

# Social Media Data Collection and Analysis: Overview and Demonstration

**Social Science Librarians Boot Camp 2019**

Brittany Andersen

PhD Candidate at Boston University / UX Researcher at Ancestry

# SSLBC 2019 Session Agenda

---

- 1. Introduction to Social Networks**
- 2. Data Collection Methods and Limitations**
- 3. Overview of Social Listening**
- 4. Live Demonstration of Crimson Hexagon (Paid Platform)**
- 5. Q&A**

# Introduction to Social Networks

# First, let's define social networks

---

- What is a **network**?
  - A group or system of interconnected people or things.
- What is **social network**?
  - Social structure made up of interconnected people specifically. Emphasis on interactions.



# Examples of online social networks

- Community
- Organizational
- Interest groups (health, music, politics, activism)
- Familial
- Professional



# Social Networking Platforms

---



# Why should we research social media?

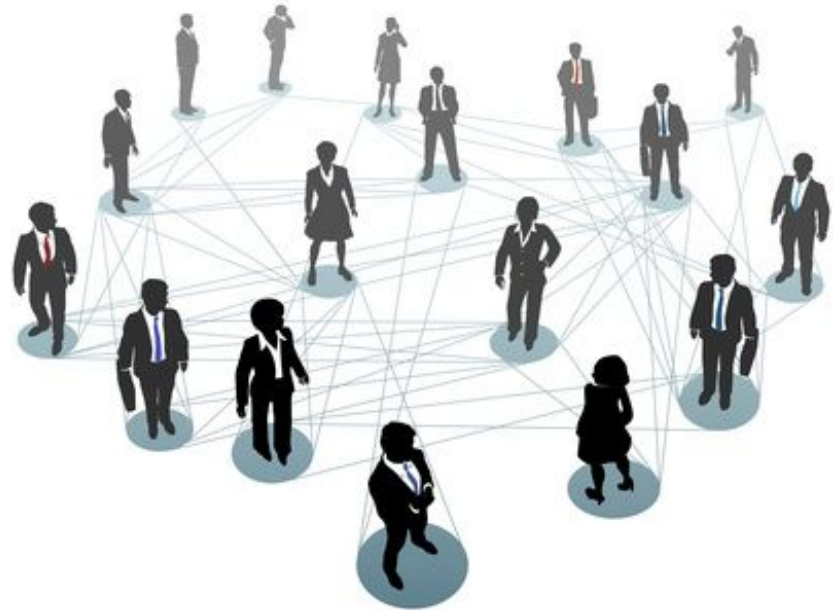
---

- Social media seems to intersect with **nearly every aspect** of personal and public life in contemporary society.
- Platforms are not discipline specific (like other methods such as ethnographies, surveys, etc.)
- **Data-driven** research that relies on analyses of social media data has grown tremendously in volume and rigor.
- Why: This data provide access to **what people do**, not what they report to have done (in terms of behaviors, content, engagement, sentiment).

# Why should we research social media?

---

- Staying up to date on **new developments** and events in a field (such as health/political) or social circle.
- Gain insight into how information (or misinformation) spreads
- Understanding **hierarchy** and influence in social settings or fields.
- Building strategic connections important in **any job field** or enterprise.





# To keep in mind when exploring:

---

- Not all phenomena will resonate as much on social media – (many things still happen elsewhere).
- Understanding social media depends on data, which can be **incomplete** and/or **expensive**.
- Platforms are different and may require varying approaches or tools, think “**medium specificity**.”

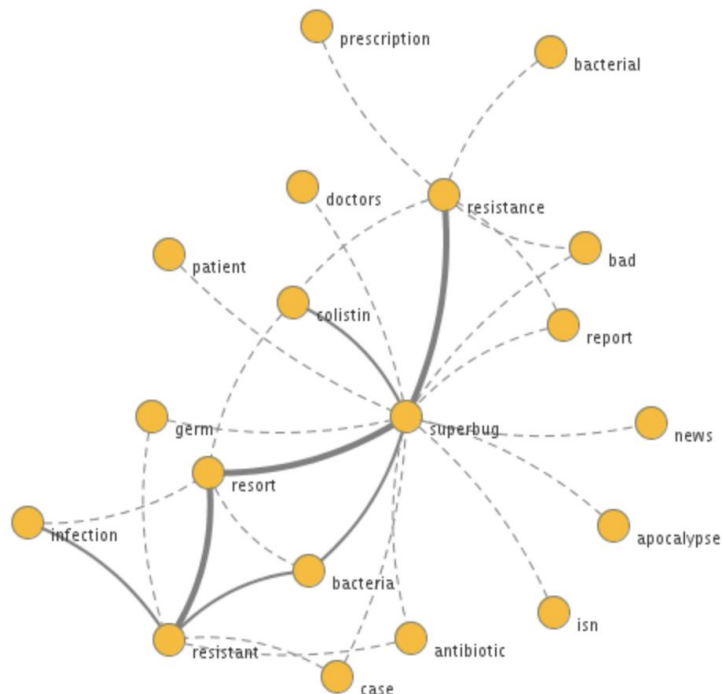
# We are explorers!

