

INSIdE nano - Online learning

Simone Romano¹

¹Università degli Studi di Salerno

Supervisori:

Prof. Roberto Tagliaferri

Dott. Angela Serra

Esame Bioinformatica - 2015/2016



- 1 **Introduzione**
- 2 **Approcci utilizzati**
- 3 **Risultati**
- 4 **Conclusioni**



Outline

1 **Introduzione**

- Dataset
- Approcci utilizzati

2 Approcci utilizzati

3 Risultati

4 Conclusioni

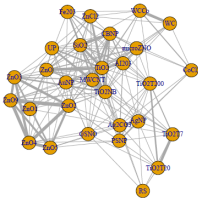


INSIdE nano

Il dataset analizzato contiene informazioni sulle relazioni tra:

- nanomaterials
- drugs
- chemicals
- diseases

Le relazioni sono valutate in base all'effetto di ciascuna entità sull'**espressione genica**



Costruzione network

Informazioni iniziali:

- Per ogni entità (nanomateriale, farmaco, prodotto chimico e malattia) si ha a disposizione una **lista di geni correlati**

In particolare:

- chemical e disease: **set di geni associati**
- drug e nanomaterial: **lista ordinata (per espressione) di geni**

Per arrivare alla network è stata calcolata la similarità tra tutte le coppie di entità disponibili



Calcolo distanze

Distanza tra **insiemi di geni**:

Jaccard Index

Dati due insiemi A e B l'indice di Jaccard è dato dalla cardinalità dell'intersezione diviso la cardinalità dell'unione

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

Distanza tra **liste ordinate di geni**:

Kendall Tau

$$K(T_1, T_2) = |\{(i, j) : i < j, (T_1(i) < T_1(j) \wedge T_2(i) > T_2(j)) \vee (T_1(i) > T_1(j) \wedge T_2(i) < T_2(j))\}|$$

Distanza tra **liste ordinate ed insiemi di geni**:

- Gene Set Enrichment Analysis - GSEA



CLR

Partendo dalla matrice di similarità \mathbf{M} così ottenuta, si è arrivati alla network finale \mathbf{W} applicando il CLR come segue:

CLR

Per ogni m_{ij} è stato calcolato il peso dell'arco w_{ij} come segue:

$$w_{ij} = \sqrt{(z_i^2 + z_j^2)} \cdot \text{sign}(M(i, j))$$

dove z_i^2 e z_j^2 sono così definiti:

$$\begin{cases} M(i, j) > 0, s = +, z_i = \max\{0, \frac{M(i, j) - \mu_i^s}{\sigma_i^s}\} \\ M(i, j) < 0, s = -, z_i = \min\{0, \frac{M(i, j) - \mu_i^s}{\sigma_i^s}\} \end{cases}$$



Network finale

La network finale:

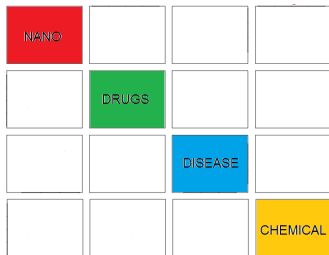
NANO			
	DRUGS		
		DISEASE	
			CHEMICAL

I valori sono stati normalizzati in $[-1;1]$ per poter essere confrontati



Problematica affrontata

La network descritta può variare con l'**aggiunta/rimozione** di elementi

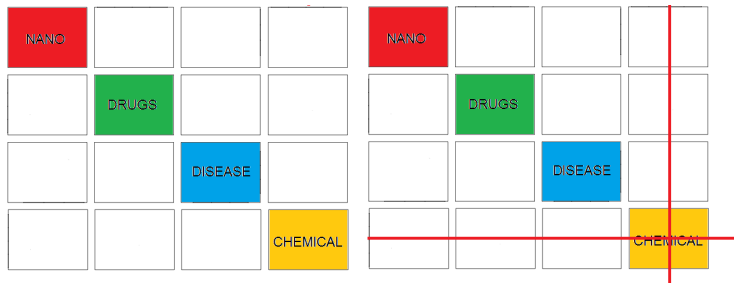


Come cambia la struttura della network dopo aver apportato le modifiche?



