

UNIVERSITÀ DEGLI STUDI DI MILANO-BICOCCA
DIPARTIMENTO DI INFORMATICA, SISTEMISTICA E COMUNICAZIONE
CORSO DI LAUREA MAGISTRALE IN INFORMATICA



Trenord Transport Network

Marco Piazza - 829588 Cazzaniga Elisa - 829914

Indice

Acronimi	6
1 Introduzione	7
1.1 Descrizione del problema	7
1.2 Dataset	7
2 Costruzione della rete	9
2.1 Interpretazioni	9
2.1.1 Rappresentazione uno	9
2.1.2 Rappresentazione due	10
3 Task	13
3.1 Vulnerabilità della rete	13
3.1.1 Analisi a priori della rete	13
3.1.2 Degree Centrality e Betweenness Centrality	16
3.1.3 Strategie di attacco	16
3.2 Carico giornaliero	17
4 Case study - Task 1	19
4.1 Analisi a priori	19
4.2 Degree Centrality e Betweenness Centrality	21
4.3 Strategie di attacco	22
4.3.1 Random Attack	22
4.3.2 Targeted Attack	23
4.3.3 Cascading Failure - Betweenness	25
4.3.4 Cascading Failure - Random	26

5	Case study - Task 2	29
5.1	Analisi a priori	29
5.2	Analisi del carico sui nodi della rete	31
5.3	Analisi delle tratte	35
5.3.1	Categorie di tratte	37
5.4	Analisi delle fasce orarie	40
5.4.1	Analisi delle tratte per fasce orarie	44
6	Conclusione	47
6.1	Task 1	47
6.1.1	Analisi a priori	47
6.1.2	Strategie d'attacco	47
6.2	Task 2	48

Elenco delle figure

2.1	Grafo rete Trenord	10
2.2	Schema logico	11
2.3	Come agiscono gli schemi di costruzioni delle reti sullo stesso viaggio	12
4.1	Distribuzione di grado	20
4.2	Il grafico a sinistra rappresenta la degree centrality, mentre quello a destra la betweenness centrality	21
4.3	Efficienza e Giant Connected Component durante l'attacco random	22
4.4	Efficienza e la Giant Connected Component Giant Connected Component (GCC) durante l'attacco mirato	23
4.5	La rete a sinistra rappresenta l'attacco tragettizzato basato sulla degree centrality. La rete a destra rappresenta l'attacco tragettizzato basato sulla betweenness centrality. Nella figura i punti neri rappresentano i vertici rimossi.	24
4.6	La rete a sinistra rappresenta l'attacco tragettizzato basato sulla degree centrality. La rete a destra rappresenta l'attacco tragettizzato basato sulla betweenness centrality. Nella figura i punti neri rappresentano i vertici rimossi.	25
4.7	Rete prima e dopo l'effetto Cascading Failure. Nella figura i punti neri rappresentano i vertici rimossi. Il punto giallo è il nodo che genera il fallimento a cascata.	26
4.8	Nella rete a sinistra il nodo che ha provocato il fallimento a cascata è Sesto San Giovanni. Nella rete a destra il nodo che ha provocato il fallimento a cascata è Dubino. Nelle figure i punti neri rappresentano i vertici rimossi. Il punto giallo è il nodo che ha generato il fallimento a cascata.	27
4.9	28
5.1	Distribuzione di grado	30
5.2	Distribuzione di grado in ingresso e uscita	31
5.3	Carico in ingresso e uscita	32
5.4	Carico dei nodi - disposizione geografica	33

5.5	Confronto tra l'andamento del carico sui nodi e la centralità di grado	34
5.6	Confronto tra l'andamento del carico e la betweenness dei nodi	35
5.7	Prime 20 tratte per carico con numero di stazioni attraversate e distanza percorsa .	36
5.8	Confronto tra carico, numero di stazioni attraversate e distanza	36
5.9	Confronto tra carico, numero di stazioni attraversate e distanza su un unico grafico	37
5.10	Confronto tra il carico sulle varie categorie	38
5.11	Confronto tra il numero di stazione attraversate	39
5.12	Confronto tra la distanza percorsa	39
5.13	Confronto tra l'andamento delle varie misure	39
5.14	Confronto carico totale tra le diverse fasce orarie	41
5.15	Confronto carico sbilanciato in alcune città	42
5.16	Carico stazioni complessivo (sinistra) - Carico stazioni prima 9:59 (destra)	43
5.17	Carico stazioni 10:00 - 12:59 (sinistra) - Carico stazioni 13:00 - 16:59 (destra) . . .	43
5.18	Carico stazioni 17:00 - 19:59 (sinistra) - Carico stazioni dopo 20:00 (destra)	44
5.19	Andamento delle categorie di tratte	45
5.20	Divisione per fasce orarie del carico sulle varie tratte	45

Elenco delle tabelle

4.1	Numero di nodi, archi, grado minimo e massimo, shortest path medio, densità e assortatività della rete	19
4.2	Valori di efficiency(e), dimensione della GCC(S), spectral gap (s) e algebraic connectivity (λ_2)	20
5.1	Numero di nodi, archi, grado minimo, massimo, medio, shortest path medio, densità, assortatività della rete, reciprocità	29
5.2	Dimensioni delle diverse categorie di tratte	38
5.3	Correlazione tra il carico e le altre misure diviso per categorie	40

Acronimi

GCC Giant Connected Component

PTN Public Transport Network

Capitolo 1

Introduzione

1.1 Descrizione del problema

L'obiettivo di questo progetto è stata l'analisi del grafo, basandosi sulle assunzioni legate alla teoria dei grafi, rappresentante la rete ferroviaria di Trenord. L'analisi si è sviluppata secondo due maggiori direttrici: la prima riguardante l'analisi della robustezza delle rete e della resilienza, ovvero la capacità di resistere a cadute di uno o più nodi; la seconda ha riguardato l'analisi del carico sulla rete, quindi è stato considerato per ogni stazione e per ogni tratta il numero di treni e di tratte che vi transitano. Pensiamo che l'analisi di questa rete possa produrre insight e indicazioni utili al fine di migliorare la viabilità sulla rete e in caso di ampliamento della stessa possa suggerire quali scelte seguire.

1.2 Dataset

Il dataset di partenza utilizzato per la costruzione della rete è disponibile al seguente link:

<https://transitfeeds.com/p/trenord/724/latest>

Contiene diversi dati legati alla rete ferroviaria Trenord. In particolare, è costituito da 7 file csv:

- *agency.csv*: contiene i dati dell'agenzia ferroviaria
- *calendar.csv*: contiene i giorni della settimana in cui viene effettuata una specifica corsa.
- *calendar_dates.csv*: contiene le corse con la relativa problematica riscontrata.

- *routes.csv*: contiene tutte le tratte.
- *stop_times.csv*: contiene le fermate di ogni viaggio.
- *stops.csv*: contiene tutte le stazioni ferroviarie di Trenord.
- *trips.csv*: tabella riassuntiva che contiene i collegamenti tra le altre tabelle.

I dati contenuti nei vari file sono stati estratti e sono stati rielaborati per costruire reti diverse, come verrà presentato in seguito.

Questo documento avrà la seguente struttura:

- Nel secondo capitolo verranno presentate le due interpretazioni e le rispettive costruzioni della rete che abbiamo realizzato.
- Nel terzo capitolo verranno presentati da un punto di vista teorico i task che abbiamo svolto.
- Nel quarto capitolo verranno presentate le analisi svolte relative al primo task, con i rispettivi risultati ottenuti.
- Nel quinto capitolo verranno presentate le analisi svolte relative al secondo task, con i rispettivi risultati ottenuti.
- Nel sesto capitolo verranno presentate le conclusioni.

Capitolo 2

Costruzione della rete

2.1 Interpretazioni

Seguendo l'esempio di altri lavori presenti in letteratura, come [von Ferber et al., 2009b] abbiamo ipotizzato che non esistesse una modalità univoca per rappresentare la nostra realtà di interesse in forma di grafo. Partendo dallo stesso dataset, abbiamo deciso di creare due grafi differenti per effettuare le due analisi che ci eravamo prefissati.

2.1.1 Rappresentazione uno

Il primo grafo rappresenta la struttura vera e propria della rete, ovvero come sono collegate fisicamente tra di loro stazioni adiacenti. Per ogni viaggio (che nel dataset rappresenta un viaggio reale intrapreso da un treno) abbiamo ordinato le stazioni in base alla sequenza delle fermate e abbiamo collegato tra loro due stazioni se e solo se il numero di stazioni intermedie è uguale a 0 (ovvero solo se due stazioni hanno *sequence_id* consecutivi nello stesso viaggio). In questa rappresentazione non è possibile avere più di un arco tra due nodi. Ogni stazione presente nella tabella *stop.csv* costituisce un nodo della rete. I nodi sono caratterizzati dai seguenti attributi:

- *name*: corrisponde al nome della stazione;
- *stop_id*: corrisponde all'id della stazione.
- *stop_lat*: corrisponde alla latitudine della stazione.
- *stop_lon*: corrisponde alla longitudine della stazione.

Gli archi invece non presentano attributi e sono non orientati, in quanto consideriamo che il collegamento tra stazioni è bidirezionale, se esiste in un senso esiste anche nell'altro. Questa rappresentazione ci è stata utile per la parte di analisi sulla robustezza e resilienza della rete.



Figura 2.1: Grafo rete Trenord

2.1.2 Rappresentazione due

Il secondo grafo rappresenta la rete delle stazioni che vengono toccate successivamente dalle varie corse dei treni di diverse tratte ferroviarie. Per ogni viaggio abbiamo ordinato le stazioni in base alla sequenza delle fermate. Ogni coppia di stazioni adiacenti nella sequenza di fermate del viaggio costituisce un arco del grafo (indipendentemente dalla distanza dei *sequence_id*). Ogni stazione presente nella tabella *stop.csv* costituisce un nodo della rete. I nodi sono caratterizzati dagli stessi attributi visti nella rappresentazione precedente:

- *name*: corrisponde al nome della stazione;
- *stop_id*: corrisponde all'id della stazione.
- *stop_lat*: corrisponde alla latitudine della stazione.

- *stop_lon*: corrisponde alla longitudine della stazione.

Anche gli archi presentano degli attributi:

- *route*: corrisponde al nome della tratta rappresentata dall'arco;
- *size*: rappresenta il numero di corse giornaliere per una data tratta su un dato arco.

Inoltre gli archi in questo caso sono orientati, perchè abbiamo osservato che può esistere il collegamento tra due stazioni in un verso ma non nell'altro, consideriamo quindi che questa scelta rappresenti meglio il dominio di interesse.

In questa rappresentazione è possibile avere più percorsi, anche di dimensioni diverse tra le stesse coppie di nodi, ci è stata utile per il task di misura delle performance e dei carichi sui vari nodi della rete.

Per un ulteriore chiarimento abbiamo riportato lo schema logico delle due tabelle coinvolte e un esempio per ciascuna costruzione dei grafi.

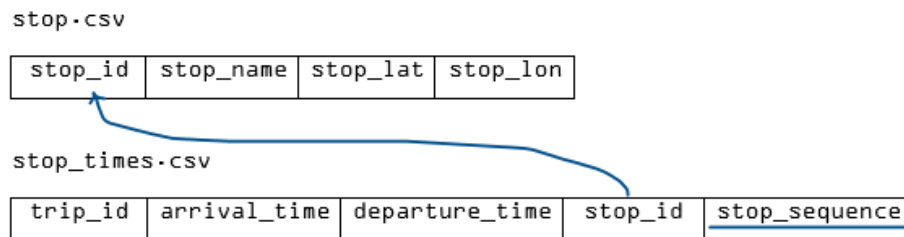


Figura 2.2: Schema logico

trip_id	arrival_time	departure_time	stop_id	stop_sequence		trip_id	arrival_time	departure_time	stop_id	stop_sequence
13497685	06:09:00	06:09:00	1139	1		13497685	06:09:00	06:09:00	1139	1
13497685	06:12:00	06:13:00	2151	2		13497685	06:12:00	06:13:00	2151	2
13497685	06:17:00	06:18:00	1258	3		13497685	06:17:00	06:18:00	1258	3
13497685	06:22:00	06:23:00	1228	4		13497685	06:22:00	06:23:00	1228	4
13497685	06:28:00	06:29:00	618	5		13497685	06:28:00	06:29:00	618	5
13497685	06:33:00	06:38:00	2993	6		13497685	06:33:00	06:38:00	2993	6
13497685	06:41:30	06:42:00	1329	7		13497685	06:41:30	06:42:00	1329	7
13497685	06:48:00	06:49:00	1575	8		13497685	06:48:00	06:49:00	1575	8
13497685	06:53:30	06:54:00	1934	9		13497685	06:53:30	06:54:00	1934	9
13497685	06:59:00	07:00:00	1632	10		13497685	06:59:00	07:00:00	1632	10
13497685	07:06:00	07:07:00	365	11		13497685	07:06:00	07:07:00	365	11
13497685	07:17:00	07:19:00	1550	12		13497685	07:17:00	07:19:00	1550	12
13497685	07:26:00	07:27:00	807	15		13497685	07:26:00	07:27:00	807	15
13497685	07:37:00	07:38:00	1066	18		13497685	07:37:00	07:38:00	1066	18
13497685	07:43:00	07:44:00	905	20		13497685	07:43:00	07:44:00	905	20
13497685	07:54:00	07:55:00	1841	22		13497685	07:54:00	07:55:00	1841	22
13497685	08:07:00	08:07:00	1728	25		13497685	08:07:00	08:07:00	1728	25

Figura 2.3: Come agiscono gli schemi di costruzioni delle reti sullo stesso viaggio

Capitolo 3

Task

In questo capitolo verranno discussi i task che vorremmo realizzare per questo progetto. L'analisi della rete seguirà due direzioni principali: la prima legata alla robustezza e resilienza della rete e la seconda legata al carico giornaliero sulla rete.

3.1 Vulnerabilità della rete

3.1.1 Analisi a priori della rete

In un sistema complesso come quello rappresentato dalla rete rappresentate il sistema ferroviario di Trenord, come ogni altra rete di trasporti pubblici, due caratteristiche fondamentali sono rappresentate dalla robustezza, ovvero la capacità di sostenere guasti o cadute di nodi e la resilienza, ovvero la capacità di sostenere le cadute ed adattarsi per continuare a funzionare. Queste due misure danno informazioni importanti circa la suscettibilità della rete a guasti e quanto si riduce il servizio offerto in caso di malfunzionamenti [Berdica, 2002]. Per stimare queste dimensioni sono disponibili in letteratura misure alternative, ognuna delle quali fornisce informazioni diverse sul sistema. Gli approcci utilizzati per la realizzazione di questo elaborato sono stati estratti dalla letteratura esistente sull'argomento, in particolare [Berche et al., 2010],[von Ferber et al., 2009a], [von Ferber et al., 2009b], [Candelieri et al., 2019].

