

Primo Report sull'analisi della violenza contro le donne veicolata dai social media

Dal 2020 L'Istituto Nazionale di Statistica, in collaborazione con il Dipartimento per le Pari Opportunità (DPO), ha iniziato ad analizzare il fenomeno della violenza di genere nell'ambito dei *social media*, al fine di osservare come questo fenomeno viene rappresentato e analizzare come gli stereotipi di genere sono veicolati nello spazio virtuale.

In particolare, è stato promosso un progetto di analisi del *sentiment*¹ sulle interazioni generate dai *social media* (*Twitter - X*, pagine pubbliche di *Instagram* e *Facebook*, *Webnews*) volto a osservare come i *social* producano e/o riproducano stereotipi di genere, amplifichino il linguaggio violento oppure generino indignazione e quali nuove forme di violenza di genere possono generarsi *online* (cyber-violenza).

La violenza sulle donne raccontata dai social

I *social network* sono " moderne agorà " e, durante la pandemia, sono diventati uno spazio sociale dove la violenza contro le donne si è diffusa trasferendo molte delle sue forme dal reale al virtuale. I *social media* costituiscono una nuova fonte di dati per comprendere, insieme ad altre fonti, la natura mutevole della violenza di genere. UNWOMEN in una nota del 2022² sottolinea la necessità di colmare il divario informativo sulle nuove forme di violenza di genere (definite come *technology facilitated gender based violence* – TFGBV). Tali nuove forme riguarderebbero il 38% delle donne in forma diretta e l'85% in forma indiretta³. Individuare metodologie e tecniche in grado di utilizzare queste nuove fonti di *Big Data* è essenziale per comprendere e monitorare il fenomeno, e capirne le sue nuove forme di diffusione *online*. La stessa violenza contro le donne, infatti, si sviluppa sempre più nel mondo virtuale e trova nei *social media* un canale importante di diffusione. Questa forma di violenza condivide le stesse cause che determinano le forme di violenza *offline*, ovvero gli stereotipi culturali sui ruoli di genere, su cui Istat ha già indagato⁴, che riflettono le disuguaglianze di genere nella società.

L'Istat, a questo proposito, ha elaborato nuove analisi sulla violenza di genere, basate sull'analisi del *sentiment* e delle emozioni dei contenuti dei *social media* finalizzate a rilevare le modalità con cui gli utenti dei social reagiscono alla violenza contro le donne e/o generano discussioni intorno ad essa. Il metodo adottato aiuta a capire quali sono i messaggi che scatenano la discussione collettiva, fornendo una mappa degli argomenti (come ad esempio

¹ La *Sentiment Analysis* (SA), definita come lo studio delle opinioni e dei sentimenti espressi da dati testuali (Liu, 2012), è una tecnica in rapida crescita nell'ambito della ricerca del NLP, grazie anche all'ampia gamma di applicazioni effettuate in diversi campi di analisi. Le tecniche di SA vanno da regole e metodi relativamente semplici a procedure avanzate di *deep learning* (per una rassegna dettagliata si veda: Liu 2012; Medhat, Hassan e Korashy 2014).

² UNWOMEN, *Accelerating efforts to tackle online and technology facilitated violence against women and girls (VAWG)*, <https://www.unwomen.org/en/digital-library/publications/2022/10/accelerating-efforts-to-tackle-online-and-technology-facilitated-violence-against-women-and-girls>, accesso il 13-07-2023

³ The Economist Intelligence Unit, "Measuring the prevalence of online violence against women", 2021. <https://onlineviolencewomen.eiu.com> accesso il 13-07-2023

⁴ Si veda a questo proposito l'indagine Istat sugli stereotipi di genere e l'immagine sociale della violenza. <https://www.istat.it/it/violenza-sulle-donne/il-fenomeno/stereotipi>

il *body-shaming*, il femminicidio e lo stupro), che sono rilevati attraverso la *sentiment⁵ ed emotion⁶ analysis* al fine di osservare le polarità (positive, neutre e negative) e lo spettro delle emozioni associate. I risultati della sperimentazione evidenziano una interessante coerenza con gli andamenti registrati con un'altra fonte di dati, [il numero di pubblica utilità contro la violenza e lo stalking \(1522\)](#), anch'esso recentemente entrato tra le fonti utilizzate. Questa ultima fonte permette di monitorare in tempo reale l'andamento delle diverse forme di violenza (violenze fisiche, sessuali, psicologiche, economiche, ecc.), sulla base delle chiamate pervenute al 1522.

