Archive Of Our Owns
Fandoms and Tags

同人小說領域中粉絲圈與標籤的關係

#內容警告

#同人術語 #色情描寫 #少許中二內容 #可能... 只是可能...這份簡報中出現的某些標籤剛好是劇透，所以...#劇透警告

介紹

Archive Of Our Owns (AO3)

一個非盈利且開源的同人小說存檔網站

網站上的作品數量已經超過450萬

標籤 (tags)

在網路上，標籤是拿來標註並描述一份資訊的字詞，它能使資訊方便瀏覽及搜索。

通常是由資料的創建者或其觀看者加上的。

AO3上的標籤類型

Rating (分級)
◦ None、General、Teenage And Up、Mature、Explicit

Category (配對類別)
◦ None、Gen、M/M、F/M、F/F、Multi、Other

Warnings (警告)
◦ 例如: "Graphic Depictions Of Violence"

Character (角色)
◦ 例如: "Peter Parker", " Tony Stark"

Relationship (關係)
◦ 例如: "Peter Parker/Tony Stark"

Freeform (自由發揮)
◦ 例如: "Time travel", "Alternate Universe"

資料來源

找AO3網站上作品數量最多的前20名粉絲圈

每個粉絲圈抓下約1000份最新的作品，按讚數至少100以上

並記錄每一個作品的粉絲圈與標籤，但不包含 "Character"和"Relationship"

共20155筆資料，預處理後總共有6465種標籤

fandom	tags
Harry Potte	{'not rated', 'alternate universe - time travel', 'gen, m/m', 'master of death harry potter'}
Harry Potte	{'marauders era (harry potter)', 'wolfstar', 'the noble and most ancient house of black', 'general audiences', 'marauders', 'siblings', 'jily', 'hogwarts', 'brothers', 'slytherin', 'm/m', 'gryffindor'}
Harry Potte	{'gryffindor bashing', 'sweet girls who love draco', 'viktor is such a sweetheart in love', 'draco malfoy & harry potter friendship', 'confused harry potter', 'so many ships', 'general audiences'
Harry Potte	{'explicit', 'tumblr prompts', 'forbidden prompts', 'm/m, f/m, multi', 'tags in each chapter'}
Harry Potte	{'gotham', 'f/m, gen', 'devil may cry 3', "marauders' era", 'superwholock', 'marauders', 'nightwatch', 'xreader', 'crossover', 'game of thrones - freeform', 'teen and up audiences', 'reader insert
Harry Potte	{'ftm ron weasley', 'rare plants', 'alternate universe - college/university', 'alternate universe - modern setting', 'm/m', 'not rated', 'alternate universe - non-magical', 'pining', 'theatre class', 'trar
Harry Potte	{'alternate universe - canon divergence', 'f/m, gen', 'hogwarts inter-house rivalries', 'book 6: harry potter and the half-blood prince', 'good albus dumbledore', 'good slytherins', 'sirius black
Harry Potte	{'psychological trauma', 'past child abuse', 'angst and hurt/comfort', 'm/m', 'abusive past', 'mature', 'heavy angst'}
Harry Potte	{'forensics', 'murder mystery', 'courtroom drama', 'past alcohol abuse/alcoholism', 'f/m', 'severus snape lives', 'miscarriage', 'suicide', 'mature', 'alternate universe', 'american legal system'}
Harry Potte	{'lily evans potter lives', 'bottom damon salvatore', 'albus dumbledore bashing', 'parent damon salvatore', 'james potter bashing', 'lily evans potter bashing', 'child harry potter', 'adopted harr
Harry Potte	{'goes through most of her years', 'may alan rickman rip', 'severus snape has a heart', 'patronus', 'potter', 'malfoys - freeform', 'professor/student', 'f/m', 'she dates bill snape is jealous', 'this is
Harry Potte	{'gay draco malfoy', 'background relationships', 'past harry potter/ginny weasley', 'mentioned cedric diggory - freeform', 'teen and up audiences', 'bisexual harry potter', 'm/m'}

預處理

1. 將所有字母轉成小寫、刪除所有的[\?\!\.\,]

2. 人工找出一些常被組合使用的詞，將這些組合詞拆開
   (像是 "A"和"B"兩個標籤可能組合成: "A with B" 或 "A and B" 或 "A/B" 之類)

