

Reti biologiche: Analisi e proprietà

Giovanni Scardoni

CBMC (center for biomedical computing)

Università di Verona

giovanni.scardoni@gmail.com

Prof. Carlo Laudanna, dip. Patologia Università di
Verona, CBMC

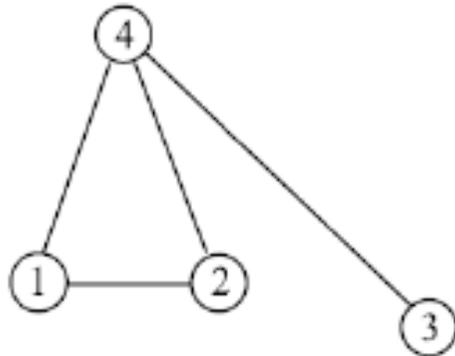
Network: definizione matematica

- Grafo (network) $G=(N,E)$
 - N =insieme dei nodi
 - E =insieme degli archi (edges). Gli archi sono coppie di nodi
- Coppie ordinate $(a,b) \neq (b,a)$:
 - grafo diretto
- Coppie non ordinate $(a,b) = (b,a)$:
 - grafo diretto

Network: definizione matematica

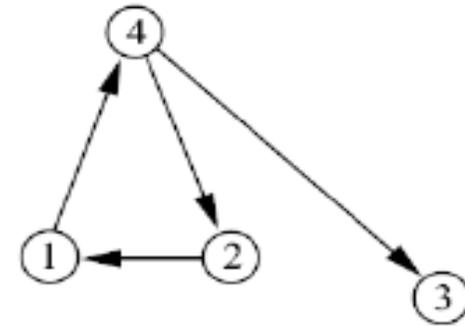
- Matrice di adiacenza

$$A = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$$



Grafo indiretto

$$A = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{pmatrix}$$



Grafo diretto

Network: definizione

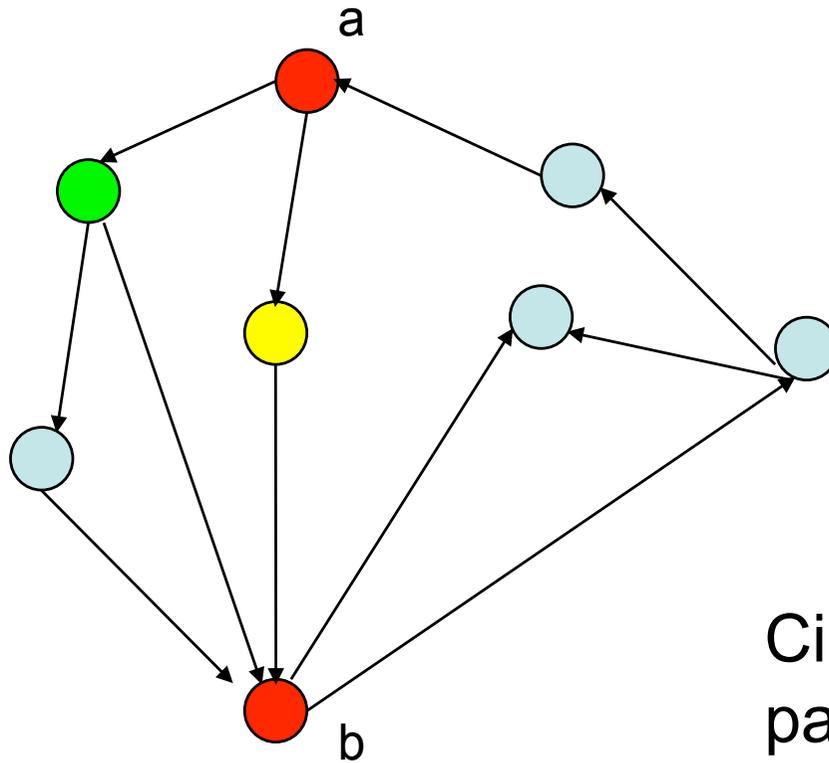
- Parametri fondamentali

- Grado (degree): numero di archi connessi al nodo (distinguere in e out degree nei grafi diretti):

grado di k
$$k_i = \sum_{j=1, n} a_{ij}$$

- Cammino minimo (Shortest Path): numero di archi tra un nodo a l'altro (chiamata anche distanza)
- Diametro: distanza massima tra due nodi del grafo

Shortest path

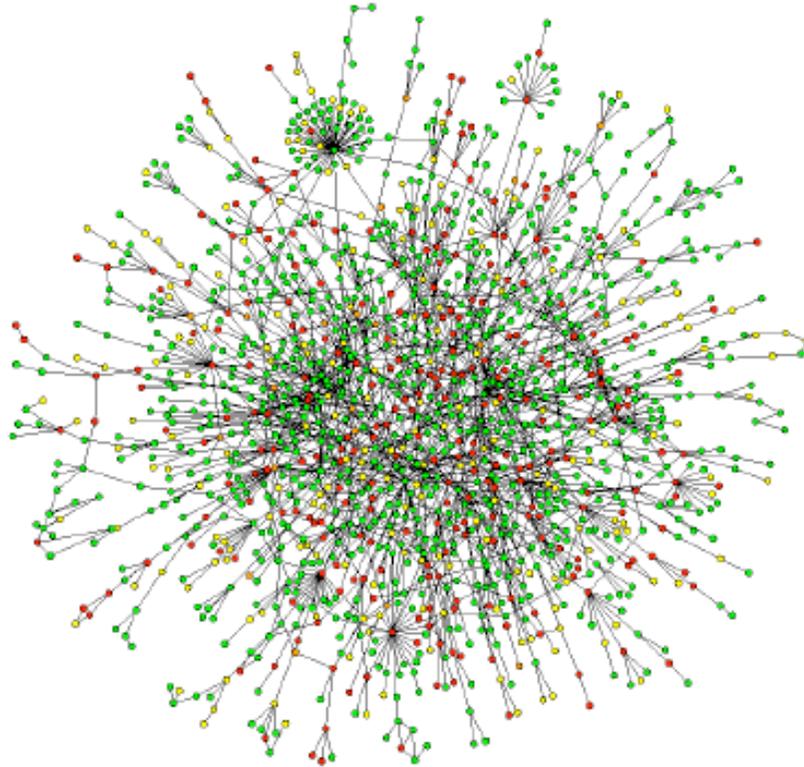


Ci sono due shortest path tra a e b

Reti: esempi

- Sociali (relazioni tra persone, amicizie, parentele, facebook, epidemie)
- Economiche (relazioni tra aziende, tra consumatori, aziende-consumatori)
- Internet
- Reti biologiche