Il metodo sperimentato può aiutare anche a monitorare le nuove forme di violenza *online* contro le donne e ad aprire una nuova strada volta a integrare diverse fonti di dati (Van den Brakel J., Söhler E., Daas P., Buelens B., 2017; Daas P., Puts M., 2014), arricchendo in tal modo il quadro informativo predisposto dall'Istat in costante aggiornamento sulla [Violenza sulle donne](#).

Attraverso un processo di cattura dei contenuti veicolati dai *social*, sulla base della presenza di almeno una parola appartenente ad un insieme di parole filtro, predisposto da esperti di dominio, è possibile raccogliere ed osservare le opinioni e i contenuti di "*post*", *tweet* e messaggi, attraverso cui si misura il modo in cui il fenomeno della violenza viene rappresentato. Inoltre, l'analisi consente di osservare come tali messaggi siano utilizzati per contrastare, condannare o isolare la cultura dello stereotipo e della violenza di genere, oppure se, al contrario, ne esaltano il lato negativo, amplificando la portata di odio e di violenza. Questa analisi fornisce informazioni su:

1. Quanto il fenomeno della violenza di genere viene rappresentato sui contenuti *social* (*Twitter, Facebook, Instagram e rassegna stampa Web*), attraverso grafici temporali relativi al numero dei *post* sia prodotti sia commentati o ripostati, relativi ad un anno di osservazione. Il periodo considerato in questo report è compreso tra il 1° novembre 2021 e il 30 novembre 2022. L'analisi di questi volumi permette di osservare gli andamenti in relazione ad eventi particolarmente centrali nella produzione di messaggi *social* sul tema (ad esempio eventi celebrativi come la Giornata internazionale per l'eliminazione della violenza contro le donne del 25 novembre) e i picchi connessi ad eventi di cronaca o altri fatti legati alla violenza di genere (come i provvedimenti giuridici) che generano reazioni contrastanti tra gli utenti.
2. La rappresentazione di come tali conversazioni utilizzino un linguaggio violento o di indignazione nei confronti dei temi legati alla violenza di genere.
3. L'analisi delle emozioni connesse agli atteggiamenti di indignazione espressi tramite i *social*.

⁵ La *sentiment analysis* (anche conosciuta come *opinion mining*) è lo studio computazionale delle opinioni, delle valutazioni, degli atteggiamenti e delle emozioni delle persone nei confronti di entità, individui, questioni, eventi, argomenti e dei loro attributi (Liu, B., Zhang, L. (2012). *A Survey of Opinion Mining and Sentiment Analysis*. In: Aggarwal, C., Zhai, C. (eds) *Mining Text Data*. Springer, Boston, MA, (pag. 415). I risultati dello studio descrivono la prevalenza di *sentiment* positivo, negativo e neutro. Si veda a questo proposito la nota metodologica.

⁶ L'*emotion detection*, similmente alla *sentiment analysis*, di cui condivide metodologie e tecniche di analisi, descrive la prevalenza di 6 emozioni (Amore, Gioia, Neutro, Tristezza, Paura, Rabbia), con cui le persone si esprimono sui *social*. Sulla scelta delle emozioni e sul processo di estrazione delle informazioni dai *social* si veda la nota metodologica.

4. L'elaborazione di una *topic analysis* sui messaggi che costituiscono i picchi di discussione per evidenziare quali siano gli argomenti trattati nell'ambiente virtuale.