3. 刪除所有出現次數小於某一定值的標籤

分析

統計圖表
- 標籤數量排名
- 標籤特徵
- 粉絲圈特徵

Association Rule Mining
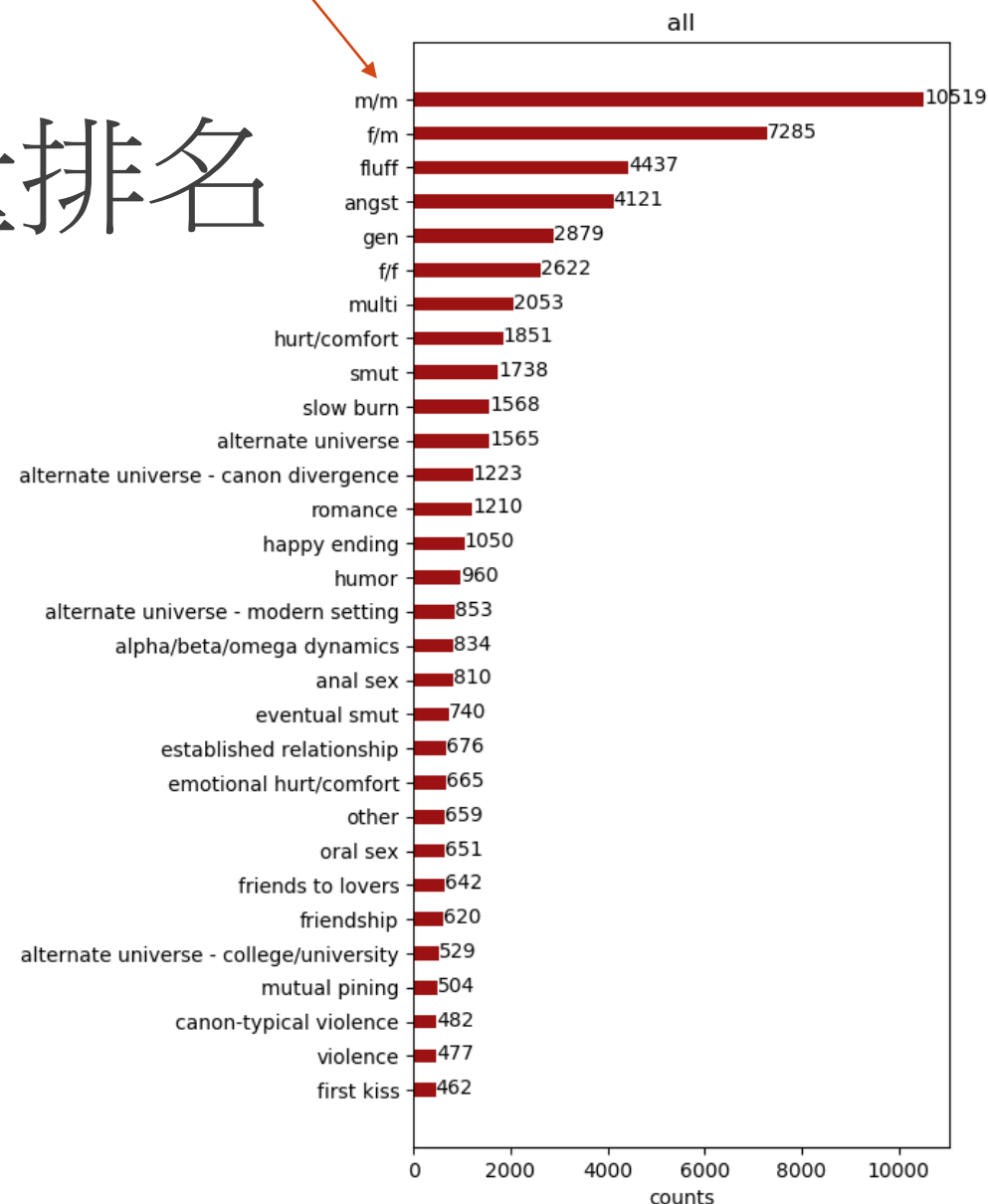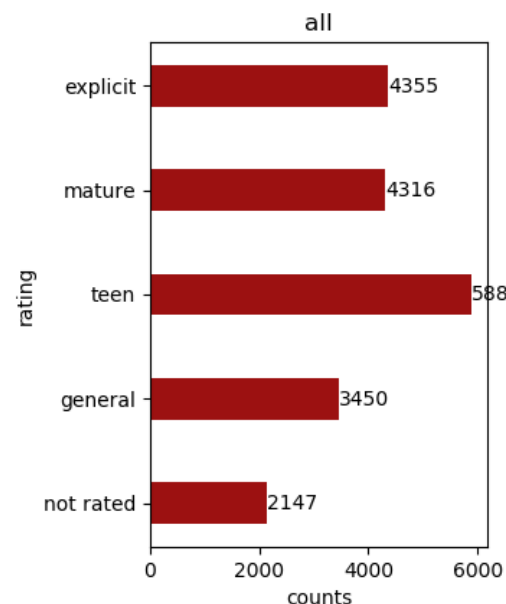