



Manuale Utente

Gruppo QuaranTeam - Progetto HD Viz

quaranteam2021@gmail.com

INFORMAZIONI SUL DOCUMENTO

Versione	1.0.0
Approvatore	Veronese Luca
Redattori	Gibellato Alice Sinigaglia Matteo
Verificatori	Mason Damiano Chiarello Federico
Stato	Approvato
Uso	Esterno
Destinato a	Prof. Vardanega Tullio Prof. Cardin Riccardo Zucchetti S.p.A. QuaranTeam

Descrizione

Manuale utente per l'installazione e l'utilizzo di *HD Viz* realizzato dal gruppo *QuaranTeam*.

Registro delle modifiche

Versione	Data	Descrizione	Autore	Verificatore
1.0.0	2021-05-10	Approvazione del documento	Veronese Luca	
0.6.0	2021-05-04	Stesura §A.	Gibellato Alice, Sinigaglia Matteo	Chiarello Federico
0.5.0	2021-05-03	Stesura §4.	Sinigaglia Matteo	Chiarello Federico
0.4.0	2021-05-02	Stesura §3.	Sinigaglia Matteo	Mason Damiano
0.3.0	2021-04-28	Stesura §2.	Gibellato Alice	Chiarello Federico
0.2.0	2021-04-24	Stesura §1.	Gibellato Alice	Mason Damiano
0.1.0	2021-04-24	Creazione del documento.	Veronese Luca	Chiarello Federico

Indice

1	Introduzione	1
1.1	Scopo del documento	1
1.2	Scopo del prodotto	1
1.3	Glossario	1
1.4	Riferimenti	1
1.5	Supporto tecnico	2
2	Installazione	3
2.1	Requisiti	3
2.2	Installazione e avvio	3
2.2.1	Preliminari	3
2.2.2	Avvio del client	4
2.2.3	Configurazione ed avvio del server (opzionale)	4
3	Guida all'utilizzo	6
3.1	Operazioni sui grafici	6
3.1.1	Inserimento	6
3.1.2	Modifiche al grafico	8
3.1.2.1	Modifica titolo	8
3.1.2.2	Interazioni con il grafico	8
3.1.3	Salvataggio di un grafico	9
3.1.4	Rimozione di un grafico	9
3.2	Operazioni sul database	10
3.2.1	Importazione dati	10
3.2.2	Aggiunta di un dataset	10
3.2.3	Eliminazione di un dataset	10
4	Componenti e Funzionalità	12
4.1	Caricamento dati	12
4.2	Scelta di una riduzione dimensionale _G	13
4.3	Scelta di una funzione per il calcolo delle distanze	14
4.4	Funzionalità dei grafici	14
4.4.1	Force Field	14
4.4.1.1	Modalità di interazione	14
4.4.1.2	Parametri della visualizzazione	15
4.4.2	Heat Map	15
4.4.2.1	Modalità di interazione	16
4.4.2.2	Parametri della visualizzazione	16
4.4.3	Scatter plot Matrix	16
4.4.3.1	Modalità di interazione	16
4.4.4	Proiezione Lineare Multi Asse	17
4.4.4.1	Modalità di interazione	17
A	Glossario	19

Elenco delle figure

1	Importazione di un dataset	6
2	Inserimento di un grafico	7
3	Modifica al nome di un grafico	8
4	Importazione di un dataset dal database	10
5	Aggiunta di un dataset al database	11
6	Eliminazione di un dataset dal database	11
7	Modifica parametri Force Field	15
8	Modifica parametri Heat Map	16
9	Interazione con Scatter plot Matrix	17
10	Interazione con Proiezione Lineare Multi Asse	18

Elenco delle tabelle

2	Manipolazioni possibili per ogni grafico	7
3	Interazioni possibili per ogni grafico	8
4	Opzioni disponibili per ogni riduzione dimensionale	13

1 Introduzione

1.1 Scopo del documento

Il presente documento rappresenta il manuale utente per l'applicazione web *HD Viz*. Vengono descritte dettagliatamente tutte le caratteristiche dell'applicativo utilizzabili dall'utente.

Tale manuale è suddiviso in sezioni:

- in §2 vengono riportate le istruzioni per l'installazione;
- in §3 vengono descritte le operazioni principali eseguibili dall'utente di *HD Viz*;
- in §4 il focus è rivolto verso le componenti visibili dell'applicazione, quindi vengono descritte le loro caratteristiche, le loro funzionalità e le loro modalità di interazione.

Il manuale sarà consultabile alla sezione "Aiuto" accessibile dal menu della pagina principale di *HD Viz*.

1.2 Scopo del prodotto

HD Viz è un'applicazione web avente scopo di fornire uno strumento per la visualizzazione di dati con molte dimensioni a supporto della fase esplorativa dell'analisi dei dati. *HD Viz* è in grado di rappresentare dati che possono avere almeno 15 dimensioni e fornisce 4 diversi tipi di visualizzazione a tale scopo.

1.3 Glossario

Il documento contiene termini che possono presentare significati ambigui. Viene quindi fornito un glossario individuabile nell'appendice §A, all'interno del documento, contenente tutti i termini definiti ambigui e la loro spiegazione. Nel documento vengono identificati con una G a pedice.