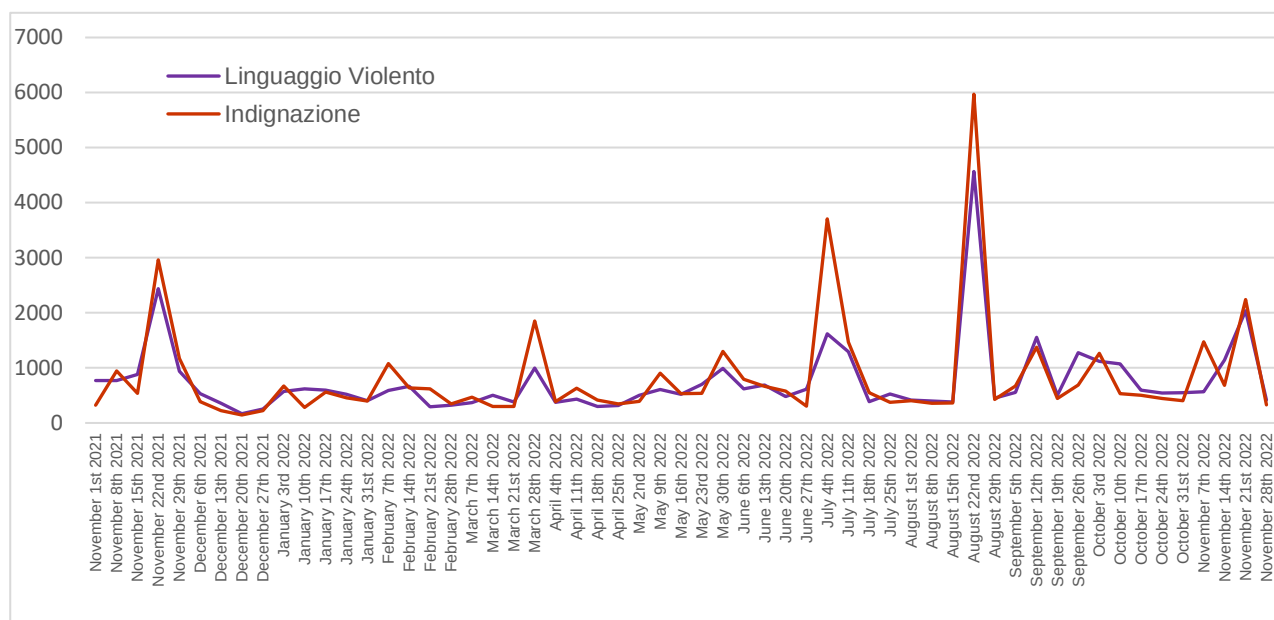
I dati dal 1° novembre 2021 al 30 novembre 2022 hanno riguardato un totale di 1.231.385 messaggi sulla violenza di genere: in prevalenza sono 1.012.110 *X-Twitter*, seguiti dai 103.442 messaggi web, 55.358 post *Instagram* e 35.205 commenti *Facebook*. I *tweet* assorbono l'82,2% dei messaggi.

Non sempre il messaggio *social* contiene opinioni o esperienze personali. A volte il contenuto di un messaggio rappresenta un'informazione oggettiva (riferimenti a notizie, *tweet* promozionali o informativi, ecc.). La metodologia adottata consente di controllare se i messaggi sono di natura prettamente informativa e giornalistica oppure se siano solo espressione di opinioni/giudizi. In questo caso, l'adozione di un classificatore binario consente di identificare due tipi molto diversi di contenuto e di realizzare l'analisi solo su *post* non strettamente informativi (si veda nota metodologica allegata) al fine di rilevare con maggior evidenza quale sia il sentimento che emerge dai *social* su determinati *topic*.

Il focus di seguito presentato si concentra sul modo con cui “si parla” *online*: esso è stato effettuato sulle reazioni di odio, di aggressività e di violenza che gli eventi e/o fatti di cronaca generano nelle conversazioni, e sul senso di indignazione che gli stessi eventi provocano tra gli *user* dei *social* e del *web*. Tale focus è stato realizzato attraverso la creazione di un *dataset* selezionato di contenuti “non informativi” estratto sulla base di un sotto insieme di *keyword* volte a cogliere le opposte categorie di odio e indignazione.

