



עקסט נתונים - עיבוד שפה טבעיות וניתוח רשותות

פרויקט הקורס



מרצה: ד"ר מיה שטמר

מגישים:

יונתן גולן - 387380632
יואב אילן - 101019490209



תוקן עניינים

3.....	1. בחירת הנושא והציגת הנתונים.....
4.....	2. ניתוח רשותות.....
4.....	2. נתונים כלליים על הרשות..... 2.1
5.....	2.2. ויזואלייזציה.....
9.....	2.3. מטריקות עיקריות.....
10	2.4. מדדי מרכזיות.....
11.....	2.5. קהילות
12	3. עיבוד שפה טבעית.....
12	3.1. עיבוד טקסט מקדים.....
14	3.2. ניתוח סנטימנט.....
16	3.3. מילוט מפתח.....
17	3.4. ניתוח נושאי.....
18	3.5. ענן מילים.....
20	4. עיבוד שפה ודינמיקה חברתית.....
21	נספחים
21	נספח 1 - פירוט רשימת הדמיות
22	נספח 2 - פירוט רשימת החברות המתחרות
23	נספח 3 - רשימת המילים TF-IDF
23	נספח 4 - ניתוח נושאי



1. בחירת הנושא והצגת הנתונים

התמקדנו בניתוח כולל של הקשרים החברתיים סביב חברת טסלה ודמויות מובילות בתעשייה הרכבת החשמלי והטכנולוגית בכלל בארצות הברית, כפי שהתבטאו ב-subsreddits reddit ייעודיים ב-[Reddit](#).

ה-subsreddits מתחום לקחנו את הפוסטים הם - [TeslaInvestorsClub](#), [TeslaMotors](#), [elonmusk](#), [RealTesla](#), [politics](#), [technology](#), [electricvehicles](#), [investing](#), [wallstreetbets](#)

הדמות העיקריות בהם בחרנו לחפש את הקשרים הם -

Elon Musk, Jeff Bezos, Larry Page, Donald Trump, Mark Zuckerberg, Joe Biden, Cathie Wood, Jim Cramer, Chamath Palihapitiya, Michael Burry, Gavin Newsom, Alexandria Ocasio-Cortez, Pete Buttigieg, Bernie Sanders ([בנוף 1](#) מוצג פירוט על הדמויות).

החברות מתחנות וחברות בעלות עניין הם -

[Rivian](#), [NIO](#), [Lucid](#), [BYD](#), [Ford](#), [GM](#), [Apple](#), [Meta](#), [Palantir](#).

בכל אחד מה-subsreddits שהזכרנו חיפשנו אזכורים של כל דמות או חברת מתחנה באמצעות רישימת היכיניות והכינויים החלופיים שהגדרנו (למשל "CEO", "Mus�", "Tesla CEO", "Elon" לאילון מאסק), כדי ללווד כמה שיוור שיח רלוונטי על כל אחד מהם.

בסוף השתמשנו ב- API Reddit כדי להביא את המידע הנדרש והגדרנו שנביא עבור כל subsreddits 50 פוסטים מוביילים ועבור כל פוסט את 20 התגובה המוביילים. סך הכל השתמשנו ב-218 פוסטים וב- 1973 תגובות.

חשוב לציין ש עקב הגבלה של השימוש ב-API שמרנו לקליטת את הנתונים כדי שנוכל לנתח אותם בצורה מהירה ולא צריך לכריית הנתונים מחדש, עם זאת הקוד שבינו יוכל לקחת את הנתונים מחדש ולدرוס את הנתונים הקודמים כדי לנתח את הנתונים העדכנים יותר.

כדי לעבוד בצורה מודולרית יצרנו מספר מחלקות שמהותן כל אחת מהן שונה מבחינת השירות אותה היא מספקת. כל מחלוקת יוצרת אובייקט שונה שמטפל במקרה אחר של העבודה לדוגמא - עיבוד מקדים, צייר גרף, ניתוח טקסט וכדומה.



2. ניתוח רשות

2.1. נתוניים כלליים על הרשות

יצרנו 4 רשותות כאשר כל רשות מנתחת את הקשרים בין הדמיות המובייליות והחברות המתחרות בתקופות זמן שונות. תקופות הזמן הן - לפני 2020, 2020-2021, 2022-2023, 2024-2025.

הצמתים בכל גרפּה הם הדמיות והחברות. בכל גרפּ יש עד 23 צמתים (יש שנים שבהם אין אזכור לדמיות או לחברות). הקשרים בין הצמתים מייצגים הופעה משותפת של הדמיות והחברות באותו פוטט או תגובה. עובי הקשת משקף את תדירות ההופעה המשותפת - ככל שמספר הפסעים או התגובהות שבהם הוזכרו יחד גדול יותר, כך הקשת המחברת ביניהם תהיה עבה יותר. צבע הצומת הוא הסנטימנט הממוצע של אותה דמות/חברה. באופן דומה, צבע הקשת הוא הסנטימנט הממוצע בין 2 צמתים. ברשותות אין משמעות לכיוון של הקשר, כלומר הגרפים לא מכונים. כל הגרפים קשורים.

שאלות מחקר:

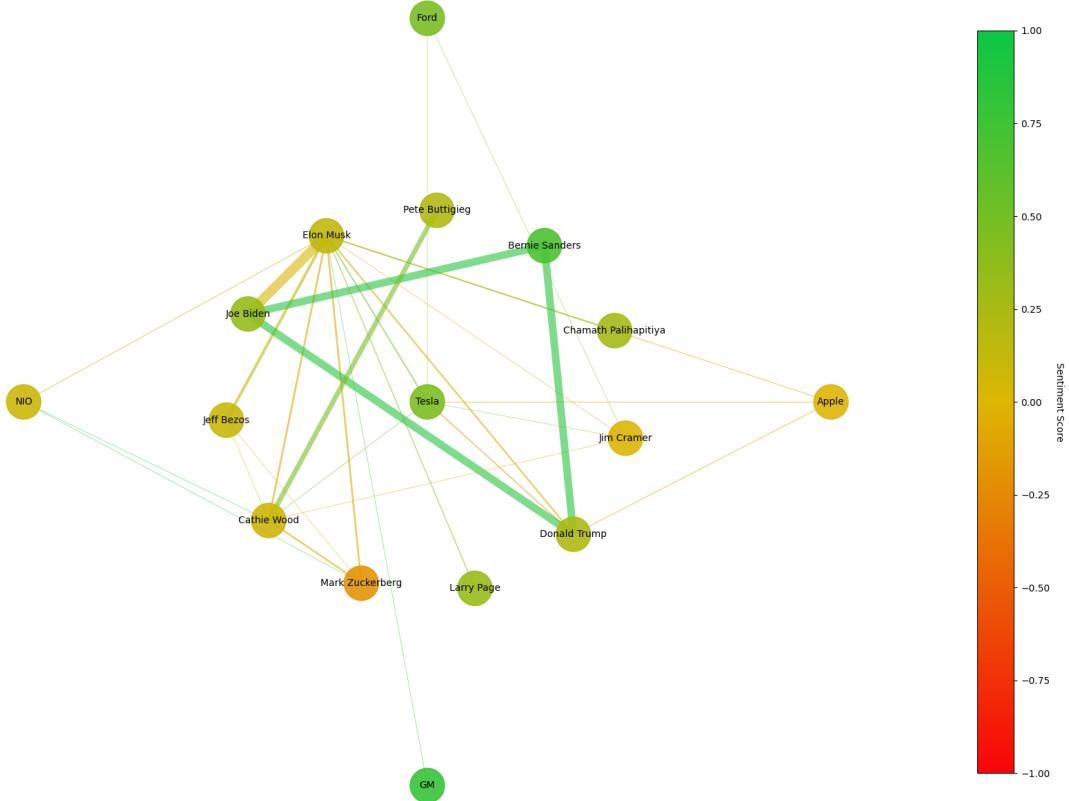
- מה מעמדו ותפקידו של אילון מספק, כדמות מרכזית בטולה ובשייב subreddits השונים וכייזד מתבטאת מרכזיותו ברשות לארוך התקופות?
- כיצד משתנה הדינמיקה בין הדמיות ברשות לארוך זמן, ומהם השינויים הבולטים בקשרים ובסנטימנט ביניהם?
- האם ניתן לזהות קשרים חריגים ומה המשמעות של הופעת קשרים חיוביים ושליליים מנוגדים בין תקופות שונות?
- כיצד בניו מבנה הרשות לארוך התקופות - האם ניתן לזהות קבוצות מובחנות, והאם קיימות דמיות שימושísticas כמחויבות בין קהילות שונות?



2.2. ויזואלייזציה

כאמור, צרנו 4 רשותות ע"פ תקופות הזמן השונות.