Problematica affrontata

La network descritta può variare con l'**aggiunta/rimozione** di elementi

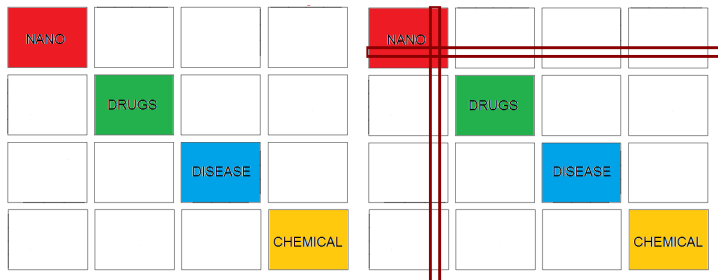


Come cambia la struttura della network dopo aver apportato le modifiche?



Problematica affrontata

La network descritta può variare con l'**aggiunta/rimozione** di elementi



Come cambia la struttura della network dopo aver apportato le modifiche?



Outline

1 Introduzione

2 Approcci utilizzati

3 Risultati

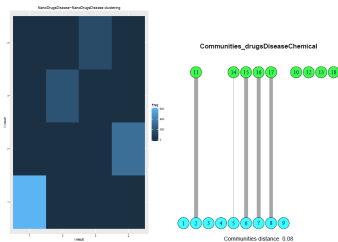
4 Conclusioni



Comparazione network

Per valutare la differenza tra le network ottenute sono state fatte le seguenti valutazioni:

- **clustering** delle varie sottomatrici
 - nano-nano,nano-drugs,...,chemical-chemical
- calcolo della **degree centrality** su tutte le subnetwork
- ricerca delle **communities** basato sulla modularità



	degreeNanoNano_distance		
KendallTau Distance	1		
DEGREE (no nano)	NAME	DEGREE (with nano)	NAME
15	AuNP	15	AuNP
15	TiO2	15	TiO2
12	Ag2CO3	12	Ag2CO3
12	ZnO2	12	ZnO2
11	AuNP	11	AuNP

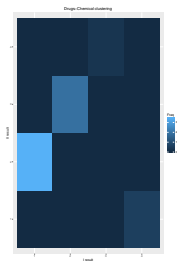


Clustering

Per ogni subnetwork è stato applicato il **kmeans** con distanza euclidea per:

- Network iniziale
- Network modificata

È stato effettuato il plot delle confusion matrix per studiare la stabilità dei cluster



Degree centrality

Per ogni subnetwork è stata calcolata la **degree centrality** di ogni nodo

- Ordinamento liste per degree decrescente
- calcolo Kendall Tau

	degreeNanoNano_distance		
KendallTau Distance	0.862068965517241		
DEGREE (no chemical)	NAME	DEGREE (with chemical)	NAME
24	TiO2	23	TiO2
23	TiO2NB	22	TiO2NB
22	MWCNT	21	MWCNT
18	AuNP	19	SiO2
18	SiO2	18	AgNP
18	ZnO2	17	AuNP
17	AgNP	16	Ag2CO3
17	Ag2CO3	16	Al2O3
17	Al2O3	16	CBNP
17	ZnO	15	UP
16	CBNP	15	ZnO
16	UP	15	ZnO2



Communities

Community basate sulla **modularità** [1]

Modularity

Data una network ed una sua suddivisione in community, tale divisione è buona quando si hanno

- molti archi all'interno di una community
- pochi archi tra le community

Output:

- calcolo similarità tra community utilizzando la metrica **VI** (Variation of information) [2]
- grafo bipartito che mostra le community per le varie coppie di subnetwork