- minsup = 0.01~0.02
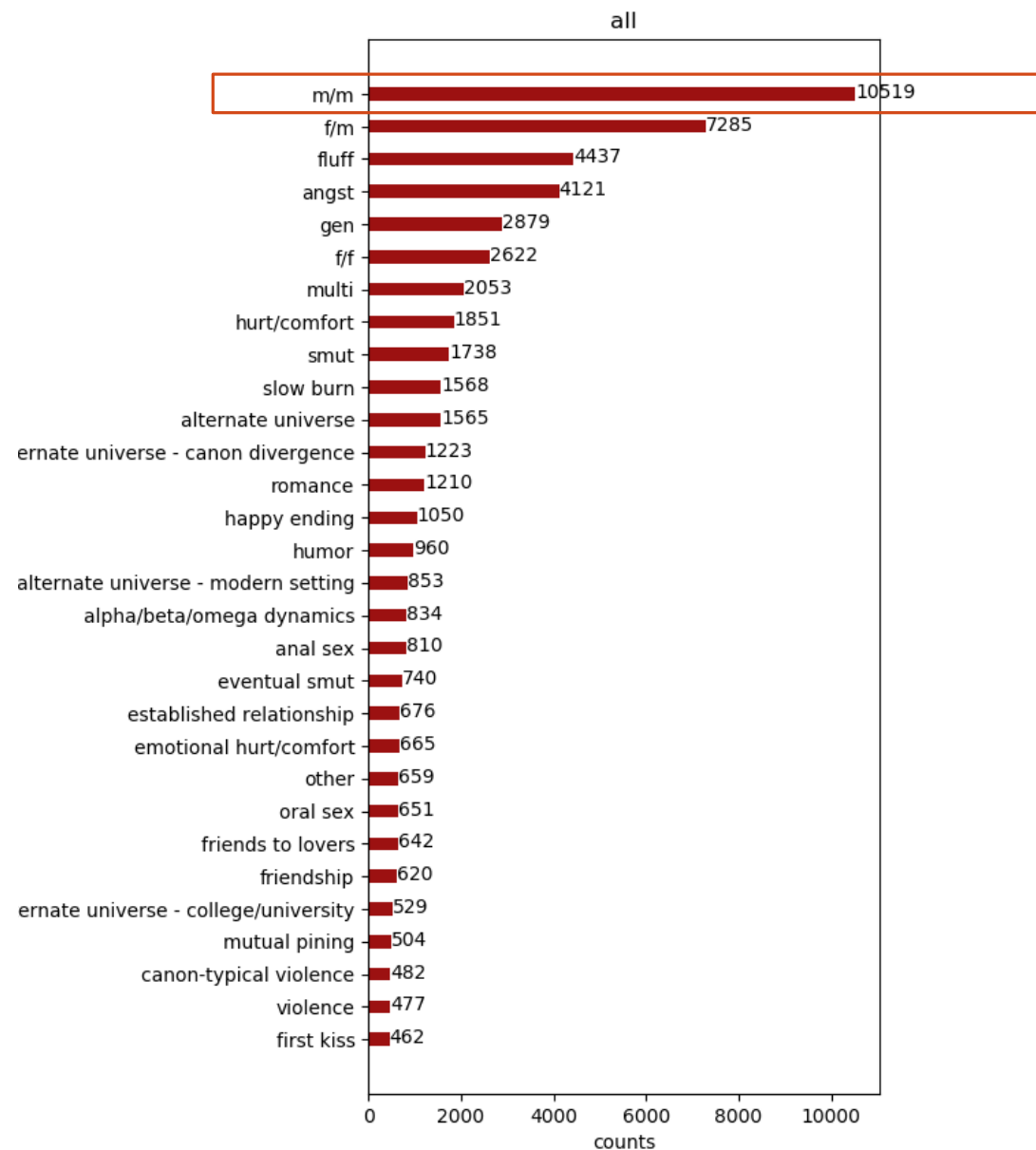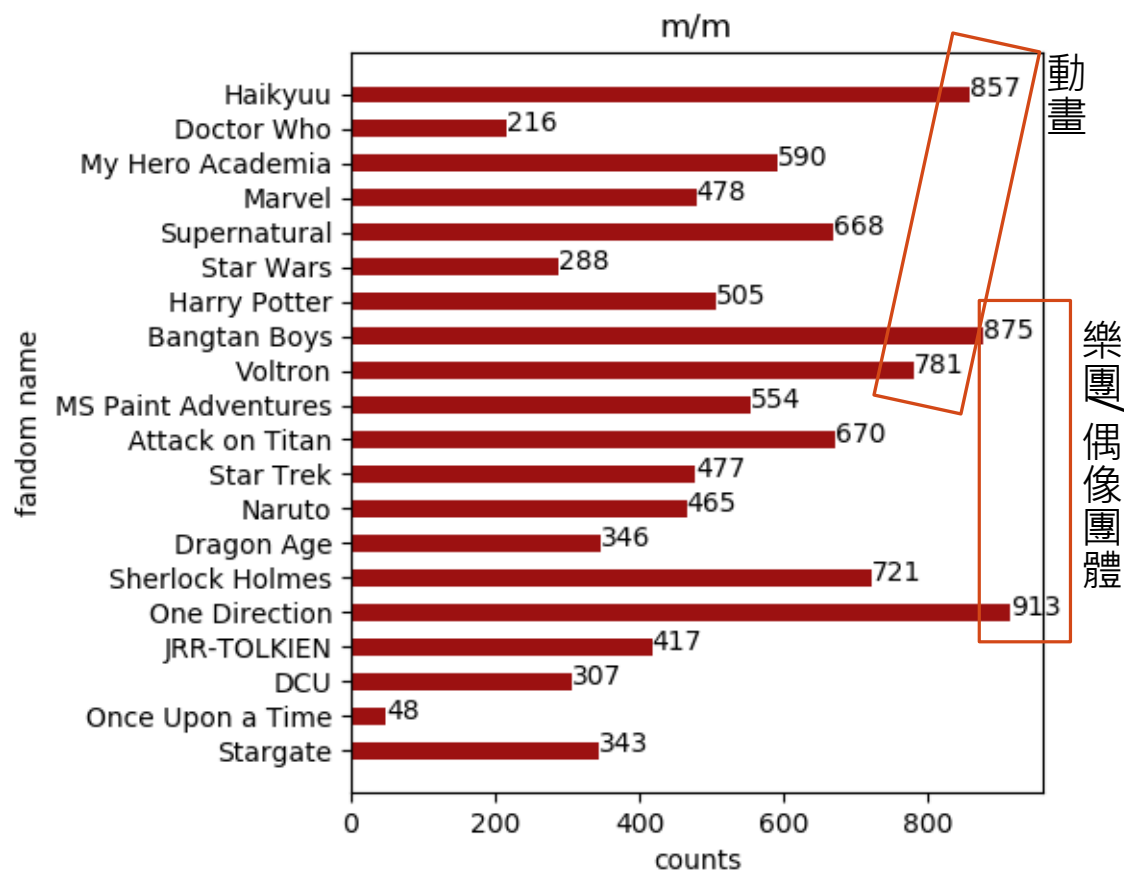- minconf = 0.6~0.8

