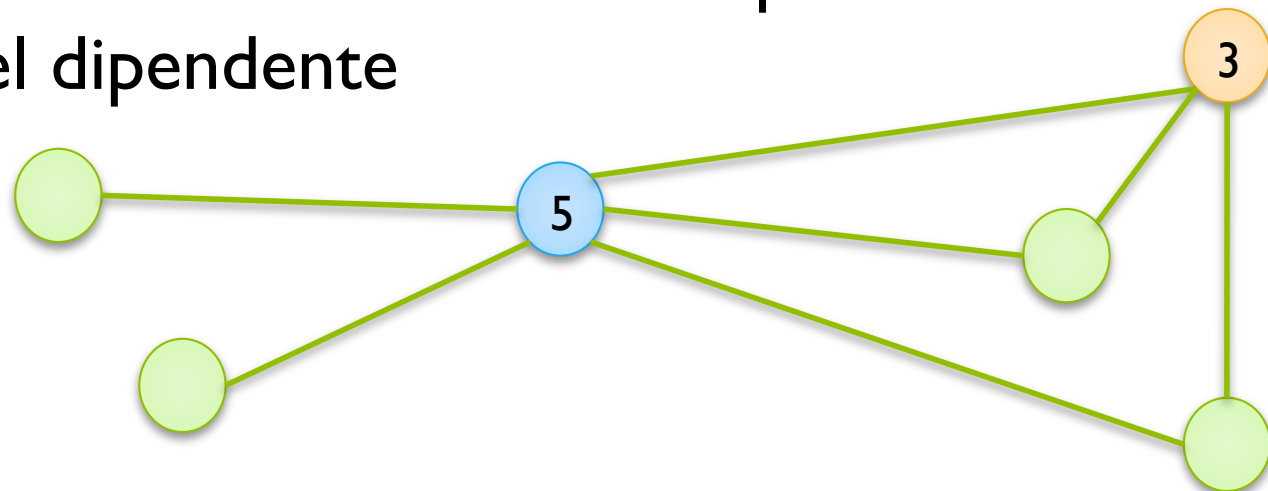


# Indici di centralità e di flusso

- L'importanza di un nodo dipende dal contesto
- Un nodo può essere importante se:
  - È connesso con molti nodi (degree centrality),
  - È raggiungibile nel minor tempo (o costo) da ogni altro nodo (closeness centrality)
  - È posto sui cammini minimi fra molte coppie di nodi (betweenness centrality)
  - È vicino a nodi importanti (eigenvector centrality).
- Alcuni di questi indici fanno ipotesi esplicite su come fluisce l'informazione sulla rete

# Indici di centralità - Degree centrality

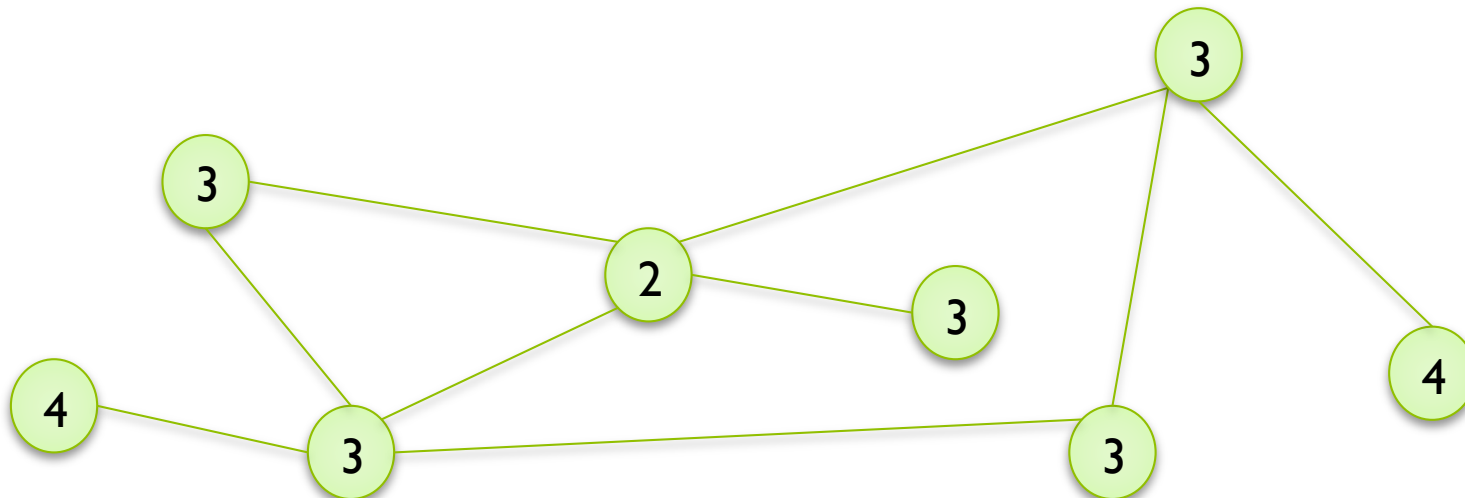
- Per ogni nodo e' il suo grado
- Misura la capacità di un nodi di connettersi con i vicini
- **Esempio:**
  - Nel grafo delle collaborazioni fra dipendenti grado alto indica buona capacità relazionale del dipendente