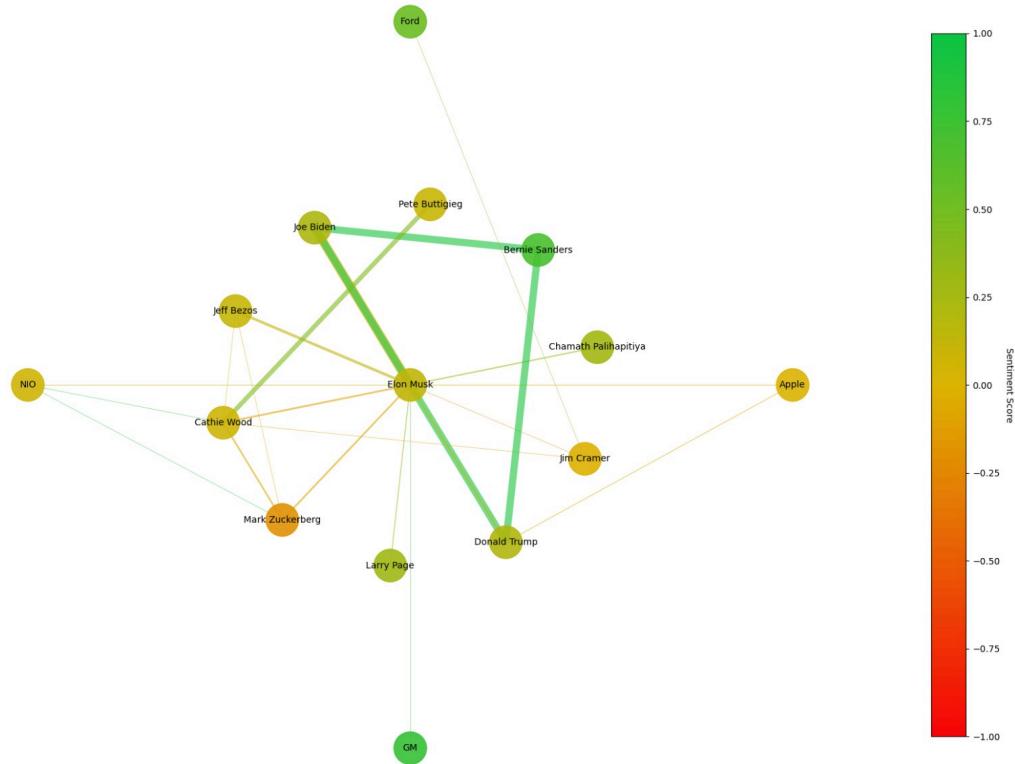
Tesla & Influencers for period <2020 (Node colors based on sentiment)



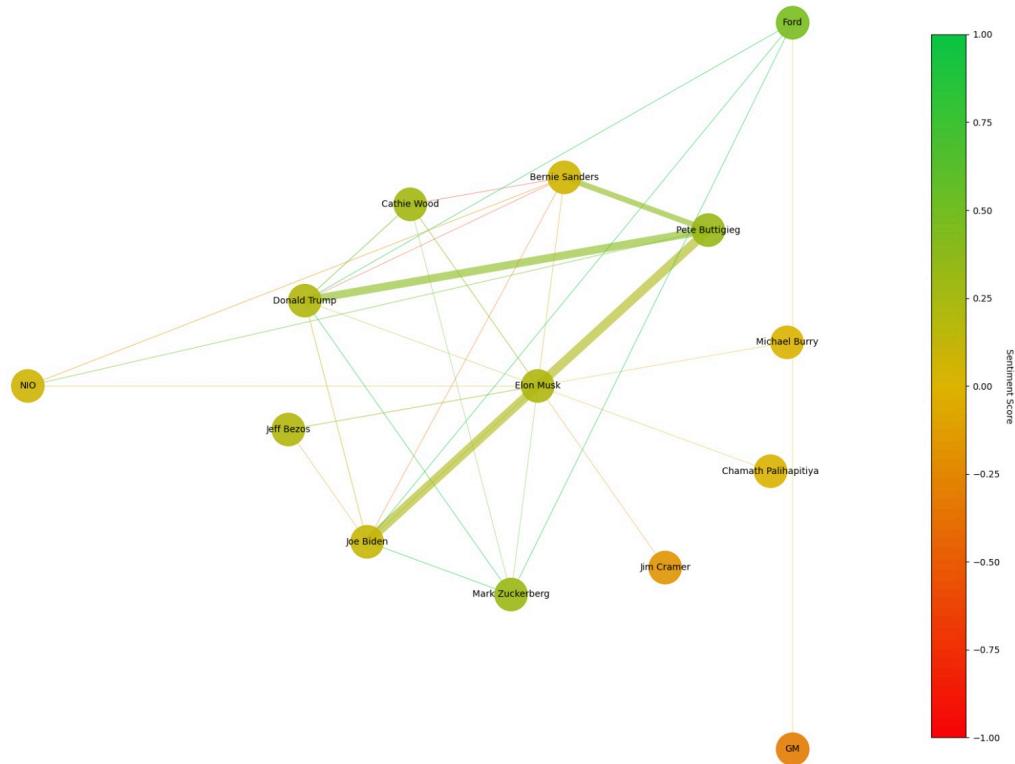
לאחר ניתוח ראשוני של הרשותות, מצאנו כי אילון מסלק מזוואה את הצומת המרכזית והמשפיע ביותר בשיח, אף יותר מהמותג טסלה עצמה. לכן, בהתאם לממצאים שיצגו במדדי המרכזיות, בחרנו להציג את הגרפים כאשר מסלק נמצא במרכז. גישה זו מאפשרת להבין טוב יותר את מבנה הרשותות ואת הקשרים שנרקמים סביבו.



Elon Musk & Influencers for period <2020 (Node colors based on sentiment)

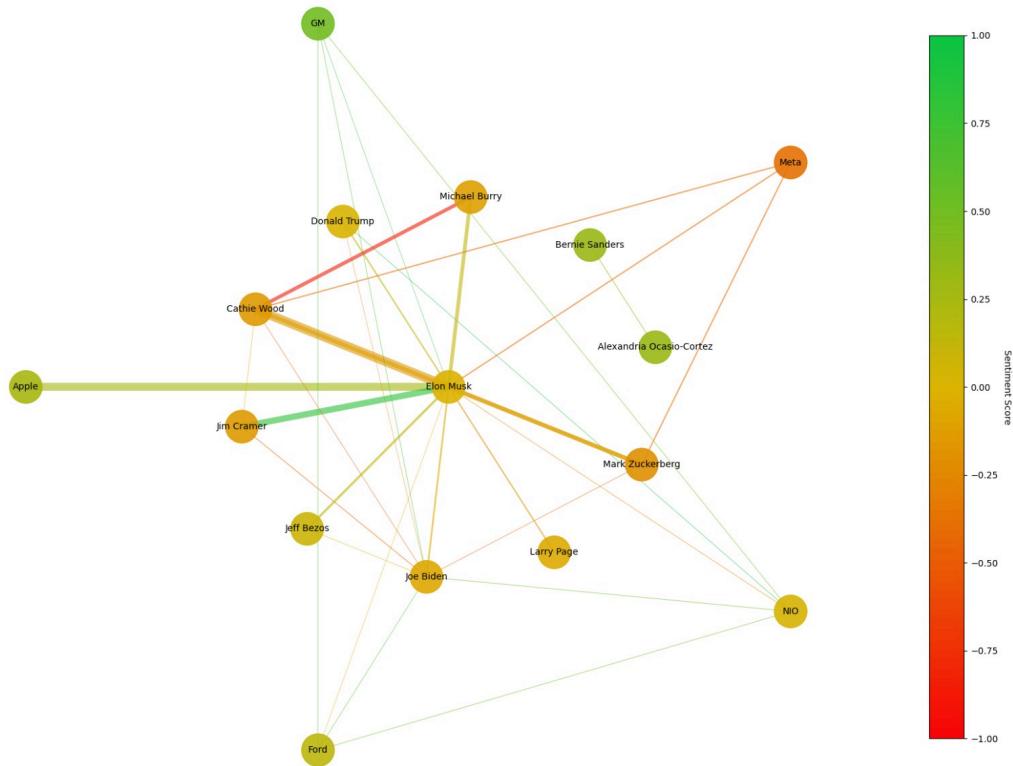


Elon Musk & Influencers for period 2020-2021 (Node colors based on sentiment)

