From directed interactions in social media to insights about human behaviour

Andreas Kaltenbrunner

Social Media Research Group, Barcelona Media, Barcelona, Spain

Nadine Meeting, June 14th, 2013



Outline



- Political Affiliation on Wikipedia
- Propagation of emotional styles on Wikipedia
- Friendship and online discussion
- 5 Geographical distance and Friendship
- 6 Links between biographies on Wikipedia



Outline



- 2 Political Affiliation on Wikipedia
- Propagation of emotional styles on Wikipedia
- 4 Friendship and online discussion
- 5 Geographical distance and Friendship
- 6 Links between biographies on Wikipedia



Analysis of the Spanish General Elections of 2011

Introduction

Research Questions

- Do political parties interact on Twitter?
- Do political parties use Twitter to engage in conversations or as one-way flow broadcast medium?
- Are there differences between the parties?

Dataset collected between Nov 4 and 24, 2011

- ~ 3 million tweets.
- ~ 380.000 users.

Results will be published in

P. Aragón, K. Kappler, A. Kaltenbrunner, D. Laniado and Y. Volkovich. Communication Dynamics in Twitter During Political Campaigns: The Case of the 2011 Spanish National Election,

Policy & Internet, 5 (2), 2013.

Retweets

Users almost exclusively propagated contents from members of their own party



@akalten bci

Replies



Some amount of communication also among members of

• PP - PSOE

• IU - UPyD - EQUO

ERC - CiU

Andreas Kaltenbrunner

Insights about human behaviour from directed interactions

Conclusions and Future Research

Conclusions

- Retweets: Balkanisation of Spain's (online) political sphere
- Replies: Inter-party communication happens but most of the interactions still occur within the parties.
- Political parties use Twitter as a one-way flow broadcast.
 - Low number of replies by candidate and party profiles
 - Low ratio between sent and received replies.
- New and minor parties tend to be more clustered and better connected ⇒ a more cohesive community.

Future Research

 In-depth analysis of the topological patterns of party networks to characterize the different party apparatus (centralised, decentralised, or distributed).

Outline



- Political Affiliation on Wikipedia
- Propagation of emotional styles on Wikipedia
- 4 Friendship and online discussion
- 5 Geographical distance and Friendship
- 6 Links between biographies on Wikipedia



Motivation

Does political polarisation also take place in Wikipedia?

• Obtain a deeper understanding of online interaction and collaboration among members of distinct political parties.

Research questions

- Do political users in Wikipedia exhibit a preference for interacting with members of their same political party?
- Do we see a division in patterns of participation along party lines?

Results published in

J. J. Neff, D. Laniado, K. E. Kappler, Y. Volkovich, P. Aragón & A. Kaltenbrunner. Jointly They Edit: Examining the Impact of Community Identification on Political Interaction in Wikipedia.

PLoS ONE, vol. 8, no. 4, page e60584, 2013.

Introduction Wikipedia visible side



Article	Discussion	Read	Edit	View history	Search

The Presidency of Barack Obama began at noon EST on January 20, 2009 when he became the 44th President of the United States. Obama was a United States Senator from

Illinois at the time of his victory over Arizona Senator John McCain in the 2008 presidential election. Obama became the first African-American president of the United States.

Presidency of Barack Obama

See also: Timeline of the Presidency of Barack Obama

The Free Encyclopedia

Main page Contents Featured content Current events Random article Donate to Wikipedia

 Interaction Help About Wikipedia Community portal Recent changes Contact Wikipedia
 Toolbox
 Print/export

Languages
 Česky
 Español
 Français
 Èнәоl
 Kiswahili
 Norsk (bokmål)
 Русский
 Svenska
 Tinu

