Finding Like-minded Communities of Highly Interactive Individuals on Twitter

Presented by: Amitava Datta
Co-prepared by: Kwan Hui Lim
School of Computer Science and Software Engineering
The University of Western Australia
What is Twitter?
What is Twitter?

- Popular micro-blogging service

Source: http://blog.twitter.com
What is Twitter?

- Popular micro-blogging service
- Tweets of up to 140 characters

Source: http://blog.twitter.com
What is Twitter?

- Popular micro-blogging service
- Tweets of up to 140 characters
- Users follow one another

Source: http://blog.twitter.com
What is Twitter?

- Popular micro-blogging service
- Tweets of up to 140 characters
- Users follow one another
  - May not be reciprocal

Source: http://blog.twitter.com
Why Twitter?

Okay, Marion...
...this Twitter thing has gone far enough.

Source: http://www.bradfitzpatrick.com
Why Twitter?

↳ Large user base

Source: http://www.bradfitzpatrick.com
Why Twitter?

¬ Large user base
  • 500 million users [1]

Source: http://www.bradfitzpatrick.com

Why Twitter?

- Large user base
  - 500 million users [1]
- High activity level


Source: http://www.bradfitzpatrick.com
Why Twitter?

- Large user base
  - 500 million users [1]

- High activity level
  - 2,200 tweets/second [2]

Source: http://www.bradfitzpatrick.com

Part of everyday Life?
NEW OPPORTUNITIES FOR STUDENTS

WHERE IS YOUR HOMEWORK?

DIDN'T YOU GET IT? I TWITTERED IT TO YOU

WILL YOU MARRY ME?

I HAVE TO ASK MY FOLLOWERS

CROWDSOURCING ’09

Source: http://geekandpoke.typepad.com
Why detect like-minded communities?
Why detect like-minded communities?

- Marketing and advertising
Why detect like-minded communities?

- Marketing and advertising
  - Like-minded = Same product interest
Why detect like-minded communities?

➤ Marketing and advertising
  • Like-minded = Same product interest
  • Interactive individuals = Information diffusion
Criteria for community detection
Criteria for community detection

Detect users with the same interest
Criteria for community detection

- Detect users with the same interest
  - Same interest in product area
Criteria for community detection

- Detect users with the same interest
  - Same interest in product area

- Detect users that are well-connected
Criteria for community detection

- Detect users with the same interest
  - Same interest in product area

- Detect users that are well-connected
  - Facilitates word-of-mouth spreading
Definitions
Definitions

Followership link \((i, j)\) : user \(i\) follows user \(j\)
Definitions

Followership link $(i, j)$: user $i$ follows user $j$
Definitions
Definitions

¬ Friendship link $Fr_{i,j} : \text{link} (i, j) = \text{link} (j, i)$
Definitions

Friendship link $F_{r_{i,j}} : link (i, j) = link (j, i)$
Definitions
Definitions

Celebrity: a user with >10,000 followers
Definitions

Celebrity: a user with >10,000 followers
Definitions
Definitions

Interest $Int_{cat}$: No. of celebrities a user follows
Definitions

Interest $Int_{cat}$: No. of celebrities a user follows

Celebrities

Fans

$Int_{sports} = 2$

$Int_{sports} = 4$
Definitions
Definitions

Followership link \((i, j)\) : user \(i\) follows user \(j\)
Definitions

- Followership link \((i, j)\) : user \(i\) follows user \(j\)

- Friendship link \(Fr_{i,j} : link (i, j) = link (j, i)\)
Definitions

- Followership link \((i, j)\) : user \(i\) follows user \(j\)
- Friendship link \(Fr_{i,j}\) : \(link (i, j) = link (j, i)\)
- Celebrity : a user with \(>10,000\) followers
Definitions

- Followership link \((i, j)\) : user \(i\) follows user \(j\)

- Friendship link \(Fr_{i,j} : link (i, j) = link (j, i)\)

