

‘What is.... Web Content Analysis? ’

Rachel Gibson (University of Manchester)

Methods@Manchester ‘What is..’Seminar Series

Web-based methods for capturing and analysing online campaigns (supply-side)

Adapted from offline - `digitized methods` (Rodgers, 2010)

- Online interviews, elite web-surveys
- Online discourse / ethnographic analysis of web `texts`
- Online content/feature analysis
 - 'Classic 'offline approaches - textual/semiotics - online discussion/ads
 - 'Web specific' - 'whole site' approach, covers new types of content and design - documents, email contact, e-news, e-petitions, chat rooms, links, use of flash, accessibility.

Online specific - 'digital methods'

- Hyperlinking and webspheres -VOSON, Issue Crawler, SocScibotOnline
- Blogs -Technorati, BAT, Blogpulse
- Twitter and Facebook - Infoscape Lab, 140kit.com, Twapperkeeper, Discovertext
- YouTube - Tubemogul, Tubekit
- Wikipedia - Wikiscanner

Methodologies for analysing impact of online campaigns ('demand' side)

User Focused

- Focus groups: Price and Capella (2001); Stromer-Galley and Foot (2002)
- Opinion Surveys: Bimber and Davis (2003); Gibson & McAllister (2007)
- Experiments: Iyengar (2002); Lupia & Philpott (2005)
- User statistics: Hitwise, Sitemeter, Alexa, Google Analytics
- Computer tracking devices: Phorm (Chen and Vromen, 2010)

Online discourse analysis / ethnography of Web Campaigns

- Focus of the analysis is on web content as socio-cultural 'texts'.
- Described and explored from a qualitative perspective to uncover underlying meaning and significance.
- Examples of studies:
 - Markham (1998) Howard (2002) Hakken (1999) Hine (2000)
 - Benoit and Benoit (2000) 'The Virtual Campaign: Presidential Primary Websites in Campaign 2000' *American Communication Journal*
 - Warnick (1998) 'Appearance or Reality? Political Parody on the Web in Campaign 96' *Critical Studies in Mass Communication*

Web specific content/feature analysis

- Focus is on the individual web page/site as the unit of analysis
- Systematic set of criteria based on items or features of web pages developed into a coding scheme that is applied to sites to provide quantitative measures of content, functionality, usability, and design features.
- Key Questions:
 - Campaign Style/ Approach - is it making elections and parties more participatory?
 - Inter-party power - is it raising the profile of smaller / marginal parties?

Data collection

Archive-It

A subscription service that allows institutions to build and preserve collections of digital content. Archive-It partners harvest, catalog, manage, and browse their archived collections. Collections are hosted at the Internet Archive data center and are accessible to the public with full-text search. <http://www.archive-it.org>

Sites to include?:

- Organisations - all regular and campaign sites for parliamentary parties plus interesting 'other'. BNP, Greens, ?
- Individual sites - all leader sites, leading ministers? shadow cabinet?
- Blogs - Party/Leader (official); Conshome, Labhome, Libdem voice + A-list?
- Official YouTube channels - Party /Leader
- Facebook profiles - Party/Leader
- News organisations' election sites - BBC, Guardian, Times/Telegraph

Fully automated quantitative data collection and analysis

Internet Research: Electronic Networking Applications and Policy

Volume 10 · Number 1 · 2000 · 31–43

Table I Domain-independent, automatically collected classification criteria

Criteria	Variables
Content	No. documents [total]
	Kbytes downloaded [total/text only]
	No. file types [distinct extensions]
	No. images [total/distinct]
Interactivity	No. forms [total/distinct/fields]
	No. documents w/JavaScript [total]
	No. Java applets [total/distinct]
	No. MailTo-links [total/distinct]
Navigation	Frames [yes/no]
	No. internal links [total/distinct/broken]
	No. external links [total/distinct/broken]
	No. anchors [total]
	No. links to anchors [within/between documents]

Figure 5 Top left corner of the WebAnalyzer's two-dimensional output matrix

	D	E	F	G	H	I	J	K	L	
		anchor count	anchor linkcount	applet count	external linkcount	form count	formfield count	frames	image count	int
1	Name									
2	http://www.2020vision.org/	70	16	0	790	4	58	0	248	
3	http://www.arcweb.org/	8	17	0	60	0	0	0	835	
4	http://www.cec.org/english/	1897	2012	0	1454	8	115	1	2253	
5	http://www.cfact.org/	1	2	0	71	4	35	1	183	
6	http://www.conservation.org/	207	237	6	556	27	125	0	26938	
7	http://oneearth.org/fs_index.htm	0	0	0	3	0	0	1	78	
8	http://www.earthisland.org/	301	356	183	1229	62	866	1	6053	
9	http://www.earthpledge.org/	7	11	0	72	1	7	1	780	
10	http://www.earthsummitwatch.org/	259	290	0	523	1	1	1	1015	
11	http://gaia.earthwatch.org/	58	68	0	827	20	229	1	6099	
12	http://www.igc.org/igc/econet/	0	0	0	48	2	2	0	14	
13	http://www.envirolink.org/	604	653	0	12538	32	327	1	1537	
14	http://www.edf.org/	489	521	0	158	86	302	0	8380	
15	http://www.eli.org/	279	32	0	574	3	53	1	1260	
16	http://www.foe.org/	966	911	0	842	5	31	0	6011	
17	http://www.greenpeace.org/	1738	1982	4	2157	53	259	1	21867	
18	http://www.worldforum.org/	11	0	0	209	61	112	0	5729	
19	http://www.irm.org/	1029	676	0	408	47	316	1	3146	

