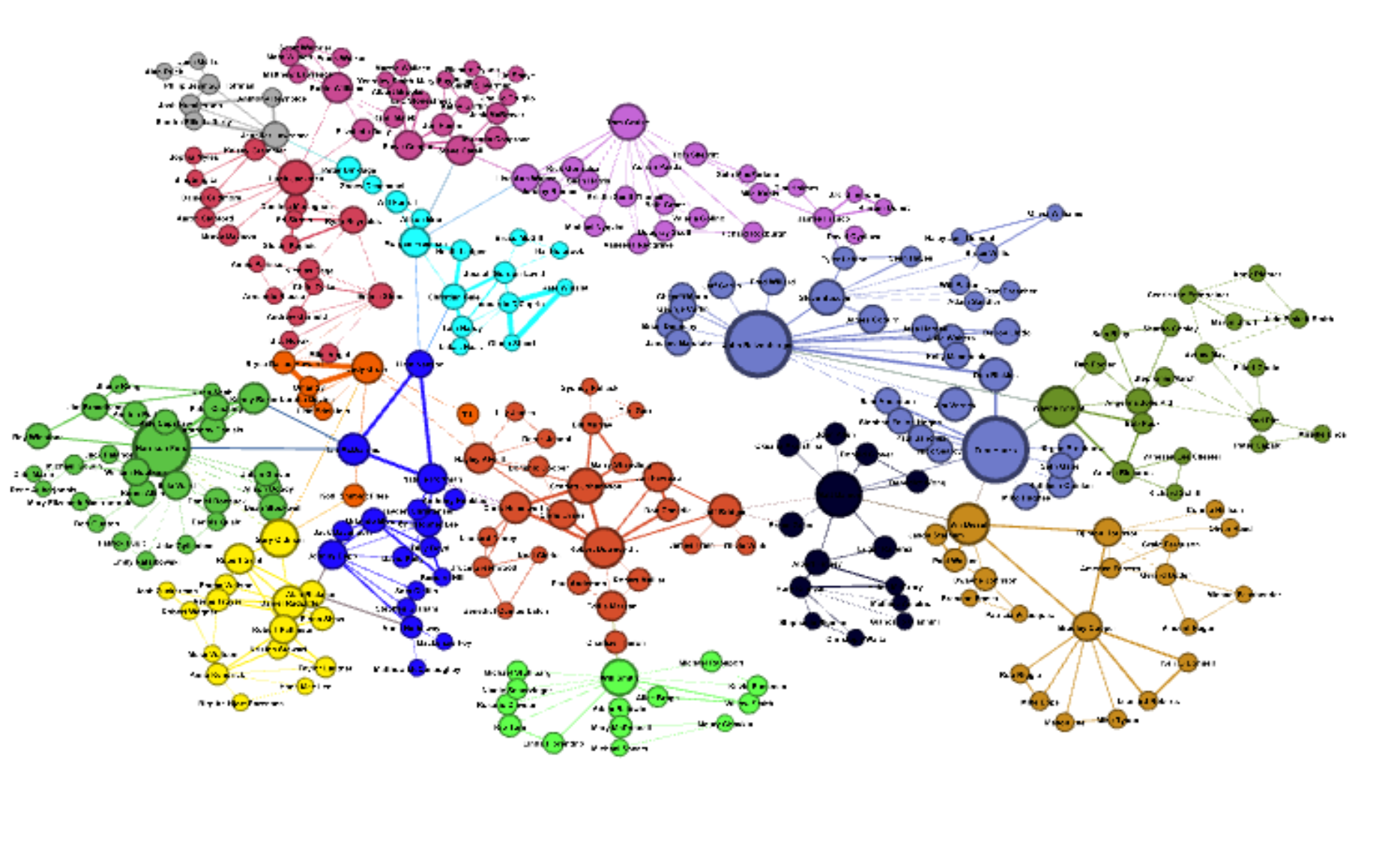
**IMDB Actor Network Analysis**

By: Andy Dotter, Chris Nobblitt, Jake Lasky, Kyuhyun Jeong

**Data:** We used a dataset containing information pulled from IMDB. Our dataset consisted of 5000 movies, the 3 biggest actors by role in each of those movies, the IMDB score, the gross revenue from the movie along with other various attributes. For the purpose of our Network Analysis we isolated to only look at the movie titles, actors, IMDB scores and gross revenue for each movie.

