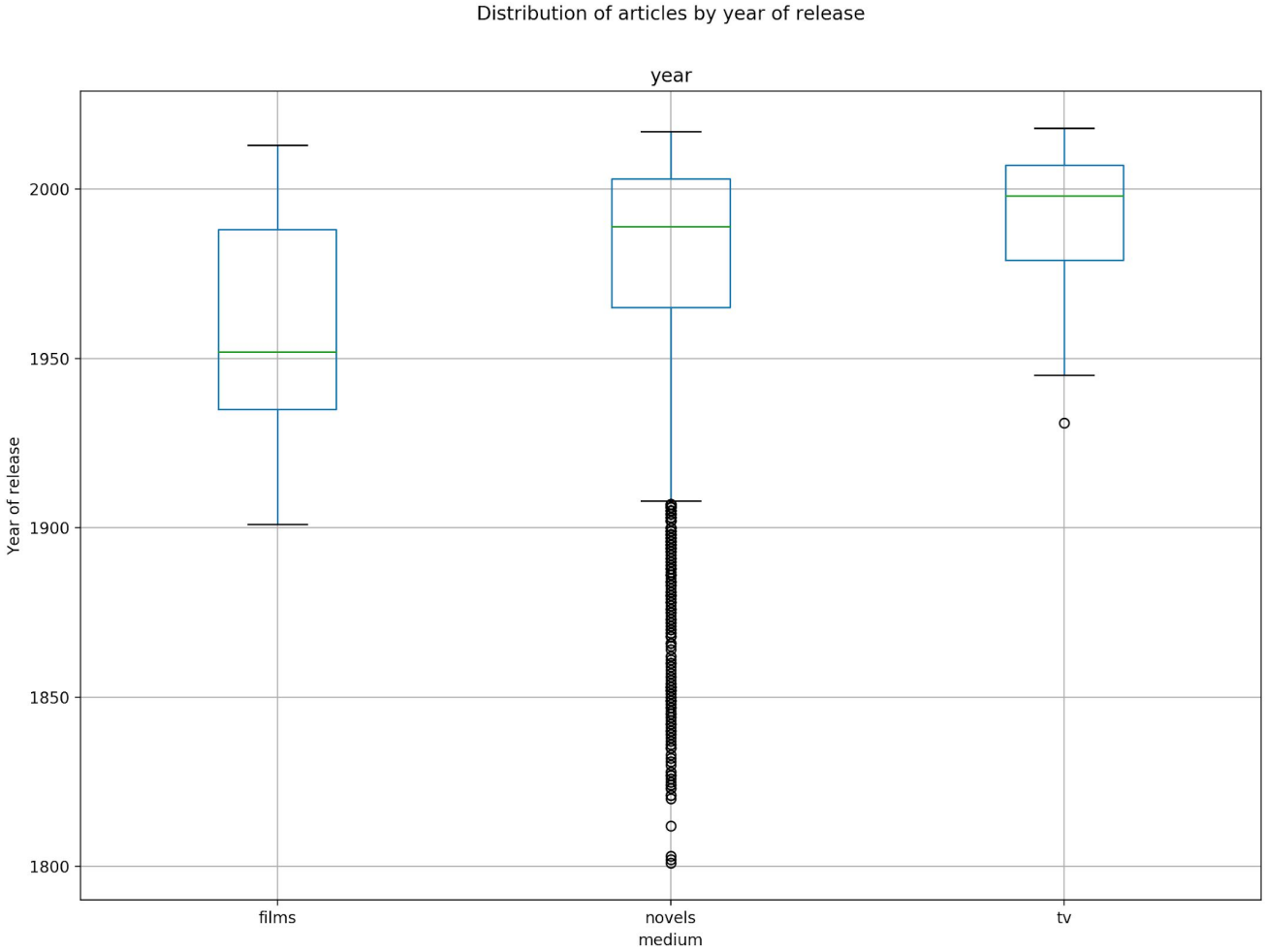
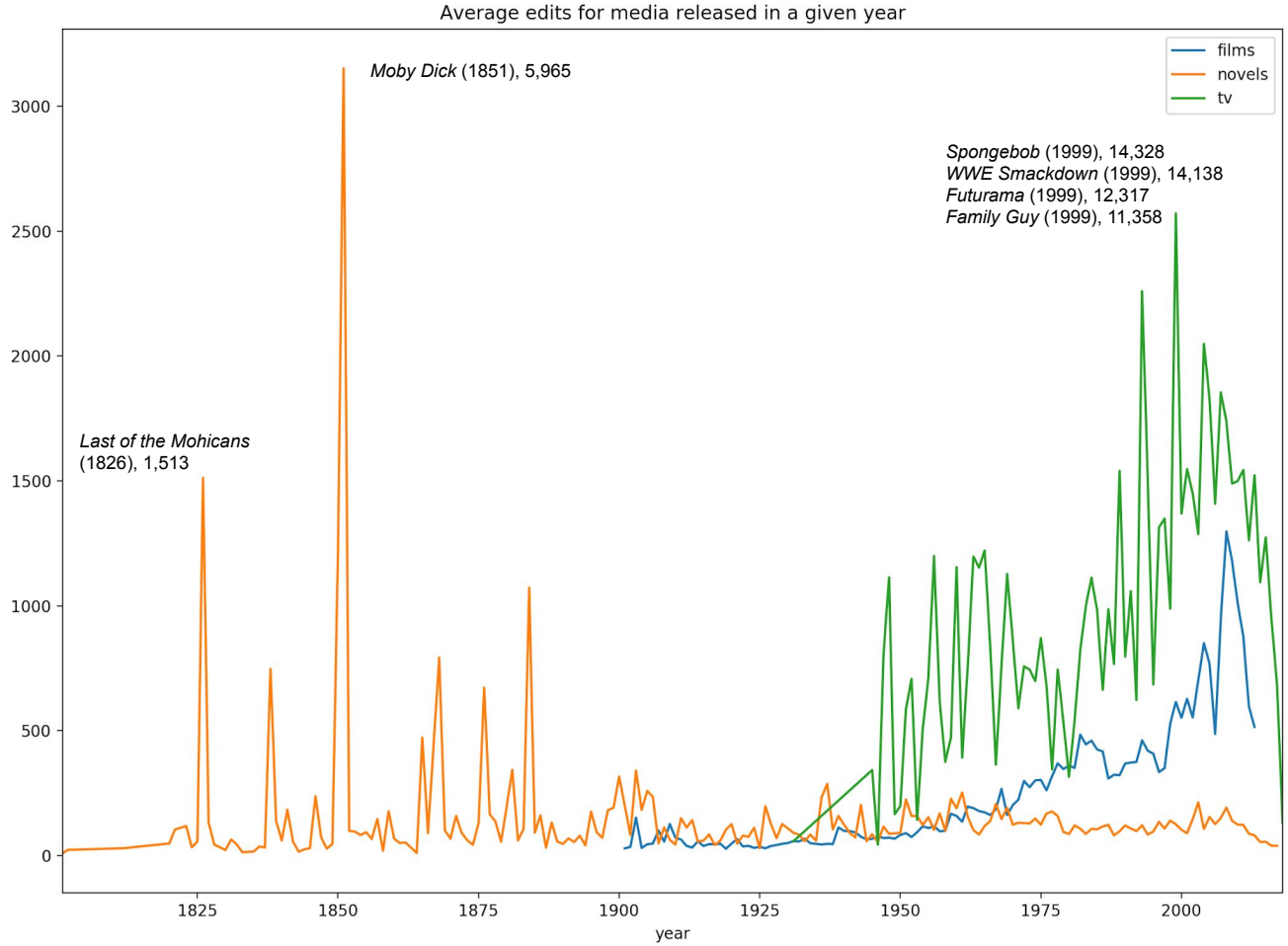
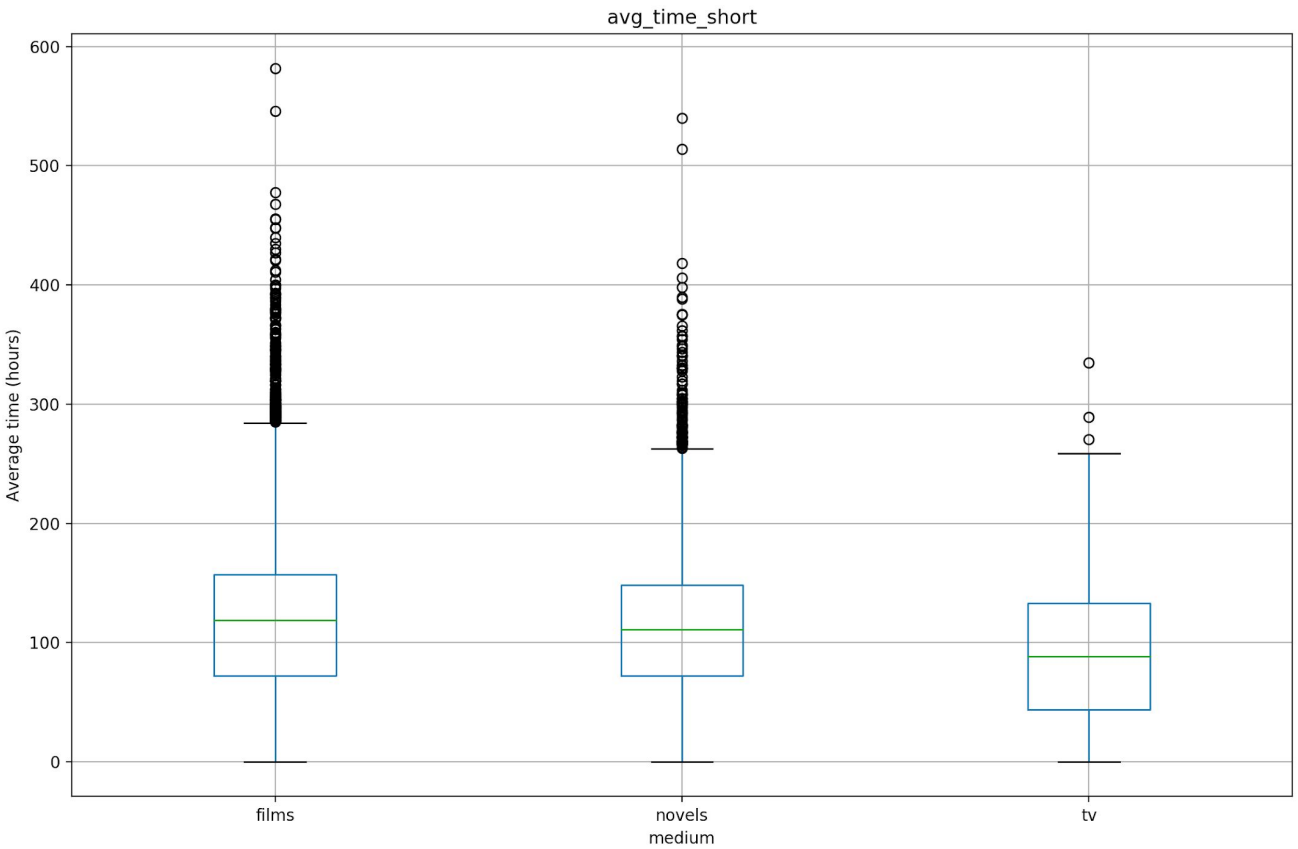
Wikipedia