Mixed method

Manual coding/ data collection, quantitative measures and analysis

3 Stages / Approaches

- Web 1.0 static/fixed homepages -2 Dimensional 'Flat'
 - Large N - Latent Variable Analysis Approach Gibson and Ward (2000)
 - Small N - Structural-Functional Approaches (Gibson and Ward, 2002) Foot & Schneider (2002)
- Web 2.0 or 1.5? - Updated web 1.0 schemes to incorporate web 2.0 elements.
 - Party and Campaign home pages - Lilleker and Jackson (2010) Gulati & Williams, (2006)
 - NSM - Stein, 2009;
- 'Action Centers' - MyBO, Membersnet, MyConservatives, LibDemACT (Gibson, 2010)

TABLE 1
Indicators for Information and Communication Flows

<i>Feature</i>	<i>Measure</i>
Downward Information Flows	
Organizational history	→ word count
Documents (e.g., manifesto)	→ summed word count
Policies	→ summed word count
Values/ideology	→ word count
Structure	→ word count
Who's who	→ average word count per person
Newsletters	→ number available (including archived copies)
Media releases (e.g., speeches)	→ number available (including archived copies)
Candidate profiles	→ average word count per person
Leader profile	→ word count
Election information	→ summed word count
Event calendar	→ present (1), absent (0)
Frequently asked questions	→ summed word count
Negative campaigning	→ summed word count
Credit claiming	→ summed word count
Targeted pages	→ number of groups targeted
Upward information flows	
Donation	→ download form (1), online inquiry form (2), online transaction, (3) no reference (0)
Merchandise	→ download form (1), online inquiry form (2), online transaction (3), no reference (0)
Cookies	→ present (1), absent (0)
Lateral/horizontal information flows	
Partisan links	→ number of groups supportive of organization goals
Reference links	→ number of general information sites
Internal links	→ number of suborganizational groups
Interactive information flows: Asynchronous	
Download logos/posters	→ number of opportunities
Site search	→ present (1), absent (0)
Online games/gimmicks	→ present (1), absent (0)
E-mail contact	→ number of addresses offered
E-mail feedback	→ e-mail address (1), e-mail requesting comments (2), online form/poll (3), no reference (0)
Join e-mail list	→ present (1), absent (0)
Join online campaign	→ present (1), absent (0)
Membership	→ download form (1), online inquiry form (2), online transaction (3), no reference (0)
Absentee voting	→ information (1), e-mail inquiry contact (2), download inquiry/registration form (3), no reference (0)
Bulletin board	→ present (1), absent (0)
Interactive information flows: Synchronous	
Chat room	→ present (1), absent (0)
Online debate	→ number of opportunities to debate candidates/leaders (set time period)

Information provision

Additive index—1 point assigned for each item present (0–13)

Organisational history

Structure

Values/ideology

Policies

Documents (ie manifesto, constitution)

Newsletters

Media releases (ie speeches, statements, interview transcripts, conferences)

People/who's who

Leader profile

Candidate profiles

Electoral information (statistics, information on past performance)

Event calendar

Frequently asked questions

Resource generation

Cumulative index (0–9) comprised of three ordinal indices

(i) Donation index 0–3

(ii) Merchandise purchase index 0–3

(iii) Membership index 0–3

For each index (1) download form and post (2) online enquiry (specific e-mail or online form) (3) online transaction (0) no references made.

Networking

Total count of all links out of the site to other *independent* sites (0–n)

- (i) **Internal links** include those to local parties, MP's home pages, international branches *if they are independent* of the main party site.
- (ii) **Partisan links** are those to other parties and organisations that are supportive of the party's goals. For example, does the Labor Party link to the Trades Union Congress site or the American Democratic Party?
- (iii) **Reference links** are those to neutral or news/educational sites such as news broadcasters, newspapers, parliamentary/government sites, national libraries, etc.

Participation

Two ordinal indices and one count. Non-cumulative

- Openness** (0–n): Total count of all e-mail contacts listed on site.
- Feedback index** (0–3): Ordinal (1) e-mail address on the site; (2) e-mail address explicitly focused on soliciting comments; (3) an online form or poll to submit views offered.
- Interaction index** (0–4): Ordinal (1) games/gimmicks to play; (2) bulletin board or guestbook to post views; (3) chat room for real-time discussion; (4) opportunity for online debate with leader/senior organisation figures.

Campaigning

Divided into two components, both of which centre on parties' explicit efforts to recruit votes on their site. The passive or 'pull' campaigning index captures aspects of the site designed to pull voters in to reading the party message. The active or 'push' campaigning index refers to items that seek to push the party message onto the voter.