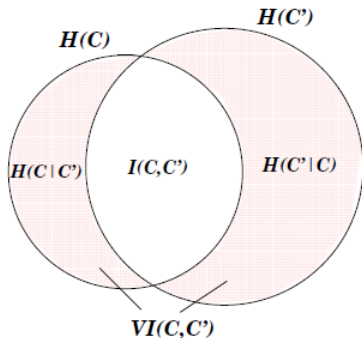


Communities

Variation of information

Dati due clustering C , C' , la variazione di informazione è data da:

$$VI(C, C') = [H(C) - I(C, C')] + [H(C') - I(C, C')]$$



Outline

1 Introduzione

2 Approcci utilizzati

3 Risultati

- Network senza chemical
- Network senza nano

4 Conclusioni



1 Introduzione

2 Approcci utilizzati

3 Risultati

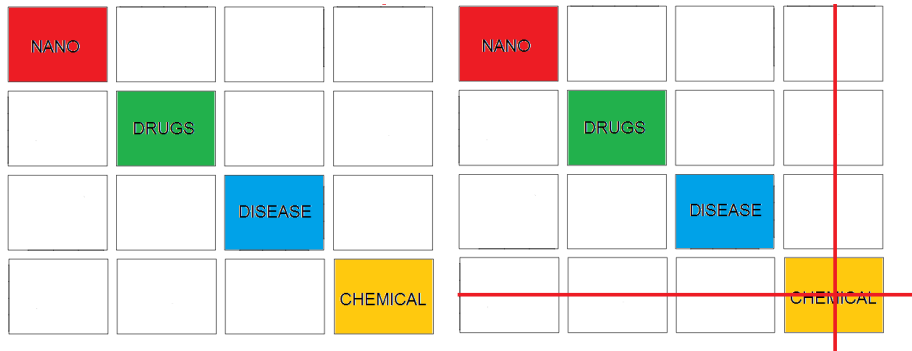
- Network senza chemical
- Network senza nano

4 Conclusioni



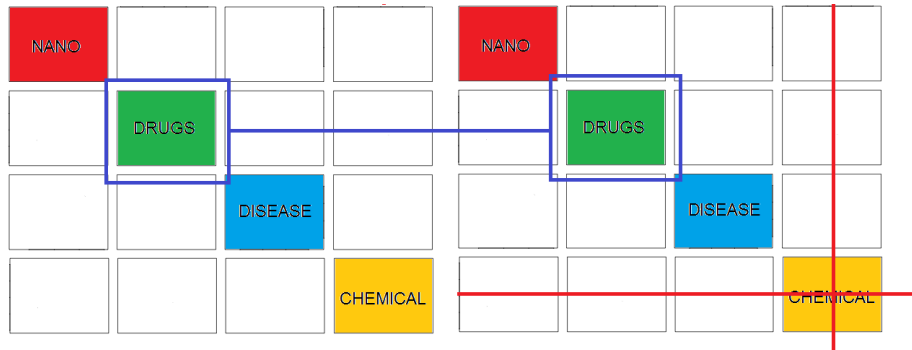
Clustering

Confronto subnetwork **con/senza chemical**:



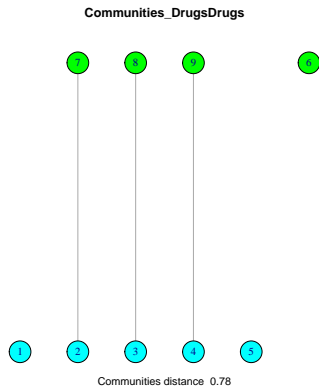
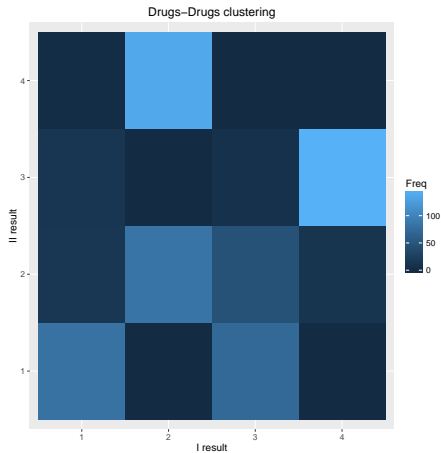
Clustering

Confronto subnetwork **con/senza chemical**:



Clustering

Confronto subnetwork **con/senza chemical**:



