Project 1:

IS620 Student: Neil Acampa Date: 10/01/16

**The data was found at www.imdb.com**

The Movie database consists of movie title, length in minutes, year released, average rating and movie genre.

Two nodes were created, one **Movie** with attributes movie title, year and rating

The second is **Genre** consisting of 7 Genres (Action, Animation, Comedy, Drama, Documentary, Romance and Short.

An **Edge i**s created linking Movie and Genre based on a 0,1 entry in file with Edge properties: year and rating

Two centrality measures are calculated: **Degree and Eigenvector**

**Degree Centrality**: Is a measure of the number of edges or links incident upon a node.

In Degree Centrality a nodes importance increases with the number of in or outbound edges. Inbound edges indicate popularity, output indicate influence. A specific node: Genre is important because there are many movies linked to that Genre

**Eigenvector Centrality**: Measures the importance of a node based on its connectivity to other important nodes.

A nodes status increases by being connected to high status members and conversely decreases by being connected to low status members.

Example: if 'Comedy' has high degree centrality and 'Romance' has low degree centrality. A database with many 'Romantic Comedies' may have a higher relative 'Romance' eigenvector measure by virtue of its association with the 'Comedy'.

Shorts may also have a higher eigenvector measure due to the association with Comedy or Drama

Example of a **Movie node**: {'rating': '4.3', 'movietitle': "Aujourd'hui ou jamais", 'movieyear': '1998'}

Example of a **Genre node**: {'genre': 'drama'}

Example of an **Edge**: {'Drama': {'year\_released': '1998', 'avg\_user\_rating': '4.3'}}

Example of a **Movie node**: {'rating': '5.6', 'movietitle': 'Mrs. Barrington', 'movieyear': '1974'}

Example of an **Edge**: {'Drama': {'year\_released': '1974', 'avg\_user\_rating': '5.6'}}

**Statistics:**

**One sided T-tests have been run at a 99 percent CI alpha of .01 with CV: 3.143**

**and at a 95 percent CI alpha of .05 with CV: 1.943**

Statistics have been gathered on a random sample of **150** data points.

on a random sample of **300** data points

and for the entire db of **58,789** data points

**Genre Statistics: One sided T-Test at a 99 percent CI for 150 records**

Avg Degree Centrality: **25.29** Stdev: **18.10** Standard Error: **6.84**

Avg Eigen Centrality: **0.20** Stdev: **0.18** Standard Error: **0.07**

Critical Value: **3.143**

99 percent CI for Degree is**: Lower (3.78) Upper (46.79)**

99 percent CI for Eigen Centrality is**: Lower (-0.01) Upper (0.41)**

**Significant “Hub” Genre at .01 alpha indicated by '\*\*\*'**

**Genre Degree Deg Per Sig Eign Sig**

Action 13 0.09 0.08

Animation 7 0.05 0.02

**Comedy 44 0.29 0.38 (Increased 9%)**

**Drama 55 0.37 \*\*\* 0.50 \*\*\* (Increased 13%)**

Documentary 11 0.07 0.05

Romance 10 0.07 0.06

**Short 37 0.25 0.31 (Increased by 6%)**

**Genre Statistics: One sided T-Test at a 95 percent CI for 150 records**

Avg Degree Centrality: **25.29** Stdev: **18.10** Standard Error: **6.84**

Avg Eigen Centrality: **0.20** Stdev: **0.18** Standard Error**: 0.07**

Critical Value: **1.943**

95 percent CI for Degree is**: Lower (11.99) Upper (38.58)**

95 percent CI for Eigen Centrality is: **Lower (0.07) Upper (0.33)**

Significant Hub Genre's at .05 alpha indicated by '\*\*'

