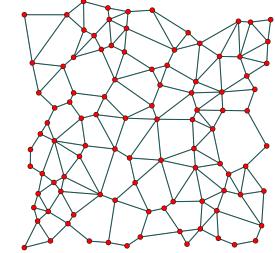


---

# The Musical Collaboration Graph



---

— Akash Levy, Erica Wu, Sunny He,  
Vincent Po, Daniel Wood

---

# Motivation

## Who's the most influential pop artist in music?

- Which artists know the most other artists/are the most central in music?
- Is there a correlation between an artist's popularity and his/her collaborations?
- Examine whether collaborations might be a useful metric for how influential an artist is
- Potential tool for artists and labels who want to understand how collaboration is related to popularity



# Introduction to Spotify

- Music, podcast, and video streaming service
- Large database of artists and songs
- Accurate artist collaboration info
- API is great:
  - Simple, found a nice library for Python
  - Fast response time
  - Almost zero rate-limiting
  - No need to login
- Bonus: artist profile pictures available through API

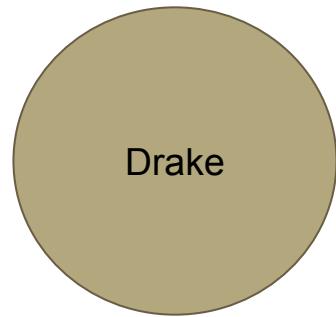


# Methodology

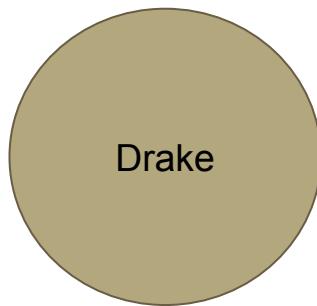
- Extract a collaboration graph from Spotify's immense music catalog
- Scrape a subset starting from a specific artist
- Build graph using networkX



# Spotify Scraping



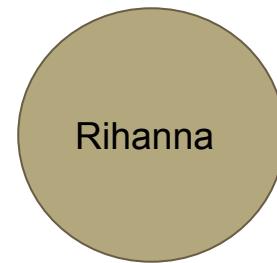
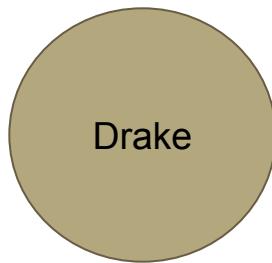
# Spotify Scraping



