

# NLP Network Analysis of Twitter (X) Data

Nike, Adidas, and Lululemon



MSDS DSTA 5800 Final Project

# About this Presentation

- Very Brief: **5 Minutes, won't all through everything in the slides. Only key finding.**
- Main Focus: **Deliver Insights**
- **Technical Part:** No time for details but **will quickly go through** at very top level.
  - For details, please refer to the **notebook**

# Project Overview

## Twitter (now X)\* Data from API :

- The [database](#) is created using the **Twitter API (Search Endpoint)** to retrieve Tweets that mention the three brands **Nike, Lululemon, Adidas** for a **3-month span**.
- Data is in "[json](#)" format. For structure of data, see [here](#).

(😊 I would like to mention Twitter (now X) as just Twitter from now on. )

### Project Tasks:

1. Data exaction
2. Twitter mention network graph and business insights
3. Text preprocessing
4. Semantic network graph and business insights

(Degree Centroids and Bridgers)

### Main Tools & Environment:

1. **Google Colab** (and file savings etc)
2. File manipulation packages: **gzip, json**
3. NLP packages: **nltk**
4. Graph and visualization: **nx, pyvis**

# 1. Data Extraction

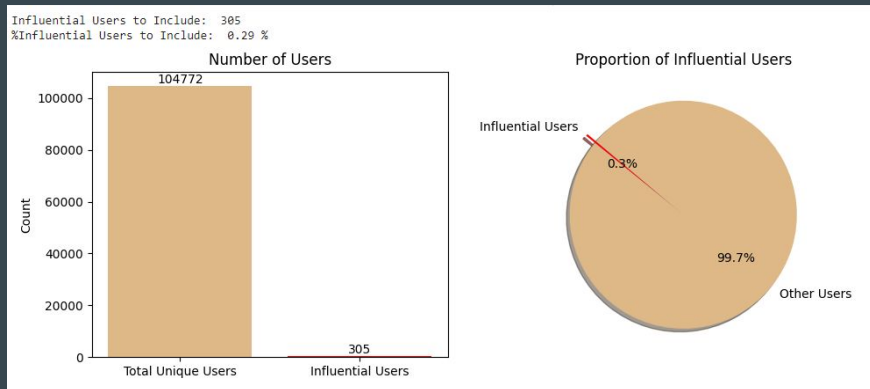
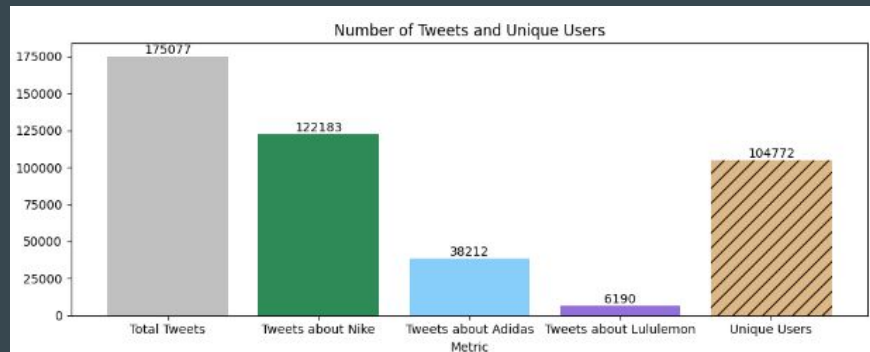
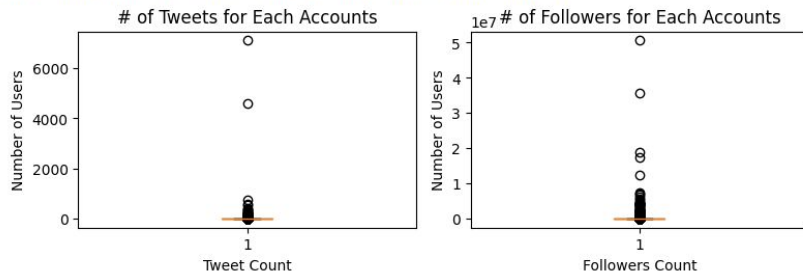
Original data:

- ~175k total tweets
- ~105k unique user accounts

## Influential Accounts

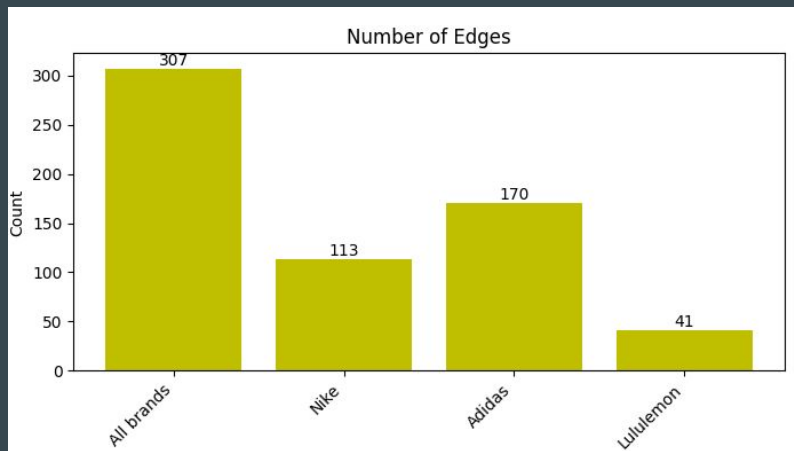
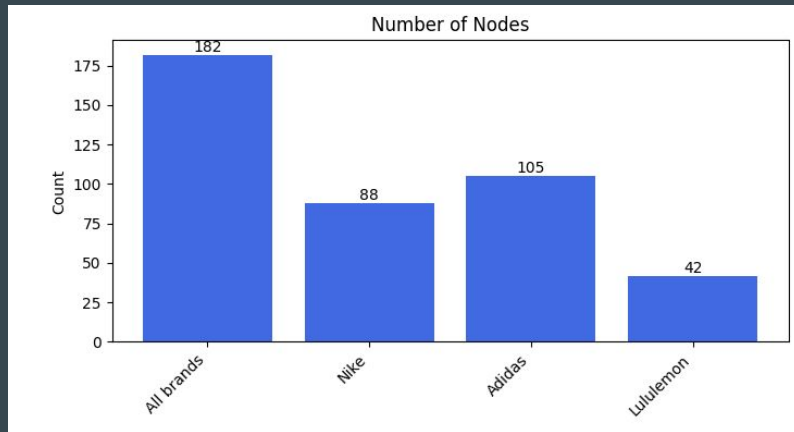
- Accounts with 2+ tweets and >100k followers (>10k for Lululemon)
- Reduced users to be 305 (0.3%)