Approcci basati su valori

I primi approcci che verranno presentati sono basati su valori. Un valore importante è quello rappresentata dall'efficienza come presentata da [Latora and Marchiori, 2001]. Questo coefficiente permette di stimare quanto una rete è efficiente nello scambio delle informazioni. La formula per il calcolo è definita:

$$E = \frac{1}{N(N-1)} \sum \frac{1}{d_{ij}} \quad (3.1)$$

dove d rappresenta la distanza minima dal nodo i al nodo j e va calcolata per ogni coppia di nodi i, j del grafo.

Se la rete non è pesata il valore è normalizzato e assume valore 1 nel caso di una rete completa, ma in caso di reti che rappresentano problemi reali di Public Transport Network (PTN) difficilmente raggiunge il valore massimo. È importante stimare questo valore inizialmente per effettuare delle ipotesi a priori sulla rete, inoltre è interessante osservare l'andamento in corrispondenza della rimozione successiva di nodi.

Il secondo valore utilizzabile è rappresentato dalla dimensione della GCC, normalizzata rispetto al numero di nodi della rete.

$$S = \frac{N_1}{N} \quad (3.2)$$

In caso di reti connesse questo valore è 1. Assume rilevanza la variazione di questo valore correlata al numero di nodi rimossi della rete, come visto precedentemente con l'efficienza.

Il terzo e ultimo valore che può fornire informazioni circa la rete è la lunghezza media dello shortest path tra i vari nodi della rete e il suo andamento ad ogni caduta di un nodo. Infatti permette di capire quanto può essere complicato raggiungere tutti i nodi della rete e se questo valore cambia con la rimozione successiva di nodi. È un valore che fornisce informazioni importanti circa la transitabilità della rete.

Node degree

Un'altra metrica che può fornire informazioni importanti circa la vulnerabilità della rete è la distribuzione di grado. Le reti denominate *scale-free* presentano una distribuzione di grado esponenziale negativa, con la presenza quindi di un numero elevato di spoke (nodi con pochi archi connessi) e di un numero minore di hubs (nodi con molti archi connessi). Questa tipologia di rete è molto efficace nel resistere ad attacchi random, mentre ha più difficoltà a resistere nel caso di attacchi mirati. (Intuitivamente questo significa che un attacco random ha più probabilità di colpire nodi spoke, quindi poco importanti ma più numerosi, questa tipologia di attacco non avrà però grosse ripercussioni sulla rete. Un attacco mirato invece per prima cosa rimuoverà nodi hub, precludendo quindi l'intera comunicazione sulla rete)

Fall of efficiency

Un'altra possibile misura che abbiamo utilizzato per stimare gli effetti di un attacco sulla rete è la Fall of efficiency [Candelieri et al., 2019]. Questa misura è basata sulla misura di efficiency

presentata prima e misura la differenza di questo valore provocata dalla rimozione di un dato arco, rispetto al valore di efficiency iniziale. Questa misura può essere quindi così definita:

$$C_{\Delta}^E = \frac{E(G) - E(G_i)}{E(G)} \quad (3.3)$$

Inoltre il valore massimo ottenuto fornisce informazioni circa il nodo più "importante" della rete dal punto di vista dell'efficienza nel trasferimento di informazioni nella stessa.

Tecniche spettrali

Informazioni aggiuntive sugli aspetti analizzati della rete possono essere forniti da un'analisi spettrale della rete. Per effettuare questa tipologia di analisi è necessario lavorare sulla matrice di adiacenza e sui suoi autovalori. In particolare l'autovalore di modulo massimo è definito raggio spettrale (ρ). La prima misura introdotta è quella denominata spectral gap, che misura la differenza tra il raggio spettrale della matrice e l'autovalore di modulo inferiore:

$$s(G) = \rho G - \mu_{n-1}(G) \quad (3.4)$$

Questo valore fornisce informazioni molto importanti in termini della connettività della rete. In particolare, in caso di reti poco connesse, o in caso di presenza di colli di bottiglia o nodi che svolgono la funzione di ponte questo valore è basso. Una bassa connettività nella rete e la presenza di colli di bottiglia o nodi ponte nella rete può pesare negativamente sulla robustezza e resilienza della stessa. L'ultimo valore considerato nell'analisi della robustezza della rete è denominato algebraic connectivity. Per questo è necessario introdurre il concetto di Laplacian matrix. Questa matrice è molto importante e largamente usata nell'analisi di grafi per alcune sue caratteristiche peculiari. La matrice è definita come:

$$L(G) = D(G) - A(G) \quad (3.5)$$

Dove $D(G)$ rappresenta una matrice diagonale in cui ogni elemento i,i corrisponde al grado del nodo i del grafo e $A(G)$ rappresenta la matrice di adiacenza del grafo G . Gli autovalori di questa matrice sono chiamati autovalori laplaciani e sono molto importanti. Il primo ha sempre valore 0 e la sua molteplicità è uguale al numero di componenti connesse della rete. Il secondo viene chiamato algebraic connectivity, in particolare è considerato un parametro molto importante per stimare la robustezza della rete, infatti ad un valore elevato di questo valore corrisponde una rete robusta, quindi difficile da disconnettere nelle singole componenti connesse e viceversa.

3.1.2 Degree Centrality e Betweenness Centrality

Tra le misure di centralità esistenti, abbiamo deciso di studiare la degree e la betweenness sui nodi della nostra rete. La degree centrality ci dà informazioni riguardanti quanti collegamenti ha una stazione con quelle adiacenti.

$$C_D(i) = \sum_j x_{ij} \quad (3.6)$$

La betweenness centrality ci dice quanto una stazione funge da collegamento tra altre stazioni.

$$C_B(i) = \sum_{j < k} g_{jk}(i) / g_{jk} \quad (3.7)$$

dove $g_{jk}(i)$ è il numero di percorsi geodesici tra i e k che passano attraverso i .

Quindi se una stazione è centrale secondo una di queste misure si può dire che ha un ruolo importante all'interno della rete.

3.1.3 Strategie di attacco

Per quanto riguarda la vulnerabilità strutturale, è possibile studiare come cambiano le caratteristiche di una rete complessa quando vengono rimossi alcuni suoi elementi. L'origine dell'attacco e il suo scenario possono differire in larga misura, passando da un attacco casuale, ad un attacco mirato. Una volta che un nodo, in modo casuale o mirato, viene attaccato e rimosso dalla rete, ci sono due modi diversi per continuare e/o analizzare l'attacco: l'approccio statico [von Ferber et al., 2009a], e l'approccio dinamico [Zou et al., 2013]. Nell'approccio statico, si presume che il fallimento di un nodo non provochi il fallimento di altri nodi. Nell'approccio dinamico, si presume che la rimozione del nodo possa causare il fallimento di altri nodi. In condizioni di limite di capacità della rete, un guasto di un nodo porterà alla ridistribuzione del carico dell'intera rete, facendo sì che il carico di alcuni nodi superi la sua capacità e provochi il possibile guasto di altri nodi.

Random attacks

In un attacco random, uno o più nodi della rete vengono rimossi in modo casuale.

Target attack

Uno o più nodi della rete (i componenti della rete più influenti) vengono rimossi in base alle loro caratteristiche di funzionamento. È possibile misurare continuamente le caratteristiche dei nodi dopo ogni passaggio e modificare tale elenco nel corso di un attacco. Gli attacchi basati su elenchi ricalcolati spesso si rivelano più dannosi delle strategie di attacco basate sull'elenco

iniziale, suggerendo che la struttura della rete cambia quando vengono rimossi nodi o collegamenti importanti.

Cascading failure

Un fallimento a cascata è un effetto domino che si origina quando la caduta di un certo nodo innesca successivi fallimenti di uno o più nodi, che a loro volta innescano altri fallimenti. Pertanto, il numero di nodi guasti o stressati aumenta, propagandosi in tutta la rete. Un esempio di guasto a cascata nel mondo reale è un'interruzione di corrente su vasta scala. Un semplice modello per un guasto a cascata su una rete presuppone che ogni nodo trasmetta un'unità di una certa quantità (energia, informazioni, volume di passeggeri, ecc.) a ogni altro nodo attraverso il percorso più breve che li collega. Di conseguenza, ogni nodo i è caratterizzato da un certo carico L_i , che descrive l'importanza del nodo nella rete. Generalmente, per calcolare il carico dei nodi si utilizza il valore di betweenness, che rappresenta il numero di cammini minimi passanti per quel nodo. Ogni nodo i è caratterizzato da una certa capacità Ψ_i , che è il carico massimo che può essere gestito da quel nodo. La capacità assegnata a ciascun nodo è proporzionale al carico che quel nodo deve gestire.

$$\Psi_i = (1 + \alpha) * L_i(0) \forall i \in V^0 \quad (3.8)$$

dove $\alpha > 0$ è un parametro di tolleranza che quantifica il carico in eccesso che un dato nodo può gestire e V^0 è l'insieme dei nodi del grafo originale.

I fallimenti a cascata avvengono quando un nodo pesantemente caricato cade per qualche motivo e il suo carico deve essere ridistribuito sugli altri nodi della rete. La rimozione del nodo simula la perdita del nodo. La ridistribuzione del carico agli altri nodi della rete richiede il ricalcolo del valore di carico $L_i(t + 1)$ per ogni nodo i . Questo può alterare radicalmente i valori dei carichi sulla rete. Alcuni nodi possono avere un carico inferiore al loro valore precedente e altri possono averne uno maggiore. In particolare, se un nodo i ha un nuovo carico $L_i(t + 1) > \Psi_i$, allora anche questo nodo fallisce. Questi guasti possono a loro volta innescare più guasti, portando così ad una cascata di fallimenti. Il processo viene iterato fin quando non devono più essere rimossi nodi dalla rete.

3.2 Carico giornaliero

Il secondo task si pone l'obiettivo di analizzare la rete ottenuta dalle stazioni di Trenord e dai loro collegamenti da un punto di vista differente rispetto al task precedente. I nodi della rete rappresentano le stazioni della rete ferroviaria, gli archi invece rappresentano una tratta che tocca entrambe le stazioni, l'arco ha un peso che indica il numero di treni che quotidianamente passano

tra le due stazioni per la tratta. Inoltre sono state create altre reti, in diverse fasce orarie della giornata per effettuare i confronti sul carico di treni a seconda dell'ora su cui si sta osservando. Sono state identificate cinque fasce orarie su cui effettuare i confronti:

- Fascia 1: fino alle 9:59
- Fascia 2: dalle 10 alle 12:59
- Fascia 3: dalle 13 alle 16:59
- Fascia 4: dalle 17 alle 19:59
- Fascia 5: dopo le 20

Anche in questo caso è stata effettuata un'analisi a priori della rete, volta ad indagare la sua struttura. In questo caso non è stato trovato alcun lavoro simile in letteratura, perciò non si è potuto partire da conoscenze pregresse o trarre informazioni. Si è quindi proceduto con le conoscenze di base della teoria dei grafi. Per l'analisi del task vero e proprio è usato il peso degli archi per stimare il valore del carico di ogni tratta e di ogni nodo della rete. In questo lavoro è stato definito carico sulla rete ferroviaria il numero quotidiano di treni che vi transitano.

Capitolo 4

Case study - Task 1

In questa sezione verrà presentato l'implementazione del task uno e i risultati ottenuti dall'analisi della rete rappresentante le stazioni Trenord. In primo luogo è stata condotta un'analisi a priori sulla rete per valutarne la robustezza, quindi sono state valutate le prestazioni della stessa a seguito di diverse strategie di attacco e fallimento.

4.1 Analisi a priori

La rete è stata analizzata a priori per conoscerne la struttura e la topologia, nonché per effettuare ipotesi e previsioni sulla sua robustezza e la capacità di adattarsi a potenziali cadute di nodi. Per stimare robustezza e resistenza sono stati utilizzati i parametri presentati in nella sezione 3.1.1 inoltre sono state effettuate altre misure e analisi per conoscere meglio la rete. Sono state effettuate alcune misure di base sulla rete che verranno riportate e discusse nella tabella seguente:

Nodi	Archi	$\min(d_k)$	$\max(d_k)$	\bar{l}	d	a
415	444	1	5	23.34073	0.00516	0.23966

Tabella 4.1: Numero di nodi, archi, grado minimo e massimo, shortest path medio, densità e assortatività della rete

Per questa rete abbiamo deciso di utilizzare gli archi non orientati, in quanto pensiamo che il collegamento tra due stazioni sia bidirezionale. Come si può vedere dalla tabella il numero di nodi e il numero di archi non differenzia di molto, questo significa che ogni nodo presenterà pochi collegamenti, questa ipotesi è anche confermata dal valore di grado massimo. Un'altra conseguenza della forma che la rete assume è l'elevato valore di shortest path medio, infatti è ipotizzabile che per

collegare due nodi sia necessario passare a sua volta attraverso molti altri, visti i pochi collegamenti che ognuna di essi presenta. Per tutti questi motivi anche il valore di densità, ovvero il rapporto tra il numero di archi e il numero massimo possibile di archi è molto basso. Infine il valore di assortatività della rete è positivo, quindi la rete presenta una tendenza assortativa, ovvero gli hub tenderanno a legarsi con altri hub, anche se questo valore può essere influenzato dal valore molto basso di grado massimo.

Successivamente è stata indagata la distribuzione di grado, per capire se la rete può essere considerata una rete *scale-free* o meno, ovvero se la distribuzione di grado, asintoticamente presenta una distribuzione di tipo esponenziale.

Fig. 4.1 Nella prima immagine distribuzione di grado con scala normale, nel secondo scala logaritmica.

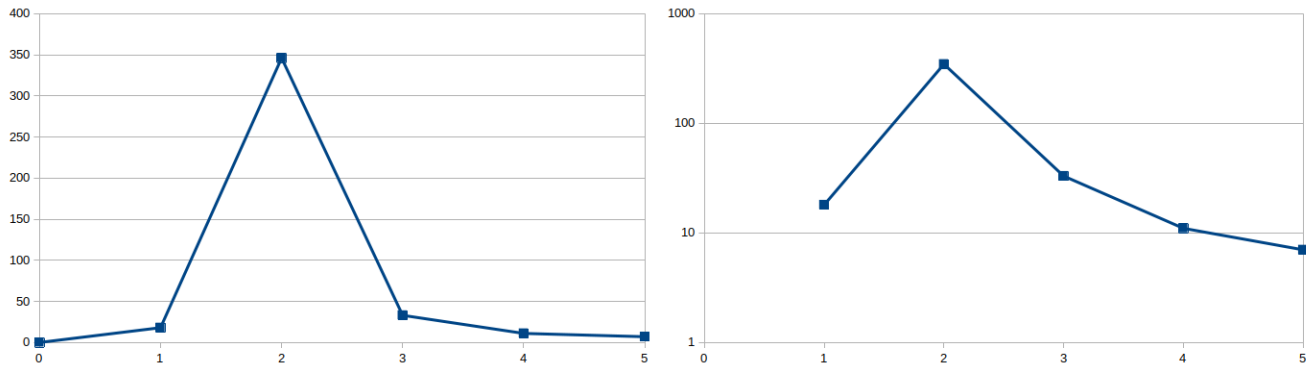


Figura 4.1: Distribuzione di grado

Come si può riconoscere dai due grafici l'andamento delle curve tende ad avvicinarsi a quello esponenziale, tuttavia avendo un valore di grado massimo così contenuto non è possibile osservare l'andamento a livello asintotico. È stato comunque ipotizzato che la rete si possa comportare come una rete di tipo *scale-free*, in quanto esistono molti nodi che valore di grado limitato e pochi nodi con i valori più elevati di grado.

