



UNIVERSITÀ DEGLI STUDI DI SALERNO

Dipartimento di Informatica

Corso di Laurea Magistrale in Informatica

TESI DI LAUREA MAGISTRALE IN INFORMATICA

Impatto delle dispersioni culturali sull'emergere di Community Smell: confronto tra diversi modelli culturali

RELATORE

Prof. Fabio Palomba

Dott. Stefano Lambiase

Università degli Studi di Salerno

CANDIDATO

Saverio De Stefano

Matricola: 0522501109

Questa tesi è stata realizzata nel



*La conoscenza e il sapere sono le uniche risorse che potete utilizzare e commerciare sempre, e
che nessuno potrà mai togliervi.*

Abstract

L'Ingegneria del Software negli ultimi anni ha subito enormi cambiamenti grazie allo sviluppo di tecnologie e infrastrutture di comunicazione che hanno permesso di mettere in contatto persone provenienti da diverse parti del mondo. Tuttavia, mettere in contatto persone con differenti culture risulta un'attività molto complessa per i manager, in quanto, se non gestite, potrebbero portare al fallimento del progetto in corso. Le problematiche che potrebbero insorgere possono essere individuate attraverso fenomeni noti come "Community Smell". Nel contesto di questa tesi sono state confrontate le dimensioni di tre framework culturali come Globe, Hofstede e Trompenaars. Lo scopo è quello di fornire ai manager nuovi strumenti decisionali per formare al meglio il team di sviluppo, cercando di comprendere quanto i team open source di GitHub siano culturalmente dispersi e se esiste una correlazione tra il verificarsi di Community smell (Black Cloud, Lone wolf, Organisation Silo, Radio Silence) e la presenza di differenze culturali, quest'ultima è stata verificata applicando un modello di regressione lineare misto. I risultati ottenuti affermano che le comunità software open-source non risultano culturalmente disperse e che le differenze culturali influiscono nel verificarsi di community smell, principalmente influenzano 'lone wolf' e 'radio silence'.

Indice

Elenco delle Figure	iv
Elenco delle Tabelle	v
1 Introduzione	1
1.1 Contesto Applicativo	1
1.2 Motivazioni e Obiettivi	2
1.3 Risultati ottenuti	2
2 Background e stato dell'arte	5
2.1 Global Software Engineering	5
2.2 Community Smell e Social Debt	6
2.3 Cultura nella Software Engineering	7
2.3.1 Hofstede	7
2.3.2 Globe	8
2.3.3 Trompenaars	10
3 Metodologia di ricerca	13
3.1 Ipotesi e domande di ricerca	13
3.2 Contesto di studio	14
3.3 RQ₁ - Dispersione culturale	15

3.4	RQ₂ - Costruzione del modello statistico	18
3.4.1	Variabili Indipendenti	18
3.4.2	Variabile di Risposta	19
3.4.3	Variabili di controllo	20
3.4.4	Verifica delle assunzioni	21
3.4.5	Costruzione del modello di regressione lineare	22
4	Analisi dei dati e risultati	23
4.1	RQ₁ - Dispersione culturale	23
4.2	RQ₂ - Analisi della dispersione culturale per Community Smell	27
4.3	RQ₂ - Analisi dei risultati del modello statistico	31
4.4	Risultati Black Cloud	31
4.4.1	Globe	31
4.4.2	Hofstede	33
4.4.3	Trompenaars	33
4.5	Risultati Lone wolf	35
4.5.1	Globe	35
4.5.2	Hofstede	36
4.5.3	Trompenaars	36
4.6	Risultati Organisation Silo	39
4.6.1	Globe	39
4.6.2	Hofstede	40
4.6.3	Trompenaars	40
4.7	Risultati Radio Silence	44
4.7.1	Globe	44
4.7.2	Hofstede	44
4.7.3	Trompenaars	45
5	Threats to validity	48
5.1	Threats to Construct Validity	48
5.2	Threats to Conclusion Validity	49
5.3	Threats to External Validity	49

6 Conclusioni	50
Bibliografia	53

Elenco delle figure

3.1	Panoramica sul processo di ricerca.	15
3.2	Panoramica sul processo di computazione.	17
4.1	Distribuzione della dispersione culturale misurata con globe.	24
4.2	Distribuzione della dispersione culturale misurata con hofstede.	25
4.3	Distribuzione della dispersione culturale misurata con trompenaars.	26
4.4	Dispersione Globe BC	27
4.5	Dispersione hofstede BC	27
4.6	Dispersione Trompenaars BC	27
4.7	Dispersione Globe LW	28
4.8	Dispersione hofstede LW	28
4.9	Dispersione Trompenaars LW	28
4.10	Dispersione Globe OS	29
4.11	Dispersione hofstede OS	29
4.12	Dispersione Trompenaars OS	29
4.13	Dispersione Globe RS	30
4.14	Dispersione hofstede RS	30
4.15	Dispersione Trompenaars RS	30

Elenco delle tabelle

4.1	Dimensioni e coefficienti di variazione per globe	24
4.2	Dimensioni e coefficienti di variazione per hofstede	25
4.3	Dimensioni e coefficienti di variazione per Trompenaars	26
4.4	Community smell	31
4.5	Tabella riassuntiva dei risultati del modello Black Cloud per il dataset Globe	32
4.6	Tabella riassuntiva dei risultati del modello Black Cloud per il dataset Hofstede	34
4.7	Tabella riassuntiva dei risultati del modello Black Cloud per il dataset Trompenaars	35
4.8	Tabella riassuntiva dei risultati del modello Lone Wolf per il dataset Globe	37
4.9	Tabella riassuntiva dei risultati del modello Lone Wolf per il dataset Hofstede	38
4.10	Tabella riassuntiva dei risultati del modello Lone Wolf per il dataset Trompenaars	39
4.11	Tabella riassuntiva dei risultati del modello Organisation Silo per il dataset Globe	41

4.12 Tabella riassuntiva dei risultati del modello Organisation Silo per il dataset Hofstede	42
4.13 Tabella riassuntiva dei risultati del modello Organisation Silo per il dataset Trompenaars	43
4.14 Tabella riassuntiva dei risultati del modello Radio Silence per il dataset Globe	45
4.15 Tabella riassuntiva dei risultati del modello Radio Silence per il dataset Hofstede	46
4.16 Tabella riassuntiva dei risultati del modello Radio Silence per il dataset Trompenaars	47

CAPITOLO 1

Introduzione

1.1 Contesto Applicativo

Questo lavoro si colloca nel contesto dell'Ingegneria del Software, che, negli ultimi anni ha subito enormi cambiamenti grazie allo sviluppo di tecnologie e infrastrutture di comunicazione che hanno permesso di mettere in contatto persone provenienti da diverse parti del mondo. Questo fenomeno è noto come Global Software Engineering (GSE) ed è in rapida diffusione, anche grazie alla possibilità di accedere a forza lavoro a costi ridotti in varie parti del mondo. Tuttavia, la GSE presenta diverse problematiche dovute alle differenze nelle distanze temporali, culturali e geografiche delle persone che compongono il team di sviluppo. Se non gestite correttamente, queste problematiche rischiano di portare al fallimento del progetto in corso. I progetti software sono delle attività ad alto contenuto umano e, in quanto tali, possono presentare i problemi che accomunano gli esseri umani; una possibile causa di questo fenomeno potrebbero essere le differenze culturali che le varie popolazioni del mondo condividono. Le problematiche che potrebbero insorgere possono essere individuate attraverso fenomeni noti come "Community Smell", i quali, per definizione, "rappresentano caratteristiche organizzative e socio-tecniche non ottimali all'interno di una comunità software, che potrebbero portare a costi

aggiuntivi". Uno studio condotto da Lambiase et al. [1] ha misurato le differenze culturali dei progetti software open source tramite il modello a 6-D di Hofstede. Tale studio ha dimostrato sia quanto le comunità open-source siano disperse, sia quanto queste differenze culturali possano influire sui community smell.

1.2 Motivazioni e Obiettivi

Questo lavoro ha come obiettivo quello di estendere il lavoro di Lambiase et al. [1] precedentemente descritto. Nel contesto di questa tesi le dimensioni di Hofstede saranno confrontate con quelle di Globe e Trompenaars. Lo scopo è quello di fornire ai Project Manager nuovi strumenti decisionali in modo tale da poter formare al meglio i team di sviluppo. Per raggiungere l'obiettivo di ricerca si è deciso di dividere il lavoro in due RQ:

- **RQ₁** - In che misura le comunità open source sono culturalmente disperse?
- **RQ₂** - In che misura la dispersione culturale all'interno dei team influenza il numero di community smell?

La prima RQ mira a fornire una visione della dispersione culturale delle comunità open source presenti su GitHub tramite la computazione delle metriche culturali fornite dai modelli a **6-D Hofstede, Globe, Trompenaars**.

La seconda RQ mira a trovare una possibile correlazione tra il verificarsi di Community smell e la presenza di differenze culturali, quest'ultima sarà verificata applicando un modello di regressione lineare misto per ogni dataset di community smell e utilizzando ciascun framework culturale.

1.3 Risultati ottenuti

Per rispondere alla **RQ₁**, è stato utilizzato un dataset rappresentante le comunità di GitHub, su cui sono state successivamente calcolate le dispersioni culturali utilizzando i tre framework in esame: Globe, Hofstede e Trompenaars. Successivamente, tali dispersioni sono state rappresentate graficamente tramite dei violin plot,

evidenziando che le comunità open source prese in esame non presentano dispersioni eccessive, indipendentemente dal modello di misurazione della cultura (Globe, Hofstede, Trompenaars).

Per quanto concerne la **RQ₂**, sono stati utilizzati quattro dataset creati da Catolino et al. [2], relativi a quattro "community smell" considerati nello studio (Black Cloud, Lone wolf, Organisation Silo, Radio Silence). Inizialmente, sono state mostrate le dispersioni tramite dei violin plot misurati con i diversi modelli culturali. Successivamente, per comprendere meglio la correlazione tra la dispersione e l'emergere di "community smell", è stato costruito un modello di regressione lineare misto [1], utilizzando come variabili di controllo tutte le variabili che Catolino et al. [2] avevano identificato come influenti sull'emergere di "community smell".

È stato osservato che alcuni "community smell", come "radio silence" e "lone wolf", sono correlati sia in modo positivo che negativo all'aumentare della dispersione di determinate caratteristiche culturali. A seconda del modello utilizzato, sono emersi risultati leggermente diversi:

- **Globe:** Si è notato che avere personale con idee diverse riguardo alla dimensione 'Collectivism II Societal Practices (In-group Collectivism)' potrebbe portare a una riduzione delle istanze di 'lone wolf', mentre le dimensioni 'Future Orientation Societal Practices', 'Collectivism I Societal Practices (Institutional Collectivism)', 'Performance Orientation Societal Practices' e 'Assertiveness Societal Practices' sono rilevanti nella variazione del numero di istanze di 'radio silence'.
- **Hofstede:** Si è visto che differenze maggiori in termini di 'power distance' e 'individualism' possono ridurre le istanze di 'lone wolf', mentre maggiori differenze in termini di 'indulgence' potrebbero portare ad incrementi di fenomeni di "lone wolf". Inoltre, differenze maggiori in termini di 'individualismo' potrebbero portare a diminuzioni di istanze di "radio silence".
- **Trompenaars:** Si è notato che avere diverse idee sulla dimensione 'sequential/syncronic', ovvero, differenze in termini di multi-tasking e propensione a lavori sequenziali, potrebbe portare a una riduzione delle istanze di 'lone wolf'. Inoltre, sebbene con una bassa significatività, maggiori differenze sulle

dimensioni 'Neutral/Affective' e 'Internal/External' potrebbero portare ad aumenti di 'radio silence'.

CAPITOLO 2

Background e stato dell'arte

In questo capitolo andremo ad analizzare la ricerca scientifica rispetto alla cultura nella Software Engineering, ai community smell e alla Global Software Engineering.

2.1 Global Software Engineering

Negli ultimi anni la Global Software Engineering (GSE) è diventata una pratica molto diffusa per diverse ragioni, tra cui lo sviluppo di internet e di infrastrutture di telecomunicazioni internazionali a basso costo, oppure per la disponibilità di ingegneri del software in località a basso costo (Europa dell'est, America latina e estremo Oriente). In altri casi le organizzazioni hanno delocalizzato le loro sedi in economie emergenti dove il costo del lavoro risulta inferiore [3]. Nel mondo odierno quasi tutti i prodotti software sono sviluppati in maniera distribuita a livello globale, purtroppo molti progetti non riescono a raggiungere i risultati prefissati e vengono annullati [4, 3]. La ricerca fatta da Richardson et al. [3] ha dimostrato che molte aziende non riescono ad attuare la GSE per diverse ragioni, come la distanza temporale, culturale e geografica. Ad esempio, le difficoltà culturali, di gestione dei progetti e di comunicazione causano continuamente problemi agli ingegneri del software e ai project manager.

2.2 Community Smell e Social Debt

È facile pensare allo sviluppo software come un'attività puramente tecnica in quanto nello sviluppo sono coinvolte persone con elevate capacità tecniche, ed è proprio per la natura socio-tecnica nel processo che è necessario considerare anche l'aspetto umano oltre che quello tecnico. Infatti, l'essere lo sviluppo software un'attività sociale fa sì che possano emergere Community Smell e Social Debt [5]. Il Social Debt nell'ingegneria del software rappresenta l'emergere di un costo imprevisto connesso ad una certa comunità di sviluppo (come ad esempio l'abbandono di un membro). Anche il debito sociale, come il debito tecnico, può avere un impatto importante sulla buona riuscita di un progetto software [6]. I community smell "rappresentano caratteristiche organizzative e socio-tecniche non ottimali all'interno di una comunità software che potrebbero portare a costi aggiuntivi" [7]. Quelli presi in considerazione all'interno di questo lavoro di tesi e da altra letteratura sul tema [7, 2, 1]:

- **Black Cloud:** questo Community Smell riflette un sovraccarico di informazioni dovuto alla mancanza di comunicazioni strutturate o di una governance di cooperazione.
- **Lone Wolf:** questo community Smell si verifica quando un determinato soggetto nella comunità software sfiduciato e avendo una bassa considerazione dei propri colleghi, porta avanti il suo lavoro in solitaria senza fornire aggiornamenti.
- **Organisation Silo:** questa forma di debito sociale si riferisce alla presenza di aree della comunità separate che non comunicano, se non attraverso uno o due dei loro rispettivi membri;
- **Radio Silence:** questa è un'istanza del problema del "unique boundary spanner" nell'analisi delle reti sociali: un membro si inserisce in ogni interazione formale tra due o più sub-comunità con scarsa o nessuna flessibilità per introdurre altri canali paralleli.

