Indice

1 Introduzione **5**
1.1 Contesto e motivazioni 5
1.2 Obiettivi e contributi della tesi 6
1.3 Struttura della tesi 7

2 Stato dell'arte **9**
2.1 La comunitá scientifica 9
2.2 Scientometria 11
2.2.1 Bibliometria 12
2.2.2 H-Index 13
2.3 Correlazione e relativi indici di misurazione 14

3 Modalitá di raccolta dei dati **17**
3.1 Descrizione generale 17
3.2 Il questionario 17
3.2.1 Com'é stato costruito 17
3.2.2 A chi é stato somministrato 17
3.2.3 Modalitá di somministrazione 17
3.3 Considerazioni 17

4 Differences between Community Opinions **19**
4.1 Descrizione del problema 19
4.2 Criteri di analisi e ipotesi sui risultati 20
4.3 Algoritmo "Diffcomm" 22

4.3.1 Input 22
4.3.2 Output 23
4.3.3 Algoritmo 24
4.3.4 Un esempio esplicativo 24
4.4 Risultati e analisi dei dati 28

5 Metriche scientifiche e valutazioni reali 31
5.1 Descrizione del problema 31
5.2 Ipotesi sui risultati e criteri di analisi 32
5.3 Risultati e analisi dei dati 33

6 Conclusioni e considerazioni finali 37

A Dati raccolti e analisi svolte 39
A.1 AAI 39
A.1.1 Dati raccolti 39
A.1.2 Analisi della correlazione con h-index 39
A.1.3 Analisi dei voti della comunitá 40
A.2 BPM.ICSOC.ICWS 40
A.2.1 Dati raccolti 40
A.2.2 Analisi della correlazione con h-index 40
A.2.3 Analisi dei voti della comunitá 41
A.3 ER 42
A.3.1 Dati raccolti 42
A.3.2 Analisi della correlazione con h-index 42
A.4 EUROCRYPT 42
A.4.1 Dati raccolti 42
A.4.2 Analisi della correlazione con h-index 43
A.5 ICSE 43
A.5.1 Dati raccolti 43
A.5.2 Analisi della correlazione con h-index 43
A.6 ICWE.WISE 43
A.6.1 Dati raccolti 43
A.6.2 Analisi della correlazione con h-index 43
A.6.3 Analisi dei voti della comunitá 44

A.7 VLDB 45
A.7.1 Dati raccolti 45
A.7.2 Analisi della correlazione con h-index 45
A.7.3 Analisi dei voti della comunitá 45
A.8 GENERAL 45
A.8.1 Dati raccolti 45
A.8.2 Analisi della correlazione con h-index 45

B Algoritmo Diffcomm **55**
B.1 Le classi 55
B.2 I file di input 56

Bibliografia **59**

Capitolo 1

Introduzione

1.1 Contesto e motivazioni

In questi ultimi anni abbiamo potuto osservare come, con l'avvento di Internet e delle nuove tecnologie, sia cambiato non solo il modo di comunicare ma anche il modo in cui le persone si relazionano tra di loro.

Anche le comunitá scientifiche si sono adattate a questo cambiamento diventando col tempo un soggetto complesso e in continua evoluzione. Oggigiorno, infatti, non esistono piú comunitá isolate, ma grandi insiemi di comunitá scientifiche strettamente collegati e sovrapposti tra di loro. Una qualsiasi entitá scientifica, infatti, puó non solo essere di interesse comune ma anche appartenere o collaborare con divese comunitá. Da qui nasce la necessità di conoscere meglio le comunitá scientifiche: come nascono, si evolvono nel tempo ma soprattutto come si relazionano tra di loro e quanto influenzino i loro membri.

Ma non solo, come suggerito in [13] una maggiore conoscenza delle comunitá scientifiche ci puó aiutare a migliorare due importanti attivitá nell'ambito della ricerca scientifica: la *ricerca* di contributi scientifici e la *valutazione* dei ricercatori. Grazie a tool come il Community Engine Tool (CET) [13] nato e sviluppato all'interno del progetto Liquidpub [11], oggi siamo in grado di avere un'idea chiara e precisa delle diverse comunitá scientifiche esistenti e

dei loro membri, ma non esiste ancora un indice di valutazione che tega conto del contesto, ossia la comunitá, in cui una determinata ricerca si inserisce.

Le attuali metriche di valutazione, come h-index [21] e g-index [21], sono basate sulle citazioni: il numero di citazioni che una determinata pubblicazione scientifica ha ricevuto determina il contributo che questa porta alla comunitá e, maggiore è il contributo, maggiore sará anche la valutazione degli autori. Come detto precedentemente, il contesto non viene preso in considerazione. In molti altri casi, inoltre, come mostrato in [12] il risultato di una valutazione ottenuta con le attuali metriche puó non riflettere correttamente la reputazione che un gruppo di ricercatori e/o un determinato contributo scientifico ha all'interno di una specifica comunitá.

É a partire da questo scenario che nasce l'idea che sta alla bese di questa tesi. Attraverso un sondaggio, con cui abbiamo invitato alcune persone ad esprimere la loro opinione su un gruppo di candidati, abbiamo calcolato la correlazione esistente fra le opinioni delle diverse comunitá scientifiche. Questo ci ha permesso, non solo di avere un dato circa la *reale* reputazione di un determinato individuo in modo da poter poi verificare l'attendibilitá degli attuali indici di valutazione dei ricercatori, ma anche di capire la distanza e le relazioni che intercorrono fra le diverse comunitá. Cosa che puó essere molto utile se si vuole, per esempio, capire i meccanismi che sottendono l'elezione di un preside di una particolare facoltá o il rettore di un'universitá.

1.2 Obiettivi e contributi della tesi

Gli obiettivi pricipali di questa tesi sono essenzialmente due:

- **Verificare l'attendibilitá delle attuali metriche di valutazione** della reputazione scientifica calcolando e analizzando la correlazione esistente fra questi indici e la reale opionione degli altri scienziati;

- **Studiare** la natura delle **diverse opinioni fra le varie comunitá scientifiche** calcolando ed analizzando la correlazione fra le opinioni.

Entrambi gli obiettivi hanno come scopo quello di *incrementare la conoscenza*:

- sulle **comunitá scientifiche**: proponendo un algoritmo che sia in grado di rintracciare le differenze di opinioni e le relazioni che intercorrono fra le varie comunitá;

- sugli attuali **indici di valutazione** per poter proporre, in futuro, un nuovo indice che tenga conto del contesto in cui un'entitá scientifica opera.

1.3 Struttura della tesi

In questa tesi viene presentata la ricerca e l'analisi dei dati che abbiamo effettuato per raggiungere i nostri obiettivi con particolare riferimento allo sviluppo dell'algoritmo "Diffcomm" in grado di analizzare le differenze di opinioni fra le varie comunitá scientifiche.
Di seguito una breve panoramica dei vari capitoli.

Capitolo 2

Parleremo di comunitá scientifica, scientometria e correlazione dandone le definizioni e mostrando lo stato dell'arte relativo a questi argomenti.

Capitolo 3

Parleremo del questionario. Prima in generale e poi soffermandoci su come abbiamo costruito i questionari che abbiamo utilizzato nel nostro lavoro.

Capitolo 4

Parleremo della differenza di opinioni e del legame esistente fra le comunitá scientifiche presentando un nuovo algoritmo ("Diffcomm") in grado di quantificare tutto ció.

Capitolo 5

Parleremo degli indici di valutazione del lavoro scientifico e della loro attendibilitá.

Capitolo 6

Vengono presentati brevemente i risultati ottenuti con questo lavoro.

Capitolo 2

Stato dell'arte

In questo secondo capitolo si passa in rassegna lo stato dell'arte sulle comunitá scientifiche e le attuali metriche di valutazione utilizzate nel mondo scientifico. Nella sezione 2.1 viene data una definizione di comunitá scientifica e presentato il lavoro di ricerca svolto fino ad ora all'interno del progetto Liquidpub [11]. Nella sezione 2.2 vengono introdotte le metriche di valutazione piú famose ed utilizzate andando a sottolineare quelli che sono i loro punti deboli. Infine, nella sezione 2.3, viene presentata la correlazione e fatta una panoramica sui relativi indici di misurazione.

2.1 La comunitá scientifica

Il concetto di comunitá é uno di quei concetti fondamentali che appartengono a piú discipline contemporaneamente. Pur essendo usato principalmente in sociologia ed in antropologia, questo é usato piú volte anche in filosofia e in diritto. Nell'ambito delle scienze scociali il termine comunitá, introdotto per la prima volta da Ferdinand Tönnies[1] in [7], grazie anche agli studi della sociologia moderna, puó essere attualmente inteso in due modi differenti [20]:

- **Comunitá geografica:** la comunitá geografica é una comunitá dove le persone hanno in comune il luogo, inteso come territorio, in cui vivono;

[1] 26 July 1855, near Oldenswort (Eiderstedt, North Frisia) - 9 April 1936, Kiel, Germany

- **Comunitá d'interessi:** in questo tipo di comunitá i diversi individui sono accomunati dai loro interessi, siano questi di tipo religioso, lavorativo, etc.;

Prendendo spunto da queste idee di comunitá, possiamo definire la **comunitá scientifica** come una *comunitá d'interessi*, dove l'interesse comune dei diversi soggetti é rappresentato dal loro lavoro di ricerca scientifica, e si riflette sulle molteplici interazioni fra gli individui che vi appartengono.
Una definizione di comunitá scientifica puó essere ritrovata in [15], dove viene considerata **comunitá scientifica** un insieme caratterizzato da un nome di entitá scientifiche come scienziati, eventi scientifici, contributi scientifici, etc. che sono strettamente legati fra di loro. Piú formalmente:

$$C_i = (nome, \{e^{[w]}, \{w_t\}\}, \{t\})$$

dove:

- C_i identifica la comunitá i-esima in un insieme di comunitá;
- *nome* é il nome della comunitá che corrisponde ad una determinata area di studio (e.g. Intelligenza Artificiale,...);
- e é un'entitá scientifica qualsiasi fra le seguenti: contributo scientifico, persona, evento o collezione;
- w é un coefficiente globale di partecipazione che indica il grado di partecipazione dell'entitá all'interno della comunitá scientifica;
- $\{w_t\}$ é un insieme di regole o vincoli che definiscono il tipo di relazione che deve essere presa in considerazione per definire il rapporto di appartenenza di un entitá ad una comunitá;
- $\{t\}$ é un insieme di *nano-topics* che espandono la definizione del *nome della comunitá*.

Questa definizione formale di comunitá scientifica nasce all'interno dell'area di ricerca sulle comunitá scientifiche del progetto Liquidpub [11], ideato per permettere la divulgazione della conoscenza scientifica, in maniera facile

ed efficiente, attraverso il Web. Come spiegato in [16], quello che si sta cercando di fare é creare un modello per le comunitá scientifiche e sviluppare un tool in grado di rintracciare tutte le comunitá attualmente esistenti.

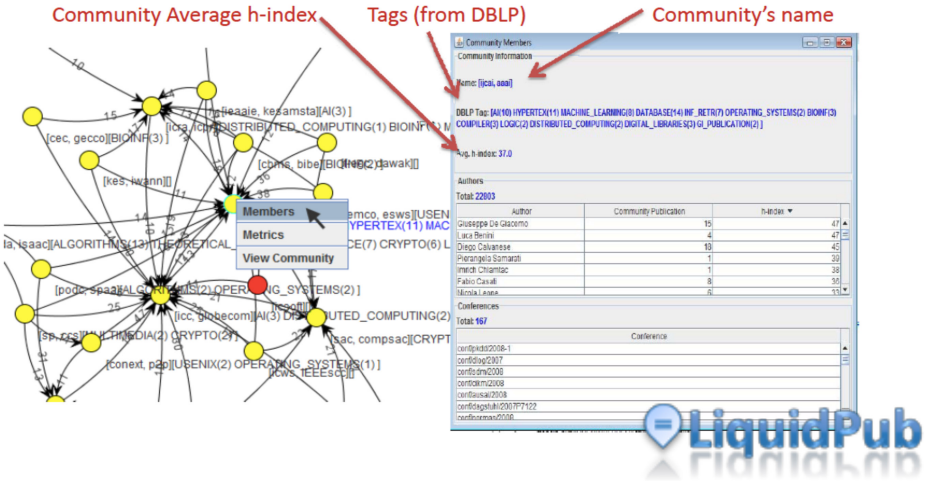


Figura 2.1: Community Engine Tool (CET)

Attualmente grazie al Community Engine Tool (CET) [13], un'applicazione desktop, siamo in grado di rintracciare le diverse comunitá scientifiche e di creare un "Community Network", attravreso il quale é possibile effettuare un'analisi delle diverse comunitá, come il loro impatto e la loro produttivitá. É, infine, possibile avere un grafo delle diverse comunitá scientifiche in grado di mostrarci le diverse relazioni che intercorrono tra di loro.