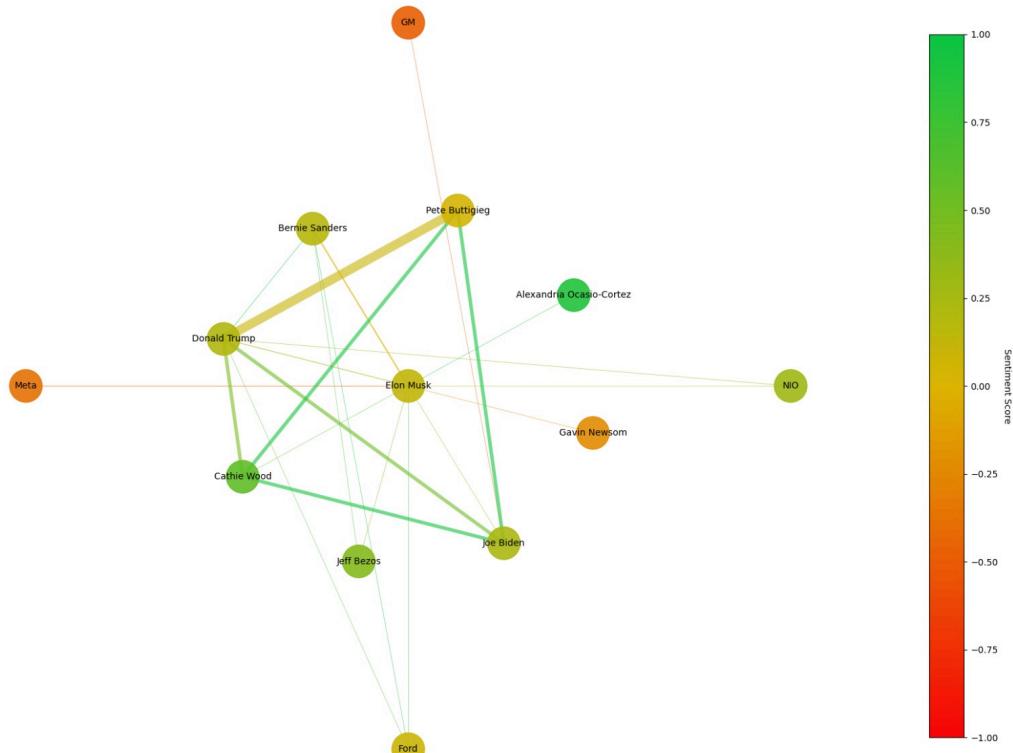




Elon Musk & Influencers for period 2022-2023 (Node colors based on sentiment)



Elon Musk & Influencers for period 2024-2025 (Node colors based on sentiment)





ניתן להבחן וללמוד מהרשאות כמה דברים על שאלות המחקר ובאופן כללי:

- בכל ארבעת הגרפים ניתן לראות באופן ברור את מסיק צומת בעל מספר הקשרים הרב ביותר, עם קשרות עבות המיצגות תדיות הופעה גבוהה. הדבר משקף את מקומו הדומיננטי בשיח למשך זמן, הרבה מעבר לקשר היישר לטלה.
- בגרף שלפני 2020 מופיעים פחות קשרים, אך רבים מהם עבים ומובהקים. החל מ-2020-ל-2021, ניתן לראות עלייה במספר הקשרים אך בעובי מופחת - דבר שיוביל לעמיד על פיזור תשומת הלב בין יותר שחקרים או ירידה בתדיות הופעה משותפת של דמיות בכל קשר.
- בגרף שלפני 2020 בולטים הקשרים החזקים והחויבים בין ג'ו ביידן, دونלד טראמפ וברני סנדרס. שלושתם מקשרים זה לזה בקשרות עבות ובנטימנט ירוק, מה שמرمץ על שיח ציבורי נרחב סבירם, יתכן בהקשר פוליטי-כלכלי, סביב רגולציה או חדשנות.
- הגרף של 2023-2022 מבילוט רשת שבה רוב הקשרים צבעים בגווני כתום-אדום, המעידים על סנטימנט שלילי מובהק. בהשוואה לשאר התקופות, זו התקופה שבה מתקייםיחס שלילי רחוב כלפי הדמויות המרכזיות, מה שעשו לעמיד על שינוי בגישה הציבורית או על אירועים שליליים הקשרים אליהם.
- בגרף של 2023-2022 הקשר בין מסיק לג'ים קרמר בולט כחריג גם בצבע י록 חיובי וגם בעובי, זאת בתוך רשת שבה רוב הקשרים נראים ניטרליים או שליליים. קשר זה עשוי להצביע על אפיוזה תקשורתית מבודדת אויחס אישי וצא דוף.
- ניתן לראות שדמיות מסוימות, כמו دونלד טראמפ ופייט בטיג'ג', מקשרות ביניהן באופן חיובי בתקופה 2020-2021, אך ב-2024-2025 הקשר ביןיהם מקבל גוון צהוב-כתום, המעיד על שינוי מגמה בסנטימנט. מדובר בדוגמה לשינוי נקודתי ביחס הציבור לדמויות, גם כאשר הן שמורות על נוכחות בשיח.

```
def visualize(self):
    fig, ax = plt.subplots(figsize=(18, 14))

    central_node = ["Elon Musk"]
    influencer_nodes = [n for n in self.graph.nodes if n in CONFIG.INDIVIDUALS and n != "Elon Musk"]
    competitor_nodes = [n for n in self.graph.nodes if n in CONFIG.COMPARATIVE_COMPANIES]
    other_nodes = [n for n in self.graph.nodes if n not in central_node + influencer_nodes + competitor_nodes]

    shells = [central_node, influencer_nodes, competitor_nodes + other_nodes]
    pos = nx.shell_layout(self.graph, shells)

    edge_colors = [data['color'] for _, _, data in self.graph.edges(data=True)]
    edge_weights = [data['normalized_weight'] for _, _, data in self.graph.edges(data=True)]
    node_colors = [self._get_node_sentiment_color(node) for node in self.graph.nodes]

    nx.draw_networkx_nodes(self.graph, pos, node_color=node_colors, node_size=1300, alpha=0.9, ax=ax)
    nx.draw_networkx_labels(self.graph, pos, font_size=10, ax=ax)
    nx.draw_networkx_edges(self.graph, pos, edge_color=edge_colors, width=edge_weights, alpha=0.6, ax=ax)

    sm = plt.cm.ScalarMappable(cmap=self.custom_cmap, norm=plt.Normalize(vmin=-1, vmax=1))
    sm.set_array([])
    cbar = plt.colorbar(sm, ax=ax, shrink=0.8, aspect=20)
    cbar.set_label('Sentiment Score', rotation=270, labelpad=15)

    ax.set_title(f"Elon Musk & Influencers for period {self.period} (Node colors based on sentiment)", fontsize=16)
    ax.axis('off')
    plt.tight_layout()
    plt.show()
```



2.3. מטריקות עיקריות

בעבלה הבאה מתוארים המטריקות העיקריות בכל גרפ -

גרף	צמתים	קשנות	CAFIFOT	דרגה ממוצעת	SENTRYMENT ממוצע
2020>	16	29	0.2417	3.62	0.2534
2020-2021	15	35	0.3333	4.67	0.3192
2022-2023	17	42	0.3088	4.94	-0.0263
2024-2025	14	30	0.3297	4.29	0.1707

מהובליה עולה מגמה של עלייה במספר הצמתים והקשנות עד 2022-2023, יחד עם עלייה בדרגה הממוצעת - מה שמעיד על רשת שהופכת צפופה ומקושת יותר לאורך הזמן. בתקופה זו נרשם גם הסנטימנט הממוצע הנמוך ביותר, ואף שלילי, מה שעשו לرمוז על גידול באינטראקטיות ביקורתית או שליליות סביב הנושאים שנדונו, כפי שנראה גם במבנה הגרפ. בנוסף, ניכרת ירידה קלה בקוטר הרשת - ככלומר, מספר הצעדים המקיים לצורך כדי לעבור בין שני צמתים כלשהם קטן. נמצא זה מעיד על כך שהרשת הפכה קצר יותר "קטנה" במוני מרחקים, ויתכן שהדמויות המרכזיות בה הפכו למתחמות ייעילות יותר בין קבוצות שונות.

נקודות חיתוך -

בניטוח הגרפ עבר התקופה שלפני 2020, דחוו שתי נקודות חיתוך - אילון מסק וקתי ווד. מיקומו של אילון מסק כצומת חיתוך אינו מפתיע, שכן הוא מהו מוקד מרכזי בשיבוב טסלה וקשרו למספר רב של דמיות וחברות - היסתו מביאה לפירוק ממשועוט של הרשת. לעומת זאת, קתי ווד נראית פחות מרכזית מבחן מספר הקשרים, אך היא משמשת כצומת ייחודה המחברת בין פיט בוטייג'יג' לרשת. בתקופה זו, פיט עדין לא היה דמות בולטת בשיבוב ציבורי והואזכר בעיקר בהקשרים של מדיניות תחבורה ותמכה בטכנולוגיות י록ות - נושאים שימושיים לאסטרטגיית ההש侃ות של קתי ווד בטסלה. ניתן שזו הסיבה לכך שככל האזכורים שלו הופיעו יחד עמה.

בכל הגրפים האחרים, אילון מסק ממשיך להופיע כצומת חיתוך, ובנוסף מזוהים צמחי חיתוך נוספים כמו Ford, Alexandria Ocasio-Cortez ברשות לצומת אחת בודדת, כך שהרטמת יוצרת ניתוק מבני. עם זאת, לאחר שתפקידם במבנה הרשת פחות מהותי ואינו מצביע על מוקד עניין מובהק, לא נרחב עליהם בניתוח הנוכחי.

גשרים -

בניטוח הגרפים, מרבית הגשרים שנמצאו מקשרים בין אילון מסק לבין צמתים שלילים, שמחוברים לרשת אך דרך דרכו. תופעה זו טبيعית במבנה שבו מסק מהו צומת מרכזית, והקשרים אליו מחוברים בין אזור מסוים היחיד של דמיות מסוימות לשיבוב כללי. בהתאם לכך, רבים מהגשרים תואמים גם את מיקומו כנקודות חיתוך. עם זאת, ניתן לציין גשר אחד - גשר מתחבר המחבר את אילון מסק לאלכסנדריה אוקסיו-קורטז, וממנה לבני סנדרס. גשר זה יוצר מעין מעבר רעוי בין מרכז השיבוב על טסלה לבין קצה שבו מופיעים נושאים של ביקורת חברתית וכלכליות, והוא מדגים כיצד גשר ייחיד יכול לחבר בין תתי-רשתות עם תוכן שונה.