Overview of the dataset

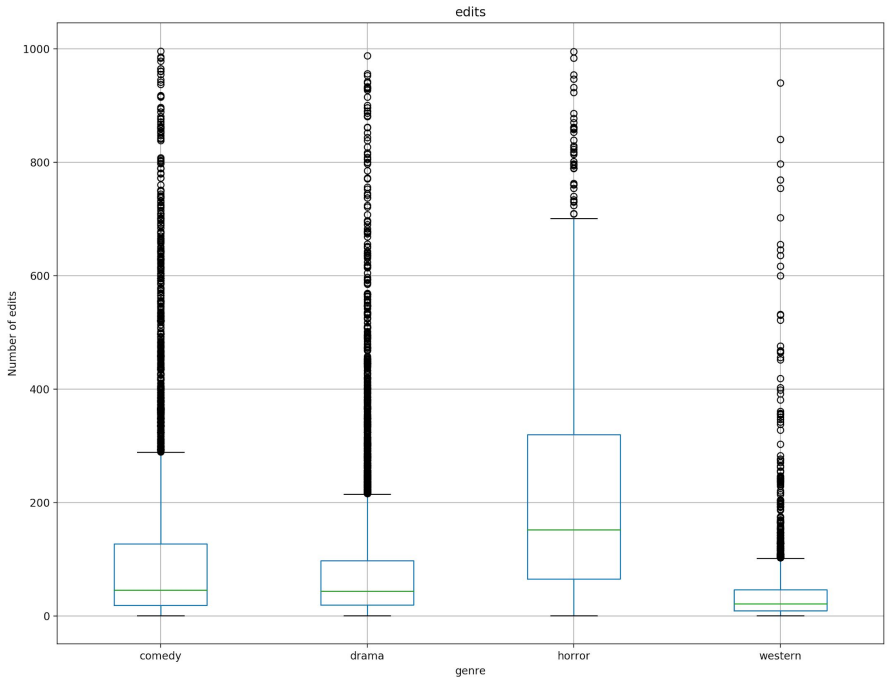
- Complete histories of Wikipedia pages for:
  - 25,355 films (1901-2013), via en.wikipedia.org/wiki/Lists_of_American_films
  - 10,523 novels (1801-2017), via en.wikipedia.org/wiki/Category:American_novels
  - 2,324 tv shows (1933-2018), via en.wikipedia.org/wiki/List_of_American_television_programs


Distribution of articles by year of release
year
Year of release
2000
1950
1900
1850
1800
films
novels
tv
medium

