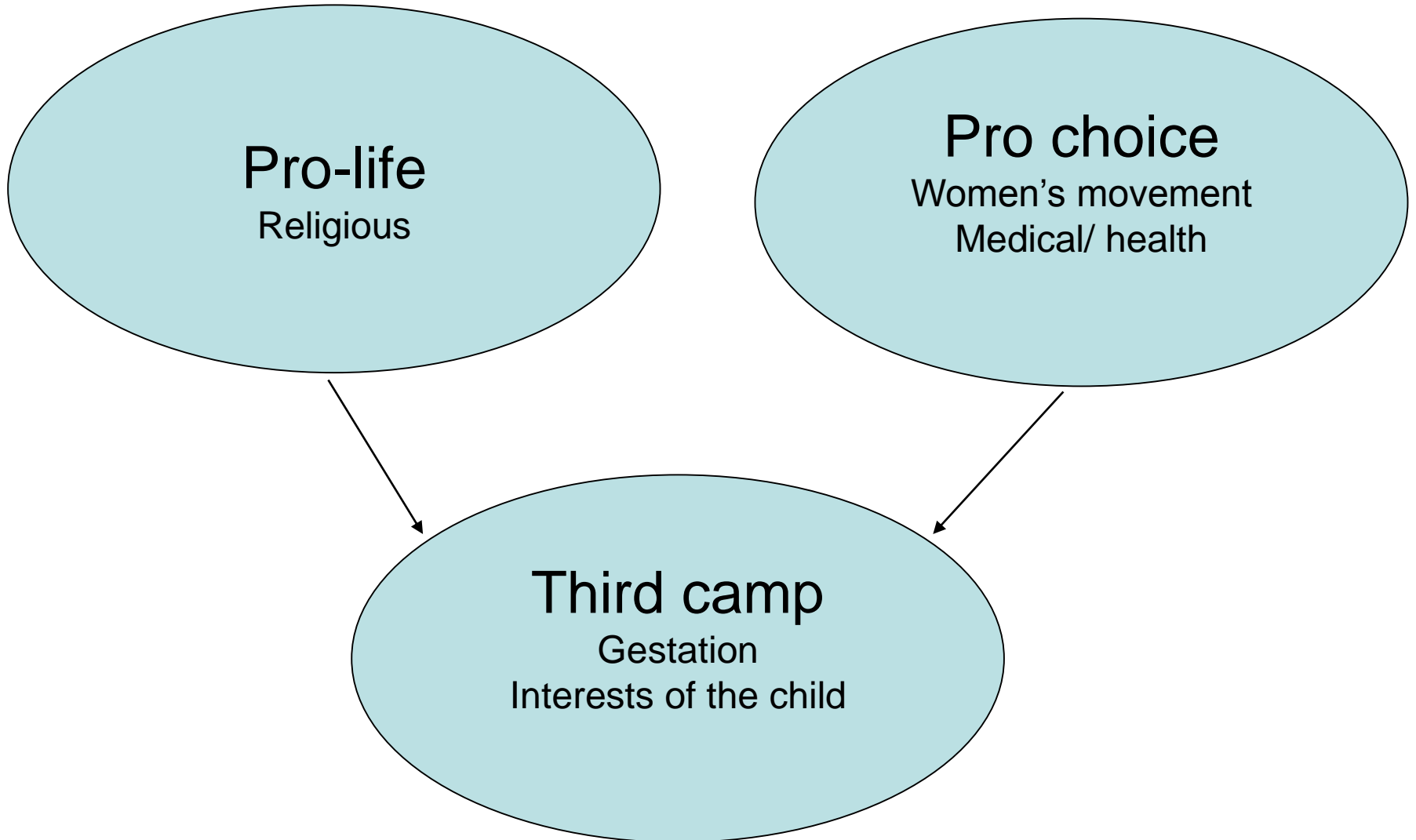


The Visibility of Abortion-Related Information on the World Wide Web

TASA 6 Dec 2005 Hobart
Robert Ackland and Ann Evans
Research School of Social Sciences
The Australian National University

The abortion debate



Research questions

- To what extent do the two sides of the debate link to each other on the web?
- Who is more visible on the web? Pro-life or pro-choice
- What terms characterise the discourse of each side of the debate on the web?