Contents [hide]	
1 Transition period	
2 Inauguration	
3 First 100 days	
3.1 Expectations	
3.2 Legislation and executive orders	
Approval ratings and opinion	
4.1 2009	
4.1.1 Handling of the economy	
4.2 2010	
Major legislation	
5.1 Legislation signed	
5.1.1 2009	
5.1.2 2010	
5.1.3 2011	
Personnel	
6.1 Cabinet appointees	
6.2 Notable non-Cabinet positions	
6.3 Judicial nominees	
6.3.1 Supreme Court	
6.3.2 Other courts	
Policies	
7.1 Economy	
7.2 Etnics	
7.2.1 Lobbying reform	
7.2.2 transparency	
7.3 Foreign policy	
7.5.1 Guantanamo Bay detention camp	
7.3.2 Overseas Contingency Operation	
7.4 Gun control	
7.5 Science and technology	

Presidency of Barack Obama

& Log in / create account



Andreas Kaltenbrunner

7.5.3 NASA

Introduction Article talk pages



@akalten bcn

Insights about human behaviour from directed interactions

Example Structure

Discussion tree for article "Presidency of Barack Obama"



- red \rightarrow root (the article)
- blue \rightarrow structural nodes
- green → anonymous comments
- grey \rightarrow registered comments

More details in:

D. Laniado, R. Tasso, Y. Volkovich, and A. Kaltenbrunner.

When the Wikipedians talk: Network and tree structure of Wikipedia discussion pages.

In Proc. of ICWSM, 2011.

Interactions of partisan users on article talk pages

User-boxes \Rightarrow Party assign.

Democrats

dem This user supports the U.S. Democratic Party.

Republicans



This user supports the Republican Party of the United States.

Cross-party interactions

- Shuffle test indicates neutral mixing.
- ⇒ no stat. significant preference for neither internor intra-party interaction.



Insights about human behaviour from directed interactions

Top 10 Articles per number of distinct editors

Partisan users are more interested in politics

#	Democrats	Republicans	All Users
1.	Barack Obama	George W. Bush	George W. Bush
2.	Unites States presidential election, 2008	Unites States presidential election, 2008	Wikipedia
3.	George W. Bush	United States	United States
4.	Unites States	Republican Party (United States)	Barack Obama
5.	Bill Clinton	John McCain	Adolf Hitler
6.	Democratic party (United States)	Barack Obama	Michael Jackson
7.	Wikipedia	Wikipedia	Britney Spears
8.	Britney Spears	Ronald Reagan	Jesus
9.	Hillary Rodham Clinton	Virginia Tech Massacre	World War II
10.	Al Gore	Adolf Hitler	PlayStation 3

Articles in bold deal with political topics.

Conclusions

- Wikipedian identity seems to predominate over party identity.
- Despite the increasing political division of the U.S., there are still areas in which political dialogue is possible and happens.



Outline

- Political User interaction on Twitter
- Political Affiliation on Wikipedia
- Propagation of emotional styles on Wikipedia
- 4 Friendship and online discussion
- 5 Geographical distance and Friendship
- Links between biographies on Wikipedia



Analysis of emotions expressed in talk pages

Goal:

Study the emotional dimension in a large peer production community

Research questions: How are the emotional styles of editors ...

- affected by their level of experience?
- affected by their gender and the topics they choose to work on?
- affected by interacting with others (emotional congruence)?
- related to those of the editors they interact more frequently with (emotional homophily)?

Results are partly published in

Laniado, D., Castillo, C., Kaltenbrunner, A., and Fuster Morell, M. F. (2012) Emotions and dialogue in a peer-production community: the case of Wikipedia. 8th International Symposium on Wikis and Open Collaboration, WikiSym'12

User gender labelling

- $\approx 12\ 000$ users wrote ≥ 100 comments in articles talk pages
- Gender identified through Wikipedia API for $pprox 2\ 000$ of them
- Out of the remaining ones, a sample of 1 385 users for manual labelling through crowd-sourcing (Crowdflower)



Measuring the Emotional Content of Discussions

Example for the results of different emotional lexica

Andre

Table: Example messages with their corresponding Valence, Arousal, and Dominance(ANEW) or positive & negative scores (LIWC, SentiStrength).