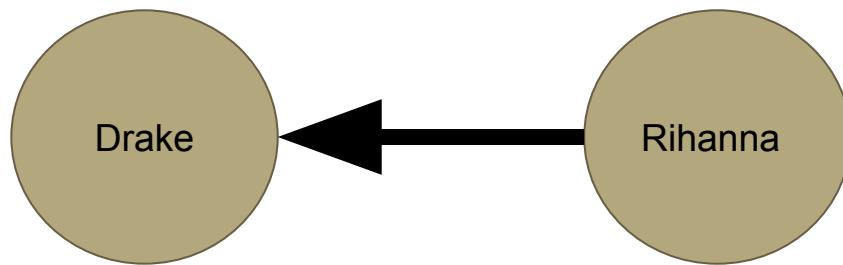
Drake

**Too Good**  
By Drake  
  
Feat. Rihanna

# Spotify Scraping

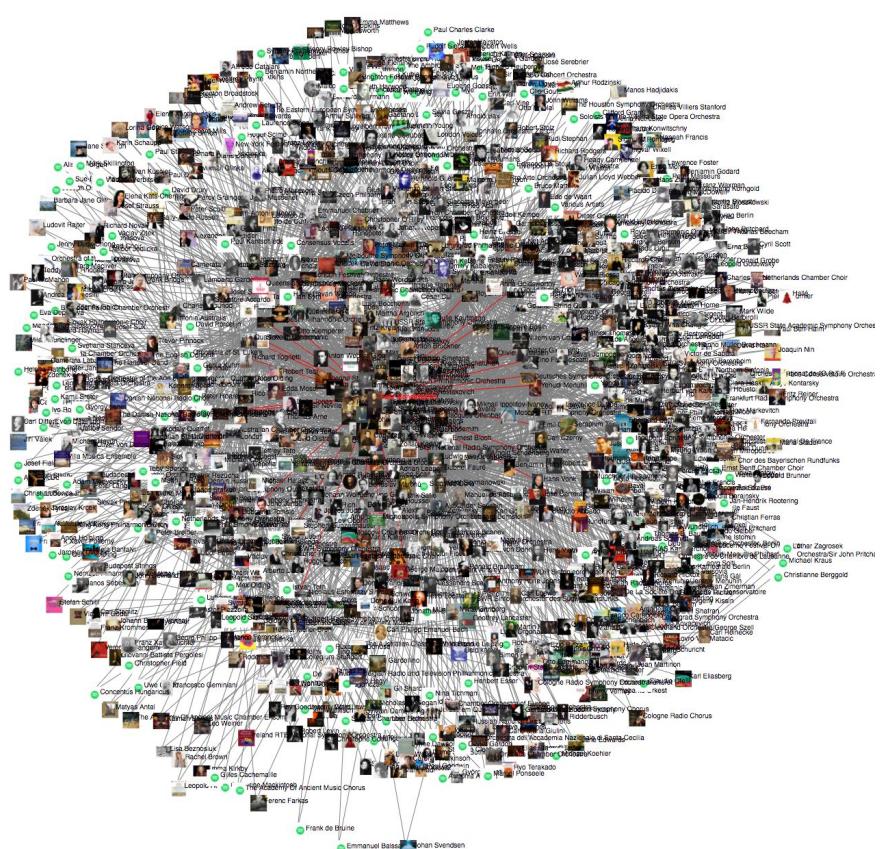


# Spotify Scraping



# Visualization Techniques

- D3js Visualization Library
- Interactive Force simulation
  - Edges pull nodes together with force proportional to number of collaborations
  - Nodes repel each other with fixed force
- Highly connected nodes tend to self-center



# The Subgraph Boundary Problem

- Impossible to extract entire Spotify collaboration graph
- Have to stop somewhere
- Always biased towards central nodes i.e. central nodes have greater degree
- Can introduce problems at boundaries of the graph
- Options we came up with:
  - Keep all nodes anyways
  - Keep all nodes that have at least two links (i.e. degree  $>1$ )
  - Keep only nodes in the 1000 iterations of BFS
  - Keep only nodes in the 1000 iterations of BFS that have at least two links

# GRAPH 1: Keep all nodes

**Number of nodes:** 12813

**Number of edges:** 19991

**Average degree:** 3.1204

**Clustering Coefficient:** 0.0537585183177

- Graph has too many nodes (12x what we wanted)
- Average degree is extremely low due to boundary nodes