Infine abbiamo provato a stimare la robustezza e la resilienza a priori mediante alcune misure definite precedentemente di cui vengono riportati i valori.

Tab. 4.2 Misure iniziali per robustezza rete

$e(G)$	$S(G)$	$s(G)$	λ_2
0.05909	1	0.52067	0.00185

Tabella 4.2: Valori di efficiency(e), dimensione della GCC(S), spectral gap (s) e algebraic connectivity (λ_2)

Il valore dell'efficienza è relativamente basso, tuttavia è noto che viene assunto valore 1 in corrispondenza di reti complete, il valore rilevato per la rete in esame è quindi completamente atteso, visto il basso numero di archi a confronto dei nodi. La rete allo stato iniziale risulta connessa, quindi il rapporto tra la dimensione della GCC e dell'intera rete è 1. Invece il valore di spectral gap non essendo troppo basso potrebbe suggerire che la rete non sia debolmente connessa e la rete non presenti situazioni di colli di bottiglia o ponti. Questo va un in contro tendenza alle ipotesi fatte inizialmente sulla rete, infatti si credeva che gli hub della rete rappresentassero possibili colli di bottiglia. L'ultimo valore è quello dell'algebraic connectivity, come si può vedere il valore è molto basso, questo significa che la rete non è robusta ed è facile separarla in più componenti disconnesse tra loro.

4.2 Degree Centrality e Betweenness Centrality

Abbiamo calcolato la degree centrality e la betweenness centrality per tutti i nodi della rete. Di seguito riportiamo due figure che mostrano i nodi della rete colorati in base al valore di centralità che assumono. Partendo dal colore più basso, la scala di colorazione è la seguente: bianco - giallo - arancione - rosa - rosso.



Figura 4.2: Il grafico a sinistra rappresenta la degree centrality, mentre quello a destra la betweenness centrality

Come si può notare dalla figura e come ci aspettavamo, i nodi con degree maggiore sono situati nell'area di Milano e limitrofi. Poi seguono le stazioni che fungono da collegamento tra le

varie tratte. Per quanto riguarda la betweenness i nodi con valore maggiore rimangono comunque nell'area milanese, ma sono presenti alcune stazioni nelle province di Lecco, Bergamo e Brescia che fungono da collegamento per le stazioni più isolate con il resto della rete.

4.3 Strategie di attacco

Abbiamo sottoposto le rete alle tre tipologie di attacco presentate in precedenza per analizzare i vari comportamenti.

4.3.1 Random Attack

Per analizzare un possibile attacco random abbiamo simulato la chiusura di un certo numero di stazioni (75%) in modo randomico. I risultati dell'attacco sono riportati nelle figure seguenti, in cui mostriamo il cambiamento dei valori di Efficiency (E) e la Giant Connected Component GCC (S) in funzione del numero di nodi rimossi.

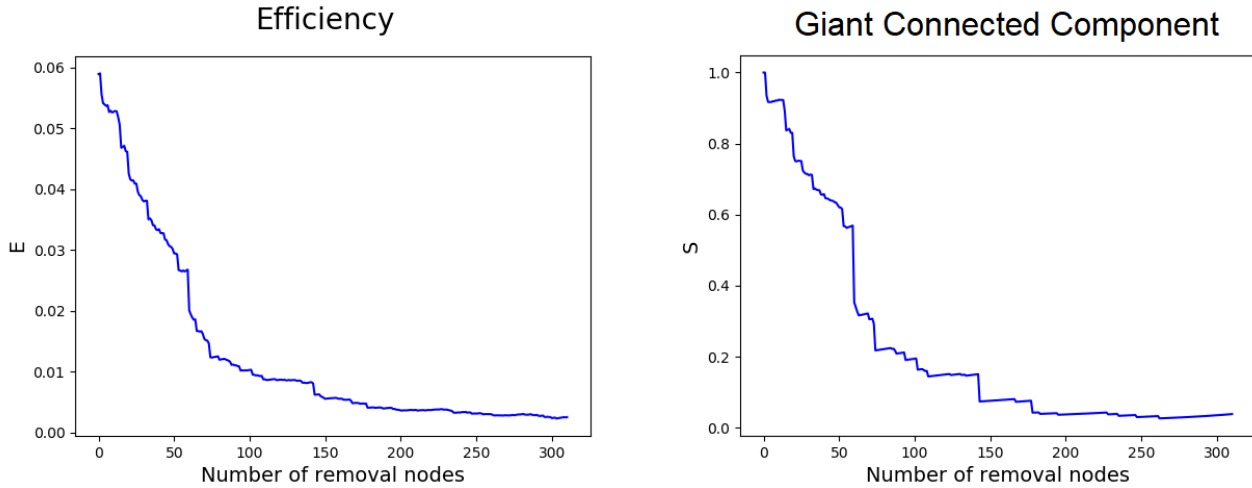


Figura 4.3: Efficienza e Giant Connected Component durante l'attacco random

Possiamo osservare che entrambe le curve presentano un andamento quasi esponenziale, ovvero abbiamo una notevole riduzione dei valori rimuovendo un numero di nodi abbastanza contenuto. Per quanto riguarda la misura S, sono presenti alcuni picchi di decremento che portano un danno elevato.

4.3.2 Targeted Attack

Per analizzare un possibile attacco mirato abbiamo simulato la chiusura di un certo numero di stazioni (75%) in accordo a due misure decrescenti: degree e betweenness. Queste due liste di nodi da rimuovere sono state preparate all'inizio dell'attacco. I risultati di questi attacchi sono riportati nelle figure seguenti, in cui mostriamo il cambiamento dei valori di Efficiency (E) e la Giant Connected Component GCC (S) in funzione del numero di nodi rimossi. La curva viola e la curva azzurra rappresentano gli attacchi effettuati utilizzando rispettivamente la degree centrality e la betweenness centrality.

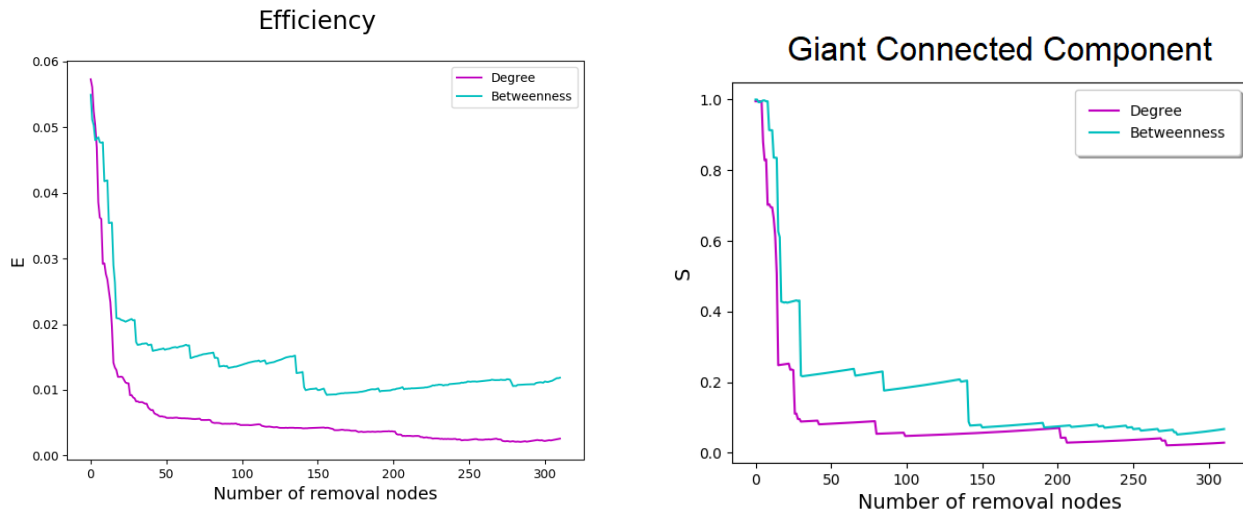


Figura 4.4: Efficienza e la Giant Connected Component GCC durante l'attacco mirato

La diminuzione più significativa dell'efficienza si verifica per le curve di grado, questo significa che un attacco mirato ha maggiore successo di riduzione del livello di efficienza della rete, quando l'obiettivo sono i nodi con grado più alto. Le curve hanno un andamento esponenziale. Il comportamento di S appare un po' diverso. La decrescita del valore di S ha comunque un andamento simile ad un'esponenziale, ma presenta diversi picchi di decremento che portano un danno elevato. Anche in questo caso, si verifica una diminuzione maggiore per quanto riguarda la curva di grado.

In una circostanza reale, generalmente, possiamo assumere che non ci siano numerosissimi fallimenti contemporaneamente. Nelle immagini che seguono abbiamo simulato il fallimento dello 0,02% delle stazioni, con entrambe le misure presentate in precedenza.

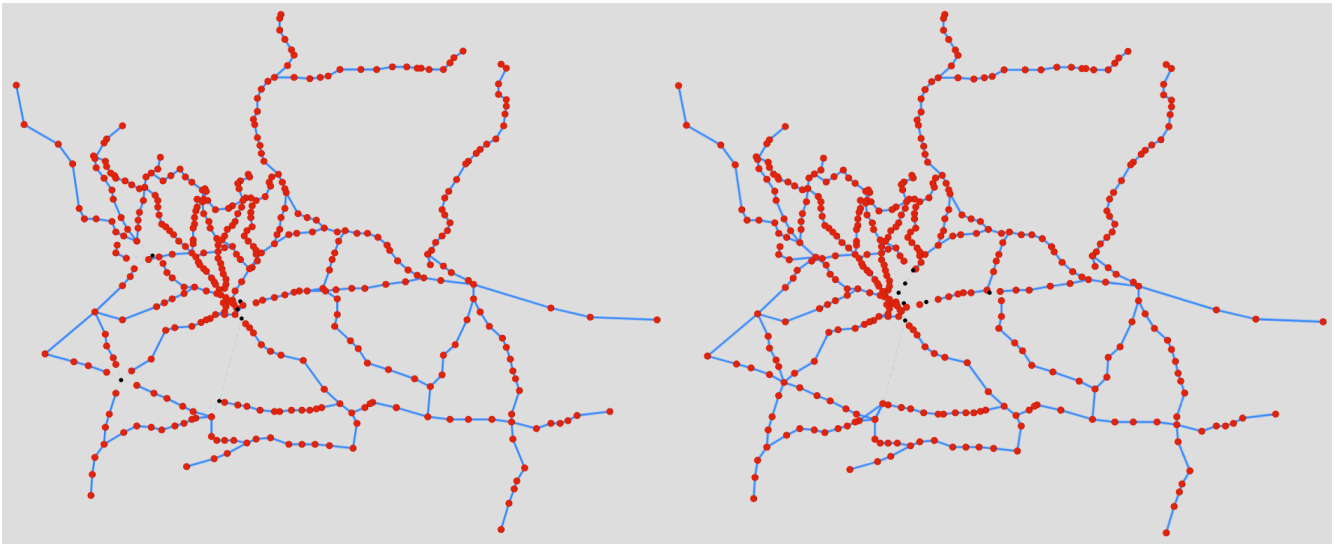


Figura 4.5: La rete a sinistra rappresenta l'attacco tragettizzato basato sulla degree centrality. La rete a destra rappresenta l'attacco tragettizzato basato sulla betweenness centrality. Nella figura i punti neri rappresentano i vertici rimossi.

Possiamo osservare che il danno è abbastanza contenuto ed essendo il numero di nodi rimossi ridotto, la differenza tra il peggioramento effettuando un attacco basato sulla degree piuttosto che sulla la betweenness non è ancora evidente.

Per riuscire ad osservare che un attacco mirato ha maggiore successo di riduzione del livello di efficienza e di suddivisione della rete, quando l'obiettivo sono i nodi con grado più alto, mostriamo le rete in cui falliscono circa il 30% dei nodi.

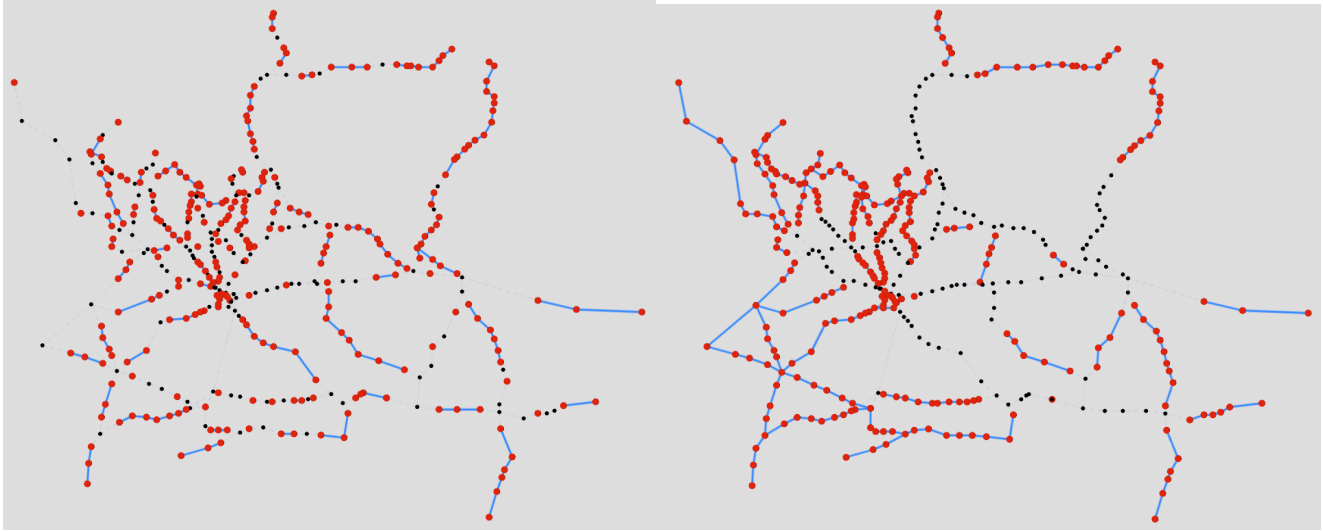


Figura 4.6: La rete a sinistra rappresenta l'attacco tragettizzato basato sulla degree centrality. La rete a destra rappresenta l'attacco tragettizzato basato sulla betweenness centrality. Nella figura i punti neri rappresentano i vertici rimossi.