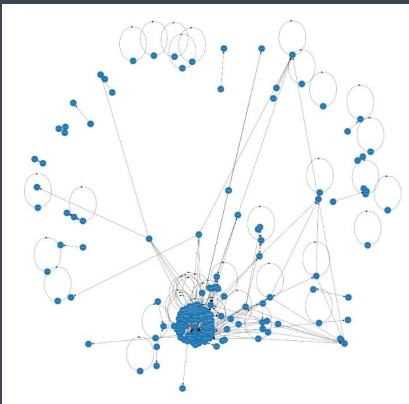
Number of tweets per users in the last 3 months - median: 1.0, min: 1, max: 7118  
Number of followers per users - median: 362.0, min: 0, max: 5061483



## 2. Mention Network Graph

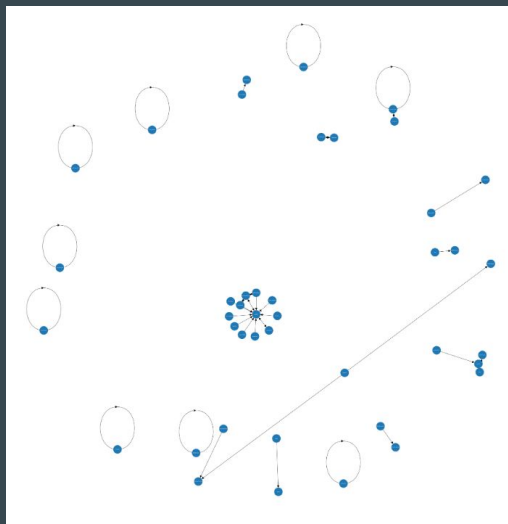
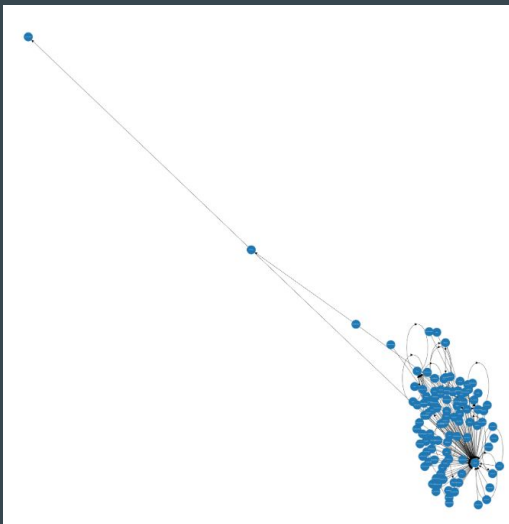
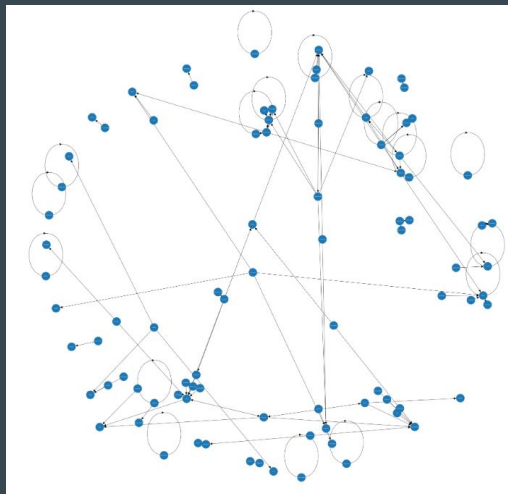
- **Directed Graph**
- Only include nodes in influential accounts (An influential account only tweets about non-influential accounts would be excluded)
- **182 nodes** (from 305 accounts)
- **307 edges**
- Graphs for each brand too