2.4. ממדים מרכזיות

בטבלה הבאה מתוארים ממדדי המרכזיות - Degree, Betweenness, Closeness, של שלושת הזרים
המובילים בכל גף -

Closeness	Betweenness	Degree	גרף
Elon Musk (0.83) Tesla (0.62) Cathie Wood(0.62)	Elon Musk (0.62) Cathie Wood (0.18) Tesla(0.12)	Elon Musk (0.80) Cathie Wood (0.47) Tesla(0.40)	2020>
Elon Musk (0.78) Donald Trump (0.70) Joe Biden(0.70)	Elon Musk (0.47) Ford (0.14) Joe Biden(0.13)	Elon Musk (0.79) Donald Trump (0.57) Joe Biden(0.57)	2020-2021
Elon Musk (0.94) Tesla (0.70) Joe Biden(0.70)	Elon Musk (0.59) Alexandria Ocasio-Cortez (0.12) Tesla(0.07)	Elon Musk (0.94) Tesla (0.62) Joe Biden(0.62)	2022-2023
Elon Musk (0.87) Donald Trump (0.72) Tesla(0.72)	Elon Musk (0.43) Tesla (0.18) Donald Trump(0.15)	Elon Musk (0.85) Donald Trump (0.62) Tesla(0.62)	2024-2025

לאורך כל התקופות, אילון מאסק ממשיך להופיע בעקבות צומת המרכזיות ביותר בכל שלושת ממדדי המרכזיות - מה שמעיד על מעמדו המרכזי והמתמשך בשיח סיבוב טסלה. במיוחד, ניתן לזהות עלייה הדרגתית בנסיבות של ג'ו ביידן, שמתחליה בתקופה של 2020-2021 ונמשכת גם לשנים שלאחר מכן. הדבר משקף את כניסה של איזה הפוליטית הרלוונטי עם מינויו לנשיא, ואת מעורבותו הגדולה בנסיבות הקשורות לרכיבים חשמליים. במקביל, דונלד טראםף תופס מקום בולט בעיקר במדדי קירבה ותיווך בשנים 2020-2025, מה שמעיד על תפקידו כשחקן בעל נראות והשפעה בשיח, צומת שמקשר בין אזורים שונים בראשת.

בנוסף, ניתן לראות את הפרש הגדול בין מאסק לכל השאר, בעיקר במדד Betweenness - פער זה מדגיש את התלות של הרשות בו כצומת מקשר בין קהילות שונות, ומצביר על קר שכםעט כל אינטראקציה מהותית ברשותות עברה דרכו. נמצא זה מהוות נקודת מפנה בהבנת השיח - למרות שניתחנו רשותות שהתבססו על subreddits של טסלה, התברר שמאסק הוא הדמות המרכזית והמשמעותית ביותר בשיח ולעתים אף יותר מהמוחג עצמו. לאור זאת, המשך הנition יתמקד גם באילון מאסק ובפתרונות החברתיים שנרקמים סביבו.



2.5. קהילות

בגרף קהילה היא קבוצה של צמתים שמקושרת יותר חזק בתוכה מאשר כלפי חוץ. זיהוי קהילות מאפשר להבין טוב יותר את מבנה הרשת, לחושף קבוצות של משתתפים עם תחומי עניין משותפים, ולהציג על מבניים חברתיים סטטוטים. אנחנו בחרנו לשימוש באלגוריתם Girvan-Newman שמצויה קהילות על ידי הסירה הדרגתית של קשתות עם ריכוזיות קשותות גבוהה (edge betweenness) מתוך הנחה שקשותות אלו מחברות בין קהילות שונות. בשיטה זו ניתן לעקוב אחרי התפרקות הדרגתית של הגרף ולזהות את המבנה הקהילתי שלו.

בגרף של התקופה שלפני 2020 קיבלנו קהילה שמורכבת מגו' בידן, دونלד טראמפ וברני סנדרס - תוצאה הגיונית לאור המבנה שראינו בגרף, שבו שלוש הדמויות יצרו קבילה מבודדת יחסית עם קשרים חזקים ביןיהם ומעט מאוד קשרים כלפי חוץ. המבנה זהה תואם את האופן שבו האלגוריתם מזיהה קהילות - הקשתות שמחברות את הקבילה לשאר הרשת הן בעלות ריכוזיות גבוהה, ולכן הן אלו שנחlected ראשונות וمبליות את הקבילה כמבנה נפרד בתוך הרשת.

בגרפים של לאחר התקופה של 2020 ניתן לראות שרשת הקשרים בין היישויות צפופה ובעל קשותות רבה יותר בהשוואה לתקופה הקודמת, ורוב הקשרים עוברים דרך אילון מאסק שמשמש כצומת מרכזי ברשת. במבנה זהה, שבו כמעט כל המסלולים הקצרים ביותר עוברים דרך אותו צומת מרכזי, האלגוריתם שմבוסס על ריכוזיות קשותות (edge betweenness) מתקשה לזהות קהילות משמעותיות. הסיבה לכך היא שהקשתות שמחברות את מצוקה לצמתים האחרים יקבעו ערכיו ריכוזיות גבוהים וייכרתו ראשונות, אך פירוק זה לא בהכרח יחשוף מבנה קהילתי מובהק אלא בעיקר משקף את מרכזיותו של מאסק. לכן, בניתוח שלנו לא מצאנו משמעות ברווח או חלוקה משמעותית בקהילות שהאלגוריתם ייצר עבור התקופות הללו.

```
def get_communities(self):
    comp = nx.algorithms.community.girvan_newman(self.graph)
    for _ in range(1):
        communities = next(comp)
    return list(communities)
```



3. עיבוד שפה טבעית

3.1. עיבוד טקסט מקדים

על מנת לאפשר ניתוח מדוקין, אחיד ומשמעותי של השיח ברמה הלשונית והרגשית, ביצענו תהליך עיבוד מקדים לאוסף הטקסטים שנאסף מהפוסטים והתגובהות. מטרת העיבוד הייתה להסיר רעש לשוני, לאחד יצוגים שונים של אותו תוכן, ולשמור מידע רלוונטי לנתח סנסטימנט ומבנה שפה.

תחילה, המרנו כינויים לשמות המקורים של הדמויות והחברות אשר בוצעו על פי מילון ייעודי שהוגדר מראש, כך שכל מופע של שם, גם אם נכתב בצורה לא רשמי, יאותד תחת יצוג אחיד (למשל “Elon”, “Elon Musk”, “Tesla CEO” וכו’). לאחר מכן, הטקסטים פורקו לתוכנים תוך הסירה של מילים עצירה באנגלית, תוך שמירה רק על מילים באורך של שלוש אותיות ומעלה.

כדי לשמר את הממד הרגשי של השיח, בחרנו שלא להתעלם מאיםוגים אלא להמיר אותם לתיאורים טקסטואליים שימושיים, כך שהאינפורמציה הרגשית תישמר ותוכל להופיע על נתחם הסנטימנט. בנוסף, כל מילה עברה תהליכי Lemmatization, החזרה לצורת השורש הדקדוקית שלה. תהליכי זה מאפשר לצמצם וריאציות לשונות של אותה מילה (למשל הטוויות של פעלים), ולבסס אוסף אחיד ומובנה יותר מבחינה לשונית.

כמו כן הורדנו דברים שלא תורמים ואף מפריעים לנתח השפה, כמו כתובות אתרים, תווים בגיבריש, סימני קיראה ושאלה, ורווחים מיוחדים.

```

def normalize_text(self, text):
    text = emoji.demojize(text, delimiters=( "", " "))
    text = re.sub(r'\b[\w\.-]+@[^\w\.-]+\.\w+\b', ' ', text)
    text = re.sub(r'https?://\S+|www\.\S+', ' ', text)
    text = unicodedata.normalize('NFKD', text).encode('ASCII', 'ignore').decode('utf-8')
    text = re.sub(r'[\x00-\x1f\x7f-\x9f]', ' ', text)
    text = re.sub(r'[\^\\w\\s\\.,!?\\"\\-]', ' ', text)
    text = re.sub(r'\s+', ' ', text).strip()
    doc = nlp(text)
    text = ' '.join([token.lemma_ for token in doc])
    return text

def _replace_aliases(self, text):
    if not text:
        return ""
    text_lower = text.lower()
    for alias, canonical in self.alias_lookup.items():
        canonical_token = canonical.replace(" ", "_")
        pattern = re.compile(rf'\b{re.escape(alias)}\b', re.IGNORECASE)
        text_lower = pattern.sub(canonical_token, text_lower)
    return text_lower

def _tokenize(self, text):
    if not text:
        return []
    tokens = re.findall(r'\b[a-z_]+\b', text.lower())
    tokens = [token for token in tokens if token not in self.stop_words and len(token) > 2]
    ngram_tokens = []
    for n in range(self.ngram_range[0], self.ngram_range[1] + 1):
        ngrams = zip(*[tokens[i:] for i in range(n)])
        ngram_tokens.extend(['_'.join(ngram) for ngram in ngrams])
    return ngram_tokens