2.2 Scientometria

A partire dalla fine degli anni Sessanta, inizi anni Settanta, del XX Secolo inizia a svilupparsi e prendere piede una nuova scienza chiamata **Scientometria**. La scientometria é definita come la scienza che si occupa della misurazione e dell'analisi della scienza e delle produzioni scientifiche [21], ossia, come spiegato in [4], *la scienza per la misura e l'analisi della scienza.* Nata dagli studi di quello che viene cosiderato oggi il suo padre fondatore,

Derek John De Solla Price[2] [5] [6] e di Eugene Garfield[3], fondatore delle teorie dell'Institute for Scientific Information ISI e ideatore dell'Impact Factor, la scientometria moderna è divenuta un campo di studio molto complesso e specialistico che va oltre il semplice studio e l'analisi delle citazioni per arrivare a coinvolgere numerosi ricercatori di diverse discipline di tutto il mondo. Nel 1993 viene fondata l'*International Society for Scientometrics and Infometrics* (ISSI) la quale ha come suoi membri numerosi studiosi pionieri della scientomentria come lo stesso Garfield e molti altri.

Aspetti fondamentali della scientometria sono la valutazione e la misurazione delle pubblicazioni scientifiche. Questi possono essere considerati due elementi principali nel processo di crescita e di diffusione della conoscenza scientifica. Valutare la ricerca significa stabilire norme e criteri per misurare la quantità ed esprimere giudizi sulla qualità di una produzione scientifica. Tale valutazione si basa su diversi metodi fra cui:

- l'*analisi bibliomentrica*, aspetto fondamentale della bibliometria che effettua un'analisi di tipo quantitativo dei dati;
- la *peer-review*, ossia la revisione dei pari, che effettua un'analisi di tipo qualitativo di una determinata pubblicazione scientifica attraverso una valutazione fatta da specialisti del settore.

2.2.1 Bibliometria

La bibliometria è una disciplina che applica metodi matematici e statistici allo studio della comunicazione scientifica, della sua struttura e delle sue dinamiche. L'analisi bibliometrica si concentra su articoli di riviste scientifiche, libri, bibliografie, con l'obiettivo sia di scoprire le relazioni quantitative tra pubblicazioni, autori ed aree di ricerca, sia di misurare la qualità (o performance) dei ricercatori [19].
Nata negli ultimi vent'anni per far fronte alla enorme quantità di dati gestiti da banche di dati online, la bibliometria deve il suo successo alla ricerca, da

[2] Inghilterra, 22 Gennaio 1922 –Stati Uniti d'America, 3 Settembre 1983
[3] New York City, 16 Settembre 1925

parte delle comunitá scientifiche, di strumenti in grado di misurare oggettivamente il lavoro degli scienziati: gli indicatori bibliometrici.
Come spiegato in [19], gli indicatori bibliometrici piú noti ricavabili dagli attuali database citazionali sono:

- **Numero di citazioni** é il numero di volte che un autore o un documento é stato citato da un gruppo selezionato di fonti. Nonostate sia il mattone fondamentale per la costruzione degli altri indici di citazione, quest'indicatore dipende da fattori che non c'entrano niente con la qualitá dell'autore o del documento;

- **Impact Factor (IF)** é il rapporto tra il numero complessivo di citazioni ricevute in un dato anno dagli articoli pubblicati da una certa rivista nei due anni precedenti e il numero di questi ultimi: in pratica il numero medio di citazioni ricevute da un singolo articolo pubblicato da una determinata rivista in un certo periodo di tempo [17].
  Inventato da Eugene Garfield negli anni '50, l'impact factor nasce con lo scopo di valutare il peso di una rivista all'interno del suo settore specifico. L'IF é pubblicato a cadenza annuale nel Journal Citation Reports (JCR) e viene calcolato per migliaia di riviste scientifiche inserite nelle banche dati citazionali di Thomson Reuters [21];

- **H-index** é un indicatore di performance individuale introdotto dal fisico Jorge E. Hirsch in un articolo del 2005. Vedi la Sezione successiva per maggiori chiarimenti.

2.2.2 H-Index

Per **h-index** intendiamo un indice introdotto nel 2005 dal fisico Jorge E. Hirsh in [9], utilizzato per quantificare la prolificitá e l'impatto del lavoro degli scienziati, basandosi sia sul numero delle loro pubblicazioni che sul numero di citazioni ricevute. La definizione di Hirsch é la seguente:
Uno scienziato possiede un indice h se h dei suoi N_p lavori hanno almeno h citazioni ciascuno e i rimanenti $(N_p - h)$ lavori hanno ognuno al più h citazioni.

Oltre ad essere estremamente facile da calcolare, h-index é strutturato per verificare con un processo relativamente semplice non solo la produzione, ma anche l'influenza di uno scienziato, distinguendolo da chi avesse pubblicato molti articoli ma di scarso interesse. Inoltre l'indice non é troppo influenzato da singoli articoli di grande successo.

Nonostante tutto ció e la sua popolaritá l'h-index é un indicatore molto discusso in quanto, non solo penalizza i ricercatori piú giovani, ma non tiene conto delle autocitazioni e del multi-autorship. Una serie di varianti dell'h-index, quali g-index, hg-index, A-index, m-index, sono state proposte allo scopo di correggere alcuni evidenti difetti dell'indicatore di Hirsch, quali la scarsa sensibilità al numero di citazioni, alla dimensione temporale ed alle differenze disciplinari.

2.3 Correlazione e relativi indici di misurazione

In statistica, per **correlazione** si intende una relazione tra due variabili casuali tale che a ciascun valore della prima variabile corrisponda con una certa regolarità un valore della seconda [21].

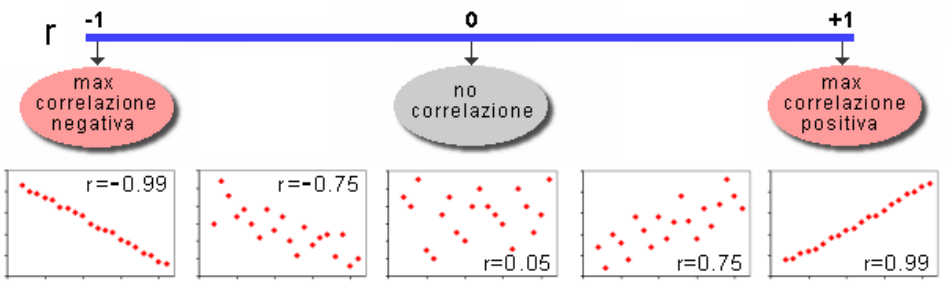


Figura 2.2: Esempi di insiemi di punti (x, y) con relativo coefficiente di correlazione. (http://www.quadernodiepidemiologia.it/epi/cause/coef_cor.htm)

La correlazione puó essere:

- **positiva o diretta** quando al crescere (o al diminuire) di un valore di una variabile, corrisponde un crescita (o una diminuzione) del va-

lore dell'altra variabile. E.g. a una crescita del prezzo del petrolio corrispone una crescita del prezzo della benzina;

- **negativa o indiretta** quando al crescere (o al diminuire) di un valore di una variabile, corrisponde una diminuzione (o una crescita) del valore dell'altra variabile. E.g. alla diminuzione di produzione di grano corrisponde un aumento del prezzo della pasta;

- **nulla** quando non vi é correlazione tra le variabili.

Il grado di correlazione tra due variabili viene espresso tramite degli indici o coefficienti di correlazione che assumono valori compresi tra **-1** e **1**, dove **0** é l'indice di correlazione nulla. Vi sono numerosi coefficienti di correlazione. I piú usati e famosi sono i seguenti:

- **Coefficiente ρ di correlazione per ranghi[4] di Spearman.** É un indice di correlazione usato in statistica non parametrica per calcolare la correlazione fra due diversi ranking o ranghi. Il coefficiente ρ di Spearman é definito nel seguente modo:

  $$\rho = 1 - \frac{6 \sum_i D_i^2}{N\,(N^2 - 1)}$$

  dove:

  - $D_i = r_i - s_i$ é la differenza dei ranghi con r_i e s_i rispettivamente il rango della prima variabile e della seconda variabile circa l'i-esima osservazione;
  - N é il numero complessivo di osservazioni.

- **Coefficiente τ di correlazione per ranghi di Kendall.** É un indice di correlazione usato in statistica non parametrica per calcolare appunto la correlazione fra due diversi ranking o ranghi. Il coefficiente τ di Kendall é definito nel seguente modo:

  $$\tau = \frac{(numero\,di\,coppie\,concordanti)\ -\ (numero\,di\,coppie\,discordanti)}{\frac{1}{2}\,n(n-1)}$$

  dove, preso un insieme di informazioni come $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$, abbiamo che:

[4] Studio della correlazione fra differenti rankings sullo stesso insieme di oggetti [21].

- Una coppia si dice **concordante** se $x_i > x_j$ e $y_i > y_j$, oppure se $x_i < x_j$ e $y_i < y_j$;
- Una coppia si dice **discordante** se $x_i > x_j$ e $y_i < y_j$, oppure se $x_i < x_j$ e $y_i > y_j$;

Capitolo 3

Modalitá di raccolta dei dati

3.1 Descrizione generale

3.2 Il questionario

3.2.1 Com'é stato costruito

3.2.2 A chi é stato somministrato

3.2.3 Modalitá di somministrazione

3.3 Considerazioni

Capitolo 4

Differences between Community Opinions

In questo quarto capitolo analizzeremo il problema delle differenti opinioni fra le comunitá scientifiche. Dopo una descrizione generale (4.1), andremo ad esaminare i criteri di analisi dei dati che abbiamo utilizzato (4.2) soffermandoci, in particolare, sul funzionamento dell'algoritmo "Diffcomm" (4.3). Infine, nella sezione 4.4 presenteremo i risultati del nostro lavoro.

4.1 Descrizione del problema

Il problema dello studio delle differenze di opinioni fra le varie comunitá scientifiche nasce dalla necessitá di conoscere meglio le comunitá e i meccanismi insiti al loro interno. In particolare, ció che ci interessa é capire non solo la *reale* opinione delle diverse comunitá scientifiche circa un gruppo di candidati e i relativi contributi, ma anche e soprattutto le affinitá o le differenze che esistono fra i vari gruppi di ricerca.

Lo scopo del nostro lavoro é stato quello di ricavare un dato chiaro e preciso in grado di quantificare la distanza fra le opinioni di due comunitá scientifiche e proporre un algoritmo in grado di calcolare tale dato.

In ambito accademico le applicazioni di questo algoritmo sono numerose. Dalla previsione e l'analisi dei risultati delle elezioni del preside di una facoltá o del rettore di una universitá attraverso lo studio delle opinioni dei vari gruppi di professori e ricercatori presenti, all'analisi delle collaborazioni fra le diverse comunitá scientifiche. É, infatti, molto piú probabile che vi sia una collaborazione fra due comunitá con opinioni vicine fra di loro, rispetto ad una collaborazione fra due comunitá che non hanno niente in comune.

Questo algoritmo, inoltre, puó risultare utile anche al di fuori dell'ambito accademico. Puó essere usato, per esempio, in politica per capire e avere un'idea delle opinioni che diversi gruppi di cittadini hanno circa un candidato o una scelta effettuata; oppure in ambito aziendale per operare una scelta piú consapevole degli investimenti.

4.2 Criteri di analisi e ipotesi sui risultati

In questa sezione presenteremo alcune ipotesi sui risultati e faremo una breve panoramica su come abbiamo impostato il nostro lavoro ponendo l'attenzione e spiegando il motivo di alcune scelte che abbiamo effettuato.

Abbiamo sottoposto un questionario (vedi Capitolo 3) durante alcune recenti conferenze ad un gruppo di partecipanti per capire l'impatto di alcuni ricercatori nell'ambito della computer science. In ogni questionario veniva chiesto al soggetto di esprimere la propria opinione circa l'importanza e l'influenza del lavoro di alcuni candidati, nonché di fornire alcuni dati relativi la comunitá scientifica di appartenenza.

Essendo lo scopo del nostro lavoro studiare la natura delle diverse opinioni fra le comunitá scientifiche, era necessario avere un'idea del punto di vista di una determinata comunitá. Per fare ció, abbiamo suddiviso le opinioni dei singoli individui per comunitá di appartenenza ed abbiamo effettuato una media ponderata basata sul numero di pubblicazioni. Nel calcolare l'opinione generale di una determinata comunitá, infatti, per ogni ricercatore si é considerato il numero di pubblicazioni effettuate allo scopo di garantire un maggior peso e, di conseguenza, una maggiore importanza al voto di un

ricercatore molto attivo rispetto a quello di uno meno attivo nella community. Nella realtá, peró, per come é stato costruito il nostro questionario, non é stato possibile "pesare" il voto dei vari individui riducendo, cosí, questa operazione ad una semplice media aritmetica.

Grazie a tool statistici come "R" [18] possiamo poi calcolare un indice, come il coefficiente di correlazione Kendall's tau [10], che misura la corrispondenza e la distanza fra diversi ranking, nel nostro caso le diverse opinioni delle varie comunitá scientifiche. In statistica, come mostrato in 2.3, esistono diversi coefficienti di correlazione. La nostra scelta peró é ricaduta sul coefficiente tau di correlazione per ranghi di Kendall. I motivi di questa scelta sono i seguenti:

- Noi lavoriamo con piccoli campioni di dati (< 100) dei quali é estremamente difficile conoscere i parametri della distribuzione statistica di appartenenza. Questo fa si che il nostro studio sia molto vicino alla statistica non parametrica [1] di cui il coefficiente di correlazione Kendall's tau é uno dei principali metodi;

- Ha una interpretazione semplice ed intuitiva in termini della probabilitá di osservazione di coppie di dati concordanti o discordanti [3, 14];

- É un metodo molto rubusto, anche in presenza di notevoli variazioni sui dati, perché basato sui ranghi;

- Inoltre, le sue proprietá statistiche sono migliori rispetto a quelle degli altri coefficienti di correlazione della statistica non-paramentrica [3].