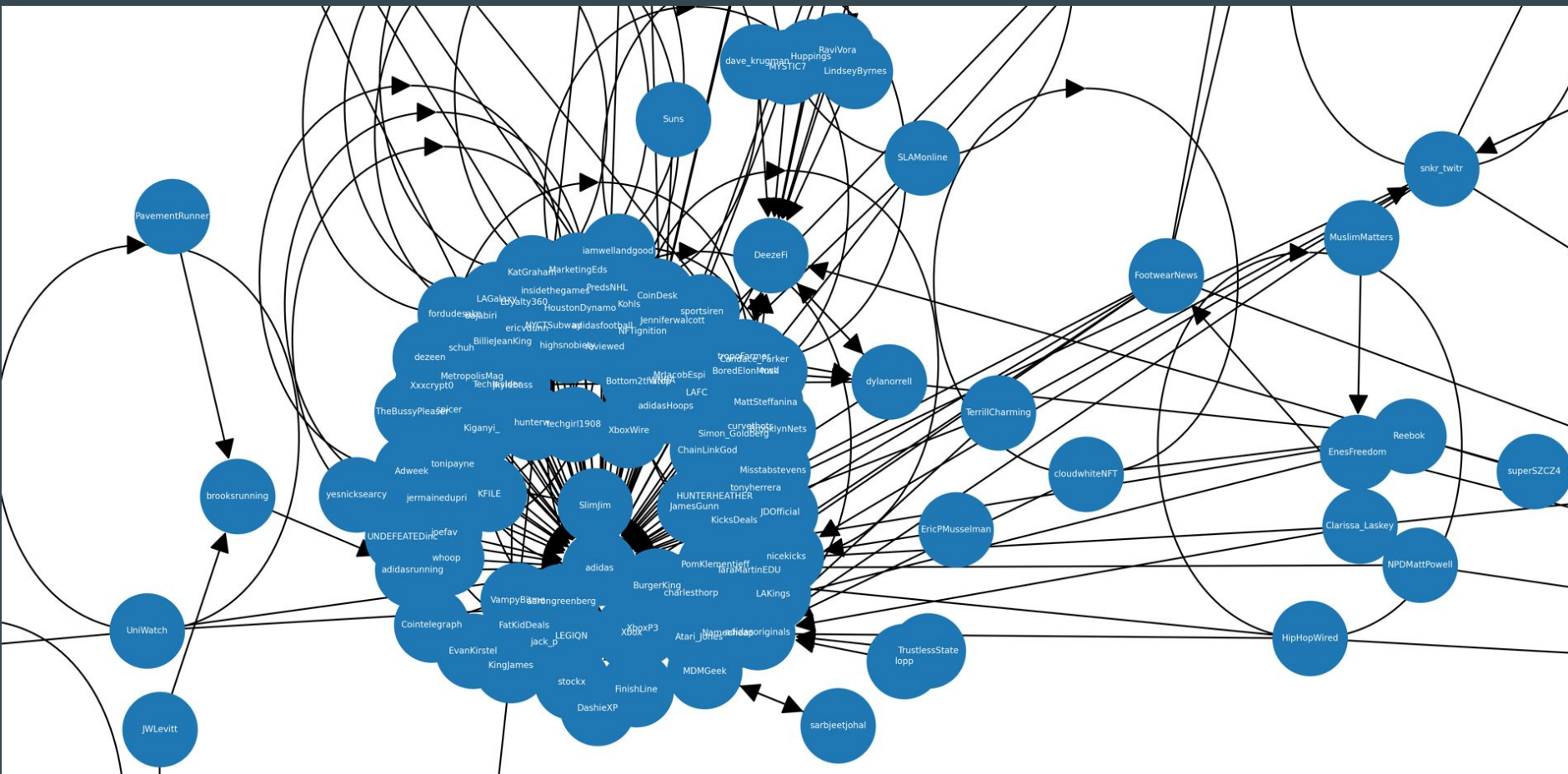
### Shapes are different

- Some have one major cluster, and small minor ones links to it
- Some have more even cluster sizes
- Some clusters are more linked, some are less linked
- Directed graph





## A Close Look At Center Part of the Graph



# Central and Bridge Accounts

From a marketing perspective, understanding **central accounts** and **bridge accounts** is crucial for developing effective business strategies. These accounts play key roles in brand visibility, audience reach, and influence across different communities.

## Central Accounts (high degree centrality scores)

These are influential or loyal accounts and can act as brand ambassadors or advocates. A brand could offer them **exclusive access/discount/events** if they are **individual persons/business or even try to collaborate with them**. If they are **brands or a famous person collaborated**, it's **a sign that the collaboration is successful**.

## Bridge Accounts (high between centrality scores)

These accounts might not have the highest follower count or engagement, but they connect different communities/clusters. These accounts can help a brand tap into new audiences or communities that they wouldn't otherwise reach (Ex: sports to fashion/art audiences). A brand could use bridges to discover **potential new communities/trends/subcultures** to broad the business.

## All Brand Graph



- Have **one major cluster** where linked between each accounts are **very strong**.

### Central Accounts:

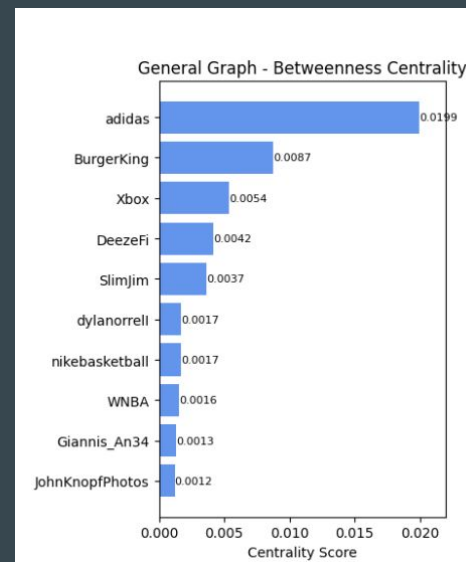
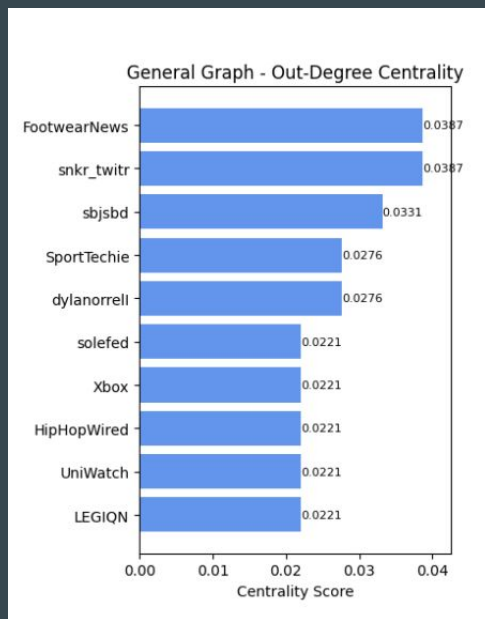
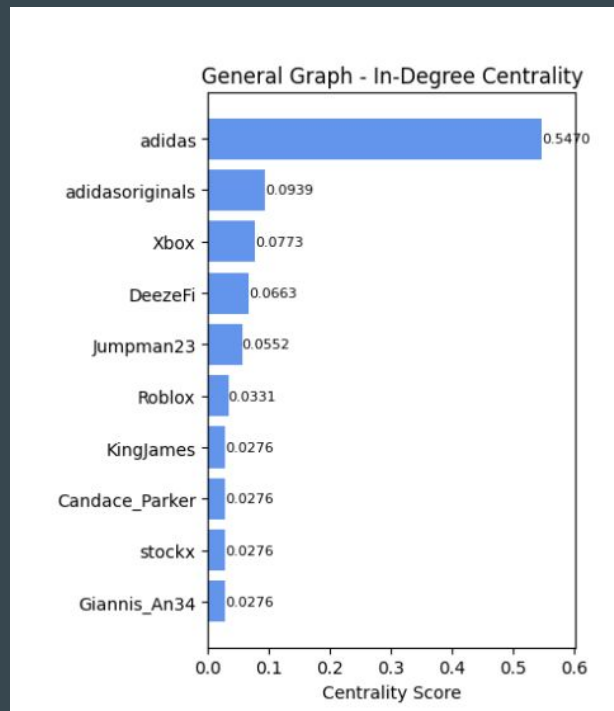
- Mostly **brands itself, companies collab with the brands, famous athletes, famous sports teams, sports news, and online stores/deals**. A few are **popular social media** accounts.
- Adidas seems to be the most popular among all 3 brands

### Bridges

- A few collab company are bridges. **Burger King** collaborates with both brands and sports teams. A brand could **co-create events or promotions** with Burger King that cater to both fast food lovers and sports fans, offering **exclusive deals** or co-branded products targeted at both audiences.
- A few bridges are **influential** social media individuals. For instance, "JohnKnopfPhotos", known for connecting photography enthusiasts and artistic communities. Collaborating with such an account allows a brand to **reach artistic audiences** they wouldn't typically engage with.