1.4 Riferimenti

Per una trattazione più approfondita di funzioni matematiche, algoritmi e grafici sono state selezionate le seguenti fonti:

- **Funzioni di riduzione dimensionale_G:**
 - FASTMAP:
<https://dl.acm.org/doi/10.1145/568271.223812>
 - ISOMAP:
<https://science.sciencemag.org/content/290/5500/2319>
 - t-SNE:
<https://www.jmlr.org/papers/v9/vandermaten08a.html>
 - LLE:
<https://science.sciencemag.org/content/290/5500/2323>
 - UMAP:
<https://arxiv.org/abs/1802.03426>
- **Funzioni di calcolo delle distanze:**

- Distanza_G di Canberra:
https://en.wikipedia.org/wiki/Canberra_distance
- Distanza_G Euclidea e Euclidea Quadrata:
<https://xlinux.nist.gov/dads/HTML/euclidndstnc.html>
- Distanza_G di Chebyshev:
https://en.wikipedia.org/wiki/Chebyshev_distance
- Distanza_G Cosine:
<https://www.itl.nist.gov/div898/software/dataplot/refman2/auxillar/cosdist.htm>
- Distanza_G di Manhattan:
<https://xlinux.nist.gov/dads/HTML/manhattanDistance.html>

- **Grafici:**

- Force Field:
<https://observablehq.com/@d3/force-directed-graph>
- Heat Map:
<https://observablehq.com/@bstaats/matrix-diagram>
- Scatter plot Matrix:
<https://observablehq.com/@d3/scatterplot-matrix>
- Proiezione Lineare Multi Asse:
<https://orange3.readthedocs.io/projects/orange-visual-programming/en/latest/widgets/visualize/freeviz.html>

1.5 Supporto tecnico

Qualora venissero riscontrati bug_G o malfunzionamenti, si prega di segnalarli per mail all'indirizzo:

quaranteam2021@gmail.com

È preferibile l'utilizzo della seguente struttura:

Oggetto: HD Viz - bug - (Evento da segnalare);

Allegato (opzionale): è possibile allegare immagini o video esplicativi;

Corpo:

- browser_G e sistema operativo utilizzati;
- descrizione del problema;
- versione di HD Viz utilizzata.

2 Installazione

2.1 Requisiti

Sistema operativo

- Windows 10;
- Ubuntu 20.04 o superiore, o distribuzioni Linux derivate.

Browser

- **Mozilla Firefox** v85.0 o superiore;
- **Google Chrome** v87.0 o superiore.

È consigliabile avere impostato uno dei due browser_G elencati come browser_G predefinito, perché la procedura indicata in §2.2.2 (Installazione - Avvio del client) apre in automatico una pagina del browser_G predefinito. È comunque possibile aprire l'applicazione da un browser_G non predefinito in seguito, al seguente URL:

`http:localhost/<numero_porta>`

Non è garantita la piena compatibilità per browser_G derivati dai suddetti o per versioni precedenti dei suddetti.

Requisiti software

- **Node.js_G** 14.16.0 o superiore;
- **npm_G** 7.6.3 o superiore;
- **PostgreSQL_G** 13.2 o superiore (opzionale¹).

Requisiti hardware

- **RAM:** necessari almeno 4 GB, consigliati almeno 8 GB;
- **connessione a Internet.**

2.2 Installazione e avvio

2.2.1 Preliminari

1. Aprire la pagina del repository_G GitHub_G di *HD Viz* al seguente link:

<https://github.com/QuaranTeam2021/HD-Viz>

2. Selezionare **Code > Download Zip**;
3. Salvare l'archivio *.zip* ed estrarre i file nella locazione desiderata;
4. Aprire la cartella *HD-Viz-main*.

¹necessario solo per il collegamento con il database_G

In alternativa, in un sistema con `wget` e `unzip` installati² aprire il terminale³ e digitare le seguenti istruzioni:

```
wget https://github.com/QuaranTeam2021/HD-Viz/archive/main.zip
unzip main.zip
cd HD-Viz-main
```

2.2.2 Avvio del client

Una volta posizionati all'interno della cartella `HD-Viz-main`:

1. Aprire la cartella `client` e aprire un terminale posizionato nella cartella `client` (se non già aperto);
2. Se si cerca di avviare per la prima volta *HD Viz*, digitare il comando:

```
npm install
```

3. Eseguire la `buildG` dell'applicazione con il comando:

```
npm run build
```

4. Se si cerca di avviare per la prima volta *HD Viz*, digitare il comando:

```
npm install -g serve
```

5. Digitare infine il comando:

```
serve -s build -l 3000
```

6. Attendere l'apertura automatica del browser_G predefinito.

2.2.3 Configurazione ed avvio del server (opzionale)

Il server_G viene sfruttato per interagire con un database_G esterno, pertanto la sua installazione ed esecuzione è opzionale. Il database_G utilizzato è PostgreSQL, scaricabile al seguente link:

<https://www.postgresql.org/download/>

Una volta posizionati all'interno della cartella `HD-Viz-main`:

1. Aprire la cartella `server` e aprire un terminale posizionato nella cartella `server` (se non già aperto);
2. Creare un file `.env` rispettando la seguente notazione:

```
USER = postgres
PASSWORD = postgres
HOST = localhost
PORT = 5432
DATABASE = hdviz
```

²installati di default in ogni distribuzione Linux, necessaria installazione manuale per Windows

³terminale per Linux o Windows PowerShell per Windows

Inserire come valore dei campi USER, PASSWORD e DATABASE rispettivamente il nome utente, la password e il nome del database_G a cui ci si vuole connettere.

3. Se si cerca di avviare per la prima volta la componente server_G di *HD Viz*, digitare il comando:

```
npm install
```

4. Avviare il server_G con il seguente comando:

```
node server
```

Il server_G sarà in esecuzione alla porta 5000.

3 Guida all'utilizzo

3.1 Operazioni sui grafici

3.1.1 Inserimento

È possibile importare un dataset dalla memoria locale o dal database_G, scegliendo tra i due pulsanti mostrati in figura 1.

1.a) Cliccare sul pulsante "IMPORTA" e scegliere il file .csv_G o .tsv_G da cui prelevare i dati.

Per ulteriori informazioni su caratteristiche obbligatorie dei file .csv_G o .tsv_G si rimanda a §4.1 - Caricamento Dati.