```



```

def process_all_text(self):
    word_scores = defaultdict(list)
    seen_contexts = set()

    for post in self.posts:
        if post.get('title'):
            sentiment = post.get('sentiment_score', 0)
            title_text = self._replace_aliases(post['title'])
            tokenized = self._tokenize(title_text)
            for word in set(tokenized):
                self.word_counts[word] += tokenized.count(word)
                word_scores[word].append(sentiment)
                key = (word, post['title'])
                if key not in seen_contexts:
                    self.word_contexts[word].append((post['title'], sentiment))
                    seen_contexts.add(key)

        for comment in post.get('comments', []):
            if comment.get('body'):
                sentiment = comment.get('sentiment_score', 0)
                body_text = self._replace_aliases(comment['body'])
                tokenized = self._tokenize(body_text)
                for word in set(tokenized):
                    self.word_counts[word] += tokenized.count(word)
                    word_scores[word].append(sentiment)
                    key = (word, comment['body'])
                    if key not in seen_contexts:
                        self.word_contexts[word].append((comment['body'], sentiment))
                        seen_contexts.add(key)

    self.averaged_scores = {
        word: np.mean(scores) for word, scores in word_scores.items()
    }
}

```



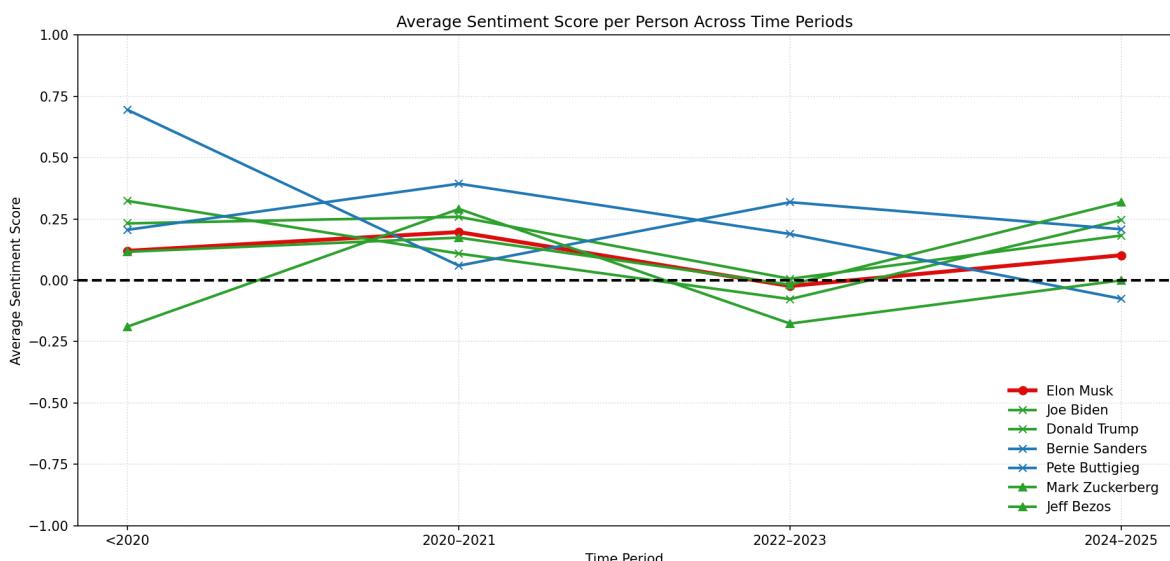
3.2. ניתוח סנטימנט

ניתוח סנטימנט הוא חלק חשוב בעבודה שלנו, ושילב בין הבנת התוכן הרגשי של השיח לבין מבנה הקשרים ברשות. לאחר וההתקדמות שלנו הייתה לא רק בדמויות ובחברות שומות ייחד, אלא גם באופי שבו הן מוזכרות, חיובי, שלילי או נייטרלי, החלנו לשלב את הסנטימנט ישירות בגרפים. דבר זה התרטט במצב של הצמתים (לפי הסנטימנט כלפי הדמות או החברה) ובצבע של הקשתות (לפי הסנטימנט המוצע של האזוריים המשותפים בין שני צמתים), כפי שפירטנו בחלק הקודם הקודם של העבודה.

לצורך כך השתמשנו בספרייה `vaderSentimentIntensityAnalyzer`, שמתאמת במיוחד לעקסטים קצרים ובלתי רשמיים כמו פוסטים ותגובה ב-Reddit. עבור כל טקסט חושב ציון בין -1 ל-1, ובגענו ספר לפיו טקסט עם ערך מעל 0.07 נחשב חיובי, מתחת ל-0.07 – נחשב שלילי, ובתווך – נייטרלי. הספרים האלה נבחרו לפי ניסוי ובדיקה כדי להגיע לשאלה שנראתה לנו מאוזנת ואמינה.

מעבר לשילובו הויזואלי בגרפים, ניתוח הסנטימנט פותח פתח להבנה טובה יותר של האופן שבו דמויות והחברות נתפסות בשיח. הוא מאפשר לבחון את האווירה הרגשית הכללית, לזהות הבדלים בטון בין גורמים שונים, ולעקב אחר שינויים בהקשר ובגישה לאורך תקופות שונות. נציג כעת כמה נקודות שעלו מתוך ניתוח הסנטימנט –

- סנטימנט ממוצע של דמויות מרכזיות לפי תקופות:



הגרף מציג את ממוצע הסנטימנט עבור חלק מהדמויות המרכזיות שנבחנו לאורך ארבעת תקופות הזמן. לצורך בהירות, חילקו את הדמויות לשתי צורות ייצוג: צורת הסמן (marker) מייצגת את סוג הדמות – אילון מאסק מסוכן בעיגול, פוליטיקאים (ג'ו ביידן, دونלד טראמפ, ברני סנדרס ופיט בוטיג'ג') מסומנים באיקס, ואנשי טכנולוגיה (ג'פ בוזס ומארק צוקרברג) מייצגים באמצעות משולש. בנוסף, צבעי הקווים נבחרו בהתאם לחשיבותם הסנטימנט דומות לאורך התקופות, כפי שנמצא בנתונים. ניתן לראות כי לרוב הדמויות ברוב תקופות הזמן הסנטימנט המוצע הוא חיובי חלש. מבחינה תוכנית, הגרף מאפשר לבחון כיצד הסנטימנט סביב דמויות שונות משתנה לאורך זמן. למשל, בתקופה של 2023-2022 ניכרת ירידת חדה בסנטימנט אצל מרבית הדמויות, תוצאה שמתווספת בשיתות היבט עם הממצאים הקודמים מניתוח הגרפים, בהם התקופה הזאת בלטה כנקודת שפל רגשית בשיח. כמו כן, ניתן לראות שקיימת קרבה בין סנטימנט של דמויות שונות – מאסק מציג מגמה



דומה לו של בידן וטראמפ (ירידה חדה ולאחריה התאוששות), אך גם תנודות אחרות שמאפיינית את בזוס צוקרברג.

מאסק מהו דוגמה מובהקת לדמות רגשיות מורכבת - לא למגרי פוליטיקאי אך גם לא רק טכנולוג, ולפיכך חוצה גבולות בין קבוצות שונות, גם מבחינת תפקידו ברשות וגם לפי הדריך בה מדברים עלי).

- הקשר בין אילון מאסק וג'ים קרמר בתקופה 2022-2023:
כפי שראינו, הגраф של תקופה 2022-2023 מתאפיין בסנטימנט שלילי במיוחד. יחד עם זאת, בולטת קשת יrokeה חיובית בין אילון מאסק לג'ים קרמר. ניתוח איקונטני העלה כי היו ביניהם חמש אינטראקציות ושירות בתקופה זו, שבהן חזרה המילה "אופטימי" שלוש פעמים. למילה זו סנטימנט של 0.32, והיא שתרמה באופן משמעותי לחיבובי של הקשר - מה שמעיד לכך מונח רגשי בודד יכול להשפיע על אופיו הכללי של שיח בין דמויות.
- הקשר בין ג'ו בידן, دونלד טראםפ, וברני סנדרס בתקופה שלפני 2020:
שלישיה זו זההתה כקליקה וקילילה גם בניתוח הרשות, אשר ניתן לראות שיש בהם קשרים חזקים וחיבוביים. ביצענו ניתוח לתוך כדי לבדוק את הסנטימנט שהתקבל מהרשות - מילים כמו, (win 0.58) (great 0.62), (support 0.4).
יתכן שהשיכח שבבים, למרות הבדלים אידיאולוגיים, התמקד בתקופה זו בעיקר בנושאים כלליים כמו תקווה לשינוי או הערכה למעורבות הפוליטית - וכן נשמעו חיבובי יותר. יתכן גם שהשיכח היה מרכיב או ביקורת יותר, אך הפרוק למילים בודדות בעלות מטען רגשי חיבובי הראה את תוצאה ניתוח הסנטימנט והציג את הקשר חיובי, גם אם ההקשר הכללי היה שונה (למשל, סרקסטי, אירוני או אמביוולנטי).

```
def _analyze_sentiment(self, text):
    score = self.sia.polarity_scores(text)[ 'compound' ]
    threshold = CONFIG.SENTIMENT_THRESHOLD

    if score >= threshold:
        sentiment = CONFIG.SENTIMENT_LABELS.get('positive', 'positive')
    elif score <= -threshold:
        sentiment = CONFIG.SENTIMENT_LABELS.get('negative', 'negative')
    else:
        sentiment = CONFIG.SENTIMENT_LABELS.get('neutral', 'neutral')

    return score, sentiment
```



3.3. מילוט מפתח

אפשרות מפתח זה כל' חשוב להבנת מוקדי השיח בראשת ויזיהו הנושאים והערכיהם שמייחדים קהילה או דיוון מסוים. שימוש במדד כמו IDF-TF שבודק עד כמה מילה נפוצה במסמך מסוים אך נדירה במסמכים אחרים, מאפשר לחושף דפוסי שיח, רעיונותבולטים והקשרים מרכזים בין דמיות ונושאים.

