



**Consiglio Nazionale delle Ricerche  
Istituto di Fisiologia Clinica  
Area della Ricerca S.Cataldo Pisa  
Sez. Epidemiologia e Ricerca sui Servizi Sanitari**

## **Technical Report / Apr. 2012**

**BINET – Analisi di dati sanitari mediante SNA**  
**Validità degli indicatori statistici ottenuti rispetto a metodologie tradizionali**

**Versione 1.0  
Aprile 2012**

**Autori  
S. Pieroni, F. Mariani, M. Franchini, S. Molinaro**

## INDICE dei contenuti

Introduzione.....	3
Materiali e metodi .....	3
Il caso studio .....	4
Misure quantitative della rete .....	8
Confronto con la metodologia tradizionale .....	9

## Introduzione

Il progetto BINET si propone il disegno e la progettazione di un framework di Business Intelligence (BI) per il supporto alle decisioni basate sulle Reti Sociali.

Una delle fasi di progetto è stata dedicata alla sperimentazione delle metodologie proposte al fine di poter effettuare una validazione scientifica della loro efficacia ed efficienza in situazioni reali. La valutazione è avvenuta utilizzando i dati dei servizi sanitari e l'esperienza di un gruppo di ricerca afferente l'Istituto di Fisiologia Clinica del CNR di Pisa, uno dei partner di progetto..

Nel presente rapporto, a seguito della sperimentazione del framework applicativo messo a disposizione da BINET su basi di dati sanitari, si effettua un confronto della validità e accuratezza degli indicatori statistici ottenuti rispetto a metodologie tradizionali.

## Materiali e metodi

Relativamente all'anno 2007, sono stati caricati nel database dell'applicazione i dati dei cittadini residenti nei nove comuni della Zona Pisana mostrati in Figura 1.

Si tratta di 198.829 soggetti, che hanno generato 27.955 ricoveri, 1.765.744 prescrizioni farmaceutiche e 2.200.000 prestazioni di specialistica ambulatoriale.



Figura 1: Popolazione di riferimento

Il modello dati è rappresentato in Figura 2. Il grafo mostrato indica che ogni soggetto della rete è collegato:

- ad una o più classi di farmaco (classe ATC4, il sottogruppo chimico-terapeutico farmacologico della classificazione ATC Anatomical Therapeutic Chemical Classification System) sulla base delle prescrizioni ricevute attraverso il suo Medico di Medicina Generale (MMG)
- ad uno o più gruppi di prestazioni di specialistica ambulatoriale
- ad una o più diagnosi (principale e secondaria) ricevute durante ricoveri
- ad uno o più grandi gruppi di patologia attribuita secondo criteri di inclusione che prendono in esame Esenzioni per patologia, DRG e diagnosi effettuate durante i ricoveri, ATC e DDD dei farmaci assunti e branche specialistiche delle prestazioni ricevute

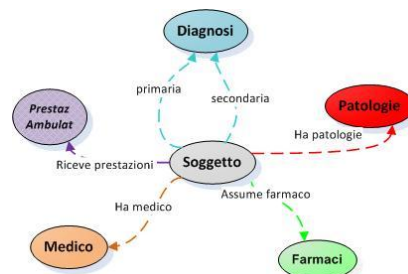


Figura 2: Modello Dati

La rete viene analizzata secondo due approcci:

- la link analysis (analisi di legame), un approccio descrittivo di esplorazione dei dati, che aiuta ad identificare le relazioni tra i nodi della rete, attraverso regole di associazione rilevanti da un punto di vista statistico. Tali regole sono implementate dagli algoritmi di ranking disponibili attraverso la piattaforma BINET, come il

pageRank che consiste nell'inferire l'importanza di un nodo basandosi sulla struttura topologica del grafo, analizzando gli archi uscenti ed entranti su ogni nodo.

- Gli indici di centralità che permettono di definire il posizionamento di un nodo nel proprio network in termini puramente relazionali. La centralità di un nodo può essere espressa almeno in tre modi diversi, ovvero basandola sul suo grado, sulla sua interposizione, ossia sul suo essere intermedio tra ogni nodo ed ogni altro, ed infine sulla sua vicinanza o prossimità a qualsiasi altro nodo del grafo e quindi sull'efficienza con cui esso può raggiungere tutti gli altri nodi della rete.