# Indici di centralità - Eccentricità

- L'eccentricità di un nodo è la massima distanza fra il nodo ed ogni altro nella rete.
- L'indice di centralità basato sull'eccentricità valuta come migliori i nodi con minor massima distanza.
- I nodi identificati da questo indice sono quelli appartenenti al centro del grafo.

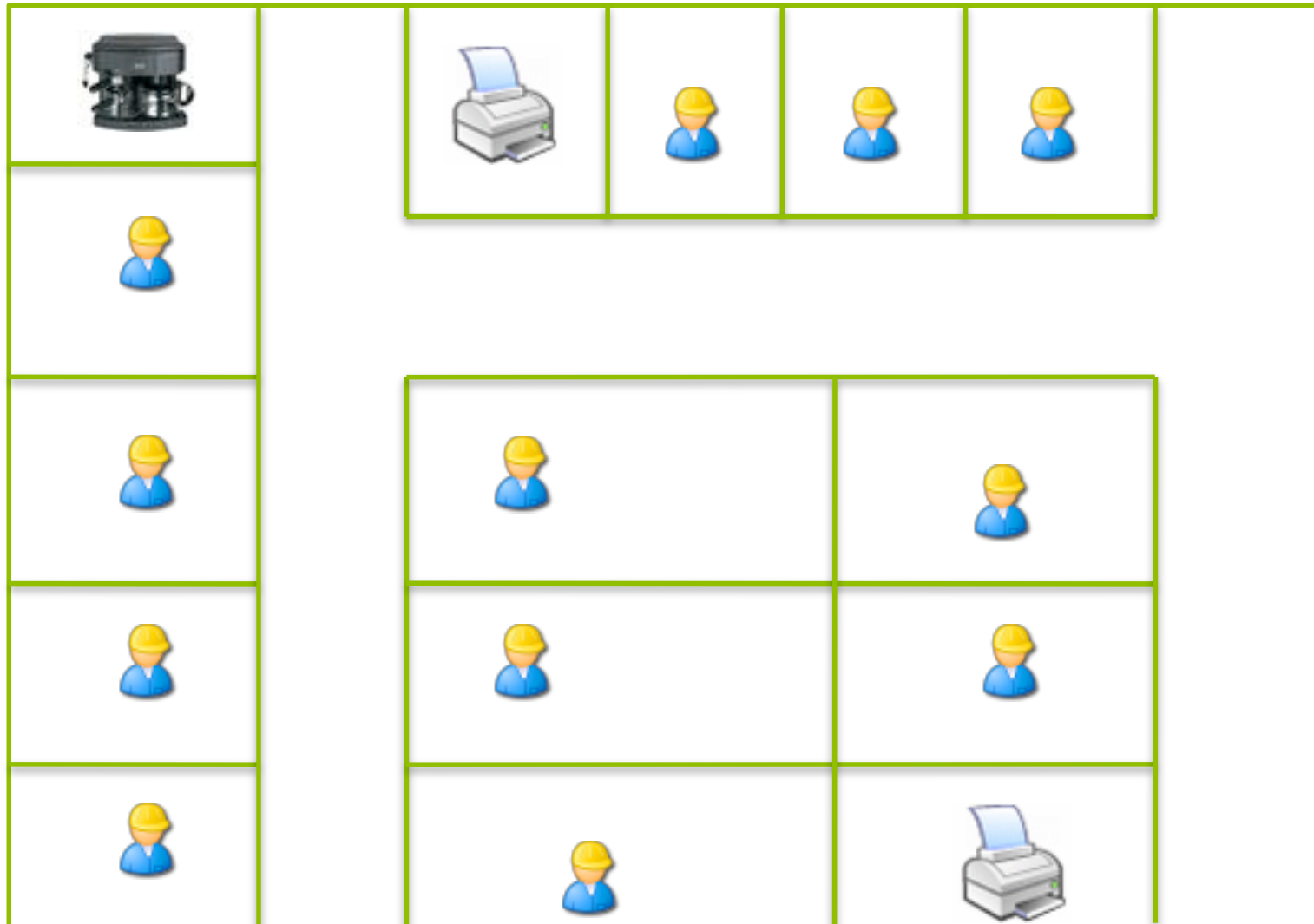




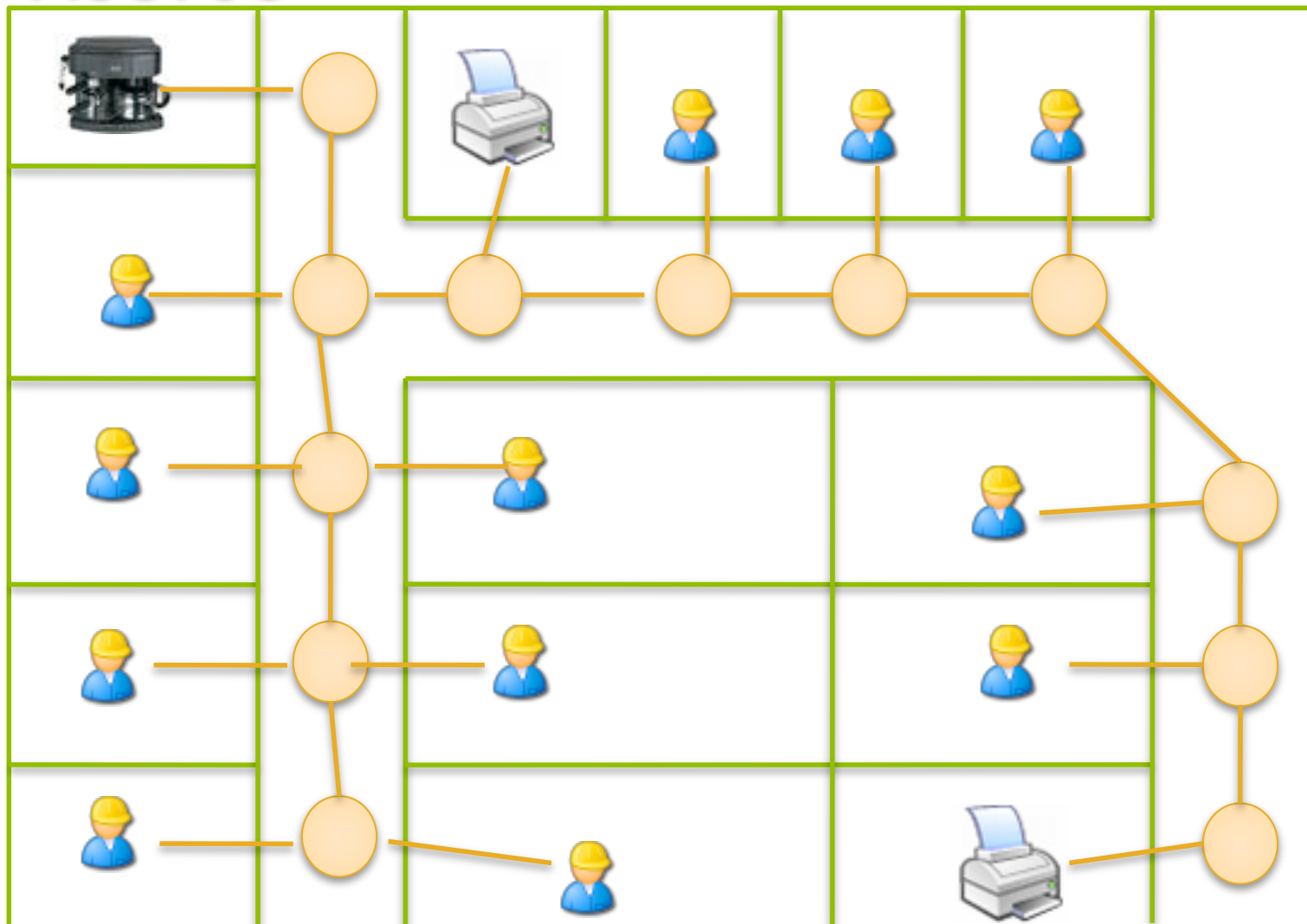
# Indici di centralità - Closeness

- Somma della distanza di un nodo da tutti gli altri
- Ben definito solo per grafi connessi
- I nodi più centrali sono quelli che hanno valori minori per l'indice.
- Questo indice assume che l'informazione fluisca attraverso un cammino minimo
- Può essere usato nei problemi di collocamento di una risorsa quando vogliamo che essa sia nel luogo più facilmente accessibile da ogni altro nodo della rete.