訓練簡易神經網路

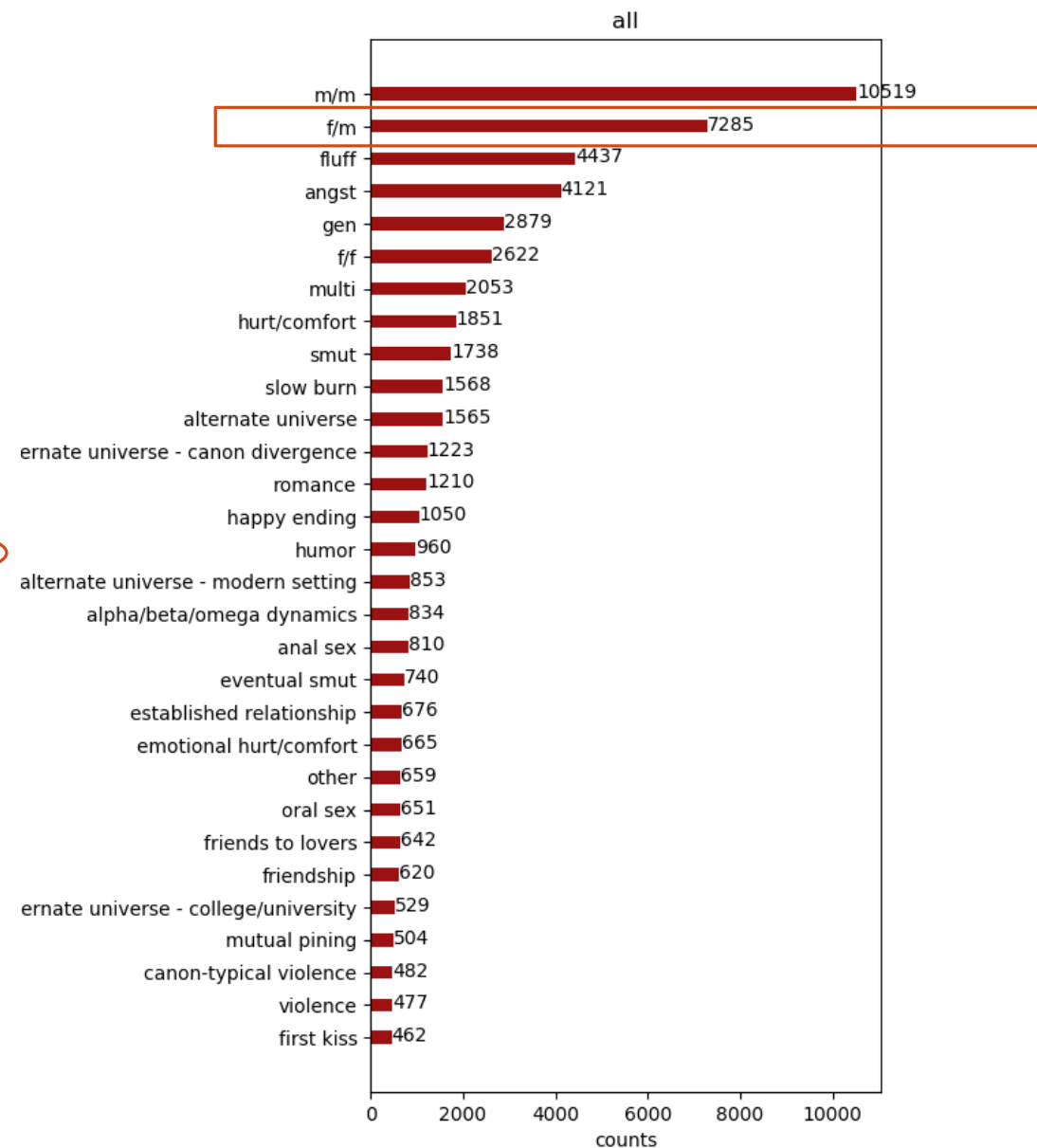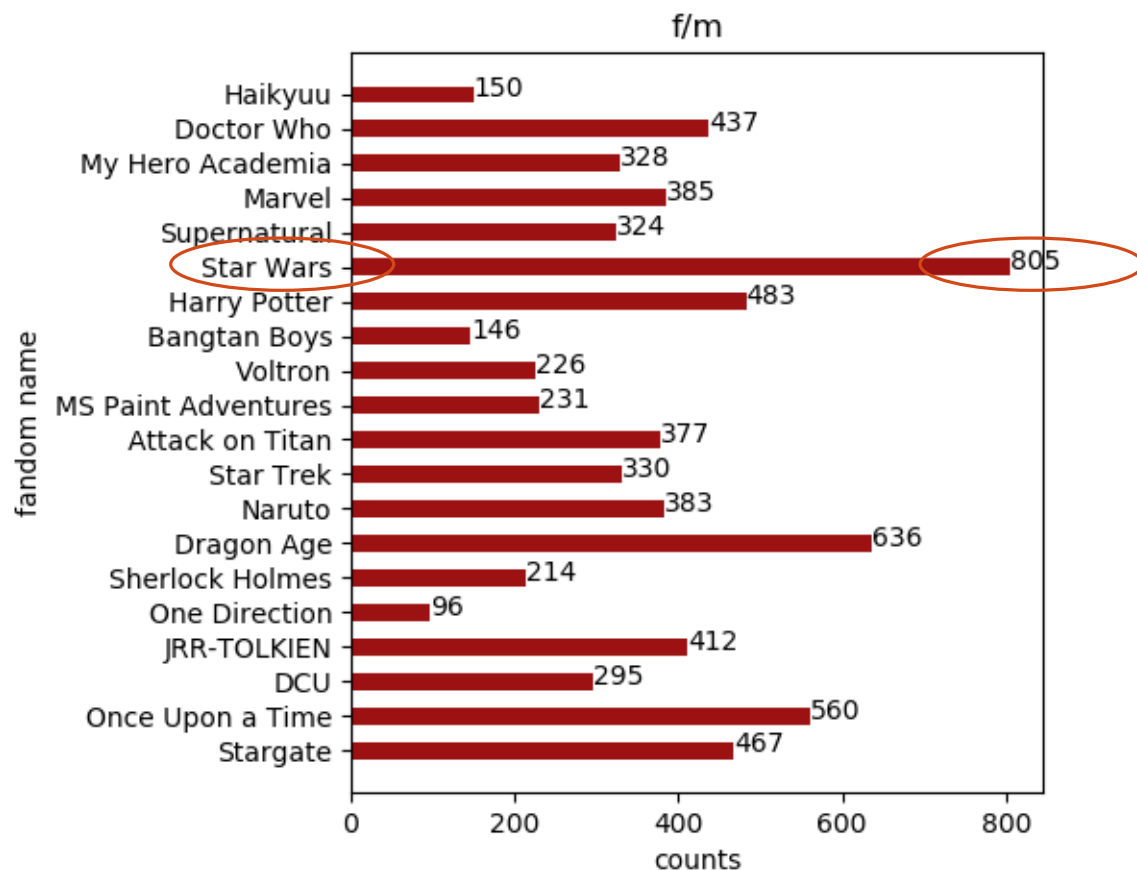# 統計圖表 – 標籤數量排名