כדי להבini מהם הנושאים שעמדו במקור השיח סיבי דמיות מרכזיות בתקופה שלפני 2020, בחרנו להתמקד בקבוצה שהייתה קliquה בולטת ובעל קשרים חיוביים במיוחד בראשת - דונלד טראמפ, ג'ו ביידן וברני סנדרס. לקחנו את כל הופעותיהן והתגבותן שבמה הזכרוCMDI הדמיות האלו בתקופה הנ"ל, וביצענו עליהם ניתוח מילוט מפתח באמצעות מודד-IDF TF שבודק אילו מילים ייחודיות וחשובות מאפיינות את השיח במסמכים האלו. כדי למקדד את הניתוח בנושאים עצם ולא באנשים, הסרנו את שמות הדמיות מהשיח באמצעות stop words מותאמת אישית.

מהותצאות אפשר להבין שהשיח סביר הדמיות האלו עסוק לא רק בפוליטיקה עצמה, אלא גם בשאלות רחבות יותר של מוסר, חברה וכלכלה. מהניתוח אפשר לזהות את המילויים והנושאים שהפכו את השיח עליהם ליחסוני ומשונה מהשיח הכללי בראשת וראה מכם המודדים המרכזים שעלו בדיונים טריים.

חשוב לציין שהתייחסנו בכל הנি�וח לכל פופט או תגובה כ"מסמר", וכך התוצאות משקפות בעיר את הנושאים המרכזיים שעליהם בפוטרים עצם ולא בהכרח את המורכבות המלאה של השית. יחד עם זאת, הניות מאפשר לחתות את הרעיונות והמושגים הבודדים שאפיינו את הדינזטים סביר הדמיות הללו.

```
jim_elon_2022_2023 = find_co_mentions(processed_posts, "Jim Cramer", "Elon Musk", aliases=CONFIG.ALIASES, period="2022-2023")
biden_senders_before_2020 = find_co_mentions(processed_posts, "Joe Biden", "Bernie Sanders", aliases=CONFIG.ALIASES, period=<2020")
trump_sanders_before_2020 = find_co_mentions(processed_posts, "Donald Trump", "Bernie Sanders", aliases=CONFIG.ALIASES, period=<2020")
trump_biden_before_2020 = find_co_mentions(processed_posts, "Donald Trump", "Joe Biden", aliases=CONFIG.ALIASES, period=<2020")

threesome = [biden_senders_before_2020, trump_sanders_before_2020, trump_biden_before_2020, trump_pete_20_21, trump_pete_24_25] # 20 sentences

def get_top_tfidf_words(documents: list[str], top_n=20):
    vectorizer = TfidfVectorizer(
        stop_words=list(stopwords) + ["elon", "musk", "jim", "cramer", "biden", "sanderson", "trump", "bernie"],
        ngram_range=(1,1),
        token_pattern=r'\b[a-z_]{3,}\b',
        lowercase=True
    )
    tfidf_matrix = vectorizer.fit_transform(documents)
    mean_scores = np.asarray(tfidf_matrix.mean(axis=0)).flatten()
    tfidf_scores = dict(zip(vectorizer.get_feature_names_out(), mean_scores))
    top_tfidf = sorted(tfidf_scores.items(), key=lambda x: x[1], reverse=True)[:top_n]
    return top_tfidf
```



3.4. ניתוח נושאי

בניתוח נושאי המטרה הייתה לזהות את הנושאים המרכזיים שמרכזים את השיח בטקסטים ולנסות להבין מהם הריעונות המרכזיים שחוורים על עצם בזורה עקבית. לשם כך השתמשנו באלגוריתם (Latent Dirichlet Allocation) שמצויה קבוצות מילים שונות להופיע יחד במסמכים ומגדיר אותן כנושאים סמיוטיים. את האלגוריתם הפעילו על הטקסטים לאחר שעברו עיבוד מקדים, כך שהמילים שנשארו משקפות טוב יותר את התוכן המקורי של השיח. התוצאה שהתקבלה היא רשימה של נושאים, כשל נושא מתואר באמצעות אוסף מילים מרכזיים עם משקלים מסוימים שמייצגים עד כמה כל מילה מאפיינת את אותו נושא - ככל שהציוון גבוה יותר, כך המילה בולעת ומשמעותית יותר לנושא ביחס למילים אחרות. רשימת הנושאים מופיעעה [בנספח](#).

אילו מספק היה דמות מרכזית לאורך כל התקופות, אך ההקשר שבו הופיע השינה בזורה חזה. לפני 2020 השיח סבבו התמקד ביזמות וחדשנות, כפי שנראה בנושאים המרכזיים של התקופה שעיסקו בדמויות מובילות ובטכנולוגיות מתקדמות (למשל ב-1 Topic 0, Topic 1-2).

ב-2020-2021 השיח הפר מרכיב יותר, כאשר מספק הופיע בנושאים ששילבו שאלות מוסריות-חברתיות וככללות, כמו נושאים שעסקו בmundi כלכלי, מיסוי ואחריות ציבורית (למשל ב-1 Topic 1-2). ב-2022-2023 וב-2024-2025 השיח סבבו השתלב בנושאים בעלי אופי פוליטי ופוליטי, שהתחמקדו ב זהויות פוליטיות, עימותים ציבוריים ושיח תרבותי מוקוב (כמו בנושאים שעסקו ב涅גדים בין ימין לשמאלי, חופש ביטוי וחלוקת ציבורית - 4 Topic 0, Topic 2, Topic 4 של אותן התקופות).

כמו שראינו בניתוח הסנטימנט את השני באופי השיח, משיכ חיים ומעירך שהפר עם הזמן לשלי וטענו יותר, גם בניתוח הנושאי אנו רואים מגמה דומה. השיח המוקדם התמקד בחדשנות, טכנולוגיה ודמויות ידניות בהקשרים חיוביים ומעוררי השראה, אך בהמשך התקופות הוא הפר לעסוק יותר ויתר בשאלות חברתיות, כלכליות ופוליטיות. הנושאים הפכו טעוניים ומוסוגים, והדמויות המרכזיות נתפסו יותר ויותר כדמות של שמעוררות מחלוקת ופחות כSAMPLEים לחידשות. השיח התרחק מהתחמקדות בטכנולוגיה ועבר לזרות של עימותים ציבוריים וביקורת חברתית.

הניתוח הנושאי חושף הבדל عمוק לא רק באופי השיח אלא גם במבנה שלו. התקופה שלפני 2020 התאפיינה בשיח שבו נושאים חזו סביר ערכיהם של חדשנות, יזמות והערכתה לדמויות מובילות. לעומת זאת, בתקופות המאוחרות, ובמיוחד 2022-2023 ו-2024-2025, השיח הפר למבוזר וdispersed יותר, עם נושאים שלא פעם חופפים או מתערבבים בין פוליטיקה, כללה, תרבות ושיח קהילתי פנימי. המגנון הזה מצביע על אובדן מוקד ברור בשיח, וייתכן שהוא משקף את המורכבות של השיח הציבורי בעידן שבו הגבולות בין טכנולוגיה, פוליטיקה וחברה היטשטשו.

בדיעבד, בהסתכלות רחבה יותר, אפשר להבין את ההבדלים האלה על רקע אירועים חיצוניים שה השפיעו על התודעה הציבורית - כນיסוחן של הדמויות המרכזיות (ובעיקר אילו מספק) לזרה הציבורית והפוליטית באופן פעיל, עלית הדין סביר חופש ביטוי ושליטה בתקשורת החברתית, והמתוך הגובר סביר עצמת הahn והתשפעה על החברה. אלו לא רק שינוי את הדימוי של הדמויות, אלא גם את הדרך שבה הקהילה עצמה בחרה לנוהל את הדין.



3.5. ענן מילימ

ען מילים הוא קל ויזואלי בניתו שפה, המציג את המילים השכיחות ביותר במאגר טקסט בגודל יחסית לשכיחותן - ככל שטילה מופיעה יותר, היא תציג גודלה יותר. ען המילים מאפשר לzechot ב מהירות את הנושאים המרכזיים בשיח, דמיות בולטות ומונחים טעונים רגשית. בפרויקט זה שילבנו צבעים לפי ניתוח הסנטימנט שביצענו, וכן קיבלנו תמונה אינטואיטיבית וברורה של השיח. תרומתו של ען המילים היא בהמחשת השינויים בשיח לאורך הזמן, דהיינו מוקדים של עניין או מחלוקת, והציגת המגמות בצורה חזותית גם ללא ניתוח סטטיסטי עמוק.