# Esempio

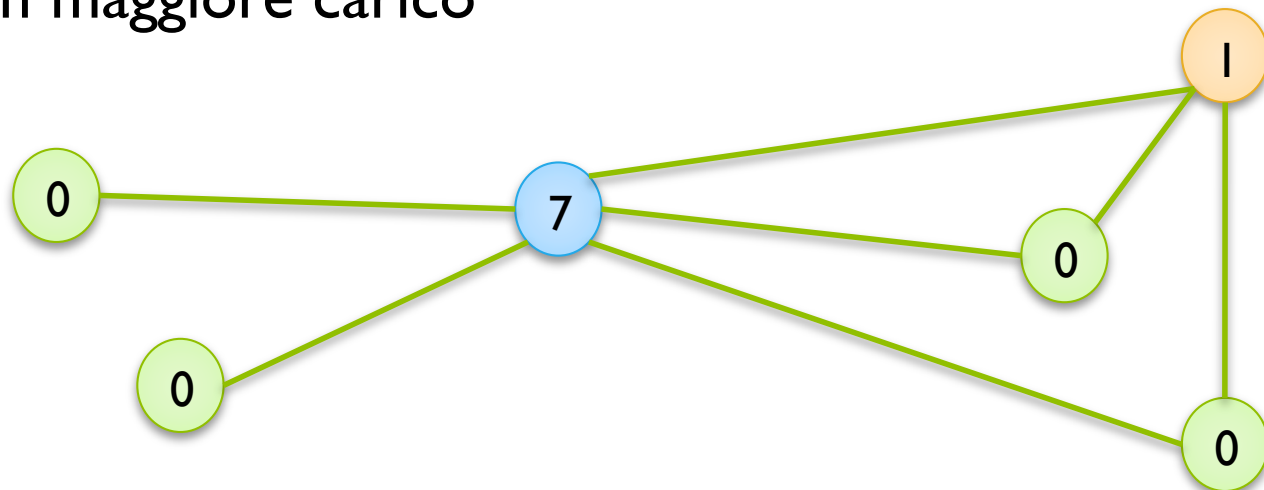


# Esempio – grafo delle distanze tra le risorse



# Indice di flusso - betweenness

- Per un nodo  $v$  e' la somma del numero di cammini tra tutte le coppie di nodi nel grafo che passano per  $v$
- Questo indice dà una misura del controllo che il vertice  $v$  ha sul flusso dell'informazione nelle rete fra altri nodi.
- **Esempio:**
  - Nel grafo dell'iter tra uffici di una pratica indica quello con maggiore carico

