UNIVERSITÀ DEGLI STUDI DI UDINE

Thesis Title

di

Andrea Beggiato

Tesi presentata per il conseguimento
della laurea magistrale

in

Facoltà di scienze MM.FF.NN.
Comunicazione multimediale e tecnologie dell'informazione

Febbraio 2014

Dichiarazioni dell'autore

Io, ANDREA BEGGIATO, dichiaro che la tesi intitolata 'THESIS TITLE' ed il lavoro presentato in essa sono frutto di . Confermo che:

- Questo lavoro stato fatto interamente durante la frequentazione del corso "Comunicazione multimediale e tecnologie dell'informazione" presso questa Universit.
- Qualora sia stata presentata in precedenza qualsiasi parte di questa tesi di laurea presso questa Universit o qualsiasi altra istituzione, questo stato chiaramente affermato.
- Dove ho consultato il lavoro pubblicato di altri, questo sempre chiaramente attribuito.
- Dove ho citato lavoro altrui, la fonte sempre data. Con l'eccezione di tali citazioni, questa tesi interamente lavoro personale.
- Qualora la tesi si basa sul lavoro svolto da me insieme ad altri, espressamente sottolineato ci che stato fatto da altri e quello a cui io ho contribuito.

Firmato:

Data:

"Prediction is very difficult, especially if it's about the future."

Niels Bohr

UNIVERSITÀ DEGLI STUDI DI UDINE

Abstract

Facoltà di scienze MM.FF.NN.

Comunicazione multimediale e tecnologie dell'informazione

Doctor of Philosophy

di Andrea Beggiato

The Thesis Abstract is written here (and usually kept to just this page). The page is kept centered vertically so can expand into the blank space above the title too...

Riconoscimenti

The acknowledgements and the people to thank go here, don't forget to include your project advisor...

Contents

List of Figures

List of Tables

For/Dedicated to/To my...

Chapter 1

Introduzione

Chapter 2

Stato dell'arte

In questo capitolo sarà descritto lo stato dell'arte riguardanti gli studi e le analisi effettuate relative ad informazioni di tipo geolocalizzato riguardanti gli esseri umani; questa tipologia di dati abbraccia diversi ambiti, tra i quali l'ambito sociale e l'ambito scientifico.

L'interesse che gli accademici che si occupano di sociologia e comportamento umano è dovuto principalmente al modo in cui gli spostamenti delle persone condizionino le amicizie e relazioni tra loro, ma soprattutto come l'analisi dei dati sia spesso contrastante con l'impressione che hanno le persone stesse dei loro spostamenti, come esposto da Nathan Eagle ed altri in [1].

Un'altro ambito in cui lo studio di informazioni di tipo geolocalizzato è quello scientifico, dove si può trovare l'applicazione di diverse discipline matematiche per effettuare l'analisi di dati grezzi; tra queste la teoria dei grafi e la statistica sono sicuramente le maggiormente utilizzate. Gli studiosi che si occupano di scienze sono interessati all'aggregazione dei dati grezzi in dati pi strutturati, per poter, ad esempio, essere in grado di predirre sia le posizioni delle persone nel futuro, sia le relazioni di amicizia tra gli utenti nel tempo.
Infine, utilizzando i dati strutturati in maniera adeguata, è possibile creare un

sistema di raccomandazione e stimare la similitude tra utenti basandosi solamente su informazioni di tipo geolocalizzato.

2.1 Ambito sociale

Lo studio di Nathan Eagle ed altri in [1], il quale durato complessivamente nove mesi, si basa su 94 soggetti dello stesso gruppo di lavoro dotati di smartphone che hanno installato al loro interno alcune applicazioni che permettono di registrare ed inviare ad i ricercatori diverse informazioni, tra cui il log delle chiamate, l'identificativo dei dispositivi Bluetooth che sono stati a meno di cinque metri dal soggetto e l'identificativo della cella attraverso la quale lo smartphone riceve il segnale.

Oltre a questi dati analitici, che possiamo definire comportamentali, ad ogni soggetto è stato chiesto di compilare alcuni questionari mensili che mirano a raccogliere informazioni personali riguardanti le relazioni d'amicizia e la durata approssimativa della vicinanza con altri soggetti; i dati emersi da questi questionari vengono definiti autodichiarati.

L'analisi di tutti i dati raccolti divisa in tre fasi:

- analisi della relazione tra dati comportamentali ed autodichiarati
- analisi della presenza di dati comportamentali che caratterizzano l'amicizia

2.1.1 Relazione tra dati comportamentali ed autodichiarati

Negli ultimi trent'anni si è discusso molto sull'affidabilità delle misurazioni esistenti per le relazioni, osservando soprattutto che le osservazioni comportamentali sono debolmente correlate con le interazioni riportate dai soggetti; alcuni studi [**?**] hanno dimostrato come le persone riescono a ricordare meglio le strutture sociali a lungo termine rispetto a quelle nel breve periodo. Si possono riscontrare due diverse tipologie di bias, uno basato sul ricordo degli eventi recenti denominato *recency bias*, l'altro basato sul ricordo degli eventi pi importanti denominato

salience bias; attraverso i dati raccolti si pu quindi assimilare l'effetto di *recency bias* alla quantità di interazioni in un periodo prefissato antecedente al questionario e l'effetto di *salience bias* alla presenza o meno di una relazione di amicizia tra due soggetti.