# Centrals and Bridges (All Brands)



# Some Popular Centrals and Bridges of the Brands - Who they are? 🙄

## Companies Collab with brand

- [Roblox](#): Video game company. Collab with Nike.
- [Xbox](#): Video game company. collab with Nike and Adidas.
- [BurgerKing](#): fast food company. Collab with Adidas.
- [Slimjim](#): Food company. Collab with Nike.

## Famous Athletes

- [jumpman23](#): Michael Jordan account
- [Candance\\_Parker](#): Kobe Bryant account
- [KingJames](#): LeBron James account

## Sports Teams

- [Giannis\\_An34](#): Giannis Antetokounmpo (NBA athlete)
- [PredNHL](#): Nashville Predators are a professional ice hockey team based in Nashville, Tennessee.
- [NWSL](#): National Women's Soccer League
- [LAGalaxy](#): Football team

## Sports leaders/ Performance coaches

- [MYSTIC7](#): Self-proclaimed Leader of Team Mystic
- [RealClark25](#): sports training and performance analysis program.

## Sports Discussions

- [LEGIAN](#): tweets about sports

## Online Sports stores

- [stockx](#): Online sneaker and other sports apparel store
- [nickkicks](#): Online sneaker store

## Online Sports stores

- [stockx](#): Online sneaker and other sports apparel store
- [nickkicks](#): Online sneaker store

## Owner of sports stores

- [bajabiri](#): BA,MA UDSM|Founder Bajabir Logistics |Marketing Manager JustFit Sports Gear |StreetSoul Clothing Stores

## Sportswear deals

- [solefed](#): Deal tweet account post about deals of sneakers and other sports apparel.

## Tweets about Business

- [FastCompany](#): Inspiring readers to think beyond traditional boundaries & create the future of business. Subscribe to our daily newsletter

## Popular social accounts

- [DeezeFi](#): Humorous artist
- [JohnKnopfPhotos](#): Artist
- [coachcodutti](#): Head Football Coach
- [JWLevitt](#): Sports sales entrepreneur
- [LindseyByrnes](#): Photographer and artist
- [Atri\\_Jones](#): Entrepreneur
- [MrLeonardKim](#): Individual talks about digital marketing

# Nike Graph



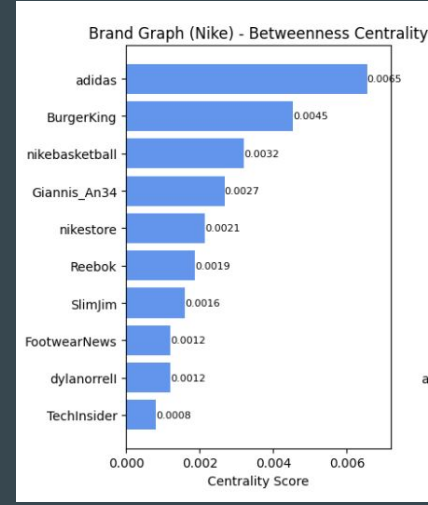
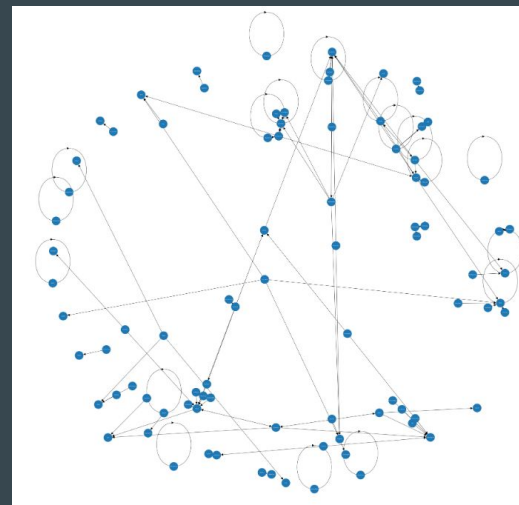
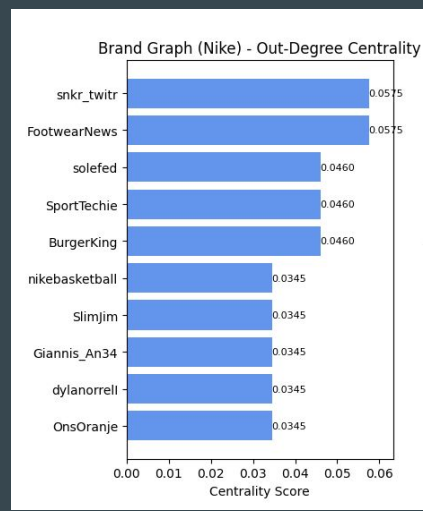
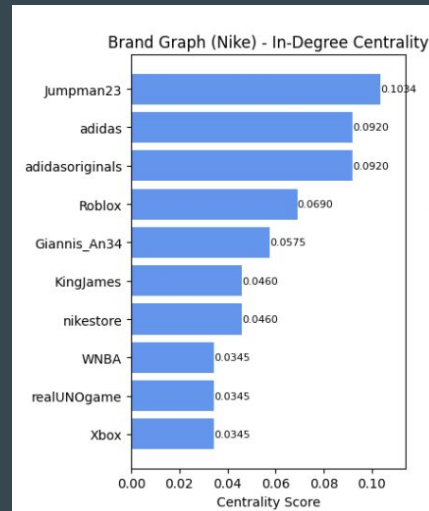
- **Many small, weakly connected clusters**, indicating Nike could focus on **linking its fragmented communities to have a stronger/broader network**

## Central Accounts:

- **Famous athletes/teams** Jordan/James/Giants ("jumpman23"/"Kingjames"/"Giannis\_An34") are major in-degree centrals/ They play an important part of brand engagement.
- Involved with **Basketball and football players and teams** a lot.
- For **out-degree cereals**, there are lots of **footwear** stores/deals accounts. Footwear seems to dominate Nike sales.
- **Adidas brand accounts are central accounts** in Nike graph while **Nike isn't** (although "nikestore" is but ranked lower). This is a sign account talk about nike also talks about Adidas, and probably talk about Adidas even more. (not good!) Nike might need to try to be engaged more.