È evidente come nell'attacco basato sulla degree, la rete venga suddivisa in un numero di componenti maggiori e più piccole. A seguito dell'attacco basato su betweenness, la rete risulta comunque molto danneggiata, ma esiste ancora qualche tratta che potrebbe funzionare.

4.3.3 Cascading Failure - Betweenness

Per analizzare questo fenomeno abbiamo simulato la chiusura della stazione con il più alto valore di betweenness. Abbiamo rimosso questo nodo impostando la capacità $\alpha = 0,01$. Il ricalcolo del betweenness per ogni nodo permette di identificare i nuovi nodi che falliscono in cascata (cioè nodi con capacità inferiore al carico attuale). Questi nodi vengono rimossi e il processo viene ripetuto fino a quando nessun altro nodo fallisce. Nelle figure seguenti viene mostrato il confronto della rete prima e dopo il fallimento a cascata. I punti neri rappresentano i vertici rimossi.

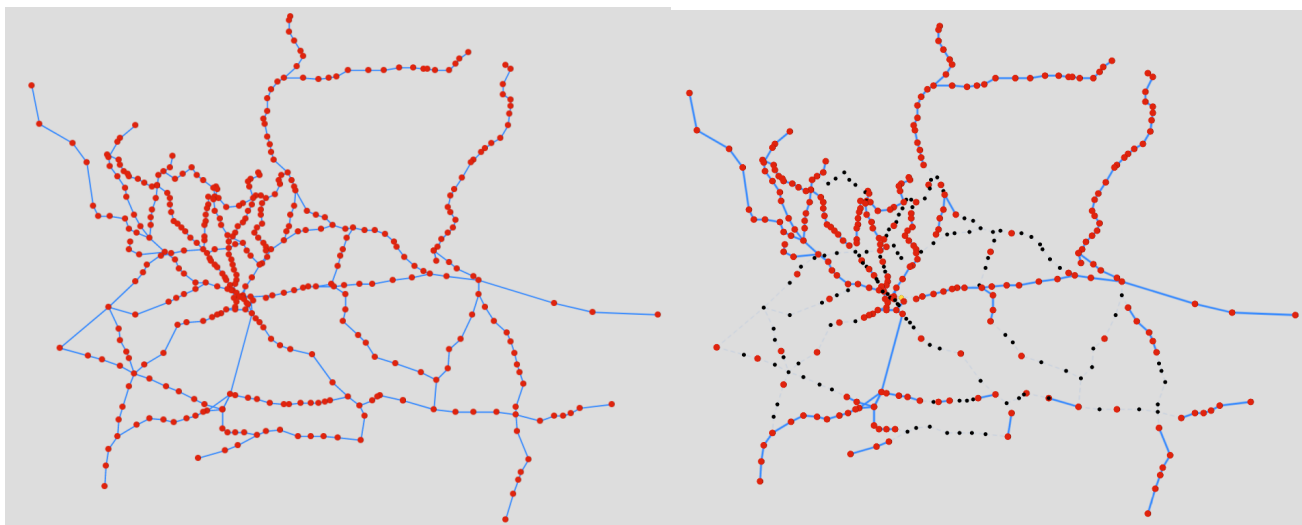


Figura 4.7: Rete prima e dopo l'effetto Cascading Failure. Nella figura i punti neri rappresentano i vertici rimossi. Il punto giallo è il nodo che genera il fallimento a cascata.

La maggior parte dei nodi viene rimossa perchè il carico supera la capacità dopo la rimozione del primo nodo. Possiamo notare che la maggior parte dei nodi rimossi è localizzata al centro, in prossimità del nodo iniziale del fallimento a cascata.

4.3.4 Cascading Failure - Random

Abbiamo analizzato il fallimento a cascata anche con la scelta di un nodo random che causa il fenomeno. Questo per confrontare il comportamento della rete in base a qual è il nodo che genera il fallimento. Nelle immagini che seguono abbiamo scelto di riportare il fallimento di un nodo con un valore di centralità di betweenness abbastanza elevato e uno con un valore di centralità di betweenness basso.

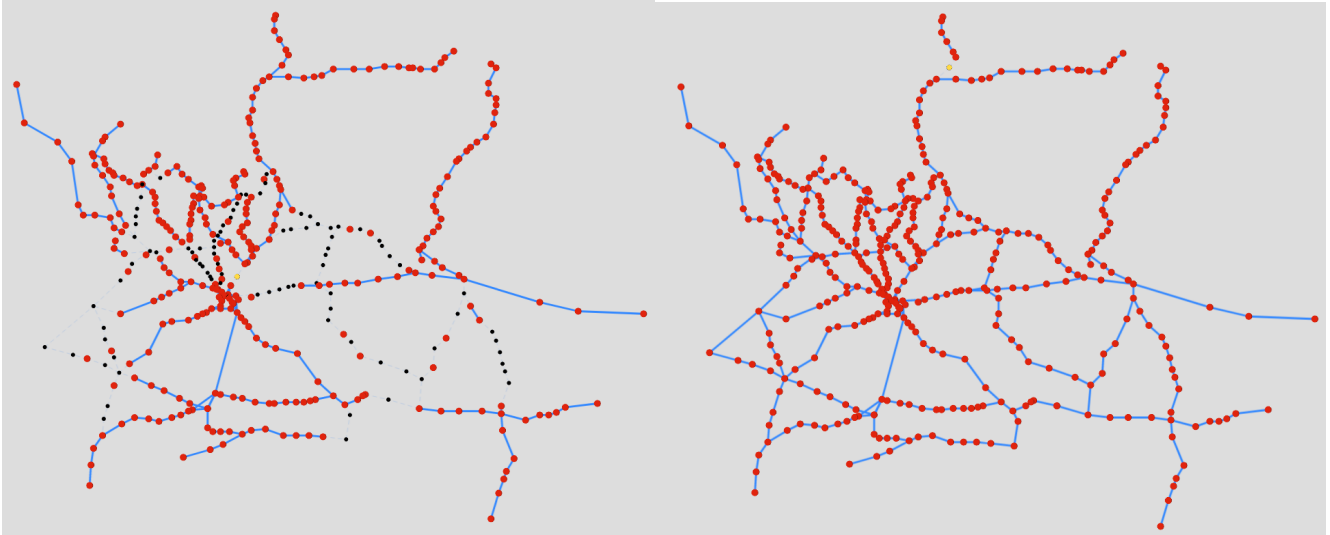


Figura 4.8: Nella rete a sinistra il nodo che ha provocato il fallimento a cascata è Sesto San Giovanni. Nella rete a destra il nodo che ha provocato il fallimento a cascata è Dubino. Nelle figure i punti neri rappresentano i vertici rimossi. Il punto giallo è il nodo che ha generato il fallimento a cascata.

Possiamo notare che nel primo caso, l'effetto del fallimento sulla rete è molto più importante. La rete viene suddivisa in diverse sottocomponenti. Nel secondo caso, il fallimento è contenuto. Viene interrotta solamente la tratta che comprende il nodo fallito.

Riportiamo di seguito un esempio di confronto di E ed S di un fallimento a cascata basato sulla betweenness e di un fallimento a cascata random. È stato scelto un esempio significativo in cui anche con il fallimento a cascata random vengono rimossi numerosi nodi. Infatti, in un fallimento a cascata random, potrebbe capitare, che rimuovendo un nodo poco importante non ci siano ripercussioni sul resto della rete.

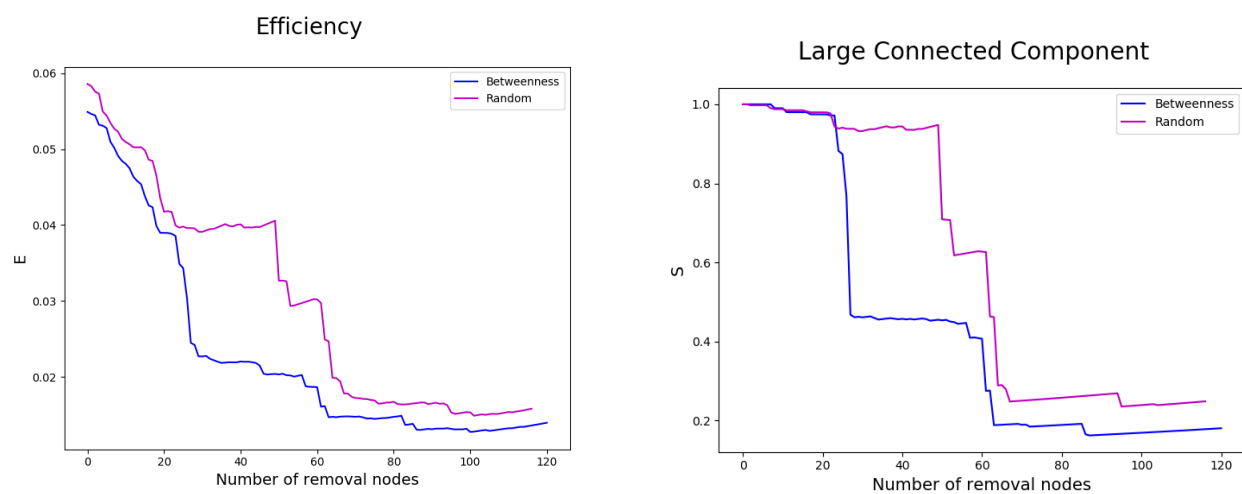


Figura 4.9: .

In generale, possiamo notare che per entrambe le misure, la betweenness porta un peggioramento più significativo. Per quanto riguarda S , fino alla rimozione di una ventina di nodi, in questo specifico caso, non è presente differenza tra le due curve.

Capitolo 5

Case study - Task 2

In questo capitolo verrà presentata l'implementazione e i risultati ottenuti sul task 2, ovvero l'analisi della rete dal punto di vista del carico. Il valore di carico è stato definito come numero di treni che quotidianamente circola sulla rete. E' stata condotta un'analisi a priori sulla rete, volta ad indagare la sua struttura di base, quindi sono state effettuate le stime del carico. Successivamente si è approfondita l'analisi studiando le possibili differenze tra diverse fasce orarie e tra le diverse tratte.

5.1 Analisi a priori

La rete è stata analizzata a priori per conoscerne la struttura e la topologia, nonché per effettuare ipotesi sui possibili risultati delle analisi successive. Per questo task è stato scelto di utilizzare un grafo orientato perchè è stato osservato che non tutti gli archi hanno il reciproco, di conseguenza non si voleva precludere questa parte di informazione.

Tab. 5.1 Struttura della rete

Nodi	Archi	$\min(d_k)$	$\max(d_k)$	\bar{d}	\bar{l}	d	a	r
415	1596	2	52	7.69156	10.43470	0.00928	0.53525	0.90726

Tabella 5.1: Numero di nodi, archi, grado minimo, massimo, medio, shortest path medio, densità, assortatività della rete, reciprocità

Come si può vedere dalla tabella il numero di nodi è nettamente inferiore al numero di archi. La reciprocità è molto elevata 90%, questo spiega la ragione per cui si è scelto di utilizzare archi orientati. È stato deciso di riportare il valore di grado considerando la somma tra l'outdegree e l'indegree perchè è stato osservato che in 78 casi i due valori sono diversi, in questo modo si pensa

ci sia più uniformità. La lunghezza media dello shortest path differenzia molto da quella osservata precedentemente, questo può spiegare che il collegamento tra due stazioni realmente può essere più efficiente, di quello che si ottiene semplicemente seguendo i binari tra esse. Infine è stato osservato come la rete sia assortativa, con un valore ben più elevato dello 0, questo vuol dire che i nodi con grado elevato (hub) tendono ad essere collegati tra loro. È possibile ipotizzare una spiegazione: le stazioni più importanti della rete ferroviaria avranno un collegamento diretto tra esse, mentre le sequenze di stazioni poco importanti su tratti di ferrovia in periferia presenteranno tanti nodi spoke collegati.

Anche per questa rete è stata presa in esame la distribuzione di grado.

Fig. 5.1 Nella prima immagine distribuzione di grado, nella seconda la sua versione normalizzata

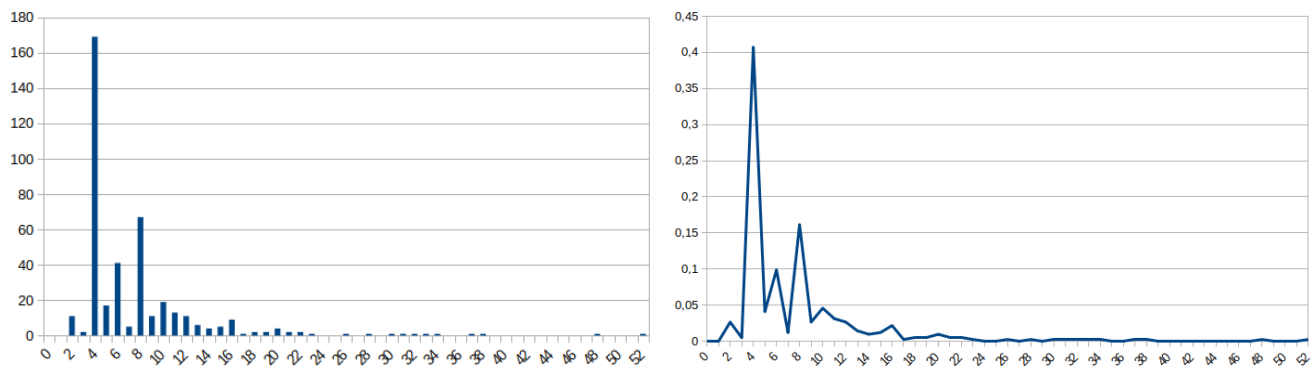


Figura 5.1: Distribuzione di grado

Dalle due immagini è possibile notare come la rete presenti un elevato numero di stazioni con un numero modesto di collegamenti e un numero basso di stazioni con molti collegamenti. Inoltre come già detto precedentemente la tendenza dimostrata dal valore di assortativity è quella che i nodi hub si leghino tra di loro e i nodi spoke facciano altrettanto. Inoltre è stata condotta un'indagine circa l'indegree o l'outdegree, perché è chiaro che ci siano delle differenze, ma si voleva capire quanto fossero marcate queste differenze.