2.3 Cultura nella Software Engineering

Come detto in precedenza la Software Engineering è un'attività principalmente umana, e in quanto tale soffre, dei problemi a cui sono sottoposti gli esseri umani. Uno dei problemi analizzati in precedenza è la differenza culturale, nel tempo sono stati sviluppati dei framework che permettono di misurare queste differenze. Uno dei principali lavori in questo ambito è stato svolto da Lambiase et al. [1] dove ci si è chiesti se le differenze culturali misurate con le 6-D di Hofstede potessero in qualche modo impattare sull'insorgere di community smell; ciò che è emerso è che la dispersione culturale è un fattore rilevante per la nascita di Radio Silence e Lone Wolf, mentre la Dispersione Geografica è rilevante per tutti gli smell, tranne che per Lone Wolf.

2.3.1 Hofstede

Secondo Hofstede la cultura è *"La programmazione della mente umana con cui un gruppo di persone si distingue da un altro gruppo"*. La cultura è composta da vari strati e può essere paragonata ad una cipolla. Gli strati esterni della cipolla sono costituiti da simboli come cibo, loghi, colori o monumenti. Il livello successivo è costituito da eroi e può includere personaggi pubblici della vita reale, come statisti, atleti o fondatori di aziende, o figure come Superman nella cultura popolare. Sul terzo strato, più vicino al nucleo, si trovano i rituali, come sauna, karaoke o riunioni. [8] Hofstede consiste in 6 dimensioni numeriche che permettono di modellare gli aspetti culturali di un paese e dei suoi abitanti.

1. **Power Distance (PDI):** questa dimensione esprime il grado in cui i membri meno potenti di una società accettano e si aspettano che il potere sia distribuito in modo diseguale [9].
2. **Individualism (IDV):** rappresenta il grado in cui le persone di una società sono integrate in gruppi. Un livello elevato di questa dimensione indica una società in cui ci si aspetta che gli individui si prendano cura solo di se stessi e dei propri familiari. Al contrario, un livello basso indica una preferenza per un contesto sociale molto unito, in cui gli individui possono aspettarsi che i propri

parenti o i membri di un particolare gruppo si prendano cura di loro in cambio di una fedeltà indiscussa [1].

3. **Masculinity vs Femininity (MAS):** il lato mascolinità di questa dimensione rappresenta una preferenza nella società per il successo, l'eroismo, l'assertività e le ricompense materiali per il successo. La società in generale è più competitiva. Il suo opposto, la femminilità, rappresenta una preferenza per la cooperazione, la modestia, la cura per i deboli e la qualità della vita. La società in generale è più orientata al consenso [9].
4. **Uncertainty Avoidance (UAI):** esprime il grado in cui i membri di una società si sentono a disagio con l'incertezza e l'ambiguità. La questione fondamentale qui è come una società affronta il fatto che il futuro non può mai essere conosciuto: dovremmo cercare di controllare il futuro o semplicemente lasciare che accada? [9]
5. **Long Term Orientation (LTO):** le società che ottengono un punteggio basso su questa dimensione, ad esempio, preferiscono mantenere tradizioni e norme consolidate nel tempo, mentre guardano con sospetto al cambiamento sociale. Quelli con una cultura che ha un punteggio elevato, d'altra parte, adottano un approccio più pragmatico: incoraggiano la parsimonia e gli sforzi nell'educazione moderna come un modo per prepararsi per il futuro. [9]
6. **Indulgence (IVR):** si riferisce al grado di libertà che le norme sociali concedono ai cittadini per soddisfare i loro desideri umani. Un livello alto indica una società che permette una gratificazione libera dei desideri umani fondamentali e naturali legati al godimento della vita e al divertimento. Al contrario, un livello basso indica una società che controlla la gratificazione dei bisogni e la regolamenta utilizzando norme sociali rigorose [9].

2.3.2 Globe

Un secondo importante framework culturale, il progetto Global Leadership and Organizational Behavior Effectiveness (GLOBE), fornisce ai manager un'ulteriore lente attraverso la quale comprendere meglio come ottenere buoni risultati in un

ambiente internazionale. Mentre il framework di Hofstede è stato sviluppato negli anni '60, il progetto GLOBE, sviluppato negli anni '90, è un tentativo più recente di comprendere le dimensioni culturali. Il progetto GLOBE coinvolge 170 ricercatori di oltre 60 Paesi che hanno raccolto dati su 17.000 manager di 62 Paesi del mondo. Analogamente a Hofstede, i ricercatori del GLOBE hanno individuato nove dimensioni culturali. Tuttavia, basando il loro lavoro sulle dimensioni culturali di Hofstede, infatti 5 delle 9 dimensioni sono simili a quelle individuate da Hofstede. Il progetto Globe per facilitare il lavoro dei manager che gestiscono persone di diversi paesi ha identificato dei cluster con caratteristiche culturali simili [10]. Di seguito vediamo le nove dimensioni del framework Globe:

1. **Performance Orientation:** il grado in cui un collettivo incoraggia e premia (e dovrebbe incoraggiare e premiare) i membri del gruppo per il miglioramento delle prestazioni e l'eccellenza [11].
2. **Assertiveness:** il grado in cui gli individui sono (e dovrebbero essere) assertivi, conflittuali e aggressivi nella loro relazione con gli altri [11].
3. **Future Orientation:** la misura in cui gli individui si impegnano (e dovrebbero impegnarsi) in comportamenti orientati al futuro come la pianificazione, l'investimento nel futuro e il ritardo della gratificazione [11].
4. **Humane Orientation:** il grado in cui un collettivo incoraggia e premia (e dovrebbe incoraggiare e premiare) gli individui per essere giusti, altruisti, generosi, premurosi e gentili con gli altri [11].
5. **Institutional Collectivism:** il grado in cui le pratiche istituzionali organizzative e sociali incoraggiano e premiano (e dovrebbero incoraggiare e premiare) la distribuzione collettiva delle risorse e l'azione collettiva [11].
6. **In-Group Collectivism:** il grado in cui gli individui esprimono (e dovrebbero esprimere) orgoglio, lealtà e coesione nelle loro organizzazioni o famiglie [11].
7. **Gender Egalitarianism:** il grado in cui un collettivo minimizza (e dovrebbe minimizzare) la disuguaglianza di genere [11].

8. **Power Distance:** la misura in cui la comunità accetta e approva l'autorità, le differenze di potere e i privilegi di status [11].
9. **Uncertainty Avoidance:** la misura in cui una società, un'organizzazione o un gruppo si basa (e dovrebbe fare affidamento) su norme, regole e procedure sociali per alleviare l'imprevedibilità degli eventi futuri. Maggiore è il desiderio di evitare l'incertezza, più le persone cercano ordine, coerenza, struttura, procedure formali e leggi per coprire le situazioni della loro vita quotidiana [11].

I ricercatori di GLOBE hanno identificato le dimensioni culturali universalmente desiderabili come "*Performance Orientation*" e le dimensioni universalmente non desiderabili come la "*Power Distance*" [11].

2.3.3 Trompenaars

Questo modello è stato sviluppato da Fons Trompenaars e Charles Hampden-Turner andando a sviluppare un sondaggio su 8.841 manager e coinvolgendo 43 paesi. Questo framework culturale ha sette dimensioni. Ci sono cinque orientamenti che coprono i modi in cui gli esseri umani si relazionano l'uno con l'altro, uno che ha a che fare con il tempo e uno che si occupa dell'ambiente [12].

1. Universalism vs particularism:

- Universalismo: culture che considerano le leggi, le regole e gli obblighi più importanti delle relazioni.
- Particularismo: queste culture credono che le relazioni determinino le regole con cui vivono e che ogni situazione possa richiedere una risposta diversa. [13]

2. Individualism vs communitarianism:

- Individualismo: dove la libertà e la realizzazione personale sono celebrate, premiate e collegate ai bisogni degli altri o dell'organizzazione nel suo complesso. I dipendenti sono autonomi, creativi e autorizzati a commettere errori.

- Comunitarismo: la convinzione che il gruppo sia più importante dell'individuo, fornendo sicurezza e assistenza in cambio di lealtà. Gli individui non sono elogiati pubblicamente e non viene mostrato alcun favoritismo. [13]

3. Neutral vs emotional:

- Neutrale: i dipendenti sopprimono le loro emozioni e tendono a non rivelare ciò che stanno pensando o sentendo.
- Emotivo: descrive i dipendenti che sono più disposti a esprimere emozioni sul lavoro in modo spontaneo. Le emozioni costruiscono fiducia e rapporto e possono anche essere utilizzate per gestire il conflitto prima che si intensifichi. [13]

4. Specific vs diffuse:

- Specifico: questi dipendenti mantengono separate le loro vite personali e professionali e credono che i colleghi possano lavorare insieme senza necessariamente piacersi.
- Diffuso: i dipendenti sovrappongono la vita personale con quella professionale [13].

5. Achievement vs ascription:

- Successo: vengono premiate le prestazioni indipendentemente dall'anzianità. Il valore di una persona è correlato a cosa sa e cosa fa rispetto ai titoli che possiede.
- Ascrizione: le persone vengono valutate in base al potere o al titolo, chi ha l'autorità ha il massimo rispetto. [13]

6. Sequential vs synchronic:

- Tempo sequenziale: i dipendenti apprezzano che gli eventi si verifichino in ordine logico, considerando: pianificazione, impegno e puntualità.

- **Tempo singolo:** i dipendenti considerano il passato, il presente e il futuro come intrecciati. Sono felici di lavorare su più progetti contemporaneamente e non vedono i piani come rigidi o immutabili. [13]

7. Internal vs external control:

- **Interna:** coloro che hanno una direzione interna credono di avere il controllo sul proprio ambiente e su come lavorano con altri individui e team.
- **Esterna:** quelli con direzione esterna credono che il loro ambiente li controlli. Le loro azioni ruotano attorno all'elusione o alla minimizzazione dei conflitti. [13]

CAPITOLO 3

Metodologia di ricerca

In questo capitolo si mostrano le domande di ricerca e le metodologie di ricerca utilizzate per dare risposta alle domande. L'obiettivo della tesi si rifà al lavoro svolto da Lambiase et al. [1], in cui i ricercatori si sono chiesti quanto le differenze culturali potessero impattare nell'emergere di community smell; nel caso della ricerca di Lambiase et al. [1] è stato utilizzato esclusivamente il framework **Hofstede 6-D**. Nel contesto di questa tesi le dimensioni di Hofstede saranno confrontate con quelle di **Globe e Trompenaars**. Lo scopo è quello di fornire ai Project Manager nuovi strumenti decisionali in modo tale da poter formare al meglio i team di sviluppo.

3.1 Ipotesi e domande di ricerca

L'ipotesi alla base del lavoro è la seguente:

La presenza di differenze culturali all'interno del team di sviluppo è correlata alla presenza di community smell.

Per raggiungere l'obiettivo di ricerca si è deciso di dividere la ricerca in due RQ:

- **RQ₁** - In che misura le comunità open source sono culturalmente disperse?

- **RQ₂** - In che misura la dispersione culturale all'interno dei team influenza il numero di community smell?

La prima RQ mira a fornire una visione della dispersione culturale delle comunità open source presenti su GitHub tramite la computazione delle metriche culturali fornite dai modelli a **6-D Hofstede, Globe, Trompenaars**.

La seconda RQ mira a trovare una possibile correlazione tra il verificarsi di Community smell e la presenza di differenze culturali, quest'ultima sarà verificata applicando un modello di regressione lienare per ogni dataset di community smell e utilizzando ciascun framework culturale.

3.2 Contesto di studio

Il contesto dello studio consiste dal dataset delle comunità open source creato da Vasilescu et al. [14] composto da 93.057 record, dalle dimensioni dei diversi framework culturali, dove per globe e Hofstede sono disponibili dei dataset online mentre per Trompenaars è stato necessario costruire un nuovo dataset partendo dalle informazioni presenti sul sito [15]. Inoltre si sono utilizzati i dataset contententi le informazioni riguardanti i community smell di Catolino et al. [2], nella figura 3.1 si mostrano le operazioni effettuate per raggiungere l'obiettivo di ricerca.