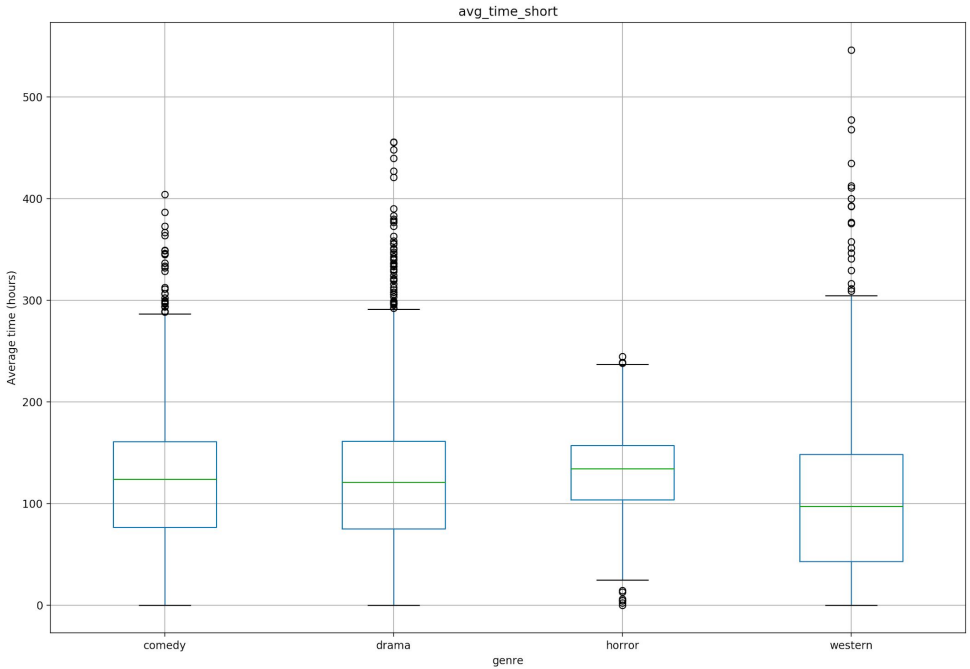

Average edits for media released in a given year
films
novels
tv
Moby Dick (1851), 5,965
Spongebob (1999), 14,328
WWE Smackdown (1999), 14,138
Futurama (1999), 12,317
Family Guy (1999), 11,358
Last of the Mohicans (1826), 1,513
Wall-E (2008), 9,967
The Dark Knight (2008), 12,766
High School Musical 3 (2008), 8,904
Breaking Dawn (2008), 5,917
Brisingr (2008), 4,616
The Hunger Games (2008), 3,845
3000
2500
2000
1500
1000
500
0
1825
1850
1875
1900
1925
1950
1975
2000
year


Time between edits for edits of less than one month across all mediums
avg_time_short
Average time (hours)
600
500
400
300
200
100
0
films
novels
tv
medium

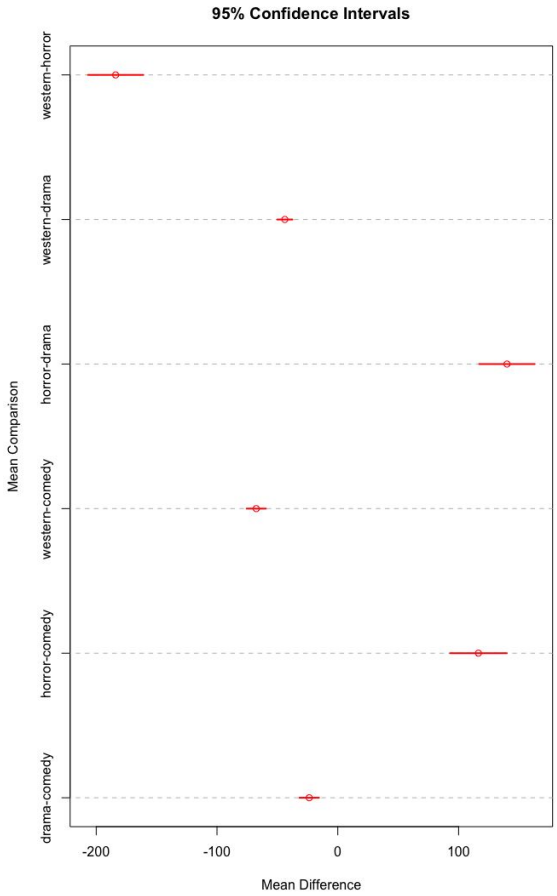
Number of edits for articles on films grouped by genre



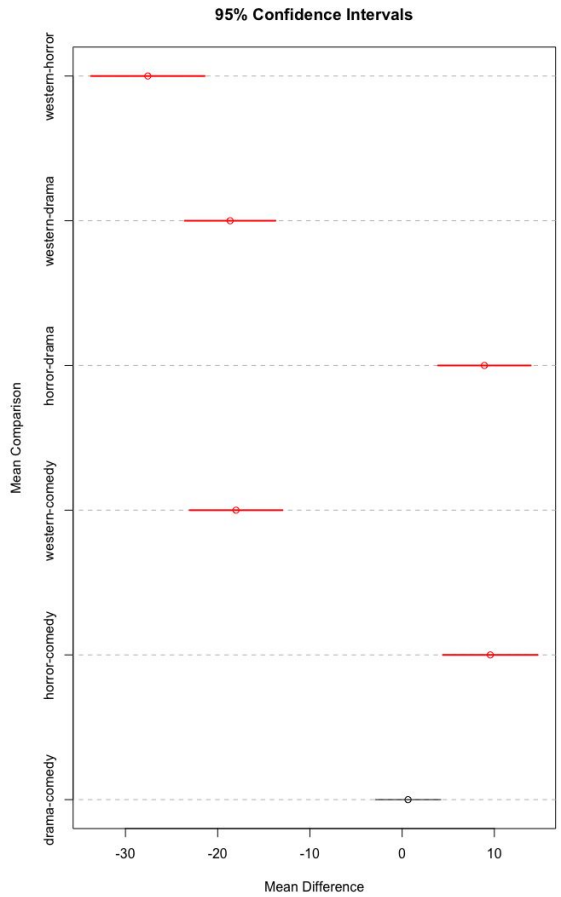
Time between edits for articles on films grouped by genre