Oppure

1.b) Cliccare sul pulsante "DATABASE". Nella finestra che si aprirà saranno presenti i seguenti due menù a tendina:

- **Tabelle:** in cui è possibile selezionare uno dei dataset presenti nel database_G;
- **Colonne:** in cui è possibile selezionare le dimensioni del dataset selezionato che si vogliono importare.

Per ulteriori dettagli sull'importazione di dataset dal database_G si rimanda a §3.2.1 - Importazione Dati.



Figura 1: Importazione di un dataset

Una volta importato un dataset, è possibile selezionare una visualizzazione ed eventualmente impostare le opzioni di manipolazione dati disponibili, attraverso l'interfaccia di figura 2:

- 2) Selezionare una tra le possibili visualizzazioni;
- 3) Selezionare una o più colonne da visualizzare;
- 4) Selezionare una colonna per il raggruppamento;
- 5) Selezionare le opzioni per la manipolazione dati. Tali opzioni dipendono dal grafico selezionato al punto 2. La gamma di opzioni presenti per la manipolazione dati per ogni grafico è riassunta nella tabella 2.
- 6) Premere il pulsante "CONFERMA" per aggiungere il grafico alla pagina di visualizzazione ed essere rimandati a tale pagina.

Grafico:

☒ Scatterplot Matrix

☐ Scatterplot

☐ HeatMap

☐ Force Field

☐ Proiezione Multiassi


Colonne:

☒ sepal_length

☒ sepal_width

☒ petal_length

☐ petal_width

Colonna grouper: 

☒ species

Algoritmo:

☒ Nessuna riduzione

☐ UMAP

☐ FASTMAP

☐ ISOMAP

☐ T-SNE

☐ LLE

CONFERMA

Figura 2: Inserimento di un grafico

Tabella 2: Manipolazioni possibili per ogni grafico

	Scatter plot Matrix	Force Field	Heat Map	Proiezione Lineare Multi Asse
Visualizzazione di una riduzione dimensionale _G	Sì	No	No	Sì
Visualizzazione di una matrice delle distanze	No	Sì	Sì	No
Visualizzazione di dati non manipolati	Sì	No	No	Sì

3.1.2 Modifiche al grafico

3.1.2.1 Modifica titolo

- 1) Cliccare sul pulsante con l'icona di modifica (figura 3.a);
- 2) Inserire il nome desiderato per il grafico;
- 3) Premere l'icona di conferma (figura 3.b).

A questo punto il nuovo titolo viene visualizzato. È possibile modificare nuovamente il titolo.

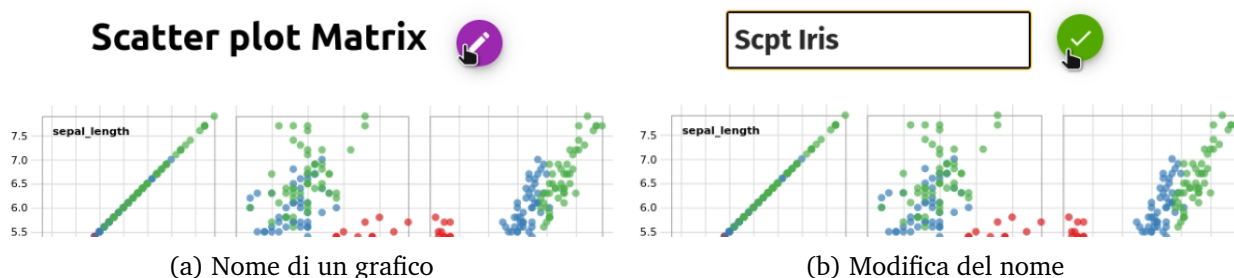


Figura 3: Modifica al nome di un grafico

3.1.2.2 Interazioni con il grafico

Uno dei punti di forza di *HD Viz* è la presenza di grafici *dinamici*, che rispondono alle interazioni con l'utente. Le opzioni dipendono dallo specifico tipo di grafico, pertanto si rimanda alla sezione §4.4 che illustra in dettaglio le caratteristiche e le funzionalità presenti per ogni grafico. Un riassunto delle interazioni possibili è il seguente:

Tabella 3: Interazioni possibili per ogni grafico

	Scatter plot Matrix	Force Field	Heat Map	Proiezione Lineare Multi Asse
Evidenziamento di dati	Sì	No	No	No
Aggiunta dimensione	Sì	No	No	Sì
Rimozione dimensione	Sì	No	No	Sì
Ordinamento degli elementi	Sì	No	Sì	No
Trascinamento di un nodo	No	Sì	No	No
Modifica della funzione di forza applicata	No	Sì	No	No

3.1.3 Salvataggio di un grafico

È possibile salvare un grafico come immagine (non animata) in formato `.svgG`.

- 1) Cliccare sul pulsante "SALVA";
- 2) Selezionare il percorso nel quale salvare il file e assegnare un nome al file;
- 3) Premere conferma.

A questo punto è possibile aprire il file con un visualizzatore di immagini o con un browser_G. Entrambi i browser_G consigliati supportano la funzionalità di apertura di un file in formato `.svgG`.

3.1.4 Rimozione di un grafico

È possibile rimuovere un grafico dalla lista dei grafici visualizzati:

- 1) Posizionarsi nel grafico;
- 2) Premere il pulsante "RIMUOVI".

3.2 Operazioni sul database

3.2.1 Importazione dati

Dopo aver cliccato sul pulsante "DATABASE" come modalità di importazione dei dati scelta, si aprirà una finestra con i seguenti due menù a tendina (figura 4):

- **Tabelle:** in cui è possibile selezionare uno dei dataset presenti nel database;
- **Colonne:** in cui è possibile selezionare le dimensioni del dataset selezionato che si vogliono importare.

Nel caso in cui non vengano selezionate delle colonne tramite il menù a tendina "Colonne", verrà importato l'intero dataset con tutte le dimensioni presenti nel database_G.

Al termine delle selezioni è sufficiente cliccare il pulsante "SCEGLI QUESTI DATI" per completare l'operazione di importazione.