## Il caso studio

In figura 3 è riportato il grafo di una rete di 60 soggetti neoplastici affetti anche da diabete, da problemi cardiovascolari e da broncopneumopatie, rappresentata da un grafo a layout circolare, ottenuto attraverso la piattaforma di progetto. La rete è costituita di 217 nodi e 577 archi.

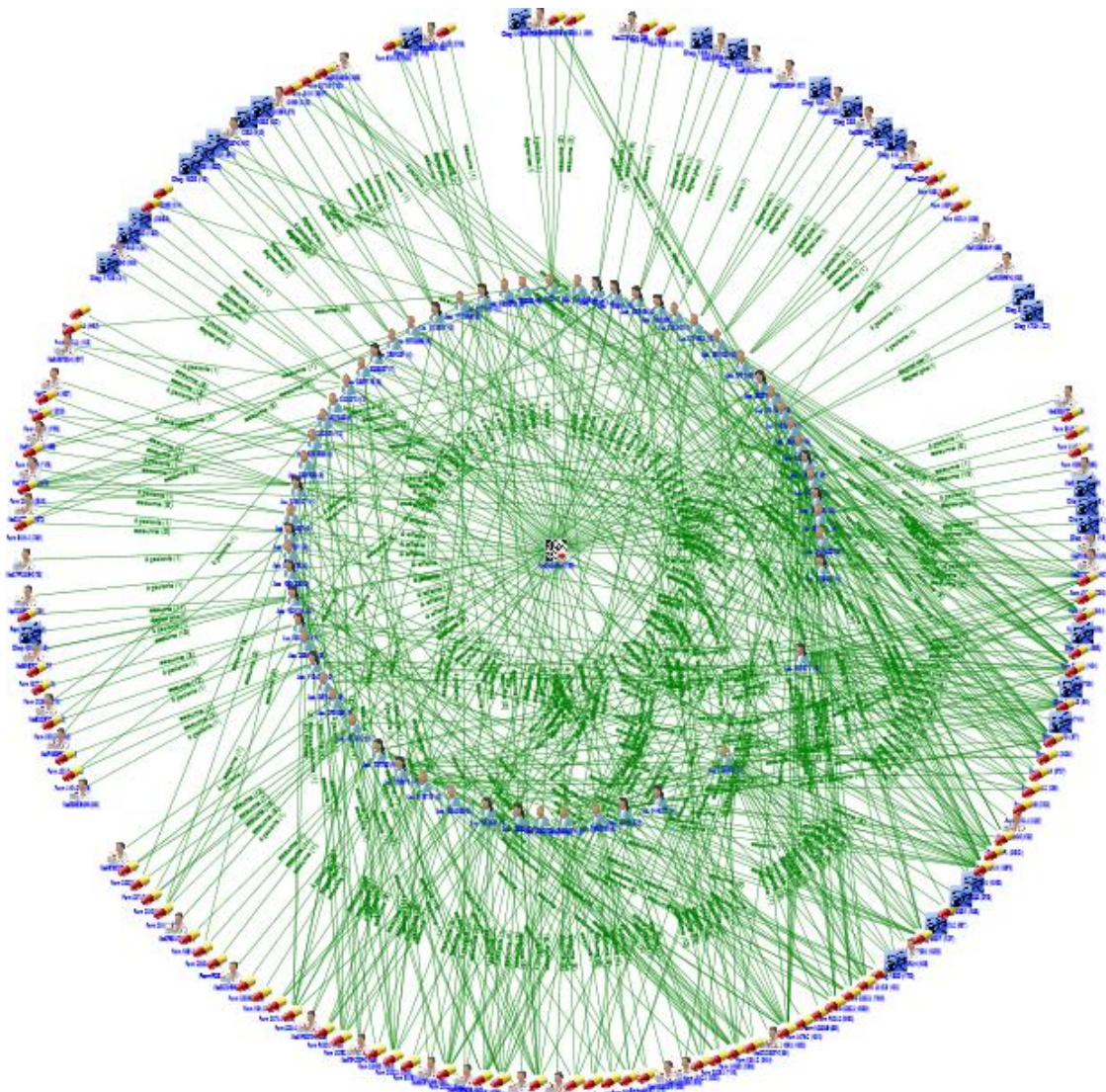


Figura 3: La rete in studio, layout circolare

L'iconografia del grafo mostra che i 60 soggetti, visibili sull'anello centrale sono trattati prevalentemente in modo farmacologico (classi di farmaci visibili sull'anello esterno), sono infatti utilizzate 82 diverse classi di farmaco.