	ANEW			LIV	LIWC Se		entiSt.	
	V	А	D	+	-	+	-	
Sounds like a good challenge - to be proven or disproven. I'm happy if it can be shown to go further using closed cubic polynomial solu- tions. The nice thing about these are that they are pretty easy to test numerically -in "Exact trigonometric constants"	7.4	5.3	6.2	15	0	3	-2	
Seems you have not yet seen female lover after having sex who do not wish to have sex with the same lover any more :) Once you've seen it, you understand very well what war of Venus means compared to war of Mars. -in <i>"House (astrology)"</i>	5.5	7.0	5.2	6.8	4.5	4	-3	
What about the whirlie hazing, the alcohol abuse , the emotional poverty , the suicide in 1995/6, the biotech plans which were stopped by pitzer protests	1.6	5.8	3.5	4	8	1	-4	
-in "Harvey Mudd College"			⇒ ∢	6 ≻ ∢	$\Xi \succ i$	(国)	-≣ - ≮	
as Kaltenbrunner @akalten bcn Insights about human be	ehaviour	from d	irected	l interac	tions			

Emotions, Status and Gender

Similar results with different Lexica

Emotions and Status

- Admins express, on average, more positive emotion (p < 0.001).
- Admins also express less negative emotion (p < 0.001).
- Non-admins express more affect, in particular, more anxiety, anger and sadness (all with *p* < 0.001) compared to admins.

Emotions, Status and Gender

- Significant difference between male admins and non-admins (*p* < 0.001)
- No significant difference between female admins and non-admins.



Emotions and gender



- ANEW Words more used by females and males
- Size accounts for difference in frequency



Topics, emotions and gender



Figure: Mean valence for discussions of articles in different topic categories, 2 vs the proportion of comments written by male editors

Replies are more positive

On average, editors tend to reply with:

- higher valence: +0.05 (*p* < 0.01)
- higher dominance: $+0.04 \ (p < 0.01)$
- no statistically significant differences for arousal

• Users tend to be more positive and dominant when replying, but without recurring to words evoking stronger sentiments.



Emotional homophily

Mixing patterns: do users interact preferentially with similar users?

• Assortative by emotional style: users interact more with others expressing similar emotions.



- edges connecting users who have exchanged at least 10 replies
- red nodes $\rightarrow 15\%$ users expressing higher valence in article discussions
- black nodes $\rightarrow 15\%$ users expressing lower valence in article discussions
- size → proportional to the number of connections



Outline

- Political User interaction on Twitter
- Political Affiliation on Wikipedia
- Propagation of emotional styles on Wikipedia
- Friendship and online discussion
- 5 Geographical distance and Friendship
- Links between biographies on Wikipedia



Contrast of two directed social networks

- explicit (articulated) friendship network
- implicit (behavioural) interactions via comments (reply network)

Data from Spanish Digg-like social news website meneame.net

- 33,194 users in reply network
- 8,673 users in friendship network

Results published in

A. Kaltenbrunner, G Gonzalez, R Ruiz de Querol and Y. Volkovich.

Comparative analysis of articulated and behavioural social networks in a social news sharing website,

New Review of Hypermedia and Multimedia, 17 (3), 243-266, 2011.

Main page



Example for comments about and article





Interactions in discussions

Types of interactions

- only friendship link is present
- interaction in both networks
- only reply interaction

Measure reply reciprocity

$$\rho_{\text{reply}}(\text{user } A, \text{ user } B) =$$

 $\log \frac{|\text{replies of } A \text{ to } B| + 1}{|\text{replies of } B \text{ to } A| + 1}$

- symmetrical around 0.
- \(\rho_{\mathbf{reply}} = 0\) if interactions
 are balanced

Interaction Network



Andreas Kaltenbrunner

Combined analysis of reply and friendship networks

Results

- Relatively high probability of replies from a user to his/her friends.
- but the number of reply interactions between these users is relatively small.
- Users who interact repeatedly with others are mostly non-friends.