術語解釋
Fluff – 甜
Angst – 虐
Smut – 肉(色情)
Hurt/Comfort – 傷害與安慰
Slow Burn – 劇情慢熱

BL佔了整個資料集的一半 10519/20155
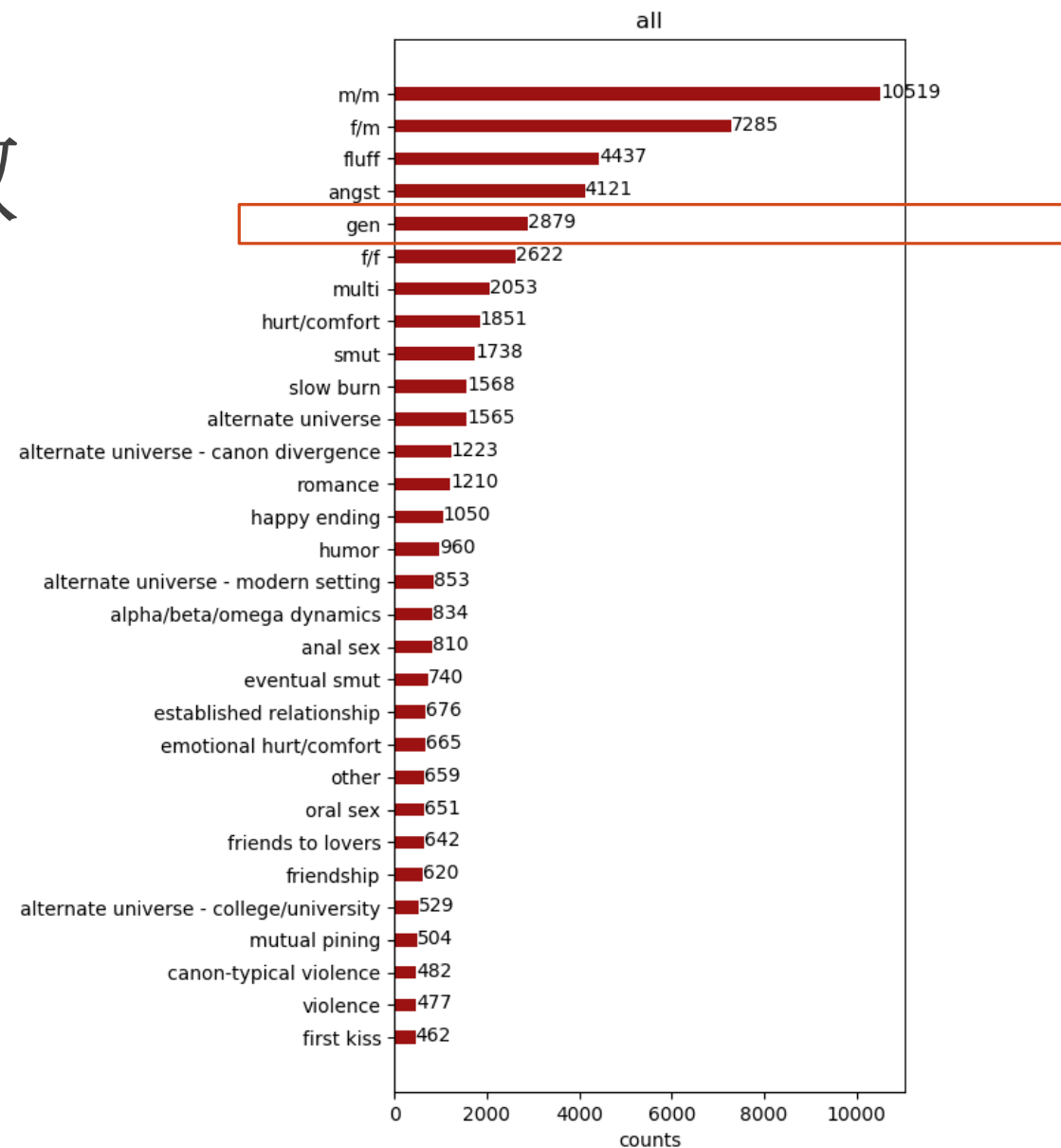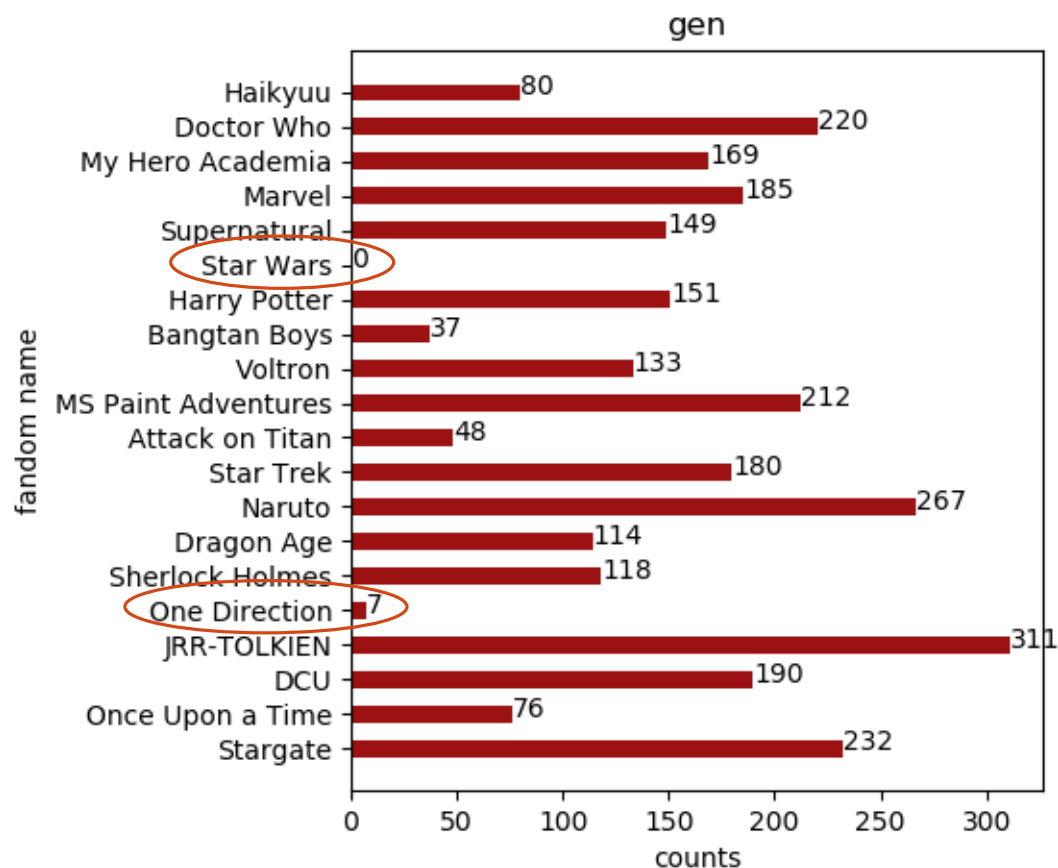
統計圖表 – 標籤特徵