Utilizziamo gli indici di centralità per descrivere quali sono i farmaci più utilizzati, in particolare l'**Eigenvector Centrality**, che misura l'importanza dei nodi nella rete assegnando uno score relativo a tutti i nodi basandosi sul principio che le connessioni agli high scoring nodes contribuiscono di più allo score del nodo in questione che le stesse connessioni ai low scoring nodes.

In altri termini, un farmaco che viene dato in associazione con un altro farmaco molto utilizzato per la terapia di questi soggetti, prende più peso che un farmaco dato in associazione ad un farmaco poco utilizzato. Quindi i farmaci con più alto valore di Eigenvector Centrality sono i farmaci che vengono dati più frequentemente in associazione fra loro.

Il calcolo di questo indice ci fornisce il risultato mostrato in figura 4:

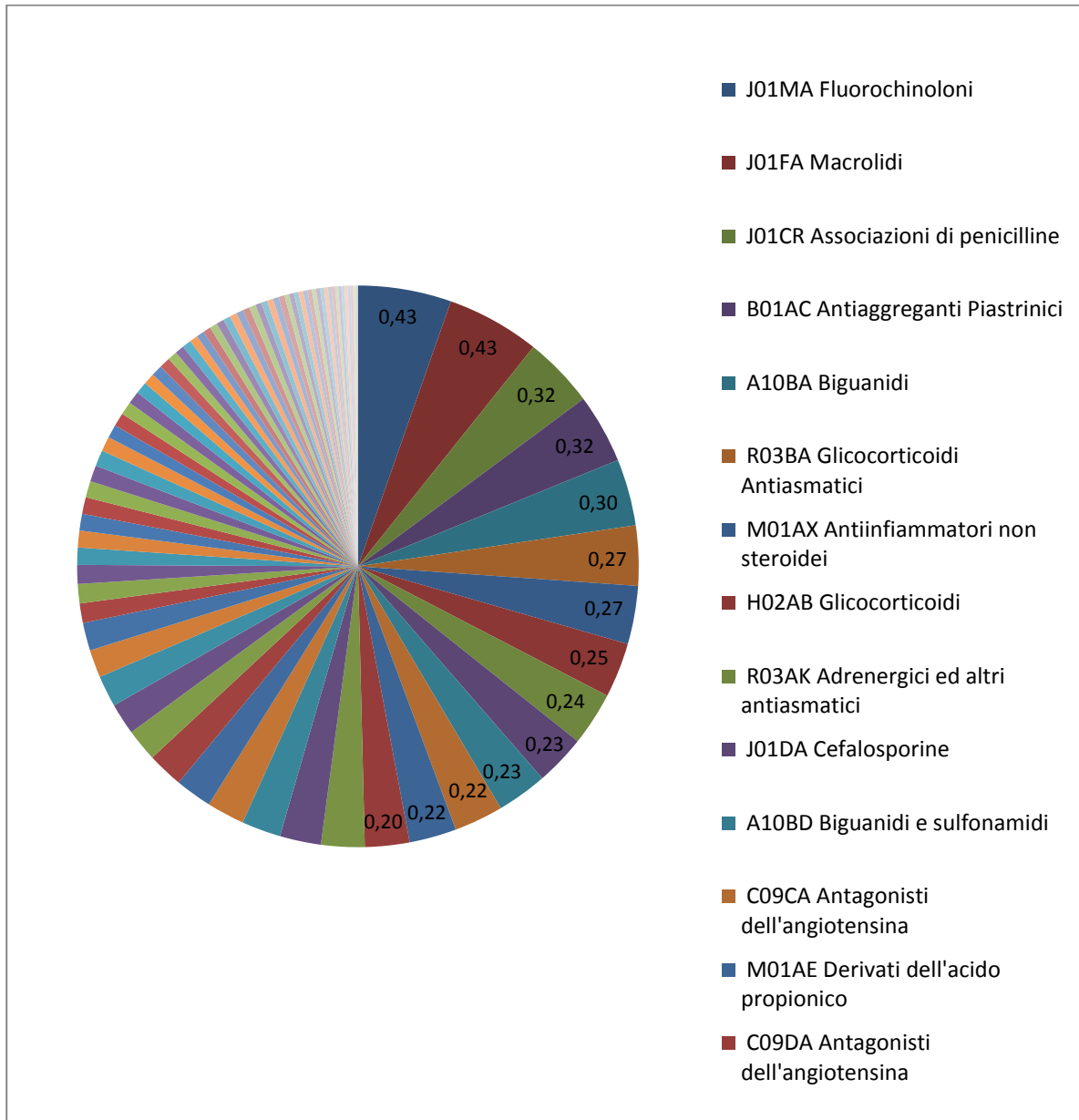


Figura 4: l'indice eigenvector dei farmaci

Vengono somministrati innanzitutto antibiotici ed antibatterici, seguiti da penicilline (farmaci legati evidentemente ai problemi delle infezioni batteriche che sostengono le broncopneumopatie), seguiti dagli antiaggreganti piastrinici (farmaci utilizzati principalmente per problemi cardiovascolari) e biguanidi (farmaci tipicamente usati per la cura del diabete). I 14 farmaci mostrati nel grafico coprono circa la metà delle somministrazioni.

Utilizzando il layout a campo di forze si ottiene una rappresentazione coerente con questo risultato, come mostrato in figura 5:

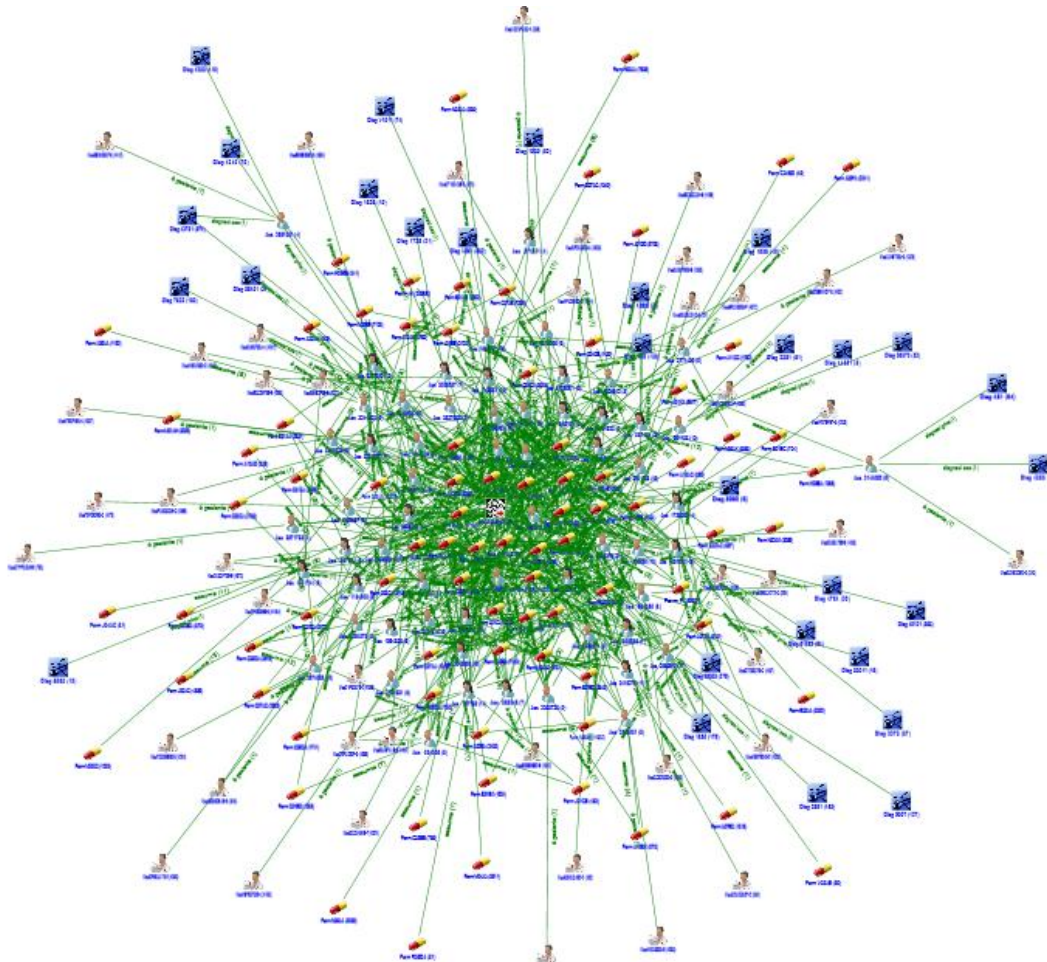



Figura 5: La rete in studio, layout a campo di forze

I farmaci rappresentati centralmente sono quelli con indice eigenvector più elevato.

La rappresentazione grafica sembra indicare anche il fatto che l'ospedalizzazione dei pazienti non è molto significativa.

Le icone celesti  rappresentano le diagnosi principali e secondarie ricevute dai pazienti durante ricovero, non rappresentano dunque il ricovero in se.

Se eliminiamo dal grafo tutti i nodi tranne assistiti e diagnosi e lasciamo i soli archi che collegano gli assistiti alle diagnosi otteniamo la rappresentazione stilizzata visibile in figura 6, dove i nodi in grigio sono i pazienti ed i nodi in blu le diagnosi. Da questa rappresentazione si evince che 17 su 60 sono i pazienti che hanno ricevuto un ricovero e durante tali ricoveri sono state formulate 30 diagnosi principali o secondarie diverse:

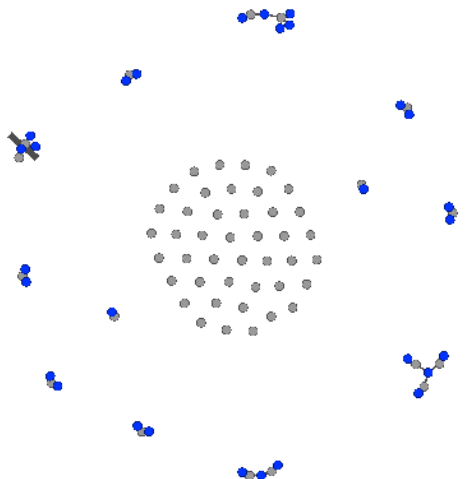


Figura 6: sottorete delle diagnosi

Dalla sottorete delle diagnosi si evidenziano alcuni sottografi:



la chemioterapia (nodo blu centrale) è comune a 3 pazienti aventi ognuno una diversa diagnosi di neoplasia (tumori maligni del colon, tumori maligni della trachea dei bronchi e dei polmoni, anamnesi personale di tumore maligno).



L’aritmia (nodo blu centrale) è comune a due pazienti a cui è stato diagnosticato anche scompenso cardiaco e malattia dell’endocardio.



