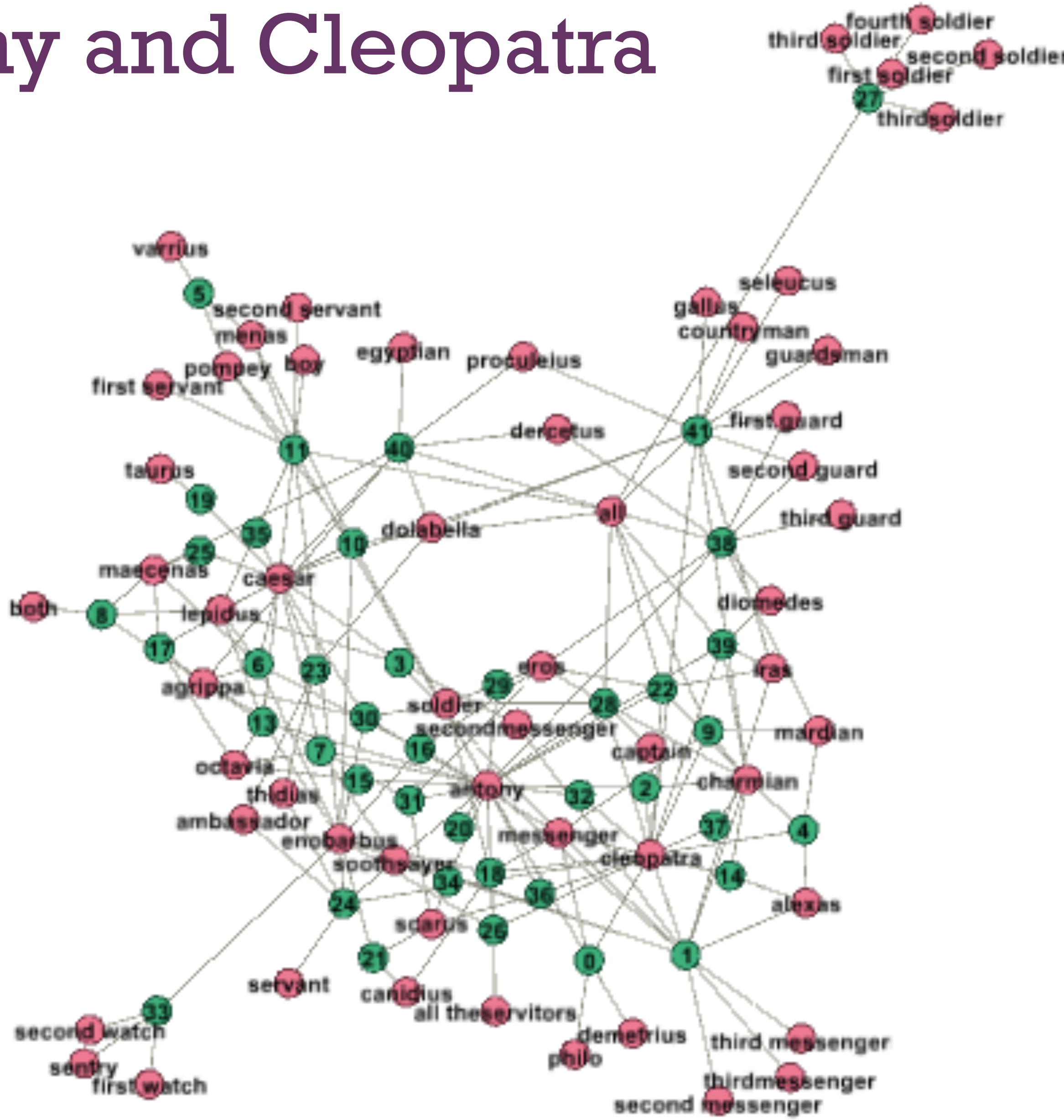
Recommender Systems
Random Walk Recommendation

Professor Robin Burke
Spring 2019

Bipartite Networks

- Special kind of network
- Two types of nodes
  - Think users and items
  - But many other applications
- Edges only allowed between different types of nodes
  - User-item edges
  - No user-user or item-item edges