統計圖表 – 標籤特徵

統計圖表 – 標籤特徵

gen

all
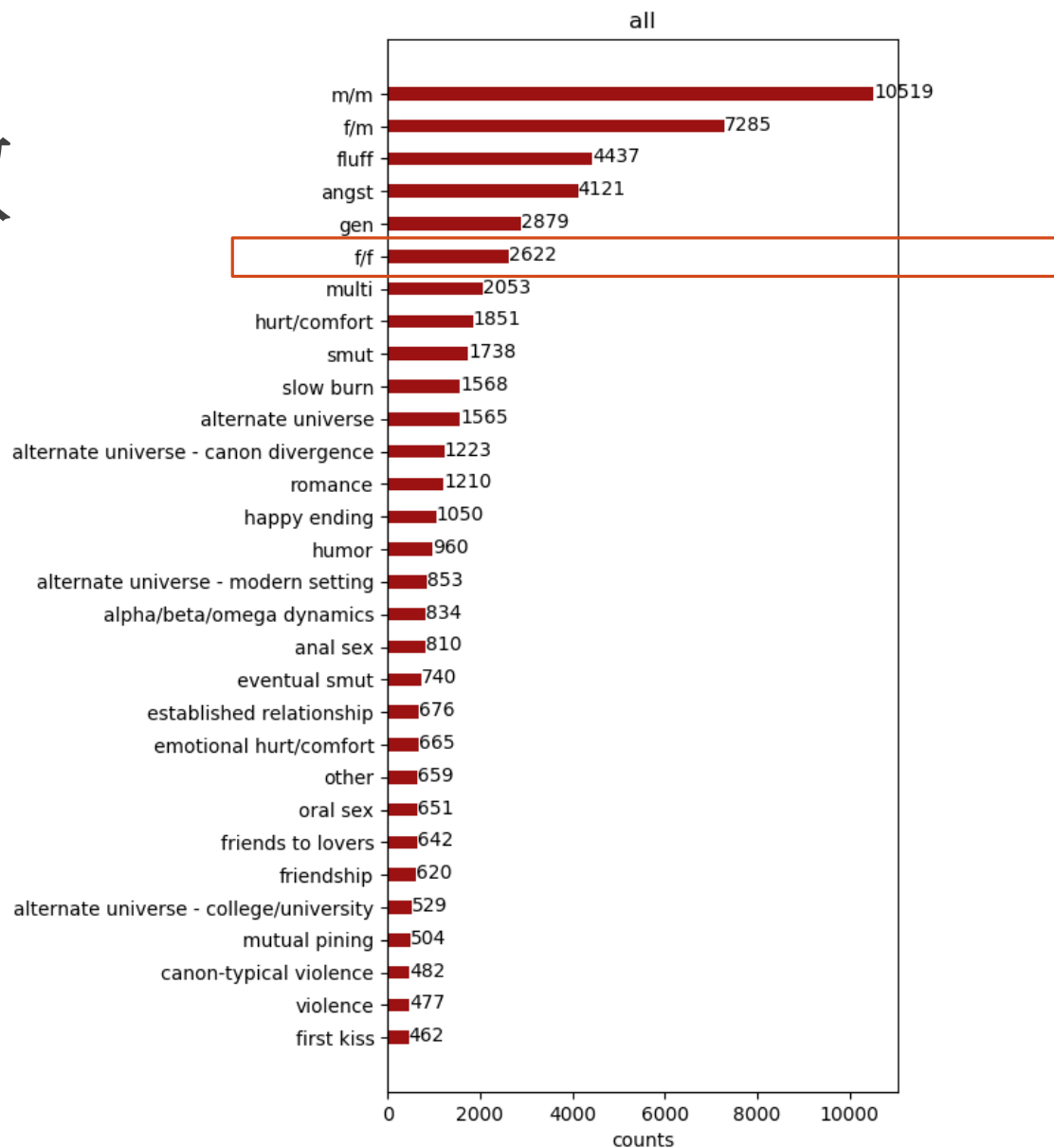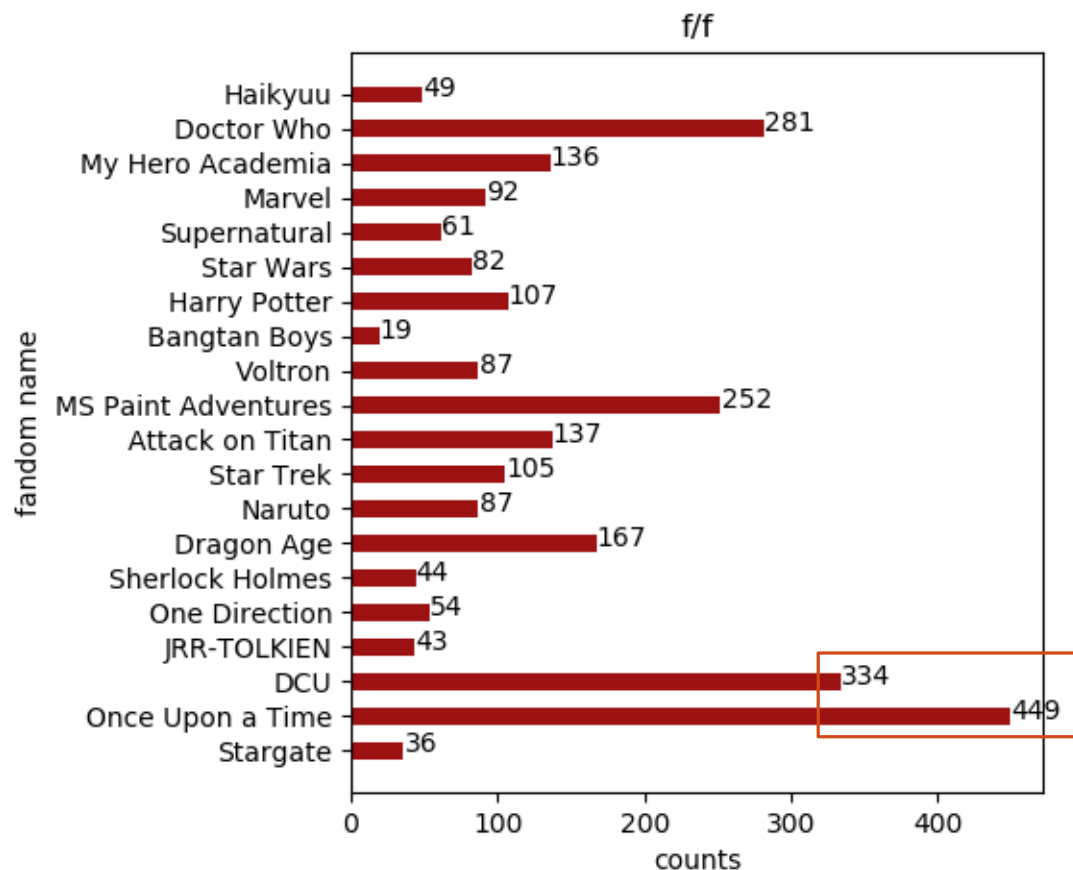
統計圖表 - 標籤特徵

