



# Extracting data from MySpace

-an example of large scale analysis of information about individuals

Mike Thelwall  
Statistical Cybermetrics Research Group  
University of Wolverhampton, UK

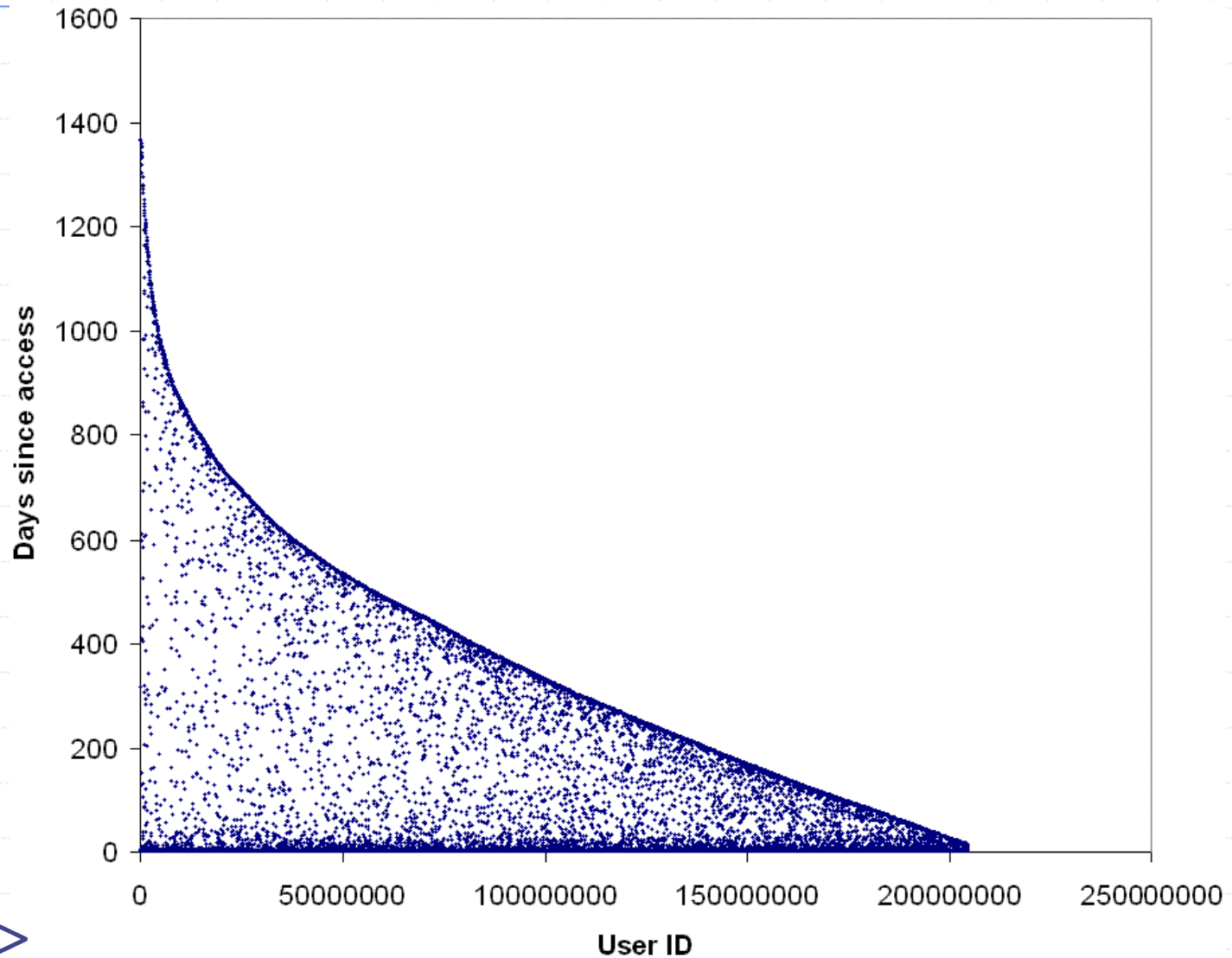
# MySpace members data



1. a random sample of 15,043 members
  2. a systematic sample of 7,627 members who joined on July 3, 2006
  3. 403 members from July 3, 2006
    - ◆ excluded: music sites, ex-members
    - ◆ excluded (normally): members with 0 or 1 friends
- all information on home pages was automatically downloaded and harvested by SocSciBot 4 -> Excel