Data collection - seed sites

- Collected first 470 results (pages) from Google query (on 23/10) with keywords: Abortion, Australia. This is referred to as the “seed set”
- 470 seed pages were grouped into 348 “pagegroups” or “sites”
 - For example, the Women Electoral Lobby has two pages in the seed set:
<http://www.wel.org.au/issues/abortion/>,
<http://www.wel.org.au/inkwel/ink956/956aborn.htm>
 - These were placed in the pagegroup
<http://www.wel.org.au/>

Categorisation of seed sites

- `uberlink` automatically codes pages according to country top-level domain (TLD) (.au, .uk etc.) and generic TLD (.edu, .com etc.)
- Seed sites were manually categorised:
 - *abortion stance*: neutral, pro-choice, pro-life, unrelated
 - *site type*: media, government, NGO, political party, religious organisation, abortion provider etc

Composition analysis - seed sites

	cat1				
cat2	neutral	pro-choice	pro-life	unrelated	total
NGO	<u>7</u>	<u>36</u>	<u>26</u>	<u>1</u>	<u>70</u>
abortion provider	<u>0</u>	<u>8</u>	<u>0</u>	<u>0</u>	<u>8</u>
academic-other	<u>20</u>	<u>2</u>	<u>0</u>	<u>0</u>	<u>22</u>
academic-publication	<u>26</u>	<u>11</u>	<u>0</u>	<u>0</u>	<u>37</u>
blogsite	<u>2</u>	<u>3</u>	<u>7</u>	<u>0</u>	<u>12</u>
commercial	<u>6</u>	<u>1</u>	<u>0</u>	<u>5</u>	<u>12</u>
directory/portal	<u>28</u>	<u>5</u>	<u>5</u>	<u>1</u>	<u>39</u>
government	<u>13</u>	<u>1</u>	<u>0</u>	<u>0</u>	<u>14</u>
individual	<u>1</u>	<u>2</u>	<u>6</u>	<u>2</u>	<u>11</u>
info-discussion	<u>13</u>	<u>4</u>	<u>2</u>	<u>0</u>	<u>19</u>
media	<u>37</u>	<u>4</u>	<u>1</u>	<u>0</u>	<u>42</u>
political party	<u>1</u>	<u>6</u>	<u>1</u>	<u>0</u>	<u>8</u>
politician homepage	<u>2</u>	<u>1</u>	<u>1</u>	<u>0</u>	<u>4</u>
religious organisation	<u>0</u>	<u>2</u>	<u>30</u>	<u>0</u>	<u>32</u>
religious-media	<u>1</u>	<u>0</u>	<u>17</u>	<u>0</u>	<u>18</u>
total	<u>157</u>	<u>86</u>	<u>96</u>	<u>9</u>	<u>348</u>

Composition analysis - seed sites (2)

	cat1				
cat3	neutral	pro-choice	pro-life	unrelated	total
Australia	<u>63</u>	<u>37</u>	<u>31</u>	<u>4</u>	<u>135</u>
Canada	<u>3</u>	<u>2</u>	<u>1</u>	<u>0</u>	<u>6</u>
Cocos (Keeling) Islands	<u>1</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>1</u>
Fiji	<u>0</u>	<u>1</u>	<u>0</u>	<u>0</u>	<u>1</u>
Germany	<u>1</u>	<u>0</u>	<u>1</u>	<u>0</u>	<u>2</u>
Nepal	<u>1</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>1</u>
Netherlands	<u>0</u>	<u>1</u>	<u>0</u>	<u>0</u>	<u>1</u>
New Zealand	<u>4</u>	<u>1</u>	<u>0</u>	<u>0</u>	<u>5</u>
Singapore	<u>0</u>	<u>0</u>	<u>1</u>	<u>0</u>	<u>1</u>
Switzerland	<u>1</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>1</u>
United Kingdom	<u>2</u>	<u>4</u>	<u>5</u>	<u>0</u>	<u>11</u>
United States	<u>1</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>1</u>
unknown	<u>80</u>	<u>40</u>	<u>57</u>	<u>5</u>	<u>182</u>
total	<u>157</u>	<u>86</u>	<u>96</u>	<u>9</u>	<u>348</u>

Google rank analysis

- The top-ranked site is pro-life, but otherwise pro-choice sites appear to have higher visibility on Google:
- 34% of the sites in the 1st decile (sites ordered according to google ranking) are pro-choice, compared with 29% being pro-life.