Il diabete mellito (nodo blu centrale) è comune a due pazienti. Ad uno di questi è stato diagnosticato anche malattie delle valvole mitrale e aortica, mentre all’altro tre ulteriori diagnosi: tumori benigni delle altre ghiandole endocrine e strutture connesse, bronchite cronica ed altre malattie del polmone.

L’eigenvector centrality delle diagnosi principali e secondarie ricevute dai pazienti durante ricovero è

Tumori maligni della vescica	0,0548
Chemioterapia	0,0463
Diabete Mellito	0,0463
Altre malattie del polmone	0,0268
Tumori benigni di altre ghiandole endocrine	0,0268
Bronchite cronica	0,0268
Tumori maligni della prostata	0,0233
Tumori di comport incerto apparati digerente e respiratorio	0,0233
Aritmie cardiache	0,0219
Altre e non specificate anemie	0,0218
Altre patologie dell uretra e del sistema urinario	0,0218
Traumatismi intracranici di altra e non specificata natura	0,0209
Malattie delle valvole mitrale e aortica	0,0194
Anamnesi personale di tumore maligno	0,0181
Altre malattie vascolari periferiche	0,0181
Diabete Mellito	0,0181
Tumori maligni del colon	0,0160
Melanoma maligno della cute	0,0151
Altre malattie dell endocardio	0,0141
Psoriasi e affezioni similari	0,0135
Tumori maligni del colon	0,0132
Tumori maligni del colon	0,0132
Altre malattie delle vie respiratorie superiori	0,0126
Tumori maligni della trachea, dei bronchi e dei polmoni	0,0122
Emorragia cerebrale	0,0084
Postumi delle malattie cerebrovascolari	0,0084
Insufficienza cardiaca (scompenso cardiaco)	0,0078