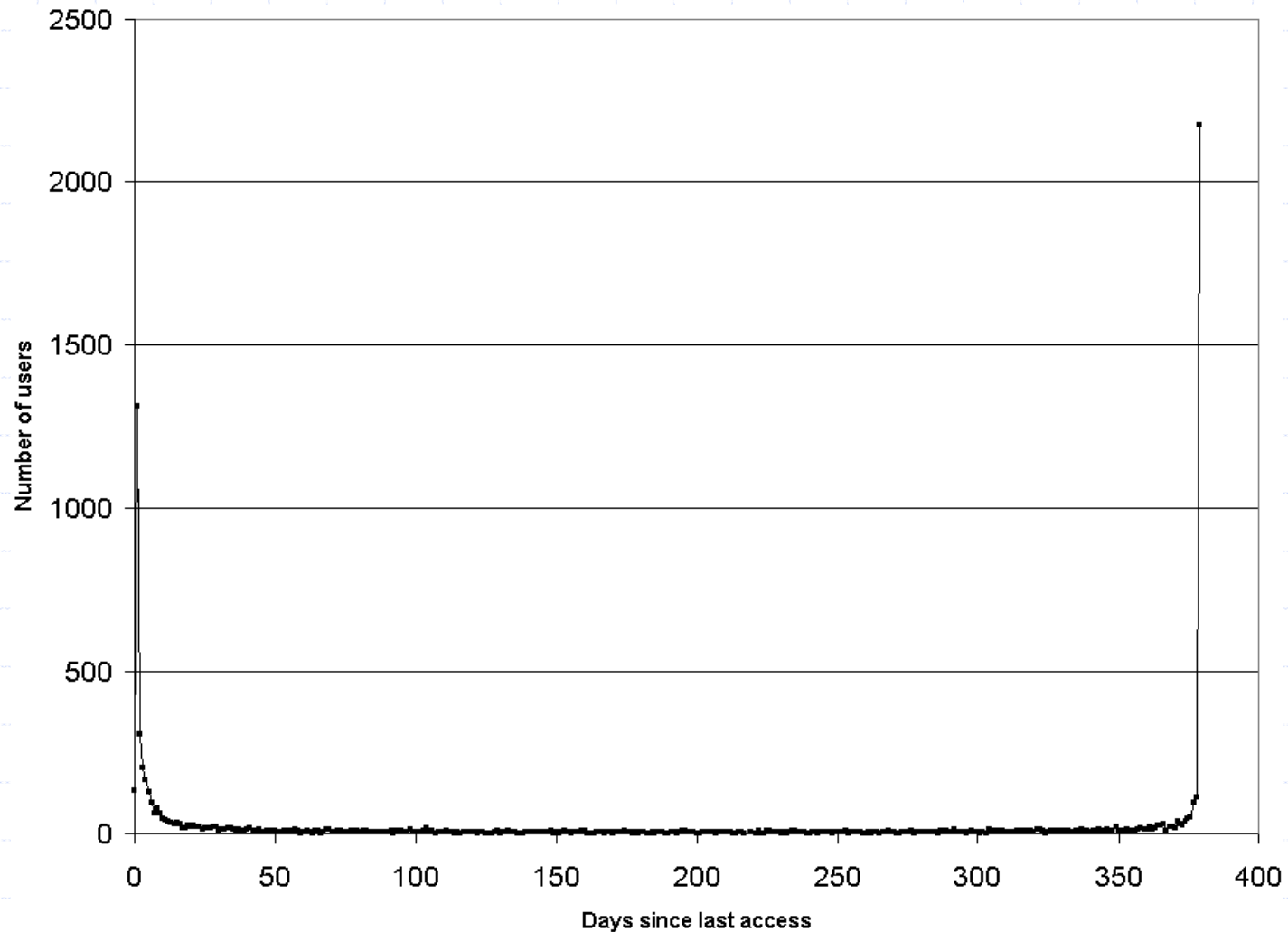
Can access profiles of 1/3 of members with public profiles.