La figura 1 riporta l'andamento, nel periodo considerato, di entrambe le reazioni ed evidenzia una predominanza del senso di indignazione, segnale della presenza, tra gli utenti *social*, di una elevata consapevolezza di dover contrastare la violenza basata sul genere e sugli stereotipi. La non accettabilità degli episodi di violenza di genere appare quindi più diffusa rispetto al linguaggio violento, e ciò non solo in occasione di eventi che attirano maggiormente la produzione di contenuti (come, ad esempio, il 25 novembre), ma anche in coincidenza con fatti di cronaca e/o semplici contenuti di condanna della violenza basata sul genere. È interessante, inoltre, evidenziare come la distanza tra l'espressione di indignazione e l'uso di un linguaggio violento aumenti in presenza dei picchi legati ad episodi, informazioni e notizie che hanno provocato una più intensa discussione sui *social*.

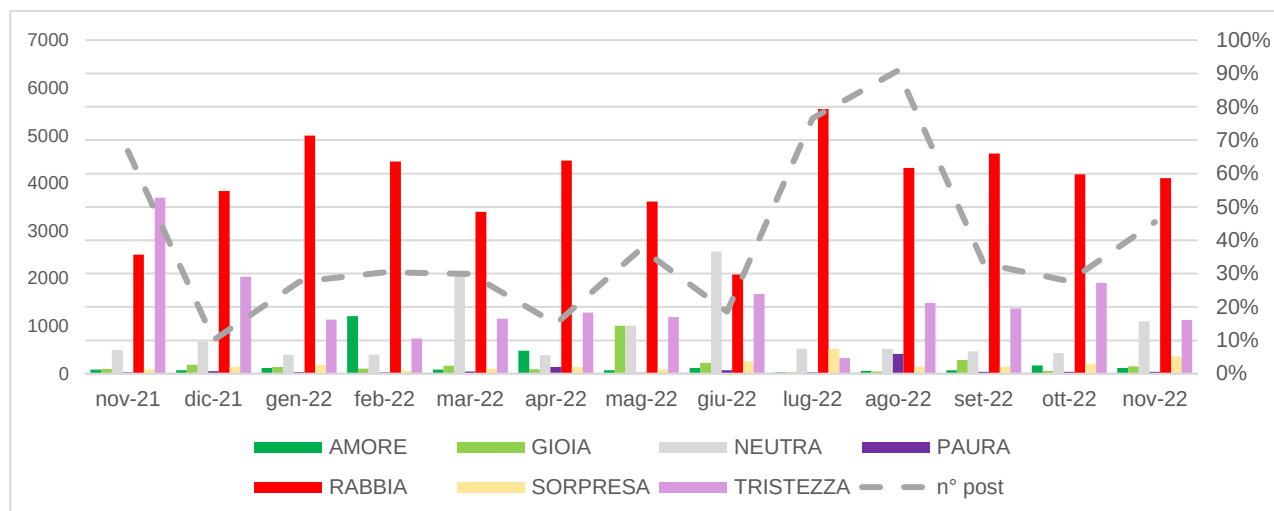
Figura 1 – Andamento giornaliero sui social del linguaggio violento e di indignazione sui temi di violenza di genere (Periodo 1 novembre 2021 – 30 novembre 2022). Valori Assoluti



Il linguaggio violento, sebbene numericamente meno presente rispetto alle espressioni di indignazione, evidenzia comunque quanto nel dibattito pubblico sia ancora molto diffusa una cultura volta a rafforzare lo stereotipo di genere, che costituisce la radice socioculturale della violenza contro le donne.

La metodologia adottata consente di andare ancora più in profondità nell'analisi per capire che tipo di emozioni si associano alle reazioni di indignazione in concomitanza dei picchi (eventi/episodi). La Figura 2 riporta queste emozioni: sono la rabbia e la tristezza ad animare vistosamente il tenore delle discussioni. In totale, infatti, le emozioni di "Rabbia" sono il 58,7% del totale, quelle di "Tristezza" il 21,3% e solamente l'1,5% dei messaggi di indignazione evidenzia un'emozione legata alla "Paura".