## Bridges:

- Once again, **Adidas acts as a top bridge account** in the Nike network. This suggests Adidas is also likely driving engagement across different groups or clusters within the Nike-related conversations.



# Adidas Graph



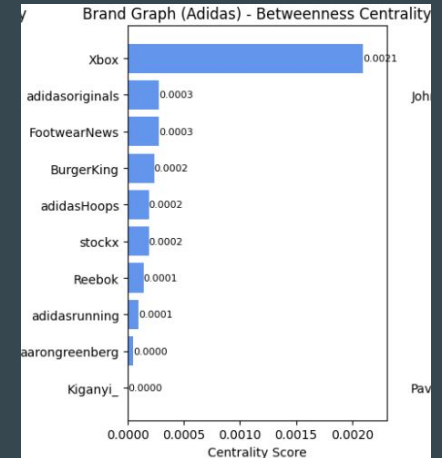
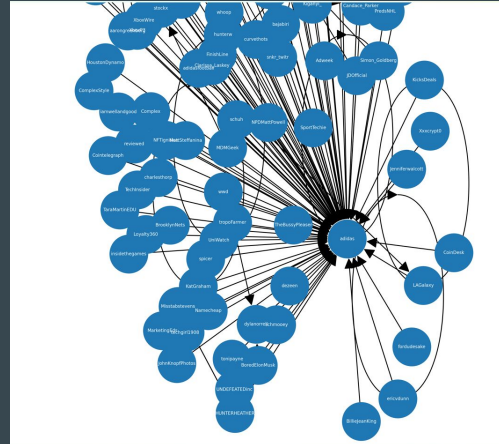
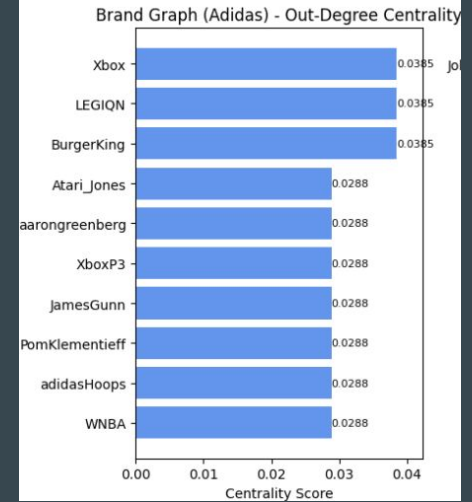
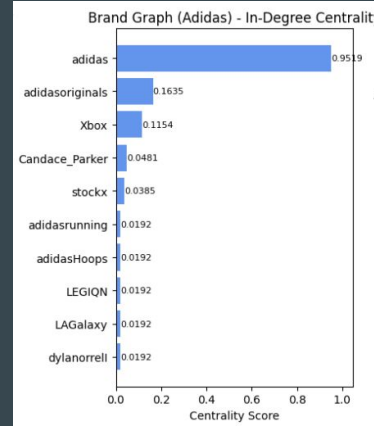
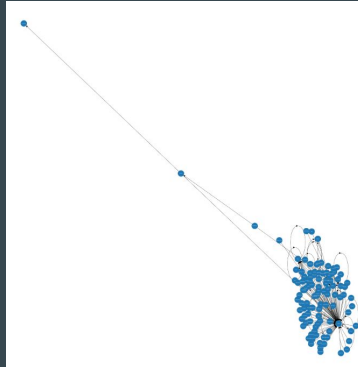
- Big Cluster centered at "adidas".

## Central Accounts:

- Several adidas brand accounts are central accounts ("adidas", "adidasoriginal", "adidasrunning", "adidasHoops"). Adidas seems to do well on brand engagement/awareness and have varieties of activities/methods/account to engage customers and the engagement is successful.

## Bridges:

- Xbox ranked very high in both central accounts and bridge accounts. It's a very successful collaborate company for Adidas.

