

Smart Working

Social Network and Content Analysis

Agazzi Ruben 844736
Cominetti Fabrizio 882737

Sommario—Negli ultimi anni il tema legato allo *smart working* ha assunto una caratura di primo rilievo nelle discussioni in ambito lavorativo. Dalla necessità nel periodo di pandemia alle richieste di lavoratori che, una volta ristabilita la situazione sanitaria post-pandemia, non hanno voluto rinunciare a tale possibilità, considerata un vantaggio nel proprio equilibrio tra vita privata e lavoro. Ma il tema è discusso anche per i vantaggi offerti alle aziende, soprattutto in termini di risparmio su strutture ed energie, oppure ancora per questioni legate a traffico e inquinamento. Insomma, lo *smart working* non può più essere considerato un tema di secondo piano. Per questo motivo, abbiamo voluto analizzare il sentiment e le community degli utenti relativamente al tema, specialmente dopo gli ultimi interventi a livello governativo che hanno parzialmente ridotto la possibilità di usufruire dello *smart working*.



INDICE

1 Introduction

2 Data Collection

2.1 Data Pre-Processing

3 Analysis

3.1 Social Network Analysis .

3.1.1 Graph building .

3.1.2 Nodes degree . .

3.1.3 Assortativity . .

3.1.4 Community detection

3.1.5 Visualizations .

3.2 Social Content Analysis . .

4 Visualization

5 Summary

1 INTRODUCTION

Lo *Smart Working*, anche conosciuto come *lavoro agile*, indica una modalità di esecuzione del rapporto di lavoro subordinato caratterizzato dall'assenza di vincoli orari o spaziali e un'organizzazione per fasi, cicli e obiettivi, stabilita mediante

accordo tra dipendente e datore di lavoro; una modalità che aiuta il lavoratore a conciliare i tempi di vita e lavoro e, al contempo, favorire la crescita della sua produttività [9] [4].

Lo *smart working* è un modello organizzativo in grado di portare notevoli vantaggi alle organizzazioni che lo adottano: in termini di produttività, di raggiungimento degli obiettivi, ma anche in termini di welfare e qualità della vita del lavoratore.

Già nel 2015, il British Standards Institution affermava che: *"I principi dello smart working riconoscono che la tecnologia e i modelli di lavoro flessibile stanno cambiando in meglio il modo in cui lavoriamo* [2].

Guardando all'altra faccia della medaglia, ricerche come quella condotta da NordVPN nel 2020, hanno portato alla luce alcuni rischi per i lavoratori, tra cui un involontario aumento di ore lavorative, l'isolamento dell'individuo e la difficoltà dello stesso di separare vita sociale e attività lavorativa, oltre a pericoli legati al tema privacy e sicurezza [3].

Tra fine 2021 e inizio 2022, si stimavano intorno ai 2,9 milioni i cosiddetti "smart worker" in Italia, in aumento rispetto ai 1,15 milioni di fine 2019 ma in calo al giorno d'oggi e su un totale stimato di 8 milioni di potenziali usufruttori.

La stessa ricerca mostra inoltre come l'Italia sia fanalino di coda nell'adozione dello smart working in Europa e abbia una media di gran lunga inferiore rispetto alla media UE considerando numero di individui e totale di giorni a settimana di lavoro agile [7].

A partire da febbraio 2020, a seguito del diffondersi dell'epidemia Covid-19 del Coronavirus, sono stati emanati dal Governo una serie di provvedimenti per semplificare l'accesso allo Smart Working e diffonderne al massimo l'utilizzo. La pandemia ha infatti costretto numerosi lavoratori a fronteggiare l'emergenza sanitaria tramite lo smart working.

I numeri enunciati sopra fanno pensare che, una volta finita la fase più dura dell'emergenza - con misure tra le quali il lockdown - la maggior parte di aziende e lavoratori abbiano scelto di tornare alle modalità di lavoro tradizionali.

L'opinione pubblica è divisa tra chi ha preferito continuare ad usufruire di una modalità di lavoro agile anche in seguito, sottolineandone i numerosi benefici, e chi, con un parere opposto, sostiene l'importanza della comunicazione face-to-face e considera insostituibile l'interazione umana giornaliera, oppure ancora vorrebbe cedere il passo a nuovi strumenti e tecnologie.

In aggiunta a tutto ciò, è anche opportuno precisare che solo alcune tipologie di lavoratori possono svolgere con continuità una modalità di lavoro agile, principalmente i lavoratori d'ufficio, mentre altri hanno dunque un interesse marginale per l'argomento.

Oltre alla già citata pandemia, le direttive annesse per i lavoratori e le varie proroghe allo stato d'emergenza e allo smart working in regime semplificato, hanno portato alla ribalta, una volta di più, il fenomeno.