Quindi alla fine possiamo dire che, nonostante i problemi espressi in [2], il coefficiente di correlazione Kendall's tau é un buon metodo per il lavoro che abbiamo svolto.

[1] La statistica non parametrica é quella branca della statistica che si occupa di analizzare dati che non si adattano ad una distribuzione normale e, di conseguenza, non dipendono dai suoi parametri principali: la media e la deviazione standard.

Infine, abbiamo analizzato il coefficiente di correlazione appena ricavato per poter verificare il tipo di correlazione esistente fra le varie parti.

Arrivati a questo punto possiamo provare a formulare delle ipotesi sui risultati del nostro lavoro relativo a questa problematica. Possiamo supporre che, *date due comunitá scientifiche, maggiore sará il legame presente fra queste due comunitá, maggiore sará la correlazione fra le loro opinioni.* Verificheremo poi se tale ipotesi é corretta oppure no.

4.3 Algoritmo "Diffcomm"

L'algoritmo "Diffcomm" é un algoritmo iterativo che calcola la correlazione esistenze fra le opinioni di diverse comunitá scientifiche. I passi principali dell'algoritmo sono i seguenti:

1. Calcolare la media ponderata delle opinioni dei vari ricercatori appartenenti ad una determinata comunitá scientifica, in modo da ottenere l'opinione generale della comunitá;

2. Calcolare il coefficiente di correlazione Kendall's tau per ogni coppia di comunitá.

Andiamo ora ad analizzare nel dettaglio l'input, l'output e il funzionamento dell'algoritmo.

4.3.1 Input

L'algoritmo "Diffcomm" prende in input un gruppo di insiemi e funzioni:

- **Insieme dei Candidati**: $X = \{x_i : i = 1...m\}$
  Questo insieme contiene l'elenco dei candidati selezionati per un determinato sondaggio.

- **Insieme dei Votanti**: $V = \{v_i : i = 1...n\}$
  Questo insieme contiene l'elenco dei votanti selezionati per un determinato sondaggio.

- **Insieme delle Comunitá**: $C = \{c_i : i = 1...k\}$
  Questo insieme contiene l'elenco delle comunitá scientifiche a cui appartengono i votanti selezionati.

- **Funzione Membership**: $M : (V \cup X) \rightarrow C$
  Questa funzione prende in input un candidato $x \in X$ (o un votante $v \in V$) e ritorna la comunitá scientifica a cui appartiene. M(x) sará perció la comunitá scientifica del candidato o del votante $l \in (V \cup X)$.

- **Funzione Opinion**: $S : (V \times X) \rightarrow \Re$
  Questa funzione prende in input un votante $v \in V$ e un candidato $x \in X$ e ritorna il numero reale relativo all'opinione del votante sul candidato. S(v,x) sará perció il numero reale relativo all'opinione del votante $v \in V$ circa il candidato $x \in X$.

Si veda B.2 per avere un'idea dell'implementazione dell'input dell'algoritmo.

4.3.2 Output

L'algoritmo "Diffcomm" ritorna l'**insieme dei risultati R** per cui

$$R = \{diff(c_1, c_2) : c_1, c_2 \in C\}$$

dove *diff* é la **funzione Difference**:

$$diff : (C \times C) \rightarrow \Re$$

Questa funzione prende in input due comunitá scientifiche $c_1, c_2 \in C$ e ritorna il numero reale relativo alla correlazione esistente fra le opinioni delle due comunitá. $diff(c_1, c_2)$ sará il numero reale che rappresenta la differenza sulla media delle opinioni della comunitá c_1 e della comunitá c_2 rispetto a tutti i candidati. R sará dunque un insieme di numeri reali relativi alle differenze di opinioni su tutti i candidati di tutte le coppie possibili fra le diverse comunitá scientifiche.

Un eventuale modifica che puó essere apportata alla funzione *diff* é la seguente:

$$diff_c : (C \times C) \times X \rightarrow \Re$$

Questa funzione, molto simile a *diff*, prende in input due comunitá scientifiche $c_1, c_2 \in C$, un candidato $x \in X$ e ritorna il numero reale relativo alla correlazione esistente fra le opinioni delle due comunitá rispetto al candidato $x \in X$. Sostituendo $diff_c$ a *diff*, R diventerá l'insieme dei numeri reali relativi alle differenze di opinioni sul candidato $x \in X$ di tutte le coppie possibili fra le diverse comunitá scientifiche.

4.3.3 Algoritmo

Si veda l'*algoritmo 1* nella pagina successiva.

4.3.4 Un esempio esplicativo

Proponiamo ora un esempio esplicativo sul funzionamento dell'algoritmo "Diffcomm". I dati che esamineremo sono alcuni dei dati raccolti tramite dei sondaggi che abbiamo effettuato ad alcune conferenze (ICWE [2], ICPS [3] e BPM [4]). Abbiamo chiesto ad un gruppo di votanti di esprimere la loro opinione circa il lavoro di un gruppo di candidati precendentemente selezionati (si veda il Capitolo 3 per maggiori chiarimenti circa il modo in cui sono stati selezionati i candidati).

La lista dei candidati é la seguente:

- Jeffrey D. Ullman (A)
- Michael Stonebraker (B)
- Stefano Ceri (C)
- Giuseppe De Giacomo (D)
- Dieter Fensel (E)
- Frank van Harmelen (F)

[2] ICWE: 10th International Conference on Web Engineering, July 5-9 2010.
[3] ICPS: 7th International Conference on Pervasive Services, July 13-15 2010.
[4] BPM: 8th International Conference on Business Process Management, Sept. 13-16 2010.

Algorithm 1 Diffcomm algorithm

```
function DIFFCOMM(X, V, C, M, O)
    R ← ∅                    ▷ R é l'insieme dei risultati.
    for all c ∈ C do
        N_c ← ∅              ▷ N_c è l'insieme delle opinioni della community c.
    end for
    ▷ Suddividiamo le diverse opinioni per community di appartenenza
        dei diversi voters e ne calcoliamo la media ponderata.
    for all o ∈ O do
        for all c ∈ C do
            ▷ M(o.v) é la community del voter v che ha espresso
                l'opinione o.
            if M(o.v) = c then
                N_c ← N_c ∪ o
            end if
        end for
    end for
    for all c ∈ C do
        N_c ← effettuo la media ponderata di tutte le o ∈ N_c
    end for
    ▷ Calcoliamo le correlazioni fra le opinioni delle diverse comunitá.
    k ← |C|
    for i ← 1 to k do
        for j ← 1 to k do
            if j > i then
                diff(i, j) ← Nan
            else
                diff(i, j) ← coefficiente tau di Kendall (N_i, N_j)
            end if
            R ← R ∪ diff(i, j)
        end for
    end for
    return R
end function
```

- David Maier (G)
- Geoffrey E. Hinton (H)
- Alon Y. Halevy (I)
- David J. DeWitt (J)
- Christian S. Jensen (K)
- Ahmed K. Elmagarmid (L)
- Deborah L. McGuinness (M)
- Hector Garcia-Molina (N)
- Amit P. Sheth (O)
- Sushil Jajodia (P)
- Steffen Staab (Q)

Nella *tabella 4.4* a pagina 27 é possibile vedere i risultati dei nostri sondaggi. Naturalmente, per questione di privacy, i nomi dei votanti non sono stati raccolti.

L'algoritmo va ora a calcolare le opinioni delle diverse comunitá scientifiche[5] di appartenenza facendo la media ponderata delle opinioni raccolte. Il risultato é mostrato nelle tabelle seguenti:

COMUNITÁ	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**
ICWE	3,38	3,25	3,08	2,75	2,38	2,50	2,40	3,00	3,25
ICPS	4,00	3,50	3,00	1,00	1,00	2,00	3,00	0,00	3,00
BPM	3,73	3,57	3,27	2,8	2,42	2,6	3,5	3	3,25

Tabella 4.1: Opinioni delle comunitá - 1 -

[5]Non avendo avuto dati esaurienti circa la comunitá scientifica di appartenenza e h-index dei vari votanti, abbiamo utilizzato come comunitá scientifica la conferenza a cui quest'ultimi hanno partecipato.

COMUNITÁ	**J**	**K**	**L**	**M**	**N**	**O**	**P**	**Q**
ICWE	2,50	2,50	2,50	2,71	3,42	2,58	2,50	2,00
ICPS	3,00	2,50	3,00	2,00	3,00	2,00	2,00	2,00
BPM	3	2,8	3	2,75	3,6	2,33	2,5	2,4

Tabella 4.2: Opinioni delle comunitá - 2 -

Calcoliamo ora la correlazione fra le varie comunitá scientifiche da cui ricaviamo la seguente matrice di correlazione (Tabella 4.3).

	ICWE	ICPS	BPM
ICWE	1	0.2794118	0.5735294
ICPS	-	1	0.5882353
BPM	-	-	1

Tabella 4.3: Matrice di correlazione.

Infine, grazie a Tab 4.3 possiamo vedere che la correlazione esistente fra ICWE-BPM é molto simile a quella fra ICPS-BPM.

	VOTERS	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
ICWE	icwe1	-	-	-	-	2	2	-	-	4	-	-	-	3	4	3	-	-
	icwe2	4	-	-	-	3	-	-	-	-	-	-	-	-	3	-	-	-
	icwe3	-	-	3	2	2	2	2	-	4	-	-	-	-	4	3	-	-
	icwe4	4	4	4	4	2	1	3	3	2	1	1	3	3	4	2	3	1
	icwe5	2	1	1	-	1	1	1	2	-	-	-	-	1	2	2	-	1
	icwe6	-	-	4	4	1	3	-	-	4	-	-	-	3	4	-	-	-
	icwe7	4	3	2	1	-	-	2	-	2	3	3	2	-	2	2	-	1
	icwe8	-	-	4	-	3	3	-	-	4	-	-	-	-	4	3	-	-
	icwe9	-	-	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	icwe10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	icwe11	4	4	4	-	4	4	-	-	4	-	4	-	-	4	4	-	2
	icwe12	3	4	3	-	2	-	-	-	-	-	2	3	-	3	2	-	2
	icwe13	4	4	3	-	2	-	-	-	2	2	3	-	-	3	1	-	-
	icwe14	-	4	3	-	4	4	-	-	-	-	-	-	3	-	3	-	2
	icwe15	2	2	-	-	2	-	4	4	-	4	2	2	3	4	3	-	3
	icwe16	-	-	3	-	3	-	-	-	-	-	-	-	3	-	3	2	4
ICPS	icps1	4	4	3	1	1	2	3	-	3	3	3	3	2	3	2	2	2
	icps2	-	3	-	-	-	-	-	-	-	-	2	3	-	3	-	-	-
	icps3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	icps4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	icps5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
BPM	bpm1	-	3	3	-	3	-	-	-	-	-	2	-	-	-	-	-	2
	bpm2	-	-	-	-	2	-	-	-	-	-	-	2	-	4	2	-	-
	bpm3	-	-	-	-	3	-	-	-	-	-	-	-	-	-	-	-	3
	bpm4	4	4	2	-	3	-	4	-	1	4	3	4	-	4	1	-	-
	bpm5	4	-	4	-	-	-		-	4	-	-	-	-	4	4	-	-
	bpm6	4	3	4	3	-	-	4	-	3	3	-	4	4	4	2	-	-
	bpm7	4	-	3	-	3	-	-	-		-	-	-	-	-	-	-	-
	bpm8	-	-	4	-	3	2	-	-	4	-	-	-	-	4	-	-	-
	bpm9	3	-	3	2	2	3	-	-	4	-	-	-	-	3	2	-	-
	bpm10	3	-	3	-	3	2	-	-	-	-	-	-	2	-	2	-	2
	bpm11	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	bpm12	4	4	3	2	1	3	3	-	3	2	3	2	2	3	3	2	2
	bpm13	-	4	4	-	2	-	-	-	3	3	3	-	-	4	-	-	-
	bpm14	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	bpm15	4	3		3	3	3	3	3	4	3	-	4	3	4	3		3
	bpm16	-	4	3	-	1	-	-	-	-	-	-	2	-	2	2	3	-
	bpm17	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	bpm18	-	-	-	4	-	-	-	-	-	-	3	-	-	-	-	-	-

Tabella 4.4: Results of surveys conducted on ICWE, ICPS and BPM conferences

4.4 Risultati e analisi dei dati

Presenteremo ora molto brevemente i risultati dell'applicazione dell'algoritmo "Diffcomm" ad alcuni sondaggi online da noi svolti. La versione integrale dei dati é visibile in Appendice A nelle tabelle A.5, A.9 e A.10, rispettivamente a pagina 47, 51 e 52.