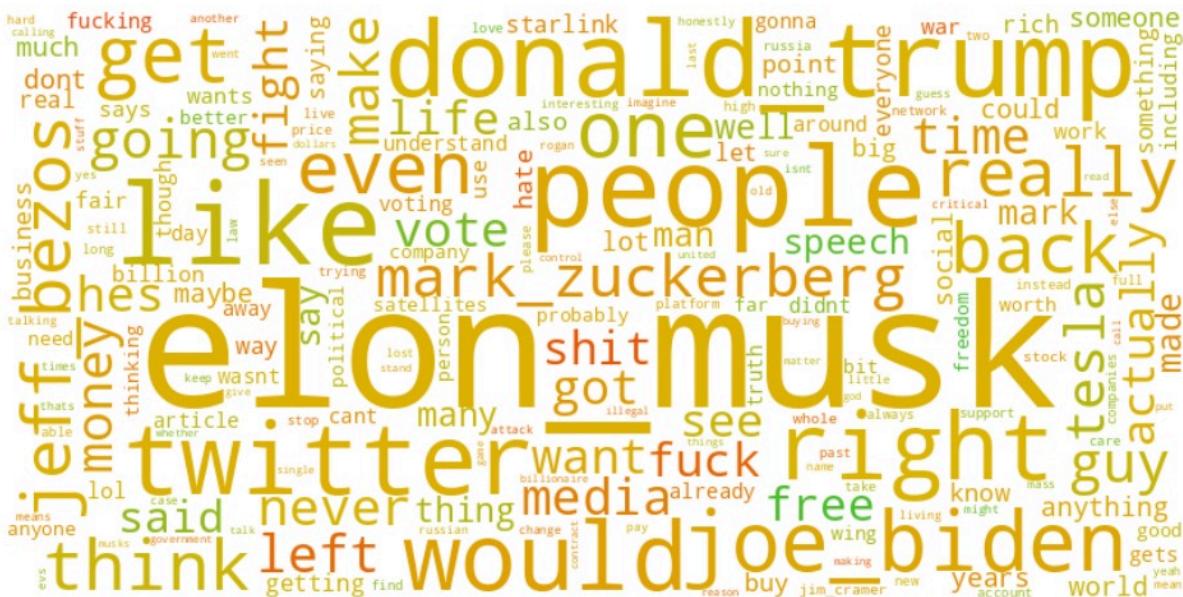
Word Cloud Colored by Sentiment for <2020



בען המילים הראשון (בתקופה שלפני 2020) בולטים במיוחד השמות Elon Musk, Jeff Bezos, Donald Trump ו-Mark Zuckerberg. מה שמעיד שהשיח התמקד בעיקר בדמותות מובילות בעולם הטכנולוגיה והעסקים. לצדDEM מופיעות מילים כלליות כמו get, think, like, would, people, think שמרמזות על שיח פתוח וمبוסס הbhut דעתות והשואות. נושאים טכנולוגיים כמו space, force, technology ועוד, מה שמעיד שהם היו חלק מהשיח אך לא במרקז. ניתן לראות לפי הצבעים שהסentiiment הוא חיובי-ניטרילי ברובו. ניתן להגיד שעולה תמונה של שיח שהתמקד ביזמות, טכנולוגיה והערכתה להישגים של דמויות כמו מאסק ובוזס, שנראו איזבעיקר כחדשנים ומובילי עסקים. הpolicy היה נוכח, אך לא במוקד, והשיח כולל שידר עניין, סקרנות וגישה חיובית יותר כלפי הדמויות והנושאים שעליהם לדין.



Word Cloud Colored by Sentiment for 2022-2023



בענין המילוי השני (בתקופה 2023-2022) בולטים לעין כמה שינויים מרכזים בשיח. הטען הכללי של הדיונים הפך לשלייל ווענן הרבה יותר, כפי שניתן לראותו בצלביהם הדומיננטיים של המילויים - גוננים של כתום ואדום שמצביעים על נעס, תפסול וביקורת. הדמויות המרכזיות, ובראשן Elon Musk, ממשיכות לתרוף מקום מרכזי בענין, אך השיח סבירו השתנה בצורה ברורה - מאסק כבר לא נתפס רק כיזם טכנולוגי, אלא כדמות שמעוררת מחולקות עמוקות סביב נושאים כמו חופש ביוטי, פוליטיקה וזהות.

מילים שהוזכרו מהתקופה הקודמת כמו would, people, like מופיעות כתעת בצלביהם של לילים יותר, מה שמייד על שינוי בಗוון השיח - פחתה הערכאה וסקרנות, יותר ביקורת, לעג או ציניות. יתרון זהה דבר נובע מהקשרי הדיוון החדשניים, בהם משתמשים במילים האלו בהקשרים של עימותים, האשמות והשואות טענות כלפי דמויות ונושאים. לצד זה בולטות עליה ברורה במילים בוטות כמו fuck, shit, fight שמחזקות את התחשוה של שיח לוחמני וקשה. המילים twitter, media בולטות גם הן וממחישות את התפקיד של הפלטפורמות הללו לא רק כמקומות להפצת רעיונות, אלא גם ככר פורה לשיח מקובל, כועס ולעתים פוגעני.

הניתוח באמצעות ענן מילימ חישב בצורה ברורה את המעבר של השיח בקהילה מדין טכנולוגיית סקרני וחיבוי לשיח טעון, ביקורתית ומפוגג סביב סוגיות פוליטיות וחברתיות. ראיינו זאת דרך המילימ המרכזיות שהתגלו בכל תקופה, ודרכן האופן שבו ניתוח המילימ והציגן ענן מילימ אפשרו לזהות את השינויים בדמותם, בנושאים ובطن השיח שהשתנו לאורך הזמן.



4. עיבוד שפה ודינמיקה חברתית

השילוב בין ניתוח רשותות לנתח שפה מאפשר הבנה עמוקה יותר של הדינמיקה החברתית בשיח. בעוד שניתוח הרשותות מציג את הקשרים בצורה ייזואלית (מי מקשר למי, כמה חזק הקשר, והאם ניתן לזהות קהילות או קלייקות) ניתוח השפה מאפשר לנו לחדר לעומק התוכן ולבחון את האופי ומה עומד מאחוריו אוטם קשרים. כך למשל, גם כאשר שני גורמים מוקשרים בצורה חזקה בראשת, רק ניתוח השפה יכול לחושר האם הקשר ביןיהם הוא חיובי, שלילי, סרקיסטי או אמביוולנטי. כאמור, הרשות מציגה את השילד של הקשרים, ואילו השפה מוסיפה את הרבדים הרגשיים, הערכיים והמשמעותיים שמרכיבים את אותם קשרים. לדוגמה, ראינו שאלון מסלק נמצא במרכז הרשות ומקשר לרבים, אך רק ניתוח השפה והנושאים חשף שהמרכזיות שלו לא נובעת רק מכך שהוא דמות מובילה, אלא גם מכך שהוא עומד בלב השיח הציבורי סביב מחולקות, דיבונים פוליטיים וביקורת חברתית.

בנוסף, החיבור בין שני התחומים מאפשר להיותות תופעות שלא ניתן לראות בכל אחד מהם בנפרד. ברשותות דיהינו קבוצות של דמויות שנראות מ הקשורות זו לזו, אך רק דרך ניתוח השפה והסנטימנט ניתן להבין את טיב הקשרים הללו - האם מדובר בהערכה הדדית, שיח פוליטי טעון, ביקורת, או אולי אזכורים חיוביים מפתיעים בין יריבים. השפה מאפשרת לנו להבין איך הסנטימנט בא לידי ביטוי בשיח עצמו, ולא רק במבנה הרשות. היא חשפה כיצד קשרים שנראו חזקים בראש מლויים לעיתים בשיח טעון שהופך להיות יותר ביקורת ופוליטי ככל שהתקפות מתקדמות. כך עיבוד השפה מאפשר לנו להעמיק בתמונה שהרשות סיפקה, לחזק את הממצאים ולזהות היבטים רגשיים ומשמעותיים שלא היו נראים במבנה הקשרים בלבד.

עיבוד השפה מהוות כלי משלים שמחזק את ממצאי ניתוח הרשות - הוא משתמש בתובנות קיימות, מצירחריגות מעניינות, ולייעיתם אף סותר את המסקנה הראשונית ומציע פרשנות חלופית. השילוב הזה תורם להבנה מלאה של הרשות ושל הקשרים, כאשר אם ננתח רק בדרך אחת את הקשרים, נקבל תמונה חלקית בלבד.