Il Governo ha infatti prorogato negli ultimi giorni del 2022 il diritto allo smart working con un emendamento dedicato in Manovra soltanto per i lavoratori considerati "fragili" del pubblico e del privato. Dunque, dicendo addio ai giorni di lavoro da casa per i genitori di figli under 14, fino a prima di questo emendamento compresi nella misura varata dal precedente Esecutivo, e che dall'1 gennaio sono dovuti tornare a contrattare individualmente i giorni

di lavoro da remoto con la propria azienda.

La decisione del Governo di escludere dalla misura i genitori di figli under 14 è spiegata dal venire meno delle misure anti-Covid adottate durante la pandemia, sulle quali si basava la necessità di ricorrere al lavoro da remoto per milioni di lavoratori.

La nostra analisi vuole dunque concentrarsi su questo periodo finale, tra la fine del 2022 e l'inizio del 2023, per indagare le reazioni, il sentiment e le emozioni riversate sulla rete - Twitter nel nostro caso - in seguito all'annuncio di queste decisioni.

Per concludere, il progetto vuole rispondere a domande quali *Come è rappresentato il panorama italiano intorno al tema su Twitter? Quali sono gli utenti più influenti al suo interno? Qual è stato il sentiment generale nel periodo preso in considerazione? Quali sono state le emozioni e le diverse reazioni nei giorni intorno all'annuncio del Governo sull'ulteriore riduzione dello Smart Working?*

2 DATA COLLECTION

Per costruire il dataset, necessario per rispondere alle domande di progetto, abbiamo scaricato un totale di 2667 tweet dal social network Twitter, tramite l'utilizzo dell'API fornita dallo stesso social.

I dati sono stati scaricati utilizzando un jupyter notebook e la libreria Tweepy, che permette di effettuare il download di tweet corrispondenti ad uno o più hashtag o keyword. Nel nostro caso, abbiamo scelto di utilizzare le seguenti keyword: "smartworking" OR "remotework" OR "lavoroagile" OR "hybridwork" OR "smart working" OR "remote work" OR "lavoro agile".

Siccome il tema è dibattuto e riguarda in particolare modo l'Italia, abbiamo scelto di svolgere lo studio su tweet in lingua italiana, scaricando dunque solamente i tweet nella lingua selezionata.

L'API permette di scaricare tweet con alcune limitazioni relative al numero di giorni e alle richieste di download consecutive. Il periodo d'interesse va dal 26 dicembre 2022 al giorno 5 gennaio 2023. Il periodo considera dunque anche giorni iniziali in cui alcune voci in merito iniziavano ad emergere e un periodo successivo

dopo l'annuncio ufficiale. Come possiamo osservare dalla seguente distribuzione, il giorno con il maggior numero di tweet pubblicati è il giorno 30/12/2022, penultimo giorno di lavoro agile concesso anche ai genitori di figli U14 e giorno in cui l'opzione - obbligatoria - di tornare in ufficio si avvicinava sempre più per moltissimi italiani.

Tweet per day:	
date	
2022-12-26	59
2022-12-27	386
2022-12-28	331
2022-12-29	204
2022-12-30	562
2022-12-31	276
2023-01-01	147
2023-01-02	158
2023-01-03	169
2023-01-04	227
2023-01-05	148
Name: id, dtype: int64	

Figura 1: Number of Tweets per Day

Il dataset iniziale, esportato e salvato in csv, si compone dunque di un totale di 10 variabili:

'date', 'id', 'text', 'like', 'n_rt', 'author', 'location', 'retweeted', 'user_mentions', 'hashtags'

- *date* : data e ora di pubblicazione del tweet
- *id* : identificatore unico del tweet
- *text* : testo contenuto nel tweet, estratto tramite l'opzione extended
- *like* : numero di like
- *n_rt* : numero di retweet
- *author* : nome utente dell'autore del tweet
- *location* : localizzazione di pubblicazione del tweet, non sempre disponibile
- *retweeted* : valore booleano, 'true' se si tratta di un retweet, 'false' altrimenti
- *user_mentions* : menzioni di altri utenti contenute nel tweet
- *hashtags* : hashtags utilizzati nel tweet

2.1 Data Pre-Processing

La prima operazione effettuata è stata di verificare l'eventuale presenza di duplicati, e l'eventuale rimozione degli stessi. Non sono presenti

tweet duplicati, dunque il totale di tweet resta di 2667 elementi.

Il formato della variabile 'date' è stato processato e sono state estratte le variabili 'day', 'month', e 'year' per visualizzare il numero di tweet effettuati per giorno, come visto nell'immagine di cui sopra.