Abbiamo analizzato i dati relativi alle comunitá BPM, ICWE e VLDB prendendo come candidati i nove candidati comuni alle tre comunitá. Il risultato é il seguente:

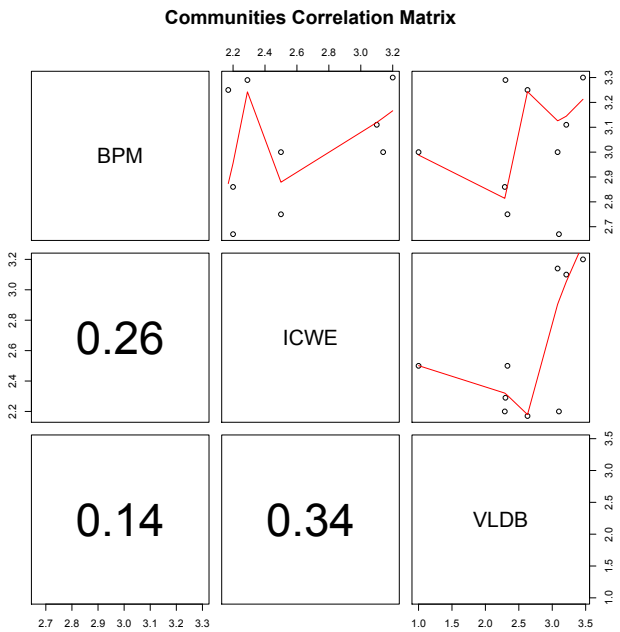


Figura 4.1: Correlazione BPM, ICWE, VLDB conferences

Come possiamo vedere dalla figura 4.1, la correlazione esistente fra le varie comuntá é positiva. Il che sta a significare che **le opinioni delle tre comunitá circa i candidati proposti sono abbastanza simili**. Diciamo abbastanza perché possiamo vedere che, per esempio, la somiglianza fra le opinioni delle comunitá BPM-VLDB (0,14) é molto minore rispetto a quella fra ICWE-VLDB (0,34) che a sua volta risulta essere distante dal valore

dalla correlazione massima che é 1.

In generale, possiamo dire che sia in questo caso, che nel precedente (vedi Sezione 4.3.4 a pagina 24), la correlazione fra le comunitá scientifiche esaminate é sempre positiva. Questo sta ad indicare che, se pur in modo diverso, vi é un legame di somiglianza fra le opinioni delle varie comunitá prese in esame.

Per concludere, vogliamo rivolgere la nostra attenzione su due dati che abbiamo ricavato: la correlazione fra le comunitá BPM-ICWE presente nella tabella 4.3.4 a pagina 27 e quella presente nel grafico 4.1 a pagina 28. Come possiamo vedere, nonostante le comunitá esaminate siano esattamente le stesse, il dato della correlazione é notevolmente diverso. Si passa da un 0,57 (pagina 27) ad un 0,26 (pagina 28). Questa diversitá puó essere causata da diversi fattori tra cui sicuramente:

1. I diversi candidati sui quali si é chiesto di esprimere una opinione;
2. Il diverso numero di votanti;
3. Il diverso contesto in cui é stato somministrato il questionario.

Tutto ció é stato fatto notare per sottolineare l'importanza della rigorositá in un lavoro di ricerca scientifica ma soprattutto per evidenziare l'attenzione che deve essere riposta nella veridicitá di alcuni dati.

Capitolo 5

Metriche scientifiche e valutazioni reali

> Non tutto ció che puó essere contato conta, e non tutto ció che conta puó essere contato.
>
> ALBERT EINSTEIN
> (1879-1955)

In questo quinto capitolo analizzeremo il problema dell'attendibilitá delle attuali metriche di valutazione scientifiche. Dopo una descrizione generale (5.1), andremo ad esaminare, nella sezione 5.2, i criteri di analisi dei dati che abbiamo utilizzato. Infine, nella sezione 5.3 presenteremo i risultati del nostro lavoro.

5.1 Descrizione del problema

Dalla metá degli anni Cinquanta quando Eugene Garfield propose per la prima volta il Citation Index [8] ad oggi, il ruolo degli indici di valutazione scientifica é notevolmente mutato, passando dal semplice supporto alla ricerca di articoli di interesse ad essere uno strumento di valutazione di documenti e persone allo scopo di ripartire i fondi per la ricerca, attrarre finanziamenti e, in ambito accademico, reclutare docenti. [17] Come dice Alberto Baccini in [1] esiste oramai la convinzione che la valutazione possa

essere condotta attraverso indicatori quantitativi in grado di approssimare la qualità, vale a dire il merito di una ricerca e la sua influenza all'interno della comunità scientifica, e l'impatto, ossia gli effetti che essa produce.

Tutti questi indici, peró, hanno limiti notevoli e ben conosciuti che possono trarre in inganno un utilizzatore poco esperto. Un esempio inequivocabile che sottilinea un limite dell'H-index [21] è quello citato in [4] dove viene sottolineata la sua debolezza di fronte ad autori importanti che hanno avuto una carriera molto breve. Si pensi ad Evariste Galois[1], morto in duello a soli ventuno anni, il cui indice H é pari a 2. La sua unica opera, di notevole importanza in campo matematico, venne pubblicata nel 1843.

Nonostante in questi ultimi anni si stia svolgendo una cospicua attivitá di ricerca per trovare una soluzione a questi limiti, ad oggi, non siamo ancora in grado di capire quanto questi indici rispecchino la realtá, ossia quanto un determinato indice sia rappresentativo o meno della reale reputazione di un individuo all'interno di una comunitá scientifica e, piú in generale del mondo scientifico. Quello che abbiamo fatto col nostro lavoro é stato, attraverso il calcolo della correlazione, fornire un dato reale che ci dia un'idea di questa relazione.

5.2 Ipotesi sui risultati e criteri di analisi

In questa sezione, come nel precedente capitolo, presenteremo alcune ipotesi sui risultati e faremo una breve panoramica su come abbiamo impostato il nostro lavoro.

Il nostro scopo era quello di verificare quanto gli attuali indici di valutazione rispecchino le reali opinioni che le varie comunitá scientifiche hanno circa uno scienziato. Per fare ció abbiamo sottoposto un questionario online (vedi Capitolo 3) ad un gruppo di partecipanti ad alcune recenti conferenze. In ogni questionario veniva chiesto al soggetto di esprimere la propria opinione circa l'importanza e l'influenza del lavoro di alcuni candidati, nonché

[1] Bourg-la-Reine, 25 ottobre 1811 –Parigi, 31 maggio 1832

di fornire alcuni dati relativi la comunitá scientifica di appartenenza.

Per ogni sondaggio abbiamo calcolato la media dei voti ricevuti dai vari candidati e abbiamo valutato la correlazione esistente fra quest'ultima e il relativo h-index. Il dato così ottenuto, ci indicherá l'attendibilitá dell'indice h relativo ad un gruppo di candidati scelti appositamente per quel sondaggio relativo ad una determinata comunitá scientifica.

Inoltre, per poter avere un dato piú generale, indipendente dalle varie comunitá a cui sono stati sottoposti i questionari, abbiamo calcolato la media dei voti ricevuti dai vari candidati su tutti i questionari in cui erano presenti e ne abbiamo valutato la correlazione con l'indice h e altri indici di valutazione scientifica. Quest'ultimo dato, a differenza del precedente, essendo indipendente dalle varie comunitá scientifiche, é in grado di fornirci un'idea piú precisa sulla veridicitá degli indici presi in esame.

Per concludere, possiamo dire che quello che ci aspettiamo di trovare dalla nostra ricerca é una serie di correlazioni positive, segno che l'indice h preso in esame rispecchia le opinioni reali dei vari membri delle diverse comunitá scientifiche.

5.3 Risultati e analisi dei dati

Presenteremo ora i risultati del calcolo della correlazione per ogni comunitá scientifica esaminata. I dati integrali dei sondaggi sono visibili in Appendice A a pagina 39.

Quello che possiamo vedere immediatamente dalla tabella 5.1 a pagina 34 é che su otto "comunitá scientifiche" prese in esame, la correlazione di quattro di queste é positiva (righe di colore giallo), mentre la correlazione delle rimanenti quattro é negativa (righe di colore blu). Possiamo, inoltre, notare che l'intervallo di appartenenza dei valori é

$$-0,18 \leq x \leq 0,25$$

COMUNITÁ	CORRELAZIONE con H-Index
AAI	0,06
BPM	0,22
ER	0,15
EUROCRYPT	-0,18
ICSE	-0,08
ICWE	0,25
VLDB	-0,09
GENERAL	-0,17

Tabella 5.1: Tabella delle correlazioni fra i voti dei questionari online e l'indice h dei candidati.

ossia molto ristretto attorno allo 0 (no correlazione) rispetto ad un normale intervallo dei valori di correlazione ($-1 \leq x \leq 1$).

A causa dell'esiguo numero di partecipanti a questi sondaggi, possiamo dire che non tutti i valori riportati nella tabella 5.1 sono da considerarsi significativi in quanto le opinioni raccolte sono magari l'espressione solo di un paio di membri della comunitá. Per questo motivo abbiamo deciso di modificare la tabella precedente mostrando solamente le correlazioni ricavate dai questionari che hanno avuto un numero di votanti almeno pari a 14. I valori sono riportati nella seguente tabella:

COMUNITÁ	CORRELAZIONE con H-Index
BPM	0,22
ICWE	0,25
VLDB	-0,09
GENERAL	-0,17

Tabella 5.2: Tabella delle correlazioni fra i voti dei questionari online con numero di votanti ≥ 14 e l'indice h dei candidati.

Da qui si vede chiaramente come via sia una correlazione abbastanza positiva fra le opinioni dei membri delle comunitá BPM e ICWE e gli indici

h dei candidati scelti per questi due sondaggi. Si nota, invece, come per le rimanenti due comunitá (VLDB, GENERAL) vi sia una correlazione leggermente negativa. Questi dati, purtroppo, non ci aiutano nel nostro compito. Possiamo solo dire che, *l'h-index risulta essere piú o meno valido a seconda della comunitá scientifica presa in esame.*

Per poter avere un dato indipendente dalle comunitá scientifiche, come abbiamo detto precedentemente, abbiamo calcolato la correlazione esistente fra la media di tutti i voti ricevuti dai vari candidati su tutti i sondaggi e alcune delle piú note metriche di valutazione del lavoro e delle persone del mondo della scienza. I risultati sono riportati nella tabelle 5.3 e 5.4 a pagina 36.

Anche qui, come nelle tabelle precedenti, possiamo notare che l'intervallo dei valori $(-0,28 \leq x \leq 0,36)$ é sempre molto ristretto rispetto ad un normale intervallo di valori di correlazione. Questo significa che le correlazioni trovate, siano queste negative o positive, non sono forti ma solamente accennate. Gli indici che rispecchiano maggiormente le opioni dei votanti risultano essere: Citations/Paper (RESEVAL) e # Publications (DBLP). Al contrario, l'indice che si discosta maggiormente dalle opinioni riscontrate risulta essere: # Publications (RESEVAL). L'indice H fino ad ora esaminato, indicato in blu nelle tabelle, risulta avere una correlazione positiva, segno che é abbastanza rappresentativo dell'opinione comune.

Concludendo, possiamo dire che la maggior parte degli indici esaminati (quasi il 60%) risultano essere abbastanza rappresentativi delle opinioni dei vari membri delle comunitá scientifiche prese in esame.

[2]http://www.cs.ucla.edu/∼palsberg/h-number.html
[3]http://dblp.uni-trier.de/ - http://www.inormatik.uni-trier.de/∼ley/db/
[4]http://reseval.org - http://project.liquidpub.org/reseval
[5]http://readermeter.org

METRIC	SOURCE	CORRELATION
H-Index	**Palsberg List**[2]	**0,124**
# Publications	DBLP[3]	0,350
H-Index	RESEVAL[4]	-0,009
G-Index	RESEVAL	-0,045
# Citations	RESEVAL	0,022
# Publications	RESEVAL	-0,283
Citations/Paper	RESEVAL	0,360
HR-Index	ReaderMeter[5]	0,019
GR-Index	ReaderMeter	0,004
Most Read Pub.	ReaderMeter	0,076
# Publications	ReaderMeter	-0,112
# Bookmarks	ReaderMeter	-0,052

Tabella 5.3: Tabella delle correlazioni fra i voti dei questionari online e varie metriche di valutazione.

METRIC	SOURCE	CORRELATION
Citations/Paper	RESEVAL	0,360
# Publications	DBLP	0,350
H-Index	**Palsberg List**	**0,124**
Most Read Pub.	ReaderMeter	0,076
# Citations	RESEVAL	0,022
HR-Index	ReaderMeter	0,019
GR-Index	ReaderMeter	0,004
H-Index	RESEVAL	-0,009
G-Index	RESEVAL	-0,045
# Bookmarks	ReaderMeter	-0,052
# Publications	ReaderMeter	-0,112
# Publications	RESEVAL	-0,283

Tabella 5.4: Tabella ordinata delle correlazioni fra i voti dei questionari online e varie metriche di valutazione.

Capitolo 6

Conclusioni e considerazioni finali

In questa tesi abbiamo presentato il lavoro che é stato svolto per incrementare la nostra conosenza sulle comunitá scientifiche e i legami esistenti tra quest'ultime, nonché verificare l'attendibilitá degli attuali indici di valutazione dei lavoro e delle persone appartenenti al mondo scientifico.