In risposta alla prima domanda di ricerca il dataset di Vasilsecu et al. [14] è stato filtrato in quanto il campo 'countries' conteneva un gran numero di valori 'none', questo perché l'informazione relativa alla nazionalità su GitHub è qualcosa di non obbligatorio, quindi si è rivelato necessario filtrare per un certo numero massimo di valori 'none', si è scelto quindi di non superare il 25%, l'operazione ha restituito un dataset contenente 30.451 record. Successivamente sono state computate le metriche culturali per ogni framework a disposizione, dando vita a 3 nuovi dataset aventi l'informazione della distanza culturale ed è stato poi necessario normalizzare le distanze culturali dato che ogni framework esprimeva la cultura con grandezze numeriche diverse tra loro, per far ciò è stato utilizzato il coefficiente di variazione.

$$CV = \frac{\sigma}{\mu} = \frac{\text{Deviazione standard}}{\text{Media}}$$

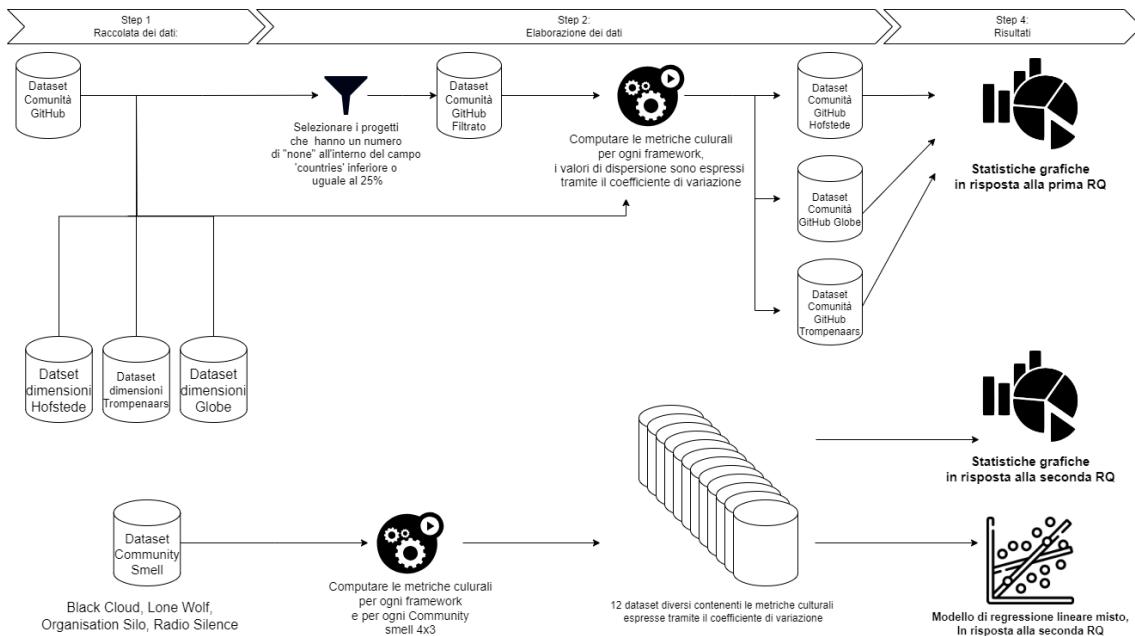


Figura 3.1: Panoramica sul processo di ricerca.

Successivamente sono stati eliminati i progetti che avevano solo un membro, poiché avrebbero causato una dispersione pari a 0. Infine sono stati sviluppati dei violin plot per mostrare graficamente come le comunità open source erano culturalmente disperse. Per quanto riguarda la seconda RQ si sono computate le metriche culturali per ogni dataset dei community smell e per ogni framework culturale dando origine a 12 nuovi dataset. Infine su questi dati sono stati applicati dei modelli di regressione lineare mista per verificare possibili correlazioni tra la dispersione culturale ed i community smell.

3.3 RQ₁ - Dispersione culturale

Q RQ₁. In che misura le comunità open source sono culturalmente disperse?

Per affrontare la RQ₁ si è partiti dal dataset delle comunità GitHub creato da Vasilescu et al. [14] filtrato rispetto ai valori 'none' del campo 'countries' e successivamente sono stati utilizzati i dataset, precedentemente trasformati in json, contenenti le dimensioni di ogni framework.

I file json dei framework culturali contengono le seguenti informazioni:

- **Globe:** in questo caso si è deciso di effettuare un mapping in modo da semplificare la lettura dei file risultanti

- "Uncertainty Avoidance Societal Practices": A1
- "Future Orientation Societal Practices": A2
- "Power Distance Societal Practices": A3
- "Collectivism I Societal Practices (Institutional Collectivism)": A4
- "Humane Orientation Societal Practices": A5
- "Performance Orientation Societal Practices": A6
- "Collectivism II Societal Practices (In-group Collectivism)": A7
- "Gender Egalitarianism Societal Practices": A8
- "Assertiveness Societal Practices": A9

- **Hofstede:**

- pdi
- idv
- mas
- uai
- ltows
- ivr

- **Trompenaars:**

- Universalism/Particularism
- Individualism/Communitarianism
- Specific/Diffuse
- Neutral/Affective
- Achievement/Ascription
- Past,Present,Future

- Sequential/Synchronic
- Internal/External

Grazie ai dati precedentemente descritti è stato possibile computare la dispersione culturale andando a mappare le nazionalità di ogni sviluppatore presenti nel campo 'countries' con il relativo valore della dimensione culturale, in figura 3.2 è possibile vedere il dettaglio delle operazioni.

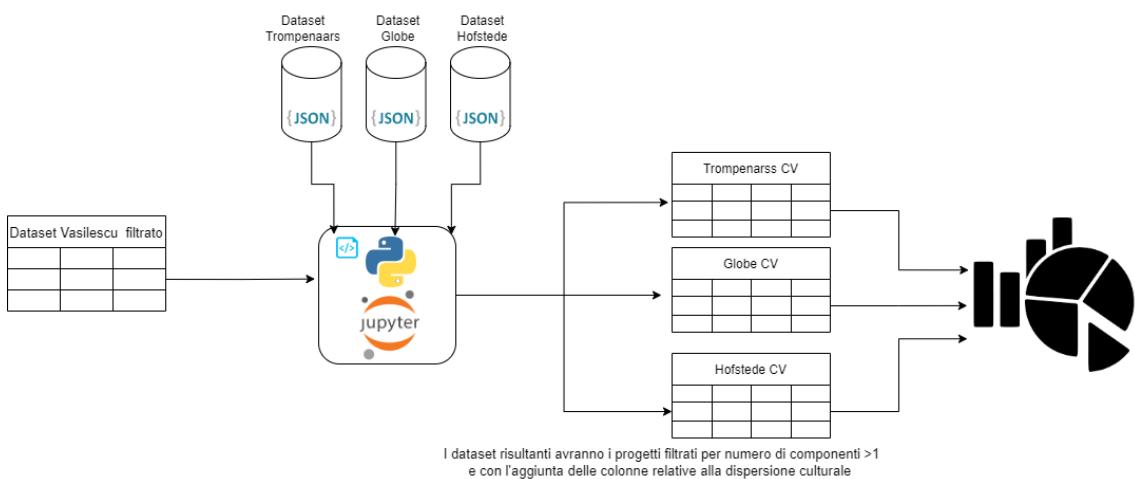


Figura 3.2: Panoramica sul processo di computazione.

I tre dataset risultanti sono quindi composti da nuove colonne che rappresentano le metriche di dispersione culturale espresse con diverse misure, di seguito la descrizione dei dataset:

- **trompenaars_metrics.csv**: dopo la computazione il dataset contiene 24 colonne in più date dalla computazione di media, deviazione standard e coefficiente di variazione per ogni dimensione del framework Trompenaars, quindi 3x8.
- **globe_metrics.csv**: dopo la computazione il dataset contiene 27 colonne in più date dalla computazione di media, deviazione standard e coefficiente di variazione per ogni dimensione del framework Globe, quindi 3x9.
- **hofstede_metrics.csv**: dopo la computazione il dataset contiene 18 colonne in più date dalla computazione di media, deviazione standard e coefficiente di variazione per ogni dimensione del framework Hofstede, quindi 3x6.

I nuovi dataset risultanti sono stati utilizzati per effettuare statistiche descrittive sulle comunità software prese in esame. Nello specifico sono stati sviluppati grafici a barre e violin plot.

3.4 RQ₂ - Costruzione del modello statistico

Q RQ₂. *In che misura la dispersione culturale all'interno dei team influenza il numero di community smell?*

Per rispondere alla **RQ₂** sono state computate le distanze culturali sui dataset dei community smell (Balck Cloud, Lone Wolf, Organisation Silo, Radio Silence) [2], utilizzando i tre framework culturali (Globe, Trompenaars, Hofstede). Questa operazione ha contribuito a generare 12 nuovi dataset, 3 per ogni community smell. Successivamente è stato costruito il modello di regressione lineare misto.

3.4.1 Variabili Indipendenti

Le variabili indipendenti considerate sono i coefficienti di variazione di ogni dataset, le quali si differenziano a seconda del framework utilizzato:

- **Globe:**

- Coefficiente di variazione della dimensione: Uncertainty Avoidance Societal Practices;
- Coefficiente di variazione della dimensione: Future Orientation Societal Practices;
- Coefficiente di variazione della dimensione: Power Distance Societal Practices;
- Coefficiente di variazione della dimensione: Collectivism I Societal Practices (Institutional Collectivism);
- Coefficiente di variazione della dimensione: Humane Orientation Societal Practices;
- Coefficiente di variazione della dimensione: Performance Orientation Societal Practices;

- Coefficiente di variazione della dimensione: Collectivism II Societal Practices (In-group Collectivism);
- Coefficiente di variazione della dimensione: Gender Egalitarianism Societal Practices;
- Coefficiente di variazione della dimensione: Assertiveness Societal Practices;

- **Hofstede:**

- Coefficiente di variazione della dimensione: Power Distance (PDI);
- Coefficiente di variazione della dimensione: Individualism (IDV);
- Coefficiente di variazione della dimensione: Masculinity vs Femininity (MAS);
- Coefficiente di variazione della dimensione: Uncertainty Avoidance (UAI);
- Coefficiente di variazione della dimensione: Long Term Orientation (LTO);
- Coefficiente di variazione della dimensione: Indulgence (IVR);

- **Trompenaars:**

- Coefficiente di variazione della dimensione: Universalism/Particularism;
- Coefficiente di variazione della dimensione: Individualism/Communitarianism;
- Coefficiente di variazione della dimensione: Specific/Diffuse;
- Coefficiente di variazione della dimensione: Neutral/Affective;
- Coefficiente di variazione della dimensione: Achievement/Ascription;
- Coefficiente di variazione della dimensione: Past,Present,Future;
- Coefficiente di variazione della dimensione: Sequential/Synchronic;
- Coefficiente di variazione della dimensione: Internal/External;

3.4.2 Variabile di Risposta

Per verificare quanto la dispersione culturale possa incidere sul verificarsi di un determinato community smell bisogna scegliere come variabili di risposta proprio

quelle che indicano la presenza o meno di community smell; a seconda del dataset utilizzato possiamo identificare quattro variabili dipendenti:

1. **Dataset Balck Cloud:** black;
2. **Dataset Lone wolf:** lone;
3. **Dataset Organisation Silo:** organisation;
4. **Dataset Radio Silence:** radio;

3.4.3 Variabili di controllo

Per selezionare le variabili di controllo si è fatto riferimento ai lavori di Lambiase et al. [1] e di Catolino et al. [2] dove si afferma che le seguenti variabili di controllo siano correlate all'emergere di community smell:

- **Number of Committers:** numero di persone che hanno effettuato almeno un commit, in quanto potrebbe influenzare l'emergere di community smell.
- **Number of Commits:** il totale di tutti i commit di un progetto, dato che potrebbe influenzare la presenza di community smell.
- **Team Size:** numero di membri del team in un intervallo di tempo, incrementando i membri potrebbero aumentare i problemi.
- **Turnover:** un alto turnover indica che i membri della comunità cambiano frequentemente, il che può portare a problemi di comunicazione.
- **Project Age:** si tratta della differenza tra l'indice massimo e l'indice dell'intervallo temporale di 90 giorni durante il quale è stato registrato il primo commit. I progetti più vecchi e le loro squadre potrebbero avere sperimentato tendenze o abitudini di lavoro diverse, e questi cambiamenti potrebbero influenzare la presenza di "community smells".
- **Tenure diversity:** sono state utilizzate per misurare l'esperienza degli sviluppatori in campi differenti.

- **Tenure median:** per rappresentare una media di tenure di sviluppatori di una comunità, è stata presa in considerazione anche la mediana di tenure del progetto e dei commit.
- **Number of women in a team:** Catolino et al. [2] mostrano come la presenza di donne può influenzare l'insorgere di determinati community smell, come "Black Cloud" e "Radio Silence".
- **Blau-Index:** l'indice di diversità di Blau $1 - \sum(P_i^2)$, dove P_i si riferisce alla percentuale di membri del team femminili. I valori variano tra 0 e 0,5, in cui c'è la stessa percentuale di membri del team maschili e femminili e quindi la diversità è massimizzata.
- **Socio-Technical Congruence (STC):** rappresenta "lo stato in cui un'organizzazione di sviluppo software ha sufficienti capacità di coordinamento per soddisfare le esigenze di coordinamento dei prodotti tecnici in sviluppo" [16], numero di collaborazioni di sviluppo che comunicano rispetto al numero totale di collegamenti di collaborazione presenti nella rete di collaborazione.
- **Truck Factor (TF):** misura il numero minimo di membri di una comunità che devono abbandonare prima che il progetto fallisca.
- **Centrality:** la centralità è una misura della forza di una comunità basata su misure di modularità. Un valore superiore a 0,3 indica che la comunità è altamente modulare, con una chiara distinzione tra le sottocomunità nella rete di sviluppo. Al contrario, un valore inferiore a 0,3 indica che non ci sono distinzioni chiare tra le sottocomunità, suggerendo una comunità meno strutturata.

3.4.4 Verifica delle assunzioni

Prima di procedere con la costruzione del modello di regressione lineare è bene verificare se i dati in uso soddisfano le seguenti assunzioni: [17]

- Linearità
- Omoscedasticità dei residui

- Normalità degli errori
- I residui non dovrebbero essere autocorrelati
- Nessuna multicollinearità nei dati

In questa fase si è visto che alcune assunzioni per i vari dataset non erano verificate come la Linearità, inoltre i risultati mostrano segni di autocorrelazione e alcuni fenomeni di multicollinearità positivi.