**Problem Statement:** Our goal was to analyze IMDB data to find anything interesting about the relationship of actors who played in the same movies. Particularly how did these actors group together in terms of revenue and IMDB score, who were the most influential actors, and were their clear clusters of main actors and their supporting cast.

**Graph 1:** Actors relationship for the largest component of a social network graph for the top 250 movies ranked by Gross Revenue. Actors are Nodes, Edges are weighted based on gross revenue from the movies the actors were in.



**Clusters using Modularity**

When calculating the modularity maximization with edges weighted by gross revenue for movies and actors. When we clustered we ended up with 15 clusters. Each of these clusters had a main actor in each group, the main actors for each group can be seen in the network graph by the largest node in each cluster (color). These actors are:

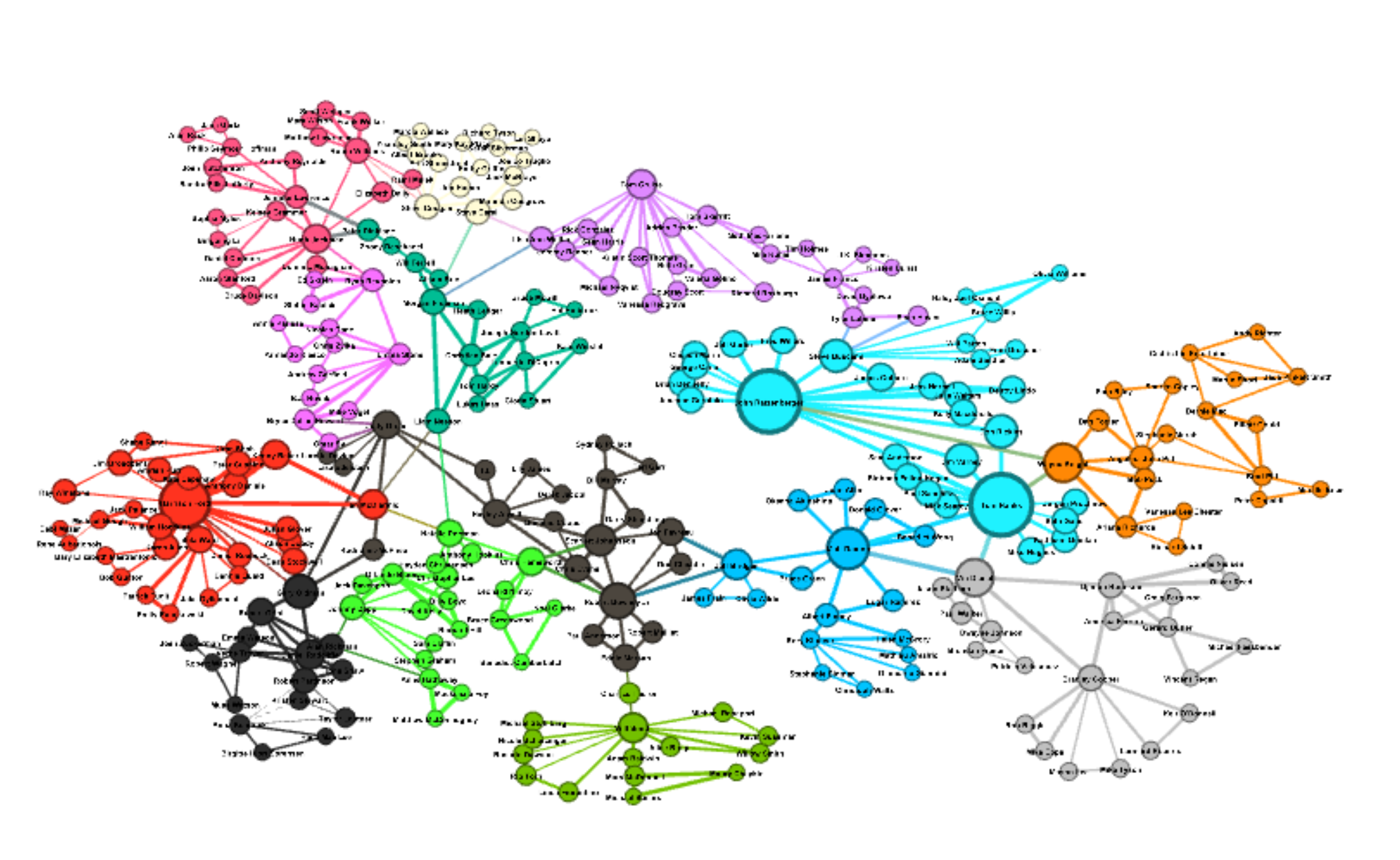
1. Matt Damon
2. Gary Oldman
3. Ian McDiarmid
4. John Ratzenberger
5. Hugh Jackman
6. Robert Downey Jr.
7. Morgan Freeman
8. Jennifer Lawrence
9. Vin Diesel
10. Judy Greer
11. Robin Williams
12. Tom Cruise
13. Will Smith
14. Harrison Ford
15. Wayne Knight