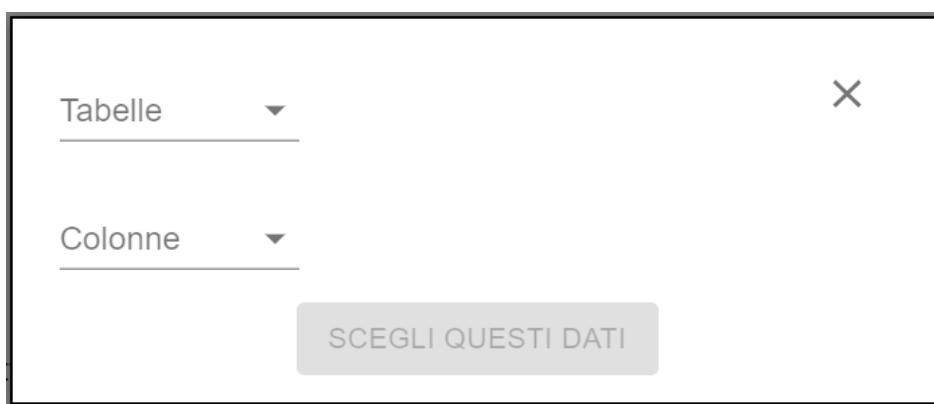


Figura 4: Importazione di un dataset dal database

3.2.2 Aggiunta di un dataset

Accedendo alla sezione "Gestisci Database" è possibile caricare i dati contenuti in un file CSV_G o TSV_G nel database_G esterno tramite i seguenti passaggi (figura 5):

1. Cliccare sul pulsante di aggiunta (+) e scegliere il file .csv_G o .tsv_G contenente i dati che vogliamo caricare nel database_G. Per ulteriori informazioni su caratteristiche obbligatorie dei file .csv_G o .tsv_G si rimanda a §4.1;
2. Inserire il nome con cui si vuole salvare il dataset nel database_G nell'area di testo apposita. Nel caso in cui non venga esplicitamente inserito un nome verrà utilizzato il nome del file caricato;
3. Confermare l'operazione premendo il tasto "INVIA".

3.2.3 Eliminazione di un dataset

Accedendo alla sezione "Gestisci Database" è possibile eliminare un dataset contenuto nel database_G tramite i seguenti passaggi (figura 6):



A form for adding a dataset. It features a blue plus sign icon on the left. To its right is a text input field with a light gray border. Above the input field is the label "Nome" in a small, light gray font. Inside the input field, the word "iris" is typed in a dark gray font. To the right of the input field is a purple rectangular button with the word "INVIA" in white, uppercase letters.

Figura 5: Aggiunta di un dataset al database

1. Cliccare sul pulsante con l'icona di eliminazione situato a fianco del nome del dataset che vogliamo eliminare;
2. Confermare l'operazione cliccando il tasto "CONFERMA".
Alternativamente annullare l'operazione cliccando il tasto "ANNULLA".



Figura 6: Eliminazione di un dataset dal database

4 Componenti e Funzionalità

4.1 Caricamento dati

L'utente preme sul pulsante importa file e seleziona un file con estensione `.csvG` o `.tsvG`. Questo file deve contenere *obbligatoriamente*:

- **una e una sola** riga di intestazione;
- **una o più** righe di dati.

Un esempio di file `.csvG` ben formato è il seguente:

Listing 1: formattazione corretta CSV

```
sepal_length,sepal_width,petal_length,petal_width,species
5.1,3.5,1.4,0.2,setosa
4.9,3.0,1.4,0.2,setosa
4.7,3.2,1.3,0.2,setosa
4.6,3.1,1.5,0.2,setosa
5.0,3.6,1.4,0.2,setosa
```

Un esempio di file `.tsvG` ben formato è il seguente:

Listing 2: formattazione corretta TSV

sepal_length	sepal_width	petal_length	petal_width	species
5.1	3.5	1.4	0.2	setosa
4.9	3.0	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa
5.0	3.6	1.4	0.2	setosa

Il contenuto del file in input deve sottostare alle seguenti regole:

- il carattere delimitatore per un file `.csvG` deve essere una virgola ',' o un punto e virgola ';' ;
- il carattere delimitatore per un file `.tsvG` deve essere una tabulazione;
- il delimitatore per le cifre decimali deve essere un punto;
- non deve essere presente un delimitatore per raggruppare le migliaia. Ciò implica che:
 - 10000 è accettato e considerato come 10^4 ,
 - 10.000 è accettato e considerato come 10^1 ,
 - 10'000 non è accettato,
 - 10,000 viene interpretato come il contenuto di due celle contigue;
- ogni colonna deve contenere dati dello stesso tipo;
- il formato di codifica accettato è UTF-8.

Una volta importati i dati l'utente può selezionare tra le colonne numeriche quelle da considerare per la visualizzazione senza manipolazioni, per la riduzione dimensionale_G o per il calcolo della matrice delle distanze_G.

In seguito è possibile selezionare la colonna (numerica o non numerica) per il raggruppamento. È possibile validare ed eventualmente correggere un file `.csvG` all'indirizzo

<http://www.fixcsv.com/>

4.2 Scelta di una riduzione dimensionale_G

Dopo aver scelto una visualizzazione Scatter plot Matrix o Proiezione Lineare Multi Asse, è possibile selezionare **una sola o nessuna** delle seguenti riduzioni dimensionali_G e le relative opzioni:

- FASTMAP, con opzioni:
 - numero di dimensioni in output;
 - funzione di calcolo della distanza_G associata;
- ISOMAP, con opzioni:
 - numero di dimensioni in output;
 - funzione di calcolo della distanza_G associata;
 - parametro *neighbors*;
- t-SNE, con opzioni:
 - numero di dimensioni in output;
 - funzione di calcolo della distanza_G associata;
 - parametro *neighbors*;
 - parametro *perplexity*;
 - parametro *epsilon*;
- LLE, con opzioni:
 - numero di dimensioni in output;
 - funzione di calcolo della distanza_G associata;
 - parametro *neighbors*;
- UMAP, con opzioni:
 - numero di dimensioni in output;
 - parametro *neighbors*.