Fig. 5.2 Nella prima immagine distribuzione di grado in ingresso (indegree), nella seconda la sua versione in uscita (outdegree)

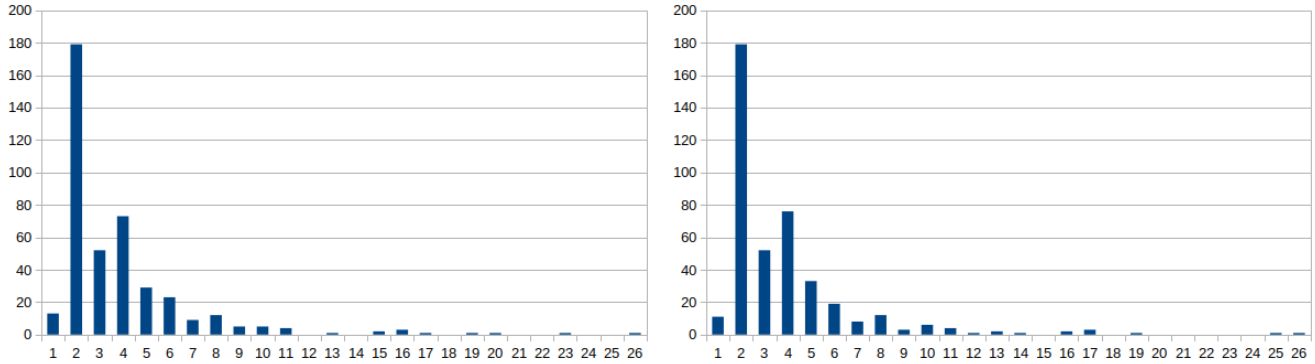


Figura 5.2: Distribuzione di grado in ingresso e uscita

Come è possibile vedere dal confronto tra i grafici le distribuzioni sono ressochè identiche, quindi le variazioni saranno minime. Infatti è possibile ipotizzare che le variazioni siano nell'ordine di 1 o 2, in quanto è possibile che uno tratta entri in una stazione da un'altra, ma poi non riparta più, ma questo fenomeno non può accadere per un numero elevato di tratte.

5.2 Analisi del carico sui nodi della rete

Per analizzare il carico sulla rete, in forma di treni che circolano al giorno, abbiamo utilizzato la rappresentazione due, come spiegato precedentemente, quindi esiste un arco tra due nodi se esiste almeno una tratta che tocca un nodo e poi l'altro. L'arco sarà etichettato con due valori: il nome della tratta e il numero di treni giornalieri. L'arco è orientato, perchè esistono lievi differenze tra ingresso ed uscita. Come prima analisi abbiamo controllato per ogni nodo quale fosse il suo carico in ingresso e il carico in uscita. Il risultato è che questi due valori differiscono di poco per ogni stazione.

Fig. 5.3 Confronto tra i valori di carico in ingresso e in uscita, valori ordinati per carico totale.

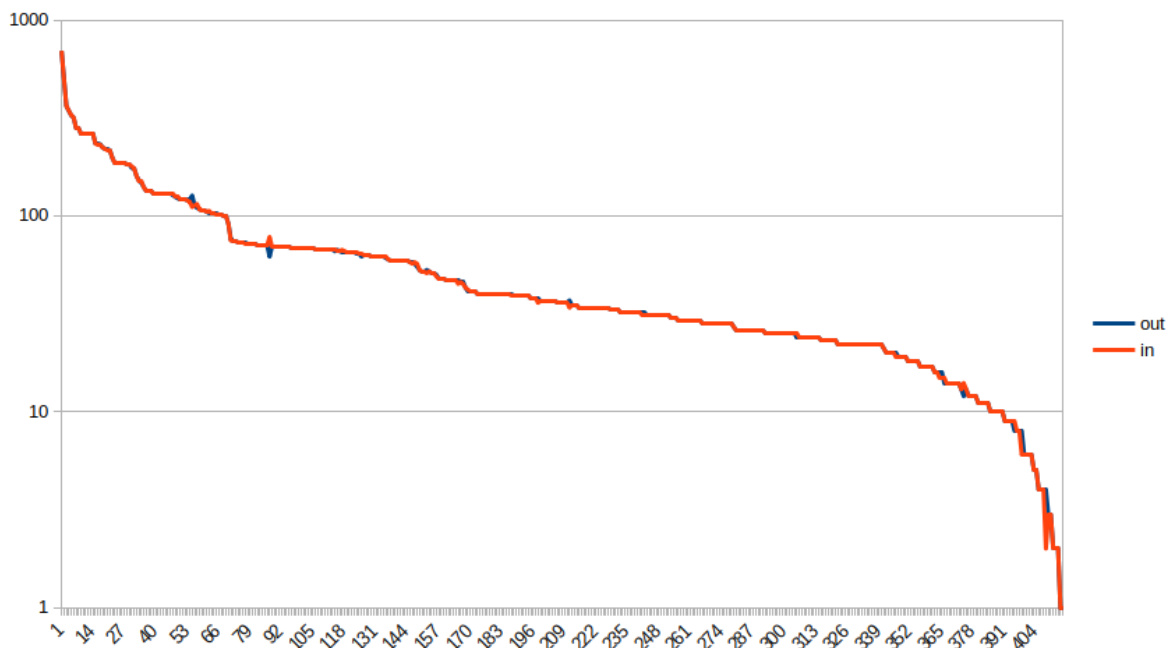


Figura 5.3: Carico in ingresso e uscita

Successivamente abbiamo indagato circa la natura dei nodi che presentano i valori più elevati e meno elevati. In particolare abbiamo indagato la posizione geografica di questi all'interno della rete, e la correlazione con due misure di centralità: degree e betweenness. Questo per verificare se i nodi più importanti nella rete per la circolazione dei treni siano anche quelli attraverso cui circolano più treni o no.

Nell'immagine che segue sono riportati i nodi colorati in base al loro carico. I nodi gialli sono i 50 nodi con carico minore e i nodi blu sono i 50 con carico maggiore.

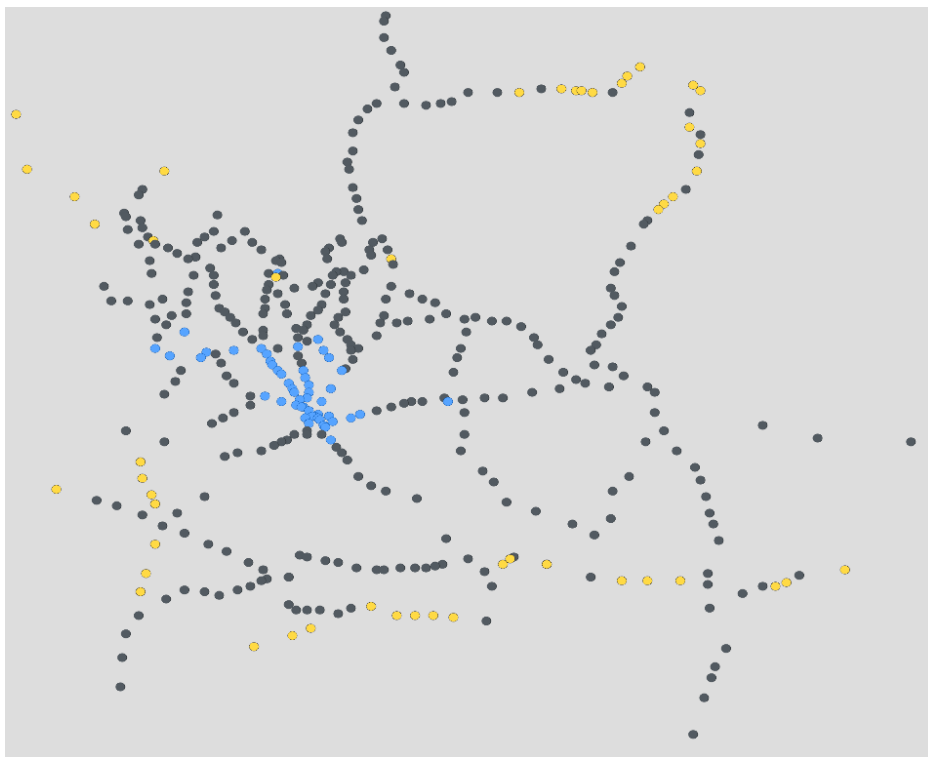


Figura 5.4: Carico dei nodi - disposizione geografica

Come possiamo osservare dalla rete in figura, le stazioni che presentano un carico maggiore sono situate nell'area di Milano e limitrofi. Le stazioni con i carichi minori appartengono alla parte finale di alcune tratte o comunque in zone ai bordi della rete.

Abbiamo ipotizzato di utilizzare due misure di centralità per ricavare i nodi più importanti della rete: centralità di grado e betweenness. La prima permette di ricavare i nodi in base al numero di collegamenti che questi hanno. Nella vita reale si può ipotizzare che una stazione che abbia tanti collegamenti con altre stazioni sia importante, in quanto trafficata sia da treni che da persone. La seconda permette di ricavare i nodi in base a quanti percorsi che collegano diverse stazioni vi passano attraverso. Nella rete reale si può ipotizzare che una stazione attraverso cui passano tante collegamenti tra coppie di stazioni possano essere considerati importanti.

Sono state fatte varie analisi, considerando un subset dei nodi di dimensione 100 per rendere più chiaro e leggibile il grafico. In tutte le seguenti analisi abbiamo utilizzato il carico come somma della dimensione dei nodi in ingresso e in uscita. Come prima analisi abbiamo scelto i primi nodi per carico e abbiamo analizzato il comportamento della betweenness e del grado. In questo primo grafico abbiamo riportato il confronto tra l'andamento del carico sui nodi (arancione) e la centralità di grado del nodo (blu).

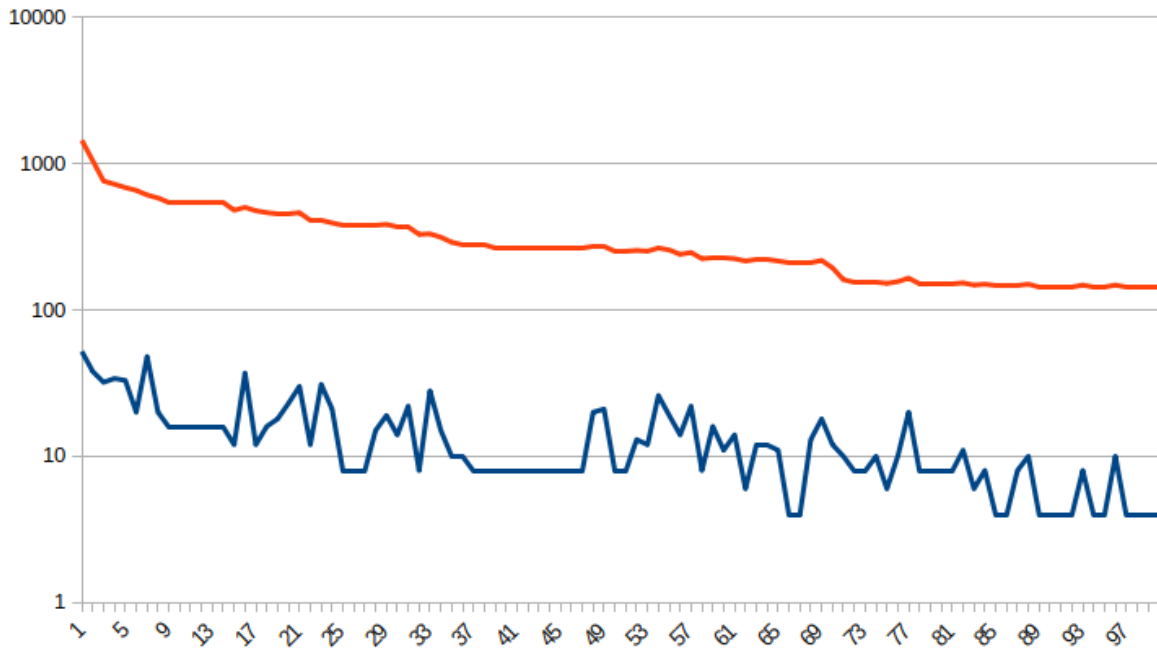


Figura 5.5: Confronto tra l'andamento del carico sui nodi e la centralità di grado

È possibile notare il grado presenta un andamento molto più irregolare rispetto al carico, tuttavia è possibile riconoscere un di andamento simile, almeno a livello asintotico, infatti al decrescere del carico decresce anche il grado. Si può quindi affermare che può esserci una correlazione tra i due valori. Questa ipotesi è confermata dal calcolo del coefficiente di correlazione secondo che è stato ottenuto essere 0.78. Questo risultato corrisponde a quello atteso inizialmente, in quanto è molto probabile che una stazione che ha tanti collegamenti con altre ha anche una più alta circolazione di treni durante la giornata.

Nel grafico successivo viene riportato il confronto tra l'andamento del carico sui nodi e la centralità rispetto alla misura di betweenness.

Fig. 5.6 Confronto tra i valori di carico totale, e centralità rispetto alla betweenness.

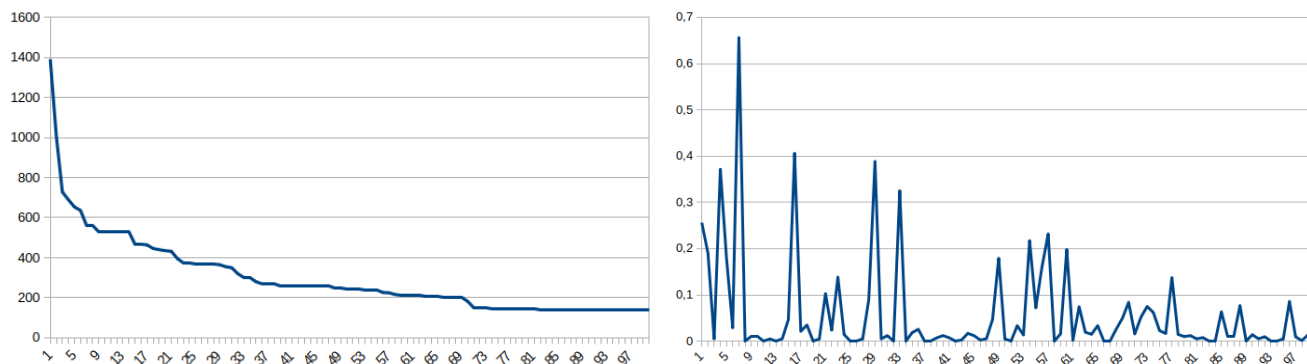


Figura 5.6: Confronto tra l'andamento del carico e e la betweenness dei nodi

Come è possibile osservare il grafico della betweenness presenta un andamento molto irregolare, con parecchi picchi. Risulta difficile fare un confronto con l'andamento molto più regolare del grafico del carico. Analizzando i due valori attraverso il coefficiente di correlazione è emerso che risultano debolmente correlate, infatti il risultato del calcolo è stato: 0.39.

Lo stesso ragionamento è stato effettuato prendendo i primi 100 nodi per grado e betweenness e confronta quindi con l'andamento del carico, ma i risultati ottenuti sono stati identici a quelli presentati sopra.

5.3 Analisi delle tratte

La rete ferroviaria di Trenord è composta da 56 tratte divise in 4 categorie: Regionale, RegionaleExpress, Suburbani e Malpensa Express. Abbiamo voluto analizzare singolarmente il carico sulle varie tratte, quindi le differenze tra le varie categorie. In questa analisi abbiamo anche considerato il numero di stazioni dalla tratta e la distanza (in km) tra la stazione di partenza e di arrivo.