Pull campaigning: Cumulative index (0–8) combining three measures:

- (i) **Negative campaigning** 0 = absent 1 = present
- (ii) **Credit claiming** 0 = absent 1 = present
- (iii) **Targeting index** 0–6 additive index, one point assigned for each of the following groups of voters targeted on the site:
 - Marginal constituency
 - Women
 - Business
 - Labour
 - Issue-oriented (eg environmental)
 - Identity-based (eg women, ethnic minorities)
 - Younger voters

Push campaigning: Additive index—1 point for each item present (0–5):

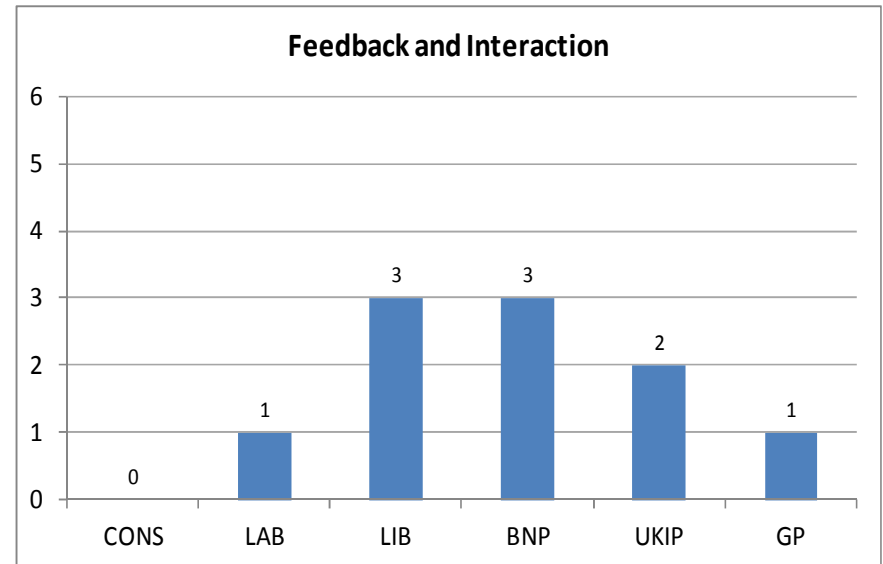
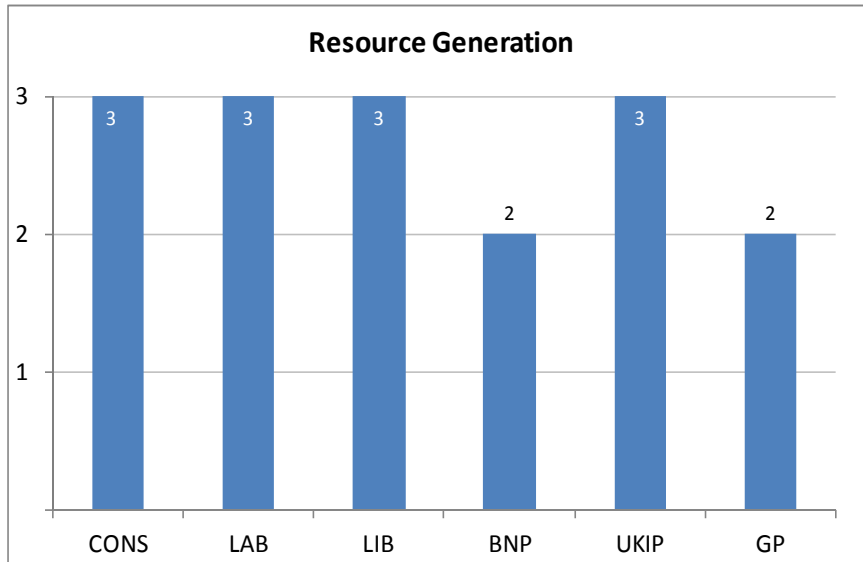
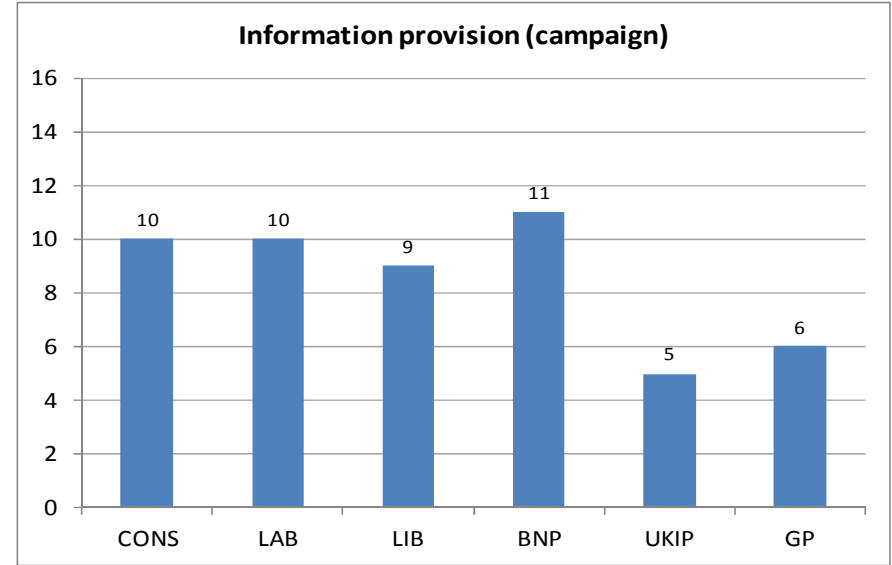
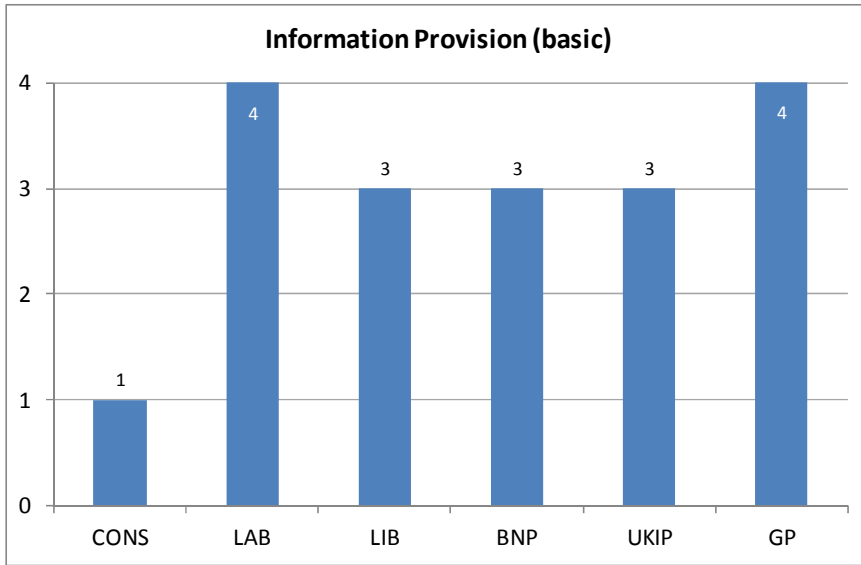
- Cookie
- Join an e-mail update list
- Become online volunteer