Reciprocity of replies

- Non-friends who reply to each other are more likely to have similar number of replies.
- → The social norm of reciprocity [Gouldner 1960] suggests a negative connotation of replies between non-friends.
- Replies between friends may be more neutral and do not provoke replies to the same extend.

Outline

- Political User interaction on Twitter
- Political Affiliation on Wikipedia
- Propagation of emotional styles on Wikipedia
- Friendship and online discussion
- 5 Geographical distance and Friendship
- Links between biographies on Wikipedia



online tools and long-distance travel \Rightarrow *death of distance*?

- individuals try to minimise the efforts to maintain a friendship by interacting more with their spatial neighbours.
- probability of a social interaction quickly decays as an inverse power of the relative geographic distance [Stewart 1941].
- probability of connections between two individuals on online social networking services still decreases with their geographic distance [Backstrom 2010, Liben-Nowell 2005].

Results published in



A. Kaltenbrunner, S. Scellato, Y. Volkovich, D. Laniado, D. Currie, E. J. Jutemar & C. Mascolo. Far from the eyes, close on the Web: impact of geographic distance on online social interactions. In Proceedings of ACM SIGCOMM Workshop on Online Social Networks (WOSN 12), ACM, 2012

Y. Volkovich, S. Scellato, D. Laniado, C. Mascolo & A. Kaltenbrunner.

The length of bridge ties: structural and geographic properties of online social interactions. In ICWSM-12 - 6th International AAAI Conference on Weblogs and Social Media. The AAAI Press, 2012.

Dataset from Tuenti

"Spanish Facebook", a Spain-based social networking website

- ~ 10 million users
- 1174 million friendship links
- ~500 directed messages exchanges during 3 months;



Andreas Kaltenbrunner

Geographic properties

The effect of distance on friendship

Probability of connection as function of geographic distance



Interaction Analysis

Interactions and distance

Probability of message exchange between friends



 high-intensity communication takes place on social connections regardless of their geographic distance

Andreas Kaltenbrunner @akalten_bcn Insights about human behaviour from directed interactions

Conclusions

The effect of geographic distance on online social interactions

- Spatial proximity greatly affects how users establish their connections on online social platforms.
- Social interactions are only weakly affected by distance.
- Geography affects whom we interact with, however it does not influence how much we interact.

Applications

- Ink prediction,
- tie strength modelling,
- user profiling.



Andreas Kaltenbrunner

@akalten bc

Outline

- Political User interaction on Twitter
- Political Affiliation on Wikipedia
- Propagation of emotional styles on Wikipedia
- Friendship and online discussion
- 5 Geographical distance and Friendship
- 6 Links between biographies on Wikipedia



Motivation

Wikipedia as global collective memory place allows ...

- to extract from biographies how social links are recorded ...
- to generate networks of links between biographical articles.

Research questions

- Who are the most central characters in these networks?
- Do culture related peculiarities exist?
- Which cultures are more similar?
- What is the shared knowledge about connections between persons across cultures?

Results published in

P. Aragón, A. Kaltenbrunner, D. Laniado & Y. Volkovich. Biographical Social Networks on Wikipedia - A cross-cultural study of links that made history. In Proc. of the 8th Int. Symp. on Wikis and Open Collaboration (WikiSym'12). ACM, 201

- Selected the 15 largest language editions of Wikipedias
- Starting point: 296 511 biographies from the English Wikipedia (from DBpedia)
- Identified the corresponding articles (when existing) on the remaining 14 languages
- Generated a directed network for each language version:
 - $\bullet \ nodes \rightarrow persons$
 - $\bullet~$ edges \rightarrow links between the articles of the corresponding persons
- Manage alternative titles of articles: track redirects
- Data collected through Wikipedia APIs between September 8th and 13th, 2011