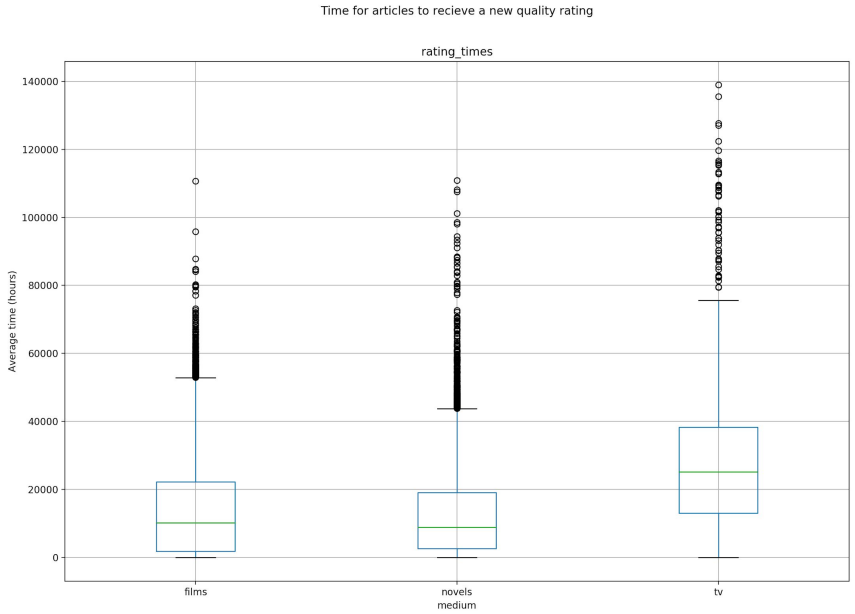


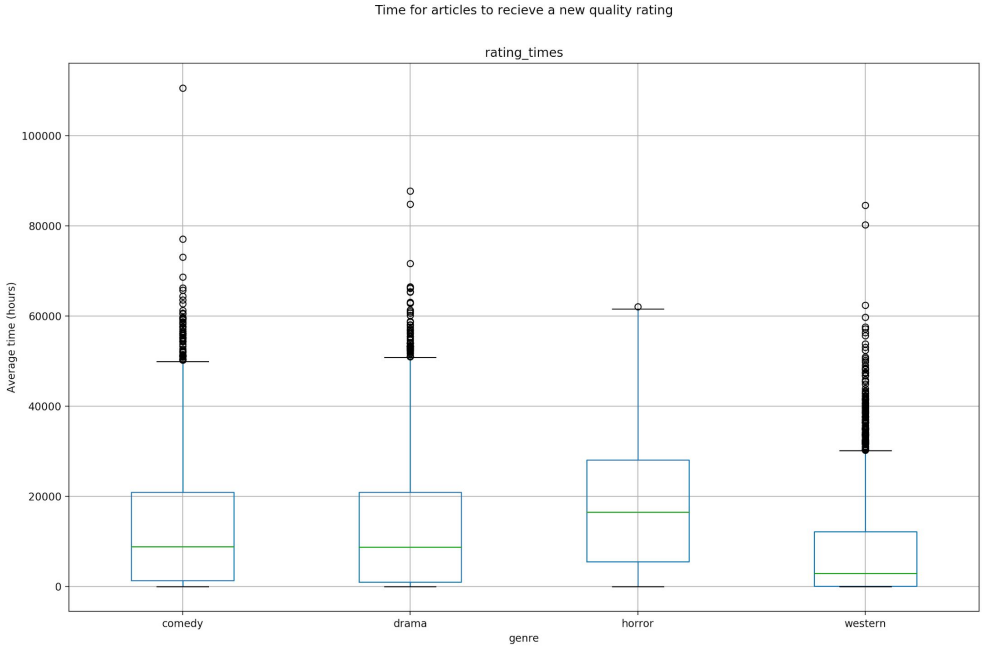
Comparison of means for number of edits by genre



Comparison of means for time between edits by genre




Time for articles to recieve a new quality rating
rating_times
Average time (hours)
0
20000
40000
60000
80000
100000
120000
140000
films
novels
tv
medium


Time for articles to recieve a new quality rating
rating_times
Average time (hours)
0
20000
40000
60000
80000
100000
comedy
drama
horror
western
genre

Measuring community coherence

Louvain community detection

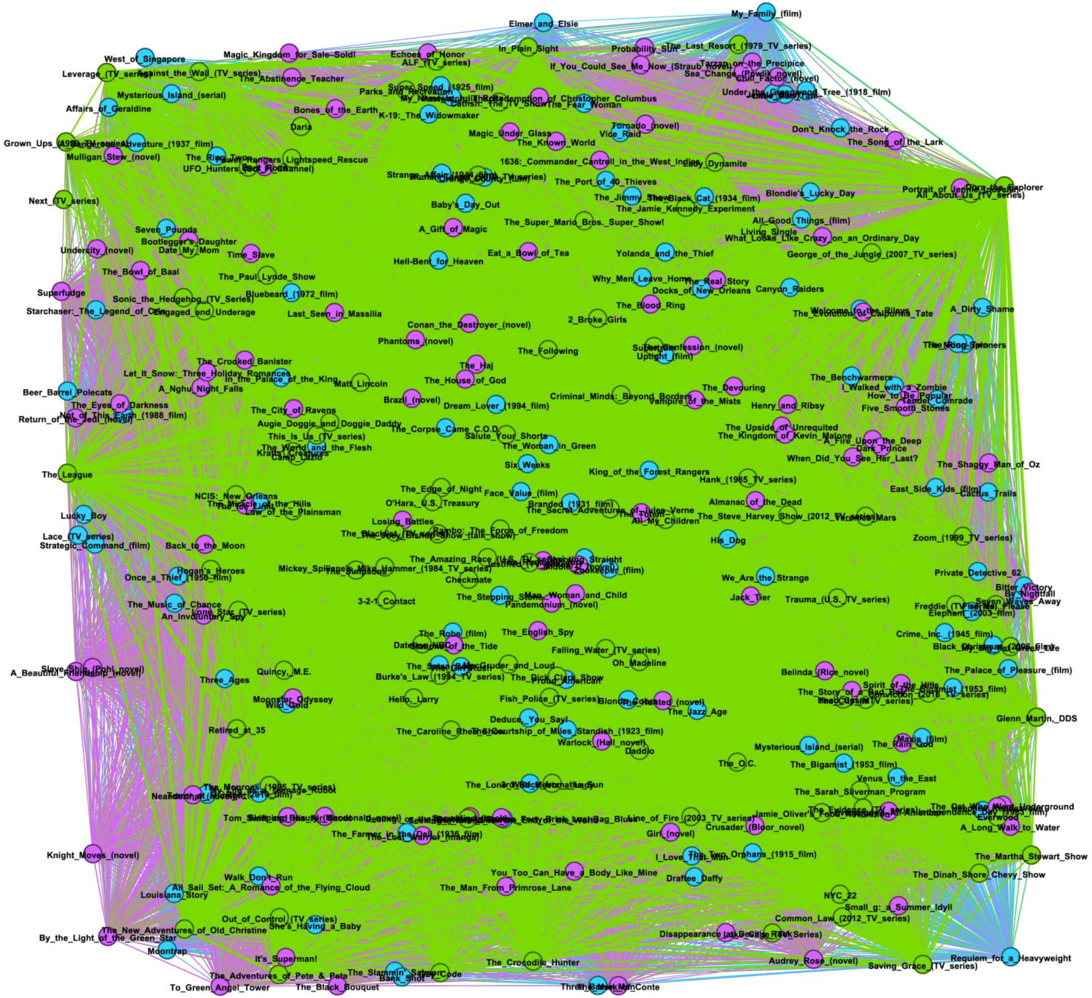
- Unsupervised
- Modeled on user behavior