Abbiamo visto che, nei casi presi in esame, la correlazione esistente fra le opinioni delle comunitá scientifiche é positiva. Questo sta a significare che le idee di queste comunitá sono abbastanza simili e che, di conseguenza, esiste, se pur in forma differente, un legame fra quest'ultime. Legame che abbiamo potuto anche quantificare attraverso l'uso dell'algoritmo "Diffcomm" presentato nel Capitolo 4.

Abbiamo, inoltre, appurato che, la correlazione esistente fra l'indice H e le opinioni raccolte, é positiva. Quest'indice, perció, puó essere ritenuto abbastanza rappresentativo delle idee delle varie comunitá scientifiche. In generale, possiamo dire che la maggior parte delle metriche di valutazione esaminate (60% circa) risultano essere abbastanza attendibili; abbastanza perché, come spiegato nel Capitolo 5, i valori di correlazione sono molto bassi.

I benefici portati da questo lavoro, perció, possono essere riassunti qui di seguito:

- Abbiamo verificato l'esistenza e quantificato il legame presente fra alcune comunitá scientifiche;
- Abbiamo proposto un nuovo algoritmo "Diffcomm" che, presi in input una serie di dati legati alle opinioni di alcuni membri di diverse comunitá scientifiche, calcola la correlazione esistente fra quest'ultime;
- Attraverso i dati raccolti da alcuni questionari online, abbiamo analizzato la costituzione, la distribuzione dei voti e il modo di votare di alcune comunitá scientifiche (vedi Appendice A);
- Per ogni comunitá presa in esame, abbiamo calcolato la correlazione esistente fra l'indice H (Palsberg List) e le opinioni dei suoi membri relative ad uno specifico gruppo di candidati;
- Infine, abbiamo calcolato la correlazione esistente fra una serie di indici di valutazione scientifica e le opinioni ricavate da tutti i sondaggi svolti.

Tutto ció, inoltre, potrá essere molto utile in futuro per poter proporre un nuovo metodo di valutazione dei ricercatori e del lavoro scientifico in grado di esaminare l'impatto reale di una ricerca tenendo conto del contesto delle comunitá in cui questa si va ad inserire.

Appendice A

Dati raccolti e analisi svolte

In questo capitolo presentiamo integralmente i dati ricavati dai sondaggi online da noi svolti, nonché una breve analisi, inerente a quanto detto nei Capitoli 4 e 5, di questi risultati. Tutti i seguenti sondaggi sono stati svolti in forma anonima anche se é stato chiesto a chi vi partecipava di fornire alcune informazioni essenziali come la fascia di etá di appartenenza, il genere e la posizione. Ogni sondaggio ha una lista di candidati che si diversifica a seconda delle diverse conferenze a cui sono legati i sondaggi.

A.1 AAI

A.1.1 Dati raccolti

Il sondaggio effettuato é visibile al sito:

$http://FreeOnlineSurveys.com/rendersurvey.asp?sid = ay1i06gbp6h94ha813237$

I dati raccolti sono visibili nella tabella A.4 a pagina 46.

A.1.2 Analisi della correlazione con h-index

Abbiamo calcolato la correlazione esistente fra l'indice h dei candidati e le opinioni relative a quest'ultimi ricavate dal nostro sondaggio. Il risultato é il seguente: **0.05526571**.

A.1.3 Analisi dei voti della comunitá

Vista la scarsa quantitá di dati non possiamo dire molto circa le opinioni dei membri di questa comunitá. Quello che possiamo vedere é che il campione a cui il nostro sondaggio é stato sottoposto, é composto prevalentemente da professori maschi con un'etá media attorno ai 40 anni e che quasi il 40% dei voti, come si puó vedere nel grafico A.1, é il massimo fra quelli consentiti dal sondaggio.

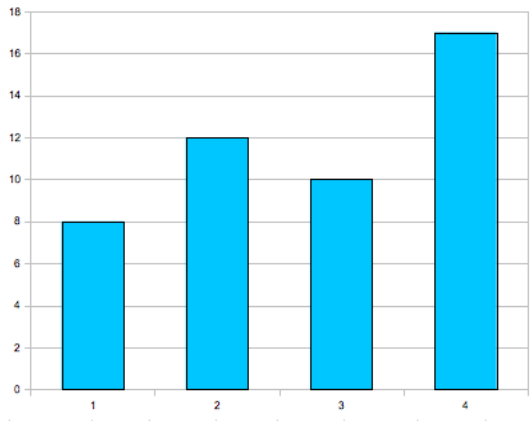


Figura A.1: Grafico distribuzione voti AAI conference

A.2 BPM.ICSOC.ICWS

A.2.1 Dati raccolti

Il sondaggio effettuato é visibile al sito:

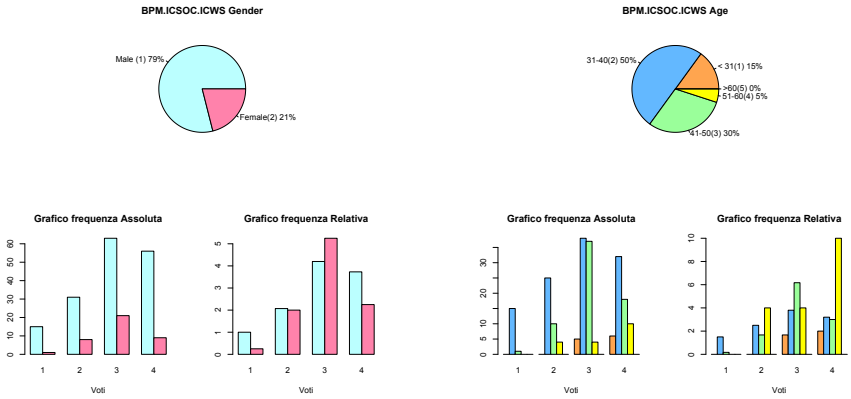
$http://freeonlinesurveys.com/rendersurvey.asp?sid = b1do6v0ur491dtu808890$

I dati raccolti sono visibili nella tabella A.5 a pagina 47.

A.2.2 Analisi della correlazione con h-index

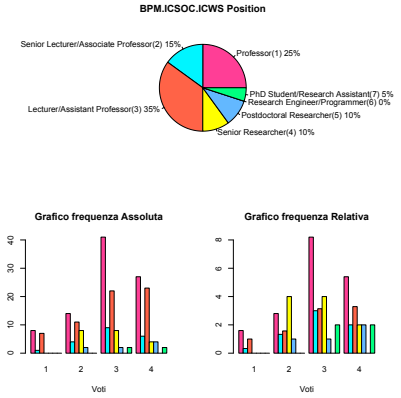
Abbiamo calcolato la correlazione esistente fra l'indice h dei candidati e le opinioni relative a quest'ultimi ricavate dal nostro sondaggio. Il risultato é il seguente: **0.2194013**.

A.2.3 Analisi dei voti della comunitá



(a) Genders

(b) Age



(c) Position

Figura A.2: Grafici distribuzione voti BPM.ICSOC.ICWS cenferences

Da questi grafici possiamo notare alcune cose:

- Il 21% del nostro campione é costituito da femmine il cui voto é stato in prevalenza 3, mentre il 79% é costituito da maschi. Anche in questo caso il voto prevalente é stato 3 ma, diversamente dalle prime, la differenza fra la frequenza del voto 3 e gli altri voti é minore;
- Il 50% del campione é costituito da persone con una etá compresa fra i 31-40 anni. I voti di questa fetta di persone si distribuscono abbastanza equamente fra la scala di indici a disposizione. Da notare in questo grafico é la spiccata prevalenza di voti alti(4) nel 5% della popolazione

composta da persone fra i 51-60 anni. Quest'ultimo dato, nonostante sembri essere significativo percentualmente parlando, in realtá perde di importanza nel momento in cui si nota che questo 5% indica una sola persona e che quindi il voto si riferisce alle opinioni di una sigola persona.

- In generale, vi é una spiccata prevalenza nell'usare valori alti nella votazione (3-4) o nel non votare affatto. Poca é la frequenza di voti bassi come 1 o 2.

A.3 ER

A.3.1 Dati raccolti

Il sondaggio effettuato é visibile al sito:

$http://FreeOnlineSurveys.com/rendersurvey.asp?sid = hbas7r98593zdbs812397$

I dati raccolti sono visibili nella tabella A.6 a pagina 48.

A.3.2 Analisi della correlazione con h-index

Abbiamo calcolato la correlazione esistente fra l'indice h dei candidati e le opinioni relative a quest'ultimi ricavate dal nostro sondaggio. Il risultato é il seguente: **0.1510995**.

A.4 EUROCRYPT

A.4.1 Dati raccolti

Il sondaggio effettuato é visibile al sito:

$http://FreeOnlineSurveys.com/rendersurvey.asp?sid = 3bq8m3f7tmwqedd812396$

I dati raccolti sono visibili nella tabella A.7 a pagina 49.

A.4.2 Analisi della correlazione con h-index

Abbiamo calcolato la correlazione esistente fra l'indice h dei candidati e le opinioni relative a quest'ultimi ricavate dal nostro sondaggio. Il risultato é il seguente: **-0.1830964**.

A.5 ICSE

A.5.1 Dati raccolti

Il sondaggio effettuato é visibile al sito:

$http://FreeOnlineSurveys.com/rendersurvey.asp?sid = bmsv4bjfbli9tdl824346$

I dati raccolti sono visibili nella tabella A.8 a pagina 50.

A.5.2 Analisi della correlazione con h-index

Abbiamo calcolato la correlazione esistente fra l'indice h dei candidati e le opinioni relative a quest'ultimi ricavate dal nostro sondaggio. Il risultato é il seguente: **-0.07583338**.

A.6 ICWE.WISE

A.6.1 Dati raccolti

Il sondaggio effettuato é visibile al sito:

$http://freeonlinesurveys.com/rendersurvey.asp?sid = 6uopawl0dp82gz2812399$

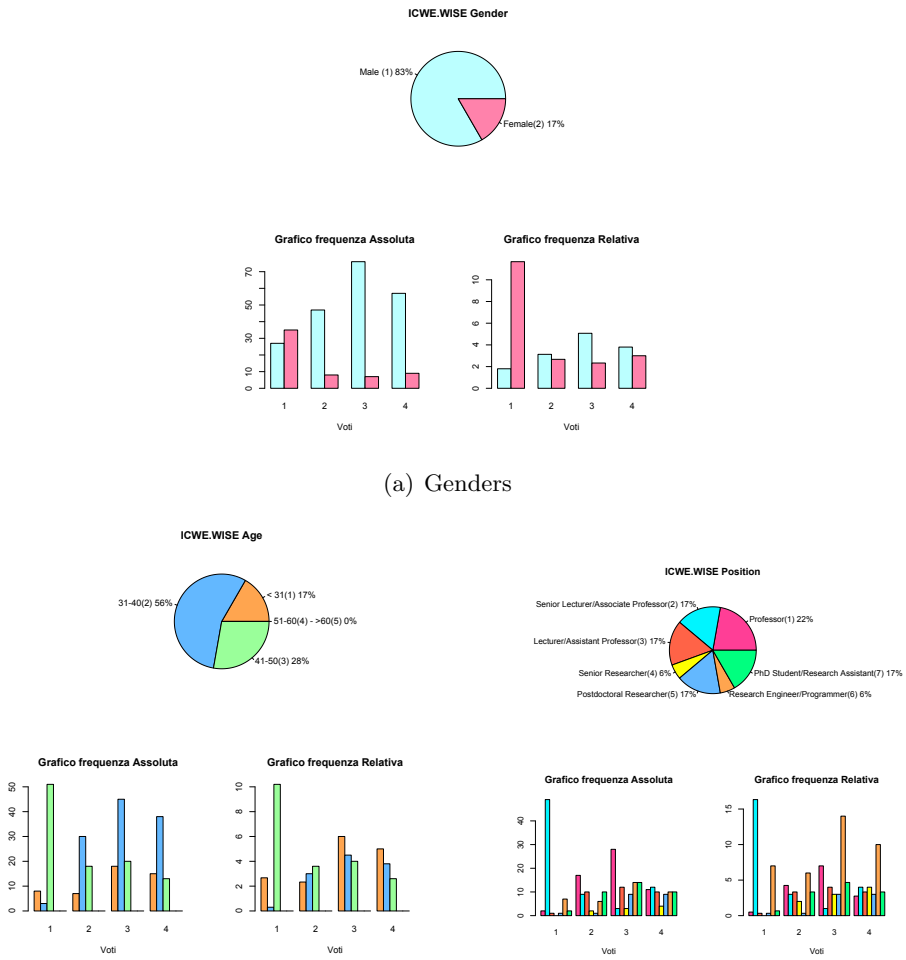
I dati raccolti sono visibili nella tabella A.9 a pagina 51.

A.6.2 Analisi della correlazione con h-index

Abbiamo calcolato la correlazione esistente fra l'indice h dei candidati e le opinioni relative a quest'ultimi ricavate dal nostro sondaggio. Il risultato é il seguente: **0.2511704**.