TABLE 2
Indicators for Site Delivery

<i>Feature</i>	<i>Measure</i>
Graphics/flashiness	→ total number of images or pictures
Multimedia/dynamism	→ moving icons (1), audio (2), video (3), live streaming (4)
Freshness	→ updated daily (6), 1 to 2 days (5), 3 to 7 days (4), every 2 weeks (3), monthly (2), 1 to 6 months (1), more than 6 months (0)
Accessibility (in principle)	→ no frames option (+1), text-only option (whole site) (+1), text-only documents to download/print (+1), foreign language translation (+1), blind/visually impaired software (+1)
Accessibility (in practice)	→ (a) site working (1), inaccessible (0) (b) size of home page in Kb
Navigability	→ navigation tips (+1), number of search engines (+n), home page icon on each page (+1), major site area links/menu bar on each page (+1), site map/index (+1)
Responsiveness (speed)	→ same day (5), 1 to 2 days (4), up to 1 week (3), up to 2 weeks (2), up to 1 month (1), more than 1 month (0)
Responsiveness (quality)	→ number of words, (0) if irrelevant to query
Visibility	→ (a) number of links in (b) yahoo party index, present (1), absent (0)

Web 1.0: UK parties General Election 2010

Feature	Items	Index																
Information provision (basic)	List of local/regional branches – address/phone List of local/regional branches – website List of local/regional branches – email E-campaigns other than election	0-4																
Information provision (campaign)	<table border="0"> <tr> <td>E-news sign up</td> <td>Schedule/calendar</td> </tr> <tr> <td>RSS/WAP sign up</td> <td>Archive reports, news, speeches</td> </tr> <tr> <td>Downloads offline</td> <td>Audio podcasts</td> </tr> <tr> <td>Manifesto (any format)</td> <td>Party TV adverts</td> </tr> <tr> <td>List of candidates</td> <td>Press/media centre</td> </tr> <tr> <td>Policy matching tool</td> <td>National events</td> </tr> <tr> <td>Opponent dismissal in specific section</td> <td>Local events</td> </tr> <tr> <td>Issue/policy positions</td> <td>General election info (any kind)</td> </tr> </table>	E-news sign up	Schedule/calendar	RSS/WAP sign up	Archive reports, news, speeches	Downloads offline	Audio podcasts	Manifesto (any format)	Party TV adverts	List of candidates	Press/media centre	Policy matching tool	National events	Opponent dismissal in specific section	Local events	Issue/policy positions	General election info (any kind)	0-16
E-news sign up	Schedule/calendar																	
RSS/WAP sign up	Archive reports, news, speeches																	
Downloads offline	Audio podcasts																	
Manifesto (any format)	Party TV adverts																	
List of candidates	Press/media centre																	
Policy matching tool	National events																	
Opponent dismissal in specific section	Local events																	
Issue/policy positions	General election info (any kind)																	
Resource generation	Donation Membership Merchandise	0-3																
Feedback and interaction	Contact details – address/phone Contact details – email Contact – online form E-poll Discussion forum Chat room	0-6																



Web 1.0 to Web 2.0 or 1.5?

Web 1.0 era – static coding schemes worked well, sites were relatively ‘flat’ or two dimensional. Downward, upward, lateral communication flows

Web 2.0 era – more problematic to code

- More sites/platforms

- More interactivity – i.e. content changes more quickly

- 3 D – inter-operable, embedded in wider social media space, actions more diverse and linked to offline sphere.

Web 1.5? (Lilleker and Jackson, 2009)

- Schemes do incorporate interactivity and above features to an extent.

- Still focus on individual websites rather than wider web 2.0 presence

- A ‘one-off’ in the campaign so don’t capture dynamic nature of blogs, twitter, facebook profiles etc.