Attraverso l'incrocio dei dati comportamentali ed autodichiarati è emerso che la maggiorparte della prossimit è misurata è stata invece dichiarata dal soggetto come non vicinanza; inoltre, quando i due dati non erano in contrasto, il tempo di contatto sempre stato sovrastimato, essendo la media dei dati comportamentali di 33 minuti al giorno contro la media dei dati autodichiarati di 87 minuti al giorno. Infine si è osservato che i dati riportati da soggetti che si reputano amici sono molto pi precisi rispetto ai dati riportati da soggetti che non si considerano amici.

2.1.2 Analisi della presenza di dati comportamentali che caratterizzano l'amicizia

Analizzare i dati comportamentali per evidenziare il grado di relazioni tra due soggetti, come l'amicizia, è diverso da misurarne la vicinanza; infatti è plausibile che due persone anche essendo amiche, siano distanti anche per periodi di tempo piuttosto lunghi. Ad ogni modo il contesto, sia spaziale che temporale, pu aiutare a definire alcuni pattern per la predizione delle amicizie, ad esempio l'aver passato con un'altra persona poche ore un sabato sera in un posto diverso dal luogo di lavoro indica una relazione differente rispetto all'aver passato quattro ore nel luogo di lavoro di un mercoledì pomeriggio.

In figura 2.1 è rappresentata graficamente la distribuzione della probabilit di vicinanza sia all'interno del luogo di lavoro, sia all'esterno, tra persone che si reputano amici reciprocamente, persone tra le quali solamente una delle parti si reputa amica dell'altra parte e persone che non si reputano amici a vicenda; si nota immediatamente come la vicinanza è pi probabile tra le prime due categorie di

persone, ma, essendo il luogo d'incontro un fattore determinante, si può osservare come la vicinanza misurata all'esterno del luogo di lavoro sia maggiore per chi si reputa amico reciprocamente.

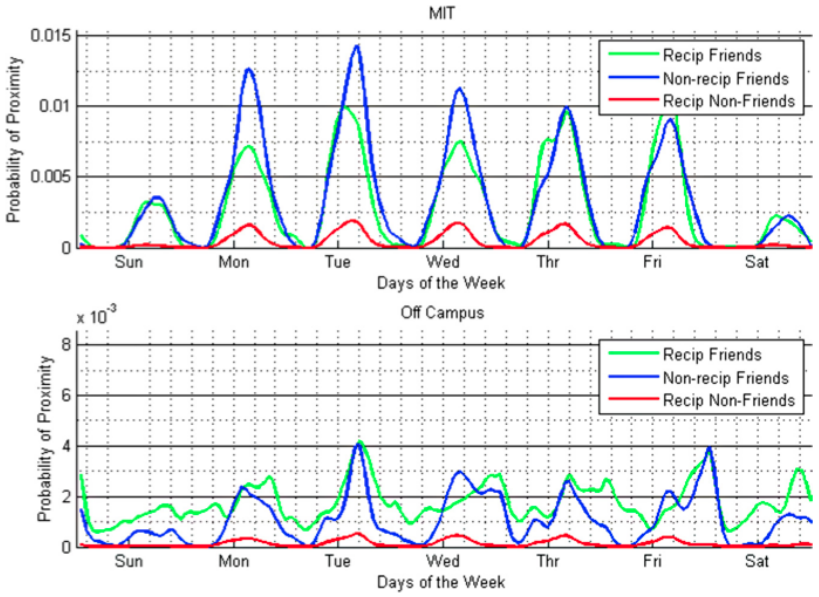

FIGURE 2.1: Distribuzione della vicinanza nel tempo e nello spazio

Nathan Eagle ed altri hanno successivamente classificato la vicinanza in diverse variabili corrispondenti alla vicinanza all'interno o all'esterno del campus, alla vicinanza di giorno o di notte e alla vicinanza nei giorni lavorativi o nei fine settimana; una fattorizzazione di queste variabili ha evidenziato come esistano solamente due fattori discriminanti, ovvero la vicinanza durante le ore del giorno nel luogo di lavoro e la vicinanza nelle ore serali o nei fine settimana all'esterno del campus. Solamente utilizzando il secondo fattore è possibile predirre il 96% percento dei rapporti di non amicizia reciproca ed il 95% percento dei rapporti di amicizia reciproca, potendo quindi tralasciare i dati autodichiarati di amicizia, quest'ultima risulta stimabile correttamente utilizzando solamente i dati comportamentali.

2.2 Ambito scientifico

2.2.1 Predizione delle traiettorie

2.2.2 Predizione delle connessioni tra utenti

2.2.3 Sistema di raccomandazione

2.2.4 Similitude tra utenti

Chapter 3

Setup dell'esperimento

3.1 Datasets

3.1.1 Foursquare

3.1.2 Geolife

3.2 Strumenti utilizzati

3.2.1 Python

3.2.2 Networkx

3.2.3 Google Maps API

Chapter 4

Dettaglio dell'esperimento

4.1 Raccolta dati grezzi

4.1.1 Visualizzazione dati grezzi

4.2 Creazione del grafo

4.2.1 Algoritmo basato sulla densit

4.2.2 Algoritmo basato sulla suddivisione

4.2.3 Algoritmo basato sul tempo e sullo spazio

4.2.4 Creazione del grafo aumentato

4.3 Analisi dei grafi

4.3.1 Analisi dei grafi personali

4.3.1.1 Analisi dei grafi giornalieri

4.3.1.2 Analisi dei grafi periodici

Chapter 5

Discussione dei risultati

Chapter 6

Conclusioni e sviluppi futuri

Bibliography

[1] David Lazer. Inferring friendship network structure by using mobile phone data.