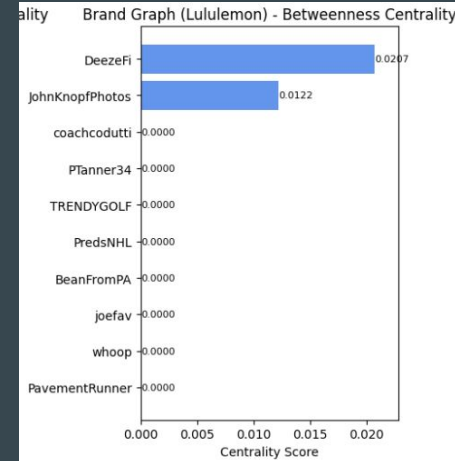
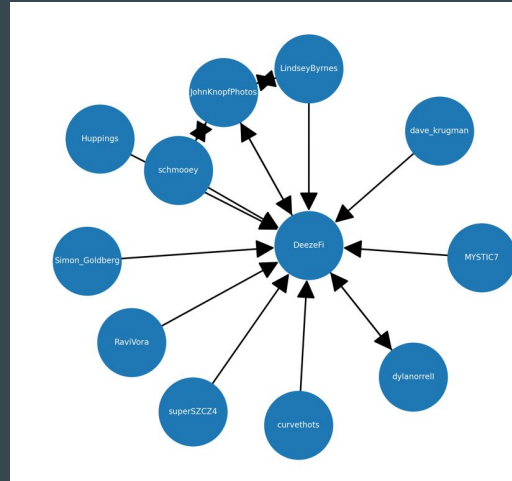
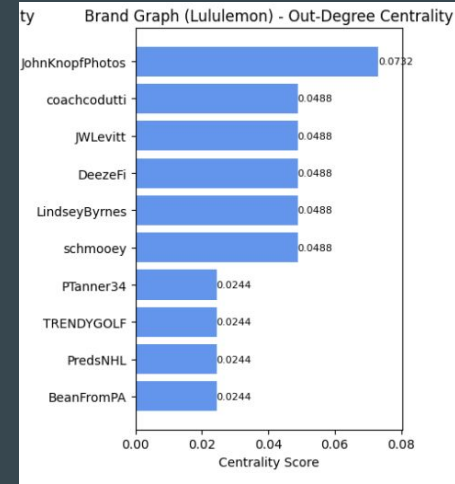
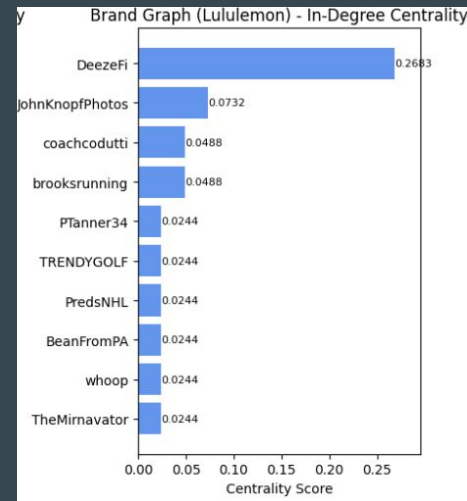
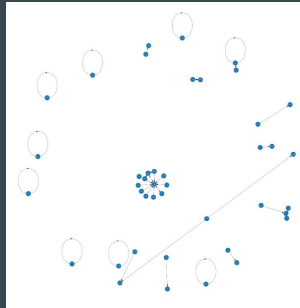


# Lululemon Graph

- Like Nike, has **many small, weakly connected clusters**, except one account “DeezeFi” has strong in-degree.
- **Way less popular influencers**: have to adjust threshold to include more accounts (influential accounts as >100k Nike and Adidas followers vs. >10k for Lululemon)

## Central and bridge Accounts:

- More **popular (but not celebrity) social media**. They are mostly **artist, photographers, entrepreneurs, and other content makers**,
- Centrals/Bridges include more **health/fitness**, and sports other than basketball (**running/golf/ice hockey**) related accounts



We just finished with  
mention graph. 🙌

Now let's discuss  
Semantic network graph.

👉 But first, some text  
preprocessing....



# 3. Text Preprocessing

## First Round Cleaning:

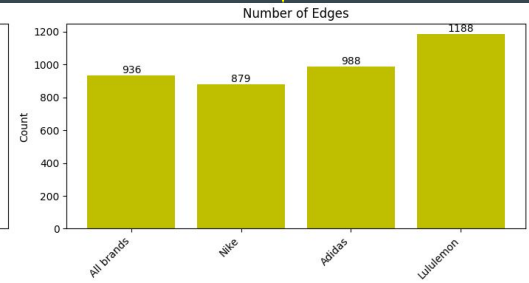
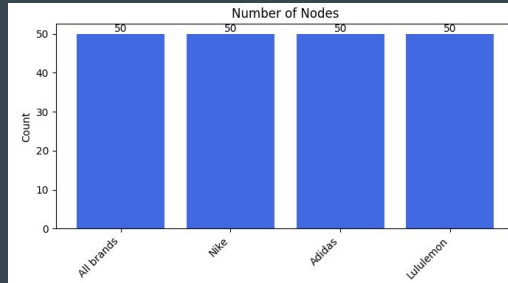
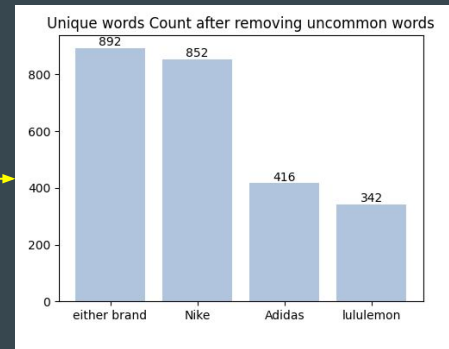
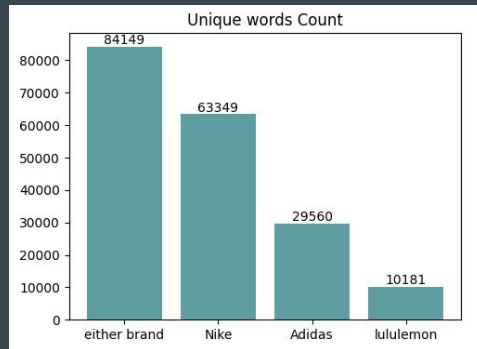
- Tokenization
- Lower case
- Remove links, stop words, punctuations, emoji, short words
- Lemmatization
- Unique words count is **~84k**

## Second Round Cleaning:

- Remove uncommon words: minimum word count  $\geq 300$  for all brands, proportion for each brand based on number of unique words)
- Word count reduced to **892**