Dopodiché, abbiamo creato un dataframe all'interno di cui è stato svolto un conteggio di numero di tweet effettuati dallo stesso autore. Il risultato è visibile nell'immagine sottostante:

mrmcp histo	44
infoit interno	41
aleale2780	29
saragra98354743	23
lucague46760617	22
Name: author, dtype: int64	

Figura 2: Number of Tweets per Author - Ordered

Da ogni tweet, abbiamo poi estratto la lista degli hashtags e delle menzioni contenute, inserite rispettivamente in due colonne del dataframe principale: *'hashtagsist', 'mentions'*.

Per ogni documento presente nel corpus, sono state svolte le operazioni di pre-processing necessarie per analizzare il testo. In particolare:

- *Remove Numbers* : rimozione dei numeri all'interno del testo
- *Lower Text* : testo convertito in carattere minuscolo ('lower case')
- *Lemmatization* : x
- *Remove Punctuation* : x
- *StopWords Deletion* : rimozione delle stopwords italiane
- *Tokenization* : x

3 ANALYSIS

3.1 Social Network Analysis

For the social network analysis we took the dataset of tweets we obtained from the twitter web api, and we kept only the columns relative to the tweet's author username and the list of users mentioned in the tweet.

3.1.1 Graph building

We built our graph using the python library "NetworkX[6]". Specifically we took every tweet in our dataset and for every tweet we took the tweet's author and users mentioned in the tweet, which becomes nodes of the graph, and then we make an edge between them. After this step we created inside our graph a total of 1820 nodes and 2568 edges between nodes.

3.1.2 Nodes degree

After building the graph we calculated the degree of every node inside the graph. The degree of a node of a graph is the number of edges that are incident to the node. After calculating the degree for every node we found out that the maximum degree found inside the graph is equal to 210 and the minimum degree found is equal to 1. We also decided to calculate the

Node name	Degree
BonomiAllegra	210
mrmcphesto	145
GiorgiaMeloni	141
AlbertoBagnai	118
ClaudioDurigon	82

Tabella 1: First five nodes ordered by degree

average degree of the graph, which is equal to 2.82. Because the average degree is greater than 1 we already assumed that there will be a giant component inside the graph. A giant component in graph theory is a connected component of a graph that contains a finite fraction of the entire graph's vertices.

3.1.3 Assortativity

Assortativity is a measure of how nodes inside a network connect to each other. If an interaction network is assortative nodes of comparable degree tend to link each other: bigger nodes(hubs) will tend to connect to hubs and smaller nodes will tend to connect to each other. If an interaction network is disassortative hubs will tend to connect to small degree nodes, and small degree nodes will tend to connect to hubs. If the interaction network is neutral nodes will link to each other randomly. We can calculate an assortativity coefficient of a graph; if this coefficient is greater than 0 the network

is assortative, if is equal to 0 is neutral and if is less than 0 it is disassortative. In our case the assortativity coefficient is equal to -0.244, which makes our network disassortative.

3.1.4 Community detection

Communities are parts of the graph that are more densely connected, meaning they interact a lot with each other. In this case we used the python library "NetworkX[6]" to find the community with modularity maximization. Local modularity is based on the hypothesis that random network does not have communities. Local modularity[5] compares the number of links between the real network and the one expected on a random network. We can calculate the local modularity in the following way:

$$M_c = \frac{1}{2L} \sum_{i \in C, j \in C} (A_{ij} - \frac{d_i d_j}{2E})$$

Where:

- A_{ij} : Is the actual number of edges between nodes i and j.
- $\frac{d_i d_j}{2E}$: is the expected number of edges in a random network for nodes i and j.

If the modularity is less than zero the network is randomly connected, and if the modularity for a community of nodes is less than zero then it is not considerable as a community. After running the modularity maximization algorithm we found 197 different communities, and the biggest community is formed by 234 nodes.

Community number	Number of nodes
Community 0	234
Community 1	210
Community 2	179
Community 3	139
Community 4	87

Tabella 2: First 5 communities by node count

3.1.5 Visualizations

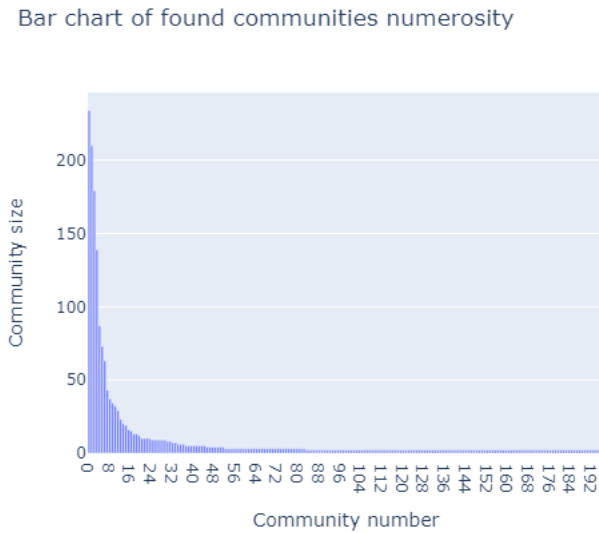


Figura 3: Bar chart of the communities numerosity.