統計圖表 – 標籤特徵

統計圖表 - 標籤特徵

Association Rule Mining

結論：

本本情節終究是最有規則的
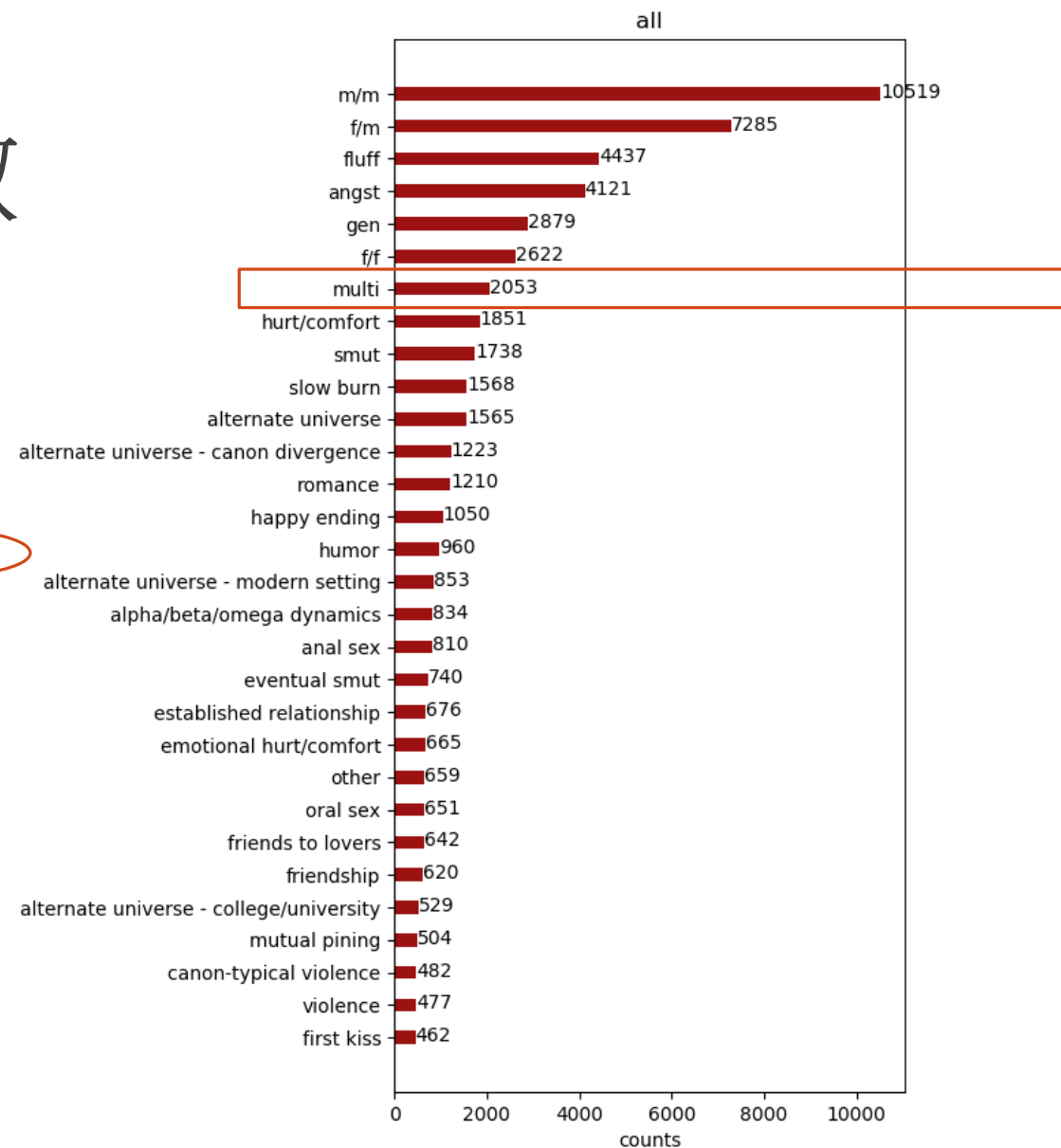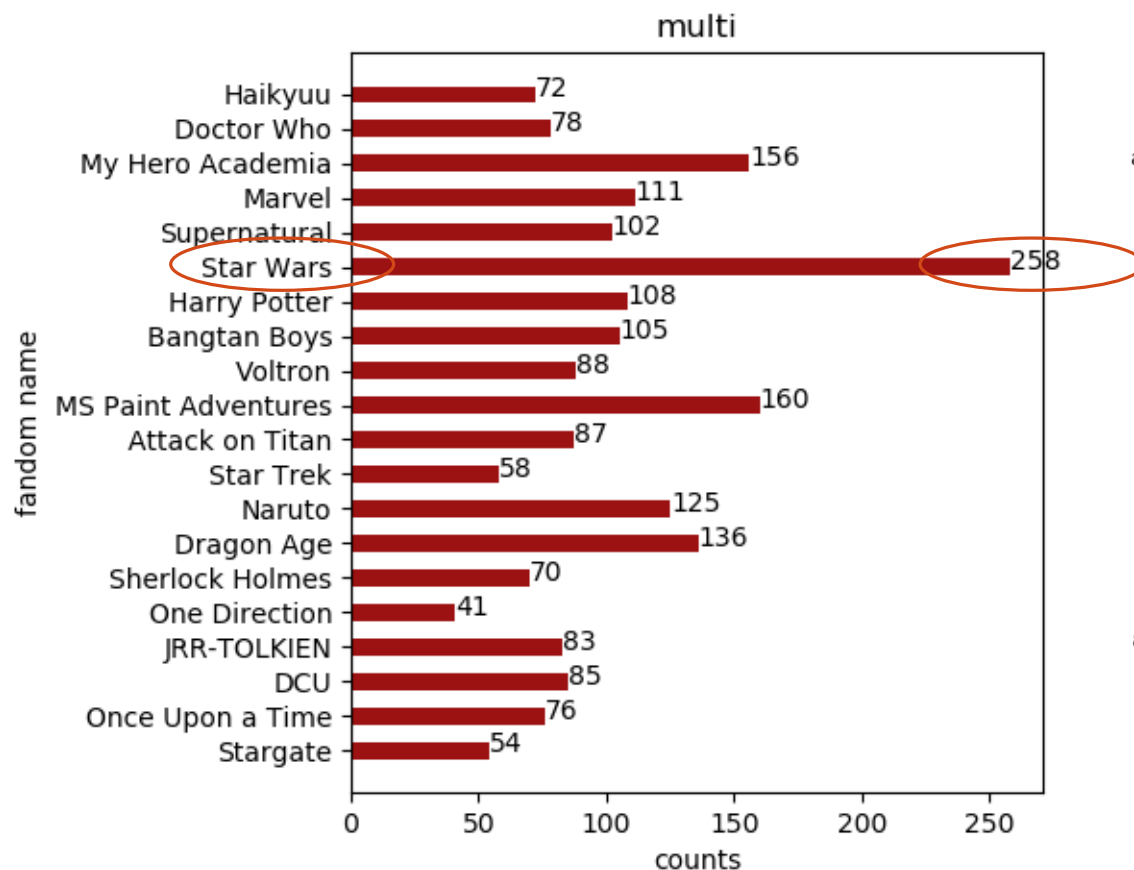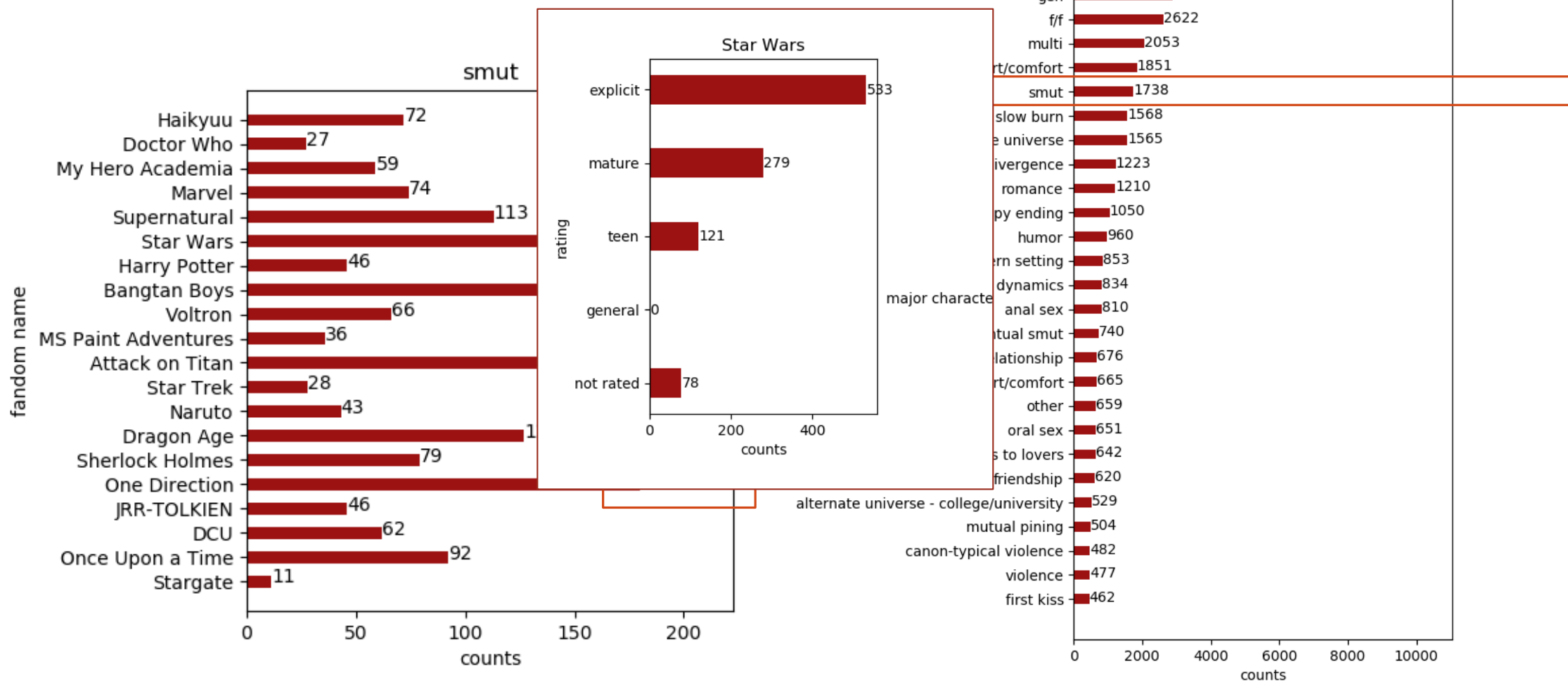
Association Rule Mining