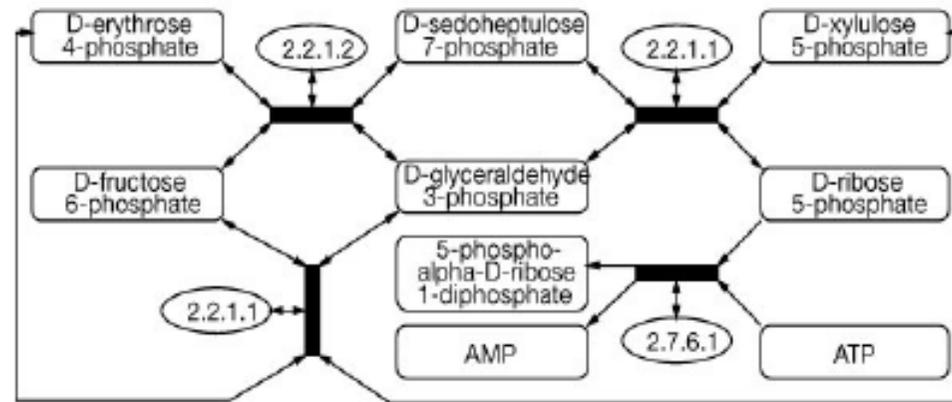
Protein interaction network



- Nodi: proteine
- Archi: interazioni proteina-proteina

Map of yeast protein-protein interactions, by Hawoong Jeong

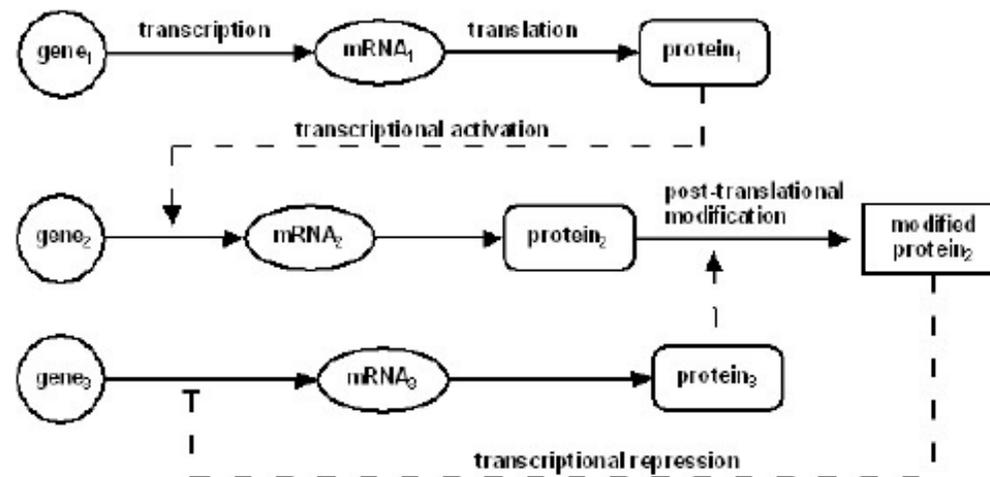
Biochemical reactions network



- I nodi possono essere reagenti, substrati, prodotti di reazioni, enzimi, o complessi (reagente enzima)
- Gli archi sono le reazioni (direzionali, es. da reagenti al prodotto della reazione o viceversa)

Gene regulatory network

- Descrivono le interazioni che influenzano l'espressione genica
- I nodi possono essere geni, mRNA, proteine
- Gli archi
 - Possono indicare inibizione o attivazione di reazioni
 - Possono indicare produzione di mRNA o proteine



Network analysis

Ito (yeast): 8868 interactions between 3280 proteins

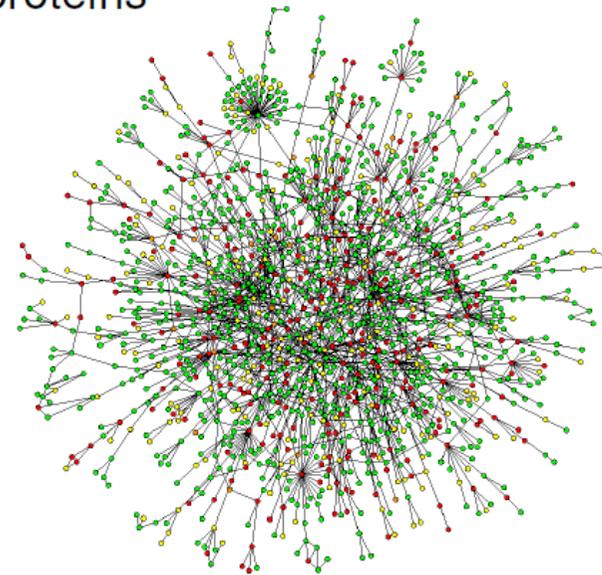
Uetz (yeast): 4480 interactions, 2115 proteins

Giot (Drosophila): 4780 interactions among 4679 proteins

Li (C. elegans): 5534 interactions, 3024 proteins

Rual (human): 2800 interactions, 8300 proteins

- Protein interactions are considered symmetrical
- Many untested interactions, false positives – **problem**
- All networks have giant connected components.
- The topological properties of diverse protein interaction networks are similar.



H. Jeong et al. Nature 411, 41-42 (2001)

S.-H Yook, Z.N. Oltvai, A.-L. Barabasi, Proteomics 4, 928 (2004)

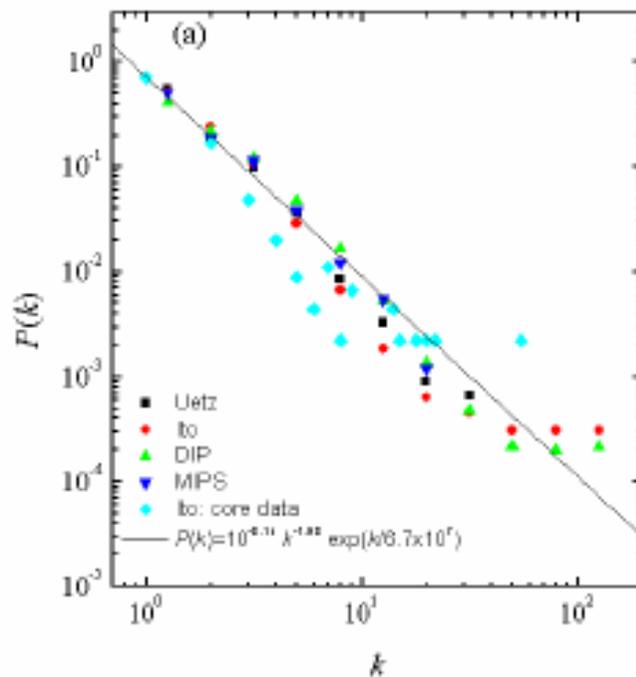
Sono necessari software per l'analisi dei network!