Most central persons in the English Wikipedia

sorted by in-degree. Ranks for out-degree, betweenness and PageRank in parenthesis

person	in-degree	out-	degree	betw.	PageRa	ank
George W. Bush	2123	89	(107)	(1)	0.00209	(1)
Barack Obama	1677	51	(710)	(8)	0.00162	(2)
Bill Clinton	1660	74	(205)	(4)	0.00156	(4)
Ronald Reagan	1652	90	(103)	(2)	0.00156	(3)
Adolf Hitler	1407	119	(26)	(3)	0.00149	(5)
Richard Nixon	1299	86	(127)	(7)	0.00136	(6)
William Shakespeare	1229	25	(4203)	(63)	0.00113	(9)
John F. Kennedy	1208	104	(53)	(5)	0.00123	(8)
Franklin D. Roosevelt	1052	71	(237)	(15)	0.00131	(7)
Lyndon B. Johnson	1000	106	(50)	(12)	0.00108	(11)
Jimmy Carter	953	80	(158)	(9)	0.00113	(10)
Elvis Presley	948	82	(142)	(27)	0.00063	(24)
Pope John Paul II	941	59	(444)	(11)	0.00083	(18)
Dwight D. Eisenhower	891	55	(564)	(22)	0.00095	(14)
Frank Sinatra	882	108	(47)	(18)	0.00056	(28)
George H. W. Bush	878	87	(118)	(19)	0.00096	(13)
Abraham Lincoln	846	54	(593)	(40)	0.00089	(16)
Bob Dylan	835	151	(11)	(14)	0.00055	(30)
Winston Churchill	748	84	(136)	(10)	0.00092	(15)
Harry S. Truman	743	81	(145)	(24)	0.00099	(12)
Joseph Stalin	723	69	(265)	(43)	0.00089	(17)
Michael Jackson	663	71	(237)	(34)	0.00042	(51)
Elizabeth II	653	52	(665)	(6)	0.00074	(19)
Jesus	572	38	(1595)	(51)	0.00068	(20)
Hillary Rodham Clinton	554	87	(118)	(32)	0.00063	(25)



Most central persons in different language Wikipedias

Top 5 most central persons for each language by betweenness

lang	#1	#2	#3	#4	#5
en	George W. Bush	Ronald Reagan	Adolf Hitler	Bill Clinton	John F. Kennedy
de	Adolf Hitler	George W. Bush	Martin Luther King, Jr	Barack Obama	Frank Sinatra
fr	Adolf Hitler	George W. Bush	William Shakespeare	Barack Obama	Jacques Chirac
it	Frank Sinatra	George W. Bush	Pope John Paul II	Michael Jackson	Elton John
es	Michael Jackson	Fidel Castro	William Shakespeare	Che Guevara	Adolf Hitler
ja	Adolf Hitler	Michael Jackson	Ronald Reagan	Yukio Mishima	Barack Obama
nl	Elvis Presley	Adolf Hitler	Bill Clinton	Joseph Stalin	William Shakespeare
pt	Michael Jackson	Richard Wagner	Adolf Hitler	Ronald Reagan	David Bowie
sv	George W. Bush	Winston Churchill	Elizabeth II	Michael Jackson	Adolf Hitler
pl	Elizabeth II	Pope John Paul II	Margaret Thatcher	George W. Bush	Ronald Reagan
fi	Barack Obama	Adolf Hitler	Michael Jackson	George W. Bush	Benito Mussolini
no	Marilyn Monroe	Adolf Hitler	John F. Kennedy	Bob Dylan	Bill Clinton
ru	William Shakespeare	Napoleon II	Kenneth Branagh	Elton John	Joseph Stalin
zh	Chiang Kai-Shek	William Shakespeare	Barack Obama	Deng Xiaoping	Adolf Hitler
ca	Adolf Hitler	Che Guevara	Juan Carlos I	Michael Schumacher	Juan Manuel Fangio

Most are known to be (or have been) highly influential

- We find political leaders, revolutionaries, famous musicians, writers and actors.
- Hitler, Bush, Obama dominate in almost all top rankings.
- Top ranked in many languages reflect country peculiarities.