---

- While we can use these methods for hypothesis testing, we are currently undertaking **exploratory data analysis**.
  - We don't know what we will find!
- We need methodological **flexibility**.
  - Adjusting keyword searches, filtering out terms, consideration of other tools

# Investigating social media content

- **Which users are talking?**
  - (Who are they, what do they do)
- **Where are the users from?**
  - (Country of origin/geolocation)
- **What are they talking about?**
  - (Politics, health topic, world events)
- **When are they talking about it?**
  - Timeframe, spikes in dataset
- **How are they talking about it?**
  - (Sentiment, tone, writing style, grammar versus chat speak)
- **Characteristics of conversation?**
  - (Retweets, link sharing, @mentioning specific users, hashtags associated with topic)

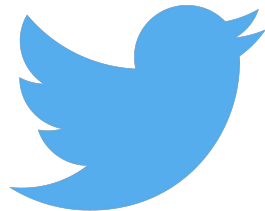


# Data Collection: Methods and Limitations

# How can we collect social media data?

---

- In social media monitoring, we mostly rely on **Twitter data** due to widespread accessibility.
- Limitations due to accounts with more **privacy** restrictions.
  - Collection from Facebook pages than personal accounts.
- Collection for research is typically through the source (e.g. Twitter), via **third-party**, or even hand-collected



# Free vs. Premium Options and Limitations

— — —

## *Twitter Search (Free)*

- Free
- Data comes in near-real-time
- No access to **historical data** or visualization options
- **Limitations on queries**
  - Basic keywords, limited filters
- Requires you to think and plan ahead
  - Like a radio, you cannot replay a song that has already played

## *Twitter & Third Party (Premium)*

- Expensive
- Twitter Search (Premium/Enterprise)
- Third-party site (Sysomos, Crimson Hexagon)
- Access to historical data + visualization options (third-party)
- More flexibility with research queries
  - Advanced Boolean strings and search filters

# Overview of Social Listening

# What is social listening?

---

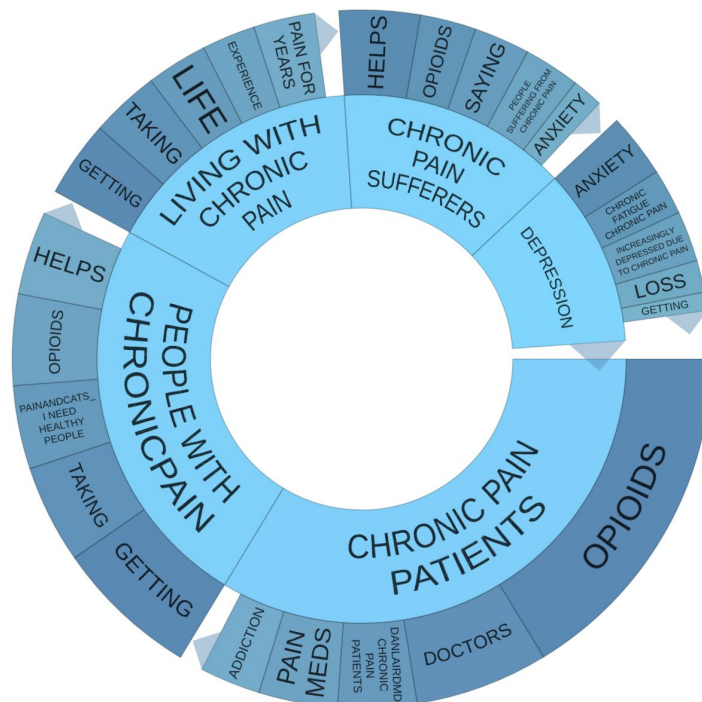
- The process of monitoring and tracking digital social conversations
- Widely-used platforms: **Sysomos** and **Crimson Hexagon**





# What is social listening?

- **Analyzing users**
  - Who is partaking in discussion?
  - Who are the influencers?
  - Demographics
- **Identifying popular content**
  - What is being shared?
  - Links, retweets, original comments?
- **Topics of interest**
  - Public health -> antimicrobial resistance
  - Chronic pain -> depression
- **Post volume**
  - Spikes in conversation?
- **Media**
  - Videos or photos shared



# Live Demonstration: Crimson Hexagon

# Thank you! Questions?

Brittany Andersen

bla@bu.edu

PhD Candidate at Boston University / UX Researcher at Ancestry