**Most influential by profit**

We ran an Eigenvector Centrality on my network analysis with the weight of the edges as the gross revenue from the associated movies and actors. The actors that are most influential from a gross revenue standpoint are:

* + 1. John Ratzenburger (1.0)
    2. Tom Hanks (0.993)
    3. Harrison Ford (0.828)
  1. The most influential actor that is part of only one triangle is Don Rickles who is part of both John Ratzenburg and Tom Hanks triangle and no other actors. This movie is Toy Story 3.

**Graph 2:** Actors relationship for the largest component of a social network graph for the top 250 movies ranked by Gross Revenue. Actors are Nodes, Edges are weighted based on the IMDB rating of the movie that the actors were in.



**Clusters using Modularity**

When calculating the modularity maximization with edges weighted by IMDB score for movies and actors. When we clustered we ended up with 14 clusters. Each of these clusters had a main actor in each group, the main actors for each group can be seen in the network graph by the largest node in each cluster (color). These actors are:

1. Chris Hemsworth
2. Scarlett Johansson
3. Matt Damon
4. Ryan Reynolds
5. Wayne Knight
6. Hugh Jackman
7. Vin Diesel
8. Morgan Freeman
9. Gary Oldman
10. Steve Carell
11. Harrison Ford
12. Robert Downey Jr.
13. Tom Cruise
14. John Ratzenberger

**Most influential by profit**

We ran an Eigenvector Centrality on my network analysis with the weight of the edges as the IMDB rating from the associated movies and actors. The actors that are most influential from a IMDB standpoint are:

* + 1. John Ratzenburg (1.0)
    2. Tom Hanks (0.987)
    3. Harrison Ford (0.73)

Again, the most influential actor that is part of only one triangle is Don Rickles who is part of both John Ratzenburger and Tom Hanks triangle and no other actors. This movie is Toy Story 3.

**Conclusion & Next Steps:**

As you will see both graphs are similar when we analyzed them based on Gross Revenue or by IMDB rating. Some of the similarities of these graphs is due to isolating for the largest component from the top 250 movies based on revenue. If we expanded the number of movies we included and then focused on the largest component we may see a larger difference when we weight the edges by gross revenue or IMDB score. One thing that is interesting to note is that when we look at the individual with the highest eigenvector centrality in each group, the over lap between the graphs is approximately 60%, i.e. 60% of the actors that had the highest eigenvector centrality based on gross revue also had it the highest for IMDB rating. One thing is clear, there is a small number of actors that have a large influence on the success of a movie.