3.4.5 Costruzione del modello di regressione lineare

Per costruire il modello in questo lavoro di tesi si è fatto riferimento al lavoro effettuato da Lambiase et al. [1, 18], in questo caso però, sono stati utilizzati i 12 dataset risultanti dall'operazione di computazione delle metriche culturali. Sono stati costruiti quattro modelli di regressione lineare mista i quali hanno consentito la regressione lineare pur non rispettando le assunzioni principali [19], uno per ogni community smell "Black Cloud, Lone Wolf, Organisation Silo, Radio Silence" dove le variabili di risposta sono rispettivamente: Black, Lone, Organisation, Radio. La costruzione del modello è stato fatta con la libreria lme4 di R con le funzioni lmer e lmer.test. Per risolvere la multicollinearità si è fatto uso del pacchetto CAR (Companion applied Regression) nello specifico è stata utilizzata la funzione *vif*. Successivamente sono stati costruiti altri due modelli statistici, uno avente tutte le variabili di controllo e quelle casuali ma senza le variabili indipendenti ed un altro avente solo le variabili casuali ovvero la variabile che indica la finestra temporale. Come fatto in [1, 18] si sono confrontati i precedenti modelli di base tramite AKAIKE INFORMATION CRITERION e BAYESIAN INFORMATION CRITERION, che permettono di valutare e stimare la qualità dei modelli.

CAPITOLO 4

Analisi dei dati e risultati

In questo capitolo si mostreranno i risultati relativi alle due domande di ricerca.

4.1 RQ₁ - Dispersione culturale

Q RQ₁. *In che misura le comunità open source sono culturalmente disperse?*

Seguendo ciò che si è detto in 3.3, in questa sezione sono stati sviluppati dei violin plot che permettono di avere una visione grafica di come le comunità open source siano culturalmente disperse.

Ai fini di comprendere meglio i grafici si mostrano le Tabelle 4.1, 4.2, 4.3 riguardanti le dimensioni dei vari framework culturali .

Nel caso del framework Globe in Figura 4.1 si possono osservare 9 violin plot uno per ogni dimensione rappresentati i coefficienti di variazione (CV), ciò che si può notare è che ci sono molti CV pari a 0, ovvero, che molte comunità software non sono eccessivamente disperse. È possibile fare le stesse osservazioni per Hofstede in Figura 4.2 e per Trompenaars in Figura 4.3. In generale quindi si può affermare che le comunità non risultano eccessivamente disperse e che tutti i framework forniscono valori di dispersione simili tra loro.

Tabella 4.1: Dimensioni e coefficienti di variazione per globe

cv	Dimensione
CV_1	Uncertainty Avoidance Societal Practices
CV_2	Future Orientation Societal Practices
CV_3	Power Distance Societal Practices
CV_4	Collectivism I Societal Practices (Institutional Collectivism)
CV_5	Humane Orientation Societal Practices
CV_6	Performance Orientation Societal Practices
CV_7	Collectivism II Societal Practices (In-group Collectivism)
CV_8	Gender Egalitarianism Societal Practices
CV_9	Assertiveness Societal Practices

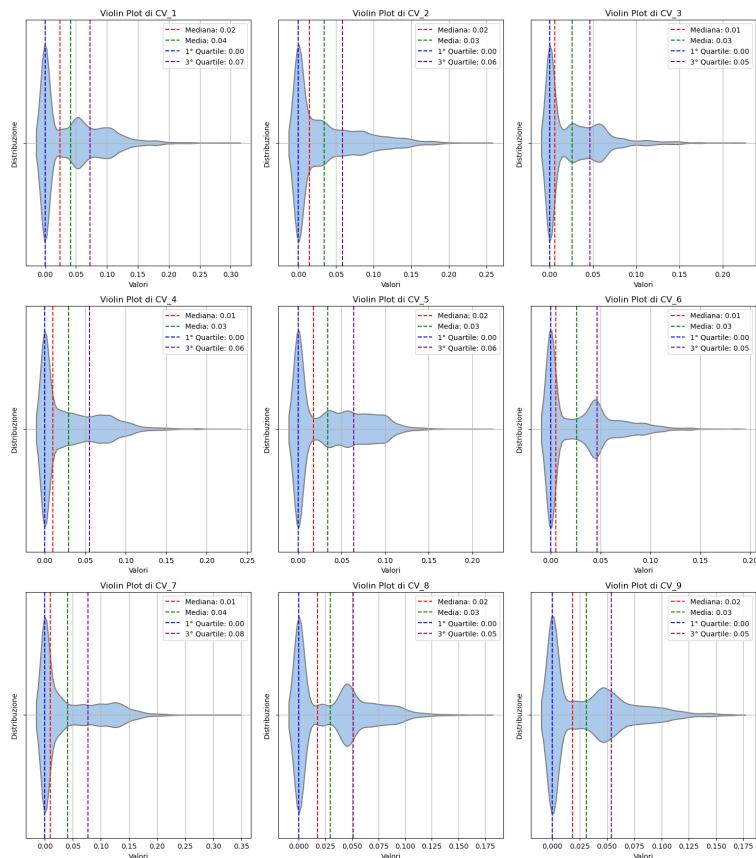
**Figura 4.1:** Distribuzione della dispersione culturale misurata con globe.

Tabella 4.2: Dimensioni e coefficienti di variazione per hofstede

cv	Dimensione
CV_pdi	Power Distance (PDI)
CV_idv	Individualism (IDV)
CV_mas	Masculinity vs Femininity (MAS)
CV_uai	Uncertainty Avoidance (UAI)
CV_ltowvs	Long Term Orientation (LTO)
CV_ivr	Indulgence (IVR)

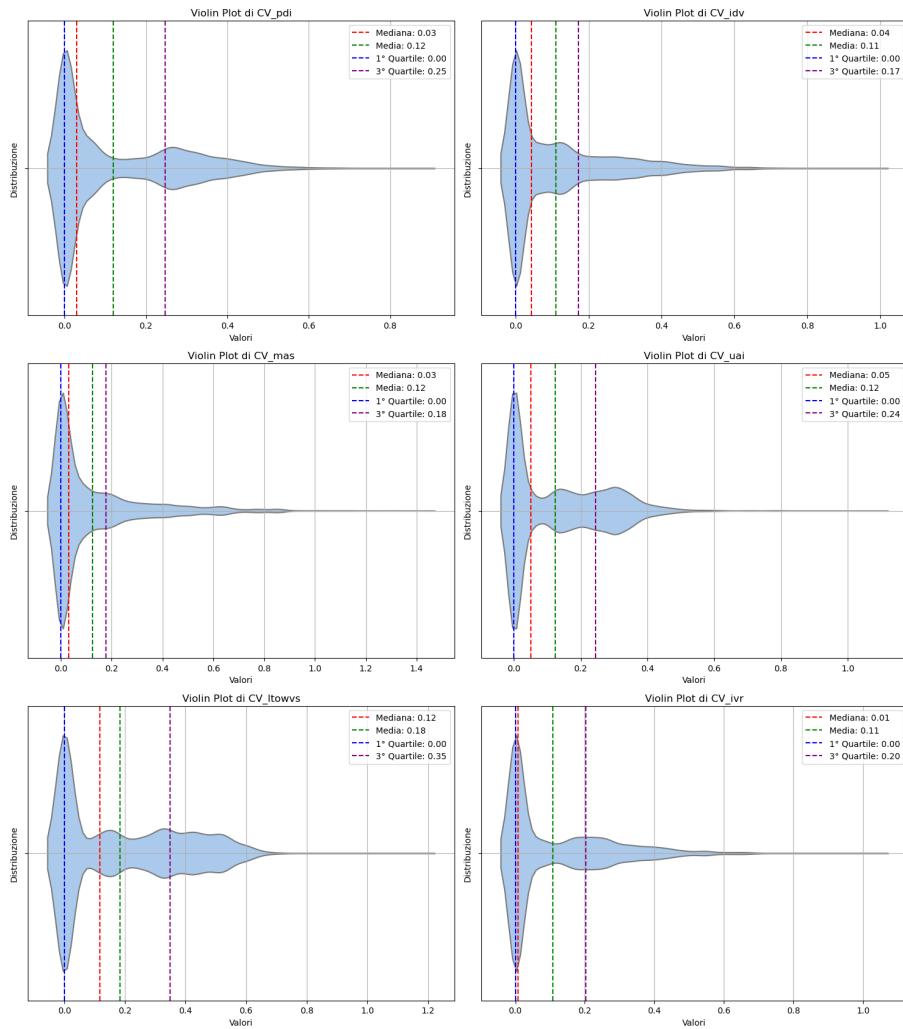
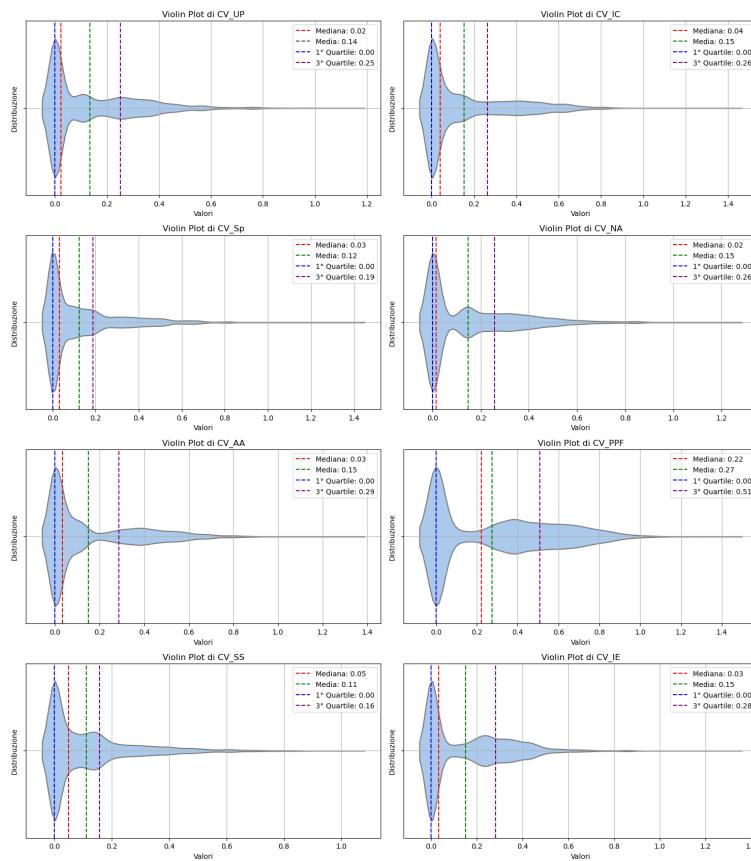

Figura 4.2: Distribuzione della dispersione culturale misurata con hofstede.

Tabella 4.3: Dimensioni e coefficienti di variazione per Trompenaars

cv	Dimensione
CV_UP	Universalism/Particularism
CV_IC	Individualism/Communitarianism
CV_Sp	Specific/Diffuse
CV_NA	Neutral/Affective)
CV_AA	Achievement/Ascription
CV_PPF	Past,Present,Future
CV_SS	Sequential/Synchronic
CV_IE	Internal/External


Figura 4.3: Distribuzione della dispersione culturale misurata con trompenaars.

4.2 RQ₂ - Analisi della dispersione culturale per Community Smell

In questa sezione si analizza, tramite dei violin plot, la dispersione culturale misurata sui dataset dei Community Smell.

Black Cloud Nelle Figure 4.4, 4.5, 4.6 si mostrano i violin plot relativi al Community Smell "Black Cloud", come si può notare le comunità sono mediamente disperse ed i diversi framework forniscono dei risultati coerenti tra loro.

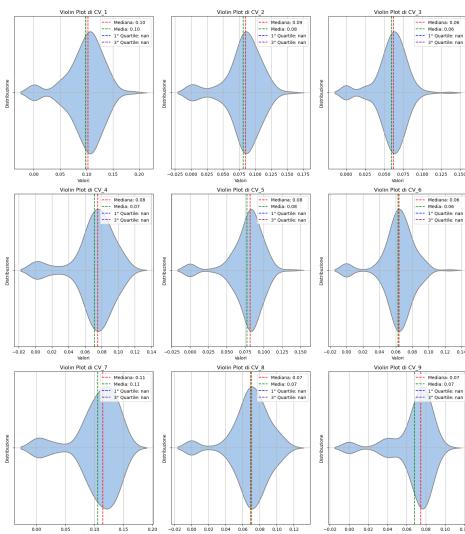


Figura 4.4: Dispersione Globe BC

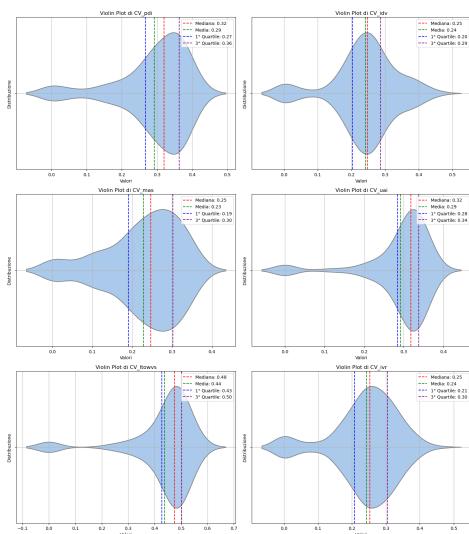


Figura 4.5: Dispersione hofstede BC

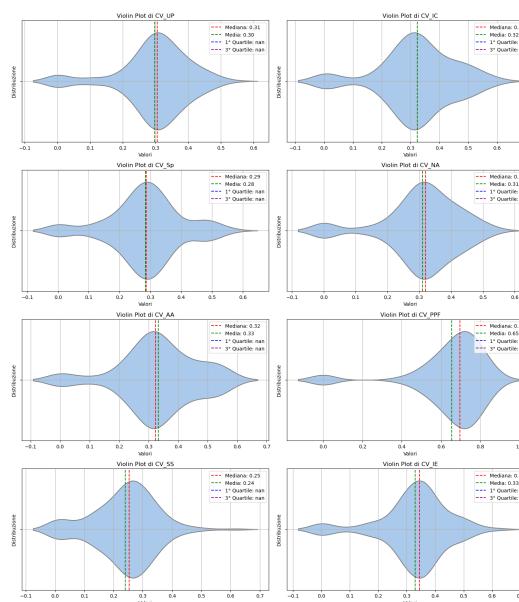


Figura 4.6: Dispersione Trompenaars BC

4.2 – RQ₂ - Analisi della dispersione culturale per Community Smell

Lone Wolf Nelle Figure 4.7, 4.8, 4.9 si mostrano i violin plot relativi al Community Smell "Lone Wolf", come si può notare le comunità sono mediamente disperse, i diversi framework non forniscono dei risultati coerenti tra loro, in quanto è possibile notare una dispersione maggiore nei grafici relativi alle misurazioni con Hofstede.