Clustering

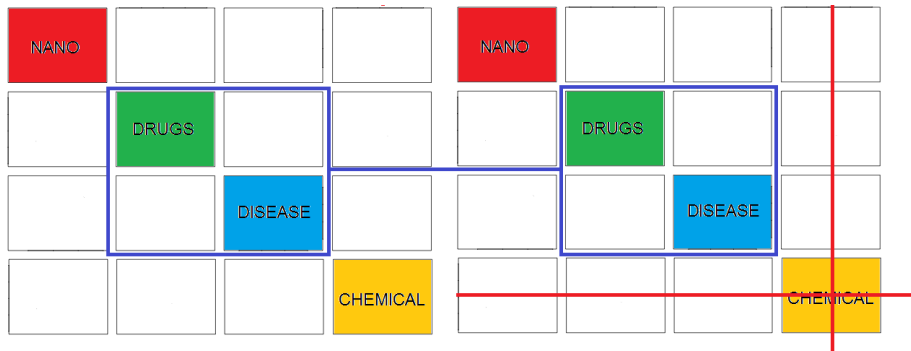
Confronto subnetwork **con/senza chemical**:

	degreeDrugsDrugs_distance		
KendallTau Distance	0.995858160536003		
DEGREE (no chemical)	NAME	DEGREE (with chemical)	NAME
519	nomifensine	492	reserpine
515	cefazolin	490	articaïne
503	fipexide	489	azapropazone
503	reserpine	485	gliquidone
502	ornidazole	477	cefoperazone
499	metampicillin	473	ketorolac
498	oxytetracycline	470	alverine
496	alverine	466	ceftazidime
496	articaïne	466	methotrexate
493	betamethasone	465	proxiphylline
493	urapidil	463	cefotetan
492	methotrexate	462	artemisinin
492	nicergoline	462	oxymetazoline
490	oxymetazoline	461	oxytetracycline



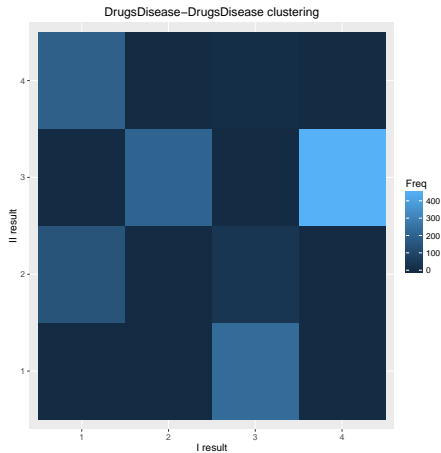
Clustering

Confronto subnetwork **con/senza chemical**:



Clustering

Confronto subnetwork **con/senza chemical**:



Communities_DrugsDisease



Communities distance 0.15



Clustering

Confronto subnetwork con/senza chemical:

	degreeDrugsDisease_distance		
KendallTau Distance	0.997811665179753		
DEGREE (no chemical)	NAME	DEGREE (with chemical)	NAME
1027	Lung Injury	977	Neurobehavioral Manifestations
1021	Memory Disorders	974	Oligospermia
1020	Anorexia	968	Anorexia
1020	Carcinoma, Squamous Cell	964	Seizures
1018	Neurobehavioral Manifestations	963	Exanthema
1018	Oligospermia	962	Infertility, Male
1017	Liver Diseases	960	Atrophy
1016	Seizures	960	Hemorrhage
1015	Micronuclei, Chromosome-Defective	960	Lung Injury
1015	Bradycardia	959	Cardiovascular Diseases
1014	Adenocarcinoma	959	Liver Diseases
1013	Hemorrhage	959	Micronuclei, Chromosome-Defective
1013	Liver Cirrhosis, Experimental	959	Pneumonia
1013	Proteinuria	959	Bradycardia
1012	Neural Tube Defects	959	Anxiety Disorders
1012	Skin Neoplasms	958	Cardiomegaly
1011	Carcinoma, Hepatocellular	958	Muscular Diseases
1011	Cognition Disorders	957	Skin Neoplasms
1011	Ovarian Neoplasms	955	Carcinoma, Hepatocellular
1010	Kidney Diseases	954	Abnormalities, Drug-Induced
1010	Cardiomegaly	954	Adenocarcinoma