- Celebrity : a user with >10,000 followers

- Interest \(Int_{cat} :\) No. of celebrities a user follows
Overall Framework
Overall Framework

1. Identify and classify celebrities
Overall Framework

1. Identify and classify celebrities
   • Representing an interest category
Overall Framework

1. Identify and classify celebrities
   • Representing an interest category

2. Select users with common interest
Overall Framework

1. Identify and classify celebrities
   • Representing an interest category

2. Select users with common interest

3. Detect communities among them
Overall Framework

1. Identify and classify celebrities
   - Representing an interest category

2. Select users with common interest

3. Detect communities among them
   - Based on topological links
Overall Framework

1. Identify and classify celebrities
   • Representing an interest category

2. Select users with common interest

3. Detect communities among them
   • Based on topological links
   • Based on interaction links
Identifying and classifying celebrities
Identifying and classifying celebrities

1. Identify popular twitter celebrities
Identifying and classifying celebrities

1. Identify popular twitter celebrities
   • Based on number of followers

Source: https://twitter.com/
Identifying and classifying celebrities

1. Identify popular twitter celebrities
   - Based on number of followers

2. Classify interest categories of celebrities

Source: https://twitter.com/
Identifying and classifying celebrities

1. Identify popular twitter celebrities
   • Based on number of followers

2. Classify interest categories of celebrities
   • Using information on Wikipedia
Identifying and classifying celebrities
Identifying and classifying celebrities

1. Identify popular twitter celebrities
Identifying and classifying celebrities

1. Identify popular twitter celebrities
   - Based on number of followers
Identifying and classifying celebrities

1. Identify popular twitter celebrities
   - Based on number of followers

The Twitaholic.com Top 100 Twitterholics based on Followers

<table>
<thead>
<tr>
<th>#</th>
<th>Name (Screen Name)</th>
<th>Location</th>
<th>URL</th>
<th>Followers</th>
<th>Following</th>
<th>Updates</th>
<th>Joined</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.</td>
<td>Justin Bieber (justinbieber)</td>
<td>All Around The World</td>
<td><a href="http://www.youtube.com/justin">http://www.youtube.com/justin</a>...</td>
<td>25830123</td>
<td>123216</td>
<td>17446</td>
<td>41 months ago</td>
</tr>
<tr>
<td>4.</td>
<td>Rihanna (rihanna)</td>
<td>LA BABY!</td>
<td><a href="http://www.rihannarow.com">http://www.rihannarow.com</a></td>
<td>23480469</td>
<td>839</td>
<td>6002</td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td>Britney Spears (britneyspears)</td>
<td>Los Angeles, CA</td>
<td><a href="http://facebook.com/britneysp">http://facebook.com/britneysp</a>...</td>
<td>19083490</td>
<td>414028</td>
<td>1507</td>
<td>47 months ago</td>
</tr>
</tbody>
</table>

Source: http://twitaholic.com
Identifying and classifying celebrities
Identifying and classifying celebrities

2. Classify interest categories of celebrities
Identifying and classifying celebrities

2. Classify interest categories of celebrities
   - Using information on Wikipedia
Identifying and classifying celebrities

2. Classify interest categories of celebrities
   • Using information on Wikipedia

Yao Ming

From Wikipedia, the free encyclopedia

This is a Chinese name; the family name is Yao (姚).

Yao Ming (born September 12, 1980) is a retired Chinese professional basketball player who last played for the Houston Rockets of the National Basketball Association (NBA). At the time of his final season, he was the tallest active player in the NBA, at 2.29 m (7 ft 6 in).[1]

Source: http://en.wikipedia.org
Identifying and classifying celebrities

2. Classify interest categories of celebrities

- Using information on Wikipedia

"Occupation" Field

Yao Ming

From Wikipedia, the free encyclopedia

This is a Chinese name; the family name is Yao (姚明).