# GRAPH 2: Keep all nodes with at least two links

Number of nodes: 2302

Number of edges: 9325

Average degree: 8.1017

Clustering Coefficient: 0.317



- Extremely dense
- High average degree
- Lots of nodes

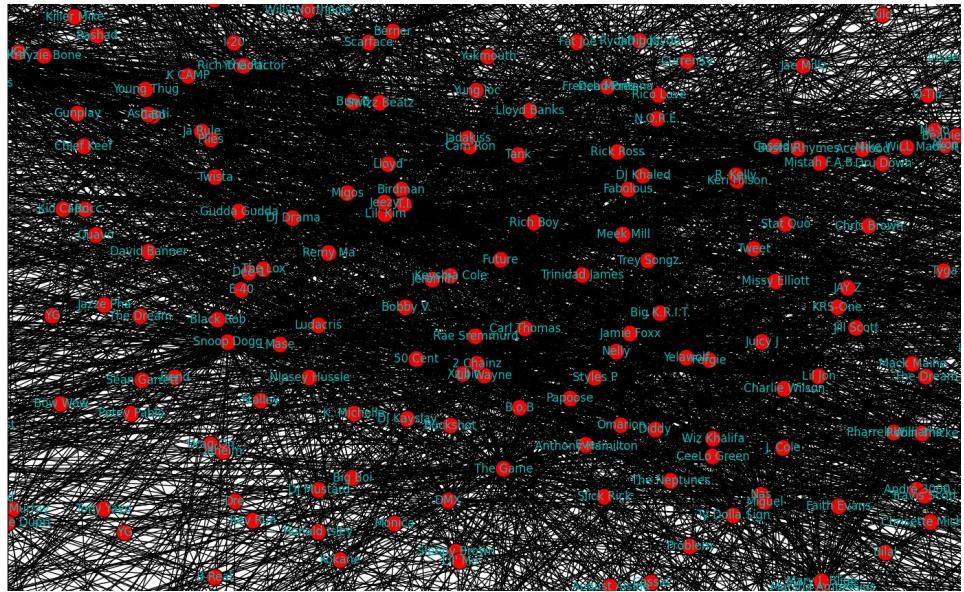
# GRAPH 2: Keep all nodes with at least two links

Number of nodes: 2357

Number of edges: 9535

Average degree: 8.0908

- Accurate degree in center
- Less accurate at boundary



# GRAPH 3: Keep nodes in 1000 iterations of BFS

Number of nodes: 999

Number of edges: 4999

Average degree: 10.0080

Clustering Coefficient: 0.231

- Can see clustering
- Outskirt artists are prolific



# GRAPH 3: Keep nodes in 1000 iterations of BFS

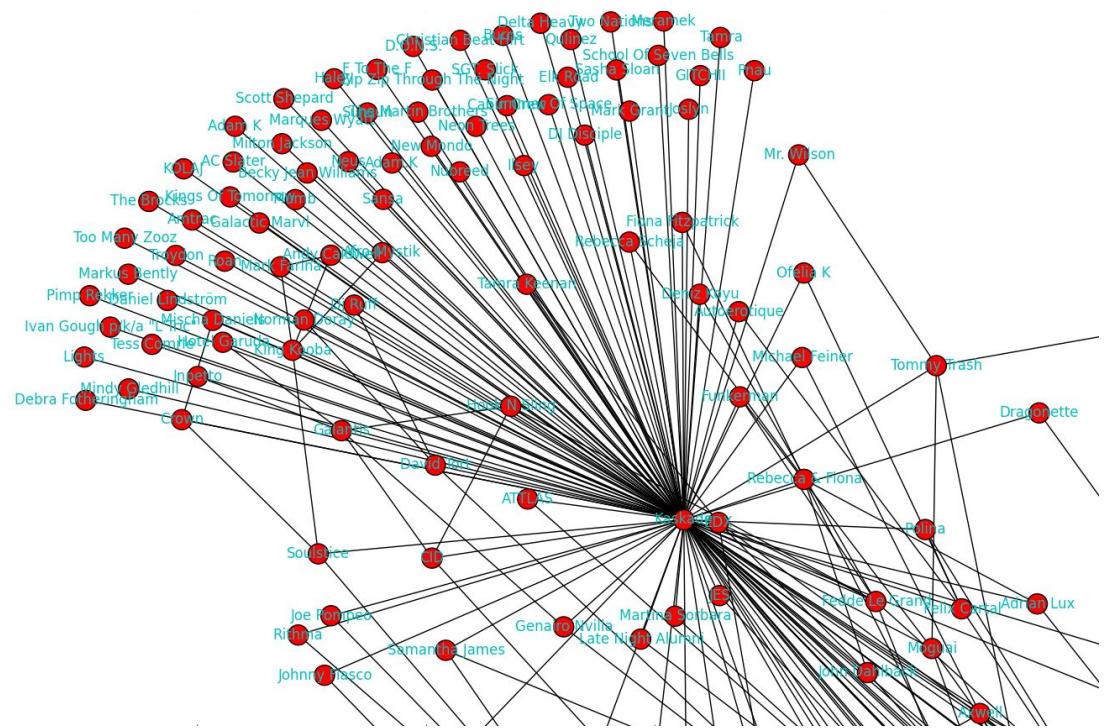
**Number of nodes: 999**

# Number of edges: 4999

Average degree: 10.0080

## Clustering Coefficient: 0.231

- Can see clustering
  - Outskirt artists are prolific
  - Zoomed in on outskirts