Tabella 1: eigenvector centrality delle diagnosi

Si può notare come i valori di questo indice siano sensibilmente più bassi degli stessi valori applicati alla classi di farmaci somministrati ai pazienti della rete. Questo è evidente data la frequenza molto più elevata della ricettazione rispetto al ricovero (ordine di grandezza 0,5 vs 0,05).

## Misure quantitative della rete

La rete in studio è una rete poco densa, il grado di densità è rappresentato da numero di connessioni reali diviso per il numero di tutte le possibili connessioni (è quindi un numero sempre compreso tra 0 e 1). La rete ha densità 0,025.

Spesso una misura di coesione più intuitiva rispetto alla densità è il grado medio dei nodi della rete che indica il numero medio dei legami per ogni singolo nodo (Average Degree). L'average degree è 5,318

Il coefficiente di clustering medio, ovvero la probabilità media che 2 nodi connessi allo stesso nodo siano anche connessi fra loro è 0

Grado di densità	0,025
Average Degree	5,318
Coefficiente clustering medio	0

Tabella 2: Misure quantitative

Questi valori indicano una **rete ad invarianza di scala**, molti nodi con pochi link e pochi nodi con molti link: la distribuzione dei nodi in funzione del loro grado segue una legge di potenza esponenziale verso il basso (Figura 7)

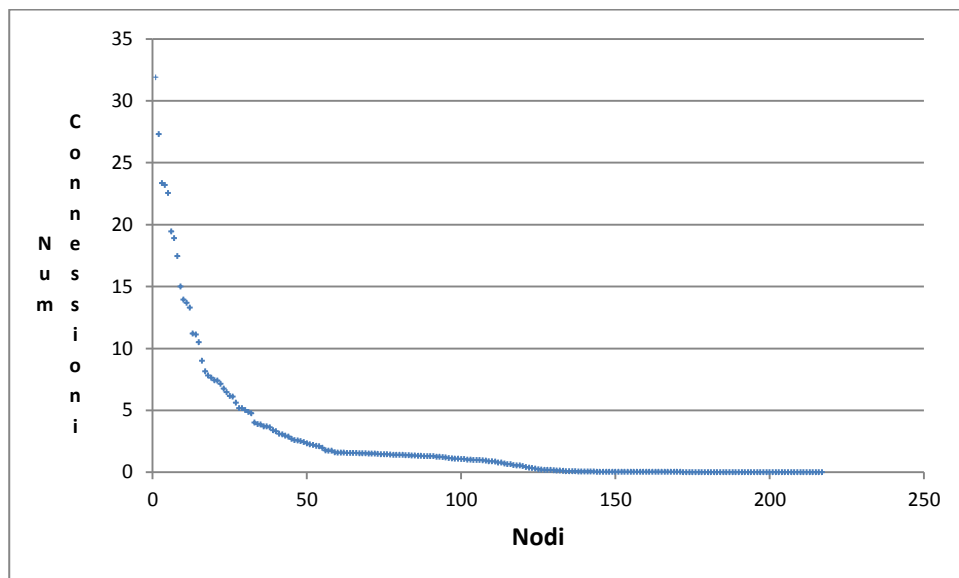


Figura 7: Distribuzione degree

In una rete ad invarianza di scala quando un nodo deve stabilire un nuovo collegamento, preferisce farlo verso un nodo che ne ha già molti, portando questi ad una crescita esponenziale con l'aumentare del numero dei collegamenti della rete. Questo concetto, nel contesto di questa rete sanitaria, può essere letto in questo modo: i pazienti che assumono molti farmaci tendono ad assumerne sempre di più mentre quelli che ne assumono meno sempre meno (in proporzione). In questo tipo di rete i *pochi* nodi con degree *molto* alto vengono detti hub, con pochi singoli attacchi mirati agli hub è possibile abbattere la funzionalità della rete. Gli hub<sup>1</sup> sono nodi altamente connessi e statisticamente rari che tengono insieme le reti sociali.