Proprietà dei network: scale free

- Le reti biologiche (e molte altre) sono scale free:
 - $P(k)$ è la frazione di nodi del network che hanno grado k
 - $P(k) \sim k^{-\gamma}$ con γ con $2 < \gamma < 3$
 - Molti nodi con pochi “vicini” pochi nodi con molti “vicini”
 - Alcuni nodi sono più importanti di altri

Scale free network

Yeast interaction network

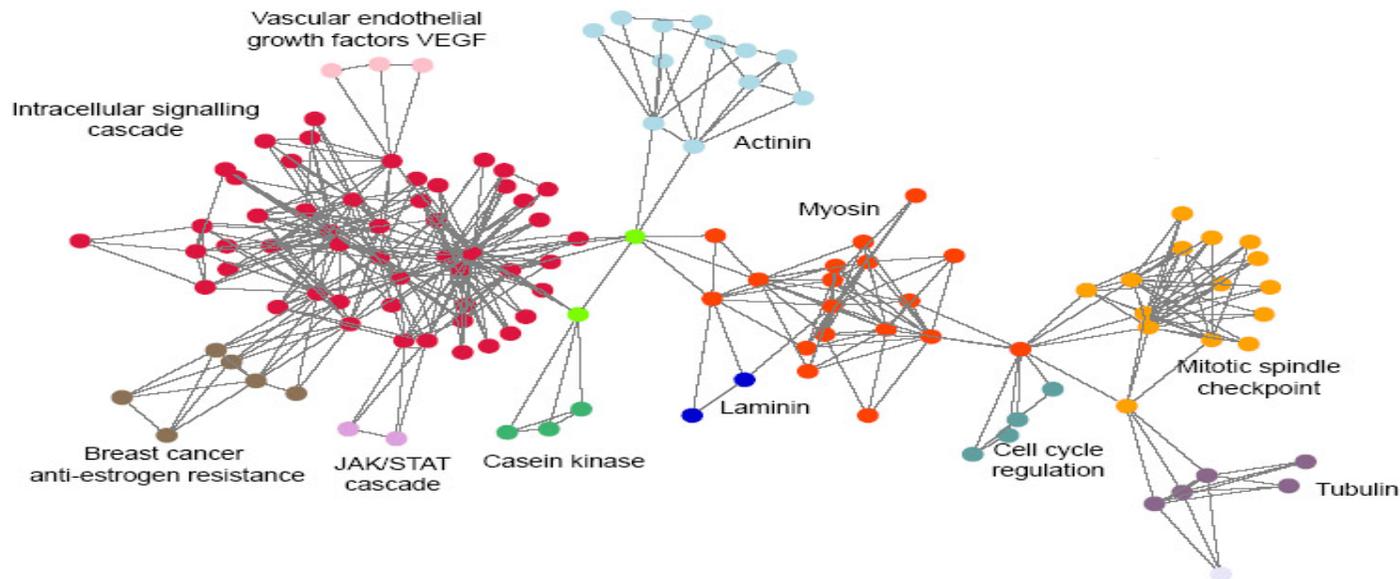


$$P(k) \sim k^{-2,5}$$

- I nodi con più interazioni sono determinanti
- Vale anche per
 - Social network (vip, politici ecc...)
 - World wide web (siti più linkati, google)

Proprietà dei network: Clusters

- Tendenza a clusterizzare
 - Cluster = agglomerato di nodi con alto numero di interazioni tra loro



Come identificare i Clusters?

Clusters

- Coefficiente di clusterizzazione

$$C_i = \frac{1}{k_i(k_i - 1)/2} \sum_{jk} a_{ij} a_{ik} a_{jk}$$

$$C_i = \frac{\textit{n}^\circ \textit{ di triangoli connessi a } i}{\textit{n}^\circ \textit{ di terzine centrate in } i}$$

Clusters

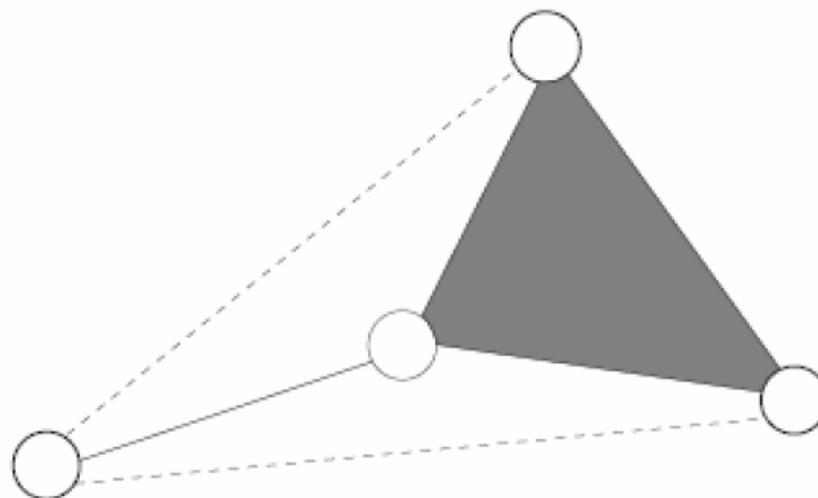
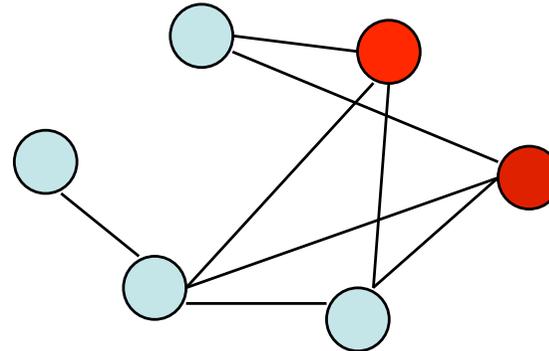
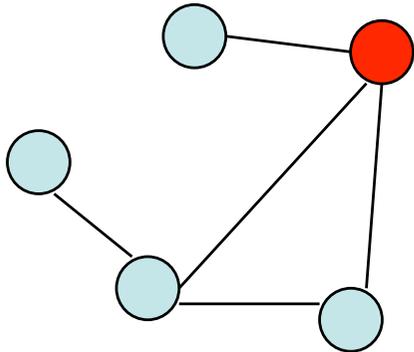


Fig. 1.5 The clustering coefficient of the central vertex is $1/3$. This is because its degree is three and its neighbours can be connected each other in three different ways. Of these possibilities (dashed line) only one is actually realized (solid line) and therefore $C_i = 1/3$. The three connected vertices form the coloured triangle. For that reason, sometimes the clustering coefficient of a vertex is defined through the number of triangles it belongs to.

Clusters: quale origine?

- Esempi:
 - Duplicazione del gene
 - Copia di un sito web
 - Amici degli amici



Proprietà dei network: small world property

- Ciascun nodo può raggiungere gli altri con un numero limitato di passi
- Esperimento di Milgram (1968):
 - Si prendono due cittadini a caso e si conta il numero di contatti che intercorrono tra loro
 - Viene detto al primo di inviare una lettera al destinatario (se conosciuto) oppure a qualcuno che lui ritenga conoscere il destinatario.
 - Milgram sceglie Omaha, Nebraska e Wichita, Kansas come città di partenza e Boston, Massachusetts come destinazione finale
 - Tali città sono molto distanti tra loro sia geograficamente che socialmente.

Proprietà dei network: small world property

- Risultati
 - La distanza media è di circa 5 o 6 passaggi
 - Se ciascuno conosce 100 persone:
 - $100^5 = 10.000.000.000$ (vanno tolte le conoscenze in comune)
 - Successivi esperimenti confermano l'ipotesi
 - Duncan Watts e Steven Strogatz studiano matematicamente il fenomeno (Nature 1998)
 - Curiosità: six degrees of Kevin Bacon (<http://oracleofbacon.org/>)

Proprietà dei network: small world property

La small world property è verificata per i network biologici!