## Graph:

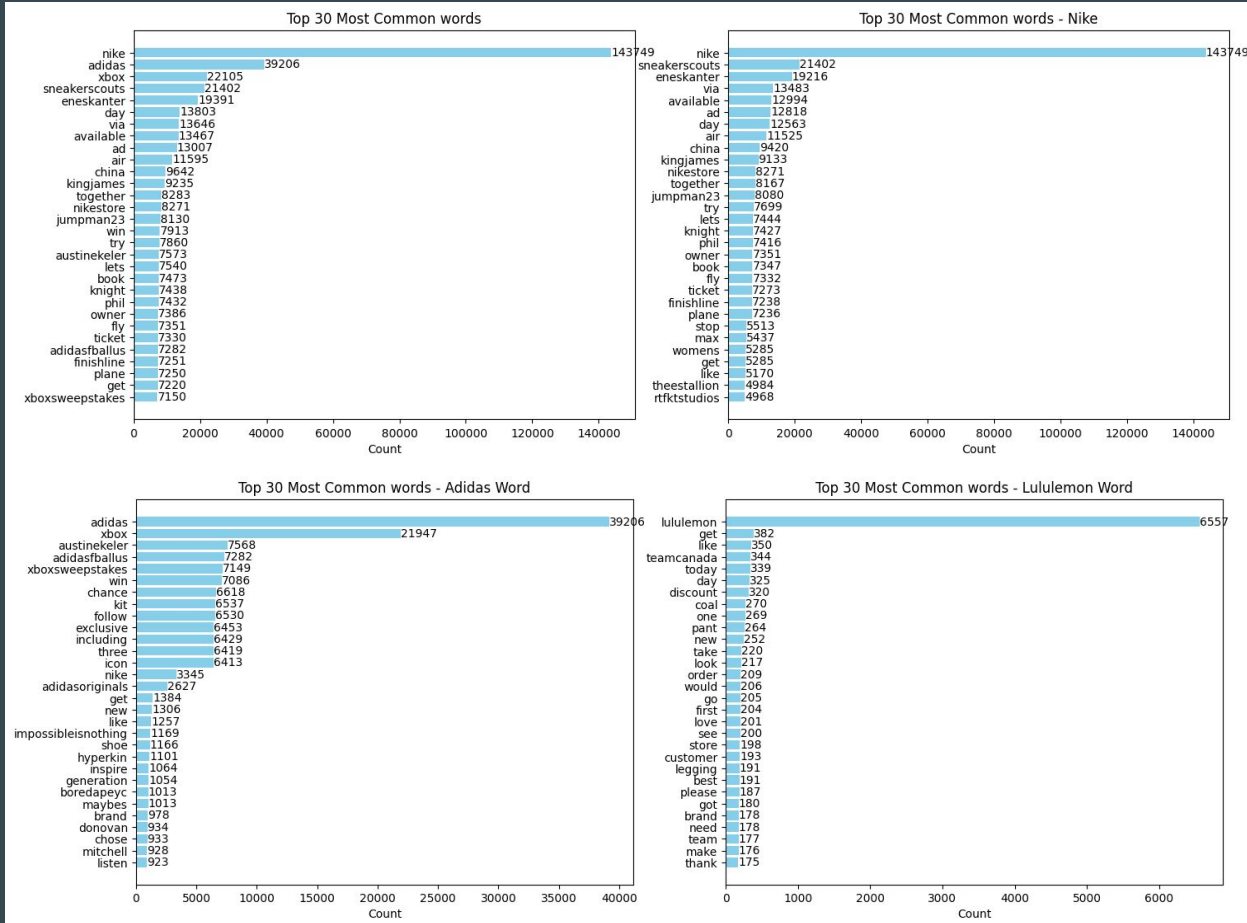
- Nodes: keep only **top 50 most common** words (for all brands and each brands)
- Edges: **936** for all brands



```
MINCOUNT = 300
MINCOUNT_NIKE = MINCOUNT * len(uniquewords_nike)/ len(uniquewords)
MINCOUNT_ADIDAS = MINCOUNT * len(uniquewords_adidas)/ len(uniquewords)
MINCOUNT_LULULEMON = MINCOUNT * len(uniquewords_lululemon)/ len(uniquewords)
```

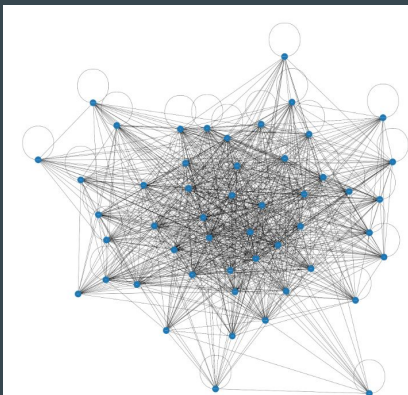


# Top 30 Most Common Words



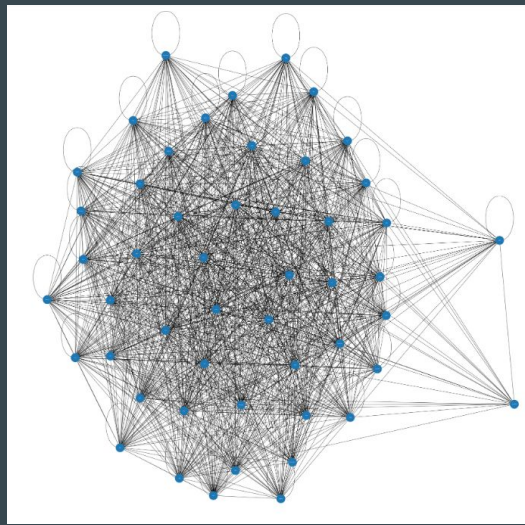
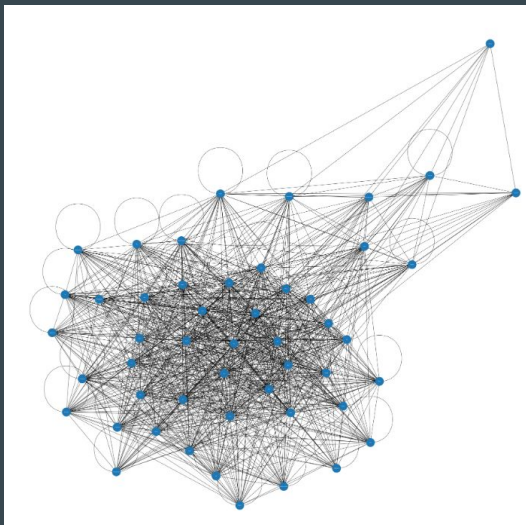
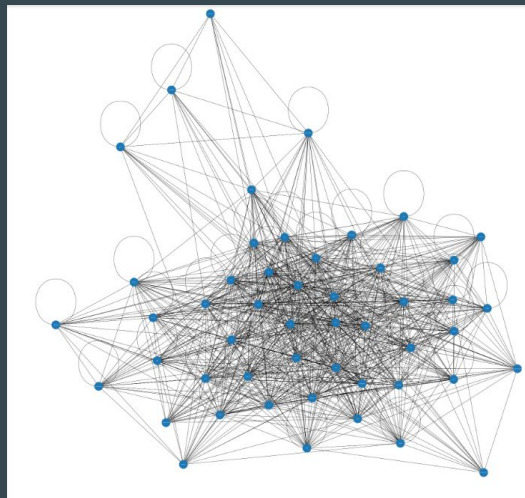


