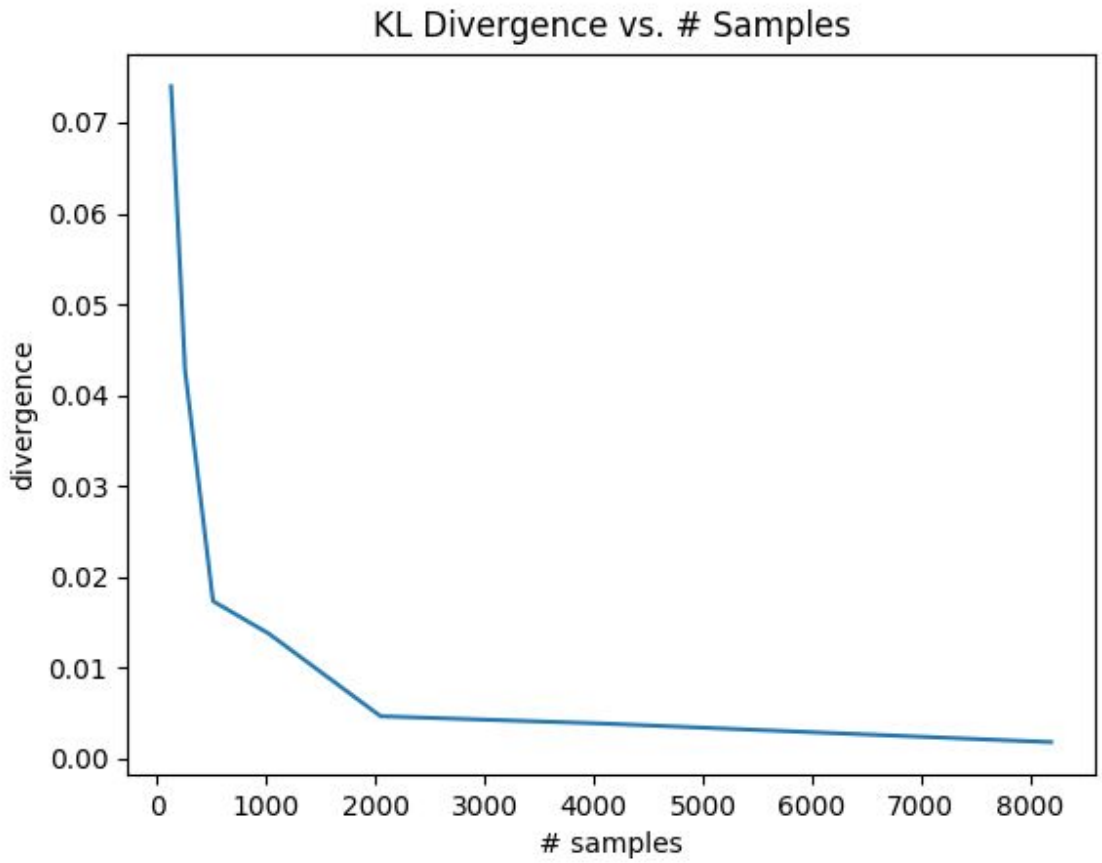
Milo Knowles
6.008 Fall 2018
December 6, 2018

Lab 5

a. See code for implementation.



b. The actors who appear consistently in the Pagerank Top 20 (all three trials):

De Niro, Robert
Wilson, Owen
Jenkins, Richard
Willis, Bruce
Cage, Nicholas
Kidman, Nicole

Trial #1
0.007400: Jackson, Samuel L.
0.006500: De Niro, Robert
0.005350: Wilson, Owen (I)
0.005300: Reilly, John C. (I)
0.005200: Jenkins, Richard (I)
0.005150: Hoffman, Philip Seymour
0.005050: Willis, Bruce
0.005050: Cage, Nicolas
0.004950: Sutherland, Donald (I)
0.004900: Stiller, Ben
0.004800: Baldwin, Alec
0.004650: Rhames, Ving
0.004650: Hopkins, Anthony (I)
0.004650: Freeman, Morgan (I)
0.004600: Walken, Christopher
0.004600: Kidman, Nicole
0.004500: Moore, Julianne
0.004450: Damon, Matt
0.004450: Buscemi, Steve
0.004400: Affleck, Ben

Trial #2
0.006300: Tucci, Stanley (I)
0.006050: Willis, Bruce
0.005950: Jackson, Samuel L.
0.005750: De Niro, Robert
0.005700: Jenkins, Richard (I)
0.005600: Cusack, John
0.005400: Cage, Nicolas
0.005350: Moore, Julianne
0.005050: Hackman, Gene
0.005000: Kidman, Nicole
0.004750: Wilson, Owen (I)
0.004700: Stiller, Ben

0.004650: DeVito, Danny
0.004550: Sutherland, Donald (I)
0.004550: Freeman, Morgan (I)
0.004500: Travolta, John
0.004500: Hoffman, Philip Seymour
0.004450: Platt, Oliver
0.004400: Reilly, John C. (I)
0.004400: Hopkins, Anthony (I)

Trial #3
0.007300: De Niro, Robert
0.006850: Jackson, Samuel L.
0.005550: Cusack, John
0.005400: Buscemi, Steve
0.005300: Wilson, Owen (I)
0.005100: Willis, Bruce
0.005100: Walken, Christopher
0.005100: Moore, Julianne
0.005050: Hackman, Gene
0.005000: Jenkins, Richard (I)
0.004850: Tucci, Stanley (I)
0.004700: Travolta, John
0.004600: Kidman, Nicole
0.004600: Keitel, Harvey
0.004600: Cusack, Joan
0.004600: Cage, Nicolas
0.004550: David, Keith (I)
0.004400: Sarandon, Susan
0.004400: Diaz, Cameron
0.004400: Baldwin, Alec

Top 20 Actors by Number of Movies
[('Jackson, Samuel L.', 53),
('De Niro, Robert', 45),
('Willis, Bruce', 44),
('Cage, Nicolas', 38),
('Moore, Julianne', 37),
('Hopkins, Anthony (I)', 35),
('Buscemi, Steve', 34),

('Baldwin, Alec', 33),
('Depp, Johnny', 33),
('Walken, Christopher', 32),
('Williams, Robin (I)', 32),
('Freeman, Morgan (I)', 32),
('Wilson, Owen (I)', 31),
('Washington, Denzel', 31),
('Damon, Matt', 31),
('Cox, Brian (I)', 31),
('Goodman, John (I)', 30),
('Neeson, Liam', 30),
('Kidman, Nicole', 30),
('Pitt, Brad', 30)]

c.

i. The Markov chain is *not* fully communicating, because not every page is reachable from every other page. For example, we may follow a hyperlink to a webpage with no links on it, making it impossible to transition to any other state.

ii. With randomization, the Markov chain is fully communicating. There is always a nonzero probability of transitioning to any other page, so it is now possible to reach any page from any other. Because the chain is fully communicating with randomization, it has a stationary distribution.

iii. To increase page rank score, a web page owner could either add self-links to their own page, or convince other webpages to add links to their site. In particular, they should try to get incoming links from other pages with high Page Rank, since these other pages will be visited more frequently. As we can see from the results on the actor dataset, actors that starred in many movies (i.e have many incoming edges from other actor nodes) also tend to have high page rank.