Figura 2 - Emotion detection registrata sui messaggi con reazioni di indignazione (Periodo 1 novembre 2021 – 30 novembre 2022). Valori Assoluti e percentuali sul numero di post totali di indignazione



La reazione dei *social* ai casi di vittimizzazione secondaria

Escludendo i picchi connessi ad eventi celebrativi (come la Giornata internazionale della donna dell'8 marzo e, ancora di più, la Giornata mondiale contro la violenza sulle donne del 25 novembre), che generano naturalmente una intensa discussione *social* su questi temi, si propone e un approfondimento sui picchi generati da notizie che hanno per lo più animato la discussione su *Twitter* - *X* e che evidenziano come queste nuove *agorà* rafforzino le forme di violenza di genere subita. L'approfondimento riguarda un caso di vittimizzazione secondaria, ovvero le situazioni e le informazioni/comunicazioni che portano a considerare la vittima colpevole dell'accaduto, fenomeno che ovviamente genera su di essa ulteriori conseguenze dannose. Si tratta del picco osservabile nel mese di luglio 2022, che riguarda una sentenza della Corte d'Appello che assolve un giovane dall'accusa di stupro con la seguente motivazione: "La ragazza lo ha indotto ad osare" in quanto ubriaca. Il picco evidenzia la difficile posizione della donna che nel momento della denuncia viene messa in discussione nei suoi comportamenti, considerati nei *social* la causa dell'evento delittuoso.

Si rende in questo modo la vittima colpevole del crimine subito. Inevitabilmente i *social* s'inflammiano di *post* carichi di rabbia contro una nuova sentenza considerata ingiusta per la vittima. Due in particolare sono i *post* che generano la discussione: il primo (che ha generato il picco di luglio 2022), di indignazione genera 1.267 *retweet*, 130 *tweet* di citazione e 10.673 mi piace; il secondo (del giorno successivo a quello osservato), dello stesso *sentiment*, è "retwittato" 627 volte, ha 22 *tweet* di citazione e 4.930 mi piace.

È possibile approfondire l'analisi di questo picco attraverso l'adozione di tecniche di *topic modelling*. Queste tecniche utilizzano una serie di modelli statistici di analisi di testi in linguaggio naturale che vengono utilizzati per estrarre le strutture semantiche in essi presenti. Di seguito si mostrano i risultati relativi al picco preso in considerazione (dal 7 all'11

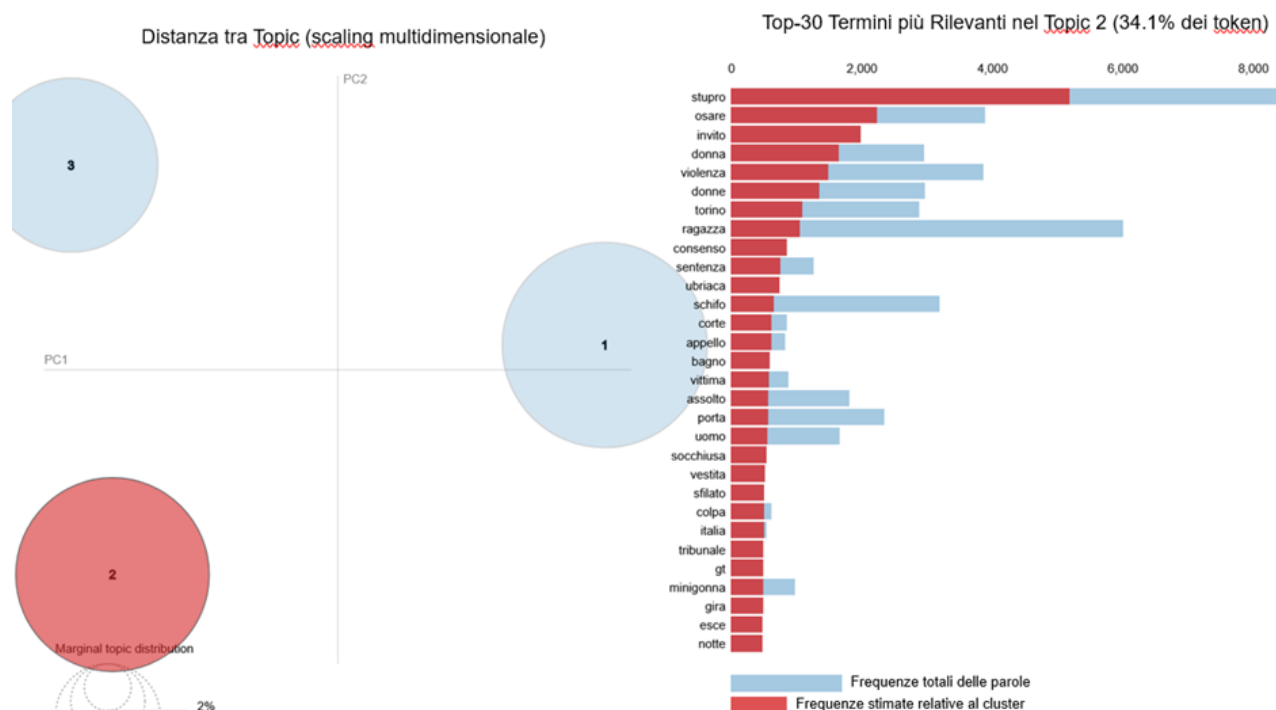
luglio 2022) ottenuti utilizzando un modello basato sulla *Latent Dirichlet Allocation (LDA)* (vedi nota metodologica). Queste classi di modelli a variabili latenti hanno il vantaggio di far emergere in modo automatico dei *pattern* semantici che possono essere interpretati come *topic* impliciti per analisi di grandi quantità di testo (di qui la dizione “*Topic analysis*”). È compito dell’analista validare ed interpretare il tema prevalente del *cluster*. Rispetto ad altre tecniche di analisi semantica non supervisionata, questa tecnica inoltre offre il vantaggio di quantificare le dimensioni del *cluster*, ovvero quante parole della collezione di testi esaminati sono in esso contenute, nonché stimare le frequenze delle parole stesse contenute nel *cluster*. Altre tecniche permettono di stimare solo qualitativamente le relazioni semantiche tra parole.