Le opzioni possibili sono riassunte nella tabella in seguito:

Tabella 4: Opzioni disponibili per ogni riduzione dimensionale

	FASTMAP	ISOMAP	t-SNE	LLE	UMAP
Numero di dimensioni in output	Sì	Sì	Sì	Sì	Sì
Funzione di calcolo della distanza	Sì	Sì	Sì	Sì	No
Parametro <i>neighbors</i>	No	Sì	Sì	Sì	Sì
Parametro <i>perplexity</i>	No	No	Sì	No	No
Parametro <i>epsilon</i>	No	No	Sì	No	No

Le varie opzioni appaiono e scompaiono a video dinamicamente in seguito alla selezione della riduzione dimensionale_G. Tutte le opzioni sono selezionabili tramite slider_G numerico, ad eccezione della scelta della funzione di calcolo della distanza_G associata, selezionabile tramite lo stesso widget utilizzato per la scelta di una funzione per il calcolo delle distanze_G utilizzato per il calcolo della matrice delle distanze_G.

È consentito selezionare un numero di dimensioni in output minore, uguale o maggiore del numero di dimensioni in input.

Per una trattazione approfondita dei vari algoritmi di riduzione dimensionale_G si rimanda a §1.4 - Riferimenti.

4.3 Scelta di una funzione per il calcolo delle distanze

Dopo aver selezionato una visualizzazione Force Field o Heat Map è necessario scegliere **una e una sola** tra le funzioni disponibili per il calcolo delle distanze_G.

Le opzioni possibili sono le seguenti:

- distanza_G Euclidea;
- distanza_G di Manhattan;
- distanza_G Cosine;
- distanza_G di Chebyshev;
- distanza_G Euclidea quadrata;
- distanza_G di Canberra;

Sono selezionabili tramite un elenco di checkbox_G.

Per una trattazione approfondita delle varie funzioni di calcolo della distanza_G si rimanda a §1.4 - Riferimenti

4.4 Funzionalità dei grafici

4.4.1 Force Field

Il grafico Force Field permette di visualizzare un integratore numerico di velocità Verlet_G per la simulazione di forze fisiche sulle particelle, utile per lo studio di reti.

4.4.1.1 Modalità di interazione

È possibile *trascinare e fissare* il nodo in un punto dello schermo, **mantenendo premuto** il cursore sopra un nodo e **spostando** il nodo nel punto desiderato. Al termine dell'operazione il bordo del nodo assume il colore nero.

È possibile *rilasciare* un nodo precedentemente fissato, cliccando sul nodo stesso. Al termine dell'operazione il bordo del nodo assume il colore bianco.

È possibile *visualizzare il nome* associato a un nodo, **posizionando il cursore per qualche secondo** sopra il nodo stesso.

4.4.1.2 Parametri della visualizzazione

Attraverso il pannello di controllo della visualizzazione si possono modificare i seguenti parametri:

- **Intensità della forza_G**: l'utente può impostare l'attrazione reciproca esercitata dai nodi. Se impostata ad un valore positivo i nodi si attraggono tra loro, se impostata ad un valore negativo i nodi si respingono. Il suo valore di default è -30 .
- **Distanza_G minima**: distanza_G minima tra i nodi affinché la forza_G venga applicata alla coppia di nodi. Il suo valore di default è impostato a 1.
- **Distanza_G massima**: distanza_G massima tra i nodi affinché la forza_G venga applicata alla coppia di nodi. Il suo valore di default è impostato a 2000.
- **Soglia minima**: intensità minima dei collegamenti da visualizzare.

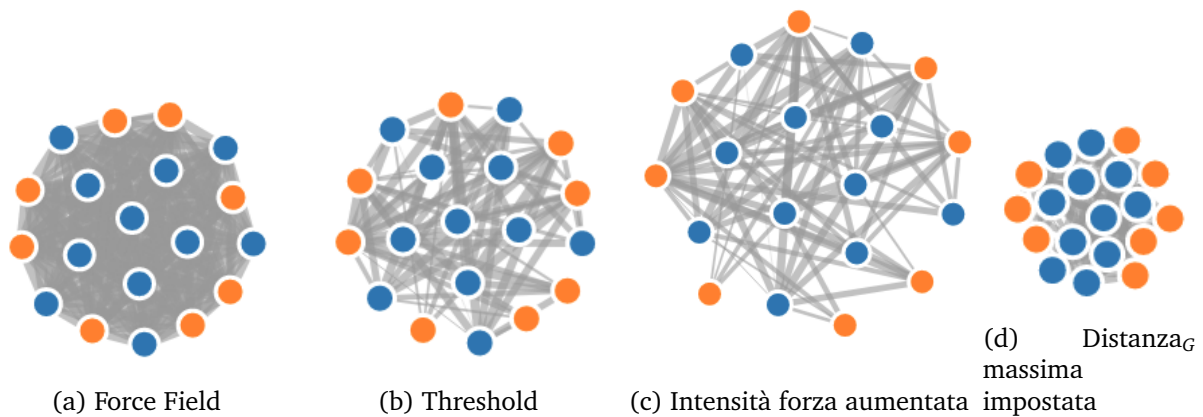


Figura 7: Modifica parametri Force Field

4.4.2 Heat Map

Il grafico Heat Map permette di visualizzare una matrice bidimensionale delle distanze, contenente rettangoli colorati.

Ogni rettangolino posizionato *al di fuori* della diagonale_G rappresenta un arco di un grafo_G. La riga e la colonna di appartenenza indicano ciascuna un nodo che è estremo dell'arco, il cui nome è riportato sopra la riga o la colonna. Ogni rettangolino ha un colore tendente al viola, di opacità proporzionale all'intensità del collegamento tra i nodi rappresentanti riga e colonna.