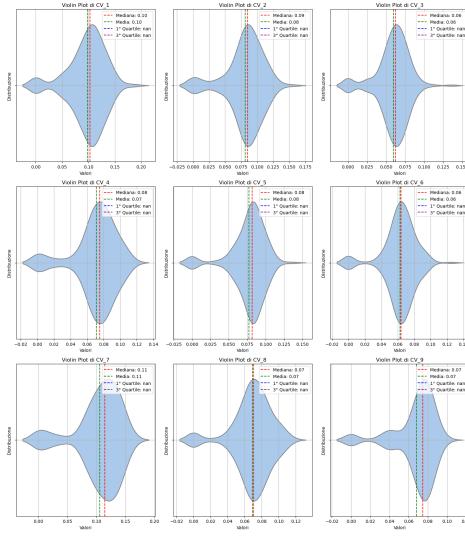


Figura 4.7: Dispersione Globe LW

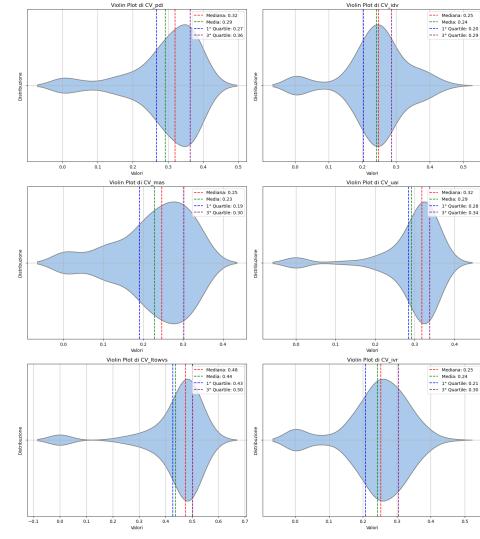


Figura 4.8: Dispersione hofstede LW

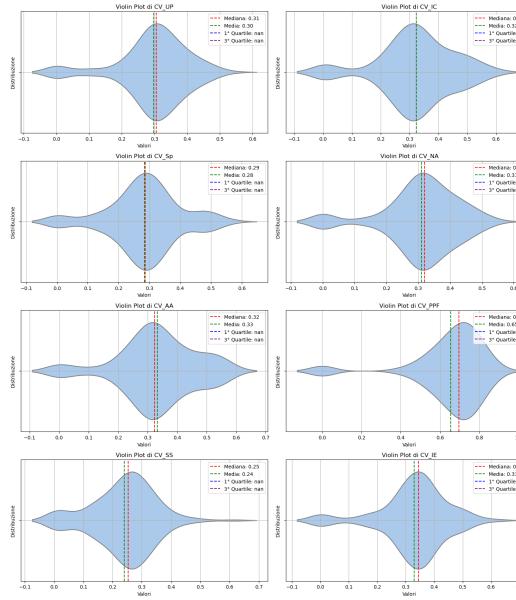


Figura 4.9: Dispersione Trompenaars LW

Organisation Silo Nelle Figure 4.10, 4.11, 4.12 si mostrano i violin plot relativi al Community Smell "Organisation Silo", come si può notare le comunità sono disperse

4.2 – RQ₂ - Analisi della dispersione culturale per Community Smell

in modo simile per le misurazioni con Globe e Trompenaars, ma maggiormente disperse per le misurazioni effettuate con hofstede.

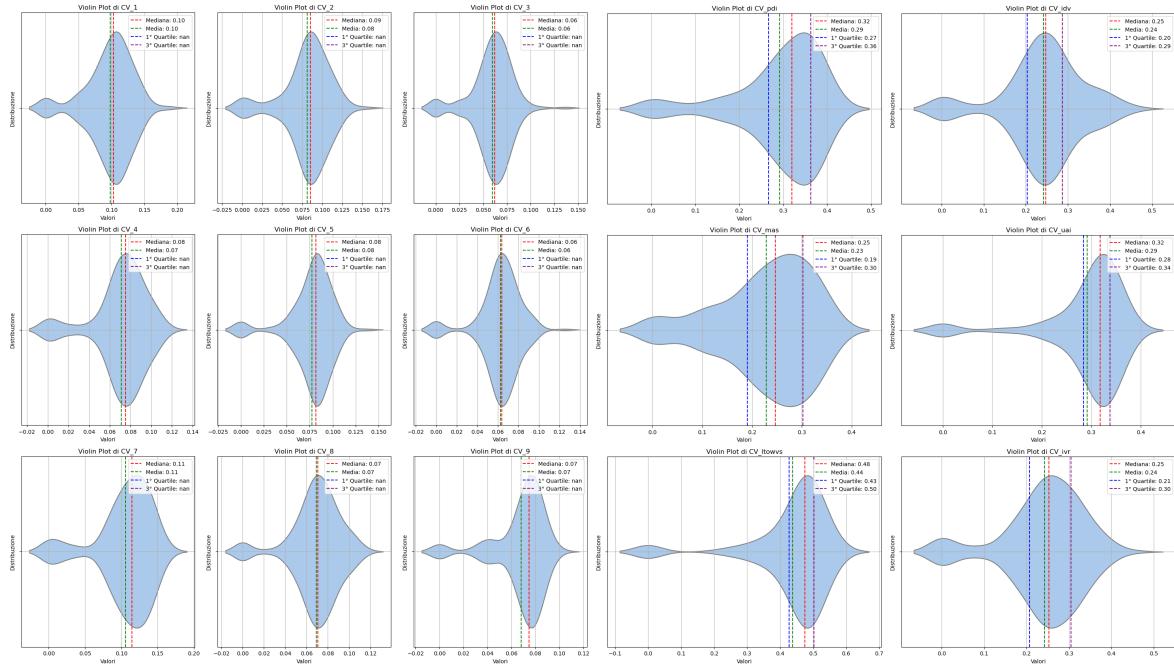


Figura 4.10: Dispersione Globe OS

Figura 4.11: Dispersione hofstede OS

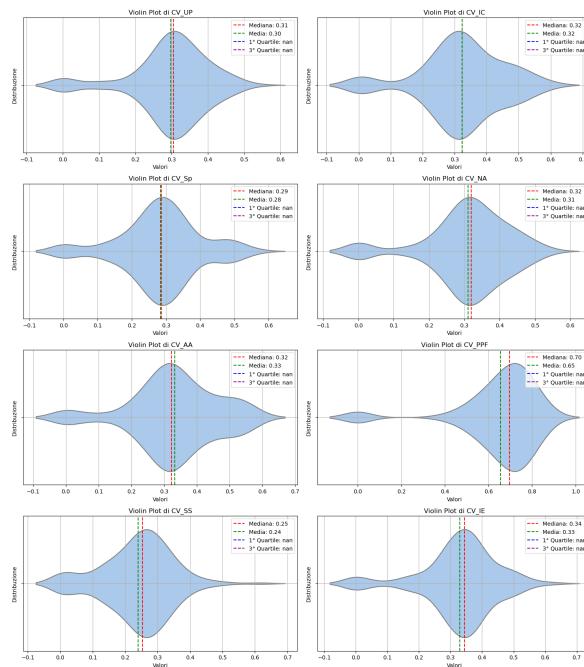


Figura 4.12: Dispersione Trompenaars OS

Radio Silence Nelle Figure 4.4, 4.5, 4.6 si mostrano i violin plot relativi al Community Smell "Radio Silence", anche in questo caso le misurazioni effettuate con il

4.2 – RQ₂ - Analisi della dispersione culturale per Community Smell

modello Hofstede forniscono una maggiore dispersione rispetto ai modelli Globe e Trompenaars.

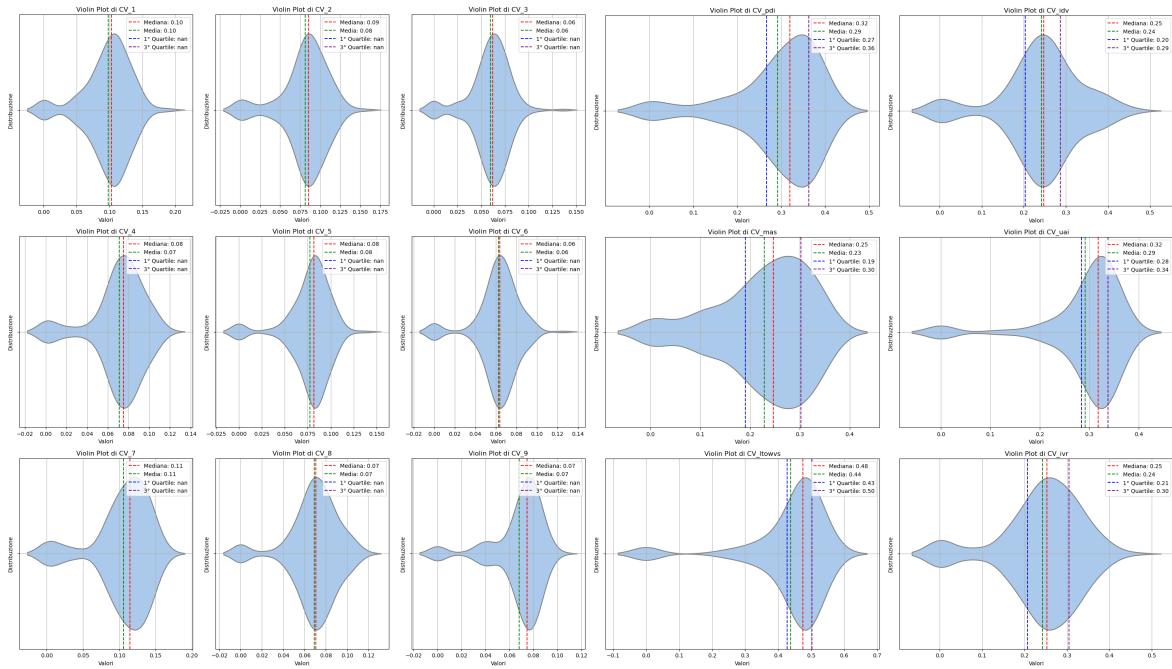


Figura 4.13: Dispersione Globe RS

Figura 4.14: Dispersione hofstede RS

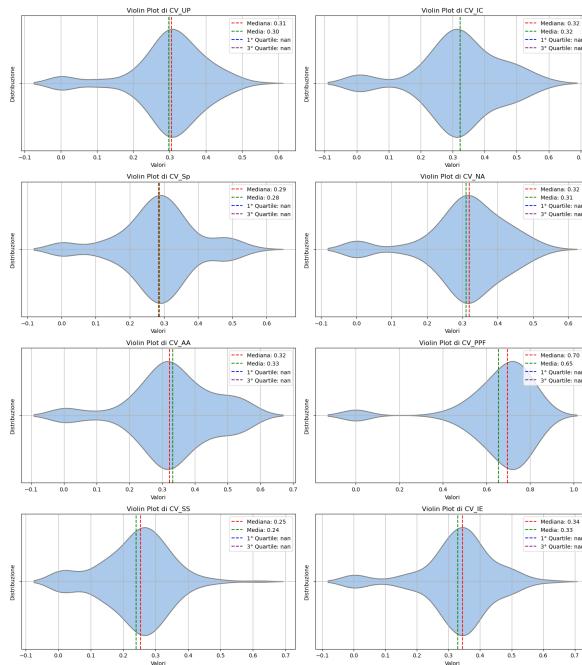


Figura 4.15: Dispersione Trompenaars RS

4.3 RQ₂ - Analisi dei risultati del modello statistico

Q RQ₂. In che misura la dispersione culturale all'interno dei team influenza il numero di community smell?

Di seguito si mostrano i risultati ottenuti dai quattro modelli di regressione lineare sviluppati, uno per ogni community smell, applicati ai dataset di ogni framework culturale. In Tabella 4.4 si ricordano i community smell considerati nello studio.