Yao Ming (born September 12, 1980) is a retired Chinese professional basketball player who last played for the Houston Rockets of the National Basketball Association (NBA). At the time of his final season, he was the tallest active player in the NBA, at 2.29 m (7 ft 6 in).[1]

Source: http://en.wikipedia.org
Identifying and classifying celebrities

2. Classify interest categories of celebrities

- Using information on Wikipedia
Selecting users with common interest
Selecting users with common interest

1. Identify $k$ celebrities, representing an interest category
Selecting users with common interest

1. Identify $k$ celebrities, representing an interest category

2. Retrieve all followers of these $k$ celebrities
Selecting users with common interest

1. Identify $k$ celebrities, representing an interest category
2. Retrieve all followers of these $k$ celebrities
3. Construct Set $P$: users who follow all $k$ celebrities

\[ P = \bigcap \left( \bigcup_{i} \text{link}(i, c_j) \right), \text{ for } 1 \leq j \leq k \]
Selecting users with common interest
Selecting users with common interest

1. Identify $k$ celebrities, representing an interest category
Selecting users with common interest

1. Identify $k$ celebrities, representing an interest category
Selecting users with common interest
Selecting users with common interest

2. Retrieve all followers of these $k$ celebrities
Selecting users with common interest

2. Retrieve all followers of these $k$ celebrities
Selecting users with common interest
Selecting users with common interest

3. Construct Set $P$: users who follow all $k$ celebrities
Selecting users with common interest

3. Construct Set $P$: users who follow all $k$ celebrities
Selecting users with common interest
Selecting users with common interest

3. \[ P = \bigcap \left( \bigcup_{i} \text{link}(i, c_j) \right), \text{for } 1 \leq j \leq k \]
Selecting users with common interest

3. \[ P = \bigcap \left( \bigcup_{i} \text{link}(i, c_j) \right), \text{for } 1 \leq j \leq k \]
Detect communities among Set $P$
Detect communities among Set $P$

1. Retrieve the linkages among Set $P$
Detect communities among Set $P$

1. Retrieve the linkages among Set $P$
   - Based on topological links
Detect communities among Set $P$

1. Retrieve the linkages among Set $P$
   - Based on topological links
   - Based on interaction links
Detect communities among Set $P$

1. Retrieve the linkages among Set $P$
   - Based on topological links
   - Based on interaction links


Retrieve the linkages among Set $P$
Retrieve the linkages among Set $P$

1. Topological links $\rightarrow$ Friendship links
Retrieve the linkages among Set $P$

1. Topological links $\rightarrow$ Friendship links
   - Common Interest Community Detection (CICD)
Retrieve the linkages among Set $P$

1. Topological links $\rightarrow$ Friendship links
   - Common Interest Community Detection (CICD)
   - Reciprocal links reflect real-life relationships
Retrieve the linkages among Set $P$

1. Topological links $\rightarrow$ Friendship links
   - Common Interest Community Detection (CICD)
   - Reciprocal links reflect real-life relationships

2. Interaction links $\rightarrow$ @mention links
Retrieve the linkages among Set $P$

1. Topological links $\rightarrow$ Friendship links
   - Common Interest Community Detection (CICD)
   - Reciprocal links reflect real-life relationships

2. Interaction links $\rightarrow$ @mention links
   - Highly Interactive Community Detection (HICD)
Retrieve the linkages among Set $P$

1. Topological links $\rightarrow$ Friendship links
   - Common Interest Community Detection (CICD)
   - Reciprocal links reflect real-life relationships

2. Interaction links $\rightarrow$ @mention links
   - Highly Interactive Community Detection (HICD)
   - Filtering based on number of @mentions

\[ I_{i,j} = M_{i,j}, \text{ for } i, j \in \mathcal{P} \]
Detect communities using CPM / Infomap
Detect communities using CPM / Infomap

✓ Clique Percolation Method
Detect communities using CPM / Infomap

- Clique Percolation Method
  - Communities are a series of adjacent $k$-cliques (fully interconnected $k$ nodes)
Detect communities using CPM / Infomap