<sup>1</sup> Definizione di hub data da Albert-Laszlo Barabasi nel libro *Link La scienza delle reti* di cui è autore



## Confronto con la metodologia tradizionale

Per effettuare un primo confronto in termini di validità ed accuratezza degli indicatori ottenuti dalla analisi di rete rispetto a quelli ottenuti con tecnologia tradizionale, abbiamo estratto lo stesso data set caricato sulla piattaforma Binet e lo abbiamo analizzato con metodi di statistica descrittiva standard.

I grafici in figura 8 mostrano le distribuzioni di frequenza di prescrizioni e ricoveri nell'anno 2007 considerando lo stesso campione di popolazione e gli stessi flussi sanitari.

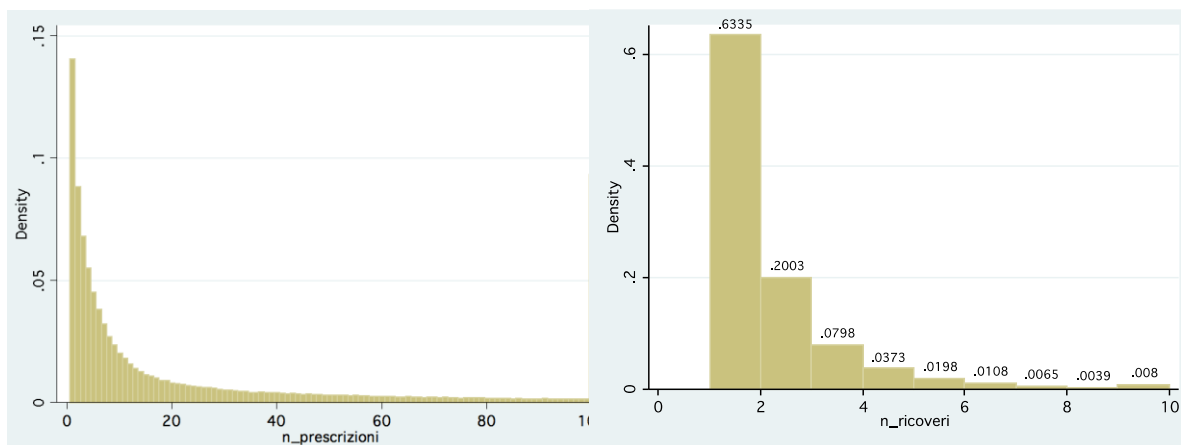


Figura 8: Distribuzione prescrizioni farmaceutiche e ricoveri

Questo primo confronto indica coerenza tra i risultati delle analisi effettuate con la metodologia della Social Network Analysis e la metodologia statistica descrittiva, si può osservare infatti che le distribuzioni rappresentate graficamente in figura 7 e figura 8 seguono la stessa legge di potenza esponenziale verso il basso. Vanno però effettuate alcune osservazioni: in questo esempio di rete sanitaria costituita di pazienti neoplastici affetti da patologie concomitanti (diabete, problemi cardiovascolari e broncopneumopatie) non abbiamo inserito i nodi che rappresentano le prestazioni ambulatoriali, principalmente per limitazioni temporali. Pertanto siamo consapevoli che questa limitazione potrebbe modificare alcune delle misure, così come il modello adottato per rappresentare le connessioni potrebbe limitare l'espressività della rete; inoltre va considerato anche che il fatto che il framework è un primo risultato del progetto di ricerca ancora in evoluzione.

Inoltre, a completamento di questa prima indagine intendiamo effettuare ulteriori confronti utilizzando metodi di analisi statistica più complessi rispetto alle descrittive standard.

Un'ipotesi di lavoro è quella di utilizzare tecniche quali l'analisi delle corrispondenze o delle componenti principali, mediante l'ausilio di software statistico specialistico.