# days since last access - members use MySpace once or frequently



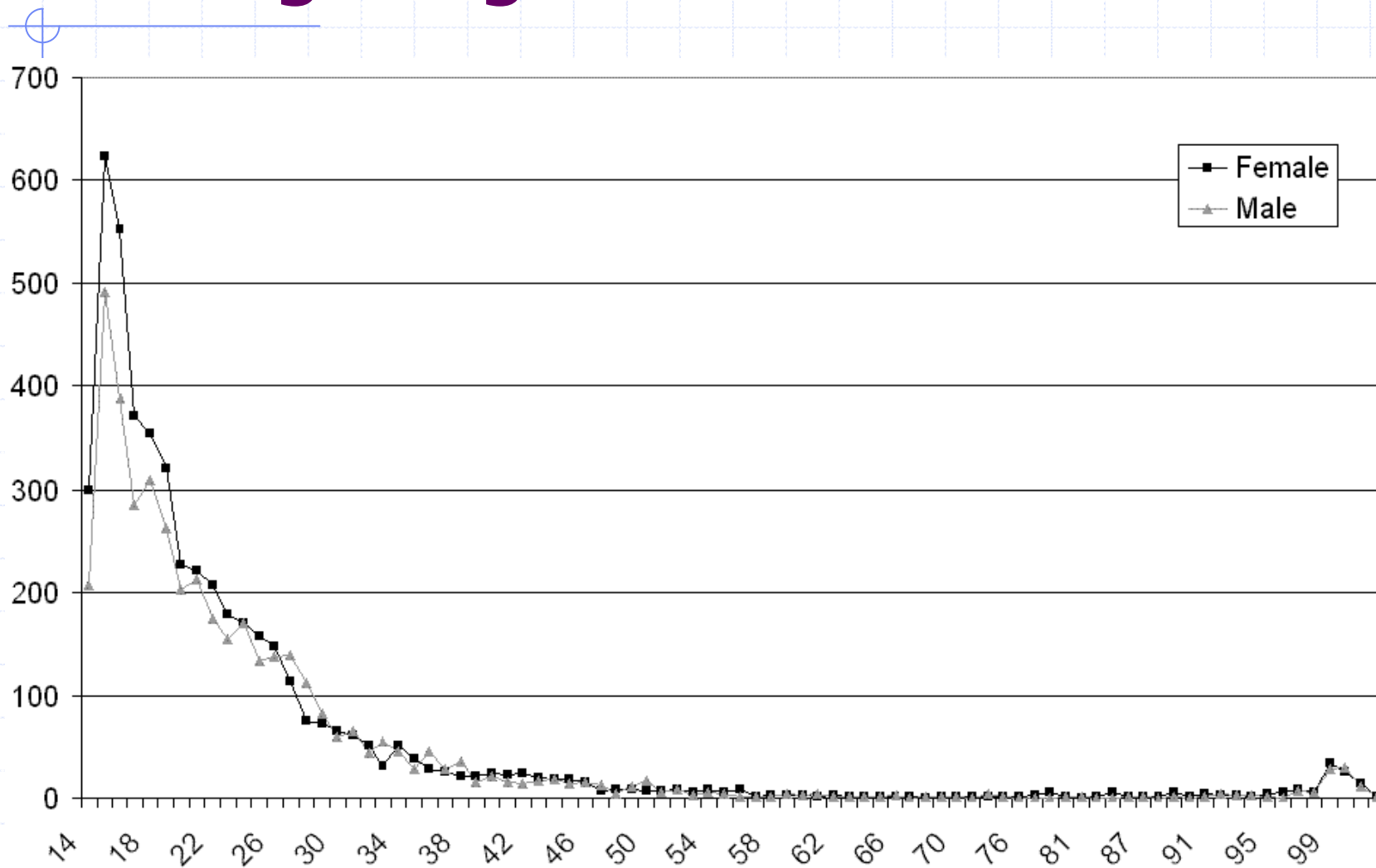
all  
Members >>

# days since last access -members use MySpace once or frequently

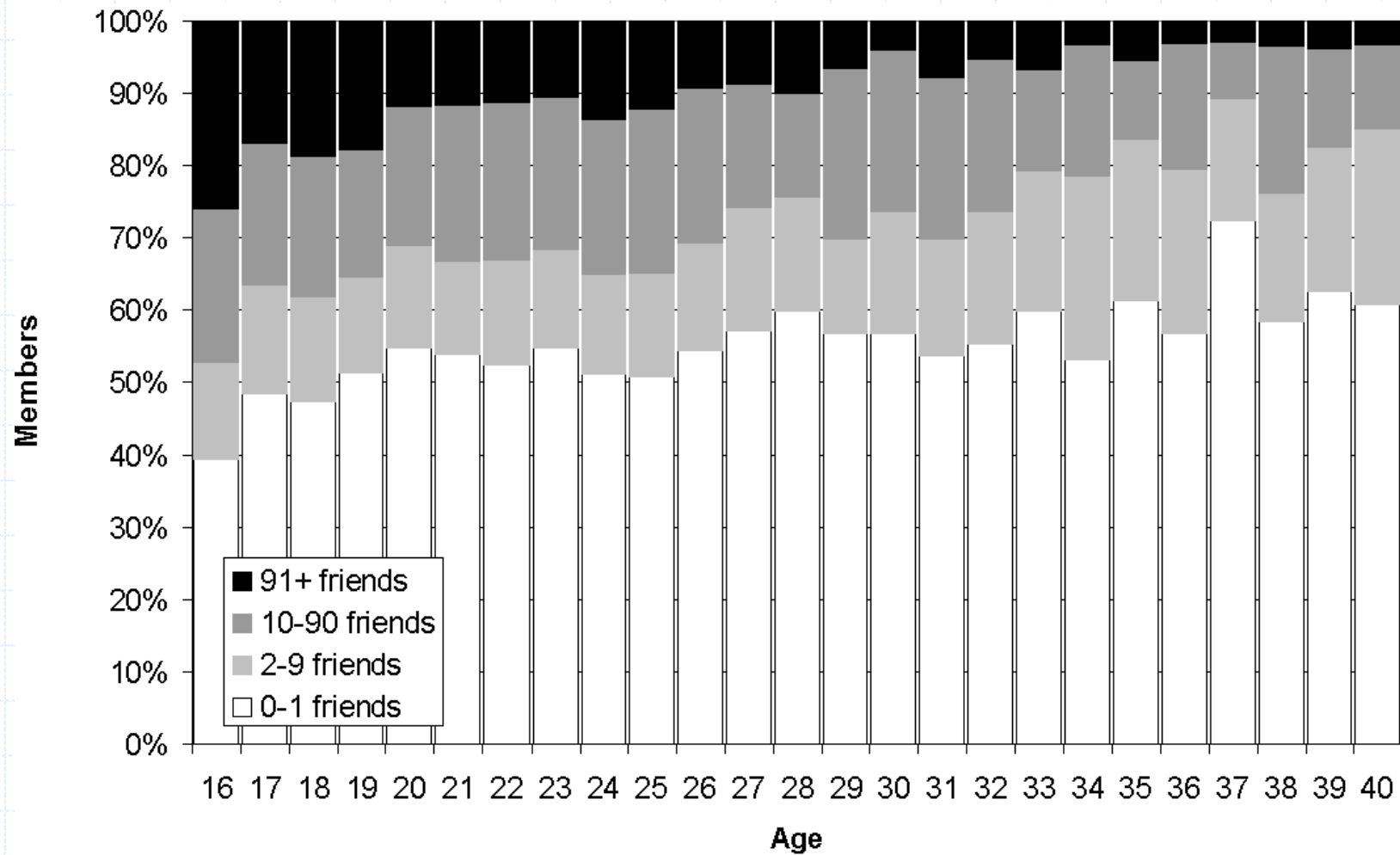


July 3  
members

# MySpace "age" profile: average "age" is 21



# younger members have more MySpace friends



# gender factors



- ◆ female users more likely to be “here for” friendship and male users more likely to be “here for” dating (but only a minority)
- ◆ males and females both preferred to have more female friends and top 8 friends
- ◆ females preferred a greater proportion of female Top 8 friends

(403 data set)

women make the best friends!