- Focus on online-specific forms of interactivity

- Conclude normalization and that parties/candidates still ‘monologic’ top-down, non-participatory

Source: Lilleker and Jackson (2010) *Internet Research* 20(5) 527-544

Feature	Party ^a	Candidate ^b
Contact details	92	65
Search engine	43	18
Enmeshing	48	15
Navigation aids	10	3
Questionnaires	5	0
Visitor's questionnaires	0	0
Polls	5	5
Visitor's polls	0	0
Petitions	19	3
Visitor's petitions	0	0
Video/picture sharing	71	15
RSS	38	13
Weblog	48	30
Twitter	48	10
Social network link	52	15
Podcasts	57	10
Ability for visitors to upload videos/files	33	25
Forum	21	10
Ability for visitors to share information	70	34
Ability for visitors to update information	38	25
Private conversations allowed	82	61
Public conversations allowed	57	38
Overall percentage of Web 2.0	38	18

Note: ^a Percentage of sites exhibiting feature ($n = 23$); ^b percentage of sites exhibiting feature ($n = 38$)

Table II.
Party and candidate use
of Web 2.0 applications

Source: Gulati and Williams (2007) Social Science Computer Review 25(4)

Table 8
Mobilization Tools on 2006 House Campaign Web Sites by Party

	All Candidates	Major-Party Candidates	Minor-Party Candid
Online volunteer form	69	81 ^a	17 ^a
Downloadable materials	18	18	16
Event form	16	19 ^a	2 ^a
Campaign materials and merchandise	15	16 ^a	10 ^a
Tell a friend	12	14 ^a	4 ^a
Letter to the editor	8	10 ^a	3 ^a
Foreign language content	4	4	3

a. The difference (χ^2) between major- and minor-party candidates is statistically significant at the .05 level

FEATURES OF ACTION AND MOBILIZATION	FREQUENCY	PERCENT	ALPHA
Denial of service attacks	1	1.2	.70
Online petitions	11	12.8	.98
Email campaigns	12	14	.82
Coordinates online actions	16	18.6	.69
Surface letter-writing campaigns	16	18.6	.87
Plans international actions	20	23.3	.86
Urgent action alerts	23	26.7	.87
Plans local actions	27	31.4	1.00
Plans national actions	36	41.9	1.00
Calendar of events	37	43	.77
Coordinates offline actions	51	59.3	.90
Project or campaign descriptions or news	71	82.6	.73

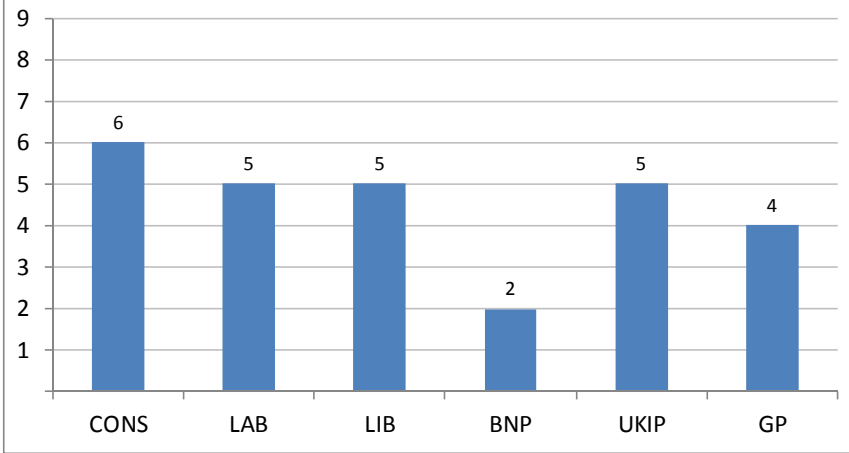
• Table 3 Interaction and dialog

FEATURES OF INTERACTION AND DIALOG	FREQUENCY	PERCENT	ALPHA
Discusses organizational issues online	1	1.2	.85
Holds meetings online	2	2.3	1.00
Weblogs	5	5.8	.71
Online polls and surveys	6	7	.78
Member or supporter profiles	7	8.1	.64
Member or supporter contact information	10	11.6	.83
First-person accounts of movement activities	17	19.8	.81
Discusses strategies online	17	19.8	.82
Participatory forums	18	20.9	.85
Support services or advice	29	33.7	.79
Organization contact information	81	94.2	.97

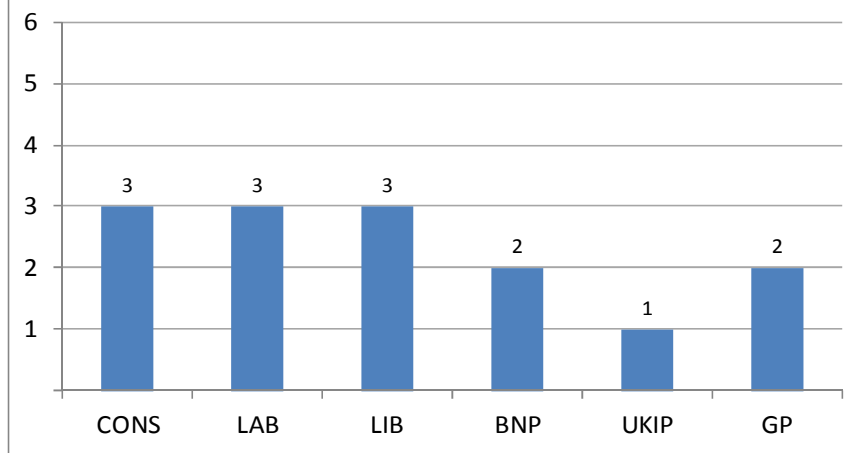
Web 2.0: UK parties General Election 2010