Constructing a connectivity database

- 260 pages in the seed set were crawled
 - Pro-life and pro-choice
- The crawler automatically follows “internal” links on each page, and therefore can potentially crawl the whole site
 - only links to “external” pages were collected
 - new pages found by crawler (i.e. pages not in seed set) were put into connectivity database - these comprise the “first ring” of the connectivity database

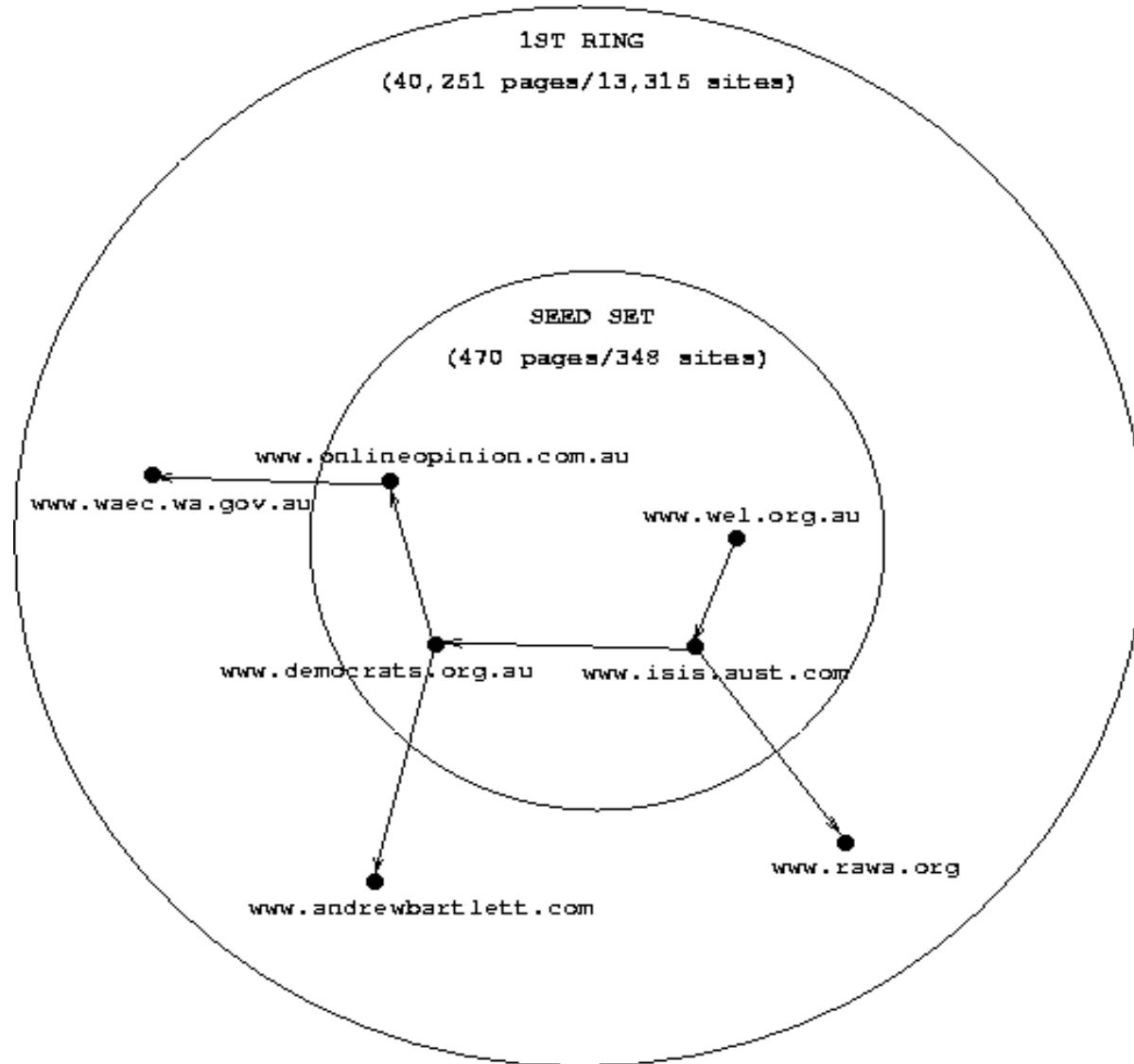
Structure of connectivity database

- Connectivity database
 - rows (observations) - web pages. Each web page assigned a unique identification number
 - columns (fields or variables) - attributes of web pages (e.g. country TLD, generic TLD, page content).
 - One field contains the id numbers of the pages that page i links to.