L’analisi è stata svolta su un totale di 26.320 *tweet* della settimana di riferimento. Nel seguito si indica con “*corpus*” l’insieme di tutti i *token* (parole) presenti nei *tweet* considerati. La metodologia *LDA* crea un modello in cui ciascun *topic* identificato dall’analisi è descritto da una combinazione di parole del corpus; le parole presenti nel corpus vengono associate dal modello ai tre *topic* con una certa probabilità (frequenza stimata).

L’output dell’analisi è mostrato nella Figura 3, in cui vengono riportati i tre *cluster*. In particolare, le aree semantiche individuate nei *tweet* relativi al periodo di riferimento si caratterizzano come: “1 - *Body shaming*”, “2 - Vittimizzazione secondaria” e “3 - Discussione sulla sentenza per stupro”. A sinistra, i *topic* sono rappresentati come cerchi in uno spazio bidimensionale tramite *scaling* multidimensionale, una tecnica statistica che permette di rappresentare le differenze tra gruppi in uno spazio N-dimensionale con N fissato: il raggio di ogni cerchio è proporzionale alla frequenza delle parole nel topic. La distanza tra i topic diminuisce o aumenta in funzione del numero di parole che i topic hanno in comune.

A destra, sono riportate le frequenze stimate relative al totale dei *tweet* nel *corpus* (in blu) e relative al cluster di riferimento (in rosso). La Figura 3 illustra il cluster numero 2, relativo al tema della “vittimizzazione secondaria”. La percentuale di occorrenza delle parole associate al cluster è stimata pari al 34,1 % del *corpus*.

Figura 3 –Topic analysis (LDA) relativa al picco di luglio 2022



Nel grafico tra le parole che caratterizzano il cluster 2, ordinate per frequenza stimata, compaiono “stupro”, “osare”, “invito”, “violenza”, “donne”, “Torino”, “ragazza”, “consenso”, “sentenza”, “ubriaca”. In particolare, la parola “invito”, “consenso” e “ubriaca” sono presenti esclusivamente nel cluster 2 e sono le parole che identificano l’atteggiamento di coloro che ritengono la vittima di violenza co-responsabile, almeno in una certa misura, dell’episodio violento (per l’appunto, è l’atteggiamento che dà origine al fenomeno della “vittimizzazione secondaria”).