- Clique Percolation Method
  - Communities are a series of adjacent $k$-cliques (fully interconnected $k$ nodes)
Detect communities using CPM / Infomap

Clique Percolation Method

- Communities are a series of adjacent $k$-cliques (fully interconnected $k$ nodes)
Detect communities using CPM / Infomap

- Clique Percolation Method
  - Communities are a series of adjacent $k$-cliques (fully interconnected $k$ nodes)
Detect communities using CPM / Infomap

✓ Clique Percolation Method

• Communities are a series of adjacent $k$-cliques (fully interconnected $k$ nodes)
Detect communities using CPM / Infomap

- Clique Percolation Method
  - Communities are a series of adjacent $k$-cliques (fully interconnected $k$ nodes)
Detect communities using CPM / Infomap

- Clique Percolation Method

  - Communities are a series of adjacent \( k \)-cliques (fully interconnected \( k \) nodes)
Detect communities using CPM / Infomap

- Clique Percolation Method
  - Communities are a series of adjacent $k$-cliques (fully interconnected $k$ nodes)
Detect communities using CPM / Infomap

<UFunction

• Communities are a series of adjacent $k$-cliques (fully interconnected $k$ nodes)
Detect communities using CPM / Infomap
Detect communities using CPM / Infomap

Infomap Algorithm
Detect communities using CPM / Infomap

Infomap Algorithm

• Communities are nodes where a random walker spends most time traversing
Detect communities using CPM / Infomap

Infomap Algorithm

- Communities are nodes where a random walker spends most time traversing
Detect communities using CPM / Infomap

Infomap Algorithm

• Communities are nodes where a random walker spends most time traversing
Detect communities using CPM / Infomap

Infomap Algorithm

- Communities are nodes where a random walker spends most time traversing.
Detect communities using CPM / Infomap

Infomap Algorithm

- Communities are nodes where a random walker spends most time traversing
Detect communities using CPM / Infomap

¬ Infomap Algorithm

• Communities are nodes where a random walker spends most time traversing
Detect communities using CPM / Infomap

Infomap Algorithm

• Communities are nodes where a random walker spends most time traversing
What kind of communities are formed by users with common interest?
Dataset
Dataset

- Used Twitter dataset collected by Kwak et al. [5]

Dataset

- Used Twitter dataset collected by Kwak et al. [5]
  - Collected in Jun 2009

Dataset

- Used Twitter dataset collected by Kwak et al. [5]
  - Collected in Jun 2009
  - 41.7 million users

Dataset

Used Twitter dataset collected by Kwak et al. [5]

- Collected in Jun 2009
- 41.7 million users
- 1.47 billion links

Investigating Common Interest
Investigating Common Interest

Identifying popular interest categories
Investigating Common Interest

Identifying popular interest categories

1. Chose the top 100 celebrities
Investigating Common Interest

Identifying popular interest categories

1. Chose the top 100 celebrities
   – Based on number of followers
Investigating Common Interest

Identifying popular interest categories

1. Chose the top 100 celebrities
   - Based on number of followers

2. Determined the categories they represent
Investigating Common Interest

- Identifying popular interest categories
  1. Chose the top 100 celebrities
     - Based on number of followers
  2. Determined the categories they represent
  3. Selected the most popular categories
Investigating Common Interest

⚠ Identifying popular interest categories

1. Chose the top 100 celebrities
   – Based on number of followers

2. Determined the categories they represent

3. Selected the most popular categories
   – Based on frequency of occurrence
Investigating Common Interest
Investigating Common Interest