Logistic regression

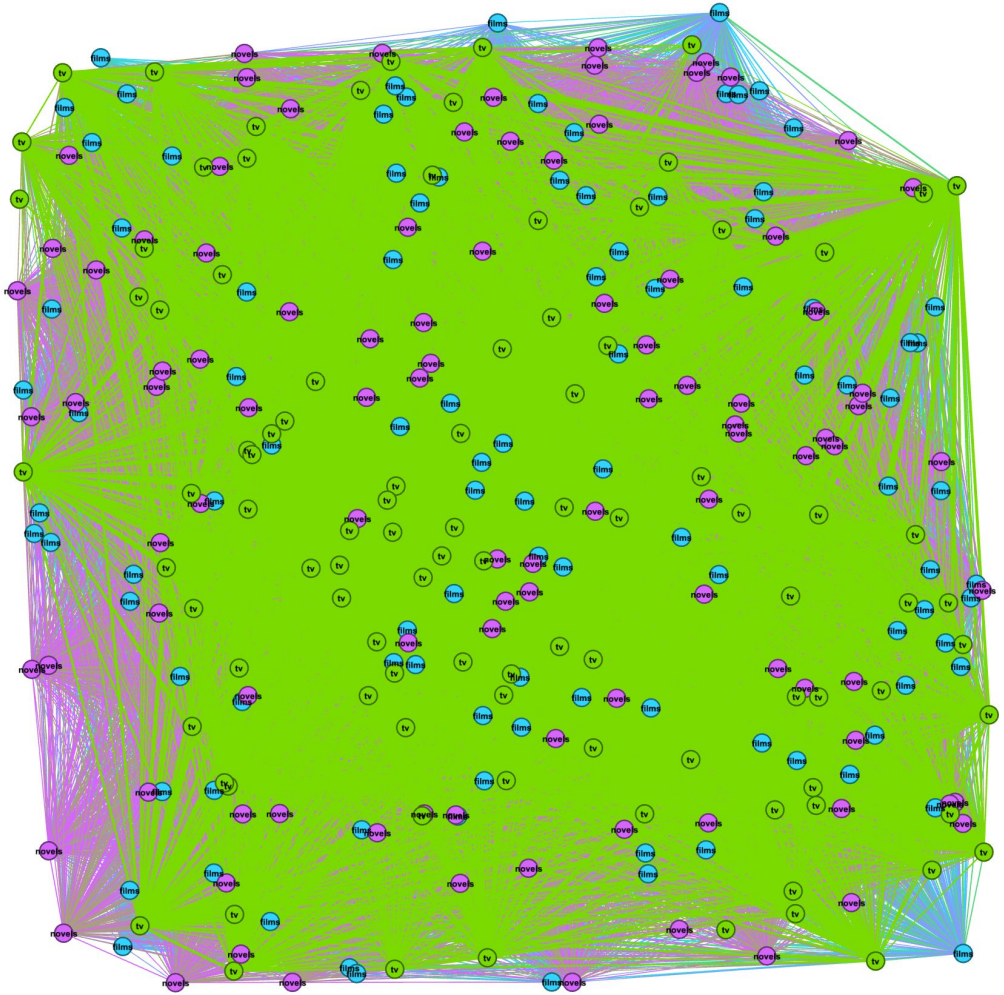
- Supervised
- Modeled on revision semantics

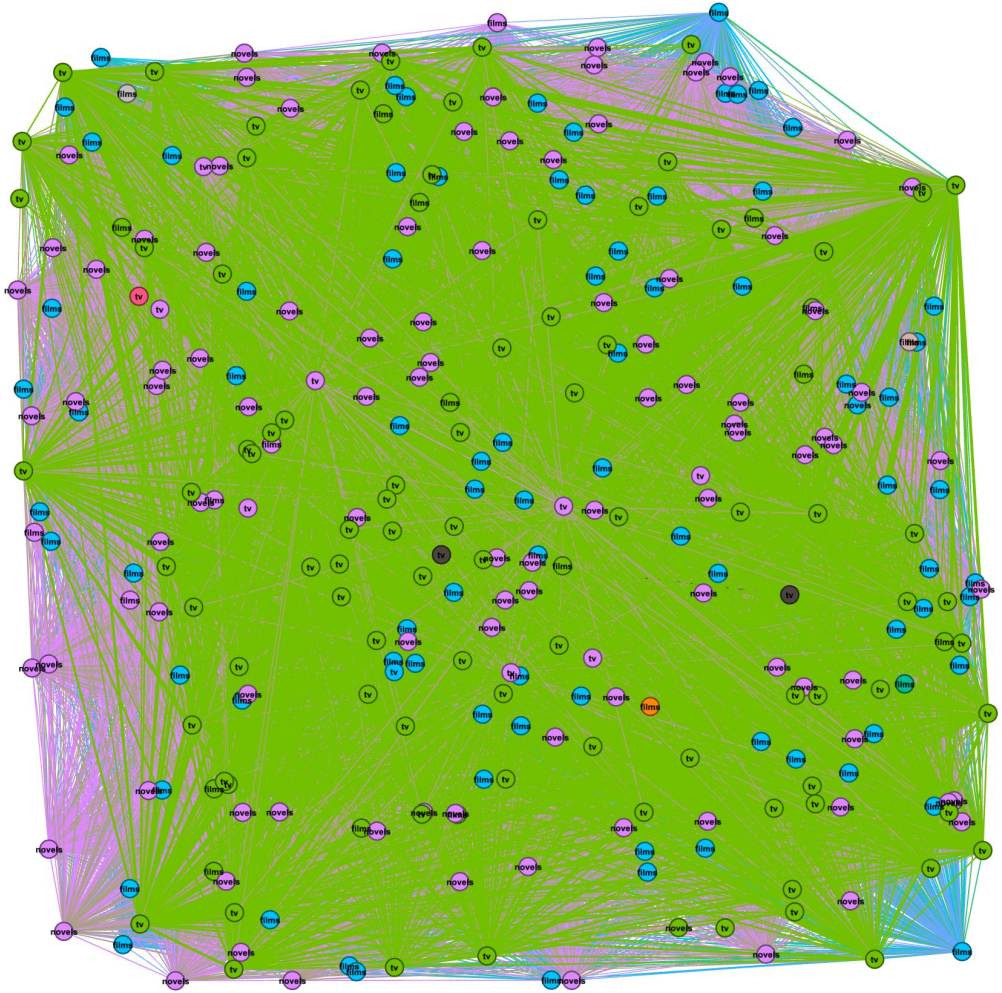
Louvain

- Network representation of editor behavior
  - Nodes represent articles, weighted edges represent the number of common editors between any two articles
- Structure
  - Graphs generated through random selection of articles
  - To test coherence by medium: 50 graphs, each with 100 nodes from each medium
  - To test coherence by period and genre: 150 graphs (50 representing each medium), each with 300 nodes from the same medium