Feature	Items	Index
Web 2.0 interoperability	Links to: SNS Twitter Blog official party/leader, external Blog unofficial activists/members	YouTube Flickr Delicious Digg Campaign site
Online campaigning	Volunteer Register as a supporter Share/forward/embed Letter to editor Downloads apps smartphone Uploads: content creation	0-6

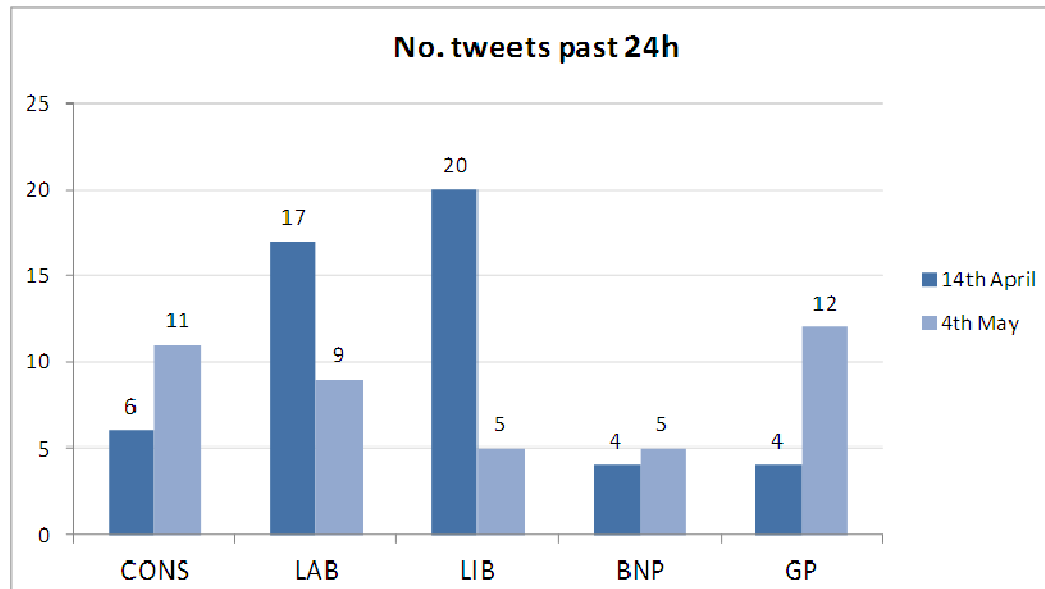
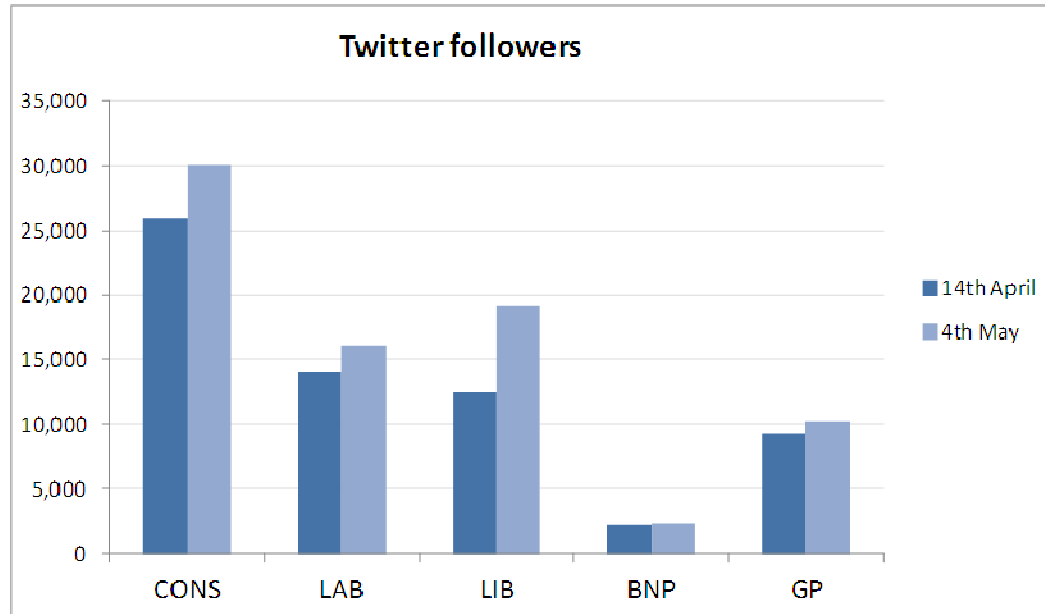
Web 2.0 Inter-operability