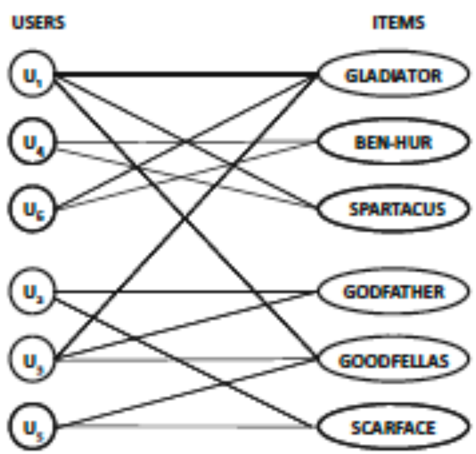
Antony and Cleopatra



User-item networks

- Rating-matrix = Incidence matrix
  - Different from adjacency matrix
  - Types of nodes on axes
- Bipartite network

	GLADIATOR	GODFATHER	BEN-HUR	GOODFELLAS	SCARFACE	SPARTACUS
U_1	1			5		2
U_2		5			4	
U_3	5	3		1		
U_4			3			4
U_5				3	5	
U_6	5		4			



The incidence matrix connects users and items, each of which are types of nodes, so it is a segment of the adjacency matrix (which has rows and columns for all the nodes). What parts are not present in the incidence matrix?

- A. The rest of the adjacency matrix would just be copies of this part.
- B. The parts of the adjacency matrix that represent the similarity between items
- C. The parts of the adjacency matrix that are zero because nodes of the same type can't connect to each other.

Paths in user-item network

- Length = 3
  - User 1 – Goodfellas – User 6 – Scarface
- Length = 5
  - User 1 – Gladiator – User 5 – Godfather – User 4 – Scarface
- Other items linked to users who share an item
  - Essentially what collaborative filtering is giving us

Random walks

- One option is to consider all odd-length random walks
  - Rank items by how often they are encountered
  - But this does not have a nice PageRank type solution
- Katz measure
  - Number of random walks between two nodes

$$Katz(i,j) = \sum_{t=1}^{\infty} \beta^t \cdot n_{ij}^{(t)}$$

  - Purpose of $\beta < 1$ is to discount longer paths
  - Matrix formulation

$$K = \sum_{i=1}^{\infty} (\beta A)^i = (I - \beta A)^{-1} - I$$

  - Has a solution as long as β is less than 1/(largest eigenvalue in A)
  - But long paths are probably not good in general
    - Doesn't take bipartite nature into account

+ P^3 and P^3_α

- How many times would item j be encountered in random walks of length 3 from user i?

- Probability of moving from node i and j
  - Provided there is an edge
  - a_{ij} / d_i

- Let D = diagonal matrix with vertex degrees d_i on the diagonal
  - D^{-1} = diagonal metric with $1/d_i$ on the diagonal

- One-step of the random walk
  - $D^{-1} A$
  - Probability matrix
  - $P^3 = (D^{-1} A)^3$

- Turns out better results come from $(a_{ij} / d_i)^\alpha$

  - α = number between 1 and 2, ex. 1.5, 1.8

P^3 uses length 3 random walks in the user-item bipartite network. You could create a similar algorithm P^5 using length 5 random walks, which also end at items. The recommendations from P^5 would be different in what way(s) from P^3?

- A. They would be the same or very similar.
- B. They would be completely different and most likely completely wrong because you are no longer in the target user's neighborhood (people who rated the same items)
- C. They would be similar to the P^3 recommendations, but maybe a bit more diverse and a bit less accurate.
- D. You can't get to length 5 in a random walk in a bipartite network.

