



UNIVERSITÀ DI PISA

Social Network Analysis – Pseudo Report

Giulia Calvo

Ukraine War Analysis on Twitter

Premessa

Il seguente report è solo un estratto dell'intero lavoro svolto dal gruppo. I nodi e i link che verranno descritti di seguito sono frutto di un dataset creato usando i dati che abbiamo estrapolato da Twitter per poter analizzare il fenomeno "Ukraine war".

1) Data preparation

a) Idea

La nostra analisi mira a costruire una rete ispirata alla diffusione del dibattito (tweet e retweet) sullo scoppio del conflitto in Ucraina.

Abbiamo iniziato ad analizzare i tweet in inglese dal 15 febbraio 2022.

Tendendo conto delle conoscenze che avevamo sul conflitto basate sulle principali testate giornalistiche e delle dinamiche sui social che frequentiamo abbiamo optato per la seguente divisione degli utenti in 4 gruppi:

- "Pro Ukraine"
- "Pro Russia"
- "Pax Romana"
- "I don't care"

L'obiettivo è categorizzare gli utenti esaminando i loro tweet.

L'analisi che effettueremo sui tweet comporterà, oltre all'osservazione degli hashtags in essi contenuti, anche l'analisi dell'andamento degli stessi con l'obiettivo finale di andare a individuare quelli che potrebbero essere potenziali bot all'interno della rete.

a. Twark

Attraverso la libreria di Twark abbiamo implementato diverse funzioni che ci hanno permesso di svolgere la nostra ricerca dei vari tweet:

- SocialETL, ci permette di estrarre con l'utilizzo di una query una determinata keyword
- Count, ci permette di individuare con quale frequenza una determinata keyword appare su Twitter
- hashtags, ci permette di estrarre gli hashtags dai tweet

Tutte queste funzioni vengono considerate all'interno di un range temporale che va dal 15 febbraio in poi.

METODOLOGIA PER LA CATEGORIZZAZIONE

- creare una lista per ogni categoria
- individuare un hashtag padre per ogni categoria, il quale andrà inserito nella lista appartenente
- selezionare i figli di ogni padre che verranno inseriti nella lista appartenente

Hashtags-padre

Nel nostro caso abbiamo individuato 3 principali categorie con i seguenti hashtag padre:

- a) Pro Ucraina - #slavaukraini
- b) Pro Russia - #istandwithputin
- c) Pax Romana - #stopwarinukraine

Gli hashtag padre sono stati selezionati utilizzando la funzione Count, attraverso la quale abbiamo ricercato gli hashtag più utilizzati nei tweet.

Per ogni categoria abbiamo individuato dei possibili candidati attraverso ricerche sui principali social e attraverso i motori di ricerca.

Per identificare l'hashtag "padre" per ogni categoria attraverso la funzione Count abbiamo selezionato il più "twittato".

Pro Ucraina		Pro Russia		Pax Romana	
	Count		Count		Count
#slavaukraini	871114	#istandwithputin	259662	#stoprussia	1887070
#IStandWithUkraine	633918	#istandwithrussia	153669	#stopwarinukraine	250781

Per quanto riguarda la categoria PaxRomana, anche se l'hashtag candidato #stoprussia ha raggiunto un numero maggiore di tweet, abbiamo deciso di scegliere l'hashtag #stopwarinukraine in quanto ci sembrava semanticamente più affine alla linea di pensiero della categoria che andrà a rappresentare. L'hashtag #stoprussia ci sembrava troppo "filo Ucraino" per rappresentare la categoria PaxRomana.

Hashtags- figli

Gli step che abbiamo seguito sono i seguenti

1. Abbiamo creato le liste di hashtag per categorizzare tra pax, pro-russia e pro- ucraine. Dopo aver individuato il padre di ogni categoria abbiamo ricercato tutti i tweet contenenti tali hashtag padre, e gli ulteriori hashtag che trovavamo nel tweet vengono inseriti in una lista (figlio- count del figlio) che li collegava al relativo padre.
2. Abbiamo usato una query per cercare i tweet che contenevano come parola chiave hashtag correlati al conflitto e non. Abbiamo quindi aggiunto una colonna nel dataset per assegnare una determinata categoria al tweet.
3. Le liste di hashtag create prima ci sono servite a distinguere tra le tre categorie pax, pro-Russia e pro- ucraine e la quarta categoria "I dont care"; tenendo conto di quest'ultima un utente tipo è una persona che non ha usato nessuno degli hashtag individuati al punto 1.
4. Arrivati a questo punto dovevamo categorizzare il nodo, quindi utilizzando l'ID dell'utente contenuto nelle info dei tweet un utente è categorizzato se la maggior parte dei tweet che ha pubblicato sono di una determinata categoria.
5. Abbiamo creato i link in base a retweet e discussioni tra soggetti