- Interattoma umano (Laudanna, CBMC):
 - Decine di migliaia di proteine
 - Centinaia di migliaia di connessioni
- La distanza massima è stata calcolata in 8 passaggi (alcune connessioni sono mancanti)

Proprietà dei network: Network motifs

- Sottostrutture (sottografi) che si ripetono con frequenza alta nelle reti
- Si prendono network reali (www, food network, social network, circuiti elettronici, reti biologiche)
- Il confronto viene fatto con reti generate casualmente
- Si trovano diversi tipi di motifs che sono molto più frequenti nelle reti reali che non in quelle generate casualmente
- I network motifs sottendono a particolari funzioni nella rete

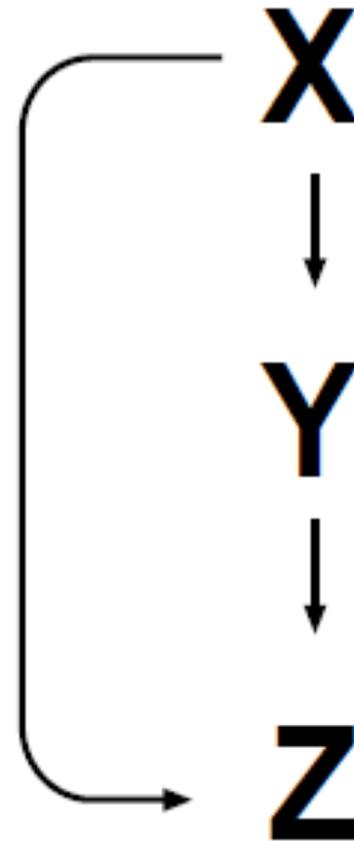
Network motifs: esempi

Feedforward loop

Funzione: regolare nel tempo l'attivazione di Z

Z è attivato solamente se X è attivato abbastanza a lungo

Es: reti biologiche (Escherichia coli), circuiti elettronici



Shen –Orr, Milo, Mangan, Alon, Nature Genetics 2002

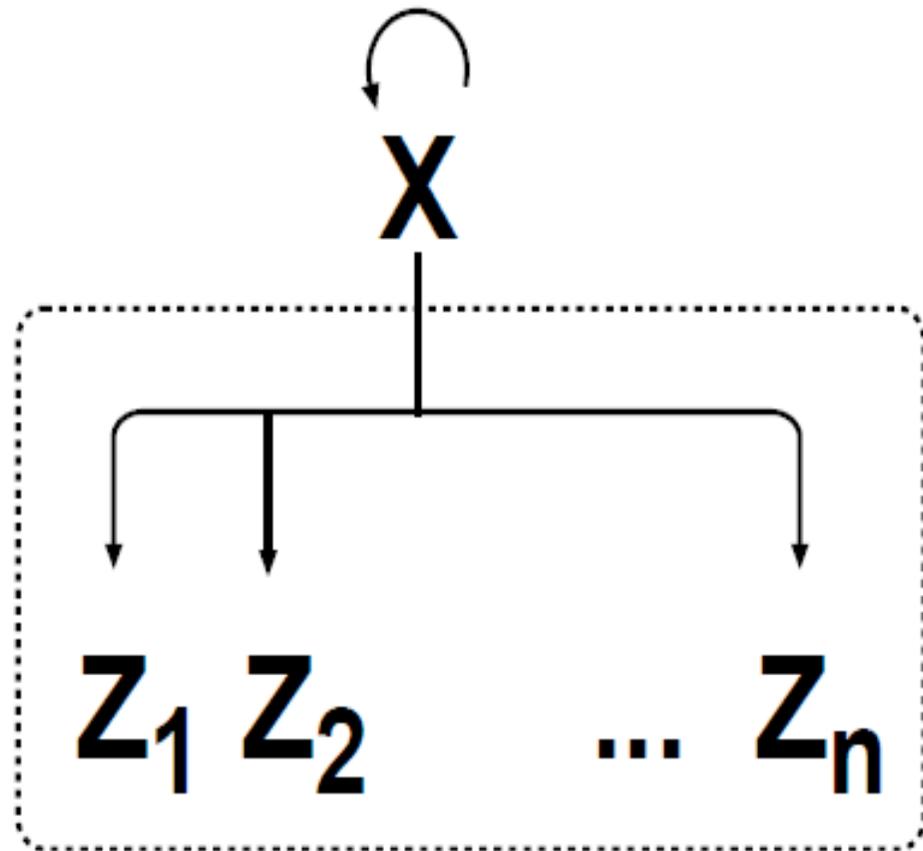
Network motifs: esempi

SIM: single input module

Funzione: sincronizzare
l'attivazione di $Z_1 \dots Z_n$

Una volta attivato X , $Z_1 \dots Z_n$
vengono attivati in tempi
precisi a seconda delle
rispettive soglie di
attivazione

Es: reti biologiche
(*Escherichia coli*), circuiti
elettronici



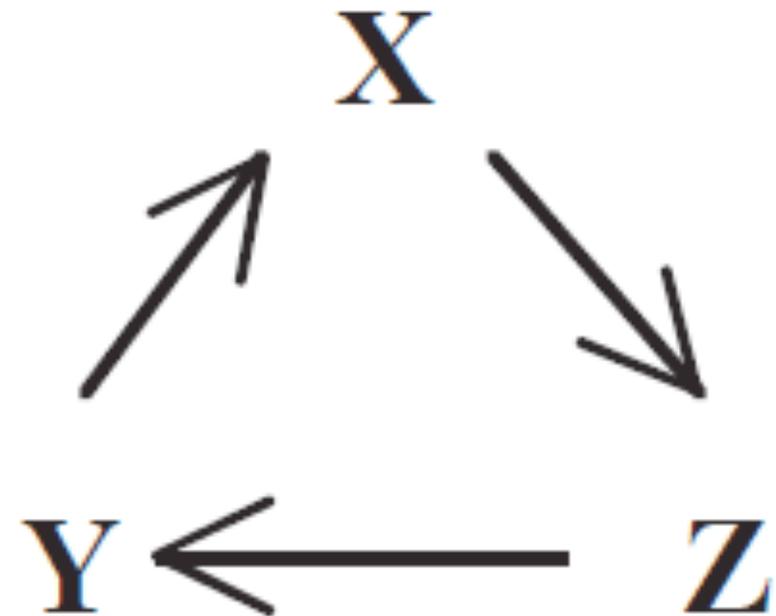
Network motifs: esempi

Feedback loop

Funzione: fenomeni oscillatori,
fenomeni permanenti nel
tempo

X attiva Z ed è a sua volta
attivato da Y

Es: reti biologiche (ritmo
circadiano, memoria,
fenomeni oscillatori)