Popular interest categories
Investigating Common Interest

Popular interest categories

- Film & TV, Music, Hosting, News, Blogging

- Film & TV (21)
- Online Media (11)
- News (7)
- Commerce (7)
- Comedian (4)
- Author (4)
- Entrepreneur (3)
- Government (2)
- Magazine (2)
- Medicine (1)
- Gaming (1)
- Search Engine (1)
- Music (20)
- Hosting (8)
- Blogging (7)
- Politics (4)
- Sports (4)
- Journalist (3)
- Twitter (3)
- Model (2)
- F&B (1)
- Film Maker (1)
- Comics (1)
Investigating Common Interest
Investigating Common Interest

Identified six celebrities to represent each category
Investigating Common Interest

- Identified six celebrities to represent each category

<table>
<thead>
<tr>
<th>Film &amp; TV</th>
<th>Music</th>
<th>Hosting</th>
<th>News</th>
<th>Blogging</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ashton Kutcher</td>
<td>Britney Spears</td>
<td>Ellen DeGeneres</td>
<td>CNN Breaking News</td>
<td>Perez Hilton</td>
</tr>
<tr>
<td>Demi Moore</td>
<td>John Mayer</td>
<td>Oprah Winfrey</td>
<td>The New York Times</td>
<td>Mashable</td>
</tr>
<tr>
<td>Jimmy Fallon</td>
<td>Sean John Combs</td>
<td>Ryan Seacrest</td>
<td>The Onion</td>
<td>Dooce</td>
</tr>
<tr>
<td>Miley Cyrus</td>
<td>Miley Cyrus</td>
<td>Jimmy Fallon</td>
<td>Good Morning America</td>
<td>Ana Marie Cox</td>
</tr>
<tr>
<td>Perez Hilton</td>
<td>Coldplay</td>
<td>Chelsea Handler</td>
<td>ABC News Nightline</td>
<td>Brandon Mendelson</td>
</tr>
<tr>
<td>50 Cents</td>
<td>DeAndre Cortez Way</td>
<td>Veronica Belmont</td>
<td>Breaking News</td>
<td>Sockington</td>
</tr>
</tbody>
</table>
Investigating Common Interest
Investigating Common Interest

Trade-off between interest level and user group size
Investigating Common Interest

△ Trade-off between interest level and user group size

• Higher interest level, smaller Set $P$
Investigating Common Interest
Investigating Common Interest

- Comparing interest groups to control group
Investigating Common Interest

- Comparing interest groups to control group
  - More communities found

![Graph of Total Communities Detected](image1)

![Graph of Largest Community Detected](image2)
Investigating Common Interest

- Comparing interest groups to control group
  - More communities found
  - Largest community is significantly larger
Investigating Common Interest
Investigating Common Interest

- Comparing interest groups to control group
Investigating Common Interest

- Comparing interest groups to control group
  - Higher clustering coefficient or link degree

<table>
<thead>
<tr>
<th>Category</th>
<th>Control</th>
<th>Film&amp;TV</th>
<th>Music</th>
<th>Hosting</th>
<th>News</th>
<th>Blogging</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg. Path Length</td>
<td>2.83</td>
<td>3.03</td>
<td>2.82</td>
<td>3.09</td>
<td>3.35</td>
<td>3.09</td>
</tr>
<tr>
<td>Avg. Clustering Coeff.</td>
<td>0.60</td>
<td>0.62</td>
<td>0.63</td>
<td>0.59</td>
<td>0.58</td>
<td>0.62</td>
</tr>
<tr>
<td>Diameter</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>Avg. Degree</td>
<td>7.81</td>
<td>6.80</td>
<td>7.29</td>
<td>8.17</td>
<td>9.15</td>
<td>7.51</td>
</tr>
</tbody>
</table>
What are the effects of deepening or specializing interest?
Specializing of Interest
Specializing of Interest

- Identified six celebrities to represent the specialized interest group (Country Music)
Specializing of Interest

✓ Identified six celebrities to represent the specialized interest group (Country Music)

• Based on winners of Country Music Awards
Specializing of Interest

- Identified six celebrities to represent the specialized interest group (Country Music)
  - Based on winners of Country Music Awards