נספחים

נספח 1 - פירוט רשימת הדמויות

שם הדמות	תיאור
אלון מאסק	מנכ"ל טסלה. מייסד, אדריכל מוצר, ומקבל החלטות המרכזי בטסלה.
ג'ף בזוס	מייסד אמזון. יריב עסקי וחלי של מאסק (Blue Origin & SpaceX) ומשווה לעתים קרובות למאסק.
دونלד טראמפ ג'ו ביידן	נשיא ארה"ב. משפיע על דעת קהל ומדיניות; התייחס לטסלה ולמאסק. נשיא ארה"ב לשעבר. בתחילת התעלם מטסלה בדיוני רכב חשמלי, בהמשך הכיר בהבולטה בתחום.
לארי פיג'	מייסדי גугл. חבר ותיק של מאסק. Waymo של גугл מתחרה בטסלה בתחום נהיגה אוטונומית.
مارك צוקרברג	מנכ"ל מטא. מתחרה במאסק בתחום הרשותות החברתיות; הביע עמדות בנוגע לטסלה.
פיט בוטיג'gi	שר התעשייה של ארה"ב. דיבר על תרומות טסלה לaimoz רכבים חשמליים ותשתיות.
קתי ווד	מנכ"לית ARK Invest. משקיעה גדולה בטסלה, צופה לטסלה שווי שוק גבוה מאוד.
ג'ים קריימר	מנחה ב CNBC-שינה עמדות כלפי טסלה; ביום תומך אך שני בחלוקת בקשר להילת טסלה.
מייקל ברוי	מכור מ"ה מכירה הגדולה". הימר נגד טסלה; סקפטו לגבי ערכה.
גווין ניוסום	מושל קליפורניה. מדינתו היא בסיס טסלה; התייחס אליה בתקורת.
אלכסנדריה פרוגרסיבית	חברת קונגרס. ביקרה את מאסק בנושאים חברותיים; מייצגת דמוקרפיה אוקסי-קורטז
צ'יאמאת פלייפיטיה	משכיע ויזם. תומך פומבי בטסלה ובמאסק; מקדם את טסלה בתקורת.
ברני סנדס	סנאטור. מבקר את מאסק בשל עשרו והשפעתו; מייצג גישה פרוגרסיבית לאו-שוון.



נספח 2 - פירוט רשימת החברות המתחרות

שם החברה	תיאור
Ford	יצרנית רכב אמריקאית ותיקה. נכננה חזק בתחום הרכיב החשמלי עם דגמים כמו Mustang Mach-E, F-150 Lightning
GM	יצרנית הרכב הגדולה בארה"ב. מתכננת מעבר לרכבים חשמליים בלבד עד 2035, עם מותגים כמו Cadillac, Chevrolet.
BYD	יצרנית סינית מהגדולות בעולם בתחום הרכיב החשמלי. פעילה גם בתחום תחבורה ציבורית יroke.
Lucid	יצרנית אמריקאית של רכבי יוקרה חשמליים, עם ביצועים גבוהים וטוווח נסיעה מרשים. נוסדה ע"י יוצאי Tesla.
Rivian	יצרנית אמריקאית חדשה בתחום רכבים חשמליים, בדגש על טנדרים ורכבי שטח.
NIO	יצרנית רכבים חשמליים סינית, מציעה גם תחנות להחלפת סוללה. נחשבת מתחילה מבטיחה לטסלה בשוק הסיני.
Apple	ענקית טכנולוגית. מנהלת פרויקט חשאי לרכב חכם ("Apple Car") עדין לא יצא עם מוצר בפועל.
Meta	חברת האם של פייסבוק. לא פעילה ישירות בתחום הרכיב, אך מתמודדת מול מאסק בזירת הבינה המלאכותית והרשאות החברתיות.
palantir	חברת תוכנה לניטוח נתונים ובייה מלאכותית. מושרת לשיח טכנולוגי מתקדם ולעיתים מוזכרת לצד טסלה ומאסק.



נספח 3 - רשימת המילים TF-IDF

abortion: 0.1543	people: 0.1018	want: 0.0817
read: 0.1331	like: 0.0940	vote: 0.0792
help: 0.1230	support: 0.0918	focus: 0.0790
medium: 0.1128	need: 0.0892	president: 0.0749
life: 0.1115	course: 0.0852	come: 0.0713
chain: 0.1078	great: 0.0852	work: 0.0713
create: 0.1078	innovation: 0.0852	

נספח 4 - ניתוח נושא

תקופה שלפני: 2020

Topic 0 = 0.880 * "elon_musk" + 0.249 * "elon_musk_elon_musk" + 0.136 * "jeff_bezos" + 0.122 * "like" + 0.107 * "donald_trump" + 0.088 * "would" + 0.078 * "think" + 0.076 * "people"

Topic 1 = 0.759 * "force" + 0.238 * "gravity" + 0.232 * "together" + 0.232 * "associated" + 0.186 * "known" + 0.185 * "holds" + 0.142 * "called" + 0.140 * "electric"

Topic 2 = 0.455 * "like" + 0.324 * "would" + 0.250 * "people" + -0.238 * "elon_musk_elon_musk" + -0.205 * "jeff_bezos" + -0.172 * "elon_musk" + 0.162 * "think" + 0.133 * "charities"

Topic 3 = -0.723 * "jeff_bezos" + -0.305 * "jeff_bezos_jeff_bezos" + -0.175 * "like" + -0.167 * "see" + 0.152 * "elon_musk" + -0.133 * "would" + -0.131 * "high" + -0.120 * "carbon"

Topic 4 = 0.285 * "jeff_bezos" + -0.272 * "donald_trump" + 0.254 * "people" + -0.213 * "solar" + 0.175 * "democracy" + -0.173 * "carbon" + -0.173 * "emissions" + 0.172 * "think"

:2020-2021

Topic 0 = 0.345 * "abortion" + 0.232 * "elon_musk" + 0.207 * "lives" + 0.196 * "life" + 0.164 * "one" + 0.158 * "donald_trump" + 0.157 * "say" + 0.152 * "joe_biden"

Topic 1 = -0.371 * "elon_musk" + -0.206 * "tax" + -0.189 * "pay" + -0.187 * "money" + 0.170 * "please" + -0.156 * "billion" + -0.148 * "year" + 0.146 * "abortion"

Topic 2 = -0.284 * "elon_musk" + -0.202 * "please" + 0.201 * "abortion" + -0.153 * "tax" + -0.137 * "pay" + -0.134 * "money" + -0.133 * "subreddit" + -0.133 * "questions"



Topic 3 = 0.650 * "elon_musk" + 0.223 * "elon_musk_elon_musk" + 0.207 * "jeff_bezos" + -0.204 * "pay" + -0.191 * "money" + -0.173 * "tax" + -0.149 * "year" + -0.137 * "billion"

Topic 4 = 0.269 * "like" + -0.225 * "elon_musk" + 0.202 * "would" + 0.191 * "tax" + 0.185 * "people" + 0.182 * "get" + 0.146 * "good" + -0.127 * "abortion"

:2022-2023

Topic 0 = -0.804 * "elon_musk" + -0.200 * "donald_trump" + -0.180 * "elon_musk_elon_musk" + -0.150 * "twitter" + -0.128 * "like" + -0.117 * "would" + -0.112 * "people" + -0.110 * "vote"

Topic 1 = 0.286 * "like" + -0.243 * "elon_musk" + 0.210 * "left" + 0.146 * "years" + 0.142 * "past" + 0.138 * "constantly" + 0.123 * "back" + 0.106 * "made"

Topic 2 = -0.307 * "people" + -0.284 * "right" + 0.216 * "elon_musk" + -0.195 * "stock" + -0.188 * "like" + -0.155 * "think" + -0.153 * "buy" + -0.132 * "say"

Topic 3 = 0.361 * "jeff_bezos" + -0.318 * "donald_trump" + 0.218 * "contract" + 0.212 * "board" + 0.211 * "lawsuit" + -0.177 * "vote" + 0.173 * "company" + 0.168 * "stock"

Topic 4 = 0.317 * "right" + -0.283 * "stock" + 0.216 * "left" + -0.213 * "buy" + -0.187 * "back" + -0.161 * "pressure" + -0.142 * "vote" + 0.127 * "people"

:2024-2025

Topic 0 = 0.207 * "every" + 0.166 * "year" + 0.164 * "days" + 0.164 * "platoon" + 0.164 * "ammo" + 0.132 * "would" + 0.129 * "want" + 0.124 * "following"

Topic 1 = -0.738 * "elon_musk" + -0.353 * "donald_trump" + -0.142 * "elon_musk_elon_musk" + -0.121 * "like" + -0.114 * "people" + -0.099 * "even" + -0.081 * "one" + -0.075 * "bernie_sanders"

Topic 2 = -0.299 * "please" + -0.224 * "subreddit" + -0.150 * "questions" + -0.149 * "discussion" + -0.149 * "moderators" + 0.126 * "elon_musk" + -0.076 * "rule" + -0.076 * "attack"

Topic 3 = -0.194 * "theft" + -0.194 * "corruption" + 0.173 * "elon_musk" + -0.135 * "far" + -0.133 * "years" + -0.129 * "operations" + -0.129 * "contractor" + -0.129 * "scheduling"

Topic 4 = 0.519 * "donald_trump" + -0.376 * "elon_musk" + 0.265 * "president" + 0.260 * "hes" + 0.210 * "pete_buttigieg" + 0.205 * "even" + 0.194 * "donald_trump_donald_trump" + 0.148 * "aware"