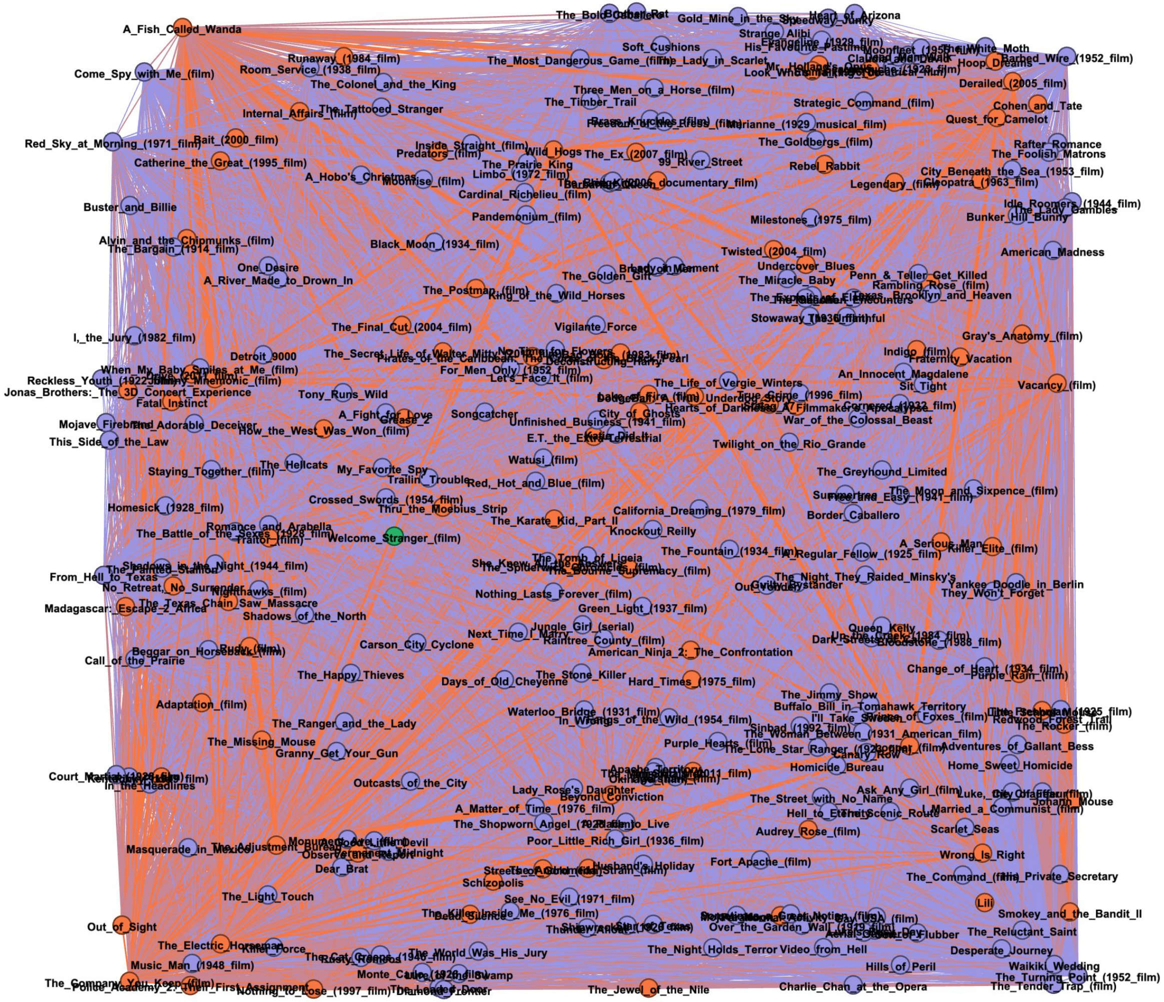
Elmer_and_Elsie
My_Family_(film)
Magic_Kingdom_for_Sale–Sold!
Echoes_of_Honor
ALF_(TV_series)
In_Plain_Sight
The_Last_Resort_(1979_TV_series)
West_of_Singapore
The_Abstinence_Teacher
Tarzan_on_the_Precipice
Chill_Factor_(novel)
Mysterious_Island_(serial)
Affairs_of_Geraldine
Super_Speed_(1925_film)
Parks_and_Recreation
Bones_of_the_Earth
K-19:_The_Widowmaker
Daria
Tornado_(novel)
Don't_Knock_the_Rock
Magic_Under_Glass
The_Known_World
The_Song_of_the_Lark
Mulligan_Stew_(novel)
Power_Rangers_Lightspeed_Rescue
1636:_Commander_Cantrell_in_the_West_Indies
Lady_Dynamite
The_Port_of_40_Thieves
Next_(TV_series)
Blondie's_Lucky_Day
Baby's_Day_Out
The_Jamie_Kennedy_Experiment
Seven_Pounds
A_Gift_of_Magic
The_Super_Mario_Bros._Super_Show!
All_Good_Things_(film)
Bootlegger's_Daughter
Date_My_Mom
Undercity_(novel)
Time_Slave
Eat_a_Bowl_of_Tea
Yolanda_and_the_Thief
George_of_the_Jungle_(2007_TV_series)
The_Bowl_of_Baal
Hell-Bent_for_Heaven
The_Paul_Lynde_Show
Why_Men_Leave_Home
Docks_of_New_Orleans
Canyon_Raiders
Superfudge
Sonic_the_Hedgehog_(TV_Series)
Bluebeard_(1972_film)
The_Blood_Ring
A_Dirty_Shame
Last_Seen_in_Massilia
2_Broke_Girls
Conan_the_Destroyer_(novel)
Phantoms_(novel)
The_Following
The_Crooked_Banister
The_Haj
The_House_of_God
The_Devouring
The_Benchwarmers
I_Walked_with_a_Zombie
How_to_Be_Popular
A_Nghu_Night_Falls
Matt_Lincoln
Beer_Barrel_Polecats
Brazil_(novel)
Dream_Lover_(1994_film)
Vampire_of_the_Mists
Henry_and_Ribsy
Five_Smooth_Stones
The_Eyes_of_Darkness
The_City_of_Ravens
Augie_Doggie_and_Doggie_Daddy
The_Corpse_Came_C.O.D.
The_Upside_of_Unrequited
This_Is_Us_(TV_series)
Salute_Your_Shorts
The_Kingdom_of_Kevin_Malone
The_Woman_in_Green
Dark_Prince
Camp_Lazlo
When_Did_You_See_Her_Last?
Six_Weeks
The_Shaggy_Man_of_Oz
The_League
King_of_the_Forest_Rangers
Hank_(1965_TV_series)
The_Edge_of_Night
Face_Value_(film)
East_Side_Kids_(film)
Cactus_Trails
O'Hara,_U.S._Treasury
Almanac_of_the_Dead
NCIS:_New_Orleans
Law_of_the_Plainsman
All_My_Children
Lucky_Boy
Losing_Battles
His_Dog
Lace_(TV_series)
Strategic_Command_(film)
Back_to_the_Moon
Zoom_(1999_TV_series)
Hogan's_Heroes
Once_a_Thief_(1950_film)
Checkmate
We_Are_the_Strange
Private_Detective_62
Bitter_Victory
The_Music_of_Chance
3-2-1_Contact
Pandemonium_(novel)
Jack_Tier
Trauma_(U.S._TV_series)
An_Involuntary_Spy
The_Robe_(film)
The_English_Spy
Crime,_Inc._(1945_film)
Falling_Water_(TV_series)
Oh_Madeline
A_Beautiful_Friendship_(novel)
Quincy,_M.E.
Burke's_Law_(1994_TV_series)
Belinda_(Rice_novel)
The_Palace_of_Pleasure_(film)
Three_Ages
Moonstar_Odyssey
Hello,_Larry
Fish_Police_(TV_series)
The_Jazz_Age
Deduce,_You_Say!
Retired_at_35
Glenn_Martin,_DDS
Warlock_(Hall_novel)
Daddio
Mysterious_Island_(serial)
The_O.C.
The_Bigamist_(1953_film)
Venus_in_the_East
The_Sarah_Silverman_Program
Line_of_Fire_(2003_TV_series)
Crusader_(Bloor_novel)
A_Long_Walk_to_Water
Girl_(novel)
Knight_Moves_(novel)
The_Martha_Stewart_Show
Walk,_Don't_Run
You_Too_Can_Have_a_Body_Like_Mine
Draftee_Daffy
The_Dinah_Shore_Chevy_Show
All_Sail_Set:_A_Romance_of_the_Flying_Cloud
Louisiana_Story
The_Man_From_Primrose_Lane
NYC_22
Small_g:_a_Summer_Idyll
Out_of_Control_(TV_series)
Common_Law_(2012_TV_series)
The_New_Adventures_of_Old_Christine
She's_Having_a_Baby
By_the_Light_of_the_Green_Star
Moontrap
It's_Superman!
The_Crocodile_Hunter
Audrey_Rose_(novel)
Saving_Grace_(TV_series)
Requiem_for_a_Heavyweight
The_Adventures_of_Pete_&_Pete
To_Green_Angel_Tower
The_Black_Bouquet