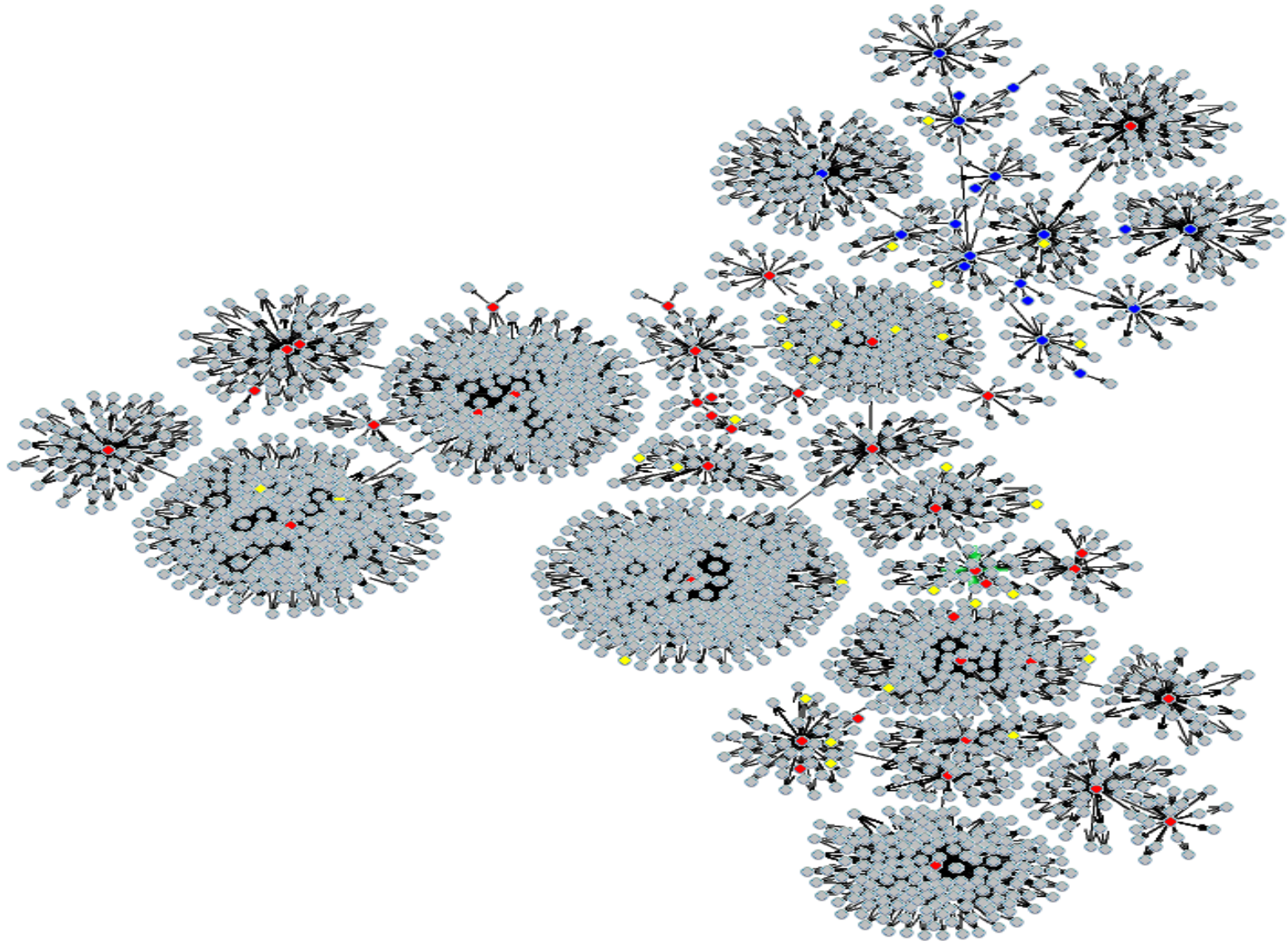
Rings of connectivity database

- Connectivity database resembles concentric rings
 - seed set - 470 pages=348 sites returned by Google
 - first ring - 40,251 pages=13,315 sites linked to by the seed sites. Note: these are *new* pages (i.e. not already in seed set)