Sampling

- Matrix multiplication with really large sparse matrices
  - Not very memory efficient
- In these cases, random walk problems are solved by sampling
  - Exactly what it sounds like
  - Generate a bunch of random walks and count hits
  - This can be very efficient
- Can even do selective updating
  - A user adds a new rating
  - Re-do just that user's random walks

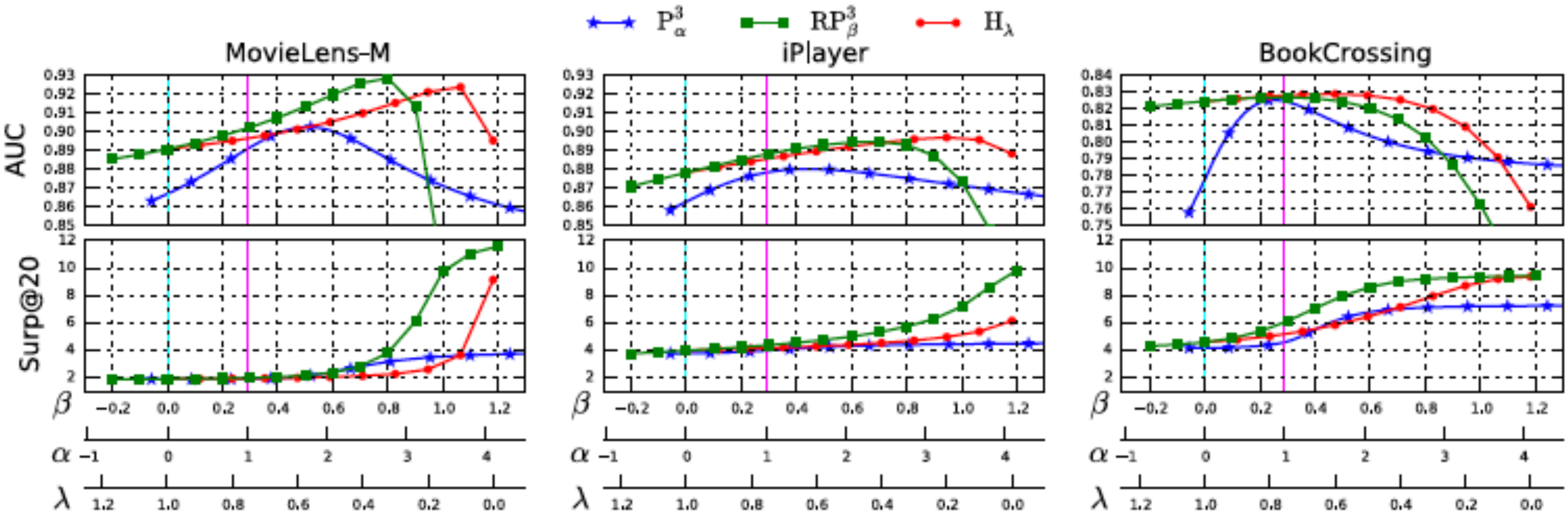
Popularity re-ranking

- Consider a user v_L with low degree
  - Rated two items j and k
- Every path through v_L will either exit to j or k
  - These are high probability edges
- Consider user v_H with high degree
  - Rated 200 items
- A path through v_H could go to any of 200 items
- There is a bias in the algorithm towards items rated by cold-start users
  - If we hit one of these users, we will go to one of their handful of items

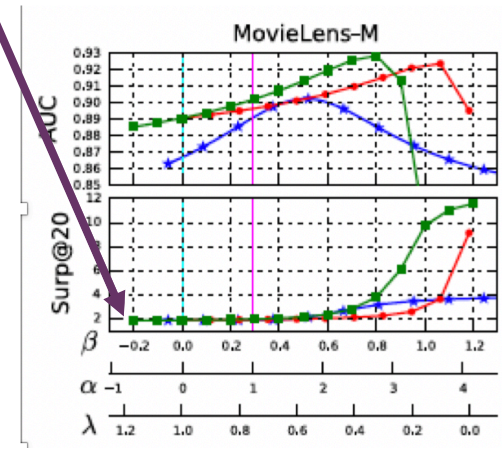
+ RP^3_β

- To adjust for the degree bias
  - Divide the score by the probability of following each path
- Probability of the path L =
  - $\frac{1}{\prod_{i \in L} d_i}$
- Apply this adjustment to all the resulting probabilities
  - This works better in experiments
  - Both accuracy and diversity
    - Christoffel et al., 2015

Some results



What punishment would be appropriate for the author who superimposed three different x-axes on top of each other in the previous figure?



- A. Having their cell phone permanently switched to Greek language mode
- B. Permanently changing the distance readout in their GPS applications to furlongs instead of miles.
- C. An eternity grading freshman math exams.