- Control group is the general interest group (Music)
Specializing of Interest
Specializing of Interest

- Comparing specialized interest (Country Music) to general interest (Music)
Specializing of Interest

Comparer specialized interest (Country Music) to general interest (Music)

• More cohesive and higher reciprocality

<table>
<thead>
<tr>
<th>Category</th>
<th>General (Music)</th>
<th>Specialized (Country)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg. Path Length</td>
<td>2.82</td>
<td>2.10</td>
</tr>
<tr>
<td>Avg. Clustering Coefficient</td>
<td>0.63</td>
<td>0.76</td>
</tr>
<tr>
<td>Diameter</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>Avg. Degree</td>
<td>7.29</td>
<td>5.52</td>
</tr>
<tr>
<td>Reciprocality</td>
<td>18.2%</td>
<td>20.1%</td>
</tr>
</tbody>
</table>
Deepening of Interest
Deepening of Interest

Deepening interest = increasing $\text{Int}_{\text{cat}}$ value
Deepening of Interest

Deepening interest = increasing $\text{Int}_{\text{cat}}$ value

- More likely to form communities
Deepening of Interest
Deepening of Interest

∴ With increasing $\text{Int}_{\text{cat}}$ value
Deepening of Interest

- With increasing $\text{Int}_{\text{cat}}$ value
  - Increasing clustering coefficient

![Graph showing average clustering coefficient and average path length vs. number of country singers followed.](image-url)
Deepening of Interest

- With increasing $\text{Int}_{\text{cat}}$ value
  - Increasing clustering coefficient
  - Decreasing path length

![Average Clustering Coefficient](image1)

![Average Path Length](image2)
Deepening of Interest
Deepening of Interest

\[ \forall \text{Across all } \text{Int}_{\text{cat}} \text{ value} \]
Deepening of Interest

- Across all \( \text{Int}_{\text{cat}} \) value
  - Characteristics of scale-free networks

![Degree CCDF of Largest Community with \( \text{Int}_{\text{Country}} = 2 \)](image1)

![Degree CCDF of Largest Community with \( \text{Int}_{\text{Country}} = 4 \)](image2)
What are the effects of using interaction links to construct the communities?
Dataset
Dataset

¬ Collected using Twitter API
Dataset

Collected using Twitter API

- From 17 Nov 11 to 14 Jan 12
Dataset

Collected using Twitter API

- From 17 Nov 11 to 14 Jan 12
- 17,941 users
Dataset

Collected using Twitter API

- From 17 Nov 11 to 14 Jan 12
- 17,941 users
- 1.9 million tweets
Representative Celebrities
Representative Celebrities

 Identified six celebrities to represent each category
Representative Celebrities

Identified six celebrities to represent each category

<table>
<thead>
<tr>
<th>Country Music</th>
<th>Tennis</th>
<th>Dallas Mavericks</th>
<th>Chicago Bulls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Taylor Swift</td>
<td>Serena Williams</td>
<td>Lamar Odom</td>
<td>C. J. Watson</td>
</tr>
<tr>
<td>Brad Paisley</td>
<td>Rafael Nadal</td>
<td>Jason Terry</td>
<td>Carlos Boozer</td>
</tr>
<tr>
<td>Blake Shelton</td>
<td>Andy Murray</td>
<td>Dirk Nowitzki</td>
<td>Luol Deng</td>
</tr>
<tr>
<td>Miranda Lambert</td>
<td>Novak Djokovic</td>
<td>Shawn Marion</td>
<td>Kyle Korver</td>
</tr>
<tr>
<td>Kenny Chesney</td>
<td>Caroline Wozniacki</td>
<td>Vince Carter</td>
<td>Taj Gibson</td>
</tr>
<tr>
<td>Keith Urban</td>
<td>Venus Williams</td>
<td>Jason Kidd</td>
<td>Ronnie Brewer</td>
</tr>
<tr>
<td>Martina McBride</td>
<td>Andy Roddick</td>
<td>Brian Cardinal</td>
<td>Jimmy Butle</td>
</tr>
<tr>
<td>Tim McGraw</td>
<td>Sania Mirza-Malik</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Toby Keith</td>
<td>Kim Clijsters</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Interaction VS Topological Links
Interaction VS Topological Links