Rings of connectivity database



www.wel.org.au - outbound links

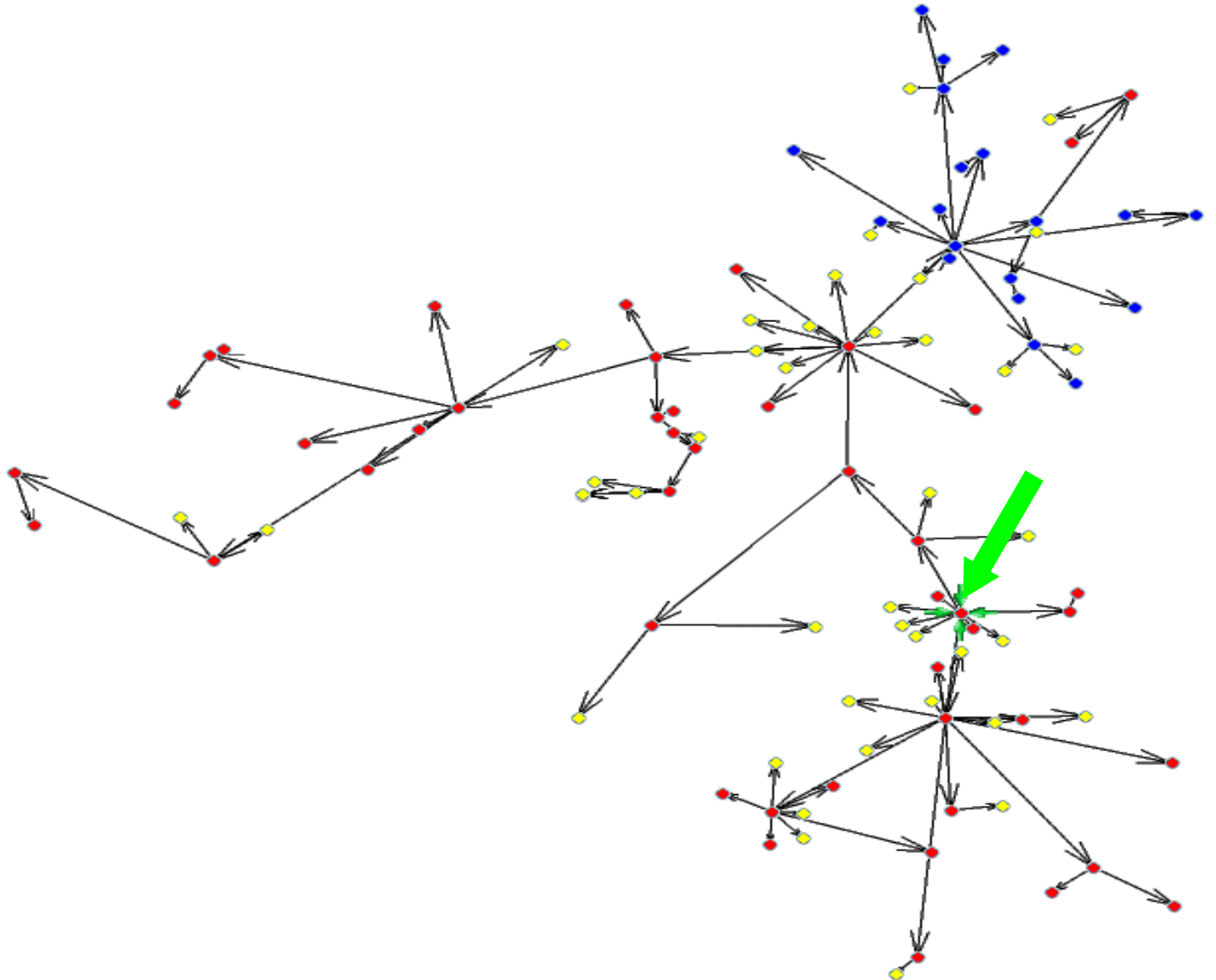


WEL – outbound links

Pro-life

Pro-choice

Neutral



Link analysis

- 86 pro-choice sites - 8221 outbound links (95.6 links per site)
 - 39.7% of links to .com
 - 35.7% to .org
 - 7.5% to .gov
- 96 pro-life sites - 10856 outbound links (113.1 links per site)
 - 58.3% of links to .com
 - 23.6% to .org
 - 5.1% to .net

Link analysis - country TLD

- Pro-choice:
 - 26.1% of links to Australian sites,
 - 3.2% to UK,
 - 1.5% to Canadian,
 - 65.3% to Unknown
- Pro-life:
 - 7.8% of links to Australian sites,
 - 3.3% to UK,
 - 1.3% to Canadian,
 - 83.0% to Unknown

Link analysis – seed site type

- 489 links from pro-choice seed sites to other seeds:
 - 31.5% of links to NGOs
 - 23.7% to media
 - 12.7% to academic-publication
- 604 links from pro-life seed sites to other seeds:
 - 33.3% of links to media
 - 22.8% to NGOs
 - 12.9% to religious organisations

Link analysis - abortion stance

- 489 links from pro-choice seed sites to other seeds:
 - 46.8% of links to neutral sites
 - 43.8% to pro-choice
 - 9.2% to pro-life
- 604 links from pro-life seed sites to other seeds:
 - 46.2% of links to neutral sites
 - 46.0% to pro-life
 - 7.6% to pro-choice

Text content analysis

- What terms characterise the discourse of each side of the debate on the web?
- Frequency counts of the keywords, grouped by abortion stance. Only keywords with frequency ≥ 3

Pro-choice keywords