Il valore presenta grandi differenze tra circa 2000 della tratta più trafficata e 24 della meno trafficata, con una media di 447.98 e una deviazione standard di 415.30 a testimoniare questa escursione di valori. Abbiamo indagato circa la correlazione tra il numero di treni che quotidianamente passano per la tratta con il numero di stazioni che compongono la tratta e la distanza chilometrica tra l'inizio e la fine della tratta. Durante la giornata potrebbero variare, di poco, i percorsi dei treni su una tratta, quindi cambia anche il numero di stazioni attraversate e il chilometraggio, abbiamo quindi deciso di utilizzare il valore medio in entrambi i casi per queste analisi.

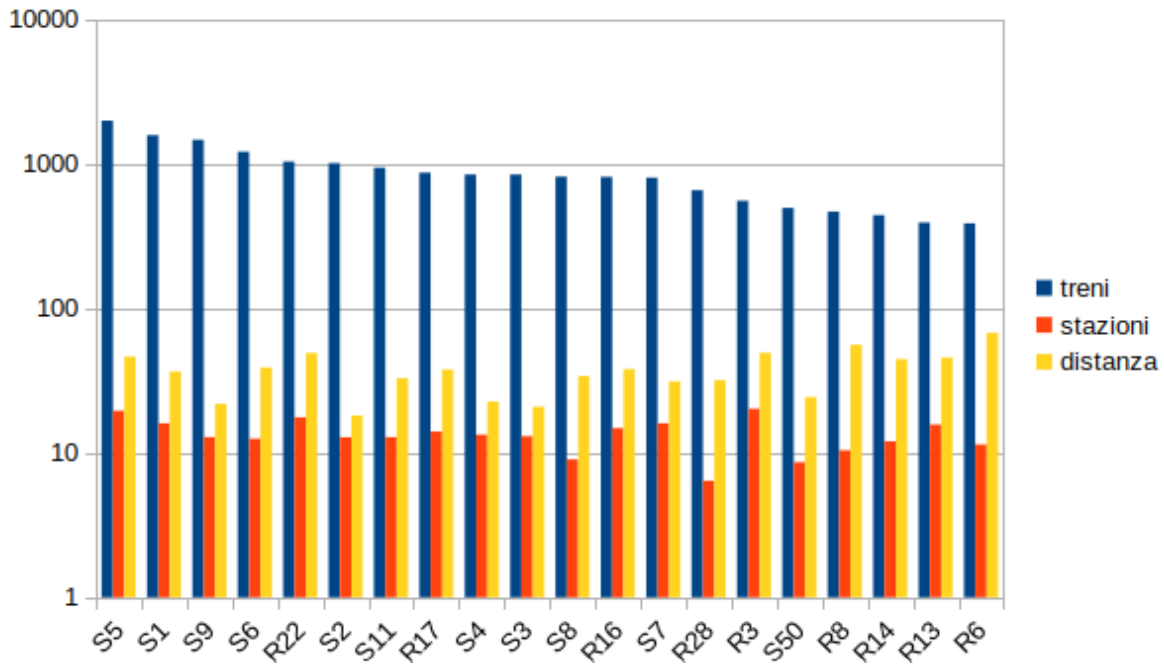


Figura 5.7: Prime 20 tratte per carico con numero di stazioni attraversate e distanza percorsa

Nel grafico abbiamo riportato le prime 20 tratte per carico di treni e questo valore è stato messo a confronto con il numero di stazioni attraversate e la distanza percorsa. Come si può iniziare a vedere sembrerebbe esserci una piccola correlazione tra i valori, in quanto seguono un andamento simile. Sono state quindi effettuate analisi più approfondite mettendo a confronto i tre valori.

I valori sono stati ordinati per carico decrescente e poi è stata confrontata la curva del carico con le altre due.

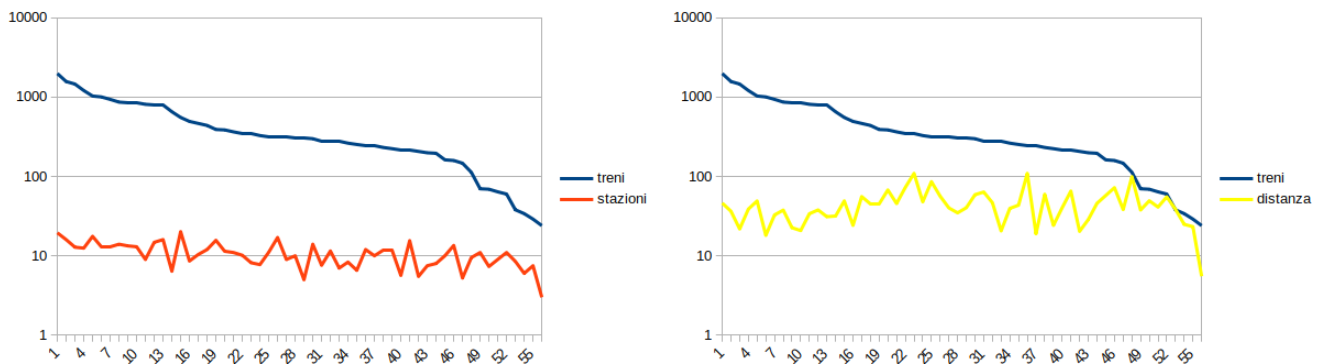


Figura 5.8: Confronto tra carico, numero di stazioni attraversate e distanza

Come si può vedere dal primo grafico l'andamento delle due curve può essere riconducibile ad una forma comune, nonostante il numero di stazioni presenti un andamento molto più irregolare. È possibile ipotizzare che i due valori siano quindi correlati, questa ipotesi è confermata dal valore ottenuto pari a 0.58 come coefficiente di correlazione. Nella realtà questo si può tradurre come l'aspetto che i treni di cui esistono più corse, quindi in teoria anche più battuti dagli utenti fermano anche in un maggior numero di stazioni. Nel secondo grafico sono riportati gli andamenti di distanza percorsa sulla tratta e il carico sulla stessa. In questo caso tra le due curve non è possibile riconoscere alcun andamento comune e soprattutto nella parte finale risulta marcata la differenza. È possibile quindi formulare l'ipotesi che i due valori non siano direttamente proporzionali, confermata dal valore del coefficiente di correlazione: -0.19. Tuttavia questa correlazione risulta essere debole, tendente al valore neutro. Questo aspetto può essere tradotto nel fatto che le stazioni più vicine hanno molti più collegamenti, in quanto spesso sono nell'intorno di grandi città, in particolare Milano, mentre quelle che distano maggiormente hanno meno collegamenti.

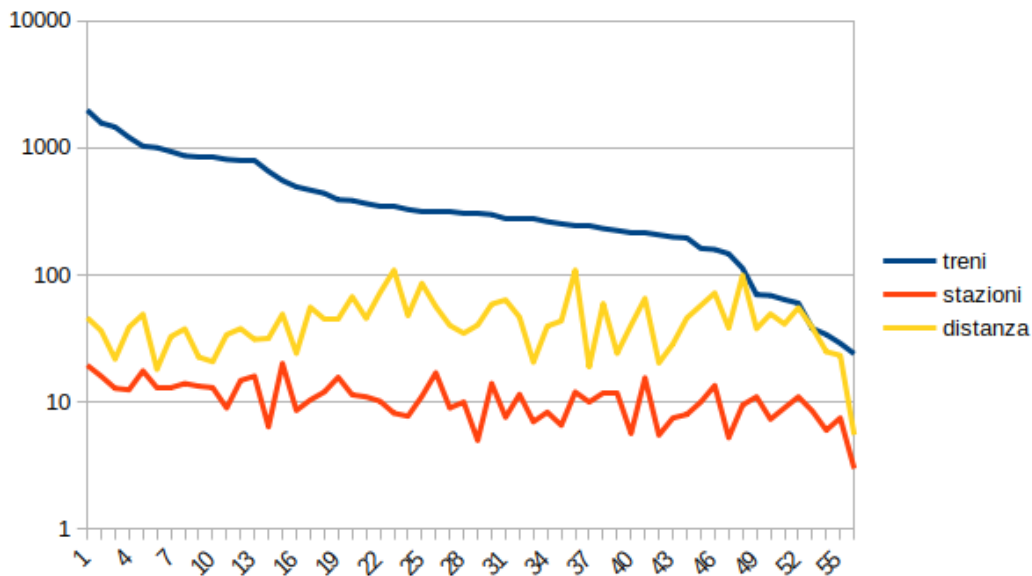


Figura 5.9: Confronto tra carico, numero di stazioni attraversate e distanza su un unico grafico

5.3.1 Categorie di tratte

È possibile dividere le tratte della ferrovia in 4 categorie: Regionale, RegioExpress, Malpensa Express e Suburbano. Ognuno ha caratteristiche diverse e viene usata per scopi diversi. Abbiamo provato a condurre un'analisi volta ad indagare le possibili differenze tra queste.

Tab. 5.3 confronto tra le dimensioni delle varie tratte.

	N°	\bar{c}	$\sigma(c)$	\bar{n}	$\sigma(n)$	\bar{d}	$\sigma(d)$
Regionale	30	337.867	242.422	10.739	3.684	44.519	14.151
RegioExpress	10	214.2	99.481	9.908	3.141	70.206	29.494
Malpensa Express	2	259.5	62.932	5.333	0.471	40.344	0.012
Suburbano	14	877.857	562.997	12.171	3.916	27.786	10.737

Tabella 5.2: Dimensioni delle diverse categorie di tratte

Come è possibile vedere dalla tabella tutte hanno un valore di deviazione standard per il carico molto elevato, mentre quello per numero di stazioni e distanza si mantengono vicine ad un valore più contenuto. Si può quindi ipotizzare che le categorie siano state create accorparendo tratte con distanza e numero di stazioni comuni, mentre esistono tratte con un elevato numero di corse giornaliere e altre meno. Confrontando tra le varie tratte si può osservare che quelle della categoria suburbano hanno valore di carico decisamente più alto, mentre le altre risultano essere abbastanza simili. Mediamente sulle tratte suburbane ci sono anche più stazioni, mentre la distanza è mediamente inferiore. I treni RegioExpress si muovono su distanze elevate. Più della metà dei treni sono quindi Regionali che si muovono su distanze medie e con grandi variazioni interne legate al numero di corse. I treni suburbani si muovono in un raggio ristretto con un altissimo numero di corse giornaliere e un alto numero di fermate, quindi collegano la zona relativa al centro della rete(Milano e limitrofi). Infine i RegioExpress si muovono su un raggio più ampio con meno corse giornaliere e mediamente un numero inferiore di stazioni, collega le periferie della rete e le periferie della rete con il centro.

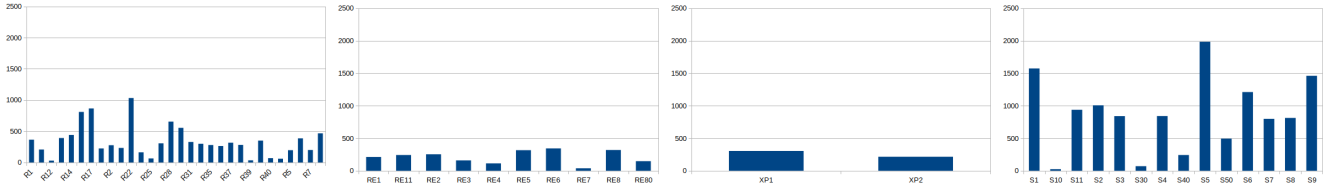


Figura 5.10: Confronto tra il carico sulle varie categorie

Per confrontare la differenza di carico tra le diverse tratte abbiamo riportato questi quattro grafici con la stessa unità di misura l'asse y. Questa immagine conferma quanto osservato precedentemente con i valori medi: il carico sulle tratte della rete suburbano è quello più elevato, anche se esistono differenze vistose tra alcune tratte della categoria. Anche per quanto riguarda le tratte Regionali esiste una grande disparità che è evidenziata anche dal valore di deviazione standard.

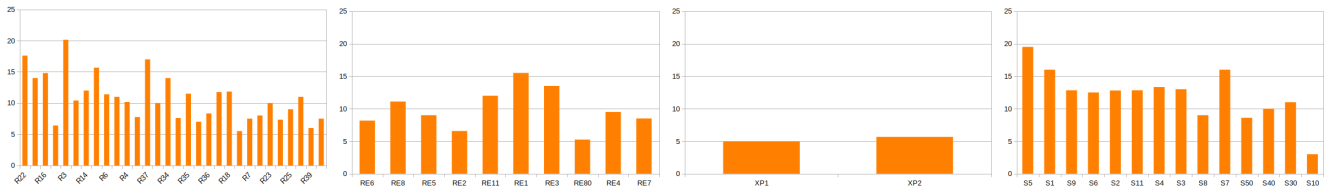


Figura 5.11: Confronto tra il numero di stazione attraversate

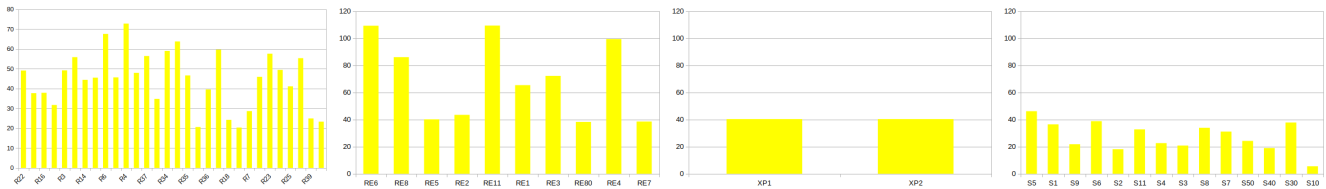


Figura 5.12: Confronto tra la distanza percorsa

Le stesse indagini sono state poi effettuate sia sul numero di stazioni che sulla distanza percorsa dai treni delle varie categorie. Anche qui si trova conferma alle ipotesi fatte precedentemente, in quanto la linea suburbana risulta avere una densità di fermate più elevata, mentre la distanza si riduce. Per quanto riguarda la linea RegioExpress invece avviene il contrario, si riduce il numero di stazioni ma aumenta la distanza percorsa, soprattutto in alcune sue tratte particolarmente lunghe. Mentre per le tratte regionali non sempre esserci alcuna differenza evidente.

Infine è stata condotta un'indagine circa l'andamento delle curve delle tre misure nelle varie categorie per capire se l'andamento segue quello generale o si presentano differenze tra le categorie. Dalle analisi presentate dovrebbero esserci differenze in quanto all'interno delle categorie esistono grandi differenze tra le misure. Per semplicità è stata omessa la categoria Malpensa Express, in quanto è costituita solo da due tratte.