A.6.3 Analisi dei voti della comunitá

Quello che possiamo notare in questo caso é la spiccata preminenza del voto 1, sia nel sesso femminile, che nelle persone fra i 41-50 anni e soprattutto fra i professori. Questa puó essere considerata una inversione di tendenza rispetto ai precedenti sondaggi, dove i voti piú usati erano quelli alti(3-4) e, di conseguenza, forse dimostrano un'indice della maggiore veridicitá delle votazioni.



(a) Genders

(b) Age

(c) Position

Figura A.3: Grafici distribuzione voti ICWE.WISE cenferences

A.7 VLDB

A.7.1 Dati raccolti

Il sondaggio effettuato é visibile al sito:

$http://FreeOnlineSurveys.com/rendersurvey.asp?sid = tu35q8rz122uh0m812398$

I dati raccolti sono visibili nella tabella A.10 a pagina 52.

A.7.2 Analisi della correlazione con h-index

Abbiamo calcolato la correlazione esistente fra l'indice h dei candidati e le opinioni relative a quest'ultimi ricavate dal nostro sondaggio. Il risultato é il seguente: **-0.09126259**.

A.7.3 Analisi dei voti della comunitá

Anche qui, come abbiamo visto precedentemente, vi é una spiccata preminenza di voti 1 sia nel sesso femminile, che nelle persone con etá inferiore ai 31 anni e soprattutto fra i "PhD student" e fra i "Research Assistent".

A.8 GENERAL

A.8.1 Dati raccolti

Il sondaggio effettuato é visibile al sito:

$http://FreeOnlineSurveys.com/rendersurvey.asp?sid = tu35q8rz122uh0m812398$

I dati raccolti sono visibili nella tabella A.11 a pagina 53.

A.8.2 Analisi della correlazione con h-index

Abbiamo calcolato la correlazione esistente fra l'indice h dei candidati e le opinioni relative a quest'ultimi ricavate dal nostro sondaggio. Il risultato é il seguente: **-0.1743310**.

AAI

	Q1a Hsinchun Chen	Q1b Daphne Koller	Q1c Alon Halevy	Q1d Tomaso Poggio	Q1e W. Bruce Croft	Q1f C. Lee Giles	Q1g Joseph M. Hellerstein	Q1h Victor Lesser	Q1i Sebastian Thrun	Q1j David Salesin	Q1k Ronald L. Rivest	Q1l Ian Horrocks	Q1m Alex Pentland	Q1n Jeffrey D. Ullman	Q1o Jiawei Han	Q1p Vipin Kumar	Q1q David Haussler	Q1r Aaron Bobick	Q1s Christos H. Papadimitriou	Q1t Jitendra Malik	Q1u Brad Myers	Q1v Daniel Sleyvorek	Q1w Moshe Y. Vardi
V1								2				2	2	4	3				4				4
V2			3									2		4					4				4
V3			3	3								3		4					4				4
V4		4		4				4	4		4	1		3					4				3
V5		3		2																			
Numbers of people who voted	0	2	2	3	0	0	0	2	1	0	1	4	0	4	1	0	0	0	4	0	0	0	4
Average rating	0	3,5	3	3	0	0	0	3	4	0	4	2	0	3,75	3	0	0	0	4	0	0	0	3,75
H-index	50	49	62	89	49	49	51	45	49	44	43	63	83	87	51	45	70	45	79	55	49	49	72

	Q1x James Hendler	Q1y Geoffrey E. Hinton	Q1z Demetri Terzopoulos	Q1{ Philip S. Yu	Q1| Frank van Harmelen	Q1} Giuseppe De Giacomo	Q1~ Douglas C. Schmidt	Q1 Thomas S. Huang	Q1ˇ Nick Jennings	Q1˘ H. V. Jagadish	Q1˙ Didier Dubois	Q1¨ Deborah L. McGuinness	Q1˚ Bernhard Schölkopf	Q1¸ Andrew Blake	Q1˝ Takeo Kanade	Q1˛ Alan Yuille	Q1ˇ Steffen Staab	Q2 Age group	Q3 Gender	Q4 Position
V1					2	2			2	2								41 – 50	Male	Professor
V2						2												41 – 50	Male	Professor
V3	1				1	3			2	4		1					2	41 – 50	Male	Senior Lecturer / Associate Professor
V4															4			31 – 40	Male	Professor
V5					1	2			1			1					1	31 – 40	Male	Professor
Numbers of people who voted	1	0	0	0	3	4	0	0	3	2	0	3	0	0	1	0	2			
Average rating	1	0	0	0	1,33	2,25	0	0	1,67	3,5	0	1,33	0	0	4	0	1,5			
H-index	58	71	64	74	52	48	46	46	82	44	41	41	71	57	68	50	61			

	Q1x James Hendler	Q1y Geoffrey E. Hinton	Q1z Demetri Terzopoulos	Q1{ Philip S. Yu	Q1| Frank van Harmelen	Q1} Giuseppe De Giacomo	Q1~ Douglas C. Schmidt	Q1 Thomas S. Huang	Q1ˇ Nick Jennings	Q1˘ H. V. Jagadish	Q1˙ Didier Dubois	Q1¨ Deborah L. McGuinness	Q1˚ Bernhard Schölkopf	Q1¸ Andrew Blake	Q1˝ Takeo Kanade	Q1˛ Alan Yuille	Q1ˇ Steffen Staab
Numbers of people who voted (> 0)	1	-	-	-	3	4	-	-	3	2	-	3	-	-	1	-	2
Average rating	1	-	-	-	1,33	2,25	-	-	1,67	3,5	-	1,33	-	-	4	-	1,5
H-index	58	-	-	-	52	48	-	-	82	44	-	41	-	-	68	-	61
Candidate position by h-index	10	-	-	-	11	15	-	-	3	17	-	19	-	-	6	-	9
Candidate position by average rating	19	-	-	-	17	13	-	-	15	8	-	18	-	-	4	-	16

Figura A.4: Results of online survey conducted on AAI conference.

BPM.ICSOC.ICWS

	Q1a Richard Baranuk	Q1b George Varghese	Q1c Stefano Ceri	Q1d Deborah L. McGuinness	Q1e Wil van der Aalst	Q1f Sushil Jajodia	Q1g Giuseppe De Giacomo	Q1h Amit Sheth	Q1i Christian S. Jensen	Q1j Frank van Harmelen	Q1k Nick McKeown	Q1l David Haussler	Q1m Avi Wigderson	Q1n Richard Snodgrass	Q1o Michael Stonebraker	Q1p Jeffrey D. Ullman	Q1q Robert Tarjan
v1			3		3											4	
v2			4		2					2							
v3			3		3		2	2		3	3					3	
v4			3	2	4					2						3	
v5					4		2	3	3	3				3	4	4	4
v6			3	2	4	2	2	3	3						4	4	
v7			4		3				3								
v8					4							3		3	3	4	3
v9				3	3	3	3	2							4	4	
v10			3		4												
v11					4		4	4	3							3	
v12			2	3	3	4	2	4	3					3	3	4	
v13		3	3		4	4	4	2	4					2	4	4	
v14			4	2	4		3	4	3								
v15			1		3		1								1	1	
v16			3		3	2											3
v17			4		4			3	4					3	3	4	4
v18			1	1	1		4	1		1							
v19			4	4	4		3	3		3				3		4	4
v20			4		4										2		4
Numbers of people who voted	1	1	15	6	19	5	9	12	7	7	1	1	0	6	9	13	5
Average rating	2	3	3	2,33	3,37	3	2,67	2,75	3,29	2,43	3	3	0	2,83	3,11	3,54	3,6
H-index	48	40	41	41	68	65	48	63	45	52	42	70	43	45	49	87	82

	Q1r Rakesh Agrawal	Q1s Andrew Blake	Q1t Pat Hanrahan	Q1u Margaret Martonosi	Q1v Won Kim	Q1w Steffen Staab	Q1x Ion Stoica	Q1y David J. DeWitt	Q1z Jim Kurose	Q1α Dieter Fensel	Q1β David Maier	Q1γ David E. Goldberg	Q1~ Hector Garcia-Molina	Q1 Maurizio Lenzerini	Q1ϑ Zoubin Ghahramani	Q1ϕ John A. Stankovic	Q1ϰ Alberto Sangiovanni-Vincentelli	Q1ϱ Milind Tambe
v1										3				2				
v2										3			4					
v3										2			3	3				
v4						2				3								
v5																		
v6	3				2	2		2		1			3	3		3		
v7	4	3						3		2			4	2				
v8	3					3							4	3				
v9	3	3						3		3	3	4	2					
v10					3					1								
v11								3						4				
v12	3				3			3		2	4		3					
v13	4				4			3			3		4	4				
v14										3			4	1				
v15								2		2			2	4				
v16										1				4				
v17	4					1				1			4	4			2	
v18						2				2			4	2			3	
v19																		
v20																		
Numbers of people who voted	7	2	0	0	4	5	0	6	0	14	4	1	12	12	0	0	0	3	0
Average rating	3,43	3	0	0	3	2	0	2,67	0	2,07	3,25	4	3,42	3	0	0	2,67	0
H-index	52	57	42	40	43	61	65	49	47	56	43	54	89	56	47	46	50	47

(a) Part a

	Q1b Geoffrey E. Hinton	Q1f Eitan Altman	Q1j Alon Halevy	Q1λ Songwu Lu	Q1θ Ahmed Elmagarmid	Q2	Q3	Q4
v1						41-50 years old	Male	Senior Lecturer/Associate Professor
v2			4			31-40 years old	Male	Postdoctoral Researcher
v3			4			31-40 years old	Female	Lecturer/Assistant Professor
v4						31-40 years old	Male	Senior Researcher
v5						less than 31 years old	Male	Postdoctoral Researcher
v6			3		2	31-40 years old	Male	Professor
v7			3			31-40 years old	Male	Lecturer/Assistant Professor
v8						31-40 years old		Lecturer/Assistant Professor
v9	3		4		4	41-50 years old	Male	Professor
v10					2	41-50 years old	Female	Senior Lecturer/Associate Professor
v11						41-50 years old	Male	Lecturer/Assistant Professor
v12						less than 31 years old	Male	Senior Lecturer/Associate Professor
v13					3	41-50 years old	Female	Professor
v14					4	51-60 years old	Male	Professor
v15			3			31-40 years old	Male	Lecturer/Assistant Professor
v16			1		2	31-40 years old	Male	Lecturer/Assistant Professor
v17			4			less than 31 years old	Female	PhD Student/Research Assistent
v18			4		3	31-40 years old	Male	Lecturer/Assistant Professor
v19						31-40 years old	Male	Professor
v20			3			41-50 years old	Male	Senior Researcher
Numbers of people who voted	1	0	10	0	7			
Average rating	3	0	3,3	0	2,86			
H-index	71	41	62	41	43			

(b) Part b

Figura A.5: Results of online survey conducted on BPM conference.

ER

	Q1a	Q1b	Q1c	Q1d	Q1e	Q1f	Q1g	Q1h	Q1i	Q1j	Q1k	Q1l	Q1m	Q1n	Q1o	Q1p	Q1q	Q1r	Q1s
	Dieter Fensel	Won Kim	David Maier	C.A.R. Hoare	Stefano Ceri	Hector Garcia-Molina	Avi Wigderson	Tuomas Sandholm	Amit Sheth	Alan Yuille	Ken Kennedy	Ian Horrocks	Deborah L. McGuinness	Randy H. Katz	Jiawei Han	Maurizio Lenzerini	Robbert van Renesse	Joachim Weickert	Christian S. Jensen
v1	1				3	3						3	1			3			1

Numbers of people who voted	1	0	0	0	1	1	0	0	0	0	0	1	1	0	0	1	0	0	1
Average rating	1	0	0	0	3	3	0	0	0	0	0	3	1	0	0	3	0	0	1
H-index	56	43	43	47	41	89	43	46	63	50	54	63	41	56	51	56	44	41	45

	Q1t	Q1u	Q1v	Q1w	Q1x	Q1y	Q1z	Q1{	Q1|	Q1}	Q1~	Q1$	Q1%	Q1^	Q1&	Q1*	Q1(	Q1)	Q1_	Q1+	Q1=	Q1[	Q1]	Q1{
	Don Towsley	Christos Faloutsos	Gaurav Sukhatme	Prabhakar Raghavan	Heikki Mannila	Michael J. Carey	Alex Pentland	Oded Goldreich	Wil van der Aalst	Steffen Staab	Richard Snodgrass	Hector Levesque	Anil K. Jain	Sushil Jajodia	Jim Gray	Giuseppe De Giacomo	Jim Kurose	Alan Newell	Douglas C. Schmidt	David J. DeWitt	Leslie Lamport
v1										2	1	4			4	3		2			

Numbers of people who voted	0	0	0	0	0	0	0	0	0	1	1	1	0	0	1	1	0	1	0	0	0	0	0	0
Average rating	0	0	0	0	0	0	0	0	0	2	1	4	0	0	4	3	0	2	0	0	0	0	0	0
H-index	77	44	45	45	45	48	83	44	68	61	45	49	86	65	45	48	47	57	46	49	57

v1 e di sesso maschile, età 41-50 anni e posizione "Senior Lecturer/Associate Professor".

Figura A.6: Results of online survey conducted on ER conference.