Languages similarity network

Every language links to the two most similar ones according to Jaccard coefficient



Definition of Jaccard coefficient J

• Given the set of links A and B of two networks

$$J = \frac{|A \cap B|}{|A \cup B|}$$

• *J* is the ratio between the number of links present in both networks (their intersection) and the number of links existing in their union.

Intersection of networks in different languages



Andreas Kaltenbrunner

@akalten bc

Insights about human behaviour from directed interactions

Conclusions and future work

Conclusions

- Global social network measures are largely similar for all networks.
- Most central persons unveil interesting peculiarities about the language communities.
- Networks are more similar for geographically or linguistically closer communities.
- Many connections which can be found in most of the analysed language Wikipedias.

Future work

- Application of the methodology to generate subnetworks of other kinds of article categories
- Consider all biographies for each language.
- Analyse links missing only in a few language Wikipedias.

Questions?



We are hiring

• Marie Curie fellow co-funded by the Catalan regional government.

 More info http://www.catalonia.com/en/technology/ tecniospring/call-for-proposals-2013.jsp

Bibliography I



P. Aragón, A. Kaltenbrunner, D. Laniado & Y. Volkovich.

Biographical Social Networks on Wikipedia - A cross-cultural study of links that made history. In Proceedings of the 8th International Symposium on Wikis and Open Collaboration (WikiSym'12). ACM, 2012.



Communication Dynamics in Twitter During Political Campaigns: The Case of the 2011 Spanish National Election. Policy & Internet, vol. 5, no. 2, 2013. In press



Lars Backstrom, Eric Sun & Cameron Marlow.

Find me if you can: improving geographical prediction with social and spatial proximity. In Proceedings of WWW 2010, Raleigh, North Carolina, USA, 2010.



A.W. Gouldner.

The norm of reciprocity: A preliminary statement. American Sociological Review, vol. 25, no. 2, pages 161–178, 1960.



A. Kaltenbrunner, G. Gonzalez, R. Ruiz de Querol & Y. Volkovich.

Comparative analysis of articulated and behavioural social networks in a social news sharing website. New Review of Hypermedia and Multimedia, vol. 17, no. 3, pages 243–266, 2011.



A. Kaltenbrunner, P. Aragón, D. Laniado & Y. Volkovich.

Not all paths lead to Rome: Analysing the network of sister cities. In Proceedings of the 7th International Workshop on Self-Organizing Systems (IWSOS 2013). Springer, 2013



D. Laniado, C. Castillo, A. Kaltenbrunner & M. Fuster-Morell.

Emotions and dialogue in a peer-production community: the case of Wikipedia. In Proceedings of the 8th International Symposium on Wikis and Open Collaboration (WikiSym'12). ACM, 2012.



Bibliography II



David Liben-Nowell, Jasmine Novak, Ravi Kumar, Prabhakar Raghavan & Andrew Tomkins.

Geographic routing in social networks.

PNAS, vol. 102, no. 33, pages 11623-11628, August 2005.



Jointly They Edit: Examining the Impact of Community Identification on Political Interaction in Wikipedia. PLoS ONE, vol. 8, no. 4, page e60584, 2013.



John Q. Stewart.

An Inverse Distance Variation for Certain Social Influences. Science, vol. 93, no. 2404, pages 89–90, 1941.



Y. Volkovich, S. Scellato, D. Laniado, C. Mascolo & A. Kaltenbrunner.

The length of bridge ties: structural and geographic properties of online social interactions. In ICWSM-12 - 6th International AAAI Conference on Weblogs and Social Media. The AAAI Press, 2012.