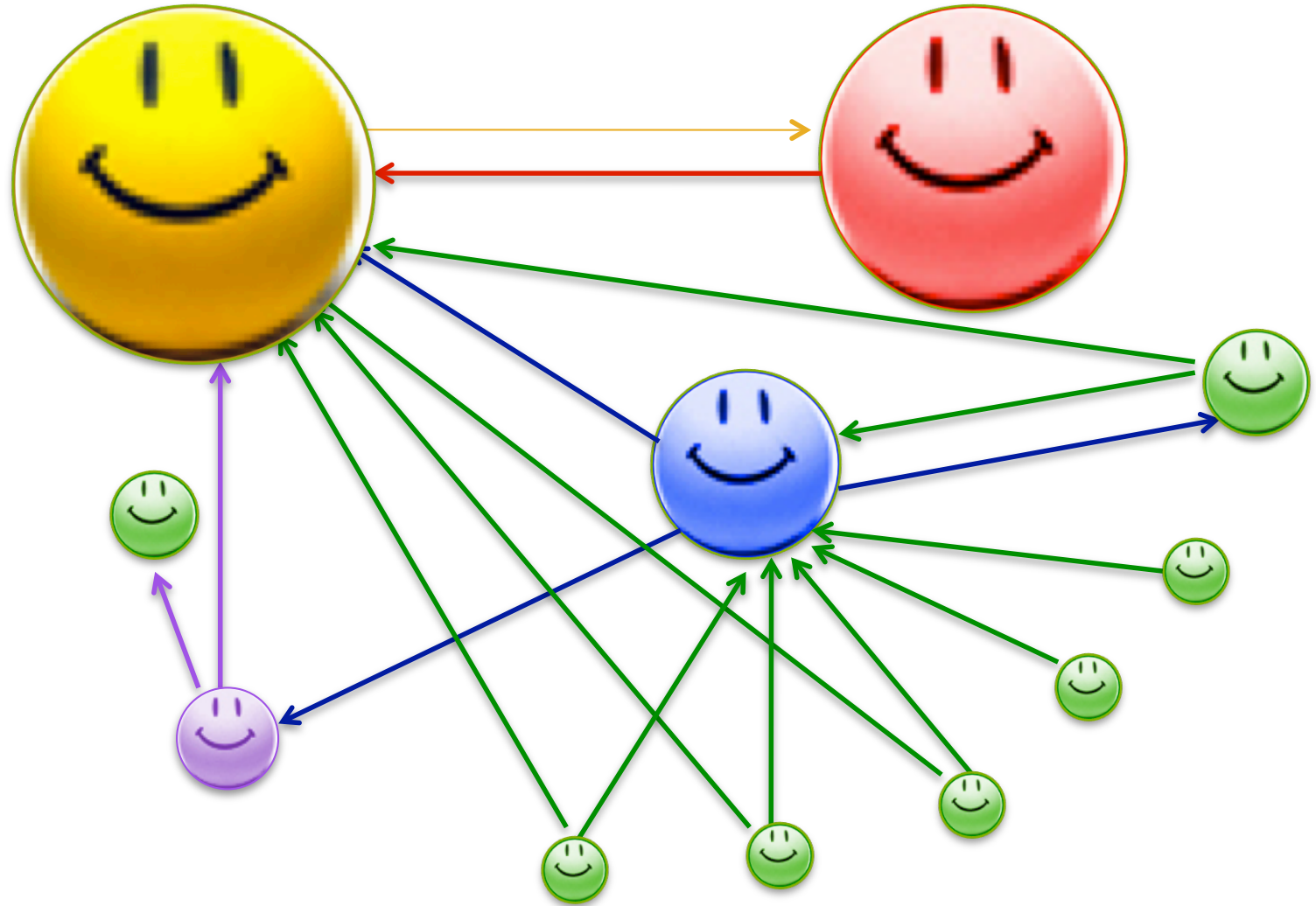




# Eigenvector centrality

- Misura l'importanza delle connessioni di un nodo.
- Nella metafora delle reti sociali avere una connessione con persone più influenti rende una persona più importante che avere una connessione con persone poco influenti
- La centralità di un nodo dipende in maniera proporzionale da quella dei vicini
- E' un calcolo ricorsivo molto pesante computazionalmente
- Il più famoso di questi indici e' pagerank

## II PageRank



# Esempio

- Sia una rete in cui i nodi sono i membri di un consiglio di amministrazione
- Supponiamo che ogni membro nelle votazioni valga 1
- Tra due nodi vi è un arco (diretto) se il membro del C.d.A. associato al nodo sorgente ha appoggiato una proposta del membro associato al nodo destinazione
- Il pageRank in questo caso misura la propensione di un membro del C.d.A. a far prevalere la sua opinione