Ogni rettangolino posizionato *nella* diagonale_G non rappresenta un arco di un grafo_G, bensì il nodo rappresentante la riga e la colonna (che hanno lo stesso indice, trattandosi di elementi sulla diagonale). È riempito con un colore opaco diverso dal viola, che rappresenta il gruppo di appartenenza del nodo.

Ciascun collegamento tra nodi è presente due volte nel grafico Heat Map: una volta a destra della diagonale_G, un'altra volta a sinistra della diagonale_G.

4.4.2.1 Modalità di interazione

È possibile *ordinare* gli elementi della matrice, selezionando la modalità di ordinamento dall'apposito menu a tendina.

Le opzioni selezionabili sono:

- **ordinamento per cluster_G**;
- **ordinamento per nome**;
- **ordinamento originale**.

4.4.2.2 Parametri della visualizzazione

Attraverso il pannello di controllo della visualizzazione si possono modificare i seguenti parametri:

- **Soglia minima**: intensità minima dei collegamenti da visualizzare.

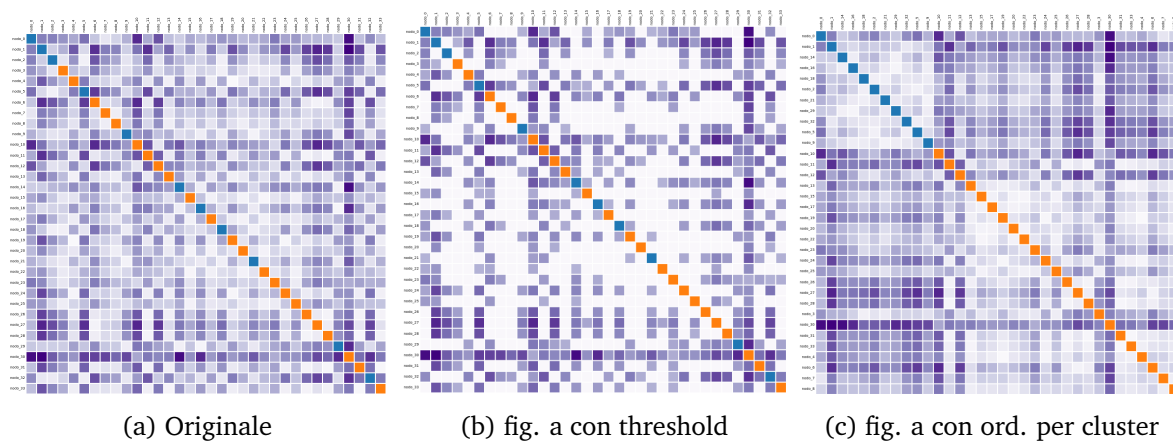


Figura 8: Modifica parametri Heat Map

4.4.3 Scatter plot Matrix

Il grafico Scatter plot Matrix permette di visualizzare una griglia di Scatter plot_G, ciascuno avente come assi una coppia di dimensioni del dataset. Ciò significa che, preso un dataset con k dimensioni, lo Scatter plot Matrix avrà k righe e k colonne, e il grafico alla i -esima riga e alla j -esima colonna è uno Scatter plot_G con la i -esima dimensione nell'asse delle ascisse e la j -esima dimensione nell'asse delle ordinate.

N.B. Ogni coppia di dimensioni distinte appare in due diversi Scatter plot_G della matrice, in posizioni speculari rispetto alla diagonale, con assi delle ascisse e delle ordinate invertiti.

4.4.3.1 Modalità di interazione

L'operazione di *brushing* permette di selezionare un insieme di punti in uno Scatter plot_G e di vedere evidenziati quei punti in tutti i restanti Scatter plot_G della matrice.

È possibile *effettuare il brush_G* dei nodi **selezionando un area rettangolare** in uno Scatter plot_G della griglia. In seguito a questa operazione in ogni Scatter plot_G sono distinguibili:

- i punti selezionati, nelle dimensioni e nei colori originali;
- i punti non selezionati, in dimensioni nettamente ridotte e in colore nero.

È possibile *spostare* il rettangolo di selezione precedentemente disegnato, **trascinandolo** con il mouse nella posizione desiderata. L'evidenziamento dei dati nei restanti Scatter plot viene aggiornato in tempo reale.

È possibile *annullare il brush_G* della selezione precedentemente creata, **clickando in un punto** del grafico **al di fuori** del rettangolo di selezione. In seguito a questa operazione tutti i punti vengono visualizzati nella loro forma originaria e non sono più visibili aree di selezione nel grafico.

È possibile *aggiungere o rimuovere* una dimensione selezionandola dal pannello in parte al grafico. Ad ogni aggiunta o rimozione di una dimensione l'eventuale brush_G applicato viene rimosso.