1 Introduzione

2 Approcci utilizzati

3 Risultati

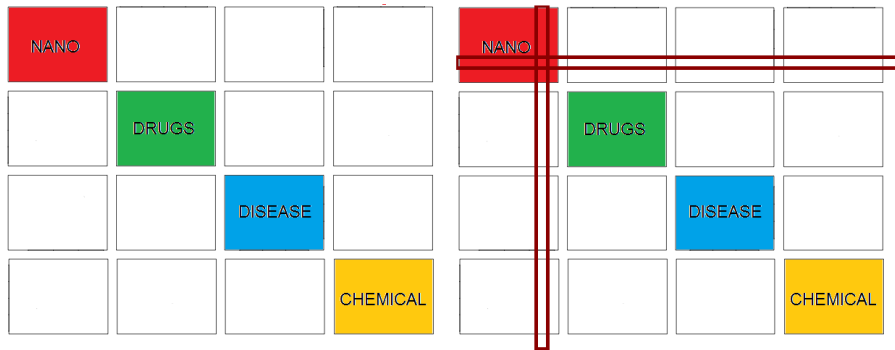
- Network senza chemical
- Network senza nano

4 Conclusioni



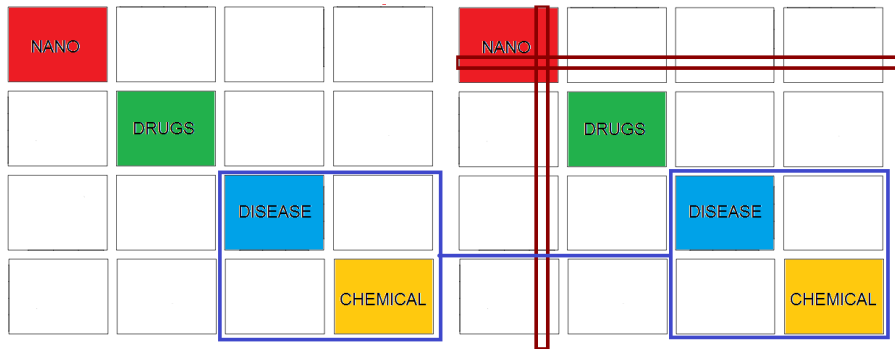
Clustering

Immagini non corrette Confronto subnetwork **con/senza nano**:



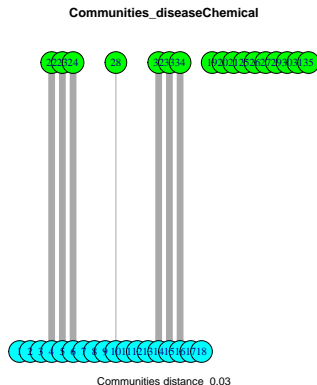
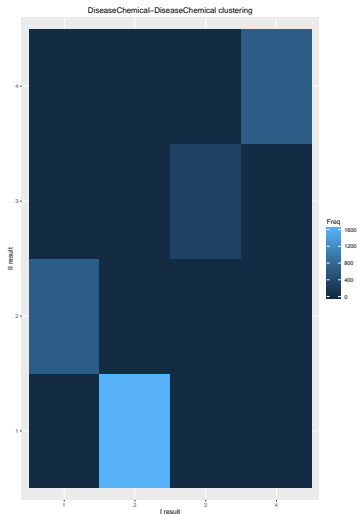
Clustering

Immagini non corrette Confronto subnetwork **con/senza nano**:



Clustering

Immagini non corrette Confronto subnetwork **con/senza nano**:



Clustering

Immagini non corrette Confronto subnetwork **con/senza nano**:

	degreeDiseaseChemical_distance		
KendallTau Distance	0.999514360195574		
DEGREE (no nano)	NAME	DEGREE (with nano)	NAME
2559	cirsiliol	2563	cirsiliol
2531	BHTOH-QM	2532	BHTOH-QM
2520	sulglicotide	2526	sulglicotide
2495	moxifloxacin	2498	fluanisone
2494	fluanisone	2497	moxifloxacin
2488	Nimustine	2490	Nimustine
2480	N2 Dental Cement	2485	Pulmonary Surfactant
2480	Pulmonary Surfactant	2481	N2 Dental Cement
2467	SB 216763	2468	Ly-364947
2460	Ly-364947	2468	SB 216763
2450	Hexadimethrine Bromi	2452	Hexadimethrine Bromi
2432	Hydrocarbons, Bromin	2436	4-aminophenol
2431	RV09 compound	2435	Hydrocarbons, Bromin
2430	4-aminophenol	2435	RV09 compound
2417	3-O-tert-butoxycarbo	2423	3-O-tert-butoxycarbo
2411	isophthalate	2415	isophthalate
2411	propamocarb	2414	kamebakaurin
2410	beta-Aminoethyl Isot	2412	beta-Aminoethyl Isot
2410	kamebakaurin	2411	propamocarb
2405	N,N-dimethylaniline	2405	N,N-dimethylaniline
2400	gamma-oryzanol	2404	gamma-oryzanol
2399	Perchlorates	2403	pirenperone
2394	latoconazole	2402	Perchlorates