Tabella 4.4: Community smell

Nome	Descrizione
Black Cloud	Questo Community Smell riflette un sovraccarico di informazioni dovuto alla mancanza di comunicazioni strutturate o di una governance di cooperazione
Lone Wolf	Questo community Smell si verifica quando un determinato soggetto nella comunità software sfiduciato e avendo una bassa considerazione dei propri colleghi, porta avanti il suo lavoro in solitaria senza fornire aggiornamenti
Organisation Silo	Questa forma di debito sociale si riferisce alla presenza di aree della comunità separate che non comunicano, se non attraverso uno o due dei loro rispettivi membri
Radio Silence	Questa è un'istanza del problema del "unique boundary spanner" nell'analisi delle reti sociali: un membro si inserisce in ogni interazione formale tra due o più sub-comunità con scarsa o nessuna flessibilità per introdurre altri canali paralleli

4.4 Risultati Black Cloud

4.4.1 Globe

Come si può evincere dalla Tabella 4.5, che mostra i risultati del modello Black Cloud per il dataset con le distanze calcolate tramite il framework Globe, la variabile

Tabella 4.5: Tabella riassuntiva dei risultati del modello Black Cloud per il dataset Globe

Factor	All Variables			Conf. Variables			Random Variables		
	Estimate	SE	Sign	Estimate	SE	Sign	Estimate	SE	Sign
(Intercept)	5.767	1.999		5.822	1.561		4.506	0.176	
totalCommitters	-0.176	0.234		-0.135	0.206				
totalcommits	0.029	0.185		0.102	0.155				
projectAge	-0.031	0.037		-0.024	0.036				
turnover	0.666	0.803		0.782	0.724				
blauGender	-8.783	2.627	***	-10.475	2.120	***			
tenureMedian	0.029	0.087		0.049	0.087				
tenureDiversity	0.004	0.062		0.012	0.061				
teamSize	0.301	0.225							
stCongruence	-0.253	0.452		-0.222	0.456				
truckFactor	-0.020	0.115		0.006	0.114				
female	0.031	0.028		0.031	0.027				
expertise	-0.239	0.520		-0.159	0.498				
centrality	-0.557	0.363		-0.622	0.361				
CV_1	0.739	18.468							
CV_2	-6.445	14.032							
CV_3	-19.779	17.238							
CV_4	12.427	19.099							
CV_5	-1.812	14.483							
CV_6	3.928	29.068							
CV_7	7.288	9.718							
CV_8	6.368	27.350							
CV_9	-11.317	20.410							

Valori sign. ***: $p < 0.001$; **: $p < 0.01$; *: $p < 0.05$; ..: $p < 0.1$

di controllo "blauGender", (la quale indica un bilanciamento tra uomini e donne nel team di sviluppo), risulta essere molto correlata con l'emergere di fenomeni di Black Cloud dati i valori di stima (-8.783) e significatività con $p < 0.001$ molto alti. Nessuna

delle variabili indipendenti sembra essere statisticamente significativa. Questi risultati, come quelli di Catolino et al. [2], mostrano che avere un team diversificato per quanto riguarda il genere permette di ridurre i fenomeni di Black Cloud.

Relativamente al valore AIC per il modello con tutte le variabili risulta essere simile a quello con le variabili casuali e leggermente superiore a quello con le variabili confondenti, mentre il valore BIC risulta essere il più alto per il modello in questione e il più basso per il modello con le variabili casuali, questi valori potrebbero indicare che le variabili indipendenti non sono di aiuto nel rappresentare la variabile di risposta.

4.4.2 Hofstede

Nella Tabella 4.6 è possibile notare valori molto simili rispetto a quelli misurati per il framework Globe, in questo caso le metriche di dispersione sono state misurate tramite il framework Hofstede e i risultati che si ottengono sono i medesimi, ovvero la variabile "blauGender" risulta avere stima e significatività rilevanti. Vale lo stesso discorso per i valori AIC e BIC dove non vengono rilevati valori eccessivamente bassi che permettono di affermare che l'aggiunta di variabili indipendenti aiuti a rappresentare al meglio la variabile di risposta.

4.4.3 Trompenaars

In Tabella 4.7 è possibile notare che i risultati ottenuti con il framework trompenaars sono leggermente diversi dai precedenti, risulta sempre esserci una forte correlazione con la variabile "blauGender" ma qui, in aggiunta, è presente una leggera correlazione positiva con la variabile "teamSize". Per quanto riguarda le variabili indipendenti è possibile vedere delle correlazioni negative per i coefficienti CV_6 e CV_7 che rappresentano rispettivamente la dispersione delle dimensioni "Past,Present,Future" e "Sequential/Syncronic". Questi risultati affermano che all'aumentare dell'eterogeneità di genere, della propensione al passato e del lavoro a singolo task e non multitasking c'è la possibilità che i fenomeni di black cloud diminuiscano, mentre all'aumentare della grandezza del team questi fenomeni aumentino.

Relativamente ai valori AIC e BIC essi non forniscono nessun risultato rilevante

Tabella 4.6: Tabella riassuntiva dei risultati del modello Black Cloud per il dataset Hofstede

Factor	All Variables			Conf. Variables			Random Variables		
	Estimate	SE	Sign	Estimate	SE	Sign	Estimate	SE	Sign
(Intercept)	5.049027	2.067613		5.822913	1.561796		4.5060	0.1766	
totalCommitters	-0.114491	0.233109		-0.135296	0.206874				
totalcommits	0.023803	0.177852		0.102238	0.155227				
projectAge	-0.020489	0.037219		-0.024549	0.036136				
turnover	0.683087	0.763294		0.782085	0.724888				
blauGender	-8.573778	2.704472	**	-10.475619	2.120067	***			
tenureMedian	0.063706	0.088142		0.049157	0.087089				
tenureDiversity	0.015988	0.061633		0.012062	0.061803				
teamSize	0.269396	0.232283							
stCongruence	-0.212388	0.456211		-0.222344	0.456140				
truckFactor	0.009162	0.114015		0.006548	0.114495				
female	0.037219	0.028296		0.031600	0.027610				
expertise	-0.207151	0.499771		-0.159103	0.498861				
centrality	-0.590572	0.365201		-0.622232	0.361357	.			
CV_1	-0.934249	3.300639							
CV_2	0.896752	4.017696							
CV_3	1.367557	2.923935							
CV_4	-1.104878	5.257865							
CV_5	-1.257845	2.563971							
CV_6	2.155360	4.461876							

Valori sign. ***: $p < 0.001$; **: $p < 0.01$; *: $p < 0.05$; .: $p < 0.1$

a favore dell'utilizzo delle variabili indipendenti ai fini di comprendere meglio la variabile di risposta.

Tabella 4.7: Tabella riassuntiva dei risultati del modello Black Cloud per il dataset Trompe-naars

Factor	All Variables			Conf. Variables			Random Variables		
	Estimate	SE	Sign	Estimate	SE	Sign	Estimate	SE	Sign
(Intercept)	5.121611	1.973793		8.813260	2.295216		8.1875	0.3456	
totalCommitters	-0.006787	0.217109		0.214510	0.308961				
totalcommits	-0.045591	0.170887		0.184696	0.231809				
projectAge	-0.020487	0.036357		0.117693	0.053584	*			
turnover	0.022607	0.759723		0.869837	1.083180				
blauGender	-8.487706	2.538158	***	-2.227819	3.153427				
tenureMedian	0.067029	0.083761		-0.062020	0.129759				
tenureDiversity	0.014801	0.059844		0.076192	0.092256				
teamSize	0.419001	0.209524	*						
stCongruence	-0.158755	0.436897		-14.353546	0.939173	***			
truckFactor	0.038243	0.109774		-0.173819	0.171079				
female	0.036555	0.027372		0.008265	0.040935				
expertise	-0.170993	0.491624		0.489398	0.745927				
centrality	-0.628640	0.354252		0.925333	0.533883	.			
CV_1	4.319710	5.350418							
CV_2	5.404610	3.846560							
CV_3	2.886457	3.185623							
CV_4	7.408224	4.214292							
CV_5	-7.396776	5.175349							
CV_6	-3.473067	1.379752	*						
CV_7	-10.465609	3.530592	**						
CV_8	1.316173	2.602450							

Valori sign. ***: $p < 0.001$; **: $p < 0.01$; *: $p < 0.05$; .: $p < 0.1$

4.5 Risultati Lone wolf

4.5.1 Globe

Come si può evincere dalla Tabella 4.8, che mostra i risultati del modello Lone Wolf per il dataset con le distanze calcolate tramite il framework Globe, le variabili

di controllo significative sono "projectAge", "teamSize" e "stCongruence". Mentre le variabili indipendenti che hanno una significatività importante sono CV_7 e CV_8, rispettivamente Collectivism II Societal Practices (In-group Collectivism), Gender Egalitarism Societal Practices. I dati riscontrati sono coerenti con il community smell, ovvero, all'aumentare delle dimensioni del team c'è la possibilità del verificarsi del lone wolf ma che una miglior congruenza socio-tecnica può diminuire il verificarsi del fenomeno. Lo stesso discorso vale per il valore CV_7 dove l'aumentare di quest'ultimo potrebbe essere utile alla riduzione di fenomeni di lone wolf. Inoltre si può evincere che progetti di età superiore abbiano maggior possibilità di sviluppare fenomeni di lone wolf. I valori AIC e BIC risultano essere inferiori per il modello con tutte le variabili, questo dato indica che l'aggiunta delle variabili indipendenti può essere significativo per capire meglio la variabile dipendente.

4.5.2 Hofstede

Come si può evincere dalla Tabella 4.9, che mostra i risultati del modello Lone Wolf per il dataset con le distanze calcolate tramite il framework Hofstede, le variabili di controllo significative sono "projectAge", "teamSize" e "stCongruence". Mentre le variabili indipendenti che hanno una significatività importante sono CV_1, CV_2, CV_6 le quali sono i coefficienti di variazione delle seguenti distanze: Power Distance, Individualism e Indulgence. Questi risultati affermano che la grandezza del team, l'età del progetto e la congruenza socio-tecnica possono essere fattori importanti nel verificarsi di fenomeni di lone wolf ma che maggiori gradi di power distance e individualism possano ridurre il fenomeno. Relativamente ai valori AIC e BIC il modello con tutte le variabili ha ottenuto valori minori, questo dato afferma che l'aggiunta delle variabili indipendenti è utile nel comprendere la variabile di risposta.

4.5.3 Trompenaars

In Tabella 4.10 sono presenti i risultati ottenuti con il dataset avente le distanze misurate con il framework Trompeenars, i quali sono simili ai precedenti e si differenziano per i risultati delle variabili indipendenti dove emerge una significatività

Tabella 4.8: Tabella riassuntiva dei risultati del modello Lone Wolf per il dataset Globe

Factor	All Variables			Conf. Variables			Random Variables		
	Estimate	SE	Sign	Estimate	SE	Sign	Estimate	SE	Sign
(Intercept)	-10.33562	4.33548		8.813260	2.295216		8.1875	0.3456	
totalCommitters	0.59835	0.31159		0.214510	0.308961				
totalcommits	0.07653	0.24967		0.184696	0.231809				
projectAge	0.10014	0.04963	*	0.117693	0.053584	*			
turnover	-0.68879	1.09557		0.869837	1.083180				
blauGender	-5.54141	3.38513		-2.227819	3.153427				
tenureMedian	-0.04912	0.11753		-0.062020	0.129759				
tenureDiversity	0.06810	0.08347		0.076192	0.092256				
teamSize	4.76374	0.87323	***						
stCongruence	-7.67400	1.50545	***	-14.353546	0.939173	***			
truckFactor	-0.15647	0.15536		-0.173819	0.171079				
female	0.01678	0.03751		0.008265	0.040935				
expertise	0.91410	0.69801		0.489398	0.745927				
centrality	0.09764	0.50227		0.925333	0.533883	.			
CV_1	-34.33837	24.92454							
CV_2	-2.70069	18.79953							
CV_3	41.19147	22.75542							
CV_4	8.64682	25.79679							
CV_5	-9.43675	19.35032							
CV_6	34.55855	39.30303							
CV_7	-58.41290	13.08040	***						
CV_8	76.52825	36.52777	*						
CV_9	-19.13391	27.20753							

Valori sign. ***: $p < 0.001$; **: $p < 0.01$; *: $p < 0.05$; : $p < 0.1$

per i coefficienti CV_7 e CV_8 che rappresentano rispettivamente la dispersione delle dimensioni "Sequential/Syncronic" e "Internal/External". Questi risultati affermano che la grandezza del team, l'età del progetto e la congruenza socio-tecnica possono essere fattori importanti nel verificarsi di fenomeni di lone wolf ma che avere membri del team con un orientamento verso il multi-tasking permette di ridurre il fenomeno.

Tabella 4.9: Tabella riassuntiva dei risultati del modello Lone Wolf per il dataset Hofstede

Factor	All Variables			Conf. Variables			Random Variables		
	Estimate	SE	Sign	Estimate	SE	Sign	Estimate	SE	Sign
(Intercept)	-8.009780	4.344510		8.813260	2.295216		8.1875	0.3456	
totalCommitters	0.429837	0.314840		0.214510	0.308961				
totalcommits	0.080659	0.244912		0.184696	0.231809				
projectAge	0.102810	0.051221	*	0.117693	0.053584	*			
turnover	0.164711	1.052333		0.869837	1.083180				
blauGender	-5.532499	3.583147		-2.227819	3.153427				
tenureMedian	-0.078096	0.121013		-0.062020	0.129759				
tenureDiversity	0.060009	0.084987		0.076192	0.092256				
teamSize	4.525513	0.878066	***						
stCongruence	-8.261058	1.497358	***	-14.353546	0.939173	***			
truckFactor	-0.150398	0.157602		-0.173819	0.171079				
female	-0.002416	0.037965		0.008265	0.040935				
expertise	0.607594	0.690377		0.489398	0.745927				
centrality	0.151354	0.515053		0.925333	0.533883	.			
CV_1	-11.214253	4.466674	*						
CV_2	-14.548135	5.485614	**						
CV_3	0.641192	4.004666							
CV_4	10.152959	7.263162							
CV_5	-2.449517	3.363413							
CV_6	13.190426	6.088696	*						

Valori sign. ***: $p < 0.001$; **: $p < 0.01$; *: $p < 0.05$; .: $p < 0.1$

Relativamente ai valori AIC, il modello con tutte le variabili ha ottenuto risultati migliori, mentre per il BIC non si notano differenze eclatanti tra i differenti modelli. Quindi si può asserire che le variabili indipendenti permettono di capire meglio la variabile di risposta.