# GRAPH 4: Keep only nodes in the 1000 iterations of BFS that have at least two links

Number of nodes: 652

Number of edges: 4652

Average degree: 14.2699

Clustering Coefficient: 0.358

- High average degree
- Fewer nodes left
- Can see influential artists



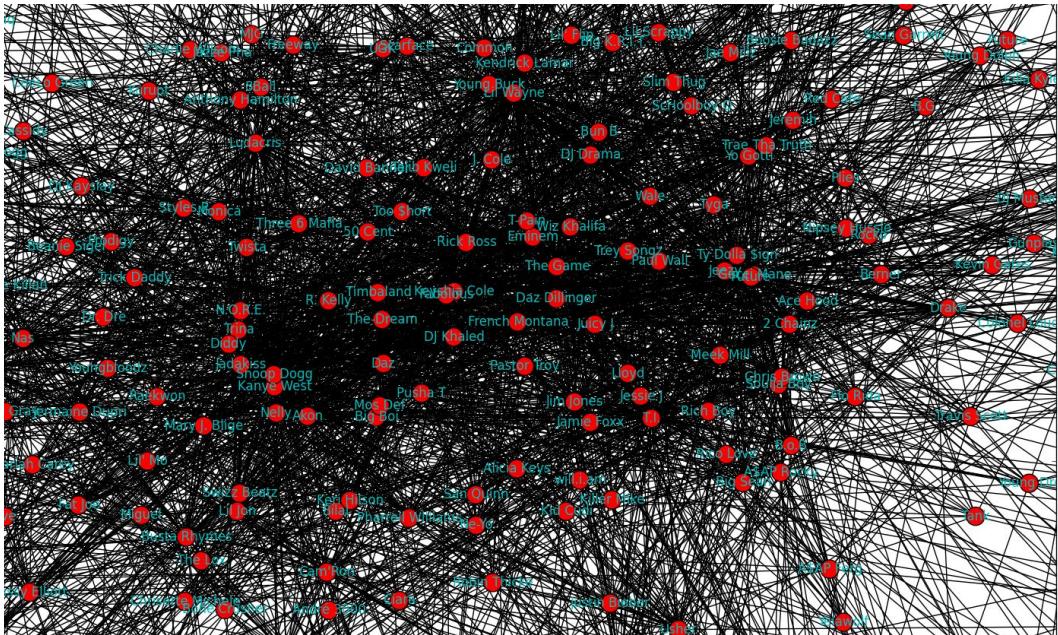
# GRAPH 4: Keep only nodes in the 1000 iterations of BFS that have at least two links

Number of nodes: 652

Number of edges: 4652

Average degree: 14.2699

- High average degree
- Fewer nodes left
- Can see influential artists



# Demo

[goo.gl/B1uOEU](http://goo.gl/B1uOEU)

# Why start with Drake for BFS?

- He started from the bottom
- Currently is #1 most-played artist on Spotify (recently and all-time!)
- Large number of collaborations
- Known to be an influential artist in his genre
- Very central in mainstream music: will almost certainly get to Drake within 1,000 breadth-first iterations from anywhere



# Measuring Artist Importance (Centrality)

Now that we have assembled a static graph, we can examine the graph topology, using different heuristics to determine the importance of artists.