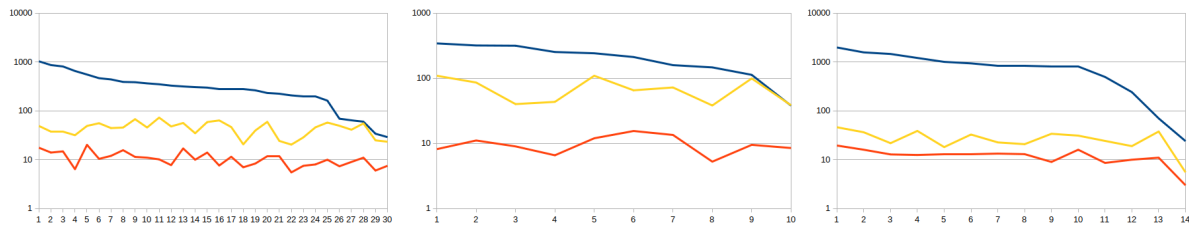


Figura 5.13: Confronto tra l'andamento delle varie misure

Come si può vedere la categoria Regionale (primo grafico) è quella che mantiene un andamento più simile all'andamento generale della rete, questo anche perchè è la più numerosa. La categoria

RegioExpress, invece presenta alcune differenze. La distanza, nonostante l'andamento irregolare sembrerebbe essere più correlata al carico rispetto alla rete generale, mentre sul numero di stazioni non sembra esserci una notevole differenza. Infine la categoria Suburbana è quella che presenta le differenze più marcate, infatti è possibile notare dal grafico come le curve seguano un andamento molto simile. Queste ipotesi sono confermate dal calcolo del coefficiente di correlazione, riportato nella tabella seguente:

	N°	d
Regionale	0.57	0.09
RegioExpress	0.04	0.31
Suburbano	0.79	0.55

Tabella 5.3: Correlazione tra il carico e le altre misure diviso per categorie

Come si può vedere anche dalla tabella la categoria Suburbana è quella che ha un grado di correlazione più elevato tra il carico e le altre due misure prese in considerazione, quindi il carico è legato sia alla distanza della tratta, che al numero di stazioni sulla stessa. Per quanto riguarda la tipologia RegioExpress invece il carico risulta poco correlato al numero di stazioni, mentre risulta correlato debolmente alla distanza, quindi ad una maggiore distanza, corrisponderà un carico di treni maggiore. Infine la categoria regionale non presenta correlazione con la distanza, è stato infatti ricavato un valore neutro, ma presenta una correlazione con il numero di stazioni, quindi all'aumentare del numero di stazioni aumenta il carico sulla rete.

5.4 Analisi delle fasce orarie

Abbiamo creato delle sottoreti della rete di Trenord originale riguardanti le diverse fasce orarie elencate in precedenza. Questo è stato fatto principalmente per raggiungere due obiettivi:

- capire quali fossero i momenti della giornata più o meno serviti,
- identificare le stazioni per le quali, in una certa fascia della giornata, sono presenti pochi treni rispetto al totale nella giornata,
- analizzare i comportamenti delle diverse tratte nelle diverse fasce orarie.

Per quanto riguarda il primo punto abbiamo calcolato il carico totale comprensivo di tutte le città suddiviso per fasce orarie. Il relativo grafico è riportato nella figura che segue.

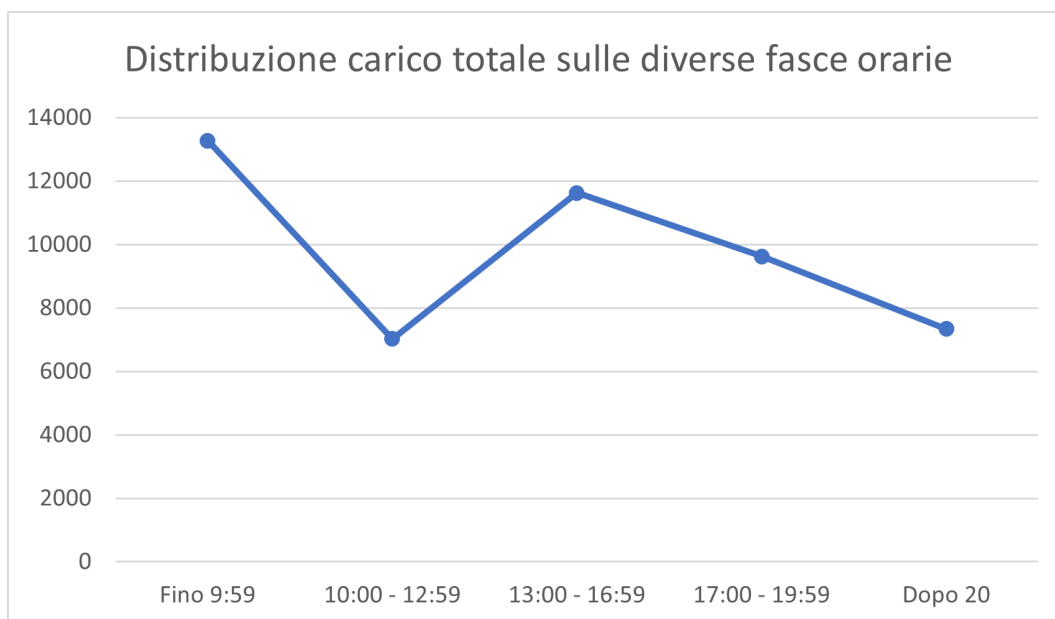


Figura 5.14: Confronto carico totale tra le diverse fasce orarie

Come possiamo notare dal grafico, la fascia oraria che comprende i treni che partono/ arrivano entro le 9:59 è quella che presenta un carico maggiore. Questo ci sembra ragionevole in quanto è in quello il momento della giornata in cui le persone raggiungono il luogo di lavoro o studio. La fascia oraria meno servita è quella che riguarda la tarda mattinata. Come ipotesi iniziale avevamo pensato correttamente che la fascia della mattina potesse essere la più trafficata, insieme a questa però pensavamo fosse la fascia pre-serale (17-19:59). Abbiamo invece ipotizzato di aver ottenuto questi risultati in quanto la fascia della mattina è utilizzata sia da studenti che da lavoratori, mentre per il ritorno queste due categorie si dividono nella fascia pomeridiana e pre-serale.

Per quanto riguarda l'identificazione delle stazioni per le quali, in una certa fascia della giornata, sono presenti pochi treni rispetto al totale nella giornata, abbiamo selezionato le 22 città in cui la problematica risulta più evidente. Abbiamo calcolato la percentuale della differenza tra la fascia oraria con più carico e quella con il minor carico sul totale del carico della giornata nella rispettiva città. Nel grafico che segue, per ognuna di queste 22 città sono riportati i rispettivi carichi nell'arco della giornata.

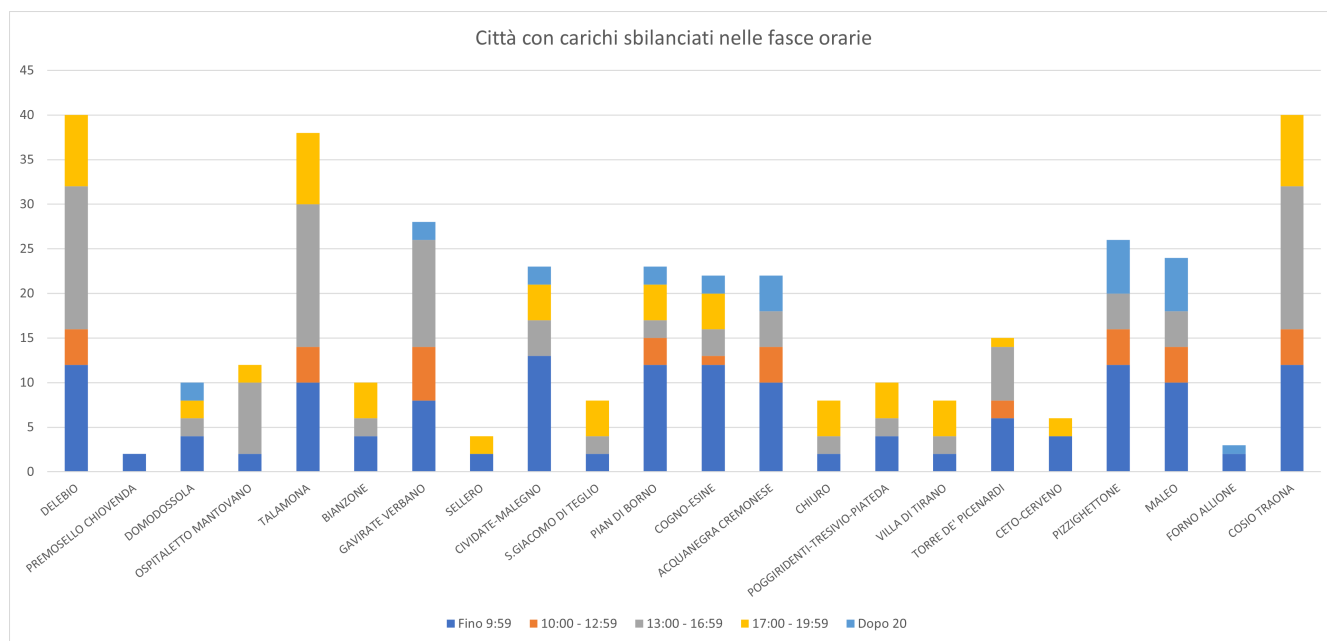


Figura 5.15: Confronto carico sbilanciato in alcune città

Pensiamo possa essere interessante, per chi si occupa di decidere le fermate e le corse dei treni analizzare più in dettaglio questi casi per valutare se effettivamente la richiesta dei passeggeri nelle relative città e momenti della giornata è soddisfatta oppure sarebbe necessaria un'integrazione del servizio ferroviario.

Infine, abbiamo considerato come variano le 50 stazioni più o meno cariche della rete ferroviaria.

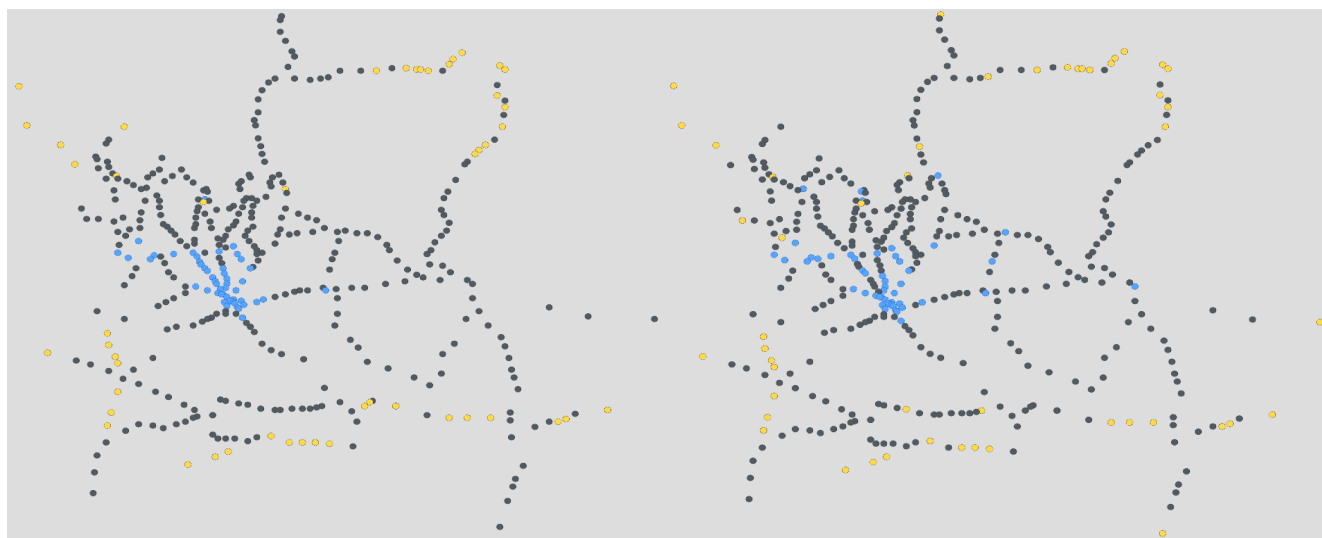


Figura 5.16: Carico stazioni complessivo (sinistra) - Carico stazioni prima 9:59 (destra)

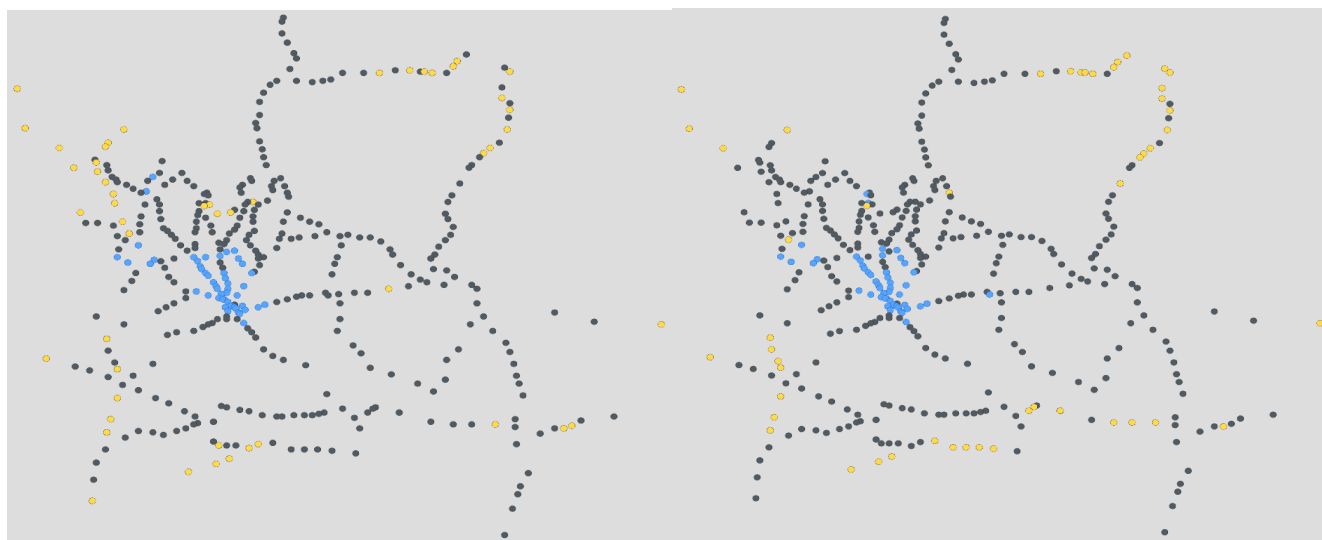


Figura 5.17: Carico stazioni 10:00 - 12:59 (sinistra) - Carico stazioni 13:00 - 16:59 (destra)