R. Milo *et al*, Science 2002

Proprietà dei network: centralità

- Sono parametri attribuiti alla rete o ai suoi nodi
- Determinano particolari caratteristiche della rete
- Permettono di individuare i nodi più importanti per il corretto funzionamento della rete

Proprietà dei network: centralità

- Network parameter: valori attribuiti all'intera rete:
 - Diametro (Δ_G): massima delle distanze tra ciascuna coppia di nodi del network
 - Valore basso -> il network è “compatto”
 - Valore alto -> il network può essere “disperso”
 - Distanza media (AvD_G): distanza media tra le coppie di nodi del network
 - Valore basso -> molti nodi vicini tra loro
 - Valore alto -> molti nodi lontani tra loro

Per distanza si intende lo shortest path

Centralità dei nodi

- Sono parametri di ciascun nodo utilizzati inizialmente in sociologia ed economia
- Aiutano a comprendere il ruolo del singolo nodo all'interno del network
- Sono calcolati in modo che un valore alto indichi un risultato positivo (vicinanza tra i nodi)
- Un nodo con centralità alte è un nodo fondamentale per il corretto funzionamento del network

Centralità dei nodi

Eccentricity:

$$C_{ecc}(v) := \frac{1}{\max\{dist(v, w) : w \in V\}}$$

Valore alto -> tutti i nodi sono vicini a v
(il nodo è “centrale”)

Valore basso -> esistono nodi lontani da v

Centralità dei nodi

- Closeness

$$C_{clo}(v) := \frac{1}{\sum_{w \in V} dist(v, w)}$$

$dist(v, w)$ = lunghezza dello shortest path tra v e w

Valore alto -> il nodo è generalmente vicino a molti altri nodi (il nodo è “centrale”)

Valore basso -> il nodo è generalmente lontano dagli altri nodi (il nodo è “periferico”)

Centralità dei nodi

Radiality

$$C_{rad}(v) := \frac{\sum_{w \in V} (\Delta_G + 1 - dist(v, w))}{n - 1}$$

Valore alto -> il nodo è mediamente vicino agli altri (rispetto al diametro del network)

Valore basso -> il nodo è mediamente lontano dagli altri (rispetto al diametro del network)

Centralità dei nodi

Centroide

$$C_{cen}(v) := \min\{f(v, w) : w \in V \setminus \{v\}\}$$

$$f(v, w) := \gamma_v(w) - \gamma_w(v)$$

$\gamma_v(w)$ = numero di nodi che sono più vicini a v piuttosto che a w

Valore alto -> il nodo è decisamente centrale rispetto agli altri nodi (ha più vicini degli altri nodi)

Valore basso -> il nodo è decisamente periferico (ha meno vicini rispetto agli altri nodi)

Centralità dei nodi

Stress

$$C_{str}(v) := \sum_{s \neq v \in V} \sum_{t \neq v \in V} \sigma_{st}(v)$$

$\sigma_{st}(v)$ = numero di shortest path passanti per il nodo v

Valore alto -> per il nodo v passano molti shortest path (v è un “passaggio obbligato” tra molti nodi)

Valore basso -> pochi percorsi passano da v . (v è una strada secondaria)

Centralità dei nodi

- Betweenness

$$C_{spb}(v) := \sum_{s \neq v \in V} \sum_{t \neq v \in V} \delta_{st}(v)$$

$$\delta_{st}(v) := \frac{\sigma_{st}(v)}{\sigma_{st}}$$

Valore alto -> per il nodo v passano molti shortest path rispetto al resto del network (v è un “passaggio obbligato” tra molti nodi)

Valore basso -> pochi percorsi passano da v rispetto al resto del network (v è una strada secondaria)

Cytoscape!

- Piattaforma per la visualizzazione e l'analisi di reti biologiche
- Collaborazione tra: Institute Pasteur, Ucsd, Ucsf, mskcc, Unilever, ISB...
- Sistema a plug-in: chiunque puo' costruire dei programmi che integrino Cytoscape
- È gratuito. Tutte le informazioni e il download su: www.cytoscape.org

Come analizzare una rete?

- Studiando la struttura della rete cerchiamo di capire i nodi fondamentali
- Tali nodi sono quelli che determinano il corretto funzionamento del network
- A seconda del problema possiamo:
 - integrare l'analisi della struttura con dati da laboratorio
 - Verificare in laboratorio ipotesi suggerite dall'analisi della rete
- Lo scopo generale è quello di guidare gli esperimenti di laboratorio:
 - Anziché provare alla cieca ci concentriamo sulle proteine che sono strutturalmente importanti per il network

Analisi di un network: procedura

- Interattoma (prof. Laudanna):
 - 14000 proteine e 101000 interazioni
 - 80% human rimanente mouse e rat
 - Ottenuto da:
 - Dati in letteratura
 - Database: HPRD, MINT, DIP, IntAct, BIND, BioGRID
 - Secondo attuali stime copre circa il 70% dell'interattoma umano