Tabella 4.10: Tabella riassuntiva dei risultati del modello Lone Wolf per il dataset Trompe-naars

Factor	All Variables			Conf. Variables			Random Variables		
	Estimate	SE	Sign	Estimate	SE	Sign	Estimate	SE	Sign
(Intercept)	-11.02558	4.36687		8.813260	2.295216		8.1875	0.3456	
totalCommitters	0.41449	0.30115		0.214510	0.308961				
totalcommits	0.29011	0.24042		0.184696	0.231809				
projectAge	0.10274	0.05059	*	0.117693	0.053584	*			
turnover	-0.38377	1.08616		0.869837	1.083180				
blauGender	-4.40217	3.37144		-2.227819	3.153427				
tenureMedian	-0.01189	0.11844		-0.062020	0.129759				
tenureDiversity	0.08668	0.08462		0.076192	0.092256				
teamSize	4.79067	0.87435	***						
stCongruence	-7.60461	1.49209	***	-14.353546	0.939173	***			
truckFactor	-0.18626	0.15562		-0.173819	0.171079				
female	0.01488	0.03778		0.008265	0.040935				
expertise	0.66945	0.69033		0.489398	0.745927				
centrality	0.01405	0.51151		0.925333	0.533883	.			
CV_1	-8.50224	7.44150							
CV_2	5.16374	5.49044							
CV_3	-4.75285	4.48460							
CV_4	4.90803	5.93534							
CV_5	0.46270	7.22820							
CV_6	-2.08982	1.93993							
CV_7	-13.15557	4.79748	**						
CV_8	9.96360	3.70005	**						

Valori sign. ***: $p < 0.001$; **: $p < 0.01$; *: $p < 0.05$; .: $p < 0.1$

4.6 Risultati Organisation Silo

4.6.1 Globe

In Tabella 4.11 è possibile notare che le variabili significative sono "totalCommitters" e "turnover", mentre nessuna variabile indipendente risulta significativa. Questi

risultati affermano che incrementi del numero di commiter portano leggeri incrementi del fenomeno di "organisation silo", mentre incrementi di turnover portano un decremento del community smell in questione.

Relativamente ai valori AIC e BIC non si riscontrano valori tali da poter affermare che l'aggiunta delle variabili indipendenti sia significativa per capire meglio la variabile di risposta.

4.6.2 Hofstede

In Tabella 4.12 sono presenti i risultati ottenuti dal modello per organisation silo con le metriche computate con le dimensioni di Hofstede; è possibile notare che le variabili significative per questo modello sono "turnover", "teamSize" e "female", mentre nessuna variabile indipendente risulta significativa. Questi risultati affermano che team che hanno un alto turnover possono avere riduzioni del fenomeno, al contrario, incrementi della dimensione del team possono portare leggeri aumenti del fenomeno in questione.

Relativamente ai valori AIC e BIC non si riscontrano valori tali da poter affermare che l'aggiunta delle variabili indipendenti sia significativa per capire meglio la variabile di risposta.

4.6.3 Trompenaars

In Tabella 4.13 sono presenti i risultati ottenuti dal modello per organisation silo con le metriche computate con le dimensioni di Trompenaars; è possibile notare che le variabili significative per questo modello sono "totalCommitter" e "projectAge", mentre nessuna variabile indipendente risulta significativa. Questi risultati affermano che team che hanno un alto di numero committer potrebbero avere un incremento dello smell in questione al contrario progetti che vanno avanti da più tempo hanno minor rischio che il fenomeno si verifichi.

Relativamente ai valori AIC e BIC non si riscontrano valori tali da poter affermare che l'aggiunta delle variabili indipendenti sia significativa per capire meglio la variabile di risposta.

Tabella 4.11: Tabella riassuntiva dei risultati del modello Organisation Silo per il dataset Globe

Factor	All Variables			Conf. Variables			Random Variables		
	Estimate	SE	Sign	Estimate	SE	Sign	Estimate	SE	Sign
(Intercept)	2.545158	1.491705		1.76031	1.18296		2.385	0.123	
totalCommitters	0.420619	0.178454	*	0.40372	0.15834	*			
totalcommits	-0.176115	0.140266		-0.14431	0.11819				
projectAge	-0.030251	0.028370		-0.02443	0.02768				
turnover	-1.639682	0.610633	**	-1.38135	0.5546	*			
blauGender	1.328480	1.975786		1.94591	1.61696				
tenureMedian	-0.125299	0.066325	.	-0.09215	0.06655				
tenureDiversity	0.001283	0.046992		0.01451	0.04720				
teamSize	0.290309	0.170566	.						
stCongruence	0.334951	0.342530		0.38007	0.34847				
truckFactor	0.034951	0.087388		0.04206	0.08775				
female	0.039667	0.020731	.	0.03992	0.02095	.			
expertise	-0.344030	0.395550		-0.37246	0.38274				
centrality	0.028069	0.273559		0.02872	0.27424				
CV_1	-7.507475	14.071473							
CV_2	-9.952546	10.619782							
CV_3	-8.910067	12.958495							
CV_4	15.770592	14.494298							
CV_5	-2.693425	10.854846							
CV_6	12.220115	22.018531							
CV_7	-1.534864	7.344964							
CV_8	12.259021	20.763052							
CV_9	25.090925	15.491484							

Valori sign. ***: $p < 0.001$; **: $p < 0.01$; *: $p < 0.05$; .: $p < 0.1$

Tabella 4.12: Tabella riassuntiva dei risultati del modello Organisation Silo per il dataset Hofstede

Factor	All Variables			Conf. Variables			Random Variables		
	Estimate	SE	Sign	Estimate	SE	Sign	Estimate	SE	Sign
(Intercept)	1.395977	1.541676		1.76031	1.18296		2.385	0.123	
totalCommitters	0.332618	0.177967	.	0.40372	0.15834	*			
totalcommits	-0.101356	0.134933		-0.14431	0.11819				
projectAge	-0.027446	0.028382		-0.02443	0.02768				
turnover	-1.207661	0.581334	*	-1.38135	0.55466	*			
blauGender	1.437807	2.040480		1.94591	1.61696				
tenureMedian	-0.099330	0.066761		-0.09215	0.06655				
tenureDiversity	0.021192	0.046804		0.01451	0.04720				
teamSize	0.354546	0.176055	*						
stCongruence	0.373505	0.346870		0.38007	0.34847				
truckFactor	0.039678	0.087027		0.04206	0.08775				
female	0.042810	0.020935	*	0.03992	0.02095				
expertise	-0.392552	0.381443		-0.37246	0.38274				
centrality	0.006655	0.276379		0.02872	0.27424				
CV_1	1.033520	2.511175							
CV_2	0.573831	3.040383							
CV_3	-1.959870	2.229393							
CV_4	-4.062589	3.997663							
CV_5	1.236888	1.950336							
CV_6	-1.903487	3.351464							

Valori sign. ***: $p < 0.001$; **: $p < 0.01$; *: $p < 0.05$; .: $p < 0.1$

Tabella 4.13: Tabella riassuntiva dei risultati del modello Organisation Silo per il dataset Trompenaars

Factor	All Variables			Conf. Variables			Random Variables		
	Estimate	SE	Sign	Estimate	SE	Sign	Estimate	SE	Sign
(Intercept)	1.06933	1.53801		1.76031	1.18296		2.385	0.123	
totalCommitters	0.34392	0.17087	*	0.40372	0.15834	*			
totalcommits	-0.07603	0.13384		-0.14431	0.11819				
projectAge	-0.01442	0.02876	*	-0.02443	0.02768				
turnover	-1.30050	0.59215		-1.38135	0.55466	*			
blauGender	1.70782	1.98984	.	1.94591	1.61696				
tenureMedian	-0.10970	0.06599		-0.09215	0.06655				
tenureDiversity	0.01354	0.04707		0.01451	0.04720				
teamSize	0.23340	0.16429							
stCongruence	0.35592	0.34412		0.38007	0.34847				
truckFactor	0.03276	0.08667		0.04206	0.08775				
female	0.03722	0.02117	.	0.03992	0.02095				
expertise	-0.28610	0.38883		-0.37246	0.38274				
centrality	0.06381	0.27606		0.02872	0.27424				
CV_1	-4.75624	4.15030							
CV_2	1.52702	3.04888							
CV_3	0.74848	2.49283							
CV_4	-5.83161	3.30305	.						
CV_5	3.72863	4.04585							
CV_6	1.19236	1.08454							
CV_7	3.19365	2.80149							
CV_8	-2.89280	2.04903							

Valori sign. ***: $p < 0.001$; **: $p < 0.01$; *: $p < 0.05$; .: $p < 0.1$

4.7 Risultati Radio Silence

4.7.1 Globe

Nella Tabella 4.14 sono presenti i risultati ottenuti per il modello radio silence con le distanze culturali calcolate secondo il framework Globe; le variabili significative che si possono osservare sono: "turnover", "blauGender", "CV_2", "CV_4", "CV_6", "CV_9". Questi risultati affermano che team con cambi frequenti di personale e con una forte eterogeneità di genere hanno un'importante influenza sull'emergere dello smell in questione. Inoltre si è visto che le dimensioni più rilevanti per l'emergere o meno del fenomeno del radio silence sono le seguenti: Future Orientation Societal Practices, Collectivism I Societal Practices (Institutional Collectivism), Performance Orientation Societal Practices e Assertiveness Societal Practices. Relativamente ai valori AIC e BIC, essi risultano simili tra il modello con tutte le variabili e quello con le variabili confondenti ma nettamente inferiore al modello con le variabili casuali; è possibile quindi dire che l'aggiunta delle variabili indipendenti migliora la comprensione della variabile di risposta ma bisognerebbe effettuare ulteriori controlli.

4.7.2 Hofstede

Nella Tabella 4.15 sono presenti i risultati ottenuti per il modello radio silence con le distanze culturali calcolate secondo il framework Hofstede; le variabili significative che si possono osservare sono: "turnover", "blauGender", "teamSize" e "CV_2". I risultati ottenuti permettono di capire che cambi di personale, forte diversità di genere e dimensioni del team possono influire sull'aumento dello smell in questione, tuttavia secondo CV_2 all'aumentare della dispersione sulla dimensione dell'individualismo potrebbe diminuire il fenomeno del radio silence.

Relativamente ai valori AIC e BIC, essi risultano simili tra il modello con tutte le variabili e quello con le variabili confondenti ma nettamente inferiore al modello con le variabili casuali; è possibile quindi dire che l'aggiunta delle variabili indipendenti migliora la comprensione della variabile di risposta ma bisognerebbe effettuare ulteriori controlli.

Tabella 4.14: Tabella riassuntiva dei risultati del modello Radio Silence per il dataset Globe

Factor	All Variables			Conf. Variables			Random Variables		
	Estimate	SE	Sign	Estimate	SE	Sign	Estimate	SE	Sign
(Intercept)	-4.784346	0.972422		-2.294259	0.648388		2.3582	0.1276	
totalCommitters	-0.025435	0.089977		-0.047530	0.086047				
totalcommits	0.085599	0.069163		0.028379	0.061202				
projectAge	-0.022277	0.014619		-0.024382	0.014861				
turnover	10.193689	0.445896	***	10.055517	0.460683	***			
blauGender	3.532005	1.427215	*	0.307979	0.836685				
tenureMedian	0.064317	0.034635	.	0.046144	0.035956				
tenureDiversity	0.006833	0.024389		0.009707	0.025311				
teamSize	0.222729	0.154841							
stCongruence	-0.261494	0.178272		-0.305010	0.188020				
truckFactor	-0.011461	0.045446		-0.015009	0.047116				
female	0.003486	0.022083		0.038041	0.012415	**			
expertise	0.200922	0.203579		0.079327	0.206509				
centrality	-0.093947	0.142747		-0.045816	0.148499				
CV_1	10.536556	6.881153							
CV_2	12.934455	5.580179	*						
CV_3	4.434118	6.564351							
CV_4	-19.682285	7.505983	**						
CV_5	4.883141	5.588744							
CV_6	-33.283238	11.530327	**						
CV_7	-6.901852	3.825182	.						
CV_8	-1.527238	10.081551							
CV_9	30.519367	8.082081	***						

Valori sign. ***: $p < 0.001$; **: $p < 0.01$; *: $p < 0.05$; .: $p < 0.1$

4.7.3 Trompenaars

Nella Tabella 4.15 sono presenti i risultati ottenuti per il modello radio silence con le distanze culturali calcolate secondo il framework Trompenaars; le variabili significative che si possono osservare sono: "turnover", "blauGender", "teamSize", "CV_4" e "CV_8". I risultati ottenuti permettono di capire che cambi di personale, forte

Tabella 4.15: Tabella riassuntiva dei risultati del modello Radio Silence per il dataset Hofstede

Factor	All Variables			Conf. Variables			Random Variables		
	Estimate	SE	Sign	Estimate	SE	Sign	Estimate	SE	Sign
(Intercept)	-4.7087343	0.9859096		-2.294259	0.648388		2.3582	0.1276	
totalCommitters	-0.0515603	0.0944308		-0.047530	0.086047				
totalcommits	0.0630342	0.0696188		0.028379	0.061202				
projectAge	-0.0153888	0.0149747		-0.024382	0.014861				
turnover	10.2055275	0.4514016	***	10.055517	0.460683	***			
blauGender	3.5645268	1.4600028	*	0.307979	0.836685				
tenureMedian	0.0553195	0.0356333		0.046144	0.035956				
tenureDiversity	0.0015467	0.0248089		0.009707	0.025311				
teamSize	0.3294942	0.1606533	*						
stCongruence	-0.2608575	0.1847352		-0.305010	0.188020				
truckFactor	-0.0002266	0.0460329		-0.015009	0.047116				
female	-0.0003984	0.0230506		0.038041	0.012415	**			
expertise	0.0879982	0.2017210		0.079327	0.206509				
centrality	-0.0776289	0.1467840		-0.045816	0.148499				
CV_1	1.8674750	1.3445396							
CV_2	-3.3745807	1.6101074	*						
CV_3	0.1379771	1.2034427							
CV_4	-0.7101090	2.0561636							
CV_5	0.7602673	0.9773745							
CV_6	1.7543906	1.7769908							

Valori sign. ***: $p < 0.001$; **: $p < 0.01$; *: $p < 0.05$; : $p < 0.1$

diversità di genere e dimensioni del team possono influire sull'aumento del fenomeno di radio silence, inoltre, maggiori differenze sulle dimensioni Neutral/Affective e Internal/External potrebbero portare aumenti di radio silence.