1. **Degree Centrality:** fraction of nodes a node is connected to
2. **Eigenvector Centrality:** “left” eigenvector centrality of in-edges
3. **Closeness Centrality:** Reciprocal of the sum of shortest path distances from a node to all other nodes
4. **Betweenness Centrality:** Sum of fraction of all-pairs shortest paths through a node

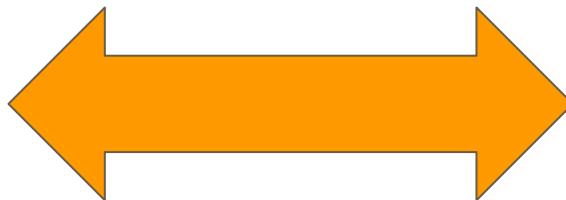
# Undirected vs. Directed Musical Networks

## UNDIRECTED

- Assumes equal participation in a song composition/performance
- Emphasizes a “who knows who” mentality in the music world
- More similar to a friend graph

## DIRECTED

- Remixes: Maker may list original artist as a collaborator (dangling node)
- For every song, there is usually a main and supporting artist



# Centrality in Undirected Graph 4: top 10

	DEGREE	EIGENVECTOR	CLOSENESS	BETWEENNESS	PAGERANK	AGGREGATE	SPOTIFY
1	Lil Wayne	Lil Wayne	T.I.	Kaskade	Kaskade	Lil Wayne	90
2	T.I.	Rick Ross	Lil Wayne	T.I.	Lil Wayne	T.I.	80
3	Rick Ross	Gucci Mane	Rick Ross	Lil Wayne	T.I.	Rick Ross	82
4	Gucci Mane	T.I.	Nicki Minaj	Beenie Man	Rick Ross	Ludacris	77
5	Ludacris	The Game	Ludacris	Beyonce	Ludacris	Gucci Mane	89
6	Snoop Dogg	DJ Drama	Usher	Nas	Gucci Mane	Kaskade	72
7	The Game	Ludacris	Snoop Dogg	Ludacris	Beenie Man	Snoop Dogg	83
8	2 Chainz	DJ Khaled	Busta Rhymes	Usher	Snoop Dogg	The Game	77
9	DJ Drama	Jeezy	Gucci Mane	Rick Ross	Nas	Beenie Man	60
10	Nas	Snoop Dogg	The Game	Nicki Minaj	2 Chainz	Nicki Minaj	91

# Centrality in Directed Graph: top 10

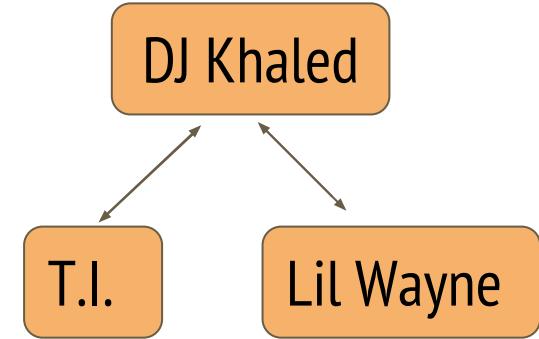
	EIGENVECTOR	CLOSENESS	BETWEENNESS	PAGERANK	HITS (Hub)	HITS (Auth)	AGGREGATE	SPOTIFY
1	Gucci Mane	Lil Wayne	Kaskade	Kaskade	Lil Wayne	DJ Drama	T.I.	80
2	DJ Drama	Rick Ross	T.I.	T.I.	Rick Ross	Gucci Mane	Rick Ross	82
3	DJ Khaled	Snoop Dogg	Lil Wayne	Gucci Mane	T.I.	DJ Khaled	Lil Wayne	90
4	The Game	Busta Rhymes	Rick Ross	Ludacris	Snoop Dogg	The Game	Gucci Mane	89
5	Ludacris	2 Chainz	Beenie Man	Cam'Ron	2 Chainz	Rick Ross	DJ Drama	67
6	T.I.	T.I.	Snoop Dogg	DJ Drama	Jeezy	Ludacris	Ludacris	77
7	Rick Ross	Nas	Nicki Minaj	JAY Z	Bun B	Trae Tha Truth	Snoop Dogg	83
8	Trae Tha Truth	Kanye West	Ludacris	Lil Wayne	Nas	T.I.	Kaskade	72
9	French Montana	Nicki Minaj	Nas	Beenie Man	Wiz Khalifa	French Montana	DJ Khaled	79
10	Yo Gotti	Pharrell Williams	JAY Z	Rihanna	Jadakiss	Nelly	The Game	77