minsup = 0.012

minconf = 0.75

我將結果分成三種
◦ 配對類別的規則
◦ 限制級的規則
◦ 其他

Association Rule Mining - 配對類別

(Supernatural, explicit,) --> (m/m,)

(One Direction,) --> (m/m,)

(One Direction, explicit,) --> (m/m,)

(Bangtan Boys,) --> (m/m,)

(Bangtan Boys, angst,) --> (m/m,)

(Bangtan Boys, mature,) --> (m/m,)

(Bangtan Boys, fluff,) --> (m/m,)

(Haikyuu,) --> (m/m,)

(Haikyuu, fluff,) --> (m/m,)

(Haikyuu, teen and up audiences,) --> (m/m,)

(Voltron,) --> (m/m,)

(Star Wars,) --> (f/m,)

(Star Wars, multi,) --> (m/m,)

(Star Wars, m/m,) --> (multi,)

Association Rule Mining - 配對類別

(alpha/beta/omega dynamics,) --> (m/m,)

(alternate universe - high school,) --> (m/m,)

(anal sex,) --> (m/m,)

(anal fingering,) --> (m/m,)

(blow jobs,) --> (m/m,)

(explicit, anal sex,) --> (m/m,)

(mpreg,) --> (m/m,)

(pining,) --> (m/m,)

(smut, anal sex,) --> (m/m,)

(teen and up audiences, m/m, f/f,) --> (f/m,)

(teen and up audiences, f/m, f/f,) --> (m/m,)

Association Rule Mining - 限制級

(oral sex,) --> (explicit,)

(masturbation,) --> (explicit,)

(oral sex, m/m,) --> (explicit,)

(vaginal sex,) --> (explicit,)

(explicit sexual content,) --> (explicit,)

(anal fingering, anal sex,) --> (explicit,)

(anal sex, m/m,) --> (explicit,)

(oral sex, f/m,) --> (explicit,)

Association Rule Mining – 其他

(smut, eventual smut,) --> (fluff,)

(angst, smut,) --> (fluff,)

(angst, smut, m/m,) --> (fluff,)

(angst, smut, f/m,) --> (fluff,)

(happy ending,) --> (angst,)

(happy ending, fluff,) --> (angst,)

(happy ending, f/m,) --> (angst,)

(happy ending, m/m,) --> (angst,)

(explicit, fluff, f/m,) --> (smut,)

(mating cycles/in heat,) --> (alpha/beta/omega dynamics,)

附錄: 數量前20名粉絲圈 –標籤名(中文翻譯)

1. Marvel (漫威)
2. Supernatural (超自然檔案)
3. Harry Potter - J. K. Rowling (哈利波特)
4. Sherlock Holmes & Related Fandoms (夏洛克·福爾摩斯及相關作品)
5. DCU (DC宇宙)
6. 방탄소년단 | Bangtan Boys | BTS (防彈少年團)
7. Star Wars (星際大戰)
8. Doctor Who & Related Fandoms (超時空奇俠/神秘博士及相關作品)
9. Voltron: Legendary Defender (聖戰士：傳奇護衛)
10. TOLKIEN J. R. R. - Works & Related Fandoms (J·R·R·托爾金 - 個人作品及相關著作)
11. Dragon Age - All Media Types (闇龍紀元及所有改編作品)
12. Star Trek (星艦迷航記)
13. One Direction (Band) (一世代 (樂團))
14. 僕のヒーローアカデミア | Boku no Hero Academia | My Hero Academia (我的英雄學院)
15. Haikyuu!! (排球少年)
16. MS Paint Adventures
17. Once Upon a Time (TV)
18. Stargate - All Media Types (星際之門及所有改編作品)
19. Naruto (火影忍者)
20. Shingeki no Kyojin | Attack on Titan (進擊的巨人)

Thank you