films
novels
tv


films
tv
novels

Louvain by medium

- Only clusters containing >50 nodes were considered for analysis
  - 47.8% of the total clusters across all graphs.
  - 3.02 large clusters per graph
- Average purity of 92.5% with 5.3% standard deviation
  - 33.1% of all clusters were majority novels
  - 33.1% of all clusters were majority films
  - 33.8\% of all clusters were majority television shows
- User behavior is extremely confined by medium


A_Fish_Called_Wanda
Come_Spy_with_Me_(film)
Red_Sky_at_Morning_(1971_film)
Buster_and_Billie
Runaway_(1984_film)
Room_Service_(1938_film)
The_Colonel_and_the_King
Internal_Affairs_(film)
The_Tattooed_Stranger
Bait_(2000_film)
Catherine_the_Great_(1995_film)
A_Hobo's_Christmas
Moonrise_(film)
The_Most_Dangerous_Game_(film)
The_Lady_in_Scarlet
Soft_Cushions
Gold_Mine_in_the_Sky
Heart_of_Arizona
Three_Men_on_a_Horse_(film)
The_Timber_Trail
Strategic_Command_(film)
Derailed_(2005_film)
Barbed_Wire_(1952_film)
Cohen_and_Tate
Quest_for_Camelot
Marianne_(1929_musical_film)
The_Goldbergs_(film)
Inside_Straight_(film)
Predators_(film)
Wild_Hogs
The_Prairie_King
Limbo_(1972_film)
The_Ex_(2007_film)
99_River_Street
Rebel_Rabbit
Rafter_Romance
The_Foolish_Matrons
City_Beneath_the_Sea_(1953_film)
Legendary_(film)
Cleopatra_(1963_film)
Idle_Roomers_(1944_film)
The_Lady_Gambles
Bunker_Hill_Bunny
Cardinal_Richelieu_(film)
Pandemonium_(film)
Milestones_(1975_film)
American_Madness
Alvin_and_the_Chipmunks_(film)
The_Bargain_(1914_film)
Black_Moon_(1934_film)
Twisted_(2004_film)
One_Desire
A_River_Made_to_Drown_In
Undercover_Blues
The_Golden_Gift
The_Miracle_Baby
Penn_&_Teller_Get_Killed
Rambling_Rose_(film)
The_Postman_(film)
King_of_the_Wild_Horses
Vigilante_Force
The_Final_Cut_(2004_film)
Gray's_Anatomy_(film)
I,_the_Jury_(1982_film)
Detroit_9000
Indigo_(film)
Fraternity_Vacation
An_Innocent_Magdalene
Sit_Tight
Vacancy_(film)
For_Men_Only_(1952_film)
Let's_Face_It_(film)
The_Life_of_Vergie_Winters
Jonas_Brothers:_The_3D_Concert_Experience
Fatal_Instinct
Tony_Runs_Wild
A_Fight_for_Love
Songcatcher
City_of_Ghosts
War_of_the_Colossal_Beast
Mojave_Firebrand
The_Adorable_Deceiver
How_the_West_Was_Won_(film)
This_Side_of_the_Law
Unfinished_Business_(1941_film)
Twilight_on_the_Rio_Grande
Watusi_(film)
Staying_Together_(film)
The_Hellcats
My_Favorite_Spy
Trailin'_Trouble
Red,_Hot_and_Blue_(film)
The_Greyhound_Limited
Homesick_(1928_film)
Crossed_Swords_(1954_film)
Thru_the_Moebius_Strip
California_Dreaming_(1979_film)
The_Moon_and_Sixpence_(film)
Romance_and_Arabella
The_Karate_Kid,_Part_II
Border_Caballero
Welcome_Stranger_(film)
Knockout_Reilly
A_Serious_Man
Killer_Elite_(film)
The_Fountain_(1934_film)
A_Regular_Fellow_(1925_film)
The_Tomb_of_Ligeia
The_Bourne_Supremacy_(film)
From_Hell_to_Texas
No_Retreat,_No_Surrender
Nighthawks_(film)
The_Night_They_Raided_Minsky's
Yankee_Doodle_in_Berlin
They_Won't_Forget
Madagascar:_Escape_2_Africa
The_Texas_Chain_Saw_Massacre
Nothing_Lasts_Forever_(film)
Green_Light_(1937_film)
Shadows_of_the_North
Jungle_Girl_(serial)
Next_Time_I_Marry
Queen_Kelly
Carson_City_Cyclone
Raintree_County_(film)
American_Ninja_2:_The_Confrontation
Call_of_the_Prairie
Change_of_Heart_(1934_film)
Purple_Rain_(film)
The_Happy_Thieves
The_Stone_Killer
Days_of_Old_Cheyenne
Hard_Times_(1975_film)
The_Jimmy_Show
Buffalo_Bill_in_Tomahawk_Territory
Adaptation_(film)
Waterloo_Bridge_(1931_film)
The_Rocker_(film)
The_Ranger_and_the_Lady
The_Missing_Mouse
Granny_Get_Your_Gun
Purple_Hearts_(film)
Adventures_of_Gallant_Bess
Homicide_Bureau
Home_Sweet_Homicide
Outcasts_of_the_City
Lady_Rose's_Daughter
Beyond_Conviction
Ask_Any_Girl_(film)
Johann_Mouse
The_Street_with_No_Name
A_Matter_of_Time_(1976_film)
Hell_to_Eternity
The_Scenic_Route
I_Married_a_Communist_(film)
Audrey_Rose_(film)
Scarlet_Seas
Masquerade_in_Mexico
Poor_Little_Rich_Girl_(1936_film)
Wrong_Is_Right
Dear_Brat
Fort_Apache_(film)
Husband's_Holiday
Schizopolis
His_Private_Secretary
The_Light_Touch
See_No_Evil_(1971_film)
Lili
Smokey_and_the_Bandit_II
Out_of_Sight
Over_the_Garden_Wall_(1919_film)
The_Reluctant_Saint
The_Electric_Horseman
The_World_Was_His_Jury
The_Night_Holds_Terror
Video_from_Hell
Desperate_Journey
Music_Man_(1948_film)
Hills_of_Peril
Waikiki_Wedding
The_Turning_Point_(1952_film)
Charlie_Chan_at_the_Opera
The_Tender_Trap_(film)
The_Jewel_of_the_Nile