Relativamente ai valori AIC e BIC, essi risultano simili tra il modello con tutte le variabili e quello con le variabili confondenti ma nettamente inferiore al modello con le variabili casuali; è possibile quindi dire che l'aggiunta delle variabili indipendenti migliora la comprensione della variabile di risposta ma bisognerebbe effettuare ulteriori controlli.

Tabella 4.16: Tabella riassuntiva dei risultati del modello Radio Silence per il dataset Trompe-naars

Factor	All Variables			Conf. Variables			Random Variables		
	Estimate	SE	Sign	Estimate	SE	Sign	Estimate	SE	Sign
(Intercept)	-4.564420	1.002534		-2.294259	0.648388		2.3582	0.1276	
totalCommitters	-0.046905	0.090202		-0.047530	0.086047				
totalcommits	0.047368	0.069466		0.028379	0.061202				
projectAge	-0.021381	0.015103		-0.024382	0.014861				
turnover	10.151464	0.453059	***	10.055517	0.460683	***			
blauGender	3.670735	1.442741	*	0.307979	0.836685				
tenureMedian	0.054018	0.035000		0.046144	0.035956				
tenureDiversity	0.006003	0.025068		0.009707	0.02531				
teamSize	0.364748	0.152296	*						
stCongruence	-0.278528	0.182264		-0.305010	0.188020				
truckFactor	-0.005453	0.045687		-0.015009	0.047116				
female	-0.001910	0.022533		0.038041	0.012415	**			
expertise	-0.012927	0.202635		0.079327	0.206509				
centrality	-0.092194	0.145609		-0.045816	0.148499				
CV_1	1.156583	2.185867							
CV_2	-1.017977	1.569737							
CV_3	-2.171341	1.297871	.						
CV_4	3.712707	1.730648	*						
CV_5	-1.612814	2.110333							
CV_6	0.118044	0.561286							
CV_7	-2.218071	1.457544							
CV_8	2.237561	1.086692	*						

Valori sign. ***: $p < 0.001$; **: $p < 0.01$; *: $p < 0.05$; .: $p < 0.1$

CAPITOLO 5

Threats to validity

5.1 Threats to Construct Validity

Questa sezione è relativa alla relazione tra la teoria e il risultato, la quale comporta le minacce della validità che sono causate dall'inesattezza delle misurazioni. Come viene espresso nel lavoro di Lambiase et al. [1] una problematica principale riguarda le misurazioni della dispersione culturale, in questo caso i framework con i quali si è calcolata sono Globe, Hofstede e Trompenaars. Le misurazioni che suscitano maggiori perplessità sono quelle effettuate con Trompenaars in quanto non è stato fornito un dataset con i valori delle dimensioni ma i dati sono stati ricavati dal sito web dell'organizzazione [15], mentre Lambiase et al. [1] affermano che Hofstede risulta uno dei migliori framework per misurare la cultura in una comunità. Globe, invece, è potenzialmente più completo del modello Hofstede in quanto è basato su quest'ultimo ma è un modello più recente che coinvolge nello studio 170 ricercatori di oltre 60 Paesi che hanno raccolto dati su 17.000 manager di 62 Paesi del mondo [10]. Per quanto riguarda i dataset utilizzati [14, 2], sebbene sia emersa la possibilità di imprecisioni delle misurazioni, questa ipotesi risulta minima in quanto i dataset in questione sono stati già stati ampiamente validati in altri studi [1].

5.2 Threats to Conclusion Validity

Questa sezione riguarda le minacce delle conclusioni, ovvero, la relazione tra treatment e outcome. Una di queste minacce riguarda la **RQ₂** dove è stato costruito il modello statistico e le sue assunzioni, infatti, come si mostra in questo lavoro di tesi alcune delle assunzioni non erano verificate, come la linearità oppure la non presenza di multicollinearità. Si è cercato di risolvere questa problematica andando ad utilizzare un modello di regressione lineare misto [1], oppure valutando i valori AIF e BIF per la problematica della multicollinearità. Infine, come per lo studio di Lambiase et al. [1], è stato utilizzato il test ANOVA per misurare la bontà dei modelli.

5.3 Threats to External Validity

Questa sezione mostra le minacce delle validità esterne relative alla generalizzazione dei risultati dell'esperimento nel mondo reale. In questo caso, come nel lavoro di Lambiase et al [1], la minaccia principale riguarda il fatto che i dataset utilizzati [14, 2] contengono nel campo 'countries' molti dati incompleti o 'none'. Per risolvere questa minaccia, nella **RQ₁**, sono state eliminate tutte le righe che nel campo 'countries' contenevano valori 'none' maggiori del 25%, questa operazione ha permesso di non lavorare con comunità di sviluppo con eccessivi valori mancanti, al fine di avere una dispersione culturale quanto più precisa possibile.

CAPITOLO 6

Conclusioni

Il principale obiettivo di questa tesi è stato analizzare empiricamente la dispersione culturale nelle comunità di sviluppo open source e valutare l'eventuale impatto delle dispersioni culturali sull'insorgere di problemi noti come 'community smell'. A questo scopo, sono state considerate due ipotesi, dalle quali sono emerse le seguenti domande di ricerca:

1. **RQ₁** - In che misura le comunità open source sono culturalmente disperse?
2. **RQ₂** - In che misura la dispersione culturale all'interno dei team influenza il numero di community smell?

Per rispondere alla **RQ₁**, è stato utilizzato un dataset rappresentante le comunità di GitHub, su cui sono state successivamente calcolate le dispersioni culturali utilizzando i tre framework in esame: Globe, Hofstede e Trompenaars. Successivamente, tali dispersioni sono state rappresentate graficamente tramite dei violin plot, che hanno evidenziato che le comunità open source prese in esame non presentano dispersioni eccessive, indipendentemente dal modello di misurazione della cultura (Globe, Hofstede, Trompenaars).

In merito alla **RQ₂** si sono utilizzati i quattro dataset, creati da Catolino et al. [2], relativi a quattro community smell presi in considerazione nello studio (Black

Cloud, Lone wolf, Organisation Silo, Radio Silence). In primo luogo si sono mostrate le dispersioni tramite dei violin plot e misurate con i diversi modelli culturali, successivamente per comprendere al meglio la correlazione tra la dispersione e l'emergere di community smell è stato costruito un modello di regressione lineare misto [1] utilizzando come variabili di controllo tutte quelle variabili che come visto da Catolino et al. [2] avevano un impatto sull'emergere di community smell. Ciò che si è osservato è che alcuni 'community smell', come 'radio silence' e 'lone wolf', sono correlati sia in modo positivo che negativo all'aumentare della dispersione di determinate caratteristiche culturali. A seconda del modello utilizzato, sono emersi risultati leggermente diversi:

- **Globe:** Si è notato che avere personale con idee differenti riguardo alla dimensione 'Collectivism II Societal Practices (In-group Collectivism)' potrebbe portare a una riduzione delle istanze di 'lone wolf', mentre le dimensioni 'Future Orientation Societal Practices', 'Collectivism I Societal Practices (Institutional Collectivism)', 'Performance Orientation Societal Practices' e 'Assertiveness Societal Practices' sono rilevanti nella variazione del numero di istanze di 'radio silence'.
- **Hofstede:** si è visto che differenze maggiori in termini di 'power distance' e 'individualism' possono ridurre le istanze di "lone wolf", mentre maggiori differenze in termini di 'indulgence' potrebbero portare ad incrementi di fenomeni di "lone wolf". Inoltre, differenze maggiori in termini di 'individualismo' potrebbero portare a diminuzioni di istanze di "radio silence"
- **Trompenaars:** si è notato che avere diverse idee sulla dimensione 'sequential/syncronic', ovvero, differenze in termini di multi-tasking e propensione a lavori sequenziali, potrebbe portare a una riduzione delle istanze di 'lone wolf'. Inoltre, sebbene con una bassa significatività, maggiori differenze sulle dimensioni 'Neutral/Affective' e 'Internal/External' potrebbero portare ad aumenti di 'radio silence'.

Infine, i contributi che questo lavoro aggiunge sono:

1. Panoramica sulla dispersione delle comunità open source confrontata con i diversi modelli culturali.
2. Nuove relazioni tra dispersione culturale e community smell computando la cultura con modelli culturali differenti.

Sviluppi Futuri I principali sviluppi futuri riguardano il modello Trompenaars, la cui debolezza sta nella la fonte dei dati, che, sebbene siano stati prelevati dal sito ufficiale [15], sono per certi versi incompleti dato che l'organizzazione proprietaria del modello non mette a disposizione un dataset da cui attingere.

Replication package Al fine di rendere il lavoro replicabile da altri è stata pubblicata su *GitHub* un'appendice online contenente i vari dataset e script utilizzati [20].

Bibliografia

- [1] S. Lambiase, G. Catolino, D. A. Tamburri, A. Serebrenik, F. Palomba, and F. Ferrucci, "Good fences make good neighbours? on the impact of cultural and geographical dispersion on community smells," in *Proceedings of the 2022 ACM/IEEE 44th International Conference on Software Engineering: Software Engineering in Society*, ser. ICSE-SEIS '22. New York, NY, USA: Association for Computing Machinery, 2022, p. 67–78. [Online]. Available: <https://doi.org/10.1145/3510458.3513015> (Citato alle pagine 2, 3, 6, 7, 8, 13, 20, 22, 48, 49 e 51)
- [2] G. Catolino, F. Palomba, D. A. Tamburri, A. Serebrenik, and F. Ferrucci, "Gender diversity and women in software teams: How do they affect community smells?" in *2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Society (ICSE-SEIS)*, 2019, pp. 11–20. (Citato alle pagine 3, 6, 14, 18, 20, 21, 33, 48, 49, 50 e 51)
- [3] I. Richardson, V. Casey, J. Burton, and F. McCaffery, "Global software engineering: A software process approach," *Collaborative Software Engineering*, p. 35–56, 2010. (Citato a pagina 5)
- [4] C. Ebert, M. Kuhrmann, and R. Prikladnicki, "Global software engineering: Evolution and trends," in *2016 IEEE 11th International Conference on Global Software Engineering (ICGSE)*, 2016, pp. 144–153. (Citato a pagina 5)

- [5] E. Caballero-Espinosa, J. C. Carver, and K. Stowers, "Community smells—the sources of social debt: A systematic literature review," *Information and Software Technology*, vol. 153, p. 107078, 2023. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0950584922001872> (Citato a pagina 6)
- [6] D. Tamburri, P. Kruchten, P. Lago, and H. Vliet, "What is social debt in software engineering?" 05 2013. (Citato a pagina 6)
- [7] D. A. Tamburri, F. Palomba, and R. Kazman, "Exploring community smells in open-source: An automated approach," *IEEE Transactions on software Engineering*, vol. 47, no. 3, pp. 630–652, 2019. (Citato a pagina 6)
- [8] (Citato a pagina 7)
- [9] [Online]. Available: <https://www.hofstede-insights.com/intercultural-management> (Citato alle pagine 7 e 8)
- [10] D. S. Bright, A. H. Cortes, E. Hartmann, K. P. Parboteeah, J. L. Pierce, M. Reece, A. Shah, S. Terjesen, J. Weiss, M. A. White, and et al., 6.3. OpenStax, Rice University, 2019. (Citato alle pagine 9 e 48)
- [11] globe project, "2004, 2007 Studies - GLOBE Project — globeproject.com," https://www.globeproject.com/study_2004_2007?page_id=data#surveyed, [Accessed 29-09-2023]. (Citato alle pagine 9 e 10)
- [12] F. Trompenaars and C. Hampden-Turner, *Riding the waves of culture: Understanding diversity in global business.* Nicholas Brealey International, 2011. (Citato a pagina 10)
- [13] gennaro cuofano, "Trompenaars' cultural dimensions - fourweekmba," <https://fourweekmba.com/trompenaars-cultural-dimensions/>, (Accessed on 09/29/2023). (Citato alle pagine 10, 11 e 12)
- [14] B. Vasilescu, A. Serebrenik, and V. Filkov, "A data set for social diversity studies of github teams," in *2015 IEEE/ACM 12th working conference on mining software repositories.* IEEE, 2015, pp. 514–517. (Citato alle pagine 14, 15, 48 e 49)

- [15] h.-t. Trompenaars, "CultureXplore — htconsulting.com," <https://www.htconsulting.com/culture-factory/culture-explore/>, [Accessed 03-10-2023]. (Citato alle pagine 14, 48 e 52)
- [16] G. Valetto, M. Helander, K. Ehrlich, S. Chulani, M. Wegman, and C. Williams, "Using software repositories to investigate socio-technical congruence in development projects," in *Fourth International Workshop on Mining Software Repositories (MSR'07:ICSE Workshops 2007)*. IEEE, May 2007. [Online]. Available: <https://doi.org/10.1109/msr.2007.33> (Citato a pagina 21)
- [17] Dec 2022. [Online]. Available: <https://www.geeksforgeeks.org/assumptions-of-linear-regression/> (Citato a pagina 21)
- [18] S. Lambiase, "Cultural and geographical dispersions impact on collaboration and communication of software development teams," 2020/2021, available at https://github.com/SeSaLabUnisa/Tesi/blob/main/Altre/lambiase_community_dispersion/Tesi.pdf. (Citato a pagina 22)
- [19] "Intro," https://elearning.unimib.it/pluginfile.php/1055988/mod_resource/content/2/Lezione11.pdf, (Accessed on 10/04/2023). (Citato a pagina 22)
- [20] Savedest, "Savedest/impatto-delle-dispersioni-culturali-sull-emergere-di-community-smell: Appendice online del lavoro di tesi: "impatto delle dispersioni culturali sull'emergere di community smell: Confronto tra diversi modelli culturali"." [Online]. Available: <https://github.com/savedest/Impatto-delle-dispersioni-culturali-sull-emergere-di-Community-Smell.git> (Citato a pagina 52)