# who swears most?

- ◆ for US MySpace home pages:
  - male = more likely to contain strong swearing
- ◆ for UK MySpace home pages
  - male = more likely to contain moderate swearing
  - no difference in strong swearing - possibly more strong swearing in female home pages in the younger age groups
- ◆ apparent *reversal* in gendered strong swearing in the UK for young people

# percentage of profiles containing swearing

	moderate	strong	very strong	sample size
US males 16-19	10%	47%	2%	1,530
US females 16-19	11%	38%	2%	1,287
UK males 16-19	33%	33%	8%	171
UK females 16-19	18%	38%	3%	130

(typical sample size 20-148 for non-web swearing research)

emphatic adverb/adjective OR adverbial booster  
OR premodifying intensifying negative adjective  
(36% of swearing)

- and we r guna go to town again n make a ryt fuckin nyt of it again lol
- see look i'm fucking commenting u back
- lol and stop fucking tickleing me!!
- Thanks for the party last night it was fucking good and you are great hosts.
- That 50's rock and roll weekender was fucking mint!
- Fuckin my space, my arse
- 1/2 d ppl cudnt even speak fuckin english!
- yeah so me and sarah broke up and everythings fucking shit

# personal insult referring to defined entity (28% of swearing)

- ◆ tehe i am sorry.. i m such a sleep deprived twat alot of the time! lol
- ◆ Maxy is the soundest cunt in the world!!!!
- ◆ 3rd? i thought i was your main man number one? Fucker
- ◆ write bak cunt xxx
- ◆ You Godless bastard!
- ◆ You evil cunt! Haha
- ◆ CHEEKY LITTLE CUNT !
- ◆ lucky fuck

## idiomatic set phrase OR figurative extension of literal meaning (23%, mostly male)

- ◆ think am gonna get him an album or summet fuck nows
- ◆ got another copy of the reaction CD (will had fucked the last one lol)
- ◆ qu'est ce que fuck?
- ◆ what the fuck pubehead whos pete and why is this necicery mate
- ◆ Heh long story.. cant be fucked to explain :D

# geography of MySpace Friends

- ◆ [Tobias Escher's 2007 presentation...](#)
- ◆ Tobias is also helping to develop a Perl module [WWW::Myspace](#) for automatic extraction of information

# emotion in Friend comments

- ◆ The majority of comments contain positive emotion (including formal expressions, such as "Love, Mike" or "mike x")
- ◆ A minority (20%) contain negative emotion
- ◆ & People who give positive emotion tend to get it back

Emotion	+ve	-ve
1 (none)	34%	80%
2	28%	6%
3	35%	11%
4	3%	2%
5 (strong)	0%	1%

Emotion strength in 819 random comments

# emotion in Friend comments

- ◆ Positive emotion mainly used by females *and* mainly directed at females
- ◆ No gender difference in negative emotions

	From female	From male
To female	2.4 (+) 1.3 (-)	2.0 (+) 1.3 (-)
To male	2.2 (+) 1.3 (-)	1.7 (+) 1.5 (-)

Average emotion strength in 819 random comments



# gendered MySpace comments?

	Female sender	Male sender
Female recipient	i love you!!!!!!	YOU HAVE A PRETTY COOL PAGE.....:)
Male recipient	i can't believe u would say goodbye to the ducky--he's so cute :( lol	you wish lol ill be there tuesday if you will

# conclusions

- ◆ there are significant gender differences in social network site communication
- ◆ language is very informal but this does not mean that it is wrong in context
- ◆ social network sites are apparently a valuable source of social support



# other references

- ◆ **Thelwall, M. (2008). Social networks, gender and friending: An analysis of MySpace member profiles, *Journal of the American Society for Information Science and Technology*, 59(8), 1321-1330.**
- ◆ **Thelwall, M. (2008). Fk yea I swear: Cursing and gender in a corpus of MySpace pages, *Corpora*, 3(1), 83-107.**
- ◆ **Thelwall, M., Wilkinson, D. & Uppal, S. (in press). Data mining emotion in social network communication: Gender differences in MySpace, *Journal of the American Society for Information Science and Technology*.**
- ◆ All available online – Google Thelwall