• abortion	26	• women	5
• abortion clinic	11	• health	3
• TOP	3	• policy	3
• RU486	3	• Tony Abbott	3
• pregnancy	8	• Australian politics	6
• unwanted pregnancy	3	• pro-choice	8
• birth control	5	• choice	3
• emergency contraception	4	• pro-life	3
• sterilisation	3	• society	3
• contraception	3	• environment	3
		• queensland	3
		• Australia	12

Pro-life keywords

• abortion	23	• Marriage	3
• pro-choice	3	• Family	16
• anti-abortion	3	• Euthanasia	6
• pro-life (or RTL)	15	• human life	3
• partial birth abortion	3	• justice	4
• post abortion syndrome	7	• stem cell research	3
• Catholic(ism)	28		
• Christian(ity)	15	• Australia	21
• Anglican	3	• Sydney	4
• Christ	3	• Women's Weekly	3
• church	3	• statistics	3
• clergy	4	• stories	3
• doctrine	4		
• God	6		
• Jesus	9		
• John Paul II	3		
• Pell	3		
• Pope	7		

Summary

- Pro-choice *more* visible on Google
- Pro-life *more* outbound links per site
- Pro-life link to media and religious orgs,
Pro-choice to NGOs and academic pubs
- Pro-choice more likely to link to pro-life
than vice-versa
- Pro-choice dominate by health keywords
- Pro-life dominated by religious, family/life
keywords

Where to from here?

- Content analysis of entire page – not just keywords
- Prediction of abortion stance based on content (machine learning methods)
- Statistical analysis of linking behaviour (p* logit models for social networks)
 - Are there significant differences in linking behaviour of pro-choice versus pro-life groups?