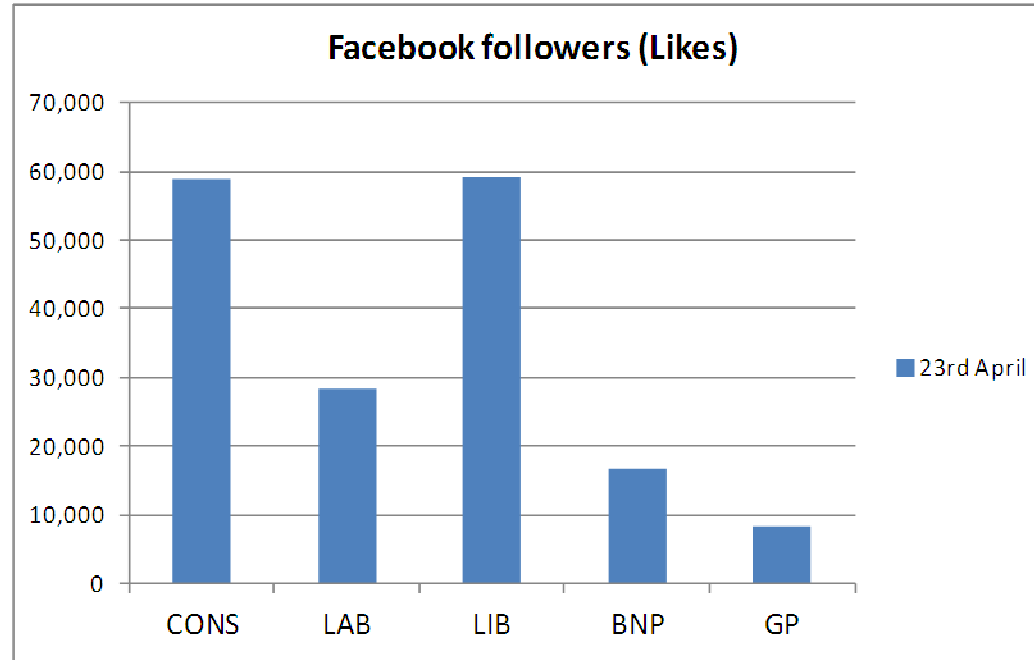
Online Campaigning



Twitter



Facebook



<i>23rd April</i>	CONS	LAB	GP
Posts by party past 24h	6	12	4
No. comments (total)	594	480	3
Average comments per post	99	40	0.75

YouTube

<i>26th April</i>	LAB	LIB	UKIP	GP
Date started	12/02/2007	31/10/2006	04/11/2008	31/07/2007
Channel views	278318	118016	49485	38821
Subscribers	3233	2751	1571	757
Channel comments	No comments	590	370	76
Videos past 24h	2	2	2	0

Evolution of Web Campaign Content Analysis:

New Challenge: Party and Candidate 'Web Action Centers'

New developments driven by parties to adapt to Web 2.0 environment

Activist Hubs for members/non-members

- Require registrations so cannot be archived
- Incorporate range of web 2.0-like features – blogs, email, updates, groups, events
- Content changing not just daily but hourly
- Focus is on promoting offline rather than online forms of campaign activism
- Conclude normalization but not that parties are 'monologic'



Today, Michelle and I join every American in sending our thoughts and prayers to Elizabeth and John and the entire Edwards family. We all admire Elizabeth's strength and determination and the deep love they so obviously share.

~Barack Obama

**HOPE.
ACTION.
CHANGE.**

Community Kickoff
March 31, 2007

[Learn More >](#)

[My.BarackObama.com](#)

This Campaign is About You.



Create Your Own Profile



Network With Your Friends



Find Supporters Near You



Become A Fundraiser



Plan and Attend Events



Write Your Own Blog



BARACK TV



ANNOUNCEMENT
Watch Senator Obama make his presidential announcement to 15,000+ people in Springfield, IL.



CONCORD, NH
Get the word on the street from New Hampshire natives about Barack Obama.



OBAMA IN AMES, IOWA
See footage from Ames on Senator Obama's first swing through Iowa.

Watch BARACK TV

In the News

- ▶ **Obama pushes Iraq pullout**
Portsmouth Herald
- ▶ **Obama says he's always opposed the war**
Sioux City Journal
- ▶ **Obama Draws Enthusiastic Crowd in Okla.**
NY Times



Click here to support Senator Obama's Iraq De-Escalation Act

Supporter Count: 50,043



Obama Store

Upcoming Events

- ▶ **Mar. 31 | Hope.Action.Change. Community Kickoff**



Click here for photos from the road...

membersnet is a tool that helps Labour members and supporters to organise, share and connect

Login



Remember me
[Forgotten Your Details?](#)



Keep in touch

Membersnet helps you communicate with friends & other members.



Organise & promote your events

Membersnet lets you organise events simply and easily and puts you in touch with like-minded people.



Social networking from one place

If you already use Facebook or Twitter why not link them with membersnet?

Upcoming Events what's this ?

Campaign Groups what's this ?

[View more groups](#)

ACT

The Lib Dem Network

now for change that works for you

BUILDING A FAIRER BRITAIN

Home My Page Groups Videos Photos Huddle

Take Action



DONATE



ORGANISE



VOLUNTEER



WHAT IS ACT?

Latest Activity



David Harris joined Jennifer Speller's group

Let's get AV

13 hours ago



Leon Keith Greenwell added 2 discussions to the group Totally dismantle prohibitionism NOW!

Slaughter of the Army of God -

Liberal Democrat fairer votes

Yes!

Join the Lib Dems' campaign for fairer votes

[Find out more >](#)

Welcome to Lib Dem Act

Sign Up
or Sign In

Twitter Feed

libdems: Photos from today's Liberal Democrat Yes! to Fairer Votes launch in Manchester <http://j.mp/ib9zX9> #LibDemYes

libdems: Today @TimFarron launched the Lib Dem Yes! to Fairer Votes campaign – find out what you can do to help <http://j.mp/eGNdYp> #LibDemYes

libdems: RT @libdemvoice: Paul Burstow writes... Labour shouldn't play party politics with the NHS <http://dv.org.uk/23459>

New Groups



Inequality Policy Workin...

15 members



ENGAGE - The Lib Dem Pol...