**Genre Degree Deg Per Sig Eign Sig**

Action 13 0.09 0.08

Animation 7 0.05 0.02

**Comedy 44 0.29 \*\* 0.38 \*\***

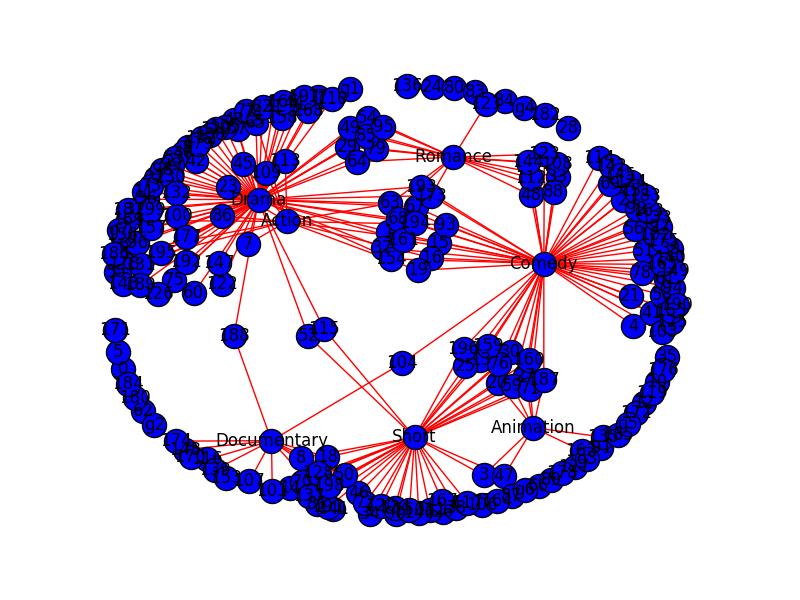
Documentary 11 0.07 0.05

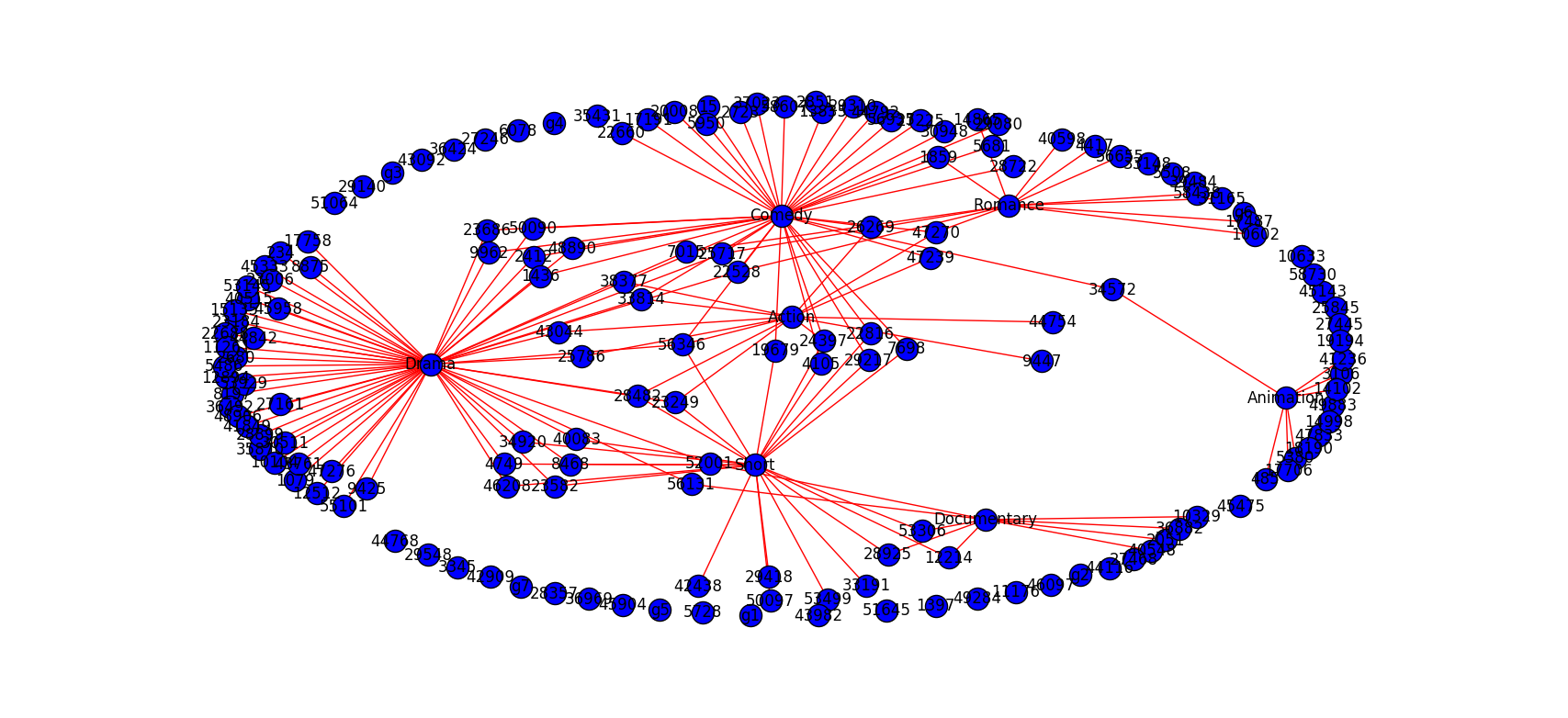
**Drama 55 0.37 \*\* 0.50 \*\***

Romance 10 0.07 0.06

**Short 37 0.25 0.31 (Increased by 6%)**

**Movie Genre Graph for 150 records**

****

****

**Genre Statistics: One sided T-Test at a 99 percent CI for 300 records**

Avg Degree Centrality: 45.57 Stdev: 36.85 Standard Error: 13.93

Avg Eigen Centrality: 0.19 Stdev: 0.19 Standard Error: 0.07

Critical Value**: 3.143**

99 percent CI for Degree is**: Lower (1.79) Upper (89.35)**

99 percent CI for Eigen Centrality is: **Lower (-0.04) Upper (0.41)**

Significant Hub Genre at .01 alpha indicated by '\*\*\*'

**Genre Degree Deg Per Sig Eign Sig**

Action 27 0.09 0.08

Animation 11 0.04 0.00

Comedy 85 0.28 0.39

Drama 117 0.39 \*\*\* 0.56 \*\*\*

Documentary 16 0.05 0.03

Romance 27 0.09 0.12

Short 36 0.12 0.13

**Genre Statistics: One sided T-Test at a 95 percent CI for 300 records**

Avg Degree Centrality**: 45.57** Stdev: **36.85** Standard Error: **13.93**

Avg Eigen Centrality: **0.19** Stdev: **0.19** Standard Error: **0.07**

Critical Value: **1.943**

95 percent CI for Degree is: **Lower (18.51) Upper (72.64)**

95 percent CI for Eigen Centrality is: **Lower (0.05) Upper (0.33)**

Significant Hub Genre's at .05 alpha indicated by '\*\*'