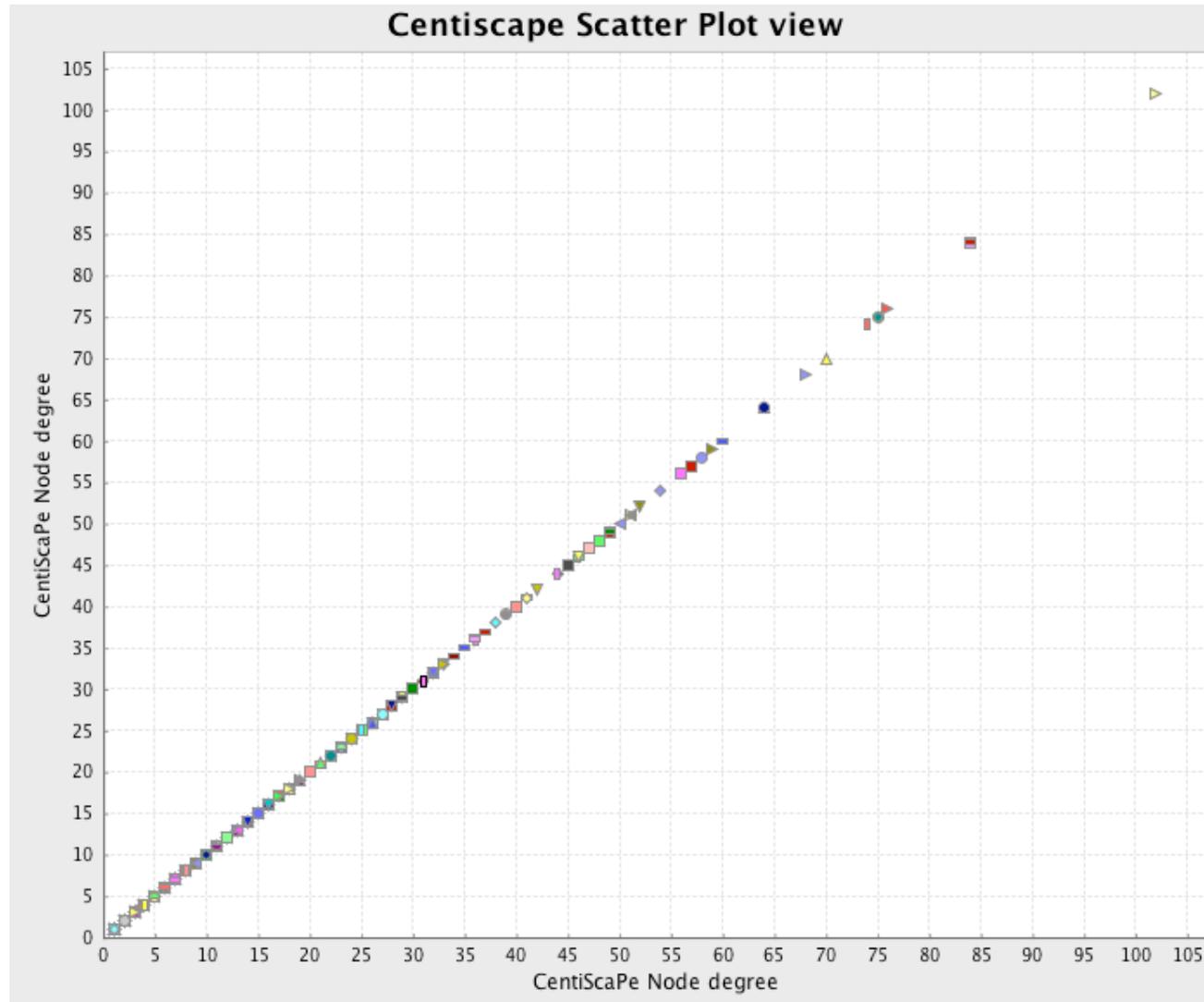
Analisi di un network: procedura

- Dall'interattoma umano si estrae il subnetwork relativo al fenomeno che ci interessa
- Si analizza il network da un punto di vista strutturale (topologico)
- Si integra l'analisi con i dati sperimentali
- Si determina un nuovo esperimento in laboratorio (es: verifica dell'attivazione di altre proteine)

Esempio: human kino-phosphatome

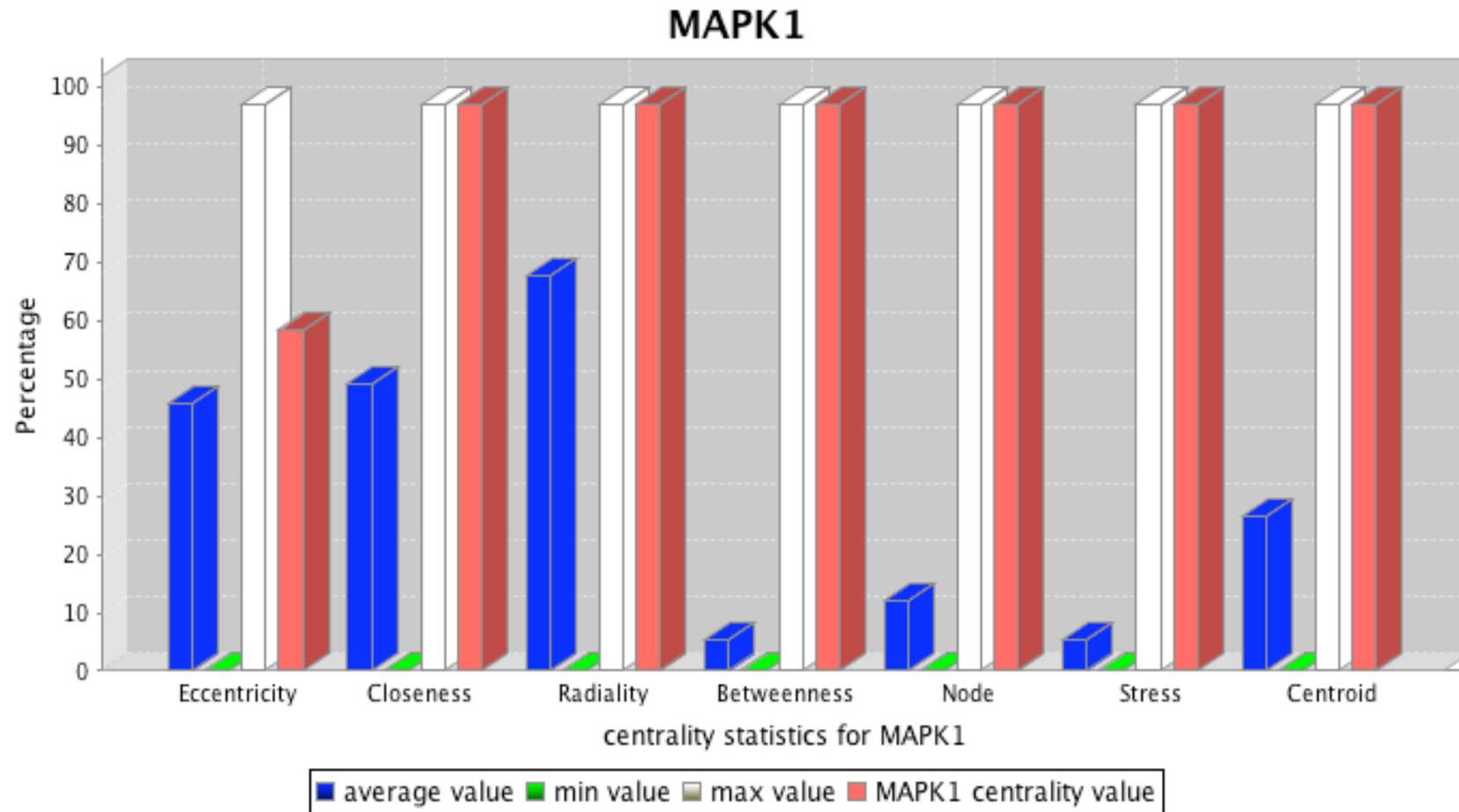
- Dall'interattome umano sono state estratte le interazioni conosciute tra le chinasi e le fosfatasi
- Il network risultante è costituito da 549 nodi e 3844 interazioni
- 406 kinasi e 143 fosfatasi