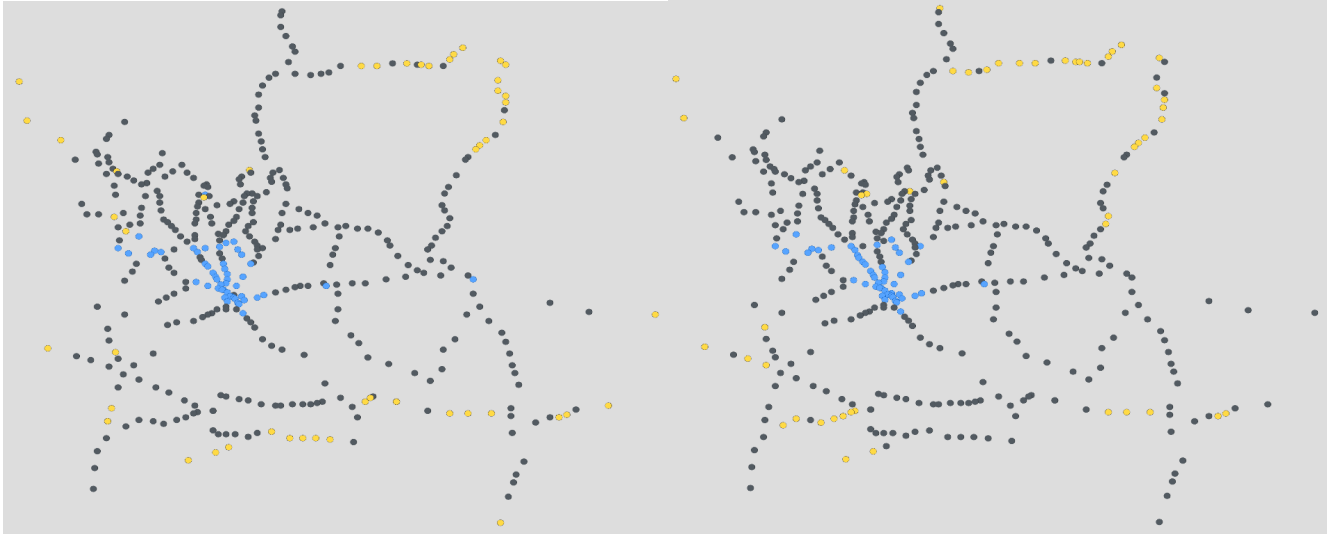


Figura 5.18: Carico stazioni 17:00 - 19:59 (sinistra) - Carico stazioni dopo 20:00 (destra)

In generale, possiamo dire che la zona delle stazioni con carico maggiore rimane quella dell'area milanese. La fascia che presenta una variazione maggiore è quella delle corse prima delle 9:95 che include alcune altre città come Lecco, Brescia, Bergamo, ... Le 50 stazioni con meno carico variano leggermente in base alla fascia considerata, ma rimangono comunque nei bordi della rete ferroviaria.

5.4.1 Analisi delle tratte per fasce orarie

Come ultima analisi abbiamo indagato il comportamento delle diverse tratte in corrispondenza delle fasce orarie. In primo luogo ci siamo occupati dell'andamento delle categorie di tratte (Regionale, RegioExpress, Suburbano e Malpensa Express) sui diversi orari per effettuare un confronto con l'andamento generale della rete.

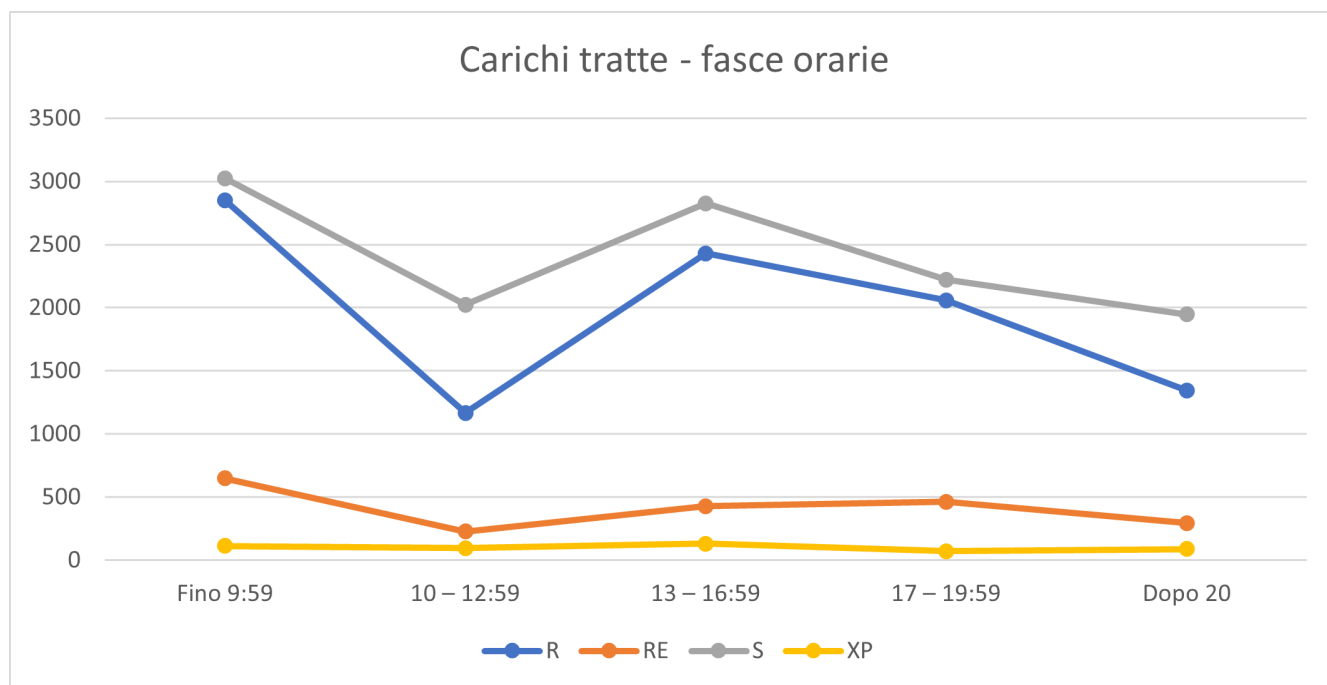


Figura 5.19: Andamento delle categorie di tratte

Come si può vedere dalla figura per le due categorie più numerose (Regionale e Suburbano) l'andamento è uguale a quello della rete globale, infatti il loro apporto è maggiore in termini numerici. Tra loro abbiamo la differenza che la flessione nella fascia della tarda mattinata è più marcata per quanto riguarda i treni regionali e anche nella fascia pre-serale i treni regionali hanno un leggero picco. Le altre due fasce, invece, seguono un andamento più regolare mantendosi più costante durante tutto l'arco della giornata. L'unica differenza per la categoria RegioExpress riguarda la fascia pre-serale dove si può notare un leggero picco verso l'alto.

Successivamente è stata condotta un indagine interna alle varie categorie per osservare se esistono variazioni sulle specifiche tratte.

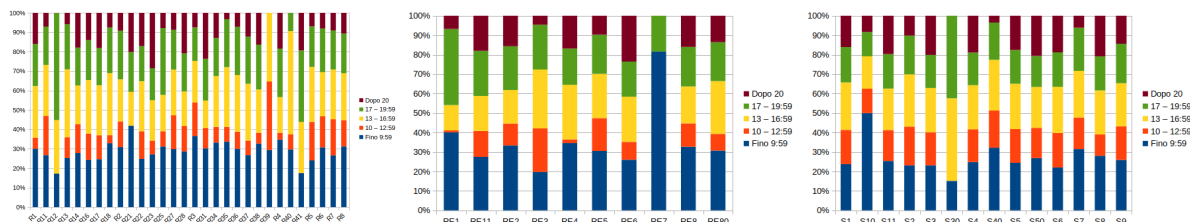


Figura 5.20: Divisione per fasce orarie del carico sulle varie tratte

Nella figura viene mostrato come è ripartito il carico sulle varie fasce orarie per ogni tratta delle tre categorie più numerose: Regionale, RegioExpress e Suburbano. Come si può vedere per la categoria Regionale si ha una netta predominanza della prima fascia, mentre per le altre due questo aspetto è meno marcato. La categoria RegioExpress è quella che segue maggiormente l'andamento atteso con una distribuzione del carico maggiore sulla fascia mattutina e pre-serale. Per la categoria suburbano si ha una predominanza della fascia mattutina e pomeridiana. Per quanto riguarda tutte e tre le categorie si hanno dei valori che non rispecchiano queste riflessioni e presentano valori molto diversi dagli altri e lontani da quelli attesi, potrebbero essere quindi tratte che svolgono un servizio particolare o collegano stazioni con necessità specifiche.

Capitolo 6

Conclusione

6.1 Task 1

6.1.1 Analisi a priori

Da un'analisi a priori della rete è emerso che può avere un comportamento di tipo *scale-free*, ovvero resistere bene ad attacchi di tipo casuale, ma non riuscire a resistere ad attacchi di tipo mirato. Il valore di shortest-path medio elevato ha suggerito la possibilità di avere interruzioni sul servizio anche con la caduta di pochi nodi, tuttavia pensiamo che questo sia un risultato della specifica topologia della rete, in quanto in caso della caduta di una stazione non è possibile per un treno attraversarla. Allo stato iniziale la rete è connessa, quindi il rapporto con la giant component è pari ad 1. Il valore di $s(g)$ potrebbe suggerire l'assenza di colli di bottiglia o ponti, però il valore di $e(g)$ e λ_2 suggeriscono una rete poco connessa e connessa nella maniera errata. Anche questo risultato è sicuramente dovuto alla peculiarità di questa rete di trasporto pubblico che segue la struttura del luogo in cui è ubicata e non può saltare nodi ma è obbligata a seguirne la sequenza.

6.1.2 Strategie d'attacco

Per quanto riguarda gli attacchi di tipo statico, possiamo concludere che, complessivamente, i più efficaci, dal punto di vista di riduzione dell'efficienza, sono quelli mirati basati sulla degree centrality. Questo perché con la rimozione di un numero abbastanza basso di stazioni provocano una diminuzione dell'efficienza più repentina. Inoltre, gli attacchi mirati sono più efficaci rispetto a quelli random. Questo perché, essendoci diverse stazioni poco importanti e meno stazioni molto importanti, con un attacco random la probabilità di selezionare un nodo meno rilevante per l'eliminazione è più elevata, quindi la riduzione dell'efficienza e della GCC sarà minore. Al contrario,

in un attacco mirato, partendo dalla rimozione dei nodi più importanti la rete viene danneggiata in modo più significativo.

Per quanto riguarda i fallimenti a cascata, possiamo concludere che la rete risulta maggiormente danneggiata se a generare il fallimento è il nodo che presenta una *betweenness* maggiore piuttosto che un nodo random. Selezionando un nodo random potrebbe anche capitare che il fallimento non si propaghi al resto della rete, ma rimanga locale.

6.2 Task 2

Come abbiamo visto dalla presentazione dei risultati il carico in termini di treni sulla rete non è bilanciato, in quanto esistono nodi con un elevato carico e altri con un carico modesto. Questo risultato però è perfettamente in linea con l'utilizzo della rete, in quanto non tutte le aree o le stazioni hanno lo stesso bisogno di collegamento. Come è stato visto in particolare le stazioni più stressate dal punto di vista del carico sono quelle che geograficamente sono al centro, quindi nella zona di Milano. Mentre quelle che hanno un carico inferiore risultano ai lati del grafo. Il carico sui nodi, come è facilmente intuibile è strettamente legato al grado dello stesso, in quanto una stazione che ha maggior collegamenti con le altre sicuramente avrà anche un carico più elevato, mentre non è legato all'importanza della stazione, infatti non c'è molta correlazione con il valore di *betweenness* che identifica le stazioni più importanti nello svolgere il ruolo di ponte tra le varie parti della rete. Per quanto riguarda l'analisi della tratte è stato osservato che generalmente il carico è direttamente proporzionale al numero di stazioni attraversate durante il percorso, mentre è inversamente proporzionale al numero di chilometri percorsi durante lo stesso; infatti per tratte lunghe immaginiamo che ci siano meno collegamenti giornalieri, mentre le tratte brevi soprattutto nell'intorno di Milano hanno molti più collegamenti e molte più stazioni da attraversare. L'analisi delle singole categorie di tratte ha permesso di identificare inoltre i ruoli distinti svolti da ognuna di esse: i treni Regionali riguardano vari collegamenti con varie distanze e numero di stazioni, sono la maggior parte delle tratte; i RegioExpress riguardano treni a lunga percorrenza con un modesto numero di stazioni e corse giornaliere, quindi riguardano treni che collegano luoghi importanti ma distanti; i Suburbani hanno percorsi più brevi con un numero maggiore di stazioni attraversate e un elevato carico giornaliero, sono quindi i treni di collegamento tra Milano e l'hinterland. Infine l'analisi delle fasce orarie ha permesso di confermare alcune delle ipotesi iniziali, infatti la fascia più trafficata è risultata essere quella della prima mattina (prima delle 10), mentre per quanto riguarda la fascia pre-serale (17-19:59) ha disatteso le aspettative iniziali. Una possibile motivazione potrebbe essere identificata nel fatto che la fascia mattina è utilizzata sia da studenti che da lavoratori che vengono poi invece suddivisi nel rientro a casa: studenti più sul primo pomeriggio e lavoratori nella fascia pre-serale. L'analisi delle tratte in corrispondenza delle diverse fasce orarie ha permesso di confermare quanto affermato sulla rete nel suo complesso.

Bibliografia

- [Berge et al., 2010] Berge, B., von Ferber, C., Holovatch, T., and Holovatch, Y. (2010). Public transport networks under random failure and directed attack.
- [Berdica, 2002] Berdica, K. (2002). An introduction to road vulnerability: what has been done, is done and should be done. *Transport Policy*, 9(2):117–127.
- [Candelieri et al., 2019] Candelieri, A., Galuzzi, B. G., Giordani, I., and Archetti, F. (2019). Vulnerability of public transportation networks against directed attacks and cascading failures. *Public Transport*, 11(1):27–49.
- [Latora and Marchiori, 2001] Latora, V. and Marchiori, M. (2001). Phys Rev Lett Efficient behavior of small-world networks. *Phys Rev Lett*, 87(19):198701.
- [von Ferber et al., 2009a] von Ferber, C., Holovatch, T., and Holovatch, Y. (2009a). Attack vulnerability of public transport networks. In Appert-Rolland, C., Chevoir, F., Gondret, P., Lassarre, S., Lebacque, J.-P., and Schreckenberg, M., editors, *Traffic and Granular Flow '07*, pages 721–731, Berlin, Heidelberg. Springer Berlin Heidelberg.
- [von Ferber et al., 2009b] von Ferber, C., Holovatch, T., Holovatch, Y., and Palchykov, V. (2009b). Public transport networks: empirical analysis and modeling. *The European Physical Journal B*, 68(2):261–275.
- [Zou et al., 2013] Zou, Z., Xiao, Y., and Gao, J. (2013). Robustness analysis of urban transit network based on complex networks theory. *Kybernetes*, 42(3):383–399.