Outline

1 Introduzione

2 Approcci utilizzati

3 Risultati

4 Conclusioni



Discussioni

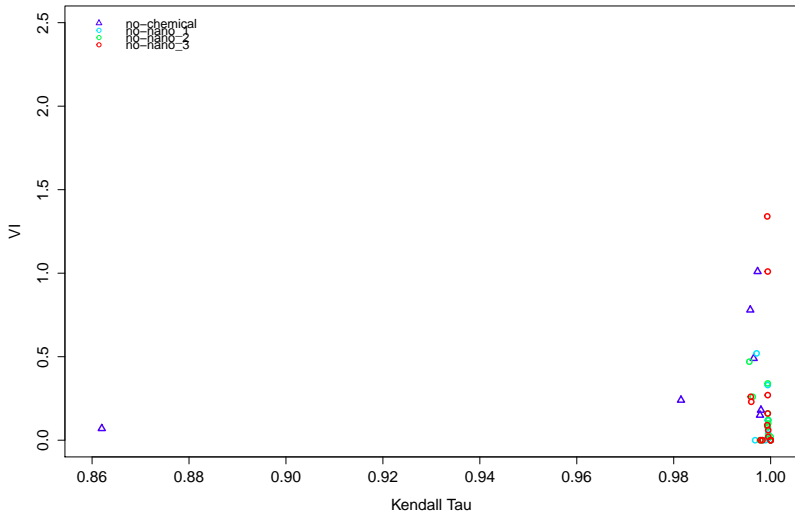
Dal lavoro svolto si evince che:

- **Cluster**: gli elementi vengono raggruppati in maniera significativamente differente quando si rimuove una intera categoria
- **Community** e **Degree centrality**: rimangono abbastanza stabili sia quando si rimuove una intera categoria sia quando si rimuovono alcuni elementi in una categoria



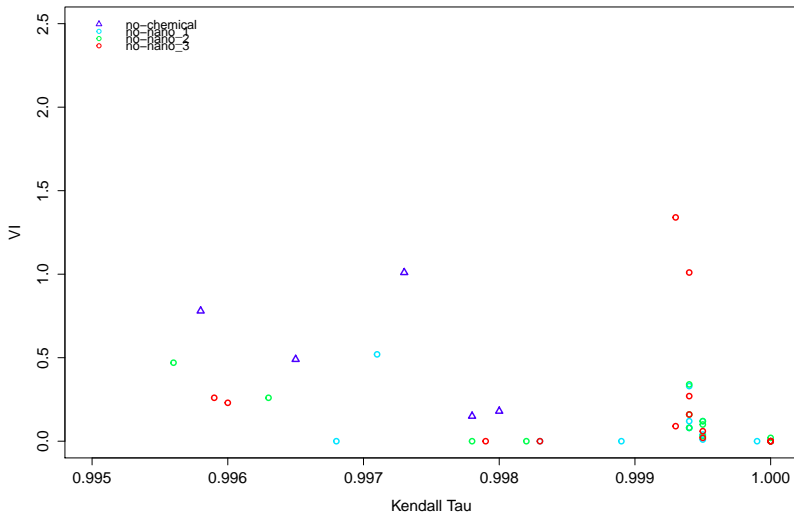
Discussioni - 1

Results



Discussioni - 2

Results



Riferimenti I



Aaron Clauset et al.

Finding community structure in very large networks.

<http://arxiv.org/pdf/cond-mat/0408187v2.pdf>



Marina Meilă

Comparing clusterings.

<http://www.stat.washington.edu/mmp/Papers/compare-colt.pdf>



Grazie per l'attenzione