198 members



Say No to Trident

262 members



Let's get AV

175 members



Bring Back Sue

28 members

[View All](#)

New Videos

5 Functions of Web Action Sites

- Community building
- Internal resource mobilization
- External voter mobilization (gotv)
- Message production
- Message distribution

Measuring 5 Functions of Web Action Sites

Community Building

- Setup profile
- Photo
- Biography
- Why joined
- Setup/join Groups
- Setup Blog
- Setup Wiki
- Email/msg system
- Externally promote profile

Subtotal (additive 0-9)

Internal Mobilization

- Personal fundraising
- Promote membership
- Sign up as local organiser
- Sign up as candidate
- Organize / add event
- Vote leaders to attend events

Subtotal (additive 0-6)

External Mobilization

GOTV offline

- Access phonebank
- Sign up for f2f canvassing
- Sign up to discuss with f&f
- Leaflets download
- Externally promote event

GOTV online

- Send email
- Post to FaceBook
- Post to Twitter
- GOTV i-phone app

Elite Mobilization

- Email forward to MP/newspp

- Start e-petition

Subtotal (additive 0-11)

Message create

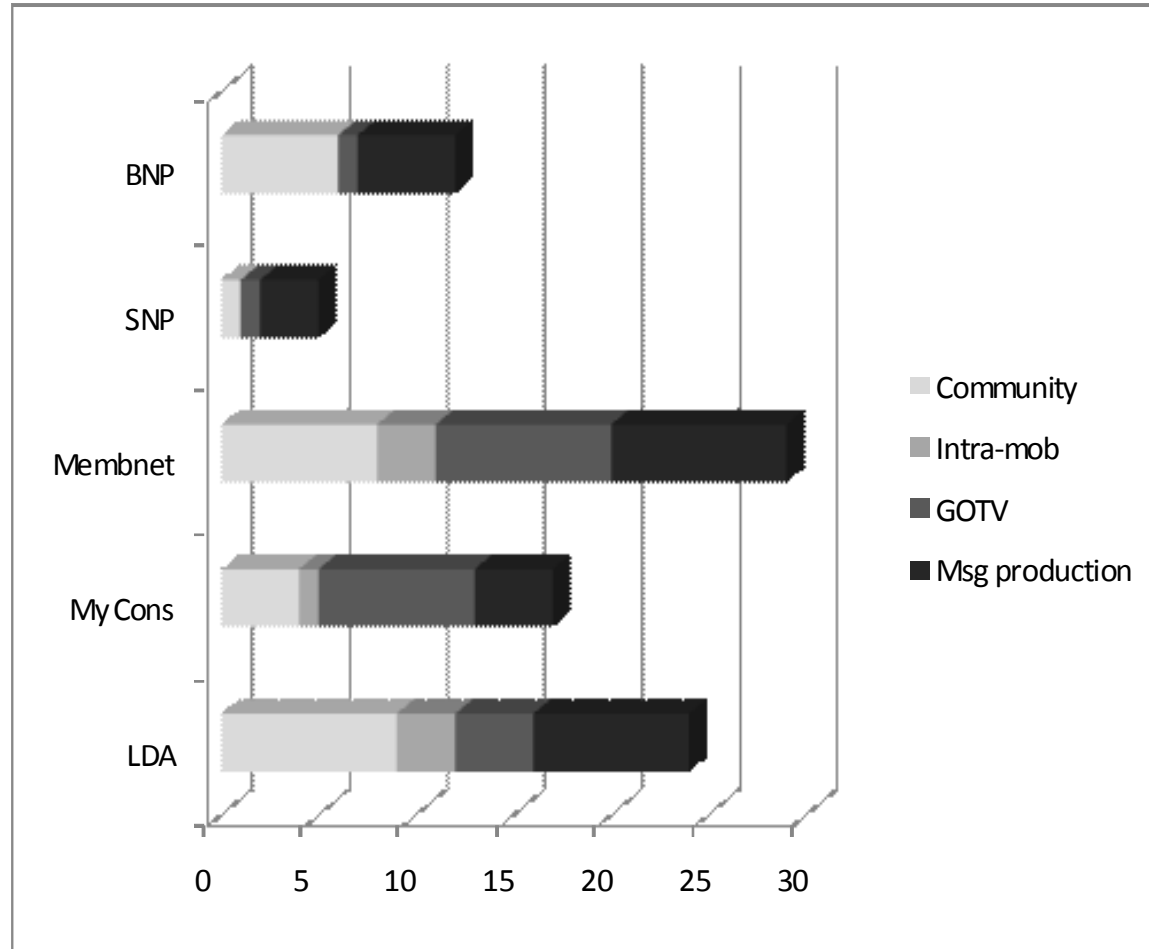
- Policy email fwd/customize
- Poster/leaflet reate/customize
- Manifesto input/feedback

Message distribute

- Web banners/ads d-load
- Posters/leaflets d-load
- Email/share policy docs
- RSS feed to website
- Share blog posts ext.
- Link to SNS profile
- Link to Twitter account
- Import email contacts

Subtotal (additive 0-11)

Figure 1: Citizen-initiated Campaigning Emphasis by Party



Conclusions and Future Directions

- Automating site data collection and analysis – following the Bauer and Scharl model?
 - Comparative analysis – Large N but does this really capture the full content?
- Standardized measures applied – benchmarking
- Archiving challenges – dynamic quality
- Mobile applications
- Online to offline
- Standardization of tools to capture and analyze social media use – Twitter, YouTube, Facebook