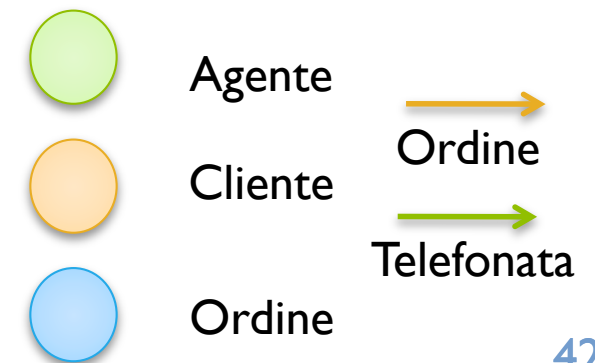
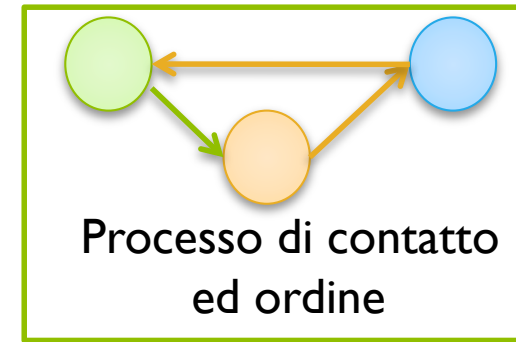
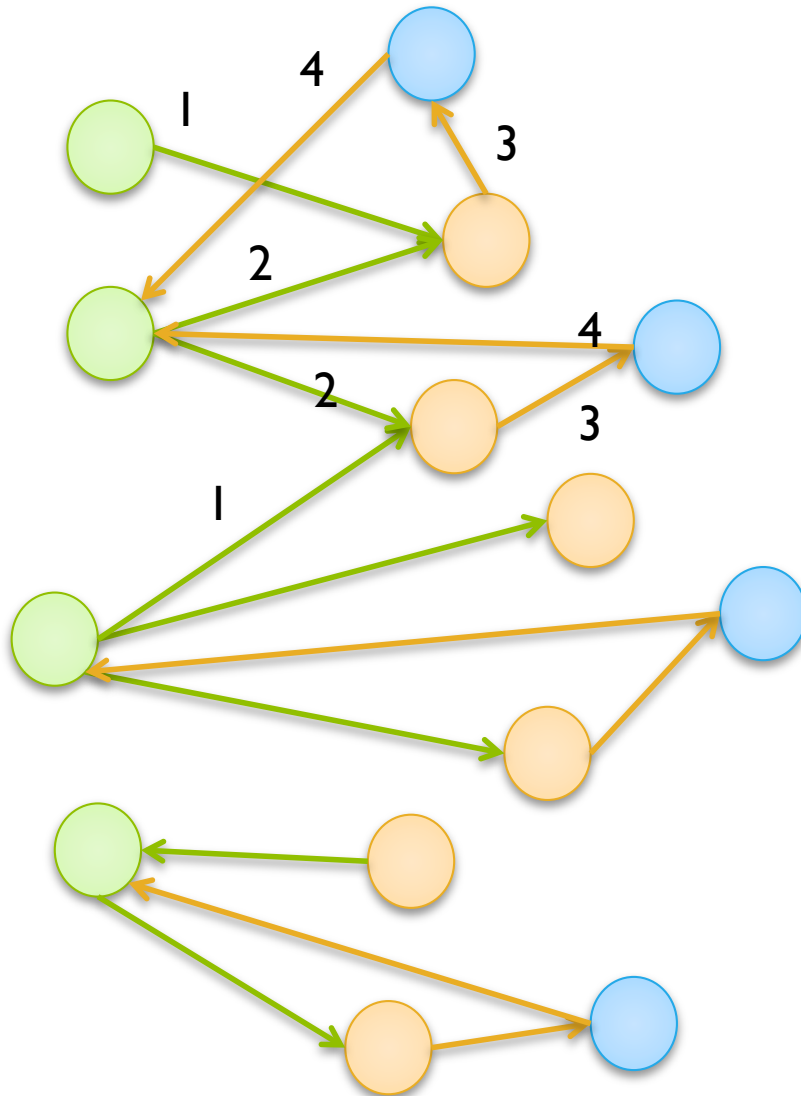




# Ricerca di pattern

- Permette di individuare fenomeni che si ripetono di frequente
- **Esempio:**
  - Grafo in cui i nodi sono agenti, clienti ed ordini
  - C'e' un arco tra agente e cliente (o viceversa) se uno chiama l'altro
  - Ad ogni arco associo un timestamp
  - Secondo le regole aziendali un cliente e' associato solo ad un agente

# Esempio



# Sottografi densi

- Indicano forti correlazioni tra fenomeni
- Spesso indicano comunità
- **Esempio:**
  - nella rete degli articoli acquistati insieme rappresentano oggetti che vengono usati spesso insieme
    - Latte, biscotti, cereali, zucchero, caffè
    - Posso volere posizionarli vicino anche se appartengono a categorie merceologiche diverse
  - Una rete in cui i nodi sono i membri del C.d.A. e gli archi (orientati) il fatto di aver votato una mozione proposta da un altro
    - Il sottografo identifica lobby di potere