Louvain by period

- Films
  - Average purity of 78.6% (9.2% stdev)
  - 50% were majority early period (71% average purity), 50% were majority late period (85% average purity)
  - Strong division based on period: Editors divide equally as either early- or late-period and stay more strictly confined to editing media from their period of choice
- TV
  - Average purity of 70.8% (11.2% stdev)
  - 60.7% were majority early period (70.3% average purity), 39.3% were majority late period (71.6% average purity)
  - Division based on period: Editors cluster more strongly around early period and tend to edit only either early- or late-period TV shows

Louvain by period

- Novels
  - Average purity of 57.9% (6.4% stdev)
  - 72.9% were majority early period (59.2% average purity), 27.1% were majority late period (54.4% average purity)
  - Little to no division based on period: Editors contribute equally to early- and late-period articles, with slightly stronger clustering around early period novels

Louvain by genre

- Clustering only using articles tagged as either dramas or comedies
- Films
  - 57.5% purity
  - 90.0% majority drama
  - This split matches the division between comedies and dramas in the dataset: 55.8% of films in the set are tagged by genre. This suggests little to no division of editor behavior by genre for films.
- Television
  - 76.4% purity
  - 100% majority comedy
  - This split also matches the division between comedies and dramas; 77.3% of TV articles in this set are comedies. Again, little to no division of editor behavior by genre.

Logistic regression: Accuracy

- Pairwise medium comparisons
  - 93.3% accurate between novels and films
  - 95.2% accurate between novels and TV
  - 96.6% accurate between TV and film
- Pairwise period comparisons (between sets split by the median release year)
  - 86.0% accurate for films
  - 82.0% accurate for television
  - 76.3% accurate for novels
- Pairwise genre comparisons (between drama and comedy)
  - 88.9% accurate for television
  - 88.4% accurate for films

Distinctive features by medium

Medium specific
Genre specific
Of note

- As the intersection of each set of pairwise comparisons
  - Television: lasting, cancellation, rancher,cancels, premiering, decadence, and resale
  - Novels: publishers, publishing,paperboy, and novelty
  - Films: silents, stars,directing, mayhem, and westerners

Between each medium

Medium specific
Genre specific
Of note

- Novels compared to television
  - TV: stars, liveliest, direly, completely, productively, hostess, clearance, variously, scheduling, starry, and producers
  - Novels: authoress, spoiling, novel-ettes, trimester, adapting, thrilling, booksellers, deli, harebrained, suicides, presser, bookcase, considerably,and message
- Television compared to films
  - TV: reruns, site, revolvers, dirge, thirties, lift, stubble, amazons, boldest, focusing, and syndicating
  - Films: bangs, festive, spoiling, rotter, tabor, republicanism, weaned, schemer, romanticism, permeate, sales, picturing, and screen
- Novels compared to films
  - Novels: finality, repudiate, harden, serials, reprints, beta, bestselling, locusts, illustrating, torched, prosecute, reciprocal, novelization, omnipotence, nebulous, and sordidness
  - Films: forcing, archives, temples, billet, monograph, melodramatic, dramatic, beeswax, distributes, digresses, comer, uncritically, documented, and tomboy

Medium specific
Genre specific
Of note

Distinctive features by period

- Novels (1989)
  - Early period: pseudonyms, serials, illustrating, adapting, originate, reprints, spying, editor, swiftness, rulers, westerners, drugstores, repudiate, chickens, publishers, coping, scientific, stover, femme, and adventures
  - Late period: ween, scheduling, announcements, fantasies, positives, torched turvy, harden, ailed, officials, entertains, graphically, wed, writhe, virtue, cabaret, synod, measly, adulterer, and homeland
- Films (1950)
  - Early period: silents, preserves, profitable, domesticated, melodramatic, beeswax, censured, twerp, survivor, codebooks, remainder, overheated, reelection, fecal, dome, smilingly, corporatism, crisply, congressman, and printed
  - Late period: engulfing, devouring, externally, theatrics, festive, bulletproof, spoiling, erotically, globetrotters, thrilling, furbelows, prostitutes, rocked, tomboy, cultists, heisting, rotter

Medium specific
Genre specific
Of note

Distinctive features by period

- Television
  - Early period: runaways, starry, releasing, regularly, including, syndrome, programing, appears, chill, releases, yeast, hag, broadcasters, dusk, answer, beck, policemen, fatherhood, seeped
  - Late period: premieres, realizations, millions, fond, channeling, beep, follows, relative, comas, seconded, whim, defends, killed, anyhow, premiering, willful, islanders, lookalikes, and beached

Medium specific
Genre specific
Of note

Distinctive features by genre

- Films
  - Dramas: dramatic, melodramatic, dramatics, reigning, thrilling, feelings, informative, novelist, prisoners, activate, formidable, lovesick, gibbon, holes, possum, pennies, wardens, understandably, someplace, and wellbeing
  - Comedies: comer, tickle, gawker, sequence, functional, nip, canes, superb, originally, harebrained, opposed, reinforced, watery, heroic, manslaughter, serious, doves, thirds, turning, and convincingly
- TV
  - Dramas: dramatic, produces, seriously, italics, historians, offbeat, newscaster, outraged, inaccuracy, part, laurel, teaming
  - Comedies: site, comers, upbringing, filming, youngest, players, allegations, hammering, west, showcase, songwriter, episodic, brace, one, frontier, hind, washed, sheen, headache, and them

Further work

- Tracing distinctive features
  - How does usage of distinctive features for genres and mediums change over time?
- Topic modeling
  - Topic model the history of each document, compare topics for high-frequency and low-frequency pages to understand differences between high activity and low activity pages
- Intra- and extra-Wikipedia comparisons
  - How do editing habits on Wikipedia compare to academic writing in journals?
  - How do other social domains on Wikipedia compare to cultural domains?