Esempio: human kino-phosphatome



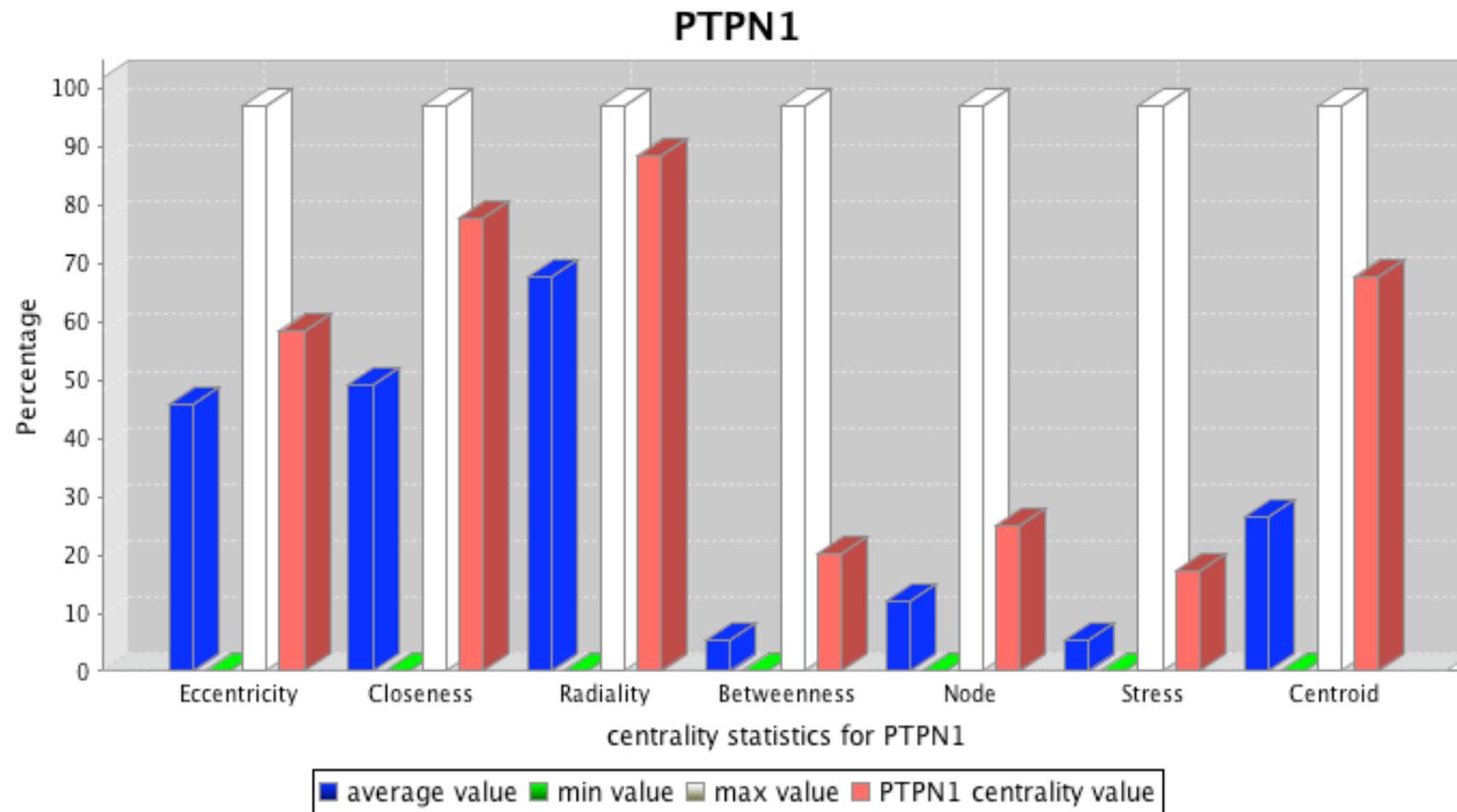
- Alcuni nodi hanno degree più alto (scale free)

Esempio: human kino-phosphatome



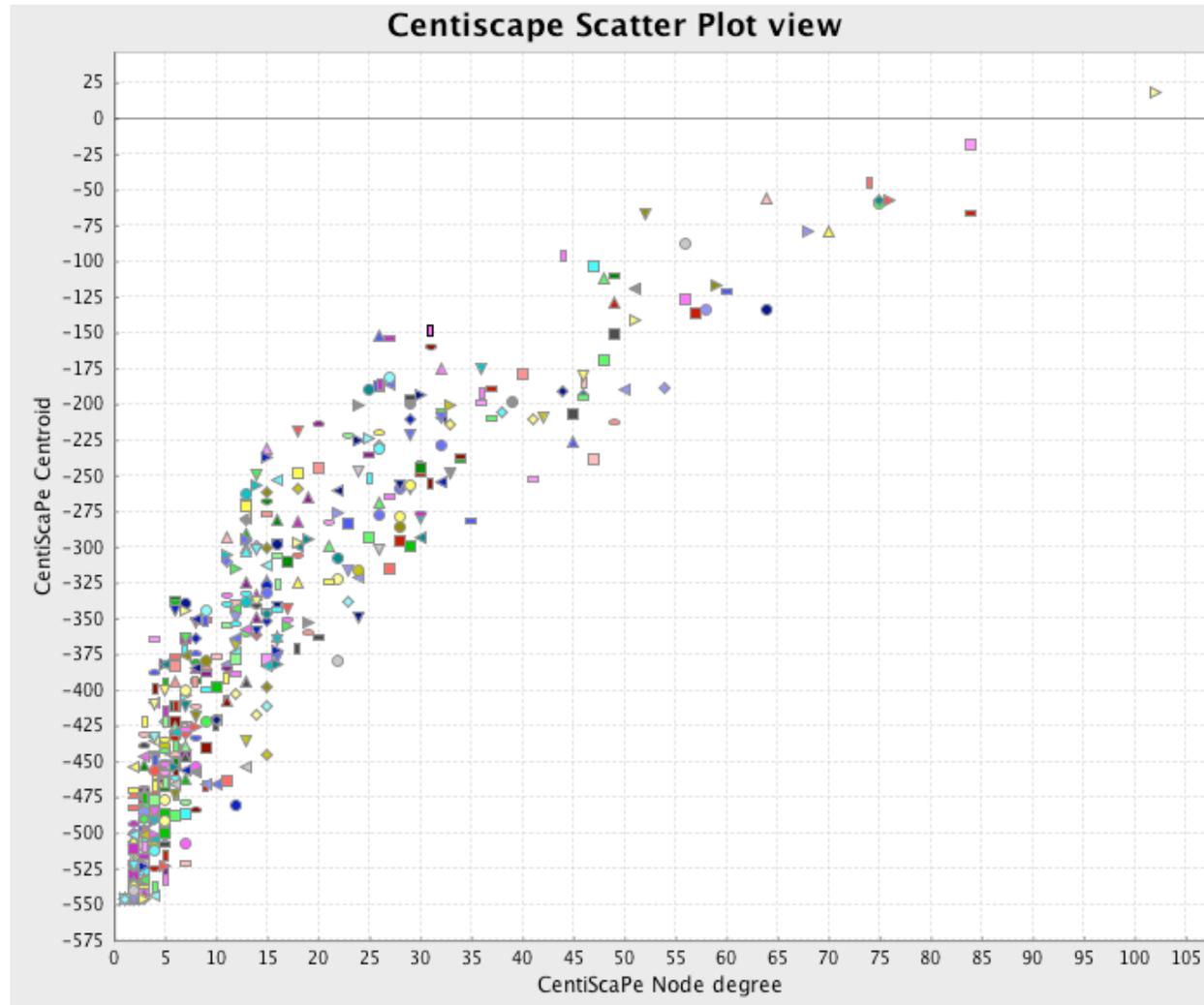
La kinasi mapk1 ha i valori più alti di centralità