**Genre Degree Deg Per Sig Eign Sig**

Action 27 0.09 0.08

Animation 11 0.04 0.00

Comedy 85 0.28 \*\* 0.39 \*\*

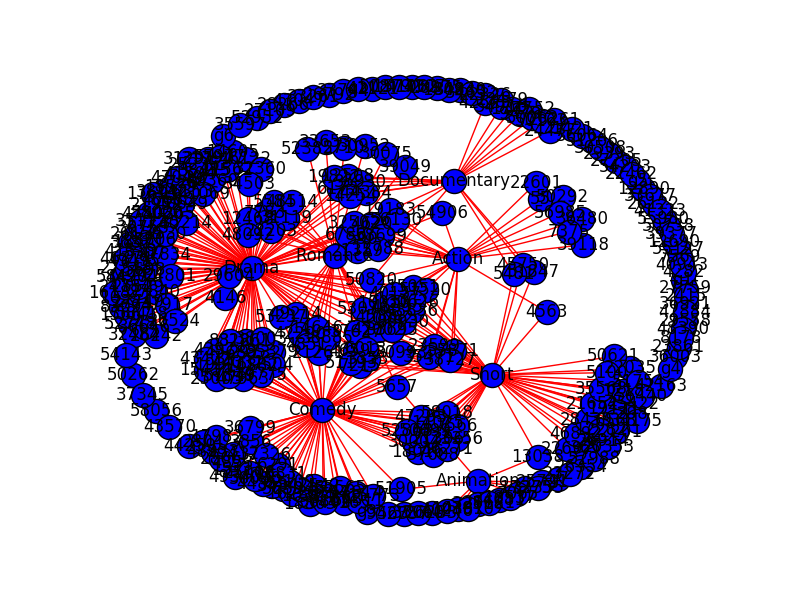
Drama 117 0.39 \*\* 0.56 \*\*

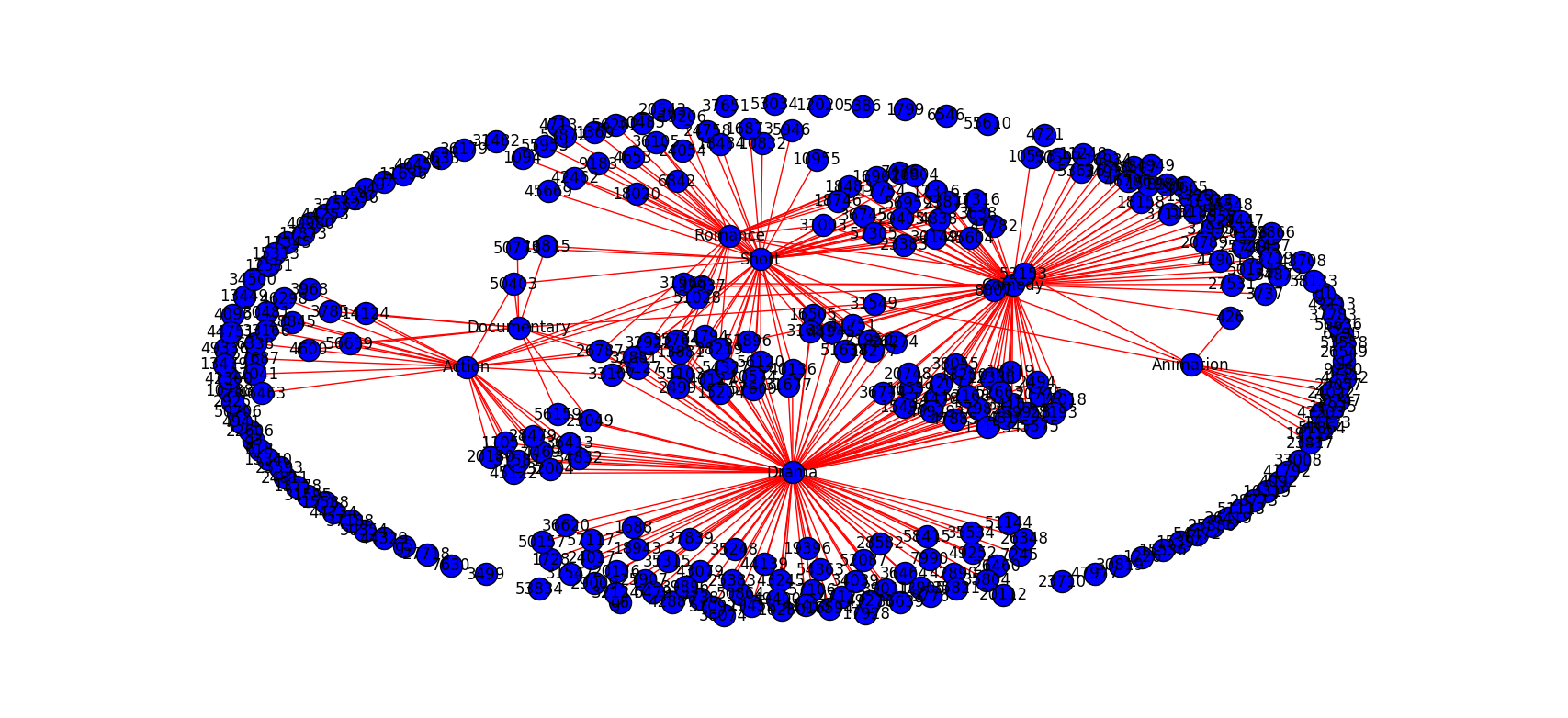
Documentary 16 0.05 0.03

Romance 27 0.09 0.12

Short 36 0.12 0.13

**Movie Genre Graph for 300 records**



****

**Genre Statistics: One sided T-Test at a 99 percent CI for 58789 records**

Avg Degree Centrality**: 9304.86** Stdev: **6840.55** Standard Error: **2585.49**

Avg Eigen Centrality: **0.19** Stdev: **0.19** Standard Error**: 0.07**

Critical Value**: 3.143**

99 percent CI for Degree is: **Lower (1178.68) Upper (17431.04)**

99 percent CI for Eigen Centrality is**: Lower (-0.03) Upper (0.41)**

Significant Hub Genre at .01 alpha indicated by '\*\*\*'

**Genre Degree Deg Per Sig Eign Sig**

Action 4688 0.08 0.07

Animation 3690 0.06 0.03

Comedy 17271 0.29 0.40

Drama 21811 0.37 \*\*\* 0.54 \*\*\*

Documentary 3472 0.06 0.04

Romance 4744 0.08 0.08

**Short 9458 0.16 0.19 (Increased by 3%)**

**Genre Statistics: One sided T-Test at a 95 percent CI for 58789 records**

Avg Degree Centrality: **9304.86** Stdev: **6840.55** Standard Error**: 2585.49**

Avg Eigen Centrality**: 0.19** Stdev: **0.19** Standard Error**: 0.07**

Critical Value**: 1.943**

95 percent CI for Degree is**: Lower (4281.26) Upper (14328.46**)

95 percent CI for Eigen Centrality is: **Lower (0.06) Upper (0.33)**

Significant Genre's at .05 alpha indicated by '\*\*'

**Genre Degree Deg Per Sig Eign Sig**

Action 4688 0.08 0.07

Animation 3690 0.06 0.03

**Comedy 17271 0.29 \*\* 0.40 \*\***

**Drama 21811 0.37 \*\* 0.54 \*\***

Documentary 3472 0.06 0.04

Romance 4744 0.08 0.08

**Short**  9458 **0.16** **0.19 (Increased by 3%)**