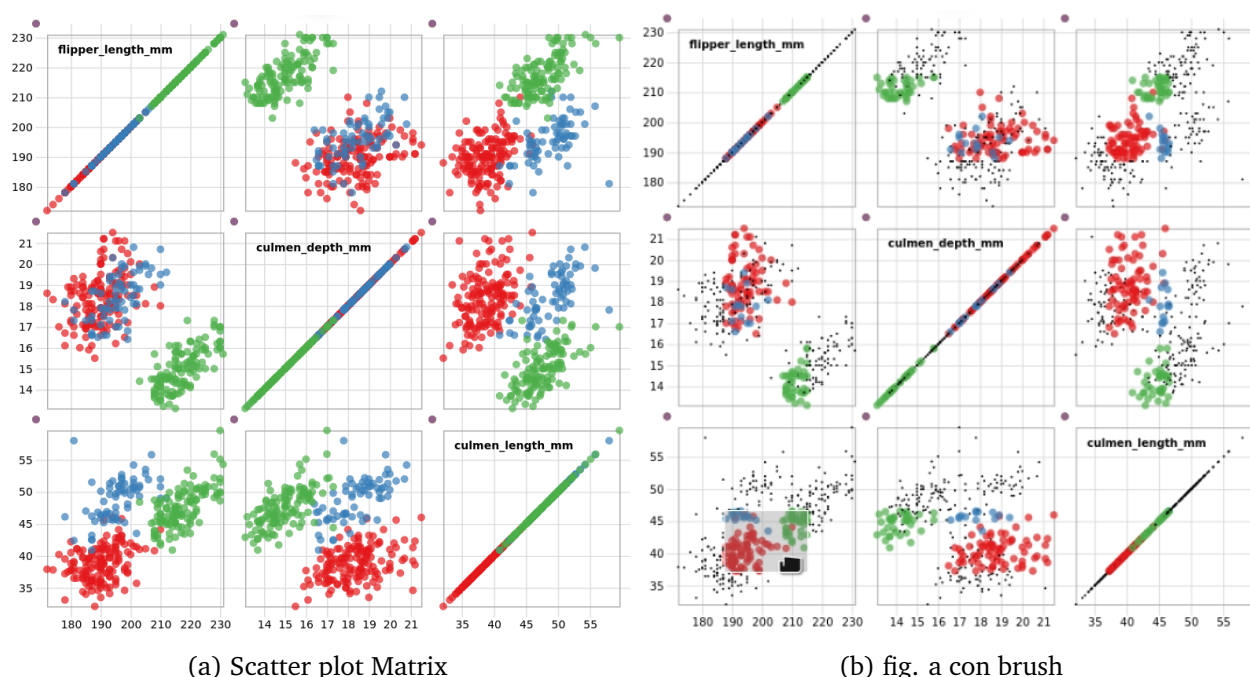


Figura 9: Interazione con Scatter plot Matrix

4.4.4 Proiezione Lineare Multi Asse

La Proiezione Lineare Multi Asse è una variante dello Scatter plot_G tradizionale, nella quale gli assi sono visibili e trascinabili.

Il trascinamento di un asse mostra a schermo una differente proiezione_G bi-dimensionale del dataset multi-dimensionale_G originario.

L'asse delle ascisse e delle ordinate rappresentano le componenti principali 1 e 2. Una griglia posizionata in secondo piano rispetto al grafico aiuta a dedurre le coordinate dei punti rispetto alle due componenti principali. L'origine_G è rappresentata dall'intersezione degli assi.

4.4.4.1 Modalità di interazione

È possibile *visualizzare una diversa proiezione_G* del dataset originario **trascinando** l'estremità di un asse in un altro punto del grafico avente distanza_G uguale (o simile) dall'origine_G.

È possibile *comprimere o allungare* un asse **trascinando** l'estremità dell'asse in un altro punto del grafico più vicino o lontano dall'origine_G, senza effettuare rotazioni.

Un'operazione di trascinamento può produrre uno solo o entrambi gli effetti descritti sopra, poiché l'utente non è limitato a muovere un'estremità di un asse lungo una stessa orbita circolare o lungo la stessa direzione più o meno lontana dall'origine_G.

È possibile *aggiungere o rimuovere* un asse selezionandolo dal pannello in parte al grafico. Ad ogni aggiunta o rimozione di un asse la proiezione_G viene ricalcolata, dunque gli assi già presenti o rimasti potrebbero assumere una posizione differente.

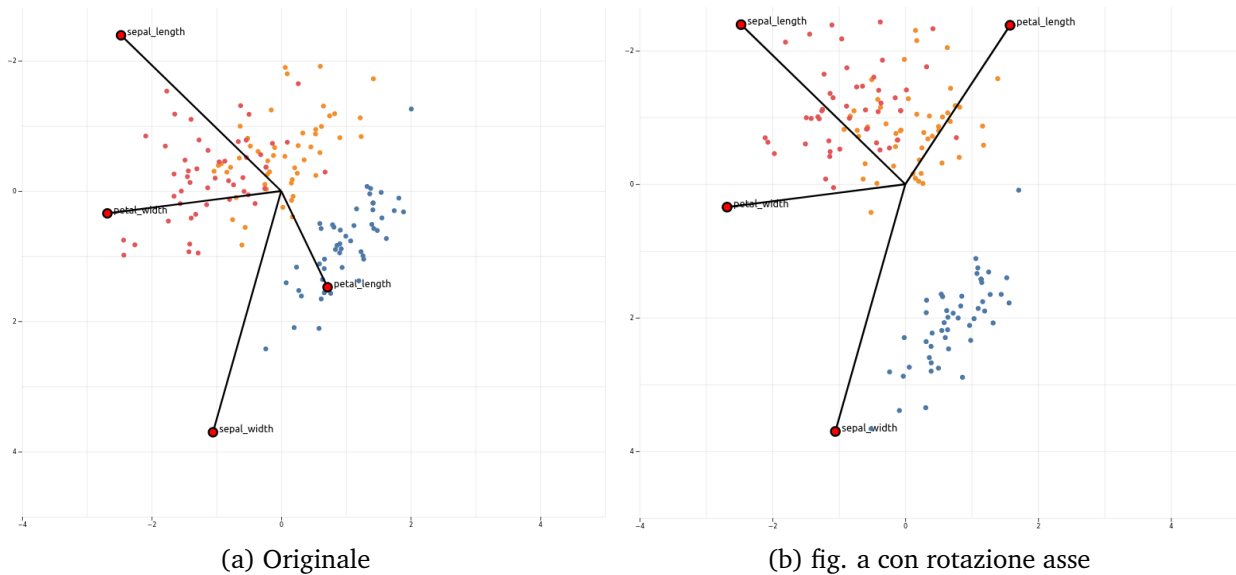


Figura 10: Interazione con Proiezione Lineare Multi Asse

A Glossario

Di seguito non sono riportate le definizioni di funzioni matematiche e algoritmi per i quali si è ritenuto poco utile fornire una definizione concisa e povera di formalismi matematici che si adattasse a un glossario.