Esempio: human kino-phosphatome



La ptpn1 ha i valori di centralità più alti tra le fosfatasi

Esempio: human kino-phosphatome



Degree
Vs
centroid

Mapk1 e ptpn1 si trovano in alto a destra

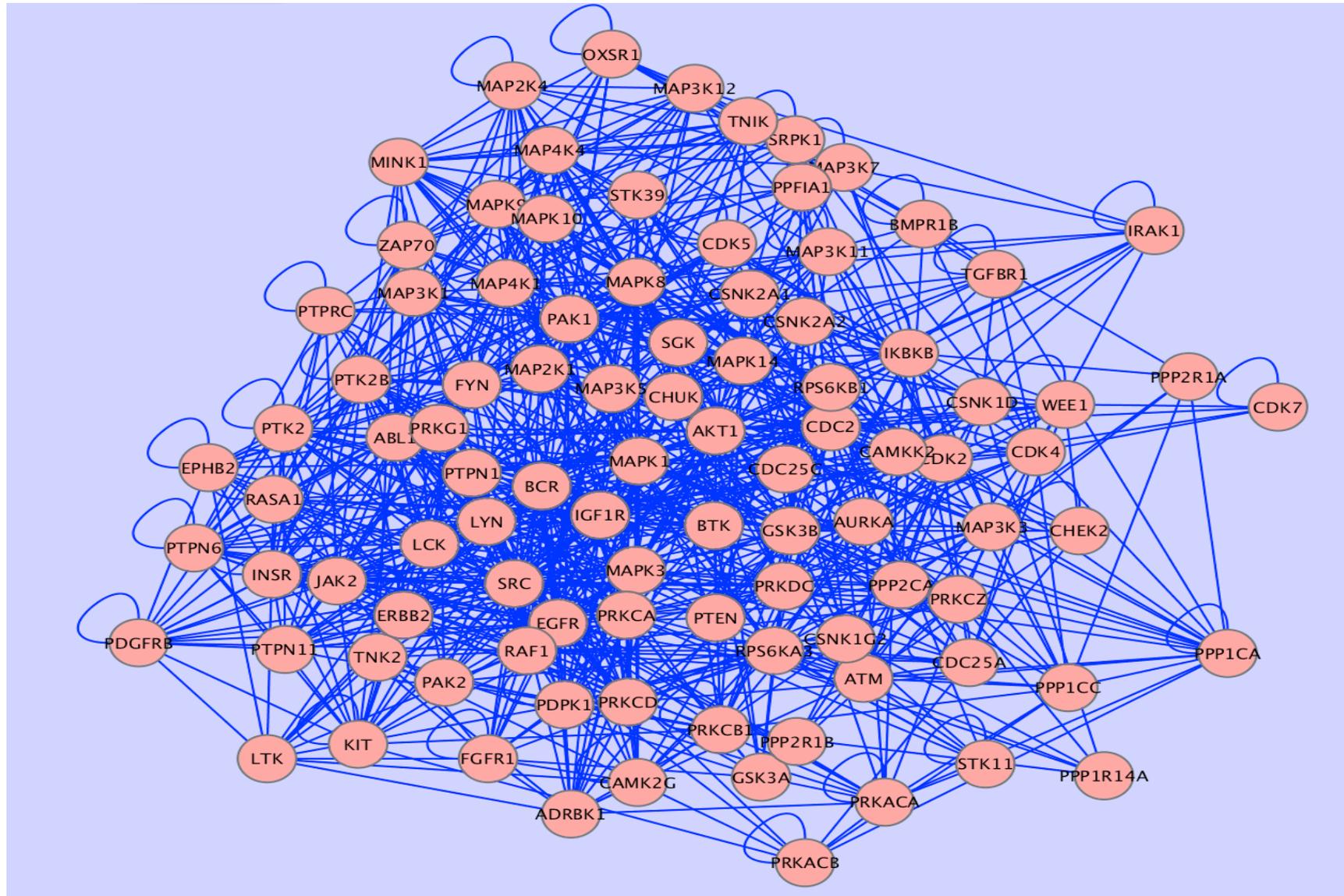
Esempio: human kino-phosphatome

- Mapk1 e ptpn1 evidentemente giocano un ruolo fondamentale nel network
- Costituiscono rispettivamente la chinasi e la fosfatasi più importanti

Esempio: human kino-phosphatome

- Estraiamo ora i nodi che hanno tutte le centralità sopra la media
- Otteniamo un subnetwork costituito da:
 - 97 nodi (83 kinasi and 15 fosfatasi)
 - 962 interazioni

Esempio: human kino-phosphatome



Esempio: human kino-phosphatome

- Le chinasi e le fosfatasi estratte giocano evidentemente un ruolo critico nella regolazione della fosforilazione nelle cellule umane
- Successivi esperimenti in laboratorio dimostrano che proteine con alte centralità presentano alti livelli di attivazione

Analisi di un network 1: comprendere i meccanismi di una malattia

- In alcune malattie mancano o sono inibite alcune proteine nel network
- Tale mancanza determina una diversa struttura della rete
- L'analisi permette di capire come si ristrutturata la rete in assenza di alcune proteine
- Dall'analisi possiamo capire quali proteine mancanti sono fondamentali e quali sono ininfluenti

Analisi di un network 2: comprendere i meccanismi di una malattia

- In alcune malattie alcune proteine sono sovraesprese
- L'analisi del network permette di individuare quali sono le proteine che possono determinare tale sovraespressione (es: alterazione nello shortest path)
- Possiamo agire sulle proteine che inibiscono la proteina sovraespressa

Analisi di un network 3: comprendere i meccanismi di una malattia

- Sappiamo che alcune proteine hanno livelli di attivazione alti
- Cerchiamo altre proteine che siano attivate
- Anziché provare alla cieca cerchiamo:
 - Le proteine vicine a quelle attivate
 - Le proteine con alti valori di centralità
- Verifichiamo sperimentalmente tali attivazioni

Analisi di un network 4: comprendere i meccanismi di una malattia

- Con le strumentazioni attuali siamo in grado di fotografare le proteine presenti in una cellula prelevata da un paziente
- Queste proteine danno luogo a un network
- Che differenza c'è con il network di un paziente sano?
- In un caso sono state trovate delle sottostrutture molto simili tra pazienti con la stessa malattia (leucemia linfatica cronica B), confrontando dati da laboratorio e da analisi topologica:
 - Le stesse proteine attivate
 - Le stesse proteine con i medesimi valori di centralità alti

Analisi del network 5: simulare l'utilizzo di un farmaco

- Utilizzare un farmaco può significare introdurre un nodo nella rete (con nuovi archi)
- Come variano i parametri della rete con il nuovo nodo?
- Possiamo comprendere quali cambiamenti il farmaco introduce nella rete
- Possiamo prevedere effetti collaterali

Analisi di un network 6: simulare l'utilizzo di un farmaco

- Un farmaco può servire a inibire una proteina
- Questo significa eliminare un nodo (e i suoi archi) dal network
- Come cambia l'analisi del network?
- Che cambiamenti porta alla rete?
- Quali effetti collaterali?

Problemi e prospettive

- Mancanza di dati (archi o direzione degli archi)
- Simulazione dinamica
- Applicazioni in farmacologia
- Medicina specifica per ogni paziente