EUROCRYPT

	Q1a	Q1b	Q1c	Q1d	Q1e	Q1f	Q1g	Q1h	Q1i	Q1j	Q1k	Q1l	Q1m	Q1n	Q1o	Q1p	Q1q	Q1r	Q1s
	Avi Wigderson	Andrew Zisserman	Kurt Keutzer	Adi Shamir	Jon Crowcroft	Adrian Perrig	Francisco Herrera	B. R. Badrinath	John A. Stankovic	Rajeev Motwani	David L. Dill	David Haussler	Oded Goldreich	Moni Naor	Jason Cong	Christos H. Papadimitriou	David S. Johnson	Hongjiang Zhang	Gene Tsudik
V1	2	1	1	4	1	1	1	1	1	4	1	1	3	3	1	1	1	1	2
V2	4			4									4	4		4			
V3	3			4									3	2					
Numbers of people who voted	3	1	1	3	1	1	1	1	1	1	1	1	3	3	1	2	1	1	1
Average rating	3	1	1	4	1	1	1	1	1	4	1	1	3.33	3	1	2.5	1	1	2
H-index	43	72	46	53	49	56	55	45	46	45	56	70	44	49	51	79	40	73	45

	Q1t	Q1u	Q1v	Q1w	Q1x	Q1y	Q1z	Q1{	Q1|	Q1}	Q1~	Q1	Q1a	Q1b	Q1c	Q1d	Q1e	Q1f	Q1g
	Prabhakar Raghavan	Ronald L. Rivest	George Varghese	Leonidas J. Guibas	Martin Vetterli	Chandrajit L. Bajaj	Baruch Awerbuch	David Culler	Robert Tarjan	Noga Alon	Rafail Ostrovsky	Jon M. Kleinberg	Barbara Liskov	Heikki Mannila	Sally Floyd	Deborah Estrin	Mihir Bellare	Dinesh Manocha	Robert Kraut
V1	1	4	1	1	1	1	1	1	1	1	2	1	1	1	1	1	4	1	1
V2		4									3						4		
V3		4									3			3			3		
Numbers of people who voted	1	3	1	1	1	1	1	1	1	1	3	1	1	2	1	1	3	1	1
Average rating	1	4	1	1	1	1	1	1	1	1	2.67	1	1	2	1	1	3.67	1	1
H-index	45	43	40	49	64	43	51	70	82	49	43	41	43	45	53	94	57	55	58

v1 è di sesso maschile, età 31-40 anni, posizione "Professor".
v2 è di sesso maschile, età 31-40 anni, posizione "Research Group Leader".
v3 è di sesso maschile, età minore di 31 anni, posizione "PhD Student/Research Assistant".

Figura A.7: Results of survey conducted on EUROCRYPT conference.

ICSE

	Q1a	Q1b	Q1c	Q1d	Q1e	Q1f	Q1g	Q1h	Q1i	Q1j	Q1k	Q1l	Q1m	Q1n	Q1o	Q1p	Q1q	Q1r	Q1s
	Sushil Jajodia	Ian Horrocks	David Garlan	HongJiang Zhang	Jitendra Malik	Amir Pnueli	Eduardo Sontag	David Karger	Baruch Awerbuch	Joseph M. Hellerstein	Thomas W. Reps	Michael Stonebraker	Gregory D. Abowd	David Harel	Douglas C. Schmidt	Carl Kesselman	Zbigniew Michalewicz	Rajeev Rastogi	David Salesin
v1	1	1	3	1	1	4	1	1	1	1	1	1	1	4	3	1	1	1	1
v2			3			4								4					
Numbers of people who voted	1	1	2	1	1	2	1	1	1	1	1	1	1	2	1	1	1	1	1
Average rating	1	1	3	1	1	4	1	1	1	1	1	1	1	4	3	1	1	1	1
H-index	65	63	56	73	55	54	53	52	51	51	51	49	46	46	46	45	45	44	44

	Q1t	Q1u	Q1v	Q1w	Q1x	Q1y	Q1z	Q1{	Q1|	Q1}	Q1~	Q1	Q1	Q1	Q1	Q1	Q1	Q1	Q1
	Erol Gelenbe	Barbara Liskov	Krithi Ramamritham	Barry Boehm	Thomas A. Henzinger	Alex Aiken	Paul Dourish	Monica S. Lam	Craig Chambers	Edmund M. Clarke	David S. Johnson	Ahmed Elmagarmid	Badri Nath [B R Badrinath]	Bernhard Schölkopf	Brad Myers	C.A.R. Hoare	Chandrajit Bajaj	Ian Foster	John Stasko	Nancy Lynch	Victor Basili
v1	2	4	1	4	3	1	1	1	1	4	1	1	1	1	2	4	1	2	1	3	4
v2										4											3
Numbers of people who voted	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	2
Average rating	2	4	1	4	3	1	1	1	1	4	1	1	1	1	2	4	1	2	1	3	3,5
H-index	44	43	43	42	42	41	41	41	40	40	40	43	45	71	49	47	43	90	42	40	44

v1 è di sesso maschile, età 41-50 anni, posizione "Senior Lecturer/ Associate Professor".
v2 è di sesso maschile, età 31-40 anni, posizione "Lecturer/ Assistant Professor".

Figura A.8: Results of survey conducted on ICSE conference.

ICWE.WISE

	Q1a Jeffrey D. Ullman	Q1b Vipin Kumar	Q1c Michael Stonebraker	Q1d Yern Pavson	Q1e C. A. R. Hoare	Q1f Carl Kesselman	Q1g W. Bruce Croft	Q1h Stefano Ceri	Q1i Giuseppe De Giacomo	Q1j Dieter Fensel	Q1k Luca Cardelli	Q1l Nicholas R. Jennings	Q1m Daphne Koller	Q1n Ian T. Foster	Q1o C. Lee Giles	Q1p Frank van Harmelen	Q1q Jason Cong	Q1r Benjamin C. Pierce	Q1s David Maier	Q1t Christos Faloutsos
v1	3	1	1	1	1	1	1	4	1	1	1	1	1	1	1	1	1	1	1	1
v2	4		4		4			3			3									
v3					4					2	3	4				2				
v4	4				4					3				3				3		
v5	4	3	4	3	3	3	3	3	2	2						2		2	2	
v6	2	1	1		2	1	1	4	4	2		1	2	1	3	1	3	2	3	4
v7								1		1	1	1		3		1		1	1	2
v8	4	3	3		4			4	4	1	4	3	3			3	4	3	2	
v9								2			2									
v10								4		3	4									
v11								3						3		3				
v12	4		4		3	4			4		3	3			2					
v13	3		4		3			3	2					2						
v14	4		4		4	3		3	2		3					4				
v15								3	4											
v16	2	2	2	2	3	1	3	3	2		3		4	4	4			2	4	
v17									2			2								
v18								3	3											
Numbers of people who voted	10	4	10	3	11	7	3	14	5	14	9	7	3	8	3	9	3	7	6	3
Average rating	3.4	2.25	3.1	2	3.18	2.29	2.33	3.14	2.4	2.29	2.56	2.14	2	2.38	2.67	2.33	2.67	2	2.17	2.33
H-index	87	45	49	43	47	45	49	41	48	55	58	82	49	90	49	52	51	40	43	44

	Q1u Geoffrey E. Hinton	Q1v Alon Y. Halevy	Q1w Joachim Weickert	Q1x Edmund M. Clarke	Q1y David J. DeWitt	Q1z Christian S. Jensen	Q1! Ahmed K. Elmagarmid	Q1@ Serge Abiteboul	Q1# Deborah L. McGuinness	Q1$ Hector Garcia Molina	Q1% Amit P. Sheth	Q1^ Subhi Jajodia	Q1& Demetri Terzopoulos	Q1* Ben Bederson	Q1(Ramesh Govindan	Q1) Nancy Lynch	Q1[David Harel	Q1] Steffen Staab	Q1{ Mario Gerla	Q1} Michael J. Carey
v1	1	3	1	1	1	1	1	2	1	1	3	1	1	1	1	1	1	1	1	2
v2		3						2	1	3	4	2					3			
v3		4						4		4	3									
v4								3		3	3									
v5								4	3	4										
v6	3	4			1	1	3	4	3	4	3		4	3	2			3	1	4
v7	2	2						4	1	2						1		1	1	
v8		4						4	3	2										4
v9		2						2		4										
v10		4				3	2	3		2		3					4			
v11								4		3										
v12		4				4		4		4							4			
v13		2				2	3	3		3							2			
v14						3		3		3							2			3
v15										1										
v16										3										
v17	4					2	3	3	3	3					4	4	2	2	3	1
v18				2		2	2	3	3	3					4	3	2	4		3
Numbers of people who voted	4	10	2	2	5	7	5	13	8	14	14	3	2	2	4	4	10	9	2	7
Average rating	2.5	3.2	1	1.5	2.2	2.29	2.2	2.92	2.5	3.21	2.5	2	2.5	2	2.75	2.25	2.6	1.89	2.5	2.71
H-index	71	62	41	40	49	45	43	49	41	69	63	65	64	41	43	40	48	61	81	48

(a) Part a

	Q2	Q3	Q4
v1	41-50 years old	Female	Senior Lecturer/Associate Professor
v2	41-50 years old	Male	Professor
v3	41-50 years old	Male	Senior Researcher
v4	31-40 years old	Male	PhD Student/Research Assistant
v5	31-40 years old	Female	Lecturer/Assistant Professor
v6	< 31 years old	Male	Research Engineer/Programmer
v7	41-50 years old	Male	Senior Lecturer/Associate Professor
v8	< 31 years old	Female	Postdoctoral Researcher
v9	41-50 years old	Male	Professor
v10	31-40 years old	Male	Postdoctoral Researcher
v11	< 31 years old	Male	PhD Student/Research Assistant
v12	31-40 years old	Male	Postdoctoral Researcher
v13	31-40 years old	Male	Senior Lecturer/Associate Professor
v14	31-40 years old	Male	Professor
v15	31-40 years old	Male	Lecturer/Assistant Professor
v16	31-40 years old	Male	Lecturer/Assistant Professor
v17	31-40 years old	Male	PhD Student/Research Assistant
v18	31-40 years old	Male	Professor

(b) Part b

Figura A.9: Results of online survey conducted on ICWE conference.

VLDB

	Q1a Edmund M. Clarke	Q1b Thomas A. Henzinger	Q1c Patrick W. Patterson	Q1d Bruce Croft	Q1e Rakesh Agrawal	Q1f Alon Halevy	Q1g Stefano Ceri	Q1h Erol Gelenbe	Q1i Gaetano Borriello	Q1j Jose Joaquin Garcia-Luna-Aceves	Q1k Gerhard Tsudik	Q1l Ben Shneiderman	Q1m Amit Sheth	Q1n Ricardo Baeza-Yates	Q1o C. Lee Giles	Q1p Christos Faloutsos	Q1q David J. DeWitt	Q1r Thomas W. Reps
v1				3	4	4	4						3	4		3	4	
v2					3	4	3							4				
v3					3	3	3							3				
v4					3	3	3							3				
v5					3	3	3							3				
v6						2	2							3			3	
v7			3	3	4	4	4	4						4		4	4	
v8			4	2	3	4	3						3	3		2	3	
v9			4	3	3		3	2				4	3	3		3	4	
v10					2								1	3		2	3	
v11	1	1	1	1	3	3	3	1	1	1	1	1	1	1	2	1	1	1
v12			3		4	4	4						3	4			3	
v13			3		4	4	2								1	4	3	
v14	3		3		4	4	3				3			3	3	4	3	
Numbers of people who voted	2	1	7	5	12	13	13	3	1	1	2	2	6	13	3	8	10	1
Average rating	2	1	3	2,4	3,17	3,46	3,08	2,33	1	1	2	2,5	2,33	3,15	2	2,88	3,1	1
F-index	40	42	42	49	52	62	41	44	47	57	45	47	63	43	49	44	49	51

	Q1s Henry C.K. Korth	Q1t Erkki Ukkonen	Q1u Dirk Olle	Q1v David S. Johnson	Q1w Ahmed Elmagarmid	Q1x Daphne Koller	Q1y Dunesh Mensotra	Q1z Vern Paxson	Q1{ Geoffrey E. Hinton	Q1| Michael Stonebraker	Q1} Mirage Magazu	Q1~ Kim G. Larsen	Q1 David Oulles	Q1€ Philip S. Yu	Q1‚ Amir Pnueli	Q1ƒ Stanley Osher	Q1„ Jon Klenberg	Q1… Jim Gray	Q1† Bernhard Scholkopf	Q1‡ David Meyer	Q1ˆ S. Jensen	Q1‰ Christian Vanbat	Q1Š Amnir
v1										4				4			4	3			3		
v2										2				3			4			2			
v3										3					4		4					1	
v4										3					4		4					1	
v5										3					4		4					1	
v6										3					3		3			3			
v7					2					4			4		4		4			3		3	
v8			4		2					4			3				4			3		3	
v9		4			3					4		4					4			3			
v10					1					4		1	1				4			3		2	
v11	2	1	1	1	1		1		1	1	1	1	2	1	1	1	1		1	1		1	1
v12					3					3							4						
v13	2			4	4					4				3			3			3		4	
v14										3				4			3					4	
Numbers of people who voted	2	2	1	3	7	1	1	1	1	14	1	1	1	8	7	1	3	14	1	8	10	1	1
Average rating	2	2,5	1	3	2,29	1	1	1	1	3,21	1	1	1	3	3	1	2,67	3,57	1	2,63	2,3	1	1
F-index	40	47	42	40	43	49	55	43	71	49	40	46	70	74	54	52	41	45	71	43	45	51	

(a) Part a

	Q2	Q3	Q4
v1	41-50 years old	Male	Senior Lecturer/Associate Professor
v2	31-40 years old	Male	Lecturer/Assistant Professor
v3	41-50 years old	Male	Senior Lecturer/Associate Professor
v4	41-50 years old	Male	Senior Lecturer/Associate Professor
v5	41-50 years old	Male	Senior Lecturer/Associate Professor
v6	31-40 years old	Male	Postdoctoral Researcher
v7	51-60 years old	Male	Professor
v8	41-50 years old	Male	Professor
v9	more than 60 years old	Male	Professor
v10	51-60 years old	Male	Professor
v11	< 31 years old	Female	PhD Student/Research Assistant
v12	31-40 years old	Male	Lecturer/Assistant Professor
v13	31-40 years old	Male	Professor
v14	41-50 years old	Male	Professor

(b) Part b

Figura A.10: Results of online survey conducted on VLDB conference.