Si rimanda dunque a §1.4 - Riferimenti per la definizione dei seguenti termini:

- ALGORITMI DI RIDUZIONE DIMENSIONALE_G: FASTMAP, ISOMAP, LLE, t-SNE, UMAP;
- FUNZIONI DI CALCOLO DELLE DISTANZE_G: distanza di Canberra, distanza di Chebyshev, distanza Cosine, distanza Euclidea, distanza Euclidea Quadrata, distanza di Manhattan.

B

Browser

Applicazione per la fruizione di contenuti web.

Brush

In uno Scatter plot Matrix, l'operazione di Brush permette di selezionare un insieme di punti in uno Scatter plot e di vedere evidenziati quei punti in tutti i restanti Scatter plot della matrice.

Bug

Guasto che porta al malfunzionamento del software, tipicamente dovuto ad un errore nella scrittura del codice sorgente di un programma.

Build

Trasformazione del codice sorgente in codice eseguibile.

C

Checkbox

Controllo grafico con cui l'utente può effettuare selezioni multiple.

Cluster

Gruppo di oggetti di un insieme. Può essere dato in partenza o può essere ricavato da particolari algoritmi di apprendimento non supervisionato, detti di *clustering*.

CSV

Acronimo di *Comma-Separated Values*, è un formato di file utilizzato per la rappresentazione di una tabella di dati, che utilizza il carattere delimitatore virgola ',' o punto e virgola ';' per separare le celle.

D

Database

Insieme organizzato di dati, gestito da un DBMS (DataBase Management System).

Dati multi-dimensionali

Dataset in cui il numero di features per ogni record del dataset è $\gg 3$.

Diagonale

In una matrice A , la diagonale è l'insieme di elementi $A_{i,j}$ tali che $i = j$.

Distanza

Misura della "lontananza" tra due punti di un insieme al quale si possa attribuire qualche carattere spaziale.

DSV

Acronimo di *Delimiter-Separated Values*, è una famiglia di formati di file utilizzati per la rappresentazione di una tabella di dati, che utilizzano un carattere delimitatore per separare le celle. CSV_G e TSV_G sono formati DSV.

E**EDA**

Acronimo di *Exploratory Data Analysis*, parte dell'attività di Data Analysis, nella quale l'analista ricerca associazioni, sequenze ripetute nascoste o pattern nei dati.

F**Forza**

Grandezza fisica vettoriale che si manifesta nell'interazione reciproca di due o più corpi sia a livello macroscopico, sia a livello delle particelle elementari.

G**Grafo**

In Matematica Discreta, un grafo è una tripla (V, E, f) , dove V è detto insieme dei nodi, E è detto insieme degli archi e f è una funzione che associa ad ogni arco e in E due vertici u, v in V .

Git

Sistema di versionamento_G distribuito multi-piattaforma. Permette di versionare i sorgenti software_G, documenti di testo e di collaborare alla loro realizzazione.

GitHub

Servizio di hosting per progetti software_G, che implementa lo strumento di controllo versione distribuito Git_G.

I

Integratore di Verlet

Metodo numerico per il calcolo delle Equazioni del moto di Newton.

M

Matrice delle distanze

Matrice quadrata contenente le distanze tra ogni coppia di un insieme.

N

Node.js

Framework open-source multi-piattaforma orientato agli eventi per l'esecuzione di codice JavaScript.

npm

Acronimo di *Node.js Package Manager*, gestore di pacchetti predefinito per Node.js.

O

Origine

Nel sistema di riferimento cartesiano, il punto nel quale si intersecano tutte le rette orientate dette *assi*.

P

PostgreSQL

Database ad oggetti e relazionale open-source_G, focalizzato sulla scalabilità e sulla massima aderenza allo standard SQL.

Proiezione

Trasformazione lineare definita da uno spazio vettoriale in sé stesso.

R

Radio button

Controllo grafico che consente all'utente di effettuare una scelta singola esclusiva nell'ambito di un insieme predefinito di opzioni o possibili scelte.

Repository

Archivio o sito web nel quale sono raccolti e conservati dati ed informazioni in formato digitale.

Riduzione dimensionale

Algoritmo di trasformazione di dati da uno spazio multi-dimensionale_G ad uno spazio con meno dimensioni. Un algoritmo di riduzione dimensionale cerca di mantenere le caratteristiche dello spazio multi-dimensionale_G anche nello spazio con meno dimensioni.

S

Scatter plot

Tipo di grafico in cui due variabili di un set di dati sono riportate su uno spazio cartesiano. Viene chiamato anche grafico di dispersione.

Scatter plot Matrix e Proiezione Lineare multi asse sono varianti di Scatter plot.

Server

Componente o sottosistema informatico di elaborazione e gestione del traffico di informazioni che fornisce, a livello logico e fisico, un qualunque tipo di servizio ad altre componenti che ne fanno richiesta attraverso una rete di computer, all'interno di un sistema informatico o anche direttamente in locale su un computer.

Slider

Componente grafico con il quale un utente può impostare un valore muovendo un indicatore, solitamente con uno spostamento orizzontale.

SVG

Acronimo di *Scalable Vector Graphics*, indica un particolare formato che è in grado di visualizzare oggetti di grafica vettoriale e quindi di salvare immagini in modo che siano ridimensionate a piacere senza perdere in risoluzione grafica.

T

TSV

Acronimo di *Tab-Separated Values*, è un formato di file utilizzato per la rappresentazione di una tabella di dati, che utilizza il carattere delimitatore tabulazione per separare le celle.

V

Validatore

Un validatore è uno strumento in grado di controllare la conformità di un file rispetto a uno standard.