 Comparing HICD (Interaction) to CICD (Topological)
Interaction VS Topological Links

- Comparing HICD (Interaction) to CICD (Topological)
  - More communities found (Infomap)
Interaction VS Topological Links

- Comparing HICD (Interaction) to CICD (Topological)
  - More communities found (Infomap)
  - Largest community is smaller (except Bulls)
Interaction VS Topological Links
Interaction VS Topological Links

✓ Comparing HICD (Interaction) to CICD (Topological)
Interaction VS Topological Links

- Comparing HICD (Interaction) to CICD (Topological)
  - Higher clustering coefficient
Interaction VS Topological Links

Compared HICD (Interaction) to CICD (Topological)

- Higher clustering coefficient
- Less average degrees

![Graphs showing Clustering Coefficient and Average Degree for different categories: Country Music, Tennis, Mavericks, Bulls. The graphs compare HICD and CICD.]
Increasing Threshold $T$ of $I_{i,j}$
Increasing Threshold $T$ of $I_{i,j}$

Smaller but more cohesive communities
Increasing Threshold $T$ of $I_{i,j}$

- Smaller but more cohesive communities
  - Increasing clustering coefficient

<table>
<thead>
<tr>
<th>Threshold $T$ of $I_{i,j}$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>No of Nodes</td>
<td>474</td>
<td>313</td>
<td>188</td>
<td>108</td>
<td>70</td>
<td>42</td>
</tr>
<tr>
<td>Avg. Path Length</td>
<td>2.84</td>
<td>2.63</td>
<td>2.64</td>
<td>2.52</td>
<td>2.68</td>
<td>2.49</td>
</tr>
<tr>
<td>Avg. Clustering Coefficient</td>
<td>0.70</td>
<td>0.72</td>
<td>0.74</td>
<td>0.77</td>
<td>0.75</td>
<td>0.77</td>
</tr>
<tr>
<td>Diameter</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>5</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Average Degree</td>
<td>6.20</td>
<td>6.27</td>
<td>5.67</td>
<td>5.28</td>
<td>4.66</td>
<td>4.52</td>
</tr>
</tbody>
</table>
Content of Tweets
Content of Tweets

Top 10 #hashtags

• More #hashtags about common interest
Content of Tweets

⚠️ Top 10 #hashtags

- More #hashtags about common interest

<table>
<thead>
<tr>
<th>Set P</th>
<th>Com\textsubscript{CICD}</th>
<th>Com\textsubscript{HICD}</th>
</tr>
</thead>
<tbody>
<tr>
<td>#FF</td>
<td>#FF</td>
<td>#FF</td>
</tr>
<tr>
<td>#fb</td>
<td>#fb</td>
<td>#CMAawards*</td>
</tr>
<tr>
<td>#NowPlaying</td>
<td>#NowPlaying</td>
<td>#nowplaying</td>
</tr>
<tr>
<td>#nowplaying</td>
<td>#CMAawards*</td>
<td>#fb</td>
</tr>
<tr>
<td>#CMAawards*</td>
<td>#nowplaying</td>
<td>#PeoplesChoice</td>
</tr>
<tr>
<td>#iTunes</td>
<td>#jesustweeters</td>
<td>#cmchat*</td>
</tr>
<tr>
<td>#PeoplesChoice</td>
<td>#iTunes</td>
<td>#ff</td>
</tr>
<tr>
<td>#ff</td>
<td>#concert*</td>
<td>#CMTAOTY*</td>
</tr>
<tr>
<td>#jesustweeters</td>
<td>#DT</td>
<td>#countryartist*</td>
</tr>
<tr>
<td>#concert</td>
<td>#Nashville</td>
<td>#ACAs*</td>
</tr>
</tbody>
</table>