# Homophily in Pop Music?

Do pop artists with similar styles influence each other and collaborate more often? We examined the (in)dependence of artists in the Spotify network

- Global coefficient indicates overall clustering in network
- How does the topology of the Spotify network compare to a social network such as Facebook?

Average Clustering Coefficient: 0.358



# Clique Analysis

**Question:** Are there big cliques (fully connected subgraphs) in the music industry? Who's in them and how big are they?

**Answer:** Yes, the maximum size is **10**. There are a lot of 9s and 8s as well.

Kanye West, Rick Ross, Lil Wayne, DJ Khaled, Jeezy, Gucci Mane, T.I., Ludacris, Future, Usher

Kanye West, Rick Ross, Lil Wayne, DJ Khaled, Jeezy, Gucci Mane, T.I., Ludacris, Future, The Game

Kanye West, Rick Ross, Lil Wayne, DJ Khaled, Jeezy, Gucci Mane, Nicki Minaj, Yo Gotti, French Montana, 2 Chainz

Kanye West, Rick Ross, Lil Wayne, DJ Khaled, Jeezy, Gucci Mane, Nicki Minaj, Yo Gotti, French Montana, Jadakiss

Kanye West, Rick Ross, Lil Wayne, DJ Khaled, Jeezy, Gucci Mane, The Game, Yo Gotti, 2 Chainz, Future

Kanye West, Rick Ross, Lil Wayne, DJ Khaled, Jeezy, Gucci Mane, The Game, Yo Gotti, 2 Chainz, French Montana

Kanye West, Rick Ross, Lil Wayne, DJ Khaled, Jeezy, Nas, Nicki Minaj, Chris Brown, Usher, Ludacris

Kanye West, Rick Ross, Lil Wayne, DJ Khaled, Jeezy, Chris Brown, T.I., Ludacris, Future, Usher

Kanye West, Rick Ross, Lil Wayne, DJ Khaled, Jeezy, Chris Brown, T.I., Ludacris, Future, The Game

DJ Khaled, Lil Wayne, Rick Ross, Gucci Mane, Nicki Minaj, French Montana, Yo Gotti, Wale, 2 Chainz, Wiz Khalifa

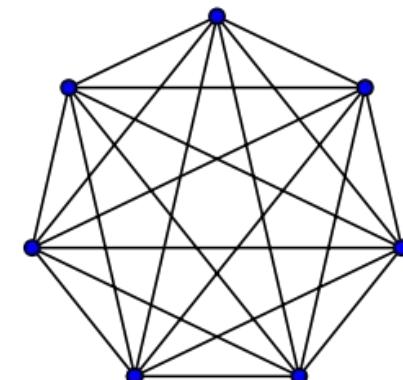
DJ Khaled, Lil Wayne, Rick Ross, Gucci Mane, Nicki Minaj, French Montana, Tyga, Wiz Khalifa, 2 Chainz, Wale

DJ Khaled, Lil Wayne, Rick Ross, Gucci Mane, The Game, French Montana, Wale, 2 Chainz, Wiz Khalifa, Yo Gotti

DJ Khaled, Lil Wayne, Rick Ross, Gucci Mane, The Game, French Montana, Wale, 2 Chainz, Wiz Khalifa, Tyga

DJ Khaled, Lil Wayne, Rick Ross, Chris Brown, Wale, Tyga, 2 Chainz, Wiz Khalifa, French Montana, Nicki Minaj

DJ Khaled, Lil Wayne, Rick Ross, Chris Brown, Wale, Tyga, 2 Chainz, Wiz Khalifa, French Montana, The Game



# Demo