## 5. Semantic network graph.

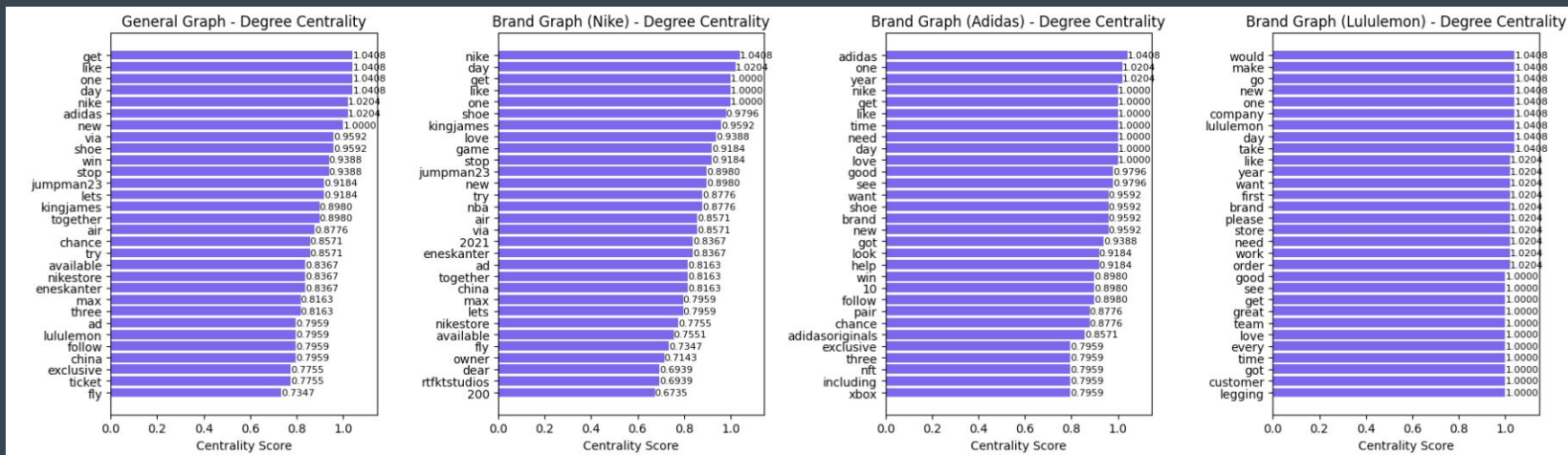


Shapes are more similar (at first glance)

- Many more edges than mention graph
- Words related to each other more than accounts
- Undirected graph



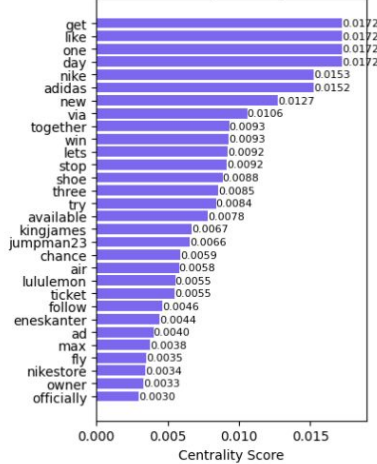
# Central Words



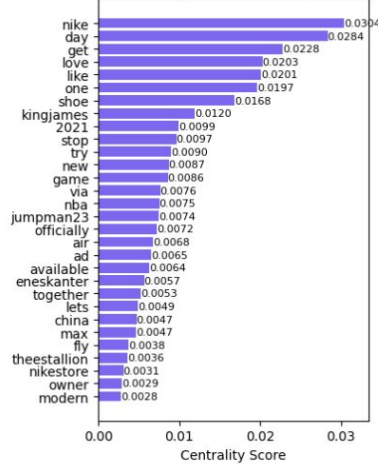
- **Xbox** collaborate with both Nike and **Adidas**, but it's **not under top 30 central words** of **Nike**. It could be sign collaboration of **Xbox with Adidas is more successful**.
- **Lululemon** has a few central words that could **indicate room for improvement** of products. For example, **"would"** and **"need"** could indicate the customer prefer some improvement or some other features. **"customer"** could link to customer service dissatisfaction. Brand could dive deeper in tweets with those keywords.

# Bridge Words

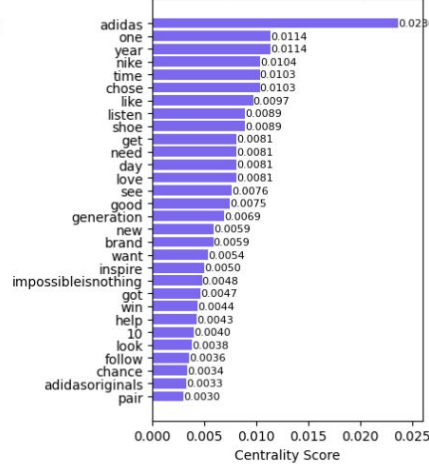
General Graph - Bridge Words



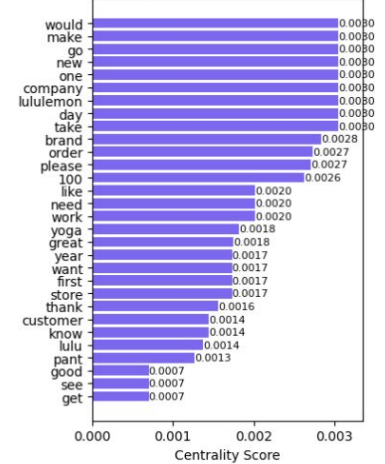
Brand Graph (Nike) - Bridge Words



Brand Graph (Adidas) - Bridge Words



Brand Graph (Lululemon) - Bridge Words



- The word **"nba"** serves as a unique bridge word for Nike, distinguishing it from other brands. This may highlight Nike's strong association with basketball culture.
- For Adidas, **"listen"** and **"impossibleisnothing"** are unique bridge words. These words don't appear as bridge words for other brands, suggesting they reflect Adidas' connection to music and its inspiring brand message, "Impossible is Nothing."
- For Lululemon, **"work," "yoga,"** and **"pant"** stand out as unique bridge words. These words suggest that Lululemon is closely associated with workwear and yoga, with pants being a particularly popular product category for their customers.