* Related to Country Music
Content of Tweets
Content of Tweets

Top 10 @mentions

• More @mentions about celebrities related to common interest
Content of Tweets

Top 10 @mentions

- More @mentions about celebrities related to common interest

<table>
<thead>
<tr>
<th>Set P</th>
<th>Com$_{CICD}$</th>
<th>Com$_{HICD}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>youtube</td>
<td>youtube</td>
<td>blakeshelton*</td>
</tr>
<tr>
<td>blakeshelton*</td>
<td>blakeshelton*</td>
<td>davidnail*</td>
</tr>
<tr>
<td>YouTube</td>
<td>YouTube</td>
<td>Miranda_Lambert*</td>
</tr>
<tr>
<td>GetGlue</td>
<td>taylorswift13*</td>
<td>ladyantebellum*</td>
</tr>
<tr>
<td>taylorswift13*</td>
<td>Miranda_Lambert*</td>
<td>GetGlue</td>
</tr>
<tr>
<td>justinbieber</td>
<td>davidnail*</td>
<td>ScottyMcCreery*</td>
</tr>
<tr>
<td>Miranda_Lambert*</td>
<td>GetGlue</td>
<td>ChrisYoungMusic*</td>
</tr>
<tr>
<td>ScottyMcCreery*</td>
<td>BradPaisley*</td>
<td>Lauren_Alaina*</td>
</tr>
<tr>
<td>BradPaisley*</td>
<td>JimmyWayne*</td>
<td>taylorswift13*</td>
</tr>
<tr>
<td>jakeowen*</td>
<td>jakeowen*</td>
<td>SUGARLAND4EVER</td>
</tr>
</tbody>
</table>

* Related to Country Music
Trends in Tweeting
Trends in Tweeting

 Usuarios with high tweeting activity selected
Trends in Tweeting

_users with high tweeting activity selected_

- HICD (Interaction) > CICD (Topological) > Set $P$
Temporal Analysis of Links
Temporal Analysis of Links

.Users more active in creating than deleting links
Temporal Analysis of Links

- Users more active in creating than deleting links
  - HICD (Interaction) > CICD (Topological) > Set P

Time Analysis of Created Links

Time Analysis of Deleted Links
Choice of Links
Choice of Links

- Advantages and disadvantage
## Choice of Links

### Advantages and disadvantage

<table>
<thead>
<tr>
<th>Link Type</th>
<th>Topological (CICD)</th>
<th>Interaction (HICD)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pros</strong></td>
<td>• Easier to collect data (followership links)</td>
<td>• More active community</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Frequently talks about common interest</td>
</tr>
<tr>
<td><strong>Cons</strong></td>
<td>• Community may not communicate frequently about common interest</td>
<td>• Requires data (tweets) gathering over time</td>
</tr>
</tbody>
</table>
Conclusion
Conclusion

✓ Presented an approach to detect communities that share common interest
Conclusion

- Presented an approach to detect communities that share common interest
- Studied the effects of specializing and deepening of interest on community structure
Conclusion

- Presented an approach to detect communities that share common interest
- Studied the effects of specializing and deepening of interest on community structure
- Studied the topological and interaction patterns of these communities
Conclusion
Conclusion

✔ Applicable to other social networks (e.g. Facebook)
Conclusion

- Applicable to other social networks (e.g. Facebook)
  - Celebrity = Pages of the celebrity
Conclusion

✓ Applicable to other social networks (e.g. Facebook)
  • Celebrity = Pages of the celebrity
  • Topological Links = User “likes” on that page
Conclusion

- Applicable to other social networks (e.g. Facebook)
  - Celebrity = Pages of the celebrity
  - Topological Links = User “likes” on that page
  - Interaction Links = Posts on user’s wall
Thank You

Amitava Datta
amitava.datta@uwa.edu.au