THE SOCIAL LIFE OF ORPHANS:
APPLYING ECONOMIC ACTIVITY AND GRAPH ANALYSIS TO ORPHAN WORK SEARCHES

Bruce Nash

Bruce.Nash@OpusData.com
Nash Information Services

Web Services

Data Collection and Classification

Analysis

Data Services

Consulting
Measuring Economic Activity

- All works experience a lifecycle of economic activity
- For films
  - Theatrical release
  - Home video, Pay-per-view, hotels, airlines
  - Cable TV
  - Network TV
  - Eventually, recirculation (resale of DVDs etc.)
Spot the Orphans

Economic activity level has implications for the likelihood that a work is orphaned/public domain etc.

Significant economic activity strongly implies that a work is being exploited by a rights holder.

An economic activity search can reduce the chances of a false positive.

$p(\text{Orphan}) = f(\text{Economic Activity})$
Creative Social Graphs

Steven Spielberg

$57,854,992

Catch Me if You Can .......... $71,503,293
Saving Private Ryan .......... $67,321,927
The Terminal .......... $39,555,810

Tom Hanks
Analyzing Creative Social Graphs

- Applying the PageRank* algorithm identifies largest contributors to economic activity

Analyzing Creative Social Graphs

- Applying the PageRank* algorithm identifies largest contributors to economic activity

Movie Industry Power Ranks, April, 2012

Based on annual contribution to domestic box office receipts. (Read more about our methodology here.)

1. Steven Spielberg □ □ □ □ □ □ $13,132,324
   Director: War of the Worlds, War Horse, The Adventures of Tintin, more...
   Also Executive Producer, Producer, Screenwriter
   Movie profile: Original Screenplay, Science Fiction, Live Action, Action, PG-13
   Major Collaborators: Kathleen Kennedy, Tom Cruise, Ian Bryce, Kyle Chandler, Tom Hanks, John Williams

2. Johnny Depp □ □ □ □ □ □ $10,438,431
   Actor: Alice in Wonderland, Pirates of the Caribbean: The Curse of the Black Pearl, more...
   Also Producer
   Movie profile: Based on Theme Park Ride, Fantasy, Live Action, Adventure, PG-13
   Major Collaborators: Tim Burton, Gore Verbinski, Helena Bonham Carter, Mia Wasikowska, Richard D. Zanuck, Mike Stenson

3. John Lasseter □ □ □ □ □ □ $10,185,533
   Executive Producer: Up, Tangled, Finding Nemo, Tangled, WALL-E, more...
   Also Director, Story
   Movie profile: Original Screenplay, Kids Fiction, Digital Animation, Adventure, G
   Major Collaborators: Andrew Stanton, Bob Peterson, Owen Wilson, Darla Anderson, Brad Bird, John Ratzenberger

Adding Rights Holders (and other metadata)

- Black and White
- Digital Animation
- Environmental Documentary
Rights Holder Prediction using nRANK

Johnny Depp

Tim Burton

Pirate Movies

Walt Disney Pictures

Digital Animation

John Lasseter

* See also SimRank++: I. Antonellis, H. Garcia-Molina and C.-C. Chang. Simrank++: Query Rewriting through Link Analysis of the Click Graph. In VLDB '08: Proceedings of the 34th International Conference on Very Large Data Bases, pages 408--421.
Observations

- nRank can work with any number of inputs
  - e.g., Find most likely publisher based on many characteristics (genre + author + publication date + …)
- Produces results with very little input information
  - e.g., Even “it’s a black and white film from 1943” would produce some output
  - Lower quality input will imply less confident output
- Weighting by economic activity increases the probability of finding someone with an economic interest in a work
  - Relevant for an orphan work search
- Weighting is not strictly necessary
- Any collection of metadata could potentially be used
  - Applicable to books, photographs, music and other creative works
- Graph quality is important, though
  - A biased set of metadata will skew results (e.g., the “IMDB keyword problem”)
Further Work

- Full nRank implementation
- Devise weighting mechanisms for other creative works
  - e.g., Lending activity at libraries? Web searches/traffic? Citations?
- Optimization strategies for graph weights
  - What weighting algorithm produces the best predictions?
  - Over-connected nodes can cause random walks to become “lost”