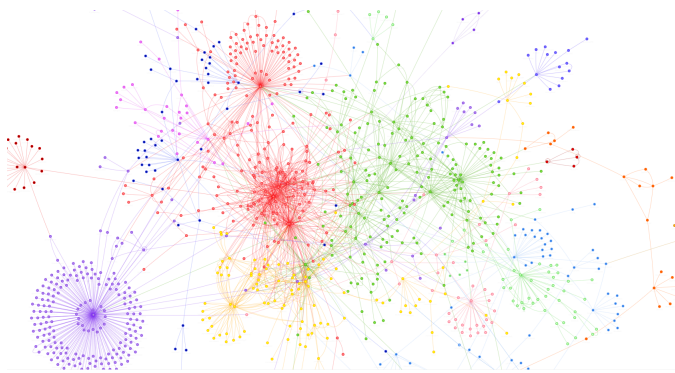


Figura 4: Visualization of part of the graph with different colors representing different communities.

3.2 Social Content Analysis

Per quanto riguarda la *Social Content Analysis*, è stato utilizzato FEEL-IT, un dataset e un pacchetto per effettuare la *Sentiment Analysis* e l'*Emotion Recognition* in lingua italiana. FEEL-IT è disponibile tramite un pacchetto Python [1] e utilizza BERT, ...

Per quanto riguarda la *Sentiment Analysis*, il pacchetto divide i tweet in 'positive' e 'negative', valori che sono stati assegnati ad una nuova colonna, chiamata 'sentiment_BERT'. Inoltre, a seconda del valore assegnato dal modello, sono

state create altre tre variabili, 'positive', 'negative', con valore 1 in caso di valore corrispondente al sentiment (quindi 1 alla colonna 'positive' in caso di sentiment individuato positivo) e 0 altrimenti, e 'ratio', di valore 1 in caso di sentiment positivo e -1 in caso di sentiment negativo. Il dataset così composto è stato salvato in un file csv per ulteriori analisi.

Il modello di *Emotion Recognition* è in grado di individuare invece quattro emozioni: 'joy', 'fear', 'anger', 'sadness'. Il procedimento svolto è il medesimo per la sentiment analysis, e sono state dunque create quattro variabili, una per ogni emozione, con valore 1 in caso di tweet corrispondente alla stessa emozione, 0 altrimenti. Anche in questo caso, il dataset è stato esportato in file csv per ulteriori analisi.

4 VISUALIZATION

...

5 SUMMARY

...

RIFERIMENTI BIBLIOGRAFICI

- [1] Federico Bianchi, Debora Nozza e Dirk Hovy. *FEEL-IT: Emotion and Sentiment Classification for the Italian Language*. URL: <https://github.com/MilaNLPProc/feel-it>.
- [2] BSI. *New Smart Working Code of Practice announced*. URL: <https://www.gov.uk/government/news/new-smart-working-code-of-practice-announced>.
- [3] Forbes. *Working From Home: VPN Use Reveals Longer Hours And May Hide Privacy Threat*. URL: <https://www.forbes.com/sites/zakdoffman/2020/03/24/coronavirus-work-from-home-longer-hours-more-distractions-and-this-surprising-privacy-threat/?sh=5eb2527c7363>.
- [4] Ministero dell'Istruzione e del Merito. *Lavoro agile*. URL: <https://miur.gov.it/lavoro-agile>.
- [5] *Modularity (networks)*. URL: [https://en.wikipedia.org/wiki/Modularity_\(networks\)](https://en.wikipedia.org/wiki/Modularity_(networks)).

- [6] NetworkX. URL: <https://networkx.org/>.
- [7] Rai News. *In Italia lo smart working è al palo: solo 2,9 milioni lavorano da casa*. URL: <https://www.rainews.it/articoli/2022/07/gli-italiani-tornano-a-lavorare-in-ufficio-negli-altri-paesi-europei-sale-il-numero--c0201a04-5af8-40b4-b041-dbf89f4eebdc.html>.
- [8] Osservatori. *SMART WORKING - Il lavoro agile dalla teoria alla pratica*. URL: https://blog.osservatori.net/it_it/smart-working-cos-e-come-funziona-in-italia.
- [9] Gazzetta Ufficiale. *Legge 22 maggio 2017, n. 81*. URL: <https://www.gazzettaufficiale.it/eli/id/2017/06/13/17G00096/sg>.