GENERAL

	Q1a Noga Alon	Q1b Sangur Sinhari	Q1c Ron Fedkiw	Q1d Sushil Jajodia	Q1e Anil K. Jain	Q1f Alex Nicolau	Q1g Xin Yao	Q1h Henri Prade	Q1i Joachim Weickert	Q1j Judea Pearl	Q1k Rajeev Alur	Q1l George Varghese	Q1m Andrew B. Kahng	Q1n Frank Pfenning	Q1o Joseph M. Hellerstein	Q1p Victor Basili	Q1q Brad Myers	Q1r Miron Livny	Q1s Baruch Awerbuch	Q1t Robert Tarjan	Q1u Andrew Blake	Q1v Luis Gravano
v1	1	1			1							1		1		1	1		1		1	1
v2																						
v3																						
v4															4							3
v5							3									2						
v6																						
v7					4																	
v8																	3					
v9																						
v10											4											
v11									3		4											
v12																						
v13																						
v14																						
v15																						
v16							3		3													
v17																						
v18																						
v19																						
v20	2	2	2	2	2	3	1	2	3	2	1	2	3	2	2	1	1	2	2	2	2	3
v21																	3					
v22																						
v23												3					3					
v24																						
v25																	3					
v26																						
v27										4						1						
v28																						
Numbers of people who voted	2	2	2	2	3	2	4	2	2	5	2	3	2	2	3	7	6	2	2	2	2	3
Average rating	1.5	1.5	1.5	1.5	2.33	2	2	1.5	2	2.6	1	2	2	1.5	2.33	2.43	2.33	1.5	1.5	1.5	1.5	2.33
H-index	49	40	45	65	86	44	40	84	41	47	49	40	50	46	51	44	49	42	51	82	57	41

	Q1w Edmund M. Clarke	Q1x Ricardo Baeza-Yates	Q1y David Culler	Q1z Mihalis Yannakakis	Q1{ Moshe Y. Vardi	Q1| Jeffrey Vitter	Q1} Didier Dubois	Q1~ Stanley Osher	Q1 David Gartan	Q1ˇ Adrian Perrig	Q1ˇ Joseph Goguen	Q1ˇ Kurt Keutzer	Q1ˇ John A. Stankovic	Q1ˇ Pierre Baldi	Q1ˇ Nancy Lynch	Q1ˇ Miron Livny	Q1ˇ Daniel Sewiorek	Q1ˇ Zhengyou Zhang	Q1ˇ Norman I. Badler
v1		4	1	1	2	1	1	1	1	1	1	1	1	3			1	1	1
v2																			
v3	4				4														
v4				4	4														
v5		4																	
v6	4								4										
v7																			
v8		3																	
v9																			
v10																			
v11	4				3						4								
v12																			
v13																			
v14																			
v15																			
v16																			
v17																			
v18																			
v19																			
v20	3	3	2	1	2	2	2	2	3	1	2	2	1	3	2	2	2	2	3
v21	4		4	2	2								4	3					
v22																			
v23																			
v24	4	3	4					2							4				3
v25		3												3					
v26																			
v27					4														
v28		4			4														
Numbers of people who voted	6	7	4	4	8	2	2	2	3	2	3	2	3	4	3	2	2	4	3
Average rating	3.83	3.43	2.76	2	3.13	1.5	1.5	1.5	2.67	1	2.33	1.5	2	3	3	2	1.5	2.75	2.33
H-index	40	43	70	48	72	55	41	52	56	56	43	46	46	48	40	40	49	40	46

(a) Part a

	Q2	Q3	Q4	Q5
v1	< 31 years old	Male	Postdoctoral Researcher	25
v2	51-60 years old	Male	Professor	1
v3	< 31 years old	Female	PhD Student/Research Assistant	22
v4	31-40 years old	Male	Lecturer/Assistant Professor	7
v5	< 31 years old	Male	PhD Student/Research Assistant	3
v6	31-40 years old	Male	Postdoctoral Researcher	
v7	< 31 years old	Male	Lecturer/Assistant Professor	
v8	< 31 years old	Male	PhD Student/Research Assistant	12
v9	31-40 years old	Male	Just fineshed MSC in computer Science	24
v10	31-40 years old	Male	PhD Student/Research Assistant	24
v11	< 31 years old	Male	Postdoctoral Researcher	3
v12	< 31 years old	Male	PhD Student/Research Assistant	24
v13	< 31 years old	Male	PhD Student/Research Assistant	24
v14	31-40 years old	Male	Lecturer/Assistant Professor	17
v15	31-40 years old	Male	Lecturer/Assistant Professor	7
v16	31-40 years old	Male	Professor	3
v17	< 31 years old	Female	Master Student	13
v18	< 31 years old	Female	Professor	16
v19	< 31 years old	Male	Postdoctoral Researcher	9
v20	31-40 years old	Male	Professor	17
v21	< 31 years old	Male	PhD Student/Research Assistant	20
v22	< 31 years old	Female	PhD Student/Research Assistant	1
v23	< 31 years old	Female	Senior Lecturer/Associate Professor	12
v24	41-50 years old	Female	Professor	9
v25	41-50 years old		Senior Lecturer/Associate Professor	12
v26	< 31 years old	Male	PhD Student/Research Assistant	13
v27	41-50 years old	Male	Senior Lecturer/Associate Professor	1
v28	41-50 years old	Male	Senior Researcher	13

(b) Part b

Figura A.11: Results of online survey conducted on GENERAL conference.

Appendice B

Algoritmo Diffcomm

In questo capitolo illustreremo brevemente le componenti principali dell'implementazione in Java dell'algoritmo "Diffcomm".

B.1 Le classi

Il progetto contiene i seguenti file:

- **diffcomm.Community.java:** Contiene il nome della comunità scientifica e l'arrayList con le opinioni dei voters che appartengono a questa comunità;
- **diffcomm.CommunityOpinion.java:** Contiene il nome della comunità scientifica e l'array con i voti ricavati facendo la media ponderata delle diverse opinioni dei voters appartenenti a questa comunità;
- **diffcomm.DiffCommAlgorithm.java:** É la classe principale che contiene il main e il cuore dell'algoritmo diffcomm;
- **diffcomm.Opinion.java:** Contiene l'identificativo del soggetto che vota, l'id del candidato su cui ci stiamo esprimendo e il voto espresso;
- **diffcomm.Voter.java:** Contiene l'id, il nome del voter, la lista delle comunità scientifiche e del numero di pubblicazioni del voter in quella determinata comunità.

B.2 I file di input

I file che contengono i dati che verranno presi in input dal nostro algoritmo sono in formato CSV e sono:

- ***Candidates.csv:*** (CANDIDATE, NAME)
  Contiene la lista dei candidati nella forma **candidate id, "candidate name"**. É possibile vedere un esempio nella tabella B.1.

CSV
"CANDIDATE","NAME"
1,"Mario Rossi"
2,"Roberto Bianchi"

Tabella B.1: Candidates.csv

- ***Communities.csv:*** (COMMUNITIES, NAME)
  Contiene la lista delle comunitá nella forma **"community id","community name"**. É possibile vedere un esempio nella tabella B.2.

- ***Voter.csv:*** (VOTER, NAME)
  Contiene la lista dei votanti nella forma **voter id,"voter name"**. É possibile vedere un esempio nella tabella B.3.

- ***VoterMembership.csv:*** (VOTER, COMMUNITY, NUMBERS_OF_PUBBL)
  Contiene la lista delle relazioni voter-community nella forma **voter id,"community id",numbers of pubblications**. É possibile vedere un esempio nella tabella B.4.

- ***Opinions.csv:*** (VOTER, CANDIDATE, VOTE)
  Contiene la lista delle opinioni nella forma **voter id,candidate id,vote**. É possibile vedere un esempio nella tabella B.5.

CSV
“COMMUNITIES”,”NAME”
“a”,”[edoc]”
“b”,”[sat]”
“c”,pippo

Tabella B.2: Communities.csv

CSV
“VOTER”,”NAME”
1,”Martina Verdi”
2,”Voter2”
3,”Voter3”

Tabella B.3: Voter.csv

CSV
“VOTER”,”COMMUNITY”,”NUMBERS_OF_PUBBL”
1,“a”,2
2,“a”,2
2,“b”,4
3,“c”,1

Tabella B.4: VoterMembership.csv

CSV
“VOTER”,”CANDIDATE”,”VOTE”
1,1,2
1,2,4
2,1,0
2,2,1
3,1,2
3,2,3

Tabella B.5: Opinions.csv

Bibliografia

[1] BACCINI A. *Valutare la ricerca scientifica. Uso e abuso degli indicatori bibliometrici.*, Il Mulino, Bologna, 2010.

[2] CARTERETTE B. *On Rank Correlation and the Distance Between Rankings*, SIGIR 09.

[3] CONOVER W.J. *Practical Non-Parametric Statistics*, John Wiley and Sons, New York, second edition, 1980.

[4] DE ROBBIO A. *Analisi citazionale e indicatori bibliometrici nel modello Open Access.*, Bollettino AIB, XLVII, 3, pp. 257?287, 2007. Website: http://eprints.rclis.org/11999/.

[5] DE SOLLA PRICE, D. J. *Little science, big science.*, Columbia University Press, New York, 1963.

[6] DE SOLLA PRICE, D. J. *Networks of scientific papers. The pattern of bibliographic references indicates the nature of the scientific front.*, Science, 149 (3683): 510-515, 1963.

[7] (ED. JOSE HARRIS), F. T. *Community and Civil Society.*, Cambridge University Press, 2001.

[8] GARFIELD E. *Citation Indexes for Science: a new dimension in documentation through association of ideas.*, Science, CXXII, pp. 108–111, 1955. Website: http://garfield.library.upenn.edu/essays/v6p468y1983.pdf.

[9] HIRSCH, J. E., *An index to quantify an individual's scientific research output.*, Proceedings of the National Academy of Science, 102, 16569, 2005, Website: http://arxiv.org/abs/physics/0508025v5.

[10] KENDALL M. *Rank Correlation Methods*, Griffin, London, UK, fourth edition, 1970.

[11] LIQUIDPUB PROJECT. *Scientific Community.* Website: http://project.liquidpub.org/research-areas/scientific-community.

[12] MAKINO JUNICHIRO. *Productivity of Research Groups - Relation between Citation Analysis and Reputation within Research Communities*, 1998.

[13] MUSSI A. *Discovering and analyzing scientific communities using conference network.* Universidad Católica Nuestra Segñora de la Asunción-Paraguay, 2010.

[14] NOETHER G.E. *Why Kendall Tau?* Website: http://rsscse.org.uk/ts/bts/noether/text.html.

[15] PARRA C. *Infraestructura web para comunidades y recursos cientificos.* Modelo, servicios y métricas. Universidad Nacional de Asunción-Paraguay, 2009.

[16] PARRA C., CASATI F., CERNUZZI L. *Modelling, Managing and Analyzing Scientific Communities* Website: https://dev.liquidpub.org/svn/liquidpub/communities/pub/Communities-ResearchProposal.pdf.

[17] PIAZZINI T. *Bibliometric indicators: spread reflections for a focusing and conscious use.* JLIS.it, North America, 1, apr. 2010. Website: http://leo.cilea.it/index.php/jlis/article/view/24.

[18] R. *The R project for statistical computing.* Website: http://www.r-project.org.

[19] SISTEMA BIBLIOTECARIO DI ATENEO (SBA) *Indicatori Bibliometrici*, Universitá degli studi di Modena e Reggio Emilia. Website:

http://www.sba.unimore.it/Inside/Biblioteca%20Digitale/Indicatori %20bibliometrici.asp.

[20] SMITH M. K. *Community. a review of the theory.* Website: http://www.infed.org/community/community.htm, 2001.

[21] WIKIMEDIA FOUNDATION. *Wikipedia, the free encyclopedia that anyone can edit.* Website: http://www.wikipedia.org.